Winter 12-2019

Predictive Cost Analytics of Vehicle Assemblies Based on Machine Learning in the Automotive Industry

Frank Bodendorf

Stefan Merbele

Joerg Franke

Predictive Cost Analytics of Vehicle Assemblies Based on Machine Learning in the Automotive Industry

Frank Bodendorf
Friedrich-Alexander-University of Erlangen-Nuremberg (FAU)
Institute for Factory Automation and Production Systems
Erlangen, Germany
frank.bodendorf@faps.fau.de

Stefan Merbele
Friedrich-Alexander-University of Erlangen-Nuremberg (FAU)
Erlangen, Germany
stefan.merbele@fau.de

Joerg Franke
Friedrich-Alexander-University of Erlangen-Nuremberg (FAU)
Institute for Factory Automation and Production Systems
Erlangen, Germany
joerg.frank@faps.fau.de

Abstract

Due to the high pace of development in the automotive industry there is a need for innovating cost engineering. A methodology for intelligent cost estimation in the early stages of the product life cycle is introduced. In a first step it is shown how significant economic and technical parameters for cost prediction can be prepared and filtered from historical calculation data. Subsequently, it is shown how cost prediction models can be developed using machine learning algorithms. Learning data and practical use cases come from a large automotive manufacturer in Germany. The models predict costs of car parts and assemblies of increasing complexity. Seven different machine learning models are trained and optimized. Based on the test data of the use cases these models are assessed and compared. Finally, the prediction results obtained are evaluated from different perspectives, demonstrating the practical applicability of the most suitable methods explored.

Keywords

Cost Engineering; Machine Learning; Predictive Analytics; Comparative Study; Automotive Industry.

Introduction

The growing innovation pressure associated with digitization is increasingly influencing the strategic actions of many companies. In the automotive industry, the pressure to innovate is particularly high and forces Original Equipment Manufacturers (OEMs) to continually adapt to new technological trends. The shortening of product life cycles, the boost of product complexity, and the flexible adjustment to customer requirements increasingly determine the way in which material components and assemblies are procured and calculated (Cooper & Edgett, 2008). The success of a purchasing process depends on the interplay of benefit, quality, and price (Chan, 2011; Cooper & Edgett, 2008; Relich & Bzdyra, 2014; Spalek, 2013). This

is exactly where cost analysis tries to help the buyer to fulfil these success criteria via traditional overhead calculation methods (Bottler & Engel, 1977; Cooper & Edgett, 2008; Trott, 2017; Ulrich, 2016).

The review of literature and reports of cost analyses in practice lead to the following insights:

- Cost engineering is a time-consuming and work-intensive process (Scanlan et al.).
- Tools for cost calculation cope with a trade-off between the accuracy of the estimate and the effort of tool application (VDI-Fachbereich Produktentwicklung und Mechatronik, 1997).
- In cost accounting the potential of modern information technology has hardly been realized so far (Simen, 2015).
- Traditional cost prediction tools are not reusable, given changes in product design or the technical characteristics (Newnes, Mil & Hosseini-Nasab, 2007).
- Accuracy of cost prediction depends largely on the timing of calculation (Castagne et al., 2008; Curran et al., 2007; Early, Price, Curran & Raghunathan, 2012; Kundu, Raghunathan & Curran; Price, Raghunathan & Curran, 2006).

Above statements motivate the necessity of a new analytics approach for cost calculation. The concept of an intelligent cost prediction procedure is presented and validation results coming from pilot tests are shown. The predictive model is based on machine learning and enables robust estimations of component-specific or assembly-specific cost values in an early product development phase and thus avoids the high data acquisition and implementation efforts of traditional cost calculations.

The potential of the developed model is first motivated by a comparison to the state of the art. The machine learning approach for intelligent cost prediction is explained in Section "*Machine Learning Models for Cost Estimation"* based on the methodology introduced in Section *"Research Goal and Methodology"*. Section *"Use Cases and Comparative Studies"* outlines pilot studies based on two selected forecast projects. Section *"Conlusion"* summarizes the results and discusses implications for practice.

Research Goal and Methodology

Overhead costing mentioned in "Introduction" and detailed in Bottler and Engel (1977) requires a large amount of information about material, production, and overhead rates. Therefore, such a cost prediction approach is only suitable to a limited extent in the early phase of the development process, since most of the information is not yet available at this point. Niazi et al. report first attempts to solve the problem and provide a detailed overview of ideas for appropriate cost estimation models (Niazi, Dai, Balabani & Seneviratne, 2006). A distinction is made between quantitative and qualitative methods. Qualitative methods include intuitive (Ahn et al., 2014) and heuristic methods (VDI-Fachbereich Produktentwicklung und Mechatronik, 1987), while qualitative methods involve analytical (Gupta & Galloway, 2003), parametric (Cavalieri, Maccarrone & Pinto, 2004), and synthetic methods (Coenenberg, Fischer & Guenther, 2016; Guenther & Schuh, 1998). However, in practice it turns out that these methods are not applicable. Experiments in real use case scenarios at a large German automotive manufacturer in Bavaria (detailed information is confidential) exhibit serious deficiencies. Reasons for this are, among others, an insufficient fulfillment of accuracy requirements, a time consuming method execution, a missing degree of detail of input parameters, and a difficult transferability to new use cases.

