POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

DEVELOPMENT OF A METHODOLOGY FOR DESIGNING PRODUCTS FOR DE-MANUFACTURING EXPLOITING ARTIFICIAL INTELLIGENCE

TESI MAGISTRALE IN MANAGEMENT ENGINEERING – INGEGNERIA GESTIONALE

AUTHOR: ELENA ZERBI

ADVISOR: MARCELLO COLLEDANI

ACADEMIC YEAR: 2020-2021

1. Introduction

Digitalization, product servitization, big data and machine learning are now part of the ordinary language when it comes to manufacturing industries. Unfortunately, these trends are not exploited for their whole potential and, too frequently, are not employed at all by companies. This work has the ambition to provide clear guidelines and a framework to help manufacturing companies exploiting data collection and analysis for accelerating and enabling the large-scale introduction of Circular Economy related strategies. Most of the issues and challenges affecting European manufacturers and customers, indeed, have the potential to be solved with the introduction of circular techniques such as Reuse, Repair, Remanufacturing and Recycling, as long as they are supported by an integrated transformation of the entire business.

The methodology presented in this work will deal with the strategic problems and implications of introducing AI and Machine learning in one of the most critical and relevant activity in a manufacturing company: product design, with a lifecycle perspective.

2. Context

The starting point is the consciousness about the need, at manufacturing industry level, to change perspective and to move to a more sustainable and efficient management of resources, mainly intended as time and materials. This is due to the combination of megatrends such as demographic growth and social change, emerging markets, climate change and scarcity of resources which lead to the need to revolutionize the way value is created.

Circular Economy approaches represent the path to be followed. [1] However, their complexity still has to be tackled with a systematic and structured method. Circular economy processes like re-manufacturing, disassembly and recycling are still far from being deployed at a large scale due to issues related to the inherent complexity of products and to the difficulties in managing and organizing the reverse chain and the collection of

returned products. These issues can be faced through the combination of different tools and strategies, which are reported in this work. Among all the complex products that are now present on the market, specific mention is done for Lithium-Ion battery packs for Electric Vehicles, due to their increasing importance in current and future scenarios and to the high complexity related to their end-of-life management. The key finding, and the starting point of the work, is the major role that the design of products has in determining the feasibility and the easiness of adoption of the circular economy techniques. [2] Product design strongly influences the ability of a product to be handled, disassembled, inspected, repaired, and remanufactured. For this reason, product design is the target of the developed methodology. The first strategy which is introduced in the work is the design for product lifecycle. Particular attention is posed on the dynamics linking product to processes and system, dealing to the need to strategically consider the re-definition of companies' processes and systems when acting on the product. Tools for achieving this goal are represented by the introduction of Artificial Intelligence technologies and in particular Machine Learning. Major issues in the adoption of this technology mostly regard the proper identification of needed data and their collection.
In current scenario, attempts of using machine learning for solving circular economy issues have been made but are still very far from being deployed. Also, they do not focus on the design of products, rather on specific operational issues, with a very poor level of integration.
On the other hand, new research are under development as concerns the introduction of machine learning in product design phase. Although they do not specifically target De-manufacturing purposes, an interesting example of such kind of implementation is provided in [3] and [4].
In this variegated and challenging context, this work has the aim to combine the mentioned aspects in a systematic way, trying to couple with their related issues.
According to CIMO logic, he aim is to develop, considering manufacturing firms (Context), a structured methodology and framework (Intervention) allowing the exploitation of machine learning and strategic tools (Mechanism) to design CE oriented products (Outcome).

3. Gaps and Objectives

As concerns the use of machine learning for solving different de-manufacturing issues, no practical implementations nor clear frameworks illustrating the path to be followed are present. What is missing is a proper **stage setting**, i.e., a preliminary step to be performed in order to allow a systematic implementation of enabling technologies for solving specific operational and tactical needs in the adoption of Circular Economy strategies. To make and example, it is impossible to have an algorithm able to perform hard disassembly tasks if the product was not designed for enabling this kind of activity.
This work poses itself at strategic level, providing a framework for supporting the design of products for Circular Economy purposes, i.e., for enabling future ML implementations at operational level through the design of the entire lifecycle of the product.
This is carried out considering the design phase as a multi-stage process, which requires many inputs to be considered and a high level of integration and collaboration between companies.
Until now, research and experiments on the introduction of AI in designing products are built considering the design phase as a simple, single stage one. This do not take into account the always increasing complexity of products' structure and the unbreakable relation linking product to process and system.
This work combines

- Structured and multi-stage design approach in a co-evolutionary perspective
- Lifecycle oriented design
- Exploitation of machine learning and generative design

For enabling the future use of machine learning for de-manufacturing processes, and in general for facilitating the introduction of CE strategies.

4. Framework

Built considering product design, process, and system together, the developed framework is illustrated in Figure 1.

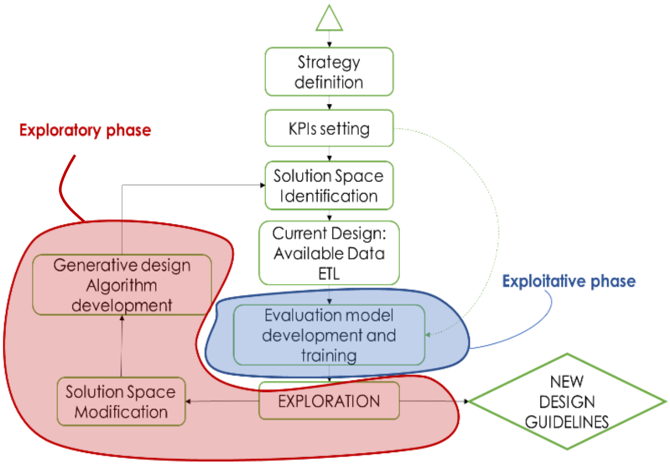


Figure 1- Framework of reference

The framework starts with the definition of the company strategy and the identification of the main business areas which will be impacted by the new pursued product design. This phase deals with the decisions about the supply chain organization, the reverse chain management, the specific circular economy approach that is willing to adopt, the key partners and activities, and with the possible re-thinking of the revenue stream and the way relationships with customers are built. In other words, the required level of integration and synergies between product, process and system leads to the redefinition of the entire business model. The second phase is a proper setting of measures for evaluating and assessing the performances of the company at every point in time, during the undertaken path. KPIs must cover

- system performances,
- processes efficiency and effectiveness,
- product design compliance to requirements

under the **manufacturing** and **remanufacturing** perspectives. This means that the set of measures must be chosen in such a way to guarantee the right assessment of circular economy related performances together with traditional economical and operational indicators. These last indicators are for example ROE, EBITDA, lead time, time to market ecc., while the circular economy related performances must be assessed with an integrated, holistic and scalable set of measures considering product, process and system together.

After these preliminary phases, the framework continues with the definition of the current space of solutions, intended as the set of features and requirements that the product must satisfy. This space of feasible solutions can be divided into three main areas:

- **Product Definition**, i.e., features of the product responding to specific customers' needs.
- **Design Variables**, i.e., technical, engineering specifications of the product.
- **Constraints** represented by Circular Economy requirements.

Once drafted, the list of features and attributes defining the solution space must be structured and transformed in a ordinate and readable set of data. The big effort in this phase lies in the translation of the unstructured information in a set of data which can be collected in an ordered and systematic way, and in the assignment of a relative importance to these data.

The next step is a benchmark between needed data, i.e., those defined in the previous stage, and currently available ones. Due to issues related to data collection, it is common that companies are not provided with the tools for collecting the entire set of data they need for developing the method. In this case, a preliminary iteration of the method can be concluded with the modification of the informative system of the company, with the introduction of new tools for gathering data, with the collaboration with other companies or with the initial modification of the product design, in a way which allows needed data collection. In this sense, the developed method tackles the issues related to the lack of clear guidelines and systematic approaches for data collection, providing a tool for identifying the needed data and consequently gather them. Once fundamental data about the current product design are available, they have to be properly cleaned and pre-processed, following the traditional Data Mining activities, illustrated in Figure 2.

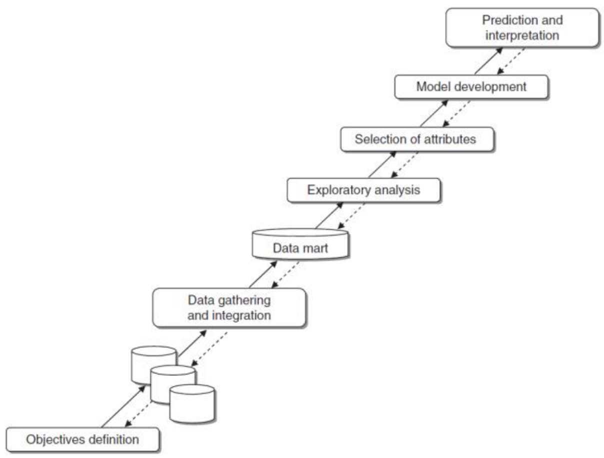


Figure 2- Data mining process

Data present in the dataset (data mart) are split into categorical and numerical attributes, in order to be properly treated in the phase of "exploratory data analysis".
This is done to perform the core part of the method, which consists in the **evaluation** of current design configurations under the defined measures of interest. In this phase, the defined KPIs specifically targeting the product are used for modelling configurations to evaluate them and assess their level of adherence to requirements.

4.1. Evaluation phase

For evaluating existing configurations, the prepared data are used as input to feed regressive algorithms able to learn the existing relations between product characteristics and specifications and the related KPI. **n** different algorithms have to be trained, being **n** the number of defined measures used to assess the goodness of the configuration (i.e., the KPIs). The stage of attributes selection will be different for the trained algorithms since features are selected basing on their relevance and impact on the target variable. Changing the target variable, which is the specific KPI, different attributes will be considered relevant and will be selected. In this direction, the development of these algorithms also acts as a validation of the choice of data. The outputs of the training and development on the n algorithms will be **n** different models which describe the same configurations under different perspectives. In this way, future configurations can be easily evaluated through the trained algorithms and evidence of their ability to satisfy circular needs and customer requirements will be given. The nature of machine learning approach requires that the best algorithm, i.e., the one that performs better in assigning the right value of the KPI to each observation while guaranteeing the right level of generalization, is not known a priori. A grid search must be performed for testing different algorithms in a "try and error" perspective. Different algorithms are evaluated basing on defined measures which are typically:

- MAE, which is the mean average error
- MSE, which is the mean squared error
- RMSE, which stands for root mean squared error
- R squared.

In case of very few data available, clustering can be used for grouping different configurations basing on the degree of their similarity and allowing further analyses to be performed at single cluster level.
The case of new product design requires different approaches to be followed. In this situation, no available configurations are present in the market, and the only input is given by the upstream stage output, i.e., solutions provided by the Generative Design Algorithm. In this case, instead of using Black box machine learning algorithms, the best solution could be that of creating ad-hoc white box models, exploiting the knowledge of the physics behind the object design. White-box models, indeed, are based on known physical laws which are able to model the drafted product configuration (general representation) and link it to specific measures of interest. This process requires big efforts in the study of existing relations between variables, but it allows to extend the method to products for which old configurations are not available, or to companies which have a deep knowledge on the physics behind a product and prefer to apply them for evaluating solutions.
Even in case of already existing products, each company can decide whether to develop white or black box models, or even to go for a double evaluation, trying to develop both approaches for improving the robustness of results.
At the end of this stage, the output is a set of models describing the behavior of different configurations basing on n relevant metrics.
This is fundamental for any kind of design process since it provides a clear and structured approach

for evaluating feasible or already existing configurations.

4.2. Exploratory phase

Exploration of results includes the merging of the different evaluations in a unique result, allowing to choose the best configuration or configurations.
This can be done with the creation of a multi-objective function, which comprehends all the measures of interest and synthesizes them in a single objective.
After that, decision must be taken on whether to be satisfied with existing solutions or to generate new product configurations, after a proper adjustment of the solution space.
In the first case, design guidelines are set, and a standard can be introduced, corresponding to product specifications related to the best configuration.
In the other case, the new space of feasible solutions can be fed into a generative design algorithm.

4.3. Generative Design Phase

Generative design is an extensive explorative design process which consists in giving design goals as input to the generative design process, along with parameters such as performance, spatial requirements, materials, manufacturing methods, cost constraints etc. Unlike optimization, the system explores all possible permutations of a solution by quickly generating many design alternatives. The system learns through testing and receiving feedback on the various iterations of a solution, and applies updates based on that feedback to the next iteration, until the design satisfies the objectives required.
Depending on the effort the company wants to put in preliminary phases and in training the algorithm, different approaches can be selected, which require different inputs and which consequently lead to diverse levels of quality of results. Lowest quality is gained with random sampling approach, which typically relies on pseudo-random number generators. In this case, bigger effort should be put in the next phase of evaluation and exploration of results. As shown in Figure 1, indeed, after generative design algorithm implementation, the process should be re-iterated for testing and evaluating the generated solutions.

5. Lithium-ion Batteries

The described method is applied to Lithium-ion batteries for Electric vehicles. The design complexity of this product and its logistic system, its complicated end-of-life management, and its increasing importance due to the growth of electric mobility market leads to the need to define a systematic way to tackle all these issues.
After a careful selection of needed data defining the space of solutions, carried out through an in-depth study on the functioning and architecture of battery packs, the phase of evaluation of current configuration has been performed. This was done considering a peculiar measure of interest when dealing with circular economy purposes: the **easiness of disassembly** of the product.
Regressive algorithms were trained, feeding them with the relevant data selected and collected starting from existing electric vehicles models. The algorithm which performed best in learning and predicting the relationships between independent input variables (i.e., attributes corresponding to the product definition, the design variables and the CE constraints) and the easiness of disassembly was a Lasso regression algorithm, whose equation is reported in Equation 1. Performances of the algorithm and its hyperparameter are reported in Table 1.

$$\min_{\underline{w}} RR\,(\underline{w}\ , D) = \min_{\underline{w}} \lambda\,|w| + \sum_{i=1}^{m} (yi - \underline{w}'xi)^2$$

Equation 1 - Lasso regression

Best score: Lasso Regressor	*Negative Mean Squared Error = -0.550333*	
Best hyperparameters	Generalization term λ= 0.01	Normalize = False
PERFORMANCES	**Train set**	**Test set**
Mean Absolute Error	0.361	0.452
Mean Squared Error	0.215	0.333
Root Mean Squared Error	0.463	0.577
R squared	0.763	0.673

Table 1 - Selected algorithm Performances

This was part of the evaluation phase, done for a single KPI (easiness of disassembly). The method can be fully performed, with the training of the other regressive algorithms able to evaluate configurations according to all the measures of interest identified. In the work, 10 KPIs are proposed, which cover all the aspects related to customers satisfaction (both B2B and B2C) and Circular Economy needs.

Next steps of the framework application would lead to the exploration of results and se consequent decision on whether to be satisfied with an existing solution or to generate new design alternatives, as shown in Figure 3.

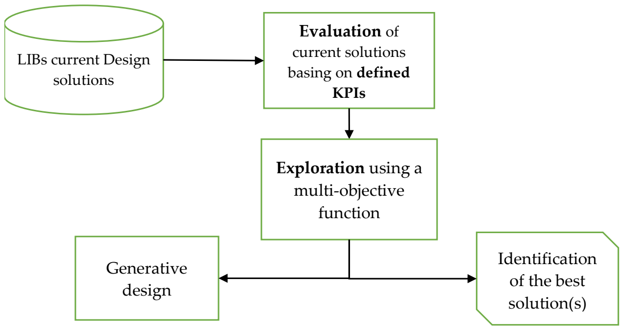


Figure 3 - Framework application to LIB packs

6. Conclusions

The developed framework poses the attention on three important aspects: the fact that the phase of product design is complex and multi-stage and must be tackled accordingly, in a systematic way; important improvements should be addressed in a collaborative way, with the co-operation between all the companies involved; the introduction of enabling technologies like machine learning and generative design must be intended as a tool to support important decisions, always taken by humans.

Compared to related research in generative design and in the use of AI for designing products, this work contributes by including all the aspects related to the business management and to the exploitation of synergies for manufacturers, by considering Circular Economy requirements as fundamental constraints when designing a product in order to simultaneously design its lifecycle, and by creating a more generic framework that shows the technical and strategical workflow of the generative design system.

It also contributes by further exploring the effects of the framework adoption on potential future iterations, explaining the benefits of a closed loop methodology aimed at continuously improve results.

Finally, the field of generative design and its application in the lithium-ion batteries for EV context shows promises and has the potential to be a part of a future designer's toolkit. The main findings in this direction regard the

- validation of the relevant set of variables for describing the easiness of disassembly behavior (mix of categorical and numerical variables), and
- the values assumed by these variables which characterize a configuration that is easy to disassemble. Furthermore, the
- identification of a precise model for describing such behavior, which can be easily replicated for other behaviors.

Bibliography

[1] Ellen McArthur Foundation, "Ellen McArthu Foundation," [Online]. Availabl https://ellenmacarthurfoundation.org/publicati ns.. [Accessed April 2021].

[2] A. B. M. C. S. K. G. S. J. D. O. B. S. T. T. Tolli "Design, Management and Control Demanufacturing and Remanufacturin Systems," *CIRP Annals Manufacturing Technolog* 2003.

[3] N. N. E. Konno, "Application of Artifici Intelligence Technology in Product Design *Fujitsu Scientific & Technical Journal,* July 2017.

[4] J. M. a. M. Sandberg, "Architectural Desig Exploration Using Generative Desig Framework Development and Case Study of Residential Block," *buildings,* p. Swede November 2020.

POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Development of a methodology for Designing Products for De-manufacturing exploiting Artificial Intelligence.

TESI DI LAUREA MAGISTRALE IN
MANAGEMENT ENGINEERING-INGEGNERIA GESTIONALE

Author: **Elena Zerbi**

Student ID:	945111
Advisor:	Prof. Marcello Colledani
Co-advisor:	Dr. Marco Diani, Dr. Luca Gentilini
Academic Year:	2020-21

Abstract

The multitude of changes that are happening in the world must be tackled in a structured and rigorous way. Given the importance of manufacturing industry in the European context, the impacts that these changes have on it are huge and should be addressed. Scarcity of resources, demographic growth and social changes, environmental issues and climate change are just some of the most important megatrends which are leading to the need to revolutionize the way value is created and delivered to customers.

The adoption of Circular Economy approaches represents the solution to most of these issues and must be supported by a strict strategy which should start with the complete re-design of products and, consequently, of business models. Currently, products are designed and engineered for satisfying functional requirements and ensure the pace of technological change, without any particular attention posed on their lifecycle. The path toward a large-scale implementation of Circular Economy approaches like Reuse, Repair, Remanufacturing and Recycling must be undertaken and should start with a correct design for product lifecycle.

A clear framework is presented for tackling all the critical aspects related to the introduction of CE strategies in every manufacturing industry, allowing to guarantee the maximum exploitation of synergies for the manufacturers. An important boost is provided by the introduction in the framework of enabling technologies like Machine learning algorithms and Generative design, which allow to speed up and lighten some phases of the process. The proposed framework starts with the definition of the strategy that is willing to pursue, and the consequent setting of KPIs for measuring the desired performances in a co-evolutionary approach, thus considering products, processes, and system together. These preliminary stages are followed by the core part of the method which consists in the exploitation of Machine Learning tools for enabling a precise evaluation of existing design configurations under the defined measures of interest. The phase of evaluation and exploration of results is followed by the decision on whether to be satisfied with an existing solution or to generate new design configurations, through the use of Generative Design tools, modifying the existing space of feasible solutions. Particular attention is posed on the phase of relevant data identification and collection, in order to guarantee the robustness of Machine Learning built models.

The methodology is then applied to a complex and critical product which is Lithium-ion batteries for Electric Vehicles. The relevance of this product in the current and future scenarios and the complexity related to its end-of-life management makes it a perfect starting point for testing the proposed method. The phase of evaluation of solutions is carried out through the training of a regression algorithm able to learn the existing relationships between variables describing the product configurations and the related Easiness of Disassembly.

Key-words: Circular Economy, Product Design, Re-manufacturing, Machine learning, Key Performance Indicators (KPI), Lithium-ion Batteries.

Abstract in lingua italiana

I continui cambiamenti che si stanno svolgendo negli ultimi anni devono essere affrontati con un approccio sistematico e rigoroso. Vista l'importanza che l'industria manifatturiera ha nel contesto europeo, le conseguenze dei cambiamenti su quest'ultima sono molto importanti e devono essere trattate.

La scarsità di risorse, la crescita demografica ed il cambiamento sociale, il cambiamento climatico ed i problemi ambientali che ne derivano sono solo alcune delle più importanti tendenze che stanno portando alla necessità di una completa ridefinizione del modo in cui il valore è creato e offerto al cliente.

L'adozione delle tecniche e degli approcci legati all'Economia Circolare rappresenta l'unica soluzione per fronteggiare gran parte dei cambiamenti menzionati, ma deve essere supportata da un approccio strategico e rigoroso che dovrebbe partire con la riprogettazione dei prodotti, e conseguentemente del modello di business dell'azienda. Al momento, i prodotti sono pensati e disegnati per rispondere ai bisogni dei clienti, e per garantire il giusto tasso di cambiamento tecnologico. Nessuna attenzione è posta al ciclo di vita del prodotto al momento della sua progettazione. Il percorso che porta ad una introduzione su larga scala di tecniche come il riuso, il re-manufacturing ed il riciclo deve essere intrapreso e deve partire con il corretto design dei prodotti pensando al loro ciclo di vita.

All'interno dell'elaborato viene presentato un preciso framework che affronta tutti gli aspetti critici legati all'implementazione delle tecniche di Economia Circolare all'interno di una qualsiasi azienda manufatturiera, consentendo di sfruttare le potenziali sinergie per i produttori. La metodologia prevede l'utilizzo di importanti tecnologie di supporto quali il Machine learning, che rappresenta uno strumento per alleggerire e velocizzare certe fasi del processo. Il framework proposto inizia con una chiara definizione della strategia che l'azienda vuole perseguire, proseguendo con la configurazione di un insieme di KPIs volti a misurare le prestazioni seguendo un approccio di co-evoluzione, ossia tenendo conto dell'indivisibile relazione tra prodotto, processo e sistema. Queste prime fasi sono seguite dal fulcro della metodologia, che prevede l'utilizzo di algoritmi di Machine learning per valutare le configurazioni di prodotti esistenti, in base a ciascuna misura definita precedentemente. Dopo una precisa valutazione ed esplorazione dei risultati, l'azienda deciderà se esistono configurazioni in grado di rispettare le performance desiderate, o se proseguire con la fase di Generative design, dando in input al sistema le giuste variabili e vincoli.

Un'attenzione particolare è posta alla fase di selezione e raccolta dei dati necessari ad alimentare gli algoritmi, per garantirne la robustezza e l'ottimizzazione dei risultati.

Il framework è applicato ad un prodotto complesso e di particolare interesse: le batterie agli ioni di litio per i veicoli elettrici. L'importanza di tale prodotto nello scenario attuale e futuro, assieme alla complessità della sua gestione a fine vita fa sì che questo prodotto sia un perfetto punto di partenza per testare la metodologia proposta. La fase di valutazione delle configurazioni esistenti è svolta allenando un algoritmo di regressione capace di imparare le relazioni esistenti tra specifiche del prodotto ed una particolare performance che è la facilità di disassemblaggio.

Parole chiave: Economia Circolare, Progettazione del prodotto, Re-manufacturing, Machine Learning, Indicatori di Performance (KPI), Batterie agli Ioni di Litio.

Contents

Introduction

Fourth industrial revolution is completely changing the scenario in which manufacturing companies are operating. It is a complex phenomenon, affecting economical, social and environmental aspects. While other historical industrial revolutions were concentrated on the development of a particular technology, which for sure had impacts on each of the three aspects mentioned before, this fourth revolution embeds an enormous and continuously growing number of technologies and tools.

Instead of thinking about the fourth industrial revolution as a responsible of the stabilization and empowerment of the take-make-use-dispose model due to mass production and increased rate of technological change, this work will concentrate on its potential to overcome barriers and enable big changes in business models and value creation, in a Circular Economy perspective. Technological innovation plays a major role in bringing the circular economy vision to life. For instance, intelligent and connected assets can enable predictive maintenance to prolong the asset life; blockchain can create traceability and transparency in supply chains to reduce waste; and repair is made easier by 3D printing of spare parts. However, the starting point must be the products' design. AI, as an emergent 'Fourth Industrial Revolution' technology, can support and accelerate the pace of human innovation to design products, can bring together aspects of successful circular business models, and optimize the infrastructure needed to loop products and materials back into the economy. Utilizing AI capabilities could create a step change which goes beyond realizing incremental efficiency gains, helping to engineer an effective economic system that is regenerative by design.

It is important to dominate the change, to foresee and exploit opportunities and to strategically keep an eye on what determines the success/unsuccess of a company. These last responsible of the fate of a company are well represented by the ability to create a sustainable business, both under the economical and environmental point of view. This brings to the identification of two most important megatrends:

- Right understanding of customer needs,
- Attention to environmental impact of production and resources criticalities.

This is the reason why the two main topics of the work arise: Circular Economy strategies which allow to decouple resources consumption and economic growth, thus strongly and positively affecting environmental problems while guaranteeing the right level of social and economic development; Artificial Intelligence technologies as a strong tool to properly understand the real needs of the customers, combining them with the requirements coming from a circular strategy and allowing to create lean and slim processes.

It is clear how powerful can be the implementation of AI enabling technologies in the creation of a structured methodology allowing to quickly and wasteless find new circular product designs that are fully able to satisfy customers' real needs.

To synthesize and visualize the concept, Figure 1 shows the described mechanisms.

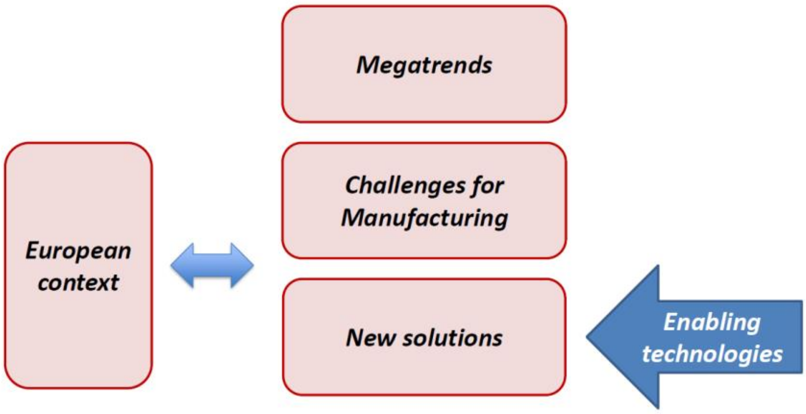


Figure 1- Problem setting

Climate change, demographic change, scarcity of resources, emerging markets, product servitization and increasing customers' requirements are some of the main megatrends; challenges for manufacturers are represented by increasing competition, volatility of raw material market, government policies and restrictions aimed at mitigating climate change. New solutions are represented by Circular Economy, and the wide range of Artificial Intelligence tools plays the role of Enabling technologies.

1. Circular Economy and De-Manufacturing

In the last years, the frequency of disruptions has risen. Disruptions may affect the technology, and we talk about technological disruption, or the entire business models.

Aim of the Circular Economy is to integrate these two types of innovation to promote energy and resource efficiency in an integrated way, in every kind of industry.

Ensuring a higher life quality with less resources extracted and consumed is one of the clearest strategical points of a circular economy approach.

The solutions allowing to decouple production and resource extraction are embedded in a sophisticated and continuously improving set of technologies, tools, processes, and knowledge-based methods aimed at recovering and re-using functions and materials from industrial waste and post-use high tech products, which goes under the name of De-manufacturing and Remanufacturing.

It is important to underline that benefits of circular economy can not be simplified and led back to purely environmental benefits. Innovations and changes linked to a circular way of making business have a strong and positive impact at social and economic level.

Equilibrium-modelling results and a comparative labor study suggest that, for the European economy at large, circular economy could produce better welfare, GDP, and employment outcomes than the current development path. This is also due to the highly volatile nature of raw materials market, to its intrinsic risk connected to politics and social instability typical of the countries in which raw material are extracted. Interrupt the dependency to this market represents a fundamental action to be undertaken.

Decoupling production and resources extraction through re-use, repair, remanufacturing, and recycling is going to delete and reduce some jobs but also to create new ones, with a positive delta and a positive impact on the life quality of workers all over the world.

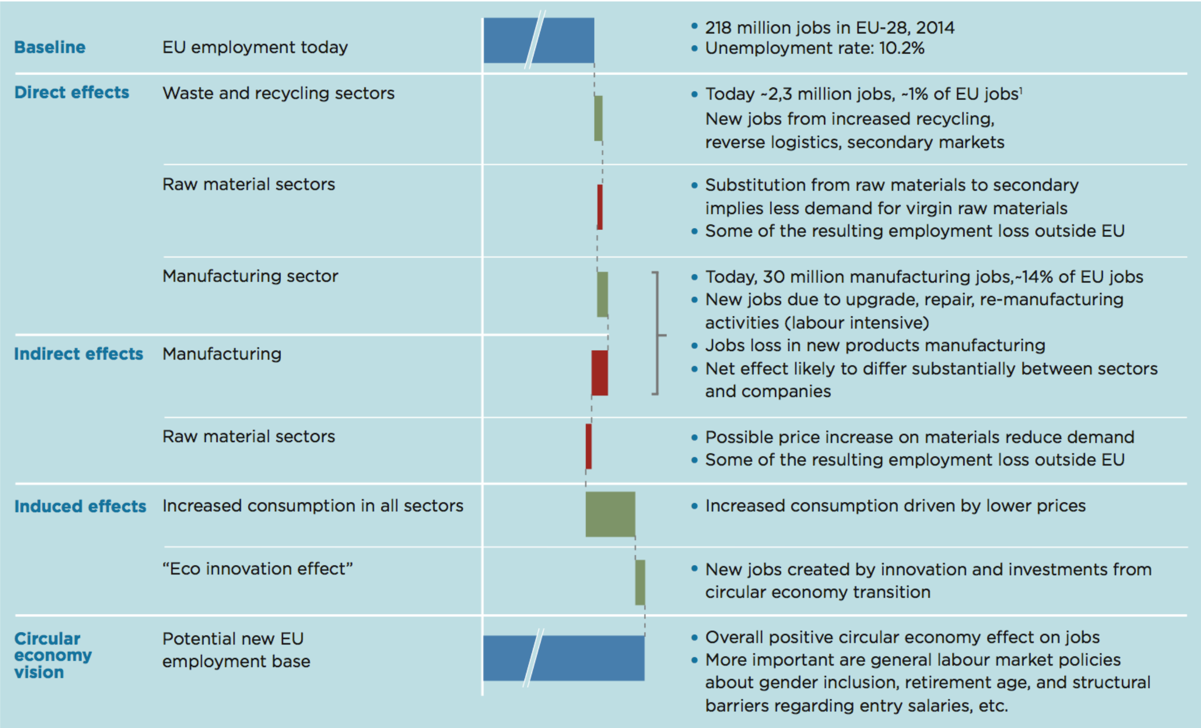


Figure 2- Qualitative employment effects of a Circular Economy transition [82] [83] [84]

The starting point is that circular economy is not about consuming less. Instead, it promotes a model in which all the people have the same chance to consume high quality products. This is achievable with a growth within strategy.

Social advantages of Circular Economy can be clearly summarized as follow.

- Creation of new jobs, as reported in Figure 2.
- Bringing of new efficient and effective technologies helping humans and allowing to reduce those tasks that are dangerous and with reduced added value.
- Improvement in the image of companies adopting such kind of strategies.
- Energy savings, since raw material extraction is much more expensive in terms of energy consumed than re-work or recycling, given the same level of purity required.
- Political benefits deriving from the independency from volatile and risky markets.
- Reduction of the gap between countries and social classes, allowing to gain a more homogeneous distribution of products and resources.

This last point is crucial and of central importance under the social point of view. It can be reached only through a large-scale deployment of activities related to De-manufacturing like remanufacturing and recycling.

CE is not moving activities along the lifecycle of the product, it is changing the entire business model, therefor the way the lifecycle is designed and created.

According to McArthur Foundation [1], value is created through four major mechanisms, which involve different strategies and different De-manufacturing tools.

1. The power of the inner circle.

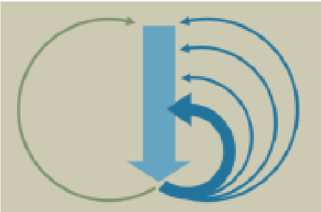


Figure 3- The power of inner circle

   Higher margins can be reached with the ability to keep the product as close as possible to direct reuse. Gains in terms of money come with savings in materials, labour, energy, capital and related externalities.

2. The power of circling longer.

Figure 4- The power of circling longer

   It's achieved by keeping the product/component/material in use as long as possible. This mechanism can be further split in two different strategies: enabling more cycles of the same product/single component or increasing the timespan of a single cycle. In the first case, attention can be posed on the entire product or on single components, allowing those components with a higher expected lifecycle to remain in use longer, through their proper implementation in different products.

3. The power of cascaded use.

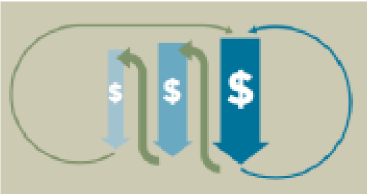


Figure 5- The power of cascaded use

This mechanism is also known as Industrial Symbiosis. It consists in recovering not only products but also energy and materials.

4. The power of pure circles.

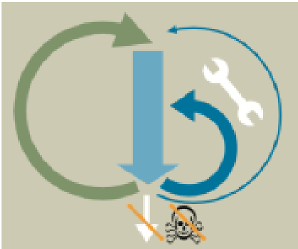


Figure 6- The power of pure circles

Cross-sectorial strategy. It enables to create value using materials discarded after the value creation by another value chain, as if they were virgin materials. This can happen exploiting different materials properties that have not been exploited in previous value-chain. Value creation happens thanks to the increased efficiency in redistribution and collection of materials through uncontaminated material streams, while guaranteeing high quality and service level.

Advantages of Circular Economy are many and involve both companies and consumers. The path through the implementation and spread of CE activities has been undertaken by some companies but, in order to reach the expected and desired results, a large-scale commitment is needed.

Such commitment must be supported by a systematic approach to tackle major challenges and critical steps, beside Cross KETs innovations in both traditional and emerging sectors. Innovations should involve:

- Digital innovations;
- Value chain building innovations;
- Enabling technologies and processes innovations.

1.1 Circular Economy in Manufacturing Industry

Concentrating on the Circular Economy enabling technologies and processes, which continuously need to innovate and improve, it is fundamental to pose the attention on the different strategies by which Circular Economy is expressed. Depending on the point at which the reverse chain intercepts the traditional value chain, different approaches can be identified.

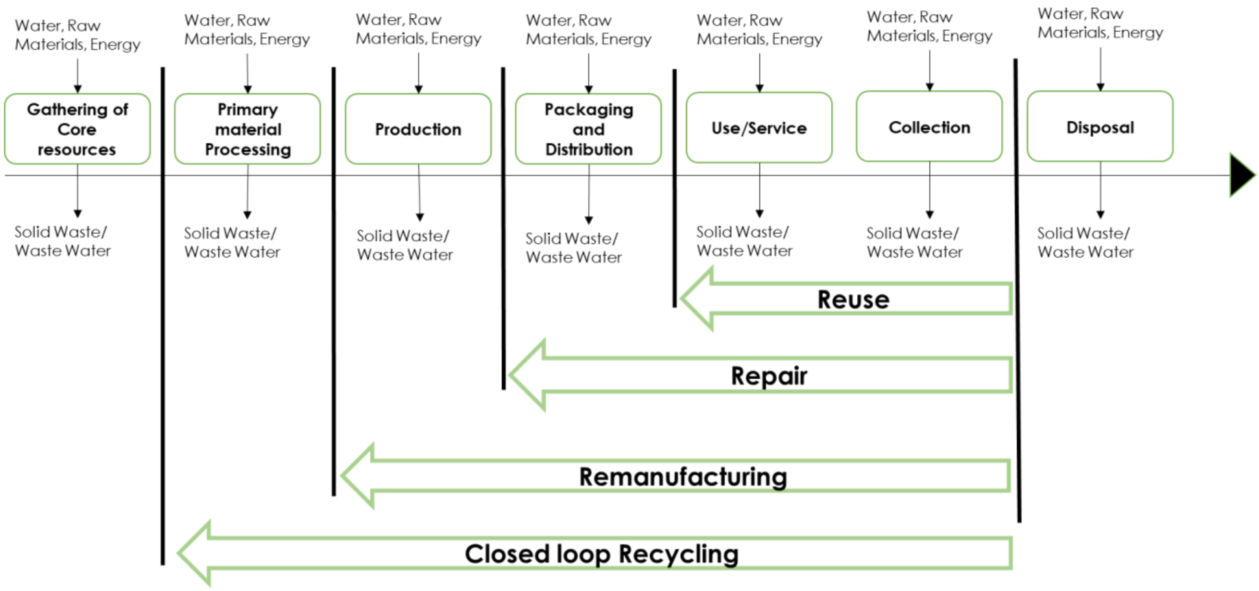


Figure 7- Circular Economy business options to close the loop

1.1.1 Reuse

This activity, which mostly exploits the power of inner circle, comprehends the set of operations aimed at putting a return product back in the market. The product keeps the same form it had in its previous cycle, with or without the need of repair or remediate it.

The product can revert to the market with the same purpose it had in its previous lifecycle, for instance being deployed in the same complex product as before, or with a completely new purpose (cross-industry reuse).

1.1.2 Repair

Repair involves the correction of faults. It comprehends the set of actions needed to return a product or component purely to a functioning condition after the detection of a failure.

Yet at this point, it starts to be clear how fundamental it is to have a proper product design, in order to easily identify and reach those components that caused the failure and directly act on them, without damaging other components.

1.1.3 Remanufacturing for Function Restore/Upgrade

Remanufacturing for function restore returns a used product to at least its original performance. From a customer viewpoint, the remanufactured product can be considered the same as a new product.

Remanufacturing for function upgrade provides new functionalities to products, extending their value and enabling the introduction of technological innovation into remanufactured products. This is done preserving as much as possible the physical resources employed in the process.

Many of the advantages of CE, listed in previous section, can be achieved through remanufacturing. This strategy has the potential to provide benefits to the remanufacturer, to the customer and to the environment at the same time.

Both in case of function restore and upgrade, remanufacturing activity is complex and multi-stage, involving many actors and requiring many features to be satisfied by the return product. Figure 8 shows the path to be followed for performing any kind of remanufacturing activity.

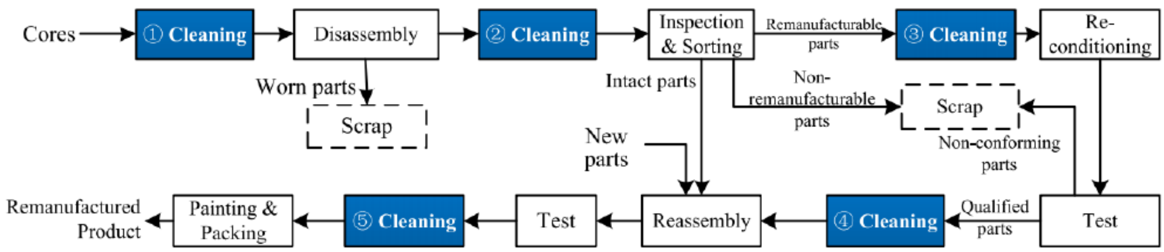


Figure 8- Remanufacturing stages

Despite the presence of many cleaning stages, each stage of the remanufacturing path is different from the others as concerns treated parts, and consequently for the adopted techniques. For this reason, remanufacturing is a very complex process which needs to be optimized and tackled with a structured approach, that should allow to rise the margins for remanufacturers.

Currently, issues and criticalities affect all the operations present in the process, especially when dealing with high value-added products.

The first issue is intrinsic in the nature of the returned products market, and it is related to the uncertainty about the conditions of the cores and returned products. Remanufacturing, indeed, is a much more complex process than manufacturing itself, due to the variety of inputs.

Stages which are mostly affected by the uncertainty of inputs are the cleaning stages. Cleaning in remanufacturing is performed on parts which are of high variability in sizes, materials, shapes, and surface conditions. Even the level of contamination and the entity of the dirtiness is highly variable and dependent on many factors comprehending the upstream works done to the part/component. It is possible to distinguish between "unusual" deposits which are deposited on the core surface because of the long-term effects of physical, chemical, or biological agents during the core's use phase, and fouling, which is introduced during the remanufacturing process.

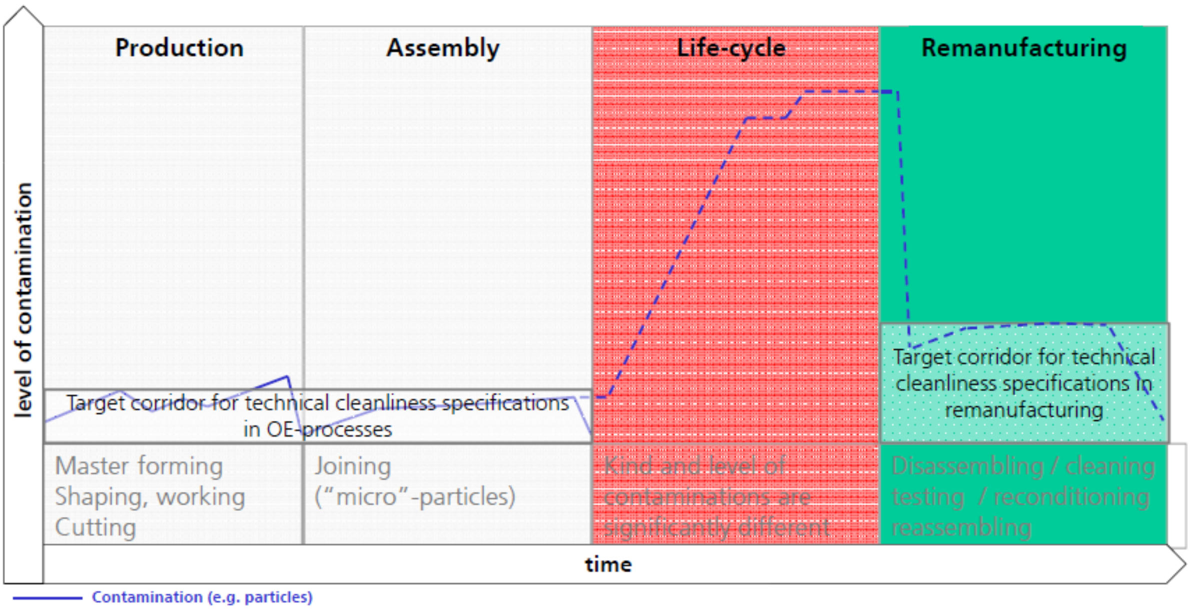


Figure 9- Comparison along the Product lifecycle of factors influencing the cleaning efforts [85]

Besides being complex, cleaning stages are upstream to the most important and critical steps of remanufacturing, having a strong impact on their efficiency and on the effectiveness of the entire process. Five different cleaning steps are identified, which are performed for different purposes and on very different parts of the product.

1	Pre-Cleaning before Disassembly
2	Cleaning of components and parts after disassembly
3	Cleaning before Reconditioning
4	Cleaning before assembling remanufactured parts
5	Cleaning before painting remanufactured parts

Table 1- Cleaning stages in Remanufacturing

Also, cleaning in remanufacturing is done for the whole mechanical parts, in order to meet the quality requirement after remanufacturing.

Disassembly represents the first step needed for remanufacturing after a rough cleaning of the core. It is the preliminary process also for recycling and recovery and as such it absorbs the most variability typical of the reverse chain. It comprehends a large set of criticalities and system optimizations needed, summarized in the concept of disassembly line design optimization.

Disassembly

Main difficulties in disassembly are related to:

- Unknown defect types at components level
- High risk of damaging components during disassembly operations
- Potential risk of some components (inflammable materials, dangerous reactions)
- Multiple core models and multiple quality classes for each core
- Need for precision for performing very difficult operations (low level of automation allowed).

Disassembly is itself a multi-stage process, which mainly involves the need to take decisions and optimize solutions subject to multiple and stochastic constraints.

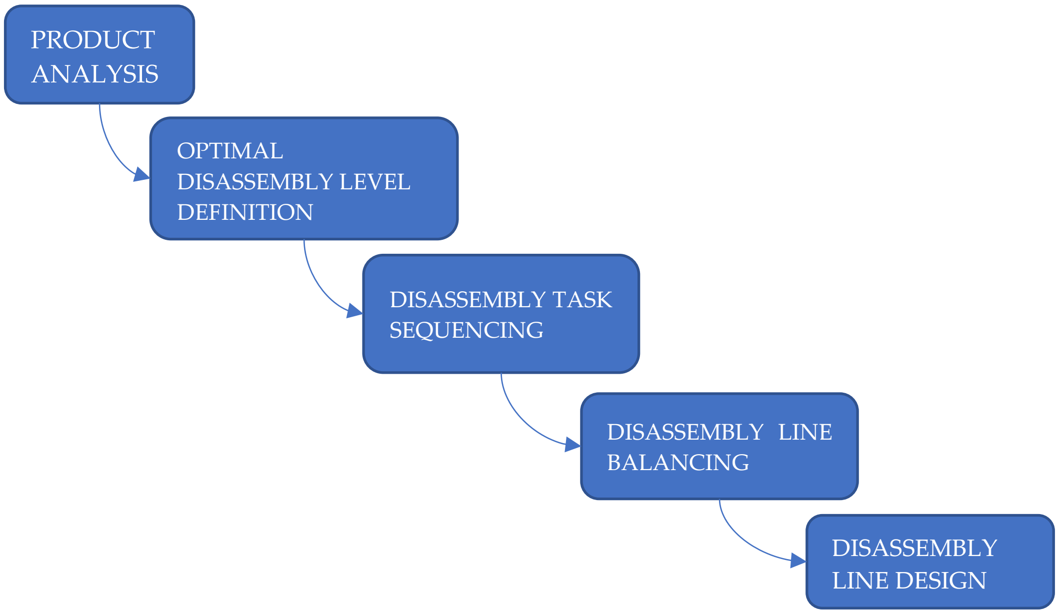


Figure 10 - Disassembly stages

1. Product analysis is carried out in order to cope with variability and requires

(i) Collection of information about product's qualitative and quantitative characteristics

(ii) Identification of KPIs for measuring such characteristics

(iii) Perform of a statistical analysis for identifying the factors that are more relevant to keep into consideration

(iv) Classification of cores in order to properly treat them

(v) Collection of information about products assembly and disassembly through graphical representation, classification and research of the disassembly mechanisms according to fastener type.

