MACHINE LEARNING FOR PARAMETRIC COST ESTIMATION OF AXISYMMETRIC COMPONENTS

Manuguerra, Luca;
Mandolini, Marco;
Germani, Michele;
Sartini, Mikhailo

UNIVPM Università Politecnica delle Marche

ABSTRACT

Machine learning (ML) is a well-established research topic in Industry 4.0 is boosting its adoption. ML is also used for manufacturing cost estimation during design. Such approaches are commonly used to estimate the cost of mass-produced parts. Many consolidated historical data are available for training the regression models. Unfortunately, very often, such a database of data is not available.

The paper defines an ML approach for parametric cost estimation of axisymmetric components. The data for training the ML model derives from automatic software for analytically estimating the manufacturing cost. With a proper set of simulations, the tool can generate a large amount of data for training. The paper presents the steps for developing a parametric cost model using ML. The approach is based on CRoss Industry Standard Process for Data Mining method. The proposed method was used to develop one cost model (to estimate the total cost that considered raw material and manufacturing cost). The obtained Relative Error is 23.52% ± 1.37%, coherent with E2516 − 11, Standard Classification for Cost Estimate Classification System.

Keywords: Design costing, Machine learning, Conceptual design, Big data

Contact:
Manuguerra, Luca
UNIVPM Università Politecnica delle Marche
Italy
l.manuguerra@pm.univpm.it

Cite this article: Manuguerra, L., Mandolini, M., Germani, M., Sartini, M. (2023) 'Machine Learning for Parametric Cost Estimation of Axisymmetric Components', in *Proceedings of the International Conference on Engineering Design (ICED23),* Bordeaux, France, 24-28 July 2023. DOI:10.1017/pds.2023.249

1 INTRODUCTION AND LITERATURE REVIEW

Machine learning (ML) advances have prompted companies to include intelligent solutions (also for cost estimation) in their enterprise software stack. These modern developments have made it possible to automate tasks that were previously thought impossible to program. Using decision-making, statistical and mathematical models to draw quick conclusions, particularly in the early stages of developing complex products and technologies, is an established industry practice.

The scientific literature on parametric cost modelling has ascertained the effectiveness of data mining, machine learning and artificial intelligence approaches for cost estimation during the preliminary design phase (Campi et al., 2021). During design, cost drivers are connected to the characteristics of CAD models. This is essential for a straightforward interpretation of the cost model. (Verlinden et al., 2008) proposed a procedure to create cost formulas using regression analysis and neural networks through a quick inspection of the CAD model. Through a precise and rapid calculation of the production cost of the various pieces of an assembly, it is possible to increase the competitiveness of products. Deep learning techniques can automatically learn the complex relationships between part characteristics and manufacturing processes within an ever-changing industrial system (Ning et al., 2020a). The method proposed by CHEN et al. (2021), applied in aviation, uses forecasting models to overcome the problem of the optimal choice of the development plan based on costs.

Machine learning for manufacturing cost prediction has made possible accurate predictions during the early product development stage by using little product information (Hennebold et al., 2022). Campi et al. (2021) presented a cost estimation methodology by testing parametric cost models based on machine learning on nine different sizes of axial compressor disks. Loyer et al. (2016) demonstrated the effectiveness of prediction models in the initial design phase, with little data available. They developed five statistical models using industrial data to estimate jet engine components' production cost. Yoo and Kang (2021) combined the geometric information from 3D CAD models with a production cost forecasting process using artificial intelligence. Thus, designers can visualise the characteristics that influence production costs. Also, Ning et al. (2020b) proposed a similar method that relates the manufacturing cost to the 3D CAD features extracted through deep learning methods.

Cost estimation can become very complex when a learning model is used for Life Cycle Costing (LCC). It requires much product information to coordinate with many different areas, from design to production (Kadir et al., 2020). Seo et al. (2002) proposed an estimation technique for LCC to be used in preliminary design. It is based on high-level information, generally known in the preliminary stage. This method promotes an integrated design process. A preliminary evaluation of production costs and energy consumption is necessary to determine the choice of a manufacturing technique. The problem is correctly considering all the information in the product life cycle. Kamps et al. (2018) proposed two integrated low-volume or high-variant sprocket production cost and life cycle assessment models. Bertoni and Bertoni (2020) offered a method to calculate the cost of design alternatives during the preliminary design, starting from information from Computer Aided Engineering (CAE).