A new concept of intelligent cost forecasting addresses the above-mentioned problems of traditional methods (see "Machine Learning Models for Cost Estimation"). The methodology is based on the approach of learning systems, especially machine learning (ML). However, there is no "silver bullet", i.e. best method that delivers always the best results even for a definable class of problems. It is necessary to find out the most suitable method or ensemble methods (Thomas G. Dietterich, 2000; Lior Rokash, 2005; Cha Zhang & Yunqian Ma, 2012) for each use case by model studies.

With the help of a large historical database of sample calculations, executed at the German automotive manufacturer, different ML models are trained and their hyperparameters optimized. The model quality in terms of accuracy and variance of the predictions is then evaluated. Table 1 shows the methodological design. The entire ML development process follows the CRISP-DM approach (Cross Industry Standard Process for Data Mining). The analogy to CRISP-DM (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment) can be traced in the presented ML model in Figure 2 (right). Correlation Analysis is used for the selection of the most influencing factors (attributes for ML training). Seven ML models for cost prediction are developed, trained by historical data, evaluated by k-

fold cross-validation and the coefficient of determination R2, and finally applied and tested with new use cases.

Approach	Method
Process Model	CRISP-DM (Nisbet, Elder & Miner, 2009; Wirth & Hipp, 2000)
Feature Selection	Correlation Analysis
ML Algorithms	- Linear Regression (Lin.Regr.) - Polynomial Regression (Poly.Regr.) - k-Nearest-Neighbor Regression (KNN) - Decision Tree (DT) - Random Forest (RF) - Support Vector Regression (SVR) - Artificial Neural Network (ANN)
Model Quality Check	- Coefficient of Determination R2 (Stocker & Steinke, 2016) - K-fold Cross-Validation

Table 1. Methodological design

Machine Learning Models for Cost Estimation

The model development process is divided into two closely interlinked components:

- Feature selection (analysis of influencing factors) to reduce complexity (see Figure 1 left)
- Model training and optimization for cost estimation (see Figure 1 right)

The training basis for the ML models is raw data of historical calculations stored in a data lake. In the first step, the raw data is processed and clustered according to product categories. This is followed by the inspection of influencing factors based on correlation analyses (see Figure 1 left). As a result, economic and technical parameters are extracted for the training, which on the one hand are highly correlated with the final cost value (to be predicted later in new use cases) and on the other hand show low correlations among each other.

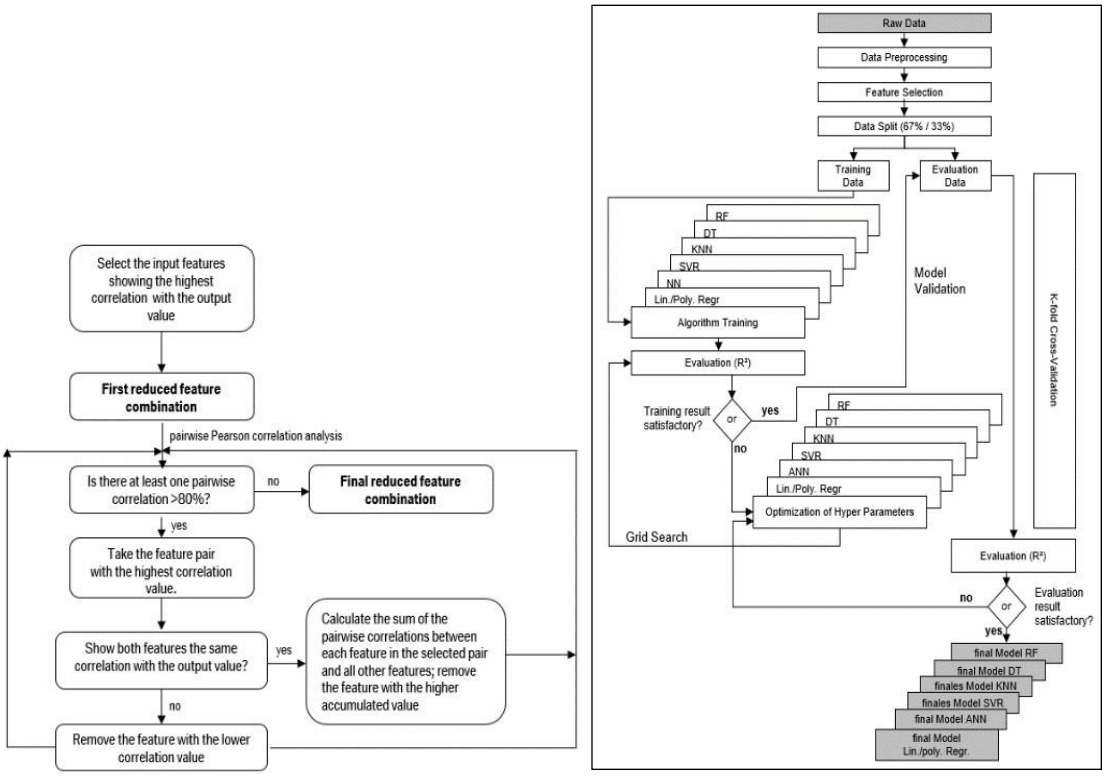

Figure 1. Feature selection (left) and development of the ML cost prediction models (right)

For feature selection the correlation values play an important role. However, in real application scenarios it is equally important whether the data required for ML training can be obtained with relatively little effort in an early development phase. Both selection criteria are illustrated by use cases in "Use Cases and Comparative Studies". Once this selection has been made, the machine learning models listed in Table 1 are implemented (see Figure 1 right).

