Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 128 (2024) 351–356

www.elsevier.com/locate/procedia

34th CIRP Design Conference

A parametric environmental impact model for manufacturing components based on machine learning techniques

Luca Manuguerra[a]*, Federica Cappelletti[a], Marta Rossi[b], Marco Mandolini[a], Michele Germani[a]

[a]*Università Politecnica delle Marche, Via Brecce Bianche 12, 60131 Ancona, Italy*
[b]*Uniecampus, Via Isimbardi, 10, 22060 Novedrate (CO), Italy*

* Corresponding author. Tel.: 071 2204402. *E-mail address:* l.manuguerra@staff.univpm.it

Abstract

Environmental sustainability-oriented design is becoming increasingly important in the industrial field partly because of the effects of climate change. Sustainable development-oriented choices are most effective at the early design stage. The design team must be able to assess approximately and quickly the environmental impact early in the design phase. From these motivations comes the need for a method that quickly and with few parameters can estimate the product environmental impact during the conceptual design phase. Machine learning techniques appear to be well suited to meet this challenge. Machine learning is an established research topic in Industry 4.0 and its adoption is increasing. The integration of machine learning within conceptual design quickly facilitates the approximate assessment of environmental impact through high-level data. In this paper, a method is proposed to obtain a parametric model for the environmental impact assessment of manufacturing components at the early design stage. It allows consistent considerations concerning environmental matters, albeit little information available during design phase.

Keywords: Machine Learning, Data Science, Sustainability, Design, Life Cycle Assessment

1. Introduction

The environmental effects of a product's life cycle must be considered when designing a product due to the damage caused by climate change [1]. The design team must be able to evaluate the environmental performance of numerous project proposals from the early design stages [2]. Furthermore, it must already be able to evaluate the different scenarios of the production chain of the product's life cycle. Life Cycle Assessment (LCA) tools allow very detailed analyses of the product environmental impact, without however giving the designer alternatives [3].The solution to the opportunity to make less impactful products from an environmental point of view is left to the designer's and engineer's skills and ability. Based on the available data and the objective of the analysis, the system boundaries can be chosen. The boundaries of the system can include the phases of raw material extraction, production, distribution, use and final disposal of the product. The reference standards are ISO 14040 and ISO 14044. The practical application of eco-design and circular economy approaches represents an opportunity to be seized in the industrial sector for the reduction of environmental impacts and the creation of economic value [4].

Producing products with less and less impact and placed within a circular process is now necessary for multiple purposes: from an environmental perspective (i.e., aware consumers), economic (i.e., competitiveness) and geo-political point of view, to reduce dependence in the supply of increasingly scarce raw materials.

10.1016/j.procir.2024.04.009

1.1. Parametric LCA for Conceptual design

Conceptual design creates new challenges for environmental assessment. Little detailed information is available, decisions and trade-offs between multiple attributes must be made quickly. Borland and Wallace [5] illustrated how the capabilities of parametric LCA models developed by environmental experts can be integrated with traditional design models and made available on request. However, the use of detailed parametric models is still of limited value for initial conceptual design, due to the amount of time and information required to develop parametric LCA [6] models. As a result, many tool developers are focusing on the early design phases [7]. Qualitative or quantitative methods, such as lists of control [8], qualitative matrices [9], abbreviated LCA [10], and LCA simplification [11], represent attempts to simplify and reduce the number of resources needed for LCA modelling. The usefulness of these methods is undeniable, but they are not suitable for initial conceptual design.

