INTERNATIONAL DESIGN CONFERENCE – DESIGN 2022
https://doi.org/10.1017/pds.2022.168

Using Machine Learning for Product Portfolio Management: A Methodical Approach to Predict Values of Product Attributes for Multi-Variant Product Portfolios

J. Mehlstäubl [1,✉], F. Braun [2], M. Denk [1], R. Kraul [2] and K. Paetzold [3]

[1] Universität der Bundeswehr München, Germany, [2] MAN Truck & Bus SE, Germany,
[3] Technische Universität Dresden, Germany

✉ jan.mehlstaeubl@unibw.de

Abstract

To satisfy customer needs in the best way, companies offer them an almost infinite number of product variants. Although, an identical product was not built before, the values of its attributes must be determined during the product configuration process. This paper introduces a methodical approach to predict the values of product attributes based on customer feature configurations using machine learning. Machine learning reduces the effort compared to rule-based expert systems and is both, more accurate and faster. The approach was validated by predicting vehicle weights using industrial data.

Keywords: machine learning, portfolio management, product development, artificial intelligence (AI), data-driven design

1. Introduction

The demand for customised products as well as the number of product variants offered by companies has grown significantly in many industries (Hochdörffer *et al.*, 2018; Schuh *et al.*, 2018). This results in an extension of complexity and expenses (Agard and Kusiak, 2004). In industry the challenge is addressed by introducing product platforms, product modules, and generic product structures (ElMaraghy *et al.*, 2013; Kreimeyer *et al.*, 2016). However, the large number of features that can be combined by customers leads to an almost infinite number of variants. For example, if a customer can choose 10 optional features, this already results in a theoretical number of 2^{10} variants. A BMW 7 series car can have for example up to 10^{17} possible configurations (Hu *et al.*, 2008). For a product as large as a truck, this is an infeasible number of product configurations to be produced (Kusiak *et al.*, 2007) and there are usually no two identical configurations sold to different customers. Nevertheless, the value of product attributes must be specified already during the product configuration process. This information is currently provided by the complex and manual definition of calculation rules. The manual process of rule definition is time consuming and requires a high level of expertise (Haug *et al.*, 2012). Due to technological progress and increasing digitalisation, companies have new technological possibilities that provide new solutions for existing problems. A particularly large potential in product portfolio and variety management is offered by machine learning approaches (Mehlstäubl et al., 2021). However, machine learning is currently hardly used for the development of products (Bertoni *et al.*, 2017). Machine learning can identify patterns in past data and make predictions about future events based on them (Murphy, 2012). In this way, knowledge of products and markets can be improved, and well-founded decisions can be made. In the context of product attribute value prediction, the application of machine learning can lead to an effort reduction as well as faster and more accurate predictions.

The described problem leads to the following research questions:

- How can values of product attributes of multi-variant product portfolios be predicted, based on new customer feature configurations using machine learning?
- What advantages can be achieved by using machine learning to predict values of product attributes of multi-variant product portfolios, compared to rule-based expert systems?

In the research approach to prepare this article, the problem and objectives were first defined based on the literature and the results of an interview study with an industry partner from the commercial vehicle sector. In the interviews the prediction of attribute values was identified as a valuable use case for machine learning. Subsequently, the use case was specified in more detail and objectives were defined through discussions with the product portfolio and variety management experts. A methodical approach to predict the values of product attributes for multi-variant product portfolios was developed based on the phases of a data analysis process. The generic approach can therefore be easily applied to different product attributes. The introduced methodical approach was validated in terms of applicability and success by predicting weights of new vehicle configurations with a real data set from industry.

In section 2, a short introduction into the development of multi-variant product portfolios as well as machine learning is given. Moreover, the state of the art of machine learning tailored to the development of multi-variant product portfolios is shown. The main result of the paper is described in section 3. There, the methodical approach to predict values of product attributes for multi-variant product portfolios is introduced and validated afterwards in section 4 with real-world data from the commercial vehicle industry. Finally, the results are discussed critically, summed up and an outlook for further research activities is given.