The model is evaluated by repeatedly splitting the sample database into training data (67% of the database) and evaluation data (33% of the database). If the value of the quality criterion R2 is satisfying after k-fold cross-validation, the models are tested using data from new use cases. The hyperparameters of the corresponding models are further optimized ("tuned"). This is done by using special modules of development packages, such as "scikit-learn" from Python (see Table 2).

Model	**Parameter**	**Parameter-values**	**Package in Scikit**
Lin./poly. Regr.	non-existent	non-existent	LinearRegression PolynomialRegression
ANN	Number of neurons in concealed layer (n)	10, 20, 30,..., 100	MLP Regression
	Learning rate (lr)	0.1	
	Momentum (mc)	0.1, ...,0.9	
	Epochs (ep)	1000,...., 10000	

SVR	Kernel	Polynominal, radial	SVR
	Degree of polynomial (d)	1, 2, 3, 4	
	Regularization parameters (C)	1, 10, 100	
	Width of kernel function (γ)	0,0.1, 0.5,...,5.0	
KNN	Number of k-nearest neighbors (k)	1, 2, 3,..., 15	KNeighborsRegressor
DT	Depth of the tree (md)	1, 2, 3,..., 10	DecisionTree Regressor
	Number of sheets (sl)	1, 2, 3,..., 10	
RF	max. depth of the tree (md)	0,1,2,..., 10	RandomForest Regressor
	Number of trees (e)	10,..., 100	
	max. features (mf)	Results from Kn	

Table 2. Tuning of hyperparameters

The test data is used to simulate future use cases and to determine the deviations between the cost values generated by the ML algorithms and the real calculation values. If the result is satisfying, the corresponding ML model can be implemented, visualized, and documented for the practical user.

Use Cases and Comparative Studies

ML models mentioned in "Machine Learning Models for Cost Estimation" are tested in pilot studies on two calculation objects. At first, simple components are used in order to facilitate a basic understanding for the mode of operation to predict costs with machine learning methods. Based on these experiences, the cost prediction is then carried out for a complex assembly.

Use Case 1: Punched Parts

In the first use case, punched parts are characterized from different angles (see Table 3). They are built in vehicles in numerous variants and thus provide a very large database for ML training (over 4200 calculations with many parameters for each).

Symbol	**Meaning**
B	component
R	(raw) material
F	fabrication/production
M	machine
A	labor
W	tool

Table 3. Feature abbreviations for punched parts

As shown in Figures 1 and 2, the pre-processed data is first used to select the most influential parameters for the forecast (feature selection). A correlation analysis is performed for this purpose. The results show how strongly a single parameter correlates with the value to be predicted. If the values are directly

proportional, the correlation value is close to 1, if the values are indirectly proportional, the correlation value approaches -1. With correlation values close to zero no correlation is assumed. For the feature selection, parameters with values close to 1 or -1 are therefore considered primarily (Winker, 2007).

First, the correlations between the features of the punched parts and the corresponding overall cost values of the parts are in the focus. The result of this examination is shown in Table 4.

Parameter	**Correlation with overall cost**
Profit (amount)	0,999
Cost of materials (raw materials and purchased parts)	0,996
Direct material costs	0,995
R1-Material (raw material and purchased parts)	0.994
R1-Cost	0.994
R1-Material direct costs	0,994
R1-Gain (amount)	0,992
F-Material costs (raw materials and purchased parts costs)	0,972
F-Material committee (amount)	0,972
Material committee (Amount)	0,972
R2-Direct material costs	0,941
R2-Cost	0,928
R3-Gain	0,826

Table 4. Correlation analysis of stamped parts' features

R1, for example, indicates the amount of material of a component in the end product. R2, on the other hand, represents the gross incoming material required for the production of a component. Since only a certain percentage of the material entering the stamping machine is used for the stamped part, the weight of R2 is always higher than that of R1. Finally, R3 refers to the scrap produced in the stamping process. This represents the difference between R2 and R1. Furthermore, Table 4 shows the production costs (F) and the costs for the required specialist (A1).

As mentioned in "Machine Learning Models for Cost Estimation", it is not only the correlation value of one parameter that determines its selection, but also the fact whether it can be collected or procured with little effort. Several parameters of the R categories such as profit and material costs correlate highly with the cost value of the component to be predicted. However, these can only be determined with high effort. Parameters like R1-quantity (component weight) or A1-quantity (working time of the craftsman) can be determined easily, but will lead to inaccurate forecast results due to their low correlation value. The following intermediate processing steps are carried out in order to keep the effort for procuring the parameters to a minimum and at the same time to maximize the quality of the forecast.