1.2. Machine learning in Life Cycle Assessment

The lack of analytical methods for integrated initial conceptual design motivated the development of a surrogate learning LCA concept to be applied to preliminary assessments of the life cycle, based on Machine Learning (ML) techniques. ML presents a complex challenge linked to the enormous amount of input factors and related uncertainties that influence the entire life cycle. ML techniques belong to a class of Artificial Intelligence (AI) techniques that can learn from data to increase their accuracy without reprogramming. Through the analysis of a training database, an algorithm is generated without user assistance, ML creates a model that can be queried [12]. ML algorithms are generally classified into four groups: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning [13]. ML can be used to fill in missing data in the Life Cycle Inventory (LCI) phase for LCA [14]. An ML model can in real time evaluate how production or process changes impact and provide, based on specific constraints, potential solutions for less environmentally harmful production. ML offers suggestions for an optimization process. Feature selection techniques in ML allow to identify the most important parameters and focus on their collection. In this way, it is possible to reduce the input parameters, favouring the speed compilation, streamlining the data collection phase, making it faster and promoting a better understanding of the model. Unlike full LCA, this makes it particularly useful in the design process. The literature offers many examples of ML applications for emissions prediction. Antanasijevic et al. [15] developed a model for predicting greenhouse gas emissions from electric vehicles used in several European countries using an Artificial Neural Network (ANN) approach with sustainability, economic and industrial indicators used as inputs. Sousa et al. [2] created a surrogate LCA model based on ML techniques to evaluate the environmental impact of energy-intensive electrical products. Wistoff et al. [16] aim to measure the overall environmental impact potential of design choices in a consumer product. This is achieved by employing a multi-layer perceptron neural network with backpropagation training, a ML technique. This method establishes a relationship between the LCA impact of 37 case study products and the attributes of the products.

1.3. Aim and scope

The method proposed in this work aims to create a predictive model that can be used in the initial design phase to analyse the environmental impact; the chosen category is Climate Change which allows to verify the kgCO2eq produced in manufacturing a product. Finally, the method is tested to verify its applicability.

2. Material and Method

The CRISP-DM (CRoss Industry Standard Process for Data Mining) method is a methodology for data mining [17] and was used as the approach for this study. The six phases are presented in the following sections (Fig. 1): I. Business understanding; II. Data understanding; III Data preparation; IV. Modeling; V. Evaluation; VI. Deployment. The primary objective of this paper is to confirm that ML techniques have made the construction of an environmental impact model specifically tailored for axisymmetric components feasible, aligning with the prerequisites of the initial phases.

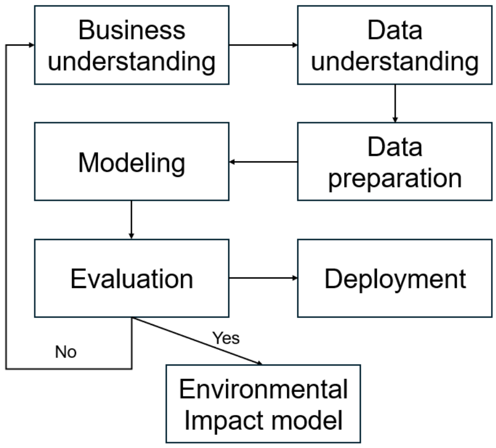

Fig. 1. Method

2.1. Business understanding

In this phase, the objectives and required needs are defined.

First, definition of the problem to study. Problem definition focuses on project requirements according to environmental objectives. This phase aims to identify lifecycle management and product design opportunities. This methodology is strongly influenced by an LCA analysis, therefore recovering the same initial inputs, such as defining the objective, in this case evaluating the environmental impact of the creation of a manufacturing product; the functional unit that is formulated as *producing a mechanical axisymmetric component*; the system boundaries, therefore the phases that to consider; the calculation method, LCIA Method, Database, Impact Category. The choice of impact categories will determine how many dependent parameters (outputs) must be predicted.

Secondly, define what accuracy is acceptable for the results that will be obtained from the model. In the case of an LCA analysis, an error of 40% can be encouraging [18].

2.2. Data understanding

Data understanding entails the identification, collection, analysis, and validation of datasets to fulfill project objectives, a concept applied to the examination of mechanical components. Specifically, this stage involves three activities. The initial activity involves identifying all components within the designated product family.

