Università degli Studi della Basilicata

Dottorato di Ricerca in
"Cities and Landscapes: Architecture, Archaeology, Cultural Heritage, History and Resources"

Titolo Progetto
"Smart Energetic Manufacturing"

TITOLO DELLA TESI
"**Innovative thermodynamic hybrid model-based and data-driven techniques for real time manufacturing sustainability assessment**"

Settore Scientifico-
Disciplinare"ING-IND/11"

Coordinatore del Dottorato

Prof. Ing. FIORENTINO Mauro

Dottorando

Dott. Ing. SELICATI Valeria

Relatore

Prof. Ing. CARDINALE Nicola

Correlatore

Prof. Ing. DASSISTI Michele

Ciclo XXXIV

Sintesi della tesi (ITALIANO)

La presente tesi di dottorato è stata realizzata grazie alla supervisione e collaborazione tra Università degli Studi della Basilicata, Politecnico di Bari e l'azienda Master Italy s.r.l. (Conversano, Italia).

I principali filoni di ricerca approfonditi e discussi nella tesi sono: la sostenibilità in generale e, più specificatamente, quella manifatturiera, il paradigma dell'Industria 4.0 legato alla smart (green) manufacturing, le tecniche di valutazione dei processi manifatturieri basate sui modelli (model-based) e le tecniche di valutazione derivate dall'analisi dei dati (data-driven). Nella tesi, queste tematiche apparentemente a sé stanti sono sviluppate in modo tale da dimostrare quanto siano fortemente interconnesse e caratterizzate da trasversalità.

Lo scopo del programma di dottorato è stato quello di implementare e convalidare i modelli di valutazione innovativi per esaminare la natura dei processi di produzione e razionalizzare le relazioni e le correlazioni tra le varie fasi del processo. Questo modello composito può essere impiegato come strumento nel processo decisionale politico sullo sviluppo sostenibile dei processi industriali e sul miglioramento continuo dei processi manifatturieri. L'obiettivo generale di questo lavoro di ricerca è proporre tecniche basate su modelli ibridi termodinamici del primo e del secondo ordine e quelli basati sui dati e l'apprendimento automatico per il monitoraggio in tempo reale delle performance e della sostenibilità manifatturiera. Il modello proposto è testato su un caso studio industriale reale attraverso un approccio sistemico: le fasi di individuazione dei requirements, l'inventario dei dati (materiali, energetici, geometrici, fisici, economici, sociali, qualitativi, quantitativi), la modellazione, l'analisi, la regolazione ad hoc degli algoritmi (tuning), l'implementazione e la validazione, sono sviluppate per il processo di pressofusione delle leghe di alluminio di una PMI situata nel sud Italia, Master Italy s.r.l., la quale progetta e produce accessori e componenti metallici per gli infissi dal 1986.

La tesi affronta il tema della sostenibilità dei processi industriali intelligenti a 360 gradi, guardando sia alla quantità di risorse impiegate, sia alla qualità del loro utilizzo durante tutto il ciclo di vita del processo di produzione. Ai modelli di analisi tradizionali della sostenibilità (come l'analisi del ciclo di vita, LCA), vengono integrati metodi basati sul secondo principio della termodinamica (analisi exergetica); a questi vengono inoltre affiancati modelli basati sulla tecnologia dell'informazione (big-data analysis). Per ciascun metodo implementato singolarmente o in maniera integrata, viene presentata una dettagliata review che ne illustra il potenziale. Dopo una descrizione delle metriche utili a qualificare il grado di sostenibilità dei processi industriali, viene illustrato il caso studio con la modellazione e analisi dettagliata dei processi, in particolare quello della pressofusione delle leghe di alluminio. Dopo la valutazione della sostenibilità dei processi produttivi basata sull'approccio model-based si passa all'applicazione in tempo reale delle analisi basate sul machine learning in cui si punta all'identificazione di fermi e guasti durante il ciclo produttivo e alla possibilità di prevederne l'accadimento con giusto anticipo a partire dai valori dei parametri termodinamici di processo raccolti in tempo reale e all'apprendimento automatico. Infine, a dimostrazione della multidisciplinarietà e la trasversalità di tali tematiche, la tesi propone l'applicazione dei modelli integrati su alcuni casi studio quali i processi di deposizione laser e la riqualificazione del patrimonio edilizio esistente, sempre in chiave sostenibile.

Il lavoro di tesi presenta interessanti spunti derivanti dall'applicazione di un approccio ibrido alla valutazione di sostenibilità dei processi produttivi, combinando insieme analisi exergetica e valutazione del ciclo di vita. Il tema proposto è assolutamente attuale e pertinente agli sviluppi più recenti inerenti la sostenibilità in ambito industriale, coniugando approcci classici model-based con approcci innovativi basati sulla raccolta di big data e sulla loro analisi con le più adatte metodologie di machine learning La tesi presenta una applicazione molto promettente delle metodiche di machine learning a dati raccolti in tempo reale allo scopo di individuare eventuali problemi alla linea di produzione partendo da metriche di sostenibilità derivate dall'analisi exergetica e dall'analisi del ciclo di vita. Come tale, presenta indubbiamente un avanzamento rispetto alle conoscenze pregresse ed illustrate nello stato dell'arte introduttivo. Infatti, le aziende manifatturiere che ad oggi implementano strategie di business basati su modelli smart e tecnologie abilitanti hanno un valore maggiore sul mercato globale in termini di qualità, personalizzazione, flessibilità e sostenibilità.

Summary of the thesis (ENGLISH)

This doctoral thesis is the result of the supervision and collaboration of the University of Basilicata, the Polytechnic of Bari, and the enterprise Master Italy s.r.l.

The main research lines explored and discussed in the thesis are: sustainability in general and, more specifically, manufacturing sustainability, the Industry 4.0 paradigm linked to smart (green) manufacturing, model-based assessment techniques of manufacturing processes, and data-driven analysis methodologies. These seemingly unrelated topics are handled throughout the thesis in such a way that it reveal how strongly interwoven and characterised by transversality they are.

The goal of the PhD programme was to design and validate innovative assessment models in order to investigate the nature of manufacturing processes and rationalize the relationships and correlations between the different stages of the process. This composite model may be utilized as a tool in political decision-making about the long-term development of industrial processes and the continuous improvement of manufacturing processes. The overarching goal of this research is to provide strategies for real-time monitoring of manufacturing performance and sustainability based on hybrid thermodynamic models of the first and second order, as well as those based on data and machine learning. The proposed model is tested on a real industrial case study using a systemic approach: the phases of identifying the requirements, data inventory (materials, energetic, geometric, physical, economic, social, qualitative, quantitative), modelling, analysis, ad hoc algorithm adjustment (tuning), implementation, and validation are developed for the aluminium alloy die-casting processes of Master Italy s.r.l., a southern Italian SME which designs and produces the accessories and metal components for windows since 1986.

The thesis digs in the topic of the sustainability of smart industrial processes from each and every perspective, including both the quantity and quality of resources used throughout the manufacturing process's life cycle. Traditional sustainability analysis models (such as life cycle analysis, LCA) are combined with approaches based on the second law of thermodynamics (exergetic analysis); they are then complemented by models based on information technology (big-data analysis). A full analysis of the potential of each strategy, whether executed alone or in combination, is provided. Following a summary of the metrics relevant for determining the degree of sustainability of industrial processes, the case study is demonstrated using modelling and extensive analysis of the processes, namely aluminium alloy die casting. After assessing the sustainability of production processes using a model-based approach, we move on to the real-time application of machine learning analyses with the goal of identifying downtime and failures during the production cycle and predicting their occurrence well in advance using real-time process thermodynamic parameter values and automatic learning. Finally, the thesis suggests the use of integrated models on various case studies, such as laser deposition processes and the renovation of existing buildings, to demonstrate the multidisciplinarity and transversality of these issues.

The thesis reveals fascinating findings derived from the use of a hybrid method to assessing the sustainability of manufacturing processes, combining exergetic analysis with life cycle assessment. The proposed theme is completely current and relevant to the most recent developments in the field of industrial sustainability, combining traditional model-based approaches with innovative approaches based on the collection of big data and its analysis using the most appropriate machine learning methodologies. Furthermore, the thesis demonstrates a highly promising application of machine learning approaches to real-time data collected in order to identify any fault source in the manufacturing line beginning with sustainability measures generated from exergetic analysis and life cycle analysis. As such, it unquestionably represents an advancement above earlier information depicted in the initial state of the art. In actuality, manufacturing companies that implement business strategies based on smart models and key enabling technologies today have a higher market value in terms of quality, customisation, flexibility, and sustainability.

CONTENTS

Nomenclature/Glossary

Symbols

A	area	[m^2]
c	specific heat	[J/kg K]
C1	first phase injection course	[mm]
C2	second phase injection course	[mm]
CC	third phase injection course	[mm]
d	diameter	[mm]
d	Euclidean distance for k-means	
DI	nondimensional Dunn index	
∂	nondimensional Carnot factor	
$\bar{e}$	specific exergy	[J/kg]
E	entropy in decision trees	
Ex	exergy	[J]
$\dot{E}x$	exergy flow	[W]
f	function	
FC	clamping force	[kN]
FR	nondimensional renewability factor	
ϕ	feature mapping	
$\phi(x)$	sigmoid activation function	
GHG	Greenhouse Gas emission	[CO_2eq]
GWP	Global Warming Potential	[kgCO_2eq]
h	optimal margin classifier	
$\bar{h}$	specific enthalpy	[J/kg]
$\dot{H}$	enthalpy flow rate	[W]
IF	nondimensional Impact Factor	
K	gaussian kernel	
L	hinge loss	
λ	thermal conductivity	[W/m K]
$\dot{m}$	mass flow rate	[kg/s]
P	probability	
PF	final pression	[bar]
PM	multiplication pressure	[bar]
PS	specific pressure	[bar]
Π	extropy	[J/K]
q	centroid	
$\dot{Q}$	heat transfer rate	[W]
r	exergy to energy ratio	[MJ_{eq}/MJ]
R	correlation coefficient	
ρ	density	[kg/dm^3]
$\sigma(z)$	softmax activation function	
$\bar{s}$	specific entropy	[J/kg K]
S	entropy	[J/K]
SM	mould thickness	[mm]
$S(t_x)$	system scenario at transformation x	
T	temperature	[K]
T1	first phase injection time	[ms]

T2	second phase injection time	[ms]
T3	solidification time	[ms]
TC	time cycle	[s]
TM	multiplication time	[ms]
U	thermal transmittance	[W/m^2 K]
V	volume	[m^3]
VA	speed casting doses	[m/s]
V1	first phase injection speed	[m/s]
V2	second phase injection speed	[m/s]
X	input variable	
$\dot{W}$	workflow rate	[W]
Y	output variable	
η	exergy efficiency	[%]
χ	nondimensional technology obsolescence index	
ψ	nondimensional life cycle quality index	
ω	distance between the support vector and the line	

Acronyms

AI	Artificial Intelligence
AM	Additive Manufacturing
AR	Augmented Reality
AUC	Area Under the Curve
BRI	Building Related Illness
CAD	Computer Aided Design
CEENE	Cumulative Exergy Extracted from Natural Environment
CERA	Cumulative Energy Requirements Analysis
CExC	Cumulative Exergy Consumption
CExD	Cumulative Exergy Demand
CF	Characterization Factor
CNC	Computerized Numerical Control
COP	Coefficient of Performance
COPD	Chronic Obstructive Pulmonary Disease
CPS	Cyber Physical System
CV	Cross-Validation
DALY	Disability-Adjusted Life Year
DED	Direct Energy Deposition
DExA	Demand of Exergy Accumulated
DLMD	Direct Laser Metal Deposition
DNN	Deep Neural Networks
EA	Exergy/Exergetic Analysis
ECEC	Ecological Cumulative Exergy Consumption
EEA	Extended Exergy Accounting
EESI	Extended Exergy Sustainability Index
ELCA	Exergetic Life Cycle Assessment
ELCD	European Reference Life-Cycle Database
ELR	Environmental Loading Ratio
EPC	Electricity Production Cost
EPD	Environmental Product Declaration
ERP	Enterprise Resource Planning

ESR	Exergy Structure Ratio
ETL	Extract-Transform-Load
EU	European Union
ExIO	Exergy based Input - Output
ExROI	Exergetic Return of Investment
EYR	Environmental Yield Ratio
FN	False Negative
FoF	Fabric of Future
FP	False Positive
FR	Renewability Factor
FU	Functional Unit
GDP	Gross Domestic Product
GPU	Graphic Processing Unit
GUI	Graphical User Interface
GWP	Global Warming Potential
HSH	Healthy Sustainable Home
I4.0	Industry 4.0
I5.0	Industry 5.0
IAQ	Indoor Air Quality
ICE	Internal Combustion Engine
ICEC	Industrial Cumulative Exergy Consumption
ICT	Information and Communication Technology
IDEA	Institute for Democracy and Electoral Assistance
IoE	Internet of Everything
IoT	Internet of Things
IoS	Internet of Services
IPCC	Intergovernmental Panel on Climate Change
JRC	Joint Research Center
K-NN	K-Nearest Neighbour
KDD	Knowledge Discovery Database
KET	Key Enabling Technology
LCA	Life Cycle Assessment
LCEA	Life Cycle Exergy Analysis
LCI	Life Cycle Inventory
LCIA	Life Cycle Impact Assessment
LDA	Latent Dirichlet Allocation
MaaS	Manufacturing-as-a-Service
MAE	Mean Absolute Error
MES	Manufacturing Executions System
MFA	Material Flow Analysis
ML	Machine Learning
MPP	Massive Parallel Processing
MRL	Minimum Risk Level
MSE	Mean Squared Error
OEE	Overall Equipment Effectiveness
OLE	Overall Labour Effectiveness
OLS	Ordinary Least Squares
PaaS	Product-as-a-Service
PCA	Principal Component Analysis

PHM	Prognostic Health Management
PLC	Programmable Logic Controller
PPD	Dynamic Payback Period
PPP	Public Private Partnerships
PVC	Poly Vinyl Chloride
R&D	Research and Development
REPA	Resource and Environmental Profile Analysis
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic
SBS	Sick Building Syndrome
SCADA	Supervisory Control and Data Acquisition
SDG	Sustainable Development Goal
SECR	Social Exergy Conversion Rate
SETAC	Society of Environmental Toxicology and Chemistry
SME	Small and Medium Enterprise
SPECO	Specific Exergy Costing
SVD	Support Vector Decision
SVM	Support Vector Machine
SVR	Support Vector Regressor
TBL	Triple-Bottom-Line
TCExL	Total Cumulative Exergy Loss
TE	Thermoeconomics
TEC	Thermo Ecological Cost
TN	True Negative
TP	True Positive
UEV	Unit Emergy Value
UN	United Nations
VOC	Volatile Organic Compound
VSL	Value of Statistical Life
WHO	World Health Organization
WSS	Within Sum of Squares

Subscripts and Superscripts

0	dead state
a	product/process a
al	aluminium
b	product/process b
c	number of total enthalpy flows
ch	chemical
d	number of total workflows
e	equilibrium
eco	ecological
eq	equivalent
g	general
i	state point at the inlet of system/sub-system
k	number of total mass flows
l	liquid
loss	flow rate loss during the sub-processes
m	melting

n	net use
nd	global
nr	non-renewable
o	state point at the outlet of system/sub-system
p	number of total heat transfer flows
r	renewable
res	resource
tot	total
th	thermal
y	years or number of energy flows for CexD

KEYWORDS: *data-driven approach; exergy; I4.0; integration modelling; life cycle assessment; model-based approach; smart manufacturing; sustainability assessment.*

INTRODUCTION

The shift to sustainability is becoming increasingly important in manufacturing, particularly in resource and energy-intensive industries. Furthermore, the Industry 4.0 (I4.0) paradigm gives up new prospects for sustainable growth. In recent years, the subject of sustainability in industrial contexts has taken an essential point on the legislative agendas of many governments and in public opinion, with the latter becoming increasingly sensitive to companies' commitment to this problem (Misopoulos et al., 2018). For this reason, manufacturers have included new sustainability routes into their production processes and raised the amount of communication about these practices to consumers and stakeholders. I4.0 technology have made production processes more efficient and less impactful for manufacturing companies (Hong et al., 2019), moving from centralised to decentralised production. To make things even better, the new I4.0 management and data collecting technologies, which are able to gather timely process data, can assist companies in reviewing the sustainable measures adopted. However, the company's value creation strategy must be considered as part of this sustainable transformation process, which is assisted by the I4.0 paradigm. Implementing sustainability is generally a complicated process that involves a medium-long-term strategic vision and efficient communication between senior management and operational business divisions, as well as between the company and its stakeholders in order to be effective (García-Muiña et al., 2020).

Manufacturers are joining the massive digital transition and are working to implement I4.0 innovations. I4.0 is a series of different innovations, Internet of Everything (IoE), Cyber-physical Systems (CPS), Digital Twin, and Smart Factories, that work together to build the next wave of production plants. This movement seeks to turn factories into "smart factories" that have to be flexible, decentralized, and integrated in order to reach greater level of automation and versatility (Kusiak, 2018).

Data is a key enabler in smart manufacturing (Siddiqa et al., 2016). However, data in its native format is not particularly useful for providing knowledge. This data must be transformed into something more meaningful, which is normally achieved in steps.

According to McKinsey and Company, 86% of companies surveyed felt that the data and analytics program was only partially effective in meeting its primary goal. They also discovered that the main technical challenge impeding success is data management (McKinsey and Company, 2016).

The aim of the doctoral program is to implement and validate innovative assessment models to examine the nature of as-is manufacturing processes and rationalize the relationships and correlations between the various phases of the process. The dual strategy of top-down research and bottom-up experimentation must guarantee that models-based become a tool and a guide for the development of ad hoc measurement and monitoring systems. The structured data provided will be processed (potentially in real-time) using a data-driven methodology and machine learning techniques to deliver more precise and sophisticated information than the ones that a single operator could manage alone, only with his expertise. In complex systems such as manufacturing processes, data-driven modelling methodologies enable the integration of parameters from several domains (e.g., product, process, and logistics) into models that would be impossible to design using purely mathematical theoretical models.

In order to achieve this goal, a functional model that can integrate and analyse large sets of heterogeneous data, that is, data with different characteristics such as different time scales, discrete data and continuous data, data derived from different technological or physical phenomena, will be required (Montáns et al., 2019). This model must be capable of producing analytical results with scientific, physical, and technical importance, as well as serving as a tool in policy decision-making on the sustainable development of industrial processes.

There are two kinds of models: quantitative models and empirical models. The quantitative model is made up of an integration of linear models of the first order and non-linear models of the second order. Specifically, the Life Cycle Assessment (LCA) is characterized by a logic additive consumption of resources, based on the principle of mass and energy conservation, and the exergetic analysis (EA) takes into account not only the typical quantitative allocations of a linear model such as LCA, but also the quality of resource use, yielding a more accurate and faithful vision of sustainability and production performance (Selicati and Cardinale, 2020).

The models based on thermodynamic analysis of manufacturing technologies are based on the study of discrete and continuous variables and represent an interesting and entertaining advanced element for maximizing the sustainability and performance of the individual process and, thus for the synergistic implementation of the entire manufacturing process. The EA technique is the gold standard for industrial sustainability. With a holistic view but a broader context, the goal of decreasing Exergy Losses promotes the maximizing of the intelligent use of resources within the manufacturing process (Abadias Llamas et al., 2019). The second-law thermodynamics assessment of energy has the benefit of being applicable to natural resources, fuels, and products using a standard measuring scale. It is applicable to individual processes, industries, and entire national economies. It provides a solid foundation for evaluating the impact of policy actions aimed at increasing energy, resource, and climate efficiency.

The company test case on which process modelling, analysis, tuning, testing, and validation are performed is for an SME in southern Italy. Master Italy is a company that is designing and manufacturing accessories and components for aluminium window frames since 1986. It follows that the most important process to investigate is definitely the die-casting aluminium alloys. The project entitled "Smart Energetic Manufacturing: Master Twin" has the ultimate goal of researching innovative and sustainable solutions for the efficiency of industrial manufacturing processes within the I4.0 and Smart Manufacturing paradigm, starting from the laws of thermodynamics, up to an integrated Cyber Physical System (CPS). Master Italy is an extremely dynamic and constantly developing SME. Although it has management platforms for production and quality control, the human influence (of operators) on all processes is still very strong.

'Increasing performance processes' usually refers to the practice of lowering energy consumption and materials, waste, the speed of the time cycle, and the ideal relationship between good and discarded pieces. However, the proposed model, which analyses the state variables used, enables to interpret the processes in the logic of product and process optimization and thus recognize the end of the interpretation of operational reality via the Key Enabling Technologies (KETs) of I.40, obtaining a true predictive production. The definition of innovation is precisely the achievement of an automatic continuous improvement. The models that define this paradigm must be consistent in their analytical formulation, empirical relationships, discrete and continuous analysis, and probabilistic analysis (Wynn and Clarkson, 2018).

The question is whether, given enough data, machine learning algorithms generated from the data-driven method may provide additional insight into process parameters or combinations of process variables useful in forecasting casting quality or performance metrics from production process data. The industry requires technology capable of detecting patterns that are too fine for humans to identify. The capacity of machine learning algorithms to discover patterns and correlations between inputs and outputs for high-dimensional datasets is a critical feature The data from the casting process is multidimensional, with various inputs like as temperatures, velocities, pressures, timings, and chemical composition. More dimensions can be added (Blondheim, 2021), but having too many adds confusing noise to the data. There is a need to start studying the data that the sector is now gathering and determine which metrics are important and which are not. Experts believe that adopting machine learning into automation is critical for organizations to maintain and improve their competitiveness (Wang et al., 2018). However, in real-world applications, this will take time. One issue is that gathering the necessary data might be difficult, owing to the associated costs in both collecting data in production systems and manually preparing it for training purposes (the definition of training data known as labelling). The data might also be noisy, i.e. not exact enough. Another reason is that incorrect predictions may arise, resulting in system faults. This might have serious consequences, such as equipment damage, full failure, or, worse, danger to individuals engaged in the manufacturing process. As a result, a number of needs that are crucial for incorporating machine learning into automation technologies have recently arisen.

This study discusses some of the techniques that can be used to construct a method that can assist with greater control over the die-casting process while also adding benefit to the production process. This, in fact, could improve data-driven decision-making policies. The aim is to provide a system that can convert a large volume of data into knowledge that can be used by plant operators in particular. As a result, in keeping with the goals of the current Industry 5.0 (I5.0) model (European Commission, 2021a, p. 0), this work focuses on and encourages both the automation aspect as well as the human learning process. Furthermore, since the workflow starts with data collection from a completely or partially automated process, data processing is dependent on the details of a precisely defined use case. As a result, domain knowledge, or the experience of field operators, is still critical.

The work is structured as follows: the first section investigates the connection between the concepts of sustainability, I4.0, and Smart Manufacturing. A reference to the upcoming I5.0 paradigm is made. In the second section discusses the model-based approach (specifically Life Cycle Assessment, Exergetic Analysis, and hybrid methods of them), as well as the difficulties of integrating such models through a systemic view of the problem, and the data-driven approach, with a focus on big-data and the most common Machine Learning techniques to process and extract knowledge from them. The section concludes with a consideration of the Research Gaps that characterize both methods. The third section contains a state-of-the-art on performance metrics, which are used to understand the findings of the analysis as well as to evaluate the models developed qualitatively and quantitatively. The fourth section covers the whole test case, from the operational context through the implementation of both model-based and data-driven approaches, as well as a discussion of the outcomes of the analyses and tests performed. The fifth section is a synthesis of different applications that highlight the transversality and multidisciplinarity of the knowledge acquired during the PhD program. Finally, the overall conclusions bring the dissertation to a close.

1. IN THE ERA OF INDUSTRY 4.0, WHAT DOES IT MEAN TO BE SUSTAINABLE?

Sustainability has had many definitions over the years from a broad variety of disciplines. We utilize the intergenerational philosophy based on meeting the needs of current generations without compromising the ability of future generations to meet theirs (WCED, 1987). We also rely on the multidimensional concept of the triple-bottom-line (TBL) (Elkington, 1998), depicted with his main features in Figure 1.1. The three main pillars of TBL are economic, environmental, and social dimensions. We also link sustainability to the United Nations Sustainable Development Goals (SDGs) (United Nations, 2015).

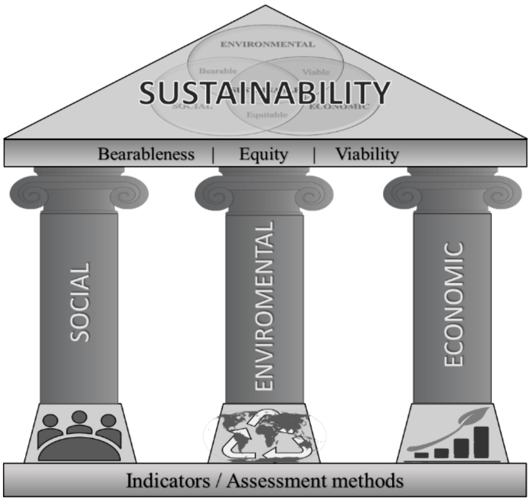


Figure 1.1 - Three pillars of sustainability concept (TBL)

While it describes a noble goal, this concept of sustainable development gives no instruction on how to accomplish it. Furthermore, a plethora of concepts, processes, standards, and metrics have been created to assist organisations in approaching sustainability (Lozano, 2008; Seliger et al., 2011), yet what “sustainability is” is indeed open to debate (Clifton and Amran, 2011; Eslami et al., 2020).

1.1. I4.0 AND SMART MANUFACTURING

The German government coined the term "I4.0" (I4.0) in 2011 to describe a set of technological changes in manufacturing systems brought about by automation and ICT (Information and Communication Technologies), such as Cyber-Physical Systems, Internet of Things, Simulation and Modelling, Big Data Analytics, Augmented Reality, Additive Manufacturing, Robotics, Cloud Computing, and now Blockchain. It aims to assist in the incorporation and merging of autonomous devices, humans, physical objects, and processes through operational stages in order to create various forms of digital data, functional, and high agility value chains throughout the entire life cycle of a product, process, or activity (Tay et al., 2018). To that purpose, (Liao et al., 2013) conducted a complete systematic literature review of I4.0 in all of its aspects, with 224 publications out of 346 prospective ones proceeding to the data collecting stage for qualitative and quantitative analysis. Smart manufacturing is also focused on the product, in fact another statement worthy of note is that date from (Doyle-Kent and Kopacek, 2020) about the product made by this new industrial time in history: production will be automated, digital, and data driven, not only agile and lean. Products will be of exceedingly high quality and more reasonably priced. These items must be of the highest quality, and the supply chain must be optimized to facilitate manufacture.
The I4.0 paradigm is performed in three dimensions (Vaidya et al., 2018):
<u>Horizontal integration</u> across the full network of value production. It refers to the integration of numerous IT systems utilized in various phases of manufacturing and business planning processes, involving the interchange of goods, resources, and information;

End-to-end digital integration across the whole product life cycle. It enables the incorporation of smart business processes throughout the supply chain, including the production floor and CPS services. Intelligent cross-linking and digitalization include the use of an end-to-end solution based on ICT integrated in the cloud;

Vertical integration and networked manufacturing systems. It refers to the integration of various IT systems at the company's many hierarchical levels during the production process, from product development to manufacture, logistics, and sales.

A collection of essential principles of the I4.0 paradigm that lead to smartness follows (Lasi et al., 2014; Hermann et al., 2016):

a) *automation* of repetitive activities, implying that employees must concentrate on innovative, inventive, and communicative tasks. It is required to halt manufacturing in order for production to stay continuous. It entails outfitting each machine or assembly line with a system that supports them. Each worker must be able to cease the production process as soon as the support-aided system detects an anomalous status;

b) *decentralization*, as the capacity of enterprises, operations department, and even equipment to make decisions on their own rather than relying on a centralized computer system of a decision-making body. This idea promotes quicker decision-making and more flexibility. It is an ideal organizational structure for meeting the increased demand for highly customized services;

c) *real-time data acquisition, processing and communication*, as big data technologies improve organizations' capacity to operate in real time. The big data collected from plants about machines, equipment, and products, as well as customer data collected from various sources such as social media, direct selling points, and data received from suppliers, when analysed in real-time, changes the way decisions are made and has an impact on the industry's profitability;

d) *virtualization*, referring to the process of generating a virtual duplicate of the physical system. Process monitoring and machine-to-machine communication are aided by virtualization. The sensor data is integrated to these simulation-based virtual models. The virtualization aids in warning the human of system malfunctions and improves safety requirements;

e) *modularity*, referring to modular manufacturing systems that can be easily customized by changing and extending specific components. System flexibility allows for capacity adjustments in the event of seasonal variations or changes in product increase in output. Modularity also allows for the simulation of multiple manufacturing processes, such as product design, production planning, production and production engineering, and services, as separate processes and then tightly integrating them to provide interchangeability;

f) *flexibility*, Ad hoc networking based on CPS allows for the dynamic setting of many elements of business operations such as quality, time, risk, resilience, price, and sustainability. This allows for ongoing material and supply chain optimization. It also implies that engineering processes can be made more flexible, manufacturing processes can be modified, momentary shortages (due to supply concerns, for example) may be compensated for, and massive increases in output can be accomplished in a short period of time;

g) *agility*, as a company dynamic capability that enables it to manage change and uncertainty in the environment. There are two kinds of agility (Mrugalska and Ahmed, 2021): facility agility and flexibility agility. The capacity of a production facility or shop floor to handle any unpredictable change in product manufacturing preference is characterized as facility agility. While agility refers to the organizational competence required to work on many tasks at the same time. Smart manufacturing, cyber-physical systems, big data and analytics, cloud computing, and IoT enable businesses to improve their agility in both value and supply chains;

h) *efficiency*, providing the greatest feasible output of products from a given volume of resources (resource productivity) and utilizing the fewest resources possible to produce a certain outcome (resource efficiency). It enables case-by-case optimization of manufacturing processes across the whole value chain. Furthermore, rather than needing to halt production, systems may be constantly optimized throughout production in terms of resource and energy usage, as well as emissions reduction;

i) *interoperability*, as the capacity to execute the same activity even after switching machines and equipment from various manufacturers creating a trustworthy environment by connecting numerous networks in a production line;

j) *service orientation*, as the entities in the production system are all interconnected, making the establishment of the product-service system easier. Because of the flexibility and agility gained as a result of service orientation, companies can adjust to market changes more rapidly. This enables the businesses' many stakeholders to collaborate and co-create value for their customers. It refers to Manufacturing-as-a-Service (MaaS) and Product-as-a-Service (PaaS) concepts (Kusiak, 2020).

I4.0 is now supported in all fields, not only manufacturing. Examples include logistics, construction, transportation, medical and surgery, food production, home automation, and so on, as well as cell phones and watches in our daily lives.

While it is difficult for researchers to settle on the appropriate and cohesive notion of I4.0 and its related supporting technologies, the research affirms that the large network of useable and open to everyone sensors, as well as Cloud Computing, is at the heart of this paradigm.

It is clear that the driving ideas of I4.0 were first focused on boosting efficiency and profitability rather than proposing answers to the environmental concerns created by manufacturing.

The associated benefits are numerous (Kiel et al., 2017), including reduced pollution and environmental dangers, as well as improved financial performance as a result of new overseas market opportunities. In reality, an environmentally conscious firm would be able to gain environmental certification, as well as the corresponding rise in status. I4.0 would be a step forward in the creation of more competitive manufacturing value. This phase is primarily characterized in contemporary literature as a dedication to the environmental element of sustainability. The distribution of services, such as products, supplies, electricity, and power, may be made more efficient by utilizing smart cross-linked value creation modules (Stock et al., 2018). To yet, the qualitative assessment of the potential for long-term value creation in I4.0 has not been addressed in a systematic and formal approach (Kamble et al., 2018).

(Bonilla et al., 2018) undertook an intriguing prospective study on both positive and negative cause-effects that all of the elements of I4.0 would bring in the short and long term in the manufacturing area, using the ideal point of sustainability as a threshold. In general, the trends of long-term environmental consequences as a result of I4.0 implementation are stage-dependent, with the tendency being negative during the deployment stage and positive during the operating stage. In the long run, and to summarize ideas, smart manufacturing would bring some positive aspects on environmental sustainability (Selicati and Cardinale, 2021a), such as :

- Creating significant effects on sustainability throughout the entire supply chain;
- Increasing the productivity with cost reductions;
- Inventory reductions through real-time smart inventory management and traceability;
- Real-time supply chain optimization & supplier’s integration that will enhance the development of a circular economy.
- Decentralization of the collection of goods and services;
- Development of strategies and goods that take into account customers' lifestyles;
- Acquisition of new ecological market awareness;
- Achievement of shorter production time cycles;
- Processing an amount of production calibrated to predicted needs, without further depletions;
- Monitoring and control of CO_2 emissions.

In the Figure 1.2 below it is schematized the context that turns around the I4.0 paradigm

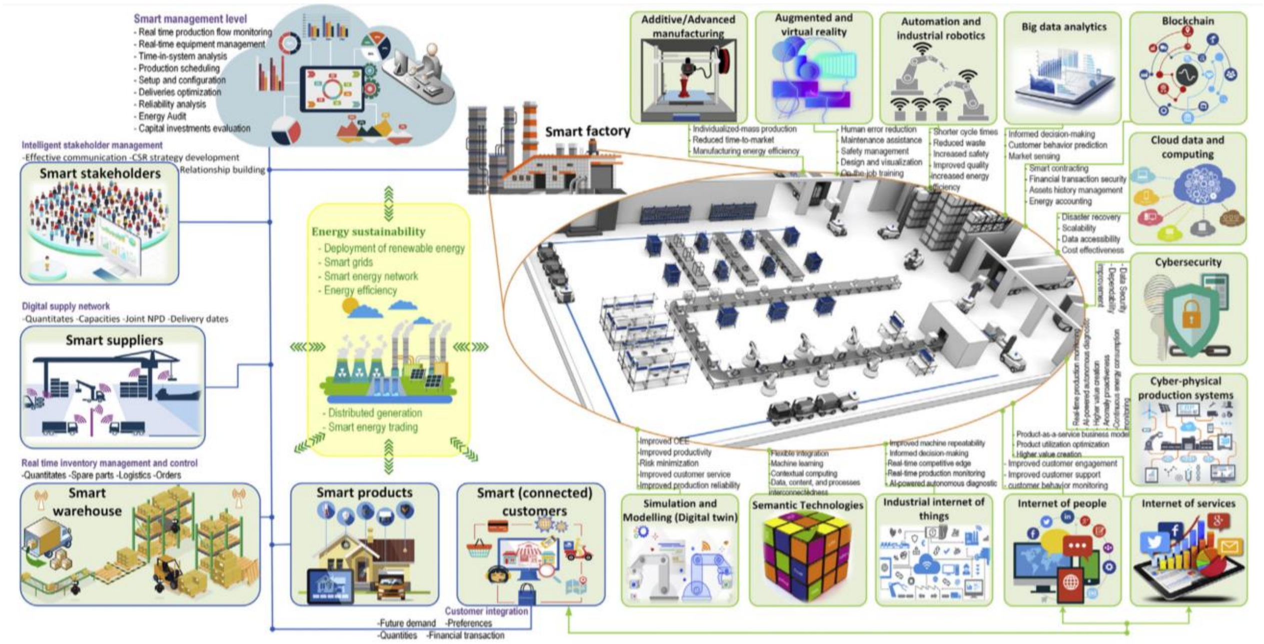


Figure 1.2 - I4.0 context (Ghobakhloo and Fathi, 2021)

Within the I4.0 paradigm, let us briefly describe the key enabling technologies (KETs) of I4.0, as depicted in Figure 1.3:

Internet of Things (IoT) (Brous et al., 2020). This technology allows companies to connect several remote devices utilizing sensors and microprocessors driven by software systems capable of relaying data across networks. In this regard, it is necessary to emphasize that such devices are internal to production machines and that they can be built even after the latter has been completed, owing to the idea that, in the era of I4.0, any physical entity has the potential of being smart with the intention of sharing information on its own state and the state of the world in which it is located. The Internet of Services (IoS) is related with the strategic use of the Web and a novel method of generating demand via the materialization of the PaaS business model. Currently, manufacturers of commercial products are attempting to establish a clear relationship with customers and strengthen their strategic advantage by providing complementary services and generating new revenue streams, and IoS is providing the necessary technical infrastructure.

Big Data Analytics (Belhadi et al., 2019) refers to a new wave of technologies and architectures that enable enterprises to economically extract value by detecting, collecting, and analysing massive amounts of data. Big data analytics helps modern businesses to extract more value from the huge volumes of data they currently have by predicting what will happen next and determining what actions should be done to obtain the best results. It eventually leads to Artificial Intelligence (AI).

Cybersecurity (Junior et al., 2021). I4.0 requires access to the environment in order to facilitate integration of various processes. While it is critical to re-establish communication methods in order to exchange information, it is also critical to monitor this sharing in order to secure data flows. Companies require cybersecurity measures to better safeguard a device or a device collection in terms of knowledge exchange and data privacy.

Blockchain (Khanfar et al., 2021). Often referred to as distributed ledger technology, it serves as the foundation for cryptocurrencies such as Bitcoin and Ethereum, although its capabilities extend well beyond that. Blockchain is permanent, decentralized, and redefines trust by enabling open, secure, efficient, and timely public or private solutions.

Augmented reality (AR) (Posada et al., 2015). Virtual reality is being touted as one of the most revolutionary uses in I4.0 using 3D modelling, CAD, and projection technologies. A three-dimensional model capable of housing a human operator is mentioned, with the purpose of evaluating the process in order to enhance it throughout the design and commissioning phases, as well as to assist worker training. In the case of Augmented Reality, however, mention is made to the concept of leveraging unique viewers to gain additional information about the object merely by framing it. In I4.0, this notion translates into the possibility of gaining access to automated and intelligent product logistics, which aids in finding them in the production and tracking order enforcement in real time. This approach allows for the testing of items from an aesthetic and functional standpoint, as well as the simulation of their placement in the reference environment.

Robotics and Advanced Manufacturing Solutions (Matheson et al., 2019). One of the primary triggers is and should be robots, which are viewed as human operators' collaborators. Such technologies have the potential to improve production processes and boost the productivity of organizations who adopt them. Human engagement in operations involving interactions between automatic and manual systems aids integrated and automated techniques. Throughout this example, robots are true interactive gadgets capable of sharing knowledge with other devices and humans while remaining autonomous and configuring paths based on the output flow's demands.

Additive manufacturing (Hernández Korner et al., 2020). It is a technique that can print a product by adding material after starting with a computer drawing (assisted by a CAD) of the thing to be manufactured. To construct any shape, the nozzle may melt tiny layers of powder and put one layer of material, either plastic or metal, on top of another. The great potential of this advancement is thus the ability to travel directly from the digitally codified concept to the product without having to go through intermediary steps, so making way for new business models where pieces may be produced on demand.

Simulation and modelling techniques (Bárkányi et al., 2021). Simulation is a term that alludes to the notion of a digital twin, which is defined as a mathematical model capable of modelling a process, product, or service in order to conduct an analysis and use predictive performance strategies. This is the development of an actual process model in order to collect useful knowledge that may help businesses cut production costs, improve the efficiency of the end product, and shorten time-to-market. Simulation and modelling would be necessary in smart factories to use real-time data to mimic the actual environment in a simulated model that may include computers, products, and humans.

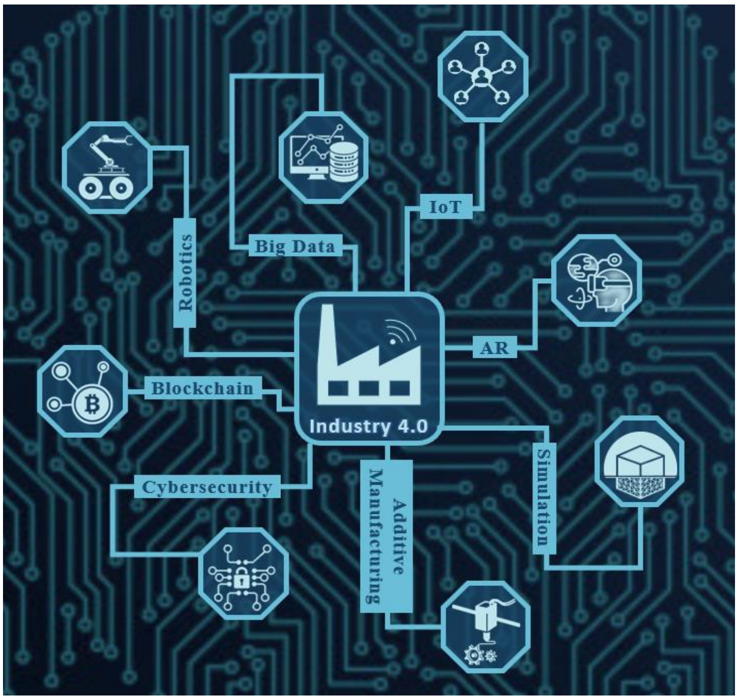


Figure 1.3 - I4.0 KETs

1.2. SUSTAINABLE SMART MANUFACTURING

In contrast to the linear model of industrial economics, in which a resource is used and the remainder is discarded, the circular economic system undermines the vision of mass consumerism through new inputs. Reuse and recycling are good places to start avoiding wasting a lot of resources and thus reduce the negative externalities on the environment (Stahel, 2016). Contamination, pollution, and conflicts over supply control have all been prevalent throughout the history of industrialization. All manufacturing processes reduce the availability of future resources, resulting in a reduction in future production capacity. To protect tomorrow, we must pay attention to the proper use and application of what we have today (WCED, 1987).

Technology and its evolution are having an ever-increasing impact on our world in the social, economic, and environmental dimensions. This paragraph will provide an overview of the current phenomenon, as well as examples of the changes it has caused. Many aspects of the industrial and economic fields have changed as a result of the fourth industrial revolution, such as the creation of new production models, new organizational forms of work, and, as a result, an evolution of the professional figures involved. The potential for manufacturing unemployment as a result of the trend to replace workers with machines is one of the nerve centres affecting the already underway I5.0 (Longo et al., 2020), along with the various approaches with which the European Union and its members are dealing with changes. The shift in the relationship between business and consumer on one hand, and consumer and product on the other, as new technologies enable greater product customization. Innovations are changing the way we think about products, characterizing them with technologies that aim to simplify the consumer's life while also protecting the environment. One of the difficulties that smart production is facing is efficiently managing the trade-off between profitability and sustainability (Lao et al., 2014).

The term "smart industry" refers to a new concept of intelligent industry. This characteristic has come to life and joined entrepreneurship in recent years as a result of new, more cautious, and efficient technologies. The concept of intelligent industry is closely related to the sustainability and sustainable development with which the company must deal, as well as technology and process efficiency (Ellen MacArthur Foundation, 2017). The industry takes advantage of the benefits that technological progress provides.

The Internet of Things enables innovation in industrial processes by interconnecting processes, using intelligent machines, collecting, storing, and sharing data (big data and cloud), and transforming them into smart, intelligent ones. There is "smart production" through interactions between machine and human, "smart services" through integrated systems aimed solely at customer needs to which they should respond as efficiently as possible, and "smart energy" which focuses on consumption, monitoring cycles of use, and waste of energy resources. The new intelligent entrepreneurship will strive to maximize results by meticulously controlling every variable in its value chain, achieving a balance of economic results and sustainability through efficiency (Youssef et al., 2017).

The new industrial progress 4.0 calls for the centralization of information, its preservation, and its application as a starting point for new programming. The intelligent industry employs machines that learn on their own and improve their performance by collecting, analysing, and using data as a foundation for learning and implementing their activities. Digitization is a critical tool for industrial responsibility. Interactivity is a significant development line in the design of new intelligent production processes, where the workflow is facilitated toward clean, effective, but most importantly efficient, and sustainable enforcement (Stock and Seliger, 2016). The 2030 Agenda, a plan of action signed by the governments of the 193 UN member countries in September 2015, includes 17 Sustainable Development Goals (SDGs) addressing people, the planet, and prosperity (Rosa, 2017). It is appropriate, according to point 9.5 of the 2030 Agenda: "Enhance scientific research, upgrade the technological capabilities of industrial sectors in all countries, in particular developing countries, including, by 2030, encouraging innovation and substantially increasing the number of research and development workers per 1 million people and public and private research and development spending". The new Smart Industry is aimed squarely at achieving these goals. The 2030 Agenda identifies the guidelines for addressing the problem of unsustainable development, which has received far too much attention in the past

(e.g., UN conferences in Stockholm in 1972, Rio de Janeiro in 1992, and Johannesburg in 2002), but has yet to yield implementation directives and deadlines. The integrated solutions agreed upon at previous summits are now being directed toward the targets to which they can respond, owing to the new technological-industrial progress we are witnessing. In terms of innovation, the combination of circular economy goals and I4.0 appears to provide excellent answers. Goal 12 in particular aims to ensure “sustainable production and consumption patterns”. Consumption and sustainable production are aimed at doing more and better with less, increasing the benefits in terms of well-being derived from economic activities while reducing necessary resources, degradation, and pollution throughout the production cycle, thereby promoting the improvement of life quality. In this regard, technology has the potential to make a significant contribution to this change. The equipment of I4.0 enables the strict control of production cycles, with optimal use of each production source. As a result, the model of planned obsolescence has become unbalanced, forcing the transition from the old to the new. Furthermore, reusing is possible thanks to many technologies designed for the new Industry model. To enhance production while preserving environmental, economic, and social sustainability, manufacturers have access to a wide range of information and learning techniques. This is not the first time that data analysis has been used to enhance goods and processes, though. Instead of people being unable to interpret huge volumes of data and computers being better at analysing large amounts of data, big data analytics in manufacturing allows practitioners to extract inherent knowledge utilizing this current approach based on computer analysis techniques. Horizontal integration also allows to create a sustainable industrial environment (Ejsmont et al., 2020), in addition the pull concept is used in the smart factory’s logistics, which ensures that raw materials or semi-finished manufacturing materials are demanded on request (Waibel et al., 2017). There are numerous research papers in the literature that link sustainability to the I4.0 paradigm. Process and information flow integration is one of them (Carvalho et al., 2018).
It should be noted that the approach to sustainability convergence in I4.0 is changing. Indeed, I4.0 began by concentrating solely on increasing efficiency and productivity in order to maximize profits and competitiveness. Since now, even with the many obstacles that this sector faces, such as the unification of laws, corporate protocols, the competition for trained labour, and the introduction of a compliant regulatory system, the current technological transition has centred on manufacturing rather than an environmentally friendly framework (Rajnai and Kocsis, 2018).
(Ghobakhloo, 2020) clarified the relationships between various sustainability functions of I4.0 in order to understand the opportunities for sustainability provided by the digitalization.
(Vrchota et al., 2020) in their extensive review concluded that I4.0 technologies are an auxiliary tool for achieving sustainability in all three dimensions, environmental, economic, and social. Furthermore, the most important enabling technologies for achieving high levels of sustainability are identified. (Müller and Hopf, 2017) suggest a concept focusing on the TBL, which is a model that includes the problems and opportunities involved with the implementation of I4.0. (Oláh et al., 2020) have deepened the link between I4.0 and sustainability from a different perspective: they have not evaluated the usefulness of enabling technologies in the development of sustainable manufacturing, but they have observed how these technologies can have a negative impact on environmental sustainability due to the immeasurable use of non-renewable resources and pollution. Another method for assessing the relationship between I4.0 and sustainability is the one used by (Bai et al., 2020) who tested the enabling technologies as well as their intended application using an evaluation scheme focused on the triple-bottom-line principle. Another interesting contribution is provided by (Kamble et al., 2018) who, following a special review, have proposed a personal Sustainable I4.0 framework that relates the KETs to the sustainable outcomes for the sustainable manufacturing decision-making policies in accordance with the foundation principles of I4.0.
From this brief literature examination, it is clear that most of the authors are confident about the fact that there is an important and positive relationship between the application of I4.0 and its environmental benefits so that companies tend to adopt this technology more given its benefits and regardless of the company size and industry sector.

I4.0 is an excellent option for long-term manufacturing sustainability. The architecture necessitates massive amounts of data processing, retrieval, and examination on cloud computing. As an optimal starting point, practitioners suggest Starting with a historical data approach to training the experience provided by ERP, MES, PLC systems and then also integrate sensing, actuation, and specific levels of real-time tracking and regulation. Predictive analytics and automation are viewed as vital technologies needed for sustainable manufacturing in R&D (de Sousa Jabbour et al., 2018; Ren et al., 2019).
If all goes as planned, such positive prospects may be capitalized on and be profitable in the long run. Enhanced production effectiveness is directly proportional to enhanced sustainability (Tiwari and Khan, 2020).The choice of Long Run is no accident. Considering a hypothetical life cycle of an I4.0 application, it is useful to split the design part, the installation of the equipment, which undoubtedly have a negative impact on sustainability, and the operating part, in which technologies are now ridden and become in effect Tools for the efficiency of process performance, both at the quality and sustainability points.
The challenges and opportunities associated with implementing I4.0 are, however, for the moment unknown, as environmental sustainability technologies associated with this sector have not been adequately explored, because new technologies still exist; thus, gaps exist in the way in which we can integrate effective use of scarce resources, raw materials and other resources.

1.3. THE DAWN OF INDUSTRY 5.0 PARADIGM

Despite the fact that the Industry 4.0 Paradigm is still being deployed, particularly for small company realities, the notion of Industry 5.0 (I5.0) has been proposed since last year.
This notion arose from a question that began when all contingent technologies to the I4.0 phenomena threatened the function and utility of human inside the smart factory: are we certain that the human will continue to play a valuable role within the industrial value chain? The wave of change in industry will have long-reaching consequences that will extend well beyond the technical changes on the production floor. A revolutionized industry will also have a disruptive effect on society. This is especially true for industrial workers, whose jobs may be altered or even endangered. However, no matter how advanced technology becomes, people will always play a critical role. To function, data analysis necessitates the use of human skills and interpretation in order to make intelligent decisions at the end of the analysis. Companies must strike the proper balance between self-organizing and autonomous systems and the human capital they already have. Humans are then reintroduced into the spiral in this new paradigm, boosting their collaboration with intelligent machines to the point of working side by side on the production floor. I5.0 may provide the best of both worlds, combining the well-known benefits of robots with improved cognitive skills of humans in areas such as critical thinking. Production lines may become even smarter in this more complex environment, with people able to oversee far greater degrees of product customisation. That's an interesting concept in industries as diverse as automotive, consumer electronics, and jewellery, or even for items like craft beer, where subtle labelling touches may result in increased customer appeal. From the first industrial revolution to the current day, the significance of humans in industry, particularly in manufacturing, has been emphasized (i.e., the Taylorism in USA (Taylor, 2003)). The I5.0 paradigm is essentially an extension of the I4.0 paradigm, with a concentration on automation and robotics: Multi Robot Systems are made up of cooperative, industrial, mobile, and humanoid intelligent robots that work together to perform basic functions. Collaborative robots are not intended to replace people, but rather to assist them (Durakbasa and Gençyılmaz, 2021). It is a description of a human-centered system. There are those who speak of I4.0 Joint to Society 5.0 (Polat and Erkollar, 2020).
The Figure 1.4 depicts a comparison of I4.0 enabling technologies and their advancements with the introduction of I5.0.

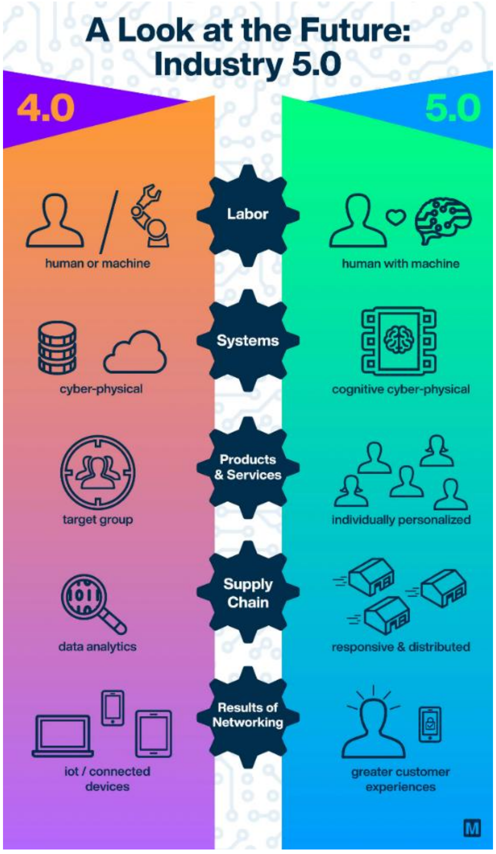


Figure 1.4 - Improved principles and technology as a result of the I5.0 paradigm revolution (Hibbert, 2020)

It is undeniable that the operator who will interact with the robot must be suitably skilled, say digitally competent, to adapt to a highly automated, ever changing workplace; he must be an expert in mechatronics, so that the figure of man is still useful in the smart factories, it will have to take the role that according to the i5.0 will be defined as "Robocollaborator " or" Coboter" (Doyle-Kent and Kopacek, 2021). Education is crucial in this regard. This new operator falls between a traditional operator and one who has some mechatronic competence. Furthermore, due of the significant advancement in this subject, they should be taught on a yearly basis. This, in turn, entails significant costs for the company (much more to SMEs), which will be required to implement continuous updating courses for all employees. Based on what has been discussed, the initial pillars of the I5.0 paradigm that can be counted will be: Organization (management), Technology (cobots), People (coboters), and Tasks (job specification). Considering factors from a social standpoint, greater automation has resulted in higher worker safety as the I4.0 production environment has substantially improved. Ergonomically built workstations enhance working conditions. Collaboration increases when consistent data is available. Resource optimization, such as more energy-efficient machinery operation, also improves environmental protection.

The European Commission also spoke out over the I5.0 paradigm (European Commission, 2021a): “I5.0 provides a vison of industry that aims beyond efficiency and productivity as the sole goals, and reinforces the role and the contribution of industry to society. It places the wellbeing of the worker at the centre of the production process and uses new technologies to provide prosperity beyond jobs and growth while respecting the production limits of the planet. It complements the existing I4.0 approach by specifically putting research and innovation at the service of the transition to a sustainable, human-centric and resilient European industry”. In fact, these are the three keywords, and Figure 1.5 schematizes them. For industry to become the provider of true prosperity, the definition of its true purpose must include social, environmental and societal considerations. This includes responsible innovation, not only or primarily aimed at increasing cost-efficiency or maximising profit, but also at increasing prosperity for all involved: investors, workers, consumers, society, and the environment (Xu et al., 2021). As already stated, a ***human-centric*** strategy places essential human demands and interests at the centre of the manufacturing process. Instead of asking what we can accomplish with new technology, we ask what it can do for us. Rather than asking workers in the industry to adapt their abilities to the demands of quickly expanding technology, we want to utilize technology to adapt the production process to the needs of the workers, such as guiding and training them. It also entails ensuring that the use of new technology does not jeopardize employees' fundamental rights, such as the right to privacy, autonomy, and human dignity. This description clearly reports the fundamental definition of sustainability; in particular, the factory of the future (already one of the I4.0 paradigm's pillars) must be *sustainable*: generate circular manufacturing systems for reusing, repurposing, and recycling natural resources, reducing waste and environmental effect. Lastly, a *resilient* industry refers to the requirement to improve industrial production's robustness, equipping it better against interruptions and ensuring it can offer and sustain key infrastructure in times of crisis; it should be balanced by building sufficiently robust strategic value chains, adaptive manufacturing capacity, and flexible processes, particularly in value chains that support essential human needs such as healthcare or security.

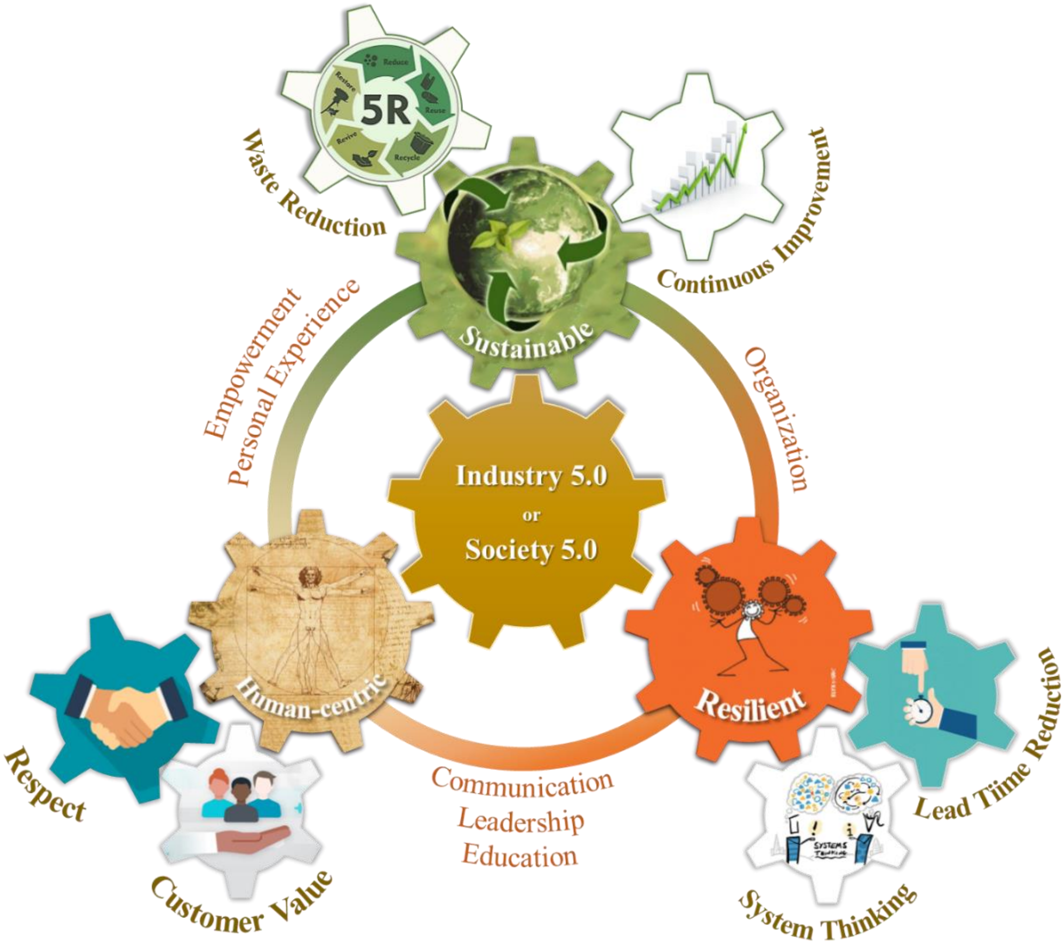


Figure 1.5 - The fundamental elements upon which the notion of Industry 5.0 is centred

However, such an approach must take into account society's diverse perceptions of values and requirements, as assessing and quantifying environmental and, especially, social worth remains problematic.

When life sciences technologies are coupled with engineering and computer technology disciplines, a more systematic innovation strategy that combines multiple views and takes a holistic view of complete ecosystems is required. Furthermore, the systems created will be very sophisticated, interconnected, and interdependent, as well as dealing with heterogeneous data sets. Economic goals such as productivity and competitiveness must not be overlooked, but must be established within the context of agreed-upon ecological and social objectives. This may be accomplished through economic models that place a premium on the development of ecological and social value, as well as through legislative incentives.

Apart from terminology and issues originating from complexity or technology concerns, such a notion must incorporate society and the majority of the industrial landscape. Customers and whole supply chains, all the way up to SMEs, must thus be better connected to guarantee broad-scale application and value development toward prosperity.

For the sake of brevity, and because this thesis does not deal with the emerging I5.0 paradigm, refer back to the European Commission's Report (European Commission, 2021b) for further insight, particularly on the KETs that characterize this trend.

2. MODELS OF ASSESSMENT: STATE OF ART

Manufacturing systems necessitate effective techniques to modelling the complexity of industrial systems and the performance criteria associated with them. Several techniques and tools for modelling complex systems have been proposed and developed in the literature. Model-based methods and tools provide a way to use models to design and formalize the behaviour of a real physical system; data-driven methods provide a way to detect patterns from data, intrinsic and explicit features of a real system; and combined approaches that merge models and data to emulate the physical real system with intelligent models (Tidriri et al., 2016). taking into account a systemic thinking of manufacturing processes (which characterizes the guideline of the topics discussed in this dissertation), Figure 2.1 depicts the relationship between data-driven, stated as a bottom-up approach, and model-based, stated as a top-down approach, implemented on the process/machine under study. While the first leads to problem solution of the issue from a local to a global scale (e.g., using the invariant and automated model), the second begins with a problem description based on the company's needs to answer the question "what do we need to meet these requirements?". The two techniques, when integrated and applied simultaneously, cross right at the model design phase, when the required model (the automated model) must mirror as accurately as possible the physical model that represents the actual behaviour of the process/machine. Despite the benefits and promise of hybrid techniques, the biggest limitation is a lack of a general framework for hybrid approaches.

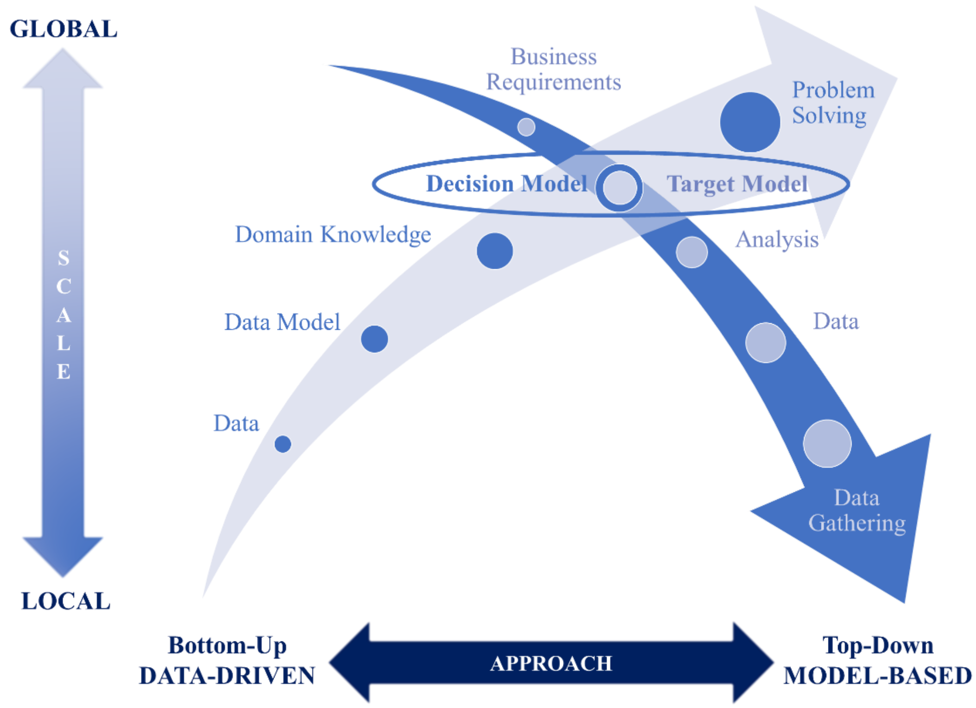


Figure 2.1 - Bottom-up (data-driven) approach integration scheme with top-down (model-based) approach

2.1. MODEL-BASED APPROACH

Model-based has recently garnered a lot of attention in the manufacturing industry. In a manufacturing environment, this notion refers to the development of items utilizing a form of digital model from which additional outlying actions to manufacture the product may be derived. A model is a depiction of an actual thing's structure, object, operation, or idea. Model-based engineering is a way of putting into action a collection of interconnected models that allow for the definition, design, and documentation of a system under development. Until recently, most engineering and manufacturing techniques depended on conventional ways

of transmitting engineering data and driving production processes, such as hardcopy and digital documents. Model-based approach is essentially an application of modelling principles, design, analysis, validation, and verification procedures. It provides a comprehensive best value solution for studying and documenting a system's attributes. Models, though not an ideal depiction of a physical system, produce the necessary information as well as quick feedback (Frechette, 2011). Fortunately, with the introduction of contemporary industrial data handling and powerful engineering programs, it is now possible to accomplish the majority of production tasks using data models. Model-based engineering is a modern way of handling production data that relies on models rather than traditional papers for all production procedures throughout a product's life cycle. Recently, the manufacturing sector has been focused on addressing the factors related to model implementation utilizing computer simulation experiments, in order to close the gap between model definition and simulation program (software).
Models based on thermodynamic analyses, for example, provide a novel and fascinating method for assessing and optimizing the sustainability of manufacturing system performances, therefore easing the management of smart manufacturing processes. The goal of this study is to provide an understanding of how thermodynamic principles may assist improve energy efficiency while also providing methodological support to develop towards a smart sustainable process.

2.1.1. Life Cycle Assessment

The LCA is an objective and reliable methodology to get a comprehensive and holistic approach of assessing environmental damage related to the building even when it is used to support decision making for the definition of policies strategic in this sector. However, because a detailed LCA study can be costly in terms of economics and time, as well as complex to carry out (a significant amount of environmental data must be acquired during each phase of the life cycle, as well as knowing in depth both the standardized methodology and the support tools such as software and databases), researchers are increasingly developing "Simplified LCA" tools (Oregi et al., 2015) that allow an immediate assessment of the life cycle of the products even by those who are not experts (Christiansen, 1997).
The LCA may be traced back to the early 1960s, with the release of studies on energy loads associated with various industrial outputs. During this time, a concept that encompasses the full life cycle, known as "Environmental Life Cycle Thinking," began to gain traction. The challenge of raw material and energy resource exhaustibility prompted more research in the next decade, with a primary focus on improving energy resource management. Between the late 1960s and the early 1970s, there was a gradual shift from analysis that focused primarily on energy use to analysis that examined both raw material and energy resource usage. Two major publications from this time are "The Limits to Growth" by Meadows et al. in 1972 and "A Blueprint for Survival" by Goldsmith et al. in 1972, both of which attempted to forecast the consequences of an increasing global population on the demand for raw materials and energy. During this time, the "from cradle to grave" concept was also adopted, which quantifies the consumption of resources and the discharge of contaminants into the environment throughout the product life cycle. The measurement of resource consumption and environmental consequences of goods was established in the United States under the acronym REPA (Resource and Environmental Profile Analysis), while in Europe it was known as eco-balance.
In the late 1970s, the concept of "sustainable development" emerged, and at the same time, the "Handbook of Industrial Energy Analysis" by Bounstead and Hancock (1979) was published in Europe, marking a watershed moment in the history of LCA methodology in that it was the first document to offer a description of an operational nature of the analytical procedure, which is to be considered a fundamental part of the current technique (Dealy, 1980).
The most comprehensive description of LCA was offered by SETAC in its article "Guideline for Life-Cycle Assessment: a Code of Practice" (Fava et al., 2014): "A process to evaluate the environmental burdens associated with a product, process, or activity by identifying and quantifying energy and materials used and

wastes released to the environment; to assess the impact of those energy and material uses and releases to the environment; and to identify and evaluate opportunities to effect environmental improvements".

At the European level, the European Platform on the Evaluation of the Life Cycle was formed in 2005, supervised by the Institute for Environment and Sustainability of the European Commission's JRC (Joint Research Center) and the Directorate-General for Environment.

Among the most significant works of this collaboration is the publishing in 2010 of the ILCD Handbook (International Reference Life Cycle Data System) (Chomkhamsri et al., 2011), which promotes the application of ISO standards in the field of LCA by exploring many elements of the approach.

A suitable instrument for detecting key environmental features is expressly stated in the COM 2001/68/EC Green Paper and the COM 2003/302/EC on Integrated Product Policy, and is implied, at least indirectly, in the European EMAS (1221/2009) and Ecolabel (66/2010) regulations.

The ISO standards series regulates the LCA. The initial edition of ISO standards was revised several times, the most recent in 2006. In reality, the following 14040 series standards are presently the international regulatory reference for the production of LCA studies:

- UNI EN ISO 14040:2006 ("UNI EN ISO 14040," 2006) Environmental Management - Life Cycle Evaluation - Principles and Frameworks, which provides a general framework for the practices, applications, and limitations of the LCA, and is aimed at a wide range of potential users and interested parties, even those with limited knowledge of life cycle evaluation;
- UNI EN ISO 14044:2006 ("ISO 14044," 2006) Environmental management - Life cycle evaluation - Requirements and guidelines, which was developed for the preparation, management, and critical assessment of the life cycle and serves as the primary support for the actual execution of an LCA research.

Furthermore, the two technical studies stated below are available to support the UNI EN ISO 14040 standards:

- ISO/TR 14047:2003 ("ISO/TR 14047," 2003a) "Environmental management – Life cycle impact assessment Examples of ISO 14042 implementation".
- ISO/TR 14049: 2000 ("ISO/TR 14049," 2000b) "Environmental management – Life cycle assessment – Examples of ISO 14041 application to goal and scope formulation and inventory analysis".

The technical specification ISO / TS 14048:2002 ("ISO/TS 14048," 2002c) "Environmental management – Life cycle assessment – Data documentation format" is also available. Its purpose is to provide the requirements and structure relating to the format of the data, which is used for the documentation and exchange of these during the inventory phase, as well as during the evaluation of the life cycle itself.

The standards are provided by the ISO describe the four main phases of an LCA (Figure 2.2):

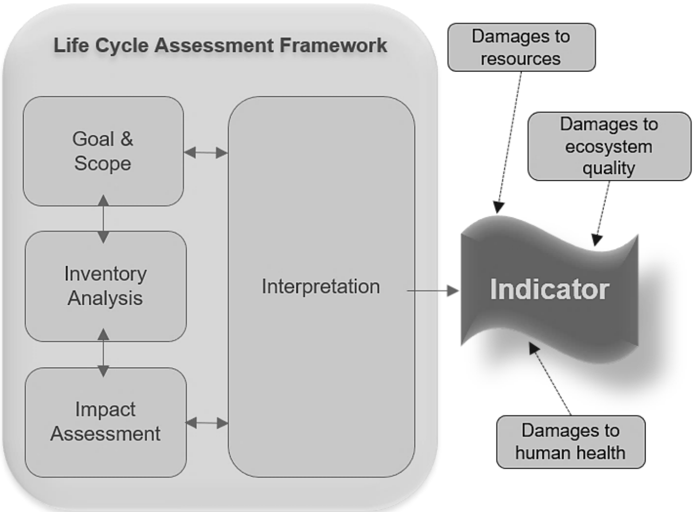


Figure 2.2 - Scheme of the phases of the LCA

1. Goal and scope definition
2. Inventory analysis
3. Impact assessment
4. Interpretation

While the following Figure 2.3 depicts the entire life cycle of a product, process, or activity, including raw material extraction and treatment, manufacturing, transportation, distribution, reuse, recycling, and final disposal, as well as input flows and outputs that are typically considered during the inventory phase.

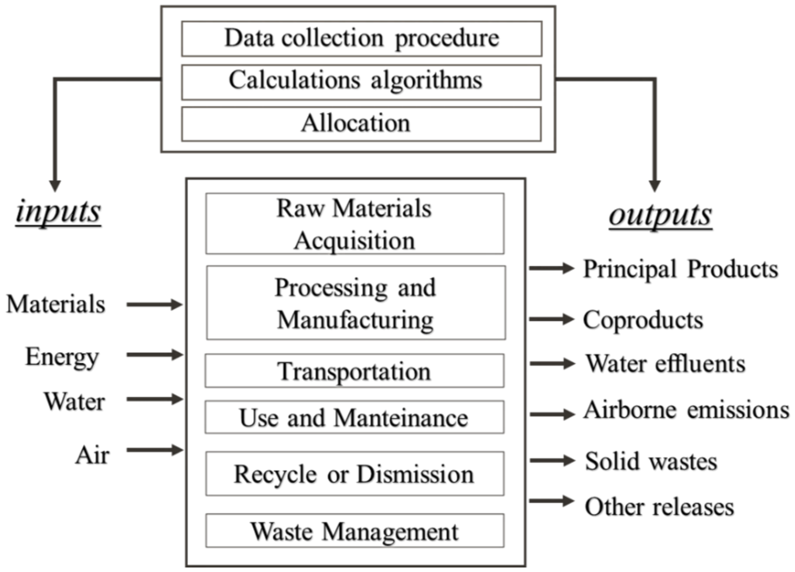


Figure 2.3 - Life cycle of a generic product, process or activity with its inlet and outflows.

LCA is an iterative process that allows to refine things along the analysis. For example, the initial phase of analysis may indicate that more data is required. Alternatively, the evaluation results or interpretation may suggest to change the goal and scope. In this manner, each LCA not only provides vital recommendations for making changes in organization, but it also shows how to effectively adjust the LCA to learn even more.
There are several varieties of LCA. The general rule is that the more detail you desire, the more detailed your LCA must be (Production Engineering, 2019). A report for internal usage has fewer criteria than a report for marketing or other forms of external communication. There are also several LCA-related evaluations, such as Environmental Product Declarations (EPDs), studies that are compliant with product- or sector-specific standards, single-issue analyses such as the carbon or water footprint, social LCA, and long-term monitoring studies. The wonderful thing about a life cycle model is that it can be used to undertake a number of assessments, depending on what your organization requires right now. Every phase of a product's life cycle (extraction of materials from the environment, production of the product, use phase, and what happens to the product after it is no longer used) can have a significant impact on the environment. LCA allows to assess the environmental impact of your product or service from the beginning to the end, or from cradle to grave.

STEP 1 - Goal and scope definition

This initial phase is critical because it emphasizes the fundamental rationale for doing this study, permits understanding of how the data will be used, and specifies the amount of information sought. First, it must be thoroughly described: 1) the desired use; 2) the motives for conducting the study; 3) the sort of audience for whom it is intended; and 4) if the results are to be utilized to perform comparison statements for public dissemination. Before any data is gathered, the purpose and scope of the study are determined (Curran, 2017). It should also be noted that when new information about the product system is collected during the analysis, the purpose and scope will need to be revisited and revised. Issues and facts that were unknown or could not be predicted at the start of the project would necessitate rewriting the target. For example, the initial target established for the test could not be met. This exemplifies LCA's iterative nature (Villares et al., 2017).

SUB-STEP 1.1 - Functional unit definition

The functional unit (FU) is defined as the performance of a system's functional outputs ("UNI EN ISO 14040," 2006). A valid functional unit should be physically quantifiable and accompanied with an appropriate unit of measurement, which might be a unit of time or a well-defined amount if product durability is required. In other words, the coherence of the FU selection must be compatible with the baseline situation and transformation objectives. The significance of a clear definition of the functional unit cannot be overstated, because it is required for the right development of scenarios and the correct normalizing of all reference flows, and therefore for the final outcomes. The FU definition should make it simple to compare the evaluation findings to other values regarded as benchmarks.

SUB-STEP 1.2 - Reference flow and system boundaries

A reference flow is the quantity and kind of energy and materials required by a product, process, or activity to generate the functional unit's performance (Weidema et al., 2004). Several authors (Cooper, 2003; Reap et al., 2008a; Gandiglio et al., 2019) believe the reference flow definition to be inextricably linked to the functional unit definition. The problems concern the specification of product lifespan, performance, and interdependence of the product, process, or activity under consideration. The suggested technique begins with designing and defining the system using a process flow diagram that depicts the links between the unit-processes and the reference flows. System boundaries ((European Committee for Standardization, 2011)) depicted in Figure 2.4.define which processes and activities will be included or omitted from the evaluation. Boundaries have an influence on the breadth and depth of the evaluation, as well as the dependability of the results (Liu and Müller, 2012). The exclusion (cut-off) of factors, i.e. features of the process, product, or activity to be analysed, that are not regarded to affect the impact of the system on the scope of the analysis (Goedkoop et al., 2016), is the determination of system boundaries. The cut-offs should be determined based on the results of a sensitivity analysis of the process or activity's impact on the LCA results (L'Abbate, 2018). The problems in including economic and social components into the LCA are mostly the result of: Social (Dreyer et al., 2006) - dispute on measurements, contextual techniques, and reliance not just on the life cycle but also on company conduct. Economic (Guinée et al., 2011) - lack of scientific or procedural consensus on terms, methodology, and so on; challenges in dealing with externalities, cost allocations, system border compatibility, and possible cost forecast.

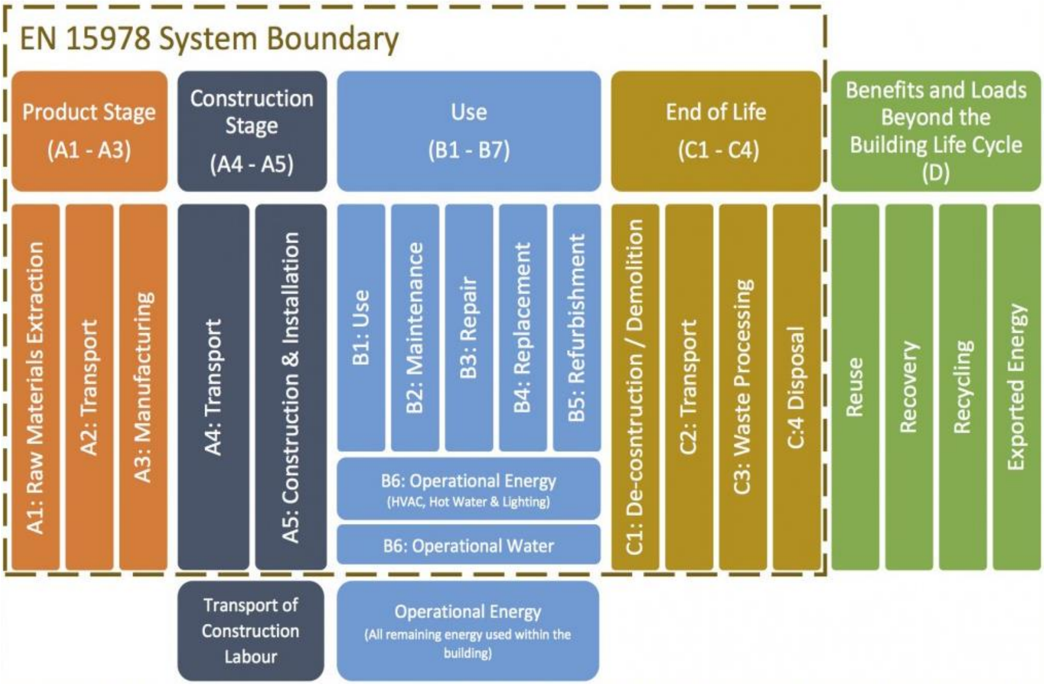


Figure 2.4 - System Boundaries of an LCA according to EN 15978

STEP 2 – Life Cycle inventory analysis

The most challenging step of an LCA is the life cycle inventory (LCI). LCI analysis is the computation of all raw materials, energy resources, and machinery (inputs) utilized in the process, as well as the estimation of their emissions to air, water, and soil (outputs) created across the whole life cycle, using the functional unit as a reference.

Figure 2.5 illustrates the questions (requirements) that must be asked in order to obtain a full inventory of input and output data (information) flows.

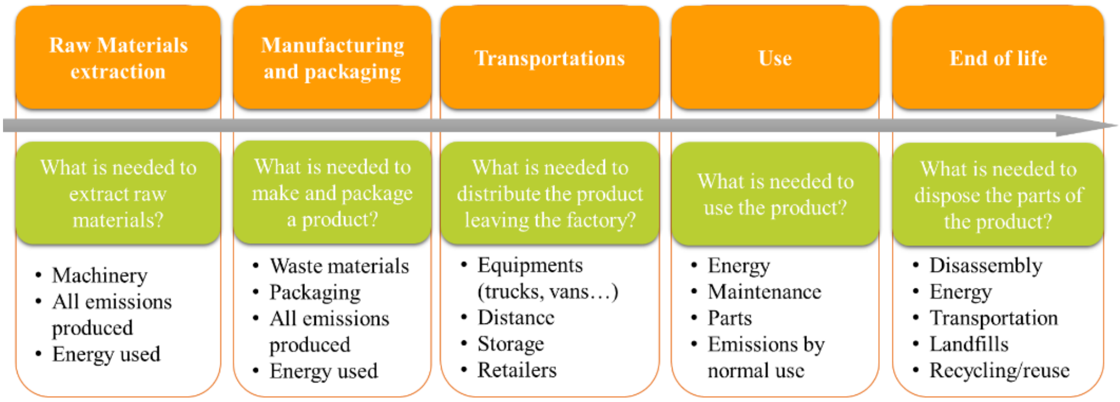


Figure 2.5 - Requirements for a LCI

SUB-STEP 2.1 - Choice of the database

The LCI must give the most accurate and objective picture of reality possible [21]. The accuracy of the data is critical in the model's implementation. An LCI analysis necessitates a thorough examination of each process flow and the emissions that emerge from it. Numerous databases have been developed over the years to assist the collecting of data on raw materials and auxiliary materials, energy, transportation, machinery, and so on by research centers (e.g., ELCD (Mathieux et al., 2013)), organizations, and volunteers (e.g., Ecoinvent ("Ecoinvent," 2019) or WorldSteel (World Steel Assoctiation, n.d.)). Some existing datasets have formerly been incorporated into LCA tools such as SimaPro® or GaBi®. The database enables the selection of raw materials, transportation networks, energy mixtures, and even complete processes. Each item on the list already includes a set of information (Input/output matrix) regarding the steps of raw material extraction and processing, making inventory analysis a lot easier. The availability of data at the local and sectoral levels is very important. Before beginning the real inventory analysis, it is best to pull every data point to be included to the inventory from the database that best suits its context, keeping in mind the declared purpose and scope of the evaluation. Considering the production procedures of items. Table 2.1 provides a selection of the most useful databases. It is important to emphasize that, in order to justify any lack of consistency in the data (which reflects on the results of the analysis), specifying the version of the reference database is essential during the Inventory phase, because the Input/output matrices are constantly updated and can drastically influence the magnitude of the emissions and environmental impacts.

SUB-STEP 2.2 - Allocation

The allocation refers to techniques for allocating the environmental load of operations with several outputs. To prevent allocation, ISO 14040 suggests separating operations into sub-processes or widening system boundaries. In LCA, consequential modelling avoids allocation, whereas the allocation technique is known as attribution modelling. See (Schrijvers et al., 2016) for further approaches. Attributional LCA tries simply to reduce the product's worldwide environmental effect. The purpose is to explain the relevant physical flow in the environment. It is necessary to utilize average data. Consequential LCA tries to record changes in environmental effect caused by a specific activity and, as a result, create information on the consequences of activities. It is necessary to use marginal data.

SUB-STEP 2.3 - Local technical uniqueness

Disparities in location can represent differences in extraction, production distribution, and end-of-life technology. These changes in geography, businesses, facilities, and production lines might have a significant influence on the Inventory phase and ultimately the LCA outcomes. To have as little uncertainty as possible, the data for the LCI should be geographically characterized (see, for example, (Nilsson et al., 2010)). Countries all across the world have produced a diverse set of rules and regulations. The preceding are well-known country-specific standards (British Standards Institution, 2011; European Commission and Joint Research Centre, 2010; “Requirements for the EcoLeaf PCR,” n.d.), which give clear definitions and criteria to reduce choice flexibility and support the accuracy of LCA outcomes and quality assurance related to these throughout their life cycle.

According to ISO 14040, the inventory analysis in conducted following the equation:

$$\underset{\text{(MJ or Kg)}}{\text{Amount}} \times \underset{\text{(g/MJ or g/kg)}}{\text{Emissions}} \times \underset{\text{(from databases)}}{\text{Characterisation factor}} = \underset{\text{(g-eq)}}{\text{Equivalents}} \qquad \text{Eq. 2.1}$$

Table 2.1 - Benchmark of databases for LCI in manufacturing context up to the end of 2020

	Database Name	Link	Last update	Open source	Category of materials and processes	Nr of data items
	Ecoinvent	https://www.ecoinvent.org	2020	no	Metals, aluminium, paints, energy, transports, textile industry, waste management	~17.000
	ICE	http://www.bath.ac.uk/mech-eng/sert/embodied	2019	yes	Construction materials	~1.700
	ELCD	https://eplca.jrc.ec.europa.eu/ELCD3/	2013	no	Transports, wastes, organic and inorganic materials, wood, semi-metals	~500
	USLCI	https://www.nrel.gov/lci/	2009	yes	Processing metals, finishing, washing processes, paintings	~100
	AusLCI	http://www.auslci.com.au/	2016	yes	Agricultural products, energy, fuels	~400
	IDEA	http://idea-lca.com/?lang=en	2016	no	Forestry, Mining, ceramics, gas, water, sewerage	~3.900
	Material Universe	https://grantadesign.com	2019	no	Polymers	~3.700
Industry Data	Plastics Europe	https://www.plasticseurope.org/it	2010	no	Plastics	~85
	WorldSteel	https://www.worldsteel.org/	2018	no	Steels	~45
	ERASM	http://erasm.org/index.php/about-surfactants/value-chain	2016	yes	Detergents, surfactants, chemicals	~70

STEP 3 - Life cycle impact assessment

The goal of life cycle impact assessment is to establish a link between the right burdens and the right affects at the right time and location (Reap et al., 2008b).
Three matrix equations can be used to represent life cycle assessment (LCA) (Heijungs and Sun, 2002):
The first equation is used to convert process data into a manufacturing system:

$$s = A^{-1} \cdot f \qquad \text{Eq. 2.2}$$

Where
A: database of process flows and manufacturing processes
f: final demand vector, or the intended output from the system
s: scaling vector, which represents the intensity of manufacturing processes
The scaling vector derived from the first equation is used to calculate the intensity of emissions from unit processes

$$\mathrm{g} = \mathrm{B} \cdot \mathrm{s} \qquad \text{Eq. 2.3}$$

Where
B: unit emission matrix (a database of process values)
g: emission inventory vector detailing the emissions produced by the entire system
The third equation is used to convert emissions into environmental consequences (e.g., CO_2 emissions into climate warming potential)

$$h = Q \cdot g \qquad \text{Eq. 2.4}$$

Where
Q: matrix of characterisation (impact intensity characterization values)
h: a vector indicating the system's environmental impacts.
To make the obtained affects more understandable, they might be normalized by dividing by a geographical reference impact vector (e.g., The climate change potential of the EU countries in 2004).
Finally, the normalized results may be weighted using value-based weights to provide a single environmental performance indicator.
The data is transformed from inventory to potential impact ($EP(j)_i$) by multiplying the input/output of a specific substance (Q) with its equivalent factor ($EQ(j)_i$) (characterisation):

$$EP(j)_i = Q \times EQ(j)_i \qquad \text{Eq. 2.5}$$

While the total emission is the sum of every potential impact during each phase of life cycle:

$$\sum_i^n EP(j)_i = Q \times EQ(j)_i \qquad \text{Eq. 2.6}$$

SUB-STEP 3.1 - Impact category selection

This stage involves associating data acquired during the LCI with the appropriate effect category, such as global warming, acidification, human toxicity, waste resource use, and so on. The main challenges in selecting an impact category are the absence of standardization, the disuse, and the selection of a mid-/end-point.

SUB-STEP 3.2 - Space characterization

The computation of impacts and their impact on the environment might be heavily reliant on spatial characterization. The characterization of meteorological, topology, hydrology, and land use status all influence the estimation of consequences, such as acidification, eutrophication, and health implications. (Bartolozzi et al., 2013) provides an example of environmental uniqueness consideration. For example, the outcome of resource extraction or pollution might vary based on the distinctiveness of the local environment, such as soil characteristics or population density.

SUB-STEP 3.3 - Time characterization

This problem is about the temporal characterization of transformations and the effect evaluation criteria. During the analysis, the duration of the transformations in each unit-process and, ultimately, the useful life of the functional unit should be considered as a change in emissions caused in narrow or prolonged time arcs, thereby also altering potential environmental damage in the short, medium and long term.

STEP 4 - Life cycle interpretation

The life cycle interpretation process entails the study of inventories and the implications to support a product/process selection, an enhancement of some product feature or process element, and so on (Hellweg et al., 2005). According to ISO 14044 ("ISO 14044," 2006), the interpretation means to: a) identify significant issues based on the LCI and LCIA phases; b) critically evaluate the overall analysis itself, how complete it is, and if it is done sensitively and consistently to the goal and scope and the determined requirements; and c) provide conclusions, limitations, and recommendations.
Understanding the precision of the data and ensuring they achieve the study's objective are the first steps in interpreting LCA findings. This is achieved by determining the data elements that relate significantly to each effect group, determining the importance of these significant data elements, assessing the completeness and accuracy of the analysis, and drawing conclusions and guidelines based on a detailed view of how the LCA was performed and the findings were developed (Hauschild et al., 2018).

2.1.2. Exergy Analysis

2.1.2.1. Real thermodynamic systems and their irreversibility

To evolve, life has always taken use of whatever resource that the Earth has offered, even if it has been compensated by solar energy, creating trash with a higher entropy content than the starting state. Outputs can become more resources, but this cycle cannot be infinite since the gap between initial and final availability becomes too large at some point, i.e., the necessary resources are no longer sufficient for further transformation.
Real transformations are distinguished by irreversibility: if a system undergoes a transformation that moves a state variable from state A to state B (Figure 2.6), it can be returned to state A, resulting in a thermodynamic cycle. The system will be returned to its previous level of entropy, but the environment's entropy will grow.

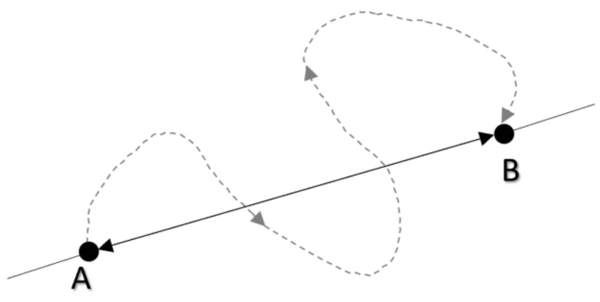


Figure 2.6 - Ideal transformation (in continue line) and real transformation (in dotted line) from state A to state B

Time appears to be unimportant in the study of the first law of thermodynamics; but, in thermodynamics, we deal about variables and state functions that vary and transform so quickly that in irreversible thermodynamics, considerable emphasis is paid to the 'speed of a process' (Petrescu et al., 2016). Real processes change with the largest variation in entropy in the quickest period, therefore designing a

sustainable process implies creating the least amount of entropy (and hence causing the least number of environmental difficulties) over its life cycle.
Entropy increase principle is equivalent to the extropy decrease principle formulation of the Second Law (Martinas and Frankowicz, 2001): imagine the Earth as a reservoir containing material resources, energy, and the ecosystem, as shown in Figure 1.

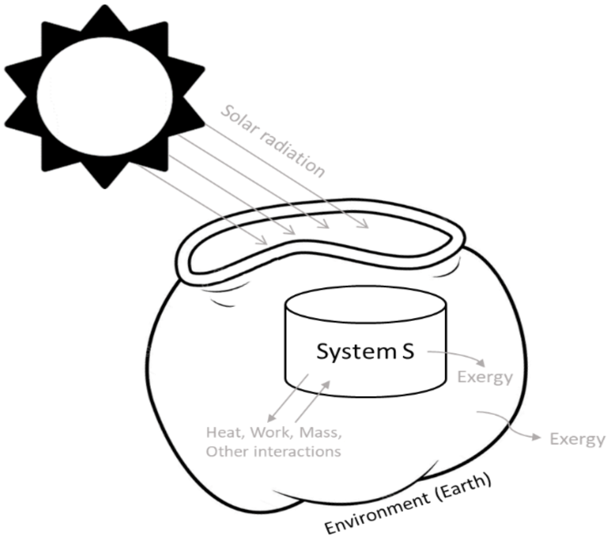


Figure 2.7 - Theoretical model of the system and the environment interactions

The reservoir is classified as an open system in thermodynamics because it exchanges energy and matter with its surroundings (Bertalanffy, 1950). Following a particular transition over time, the reservoir would contain matter and energy in states other than the initial ones, causing an imbalance in the environment that would have an effect. In an open system, there is no possibility of thermodynamic equilibrium; instead, it is all about thermodynamic non-equilibrium (Martinás, 1997).
In an ideal world, the entropy of the system under study and the entropy of the system with which it interacts are identical in form but have opposite signs, because one emits heat and the other absorbs it. As a result, the system's cumulative shift is zero. In practice, the net increase in entropy is positive because the entropy value of the system that produces work (which is positive) is larger than that of the system under examination (which is negative). As a result, in the actual world, a change that occurs in a non-isolated system generates a decrease in entropy in the physical system and an increase in entropy throughout the universe. Because manufacturing processes are artificial non-equilibrium systems, the same idea applies. The process or product are closed systems in and of themselves, but because the production pulls resources from the environment in which it operates and returns losses and waste to it, it must be regarded an open system.
Combining this principle with sustainability thinking, while entropy is not a natural state variable and contains statistical assumptions in its definition (Martinás and Grandpierre, 2007), extropy is more 'physically sound' because it quantifies the distance from equilibrium, i.e. the degree of irreversibility (Poór, 2005). Because thermodynamic reactions have a preferred path to achieve equilibrium, this distance is not regular. Extropy is closely connected to exergy: the creation of extropy is a measure of the extent of a process's irreversibility. For a steady-flow process, entropy generation is expressed as (Prasad et al., 2009):

$$S_{gen} = \sum_{out} m_e \cdot S_e - \sum_{in} m_i \cdot S_i - \sum_i \frac{Q_i}{T_i} \geq 0 \qquad \text{Eq. 2.7}$$

The first two terms on the right are the sums of exergy outputs and inputs, and the third is the rate of entropy transfer over the part of the control surface where the instantaneous absolute temperature is T_i. $Ex_{loss} = T_0 \cdot S_{gen}$ may be used to calculate the exergy. Extropy is a measure of the entropy produced by the fictive process that brings a physical system into balance with its surroundings. As a result, it is equal to the difference between the maximal entropy of the system and the total entropy of the system and the environment:

$$\Pi = S_e^S + S_e^E - S^S - S^E \qquad \text{Eq. 2.8}$$

Because manufacturing processes are artificial non-equilibrium systems, the same idea applies. The process or product are closed systems in and of themselves, but because the production pulls resources from the environment in which it operates and returns losses and waste to it, it must be regarded an open system. These considerations allow to define the efficiency indicator (η_l) as the outcome of a linear analysis (shown below as the horizontal line in Figure 2.8): a simple ratio of output to input, final state to beginning state (state B and A as described above). While the vertical arrow, which represents the conceptual model of the transformation under consideration, investigates the dynamic features of the transformation and subsequently investigates its thermodynamic laws (which would mean constructing the real trajectory of the transformation).

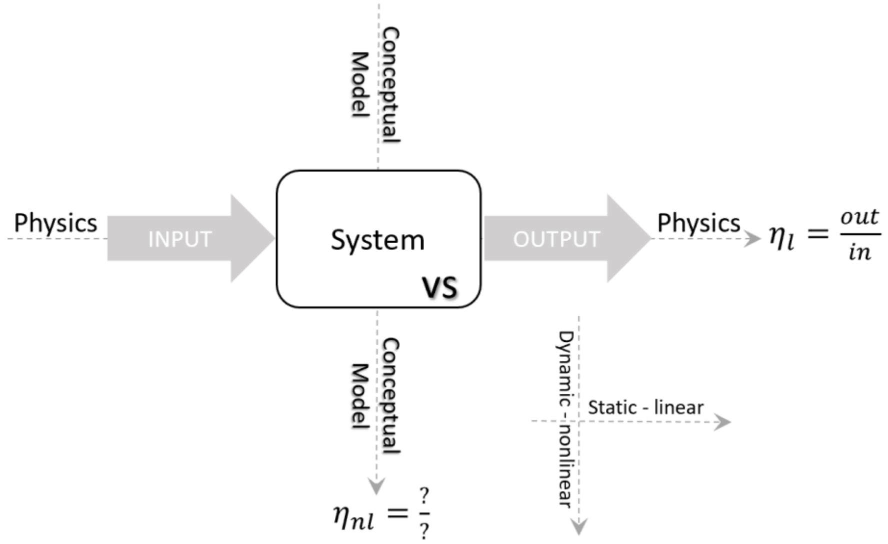


Figure 2.8 - General scheme of linear and non-linear features in an industrial process

Complex systems' non-linear dynamics make it exceedingly difficult to anticipate their evolution over time. Hybrid evaluation models, such as those provided in this paper, generally depict the dynamics of a system and may be used to forecast multidimensional system trajectories by working on a wide range of parameters. As a result, the transformation's probable consequences may be predicted.

2.1.2.2. Exergetic Analysis

When we think about energy, we think in terms of quantity. However, in a resource-constrained reality, energy must also be valued in terms of quality, which is basically a measure of its utility, or capacity to do labour. Exergy must be measured in order to account for the quality of energy rather than just the amount. Exergy analysis may be applied to individual processes, companies, and even whole national economies (Sousa et al., 2017). It provides a solid foundation for evaluating the impact of policy actions aimed at increasing energy, resource, and climate efficiency. In the future, consumers may be told about items and services based on their exergy-destruction footprint in the same manner that they are educated about carbon emissions (Brockway et al., 2016).

The notion of exergy and its application to energy efficiency was discussed in 2015 at (Science Europe Scientific Committee for the Physical, Chemical and Mathematical Sciences, 2015). In doing so, the Committee appealed to policymakers to establish an International Exergy Panel to: bridge the gap between energy science and energy policy, resulting in the systematic use of the concept of exergy where appropriate; provide an evidence-base for interrelated energy-climate change and economic policies; and drive interdisciplinary research and development on the causes of exergy destruction and how we can minimize this destruction, from the molecular to the global scale; direct the development of exergy footprints for commodities and services; and engage with the Intergovernmental Panel on Climate Change (IPCC).

The idea of exergy is intricately linked to the fundamental laws of thermodynamics. These laws must not be disregarded: they are vital. First law: energy is preserved. Second law: heat cannot be completely turned into useful energy. The second law deals with the idea of exergy. Exergy is destroyed in every energy-conversion process.

Exergy is defined as the maximum amount of work which can be delivered by a system or a flow of matter or energy when it reaches the equilibrium with a reference environment through a sequence of reversible processes in which the system can only interact with (Rant, 1956). Due to the irreversibility of the process, some of the exergy is wasted or destroyed during the process (Kotas, 2013). Exergy Analysis in a manufacturing system seeks to find and analyse thermodynamic flaws (irreversibilities) and to reveal opportunities for improvement. Furthermore, exergy efficiency is an essential criterion for determining the sustainability of a process or product. According to (Duflou et al., 2011) and (Renaldi et al., 2011), the application of exergy in manufacturing systems allows for the detection and evaluation of thermodynamic flaws as well as the identification of possibilities for improvement. The measurement of resource consumption and the consequences of emissions may be represented on a single objective scale using the second law of thermodynamics, distinguishing it from the LCA, which examines the diverse effects by quantifying them on several scales.

While exergy destruction is never zero in any process, it may be minimized. Every process leaves a distinct exergy destruction footprint. This footprint may be used to rationalize resource selections before to production and to monitor the usage of energy and resources throughout production. It may be utilized in a whole life-cycle approach to examine a product's overall energy and resource 'cost': basically, its exergy destruction footprint. It is critical to remember that there can be no output without an exergy destruction footprint. When designing more environmentally friendly technology, a deliberate attempt to decrease exergy degradation to a bare minimum is a goal to strive towards.

Given the triple approach to sustainability (TBL), even exergetic analysis bridges well this concept; in fact, (Morosuk and Tsatsaronis, 2012) graphically depicted some conceivable interdependencies among exergy, economics, and environment already in 2012. These dependencies are shown in Figure 2.9

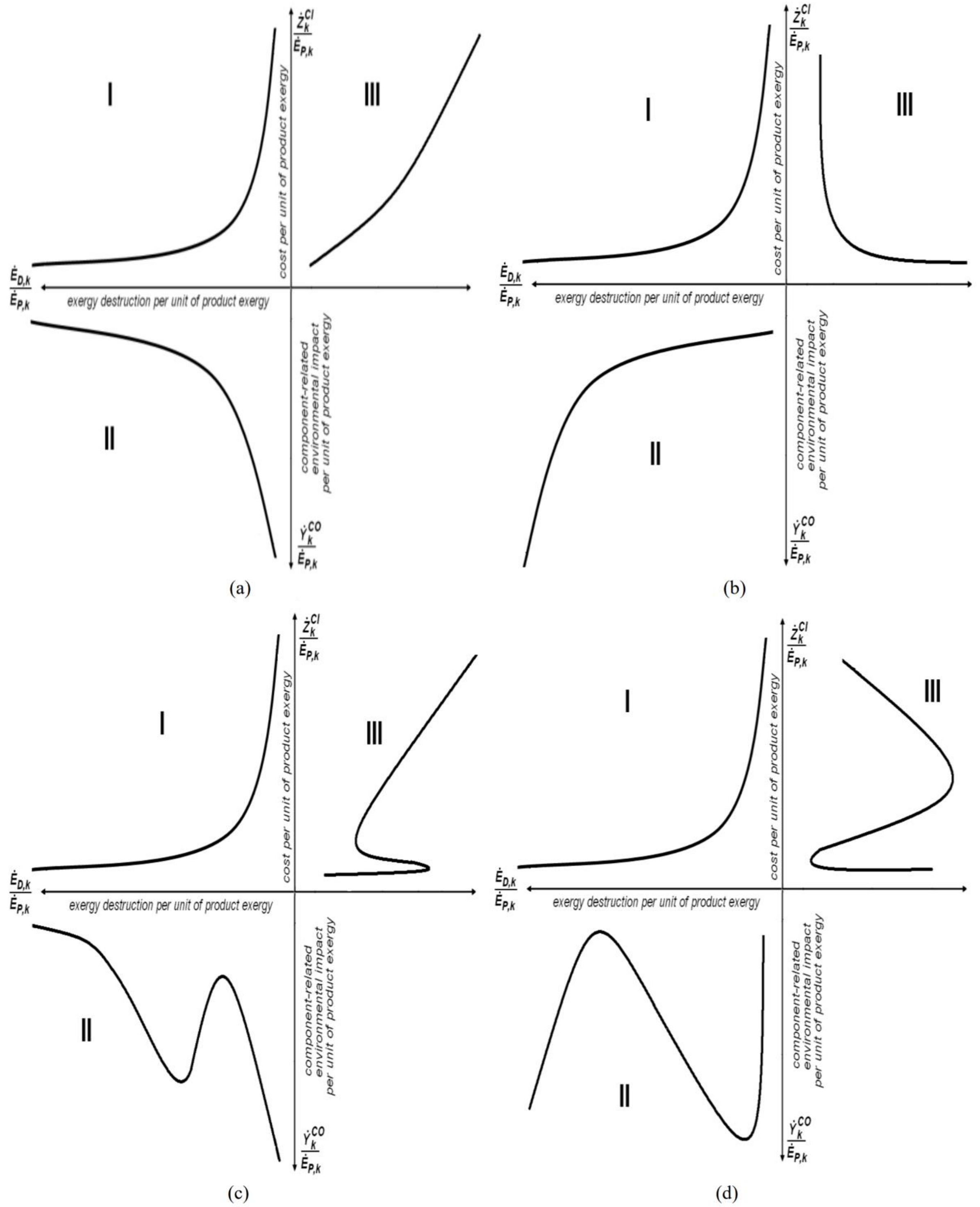


Figure 2.9 - Interdependencies between exergy destruction and the three dimensions of the TBL paradigm

A first approach may appear to be that exergetic analysis provides a more abstract interpretation of concepts already known as the thermal efficiency of a thermodynamic cycle, but exergetic analysis is a useful investigation tool to identify which They are the causes that determine the plant's thermal efficiency reduction. The use of exergetic analysis is exactly in its capacity to detect exergy to the extent The potential to create work and, thus, as a measure of energy quality. The other significant attribute is its relationship with thermodynamic properties (such as pressure and temperature), for which it is also a thermodynamic property. It is therefore feasible to measure the exergetic loss owing to the

irreversibility of the Transformation in complex systems, offering indications on its origin and impacts on the whole process, in order to pinpoint the components on which intervention is most suitable.
The fundamental principle of thermodynamics affirms the ultimate conservation of energy: it, like matter, cannot be generated or destroyed; nevertheless, it may be controlled to change the form in which it appears. In actuality, throughout thermodynamic processes, entropy is generated, and as a result, the quality of energy is degraded, as defined by the second thermodynamic principle. Consider the energy of a material flow as a measure of its potential utility, and you will be more easily induced. If you examine the same energy term of electricity and hot water, they have a very different quality of energy: unlike the electric current flow, the utility that can be gained from the scope of hot water is principally subject to its temperature and environment of reference. In reality, while it is theoretically conceivable to transform a full electrical flow into mechanical work, the same cannot be said for hot water, from which reversible work may be derived based on its Carnot limit (Boateng, 2016).

$$W_{max} = Q \cdot \frac{T - T_0}{T} \qquad \text{Eq. 2.9}$$

For these reasons, exergy, defined as "equivalent reversible work," has become a standard for measuring not only the quantitative but also the qualitative aspects of energy exchanges between many systems or between a system and the environment over the years.

Figure 2.10 depicts a generic open system in equilibrium, the state of which is characterized by specific values of its physical and chemical attributes. The system interacts with its reference environment, which has certain physical and chemical features.

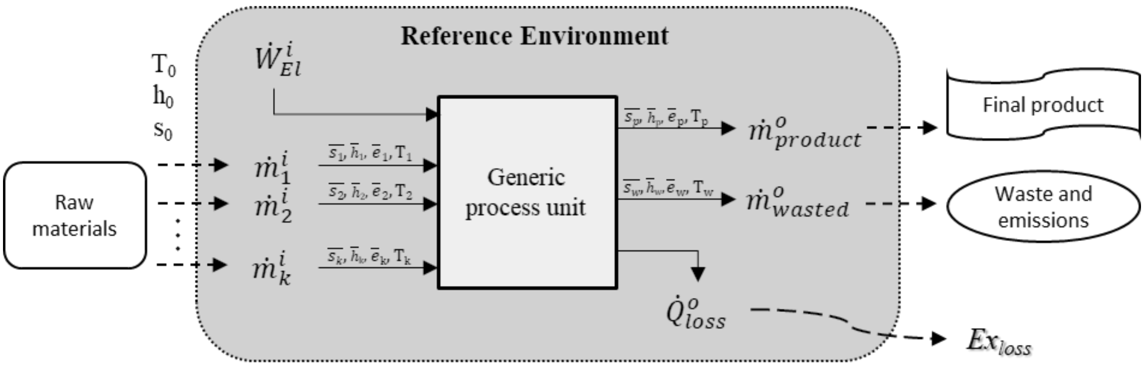


Figure 2.10 - Generic open system control volume

In addition, a generic system, such as the one seen in Figure 2.10, can interact with its reference environment in two ways:

1) Non-material interaction: any interactions that do not include a flowing flow rate. Because exergy is defined as the reversible work that may be taken from a system, mechanical work and exergy are equivalent. They can be further subdivided into:
   - Thermal interactions: the system can exchange heat via conductive, convective, or radiative modes on its second side. In this scenario, the maximum work done by a thermal exchange may be specified. It is feasible to acquire from the thermal flow studied, utilizing the reference environment as a thermal tank with an unlimited thermal capacity; heat is the word given to exergy associated with a thermal flow. Considering the environment as a thermal well at the reference temperature T_0, with a given thermal power Q and a given control surface temperature T, heat exergy is computed as $Ex_Q = \dot{Q} \cdot \partial$, where ∂ it is referred to the Carnot factor, or non-dimensional exergetic temperature. When the temperature of the

control surface through which the heat exchange occurs is T, the heat exergy describes the transfer of exergy associated with the transmission of heat. The direction of the flow of heat exergy will be determined by the following factors: I) the sign of ∂: the Carnot factor will be positive if $T > T_0$, and negative if $T < T_0$; II) the direction of the thermal flow Q: positive or negative according to the signed conventions employed.

- Mechanical interactions: As a result of the work transfer, exergy transfers can be characterized in the direction and form to which they correspond.

  The work exergy is defined as $Ex_W = \dot{W} - P_0 \cdot \frac{dV}{dt}$. According to the signed conventions, an emphasis on the system has a positive sign and corresponds to outbound exegetic power, which is likewise positive. If the system is rigid, the second-term is null, and the work is completely accessible, matching to exergy. Mechanical work can take many forms, including system volume fluctuation work, shaft rotation work, electrical work, magnetic work, and so on. If the system is not rigid, it can mechanically interact with its surroundings by modifying its volume. Interactions between mechanical elements can, however, occur in the presence of a non-rigid system, such as spinning trees or electricity fluxes;

2) Material interaction: it occurs whenever a state material flow and chemical composition different from those of the ambient pass through the system, interacting by work, heat, or chemical species exchanges. In the absence of nuclear, magnetic, electric, and surface tension effects, exergy associated with a material flow is a property of the system; it can be expressed as the sum of kinetics, potential, physical, and chemical exergy. Because it is related to a moving mass, exergy may also be represented as a specific measure.