Parametric models are usually used to calculate the unit cost of a product (Niazi et al., 2006). In a parametric model, the cost is a function of the component characteristics (geometric or non-geometric). These parameters are known as cost factors. Cavalieri et al. (2004) compared the results obtained from parametric and artificial neural network techniques, applying them to evaluate the unit production costs of a brake disc. Boothroyd and Reynolds (1989) concluded that designers could use a cost model to decide on materials and manufacturing techniques. Their case study is based on a parametric costing technique. The cost is estimated considering the volume of turned parts as a cost factor. COSYSMO 3.0 is an example of a tool that calculates the engineering costs of parametric systems. An open model allows users to numerically understand how its parameters affect cost estimates (Alstad, 2019). The literature shows an example of a tool capable of estimating production costs. This solution consists of two models. The first predicts different costs at various levels of any production plant. The second calculates the unit cost of future integral bladed disc designs used by the aerospace industry in gas turbine compressors (Langmaak et al., 2013). Masel et al. (2010) proposed an approach that uses process-based CERs (cost estimate reports) that have been created to evaluate the cost of jet engine parts.

Artificial intelligence (AI) is often used as a black box. Input data is provided, and outputs are given without understanding the reasons. In literature, indeed, AI is criticised for the impossibility of

https://doi.org/10.1017/pds.2023.249 Published online by Cambridge University Press

interpreting the results (Hihn and Menzies, 2015). Features selection, particularly the Feature Permutation Importance (FPI) method, can overcome this problem. The behaviour of a machine learning model can be understood using the global explanation approach of the importance of the permutation of features. Based on each attribute's influence on the predictions made by the trained machine learning model, it calculates and ranks the importance of each feature. The challenges that can arise when creating a cost model are related, for example, to the selection of hyperparameters and overfitting problems. Statistical learning approaches are often applied during model construction to understand the relationship between the main cost drivers and project costs (Elmousalami, 2021).

The paper presents a parametric cost model based on machine learning techniques for costing axisymmetric parts in the preliminary design phase. The method used for developing the cost model is based on CRISP-DM: CRoss Industry Standard Process for Data Mining. The approach overcomes the limitations highlighted in the literature review addressing the following challenges:

- Machine learning techniques require a large amount of data for training, which is not always available. Often, historical data are not as vast and coherent as requested. Their normalisation, when required, is complex but not always possible. In the method presented by the authors, training data are retrieved from an automatic cost estimation software tool based on an analytical approach. This solution allows for a vast database of coherent data by simulating the cost of different combinations (e.g., shapes, dimensions, materials, manufacturing processes).
- The method presented in this work proposes feature selection techniques that prevent the high correlation between parameters.
- FPI gives designers a list of product features influencing product cost. In this way, the cost model is less perceived as a black box.

2 METHODOLOGY

The approach is based on CRISP-DM (CRoss Industry Standard Process for Data Mining) method because it is an existing methodology for data mining. It comprises six steps, presented in the following sections: I. Business understanding; II. Data understanding; III Data preparation; IV. Modelling; V. Evaluation; VI. Deployment. The technical scope of this paper is to verify that, through machine learning techniques, it was possible to create a reasonably accurate cost model for axisymmetric components based on preliminary stages requirements.

2.1 Business understanding

In this phase, the conceptualisation of the objectives and needs takes place. The proposed approach aims to calculate the manufacturing cost of axisymmetric parts through machine learning techniques in the preliminary design phase. The output of this analysis is the understanding of acceptable model performance values and the selection of dependent parameters for prediction. The admissible estimation relative error for the model developed in this study is estimated at 30%. At the same time, only one output is expected for the cost model.

2.2 Data understanding

Data understanding involves identifying, collecting, analysing and verifying data sets to achieve the project objectives, translating into studying axisymmetric mechanical components. In detail, this phase is carried out through two activities:

The *first* task requires identifying all the parts of the specific product family. In this case, the model to be generated is very general and will consider only axisymmetric components.

The *second* task concerns the collection of data and information, which are:

- *3D CAD models*: 3D CAD models of mechanical components are necessary to assess the cost through an analytical software tool (LeanCOST by Hyperlean, Italy). For this study, 73 different models were collected (e.g., shafts, discs, pins, flanges, washers). Axisymmetric components are parts with rotational symmetry, mainly produced for turning operations. Although it is also possible to identify other processes, such as cutting, drilling, and milling, turning is the most crucial manufacturing phase.
- *Technical info*: the geometry can be described by several configuration parameters. The parametrises chosen must be common and able to describe all the parts analysed. The following geometric and product manufacturing information (PMI) have been included:

- Overall dimensions D1, D2, D3 [mm]. D1 and D2 are the maximum radial extensions, while D3 is the axial length;
- Type of finished [-]: if bulk or hollow;
- Internal diameter [mm]: ≠0 if hollow, =0 otherwise;
- Volume [mm^3];
- Surface [mm^2];
- Number of surfaces [-];
- Minimum tolerance [IT];
- Minimum roughness [µm];
- General roughness [µm];
- Surface with minimum roughness [mm^2];
- Surface with standard roughness [mm^2];
- Presence of non-coaxial geometries (e.g., holes non-coaxial to the turning axis) [yes/no];
- Presence of general milling geometries (i.e., any surface impossible to realise by turning) [yes/no];
- Presence of grooves for rings [yes/no];
- Presence of key sets [yes/no].

The first two activities provide all the data that can be grouped and sorted in a first database (or code list). All the associated configuration parameters describe each part, expressed separately as independent parameters. For this work, Microsoft Excel was used to create the database.

2.3 Data preparation

Data preparation can be considered the most critical phase of the methodology. Preparing the final data sets for modelling means building an extensive, structured data set to obtain accurate algorithms with low error. The data preparation phase can be divided into five different activities. In the end, the database will be ready for the modelling phase.

The *first* activity is to define all the possible geometrical and non-geometrical parameters that can drive the cost. In part, they have been described in the data understanding. The geometric parameters are the configuration parameters extracted from technical information and 3D CAD models. Non-geometric parameters are independent of the geometry (e.g., material and cost per weight, the production batch, and country). Thus, all these parameters, if considered, will represent the input to the parametric cost model. Therefore, independent parameters linked to non-geometric information are also added to the database obtained at the end of the data understanding.

The *second* activity is the extension of the independent parameter values. The accuracy and robustness of the results obtained by the parametric cost model prediction algorithm can be determined by the characteristics of the database used to train the model. The more the records, the better the training. The database generated for the axisymmetric pieces currently has 73 records, making it insufficient to conduct an appropriate analysis. The difference in values within the same parameter can help reduce the risk of overfitting. The currently obtained database does not satisfy this aspect. This activity aims to extend the number of records. In this study, it was preferred to focus on the geometric elements to analyse how these affect the cost. The non-geometric parameters have been fixed and not used to extend the model. Thus, the material (S235JR), production country (Italy), and production batch (one) are the same for all the components. To extend the database and focus on the geometric aspects, the parts were scaled by choosing the appropriate scaling factor for each. The scaling is done to enlarge the field of applicability of the model. The new models are valid (manufacturable) as the original parts. Scaling involves assigning factors to the dimensions that govern 3D models. In other words, these factors are values multiplied by all three parameters with which the model's measurements in space can be defined. Scaling can be homogeneous if all geometric parameters of a given model are multiplied by the same factor or heterogeneous if at least two dimensions are multiplied by different factors. It can then be an enlargement or reduction scaling if the factor is greater or lower than one. A homogeneous scaling has been applied so that the shape has not changed compared to the original models. For tiny models, only magnification factors were used. For large models, only reductions were applied. The goal was to achieve a homogenous distribution of parts in various sizes. To achieve this, all components have been divided into four clusters based on D1. Two different scale factors have been applied for each group (diverse for each category):

- *First group*: 0.05 – 0.25 for the first category of “large” objects, i.e. those with D1 > 250 mm. These factors lead to a significant reduction in size;
- *Second group*: 0.25 – 2.00 for the second category of “medium-large” objects (i.e. those with D1 > 100 mm and D1 <=250);
- *Third group*: 2.00 – 4.00 for the third category of “medium-small” objects (i.e., those with D1 > 50 and D1 <= 100 mm);
- *Fourth group*: 2.25 – 6.00, which involves the highest magnifications for the fourth category of “small” objects (i.e. those having D1 <= 50 mm).

D1 was chosen because it was considered the most impact dimension for the machining cost.

Table 1. Scaled parts

Group	Number of original parts	D1 range	Number of scaled parts
First	6	From 255 mm to 528 mm	12
Second	18	From 104 mm to 250 mm	36
Third	13	From 58 mm to 100 mm	26
Fourth	36	From 4 mm to 50 mm	72

At the end of this procedure (Table 1), the database has 219 parts (73 originals and 146 scaled parts). Figure 1 shows the distribution of records for parameters D1, D2 and D3. In this way, the applicability of the cost model's three dimensions (D1, D2, D3) is visible. For sizes larger than those indicated, relevancy is not guaranteed.