The second task requires carrying out a complete LCA analysis of at least one component chosen as a reference or a set of representative components of the entire database. To reduce the time of this phase, LCA analyses previously carried out on components could be used, it is not necessary to do new ones, or LCA analyses from the literature can be used, when representative of the component. This preliminary analysis is used to understand the life cycle of the component, the phases that will be considered, and therefore the information and parameters that will be collected. Furthermore, it also allows to understand what the critical issues and limitations of the analysis are. This step allows to write down all the hypotheses that have been made to solve problems and limitations. The model will therefore be trained on data that have been deemed reliable and of which the limits and critical issues are known.

The third task concerns the collection of data and information, which is divided into geometric and non-geometric information:

- Geometric information is collected from 3D CAD models or 2D drawings. The parameters chosen must be common and capable of describing all the parts analyzed.
- Non-geometric information concerns, for example, the manufacturing processes to obtain the semi-finished product (e.g. forging, casting) and where the material, manufacturing processes to obtain the finished product (e.g. milling, drilling, turning) take place. Manufacturing cycles, with which it is possible to identify manufacturing processes.

These activities provide all the data that can be grouped and sorted in a first database. All associated parameters describe each part, expressed separately as independent parameters.

2.3. Data preparation

Preparing data for modeling means building a large and structured data database to be used to obtain algorithms that predict an acceptable error. The data preparation phase can be divided into three different activities, at the end of which, the database will be ready for the modeling phase.

The initial step involves outlining all conceivable geometric and non-geometric parameters that could influence the impact assessment, some of which were detailed in the data understanding phase.

The subsequent task involves broadening the range of independent parameter values. The precision and reliability of outcomes generated by the predictive algorithm of the parametric model hinge on the characteristics of the database employed for model training. A larger volume of records translates to more effective training and introducing variations in values within the same parameter aids in mitigating the risk of overfitting. This activity is aimed at expanding the number of records. The first aspect is to consider all possible values for certain parameters. Once determined, the database is extended considering all possible combinations. The extension can concern both geometric and non-geometric parameters.

The third activity consists in calculating the environmental impact for each record and therefore for each impact category. In this case, environmental impact analysis software tools (e.g. SimaPro, GaBi, OpenLca) are used.

2.4. Modeling

Modeling is the phase in which diverse models are built and assessed using various techniques. Regression models and supervised algorithms are used. The objective is to identify the optimal ML algorithm for predicting the dependent parameter. The initial step involves selecting the algorithm to be tested. During this stage, the various models generated using different algorithms are juxtaposed. Subsequently, models are generated in the second step. The third step involves evaluating the obtained model using performance indicators. The results obtained in this phase inform the choice of the best algorithm. These activities are iterated for each algorithm under consideration. By directly comparing performance indicators, the algorithm with the least error is selected.

2.5. Evaluation

This task assesses whether the accuracy of the model meets the criteria established in the Business Understanding phase.

2.6. Deployment

During the deployment process, users can manage the parametric model and evaluate the impact of each independent parameter on the environmental impact. This process embodies model interpretation, emphasizing the goal of providing a comprehensible model rather than a black box. A comprehensive exploration of feature importance is undertaken, enabling users to grasp the significance of each independent parameter and its influence on the environmental impact. Various methods can be employed to ascertain the significance of each feature.

3. Case study

This section will go over the method by applying it to the case study, namely manufacturing axisymmetric parts.

Business Understanding. The objective of the proposed model aims to calculate the environmental impact for the Climate Change category [kgCO2eq] to manufacturing axisymmetric parts. The functional unit is to produce a mechanical axisymmetric manufacturing part. The system boundaries consider the material, pre-production, transport and manufacturing phases (cradle to gate). The pre-production is the phase where the semifinished part is produced. The manufacturing phase is the phase where the part is finished and refined. The Calculation method: ReCiPe Midpoint (H) V1.12 / Europe Recipe H, Ecoinvent 3 – allocation, default, Climate Change [kg CO2 eq]. The allowable estimate error is 40%.