The interaction between a system and its reference environment can thus occur through the system control surface in different modes: energy exchanges (decreased in work or heat) or mass exchanges.
Each of these interactions is associated with a certain quantity of exergy, which represents the amount of reversible work that may be achieved from the interaction or as a group of interactions considered in the analysis. The overall exergy associated with the system is therefore sum of three contributions associated with work, heat and mass flows.
Instead, in the case of manufacturing systems, which is the major issue of this thesis, it is conceivable to neglect specific potential and kinetic exergy contributions and focus just on specific chemical and physical exergy (referring to material and energy flows).
The foundation of EA is stated by the first and second principles of thermodynamics. The first law deals with energy conservation, while the second deals with the quality of energy and materials. These thermodynamic rules underlying the EA are critical for tracing the set of parameters that must be measured and monitored throughout the process, as well as the variables that may be derived. According to Szargut's research (Szargut et al., 1987), reference flows can be uniquely recognized in the balancing equations below.
Eq. (4)'s mass flow balance explains the balance for the investigated system of in and out material flows.

$$\sum_k \dot{m}_k^i = \sum_k \dot{m}_k^o \qquad \text{Eq. 2.10}$$

Energy flow balance is reported in Eq. 2.11. Because energy is a broad variable, the energy of a system in a given state equals the sum of the energies of all subsystems that can be identified as being a component of a particular system (Bakshi et al., 2011). The entire energy content of an isolated system cannot change, as stated by the first rule of thermodynamics: energy is conserved (Terzi, 2018). As a result, energy may only be transformed or converted from one kind to another, with no regard for energy quality loss. To identify and quantify the irreversibility, an EA must be performed. Closed material and energy flow balancing with energy interactions (work and heat) between incoming and outbound flows from the system boundaries must be performed to accomplish it.

$$\sum_c \dot{H}_c^i + \sum_d \dot{W}_d^i + \sum_p \dot{Q}_p^i = \sum_c \dot{H}_c^o + \sum_d \dot{W}_d^o + \sum_p \dot{Q}_{p,tot}^o \quad \text{Eq. } 2.11$$

Exergy flow balance is stated in Eq. 2.12. Because it is neither visible or even feasible in nature, the idea of equilibrium is frequently questioned. Unlike energy, exergy is not preserved. In any genuine process, it is instead consumed or destroyed to some extent. As a consequence, the quantity of exergy destroyed by the system may be calculated by accounting for all of the exergy streams in the system. The exergy loss is proportional to the entropy created; the lost exergy, or produced entropy, is responsible for the system's less-than-theoretical efficiency. When work is exchanged or heat transfers occur, the exergy destruction rate ($\dot{E}x_{loss}$) may be calculated by balancing the exergy between inbound and outgoing flows.

$$\sum_c \dot{E}x_c^i + \sum_d \dot{W}_d^i + \sum_p \left(1 - \frac{T_0}{T_e}\right) \dot{Q}_p^i = \sum_c \dot{E}x_c^o + \sum_d \dot{W}_d^o + \sum_p \left(1 - \frac{T_0}{T_e}\right) \dot{Q}_p^o + \dot{E}x_{loss} \quad \text{Eq. } 2.12$$

In the term $\left(1 - \frac{T_0}{T_e}\right)$, T_0 is the dead state reference temperature and T_e is the equilibrium temperature, as defined in Eq. 2.13.

$$T_e = \frac{T_k^i + T_k^o}{2} \quad \text{Eq. } 2.13$$

The following equations are used to compute the enthalpy flow rate (Eq. 2.14), specific entropy (Eq. 2.15), and exergy (Eq. 2.16):

$$\dot{H} = \dot{m} \cdot c \cdot (T - T_0) \quad \text{Eq. } 2.14$$

$$\bar{s} = c \cdot \ln\left(\frac{T}{T_0}\right) \quad \text{Eq. } 2.15$$

$$\dot{E}x = \dot{m} \cdot [\bar{h} - \bar{h}_0 - T_0 \cdot (\bar{s} - \bar{s}_0)] \quad \text{Eq. } 2.16$$

The process or its components' performance indicators are stated in the following net and general efficiencies (Eq. 2.17 and Eq. 2.18), depending on whether the goal is to assess the percentage of relevant exergy for the realization of the final product or the total exergy of the process:

$$\eta_n = \frac{\sum Ex^{product}}{\sum Ex^{in}} \quad \text{Eq. } 2.17$$

$$\eta_g = \frac{\sum Ex^{out}}{\sum Ex^{in}} \quad \text{Eq. } 2.18$$

The optimization criteria entail minimizing the term Exloss, which is the source of the process's less-than-theoretical efficiency. Temperature changes are important in the exergetic equilibrium. The bigger the temperature difference between two transition phases, the more energy is generated. The energy balance in Eq. (6) is also significant for optimizing product quality. This means that energy analysis makes possible at the same time to increase the quality of the finished product, to control its characteristics and to reduce the energy costs of the process (Kamps et al., 2018).

Each complex manufacturing system is distinguished by the combination of multiple elementary subsystems. Attention should be made to its subsystems as well as the overall system in order to evaluate it from an energy standpoint. It is an issue of accurately identifying the control volume each time, since the object of the research varies: the writing of the exergetic budget and the computation of its efficiency

are consequences and functions of the control volume choice. It's worth noting that, while the energy-type efficiency yield or characteristics allow to compare machines of the same kind; the exergetic yield, which also expresses energy quality, allows you to compare machines of various types.
The following are the primary benefits of using exergetic analysis:

1) The ability to compare different energy systems, such as direct cycles and inverse cycles;
2) Possibility of locating and quantifying the real sources of system inefficiency, giving helpful information, and properly resolving the resource expenditure to improve the system's effectiveness.

To summarize and conclude, exergy is a unit of measurement for usable energy. The true efficiency of an energy system or process is referred to as exergy efficiency. In this regard, the second-law exergy technique, when contrasted to standard first-law thermodynamics energy approaches, can uncover and quantify the reasons of inefficiencies. Exergy is thus the appropriate statistic for valuing energy consumption and resource scarcity. However, the practical implementation of exergy for resource evaluation meets numerous challenges: first, any exergy accounting involves the establishment of a reference environment that is in thermodynamic equilibrium (Gaudreau et al., 2012). This proves to be rather difficult: can the planet in its current condition (not in equilibrium) be used as a reference setting, or should a hypothetical earth in thermodynamic equilibrium/maximum entropy be assumed? Second, because chemical exergy comprises both enthalpy and resource concentration, exergy combines the two characteristics of energy content (formation enthalpy/heating value) and availability (concentration). This is consistent with the nature of exergy, but it might be deceptive when used to resource estimation (Finnveden et al., 2016). Energy resources have a high exergy value, whereas even rare but very inert (= low enthalpy) substances have a low exergy value that can only be estimated by looking at their concentration exergy. Although exergy continues to be of relevance for resource depletion accounting, being a scientific thermodynamic method with few assumptions, it is objective (Peters, 2021).

2.1.3. Hybrid Exergy-Life Cycle Assessment

Sustainability is defined by a dynamic multidimension, and no apparent, simple solution appears to be capable of dealing with its entire complexity. To collect detailed knowledge from a manufacturing process in terms of productivity, performance, quality, and reversibility, a variety of evaluation approaches are used independently or in a hybrid way throughout the multidimensional sense of sustainable development. Hybrid modelling is another viable option for balancing bottom-up and top-down evaluation methodologies.
According to the preceding paragraphs, while traditional LCA tools place a strong emphasis on emissions, EA focuses on resource and product availability and utility, and is thus efficiency oriented (Moya et al., 2013). Each approach reflects on the same problem: integrating two distinct points of view may lead to the usage of the combined methods' strengths while decreasing the flaws of the individual ones (Milanovic et al., 2017).
The following issues were raised concerning hybrid methods: how are they employed in case studies? What is the advantage of a hybrid analysis over a traditional one? How effectively are LCA and EA mathematically and in terms of the flows to be evaluated integrated? Which starting hypothesis governs the selection of the best hybrid method? Is there a superior one than the others?
The dictionary definition of “Hybrid” (as a noun in the early 17th century) is: "from Latin 'Hybrida,' meaning 'bastard,' of unclear derivation." Something created by merging two or more distinct parts. Characteristics that are conflicting. A term made from of components from other languages, such as television (tele from Greek, vision from Latin). A vehicle powered by both a gasoline engine and an electric motor" (Cambridge English Dictionary, n.d.). Composite, cross-bred, interbred; compound,

combined, blended, mongrel, impure are synonyms. As a result, a hybrid technique of evaluation entails combining, merging, or cross-breeding one or more approaches to offer a comprehensive picture of the same system or phenomena defined by diverse data types and dynamics. Otherwise, it may be thought of as a model involving linked approaches based on many methodologies, each expressing one aspect of the system in the most appropriate way in relation to reality, with a level of interconnection ranging from basic comparison to entire fusion (Vincenot et al., 2017).

2.1.3.1. Hybrid models state of art

Since now, there have been several hybrid techniques that integrate EA and LCA that are available in the literature. In Figure 2.11, the hybrid techniques indicated below, together with their respective authors, are shown in sequential sequence.

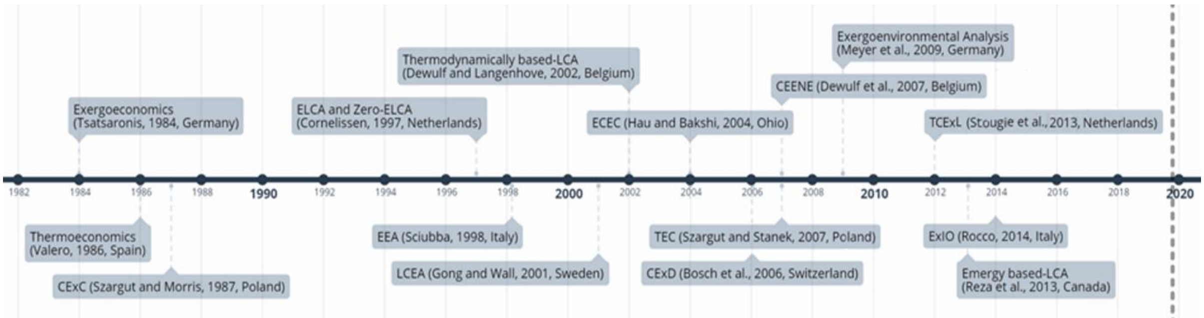


Figure 2.11 - Timeline of hybrid EA and LCA methodologies discovered in the literature

- (Valero, 1986) developed *Thermoeconomics (TE)*, a monetary costing system that blends the ideas of the second law with traditional cost accounting methodologies. He began by considering that the economic dimension, because it involves matter, energy, entropy, and other externalities, should also be modelled as a thermodynamic process, and thus integrates thermodynamics and cost accounting principles, measuring the expense of the product or process with either exergy accounting. Exergy is referred to as TE because, according to the 'exergy costing theory,' exergy represents the true usage value of goods. The objective is to establish the optimal balance between system performance and overall costs. One disadvantage of TE is that the costs of environmental restoration are inconsistent since the exergy of an outflow and its toxicity may not have a thermodynamic-physically determined relationship (Sciubba et al., 2012).
- (Szargut and Morris, 1987) made the first attempt to merge EA and LCA in 1987, developing the *Cumulative Exergy Consumption (CExC)*. This method, which is equivalent to Valero's exergy cost method (as stated by (Stougie, 2011)), consists of a set of balance equations expressing the cumulative consumption characterizing the process as a sum of the cumulative consumption connected to the natural resources extracted directly from the environment. Szargut was able to calculate the process's Cumulative Degree of Perfection (CDP) using this approach.
- The *Cumulative Exergy Demand (CExD)*, created by (Bösch et al., 2007), is an analogous approach of CExC that is currently included as an effect assessment method in all main LCA software packages. It calculates the overall exergy drawn from nature to produce a product by adding the exergy (chemical, kinetic, hydro-potential, nuclear, solar-radiative, and thermal) of all energy carriers used in the process. The last impact category is divided into eight resource groups (fossil, nuclear, renewable, hydropower, biomass, water, minerals, and metals). Its interpretation entails assessing both the quality and amount of the resources required by the process in order to produce a certain functional unit. One of the fundamental disadvantages of this technique is that the social need for a resource, as well as its technical supply or scarcity, are not taken into account in CExD.

- *Exergetic Life Cycle Assessment (ELCA)* and *Zero-Exergy Emission ELCA (Zero-ELCA)* methods were developed by (Cornelissen, 1997) in his PhD dissertation. The ELCA uses the same structure as the LCA, but is evaluated in terms of life cycle irreversibility, i.e., the exergy loss of a product across its entire life cycle. ELCA's inventory analysis is more comprehensive than LCA's; there is no need for classification during the impact assessment since the process allows for the computation of energy and matter flow in a single unit of measurement. The ELCA is a time-based extension of CExC in which the material and energy streams are examined throughout the system's life cycle, including maintenance and deconstruction. In Zero-Exergy Emission ELCA, the primary exergy cost is calculated while additionally taking into account the environmental effect of pollutant emissions from emission abatement techniques. CExC, ELCA, and Zero-ELCA, like CExC, do not contain labour and capital fundamentals.
- Another remarkably interesting concept is Emergy. Emergy analysis describes a product or process in terms of solar energy equivalents, or how much energy would be required directly or indirectly to generate an output if solar radiation were the sole input. Because emergy is thought to be embodied in the product or processes, the term embodied energy was simplified to emergy. (Odum and Odum, 2000), who was concerned in environmental and energy quality issues, established this notion. Matter and energy flows are quantified in solar equivalent joules (sej) using a conversion factor known as 'Solar Transformity' or Unit Emergy Value (UEV), which is the amount of Emergy necessary to produce one unit of a specific product or service. Transformity is important because it allows for a hierarchical structure between energy flows and different sub-systems. The superior output energy flow is hierarchically superior. The higher an energy flow's hierarchical rank, the more transformations are required to obtain it (Odum, 1988). Because of its ability to capture the dynamics of large open systems, Emergy is a viable accounting tool for supporting environmental management initiatives (Jiang et al., 2019). Emergy reflects the energy costs of a system in terms of solar power throughout each transition in the life cycle, and it is easily comprehended in monetary terms as well. Several publications in the literature link Emergy directly to LCA; one of these hybridizations is called *Emergy Based-LCA*, and it is formally provided by (Reza et al., 2014a), who say that Emergy is a beneficial supplemental tool to LCA rather than an alternative technique. Actually, it has been explored about coupling emergy analysis with LCA since (Li and Wang, 2009)but the technique known as Emergy Based-LCA was better formalized in 2013. Because the Emergy study is capable of systematically evaluating the role of environmental, economic, and social impacts in an energy-based framework, the indirect effects of raw materials and energy carriers as environmental support for the production of any output, including monetary capital, can be measured and identified using a single metric.
- (Wall and Gong, 2001) introduced a method called *Life Cycle Exergy Analysis (LCEA)* to solve the disadvantages of the LCA's multidimensional approach. The distinction between LCEA and ELCA is ephemeral: it is at the level of aggregate. The first combines all exergetic contributions at each stage (high level of aggregation), whereas the second disaggregates any exergetic input at each stage of the life cycle to emphasize local irreversibilities. Another contrast is the separation between renewable and non-renewable resources.
- *Extended Exergy Accounting (EEA)* by (Sciubba, 2001) is an approach to performing design and configuration optimization of a system evaluating overall resource consumption because it enriches the energy and matter flow with some other 'externalities' (Rocco et al., 2014) as capital flow, environmental damage remediation costs flow, and labour flow, always in exergetic terms, where the calculations are done along all system's Life Cycle phases. All terms of the production cost function are translated into exergy flows in order to generate an exergy cost function in a single measure. One disadvantage of this technique is that substantial assumptions must be made in the

computation of the conversion factors since they must be viewed as primary resource flow. EEA, like ECEC, allows for the provision of economic data in primary exergy equivalents. The extended exergy cost function produced from EEA may be used to replace the economic cost function across the Thermoeconomics framework to identify and optimize the utilization of a system's natural resources. Therefore, should be noted that EEA is a newer approach than others that are well established, and it requires more validation before it can be considered a standard assessment method.

- (Hau and Bakshi, 2004) suggested an extension of CExC termed *Industrial/Ecological Cumulative Exergy Consumption (ICEC/ECEC)*. The primary distinction between ICEC and ECEC is that the former does not take into account the exergy losses of ecological processes in its calculations, but just the non-renewable natural resource use in terms of exergy losses. ECEC is specifically related to Emergy Analysis and consists in including in the assessment the exergy consumed by ecological processes for the processing of raw materials, the dissipation of pollutants, and the operation of industrial systems, whereas CExC focuses solely on natural resources, ignoring ecological products and services. The ECEC analysis is predicated on the concept that ecosystems produce all products and services at the cost of one solar equivalent Joule. With the ECEC, it is feasible to evaluate the system from a monetary standpoint thanks to emergy balances. (Yang et al., 2013, 2015) were successful in including economic and environmental considerations into ECEC computation. Currently, the model based on ecological LCA considers three factors: resource use, economic capital, and environmental effect. Because ECEC is a very advanced technology, information about it is still evolving, and it has limited use in the industrial area. Another flaw in ECEC analysis is the unpredictability in emergy transformities.
- Another more recent technique that deviates significantly from the CExC is the *Cumulative Exergy Extraction from Natural Environment (CEENE)*, which adds the cumulative exergy expenses of land occupancy to standard CExC accountings. According to this theory, land usage is just as significant as other categories since the land uses solar irradiation (represented in exergetic effects) to exist and sustain itself, hence occupying areas of land reduces the odds of capturing solar energy. (Dewulf et al., 2007) created this technique and immediately offered the methodological support to integrate it with LCA; in fact, it was born with the intention of being compatible with current LCA datasets.
- A different method presented by (Dewulf and Van Langenhove, 2002) is termed *Thermodynamically Based-Life Cycle Analysis*, and it presents a framework for measuring all of a process's or product's effects on the Ecosphere and population on a single objective scale. The authors propose an open system in which the Ecosphere, Technosphere, and society interchange exergy created by solar energy, resources, products, wastes, and a fraction of the heat irradiation associated with process irreversibility through time. This method integrates life cycle impact assessment, exergy analysis, and socio-economic aspects into one large exergetic equilibrium of pollution and related human and ecotoxicological impacts across the entire life-cycle in terms of exergetic losses, i.e., the risk of opportunity depletion for current and future generations.
- (Szargut and Stanek, 2007) later proposed the *Thermo Ecological Cost (TEC)*. They expanded the Exergy Analysis into the environmental dimension based on CExC to account for the cumulative consumption of natural resource costs in terms of environmental consequences. This technique specifies environmental costs in order to limit the adverse effects of pollutants discharged into the natural environment (Stanek et al., 2014). The original TEC was concerned with the investigation of the single operational phase under the Life Cycle concept. The first TEC investigated the only activity inside the Life Cycle thinking. In recent years, TEC has been partnered with or integrated

into LCA in a number of works to quantify the worldwide impact of a given manufacturing process (Domínguez et al., 2014; Stanek, 2018).

- *Exergoenvironmental Analysis*, pioneered by (Meyer et al., 2008), evaluates the environmental effect of each component of a system as well as the true sources of the impact by integrating exergy analysis and LCA. The Exergoenvironmental Analysis incorporates both traditional exergy analysis and environmental analyses such as LCA to assess the impact of exergy losses and exergy destruction on environmental sustainability. The first step is to do an exergy analysis on all of the material and energy sources. The second stage includes the LCA of each component or subprocess all the way up to the input streams to the overall system. The environmental implications of LCA are assigned to the exergy streams in the third and final stage (Tsatsaronis, 2007).
- Exergoenvironmental Analysis is frequently used in tandem with *Exergoeconomics*, which was introduced earlier than the latter by (Tsatsaronis, 1984). Exergoeconomic analysis combines thermodynamics and economics to determine the cost of a product or process. It correlates exergy losses with capital expenses in order to build an economically viable efficient system that is not possible with traditional solitary energetic or economic analyses. (Aghbashlo and Rosen, 2018a) dubbed the integrated framework of exergy analysis, economic principles, and environmental evaluation 'Exergo-econo-environmental analysis'. Advances have been made by integrating Emergy with traditional Exergoenvironmental and Exergoeconomic Analysis (Aghbashlo and Rosen, 2018b): the monetary term and the environmental impact score are substituted by the solar emergy joule (sej) going to standardize the metric of their outputs. This allows to better understand and interpret the results e provide an accurate picture of the same phenomenon in both environmental and economic dimensions.
- (Stougie, 2012) introduced the *Total Cumulative Exergy Loss (TCExL)* procedure, which was previously known as CExL. The TCExL is a technique that combines or extends the CExC, CEENE, and ELCA approaches. The system, which is based on basic mathematical concepts, has been designed to take into consideration as many aspects of sustainability as feasible. By analysing the net total exergy loss emerging from a technical system, the TCExL method indirectly addresses resource degradation and scarcity, exergy loss produced by waste flow and emissions management systems, and land use systems. One benefit of TCExL, according to (Stougie and van der Kooi, 2016), is its independence from time and weighting considerations. It would reduce the objectivity of the overall process since it involves factors and equations that are not derived from thermodynamic rules. As a result, it does not expressly strive to combine the economic and social components of sustainability, but they are regarded indirectly since they represent a certain degree of indirect exergy flow.
- (Rocco, 2014) defined an integrated technique termed *Exergy based Input - Output analysis (ExIO)* in his doctoral dissertation. Its theoretical foundation is based on Leontief's Input-Output Analysis (IOA) (Bjerkholt and Kurz, 2006), which is one of the most widely used methods in both Environmental Impact Analysis and Economics, as well as Exergy Analysis, which is applied in a Life Cycle perspective. The fact that the creation of products and services in modern economies may result in significant indirect resource depletion or other externalities that are missed by traditional methodologies was the spark. The ExIO is unusual in that it has numerous unique 'extensions' cantered on the case study's features, one of which is the *Hybrid Exergy-based Input-Output Analysis (H-ExIO)*, in which the exergy is utilized to characterize both external resources and the foreground system. As a result, the system boundaries include direct primary exergy requirements as well as indirect primary exergy requirements due to the system's supply chains, i.e., the estimation of the product or process's primary exergy costs as well as the associated costs of exergy losses throughout the transformation.

2.1.3.2. Integration modelling degree

The assessment techniques listed above each propose a hybridization strategy between traditional methodologies in their unique way. Their integration levels differ, and this issue has never been discussed previously. Because the review was done from a different point of view than other studies, the findings of this part reflect the heart of this study as well as an upgrade to the literature's state-of-the-art. In the generic functional system model shown in Figure 2.12, the structure of an analysis is always determined by the context in which it is performed, i.e., the purpose and scope of the analysis, as well as the specification of the functional unit to be studied. It also relies on the type of streams to be studied, the data available, and the amount of depth of the study. To this purpose, EA-LCA hybrid methodologies were used to examine the following for each case study:

- The fundamental aspects of any assessment, namely the starting assumptions, the functional unit, system boundaries, operating context, and related goal and scope definition;
- The types of input and output streams considered in the analyses (which have been generalized and schematized in Figure 2.12) as well as the reference units of the streams to comprehend their physical nature;
- The databases that were utilized during the inventory phase;
- What characteristics of sustainability were taken into account in the analyses;
- What dimensions of sustainability were taken into account in the analyses;
- The approximations made throughout the analyses and how they were justified;
- The mathematical equations on which each evaluation is based, the analysis of which has permitted the most consistent formalization of interoperation models;
- The metrics used to depict the results;
- Where applicable, the indicators and/or multi-criteria analyses utilized to deal with the finding's interpretation phase.

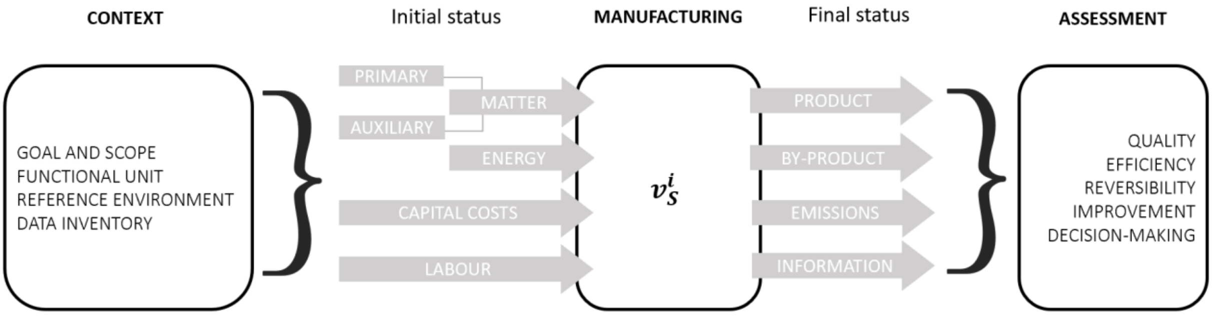


Figure 2.12 - Generic functional system model

The examined characteristics indicated four recurring degrees of integration. Let us add an auxiliary notation to focus on the integration models (schematized, to be clearer, in Figure 2.13): within the framework of the analysis (braces), round brackets indicate the LCA contribution, and square brackets reflect the EA contribution.

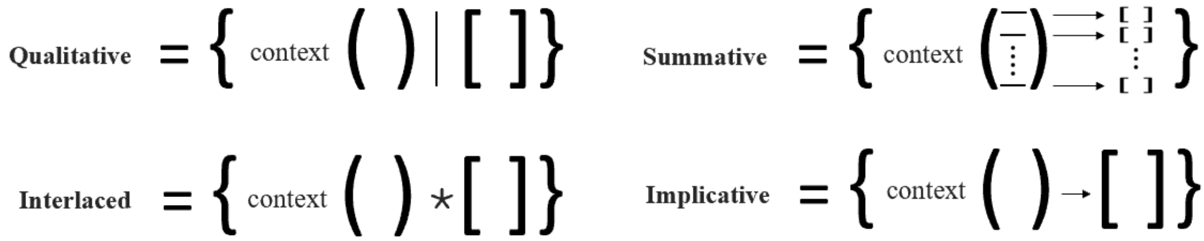


Figure 2.13 - Scheme of the integration models. Exergy is [ ] and LCA is ( )

QUALITATIVE MODEL

The weakest integration occurs following the qualitative model, in which EA and LCA are carried out independently. During the interpretation step, their outcomes are simply compared. This model is typically used when the goal of the research is to acquire a more detailed knowledge of the effective utilization and loss of resources and wastes in order to establish the best approach for optimizing both the process and the product. Furthermore, because of the well-known variability of the presented results and their poor comparability, auxiliary multi-criteria evaluation approaches are frequently used, see (Wang et al., 2015) for results using a multi-objective optimization model; (Mejia et al., 2012) for decision-making using the PROMETHEE-GAIA multi-criteria model; (C. Zhang et al., 2019) for multi-factor evaluation and decision making method for trans-critical ORC through a Fuzzy Analytic Hierarchy Process; and (Arnal et al., 2020) Another option is to extract dimensionless indications from both approaches and then compare them (Medyna et al., 2009a; A.F.C. Fortes et al., 2018).
There are various situations in the literature where this approach has been used. (Beccali et al., 2003) conducted an ELCA on the production process of two different plaster products. Throughout the interpretation phase, the impact categories derived from the LCA were supplemented by an exergetic index to produce a thorough multi-criteria summary of effective resource consumption and the associated environmental impacts (with LCA) and the extent of wastes and depletions along the processes (with EA). (Mejia et al., 2012) did a similar case study. The authors were able to select the best alternative material to conventional plastic for shopping bags and bottles using a multi-criteria method that included LCA's GWP and EA's exergy loss. (Contreras et al., 2009) and (Moya et al., 2013) compared the LCA results from dominant impact categories of Eco-Indicator 99 and CExC of four different cane sugar production processes for by-products valorization in order to gain information both on their environmental impacts and the resources consumption efficiency. Another very similar study was performed by (Shirkhani et al., 2018) to determine the environmental sustainability of the Iranian cement production plant, which studied the effects and risk parameters based on CML baseline, IMPACT 20021, CExD, and Eco-indicator 99 methods. (Milanovic et al., 2017) made the first effort to utilize ELCA in a mobile communication network by comparing the solar-powered hybrid base transmitter with the standard model, but the study is insufficient owing to a lack of reliable and consistent data for disposal situations. In this case, eleven CML-IA LCA and CExD effect categories were compared. (Finnveden et al., 2016) focused their research on the effective use of resources and wastes on two case studies: ferrous waste recycling and the manufacturing and usage of a laptop computer. They argued that the contrast between CExD and EA may lead to better interpretations if CExD is a tool that quantifies all sorts of resources in exergetic terms and the thermodynamic approach is based on strong research and gives conclusions that are useful to decision-making. CExD and EA are also compared to other LCA evaluation methods such as ReCiPe, Eco-Indicator 99, CML-IA, and CED in

order to give a holistic picture of environmental consequences and resource depletion. The similar strategy may be observed in (Khanali et al., 2017)'s study of the Iranian saffron production process. CML and CExD effect categories were analyzed to determine the major hotspot during cultivation. (Mehmeti et al., 2018)'s work is unique in that they included information regarding the economic elements of the procedure in a 'multi-impact assessment'. The findings of the LCA ReCiPe procedure's environmental evaluation of the molten carbonate fuel cell system (MCFC), the findings of the CEENE research to calculate the MCFC resource footprint in an exergetic method, and the leveling cost of energy to analyze the process's economic feasibility are compared. (N. Faleh et al., 2018) did a thermo-environmental LCA study on the biodiesel production system by transesterification of mutton tallow to choose the optimal mixture of different biodiesel structural components. The results of the process simulation using the Aspen Plus™ software were used for the EA as well as inputs to the Life Cycle Inventory Analysis and, finally, the LCA. Finally, the results of standard EA and LCA were compared to thoroughly evaluate the process from thermodynamic and environmental perspectives. (Stougie et al., 2018) evaluated the environmental sustainability of three different biomass power production systems using the ReCiPe impact categories and the exergetic sustainability using the TCExL. This technique may compare not only the systems, but also the assessment procedures, based on the data gathered. Furthermore, the authors investigated how the various ways in which the goods were assigned and avoided in the research may have an impact on the outcomes. Another example of a qualitative technique is the use of Thermoeconomic Analysis. (A.F.C. Fortes et al., 2018) conducted an illustrative example. The goal of this research was to divide the environmental consequences and costs of food dehydration and water generation by a heat pump among its components. LCA was performed to examine the emissions connected with the process. EA was performed to identify the components with the biggest exergy losses, and economic analysis was performed to determine the monetary cost. The outcomes of each component were then compared. (Domínguez et al., 2011) conducted a combined economic and environmental analysis (ELCA) of various energy sources for electric power generation in order to evaluate the relationship between non-renewable and renewable resources over the entire life cycle of each energy source considered, including economic aspects. They began with a traditional LCA using Eco-Indicator 99 effect categories. The overall investment expenditures of each energy system were then compared to the Demand of Exergy Accumulated (DExA), which indicates the complete removal of exergy from nature throughout the creation of a system product. The authors were able to select the optimum alternative method for electric power generation by comparing all of the outcomes.

According to what has been discussed so far, the findings of the impact categories in LCA assessment methodologies (for example, Eco-Indicator 99 or ReCiPe) and the EA outcomes in such studies are not precisely similar. Aside from the many units of measurement that can be standardized and transformed into non-dimensional units, the exergy analysis does not explicitly account for the impacts of emissions or analyze land use. On the contrary, LCA does not include information on the quality of processes or the true efficiency of resource consumption, but rather focuses on their repercussions. ELCA is always applied via a gentle interface between Exergy Analysis and current LCA tools and databases.

SUMMATIVE MODEL

The summative integration model is the second. The technique is slightly different from the previous one in that the EA is conducted at each step of the life cycle, which is included in the system boundaries of the specific analysis to be performed. In this case, the economic dimension may be easily included as long as it is turned into energy flow, as detailed in the first place in (Cornelissen and Hirs, 2002) and afterwards in more recent ways such as (Açıkkalp et al., 2018). The exergetic losses at each step of the life cycle may be examined separately, summed collectively, or combined, depending on the scope of the analysis. The same is true for the efficiency indices. This method should be used while doing research to determine which stage of the process or product's life cycle is the most energy-efficient. In the event of an early design phase, the process system or the product itself may have already been enhanced; in the case of an existing analysis, efficiency interventions on the plant or production cycles may be planned.

The summative model is commonly used with ELCA. (Cornelissen and Hirs, 2002), for example, did an ELCA on four different types of wood waste treatment to evaluate both the usage and depletion of the natural resources necessary for the procedures. The scientists were able to pinpoint where and when natural resource depletion occurred by estimating the quantity of exergy losses for each component at each life cycle stage. The ELCA results were also compared with traditional LCA results in this study to better understand the environmental implications of waste and natural resource depletion. (Wang et al., 2011) implemented ELCA on a CO_2-zero-emission energy plant comprised of a coal-fired power system and a CO_2 abatement system using a very different architecture. The goal of using exergy as a basic physical parameter of energy and resource consumption and environmental impact management in the inventory analysis and impact assessment rather than the conventional LCA is to use exergy as a basic physical parameter of energy and resource consumption and environmental impact management in the inventory analysis and impact assessment. The authors computed the CExC and pollution abatement (AExC) of the facility during the phases of building, operation, and decommissioning. (Aleksic and Mujan, 2016) used ELCA on electrical devices in smart grids. They computed the embodied and operational exergy losses of each component of each sub-process, as well as the losses during each life cycle step. Encourages the presence of the most influential material/component in smart grid equipment and at the most consuming stage. As a result, it is simple to select the most appropriate policies for improvement. (Rocco and Colombo, 2016) used the ELCA to an electrical production Waste-to-Energy system. They calculated the basic energy expenses as well as the monetary costs of the life cycle during the building, operating, and maintenance stages. When describing the stages of the life cycle in the evaluation, (Dincer and Rosen, 2013) took a different approach than Rocco and Aleksic. Their ExLCA was evaluated using fossil fuel supply, dividing exergy consumption into direct exergy lost during transformation and indirect exergy loss from embodied exergy released by construction materials and machinery during all life-cycle stages. The economic ramifications of indirect exergy losses are also included in this research, along with an independent capital investment efficiency factor. The authors adopted the exergy technique in this research to reduce the irreversible character of the manufacturing life cycle, but it is important to note that the exergy is also impacted by a certain degree of uncertainty due to the irreversibility and non-linearity of the real system itself. (Lettieri et al., 2009) created a set of advantages and cons for using exergy in LCA using the Life Cycle Exergy Analysis (LCEA) framework in 2008. Among the benefits, the most important is that this information will identify circumstances in which urgent technological improvements may be made, which should be allocated to maintenance processes, efficiency improvements, or optimizations. The authors offered a demonstration by applying this strategy to Computer Servers in a Data Center as a case study. They measured the server's exergy consumption at each stage of its life cycle, from raw materials extraction to disposal and recycling. (Wall, 2011), who used LCEA to a wind power facility, found the same benefits highlighted by Lettieri et al. a few years earlier. His LCEA framework consisted of computing the exergetic balances of the wind power plant's key life-cycle stages, from turbine production through destruction and sanitation. LCEA's application to various wind generating systems provided an excellent description of the exergy fluxes involved.

The inclusion of exergy in the LCA in analyzing the use of natural and non-natural resources, renewable and non-renewable resources, allows the evaluation to be more objective and robust, which is a noteworthy consequence of this study.

Aside from (Hau and Bakshi, 2004), (Yang et al., 2015) and (Wang et al., 2019) conducted ECEC analyses on a Chinese raw-coal manufacturing facility and an Organic Rankine Cycle for waste heat power generation, respectively. The former used their new expanded ECEC, which included ecological, environmental, and economic considerations, to determine the best optimization method for the process. The authors emphasized, particularly in terms of economics, that this technique supplements traditional economic analyses such as the Life Cycle Cost (LCC). Because this was the first attempt to adopt expanded ECEC, there were ambiguities and difficulty in organizing all of the data (information). (Wang et al., 2019) also attempted to adapt ECEC analysis to the ORC system in order to examine the environmental, resource consumption, and economic implications of 1 kWh of energy generation over the whole life cycle. The outputs of the thermodynamic (energetic), economic, and sustainable (resource and environmental) goals managed to place the emphasis on diverse optimization goals and their viability in each of the works described above. EEA takes a very different approach to merging LCA and EA than the previous example study, owing to its differing monetary cost methodology. (Rocco et al., 2014) tested the practical use of EEA on an electric power transmission line with three different wire diameters. In the assessment they summed up energy and materials flow assessed with CExC, the environmental remediation costs of the pollutant pollution estimated by LCA and the labor and capital equivalents by exergy. The building, operating, and dismantling phases of the life cycle have been analyzed in this approach and according to the functional unit. The authors were able to estimate the ideal diameter using this technique, which gave the best compromise for a decrease in overall energy consumption, an environmental balance, and a large monetary savings throughout the operating period. The authors also discussed the benefits and downsides of the EEA technique in comparison to standard LCA or ELCA. (Dai et al., 2012) used EEA to elucidate thermodynamically the environmental-social-economic link between the key seven Chinese industrial sectors. Their objective was to demonstrate that the EEA is applicable to all generic manufacturing systems, not simply energy generation processes.
The data identified the most influential sector as well as a network of hierarchical reliance among these seven sectors. Throughout 2014, they released a number of particular indicators for the EEA-based sustainability assessment of key Chinese industrial sectors (Dai et al., 2014).

IMPLICATIVE MODEL

The strategy entails doing an LCA evaluation on a substance or method first, and then conducting an exergetic review just on the component or sub-process that has caused the greatest environmental effect amount. Such intervention is explained by the idea that increased usage of resources and energy has a higher environmental impact. This model is useful when it is necessary to simplify the analysis owing to the various sub-processes or multiple components of the product manufacturing. To date, this technique has been used in two works (Dassisti et al., 2019; Selicati and Cardinale, 2021a). The authors used traditional LCA on die-casting process at an Italian industrial SME. The LCA research conducted allowed for the identification of the most essential product in terms of resource usage and emissions. The Exergy Analysis was then performed to the chosen completed product, determining the exergy loss for each sub-component. The system boundaries in measuring mass flow and energy flow in the LCA were different, which was fascinating. The implicative path in this study became a strategic framework to discover process optimization options, as well as prospective enhancements to the manufacturing process, and to establish an IoT monitoring strategy on the technologies.

INTERLACED MODEL

The final model, the interlaced one, incorporates LCA and EA the most. Integration can be accomplished in a number of ways. As demonstrated by (Portha et al., 2010) and (Hamut et al., 2014), it is feasible to integrate LCA characteristics with Exergy or Emergy traits in a unique formulation or method. In this case, the conclusion is frequently expressed in energetic terms, although assumptions and approximations must nearly always be made (Dewulf and Van Langenhove, 2002). Furthermore, no case studies exist in which the exergy is transformed into the effect categories that characterize the LCA (e.g., Human Toxicity or Global Warming Potential). It is a one-sided change. The second strategy is to combine LCA and EA elements in a single indicator, as (Casas-Ledón et al., 2017) and (Gao et al., 2018) have done. Socioeconomic factors may be incorporated into the formulation as well as the indicators using the interlaced model. (Dewulf and Van Langenhove, 2002) provided one of the most complete frameworks recognized in the literature for the thermodynamic treatment of emissions and their human and ecotoxicological consequences. The authors examined several synthetic organic polymers for building insulation over time to establish their sustainability in terms of global exergy lost in the ecosphere, the technosphere, and the population, showing the rate of loss of possibilities for current and future generations. The authors presented a formulation in which renewal rates and particular exergy contents are assessed in time using a single thermodynamic measure. The former is derived from a life cycle evaluation of resource consumption and emissions, while the latter is derived from a phrase that measures the exposure to ecological and human damaging consequences. The latter is derived from an exergy study that includes some statistical thermodynamics to account for complicated biological systems. The main constraint of this study is the many assumptions and approximations made in the evaluation and application of statistical input data, particularly for society and biodiversity, but the case study indicated that it may be utilized in practice. (Portha et al., 2010) applied a coupled Exergy-LCA to two petrochemical processes to generate fuel and energy. They argued that when assessing a process involving the transfer of energy carriers, exergy is the ideal supplemental technique to LCA since it identifies process irreversibilities that LCA cannot solve on its own. They estimated the effects of greenhouse gas emissions by combining direct and indirect emissions. The former is produced by a mass balance of pollutants represented in CO_2eq of the GWP of the LCA. The latter are derived from the quantity of exergy lost in the considered system as a result of unit creation or disassembly, electricity, and heat usage. Aside from this computation, the authors devised a 'quality factor' to assess the economic worth of the flows under consideration. Combining exergy with LCA allows you to compare two alternative processes with the same purpose in terms of resource depletion, climatic change, and monetary value.

The literature has a number of case studies in which Exergoenvironmental research is conducted, frequently in conjunction with an Exergoeconomic evaluation. Following the first publication of the Exergoeconomic and Exergoenvironmental analysis by (Tsatsaronis, 1984), (Tsatsaronis and Morosuk, 2008a, 2008b) proposed a 'advanced' Exergoenvironmental and Exergoeconomic analysis implemented on a gas-turbine based cogeneration system in order to evaluate the real energetic and economic potential for improving the system and its components by splitting capital costs, environmental impacts, and exergy destruction into end Because the unavoidable components cannot be further decreased owing to technological limits of the system or the reference environment, the authors have concentrated on decreasing the avoidable-endogenous parts. Because the unavoidable components cannot be further decreased owing to technological limits of the system or the reference environment, the authors have concentrated on decreasing the avoidable-endogenous parts. The endogenous component represents the efficiency of the actual system. The authors conducted and contrasted both ordinary and advanced exergoeconomic and exergoenvironmental analyses in the case study. The main disadvantage that they emphasized in the advanced study was the use of more or less arbitrary facts and assumptions, which were necessary for the division of the subsystems into avoidable and unavoidable components. Also (Buchgeister, 2010) used only traditional Exergoenvironmental analysis to generate electricity from a high-temperature solid oxide fuel cell, with the goal of demonstrating that this approach is a powerful support tool for detecting the interdependencies between thermodynamic behavior of process components and environmental impacts. He stated that while this approach can be applied to a variety of processes (chemical, manufacturing, etc.), a unique and more practical method for conducting this analysis is still required. (Hamut et al., 2014) performed the first Exergoenvironmental analysis on a hybrid electric vehicle's thermal management system. Furthermore, the LCA is performed using SimaPro software and the Eco-Indicator 99 evaluation technique to identify the component with the greatest environmental effect. The authors indicated that the study needed to be improved by introducing a multi-objective optimization phase into the advanced framework to provide more objective data and information on trends and design improvements. (Cavalcanti et al., 2019) calculated exergy efficiency, specific cost, specific environmental impacts, and exergoeconomic and exergoenvironmental impact factors in order to evaluate a diesel engine with five diesel–biodiesel blends and three different load capacities and achieve the best conditions for electricity production. They proved that the choice of the optimal fuel is greatly impacted by the desired purpose, rather than being offered immediately and unequivocally by the comparison of outcomes.

Because solar energy has gained tremendous interest in both electricity and heat generation over the last ten years, and because solar energy is one of the largest sources of renewable energy with no environmental impacts (Dincer and Rosen, 2017), many authors have focused their attention on this field, attempting to improve the efficiency of generation systems through environmental and economic analysis. (Casas-Ledón et al., 2017) were the first to apply the Exergoenvironmental study to the integration of municipal solid waste gasification with a combined power system in Chile (waste-to-energy system). The system and its components were simulated using the Aspen One v9.0 program, however several assumptions were made in order for the simulation to operate. The research is complemented for these aspects by a sensitivity analysis, which also attempts to provide insight into the most significant concerns discovered at the component level during the evaluation of the Exergoenvironmental results. (Gao et al., 2018) performed exergy and Exergoeconomic evaluations on a coal-fired combined heat and power plant, focusing on better residue and CO_2 allocation methodologies. The authors examined the cost of flue gas cleaning in the study for the first time, however the waste heat recovery system owing to recycling potential was not addressed. The following year, and similarly to Gao et al., (C. Zhang et al., 2019) conducted the same study on the coal-fired combined heat and power plant, including the waste heat recovery system. In the same year, (Yang et al., 2019) conducted the same study on the refrigeration, heating, and power system, focusing on a dual-fuel CCHP device based on biomass and natural gas and analysing the system's thermodynamic efficiency and stream costs. (Montazerinejad et al., 2019) did the same thing on a novel CCHP hybrid solar system. A novel and fascinating research was carried out by (Okonkwo et al., 2019), which provided added value to the established studies on parabolic trough solar collector by evaluating the impact of its irreversibilities from an economic and environmental point of view via an Exergoenvironmental and Exergoeconomic analysis. The purpose of the work was to compare the traditional absorber tube with the innovative characterized with a converging-diverging geometry. (Aghbashlo and Rosen, 2018b) recast the Exergoenvironmental and Exergoeconomic studies in terms of emergy to unify the unit of measurement of their outputs and produce more ecologically sound conclusions. The authors also used the innovative framework to a cogeneration system based on a gas turbine. They employed the specific exergy costing (SPECO) approach to establish the conversion factors for each flow of matter, energy, and money in solar emergy joule, and then performed Exergoenvironmental and Exergoeconomic studies.

2.1.3.3. Integrated EA-LCA Highlights

As previously said, EA and LCA are two recognized and documented feasible sustainability estimators, each with potential and limitations, similarities and distinctions. This work is far from a complete discussion of their theoretical basis, instead focusing on the most relevant practical aspects in order to explain why the two techniques should be integrated.

Life Cycle Assessment is goal and scope oriented and based on steady-state calculations (ISO 14040, 2006), Exergy Analysis is goal and scope oriented too, but its empirical usefulness varies whether the approach is product-specific or process-driven analysis (Liao et al., 2013); EA talks about dynamic equilibrium. LCA is a linear method while EA is a non-linear method due to process irreversibilities. Both procedures are time-dependent, although even the time range evaluated during an LCA is often greater than that considered during an EA. It is critical to create the framework of the research throughout the LCA and the EA, for which the LCA is titled "System Boundaries" and the EA refers to the defining of the reference environment. Both LCA and EA involve mass and energy balances, but LCA does not have a single measure; in fact, many writers regard it as a multidimensional approach. EA has a single measure, which improves practitioner comparability and comprehension of outcomes(Romero and Linares, 2014). LCA is based on cause-and-effect linkages and seeks to comprehend the environmental effects of occurrences (Hellweg and Mila i Canals, 2014). However, it is insufficient to define indirect consequences such as the connection between goods, processes, services, sectors, and so on and their surroundings (social, economic, environmental, and innovative goals) (Onat et al., 2017). On the other hand, according to (Hammond, 2004), the relationship between exergy consumption and resource efficacy is not yet direct, hence exergy analyses are insufficient to establish whether a system is sustainable or not. According to (Gaudreau et al., 2009), "despite various attempts, there is no empirical proof that there is a direct link between the quantity of exergy contained in the wastes of a process and the potential harm that this exergy is capable of inflicting on the environment." According to (Cleveland et al., 2000), subjectivity affects LCA outcomes owing to too many assumptions and approximations throughout the assessment, whereas EA ignores human preferences and needs. The consistency of the reference databases for the design of input and output flows greatly influences the accuracy and completeness of the LCI phase. The dependability of the databases has been extensively debated in the literature, see (Hellweg et al., 2001; Edelen and Ingwersen, 2006). The same is true for the EA, which is accompanied with databases. (Alvarez, 2013) and Szargut before her conducted research on similar databases, however his research was limited to chemical Exergy (Szargut and Morris, 1987)

It is clear from this brief summary of the primary functional elements of LCA and EA that they are complimentary techniques. Other writers, see, agree with this assumption (Pati et al., 2009; Portha et al., 2010). It would be beneficial to apply these two methodologies in tandem, using hybrid approaches, for a systematic and fair assessment of sustainability.

According to the literature, there are strong perspectives on the utility of exergy in conjunction with LCA as a measure of sustainability. Exergy inefficiencies are frequently employed as an extra effect category in the existing LCA paradigm; however, these impact categories are not directly comparable (Stougie and Weijnen, 2014). (Ozbilen et al., 2012; M.A. Rosen et al., 2012) shown the possibility of adopting EA as a single indicator in tackling environmental sustainability concerns. According to (Hernandez and Cullen, 2019), EA is a holistic, flexible, integrated, and transparent method that evaluates both quantity and quality of energy and resources and can be seen as an added value to traditional life cycle assessment; however, no simple guides, training, or software tools exist to facilitate its wider use. (Alvarenga et al., 2013) provide another point of view, focusing on another major aspect of LCA: the completeness of the characterization elements. They developed new exergy-based spatial characterisation parameters for land usage as a resource within man-made or natural systems in this work, allowing for greater geographical distinction without double-counting resources. Characterization factor issues can also be encountered in (Zarei, 2020).

Because sustainability may be assessed not just in environmental terms, but also in economic and social ones (Purvis et al., 2019), and because these dimensions are not immune to change over time, these other two aspects can be included in the evaluation.

Nonetheless, economic and, notably, social issues are less studied in literature than environmental ones. A business, a technique, or an environmental alteration have all had a societal impact. The social factor is rarely analysed in the context of sustainable development strategies, or it is evaluated with a statistically incomplete, confusing, or incomplete weak connection (Assefa and Frostell, 2007). The reason is that there is a lack of knowledge about its implications in environmental and economic performances. (Clarke-Sather et al., 2011) offer an insightful perspective on the societal, environmental, and economic aspects of SMEs. Because economic consequences are critical to decision-making, it is helpful to take account of the reality that exergy efficiency is directly linked to the costs of the operating process or activity. EA is prone to be combined with economic considerations to find cost-effective and realistic improvement options (Salehi et al., 2018).

Concerning the other aspects of sustainability, the economics may be an essential feature for a factory's production system, but an economic input-output analysis alone does not show how to accomplish low-cost cleaner manufacturing. Exergy, on the other hand, is one of the pure thermodynamic metrics, but its interpretations are insufficient to address mitigating issues that are compatible with its laws. Exergy, emergy, and LCA seldom cover social aspects, despite the fact that society is one of the foundations of sustainability, and when they do, substantial approximations are made during the early and intermediate phases, resulting in unscientifically and inconsistent findings. Alternatively, the socioeconomic issues are handled using entirely different methodologies, such as in (Švajlenka and Kozlovská, 2020).

Table 2.2 summarizes the case studies examined in this review: The relationship between the EA-LCA hybrid techniques and their respective integration model is classified, as is the functional system model of assessment (e.g., the general kind of streams that each approach analyses) and the sustainability dimension.

This part allows you to respond to some of the questions raised in the paper's opening paragraph. All of the hybrid techniques studied utilised neither a unique computational model nor a shared calculation framework. Sometimes the authors' answers are too specialized for each case study, and a common approach will result in a large number of assumptions and hence not an adaptable model. This remark is especially pertinent since the review discovered that a high level of collaboration between EA (or Emergy) and LCA led to stronger and more strict hypotheses both early in the assessment and throughout. The same thing does not happen when EA, Emergy, or LCA are regarded to be separate contributions.

Most hybrid techniques find it simpler to refer to Szargut for EA, to utilize Eco-indicator 99 for LCA, or to examine simply life cycle phases for EA, or to make supplemental use of emergy balances to add economic or social flows to the evaluation.

It is unclear if the interlaced model is sufficiently integrated to ensure a full knowledge of a system process or product's sustainability and quality. As a result, we may conclude that there is no better hybrid approach than any other because each has its own set of advantages and disadvantages. It is more crucial to understand all of the ways and how to select the one that best meets company demands each time. The discrepancies are inherent in the chosen interoperation paradigm. This is the novel outcome that has been highlighted in this paper. One notable characteristic that considerably restricts the arbitrary selection of the optimal assessment technique is that the stronger and more restricted the assumptions are at the outset, the higher the interaction between EA and LCA.

Because LCA is based on linear assumptions while EA is based primarily on non-linear assumptions (such as the second law of thermodynamics), only one way of interaction appears to be possible: LCA to EA. Furthermore, the disadvantages of hybrid techniques are overcome by an ad hoc interpretation arrangement (e.g., multi-criteria analysis). It is insufficient since important information is lost during these procedures. It is considerably simpler to locate studies in the literature on how to optimize traditional LCA on several dimensions for assessing emerging technologies (for example,(Cucurachi et al., 2018) or (van der Giesen et al., 2020)) than research on optimizing the frameworks of more extensive hybrid methods.

Based on all of the methodologies discussed, it can be concluded that a perfect union of exergy and life cycle thinking is still difficult to achieve. The idea of developing a standardized and consistent assessment system capable of providing a comprehensive, detailed evaluation of process reversibility and its environmental implications is still a long way off.

Exergy Analysis remains a viable tool for optimizing production processes in the absence of a standardized, comprehensive, and rigorous method of assessment. When a quality problem emerges, the first step is to trace the physical and dynamic nature embedded in those variables by checking controllable and non-controllable characteristics. It will also allow greater efforts to be made to strengthen the specific EA-LCA approach as a standard, expand both EA and LCA databases, and, as a result, eliminate some operational uncertainties.

To conclude and summarize: neither of these have been carried out from the perspective of the authors of this review, i.e., to investigate the degree of integration, completeness, and effectiveness of hybrid methods in line with the objectives of the analyses themselves through a plethora of case studies addressing these methodologies. The evaluation identified further issues regarding the total integration and interoperability of EA and LCA, which would necessitate further examination, as well as additional research into the complicated topic of indicators as an interpretative model of hybrid analytic findings.

This review will encourage both researchers and practitioners to choose the best model approach for their goals, the streams to be considered, interpret the findings, and build a transdisciplinary understanding of the case study system's information in order to determine the best approaches for process enhancements.

Table 2.2 - A summary of hybrid approaches, the degree of integration of EA and LCA, the streams considered, and the sustainability dimension addressed

Method	Integration Model	Reference	Streams				Sustainability Dimension			Indicator provided?	Multi-criteria Analysis
			Matter	Energy	Capital	Others	Environmental	Economic	Social		
ELCA/Zero-ELCA	Qualitative	(Ayres et al., 1996, 1998)	x	x			x				
		(Cornelissen, 1997)	x	x	x		x	x			
		(Beccali et al., 2003)	x	x			x			x	x
		(Domínguez et al., 2011)	x	x			x			x	
		(Mejia et al., 2012)	x	x			x				
		(Milanovic et al., 2017)	x	x			x				
	Implicative	(Dassisti et al., 2019)	x	x			x				
	Interlaced	(Portha et al., 2010)	x	x			x				
		(Rubio Rodríguez et al., 2011)	x	x			x			x	
	Summative	(Cornelissen and Hirs, 2002)	x	x			x				
		(Wang et al., 2011)	x	x		x	x			x	
		(Ozbilen et al., 2012)	x	x			x			x	
		(Marc A. Rosen et al., 2012)	x	x			x				
		(Dincer and Rosen, 2013)	x	x	x		x	x		x	
		(Koroneos and Stylos, 2014)	x	x			x				
		(Rocco and Colombo, 2016)	x	x			x				
		(Aleksic and Mujan, 2016)	x	x			x				
LCEA	Summative	(Gong and Wall, 2001)	x	x			x				
		(Lettieri et al., 2009)	x	x			x				
		(Wall, 2011)	x	x			x				
Exergoenvironmental Analysis	Qualitative	(Meyer et al., 2009)	x	x	x		x	x		x	x
		(Buchgeister, 2010)	x	x			x				
		(Tsatsaronis, 2011)	x	x	x		x	x			
		(Morosuk and Tsatsaronis, 2014)	x	x			x			x	
		(Restrepo and Bazzo, 2016)	x	x			x			x	
		(Nahla Faleh et al., 2018)	x	x			x				
	Interlaced	(Hamut et al., 2014)	x	x			x				
ExIO/H-ExIO/B-ExIO	Summative	(Rocco, 2014)	x	x	x	x	x	x		x	
		(Rocco et al., 2017)	x	x	x		x	x			
CExC/CExD	Qualitative	(Szargut and Morris, 1987)	x	x			x			x	
		(Bösch et al., 2007)	x	x			x				
		(Medyna et al., 2009a, 2009b)	x	x			x			x	x
		(Finnveden et al., 2016)	x	x			x				
		(Khanali et al., 2017)	x	x			x				
		(Shirkhani et al., 2018)	x	x			x				
	Summative	(Moya et al., 2013)	x	x			x				
		(T. Gulotta et al., 2018)	x	x		x	x			x	
ICEC/ECEC	Qualitative	(Wang et al., 2005)	x	x	x		x	x			x
	Summative	(Hau and Bakshi, 2004)	x	x	x	x	x	x		x	
		(Yang et al., 2013)	x	x	x		x	x			
		(Yang et al., 2015)	x	x	x	x	x	x	x	x	
		(Wang et al., 2019)	x	x	x		x	x			
CEENE	Qualitative	(Dewulf et al., 2007)	x	x			x				

		(Mehmeti et al., 2018)	x	x	x		x	x			
	Summative	(Huysveld et al., 2013)	x	x			x				
TCExL	Qualitative	(Stougie, 2012; Stougie and van der Kooi, 2016)	x	x	x		x	x			
		(Stougie et al., 2018, 2019)	x	x			x				
TEC	Qualitative	(Stanek et al., 2015, 2018)	x	x	x		x	x			
	Summative	(Domínguez et al., 2014)	x	x	x		x	x			
EEA	Qualitative	(Sciubba, 2001)	x	x	x	x	x	x		x	
	Summative	(Sciubba, 2012)	x	x	x		x	x			
		(Rocco et al., 2014)	x	x	x	x	x	x	x	x	
		(Dai et al., 2012, 2014)	x	x	x	x	x	x	x	x	
Thermodynamically based-LCA	Interlaced	(Dewulf and Van Langenhove, 2002)	x	x		x	x		x		
Exergoeconomics	Interlaced	(Groniewsky, 2013)	x	x	x		x	x			
		(Açıkkalp et al., 2018)	x	x	x		x	x		x	
		(Grisolia and Lucia, 2019)	x	x	x		x	x		x	
Thermoeconomics	Qualitative	(González et al., 2003)	x	x			x				
		(Anderson Felipe Chaves Fortes et al., 2018)	x	x	x		x	x			x
	Summative	(Sieniutycz and Salamon, 1990)	x	x			x				
		(Bakshi et al., 2011)	x	x	x		x	x		x	
Exergoenvironmental + Exergoeconomic Analysis	Interlaced	(Tsatsaronis and Morosuk, 2008a, 2008b)	x	x	x		x	x		x	
		(Casas-Ledón et al., 2017)	x	x	x		x	x			
		(Aghbashlo and Rosen, 2018a, 2018b)	x	x	x		x	x			
		(Gao et al., 2018)	x	x	x		x	x			
		(Q. Zhang et al., 2019)	x	x	x		x	x		x	
		(Montazerinejad et al., 2019)	x	x	x		x	x			
		(Okonkwo et al., 2019)	x	x	x		x	x			
		(C. Zhang et al., 2019)	x	x	x		x	x		x	x
		(Cavalcanti et al., 2019)	x	x	x		x	x			
Emergy based-LCA	Interlaced	(Reza et al., 2014a, 2014b)	x	x	x	x	x	x	x		
	Qualitative	(Li and Wang, 2009)	x	x			x			x	
	Summative	(Niccolucci et al., 2009)	x	x		x	x		x		
?	**Fully Integrated**	**None**	x	x	x	x	x	x	x	**Overall Reversibility?**	

2.1.4. Exergetic-Life Cycle Assessment Heterogeneities addressed with the System Thinking

The goal of any sustainability assessment approach is to quantify the environmental, social, and economic harm caused by product, process, or activity life cycles. Thus, the comparability of numbers is a vital problem for guiding informed decision-making.

The consistency of the sustainability assessment results stems from a variety of practitioner decisions and assumptions, which vary depending on materials, technical processes, and geographical environment, all of which can be causes of variability. Addressing the sources of heterogeneity in sustainability assessment resulting from data contexts, procedural decision-making methods, and measurement procedures is critical, given that all existing methodologies, such as Life Cycle Assessment, Exergy Analysis, and methods derived from their combinations, leave a significant role to the practitioner's subjective judgment. Heterogeneity is thus related to the possibility of different perspectives in decision-making, which may, for example, reflect in the selection of dissimilar inventorying for the impact evaluation as well as in the selection of the most appropriate impact category: the effect is uncertainty in the final results. The issue of heterogeneity can be ascribed to a variety of factors, including data sources, indicators, and subjective perspectives.

At this stage of the thesis, it is almost evident that both LCA and EA have similar aims, both focusing on process sustainability and relying on observable data. Both strategies might be complimentary if they work toward the same aim of producing sustainability measurements but from different perspectives. Indeed, several attempts have been made to hybridize the two methods thus far, see the previous paragraph 2.1.3. However, current methodologies are either distant from the original thermodynamic idea of exergy or considerably too wide and unsuited for resource accounting (Peters, 2021). However, there are challenges with identifying the best suited approach based on the purpose, i.e., more than one procedural alternative option that leads to distinct outcomes. As a result, two practitioners evaluating the same system may arrive at different conclusions. The EA performed in tandem with the LCA encourages the practitioner to make more strict procedural decisions, eliminating subjectivity difficulties that might arise throughout each phase of the analysis (Bakshi et al., 2011). Because of its multidimensional character, heterogeneity is an imprecise phrase that applies to a variety of contexts. A system's heterogeneity is defined etymologically as a composite of varied pieces that are typically incomparable (“Heterogeneity. A Dictionary of the English Language - Samuel Johnson,” 2017). It refers to any discrepancy in analysis in terms of procedural activities that results in disparities in ultimate environmental impacts (Higgins and Thomas, 2019). The heterogeneity issue in sustainability assessment is widely discussed in the literature, and is sometimes referred to as "unresolved problems" (Reap et al., 2008b), "limitations" (Curran, 2014) or "ambiguities" (Werner, 2005), "technical emerging challenges" (Hellweg and Mila i Canals, 2014), and "granularity" of data (Ross and Cheah, 2019), all of which refer to the same procedural and interpretative issues. In this paragraph, we seek to define the meaning of disparities in timeframe references, data context (temporal, geographic, economic, social), procedural decision-making processes, and measurement procedures. As a result of heterogeneity, various subjective options may emerge, resulting in multiple decision-making outcomes at each stage of the EA and LCA. As a result, the quality and validity of the outcomes of the two distinct practitioners' analyses might be quite questionable. If LCA and EA are used as practical frames to answer the question of ‘how to classify and quantify emissions’, the system design view is proposed as a conceptual frame to answer the question of ‘how they can be strategically applied’ to create a context-specific sustainable strategy. In this paragraph we discuss the causes of heterogeneity in sustainability assessment, with a focus on LCA and EA, using system thinking to allow the evolution of practitioner's subjective decisions into coherent best choices, as well as to provide a possible procedural guideline for reducing heterogeneity. As a consequence, the practitioner's subjective viewpoint and other sources of variability may be reduced. System thinking is useful for delving into complex situations. It comprises a methodical approach that utilizes a variety of methodologies to investigate the activities of wholes and the numerous relationships between the components. Any of these ways is rigorous and ordered, but systematic thought is more typical in the reductive approach, where situations are broken down into constituent elements and mostly fundamental, linear cause and effect linkages are explored. The literature has a variety of system

ideas, the most important of which are given in (Kumanyika et al., 2010). In general, system design or system thinking is a method for determining how things (components and systems) are linked and how they impact one another. This strategy needs easy framing as well as the ability to describe what and how is investigated while maintaining consistency and transparency (Cabrera and Cabrera, 2019). As stated in the introduction, seeing systematically refers to a means of organizing and making sense of our thoughts about what is conceivable. The aim is to demonstrate why LCA and EA approaches are used to reduce the variability of sustainability evaluations. Essentially, the notion is that employing system thinking to apply both methodologies in parallel and simultaneously can minimize the variability of the sources. Following, we will discuss some essential notions beneficial for this goal known as "procedural choices." A system is a static model that simulates a real product or industrial process (Andersson, 2013) by utilizing unit-processes, each of which represents one or more activities (e.g., manufacturing, assembly, transportation, etc.) and the initial assumptions. These system requirements define its boundaries (Curran, 2012). The reference flows connect the unit-processes: the quantities of particular product flows required by each of the compared systems to create one unit of the function. The reference flow then serves as the foundation for developing product system models ("UNI EN ISO 14040," 2006). We define scenario S at time t_x ($S(t_x)$) as "a description of a hypothetical future condition useful for certain sustainability analysis applications, based on specified future assumptions, and (where applicable) also containing the depiction of the progression from the present to the future" (Pesonen et al., 2000). Within system thinking, a transformation is defined as any process or action that resulted in a change in the state of a system, such as from $S(t_1)$ to $S(t_2)$.
The system is referred to in order to facilitate the development of various transformation pathways from $S(t_0)$ to $S(t_x)$. Each system able to meet the requirements can be considered an alternative solution. Each system capable of meeting the requirements might be regarded as an alternative option. See Figure 2.14 for a schematic explanation of a generic system Each alternative system has a needed transformation (from $S(t_0)$ to $S(t_x)$) and one or more criteria that are used to evaluate the alternative systems. This evaluation helps you to choose the best alternative analytical path. To make each alternative comparable, practitioners must follow the same transformation path.

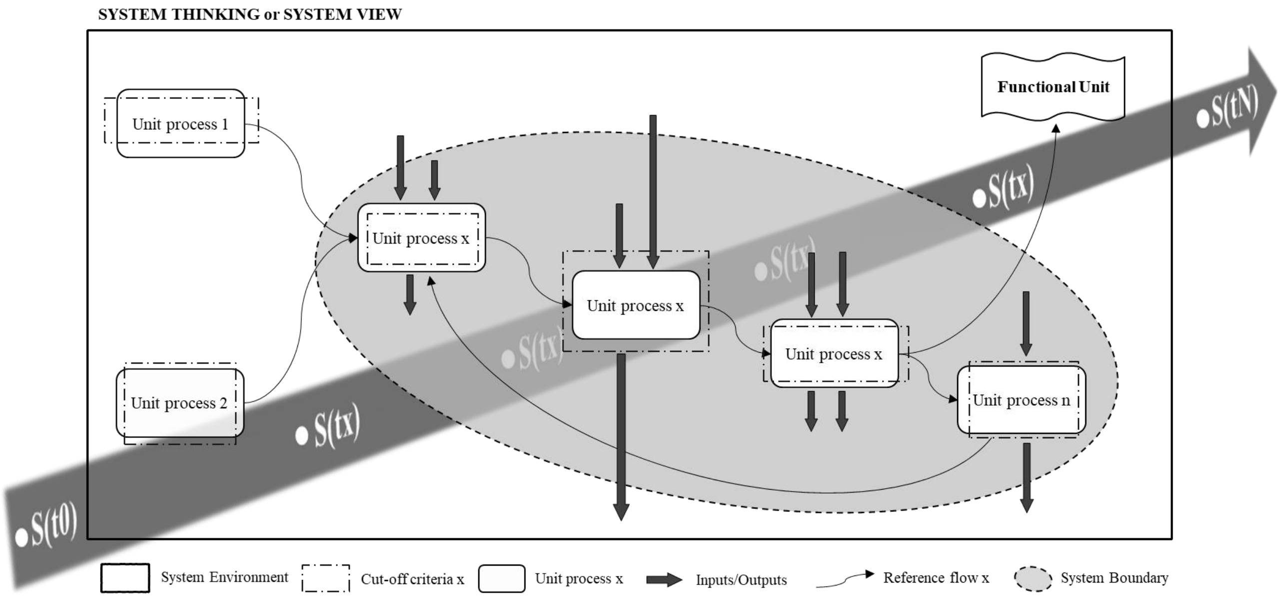


Figure 2.14 - The model of a generic system within the system thinking

The primary cause of heterogeneity in sustainability assessment is subjectivity generated from procedural decisions made during each phase of the evaluation, which is not an easy problem to overcome. Due to this subjectivity, comparing two or more LCAs of the same system (product, process, or activity) may be impracticable. The influence of any procedural option is dependent on the 1) initial scenario while contemplating a sustainability evaluation in a system thinking ($S(t_0)$), the 2) transformation objectives ($S(t_x)$) and the 3) criteria for choosing the finest makeover. As a result, a comparison is only conceivable if the processes, products, or activities being compared have the same identical beginning context, transformation objectives, and assessment criteria (coherence criteria). The EA method, on the other hand, makes this definition easier by utilizing the exergy-loss measure as a unique reference, with the optimal alternative being the one that decreases this term as much as feasible. By combining LCA and EA methodologies, a well-posed assessment framework should be offered, provided that this is a bottom-up approach directed by product advancement along the production settings. Thus, the challenge of heterogeneity is to provide an efficient impact measurement in which the acceptable measurement error is less than the uncertainty of the total computed value (Giovannini, 2015).

The paragraphs that follow are grouped in accordance with paragraph 2.1.1 as s generic actions to be taken when implementing an LCA and EA sustainability evaluation. Thus, heterogeneity sources will be explained and treated through system thinking. In conclusion, system design allows practitioners' judgments to be turned into system design activities that fulfil the demands of stakeholders. Stakeholders formalize their demands via the formulation of a list of requirements. The requirements reflect the limits that the system must follow. In this context, the EA aids in limiting the LCA's inherent heterogeneity sources that are not directly dependent on this set of requirements.

STEP 1 - Goal and scope definition

The system view encourages practitioners to share the same way of reasoning by answering what, where, and when to measure questions in a unique and unequivocal manner, reducing the sources of heterogeneity related to functional unit and reference flow choices, the scenario, system boundaries, and temporal and geographical boundaries.

SUB-STEP 1.1 - Functional unit definition

The FU must be compatible with the initial scenario concept and transformation objectives. The significance of precisely defining the functional unit cannot be overstated. It is evident how different FUs produce distinct LCA findings, explaining the resultant heterogeneity.
Adopting exergy, on the other hand, significantly reduces this difficulty because EA may look at any period of the life cycle, but only via energy measures, necessitating the examination of very particular physical phases that do not entail heterogeneities in data selections. Examples for several industrial domains may be found in (Hischier and Reichart, 2003)or (Panesar et al., 2017; Gandiglio et al., 2019).When LCA and EA are combined, their FU must be the same, which implies that the input and output data gathered during the inventory stage must be compatible with both LCA and the EA metrics, as well as the performance to be measured in line with the criteria.

SUB-STEP 1.2 - Reference flow and system boundaries

Main criticisms of the ISO 14040 and 14044 recommendations have been raised, primarily by (Raynolds et al., 2000; Jolliet et al., 2015): 1) the definition of the boundaries is influenced by data availability or ease of retrieval, but the approach of excluding or including unit-processes in the analysis would result in respectively incomplete assessment and a false sense of completeness with insignificant added information; and 2) the approach of excluding or including unit-processes in the analysis would lead to respectively incomplete assessment and a false sense of (without considering the time loss by the practitioner in analysing extra data and processes). 2) The ratio between mass and energy in terms of environmental effect. 3) Selecting between process-based and input/output LCA. The former has a high rate of truncation mistakes and excludes capital goods. The latter has issues with data resolution and the unjustifiable shortening of recycling industrial sectors and process life cycle phases (Majeau-Bettez et al., 2011). 4) The ISO ignores the economic and social consequences. Once the functional unit, the reference flow, the system boundaries, and the beginning scenario S(t0) have been identified inside the system thinking, it is feasible to unambiguously determine the set of unit-

processes to consider in order to satisfy the objectives $S(t_x)$ out of $S(t_0)$.
Furthermore, EA, like LCA, requires a system boundary definition that corresponds to the description of the reference environment. The reference environment is chosen based on the system under consideration. The reference environment must be immediately accessible to the system and its attributes must not change as a result of interactions with the system. The exergetic performance of the system as a whole may be viewed by establishing the reference environment. Each subsystem functions within these parameters, with exergy input and output dictated by the potential in relation to the surrounding environment (Lozano, 2008). Gaudreau et al., in (Gaudreau et al., 2012) conducted a comprehensive investigation of the reference environment to have a better grasp of how exergy influences decision-making. EA addresses the heterogeneity issue mentioned above by focusing on the unique flows of material, work, and energy that occur throughout every process/activity, as it is required to construct mass, energetic, and exergetic flow balances. When EA is combined with LCA from a system thinking perspective, the LCA system boundaries must correspond with the reference environment in EA.

STEP 2 – Life Cycle inventory analysis

Even if LCA is combined with EA, the life cycle inventory (LCI) is the most onerous step in which material and energy flows are collated and measured.
Because the information acquired will aid in the anticipated effect assessment, the inventory review should be guided by the selection of impact assessment metrics. If the effect computation is an exergy research, the data acquired may be appropriate, implying that gathering a large amount of pollution data may be overlooked since they are either toxic or greenhouse gases but do not contribute significantly to exergetic losses.
LCI analysis is the computation of all raw materials, energy resources, and machinery (inputs) utilized in the process, as well as the estimation of their emissions to air, water, and soil (outputs) created across the whole life cycle, using the functional unit as a reference.

SUB-STEP 2.1 - Choice of the database

Both LCA and EA suffer from the quality of available data (mainly related to specific data, i.e., derived from the physical site of the system analysed), driven from databases, in the reasoning about heterogeneity of sustainability assessment, because the nature of these methods has a subjective component. Several difficulties have been highlighted in LCA studies including data quality and a lack of understanding (Edelen and Ingwersen, 2006). It is standard practice to fill gaps in inventory data with assumptions or statistical data. Problems with data quality can arise from either the practitioner's access to data (e.g., confidentiality) or an impartial lack of understanding of the whole system (e.g., available measures of process parameters, measurement quality. This, in turn, may cause uncertainty in analysis results: this should be maintained less than the tolerance permitted for the criteria employed in the LCA research (specified as starting assumptions or restrictions on the analysis) or the total tolerance admitted in the exergy calculation. The availability of data at the local and sectoral levels is very important. Although there is still a need for a database that is universally consistent and dependable for the LCI phase of a wide range of diverse manufacturing processes (Kellens et al., 2017). In the absence of databases belonging to the country or geographical region in which the assessment must be addressed, the practitioner who is obliged to utilize databases of other nationalities must pay attention to the data source. The databases, in fact, may exhibit a great diversity of impact outcomes, even if they are identical or even if they are two distinct update versions. As a result, the decision can entirely alter the LCA research and become a source of additional subjectivity in the evaluation, whilst, realistically, it translates into a range of variability in the dependability owing to a random-error created and transmitted along the LCI and the LCIA.

SUB-STEP 2.2 - Allocation

To prevent allocation, ISO 14040 advises breaking operations into sub-processes or widening system boundaries. Obviously, the independence of the sub-processes determines the quality of the physical and economic process decomposition (Ekvall and Finnveden, 2001). The system boundary expansion, on the other hand, results in a bigger and more intricate model. This latter strategy necessitates a thorough understanding of the causalities underlying processes, since utilizing alternative causality principles for allocation can result in considerable disparities in LCA outcomes. The decision between consequential modelling and attributional modelling, according to system thinking, is determined by the purpose and scope. EA is also susceptible to allocation, such as appropriately allocating the amount of electrical energy required annually by a specific machine that generates a variety of goods over the course of a year, and it is also subject to periodic power-on and power-off cycles.

SUB-STEP 2.3 - Local technical uniqueness

Neglecting technical uniqueness is equivalent to ignoring system attributes that might affect system costs. The challenges with local uniqueness are caused by a lack of accuracy in the characterization of the qualities of the transformation to be assessed. To have as little ambiguity as feasible, the data for the LCI should be geographically characterized (see, for example, (Nilsson et al., 2010)). On the contrary, unless the economic implications of energy consumption assessed by the analysis are taken into account, EA is not subject to geographical classification.

STEP 3 - Life cycle impact assessment

SUB-STEP 3.1 - Impact category selection

This problem is caused by a practitioner's or an impact's lack of information. Distinct information might lead to different system options while starting with the same purpose and scope. As a result, differing bodies of knowledge for the LCA will always result in incomparable LCA outcomes, as demonstrated by (Müller et al., 2005).
The lack of consistency across the many proponent groups of these categories is a significant source of variation in the LCA findings. The modest discrepancies between the suggested groups are mainly due to the modelling technique used, e.g., midway versus endpoint. Endpoints are less comprehensive and have greater degrees of uncertainty (UNEP, 2003), whereas midpoints are more difficult to understand since they are not immediately tied to an area of protection (Haes et al., 2002). Unfortunately, even in a system thinking, EA does not tackle this issue, because the final exergy flow analysis accounts for the quality of natural resource degradation but not the character of output goods or the possible harm to the environment.

SUB-STEP 3.2 - Space characterization

As for the system boundary definition, the issue here is the establishment of a tolerance value for the transformation to take into account, as well as its long-term performance. This element has no bearing on EA since its methods are not dependent on the time or spatial characterization of the process under investigation; hence, EA with LCA will avoid both spatial variation and geographical characterization.

SUB-STEP 3.3 - Time characterization

Even ISO 14042 (“ISO 14042,” 2000) admits that failing to include the dynamic nature of industrial and environmental factors might diminish the usefulness of LCA results. The decision is impacted by the declared objective and scope, therefore temporal characterization of the transformation should be regarded part of the limitations contained in the specification of $S(t_0)$ and $S(t_x)$. As a result, various temporal characterizations require distinct transformation characteristics to be considered, resulting in incomparable LCA outcomes. Time characterisation by the unit control volume is a common feature of EA.
The time-dependency of the affects refers to the irregularity of emissions across time, repercussions that take years to manifest, and impact comparisons that have resulted in such changes throughout the studied time horizon (Field et al., 2000). This time-dependence is usually neglected, and the effects are averaged. (Hellweg et al., 2005)demonstrated how the selection of two distinct time periods might influence the environmental effect profile of the same operation, namely trash incineration and groundwater pollution.

STEP 4 - Life cycle interpretation

The interpretation is a process that implies a work of subjectivity because when a decision must be made, the type of aggregation of the results, possible weighting and normalization are thus inevitable. To this aim, significant is the table 12.1 in Hauschild et al. book section (Hauschild et al., 2018). Depending on how the results are shown and argued, a practitioner can interpret them in one or multiple forms, and they condition the interpretation of the results by emphasizing the importance of certain phenomena rather than others, for example the formation of the ozone layer depletion rather than acidification.

In the case of an LCA, identifying a unique criterion for comparing alternatives is a more difficult task than in EA. The clear specification of the stakeholders' needs in a system thinking allows the selection between the system alternatives to be reduced to a question of production costs if the performances of the system alternatives fulfil the criteria equally.

On the other hand, EA provides an ideal framework for evaluating this work since it is a measure of the reversibility of the processes, possibly decreasing the interpretation heterogeneities inherent in the LCA. The material, energy, and other streams all participate in the process and are changed into the product and waste streams. The exergetic yields associated to the exergetic balances of the process/activity itself provide the performances (or yields) of a specific process or activity. To summarize, it is typically difficult to analytically express the many types of uncertainty outlined in LCA. Probability distributions may be used to characterize random variability in input parameters, missing or incomplete data, and other sources of uncertainty in LCA, which are likely the most prevalent sources of uncertainty. A variety of LCA uncertainty analysis strategies have used probability distributions (Geisler et al., 2005; Guo and Murphy, 2012; Ross and Cheah, 2019; Lima et al., 2020). So far, no LCA uncertainty analysis approach has been devised to assist the characterization of this trade-off in order to make the evaluation as comprehensive as possible while being tractable in terms of decision making. The analysis of uncertainty on EA, on the other hand, on the other hand, is faster since the major source of error propagation is related to the sensitivity of the measuring devices of the in and out flows, and missing or incomplete data-sets are unlikely to occur. Using efficiency reports also substantially simplifies the comprehension of the Overall Analysis. In this scenario, there are additional assessments of uncertainty and sensitivity in Exergy Analysis in the literature (Ege and Sahin, 2014; Boyaghchi and Molaie, 2015; Javadi et al., 2020).

To summarize and conclude, addressing variability in sustainability assessments is an appealing issue, and eliminating it is a lofty goal. The job is mostly useful to the business partner and stakeholder. The problem of heterogeneity and subjective decisions is addressed by attempting to use system thinking as a procedural guideline. Systemic thinking provides a fresh method of thinking about existing problems by removing the traditional glasses of mental models, assumptions, and beliefs that allow us to perceive problems from the same perspective all of the time. To begin looking at situations holistically, it is a process of discovery that takes time. We demonstrated that including Exergy-based Analysis in Life Cycle Thinking is a useful method for reducing sources of variability throughout the whole process design. While introducing EA into the LCA approach helps to reduce sourced heterogeneity, it does not guarantee comprehensive coverage of the system under evaluation. Recommendations have been developed as a series of actions to take, resulting in a kind of reference operational framework that may decrease decisional inconsistencies that arise from several practitioners examining the same system.

Using the system thinking, that systems of different types with varied beginning circumstances and/or aims might be similar. By formalizing the links between the many unit-processes and the overall process and their proportional impacts, it is feasible to harmonize the data. The causes of heterogeneity have been identified owing to system thinking, which provides a cohesive holistic perspective on the manufacturing processes. This viewpoint appears to address the subjective practitioner's choices, which raises a coherence issue between the decision at each stage of sustainability assessment and its inputs (the initial scenario, the objectives of the transformation and the sustainable performances to be assessed). The work has also offered useful fuel for thought, which may improve the interpretability of current ISO standards. Furthermore, the study could be expanded and refined with a quantitative assessment (rather than only qualitative, as has been done thus far in this work) of the interactions between the system's components and the impact they have on the degree of uncertainty in the final results, and thus on sustainability.

2.2. DATA-DRIVEN APPROACH

Each research and development path is a dynamic journey of learning through study and experience of the environment, from which we gather property details, often readily quantifiable, often qualitative. Observation also allows one to link events to property records. It is through repeated interactions that we can deduce certain patterns that link events to data and data into actions. In the case of scientific discovery, these patterns and relations are formalized as rules and equations, the data as properties and factors, and the results as event measurements.

Because of the different operating needs and limits of the underlying industrial processes, the design of model-based process monitoring and fault detection systems has been a fascinating research issue for several decades. The well-established model-based methodologies have been effectively deployed to a variety of processes for industrial electronics (He et al., 2013), automatic control systems (Kuestenmacher and Plöger, 2016), and so on, based on physical and mathematical understanding of the industrial processes. Model-based systems need significant physical and mathematical understanding of the process. The well-established model-based methodologies might be successfully utilized after the design of the process model based on the fundamental principles. On the other hand, a data-driven approach helps to extrapolate further knowledge from reality that human experience alone is not able to capture, and has the added benefit of checking associations between various variables and findings, learning unforeseen trends in nature, and helping one to uncover new science laws or, still more, performing predictions in the absence of those laws (Montáns et al., 2019). Data-driven approach have been used for some years in industrial processing both for monitoring, maintenance and for predicting (Solomatine and Ostfeld, 2008; Ding et al., 2011; Sutharssan et al., 2015; Bousdekis et al., 2021).

There is considerable overlap between data-driven techniques and data mining (Clarke et al., 2009). Data mining is the analytical stage in the "knowledge discovery in databases" (KDD) process, which entails using data analysis and discovery algorithms to uncover patterns in big data sets (U. Fayyad et al., 1996). Data mining is the process of extracting useful information from massive volumes of data contained in databases. It is the extraction of knowledge from data. Data-driven techniques relate to the capacity to teach computers to learn without explicitly programming them.

Data, which are generated by sensors, measuring devices, machinery, and quality control tools, can be transformed into powerful tools for improving production planning, optimizing operational processes, and influencing decision-making by accurately analysing all available data. Data analysis has always been a key-activity in the management of a typical manufacturing process, and its importance has grown as the number of flows and sources responsible for generating them has increased. Sensors, measuring devices installed along production lines, industrial plants, and quality control tools all contribute to their creation today: a plethora of sources capable of producing data so important to businesses that it can influence not only production planning, but also operational process optimization and decision-making process correction through accurate analysis. In terms of production planning, Enterprise Resource Planning (ERP), Programmable Logic Controller (PLC) systems and Manufacturing Execution Systems (MES) can now provide extremely detailed views of all industrial processes, reaching previously unthinkable levels of granularity (Khan et al., 2017).

With the rapid growth of automation and information, data collecting devices are extensively utilized in smart factories, and manufacturing data in factories are becoming larger (Volume), fast (Velocity), and diverse (Variety) (Katal et al., 2013). In general, factories are motivated by a causal relationship and use factory simulation models and algorithms to increase production efficiency, product quality, and other workshop performance.

To gain adaptive control of a manufacturing process, knowledge of both the process and the environment must be gained. This expertise can be gained by mining vast quantities of data gathered during the production process monitoring. This allows for the investigation of process parameters as well as the correlations of process parameters, environmental parameters and machine faults. This allows for the establishment of information about the process and its relationship to the environment, which can then be used for adaptive process management.
The data culture has gone through various “ages” of development and innovation, with each subsequent phase incorporating the previous technology. Figure 2.15 contains a summary. On the horizontal axis, there is a timeline marked in years with two types of related references: manufacturing-related technological solutions (for example, I4.0 around the year 2012) and technological information solutions (such as, for example, IoT around 2000). On the vertical axis data’s increasing volume, variety, and complexity are showed. The graph’s dial has four sets, one inside the other, that distinguish the sequence of ages over time. In addition, the types of data that characterized each age are indicated. As can be seen, this evolution includes the digitalization process.

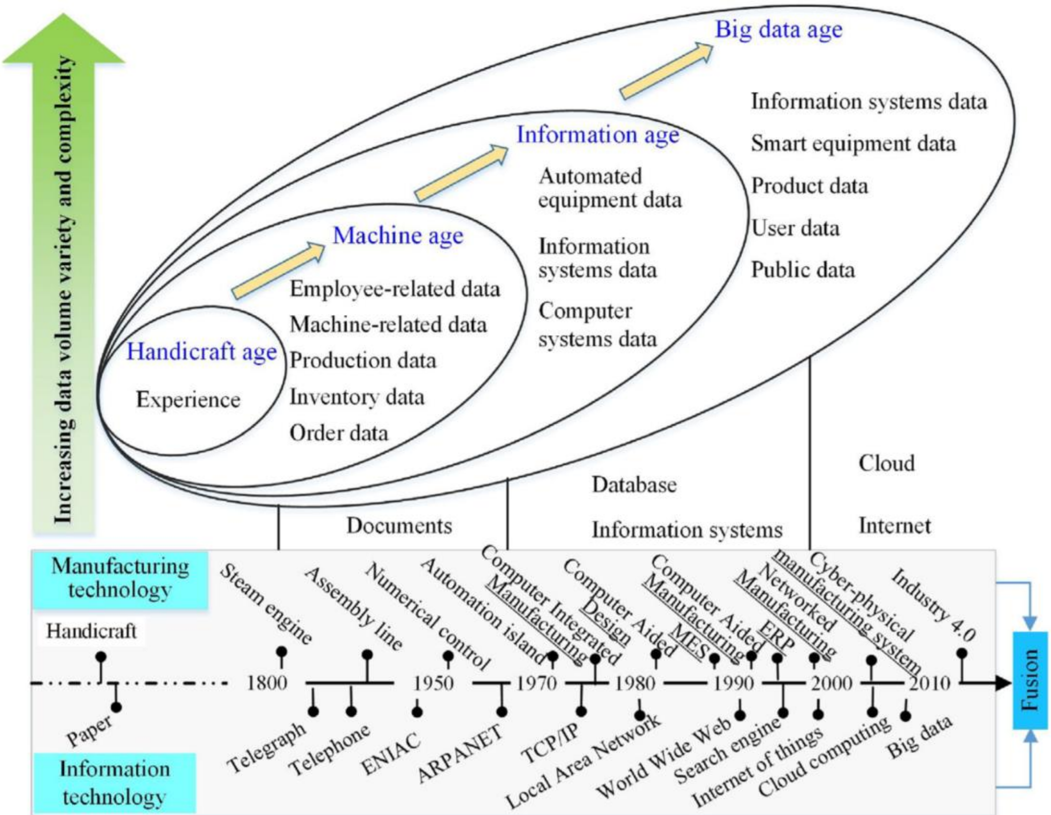


Figure 2.15 - Evolution of data in manufacturing (Tao et al., 2018)

Companies who do not process data limit their access to the very data that might sharpen their competitive edge and provide important business insights. That is why it is critical for all businesses to grasp the importance of processing all of their data, as well as how to do it.
Every data-driven approach starts with data in its raw form and converts it into a more readable format by the specific algorithms. Data processing methods for big data analytics involve (Bhatnagar, 2018):

1. Data gathering. The type of raw data collected has a huge impact on the output produced. Hence, raw data should be gathered from defined and accurate sources so that the subsequent findings are valid and usable.
2. Data storage. The data must be stored in a data warehouse, data vault, or data lake. Here, data and metadata are stored for further use. This allows for quick access and retrieval of information whenever needed, and also allows it to be used as input in the next data processing cycle directly.

3. Data preparation or data cleaning. Involves improving data quality. It is the process of sorting and filtering the raw data to remove unnecessary and inaccurate data. Raw data is checked for errors, duplication, miscalculations or missing data, and transformed into a suitable form for further analysis and processing. This is done to ensure that only the highest quality data is fed into the processing unit. Quello che si deve andare a verificare è che i
4. Data input. Clean raw data are converted into machine readable form and fed into the processing unit. This can be in the form of data entry through a keyboard, scanner or any other input source.
5. Data analysis. Raw data is subjected to various data processing methods using machine learning and artificial intelligence algorithms to generate a desirable output. This step may vary slightly from process to process depending on the source of data being processed (data lakes, online databases, connected devices, etc.) and the intended use of the output.
6. Data presentation. data is finally transmitted and displayed to the user in a readable form like graphs, tables, vector files, audio, video, documents, etc. This output can be stored and further processed in the next data processing cycle.

2.2.1. Coping with Big Data and their Heterogeneities

When it comes to gathering data for industrial IoT in the manufacturing world, it's not just about quantity. It all comes down to the accuracy of the data being collected from various machines; data that allows for interpretation and decision-making now boosts both productivity and strategic edge in the long run (World Manufacturing Foundation, 2020).

Because we observe the temporal evolution of a system, real manufacturing processes are dynamic, not static. The added richness of dynamic data allows for better understanding and intrinsic knowledge, but it can be difficult to figure out how to use the richer temporal data to gain new insights into a system's behaviour and structure (Hu, 2020).

The development of advanced calculation instruments has given rise to new forms of data collection. As a consequence, data often have dynamic dependency structures. These dynamic systems usually necessitate non-standard statistical methods that are computationally intensive (McGoff et al., 2012). Conventional tools usually assume that the data, or any appropriate transformations of it, obey a normal distribution. This presumption no longer holds expressly in these situations. There have been remarkable advances in mathematical methods for analysing such data over the last twenty years. Unfortunately, the advancement of computational applications and devices has not remained static with these methodological advancements, but practitioners now have a plethora of highly advanced methods at their disposal for dealing with complicated results. This has enabled their acceptance and implementation in the solution of important substantive problems in a variety of disciplines, especially engineering and finance, as well as medicine and health. Multivariate data, which consists of a combination of discrete (i.e., categorical, binary, count) and continuous variables, is an especially typical example of non-standard correlated data in use. (de Leon and Chough, 2013).

Furthermore, as measuring methods develop, data collection becomes less expensive and simpler. Data is often obtained from various channels or networks on the same sample collection, and is referred to as multi-view or multi-modal data. One of the main challenges associated with the analysis of multi-view data is that measurements from different sources may have heterogeneous types, such as continuous, binary, and count-valued.

In the manufacturing sector, big data analysis aids in the correction of parameters that underpin individual production processes. In the case of complex operations, i.e., those influenced by a large number of parameters, manufacturing companies can actually use the data provided in real time by sensing systems to modify these same parameters, thereby improving productivity, quality, and operational efficiency (Nagorny et al., 2017). Companies must be able to integrate all types of information (which may cause heterogeneities in the construction of an all-inclusive database) in order to conduct comprehensive analyses, regardless of the geographical location of the machinery and plants that generate it.

If we wish to extract valuable information and knowledge from a manufacturing process, we must first address data integration issues so that we can apply effective empirical techniques to detailed and uniform data. This type of practice is referred to as data process knowledge discovery (U. M. Fayyad et al., 1996).

Dealing with structured, semi-structured, and unstructured data at the same time is referred to as data heterogeneity. The ultimate goal of this task is often to obtain a coherent or cohesive understanding of the real-world institutions described in the data sources.

Every stage of Big Data analytics has its own set of challenges. These involve real-time computation, dealing with diverse data types, and parallel data processing, among other things. Big Data provides access to a vast amount of dynamic, heterogeneous, and informative data that is not always reliable.

The emergence of too many data from various data providers opens up a plethora of new doors for knowledge discovery. At the same time, it poses a new problem that can only be overcome if the data sources of interest are analysed in a coherent and interconnected manner. Indeed, scientists have accepted that the study of single-type datasets cannot explain the whole-science phenomena (Mandreoli and Montangero, 2019).

As a consequence, there are some critical problems to solve when dealing with Big Data, and new ideas are needed. This issue is called 4Vs (Cappa et al., 2021):

1. the *volume* of data coming from various sources. This is true for the scale of the data collection. It's the most prominent characteristic associated with big data. The term "volume" refers to the massive datasets that organizations are trying to use to improve decision-making policies;
2. the *velocity* at which data is generated and updated: data sources can be flexible and changing. This applies to the rate at which data is obtained as well as the rate at which it should be analysed and based;
3. the *variety* of data from multiple sources, even though they describe the same objects. This applies to a data set's structural heterogeneity. Companies will now use structured, semi-structured, and unstructured data owing the know-how to technological advancements;
4. the *veracity* of data. This is a reference to data uncertainty. Since data sources may be in conflict and data may be replicated from one source to another, the degree of authenticity associated with some types of data is known as veracity (Chen et al., 2016). To reduce confusion, analysts shall create context around the data.

In literature this issue has been deeply investigated such as the 4Vs enlarged to 10Vs, adding to the first four (Khan et al., 2018):

5. the *variability* of data. It can refer to a variety of things. The number of inconsistencies in the data is one. In order for any useful analytics to take place, they must be discovered using anomaly and outlier detection tools. Big data is also changeable due to the plethora of data dimensions arising from a variety of distinct data kinds and sources. Variability may also refer to the erratic rate at which huge data is put into the database, as opposed to velocity;
6. the *validity* of data. It relates to how accurate and precise the data is for the purpose for which it is designed. According to Forbes, data scientists spend an estimated 60% of their time cleaning their data before performing any analysis (Forbes Press, 2020). The benefit from big data analytics is only as good as its underlying data, proper data governance processes must assure uniform data quality, standard definitions, and metadata;
7. the *vulnerability* of data. Big data brings new security concerns. After all, a data breach with big data is a big breach;
8. the *volatility* of data due to the velocity and volume of big data. It is essential to define criteria for data currency and availability, as well as to assure speedy retrieval of information when needed, because the costs and complexity of a storage and retrieval procedure are increased with massive data;
9. data *visualization*. Current large data visualization solutions suffer technological issues owing to in-memory technology restrictions, as well as insufficient scalability, functionality, and reaction time.

Combining this with the plethora of variables coming from big data's diversity and velocity, as well as the intricate interactions between them, creating a meaningful representation is difficult;

10. data *value*. The primary purpose of any data analysis is to derive strategic value from data. Big data may provide significant value in a variety of ways, including better understanding your consumers, targeting them appropriately, streamlining processes, and boosting machine or company performance. Before beginning on a big data plan, you must first comprehend the possibilities as well as the more difficult qualities.

One method is data fusion, which involves merging several less reputable databases to produce a more precise and usable data point, such as social comments appended to geospatial location information. Advanced mathematics that accepts complexity, such as rigorous optimization methods and fuzzy logic approaches, is another way to treat it.

Consequently to the 4Vs issue, there are following types of data heterogeneity (Jirkovsky and Obitko, 2014):

• Syntactic heterogeneity occurs when two data sources are not expressed in the same language.

• Conceptual heterogeneity, also known as semantic heterogeneity or logical mismatch, denotes the differences in modelling the same domain of interest.

• Terminological heterogeneity stands for variations in names when referring to the same entities from different data sources.

• Semiotic heterogeneity, also known as pragmatic heterogeneity, refers to people's differing interpretations of events.

• Spatial-temporal heterogeneity happens more often while observing over a long period of time, or when the time stamp is used to distinguish specific phenomenon within the method. It is thought that the instances are structurally similar to one another in various spatial and temporal domains (e.g., different regions on the machine but with different sampling frequencies). Ignoring these dependencies during data processing may result in findings that are inaccurate and difficult to understand (Atluri et al., 2017).

• Data source heterogeneity draws attention to the fact that each data source can have a different data model. Data with identical meanings can be represented differently in each data source. Furthermore, they can contain contradictory information.

• Dependency heterogeneity, which results from the assumption that data are always isolated elements that are not optimized for use in a data integration framework. They cannot be compelled to behave in such respects. As a normal result, they will alter their data or functionality without warning.

• Distribution heterogeneity that refers to spatial distribution of data sources. The required system configuration should allow for the potential delay in communicating with data sources.

2.2.2. Data Quality and Significance

To ensure that the data used is precise, consistent, and complete, data quality management is critical. To compound the complexity of handling data quality, data is rapidly evolving, with increasing sizes, shifting formats, and different distribution methods. Data may become obsolete and unusable if it is not properly maintained. The following properties must characterize high-quality data (Madhikermi et al., 2017):

- Validity – a measure of how well data conforms to required value attributes and ensuring that data obtained are in the proper format and style. An example metric for validity is finding the percentage of data that have values within the domain of acceptable values.
- Completeness – data monitoring to ensure that data requirements allow for lost or insufficient data. The measure of completeness can be assessed in two ways: at the record level or at the attribute level. An example metric for completeness is the percent of data fields that have values entered into them
- Accuracy – sufficient accuracy for the intended purpose when taking into account expense, usage, and effort. Data accuracy is critical in large organizations, where the penalties for failure are high. An example metric for accuracy is finding the percentage of values that are correct compared to the actual value.

- Relevance – entails being appropriate for the intended reasons, as well as having a proper feedback mechanism and quality assurance.
- Integrity – to maintain all the data quality metrics when they are moved or merged between different systems. Typically, data stored in multiple systems breaks data integrity. An example metric for integrity is the percent of data that is the same across multiple systems.
- Consistency – to maintain synchronicity between different databases. An example metric for consistency is the percent of values that match across different records/reports.
- Reliability – a continuous data collection method over time and within systems.
- Timeliness – reflects the accuracy of data at a specific point in time, that is, being available at the appropriate frequency to allow for timely decision making. An example metric for timeliness is the percent of data you can obtain within a certain time frame, for example, weeks or days.
- Auditable – modifications to a collection of data must be traceable, and data transformation must be verifiable.
- Replicability – allowing a data operation to be reproduced, either by the same practitioner or by another as well.

The term Data Quality refers to activities and procedures that are aimed at analysing (and possibly improving) the quality of a data collection. To that end, the size of the data quality is proposed as a tool (qualitative) for evaluating data quality. First, the logical structure used to represent the data is examined to ensure that it is adequate and appropriate for obtaining a provided with the requisite quality characteristics (Batini et al., 2009). First, the logical structure used to represent the data is examined to ensure that it is sufficient and appropriate for obtaining data with the necessary quality characteristics. Similarly, the process-level analysis ensures that the method used to observe or collect data is appropriate. Data-level analysis, on the other hand, analyses stored data directly, without regard for the type or manner in which it was obtained. It is critical to emphasize that the quality at the scheme level influences the quality at the process level, which in turn influences the quality of the final data. However, analysing and correcting qualitative inconsistencies at the model or process level is not always feasible. In these situations, data-level analysis is thus the only viable option.

Before handing over the dataset to the algorithms, various techniques should be used to preserve data quality and guarantee that any erroneous data is found as soon as feasible and then manually or automatically repaired (Mills, 2009).

- <u>Data discovery</u> is a frequently disregarded and undervalued component of any data-related activity. People frequently make incorrect assumptions about their data since most people only see the facts from their own point of view. Data discovery, on the other hand, is an important aspect of the design process since it gives input for scope definition and project estimates. Discovery is a rigorous investigation of the data itself, with the goal of discovering correlations inside and between datasets. All applications that use the data and need to be adjusted or updated must also be considered, especially in the event of data migrations. As a result, data discovery in bigger businesses may be a multi-team effort that frequently crosses departmental boundaries.
- <u>Data cleansing</u> is the process of cleaning "dirty" data in its original place before using it in any data transformation. The data cleaning process is frequently integrated into the business logic, with the data being cleaned in the transformation but remaining intact in the source system. During data discovery, you will frequently realize that the data cannot be used in its present state and must first be cleaned. Low data quality may be caused by a variety of factors, ranging from basic ones (anything involving human data entry is likely to contain mistakes such as typos, missing data, data abuse, and so on) to complicated difficulties caused by incorrect data handling techniques and software defects.
- <u>Data validation</u>, which works hand in hand with data cleansing, is a key procedure in ensuring appropriate data quality in the target system. Any data that does not fulfil the validation requirements is flagged by the validation process. Data that fails the validation phase is flagged for clean-up. It might be either a manual or an automated procedure. Of course, for bigger systems, a high degree of

automation is desired, if not required. With automated data validation, the great majority of data may be rectified without the need for human interaction, resulting in fewer mistakes and the elimination of bottlenecks. Nonetheless, some manual validation may be required, particularly for the most critical data that cannot be permitted to pass without human inspection or correction.

- Reporting is an important aspect of ensuring data quality. Well-executed reporting guarantees that stakeholders receive all status information immediately and can respond in a timely manner. This, in turn, reduces the time required to rectify any data quality concerns and improve any processes that consistently result in bad quality data, allowing to manage the bad data that are inevitably in the systems.

To sum up all the gaps and challenges related to the big data features, (Zhang et al., 2015) and (Wang, 2017) schematized a complete picture that is reported and adapted in the following Table 2.3

Table 2.3 - Big data issues, gaps and challenges

Big data 10Vs issues	Gaps	Challenges
1. Volume 2. Velocity 3. Variety 4. Veracity 5. Variability 6. Validity 7. Volatility 8. Vulnerability 9. Visualization 10. Value	a) Data Consistency b) Data Integrity c) Data Identification d) Data Aggregation e) Data Confidentiality f) Data Interpretability g) Data Complexity h) Data Heterogeneity	- Data Scale - Incomplete Data - Data Usefulness in decision making - Data Processing and Parallel Processing - Data Quality - Data Durability - Structured, Semi-structured, Unstructured Data - Illegally Tampered Data - Human Collaboration

(Zhou et al., 2017) provided an outstanding review, and the next part examines each of the aforementioned pre-processing difficulties, as well as the obstacles and recommended methods to decrease the risk associated with them.

An application for discussing and resolving some of these difficulties is in (Stief et al., 2019) in a case study, they attempted to fill the gap with a heterogeneous benchmark dataset based on an industrial-scale multiphase flow facility. The study gathered data from varied operational situations, both with and without generated faults, to create a multi-rate, multi-modal dataset and highlight the relevance of the pre-processing step in big data aggregation and analysis.

The Internet of Things (IoT) is directly connected to big data and their life cycle in industrial process management. The advancement of internet technology enabled the possibility of a more extensive and robust network communication between the items. Every object in IoT is recognized as a node and is connected to each other in a network; this type of system enables information sharing such as receiving and transmitting.

The design and implementation of IoT for unique applications may vary, but there is a common architecture approach to be followed for IoT project execution. The best, fast, reliable, and secure convergence of the information technology and communication technology will only happen when an effective IoT architecture layer is built. (Kumar and Mallick, 2018) in their work illustrated how the layer architecture of IoT changed over time and produced a graphic comparison (see Figure 2.16): One of the first and most fundamental IoT designs introduced is three-layer architecture. It is really handy and simple to apply. The perception layer, network layer, and application layer are the three layers present in the architecture. The stated three levels describe the operation of IoT; however, they cannot provide a trustworthy solution due to the higher aspects of IoT.

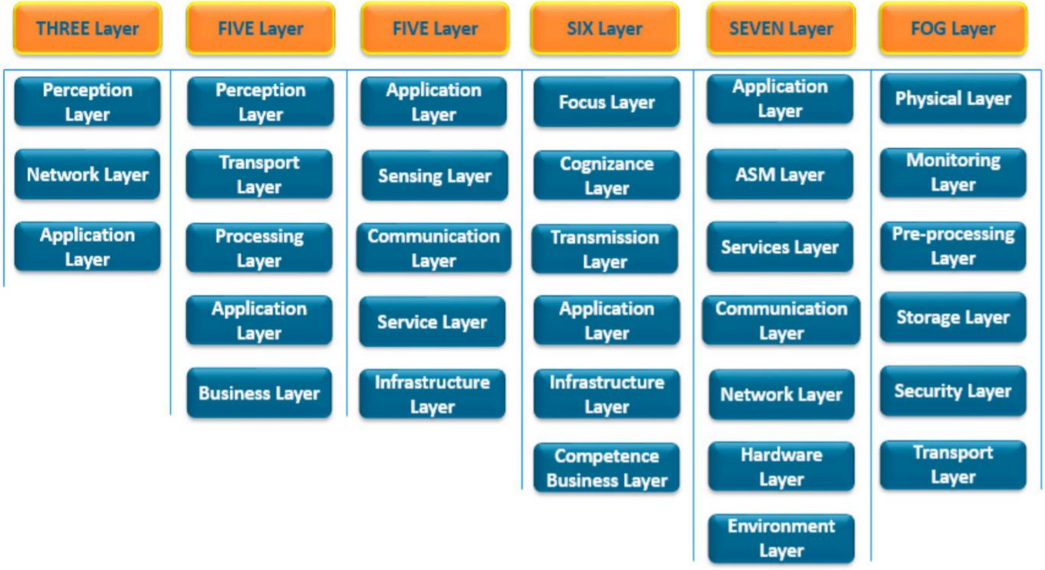


Figure 2.16 - IoT architecture layers evolution (Kumar and Mallick, 2018)

Cloud computing (FOG Layer in Figure 2.16) appears to be the tipping point: it is a more flexible and scalable method that enables numerous services for IoT systems. These services include data storage choices, software tools and analytics, an appropriate platform, and core development infrastructure. With the cloud facility, users may have visualization, machine learning, and data analytics choices for larger volumes of data. Because of the ambiguity of the information detected and produced in the form of data by IoT sensors, cloud-based architecture became popular in IoT systems. Most IoT architectures use cloud-based data processing technologies to provide centralized control over data. In this sense, it emerged as a basic layer inside the IoT architecture as a specialized layer between data sources and the structured database: a data integration layer that handles ETL processing (pre-processing layer in Figure 2.16).

To conclude, it is critical to implement effective big data cleaning methods in order to increase data quality. Data virtualization and data lakes are effective methods for facilitating data integration. In Big Data analytics, traditional data mining and deep learning approaches have drawbacks. Deep learning can analyse and learn from vast volumes of unsupervised data; hence, it has potential in Big Data analytics where raw data is mainly unlabelled and un-categorized.

The following figure closes this paragraph with an overview to the life cycle framework of manufacturing data, highlighting the IoT layers and each phase/operating level to reach data quality in big data analytics.

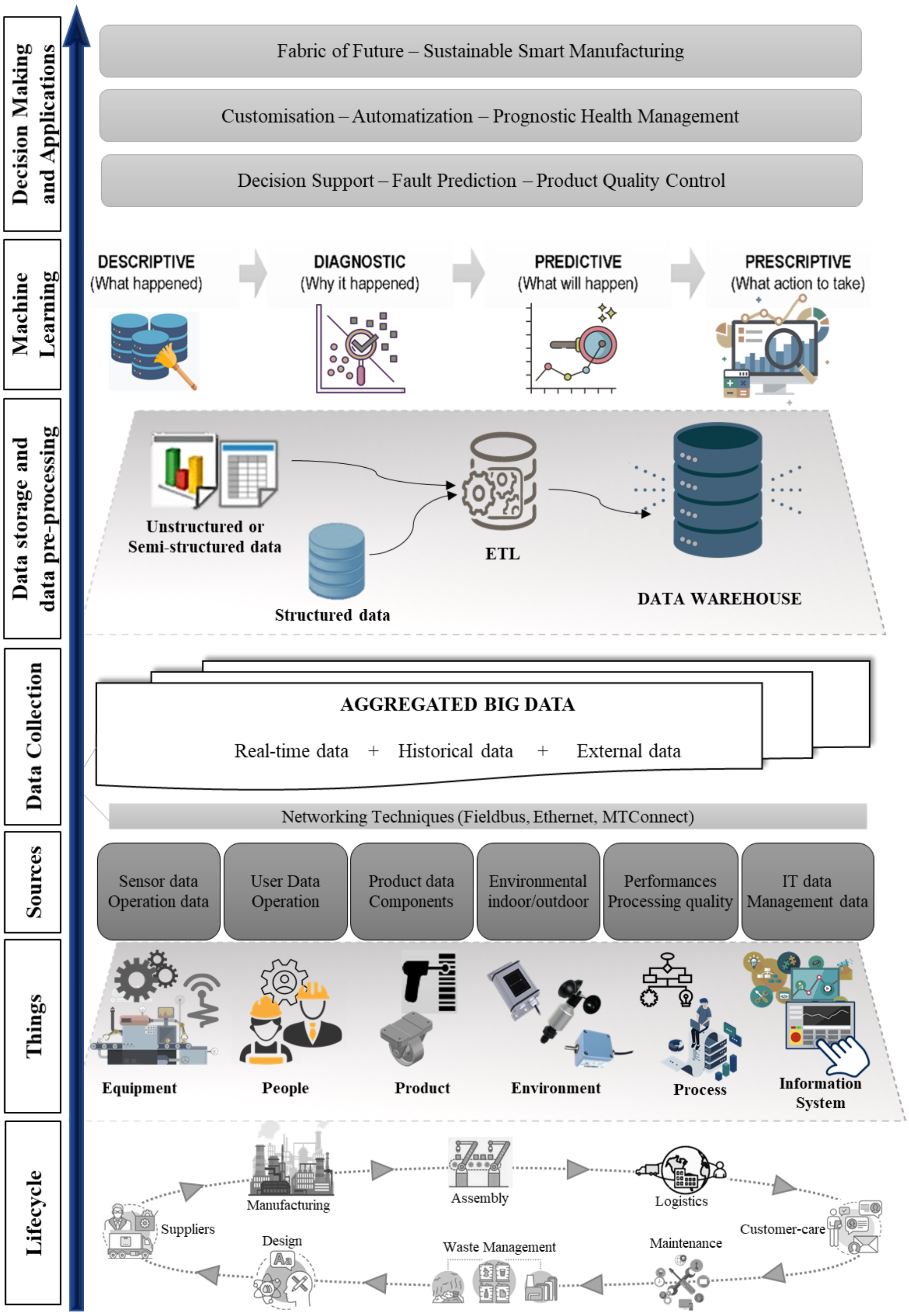


Figure 2.17 - Life cycle framework of manufacturing data

2.2.3. Machine Learning techniques for Manufacturing Systems

Machine Learning (ML) is the most growing field in Computer Science. Everyone is now talking about ML-based solution techniques for a specific issue set. ML is a subset of AI in which computer algorithms are used to learn independently from data and information.

ML is a critical component of smart manufacturing since it provides accurate insights for better decision making. As illustrated in Figure 2.18, ML has been extensively researched at many phases of the manufacturing lifecycle, including concept, design, evaluation, production, operation, and sustainment (Zhang et al., 2017). (Harding et al., 2006) examined data mining applications in industrial engineering, focusing on several areas such as production processes, operations, defect detection, maintenance, decision support, and product quality enhancement. (Esmaeilian et al., 2016; Kang et al., 2016) examined the evolution and future of manufacturing, highlighting the role of data modelling and analysis in manufacturing intelligence. To satisfy present and future demands for efficient and reconfigurable production, smart manufacturing also needs prognostics and health management (PHM) capabilities (Vogl et al., 2019).