Step 1: Forecasting the R2-Quantity

R2-Quantity shows a high correlation value in Table 4 but is difficult to gather directly. A new correlation analysis filters out highly correlating, easily obtainable parameters that are suitable for predicting the R2-

Quantity (gross material requirement). These are shown in Table 5. Due to the high correlation value, the parameter MGK (material overhead) should be selected first. However, since this material overhead cannot be assigned to the individual cost objects (products), this parameter is difficult to determine and therefore cannot be taken into account (Dahmen, 2014). Thus, in several steps, highly correlating, easily to obtain parameters are filtered out which are suitable for predicting the R2-Quantity (gross material requirement). R1-Quantity (component weight) is easy to determine and is immediately used for forecasting. W1-Lifetime-Invest describes the costs for the stamping tool used and thus implicitly reflects the complexity of the produced component. Since this value can be procured with little expenditure, also this one is used for the R2 forecast.

Parameter	**Correlation (R2-Quantity)**
MGK (amount)	0,963
R1-Quantity	0,945
R2-Gain (Amount)	0,801
W1-Lifetime-Invest	0,797
R3-Quantity	0,796
M-Quantity	0,788
W1-Quantity	0,788
R2-Costs	0,718
R2-Material-Costs	0,701

Table 5. Correlation of parameters with R2-Quantity

After having selected the most promising parameters (R1-Quantity, W1-Lifetime-Invest) for predicting the R2-Quantity, seven machine learning models for R2 prediction are developed, i.e. trained, evaluated, and optimized. These are Artificial Neural Network (ANN), Support Vector Regression (SVR), Polynomial Regression (Poly. Regr.), Linear Regression (Lin Regr.), K-Nearest-Neighbor (KNN), Random Forest (RF), and Decision Tree (DT). To compare these developed models to each other, they are tested with new use case data as described in "Machine Learning Models for Cost Estimation". Data that is completely unknown to the models from training is used for this purpose. The results are shown in Figure 2 (left).

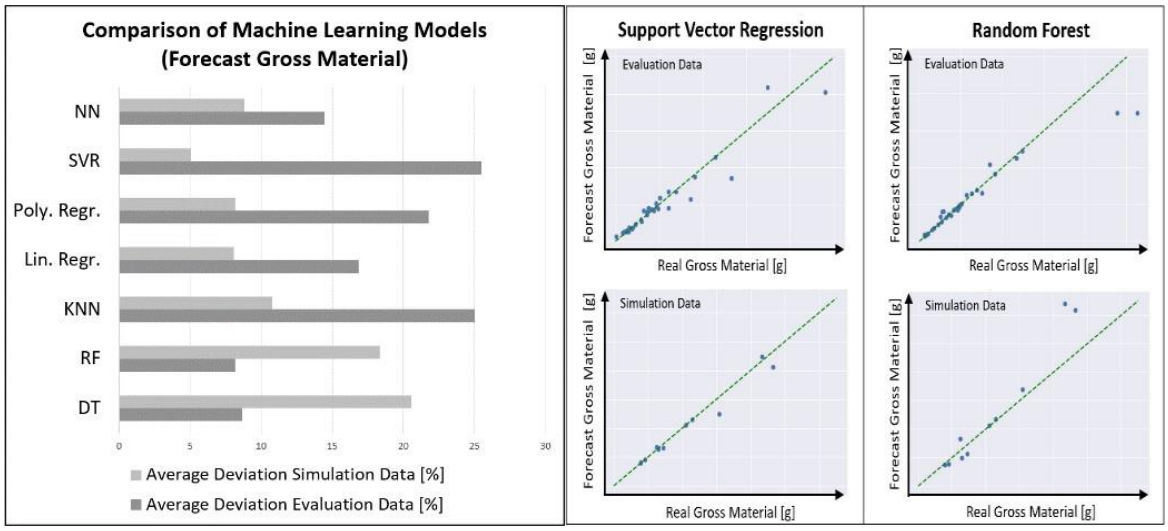

Figure 2. ML prediction quality of R2-Quantity (left) and SVR and RF predictions of R2-Quantity (right)

The light grey bars represent the average deviation of the model results, i.e. forecasts for R2-Quantity, from the "real" results of the use case, unknown to the model. The dark grey bars represent the average deviation of the model results from the evaluation data during the model training, using the historic sample data. It can be seen that there is a high variation of the differences between grey and black bars. The models with the lowest deviations based on the evaluation data, i. e. dark grey bars (Random Forest, 8.17%), and on unknown data, i.e. light grey bars, (Support Vector Regression, 5.05%), are used in the following to take a closer look.

The upper graph of Figure 2 (right) shows the model accuracy using evaluation data and the lower graph the accuracy using unknown data (simulation of new use cases). The abscissa indicates the real value for the respective data point and the ordinate the forecast value. Thus a straight line with the slope 1 would be optimal, since thereby the values would coincide. If the data point lies below this line, the predicted value is too low. If the point is above it, the model estimates the value to be too high.