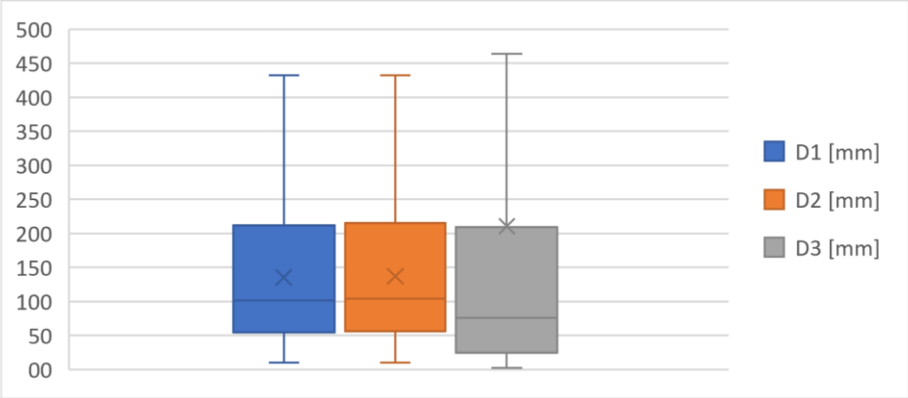

Figure 1. Distribution of records for parameters D1, D2 and D3

The *third* activity is to clarify how many dependent parameters (outputs) must be expected to define the number of parametric cost models. This study considers a single model for the total cost (material and manufacturing). Taking the database, the dependent parameter to predict is added. Figure 2 shows three mechanical components analysed and present in the database.

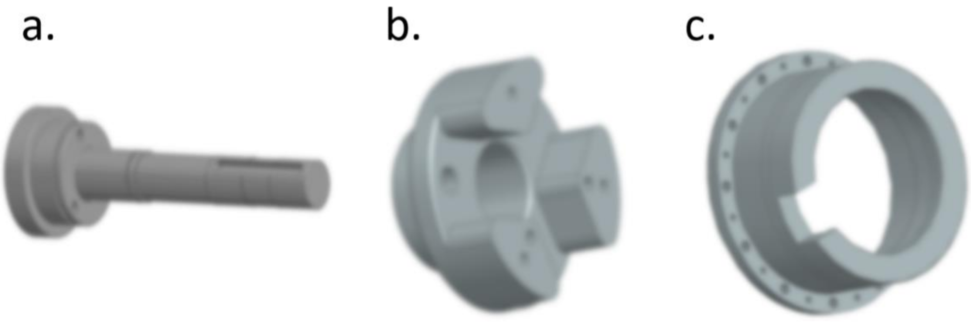

Figure 2. CAD 3D mechanical components: a) Shaft; b) Transmission coupling; c) Case

Starting from the extensive databases obtained, the *fourth* activity means completing the database by calculating the dependent parameters' values. The dependent metric is the cost of each component. An analytical costing software for mechanical components was used to obtain the cost values of each record. In

https://doi.org/10.1017/pds.2023.249 Published online by Cambridge University Press

this study, LeanCOST was used. The automatic analyses started from the 3D CAD models with embedded PMIs. LeanCOST evaluates the production cost, considering the material, investment, set-up, machining, and other expenses. In addition to the total cost [€], other valuable data were extracted to expand the database and to conduct analyses on the factual correctness of the research performed by the analytical software:

- *Cost of the raw material* [€]: Its value depends on the material specified in the model and the amount of material, i.e. by its size;
- *Process cost* [€]. Cost for the manufacturing phases (e.g., turning, saw cutting);
- *Unit cost of the material* [€/kg].

Since there is a single material, the cost per weight is expected to be similar for each component. This hypothesis is not confirmed because, for example, the raw material price depends on the stock dimensions and type. Of the 219 parts, LeanCOST could not analyse three of these, as it failed to assign the reference blank. Thus, these three parts were discarded.
Before finishing the data preparation phase, validating the results obtained from the analytical software tool is necessary for the *fifth* activity. Outliers must be identified. The methodology for identifying outliers is the Numeric Outlier Quartile Method. In this case, the analysis of the outliers consists of verifying that the €/kg ratio (total cost to the mass of the finished product) has a linear trend for increasing mass without anomalous jumps. Outliers are deleted from the database (Figure 3). From Figure 3 at record 188 (red point), there is a jump in the cost-to-mass ratio, which causes the loss of linearity. For this reason, we exclude these components from this point onwards.