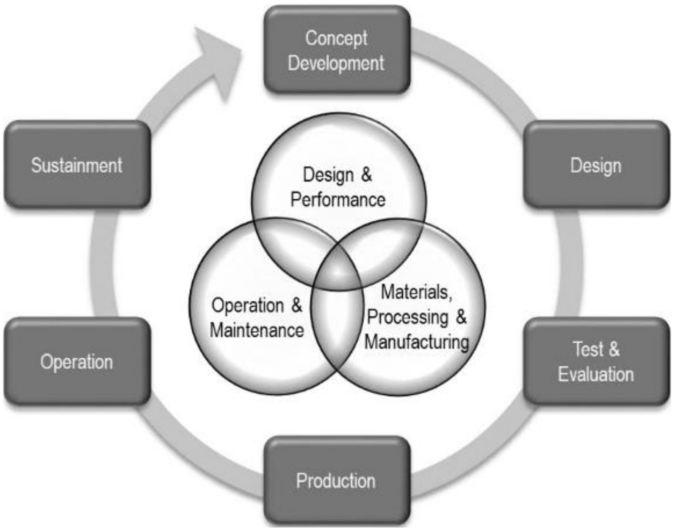


Figure 2.18 - Typical application scenarios of machine learning in smart manufacturing (Wang et al., 2018)

What are the mathematical foundations at the base of ML? In summary, it starts with a non-parametric model trained on existing data. The training data is supplied into an untrained or partially trained model, and then computations are done to calculate the deviation from the expected outcome.
ML techniques are defined as discovering a target function (f) that optimally maps input variables (X) to output variables (Y).

$Y = f(X)$

This is a common learning problem in which the aim is to predict the future (Y) given new instances of input variables (X). The function (f) is unknown as well as how it looks like or what its form is. If we did, we would utilize it directly rather than learning it from data through machine learning methods. It's more difficult than that. There is also error (e), which is unrelated to the input data (X).

$Y = f(X) + e$

This issue might be due to a lack of characteristics to adequately define the optimum mapping from X to Y. This error is known as irreducible error since it cannot be reduced no matter how proficient the model is at predicting the target function (f). To put it another way, learning a function from data is a challenging challenge, which is why the discipline of machine learning and ML algorithms exists.
The most frequent form of ML is learning the mapping Y=f(X) in order to predict Y for fresh X. This is known as predictive modelling or predictive analytics, to create the most accurate forecasts possible. As a result, the practitioner is less concerned with the shape and form of the function (f) and more concerned with the fact that it produces correct predictions. To learn more about the relationship in the data, we may discover the mapping of Y=f(X). This is known as statistical inference.
When learning the function (f), it involves estimating its shape based on the facts supplied. As a result, there will be some inaccuracy in this estimate. It will not be an exact estimation of the underlying hypothetical optimum mapping from Y onto X. Much work in applied machine learning is spent attempting to improve the estimate of the underlying function and, as a result, the performance of the model's predictions.
The error is reduced by altering the model's parameters in accordance with predefined mathematical criteria. In ML, the training process is essentially an error minimization procedure in which the model parameters are defined in such a way that the model can duplicate the training data as precisely as possible. However, the underlying model will also be capable of predicting input data that was not included in the training data. In other words, the model has the ability to generalize. The talent in ML is in picking a suitably sophisticated but

not too huge model that can accomplish good generalization across the complete range of relevant input data with only a little quantity of training data.

Some prominent ML techniques/methods include dimensionality reduction (Richter et al., 2015), clustering (Dietrich et al., 2015), association rules (Dietrich et al., 2015), classification (Mikut and Reischl, 2011), regression (Mikut and Reischl, 2011), topic modelling (Dietrich et al., 2015), time series analysis (Dietrich et al., 2015) and collaborative filtering (Twardowski and Ryzko, 2014). These are used to do analytics and forecast future trends based on current patterns and correlations between data in a particular dataset.

(Fahle et al., 2020) performed an extensive review of the potential uses of ML algorithms in manufacturing for handling particular issues such as scheduling, cost and energy projections, quality control, predictive maintenance, logistics, and so on.

There is a significant distinction between classification and regression problems. Basically, classification is concerned with predicting a label, whereas regression is concerned with predicting a quantity. A classification problem can be turned to a regression problem in some instances. A label, for example, can be turned into a continuous range. Some algorithms already achieve this by forecasting a probability for each class, which may then be scaled to a given range. If the class labels in a classification issue do not have a natural ordinal connection, converting from classification to regression may result in unexpected or bad performance because the model may learn a false or non-existent mapping from inputs to the continuous output range.

While Supervised Learning, Unsupervised Learning, Semi-supervised Learning, Reinforcement Learning, and Deep Learning are the primary categories into which ML approaches may be classified (Zhou et al., 2017).

- **Supervised learning:** Occurs when an algorithm learns from example data and associated target responses that can consist of numeric values or string labels — such as classes or tags — in order to later predict the correct response when posed with new examples. The supervised approach is, indeed, similar to human learning under the supervision of a teacher. The teacher provides good examples for the student to memorize, and the student then derives general rules from these specific examples.
- **Unsupervised learning:** Occurs when an algorithm learns from plain examples without any associated response, leaving the algorithm to determine the data patterns on its own. This type of algorithm tends to restructure the data into something else, such as new data features that may represent a class or some new values helpful for additional analysis or for the training a predictive model.
- **Semi-supervised learning:** it is a learning issue with a few labelled instances and a huge number of unlabelled examples. Learning issues of this sort are difficult to solve because neither supervised nor unsupervised learning algorithms can effectively employ combinations of labelled and untellable data. As a result, semi-supervised learning algorithms with particular features are required. The purpose of semi-supervised learning is to understand how mixing labelled and unlabelled input affects learning behaviour and to create algorithms that take advantage of this combination.
- **Deep Learning:** in terms of feature learning, model design, and model training, it differs from classical machine learning. Deep learning combines feature learning and model creation in a single model by using different kernels or tuning the parameters through end-to-end optimization. Its deep neural net design with multiple hidden layers is essentially multi-level non-linear computations. Deep Learning and Neural Networks are computing systems inspired by the human brain. Health assessment, performance prediction, and defect detection are the key applications of utilizing neural networks and deep learning in manufacturing. Its goal is to make complex manufacturing fully autonomous.
- **Reinforcement learning:** Occurs when you sequentially present the algorithm with examples that lack labels, as in unsupervised learning. The machine is placed in an environment where it is constantly trained through trial and error. The learning agent, in particular, interacts with an environment and learns the best policy on the fly based on feedback from that environment. At each time step, an agent examines the state of the environment, selects an action, and monitors the input it receives from the environment. There are several key components to the feedback from an agent's activity. The ensuing condition of the environment after the agent has acted on it is one component. Another factor is the

reward (or penalty) that the agent receives for executing that specific action in that specific condition. The incentive is carefully set to coincide with the goal for which the agent is being trained. The agent modifies its decision-making policy based on the state and reward to maximize its long-term reward. In contrast to supervised and unsupervised learning approaches, the machine learns from the past and attempts to capture the greatest available information in order to make appropriate business judgments. Reinforcement learning is connected to applications for which the algorithm must make decisions (so that the product is prescriptive, not just descriptive, as in unsupervised learning), and the decisions bear consequences.

- **Ensemble learning**: it is an approach used because of its shown improvement in performance in terms of predictions. It uses the idea of using several weak learners and combining them to form a strong learner. It takes a majority vote approach in terms of classification, and this is what makes the approach more robust as compared to using single classification algorithms independently. There are different types of ensembles learning approaches, mainly Bagging and Boosting. Bagging is a method in which multiple trees are being built over different subsets of the data. These subsets are drawn from the original dataset, with replacement. Hence, Bootstrapping is done and a model is built on each of the subsets individually. Boosting, on a high level uses algorithms that use weighted averages to convert weak learners into strong ones.

According to what has been described above, Figure 2.19 depicts a grouping of the key ML methods and the corresponding main algorithms, with the goal of schematizing and summarizing the universe of ML. There are obviously many alternative algorithms, but they are not relevant to the topic of this dissertation. A more complete list of ML algorithms can be found in (Brownlee, 2016).

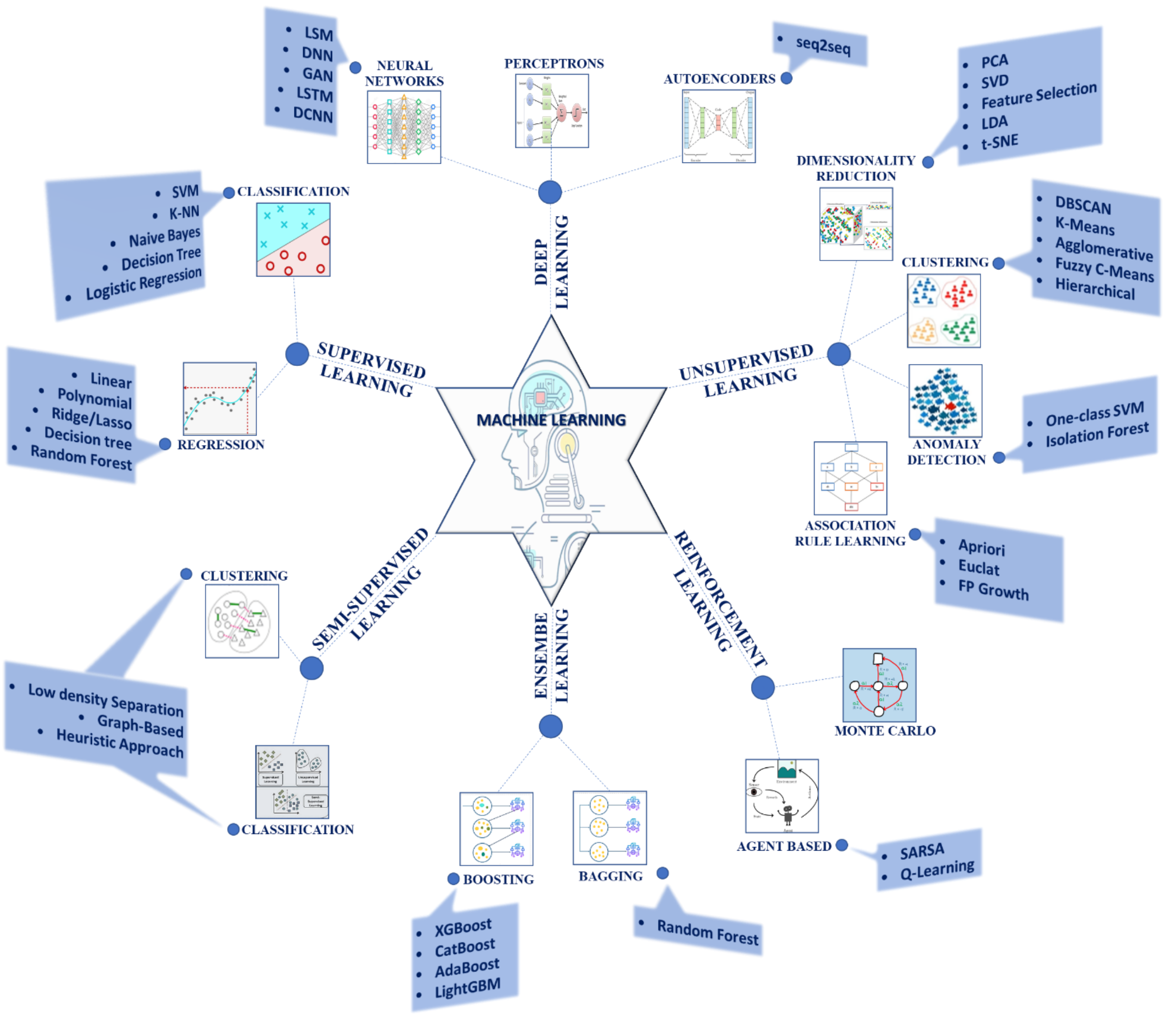


Figure 2.19 - Overview of ML techniques and their core algorithms

The next subsections will conceptually discuss the approaches and algorithms that are most commonly used in manufacturing sector challenges.

2.2.3.1. Classification

Classification modelling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (Y). The output variables are frequently referred to as labels or categories. For a given observation, the mapping function predicts the class or category. For example, a text email can be categorized into one of two categories: "spam" or "not spam". A classification assignment involves categorizing instances into one of two or more groups. It may accept input variables that are either real or discrete. A multi-class classification issue is one that has more than two classes. A multi-label classification issue is one in which one example is allocated to many classes. Classification models frequently forecast a continuous value as the likelihood of a given example belonging to each output class. The probabilities represent the likelihood or confidence that a particular example belongs to each class. By picking the class label with the highest likelihood, a projected probability may be transformed into a class value. A classification process is determined by:

• Input: a training dataset including objects with characteristics, one of which is the class label;

• Output: a model (classifier) that assigns a specific label to each item (classifies the object in one category) depending on the other attributes.

There are several methods for estimating the quality of a classification prediction model, the most frequent of which is to compute classification accuracy. The classification accuracy is the proportion of successfully categorized instances among all predictions. In paragraph 3.2 of this dissertation there is a more extensive discussion.

The Table 2.4 covers some of the issues and approaches for selecting the most suitable classifier.

Table 2.4 - Overview for selecting the most suitable classifier

	Issue	**Algorithm**
1	In addition to class labels, the classification output should provide class probabilities	Logistic Regression, Decision Tree, K-NN
2	Analysts seek to understand how the factors impact the model	Logistic Regression, Decision Tree
3	The issue is multidimensional	Naive Bayes
4	The variables in the data are of several sorts	Logistic Regression, Decision Tree
5	Nonlinear data or discontinuities in the input variables would impact the outcome	Decision Tree, SVM, K-NN
6	The information includes categorical variables with a significant number of levels	Decision Tree, Naive Bayes, SVM
7	Some of the input variables may be connected to one another	Logistic Regression, Decision Tree
8	Some of the input variables may be irrelevant	Decision Tree, Naive Bayes

Support Vector Machines (SVM)

The goal of support vector machines is to find the line that maximizes the minimum distance to the line of the instances (Figure 2.20).

Optimal margin classifier, Hinge loss and Kernel There are three types of conditions that the model must respect.

The optimal margin classifier (h) is such that:

$$h(x) = \mathrm{sign}(\omega^T x - b) \qquad \text{Eq. 2.19}$$

where $(\omega, b) \in \mathbb{R}^n \times \mathbb{R}$ is the solution of the following optimization problem: $\min \frac{1}{2}||\omega||^2$ such that

$$y^{(i)}(\omega^T x^{(i)} - b) \geq 1 \qquad \text{Eq. 2.20}$$

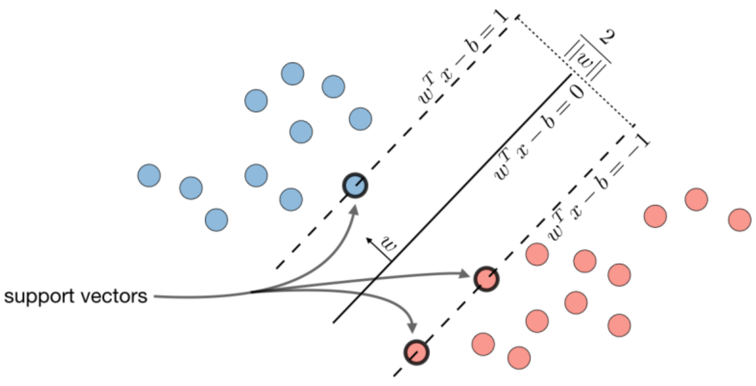


Figure 2.20 - SVM

The line is defined as $\omega^T x - b = 0$ Eq. 2.21

The hinge loss (L) is used in the setting of SVMs and is defined as follows:

$$L(z, y) = [1 - yz]_+ = \max(0,1 - yz) \quad \text{Eq. 2.22}$$

While, given a feature mapping ϕ, kernel (K) is defined as:

$$K(x, z) = \phi(x)^T \cdot \phi(z) \quad \text{Eq. 2.23}$$

In practice, the kernel K defined by $K(x, z) = \exp(-\frac{||x-z||^2}{2\sigma^2})$ is called the Gaussian kernel and is commonly used to compute the cost function using the kernel since we don't need to know the exact feature mapping ϕ, which is frequently highly difficult. Rather, just the values K(x,z) are required.

Decision Trees

A decision tree classification approach uses a training dataset to stratify or split the predictor space smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. Each of these areas only comprises a subset of the training dataset. The final result is a tree (like the one shown in Figure 2.21) with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

To anticipate the outcome of a certain (test) observation, first determine which of these zones it belongs to. Once discovered, its outcome class is predicted to be the same as the mode of all training observations included in that region. The ideas employed to stratify the predictor space may be described visually in a tree-like flowchart, thus the algorithm's name. The primary distinction is that these decision trees are drawn in the other way. (Quinlan, 1986) introduced the ID3 technique for generating decision trees performs a top-down, greedy search across the space of feasible branches with no backtracking. ID3 builds a decision tree using Entropy and Information Gain.

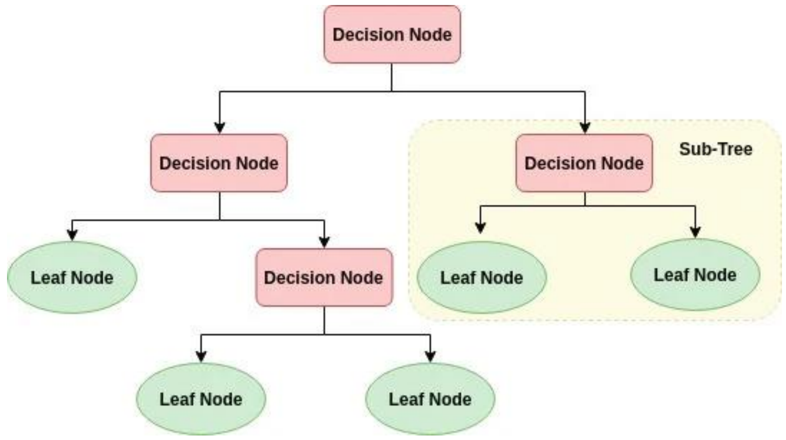


Figure 2.21 - Decision Tree

ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one. To build a decision tree, two types of entropy must be calculated as follows:
a) Entropy using the frequency table of one attribute:

$$E(S) = \sum_{i=1}^{c} p_i \log_2 p_i \qquad \text{Eq. 2.24}$$

b) Entropy using the frequency table of two attributes:

$$E(T, X) = \sum_{c \in X} P(c)E(c) \qquad \text{Eq. 2.25}$$

The information gain, on the other hand, is based on the decrease in entropy when a dataset is divided on an attribute. Creating a decision tree is all about determining which attributes provide the most information gain (i.e., the most homogeneous branches).
The procedure involves: (1) Calculating entropy of the target. (2) The dataset is then segmented based on the distinct properties. Each branch's entropy is computed. The total entropy for the split is then added proportionately. Before the split, the resultant entropy is deducted from the entropy. The consequence is an increase in information or a decrease in entropy. (3) Using the property with the highest information gain as the decision node, divide the dataset into branches and repeat the method on each branch. (4) If a branch's entropy is equal to zero, it is a leaf node; otherwise, the procedure continues with additional splitting. (5) The ID3 algorithm is run recursively on the non-leaf branches, until all data is classified.
Without the requirement for dummy variables, decision tree classification algorithms can effectively accommodate qualitative predictors. Missing values are also not an issue. Surprisingly, decision tree techniques are also employed in regression models. The same library that you would use to develop a classification model may also be used to generate a regression model when some parameters are changed.
Although decision tree-based categorization models are simple to understand, they lack robustness. The huge variance of decision trees is a key issue. A little modification in the training dataset can result in a completely different decision tree model. Another drawback is that they have lesser prediction accuracy than other classification models, such as Random Forest models (for which decision trees are the building blocks).

Naïve Bayes

Naive Bayes is one of the most common machine learning algorithms that is often used for classifying text into categories. Naive Bayes is a probabilistic classification algorithm as it uses probability to make predictions for the purpose of classification. Naive Bayes is one of the easiest classification algorithms. The Bayes Theorem underpins the Naive Bayes Classifier (Bishop, 2016). According to the Bayes Theorem, the conditional probability of a result may be estimated using the conditional probability of the outcome's cause.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \qquad \text{Eq. 2.26}$$

Where
P(A) is the prior probability of A, i.e., the likelihood of the occurrence without taking the B into account. The event is also known as the marginal probability of A.
P(B) is the prior probability of B, i.e., the likelihood of event B without taking event A into account. It is also known as the marginal probability of B.
P(A|B) is the event's conditional probability given the information about the B event. It is also known as the a posteriori probability of the occurrence since it is affected by the value of B.
P(B|A) represents the conditioned probability of event B given the knowledge about event A. It is also known as the a posteriori likelihood of event B since it is affected by the value of A. The Naive Bayes classifier selects the class with the best posterior probability based on the input variable.
Because it makes an assumption about the distribution of the data, the method is referred to be naïve. Gaussian, Bernoulli, or Multinomial distributions are all possible. Another disadvantage of Naive Bayes is that

continuous features must be pre-processed and discretized by binning, which might result in the loss of important information.

2.2.3.2. Regression

Regression predictive modelling is the task of approximating a mapping function (f) from input variables (X) to a continuous output variable (y).
A real-value, such as an integer or a floating-point value, is a continuous output variable. These are frequently numbers, such as amounts and sizes. A regression problem necessitates the estimation of a quantity. Real-valued or discrete input variables can be used in a regression. A multivariate regression issue is one that has several input variables. A time series forecasting issue is a regression problem in which the input variables are sorted by time. Because a regression predictive model predicts a quantity, the model's skill must be expressed as a prediction error.
There are several methods for estimating the quality of a regression prediction model, but one of the most frequent is to compute the root mean squared error, abbreviated as RMSE (see paragraph 3.2.3 for a better explanation).
The difficulties associated with classification predictive modelling differ from those associated with regression predictive modelling. Regression is the process of predicting a continuous quantity. For example, there is some overlap between classification and regression techniques: A continuous value can be predicted by a classification method, however the continuous value is a probability for a class label. A regression approach can forecast a discrete value, but only as an integer number. Some approaches, such as decision trees and artificial neural networks, may be utilized for classification and regression with little modifications. Some approaches, such as linear regression for regression predictive modelling and logistic regression for classification predictive modelling, cannot or should not be used for both types of issues.
In some circumstances, a regression problem can be converted to a classification problem (Salman and Kecman, 2012). The quantity to be anticipated, for example, might be translated into discrete buckets. This is known as discretization, and the resulting output variable is a classification with an ordered connection between the labels (called ordinal).
Linear Regression and Logistic Regression are the most common algorithm for regression.

Linear Regression
Linear regression is a method of modelling the connection between a continuous dependent variable y and one or more predictor variables X. One fundamental assumption is that the connection between an input variable and an outcome variable is linear. Although this assumption may appear to be restrictive, it is frequently feasible to alter the input or result variables in order to obtain a linear connection between the adjusted input and outcome variables. The relationship between y and X may be represented linearly as follows (MathWorks, 2020a):
Given the training examples $\{x_i, y_i\}_{i=1}^{N}$, the parameter vector β can be learnt.

$$y = \beta_0 + \sum_{i=1}^{N} \beta_i^T X_i + \varepsilon \qquad \text{Eq. 2.27}$$

where:
y is the outcome variable
x_i are the input variables, for i = 1, 2,…,i-1
β_i is the change in y based on a unit change in x_i for i = 1, 2,…,i-1
ε is a random error term representing the difference between the linear model and a specific observed value y.
The Ordinary Least Squares (OLS) method is a popular methodology for estimating parameters. The Regression line is a straight line that best fits the data, with the total distance from the line to the dots (variable values) depicted on a graph being the shortest. The Figure 2.22 describes all of the definitions for linear regression.

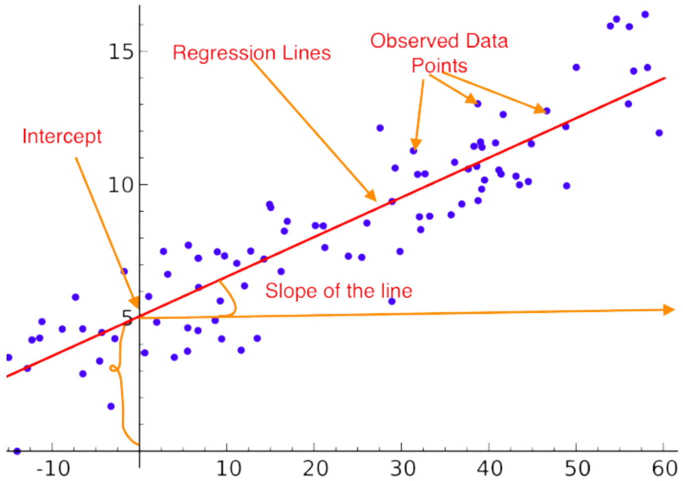


Figure 2.22 - Linear Regression

The best-fitting regression line has the equation
y = a + bx,
where:
a is the y-intercept.
b is the slope of the line
x is an explanatory variable.
y is a dependent variable

Logistic Regression

If the dependent variable is not continuous but categorical, linear regression can be transformed to logistic regression using a logit link function (Subasi, 2020).
Logistic regression is based on the logistic function f(y), as given in the following equation.
In the meantime, in Figure 2.23 is depicted how logistic regression works.

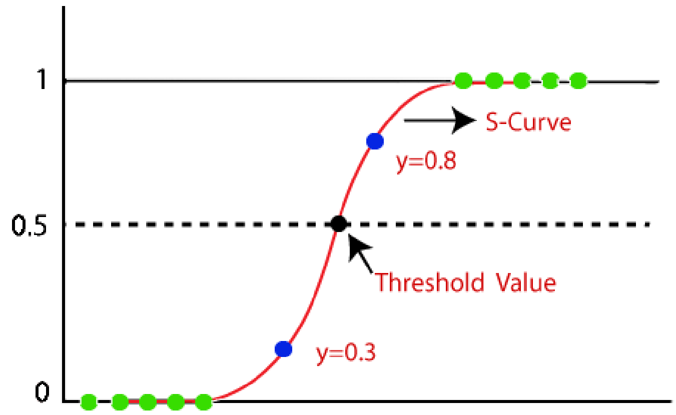


Figure 2.23 - Logistic Regression

$$f(y) = \frac{e^y}{1 + e^y} \qquad \text{for} - \infty < y < +\infty \qquad \text{Eq. 2.28}$$

That means when y → +∞, f(y) = 1 and when y → -∞, f(y) = 0. That's why the range is between 0 and 1. Because f(y) has a range of (0, 1), the logistic function looks to be an acceptable function to predict the likelihood of a specific result occurring. As the value of y grows, so does the likelihood of the result occurrence. To forecast the likelihood of a result in any given model, y must be a function of the input variables. y is stated as a linear function of the input variables in logistic regression.

When the input variables are continuous or discrete, including categorical data types, but the result variable is continuous, linear regression is appropriate. Logistic regression is a preferable alternative if the outcome variable is categorical. Both approaches require that the input variables are linear additive functions. If this assumption is not met, both regression approaches perform badly. Furthermore, the assumption of normally distributed error terms with a constant variance is significant in linear regression for many of the statistical conclusions that might be explored. If the different assumptions do not appear to hold, the data must be transformed appropriately. Although a set of input factors may be a reasonable predictor of the end variable, the analyst should not conclude that the input variables are directly responsible for the outcome.

Multicollinearity occurs when multiple of the input variables are substantially connected with one another. Multicollinearity frequently results in coefficient estimates that are relatively big in absolute magnitude and may be pointing in the wrong direction (negative or positive sign). When feasible, eliminate the bulk of these correlated variables from the model or replace them with a new variable that is a function of the associated variables. In the situation of multicollinearity, it may be necessary to limit the magnitudes of the calculated coefficients.

Ridge regression, which penalizes the magnitude of the coefficients, is one strategy that may be used. The goal of fitting a linear regression model is to determine the coefficient values that minimize the sum of the residuals squared. A penalty term proportionate to the sum of the squares of the coefficients is added to the sum of the residuals squared in ridge regression.

A comparable modelling approach is lasso regression, in which the penalty is proportional to the total of the absolute values of the coefficients. In the usage of logistic regression, only binary outcome variables were investigated. Multinomial logistic regression can be used if the result variable has more than two states.

2.2.3.3. Clustering

Clustering techniques are unsupervised in the sense that the labels to be applied to the clusters are not determined in advance by the data scientist. The data structure specifies the things of interest and dictates how the objects should be grouped. Clustering is a technique that is frequently used in exploratory data analysis. There are no predictions made during clustering. Clustering algorithms, on the other hand, discover similarities between items based on their qualities and arrange the comparable objects into clusters. Clustering methods are used in marketing, finance, and a variety of scientific fields. A popular clustering method is k-means.

K-means

Given a collection of items, each having n quantifiable properties, k-means (Tan et al., 2019) is an analytical approach that discovers k clusters of objects based on their closeness to the centre of the k groups for a given value of k. The arithmetic average (mean) of each cluster's n-dimensional vector of characteristics is used to calculate the centre. This section discusses the algorithm for calculating the k means as well as how to apply this technique to various use situations. After identifying the clusters, labels may be applied to each cluster to categorise each group depending on its features. To explain the procedure, consider each item that corresponds to the location (x, y), where x and y signify the two qualities and i = 1, 2,... M. A centroid is the location that corresponds to the mean of a particular cluster of m points (m ~ M). A centroid is a location in mathematics that corresponds to an object's centre of mass. The k-means approach for finding k clusters may be broken down into six parts. (Figure 2.24). (a) The data points to be clustered (solid blue circles) in a bidimensional feature space. (b) For random cluster center placements (aqua, green, and red hollow circles), each data point can be assigned to the nearest center. (c) Three decision boundaries split the bidimensional space into three parts (black dashed lines). (d) Each center advances to the centroid of the data points allocated to it at the time (movements shown by the black arrows). (e) The data points' revised cluster assignments are derived based on the new center positions. Steps (c) and (d) are repeated until convergence is reached. (f) finally, the cluster allocations.

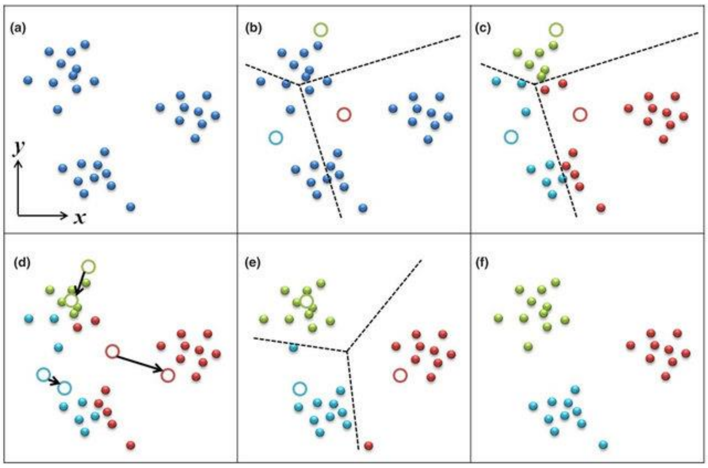


Figure 2.24 - k-means steps with a bidimensional example (Chen and Lai, 2018)

To extend to n dimensions, consider M objects, each of which is characterized by n characteristics or property values (p_1, p_2,…p_n). Then, for i = 1,2,...,M, object i is described by (p_{i1}, p_{i2},…p_{in}). In other words, there is a matrix with M rows for the M objects and n columns for the attribute values. To extend the previous procedure of locating the k clusters from two dimensions to one dimension, the following equations offer formulae for determining the distances and positions of the centroids for $n \geq 1$. For a given point, p_i,at (p_{i1}, p_{i2},…p_{in}) and a centroid, q, located at (q_1, q_2,…q_n). Equation expresses the Euclidean distance, d, between p_i and q.

$$d(p_i, q) = \sqrt{\sum_{j=1}^{n}\left(p_{ij} - q_j\right)^2} \qquad \text{Eq. 2.29}$$

Equation shows how to determine the centroid, q, of a cluster of m points (p_{i1}, p_{i2},…p_{in}).

$$(q_1, q_2, \dots q_n) = \left(\frac{\sum_{i=1}^{m} p_{i1}}{m}, \frac{\sum_{i=1}^{m} p_{i2}}{m}, \dots, \frac{\sum_{i=1}^{m} p_{in}}{m}\right) \qquad \text{Eq. 2.30}$$

The value of k can be determined by a reasonable guess or a predetermined criterion. The structure of the data would be explained by k + 1 clusters. Following that, a heuristic based on the Within Sum of Squares (WSS) metric is evaluated to get a relatively optimal value of k. WSS is defined as shown in Equation using the distance function given in Equation.

$$WSS = \sum_{i=1}^{M} d\left(p_i, q^{(i)}\right)^2 = \sum_{i=1}^{M}\sum_{j=1}^{n}\left(p_{ij}, q_j^{(i)}\right)^2 \qquad \text{Eq. 2.31}$$

WSS is the sum of the squares of the distances between each data point and the nearest centroid. The term $q^{(i)}$ refers to the centroid that is closest to the ith point. The WSS is modest if the points are substantially near to their respective centroids. As a result, if k + 1 clusters do not significantly lower the value of WSS compared to the situation with just k clusters, adding another cluster may be of little advantage.

The k-means method is sensitive to the first centroid's starting position. As a result, it is critical to conduct the k-means analysis numerous times for a given value of k to guarantee that the cluster results represent the overall minimum. WSS

Other functions to consider are the cosine similarity and Manhattan distance functions. The cosine similarity function is frequently used to compare two papers based on the frequency of each word in both publications. Equation expresses the Manhattan distance, d_1, between p and q.

$$d_1(p, q) = \sum_{j=1}^{n}\left|p_j - q_j\right| \qquad \text{Eq. 2.32}$$

If the Manhattan distance is required for a clustering study, the median is a better choice for the centroid than the mean.

Hierarchical Clustering

Hierarchical Clustering is an alternative approach to k-means clustering that uses the approach of finding groups in the data such that the instances are more similar to each other than to instances in other groups. This measure of similarity is generally a Euclidean distance between the data points, but Citi-block and Geodesic distances can also be used.

The data is broken down into clusters in a hierarchical fashion. The number of clusters is 0 at the top and maximum at the bottom. The optimum number of clusters is selected from this hierarchy. Hierarchical partitions can be visualized using a tree structure (a dendrogram). It does not need the number of clusters as an input and the partitions can be viewed at different levels of granularities (i.e., can refine/coarsen clusters) using different K.

To measure the dissimilarity between two or more clusters of observation, a number of different cluster agglomeration methods (i.e, linkage methods) have been developed to answer to this question (Scikit-learn, 2020). The most common types methods are explained and then depicted with 2-clusters example in Figure 2.25:

- Maximum or full linkage clustering: It computes all pairwise dissimilarities between elements in cluster i and elements in cluster j, and uses the biggest value (i.e., maximum value) of these dissimilarities to calculate the distance between the two clusters. It produces more compact clusters.
- Minimum or single linkage clustering: This method computes all pairwise dissimilarities between elements in cluster i and elements in cluster j, and uses the least of these dissimilarities as a linkage criteria. It produces lengthy, "loose" clusters.
- Mean or average linkage clustering: This method computes all pairwise dissimilarities between elements in cluster i and elements in cluster j, then uses the average of these dissimilarities to calculate the distance between the two clusters.
- Centroid linkage clustering: This method computes the dissimilarity between the centroid of cluster i (a mean vector of length p variables) and the centroid of cluster j.
- Ward's approach of minimizing variance: It reduces the overall within-cluster variance. Each phase, the pair of clusters with the shortest between-cluster distance is merged.

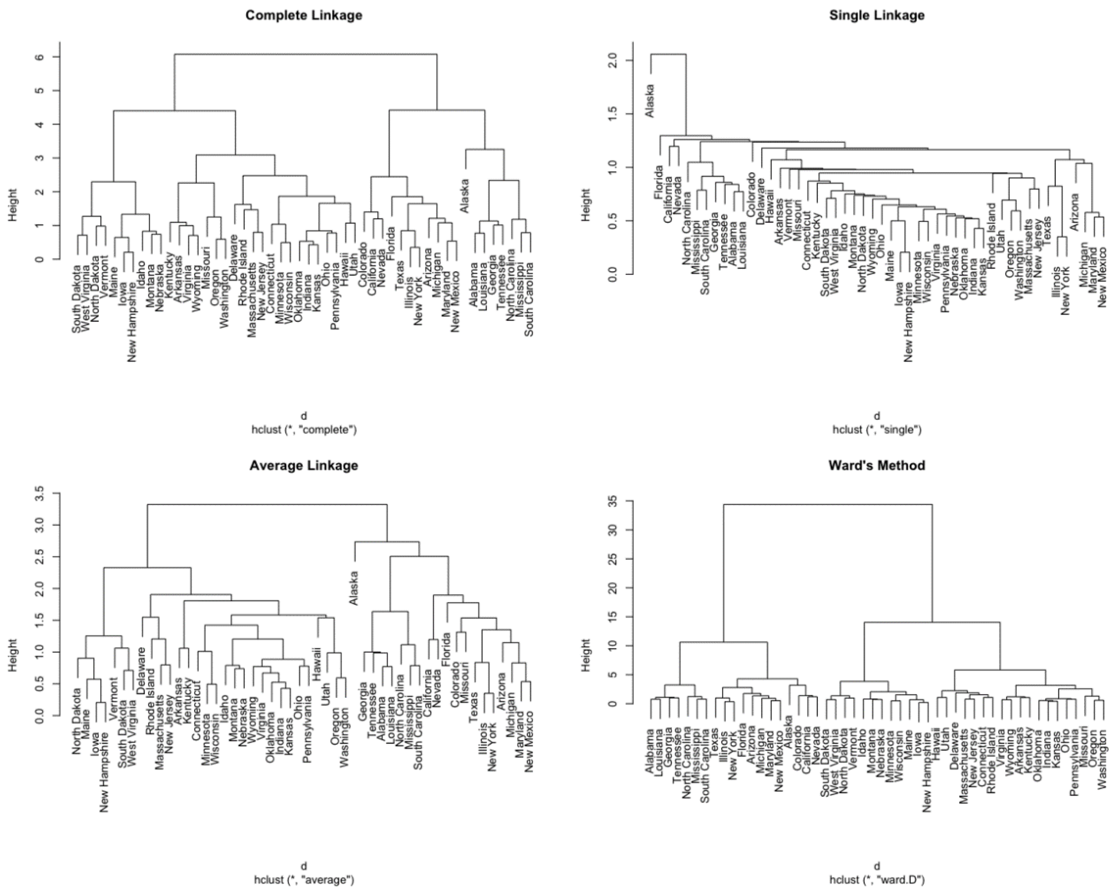


Figure 2.25 - Hierarchical Clustering - Types of linkage

DBSCAN
Partitioning methods and hierarchical clustering work for finding spherical-shaped clusters or convex clusters. In other words, they are suitable only for compact and well-separated clusters. Moreover, they are also severely affected by the presence of noise and outliers in the data. When the number of clusters, k, is not specified, DBSCAN (density-based spatial clustering of applications with noise) (Ester et al., 1996) can be used to link samples using density diffusion. A dense zone is formed by points that are x distance apart and create a collection of core points. A cluster is formed by points that are x distance apart, both core and non-core. Points that cannot be reached from any of the core points are referred to as noise points. It is a clustering technique based on density that detects dense regions in data as clusters. Dense regions are described as places where points may be reached from one another. The method employs two variables: ε and minpts. ε defines the neighbourhood around a data point while minpts is the minimum number of neighbours (data points) within ε radius. Larger the dataset, the larger value of minpts must be chosen. Figure 2.26 shows an example of the DBSCAN functioning.

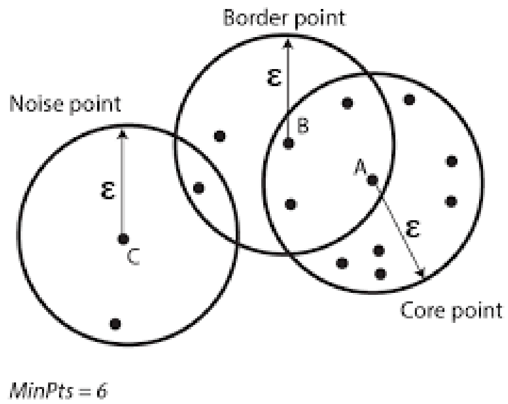


Figure 2.26 - DBSCAN

As a general rule, the minimum minpts can be derived from the number of dimensions D in the dataset as, minpts >= D+1. The minimum value of minpts must be chosen at least 3.If the distance between two data points is less than ε, they are within reach of each other. A cluster must also have a certain number of points in order to be termed a cluster. Core points are those that have the fewest number of points within ε distance. Noise points are spots that cannot be reached by any cluster. The density-based nature of DBSCAN makes it resistant to outliers. It does not, however, perform well when dealing with clusters of varied density.

2.2.3.4. Anomaly Detection

The technique of discovering outer points or observations that diverge significantly from the rest data is known as anomaly detection, sometimes known as outlier identification. Anomaly detection can range from simple outlier detection to complex machine learning algorithms trained to uncover hidden patterns across hundreds of signals. Usually, these outer points have a fascinating history to tell, and by analysing them, one may grasp the system's severe functioning conditions. It is a useful technique to deal with imbalanced data sets in which anomalies are often difficult to detect visually from raw data The dataset is trained on a single class (normal or majority class), and the method draws a line along this class. During the training, the other class is entirely ignored. Anything discovered outside of this decision threshold is referred to as a novelty or an outlier (Lee and Cho, 2006). To minimize too many false positives during the anomaly discovery process, proper anomaly detection should be able to discern signal from noise.
Several approaches to designing anomaly detection algorithms require little or no anomalous data (MathWorks, 2020b):

- Thresholding. Thresholding detects anomalies when data crosses a certain threshold on a statistical indicator. Examples include calculating the standard deviation over recent windows in time series data, applying a control chart to a signal, detecting abrupt changes in a signal using change point detection, and obtaining robust estimates of the data distribution and identifying anomalies as samples on the distribution's fringes. Thresholding on statistical measures is an excellent place to start, but it is more difficult to apply to multivariate data and is less robust than machine learning techniques to anomaly identification. Statistical estimates that are resistant to outliers, such as robust covariance, will produce superior results.

- SVMs with a single class. Support vector machines with a single class find separation hyperplanes that minimize the distance between classes. Training only one class produces a model of data that may be deemed normal, allowing you to identify anomalies in the absence of labelled abnormalities. This method, like others based on distance, requires numeric characteristics as input and will not function well with high-dimensional data.
- Isolation forests grow trees that isolate each observation into a leaf, and an anomaly score is calculated as the average depth of your sample: normal samples make fewer judgments than anomalous samples. This approach works with high-dimensional data and supports a combination of numeric and category variables.
- Autoencoders. Autoencoders are neural networks that have been trained on normal data and seek to recover the original input. A normal input will be properly reconstructed by the trained autoencoder. A significant disparity between the input and its reconstruction might suggest an error. Signal and picture data may both be encoded using autoencoders.

Depending on whether the data can be labelled, anomaly detection can be addressed in either a supervised or unsupervised manner.

2.2.3.5. Random Forest

It is a tree-based approach that employs a large number of decision trees constructed from randomly chosen sets of features. It is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Sankhye and Hu, 2020). A special case of random forest uses bagging on decision trees, where samples are randomly chosen with replacement from the original training set. It is very uninterpretable, in contrast to the basic decision tree, but its overall strong performance makes it a popular algorithm. Ensemble approaches include random forests.The Figure 2.27 fully expresses the Random Forest's procedure and highlights the distinction between the classifier and the regressor: once the nth Decision Trees are computed, the class with the highest number of votes is determined for the classifier, whereas for the regressor, average votes are used to obtain the prediction.

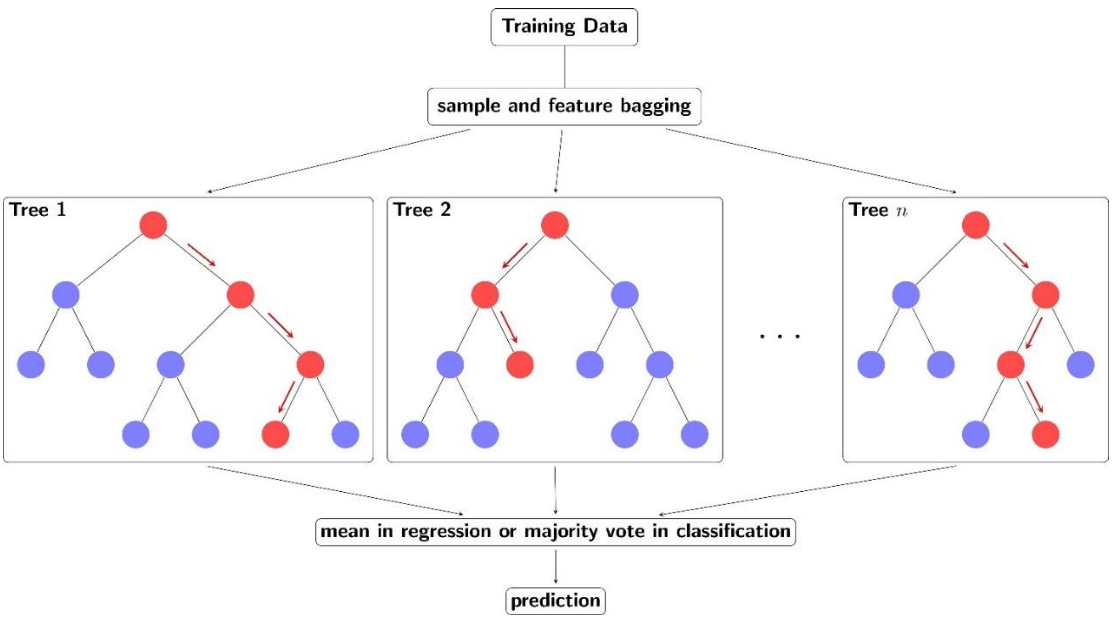


Figure 2.27 - Random Forest

2.2.3.6. Dimensionality Reduction

Feeding a high number of features straight into a machine learning algorithm is often counterproductive since some features may be useless or the "intrinsic" dimensionality may be less than the number of features. It is one of the non-supervised dataset pre-processing techniques in automated learning. It is important in Machine Learning to remove redundant (related) information from the dataset that is less or not very relevant to the issue to be addressed. It is undeniably simpler and less expensive to train an algorithm with a smaller data space. So it is a workaround for the curse of dimensionality (Chen, 2014).

Reducing data dimensionality entails not just removing some size (noise), but also integrating redundant and relevant information.

Dimension reduction can be accomplished using Features Selection and Features Extraction (with Principal Component Analysis (PCA) and Singular Value Decomposition (SVD)) (Huang et al., 2019; Velliangiri et al., 2019).

Feature selection

Feature selection is used to remove characteristics that are irrelevant or redundant from your dataset. The primary distinction between feature selection and extraction is that feature selection retains a subset of the original characteristics while feature extraction generates entirely new ones. Some supervised algorithms, such as Regularized Regression and Random Forests, already include feature selection. Feature selection can be unsupervised (e.g., Variance Thresholds) or supervised as a stand-alone activity (e.g. Genetic Algorithms). Variance thresholds exclude characteristics whose values do not vary much from observation to observation (i.e. their variance falls below a threshold). These features are of limited use. Because variance is scale dependent, normalization is required. Genetic algorithms (GA) are a large family of algorithms that may be tailored to specific needs. They are evolutionary biology and natural selection-inspired search algorithms that combine mutation and cross-over to rapidly explore huge solution spaces. GAs have two key applications in ML. The first is for optimization, such as determining the ideal neural network weights. The second is for the selection of supervised features. In this application, "genes" represent individual characteristics, whereas "organisms" represent a potential group of traits. Each organism in the "population" is assigned a fitness score, which is based on model performance on a hold-out set. The most fit creatures survive and reproduce, and this cycle continues until the population converges on a solution several generations later.

Feature extraction

Feature extraction is used to create a new, smaller collection of features that captures the majority of the important information. Again, feature selection retains just a subset of the original features, whereas feature extraction generates new ones. Some algorithms, such feature selection, already include feature extraction. Deep Learning is the finest example, since each buried neural layer recovers progressively meaningful representations of the raw input data.

PCA

PCA translates the original data space into a lower-dimensional space while retaining as much information as feasible. The PCA simply selects a subspace that best preserves the data variance, with the subspace determined by the data's covariance matrix's dominant eigenvectors.

SVD

The SVD is linked to PCA in that it produces the dominating left singular vectors that form the same subspace as PCA for the cantered data matrix (features against samples). SVD, on the other hand, is a more versatile method since it can perform things that PCA cannot. It might be the most popular technique for dimensionality reduction when data is sparse. Sparse data refers to rows of data where many of the values are zero. This is often the case in some problem domains like recommender systems where a user has a rating for very few movies or songs in the database and zero ratings for all other cases. Both techniques work on a linear mapping but identify a completely different segment. Therefore, alternative and complementary solutions with pros and cons. The PCA technique preserves the information while the LDA technique better distinguishes the two classes.

2.2.3.7. Deep Learning with Neural Networks

Neural networks spread in 1985 due to their parallel and distributed processing ability. However, research in this sector has been hampered by the ineffectiveness of the back-propagation training technique, which is extensively employed to optimize neural network parameters. In ML, SVM and other simpler models that can be readily taught by addressing convex optimization problems rapidly supplanted neural networks.

New and improved training strategies, such as unsupervised pre-training and layer-wise greedy training, have sparked renewed interest in neural networks in recent years. Increased computational capability, such as graphics processing units (GPU) and massively parallel processing (MPP), has also fuelled the resurgence of neural networks. The resurgence of neural network research has resulted in the development of models with hundreds of layers. Shallow neural networks, in other words, have developed into deep learning neural networks. Deep neural networks have had a lot of success with supervised learning. Deep learning performs as good as, if not better than, humans in voice and picture recognition. Deep learning, when applied to unsupervised learning tasks such as feature extraction, pulls features from raw pictures or voice with far less human involvement. A neural network is made up of three layers: the input layer, the hidden layers, and the output layer. The input and output layers are defined by the training samples. When the output layer is a categorical variable, the neural network is a technique for dealing with classification difficulties. The network may be used to do regression when the output layer is a continuous variable. When the output layer and input layer are the same, the network may be used to extract inherent characteristics. The model complexity and modelling capability are defined by the number of hidden layers. The activation function is a very important feature of an artificial neural network, they basically decide whether the neuron should be activated or not. A linear equation is simple to solve but is limited in its capacity to solve complex problems and have less power to learn complex functional mappings from data. A neural network without an activation function is just a linear regression model. The activation function does the non-linear transformation to the input making it capable to learn and perform more complex tasks.

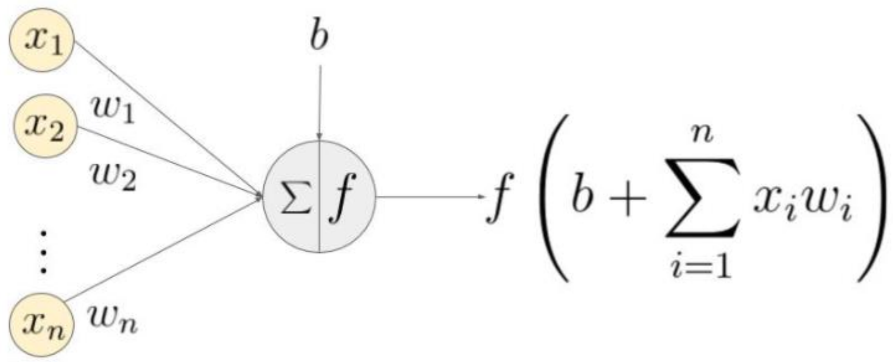


Figure 2.28 - Simple Neural Network functioning

In the above Figure 2.28, $(x_1,x_2,\ldots,x_n)$ is the input signal vector that gets multiplied with the weights $(w_1,w_2,\ldots,w_n)$. This is followed by accumulation (i.e., summation + addition of bias b). Finally, an activation function f is applied to this sum.

Some activation functions are (Gupta, 2020):

- Identity activation function (linear function) with equation $f(x) = x$ and range $(-\infty,+\infty)$. As shown in the above figure the activation is proportional to the input. . This can be applied to various neurons and multiple neurons can be activated at the same time.
- Sigmoid activation function (non-linear function) with equation $\phi(x) = \frac{1}{1+ e^{-x}}$ and range (0,1) so that the function curve looks like a S-shape. Sigmoid have major drawbacks and difficulties in application than linear activation function.
- Hyperbolic tangent activation function with equation $f(x) = \frac{1-e^{-2x}}{1+ e^{-2x}}$ and range (-1,1). Here optimization criteria is easier than Sigmoid, but it suffers vanishing gradient problem.

- Softmax activation function with equation $\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$ where z is a vector of the inputs to the output layer, j indexes the output units, so j = 1, 2, …, K. It is also a type of sigmoid function but it is very useful to handle classification problems having multiple classes and it ideally used in the output layer of the classifier where we are actually trying to attain the probabilities to define the class of each input

Deep Neural Networks (DNN) are in the frontline of data-driven approaches. DNNs have also been shown to be useful for predicting the potential state of dynamical systems (Raissi et al., 2017). The lack of interpretability of the resulting model is a major drawback of DNNs and related data-driven methods; they are based on estimation and do not have governing equations or easily interpretable models in terms of the original variable collection.

2.2.3.8. Gradient Boosting

Gradient Boosting Regression Trees is one of the most effective machine learning models for predictive analytics because they are a flexible, non-parametric learning strategy for classification and regression (Zhang et al., 2021). The strengths of two techniques, regression trees and boosting approaches, are combined in boosted regression trees. Boosted regression trees contain key benefits of tree-based approaches, such as managing diverse types of predictor variables and accepting missing data. They do not require previous data transformation or outlier removal, can fit complicated nonlinear relationships, and automatically handle predictor interaction effects. Figure 2.29 shows the learning process through gradient boosting: (1) The first distinguishing feature of gradient boosting is that it begins with a dummy estimator. Basically, it computes the average value of goal values and generates preliminary projections. Calculates the difference between the predicted and actual value using predictions. This is referred to as residues. (2) Instead of training a new data estimator to forecast the target, train an estimator to predict the first predictor's residues. This predictor is often a decision shaft with particular constraints, such as the maximum number of leaf knots permitted. If the majority of the instance residues are in the same leaf knot, the leaf node value is taken from their media and the USA. (3) To generate predictions, for each instance, add the value of the basic estimator to the predicted residual value of the instance's decision shaft to get a new forecast. Then, compute the residues between the predicted and actual values once more. (4) This procedure is continued until either a specified threshold is attained or the residual difference is extremely tiny. (5) To forecast an unseen instance, it sends it to each decision-making tree, sums their forecasts, and adds the value of the fundamental estimator.

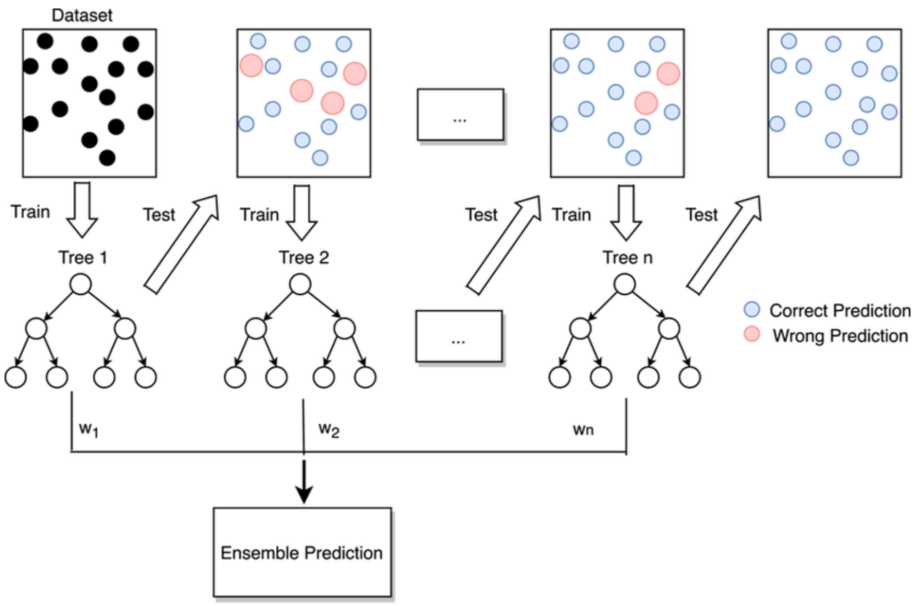


Figure 2.29 - Gradient Boosting (Zhang et al., 2021)

2.2.3.9. Other techniques

Bagging (or bootstrap aggregating) (Sutton, 2005) use the bootstrap approach, which samples a dataset with replacement from a uniform probability distribution. "With replacement" signifies that when a sample is chosen for a training or testing set, it is maintained in the dataset and can be chosen again. Because of the replacement sampling, certain samples may appear several times in a training or testing set, while others may be absent. On each bootstrap sample, a model or base classifier is trained individually, and a test sample is allocated to the class with the most votes.

Boosting (or Ada Boost) (Sutton, 2005), like bagging, utilizes classification votes to aggregate the output of individual models. Furthermore, it merges models of the same type. Boosting, on the other hand, is an iterative technique in which the performance of prior models influences the performance of the new model. Furthermore, boosting gives each training sample a weight that indicates its value, and the weight can alter adaptively at the conclusion of each boosting session. Bagging and boosting have been demonstrated to outperform decision trees.

Using association rules, patterns can be discovered from the data that allow the association rule algorithms to disclose rules of related items. The research of association rules started as early as the 1960s. Early research by (Hájek et al., 1966)introduced many of the key concepts and approaches of association rule learning, but it focused on the mathematical representation rather than the algorithm. The framework of association rule learning was brought into the database community by (Agrawal et al., 1993). Apriori is the main focus of the discussion of association rules. Apriori is one of the earliest and the most fundamental algorithms for generating association rules. It pioneered the use of support for pruning the itemset and controlling the exponential growth of candidate itemset. Shorter candidate item sets, which are known to be frequent item sets, are combined and pruned to generate longer frequent itemset. This approach eliminates the need for all possible item sets to be enumerated within the algorithm, since the number of all possible itemset can become exponentially large. One major component of Apriori is support.

2.2.3.10. ML techniques comparison

The literature abounds with reviews or case studies that compare various ML techniques, both qualitatively and quantitatively, also in production processes (Almanei et al., 2021; Amornsamankul et al., 2019; Baumann

et al., 2019; Fahle et al., 2020; Karmaker Santu et al., 2020; Komputer et al., 2019; Rosalina, 2019; Stief et al., 2019; Wuest et al., 2016).
Taking into account the primary techniques outlined above, Table 2.5 summarizes some of the most essential characteristics that a practitioner must consider when selecting the most suitable algorithm.

Table 2.5 - Ranking of the main ML algorithm according to the most important characteristics

Algorithm	Training speed	Prediction or Classification speed	Quantity of data handling	Tolerance to noise/outliers	Tolerance to missing, unrelated, redundant attributes	Ease of interpretation	Ease of use	Power
Linear Regression	9	9	9	5	8	9	9	5
SVM	7	7	5	6	9	6	6	8
K-NN	10	2	6	5	8	9	8	5
K-Means	7	7	7	4	8	5	6	6
Naive Bayes	9	8	9	8	2	8	7	4
Decision Trees	8	9	8	4	4	8	7	8
Gradient Boosting	6	6	7	6	6	6	6	9
Graph-based	4	4	3	8	6	9	3	7
Random Forest	3	2	9	7	4	4	6	9
Neural Networks	3	7	4	7	3	3	4	10

The comparison was performed using the literature and a summary of the primary benefits and drawbacks of each of the following techniques. The chart illustrates that Linear Regression gets the highest score, whereas Gradient Boosting and Neural Networks are powerful tools but are difficult to apply, tune, and interpret.

2.3. ADVANTAGES AND RESEARCH GAPS

2.3.1. Model-Based Approaches

The models based on thermodynamic analyses, thus based on discrete variables, represent an innovative and interesting strategy for maximizing the sustainability of manufacturing system performances while facilitating the management of new smart manufacturing processes, thus driving practitioners to employ a suitable sensing system and information structure for real-time monitoring, thus combining model-based approaches with data-driven ones and gathering a comprehensive picture.
LCA, described in paragraph 2.1.1, is regarded as a leading eco-design technique since it allows for in-depth analysis of each component of a product or service, allowing for an exploration of the nature of the whole life cycle. It aids in the identification of the most influencing systems and stages, as well as providing a clear image of the issues that must be remedied by the action goals. It may be used to improve existing products or to guide decision-making in the development of new ones (Alberto Navajas et al., 2017). Measurement of consumption and effects, which enables ongoing development and improvement of goods and processes not only from a technical but also from an environmental standpoint, is an expression of responsibility for all stakeholders. Because it sums up quantities, this approach is based on linear equations. LCA has certain limitations as well. The first is that it is more oriented toward the quantification of resources depleted during the process, but does not provide information on efficiency and potential margins for improvement; the second is that it relies on various datasets relating to general or generic results, regardless of the specific process assessed. If data collection is inadequate or there aren't enough data accessible, the study won't be able to draw meaningful results. The third point is that LCA evaluations are focused on assumptions and scenarios since they employ a simplified model to represent the local environment.
The inventory phase is, undoubtedly, the most time-consuming and resource-intensive, as it may include both upstream and downstream activities (resource collection, processing, and transport) (product consumption and disposal). Upstream and downstream data might be available in opensource or payment databases, such as Ecoinvent (payment, but the most extensive), ELCD (payment), USLCI (opensource), and so on, to enable data gathering and full implementation of the LCA. SimaPro, Gabi, and OpenLCA software assist users in

doing the evaluation in a more direct and straightforward manner (Dincer and Bicer, 2018). They obviously require databases to function and are thus inaccessible to any practitioner.

With paragraph 2.1.2, the benefits of employing exergy for analysing efficiency, environmental effect, and sustainability have been illustrated. Exergy ideas are thought to play an important role in assessing and growing the usage of sustainable energy and technology. Although decisions about the design and modification of energy systems are typically concerned with not only efficiency but also economics, environmental impact, safety, and other issues, exergy should be useful to engineers and scientists, as well as decision and policy makers, in design and improvement activities.

An improved knowledge of energy-related environmental concerns provides a substantial challenge, both in terms of allowing problems to be handled and ensuring that solutions are beneficial to society and energy policymaking. The potential use of exergy analysis in addressing and resolving energy-related difficulties is enormous, and exergy can play a role in energy-related decision and policy making.

The paragraph 2.1.3 provided a method for analysing the benefits and drawbacks of both EA and LCA, as well as their differences and similarities, as well as an analysis of the probable manner of interaction between the two. It should be highlighted that none of the methodologies investigated are comprehensive in every area of any industrial scenario. Both procedures cannot be substituted for one another; rather, they are complimentary and should be carried out concurrently, or better, in an integrated manner. The most consistent approach is to conduct both analyses in a systematic manner, including the same initial assumptions, objectives, phases, processes, streams, boundaries, and so on, and then, if necessary, combine the results of both analyses to make the fewest assumptions and approximations and lose as little information as possible. The hybrid EA-LCA methods were developed to achieve a holistic view of the system/process to be analysed; however, the lack of appropriate indicators and a well-established set of calculations, as well as the lack of complete and up-to-date data to overcome uncertainty analysis, are frequently problematic, resulting in a poor scientific consistency in the evaluation. However, discovered ambiguities and weaknesses can serve as a solid foundation for further refining current procedures, making it simpler to select the best suited approach based on the practitioner's needs. These types of data are not immediately deductible from a multi-criteria analysis or an individual performance/sustainability indicator.

To be fully aware of the sustainability assessment, or rather, the environmental performance of the process/system, it is necessary to strengthen the physical and mathematical concepts that tie the EA to the LCA. The issue of EA's complete integration with LCA remains unresolved, and a solution appears to be a long way off since a mathematical convolution of linear and non-linear laws (that does not need significant assumptions and approximations) has not yet been resolved.

Consider the first law of thermodynamics and the generic system depicted in Figure 2.30:

$$dU = \delta Q - \delta W \qquad \text{Eq. 2.33}$$

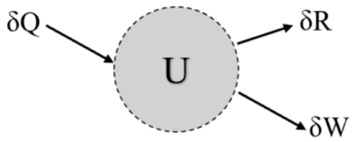


Figure 2.30 - System status and its internal energy U

where d is an exact differential associated with state functions (the state functions do not rely on the route of the transformation), and is a non-exact differential associated with values that are not state functions but are dependent on the path of the transformation (Smith et al., 2001). U is the system's internal energy, Q is the heat contributed to the system, W is the work done by the system, and R is the heat wasted owing to irreversible processes.

For a reversible process:

$$\delta Q = T \cdot dS \qquad \text{Eq. 2.34}$$

and

$$\delta W = P \cdot dV \qquad \text{Eq. 2.35}$$

so

$$dU = T \cdot dS - P \cdot dV \qquad \text{Eq. 2.36}$$

According to the second law of thermodynamics, the entropy of every isolated system always rises. Isolated systems grow spontaneously toward thermal equilibrium. This is referred to as the system's highest entropy state.

$$dQ \leq T \cdot dS \qquad \text{Eq. 2.37}$$

The first law of thermodynamics ($dU = dQ + P \cdot dV$) and the second law of thermodynamics ($dQ \leq T \cdot dS$) (Feynman et al., 2011) may now be merged into a single mathematical statement known as the combined law of thermodynamics (Eq. 2.38)

$$dU - T \cdot dS + P \cdot dV \leq 0 \qquad \text{Eq. 2.38}$$

It is worth noting that these equations were developed for a reversible adiabatic process, but they only depend on the final state, therefore the change in characteristics during a given change of state is the same for an irreversible process as for a reversible process. As a result, they can also be utilized for irreversible processes.

$$dU - T \cdot dS + P \cdot dV + \delta R \leq 0 \qquad \text{Eq. 2.39}$$

For increasingly complex systems, adding extra material to the system itself can also affect the status. When generalized forces act or species migrate beyond the system border, the generalized combined law of thermodynamics assumes the form (Gladyshev, 2015). The combined law of thermodynamics for complex systems is the Eq. 2.40

$$T \cdot dS \ \geq dU + P \cdot dV + \ \textstyle\sum_k X_k \cdot dx_k + \ \sum_k \mu_k \cdot dm_k \qquad \text{Eq. 2.40}$$

Here, T denotes temperature, S the entropy, U the internal energy, P pressure, V volume, X_k any other generalized type of work except pressure, x_k any generalized coordinate except volume, μ_k chemical potential, m_k the mass of the k-th substance, which can be replaced by the number of moles.
Starting from Clausius-Planck inequality for the definition of entropy (Eq. 2.41)

$$\Delta S \ \geq \int_A^B \frac{dQ}{T} \qquad \text{Eq. 2.41}$$

that in differential form it becomes

$$\rho s + \nabla \frac{q}{T} - \rho \frac{r}{T} \ \geq 0 \qquad \text{Eq. 2.42}$$

Where q denotes the internal energy transfer for conduction and r denotes the particular rate of energy supply or energy loss by radiation. The Helmholtz free energy ($a = u – Ts$) and the first law of thermodynamics for continuous systems ($\rho u = – \nabla q – P : \ \nabla v + \rho r$).
The differential form of the inequality becomes

$$D = -\rho(a + sT) - P : V - \frac{q\nabla T}{T} \ \geq 0 \qquad \text{Eq. 2.43}$$

The Clausius-Duhem inequality (Demirel, and Sieniutycz, 2003), depicted in Eq. 2.43, gives the general form of the combined law of thermodynamics articulated for a continuous system, where D is the dissipation and the three components reflect energy, mechanical, and thermic dissipation, respectively. The continuous form can only be resolved in a punctual, distinct manner during a fixed time span.

This inequality expresses the irreversibility of natural processes, particularly when energy dissipation is involved. The dissipation inequality is only expressed for the entire system or for each control volume (sub-components). Clausius inequality is a proposition that applies to closed systems. In his study (Bhalekar, 1996) established that it is not possible to apply the Clausius inequality equivalent for open systems, but there is an operational counterpart of the Clausius inequality at the local level that is clearly sufficient to build an irreversible thermodynamic analysis. This implies that the inequality does not apply to all non-equilibrium settings, a claim made by Meixner in 1975 (Meixner, 1973).
In an ideal case the entropy of the system under consideration and the entropy of the system with which it interacts (and that it performs work on it) are equal in the form, but they have opposite sign, because one yields heat and the other acquires it. As a result, the total change in the system is zero. In the real case, however, the total change of entropy is positive, since the value of entropy of the system that performs work (which is positive) is greater than that of the system under consideration (negative). Therefore, in a real case, a transformation that takes place in a non-isolated system causes a decrease in entropy in the physical system, and an increase in entropy in the universe.
In an ideal instance, the entropy of the system under study and the entropy of the system with which it interacts (and on which it does work) are identical in the form, but have opposite signs, because one yields heat and the other gains it. As a result, the overall change in the system is zero. In practice, however, the overall change in entropy is positive since the value of entropy of the system that produces work (which is positive) is larger than that of the system under consideration (negative). In practice, a transformation that occurs in a non-isolated system generates a drop in entropy in the physical system and an increase throughout entropy in the universe.

2.3.2. Data-Driven Approaches

Data-driven approaches allow to analyse parameters within different fields, e.g. product, process and logistics, and enable the extrapolation of forms of cause-effect interactions that traditional methodologies (i.e. statistical models, physical models) cannot identify on their own. In this way, quality issues may also be defined and managed along with sustainability concerns.
While model-based approaches rely on equations of states and boundary conditions to describe reality, data-driven approaches can discover many hidden relationships between data and gather previously unknown knowledge. But yet, they have their drawbacks: the most significant disadvantage is the difficulty in interpreting the results. So, these two approaches are not mutually exclusive, in fact more and more they are used in conjunction to solve problems.
Among the benefits of using data-driven techniques in manufacturing businesses are the elimination of mistakes and biases in decision-making, increased efficiency, improved communication, and the promotion of transparency and understanding. These correspond to drawbacks, such as the fact that each model must be customized for each project. There is currently no invariant model that can adapt to any environment without human intervention. Furthermore, a data-driven business culture might cause people to underestimate their own judgment and experience. During data analysis, it's crucial to maintain a healthy scepticism about results that look too wonderful to be true, or, conversely, too bad to be real. If something appears to be off, it is a good indication that it is. When you look at your data, you want to make sure that everything is in order and that nothing is illogical. Mistakes happen, and data is no exception. A fully automated system complicates this process and opens the door to possible misinterpretations of the procedure's reality. It takes time to learn the ability to correctly read and analyse data, and the potential of working with low-quality data is extremely high. It is, nevertheless, critical if you want your data to realize its maximum potential. Data visualization tools may assist you in quickly identifying linkages, trends, and correlations in your data. It draws the most significant information to the user's initial attention, allowing even individuals who are unfamiliar with data to better properly grasp its meaning (Dahlin, 2021).

Further issues address the economic factor: while it helps to increase the quality and efficiency of resource production and consumption, the initial investment capital is costly. The development of these technologies necessitates a significant investment. Each step along the process necessitates an investment to ensure success. For starters, the algorithms must be created by a team of developers. Then there's the portion where you have to teach new individuals on the machine learning language and the implementation process. Finally, you will require industry-specific machinery. And all of this comes at a hefty price.
Among Smart Technologies, IoT and AI enable businesses to optimize the outcomes obtained from the data-based approach. On a practical level, the application lines for AI technologies are primarily three: increased operational efficiency, improved product quality and safety, and a lower environmental impact of the entire production chain. The most complex challenges arise when it comes to having a non-intrusive architecture to gather data from older industrial plants (Farooqui et al., 2020). Exploiting big data intelligence to gain a competitive edge would necessitate addressing new problems, such as how to ensure data accuracy and significance across lengthy and dynamic supply chains (Bogle, 2017).
IoT technology has created exciting opportunities to develop powerful tools for monitoring and management production through sensor systems. However, there are still difficulties in the application of this technology, such as, for example, what kind of data needs to be collected and how to properly acquire it, because there are multiple consumption points and lots of sub-processes in the system, and how to specifically analyse the data collected from multi-sensors in order to determine the real operational state of the overall system and all the sub-systems as well, as the data acquired by a single sensor (e.g. a power sensor, a temperature sensor) may not provide sufficient information. For a hierarchical structure such as those developed by (Lee et al., 2015), process inefficiencies and effect-relationships can be detected and optimized automatically.
Finally, there is no standardized method for organizing datasets, analytical models, and interpreted results. More intelligence functions will be required to manage the algorithms and their input/output if the platform presented in this work is used by other organizations. Concerning the algorithms, as the models evolved, they got increasingly sophisticated. So, one looks at a model simply from a performance standpoint, neural networks, Gradient Boosting, and so on are typically the best models since they are relatively new. However, various models perform better with different types of data. For example, if features are very independent, Naive Bayes will perform well. SVM is useful when there are too many characteristics and the dataset is middling in size. If the dependent and independent variables have a linear relationship, linear regression, logistic regression, and SVM are appropriate. k-NN can be implemented if the dataset is tiny and the link between the dependent and independent variables is unknown. As a result, before deciding on which ML method to utilize, one must first understand and analyse the data. If you can't decide on a single machine learning algorithm, you can analyse all of them and compare their accuracies on training and test sets before settling on one.
There are methods that can be used for multiple applications, such as regressors and classifiers, which may be employed supervised or unsupervised. A beginner practitioner has a tough time determining what strategy to take right away, especially if the goal of the study is not completely stated and the raw dataset does not provide any information on the sort of data on which you must work. Many websites and platforms now offer free cheat-sheets that can help to choose the best algorithm according to the requirements or to the properties of the dataset.
However, the first implementations should not be considered as certain because an initially inadequate approach may be appropriate if performed with the necessary data pre-processing and/or with the correct tuning of hyperparameters.
The graphic in Figure 2.31 below is a Scikit-Learn cheat-sheet example, which is quite handy if you are in the initial weaponry and dataset and know essentially the size and kind of target.

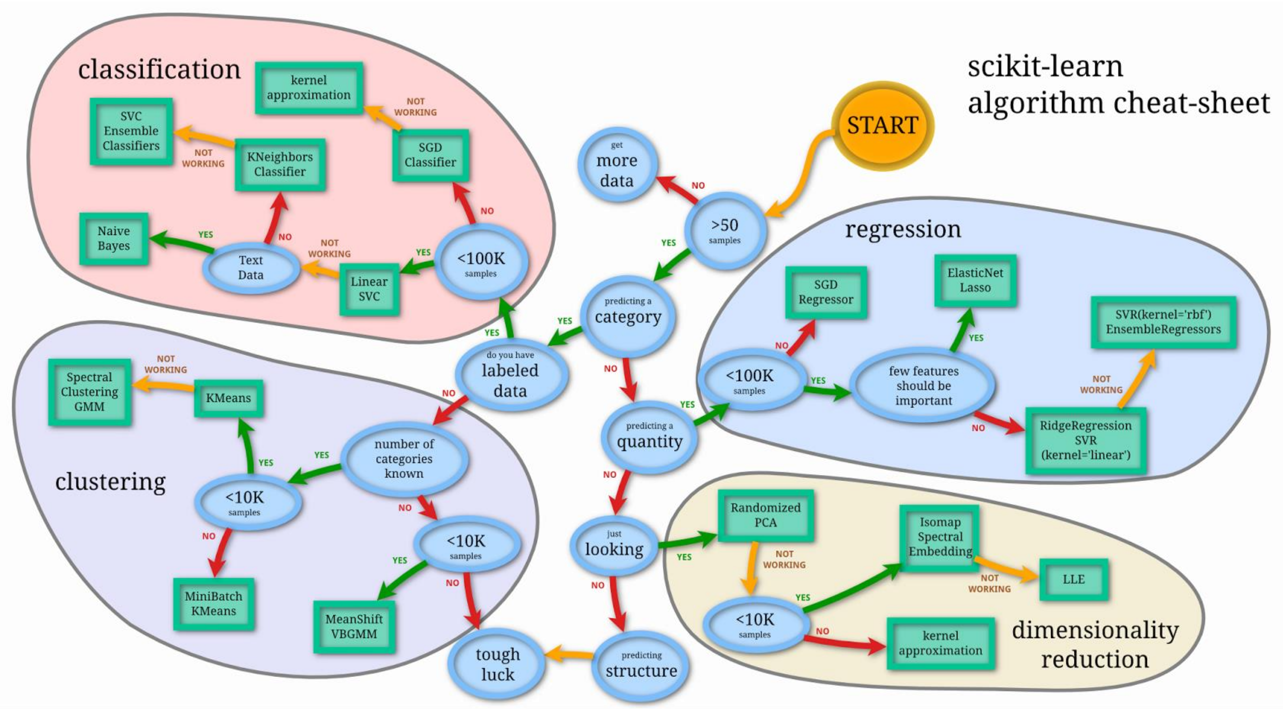


Figure 2.31 - scikit-learn algorithm cheat-sheet (https://scikit-learn.org/stable/_static/ml_map.png)

Instead, figure 2 indicates which of the proposed methods performs better in terms of implementation speed or forecast accuracy. It also considers dimension reduction to be a pre-processing phase if the amount of data is large.

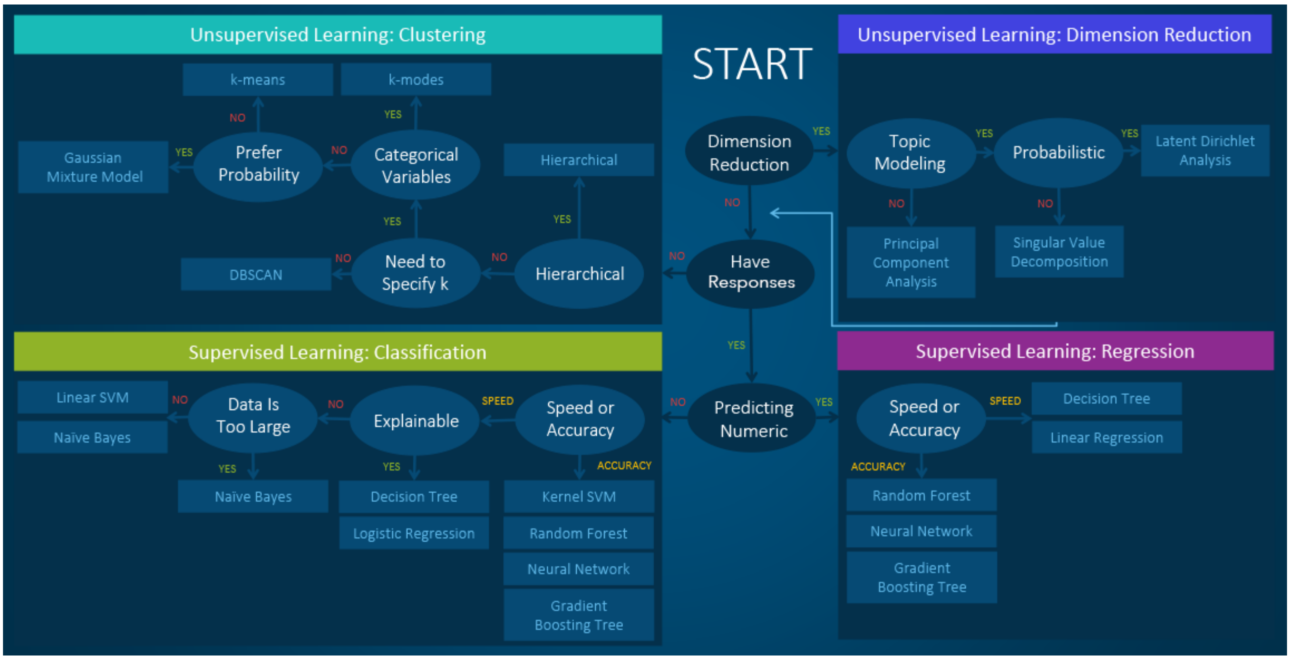


Figure 2.32 - Cheat sheet for ML algorithms depending on accuracy or implementation speed (https://blogs.sas.com/content/subconsciousmusings/files/2017/04/machine-learning-cheet-sheet-2.png)

Because the world of ML is made up of so many tests, attempts, and adjustments to every single case study, it is still impossible to talk about the approach and model being invariant and completely adaptive to a reality as dynamic as industrial manufacturing. The same is true for various stakeholders, each of whom brings their own set of skills, tools, methodologies, datasets, and objectives. The numbers are too sectoral and, at times, unable to speak with one another or derive a meaningful decision-making policy from their unique analyses (Karmaker Santu et al., 2020).

It’s worth noting that smart manufacturing should focus not only on enhancing economic and environmental sustainability, but also on improving social sustainability. As a result, process safety risk prediction can be an integral part of smart process manufacturing (Gobbo et al., 2018; Moktadir et al., 2018). The risk evaluation should be the first and most important phase in irregular situation management in order to develop a preliminary profile of the risk situations to be handled. Alarm detection, workflow control, equipment fault detection, and human activity tracking can all be effectively combined in the sense of big data analytics (Yuan et al., 2017).

While businesses are working to improve internal know-how, this first section of the dissertation demonstrated that manufacturing engineers need a basic method for data analytics rather than complex algorithms. These practitioners may need assistance in determining the algorithm is best suited to their problems, what kind of data should be collected for the algorithm, and how the collected data should be pre-processed.

3. INTERPRETING SUSTAINABILITY

The interpretability of the results provided downstream of thermodynamic and/or LCA analysis, as well as the metrics returned by the implementation of machine learning algorithms on process datasets, was systematically addressed during the doctoral programme. In the following two paragraphs, what can be deduced from the state of art analysis will be examined in detail. Following that, in the case study, explanatory applications of these metrics and the resulting interpretations will be reported to facilitate decision-making on the business strategies to be implemented to improve the quality and sustainability performance of the analysed industrial processes.

3.1. INDICATORS OF REVERSIBILITY

In manufacturing, sustainable development is the process of continuously improving environmental, social, and economic (cost-benefit) performance through time.
The greatest challenge in adopting a sustainable strategy is from the difficulty of accurately assessing and evaluating one's work's economic, environmental, and social effect. Measuring, monitoring, evaluating, and communicating sustainability is an important step in policymaking.
The first step that a practitioner must do is to map all the sustainability components of the system to be analysed. Once all relevant areas have been identified, the improvement targets must be established. These goals include, for example, limiting resources consumed and, as a result, maximizing value through decreasing energy consumption, optimizing the plant, lowering CO_2 emissions, and so on. The downstream interpretation of these interventions, as well as the resulting strategic measures to be taken, may be carried out using suitable indicators to measure the system's success versus its objectives (i.e., the level of performance to be achieved). "Only what gets measured gets managed", stated Peter Drucker (Klaus, 2015).
Environmental stewardship, economic growth, social well-being, technical innovation, and performance management are the five factors to be evaluated in the manufacturing industry, increasing from three to five. Technology advancement accounts for firms' propensity to foster technological improvement via R&D conscription, investment, and high-tech items. Performance management is concerned with the execution of sustainability initiatives and policies, as well as regulatory compliance (Joung et al., 2013).
Evaluation of various dimensions, both traditional and novel, need sophisticated evaluation methodologies, and in this case, an Exergetic Analysis (EA) combined with a Life Cycle Assessment (LCA) yields a solid implementation plan. Among the modelling and sustainability analysis approaches available in the literature, this article focuses on the strategic coupling of thermodynamic laws, and therefore the Exergetic Analysis, and the Life Cycle of the product, process, service/activity. These two methodologies had been hybridized in many ways and on multiple levels, making it difficult to immediately evaluate the data in order to generate the best decision-making strategies for the instance analysed. (Selicati et al., 2021a).
The goal of this paragraph is to first present a thorough analysis of all metrics linked to the hybrid or combined usage of exergy and LCA, their significance, and their application in specific use scenarios. The second goal is to give a broader definition of the measure as a tool to aid in the understanding of the assessment findings.
Many indicators or sets of indicators published in the literature appear to be designed to offer trustworthy information on various parts of the global sustainability environment, but it is always a difficulty when aggregation of findings is required as an integrated measure. Their accuracy in assessing environmental, but mainly social and economic, factors is yet unknown. The absence of full and up-to-date data and uncertainty analysis is frequently problematic, as is the lack of scientific consistency in the assessment's interpretation.
Manufacturing processes generate material riches for people, but they also generate a lot of waste and use a lot of resources. Waste created throughout the production process, during product usage, and after the product has reached the end of its useful life contributes to environmental damage. As a result, decreasing resource consumption and the environmental effect of production systems has become increasingly crucial. As a result, striving towards sustainable production is crucial for manufacturing businesses.

To characterize the indicators in terms of their relevance and value to sustainable production, a thorough grasp of sustainable manufacturing is essential. Although there is no common idea of sustainable manufacturing, the US Department of Commerce explains it as follows: “The creation of manufactured products that use processes that minimize negative environmental impacts, conserve energy and natural resources, are safe for employees, communities, and consumers and are economically sound” (Howard, 2011).

Sustainable production is defined as a system that integrates product and process design issues with manufacturing, planning, and control issues in order to identify, quantify, evaluate, and manage the flow of environmental waste, with the ultimate goal of reducing the environmental impact of the Earth's self-recovery capability while also attempting to manage it. As a result, the sustainable approach must be linked to a change policy in order to attain this aim with a consistent effort in a realistic time frame for current and future generations. The 6R technique might help to achieve this shift toward new linear thinking: Reduce, Reuse, Recycle, Redesign (or Rethinking), Recover, and Remanufacture. Reduce especially relates to the manufacturing process, when the quantity of energy, material, and waste should be kept to a minimum. This is related to the reuse of components after their original life cycle in order to decrease raw resource usage. Recycling waste materials is another technique to reduce the consumption of raw materials. Recover is a method of extending a product's life cycle, which might be extended to remanufacturing, which is the process of restoring the product to its original state. All of this can only be accomplished by rethinking the product and the life cycle with a long-term goal in mind (Jayal et al., 2010). The ISO standards are arranged into four parts based on a process-based LCA approach: aim and scope definition, inventory analysis, impact assessment, and interpretation. Because this paragraph focuses on the interpretation of sustainability, the last step in the process takes the results of the previous three processes and makes recommendations for improving the environment of the product or process under examination. In an ideal world, this information would give direct guidance to constructive measures, such as the development of environmental projects. EA is used to track processes or a specific product from a thermodynamic standpoint.

Among the optimization parameters is the reduction of exergy loss owing to system irreversibilities. EA combined with LCA provides significant benefits: first, they provide more objective evaluation results; second, they become a valuable tool for decision-making policies aimed at creating a retrofitting solution, allowing the system to automatically avoid any potential failure. Furthermore, they are important tools for understanding process management options in order to enhance and develop industrial process technologies. Among the optimization parameters is the reduction of exergy loss owing to system irreversibilities (Cornelissen and Hirs, 2002). The hybridization method is also quite effective for analysing the data. While most sustainability indicators (such as Carbon Footprint or Global Warming Potential) must be contextualized within international regulatory processes and frameworks in order to meet the requirements of scientificity, reproducibility, and reliability, the indicators (or, in this case, efficiencies) returned by the exergetic analysis or when combined with the LCA are self-explanatory and simple to understand.

3.1.1. Sustainable Manufacturing Indicators Aspects

Indicators have been identified in a variety of approaches in the literature. (see the reviews in (Heink and Kowarik, 2010) and (Singh et al., 2009)). In most cases, the term indicator refers to a tool that may convey information in a synthetic form that is simpler than a more complicated occurrence but has a larger significance. As a result, it is a tool capable of bringing to light a pattern or phenomena that is not immediately apparent. An indicator is a metric or aggregation of measurements that may be used to draw inferences about the phenomena of interest based on the objective set.

In order to be scientifically legitimate, any indicator must have the following properties (United Nations, 2019):

- Ease and comprehensibility: if an indication is not instantly intelligible owing to extremely complicated or inconsistent measurements, its use as an instrument of internal governance and a means of external communication is severely restricted. Furthermore, an unintelligible indication might lead to misunderstandings.

- Importance and comprehension: an indicator should enhance decision-making by highlighting areas for improvement. It should not include too many technical characteristics, nor should it be overly wide; it should encompass all of the important aspects and relevant repercussions in connection to the purpose of the study.
- Manageability and comparability: the performance standards of the indicators should be evaluated by assuring their comparability and replicability, therefore providing a continual standard in the area to which they belong.
- Controllability: a practitioner must be completely aware of what he is going to measure and calculate in order to deliver precise and timely signals to stakeholders.
- Consistency: in order to prevent invalidating the analysis, the indicator must be regularly reviewed and, if necessary, changed in reaction to changes.
- Efficiency: one of the most essential features, as an indication for which exorbitantly costly data gathering is necessary, or data collection is technically impossible, has a negative influence on the whole performance of the analysis, including the final phase of interpretation.

In general, and hence in terms of production sustainability, indicators can be defined in a variety of ways. The first significant distinction is established between indicators that relate to immediately quantifiable events and indicators that relate to phenomena that cannot be measured directly. Many of the phenomena that impact sustainability may be measured. Some are directly measurable chemical and physical phenomena (for example, CO_2 emissions); others, on the other hand, are characteristics for which we do not have direct measuring instruments but which can always be expressed quantitatively by reference to an appropriate and considered intensity scale (Boulanger, 2008). There is a distinction to be made between physical indicators, units of measurement, and levels of variables designated as relevant; and multidimensional indicators or indices, which consist of the aggregation of indicators and data of the same or different categories. In concrete terms, the former provides basic information on the system's components, such as quantities and flows (for example, annual waste generation in a region), whereas the latter allows the information presented to be condensed into several parameters in order to better communicate and facilitate knowledge (e.g., coupling between waste production and economic well-being measured by the ratio of the waste produced to the gross domestic product of a region). Indices, for example, emphasize the link between system components; moreover, indices can be given in absolute numbers via standardization and aggregation of the beginning information.

Indicators are commonly used to better evaluate and explain the results of a hybrid study, as well as to quickly compare various production or multiple systems with different units of measurement. They can also give aggregate data. They are simple to comprehend since their value may range from zero (worst case scenario) to one (best case scenario) (ideal conditions). Indicators are a good method for swiftly and intuitively recognizing changes in the energy efficiency and quality of time-dependent operations. In addition to satisfying a variety of scientific problems, environmental indicators and the weighting system must represent the aims of the many environmental issues, priorities that are clearly related to the belief system of individuals who define them and thus subjective. Some nations have developed their own scale of environmental goals, allowing professionals to develop an effective set of indicators.

3.1.2. State of Art of Sustainability Indicators in Manufacturing field related to Exergy and LCA

Many researchers stress the value of utilizing exergy losses as an indication since it gives a consistent metric for comparing and evaluating diverse processes (Bakshi and Fiksel, 2003; M.A. Rosen et al., 2012). Exergy-based indicators provide useful sustainability measures for assessing the exploitation of material resources and energy, as well as evaluating the side effects of ecological and socioeconomic behaviours in complex systems. The indicators from LCA presented following the characterization and normalisation of the assessment, on the other hand, have traditionally been deemed erroneous due to the subjectivity that characterizes this stage

(Cleveland et al., 2000), furthermore, only a handful of the LCA evaluation methodologies allow for the generation of dimensionless indicators of the necessary effect categories. Many indicators or sets of indicators have been proposed in the literature that are supposed to provide reliable information about various aspects of the global context of sustainability, but it is always a problem when an aggregation of results, such as an integrated index of LCA and EA results, is required: their accuracy as environmental, social, and economic indicators is still not clear. (Böhringer and Jochem, 2007).

Material, energy, and other streams participate in the process inside the system, and they are changed into the product and waste streams. The exergetic yields associated to the exergetic balances of the process/activity itself provide the performances (or yields) of a specific process or activity. The classic exergy efficiency rate informs about the ratio of benefits to expenses or losses. The losses are equivalent to the difference between what is offered and how much is gained, and they are associated with the irreversible destruction of exergy (Bakshi et al., 2011).

The formulations and meanings of the most typical metrics that we will examine are discussed in Table 3.1.

The traditional exergy efficiency rate provides information about the ratio between benefits and costs or losses. The losses are equal to the difference between what is provided and how much it is obtained and identified with the destruction of exergy due to irreversibility (Bakshi et al., 2011). The most often used exergetic indicators are the output/input exergy ratio (for assessing efficiency) and exergy per unit of product (for sustainability assessment). The performance metrics of the process or its components are defined in the following net use and general efficiencies, depending on whether the goal is to evaluate the portion of useful exergy for the realization of the final product or to evaluate the overall exergy of the process: output/input exergy ratio (for efficiency assessment) and exergy for unit of product (for sustainability assessment). The performance metrics of the process or its components are specified in the following net and general efficiencies, respectively, depending on whether the goal is to assess the percentage of useable exergy for the realization of the final product or to evaluate the total exergy of the process, respectively η_ε and η_g. The former is the ratio of the system's useable exergy to the total exergy supplied to the system, whereas the latter is the ratio of total exergy production to total exergy given to the system. The ratio is proportionate to the system's intrinsic exergy destruction. The Global Warming Potential (GWP) was created to enable comparisons of the global warming impacts of various resources (IPCC, 2006). It is a measure of how much energy a ton of a resource may utilize over a specific time period in contrast to a ton of CO_2 emissions (CO_2). The higher the GWP, the more a particular gas heats the Earth in compared to CO_2 during the same time period. The most frequent time span for GWPs is 100 years. GWPs provide a standard unit of measurement that allows analysts to add up emissions figures for different gases (for example, to compile a national GHG inventory) and policymakers to compare emissions reduction opportunities across industries and gases.

Cumulative exergy extracted from the natural environment (CEENE), introduced for the first time by (Dewulf et al., 2007), is a resource accounting system that quantifies diverse types of resources per functional unit in a single unit (exergy). The quantity of energy equivalent to each input in each process is computed by multiplying the resource inputs by the CEENE factor of the reference flow. The CEENE model is based on general global characteristics. It takes into account the depletion caused by the extraction of useful exergy embedded in resources when they are extracted from their natural environment, such as abiotic renewable resources, fossil fuels, nuclear energy, metal ores, minerals and mineral aggregates, water resources, land and biotic resources, and atmospheric resources. Many authors have utilized CEENE in their studies, including (Mehmeti et al., 2018) , which used CEENE to quantify the life cycle resource footprint (upstream effects) of a Molten Carbonate Fuel Cell power plant, (Alvarenga et al., 2013) who proposed and implemented a new framework for calculating exergy-based spatial explicit characterization factors (CF) for land as a resource, which deals with both biomass and area occupied on a global scale by creating a schematic overview of the Earth, dividing it into two systems (human-made and natural), allowing it to account for what is actually extracted from nature, i.e., the biomass content was set as the elementary flow to be calculated We were able to develop CF for land resources for these two separate systems using exergy. The novel CF's applicability was evaluated for a variety of biobased goods. And (Taelman et al., 2014), who included the CEENE method

in the LCIA method and was capable of analysing the environmental impact (and, more specifically, the resource footprint) of marine area occupation in two case studies: comparing resource consumption of on- and offshore oil production, and fish and soybean meal production for fish feed applications.
In their recent work (Lucia and Grisolia, 2019) introduced other two exergy-based indicators, a modification of the classic exergy efficiency ratio, were introduced to quantify the technical level of a process in relation to its unavailability. The goal was to assess the equivalent primary wasted resources, technological features, and advanced level of industrial processes by calculating the cost of the wasted exergy required to support workhours and generate capital flow, as well as the quantity of production expressed by mass and moles of CO_2 for wastes. Also (Cleveland et al., 1984) adapted the exergy efficiency naming it Energy (Exergy) Return on Energy (Exergy) Investment (EROEI or ExROI), which is defined as the ratio between the net exergy generated by the system and the embodied non-renewable exergy necessary to develop the system itself. If the ratio turns out to be less than the unit, the expenditure outweighs the gain. In the aforementioned study (Beccali et al., 2003) created an 'exergetic index' by dividing the entire consumption of exergy (in MJ) by the mass of the product that represents the functional unit of the case study (in kg). It is a particularly valuable tool for assessing the potential for technological development of processes and gauging quality. In the context of multi-criteria or multi-factor decision making, in their study (C. Zhang et al., 2019) calculated eight multi-factor indicators representing exergetic, energetic, economic and environmental elements of the Organic Rankine Cycle for water heat recovery in their study. The eight indicators are in the energetic context: net power output (W_{net}) and thermal efficiency (η_{th}); in exergetic context: total exergy loss (I_{tot}) and exergy efficiency (η_{ex}). In the economic context: cost per unit of time (Z), electricity production cost (EPC) and dynamic payback period (DPP). In the environmental context: CO2-equivalent emissions (ECE). Weighting and normalisation were used to construct the Feasibility Level, which represents the total influence of the eight indicators. Medyna et al. conducted an early environmental evaluation assessment (Medyna et al., 2009c, 2009a) comparing the three major impact categories of Eco-Indicator 99 (Human Health, Ecosystem Quality and Resource Consumption) with new three dimensionless indicators $\prod$ derived from EA (primary exergy conversion efficiency, material and resource consumption efficiency and environmental impact efficiency), in order to offer a possible solution to the heterogeneity metrics problem during the interpretation of the results. The difference of the meaning of these three new indicators lies in the considered exergetic terms for the ratio. (Rubio Rodríguez et al., 2011), in the energy systems context, presented a dimensionless sustainability index SIC in the context of energy systems to assess alternative to various end services that imply distinct metrics and magnitudes but referred to the same functional unit. The index indicates the environmental damage averted by selecting the best solution. Another point of view is provided by (Domínguez et al., 2011), who introduced an indicator called the 'renewability factor' (FR) in order to evaluate the relationship between non-renewable and renewable resources throughout the entire life cycle of each energy source considered for electric power generation. It represents the ratio between the cumulative exergy demand for renewable resources to the cumulative exergy demand for non-renewable resources. (Dai et al., 2014) in their work provided a list of six EEA-based indicator for the evaluation of the effective use of resources and energy in complex systems of some industrial sectors including environmental, social, and economic dimension. Another way to evaluate the sustainable use of the resources is given in 2006 (Toxopeus and Lutters, 2006) and it was used also by Koroneos and Stylos in 2014 in their implementation of an ELCA on polycrystalline photovoltaic system in energy generation context (Koroneos and Stylos, 2014). They introduced an ELCA-based exergetic eco-efficiency indicator to account the efficiency of consumption for both renewable and non-renewable sources along the Life Cycle of the product or process under study. It relates the exergetic efficiency of total input and output flows to the distinction between renewable and non-renewable flows throughout the Life Cycle. The large disparity between conventional and new exergetic efficiency values is due to the amount of renewable exergy (solar radiation) in the production of total incoming exergy, which the traditional indicator cannot capture. (Restrepo and Bazzo, 2016) addressed the Exergoenvironmental study from a systematic approach on co-firing power plants in 2016. The writers concentrated solely on the operational phase. They developed the Exergoenvironmental-based Global Greenhouse Gases index for a variety of co-firing scenarios in order to assess the extent of the power plant's

improvement. The index's objective is to compare the exergoenvironmental impact of the real process under investigation to the impact of the identical ideal process (under Carnot cycle condition). A higher index value indicates a more sustainable process. (T. M. Gulotta et al., 2018) integrated EA in the LCA in their study by adding three new indices focusing on quality, irreversibility, and technical obsolescence to assist decision-makers in comparing similar technologies. The Life Cycle Irreversibility Index might reveal potential exergetic inefficiencies of the process or technologies and necessary retrofit measures by comparing the usable cumulative exergy associated with all sub-processes and the total cumulative exergy demand. The Technology Obsolescence index facilitates the comparison of identical processes and products that share the same functional unit. Technology obsolescence may be a valuable criterion in policy decision-making to assess how much more inventive one technology is than another by recognizing which new technology might lower current irreversibilities from production through end-of-life, decreasing natural resource extraction. In general, the feature of technology obsolescence is still visible in the examination of industrial processes.

Table 3.1 - List of the most representative indicators found in literature for manufacturing-sustainability

Indicator's name	Indicator's ratio	Meaning
Coefficient of resources-use performance (Bakshi et al., 2011)	$\eta_p = \frac{\dot{E}x^{product}}{\dot{E}x^{i}}$	The useful exergy produced by the system divided by the total exergy provided to the system.
Net use efficiency (Bakshi et al., 2011; Ozbilen et al., 2012)	$\eta_\varepsilon = \frac{\sum \dot{E}x^{j}}{\sum \dot{E}x^{i}}$	Total exergy output divided by total exergy given to the system. The ratio is proportionate to the system's intrinsic exergy degradation.
Equivalent wasted primary resource (1) (Lucia and Grisolia, 2019)	$EI_\lambda = \frac{T_0 S_g}{n_h n_w}$	The proportion of exergy lost to working hours per worker. It calculates the cost of squandered exergy necessary to support work hours and produce capital flow.
Equivalent wasted primary resource (2) (Lucia and Grisolia, 2019)	$EI_\lambda = \frac{T_0 S_{g,PS}}{\dot{m}_{CO2}} \dot{m}_{product}$	The exergy loss-to-wasted-product mass ratio expressed in CO2 and the mass generated in a day.
Exergy Return on Investment (Cleveland et al., 1984) adapted by (Rocco, 2014)	$ExROI = \frac{Ex_{net}}{Ex_{needed}}$	The quantity of net exergy obtained from a process divided by the amount of exergy required (or its equivalent from another source) to produce it.
Exergetic index (Beccali et al., 2003)	$ex_c = \frac{Ex_c}{m_{pd}}$ [MJ/kg]	The ratio of entire exergy loss, including environmental emissions, to a specific quantity of product representing the functional unit.
Global Warming Potential (IPCC, 2006)	$GWP = \frac{\int_0^{yn} F_{res}(t)dt}{\int_0^{yn} F_{CO2}(t)dt}$ or $GWP = \sum_{i=1}^{n}(m_i \cdot IF_i)$ [kgCO2eq]	The impact of a resource over a given time period when compared to the same amount of carbon dioxide (CO2) over the same time period.
Cumulative Exergy Extracted from Natural Environment (Mehmeti et al., 2018)	$CEENE_J = \sum^{i}(X_i \cdot a_{i,j})$ [MJeq]]	Accounting for several sorts of resources (measured in different units) per functional unit, all represented in exergy terms with a reference factor. The extraction of usable exergy contained in resources results in resource depletion.
Feasibility Level (C. Zhang et al., 2019)	$FL = \sum_{i=1}^{8}(x_i \cdot w_i) \quad x_i = \begin{bmatrix} \frac{X_i}{X_{opt}} & X_i \in (En, Ex) \\ \frac{X_{opt}}{X_i} & X_i \in (Ex, Eco, Env) \end{bmatrix}$	Economic (Eco), environmental (Env), energetic (En), and exergetic (Ex) measures are all used to calculate the overall influence of eight components.
Primary exergy conversion efficiency (Tsatsaronis and Morosuk, 2008a, 2008b)	$\prod_{PECE} = \frac{Ex_{prod-i} + Ex_{biprod-i}}{Ex_{material} + Ex_{supply}}$	The ratio of the usable outcome to the sum of the inputs that worked together to produce it.
Material and resource consumption efficiency (Tsatsaronis and Morosuk, 2008a, 2008b)	$\prod_{MRCE} = \frac{Ex_{prodi} + Ex_{env-standard}}{Ex_{mat} + Ex_{supply} + Ex_{recy} + Ex_{biprodi}}$	The output, minus the exergy loss, to the total of the inputs minus the regenerated biproducts.

Environmental impact efficiency (Tsatsaronis and Morosuk, 2008a, 2008b)	$\prod_{EIE} = \frac{Ex_{env-mixing}}{Ex_{mat} + Ex_{supply} + Ex_{recy} + Ex_{biprodi}}$	The ratio of the sum of the inputs to the exergy of mixing.
Sustainability index (Rubio Rodríguez et al., 2011)	$S_{IC} = \frac{\Delta k_E}{k_T}$	The ratio of the environmental exergetic cost of two alternatives to the indirect exergetic cost.
Renewability Factor (Domínguez et al., 2011)	$FR = \frac{CExD_{ren}}{CExD_{non-ren}}$	The ratio between cumulative exergy demand of renewable resources to cumulative exergy demand of non-renewable resources.
Exergy Structure Ratio (Dai et al., 2014)	$ESR = \frac{CEC_i}{E_{wi} + E_{ci}}$	Exergy consumption structure in various productions, derived by comparing material-based exergy consumption to social supporting exergy within sectoral size.
Social Exergy Conversion Rate (Dai et al., 2014)	$SECR = \frac{E_w + E_c}{\sum CEC_i}$	Net social exergy conversion level by intaking material-based exergy, calculated by the ratio of labour and capital equivalent exergy to exergy input into the system.
Exergy Deliver Efficiency (Dai et al., 2014)	$EDE = \frac{\sum f_{ij}}{CEC_j}$	Ability in production sectors to deliver exergy into the system from the environment, calculated by exergy output from production sector j divided by exergy input into production sector j from the surrounding.
Environmental Yield Ratio (Dai et al., 2014)	$EYR = \frac{CEC}{E_{in}}$	Ability of a process to exploit available locally renewable and non-renewable resources by investing outsider sources. The higher the value of this index, the greater is the return obtained per unit of exergy invested.
Environmental Loading Ratio (Dai et al., 2014)	$ELR = \frac{E_N + CEC_N}{E_R + CEC_R}$	Outside causes of disruption to the local drive are possible. The smaller the ratio, the lesser the environmental stress.
Extended Exergy Sustainability Index (Dai et al., 2014)	$EESI = \frac{EYR}{ELR}$	Index aggregation based on interaction with the surrounding environment as well as renewability.
Exergetic Eco-Efficiency (Toxopeus and Lutters, 2006)	$\eta_{eco} = \frac{\eta_{exergetic} \cdot (F_{n-r} + F_r)}{F_{n-r} + \eta_{exergetic} \cdot F_r}$	The efficiency with which renewable and non-renewable resources are used during the full Life Cycle of the product or process under consideration.
Global Greenhouse Gases index (Restrepo and Bazzo, 2016)	$i_{GHG}^{Global} = \frac{i_{GHG\ (carnot\ condition)}^{total}}{i_{GHG\ (real\ condition)}^{total}}$	Degree of improvement in relation with the impact category (focused in the GHG emission).
Life Cycle Quality Index (T. M. Gulotta et al., 2018)	$\psi = \frac{UCEx}{CExD}$	The ratio of the beneficial impacts of a process or product to the entire cost of providing that process or product.
Life Cycle Irreversibility Index (T. M. Gulotta et al., 2018)	$X = 1 - \psi = \frac{CExD - UCEx}{CExD}$	Complementary to ψ.
Technology Obsolescence Index (T. M. Gulotta et al., 2018)	$X_{i,j} = \frac{X_i}{X_j}$	Which revolutionary technology, when compared to existing technologies, has the potential to lower the irreversibilities of the process or product under consideration.

The authors provide these indicators under a variety of titles, but the general pattern is a comparison of the system's output flows with its input flows, with certain special idiosyncrasies for each case study. Furthermore, their significance is stated in many ways as an indicator/index of quality, performance, efficiency, and sustainability.
In this thesis, the ideal interpretation for this sort of indication is ***'indicator of reversibility'*** (Selicati and Cardinale, 2021b). The choice is endorsed first by the definition of sustainability as stated the Brundtland Report in 1987 (WCED, 1987): "is the development that meets the needs of the present without compromising the ability of future generations to meet their own needs"; second by Dewulf et al. in (Dewulf et al., 2000) who argued that a technological process is sustainable only if its resource supplying, production and resource depletion or wastes won't damage the ecological balance in the ecosphere. This implies ensuring that the process consumes raw materials from the environment at a rate lower than their potential to regenerate. Third, by (Valero et al., 2013) that established an indicator called "exergy replacement cost," which is the amount of exergy required to return the resources to their initial state (equilibrium).
To summarize this paragraph, sustainable manufacturing is the most crucial component that all production engineers must identify, not because it is a cultural trend, but because it is a mandate as a duty to the environment in which we live. The study of the product life cycle has become a popular tool for determining the environmental impact of items, processes, or activities. To reach the goal of earth's self-recovery capabilities, the three key ideas to be addressed are minimizing the use of resources in the process, using environmentally friendly materials, reducing all sorts of waste, and reusing and recycling as much material as feasible.
According to the findings of the state of the art and the case study, it is not feasible to establish an indicator that individually and thoroughly assesses the degree of manufacturing sustainability, not one of the steel corner production methods. Despite the lack of a defined, thorough, and widely used assessment model, Exergy Analysis within Life Cycle thinking remains an effective technique for optimizing industrial processes.
The multidimensional nature of the measures described in this paper emphasizes how difficult the topic of sustainable manufacturing is. The lack of suitable metrics and a well-established collection of equations for a set of sustainability challenges, as well as a lack of complete and up-to-date data and uncertainty analysis, are frequently troublesome, resulting in a poor level of scientific accuracy in the evaluation.

3.2. METRICS FOR DATA-ANALYSIS

Predictive models rarely predict everything perfectly, so there are many performance metrics that can be used to analyse the models.
Aside from a simple comparison of the analysis findings and actual data from the factory, several methodologies are necessary to evaluate the accuracy and performance of the generated algorithms.
The algorithm's performance may be measured in a variety of ways. Classifiers are frequently evaluated using accuracy, precision, recall, and F1-score. Regression techniques can be scored using mean absolute error (MAE) and mean squared error (MSE) metrics. While the Silhouette coefficient and Dunn's Index are the two most often used metrics for evaluating clustering techniques. Regardless of the algorithm, the error on the training data is often less than the error on the test data. The model is considered to be overfit to the training data when the difference between the two is considerable. Over-fitting is a concern for data scientists since such a model does not generalize. The model performs admirably on training data, but when fed fresh data, the algorithm's predictions become untrustworthy.