It is noticeable that the forecast of the Support Vector Regression "fans out" with increasing values, i.e., higher cost values diverge more, particularly are estimated to be too low. With Random Forest, on the other hand, the costs are predicted very accurately over a larger range of values until only the highest range finally shows strong deviations.

Step 2: Calculation of further parameters

The parameters R1- and R2-Quantity are now available. Further highly correlating parameters (see Table 4) are added in the second step. As described above, the scrap quantity is the difference between the R1- and R2-Quantities:

$$R3\text{-}Quantity = R2\text{-}Quantity - R1\text{-}Quantity \quad (1)$$

The R2-Material-Costs are calculated as follows using the current stock exchange price for steel:

$$R2\text{-}Material\_Costs = R2\text{-}Quantity \times steel\ price \quad (2)$$

Similarly, the R3 material costs are calculated with the current scrap price:

R3-Material-Costs= R3-Quantity x scrap price *(3)*

It should be recalled that the R3-Material-Costs are actually a revenue coming from scrap sales. Thus, the R1-Material-Costs are reduced by the R3 scrap revenue:

R1-Material-Costs = R2-Material-Costs – R3-Material-Costs *(4)*

Step 3: Forecasting the costs of punched parts

For the final forecast of the overall cost of the punched part, the different machine learning models are assessed regarding the forecast quality and compared to each other (see Figure 3 left). Here, it can be seen that the forecasts differences based on evaluation data from the sample data and unknown data from the new use cases are very small. Only the Decision Tree model shows strong deviations. The Neural Network and the Random Forest offer the smallest deviations. Therefore, they are analyzed in more detail.

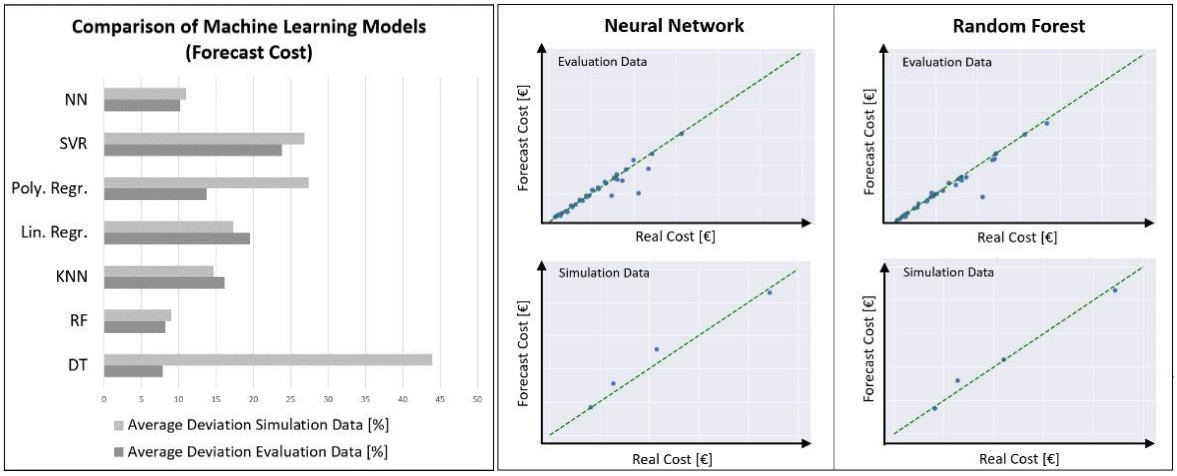

Figure 3. ML prediction quality of overall punch part costs (left) and ANN/RF predictions and comparisons of real punch part costs (right)

Figure 3 (right) shows that both the Artificial Neural Network and the Random Forest predict very correctly, based on the evaluation data as well as unknown data. Only for values close to 1 Euro both models predict values that are clearly too low. With smaller and larger values, however, the deviations are very small.

Use Case 2: Engine Covers

The findings of the models for predicting costs of single components are used to estimate the costs of vehicle assemblies. As shown in Figure 4, the engine cover consists of seven individual components with one assembly.