2. The definition of the optimal disassembly level is mostly an economical issue aimed at minimizing total costs. Costs comprehend cost of disassembly (both fixed costs and variable costs dependent on the task time), cost of non-disassembly, and set up costs. It is usually carried out through a table identifying the *delta* between costs and benefit of disassembling each component.

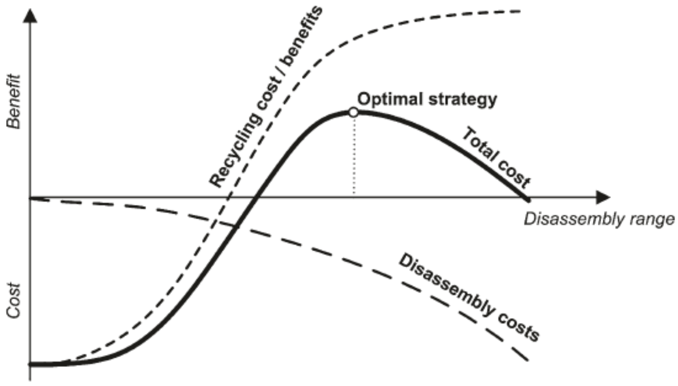


Figure 11- Optimal disassembly level

3. Disassembly task sequencing is mostly given by product analysis and economical convenience. It deals with the optimization of the sequence in which activities have to be performed in order to minimize set up costs, under the constraints given by the structure of the product.
4. Disassembly line balancing is the most critical and time-consuming activity, and it can be performed at different levels of approximation and adherence to reality. The more the modelled problem is similar to real issue, the higher the number of constraints and the higher their stochastic and uncertain nature.
   It is about computing the minimum number of stations required to perform the tasks and allocate each task to workstations according to different measures and constraints.
5. Disassembly line design adds the complexity related to coupling between machines, adding buffers and optimize their size and number to handle variability.

Despite being sequential according to a purely logical perspective, the mentioned steps must be handled in an integrated way. It is not convenient nor suggested to solve each point independently or separately from the others. In every Circular economy implementation, the links between single stage (process) and system must be always considered and tackled. The risk is ending up with a sub-optimal or even infeasible solution.

As mentioned, disassembly is just the starting point of the complex and articulated remanufacturing process. After the second cleaning stage, which involves the disassembled parts, inspection and sorting of these parts must be performed.

This stage represents another key point. Errors in this phase have a strong impact on the efficiency of the entire system and on the effectiveness of remanufacturing. In remanufacturing in general, inspection can have three targets: cores, parts, and final remanufactured product. Every inspection step is followed by a testing phase needed to validate results and verify the correct sorting of cores and parts and the correct functioning of final products.

- Cores inspection and testing is mostly needed for an economical assessment. Aim of this stage is to remove cores that would be uneconomical or impossible to remanufacture due to technical constraints. It is carried out through visual inspection, physical inspection, or identification inspection.
- Part inspection and testing is carried out after disassembly phase, and it is fundamental for selecting only those parts that are worthy to be remanufactured. The more the structure of the product is modular and standardized, the easier the way this stage can be performed. Another time, it is clear how the design of the return product strongly affects the ability to perform these tasks in an easy and automatic way. Functional and geometric inspection are used for this purpose.
- Final product inspection and testing is the very last stage before the product is put back in the market. Warranties on product functionalities needed for putting remanufactured products back in the market are delivered through this step.

The stage which characterizes the single remanufacturing process is the re-conditioning one. Despite being the core of the strategy, it is less crucial and critical than other stages. Types of reconditioning processes can be classified into 5 categories:

- Remove surface and shape defects;
- Material addition or surface replacement;
- Restore material properties;
- Assembly and fastening manipulation;
- Surface finishing.

1.1.4 Recycling

Recycling is the most expensive and energy consuming activity among the CE approaches. It requires well-structured processes and in-depth studies for optimizing the systems and tools exploited for this aim. Nevertheless, in some cases, when reuse or remanufacturing are not possible nor convenient, recovering materials through recycling can be a good and sustainable option for the business. Recycling systems are multi-stage systems including a wide range of possible operations grouped in four main categories:

1. Size reduction: it deals with only one input and one output streams. This stage changes the properties of materials in input, thus the belonging of a particle to a certain class defined by three orthogonal features: size, liberation, and shape.
2. Separation: it deals with one input stream and multiple output flows. This stage does not change the state of nature of the particles but only their concentration.
3. Mixing: has multiple input and just one output stream since it merges incoming input flows in just one flow. It changes the concentration of particles in the output flow.
4. Splitting: single input and multiple output flows since it separates the incoming flow in order to reduce the total throughput entering the downstream stages. Output flow rates are thus fractions of the total incoming flow rates. Splitting has no impacts on the concentrations of different materials in the output streams.

The objective of recycling systems is to process an incoming input (composite product or component) in order to obtain as output separated flows of pure materials which can be re-used as secondary raw materials in any other manufacturing process.

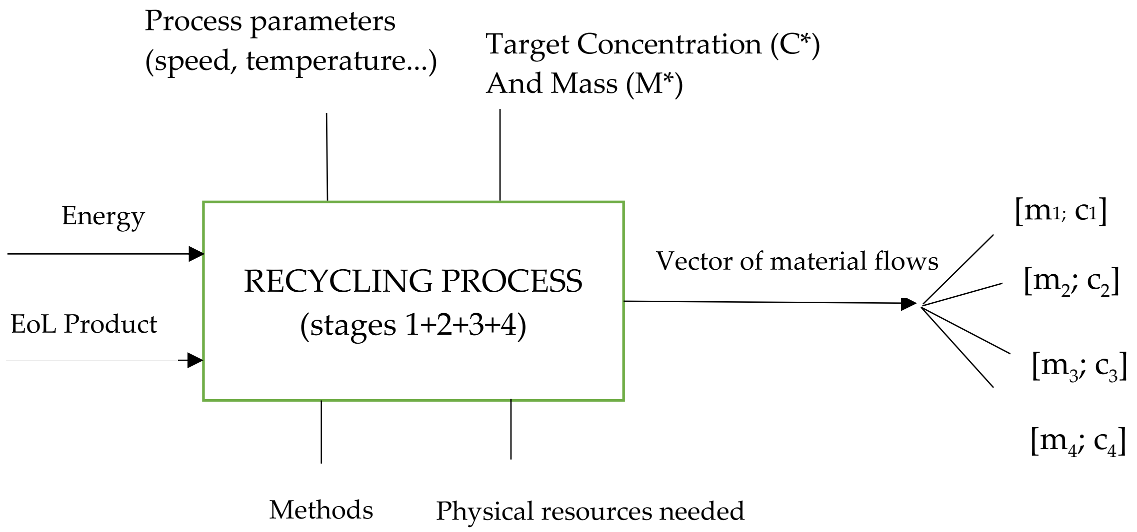


Figure 12- Recycling systems

Size reduction, also called comminution or shredding, is the only stage which brings transformations on particles structure. It brakes large particles into small particles until they reach a certain size threshold which allows them to be more easily separated through mechanical processes. This stage increases the fraction of liberated particles (i.e., homogeneous as concerns the type of composing material). The efficiency of size reduction depends on

- Design parameters, which are non-dynamically controllable. It means that are static parameters which can not be changed once the design of the stage is set. Such parameters are for example the kind of mechanical process used, the size and geometry of tools, the chamber capacity.
- Operational parameters, which instead can be changed and controlled after having designed the process. Such parameters are for instance the shaft speed, the grid size etc.

Separation stage changes the concentration in output of single flows. It splits a mixed input stream into two or more output streams in which the concentration of target material is greater than in the input flow. In general, it exploits a particular property of the target material. Using this property, an environment is created in which particles with high value of the property are forced to move in a different direction to that of particles with low value of the property. A taxonomy is given for categorizing different separation processes:

- According to the objective: Extraction, in which the target is represented by high purity of a single output stream; Separation; De-contamination, in which the target is high purity of one or more hazardous materials.
- According to the type of process applied: chemical, thermal, or mechanical/physical.
- According to the material property exploited by the process: magnetic susceptibility, electric conductivity, density, particle shape or size, colour, transparency.
- According to the environment in which the separation takes place: wet or dry.

It is not uncommon that different separation processes are used in the same system. This is what happens for critical products like PCBs, in which a thermal separation (Pyrolysis) process is performed upstream to a mechanical size reduction and a mechanical pre-treatment separation (electrostatic separation), which are followed by another size reduction based on chemical treatment (hydrometallurgical process).

Both separation and size reduction efficiency and effectiveness are strongly affected by a huge number of factors. This is the main reason why a multi-stage approach is needed.

In particular, the efficiency of separation is undermined by the presence of multi-layer particles, which creates impacts among them and prevents the separation to be accurate. The probability of having impacts increases with the flow rate increase, and therefore splitting stage is sometimes needed upstream to separation process. Splitting, indeed, has no impact on particles structure or concentration, it only reduces the quantity of mass present in the flow.

Also, particles shape and liberation have a strong impact on the selectivity of the process.

Separation process is also influenced by static parameters (given by the choice of the machine) and dynamically controllable ones, like size reduction stage.

For better comprehending the complexity linked to such kind of process, and to the design of a recycling system in general, the notions of recovery and grade must be explained, which are the performance measures of interest for a real separation process.

- Recovery of a target material belonging to a mixture i, in the output flow j is the quantity of target mass in output on the total mass in input.

$$R_{i,j} = \frac{m_{i,j}'}{m_i} \qquad \frac{[output]}{[input]}$$

Equation 1- Recovery of target material in the mixture i

- Grade of a target material belonging to mixture i in the output flow of interest j is the mass of target material in output from the mixture i on the total mass of target material in output.

$$G_{i,j} = \frac{m_{i,j}'}{m'_j} \qquad \frac{[output]}{[output]}$$

Equation 2- Grade of target material in mixture i

While Recovery is a measure of the quantity of material recovered, Grade is a measure of the quality, i.e., the level of purity of the target material.

Since no real process can provide an ideal separation, it is important to design a process allowing to gain the right trade-off between grade and recovery. Optimal trade-off is not fixed, rather, it changes with market requirements and volatility of prices of materials. Therefore, a sustainable and profitable recycling system should provide the right level of flexibility, allowing to match the best economic point for the specific point in time.

Considering the whole recycling system, composed by a specific sequence of all the stages described, important KPIs are:

- Grade $_{i,j}$ (z) i.e., grade of flow between stage i and stage j of material z
- Recovery $_{i,j}$ (z) i.e., recovery of flow between stage i and j of material z
- Total throughput E $_{i,j}$ [kg/hour] i.e., the total (average) amount of material crossing a branch (i,j)
- Effective throughput $E^{EFFECTIVE}{}_{i,j}$ i.e., the subset of the total throughput represented by the only target materials. It is a very important KPI since it combines the quantity of material crossing a branch with the quality of material in output.
- Specific energy
- Work in progress
- Profit

These KPIs can be applied to one stage at a time. However, in order to measure the overall system efficiency, the definition of a clear strategy is needed. This strategy deals with the design of the recycling process, i.e., the setting of a feasible sequence of visiting different stages and their definition/design. It is important to keep in mind the strong link existing between single process level and system level, since the decisions taken at every single stage have a huge impact on the downstream stage and thus on the overall result.

To the aim of this work, explaining the complexity of recycling processes has a meaningful scope since it allows to understand how a structured and clear strategy plays a fundamental role when it comes with turning complex activities into sustainable and profitable ones.

In recycling case, in fact, suitable systems can be built through the ability to predict the particle features in output from a certain stage together with the energy consumption of that stage, knowing the input particle features and the machine parameters. Also, the knowledge of the effects that upstream stages have on the performance of downstream stages are needed to determine the right choice of machine parameters.

The development of a methodology aimed at allowing this prediction, and gaining this knowledge, is a key activity which involves the exploitation of mathematical tools and complex equations, and which can be supported and enhanced by the collaboration between humans and enabling technologies like machine learning.

This strategy can be extended to many other purposes, including the broader aim of this thesis. The concept of integration between system and single process will be extended to the whole company, considering as process the single company function/operation and as system the overall enterprise. Synergies and relations will be investigated within the development of a framework.

The first point to highlight is that, despite the big efforts required by these activities, and the need to invest time and money in designing flexible and reconfigurable systems, results in terms of monetary savings deriving from their implementation will always be positive for manufacturers.

Manufacturers, indeed, already have a big advantage deriving from the knowledge of the product, beside owning infrastructures and machines to be exploited also for reverse chain activities.

1.1.5 Manufacturer-centered approach

A conceptual model which synthesizes the existing synergies and mechanisms between manufacturers and CE approaches is present, known as manufacturer centered approach. It provides the right key to understand how business models and value chains must be adapted to circular economy requirements in order to create a concrete advantage for the manufacturer.

The key concept is that manufacturers have a favorable position with respect to independent remanufacturers since they own the knowledge base related to how the product is built and commercialized and why it had been thought and designed in a certain way. Manufacturers have the potential to exploit synergies. This potential can turn into effective gain only if they are able to properly modify their value chain and their business model, besides to correctly re-think and think new products that are compliant to circular specifications.

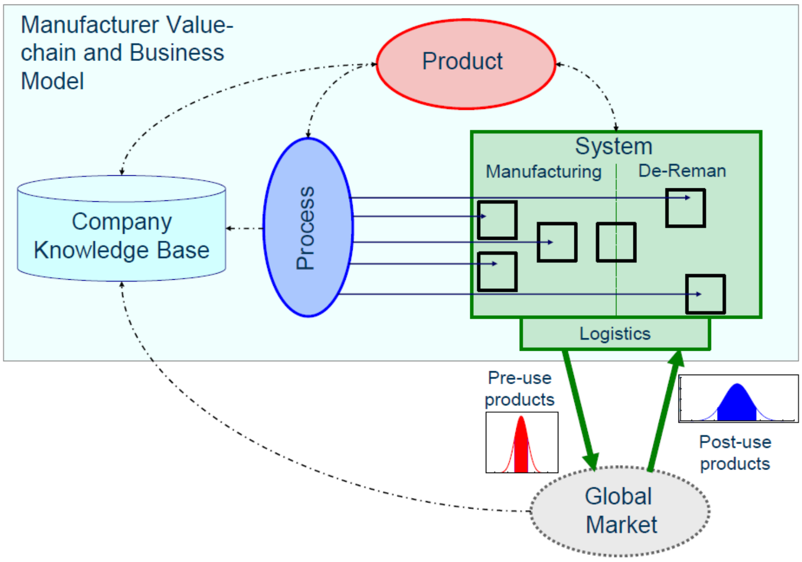


Figure 13- Manufacturer-centered Approach

The framework (Figure 13) matches the dynamics of the product, the process, and the system life cycle, and it was designed to address the request of improving manufacturing performance in dynamic and unpredictable environments. The three elements (product, process, and system) should be profoundly linked and evolve in a coordinate manner to achieve the said goal. Product area comprehends the product design, while in the process area there are the methodologies and technologies to achieve production, fed by the knowledge base which is founded on the product characteristics as well. The system level comprehends both manufacturing and De-manufacturing activities, particularly the ones illustrated in remanufacturing and repair processes, that can be located under one roof to exploit resources synergies. Another important stage connecting the manufacturer with other actors outside the company is logistics: for companies implementing De-manufacturing operations logistic is bi-directional, that is, it has the role of shipping manufactured and remanufactured products to the global market, and to recover used products from it. The post-use product conditions will be more diverse compared to pre-use products. Another time, in order to gain the maximum from a circular approach, it is necessary to carry out a value chain and business model reconfiguration.

Today, there are few virtuous producers that are worthy to mention since they represent a good example of implementation of the manufacturer centric approach.

Talking about remanufacturing for function restore, that is of primary interest in this work, one of the most involved industries is the automotive one, and products that are more affected are the mechatronics components, since they are increasing their presence in new vehicles. It's interesting to know that the only 17% of profit in automation industry comes from the selling of new products; and as far as the 40% of profit comes from post-use services. [2] Remanufacturing has the potential to have an impact and improve the main source of profit of automotive industry.

One of the consequences that remanufacturing implementation has on the supply chain is the creation of competitiveness between suppliers and the original equipment manufacturers that are both in possession of the knowledge base about cores features; another actor that takes part to the competition is the independent remanufacturer. However, he starts with a disadvantage since he has to collect lot of information about the cores and components and thus is placed in a disadvantageous position.

An example of supplier who put an effort in the creation of a sustainable re-manufacturing system is Knorr-Bremse, which reconfigured its business model in order to remanufacture EBS (electronic braking systems) for trucks. The stages related to re-manufacturing activities (Figure 8) are performed in a dedicated plant in Czech Republic, exploiting the work of high skilled operators. Indeed, due to the high variability which characterizes returned products market, human intensive operations are needed. Nevertheless, the Re-Assembly operation of re-manufacturing components is done in Germany, where the new products are produced. This is done to guarantee the same high-quality of re-manufactured components, which are tested with the inspection machines used to test new products. This also avoids additional costs related to the purchasing of other inspection machines which are extremely expensive. Knorr-Bremse exploits the synergies between manufacturing and de-manufacturing within the same organization, but in different production sites, while it conducts the expensive testing and inspection activities in the main plant in Germany.

It's interesting to notice that, when improvements in design for product lifecycle will allow less human-intensive activities to be carried out for remanufacturing, all the stages of the reverse-chain could be performed in the original manufacturing plant, allowing to save additional costs and fully exploit the knowledge-base.

Caterpillar, the largest Construction-equipment manufacturer, is developing a decision support system for future re-manufacturing operations. This system is composed by in-line inspection and core classification as upstream stages to the remanufacturing process. This will allow to have more accurate information on the way products have been used, reducing the uncertainty about core conditions, and allowing a better understanding of the operations to be carried out. With an overall speed up of the process.

Shifting the attention to recycling, always remaining in the automotive sector, it's interesting to have an overview on the recycling systems for end-of-life vehicles. ELV, in fact, present a large variety of critical commodity materials that can be recovered and used for different applications. ELV are usually treated with an integrated de- and remanufacturing process flow, meaning that as soon as the materials are removed/disassembled, they are sent to re-manufacturing or recycling steps. The last step of this integrated process is shredding of the remaining part (ASR= automotive shredded residuals). Here opens the large issue briefly introduced and not tackled in this work which regards the way recycling systems are designed. Recycling systems must be extremely flexible and smart in order to adapt to the volatile and changing needs of the market, beside of course being efficient and effective in recovering target materials.

It is straightforward that some components are more critical than others, either to be disassembled, either to be remanufactured/recycled. Batteries in ELV represent one of the most critical components.

As an example of manufacturer-centric approach application, Renault created its own network of dismantlers, recyclers, and remanufacturer partners (Indra Group). The aim of the French company is to keep everything inside in a complete growth-within perspective, enabling the re-use of its own materials. In this way the added value and the gain of implementing such a strategy are fully exploited. CEO of Renault stated: *"Who is better than the producer of the goods (and corresponding services) to detect potential resources in EoL products and safeguarding their technical and economic value".* This can be made through a careful control, quality, traceability, and optimization process.

Renault also adopted the design for disassembly and design for re-manufacturability strategies, adapting the design of vehicles to specific needs imposed by De-manufacturing processes. This kind of action brings new constraints in the design process since it forces to consider the whole product lifecycle while designing it. These strategies will be briefly tackled in next sections.

Indra group developed and industrialized advanced engineering applications for disassembly lines optimization, while improving the overall recycling rate of materials coming from the core business. The challenge for Renault is to keep the pace of the technological change: the development of new dismantling procedures for hybrid and electric vehicles is in progress together with the establishment of a proper recovery network for End-of-Life batteries.

Separate mention must be done for the treatment of WEEE (Waste of Electric and Electronic Equipment). In this case the collection per se is a problem. In Italy the current collection system works but it is not efficient and not sustainable overtime. This is because manufacturers do not know anything and do not care about the collection and the reverse part of the chain. The need is to involve manufacturers in a more sustainable and conscious way of doing their business and producing their products, also given the potential advantage they could gain. Thanks to the implementation of a proper and structured methodology, WEEE, in particular PCBs, could finally be recycled and treated in the most efficient way. Problems related with PCBs are many and regard all the stages of the value chain. The starting point for overcoming these issues could be a smart and circular oriented design, enabled by the exploitation of data and machine learning tools, which could help for example in solving the issue related to different concentration of critical materials in different parts of the product, with key metals highly concentrated in small parts of PCBs. Furthermore, the reuse and repair that nowadays are not supported at all, could find their scope thanks to proper data collection. However, in conventional Linear Economy, electronic manufacturing is well a matured industry operating at very low cost and re-usability workflow and detachable provision for certain long-life chipsets or devices have not been included due to additional cost incurred during the production. To mitigate certain issues associated, it is mandatory to understand the customer needs aligned with CE and therefore Quality Function deployment (QFD) is an efficient tool to design a product that has reusability function. The very starting point, however, is the commitment of the manufacturer and his willing to undertake a circular approach. [3]

To put this into context, across all e-waste categories, 48% of the monetary value is embedded in the PCBs fraction, yet PCBs account for only 8% of the overall e-waste mass. For IT and Telecommunications equipment, or consumer electronics, this PCB mass fraction is even higher, typically 13% – 14%. PCB fraction recovery rates can vary greatly. While 100% recovery can be achieved through labor intensive manual disassembly and separation, mechanical shredding or crushing, if coupled with automated flake sorting, results in poor recovery – typically 30% – 80%. About 20% of precious metal content is lost to non-recoverable output side-streams such as plastics, process residues or saleable metals when e-waste items are mechanically pre-processed.

Through a proper design of e-products and PCBs, this issue could be overcome, and an easy and automated efficient material recovery could be implemented.

1.1.6 Open Issues

As clearly inferable from the PCBs example, the complexity related to the reverse chain management and to circular manufacturing approach still have to be tackled. This complexity is the result of hundreds of years of development of a linear model of production and consumption all over the world. European economy, indeed, is surprisingly wasteful in its value creation model.

All the strategies related to a circular approach, which involve the introduction of a closed loop at different levels, have the potential to become a profitable solution for companies and for customers. Profitability comes with a proper market transformation, which requires a structured and rigorous approach to be followed.

For this reason, complexity can be classified basing on the level of integration considered. Starting from the highest degree of integration, at Strategic level, the previously mentioned challenges related to the change management to a new business model and to the creation of a new value chain can be mentioned. The change must happen both at technological and business model level, with a wave of disruptions targeting all the aspects related to value creation. One of the main problems and challenges of such a change lies in the paradigm shift to a growth within model. Also, the need to build flexible systems and processes able to follow the market requirements and to be sustainable and feasible under the economical point of view.

Complexity at strategic level	Complexity at tactical/operational level
Build reverse chain and manage reverse logistic	Cores classification, also due to increasing product variety
Predict remanufacturing demand	Manage cores variability (quantity and quality)
Forecast cores quantity	Manage preliminary stages like sorting and inspection
Adapt the business model	Deal with optimal disassembly level and disassembly line design
Adapt the product design	Manage cleaning processes
Build flexible recycling systems	Deal with high investments and costs for size-reduction processes
Build flexible and profitable remanufacturing systems	Deal with criticalities in separation processes
Choose the right statical parameters in recycling processes	Choose the right dynamic parameters in recycling processes
Development of a proper performance measurement system	

Table 2- Taxonomy of Circular Economies criticalities

At tactical and operational level, all the problems related to the choice of machines' dynamic parameters and optimization of functions related to disassembly.

Critical stages of de-manufacturing strategy, indeed, still require huge number of resources seen as humans and time, thus compromising the perceived potential gain of this model implementation.

Main issues are summarized in Table 3, labelled according to the De-manufacturing process they refer to.

Challenges and issues	Process
High variability in the conditions of end-of-life products	Inspection, Disassembly
Poor information about return products	Inspection
Increasing product variety	Inspection, Disassembly
High cost and amount of resources required for cleaning	Cleaning
High cost of manual labor-based disassembly	Disassembly
High cost of size reduction process	Size reduction
Randomness of separation process	Separation
High fluctuation in materials' value	Recycling
Input uncertainty for remanufacturers	Remanufacturing, Logistics

Table 3- Challenges and issues of De-manufacturing at operative level

Another important gap at strategic level that has risen in recent studies and that prevents CE from being an established strategy, is the lack of an integrated, holistic, and scalable framework for measuring Circular Economy performance. The measurement of performance is paramount to track progress and foster the implementation of the CE paradigm [4] . Despite all the reviewed contributions are focusing on issues of interest to the industrial decision-maker (IDM), they take different perspectives. The majority address frameworks at the single-product or the materials-and-resources level, some consider the firm level, others consider a system perspective, while only a few studies analyze different levels simultaneously. The use of an index, i.e., a combination of indicators providing a snapshot of a given performance area, presents several benefits: it is easy to understand, communicate, and benchmark efforts towards CE. Among the most common indexes, it is possible to cite the Circular Economy Indicator Prototype [5], the Circular Economy Toolkit [6], the Material Circularity Indicator [7]. All three are nonetheless related to the product level and focus mainly on environmental aspects, although business opportunities are described by the Circular Economy Toolkit. Despite being straightforward in their use, indexes present drawbacks in their application, as they neither distinguish between different loops (e.g., reuse, remanufacturing, recycle) nor provide guidance for circular product development. The measurement of performance should allow internal improvement, communication with external stakeholders, and benchmarking with peers. In this way, a performance-measurement system should be general enough to be applied in different contexts, such as sector and geographical area, while also allowing a tailored approach to possible distinct needs. To reduce the complexity of the measurement process, it is suggested for an effective performance-measurement system to meet all the above-mentioned features (integrated, holistic, and scalable).

As for integration, an effective performance-measurement system for CE should provide clear indications regarding the simultaneous coverage of other paradigms within the manufacturing firms. For developing such an integrated system, a great and deep understanding of the interrelations and overlaps between the paradigms would be required, and additional value could derive from the simultaneous consideration of the perspectives of multiple IDMs within the same manufacturing firms and their industrial systems.

As for the holistic perspective, an effective performance-measurement system for CE should thus provide coverage of different CE levels and approaches, understanding the interrelations among them. Again, it is advisable to have a single, unique system for measuring performance at different levels, rather than separate ones.

As for scalability, an effective performance-measurement system for CE should be adapted to different firms, specifically SMEs and New Adopters, according to their characteristics and their evolving needs, while also simultaneously allowing for internal performance measurement and benchmarking activities.

McKinsey, [8], provides a very general and aggregated framework for measuring CE performances at very high level.

	KEY PRINCIPLE	PRIMARY METRIC	SECONDARY METRICS
1	Preserve and enhance natural capital by controlling finite stocks and balancing renewable resource flow	Degradation-adjusted net value add (NVA)	Annual monetary benefit of ecosystem services
			Annual degradation
			Overall remaining stock
2	Optimize resource yields by circulating products, components and materials in use at the highest utility	GDP generated per unit of net virgin finite material input	Product utilization
			Product depreciation/lifetime
			Material value retention ratio
3	Foster system effectivenes by revealing and designing out negative externalities	Total cost of ecternalities and oppurtinity cost	Cost of land, air, water, noise pollution
			Toxic substances in food system
			Climate Change, congestion and health impacts

Table 4- Aggregated CE measures

Structured methods and frameworks are needed for solving all these issues. Important tools can help in the achievement of results, allowing to save time and resources. Artificial intelligence represents one of these key instruments, being a support for speeding up processes and reduce wastes and losses, while allowing to consider many aspects at a time, in an integrated perspective.

2. Artificial Intelligence

"*Artificial intelligence enabled big data analytical tool coupled with data compression could revolutionize the IoT industry and move at the edge for real-time decision making. In a nutshell, IoT is the senses, Big Data is the fuel, and artificial Intelligence is the brain to realize the future of a smart connected world*." [3]

Artificial intelligence is an overarching term for a collection of technologies, dealing with models and systems that perform human-like cognitive functions such as reasoning and learning. AI helps to solve problems through pattern recognition, prediction, optimization, and recommendation generation, based on data from videos, images, audio, numeric, text and more.

Wide set of technologies cover a huge number of fields of application. However, common path can be identified, as a general framework that is necessarily followed by all the Artificial Intelligence technologies.

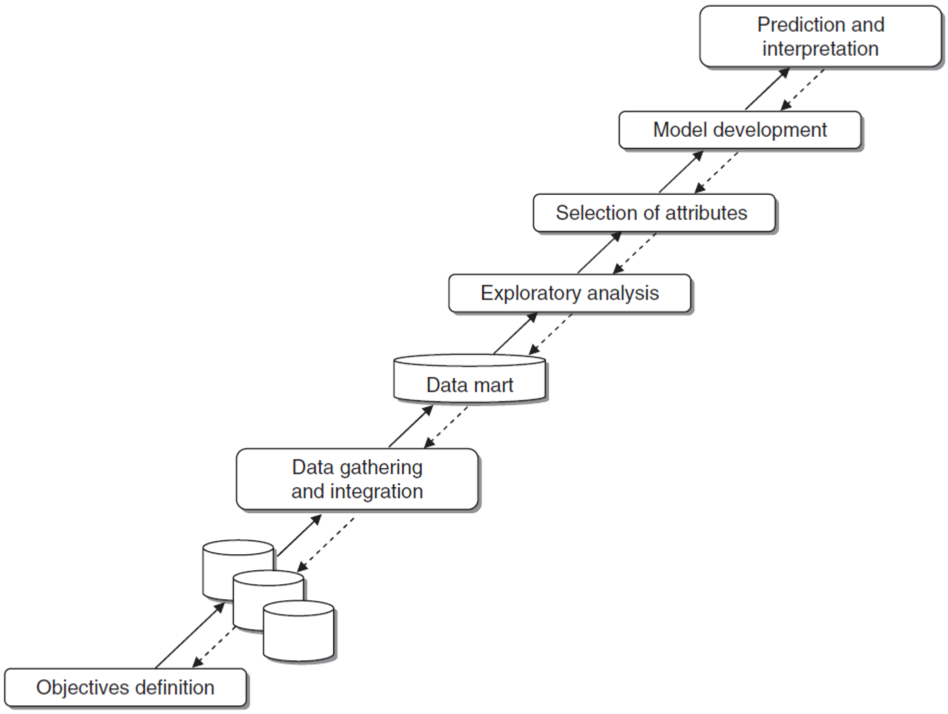


Figure 14- Data mining Process [9]

After the problem definition and formalization, the required data are collected through capturing images and other metadata collected through sensors. Data may come from different sources and therefore require integration. Data sources can be internal, external, numerical, or categorical. Data integration aims at giving data a common structure to carefully represent the real-world application under investigation. Data are then consistently labelled and engineered into a format which is machine-readable. Preliminary and qualitative conclusions are drawn for any kind of data in the **exploratory** analysis phase, while the selection of attributes usually involves a more quantitative and automated process, supported by mathematical tools. After that, an algorithm is developed. Different types of algorithms can be employed depending on the use case.

All these steps require a certain amount of human work, which depends on the specific application field and starting condition. Future of IoT will trigger massive amount of structured, unstructured, real-time, images, videos, information data into the cloud network. The staggering amount of data needs to be stored, processed and hence it is a critical challenge to process data when it is still in motion and extract valuable information from it. Big Data Analytical tool coupled with Artificial intelligence will be employed to derive conclusion, examine raw data with the purpose of finding patterns by deep learning algorithm. For example, understanding the information/data from the sensors/ things connected and running through several data sets to look for meaningful correlations between each other to positively impact businesses.

AI is not new, its application can be draw back to the beginning of the 50s. What is changing now is the ability to collect and store amounts of data which are not comparable to those available in the past. Consequently, the amount of work required to process, clean, structure these data is continuously growing. There is a common misconception that AI algorithms are 'smart' by themselves. In fact, AI is dependent on humans to clearly establish the inputs and outputs for a model (piece of software) before a machine can solve it. As mentioned, AI encompass a multitude of specific applications which mainly differ for the "Model development" stage. Indeed, each algorithm is trained to perform a very specific function, such as object detection for autonomous driving, identifying fraudulent financial transactions or delivery route optimization.

AI Application	Functionalities
Pattern recognition	Financial Risk estimation Classification Data model building Music and voice recognition …
Prediction	Medical diagnosis Targeted advertising Recommendation engines …
Optimization and Planning	Route planning Spend optimization Dynamic pricing
Integrated solutions with robots	Autonomous driving Robotic surgery Household Robots

Table 5- AI applications [10]

Within all the different fields, the work concentrates on industrial applications, which mainly relates to Prediction applications, pattern recognition and Integrated solutions with robots, involving a specific branch of AI which is Machine Learning.

Within the era of the fourth industrial revolution, in fact, it is not uncommon to hear about artificial intelligence when talking about manufacturing industries. The reason is fully explainable starting from one of the definitions of the technology: it is a way to make objects learn and predict what is happening and so enabling them to help humans in performing every kind of task.

As an example, rather than building a control system that works to rigid tolerances, pre-defined by human analysis, Machine Learning (ML) could be used to define tolerances for industrial control systems with little data. In addition, ML could be used to progressively improve the performance of a specific task using an ever-increasing amount of data captured. Machines could be able to 'intelligently' identify and disassemble e-waste items and sort sub-assemblies and components into categories without human intervention, only after a proper and structured stage setting. This work will provide the basis for preparing the stage for a full exploitation and deployment of such functionalities.

2.1 Machine Learning

Machine learning is a sub field of computer science, a type of Artificial Intelligence (AI), that provides machines with the ability to learn without explicit programming. Machine learning evolved from pattern recognition and Computational Learning Theory and collocates itself in the interface between mathematics and computer science.

Following the typical AI path described in Figure 14, ML has the aim to learn from experience through a specific learning process.

This process is synthesized in the last steps of the path (represented in Figure 15) and comprehends a set of stages which involve different algorithms and approaches according to the specific situation and aim, and in particular to the quantity and quality of data available.

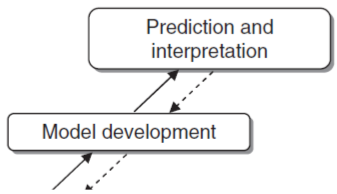


Figure 15- last stages of Data Mining

The steps of the learning process are:

- Training of a selected model with the previously transformed data
- Testing of the model using a sub-set of data in order to validate performances linked to the ability of the model to be accurate and general at the same time, with the right trade-off.
- Eventual comparison between different models trained and tested, using specific performance indicators.
- Feeding the developed model with new data in order to derive knowledge and conclusions.

Before entering the details of the possible different ways these steps can be performed, it is worthy to concentrate on the upstream stages of the data mining process, which are common for every kind of machine learning approach, independently on the availability of data and the specific purpose.

After a proper setting of the objective, and the identification of needed information, the unstructured and not easily interpretable amount of data must be structured and ordered. Here starts the Data preparation phase.

2.1.1 Data preparation phase

It is aimed at solving typical problems related to collected data, such as:

- Incomplete data
- Noisy data (Outliers)
- Inconsistencies
- Excessive amount of variables/records

Incomplete data

Some records may contain missing values corresponding to one or more attributes, and there may be a variety of reasons for this. It may be that some data were not recorded at the source in a systematic way, or that they were not available when the transactions took place. In other instances, data may be missing because of malfunctioning recording devices. It is also possible that some data were deliberately removed during previous stages of the gathering process because they were deemed incorrect. Incompleteness may also derive from a failure to transfer data from the operational databases to a data mart used for a specific business intelligence analysis.

In order to treat these data, it is important to have an acknowledgment of the reason behind, and consequently operate in a consistent way. Since missing data have to be managed in order to be recognized by the machine and the algorithm, here are listed the possible solutions:

- **Elimination.** It is possible to discard all records for which the values of one or more attributes are missing. In the case of supervised learning, it is essential to eliminate a record if the value of the target attribute is missing.
- **Inspection.** Alternatively, it is possible for experts in the application domain to inspect each missing value to obtain recommendations on possible substitute values. Obviously, this approach suffers from a high degree of subjectivity, and is rather burdensome and time-consuming for large datasets. On the other hand, experience indicates that it is one of the most accurate corrective actions if skillfully exercised.
- **Identification.** As a third possibility, a conventional value might be used to encode and identify missing values, making it unnecessary to remove entire records from the given dataset. For example, assigning the value $\{-1\}$ to all missing data. This solution is particularly interesting in case of categorical attributes.
- **Substitution.** Several criteria exist for the automatic replacement of missing data, although most of them appear somehow arbitrary. For instance, missing values of an attribute may be replaced with the mean of the attribute calculated for the remaining observations. This technique can only be applied to numerical attributes, but it will clearly be ineffective in the case of an asymmetric distribution of values. Finally, the maximum likelihood value can be used, estimated using regression models or Bayesian methods, but it can become rather complex and time-consuming for a large dataset with a high percentage of missing data.

Noisy data (outliers)

Data may contain erroneous or anomalous values, which are usually referred to as *outliers*. Other possible causes of noise are to be sought in malfunctioning devices for data measurement, recording and transmission. The presence of data expressed in heterogeneous measurement units, which therefore require conversion, may in turn cause anomalies and inaccuracies.

Ways for identifying outliers are usually suitable only in case of numerical attributes having a normal distribution. These ways are based on the computation of the dispersion:

- Central limit theorem [9]
- Z-index [9]

Extension to other distributions is provided by the Tchebysheff theorem, allowing to identify those records that are farer from the sample mean than a certain acceptable threshold.

Multidimensional analysis, comprehending more than one attribute at a time, can be done through clustering techniques which derive homogeneous groups of observations. The ones which remain alone are more likely to be considered outliers. Clustering techniques will be further explored in next section as they represent one learning approach based on unsupervised learning.

Only in case of certainty about the reasons behind, noisy data can be substituted with the values that would have been expected if the particular event which caused the presence of outliers never occurred.

Inconsistencies

Sometimes data contain discrepancies due to changes in the coding syntax used for their representation, and therefore may appear inconsistent. Also, in case of numerical attributes, inconsistencies may occur when different variables have completely different order of magnitude.

To solve this last problem, some techniques such as (i) Decimal scaling, (ii) Min-max method (normalization), (iii) Z-index method (also known as standardization, which is only applicable to normally distributed data), can be applied. [9]

Excessive amount of variables/records

When dealing with a small dataset, the transformations described above are usually adequate to prepare input data for a data mining analysis. However, when facing a large dataset, it is also appropriate to reduce its size to make learning algorithms more efficient, without sacrificing the quality of the results obtained. In these cases, Data reduction is a fundamental step with three main objectives:

- Increase the efficiency in model identification.

  The reduction of the number of total tuples of a dataset can significantly decrease the computational time for each model training iteration, and this computational time reduction is exponential in the training phase if the algorithm uses functions of higher complexity.

- Preserve model accuracy.
  In most applications, the accuracy of the models generated represents the main criterion followed to select one class of learning methods over another. Data reduction techniques should not significantly compromise the accuracy of the model generated. In some cases, data reduction techniques based on attribute selection will lead to create models with a higher generalization capability on future records.
- Achieve simpler models.
  In some data mining applications, models should produce rules easily interpretable by experts in the application domain. Data reduction often represents an effective technique for deriving models that are more easily interpretable.

Depending on the specific aim, data reduction can be performed in different directions.

- Sampling techniques can be applied for **reducing the number of records**, distinguishing between pure sampling and stratified sampling in those supervised learning cases in which the analyst wants to preserve the same proportion between different classes.
  In general, a sample comprising a few thousand observations is adequate to train most learning models. It is also useful to set up several independent samples, each of a predetermined size, to which learning algorithms should be applied. In this way, computation times increase linearly with the number of samples determined, and it is possible to compare the different models generated, in order to assess the robustness of each model and the quality of the knowledge extracted from data against the random fluctuations existing in the sample. It is obvious that the conclusions obtained can be regarded as robust when the models and the rules generated remain relatively stable as the sample set used for training varies.

- **Attribute reduction** can be performed in different ways. The idea is to keep only those attributes that are significant for the learning process, i.e., those features that are able to describe and influence the desired output of the model. Attribute selection can be done by the user through filtering methods (which make use of linear correlation index) and present some disadvantages linked to a simplistic management of the problem and to the fact that the algorithm is only applied after the attributes have been discarded. A more accurate set of methods performed by the user is called Wrapper Methods and consists in continuously applying the algorithm and evaluating each time the consistency between the model and the selected attributes. Relevant attributes are chosen after running several times the model. The drawback of this approach is surely the complexity and the amount of time and effort required.

A great advantage in this direction is given by the so-called Embedded methods: some algorithms automatically perform the feature selection in the same moment they are trained. The problem with these methods is that they use heuristic and greedy search scheme like forward elimination, backward and forward-backward elimination. In this way there is no guarantee that the selected variables are the most relevant ones.

In practice, filter methods are the best choice when dealing with very large datasets, whose observations are described by many attributes. In these cases, the application of wrapper methods is inappropriate due to very long computation times. Moreover, filter methods are flexible and in principle can be associated with any learning algorithm. However, when the size of the problem at hand is moderate, it is preferable to turn to wrapper or embedded methods which afford in most cases accuracy levels that are higher compared to filter methods.

- Dimensionality reduction

  Principal component analysis (PCA) is the most widely known technique of *Dimensionality reduction* by means of projection. The purpose of this method is to obtain a projective transformation that replaces a subset of the original numerical attributes with a lower number of new attributes obtained as their linear combination, without this change causing a loss of information. The idea is to project attributes in the direction(s) of maximum variation, in order to have a clear representation of data with less space needed. Experience shows that a transformation of the attributes may lead in many instances to higher accuracy in the learning models subsequently developed, and a decreased training phase time of processing since less attributes are present.

Before applying the principal component method, it is expedient to standardize the data to obtain for all the attributes the same range of values, usually represented by the interval [−1,1], with the mean of each attribute made equal to 0 by applying the transformation.

PCA is widely used in case of image data. On the other hand, it is discarded in those cases in which a clear interpretation of the model is needed.

- Data discretization. For reducing the total number of records, in case of numerical attributes one can build classes of equal size or equal width (related to values assumed) and substitute the classes with just one value. In case of categorical attributes, a hierarchical discretization can be performed.

2.1.2 Exploratory data analysis

The primary purpose of exploratory data analysis is to highlight the relevant features of each attribute contained in a dataset using graphical methods and calculating summary statistics, and to identify the intensity of the underlying relationships among the attributes. Exploratory data analysis includes three main phases:

1. **Univariate analysis**. The properties of each single attribute of a dataset are investigated. Attention must be posed on the fact that categorical and numerical variables must be treated in distinct ways. It is important in this phase, with the support of the domain expert, to understand if the distribution of each numerical variable is coherent with the phenomenon under investigation. It is crucial to understand if there is any biased trend in the data with respect to a nominal condition, and in some cases you want to train the model on the most general data points in order to generalize well for future example of data points, so some transformations could be suggested by the domain experts in order to fit the data into a more typical distribution.

   For categorical attributes, main verifications concern the computation of *empirical frequencies*, and the level of heterogeneity between different classes, which is measured through Gini index or Entropy index. The best case is that in which one has maximum homogeneity.

   Numerical continuous variables are visually analyzed through histograms and can be quantitatively evaluated thanks to a large set of indicators covering (i) measures of central tendency, (ii) measures of dispersion, and (iii) indicators of relative location.

   Another important and very used visual tool is the box and whisker plot, useful for visually detecting outliers and understand the right distribution of the attribute under analysis.

Another role of box-and-whisker plot is the assessment of symmetry/asymmetry of data.
Non graphical tools exist for the same purpose, like asymmetry index (also known as skew index).

2. **Bivariate analysis**. Pairs of attributes are considered to measure the intensity of the relationship between them. For supervised learning models, it is of particular interest to analyze the relationships between the explanatory attributes and the target variable. Even relationships between pairs of independent features are of particular interest and they are usually evaluated through a widely used graphical representation method which is scatterplot diagram. Another graphical tool is the LOESS diagram (local regression diagram), which tries to identify the trend of the two considered variables in a more detailed way compared to simple scatterplot, through the use of parameters describing the degree of the polynomial and the size of the neighborhood.
3. **Multivariate analysis**. The relationships holding within a subset of attributes are investigated. Widely used methods for such kind of evaluation are (i) scatterplot matrix (n dimensions), (ii) star plots which consist in designing each observation for every attribute in order to have a clear view of the presence of recurring patterns among distributions of the observations, and (iii) spider web charts. These last represent the most effective approach in those cases in which the dataset contains very large number of records.

2.1.3 Selection of attributes and feature extraction

After the exploratory data analysis, evidence could arise of the need to further eliminate some variables due to redundancy of information, or to correlation. Also, the need to extract features from available attributes can be satisfied through the creation of additional columns of data.

At this point, the peculiar stages related to the learning process itself are ready to be performed. As mentioned, different approaches correspond to different objectives, but also to inequal availability of data.

2.1.4 Model development

Machine learning algorithms can be divided into categories basing on their scope. *Predictive* models are used for those applications in which the prediction of a value is required, and it is derived through the use of other variables present in the dataset. These algorithms manage to discover and model the existing relationship between the target variable (the one whose value is going to be predicted) and other features. Even though the word "prediction" can remind to some kind of forecast, it is not always true that predictive models are used to foresee future events or values assumed by some variables. Indeed, predictive models can be used to model past events, like for instance deriving the time at which a command was issued, given the start of the inducted action. Also, these algorithms can be used to model real time events. These algorithms know exactly the variable whose values are going to be modeled and receive precise instructions during the training phase of the behavior of this target variable. For this reason, such kind of models are called ***supervised learning models***.