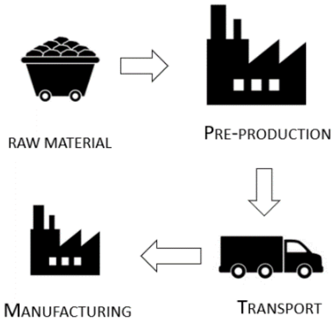

Fig. 2. System boundaries.

Data Understanding. For this study, 73 different models were collected (e.g., shafts, discs, pins, flanges, and washers). Axisymmetric components are parts with rotational symmetry. Two comprehensive LCA analyses were conducted on two components chosen as representative of the database (Fig. 3), through LCA software SimaPro.

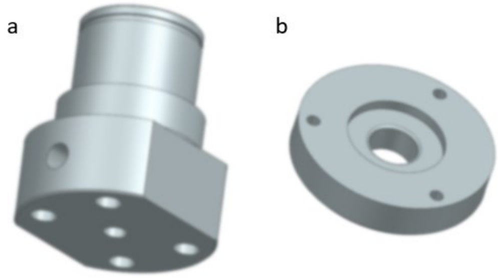

Fig. 3. Axisymmetric representative products: pin (a), disc (b).

The pin (Fig. 3 a) was made of chromium steel and it was made through casting processes in China, then transported to Italy where it underwent mainly turning operations, and then drilling operations. Transportation from China to Italy was by ship. The disc (Fig. 3 b) was made of low-alloyed steel and it was made through forging processes in Italy, then transported to the U.S. where it underwent mainly milling, and then drilling processes. Transportation from Italy to the U.S. was by air.

Fig. 4 shows the results of the comprehensive LCA analysis that was conducted on these two components. From this analysis, is evident that all the considered phases are not negligible and the impact and importance of the phase may vary depending on the input data.

Data Preparation. Input parameters are divided into geometric and non-geometric. The geometric ones are:

- Dimensions D1, D2, [mm] are the max radial extensions;
- Dimension D3 [mm] is the max axial length;
- D_Inner [mm] is internal diameter;
- F_Type (Hollow/Solid) indicates whether the finite is hollow or solid;
- F_Mass [kg] is the finished mass;
- Ring_Seat, Keys_Seat (Yes/No) indicates whether there is this specific machining on the component;
- N_Thread (Yes/No) indicates the number of this specific machining on the component;
- S_Type (Round/Sheetmetal/Hexagon/RoundTube) indicates the type of the starting raw.

The non-geometric are:

- Material (Chromium steel/ Steel low alloyed/Steel unalloyed);
- Turning, Milling, Drilling (Yes/No) indicates whether there is this specific machining on the component;
- Preproduction (Casting/Forging) indicates the type of manufacturing process of the semi-finished product;
- Country_Preproduction (Italy/China/USA) and Country_Machining (Italy/USA) indicate the country where the productive plant is present.

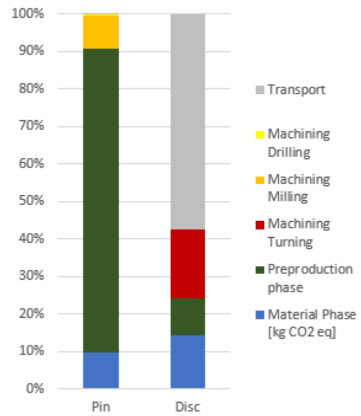

Fig. 4. Environmental impact assessment

For database extension, it was decided as a first task to geometrically scale the 3D CAD models. Scaling means enlarging or decreasing parts in all directions according to scaling factors. From 73 parts, 219 parts were obtained. For database extension, non-geometric parameters were also used:

- the material (3), considering that each part can be made of Chromium steel, Steel low alloyed and Steel unalloyed;
- the preproduction process (2), considering that each part can be made by casting or forging;
- the nation where the preproduction process takes place (3), considering Italy, China and the United States;
- the nation where the machining process takes place (2), considering Italy and the United States.