3.2.1. Metrics for Clustering

The most commonly used type of unsupervised learning is clustering. The dataset contains no labels for clustering, only a collection of observational characteristics whose objective is to build groups with similar observations aggregated together and different observations separated as much as feasible. Evaluating the

performance of a clustering method is more complicated than just calculating the number of mistakes, accuracy, and recall, as is done with supervised learning algorithms.
The clusters' consistency is assessed using a similarity or dissimilarity metric, such as the distance between cluster points. The clustering method has worked well if it isolates different observations and groups like observations together.
The mean of the Silhouette Coefficients for each sample is used to calculate the Silhouette Coefficient for a set of samples (scikit-learn, 2021).

$$s = \frac{b-a}{\max(a,b)} \qquad \text{Eq. 3.1}$$

Where
- s is the Silhouette Coefficient
- b is the average distance between a sample and all other points in the closest cluster
- a is the average distance between a sample and all other points in a cluster

The score ranges from -1 for inaccurate clustering to +1 for robust clustering. Scores close to 0 denote overlapping clusters. The score is greater when clusters are dense and well-spaced, which corresponds to a conventional cluster idea.

Given the information of the samples' regression coefficients class assignments, conditional entropy analysis may be used to establish any understandable measure.
The following two desired goals for each cluster assignment can be converted into scores:
- homogeneity: each cluster only contains members of one class.
- completeness: every member of a particular class is allocated to the same cluster.

which are both constrained by 0.0 (worst clustering) and 1.0 (best clustering):
Their harmonic mean, known as the V-measure, is determined (Rosenberg and Hirschberg, 2007)

$$v = 2 \cdot \frac{\text{homogeneity} \cdot \text{completeness}}{\text{homogeneity} + \text{completeness}} \qquad \text{Eq. 3.2}$$

Clustering with a poor V-measure can be qualitatively evaluated in terms of homogeneity and completeness to get a clearer sense of the 'kind' of errors committed by the assignment.
There are no assumptions about the cluster structure: it may be used to compare the results of clustering algorithms such as k-means, which assumes isotropic blob forms, with the results of spectral clustering algorithms, which can identify clusters with "folded" shapes. As a disadvantage, the metrics are not standardized in terms of random labelling: depending on the number of samples, clusters, and ground truth classes, a totally random labelling will not always produce the same values for homogeneity, completeness, and therefore v-measure. Random labelling, in particular, will not produce zero scores, especially when the number of clusters is considerable. Furthermore, these metrics need knowledge of ground truth classes, which is usually never accessible in practice or necessitates manual assignment by human annotators (as in the supervised learning setting).

Another metric for assessing a clustering technique is Dunn Index (DI) (Stein et al., 2003). It is a metric for evaluating clustering algorithms, is an internal evaluation scheme, where the result is based on the clustered data itself. Like all other such indices, the aim of this Dunn index to identify sets of clusters that are compact, with a small variance between members of the cluster, and well separated, where the means of different clusters are sufficiently far apart, as compared to the within cluster variance. Dunn Index is calculated by dividing the lowest inter-cluster distance by the maximum cluster size. It is worth noting that larger inter-cluster distances (better separation) and smaller cluster sizes (more compact clusters) result in a higher DI value. A greater DI indicates improved clustering. It is assumed that superior clustering implies that clusters are compact and well-separated from one another. It also has some drawbacks. As the number of clusters and dimensionality of the data increase, the computational cost also increases.

Other metrics about clustering approach can be found in (Desgraupes, 2017)

3.2.2. Metrics for Classification

The most important categorization statistic is accuracy. It is rather simple to grasp. And it's well-suited for both binary and multiclass classification problems. The proportion of genuine outcomes among the total number of cases studied is referred to as accuracy. Data can be divided into: true positives are data points classed as positive by the model that are actually positive (meaning they are correct), whereas false negatives are data points labelled as negative by the model that are actually positive (incorrect).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FN+TN+FP}} \qquad \text{Eq. 3.3}$$

Where

- TP is the number of true positives
- TN is the number of true negatives
- FP is the number of false positives
- FN is the number of false negatives.

Accuracy itself is misleading when it comes to finding metrics for classifier performance evaluation on an unbalanced data set. This is so because, even if the classifier predicts all of the instances as the majority class, even then the accuracy will be very high. However, in case of unbalanced datasets, identifying the more important minority class is the goal. This can be tested and achieved by using balancing techniques and using other performance metrics for classification evaluation, such as, precision, recall, area under the ROC curve.

The capacity of a model to detect all relevant cases within a data collection is the precise definition of recall. It is computed mathematically by dividing the number of true positives by the number of true positives + the number of false negatives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad \text{Eq. 3.4}$$

Where

- TP is the number of true positives
- TN is the number of true negatives
- FP is the number of false positives
- FN is the number of false negatives.

Precision is defined as a classification model's ability to identify only relevant data points. It is defined mathematically as the number of true positives divided by the number of true positives and the number of false positives. False positives are casing the model incorrectly labels as positive that are actually negative,

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad \text{Eq. 3.5}$$

While recall represents the capacity to locate all relevant occurrences in a dataset, precision expresses the proportion of data points that our model said were relevant that were truly relevant.

The measures we pick to optimize, like most notions in data science, have a trade-off. In the case of recall, increasing the recall reduces the precision. The idea underlying the precision-recall trade-off is that changing the threshold for detecting whether a class is positive or negative will cause the scales to tip. That is, it will cause precision to rise but recall to decrease, or vice versa. Classifier computes a decision score for each instance, and if the decision score is equal to or greater than the threshold value, it predicts a positive class, indicating that the instance belongs to the class, target, or output. If the decision score is less than the threshold, the instance is in the negative class, target, or output. The majority of the classifier employs a threshold of 0.

An example of a relationship between Recall and Precision is plotted in Figure 3.1. The graph illustrates that if we require greater precision, we should set threshold higher than the default threshold value, and if we need higher recall, we should set threshold lower than the default threshold value.

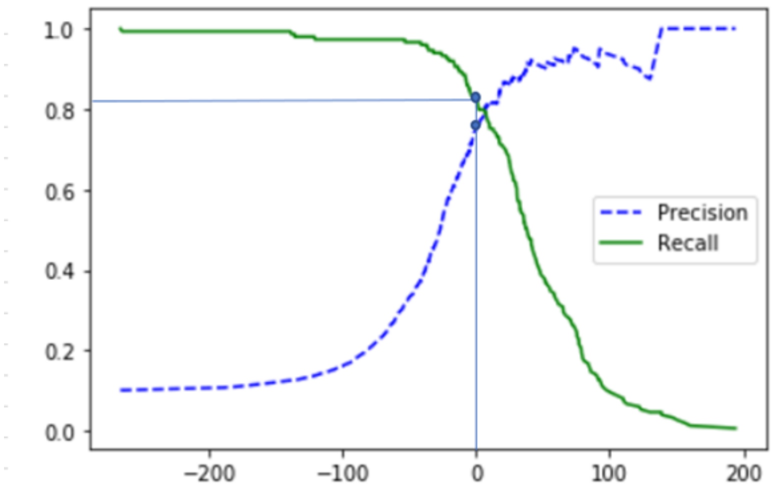


Figure 3.1 - Recall vs Precision with threshold equal to zero

In certain cases, it is unclear if the aim is to maximize recall or precision at the expense of the other statistic. When it is necessary to find an ideal balance of accuracy and recall, the F1 score combines the two variables. The F1 score is calculated by taking the harmonic mean of accuracy and recall.

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \qquad \text{Eq. 3.6}$$

Because it excludes extreme values, the harmonic mean is utilized instead of a simple average. A classifier with 1.0 accuracy and 0.0 recall has a simple average of 0.5 but an F1 score of 0. The F1 score gives equal weight to both criteria and is a subset of the generic Fβ metric that may be altered to give greater weight to either recall or accuracy by adjusting β. Other metrics for combining precision and recall, such as the Geometric Mean of accuracy and recall, are available, but the F1 score is the most widely employed. We strive to maximize the F1 score if we wish to develop a balanced classification model with the best balance of recall and precision.
The methods for calculating these metrics are well represented visually in the following Figure 3.2:

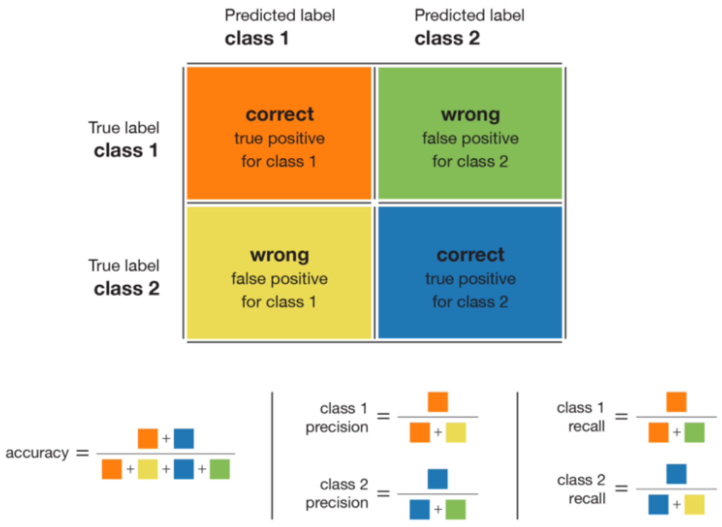


Figure 3.2 - Confusion matrix for binary classification and metrics calculation

The diagonal entries of the confusion matrix are the true predictions of both classes. The greater the number, the more accurate the predictions, and therefore the classifier. The misclassifications predicted wrongly by the classifier are represented by the diagonal elements. When compared to metrics such as accuracy, confusion matrix is a good metric to use in cases of unbalanced data sets. The reason for this is demonstrated with a simple example: suppose in a data set there are 99 good samples and 1 bad sample, then the classifier would be biased towards the majority class and make all predictions as belonging to the good class. If this is the case, the classifier's accuracy remains at 99%. In this scenario, it is clear that the classifier failed to identify the more relevant minority class. Similarly, finding the components that will fail the quality test is more critical in metal casting data than identifying the majority or passed class samples. The confusion matrix allows us to easily see the classifier's predictions on the minority class. The confusion matrix illustrates how many samples are erroneously recognized. They can also provide us with indices of the components that will fail. Instead of merely giving us the raw numbers of accurate and erroneous predictions for a certain data set, this will tell us the percentage of correct and incorrect forecasts.
In general, these measurements are interpreted as follows (Shivaprasad, 2020):
• High recall or precision values indicate that the model manages the class well;
• Low recall or precision values indicate that the model is reliable when it predicts a positive one for that class, but that the model is unable to identify the members of that specific class (it is true in our case for the class of the class "1" As the prediction of "1" is usually correct but does not identify several)
• A high recall value and a low precision value imply that the model correctly recognizes the class but also contains components from other classes (it is true in our case for the class of "0" as it intercepts almost all but predicts "0" also for many "1")
• Low recall or precision levels indicate that the model performs poorly.

<u>AUC</u> (which stands for "Area Under the Curve") is another metric for a classifier, where the curve in issue is termed ROC (Receiver Operating Characteristic curve). A model's ROC curve may be quantified by computing the overall Area Under the Curve (AUC), a statistic that ranges between 0 and 1, with a greater value indicating better classification performance. AUC ROC indicates how well the probabilities from the positive classes are separated from the negative classes. In essence, the curve depicts the trend of the probability threshold "T" above and below which a positive and negative is characterized in terms of belonging to a given class. For example, if it is determined that 0.7 is the value of the probability threshold above which a positive one is credited, we will have a given percentage of false positives and a certain rate of true positives at this threshold.

For all values of "T," the function that depicts the true positives vs. false positives, the trend of the sort is depicted in Figure 3.3:

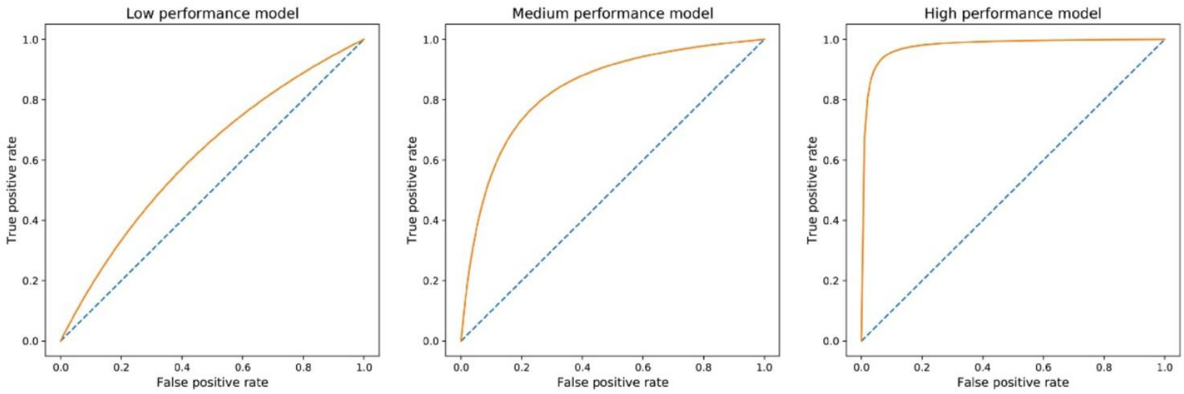


Figure 3.3 - Illustration of possible ROC curves depicting different model performance. From the left: the model has to sacrifice lot precision to get high recall; to the right: the model is highly effective and it can reach high recall while maintaining high precision

The true positive rate on the y-axis is shown against the false positive rate on the x-axis in a ROC curve. The recall is represented by the true positive rate (TPR), while the chance of a false alert is represented by the false positive rate (FPR). The confusion matrix may be used to compute both of these. The representation on the left indicates that a model with this property, in order to intercept a significant number of true positives, must inevitably produce a big number of false positives (the worst case is the dotted blue colour in which one is substantially a 50% of real positive).
Instead, in the case of the model on the right, with a low rate of false positives (the optimum scenario is in the top left corner), it has a large number of actual positives, and so, the more the curve has this tendency, the better the curve reacts. The AUC, and the diagram on the right will have a value of the order of 0.95 (which is already a very excellent number), but the diagram on the left will have a value of the kind of 0.6. This indicates that AUC is not affected by scale. It evaluates how well predictions are rated rather than their absolute values; it is classification-threshold-invariant: it assesses the quality of the model's predictions regardless of the classification threshold used, unlike F1 score or accuracy, which are affected by the threshold used.

3.2.3. Metrics for Regression

Regression is a term used to describe predictive modelling challenges that entail forecasting a numerical value. It differs from classification, which entails anticipating a class label. In contrast to classification, accuracy cannot be used to evaluate a regression model's predictions. Instead, there are error metrics that are particularly built for analysing regression predictions. Error addresses on average how close predictions were to their expected values (Naz et al., 2019).
There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE)
The Mean Squared Error, or MSE is calculated as the mean or average of the squared differences between predicted and expected target values in a dataset.

$$\mathrm{MSE} = \frac{1}{\mathrm{n}} \cdot \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{Y_i} - \hat{\mathrm{Y}}_\mathrm{i})^2 \qquad \text{Eq. 3.7}$$

Where

- n is the number of samples in the dataset
- $\mathrm{Y_i}$ is the actual value of the output (the one expected)

- $\widehat{Y_i}$ is the predicted value of the output (that ideally should be equal to the actual value)

The difference between these two numbers is squared, which removes the sign and results in a positive error value. Squaring has the additional consequence of inflating or amplifying big mistakes. That is, the greater the gap between the predicted and actual values, the greater the squared positive error. When MSE is employed as a loss function, this has the effect of "punishing" models more for higher mistakes. When employed as a measure, it also has the effect of "punishing" models by raising the average error score. A perfect mean squared error value of 0.00 indicates that all forecasts completely matched the expected values. This nearly never happens, and if it does, it indicates that the predictive modelling problem is easy.

The Root Mean Squared Error (RMSE) is an extension of the mean squared error. When the square root of the error is calculated, it indicates that the RMSE units are the same as the original units of the predicted target value.

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (Y_i - \widehat{Y_i})^2} \qquad \text{Eq. 3.8}$$

Where

- n is the number of samples in the dataset
- Y_i is the actual value of the output (the one expected)
- $\widehat{Y_i}$ is the predicted value of the output (that ideally should be equal to the actual value)

It is important to note that the RMSE cannot be determined by taking the average of the square root of the mean squared error values. This is a common blunder.

MSE use the square operation to eliminate the sign of each mistake value and to penalize big errors. This procedure is reversed by the square root, but the outcome remains positive. Like for MSE, a perfect root mean squared error value of 0.00 indicates that all forecasts completely matched the expected values.

Mean Absolute Error, or MAE, is a common metric because, like RMSE, the units of the error score correspond to the units of the anticipated target value.

Changes in MAE, unlike RMSE, are linear and hence intuitive.

In other words, MSE and RMSE penalize greater errors more severely than smaller ones, inflating or magnifying the mean error score. This is due to the incorrect value being squared. The MAE does not give distinct sorts of mistakes more or less weight, and instead, the scores grow linearly as the error increases.

The MAE score is derived as the average of the absolute error values, as the name implies.

$$\mathrm{MAE} = \frac{1}{n} \cdot \sum_{i=1}^{n} |(Y_i - \widehat{Y_i})| \qquad \text{Eq. 3.9}$$

Where

- n is the number of samples in the dataset
- Y_i is the actual value of the output (the one expected)
- $\widehat{Y_i}$ is the predicted value of the output (that ideally should be equal to the actual value)

When computing the MAE, the difference between an expected and forecast value might be positive or negative, and the abs forces it to be positive.

Like for MSE and RMSE, a perfect mean absolute error value of 0.00 indicates that all forecasts completely matched the expected values

R-Squared (R^2) represents the amount of variability in the dependent variable that the model can explain. It is the square of the Correlation Coefficient (R), thus the name R-Squared. R^2 is derived by squaring the prediction error and dividing it by the entire sum of the squares that replace the calculated forecast with mean. It ranges from 0 to 1, with a higher value indicating a better match between forecast and actual value.

$$R^2 = 1 - \frac{SS_{residuals}}{SS_{total}} = \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y_i})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \quad \text{Eq. 3.10}$$

Where

- n is the number of samples in the dataset
- SS means sum of squares
- Y_i is the actual value of the output (the one expected)
- $\widehat{Y_i}$ is the predicted value of the output (that ideally should be equal to the actual value)
- $\overline{Y}$ is the mean value of the entire samples in the dataset

R^2 is a useful metric for determining how well a model fits the dependent variables. R-squared values range from 0 to 1 and are commonly stated as percentages from 0% to 100%. In the best-case scenario, the modelled values precisely match the observed values, resulting in $SS_{residuals} = 0$ and $R^2 = 1$. While $R^2 = 0$ for a baseline model that always predicts $\overline{Y}$. Models with poorer predictions have a negative R^2 value.
Figure 3.4 shows plotted graphs, one with a high value of R^2 and the other a low value of R^2.

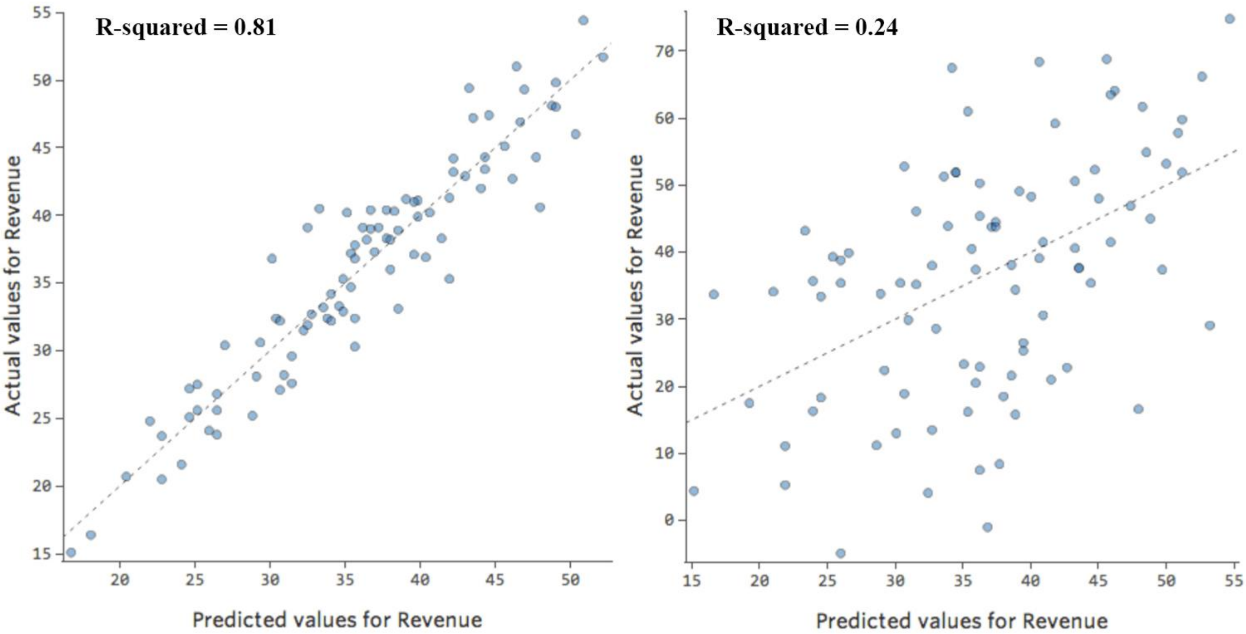


Figure 3.4 - Plot of a good regression prediction model vs a bad regression prediction model

However, it does not account for the overfitting problem. Because the model is overly sophisticated if the regression model includes numerous independent variables, it may fit extremely well to training data but perform poorly on testing data. That is why Adjusted R^2 is introduced; it penalizes the inclusion of new independent variables to the model and adjusts the measure to avoid overfitting difficulties.
R^2 is a convenient, apparently straightforward metric for determining how well your linear model fits a collection of facts. However, R^2 doesn't reveal the whole picture. It cannot identify whether the coefficient estimations and projections are skewed, which is why the residual plots must be evaluated: it does not indicate the suitability of a regression model. A good model can have a low R^2 value, while a model that does not match the data can have a high R^2 value. In summary, it estimates the strength of the association between your model and the response variable, but it does not give a formal hypothesis test for this relationship.

Through a measure called "Permutation Feature Importance" (Breiman, 2001), which shows the sensitivity of the model score to fluctuations in the values of the individual features, it is possible to evaluate what R^2 is reliant on each of the features (thus assessing a "score" of the features for the model). This model inspection approach may be applied to any estimator with tabular data that has already been trained. Essentially, the approach allows you to calculate the loss in model score (when compared to training) if the values from a

certain feature are cancelled arbitrarily. This process will provide a score for the lowest model achieved (e.g., the predictions with the data of a scrambled feature will be less exact), and the drop in the model score will represent an indication that measures how much that performance depends on a specified feature. In essence, the approach destroys the association between the feature values and the accompanying labels and analyses the resulting drop in the model's score. The approach may be used to regressors or classifiers, as well as to test and training data. When the relevance of features on test data differs significantly from what is obtained on training data, this is seen as model overfitting. The final key consideration is the clarification that technical outcomes are model-specific:
1) It is possible that features are unimportant in models with low scores but extremely significant in models with high scores (with equal estimators). It is usually preferable to base the evaluation of the Permutation Feature Importance on models derived from optimal parameters acquired using the cross-validation procedure.
2) Because the approach is model-specific, it is not essential to determine how much more or less high the inherent predictive value of a feature is, but rather how relevant the feature is for the given model.

Cross-validation, also noted CV, is a method that is used to select a model that does not rely too much on the initial training set. The different types are summed up below:
k-fold
- Training on k − 1 folds and assessment on the remaining one
- Generally k = 5 or 10
Leave-p-out
- Training on n − p observations and assessment on the p remaining ones
- Case p = 1 is called leave-one-out

The most often used approach is k-fold cross-validation, which divides the training data into k folds in order to verify the model on one-fold while training the model on the remaining k folds, all k times. The error is then averaged over the k folds and referred to as the cross-validation error.

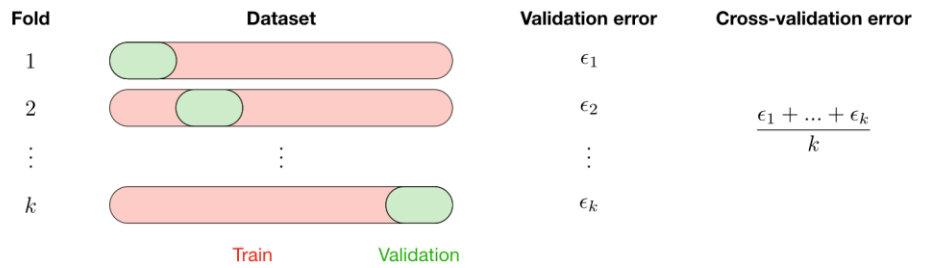


Figure 3.5 - k-fold Cross-Validation

There are many additional metrics that may be used in practice today, but a comprehensive treatment of them would be too time consuming and, in any case, would be inconsistent with the topic of the doctoral thesis and with the case study that will be discussed in the next paragraph. It is recommended a careful reading of for all other procedures in (Pedregosa et al., 2011)

4. COMPANY TEST CASE: MANUFACTURING-SUSTAINABILITY ASSESSMENT

4.1. MASTER ITALY S.R.L.

Master Italy s.r.l, is an Italian SME, that designs, produces and sells all of the small accessories for window and door frames (Mster Italy, 2019), with a research, investment, and study of aluminium cultural process, and reserving a profound Attention to the quality of materials, innovative technologies (by 97% of the value in house), identifying new market demands (national and worldwide), and customer satisfaction and care (on Time Delivery 95%).
Since 2013, Master has chosen to embark on a continuous improvement program with the goal of ensuring customer satisfaction and managing the rising company complexity.
The logics at the base of the program master improvement are those of Lean Thinking, that is, the battle against waste through process simplification, staff participation, and a drive toward the construction of synchronous process flows. To support the activities, enabling technologies, such as integrated automations, machines, and systems, are being used, allowing for the maintenance of proper working conditions and the analysis of real-time performance. As a result, the groundwork is laid for challengers' ever-increasing improvement activities. After 5 years of implementing Lean logics, the company capitalized on its experiences by developing the Master Italy Program System, which is a dynamic collection of lean techniques and methods to be used in various operating areas, human resource development tools, and Best Practices to be inspired by new projects. The areas within which it is spaced include security and the environment, improvement and ongoing innovation, digital transformation, and competence development (Master Italy, 2018).
As an additional value, Master Italy has chosen to begin an activity of study of the production chain and reduction of the environmental effect of the goods in the domain of window components in 2011. This is the path that the Master Group has followed to position itself as a virtuous model of circular economy by thinking that a survey of environmental consequences is a must, demonstrating its commitment to the long-term growth of the firm. The monitoring of consumption and impacts, which enables ongoing action and improvement of its goods and processes, not only from a technological but also from an environmental standpoint, is therefore an acceptance of duty towards all stakeholders.
The industrial structure of the Master Italy is exceptionally varied, as it is distinguished by a large number of machines (72) split among around 15 manufacturing departments spread throughout three sheds and an administration office building, with a total area of 37.000 m^2 (see Figure 4.1). Furthermore, the company is distinguished by a vertically integrated organization, which results in a diverse range of technological processes: aluminium die-casting, zamak die-casting, plastic molding, steel shearing, aluminium shearing, shot blasting, washing and tumbling, painting, drilling and threading, automatic and manual assemblies. The supply chain may be separated into two macroblocks: 1. The first is for raw material processing (aluminium die-casting, zamak die-casting, plastic molding, steel shearing, aluminium shearing, shot blasting, washing and tumbling, drilling and threading), which is distinguished by a high production capacity, large production batches, a large number of machines, high plant and equipment costs, and so on; 2. The second processing block consists of painting (and other outdoor surface finishes) and automatic and manual assemblies, which have distinct characteristics and organizational needs from the first block and are highly heterogeneous within it due to the large number of references and the high variability of production batches (from a few pieces to thousands).

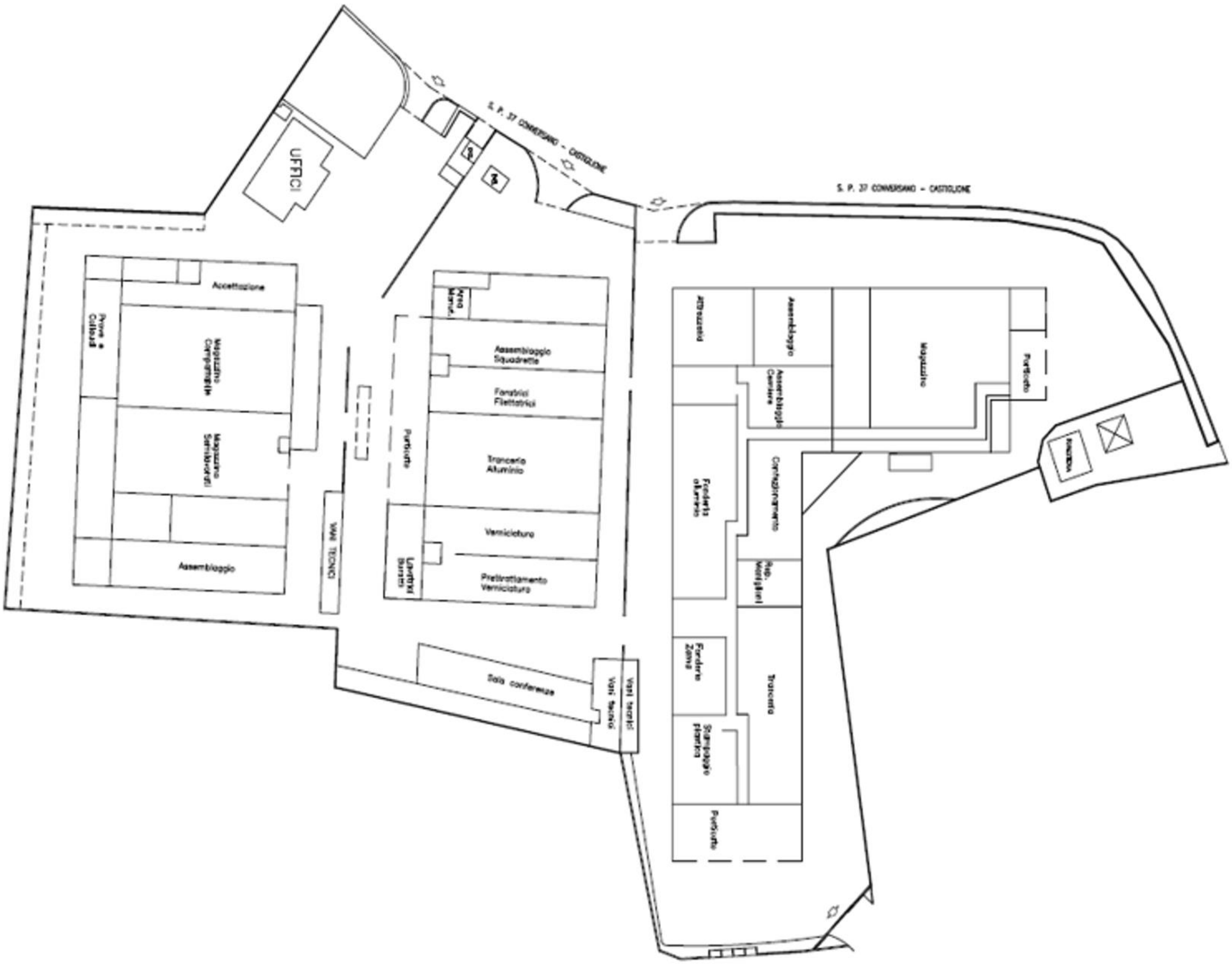


Figure 4.1 - Plant of Master Italy's spatial distribution

Since 2013, the company's Lean Manufacturing approach has enabled it to arrange processes and resources according to the "value stream" logic, especially regarding:

• The processes have been categorized according to homogenous product families, both in terms of final product families, material absorption, and comparable processes (belonging to both block 1 and block 2, with a view to the "flow";

• The key process management and measurement techniques (OEE, OLE, SMED, 5S, Kanban, Action Plan, Problem Solving, and so on...) have been implemented in various departments in order to monitor process efficiency and effectiveness and to introduce the logic of continuous improvement.

Due to the lack of technological solutions on the production lines, production data is managed using manual or low-tech tools, which are plagued by issues of timing (timely availability of data to support both low- and high-level decision making) and quality (reliability and quality of the data), as well as a high risk of operational error due to the large number of manual loading and management activities.

The issues surrounding data generation and storage included: the time-consuming nature of data collection, validation, and entry into the system, which is often done manually; the risk of data entry or transmission errors due to manual skills, and the fragmentation of data across multiple information mediums, including paper, which is not always integrated with the company's management system;

Data management, on the other hand, created the following issues:

1) The inability of the system to automatically update production data, which has apparent negative consequences for data dependability and correctness.

2) the inability of the Production Manager and value stream leaders to control machine activities and key productivity indicators of operators, machinery, and other production resources in real time;

3) the lack of a centralized corporate reporting system;

4) the lack of congruity analysis and data certification.
The virtualization of the organization necessitates the development and use of digital models known as digital twins, which reflect the collection of resources and knowledge of real-world operations. Because virtual replication of a physical system is quite complicated, a substantial quantity of data and models that reflect the operational semantics of the simulated pieces must be available.
A lack of knowledge of the genuine goal to be reached still persists. Companies in manufacturing that seek to analyse Big Data to improve the perception of what they are doing, for example, by being aware of the aim they want to attain, are still few. Master Italy is not one of them; it no longer wants to sell components; instead, it wants to sell a full service that includes AI ("Cafagna, Gruppo Master Italy Srl," 2018).
As a result, technical solutions (MES and scheduler - machine level production planning simulation software) were implemented in 2017 to optimize production process management and monitoring along the following intervention axes:
1. Automated production planning and scheduling
2. Real-time updates on the status of production orders and deviations from the scheduled one.
3. Elimination of errors in paper compilation and manual data entry.
4. The created pieces are automatically registered.
5. Manufacturing waste is automatically documented.
6. Automatic recording of active and idle machine times.
7. Machine status is automatically updated in real time.
8. Consistent resource monitoring throughout the production process.
9. Processing and monitoring of performance indicators in real time and automatically (e.g. OEE, Machine production capacity).
The thesis work is part of a larger and long-term project begun by the company in 2017 to report the phases of analysis concerning the implementation of the project called "Master Twin" which aims at an application of ML aimed at predicting, with a certain advance, the machine faults of a manufacturing plant. The digital twin of a product or process allows for the emulation, analysis, and simulation of the evolution of its actual twin, assisting in the prevention of issues or improving performance through real-time data analysis. It is built on three pillars (Abusohyon et al., 2021):
- Big data, IoT, and sensors: they enable continuous data interchange across goods, systems, and processes.
- Digitalization: the digital copy provides a virtual environment for evaluating ideal operating conditions, highlighting deviations, and testing remedial solutions.
- Intelligence: the digital twin provides tangible support for informed and timely strategic and operational decision-making.

The output of the prototype phase is thus represented by the demonstration on field of the applicability of Machine Learning technologies to the operational context of the Master company, and that the application of these technologies can descend a concrete benefit of greater plant efficiency, higher productivity, lower production waste, and so on. Because it was a purely exploratory area, the characteristics of the prototype version were represented by a limited operational context to a single machine, the use of historical data (not updated in real time), and access to external support for the component of artificial intelligence technology.
The end result is a computer application that, based on real-time readings of the machine's operating parameters, is able to visually notify the operator that the conditions for a detained machine are being generated, to indicate the peak of the machine where the malfunction is occurring (based on the parameters of the pattern found), and to suggest a preventive solution. It will be possible, using an evolved human-machine interface (HMI), to alert the employees in charge of supervising the plant and suggest interventions aimed at preventing the critical event that is being announced, with obvious benefits on plant efficiency and productivity.

4.1.1. Context

The “Master Twin” project had a breakthrough in 2020, with a development of a static prototype to test the applicability of artificial intelligence technologies to the context of company operations.

In 2019, thanks to investments made by the company in accordance with the Smart-Factory philosophy, a capillary field data acquisition layer that, using IoT technologies, conveys large databases collected along the entire Production line to a centralized system monitoring system will be available (SCADA). This massive amount of field data constitutes a massive informational legacy (Big Data) that can be profitably used for ML purposes in order to train algorithms capable of predicting events based on large databases gathered.

An application solution capable of bridging data from machine sensors on a single die-casting machine with machine event casualization from the production management system has been introduced. With this prototype, we want to demonstrate that, by applying ML methodologies to the data in question, artificial intelligence (AI) technology is capable of extracting, from the collected data, a series of patterns capable of predicting events whose correspondence has been found in the causals of critical events (microstops, production waste, faults) relating to the same machine on the MES system.

The Master Twin project is divided into the following phases:

1. Project definition: an awareness on the organization's goals, priorities, and resources in order to identify the goal that the predictive analysis must reach by calculating the costs of implementation and the associated timing. This step entails conducting research on the aluminium die-casting process, which will be followed by research on the other raw material departments and the painting department, as well as the subsequent application of the model on them, in order to make a technological change and manage production in a flexible and predictive manner.
2. Data preparation: it is required to collect all relevant data so that it may be utilized later in the analysis. The majority of Master's machines are outfitted with a PLC with basic control capabilities and a set of machine edge sensors capable of monitoring some crucial process parameters. The following operational steps are used to gather, historicize, and clean data:
    - Data Selection: choosing a data collection necessitates understanding of the domain from which the data is drawn. Removing irrelevant data from the data set minimizes the research space during the data mining phase, resulting in a shorter analysis time.
    - Data pre-processing: his phase involves cleansing the information and eliminating discrepancies that may create difficulties throughout the data analysis process. This phase involves the development of E-R (entity-relationship) models, which are conceptual models for the conceptual and graphical representation of data as well as the relationships between data databases.
    - Data Transformation: the data is translated into forms suited for data mining techniques analysis. During this phase, the diversity of data is reduced while the quality of the data is maintained.
3. Data analysis and exploration: the extraction of information from current data sets in order to create models and forecast future outcomes and/or trends. The best algorithms for analysis are defined at this stage. This stage entails evaluating and implementing machine sensors (e.g., machine learning algorithms or data mining techniques). The two primary roles of data mining are extraction and analysis. The approaches employed in the first extract implicit and tacit information to make it useable; in the second, the data is investigated and analysed mechanically or semi-automatically to uncover patterns and regularity in the behaviour being studied. These are referred to as patterns. A pattern is a synthetic and semantically rich representation of a set of data that encapsulates a recurring pattern in the data.
4. Digital model construction: it entails creating a very accurate virtual representation of the actual thing. The model collaborates with the appropriate physical system to undertake behaviour analysis, assessment, and prediction.

Users must develop the following models to generate a full mirror replica of the physical system:

1. The geometric model was built as a solid 3D model.

2. a physical model that simulates the physical features of a physical system (e.g., stress on the gear and temperature, working pressure, etc.).
3. The behavioural model depicts the behaviour of a physical system that is governed by driving factors (for example, production orders) or disturbing factors (e.g., human interference factors).
4. A rule model contains restrictions and relationships that enable assessment, optimization, and/or prediction.
5. Model Validation and Model Use: To validate the developed digital model, tests must be done. Statistical analysis, heuristic analysis, experimental analysis, and human analysis are some of the methodologies that may be utilized for evaluation.

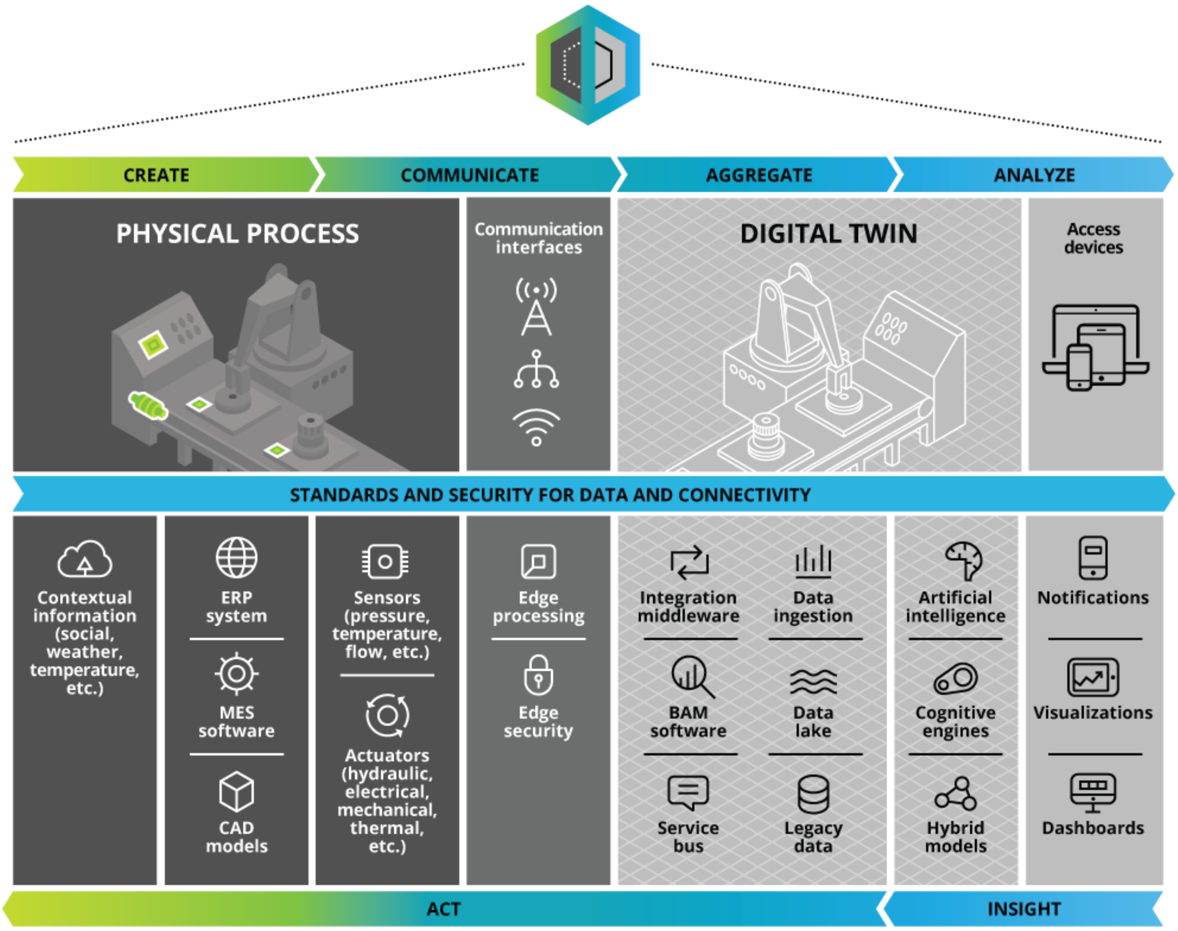


Figure 4.2 - Digital Twin conceptual architecture (Parrot and Warshaw, 2017)

As a result, the digital twin's ability is to emulate three human brain capacities: conceptualization, comparison, and collaboration. At the following stages, virtual transformation enables the following benefits, confirming what has been said by (Becue et al., 2020):

- Quality: reduced processing waste due to proper system control;
- Productivity: reduced setup times, failures, and machine downtime;
- Competitiveness: reduced testing times;
- Innovation: highly technological products and processes.
- Real-time communication between the digital twin and the real system.
- Digital twin implementation and utilization
- Documentation and interpretation of the preceding stages' outcomes It may be required to return to prior phases to refine or change the information obtained in response to the user's most pressing demands.

4.1.2. Die-casting Process

Die-casting is a process that involves melting metal pans in high-temperature furnaces, then pouring the liquid metal into a mold of the appropriate shape and allowing it to solidify. When the final solidified object, also known as a casting, is expelled from the mold, the cycle is complete.

The casting process is comprised of one cycle, from mold cleaning to casting, and it takes around 4–120 seconds, depending on the casting size. As a result, it has a high productivity, with the ability to manufacture 30–1000 castings per machine per hour; the casting duration is determined by the mold clamping force. Die-casting offers the most price competitiveness in mass production components since it is a high-productivity technique. The traditional die-casting method, on the other hand, has drawbacks such as a lack of pressure resistance owing to high internal gas content, surface flaws due to the breaking layer, losses in strength and airtightness due to internal shrinkage, and difficulties in undercut processing. A furnace, metal, die-casting machine, and die are all used in the die-casting process. The metal, which is usually a nonferrous alloy such as aluminium or zinc, is melted in the furnace and then injected into the dies using a die-casting machine.

Die-casting technology is quite complicated in practice, and it is critical to select the parameters of the HPDC process in order to obtain appropriate mechanical qualities and the desired performance for the made aluminium-alloy components.

operating parameters during the cycle (strokes of the mechanical parts C1, C2, and CC, execution times of the various phases of the T1, T2 cycle, speed V1, V2, pressures PS, PF, and PM, closing force FC, and measured sprue thickness in the SM mold).

One cycle of die-casting process can be divided into four distinct phases:

1. Melting: Aluminium enters as a solid and departs as a molten state. For injection, die-casting needs that aluminium be heated deep into its liquid phase. Aluminium has a melting point of 680-700 °C. The aluminium is transported to each die-casting machine once it has been melted and heated to the right temperature. Each die cast machine has its own holding furnace, which keeps the molten aluminium at the proper temperature as it waits to be used in the die cast machine.

2. Injection: Molten aluminium is pumped into the mold via a plunger (Figure 4.3). The molten metal is then moved into a chamber where it may be injected into the die after being kept at a constant temperature in the furnace. When a die cast machine is ready for its next cycle (die closed and ready for shot), an automated ladle pulls a predetermined volume of molten aluminium from the holding furnace and pours it into the mold. When the pouring is finished, the injection step begins. The first injection phase is the sluggish phase (T1), in which the plunger goes slowly forward (V1).

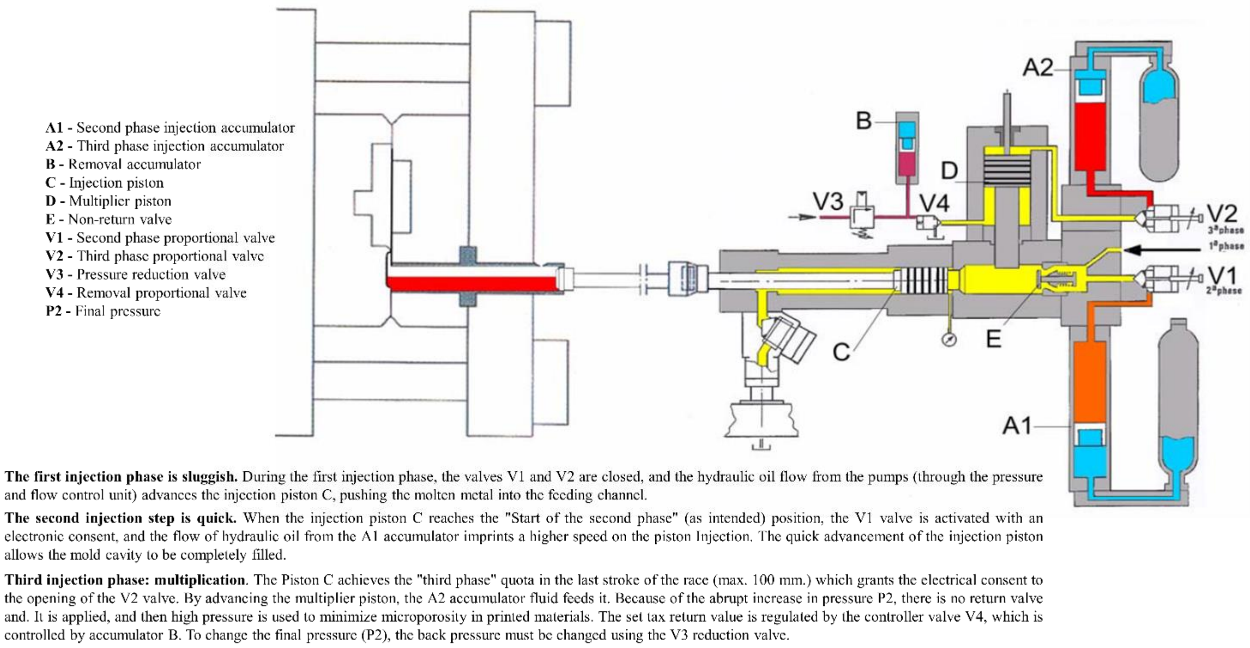


Figure 4.3 – Injection scheme

After a predetermined distance (C1), the plunger moves (C2) with an intermediate speed phase (V2), during which the speed is raised to fill the mold. When this is finished (T2), the machine enters a rapid phase in which the speed is considerably increased in order to fill the component cavity with aluminium (CC).

3. Molding: The molten aluminium hardens in the cavity of the mold. After the cavity has been filled and the plunger has stopped moving, the hydraulic cylinder that is pushing the plunger is inflated to a greater pressure (PM). During solidification, this pressure keeps the molten metal in the dies. The final shape of the casting is established when the whole cavity is filled and the liquid metal hardens. The die cannot be opened until after the cooling time has passed and the casting has hardened. To hold the die tightly closed while the metal is injected, clamping force (FC) must be provided. The die opens the ejector or moves half of the die after a certain length of time (TC).

4. Extraction: The casting is pushed out of the mold chamber by an ejection device. When the injection cycle is finished and the machine is fully open, the die cast is pushed out and the thickness of the die cast (SM) is regulated to avoid quality faults. At this point, the mold is lubricated and a new cycle can start over.

5. While the next cycle begins, the previous print has finished cooling and is moving towards the final steps of processing. Depending on the intricacy of the component, they might be more or less articulated.

6. The cleaning of the mold comes after the molding and cooling phases. This process entails removing the meshes. To remove the part of the jet that has hardened and remained connected to the product component throughout this step of the process. It is a variable procedure dependent on the size of the component to be separated from the castable channel: for little pieces (as is typically the case with zamak products), bending the break point is often sufficient to break the connection; however, for big parts, a specific cutting equipment is necessary, which can be automatic or manual / robotized, depending on the component in production.

A die cast is a representation of an injection cycle. There are 36 pieces in the die cast. This means that every 33 seconds, 36 components are created for each die-casting cycle.

The quality of a die-cast product is affected by a number of elements. It is mostly determined by the properties of the alloy materials, the casting parameters, and the design circumstances of the molds and components. In general, a die-cast product is built with component thickness, charge time, local overheating temperatures, and surface conditions in mind. To reduce casting defects and produce high-quality die-cast goods, the equipment's working conditions must be adjusted, and the size and position of the gate speed, runner, overflow, and ventilation must be appropriately addressed while constructing molds. Among them, casting procedures or casting circumstances have a greater impact on the mechanical characteristics of die-cast aluminium alloys than the alloys utilised. Several thermodynamic parameters impact on die-casting process. They increase the quality of die-cast items when appropriately determined and adjusted. Mold temperature, dose volume, slow and quick injections, injection pressure, set up pressure, chemical composition, and liquid metal temperature piston velocity, metal temperature, filling time, and hydraulic pressure (Verran et al., 2008).

A die-casting process's environmental and technological performance is the consequence of a large number of criteria. Some of these settings may be changed, while others are random. On a daily basis, modern foundries may collect massive amounts of process data. These contain information about molten metal preparation, casting process parameters, simulation findings, component shape, Non-Destructive Evaluation (NDE) data, and so forth. In an industry firm, there are many various sources of information, such as data from the PLC system, MES, manufacturing line, and from outside the company. The ERP system understands what consumers want, the MES system understands how to create it, and data from production line sensors demonstrates how the production system functions. Enterprise data source systems are frequently produced by various suppliers and may not speak the same language. Data is acquired at several phases of the process. These data are held in silos, and their value is frequently restricted until they are combined. This would be a missed opportunity for the foundry unless there was a means to combine, fuse, and analyse the data.

Each die-casting process cycle is monitored and recorded on the cells during the operations. Each cycle record is made up of values relating to the input and output process parameters.

Over time, the types of data and collecting methods utilized by several departments within the same institution have developed. Methods range from high-tech automated uploading to a cloud database to handwritten notebook entries. As a result, merging the numerous sources into a coherent dataset presents issues because collection frequency and identities frequently differ. Communication among stakeholders is crucial throughout the process to determine whether, how, and how frequently data should be gathered to provide the best description of the system to be modelled. Integrated data is required for conducting machine learning, and it is a missed opportunity for the foundry business if no attempt is made to assemble, fuse, and analyse this data in order to better understand the process elements impacting casting quality.

4.1.2.1. Company die-casting process research challenges

1. Real-time thermodynamic process analysis. It is important to construct a network of sensor meets and a data integration system (IOT architecture) to gather all operations in real time and do automate thermodynamic analysis using a built-in ad-hoc algorithm.
2. Being able to incorporate an EA and LCA study into a single automated analysis.
3. Solve the challenge of heterogeneous data sources and formats by combining raw data sets from MES, PLC, and environmental data into a single structured database. The ability to acquire deep PLC data, MES data, and data from sensors, which monitor more particular objects that may not be available via the PLC or MES, is critical to high quality analytics and outcomes (Khan et al., 2017).
4. Managing an Unbalanced Dataset. Handling unbalanced data for machine learning is a major study topic, and various studies have been conducted to address the skewed data set issue, since many real-world datasets are severely imbalanced. Considering the causalisations reported on the MES, it is discovered that about 80% of the dataset influences the cause for microstops, 10% for mechanical faults, and 5% for both electrical faults and equipment faults, resulting in severely uneven data training.
5. Apply various anomaly detection approaches, like as machine learning, to find minority class samples that are oddities or outliers.
6. Fine-tune algorithm performance with additional data and verify models using previously unknown test data.

4.2. COUPLED EA-LCA FOR THE THERMODYNAMIC MODELLING

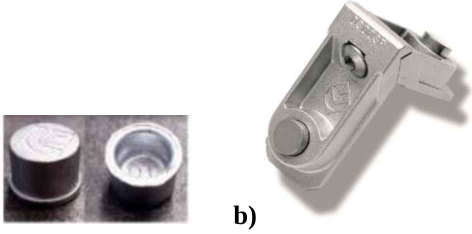
In this paragraph, an EA-LCA coupled analysis is performed on the company industrial test case, in accordance with what is stated in paragraph 2.1.4, thus following the System Thinking. Since the company's dedication to the long-term growth of the business for all stakeholders, it has been feasible to gather data formatted according to our needs and undertake direct measurements on the field of the process under investigation. LCA, as the name implies, is commonly used to assess a product's complete life cycle in a so-called "cradle-to-grave" assessment. LCAs are sufficient for analysing environmental impacts over the whole product life cycle. LCAs, on the other hand, do not have a broad influence on a certain stage of life. Because just a few customers have complete influence over a product over its entire life cycle, thorough reviews are vital. The application will allow you to clarify the systematic approach outlined in the preceding paragraph in further depth and identify any gaps that remain unbridged. This case study is typical since the whole process may be characterized using alternative beginning assumptions; also, it is composed of many sub-processes with distinct flows, increasing the risk of meeting heterogeneities.

STEP 1 - Goal and scope definition

The goal and scope derive from the company's requirements in quest of product system optimization: to decrease scraps created, to limit the environmental effect in terms of CO_2 in order to get the EPD declaration on the company's products ("EPD International," 2020), and to minimize energy consumption. To achieve these objectives, the LCA and EA will be built individually, followed by a hybrid EA-LCA. The study is performed on a single component (a safety pin, manufactured by die-casting zamak) and then on the full manufacturing chain of the finished product ready for sale (a handle). In terms of research, the purpose is to demonstrate the differences in the level of complexity of the analyses done on a particular process vs the complete production chain. As a result, the amount of heterogeneity is more or less significant. The two analyses will be conducted in parallel following the system thinking.

SUB-STEP 1.1 - Functional unit

It is simple to pick the functional unit for both the LCA study and the EA based on the company's performance requirements. The functional unit of the product system under examination in the current case study was found in one piece of safety pin with a diameter d = 9.6 mm and weight m = 20.1 gr, which is the same for all scenarios. While the functional unit chosen for the assessment of the entire manufacturing process is the steel corner square (whose safety pin is a component) with dimensions L x P x H = 24 mm x 14 mm x 24 mm and weight of m = 121 gr. This option normalizes the input and output data for the reference flows, proportional to the functional unit of the investigated product system. The functional units (depicted in Figure 4.4) provide a structure for the standardization (in a computational sense) of input and output data on the basis of which the performance of the analyzed process may be defined.



a) **b)**

Figure 4.4 - Functional Units. a) the safety pin; b) the steel corner square

SUB-STEP 1.2 - Reference flow, scenario and system boundaries

The recommended initial step is to create a process flow diagram depicting the relationship between the system unit-processes, as illustrated in Figure 4.5. According to the system thinking, the benefit of studying the process as a first step is the identification of its main phases, as well as the definition of its primary and secondary elements to be considered in the analysis, the estimation of the total cycle time, and the apparent criticalities that may arise during the analysis. As it corresponds to a distinct technological cycle, the schematization becomes objective, eliminating heterogeneity.

Third-party suppliers' zinc alloy (zamak) (“UNI EN 1774,” 1999) panels (or ingots) are unloaded using an electric forklift in a dedicated zone of the production plant. The individual panels are then placed into the furnace. The alloy is melted in an electric oven at a temperature of 400-420 °C. Molten metal is pumped into steel molds during die-casting. After that, the molten material is squeezed into steel molds. To cool melting systems and trim, closed cycle water is employed. While reusable scraps are reintroduced upstream of the furnace, completed items are retrieved, put into boxes, and transferred to the assembly section. The die-casting process is characterized as a basic sequence of activities indicated in Figure 4.5: melting (furnace), die-casting (injection and die phases), cooling and trimming, and final extraction for machining.

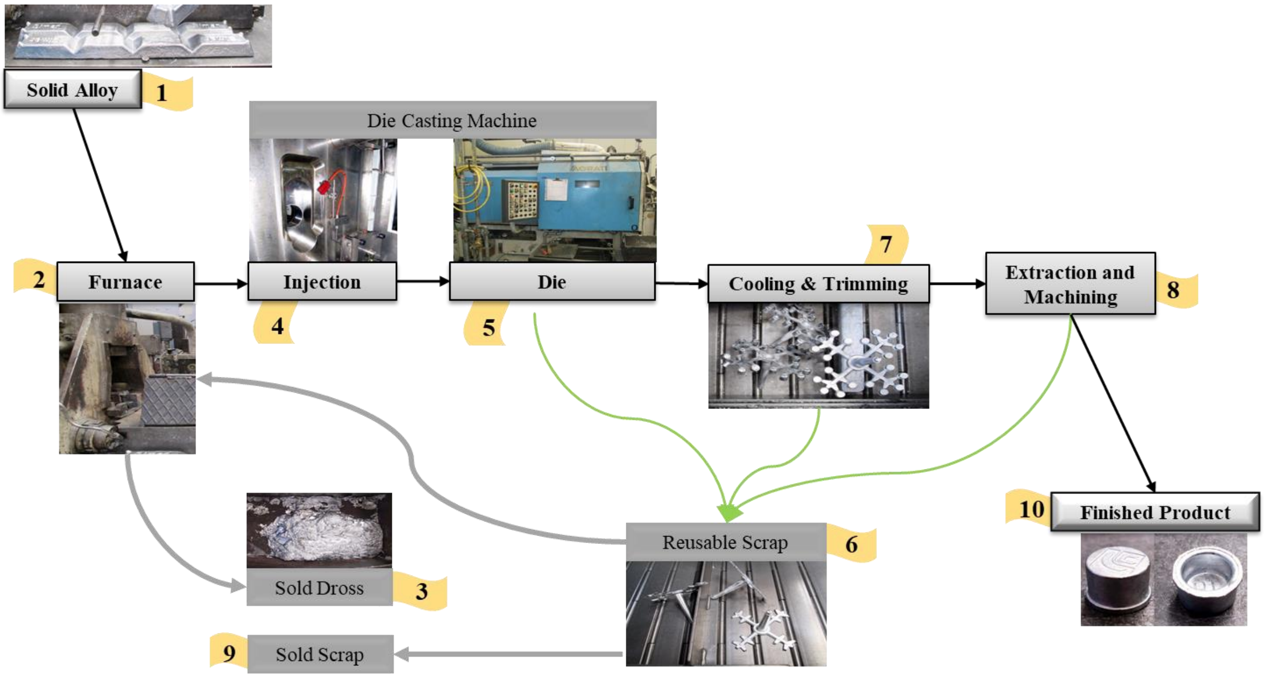


Figure 4.5 - Die-casting sub-processes (or sub-systems according to the system thinking)

While, concerning the analysis relating to the steel corner square, one final product, 1 pc, is made up of six components (see Figure 4.6**Errore. L'origine riferimento non è stata trovata.** for the overall manufacturing process flow chart):

- a safety pin produced by the zamak die-casting and vibro-tumbling processes. Die-casting zamak is a method of producing zamak (zinc alloy) items by injecting molten metal into steel molds. In an electric oven, the alloy is fused at a temperature of 400-420 °C. The molten substance is then injected and squeezed into steel molds. Following that, the printed material is discharged into boxes and allowed to cool to room temperature. The sprues are reintroduced into the furnace prior to the melting process. The semi-finished product is next transferred to the vibro-tumbling step, which is a mechanical scouring of metal surfaces that also allows the piece's sharp edges to be removed. The procedure is carried out by immersing the parts to be treated in a heterogeneous mass of moving granules or spheres that perform metal removal / sanding by sliding along the surfaces of the component.
- a spring block plate produced by stainless steel shear presses, washing, and vibro-tumbling processes Stainless steel coils are physically treated with additives in the metal pressing and sheathing section utilizing eccentric presses for shearing machining supplied with steel pitch molds. The steel belt is then forced into the mold, where it is molded and cut to fit the plate to be created. Dirty components are delivered to an industrial washing machine with a centrifuge. The procedure is intended to remove

processing residues (oil, pastes, fats, dust, etc.) from metal semi-finished goods. The washing system operates by dissolving appropriate detergents in hot aqueous washing solutions maintained at roughly 70 °C.

- a female wing produced by die-casting and shot blasting, and a male wing produced by die-casting, shot blasting, drilling, and threading. In order to make aluminium alloy components, molten metal is injected into steel molds during the die-casting process. The raw material for aluminium alloy loaves is delivered within melting furnaces and heated at the melting temperature of the alloy (660-700 °C). The molten substance is then injected and squeezed into steel molds. Robotic arms are then used to place the printed material in boxes and bring it to room temperature. Shot blasting is carried out by pushing steel metal balls with a diameter of 0.5 mm at high speed against the objects to be treated, eliminating any leftover burrs created by molding. The item is then transported to the department of drilling and threading. The section is comprised of a range of machine tools that can work dry as well as with lubricants and chemicals. In the latter situation, fully automated machines run a continuous cycle in a closed cab, recycling the emulsion after filtering.
- third-party contractors are used to purchase springs and screws.

Following that are the semi-automatic assembly and hand packing (with labels and cardboard) steps.
To make the findings of LCA and exergetic analysis more similar in terms of indicators, the stages of transport (typical of LCA), assembly, and packing will be excluded from the study. This decreases the amount of variation in process definition, inventory, and computations. Finally, the purpose and scope of the two independent studies are the same, the system boundaries are the same, and the inventory analysis is performed on the same flows of matter and energy.

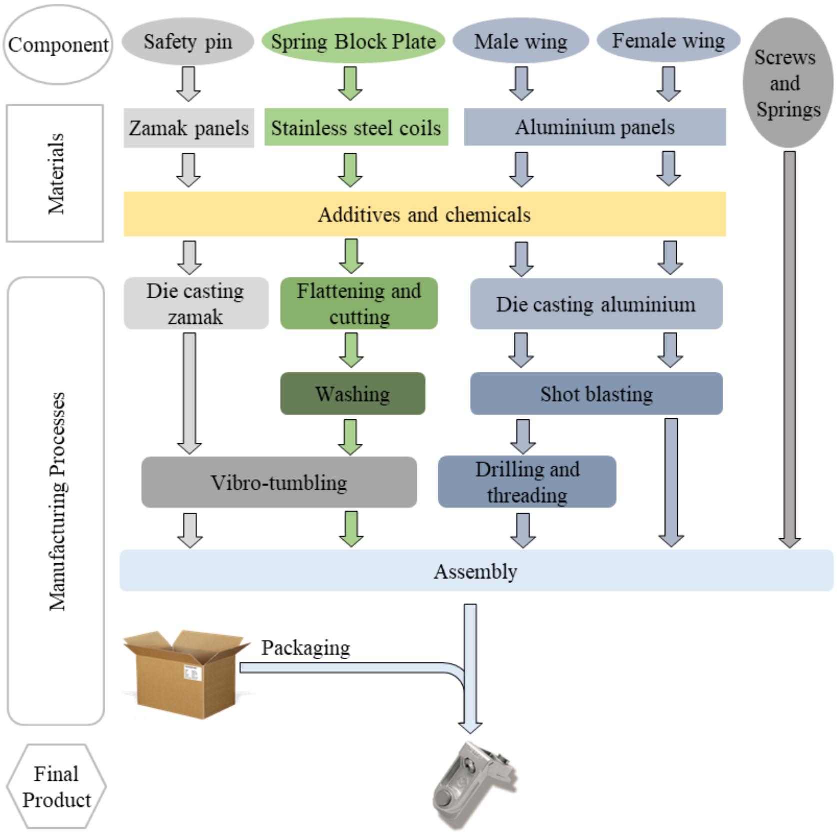
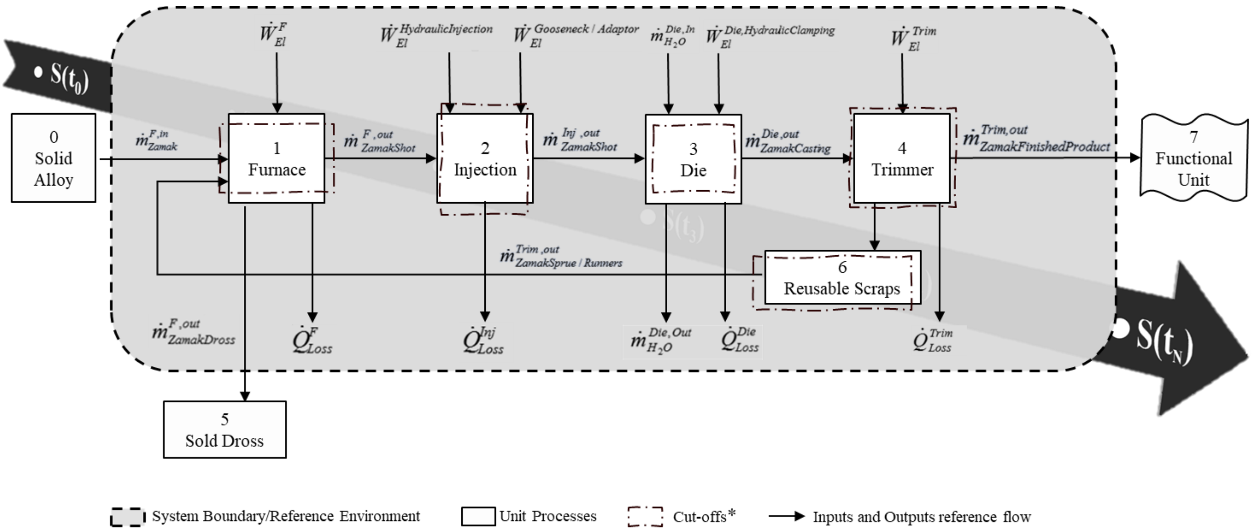


Figure 4.6 - A flow chart depicting the manufacturing processes of corner square

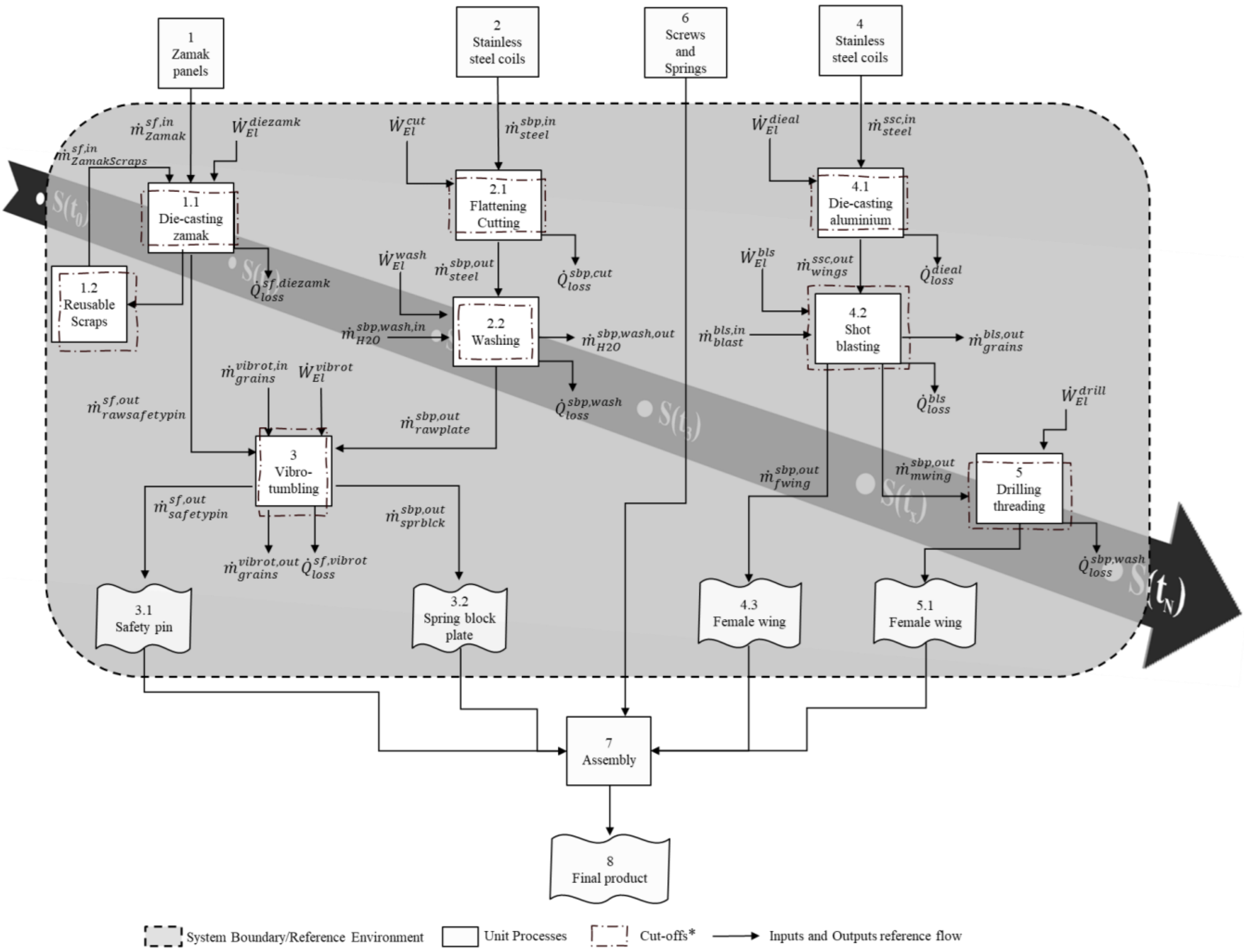
The scenario analysis is linked to the sensitivity analysis in this example, which is a function of the possible alterations in the percentage allocation of energy consumption during the production phases. EA substantially assists the quest for technical or operational process improvements. In this situation, EA aids in the definition of a single scenario, eliminating potential heterogeneities by identifying sub-systems (unit-processes)

whenever a state change happens. Concerning the reference flows, non-compliant items were included in the analysis as long as the overall entering raw materials and their consequences were included. Waste treatment methods were also provided. The EA uniquely identifies reference flows in its balance equations (see paragraph 2.1.2), eliminating analytical heterogeneity.
The system boundaries were chosen based on the analysis's goal and scope. The defined boundaries are as follows: i) a temporal boundary (2020); ii) a geographical boundary (the main production setting place is Conversano - Italy); and iii) a technological boundary, which includes all in-plant production processes necessary to manufacture the two functional units, excluding raw material production and transportation (buy) and packaging (made in another company's department). Due to a lack of precise data, only the manufacturing phase has been included in the system boundaries for the hybrid analysis done here.
Because EA was used as a guideline to identify the unique intermediate path from $S(t_0)$ to $S(t_N)$, heterogeneity was severely reduced. The system thinking aims to simplify the search for different production of $S(t_x)$.
The optimization criteria in the EA is a guidance to this goal by decreasing the term Ex_{loss}, which is the source of the system's less-than-theoretical efficiency. Heterogeneity is decreased in EA and LCA by describing our industrial system as a thermodynamic model. It is probable that LCA is an estimate of the distance between the end state $S(t_N)$ stated in terms of environmental effect rather than thermodynamic potential, and that EA is a thermodynamic approach that also helps to measure process inefficiencies. The thermodynamic model is built of four unit-processes, each of which may be regarded as a separate thermodynamic system (more control volumes) with associated mass and energy input and output fluxes determined by the functional unit and the system boundaries/system environment. Figure 4.7 and Figure 4.8 depict a graphical depiction of the two system models. The models were built on the basis of the "ontological reference model" approach used by (Cao et al., 2018), which states that when there is a variation from the previous state, the system must be divided into sub-processes; this means having a technological view of the production cycle and the product.



* Cut-offs in this scheme have only graphic purposes and are not accurately proportionate to the real assumptions made in the case study. They qualitatively indicate some process parameters not monitored in the present analysis.

Figure 4.7 - The model of the die-casting process according to the system thinking



* Cut-offs in this scheme have only graphic purposes and are not accurately proportionate to the real assumptions made in the case study. They qualitatively indicate some process parameters not monitored in the present analysis.

Figure 4.8 - The model of the entire manufacturing process of the steel corner square according to the system thinking

STEP 2 - Inventory analysis

It is known which data must be inventoried now that the goal and scope specification process has concluded. The physical characteristics of raw materials and secondary materials (zinc, water) may be discovered in technical standards datasheets. The amount of processed materials intake and output, as well as their relative temperatures, have been monitored throughout time. Partial and whole-time cycles have also been tracked. Over the course of 2020, the electrical power consumption by machines was measured and proportionately assigned to the functional unit.

SUB-STEP 2.1 - Choice of the database

Equivalent items must be discovered and selected from the appropriate database for all data linked to the functional unit. Because the availability of data related to each specific database was limited, as was the number of items recorded (Table 2.1) and the poor geographical specification that these datasets provide, the choice for the most suitable database for our case study fell solely on Ecoinvent v.2.2 uploaded on the SimaPro® v.7.3 software. Concerning EA, the chemical exergies associated to zinc and water were gathered from the most recent list of chemical exergies (Szargut, 1989; Rivero and Garfias, 2006).

SUB-STEP 2.2 - Allocation

Allocation criteria for energy consumption were assessed based on the nominal power of each machine, rather than machining costs (economic allocation). As a result, an attributional LCA was performed using average data on power usage. The same allocation criteria were used for EA. To address the issue of average data, expanding the borders would be avoided since some information would be lost; instead, it would be a good practice to place an electrical meter on each machine.

SUB-STEP 2.3 - Local technical uniqueness

The topic of location in LCA encompasses more than just changes in geographical, topographic, and climatic geometry. Each ecosystem touched by resource exploitation or pollution is unique to some extent. As a result, a local ecosystem is particularly sensitive to the constraints imposed by a single product system's life cycle. Local distinctiveness is demonstrated by investigations into the influence of site-specific evidence and testing programs aimed at incorporating local sensitivities in LCA. According to the system thinking, data relating to the geographical position have been picked in the database: primarily data related to Italy, if accessible, or average data connected to European territory in the remaining situations. In contrast to LCA, EA is not susceptible to geographical requirements that might alter the technical application of this technique of analysis, limiting the potential heterogeneities that may develop when picking data based on local specifications in LCA. The geographical location of the system/process under investigation has no direct influence on the exergetic analysis since what counts is the temperature differential between the material fluxes in, out, and with regard to the dead state. It signifies that the ability to create productive work through the absorption of power remains the same. When varied ambient conditions are considered, it will take more time and energy to melt the zamak panels, although this is not a reason that directly impacts exergy, in terms of spatial-local specification, even a strong conditioning system in the plant can degrade the efficiency of the foundry.

In terms of the inventory connected to the safety pin, Table 4.1 shows the primary flows of materials, energy, and wastes for each sub-process of the die-casting zamak. These figures have previously been computed in terms of the functional unit, which is one pc of safety pin.

Table 4.1 - Main process parameters, i.e., process flows of safety pin manufacturing

Sub-process	Materials		Electrical Energy		Wastes	
	Type	Quantity	Non-renewable	Photovoltaic	Type	Quantity
Furnace	Zamak panels + reusable scraps	0.0014 kg	0.00035 kWh	0.00003 kWh	Metal dross	0.00003 kg
	Additives/Chemicals	0.0001 kg			VOC	0.00007 kg
Injection	Zamak shot	0.0248 kg	0.00017 kWh	0.00002 kWh	Metal	0.00003 kg
	Additives/Chemicals	0.00005 kg			VOC	0.00002 kg
Die	Zamak shot	0.0011 kg	0.00054 kWh	0.00005 kWh	Metal	0.00003 kg
	Additives/Chemicals	0.00035 kg			Oil	0.0001 kg
Trimming	Zamak casting	0.0013 kg	0.00014 kWh	0.00001 kWh	Dust	0.00001 kg
					Metal Scraps	0.00003 kg

In terms of the inventory connected to the steel corner assessment, Table 4.2 shows the primary flows of materials, energy, and wastes for each sub-process. These figures have previously been computed in terms of the functional unit, which is one pc of steel corner.

Table 4.2 - Main process parameters, i.e., process flows of steel corner square manufacturing

Sub-process	Materials		Electrical Energy		Wastes	
	Type	Quantity	Non-renewable	Photovoltaic	Type	Quantity
Die-casting zamak	Zamak panels	0.0014 kg	0.0012 kWh	0.00011 kWh	Metal	0.0001 kg
	Additives/Chemicals	0.0005 kg			VOC	0.00011 kg
Die-casting aluminium	Aluminium panels	0.0248 kg	0.00067 kWh	0.00003 kWh	Metal	0.00048 kg
	Natural gas	0.012 m^3			VOC	0.00002 kg
	Additives/Chemicals	0.001 kg			Oil mist	0.0009 kg

Flattening and cutting	Stainless steel coils	0.0011 kg	0.00082 kWh	0.00008 kWh	Metal	0.00005 kg
	Additives/Chemicals	0.0003 kg			Oil	0.0001 kg
Washing	Water	0.0976 l	0.0022 kWh	0.00021 kWh	Sludge	0.095 l
	Natural gas	0.0018 m^3			Formaldehyde	0.000001 kg
	Additives/Chemicals	0.0003 kg			Chemicals	0.0003 kg
Vibro-tumbling	Abrasive grains	0.0002 kg	0.00082 kWh	0.00008 kWh	Grains	0.0002 kg
Shot blasting	Abrasive blasting	0.0003 kg	0.0016 kWh	0.00014 kWh	Grains	0.0003 kg
Drilling and threading	Additives/Chemicals	0.0001 kg	0.000034 kWh	0.000006 kWh	Metal	0.0001 kg

STEP 3 - Life cycle impact assessment

SUB-STEP 3.1 - Impact category selection and cut-off rules

The environmental performance comprises data on resource usage, energy consumption, pollutant emissions across the product's life cycle, and possible environmental consequences in kgCO_2eq. The IPCC-GWP-100_y impact category was chosen, with a cut-off value of less than 1%. The firm's goal of obtaining EPD certification on their products, which demands exactly the environmental effect represented in kgCO_2eq, compelled the corporation to choose this impact category ("The International EPD® System - The International EPD® System," n.d.). The sole metric for EA is the amount of exergy loss represented in Joules, or dimensionless efficiency. Cut-off was simple in this situation since the EA previously set the tolerance of the performance parameter to +/- 1 [J]. This decision is solely based on the goal and scope previously defined.
Because the information acquired will aid in the anticipated effect assessment, the inventory review should be guided by the selection of impact assessment metrics. If the effect calculation is exergy research, the data acquired may be suitable to this, implying that gathering a large amount of pollution data should be overlooked since they are either damaging or greenhouse gases but do not contribute significantly to Exergy losses. Consider that EA alone is incapable of delivering full information regarding the system under study's environmental sustainability. Just when coupled with LCA and adding information pertaining to process performance, not only environmental, but also technical/technological, does exergetic analysis become a helpful instrument for the cause, as in Exergoenvironmental and Exergoeconomic analysis, or CExD. As a result of being more limiting in terms of system boundaries and flows to be evaluated than the LCA, exergy minimizes some sources of uncertainty and heterogeneity inherited in the LCA. When a broad-spectrum examination of probable environmental consequences is needed, the LCA impact categories, such as Recipe or CML-IA, are recommended. CExD is the best assessment method for integrating LCA and EA because it allows for the consideration of all upstream and downstream flows (as well as related primary and secondary data) that characterize the LCA to be expressed in a single more comprehensive metric, such as exergy demand (MJ).

SUB-STEP 3.1 - Space and time characterization

The bounds were chosen to address space characterization. In terms of time characterization, the system view only applies to the production stage, reducing heterogeneity caused by the limitation of different alternatives that would occur at each sub-step of the analysis, i.e., additional alternatives in functional units, system boundaries, and even more during LCI, where similar data in the datasets may lead to inaccuracies.

STEP 4 - Results and interpretation

ISO 14044 state that "The selection of impact categories shall reflect a comprehensive set of environmental issues related to the product system being studied, taking the goal and scope into consideration" ("ISO 14044," 2006). Indeed, even if it investigates the whole product system, a study that focuses primarily on one type of effect, such as carbon footprint or water footprint reports, is not considered a life cycle assessment in the meaning of the ISO standard. However, from a system standpoint, where the practitioner must answer to the

company's needs, which constitute a constraint, the purpose is realized by the measurement of CO_2eq emissions, which is the quantity necessary to receive the EPD certificate. Furthermore, as established at the start, the EA allows to reduce some of the inherent heterogeneities of the LCA technique while simultaneously providing us with relevant information on process and system performance. The general character of ISO standards is a limitation: more contextualization or a set of representative test cases would be preferred to avoid free interpretation.
As concerns the case study, the life cycle results, the emissions have been summarized per each component in the following Figure 4.9 illustrating the normalized overall impacts of one safety pin. The evaluation of the impacts for the GWP reported a total value of 0.0775 $kgCO_2$eq/pc. While the normalized overall impacts for the steel corner square are depicted in Figure 4.10, reporting a total value of 0.3186 $kgCO_2$eq/pc

	IPCC GWP 100a	
	Unit	kgCO2eq
Subsystem	1 - Furnace	0.02934
	2 - Injection	0.01511
	3 - Die	0.02207
	4 - Trimming	0.01100
	Total	**0.07752**

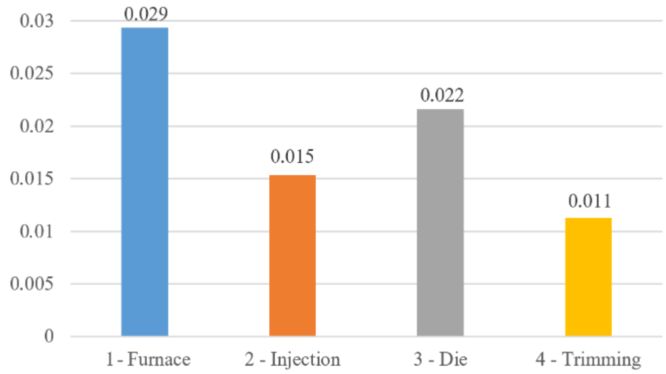


Figure 4.9 - Safety pin: results of IPCC GWP100y performed with LCA

	IPCC GWP 100y	
	Unit	kgCO2eq
Subsystem	Die casting zamak	0.0731
	Die casting aluminium	0.1674
	Flattening and cutting	0.0107
	Washing	0.0218
	Vibro-tumbling	0.0099
	Shot blasting	0.0321
	Drilling and threading	0.0036
	Total	**0.3186**

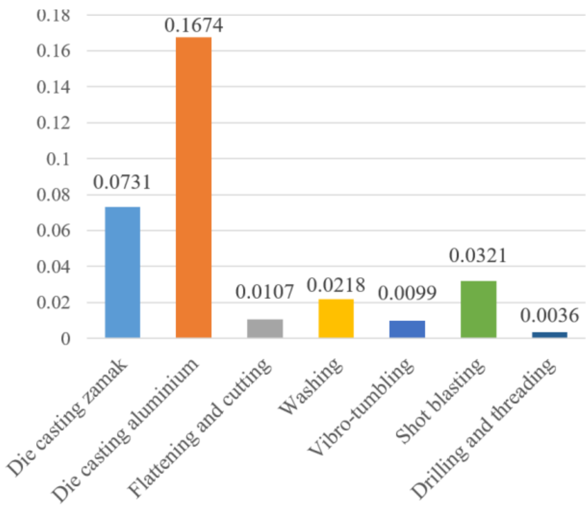


Figure 4.10 – Steel corner square: results of IPCC GWP100y performed with LCA

When the EA is performed from a systemic perspective, the total sustainability evaluation becomes less uncertain than the LCA. The retrofit work has been examined indirectly by taking into account the exergy loss flow in the Sankey diagram in Figure 4.11. Despite the unambiguous specification of all presumptions and assumptions made for both analyses, the greatest exergy loss happens in the 3-Die sub-process, which contradicts the conclusions obtained by using the LCA.

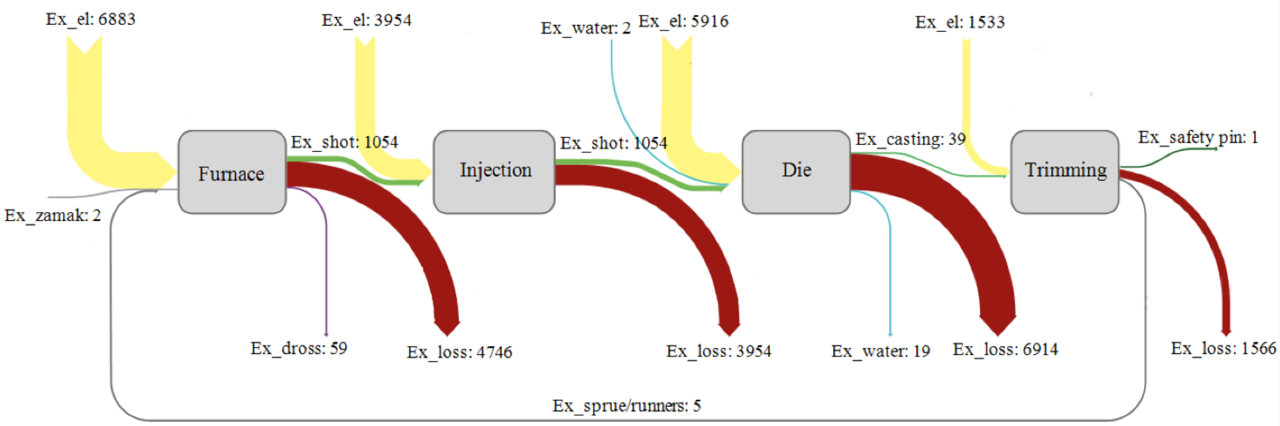


Figure 4.11 - Sankey diagram of exergy flows expressed in Joule

The following figures were produced from the EA for the sustainability performances based on the exergy efficiency indices determined for each sub-process, as shown in Figure 4.12. While the metrics for the complete manufacturing process of the steel corner square are presented in the next paragraph in Table 4.3

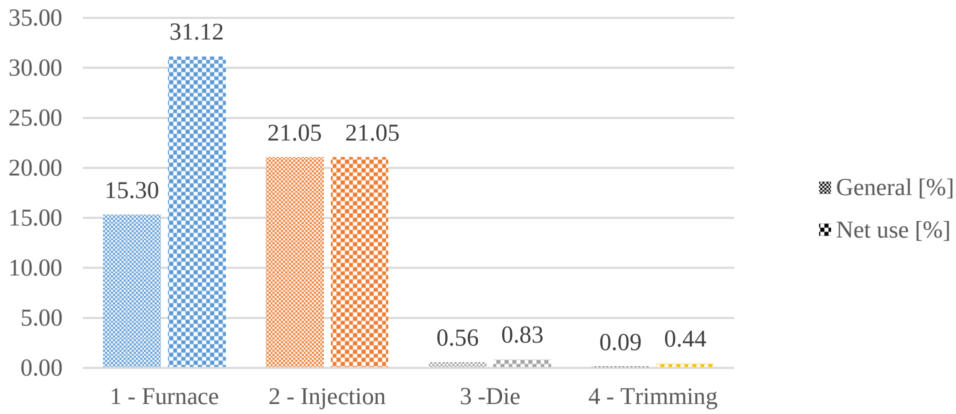


Figure 4.12- Exergy efficiencies for each sub-process

Because of the quantity of destroyed exergy, heat losses, and wastes, net use exergy efficiency differs from general exergy efficiency. The gap between these two efficiency indicators implies that the melting operation's exergy efficiency may be enhanced further, i.e., there is potential for alternate process scenarios $S(t_N)$.
As a reference, a more integrated EA-LCA was also performed in SimaPro®, integrating Exergy and LCA. The CExD approach was chosen as the analytical method. The goal of CExD is to calculate the total exergy extracted from nature by adding the exergies of all resources (both material and energy) necessary to supply a product or process. The evaluation occurs at several stages of the life cycle. This approach assesses the quality of energy demand in the investigated system by using exergy as a measure of the possible loss of valuable energy. The CExD assessment results, presented in Figure 4.13 and Figure 4.14, were consistent with independent LCA and EA evaluations.

CExD	
Unit	MJ
Subsystem	
1 - Furnace	9.50132
2 - Injection	4.13134
3 - Die	8.26744
4 - Trimming	3.25209
Total	**25.15219**

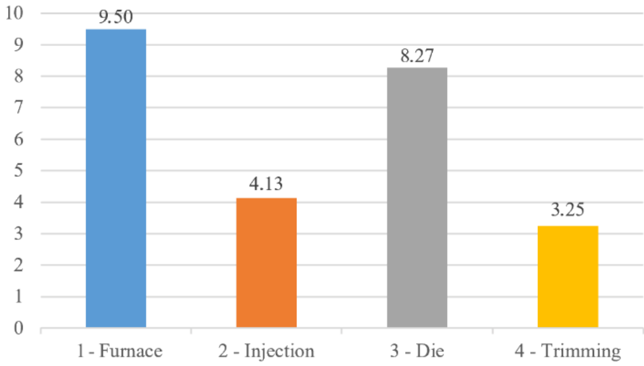


Figure 4.13 - Safety pin: results of CExD expressed in MJ

CExD	
Unit	MJ
Subsystem	
Die casting zamak	25.59
Die casting aluminium	43.31
Flattening and cutting	20.03
Washing	17.49
Vibro-tumbling	4.56
Shot blasting	6.15
Drilling and threading	7.18
Total	**124.31**

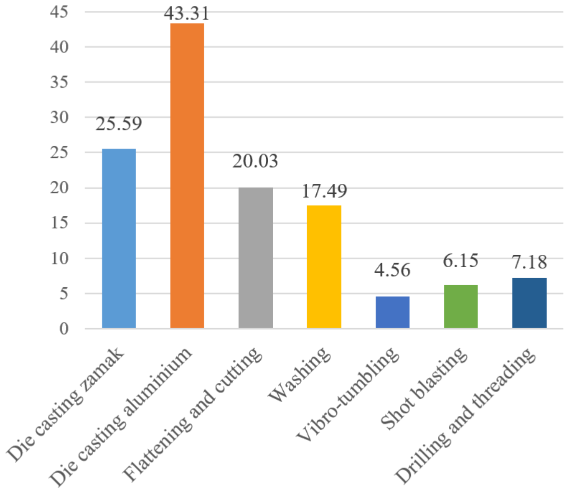


Figure 4.14 - Steel corner square: results of CExD expressed in MJ

Because they use quite different measures, the sustainability assessment results obtained through the LCA alone are not directly comparable to those obtained by the EA alone. The case study results utilizing the two methodologies independently are contradictory, because LCA deems the furnace to be the most influential sub-process, whilst the EA turns out to be die-casting. Because the aluminum endures a significant drop in temperatures in a relatively short period of time during die-casting, the value of the Carnot efficiency component $\left(1-\frac{T_0}{T_K}\right)\dot{Q}_e^i$ of Eq. 2.12 turns out to be greater in this sort of Exergy-based research. Assuming that the CExD was the appropriate compromise between LCA and EA, the following findings are consistent and predictable: SimaPro® software does CExD analysis, hence it relates to its method's characterisation factors (see (PRé, various authors, 2019) for a deep insight). CExD's findings, but as shown in Figure 4.13, the impacts of the die-casting sub-process are proportionately bigger than those of the LCA alone in Figure 4.9. This implies that exergy has helped to achieve a better balance between the significance of data flows and the emissions associated with them, as well as the environmental consequences.
The comparability of the analysis's results is then ensured. Another way to look at it, always in accordance with the system thinking, is to conceive of a production batch, i.e., one printed (a cycle) made up of 32 safety pins, as a functional unit (in Figure 4.5 is visible the whole printed extract before trimming). We may analyse the integral physics, the genuine physical quantities that are not distributed equally among the 32 pcs, in this approach, but information about the processing on the single piece may be lost.

4.2.1. Process Performance Metrics

The purpose of this paragraph is to offer an overview of the value of the most important indicators mentioned in the state of the art (see paragraph 3.1.2) concerning the case study.

The indicators will be calculated on the basis of the manufacturing process for the production of a corner square (which therefore represent the functional unit, as well as the finished product of the analysis). As a result, in this situation, the analysis is carried out not only on the die-casting process, but also on the full chain of processes required for the manufacturing and supply of the completed product and packaging. This process's information and assumptions have already been explained in the preceding paragraph 4.2.
The results of the analyses per unit of manufacturing process, all of which are connected to the functional unit, will be displayed as indicators of process performance in terms of sustainability and technological quality.
The indicators discussed below were computed using the equations provided in Table 3.1. As a result, the LCA and EA performed are useful in determining the parameters needed to calculate the indicators. The results are shown in Table 4.3.
LCA's GWP100y, which is widely used as a benchmark for obtaining EPD certifications for a sustainable product. The greenhouse gas potential (GWP) represents a greenhouse gas's contribution to the greenhouse effect in proportion to the CO_2 impact, which has a reference potential of 1. Each GWP value is determined for a specific time period of 100 years. The case study reported a total value of GWP of 0.3186 kgCO_2eq/pc net of assembly and packaging for the case study carried out with SimaPro® and Ecoinvent v.3 database. The meaning of this measure is to determine the quantity of CO_2 equivalent created for each sub-process. The more energy-intensive the process, the higher the quantity generated. As a result, according to $GWP100_y$, die-casting aluminium is the most energy-intensive subprocess (considering, however, that this process is called into question twice, for the production of both wings, male and female). Die-casting methods, in general, have the biggest environmental effect.
Exergy losses are irreversible poor uses of available energy, or squandered labour potential. This is also known as dissipated energy, and it may be decreased with the right retrofit solutions. The Ex_{loss} number has also been included in the table since it is commonly mistaken as a metric for evaluating which sub-process consumes the most energy and hence has the greatest opportunity for improvement in terms of technological quality and sustainability.
the overall Exloss of 1.315 MJ represents the sum of exergies lost in sub-processes. When the Ex_{loss} values of sub-processes are examined, it is clear that there are cases where the result is consistent with that expressed by $GWP100_y$, at least from a ranking standpoint, and other cases where a sub-process, such as washing, finds itself less energizing than vibro-tumbling, despite the GWP's assertion to the contrary. Ex_{loss} has repeatedly classified the two die-casting methods as less sustainable.
The coefficient of resource-use performance, η_p, is defined as exergy efficiency. It is a non-dimensional measure that may be stated in percentage form. In order to increase performance, exergy efficiency highlights the significance of measuring losses and internal irreversibilities. Higher exergy efficiency reflects more energy content employed in the system, making it more sustainable, whereas lower exergy efficiencies represent energy losses and internal irreversible processes, resulting in inferior energy quality and a worse sustainable rating. It is calculated by dividing the total exergy intake by the useable exergy output. The overall efficiency of the process under consideration is 50.70%, placing steel corner production on a medium sustainable path. As a consequence of pure exergetic analysis, the η_p findings are totally compatible with Ex_{loss}, indicating die-casting operations as the most energy-intensive sub-processes. Because η_p in this situation is not cumulative, the process efficiency is represented by the value 12.34% of the die-casting aluminium, independent of the number of pieces produced. It has a little better efficiency than zamak, which contradicts the $GWP100_y$.
The Exergetic Eco-Efficiency, η_{eco}, is a measure used to compare two identical processes. It is worried with the possible effect differential between exergy generated from renewable sources and exergy created from non-renewable sources. Thus, the more specific the definition of which streams in the process come from renewable resources and which come from non-renewable resources, as opposed to the intelligent use of recyclable materials, the more accurate this indicator becomes. The most notable difference in the case study in question is the amount of electricity absorbed by the machines in the various sub-processes, which is generated for approximately 9% by photovoltaics and the remainder purchased from grids, consisting of 20% coal, 1.1% oil,

61.2% natural gas, 5.1%nuclear, 8.7% renewable, and the remainder from a combination of sources. In light of the foregoing, and depending on the source of raw materials, the overall process has an average eco-efficiency of about 0.56, which is not directly comparable to simple energy efficiency, but when compared to the hierarchy of sub-processes, it is very consistent with what the GWP has expressed. Furthermore, in terms of environmental effect, the two die-casting operations are the worst sub-processes.
The Life Cycle Irreversibility Index, χ, complementary of the Life Cycle Quality Index ψ, supports to the comparison of processes and products having the same functional unit. In contrast to the previous indicator, this one emphasizes the value of useful exergy generated throughout the life cycle, as well as the recycling potential of waste materials, and thus the exergy that can be recovered rather than that which is completely lost. The Life Cycle irreversibility index considers the exergy inefficiency, however if a real system is compared to the latest technological innovations, or to an ideal Carnot machine, an index that measures technology obsolescence may be implemented. In terms of the case study, the Life Cycle Irreversibility Index verifies what the other indicators have already said: die-casting procedures are the most impactful, and in this case, the most irreversible. It should be highlighted that the majority of sub-processes are reasonably valuable. This is owing to the fact that all genuine processes are irreversible, especially those involving abrupt temperature changes, state changes, or large waste of auxiliary material that does not contribute to the increase in useable exergy created. Because there are no substantial temperature variations or material waste, procedures with reduced irreversibility include vibro-tumbling, drilling, and threading. On average, the whole production cycle of the steel corner component is 60% irreversible. The Renewability Factor is calculated by dividing the cumulative exergy demand of renewable resources by the cumulative exergy demand of non-renewable resources. This metric is equivalent to the η_{eco}, but since it is calculated entirely on SimaPro using the hybrid CExD method, it eliminates the uncertainty that may arise when integrating the LCA and EA only at the end, rather than from the beginning. In the case study, FR also confirms that die-casting processes are the least sustainable, with a low renewability factor. However, it is inconsistent with η_{eco} on the other sub-processes as well. On average, the overall production process has a low renewability factor of 0.348, where 1 is the optimum process.

Table 4.3 - List of the main indicators' results for each sub-process

Sub-process	**Metric**					
	$GWP100_y$ [$kgCO2_{eq}$]	Ex_{loss} [MJ]	η_p [%]	η_{eco} [-]	χ [-]	FR [-]
Die-casting zamak	0.0731	0.217	11.2	0.137	0.92	0.149
Die-casting aluminium	0.1674	0.445	12.34	0.114	0.89	0.229
Flattening and cutting	0.0107	0.133	47.89	0.564	0.74	0.46
Washing	0.0218	0.085	79.08	0.721	0.65	0.355
Vibro-tumbling	0.0099	0.178	52.47	0.821	0.24	0.371
Shot blasting	0.0321	0.186	63.28	0.68	0.58	0.431
Drilling and threading	0.0036	0.071	88.65	0.873	0.17	0.444
Overall	0.3186	1.315	50.70*	0.559*	0.60*	0.348*

*average between the values of each sub-process

In conclusion, while all metrics agree that the zamak and aluminium die-casting processes are the most energy-intensive and least sustainable, when compared to other sub-processes, this agreement is not clear.
Because the inventory created during the LCA, which is done with SimaPro, is made up of background data that already has certain pre-set processing, the findings of the LCA and EA analyses cannot be directly compared. Here, we're talking about metrics and orders of magnitude that are so dissimilar that SimaPro®'s CExD values and the findings of pure exergetic analysis do not converge to equivalent conclusions. All of this raises the level of uncertainty in interpreting the data for effective consumption reduction and process improvement measures. As a consequence, it would be reasonable to compare consistent measures in terms of

inventory and process type with one another. Indicators obtained from combined EA-LCA evaluations may appear to be a more consistent solution as an outcome.

4.3. DIE-CASTING PROCESS THERMODYNAMIC MODEL

These studies conducted in the preceding paragraph help as a reference point for determining which factors should be regulated and where to begin constructing one's approach. In reality, the die-cast model was required to emphasize process parameters, their physical linkages, and the type of source to be utilized for control, monitoring, and validation. The Table 4.4 below depicts an overview from the Excel model, which will serve as the foundation for the later design phase of the measuring and monitoring system, as well as the automation of the real time EA.

Master Italy is an extremely dynamic and constantly developing SME. Although it has management platforms for production and quality control, the human influence (of operators) on all processes is still very strong. The continuous improvement of processes through targeted interventions identified by EA and LCA analyses, as well as the company's management system platforms, attempts to minimize this aspect, which is still not easily measurable in terms of quality loss and energetic effort (we can call it "anthropic entropy"). This is supported by the fact that there is currently no technique of objective and accurate study in the literature that completely characterizes the contribution of the social dimension within the paradigm of both sustainability and quality of processes and/or finished products sold by the company.

Table 4.4 - Overview of the process's Excel thermodynamic model, including the sources from which to monitor each parameter

MELTING

Process parameter	Symbol	m.u.	Source
Solid state temperature of ingots	T_s	°C	Thermo cam
Melting temperature of the aluminium alloy	T_f	°C	Datasheet
Aluminium alloy casting temperature	T_{al}	°C	PLC
Furnace capacity volume	V_{al}	m³	PLC
Density of the alloy	ρ	kg/dm³	Datasheet
Specific heat of the alloy	cs	cal/g°C	Datasheet
Fusion latent heat	cl	cal/g	Datasheet
Furnace temperature	$T_{furnace}$	°C	PLC
Methane gas	M_{gas}	m³	PLC
Room pressure	P_0	Pa	Air quality sensor

Simulation	Symbol	m.u.	Check	Source
Quantity of heat for melting	Q_m	J	$Q_m = p \cdot V_{al} \cdot cs \cdot (T_f - T_s)$	Pyrometer
Quantity of heat for transformation from solid to liquid	Q_l	J	$Q_l = p \cdot Val \cdot cl$	Pyrometer
Quantity of heat for casting	Q_c	J	$Q_c = p \cdot V_{al} \cdot cs \cdot (T_{al} - T_f)$	Pyrometer
Total amount of heat	Q_t	J	$Q_t = Q_m + Q_l + Q_c$	Pyrometer

CASTING

Process parameter	Symbol	m.u.	Source
Alloy volume to be injected	V_g	m³	Type of product to be manufactured
Alloy volume deposited on the mould by friction	V_t	m³	CCD Camera
Mould temperature	T_{cup}	°C	Thermocam
Container temperature	T_{cont}	°C	Thermocam
Cup inclination	%	%	PLC
Cast translation speed	v_t	mm/msec	PLC
Pouring speed	v_v	mm/msec	PLC
Machine closing time	t_{ch}	msec	PLC
Alloy pouring time	t_v	msec	PLC

Simulation	Symbol	m.u.	Check	Source
Melted alloy volume in the spilling bath	V_b	m³	$V_b = 2/3 \cdot V_{al}$	CCD Cam
Alloy volume injected into the container in the pouring phase	V_l	m³	$V_l = V_g - V_t$	Flow meter
Alloy injection temperature	T_l	°C	$T_l = T_{al} - T_s - T_t - T_{cont}$	Pyrometer
Aluminium mass	m_{al}	°C	$m_{al} = V_l \cdot p$	Mass flow meter

FIRST PHASE INJECTION

Process parameter	Symbol	m.u.	Source
First phase injection course	C1	mm	PLC
First phase time	T1	msec	PLC
First phase accumulator pressure	p_c1	Pa	PLC
Hydraulic cylinder diameter	d_c	mm	Type of product to be manufactured
Injection piston diameter	d_p	mm	Type of product to be manufactured
Container length	L_{cont}	mm	Type of product to be manufactured

Simulation	Symbol	m.u.	Check	Source
Container volume	V_{cont}	m³	$V_{cont} = L_{cont} \cdot A_p$	CCD Cam
Filling rate	f_{cont}	%	$f_{cont} = (V_{cont}/Vl) \cdot 100$	PLC
First phase course	C1	mm	$C1 = L_{cont} \cdot f_{cont}$	PLC
First phase injection speed	V1	mm/msec	$V1 = C1/T1$	PLC
Hydraulic cylinder area	A_c	mm²	$A_c = d_c^2 \cdot (3.14/4)$	CCD Cam
Injection piston area	A_p	mm²	$A_p = d_p^2 \cdot (3.14/4)$	CCD Cam
Alloy flow in first phase injection mould	$Q_{vol}1$	m³/sec	$Q_{vol}1 = Vl/T1$	Flow meter
First phase piston force	F_p1	N	$F_p1 = p_c1 \cdot A_c$	Inductive meter
Agent pressure on the metal in the first phase	$p_{al}1$	Pa	$p_{al}1 = F_p1/A_p$	Inductive meter
First phase aluminum force	$F_{al}1$	N	$F_{al}1 = p_{al}1 \cdot A_p$	Inductive meter

SECOND PHASE INJECTION

Process parameter	Symbol	m.u.	Source
Second phase injection course	C2	mm	PLC
Second phase injection time	T2	msec	PLC
Second phase accumulator pressure	p_c2	Pa	PLC
Printable area width	s_p	mm	CCD Cam
Long casting length	L_{ac}	mm	CCD Cam

Simulation	Symbol	m.u.	Check	Source
Second phase injection speed	V1	mm/msec	$V1 = C2/T2$	PLC
Second phase injection course	C2	mm	$C2 = Vl/A_p$	PLC
Second phase range	Q2	m³/sec	$Q2 = V2 \cdot Ap$	Flow meter
Alloy flow in second phase injection mould	$Q_{vol}2$	m³/sec	$Q_{vol}2 = Vl/T2$	Flow meter
Long casting area	A_{cc}	mm²	$A_{ac} = s_p \cdot L_{ac}$	CCD Cam
Long casting speed	V_a	mm/msec	$V_a = V2 \cdot (A_p/A_{cc})$	PLC
Second phase piston force	F_p2	N	$F_p2 = p_c2 \cdot A_c$	Inductive meter
Agent pressure on the metal in the second phase	$p_{al}2$	Pa	$p_{al}2 = F_p2/A_p$	Inductive meter
Second phase aluminium force	$F_{al}2$	N	$F_{al}2 = p_{al}2 \cdot A_p$	Inductive meter
Total course	C_{tot}	mm	$C_{tot} = C1 + C2$	PLC

THIRD PHASE INJECTION								
Process parameter	**Symbol**	**m.u.**	**Source**	**Simulation**	**Symbol**	**m.u.**	**Check**	**Source**
Multiplication time	TM	msec	PLC	**Multiplication pressure**	PM	Pa	$PM = ((p3 \cdot A3 - p_{contr} \cdot A_{contr}) \cdot A2)$	PLC
				Specific pressure	PS	Pa	$PS = PM \cdot (A_p/A_c)$	PLC
				Third phase injection course	CC	mm	$CC = C_{tot}-(d_p/2)$	PLC

SOLIDIFICATION								
Process parameter	**Symbol**	**m.u.**	**Source**	**Simulation**	**Symbol**	**m.u.**	**Check**	**Source**
Mould temperature	T_{mould}	°C	Thermocam	**Thermal content**	CS	Kcal	$CS = (c_l \cdot m_{al}) + (c_s*(Tl - T_e) \cdot m_{al})$	Thermocam
				Opening force	F_{open}	N	$F_{open} = PS \cdot A_s$	PLC
				Clamping force	FC	N	$FC = F_{open} \cdot 1.2$	PLC
				Thickness	SM	mm	$SM = C_p - C_{tot}$	PLC

LUBRICATION			
Process parameter	**Symbol**	**m.u.**	**Source**
Lubrication time	t_l	sec	PLC
Detachment level	DL	%	PLC

EXTRACTION			
Process parameter	**Symbol**	**m.u.**	**Source**
Extraction temperature	T_e	°C	Thermocam
Cycle time	TC	sec	MES
Good pieces	pc conf	pc	MES
Discarded pieces	pc non conf	pc	MES

Finally, to complete this modelling phase, the developed model allowed for the simulation of the piston race in its three phases, based on the variation of the technological parameters associated with it. Matlab Simulink's Model Predictive Control (MPC) was used to simulate the die-casting process (MathWorks, 2019). Figure 4.15 depicts an overview of the model and the simulation results. On Simulink (Bemporad et al., n.d.), the mathematical model is called ‘plant’ and interacts with the MPC controller. The plant consists of a piston, considered as a material point of mass (m) to which a force (F) is applied, subject to damping due to liquid metal. So, the piston is subjected to the applied force F, damping force and inertia force. Assuming that the system is in equilibrium, an equation of the second order at constant coefficients is obtained, which is the mathematical law behind the simulation. The plant integrated and connected to the MPC Controller, provides as output the velocity and the position of the piston in time. The velocity is then kept constant by the MPC through the variation of the force F. Lastly, is introduced a disturbance to the applied force F to simulate the effects of the uncontrolled parameters (i.e. inhomogeneity of molten aluminium). These inhomogeneities are completely random and not measurable in a deterministic way, depending on factors external to the process analysed, such as quality of the raw material and times of supply of the metal from the melting bath to our system.