Descriptive models are used for applications which would gain from a knowledge obtained through a new and structured representation of data. Differently from predictive models, which observe and predict a specific variable they are interested to, in descriptive models no target variable is present, and each feature of the model assumes the same importance in defining the objective. For this reason, descriptive algorithms are called ***unsupervised learning models***. These models are widely used both in standalone applications, and as preliminary steps to supervised models.

2.1.5 Learning Process

As mentioned, given a set of data, a supervised learning algorithm attempts to optimize a function (i.e., the model) in order to find the combination of values of the features which produce the output representing the target. During the training phase, the target variable is fed into the algorithm together with the independent attributes, so that the algorithm can develop and improve the model (learning process). After the training phase, by feeding the algorithm with unlabeled observations, i.e., set of attributes without the target variable, it should be able to predict the target according to the learnt relationships.

Unsupervised learning algorithms usually rely on iterative processes, which aim to strengthen and get better at each iteration through the adjustment of important parameters.

Supervised learning

Supervised learning algorithms can be divided basing on the nature of the target variable they want to predict. In case of categorical variable, *classification* is used. As suggested by the name, the target variable represents the class to which the related observation belongs to. If the target variable is numerical and continuous, the selected approach is *regression*.

Classification:

Classification usually has a twofold objective:

- Make predictions on new data coming from the same source of the training set;
- Better understand the domain involved in the analysis.

The purpose of each classification algorithm is to define a set of possible functions able to learn the separation of observations in different classes. It is common to automatically think about binary problems, i.e., separations done basing on binary targets such as black-white, yes-no, ill-healthy, etc. however, it is also frequent to have multi-category classification problems. However, classification algorithms can only perform with binary target, thus, any multi-category problem can be reduced to a set of binary problems. More specifically, two approaches can be followed to split the multi-category problems into a set of binary ones: (1) One-against-all: consists in training k= number of classes different algorithms. (2) One-against-one: the effort is greater since $\frac{k(k-1)}{2}$ binary classifiers must be trained, but it can lead to better results.

In Figure 16, a synthesized description of the structure of any classification method is provided.

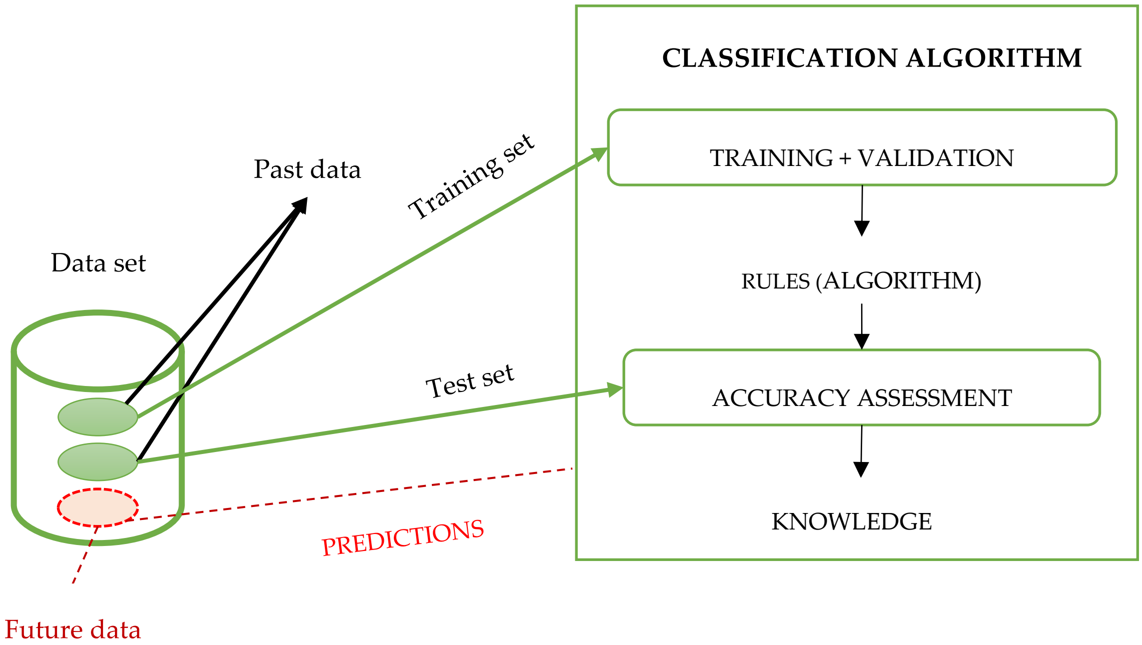


Figure 16- Classification methods

Assessing the goodness of classification methods is a key activity since the drown conclusions about the ability of the algorithm to learn correctly strongly depend on the criterion used.

When the future becomes present and then past, it is possible to perform a check to verify whether the predictions were aligned with the expected *Accuracy*.

To judge the performances of classification algorithms, thus their prediction accuracy, the suggestion is to consider different rules at a time.

The loss function (L) counts the number of misclassifications, i.e., the observations falling in the wrong class.

Empirical error (R_{emp}) is an average measure of the error, expressed by the sum of all the errors divided by the total number of observations.

Accuracy is equal to 1 - R_{emp}, thus providing a measure of the goodness of the method.

Nevertheless, the accuracy measured through these indicators does not allow to have a view on the actual quality of prediction. Indeed, the abovementioned measures can be suitable only if the two considered classes have almost the same size. On the contrary, they are quite useless in case of unbalanced population classes.

More reliable ways to assess accuracy exist, and they mostly rely on graphical representation.

a) CONFUSION MATRIX

It indicates the number of records correctly classified and those that have been misclassified, considering the proportion with the total number of observations actually falling in the different classes.
Rows represent the real values, while in the columns are registered the predicted ones.

		predictions		
		−1 (negative)	+1 (positive)	total
examples	−1 (negative)	p	q	$p+q$
	+1 (positive)	u	v	$u+v$
	total	$p+u$	$q+v$	m

Figure 17- Confusion Matrix

In particular, it allows to derive important indicators such as:

- Precision. Looking at the column of the positive:

$$\frac{T_p}{T_p + F_p}$$

T_p = true positive (those records classified as 1 that are actually 1);

F_p = false positive (those records classified as 1 that are actually 0).

- Recall, also called true positive rate:

$\frac{T_p}{T_p + F_n}$	T_p = true positive (those records classified as 1 that are actually 1); F_n = false negative (those records classified as 0 that are actually 1)

- F-measure is defined as:

$$\frac{(\beta_2 - 1) \cdot T_p \cdot precision}{\beta_2 \cdot precision + T_p}$$

Where $\beta_2 \in [0, \infty)$ regulates the relative importance of the precision with respect to the true positives rate. Of course, the higher this value, the better the accuracy of prediction. The F-measure can also be equal to 0 if all the predictions are incorrect.

b) ROC Curve

Receiver operating characteristic (ROC) curve charts allow the user to visually evaluate a classifier's accuracy and compare different classification models. A ROC chart is a two-dimensional plot with the proportion of false positives (fp) on the horizontal axis and the proportion of true positives (tp) on the vertical axis. The point (0,1) represents the ideal classifier, which makes no prediction error since its proportion of false positives is null $\left(F_p = 0\right)$ and its proportion of true positives is maximum $(T_p = 1)$. The point (0,0) corresponds to a classifier that predicts the class $\{-1\}$ for all the observations, while the point (1,1) corresponds to a classifier predicting the class $\{1\}$ for all the observations.

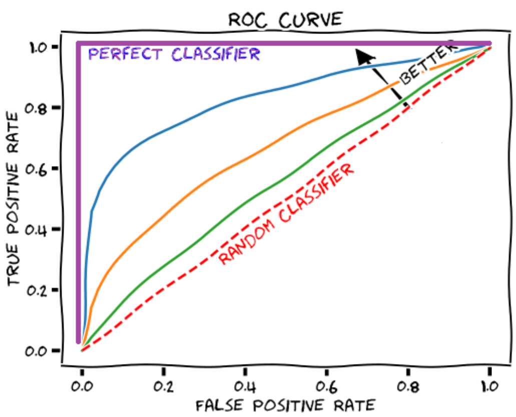


Figure 18- ROC curve and underneath Area [11]

c) Cumulative Gains

Graphically represent the gain achieved in the prediction using different percentages of the records in the dataset. It is also a measure of speed and robustness of the method, since it allows to understand whether it's worthy or not to use the entire dataset.

d) Lift

This plot is derived from the Cumulative Gains one. On the x-axis are still reported the percentages of the records taken into consideration, while on the y-axis, ratios are computed between the values on the y-axis of the cumulative gain chart and their related values on the abscises. It is a measure of how better your algorithm is with respect to the random selection method.

Analyzing the classification method illustrated in Figure 16, attention must be posed on how to split the set of past data used to run the algorithm between training and test set. The training set is used to practically make the algorithm learn the relations between variables. Errors and the abovementioned measures of the goodness of the model are computed for the predictions made using this subset and it's usually greater in terms of number of records than the test set. This last is used to double-check the developed algorithm, always computing and drafting the graphs needed to evaluate the method.

It is straightforward that the decision on how to split training and test sets has an influence on the related results. Different methods are possible, having pros and cons leading them to be more or less suitable for specific cases.

1. Holdout methods randomly select test set from the entire available dataset. It is not robust since the measured accuracy strongly depends on the random extraction of the dataset.
2. Repeated random sampling methods are more robust since the accuracy is measured many times, varying at each iteration the choice of the test set. Nevertheless, the time and effort required are greater than for the holdout method.
3. K-fold cross validation is the best practice since it is a systematic procedure, less dependent on the random extraction. The method is the following: the entire dataset is divided into K folds (groups) and at each iteration, one-fold is used as test and all the others as training set. This method can be used in case of small/medium datasets (maximum 1000 observations), with an optimal K

almost equal to 10. In case of big datasets, it is however preferable to go with holdout methods, allowing to have a large test set which avoids fluctuations.

4. A particular case of K-fold cross validation is the "Leave-one-out" method. K is set equal to m, i.e., the number of observations. In this way, each single observation is used as test once. It is obviously suitable only in case of very small datasets.

The learning process can be very different considering the set of existing algorithms performing classification task.

In almost every algorithm, preliminary decisions must be taken by the data analyst about some values to be set. These values are called **hyperparameters** and represent levers to play with in order to increase the performances and the potential goodness of each algorithm.

CLASSIFICATION ALGORITHMS:

- **K-Nearest Neighbour** classifier. It is a heuristic method which makes use of the notion of distance in order to determine whether observations are similar or not.

  Hyperparameters to be tuned in this case are:
  - **K**, which is the number of observations that are considered as "neighbourhood", i.e., the number of observations taken into consideration when assigning the next observation to a certain class.
  - **The notion of distance**. Distance can be measured through different methods like Euclidean distance, Mahalanobis, ecc.

- **Classification trees** are heuristic, iterative and greedy algorithms. Starting from the whole set of data, they compute the impurity index on the root node (i.e., the one embedding all the observations), and it is set as the upper bound for the next split, in order to compute the information gain.

$$\begin{aligned}\Delta(q, q_1, q_2, \ldots, q_K) &= I(q) - I(q_1, q_2, \ldots, q_K) \\ &= I(q) - \sum_{k=1}^{K} \frac{Q_k}{Q} I(q_k).\end{aligned}$$

Equation 3- information gain computed through impurity index

Iteratively, impurity indexes are computed for the nodes generated by the splits done basing on all the different attributes present in the dataset. Through the impurity index, the information gains achieved through the different attributes are also computed. The attribute which maximises the information gain is selected as split criterion. In this case hyperparameters are

- **Impurity index (Gini, Entropy)**
- **A set of pruning criteria to avoid overfitting:** choice of the maximum number of leaves, maximum depth of the tree, maximum purity, and maximum gain. This last represent the threshold for avoiding further splits which would lead to an increase in the information gain which is not worthy to be reached.

- **Bayesian methods** calculate the posterior probability $P(y|\underline{x})$ through Bayes Theorem. Where x is the observation vector (of size n= number of attributes related to the observation) and y the specific target class. In this case no Hyperparameters are present, but human effort is largely required for estimating the probability that a specific observation occurs, given that it belongs to a certain class ($P(\underline{x}|y)$). In fact, $P(y|\underline{x})$ can be only computed starting from that probability and applying the MAP hypothesis (maximum a posteriori), meaning that $\underline{x}$ is assigned to the class y if and only if y is the most probable value among all the values belonging to the target variable. Huge computational power is needed to compute these probabilities, especially in case of big number of attributes (large dataset).
  More advanced versions of Bayesian methods have been developed for reducing the computational power and time required (**Naïve Bayesian method**) and for introducing some flexibility to the concept of conditional independence, allowing reticular hierarchical links which assign selected stochastic dependencies between explanatory (independent) variables (**Belief networks**).

- **Logistic regression** is a probabilistic method like Bayesian methods and it allows to convert binary classification problems into linear regression ones. It postulates that $P(y|\underline{x})$ follows a logistic function. Computing the ODDS Ratio, it ends up with the formula $z = \underline{\mathbf{w}}' \, \underline{x}$, where z is the dependent variable (target one) and $\underline{x}$ is the vector of independent variables. Weights w, which represent the linear term connecting dependent and independent variables, are computed using an iterative method aimed at maximizing the likelihood. This technique presents all the problems related to linear regression, which will be briefly explained in next section. (Especially multi-collinearity of independent variables).

At this point, an improvement in the development of algorithms is introduced.

So far, classification algorithms were aimed at minimizing the empirical error, i.e., the number of misclassified observations, without considering the generalization error. This last is fundamental in order to prevent the algorithm to perform overfitting and to overcome the problem of robustness of the solution.

Algorithms aimed at minimizing the **structural risk**, which is the sum of empirical error, overfitting and ill-posed risks have been developed.

- **Support vector machines.** The principle used is the SRM= structural risk minimization, with a modified risk function represented in Equation 4.

$$\hat{\boldsymbol{R}}(\boldsymbol{f}) = \frac{\boldsymbol{1}}{\boldsymbol{m}} \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{m}} \boldsymbol{V}\big(\boldsymbol{y_i}, \boldsymbol{f}(\boldsymbol{x_i})\big) + \boldsymbol{\lambda}\, ||\boldsymbol{f}||_{\boldsymbol{k}}^{\boldsymbol{2}}$$

Equation 4- modified risk function

Lambda is the regularization term used to tune the trade-off between empirical error and generalization capability and it is an hyperparameter of the method. SVM bases on the notion of hyperplanes which separate the dataset into correct classes. The nearest the separating hyperparameters are to the training set, the lower the generalization capability. The aim is to maximise the minimum distance in order to increase the generalization capability while keeping the right level of accuracy of predictions. Training points lying on canonical supporting hyperplanes are called support vectors.

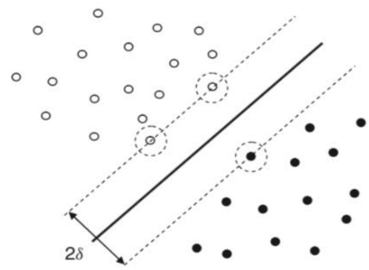


Figure 19- Canonical hyperplanes and support vectors

In the most probable case of not linearly separable points, the need is still to separate them through a linear function. For this reason, use is made of linear kernels which help to easily linearize complex functions. Kerner, indeed, are functions for which the mapping of original observations into transformed space is not explicitly computed. This allows very efficient linear separation also in infinite-dimensional space. Examples of kernels, which represent another hyperparameter of the process, are (i) Polynomial Kernels (of degree d); (ii) Radial basis function kernels (rbf); (iii) Neural Networks Kernels with hyperbolic activation function.

- **Neural networks.**

  A neural network is an oriented graph consisting of nodes which, in the biological analogy represent neurons. Nodes are connected by arcs, which correspond to dendrites and synapses. Each arc is associated with a *weight*, while at each node an *activation function* is defined which is applied to the values received as input by the node along the incoming arcs, adjusted by the weights of the arcs. The training stage is performed by analyzing in sequence the observations contained in the training set one after the other and by modifying at each iteration the weights associated with the arcs.

  The **perceptron** (Figure 20) is the simplest form of neural network and corresponds to a single neuron that receives as input the values (x_1, x_2, … , x_n) along the incoming connections, and returns an output value f(**x**). The input values coincide with the values of the explanatory attributes, while the output value determines the prediction of the response variable y. Each of the n input connections is associated with a weight w_j. An activation function g and a constant θ, called the *distortion*, are also assigned.

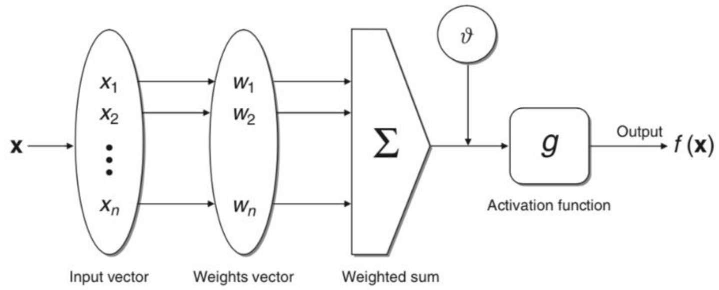


Figure 20 - Perceptron

Supposing that the values of the weights and the distortion have already been determined during the training phase, the prediction for a new observation **x** is then derived by performing the following steps.

First, the weighted linear combination of the values of the explanatory variables for the new observation is calculated and the distortion is subtracted from it, as reported in Equation 5:

$$w_1x_1 + w_2x_2 + \cdots + w_nx_n - \theta = \boldsymbol{w}'\boldsymbol{x} - \theta$$

Equation 5- linear combination of predictors net of distortion

The prediction f (**x**) is then obtained by applying the activation function g to the linear combination of the predictors:

$$f(\boldsymbol{x}) = g(w_1x_1 + w_2x_2 + \cdots + w_nx_n - \theta) = g(\boldsymbol{w}'\boldsymbol{x} - \theta)$$

Equation 6 - prediction f(x)

The purpose of the function g is to map the linear combination into the set of values H = {v_1, v_2, . . . , v_H } assumed by the target variable, usually by means of a sigmoid, relu, sign or an hyperbolic tangent function.

An iterative algorithm is then used to determine the values of the weights wj and the *bias* θ, examining the examples in sequence, one after the other. For each example **xi** the prediction f(**xi**) is calculated, and the value of the parameters is then updated using recursive formulas that take into account the error yi – f(**xi**).

A *multi-level feed-forward* neural network (Figure 21) is a more complex structure than the perceptron.

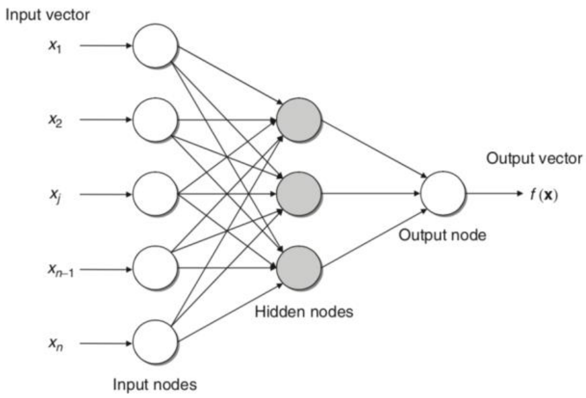


Figure 21- Multi-layer network

Each node of the network is a single perceptron, in the sense that given weights are associated with the input arcs, while each node is associated with a bias and an activation function. It is composed of one (passthrough) *input layer*, one or more layers of perceptron, called *hidden layers*, and one final layer called the *output layer*. It is called a *feed-forward network* because every layer, except the output one, includes a bias neuron and is fully connected to the next layer. In this way the signal flows only in one direction, from the inputs to the outputs.

Input nodes. The purpose of the input nodes is to receive as input the values of the explanatory attributes for each observation. Usually, the number of input nodes equals the number of explanatory variables.

Hidden nodes. Hidden nodes apply given transformations to the input values inside the network. Each node is connected to incoming arcs that exit from other hidden nodes or from input nodes, and it is connected with outgoing arcs to output nodes or to other hidden nodes.

Output nodes. Output nodes receive connections from hidden nodes or from input nodes and return an output value that corresponds to the prediction of the response variable. In classification problems, there is usually only one output node.

The method that determines the weights of all the arcs and the distortion at the nodes is called the *backpropagation* algorithm and follows a similar logic to that used for the single perceptron. The weights are initialized in an arbitrary way, for instance by setting their value equal to randomly generated numbers. The examples of the training set are therefore examined in sequence, using at each iteration the current values of the weights, to calculate the prediction and the

corresponding misclassification error. This latter is used to recursively correct the values of the weights, then used to analyze the subsequent example within the procedure. The weights are updated using a Gradient Descent algorithm, which is a variant of the gradient method.
Deep learning occurs when more than 3/5 hidden layers are set. Deep learning also learns the **representation** of the dataset (trainable features extraction) through a hierarchical abstraction of features.

- **Ensemble classifiers.** Are heuristic but very efficient methods which work combining predictions obtained by simple, heuristic classifiers like classification trees. Basing on the way they repeat and combine such methods, ensemble algorithms can be categorized distinguishing between:
  - **Bagging methods**: they split the dataset into subsets and make predictions in parallel, training different classifiers (or the same but with different hyperparameters set) using the different subsets. Then, all the trained classifiers are applied to the new observations sequentially, and the definitive class is assigned the target class through majority voting or through a weighted average of votes in case of higher importance of any trained algorithm. Examples of this method are Random Forest classifiers, which allow to gain good predictions but with very low interpretability.
  - **Boosting methods**: they make repeated predictions sequentially, using the whole dataset for each prediction or subsets of it. At the first iteration, all observations have the same weight. At each iteration, weights are updated basing on the errors performed during the previous iteration (higher weights for misclassified instances). To better explain, each observation has its own weight which is increased or decreased iteration by iteration basing on whether that observation is correctly classified or not. The weight represents the chance of being selected as part of the training set in the next iteration, thus increasing the accuracy of classification. The "learning rate" is a measure of how much the observations are weighted.

Regression:

Some of the considerations made for classification can be applied to regression. The method described in Figure 16 can also be valid for regressive algorithms. Same criteria can be used for splitting the dataset into training and test sets.

Nevertheless, regression has a slightly different purpose. Aim of this technique is to predict the numerical values assumed by a dependent variable, given a set of independent attributes describing the observations, through a proper function as shown in Equation 7. For this reason, regression appears much more complex and requires a higher degree of attention. Of course, methods for assessing the goodness of a regressive model are completely different from those used for classification.

$$\underline{Y} = f(x_1, x_2, x_3, \ldots, x_n), \text{ being } f\colon \mathbb{R}^n \Longrightarrow \mathbb{R}$$

Equation 7 - Function

A first distinction must be made regarding the fact of having just one independent variable (simple regression), more than one (general case of multiple regression). In the first case, function f is defined from $\mathbb{R} \Longrightarrow \mathbb{R}$.

Simple linear regression model can be represented by:

$$Y = w\,X + b + \mathcal{E}$$

Where $\mathcal{E}$ is the random residual variable.

Identifying the regression line means finding the right values of parameters w and b. (slope and intercept of the line).

One of the algorithms in charge of making such computation is the **Least-Squares Linear Regression**. It identifies the regression parameters (w and b) minimizing the sum of squared errors, i.e.,

$$SSE = \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} [yi - f\,(xi)]^2 = \sum_{i=1}^{m} [yi - w \times xi - b]^2$$

Equation 8- Sum of Squared Errors

However, the most common case is the case of multiple linear regression, which adds the complexity related to the fact of operating with vectors and matrices and to the possible, crucial, presence of correlations between different independent variables.

Multiple-least square linear regression applies the same principle aimed at minimizing the SSE.

This technique presents a lot of criticalities and issues:

- **ILL-POSED PROBLEM.** Stating that the problem is ill-posed means that the solution, whatever it is, will not be robust. That means that small changes in the input data (training set) result in large changes in the output (approximation and prediction). This is due to the fact that the technique is formulated for minimizing the SSE on training data, thus providing a "tailored" function which poses all the attention on accuracy and no attention on the generalization factor.

- **MULTICOLLINEARITY.** The matrix used to have pointwise estimates of the vector of regression parameters $\underline{\hat{w}}$. $(\mathbf{X'X})$ is not always invertible. However, it has to be inverted in order to find the estimates of regression parameters. The matrix is not invertible when columns are not orthogonal, thus when independent variables are actually correlated one to the other. Also, given that independent variables are not correlated, even if the matrix in in principle invertible, when the structure of its eigenvalues is very spread (i.e., there is too big difference between the biggest and the smallest eigenvalue), $(\mathbf{X'X})'(\mathbf{X'X})\neq\mathbf{I}$. This happens when the number of observation m is not enough.

For this reason, variants of the least squares regression techniques have been proposed, which take into consideration the *REGULARIZATION* term (Lambda) also seen in Classification techniques. This term is aimed at reducing the space of hypothesis, setting a limit on the norm of the function f.

Such methods are:

Ridge Regression, in which regression coefficients are obtained solving the minimization problem described by Equation 9:

$$\min_{\underline{w}} RR\,(\underline{w}\ , D) = \min_{\underline{w}} \lambda\,||\underline{w}||^2 + \sum_{i=1}^{m} (yi - \underline{w}'xi)^2$$

Equation 9- Ridge regression

Lasso Regression only differs for the norm of the vector of parameters.

Note that, with λ=0, the problem is reconduct to a Least-square regression.

Other techniques allow to adapt to the situation in which there is no linear relation between explanatory variables and target one. Gaining in flexibility, these techniques loose performances related to interpretability of results.

Such kind of algorithms are very similar to those seen for Classification problem:

- **K-NN for regression**

  Instead of having labels related to each observation, like in classification problems, numbers are associated to each point. In this case, the measure used to make the prediction is no more majority voting, rather a measure of central tendency used to assign the right number to a new observation. (Usually the average is adopted).

- **Regression Trees**

  Impurity indexes are substituted by a measure which maximises similarity among observations. Usually, this measure coincides with the **variance** S of each descendant note (i.e., the nodes resulting from a split). Basically, the algorithm searches for small variances in each leaf of the tree, allowing the observations in the same leaf to be as similar as possible in terms of values of the target variable y.
  The goal is thus to minimize the **total variance**, expressed as the sum of all the variances in all descendants.

- **SVM for regression**

  The aim is to set a fitting hyperplane and let all the points (observations) being as close as possible to it. A hyperparameter is set to define the admissible variation forming canonical hyperplanes. The distance ξ of each observation from the canonical hyperparameters is the error. It is nothing but an optimization problem like the one seen for classification: on the one hand we are trying to minimize the empirical error, on the other to increase the generalization minimizing the risk of overfitting and reducing the space of hypothesis.
- **Neural Networks for regression**

The only difference with neural networks for classification lies in the activation function. For the output node, there can only be linear activation function.

Unsupervised learning

Clustering

Purpose of clustering methods is the identification of homogeneous groups of records called clusters, by defining appropriate metrics and the induced notions of distance and similarity between pairs of observations. With respect to the specific distance selected, the observations belonging to each cluster must be close to each another and far from those included in other clusters. In other terms, each cluster has to be as homogeneous as possible within itself, while being very different from the other clusters.

Clustering methods can be classified into a few main types based on the logic used for deriving the clusters: partition methods, hierarchical methods, and density-based methods.

- **Partition methods**. Partition methods develop a subdivision of the given dataset into a predetermined number K of non-empty subsets. They are suited to obtaining groupings of a spherical or at most convex shape and can be applied to datasets of small or medium size. The most used partitioning methods are *K-means* and *K-medoids*.
- **Hierarchical methods**. Hierarchical methods carry out multiple subdivisions into subsets based on a tree structure and characterized by different homogeneity thresholds within each cluster and inhomogeneity thresholds between distinct clusters. Unlike partition methods, hierarchical algorithms do not require the number of clusters to be predetermined, but a similarity target level to be reached given in input by the user (hyperparameter). They are divided in agglomerative and divisive hierarchical methods. In both cases, they follow a dendrogram structure: in agglomerative methods each observation starts as a single cluster and then they are agglomerated into bigger clusters based on their similarity level, in the divisive process the method is the opposite, hence the whole set of observations is included in one big clusters, and at each step they are divided in two separate clusters at each step until the desired level of similarity is obtained.

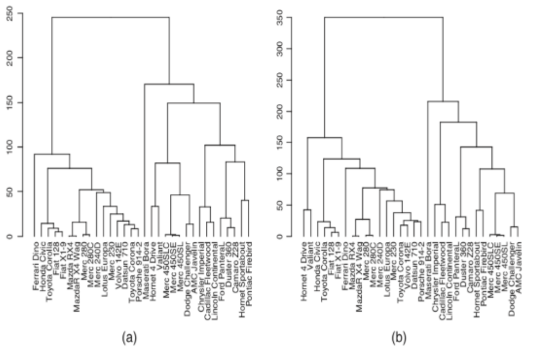


Figure 22- Dendrogram structure

- **Density-based methods**. Whereas the two previous classes of algorithms are founded on the notion of distance between observations and between clusters, density-based methods derive clusters from the number of observations locally falling in a neighborhood of each observation. More precisely, for each record belonging to a specific cluster, a neighborhood with a specified diameter must contain a number of observations which should not be lower than a minimum threshold value. Density-based methods can identify clusters of non-convex shapes and effectively isolate any possible outliers.

A second distinction can be made basing on the methods used for assigning the observations to each single cluster. It is possible to include each observation exclusively in a single cluster or to place it by superposition into multiple clusters. Furthermore, *fuzzy methods* have been developed which assign the observations to the clusters with a weight between 0 (the observation is totally extraneous to the cluster) and 1 (the observation exclusively belongs to the cluster), with the additional condition that the sum of the weights over all clusters be equal to 1. Finally, a distinction should be made between complete clustering methods, which assign each observation to at least one cluster, and partial methods, which may leave some observations outside the clusters

Associative rules

These unsupervised machine learning techniques are used for identifying regular patterns and recurrencies within a large set of Transactions. The fields of application are many and range between market basket analysis, web mining, fraud detection and healthcare services. For this reason, they are not of particular interest for the development of this work. However, they are based on an interesting principle which is called "A Priori Principle", and which allows to eliminate set of data (called itemset) exploiting the fact that if an itemset is not relevant, then all the itemsets of greater cardinality containing that itemset are not relevant too.

As mentioned, contrary to all the previously mentioned techniques, associative rules are not of relevant concerns for the aims of this work.

Novelty detection

In between the supervised learning process and the unsupervised one, Novelty detection is particularly used in industrial applications. Novelty detection can be applied for a specific area which has gained great concern in the last years: Predictive Maintenance.

This machine learning approach can be compared to a classification problem in which the available data used for training the algorithm only belong to one particular class. Algorithms deal with a supervised problem, called one-class classification problem, in which a training dataset is available, with the characteristic of presenting only (or mainly) "normal" (i.e., non-target class) behavior outcomes, and insufficient data describing the "abnormal" (target) ones.

For this reason, Novelty Detection (ND) can be defined as the task of recognizing that test data differ in some respect from the data that are available during training phase.

Approaches to novelty detection include both Frequentist and Bayesian approaches, information theory, extreme value statistics, support vector methods and neural networks.

In general, all these methods build models upon a training set that is selected to contain no examples (or very few) of the target (i.e., novel) class.

Novelty scores $z(\mathbf{x})$ are assigned to observation $\mathbf{x}$, and deviations from normality are detected according to a decision boundary that is usually referred to as the novelty threshold $z(\mathbf{x})=k$.

Different metrics are used to evaluate the effectiveness and efficiency of novelty detection methods. The effectiveness can be evaluated according to how many novel data points are correctly identified, and according to how many normal data are incorrectly classified as novel data. The latter is also known as the **false alarm rate**.

Novelty detection techniques should aim at having a high detection rate while keeping the false alarm rate low. On the other hand, the efficiency of novelty detection approaches is evaluated according to computational cost, and both time and space complexity.

As regards the Predictive Maintenance field of application, it is of particular interest in this work, due to its strong relation with Circular Economy needs. One of the objectives of CE, indeed, is to longer the product lifecycle and saving resources while gaining high quality results.

The adoption of Novelty detection methods for assessing the state of health of machinery and products represents an important enabler for achieving longer lifetimes. Collecting data about normal behavior of machinery/product while they are on health (healthy state), allows to build models which are able to identify the point in which these behaviors start to change, allowing to predict the eventual need of maintenance interventions.

In this work a preliminary issue will be tackled: the importance of the product/machinery design in determining the ability/inability of collecting such kind of data, and thus to apply Novelty detection techniques.

The approaches analyzed so far are part of the wide machine learning set of algorithms called **Discriminative models**:

Discriminative models

Discriminative models are those used for most supervised **classification** or **regression** problems. As an example of a classification problem, suppose to train a model to classify images of handwritten digits from 0 to 9. For doing so, a labeled dataset containing images of handwritten digits and their associated labels indicating which digit each image represents could be used.

During the training process, a specific algorithm will be used to adjust the model's parameters. The goal would be to minimize a loss function so that the model learns the **probability distribution** of the output given the input. After the training phase, one could use the model to classify a new handwritten digit image by estimating the most probable digit the input corresponds to, as illustrated in the figure below:

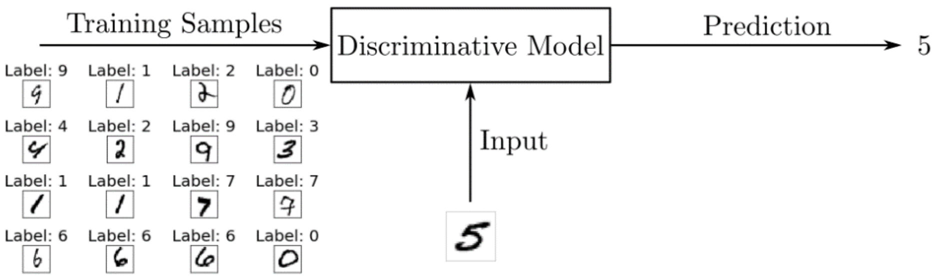


Figure 23- Discriminative model

Discriminative models for classification problems can be seen as blocks that use the training data to learn the boundaries between classes. They then use these boundaries to discriminate an input and predict its class. In mathematical terms, discriminative models in general learn the conditional probability $P(y|x)$ of the output y given the input x.

Besides these models, other objectives can be reached through the use of a different kind of algorithms, that is of particular interest for manufacturing industries: the Generative models.

Generative models

Generative models are trained to describe how a dataset is generated in terms of a **probabilistic** model. By sampling from a generative model, you're able to generate new data. While discriminative models are used for supervised learning, generative models are often used with unlabeled datasets and can be seen as a form of unsupervised learning. Using the dataset of handwritten digits, the training of a generative model to generate new digits can be performed. During the training phase, some algorithms will be used to adjust the model's parameters to minimize a loss function and learn the probability distribution of the training set. Then, with the model trained, one could generate new samples, as illustrated in the following figure:

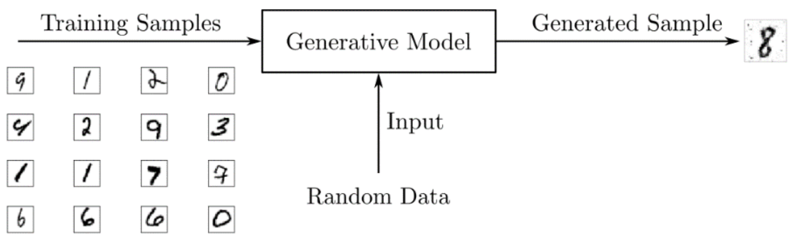


Figure 24- Generative model

To output new samples, generative models usually consider a **stochastic**, or random, element that influences the samples generated by the model. The random samples used to drive the generator are obtained from a **latent space** in which the vectors represent a kind of compressed form of the generated samples.

Unlike discriminative models, generative models learn the probability $P(x)$ of the input data x, and by having the distribution of the input data, they're able to generate new data instances.

A recent development of Generative algorithms is represented by Generative Adversarial networks.

Generative adversarial networks are machine learning systems that can learn to mimic a given distribution of data. They were first proposed in a 2014 NeurIPS paper by deep learning expert Ian Goodfellow and his colleagues.

GANs consist of two neural networks, one trained to generate data and the other trained to distinguish fake data from real data (hence the "adversarial" nature of the model).

In brief, GANs work this way: the generative algorithm is pitted against an adversary: i.e., the discriminative model that learns to distinguish whether the sample comes from the dataset or from the generative model. To provide an analogy: Generative algorithm is a team of counterfeiters, that produce fake currencies and the discriminative model is the police that is trying to distinguish whether money are fake (come from the model) or real (from dataset). Both models are trying to improve themselves until the counterfeits are indistinguishable from the real currencies.

Features can be learned automatically from the input data. In this area, mainly four different approaches can be differentiated:

(i) Multi-view images
(ii) Voxel models
(iii) Point clouds
(iv) Graphs

Simple GANs, however, generate random data, with no specified way. Conditional GANs enable feature specification to control the output of generative models. The idea of these conditional GANs was first presented in [12] where images are conditioned on their class labels.

Although GANs have received a lot of attention in recent years, they're not the only architecture that can be used as a generative model. Besides GANs, there are various other generative model architectures such as:

- Boltzmann machines
- Variational autoencoders
- Hidden Markov models
- Models that predict the next word in a sequence, like GPT-2

Furthermore, recent algorithms can generate new 3D objects. They make use of a combination of autoencoders and generative models, such as GANs, to create new 3D objects. An autoencoder is a deep learning architecture consisting of an encoder and a decoder that is trained to first encode the input into a low dimensional representation, and then to reconstruct the original input from this compressed latent representation again (learn the data representation).

2.1.6 Open Issues

Starting from general issues related to AI adoption, mentions will be made of the criticalities related to specific machine learning technologies.

General issues related to AI mostly involve barriers to adoption due to investments, security, ethics, accessibility, inclusion.

AI has the potential to accelerate shifts in market share, revenues, and profit pools. However, some issues are present as concerns the large-scale adoption of such kind of technology. AI can deliver real value to those companies that have a strong involvement in digitalization and that are able to combine strong digital capabilities with proactive strategies.

Therefore, the adoption of AI is more spread in firms and industries already on the digital frontier, but others are hesitant to act. Moreover, some key enablers to AI adoption have been identified [13]: leadership from the top, managements and technical capabilities, seamless data access. There are no shortcuts for companies for a profitable and efficient adoption of AI; changes are needed at corporate and business level. One of the strategies adopted by big giants of tech industry is the so called "acqui-hiring" process. This means that these big companies have been buying start-ups to secure qualified talents and to ensure the right technology acquisition, overcoming some of the main issues related to AI adoption.

In adopting AI, attention must be posed on ethics and security, beside keeping into account implications of such technology on humans and relations. Reliable AI has three fundamental components which should be met throughout the system's entire life cycle: (I) it should be lawful, complying with all applicable laws and regulations; (II) it should be ethical, ensuring adherence to ethical principles and values; and (III) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. In particular, the EU High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence, based on seven key requirements. [14] :

1. Human agency and oversight. Including fundamental rights.
2. Technical robustness and safety. Including resilience to attack and security, fall back plan and general safety, accuracy, reliability, and reproducibility.
3. Privacy and data governance. Including respect for privacy, quality and integrity of data, and access to data.

4. Transparency. Including traceability, explain-ability, and communication.
5. Diversity, non-discrimination, and fairness including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation.
6. Societal and environmental wellbeing. Including sustainability and environmental friendliness, social impact, society, and democracy.
7. Accountability Including auditability, minimisation and reporting of negative impact, trade-offs, and redress.

Besides purely social-oriented issues, even at operative level many companies do not have clear view of the kind of data to be collected nor the right consciousness of the importance of data collection itself.

Focusing on traditional manufacturing companies, it results evident the need to consider all the critical aspects in the definition of a structured plan for systematically collect data and feed them into the right algorithms.

The methodology presented in this work will deal with the strategic problems and implications of introducing AI and Machine learning in one of the most critical and relevant activities in a manufacturing company: product design.

Shifting to operative-level issues, many times companies are still unable to proper collect and organize their data, beside not having a clear view of which data they need to gather.

To capture value from AI, organizations need to establish digital processes, and an open culture around AI. On a technical level, appropriate processing power to handle all data inputs is needed. Machine learning algorithms can be very complicated and onerous in terms of time and computational power, therefor requiring the right infrastructures, and a certain level of efficiency in the data preparation phase.

Main issues related to Machine learning implementation are:

Insufficient Quantity of Training Data

For a human being to learn what a car is, all it takes is to point to a car and tell him that is a car (possibly repeating this procedure a few times). Now he/she can recognize cars in all sorts of colors and models. Machine Learning is not quite there yet; it takes a lot of data for most Machine Learning algorithms to work properly. Even for very simple problems, it typically needs thousands of examples, while for complex problems such as image or speech recognition millions of examples may be needed.

The Unreasonable Effectiveness of Data

In a famous paper published in 2001 [15], Microsoft researchers Michele Banko and Eric Brill showed that very different Machine Learning algorithms, including fairly simple ones, performed almost identically well on a complex problem of natural language disambiguation once they were given enough data.

As the authors put it: “these results suggest that it should be appropriate to reconsider the trade-off between spending time and money on algorithm development versus spending it on data quality.”

Nonrepresentative Training Data

To avoid the problem of having very precise predictions on the training set and bad ones on future data (test set and future observations), it is crucial that the training data are representative of the whole set of possible cases which may occur. For example, Figure 25 shows an attempt to find a linear function for modelling, thus predicting, the life satisfaction given the GDP per capita. To do so, different countries have been taken as training set (blue dots). The resulting model is not correct if other countries are considered (red squares). This is representative of the fact that, in order to build a general model, the right choice of training set is fundamental.

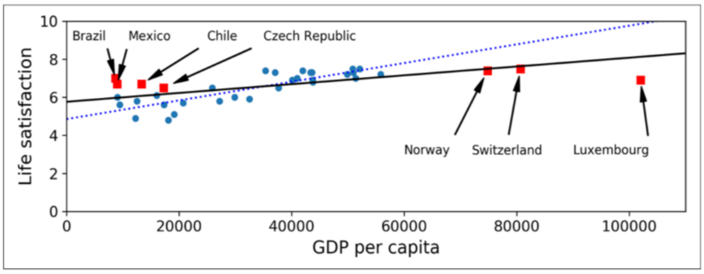


Figure 25- Predictions obtained through non-representative training data

The model trained on the uncomplete data points is represented by the blue dotted line, while the new model is represented by the black straight line. With respect to the old dataset, the model is significantly altered, thus implying that even for a simple linear regression model, a nonrepresentative training set leads to more inaccurate predictions.

It is crucial to use a training set that is representative of the situation that is willing to predict. This is often harder than it sounds: in case of too small sample data, *sampling noise* (i.e., nonrepresentative data given by chance) will occur, but even very large samples can be nonrepresentative if the sampling method is flawed. This is called *sampling bias*.

Overfitting

That's the risk when dealing with algorithms that concentrate all the effort in creating accurate models, without considering the generalization error.

Figure 26 graphically shows the effect of overfitting, and of its opposite issue, which is underfitting, in shaping the distribution of data. It is evident that, in the last scatterplot, future data may behave slightly different, and the model would not be able to properly predict this behavior.

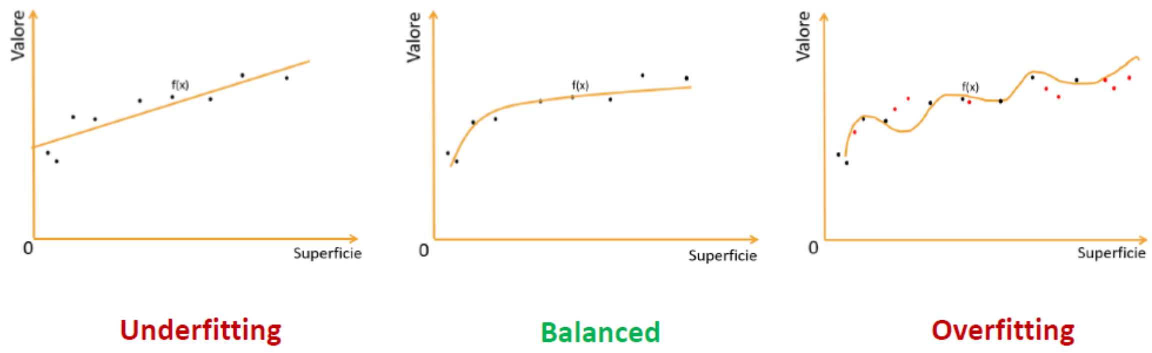


Figure 26- Overfitting and Underfitting models [16]

The correct model is the one which is able to be precise in shaping training data, but also to have a good generalization ability when it comes with new data. The solution lies in a correct choice of the algorithm and of its hyperparameters, besides the starting point which has to be a proper availability of data.

In practice, when training an algorithm, overfitting can be spotted, and consequently corrected through hyperparameters tuning, by looking at the output errors. If the model is very precise in predicting the training test (errors are low), but less precise in predicting the test set, it's almost sure that the algorithm performed overfitting.

As stated, a proper re-setting of hyperparameters can correct this issue, giving more importance to the generalization term than to the empirical error for those algorithms which have both the terms. Even in case of heuristic algorithms like Classification and regression trees, overfitting can be corrected through hyperparameters tuning, for instance reducing the maximum depth of the algorithm or reducing the maximum number of leaves.