In this way, the final database from 219 records was extended to 7884. Transportation was calculated accordingly and considering the distances and types shown in Table 1. Table 2 shows database statistical information.

Modeling. Modeling is the phase during which numerous models are generated and evaluated using various approaches. The primary goal is to identify the most effective machine learning methods for predicting specific environmental impact factors, with one dedicated to each parametric impact model.

To achieve this, RapidMiner, a data science platform developed by Altair Engineering, was used. RapidMiner offers a diverse range of machine-learning techniques.

Table 1. Transport information

From	To	Distance [km]	Type
Italy	Italy	500	Lorry
Italy	USA	6500	Air
China	Italy	9500	Ship
China	USA	13000	Air
USA	Italy	6500	Air
USA	USA	750	Lorry

The process was facilitated by the Auto Model plugin within RapidMiner Studio, which simplifies the generation and validation of models [19]. In RapidMiner there are several ML techniques, including Gradient Boosting (GB) [20], Random Forest (RF) [21], Deep Learning (DL) [22], Neural Network (NN) [23], and Linear Regression (LR) [24]. The process of constructing an environmental impact model commences by partitioning the database into training and testing sets. Software tools designed for data mining or data sciences are essential for the modeling process.

Table 2. Database statistical information

Parameter	Max	Min	Mean	Standard deviation
D1 [mm]	5.00E+02	4.00E+00	1.19E+02	1.04E+02
D2 [mm]	5.00E+02	4.00E+00	1.21E+02	1.05E+02
D3 [mm]	4.14E+03	6.25E-01	1.85E+02	4.24E+02
D_Inner [mm]	3.04E+02	0.00E+00	4.10E+01	5.38E+01
F_Mass [kg]	8.89E+03	4.70E-03	2.39E+02	8.75E+02
N_Thread	1.20E+01	0.00E+00	1.95E-01	1.08E+00

Ultimately, each algorithm produces a predictive model. A 40% hold-out set is employed for performance evaluation. Initially, the software utilizes this hold-out set as its input. Subsequently, employing a multi-hold-out-set validation approach, the software assesses performances across seven distinct subsets. The most robust and top-performing results are excluded. The average of the remaining five performances is then computed. Table 3 shows the results of the modeling process for the Relative Error [19]. The authors opt for the Relative Error because it is straightforward for engineers to comprehend and has a direct connection to the initial design phase.

Table 3. Modeling result

Model	Relative Error
Generalized Linear Model	83.4%
Deep Learning	79.0%
Decision Tree	21.7%
Random Forest	71.7%
Gradient Boosted Trees	43.5%

Evaluation. The results show that only the decision trees respect the stability constraints for the relative error during the Business Understanding phase. RapidMiner automatically tuned the hyperparameter to achieve the following value (Optimise Parameter Quadratic Operator): maximal depth: 25.

Deployment. RapidMiner allows to carry out a Feature Importance analysis to understand how parameters influence the model. Fig. 5 shows that F_Mass is the most important parameter. The models were deployed and stored in a database to be used for design by engineers through RapidMiner's simulator module.

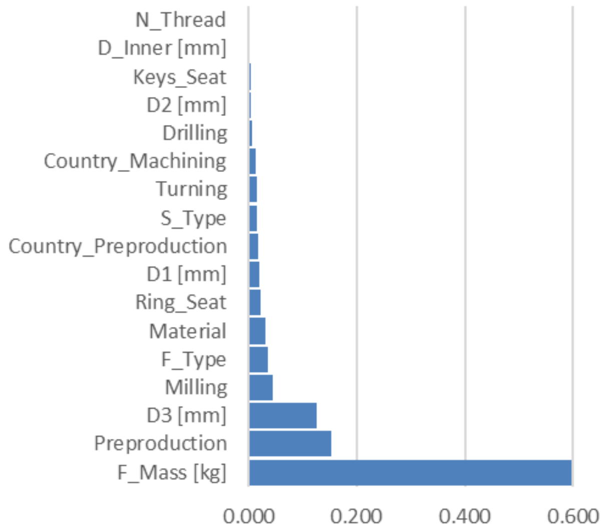

Fig. 5 Feature Importance

4. Results

The ML model allows for predicting an axisymmetric component impact assessment with a relative error of 21.7%. The error obtained is aligned with what the business understanding requires (40%).