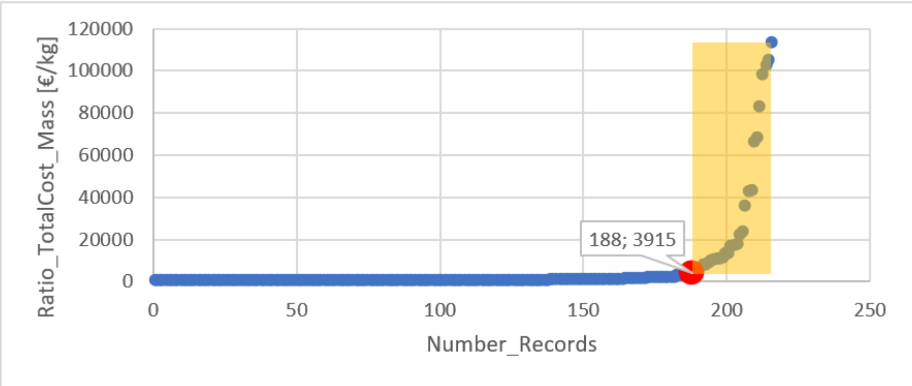

Figure 3. Outliers

2.4 Modelling

Modelling is the stage where various models are built and evaluated using different techniques. Each cost model aims to find the best machine-learning algorithm to predict the dependent parameter. For this phase, RapidMiner (by Altair Engineering) was used. It is a data science platform with many different algorithms for machine learning. In RapidMiner Studio, an extension, Auto Model, accelerates the process of building and validating models. The performance is calculated on a 40% hold-out set which has not been used for any of the performed model optimisations. This hold-out set is then used as input for a multi-hold-out-set validation where we calculate the performance for seven disjoint subsets. The largest and the highest performance are removed, and the average of the remaining five performances is reported here. Although this validation is not as thorough as full cross-validation, this approach balances runtime and model validation quality (RapidMiner Documentation, 2023). The authors chose to use the relative error because it is easily interpreted by an engineer and directly comparable with business understanding. The relative error is the average absolute deviation of the prediction from the actual value divided by the maximum of the real and forecast values. The values of the label attribute are the actual values (RapidMiner Documentation, 2023).
The *first* activity is to choose the algorithm to test. RapidMiner allows testing and comparing five algorithms: Generalised Linear Model, Deep Learning, Decision Tree, Random Forest and Gradient Boosted Trees.
Cost models are generated through RapidMiner in the *second* activity. Before launching the analysis, however, some columns were excluded because containing redundant information and were highly

https://doi.org/10.1017/pds.2023.249 Published online by Cambridge University Press

correlated to other parameters (Mass Finished and Mass Raw Material with the relative volumes; Material cost and Process cost with the Total cost). The database comprises 188 rows and 20 columns, including the dependent parameter.
In the *third* activity, the model obtained is evaluated through performance indicators. The result obtained will be used to choose the best algorithm. The authors selected for this study the relative error. Namely, it is the average absolute deviation of the prediction from the actual value divided by the real value.

Table 2. Model comparison

Model	Relative Error	Standard Deviation
Deep Learning	64.15%	±10.11%
Generalized Linear Model	53.41%	±9.87%
Decision Tree	24.41%	±2.42%
Random Forest	25.36%	±2.59%
Gradient Boosted Trees	27.89%	±2.87%

Activities of this phase are repeated for each algorithm to be tested. Through a direct comparison of the performance indicators, the algorithm providing the minor error is chosen (Table 2). The decision tree is selected for this study.
A feature selection activity has been envisaged in the *fourth* activity, which can also lead to the confirmation of the independent parameters if the results obtained from the performance indicators satisfy the requests. The choice of cost drivers is made to select the most representative parameters. The final option is left to the user, but he can receive suggestions through this method. The selection of the features can take place simply through the user's choice based on his experience, knowledge and critical sense. To support the user in choosing, there are two possible methods: filter and wrapper. Filter methods focus on the features' intrinsic properties without involving a model or a classifier (e.g. Pearson analysis). Wrapper methods involve a model, and the goal is optimising the predictive performance of a model (e.g. Sequential Forward Selection, SFS). For this study, it was decided to use SFS, a wrapper method, because it allows obtaining a higher level of accuracy than a filter method. In contrast, it requires a high computational load since it involves a prediction model.
From 19 independent parameters, the problem's dimensionality was reduced to 15. This reduction has improved the relative error, now 23.52% ± 1.37%.
The modelling phase must be repeated as often as the dependent parameters chosen. In this case study, there is one dependent parameter, the total cost. Table 3 shows the data collected for the parameters preferred by the selection feature for the three components shown in Figure 2.