2. State of the Art

2.1. Terminology

2.1.1. *Features, Components and Attributes*

In the development process, several perspectives on the product have to be taken into account. For this in the literature different terms are used. In the following, the views of customer features, product components and attributes are defined and differentiated from each other (Figure 1). Customer Features are selected by a customer when configuring a product (Braun, 2021). They form the target specifications of the product from a customer point of view. In industrial practice, these can be both characteristics and properties (Braun, 2021). The characteristics can be directly influenced or determined by the designers (e.g. structure, shape and material) (Weber, 2005). The properties describe the product's behaviour (e.g. performance, safety and reliability) and cannot be directly influenced (Weber, 2005).

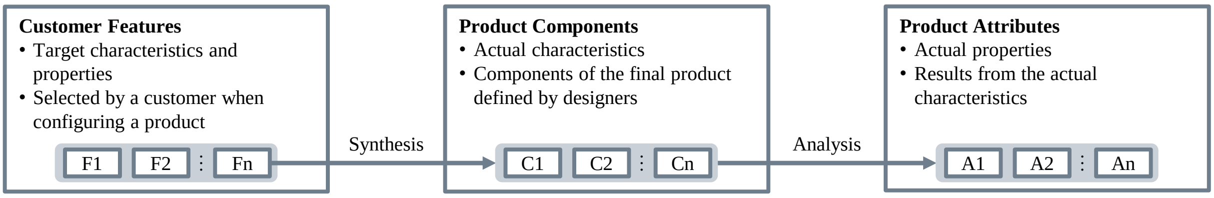


Figure 1. Features, components, attributes (adapted from Ponn, 2016)

The component structure includes all hardware and software components of the final product (Braun, 2021). In synthesis, the developers define the component structure of the product based on the required features (Ponn, 2016). The product attributes represent the actual properties of the final product. Based on the component structure, the product attributes can be determined by an analysis (estimation, simulation or test) (Ponn, 2016). This paper also distinguishes between technical product attributes (e.g. performance, weight) and non-technical product attributes (e.g. willingness to pay or country of sale).

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press

2.1.2. *Multi-variant Product Portfolios*

The product portfolio or product programme refers to the totality of all products and/or services that a company offers on the market (Jonas, 2013). Examples of multi-variant product portfolios are those of automobile or truck manufacturers with hundreds or thousands of features that can be chosen by customers in a configurator (Greisel *et al.*, 2013). To manage the variety of such multi-variant product portfolios, a generic product structure is necessary to document them in a manner that components become reusable (Kreimeyer *et al.*, 2016). In such a generic product structure, similar features (e.g. large Cab or small Cab) are grouped hierarchically in so-called feature categories (e.g. Cab). For a valid configuration, exactly one feature needs to be chosen for each feature category. The combinability is restricted due to technical and sales constraints formulated in configuration rules (e.g. four axles with a short frame). Moreover, the correct components within the component structure were selected based on the combination of features by component selection rules. Thus, the components are used in many different product configurations. The sum of all possible combinations between features forms the multi-variant product portfolio.

2.1.3. *Types of Machine Learning*

In machine learning, a basic distinction between three types of learning can be made. These are supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, data contain the desired solution, called labels, which are used to train and test the models (Géron, 2017). Common supervised learning techniques are classification and regression. Classification assigns a data sample to one of several predefined classes (Weiss and Kulikowski, 1991). In regression the relationship between a dependent variable and one or more independent variables are modelled (Backhaus *et al.*, 2016). Typical algorithms for supervised learning are e.g. linear regression or decision tree algorithm. In unsupervised learning the training data is unlabelled. The aim is to discover patterns and knowledge in the data (Murphy, 2012). Common techniques are clustering (e.g. k-means algorithm) and association rule learning (e.g. apriori algorithm). Clustering divides data points into groups or clusters so that objects in the same cluster are as similar as possible and objects from different clusters are as dissimilar as possible (Ester and Sander, 2000). Association analysis expresses rules about frequently occurring relationships in transaction data (Ester and Sander, 2000). Reinforcement learning differs from the other two learning types. It learns from an agent's direct interaction with its environment without relying on labelled examples or full models of the environment (Sutton and Barto, 2018).