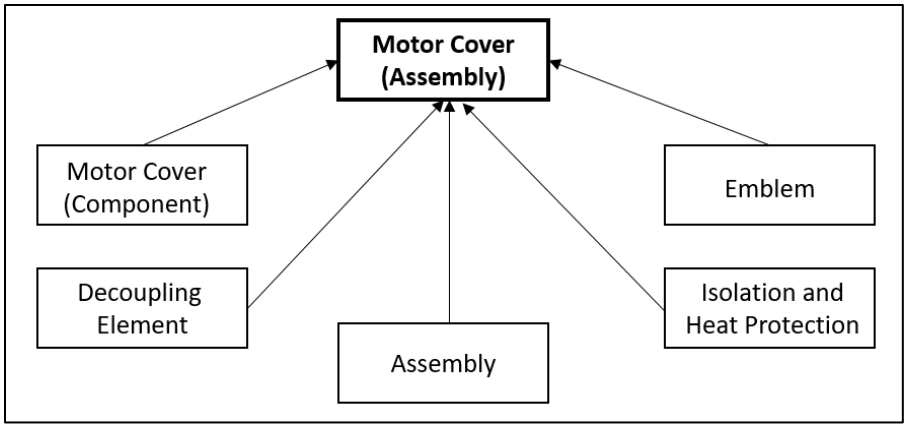

Figure 4. Components of motor cover

In order to filter out the most influential parameters, a correlation analysis is performed with all available parameters. The features that correlate closely with the overall cost are shown in Table 6. It can be seen that the total assembly has the highest correlation value. However, since the total installation represents the sum of all installation steps, it is very time-consuming to obtain this value, which is why it is not taken into account. Assembly step 1, on the other hand, represents the building of the motor cover (component) using the injection moulding process, which is relatively easy to determine. The clamping force in tons characterizes the injection moulding machine and can be used as well as a parameter. The geometric parameters and the component weight are also highly correlated and can be obtained with little effort. The emblem is regarded as a standard component and can therefore also be integrated. The approach is to use all parameters with a correlation value of about 0.6 and above, except for the assembly as a whole.

Parameter	**Correlation (Costs)**	**Parameter**	**Correlation (Costs)**
Assembly Total	0,854	Raw material costs (co)	0,688
Assembly Step 1 (ass)	0,804	emblem	0,607
Motor cover length (lgt)	0,750	Motor cover width (wdt)	0,596
Motor cover area (ar)	0,749	Assembly step 2	0,503
Locking force in tons (lf)	0,739	Assembly step 8	0,460
Assembly step 6	0,453	Assembly Step 5	0,421
Assembly Step 9	0,428		

Table 6. Correlations of features with motor cover costs

In a further consideration, it does not make sense to use all input parameters for ML training that correlate highly with each other. Figure 5 shows the results of the correlation analysis, i.e. the relationships between the input parameters. The goal is to pick those features that correlate as little as possible with each other but have a high correlation with the overall cost to be predicted. For example, the emblem correlates with

0.607 with the cost, but only with a maximum of 0.30 with the other parameters. When selecting features, the parameter Emblem seems to be important for a stable prediction.

	Em	Lgt	Wdt	Ar	LF	Com	Ass	Co
Emblem	1,00	0,44	0,24	0,41	0,32	0,35	0,28	0,30
Motor Cover Length	0,44	1,00	0,62	0,93	0,89	0,85	0,82	0,84
Motor Cover Width	0,24	0,62	1,00	0,78	0,82	0,84	0,78	0,80
Motor Cover Area	0,41	0,93	0,78	1,00	0,97	0,92	0,88	0,92
Locking Force	0,32	0,89	0,82	0,97	1,00	0,92	0,91	0,92
Component Weight Motor Cover	0,35	0,85	0,84	0,92	0,92	1,00	0,85	0,97
Assembly Step 1	0,28	0,82	0,78	0,88	0,91	0,85	1,00	0,86
Cost Raw Material	0,30	0,84	0,80	0,92	0,92	0,97	0,86	1,00

Figure 5. Multicorrelation of parameters for engine covers

In addition, parameters are extracted by which the cost of the raw material and the Assembly step 1 can be estimated. The raw material costs are predicted using the length, width, and area of the engine cover, the component weight, and the clamping force in tons. Assembly step 1 is estimated by Artificial Neural Network and Random Forest models as they show the smallest average deviations (see Figure 6 left).

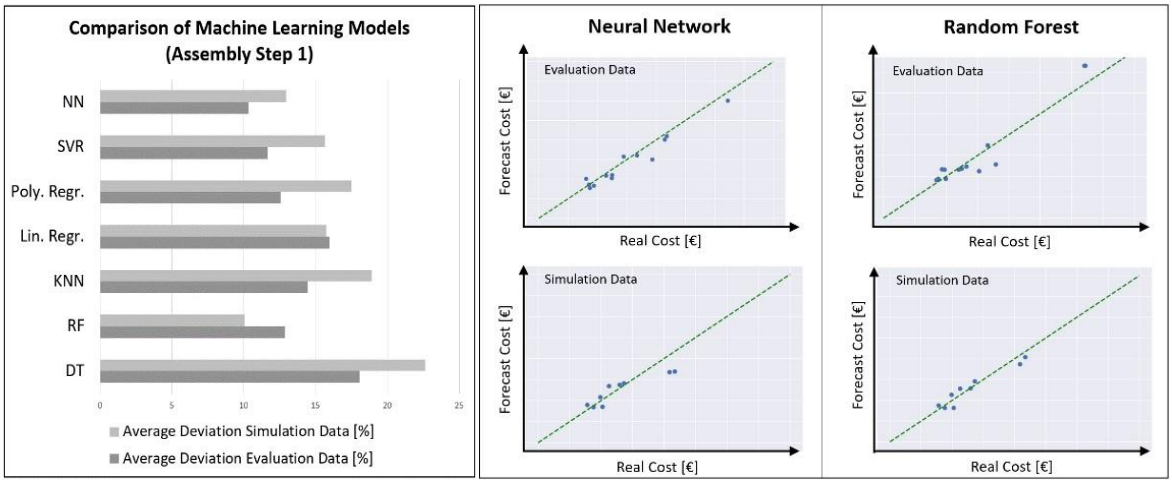

Figure 6. ML prediction quality of Assembly Step 1 (left) and ANN/RF predictions for Assembly step 1 (right)