Table 3. Cost drivers for some test components

Part	Shaft	Transmission coupling	Case
D1 [mm]	50	70	255
D2 [mm]	50	70	255
D3 [mm]	200	40	100
Volume_Finished [mm^3]	98187	62841	1122624
N_Surface	40	30	70
Min_Roughness [µm]	1.6	1.6	1.6
Std_R_Surface [mm^2]	18133	15788	141276
Milling_Geometry	YES	YES	YES
Ring_Seat	YES	NO	YES
Tab_Key_Seat	YES	NO	NO
D_Internal [mm]	0	25	160
Total_Surface [mm^2]	20645	17429	195015
Min_R_Surface [mm^2]	2512	1641	53739
Finished_Type	Full	Hollow	Hollow
Std_Roughness [µm]	6.3	3.2	6.3

https://doi.org/10.1017/pds.2023.249 Published online by Cambridge University Press

2.5 Evaluation

The cost model produced shows a higher ability to forecast that dependent parameter. This activity determines if the accuracy of the cost model satisfies the criteria set out in the Business Understanding phase. The obtained performance of the cost model is compared to that established at the outset. Using the cost model is possible if the comparison demonstrates compliance with the standards. The requirements are evaluated if the comparison reveals non-conformity with the criteria, which prevents the model from being employed. The cost model obtained has a relative error lower than the requirements of the business understanding.

2.6 Deployment

Users may control the parametric cost model throughout the deployment process and assess the effects of each independent parameter on the cost. This action represents the model interpretation. This activity aims to deliver a fully interpretable cost model, not a black box. A thorough feature importance research is conducted, allowing users to understand the relevance of each independent parameter and how it influences the cost. To determine the significance of the characteristic, many methods may be employed. The FPI method allows for identifying the factors affecting the cost of the product or process and subsequently assists the designer in using the cost model at the preliminary stage. This activity is introduced to improve the explainability of the model and not to improve performance. The FPI measures the increase in the model's prediction error after it permuted the feature's values, which breaks the relationship between the feature and the true outcome. The results of this analysis were calculated by the commercial tool RapidMiner. Figure 4 shows that the volume of the raw material (Total_Surface) is a very relevant parameter for obtaining the final cost. A change of this parameter during design significantly impacts the price.

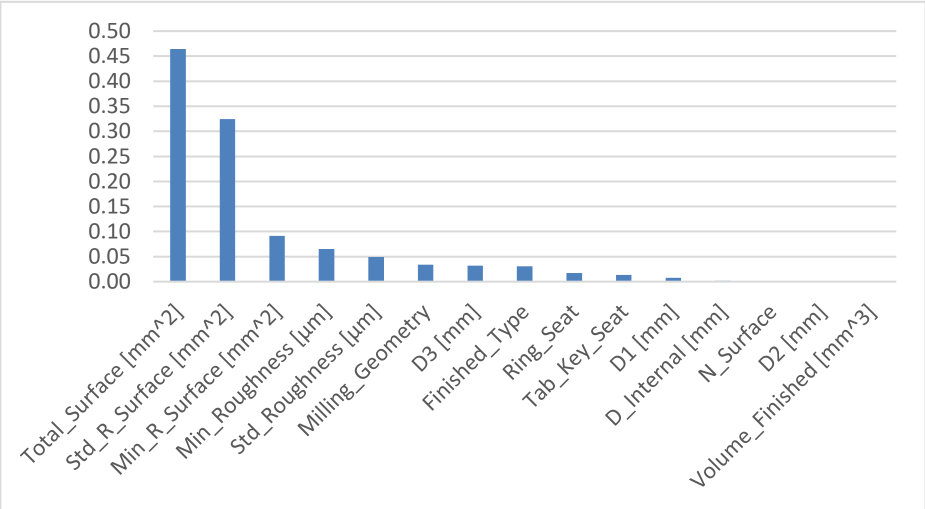

Figure 4. Feature permutation importance

3 RESULTS AND DISCUSSION

The cost model allows predicting an axisymmetric component's cost with a relative error of 23.52% ± 1.37%. The error obtained aligns with what the business understanding requires (30%). It falls within the Class 3 estimation of the degree of definition of a project as indicated by E2516 − 11, Standard Classification for Cost Estimate Classification System. The independent parameters required for the prediction are:

- Total_Surface [mm^2]: is the total surface of the part;
- Min_R_Surface [mm^2]: indicates the area destined to the minimal roughness;
- D3 [mm]: is the axial dimension of the part;
- Tab_Key_Seat: means whether machining has been carried out for the seats of keys and tongues;
- N_Surface: is the number of surfaces of the finite;
- Milling_Geometry: indicates if milling operations have been carried out;
- Std_R_Surface [mm^2]: it is the area destined to the standard roughness;

- Std_ Roughness [µm]: means the typical roughness;
- Ring_Seat: is the number of seats for rings;
- Volume_Finished [mm^3]: indicates the finished volume;
- Min_Roughness [µm]: means the minimum roughness;
- Finished_Type: is the type of the part (full or hollow);
- D1 [mm] and D2 [mm]: these are the two fundamental dimensions of the axisymmetric component;
- D_internal [mm]: indicates the internal diameter of the part.