5. Discussion

The proposed approach is not restricted to any software; it can be implemented using various data science tools and other analytical disciplines for environmental impact estimation. Moreover, each parameter chosen for predictive models is identifiable during the early stages of design. The advantage is due to the simplicity of using a prediction model and the speed with which results are obtained. The critical disadvantage of this research is that it provides only one overall estimate of the environmental impact without understanding which of the various phases has the greatest impact. For the next steps, it could be to create a model for each phase of the life cycle, in this way in addition to having greater traceability of the impact, the influence of each parameter in each phase will be managed, and a better result will be achieved also in terms of accuracy. The result obtained from this model does not replace a complete LCA analysis conducted at the end of the design. The proposed method allows the creation of a prediction model of environmental impacts with an acceptable error, to be used in the initial design phase. The model is proposed as a decision-making tool for the designer to guide him/her from the first initial steps in a conscious design oriented towards environmental sustainability. The rapid modification of both geometric and non-geometric parameters allows ecodesign modeling to be carried out, albeit at a high level, but consistent with the little information available in the initial design phase. The high generalizability of the geometric application of the model decreases the accuracy of the result. To obtain better results in terms of accuracy, it is possible to intervene in different ways, adding a parameter that identifies the family of the part or creating a specific model for each family. A model of general applicability is useful when designing components that cannot be assimilated into any family.

6. Conclusion

The research outlined a procedure for developing parametric models for manufactured parts during the conceptual or early design phases. This approach relies on machine learning algorithms trained with databases generated by a dedicated software tool designed for estimating environmental impact. Through this methodology, a model specific to axisymmetric manufactured parts was successfully established, and the strategy facilitated the selection of the most suitable machine learning algorithm for addressing this particular challenge. Future investigations should assess the method's applicability to other product items by subjecting it to testing across diverse geographical areas and production processes.

7. References

[1] F. Cappelletti, M. Rossi, and M. Germani, "How de-manufacturing supports circular economy linking design and EoL - a literature review," J Manuf Syst, vol. 63, pp. 118–133, Apr. 2022, doi: 10.1016/j.jmsy.2022.03.007.

[2] I. Sousa and D. Wallace, "Product classification to support approximate life-cycle assessment of design concepts," Technol Forecast Soc Change, vol. 73, no. 3, pp. 228–249, Mar. 2006, doi: 10.1016/j.techfore.2004.03.007.

[3] R. Marta, C. Federica, and G. Michele, "A Step Forward Life Cycle Assessment to Optimize Products and Increase Company Eco-design Competencies," 2023, pp. 62–74. doi: 10.1007/978-3-031-15928-2_6.

[4] H. K. Lamba, N. S. Kumar, and S. Dhir, "Circular economy and sustainable development: a review and research agenda," International Journal of Productivity and Performance Management, Jan. 2023, doi: 10.1108/IJPPM-06-2022-0314.

[5] N. Borland and D. Wallace, "Environmentally Conscious Product Design: A Collaborative Internet-based Modeling Approach," J Ind Ecol, vol. 3, no. 2–3, pp. 33–46, Apr. 1999, doi: 10.1162/108819899569539.

[6] I. Sousa, D. Wallace, and J. L. Eisenhard, "Approximate Life-Cycle Assessment of Product Concepts Using Learning Systems," J Ind Ecol, vol. 4, no. 4, pp. 61–81, Sep. 2000, doi: 10.1162/10881980052541954.