2.1.4. *Process Modell for Machine Learning*

The implementation of the algorithms is only one step in the whole data analysis process. In this article the Cross Industry Standard Process for Data Mining (CRISP-DM) from Wirth and Hipp (2000) is used because it was developed by several industry representatives and has a strong focus on the industrial application of data science (Figure 2). Compared to other models like the Knowledge Discovery in Databases Process (see Fayyad *et al.*, 1996), it includes two understanding phases in the beginning. In the following the individual steps of the model are briefly described. The business understanding phase focuses on understanding the business goals of the data analysis project and their translation into a machine learning problem. In the second step, data understanding, the data is collected and examined in more detail to identify data quality issues, initial insights, and interesting subsets. Data preparation includes all the activities required to transform the raw data into the final dataset (e.g. data cleansing, feature engineering and data transformation). Modelling involves selecting, implementing, and comparing different algorithms to achieve optimal results. In the Evaluation phase, the models are assessed with regard to the predefined goals and a decision on their use is made. In the final phase, the deployment, the models and knowledge are made available to the users.

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press

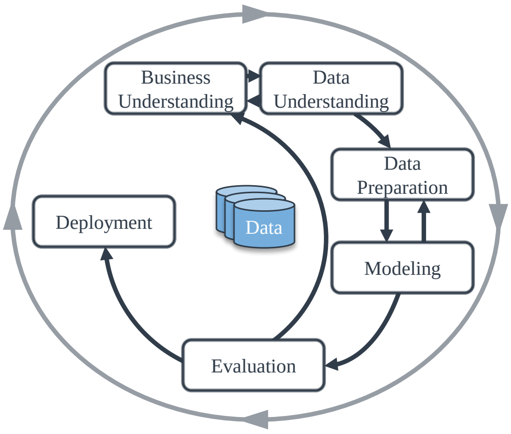


Figure 2. CRISP-DM (Wirth and Hipp, 2000)

2.2. Machine Learning in Development of Multi-variant Product Portfolios

A comprehensive review of the state of the art of data science in product portfolio and variety management is given in Mehlstäubl et al. (2021). In the following, the articles that attempt to predict values of technical or non-technical attributes of new products with the help of machine learning are presented.

In the literature, there are several articles that deal with the prediction of trends. Tucker and Kim (2011b) and Tucker (2014) forecast demand trends of individual product features from sold product configurations over time. To predict temporal evolvements, they use a time series analysis, which is a special form of regression. Ma *et al.* (2014) and Ma and Kim (2014) extend these approaches with an automatic optimisation process for model selection, parameter setting, and initial state estimation. Tucker and Kim (2011a) analyse online customer reviews to forecast demand trends. They also use a time series algorithm to model the expected trends of the individual customer features in the reviews. Ma and Kim (2016) analyse historical transaction information, which includes the selected product structure, its features, and the price for profit forecasting.

Besides the prediction of temporal trends, there are also initial approaches for the prediction of states and events. Tucker and Kim (2009) aim to assess the payment willingness for product features of the customers. They implement a decision tree algorithm to classify the product features of a survey dataset into multiple price categories. The decision tree makes it possible to determine the driving characteristics manually by analysing its structure. Boyarkin *et al.* (2019) describe the use of regression to predict the price of new feature combinations. For the regression, they train a neural network with features and the resulting prices of past product configurations.

The previous approaches do not describe a holistic and methodological procedure for the application under different conditions and to predict different technical and non-technical product attributes. Moreover, a comparison of different algorithms is missing and most of the approaches use synthetic data instead of real-world data from industry.