Most cost drivers chosen are information known during the preliminary design phases. Using SFS as a feature selection technique made it possible to reduce the problem's dimensionality. This procedure is suitable because machine learning models sometimes worsen because they are often overfitted. The relative error improvement is a direct consequence. It should be underlined that the reduction of the independent parameters leads to an improvement in computational efficiency and an increase in the ease of data collection and favours the interpretability of the model. Applying the FPI allows for understanding the cost model and the importance of the features. Therefore the cost model is not perceived as a black box. LeanCOST, an analytical cost estimation software, has made it possible to obtain a robust database for training machine learning algorithms. Its use allowed for overcoming the limitations of developing a parametric cost model by using few, fragmented and not-normalised historical data. The developed method is not tied to any software tools. The methodology can be applied using other analytical cost estimation and data science software. The main limitation of this study concerns the limited number of non-geometric attributes (e.g., materials, production batch, production country, unitary material price) considered for developing the cost model. Furthermore, the database is based on medium-complex axisymmetric parts. It should be extended to manage more complex shapes.

4 CONCLUSIONS

The proposed method has made it possible to obtain a cost model, including business requests (business understanding). Then, in the data understanding, it is analysed the characteristics of the product to be assessed. In data preparation, it is created the database which was used in the modelling phase to generate the cost model. The evaluation phase made it possible to verify that the model obtained was in line with the initial requests. Finally, the interpretation of the accepted model took place during the deployment. The cost model coherently predicts the cost of axisymmetric components with E2516 − 11, Standard Classification for Cost Estimate Classification System. The problems of model overfitting and interpretability have been adequately addressed. Future work aims to improve the model by considering more complex components and non-geometrical features. The goal is to extend the analysis to other geometries, morphologies and production processes.

ACKNOWLEDGMENTS

The authors thank Miriam Nardelli (software developer) and Matteo Cicconi (student of Università Politecnica delle Marche) for their support during the entire research project.

REFERENCES

Alstad, J.P. (2019), "Development of COSYSMO 3.0: An Extended, Unified Cost Estimating Model for Systems Engineering", *Procedia Computer Science*, Elsevier B.V., Vol. 153, pp. 55–62, https://dx.doi.org/10.1016/j.procs.2019.05.055.

Bertoni, A. and Bertoni, M. (2020), "PSS cost engineering: A model-based approach for concept design", *CIRP Journal of Manufacturing Science and Technology*, Elsevier Ltd, Vol. 29, pp. 176–190, https://dx.doi.org/10.1016/j.cirpj.2018.08.001.

Boothroyd, G. and Reynolds, C. (1989), "Approximate cost estimates for typical turned parts", *Journal of Manufacturing Systems*, Vol. 8 No. 3, pp. 185–193, https://dx.doi.org/10.1016/0278-6125(89)90040-X.

Campi, F., Mandolini, M., Santucci, F., Favi, C. and Germani, M. (2021), "PARAMETRIC COST MODELLING OF COMPONENTS FOR TURBOMACHINES: PRELIMINARY STUDY", *Proceedings of the Design Society*, Cambridge University Press, Vol. 1, pp. 2379–2388, https://dx.doi.org/10.1017/pds.2021.499.