Besides tuning the *hyperparameters*, other possible solutions to avoid overfitting are:

- To simplify the model by selecting one with fewer parameters (e.g., a linear model rather than a high-degree polynomial model), by reducing the number of attributes in the training data or by constraining the model
- To gather more training data

- To reduce the noise in the training data (e.g., fix data errors and remove outliers, data cleaning)

Nevertheless, the amount of regularization to apply during learning can be only controlled by *hyperparameters*, which represent parameters of the learning algorithm. As such, regularization is not affected by the learning algorithm itself; it must be set prior to training and remains constant during training. By setting the regularization hyperparameter to a very large value, the learning algorithm will almost certainly not overfit the training data, but it will be less likely to find a good solution. Tuning hyperparameters is an important part of building a Machine Learning system and is often performed using an iterative trial of different combinations of different hyperparameters settings (called GridSearch) in combination with K-fold Cross Validation.

3. Artificial Intelligence for Circular Economy

Previous disclosure on Circular Economy and Artificial Intelligence were fundamental to fully understand the reasons behind the developed work.

The starting point is the consciousness about the need, at manufacturing industry level, to change perspective and to move to a more sustainable and efficient management of resources, mainly intended as time and materials.

Having understood the potential of Machine learning as a tool to support this process, attention must be posed on how to deploy it in all its complexities and critical aspects.

First, in order to exploit Machine learning, high quality data are needed.

Where does the data exist?

This work builds upon the confidence that the way products are designed and thought determines the ability or inability to extract data and derive information from them, thus enabling or disabling the process of knowledge creation needed to perform any kind of upgrade in the business. In particular, the upgrade tackled in this work deals with the large-scale implementation of a Circular Economy approach in manufacturing industry.

Data exist in the product itself and in the market, intended as both end-users and B2B in case of products that are built to be deployed on more complex products, and in the entire supply chain. The criticalities lie in a proper identification of relevant data, in their collection, and in the subsequent manipulation done in order to extract knowledge. In many cases, problems arise since the very first step of this multi-stage process: companies are not able to recognize and properly collect the relevant data. The proposed method tackles this issue through the creation of an iterative, closed loop methodology. This allows to improve the gathering of data at each iteration, trying to exploit as many information as possible from already existing products currently on the market, while providing solutions for improving them through new configurations, leading to the ability to validate the selection of data, to gather them and to extract from them the related information at next iteration.

In order to build an efficient process, a clear definition of the objective is needed.

What well defined Circular Economy problems could Artificial Intelligence help to solve?

The areas of possible improvement of Circular Economy are many; Artificial intelligence has all the potential to be a change enabler. AI can play a fundamental role in the creation of efficient disposal and collection systems, in the reduction of **uncertainty** and **variability** that are typical of the reverse chain inputs, in the reduction of costs related to critical activities such as shredding for recycling materials, and in the reduction of human work required with the consequent speed up of processes like sorting and disassembling.

At the state of the art, some CE related problems have been discussed and attempts to use artificial intelligence as a tool for solving them have been made.

Nevertheless, in all the cases the proposed solutions are not well defined nor structured and occur as rough guidelines which do not consider the overall impacts at system level.

In the following, a brief overview of the problems attempted to be solved through machine learning is provided. Re-taking the taxonomy proposed in section 1.1.6 - Open issues related to CE, examples of Machine learning based solutions can be provided for almost every mentioned field. For an easier reference, CE problems that can be solved using Artificial Intelligence have been grouped into 3 categories representing the decision-making levels and the consequent degree of integration in a company:

- **Single machine level** (operational). It affects the single operation or sub-process.
- **Process level** (tactical). It affects a system of 2 or more sub-processes and their interaction.
- **System level** (strategic). It affects the whole company and / or the interaction with other companies in the supply chain.

3.1 Operational level

- **Automated cores classification**
  The issues related to the identification of valuable components and to their separation from the rest of the product have been discussed in the section dedicated to remanufacturing activity. Inspection usually consists of a

combination of visual and dimensional measurements, non-destructive tests and functional tests performed by a skilled operator who knows the core thanks to his knowledge-base about the original product, and with the support of standard operator sheets (SOS) reporting a checklist for the components' status. The challenges of this process are the high variability in the conditions of End-of-life products, the poor information about return products and increasing product variety. Machine learning is the ideal tool to face these challenges because it allows to speed up the decision-making process exploiting the machine data on processing power; indeed, it can support the operator in the process to increase his productivity (decreasing the average task time), or even substitute the operator with a completely automated process. However, as discussed in section 2.1.6, suitable training data covering all the cases must be provided to the algorithm.
Existing solutions have been searched in the manufacturing field, and the application which seems to be closest to the core classification is the surface defect detection tool of Landing AI, which is a company that provides customers with AI-powered industrial computer vision applications [17].

- **Automated disassembly**
  Product disassembly is one of the most manual labor dependent activities in remanufacturing, due to the complexity and variety of operations that must be performed on cores to obtain single components. Manual labor is generally more expensive than activities performed by machines, so the economic feasibility of disassembly tasks is hard to achieve. Also, flexibility is key for disassembly tasks and flexibility has a cost.
  Products disassembly can also be dangerous for the safety of operators due to the presence of hazardous components in products, mainly toxic materials, and sharp edges typical of a product or created by product breakages. The use of toxic materials inside products has been reduced in the recent years thanks to severer regulations, nevertheless many products are still around to be retrieved as cores (e.g., cathode-ray tube televisions containing phosphates and lead), and other products will continue to be produced even if toxic materials are required for their functioning, such as batteries.
  The most advanced solution found for automatizing disassembly is presented in the study by Vongbunyong et al. [18] regarding the prototype of an automated disassembly cell for LCD screens.
  The cell is provided with a 3D vision system that allows to: (i) detect the components of the product, (ii) detect connective components such as screws, (iii) determine changes in the disassembly state. Data collected from the vision

system are fed into the cognitive robotic agent powered by Golog, a programming language based on situation calculus. The actuator is composed of a grinder connected to a robotic arm, and a flipping table to remove disassembled components. In a series of experiments on 30 different models, the system has proven to be capable of disassembling LCDs with different components and different positioning of components as well, keeping a recognition accuracy higher than 90% and a position accuracy within 5 mm.

- **Predictive maintenance**

Predictive maintenance is a popular application of predictive analytics that can help businesses in several sectors to achieve significantly high asset utilization and savings in operational costs. Avoiding failures, a product can circle for longer in the market, following one of the four value creation mechanisms theorized by Ellen MacArthur Foundation [10].
Depending on the product and the context considered, the prediction could refer to different elements:

1. Anomalies detection in product or component performance and functionality.
2. Predict failures in the near future.
3. Estimate the remaining useful life of a product or component.
4. Identify the main causes of a failure.
5. Identify when and which maintenance actions are needed on the product or component.

Predictive maintenance potential has been considered extensively: in fact, predictive maintenance requires to continuously assess the status of components through set parameters, therefore the same information can be used to predict the quality state of a component even when it is in the hands of a customer. Moreover, the same product information gathered for predictive maintenance could feed the high-level operations planning of a company applying De-manufacturing, making the product return a quasi-deterministic variable.

Many companies are already providing predictive maintenance services based on machine learning, and they are mostly adopted for very valuable products, such as industrial equipment, trains, buildings cooling systems, etc.

- **Automated optical separation**

Separation in recycling plants is a process that can occur at two different levels depending on the company business:

High level. If the recycling company collects various end of life products and waste from the market, it consists in separating them grossly depending on the type of product and on the material of which they are prevalently made. Then the company can process the different products internally or send the aggregated similar materials to a specialized company for the recycling of such material.

Low level. Is the one described in section 1.1.4, when the recycling company, which receives in input similar products, separates the particles obtained from the grinding of such products to obtain in output a flow of particles of the same material as pure as possible.

There could be companies which perform both the two levels of separation. For example, Relight receives different types of lamps that must be treated separately because they could be made of different materials. After a high-level separation, lamps are shredded and a low-level separation is performed to separate target materials from non-target materials.

High level separation is usually performed by operators who pick manually objects from a conveyor belt and put them into bins depending on the object category or material. The throughput of this operation is constrained by the operators picking speed and the number of operators. Increasing the picking speed of operators increases the risk of alienation due to the repetitiveness and quickness of the task. The presence of hazardous materials and products to sort is another possible threat to the safety of operators.

Low level separation is usually automatically performed thanks to processes that exploit physical properties of materials (e.g., conductivity, magnetism, density, particle size, etc.). Such solutions could be useless or underperforming in some situations, for example when separating different kinds of plastics or when separating particles with different colors.

3.2 Tactical Level

- **Optimal disassembly level**

  As shown in Figure 11, setting the optimal disassembly level is one of the most critical decisions in De-manufacturing.

  To assess the optimal disassembly level, it is fundamental to determine the costs and revenues that drive this decision considering component by component. Only variable costs that arise from this decision are accounted:

Disassembly costs. They are the costs to disassemble the component, which are function of the task time for the disassembly, the setup time to have the core in the right position and to prepare the right tool to disassemble the component, and the energy required from tools to carry on such operations. Task time and setup time are linked then to the labor cost in case of manual disassembly, or to the machine usage and depreciation in case of automated disassembly.

Reconditioning costs. They are the costs to recondition the component to make it usable again and warranted like new. Examples of reconditioning costs are: (i) labor cost, dependent on the task time and the setup time, (ii) material cost, for instance if material addition is needed to refurbish surfaces, or if working fluid should be substituted or refilled, (iii) cleaning cost, which is usually performed several time during the reconditioning process, (iv) packaging cost to deliver the product to the customer, (v) energy and machine depreciation to perform various reconditioning operations previously mentioned, (vi) delivery cost to the customer.

Reconditioned components revenues. It is the gain from selling a reconditioned component.

Disposal cost. If a product cannot be reused or it is not convenient to be reused, a cost will be associated with his disposal, which can consist in recycling cost if the component's materials can be extracted and recycled, or in landfill cost. Moreover, the transportation cost could be accounted. Usually, those costs are embedded into the extended producer responsibility (EPR), which is a strategy used by many countries to add all the environmental costs associated with a product throughout the product life cycle to the market price of that product.

The use of machine learning algorithms for this issue is mainly focused on gathering high quality data to find a more accurate solution rather than on optimizing the existing solution. In fact, the optimization issue is addressed by genetic algorithms, which are widespread tools in the optimization field.

- **Flexible cores routing**
  As explained in the section dedicated to Disassembly, line configuration and setting is an important lever which acts on the profitability of the entire disassembly process.
  Keeping a fixed rule for the assignment of tasks to the disassembly stations could create strong workload unbalances generating queues, therefore

reducing throughput. Moreover, workload unbalances are a source of dissatisfaction for humans, leading to social problems which may undermine the sustainability of these solutions.

Using machine learning algorithms, it could be possible to create a flexible system for the routing of cores, gathering more precise data about cores status as explained in the optimal disassembly level section and predicting in a more reliable way disassembly task times and their variance thanks to historical disassembly data.

As regard the use of product data and other historical data to predict disassembly task times, no solution can be found in literature. However, as highlighted by Usuga Cadavid et al. [19], it is possible to find various studies about manufacturing production planning and scheduling focused on time estimation of tasks due to the increasing complexity of products and continuous pursuit of process efficiency.

- **Dynamic recycling routing**

  As explained in the dedicated section (1.1.4) the obsolescence risk of a fixed recycling system is a major threat for the owner of a plant or a possible investor. A rigid architecture could also reduce the recycling performance in case of a varying composition of the input material. Two elements are mandatory to create a flexible system with a dynamic routing:

  - A dynamically changeable transportation system which connects the stages of a recycling system that guarantees both the flexibility in the use of the system thanks to the possibility of routing selection, which allows the particles to visit only the stages needed and eventually to re-visit them, and it allows to add modules to treat different input materials following the trends of the offer of End-of-life products and components inbound.
  - An intelligent planner which can quickly define the best routing and machine parameters for each batch of materials inbound. The behavior of a recycling system stage, i.e., separation or size reduction, is hard to predict, except by using complex simulation tools, which require the modelling of physics behind the process that depend on particles characteristics (dimension, shape, material, orientation), on their interactions and on-stage parameters. A machine learning algorithm could quickly gather particles data from visual sensors and reading machine parameters to predict the output of the stage in a matter of milliseconds, therefore triggering the decision to keep the same routing and parameters or to change them.

No practical implementations nor theoretical results have been found in the De-manufacturing field. However, machine learning algorithms have been used to predict the recovery and in general the performances of material extraction processes in the mining industry.

3.3 Strategic Level

- **Selective reverse logistic**
  Machine learning algorithms could improve the reverse logistic process in two ways. The first is the prediction of the quantities of returned products at a retrieval point, which is hard to perform, and can be only made possible through a proper data collection. The second is the use of automated performances assessment solutions and predictive maintenance models to make a first rough core assessment at the collection point or even when the product is still in customer's hands.
  Reverse logistics optimization tools such as the network design process to find the optimal position and number of collection nodes minimizing shipment costs and warehouse costs, and the selection of logistic suppliers are other useful artificial intelligence tools to reduce reverse logistic costs. For example, Yanchao et al. study [20] reports a model for multi-objective optimization considering the location, frequency of collection, quality and quantity of materials collected for recycling. Another example from Jeong-Eun et al. presents a model for the optimization of reverse logistics networks through a hybrid genetic model [21].

- **Recycling exchange platforms**
  A wasted opportunity for circular businesses resides in the missing communication among companies of different sectors to create an extended supply chain. In other words, the power of cascaded use and the power of pure circles are not exploited at all. In fact, material that is considered the byproduct of a process for a company could be used as primary raw material for production by another company. Sometimes the company that treats the material as waste pays to get rid of the material that is discarded, while the company that purchases it pays to obtain the material which is extracted or obtained through an industrial process. This situation can be turned into a win-win condition if both companies would share their data on common platforms and match their requirements through machine learning algorithms, bringing systemic benefits and cost reduction to both parts. The solution here presented could be the incentive to boost downcycling processes which struggle to spread among industries.

Another scenario of value loss due to missing communication is the degradation of raw materials obtained through recycling processes that have worse properties compared to virgin raw materials. The discrepancy could be so high that the recycled material cannot be used for the same purpose as the virgin material, but it could still be used for lower value products. For example, plastics can be recycled few times, and the value depends on the polymer considered, before losing their properties such as tensile strength and transparency. The maximum number of cycles decreases if the plastic is recycled mixing different polymers or if the material is contaminated with additives.

A few waste exchange platforms exist online, however, one of the few examples of platforms using machine learning models to facilitate the match between demand and offer is WSX-BM, a European platform born to promote national and international waste supply chains for companies of different dimensions.

- **Remanufacturing demand forecast**

  The prediction capability of machine learning algorithms can be exploited also to have more reliable forecasts about customers' demand. In fact, the products demand for companies adopting remanufacturing gets more variegated with 3 main options for the customer: (i) newly manufactured products, (ii) remanufactured products, (iii) remanufactured and upgraded products. Various elements can have an impact on the customer's decision, that will have impact on the production and remanufacturing planning to make the company aligned with the market. For example, if most of the customers of a specific product are unwilling to buy that product remanufactured, it means that allocating many resources to core collection and remanufacturing is inconvenient, and that the purchase of remanufactured products should be incentivized.

  The key is to discover which data can explain returns and demand, and to collect them in a reliable way. Also, solutions can be found in a proper product design for lifecycle, allowing to quickly upgrade products even in those cases in which the rate of technological change is high, without the need to completely change its configuration and materials. Then, the strategical decision lies in aligning internal operations with returns and demand.

Having completed an overview of the potential implementations of AI for solving Circular Economy issues, it results clear that large-scale solutions are still far from being deployed.

The first step done had been finding a common factor that would combine all the above-mentioned issues.

This resulted in the confidence that the most relevant and impacting lever for solving almost every kind of issue is represented by how products are designed, and that product design must bring adequate transformations to the processes and system behind.

This wide and integrated concept embeds all the notions that are needed for tackling and targeting different issues, especially those at single machine and tactical level. Also, as explained in section 1.1.5 – Manufacturer centered approach, product re-design must be always supported by a consistent strategy and a farsighted business model setting.

In other words, acting on product design, building a structured framework which involves the use of artificial intelligence for re-thinking the way products are designed, represents a high-level decision which has impacts on all tactical and operative issues, allowing to easily deploy and implement all the solutions that are now hard to build and practice.

Also, even if it could seem unrelated, a proper product design can help in developing solutions for other strategical issues like Remanufacturing demand forecast, selective reverse logistic and recycling exchange platform. This is due to the strong relation which links product design to system and process design, bringing to the evolution of the whole value chain and business model. In this direction, product design also has the potential to transform the way different actors interact with each other and with the product and components.

Having said that, the answer to the third question is straightforward.

Which applications would benefit most right now from Artificial Intelligence?

All the Circular Economy related processes must be supported by a proper design of the products. Aim of the work is to exploit machine learning for enabling solutions which lay the groundwork for any other potential AI supported solution.

"A design is a plan or specification for the construction of an object or system or for the implementation of an activity or process, and/or the result of that plan or specification in the form of a prototype, product or process." [22]

Product design has a significant effect on de-and re-manufacturing systems since it affects the existing circular economy business options that the manufacturer can adopt as strategic actions to regain functions or materials from the post-use products. It also influences the selection of the technological solutions that should be adopted for a proper post-use product treatment, and the efficiency, effectiveness, and profitability of the de-and remanufacturing process. This lever constitutes a unique strategic competitive asset for manufacturers, especially in the case of manufacturer-centric circular economy businesses. By evaluating the effect of product design decisions on the post-use value and function recovery processes in advance, it is possible for the manufacturer to anticipate potential issues that can undermine the feasibility of the treatment and implement corrective design changes towards a sustainable business development, since the early-stage design phase of the product.

The links between product design and De and Re-manufacturing solutions have been investigated by many authors and researchers [23] [24] [25], within the broader field of "design for X" activities; in this work these research will constitute the starting point and the basis of the proposed methodology. For the sake of completeness, some important findings will be mentioned as regards the impact of particular features of the product on its ability to be easily disassembled and remanufactured. The starting point is the adoption of a 'lifecycle thinking' vision, which considers both product and process, in a co-evolutionary approach. Conclusions as regards the guidelines for a lifecycle-oriented design concern the **product materials** and **their coupling**, the **product structure** and **geometry**, the **fastening** and **joining methods** that can support an easier and more efficient disassembly and remanufacturing process [26]. In Figure 27, a synthesis of the impacts of different design aspects on Circular economy purposes is proposed.

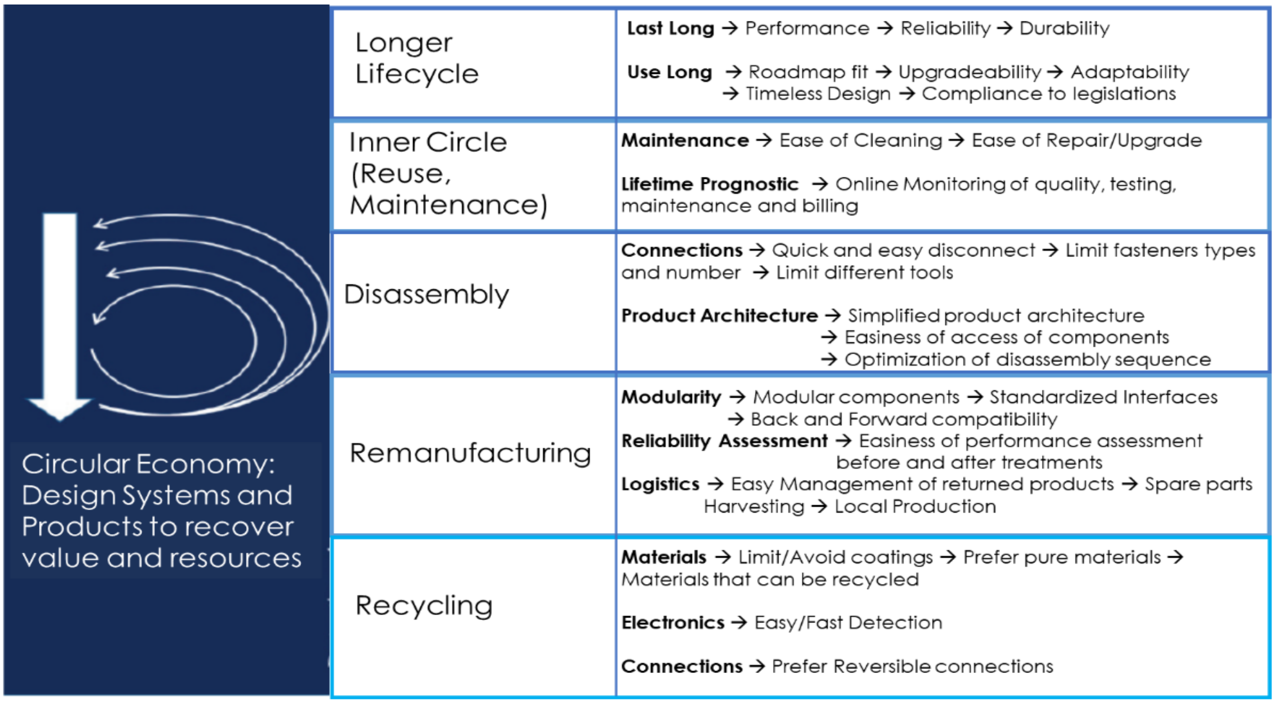


Figure 27- Circular product Design Tool [24]

Other pre-conditions to the application of a design for product lifecycle approach include the existence of a market and a demand from customers for the remanufactured product and the capabilities to simultaneously achieve benefits in terms of economic and environmental performance. The work will tackle all these aspects, presenting a methodology that is able to organize and manage all the criticalities and strategic issues.

Also, within the large field of "Design for X" research, besides the more traditional "design for disassembly" and "design for remanufacturing", supported by the concept of modular design, "Design for Upgradability" has recently been developed. It is a product design approach intended to easily upgrade product functions with the addition or the reconfiguration of one or more modules. Design for Upgradability uses modularity to conceive products that are thought to be remanufactured and upgraded. Practically speaking, the aim is to design products that can be easily cleaned, disassembled, repaired (through the substitution of broken modules) and upgraded (through the addiction of new modules or the change of modules). The idea of remanufacturing with upgrade is to extend products' value with their lifecycle, enabling the introduction of technological innovation into remanufactured products in order to satisfy evolving customers' preferences and, at the same time, preserving as much as possible the physical resources employed in the process [27]. Many authors have built models or tools to define the optimal modular structure of a product to be upgraded [28], [29], [30], [31], and [32]. Product design for upgradability should also support the diffusion of new "product-service systems" based circular economy business models, aiming at selling the product use instead of the physical product and offering product upgradability options throughout the product life cycle [33].

Design for product life-cycle methods are powerful enablers for the development of manufacturer-centric circular economy businesses through smart De and Re-manufacturing systems. It was shown through the analysis of specific real cases, that design for product life-cycle methods can improve the ability to automate De and Re-manufacturing processes [34].

However, as a matter of fact, despite the wide availability of these tools in the literature, concrete methods have not yet massively penetrated industrial practices. The main reason is the additional cost that they can introduce in the very critical phase of New Product Introduction, were most of the costs and business risks are absorbed by companies. Indeed, design for product life cycle may introduce additional constraints to the design process and this can translate into a more expensive product design process, under a short-term view. [35]

Also, the act of balancing multiple requirements and needs imposed by customers and government regulations on cost, space, aesthetics, and sustainability becomes harder.

Meeting these requirements and needs can involve novel design solutions and the expansion of the solution space beyond that of previous designs. This process can involve both exploration and exploitation, depending on the project situation and the capabilities of the design team. However, time and resource limitations can restrict the opportunities for exploration. This is the main reason why Machine Learning algorithms can help in this ambitious process, helping companies in overcoming time and resources constraints.

3.4 State Of the Art and Research Gaps

So far, state-of-the-art research have been presented as regards two macro topics: the use of Machine Learning for tackling specific Circular Economy issues and the implementation of design for product life-cycle methods. The link between these two, apparently disjoint, issues has already been explained and lies in the definition of this work. The aim is thus to exploit Machine Learning for enhancing the process of design for product lifecycle. Besides the already mentioned research and gaps in the effective implementation of design for product life-cycle methods, a literature review has been performed for identifying the current state in the utilization of machine learning for product design purposes. [36] [37] [38] [39].

Despite the huge number of articles mentioning AI, academic literature presents a clear gap as regards the adoption of artificial intelligence for industrial purposes. Some theoretical research have been published but they are usually not well disclosed and do not find any practical application/use.

The most interesting examples come from China, Korea and Japan.

3.4.1 Fujitsu

In a paper published in 2017, [40] (Artificial Intelligence applied to design), authors promote AI systematization through Fujitsu "Human Centric Artificial Intelligence Zanrai", thanks to data collected in Monozukuri.

Aim of the paper is to evaluate the effectiveness of machine learning in product design. An example is provided for machine learning applied to PCBs design. They try to estimate the number of layers of the PCBs using Support Vector Regression technique with important initiatives to continuously improve accuracy of results.

It's a very interesting example of machine learning applied to 3D structural design components based on the recognition of 3D items using the same technique applied to 2D objects. The method is as follow:

1. Collection of images for each component generated from past data

2. Extraction of features vectors from images
3. Generation of images for each new components for which similar shaped components have to be detected
4. Extraction of vector features from these last images and calculation of similarity with the features vectors stored in the database and detection of similarly shaped components.
5. Important initiatives to improve accuracy (like changing the point of view for image acquisition to avoid repetitions for symmetric components

Authors also provide a general framework of reference for applying Machine Learning to product design. The framework is called Monozukuri AI framework and is composed by:

1- Collecting training data (removing those training data that are pure noise) and classify the retained training data by applicable theme. To make the selection of training data more efficient, they have developed tools to visualize the characteristics of design data in the form of graphs.

2- After training data have been collected, one has to extract features vectors from them (feature vectors are usually numerical data that assume very different forms depending on the field and sector). Feature vectors will represent input data for the learning model.

3- Analysing the learning model created from the training data and improve accuracy.

   Even for this step, a standardized method which makes use of an integrated platform is under development. The main optimization goal is to minimize rework activities.

3.4.2 GANs

Recent article about Generative Adversary Networks in collaboration with Autoencoders for automatic design process. It is somehow one step ahead the methodology presented in this work. Also, it seems to be only applicable to quite simple products (with small number of features to take into consideration). But it is for sure a starting point and a model to be considered for the huge findings and conclusions brought. [41] The functioning of GANs have already been disclosed in dedicated section (Generative models). In this research, the target is Conditional Generative Adversarial Networks.

The formalized approach can be well synthesized as in Figure 28:

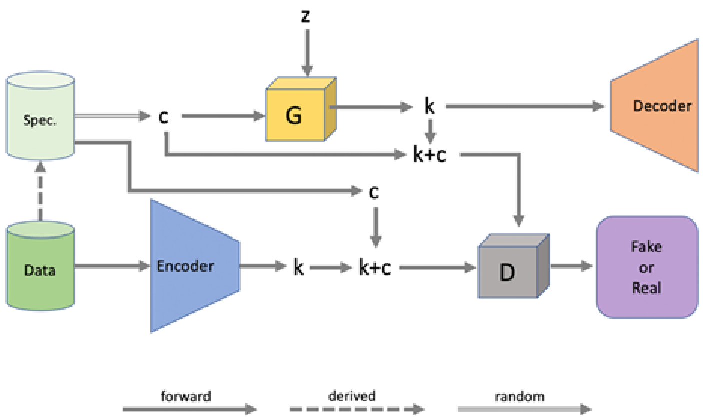


Figure 28- Conditional GAN process [41]

Specifications are derived from training Point Clouds. Point Clouds are encoded by the encoder to a latent representation k and merged with their respective specifications c to a real discriminator input k+c. The generator receives random noise input z and random specifications c to generate a fake latent object representation k, which is merged with its specifications c to a fake discriminator input k+c. Fake and real inputs are classified by the discriminator alternately. The fake latent representations k can be decoded by the decoder to receive new object Point Clouds with considered specifications.

3.4.3 Generative Design

Many articles tackle the issue of **Generative design**. Generative design is an extensive explorative design process which consists in giving design goals as input to the generative design process, along with parameters such as performance, spatial requirements, materials, manufacturing methods, cost constraints etc. Unlike topology optimization, the system explores all possible permutations of a solution by quickly generating many design alternatives. The system learns through testing and receiving feedback on the various iterations of a solution, and applies updates based on that feedback to the next iteration, until the design satisfies the objectives required [42]. This tool is useful in a circular economy perspective for minimizing the quantity of material needed to manufacture the product, to reduce the number of components connected and to use more often easy-to-remove joints.

Generative models can be used also for the design of new materials at molecular level, searching for desired properties better suited for their employment. With circular economy principles in mind, the aim could be to generate materials able to last longer, to be easily recovered and to be non-toxic for the environment [43] [44].

This design approach applied for Circular economy allows for instance to create product design that reduces the task time needed to disassemble the product using easy-to-remove **joining** systems and reduces the number of components of which the product is made, therefore improving the economic feasibility of disassembly operations. Moreover, for products presenting a simple architecture, it is easier to automate the disassembly operations using simpler systems.

Generative design algorithms are already successfully adopted by some companies. For example, Autodesk offers software solutions to apply generative design, even specifying the possibility of designing products with the aim of improving disassembly and recycling operations. The software is said to be capable of generating thousands of designs based on constraints, while a human can only generate about ten designs in the same time window. The software can create out of the box solutions hard to be conceived by humans, solving at the same time conflicting design constraints, and letting designers focus more on product innovation.

There are multiple ways thorough which a generative design approach can be realized. For example, Singh and Gu [45] presented a review of five generative design techniques (**L-systems**, **cellular automata**, **genetic algorithms**, **swarm intelligence**, and **shape grammars**). They identified potential uses of each technique given a design problem's characteristics and highlighted each technique's benefits and challenges. Based on their study, they argued that no technique could match all design problems' needs, and that those needs might change during the process itself. It was therefore suggested that a generative design system needs to be flexible and provide agency to its user to select techniques used both initially but also progressively during a generative design process. Given this, generative design has a large scope of potential implementations and applications.

Interesting research [46] discusses the implementation of generative algorithms for a particular case study in the real-estate sector. In that specific case, attention is posed on the possible utilization of two completely different generative design approaches: genetic algorithm and random sampling. The generator is described as a method that can generate solutions based on the definitions and boundaries of the solution space, with its goal being to autonomously generate a set of solutions. A **genetic algorithm** approach is driven by a set of biologically inspired processes (e.g., mutation, crossover, and selection), where each solution is evaluated by a fitness function (a function that numerically depicts the performance of a solution based on a given problem's objectives). In each iteration of a genetic algorithm approach, the biologically inspired processes are used to generate a solution set (called a 'population') that with each iteration increasingly improves the fitness of its solutions. Given the focus of genetic algorithms on minimizing or maximizing different objectives (e.g., minimizing cost), genetic algorithms' presence is typically seen in optimization problems, and has been shown to have potential for design problems with distinct and measurable objectives [47]. Despite the existence of various intelligent and advanced methods (e.g., genetic algorithms), it can also be argued that some of the methods pose a hindrance because of knowledge requirements, implementation difficulties, or time-constraints: all factors that should also be taken into consideration. For example, setting up variables and objectives is a crucial activity in genetic algorithms that can require substantial skills and knowledge. As such, simpler methods like random sampling can offer a suitable approach for conceptual design problems that are focused on exploration [48]. A random sampling approach typically relies on pseudo-random number generators to generate solution sets by selecting solutions within the solution space at random locations. Using random sampling poses few restrictions on a problem's maturity and criteria, as random sampling is not driven by well-defined measurable objectives and is arguably easier to set up and use (in contrast to genetic algorithms, for example). This also means that the responsibility of guiding the exploration's direction is up to the user, which can facilitate the inclusion of qualitative assessments of solutions (e.g., standardization of components for mentioning one CE criterion).

Many other research have been conducted and a few implementations are available.

To provide an industrial example, General Motors is trying to benefit from generative design capabilities through Autodesk software by rethinking the seat belt bracket, which secures the seat belt fastener to the seat. [49] After setting the objective and constraints the algorithm came up with 150 possibilities from which the designers could choose. The one chosen heavily changed the aspect of the component merging 8 components into just one part 40% lighter and 20% stronger compared to the previous solution. The new part was designed with the objective of cost reduction and weight reduction for a potential future use in electric vehicles; however, having one single component instead of 8 also brings positive impacts on the ease of disassembly and recycling. Results of the implementation could therefore be even greater, with Autodesk already having an embedded function to improve disassembly and recycling operations. Unfortunately, no reported cases in this field are available.

Generative design for materials has been already employed by companies as well: for example, the European Space Agency has funded a project called "Accelerated Metallurgy" with the aim of employing generative design to develop, produce and test rapidly and systematically novel alloy combinations. The new alloys were designed to be more performant and durable compared to known alloys and to be non-toxic for an improved environmental sustainability [7]. Machine learning is the best tool for this research because it can quickly examine a vast number of atoms combinations and exclude a priori most of the almost infinite combinations thanks to the learning process that is able to predict the wanted properties. Therefore, the molecule research method is organized in a new paradigm, which consists in virtually simulating a product using molecules obtained from algorithms, and finally synthesize the molecule. Generative design for molecules is also widely used in the pharma industry where it was born to reduce greatly the time to market of new products screening a wide number of molecules. [50]

3.4.4 Gaps

What is missing in all these innovative and challenging solutions is a proper "stage setting". The papers present solutions which build upon the idea that the process of designing a product can be deemed as simple, single-stage process.

This consideration unavoidably leads to the inability of implementing the proposed solutions, or at least of implementing them at large-scale in the manufacturing industry.

Covering this gap, providing a systematic and clear framework to be followed by any manufacturing company can be a strong boost through the creation of a global commitment towards circular economy.

4. Objectives

Preparing the stage for a large-scale implementation of Circular Economy principles is the main objective of this work. Starting from the way products are designed, attention will be posed on the interactions and synergies between product design and the entire system and processes composing a manufacturing company. For doing so, Machine learning solutions will be presented, and strategical issues will be tackled.

This work sets itself at a strategical level, with the aim of defining a framework of reference which can be adapted and adopted by any manufacturing company operating in different industries.

The aim is to facilitate the exploitation and effective implementation of the mentioned solutions regarding the use of Machine learning for solving specific De-manufacturing issues.

A structured approach which keeps in consideration input data definition and collection, strategic decisions, upgrades in terms of methodology implementation and related technologies will be presented.

The work is focused on the definition of a framework aimed at tackling and overcoming the existing issues and gaps in the implementation of a Circular Economy approach exploiting Product Design enabled by Machine Learning. In doing so, related issues and gaps will be tackled regarding the means used to achieve this goal. As mentioned, this means are represented by specific Artificial Intelligence technologies which are Machine Learning tools.

Many discussions and research have been conducted as regards the adoption of artificial intelligence and machine learning for automatic disassembly, cleaning, sorting, testing and re-use for high-value-added products. Upstream, these processes must be supported by a proper design of the product.

The objective is to use AI for this preliminary step: efficiently combine multiple requirements coming from different customers all along the value chain and from Circular approaches like remanufacturing and upgradability in an innovative product design.

This will be done covering the gaps related to:

- Adoption of Design for product lifecycle strategies
- Exploitation of Machine learning for product design

In order to allow the

- Exploitation of Machine learning solutions for Disassembly purposes

According to the CIMO logic [51], the aim is to develop, considering manufacturing firms (context), a structured methodology and framework (intervention) allowing the exploitation of machine learning and strategic tools (mechanism) to design CE oriented products (outcome).

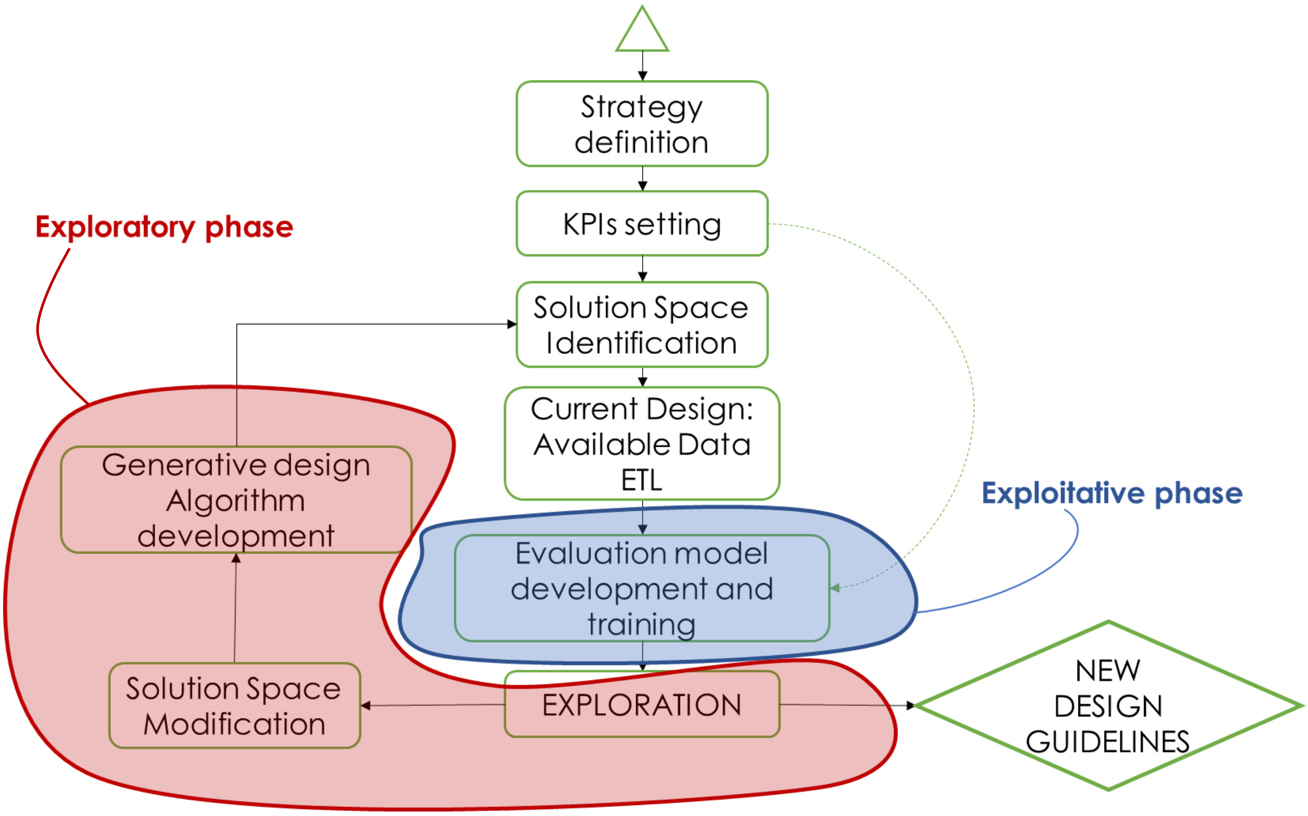
The purpose of trying to couple innovative and complex design techniques with circular economy purposes considering all the aspects involved in the path will be then applied to a product of interest which is lithium-ion batteries for electric vehicles.

5. Methodology

The developed methodology has the aim to organize, in a structured and systematic way, all the previously mentioned aspects linked to product design strategies and related circular economy needs.

The goal is to provide a clear framework which can be adopted and adapted by all the manufacturers, according to specific needs and technical constraints.

The conceptual flow, which comes as a result of in-depth studies on the mechanisms of a correct design approach, is represented in Graph 1.



Graph 1- Conceptual Flow

The starting point is a clear **strategy definition**, which considers business sustainability in all its declinations. This means that the first step in undertaking this kind of strategical path linked to product re-design is considering all the important aspects which play a role in the long-term development of the manufacturing company.

More precisely, strategy is a set of plans and policies with which a company tries to respond to market needs gaining advantages over its competitors. A company strategy must consider all the functions present in the enterprise, like Research and Development, Finance, Sales, Marketing, Strategic planning, and Production.

Although product design seems to be more related to the Production process, it has and it must have huge impacts on all the other strategic levels, especially when it is studied and engineered for circular purposes.

This is due to the intrinsic and strong relation which links product, process, and system and which forces the fact that the evolution in one of these fields (product in this specific case) necessarily means evolution in all the other fields (process and system).

5.1 Strategy Definition

The strategic aspects comprehend:

- The choice of the **main target** of the circular approach that is willing to pursue through product re-design. May it be repair, ruse for same purpose, reuse for different purposes, remanufacturing or closed loop recycling;
- The consequent adjustments needed at strategic level, strictly intended as the eventual **re-positioning** of the company in the market, its perceived **delivered value**, and the eventual need to rise entry barriers to protect the business from cannibalization or threatening new entrants. Also, decisions must be taken on how to acknowledge customers and other players of the new value that is being created and of the potential gains it brings.
- The choice of key partners; re-organizing the entire **value chain** is fundamental for building a valuable and sustainable circular approach. Value chain must be intended as both **direct and inverse value chain.**

  In the best-case scenario, indeed, Circular Economy aims at creating a circular, unified supply chain in which every actor – from the OEM to the logistics to the final customer – shares information and responsibilities. This means that, if a product is conceived in a circular perspective, the benefits of this choice will be shared and perceived by all the supply chain actors. The OEMs will have their manufacturing costs reduced thanks to the standardization of processes and the reutilization of materials and components; the customers will have costs savings coming from an extended products' life and improved quality; the middle-tiers actors will benefit from the standardization of technologies among manufacturers and from the higher level of information provided, facilitating maintenance and recovery operations. In general, the requirements that the product must satisfy are given by the necessities of a large set of stakeholders and can be expressed as a list of Redesign Requirements.

These points are well summarized by the wide concept of **business model**.

"Business Model is the conceptual and architectural implementation of a business strategy and represents the foundation for the implementation of business processes and information systems." [52]

The idea is that the methodology for implementing Circular Design in manufacturing industries should start with the re-setting of the business model. This kind of integration is fundamental in order to exploit the benefits of a product design conceived for circular purposes.

Exploiting the clearness of the *business model canvas* as a visual tool, it can be stated that the first point of the method developed in this work is the re-definition of the 9 boxes which composes such model.

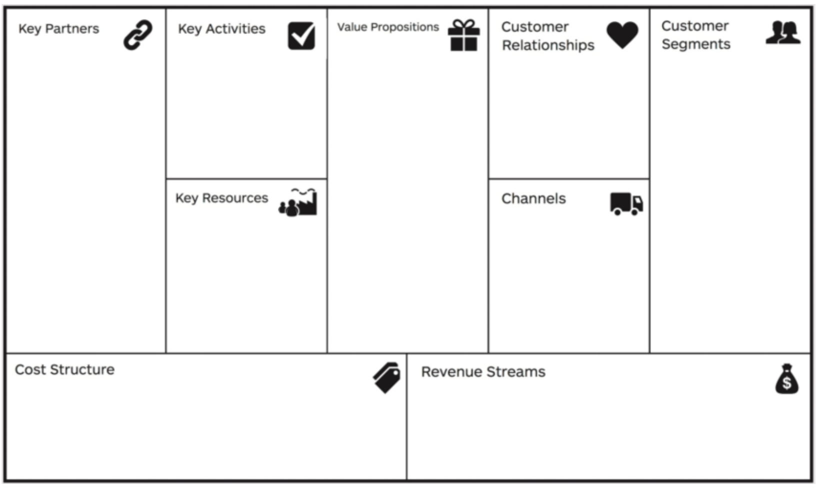


Figure 29- Business Model Canvas

Besides the already mentioned points to be defined, which are contained in the model respectively as **Value Proposition**, **Customer relationships**, **Customer Segments** and **Key Partners**, the approach should start with the clear definition of:

key activities related to the circular approach adopted;

key resources which in this case are most likely to be represented by data, data collection enablers (IIoT instruments like sensors) and data analysts;

channels through which the product is distributed and, above all, re-collected at the End-of-life.

cost structure which involves the definition of the way costs are organized, in the case in which the adoption of the new circular strategy has a strong impact on the proportion of variable and fixed costs and on the way they are distributed in time.

Revenue streams, i.e., the way the company receives the value created in terms of money. Another time, the adoption of the new circular strategy may, and should, have an impact on this. Examples of revenue streams are pay-per-use, fixed fees, only maintenance payments ecc.

Examples of new business models which may support Circular Economy approaches are present and can be grouped into five main models.

1. ***Availability guarantee***

   Especially applicable in machine tools industry. The machine tool builder keeps the responsibility of maintenance, thus has the possibility to continuously monitor the state-of-life of its manufactured product and to collect many data on it. This represents a huge advantage for manufacturers who want to undertake a circular economy approach. The machine tool builder is paid for availability, meaning that he guarantees an agreed availability rate and he's paid by the customer to maintain it.

2. ***Providing personnel assistance for customers operations***

   The manufacturer is paid for the service provided in terms of product performances guarantee and operative actions. Also in this case, operative and maintenance activities are responsibility of the manufacturers, who assist the customer with skilled personnel paid for the provided result (quality, productivity,…). Circular economy gains of this business models are always related to the possibility for the manufacturers to continuously collect data on the product conditions and state-of-health, thus enabling the implementation of solutions for keeping the product live longer like the one presented in this work.

3. ***Production services to cover peaks/smooth demand***

   It's a kind of downstream integration done by machinery tools builders. Instead of providing the customers with machinery, in those periods in which customers have to face picks in demand, manufacturers directly provide them with parts produced with the machines they usually sell. This allows a higher utilization rate of machines since manufacturers sell their parts produced to a large set of industrial customers. The revenue stream in this case is "pay per parts sold". Although this innovative business model is not strictly related to Circular Economy purposes, it is worthy to mention due to the intrinsic flexibility notion it brings. It is a clear example of the ability of a manufacturer to change its business model to adapt to strategic needs like increase productivity and make revenues form through other sources than the core business.

4. ***Build (or operate) own business models***

   The manufacturer keeps the ownership of the product and is paid for the service/value that the product delivers to the customer. Also in this case the manufacturer exploits the synergies of maintaining the product and updating it, also through its perfect knowledge of the business of the customer.