Figure 6 (right) shows why the Neural Network performs better with the evaluation data than the competing Random Forest. With the Neural Network, the data points are consistently in a narrow range close to the optimal prediction. Basically, the model slightly underestimates the data. The Random Forest, on the other hand, reproduces the data more accurately, as in previous analyses, where larger deviations can be identified again and again. Fortunately, the pattern is different for the prediction of the use case data, where very good forecasts are seen. Since the Neural Network has two major deviations in the prediction of new use case data, the results of the Random Forest are used for the final forecast of the motor cover costs. When comparing the machine learning algorithms for predicting the overall cost of the engine cover assembly, it is noticeable that the quality of the models becomes very similar (see Figure 7 left). Thus, no large differences between the individual models can be detected. As in previous considerations, however, the Neural Network (for new use case data) and the Random Forest (for evaluation data) show the best results.

The Random Forest model provides a deviation of less than 11% for evaluation data and less than 10% for unknown data.

These two models are looked at in more detail (see Figure 7 right). The Neural Network predicts very accurately in the lower value range. In the middle range, however, a strong deviation can be seen. In the higher value range, cost values are estimated clearly too small. With the use case data, on the other hand, the points are continuously assessed as minimally too low. In the high value range the data points deviate strongly downwards again. A similar pattern can be seen with Random Forest. However, this algorithm smooths the deviations of evaluation data which results in smaller numbers of deviations. But even in this case, a downward deviation trend for values in the highest range is to be expected.

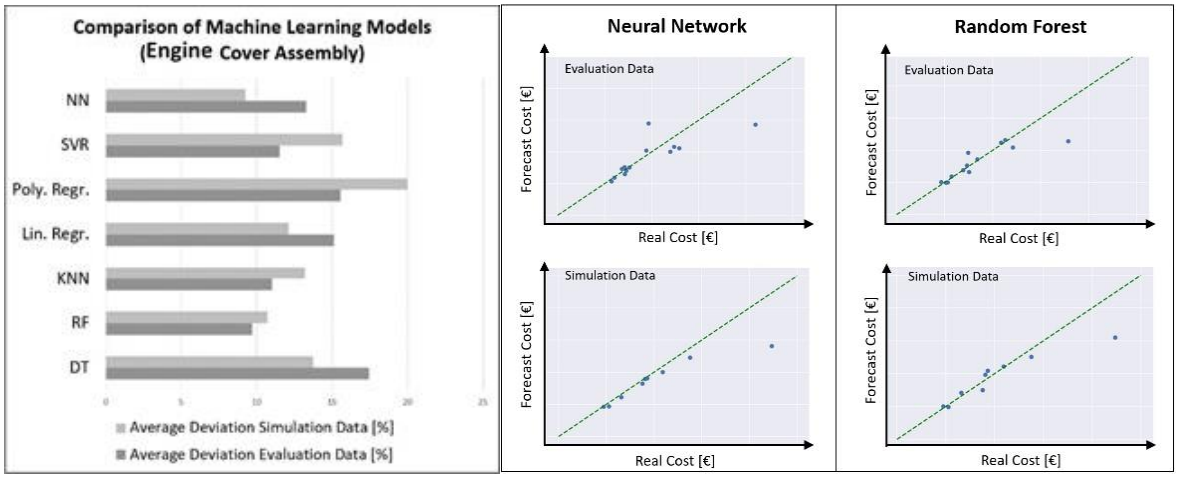

Figure 7. ML prediction quality of Engine Cover Assembly (left) and ANN/RF predictions and comparisons with real costs for motor cover (right)

Conclusions

The design and comparative studies of machine learning models show that predicting costs of components or assemblies in automotive manufacturing is a challenge that can be met by new methods coming from the field of artificial intelligence and in particular from learning systems (see "Introduction", insight 3). Machine learning allows to reduce the number of parameters used for cost estimation and thus makes the calculation procedures more efficient (see "Introduction", insight 1). Furthermore the Machine learning Model (see Figure 2) can be reused for various cost prediction use cases (punched parts total costs, motor cover total costs or material costs) (see "Introduction", insight 4). In addition, the use of parameters that are as simple as possible, such as geometrical features, enables robust and fast predictions in the early product life cycle (see "Introduction", insight 5). A set of seven machine learning algorithms is examined for this purpose. It is shown that it is possible to arrive at accurate and reliable forecasts by assessing different machine learning models and selecting the best ones. The weaknesses of the models can be ascribed to insufficient amounts of training data for specific value ranges of the output. It should be feasible to amend this by collecting more sample data.

For testing the models in use case 1 "Punched Parts", a sample of seven not yet calculated punched parts is applied. The selected ML model delivers cost values in the interval of [0.36€; 2.35€]. After a manual calculation of the components, there is a standard deviation of 3.01% between the ML and the manual results.

The same practical test is also done for use case 2 "Engine Covers". The cost forecast for the sample size of seven engine covers delivers cost values in the interval of [4.2€; 22.3€]. After a manual calculation of the same components, there is a standard deviation of 9.2% between the ML and the manual results (see "Introduction", insight 2).