Cavalieri, S., Maccarrone, P. and Pinto, R. (2004), “Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry”, *International Journal of Production Economics*, Vol. 91 No. 2, pp. 165–177, https://dx.doi.org/10.1016/j.ijpe.2003.08.005.
CHEN, X., HUANG, J. and YI, M. (2021), “Development cost prediction of general aviation aircraft using combined estimation technique”, *Chinese Journal of Aeronautics*, Chinese Journal of Aeronautics, Vol. 34 No. 4, pp. 32–41, https://dx.doi.org/10.1016/j.cja.2020.07.024.
Elmousalami, H.H. (2021), “Comparison of Artificial Intelligence Techniques for Project Conceptual Cost Prediction: A Case Study and Comparative Analysis”, *IEEE Transactions on Engineering Management*, Vol. 68 No. 1, pp. 183–196, https://dx.doi.org/10.1109/TEM.2020.2972078.
Hennebold, C., Klöpfer, K., Lettenbauer, P. and Huber, M. (2022), “Machine Learning based Cost Prediction for Product Development in Mechanical Engineering”, *Procedia CIRP*, Elsevier B.V., Vol. 107, pp. 264–269, https://dx.doi.org/10.1016/j.procir.2022.04.043.
Hihn, J. and Menzies, T. (2015), “Data Mining Methods and Cost Estimation Models: Why is it So Hard to Infuse New Ideas?”, 2015 *30th IEEE/ACM International Conference on Automated Software Engineering Workshop (ASEW)*, IEEE, pp. 5–9, https://dx.doi.org/10.1109/ASEW.2015.27.
Kadir, A.Z.A., Yusof, Y. and Wahab, M.S. (2020), “Additive manufacturing cost estimation models—a classification review”, *The International Journal of Advanced Manufacturing Technology*, Springer, Vol. 107 No. 9–10, pp. 4033–4053, https://dx.doi.org/10.1007/s00170-020-05262-5.
Kamps, T., Lutter-Guenther, M., Seidel, C., Gutowski, T. and Reinhart, G. (2018), “Cost- and energy-efficient manufacture of gears by laser beam melting”, *CIRP Journal of Manufacturing Science and Technology*, Elsevier Ltd, Vol. 21, pp. 47–60, https://dx.doi.org/10.1016/j.cirpj.2018.01.002.
Langmaak, S., Wiseall, S., Bru, C., Adkins, R., Scanlan, J. and Sóbester, A. (2013), “An activity-based-parametric hybrid cost model to estimate the unit cost of a novel gas turbine component”, *International Journal of Production Economics*, Vol. 142 No. 1, pp. 74–88, https://dx.doi.org/10.1016/j.ijpe.2012.09.020.
Loyer, J.L., Henriques, E., Fontul, M. and Wiseall, S. (2016), “Comparison of Machine Learning methods applied to the estimation of manufacturing cost of jet engine components”, *International Journal of Production Economics*, Elsevier B.V., Vol. 178, pp. 109–119, https://dx.doi.org/10.1016/j.ijpe.2016.05.006.
Masel, D.T., Dowler, J.D. and Judd, R.D. (2010), “Adapting Bottoms-up Cost Estimating Relationships to New Systems”, *SPA/SCEA Joint Annual Conference and Training Workshop*.
Niazi, A., Dai, J.S., Balabani, S. and Seneviratne, L. (2006), “Product Cost Estimation: Technique Classification and Methodology Review”, *Journal of Manufacturing Science and Engineering, American Society of Mechanical Engineers(ASME)*, Vol. 128 No. 2, pp. 563–575, https://dx.doi.org/10.1115/1.2137750.
Ning, F., Shi, Y., Cai, M., Xu, W. and Zhang, X. (2020a), “Manufacturing cost estimation based on the machining process and deep-learning method”, *Journal of Manufacturing Systems*, Elsevier B.V., Vol. 56, pp. 11–22, https://dx.doi.org/10.1016/j.jmsy.2020.04.011.
Ning, F., Shi, Y., Cai, M., Xu, W. and Zhang, X. (2020b), “Manufacturing cost estimation based on a deep-learning method”, *Journal of Manufacturing Systems*, Elsevier B.V., Vol. 54, pp. 186–195, https://dx.doi.org/10.1016/j.jmsy.2019.12.005.
RapidMiner Documentation. (2023), “https://docs.rapidminer.com/9.3/studio/auto-model/”.
Seo, K.-K., Park, J.-H., Jang, D.-S. and Wallace, D. (2002), “Approximate Estimation of the Product Life Cycle Cost Using Artificial Neural Networks in Conceptual Design”, *The International Journal of Advanced Manufacturing Technology*, Vol. 19 No. 6, pp. 461–471, https://dx.doi.org/10.1007/s001700200049.
Verlinden, B., Duflou, J.R., Collin, P. and Cattrysse, D. (2008), “Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study”, *International Journal of Production Economics*, Vol. 111 No. 2, pp. 484–492, https://dx.doi.org/10.1016/j.ijpe.2007.02.004.
Yoo, S. and Kang, N. (2021), “Explainable artificial intelligence for manufacturing cost estimation and machining feature visualization”, *Expert Systems with Applications*, Elsevier Ltd, Vol. 183, https://dx.doi.org/10.1016/j.eswa.2021.115430.