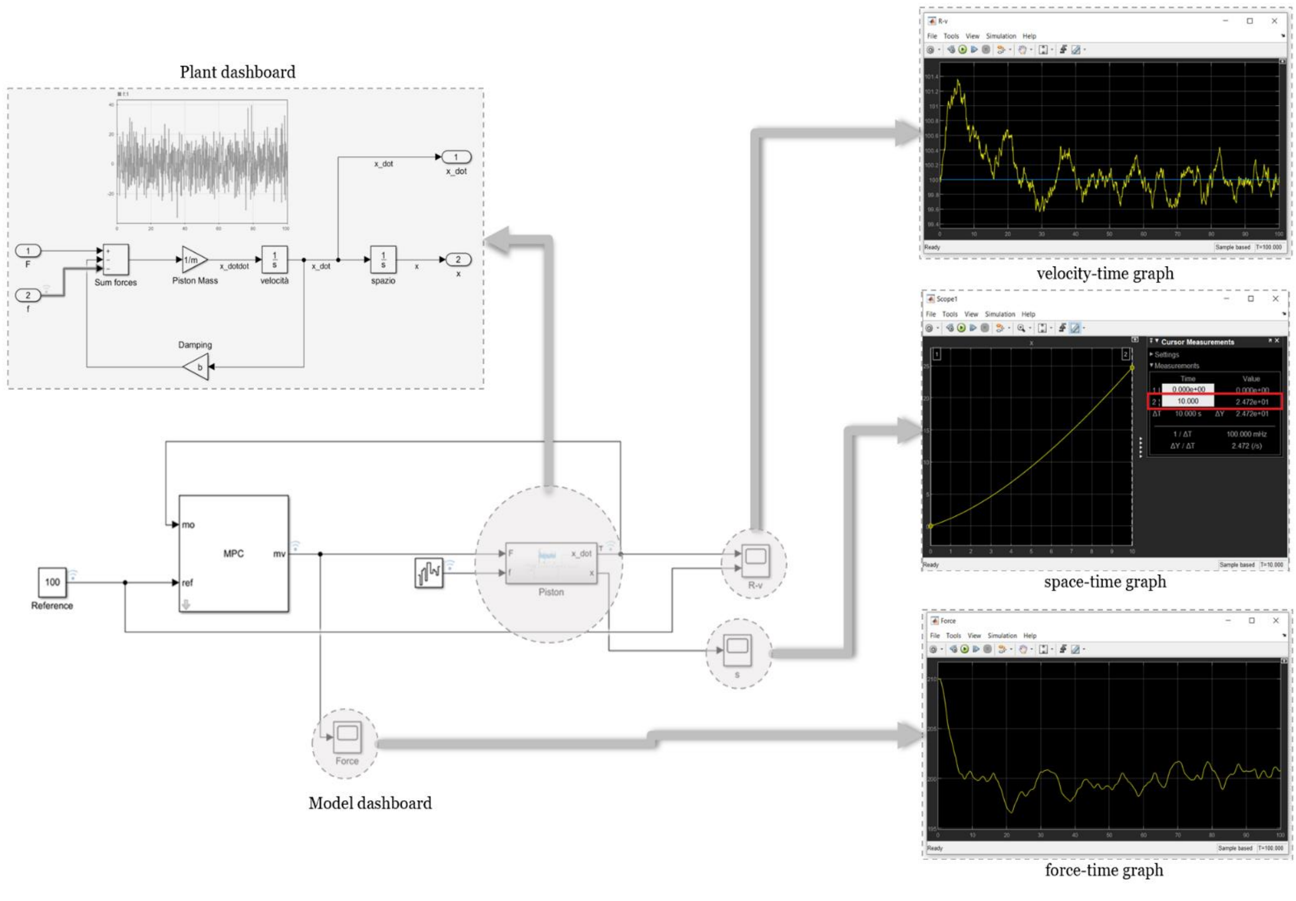


Figure 4.15 - Matlab Simulink’s MPC model of die-casting process focused on the three phases of injection-moulding

4.4. AUTOMATIZATION OF REAL-TIME PROCESS EXERGETIC ANALYSIS

In this paragraph, a novel approach to implementing the Exergy Algorithm within an online monitoring system is introduced. At each sample time, the thermodynamic variables, including the exergy, are calculated in real time. The energy and exergy efficiency index are then calculated, providing a new understanding of the evolving phenomena within the monitored process. A fast thermodynamic process was used as a test case. To put the algorithm to the test, an exergy analysis of a simple heather was performed.

To allow for iterative software implementation, the above thermodynamic equations must be rewritten in a more structured manner. The Exergy Algorithm must be applicable to generic thermodynamic streams, such as those in the chemical, industrial, electrical, and biological fields. In any case, some constraints have been imposed.

A generic thermodynamic system has been divided into three levels (Figure 4.16): Level 1 is the overall system. Reading the Table 4.5, the overall energy flux at Level 1 is calculated as the sum of all Device energy streams in the system. Sub-devices at Level 2. At Level 2, the Device represents a single machine or system with a well-defined boundary, within which one or more Level 3 material flux energy exchanges occur. An electrical energy stream may be present in a device. Finally, the device's energy input/output is the algebraic sum of materials and the electrical correspondent stream. Level 3 is the material and energy streams of the sub-device.

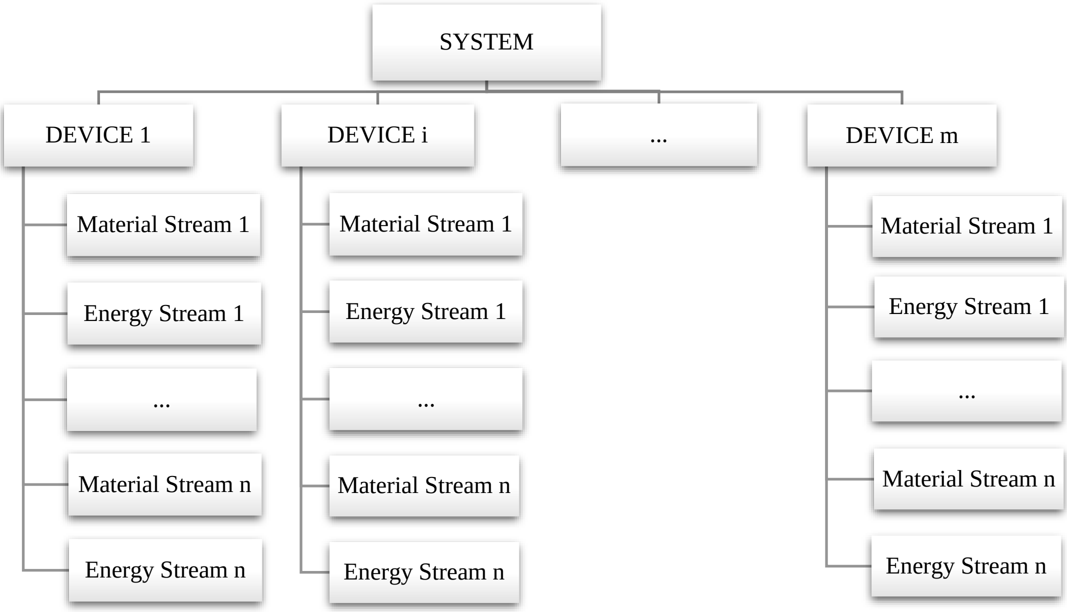


Figure 4.16 - Structure of a generic thermodynamic system

Only the following thermodynamic phenomena has been considered: heat transfer from a convective material at constant pressure and volume; while other assumptions were:

• There was no material state transformation.

• There was a constant flow of cinematic, internal, and potential energy.

• The homogeneous system outside the System boundary represents the environment. The environment is assumed to have known temperature and pressure, as well as an infinite heat capacity.

Table 4.5 - Thermodynamic system scheme

Environment	$\mathrm{DeadStateTemp}_{T0(t)}$ [K]
Level 3	**Materials**
Measured Variables	ij generic Material Material flux $\mathrm{mij}_{in(t)} \left[\frac{kg}{s}\right]$ $\mathrm{mij}_{out(t)} \left[\frac{kg}{s}\right]$ Temperatures $\mathrm{tempij}_{in(t)}$ [K] $\mathrm{tempij}_{out(t)}$ [K]

Constants	$'\mathrm{SpecificHeat}'ij$ [J/kgK] (referred to the ij material)
Specific enthalpy	$hij_{in(t)} \left[\frac{J}{kg}\right] = \mathrm{SpecificHeat}\ ij * \left(temp_{in(t)} - \mathrm{DeadStateTemp}_{T0(t)}\right)$ $hij_{out(t)} \left[\frac{J}{kg}\right] = \mathrm{SpecificHeat}\ ij * \left(temp_{out(t)} - \mathrm{DeadStateTemp}_{T0(t)}\right)$ $hij(t) \left[\frac{J}{kg}\right] = h_{out(t)} - h_{in(t)}$
Specific entropy	$sij_{in(t)} \left[\frac{J}{kgK}\right] = \mathrm{SpecificHeat}\ ij * \left(\ln\left(\frac{temp_{in(t)}}{\mathrm{DeadStateTemp}_{T0(t)}}\right)\right)$ $sij_{out(t)} \left[\frac{J}{kgK}\right] = \mathrm{SpecificHeat}\ ij * \left(\ln\left(\frac{temp_{out(t)}}{\mathrm{DeadStateTemp}_{T0(t)}}\right)\right)$ $s(t) \left[\frac{J}{kgK}\right] = s_{out(t)} - s_{in}$
Specific exergy	$exij_{in(t)} \left[\frac{J}{kg}\right] = hij_{in(t)} - \mathrm{DeadStateTemp}_{T0(t)} * sij_{in(t)}$ $exij_{out(t)} \left[\frac{J}{kg}\right] = hij_{out(t)} - \mathrm{DeadStateTemp}_{T0(t)} * sij_{out(t)}$ $exij(t) \left[\frac{J}{kg}\right] = exij_{out(t)} - exij_{in(t)}$
Material stream	$matij_{in(t)}[kg] = \sum_{0}^{t} m_{in(\tau)} * d\tau$ $matij_{out(t)[kg]} = \sum_{0}^{t} m_{out(\tau)} * d\tau$ $matij(t)[kg] = mat_{out(t)} - mat_{in(t)}$
Cumulative Enthalpy	$Hij_{in(t)}\ _{[J]} = \sum_{0}^{t} m_{in(\tau)} * h_{in(\tau)} * d\tau$ $Hij_{out(t)} = \sum_{0}^{t} m_{out(\tau)} * h_{out(\tau)} * d\tau$ $Hij(t) = H_{out(t)} - H_{in(t)}$
Cumulative Entropy	$Sij_{in}(t) = \sum_{0}^{t} m_{in(\tau)} * s_{in(\tau)} * d\tau$ $Sij_{out(t)} = \sum_{0}^{t} m_{out(\tau)} * s_{out(\tau)} * d\tau$ $Sij_{j}(t) = S_{out}(t) - S_{in(t)}$
Cumulative Exergy	$Exij_{in}(t) = \sum_{0}^{t} m_{in(\tau)} * ex_{in(\tau)} * d\tau$ $Exij_{out}(t) = \sum_{0}^{t} m_{out(\tau)} * ex_{out(\tau)} * d\tau$ $Exij\ (t) = Ex_{out}(t) - Ex_{in}(t)$
Cumulative Max Entalpy	$Hexij_{in}(t) = \sum_{0}^{t} m_{in(\tau)} * h_{in(\tau)} * eff_{max(t)} * d\tau$ $Hexij_{out}(t) = \sum_{0}^{t} m_{out(\tau)} * h_{out(\tau)} * eff_{max(t)} * d\tau$ $Hexij\ (t) = Hex_{out(t)} - Hex_{in(t)}$ Where:

	$\text{temp}_{\text{meanij(t)}} = \frac{[\,\text{temp}_{\text{in}}(t) + \text{temp}_{\text{out}}(t)]}{2}$ $\text{eff}_{\text{maxij(t)}} = \left[1 - \frac{\text{DeadStateTemp}_{T0}(t)}{\text{temp}_{\text{meanij}}(t)}\right]$

Level 2	**Devices**
Cumulative Enthalpy *Only material flux*	$H_{\text{in}_{\text{mat}_{\text{Device(t)}\,i}}} = \sum_{j=1}^{n} H_{\text{in}_{ij}}(t)$ $H_{\text{out}_{\text{mat}_{\text{Device(t)}\,i}}} = \sum_{j=1}^{n} H_{\text{out}_{ij}}(t)$ $H_{\text{mat}_{\text{Device(t)}\,i}} = H_{\text{out}_{\text{mat}_{\text{Device(t)}}}} - H_{\text{in}_{\text{mat}_{\text{Device(t)}}}}$
Cumulative Entropy *Only material flux*	$S_{\text{in}_{\text{mat}_{\text{Device(t)}\,i}}} = \sum_{j=1}^{n} S_{\text{in}_{ij}}(t)$ $S_{\text{out}_{\text{mat}_{\text{Device(t)}\,i}}} = \sum_{j=1}^{n} S_{\text{out}_{ij}}(t)$ $S_{\text{mat}_{\text{Device(t)}\,i}} = S_{\text{out}_{\text{mat}_{\text{Device(t)}\,i}}} - S_{\text{in}_{\text{mat}_{\text{Device(t)}\,i}}}$
Cumulative Exergy *Only material flux*	$Ex_{\text{in}_{\text{mat}_{\text{Device(t)}\,i}}} = \sum_{j=1}^{n} Ex_{\text{in}_{ij}}(t)$ $Ex_{\text{out}_{\text{mat}_{\text{Device(t)}\,i}}} = \sum_{j=1}^{n} Ex_{\text{out}_{ij}}(t)$ $Ex_{\text{mat}_{\text{Device(t)}\,i}} = Ex_{\text{out}_{\text{mat}_{\text{Device(t)}\,i}}} - Ex_{\text{in}_{\text{mat}_{\text{Device(t)}\,i}}}$
Cumulative Max Enthalpy *Only material flux*	$Hex_{\text{in}_{\text{mat}_{\text{Device(t)}\,i}}} = \sum_{j=1}^{n} Hex_{\text{in}_{ij}}(t)$ $Hex_{\text{out}_{\text{mat}_{\text{Device(t)}\,i}}} = \sum_{j=1}^{n} Hex_{\text{out}_{ij}}(t)$ $Hex_{\text{mat}_{\text{Device(t)}\,i}} = Hex_{\text{out}_{\text{mat}_{\text{Device(t)}\,i}}} - Hex_{\text{in}_{\text{mat}_{\text{Device(t)}\,i}}}$
Cumulative Electrical Energy	$En_{\text{el}_{\text{in(t)}_i}}[J] = \sum_{0}^{t} W_{\text{in}_j}(t)\,d\tau$ $En_{\text{el}_{\text{out(t)}_i}}\,[J] = \sum_{0}^{t} W_{\text{out}_j}(t)\,d\tau$
Cumulative Energy balance	$En_{\text{in}_{\text{Device(t)}_i}} = H_{\text{in}_{\text{mat}_{\text{Device(t)}_i}}} + En_{\text{el}_{\text{in(t)}\,i}}$ $En_{\text{out}_{\text{Device(t)}_i}} = H_{\text{out}_{\text{mat}_{\text{Device(t)}_i}}} + En_{\text{elout(t)}_i}$ $En_{\text{Device(t)}_i} = En_{\text{out}_{\text{Device(t)}\,i}} - En_{\text{in}_{\text{Device(t)}\,i}}$
Cumulative Energy Efficiency	$En_{\text{Device(t)eff}_i} = \frac{En_{\text{out}_{\text{Device(t)}_i}}}{En_{\text{in}_{\text{Device(t)}_i}}}$
Cumulative Exergy balance	$Ex_{\text{in}_{\text{Device(t)}_i}} = Ex_{\text{in}_{\text{mat}_{\text{Device(t)}_i}}} + En_{\text{el}_{\text{in(t)}_i}} + Hex_{\text{in}_{\text{mat}_{\text{Device(t)}_i}}}$ $Ex_{\text{out}_{\text{Device(t)net}_i}} = Ex_{\text{out}_{\text{mat}_{\text{Device(t)}_i}}} + En_{\text{el}_{\text{out(t)}_i}}$ $Ex_{\text{out}_{\text{Device(t)tot}_i}} = Ex_{\text{out}_{\text{Device(t)net}_i}} + Hex_{\text{out}_{\text{mat}_{\text{Device(t)}_i}}}$ $Ex_{\text{loss}_{\text{Device(t)}_i}} = Ex_{\text{out}_{\text{Device(t)}_{\text{tot}_i}}} - Ex_{\text{in}_{\text{Device(t)}_i}}$

Cumulative Exergy Efficiency	$\mathrm{Ex}_{\mathrm{gen}_{\mathrm{Device(t)}_{\mathrm{eff}}}} = \frac{\mathrm{Ex}_{\mathrm{out}_{\mathrm{Device(t)}_{\mathrm{tot}}}}}{\mathrm{Ex}_{\mathrm{in}_{\mathrm{Device(t)}_{\mathrm{tot}}}}}$ $\mathrm{Ex}_{\mathrm{net}_{\mathrm{Device(t)}_{\mathrm{eff}}}} = \frac{\mathrm{Ex}_{\mathrm{out}_{\mathrm{Device(t)}_{\mathrm{net}}}}}{\mathrm{Ex}_{\mathrm{in}_{\mathrm{Device(t)}_{\mathrm{tot}}}}}$

Level 1	**System**
Cumulative Electrical Energy	$\mathrm{En}_{\mathrm{el}_{\mathrm{in}_{\mathrm{System(t)}}}}\ [\mathrm{J}] = \sum_{i=1}^{m} \mathrm{En}_{\mathrm{el}_{\mathrm{in}_{(t)_i}}}$ $\mathrm{En}_{\mathrm{el}_{\mathrm{out}_{\mathrm{System(t)}}}}\ [\mathrm{J}] = \sum_{i=1}^{m} \mathrm{En}_{\mathrm{el}_{\mathrm{out}_{(t)_i}}}$ $\mathrm{En}_{\mathrm{el}_{\mathrm{System(t)}}}[\mathrm{J}] = \mathrm{En}_{\mathrm{el}_{\mathrm{out}_{\mathrm{System(t)}}}} - \mathrm{En}_{\mathrm{el}_{\mathrm{in}_{\mathrm{System(t)}}}}$
Cumulative Energy balance	$\mathrm{En}_{\mathrm{in}_{\mathrm{System(t)}}} = \sum_{i=1}^{m} \mathrm{En}_{\mathrm{in}_{\mathrm{Device}_i}}(t)$ $\mathrm{En}_{\mathrm{out}_{\mathrm{System(t)}}} = \sum_{i=1}^{m} \mathrm{En}_{\mathrm{out}_{\mathrm{Device}_i}}(t)$ $\mathrm{En}_{\mathrm{System(t)}} = \sum_{i=1}^{m} \mathrm{En}_{\mathrm{Device}_i}(t)$
Cumulative Energy Efficiency	$\mathrm{En}_{\mathrm{System(t)}_{\mathrm{eff}}} = \frac{\mathrm{En}_{\mathrm{out}_{\mathrm{System(t)}}}}{\mathrm{En}_{\mathrm{in}_{\mathrm{System(t)}}}}$
Cumulative Exergy balance	$\mathrm{Ex}_{\mathrm{in}_{\mathrm{System(t)}}} = \sum_{i=1}^{m} \mathrm{Ex}_{\mathrm{in}_{\mathrm{Device}_i}}(t)$ $\mathrm{Ex}_{\mathrm{out}_{\mathrm{System(t)}}} = \sum_{i=1}^{m} \mathrm{Ex}_{\mathrm{out}_{\mathrm{Device}_i}}(t)$ $\mathrm{Ex}_{\mathrm{out}_{\mathrm{System(t)}_{\mathrm{tot}}}} = \sum_{i=1}^{m} \mathrm{Ex}_{\mathrm{out}_{\mathrm{Device}_i}}(t)\quad \mathrm{tot}$ $\mathrm{Ex}_{\mathrm{loss}_{\mathrm{System(t)}}} = \sum_{i=1}^{m} \mathrm{Ex}_{\mathrm{loss}_{\mathrm{Device}_i}}(t)$
Cumulative Exergy Efficiency	$\mathrm{Ex}_{\mathrm{gen}_{\mathrm{System(t)}_{\mathrm{eff}}}} = \frac{\mathrm{Ex}_{\mathrm{out}_{\mathrm{System(t)}_{\mathrm{tot}}}}}{\mathrm{Ex}_{\mathrm{in}_{\mathrm{System(t)}_{\mathrm{tot}}}}}$ $\mathrm{Ex}_{\mathrm{net}_{\mathrm{System(t)}_{\mathrm{eff}}}} = \frac{\mathrm{Ex}_{\mathrm{out}_{\mathrm{System(t)}_{\mathrm{net}}}}}{\mathrm{Ex}_{\mathrm{in}_{\mathrm{System(t)}_{\mathrm{tot}}}}}$

4.4.1. Measuring and Monitoring System

In order to improve knowledge about all of the individual parameters at stake, industrial processes necessitate the acquisition of multiple signals.
The thermodynamic laws underlying Exergetic Analysis are important for tracing the set of parameters that should be measured and monitored throughout the process, as well as the variables that can be calculated. Reference flows are uniquely identified in its balance equations, which are based on Szargut's studies (see the equations in paragraph 2.1.2), as well as the performance metrics (see paragraph 3.1.2).
The system proposed configuration includes the following features.
Operational features to be obtained:

1. modular system
2. portability
3. adaptability
4. remotely accessible
5. platform for the cloud

Physical quantities that must be collected:
1. electrical specifications (e.g., power, energy)
2. temperatures (via thermocouples)
3. temperature (via pyrometer)
4. mV, V, A analogue signals.

Then, the monitoring system has been assembled with the following components:

- I/O Analog Acquisition Board (Figure 4.17)
- Raspberry Pi 4 (Figure 4.18)
- Type K thermocouples (Figure 4.17)
- Type T thermocouples (Figure 4.17)

Plus, the systems already in place in the department:

- PLC (quantitative information)
- MES (qualitative information)

Data elaboration tools:

- Database: MySql
- Data and plots have been elaborated on Jupyter ("Project Jupyter," 2021)

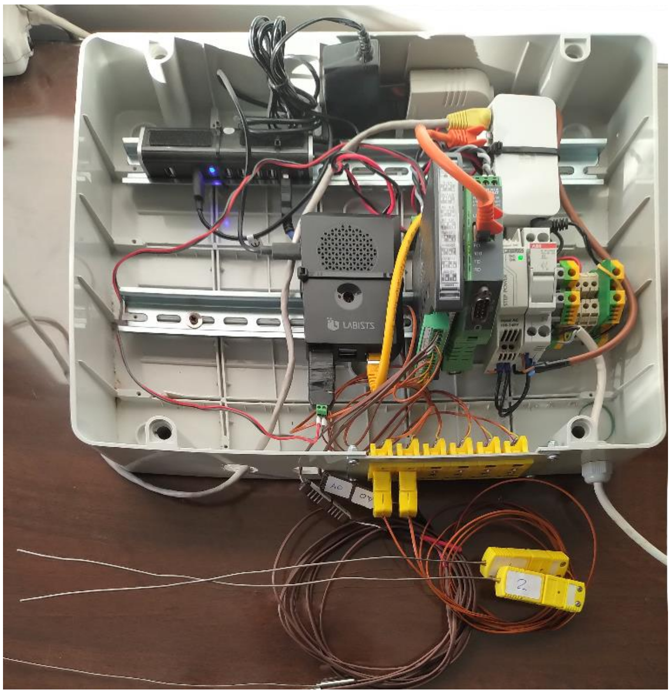


Figure 4.17 - Acquisition board

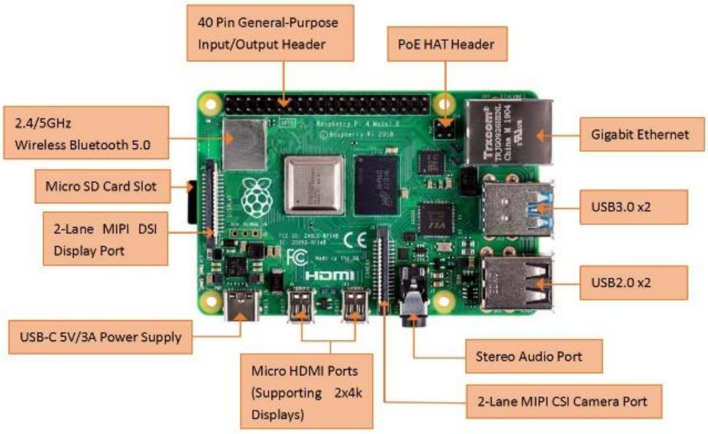


Figure 4.18 - Raspberry Pi 4 model B

The acquisition board can include up to 8 analogue sensors. Eventually energy meters, flux meters and other measurement devices can be included.

4.4.2. Implementation of the Algorithm in Python

The iterative algorithm includes a generic number of the following basic objects included in the overall System:

- Measurement instrument
- Measured variable
- Material
- Device

The EA algorithm has been implemented in Python using the following steps:

1. Read Setup file: the above objects are defined in a setup file, which contains Python dictionaries as follows in Table 4.6.
2. Data acquisition
3. Load of data
4. Check data
5. Calculate:
   i. Material Streams
      (a) Enthalpy
      (b) Entropy
      (c) Exergy
   ii. Devices Streams
      (a) Electrical Streams
      (b) Enthalpy
      (c) Entropy
      (d) Exergy
   iii. System Streams
      (a) Electrical Streams
      (b) Enthalpy
      (c) Entropy
      (d) Exergy
6. Store Data
   i. Material Streams
   ii. Device Streams

iii. System Streams

Table 4.6 - Algorithm structure in Python

Dictionary name	Filename *configSetup.py* Dictionary name: config_setup = {**…**}
Sample Data setup	‘sampleId’: ‘**Test**’, ‘timeStep’: ‘**2s**’,

Materials Specifications	‘matSpec’: {
	0: {
	‘matId’: 0,
	‘matName’: ‘ **…** ‘,
	‘matDescr’: ‘ **…** ‘,
	‘SpecificHeat’: **…**,
	‘SpecificHeat_um’: ‘**J/(kgK)**’,
	‘Density’: **…**,
	‘Density_um’: ‘**kg/m3**’,
	},
	1: {
	‘matId’: 1,
	‘matName’: ‘ **…** ‘,
	‘matDescr’: ‘ **…** ‘,
	‘SpecificHeat’: **…**,
	‘SpecificHeat_um’: ‘**J/(kgK)**’,
	‘Density’: **…**,
	‘Density_um’: ‘**kg/m3**’,
	},
	…
	n: {
	‘matId’: n,
	‘matName’: ‘ **…** ‘,
	‘matDescr’: ‘ **…** ‘,
	‘SpecificHeat’: **…**,
	‘SpecificHeat_um’: ‘**J/(kgK)**’,
	‘Density’: **…**,
	‘Density_um’: ‘**kg/m3**’,
	},
	},

Measurement devices	Structure:
	‘measurement_devices’: {
	‘idDevice’: {
	idPar_1: ‘Description Parameter’,
	…
	idPar_n: ‘Description Parameter’,
	},
	}
	‘measurement_devices’: {
	# Electric meter dev. 1
	‘e00’: {133: ‘ **… […]**’},
	
	# Temperature dev. 1
	‘t00’: {370: ‘ **…** ‘,
	371: ‘ **…** ‘,
	372: ‘ **…** ‘,
	373: ‘ **…** ‘,
	},
	},

Environment Variables	Variable declaration: ‘variable_name’: [‘idDevice_idParameter’ , k , offset],

	Where: k, offset: constants for linear transformation *v_out = k * v_in + offset* *idDevice: The measurement Device id* *idParameter: The measurement parameter id* ‘envParams’: { ‘temp_env’: [‘t00_370’, 1, **…**], ‘temp_env_um’: ‘**K**’, ‘DeadStateTemp_T0’: [‘t00_371’, 1, **…**], ‘DeadStateTemp_T0_um’: ‘**K**’, ‘DeadStatePressure_P0’: **…**, ‘DeadStatePressure_P0_um’: ‘**atm**’, },

System structure	Structure:
	
	‘sysDevices’: {
	0: {
	‘devName’: ‘**…**’,
	‘devDescription’: ‘**…**’,
	‘elecStream’: { **…** }
	‘matStream’: {
	0: {**…**}
	j: {**…**}
	…
	n: {**…**}
	}
	},
	i: {**…**},
	…
	m: {**…**}
	}

Considering a general Thermodynamic System composed of multiple machines, Level 1 includes the System model. In Level 2, each machine is represented as a "Device." Material fluxes cross the device boundary. At Level 3 of the aforementioned structure, each material flux and related energy stream is modelled.
Consider that i = 0,…,m represents the index associated with a generic j device, and j=0,…,n represents the index associated with the i_{th} device generic material streams where i is the generic device and ij the generic material stream.

5.3.2.1. Gaps in the application

The described algorithm needs to be applied and evaluated for further improvements.

1) Values of initial energy.
   The initial values of the energy and enthalpy variables have been set to zero. This causes errors in the calculation of the efficiency index.

2) Efficiency indices.
   The efficiency indices were calculated iteratively as the ratio of cumulated values at each sampling instant, beginning with the starting time. This needs to be revised because the actual initial values of internal energies were neglected. A better approach would be to calculate the efficiency index on a regular basis (hourly, daily, etc...) rather than at each sample.

3) Accumulation of energy and transients.
The exergy balance equations are only applicable to steady-state systems. During the transient, it is expected to see the dynamic behaviour of exergy balance and exergy efficiency index. This was also discovered in this Test Case. This will be looked into. The iterative algorithm must consider transient terms related to thermal energy accumulation and release phenomena. Otherwise, only the steady-state periods will be considered.

4.4.3. Model Test and Validation

The model was validated on a single sub-device, a heat exchanger installed on the company's field.
The Setup algorithm is depicted below:

```
#
# Setup Data
#
config_setup = {
    # General Sample Data Information
    'sampleId': 'Test',
    'Description': 'Test over heater',
    'location': '',

    # DateTime
    'startDate': '2021-01-09 00:00:00',
    'endDate':   '2021-01-10 00:00:00',
    'timeStep': '2s',

    # Devices
    'devices': {
        0: True,
    },

    # Materials Specifications
    'matSpec': {
        0: {
            'matId': 0,
            'matName': 'Water',
            'matDescr': '',
            'SpecificHeat': 4185,
            'SpecificHeat_um': 'J/(kgK)',
            'Density': 1000,
            'Density_um': 'kg/m3',
            'Melting_point': 0,
            'Melting_point_um': '°C',
            'latent_heat': 333500,
            'latent_heat_um': 'J/kg',

        },
        1: {
            'matId': 1,
            'matName': 'Air',
            'matDescr': 'Air',
            'SpecificHeat': 1005,
```

```
        'SpecificHeat_um': 'J/(kgK)',
        'Density': 1.2,
        'Density_um': 'kg/m3',
    },
},

# Measurement devices
'measurement_devices': {

    # Temperature dev. 1
    't00': {
            370: 'External air temperature',
            371: 'Internal air temperature',
            372: 'Heater input water temperature',
            373: 'Heater output water temperature',
            },
           },

# Variables
'envParams':
    {
        #Environment Devices
        'temp_env': ['t00_370', 1, 273.15],
        'temp_env_um': '°C',
        'DeadStateTemp_T0': ['t00_371', 1, 273.15],
        'DeadStateTemp_T0_um': 'K',
    },

'sysDevices':
    {
        # Energy Balance
        # Home heater
        0: {'devId': 0,
            'devName': 'Home_heater',
            'devDescription': 'Home_heater',

            'elecStream':   {
                          'W_in': ['nd', 1000],
                          'W_in_um': 'W',
                          'W_out':  ['nd',0],
                          'W_out_um': 'W',
                        },

            'matStream': {
                0: {
                    'matId': 0,
                    'description': 'Heater water',
                    'm_in': ['nd', 0.1],
                    'm_in_um': 'kg/s',
                    'm_out': ['nd', 0.1],
                    'm_out_um': 'kg/s',
                    'temp_in': ['t00_372', 1, 273.15],
                    'temp_in_um': '°C',
                    'temp_out': ['t00_373', 1, 273.15],
```

```
                'temp_out_um': '°C',
            },

            1: {
                'matId': 1,
                'description': 'Air',
                'm_in': ['nd', 0.3],
                'm_in_um': 'kg/s',
                'm_out': ['nd', 0.3],
                'm_out_um': 'kg/s',
                'temp_in': ['t00_370', 1, 273.15],
                'temp_in_um': '°C',
                'temp_out': ['t00_371', 1, 273.15],
                'temp_out_um': '°C',
            },
        },
    },
  }
}
```

The following diagrams depict the measured and monitored temperatures and fluxes (Figure 4.19, Figure 4.20, Figure 4.21).

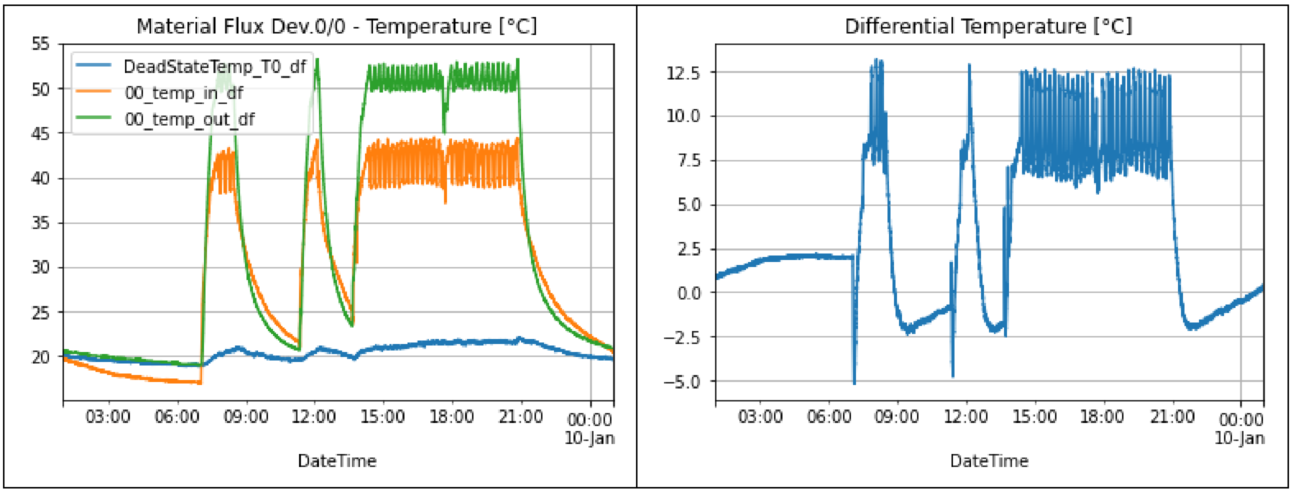


Figure 4.19 - Input/Output Temperatures of Device 0 – Material 0

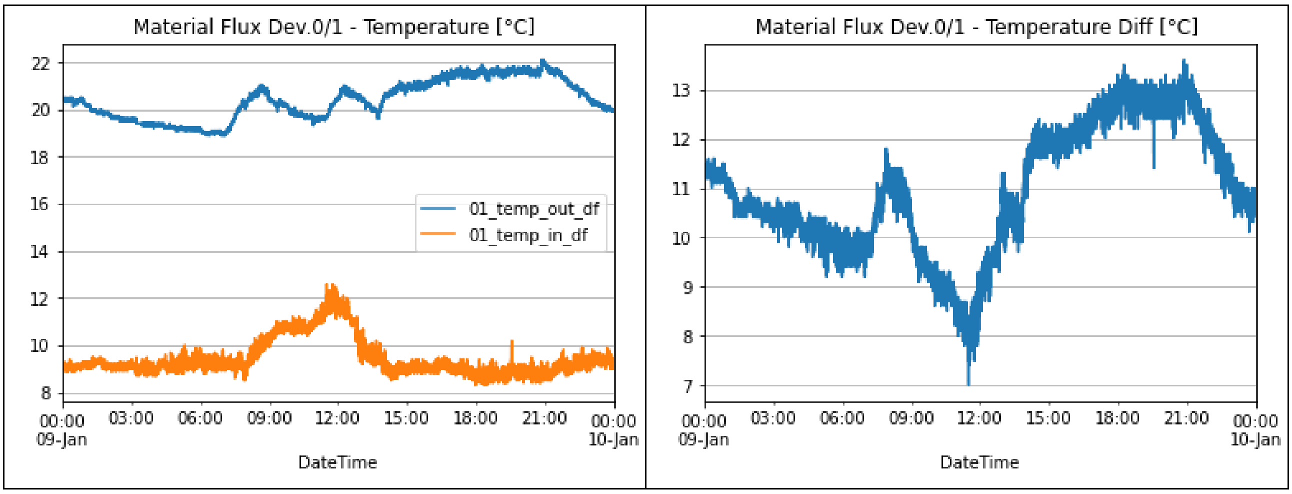


Figure 4.20 - Input/Output Temperatures of Device 0 – Material 1

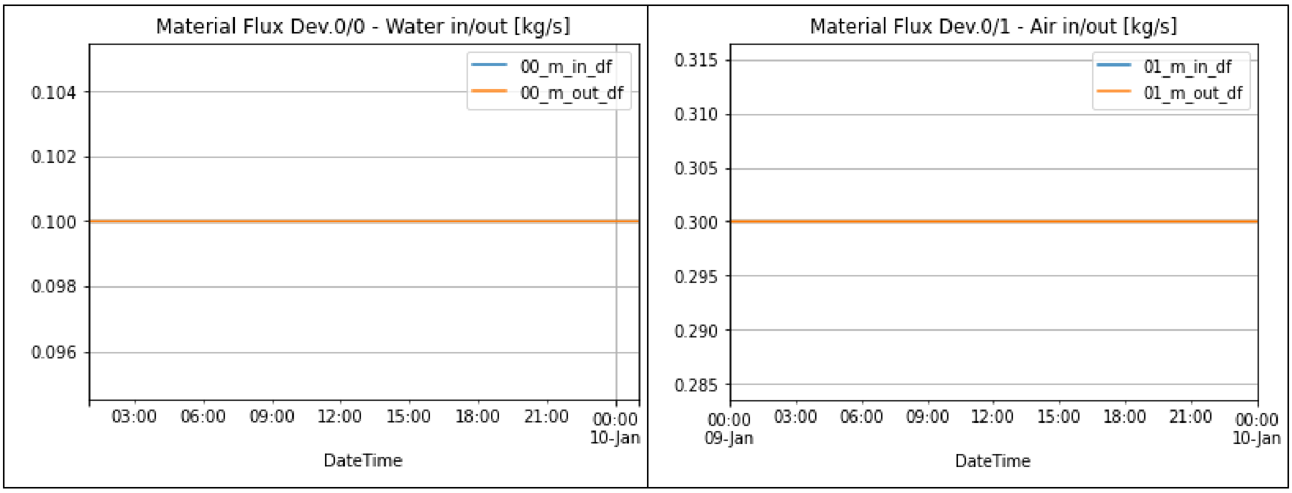
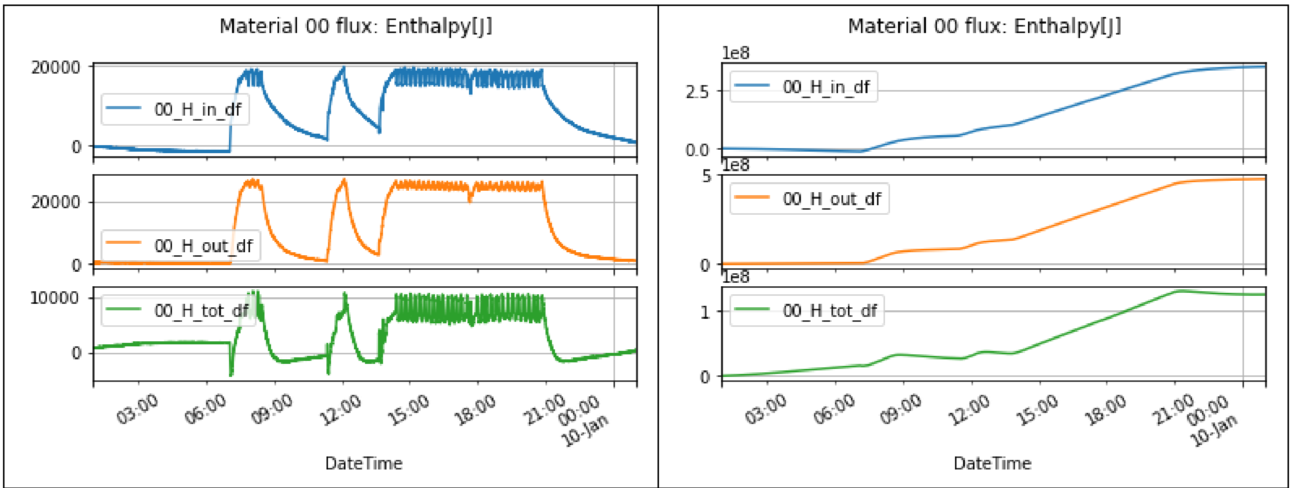


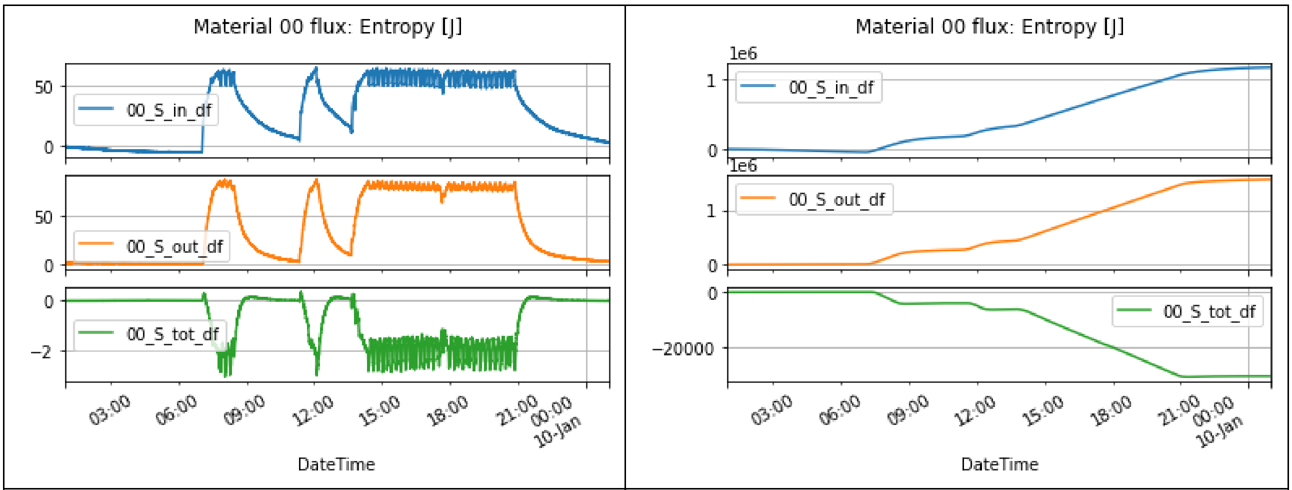
Figure 4.21 - Material Flux

The following diagrams depict the measured and monitored enthalpy, entropy and exergy (Figure 4.22, Figure 4.23 and Figure 4.24).



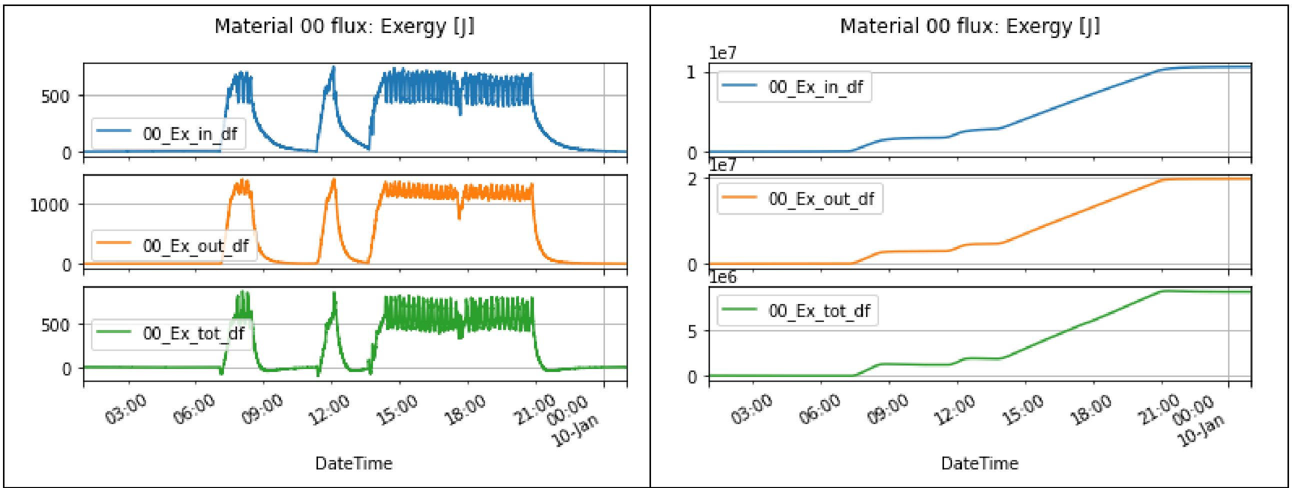
*Instant values - **dt=2 s – E(0) =0***	*Cumulative value*

Figure 4.22 - DEVICE 0 - Material 0 – Enthalpy (H)



*Instant values - **dt=2 s – E(0) =0***	*Cumulative value*

Figure 4.23 - DEVICE 0 - Material 0 – Entropy (S)



*Instant values - **dt=2 s – E(0) =0***	*Cumulative value*

Figure 4.24 - DEVICE 0 - Material 0 – Exergy (Ex)

The following diagrams depict the measured and monitored enthalpy, entropy and exergy (Figure 4.25, Figure 4.26, Figure 4.27, Figure 4.28, Figure 4.29).

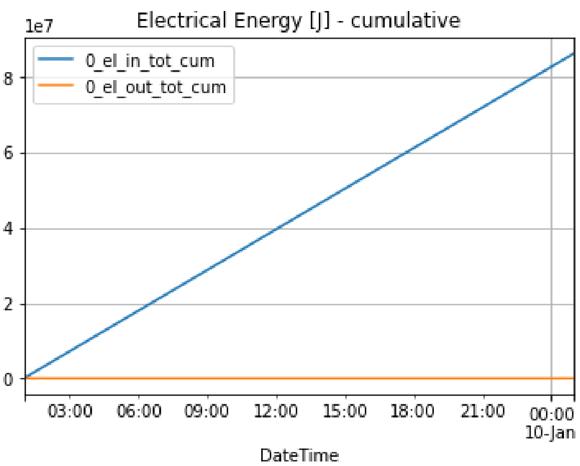
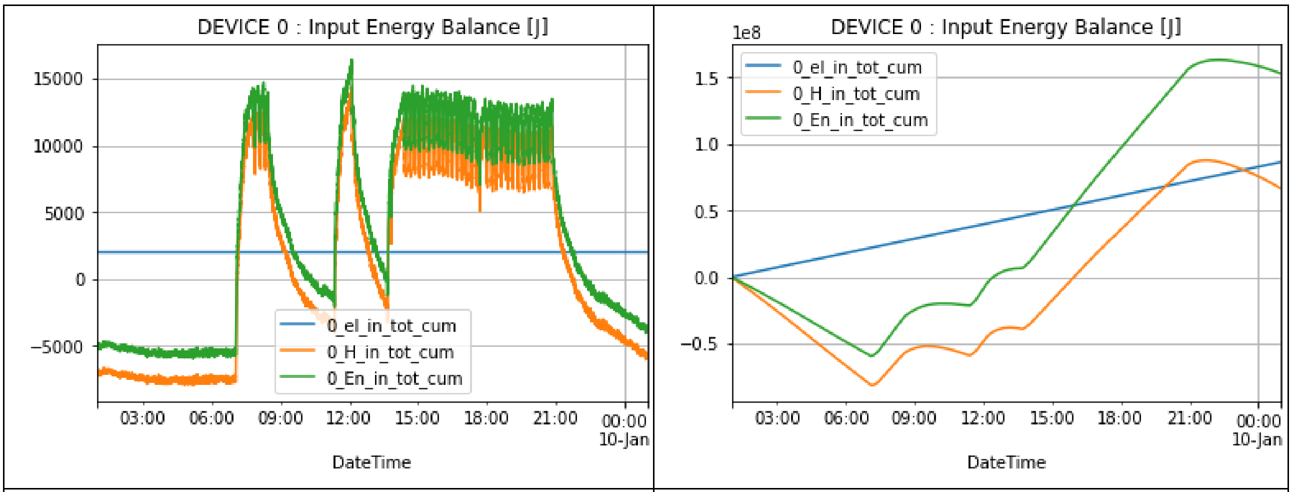
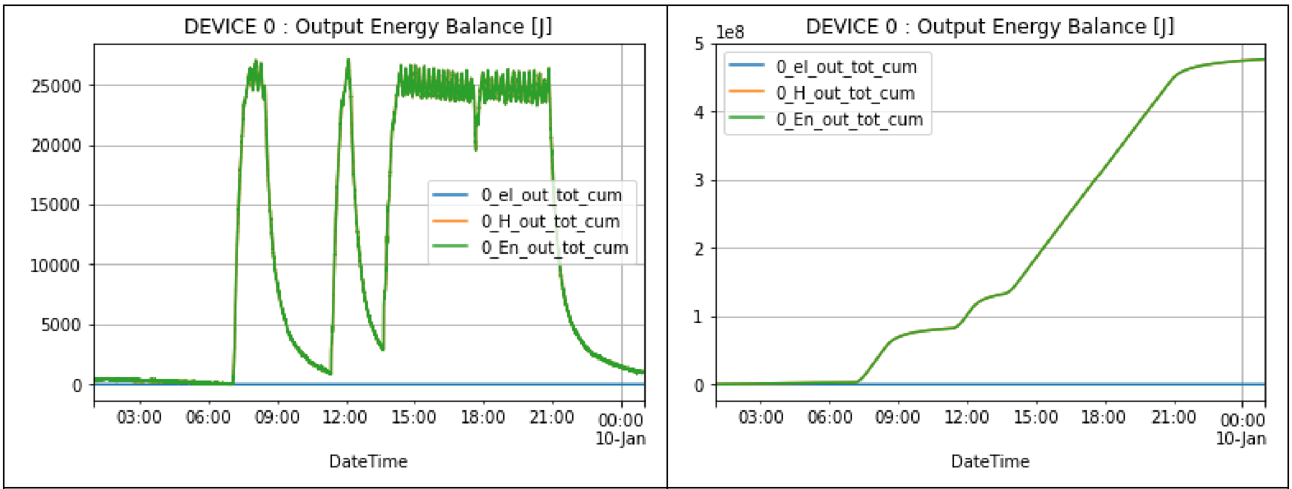


Figure 4.25 - DEVICE 0 - Electrical energy (Pconst = 1kW)



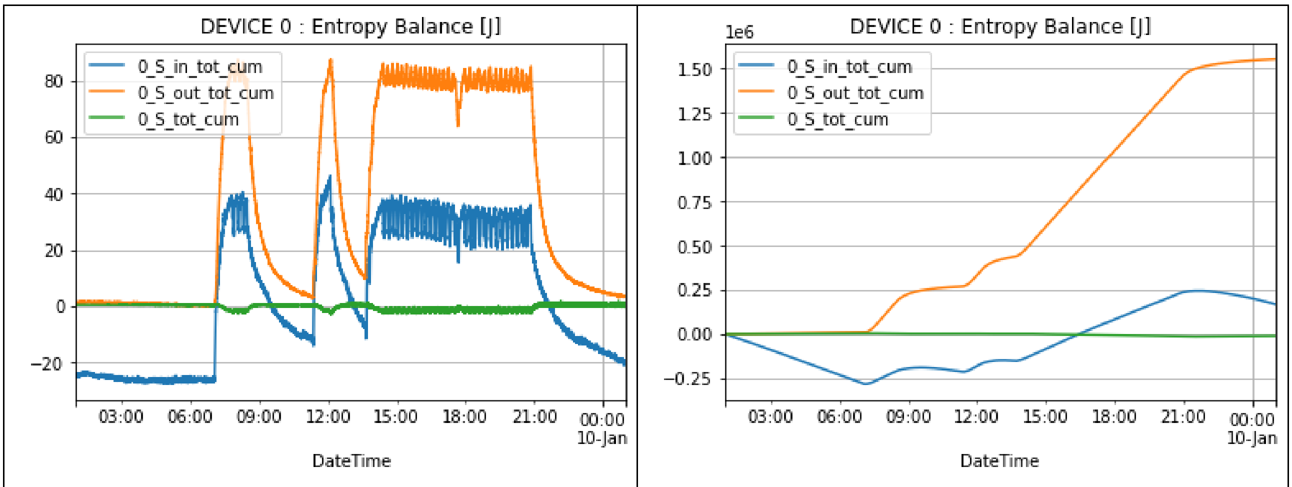
Instant values - ***E(0) = 0***	Cumulative value

Figure 4.26 - DEVICE 0 - Input balances



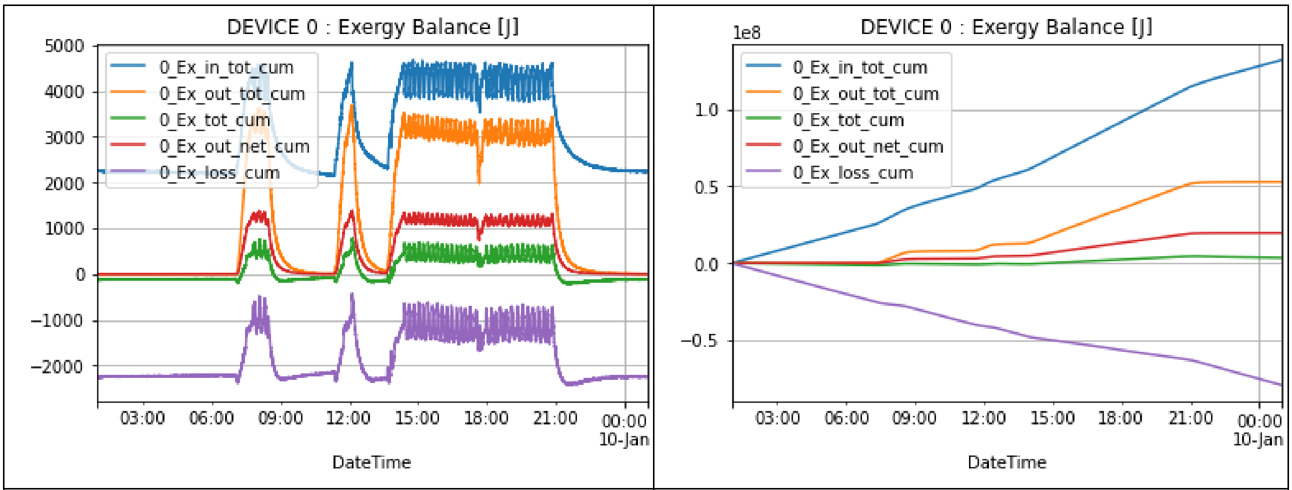
Instant values –***E(0) =0***	Cumulative value

Figure 4.27 - DEVICE 0 - Output balances



Instant values - ***E(0) =0***	Cumulative value

Figure 4.28 - DEVICE 0 - Entropy balance



*Instant values – **E(0) =0***	*Cumulative value*

Figure 4.29 - DEVICE 0 - Exergy balance

And finally, the following diagram depicts the calculated exergy net and general efficiencies (Figure 4.30).

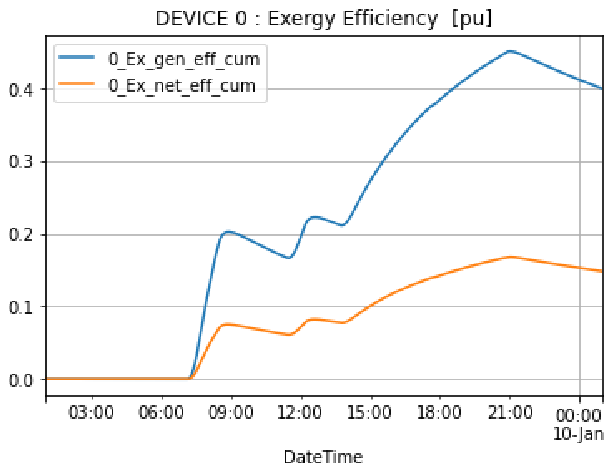


Figure 4.30 - DEVICE 0 - Exergy efficiency (general and net)

4.5. DATASETS ANALYSIS

As previously stated, the case study focuses on Master Italy aluminium die-casting process. The machine in discussion is known as “F55,” and it is located in a department floor dedicated to die-casting aluminium and zamak.

There are two data sets available for the machine being tested:

1) a collection of datasets relating to recordings of the machine’s operating parameters during the execution of each cycle detected by the PLC on board the machine and obtained from the system into which these measurements are poured on a regular basis.

2) the recording of the causes of machine downtime and anomalies performed by operators on the MES system for production progress management.

In the future, hopefully, there will also be field data from additional sensors that will monitor the internal micro-climatic conditions at the external facility climatic conditions. It will be possible to extract new patterns and correlations between process performance parameters and the latters to understand whether temperature and radiant average humidity affect the quality of the process and, more importantly, the finished product.

4.5.1. PLC

PLC is an abbreviation for Programmable Logic Controller, and it refers to a device used to control industrial processes. PLCs operate on an industrial plant by running a program and processing digital and analogical signals from sensors. It is an integral part of the "F55" machine. The process parameters continuously monitored by the machine are provided in Table 4.7. The limit values are established by trained operators and are those for which the machine goes into alert or stops completely. In that situation, the operator is compelled to act and declare (on the MES) the reason (causal) for which it was summoned. operator personnel often has to fix faulty situations manually because the possible faults and fault combinations are manifold and often need manual mechanical intervention by an operator (Vogel-Heuser et al., 2016). The parameters are congruent with what is indicated in paragraph 4.1.2. In terms of the die-casting process, the distinction between the characteristics of the first phase, second phase, molding, clamping and extraction is clear.

Table 4.7 - Example of the operating and alarm process parameter values set on "F55" machine PLC by specialized operators

Process Parameters		Lower Limits (MIN)			Upper Limits (MAX)	
Name	**Symbol**	<<	<	**m.u.**	>	>>
course first phase	**C1**	212	237	mm	263	288
time first phase	**T1**	1300	1360	ms	1532	1678
speed first phase	**V1**	0.14	0.16	m/s	0.18	0.2
piston seizure	**GP**	2	3	bar	12	20
course second phase	**C2**	95	137	mm	122	166
time second phase	**T2**	90	142	ms	106	173
average speed second phase	**V2**	0.82	0.9	m/s	1.45	1.5
max speed second phase	**VM**	1.5	1.55	m/s	2.3	2.5
course third phase (compression)	**CC**	1	2	mm	8	10
solidification time	**T3**	13	15	ms	20	22
solidification time delay	**TD**	20	23	ms	25	28
injection pression	**PM**	250	260	bar	280	290
final pression	**PF**	205	210	bar	285	310
speed casting doses	**VA**	0	0	m/s	0	0
filling pression (of the clamp)	**PR**	20	22	bar	24	26
specific pression	**PS**	830	927	bar	1025	1122
closing (or clamping) force	**FC**	5000	5391	kN	5959	6526
sprue thickness	**SM**	13	15	mm	26	30
time cycle	**TC**	30	31	s	47	48

As shown in Figure 4.31, the dataset resulting from PLC registrations contains:

• a series of references including the MeaSetId which is a sequential progressive of the PLC print, the timestamp of the printed (TimestampLocal), the machine reference (ResourceName) which is precisely the F55, the injection number N INIEZ which is also a progressive but is occasionally reset;

• the set of values assumed by the machine’s operating parameters during the cycle (strokes of the mechanical parts C1, C2, and CC, execution times of the various phases of the T1, T2 cycle, speed V1, V2, pressures PS, PF, and PM, closing force FC, and measured sprue thickness in the SM mold).

• A third set of data is shown in Figure 4.32, which represents the maximum and minimum values of each of the previous parameters set on the machine during the setup phase (MIN_C1, MAX_C1, etc).

At first glance, the elements of interest in this dataset are the progressive of printed MeaSetId, which essentially represents a progressive of machine cycle, the printed timestamp TimestampLocal, and the values assumed by the machine parameters T1, C1, SM, etc, with the others substantially constant for the given F55 machine.

MeasSetId	CompanyI	ArticleCod	Ar	Cycl	SerialNum	Reso	Wor	TimestampUTC	TimestampLocal	N_INIEZ	C1	T1	V1	C2	T2	V2	CC	PM	PF	VA	PS	FC	SM
422548	1	ATYS		10	601	F55		31/01/2019 23:00	01/02/2019 00:00	601	286	1349	21	79	69	114	4	275	274	0	1189	5917	
422549	1	ATYS		10	602	F55		31/01/2019 23:00	01/02/2019 00:00	602	285	1348	21	80	73	110	3	274	274	0	1189	5864	
422550	1	ATYS		10	603	F55		31/01/2019 23:01	01/02/2019 00:01	603	286	1351	21	79	70	113	4	274	274	0	1189	5864	
422551	1	ATYS		10	604	F55		31/01/2019 23:01	01/02/2019 00:01	604	286	1348	21	80	70	114	4	274	273	0	1184	5884	
422552	1	ATYS		10	605	F55		31/01/2019 23:02	01/02/2019 00:02	605	285	1346	21	81	73	111	3	274	273	0	1184	5874	
422553	1	ATYS		10	606	F55		31/01/2019 23:03	01/02/2019 00:03	606	285	1343	21	81	73	111	3	274	274	0	1189	5864	
422554	1	ATYS		10	607	F55		31/01/2019 23:03	01/02/2019 00:03	607	285	1344	21	81	74	109	4	274	274	0	1189	5884	
422555	1	ATYS		10	608	F55		31/01/2019 23:04	01/02/2019 00:04	608	285	1348	21	82	74	111	4	274	274	0	1189	5885	
422556	1	ATYS		10	609	F55		31/01/2019 23:04	01/02/2019 00:04	609	285	1344	21	81	74	109	4	274	274	0	1180	5895	
422557	1	ATYS		10	610	F55		31/01/2019 23:05	01/02/2019 00:05	610	285	1344	21	80	72	111	4	275	274	0	1189	5864	
422558	1	ATYS		10	611	F55		31/01/2019 23:05	01/02/2019 00:05	611	286	1351	21	81	71	114	3	274	273	0	1184	5844	
422559	1	ATYS		10	612	F55		31/01/2019 23:06	01/02/2019 00:06	612	286	1351	21	82	73	114	3	275	273	0	1189	5844	
422560	1	ATYS		10	613	F55		31/01/2019 23:06	01/02/2019 00:06	613	286	1351	21	82	73	112	3	274	274	0	1189	5864	
422561	1	ATYS		10	614	F55		31/01/2019 23:07	01/02/2019 00:07	614	285	1339	21	80	73	110	4	275	274	0	1189	5864	
422562	1	ATYS		10	615	F55		31/01/2019 23:08	01/02/2019 00:08	615	285	1339	21	80	72	111	3	273	274	0	1180	5844	
422563	1	ATYS		10	616	F55		31/01/2019 23:08	01/02/2019 00:08	616	285	1347	21	81	71	114	4	277	272	0	1197	5854	
422564	1	ATYS		10	617	F55		31/01/2019 23:09	01/02/2019 00:09	617	286	1345	21	81	71	114	4	275	276	0	1189	5864	
422565	1	ATYS		10	618	F55		31/01/2019 23:09	01/02/2019 00:09	618	285	1340	21	81	73	111	4	274	273	0	1184	5864	
422566	1	ATYS		10	619	F55		31/01/2019 23:10	01/02/2019 00:10	619	285	1339	21	79	72	110	3	274	274	0	1189	5854	
422567	1	ATYS		10	620	F55		31/01/2019 23:10	01/02/2019 00:10	620	285	1342	21	79	71	111	3	274	274	0	1184	5854	
422568	1	ATYS		10	621	F55		31/01/2019 23:11	01/02/2019 00:11	621	285	1339	21	81	74	109	4	274	273	0	1184	5864	

Figure 4.31 - PLC Dataset

PM	PF	VA	PS	FC	SM	TC	MIN_C1	MIN_T1	MIN_V1	MIN_C2	MIN_T2	MIN_V2	MIN_CC	MIN_PM	MIN_PF	MIN_VA	MIN_PS	MIN_FC	MIN_SM	MIN_TC	MAX_C1	MAX_T1
275	274	0	1189	5917	21	34	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1189	5864	22	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1189	5864	21	34	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	273	0	1184	5884	20	34	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	273	0	1184	5874	21	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1189	5864	21	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1189	5884	20	34	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1189	5885	19	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1180	5895	20	34	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
275	274	0	1189	5864	21	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	273	0	1184	5844	20	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
275	273	0	1189	5844	20	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1189	5864	19	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
275	274	0	1189	5864	21	34	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
273	274	0	1180	5844	22	34	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
277	272	0	1197	5854	19	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
275	276	0	1189	5864	19	34	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	273	0	1184	5864	20	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1189	5854	23	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	274	0	1184	5854	22	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
274	273	0	1184	5864	20	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678
273	272	0	1180	5854	22	33	212	1300	14	95	90	82	1	250	205	0	830	5000	18	31	288	1678

Figure 4.32 - Ranges of the machine set-up values

4.5.2. MES

MES stands for Manufacturing Execution System and is a system that collects and distributes information to allow for the optimization of production activities from order placement to finished product. Using real-time, up-to-date, and accurate data, the MES guides, responds, and informs about plant and production department activities as they occur. The resulting reaction time, combined with the emphasis on reducing non-value-added activities, propels plant operations and processes to maximum efficiency. By combining the optimization of production and logistics processes with the control of resource availability and product quality, the MES System bridges the “gap” between ERP systems (administrative and accounting) and machine control systems (PLC, SCADA MES).
Indeed, the lack of interaction between ERP and Control Systems turns the production process into a black box: you know the inputs and outputs of production but have no visibility into what happens in between.
MES is thus the critical component that enables system communication and provides complete visibility into what is happening in production.
Figure 4.33 depicts the dataset obtained from the MES system and containing the causes imputed by the operators in conjunction with the occurrence of an anomaly. In this dataset, the recordings have a ‘date time’ timestamp as well as the identifier of the machine to which they refer (Resource field) and a series of ancillary information such as the item produced, the number of pieces produced, the names of the machine operators, production execution times, and so on. Finally, the most important information is represented by the content of the ‘casualisation’ field and its ‘suspension description’, which reports any anomalies with associated causal factors.
The interesting aspect to notice is that there are multiple records with the same timestamp because the information is repeated for each operator and for each article code produced during the same production phase, and thus the causal anomaly occurs several times with the same timestamp. Because the information that is of

interest in this first phase is the reason for fault and the time reference at which it occurred, the first processing of this information aims to make the recordings that report the timestamp with the associated fault cause unique. A second important factor is that the MES system records a causal “99 – Fermo Macchina Generico” every time the machine stops, which is then better specified by the operator when the machine restarts (after a few minutes) with the actual fault cause. This implies that the same machine downtime will generally report two associated reasons for downtime, of which the generic reason 99 must be ignored in favour of the operator-recorded reason (see the example in Figure 4.34).

Giorno Tu	Data ora	St	Atti	Riso	Centr	Operat	Ordine	Articol	T Prep	T Lav	T Man	T Fermo	Qta Bu	Qta Sc	N Battute	T Prep Uo	T Lav Uom	Squa	Caus	Descrizione sospension	N Tu	N
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3114	20190006517	10A012M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	6
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3076	20190006517	10A012M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	6
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3114	20190006519	10A002M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3076	20190006519	10A002M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3114	20190006514	10A010F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3076	20190006514	10A010F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3114	20190006518	10A002F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3076	20190006518	10A002F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3114	20190006516	10A012F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	6
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3076	20190006516	10A012F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	6
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3114	20190006515	10A010M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 06:00:00	R	L	F55	RPRES	3076	20190006515	10A010M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3076	20190006517	10A012M	0	0	0	0	0	0	0	0	0,320564236	SQPRA	99	Fermo Macchina Generico	1	6
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3114	20190006517	10A012M	0	0,756612292	0	0	2376	0	74,25	0	0,320564236	SQPRA	99	Fermo Macchina Generico	1	6
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3076	20190006519	10A002M	0	0	0	0	0	0	0	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3114	20190006519	10A002M	0	0,630510243	0	0	1980	0	61,875	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3076	20190006514	10A010F	0	0	0	0	0	0	0	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3114	20190006514	10A010F	0	0,630510243	0	0	1980	0	61,875	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3076	20190006518	10A002F	0	0	0	0	0	0	0	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5

Figure 4.33 - MES Dataset

01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3076	20190006518	10A002F	0	0	0	0	0	0	0	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3114	20190006518	10A002F	0	0,630510243	0	0	1980	0	61,875	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3076	20190006516	10A012F	0	0	0	0	0	0	0	0	0,320564236	SQPRA	99	Fermo Macchina Generico	1	6
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3114	20190006516	10A012F	0	0,756612292	0	0	2376	0	74,25	0	0,320564236	SQPRA	99	Fermo Macchina Generico	1	6
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3076	20190006515	10A010M	0	0	0	0	0	0	0	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5
01/04/2019	01/04/2019 10:02:06	S	L	F55	RPRES	3114	20190006515	10A010M	0	0,630510243	0	0	1980	0	61,875	0	0,267136863	SQPRA	99	Fermo Macchina Generico	1	5
01/04/2019	01/04/2019 10:05:28	R	L	F55	RPRES	3076	20190006517	10A012M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	6
01/04/2019	01/04/2019 10:05:28	*	L	F55	RPRES	3114	20190006517	10A012M	0	0	0	0,010563021	0	0	0	0	0	SQPRA	400	MICROFERMI	1	6
01/04/2019	01/04/2019 10:05:28	R	L	F55	RPRES	3114	20190006517	10A012M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	6
01/04/2019	01/04/2019 10:05:28	R	L	F55	RPRES	3076	20190006519	10A002M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 10:05:28	*	L	F55	RPRES	3114	20190006519	10A002M	0	0	0	0,008802517	0	0	0	0	0	SQPRA	400	MICROFERMI	1	5
01/04/2019	01/04/2019 10:05:28	R	L	F55	RPRES	3114	20190006519	10A002M	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 10:05:28	R	L	F55	RPRES	3076	20190006514	10A010F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 10:05:28	*	L	F55	RPRES	3114	20190006514	10A010F	0	0	0	0,008802517	0	0	0	0	0	SQPRA	400	MICROFERMI	1	5
01/04/2019	01/04/2019 10:05:28	R	L	F55	RPRES	3114	20190006514	10A010F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 10:05:28	R	L	F55	RPRES	3076	20190006518	10A002F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5
01/04/2019	01/04/2019 10:05:28	*	L	F55	RPRES	3114	20190006518	10A002F	0	0	0	0,008802517	0	0	0	0	0	SQPRA	400	MICROFERMI	1	5
01/04/2019	01/04/2019 10:05:28	R	L	F55	RPRES	3114	20190006518	10A002F	0	0	0	0	0	0	0	0	0	SQPRA	0		1	5

Figure 4.34 - Double fault cause

It is clearly evident that raw data from PLCs, MES, and possibly additional sensors requires extensive cleaning and pre-processing before being exploited by ML algorithms.

4.6. DATA INTEGRATION and INTEGRATED DATASET CHARACTERISTICS

Data integration necessitates the creation of a data acquisition platform with a database capable of hosting the datasets acquired by various machines and plants, as well as the design and implementation of a series of connectors and data acquisition procedures capable of bringing field data into this database in a reasonable timeframe.

Direct connectors are used for data acquisition from SCADA and MES systems, as well as data transfer from involved systems to allow for timely analysis of the information generated by these systems. This timeliness is critical so that predictive activity about potential malfunctions or process criticalities reaches the operator in time for preventive actions to be implemented.

The following figure depicts a complete picture of the location of IoT architecture in the Smart Manufacturing system paradigm.

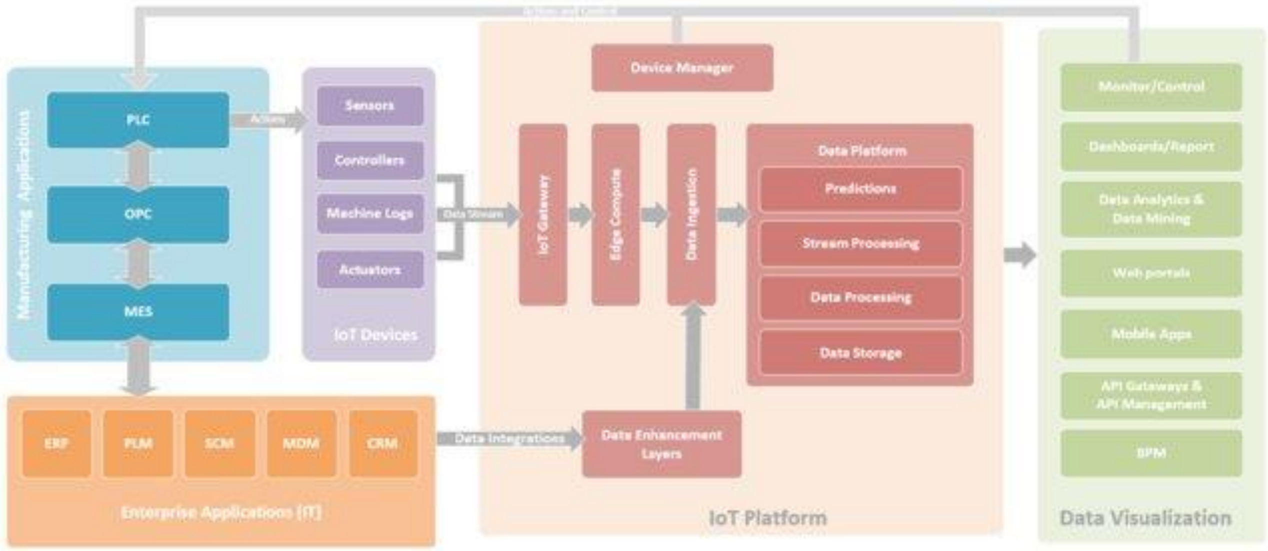


Figure 4.35 -IoT reference architecture in smart manufacturing (Illa and Padhi, 2018)

The die-casting process involves a number of variables that, on occasion, result in the production of defective castings. Productivity is lost as a result of the defects (i.e., time and cost). Identifying and removing defective castings is crucial to die-casting quality control.

Many case studies of data integration on machines dedicated to the Die-Casting Process have been published in the literature, revealing that simply monitoring the process's processing parameters is insufficient for identifying and predicting problems that may arise throughout the process. (Liu et al., 2018) provided an excellent example of monitoring die-casting energy use: an IoT system enabled technique for streaming online energy data for energy analysis of a die casting machine is presented. Real-time Ethernet was used to send energy data gathered by digital power meters and PLCs to a central server. To interpret the data and evaluate the performance of a die casting machine, a set of indicators was developed, including energy per part and energy per action. (Park et al., 2019) established an IoT system on a die-casting environment to build a link between casting parameters and production quality.

The innovative part lies in the approach used during the doctoral path that differs from the other case studies that are in the literature. Data analytics has been used to undertake some study on the methods and tools required to discover flaws in the die-casting process For example (Lee et al., 2017) have implemented analysis techniques on die-casting to address quality issues to reduce products' defect and mitigate financial burdens due to excessive expenditure of times and costs. Artificial neural networks are used by (Soundararajan et al., 2015) to prevent casting faults. (Zheng et al., 2009) use neural networks to optimize the high-pressure die-casting process parameters; in their study, an artificial neural network is introduced to generalize the correlation between surface defects and die-casting parameters such as mold temperature, pouring temperature, and injection velocity.

From the standpoint of scientific research, the issue of data quality has been tackled in any sense with the advent of I4.0 and digitization in general.

The sections that follow will explain the key approaches to the problem of analysing and quality data in relation to the topics discussed in this paper, organized by macro approaches:

- Data Collection Process Verification and Improvement: It consists of an in-depth review of the entire data collection process with the aim of identifying and enhancing the phases of the process that threaten the data's quality (eg, modifying the data entry process from manual to automatic). It is important to keep in mind two points in this context: (1) In certain real-world situations, intervening or analysing the collection process is not feasible. (2) The intervention, with potential change, of the collection process does not guarantee the quality of the data: a quality review is often required to confirm how the intervention on the collection process has resulted in an increase in the quality of the data collected.

- Integration: Allows for the integration of data by comparing it to real-world counterparties. However, since this approach is highly costly, both financially and in terms of time, it is restricted to small databases.
- Record Linkage (Fan, 2008): (also known as object identification, record matching, merge-purge problem) it entails comparing records that contain information on the same item but come from different databases (eg, data from different sources are crossed). It is important to have a key (link) that allows one or more records from different databases to be associated. Unfortunately, having databases with these characteristics is not always possible. In these cases, the technique can be used, but it is necessary to perform. (1) the Cartesian product of the domains of the two source databases (that is, all possible ways of associating all the elements of one database with all the elements of the other are generated), resulting in a large space that must be further refined; (2) Next, describing a subset of the search space on which to concentrate attention, and (3) finally, defining a decision-making model (e.g., a "distance" function) that determines whether two records from two distinct databases are similar or less related. As one would expect, the approach's applicability is constrained by the size of the search space and the proper calibration of the decision model, on the basis of which false positives/negatives can be obtained.
- Domain rules and constraints: In many real-world applications, the solution focused on data integration and record matching is inapplicable or too computationally costly. In these cases, an inspection and correction approach is used. It is the most widely used technique, and therefore the one with the most variations and improvements (Fursova, 2018). It distinguishes approaches based on (1) domain knowledge-derived rules (e.g., domain experts identify algorithms and ad-hoc tools for data analysis and quality), (2) dependencies (formalized by domain experts), and (3) learning (Neural Networks, Pattern Recognition, clustering)*:*
- ETL (Extract, Transform, Load): The term ETL refers to a process that consists of three phases (Ali and Wrembel, 2017): (1) data extraction from one or more sources; (2) data transformation (including data quality assessment and cleansing); and (3) data loading into the final database / data warehouse. In most cases, the data quality process is carried out with the assistance of business regulations (rules defined by the domain expert to manage the data quality phase). One of the challenges of this approach is the formalization of business rules, as well as the control of the side-effects (i.e., the effects that are not immediately detectable) that the implementation of these rules can have on the manipulated data. In this regard, a plethora of data profiling tools are available on the market, allowing you to conduct extensive data analysis based on the chosen dimensions.

Referring to the test case presented in this dissertation, Figure 4.36 depicts a diagram of a real-time data integration system derived from multiple platforms / sensors.

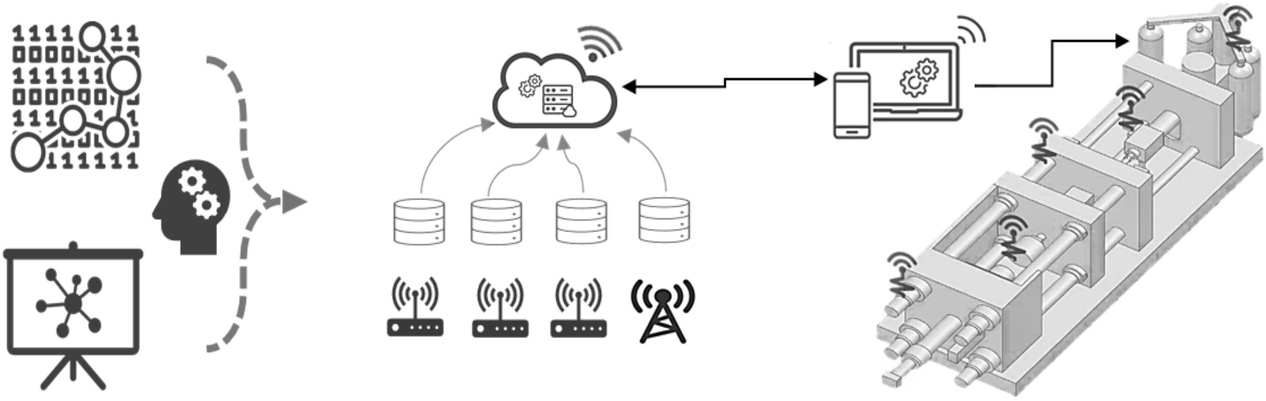


Figure 4.36 - Data Integration Prototype designed for die-casting test case

In addition to the PLC, the machine has been outfitted with additional sensors to monitor the parameters that are fundamental results during the construction of the process's thermodynamic model. PLCs, MESs, and sensors communicate with a cloud system, where they transfer the acquired data to be pre-processed (ETL) and molded based on the data analysis algorithm that will be implemented. To this end, Figure 4.37 depicts the entire date modelling procedure as a result of data acquisition via the designed IoT system.

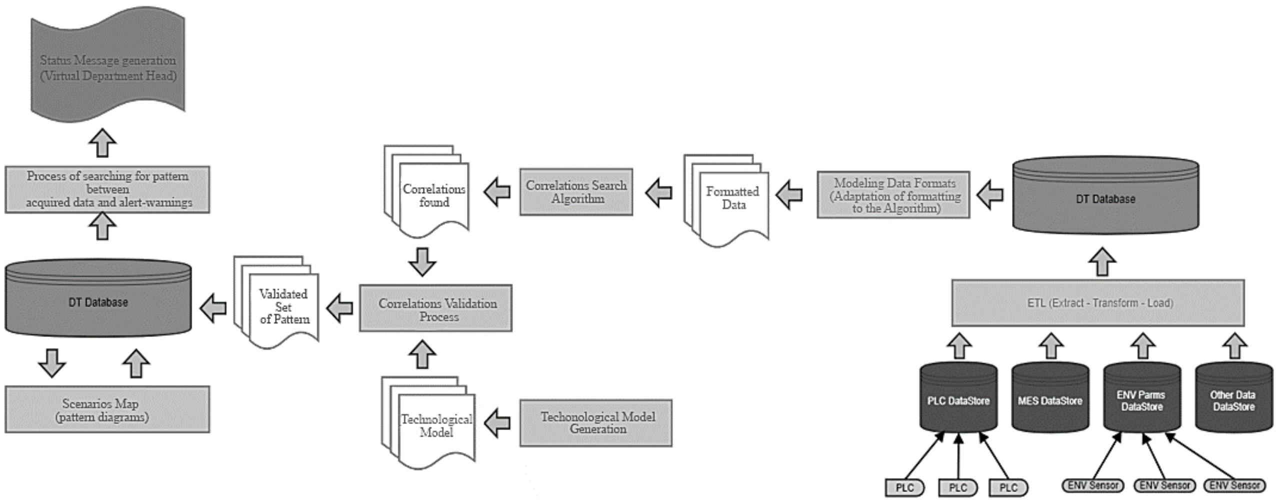


Figure 4.37 – Data modelling process following the data integration procedure

The previous analysis for the F55 machine shows that the two data sets are significantly disconnected from each other, and the only element that allows to relate the prints detected by the PLC with the reasons for downtime recorded on the MES system is represented by the resource code (F55) and the timestamp. It is also obvious that, in order to connect the printouts of the machine parameters detected by the PLC to a reason for stopping while avoiding cartesian products (duplications) between the data, the timestamp causal fault must be unique on the MES dataset. On the latter's data, the records with causal nr 99 must also be removed, taking into account only the actual reasons attributed by the operator when the machine is restarted. Finally, because it is a matter of predicting machine downtime, measures that represent the event object of the prediction must be identified. Specifically, after consolidating the PLC data with the causal from the MES via timestamps, it is considered convenient to use the minutes missing at the next machine stop as a reference measure for each PLC record by labelling the data with the reason for the immediately following machine downtime.
The decision to label the PLC records with the cause for the first machine stop immediately following stems from the desire to attribute the values of the parameters measured by the PLC to the event represented by the machine downtime immediately following. As a result, it will be possible to separate the records derived from the PLC readings by homogeneous types of machine downtime (microstops, mechanical fault, electrical fault, and so on), as it is expected that the behaviours for each type of stop will differ, and thus each type of event must be analysed separately.
As shown in Figure 4.38, machine downtime is a case study with a high rate of occurrence, and among all the reasons for blocking, the one that occurs the most is undoubtedly represented by microstops.
The causal effects with the highest frequency of occurrence immediately following (change of work and general break) are physiological in nature, as they relate to shift changes or machine equipment, whereas other causes of accidental downtime, such as mechanical fault, electrical fault, equipment fault, and so on, will also be investigated.

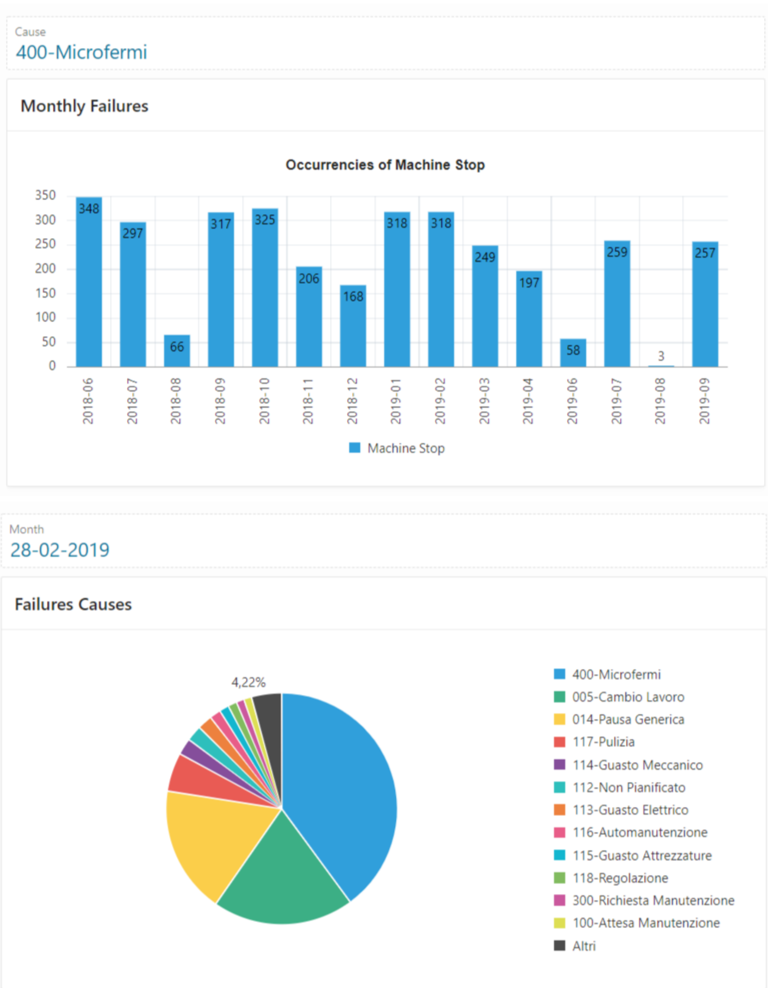


Figure 4.38 - Occurrences of machine downtime causes

All pre-processing and analysis procedures were performed on a complete dataset of 644'989 data records spanning 30 months, from August 2018 to December 2020. It contains all integrated and matched data from the PLC and the MES pertaining to the F55 machine.

The complete database consists of 14 features (independent variables), which are PLC process parameters: C1, T1, V1, C2, T2, V2, CC, PM, PF, VA, PS, FC, SM, and TC, even though VA will not be taken into account because its value is regularly set to zero, so it will only may create unbalances to the dataset during normalization or analysis. Finally, depending on whether it is classification or regression, respectively the causalisation or the missing time at the next company, regardless of whether is among those specified, the targets will be different.

In certain cases, the timestamp also serves as a feature. The total number of causalisations (i.e., anomaly descriptions) counted in the database is 12 and corresponds to the dependent variables: microstops, mechanical fault, electric fault, equipment fault, pending operator, self-maintenance, processing shift, unplanned, generic pause, cleaning, adaptation, and maintenance request. Only the first four causalisations will be investigated in the next paragraph since they are the most representative and directly depend on machine and process issues.

Regarding the remaining causals, as mentioned below, it is important to explain why they were not taken into account in the tests, classifications, and regressions: a as previously stated, the data from both the PLC and the MES are "tampered" or "dirty" by the operators as the machine generates an alarm when at least one parameter (all or even just one can be monitored, at the option of the shift operator) of the monitored process is outside of the pre-set ranges. From the time the alarm occurs, the operator has a small window of time to intervene and ensure that the process (the cycle) is completed successfully. If the operator is late, the machine stops, which increases the cycle time of the printed one. The MES detects a delay or halt in the cycle time when

compared to the optimal one and "invites" the operator to justify this delay or stop. The operator must then choose one of the above-mentioned circumstances.
In light of what has been discussed, among all the causes of downtime, there are those that are not dependent on the process's quality or the thermodynamic parameters that it takes during the cycle, and they are:

- Pending operator: the company has set up three shift bands for the die-casting section, morning-afternoon-night. This means that between one turn and the next, the operator could be absent or not fully prepared to operate on the machine if it gave alerts. Usually, at the end of the turn, the operator pauses the machine until the next operator is returned.
- Self-maintenance: because the machine is programmed to make each n cycle an internal control, it is necessary to lock between one cycle and the next. However, it does happen that the most experienced operators can detect a problem with the machine even before it stops, and in order to avoid excessive waste and material consumption, they interrupt the cycle themselves to be able to intervene. At the end, the cycle resumed from where it had been interrupted, sometimes successfully but with a substantially longer cycle duration, and sometimes discarding printed if quality final checks were not met.
- Maintenance request: can be the result of a failed machine maintenance or when an intervening operator fails to fix the problem. In both circumstances, the specialized maintainers must be contacted, and the time spent waiting for him to arrive the machine must be stopped. As a result, this halt must be justified.
- Cleaning: similar to self-maintenance, the machine must be cleaned of waste and residues after n cycle. As a result, the machine is shut off or stopped.
- Processing shift: each machine is not programmed to conduct a single process for a single product, but it is required to perform productions and a variety of components that are all extremely distinct from one another (just think of the difference between a Safety Pin and a Steel Corner). When the production program changes and the products to be cast change, the machine must be stopped, the new optimal process parameters for that processing established, and resumed.
- Adaptation: it was discovered while examining the dataset that this causalisation is directly related to the processing shift. Because the machine can sometimes execute a "break-in" cycle on the new processing, the initial cycles are not always successful.
- Unplanned: this causalisation is unintentional because it is set when the operator forgets to trigger the downtime. It is impossible to know what this capture linked to in this circumstance.
- Generic pause: when the operator needs to leave the department or the car, it is his responsibility to suspend the procedure till he returns. As a result, this causalisation is neither systematic or recurring, and it can be identified by a machine learning model trained on a data set connected to generic pause.

As a result, they will not be used in the analyses because they may impact negatively during data training and influence the model's performance.
By removing, from the dataset, the records linked to causalisations that will not be evaluated, it moves from 644'989 data records to 346'230 data records to be still cleaned, pre-processed, and analysed.

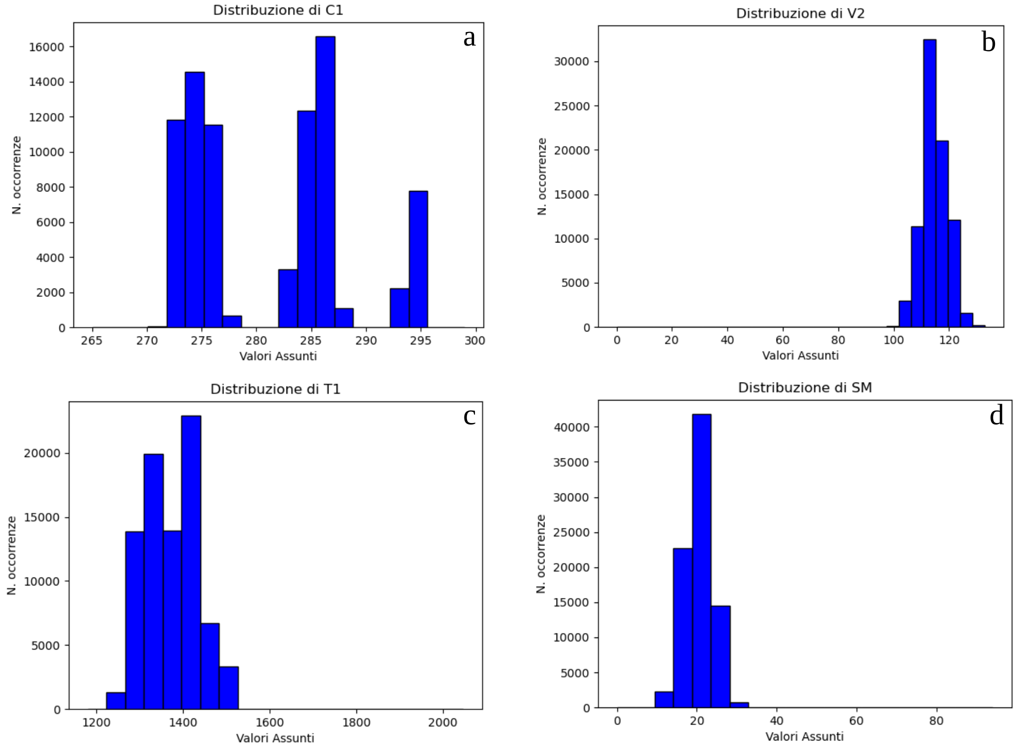
4.7. MACHINE LEARNING TECHNIQUES ON THE TEST CASE

The following approach consists of a two-phase analysis of the process's historical data: first, a characterization of the combinations of parameters that determine an alert system of the machine in relation to normal operation, and then, a prediction test on the time missing at a subsequent alarm based on the combination of values of this set of parameters.
The insight has been conducted by scikit-learn (Scikit-learn, 2020). Scikit learn is a Python library for running machine learning. Scikit-learn is an open-source library licensed under the BSD license that can be used in a variety of contexts, encouraging both academic and commercial use, in fact it is constantly updated. In Python, it provides a variety of supervised and unsupervised learning algorithms. This makes it simple to solve

regression, classification, and clustering algorithms. Data transformation, feature selection, and ensemble methods can all be accomplished in a few rows. It also enables you to easily retrieve data from a variety of sources, including SQL databases, text, CSV, Excel, JSON files, and many other less common formats. Once the data is in memory, there are dozens of operations available to analyse, transform, retrieve missing values, and clean the dataset, as well as SQL-like operations and a set of statistical functions to perform even the most basic analysis.
The first step was to query the data and determine the distribution of casualisations and parameters for each of them.

4.7.1. Microstop

The most significant case history is represented by microstops. The preliminary phases of analysis performed on microstops are described below, and they have been replicated for subsequent casualisations, but they will not be described in the thesis for the sake of brevity.
The first step is to examine the data pertaining to the values assumed by the features to determine whether or not the statistical distribution of the values is similar to a normal distribution, whether or not there are any outliers (statistically anomalous values), and so on. Figure 4.39 depicts the statistical distribution of the main feature's values, demonstrating that the distribution of values cannot be assumed always to be Gaussian, and thus standardization techniques based on the normal distribution cannot be applied to these variables (see, e.g., C1 or C2). Figure 4.39b) depicts the statistical distribution of the values of the feature V2, from which we can deduce a generally Gaussian trend in the distribution of values, as indicated by the horizontal axis extension, but also the presence of outliers. In fact, if we examine the values more closely, we find 91 instances of V2 with values less than 50 when we use a filter that extracts only values less than 50. T1 value distributions are also reported, with SM having a roughly normal distribution.



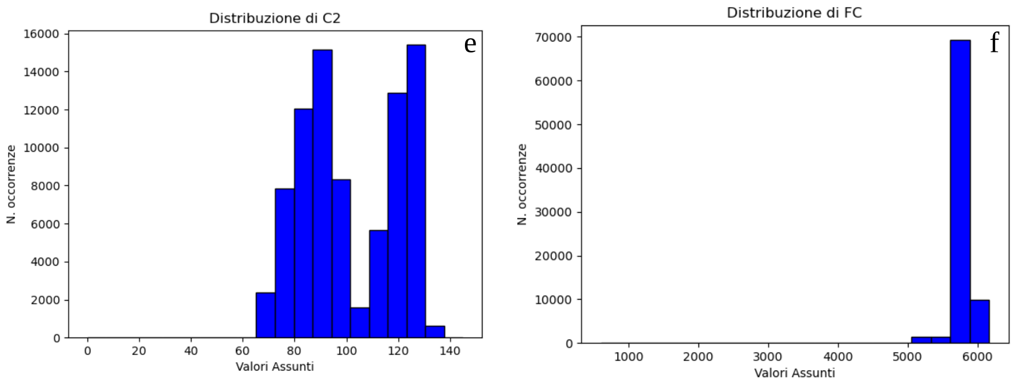


Figure 4.39 - Statistical distribution of feature: a) C1, b) V2, c) T1, d) SM, e) C2, f) FC

The scale of the histogram for SM indicates that there are also outliers in this case, as shown in Figure 4.40.

```
>>> filtroSM = datiMF['SM'] > 50
>>> datiMF[filtroSM]
       CAUSALE DESCRIZIONE_SOSPENSIONE  MIN_MANCANTI  N_CICLI_MANCANTI   C1    T1  V1  C2   T2  V2  CC   PM   PF  VA    PS    FC  SM   TC
23         400              MICROFERMI            59                84  292  1389  21   0  201   0   6    5    4   0    17  5758  92   62
24         400              MICROFERMI            56                83  290  1337  21   0  254   0   6    5    4   0    17  5801  94  217
101413     400              MICROFERMI            91               134  278  1541  18  59  237  25   2  236  235   0  1019  5590  57   37
>>>
```

Figure 4.40 - Outliers highlighting

In this case, there are a few outliers that clearly coincide with the upcoming conditions to the fault because, in addition to having a particularly low FC (the closing force of the molds), there is also an outlier for the parameter that represents the metal thickness, SM (the thickness of the sprue) equal to 0.

4.7.1.1. Effect of Outliers

An "outlier" in a data set is one that is abnormal or, in any case, isolated from the rest of the data set. Instinctively, the question of whether the presence of the unavoidable outliers in the data can affect, and possibly to what extent, the generation of the predictive model and the quality of the results produced by it arises.

Furthermore, it should be noted that the dataset covered by this document almost certainly contains abnormal data, which is frequently generated by restarting situations of the machine; the presence of this data, among other things, poses a problem if data normalization techniques are to be used. In fact, without the features of a normal statistical distribution (Gaussian) in many cases, scaling techniques based on data normalization are not applicable, but the presence of outliers has a strong influence on the different types of scaling techniques (e.g., min-max). It may then be necessary to be able to identify outliers, assess their impact on predictive models, and potentially remove them.

Python provides a functional technique for identifying outliers with the "LocaloutlierFactor" method, which is based on measuring the local density deviation of a given data sample with respect to neighbouring data. As a result, the method calculates the Euclidean distance between each element and its neighbours. Distance is used to calculate density, and what is "isolated" is an element when compared to its neighbours. Outliers can be identified by applying the method to our data and assigning a threshold (quantile) below which the element can be considered isolated (see the example in Figure 4.41).

```
from sklearn.neighbors import LocalOutlierFactor
from sklearn.preprocessing import QuantileTransformer

model = LocalOutlierFactor(n_neighbors=20)

model.fit_predict(X1)
lof = model.negative_outlier_factor_

thresh = np.quantile(lof, .003)
print(thresh)
-1.5651623634758003


index = lof<=thresh
values = X1[index]

X1.shape
X2=X1[index]

>>> X2=X1[index]
>>> X2.shape
(15, 13)
>>> X2
```

	C1	T1	V1	C2	T2	V2	CC	PM	PF	PS	FC	SM	TC
4219	272	1292	21	130	108	120	8	247	246	1067	5853	17	343
55874	273	1220	22	128	112	114	6	239	238	1032	5612	20	36
55877	273	1215	22	127	112	113	6	239	237	1028	5633	21	36
55878	273	1218	22	126	112	113	6	241	237	1041	5633	22	36
56175	274	1328	21	129	112	118	5	236	234	1019	5790	19	328
56285	273	1339	20	129	114	113	7	218	215	933	5791	18	129
139668	299	1383	21	73	198	37	3	239	236	1024	6117	21	33
158364	294	1477	20	71	70	101	4	212	211	915	5642	21	274
172188	275	1407	19	125	108	116	5	233	232	1006	5591	20	174
172230	278	1457	19	122	352	35	5	240	239	1037	5327	20	521
210730	286	1346	21	94	80	118	4	274	268	1189	5622	12	123
252179	284	1326	21	86	78	110	4	272	271	1176	5770	22	144
275157	276	1344	20	97	86	113	4	267	266	1154	5875	19	181
347186	275	1022	26	94	252	37	3	239	234	1015	6001	24	37
347187	342	1134	30	26	86	30	3	240	237	1028	6012	25	36

Figure 4.41 - Example of identifying outliers with scikit-learn

In the preceding example, a quantile threshold value of 3 per 1000 was defined, which corresponds to a threshold value of -1.56 in the measurements of the distances between the elements.

The application of this threshold filter yields 15 records from the mechanical failure dataset that are considered outliers as output. The values corresponding to the outliers are highlighted in red in the following diagram (Figure 4.42), which reports the values of two of the features.

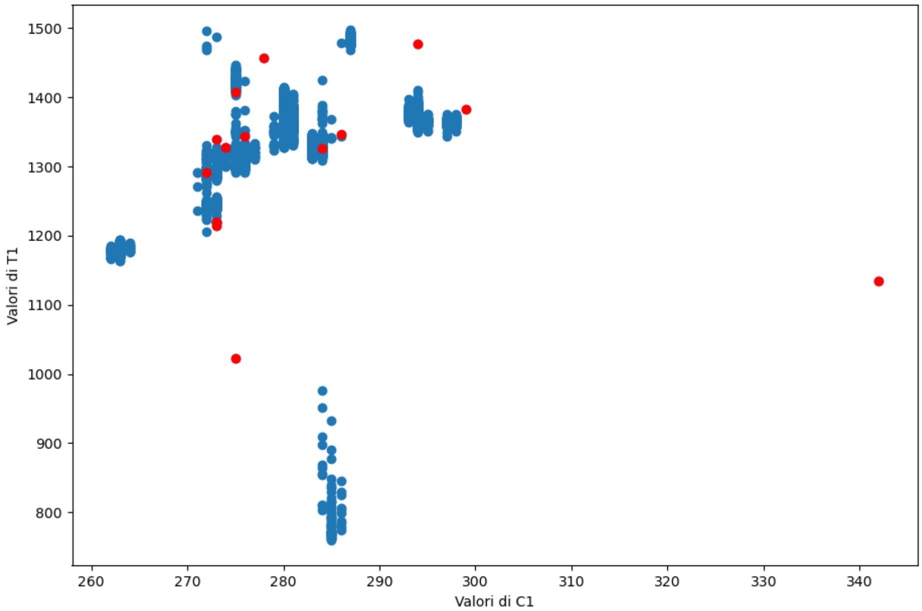


Figure 4.42 - A graphic example of identifying outliers (in red)

To assess the impact of possible outlier removal from the dataset, we hypothesize raising the quantile that defines the 1 percent threshold to -2.054. This threshold determines 50 outlier values that we will remove from the feature and label data by generating a new dataset of "X2" features and an array of "Y2" labels. We then proceed as before, determining the tests of tests and trains and training the model (using the Random Forest with the previously determined optimal parameters) on this data.

A subsequent analysis can be performed by representing the trend of the PLC parameters in relation to the number of cycles remaining before the stop, which constitute the labels of the dataset being analysed. Almost all of the parameters exhibit a specific behaviour close to 0, indicating that they begin to show anomalous values when they are nearing the end of the machine's cycle before stopping. It should also be noted that these behaviours, in many cases, are not limited to the proximity of the zero but also occur when many tens of cycles away from the zero (i.e., when one is still far from the stop), and thus, taken individually, they are unable to predict the stop with sufficient reliability.
Another point to consider is that the specific behaviours, which are not particularly visible in the sample in question, manifest themselves starting a few cycles away from zero and, because each cycle corresponds to 30 seconds, it is a matter of behaviours that occur at most in a few minutes before the microstops. In the case of TC, the behaviour is especially noticeable, but only when it is practically close to zero (which is of little significance if we consider that TC represents the cycle time and that this value will probably necessarily tend to immediately increase close to the stop). As a result, the time interval is insufficient to issue a notice sufficiently in advance to take measures capable of preventing the arrest. It can be concluded that the data analysis indicates that the algorithm we are going to identify will most likely be able to determine patterns that precede the micro-stop, but, as it is easy to hypothesize, the predictability of this event will involve a low rate of reliability. Figure 4.43 depicts the graph that describes this situation.

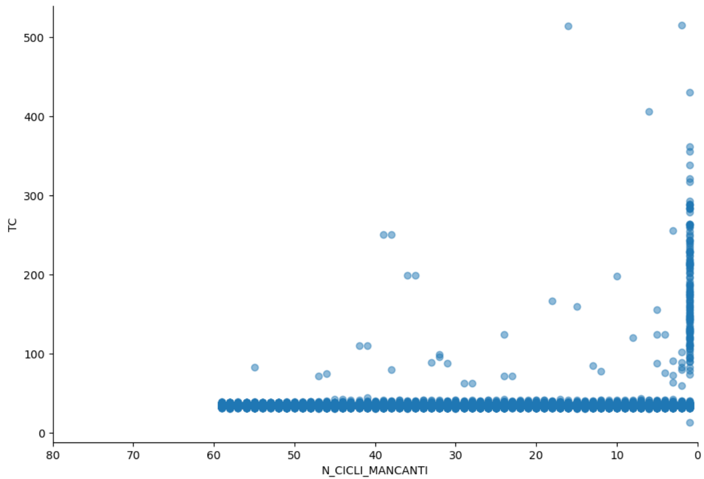


Figure 4.43 - Variation of TC vs missing cycles number

The correlation matrix and the covariance matrix, shown in the following Figure 4.44and Figure 4.45, can be used gain additional insights.

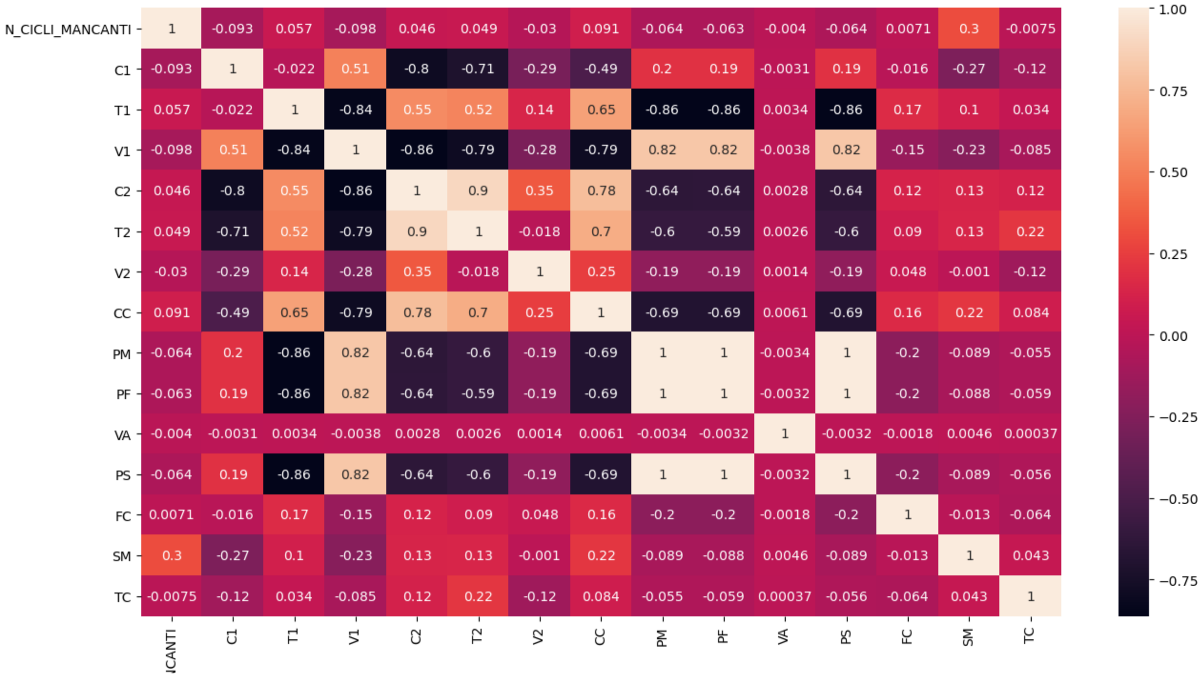


Figure 4.44 - Correlation matrix between missing cycles number and the features

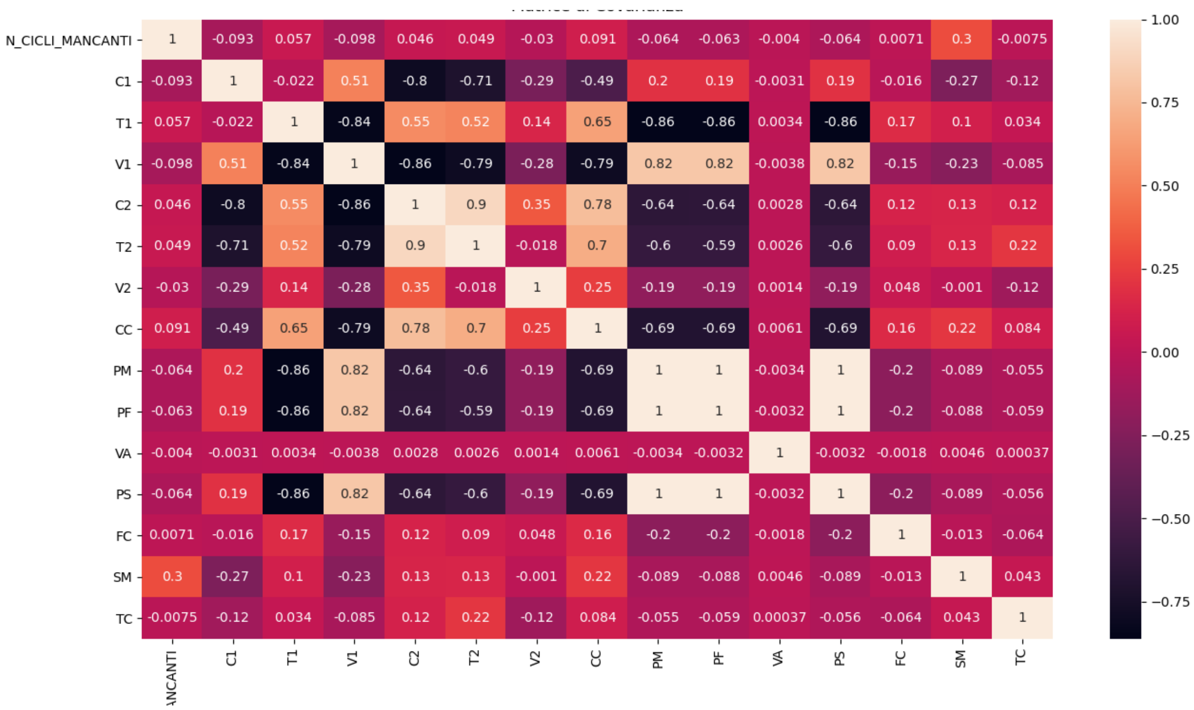


Figure 4.45 - Covariance matrix between missing cycles number and the features

As is well known, when the correlation values tend (in absolute value) to 1, it indicates that there is a direct (positive values) or inverse (negative values) correlation, or that there is a linear relationship between related values that increases in the first case and decreases in the second. It should also be noted that the absence of correlation (positive or negative values close to zero) does not necessarily imply that the two random variables are independent, but only that there is no linear dependence (but could, for example, be quadratic and, in this case, the correlation would be close to zero).