References

Ahn, J., Ji, S.H., Park, M., Lee, H.S., Kim, S., & Suh, S. (2014). The attribute impact concept: Applications in case-based reasoning and parametric cost estimation. *Automation in Construction*, 43, 195–203.

Bottler, J., & Engel, B. (1977). *Kostenträgerstückrechnung (Kalkulationsverfahren)*. Wiesbaden: Gabler Verlag.

Castagne, S., Curran, R., Rothwell, A., Price, M., Benard, E., & Raghunathan, S. (2008). A generic tool for cost estimating in aircraft design. *Research in Engineering Design*, 18(4), 149-162.

Cavalieri, S., Maccarrone, P., & 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*, 91(2), 165-177.

Chan, S.L., & Ip, W.H. (2011). A dynamic decision support system to predict the value of customer for new product development. *Decision Support Systems*, 52(1), 178-188.

Coenenberg, A.G., Fischer, T.M., & Guenther, T. (2016). *Kostenrechnung und Kostenanalyse*. Stuttgart: Schäffer-Poeschel Verlag, (pp. 138-153).

Cooper, R.G., & Edgett, S.J. (2008). Maximizing productivity in product innovation. *Research-Technology Managemen*t, 51(2), 47-58.

Curran, R., Castagne, S., Early, J., Price, M., Raghunathan, S., Butterfield, J., & Gibson, A. (2007). Aircraft cost modelling using the genetic causal technique within a systems engineering approach. *The Aeronautical Journal*, 111(1121), 409-420.

Dahmen, A. (2014). *Kostenrechnung*. München: Franz Vahlen.

Dietterich, T. (2000): Ensemble Methods in Machine Learning. In *International workshop on Multiple Classifier Systems* (pp. 1-15). Berlin: Springer.

Early, J.M., Price, M.A., Curran, R., & Raghunathan, R. (2012). Whole Life Costing for Capability. *Journal of Aircraft*, 49(3), 712-723.

Guenther, T., & Schuh, H. (1998). Näherungsverfahren für die frühzeitige Kalkulation von Produkt- und Auftragskosten. *Kostenrechnungspraxis*, 42(6), 381-389.

Gupta, M., & Galloway, K. (2003). Activity-based costing/management and its implications for operations management. *Technovation*, 23(2), 131-138.

Kundu, A., Raghunathan, S., & Curran, R. (2002): Cost Modelling as a Holistic Tool in the Multi-disciplinary Systems Architecture of Aircraft Design. The Next Frontier. 41st AIAA Aerospace Sciences Meeting & Exhibit, Reno, (pp. 327).

Newnes, L.B., Mil, A.R., & Hosseini-Nasab, H. (2007). On-screen real-time cost estimating. *International Journal of Production Research*, 45(7), 1577-1594.

Niazi, A., Dai, J.S., Balabani, S., & Seneviratne, L. (2006). Product Cost Estimation: Technique Classification and Methodology Review. *Journal of Manufacturing Science and Engineering*, 128(2), 563.

Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Burlington: Academic Press.

Price, M., Raghunathan, S., & Curran, R. (2006). An integrated systems engineering approach to aircraft design. *Progress in Aerospace Sciences*, 42(4), 331-376.

Relich, M., & Bzdyra, K. (2014). Estimating new product success with the use of intelligent systems. *Foundations of Management*, 6(2), 7-20.

Rokach, L (2005). Ensemble methods for classifiers. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 957-980). New York: Springer US.

Scanlan, J., Rao, A., Bru, C., Hale, P., & Marsh, R. (2005). Project: Cost Estimating Environment for Support of Aerospace Design Decision Making. *Journal of Aircraft*, 43 (4), 1022-1028.

Simen J.-P. (2015). *Schätzung betrieblicher Kostenfunktionen mit künstlichen neuronalen Netzen*. Hohenheim: Institution für Financial Management.

Spalek, S. (2013). Improving Industrial Engineering Performance through a Successful Project Management Office. *Ekonomika-Engineering Economics*, 24(2), 88-98.

Stocker T.C., & Steinke, I. (2016). *Statistik: Grundlagen und Methodik*. Berlin: Walter de Gruyter GmbH & Co KG.

Trott, P. (2017). *Innovation management and new product development*. Harlow: Pearson.

Ulrich, K.T., & Eppinger, S.D. (2016). *Product design and development*. New York: McGraw-Hill Professional.

VDI-Fachbereich Produktentwicklung und Mechatronik (1987). *Economical decisions during design engineering process; methods and equipment. VDI2235.* Düsseldorf: VDI-Gesellschaft Produkt- und Prozessgestaltung.

VDI-Fachbereich Produktentwicklung und Mechatronik (1997). *Design engineering methodics - Engineering design at optimum cost - Simplified calculation of costs. VDI2225 Blatt 1.* Düsseldorf: VDI-Gesellschaft Produkt- und Prozessgestaltung.

Winker, P. (2007). *Empirische Wirtschaftsforschung und Ökonometrie*. Wiesbaden: Springer.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, Citeseer, (pp. 29-39).

Zhang, C., & Ma, Y. (2012). *Ensemble Machine Learning: Methods and Applications*. Heidelberg: Springer.