5. ***Lean machine adaptation services***

   This innovative business model deals with the bringing to the market of modular and scalable products. In particular, staying in the machinery tools sector, the manufacturer shoulders the costs of flexibility instead of making the customer pay for it. In other terms, instead of flexible and expensive machines, the manufacturer sells lean machines with only basic and essential functionalities (frugal products) with the agreement to reconfigure the machines in case of new production needs of the customer. Also in this case Circular economy arises: thanks to a new design of the product (machinery tool), which must support modularity and reconfigurability, the manufacturer is able to (i) save resources for the production of next generation machines which will be built directly adding functionalities to the existing ones instead of being produced from scratch; (ii) fully and flexibly satisfy customer needs without wastes and useless costs; (iii) continuously keep an eye on the customer, thus being able to forecast his future needs and make monetary assessments of future required reconfigurations of the machine. (iv) maintain customer loyalty, thus a sustainable business.

Having understood the importance of a clear strategy setting, especially when dealing with circular economy purposes, the proposed methodology continues with the awareness of the importance of measuring results.

After a proper strategy setting, performances of the company must be measured and evaluated in order to verify whether they are aligned with the stated strategy.

For this reason, a proper set of KPIs must be defined.

5.2 KPIs Setting

Whenever a company wants to act on its business, either for improving it or just to bring it to a desired level, the fundamental need is to know:

- The exact level of performance the company wants to reach;
- The starting point;
- Performances reached in intermediate steps (i.e., keep track of the evolvement of improvements).

For doing so, a structured and precise set of measures has to be identified.

Key performance indicators are measures of the most relevant performances of a company.

For the purposes of this work, besides KPIs directly addressing the typical strategic performances of a company (financial indicators like ROA, ROE, EBITDA, and operative ones like Throughput, cycle time, average delay, etc.) [53] , specific KPIs targeting circular needs must be set.

As explained in section 1.1.6, a holistic, integrated, and scalable set of measures for circular economy is still missing, despite being essential for a large-scale implementation of CE.

A set of fundamental KPIs is here reported, starting from the general and holistic ones, to arrive to those related to specific CE approaches.

- **Cycle time**. It is the total time from the beginning to the end of a process, which includes the useful time to process a product to increase its value or to recover value from it, and the waiting time, which is time spent performing non-value-adding operations. Cycle time can be referred to a whole De-manufacturing process or to the sub-processes. Cycle time should be aligned to the lead time requested by the customer. Through the implementation of the improvements targeted in this work, artificial intelligence can be used to reduce it by acting on product design, simplifying it as much as possible to avoid duplicating resources like machines and operators and to reduce investments and variable costs.
- **Resources saturation**. Increasing saturation of resources allows to exploit completely their value, reducing variable costs of an operation. This is required for those systems which rely on expensive machinery and require very long planning time, like recycling systems. However, in high variability scenarios,

like the inverse-value chain one, increasing saturation is not always the best choice, as the idle time of a resource can be used to satisfy demand peaks or variability in the cycle time of a process.

- **Number of operators**. De-manufacturing processes are highly dependent on manual labor due to their already disclosed complexity which prevent them from being automated. The costs reduction to make De-manufacturing operations more economically feasible can be reached, thanks to the implementation of this methodology, but only through a proper measure of the costs themselves. Number of operators required, indeed, can be a good measure of the cost of performing some tasks. Moreover, some activities can be alienating to be performed by a human, like quick sorting of products. For the two reasons above, reducing the number of operators given the same amount of workload is a key factor for the feasibility of De-manufacturing processes.
- **Safety**. De-manufacturing operations could involve treatment of hazardous materials and risky tasks for operators' health. Aim of the new product design specifications and of the improvement process proposed in this work is also to reduce the presence of these kind of materials and components. Also, solutions powered by machine learning can solve these issues avoiding manual jobs or supporting operators.
- **Energy saved**. Energy is a variable cost for both manufacturing and De-manufacturing processes. Therefore, this metric considers both savings from avoided production of new products in manufacturing and energy savings in De-manufacturing processes. This metric is also strictly related to the greenhouse gases (GHG) emissions avoided, due to the still too high dependency of energy from fossil fuel. Therefore, keeping track of energy consumption with the aim of reducing it, also means reducing GHG emissions and the consequent costs related (due to government policies restrictions).
- **Virgin material saved**. This KPI highlights whether the solution allows to save more virgin material for a reasonable cost. Material saving has an impact on the environment as well, reducing the depletion of virgin material reserves and avoiding introducing new material in the system that will end up in landfills.
- **Inventory carrying costs**. They identify all the expenses related to holding and storing unsold goods. It is generally an important metric to assess, and for De-manufacturing operations characterized by uncertainty its importance could

even be greater. Companies account for different costs to obtain the total value, however the most used are: warehousing costs, opportunity costs, obsolescence costs, transportation and handling, taxes, and depreciation.

Recycling-specific metrics:

Besides the fundamental KPIs listed in dedicated chapter (1.1.4), another important indicator must be added in order to keep track of the actual profitability of recycled materials (WMV ratio).

- **Recovery**. As already explained in Chapter 1.1.4, it is the quantity recovered of target material in the output flow of interest. Increasing the recovery of a target material means to increase the quantity obtained overall, which increases the profit.
- **Grade**. As already explained in Chapter 1.1.4, it is the quality of the target material in the output flow. Increasing the grade of a material stream increases its value, because a purer material is worth more. The unitary value of a material is proportional to the square of the grade. It is important to remember that the two quantities are in a trade-off relationship, so it should be found the optimal process which maximizes the revenues from the two.
- **Waste / material value ratio**. Measure the profitability of the waste collected by the recycler considering the recycled material selling price and the cost to obtain and recycle such material:

$$WMV = \frac{C_{waste}}{P_{mat}}$$

Where: C_{waste}= Waste Acquisition and Processing Cost
P_{mat}= Recycled Material Selling Price

The target is to reach the value 0 and in general to remain below the value 1.

Remanufacturing-specific metrics [54]:

- **Core / product value ratio**. Measure the profitability of a core with the objective of sourcing high value ones and reducing the cost of core reconditioning:

$$CPV = \frac{C_{core}}{P_{comp} - C_{disp}}$$

Where: C_{core}= Core Acquisition and Processing Cost
P_{comp}= Price of Sold Components
C_{disp}= Disposal Cost of Unsold Components

The target is to reach the value 0 and in general to remain below the value 1.

- **Core class distribution**. It indicates the average quality of the cores retrieved from the market and in different stages of remanufacturing process. The higher is the quality of the cores, the higher is the potential profit obtainable. An example of a numerical representation of the metric is shown below:

$$CCD = \frac{1 \cdot N_a + w_b \cdot N_b + w_c \cdot N_c}{N_{tot}}$$

Where: w_x= Weight of Class x ($0 \leq w_x \leq 1;\ w_a = 1$)

N_x= Number of Cores of Class x

The target is to reach the value 1.

In addition to these two KPIs found in literature, other measures could be set for indicating the extent to which a particular product is worthy to be remanufactured for different purposes other than the ones it had in its previous life-cycle.

It's important to underline that KPIs defined in this section are indicators related to the overall system (and sometimes to the whole process) performances, and not to the specific results strictly related to product design.

These measures have the aim to control and evaluate the improvements brought at **Process and System level** by the implementation of the new guidelines of product design. This is due to the above-mentioned strong integration between product, process and system, according to which it is not possible to act on product design without having to modify and improve also the processes and the system in which the product is produced and sold.

A model can be introduced to formalize the concept of integration of products, processes, and production systems. [55]

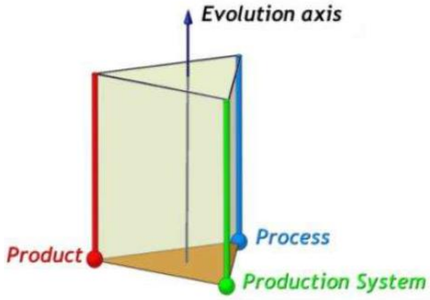


Figure 30- Co-Evolution

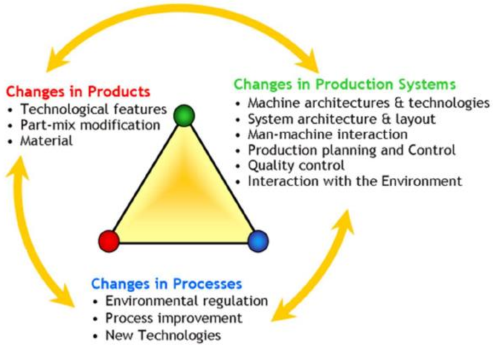


Figure 31 - Integration

A configuration approach is the entire procedure followed to configure the product, process and production system. Indeed, product, process and system must all be designed to carry out a production transformation, according to the following logic:

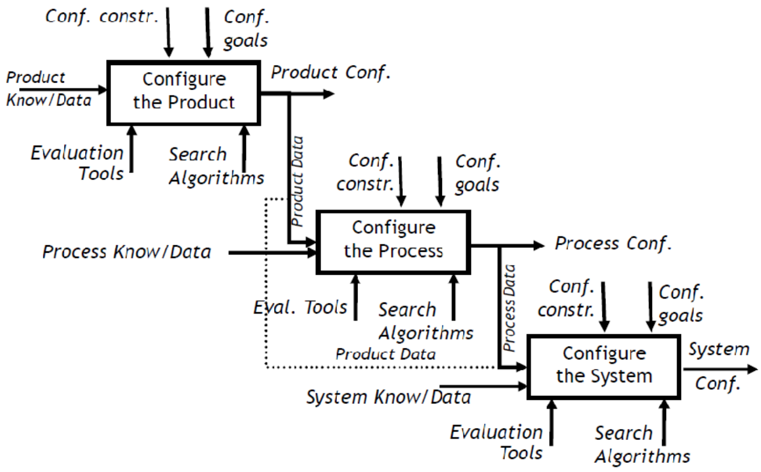


Figure 32-Sequential configuration [86]

For this reason, KPIs are set at process and system level according to the strategy defined, even if the concrete actions are taken starting from product design or re-design.

KPIs directly targeting product design are strongly dependent on the specific industry and company, since the product characteristics and functional requirements vary with the type of product considered. These KPIs must cover both performances related to customer satisfaction and circular economy requirements, according to the new strategy.

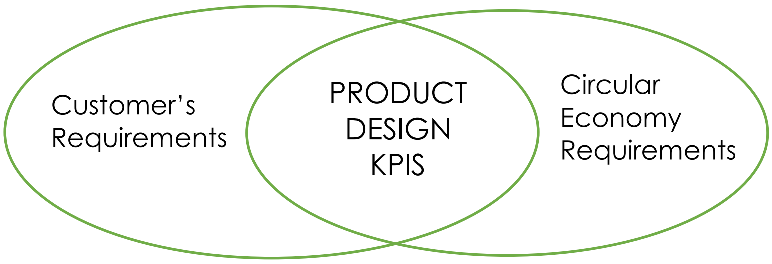


Figure 33 - Product KPIs

Once the spectrum of indicators to be considered for the evaluation and assessment of the undertaken actions has been set, the core part of the methodology starts.

In the conceptual flow in Graph 1, the next step is called *"Solution Space Identification"* and it is necessary when designing new products.

5.3 Solution Space Identification

If a manufacturing company wants to place a new product on the market, before searching for the optimal solution, it has to put an effort in identifying and defining the feasible set of solutions among which the exploitation (optimization) will take place.

As shown in Graph 1, since the proposed methodology has a continuous improvement goal, this phase of exploration of possible solutions has an iterative nature, and the solution space is continuously re-defined even for already existing products after the phase of evaluation of current solutions, in those cases in which it ends up with designers' decision to search for new solutions. In this case, Generative algorithms help humans in reaching solutions which otherwise wouldn't be reached, in a very short time compared to that required by humans and in a more efficient way. This can only happen through a careful identification of the space of solutions, i.e., the setting of the ranges of values to be assigned to the selected input variables.

At first iteration, however, a starting solution space must be provided by means of a set of features that the product must satisfy. This set of features must cover three main macro-fields:

- PRODUCT DEFINITION, i.e., features of the product responding to specific customers' needs.

- DESIGN VARIABLES, i.e., technical, engineering specifications of the product.
- CONSTRAINTS represented by Circular Economy requirements.

In case of new products introduction, after the definition of the solution space, solutions can be generated through a proper generative design algorithm, which is fed with the defined variables. In this case, in-depth study must be carried out on both the sizing of the inputs and on the proposed solutions, in order to collect data for evaluating them.

In case of already existing product design, the initial solution space is already defined, and the first iteration of the methodology is an in-depth study of the current design, which must be performed to assess the current performances. In this last case, the solution space identification coincides with the search for most relevant data describing the main features of the product and the target features of interest.

To provide another view of the framework of reference describing the methodology, summarizing the concepts explained so far, Figure 34 shows the main blocks involved in the path creation without posing any attention to the complex structure of the Machine Learning Algorithms.

As can be seen in the diagram, the space of solutions composed by the three above-mentioned fields represents the input of the Machine learning Algorithms block, which coincides with the multi-stage process of product design (later represented in Graph 2).

Inputs must be fed in the form of structured data.

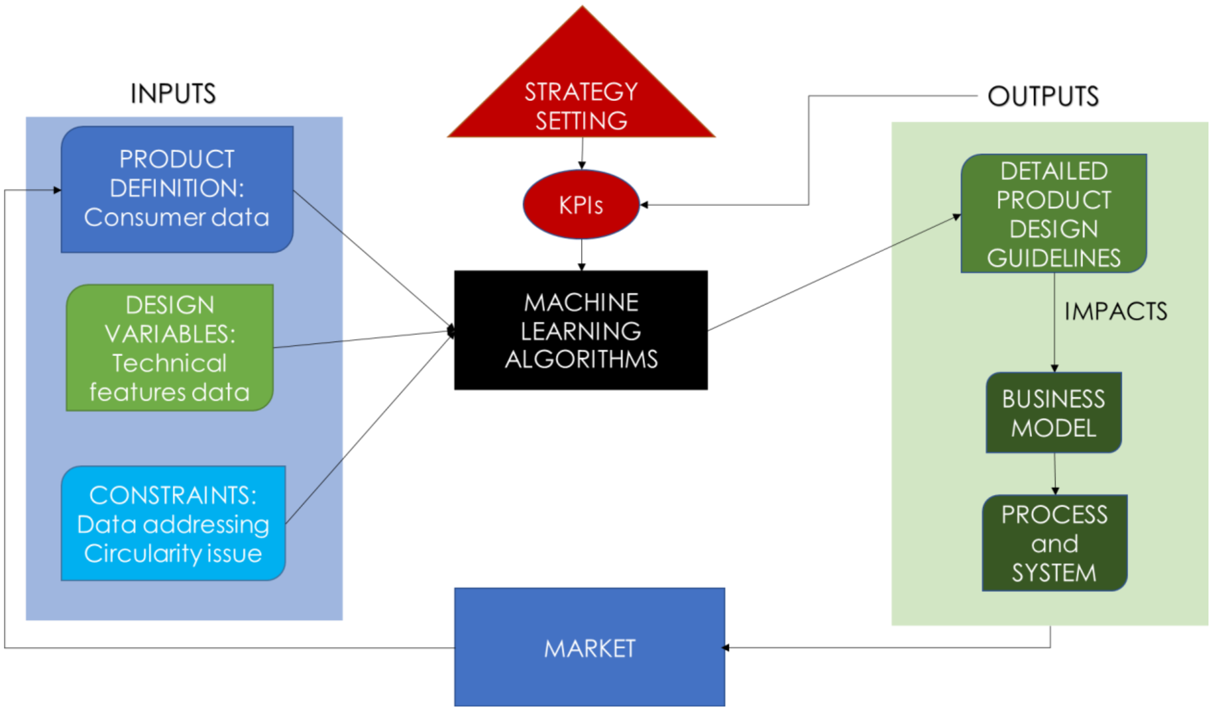


Figure 34- Framework

The fact that these structured data are available or not at the very first adoption of the method did not represent a constraint in the development of the method itself. The work has been carried out following a **PROBLEM DRIVEN APPROACH**, meaning that the solution has been developed starting from the definition of the problem, rather than from the available data.

5.4 Problem Driven Approach

This kind of approach has many reasons behind.

First, it ensures the wide-coverage and broad application field that the methodology is willing to reach. Without rigid constraints on the type of input data and on their structure, diversified companies acting on different industries, at different levels of any kind of supply chain, and with different starting points as concerns the level of digitalization and AI adoption, can be able to adapt the method and apply it. The method, indeed, provides guidelines and a path to be followed.

As a secondary aspect, the problem driven approach ensures the achievement of an optimal solution, which can be continuously upgraded as a consequence of improvements in technological development or changing requirements and constraints. Finding a solution starting from the definition of the problem itself is the best way to solve the problem, even if it requires higher efforts in the starting phases. Also, efforts are required by single companies, which have the duty to understand which kind of data they actually need, and collect them in a precise and structured way, always according to their specific target strategically defined.

Another reason which stands behind the decision to follow a problem driven approach, and which is strongly linked to the previous one, is the iterative nature of the method.

The problem driven approach allows to identify gaps in the current availability of data. Thinking about how to solve an issue only looking at the problem itself, in fact, allows to find optimal solutions, which however could not be always immediately applicable.

The strength of such approach is exactly this one: provide a clear way to understand which kind of data are missing in order to implement the right solution, and consequently which actions to undertake in order to collect these data. Actions may involve the product design itself or the entire business model definition and design. Therefore, the awareness about the lack of fundamental data do not constitute a barrier for the adoption of the method, instead it is the first step of the implementation of the method itself. Though the realization of the need of precise data, and the execution of corrective actions aimed at collecting them, a first iteration of the method can be already performed.

Many times, indeed, the major issue for companies who want to improve processes through data collection and analysis is the comprehension of which data to collect. This approach can help in solving this issue, providing a clear view of the gaps to cover in order to implement the solution.

For better explaining the concept, an example is provided:

By dint of the adoption of the method, a company gets aware of the fact that data are needed on the level of wear of a certain component due to its contact with another adjacent component in order to implement the desired solution. Following the path described, this company will think about a design of the product which allows to gather this kind of information. For instance, creating a configuration through which the two target components are easily accessible for monitoring them. Another intervention on product design could also be to put a sensor on the two components which collects data on temperature or other measures that are linked to the level of wear, or even assembling them in a way through which they can be more easily disassembled or replaced. Also, the business model can be modified in order to gain such kind of information. If the manufacturer keeps the responsibility on the product maintenance or operativity, he can continuously monitor the product thus gathering the needed data.

Therefore, following a problem driven approach allows to

1- Achieve awareness of the needed data
2- Eventually notice that some of these data are missing at the beginning
3- Take actions to gather these data

This third point represents, for all intents and purposes, a first application of the method itself, since it involves the introduction of new guidelines for the design of the product with implications on the business model and on the supply chain, and with the final goal of introducing or improving circularity in the value creation process. Nevertheless, these steps are actually a preliminary iteration of the method, aimed at collecting the needed data for applying the correct solution.

As shown in the Framework (Figure 34), indeed, the output of each iteration, which comes as new design guidelines and business model adjustments, after being properly checked in accordance with strategic requirements, is re-fed into the algorithm in the form of new data coming from the market.

Once the product is put on the market and used by consumers, indeed, it is able to provide additional data which may be the fundamental data initially needed in the cases explained before, or data which can be translated in information to continuously improve the process, allowing to expand the **space of solutions** initially defined and to pose the **exploitation phase** to next level.

To provide guidelines about the needed data for starting the core part of the approach, a list of inputs has been drafted, categorized into the three macro-fields of interest.

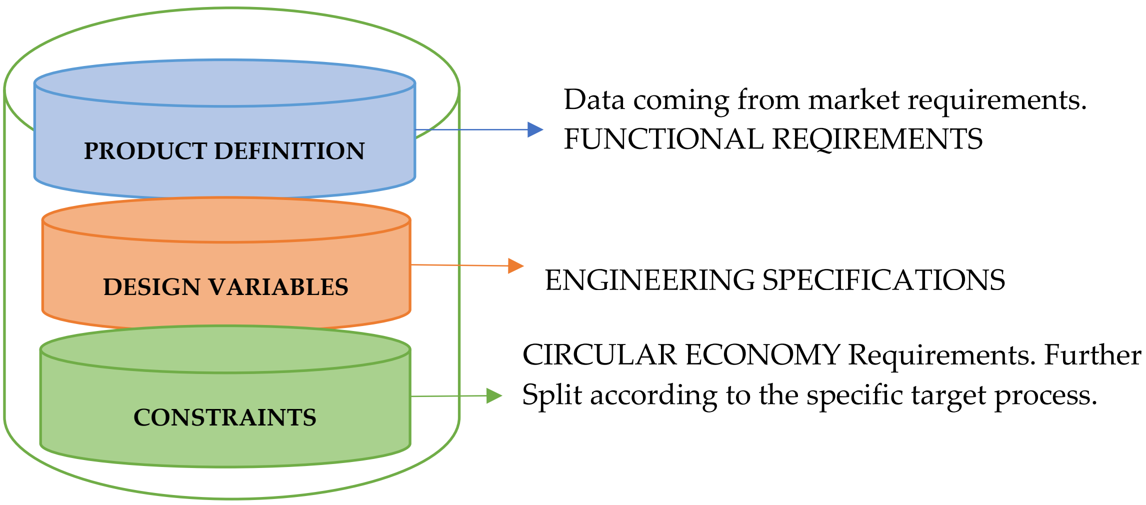


Figure 35- Input categorization

The starting point was a research done on the type of data that a product designer needs in order to start his work of engineering a new product. This is done to identify which are the most useful information for developing a product which brings improvements with respect to the existing ones, with a particular attention posed on circularity and life cycle. In this sense, the aim is to investigate and identify those data that are fundamental in order to have a deep knowledge on the product life cycle, at the moment in which it is designed and developed.

5.4.1 Product Definition

The first set contains data coming from customers and the market in general, needed for defining the product considering customer perspective. They are present in the market and constitutes those data which can be enriched through this methodology application, and which are iteratively added to the process.

In particular, they can be labelled as:

- **Data regarding the expectations of the customer.**
  Considering separately both b2b and b2c customers in case of products that have to be assembled and deployed on other complex products.
  These data, which may assume different structure and be of different type according to the specific application, are fundamental for designing a product which not only responds to customer needs, but also meets its expectations in terms of aesthetic and functionality. Such data can be represented by the usability of the product, the user interface in case of electronic products, the speed, and many other characteristics which depend on the industry and on the specific company.

- **Data regarding the way the customers use the product.**
  These data are fundamental both for a reason linked to marketing purposes, which is the reason for which these data are collected and used by now, constituting a fundamental step in order to design a product that catches and satisfies customer needs, and for a reason that is more linked to the way the methodology is created. This last "Circular Economy" reason represents the target of the entire method.
  Knowing the way the product is used, consumed and managed by the customer allows first of all to easily predict how the product will look like at the end of its useful life, thus reducing one of the main issues linked to variability of returned products. As a second aspect, these data allow to foresee whether the customer is going to keep the product until the end of its useful life or not and why, thus reducing the uncertainty in the quantity of disposed products and

their actual condition. These data will be further investigated since they represent the most part of those data that are the result of previous application of the methodology. In other words, acquiring awareness of the fact that these data are needed and of the exact shape in which these data are needed, already constitutes a hint for the next product design and business model adjustment, which should facilitate the collection of such data.
Stressing this point, product design and business model design represent strategic levers for a systematic collection of unbiased data from the customers. Starting from initial product and business configuration, the adoption of the method allows to find a way to gather this information, with benefits on the entire company.
Such a systematic approach to collect these data is significantly helpful also for the traditional marketing purpose. Until now, indeed, these kinds of data are mainly collected through surveys and other post-use and unstructured methods always based on past experience or analysis of EoL products. These methods allow to gather biased data because occur in a moment that is subsequent to the actual usage of the product.

- **Data regarding the fundamental features of the product**.
  First step in designing a product is collecting precise information concerning the key needs that have to be satisfied. Being able to distinguish between key needs that the product must satisfy and additional requirements is fundamental to avoid wastes in time and materials used. An efficient resource allocation must consider fundamental needs as high priority ones, to be accomplished and satisfied in an optimal way through an optimal product configuration.
  Also, these data are strongly dependent on the specific industry and product, and in case of IoT products, which are growing in the market, fundamental features must consider the level of intelligence of the product, its communicability with other devices/products and interoperability.

- **Data regarding those features that are not fundamental but highly recommended and expected.**
  As explained before, these data are needed in order to conduct an optimization study on how to allocate resources for the fulfilment of different requirements. Big effort should be put in the design for those functions of the product that represent the response to crucial requirements, while for the other features, always according to the strategy of the company, the designer can choose not to waste too much time and materials, always gaining good results.

On the other hand, it is necessary to collect data describing these requirements also for spotting future trends and for embracing a circular approach through product design that can be modular and upgradable, allowing to satisfy possible future requirements without the need to produce from scratch.

- **Data regarding those features and functionalities that are not fundamental and not necessary but that, if present, represent a great advantage for the product in the market.**
  These kinds of data are needed for those companies which pursue a quality leadership strategy and want to be the best on the market as concerns product functionalities. Also in this case, collecting the right data is crucial in order to avoid wastes and to avoid useless functions which do not add any value for the customer.

5.4.2 Design Variables

This set of data comprehends the engineering specifications of the product, i.e., technical data on product, materials and related characteristics which determine the product definition explained in previous chapter. Fundamental task for the product designer, indeed, is to find the relations linking engineering specifications and functional requirements, in order to precisely know whether a certain technical feature of the product is relevant or not in determining the fulfilment of a specific requirement. This analysis of the relations between technical features and functional requirements is usually performed using a visual tool called House of Quality, belonging to the management approach known as Quality Function Deployment [56]. The foundation of the House of Quality model is the belief that products should be designed to reflect customers' desires and tastes. So, marketing staff, design engineers and manufacturing operators must work closely together from the time a product is first conceived.

House of Quality has a complex structure, presented in Figure 36.

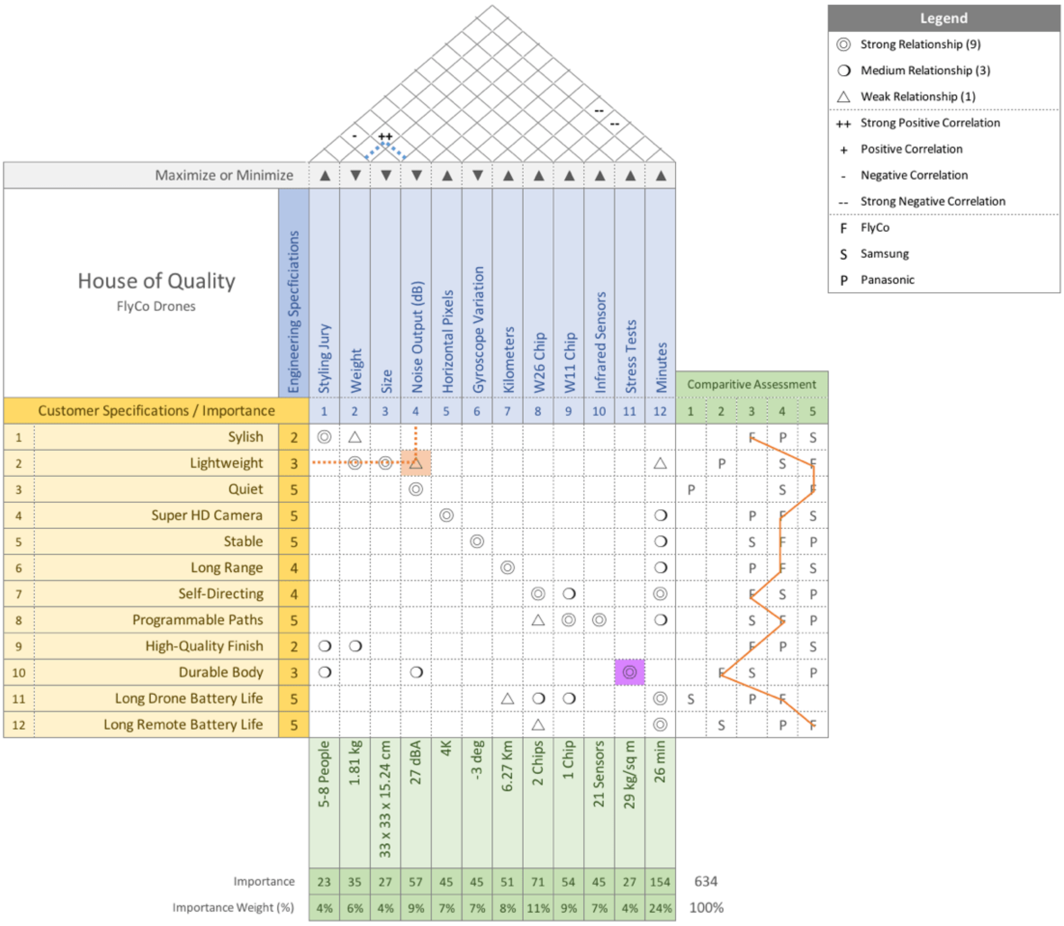


Figure 36- House of Quality [57]

As can be seen in the representation, the model is composed by different parts, each of which has a precise purpose.

- Rows of the central matrix represent customer specifications, i.e., the voice of the customers, translated in the ***Functional Requirements*** presented in section 5.4.1. According to the schema presented in the section, which includes the discrimination of features basing on their role in satisfying customer needs, each functional requirement is associated with a weight representing the level of priority/importance of that specification.
- Columns of the matrix are the ***Design Variables***, i.e., the engineering specification whose set of data will be presented in this section.
- In the middle of the house, the ***Relationship Matrix*** is the body of the whole model, as well as the visual tool in charge of explaining the above-mentioned

relations between product definition variables and design variables. Due to its strategical importance, it is filled in by a cross-functional team, and it indicates how much each engineering specification influences the customer requirements. Numbers or symbols are used to indicate the strength of these relationships, whose evaluations are based on both engineering experience and customer responses. The relationship matrix answers the question "How much do engineers influence customer-perceived value?"

- Another important aspect tackled in the model is the evaluation of the existing relations between different design variables, which will be presented in this section. This is done through the ***Roof Matrix***, a roof-shaped matrix giving the model its typical aspect of a house. This matrix is important for the engineering team since it shows the relationships between one engineering specification and all the others. If two features (A and B) are linked together for technical reasons, modifying one will inevitably require a change in the other one too. The roof matrix shows the complexity of the product architecture and answer the question "How does a change in one feature affect the others?".

- After the Relationship Matrix has been assessed, the output of the model is the evaluation of the ***Importance Weight*** of each engineering specification, usually located at the bottom of the house. Each evaluation comes from the importance of the Customer Specification multiplied by the Relationship that exists between them. In this way, the most relevant Design Variables are not the ones that affect the highest number of Functional requirements, but the ones that affect the most relevant Functional requirements.

- Additional HoQ elements are the **Comparative Assessment** box, located on the right side in the picture, and the **Objective Measures** on the bottom of the house beneath the engineering specifications (design variables) to which they pertain. The former is a benchmarking tool used to assess where the company stands with respect to competitors. To do so, customers' evaluations of competitive products are listed and compared to the target product by a trending line. The latter, instead, indicates the quantitative measures of the ES listed above.

The importance of this tool in developing this kind of methodology goes behind the marketing reasons explained before. It will be explained how the right definition of the relations between functional requirements and engineering specifications play a fundamental role also in the development of a correct circular approach.

It's important to underline how the preliminary step to the deployment of such kind of model is a correct collection of data regarding the voice of the customer, bringing back to the point disclosed in previous section (5.4.1) on the importance of creating a systematic way to collect the customer needs and feedbacks.

The engineering specifications (design variables) among which relations have to be stated and which have to influence functional requirements, can be organized in a set of data which comprehends:

- **Data about all the possible materials that can be used for each component.** (temperature of melt of each material, critical temperature, level of danger if subject to certain conditions,…)

- **Shape and dimensions of all components and of the entire product.**

- **Data on how different components mechanically and electrically interact with each other.** This is a very broad set of data comprehending the technical features which characterize specific products. Data are included regarding the assembly precedence, the way components are joint together, the presence of electronic connections and the type of connection used.

- **Costs related to different materials, components, junctions.** Even if at high level, in the design phase it is important to cross the potential materials to be used with their related cost.

Additional data not strictly linked to technical specifications can be fundamental, especially in particular industries, in the phase of design conception of products. These data can be defined as:

- **Data on product logistic and transportation** (way of transport, time of transport, places in which the product passes through). These data are important when designing a product since they strongly influence the ability of some materials of maintaining their properties.

- **Data on product distribution** (where it is sold and how).
- **Data about substitute products and competitor products.** This data may provide hints for enlarging the solution space.
- **Data about competitor SERVICES and substitute services.** Fundamental for a correct co-evolution model development, discussed in previous sections.

As mentioned, this list of data represents a guideline for companies. According to the specific aim, each company should be able to conduct an in-depth analysis for assigning correct weights to each kind of data. These weights are different form the ones present in the bottom part of the HoQ, which are related to the relevance of the specific Design Variable in determining the success of a design configuration. In this case weights indicates the importance of collecting and having the **datum** related to that variable, intended as the relevance of the engineering specification in defining the product design. Since the presented methodology has the willing to be applicable to any kind of industry, the aim is to distinguish which data are fundamental for developing this methodology and those that can be neglected, according to specific cases. In this way, through the mechanism described in in the Problem Driven Approach section, actions can be undertaken to gather those fundamental data that are missing.

The correct matching of functional requirements and technical features which comes as output of the House of Quality model, can be seen itself as a first attempt to increase the circularity of the product. By correctly identifying those features which are most impacting fundamental needs, efforts can be put in creating a product configuration which concentrates on such features, using less materials and less resources for developing features which are not relevant for satisfying important requirements. This is the reasoning behind the "frugal design" approach, [58] which can be considered an important Circular economy enabler and facilitator.

Despite the fact that the data listed so far were not specifically targeting circular purposes, they can be translated into circular constraints according to the specific product analyzed.

The third set of input data, indeed, is composed by those specific values, or ranges of values, that technical features must assume in order to satisfy specific circular needs. Besides these constraints, additional data can be added in input according to the strategy that the company is willing to pursue.

5.4.3 Circular Economy Constraints

As explained, the first step needed to undertake this kind of approach is a clear definition of the strategy that the company wants to adopt. This comprehends the setting of a target objective as concerns circular economy implementation, which must be decided a priori by the single company in compliance with features of the business and strategic objectives. Given the complexity of circular economy systems, indeed, it would be inconceivable, other than disadvantageous, to try to pursue all the circular economy strategies at a time. The methodology aimed at finding innovative product design must be tailored to one or two specific circular aims.

For this reasons, data and constraints representing circular needs have been grouped into five main macro-areas, each of which targets a specific de-manufacturing process.

Clearly, there can be overlapping in the data needed for different processes, but for the sake of completeness and clarity all the macro-areas contain the needed data to be collected in order to implement the method.

The discussed five macro-areas are:

- Increased life cycle (through reuse or repair or simply designing a product aimed at lasting longer).
- Easy and automated disassembly.
- Easy and automated sorting.
- Re-manufacturability, which comes as a conjunction of the two previous objectives, adding other constraints specifically targeting other steps of the process.
- Recyclability, which also includes disassembly and sorting and adds all the other constraints required by typical recycling processes.

1. **INCREASE PRODUCT LIFECYCLE**

   Wanting to pursue this kind of strategy, machine learning algorithms must be fed with correct constraints specifically targeting those features that have an impact on the ability of the product to last longer.
   In particular, these set of data have been identified as useful for providing correct information for the product design.

- **Data on the combined material properties of all the components of the product**. These data are fundamental for selecting right materials and assembly which can last longer enabling to increase the product lifetime. Interactions among materials and components, indeed play a fundamental role in the degradation path of components and whole product.

- **Type of materials used for joints and connections**. Also in this case, if the target is to design a product that have to last for a longer time, connections and joints must be studied in such a way to guarantee stability and hold.

- **Characteristics of joints and connections** (whether they are mostly reversible or irreversible and other features). In particular, considering only re-use and repair purposes, irreversible connections are better since on average they last longer. It will be explained in next section how this constraint completely changes when dealing with the need to disassemble the product.

- **Level of homogeneity of the product**. Intended as utilization of same material for each single component and for components having the same function.

- **Thermodynamic compatibility among metals** involved and **thermal reactions** which may cause unavoidable degradation after a certain amount of time.

- **Data on the rate of technological change of the particular product**. In case the rate in high, the reuse or repair solutions may not be the best and remanufacturing for function upgrade could be a better choice. In this case, constraints change, and the need becomes to design a modular product whose functions can be scaled and upgraded according to needs.

- The already discussed **data on customers' behaviours**. Besides the general aim explained in previous section, mainly targeting the reduction of variability and uncertainty typical of the inverse chain, these data coming from the market can be used for specific re-use and repair systems building. Information on how customers interact with the product and on how the market is evolving allow to develop products that are able to live longer both in terms of technical obsolescence and degradation of components, both in terms of customer needs that have to be satisfied. The way data are collected now can be not enough to this aim. It could be necessary to initially modify products configurations adding sensors in order to have real time data, or to re-design of the entire business model: new BM related to product servitization, in fact, which expect the producer to keep the ownership of the product and so to maintain the responsibility for maintenance and control, enable this kind of data collection. Product servitization is for sure an advantage for the producer as concerns the Circular Economy purposes.

2. EASY AND AUTOMATED DISASSEMBLY

Issues and criticalities related to this kind of process have been largely discussed in section D and in section O.

Through the exploitation of the following data, fed as input constraints to a proper machine learning algorithm, guidelines on how to design a product aimed at overcoming those issues can be found. In particular, data are:

- **Data on how the components are put together in already existing products.** (Disassembly tree). This information is needed in order to know which are the existing relations among components and, thanks to the technical and functional knowledge of the product translated in input data, make the algorithm learn of a better configuration under the disassembly perspective. Note that the fact that this new configuration will always be feasible is guaranteed by other inputs related to previous data on design variables, which define the solution space. As a consequence, **technical specification on the necessity that two components are physically connected** are needed.

- **Data on how each component interact with the other ones.** Both in terms of mechanical connections and electronic/electrical ones.

- **Number of rotations/movements needed to assemble/disassemble two components.**

- **Characteristics of the materials of connected components.**

- **Data on the number of layers needed for each component according to specific product requirements.** These data are of particular interest in case of specific products like PCBs, which are critical under the circular economy perspective.
  By knowing in advance, thanks to proper and systematic data collection, how many layers are needed for specific requirements that the product in which the PCB is deployed must have, it would be easier to design the product standardizing it as much as possible. In this way, new design would support an easy and automated disassembly.

3. **EASY AND AUTOMATED SORTING**

- **Data on the entire range of existing product as concern their composition, way of collection, usage, distribution**. This is to identify a unique and global standard for every product belonging to a specific market segment. The introduction of a standard will allow to easily sort and recognize products and components even with a different level of consumption and dirtiness. Also, it allows to reduce the number of categories in which the components have to be sorted and classified.

- **Data on the usage level of the product.** If I previously know the conditions of my product at the moment it has to be sorted, I can easily and automatically do it.

- **Data that identify a specific component/product**. RFIDs deployment on the interested component/product, for example, could help overcoming many issues related to complex products. In those cases, in which RFIDs are not present in current product configuration, their introduction could yet represent a guideline for the next product design. Algorithms will then help to reconstruct the whole product design resulting from the introduction of such technology in the product, thanks to the exploitation of data on how components must be connected, and other data listed before. As a result of this new design, simply passing through an active device, the product/component is automatically recognized and sorted.

4. RE-MANUFACTURABILITY

Data required for Re-manufacturability comprehend the sets of data required for easy disassembly and sorting, since they are stages of the complex re-manufacturing process.
All the needed data go in the direction of remanufacturability-through-design. Meaning that the only way to improve the ability of a product/component to be remanufactured is starting from its design. Remembering the steps of remanufacturing exposed in Figure 8, besides the already mentioned data and constraints needed for disassembly and sorting, the following data can be helpful:

- **Data on the rate of technological change of the product.** If this rate is high, the product designer must be able to design a modular, scalar product in order for it to be easily upgraded and so to overcome this issue.

- **Data on possible legislative restriction that may prevent the remanufacturability of the product despite its design**.

5. RECYCLABILITY

The aim in this case is to keep every material internal, through the design of a product which allows an easier implementation of the complex and multi-stage process of recycling. To this aim, a mix of previously listed data is needed, since first recycling stages also start with disassembly and sorting. Besides those data, the collection and setting of other constraints can be useful:

- **Data on each single component of the product from the moment in exit the plant to the moment it ends its life.**
  This is linked to the already mentioned concept of Design for product lifecycle (Chapter 3).
  It is important to build a business model that allows to keep track of the product avoiding that it is thrown away by the customer in an improper way. Another time, the strong link between product design and business model arises, especially in terms of value proposition delivered to the customer. The collection of these data, indeed, can be only possible through a sensibilization of customers and a proper strategy aimed at acknowledging them on advantages of proper product disposal. Instil in the customer the awareness and consciousness of the importance of proper disposal of EoL products is a fundamental step for gathering this kind of data, therefor being able to facilitate recycling.
  In this direction, the needed data are of course related to:
  - the already mentioned way in which the customer uses the product,
  - identification and tracking of each component (RFID),
  - data related to the potential danger of the materials present in the product,
  - data related to the way the product must be handled once arrived at the end of its useful life, (residual energy, dangerous materials/emissions,….)
  - data regarding the design properties of connected materials (at product and component level). This allows to predict and have an insight of what could be the liberation degree after shredding of different connected materials.
    These data are:

  - combined materials properties of the connection/component/product
  - type of material joints and connections
  - characteristics of the joints/connections
  - level of homogeneity and complexity of the product/component/connection
  - thermodynamic compatibility among metals

- Available data must cover also all the **under-developing technologies** that can be exploited in order to design a more sustainable product which do not represent a big danger in the moment it is not properly collected or disposed.

Guidelines have been presented regarding the type of data needed in input to train algorithms and implement the methodology.

It’s important to underline how the level of importance and relevance of each of these data strongly depends on the specific industry and company of reference.

Each company should perform its specific analysis, assigning a degree of importance to the input data and eventually take actions to gather them.

Following this logic, re-taking the path illustrated in Graph 1, after the definition of the feasible space of solutions, the phase of Extraction, Transformation and Loading (ETL) of currently available data have to be performed.

5.5 Current Design: Available Data ETL

This is the phase in which companies become aware of the current availability of data through a benchmarking between needed and available data. Thus, companies understand whether it is possible to proceed with the effective methodology or if preliminary actions must be performed.

After the disclosure on open issues related to AI implementation (section 2.1.6), which mostly involve the early stages of data collection, it can be supposed that not all the companies are in possession of the entire set of needed data at the first attempt to implement the methodology.

In such cases, corrective actions can be taken for gathering such data that are missing, yet crucial, for implementing the effective method. As already discussed, such kind of actions can already involve the design of the product, or even the re-thinking of the entire business model.

Even less impacting actions, which can be applicable in relatively short time could be enough in some cases in which data can be made available just bringing modifications to the way products are produced, or to the information system of the company, without directly impacting on the design of the product. For example, useful data related to the bill of material of complex products can be collected through the development of software within the ERP of a company also using already on the shelf and standardized solutions.

These preliminary actions could involve the participation of all the employees of a company, especially operators, or even the cooperation of different companies operating in a particular industry, in order to gather a wider set of data considering different configurations.

If the design of the product is implied in the preliminary stage of data gathering, the guidelines on the new product configuration to put in place it in order to collect fundamental data can be already considered a first output, therefore a preliminary iteration of the method.

When needed data become available, they must be properly transformed following all the steps described in section 2.1.1, needed for being able to feed such data inside the complex system of Machine learning algorithms involved in the methodology development.

The implementation of Machine learning algorithms strongly depends on the type of input, according to the principle of Garbage-in-garbage-out which highlights how the robustness of an output provided is undermined by the poor quality of input data. Recent studies, indeed, underline how the 50% of data analysis is composed by a careful data collection, cleaning, and transformation, together with a smart features extraction.

The choice of the models to be built and trained is a strategic decision of the company according to its specific goals and constraints. However, even in this case, this work will present guidelines and possible ways on how to proceed depending on the situation.

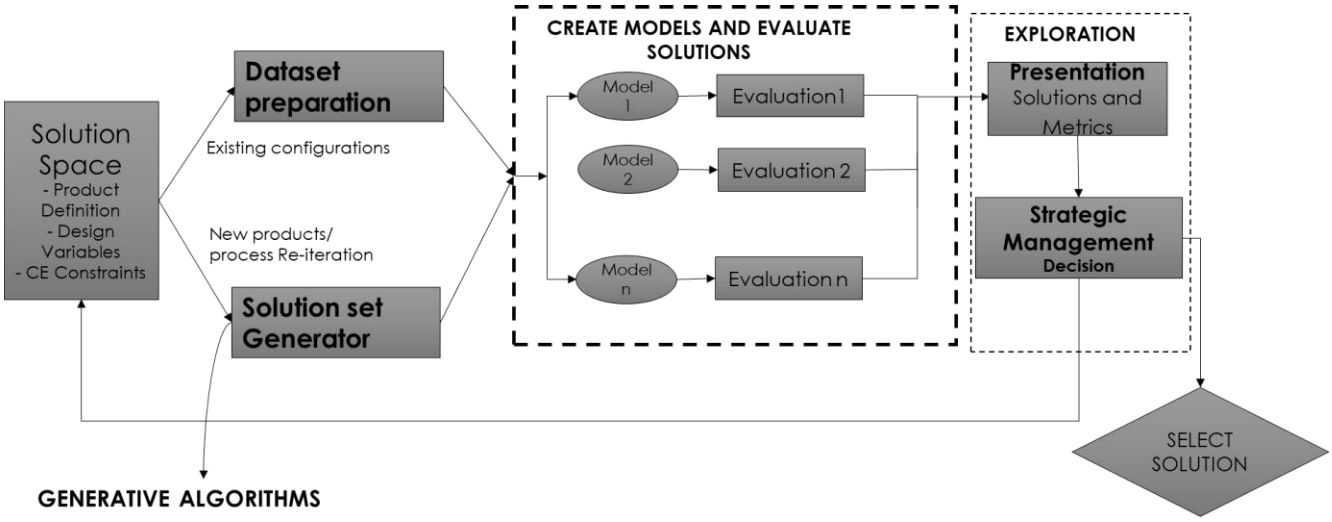
In particular, after a general framework describing the main steps of a design process (Graph 2), attention will be posed on the crucial activity which is **Evaluation (Exploitative phase)**, showing that both black-box and white-box approaches are possible.