In our case, both the correlation matrix and the covariance matrix show a weak correlation between SM and the number of missing cycles at stop, as well as an equally weak correlation between SM and C1, V1, and CC.

Overall, the correlation matrix and covariance matrix do not appear to provide useful or meaningful indications because the only obvious correlations are those that relate the run of the second phase (C2) to the relative phase time (T2) so that, as the race lengthens, the time also increases consistently; similarly, the pressures (PS, PF, PM) are significantly correlated with the speed of the first phase V1.

4.7.1.2. Feature selection

Feature selection is a method of analysis that identifies the most significant parameters that determine the values of the labels based on the data (in the case of supervised algorithms).
For the current case, represented by microstops analysis, we will employ the supervised algorithm known as "K-best." The first step is to load the period data into the dataset named "dati", which contains period records for all causal causes, as shown in Figure 4.46 below.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

dati = pd.read_csv("/dati/DATI_COMPLETI_2019T1_v2_RIDOTTO.csv", skiprows =0, sep=';')

>>> dati = pd.read_csv("/dati/DATI_COMPLETI_2019T1_v2_RIDOTTO.csv", skiprows =0, sep=';')
>>> dati
        CAUSALE DESCRIZIONE_SOSPENSIONE  MIN_MANCANTI  N_CICLI_MANCANTI   C1    T1  V1  C2  T2   V2  CC   PM   PF  VA    PS    FC  SM   TC
0            14          Pausa Generica            11                20  299  1511  19  94  97   97   1  263  253   0  1098  5800   0  335
1            14          Pausa Generica            11                19  298  1521  19  95  97   98   1  262  251   0  1089  5779   0   40
2            14          Pausa Generica            10                18  298  1526  19  94  99   95   1  259  248   0  1076  5821   0   34
3            14          Pausa Generica             9                17  298  1527  19  94  98   96   1  255  246   0  1067  5800   0   33
4            14          Pausa Generica             9                16  298  1531  19  93  96   97   1  255  246   0  1059  5833   0   34
...         ...                     ...           ...               ...  ...   ...  ..  ..  ..  ...  ..  ...  ...  ..   ...   ...  ..  ...
139116       14          Pausa Generica             3                 5  276  1295  21  94  76  124   4  273  273   0  1184  5686  22   37
139117       14          Pausa Generica             3                 4  275  1291  21  95  78  122   4  273  272   0  1180  5675  22   37
139118       14          Pausa Generica             2                 3  275  1292  21  96  79  122   4  272  272   0  1180  5717  21   38
139119       14          Pausa Generica             2                 2  275  1293  21  96  79  122   4  274  273   0  1184  5676  21   36
139120       14          Pausa Generica             1                 1  276  1298  21  95  76  125   4  273  272   0  1180  5738  21   37

[139121 rows x 18 columns]
>>> dati.head(100)
    CAUSALE DESCRIZIONE_SOSPENSIONE  MIN_MANCANTI  N_CICLI_MANCANTI   C1    T1  V1  C2  T2   V2  CC   PM   PF  VA    PS    FC  SM   TC
0        14          Pausa Generica            11                20  299  1511  19  94  97   97   1  263  253   0  1098  5800   0  335
1        14          Pausa Generica            11                19  298  1521  19  95  97   98   1  262  251   0  1089  5779   0   40
2        14          Pausa Generica            10                18  298  1526  19  94  99   95   1  259  248   0  1076  5821   0   34
3        14          Pausa Generica             9                17  298  1527  19  94  98   96   1  255  246   0  1067  5800   0   33
4        14          Pausa Generica             9                16  298  1531  19  93  96   97   1  255  246   0  1059  5833   0   34
..      ...                     ...           ...               ...  ...   ...  ..  ..  ..  ...  ..  ...  ...  ..   ...   ...  ..  ...
95      400              MICROFERMI            13                12  294  1448  20  77  72  107   3  243  242   0  1050  5842  16   32
96      400              MICROFERMI            13                11  294  1455  20  75  71  106   4  243  242   0  1050  5821  17   34
97      400              MICROFERMI            12                 9  294  1458  20  77  70  110   3  243  242   0  1050  5842  16   33
98      400              MICROFERMI            12                10  295  1452  20  76  70  109   3  243  242   0  1050  5843  16   32
99      400              MICROFERMI            11                 8  294  1455  20  77  71  108   4  243  242   0  1050  5843  15   33

[100 rows x 18 columns]
```

Figure 4.46 - Loading data

The following activity entails creating a filtered dataset called "datiMF" from the "dati" dataset by extracting only 400 causalisations relative to the microstops, as shown in Figure 4.47.

```
filtro_MF = dati['CAUSALE']==400

datiMF=dati[filtro_MF]

>>> filtro_MF = dati['CAUSALE']==400
>>> datiMF=dati[filtro_MF]
>>> datiMF.head(200)
     CAUSALE DESCRIZIONE_SOSPENSIONE  MIN_MANCANTI  N_CICLI_MANCANTI   C1    T1  V1  C2   T2   V2  CC   PM   PF  VA    PS    FC  SM   TC
20       400              MICROFERMI            91                86  299  1505  20  68   74   92   3  246  242   0  1054  5917  20   33
21       400              MICROFERMI            91                87  299  1508  19  68   74   92   3  246  242   0  1050  5906  20   33
22       400              MICROFERMI            90                85  299  1520  19  67   75   89   4  247  243   0  1054  5927  20   34
23       400              MICROFERMI            59                84  292  1389  21   0  201    0   6    5    4   0    17  5758  92   62
24       400              MICROFERMI            56                83  290  1337  21   0  254    0   6    5    4   0    17  5801  94  217
..       ...                     ...           ...               ...  ...   ...  ..  ..  ...  ...  ..  ...  ...  ..   ...   ...  ..  ...
262      400              MICROFERMI            29                44  295  1450  20  76   70  109   4  245  244   0  1059  5949  15   33
263      400              MICROFERMI            28                43  294  1452  20  78   71  110   3  244  243   0  1054  5928  15   33
264      400              MICROFERMI            28                42  295  1442  20  77   69  112   4  244  244   0  1059  5927  14   34
265      400              MICROFERMI            27                41  295  1442  20  76   69  112   4  242  244   0  1050  5927  15   34
266      400              MICROFERMI            27                40  295  1446  20  76   69  110   3  243  242   0  1050  5949  16   32

[200 rows x 18 columns]
>>>
```

Figure 4.47 - Extraction of the dataset related to microstops only

Then we separate the data of the features from the data of the labels by extracting all the columns related to the features from the dataset “datiMF.” We generate an additional dataset “X1” from the first dataset “X” by removing the column VA, which, as previously stated, does not carry any useful information (Figure 4.48).

```
X=datiMF.loc[:,'C1':'TC']

>>> datiMF.loc[:,'C1':'TC']
         C1    T1  V1  C2   T2   V2  CC   PM   PF  VA    PS    FC  SM   TC
20      299  1505  20  68   74   92   3  246  242   0  1054  5917  20   33
21      299  1508  19  68   74   92   3  246  242   0  1050  5906  20   33
22      299  1520  19  67   75   89   4  247  243   0  1054  5927  20   34
23      292  1389  21   0  201    0   6    5    4   0    17  5758  92   62
24      290  1337  21   0  254    0   6    5    4   0    17  5801  94  217
...     ...   ...  ..  ..  ...  ...  ..  ...  ...  ..   ...   ...  ..  ...
137997  276  1279  21  97   80  121   3  272  271   0  1176  5654  21   36
137998  276  1275  21  96   75  128   4  271  270   0  1171  5675  20   38
137999  276  1278  21  96   75  128   4  271  271   0  1176  5707  20   38
138000  276  1273  21  94   75  125   4  271  271   0  1176  5675  22   36
138001  276  1290  21  96   75  122   4  271  265   0  1150  5633  21  244

[82005 rows x 14 columns]

X1 = X.drop('VA', axis = 1)

>>> X1
         C1    T1  V1  C2   T2   V2  CC   PM   PF    PS    FC  SM   TC
20      299  1505  20  68   74   92   3  246  242  1054  5917  20   33
21      299  1508  19  68   74   92   3  246  242  1050  5906  20   33
22      299  1520  19  67   75   89   4  247  243  1054  5927  20   34
23      292  1389  21   0  201    0   6    5    4    17  5758  92   62
24      290  1337  21   0  254    0   6    5    4    17  5801  94  217
...     ...   ...  ..  ..  ...  ...  ..  ...  ...   ...   ...  ..  ...
137997  276  1279  21  97   80  121   3  272  271  1176  5654  21   36
137998  276  1275  21  96   75  128   4  271  270  1171  5675  20   38
137999  276  1278  21  96   75  128   4  271  271  1176  5707  20   38
138000  276  1273  21  94   75  125   4  271  271  1176  5675  22   36
138001  276  1290  21  96   75  122   4  271  265  1150  5633  21  244

[82005 rows x 13 columns]
```

Figure 4.48 - Features dataset (microstops)

Similarly, the “Labels” dataset, dubbed “Y,” is generated and shown in Figure 4.49.

```
Y=datiMF.loc[:,'N_CICLI_MANCANTI']

>>> Y=datiMF.loc[:,'N_CICLI_MANCANTI']
>>> Y
20        86
21        87
22        85
23        84
24        83
          ..
137997     5
137998     4
137999     2
138000     3
138001     1
Name: N_CICLI_MANCANTI, Length: 82005, dtype: int64
```

Figure 4.49 - Labels dataset (microstops)

At this point, the algorithm is applied to the data frames “X1” and “Y” using the score function “f_regression” and the scores of the various features are obtained (Figure 4.50)

```
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import f_regression

bestfeatures = SelectKBest(score_func=f_regression, k='all')
X_new=bestfeatures.fit(X1,Y)
X_new.scores_
```

```
dfscores=pd.DataFrame(X_new.scores_)
dfcolumns= pd.DataFrame(X1.columns)

featurescores=pd.concat([dfcolumns,dfscores],axis=1)
featurescores
```

```
>>> featurescores
     0            0
0    C1   721.425289
1    T1   271.004318
2    V1   798.147926
3    C2   176.938309
4    T2   194.556050
5    V2    72.203649
6    CC   684.945964
7    PM   334.998467
8    PF   330.619657
9    PS   332.456319
10   FC     4.153927
11   SM  8213.979976
12   TC     4.622362
>>>
```

Figure 4.50 - Feature selection with "K-Best"

According to the algorithm, the most significant parameter is SM, followed by T1, V1, and CC. A similar result is obtained by measuring the incidence of features using the Random Forest regressor, as shown in the following Figure 4.51:

```
from sklearn import ensemble
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X1, Y, test_size=0.33, random_state=42)

rf = ensemble.RandomForestRegressor(n_estimators = 188, min_samples_split = 2, min_samples_leaf = 2, max_features = 'auto', max_depth = None, bootstrap = True)
>>> print(rf.get_params())
{'bootstrap': True, 'ccp_alpha': 0.0, 'criterion': 'mse', 'max_depth': None, 'max_features': 'auto', 'max_leaf_nodes': None, 'max_samples': None,
 'min_impurity_decrease': 0.0, 'min_impurity_split': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 188, 'n_jobs': None, 'oob_score': False, 'random_state': None, 'verbose': 0, 'warm_start': False}

rf.fit(X1, Y)
RandomForestRegressor(min_samples_leaf=2, n_estimators=188)

>>>

important_features = pd.Series(data=rf.feature_importances_,index=X1.columns)
important_features.sort_values(ascending=False,inplace=True)
print(important_features.head(10))

T1    0.215444
FC    0.199758
SM    0.144783
PM    0.114771
C2    0.067838
V2    0.059960
C1    0.042513
T2    0.041613
TC    0.037055
PF    0.035130
dtype: float64
>>>
```

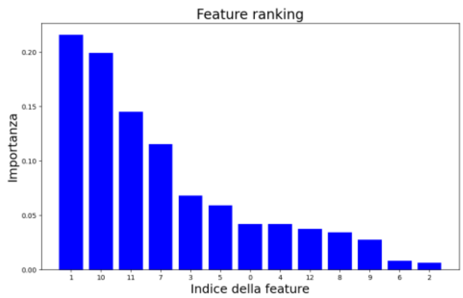


Figure 4.51 - Features score with Random Forest regressor

T1 and SM are among the most significant in this case. Whereas in the feature ranking graph, the numbers "1", "10", and "11" correspond precisely to the ordered indices of the features T1, FC, and SM.

4.7.1.3. Search for the best regressor

In this phase, we investigate various regression algorithms and evaluate them using performance metrics. The following function is defined for this purpose (Figure 4.52):

```
def find_best_model(X,Y):
    models = {
        'linear_regression' :{
            'model': LinearRegression(),
            'parameters':{

            }
        },

        'decison_tree_regressor':{
            'model': DecisionTreeRegressor(splitter='best'),
            'parameters' :{
                'max_depth' :[5,10]
            }
        },

        'random_forest': {
            'model': RandomForestRegressor(),

            'parameters' :{
                'n_estimators': [1,5,10,15,20,30,40,50,60,70,80,90,100]
            }
        },

         'svc' : {
            'model' : SVR(gamma= 'auto'),

            'parameters': {
                'kernel' : ['rbf','linear'],
                'C': [1,10,20]
            }
        }
    }
    scores = []
    cv_shuffle = ShuffleSplit(n_splits=5,test_size= 0.33,random_state=0)
    for model_names,model_params in models.items():
        gc = GridSearchCV(model_params['model'],model_params['parameters'],cv = cv_shuffle,return_train_score= False)
        gc.fit(X,Y)
        scores.append({
            'model': model_names,
            'parameters' : gc.best_params_,
            'score' : gc.best_score_
        })

    return pd.DataFrame(scores, columns=['model','best_parameters','score'])
```

Figure 4.52 - Evaluation function of different regressors

When running on the training data, the function returns a measure of each of the regressor performance:

- Linear regressor,
- Decision Tree regressor
- Random Forest regressor
- Support Vector Machine regressor (SVR)

Running the function on the training data yields the result shown in Figure 4.53, indicating that the regression based on the Random Forest algorithm yields the best results:

```
find_best_model(X_train, y_train)

                   model  best_parameters     score
0      linear_regression              NaN  0.152795
1  decison_tree_regressor             NaN  0.417818
2          random_forest              NaN  0.623486
3                    svc              NaN  0.136391
```

Figure 4.53 - Result of the function

4.7.1.4. Tuning of the Random Forest regressor

Since it is obvious that the algorithm that provides the best performance is Random Forest, it is recommended that the tuning of the hyperparameters at the heart of the algorithm's operation be performed in order to improve the results (the number of decision trees, their depth, the number of features to be considered, etc.). The best hyperparameters are generally impossible to predict ahead of time, and model optimization is the point at which machine learning transitions from science to trial-and-error engineering. The so-called "Cross Validation" method (K-Fold CV) is used for this purpose, which basically divides the training dataset into a certain number of pieces, called Fold, which will be used repeatedly to perform the training of the algorithm, sometimes as test data and sometimes as training data, and taking the metric corresponding to each iteration.

For example, if the data is divided into five pieces, the i-th piece will be used as a train dataset four times and a test dataset once for five consecutive executions. The K iteration metric is the average of the metrics from each iteration. This procedure prevents the metric from returning a value that is too dependent on how the dataset on which the algorithm is trained was packaged. The preceding procedure is repeated with different values of the algorithm's hyperparameters, evaluating the resulting metrics from time to time, to obtain at the end a set of hyperparameter values that correspond to the best values of the resulting metrics. The algorithm is then run on the entire training dataset with the hyperparameters that produced the best results, providing the metric of the cross-validation method. This technique was used to evaluate several algorithms, including the ricker-based Random Forest algorithm, for which a 5-fold cross-validation was performed by analysing different values of the hyperparameter representing the number of estimators.
The current phase goal is to fine-tune this and other algorithm parameters in order to improve the outcome. To this end, a grid of hyperparameter values is generated to be used to evaluate the algorithm's performance, as shown in Figure 4.54.

```
from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 100, stop = 200, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
print(random_grid)

>>> print(random_grid)
{'n_estimators': [100, 111, 122, 133, 144, 155, 166, 177, 188, 200],
 'max_features': ['auto', 'sqrt'],
 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
 'min_samples_split': [2, 5, 10],
 'min_samples_leaf': [1, 2, 4],
 'bootstrap': [True, False]}
>>>
```

Figure 4.54 - Parameters grid

The Random Forest regressor is then cross-validated on this grid by requesting that the parameters with the best performances be returned (Figure 4.55).

```
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = rf, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_state=42, n_jobs = 1)
# Fit the random search model
rf_random.fit(X_train, y_train)


CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True, total=  26.7s
[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True, total=  26.5s
[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True, total=  27.1s
[Parallel(n_jobs=1)]: Done 300 out of 300 | elapsed: 80.0min finished
RandomizedSearchCV(cv=3, estimator=RandomForestRegressor(), n_iter=100,
                   n_jobs=1,
                   param_distributions={'bootstrap': [True, False],
                                        'max_depth': [10, 20, 30, 40, 50, 60,
                                                      70, 80, 90, 100, 110,
                                                      None],
                                        'max_features': ['auto', 'sqrt'],
                                        'min_samples_leaf': [1, 2, 4],
                                        'min_samples_split': [2, 5, 10],
                                        'n_estimators': [100, 111, 122, 133,
                                                         144, 155, 166, 177,
                                                         188, 200]},
                   random_state=42, verbose=2)



rf_random.best_params_
>>> rf_random.best_params_
{'n_estimators': 188, 'min_samples_split': 2, 'min_samples_leaf': 2, 'max_features': 'auto', 'max_depth': None, 'bootstrap': True}
```

Figure 4.55 - Extraction of optimal parameters (tuning)

The execution of the Cross-Validation method, as shown in the previous figure, generates a set of optimal parameters with which we will perform the training-on-training data. As we can see in Figure 4.56 below, the regressor (which we instantiate by calling it "base_model") is created with the optimal parameters that we determined in the previous step, and then training data "Train_Features", "Train_Labels" is trained. Following that, the model must make predictions on the test data "X _test" and compare the predictions to the actual values of the corresponding labels "Y_test" to determine the result of the resulting metrics. Metrics reveal a mean-squared error of about 6814, a mean-absolute error of 54, and an R^2 value of 0.66.

```
train_features=X_train
train_labels = y_train
test_features = X_test
test_labels = y_test

base_model = ensemble.RandomForestRegressor(n_estimators = 188, min_samples_split = 2, min_samples_leaf = 2, max_features = 'auto', max_depth = None, bootstrap = True)
base_model.fit(train_features, train_labels)


from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
y_pred = base_model.predict(X_test)
print("Random Forest Mean Squared Error: ", mean_squared_error(y_test, y_pred))
print("Random Forest Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
print("Random Forest r-squared: ", r2_score(y_test, y_pred))


>>> from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
>>> y_pred = base_model.predict(X_test)
>>> print("Random Forest Mean Squared Error: ", mean_squared_error(y_test, y_pred))
Random Forest Mean Squared Error:  6814.62618232505
>>> print("Random Forest Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
Random Forest Mean Absolute Error:  54.14042178147523
>>> print("Random Forest r-squared: ", r2_score(y_test, y_pred))
Random Forest r-squared:  0.6601329887344464
```

Figure 4.56 - Execution metrics

It should be noted that, while the first two metrics are absolute values (and thus comparable with the respective metrics obtained by running with different hyperparameters), the last metric is a relative metric that, in general, constitutes a more significant measure because it compares the mean squared error obtained with that which would be expected. The measure of R^2 assumes a maximum value equal to 1 when the model's mean square error is 0, a value equal to "0" if the mean square error is equal to that of the hypothetical model, which always responds with the average value of the labels, and a value equal to 0 if the performance is worse than the hypothetical model (i.e., mean squared error is even greater).
In other words, for values between "0" and "1," the actual model outperforms the hypothetical model as the result approaches 1. In our case, the outcome was neither particularly good nor particularly bad.
Let's visualize the scattering diagram, which represents the predictions in relation to the actual values, to visually verify the meaning of such a result. All of the points would be arranged on a straight line inclined at 45° in the ideal case, where the predictions overlapped exactly with the real values.

As shown in Figure 4.57, we find a large scatter in our case, indicating a low precision of the predictions.

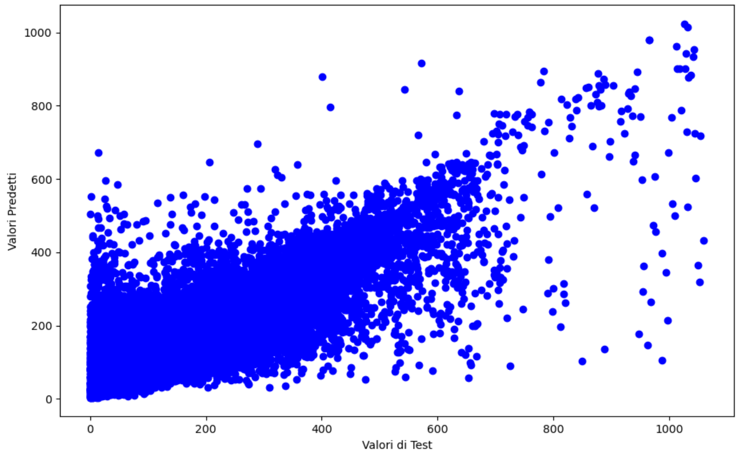


Figure 4.57 - Comparison between predict and actual values (test data)

Even close to 0 (which is the area of interest), the error is frequently very high (on the other hand the measure of absolute medium error tells us that the absolute average error is average is equal to 54 cycles or about 27 minutes).

4.7.1.5. Classification

Another way to make predictions about the approach of a micro-stop, while avoiding the possibility of predicting exactly when it will occur, is to hypothesize a time interval, such as 30 minutes (60 machine cycles), within which we will most likely enter the conditions that cause a micro-block. Assuming that the goal is to predict a micro-stop within a given time interval (i.e., within 60 machine cycles), the problem transforms from a regression problem to a classification problem because will be defined fault conditions as conditions that correspond to a micro-stop within a maximum of 60 cycles, rather than the conditions that precede the micro-stop for more than 30 minutes (characterized by a state 1). In these terms, the problem is re-posed as a binary classification problem, with the goal of predicting the 1 or predicting the machine's entry into a state where a micro-stop will be generated within the next 30 minutes (without specifying exactly when).
To accomplish this, first add a column to the dataset that represents the state of the system from the number of missing cycles to the next micro-stop:
"1" for a number of missing cycles ≤ 60
"0" for a number of missed cycles > 60
Figure 4.58 shows the syntax for running the task, as well as the resulting dataset.

```
cicli=60
datiMF['STATO'] = np.where(datiMF['N_CICLI_MANCANTI'] <= cicli, 1, 0 )


>>> datiMF
     CAUSALE DESCRIZIONE_SOSPENSIONE  MIN_MANCANTI  N_CICLI_MANCANTI   C1    T1  V1  C2   T2   V2  CC   PM   PF  VA    PS    FC  SM   TC  STATO

...      ...                     ...           ...               ...  ...   ...  ..  ..  ...  ...  ..  ...  ...  ..   ...   ...  ..  ...    ...

39       400              MICROFERMI            44                68  294  1490  19  75   69  109   4  245  241   0  1046  5885  17   34      0
40       400              MICROFERMI            43                66  294  1490  19  75   69  109   4  244  241   0  1046  5917  17   34      0
41       400              MICROFERMI            43                67  294  1496  19  73   68  107   4  244  242   0  1050  5875  19   32      0
42       400              MICROFERMI            42                64  294  1510  19  74   70  106   4  245  243   0  1054  5906  18   33      0
43       400              MICROFERMI            42                65  294  1494  19  74   69  107   4  244  243   0  1054  5864  18   32      0
44       400              MICROFERMI            41                63  294  1473  20  74   70  106   4  245  243   0  1054  5917  18   33      0
45       400              MICROFERMI            41                62  294  1484  19  74   70  106   4  245  243   0  1054  5897  18   32      0
46       400              MICROFERMI            40                60  294  1482  20  74   68  109   3  245  243   0  1054  5917  19   32      1
47       400              MICROFERMI            40                61  294  1478  20  74   70  106   4  245  243   0  1054  5939  18   34      0
48       400              MICROFERMI            39                59  294  1491  19  75   69  109   4  244  243   0  1054  5896  17   34      1
49       400              MICROFERMI            38                58  294  1489  19  75   70  107   4  244  243   0  1054  5928  17   33      1
50       400              MICROFERMI            38                57  294  1489  20  75   69  107   4  244  243   0  1050  5928  18   33      1
51       400              MICROFERMI            37                56  294  1490  19  74   68  109   3  245  243   0  1054  5927  19   33      1
52       400              MICROFERMI            37                55  294  1485  19  75   69  109   4  244  242   0  1050  5907  17   34      1
53       400              MICROFERMI            36                54  294  1485  20  74   69  109   4  244  242   0  1050  5907  18   34      1
54       400              MICROFERMI            36                53  294  1470  20  75   69  109   3  244  242   0  1050  5948  18   33      1
55       400              MICROFERMI            35                51  294  1477  20  73   68  107   4  244  243   0  1054  5907  19   32      1
56       400              MICROFERMI            35                52  294  1468  20  75   69  109   4  245  243   0  1054  5928  17   34      1
57       400              MICROFERMI            34                49  294  1473  20  74   68  109   4  245  243   0  1054  5948  18   33      1
58       400              MICROFERMI            34                50  294  1466  20  75   70  107   4  245  243   0  1054  5970  17   33      1
59       400              MICROFERMI            33                48  294  1473  20  75   70  109   3  245  243   0  1054  5927  18   33      1
60       400              MICROFERMI            32                46  294  1466  20  75   70  107   4  244  243   0  1054  5948  17   32      1
61       400              MICROFERMI            32                47  294  1460  20  75   69  109   4  244  242   0  1050  5896  17   34      1
...          ...                     ...          ...                ...  ...   ...  ..  ..  ...  ...  ..  ...  ...  ..   ...   ...  ..  ...      .
```

Figure 4.58 - Adding the "STATO" column

At this point, the dataset of features "X1" returns, as in the previous case, and it is assumed that the array of label "Y" coincides with the colonna "STATO," as shown in Figure 4.59.

```
X=datiMF.loc[:,'C1':'TC']

X1 = X.drop('VA', axis = 1)

>>> X1
         C1    T1  V1  C2   T2   V2  CC   PM   PF    PS    FC  SM   TC
20      299  1505  20  68   74   92   3  246  242  1054  5917  20   33
21      299  1508  19  68   74   92   3  246  242  1050  5906  20   33
22      299  1520  19  67   75   89   4  247  243  1054  5927  20   34
23      292  1389  21   0  201    0   6    5    4    17  5758  92   62
24      290  1337  21   0  254    0   6    5    4    17  5801  94  217
...     ...   ...  ..  ..  ...  ...  ..  ...  ...   ...   ...  ..  ...
137997  276  1279  21  97   80  121   3  272  271  1176  5654  21   36
137998  276  1275  21  96   75  128   4  271  270  1171  5675  20   38
137999  276  1278  21  96   75  128   4  271  271  1176  5707  20   38
138000  276  1273  21  94   75  125   4  271  271  1176  5675  22   36
138001  276  1290  21  96   75  122   4  271  265  1150  5633  21  244

[82005 rows x 13 columns]


Y=datiMF.loc[:,'STATO']
```

Figure 4.59 - Features and Labels dataset

On these data, the K-Fold Cross-Validation technique is used to determine the best parameters for the Random Forest classifier. As a result, the test and training datasets are created, and the parameter grid on which to run the tuning at the K-Fold Cross-Validation is defined (Figure 4.60):

```
X_train, X_test, y_train, y_test = train_test_split(X1, Y, random_state=1, stratify=Y)

from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 100, stop = 200, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
print(random_grid)


>>> print(random_grid)
{'n_estimators': [100, 111, 122, 133, 144, 155, 166, 177, 188, 200],
 'max_features': ['auto', 'sqrt'],
 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
 'min_samples_split': [2, 5, 10],
 'min_samples_leaf': [1, 2, 4],
 'bootstrap': [True, False]}
>>>
```

Figure 4.60 - Classification tuning

After a significant amount of processing time, the simulation is run for the parameter grid using 3-Fold and the following optimal parameters are obtained and shown in Figure 4.61.

```
# Use the random grid to search for best hyperparameters
# First create the base model to tune
rf = RandomForestClassifier()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = rf, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_state=42, n_jobs = 1)
# Fit the random search model
rf_random.fit(X_train, y_train)


[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True, total=  10.1s
[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=2, max_features=auto, max_depth=40, bootstrap=True, total=  10.0s
[Parallel(n_jobs=1)]: Done 300 out of 300 | elapsed: 43.1min finished
RandomizedSearchCV(cv=3, estimator=RandomForestClassifier(), n_iter=100,
                   n_jobs=1,
                   param_distributions={'bootstrap': [True, False],
                                        'max_depth': [10, 20, 30, 40, 50, 60,
                                                      70, 80, 90, 100, 110,
                                                      None],
                                        'max_features': ['auto', 'sqrt'],
                                        'min_samples_leaf': [1, 2, 4],
                                        'min_samples_split': [2, 5, 10],
                                        'n_estimators': [100, 111, 122, 133,
                                                         144, 155, 166, 177,
                                                         188, 200]},
                   random_state=42, verbose=2)


>>> rf_random.best_params_
{'n_estimators': 122, 'min_samples_split': 10, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': 40, 'bootstrap': False}
```

Figure 4.61 - Optimal classification hyperparameters

At this point, after running the Random Forest classifier with these parameters on test data to obtain predictions, you can evaluate the classifier's performance using various metrics. First and foremost, we notice that the test dataset contains a total of 20502 values, with 5522 true positive (i.e., 1) and 14980 true negative values (i.e., 0). The confusion matrix, depicted in figure 46, is the most straightforward metric to represent. The matrix reports the classes subject to the predictions (Figure 4.62):

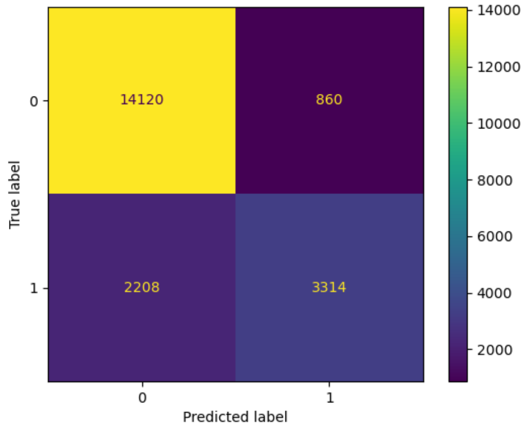


Figure 4.62 - Confusion Matrix

• For class 0 on the first line, the true positives (at the top left) account for 14120 of the 14980 actual positives, while the false positives account for 860 (at the top right). As a result, the algorithm correctly predicted 14120/14980 = 94% of class 0 positives (recall of class 1)
• On the second line, for class 1, the true negatives (lower right) account for 3314 of the actual positives (5522), while the false positives account for 2208 (lower left). As a result, the algorithm correctly predicted 3314/5522 = 60% of class 1 positives (recall of class 1 and measures the sensitivity of the class).
Looking at the matrix by columns, we can see the precision of the predictions for each class:
• The precision of class 0 (first column) is given by the ratio of the 0 correctly predicted (true positives) in a number equal to 14120 to the 0 predicted total including the erroneous (false positives) in a number equal to 2208. The accuracy of class 0 is represented by this ratio (14120 / (14120 + 2208) = 86%).
• The precision of class 1 (second column) is given by the ratio of the 1 correctly predicted (true positives) in a number equal to 3314 and the 1 correctly predicted total including the erroneous (false positives) in a number equal to 860. The accuracy of class 1 is represented by this ratio (3314 / (3314 + 860) = 79%).
Finally, a third metric defines model accuracy as the ratio of true positives across all classes (14120 + 3314) to total predictions made (14120 + 3314 + 2208 + 860), which corresponds to an 85% value in our case.

The value of these metrics is represented in the following Classification Report, which is shown in Figure Figure 4.63:

```
y_test.value_counts()

>>> y_test.value_counts()
0    14980
1     5522
Name: STATO, dtype: int64


# View accuracy score
accuracy_score(y_test, y_pred_test)
0.8503560628231392


# View confusion matrix for test data and predictions
confusion_matrix(y_test, y_pred_test)

array([[14120,   860],
       [ 2208,  3314]])

from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)  # doctest: +SKIP
plt.show()  # doctest: +SKIP

from sklearn.metrics import classification_report
#target_names = ['Outside60Cyc', 'Inside60Cyc']
target_names = ['>60Cyc', '<60Cyc']
print(classification_report(y_test, y_pred_test,target_names=target_names))

>>> target_names = ['>60Cyc', '<60Cyc']
>>> print(classification_report(y_test, y_pred_test,target_names=target_names))
              precision    recall  f1-score   support

      >60Cyc       0.86      0.94      0.90     14980
      <60Cyc       0.79      0.60      0.68      5522

    accuracy                           0.85     20502
   macro avg       0.83      0.77      0.79     20502
weighted avg       0.85      0.85      0.84     20502
```

Figure 4.63 - Classification Report

Finally, the F1 metric value represents the harmonic average of each class's precision and recall (so it is a synthetic metric per class).
Another metric is the value of AUC (which stands for "Area Under the Curve"), where the curve in question is known as ROC (Receiver Operating Characteristic).
In our case, the representation of the AUC metric yields a value of 0.771, and the best condition is obtained precisely at the value of Recall equal to 0.6, which is the sensitivity value for class 1 as shown in Figure 4.64.

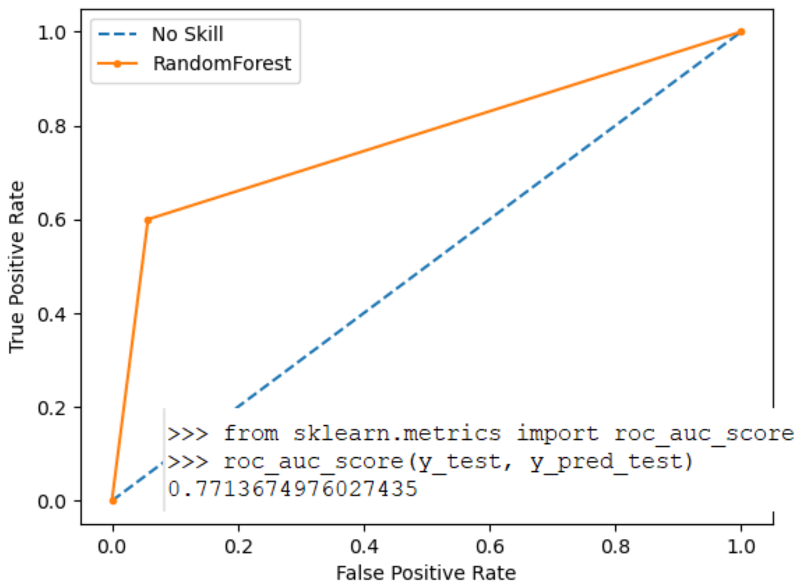


Figure 4.64 - AUC and ROC metrics

It should be noted that, by default, the representation shown in the previous figure considers actual positive labels to be equal to 1 and, thus, the representation in question refers to class 1. Instead, the value 0.771 represents the area below the curve (i.e., the value of the AUC metric).

4.7.2. Mechanical Fault

The second causalisation among machine faults is mechanical fault (causal 114), which is the subject of this paragraph.

4.7.2.1. Correlation Matrix

The methodology used to address the problem is the same as described in the case of microstops, and it begins with an analysis of the correlation matrix, in Figure 4.65, to highlight any apparent relationships between the features and the label identified, which is represented by the number of missing cycles at the next stop, as in the previous case.

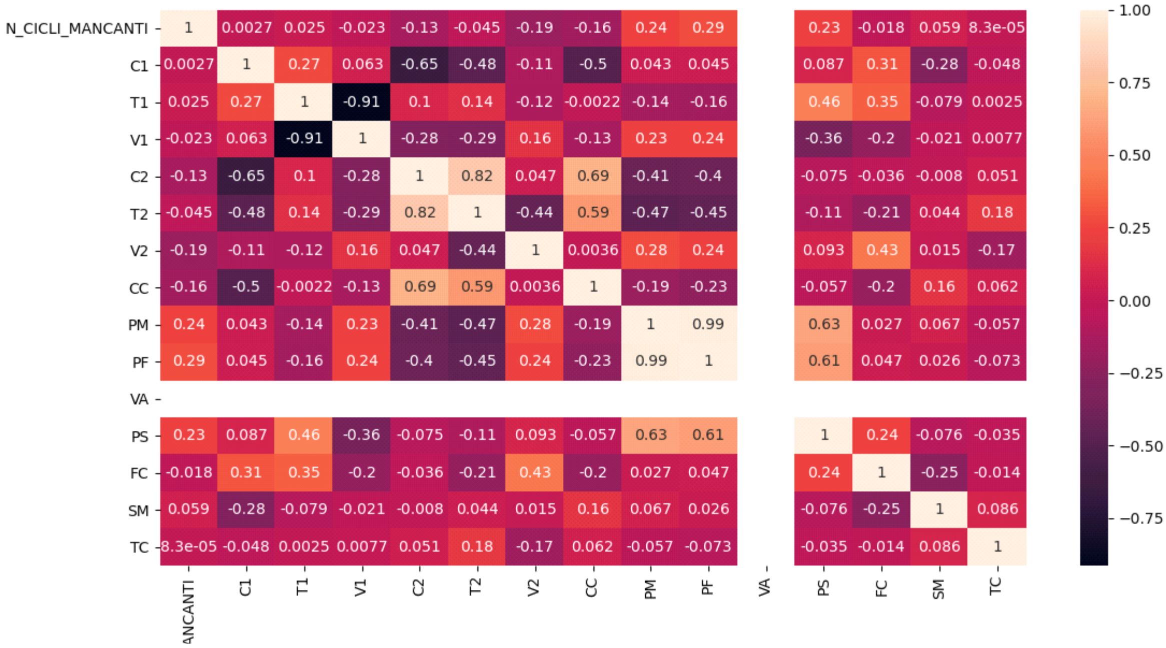


Figure 4.65 - Correlation Matrix (Mechanical fault)

The elements that emerge are represented by pressure-type quantities (PF, PS, and PM) bonds with the label and weaker bonds with the quantities V2, CC. We will represent the trends of the most correlated quantities towards the values of the corresponding labels in the same way that we did in the previous case.

Just from these representations, it is possible to notice significantly different behaviours of the quantities in question within 80-100 cycles of the machine stop, which can be considered indicators of a change in the machine's state. It should be noted that, in contrast to micro-stops, the variations in the correlated quantities are clearly visible in this case. Let's take a look at the trend of some of the other features that will be important in the analysis progression (T1 and VA) in Figure 4.66and Figure 4.67.

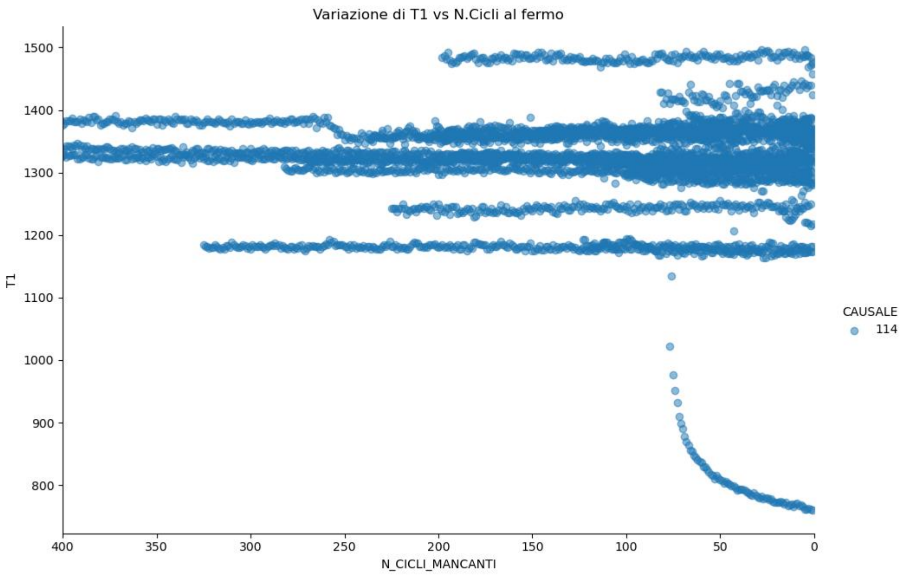


Figure 4.66 - Variation of T1 vs missing cycles number

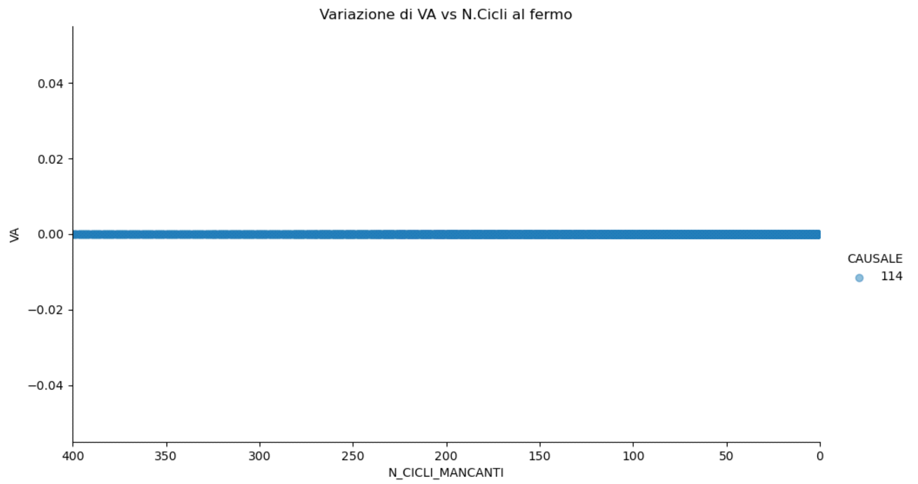


Figure 4.67 - Variation of VA vs missing cycles number

In particular, as in the previous case, VA is observed to be costing more than the label and may thus be neglected from the analysis dataset.

4.7.2.2. Feature selection

The methodology used to find the most important features is similar to the K-Best method, which was previously used for microstops. Other methodologies will be used to discover these results. Figure 4.68 depicts the application of the K-Best algorithm for mechanical faults.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#dati = pd.read_csv("/dati/DATI_COMPLETI_2019T1_v2_RIDOTTO.csv", skiprows =0, sep=';')
dati = pd.read_csv("/dati/DATI_COMPLETI_TOTALI_v2_RIDOTTO.csv", skiprows =0, sep=';')

filtro_GM = dati['CAUSALE']==114

datiGM=dati[filtro_GM]

X=datiGM.loc[:,'C1':'TC']

X1 = X.drop('VA', axis = 1)

>>> X1
        C1    T1  V1   C2   T2   V2  CC   PM   PF    PS    FC  SM  TC
2374   272  1242  21  123  101  122   6  242  239  1037  5979  24  37
2375   272  1243  21  123  102  121   7  241  241  1046  6001  23  38
...    ...   ...  ..  ...  ...  ...  ..  ...  ...   ...   ...  ..  ..
377141 280  1364  20   86   69  125   5  281  269  1167  5654  25  34
377142 280  1357  20   86   69  125   5  281  270  1171  5674  25  34

[4934 rows x 13 columns]


Y=datiGM.loc[:,'N_CICLI_MANCANTI']

from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import f_regression

bestfeatures = SelectKBest(score_func=f_regression, k='all')
X_new=bestfeatures.fit(X1,Y)
```

Figure 4.68 - Application of the K-Best algorithm (mechanical fault)

Using the K-Best algorithm on the X1 dataset of 13 features and the label array Y obtained in the previous figure, and the score function named “f_regression,” the results are shown in Figure 4.69.

```
dfscores=pd.DataFrame(X_new.scores_)
dfcolumns= pd.DataFrame(X1.columns)

featurescores=pd.concat([dfcolumns,dfscores],axis=1)
featurescores

     0            0
0   C1     7.145287
1   T1   137.729261
2   V1    75.234580
3   C2   579.806317
4   T2   591.187232
5   V2    21.731263
6   CC    99.232155
7   PM  1242.349106
8   PF  1049.742899
9   PS  1046.726575
10  FC     0.119966
11  SM   838.466813
12  TC    21.965018
```

Figure 4.69 - Results of K-Best algorithm

The algorithm thus demonstrates that the most significant features are the same as those identified by the correlation matrix or the three pressures PF, PS, and PM, followed by SM and, at a greater distance, C2 and T2.

4.7.2.3. ElasticNet regression algorithm

To determine the optimal set of algorithm hyperparameters, different regression algorithms are evaluated using, as in the previous case, the Cross-Validation technique. As a result, we prepare the grid of algorithm parameters and run the fit with the “GridSearChCV” method on a 5-fold scale. The algorithm is shown in Figure 4.70.

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
import matplotlib.pyplot as plt

X_train, X_test, y_train, y_test = train_test_split(X1, Y, test_size=0.33, random_state=1)

print('\n \n There are {} samples in the training set and {} samples in the test set'.format(X_train.shape[0], X_test.shape[0]))

 There are 3305 samples in the training set and 1629 samples in the test set


from sklearn.linear_model import ElasticNet
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import GridSearchCV, train_test_split

# Create train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=42)


my_alpha = np.linspace(.01, 1, num=5)
my_l1_ratio = np.linspace(.01, 1, num=20)
param_grid = {'l1_ratio': my_l1_ratio, 'alpha': my_alpha}

# Create the hyperparameter grid
#l1_space = np.linspace(0, 1, 30)
#param_grid = {'l1_ratio': l1_space}

# Instantiate the ElasticNet regressor: elastic_net
elastic_net = ElasticNet(max_iter=100000, tol=0.001)

# Setup the GridSearchCV object: gm_cv
gm_cv = GridSearchCV(elastic_net, param_grid, cv=5 , verbose = 2, n_jobs = 1)

# Fit it to the training data
gm_cv.fit(X_train, y_train)
```

Figure 4.70 - Optimal parameter search with ElasticNet

The results shown in Figure 4.71 are obtained by running the algorithm on test data after it has been optimized and trained on training data:

```
>>> # Predict on the test set and compute metrics
>> y_pred = gm_cv.predict(X_test)
>>> r2 = gm_cv.score(X_test, y_test)
>>> mse = mean_squared_error(y_pred,y_test)
>>> print("Tuned ElasticNet ratio/alpha: {}".format(gm_cv.best_params_))
Tuned ElasticNet ratio/alpha: {'alpha': 0.505, 'l1_ratio': 0.21842105263157896}
>>> print("Linear regression Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
Linear regression Mean Absolute Error:  82.4681220898177
>>> print("Tuned ElasticNet R squared: {}".format(r2))
Tuned ElasticNet R squared: 0.3644558478619421
>>> print("Tuned ElasticNet MSE: {}".format(mse))
Tuned ElasticNet MSE: 11734.831130397848
>>>
```

Figure 4.71 - ElasticNet algorithm results

These results show that the algorithm performance is quite poor, with a mean absolute error in the predictions of the missing time at the 84 cycles stop (about 42 minutes) and a R^2 index of very low value (0.36 about). In fact, when the dispersion diagram of the aforementioned values towards test values is represented, it is noted that the representation is very far from a linear trend that should have the optimal representation (Figure 4.72).

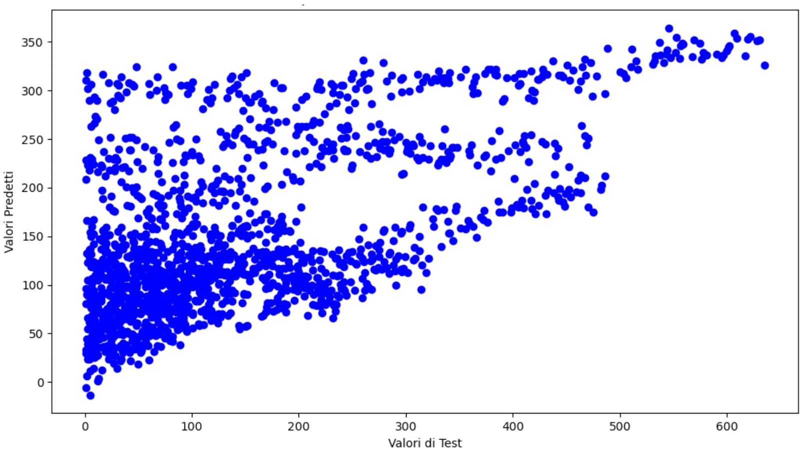


Figure 4.72 - Predictions vs label with ElasticNet (test data)

4.7.2.4. Linear regression

We now assess the performance of the linear regression algorithm, for which we do not use the cross-validation technique (at least not formally), because the algorithm lacks significant parameters for tuning (Figure 4.73).

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
import matplotlib.pyplot as plt

X_train, X_test, y_train, y_test = train_test_split(X1, Y, test_size=0.33, random_state=1)

print('\n \n There are {} samples in the training set and {} samples in the test set'.format(X_train.shape[0], X_test.shape[0]))

 There are 3305 samples in the training set and 1629 samples in the test set

lr = LinearRegression()
param_grid = {}

# Setup the GridSearchCV object: gm_cv
lr_cv = GridSearchCV(lr, param_grid, cv=5 , verbose = 2, n_jobs = 1)

# Fit it to the training data
lr_cv.fit(X_train, y_train)
```

Figure 4.73 - Execution of linear regression algorithm

In this case, based on the metrics shown in Figure 4.74, it should be noted that the model has expired performance in terms of the accuracy with which the values on the test data are predicted.

```
>>>
>>> # Predict on the test set and compute metrics
>>> y_pred = lr_cv.predict(X_test)
>>> r2 = lr_cv.score(X_test, y_test)
>>> mse = mean_squared_error(y_pred,y_test)
>>> print("Linear regression Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
Linear regression Mean Absolute Error:  82.61462637072
>>> print("Linear regression R squared: {}".format(r2))
Linear regression R squared: 0.36379130951155647
>>> print("Linear regression MSE: {}".format(mse))
Linear regression MSE: 11747.101316969804
```

Figure 4.74 - Linear regression results

The dispersion diagram (Figure 4.78) demonstrates the poor performance of the model.

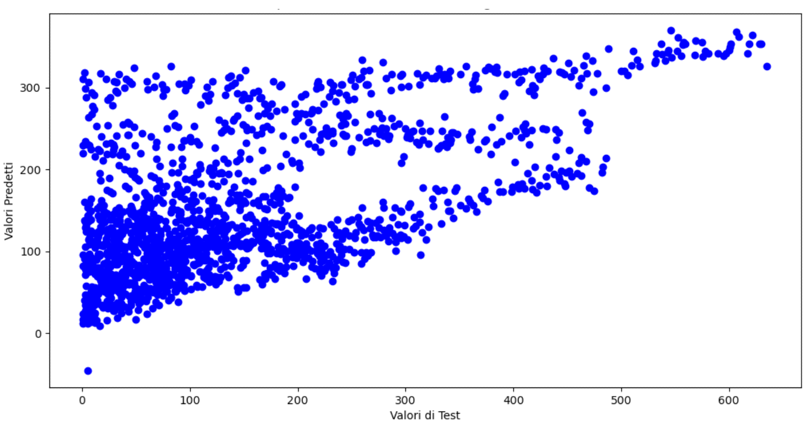


Figure 4.75 - Predictions vs Label with linear regression (test data)

4.7.2.5. SVC regression algorithm

We now assess the performance of the Support Vector Machine-based regression algorithm, using the cross-validation method to tune the parameters as usual (Figure 4.76).

```
from sklearn.svm import SVR
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
import matplotlib.pyplot as plt

X_train, X_test, y_train, y_test = train_test_split(X1, Y, test_size=0.33, random_state=1)

print('\n \n There are {} samples in the training set and {} samples in the test set'.format(X_train.shape[0], X_test.shape[0]))

 There are 3305 samples in the training set and 1629 samples in the test set

svr_best = SVR()
param_grid = {
            'kernel' : ['rbf','linear'],
            'C': [1,10,20],
            'gamma' : ('auto','scale'),
            'coef0' : [0.01,10,0.5]
            }

# Setup the GridSearchCV object: gm_cv
svr_best_cv = GridSearchCV(svr_best, param_grid, cv=5 , verbose = 2, n_jobs = 1)

# Fit it to the training data
svr_best_cv.fit(X_train, y_train)
```

Figure 4.76 - Execution of the SVC regression algorithm

The result of the algorithm’s execution, as shown in Figure 4.77, does not differ significantly from the previous reporting, and even in this case, the results are not remarkable.

```
>>>
>>> # Predict on the test set and compute metrics
>>> y_pred = svr_best_cv.predict(X_test)
>>> r2 = svr_best_cv.score(X_test, y_test)
>>> mse = mean_squared_error(y_pred,y_test)
>>> print("Tuned SVC parameters: {}".format(svr_best_cv.best_params_))
Tuned SVC parameters: {'C': 1, 'coef0': 0.01, 'gamma': 'auto', 'kernel': 'linear'}
>>> print("SVC Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
SVC Mean Absolute Error:  82.05832155917658
>>> print("SVC R squared: {}".format(r2))
SVC R squared: 0.3502103445235313
>>> print("SVC MSE: {}".format(mse))
SVC MSE: 11997.86332962648
>>>
```

Figure 4.77 - Results of the execution of the SVC algorithm

The dispersion diagram (Figure 4.78) demonstrates the poor performance of the model.

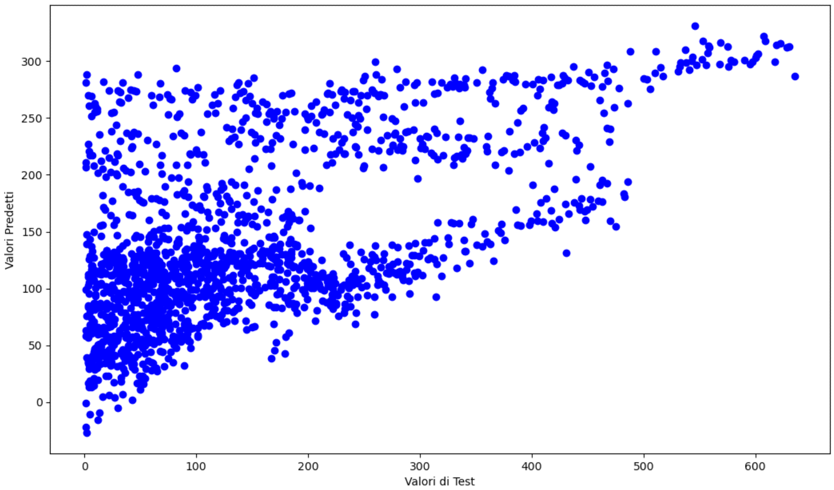


Figure 4.78 - Predictions vs Label with SVC regressor (test data)

4.7.2.6. Decision Tree regression algorithm

Another regression algorithm under consideration is one based on decision trees, on which we perform the standard tuning procedure using the Cross-Validation method (Figure 4.79).

```
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
import matplotlib.pyplot as plt

X_train, X_test, y_train, y_test = train_test_split(X1, Y, test_size=0.33, random_state=1)

print('\n \n There are {} samples in the training set and {} samples in the test set'.format(X_train.shape[0], X_test.shape[0]))

 There are 3305 samples in the training set and 1629 samples in the test set

num_leafs = [1, 5, 10, 20, 50, 100]
depths = np.arange(1, 21)

dt_best = DecisionTreeRegressor(splitter='best')
param_grid = {
             'max_depth' : depths,
             'min_samples_split': [2, 4, 6],
             'min_samples_leaf' : num_leafs
             }

# Setup the GridSearchCV object: dt_best_cv
dt_best_cv = GridSearchCV(dt_best, param_grid, cv=5 , verbose = 2, n_jobs = 1)

# Fit it to the training data
dt_best_cv.fit(X_train, y_train)
```

Figure 4.79 - Execution of decision tree regression algorithm

In this case, the algorithm execution results are noticeably better (Figure 4.80).

```
>>> # Predict on the test set and compute metrics
>>> y_pred = dt_best_cv.predict(X_test)
>>> r2 = dt_best_cv.score(X_test, y_test)
>>> mse = mean_squared_error(y_pred,y_test)
>>> print("Tuned DecisionTree parameters: {}".format(dt_best_cv.best_params_))
Tuned DecisionTree parameters: {'max_depth': 14, 'min_samples_leaf': 10, 'min_samples_split': 4}
>>> print("DecisionTree Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
DecisionTree Mean Absolute Error:  41.28495169167547
>>> print("DecisionTree R squared: {}".format(r2))
DecisionTree R squared: 0.7921264490779795
>>> print("DecisionTree MSE: {}".format(mse))
DecisionTree MSE: 3838.2243127243373
```

Figure 4.80 - Results of the execution of the algorithm Decision Tree

The model best performance is also highlighted by the dispersion diagram. As shown in Figure 4.81 below, the behaviour is much more akin to a linear trend.

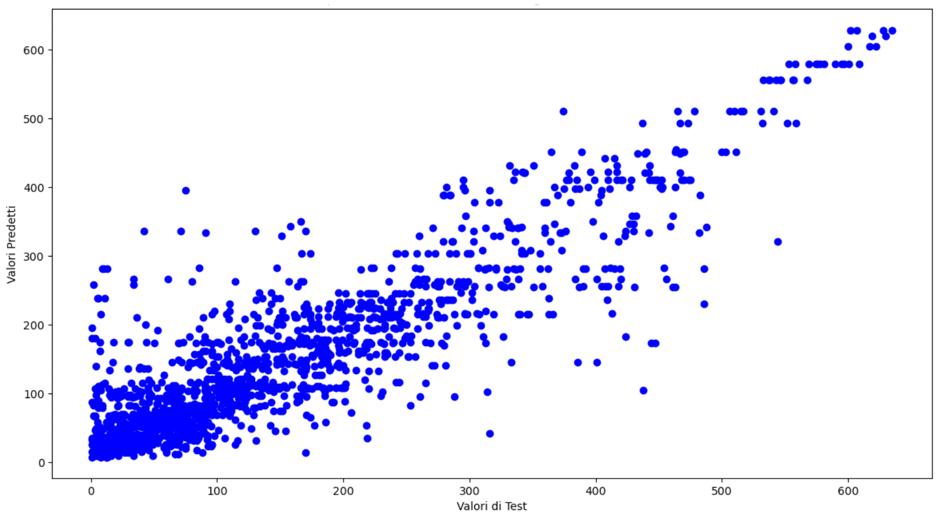


Figure 4.81 - Predictions vs Label with DecisionTree

The score attributed to the features should also be examined for this algorithm, which has produced acceptable results (Figure 4.82).

```
>>> important_features = pd.Series(data=base_model.feature_importances_,index=X1.columns)
>>> important_features.sort_values(ascending=False,inplace=True)
>>> print(important_features.head(15))
T1     0.371316
PS     0.343042
FC     0.111727
SM     0.065620
C2     0.035885
PM     0.031032
C1     0.009886
V1     0.007893
V2     0.007138
PF     0.005244
T2     0.004446
TC     0.004417
CC     0.002353
dtype: float64
```

Figure 4.82 - Score attributed to the features (DecisionTree)

It is easy to see that one of the pressures (PS) is located between the first features in order of importance, whereas the others are not considered particularly significant by the model in question.

4.7.2.7. Random Forest regressor

The Random Forest regression algorithm has highlighted the best performance for micro-stops and thus, presumably, among those evaluated, is what he has shown to better interpret the specific types of data object of the analysis. Figure 4.83 depicts the algorithm’s application to the mechanical faults causalisation.

```
from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 100, stop = 200, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
print(random_grid)


>>> print(random_grid)
{'n_estimators': [100, 111, 122, 133, 144, 155, 166, 177, 188, 200], 'max_features': ['auto', 'sqrt'], 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80,
>>>


# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = rf, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_state=42, n_jobs = 1)
# Fit the random search model
rf_random.fit(X_train, y_train)
```

Figure 4.83 - Execution of Random Forest (mechanical faults)

Also in this case, the regressor produces acceptable results that are the best among those related to the analysed models. So, even in this case, the best fitting can be considered (Figure 4.84).

```
rf_random.best_params_
>>> rf_random.best_params_
{'n_estimators': 122, 'min_samples_split': 5, 'min_samples_leaf': 1, 'max_features': 'auto', 'max_depth': 80, 'bootstrap': True}


# Predict on the test set and compute metrics
y_pred = rf_random.predict(X_test)
r2 = rf_random.score(X_test, y_test)
mse = mean_squared_error(y_pred,y_test)
print("Tuned RandomForest parameters: {}".format(rf_random.best_params_))
print("RandomForest Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
print("RandomForest R squared: {}".format(r2))
print("RandomForest MSE: {}".format(mse))

>>> # Predict on the test set and compute metrics
>>> y_pred = rf_random.predict(X_test)
>>> r2 = rf_random.score(X_test, y_test)
>>> mse = mean_squared_error(y_pred,y_test)
>>> print("Tuned RandomForest parameters: {}".format(rf_random.best_params_))
Tuned RandomForest parameters: {'n_estimators': 166, 'min_samples_split': 5, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': 80, 'bootstrap': False}
>>> print("RandomForest Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
RandomForest Mean Absolute Error:  36.06159968295527
>>> print("RandomForest R squared: {}".format(r2))
RandomForest R squared: 0.8401796610002389
>>> print("RandomForest MSE: {}".format(mse))
RandomForest MSE: 2950.95892717416
```

Figure 4.84 - Results of the execution of the algorithm Random Forest

The dispersion diagram between predicted and test values in Figure 4.85 visually demonstrates that the result is the best of the previous ones.

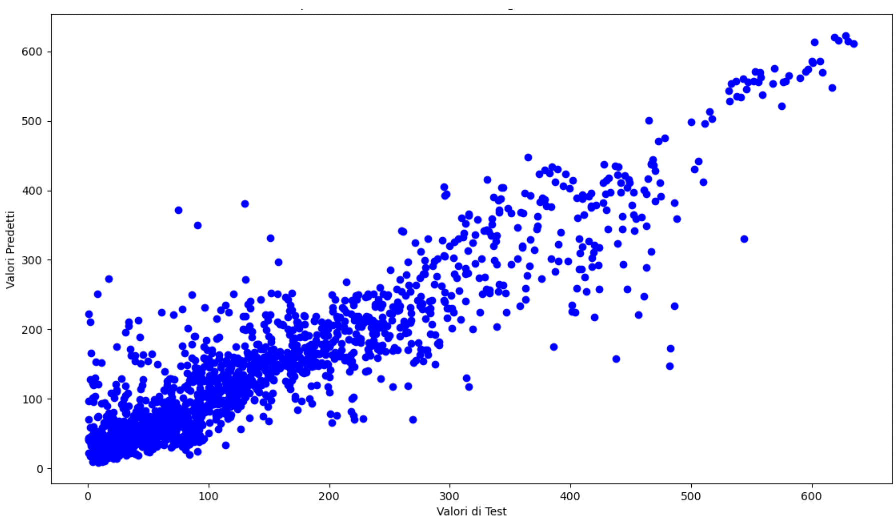


Figure 4.85 - Predictions vs Label with Random Forest

In particular, it is worthwhile to examine the trend of errors (with sign) between predicted and test values (see Figure 4.86).

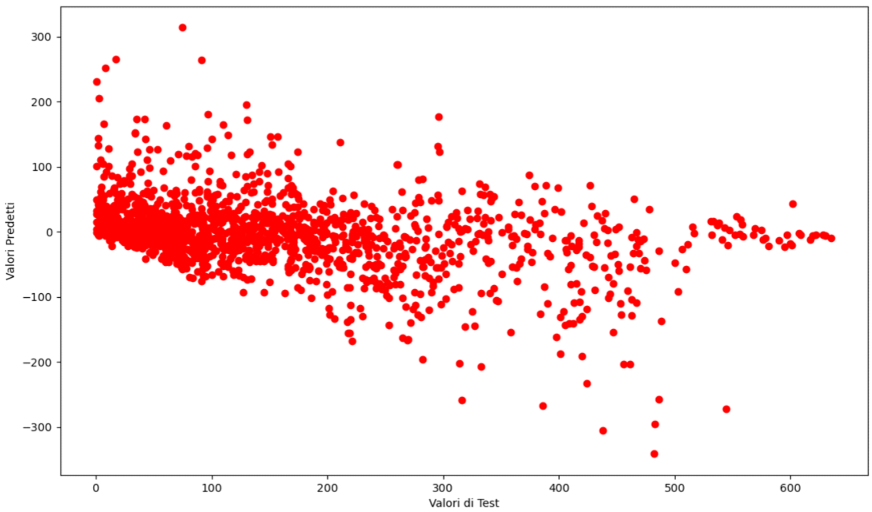


Figure 4.86 - Trend of errors between predicted and test values for mechanical faults

Specifically, the trend of errors in the immediate proximity of 0 is of particular interest: as shown in Figure 4.87, in the 60 cycles (30 minutes) preceding the machine stop, the prediction error on the test data is roughly understood (except for exceptions) to be between -20 and + 25 cycles, or an error of 10-15 minutes about the time committed to approaching the subsequent mechanical fault.

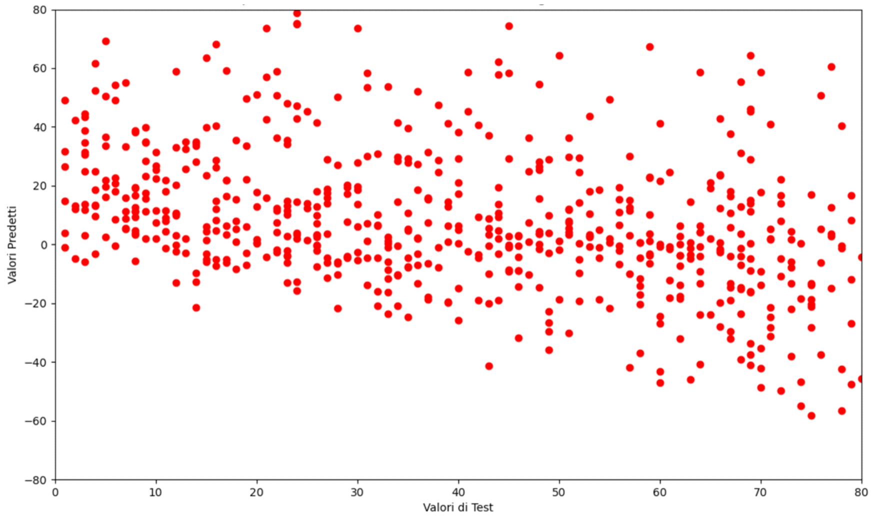


Figure 4.87 - Detail of trends of errors close to the 0

Furthermore, the model's overall score is 0.84 based on the score of the features attributed by the model (R^2). There is the option of determining how much this score is affected by each of the features. Because the Random Forest model reported a significant score, calculating the Permutation Feature Importance for this model makes sense. The ranking is shown in Figure 4.88.

```
>>> rf_best.score(X_test, y_test)
0.8454954886809413


params=rf_best.get_params()

from sklearn.inspection import permutation_importance

r = permutation_importance(rf_best, X_test, y_test,
                           n_repeats=30,
                           random_state=0)

...
T1      0.811 +/- 0.042
FC      0.220 +/- 0.011
PF      0.173 +/- 0.011
SM      0.159 +/- 0.008
PS      0.137 +/- 0.009
C2      0.049 +/- 0.005
PM      0.044 +/- 0.003
V1      0.030 +/- 0.004
V2      0.023 +/- 0.002
TC      0.011 +/- 0.001
C1      0.010 +/- 0.001
T2      0.010 +/- 0.002
CC      0.001 +/- 0.001
```

Figure 4.88 - Permutation Feature Importance for Random Forest

This rating reveals a high sensitivity of the model obtained from the T1 feature, and thus, to a lesser extent, from FC and PF. It is worth mentioning that FC and T1 were among the features with a higher score even when the decision tree regression algorithm was used.

4.7.2.8. Gradient Boosting regression

The regressor named gradient boosting is the last regression algorithm that has been evaluated. In this case, a parameter grid is also created, and the Cross-Validation method is used to determine the optimal algorithm parameters.

Figure 4.89 depicts the algorithm's application to the mechanical faults causalisation.

```
reg=GradientBoostingRegressor()

>>> reg.get_params()
{'alpha': 0.9,
 'ccp_alpha': 0.0,
 'criterion': 'friedman_mse',
 'init': None,
 'learning_rate': 0.1,
 'loss': 'ls',
 'max_depth': 3,
 'max_features': None,
 'max_leaf_nodes': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 100,
 'n_iter_no_change': None,
 'presort': 'deprecated',
 'random_state': None,
 'subsample': 1.0,
 'tol': 0.0001,
 'validation_fraction': 0.1,
 'verbose': 0,
 'warm_start': False}


param_grid = {
    'max_depth': [80, 90, 100, 110],
    'min_samples_leaf': [3, 4, 5],
    'min_samples_split': [8, 10, 12],
    'n_estimators': [100, 200, 300, 1000]
}

reg_best = RandomizedSearchCV(estimator = reg, param_distributions = param_grid, cv = 3, verbose=2, random_state=42, n_jobs = 1)
reg_best.fit(X_train, y_train)

reg_best.best_params_

>>> reg_best.best_params_
{'n_estimators': 100, 'min_samples_split': 12, 'min_samples_leaf': 4, 'max_depth': 90}
```

Figure 4.89 - Execution of Gradient Boosting (mechanical faults)

Performing with the optimal parameters so determined in Figure 4.90

```
>>> # Predict on the test set and compute metrics
>>> y_pred = reg_best.predict(X_test)
>>> r2 = reg_best.score(X_test, y_test)
>>> mse = mean_squared_error(y_pred,y_test)
>>> print("Tuned GradientBoostingRegressor parameters: {}".format(reg_best.best_params_))
Tuned GradientBoostingRegressor parameters: {'n_estimators': 100, 'min_samples_split': 12, 'min_samples_leaf': 4, 'max_depth': 90}
>>> print("GradientBoostingRegressor Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
GradientBoostingRegressor Mean Absolute Error:  35.93455977570218
>>> print("GradientBoostingRegressor R squared: {}".format(r2))
GradientBoostingRegressor R squared: 0.8350122705533382
>>> print("GradientBoostingRegressor MSE: {}".format(mse))
GradientBoostingRegressor MSE: 3046.3707944302973
>>>
```

Figure 4.90 - Results of the execution of the algorithm Gradient Boosting

As a result, the algorithm generates a model with performance comparable to that obtained with Ristress Random Forest. The dispersion diagram in Figure 4.91, which compares predicted and expected data, exhibits similar behaviour to the previous one.

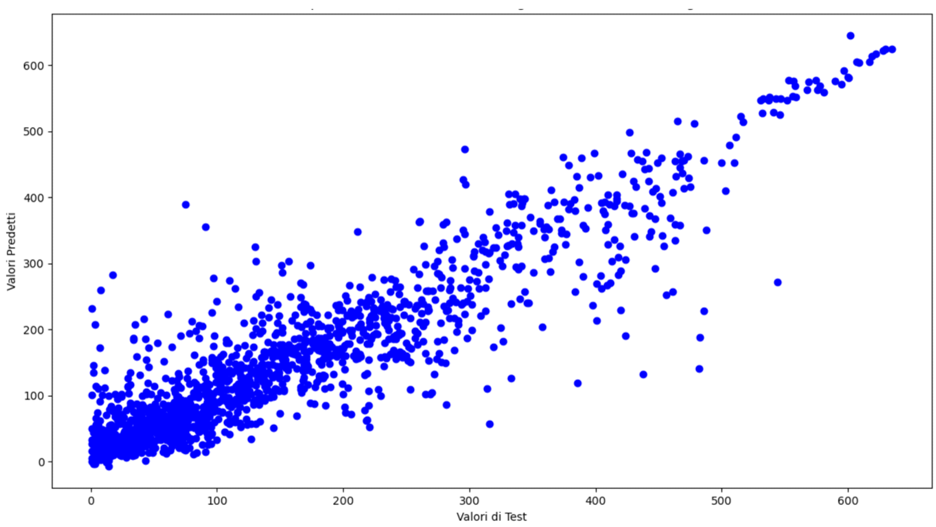


Figure 4.91 - Predictions vs Label with Gradient Boosting

Even the trend of the errors is very similar (Figure 4.92).

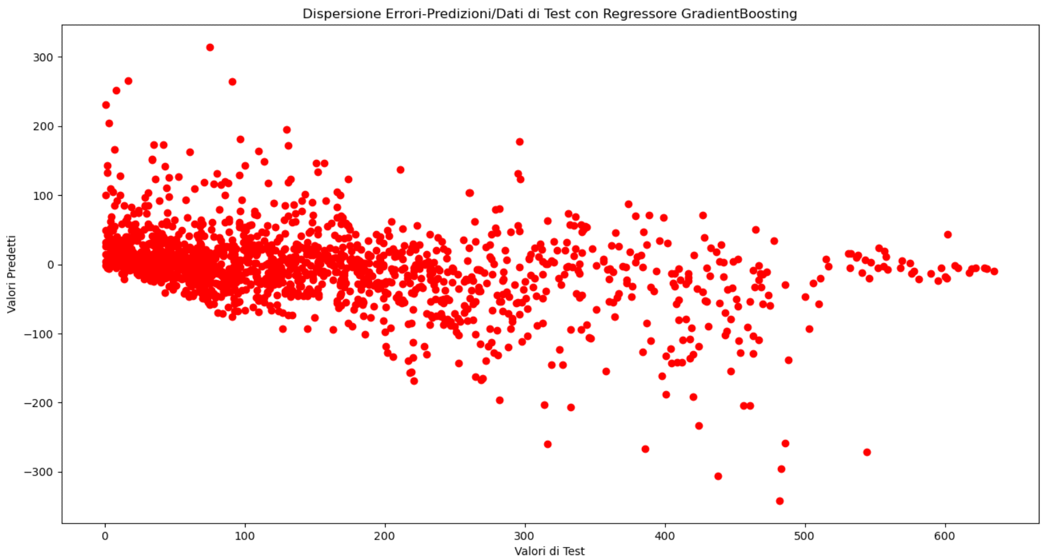


Figure 4.92 - Trend of errors between predicted and test values

In this case, we can also measure the variation of the model’s deviance (Figure 4.93) to change the number of iterations (the deviance being a dispersion index of the values obtained in relation to the average value).

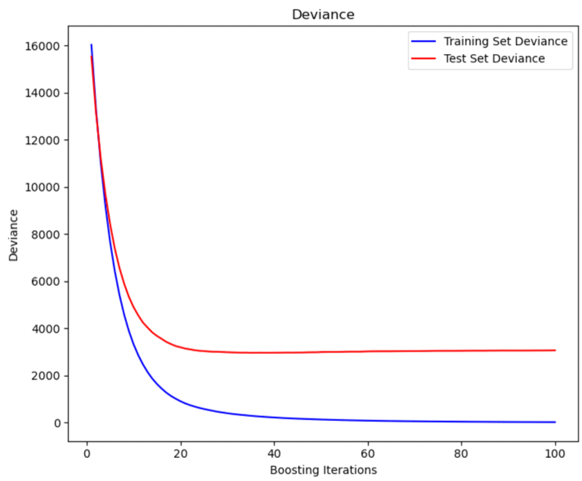


Figure 4.93 - Deviance

As a result, it is worth noting that the statistical deviance remains substantially stable over 30-40 iterations, both in the set-set test and in the training set, implying that increasing the number of iterations in the hope of achieving the best result is pointless. Another measure that we replicate in the case of gradient boosting is the relevance of the features, which is evaluated using both the model's score and the permutation importance technique that we saw previously for Random Forest regressor. The results in Figure 4.94 show that the most significant features in both cases are T1, PS, PF, and FC, which constitute results that are consistent with what was previously obtained.

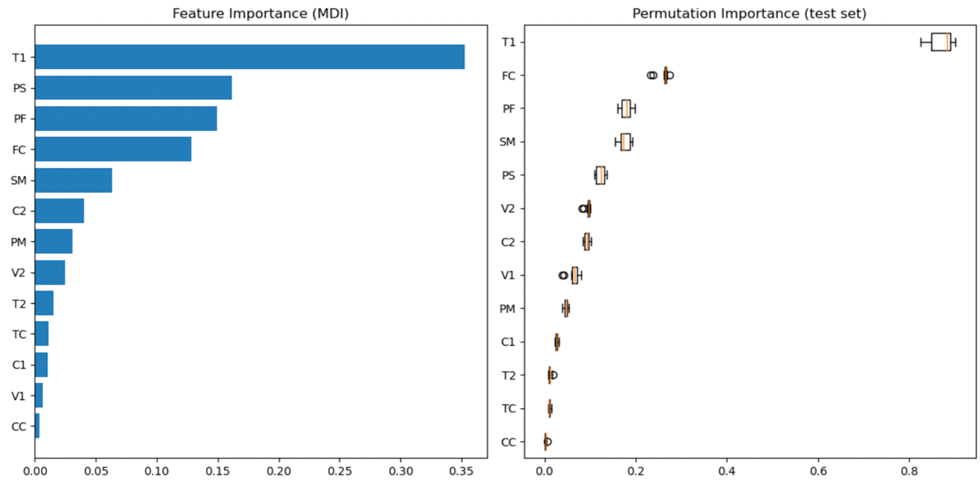


Figure 4.94 - Feature and Permutation importance with Gradient Boosting

4.7.2.9. Outliers

Python provides a method called "LocalOutlierFactor" that can be used to find and assess the impact of outliers.

To assess the impact of possible outlier removal from the dataset, we assume that the quantile that defines the 1 percent threshold is elevated, resulting in the threshold -2054. This threshold determines 50 outlier values that will be removed from the data's features, as well as generating a new label for the features X2 dataset and an array of labels Y2. It then proceeds as before, determining the set of tests and training and training the model (using the Random Forest with the previously determined optimum parameters) on these data. The following metrics values are displayed as a result of the analysis.

That is, the model's performance is slightly worse than that determined by the previous paragraph using the same hyperparameter values and the same dataset cleaned of the 50 records identified as outliers. It concludes that the values we previously considered outliers actually provide information to the model, which the model uses to refine predictions, and thus must be maintained of the metrics.

4.7.3. Equipment Fault

The third accidental machine stop case is represented by Equipment Fault (Causalisation 115), which will be the subject of this paragraph's analysis. The methodology used in this case company is completely different from that used in the previous cases.

4.7.3.1. Feature selection

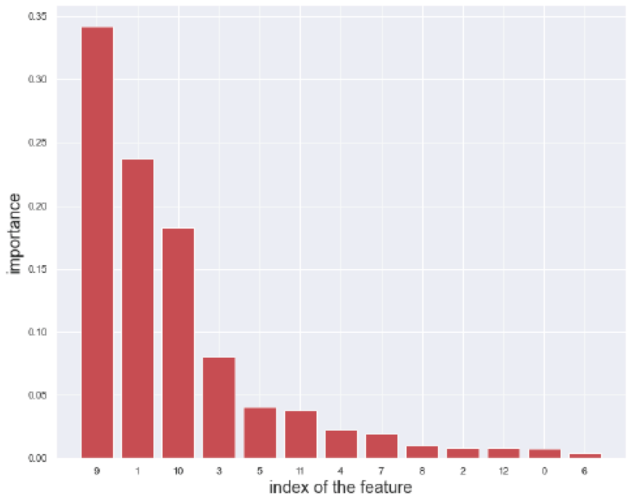
The Random Forest regressor is used once more to search for the most significant features, which, as we have seen in previous cases, has the best performance for the specific type of data. The algorithm is executed using the documented procedure shown in Figure 4.95, and the results are shown in Figure 4.96, where it is highlighted that PS, T1, and FC are the most significant features for the cases in question.

```
# We use a random forest to determine some of the most important/meaningful␣
↪features
# random forest as feature selection
# create an exhaustive random forest (200 trees up to 15 levels deep)
from sklearn import ensemble
rf = ensemble.RandomForestRegressor()
single_rf = ensemble.RandomForestRegressor(n_estimators = 200, max_depth = 15)
single_rf.fit(X,y)
y_pred = single_rf.predict(X)
print("complete")
print(X.shape[1])
```

```
complete
13
```

```
# graph feature importance
import matplotlib.pyplot as plt
importances = single_rf.feature_importances_
indices = np.argsort(importances)[::-1]
feature_names = X.columns
f, ax = plt.subplots(figsize=(11,9))
plt.title("Feature ranking", fontsize = 20)
plt.bar(range(X.shape[1]), importances[indices], color="r", align="center")
plt.xticks(range(X.shape[1]), indices)
plt.xlim([-1, X.shape[1]])
plt.ylabel("importance", fontsize = 18)
plt.xlabel("index of the feature", fontsize = 18)
plt.show()
# list feature importance
important_features = pd.Series(data=single_rf.feature_importances_, index=X.
↪columns)
important_features.sort_values(ascending=False, inplace=True)
print(important_features.head(13))
```

Figure 4.95 - Feature selection with Random Forest regressor (equipment fault)



PS	0.341660
T1	0.237401
FC	0.183147
C2	0.079857
V2	0.040674
SM	0.037803
T2	0.022572
PM	0.019268
PF	0.009578
V1	0.008619
TC	0.008406
C1	0.007202

Figure 4.96 - Feature Ranking (equipment fault)

The following analysis, like the previous one, refers to the application of various regression algorithms to the current causalisation in order to determine which one best predicts the event of equipment fault.

4.7.3.2. SVC regressor

The performance of the regression algorithm based on Support Vector Machine is evaluated using the cross-validation method (Figure 4.97).

```
# Support Vector Machines
# create holdout
import numpy as np
4
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1234)

# choose the model
from sklearn import svm
from sklearn.svm import SVR
svm = svm.SVR()

# set up 5-fold cross-validation
from sklearn import model_selection
cv = model_selection.KFold(5)

# pipeline standardization and model
from sklearn.pipeline import Pipeline
pipeline = Pipeline(steps=[('standardize', preprocessing.StandardScaler())
, ('model', svm) ])

# tune the model
my_C = [0.1, 1, 10, 100]
my_epsilon = [.01, .05, .1, .15]

# run the model using gridsearch, select the model with best search
from sklearn.model_selection import GridSearchCV
optimized_svm = GridSearchCV(estimator=pipeline
, cv=cv
, param_grid =dict(model__C = my_C, model__epsilon = my_epsilon)
, scoring = 'neg_mean_squared_error'
, verbose = 1
, n_jobs = -1
)

optimized_svm.fit(X_train, y_train)
```

Figure 4.97 - Execution of SVC algorithm (equipment faults)

The algorithm's execution yields the results shown in Figure 4.98.

```
# show the best model estimators
print(optimized_svm.best_estimator_)

# evaluate metrics on holdout
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
y_pred = optimized_svm.predict(X_test)
print("SVM Mean Squared Error: ", mean_squared_error(y_test, y_pred))
print("SVM Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
print("SVM r-squared: ", r2_score(y_test, y_pred))

Fitting 5 folds for each of 16 candidates, totalling 80 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 5.5s
Pipeline(steps=[('standardize', StandardScaler()),
('model', SVR(C=100, epsilon=0.01))])
SVM Mean Squared Error: 6072.401417693772
SVM Mean Absolute Error: 56.00609241137927
SVM r-squared: 0.6048290644506256
```

Figure 4.98 - Results of the execution of the algorithm SVC

How to understand that the results of this algorithm are not particularly satisfactory.

4.7.3.3. Gradient Boosting regressor

Another algorithm's performance is evaluated, which has given good performance of previous causalisation using a completely similar procedure to the previous one. The results are reported to Figure 4.99 below.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1234)

# choose the model
from sklearn.ensemble import GradientBoostingRegressor
gb = ensemble.GradientBoostingRegressor()

# set up 5-fold cross-validation
from sklearn import model_selection
cv = model_selection.KFold(5)

# pipeline standardization and model
from sklearn.pipeline import Pipeline
pipeline = Pipeline(steps=[('standardize', preprocessing.StandardScaler())
, ('model', gb) ])

# tune the model
my_alpha = [.5, .75, .9]
my_n_estimators= [500]
my_learning_rate = [0.005, .01]
my_max_depth = [4, 5, 6]

# run the model using gridsearch, select the model with best search
from sklearn.model_selection import GridSearchCV
optimized_gb = GridSearchCV(estimator=pipeline
, cv=cv
, param_grid =dict(model__max_depth = my_max_depth, model__n_estimators = my_n_estimators,
                   model__learning_rate = my_learning_rate, model__alpha = my_alpha)
, scoring = 'neg_mean_squared_error'
, verbose = 1
, n_jobs = -1
)
optimized_gb.fit(X_train, y_train)
```

Figure 4.99 - Execution of Gradient Boosting (equipment faults)

The execution of the algorithm produces the following results in Figure 4.100:

```
# show the best model estimators
print(optimized_gb.best_estimator_)
# evaluate metrics on holdout
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
y_pred = optimized_gb.predict(X_test)
print("Gradient Boosting Mean Squared Error: ", mean_squared_error(y_test, y_pred))
print("Gradient Boosting Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
print("Gradient Boosting r-squared: ", r2_score(y_test, y_pred))
Pipeline(steps=[('standardize', StandardScaler()),
('model',
GradientBoostingRegressor(learning_rate=0.01, max_depth=6,
n_estimators=500))])
Gradient Boosting Mean Squared Error: 3307.7305365263564
Gradient Boosting Mean Absolute Error: 36.68880856509504
Gradient Boosting r-squared: 0.7847443077699591
```

Figure 4.100 - Results of the execution of the algorithm Gradient Boosting

In contrast to the SVC, the outcome this time is acceptable.

4.7.3.4. Random Forest regressor

The performance of the Random Forest regression algorithm, which has previously produced the best results, is evaluated (Figure 4.101).

```
# random forest regression
# create holdout
import numpy as np
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1234)

# choose the model
from sklearn.ensemble import RandomForestRegressor
rf = ensemble.RandomForestRegressor()

# set up 5-fold cross-validation
from sklearn import model_selection
cv = model_selection.KFold(5)

# pipeline standardization and model
from sklearn.pipeline import Pipeline
pipeline = Pipeline(steps=[('standardize', preprocessing.StandardScaler())
, ('model', rf) ])

# tune the model
my_n_estimators = [200, 400]
my_max_features = ['auto']
my_min_samples_leaf = [1, 2, 4, 10]
my_max_depth = [20, 40]
my_min_samples_split = [2]

# run the model using gridsearch, select the model with best search
from sklearn.model_selection import GridSearchCV

optimized_rf = GridSearchCV(estimator=pipeline
, cv=cv
, param_grid =dict(model__min_samples_leaf =□
,→my_min_samples_leaf, model__max_depth = my_max_depth)
, scoring = 'neg_mean_squared_error'
, verbose = 1
, n_jobs = -1
)

optimized_rf.fit(X_train, y_train)
```

Figure 4.101 - Execution of Random Forest (equipment faults)

The results of the algorithm's execution, as shown in Figure 4.102, are interchangeable to those obtained at the previous point with the gradient boosting algorithm.

```
# show the best model estimators
print(optimized_rf.best_estimator_)
print(optimized_rf.best_params_)
# evaluate metrics on holdout

from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
y_pred = optimized_rf.predict(X_test)

print("Random Forest Mean Squared Error: ", mean_squared_error(y_test, y_pred))
print("Random Forest Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
print("Random Forest r-squared: ", r2_score(y_test, y_pred))
Fitting 5 folds for each of 8 candidates, totalling 40 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 40 out of 40 | elapsed: 2.8s finished
Pipeline(steps=[('standardize', StandardScaler()),
('model', RandomForestRegressor(max_depth=40))])
{'model__max_depth': 40, 'model__min_samples_leaf': 1}
Random Forest Mean Squared Error: 3220.57001680059
Random Forest Mean Absolute Error: 34.329578519181965
Random Forest r-squared: 0.790416413705288
```

Figure 4.102 - Results of the execution of the algorithm Random Forest

The mean absolute error of the prediction is approximately 34 cycles (about 17 minutes). Figure 4.103 depicts the differences between the aforementioned values and the actual test data values.

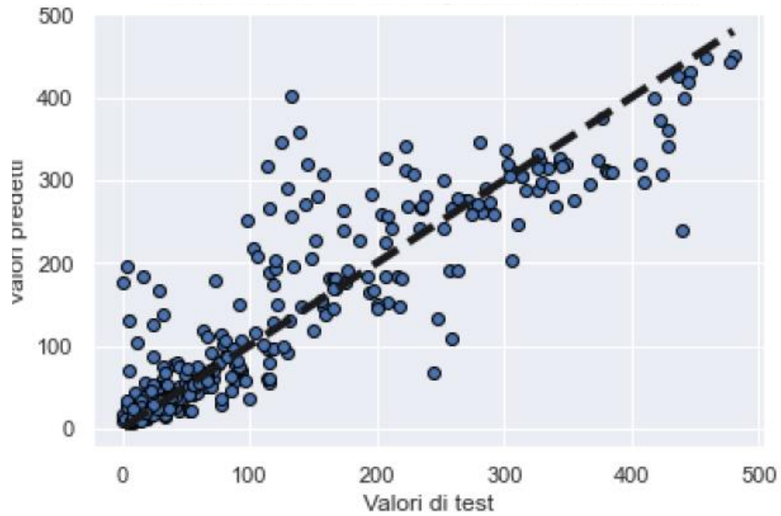


Figure 4.103 - Prediction deviations compared to test data

4.7.4. Electric Fault

The fourth c accidental machine stop case is represented by Electric Fault (causalisation 113) which will be the subject of this paragraph's analysis the methodology used is a completely similar procedure to the previous one, equipment fault.

4.7.4.1. Feature selection

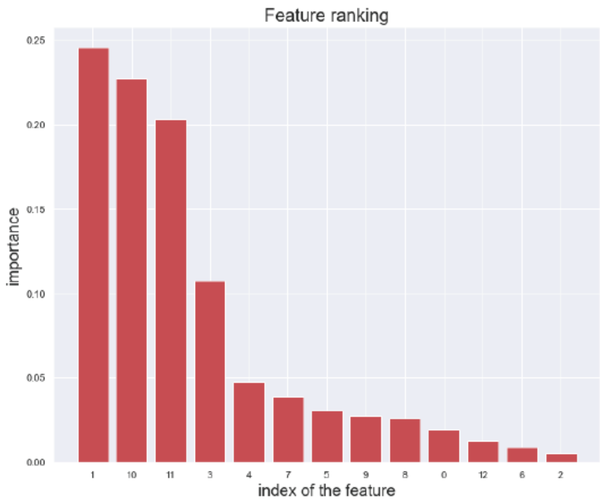
The Random Forest regressor is used once more to search for the most significant features, which, as we have seen in previous cases, has the best performance for the specific type of data. The algorithm is executed using the documented procedure shown in Figure 4.104, and the results are shown in Figure 4.105, where it is highlighted that T1, SM, and FC are the most significant features for the cases in question.

```
# We use a random forest to determine some of the most important/meaningful␣
 ↪features
# random forest as feature selection
# create an exhaustive random forest (200 trees up to 15 levels deep)
from sklearn import ensemble
rf = ensemble.RandomForestRegressor()
single_rf = ensemble.RandomForestRegressor(n_estimators = 200, max_depth = 15)
single_rf.fit(X,y)
y_pred = single_rf.predict(X)
print("complete")
print(X.shape[1])
```

```
complete
13
```

```
# graph feature importance
import matplotlib.pyplot as plt
importances = single_rf.feature_importances_
indices = np.argsort(importances)[::-1]
feature_names = X.columns
f, ax = plt.subplots(figsize=(11,9))
plt.title("Feature ranking", fontsize = 20)
plt.bar(range(X.shape[1]), importances[indices], color="r", align="center")
plt.xticks(range(X.shape[1]), indices)
plt.xlim([-1, X.shape[1]])
plt.ylabel("importance", fontsize = 18)
plt.xlabel("index of the feature", fontsize = 18)
plt.show()
# list feature importance
important_features = pd.Series(data=single_rf.feature_importances_, index=X.
 ↪columns)
important_features.sort_values(ascending=False, inplace=True)
print(important_features.head(13))
```

Figure 4.104 - Feature selection with Random Forest regressor




```
T1    0.245363
FC    0.226953
SM    0.203045
C2    0.107670
T2    0.047303
PM    0.039236
V2    0.030448
PS    0.027438
PF    0.026092
C1    0.019487
TC    0.012565
CC    0.009173
V1    0.005226
dtype: float64
```

Figure 4.105 - Feature Ranking (electric fault)

The following analysis, like the previous one, refers to the application of various regression algorithms to the current causalisation in order to determine which one best predicts the event of electric fault.

4.7.4.2. SVC regressor

The performance of the regression algorithm based on Support Vector Machine is evaluated using the cross-validation method (Figure 4.106).

```
# Support Vector Machines
# create holdout
import numpy as np
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1234)

# choose the model
from sklearn import svm
from sklearn.svm import SVR
svm = svm.SVR()

# set up 5-fold cross-validation
from sklearn import model_selection
cv = model_selection.KFold(5)

# pipeline standardization and model
from sklearn.pipeline import Pipeline
pipeline = Pipeline(steps=[('standardize', preprocessing.StandardScaler())
, ('model', svm) ])

# tune the model
my_C = [0.1, 1, 10, 100]
my_epsilon = [.01, .05, .1, .15]

# run the model using gridsearch, select the model with best search
from sklearn.model_selection import GridSearchCV
optimized_svm = GridSearchCV(estimator=pipeline
, cv=cv
, param_grid =dict(model__C = my_C, model__epsilon= my_epsilon)
, scoring = 'neg_mean_squared_error'
, verbose = 1
, n_jobs = -1
)
optimized_svm.fit(X_train, y_train)
```

Figure 4.106 - Execution of SVC (electric fault)

The algorithm's execution yields the results shown in Figure 4.107:

```
# show the best model estimators
print(optimized_svm.best_estimator_)
# evaluate metrics on holdout
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
y_pred = optimized_svm.predict(X_test)
print("SVM Mean Squared Error: ", mean_squared_error(y_test, y_pred))
print("SVM Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
print("SVM r-squared: ", r2_score(y_test, y_pred))

Fitting 5 folds for each of 16 candidates, totalling 80 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
Pipeline(steps=[('standardize', StandardScaler()),
('model', SVR(C=100, epsilon=0.15))])

SVM Mean Squared Error: 3256.3434000183142
SVM Mean Absolute Error: 40.53719130965393
SVM r-squared: 0.3775670214717143
[Parallel(n_jobs=-1)]: Done 80 out of 80 | elapsed
```

Figure 4.107 - Results of the execution of the algorithm SVC

4.7.4.3. Gradient Boosting regressor

Another algorithm's performance is evaluated, which has given good performance of previous causalisation using a completely similar procedure to the previous one. The results are reported to Figure 4.108 below.

```
# Gradient Boosting
# create holdout
import numpy as np
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1234)

# choose the model
from sklearn.ensemble import GradientBoostingRegressor
gb = ensemble.GradientBoostingRegressor()

# set up 5-fold cross-validation
from sklearn import model_selection
cv = model_selection.KFold(5)

# pipeline standardization and model
from sklearn.pipeline import Pipeline
pipeline = Pipeline(steps=[('standardize', preprocessing.StandardScaler())
, ('model', gb) ])

# tune the model
my_alpha = [.5, .75, .9]
my_n_estimators= [500]
my_learning_rate = [0.005, .01]
my_max_depth = [4, 5, 6]

# run the model using gridsearch, select the model with best search
from sklearn.model_selection import GridSearchCV
optimized_gb = GridSearchCV(estimator=pipeline
, cv=cv
, param_grid =dict(model__max_depth = my_max_depth,
                   model__n_estimators = my_n_estimators,
                   model__learning_rate = my_learning_rate,
                   model__alpha = my_alpha)
, scoring = 'neg_mean_squared_error'
, verbose = 1
, n_jobs = -1
)
optimized_gb.fit(X_train, y_train)
```

Figure 4.108 - Execution of Gradient Boosting (electric fault)

The execution of the algorithm produces the following results in Figure 4.109

```
# show the best model estimators
print(optimized_gb.best_estimator_)

# evaluate metrics on holdout
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
y_pred = optimized_gb.predict(X_test)
print("Gradient Boosting Mean Squared Error: ", mean_squared_error(y_test, y_pred))
print("Gradient Boosting Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
print("Gradient Boosting r-squared: ", r2_score(y_test, y_pred))

Fitting 5 folds for each of 18 candidates, totalling 90 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 7.6s
[Parallel(n_jobs=-1)]: Done 90 out of 90 | elapsed: 22.7s finished
Pipeline(steps=[('standardize', StandardScaler()),
('model',
GradientBoostingRegressor(learning_rate=0.01, max_depth=6,
n_estimators=500))])
Gradient Boosting Mean Squared Error: 1797.7029397763945
Gradient Boosting Mean Absolute Error: 28.53988682457056
Gradient Boosting r-squared: 0.6563785025535749
```

Figure 4.109 - Results of the execution of the algorithm Gradient Boosting

The outcome is acceptable.

4.7.4.4. Random Forest regressor

The performance of the Random Forest regression algorithm, which has previously produced the best results, is evaluated (Figure 4.110).