[7] H. Baumann, F. Boons, and A. Bragd, "Mapping the green product development field: engineering, policy and business perspectives," J Clean Prod, vol. 10, no. 5, pp. 409–425, Oct. 2002, doi: 10.1016/S0959-6526(02)00015-X.

[8] J. Fiksel, "Achieving eco-efficiency through design for environment," Environmental Quality Management, vol. 5, no. 4, pp. 47–54, Jun. 1996, doi: 10.1002/tqem.3310050407.

[9] B. Allenby, "The ontologies of industrial ecology?," Progress in Industrial Ecology, An International Journal, vol. 3, no. 1/2, p. 28, 2006, doi: 10.1504/PIE.2006.010039.

[10] T. E. GRAEDEL, B. R. ALLENBY, and P. R. COMRIE, "Matrix Approaches to Abridged Life Cycle Assessment," Environ Sci Technol, vol. 29, no. 3, pp. 134A-139A, Mar. 1995, doi: 10.1021/es00003a751.

[11] K. G. Mueller and C. B. Besant, "Streamlining life cycle analysis: a method," in Proceedings First International Symposium on Environmentally Conscious Design and Inverse Manufacturing, IEEE, 1999, pp. 114–119. doi: 10.1109/ECODIM.1999.747593.

[12] Tom M. Mitchell, Machine Learning. 1997.

[13] Batta Mahesh, "Machine Learning Algorithms - A Review," International Journal of Science and Research (IJSR), 2018.

[14] A. Ghoroghi, Y. Rezgui, I. Petri, and T. Beach, "Advances in application of machine learning to life cycle assessment: a literature review," Int J Life Cycle Assess, vol. 27, no. 3, pp. 433–456, Mar. 2022, doi: 10.1007/s11367-022-02030-3.

[15] D. Z. Antanasijević, M. Đ. Ristić, A. A. Perić-Grujić, and V. V. Pocajt, "Forecasting GHG emissions using an optimized artificial neural network model based on correlation and principal component analysis," International Journal of Greenhouse Gas Control, vol. 20, pp. 244–253, Jan. 2014, doi: 10.1016/j.ijggc.2013.11.011.

[16] A. Wisthoff, V. Ferrero, T. Huynh, and B. DuPont, "Quantifying the Impact of Sustainable Product Design Decisions in the Early Design Phase Through Machine Learning," in Volume 4: 21st Design for Manufacturing and the Life Cycle Conference; 10th International Conference on Micro- and Nanosystems, American Society of Mechanical Engineers, Aug. 2016. doi: 10.1115/DETC2016-59586.

[17] J. S. Saltz, "CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps," in 2021 IEEE International Conference on Big Data (Big Data), IEEE, Dec. 2021, pp. 2337–2344. doi: 10.1109/BigData52589.2021.9671634.

[18] M. Algren, W. Fisher, and A. E. Landis, "Machine learning in life cycle assessment," in Data Science Applied to Sustainability Analysis, Elsevier, 2021, pp. 167–190. doi: 10.1016/B978-0-12-817976-5.00009-7.

[19] RapidMiner Documentation, "https://docs.rapidminer.com/9.3/studio/auto-model/."

[20] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," Front Neurorobot, vol. 7, 2013, doi: 10.3389/fnbot.2013.00021.

[21] Y. Liu, Y. Wang, and J. Zhang, "New Machine Learning Algorithm: Random Forest," 2012, pp. 246–252. doi: 10.1007/978-3-642-34062-8_32.

[22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[23] C. M. Bishop, "Neural networks and their applications," Review of Scientific Instruments, vol. 65, no. 6, pp. 1803–1832, Jun. 1994, doi: 10.1063/1.1144830.

[24] X. Su, X. Yan, and C.-L. Tsai, "Linear regression," Wiley Interdiscip Rev Comput Stat, vol. 4, no. 3, pp. 275–294, May 2012, doi: 10.1002/wics.1198.