With black-box approach are meant all the models which do not rely on known and validated physical lows, rather on learnt relationships between a set of independent variables and a target variable of interest, following the typical learning process described in regression and classification techniques.

White-box models, instead, are physical models for which the exact independent variables needed to describe a specific behavior of the examined system are well known, and so is the relation which links them. In this case, precise data to be collected and analyzed are fixed, and the task to be performed is to properly collect, clean, manipulate them and just put them into the known equation.

Considering the complexity of the most part of product manufactured nowadays, data-driven models can be considered as a better solution for performing the evaluation phase.

To enter the details of Machine learning implementation, the following graph has been drafted to illustrate the steps of product design in which ML algorithms can be involved. Precise details on the type of algorithm will be provided next.



Graph 2- Design process supported by Machine Learning Algorithms

It can be noticed how every Design Process always comes as a combination of two, essential and crucial stages which correspond to two different ways of tackling the design issue.

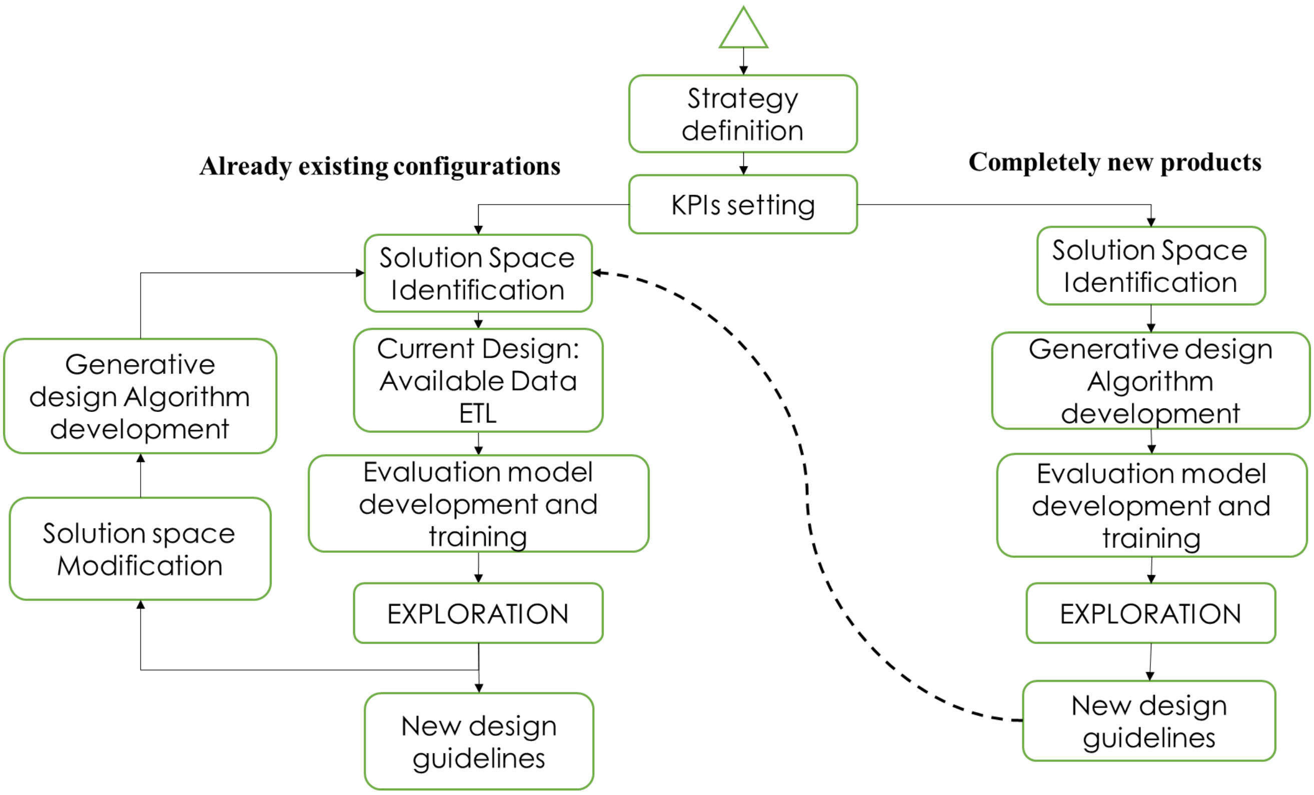
The first one is required as an initial step in case of design of a new product to be put in the market, and consists in the generation of design alternatives, given as input the feasible space of solution defined in previous section. While, in case of already existing products, this stage represents the iterative nature of the methodology, which necessarily comes after the phase of evaluation and exploration of existing solutions, i.e., once the directions of improvement have been identified. Directions of improvement are translated into modifications of the solution space, in the form of data.

In both cases, the process can be supported and accelerated by AI thanks to the use of Generative Algorithms discussed in section "**Generative models**". Specific generative algorithms developed to generate design configuration are called Generative design models, of which examples have been discussed (Section 3.4.3). Different types of approaches can be followed to generate design configurations, each of which requires different effort to be put in place in the phase of data preparation and in the training of the algorithm itself. The choice of the specific algorithm, thus, must be done in compliance with the industry of reference, the type of data available and the strategy that the company is willing to pursue, together with its availability of resources.

Examples of such models are provided in [45], which proposes the five alternative techniques represented by L-systems, Cellular Automata, Genetic Algorithms, Swarm intelligence and Shape Grammars. Another alternative must be added which is linked to random sampling, applicable in those cases in which time constraints or resources constraints do not allow a proper definition of the space of solutions and of the objectives. However, the use of such technique will require greater human efforts in the next phase of evaluation. The output of this first stage is a set of feasible solutions which are provided by the algorithm at a state of general representations, and thus have to be re-organized in order to be evaluated.

In the most common cases in which products are already on the market, and thus a current design is already available, this step can be implemented and must be implemented in order to find innovative and better solutions, but only after a proper evaluation of the current configurations. The conceptual flow presented in Graph 1 strictly refers to this last case in which the Generative Algorithm development is a consequent phase to that of Evaluation (also called exploitation). This was done due to the iterative nature of the process, for presenting the most general case. For the sake of completeness, the conceptual flow can be split into two different cases: on the right the case of design of completely new product, in which the phase of training of a Generative Design Algorithm precedes the phase of configuration Exploitation.

Anyway, this case will be then reconducted to the more general case of already existing configurations (see dotted row in Graph 3), which can be re-iterated n times.



Graph 3- Conceptual flow explicitly covering all the cases

On the left, the so far discussed case in which the target is the re-configuration of already existing products.

In any case, once data are available as concerns the possible configurations, being them in the form of output of generative design algorithms or in the form of structured data largely explained in the dedicated section, the second fundamental stage of product design must be performed.

The second stage, which comes once the configuration design has been set, data have been gathered, cleaned, and properly transformed, is an evaluation of the space of solutions and corresponds to the **Exploitative Phase**.

This phase is aimed at modelling solutions (i.e., configurations) with the objective of evaluating them under different perspectives. In other words, models are created, through the use of machine learning algorithms or other techniques, which comes as different representations of the solutions basing on different metrics.

These models build on a principle described by Sandberg et al. [59], where a general representation, or even a structured set of data describing an existing configuration, is transformed into different models to enable its evaluation. Each model contains the specific data needed for its specific purpose, and models are used to handle different representations of a solution so that it can be analyzed, evaluated, and visualized according to the requirements and objectives of the design problem and its metrics. Examples of models may include models of energy consumption, usability performances, safety of disassembly and sorting, among others. Depending on the model and on the type of measurement required, the evaluation can include calculations, simulations, analyses, or other processes deemed necessary. The evaluations are an essential part of providing input to the exploration approach in the form of metrics and potential sources of visualization of each solution, as this are the information that will guide the generator methods', together with the designers', decisions.

This is the main reason why, in the conceptual flow presented in Graph 1, this step of solution Exploitation has been put upstream to the Exploration and Generation phases, even if for sure a starting configuration is required.

This phase of modelling and evaluation can be done through the implementation of algorithms which learn the intercurrent relationships between the set of attributes describing the solution (i.e., all the data gathered in previous stages which determine the solution space), and a set of KPIs identified in order to evaluate each possible solution.

These KPIs represent the specific metric on the basis of which the model is built and are exactly the KPIs strictly related to the design of the product, defined in 5.2 (see Figure 33).

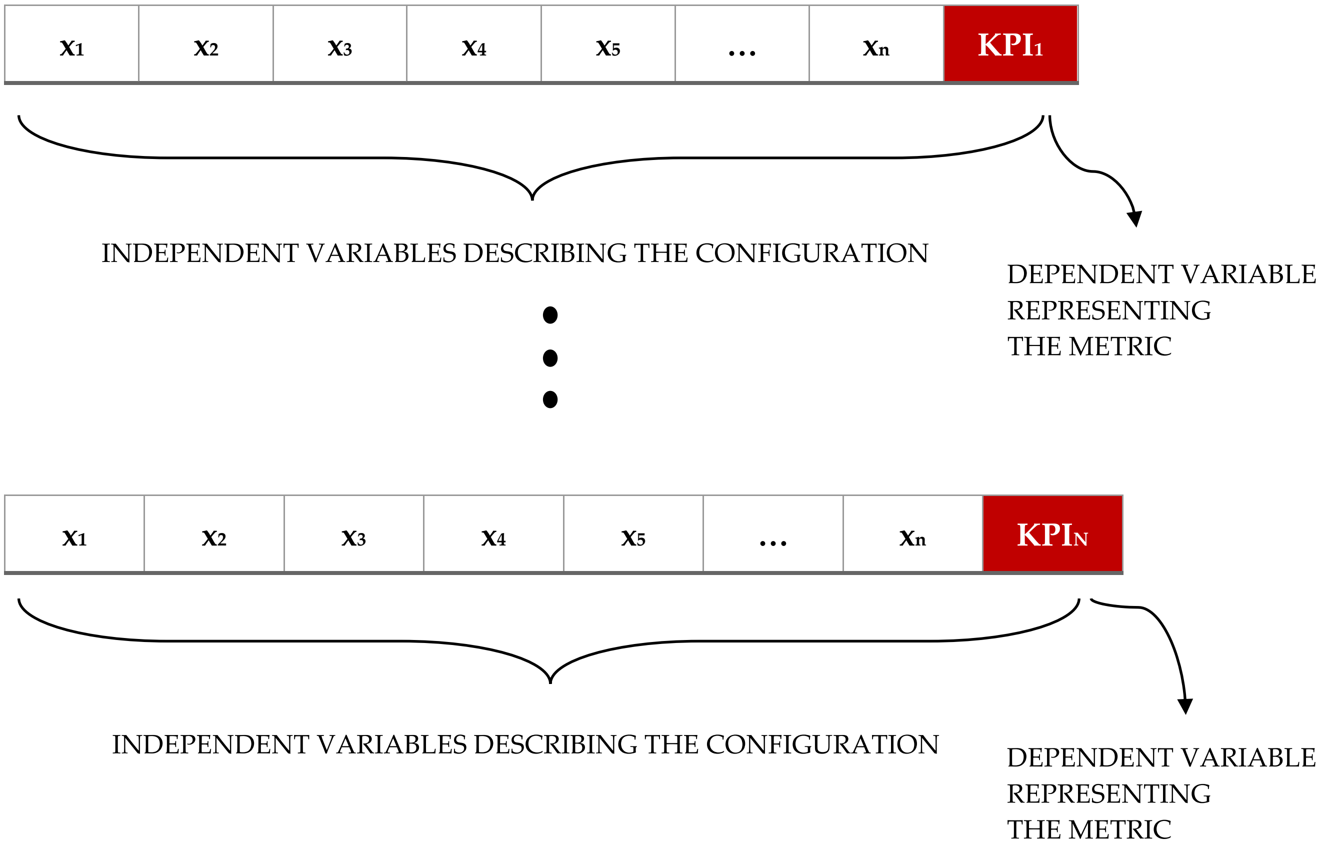
KPIs may have different natures and shapes depending on the specific field of application of the methodology but will be always related on both level of accomplishment of functional requirements and target circular needs.

Even the number of KPIs, and thus of models to be created, is a parameter which must be decided at single company level, and which can vary even depending on the iteration performed.

5.6 Evaluation Model Development and Training

For creating models aimed at evaluating solutions, the algorithms presented in this methodology are regression algorithms, which take as input the proper set of independent variables describing the solution to be evaluated, plus a target variable which indicates the value assumed by the metric (KPI) in relation to that specific observation (set of independent features).

Number of regression algorithms to be trained corresponds to the number of models to be created, thus to the number of KPIs through which solutions are wanted to be evaluated.



Regression algorithms can learn the relations between subsets of those data provided as input and explained in the "problem driven approach" section and KPIs set at strategic level.

The issue related to regression implementation is that regression algorithms necessitate of a huge amount of data in input, well-structured and cleaned. For this reason, regression is only applicable in cases of products which are already on the market, and which are available in different configurations. This is due to the need to train each algorithm using a large set of different configurations to which different values of the specific KPI are associated. In this last case, the best option to gather data could be the mentioned collaboration of companies which produce the same product following different configurations, and which share a common strategy linked to gaining circularity.

In many cases, indeed, for reaching ambitious results linked to new strategies and innovative technologies, the best option is to go for a co-opetition instead of a competition. Dealing with new technologies and wanting to improve them bringing them to their maximum level requires the commitment of all the actors playing in an industry.

Assuming to have this kind of dataset, the training of the algorithm can be done using data analysis tools and languages like Python, supported by Anaconda.

Once imported the needed libraries and data, as shown in Figure 37,

Data import

```
In [2]: import pandas as pd
        import numpy as np


        %matplotlib inline
        import seaborn as sns
        import matplotlib.pyplot as plt

In [3]: df = pd.read_csv('Configurations_KPI1'   .csv')
```

Figure 37- Data import on Python

Independent variables must be split between categorical and numerical ones, in order to properly act on them. Univariate, bivariate, and multivariate analysis must be performed before feeding the data into the algorithm, and these analyses are different basing on the nature of the variable (numerical or categorical). Even all the cleaning and preparation steps explained in dedicated sections must be performed separately for the two types of attributes. For example, numerical variables require a careful bivariate and multi-variate analysis to spot correlations between attributes which may undermine the feasibility of a model containing all of them; they also need to be properly standardized and cleaned from outliers. Categorical variables, instead, require a qualitative check on the extent of their relevance in determining the target variables and need to be properly transformed into dummies in order to be correctly evaluated and compared.

Once data have been prepared, they can be re-merged through a proper function:

```
X=pd.concat([dummies,X_numerical], axis = 1)
X.tail()
```

Figure 38- Concatenate cleaned Dummies and Numerical

And then, train and test sets must be split:

Separate Train/Test sets

```
y=df['cnt']
```

```
from sklearn.model_selection import train_test_split

#SPLIT DATA INTO TRAIN AND TEST SET
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size =0.30, #by default is 75%-25%
                                                    random_state= 123) #fix random seed for replicability

print(X_train.shape, X_test.shape)
```

Figure 39- Dataset separation into training and test sets

In this case, separation method used is the random selection called Holdout method. Nevertheless, other methods can be used, as explained in section 2.1.5.

After the split phase, training set is ready to be fed into the algorithm.

Since in most cases it is impossible to know a priori which kind of algorithm would be the best to predict target variable, usually a search grid is performed, which allows to train different algorithms and compare them basing on the measures seen for evaluating regression algorithms. The most used metric for such kind of evaluation is the MSE (mean squared error), but many other measures can be computed.

In Python, this kind of analysis is performed as illustrated in next pictures:

Models

```
: def gs_regression(model, par) :
    gs = GridSearchCV(regressor, parameters,cv=3,scoring ='neg_mean_absolute_error') #with no params it reduces to a CV
    gs = gs.fit(X_train,y_train)

    #summarize the results of your GRIDSEARCH
    print('***GRIDSEARCH RESULTS***')
    print("Best score: %f using %s" % (gs.best_score_, gs.best_params_))
    means = gs.cv_results_['mean_test_score']
    stds = gs.cv_results_['std_test_score']
    params = gs.cv_results_['params']
    #for mean, stdev, param in zip(means, stds, params):
    #    print("%f (%f) with: %r" % (mean, stdev, param))

    from sklearn import metrics
    print()
    print("MAE  train %.3f    test %06.3f" % (metrics.mean_absolute_error(y_train, gs.predict(X_train)), metrics.mean_absolute_er
    print("MSE  train %.3f test %.3f" % (metrics.mean_squared_error(y_train, gs.predict(X_train)), metrics.mean_squared_error(y_t
    print("RMSE train %.3f   test %.3f" % (np.sqrt(metrics.mean_squared_error(y_train, gs.predict(X_train))), np.sqrt(metrics.mea
    print("r2   train %.3f     test %.3f" % (metrics.r2_score(y_train, gs.predict(X_train)), metrics.r2_score(y_test, gs.predict(
```

Figure 40- Model search setting

Figure 40 illustrates the setting needed for comparing different algorithms basing on:

- MAE, which is the mean average error

- MSE, which is the mean squared error
- RMSE, which stands for root mean squared error
- R-squared.

After this step, each regression algorithm which is wanted to be trained and evaluated must be prepared, as shown in Figure 41, in order to perform a double search. Possible hyperparameters are set in input to the model, which will be trained different times using a particular value of the hyperparameter at a time.

```
In [ ]: from sklearn.linear_model import Ridge

        regressor = Ridge()
        parameters = {"alpha": [0.001,0.01,0.1,1,10], "normalize": [True, False]}

        gs_regression(regressor, parameters)
```

```
In [ ]: from sklearn.linear_model import Lasso

        regressor = Lasso()
        parameters = {"alpha": [0.001,0.01,0.1,1,10], "normalize": [True, False]}

        gs_regression(regressor, parameters)
```

```
In [ ]: from sklearn.neighbors import KNeighborsRegressor

        regressor = KNeighborsRegressor()

        parameters = {'n_neighbors': np.arange(20,50,10),
                      'p': [1,2]
                     }

        gs_regression(regressor, parameters)
```

```
In [ ]: from sklearn.tree import DecisionTreeRegressor

        regressor = DecisionTreeRegressor()
        parameters = {"max_depth": np.arange(1,8),
                      "min_samples_leaf": [5]}

        gs_regression(regressor, parameters)
```

Figure 41- Models setting

On the one hand, the output of each example shown in Figure 41 shows the performances of each algorithm as concerns the measures described in Figure 40, and on the other hand, these measures are already optimized since they are the result of hyperparameter tuning done automatically by the software. In other words, Python automatically searches for the best hyperparameters related to each specific algorithm and provides as output the errors gained with those hyperparameters, allowing a comparison of different algorithms.

In this way, optimization is performed both in terms of algorithm training for avoiding overfitting while maximizing prediction ability, and in terms of algorithm type which is most able to predict the target variable given the inputs.

After this training phase, and thanks to this comparison, the best algorithm can be selected, and next configurations can be evaluated basing on the model created.

If data are not enough to train such kind of algorithms, less precise approaches can be applied.

One option could be to use classification, which however is a supervised learning method and as such requires the target variable to be defined. The difference will be that in case of classification, the target variable will be reduced to a binary categorical one, loosing information but allowing the number of records and of attributes to be lower. Or, in case of very few data available, clustering can be used for grouping different configurations basing on the degree of their similarity and allowing further analyses to be performed at single cluster level.

The particular case of new product design requires different approaches to be followed. In this case, no available configurations are present in the market, and the only input is given by the upstream stage output, i.e., solutions provided by the Generative Design Algorithm. In this case, instead of using Black box machine learning algorithms, the best solution could be that of creating ad-hoc white box models, exploiting the knowledge of the physics behind the object design. White-box models, indeed, are based on known physical laws which are able to model the drafted product configuration (general representation) and link it to specific measures of interest. This process requires big efforts in the study of existing relations between variables, but it allows to extend the method to products for which old configurations are not available, or to companies which have a deep knowledge on the physics behind a product and prefer to apply them for evaluating solutions.

Providing an example of new product, a new component for heavy vehicles is under developing phase. Once defined the set of feasible solutions and once generated innovative configuration solutions through the use of generative design algorithms, one possible target measure for evaluating these designs could be the time that this component takes to degrade until a certain threshold value. For a correct modelling of the general representations of the product, a model can be created which links the values of temperature and air friction to which the product will be subject during its life, to the related level of wear reached in time intervals. In this way, predictions can be made on the potential values of degradation of the product depending on specific values of temperature and friction, given specific design and configuration. This example can be expanded to many other fields and industries.

Even in case of already existing products, each company can decide whether to develop white or black box models, or even to go for a double evaluation, trying to develop both approaches for improving the robustness of results.

At the end of this stage, the output is a set of models describing the behavior of different configurations basing on N relevant metrics.

This is fundamental for any kind of design process since it provides a clear and structured approach for evaluating feasible or already existing configurations.

5.7 Exploration

The next step required is a synthesis of all the metrics, i.e., the setting of a unique way for evaluating the modelled solutions, keeping into account all the KPIs.

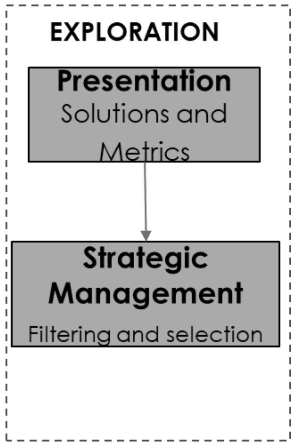


Figure 42- Exploration of results

This can be done in different ways, for example through the creation of a multi-objective function.

As stated, indeed, optimization phase always comes together with evaluation phase, following a closed loop control always aimed at controlling and improving solutions.

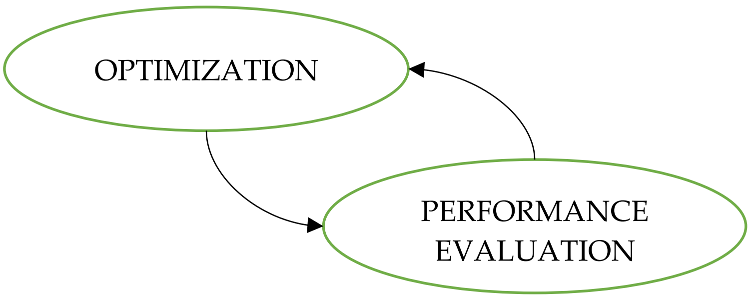


Figure 43- Exploitative phase

Following the model creation and solution evaluation in the generative design framework, is the exploration approach itself. Because the existing/generated solution set can be considerable in size, and in order to support designers and analysts in the identification of potentially best/novel solutions, the existing/generated solution set needs to be explored, basing on the above-defined metrics. The objective of the exploration is to probe a set of solutions to identify interesting regions of solutions. The outcome should either be a solution, or a set of solutions deemed satisfactory to end the generative design process, or a set of modifiers sent back to the solution space and generator to generate a new set of solutions.

The first mentioned scenario, in which the outcome of the exploration phase is a solution or set of solutions already considered satisfactory, is most likely to happen when the process shown in Graph 2 is performed from the beginning, starting from the generation of solutions through generative design algorithms. This is the case of design of new products, in which the solution space is already optimized and possible improvements may come after few years, or even the case of second iteration of the process for already existing products.

Rather, as explained, in case of first iteration of the method for already existing products, the process shown in Graph 2 starts with the phase of model creation and solutions evaluation. In other words, the flow is intercepted after the phase of generative design since configurations are already available on the market. In this, more common, case, it is most likely that the outcome of the exploration phase is a set of hints and modifiers to be sent back to the solution space.

Note that in case of large number of existing configurations, it could also be that the exploration phase identifies a solution that is better than the others as concerns all the evaluated metrics. However, since the method is applied for bringing important improvements in the direction of circular economy approaches, it is most likely that no available solutions are able to fully satisfy the set constraints. For this reason, the exploration phase will provide set of modifiers to be fed into the generative design algorithms, and the process should be re-iterated.

Anyway, this decision is taken by the company according to the designers' evaluations done on models' results. After designers have been given the presentation of each solution in the solution set, the next stage is to manage their preferences. This stage builds on the notion that generative design approaches should not be thought of as autonomous solutions, but rather as 'collaborative partners'. Besides the responsibility of each company to define the set of metrics that are considered more valuable, indeed, even the final decision on whether to be satisfied with existing solutions or to iterate the process generating new configuration is a key role restricted to people.

5.8 Generative Design Algorithm Development

As mentioned, this phase represents the iterative nature of the method.

The choice to set it at the end is derived from the need to provide the most general case, in which new solutions are generated after a proper evaluation and exploration of already existing ones.

In this case, after having derived the right kernel function which linearizes the relationships between all the KPIs or metrics and thus having explored the existing solutions basing on this synthesis measure, modification of the space of solution can be made and fed into the generative algorithm.

Once more, the extend the solution space is created in order to reach innovative configurations which allow to continuously increase the upper bound of the multi-objective function optimized through the process.

Generative design approaches have already been tackled and have to be chosen by the company depending on available instruments and resources.

As seen in the examples from literature (section 3.4.3), their development is quite simple, and strongly depends on the effort that is willing to put in the upstream and downstream phases.

After new solutions have been generated, the already discussed phases of evaluation and exploration must be performed on the new set of configurations, and conclusions must be drawn.

5.9 New Design Guidelines

The final output of the process is always the identification of a solution, or a set of solutions, which are deemed satisfactory with respect to the specific KPIs defined.

Thanks to the use of generative design algorithms, these solutions can be very different from the previous ones and important gains in terms of optimization of performances can be reached.

The advantage of implementing AI technologies, indeed, mainly lies in the innovation rate which is gained in comparison to the one that would have been reached only exploiting humans' knowledge. Also, important profit in terms of time saving can be reached through the use of machine learning and AI.

Even in the intermediate stages of evaluation, the assessment of performance is enhanced by the use of AI, which help to consider many aspects with respect to the use of other statistical or physical methods.

Further steps could be developed, which involve the creation of 3D objects downstream to the phase of design guidelines generation. This step can be undertaken only after generated solutions have been evaluated and explored, so that only the chosen solution or the chosen solutions are produced. This additional step can be performed following the approach described in [41], and already discussed in section 3.4.2.

6. Case study

In this chapter, the developed methodology will be applied to Lithium-ion batteries for Electric Vehicles.

The choice is driven by many factors regarding both design and circular economy concerns.

Considering the aspects purely linked to design, the intrinsic complexity of the product provides a challenging way to test the method. The complexity of product design is consequently reflected on the systems and processes behind, and it is a good example for assessing the co-evolution approach described in the methodology definition.

As for the circular economy aspect, the crucial position that this type of product will have in future scenarios must be tackled. Lithium-ion batteries (LIBs) have a double bond with circular economy issues since on the one hand they represent a "Green" solution allowing to cut the dependency of cars transport system from carbon fossil-fuels, with related benefits on both environment and economy.

On the other hand, also due to their design complexity, they are a critical product as regards their end-of-life treatment.

Recent studies underline how the reduced emissions during the electric vehicle's lifetime are considered to outweigh the environmental effects of the production and end-of-life phases, always assuming to use renewable energy sources. However, the end-of-life issue must be tackled and solved through the implementation of a structured approach, which should engage all the companies involved.

Also, more than any other electronic product, the electrochemical batteries represent an extremely profitable sector for the CE. On the one hand, they have a high recoverable value given by the skilled manufacturing processes required for their production and the presence of valuable metals such as Lithium, Cobalt, Nickel and others. On the other hand, their disposal will become more and more burdensome in the upcoming years, given the exponential increase of electric vehicle sales worldwide. The recovery of valuable components, the effective management of the return cycles and the reduction of the environmental impact of battery disposal may be significantly improved thanks to the deployment of CE principles and design.

6.1 Market Trends and European Legislations

By 2035, an increase of more than 50% of the market of electric vehicles is to be expected worldwide. (See Figure 44).

With Electric Vehicles sales progressively increasing, the question related to the production of their batteries and to their end-of-life management is becoming of greater importance nowadays.

For this reason, battery design and production has become a key priority for the EU commission, for reasons linked to the European automotive market in general and on Circular economy issues.

As regards the European automotive market, the concern is to be able to develop an EV industry which can become, in relatively short time, as robust as the traditional injection combustion engines (ICE) one, in order to answer to market needs and to be able to boost this solution. This can be only achieved with the development of a stable battery industry.

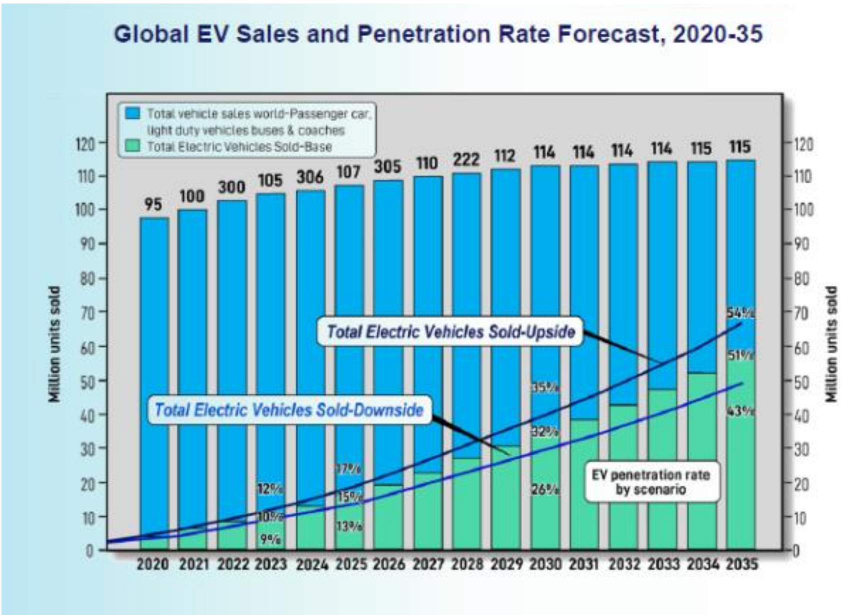


Figure 44- EV Market forecasts 2035 [60]

For building such stable industry, environmental and economical aspects must be tackled. In other words, Circular economy principles must be applied.

One of the EU Circular Economy & Green Deal principles, indeed, is the focus on keeping materials in the EU as secondary raw material and introduce back into the economy, rather than exporting these materials.

The implementation of Circular Economy strategies such as Reuse, Remanufacturing and Recycling for EV LIB packs appears to be the only viable solution for a sustainable transition towards electric mobility. To enhance such transition, important decisions must be taken at a design stage in order to develop EV LIB packs which are conceived to be easily reused, disassembled, and tested.

Providing an overview of the current EV market at global level, China is still the world largest electric car market: with 1.1 million electric cars sold in 2018 and 2.3 million units circulating, it accounts for almost half of the global electric car market share. Europe follows with 1.2 million units and then the United States with 1.1 million on the road at the end of 2018. The leader in terms of EV market share is Norway, with 46% of new electric cars in 2018, followed by Iceland (17%) and Sweden (8%). [61]

An important aspect to be mentioned is the one linked to legislations and government policies which are fundamental to consider when acting on current product design. Policies have a major influence on the growth and expansion of electric mobility. Leading countries like the ones participating in the Electric Vehicle Initiative are making progress in their policies implementation, starting with the setting of vehicle and charger standards, and then promoting economic incentives to bridge the cost gap between the EVs and the ICE vehicles. The main boost for covering this gap may come with taxation and restrictions on the use of ICE vehicles. Other actions include public procurement programs and early charging roll out.

As concerns Europe, the European commission has developed a legal framework concerning electric vehicle batteries, both for batteries as such and for batteries in cars. The End-of-Life Vehicles Directive (2000/53/EC) exists since 2000, and batteries are included. [60]

In last years, revisions have been made on this directive and the intention is to come up with a Batteries Regulation whereby the **Extended Producer Responsibility** becomes central and with new potential collection and recycling targets.

The undertaken path goes in the direction of new directives and standardized processes for the collection and disposal of EV, which ensure an efficient and effective EPR (Extended Producer Responsibility). This comprehends the creation of the ATF+ (Authorized Treatment Facility), the development of a Norm for the correct application of takeback obligation (done by FEBELAUTO) and the related release of publication of a new Belgian Standard (R 03-001) since May 2021 which marks the link between the previously used ISO standard and the FEBELAUTO norm.

The development and carry out of an efficient and effective EPR would bring advantages as concerns business opportunities and costs reduction.

However, the European efforts go in the direction of the creation of a structured and robust system for managing EoL vehicles, which has impacts on the responsibilities of producers and users and which is aimed at managing and reducing risks, while guaranteeing the highest level of recovery of batteries. This do not imply any modification in LIB packs design.

The adoption of the proposed methodology would bring additional value to the efforts put in the creation of a solid collection system.

Another important driver is the technological enhancement, promoted by an always-increasing customer demand. Research and development activities have led to big progresses in battery performances in terms of energy storage, safety and reliability. This, in turn, has caused a cost drop that is likely to continue in the following decade. As a result of the falling in prices, experts predict price parity between EVs and ICE vehicles by the mid-2020s in most segments.

All the mentioned features describing the current market show how this historical period can be a suitable moment for a large-scale implementation of the proposed methodology.

With all the OEM and LIB pack assembler collaborating for solving common issues, the battery industry can become solid and overcome problems related to both environmental and economic issues, allowing an advantage for both manufacturers and end-users.

6.2 Lithium-Ion Batteries

A battery is an electro-chemical device converting stored chemical energy into electrical energy without gaseous emissions and with high efficiency. [62]

Batteries for Electric Vehicle applications are complex and present articulated structures. However, the modularity of their architecture allows a classification of the main components that can be valid for all typologies of existing EVs. In general, an EV battery pack can be hierarchically broken down into three main levels: [63]

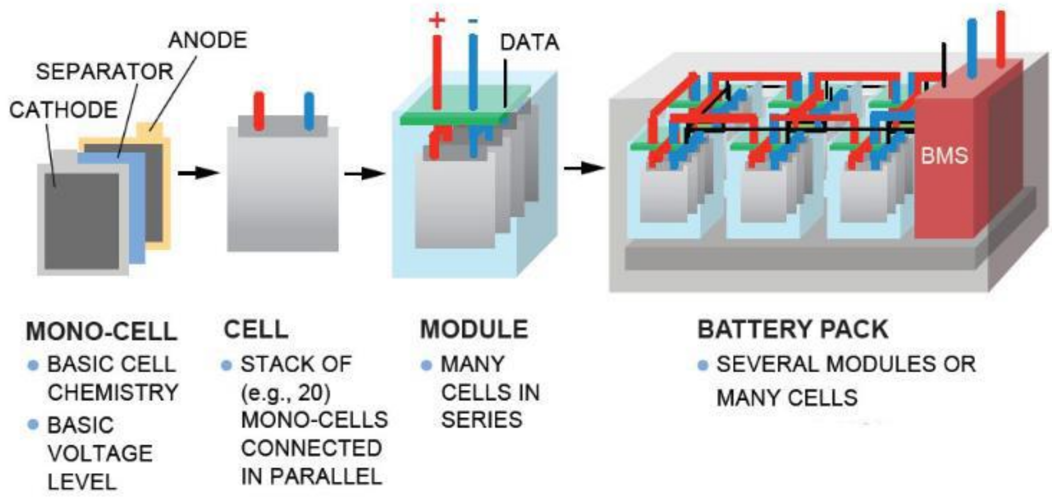


Figure 45- Structure of a battery pack- hierarchical view

(i) cell level: a single cell is primarily composed of active materials of electrodes (anode and cathode), Electrolyte (highly dielectric solvent allowing the transfer of Li+), Polymeric separator preserving electrodes from direct contact, Cu and al current collector foils;

(ii) module level: a set of multiple cells connected in series and parallel, held together by mechanical and/or physical joints;

(iii) pack level: a collection of two or more modules, which are connected in series, with sensors and controllers, encased in a housing structure [64]. The number of cells that can be connected to form a module is normally limited by the monitoring capability of the battery management system. Indeed, every battery module has to be strictly monitored through electrical and thermal control components, which are tightly packed [65].

An EV battery pack communicates with different sub-systems on multiple parameters simultaneously through various interfaces, which are presented in [66] and reported in Table 6.

INTERFACE	DEFINITION	COMPONENTS
Mechanical	Mechanical design features included for safety and functional reasons	Cell spacers, damping pads, gaskets, valves
Structural	Members that provide needed protection and isolation	Case, cover, end plates, tie rods, cross-members
Thermal	Regulates battery cell temperature	Cooling system, fans, pumps, heat exchangers

Electrical	Transmit power within, from and to the battery pack	Bus bars, cables, contactors, fuse, relays
Control	Monitor and regulate the state of battery pack	BMS (Battery management system), sensors
Support	Vehicle body parts providing additional crash worthiness	Axles, chassis, seats, vehicle floor.

Table 6- LIB pack interface system

Communication through each of these interfaces can influence the reliability and safety of the battery pack. Every battery pack should be able to deal with several issues such as thermal stability, vibration control, isolation and impact resistance at micro and macro level.

An overview of the basic structure of a battery pack is provided in Figure 46, [67]. As mentioned, main components are:

- Modules, upon which sensors are deployed.
- The cooling system which plays a fundamental role in maintaining normal operative conditions of the battery pack.
- The Battery Management System (BMS) which can be considered as the brain of the pack. It is responsible of the control of the main parameters such as temperature, current and voltage. The BMS not only actively controls the functions of the battery to maximize its life, efficiency, and safety, but also provides accurate estimations of the status of the battery such as the state of charge (SOC), state of health (SOH) and remaining useful life (RUL). The main goal of the BMS, however, is to guarantee the passengers' safety and avoid any hazards like fire, thermal shock, short-circuits or over-charge/discharge.
- The Junction block responsible for the electrical communication between the battery pack and the external electrical system.
- The Service Plug, which works as service disconnect switch. While the junction block is used for electrical switching of the heavy current circuits, Service Plug can be pulled out by hand in case of emergency. This manually operated disconnection switch allows to physically isolate the heavy current circuits from the vehicle.

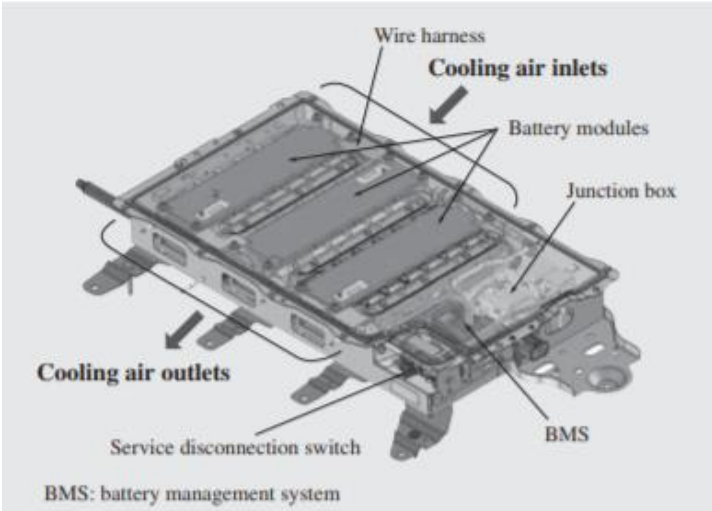


Figure 46- LIB Pack basic components

The complexity of this product would require an in-depth analysis and many aspects should be mentioned. However, for the implementation of this work, basics concepts on the structure of LIB Packs have been provided, and other data will be presented within the field of design variables that are needed for the framework development.

In a CE scenario, design should facilitate the second use and the final disposal of the product through:

- A proper LIBs labelling system (e.g., bar codes and RFID tags)
- Standardization of formats, structure and composing materials
- Reversible assembly strategy
- Clear classification of inner hazardous components [62]
- Reduced number of different materials used
- Standardized connections
- Reversible connections
- Accessibility of each component
- Modularity
- Scalability

To this aim, the proposed framework will be applied to LIBs complex architecture, starting from a clear definition of the domain of application. Given the purposes of the method, indeed, the analysis will be carried out on the aspects which mostly impact on the re-manufacturability and reuse of the product.

In particular, the aim is to find the best design to reduce the costs, the efforts, and the level of danger of remanufacturing LIBs. Nowadays, the economical unfeasibility represents the first barrier to the choice of remanufacturing for reuse in the EV sector, and other applications are preferred as more economically viable.

Therefore, the application domain will cover the design of LIBs starting from the cell considered as an atom. The chemistry of cells will not be investigated, nor further information on different available chemistries will be provided.

This is also due to the current **direct logistic** level of integration. At the moment, no cells producers are present in Europe; OEM or battery assemblers buy the cells from Chinese and Korean producer, which do not communicate anything about the chemical specifications of the cell. There is a big gap in the information system and the level of integration along the supply chain is very low, almost absent. Many times, thus, it is not possible to trace back the chemistry of a cell given the battery pack in which the cell is assembled. This fact also enhances the problems related to a fast diagnostic phase of the battery at its end-of-life. The only exception is Tesla, which produces its cells internally and which is the only OEM to deploy Cylindric cells in the LIB battery pack.

Given the aim of the methodology application, which involves the re-design for reuse and remanufacturability, better solutions will be searched for in terms of

- Redesign of some components starting from the cell level, with particular interest on joints and connections
- Assembly structure, maintaining original design for some components and modifying the design of others.

Without having impacts on single cells composition, and with no attempts to solve chemistry-related issues.

6.2.1 Current Situation in CE approaches implementation

Residual capacity of EoL LIBs from EV is on average the 80% of its initial capacity. This represents a great opportunity for reuse and remanufacturing. Despite this, due to already mentioned issues related to costs and to the difficulties in the assessment of battery features, solutions in this direction are far from being implemented.

As concerns the assessment and evaluation of batteries features, different research have been conducted to determine the remaining life and LIBs degradation phenomena and the first experimental systems were developed thanks to the collaboration between EVs manufacturers (e.g. BMW, GM and Nissan Motor) and energy management companies (e.g. Vattenfall, ABB and Sumitomo) [68].

Along with a fast and efficient diagnostic of LIBs status, another major challenge for remanufacturing and reuse is a safe and non-destructive disassembly of battery pack, based on automated process able to overcome product variability. At the moment, remanufacturing of EV battery pack is manual: it is a laborious, time-consuming, costly process which also involves safety issues given by the toxic and inflammable materials as well as the high voltages. The issue of reliability and safety is found to be one of the main rationales behind the OEMs' reluctancy to give their used battery packs to third parties. A manual process is still the only viable solution, given the highly specialized features of each battery pack model. Every manufacturer's design has peculiar features that prevent automation of the main disassembly operations required for remanufacturing. Leaving apart automation, even manual operations are far from being established. Learning curve of operators cannot reach a steady state if the product features change continuously, and if the OEMs keep their specifications secret.

Due to all these issues, so far, the only implemented CE solution for LIBs is Recycling.

As explained in the introductory part, recycling is a complex and consuming process even for simple products.

The high complexity of LIB packs increases and enhances the needs related to a proper recycling system design.

At the moment, according to the combined effect of recycling feasibility and final gain, only Cobalt (Co), Copper (Cu), Steel, Nickel (Ni) and Aluminium (Al) are recycled, while plastics are incinerated for energy recovery and Lithium, Manganese and graphite are rarely considered.

This selective nature of the current recycling system can not be considered as a sustainable solution. Also, considering, among the other things, the predominant trend to substitute cobalt in order to lower production costs, recycling processes should be developed to recover LIBs regardless their specific compositions and to balance treatment costs with the final effective revenue.

An overview on the current recycling systems is presented in [62], with state-of-the-art technologies as concerns (a) Preliminary waste preparation phase, (b) Thermal, Mechanical, Physical, Mechano-physical, Chemical and Mechanochemical pre-treatments and (c) effective Treatment for metal extraction of metals enriched separated fractions.

"In an optimized circular economy model, ***wasted LIBs management starts with product design****, developing systems easy to be reused and recycled and minimizing the amount of materials to be landfilled or incinerated. Furthermore, the residual features of end-of-life LIBs would be tested, in order to promote the reuse or to suggest remanufacturing solutions for new secondary applications.* ***Recycling processes should be used as final option****, developing treatments with the highest recovery efficiency and the lowest environmental impact, allowing primary raw material saving, economic gains, energy consumption reduction, waste minimization and safe management of harmful components."* [62]

Through the large-scale implementation of the proposed methodology, important improvements would be brought to the process of reuse and remanufacturing of LIBs pack.

Also, the aim of the method is to reduce the current space of solutions, through the identification of a unique standard, or a set of optimal configurations, which fully satisfy functional requirements and circular economy needs.

The way to proceed will strictly refer to the framework of reference presented in Graph 1 and adapted as in Figure 47:

1. Evaluation of current solutions
2. Exploration
3. Identification of solutions deemed satisfactory OR modification of the space of solution with the consequent generation of innovative designs.