3. Methodical Approach to Predict Values of Product Attributes

An overview of a methodical approach to predict values of product attributes for multi-variant product portfolios is given in (Figure 3). It consists of five steps and is derived from the CRISP-DM. First the goals of the attribute prediction are defined (Section 3.1). This includes the attribute to be predicted and the specification of the added value as well as performance criteria. Second, the required sales data, its structure as well as its typical characteristics and challenges are described (Section 3.2). In the third step, the data is cleansed and transformed so that it can be processed by a machine learning algorithm (Section 3.3). Therefore, the required activities are explained. Subsequently, regression and classification algorithms are applied (Section 3.4). In this step, criteria for the selection and testing of algorithms are presented. Finally, the evaluation compares the results with the predefined goals (Section 3.5). The deployment phase is not considered in this approach.

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press

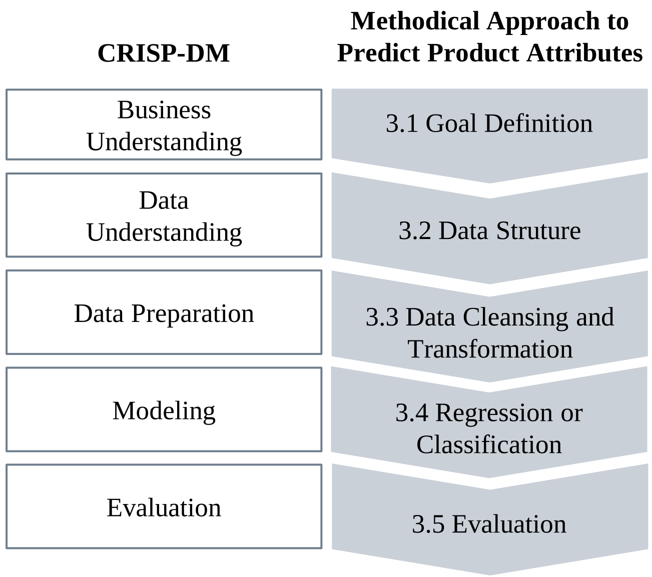


Figure 3. Overview of the methodical approach to predict product attributes for multi-variant product portfolios

3.1. Goal Definition

The overall goal of the approach is to predict the values of technical product attributes (e.g. product weight or length) or non-technical product attributes (e.g. cost or country of sales) by a configuration of customer features of the feature categories (e.g. model, axels or engine power) (see Figure 4). Depending on the kind of product attributes a regression or classification analysis is performed. If the product attribute is a continuous variable (e.g. weight in kg) a regression analysis is selected to model the relationship between the feature configuration and the values of the product attribute. If it is a categorical variable (e.g. country of sales) a classification analysis is chosen to assign the feature configuration to one of the predefined classes. In order to specify the project goals, the current procedure has to be analysed and the added value of the use of machine learning has to be identified. In this context, machine learning enables the automated derivation of correlations between customer features (input) and product attribute values (output) without the need to identify the respective components in between, as shown in (Figure 4). This reduces the effort compared to a manual rule definition. Moreover, statistical models can approximate the correlations directly on the real-world data, rather than using abstracted pre-defined rules.

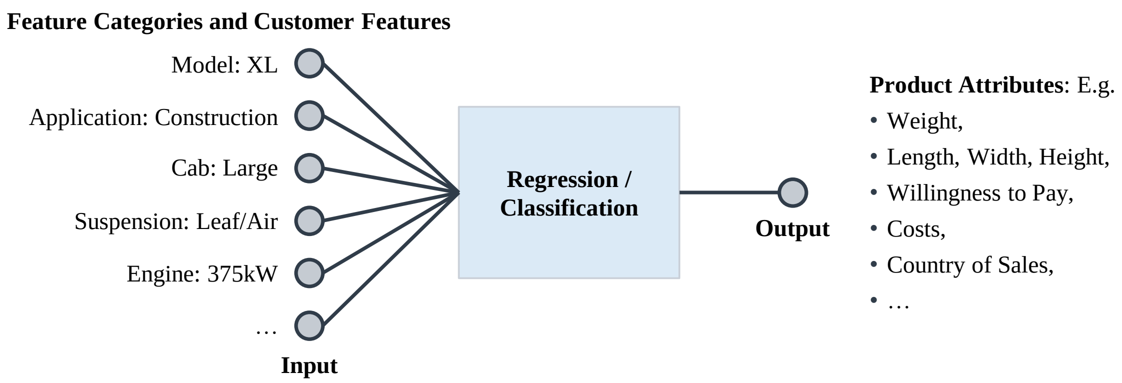


Figure 4. Basic architecture of the model

3.2. Data Structure

Sales data is used for the development and evaluation of the machine learning model. The structure of a sales data set can be found in (Table 1). It is usually stored in a filing system (e.g. an Information