```
# choose the model
from sklearn.ensemble import RandomForestRegressor
rf = ensemble.RandomForestRegressor()

# set up 5-fold cross-validation
from sklearn import model_selection
cv = model_selection.KFold(5)

# pipeline standardization and model
from sklearn.pipeline import Pipeline
pipeline = Pipeline(steps=[('standardize', preprocessing.StandardScaler())
, ('model', rf) ])

# tune the model
my_n_estimators = [200, 400]
my_max_features = ['auto']
my_min_samples_leaf = [1, 2, 4, 10]
my_max_depth = [20, 40]
my_min_samples_split = [2]

# run the model using gridsearch, select the model with best search
from sklearn.model_selection import GridSearchCV
optimized_rf = GridSearchCV(estimator=pipeline
, cv=cv
, param_grid =dict(model__min_samples_leaf = my_min_samples_leaf, model__max_depth = my_max_depth)
, scoring = 'neg_mean_squared_error'
, verbose = 1
, n_jobs = -1
)
optimized_rf.fit(X_train, y_train)
```

Figure 4.110 - Execution of Random Forest (electric faults)

The results of the algorithm's execution, as shown in Figure 4.111, are interchangeable to those obtained at the previous point with the gradient boosting algorithm:

```
# show the best model estimators
print(optimized_rf.best_estimator_)
print(optimized_rf.best_params_)

# evaluate metrics on holdout
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
y_pred = optimized_rf.predict(X_test)
print("Random Forest Mean Squared Error: ", mean_squared_error(y_test, y_pred))
print("Random Forest Mean Absolute Error: ", mean_absolute_error(y_test, y_pred))
print("Random Forest r-squared: ", r2_score(y_test, y_pred))

Fitting 5 folds for each of 8 candidates, totalling 40 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 40 out of 40 | elapsed: 8.3s finished
Pipeline(steps=[('standardize', StandardScaler()),
('model',
RandomForestRegressor(max_depth=40, min_samples_leaf=2))])
{'model__max_depth': 40, 'model__min_samples_leaf': 2}