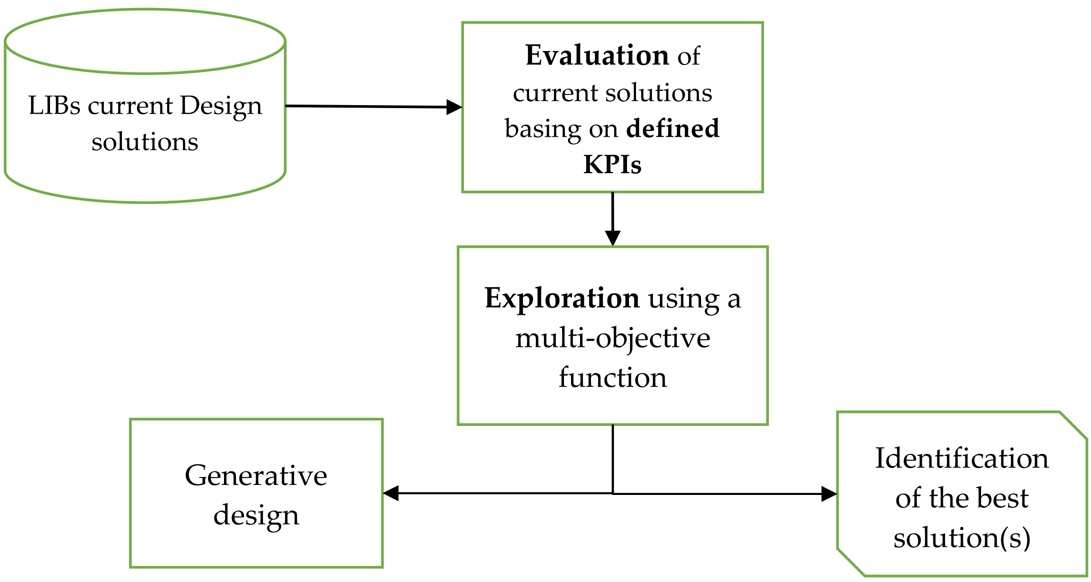


Figure 47- Framework application to LIB packs

The output of the methodology will also go in the direction of solving issues and conflicts between European commission and industrial associations like EUROBAT (Association of European Automotive and Industrial Batteries Manufacturers).

EUROBAT welcomes the initiative of the European Commission on Modernizing the EU's batteries legislation: in the past years, EUROBAT remarked several times the need to adapt the legislative framework on batteries to take into account the increased importance of batteries to decarbonize our economy. A coherent legislative framework is needed, considering the overlaps between the Batteries Directive, the End-of-Life Vehicles Directive, REACH and Occupational Safety and Health (OSH).

On the other hand, the association do not regard positive the introduction of Regulations which set a minimum threshold on batteries performance, and which are willing to introduce a standard without carefully considering all the specific industrial needs and the actual relations between batteries performances and technical specifications. [69]

6.3 Strategy Definition

6.3.1 Common Strategy

An important point to stress, which is linked to a higher strategical level with respect to the one tackled in the Methodology section, is related to the already mentioned co-opetition between companies. In this kind of industry where many configurations exist on the market since every OEM adopts its own solutions to satisfy needs and constraints, common effort should be put in the path towards the implementation of this methodology. The aim should be that of finding better solutions and configurations which allow to overcome the issues linked to the end-of-life management of batteries, therefore providing advantages to the battery industry. The shared interest in this issue should be straightforward.

Sharing information could become crucial for an efficient performance of the phases of evaluation and exploration, given the need to support these processes with machine learning algorithms which requires many data to be trained. This step is fundamental given the huge number of aspects to be considered for assessing the performances of a LIB pack.

For reaching the stability of battery pack industry, which can be only made possible through the implementation of CE principles, common effort must be shared among all the companies involved. Big innovations and improvements, indeed, can only come with a large commitment and with the change of mind-set about the concept of competition.

The co-opetition strategy, thus, deals with the need to create a stable and sustainable battery industry, given the shared interest in this issue.

After this, each company should perceive its own strategical business objectives, playing with the levers seen in section 5.1.

Besides this highly strategical aspect, indeed, matters related to the business strategy of each company have to be defined.

6.3.2 Single company strategy

Due to the huge complexity of this product, and to the number of actors involved in the issue, the strategical aspect can be tackled assuming two different points of view: that of the battery producer (assembler) and the one of the OEM. In some cases, like Tesla, these two figures coincide; in some other cases, the OEM do not physically assemble the product but provides the assembler with design guidelines, so it can be considered as the case of Tesla. There are also cases in which the design is made by the assembler, and the OEM just purchases the battery pack as a black box, only providing specifications about the size, weight, performances, safety, and accessibility). This last case will be addressed considering the OEM like a B2B customer of the process.

In any case, the strategic aspect involves the establishment of a proper system of design, production, distribution, collection, and treatment of LIB packs.

As seen in the general methodology, these aspects can be organized within the business model framework, which help in structuring the approach.

The correct choice of the business model, or the adjustment/modification of the current one, can be an important booster especially in this kind of industry. Playing with the relationships established with customers and with key partners, intended as both direct and inverse value chain suppliers, can become crucial in determining the possibility to properly implement the methodology.

Leaving apart particular boxes of the business model canvas, which must be strictly linked to single companies' decisions, and to the specific market segment they want to serve, common guidelines are presented.

The new design will be directly addressing the specific needs related to remanufacturing for function restore and upgrade, considering the case of adapting the battery to the changes and improvements in technology without completely changing the battery pack. This would bring enormous advantages for both manufacturers and end-users, given the lower cost of modifying an existing battery compared to the design and production of a new pack, which reflects in a lower price for the customers.

Another addressed approach is the reuse of end-of-life batteries considering both the same purpose (inter-sectorial approach), with the possible deployment of the battery on less-performant vehicles, and a cross-sectorial approach linked to the use of the battery for stationary charge activities. For example, a growing market opportunity can be represented by the storage of renewable energy installations or within living environments (home, office).

For doing so, the choice of key partners is fundamental and must adapt to new design specifications which will come as an output of the implementation.

The introduction of a new design, or even the consciousness about the fact that an optimal configuration exists, should be accompanied by a structured building of the value chain, and a careful management of the end-of-life and disposal system which is being tackled by European commission and by responsible producers' organizations (EPR).

The strategic point lies in the re-organization of OEM business models according to the new design created for supporting remanufacturability and reuse. Even though the new design establishment goes in the direction of a unique standard, or at least a homogenization of the configurations, differentiation between companies is not undermined, and it will be given by the decisions of different companies on how to build or adapt their business model.

To summarize, the aim of the methodology application is to find the optimal configuration in terms of

- **REMANUFACTURABILITY FOR FUNCTION RESTORE AND UPGRADE.** Which includes requirements related to disassembly, cleaning, sorting, inspection, testing, and reassembly systems that will be different from the current systems adopted for recycling

- **ENHANCEMENT OF THE LIFECYCLE WITH POTENTIAL REUSE FOR THE SAME PURPOSE OR FOR OTHER INDUSTRIES**. In this direction, the strategical aspect assumes an even more important role. With the adoption of a unique optimal standard, OEM serving the higher market segments can for example create a business model based on the continuous improvement of EV battery packs thanks to their scalability and standardization of components. In this possible scenario, a system can be created for which customers belonging to higher market segments are encouraged to bring back to the OEM their car for having their battery upgraded, while "obsolete" components can be recovered by the manufacturer and deployed to other batteries which will be sold to lower market segments. In this case requirements related to modular and scalable design with interoperable components are fundamental.

6.4 KPIs Setting

In this section, attention is posed on the definition of key performance indicators at product design level. As mentioned in the methodology part, the definition of these KPIs is linked to the specific product of reference and to the strategical decisions of the company. The set of KPIs must be a combination of measures assessing the performances of the product design in terms of:

- Defined CE needs and requirements. Which in this case are Remanufacturability and Reuse.
- Functional features which determine the success/unsuccess of the product design on the market, thus, features liked to customers satisfaction.

These metrics will be then used to evaluate existing configurations.

10 KPIs are here listed as considered to be fundamental for a complete evaluation of solutions basing on the scope of this analysis.

Since each KPI is the future output of the models trained to evaluate solutions, the output of each regression model will be a numeric variable describing the ability of each combination of features to respond to customers' and Circular Economy requirements. All the needs linked to a specific requirement have therefore to be aggregated in a few data able to synthesize many different conditions.

In particular, as concerns the Remanufacturing needs, the following table was drafted for identifying the product features mostly having an impact on each stage of the process.

Remanufacturing Stage / Related product features needed	Disassembly	Sorting	Inspection	Testing	Reassembly	Reconditioning	Cleaning
Ease of Identification	X	X	X	X			
Ease of Verification		X	X				
Ease of Access	X	X	X	X		X	X
Ease of Handling	X	X			X	X	
Ease of Separation	X	X				X	
Ease of Securing					X		
Ease of Alignment					X		
Wear Resistance	X				X	X	X

Table 7- Remanufacturing stages and related product design needs

The following key performance indicators were extracted:

1. Recharging time
2. Discharging rate
3. Maintenance effort (considering both time and money)

As concerns the end-users' requirements satisfaction.

4. Easiness of assembly

For keeping into account those cases in which OEM purchase the battery from third parties.

5. Easiness of battery performances assessment, implying the accessibility and identifiability of components, which in turn imply the standardization of components.

6. Safety and easiness of disassembly.

   This KPI is particularly crucial since it gives evidence of the presence of reversible connections and joints (Easiness of separation).
   Also, the definition of this KPI allowed to validate the hypothesis mentioned in the introductory part, for which the optimal disassembly level for a LIB, within the scope of this study, is the cell. Therefore, reversible connections are needed also between cells belonging to the same module and no investigations will be done at a higher level of detail (cell disassembly).

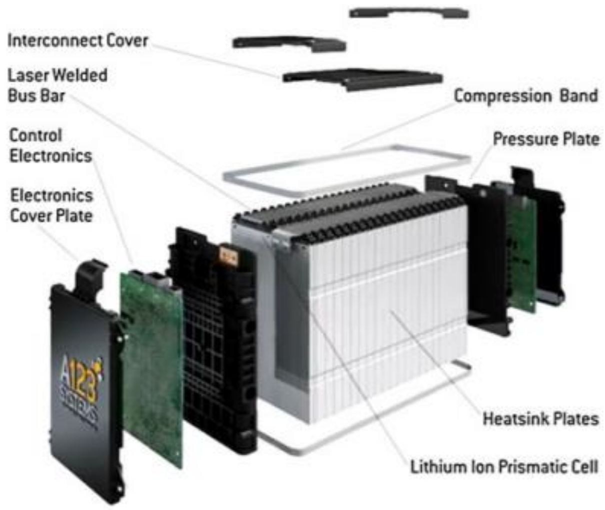
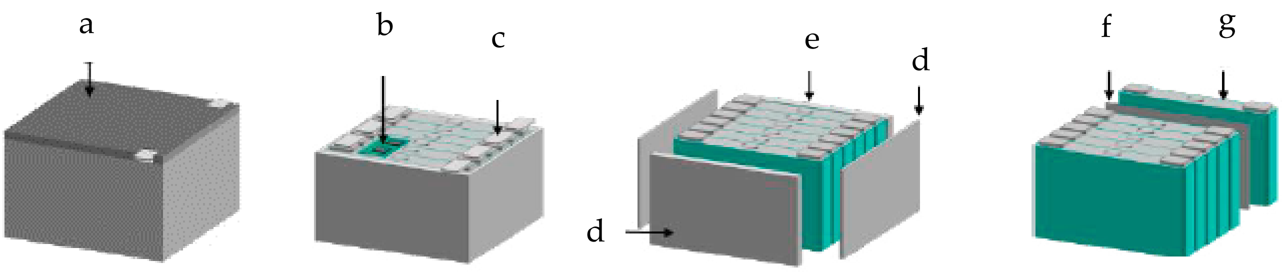


Figure 48- LIB module exploded view



a: cover
b: BMS
c: cell connector (busbar)
d: side wall
e: compressive plate
f: adhesive pad
g: prismatic hard case lithium-ion cell

Figure 49- LIB module view until cell level

7. Accessibility of electrodes at module level

8. Modularity and homogeneity of materials in components with the same function

9. Accessibility of electrodes at pack level

10. Voltage scalability

Following the methodology path, these KPIs will be assigned to the existing configurations with the aim to **Evaluate** them through a regression algorithm.

6.5 Solution Space Definition/Modification

In this section, input data will be listed using the framework of reference, thus identifying the specific information related to EV LIB packs application.

6.5.1 Product Definition: Data coming from the market

Data regarding the expectations of the customers, i.e., customer's requirements that have to be satisfied by the product.

B2B → CAR MANUFACTURERS/PRODUCERS.

- Size of the pack : numerical attribute
- Weight of the pack : numerical attribute
- Cooling system: binary attribute indicating whether the cooling system is managed through automated thermal management systems or through passive cooling techniques.
- Cooling system type of connections: categorical attribute describing mechanical joints existing between cooling system and other components
- Type of Junction Block: categorical attribute describing mechanical, electrical and electronic connections of the junction block.
- Cost of the battery: numerical attribute
- Expected cycle life
- Expected calendar life
- Type of crash protection system present in the BMS
- Type of vibration control system present in the BMS
- Type of crash thermal resistance system present in the BMS

- Type of mechanical connections of the pack: categorical attribute indicating the easiness of assembly of the pack in the car.
- Number of mechanical, electrical and electronic connections of the pack: numerical attribute

B2C → END USERS OF THE ELECTRIC VEHICLE.

- Recharging time: numerical attribute
- Discharging rate: numerical attribute
- Cost of maintenance: numerical attribute also linked to mechanical connections thus accessibility of all components
- Time of maintenance: numerical attribute also linked to mechanical connections thus accessibility of all components

Data about the way the customer uses the product

This information would be fundamental in order to develop an algorithm which optimizes the aspects linked to the customers' satisfaction. Giving as input to an optimization algorithm or to a generative one a set of constraints containing thresholds of the following variables, it would give as output configurations specifically targeting these issues. In particular optimization algorithms would most probably provide an optimized and upgraded version of already existing configurations (exploitative approach) while generative algorithm will explore innovative solutions not yet evaluated (explorative approach). These data result to be fundamental for assessing the state of the battery at its end of life.

- SOH: numerical attribute
- Average amount of kilometres run before recharging
- Average time elapsing between consecutive recharges
- Type of recharger used
- SOC at which the battery in recharged on average
- Highest and lowest external temperatures at which the battery is subject.

Data regarding the fundamental features of the product

- Ability to properly generate electric current through chemical reaction and transfer it to the electric engine. Characteristics of the reaction (we will not investigate what happens in the cell).
  As a remark, joints play a fundamental role in the satisfaction of fundamental features and safety issues. Joints are responsible of the propagation of energy, beside covering structural needs.
- Modular architecture
- Safety
- Ability to precisely communicate the level of residual energy stored (no latency and no errors)
- Functionality of the junction block. As explained, it is mainly composed by the central relay that is responsible for the supply of DC from the battery (and it is immediately de-activated in case of malfunctioning), the pre-charge relay (protecting high-voltage circuits), and the sensor measuring the battery current.

Data regarding those features that are not fundamental but highly recommended and expected.

- Lightweight
- Space optimization

Data regarding those features and functionalities that are not fundamental and not necessary but that, if present, represent a great advantage for the product in the market.

Service associated to product. In this set of data are contained all the decisions which must be taken at single company level about the business model.

For instance, maintenance is no more responsibility of the customer but of the OEM.

Integration along all the supply chain that is a fundamental feature for an optimized CE approach.

Summarizing, the main topic related to these features is maintenance; customer is maintenance free, and product is designed in such a way to guarantee fast and cheap maintenance.

6.5.2 Design Variables: Technical Specifications

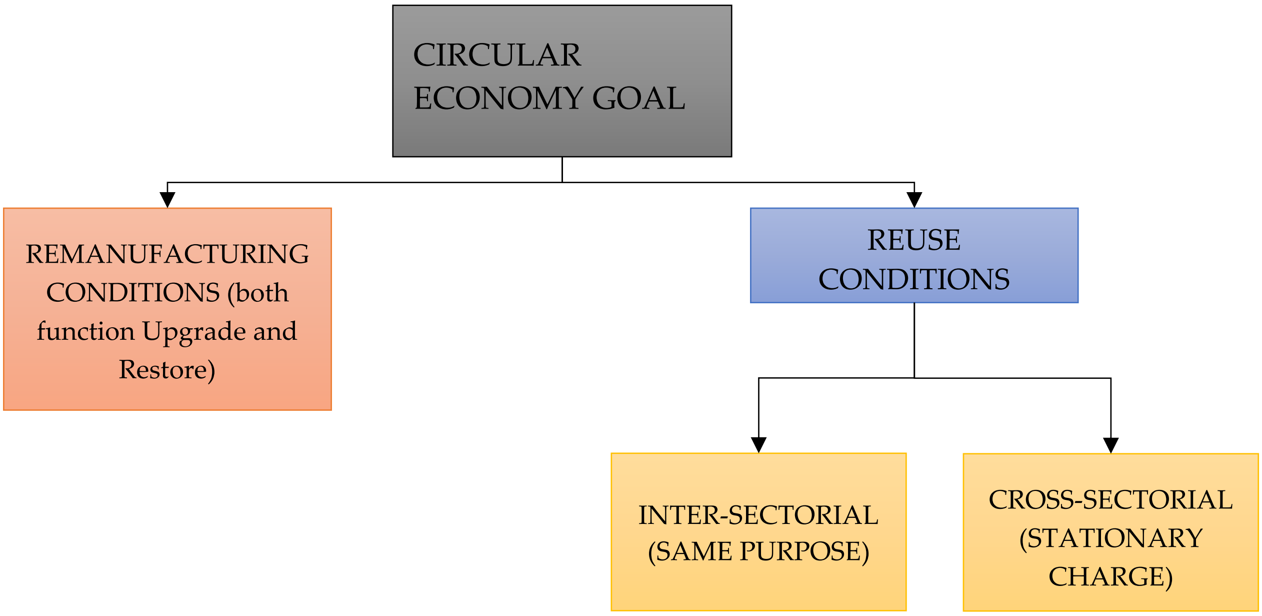
- Type of electrical and electronic connections at different interface levels:
    - Cell-cell interactions → investigate joints
    - Cells-module interaction → investigate joints and case assembly
    - Module-module interaction → investigate joints
    - Module-pack interaction → beside connections between modules and the case, junction block, BMS, service plug.
    - Pack- extern interaction → case and joints connecting pack to electric engine, control unit of the car, e-charging station, refrigerating part.

  It is a set of categorical attributes describing the current types of connections, especially those allowing the energy flow to the junction block and between modules.
- Average lifetime of electrical and electronic connections: numerical attribute linked to proper propagation and transmission of energy and to safety issues.
- Type of cells in each module: categorical attribute indicating whether cells are prismatic, cylindric or pouch.
- Number of cells in each module.
- Type and number of connections in a module: categorical attribute indicating whether cells are in parallel or in series, and the number of connections in series and in parallel. This information is linked to constraints on cells homogeneity in terms of size and capacity.
- Number of modules in a pack.
- Number of different materials used in each mechanical joint.
- Material composition of the case
- Type of assembly of the case
- BMS topology: categorical attribute indicating whether the BMS is arranged according to a distributed, centralized or modular topology.
- Types of sensors in the junction block.
- Type of cooling system: categorical attribute indicating whether the cooling system is Active Air, Passive Air or Liquid Temperature Control.
- Number of cables connecting cells and modules to the BMS.
- Module-module interface.

- Modules- cooling system interface.

6.5.3 Circular Economy Constraints

Input constraints come from the need for a battery pack to be easily reused and remanufactured, thus, to be efficiently disassembled and to quickly have access to all the main components for cleaning, testing, substituting them.



Graph 4- Goal composition

A fundamental set of inputs for the methodology development is related to constraints set on product features which mostly have an impact on remanufacturing stage and reuse. The first step, thus, is to correctly identify the impacting features.

To this aim, research about design for re-manufacturability have been reviewed and the following table have been drafted in order to clearly identify product specifications and characteristics which have an impact on different remanufacturing stages. [70]

Remanufacturing stage	Guidelines
Ease of Sorting	**i.** Reduce the variety of products and parts **ii.** Provide clear distinctive features that allow for easy recognition **iii.** Provide readable labels, text, and barcodes that do not wear off during the product's service life
Ease of Disassembly	**i.** Avoid permanent fasteners that require

	destructive removal **ii**. If destructive removal is necessary, ensure that damage to the core does not happen **iii.** Reduce the number of fasteners prone to damage and breakage during removal **iv.** Increase corrosion resistance of fasteners **v**. Reduce total number of fasteners in the unit **vi**. Reduce the number of press-fits **vii.** Reduce the number of fasteners not in direct line of sight **viii.** Standardize fasteners by reducing the number of different types of fasteners and the number of different sized fasteners.
Ease of Cleaning	**i.** Protect parts and surfaces against corrosion and dirt **ii.** Avoid product or part features that can be damaged during cleaning processes or make them removable **iii.** Minimize geometric features that trap contaminants over the service life **iv.** Reduce the number of cavities that are capable of collecting residue during cleaning operations **v.** Avoid contamination caused by wear
Ease of Inspection	**i.** Minimize the inspection time **ii.** Reduce the number of different testing and inspection equipment pieces needed and the level of sophistication required **iii.** Provide good testing documentation and specifications
Ease of Part Replacement	**i.** Prevent damage during part insertion **ii.** Provide good documentation of specifications and clear installation manuals
Ease of Reassembly	**i.** Minimize the time required to reassemble the product **ii.** Make the assembly sequence efficient with as few steps as possible **iii.** Provide good documentation of specifications and clear installation manuals **iv.** Leave surfaces available for grasping **v.** Maximize part symmetry **vi.** Eliminate tangly parts

	vii. Color code parts that are different but shaped similarly **viii.** Insert new parts into an assembly from above, or from the same direction; never require the assembly to be turned over **ix.** Proper spacing ensures allowance for a fastening tool
Standardization	**i.** Standardize and use common parts and materials **ii.** Standardize and use common fasteners **iii.** Standardize and use common interfaces **iv.** Standardize and use common tools
Reusable Components	**i.** Design a reusable platform and reusable modules **ii.** Select materials to ensure reliability and durability of the product **iii.** Make sure components are robust enough to reuse without replacement **iv.** Avoid toxic materials

Table 8- Remanufacturing stages related requirements

According to this table, synthesizing the information, for the Re-manufacturability for function upgrade and restore, here is the list of conditions that have to be satisfied, i.e., constraints that have to be set as input within the solution space definition.

- Cell geometry standardization
- Standardization of connections
- Reversible mechanical connections at single module level
- Accessible electrodes at **module** level
- No temper-resistant screws
- Minimized number and type of connections
- No hidden or non-accessible joints
- No glue and adhesives
- Have each single component labelled
- Maximise identifiability of functions
- Maximise modularity and homogeneity in materials used for each component/ for components with the same function or belonging to the same sub-assembly.
- Maximise standardization and simplicity in architecture
- Minimize disassembly directions

- Ease of stacking/storage at **module** level
- Minimize modules weight
- Provide grasping elements
- Identify high voltage components (for safety reasons linked to disassembly)
- Minimize short circuit triggering
- Maximise the separation of components with a different lifecycle in order to make them easily disassembled and reuse those components with higher lifecycle.

Requirements specifically targeting the reusability for the same purpose:

- Accessible electrodes at **pack** level
- Clear and proven testing parameters
- Ease of stacking/storage at **pack** level

Conditions for cross-sectorial reusability, in particular for stationary charging:

- Accessible electrodes at **pack** level
- Clear and proven testing parameters
- Ease of stacking/storage at **pack** level
- Scalable voltage due to the need to invert direct current (of Batteries) with alternate current (every device operates with alternate current). Inverters are not customizable and only exist at 600V or 50V and batteries are exactly in the middle (on average 300V).

6.6 Current Design: Available data ETL

After having described the key parts composing any design approach, it results evident that, in case of already defined solution space and already existing solutions within this space, improvements only come with a correct evaluation of the current situation. This comes with a proper and detailed structure of the models for evaluating existing solutions.

In this case study, before any trial to extend the solution space and find new configurations, attention is posed on a structured approach to evaluate solutions, through the setting of Key Performance Indicators (defined in section 6.4) derived from the main purposes of the study:

- Create circular products

- Meet customer needs.

Even in this practical application of the methodology, the followed approach had been a problem driven one. As stated in previous section, a broad list of data was initially drafted starting from the established strategical requirements, the examined system boarders, and consequent needed information.

Here, data have been carefully selected and formalized in order to arrive to the final set of features to be given as input to the evaluation stage.

Most relevant features identifying a specific battery pack configuration have been chosen among the complex list of attributes that can be used to describe a pack configuration, coming from the three macro areas comprising Product Definition, Design Variables and Circular Constraints.

In particular, the first, very broad, list of features was the following:

1- Pack size [m^3]
2- Pack weight [kg]
3- Module size [m^3]
4- Module weight [kg]
5- Module capacity [Ah], coinciding with battery pack capacity.
6- Module voltage [V], which is extracted from the voltage of the whole battery pack by dividing it for the number of modules since they are always connected in series.
7- Cells size (small, medium, large)
8- Cells type (Pouch, Prismatic or Cylindric)
9- Type of cell-busbar connections
10- Number of cells in a module
11- Number of series connections in a module
12- Number of parallel connections in a module
13- Number of modules in a pack
14- Type of cooling system [liquid, air]
15- Cooling system management [active, passive]

16- Cooling system - single module interface [0 if cooling system do not cross each single module, 1 otherwise]
17- Cooling system single module interface [n=number of connections between single module and cooling system]
18- Level of heterogeneity of mechanical connections between the battery case and the vehicle [high-medium-low]
19- Presence of Mechanical joints connecting modules
20- Number of Mechanical joints connecting modules and the pack
21- Type of electrical connections between the modules
22- Presence of mechanical connections between modules
23- Type of mechanical joints connecting the Junction block to the cells and to modules
24- Type of electrical joints connecting the Junction block to cells and to modules
25- Way of assembly of the pack
26- Way of assembly of the pack to the vehicle
27- BMS topology
28- Quantity of cables
29- Presence of glues/adhesives [high-medium-low]

These data have been further structured, lightened, and formalized. Also, correlations between data have been eliminated by keeping only one variable out of the ones that were clearly linearly correlated.

The resulting list of features is reported in Table 9, where data are presented in terms of range of values assumed and related unit of measure.

Design Variable	Values Assumed	Unit Of Measure
Pack weight	Numerical from 100 to 500	Kg
Module Size	Numerical from 4.500 to 49.000	Mm^3
Type of cooling system	Air / Liquid	/
Cooling system management	Active / Passive	/

Mechanical joints between battery pack and EV	From 8 to 20	Screws
Number of different screws at pack level	From 1 to 3	Different types of screws
Total number of screws at pack level	From 16 to 60	Screws
Metal sheet folding	Yes ; No	/
Cooling System – Module interface	Yes ; No	/
Junction Block mechanical joints	From 4 to 10	Screws
Junction Block electronic connections type	Plug ; Direct connection to PCB	/
Junction Block electrical connections type	Plug ; Bolt	/
Module Capacity (coinciding with Pack Capacity)	From 5 to 230	AmpèreHours [Ah]
Module Voltage	From 3,7 to 100	Volt [V]
Type of cells	Prismatic Pouch; Prismatic; Cylindric	/
Type of electrical connections between cells	Ultrasonic welding; Ultrasonic wedge bonding; Laser welding; Resistance Spot/Projection Welding; Mechanical Assembly	/
Number of cells in a module	From 4 to 1.200* (from 2 to 1200)	Cells
Series connections in a module	From 2 to 20	Number of series connections
Parallel connections in a module	From 1 to 4 for pouch and prismatic 150 for cylindric* (Tesla)	Number of parallel connections
Number of modules	From 4 to 12	Modules
Mechanical connections at module level	From 4 to 10	Screws
Electrical connections at module level	Plug; Bolt	/
Connections between modules	Yes; No	/

Sensors per module	From 2 to 10	Sensors
BMS topology	Distributed; Centralized; Modular	/
Adhesives	Yes; No	/
Glue	From 0 to 5	Level of presence of Glue (0= almost absent; 5= high amount).

Table 9- Input features

6.6.1 Data Collection

Data were collected through an analysis of the available configurations, searching through existing OEMs' Electric Vehicle models. Differences between configurations belonging to diverse manufacturers are many, and frequently differences can be also seen through LIB packs designed for different car models of the same manufacturer.

This additional complexity remarks the need for a collaboration and a cooperation between all the OEMs, in order to properly collect data that are continuously updated and validated.

For this initial implementation, data were gathered searching on the websites of Manufacturers and extracting them looking at LIBs disassembly videos. More accurate and valid data are available for OEMs without any need to modify the information system of the company, only through the sharing of information between enterprises. This potential should be exploited for a correct and complete implementation of this methodology, which will bring advantages to every actor.

As mentioned, for demonstrating the functioning of the methodology and the effectiveness of the framework, data have been gathered in a rough way, due to the impossibility to access to updated and precise data.

However, 125 observations were collected, belonging to 33 different OEM. Considered car makers were Audi, BMW, Chevrolet, Citroen, Chrysler, Fiat, Ford, Honda, Hyundai, Jeep, Kia, Lamborghini, Lancia, Land Rover, Lexus, Maserati, Mazda, Mercedes-Benz, Mini, Mitsubishi, Nissan, Opel, Peugeot, Porsche, Renault, Rolls Royce, Smart, Suzuki, Tesla, Toyota, Volkswagen, Volvo.

In particular, data were extracted from specific car reports like [71], [72], [73], [74], [75] Databases [76] [77] [78] , general reports on EV [79] and assembly/disassembly videos [80] [81].

For the 125 observations, all the features listed in Table 9 were collected (no missing values) and each raw (observation) was assigned with its related value of the KPI, basing on the acquired knowledge about the system.

All the KPIs have been transformed into numerical measures ranging from 1 (lowest level of performance) to 6 (highest level of performance).

Given the complexity of this activity, only one KPI was taken as a reference for testing the methodology, evaluating the existing configurations according to the measure "Safety and easiness of disassembly."

An important thing to mention is that the stage of assigning the correct value of the KPI to initial configuration is only needed at first iteration of the method, and do not represent a complex activity for manufacturers since they have all the needed knowledge base for performing such activity. It is needed only at first iteration for training the algorithms and creating the models. After that, generated configurations can be evaluated using the same models and scores related to each KPI will be automatically assigned by the trained algorithm.

6.6.2 Data preparation

Starting from the collected features listed in Table 9, preliminary actions were performed in a qualitative way, only basing on the knowledge base acquired through the study of LIBs functioning and dynamics.

This preliminary cleaning phase was done on the variables which were somehow linked one to the other, discarding them in order to avoid building incorrect models due to correlations. Also, the label indicating the OEM was discarded, since it is not relevant for the purposes of the analysis.

Linearly correlated features are for example the module voltage and the number of series connections in a module, since they are linked by the equation:

$$Module\ voltage = 3{,}7\ \times Series\ connections\ in\ a\ module$$

For this reason, the feature indicating the number of series connections in a module was discarded a priori.

The second important step in the data preparation phase was related to the choice of the data types. While for some variables it is straightforward since they are clearly categorical or clearly numerical, for other variables it is not immediate. There is not an absolutely correct or an absolutely wrong answer, but the decision strongly impacts on the model performances.

In this case, the variables with an ambiguous nature were

- 'Mechanical joints between battery pack and EV', which was finally treated as **Numerical.**
- 'Number of different screws at pack level', which was treated as **Categorical** since it only assumes 3 possible values, even if numerical.
- 'Mechanical connections at module level' were largely discussed and, despite assuming numerical values ranging from 4 to 10, they were treated as **Categorical** in the developed analysis. This is due to a decision based on the fact that the number of screws in this specific case actually provide a kind of "categorization" of the LIB pack, given by the way modules are assembled with the other components.
- 'Parallel connections in a module'. The other way round is valid for this variable. It assumes only value 1,2,3 or 4 for almost all the batteries deployed in cars, with exception of Tesla. Despite this, it is treated as **Numerical** since the meaning of this variable is closer to that of a numerical attribute.
- 'Sensors per module' is treated as a **Categorical**, for the same reasons valid for to the variable 'Mechanical joints between battery pack and EV'.
- Same applies for the variable 'Glue'.

For the categorical variables, the cleaning phase is, in most cases, attributable to the phase of size reduction specifically targeting attributes.

Categorical Variables

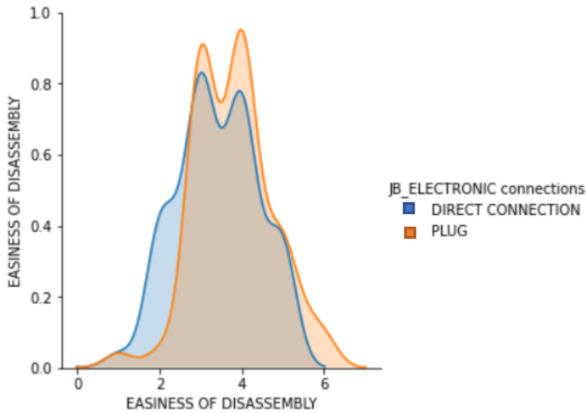
As explained, this stage of the framework consists in the creation of different models for evaluating solutions under a multitude of perspectives. Given the diversity of the perspectives adopted, the models will be different and will differ with respect to the considered attributes. This is way the phase of cleaning of categorical variables is fundamental and characterizes each trained algorithm: for each algorithm, each categorical variable is compared against the target variable, in order to assess the level of relevance and importance that the specific variable has in determining the values assumed by the target one. As a result of this, the models will be trained on a different sub-set of the entire set of data provided as input.

In this specific case, the selected KPI (target variable), i.e., Safety and easiness of disassembly., has been tested against each categorical variable at a time, in order to see how its distribution varied according to the values assumed by the categorical attributes.

This was done in two different ways in order to have a double check of results.

Beginning with a plot of the observations' distribution according to the KPI (x axis) values, divided in n graphs, being n the number of possible values assumed by the categorical variable under analysis. The graphs are visualized in one single plot for providing evidence of the differences in the distributions of data depending on the selected value of the categorical attribute.

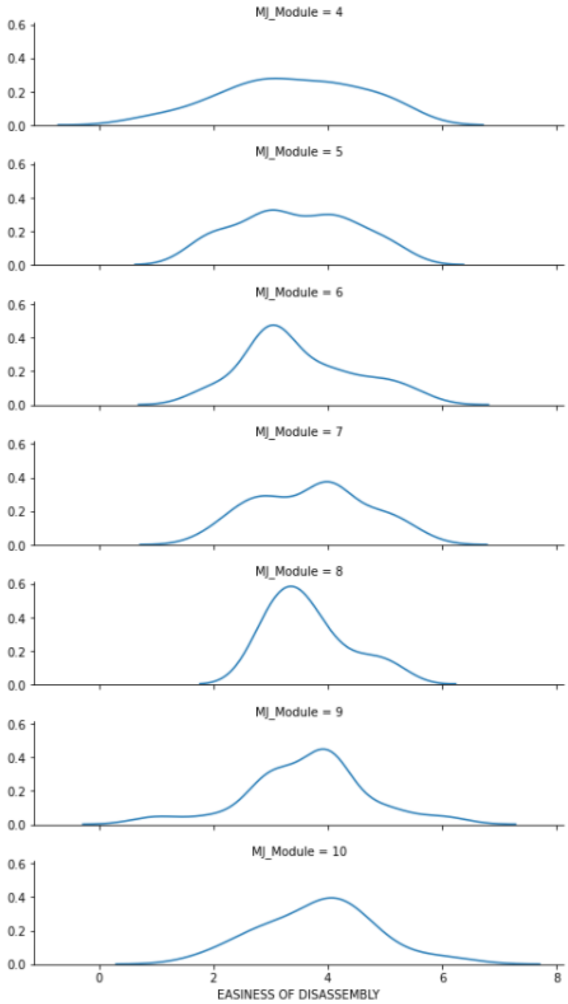
Graph 5 shows the plot built considering the categorical variable 'Junction Block electronic connections type'. In blue the distribution of observations when the value assumed by the variable was "DIRECT CONNECTION"; in orange the way data are distributed when electronic connections of the JB are "PLUG".



Graph 5- Plot of 'JB_ELECTRONIC connections' possible assumed values

The graph can be also split for having a clearer view of the different distributions of data, for instance in those cases in which the analysis is carried out on a variable which assumes more than 4 possible values.

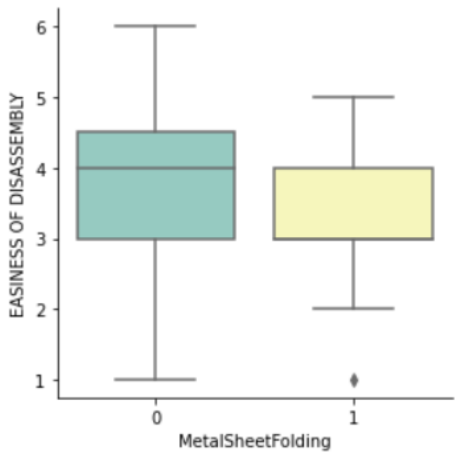
This is the case of 'Mechanical connections at module level', indicating the number of screws used to fix modules to the pack, whose plots are shown in Graph 6. On the x axis are always reported the values of the KPI.



Graph 6- Separated plot of 'MJ_Module' possible values

The analysis continued with a box-plot analysis, in which, similarly to what performed before, for each categorical variable, the distribution of data depending on its specific value assumed was plotted.

In Graph 7 is shown the box-plot obtained by the analysis carried out on the variable ‘presence of Metal Sheet Folding”. As reported in the graph, this variable is strongly influencing the distribution of values of the KPI.



Graph 7- Box-Plot 'presence of Metal Sheet Folding'

The analysis was carried out for all the attributes and the choice about the variables selection is based on the detected relevance of categorical attributes in determining the values assumed by the target one. In particular, the more the distribution of data according to the KPI varies with the values assumed by the attribute varying, the more that attribute will be relevant in determining the values assumed by the KPI.

As for the 'Presence of Metal Sheet Folding' variable, all the other categorical variables have proven to be relevant in describing the behavior of the target variable. (See Appendix B, B.1 for the remaining graphs).

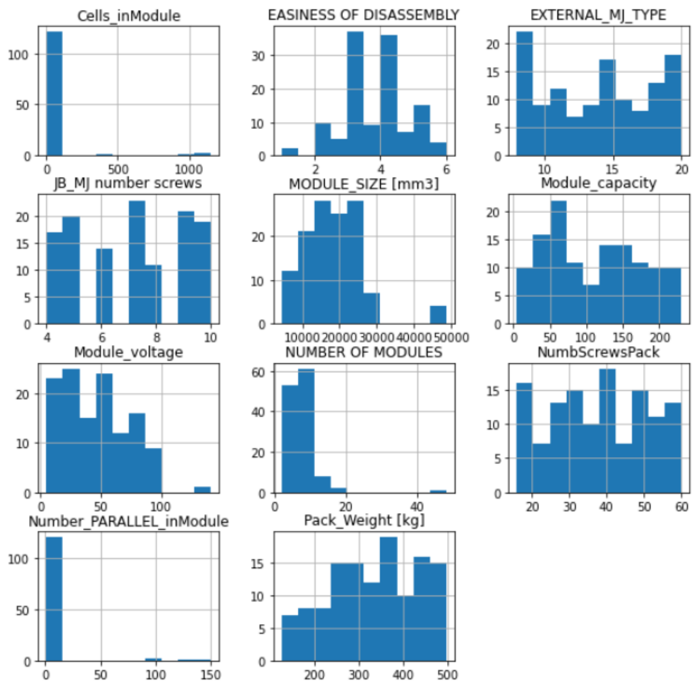
After that, selected categorical variables must be transformed into dummies in order to be readable by the algorithm.

Numerical Variables

Shifting the attention to numerical variables, different tools can be exploited for the analysis.

As a first step, numerical attributes have been plotted with simple histograms in order to see their distribution and detect the presence of normally distributed variables. As shown in Graph 8, no variables are normally distributed. Some present known distributions like 'Module Voltage' which can be considered as a Chi-squared distribution.

Some plots present peculiar shapes due to the presence in the dataset of infrequent and far-from-the- mean values. This is the case of variables indicating the number of cells in a module and the number of parallel connections in a module, which are "contaminated" by the presence in the dataset of LIB packs belonging to Tesla models. Without knowing the data and the situation, the suggested solution should be that of discarding the "strange" values, handling them as outliers. Nevertheless, having a base knowledge of the state of affairs, the best solution is to keep them in the dataset, in order to evaluate them with respect to the considered KPI.



Graph 8- Numerical Variables Histograms

Even the variable 'Number of Modules' presents a particular shape distribution given by the presence in the dataset of LIB packs having a big number of modules, like the ones deployed in the Audi e-tron and the Nissan LEAF.

No outliers have been removed for the explained reasons.

Next step in numerical variables preparation is the size reduction, performed with a correlation analysis. This stage has a double aim: on the one hand, similarly to what performed for categorical variables, the objective is to see how relevant each numerical variable is in determining the distribution of the target one. In this sense, the correlation analysis is performed in a bi-variate way considering the target variable and one numerical variable at a time.

On the other hand, through a multi-variate correlation analysis, the aim is to spot correlations between variables and consequently discard them.

As shown in Figure 50, the variables ‘Number of cells in a module’ and ‘Parallel connections in a module’ present a strong correlation. For this reason, ‘Parallel connections in a module’ has been discarded.

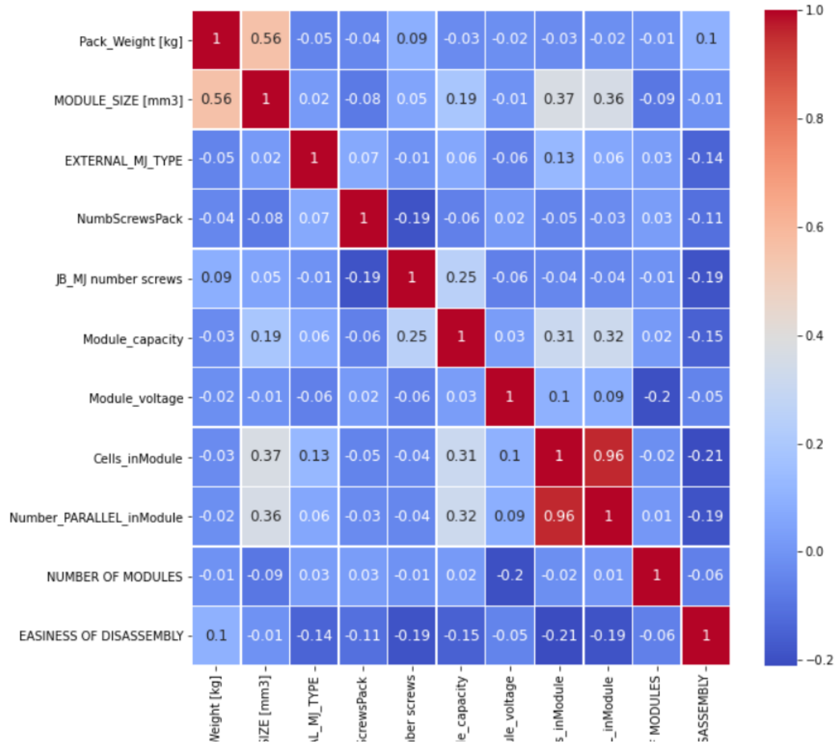


Figure 50- Heatmap of numerical variables

As for the assessment of the relevance of numerical variables in determining, or at least influencing, the distribution of the target variable, the process can not be streamlined to an observation of the linear correlation between each variable and the target one. This analysis is however helpful and a visualization of the pair-plot distributions of variables have been done and is presented in Figure 51. Y-axis always presents the value of the target variable ‘Easiness of Disassembly’.

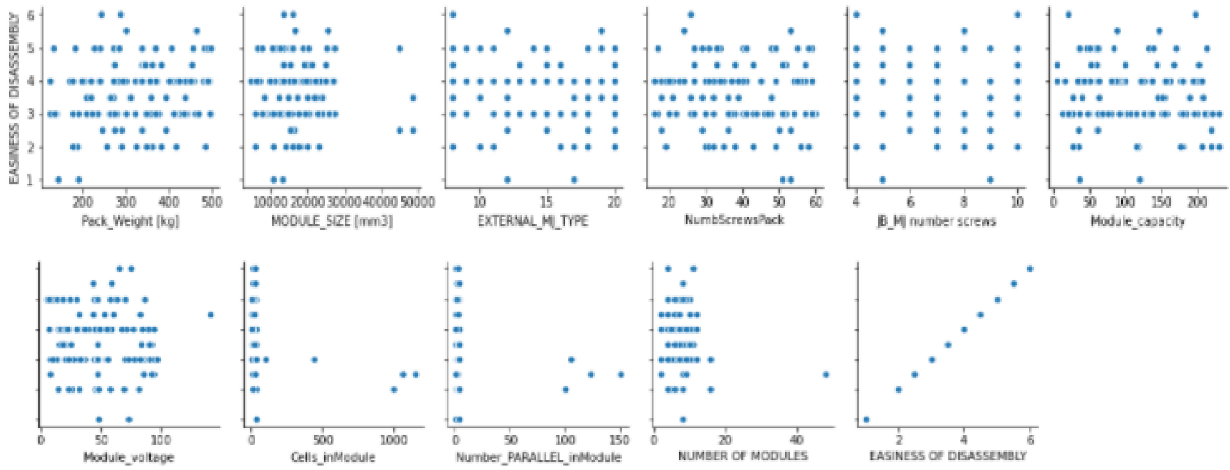


Figure 51- Pair-plot of numerical variables against 'Easiness of Disassembly'

As already clear in the heatmap presented in Figure 50, no linear correlation seems to exist between the target variable and any other numerical variable. This is due to many reasons linked to the large number of variables considered, to their consequent interaction and to the fact that these plots can only spot the linear correlation.

Given the objective of the analysis, some numerical variables have been discarded due to their irrelevant role in describing the easiness of disassembly. This lack of relevance is clearly inferable looking at Figure 51 and noticing the complete absence of patterns in some plots. These variables are: ‘Pack weight’, also given its slight correlation with the size of a module (Figure 52), Module Capacity and Module Voltage.