Random Forest Mean Squared Error: 1478.9489466081443
Random Forest Mean Absolute Error: 25.457723161171575
Random Forest r-squared: 0.7173066581603769
```

Figure 4.111 - Results of the execution of the algorithm Random Forest

The mean absolute error of the prediction is approximately 25 cycles (about 12 minutes). Figure 4.112 depicts the differences between the aforementioned values and the actual test data values.

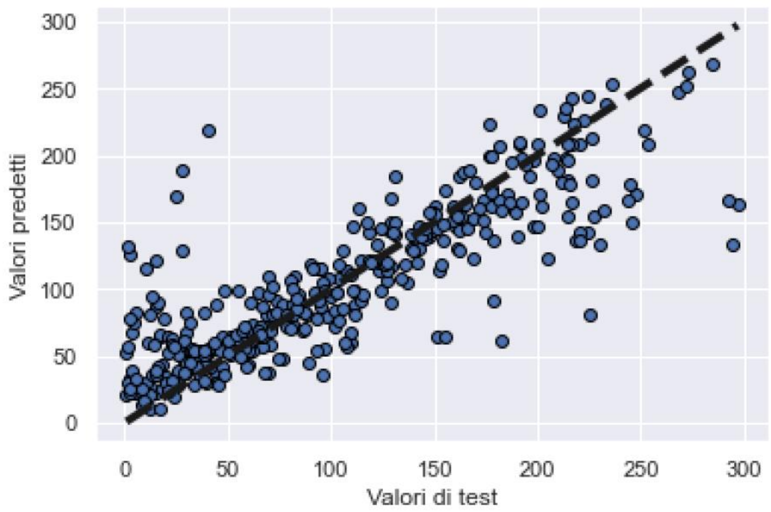


Figure 4.112 - Prediction deviations compared to test data

4.7.5. Comparison and summary of results

The findings of the performance metrics of the analyses done on the four primary causalisations that describe the die-casting process of investigation are shown in Table 4.8 below. Forecasts made on Microstops and mechanical faults yielded the best outcomes. It is not surprising given the amount of data available for training was significantly greater than that of equipment faults and electric faults, which analyses returned performance metrics with lower values, but it is satisfactory in light of all the assumptions described in the preceding paragraph. Given that the beginning dataset had a high level of variability and that the causalisations were impacted by the operators' activities, it is plausible to conclude that the random forest regression technique is the best fit for this sort of dataset.

Table 4.8 - Test case regression performance metrics

		MAE	MSE	R^2
Microstop	**Random Forest**	54.14	6814.63	0.66
	Decision tree	36.84	3422.25	0.82
	SVC	77.84	4523.09	0.55
	Gradient Boosting	35.79	4987.31	0.76
		MAE	**MSE**	**R^2**
Mechanical fault	**ElasticNet regressor**	82.47	11734.83	0.36
	Linear regressor	82.61	11747.1	0.36
	SVC regression	82.06	11997.86	0.35
	Decision Tree	41.28	3838.22	0.79
	Random Forest	36.06	2950.96	0.84
	Gradient Boosting	35.09	3046.37	0.83
		MAE	**MSE**	**R^2**
Equipment fault	**SVC regression**	56.01	6072.4	0.6
	Gradient Boosting	36.69	3307.7	0.78
	Random Forest	34.33	3220.6	0.79
		MAE	**MSE**	**R^2**
Electric fault	**SVC regression**	40.53	3256.3	0.38
	Gradient Boosting	28.53	1797.7	0.66
	Random Forest	25.46	1478.9	0.72

Instead, the outcomes of the application of classification approaches are provided in Table 4.9, which is always separated by the four primary causalisations. The characteristics (and hence the process parameters) that are associated to each other are the most accidents on that single causal for each causalisation.

Table 4.9 - Importance ranking of the most significant features for each causalisation

Causalisation	Importance ranking	KBest	Random Forest
Microstops	1	SM	T1
	2	T1	FC
	3	V1	SM
	4	CC	PM
Mechanical Faults	1	PF	T1
	2	PS	FC
	3	PM	PF
	4	SM	PS
Equipment Faults	1	PS	PS
	2	T1	T1
	3	FC	FC
	4	C2	F2
Electrical Faults	1	T1	T1
	2	SM	FC
	3	FC	SM
	4	C2	C2

In terms of data-driven methodology, due to the complexity and variety that define production systems, a single technique cannot be fitted to all of the many uses. To address issues such as defect detection and diagnosis (Zhang and Hoo, 2011), prediction or classification accuracy (Ghosh et al., 2011), hybrid techniques (Tidriri et al., 2016) have been developed. This indicates that modelling patterns must be created and established in order to accomplish a modular approach. Indeed, modularity is concerned with the transition from rigid systems and stiff production models to a smart and agile system.

4.8. FUTURE PERSPECTIVES ON THE COMPANY TEST CASE

Master Italy is a very active and ever-expanding SME. Although it has production and quality control management systems, the human effect (of operators) on all processes is still quite significant. A single technique cannot be applied to all of the numerous applications due to the complexity and variety that define production systems. To address issues like as defect detection and diagnosis, prediction, and classification accuracy, hybrid techniques have been created. This indicates that modelling patterns must be created and established in order to accomplish a modular approach. Indeed, modularity is concerned with the transition from rigid systems and inflexible production models to a smart and agile system.

The Master can choose between two paths in order to design an integrated and effective model aimed at improving the sustainability performances of a technological process and facilitating the management of smart manufacturing processes, thereby driving the company implementing an information structure for predictive manufacturing toward a transformation into an autonomous factory. The first is the development of an automated and invariant, customizable and ad hoc algorithm that is also applicable to various sorts of parameter sets resulting from the installation of additional sensors on the machine or from wholly other process types (e.g., painting) Figure 4.113 depicts the concept of a brand-new Master FoF, that is, a whole plant that is integrated and as automated as feasible.

Figure 4.113 - Master FoF

The second option is to purchase a license for a platform that fits our standards and allows any practitioner to deploy any ML approach that is required.
There are various AI platforms on the market, some of which are more user-friendly and plug-and-play, while others necessitate more programming knowledge. For each of them, the benchmark was based on trial versions, user manuals, and use cases to find: 1) Type of Data Acquisition System 2) Is it equipped with Data Transformation/ETL tools 3) Cloud or on-premises accessibility; 4) Memory accessible; 5) The quantity of data it can handle; 6) ML Algorithm that can be deployed; 7) Modelling performance measures; 8) Pricing; and 9) Case or tutorial availability.
For the sake of brevity, only the key platforms examined are listed:

- Microsoft Azure Machine Learning (https://azure.microsoft.com/en-us/overview/ai-platform/)

- SAP Leonardo Data Intelligence (https://www.sap.com/italy/products/data-intelligence.html)

- GE Predix Essential/Edge (https://www.ge.com/digital/iiot-platform)

- AWS Amazon SageMaker (https://aws.amazon.com/it/sagemaker/?c=ml&sec=srv#sm_studio)

- Salesforce Einstein Analytics (https://www.salesforce.com/eu/products/einstein/overview/)

- IBM Watson (https://www.ibm.com/it-it/watson/products-services)

- Dassault 3D Experience (https://www.3ds.com/products-services/biovia/products/data-science/pipeline-pilot/analytics-machine-learning/)

- PTC ThingWorx (https://www.ptc.com/-/media/Files/PDFs/ThingWorx/ThingWorx-Analyze-Brochure.pdf)

- Oracle Analytics (https://www.oracle.com/it/business-analytics/)

The benchmark directs the choice between two platforms currently available: Microsoft Azure Machine Learning or IBM Watson, which are the most comprehensive and integrated with the company's existing management systems, as well as the IoT architecture, which may be designed to monitor and control all processes in real-time.

5. OTHER APPLICATIONS

The know-how gained during the doctoral program, as described in previous chapters, was useful for the realization of cross-disciplinary projects, making the doctorate itself extremely multidisciplinary.

5.1. ADDITIVE MANUFACTURING: DIRECT LASER DEPOSITION

Among the I4.0 Key Enabling Technologies, Additive Manufacturing (AM) is seen as a major element in the deployment of a new production paradigm (Chiarello et al., 2018; Ruppert et al., 2018). AM is one of the Advanced Manufacturing Systems that significantly contributes in the Factory 4.0 network for the creation of prototypes or the production of custom components. The parallel development of hardware and software, as well as intensive research for the adoption of new materials (metals, polymers, ceramics, and multi-material composites), has been critical to the success of AM technologies, resulting in an extension of application sectors.
In terms of sustainability, AM technologies are usually regarded as “greener” than traditional manufacturing techniques. In reality, these adhere to some of the sustainability principles (Ford and Despeisse, 2016), such as "reduction" or "reconfiguring," as well as other advantages linked to possible societal consequences, such as new opportunities from the circular economy. Several investigations have found that the long-term viability of AM technology is not always guaranteed (Faludi et al., 2015). It should be evaluated from many angles in order to take into account all process factors such as machine tool life cycle, tooling supply chain, energy management, product use impacts, end-of-life concerns, recycling rates, disposal costs, pollutions, and so on. Various techniques have been developed throughout the years to facilitate the examination of the aforementioned components of manufacturing processes for sustainability. With such a broad range of emerging technologies, it is important to investigate which approaches are most suited for qualitative and quantitative assessment of their long-term viability.
The strategy adopted in this application entails developing a real-time monitoring framework for the Direct Laser Metal Deposition (DLMD) process, beginning with a description of its general technological features and defining the thermodynamic model for the sustainability criteria. The DLMD is a high-tech AM technique that is mostly used in high-tech industrial areas (Taddese et al., 2020). It is mostly used for fixing and refitting old components, but it is now being utilised to create 3D parts from scratch using a variety of metallic materials (Ahn, 2016).

5.1.1. The technology

The DLMD is a sub-category of the Direct Energy Deposition (DED) family of technologies, in which a laser is used as an energy source. The laser beam is focused on a metal substrate, producing extremely high temperatures that cause the workpiece's interested region to melt. In the shape of wire or powder, material is introduced to the molten pool. The latter is carried by means of an inert carrier gas (Ar, He, N_2), which also serves to protect the molten pool from corrosion. A single-track deposition is generated by moving a nozzle, and the final component is produced layer by layer by following particular pathways and methods. The laser's extraordinarily high-power density dictates its remarkable metallurgical characteristics. Temperature control during the process is critical for achieving acceptable component quality and increasing process efficiency. DLMD, like other AM technologies, can process a variety of materials, including steel alloys, aluminium alloys, titanium alloys, nickel alloys, and superalloys. The DLMD includes several process factors, and careful design is necessary to generate components that meet increased standards (Singla et al., 2021).
The laser source, which may be identified by the active medium used to create the laser beam, is at the heart of the system. All of these sources have rather low efficiency, turning a significant amount of electrical energy into wasted heat, which may jeopardise laser source functioning. As a result, heat exchangers and chillers are used to promote heat dispersion into the environment (Zobler and Mantwill, 2018). To dose and warm the

powders, the filler material is often delivered via specialised powder feeders. These are carried to the end of the deposition head by a carrier gas via a piping system. The toolpaths are carried out by 3- or 5-axis CNC machines, or, in more modern designs, by anthropomorphic robots (DebRoy et al., 2018).

5.1.2. State-of-art of Additive Manufacturing technologies sustainability assessment

This work focused on thermodynamic analysis with the EA as a major added value of the traditional and overused LCA. The reference to strategic coupled EA-LCA modelling is also justified by the fact that only two works have been found in the literature on EA on AM. The first, (Jiang et al., 2019), consists only of an Em-LCA implementation for AM sustainability evaluation. Critical problems and possible improvements are identified, but always from a sustainability-only perspective. Nagarajan and Haapala in (Nagarajan and Haapala, 2018) attempted to identify and characterize the factors affecting the systemic environmental performance of additive manufacturing, using EA, LCA and CExD as the final use of energy. There is no mention on prospective process modelling.
About LCA on AM, in (Jiang and Ma, 2020) a state of art of AM technologies is presented regarding the path planning strategies to improve printed qualities, saving materials/time and achieving objective printed properties. Another approach of using the LCA in additive manufacturing technologies is used by (Paris et al., 2016). In the early design phase, they made a comparison of alternatives for decision making, i.e. the LCA is used to choose the most sustainable solution (Guarino, 2019). No modelling was carried out. (Ribeiro et al., 2020) used three different impact assessment methods of LCA to explore the literature on AM sustainability on the three-bottom line assessment framework and its three dimensions (environmental, economic and social). A similar work has been conducted by (Arrizubieta et al., 2020) in which LCA was used to discuss on the advantages and disadvantages of AM technologies. They stated that, although AM is considered an environmentally friendlier process than traditional manufacturing systems, there are not enough LCA studies to prove it, to highlight that this technology is still little explored. They make no mention of process modelling via LCA. A research involving a technique roughly similar to the one that will be used in this paper is presented in (Meteyer et al., 2014), where a unit-process level model is created to provide a complete parametrical LCI for a further LCA of their additive manufacturing process.
To assess the long-term viability of this new disruptive technology, the EA was chosen as the best method among a number of options. It should be noted that no such thermodynamic modelling technique has been used to the DLMD process in the literature to date, making this work unique.

5.1.3. Sensing system for DLMD

To maximise overall process performance, a thorough understanding of all data relating to materials, energy, machinery, and auxiliary equipment is required. It is critical to integrate an adequate sensing and monitoring system into industrial equipment. Finally, the volume of technical data obtained in ad hoc structured datasets enables a multi-model solution to the challenge of prognostic health management (Jimenez, 2020). It may be accomplished, through the interoperability of a physics-based model, a knowledge-based model, and a data-driven model, which structures process knowledge to provide a smooth transition to I4.0 smart manufacturing modelling.
Figure 5.1(a) shows a schematic representation of the prototype system located at the Polytechnic University of Bari, consisting of:

- a fiber laser source with a nominal power of 4 kW and a wavelength of 1.070 μm (YLS 4000 IPG Photonics Ytterbium Laser System)
- a 5-axis machine equipped with a deposition head and a coaxial nozzle
- a 11.57 kW chiller system for the laser source (chiller for core from now on)
- a 1.4 kW chiller system for the nozzle and the optics (chiller for nozzle from now on)

– a 600 W chiller system for the fiber optics cables (chiller for fiber from now on)
– an external pre-heated powder feeder
– a 2.2 kW powder suction system and a gravimetric dispenser.

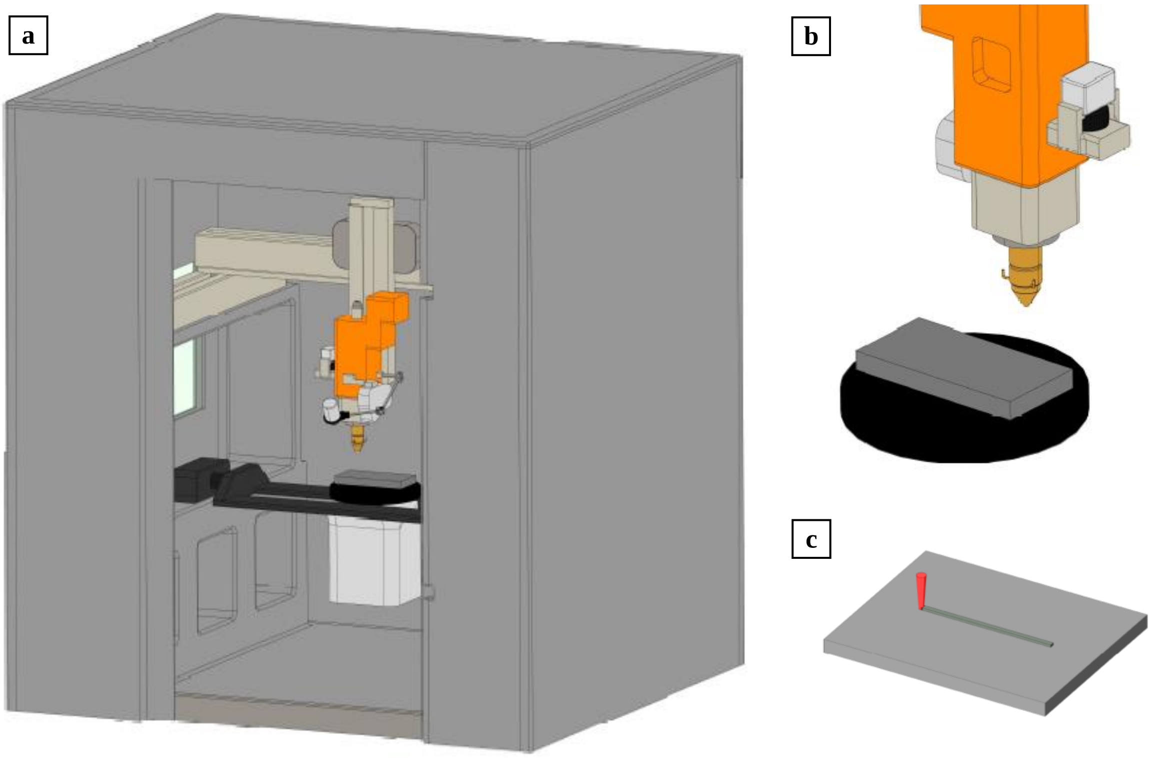


Figure 5.1 - (a) DLMD machine, (b) deposition head and (c) single-track deposition.

The working space is encircled by a glove-box chamber to protect the operators and avoid environmental contamination. In addition to the laser deposition head illustrated in Figure 5.1(b), which it has monitoring equipment such as a coaxial camera and a pyrometer. An AISI 316L stainless steel powder was placed on a substrate of the same material for this study. Argon was used as a powder transport gas as well as a shielding gas to avoid clad oxidation. A single-track deposition was carried out in order to test the viability of the suggested thermodynamic model and monitoring system, as represented in Figure 5.1(c).
The most complicated deposition patterns are built on single-track depositions, and their analysis is critical for the DLMD process. Table 5.1. shows the major process parameters utilised for the deposition test, which were chosen based on prior work on the viability of DLMD (Mazzarisi et al., 2020a), which investigated numerous materials and settings.

Table 5.1 - DLMD process parameters.

Process Parameter	**Value**	**Unit**
Laser power	600	W
Laser spot diameter	1.50	mm
Scanning speed	500	mm/min
Powder feed rate	10	g/min
Argon gas flow rate	10	l/min

5.1.4. **Thermodynamic modelling for the DLMD sensing system**

The thermodynamic model of the protype DLMD system was created, as well as its optimum sub-unit partitioning and all specific in/out flows. The model produces a collection of parameters that may be examined in real-time or at a high enough sample rate to detect patterns in energy/exergy consumption and losses that make the process less efficient and less sustainable (Selicati et al., 2021b).
Figure 5.2 depicts the design of the DLMD process model. This graphical form is also effective for emphasising the linkages within the complicated network of sub-units that comprise the model and the parameters that EA analyses. The major sub-unit is the glove-box (which is chilled by the nozzle chiller) within which the laser deposition process takes place. The laser beam is produced by a laser source that is linked to two chiller systems (chiller for core and chiller for fiber). The powder feeder regulates the powder flow in parallel. Finally, there is the suction system, which cleans the post-deposition environment. Furthermore, the image delineates what would eventually be the system boundaries for a potential LCA combined with the EA.

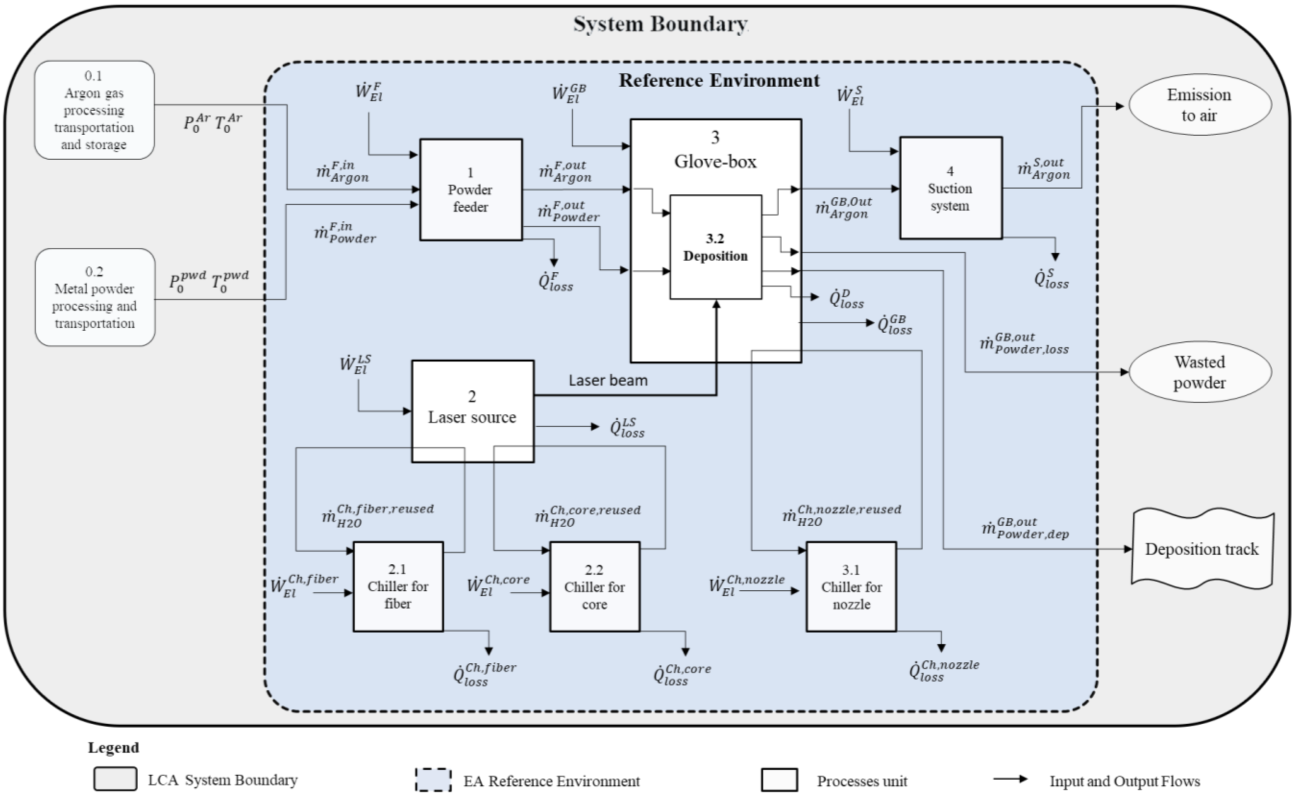


Figure 5.2 - DLMD thermodynamic model.

Table 5.2 shows the set of important thermodynamic parameters to be measured (M) and monitored (C) using thermodynamic laws based on the EA (and eventually a combined LCA) data inventory of the whole process, connected to the functional unit.

A network of sensors is required to create a database containing the values of all the parameters required for the study. A monitoring system has been developed based on Raspberry Pi ("Raspberry Pi 4 Model B specifications – Raspberry Pi," n.d.) and Python v. 3.8 ("Welcome to Python.org," n.d.) to meet the thermodynamic model requirements. The first has numerous advantages, including compact size, low cost, and high flexibility owing to the extensive range of software and hardware capabilities available

Table 5.2 - Essential thermodynamic parameters of DLMD process.

Material/Energy	**Data flow in**	**Data flow out**	**M/C**
Metal powder	mass flow [kg/s]	mass flow [kg/s]	M
	temperature [K]	temperature [K]	M
Gasses	volume flow [l/s]	volume flow [l/s]	M
	temperature [K]	temperature [K]	M
Cooling fluid	volume flow [l/s]	volume flow [l/s]	M
	temperature [K]	temperature [K]	M
Electricity		electric power [W]	M
Heat		heat loss [J]	C
Exergy		exergy loss [J]	C
Environmental Impact		GWP_{100a} [$kgCO_{2eq}$] or EDIP2003	C

. The latter is a relatively new multiplatform interpreter that is widely utilised in real-time applications, robotics, deep learning, image processing, database servers, and monitoring systems by the scientific community. It has risen to the top of the most widely used scientific software in the recent decade, owing to its simple, straightforward syntax and flexibility. In this specific application the Raspberry Pi 4 has been selected, showing the following specifications:

- Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC at 1.5 GHz
- 2 GB, 4 GB or 8 GB LPDDR4-3200 SDRAM (depending on the model)
- Gigabit Ethernet.

This option ensures both hardware and software flexibility in order to carry out the EA. Specifically, the process temperature was measured using an off-axis pyrometer (CellaTemp®, Keller ITS) connected to the deposition head to record the thermal characteristics of the whole process. The emission spectra for AISI 316L stainless steel was determined using prior works and laser processing literature data (Waqar et al., 2021). To monitor the melt pool during the deposition process, a coaxial CCD camera (IDS®, uEye RE) incorporated into the deposition head was used. It is important to evaluate the uniformity of the procedure and identify the size of the treated area with a maximum frame rate of 40 Hz.

Electrical energy was measured using a set of energy metres (Siemens Sentron PAC 3200). More than 200 parameters were stored as float numbers in the power meter. Active power [kW], reactive power [kVAr], active energy [kWh], reactive energy [kVArh], voltage and current harmonics are among the variables collected. Actual power measurement accuracy is in the order of 0.5% (Siemens, n.d.).

To determine the maximum acquisition rate, the monitoring system's communication capability was evaluated using the Siemens Sentron PAC 3200 smart meter. Following a first test, the monitoring system was able to acquire a number of 100-word registers at 149 Hz from a Modbus TCP device. This outcome was achieved as a result of a suitable acquisition strategy. To expedite the acquisition process, the Python script first constantly read data from device port 502, and then the data were elaborated and saved in a database. Figure 5.3 depicts a schematic drawing of the system architecture.

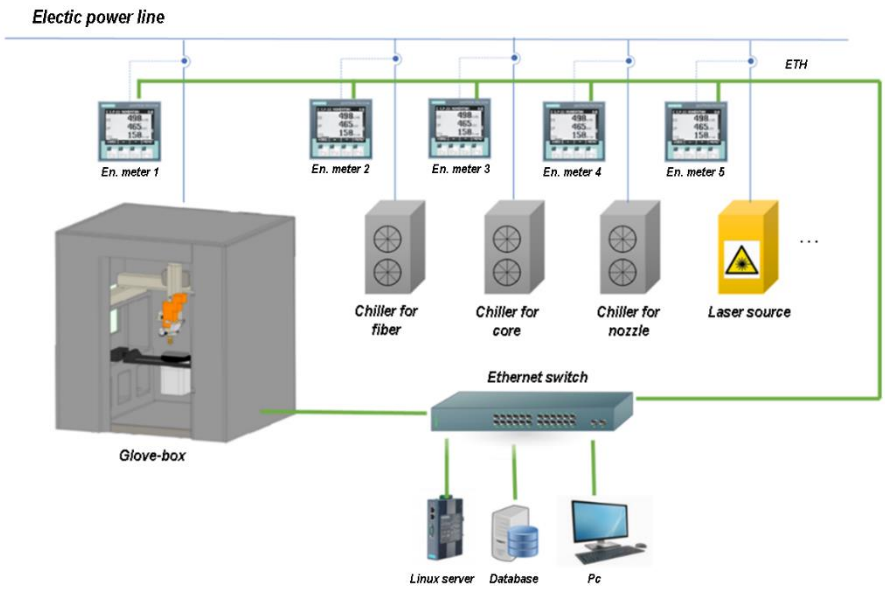


Figure 5.3 - Sensing units and monitoring system architecture.

In addition to the major sensors described above, the monitoring framework used numerous sensors integrated within each process unit. These were helpful for EA reasons since they allowed for the monitoring of secondary factors like as temperatures and water flows in chiller systems. Because these parameters exhibited smaller fluctuation regimes than the major parameters, they were treated as constants throughout the investigation. Although these have a little influence on the overall evaluation, they are critical for assessing the efficiency of each process unit. The largest measuring challenges were discovered in the real-time assessment of the powder flow during the deposition process. This parameter is critical for assessing the quality and sustainability of the laser deposition process since it has a significant influence on the deposited track. Real-time measurement of this parameter remains an open problem in the literature (Whiting et al., 2018) , in part because there are few commercial equipment capable of monitoring it. Due to the lack of a suitable sensor, the powder flow evaluation was performed through a preliminary set of flow experiments. These allowed for the calculation of the powder mass flow rate, which was kept constant throughout the deposition process. Table 5.3 summarises the key EA variables and the corresponding sensing units utilised to measure them.

Table 5.3 - Measured parameters of DLMD process.

Material/Energy	**Parameter**	**Variable**	**Equipment**
Metal powders	Mass flow [kg/s]	$\dot{m}^{in}_{Powder}$; $\dot{m}^{out}_{Powder}$	Mass flow test
	Temperature [K]	T^{in}_{Powder} ; T^{out}_{Powder}	Thermal sensor
Gases	Volume flow [l/s]	$\dot{m}^{in}_{Argon}$; $\dot{m}^{out}_{Argon}$	Flow meter
	Temperature [K]	T^{in}_{Argon} ; T^{out}_{Argon}	Thermal sensor
Cooling fluids	Volume flow [l/s]	$\dot{m}^{in}_{H2O}$; $\dot{m}^{out}_{H2O}$	Embedded Flow meter
	Temperature [K]	T^{in}_{H2O} ; T^{out}_{H2O}	Embedded Thermal sensor
Electricity	Electric power [W]	$\dot{W}_{el}$	Energy meter
Deposition track	Temperature [K]	T^{in}_{Powder} ; T^{in}_{Argon} $T^{out}_{Powder,loss}$; $T^{out}_{Powder,dep}$; T^{out}_{Argon}	Pyrometer
	Volume [mm^3]	V_{dep}	CCD camera

Temperature, pressure, and specific heat values for argon, AISI 316L metal powder, and water for chiller systems are listed in Table 5.4. The thermal equilibrium between the system and its environment is represented by the dead state.

Table 5.4 - Materials properties.

State functions	Symbol	Unit	Metal powder	Argon	Water
Dead state temperature	T_0	K	298.15	293.15	293.15
Dead state pressure	P_0	atm	0.987	4.935	0.987
Specific heat	c_p	J/kg K	500	520	4186

These data were collected in a variety of methods for each component of the system: directly, through the use of main monitoring devices such as pyrometers or monitoring systems embedded in the subunits, and indirectly, through literature, data sheets, and simulations.

5.1.5. Outcomes and discussion of the implemented EA

Regarding the graphs in Figure 5.4(a) and (b), the argon flow was deemed fully required for the creation of the deposition clad, and so, despite the fact that it is not a material component of the final component, its output quantity was not considered a loss. In contrast, according to the deposition efficiency calculation published by (Reddy et al., 2018) the powder was judged 60% usable for cladding and 40% wasted inside the glove-box and therefore considered lost mass for a coaxial nozzle.
Some assumptions have been made to simplify the process analysis: (a) the system runs under steady-state conditions, (b) pressure reductions due to all losses along the system are minimal, and (c) each processing unit is insulated, thus heat transfer to the environment is minimal. Figure 5.4 displays the outcomes of the mass balances of material flows.

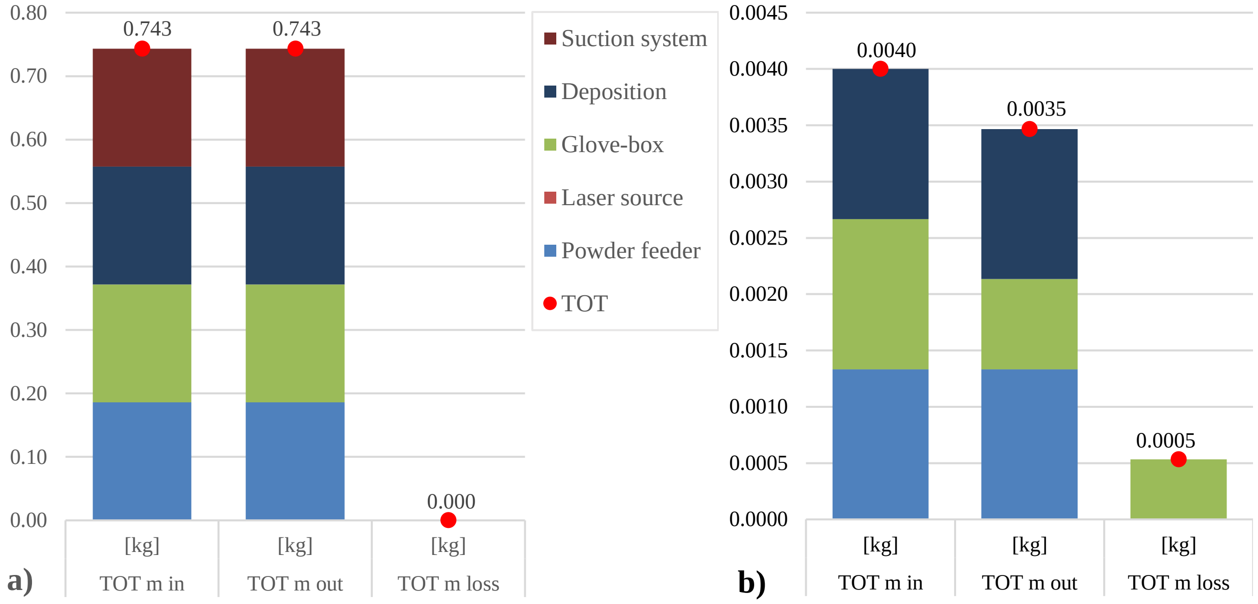


Figure 5.4 - Mass balances of (a) argon and (b) AISI 316L powder.

Taking as a reference the differentiation between exergy out (Ex_{out}) and exergy loss (Ex_{loss}), the amount of energy useful for the production of the final product ($En_{out,dep}$) and the amount of energy unnecessary for manufacturing the component and thus destinated to wasted materials and emissions ($En_{out,waste}$) have been defined. Figure 5.5 reveals that almost 70% of the entering energy (see Figure 5.5(a)) was converted to heat, which was wasted in the environment during the process and was worthless for the creation of the deposition track, and nearly 92% of the incoming exergy (see Figure 5.5(b)) was also lost. The input percent was determined by the electrical and material fluxes, which contribute to the balance via enthalpy, depending on both the specific heat and the temperature difference between the dead state and the entering temperature. The lost percentage represents all of the scattered energy that was not helpful throughout the deposition process. The three chiller systems appear to be the most energy-demanding devices, as seen by the graphs. In terms of exergies, the scenario looks to be slightly different, since the chiller for core appears to be the most active processing unit, followed by the laser source.

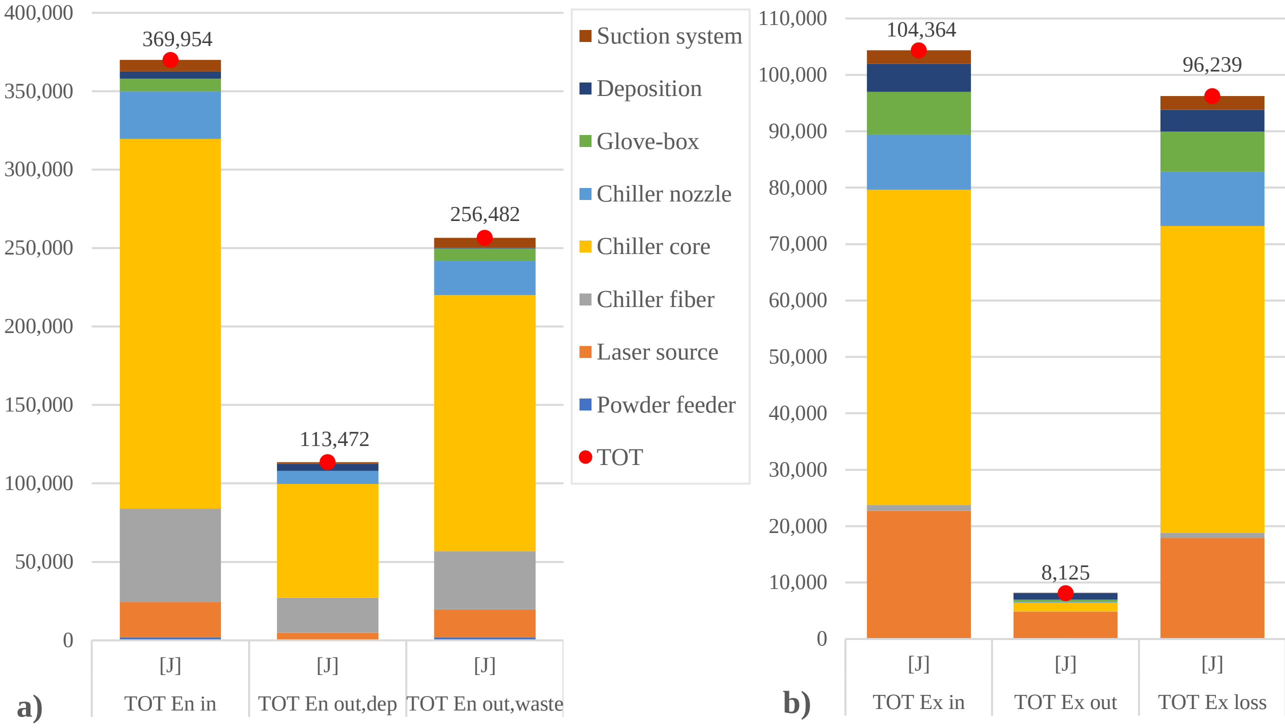


Figure 5.5 - Energy (a) and exergy (b) balances.

Figure 5.6 shows the exergetic efficiency of each processing unit; it also shows the overall efficiencies of the entire process. The laser source has an efficiency of about 21%, consistent with the literature on solid state laser sources. As could be expected, the efficiencies of the chillers turn out to be low, as these absorb more electrical power than the power needed to produce the useful work for the process. The devices are oversized, related to the process under consideration, as the system is multipurpose and must also cover the needs arising from different processes, such as laser welding.

Figure 5.6 depicts, on the right, the exergetic efficiency of the actual deposition process, which takes place within the glove-box but merits more investigation since the temperature differential of the materials may vary rapidly at this point. The value represents the energy (related to the laser beam) provided to the deposition point to create the track, but also wasted to heat up other elements involved in this phase of the process, such as the powder dispersed in the glove-box, the substrate, and the argon dropped out by the suction system. Temperatures were recorded using a pyrometer, however a reference was made to (Mazzarisi et al., 2020b) to estimate substrate overheating and quantify energy loss during deposition.

Exergetic efficiencies are more complicated than energy efficiencies because they take into account the useful work created during the process and how it relates to the maximal work of the Carnot cycle. Controlling energy yields results in greater and more consistent values: for example, the energy efficiency of chiller systems ranges from 25% to 40%.

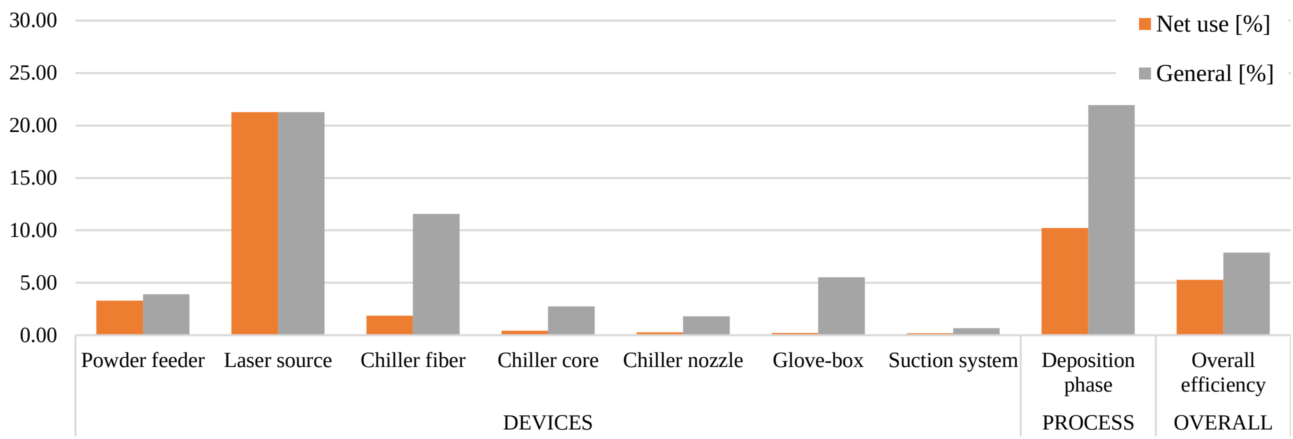


Figure 5.6 - Process units, deposition phase and overall process exergetic efficiencies.

Analysing the energy balance reveals that a greater proportion of the spent energy may be dedicated to the production of the laser beam. This finding is consistent with earlier studies (Kellens et al., 2017) , which identify the DLMD technique as the most energetic AM method currently available (considering an average energy demand of 7,779 MJ per every kg of deposited mass).

Another point to examine in light of the findings is the scarcity of bibliographical research on the environmental effect of DLMD systems that take into account all system components involved in the process. There is nearly no research on the influence of chillers on the efficiency of laser beam production and the entire process, for example. Exergetic modelling work has the benefit of analysing both the overall influence of the laser deposition process as well as the breakdown of the system into processing units in order to determine the energy and exergetic contribution for each of them. Furthermore, the efficiency of specific devices may be defined in order to plan for improvements on the most energy-intensive processing units and therefore prioritise the components that needed to be upgraded.

5.1.6. Conclusions

In conclusion, a suitable system for storing, collecting, and analysing a complete set of parameters was established. The new system framework, which was used for the first time in the field of additive manufacturing, proved to be successful in carrying out an accurate quality and sustainability evaluation of the DLMD process under consideration. The following were the major findings of the exergetic analysis:

1) The energy and exergy balance revealed that over 70% of the energy and 92% of the exergy injected into the system were lost during the process.
2) Chiller systems were the most energy-intensive sub-units of the system, requiring 88 percent of the total incoming energy and 64 percent of the incoming exergy.
3) The deposition process had a very low exergy efficiency (about 10% of net and 22% of general), while the DLMD system's efficiency fell to 5% of net and 8% of general efficiency.

The thermodynamic process modelling, as well as the analysis itself, enabled the identification of outstanding issues and criticalities linked to the monitoring and control system established in this study, with the goal of developing a further real-time monitoring and control method for the DLMD process: (a) The problem of synchronisation of sampling frequencies, as well as the possibility of integrating data from new devices into a single database, will be thoroughly investigated; (b) the analysis of complex deposition strategies, with a focus on the idle time between contiguous subphases throughout the process to detect variations in energy/exergy loss for each sub-unit, will be thoroughly investigated; (c) the analysis could even focus on environmental sustainability through the implementation of a full EA-LCA model; (d) in order to be well suited for the proposed holistic multi-sensor approach, the monitoring system architecture should combine the capabilities of all sensing units while also being effective in terms of flexibility and adaptability.

Given the advantages of combining EA and LCA for this type of process assessment, the thermodynamic model to be applied may be far more effective if connected with the LCA: At first, this can give more objective-oriented evaluation findings; later, this can become a useful tool for decision-making policies targeted at generating evolutionary solutions, allowing the process to automatically avert any potential failure. Furthermore, with a suitable sensing and monitoring system, the entire EA-LCA model may enable any practitioner to rebuild both process hardware and software to smart I4.0 standards virtually in real-time, boosting cost efficiency.

5.2. HISTORICAL HERITAGE: RETROFIT INTERVENTIONS AND LCA

In the EU the construction industry is responsible for about 42% of the final energy consumption and 50% of raw materials, produces approximately 35% of greenhouse gas emissions and 50% of waste (European Commission, 2007; Hauschild et al., 2018).
In this regard, the European Union "is committed to developing a sustainable energy system that is competitive, secure, and decarbonized by 2050," according to Directive n. 2010/31/EU on building energy performance (EPBD recast). The construction sector is crucial because it has the greatest influence on the environment in terms of resource usage, energy consumption, emissions, and waste. Also in the latest review of the Directive (July 2018) (Europan Parliament and European Council, 2018) it is clear that each UE member is tasked with developing a long-term strategy to support the renovation of residential and non-residential buildings, both public and private, in order to achieve a decarbonized and energy efficient real estate stock by 2050, facilitating the cost-effective transformation of existing buildings into nearly zero energy buildings, which translates into the need to implement solutions that reduce energy needs and wastes, particularly in the built environment. (Sartori and Hestnes, 2007).
The AiCARR (Italian organisation for air conditioning, heating, and refrigeration), which released a handbook in February 2014, was the first to emphasise the relevance of energy diagnostics in ancient structures. In compliance with new laws, for the evaluation and improvement of the energy performance of historic structures (de Santoli, 2015).
Following the broad concentration of energy efficiency, the market's and operators' focus shifted to environmental sustainability. The desire for accurate, easy-to-use indicators for building environmental evaluation has resulted in the creation of numerous tools with quite varied approaches in recent years. The first, voluntary method, in particular, resulted in the development of multi-criteria rating systems (Green Building Rating Systems) that assign a score to each criterion based on its performance. The ratings are based on a low level of sustainability, with the final evaluation based on the concept that worse environmental performance in one category can be compensated by higher environmental performance in another.
However, the regulatory path is based on a Life Cycle Thinking approach, i.e., the quantification of synthetic environmental indicators using the Life Cycle Assessment method, which is internationally recognised as a method to evaluate the environmental profile of products, encoded in international regulations, and promoted by European environmental policies. This method emphasises strong sustainability in order to demonstrate the elimination of all environmental effect at each step of the building's and its components' life cycle (Pombo et al., 2019). In this vein, the scoring systems are gradually including environmental factors into the LCA.
The aim of this work was to develop and complete previous analyses on sustainability led on historical heritage, highlighting all the issues that are still unresolved and the possible solutions to be undertaken. The historical building case of study named “Palazzo del Sedile” is located in the old town of Matera, Basilicata, Italy.
This research was funded by FESTA Project from EU Horizon 2020 research and innovation programme: Fostering Local energy investments in the Province of Matera, which promotes local energy investments for public buildings and disseminates the PPP (Public Private Partnerships) approach through innovative Energy performance contracts in the regions. The main partners were and still are the Province of Matera, which had the role of leader and coordinator of the project’s activities, University of Basilicata - DICEM which had the

role of technical partner for the scientific definition of ways and means, also innovative, that have been developed during the project and Local health care company of Matera (ASL Matera), owner of the hospital of Policoro.
Deep understanding of the energetic-environmental skills of the historical heritage of the Mediterranean area of southern Italy is critical in order to preserve their uniqueness and the original reason for being constructed. While managing to enhance thermo-hygrometric wellness and environmental quality in the context of environmental sustainability and the design of the best intervention methods. In literature, few are the data related to the sustainability or to the energy performance of the buildings that belong to the historical patrimony of the city of Matera (Gizzi et al., 2016).
In fact, there are only studies related to the evolution history of the City (Rota, 2011) or related to the structural preservation of the "Sassi" (Cardinale et al., 2013), or simply manuals (Restucci, 1998).

5.2.1. LCA regulations in construction field

The strategic importance of adopting the LCA methodology as a basic and scientifically suitable tool for identifying significant environmental aspects is clearly expressed within the COM 2001/68/EC Green Paper and the COM 2003/302/EC on Integrated Product Policy, and is suggested, at least indirectly, also within the European EMAS (1221/2009) and Ecolabel (66/2010) regulations.
For instance, a promising possibility is the integration of building systems LCA data per functional unit in the Building Information Modelling (BIM) platform (Bueno and Fabricio, 2018).
The ISO series represents the majority of national and international standard advice for LCA studies:
1) UNI EN ISO 14040 (2006) "Environmental Management - Life cycle assessment - Principles and framework";
2) UNI EN ISO 14044 (now updated to 2018) "Environmental Management - Life cycle assessment - Requirements and guidelines";
3) ISO 21931-1 (2010) "Sustainability in building construction-Framework for methods of assessment of the environmental performance of construction works Buildings";
4) UNI EN 15978 (2011) "Sustainability of construction works-Assessment of environmental performance of buildings - Calculation method" (Thibodeau et al., 2019).

5.2.2. Case study: methodology and resources

The LCA on the case study of this work has been executed using the software SimaPro® by PRé Consultants v.8.5.2.0 on energetic retrofit interventions assumed for an historical building in Matera named "Palazzo del Sedile".
The data on energetic performances, particularly throughout the assembly and usage phases of the life cycle, were obtained from previous realistic and accurate experiments conducted on the same building.
The information associated with the end-of-life phase was obtained from actual information and other related studies in the literature (Eberhardt et al., 2019; Goldstein and Rasmussen, 2017; Moslehi and Reddy, 2019; Schiavoni et al., 2017).
Palazzo del Sedile (Figure 5.7) is an historical building located in the old town of Matera, a southern city of Italy. It was built in 1540 and now is owned by the Province of Matera. It was constructed around 1540 and is presently held by the Province of Matera. In 1944, it underwent a transformation that transformed it into a focal point for the city's musical heart, becoming the primary venue of the conservatory dedicated to the composer "Egidio Romualdo Duni." Since the early 1980s, the building's basement floors have housed a contemporary theatre with a seating capacity of around 450 seats. As a result, Piazza del Sedile may be called the living room of Matera's old town, lively and active all year. It is now a site of identity, gathering, and sociability. The current construction dates from 1779, when the mezzanine level was built from scratch and the subterranean floor was restored. The mezzanine floor is shaped like a pentagon. The rooms on the ground

floor are linked to those on the first storey by an atrium. The mezzanine floor has seven rooms, all of which are administration services and toilets.


Figure 5.7 - Entrance façade of Palazzo del Sedile in Sedile's Square and a view from above

The plant covers around 345 m^2; the heights ranging from 2.40 m up to 5.05 m (where there are vaults). A similar scenario exists on the first level, which consists of thirteen rooms separated into classrooms and toilets. The building has a total of ninety persons, including students and employees. Its total area is about 590 m^2 and the heights ranging from 2.60 m up to 6.70 m at the top of the vaults. The rooms with flat floors have a ceiling height of 3 m.
The mezzanine floor facing west has different characteristics from the central body of the building, as it shown in Figure 5.7 and in the 3D model in Figure 5.8: the two portions were built at various times and in different styles, thus their energy performance is fairly different.
The building's bearing structure is made up of thick masonry septa (net of plaster) with varied widths ranging from 80 cm to 100 cm for the external walls and 40 cm to 60 cm for the interior ones. The interior floors are made of limestone blocks, as is the roof covering. Waterproofing is done using tar and bricks on walkable floors and tiles on pitched roofs.
The general parameters reported in earlier research and important in implementing the LCA are presented in Table 5.5.

Table 5.5 - General parameters of Palazzo del Sedile

Floors [n.]	**Total height** [m]	**Gross volume** [m^3]
3	13	5,550
Total area [m^2]	**Floor area** [m^2]	**Opaque walls** [m^2]
1,500	595	1,144
Transparent walls [m^2]	**Dispersant surface** [m^2]	**A/V** [m^2/m^3]
67	2,217	0.17

Figure 5.8 - DesignBuilder model of Palazzo del Sedile

The energetic retrofit on the building case of study consists in interventions on the envelope, air conditioning and heating systems and lighting, which complies with the minimum requirements imposed by Ministerial Decree June 26, 2015 (Corrado et al., 2016).
Table 5.6 depicts the hypothesised retrofit actions on the building as well as the new energy consumption reductions. With the combination of these interventions, energy savings in kWh/year are anticipated to be over 70% when compared to current use. This estimate is extremely realistic because the research was conducted in situ using data collecting and inquiry, a modelling and dynamic simulation mode provided by Designbuilder software (see Figure 5.8) and EnergyPlus™.
Dynamic simulation is a very advanced calculating approach that allows to analyse the energy performance of a structure while accounting for the inertial impacts of the casing and plants. and has a higher level of accreditation in energy certification (Selicati et al., 2019).
The process of defining the limits of the system object of research is dependent on the objective of the study itself: the same system examined with various bounds yields different findings.
The technique utilised for the case study is "from gate to cradle," also known as the "downstream module," which is the module that contains the product scenarios from the time it leaves the manufacturing company's gate until it concludes its "life" in transportation, usage, and disposal. Only the retrofit operations were subjected to LCA. According to EN 15978 specifications (European Committee for Standardization, 2011) in Figure 5.9, the stages A2 to D: the whole life cycle was evaluated, not only the raw materials stock and their assembly. The life cycle begins when the window or insulating panel is delivered to the construction site.

<table>
<tr><th colspan="3" rowspan="2">Product / Manufacture Stage [A1-A3]</th><th colspan="2" rowspan="2">Construction Process Stage [A4-A5]</th><th colspan="7">Use [B1-B7]</th><th colspan="4" rowspan="2">End-of-Life Stage [C1-C4]</th><th rowspan="2">Benefits & Loads Beyond [D]</th></tr>
<tr><th colspan="5">Building Fabric</th><th colspan="2">Operation of the Building</th></tr>
<tr><td>Raw Material Extract / Process / Supply</td><td>Transport</td><td>Manufacture</td><td>Transport to the Site</td><td>Assembly / Install in the building</td><td>Use / Application of Installed Products</td><td>Maintenance</td><td>Repair</td><td>Replacement</td><td>Refurbishment</td><td>Operational Energy Use</td><td>Operational Water Use</td><td>Deconstruction / Demolition</td><td>Transport to Waste Process</td><td>Reuse-Recovery-Recycle</td><td>Disposal</td><td>Reuse-Recovery-Recycle Potential</td></tr>
<tr><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>B1</td><td>B2</td><td>B3</td><td>B4</td><td>B5</td><td>B6</td><td>B7</td><td>C1</td><td>C2</td><td>C3</td><td>C4</td><td>D</td></tr>
<tr><td colspan="3">Cradle-to-Gate</td><td colspan="13">Gate-to-Grave</td><td rowspan="2"></td></tr>
<tr><td colspan="16">Cradle-to-Grave</td></tr>
<tr><td colspan="17">Cradle-to-Cradle</td></tr>
<tr><td colspan="17">System Boundaries</td></tr>
</table>

Figure 5.9 - Stages of the life cycle of a building in accordance with EN 15978 (Kylili et al., 2016).

As functional unit it was decided to use the m^2 floor area.

This step is defined in SimaPro under the section "Goal and scope," which is visible in the left portion of the software's GUI, and includes the following: a description of the LCA analysis that defines the author, developer, study objective, functional unit, and the library (database) with its own specific field of application. Ecoinvent3, ELCD, and Industry Data 2.0 were the datasets used for the case study (Goedkoop et al., 2016).
In terms of inventory analysis, the system in the case study is split into four subsystems, which correspond to opaque walls, floors, windows, and the heating system. The lighting system is solely assessed in terms of energy consumption during the use phase (see Table 5.6).

Table 5.6 - Retrofit interventions and energy consumption savings

Element	Before interventions	After intervention	Savings [%]
Opaque Walls	Limestone wall of 45 - 90 cm and inner plaster of 2 cm U = 0.63-1.12 W/m^{2}K	Covering wall insulation with inner coat of 5 cm Kenaf plate (λ = 0.038 W/mK). U = 0.31 W/m^{2}K	4
Roof slab	Limestone roof of 45 cm externally covered with tiles of 1 cm and internally covered with plaster of 2 cm U = 1.042 W/m^{2}K	Covering roof insulation with inner coat of 9cm Kenaf plate (λ = 0.038 W/mK). U = 0.25 W/m^{2}K	13
Windows	Wooden windows with double glass 4/6/4 or 6/12/4 U = 3.15-1.78 W/m^{2}K	PVC windows with low emissive double glass 4/8/4 with 8mm of argon interspace. U = 1.71 W/m^{2}K	20
Heating system	Boiler on the rooftop η_{nd} = 0.81	Stainless steel compression heat pump COP = 3.08 and P = 20 kW and installation of thermostatic valves	58
Lighting	Metal Iodide spotlights, Neon 1x36 W; Neon 2x36 W; Incandescent lamps	LED technology lamps	91

It has been estimated that the envelope will last 35 years, while the heating system and lights would last 15 years. This decision was taken due to the fact that the proposed retrofit interventions on Palazzo del Sedile are non-invasive and tend to just improve the energy component of this ancient structure that is obsolete.
Without thoroughly researching any possible conflict between old and new materials, it was chosen to shorten the lifespan before the necessity for additional treatments (Vilches et al., 2017).
In SimaPro, there is a process and product stage for each material or constituent that makes up the subsystem. These steps are contained in the SimaPro GUI feature Inventory. The procedures that are already in place in software are classified into categories and subcategories based on the specific area to which they pertain.
During the case study's use phase, the energy consumptions associated with the lifetime of each of the previously stated subsystems, including lighting, were taken into account. In terms of disposal, it has been proposed that the windows and heat pump be disassembled. The PVC material will be disposed of, the glass will be recycled, and the stainless-steel pump will be melted and repurposed. All other waste by-products are disposed of at a landfill. Clearly, energy consumptions for disposal and transportation of the item to be disposed of were also taken into account.
The impact assessment phase of the life cycle attempts to determine the number of possible consequences on human health and the environment based on the LCI results. The inventory data, in particular, is linked to specific categories of environmental effects and category indicators. Furthermore, LCIA offers information for the following step of interpretation, which seeks to provide meaningful recommendations in connection to the study's goals and objectives.

5.2.3. LCA Results and interpretations

The interpretation step may result in an iterative process of reviewing and revising the field of application of LCA, emphasising the limitations and possibilities of the LCA approach used in this situation. Figure 5.10 displays the tree that expresses emissions in terms of energy consumption (red lines) or energy savings (green lines) for each phase and sub-phase, with matching line thicknesses that qualitatively indicate the incidence of consumption or savings in relation to the entirety of the intervention: the LCA evaluation of the case study in all three methods of evaluation highlighted how the phase of use is the most significant, as it is considered a duration in years definitely longer than assembly and subsequent disposal phases, as expected from the behaviour of a historical building like Palazzo del Sedile. It follows the disposal phase because a greater proportion of the retired materials are disposed of in landfills; it concludes the list with the transport and assembly of the elements because the interventions to be carried out are, as previously stated, non-invasive and solely aimed at energy efficiency.

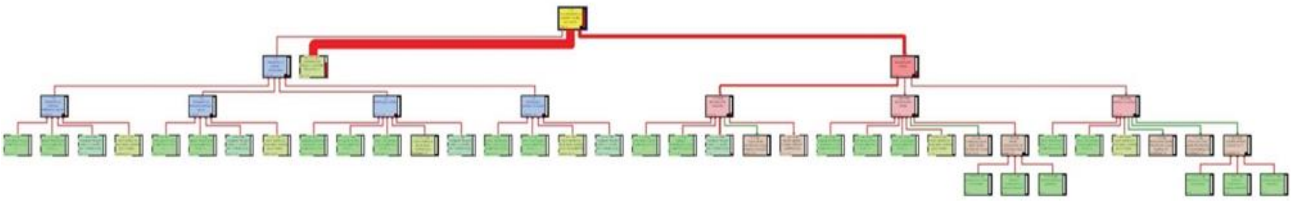


Figure 5.10 - Tree representation of the entire LCA with Eco-indicator 99 method - cut off 0.1%

The assessment methods implemented in SimaPro for the case study are (PRé, various authors, 2019):
In terms of primary energy, the assembly phase has an energy consumption of 7.15%of the total, the usage phase has an energy consumption of 75.65%, and the disposal phase has an energy consumption of 17.20% of the total. The normalised characterisation for impact categories based on the Italian energy mix supplied by Italian GSE for the year 2017 is displayed in Table 5.7, and the normalisation for impact categories is illustrated and the normalization for impact categories is shown in Figure 5.11.

Table 5.7 - Impact categories of CED method

Category	Unit	Total	Assembly	Use	Disposal
Non-renewable, fossil	MJ	7.73E+7	5.55E+6	5.85E+7	1.33E+7
Non-renewable, nuclear	MJ	5.07E+6	3.64E+5	3.84E+6	8.72E+5
Non-renewable, biomass	MJ	1.03E+6	7.42E+4	7.82E+5	1.78E+5
Renewable, biomass	MJ	3.97E+6	2.85E+5	3.00E+6	6.83E+5
Renewable, wind, solar	MJ	1.74E+7	1.25E+6	1.31E+7	2.99E+6
Renewable, water	MJ	3.31E+7	2.38E+6	2.50E+7	5.69E+6

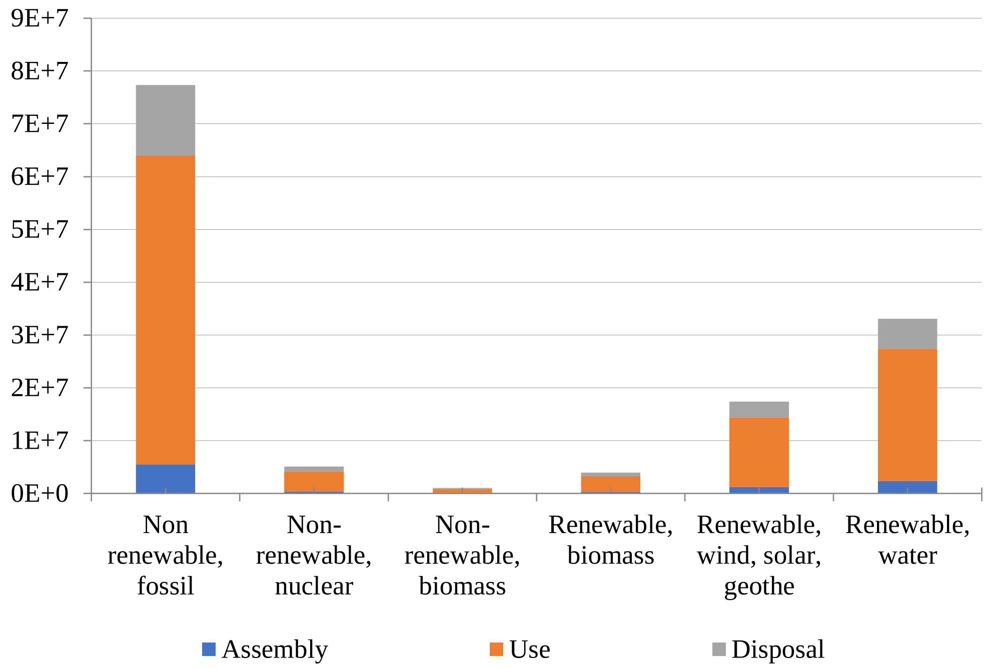


Figure 5.11 - Normalization of CED method

Eco-indicator 99: the assembly phase has an energy consumption of 2.83% of the total, the usage phase has an energy consumption of 70.17%, and the disposal phase has an energy consumption of 27% of the total. These results differ significantly from those obtained using the CED method. Table 5.8 a depicts the characterisation of impact categories, whereas Figure 5.12 depicts the normalisation of impact categories.

Table 5.8 - Impact categories of Eco-indicator 99 method

Category	**Unit**	**Total**	**Assembly**	**Use**	**Disposal**
Carcinogens	DALY	1.64	0.14	1.3	0.21
Resp. organics	DALY	3.19E-3	5.78E-5	1.75E-3	1.38E-3
Resp. inorganics	DALY	4.03	0.1	2.79	1.14
Climate change	DALY	1.34	0.01	1.03	0.3
Radiation	DALY	0.04	0.02	0.02	1.4E-3
Ozone layer	DALY	7.31E-4	3.83E-6	5.54E-4	1.73E-4
Ecotoxicity	PAF*m2yr	5.53E+6	3.18E+5	3.87E+6	1.34E+6
Acidification/Eutrophication	PDF*m2yr	1.13E+5	1.92E+3	8.76E+4	2.3E+4
Land use	PDF*m2yr	1.3E+5	1.62E+3	7.4E+4	5.43E+4
Minerals	MJ surplus	2.09E+5	4.52E+4	8.56E+4	7.86E+4
Fossil fuels	MJ surplus	8.47E+6	5.3E+4	5.72E+6	2.7E+6

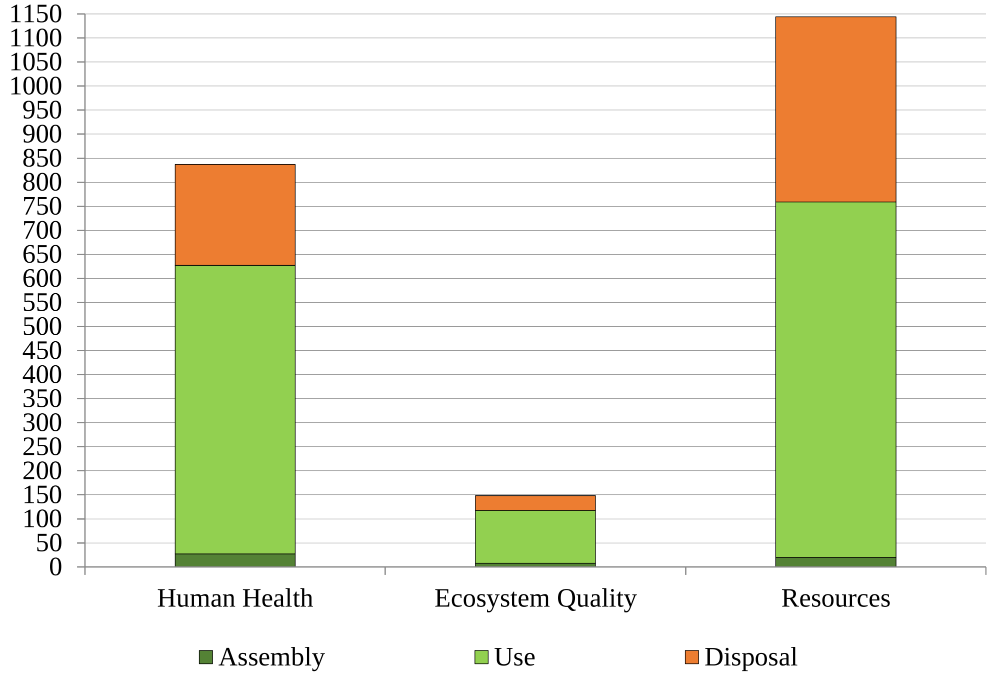


Figure 5.12 - Normalization of Eco-indicator 99 method

EDIP2003: in terms of primary energy, the assembly phase has an energy consumption of 2.32E+3 Pt (9.75% of the total), the usage phase has an energy consumption of 1.7E+4 Pt (71.43% of the total), and the disposal phase has an energy consumption of 4.47E+3 Pt (9.75% of the total) (18.78% of the total). Even though the effect categories are somewhat different, the findings are fairly comparable to the CED approach. The characterization for impact categories is shown in Table 5.9. In the table are considered only the relevant values for the interpretation of the results. The normalization for the relevant impact categories is shown in Figure 5.13.

Table 5.9 - Impact categories of EDIP2003 method

Category	**Unit**	**Total**	**Assembly**	**Use**	**Disposal**
Global warming 100a	kgCO2eq	6.44E+6	6.07E+4	4.94E+6	1.44E+6
Ozone depletion	kgCFC11eq	0.76	3.69E-3	0.59	0.17
Ozone formation	m2.ppm.h	3.11E+7	5.16E+5	2.19E+7	8.67E+6
Acidification	m2	4.87E+5	9.49E+3	4E+5	7.67E+4
Terrestrial eutrophication	m2	4.66E+5	7.24E+3	3.66E+5	9.33E+4
Ecotoxicity soil chronic	m3	1.94E+8	1.47E+8	3.44E+8	1.45E+8

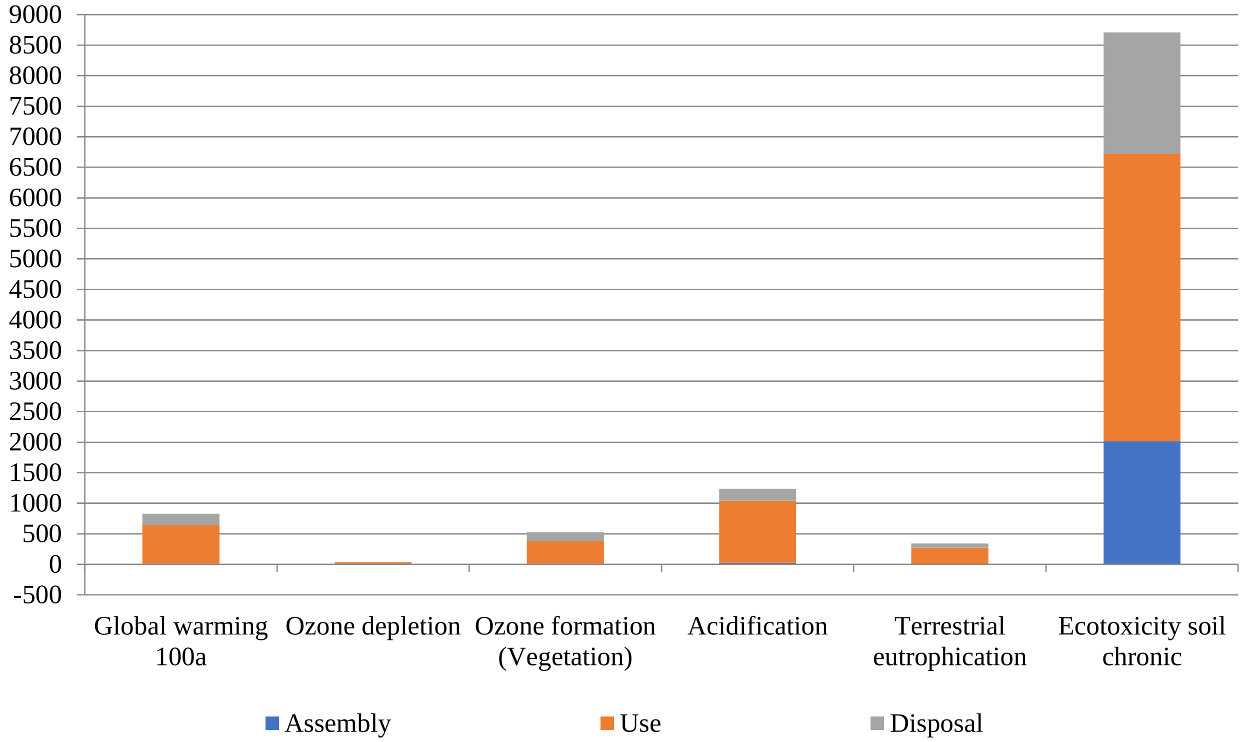


Figure 5.13 - Normalization of EDIP2003 method

These three approaches are highly diverse, and a true numerical or statistical comparison is extremely difficult. In general, when evaluating the impact assessment of a specific product and its corresponding components, the evaluation of each impact category and the calculation of the final impact score points is given by a common mathematical equation, depending on each methodology and the way the study is conducted (basic level, or advanced level using software tools) (de Gracia et al., 2010).

It is also true that the significant difference in the methodologies employed is due to the fact that two approaches to characterisation can take place along the effect route of an impact indicator within the LCIA step: midway approach and endpoint approach. Characterization at the midpoint level models the impact using an indicator located somewhere along the methodology mechanism but before the endpoint categories, whereas characterization at the endpoint level necessitates modelling until the endpoint categories described by the areas of protection (in most methodologies, the main areas of protection are eco system quality, human health and resources). EDIP 2003 is a midpoint-oriented technique, Eco-indicator 99 is an endpoint-oriented method, and CED is classified as "other based LCA methodology" since it is only concerned with energetic resource use. Table 5.10 compares the contributions of each sub-phase for each method under consideration.

Table 5.10 - Comparison between the outcomes of the three implemented methods

	CED [%]	**Eco-Indicator 99** [%]	EDIP 2003 [%]
Assembly	7.15	2.83	9.75
Use	75.65	70.10	71.43
Disposal	17.20	27.00	18.78

Looking at the percentages of energy impact allocated in the three major phases of the life cycle, there is a further demonstration of how much each method has a weighing system characterised by a quantitative factor, but also by a subjective factor: they give space to some aspects rather than others, which will affect the weighing system.

By comparing the percentage allocations of environmental effect for the three main stages, it is easy to see that there is a common framework in the calculations; otherwise, the usage phase would not have been predicted to be the most expensive. On the other side, there is no uniformity in the importance of affects being assigned. The variance is not enormous, but given that the percentages upstream of those are frequently quite large values of primary energy, a little percentage change equates to a significant energetic variation.
The link between energy savings and economic initiatives including retrofit measures is an intriguing perspective worth emphasizing. Table 5.11, which depicts the system view, contains the normalised findings. This table serves as the most essential guideline for the interpretation phase of this work: it compares environmental sustainability to another form of sustainability that serves as the true standard for decision-making, particularly if it pertains to public administrations.

Table 5.11 - System view

	Energetic savings [%]	Economic efforts [%]	η [%]
Opaque envelope	12.29	48.27	3.23
Windows	14.81	21.21	8.87
Heating system	43.18	24.78	22.13
Lighting system	29.72	5.74	65.77

The rate, which quantitatively depicts the system view, indicates the percentage ratio between the economic effort required to carry out the above-mentioned interventions and the energy savings that would result from such interventions. The table demonstrates that the lighting system is the most helpful profit of all the interventions because it provides a good compromise between energy savings and the anticipated economic expenditure required to achieve it. On the contrary, it is evident that the anticipated expenditure to refit the opaque envelope is exorbitant in comparison to the potential energy savings. This system view demonstrates that even small measures studied at the design stage (for example, the type of insulation of a HAVC system, or the conscious choice of the type of lighting system) can greatly influence the efficiency of an entire retrofit intervention, both practically, energetically, and economically. It is consequently critical to retain an overall perspective of the system and its reaction in the short and long term on the life cycle timeline.

5.2.4. Conclusions

To conclude, the environmental assessment tools (regulations, databases and software), in addition to the dynamic simulations of the building itself, are an excellent companion for achieving a sustainable result at 360 degrees. LCA has been proven to be a reliable tool: it allows for the identification of environmental criticalities in the selection of materials and building elements, as well as the calculation of built-in energy and its various contributions to global impact; this allows for the comparison of different components, evaluating the true environmental benefits of alternative design solutions. It was easy to highlight how the processes of material derivation recycling and reuse allow for significant environmental advantages at the disposal phase.
However, when comparing additional techniques, it is observed that subjectivity plays a major part in the findings, resulting in variability of the assessment, which lowers the comparability of the LCA results and, as a consequence, a difficult univocal interpretation of the assessment. The limitations of LCA with SimaPro include the prototypical nature of the building sector, an increasing complexity of the process and its phases, as highlighted by interactions between the building and external factors, including the environment; as well as the number of mini sub processes involved in a building's life cycle and the difficulties in retrieving data compatible with reality (Selicati et al., 2020). It is important to emphasise once more how the availability of an accessible and up-to-date database of materials and processes relevant to the Italian setting would improve the credibility and importance of the results produced.

Furthermore, a growing number of case studies pertaining to historical structures in Italy would allow for improved results comparability and the development of environmental sustainability benchmarks at the national level.
Finally, the reader should understand that the system perspective is a crucial guideline to be utilised throughout each decision-making step since it provides environmental sustainability with an extra interpretation key that is simpler to comprehend by any stakeholder.

5.3. INDOOR AIR QUALITY MONITORING AS A PILOT PROJECT FOR HEALTHY SUSTAINABLE HOMES

Air pollution is commonly associated with car emissions or industrial plants, but the air inside our homes is often far more polluted than the outdoor air. Given that we spend 90% of our time indoors, and 65% of that spent at home, a sick house can have detrimental effects on occupant health. Numerous studies have found strong links between housing and health. Rudolf Virchow, a famous researcher, warned in the mid-nineteenth century that indoor pollution was associated with the highest rates of transmission of virus diseases, such as the cholera epidemic that occurred in the 1930s in the United States.
After nearly a century, the problem resurfaces with the COVID-19 pandemic, which has exposed the flaws and critical issues of Italy's real estate heritage, which is characterised by old, energy-intensive, small buildings with high management costs. A property has a useful life that rarely exceeds 50 years and ages in the same way that a human being does (Passaro, 2019). As a result, it exhibits age-related symptoms. A property, a house, suffers from pathologies of various kinds over time: construction technology becomes obsolete, building materials wear out, systems lose efficiency, and air quality in the indoor environment changes due to the concentration of dangerous agents released by the building itself, the exhibition, the furniture, and the daily practises of its tenants. Getting sick in the building, for the population exposed to such environments, the risk of contracting chronic diseases increases in the long run, as people spend the majority of their time indoors throughout the day (indoors) (General (US), 2009). Numerous studies have found that the concentration of indoor pollutants is frequently higher than the optimal regulatory limits (Istituto Superiore di Sanità, 2014). Sick Building Syndrome (SBS) (Passarelli, 2009) and Building Related Illness (BRI) (Seltzer and Diego, 1994) are the most well-known indoor pollution-related diseases. Indoor pollution is frequently more prevalent than outdoor pollution. Indoor air pollution levels are 5 to 10 times higher than outdoor air pollution levels (European Environment Agency, 2013). Indoor air quality is one of the leading causes of illness, both directly and indirectly affecting the health sector. Indoor Air Quality (IAQ) (US EPA, 2014) It is determined by contamination with outside air, but more importantly by the presence of internal sources of emission and diffusion of chemical and biological contaminants. It should be noted that the health risks associated with indoor air pollution are also dependent on personal exposure and the sensitivity of each individual due to microclimatic parameters, completely random factors such as high levels of stress, and other specific discomforts that can also vary with seasonality.
According to more recent research, chronic exposure to a suboptimal indoor microclimate promotes virus spread (European Environment Agency, 2020). NO_2, PM2.5 and/or ozone concentrations in the air have been linked to an increase in the number of COVID-19 cases, the number of severe COVID-19 infections, and the risk of death from COVID-19 in China (Zheng et al., 2020), United States (Cole et al., 2020) and Europe (Travaglio et al., 2020). PMs, in particular, can act as a physical vector for the virus (Morawska and Milton, 2020), s particularly when indoors or indoors, and especially when crowded and with insufficient ventilation.
The growing emphasis on sustainability and environmental performance does not address this issue, which remains an inherent risk of living. In this context, the work presented by the World Health Organization (WHO) (World Health Organization, 2010), is extremely useful, as it has developed guidelines for indoor air quality for some specific air pollutants present in confined environments for the first time at the European level.

Indoor and outdoor air pollution was identified as a major risk factor for noncommunicable diseases in 2018, highlighting a wide range of side effects ranging from sensory distress, irritation, headache, and asthenia to serious health damage, including chronic diseases and carcinogenic effects (United Nations, 2018). Despite recent campaigns to reduce outdoor and indoor air pollution, the latter continues to have a significant impact on human health, lowering quality of life and life expectancy. It also has a significant socioeconomic impact because it increases health costs for medical visits, hospitalizations, and medication use, as well as indirect costs because psycho-physical and thermo-hygrometric malaise reduces worker productivity at work.
However, as is well known, Italy lacks legislation governing the issuance of a certificate or certification of "healthy house". Unrelated studies are being conducted by public institutions, such as the ISS in collaboration with “Gruppo di Studio Nazionale Inquinamento Indoor” (“Istituto Superiore di Sanità: Benvenuti,” n.d.), ISPRA (Istituto Superiore di Sanità, 2014),) and private projects that have made technical contributions and emergency prevention or reduction strategies but are still sectarianized in well-defined applications.
An investigation of the sector critical issues reveals that:
• The construction sector, characterised by small realities, has not favoured the use of eco-sustainable materials and technologies;
• Current state incentives for redevelopment interventions are oriented toward energy savings (Ecobonus), with the result of worsening indoor air quality (Air Quality Index);
• Many buildings in our area lack a certificate of viability, and those issued, dating back decades, do not take into account the current environment.
Politically, the problem of healthy housing has not been addressed, despite the fact that an increase in chronic diseases is reflected in an overload of the national health system, both in terms of higher costs for medicines and increased use of hospital facilities. On the other hand, as a result of today's Pandemic, people's perception of the dangers of indoor pollution in unhealthy buildings has grown.
As a result, it is possible to see the opportunity, as well as the need, to grasp systemic aspects within the building technologies recurring in the last century, to allow an integrated vision of a process that can lead to obvious improvements in terms of optimisation of the real estate market performance, allowing to grasp the critical points of habitability in relation to the healthiness of confined environments and human health, all in a sustainable way.

5.3.1. Socio-economic aspects

Regarding the purely socio-economic aspect, while significant progress has been made in recent decades to make confined environments healthy and safe places, the problem of socio-economic repercussions on both the individual and the entire community remains unsolved. It is difficult to determine the exact costs associated with this issue, let alone the potential benefits of good management and prevention. Information and forecasting of future effects expressed in monetary terms contribute significantly to policy decision-making (Camoletto, 2017).
Numerous socioeconomic studies have been conducted to date that link SBS with the health sector and the BRI, as well as the economic impact of this link. It should be noted that the majority of the research has focused on the effects of air pollution (in general) on health.
The cost of air pollution could be described as "social costs." In economic terms, social costs are the private costs borne by those directly involved, which are added to the external costs borne by third parties who are not directly involved (Health-Sanity) (CE Delft, 2020).
It is responsible for 2.7% of global disease burden (World Health Organization, 2009), with 30% attributable to childhood deaths from acute respiratory infections.
In 2011, the WHO, in collaboration with the OECD, conducted the first study on the economic cost of atmospheric pollution's effects on human health at the European level (World Health Organization, 2015). The WHO has calculated the costs of premature deaths caused by prolonged exposure to atmospheric pollution. In 2010, 600,000 premature deaths were recorded as a result of diseases caused by outdoor and indoor air

pollution, amounting to approximately $1,600 billion (on average one tenth of the total European GDP). In this case, the cost in Italy is estimated to be around 4.7% of national GDP. The economic cost of a mortality impact is calculated by multiplying the estimated atmospheric VSL (value of statistical life) by the number of premature deaths. As a result, the economic benefit of a mitigator or prevention action in this context would be calculated using the same VSL value but multiplied by the number of premature deaths avoided.
The European Commission conducted a study on the economic impact of indoor air quality (Theakston and Weltgesundheitsorganisation, 2011) , evaluating the main pathologies attributable to indoor pollutants such as asthma, pulmonary carcinoma, chronic obstructive bronchitis, infections/symptoms respirators, acute poisoning, and quantifying the impact attributable to indoor pollution in approximately 1.6 million DALY (Disability-Adjusted Life Year).
In Italy, the Indoor Commission conducted a preliminary calculation of the annual direct costs attributable to health damage caused by indoor pollution in 2001, which came to around 234 million euros per year. This report only addresses a small number of pollutants that had a more visible and serious impact on health: Allergens, benzene, carbon monoxide, radon, and passive smoke are all potential hazards (Ministero della Salute, 2014).
To summarise, in addition to workplace prevention and safety, even living prevention should not be viewed as a cost to be sustained. Despite the fact that several methods and tools have been developed in recent years to establish health costs related to Indoor Air Quality, there is still no real estimated tool available due to the complexity and multiplicity of factors to be considered, particularly locally. As a result, a focus on the balance sheet for prevention is required through the development of models for estimating costs with cost items at the individual and single pollutant level, at the level of real estate, immobile, and society as a whole. We must prioritise prevention, property investment, and maintaining optimal indoor quality. The Ministry of Health has already taken action in this regard (Ministero della Salute, 2020).

5.3.2. Goals

The first goal is to be able to respond to the question, "How safe are the places where we live?" in a timely fashion. The search for the correct characterization of specific indoor / outdoor pollutants that influence the health of a property, of any kind, is a critical component for achieving a result of improvement for potential built-in recovery interventions in a sustainable key. This goal translates operationally in monitoring the concentration levels of the most common indoor pollutants, as well as the levels of concentrations of pathogenic agents that these pollutants drag with them (consider the current Covid-19 emergency), via a network of sensors that allows to evaluate the building's "sickness" in an adequate time arc. We want to identify the level of correlation between the age of the building, its geolocation (translated in terms of outdoor environment) and any human health pathologies through the analysis of data derived from sensors and the analysis of the real estate market in the area.
Expected to increase the need for digitalization and technological innovation of processes in accordance with the I4.0 paradigm, technologies such as the IoT and CPS are also emerging in the field of buildings (Jo et al., 2020; Saini et al., 2020; Tong, 2020; Xiahou et al., 2019), this strategy aims to represent an excellent interpretative model of the operational reality to set information flows and thus calibrate a smartness model that allows the real estate sector to promote the transition to the I4.0 paradigm by inducing companies operating in the sector to deal with new challenges both at an organisational and technological level.
The second phase consists in providing the private know-how required to identify the most appropriate strategy time to the redevelopment and reclamation of the built, up to the achievement of a "Healthy House" certificate recognised at European and released levels. From the same company that provided the analysis and subsequent implementation of the building interventions. The condition sine qua non inherent in the realisation of a project of this magnitude is citizen awareness of a so delicate issue that is never explicitly discussed, neither among sector operators nor through public opinion.

HSH aspires to standardise a method for continuously monitoring physical well-being in a restricted environment, improving it over time with a 360° sustainable, innovative, and in line with the increasingly pressing need to live in a secure environment, owing to its multidisciplinary and cross-disciplinary nature.

5.3.3. Innovation degree

Consistent with the objectives stated, there is still no single body in Italian territory that provides comprehensive advice from the preliminary monitoring and analysis phase to the redevelopment and issuance of an official "Healthy House" certification. Furthermore, with regard to the Indoor Air Quality, which is directly related to human health (US EPA, 2014), there is currently no integrated and effective model to assist in the management of smart monitoring and control systems and to guide the real estate market. In the direction of the I4.0 world. The most critical point is the definition and selection of a comprehensive set of parameters, as well as the associated effective control logic of production processes in terms of predictability, leading to an automated improvement of the property's performance from a technological standpoint, including So economy, environmental impact, and, not least, sociality.
The LCA approach , which is used during the monitoring system selection phase and in greater detail during the design phase of the redevelopment interventions, aims to minimise the environmental impacts of the interventions to be carried out (Jönsson, 2000; Wu and Apul, 2015). Consistent with the proposed objectives, the LCA allows for a long-term key assessment, eroding the symbiotic relationship between a healthy environment and a sustainable environment.
HSH positions itself as an environmentally advanced project. Not only does it have scientific implications, but it has the potential to transform an entire industry, such as the construction business, which is characterized by a high pulverization of operational enterprises and a low proclivity for innovation. The initiative is in keeping with the goals of the UN's 2030 Agenda for a More Sustainable Future. The opportunity to operate in the metropolis-tana city of Bari will let the spin-off to test its capabilities on a housing stock of over 1,000,000 structures of various kinds, age of construction, and pathologies. HSH is highly scalable as a consequence of the many sensing and monitoring approaches, which are simple to install and inexpensive in cost, as it will be able to serve different sorts of stakeholders' and privates' expectations based on the varied types of buildings and pollution sources. Furthermore, unlike other short-lived academic spin-offs that are not market-oriented, the initiative is focused on a genuine need: ensuring a healthy environment that is not harmful to health. The targeted marketing, which will be created through events, publications, brochures, and citizen-science, will allow all prospective consumers of the 41 municipalities of the Metropolitan City of Bari to be reached and sensitized. This might be a trigger for improving the province of Bari's “Legambiente” ranking (Laurenti and Trentin, 2021), which now places it at the bottom of the list.

5.3.4. Application field of the case study

There are currently no national or European standards governing air quality in confined environments (Settimo, 2012). Airborne concentrations found were compared to the ATSDR (Agency for Toxic Substances and Disease) Minimum Risk Levels (MRL) to identify problematic concentrations (ATSDR, 2021). The international literature (Sagunski and Mangelsdorf, 2005) was searched for hydrocarbons (C9-C10) and aliphatic (C10-C13), but no reference for the population was found. As a result, this study, in conjunction with field investigations into pollutant-symptom-disease correlations, could provide a database with limit values relating to the regional territory.
This project's research and development will initially focus on the real estate market in the metropolitan area of Bari, the capital of the Puglia Region, before expanding to the entire national territory as a test-bed. As a result, the starting point is a preliminary analysis of the built, from a land registry, based on the year of construction of buildings, the related construction technologies, and the potential problems associated with these factors. Following that, citizens will be subjected to public awareness campaigns. It will also be beneficial

to increase the number of studio dwellings that will be subjected to preliminary technical and technological analysis, as well as sensing and assessing the health of the environment in relation to the symptoms identified by the tenants themselves.

Chemical, physical, and biological agents are all capable of altering indoor air quality; they come from the outside (outdoor air pollution, pollen), but many are produced internally. The following are the primary internal sources of pollution: occupants (humans and animals), dust (an excellent receptacle for microorganisms), structures, furnishings, systems (air conditioners, humidifiers, plumbing), and outdoor air. As a classification criterion, we distinguish biological sources, combustion processes, and building materials, as well as ventilation systems.

Several common contaminants and toxins contribute to a sick house (see Figure 5.14):

- Volatile Organic Compounds (VOCs). VOCs are defined as any organic compound with a vapour pressure of 0.01 KPa or higher at 293.15 K (20° C) (“Legislative Decree 152/2006, art.268-11,” 2006). This class includes a wide range of chemical compounds such as aliphatic, aromatic, and chlorinated hydrocarbons, aldehydes, terpenes, alcohols, esters, and ketones. The most common in residential buildings are limonene and toluene, but formaldehyde is the most toxicologically and mutagenically important. VOCs can have a wide range of effects, ranging from sensory discomfort to serious health changes; at high concentrations in indoor environments, they can affect numerous organs or systems, particularly the central nervous system. Some of them are known to be carcinogenic to humans or animals (benzene). Indoor VOC pollution has been hypothesised to pose a carcinogenic risk to subjects who spend a significant amount of time in confined environments, though these assessments are not yet conclusive due to insufficient characterization of such pollution.
- Mold. Mites, pet epidermal derivatives, cockroaches, and fungi are the most common indoor allergens. The presence of fungi in the environment is associated with high relative humidity, which promotes their growth. They can grow both inside and outside of structures. Inside, they are primarily found where there is excess moisture and poor ventilation, and they tend to develop more quickly in hot humid climates, such as summer, and in poorly lit places, on damp objects and materials, in humidifiers or air conditioning systems that are not subjected to regular cleaning and maintenance. It is important to keep in mind the possibility of certain fungal species developing in air conditioning systems. The Alternaria species causes a type of mold that grows on decaying fruits and vegetables and in particularly humid environments, releasing its spores particularly on wallpaper, carpets, and soil. Mold is one of the most common causes of allergic reactions such as asthma, conjunctivitis, rhinitis, and dermatitis. In particular, nocturnal and daytime cough, as well as the relationship with asthma and sensitization to inhalant allergens in the most crowded families.
- Formaldehyde. Formaldehyde is an organic compound in the vapour phase with a strong odour. Aside from being a by-product of combustion (tobacco smoke and other combustion sources), it is also emitted by urea-formaldehyde resins used for insulation, as well as resins used for wood chipboard and plywood, upholstery, carpeting, curtains and other textiles anti-crease treatments, and other furnishing material. In most homes, the levels range between 0.01 and 0.05 mg/m3. Also, the indoor levels of this compound are generally higher than the outdoor levels. The levels in indoor environments are typically between 10 and 50 g/m3. The highest concentrations have been found in prefabricated houses, after building interventions, and in rooms with recently laid chipboard, parquet, or carpet furniture. Formaldehyde causes eye, nasal, and throat irritation, sneezing, coughing, fatigue, and skin erythema in susceptible or immunologically sensitised individuals; however, subjects susceptible or immunologically sensitised to formaldehyde may have adverse reactions even at lower concentrations. Formaldehyde concentrations detected in homes can be of the order of those that irritate the airways and mucous membranes, especially after building interventions or the installation of new furniture or furnishings.
- Asbestos. Asbestos is a fibrous material made up of natural mineral fibres from the silicates and mineralogical series serpentine (chrysotile or white asbestos) and amphiboles (crocidolite or blue

asbestos). Mineral fibres include both natural fibrous materials, such as asbestos, and synthetic fibres, such as glass wool, stone wool, and other materials. Asbestos is typically found in a compact form, incorporated into a cement matrix (asbestos cement on the roof, chimneys, etc.) or other matrices (linoleum floors, walls, panels, etc.), but it can also be found in a crumbly, more dangerous form, when used as a soundproofing or insulating material on false ceilings and / or walls. Asbestos fibres can be released inside buildings where it is present due to slow deterioration of the constituent materials (insulators or insulators), direct damage to them by occupants, or improper maintenance. The use of asbestos is currently prohibited by law; however, the release of asbestos fibres from pre-existing structural elements inside buildings can occur due to slow deterioration of materials containing it, direct damage to them by occupants, or maintenance interventions. Even in the presence of few fibrous elements, the presence of asbestos fibres in the environment inevitably causes health problems. It is a carcinogen. The main dangers are associated with the presence of fibres in the air. When inhaled, the fibres can become lodged in the airways and pulmonary cells. The fibres that have been deposited in the deeper parts of the lung can remain in the lungs for many years, if not for the rest of one's life. The presence of extraneous fibres within the lungs can result in the development of diseases such as asbestosis, mesothelioma, and lung cancer. Its cases are strongly linked to the presence of asbestos geodispersal, and it manifests itself after 15-30 years. Also, if in a less severe form, intestinal tract pathologies and asbestos-related laryngeal exhibition have been discovered.

- Particulates. There are two major classes that can be distinguished. Particulate is classified into two types based on size and composition: coarse particulate and fine particulate. The coarse particulate is made up of particles larger than 10 μm in diameter, such as pollen and spores. They are typically retained by the upper respiratory system (nose, larynx, and trachea), and particles with a diameter of less than 2.5 micrometres (PM2.5) penetrate deeply into the lungs, particularly when breathing through the mouth. The aero dispersed particulate has the ability to adhere toxic gases and vapours to its surface. This phenomenon contributes to increasing concentrations of gaseous pollutants, which are transported by PM10 and PM2.5 particles, reaching the deepest areas of the lung. Numerous studies have found a link between acute airborne particulate matter exposure and respiratory symptoms, changes in respiratory function, hospitalizations, and mortality from respiratory diseases. Furthermore, long-term exposure to particulate matter, beginning with low doses, has been linked to an increase in mortality from respiratory diseases and diseases such as chronic bronchitis, asthma, and reduced respiratory function. Furthermore, long-term exposure is likely to increase the risk of respiratory tract cancer. Cancer has been linked to exposure to combustion particulate (fine particulate); soot has carcinogenic properties, and numerous polycyclic aromatic hydrocarbons, some of which are carcinogens, are absorbed on the fine particulate that is deeply inhaled in the lungs. It should be noted that the World Health Organization has recommended that the concentration of this pollutant be kept as low as possible; there is no threshold level below which no health effects have been demonstrated.
- Radon. Radon is a chemically inert noble gas that exists in the atmosphere as a monoatomic gas. Furthermore, because Radon has no odour or colour, his presence cannot be detected by the senses. Radon is found in nature as a result of the radioactive decay of uranium and thorium, both of which are abundant in the earth's crust. Because it is a radioactive gas, it disperses quickly in the atmosphere while concentrating in enclosed spaces and is thus considered a pollutant, typically indoors. It is primarily derived from rocks in the subsoil, particularly those of volcanic origin, or from building materials rich in natural radionuclides. Another source is water (which is soluble in it). Because the main source of Radon in a building is the soil on which it rests, the buries, basement, and all those on the ground floor are the most vulnerable to this type of pollution. The great variability of indoor radon concentration (from about 10 bq/m3 to several thousand bq/m3) is linked not only to power and physical characteristics of its main sources (soil and building materials), but also to microclimatic parameters (pressure and temperature), building construction techniques, and ventilation. Radon is thus a radioactive gas that originates primarily in the soil and is present in all buildings, albeit in

varying concentrations from one to the next. Radon produces a number of radioactive decay products, which attach to aerosol particles and only a portion of them remain free. When radon and its decay products are inhaled, they can decay inside the respiratory system, emitting high-energy ionising radiation, particularly alpha particles. In reality, radon serves primarily as a transporter and source of its decay products; the latter, particularly the particles, are primarily responsible for the health effects. Radon is the most important natural source of ionising radiation exposure for the population as a whole, and it is a significant risk factor for human health. The International Agency for Research on Cancer has classified radon gas and its decay products in the Cancerogenic Group 1, that is, in the group of substances for which there is Sufficient evidence of carcinogenicity based on human studies. Particles enter the lungs via breathing and can damage the DNA of pulmonary tissue cells, eventually transforming them into tumour cells. After tobacco smoke, radon is most likely the most important single agent inducing lung cancer. As a result, exposure to indoor radon in homes increases the risk of developing a pulmonary tumour, and it is estimated that radon is responsible for 3% to 14% of all pulmonary tumours.

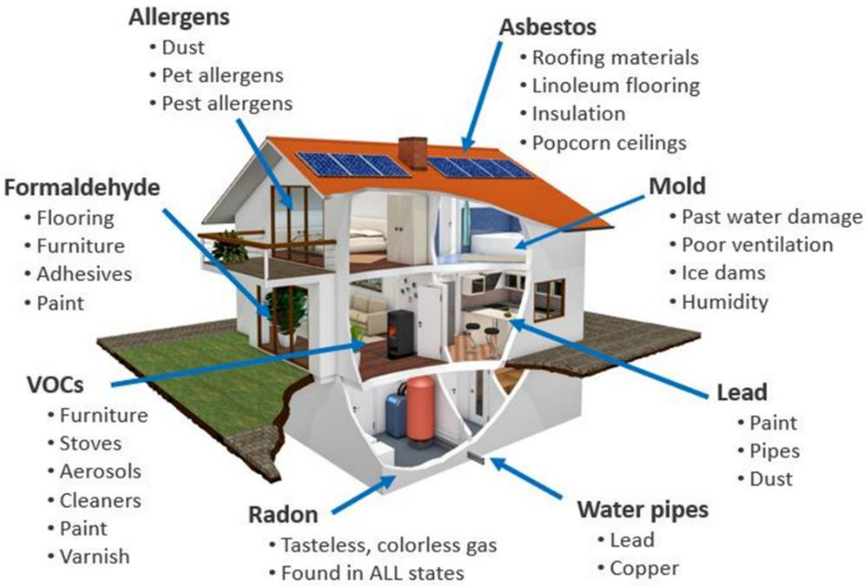


Figure 5.14 - A diagram depicting the most common contaminants and toxins, as well as their locations in a home. From ("Sick House Syndrome. Environmental Testing Products, Live Pure, Inc." 2021)

The main pollutants for each source of pollution are listed in detail in Table 5.12, to provide a more precise overview of the significant amount of both sources and pollutants emitted in our homes, as well as the problems caused by contaminated that can be found in homes with residential use and the period of development of the symptoms caused by pollution exposure. Symptoms can include headache, dry cough, itchy skin, dizziness and nausea, fatigue, difficulty concentrating, sensitivity to smells as well as eye, nose or throat irritation. The causes of the symptoms remain unclear. It can be short-term and manifests all of the symptoms that occur after a single exposure of the individual to the pollutant.

Table 5.12 - Pollutants of greatest concern for each source of pollution and their effects on health

Source	Pollutants	Effects on health
Gas or coal combustion processes for heating and/or cooking, fireplaces and wood stoves, vehicle exhaust gases	Products derived from combustion (CO, NOx, SO2, Particulate)	Increased occurrence of chronic respiratory symptoms, as well as a possible reduction in ventilatory respiratory function, sensitization thresholds to various allergens may be lowered, a higher risk of developing COPD
Building materials and insulation	Asbestos, synthetic vitreous fibres, particulate matter, radon; organic agents (due to the presence of moisture and/or powder)	diseases such as asbestosis, mesothelioma and lung cancer, possible pathologies of the intestinal tract and for the larynx
Carpets and coating materials	Formaldehyde, acrylates, VOCs and biological agents (due to the presence of moisture and/or dust)	Possible broncho reactive effects in asthmatic patients
Furnitures	Formaldehyde, VOCs and biological agents (due to the presence of moisture and/or dust)	Possible broncho reactive effects in asthmatic patients
Cleaning products and liquids	Alcohols, phenols, VOCs	sensory discomfort, effects on many organs or systems, in particular on the central nervous system
Photocopiers	Ozone (O3), toner powder, VOCs	sensory discomfort, effects on many organs or systems, in particular on the central nervous system, irritative effects on the ocular mucous membranes and the upper airways, coughing, bronco obstructive phenomena and alteration of respiratory function
Cigarette smoke	Polycyclic hydrocarbons, VOC formaldehyde, CO, fine particulate matter	sensory discomfort, effects on many organs or systems, in particular on the central nervous system, respiratory symptoms, changes in respiratory function, chronic bronchitis, asthma, anginal crisis, headache, confusion, disorientation, dizziness, impaired vision and nausea
Air conditioning systems	CO2 and VOCs (due to a low number of hourly spare parts or excess recycling), biological agents (due to lack of cleaning and maintenance)	sensory discomfort, effects on many organs or systems, in particular on the central nervous system, chronic bronchitis, asthma
Dust	Biological agents (as indoor allergens like mites)	asthma, conjunctivitis, rhinitis and dermatitis
People	CO2 and biological agents (bacteria, viruses, etc.)	sensory discomfort, effects on many organs or systems, in particular on the central nervous system, headache, confusion, disorientation, dizziness, impaired vision and nausea
Pets	Indoor allergens (hair etc.)	asthma, conjunctivitis, rhinitis and dermatitis
Natural hot springs (lava, tuff, granite, etc.)	Radon	lung diseases such as emphysema, chronic interstitial pneumonia and pulmonary fibrosis

These are effects that are usually observed during or immediately following exposure to the pollutant, and they are treatable, short-lived, and easily eliminated by simply moving away from the source of pollution. The graph in Figure 5.15 depicts examples of these types of symptoms. The other type of disease develops over time after an individual has been exposed to one or more pollutants for an extended period of time, or if the exposure has occurred repeatedly. They usually do not occur immediately after exposure, but rather after some time, even years, and can have serious consequences involving the skin (dermatitis), the respiratory system (nasopharyngeal tumours, nasal cavities), and the cardiovascular system.

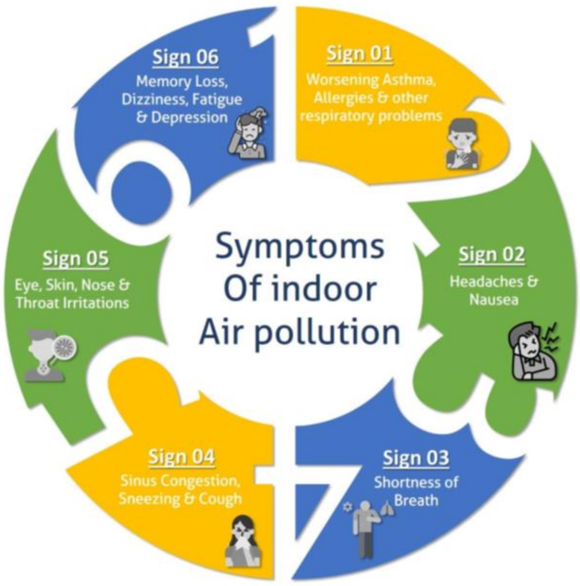


Figure 5.15 - Main symptoms of indoor air pollution

5.3.5. **Planning**

The project was scheduled to be completed in 18 months (Figure 5.16):
Phase 1: raising awareness of the initiative through events and conferences;
Phase 2: state-of-the-art techniques and problems to be realised on the field and in the literature;
Phase 3: begin monitoring actions;
Phase 4: results communication to the target, data analysis, and preliminary diagnosis;
Phase 5: creation of registers to be used in future retrofit actions;
Phase 6: implementation of planned retrofit interventions based on the results of the analysis and diagnosis;
Phase 7: submission of the results obtained from the various competent territorial planning bodies in order to obtain healthy house certification.

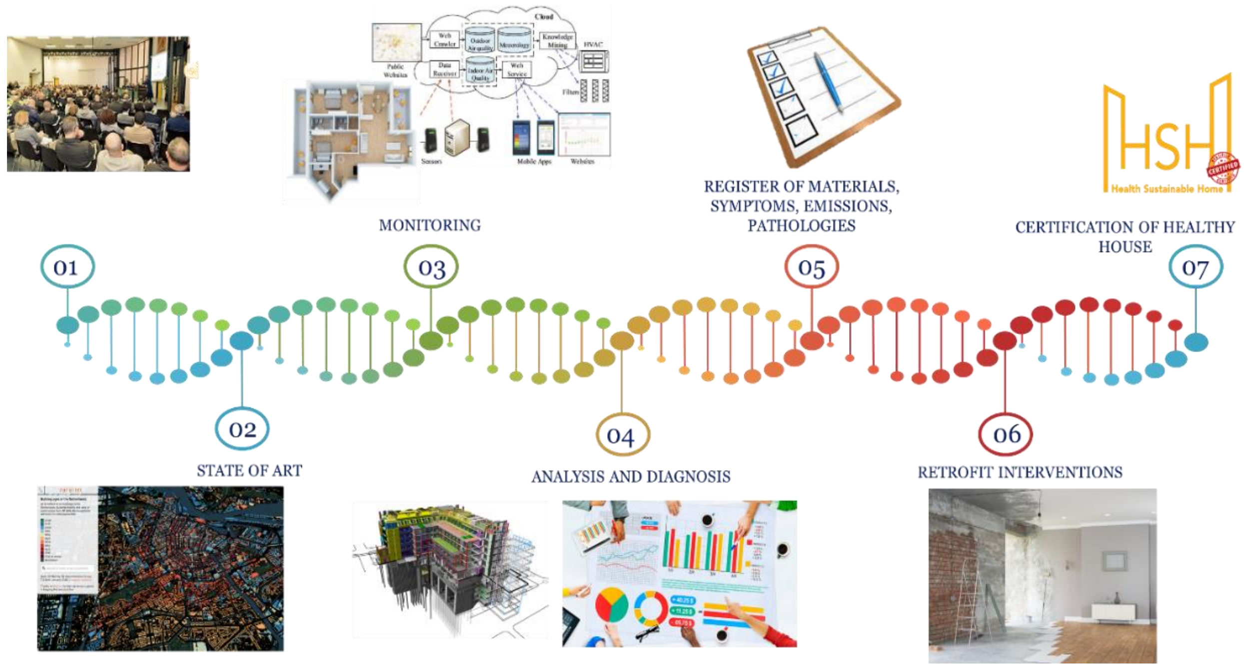


Figure 5.16 - HSH activities planning

CONCLUSIONS

In the direction of sustainable manufacturing production, the fourth industrial revolution has been fully imposed. In the driving sectors of the industrial field, encouraging experiences are emerging from various organizations across the world, particularly from SMEs, where the best practices of smart manufacturing and digital fabrication have been introduced and applied (from the local scale to the global). As previously proven, these final two principles are at the heart of the (sustainable) predictive manufacturing paradigm. According to the scientific contributions present in the literature on the subject, a massive development of Maturity Models aimed at the qualitative and quantitative assessment of the maturity and smartness level of companies in the implementation of technologies related to the I4.0 Paradigm have been realized.

The advent of the Internet of Things (IoT), based on major breakthroughs in sensor technology, has been the beginning point for most of this transition, resulting in smaller, cheaper, and better-performing sensors. These sensors, which have increased in performance, have been distributed across factory floors, allowing data to be collected from a greater range of industrial equipment. Because of pervasive connectivity, this data can be swiftly transmitted to the cloud, where it can be processed and analysed, increasingly utilizing technologies such as machine learning and artificial intelligence to provide real-time insights. This IoT-enabled process is a key element of I 4.0, where broad automation and data collecting may deliver significant commercial benefits. Manufacturing-economic sustainability is one of I4.0’s most urgent sustainability consequences. When successfully implemented, I4.0 manufacturing-economic sustainability features will lay the groundwork for I4.0’s continued commitment to environmental and socio-economic sustainability.

In the I4.0 era, the modularity, flexibility, and agility of smart factories, as well as the product customization philosophy of manufacturing digitization, increase product cycles and hasten the obsolescence of products and services. This environmentally unfavourable situation could result in increased demands for energy and services, as well as increased pollution and waste generation. This situation necessitates international policies and multilateral negotiations to mitigate the unanticipated environmental protection impacts of I4.0 and industrial digitization. Smart and autonomous production processes can help employee health and safety by taking on monotonous and routine activities, resulting in higher employee productivity and motivation. However, I4.0 advances carry with them a slew of new problems and vulnerabilities for society. The introduction of smart manufacturing systems resulted in an increase in the market for information technology professionals. Qualified workers would be needed in the information technology industry to plan, create, manage, and sustain network systems. As a result, there will be an increase in career prospects in information technology. However, in manufacturing plants, the number of machine operators and laborers would be limited. In light of what has just been said, reduced jobs, information security challenges, data sophistication, electronic waste, and low consistency, for example, may occur.

Automated systems, on the other hand, frequently struggle with low-quality data, resulting in more manual work or sub-optimal judgments with unfavourable financial effects.

What was really accomplished, in accordance with the goals established at the beginning, is detailed below:

- Continuous improvement of the entire production process and each sub-system at every stage of the life cycle, using innovative I and II order models (Life Cycle Assessment coupled with Exergy Analysis) and experimental models (Prognostic Health Management), using a top-down (model-based) and bottom-up (data-driven) approach;
- Definition of critical process parameters using a thermodynamic-technological model of the dynamic behaviour of the manufacturing process;
- Measurement and monitoring of these parameters using an ad hoc sensor system and an extremely integrated IoT architecture for the construction of a fully structured database;
- Big-data processing (very heterogeneous dataset) using Machine Learning algorithms optimally configured for such types of data available and centred on company requirements;
- Investigation of both models' performance metrics;

- Interpretation of the findings of analyses and tests conducted in order to define the most automated and enhanced business strategies.

Here are some general considerations arose following the dissertation work done.

Sustainability is not explicitly one of the core keywords for I4.0. To order for sustainability to really fit under the I4.0 framework, it is necessary to add it to the KETs of the paradigm and to incorporate the Circular Economy as one of its key-features, not only as contingent benefit. In fact, resuming the review carried out by (Bonilla et al., 2018) on the effects that the implementation of I4.0 paradigm would have on the environment, the findings suggest that the features and the KETs provide a variety of possibilities for environmental sustainability when properly designed. On the other side, the long-term scenario of I4.0 is directly related to social responses as well as public policy, regulatory systems and homogeneous distribution. Heterogeneity between nations embracing I4.0 as well as between companies with different levels of digitalized technology may build market segments of disparity and non-sustainability trends. An effective combination of data-driven and model-based techniques will be necessary, with careful consideration of the advantages and drawbacks of both model-based and data-driven approaches.

Advanced data-driven techniques should increase the applicability, capacity, and efficiency of basic data-driven process monitoring methods under varied industrial operating situations without requiring complicated system design efforts. Demand for companies to reach the best decisions based on real-time data insights have never been greater.

The modelling of complex technological systems serves as the foundation for enhancing process performance, including sustainability features (triple-bottom line). The proposed integrated model (Model-Based and Data-Driven) provides an excellent interpretative prototype of industrial reality and a high degree of predictability, in contrast to the only linear model, which has significant limitations (for example, estimation of processes that occur through the use of non-specific standardized databases compared to the geographical and time location of the analysed systems, thus providing generic results subject to excessive free interpretation). The research has revealed that the most crucial step is the definition and selection of a comprehensive set of parameters, as well as the identification of the associated processing logic and effective control in terms of predictability.

The latest cloud technologies, big-data, and IoT are exciting and promising for the present and future. However, in order for all of these technologies to be effective, they must be acknowledged, adapted to the specificity of the system under investigation, and a new style of thinking must be adopted (using an holistic approach, like the one followed in this work). Complexity is decreased as a result, and improvements can be easily seen. According, a data-driven company must view data management and analytics as a strategic pillar of the business rather than a technical factor. Being data-driven entails being guided by numbers, taking a data-driven approach, and making informed decisions based on objective facts rather than personal feelings. As a result, transforming into a data-driven company requires more than just technology, but also a change management strategy capable of bringing data culture to all levels of the organization. In such a fast-paced world, it is not enough to focus on the past, to analyse metrics and KPIs based on historical data, to generate statistics and final reports, to conduct data analysis on user behaviour, or to identify technical problems or critical events. CEOs and managers today require information to help them understand what the future holds. It is critical to have accurate, up-to-date, and frequently collected data available.

The challenges in implementing machine learning within existing manufacturing processes are: (1) data scientists must have access to an open interface (for example, for data collecting, training, and deployment) to enable interoperability across different frameworks; (2) machine learning must be easy enough to utilize without the need for specialized expertise; in other words, solutions must be able to integrate with current software infrastructure; (3) many machine learning algorithms are intrinsically inaccurate and should be handled as such. Unsuitable solutions must be modified or abandoned. This gives people confidence that the taught algorithms are trustworthy; (4) the training methods used must be somewhat robust, that is, they must operate even with tiny amounts of noisy data; (5) transparency and interpretability are critical for many businesses. The more complicated are their requirements, the more critical it is to fully comprehend the methods to be employed. This is an area where research is still in its early stages; and (6) recognize which kind of data are available and that a predictive design solution is built using a chain of tunings and algorithms rather than a single pre-set algorithm. All of the preceding conditions must be accomplished in order for machine learning and real-time control to be successfully integrated.

The activities conducted in this dissertation generates new vital awareness and cross-disciplinary knowledge that spans the macro-scopes of smart factory and I4.0, industrial applied thermodynamic-sustainability assessment, big-data analytics and machine learning, stationary and dynamic complex system holistic modelling. Developing A strategy for comprehensively defining the technological, social, environmental, diagnostic, and prognostic economic components that characterize the manufacturing process throughout its life cycle (including future forecasts) in order to assess more appropriate, intelligent, and fast policies for business planning and decision-making.

It is finally clear from the preceding considerations that, because the Master Italy production context is extremely dynamic, it will certainly be necessary to perform maintenance activities on the algorithms developed where changes in the operating conditions of the machine, entry of new plants, or expansions of the data set collected by the same machine are required. In fact, the SCADA system is already being expanded with sensors for detecting environmental parameters (humidity, temperature, air speed, etc.), with the goal of generating enough data to investigate any correlations between phenomena of production defects and specific values of environmental parameters. This will result, for example, in a new model tuning or even a new adaptation of the algorithms. Finally, an interactive dashboard that connects with field operators in real time would create (completing the circle) an end-to-end framework for monitoring data analysis, problem diagnosis, and failure prognosis. This entails laying strong foundations in order to achieve a formal-compositional sophistication that legitimizes Master Italy business plus value on the global market in terms of quality, customisation, flexibility, and sustainability.

LIST OF FIGURES

LIST OF TABLES

REFERENCES

Abadias Llamas, A., Bartie, N., Heibeck, M., Stelter, M., Reuter, M., 2019. Simulation-Based Exergy Analysis of Large Circular Economy Systems: Zinc Production Coupled to CdTe Photovoltaic Module Life Cycle. Journal of Sustainable Metallurgy 6. https://doi.org/10.1007/s40831-019-00255-5

Abusohyon, I.A.S., Crupi, A., Bagheri, F., Tonelli, F., 2021. How to Set Up the Pillars of Digital Twins Technology in Our Business: Entities, Challenges and Solutions. Processes 9, 1307. https://doi.org/10.3390/pr9081307

Açıkkalp, E., Hepbasli, A., Yucer, C.T., Karakoc, T.H., 2018. Advanced life cycle integrated exergoeconomic analysis of building heating systems: An application and proposing new indices. Journal of Cleaner Production 195, 851–860. https://doi.org/10.1016/j.jclepro.2018.05.239

Aghbashlo, M., Rosen, M.A., 2018a. Exergoeconoenvironmental analysis as a new concept for developing thermodynamically, economically, and environmentally sound energy conversion systems. Journal of Cleaner Production 187, 190–204. https://doi.org/10.1016/j.jclepro.2018.03.214

Aghbashlo, M., Rosen, M.A., 2018b. Consolidating exergoeconomic and exergoenvironmental analyses using the emergy concept for better understanding energy conversion systems. Journal of Cleaner Production 172, 696–708. https://doi.org/10.1016/j.jclepro.2017.10.205

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases, in: Acm Sigmod Record. ACM, pp. 207–216.

Ahn, D.-G., 2016. Direct metal additive manufacturing processes and their sustainable applications for green technology: A review. Int. J. of Precis. Eng. and Manuf.-Green Tech. 3, 381–395. https://doi.org/10.1007/s40684-016-0048-9

Alberto Navajas, Leire Uriarte, Luis Gandía, 2017. Application of Eco-Design and Life Cycle Assessment Standards for Environmental Impact Reduction of an Industrial Product. Sustainability 9, 1724. https://doi.org/10.3390/su9101724

Aleksic, S., Mujan, V., 2016. Exergy-based life cycle assessment of smart meters, in: 2016 ELEKTRO. Presented at the 2016 ELEKTRO, IEEE, Strbske Pleso, High Tatras, Slovakia, pp. 248–253. https://doi.org/10.1109/ELEKTRO.2016.7512075

Ali, S.M.F., Wrembel, R., 2017. From conceptual design to performance optimization of ETL workflows: current state of research and open problems. VLDB Journal 26, 777–801. https://doi.org/10.1007/s00778-017-0477-2

Almanei, M., Oleghe, O., Jagtap, S., Salonitis, K., 2021. Machine Learning Algorithms Comparison for Manufacturing Applications 6. https://doi.org/10.3233/ATDE210065

Alvarenga, R.A.F., Dewulf, J., Van Langenhove, H., Huijbregts, M.A.J., 2013. Exergy-based accounting for land as a natural resource in life cycle assessment. The International Journal of Life Cycle Assessment 18, 939–947. https://doi.org/10.1007/s11367-013-0555-7

Alvarez, I., 2013. Approach for the practical application of Exergy Analysis within branch-oriented enterprise networks, towards the realization of an Exergy Life Cycle Assessment (ELCA), in: EnviroInfo.

Amornsamankul, S., Pimpunchat, B., Triampo, W., Charoenpong, J., Nuttavut, N., 2019. A Comparison of Machine Learning Algorithms and Their Applications. International journal of simulation: systems, science & technology. https://doi.org/10.5013/ijssst.a.20.04.08

Andersson, J., 2013. Life cycle assessment in production flow simulation for production engineers. 22nd International Conference on Production Research 7.

Arnal, Á.J., Díaz-Ramírez, M., Acevedo, L., Ferreira, V.J., García-Armingol, T., López-Sabirón, A.M., Ferreira, G., 2020. Multicriteria Analysis for Retrofitting of Natural Gas Melting and Heating Furnaces for Sustainable Manufacturing and Industry 4.0. Journal of Energy Resources Technology 142, 022203. https://doi.org/10.1115/1.4044769

Arrizubieta, J.I., Ukar, O., Ostolaza, M., Mugica, A., 2020. Study of the Environmental Implications of Using Metal Powder in Additive Manufacturing and Its Handling. Metals 10, 261. https://doi.org/10.3390/met10020261

Assefa, G., Frostell, B., 2007. Social sustainability and social acceptance in technology assessment: A case study of energy technologies. Technology in Society 29, 63–78. https://doi.org/10.1016/j.techsoc.2006.10.007

Atluri, G., Karpatne, A., Kumar, V., 2017. Spatio-Temporal Data Mining: A Survey of Problems and Methods. arXiv:1711.04710 [cs].

ATSDR, 2021. Minimal Risk Levels (MRLs) for Professionals [WWW Document]. oxic Substances Portal. URL https://www.atsdr.cdc.gov/mrls/index.html (accessed 10.8.21).
Ayres, R.U., Ayres, L.W., Martinás, K., 1998. Exergy, waste accounting, and life-cycle analysis. Energy 23, 355–363. https://doi.org/10.1016/S0360-5442(97)00076-5
Ayres, R.U., Ayres, L.W., Martinas, K., 1996. EcO-THERMODYNAMICS: EXERGY AND LIFE CYCLE ANALYSIS 57.
Bai, C., Dallasega, P., Orzes, G., Sarkis, J., 2020. Industry 4.0 technologies assessment: A sustainability perspective. International Journal of Production Economics 229, 107776. https://doi.org/10.1016/j.ijpe.2020.107776
Bakshi, B.R., Fiksel, J., 2003. The quest for sustainability: Challenges for process systems engineering. AIChE Journal 49, 1350–1358. https://doi.org/10.1002/aic.690490602
Bakshi, B.R., Gutowski, T., Sekulic, D., 2011. Thermodynamics and the Destruction of Resources 524.
Bárkányi, Á., Chován, T., Németh, S., Abonyi, J., 2021. Modelling for Digital Twins—Potential Role of Surrogate Models. Processes 9, 476. https://doi.org/10.3390/pr9030476
Bartolozzi, I., Rizzi, F., Frey, M., 2013. Comparison between hydrogen and electric vehicles by life cycle assessment: A case study in Tuscany, Italy. Applied Energy 101, 103–111. https://doi.org/10.1016/j.apenergy.2012.03.021
Batini, C., Cappiello, C., Francalanci, C., Maurino, A., 2009. Methodologies for Data Quality Assessment and Improvement. ACM Comput. Surv. 41. https://doi.org/10.1145/1541880.1541883
Baumann, P., Hochbaum, D.S., Yang, Y.T., 2019. A comparative study of the leading machine learning techniques and two new optimization algorithms. European Journal of Operational Research 272, 1041–1057. https://doi.org/10.1016/j.ejor.2018.07.009
Beccali, G., Cellura, M., Mistretta, M., 2003. New exergy criterion in the “multi-criteria” context: a life cycle assessment of two plaster products. Energy Conversion and Management 44, 2821–2838. https://doi.org/10.1016/S0196-8904(03)00043-8
Becue, A., Maia, E., Feeken, L., Borchers, P., Praça, I., 2020. A New Concept of Digital Twin Supporting Optimization and Resilience of Factories of the Future. Applied Sciences 10, 4482. https://doi.org/10.3390/app10134482
Belhadi, A., Zkik, K., Cherrafi, A., Yusof, S.M., El fezazi, S., 2019. Understanding Big Data Analytics for Manufacturing Processes: Insights from Literature Review and Multiple Case Studies. Computers & Industrial Engineering 137, 106099. https://doi.org/10.1016/j.cie.2019.106099
Bemporad, A., Morari, M., Ricker, N.L., n.d. Model Predictive Control Toolbox 246.
Bertalanffy, L.V., 1950. The Theory of Open Systems in Physics and Biology. Science, New Series 111, 23–29.
Bhalekar, A.A., 1996. Does Clausius’ Inequality Analogue Exist for Open Systems? Journal of Non-Equilibrium Thermodynamics 21. https://doi.org/10.1515/jnet.1996.21.4.330
Bhatnagar, R., 2018. Machine Learning and Big Data Processing: A Technological Perspective and Review, in: Hassanien, A.E., Tolba, M.F., Elhoseny, M., Mostafa, M. (Eds.), The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018), Advances in Intelligent Systems and Computing. Springer International Publishing, Cham, pp. 468–478. https://doi.org/10.1007/978-3-319-74690-6_46
Bishop, C.M., 2016. Pattern Recognition and Machine Learning, Information Science and Statistics. Springer New York.
Bjerkholt, O., Kurz, H.D., 2006. Introduction: the History of Input – Output Analysis, Leontief’s Path and Alternative Tracks. Economic Systems Research 18:4, 331–333, 4. http://dx.doi.org/10.1080/09535310601020850
Blondheim, D., 2021. Improving Manufacturing Applications of Machine Learning by Understanding Defect Classification and the Critical Error Threshold. International Journal of Metalcasting 1–19. https://doi.org/10.1007/s40962-021-00637-0
Boateng, A.A., 2016. 9 - Mass and Energy Balance, in: Boateng, A.A. (Ed.), Rotary Kilns (Second Edition). Butterworth-Heinemann, Boston, pp. 203–229. https://doi.org/10.1016/B978-0-12-803780-5.00009-5
Bogle, I.D.L., 2017. A Perspective on Smart Process Manufacturing Research Challenges for Process Systems Engineers. Engineering 3, 161–165. https://doi.org/10.1016/J.ENG.2017.02.003
Böhringer, C., Jochem, P.E.P., 2007. Measuring the immeasurable — A survey of sustainability indices. Ecological Economics 63, 1–8. https://doi.org/10.1016/j.ecolecon.2007.03.008

Bonilla, S., Silva, H., Terra da Silva, M., Franco Gonçalves, R., Sacomano, J., 2018. Industry 4.0 and sustainability implications: A scenario-based analysis of the impacts and challenges. Sustainability 10, 3740.

Bösch, M.E., Hellweg, S., Huijbregts, M.A.J., Frischknecht, R., 2007. Applying cumulative exergy demand (CExD) indicators to the ecoinvent database. Int J Life Cycle Assess 12, 181–190. https://doi.org/10.1065/lca2006.11.282

Boulanger, P.-M., 2008. Sustainable development indicators: a scientific challenge, a democratic issue. S.A.P.I.EN.S. Surveys and Perspectives Integrating Environment and Society.

Bousdekis, A., Lepenioti, K., Apostolou, D., Mentzas, G., 2021. A Review of Data-Driven Decision-Making Methods for Industry 4.0 Maintenance Applications. Electronics 10, 828. https://doi.org/10.3390/electronics10070828

Boyaghchi, F.A., Molaie, H., 2015. Sensitivity analysis of exergy destruction in a real combined cycle power plant based on advanced exergy method. Energy Conversion and Management 99, 374–386. https://doi.org/10.1016/j.enconman.2015.04.048

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

British Standards Institution, 2011. PAS 2050. Specification for the assessment of the life cycle greenhouse gas emissions of goods and services [WWW Document]. British Standards Institution BSI. URL www.bsigroup.com/Standards-and-Publications/ How-we-can-help-you/Professional-Standards-Service/ PAS-2050/

Brockway, P.E., Kjelstrup, S., Dewulf, J., Siebentritt, S., Valero, A., Whelan, C., 2016. In a resource-constrained world: Think exergy not energy. https://doi.org/10.13140/RG.2.1.4507.0326

Brous, P., Janssen, M., Herder, P., 2020. The dual effects of the Internet of Things (IoT): A systematic review of the benefits and risks of IoT adoption by organizations. International Journal of Information Management 51, 101952. https://doi.org/10.1016/j.ijinfomgt.2019.05.008

Brownlee, J., 2016. Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch. Machine Learning Mastery.

Buchgeister, J., 2010. Exergoenvironmental Analysis - A New Approach to Support the Design for Environment of Chemical Processes? Chem. Eng. Technol. 33, 593–602. https://doi.org/10.1002/ceat.201000006

Bueno, C., Fabricio, M.M., 2018. Comparative analysis between a complete LCA study and results from a BIM-LCA plug-in. Automation in Construction 90, 188–200. https://doi.org/10.1016/j.autcon.2018.02.028

Cabrera, D., Cabrera, L., 2019. What Is Systems Thinking?, in: Spector, M.J., Lockee, B.B., Childress, M.D. (Eds.), Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy. Springer International Publishing, Cham, pp. 1–28. https://doi.org/10.1007/978-3-319-17727-4_100-1

Cafagna, Gruppo Master Italy Srl: “ecco come la digitalizzazione ha migliorato i processi” [WWW Document], 2018. . Corriere dell’Economia. URL https://www.corrieredelleconomia.it/2018/06/06/cafagna-gruppo-master-italy-srl-ecco-come-la-digitalizzazione-ha-migliorato-i-processi/ (accessed 11.28.19).

Cambridge English Dictionary, n.d. HYBRID: meaning in the Cambridge English Dictionary [WWW Document]. Cambridge English Dictionary. URL https://dictionary.cambridge.org/dictionary/english/hybrid (accessed 4.14.20).

Camoletto, M., 2017. Social Housing e misurazione degli impatti sociali: passi avanti verso un toolkit comune 29.

Cao, Q., Zanni-Merk, C., Reich, C., 2018. Ontologies for Manufacturing Process Modeling: A Survey.

Cappa, F., Oriani, R., Peruffo, E., McCarthy, I., 2021. Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance*. J. Prod. Innov. Manag. 38, 49–67. https://doi.org/10.1111/jpim.12545

Cardinale, N., Rospi, G., Stefanizzi, P., 2013. Energy and microclimatic performance of Mediterranean vernacular buildings: The Sassi district of Matera and the Trulli district of Alberobello. Building and Environment 59, 590–598. https://doi.org/10.1016/j.buildenv.2012.10.006

Carvalho, N., Chaim, O., Cazarini, E., Gerolamo, M., 2018. Manufacturing in the fourth industrial revolution: A positive prospect in Sustainable Manufacturing. Procedia Manufacturing, 15th Global Conference on Sustainable Manufacturing 21, 671–678. https://doi.org/10.1016/j.promfg.2018.02.170

Casas-Ledón, Y., Spaudo, F., Arteaga-Pérez, L.E., 2017. Exergoenvironmental analysis of a waste-based Integrated Combined Cycle (WICC) for heat and power production. Journal of Cleaner Production 164, 187–197. https://doi.org/10.1016/j.jclepro.2017.06.211
Cavalcanti, E.J.C., Carvalho, M., Ochoa, A.A.V., 2019. Exergoeconomic and exergoenvironmental comparison of diesel-biodiesel blends in a direct injection engine at variable loads. Energy Conversion and Management 183, 450–461. https://doi.org/10.1016/j.enconman.2018.12.113
CE Delft, 2020. Health costs of air pollution in European cities and the linkage with transport.
Chen, L., 2014. Curse of Dimensionality. pp. 1–1. https://doi.org/10.1007/978-1-4899-7993-3_133-2
Chen, W., Wang, R., Wu, R., Tang, L., Fan, J., 2016. Multi-Source and Heterogeneous Data Integration Model for Big Data Analytics in Power DCS, in: 2016 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). Presented at the 2016 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), IEEE, Chengdu, China, pp. 238–242. https://doi.org/10.1109/CyberC.2016.54
Chen, Y.-Z., Lai, Y.-C., 2018. Sparse dynamical Boltzmann machine for reconstructing complex networks with binary dynamics. Physical Review E 97. https://doi.org/10.1103/PhysRevE.97.032317
Chiarello, F., Trivelli, L., Bonaccorsi, A., Fantoni, G., 2018. Extracting and mapping industry 4.0 technologies using wikipedia. Computers in Industry 100, 244–257. https://doi.org/10.1016/j.compind.2018.04.006
Chomkhamsri, K., Wolf, M.-A., Pant, R., 2011. International Reference Life Cycle Data System (ILCD) Handbook: Review Schemes for Life Cycle Assessment, in: Finkbeiner, M. (Ed.), Towards Life Cycle Sustainability Management. Springer Netherlands, Dordrecht, pp. 107–117. https://doi.org/10.1007/978-94-007-1899-9_11
Christiansen, K., 1997. Simplifying LCA: just a cut? / editor: K. Christiansen... Society of Environmental Toxicology and Chemistry - Europe, Brussels, Belgium.
Clarke, B., Fokoue, E., Zhang, H.H., 2009. Principles and theory for data mining and machine learning. Springer Science & Business Media.
Clarke-Sather, A.R., Hutchins, M.J., Zhang, Q., Gershenson, J.K., Sutherland, J.W., 2011. Development of social, environmental, and economic indicators for a small/medium enterprise. Int J Acc & Info Management 19, 247–266. https://doi.org/10.1108/18347641111169250
Cleveland, C.J., Costanza, R., Hall, C.A.S., Kaufmann, R., 1984. Energy and the U.S. Economy: A Biophysical Perspective. Science 225, 890–897. https://doi.org/10.1126/science.225.4665.890
Cleveland, C.J., Kaufmann, R.K., Stern, D.I., 2000. Aggregation and the role of energy in the economy. Ecological Economics 32, 301–317. https://doi.org/10.1016/S0921-8009(99)00113-5
Clifton, D., Amran, A., 2011. The Stakeholder Approach: A Sustainability Perspective. J Bus Ethics 98, 121–136. https://doi.org/10.1007/s10551-010-0538-6
Cole, M.A., Ozgen, C., Strobl, E., 2020. Air Pollution Exposure and COVID-19 Discussion Paper No. 13367, 31. https://doi.org/Available at SSRN: https://ssrn.com/abstract=3628242
Contreras, A.M., Rosa, E., Pérez, M., Van Langenhove, H., Dewulf, J., 2009. Comparative Life Cycle Assessment of four alternatives for using by-products of cane sugar production. Journal of Cleaner Production 17, 772–779. https://doi.org/10.1016/j.jclepro.2008.12.001
Cooper, J.S., 2003. Specifying functional units and reference flows for comparable alternatives. Int J LCA 8, 337–349. https://doi.org/10.1007/BF02978507
Cornelissen, R.L., 1997. Thermodynamics and sustainable development: the use of exergy analysis and the reduction of irreversibility.
Cornelissen, R.L., Hirs, G.G., 2002. The value of the exergetic life cycle assessment besides the LCA. Energy Conversion and Management 43, 1417–1424. https://doi.org/10.1016/S0196-8904(02)00025-0
Corrado, V., Ballarini, I., Dirutigliano, D., Murano, G., 2016. Verification of the New Ministerial Decree about Minimum Requirements for the Energy Performance of Buildings. Energy Procedia, ATI 2016 - 71st Conference of the Italian Thermal Machines Engineering Association 101, 200–207. https://doi.org/10.1016/j.egypro.2016.11.026
Cucurachi, S., van der Giesen, C., Guinée, J., 2018. Ex-ante LCA of Emerging Technologies. Procedia CIRP 69, 463–468. https://doi.org/10.1016/j.procir.2017.11.005
Curran, M.A., 2017. Overview of Goal and Scope Definition in Life Cycle Assessment, in: Curran, M.A. (Ed.), Goal and Scope Definition in Life Cycle Assessment, LCA Compendium – The Complete World of Life Cycle Assessment. Springer Netherlands, Dordrecht, pp. 1–62. https://doi.org/10.1007/978-94-024-0855-3_1

Curran, M.A., 2014. Strengths and Limitations of Life Cycle Assessment, in: Klöpffer, W. (Ed.), Background and Future Prospects in Life Cycle Assessment. Springer Netherlands, Dordrecht, pp. 189–206. https://doi.org/10.1007/978-94-017-8697-3_6

Curran, M.A. (Ed.), 2012. Life cycle assessment handbook: a guide for environmentally sustainable products. Wiley/Scrivener, Hoboken, N.J.

Dahlin, E., 2021. Mind the gap! On the future of AI research. Humanit Soc Sci Commun 8, 1–4. https://doi.org/10.1057/s41599-021-00750-9

Dai, J., Chen, B., Sciubba, E., 2014. Ecological Accounting Based on Extended Exergy: A Sustainability Perspective. Environ. Sci. Technol. 48, 9826–9833. https://doi.org/10.1021/es404191v

Dai, J., Fath, B., Chen, B., 2012. Constructing a network of the social-economic consumption system of China using extended exergy analysis. Renewable and Sustainable Energy Reviews 16, 4796–4808. https://doi.org/10.1016/j.rser.2012.04.027

Dassisti, M., Semeraro, C., Chimenti, M., 2019. Hybrid Exergetic Analysis-LCA approach and the Industry 4.0 paradigm: Assessing Manufacturing Sustainability in an Italian SME. Procedia Manufacturing 33, 655–662. https://doi.org/10.1016/j.promfg.2019.04.082

de Gracia, A., Rincón, L., Castell, A., Jiménez, M., Boer, D., Medrano, M., Cabeza, L.F., 2010. Life Cycle Assessment of the inclusion of phase change materials (PCM) in experimental buildings. Energy and Buildings 42, 1517–1523. https://doi.org/10.1016/j.enbuild.2010.03.022

de Leon, A.R., Chough, K.C., 2013. Analysis of mixed data: An overview, in: Analysis of Mixed Data: Methods & Applications. CRC/Chapman & Hall, p. 12.

de Santoli, L., 2015. Guidelines on energy efficiency of cultural heritage. Energy and Buildings 86, 534–540. https://doi.org/10.1016/j.enbuild.2014.10.050

de Sousa Jabbour, A.B.L., Jabbour, C.J.C., Foropon, C., Godinho Filho, M., 2018. When titans meet – Can industry 4.0 revolutionise the environmentally-sustainable manufacturing wave? The role of critical success factors. Technological Forecasting and Social Change 132, 18–25. https://doi.org/10.1016/j.techfore.2018.01.017

Dealy, J.M., 1980. Handbook of industrial energy analysis, I. Boustead and G. F. Hancock, John Wiley & Sons, New York, 1979, 422 pages. $69.50. AIChE Journal 26, 174–174. https://doi.org/10.1002/aic.690260138

DebRoy, T., Wei, H.L., Zuback, J.S., Mukherjee, T., Elmer, J.W., Milewski, J.O., Beese, A.M., Wilson-Heid, A., De, A., Zhang, W., 2018. Additive manufacturing of metallic components – Process, structure and properties. Progress in Materials Science 92, 112–224. https://doi.org/10.1016/j.pmatsci.2017.10.001

Demirel, Y., Sieniutycz, S., 2003. Nonequilibrium Thermodynamics: Transport and Rate Processes in Physical and Biological Systems. Applied Mechanics Reviews 56, B59–B61. https://doi.org/10.1115/1.1579462

Desgraupes, B., 2017. Clustering Indices.

Dewulf, J., Bösch, M.E., Meester, B.D., Vorst, G.V. der, Langenhove, H.V., Hellweg, S., Huijbregts, M.A.J., 2007. Cumulative Exergy Extraction from the Natural Environment (CEENE): a comprehensive Life Cycle Impact Assessment method for resource accounting. Environ. Sci. Technol. 41, 8477–8483. https://doi.org/10.1021/es0711415

Dewulf, J., Langenhove, H.V., Mulder, J., Berg, M.M.D. van den, Kooi, H.J. van der, Arons, J. de S., 2000. Illustrations towards quantifying the sustainability of technology. Green Chem. 2, 108–114. https://doi.org/10.1039/B000015I

Dewulf, J., Van Langenhove, H., 2002. Assessment of the sustainability of technology by means of a thermodynamically based life cycle analysis. Environmental Science and Pollution Research 9, 267–273. https://doi.org/10.1007/BF02987502

Dietrich, D., Heller, B., Yang, B., EMC Education Services (Eds.), 2015. Data science & big data analytics: discovering, analyzing, visualizing and presenting data. Wiley, Indianapolis, IN.

Dincer, I., Bicer, Y., 2018. 1.27 Life Cycle Assessment of Energy, in: Comprehensive Energy Systems. Elsevier, pp. 1042–1084. https://doi.org/10.1016/B978-0-12-809597-3.00134-6

Dincer, I., Rosen, M.A., 2017. Optimization of Energy Systems, 1st ed. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118894484

Dincer, I., Rosen, M.A., 2013. Exergetic Life Cycle Assessment, in: Exergy. Elsevier, pp. 451–474. https://doi.org/10.1016/B978-0-08-097089-9.00022-X

Ding, S.X., Zhang, P., Jeinsch, T., Ding, E.L., Engel, P., Gui, W., 2011. A survey of the application of basic data-driven and model-based methods in process monitoring and fault diagnosis. IFAC Proceedings Volumes 44, 12380–12388.
Domínguez, A., Czarnowska, L., Valero, Alicia, Stanek, W., Valero, Antonio, 2014. Thermo-ecological and exergy replacement costs of nickel processing. Energy 72, 103–114. https://doi.org/10.1016/j.energy.2014.05.013
Domínguez, E.R., Fernández, M.C., Soria, C.G., Clavijo, D.O., Nodal, C.M., Moya, A.M.C., Morales, M.C., 2011. Exergetic Life Cycle Analysis in the selection of energetic sources for isolated communities in Latin American 10.
Doyle-Kent, M., Kopacek, P., 2021. Do We Need Synchronization of the Human and Robotics to Make Industry 5.0 a Success Story? pp. 302–311. https://doi.org/10.1007/978-3-030-62784-3_25
Doyle-Kent, M., Kopacek, P., 2020. Industry 5.0: Is the Manufacturing Industry on the Cusp of a New Revolution?, in: Durakbasa, N.M., Gençyılmaz, M.G. (Eds.), Proceedings of the International Symposium for Production Research 2019, Lecture Notes in Mechanical Engineering. Springer International Publishing, Cham, pp. 432–441. https://doi.org/10.1007/978-3-030-31343-2_38
Dreyer, L., Hauschild, M., Schierbeck, J., 2006. A Framework for Social Life Cycle Impact Assessment (10 pp). The International Journal of Life Cycle Assessment 11, 88–97. https://doi.org/10.1065/lca2005.08.223
Duflou, J.R., Kellens, K., Renaldi, Dewulf, W., 2011. Environmental Performance of Sheet Metal Working Processes. KEM 473, 21–26. https://doi.org/10.4028/www.scientific.net/KEM.473.21
Durakbasa, N.M., Gençyılmaz, M.G. (Eds.), 2021. Digital Conversion on the Way to Industry 4.0: Selected Papers from ISPR2020, September 24-26, 2020 Online - Turkey, Lecture Notes in Mechanical Engineering. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-62784-3
Eberhardt, L.C.M., Birgisdóttir, H., Birkved, M., 2019. Life cycle assessment of a Danish office building designed for disassembly. Building Research and Information 47, 666–680. https://doi.org/10.1080/09613218.2018.1517458
Ecoinvent [WWW Document], 2019. . Ecoinvent. URL https://www.ecoinvent.org/ (accessed 10.21.19).
Edelen, A., Ingwersen, W.W., 2006. Guidance on Data Quality Assessment for Life Cycle Inventory Data. U.S. Environmental Protection, EPA/600/R-16/096 47.
Ege, A., Sahin, H., 2014. Determination of uncertainties in energy and exergy analysis of a power plant. Energy Conversion and Management 85, 399–406. https://doi.org/10.1016/j.enconman.2014.05.088
Ejsmont, K., Gladysz, B., Kluczek, A., 2020. Impact of Industry 4.0 on Sustainability—Bibliometric Literature Review. Sustainability 12, 5650. https://doi.org/10.3390/su12145650
Ekvall, T., Finnveden, G., 2001. Allocation in ISO 14041—a critical review. Journal of Cleaner Production 9, 197–208. https://doi.org/10.1016/S0959-6526(00)00052-4
Elkington, J., 1998. ACCOUNTING FOR THE TRIPLE BOTTOM LINE. Measuring Business Excellence 2, 18–22. https://doi.org/10.1108/eb025539
Ellen MacArthur Foundation, 2017. New tools support manufacturers to design products for a circular economy [WWW Document]. URL https://www.ellenmacarthurfoundation.org/news/new-tools-support-manufacturers-to-design-products-for-a-circular-economy (accessed 5.4.21).
EPD International [WWW Document], 2020. . EPD International. URL https://www.environdec.com/home (accessed 12.27.21).
Eslami, Y., Lezoche, M., Panetto, H., Dassisti, M., 2020. On analysing sustainability assessment in manufacturing organisations: a survey. International Journal of Production Research 1–32. https://doi.org/10.1080/00207543.2020.1755066
Esmaeilian, B., Behdad, S., Wang, B., 2016. The evolution and future of manufacturing: A review. Journal of Manufacturing Systems 39, 79–100. https://doi.org/10.1016/j.jmsy.2016.03.001
Ester, M., Kriegel, H.-P., Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise 6.
Europan Parliament, European Council, 2018. Directive EU of the European Parliament and of the Council of 30 May 2018 amending Directive 201031EU on the energy performance of buildings and Directive 201227EU on energy efficiency.
European Commission, 2021a. Industry 5.0: towards a sustainable, human centric and resilient European industry. Publications Office, LU.
European Commission, 2021b. Enabling Technologies for Industry 5.0: results of a workshop with Europe's technology leaders. Publications Office, LU.

European Commission, 2007. A lead market initiative for Europe.
European Commission, Joint Research Centre, 2010. ILCD handbook: general guide for life cycle assessment : detailed guidance. Publications Office of the European Union, Luxembourg.
European Committee for Standardization, 2011. EN 15978. Sustainability of construction works – Assessment of environmental performance of buildings – Calculation method;
European Environment Agency, 2020. Air quality in Europe (Publication).
European Environment Agency, 2013. Indoor air quality.
Fahle, S., Prinz, C., Kuhlenkötter, B., 2020. Systematic review on machine learning (ML) methods for manufacturing processes – Identifying artificial intelligence (AI) methods for field application. Procedia CIRP 93, 413–418. https://doi.org/10.1016/j.procir.2020.04.109
Faleh, N., Khila, Z., Wahada, Z., Pons, M.-N., Houas, A., Hajjaji, N., 2018. Exergo-environmental life cycle assessment of biodiesel production from mutton tallow transesterification. Renewable Energy 74–83. https://doi.org/10.1016/j.renene.2018.04.046
Faleh, Nahla, Khila, Z., Wahada, Z., Pons, M.-N., Houas, A., Hajjaji, N., 2018. Exergo-environmental life cycle assessment of biodiesel production from mutton tallow transesterification. Renewable Energy 127, 74–83. https://doi.org/10.1016/j.renene.2018.04.046
Faludi, J., Bayley, C., Bhogal, S., Iribarne, M., 2015. Comparing environmental impacts of additive manufacturing vs traditional machining via life-cycle assessment. Rapid Prototyping Journal 21, 14–33. https://doi.org/10.1108/RPJ-07-2013-0067
Fan, W., 2008. Dependencies revisited for improving data quality, in: Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS ’08. Association for Computing Machinery, New York, NY, USA, pp. 159–170. https://doi.org/10.1145/1376916.1376940
Farooqui, A., Bengtsson, K., Falkman, P., Fabian, M., 2020. Towards data-driven approaches in manufacturing: an architecture to collect sequences of operations. International Journal of Production Research 58, 4947–4963. https://doi.org/10.1080/00207543.2020.1735660
Fava, J.A., Smerek, A., Heinrich, A.B., Morrison, L., 2014. The Role of the Society of Environmental Toxicology and Chemistry (SETAC) in Life Cycle Assessment (LCA) Development and Application, in: Klöpffer, W. (Ed.), Background and Future Prospects in Life Cycle Assessment. Springer Netherlands, Dordrecht, pp. 39–83. https://doi.org/10.1007/978-94-017-8697-3_2
Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases. AIMag 17, 37–37. https://doi.org/10.1609/aimag.v17i3.1230
Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., 1996. Knowledge Discovery and Data Mining: Towards a Unifying Framework., in: KDD. pp. 82–88.
Feynman, R.P., Leighton, R.B., Sands, M.L., 2011. The Feynman lectures on physics, New millennium ed. ed. Basic Books, New York.
Field, F., Kirchain, R., Clark, J., 2000. Life-Cycle Assessment and Temporal Distributions of Emissions: Developing a Fleet-Based Analysis. Journal of Industrial Ecology 4, 71–91. https://doi.org/10.1162/108819800569816
Finnveden, G., Arushanyan, Y., Brandão, M., 2016. Exergy as a Measure of Resource Use in Life Cycle Assessment and Other Sustainability Assessment Tools. Resources 5, 23. https://doi.org/10.3390/resources5030023
Forbes Press, G., 2020. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says [WWW Document]. Forbes. URL https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/ (accessed 1.10.21).
Ford, S., Despeisse, M., 2016. Additive manufacturing and sustainability: an exploratory study of the advantages and challenges. Journal of Cleaner Production 137, 1573–1587. https://doi.org/10.1016/j.jclepro.2016.04.150
Fortes, A.F.C., Carvalho, M., da Silva, J.A.M., 2018. Environmental impact and cost allocations for a dual product heat pump. Energy Conversion and Management 173, 763–772. https://doi.org/10.1016/j.enconman.2018.07.100
Fortes, Anderson Felipe Chaves, Carvalho, M., da Silva, J.A.M., 2018. Environmental impact and cost allocations for a dual product heat pump. Energy Conversion and Management 173, 763–772. https://doi.org/10.1016/j.enconman.2018.07.100
Frechette, S., 2011. Model Based Enterprise for Manufacturing.

Fursova, N., 2018. Methodology for data validation 1.0. European Commission.
Gandiglio, M., De Sario, F., Lanzini, A., Bobba, S., Santarelli, M., Blengini, G.A., 2019. Life Cycle Assessment of a Biogas-Fed Solid Oxide Fuel Cell (SOFC) Integrated in a Wastewater Treatment Plant. Energies 12, 1611. https://doi.org/10.3390/en12091611
Gao, J., Zhang, Q., Wang, X., Song, D., Liu, Weiqi, Liu, Wenchao, 2018. Exergy and exergoeconomic analyses with modeling for CO2 allocation of coal-fired CHP plants. Energy 152, 562–575. https://doi.org/10.1016/j.energy.2018.03.171
García-Muiña, F.E., Medina-Salgado, M.S., Ferrari, A.M., Cucchi, M., 2020. Sustainability Transition in Industry 4.0 and Smart Manufacturing with the Triple-Layered Business Model Canvas. Sustainability 12, 2364. https://doi.org/10.3390/su12062364
Gaudreau, K., Fraser, R.A., Murphy, S., 2012. The Characteristics of the Exergy Reference Environment and Its Implications for Sustainability-Based Decision-Making. Energies 5, 2197–2213. https://doi.org/10.3390/en5072197
Gaudreau, K., Fraser, R.A., Murphy, S., 2009. The Tenuous Use of Exergy as a Measure of Resource Value or Waste Impact. Sustainability 1, 1444–1463. https://doi.org/10.3390/su1041444
Geisler, G., Hellweg, S., Hungerbühler, K., 2005. Uncertainty Analysis in Life Cycle Assessment (LCA): Case Study on Plant-Protection Products and Implications for Decision Making. The International Journal of Life Cycle Assessment 10, 184–192. https://doi.org/10.3929/ethz-b-000032782
General (US), O. of the S., 2009. The Connection Between Health and Homes, The Surgeon General’s Call to Action to Promote Healthy Homes. Office of the Surgeon General (US).
Ghobakhloo, M., 2020. Industry 4.0, digitization, and opportunities for sustainability. Journal of Cleaner Production 252, 119869. https://doi.org/10.1016/j.jclepro.2019.119869
Ghobakhloo, M., Fathi, M., 2021. Industry 4.0 and opportunities for energy sustainability. Journal of Cleaner Production 295, 126427. https://doi.org/10.1016/j.jclepro.2021.126427
Ghosh, K., Ng, Y.S., Srinivasan, R., 2011. Evaluation of decision fusion strategies for effective collaboration among heterogeneous fault diagnostic methods. Computers & chemical engineering 35, 342–355.
Giovannini, A., 2015. A knowledge representation framework for the design and the evaluation of a product variety (Theses). Université de Lorraine.
Gizzi, F.T., Sileo, M., Biscione, M., Danese, M., Alvarez de Buergo, M., 2016. The conservation state of the Sassi of Matera site (Southern Italy) and its correlation with the environmental conditions analysed through spatial analysis techniques. Journal of Cultural Heritage 17, 61–74. https://doi.org/10.1016/j.culher.2015.05.002
Gladyshev, G.P., 2015. Thermodynamics of the origin of life, evolution, and aging. Advances in Gerontology 5, 55–58. https://doi.org/10.1134/S2079057015020034
Gobbo, J.A., Busso, C.M., Gobbo, S.C.O., Carreão, H., 2018. Making the links among environmental protection, process safety, and industry 4.0. Process Safety and Environmental Protection 117, 372–382. https://doi.org/10.1016/j.psep.2018.05.017
Goedkoop, M., Oele, M., Leijting, J., Ponsioen, T., 2016. Introduction to LCA with SimaPro. URL https://www.pre-sustainability.com/download/SimaPro8IntroductionToLCA.pdf.
Goldstein, B., Rasmussen, F.N., 2017. LCA of Buildings and the Built Environment, in: Life Cycle Assessment: Theory and Practice. pp. 695–722. https://doi.org/10.1007/978-3-319-56475-3_28
Gong, M., Wall, G., 2001. On exergy and sustainable development—Part 2: Indicators and methods. Exergy, An International Journal 1, 217–233. https://doi.org/10.1016/S1164-0235(01)00030-9
González, A., Sala, J.M., Flores, I., López, L.M., 2003. Application of thermoeconomics to the allocation of environmental loads in the life cycle assessment of cogeneration plants. Energy 28, 557–574. https://doi.org/10.1016/S0360-5442(02)00156-1
Grisolia, G., Lucia, U., 2019. Biocarburanti e sostenibilit`a: Analisi exergo-economica.
Groniewsky, A., 2013. Exergoeconomic Analysis of Thermodynamic Systems 6.
Guarino, S., 2019. Environmental assessment of Selective Laser Melting compared with Laser Cutting of 316L stainless steel: A case study for flat washers production. CIRP Journal of Manufacturing Science and Technology 605, 14. https://doi.org/10.1016/j.cirpj.2020.08.004
Guinée, J.B., Heijungs, R., Huppes, G., Zamagni, A., Masoni, P., Buonamici, R., Ekvall, T., Rydberg, T., 2011. Life Cycle Assessment: Past, Present, and Future. Environ. Sci. Technol. 45, 90–96. https://doi.org/10.1021/es101316v
Gulotta, T., Guarino, F., Mistretta, M., Cellura, M., Lorenzini, G., 2018. Introducing exergy analysis in life cycle assessment: A case study. MMEP 5, 139–145. https://doi.org/10.18280/mmep.050302

Gulotta, T.M., Guarino, F., Mistretta, M., Cellura, M., Lorenzini, G., 2018. Introducing exergy analysis in life cycle assessment: A case study. Mathematical Modelling of Engineering Problems 5, 139–145. https://doi.org/10.18280/mmep.050302

Guo, M., Murphy, R.J., 2012. LCA data quality: Sensitivity and uncertainty analysis. Science of The Total Environment 435–436, 230–243. https://doi.org/10.1016/j.scitotenv.2012.07.006

Gupta, D., 2020. Activation Functions | Fundamentals Of Deep Learning. Analytics Vidhya. URL https://www.analyticsvidhya.com/blog/2020/01/fundamentals-deep-learning-activation-functions-when-to-use-them/ (accessed 9.9.21).

Gupta, K. (Ed.), 2017. Advanced Manufacturing Technologies, Materials Forming, Machining and Tribology. Springer International Publishing, Cham.

Haes, U. de, Finnveden, G., Goedkoop, G., Hauschild, M., Jolliet, P., Kloffer, O., Krewitt, W., Lindeijer, W., Muller-Wenk, E., Olsen, R., 2002. Life Cycle Impact Assessment: Striving towards best practice, in: Proceedings. SETAC Press.

Hájek, P., Havel, I., Chytil, M., 1966. The GUHA method of automatic hypotheses determination. Computing. https://doi.org/10.1007/BF02345483

Hammond, G.P., 2004. Engineering Sustainability: Thermodynamics, Energy Systems and the Environment, in: Winnett, A. (Ed.), Towards an Environment Research Agenda: A Third Selection of Papers. Palgrave Macmillan UK, London, pp. 175–210. https://doi.org/10.1057/9780230554429_8

Hamut, H.S., Dincer, I., Naterer, G.F., 2014. Exergoenvironmental analysis of hybrid electric vehicle thermal management systems. Journal of Cleaner Production 67, 187–196. https://doi.org/10.1016/j.jclepro.2013.12.041

Harding, J., Shahbaz, M., Srinivas, Kusiak, A., 2006. Data Mining in Manufacturing: A Review. https://doi.org/10.1115/1.2194554

Hau, J.L., Bakshi, B.R., 2004. Expanding Exergy Analysis to Account for Ecosystem Products and Services. Environ. Sci. Technol. 38, 3768–3777. https://doi.org/10.1021/es034513s

Hauschild, M.Z., Bonou, A., Olsen, S.I., 2018. Life Cycle Interpretation, in: Hauschild, M.Z., Rosenbaum, R.K., Olsen, S.I. (Eds.), Life Cycle Assessment: Theory and Practice. Springer International Publishing, Cham, pp. 323–334. https://doi.org/10.1007/978-3-319-56475-3_12

He, D., Li, R., Zhu, J., 2013. Plastic Bearing Fault Diagnosis Based on a Two-Step Data Mining Approach. IEEE Transactions on Industrial Electronics 60, 3429–3440. https://doi.org/10.1109/TIE.2012.2192894

Heijungs, R., Sun, S., 2002. The computational structure of life cycle assessment. The International Journal of Life Cycle Assessment 7. https://doi.org/10.1007/BF02978899

Heink, U., Kowarik, I., 2010. What are indicators? On the definition of indicators in ecology and environmental planning. Ecological Indicators 10, 584–593. https://doi.org/10.1016/j.ecolind.2009.09.009

Hellweg, S., Hofstetter, T.B., Hungerbühler, K., 2005. Time-dependent life-cycle assessment of slag landfills with the help of scenario analysis: the example of Cd and Cu. Journal of Cleaner Production 13, 301–320. https://doi.org/10.1016/j.jclepro.2004.02.016

Hellweg, S., Hofstetter, T.B., Hungerbühler, K., 2001. Modeling Waste Incineration for Life-Cycle Inventory Analysis in Switzerland 17.

Hellweg, S., Mila i Canals, L., 2014. Emerging approaches, challenges and opportunities in life cycle assessment. Science 344, 1109–1113. https://doi.org/10.1126/science.1248361

Hermann, M., Pentek, T., Otto, B., 2016. Design principles for industrie 4.0 scenarios, in: System Sciences (HICSS), 2016 49th Hawaii International Conference On. IEEE, pp. 3928–3937.

Hernandez, A.G., Cullen, J.M., 2019. Exergy: A universal metric for measuring resource efficiency to address industrial decarbonisation. Sustainable Production and Consumption 20, 151–164. https://doi.org/10.1016/j.spc.2019.05.006

Hernández Korner, M., Lambán, M., Albajez, J., Santolaria, J., Ng Corrales, L.D.C., Royo, J., 2020. Systematic Literature Review: Integration of Additive Manufacturing and Industry 4.0. Metals 10, 1061. https://doi.org/10.3390/met10081061

Heterogeneity. A Dictionary of the English Language - Samuel Johnson [WWW Document], 2017. URL https://johnsonsdictionaryonline.com/page-view/?i=996 (accessed 1.21.20).

Hibbert, L., 2020. Industry 5.0 Brings Humans Back Into the Loop [WWW Document]. Mouser Elecronics. URL https://eu.mouser.com/applications/industry-5-brings-humans-back/ (accessed 12.20.21).

Higgins, J.P.T., Thomas, J., 2019. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons.

Hischier, R., Reichart, I., 2003. Multifunctional electronic media-traditional media: The Problem of an Adequate Functional Unit A case study of a printed newspaper, an internet newspaper and a TV broadcast. Int J LCA 8, 201–208. https://doi.org/10.1007/BF02978472
Hong, P., Jagani, S., Kim, J., Youn, S.H., 2019. Managing sustainability orientation: An empirical investigation of manufacturing firms. International Journal of Production Economics 211, 71–81. https://doi.org/10.1016/j.ijpe.2019.01.035
Howard, M.C., 2011. Sustainable Manufacturing Initiative (SMI):
Hu, Y.-S., 2020. The impact of increasing returns on knowledge and big data: from Adam Smith and Allyn Young to the age of machine learning and digital platforms. Prometheus 36, 10–29.
Huang, X., Wu, L., Ye, Y., 2019. A Review on Dimensionality Reduction Techniques. Int. J. Patt. Recogn. Artif. Intell. 33, 1950017. https://doi.org/10.1142/S0218001419500174
Huysveld, S., Schaubroeck, T., De Meester, S., Sorgeloos, P., Van Langenhove, H., Van linden, V., Dewulf, J., 2013. Resource use analysis of Pangasius aquaculture in the Mekong Delta in Vietnam using Exergetic Life Cycle Assessment. Journal of Cleaner Production 51, 225–233. https://doi.org/10.1016/j.jclepro.2013.01.024
Illa, P., Padhi, N., 2018. Practical Guide to Smart Factory Transition Using IoT, Big Data and Edge Analytics. IEEE Access PP, 1–1. https://doi.org/10.1109/ACCESS.2018.2872799
IPCC, 2006. Guidelines for National Greenhouse Gas Inventories, Industrial Processes and Product Use. ed. IPCC.
ISO/TR 14047. Environmental management - Life cycle impact assessment - Examples of application of ISO 14042 [WWW Document], 2003a. . ISO. URL https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/03/10/31026.html (accessed 9.28.20).
ISO/TR 14049. Environmental management - Life cycle assessment - Examples of application of ISO 14041 to goal and scope definition and inventory analysis [WWW Document], 2000b. . ISO. URL https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/02/98/29834.html (accessed 9.28.20).
ISO/TS 14048. Environmental management - Life cycle assessment - Data documentation format [WWW Document], 2002c. . ISO. URL https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/02/98/29872.html (accessed 9.28.20).
Istituto Superiore di Sanità, 2014. Rapporti ISTISAN. La qualità dell’aria indoor: attuale situazione nazionale e comunitaria. L’esperienza del Gruppo di Studio Nazionale Inquinamento Indoor.
Istituto Superiore di Sanità: Benvenuti [WWW Document], n.d. URL http://old.iss.it/indo/index.php?lang=1&anno=2020&tipo=3 (accessed 6.22.20).
Javadi, M.A., Hoseinzadeh, S., Ghasemiasl, R., Heyns, P.S., Chamkha, A.J., 2020. Sensitivity analysis of combined cycle parameters on exergy, economic, and environmental of a power plant. J Therm Anal Calorim 139, 519–525. https://doi.org/10.1007/s10973-019-08399-y
Jayal, A.D., Badurdeen, F., Dillon, O.W., Jawahir, I.S., 2010. Sustainable manufacturing: Modeling and optimization challenges at the product, process and system levels. CIRP Journal of Manufacturing Science and Technology 2, 144–152. https://doi.org/10.1016/j.cirpj.2010.03.006
Jiang, J., Ma, Y., 2020. Path Planning Strategies to Optimize Accuracy, Quality, Build Time and Material Use in Additive Manufacturing: A Review. Micromachines 11, 633. https://doi.org/10.3390/mi11070633
Jiang, Q., Liu, Z., Li, T., Cong, W., Zhang, H.-C., 2019. Emergy-based life-cycle assessment (Em-LCA) for sustainability assessment: a case study of laser additive manufacturing versus CNC machining. Int J Adv Manuf Technol 102, 4109–4120. https://doi.org/10.1007/s00170-019-03486-8
Jimenez, J.J.M., 2020. Towards multi-model approaches to predictive maintenance_ A systematic literature survey on diagnostics and prognostics. Journal of Manufacturing Systems 19.
Jirkovsky, V., Obitko, M., 2014. Semantic Heterogeneity Reduction for Big Data in Industrial Automation. CEUR Workshop Proceedings 1214, 10.
Jo, J., Jo, B., Kim, J., Kim, S., Han, W., 2020. Development of an IoT-Based Indoor Air Quality Monitoring Platform. Journal of Sensors 2020, 1–14. https://doi.org/10.1155/2020/8749764
Jolliet, O., Saade-Sbeih, M., Shaked, S., Jolliet, A., Crettaz, P., 2015. Environmental Life Cycle Assessment. CRC Press.
Jönsson, Å., 2000. Is it feasible to address indoor climate issues in LCA? Environmental Impact Assessment Review 20, 241–259. https://doi.org/10.1016/S0195-9255(99)00039-6

Joung, C.B., Carrell, J., Sarkar, P., Feng, S.C., 2013. Categorization of indicators for sustainable manufacturing. Ecological Indicators 24, 148–157. https://doi.org/10.1016/j.ecolind.2012.05.030

Junior, A.A. de S., Pio, J.L. de S., Fonseca, J.C., Oliveira, M.A. de, Valadares, O.C. de P., Silva, P.H.S. da, 2021. The State of Cybersecurity in Smart Manufacturing Systems: A Systematic Review. European Journal of Business and Management Research 6, 188–194. https://doi.org/10.24018/ejbmr.2021.6.6.1173

Kamble, S.S., Gunasekaran, A., Gawankar, S.A., 2018. Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives. Process Safety and Environmental Protection 117, 408–425. https://doi.org/10.1016/j.psep.2018.05.009

Kamps, T., Lutter-Guenther, M., Seidel, C., Gutowski, T., Reinhart, G., 2018. Cost- and energy-efficient manufacture of gears by laser beam melting. CIRP Journal of Manufacturing Science and Technology 21, 47–60. https://doi.org/10.1016/j.cirpj.2018.01.002

Kang, H.S., Lee, J.Y., Choi, S., Kim, H., Park, J.H., Son, J.Y., Kim, B.H., Noh, S.D., 2016. Smart manufacturing: Past research, present findings, and future directions. Int. J. of Precis. Eng. and Manuf.-Green Tech. 3, 111–128. https://doi.org/10.1007/s40684-016-0015-5

Karmaker Santu, S.K., Hassan, Md.M., Smith, M., Zhai, C., Veeramachaneni, K., 2020. AutoML to Date and Beyond: Challenges and Opportunities.

Katal, A., Wazid, M., Goudar, R.H., 2013. Big Data: Issues, Challenges, Tools and Good Practices, in: In Proceedings of the 6th International Conference on Contemporary Computing (IC3). Noida, August 8th-10th 2013. US: IEEE.

Kellens, K., Baumers, M., Gutowski, T.G., Flanagan, W., Lifset, R., Duflou, J.R., 2017. Environmental Dimensions of Additive Manufacturing: Mapping Application Domains and Their Environmental Implications: Environmental Dimensions of Additive Manufacturing. Journal of Industrial Ecology 21, S49–S68. https://doi.org/10.1111/jiec.12629

Khan, M., Wu, X., Xu, X., Dou, W., 2017. Big data challenges and opportunities in the hype of Industry 4.0, in: 2017 IEEE International Conference on Communications (ICC). Presented at the ICC 2017 - 2017 IEEE International Conference on Communications, IEEE, Paris, France, pp. 1–6. https://doi.org/10.1109/ICC.2017.7996801

Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A., Salehian, S., 2018. The 10 Vs, Issues and Challenges of Big Data. https://doi.org/10.1145/3206157.3206166

Khanali, M., Shahvarooghi Farahani, S., Shojaei, H., Elhami, B., 2017. Life cycle environmental impacts of saffron production in Iran. Environmental Science and Pollution Research 24, 4812–4821. https://doi.org/10.1007/s11356-016-8228-2

Khanfar, A.A.A., Iranmanesh, M., Ghobakhloo, M., Senali, M.G., Fathi, M., 2021. Applications of Blockchain Technology in Sustainable Manufacturing and Supply Chain Management: A Systematic Review. Sustainability 13, 7870. https://doi.org/10.3390/su13147870

Kiel, D., Müller, J.M., Arnold, C., Voigt, K.-I., 2017. Sustainable Industrial Value Creation: Benefits and Challenges of Industry 4.0. Int. J. Innov. Mgt. 21, 1740015. https://doi.org/10.1142/S1363919617400151

Klaus, P., 2015. The Devil Is in the Details – Only What Get Measured Gets Managed, in: Klaus, P. (Ed.), Measuring Customer Experience: How to Develop and Execute the Most Profitable Customer Experience Strategies. Palgrave Macmillan UK, London, pp. 81–101. https://doi.org/10.1057/9781137375469_7

Komputer, F., Presiden, Barat, J., 2019. A Comparison of Machine Learning Algorithms in Manufacturing Production Process [WWW Document]. URL https://www.semanticscholar.org/paper/A-Comparison-of-Machine-Learning-Algorithms-in-Komputer-Presiden/1e0ac9c84511945043407d83244f848565fc0f4a (accessed 1.9.22).

Koroneos, C., Stylos, N., 2014. Exergetic LCA of a Grid Connected, Polycrystalline Silicon Photovoltaic System 16.

Kotas, T.J., 2013. The Exergy Method of Thermal Plant Analysis. Elsevier.

Kuestenmacher, A., Plöger, P.G., 2016. Model-Based Fault Diagnosis Techniques for Mobile Robots**This work was sponsored by the B-IT foundation and the Strukturfond des Landes Nordrhein-Westfalen for the female PhD students. IFAC-PapersOnLine 49, 50–56. https://doi.org/10.1016/j.ifacol.2016.07.613

Kumanyika, S.K., Parker, L., Sim, L.J., 2010. Defining the Problem: The Importance of Taking a Systems Perspective, in: Bridging the Evidence Gap in Obesity Prevention: A Framework to Inform Decision

Making. National Academies Press (US), Washington (DC), pp. 71–88. https://doi.org/10.17226/12847
Kumar, N.M., Mallick, P.K., 2018. The Internet of Things: Insights into the building blocks, component interactions, and architecture layers. Procedia Computer Science 132, 109–117. https://doi.org/10.1016/j.procs.2018.05.170
Kusiak, A., 2020. Service manufacturing = Process-as-a-Service + Manufacturing Operations-as-a-Service. J Intell Manuf 31, 1–2. https://doi.org/10.1007/s10845-019-01527-3
Kusiak, A., 2018. Smart manufacturing. International Journal of Production Research 56, 508–517. https://doi.org/10.1080/00207543.2017.1351644
Kylili, A., Fokaides, P.A., Vaiciunas, J., Seduikyte, L., 2016. Integration of Building Information Modelling (BIM) and Life Cycle Assessment (LCA) for sustainable constructions. Journal of Sustainable Architecture and Civil Engineering 13, 28–38. https://doi.org/10.5755/j01.sace.13.4.12862
L'Abbate, P., 2018. Environmental analysis of polyester fabric for ticking. Journal of Cleaner Production 172, 8. https://doi.org/10.1016/j.jclepro.2017.10.045
Lao, L., Ellis, M., Christofides, P.D., 2014. Smart manufacturing: Handling preventive actuator maintenance and economics using model predictive control. AIChE J. 60, 2179–2196. https://doi.org/10.1002/aic.14427
Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., Hoffman, M., 2014. Industrie 4.0. WIRTSCHAFTSINFORMATIK 56, 261–264. https://doi.org/10.1007/s11576-014-0424-4
Laurenti, M., Trentin, M., 2021. Ecosistema Urbano. Rapporto sulle performance ambientali delle città.
Lee, H., Cho, S., 2006. The Novelty Detection Approach for Different Degrees of Class Imbalance, in: King, I., Wang, J., Chan, L.-W., Wang, D. (Eds.), Neural Information Processing, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 21–30. https://doi.org/10.1007/11893257_3
Lee, J., Bagheri, B., Kao, H.-A., 2015. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters 3, 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001
Lee, J.Y., Yoon, J.S., Kim, B.-H., 2017. A big data analytics platform for smart factories in small and medium-sized manufacturing enterprises: An empirical case study of a die casting factory. Int. J. Precis. Eng. Manuf. 18, 1353–1361. https://doi.org/10.1007/s12541-017-0161-x
Legislative Decree 152/2006, art.268-11 [WWW Document], 2006. URL https://www.camera.it/parlam/leggi/deleghe/06152dl5.htm#267 (accessed 11.8.21).
Lettieri, D.J., Hannemann, C.R., Carey, V.P., Shah, A.J., 2009. Lifetime exergy consumption as a sustainability metric for information technologies, in: 2009 IEEE International Symposium on Sustainable Systems and Technology. Presented at the 2009 IEEE International Symposium on Sustainable Systems and Technology (ISSST), IEEE, Tempe, AZ, USA, pp. 1–6. https://doi.org/10.1109/ISSST.2009.5156698
Li, D., Wang, R., 2009. Hybrid Emergy-LCA (HEML) based metabolic evaluation of urban residential areas: The case of Beijing, China. Ecological Complexity 6, 484–493. https://doi.org/10.1016/j.ecocom.2009.09.002
Liao, M., Ma, C., Yao, D., Liu, H., 2013. Decomposition of embodied exergy flows in manufactured products and implications for carbon tariff policies. Asia Eur J 11, 265–283. https://doi.org/10.1007/s10308-013-0357-3
Lima, R.S., de Azevedo Caldeira-Pires, A., Cardoso, A.N., 2020. Uncertainty Analysis in Life Cycle Assessments Applied to Biorefineries Systems: A Critical Review of the Literature. Process Integr Optim Sustain 4, 1–13. https://doi.org/10.1007/s41660-019-00103-9
Liu, G., Müller, D.B., 2012. Addressing sustainability in the aluminum industry: a critical review of life cycle assessments. Journal of Cleaner Production 35, 108–117. https://doi.org/10.1016/j.jclepro.2012.05.030
Liu, W., Tang, R., Peng, T., 2018. An IoT-enabled Approach for Energy Monitoring and Analysis of Die Casting Machines. Procedia CIRP 69, 656–661. https://doi.org/10.1016/j.procir.2017.11.109
Longo, F., Padovano, A., Umbrello, S., 2020. Value-Oriented and Ethical Technology Engineering in Industry 5.0: A Human-Centric Perspective for the Design of the Factory of the Future. Applied Sciences 10, 4182. https://doi.org/10.3390/app10124182
Lozano, R., 2008. Envisioning sustainability three-dimensionally. Journal of Cleaner Production 16, 1838–1846. https://doi.org/10.1016/j.jclepro.2008.02.008

Lucia, U., Grisolia, G., 2019. Exergy inefficiency: An indicator for sustainable development analysis. Energy Reports 5, 62–69. https://doi.org/10.1016/j.egyr.2018.12.001

Madhikermi, M., Buda, A., Dave, B., Framling, K., 2017. Key data quality pitfalls for condition based maintenance, in: 2017 2nd International Conference on System Reliability and Safety (ICSRS). Presented at the 2017 2nd International Conference on System Reliability and Safety (ICSRS), IEEE, Milan, pp. 474–480. https://doi.org/10.1109/ICSRS.2017.8272868

Majeau-Bettez, G., Strømman, A.H., Hertwich, E.G., 2011. Evaluation of Process- and Input–Output-based Life Cycle Inventory Data with Regard to Truncation and Aggregation Issues. Environ. Sci. Technol. 45, 10170–10177. https://doi.org/10.1021/es201308x

Mandreoli, F., Montangero, M., 2019. Dealing With Data Heterogeneity in a Data Fusion Perspective, in: Data Handling in Science and Technology. Elsevier, pp. 235–270. https://doi.org/10.1016/B978-0-444-63984-4.00009-0

Martinás, K., 1997. Thermodynamics and sustainability a new approach by Extropy. Periodica Polytechnica Ser. Chem. Eng., Periodica Polytechnica Ser. Chem. Eng. 12.

Martinas, K., Frankowicz, M., 2001. Extropy – reformulation of the entropy principle. Periodica Polytechnica Ser. Chem. Eng. 44, 10.

Martinás, K., Grandpierre, A., 2007. Thermodynamic Measure for Nonequilibrium Processes. Interdisciplinary Description of Complex Systems 5, 1–13.

Master Italy, 2018. I certificati di qualità e garanzia dell'azienda Master [WWW Document]. Master Italy. URL https://www.masteritaly.com/certificazioni-azienda-master-italy/ (accessed 11.28.19).

Matheson, E., Minto, R., Zampieri, E.G.G., Faccio, M., Rosati, G., 2019. Human–Robot Collaboration in Manufacturing Applications: A Review. Robotics 8, 100. https://doi.org/10.3390/robotics8040100

Mathieux, F., Recchioni, M., Schau, E.M., Goralczyk, M., European Commission, Joint Research Centre, Institute for Environment and Sustainability, 2013. ILCD data network and ELCD database current use and further needs for supporting environmental footprint and life cycle indicator projects. Publications Office, Luxembourg.

MathWorks, 2020a. What is linear regression? [WWW Document]. MathWorks. URL https://uk.mathworks.com/discovery/linear-regression.html (accessed 1.9.21).

MathWorks, 2020b. Anomaly Detection [WWW Document]. URL https://uk.mathworks.com/discovery/anomaly-detection.html (accessed 8.29.20).

MathWorks, 2019. Design MPC Controller in Simulink - MATLAB [WWW Document]. URL https://uk.mathworks.com/help/mpc/gs/designing-a-model-predictive-controller-for-a-simulink-plant.html (accessed 5.11.19).

Mazzarisi, M., Campanelli, S.L., Angelastro, A., Dassisti, M., 2020a. Phenomenological modelling of direct laser metal deposition for single tracks. Int J Adv Manuf Technol 111, 1955–1970. https://doi.org/10.1007/s00170-020-06204-x

Mazzarisi, M., Campanelli, S.L., Angelastro, A., Palano, F., Dassisti, M., 2020b. In situ monitoring of direct laser metal deposition of a nickel-based superalloy using infrared thermography. Int J Adv Manuf Technol. https://doi.org/10.1007/s00170-020-06344-0

McGoff, K., Mukherjee, S., Pillai, N.S., 2012. Statistical inference for dynamical systems: a review. arXiv:1204.6265 [math, stat].

McKinsey and Company, 2016. The need to lead in data and analytics [WWW Document]. URL http://www.mckinsey.com/businessfunctions/digital-mckinsey/our-insights/the-need-to-lead-in-data-and-analytics

Medyna, G., Coatanéa, E., Millet, D., 2009a. Comparative Study of Environmental Evaluation Assessment Using Exergetic LCA Implemented in Existing Software and a Novel Exergetic Approach during the Early Design Phase [WWW Document]. DS 58-4: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 4, Product and Systems Design, Palo Alto, CA, USA, 24.-27.08.2009. URL https://www.designsociety.org/publication/28667/Comparative+Study+of+Environmental+Evaluation+Assessment+Using+Exergetic+LCA+Implemented+in+Existing+Software+and+a+Novel+Exergetic+Approach+during+the+Early+Design+Phase (accessed 7.27.19).

Medyna, G., Nordlund, H., Coatanea, E., 2009b. Study of an exergy method for environmental evaluation assessment in the early design phase using comparative LCA and exergy approach. IJDE 2, 320. https://doi.org/10.1504/IJDE.2009.030178

Medyna, G., Nordlund, H., Coatanea, E., 2009c. Study of an exergy method for environmental evaluation assessment in the early design phase using comparative LCA and exergy approach. IJDE 2, 320. https://doi.org/10.1504/IJDE.2009.030178
Mehmeti, A., Pedro Pérez-Trujillo, J., Elizalde-Blancas, F., Angelis-Dimakis, A., McPhail, S.J., 2018. Exergetic, environmental and economic sustainability assessment of stationary Molten Carbonate Fuel Cells. Energy Conversion and Management 168, 276–287. https://doi.org/10.1016/j.enconman.2018.04.095
Meixner, J., 1973. The entropy problem in thermodynamics of processes. Rheologica Acta 12, 465–467. https://doi.org/10.1007/BF01503000
Mejia, L.A., Toledo, C.E.E., Rayle, B.M.R., 2012. Decision Making in Sustainable Development: Some Methods to Evaluate Energy and Nonrenewable Resources Waste When Using Some Plastics. AJOR 02, 399–407. https://doi.org/10.4236/ajor.2012.23048
Meteyer, S., Xu, X., Perry, N., Zhao, Y.F., 2014. Energy and Material Flow Analysis of Binder-jetting Additive Manufacturing Processes. Procedia CIRP 15, 19–25. https://doi.org/10.1016/j.procir.2014.06.030
Meyer, L., Tsatsaronis, G., Buchgeister, J., Schebek, L., 2009. Exergoenvironmental analysis for evaluation of the environmental impact of energy conversion systems. Energy 34, 75–89. https://doi.org/10.1016/j.energy.2008.07.018
Meyer, L., Tsatsaronis, G., Buchgeister, J., Schebek, L., 2008. Exergoenvironmental analysis for evaluation of the environmental impact of energy conversion systems. Energy 34, 75–89. https://doi.org/10.1016/j.energy.2008.07.018
Mikut, R., Reischl, M., 2011. Data mining tools. WIREs Data Mining and Knowledge Discovery 1, 431–443. https://doi.org/10.1002/widm.24
Milanovic, B., Agarski, B., Vukelic, D., Budak, I., Kiss, F., 2017. Comparative exergy-based life cycle assessment of conventional and hybrid base transmitter stations. Journal of Cleaner Production 167, 610–618. https://doi.org/10.1016/j.jclepro.2017.08.176
Mills, N., 2009. End-to-End automated analytics for CBM+. Presented at the Structural Health Monitoring 2009: From System Integration to Autonomous Systems - Proceedings of the 7th International Workshop on Structural Health Monitoring, IWSHM 2009, pp. 427–436.
Ministero della Salute, 2020. Piano Nazionale della Prevenzione 2020-2025.
Ministero della Salute, 2014. Piano Nazionale della Prevenzione 2014-2018.
Misopoulos, F., Michaelides, R., Salehuddin, M.A., Manthou, V., Michaelides, Z., 2018. Addressing Organisational Pressures as Drivers towards Sustainability in Manufacturing Projects and Project Management Methodologies. Sustainability 10, 2098. https://doi.org/10.3390/su10062098
Moktadir, Md.A., Ali, S.M., Kusi-Sarpong, S., Shaikh, Md.A.A., 2018. Assessing challenges for implementing Industry 4.0: Implications for process safety and environmental protection. Process Safety and Environmental Protection 117, 730–741. https://doi.org/10.1016/j.psep.2018.04.020
Montáns, F.J., Chinesta, F., Gómez-Bombarelli, R., Kutz, J.N., 2019. Data-driven modeling and learning in science and engineering. Comptes Rendus Mécanique 347, 845–855. https://doi.org/10.1016/j.crme.2019.11.009
Montazerinejad, H., Ahmadi, P., Montazerinejad, Z., 2019. Advanced exergy, exergo-economic and exrgo-environmental analyses of a solar based trigeneration energy system. Applied Thermal Engineering 152, 666–685. https://doi.org/10.1016/j.applthermaleng.2019.01.040
Morawska, L., Milton, D.K., 2020. It Is Time to Address Airborne Transmission of Coronavirus Disease 2019 (COVID-19). Clinical Infectious Diseases 71, 2311–2313. https://doi.org/10.1093/cid/ciaa939
Morosuk, T., Tsatsaronis, G., 2014. Exergoenvironmental Analysis is a New Tool for Evaluation of an Energy Conversion System. Енергетика і автоматика 4, 12.
Morosuk, T., Tsatsaronis, G., 2012. 3-D Exergy-Based Methods for Improving Energy-Conversion Systems. International Journal of Thermodynamics 15, 201–213.
Moslehi, S., Reddy, T.A., 2019. An LCA methodology to assess location-specific environmental externalities of integrated energy systems. Sustainable Cities and Society 46. https://doi.org/10.1016/j.scs.2019.101425
Moya, C., Domínguez, R., Van Langenhove, H., Herrero, S., Gil, P., Ledón, C., Dewulf, J., 2013. Exergetic analysis in cane sugar production in combination with Life Cycle Assessment. Journal of Cleaner Production 59, 43–50. https://doi.org/10.1016/j.jclepro.2013.06.028
Mrugalska, B., Ahmed, J., 2021. Organizational Agility in Industry 4.0: A Systematic Literature Review 23.

Mster Italy, 2019. Master progetta e realizza accessori per serramenti in alluminio [WWW Document]. Master Italy. URL https://www.masteritaly.com/chi-siamo/ (accessed 11.28.19).

Müller, A., Wambach, K., Alsema, E., 2005. Life Cycle Analysis of Solar Module Recycling Process. MRS Online Proceedings Library Archive 895. https://doi.org/10.1557/PROC-0895-G03-07

Müller, E., Hopf, H., 2017. Competence Center for the Digital Transformation in Small and Medium-Sized Enterprises. Procedia Manufacturing 11, 1495–1500. https://doi.org/10.1016/j.promfg.2017.07.281

Nagarajan, H.P.N., Haapala, K.R., 2018. Characterizing the influence of resource-energy-exergy factors on the environmental performance of additive manufacturing systems. Journal of Manufacturing Systems 48, 87–96. https://doi.org/10.1016/j.jmsy.2018.06.005

Nagorny, K., Monteiro, P., Barata, J., Colombo, A., 2017. Big Data Analysis in Smart Manufacturing: A Review. International Journal of Communications, Network and System Sciences 10, 31–58. https://doi.org/10.4236/ijcns.2017.103003

Naz, A., Javed, M.U., Javaid, N., Saba, T., Alhussein, M., Aurangzeb, K., 2019. Short-Term Electric Load and Price Forecasting Using Enhanced Extreme Learning Machine Optimization in Smart Grids. Energies 12, 866. https://doi.org/10.3390/en12050866

Niccolucci, V., Rugani, B., Bastianoni, S., 2009. Life Cycle Assessment (LCA) combined with EMergy evaluation for a better understanding of the environmental aspects associated with a crystal glass supply chain, in: Ecosystems and Sustainable Development. Presented at the ECOSUD 2009, Chianciano Terme, Italy, pp. 231–240. https://doi.org/10.2495/ECO090221

Nilsson, K., Flysjö, A., Davis, J., Sim, S., Unger, N., Bell, S., 2010. Comparative life cycle assessment of margarine and butter consumed in the UK, Germany and France. Int J Life Cycle Assess 15, 916–926. https://doi.org/10.1007/s11367-010-0220-3

Odum, H.T., 1988. Self-Organization, Transformity, and Information. Science 242, 1132–1139. https://doi.org/10.1126/science.242.4882.1132

Odum, H.T., Odum, E.P., 2000. The Energetic Basis for Valuation of Ecosystem Services. Ecosystems 3, 21–23. https://doi.org/10.1007/s100210000005

Okonkwo, E.C., Ratlamwala, T.A.H., Abid, M., 2019. Energy, exergy, exergoeconomic, and exergoenvironmental study of a parabolic trough collector using a converging-diverging receiver tube 24.

Oláh, J., Aburumman, N., Popp, J., Khan, M., Haddad, H., Kitukutha, N., Popp, J., Jammu, A., Kashmir, 2020. Impact of Industry 4.0 on Environmental Sustainability. Sustainability 4674, 21. https://doi.org/10.3390/su12114674

Onat, N., Kucukvar, M., Halog, A., Cloutier, S., 2017. Systems Thinking for Life Cycle Sustainability Assessment: A Review of Recent Developments, Applications, and Future Perspectives. Sustainability 9, 706. https://doi.org/10.3390/su9050706

Oregi, X., Hernandez, P., Gazulla, C., Isasa, M., 2015. Integrating simplified and full life cycle approaches in decision making for building energy refurbishment: Benefits and Barriers. Buildings 5, 354–380. https://doi.org/10.3390/buildings5020354

Ozbilen, A., Dincer, I., Rosen, M.A., 2012. Exergetic life cycle assessment of a hydrogen production process. International Journal of Hydrogen Energy 37, 5665–5675. https://doi.org/10.1016/j.ijhydene.2012.01.003

Panesar, D.K., Seto, K.E., Churchill, C.J., 2017. Impact of the selection of functional unit on the life cycle assessment of green concrete. Int J Life Cycle Assess 22, 1969–1986. https://doi.org/10.1007/s11367-017-1284-0

Paris, H., Mokhtarian, H., Coatanéa, E., Museau, M., Ituarte, I.F., 2016. Comparative environmental impacts of additive and subtractive manufacturing technologies. CIRP Annals 65, 29–32. https://doi.org/10.1016/j.cirp.2016.04.036

Park, S., Changgyun, K., Youm, S., 2019. Establishment of an IoT-based smart factory and data analysis model for the quality management of SMEs die-casting companies in Korea. International Journal of Distributed Sensor Networks 15, 155014771987937. https://doi.org/10.1177/1550147719879378

Parrot, A., Warshaw, L., 2017. Industry 4.0 and the digital twin [WWW Document]. Deloitte Insights. URL https://www2.deloitte.com/content/www/us/en/insights/focus/industry-4-0/digital-twin-technology-smart-factory.html (accessed 12.14.20).

Passarelli, G.R., 2009. Sick building syndrome: An overview to raise awareness. J Build Apprais 5, 55–66. https://doi.org/10.1057/jba.2009.20

Passaro, R., 2019. Life Cycle of a house: how houses age gracefully. ASHI Reporter. Inspection News & Views from the American Society of Home Inspectors, INC. 36, 10–14.

Pati, D.S.N., Pahuja, A., Selvarajan, M., 2009. INTEGRATION OF EXERGY WITH LCA FOR SUSTAINABILITY. 11th NCB International Seminar, National Council for Cement and Building Material, Ballabgarh 11.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Pesonen, H.-L., Ekvall, T., Fleischer, G., Huppes, G., Jahn, C., Klos, Z.S., Rebitzer, G., Sonnemann, G.W., Tintinelli, A., Weidema, B.P., Wenzel, H., 2000. Framework for scenario development in LCA. Int. J. LCA 5, 21–30. https://doi.org/10.1007/BF02978555

Peters, J.F., 2021. Reinventing exergy as indicator for resource depletion impacts in LCA. Matériaux & Techniques 108. https://doi.org/10.1051/mattech/2021003

Petrescu, S., Borcila, B., Costea, M., Banches, E., Popescu, G., Boriaru, N., Stanciu, C., Dobre, C., 2016. Concepts and fundamental equations in Thermodynamics with Finite Speed. IOP Conf. Ser.: Mater. Sci. Eng. 147, 012144. https://doi.org/10.1088/1757-899X/147/1/012144

Polat, L., Erkollar, A., 2020. Industry 4.0 vs. Society 5.0. https://doi.org/10.1007/978-3-030-62784-3_28

Pombo, O., Rivela, B., Neila, J., 2019. Life cycle thinking toward sustainable development policy-making: The case of energy retrofits. Journal of Cleaner Production 206, 267–281. https://doi.org/10.1016/j.jclepro.2018.09.173

Poór, V., 2005. A Concise Introduction of the Extropy 5.

Portha, J.-F., Louret, S., Pons, M.-N., Jaubert, J.-N., 2010. Estimation of the environmental impact of a petrochemical process using coupled LCA and exergy analysis. Resources, Conservation and Recycling 54, 291–298. https://doi.org/10.1016/j.resconrec.2009.09.009

Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., de Amicis, R., Pinto, E.B., Eisert, P., Dollner, J., Vallarino, I., 2015. Visual Computing as a Key Enabling Technology for Industrie 4.0 and Industrial Internet. IEEE Comput. Grap. Appl. 35, 26–40. https://doi.org/10.1109/MCG.2015.45

Prasad, T.H., Reddy, K.P., Reddy, D.R.R., 2009. Exergy Analysis of Vapour Compression Refrigeration System 22.

PRé, various authors, 2019. SimaPro Database Manual - Methods Library.

Production Engineering, T.I.A. for, 2019. CIRP encyclopedia of production engineering. Springer Berlin Heidelberg, New York, NY.

Project Jupyter [WWW Document], 2021. URL https://www.jupyter.org (accessed 9.29.21).

Purvis, B., Mao, Y., Robinson, D., 2019. Three pillars of sustainability: in search of conceptual origins. Sustain Sci 14, 681–695. https://doi.org/10.1007/s11625-018-0627-5

Quinlan, J.R., 1986. Induction of decision trees. Mach Learn 1, 81–106. https://doi.org/10.1007/BF00116251

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017. Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. arXiv:1711.10566 [cs, math, stat].

Rajnai, Z., Kocsis, I., 2018. Assessing industry 4.0 readiness of enterprises, in: 2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI). Presented at the 2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI), IEEE, Kosice, pp. 000225–000230. https://doi.org/10.1109/SAMI.2018.8324844

Rant, Z., 1956. Exergie, ein neues Wort fur "Technische Arbeitsfahigkeit" (Exergy, a new word for "technical available work"). Forschung Auf dem Gebiete des Ingenieurwesens 22: 36-37. https://doi.org/10.1007/BF02592661

Raspberry Pi 4 Model B specifications – Raspberry Pi [WWW Document], n.d. URL https://www.raspberrypi.org/products/raspberry-pi-4-model-b/specifications/ (accessed 9.15.20).

Raynolds, M., Fraser, R., Checkel, D., 2000. The relative mass-energy-economic (RMEE) method for system boundary selection Part 1: A means to systematically and quantitatively select LCA boundaries. Int. J. LCA 5, 37–46. https://doi.org/10.1007/BF02978559

Reap, J., Roman, F., Duncan, S., Bras, B., 2008a. A survey of unresolved problems in life cycle assessment: Part 2: impact assessment and interpretation. Int J Life Cycle Assess 15. https://doi.org/10.1007/s11367-008-0009-9

Reap, J., Roman, F., Duncan, S., Bras, B., 2008b. A survey of unresolved problems in life cycle assessment: Part 1: goal and scope and inventory analysis. Int J Life Cycle Assess 13, 290–300. https://doi.org/10.1007/s11367-008-0008-x

Reddy, L., Preston, S.P., Shipway, P.H., Davis, C., Hussain, T., 2018. Process parameter optimisation of laser clad iron based alloy: Predictive models of deposition efficiency, porosity and dilution. Surface and Coatings Technology 349, 198–207. https://doi.org/10.1016/j.surfcoat.2018.05.054

Ren, S., Zhang, Y., Liu, Y., Sakao, T., Huisingh, D., Almeida, C.M.V.B., 2019. A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions. Journal of Cleaner Production 210, 1343–1365. https://doi.org/10.1016/j.jclepro.2018.11.025

Renaldi, Kellens, K., Dewulf, W., Duflou, J.R., 2011. Exergy Efficiency Definitions for Manufacturing Processes, in: Hesselbach, J., Herrmann, C. (Eds.), Glocalized Solutions for Sustainability in Manufacturing. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 329–334. https://doi.org/10.1007/978-3-642-19692-8_57

Requirements for the EcoLeaf PCR [WWW Document], n.d. URL http://www.ecoleaf-jemai.jp/eng/data/system_certification.pdf

Restrepo, Á., Bazzo, E., 2016. Co-firing: An exergoenvironmental analysis applied to power plants modified for burning coal and rice straw. Renewable Energy 91, 107–119. https://doi.org/10.1016/j.renene.2016.01.048

Restucci, A., 1998. Matera, i Sassi manuale di recupero.

Reza, B., Sadiq, R., Hewage, K., 2014a. Emergy-based life cycle assessment (Em-LCA) for sustainability appraisal of infrastructure systems: a case study on paved roads. Clean Techn Environ Policy 16, 251–266. https://doi.org/10.1007/s10098-013-0615-5

Reza, B., Sadiq, R., Hewage, K., 2014b. Emergy-based life cycle assessment (Em-LCA) of multi-unit and single-family residential buildings in Canada. International Journal of Sustainable Built Environment 3, 207–224. https://doi.org/10.1016/j.ijsbe.2014.09.001

Ribeiro, Matos, Jacinto, Salman, Cardeal, Carvalho, Godina, Peças, 2020. Framework for Life Cycle Sustainability Assessment of Additive Manufacturing. Sustainability 12, 929. https://doi.org/10.3390/su12030929

Richter, A., Khoshgoftaar, T., Landset, S., Hasanin, T., 2015. A Multi-dimensional Comparison of Toolkits for Machine Learning with Big Data. https://doi.org/10.1109/IRI.2015.12

Rivero, R., Garfias, M., 2006. Standard chemical exergy of elements updated. Energy 31, 3310–3326. https://doi.org/10.1016/j.energy.2006.03.020

Rocco, M.V., 2014. Primary Exergy Cost of Goods and Services: an Input - Output Approach. Ph.D. Dissertation. Milano, Italy: Politecnico di Milano.

Rocco, M.V., Colombo, E., 2016. Exergy Life Cycle Assessment of a Waste-to-Energy Plant. Energy Procedia 104, 562–567. https://doi.org/10.1016/j.egypro.2016.12.095

Rocco, M.V., Colombo, E., Sciubba, E., 2014. Advances in exergy analysis: a novel assessment of the Extended Exergy Accounting method. Applied Energy 113, 1405–1420. https://doi.org/10.1016/j.apenergy.2013.08.080

Rocco, M.V., Di Lucchio, A., Colombo, E., 2017. Exergy Life Cycle Assessment of electricity production from Waste-to-Energy technology: A Hybrid Input-Output approach. Applied Energy 194, 832–844. https://doi.org/10.1016/j.apenergy.2016.11.059

Romero, J.C., Linares, P., 2014. Exergy as a global energy sustainability indicator. A review of the state of the art. Renewable and Sustainable Energy Reviews 33, 427–442. https://doi.org/10.1016/j.rser.2014.02.012

Rosa, W. (Ed.), 2017. Transforming Our World: The 2030 Agenda for Sustainable Development, in: A New Era in Global Health. Springer Publishing Company, New York, NY. https://doi.org/10.1891/9780826190123.ap02

Rosalina, 2019. A Comparison of Machine Learning Algorithms in Manufacturing Production Process. CommIT (Communication and Information Technology) Journal 13, 17. https://doi.org/10.21512/commit.v13i1.5177

Rosen, M.A., Dincer, I., Ozbilen, A., 2012. Exergy Analysis and its Connection to Life Cycle Assessment, in: Life Cycle Assessment Handbook: A Guide for Environmentally Sustainable Products. pp. 185–215. https://doi.org/10.1002/9781118528372.ch8

Rosen, Marc A., Dincer, I., Ozbilen, A., 2012. Exergy Analysis and its Connection to Life Cycle Assessment, in: Curran, M.A. (Ed.), Life Cycle Assessment Handbook. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 185–215. https://doi.org/10.1002/9781118528372.ch8

Rosenberg, A., Hirschberg, J., 2007. V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure, in: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). Presented at the CoNLL-EMNLP 2007, Association for Computational Linguistics, Prague, Czech Republic, pp. 410–420.

Ross, S.A., Cheah, L., 2019. Uncertainty Quantification in Life Cycle Assessments: Exploring Distribution Choice and Greater Data Granularity to Characterize Product Use. Journal of Industrial Ecology 23, 335–346. https://doi.org/10.1111/jiec.12742

Rota, L., 2011. Matera storia di una città, Edizioni Giannitelli. ed.

Rubio Rodríguez, M.A., Ruyck, J.D., Díaz, P.R., Verma, V.K., Bram, S., 2011. An LCA based indicator for evaluation of alternative energy routes. Applied Energy 88, 630–635. https://doi.org/10.1016/j.apenergy.2010.08.013

Ruppert, T., Jaskó, S., Holczinger, T., Abonyi, J., 2018. Enabling Technologies for Operator 4.0: A Survey. Applied Sciences 8, 1650. https://doi.org/10.3390/app8091650

Sagunski, H., Mangelsdorf, I., 2005. Richtwerte für die Innenraumluft: Aromatenarme Kohlenwasserstoffgemische (C 9 –C 14). Bundesgesundheitsblatt-gesundheitsforschung-gesundheitsschutz - BUNDESGESUNDHEITSBLATT-GESUND 48, 803–812. https://doi.org/10.1007/s00103-005-1071-0

Saini, J., Dutta, M., Marques, G., 2020. A comprehensive review on indoor air quality monitoring systems for enhanced public health. Sustain Environ Res 30, 6. https://doi.org/10.1186/s42834-020-0047-y

Salehi, N., Mahmoudi, M., Bazargan, A., McKay, G., 2018. Exergy and Life Cycle-Based Analysis, in: Hussain, C.M. (Ed.), Handbook of Environmental Materials Management. Springer International Publishing, Cham, pp. 1–22. https://doi.org/10.1007/978-3-319-58538-3_84-2

Salman, R., Kecman, V., 2012. Regression as classification, Conference Proceedings - IEEE SOUTHEASTCON. https://doi.org/10.1109/SECon.2012.6196887

Sankhye, S., Hu, G., 2020. Machine Learning Methods for Quality Prediction in Production. Logistics 4, 35. https://doi.org/10.3390/logistics4040035

Sartori, I., Hestnes, A.G., 2007. Energy use in the life cycle of conventional and low-energy buildings: A review article. Energy and Buildings 39, 249–257. https://doi.org/10.1016/j.enbuild.2006.07.001

Schiavoni, S., Sambuco, S., Rotili, A., D'Alessandro, F., Fantauzzi, F., 2017. A nZEB housing structure derived from end of life containers: Energy, lighting and life cycle assessment. Building Simulation 10, 165–181. https://doi.org/10.1007/s12273-016-0329-9

Schrijvers, D.L., Loubet, P., Sonnemann, G., 2016. Developing a systematic framework for consistent allocation in LCA. Int J Life Cycle Assess 21, 976–993. https://doi.org/10.1007/s11367-016-1063-3

Science Europe Scientific Committee for the Physical, Chemical and Mathematical Sciences, 2015. A Common Scale for Our Common Future: Exergy, a Thermodynamic Metric for Energy. Presented at the European Strategic Energy Technology Plan (SET-Plan) Conference on European energy policy, research and innovation, p. 14.

scikit-learn, 2021. Selecting the number of clusters with silhouette analysis on KMeans clustering [WWW Document]. scikit-learn. URL https://scikit-learn/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html (accessed 12.22.21).

Scikit-learn, 2020. User guide: contents - scikit-learn 0.24.2 documentation [WWW Document]. Scikit-learn. URL https://scikit-learn.org/stable/user_guide.html (accessed 9.2.21).

Sciubba, E., 2012. A Thermodynamically Correct Treatment of Externalities with an Exergy-Based Numeraire. Sustainability 4, 933–957. https://doi.org/10.3390/su4050933

Sciubba, E., 2001. Beyond thermoeconomics? The concept of Extended Exergy Accounting and its application to the analysis and design of thermal systems. Exergy, An International Journal 1, 68–84. https://doi.org/10.1016/S1164-0235(01)00012-7

Sciubba, E., Manfrida, G., Desideri, U., 2012. ECOS 2012 The 25th International Conference on Efficiency, Cost, Optimization and Simulation of Energy Conversion Systems and Processes (Perugia, June 26th-June 29th, 2012). Firenze University Press.

Selicati, V., Cardinale, N., 2021a. The Benefits in Coupling Exergy Analysis and Life Cycle Assessment in the Context of Sustainable Manufacturing for Industry 4.0: A Real Industrial Case. IJHT 39, 12–22. https://doi.org/10.18280/ijht.390102

Selicati, V., Cardinale, N., 2021b. Interpretation of Manufacturing Sustainability-Assessment Through Hybrid Exergetic and Life-Cycle Metrics. Tecnica Italiana-Italian Journal of Engineering Science 65, 143–150. https://doi.org/doi.org/10.18280/ti-ijes.652-401

Selicati, V., Cardinale, N., 2020. Benchmarking Sustainability on an Industrial Case Within Industry 4.0 Paradigm: Advantages of Involving Exergetic Analysis in Life Cycle Thinking. TI-IJES 64, 244–250. https://doi.org/10.18280/ti-ijes.642-418

Selicati, V., Cardinale, N., Dassisti, M., 2021a. The interoperability of exergy and Life Cycle Thinking in assessing manufacturing sustainability: A review of hybrid approaches. Journal of Cleaner Production 286, 124932. https://doi.org/10.1016/j.jclepro.2020.124932

Selicati, V., Cardinale, N., Dassisti, M., 2020. Evaluation of the Sustainability of Energy Retrofit Interventions on the Historical Heritage: A Case Study in the City of Matera, Italy. IJHT 38, 17–27. https://doi.org/10.18280/ijht.380103

Selicati, V., Cardinale, N., Rospi, G., Dassisti, M., 2019. Application of Energetic Life Cycle Assessment of Retrofit Interventions on the Historical Heritage: The Case of “Palazzo Del Sedile.” IJES 63, 235–242. https://doi.org/10.18280/ti-ijes.632-417

Selicati, V., Mazzarisi, M., Lovecchio, F.S., Guerra, M.G., Campanelli, S.L., Dassisti, M., 2021b. A monitoring framework based on exergetic analysis for sustainability assessment of direct laser metal deposition process. Int J Adv Manuf Technol. https://doi.org/10.1007/s00170-021-08177-x

Seliger, G., Khraisheh, M.M.K., Jawahir, I.S. (Eds.), 2011. Advances in Sustainable Manufacturing: Proceedings of the 8th Global Conference on Sustainable Manufacturing. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20183-7

Seltzer, J.M., Diego, S., 1994. Building-related illnesses. J ALLERGY CLIN IMMUNOL 94, 11.

Settimo, G., 2012. La qualità dell’aria in ambienti confinati: nuovi orientamenti nazionali e comunitari. Notiziario Istituto Superiore della Sanità 25, 7–10.

Shirkhani, A., Kouchaki-Penchah, H., Azmoodeh-Mishamandani, A., 2018. Environmental and exergetic impacts of cement production: A case study. Environmental Progress & Sustainable Energy 37, 2042–2049. https://doi.org/10.1002/ep.12894

Shivaprasad, P., 2020. Understanding Confusion Matrix, Precision-Recall, and F1-Score [WWW Document]. Medium. URL https://towardsdatascience.com/understanding-confusion-matrix-precision-recall-and-f1-score-8061c9270011 (accessed 12.21.21).

Sick House Syndrome. Environmental Testing Products, Live Pure, Inc. [WWW Document], 2021. . Live Pure, Inc. URL https://livepureinc.com/ (accessed 11.8.21).

Siddiqa, A., Hashem, I.A.T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., Nasaruddin, F., 2016. A survey of big data management: Taxonomy and state-of-the-art. Journal of Network and Computer Applications 71, 151–166. https://doi.org/10.1016/j.jnca.2016.04.008

Siemens, n.d. SENTRON PAC3200 - Reliable and precise monitoring of electrical power systems [WWW Document]. Datasheet Available at: URL https://www.mc-mc.com/ASSETS/DOCUMENTS/ITEMS/EN/Siemens_93_47ADAPTER_Datasheet.pdf (accessed 3.27.20).

Sieniutycz, S., Salamon, P., 1990. Finite-time thermodynamics and thermoeconomics. Taylor & Francis, New York.

Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., 2009. An overview of sustainability assessment methodologies. Ecological Indicators 9, 189–212. https://doi.org/10.1016/j.ecolind.2008.05.011

Singla, A.K., Banerjee, M., Sharma, A., Singh, J., Bansal, A., Gupta, M.K., Khanna, N., Shahi, A.S., Goyal, D.K., 2021. Selective laser melting of Ti6Al4V alloy: Process parameters, defects and post-treatments. Journal of Manufacturing Processes 64, 161–187. https://doi.org/10.1016/j.jmapro.2021.01.009

Smith, J.M., Van Ness, H.C., Abbott, M.M., 2001. Introduction to chemical engineering thermodynamics, 6. ed. ed, McGraw-Hill chemical engineering series. McGraw-Hill, Boston.

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. Journal of Hydroinformatics 10, 3–22. https://doi.org/10.2166/hydro.2008.015

Soundararajan, R., Ramesh, A., Sivasankaran, S., Sathishkumar, A., 2015. Modeling and Analysis of Mechanical Properties of Aluminium Alloy (A413) Processed through Squeeze Casting Route Using

Artificial Neural Network Model and Statistical Technique. Advances in Materials Science and Engineering 2015, e714762. https://doi.org/10.1155/2015/714762
Sousa, T., Brockway, P.E., Cullen, J.M., Henriques, S.T., Miller, J., Serrenho, A.C., Domingos, T., 2017. The Need for Robust, Consistent Methods in Societal Exergy Accounting. Ecological Economics 141, 11–21. https://doi.org/10.1016/j.ecolecon.2017.05.020
Stahel, W., 2016. Circular economy. Nature 531, 435–438. https://doi.org/10.1038/531435a
Stanek, W., 2018. Thermo-ecological cost - Szargut's proposal on exergy and ecology connection 10.
Stanek, W., Czarnowska, L., Gazda, W., Simla, T., 2018. Thermo-ecological cost of electricity from renewable energy sources. Renewable Energy 115, 87–96. https://doi.org/10.1016/j.renene.2017.07.074
Stanek, W., Czarnowska, L., Kalina, J., 2014. Application of life cycle thermo-ecological cost methodology for evaluation of biomass integrated gasification gas turbine based cogeneration. Applied Thermal Engineering 70, 1007–1017. https://doi.org/10.1016/j.applthermaleng.2014.06.029
Stanek, W., Czarnowska, L., Pikoń, K., Bogacka, M., 2015. Thermo-ecological cost of hard coal with inclusion of the whole life cycle chain. Energy 92, 341–348. https://doi.org/10.1016/j.energy.2015.05.042
Stein, B., Meyer, S., Wißbrock, E., 2003. On Cluster Validity and the Information Need of Users. Artif. Intell. Appl. 373, 216–221.
Stief, A., Tan, R., Cao, Y., Ottewill, J.R., Thornhill, N.F., Baranowski, J., 2019. A heterogeneous benchmark dataset for data analytics: Multiphase flow facility case study. Journal of Process Control 79, 41–55. https://doi.org/10.1016/j.jprocont.2019.04.009
Stock, T., Obenaus, M., Kunz, S., Kohl, H., 2018. Industry 4.0 as enabler for a sustainable development: A qualitative assessment of its ecological and social potential. Process Safety and Environmental Protection 118, 254–267. https://doi.org/10.1016/j.psep.2018.06.026
Stock, T., Seliger, G., 2016. Opportunities of Sustainable Manufacturing in Industry 4.0. Procedia CIRP 40, 536–541. https://doi.org/10.1016/j.procir.2016.01.129
Stougie, L., 2012. Electricity production from renewable and non- renewable energy sources: a comparison of environmental, economic and social sustainability indicators with exergy losses throughout the supply chain, in: Proceedings of Ecos 2012 - The 25th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems. p. 15.
Stougie, L., 2011. The relation between exergy and sustainability according to literature. Life Cycle Assessment 8.
Stougie, L., Del Santo, G., Innocenti, G., Goosen, E., Vermaas, D., van der Kooi, H., Lombardi, L., 2019. Multi-dimensional life cycle assessment of decentralised energy storage systems. Energy 182, 535–543. https://doi.org/10.1016/j.energy.2019.05.110
Stougie, L., Tsalidis, G.A., van der Kooi, H.J., Korevaar, G., 2018. Environmental and exergetic sustainability assessment of power generation from biomass. Renewable Energy 128, 520–528. https://doi.org/10.1016/j.renene.2017.06.046
Stougie, L., van der Kooi, H.J., 2016. Possibilities and consequences of the Total Cumulative Exergy Loss method in improving the sustainability of power generation. Energy Conversion and Management 107, 60–66. https://doi.org/10.1016/j.enconman.2015.09.039
Stougie, L., Weijnen, M.P.C., 2014. Exergy and sustainability: insights into the value of exergy analysis in sustainability assessment of technological systems. [Uitgever niet vastgesteld], Nederland.
Subasi, A., 2020. Chapter 3 - Machine learning techniques, in: Subasi, A. (Ed.), Practical Machine Learning for Data Analysis Using Python. Academic Press, pp. 91–202. https://doi.org/10.1016/B978-0-12-821379-7.00003-5
Sutharssan, T., Stoyanov, S., Bailey, C., Yin, C., 2015. Prognostic and health management for engineering systems: a review of the data-driven approach and algorithms. The Journal of Engineering 2015, 215–222.
Sutton, C.D., 2005. Classification and Regression Trees, Bagging, and Boosting, in: Handbook of Statistics. Elsevier, pp. 303–329. https://doi.org/10.1016/S0169-7161(04)24011-1
Švajlenka, J., Kozlovská, M., 2020. Evaluation of the efficiency and sustainability of timber-based construction. Journal of Cleaner Production 259, 120835. https://doi.org/10.1016/j.jclepro.2020.120835
Szargut, J., 1989. Chemical exergies of the elements. Applied Energy 32, 269–286. https://doi.org/10.1016/0306-2619(89)90016-0
Szargut, J., Morris, D.R., 1987. Cumulative exergy consumption and cumulative degree of perfection of chemical processes. Int. J. Energy Res. 11, 245–261. https://doi.org/10.1002/er.4440110207

Szargut, J., Morris, D.R., Steward, F.R., 1987. Exergy analysis of thermal, chemical, and metallurgical processes. Hemisphere Publishing,New York, NY, United States.
Szargut, J., Stanek, W., 2007. Thermo-ecological optimization of a solar collector. Energy 32, 584–590. https://doi.org/10.1016/j.energy.2006.06.010
Taddese, G., Durieux, S., Duc, E., 2020. Sustainability performance indicators for additive manufacturing: a literature review based on product life cycle studies. Int J Adv Manuf Technol 107, 3109–3134. https://doi.org/10.1007/s00170-020-05249-2
Taelman, S.E., De Meester, S., Schaubroeck, T., Sakshaug, E., Alvarenga, R.A.F., Dewulf, J., 2014. Accounting for the occupation of the marine environment as a natural resource in life cycle assessment: An exergy based approach. Resources, Conservation and Recycling 91, 1–10. https://doi.org/10.1016/j.resconrec.2014.07.009
Tan, P.-N., Steinbach, M., Karpatne, A., Kumar, V., 2019. Introduction to data mining, Second edition. ed. Pearson, NY NY.
Tao, F., Qi, Q., Liu, A., Kusiak, A., 2018. Data-driven smart manufacturing. Journal of Manufacturing Systems 48, 157–169. https://doi.org/10.1016/j.jmsy.2018.01.006
Tay, S., Te Chuan, L., Aziati, A., Ahmad, A.N.A., 2018. An Overview of Industry 4.0: Definition, Components, and Government Initiatives. Journal of Advanced Research in Dynamical and Control Systems 10, 14.
Taylor, F.W., 2003. Scientific Management. Routledge, London. https://doi.org/10.4324/9780203498569
Terzi, R., 2018. Application of Exergy Analysis to Energy Systems, in: Taner, T. (Ed.), Application of Exergy. InTech. https://doi.org/10.5772/intechopen.74433
The International EPD® System - The International EPD® System [WWW Document], n.d. URL https://www.environdec.com/ (accessed 11.23.20).
Theakston, F., Weltgesundheitsorganisation (Eds.), 2011. Burden of disease from environmental noise: quantification of healthy life years lost in Europe. World Health Organization, Regional Office for Europe, Copenhagen.
Thibodeau, C., Bataille, A., Sié, M., 2019. Building rehabilitation life cycle assessment methodology–state of the art. Renewable and Sustainable Energy Reviews 103, 408–422. https://doi.org/10.1016/j.rser.2018.12.037
Tidriri, K., Chatti, N., Verron, S., Tiplica, T., 2016. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. Annual Reviews in Control 42, 63–81. https://doi.org/10.1016/j.arcontrol.2016.09.008
Tiwari, K., Khan, M.S., 2020. Sustainability accounting and reporting in the industry 4.0. Journal of Cleaner Production 258, 120783. https://doi.org/10.1016/j.jclepro.2020.120783
Tong, W., 2020. Machine learning for spatiotemporal big data in air pollution, in: Spatiotemporal Analysis of Air Pollution and Its Application in Public Health. Elsevier, pp. 107–134. https://doi.org/10.1016/B978-0-12-815822-7.00005-4
Toxopeus, M.E., Lutters, E., 2006. Environmental Indicators & Engineering: an alternative for weighting factors. Presented at the 13th CIRP International Conference on Life Cycle Engineering - LCE, Duflou JR, Dewulf W, Willems B, Devoldere T, Catholic University of Leuven, Leuven., pp. 75–80.
Travaglio, M., Yu, Y., Popovic, R., Selley, L., Leal, N.S., Martins, M., 2020. Links between air pollution and COVID-19 in England 28. https://doi.org/10.1101/2020.04.16.20067405
Tsatsaronis, G., 2011. Exergoeconomics and Exergoenvironmental Analysis, in: Bakshi, B.R., Gutowski, T., Sekulic, D. (Eds.), Thermodynamics and the Destruction of Resources. Cambridge University Press, Cambridge, pp. 377–401. https://doi.org/10.1017/CBO9780511976049.019
Tsatsaronis, G., 2007. Definitions and nomenclature in exergy analysis and exergoeconomics. Energy 32, 249–253. https://doi.org/10.1016/j.energy.2006.07.002
Tsatsaronis, G., 1984. Combination of exergetic and economic analysis in energy conversion processes, in: Proceedings of the European Congress on Economics and Management of Energy in Industry. Albufeira-Algarve, Portugal.
Tsatsaronis, G., Morosuk, T., 2008a. A General Exergy-Based Method for Combining a Cost Analysis With an Environmental Impact Analysis: Part I — Theoretical Development, in: Volume 8: Energy Systems: Analysis, Thermodynamics and Sustainability; Sustainable Products and Processes. Presented at the ASME 2008 International Mechanical Engineering Congress and Exposition, ASME, Boston, Massachusetts, USA, pp. 453–462. https://doi.org/10.1115/IMECE2008-67218

Tsatsaronis, G., Morosuk, T., 2008b. A General Exergy-Based Method for Combining a Cost Analysis With an Environmental Impact Analysis: Part II — Application to a Cogeneration System, in: Volume 8: Energy Systems: Analysis, Thermodynamics and Sustainability; Sustainable Products and Processes. Presented at the ASME 2008 International Mechanical Engineering Congress and Exposition, ASME, Boston, Massachusetts, USA, pp. 463–469. https://doi.org/10.1115/IMECE2008-67219

Twardowski, B., Ryzko, D., 2014. Multi-agent Architecture for Real-Time Big Data Processing, in: Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 03, WI-IAT '14. IEEE Computer Society, USA, pp. 333–337. https://doi.org/10.1109/WI-IAT.2014.185

UNEP (Ed.), 2003. Evaluation of environmental impacts in life cycle assessment: meeting report, Brussels, 29 - 30 November 1998, and Brighton, 25 - 26 May 2000. Paris.

UNI EN 1774 [WWW Document], 1999. URL http://store.uni.com/catalogo/index.php/uni-en-1774-1999.html (accessed 7.10.19).

UNI EN ISO 14040 [WWW Document], 2006. URL http://store.uni.com/catalogo/index.php/uni-en-iso-14040-2006.html (accessed 2.5.19).

UNI EN ISO 14042. Environmental management - Life cycle assessment - Life cycle impact assessment [WWW Document], 2000. . ISO. URL http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/02/31/23153.html (accessed 7.10.19).

UNI EN ISO 14044. Environmental management - Life cycle assessment - Requirements and guidelines [WWW Document], 2006. . ISO. URL http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/03/84/38498.html (accessed 10.21.19).

United Nations, 2019. Guidelines for the Development of a Criteria and Indicator Set for Sustainable Forest Management, Geneva Timber and Forest Discussion Papers. UN. https://doi.org/10.18356/235aa26b-en

United Nations, 2018. Political declaration of the third high-level meeting of the General Assembly on the prevention and control of non-communicable diseases. Time to deliver: Accelerating our response to address noncommunicable diseases for the health and well-being of present and future generations, Resolution adopted by the United Nations General Assembly, 10 October 2018.

United Nations, 2015. Transforming our world: the 2030 Agenda for Sustainable Development.

US EPA, O., 2014. Introduction to Indoor Air Quality [WWW Document]. US EPA. URL https://www.epa.gov/indoor-air-quality-iaq/introduction-indoor-air-quality (accessed 6.22.20).

Vaidya, S., Ambad, P., Bhosle, S., 2018. Industry 4.0 – A Glimpse. Procedia Manufacturing 20, 233–238. https://doi.org/10.1016/j.promfg.2018.02.034

Valero, A., 1986. Thermoeconomics as a conceptual basis for energy-ecological analysis 25.

Valero, Alicia, Valero, Antonio, Domínguez, A., 2013. Exergy Replacement Cost of Mineral Resources. Journal of Environmental Accounting and Management 13.

van der Giesen, C., Cucurachi, S., Guinée, J., Kramer, G.J., Tukker, A., 2020. A critical view on the current application of LCA for new technologies and recommendations for improved practice. Journal of Cleaner Production 259, 120904. https://doi.org/10.1016/j.jclepro.2020.120904

Velliangiri, S., Alagumuthukrishnan, S., Thankumar joseph, S.I., 2019. A Review of Dimensionality Reduction Techniques for Efficient Computation. Procedia Computer Science, 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019 165, 104–111. https://doi.org/10.1016/j.procs.2020.01.079

Verran, G.O., Mendes, R.P.K., Valentina, L.V.O.D., 2008. DOE applied to optimization of aluminum alloy die castings. Journal of Materials Processing Tech. 1–3, 120–125. https://doi.org/10.1016/j.jmatprotec.2007.08.084

Vilches, A., Garcia-Martinez, A., Sanchez-Montañes, B., 2017. Life cycle assessment (LCA) of building refurbishment: A literature review. Energy and Buildings 135, 286–301. https://doi.org/10.1016/j.enbuild.2016.11.042

Villares, M., Işıldar, A., van der Giesen, C., Guinée, J., 2017. Does ex ante application enhance the usefulness of LCA? A case study on an emerging technology for metal recovery from e-waste. Int J Life Cycle Assess 22, 1618–1633. https://doi.org/10.1007/s11367-017-1270-6

Vincenot, C.E., Mazzoleni, S., Parrott, L. (Eds.), 2017. Hybrid Solutions for the Modelling of Complex Environmental Systems, Frontiers Research Topics. Frontiers Media SA. https://doi.org/10.3389/978-2-88945-055-8
Vogel-Heuser, B., Rösch, S., Fischer, J., Simon, T., Ulewicz, S., Folmer, J., 2016. Fault Handling in PLC-Based Industry 4.0 Automated Production Systems as a Basis for Restart and Self-Configuration and Its Evaluation. Journal of Software Engineering and Applications 09, 1. https://doi.org/10.4236/jsea.2016.91001
Vogl, G.W., Weiss, B.A., Helu, M., 2019. A review of diagnostic and prognostic capabilities and best practices for manufacturing. J Intell Manuf 30, 79–95. https://doi.org/10.1007/s10845-016-1228-8
Vrchota, J., Pech, M., Rolínek, L., Bednář, J., 2020. Sustainability Outcomes of Green Processes in Relation to Industry 4.0 in Manufacturing: Systematic Review. Sustainability 12, 5968. https://doi.org/10.3390/su12155968
Waibel, M.W., Steenkamp, L.P., Moloko, N., Oosthuizen, G.A., 2017. Investigating the Effects of Smart Production Systems on Sustainability Elements. Procedia Manufacturing 8, 731–737. https://doi.org/10.1016/j.promfg.2017.02.094
Wall, G., 2011. Life Cycle Exergy Analysis of Renewable Energy Systems. TOREJ 4, 72–77. https://doi.org/10.2174/1876387101004010072
Wall, G., Gong, M., 2001. On exergy and sustainable development—Part 1: Conditions and concepts. Exergy, An International Journal 1, 128–145. https://doi.org/10.1016/S1164-0235(01)00020-6
Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D., 2018. Deep learning for smart manufacturing: Methods and applications. Journal of Manufacturing Systems 48, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003
Wang, L., 2017. Heterogeneous Data and Big Data Analytics. ACIS 3, 8–15. https://doi.org/10.12691/acis-3-1-3
Wang, L., Törngren, M., Onori, M., 2015. Current Status and Advancement of Cyber-Physical Systems in Manufacturing [WWW Document]. URL http://namrc-msec-2015.uncc.edu/sites/namrc-msec-2015.uncc.edu/files/media/NAMRC-Papers/paper_126_framed.pdf (accessed 4.13.16).
Wang, S., Liu, C., Liu, L., Xu, X., Zhang, C., 2019. Ecological cumulative exergy consumption analysis of organic Rankine cycle for waste heat power generation. Journal of Cleaner Production 218, 543–554. https://doi.org/10.1016/j.jclepro.2019.02.003
Wang, W., Zmeureanu, R., Rivard, H., 2005. Applying multi-objective genetic algorithms in green building design optimization. Building and Environment 40, 1512–1525. https://doi.org/10.1016/j.buildenv.2004.11.017
Wang, Y., Zhang, J., Zhao, Y., Li, Z., Zheng, C., 2011. Exergy life cycle assessment model of “CO2 zero-emission” energy system and application. Sci. China Technol. Sci. 54, 3296–3303. https://doi.org/10.1007/s11431-011-4567-x
Waqar, S., Sun, Q., Liu, J., Guo, K., Sun, J., 2021. Numerical investigation of thermal behavior and melt pool morphology in multi-track multi-layer selective laser melting of the 316L steel. Int J Adv Manuf Technol 112, 879–895. https://doi.org/10.1007/s00170-020-06360-0
WCED, 1987. World Commission on Environment and Development. Our common future.
Weidema, B., Wenzel, H., Petersen, C., Hansen, K., 2004. The Product, Functional Unit and Reference Flows in LCA. Environmental News 70, 47.
Welcome to Python.org [WWW Document], n.d. . Python.org. URL https://www.python.org/ (accessed 9.10.20).
Werner, F., 2005. Ambiguities in Decision-oriented Life Cycle Inventories: The Role of Mental Models and Values. Springer Science & Business Media.
Whiting, J., Springer, A., Sciammarella, F., 2018. Real-time acoustic emission monitoring of powder mass flow rate for directed energy deposition. Additive Manufacturing 23, 312–318. https://doi.org/10.1016/j.addma.2018.08.015
World Health Organization, 2015. OECD. Economic cost of the health impact of air pollution in Europe: Clean air, health and wealth.
World Health Organization, 2010. WHO guidelines for indoor air quality: selected pollutants. WHO, Copenhagen.
World Health Organization, 2009. Global health risks: mortality and burden of disease attributable to selected major risks.

World Manufacturing Foundation, 2020. 2020 World Manufacturing Report: Manufacturing in the Age of Artificial Intelligence. p. 108.
World Steel Assoctiation, n.d. WorldSteel [WWW Document]. World Steel Assoctiation. URL http://www.worldsteel.org/ (accessed 10.21.19).
Wu, S.R., Apul, D., 2015. Framework for integrating indoor air quality impacts into Life Cycle Assessments of buildings and building related products. Journal of Green Building 10, 127–149. https://doi.org/10.3992/jgb.10.1.127
Wuest, T., Weimer, D., Irgens, C., Thoben, K.-D., 2016. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research 4, 23–45. https://doi.org/10.1080/21693277.2016.1192517
Wynn, D.C., Clarkson, P.J., 2018. Process models in design and development. Res Eng Design 29, 161–202. https://doi.org/10.1007/s00163-017-0262-7
Xiahou, R., Yi, J., He, L., He, W., Huang, T., 2019. Indoor Air Monitoring System Based on Internet of Things and Its Prediction Model, in: Proceedings of the International Conference on Industrial Control Network and System Engineering Research - ICNSER2019. Presented at the the International Conference, ACM Press, Shenyang, China, pp. 58–63. https://doi.org/10.1145/3333581.3333582
Xu, X., Lu, Y., Vogel-Heuser, B., Wang, L., 2021. Industry 4.0 and Industry 5.0-Inception, conception and perception. Journal of Manufacturing Systems 61, 530–535. https://doi.org/10.1016/j.jmsy.2021.10.006
Yang, K., Zhu, N., Ding, Y., Chang, C., Wang, D., Yuan, T., 2019. Exergy and exergoeconomic analyses of a combined cooling, heating, and power (CCHP) system based on dual-fuel of biomass and natural gas. Journal of Cleaner Production 206, 893–906. https://doi.org/10.1016/j.jclepro.2018.09.251
Yang, Shiying, Yang, Siyu, Kraslawski, A., Qian, Y., 2013. Revision and Extension of Eco-LCA Metrics for Sustainability Assessment of the Energy and Chemical Processes. Environ. Sci. Technol. 47, 14450–14458. https://doi.org/10.1021/es403987k
Yang, Shiying, Yang, Siyu, Qian, Y., 2015. The inclusion of economic and environmental factors in the ecological cumulative exergy consumption analysis of industrial processes. Journal of Cleaner Production 108, 1019–1027. https://doi.org/10.1016/j.jclepro.2015.07.039
Youssef, A.B., Boubaker, S., Omri, A., 2017. Entrepreneurship and Sustainability Goals: The Need for Innovative and Institutional Solutions. Technological Forecasting and Social Change. https://doi.org/10.1016/j.techfore.2017.11.003
Yuan, Z., Qin, W., Zhao, J., 2017. Smart Manufacturing for the Oil Refining and Petrochemical Industry. Engineering 3, 179–182. https://doi.org/10.1016/J.ENG.2017.02.012
Zarei, S., 2020. Exergetic, energetic and life cycle assessments of the modified claus process. Energy 191, 116584. https://doi.org/10.1016/j.energy.2019.116584
Zhang, C., Liu, C., Xu, X., Li, Q., Wang, S., 2019. Energetic, exergetic, economic and environmental (4E) analysis and multi-factor evaluation method of low GWP fluids in trans-critical organic Rankine cycles. Energy 168, 332–345. https://doi.org/10.1016/j.energy.2018.11.104
Zhang, J., Yang, X., Appelbaum, D., 2015. Toward Effective Big Data Analysis in Continuous Auditing. Accounting Horizons 29, 469–476. https://doi.org/10.2308/acch-51070
Zhang, Q., Gao, J., Wang, Y., Wang, L., Yu, Z., Song, D., 2019. Exergy-based analysis combined with LCA for waste heat recovery in coal-fired CHP plants. Energy 169, 247–262. https://doi.org/10.1016/j.energy.2018.12.017
Zhang, T., Lin, W., Vogelmann, A., Zhang, M., Xie, S., Qin, Y., Golaz, J., 2021. Improving Convection Trigger Functions in Deep Convective Parameterization Schemes Using Machine Learning. Journal of Advances in Modeling Earth Systems 13. https://doi.org/10.1029/2020MS002365
Zhang, W., Jia, M.-P., Zhu, L., Yan, X.-A., 2017. Comprehensive Overview on Computational Intelligence Techniques for Machinery Condition Monitoring and Fault Diagnosis. Chin. J. Mech. Eng. 30, 782–795. https://doi.org/10.1007/s10033-017-0150-0
Zhang, X., Hoo, K.A., 2011. Effective fault detection and isolation using bond graph-based domain decomposition. Computers & chemical engineering 35, 132–148.
Zheng, J., Wang, Q., Zhao, P., Wu, C., 2009. Optimization of high-pressure die-casting process parameters using artificial neural network. The International Journal of Advanced Manufacturing Technology 44, 667–674. https://doi.org/10.1007/s00170-008-1886-6
Zheng, P., Liu, Y., Song, H., Wu, C.-H., Li, B., Kraemer, M.U.G., Tian, H., Yan, X., Zheng, Y., Stenseth, N.C., Dye, C., Jia, G., 2020. Risk of COVID-19 and long-term exposure to air pollution: evidence

from the first wave in China. medRxiv 2020.04.21.20073700. https://doi.org/10.1101/2020.04.21.20073700
Zhou, L., Pan, S., Wang, J., Vasilakos, A.V., 2017. Machine learning on big data: Opportunities and challenges. Neurocomputing 237, 350–361.
Zobler, M., Mantwill, E., 2018. Cooling Solutions for Laser Applications: Why a tailor-made chiller supports the optimal performance of the laser and reduces energy costs. Laser Technik Journal 15, 50–55. https://doi.org/10.1002/latj.201800020

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005),
Fondo Sociale Europeo, Azione I.1 “Dottorati Innovativi con caratterizzazione Industriale”

UNIONE EUROPEA
Fondo Sociale Europeo

Ministero dell'Istruzione, dell'Università e della Ricerca

PON RICERCA E INNOVAZIONE 2014 - 2020