Figure 52- Extract from the heatmap

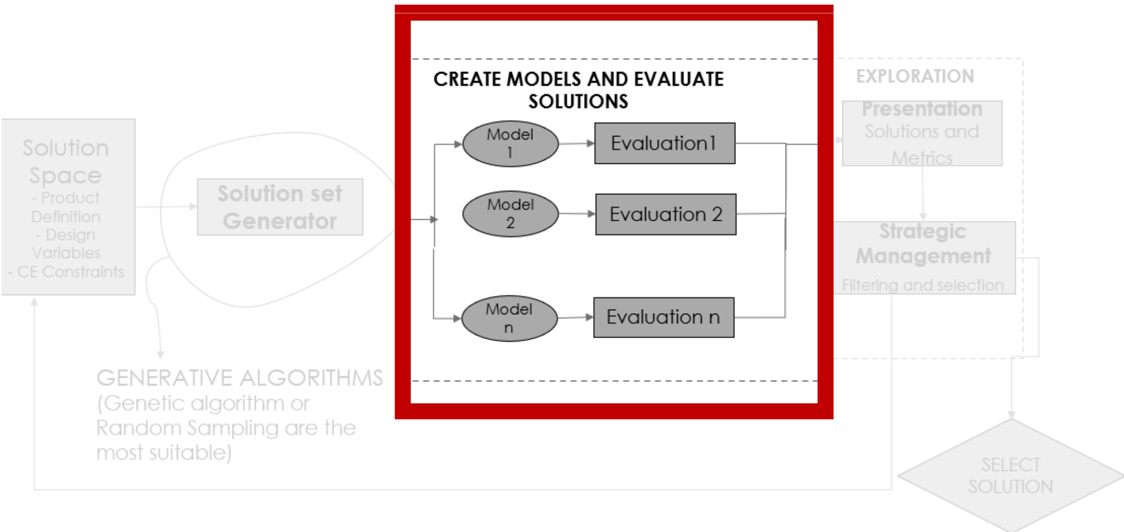
Once selected, numerical variables must be standardized in order to be treated by the algorithm without any bias.

Since the Dataset has been built ad hoc for the purposes of the analysis, no need for features extraction arose. Indeed, the needed information were already collected and the work of extracting the features was in some cases performed during the phase of data collection itself. For example, in most cases the Module Voltage was obtained from the Pack voltage, which is a more accessible information, dividing this last for the number of modules. Also, the Module capacity was mostly extracted from the total energy of the battery pack, which is one of the first information found on the websites. Energy of the battery expressed in kWh, indeed, is given by the capacity of the module (Ah) multiplied by the pack voltage (V).

With the end of the phase of data cleaning and preparation, the two subsets composed by categorical (dummies) and numerical variables have been concatenated in order to obtain a unique set of data.

The dataset is ready to feed the algorithm.

6.7 Evaluation



Graph 9- Problem setting

6.7.1 Algorithm Implementation and Selection

After having merged the cleaned and prepared numerical and dummy variables, the entire dataset must be split into training and test set.

Different models have been tested using the Grid Search tool combined with a cross validation as explained in section 5.6.

Results from Linear regression, Ridge regression, Lasso regression, K-Nearest Neighbor, Decision Trees, Random Forest, Support Vector Regressor and Multi-layer Perceptron regression algorithms implementation have been compared. (see Appendix B).

The best model results to be the one made with Lasso regressor algorithm (whose structure have been described in dedicated section - Supervised learning).

Best score: Lasso Regressor	*Negative Mean Squared Error = -0.550333*	
Best hyperparameters	Generalization term λ= 0.01	Normalize = False
PERFORMANCES	**Train set**	**Test set**
Mean Absolute Error	0.361	0.452
Mean Squared Error	0.215	0.333
Root Mean Squared Error	0.463	0.577
R squared	0.763	0.673

Table 10 - Performances of Lasso regression model

Note that the selected measure for evaluating each algorithm is the **Negative mean absolute error.** This is for a pure logical reason linked to the fact of scoring better algorithms with a higher value of the measure. Since we are dealing with errors which are measures of the badness of a model, we take the negative value in order to select the algorithm with the highest score (i.e., the lower absolute value).

As can be seen in the table presenting the results, the best Lasso model is the one with hyperparameter lambda (λ), which is the generalization term, equal to 0.01. This allow to have a very low level of overfitting, as shown by the values assumed by Mean absolute error, Mean squared error, Root mean squared error and R squared, which are pretty similar for training and test sets. R-squared is even higher for training set compared to test set, suggesting that no overfitting is present at all.

Higher levels of overfitting are achieved with the implementation of more complex models like Multi-layer perceptron, with a delta between the training and the test sets errors which reaches the 5000 percent. (see Appendix B, B.2)

The algorithm is now able to predict performances of new potential configurations, since it learned the intercurrent relations between the variables explaining the target behavior. By feeding the trained algorithm with unlabeled observations (i.e., observations with missing target variable), it would provide as output the target variable related to each new observation.

Given the poor quality of data collected, this algorithm only has been trained. Nonetheless, clear guidelines have been provided for a complete and correct development of the proposed framework. Also, tools and software are available for companies and industrial partners for performing the aggregation of the different measures in a single multi-objective function, with the aim to explore results and consequently decide whether to keep one solution or to modify the space of solutions and train the generative algorithm.

It is worthy to highlight how the development of regression models has a second very important outcome. Besides the primary purpose of evaluating different solutions and learning the existing relationships in order to predict future performances, they represent a strong instrument for validating the choice of data. This means that, through the proper training of algorithms, clear evidence of the actual significance of collected data is provided for different purposes and measures.

Resuming the general framework, another time, the information gain coming from the outcome of one iteration of the method enriches the input and the knowledge base of the system at the next iteration.

The outputs of this evaluation are a validation of the selected features and the identification of the current best configuration under the "Easiness of disassembly" point of view.

Validated variables are those which play a fundamental role in determining the performance of a configuration and are:

- Module size
- Type of cooling system
- Cooling system management
- External Mechanical joints
- Mechanical joints at pack level, module level, cells level and at Junction block level
- Different type of screws used for mechanical joints
- Presence of the external metal sheet folding
- Modules interfaces (between them and with the cooling system)
- Electric and electronic connections at junction block level, module level and cells level
- Type of cells
- Number of cells

- Number of modules
- Number of sensors per module
- Presence of Glue
- Presence of adhesives

The best configuration is the one which combines a quite small module size, a high number of modules containing a few cells, an active air cooling system which do not present interfaces with all the modules, the absence of the external metal sheet folding, the fewer number of screws used for mechanically joining the pack, the modules, the junction block and the cells, only one type of screws used for joining the same components, Plug electric connections and Laser welding when mechanical joints are not possible. Also, the absence of adhesives and glue and the fewer number of sensors per module.

6.8 Future Steps

The completion of the method can be performed with the training of the remaining 9 algorithms modelling the solutions and evaluating them under different perspectives. The steps to be followed are exactly the ones performed for the tested KPI and explained in previous section. This leads to a limited time consumption for companies for the activity completion. With the complete evaluation of solutions, after a proper exploration of results, companies should be able to decide whether to keep one configuration or to go for a modification of the space of solution and the generation of new design alternatives.

For this last task, performant software are present on the market and can be exploited by companies in order to explore the results of the evaluation phase, synthesizing them in a unique function and providing clear inputs to the generative design algorithm.

The proposed framework can be applied and extended to any other industry, providing guidelines and possible adjustments to be made for specific needs and strategies.

7. Conclusion and future development

Product design plays a fundamental role in the possibility to implement de-manufacturing at large scale. The phase of product design is complex and multi-stage; important improvements should be addressed in a collaborative way by all the companies involved. The efforts should go in the direction of a systematic identification and collection of data, which must be exploited for feeding algorithms and for deploying tools that can help to reduce the underneath complexity. Nonetheless, the exploitation of AI and machine learning can not be led back to their implementation on single operations: synergies must be exploited and a clear strategy have to be defined, which impacts on all the levels of the enterprise and of the entire supply chain. Collaboration and co-operation should be considered as important as the Machine learning tool itself, for guaranteeing the quality and relevance of data collected.

7.1 The method

The method has an iterative nature, and has been defined as follow: fundamental steps to be followed for reaching the objective are

1. a clear strategy definition, which should address issues at every level of the enterprise, in an integrated way
2. the setting of a complete and holistic set of KPIs, measuring company's performance at system, process, and product level
3. the utilization of a problem driven approach allowing a careful identification of fundamental data to be exploited in the product design phase. Such data must cover the areas identifying product definition, design variables and constraints
4. establishment of a systematic way to collect fundamental data, which is embedded in the method itself. If data are not available, design guidelines can be found for enabling the gathering of such data.
5. data cleaning and preparation for feeding algorithms able to validate such data and learn relations between product design and KPIs (Evaluation phase). This

allows to have clear and quantitative results about the ability of product configurations to satisfy those performances defined in step 2.

6. Exploration of results. In this phase human effort is fundamental for defining a unique measure which synthesizes the results coming from the upstream stage. Such results are represented by a set of models describing the behaviours of product configurations basing on different indicators. The identification of the best configuration under all the defined measures is the output of this step.
7. Decision on whether to be satisfied with the selected configuration(s) and consequently to end the process, or to re-iterate it, in a continuously improving perspective, moving on with step 8
8. modify the inputs data and feed a generative design algorithm which provides new design specifications in the form of unstructured data.
9. Go back to step 5.

The framework gives evidence of the fact that, in product design, AI supports humans in a collaborative activity. Humans have to take important decisions like selecting the most suitable methods for evaluating solutions (black box, white box), the choice of the metrics (KPIs), and the final choice on whether to re-iterate the process or to be satisfied with a certain solution.

The findings from this study also shed light on multiple aspects of generative design applications. The first of these is the importance of how a solution space is defined and how it should be given careful consideration not to be under- or over-constrained. Reducing the range or number of design variables, adding more constraints, or including more metrics is a way to potentially simplify the exploration process at the cost of potentially not finding novel solutions. Similarly, the inclusion of additional metrics, can result in a better evaluation of solutions that have the potential disadvantage of an increased complexity both in terms of the technical solution (i.e., calculating each metric) as well as in the analysis made by the users of a generative design approach. The exact selections of design variables, constraints, and metrics are considerations that might differ from project to project. Thus, the framework was designed to be adaptable and adoptable by different companies, with the potential addition of other setups than those used in this study, for reflecting the design context of other products and contexts.

7.2 Conclusions on the case study

The application of the framework to the LIBs pack design shed light to different interesting aspects mainly related to data identification and gathering. The framework was partially demonstrated through the training of a regression model developed for the evaluation of existing solutions under the "easiness of Disassembly" metrics.

The selection of variables came as a result of in-depth studies on batteries functioning and their production and distribution. The application of the evaluation phase allowed to validate the set of variables which are relevant and fundamental for assessing the easiness of disassembly of current and future configurations. Such variables are represented by:

- Module Size, which is an important attribute linked to the easiness of handling the pack and performing actions on it.
- Type of cooling system: air is better than liquid under the disassembly perspective
- Cooling system management
- Mechanical joints between battery pack and EV: the lower the better
- Number of different screws at pack level: the lower the better
- Total number of screws at pack level: the lower the better
- Metal sheet folding: better not to have it
- Cooling System – Module interface if present, is a further obstacle in the disassembly phase. Thus, better not to have it.
- Junction Block mechanical joints is always the number or screws employed, so the lower the better.
- Junction Block electronic connections type: better to have Plug.
- Junction Block electrical connections type: Plug is the best option
- Type of cells: generally Prismatic cells are associated with a higher level of easiness of assembly since they require easier-to-remove joints
- Type of electrical connections between cells: Mechanical and Laser Weld are the best options.
- Number of cells in a module: the lower the better
- Number of modules: the higher the better
- Mechanical connections at module level: the lower the better
- Electrical connections at module level: Plug is the best option
- Connections between modules: if present represent another obstacle to the disassembly activity.
- Sensors per module: the lower the better.
- BMS topology: better to have a centralized topology.

- Adhesives: if not present the disassembly is easier.
- Glue: the lower the presence of glue, the easier the disassembly.

The best configuration or configurations are those associated with a higher score of the KPI in the developed model.

Despite the implementation being incomplete, the framework has proven to be efficient and effective. The incompleteness is only due to the poor quality of accessible data. High quality data are in turn available to companies and the method can be easily implemented and exploited in all its potential.

Compared to related research in generative design and in the use of AI for designing products, this work contributes by including all the aspects related to the business management, by considering Circular Economy requirements as fundamental constraints when designing a product in order to simultaneously design its lifecycle, and by creating a more generic framework that shows the technical and strategical workflow of the generative design system.

It also contributes by further exploring the effects of the framework adoption on potential future iterations, explaining the benefits of a closed loop methodology aimed at continuously improve results.

The field of generative design and its application in the lithium-ion batteries for EV context shows promises and has the potential to be a part of a future designer's toolkit.

Bibliography

[1] "Ellen McArthur Foundation," [Online]. Available: https://ellenmacarthurfoundation.org/publications. [Accessed 2nd December 2020].

[2] Steinhilper , 2006.

[3] H. A. A. P. R. S. Tamil Selvan Ramadoss, "Artificial Intelligence and Internet of Things enabled Circular Economy," *The International Journal of Engineering and Science (IJES)*, vol. 7, 2018, Singapore .

[4] A. N. E. C. a. G. M. Marta Negri, "Circular Economy Performance Measurment in Manufacturing Firms: A Systematic Literature Review with Insights for Small and Medium Enterprises and New Adopters," *Sustainability*, 2021.

[5] S. Cayzer, P. Griffiths and V. Beghetto, "Design of indicators for measuring product performance in the circular economy," 2017.

[6] J. Evans and N. Bocken, "Developing a Tool for Manufacturers to Find Opportunity in the Circular Economy," 2014.

[7] Ellen MacArthur Foundation, "Circularity Indicators: An Approach to Measuring Circularity," 2015.

[8] McKinsey, "Growth within: a circular economy vision for a competitive Europe," 2015.

[9] C. Vercellis, Business Intelligence, John Wiley & sons, 2009.

[10] EllenMcArthur Foundation, "AI and the CE".

[11] R. Draelos, "Machine Learning," 2019.

[12] M. O. S. Mirza, "Conditional Generative Adversarial Nets.," 2014 .

[13] McKinsey Global Institute , "AI- THE NEXT DIGITAL FRONTIER?," JUNE 2017.

[14] HIGH-LEVEL EXPERT GROUP ON AI- set up by the European Commission, "Ethics Guidelines for Trustworthy AI," Brussells, 2019.

[15] B. E. Banko M, "Scaling to very very large corpora for natural language disambiguation," *Published online* , pp. 26-33, 2001.

[16] Smart Manufacturing Laboratory - State Detection, 2020.

[17] " https://landing.ai/ai-solutions-surface-defect-detection/," [Online]. [Accessed 2021].

[18] K. S. P. M. Vongbunyong S, "A framework for using cognitive robotics in disassembly automation. Leveraging Technol a Sustain World," in *Proc 19th CIRP Conf Life Cycle Eng. Published online* , 2012.

[19] L. S. G. B. P. R. F. A. Usuga Cadavid JP, "Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0.," *J Intell Manuf.* , vol. 31(6), pp. 1531-1558, 2020.

[20] L. X. L. L. Lu Y, "Multi-objective optimization of reverse logistics network based on improved particle swarm optimization," *Proc World Congr Intell Control Automation, Published online* , vol. doi:10.1109/WCICA.2008.4594085, pp. 7476-7480, 2008.

[21] G. M. R. K. Lee JE, "Network model and optimization of reverse logistics by hybrid genetic algorithm," *Comput Ind Engineering* , Vols. 156(3):951-964. doi:10.1016/j.cie.2008.09.02, p. 2009.

[22] A. Moltzau, "AI in Physical Product Design," 15 June 2019. [Online]. Available: https://alexmoltzau.medium.com/ai-in-physical-product-design-e67c02a8c2c1. [Accessed June 2021].

[23] H. A. A. P. R. S. Tamil Selvan Ramadoss, "Artificial Intelligence and Internet of Things enabled Circular economy," *The International Journal of Engineering and Science (IJES),* vol. 7, no. 9 Ver.III , pp. PP 55-63, 2018.

[24] a. A. H. Malahat Ghoreishi, "Key enablers for deploying artificial intelligence for circular economy embracing sustainable product design: Three case studies," *AIP PUBLISHING,* 2020.

[25] "Development of design for remanufacturing guidelines to support sustainable manufacturing," *Robotics and Computer-Integrated Manufacturing,* vol. 23, 2007.

[26] W. M. C. H. G. N. S. Ijomah, "Development of design for remanufacturing guidelines to Support sustainable manufacturing," *Robotics and Computer Integrated Manufacturing,* vol. 23, pp. 712-719., 2007.

[27] E. C. G. Chierici, "Remanufacturing with Upgrade PSS for New Sustainable Business Models," *Procedia CIRP,* vol. 47, p. 531–536., 2016.

[28] O. M. D. T. N. Pialot, "How to explore scenarios of multiple upgrade cycles for sustainable product innovation: The "upgrade Cycle Explorer" tool.," *Journal of Cleaner Production,* vol. 22, no. 1, pp. 19-31, 2012.

[29] N. W. D. A. R. R. A. C. Aziz, "Modelling and optimisation of upgradability in the design of multiple life cycle products: A critical review," *Journal of Cleaner Production,* 2016.

[30] Y. U. Y. T. T. Shimomura, "A proposal of upgradable design," in *First International Symposium on Environmentally Conscious Design and Inverse Manufacturing*, 1999.

[31] S. Y. C. Smith, " Green product design through product modularization using atomic theory," *Robotics and Computer-Integrated Manufacturing,* vol. 26/6, p. 790–798, 2010.

[32] Y. H. P. Kristianto, "Reprint of "Product architecture modularity implications for operations economy of green supply chains."," Research Part E: Logistics and Transportation Review, 2015.

[33] G. M. Ren G., "Servitization in Manufacturing Companies," in *Cranfield Product-Service Systems Seminar*, 2007.

[34] E. K. E. M. L. H. Sundin, "Design for automatic end-of-life processes," *Assembly Automation*, vol. 32/4, pp. 389-398, 2012.

[35] A. B. M. C. S. K. G. S. J. D. O. B. S. T. Tullio Tollio, "Design, Management and Control of Demanufacturing and Remanufacturing Systems," *CIRP Annals Manufacturing Technology* , 2003.

[36] J. P. a. G. Chryssolouris, "A neural network approach for the development of modular product architectures," *International Journal of Computer Integrated Manufacturing*, 2011.

[37] E. K. M. S. M. S. T. S. Y. K. S. G. N. Nozaki, "Application of Artificial Intelligence in Product Design," *Fujitsu Scientific & Technical Journal*, 2017.

[38] A. B. A. J. N. S. G. L. C. Krahe, "Deep Learning for Automated Product Design," *CRIP Design*, 2020.

[39] J. P.-A. M. M. B. X. D. W.-F. S. O. A. C. Y. B. Ian J. Goodfellow, "Generative Adversarial Nets," 2014.

[40] N. N. E. Konno, "Application of Artificial Intelligence Technology in Product Design," *Fujitsu Scientific & Technical Journal*, July 2017.

[41] A. B. A. J. N. S. G. L. Carmen Krahe, "Deep Learning for Automated Product Design," *Elsevier*, 2020.

[42] "autodesk.com. Generative Design.," [Online]. Available: https://www.autodesk.com/solutions/generative-design.

[43] E. M. Foundation, "Artificial Intelligence and the Circular Economy: Ai As a Tool To Accelerate," *Report. Published online 2019:39.*

[44] A.-G. A. Sanchez-Lengeling B, "Inverse molecular design using machine learning:Generative models for matter engineering.," *Science*, 2018.

[45] V. Singh and N. Gu, "Towards an integrated generative design framework," *Des. Stud.*, vol. 33, pp. 185-207, 2012.

[46] J. M. a. M. Sandberg, "Architectural Design Exploration Using Generative Design: Framework Development and Case Study of a Residential Block," *buildings,* November 2020, Sweden.

[47] F. Shadram and J. Mukkavaara, "Exploring the effects of several energy efficiency measures on the embodied/operational energy trade-off: A case study of Swedish residential buildings.," *Energy Build,* vol. 183, pp. 283-296, 2019.

[48] S. Krish, "A practical generative design method," *Comput. Des.* , vol. 43, p. 88–100., 2011.

[49] S. K., "GM And Autodesk Develop 3D-Printed Car Parts.," 2018.

[50] a. A. H. Malahat Ghoreishi, "Key enablers for deploying artificial intelligence for circular economy embracing sustainable product design: Three case studies," *published online,* 5 May 2020.

[51] D. Denyer, D. Tranfield and J. Van Aken, "Developing design propositions through research synthesis," 2008.

[52] O. a. Pigneur, "Business Model development," 2002.

[53] i. software, "30 Best Manufacturing KPIs and Metric Examples for 2021 Reporting," 2 May 2021. [Online]. Available: https://insightsoftware.com/blog/30-manufacturing-kpis-and-metric-examples/.

[54] G. P. P. Y. e. a. Graham I, "Performance measurement and KPIs for remanufacturing," J Remanufacturing., 2015.

[55] C. D. E. H. F. A. H. S. L. L. N. S. V. J. Tolio T, "SPECIES-Co-evolution of products, processes and production systems," *CRIP Annals- Manufacturing Technology,* vol. 59 (2), pp. 672-693., 2010.

[56] C. D. Hauser JR, "The House of Quality," Harvard Business Review , May 1988.

[57] "What is House of Quality / QFD Example.," [Online]. Available: https://www.whatissixsigma.net/house-of-quality-qfd/..

[58] Roland Berger , "Frugal - SIMPLY THE BEST," June 2015.

[59] M. Sandberg, J. Mukkavaara, F. Shadram and T. Olofsson, "Multidisciplinary optimization of life-cycle energy and cost using a BIM-based master model.," *Sustainability* , vol. 11, no. 286, 2019.

[60] ICBR, "Electric Vehicle Batteries takeback compliance scheme: Stage & back-stage realities," GENEVA, September 2021.

[61] International Energy Agency (IEA), "Global EV Outlook 2019," 2019. [Online]. Available: https://www.iea.org/publications/reports/globalevoutlook2019/.

[62] N. P. L. G. R. J. M. P. ,. M. C. Elena Mossali, "Lithium-ion batteries towards circular economy: A literature review of opportunities and issues of recycling treatments," *Journal of Environmental Management,* 15 June 2020.

[63] W. K. B. A. D. A. Johnsen B, "Project Blue Battery, Part I: Analysis of Fire Risk Scenarios of Existing and Upcoming Large Maritime Battery System.," 2017. [Online]. Available: https://brandogsikring.dk/files/Pdf/FogU/Project%20BLUE%20BATTERY%20-%2.

[64] L. D. W. D. G. D. Das A, "Joining Technologies for Automotive Battery Systems Manufacturing," *World Electric Vehicle Journal,* vol. 9, 2019.

[65] Y. Y. T. A. Saw LH, "Integration issues of lithium-ion battery into electric vehicles battery pack.," *J Clean Prod* , pp. 113:1032-1045., 2016.

[66] K. A. Arora S, "Mechanical Design and Packaging of Battery Packs for Electric Vehicles. Behaviour of Lithium-Ion Batteries in Electric Vehicles.," *Springer,* pp. 175-200, 2018.

[67] M. T. I. K. K. H. S. S. N. T. Namiki F, "Lithium-ion Battery for HEVs, PHEVs, and EVs.," Hitachi Review, 2014; 63(2)..

[68] N. S. M. S. P. Natkunarajah, "Scenarios for the Return of Lithium-Ion Batteries Out of Electric Cars for Recycling," *22nd CIRP Conference on Life Cycle Engineering, published on Procedia CIRP,* vol. 29, pp. 740-745, 2015.

[69] A. o. E. A. a. I. B. Manufacturers, "Position paper on Performance and durability requirements in the Batteries Regulation," EUROBAT, 1st March 2021. [Online].

Available: https://www.eurobat.org/news-publications/position-papers/487-position-paper-on-performance-and-durability-requirements-in-the-batteries-regulation. [Accessed 27 April 2021].

[70] W. D. H. H. Go T, "Multiple generation life-cycles for product sustainability: the way forward.," *J Clean Prod* , vol. 95, pp. 16-29, 2015.

[71] "Green Car Reports- Audi details Battery for 2019 e-tron electric SUV," [Online]. Available: https://www.greencarreports.com/news/1116347_audi-details-battery-for-2019-e-tron-electric-suv. [Accessed September 2021].

[72] "INSIDEEVs," [Online]. Available: https://insideevs.com/news/340435/audi-e-tron-under-the-skin-battery-pack-motors-amp-more/. [Accessed September 2021].

[73] "Audi Mission Viejo," [Online]. Available: https://www.audimv.com/etron/tesla-vs-etron-battery-truth.htm. [Accessed September 2021].

[74] "How to disassemble a 2013 Nissan leaf battery pack.," 2015. [Online]. Available: https://www.youtube.com/watch?v=CPJ1O9Qu_Bg..

[75] "Disassembling a 2012 Nissan Leaf Battery Pack.," [Online]. Available: https://www.youtube.com/watch?v=0dBjPwJ_Qaw..

[76] "Electric Vehicle Database," [Online]. Available: https://ev-database.org/#sort:path~type~order=.rank~number~desc|range-slider-range:prev~next=0~1200|range-slider-acceleration:prev~next=2~23|range-slider-topspeed:prev~next=110~450|range-slider-battery:prev~next=10~200|range-slider-eff:prev~next=100~300|r.

[77] "TESLA MODEL S BATTERY TEARDOWN.," 2017. [Online]. Available: https://www.youtube.com/watch?v=NpSrHZnCi-A..

[78] "Pics/Info: Inside the battery pack.," 2015. [Online]. Available: https://teslamotorsclub.com/tmc/threads/pics-info-inside-the-battery-pack.34934/.

[79] Total Battery Consulting , "Battery Packs of Modern xEVs- A comprehensive engineering Assessment," 2019.

[80] "New Tesla P100D Battery Pack Conceptualized," 2016. [Online]. Available: https://insideevs.com/news/330819/new-tesla-p100d-battery-pack-conceptualized/.

[81] "Nissan Leaf teardown (Part 1).," 2012. [Online]. Available: Available at: https://www.marklines.com/en/report_all/rep1049_201202#report_area_2..

[82] European Environment Agency, "Earnings, jobs and innovation: the role of recycling in a green economy," 2011.

[83] Ecorys, "Study on the competitiveness of the EU eco-industry," Eurostat, 2009.

[84] EU Commission, "Memo: Advancing Manufacturing paves way for future of industry in Europe," Brusselles, 2014.

[85] Bayreuth University, Fraunhofer.

[86] U. M. C. G, "Innovative flexibility-oriented business models and system configuration approaches: An industrial application.," *CIRP Journal of Manufacturing Science and Technologies,* 2015.

A. Appendix A – Input data (tail)

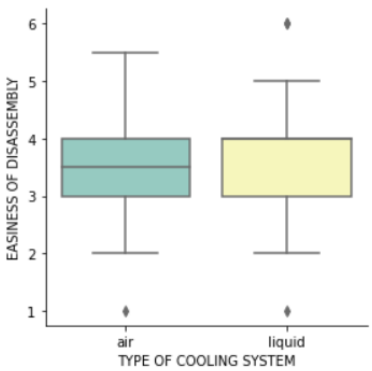
A	B	C	D	E	F	G	H	I
ID	Pack_Weight [kg]	MODULE_SIZE [mm3]	TYPE OF COOLING SYSTEM	CoolingSystem Manag	EXTERNAL_MJ_TYPE	NumbTypMechJoints Pack	NumbScrewsPack	MetalSheetFol ding
102	434	23846	liquid	ACTIVE	16	1	46	1
103	400	21978	liquid	PASSIVE	9	2	32	0
104	300	16484	air	ACTIVE	12	1	24	0
105	290	15934	air	ACTIVE	18	3	31	1
106	135	7418	liquid	ACTIVE	15	3	37	0
107	491	26978	air	ACTIVE	11	2	31	0
108	453	24890	liquid	ACTIVE	16	1	57	0
109	275	15110	liquid	ACTIVE	12	3	29	0
110	349	19176	air	ACTIVE	15	3	47	0
111	410	22528	air	ACTIVE	14	3	17	1
112	490	5806	liquid	ACTIVE	16	1	39	1
113	221	12143	liquid	ACTIVE	19	1	48	1
114	451	24780	liquid	ACTIVE	9	3	39	1
115	463	25440	air	ACTIVE	19	1	53	0
116	498	27363	air	ACTIVE	11	3	41	0
117	347	19066	air	ACTIVE	15	2	43	0
118	276	15165	liquid	ACTIVE	8	2	38	0
119	177	9725	liquid	ACTIVE	18	2	55	0
120	191	10495	air	ACTIVE	8	3	16	1
121	265	14560	air	ACTIVE	12	2	21	0
122	394	45070,2	liquid	ACTIVE	15	1	53	1
123	247	48723,12	liquid	ACTIVE	18	1	18	0
124	276	17408	liquid	ACTIVE	8	2	36	1
125	349	17408	liquid	ACTIVE	20	1	32	1

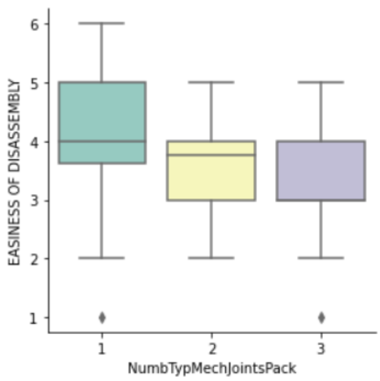
J	K	L	M	N	O	P	Q	R	S
cooling system-module_Interf	JB_MJ number screws	JB_ELECTRONIC connections	JB_ELECTRICAL connections	Module_capacity	Module_voltage	Type of cells	Electrical_Type_Cells	Cells_inModule	Number_PARALLEL_inModule
1	7	PLUG	BOLT	134	47	POUCH	ULTRASONIC WELD	34	4
1	9	PLUG	BOLT	154	15	PRISMATIC	MECHANICAL	16	4
0	8	PLUG	BOLT	147	43	PRISMATIC	LASER WELD	12	2
0	10	DIRECT CONNECTION	PLUG	177	32	POUCH	ULTRASONIC WELD	10	4
0	8	DIRECT CONNECTION	PLUG	221	10	POUCH	ULTRASONIC WELD	17	3
0	5	PLUG	BOLT	157	17	PRISMATIC	LASER WELD	36	1
1	5	PLUG	PLUG	167	82	PRISMATIC	MECHANICAL	23	1
0	7	DIRECT CONNECTION	BOLT	62	94	POUCH	ULTRASONIC WELD	26	1
0	5	DIRECT CONNECTION	BOLT	180	23	POUCH	ULTRASONIC WELD	7	4
0	9	DIRECT CONNECTION	PLUG	137	73	PRISMATIC	MECHANICAL	33	3
0	9	DIRECT CONNECTION	BOLT	80	20	POUCH	ULTRASONIC WELD	5	2
0	9	PLUG	PLUG	208	83	PRISMATIC	LASER WELD	39	3
1	5	PLUG	BOLT	180	37	PRISMATIC	LASER WELD	7	4
0	5	PLUG	PLUG	88	59	PRISMATIC	MECHANICAL	28	4
0	8	PLUG	PLUG	51	86	PRISMATIC	MECHANICAL	4	2
0	6	PLUG	PLUG	99	53	PRISMATIC	LASER WELD	7	3
1	10	DIRECT CONNECTION	PLUG	202	32	PRISMATIC	LASER WELD	25	3
1	6	PLUG	BOLT	33	22	POUCH	ULTRASONIC WELD	9	1
0	4	DIRECT CONNECTION	PLUG	154	82	POUCH	ULTRASONIC WELD	40	4
0	6	DIRECT CONNECTION	BOLT	39	90	POUCH	ULTRASONIC WELD	22	4
0	6	DIRECT CONNECTION	BOLT	220	85	CYLINDRIC	ULTRASONIC WEDGE BOI	1058	150
0	6	DIRECT CONNECTION	BOLT	220	92	CYLINDRIC	ULTRASONIC WEDGE BOI	1150	123
1	7	DIRECT CONNECTION	BOLT	230	22,8	CYLINDRIC	LASER WELD	444	105
1	8	DIRECT CONNECTION	BOLT	230	22,8	CYLINDRIC	RESISTANCE SPOT	1000	100

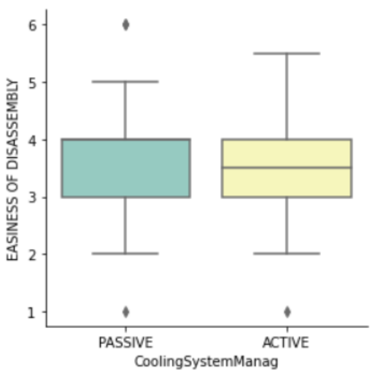
T	U	V	W	X	Y	Z	AA	AB
NUMBER OF MODULES	MJ_Module	mod-mod_interf	BMS TOPOLOGY	Type_Elettrical_Connection_modules	Adhesives	Glue	number of sensors per module	EASINESS OF DISASSEMBLY
5	6	1	DISTRIBUTED	BOLT	1	3	4	3
7	10	1	CENTRALIZED	BOLT	0	2	7	4
8	6	0	CENTRALIZED	PLUG	0	0	10	5,5
6	5	1	DISTRIBUTED	PLUG	0	5	8	2
10	9	0	CENTRALIZED	BOLT	1	1	10	3
8	6	0	DISTRIBUTED	BOLT	0	2	3	4
12	10	1	DISTRIBUTED	PLUG	0	5	7	4,5
9	7	1	DISTRIBUTED	BOLT	1	0	5	2,5
9	8	1	CENTRALIZED	PLUG	1	4	6	3
4	8	1	CENTRALIZED	PLUG	1	5	8	3
12	9	1	CENTRALIZED	BOLT	1	1	6	4
6	9	1	DISTRIBUTED	PLUG	1	1	10	3,5
9	6	0	CENTRALIZED	BOLT	1	2	2	4
8	7	0	DISTRIBUTED	BOLT	1	1	5	5,5
4	6	0	CENTRALIZED	PLUG	0	1	10	5
6	7	0	DISTRIBUTED	BOLT	1	5	10	4,5
7	5	1	CENTRALIZED	PLUG	1	0	8	4,5
7	5	0	CENTRALIZED	BOLT	0	1	4	4
7	7	1	DISTRIBUTED	PLUG	0	0	10	3
4	6	0	DISTRIBUTED	BOLT	0	1	3	3,5
2	4	1	DISTRIBUTED	BOLT	1	2	8	2,5
2	10	1	DISTRIBUTED	BOLT	1	2	5	2,5
16	10	1	DISTRIBUTED	BOLT	1	3	10	3
16	10	1	DISTRIBUTED	BOLT	1	2	5	2

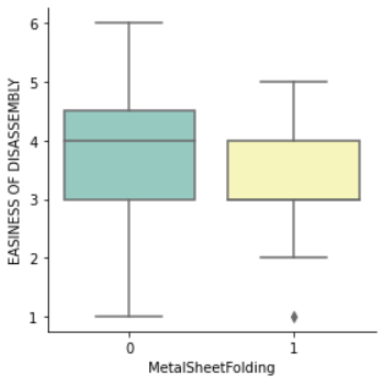
B. Appendix B

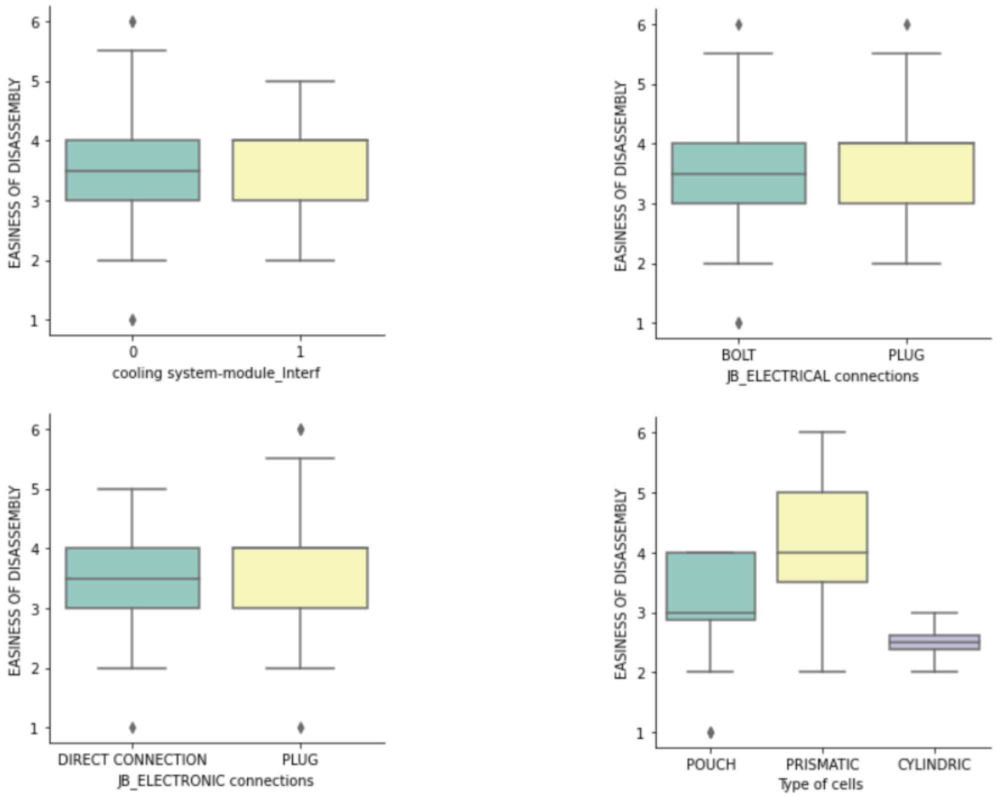
B.1 Box plots of data distributions depending on values assumed by each specific categorical variable

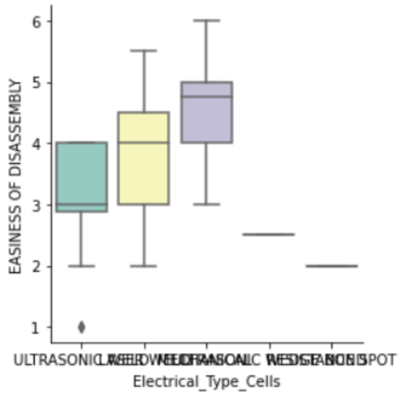


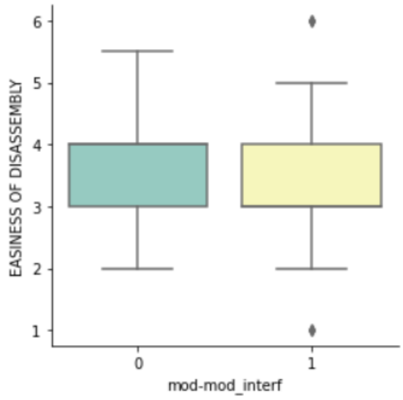


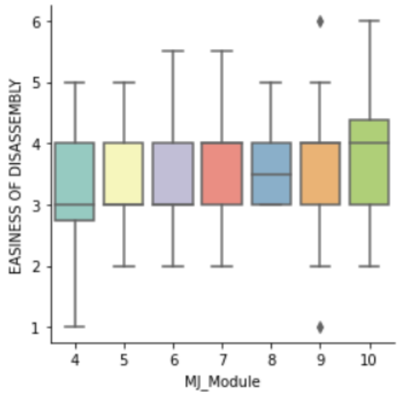


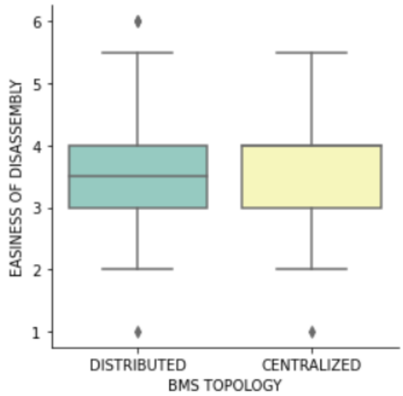


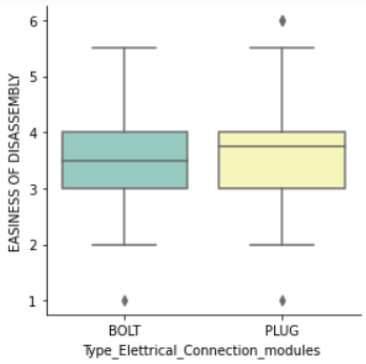

EASINESS OF DISASSEMBLY
0
1
cooling system-module_Interf
EASINESS OF DISASSEMBLY
BOLT
PLUG
JB_ELECTRICAL connections
EASINESS OF DISASSEMBLY
DIRECT CONNECTION
PLUG
JB_ELECTRONIC connections
EASINESS OF DISASSEMBLY
POUCH
PRISMATIC
CYLINDRIC
Type of cells

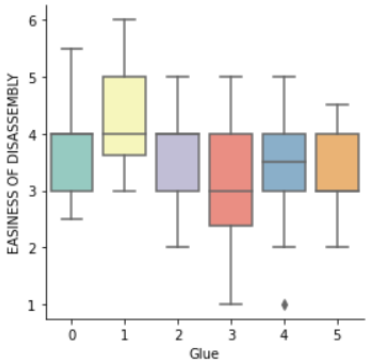

EASINESS OF DISASSEMBLY
1
2
3
4
5
6
Electrical_Type_Cells

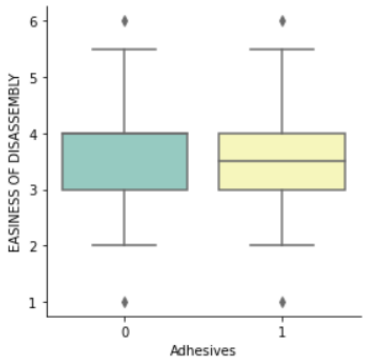

EASINESS OF DISASSEMBLY
1
2
3
4
5
6
0
1
mod-mod_interf

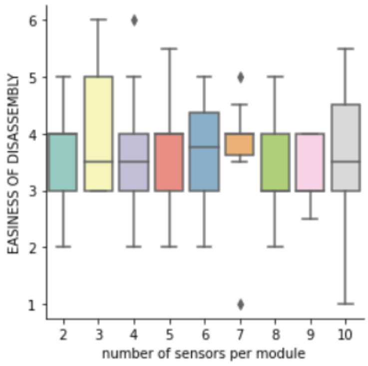

EASINESS OF DISASSEMBLY
1
2
3
4
5
6
4
5
6
7
8
9
10
MJ_Module


EASINESS OF DISASSEMBLY
1
2
3
4
5
6
DISTRIBUTED
CENTRALIZED
BMS TOPOLOGY


EASINESS OF DISASSEMBLY
1
2
3
4
5
6
BOLT
PLUG
Type_Elettrical_Connection_modules


EASINESS OF DISASSEMBLY
1
2
3
4
5
6
0
1
2
3
4
5
Glue


EASINESS OF DISASSEMBLY
1
2
3
4
5
6
0
1
Adhesives


EASINESS OF DISASSEMBLY
1
2
3
4
5
6
2
3
4
5
6
7
8
9
10
number of sensors per module

B.2 GridSearch Outputs: Trained Algorithms Performances

```
***GRIDSEARCH RESULTS***
Best score: -0.550333 using {'alpha': 0.01, 'normalize': False}

MAE  train 0.361    test 00.452
MSE  train 0.215 test 0.333
RMSE train 0.463   test 0.577
r2   train 0.763     test 0.673
```

Figure 53- Lasso Regression

```
***GRIDSEARCH RESULTS***
Best score: -0.856606 using {}

MAE  train 0.307    test 01.088
MSE  train 0.149 test 6.470
RMSE train 0.386   test 2.544
r2   train 0.836     test -5.346
```

Figure 54- Linear Regression

```
***GRIDSEARCH RESULTS***
Best score: -0.579721 using {'alpha': 10, 'normalize': False}

MAE  train 0.421    test 00.494
MSE  train 0.289 test 0.375
RMSE train 0.538   test 0.612
r2   train 0.681     test 0.632
```

Figure 55- Ridge Regression

```
***GRIDSEARCH RESULTS***
Best score: -0.642816 using {'n_neighbors': 20, 'p': 1}

MAE  train 0.612    test 00.727
MSE  train 0.596 test 0.756
RMSE train 0.772   test 0.870
r2   train 0.343     test 0.258
```

Figure 56- K-Nearest Neighbors Regression

```
***GRIDSEARCH RESULTS***
Best score: -0.628007 using {'max_depth': 1, 'min_samples_leaf': 5}

MAE  train 0.599    test 00.758
MSE  train 0.608 test 0.811
RMSE train 0.779   test 0.901
r2   train 0.331     test 0.204
```

Figure 57- Decision Trees

```
***GRIDSEARCH RESULTS***
Best score: -0.619427 using {'criterion': 'mse', 'min_samples_leaf': 10, 'n_estimators': 100, 'random_state': 42}

MAE  train 0.539    test 00.661
MSE  train 0.437 test 0.628
RMSE train 0.661   test 0.793
r2   train 0.519     test 0.384
```

Figure 58- Random Forest

```
***GRIDSEARCH RESULTS***
Best score: -0.592205 using {'C': 0.1, 'degree': 2, 'epsilon': 0.01, 'gamma': 'auto', 'kernel': 'linear'}

MAE  train 0.366    test 00.446
MSE  train 0.295 test 0.320
RMSE train 0.543   test 0.565
r2   train 0.675     test 0.686
```

Figure 59- SVM Regressor

```
***GRIDSEARCH RESULTS***
Best score: -0.652356 using {'alpha': 0.1, 'batch_size': 20, 'hidden_layer_sizes': (10, 5), 'learning_rate': 'constant', 'max_i
ter': 1000, 'solver': 'sgd'}

MAE  train 0.115    test 00.776
MSE  train 0.031 test 1.764
RMSE train 0.175   test 1.328
r2   train 0.966     test -0.730
```

Figure 60 - Multi-Layer Perceptron

List of Figures

List of Tables

List of Graphs