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press

Warehouse) after the order generation in the primary order processing system and can be exported from there. In the order processing system, every product consists of a unique ID, a feature structure as well as technical and non-technical attributes. Today, the selection of features is usually made and documented with a sales configurator. The product attributes of the sold configurations are documented after the sales process (e.g. final price) or production process (e.g. exact weight or height) is completed. The product structure in companies is in constant change due to new requirements. New customer features and entire feature categories are added, and old ones are removed from the portfolio. This leads to many missing values in the data. Also in some categories, only one feature has been sold so far and therefore, they are constant and do not provide any value for determining the product attribute.

Table 1. Exemplary sales data structure

	Cutomer Features					Product Attribute
Product ID	**Model**	**Application**	**Cab**	**Suspension**	**Engine**	**Weight**
1	XL	Construction	Large	Leaf/Air	375 kW	10 000 kg
2	S	Construction	Medium	Leaf/Leaf	294 kW	6 000 kg
3	M	Construction	Small	Leaf/Air	213 kW	8 000 kg
4	M	Beverages	Small	Air/Air	235 kW	7 500 kg
5	L	Distribution	Small	Leaf/Air	140 kW	8 500 kg

3.3. Data Cleansing and Transformation

In order to perform machine learning algorithms, the data needs to be cleansed and transformed. In this case the cleansing includes deleting columns with constant values, as these do not give any information about the composition of the product attribute, and those with many missing values. In the next step the data must be transformed through encoding. Since there is no ordinal relationship between the features of the most feature categories, a one-hot encoding is used in this method. This converts the data into a sparse matrix, where each column corresponds to a possible value of a customer feature (see Table 2).

Table 2. Exemplary encoded sales data

ID	S	M	L	XL	Construction	Beverages	Distribution	Large	Medium	Small	…	Weight
1	0	0	0	1	1	0	0	1	0	0	…	10 000 kg
2	1	0	0	0	1	0	0	0	1	0	…	6 000 kg
3	0	1	0	0	1	0	0	0	0	1	…	8 000 kg
4	0	1	0	0	0	1	0	0	0	1	…	7 500 kg
5	0	1	0	0	0	0	1	0	0	1	…	8 500 kg

3.4. Regression or Classification

In the next step, a suitable machine learning algorithm must be selected depending on various factors and trained with the training data as well as tested with the test data. When splitting the data into training and test data, it is advisable to split it between 70% / 30% and 95% / 5% training / test data depending on the amount of data (Burkov, 2019). When selecting algorithms, a variety of factors play a role in the prediction accuracy of the models. One of the most important factors is the number of data samples. Some algorithms can detect correlations with just a few data samples. Others are more suitable for a large amount of data. The number of features and the degree of linearity between the input features and the value of the attribute also play an important role in the selection of the right algorithm. In many applications of machine learning the training time is a critical factor. However, it plays a subordinate role in this approach because the model does not need to be re-trained before or parallel to use. In the prediction of the values of product attributes the prediction time is significant. It must be carried out in real-time during the product configuration process. Another relevant factor in the industrial context is

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press

transparency. For acceptance, it can be important to understand the decisions of the algorithm. This also makes it possible to understand the models and gain additional knowledge about patterns. An overview of the parameters and common supervised learning algorithms is given in (Table 3). For the implementation, it is advisable to train several promising algorithms and compare their accuracy with statistical criteria based on the predictions with the test data. Criteria with which the machine learning models are trained are primarily the mean squared error (MSE) and r-squared (R2). The mean absolute error (MAE) and mean absolute percentage error (MAPE) are recommended for the clear communication of results. The algorithm with the best accuracy is selected and evaluated in the next step.

Table 3. Factors for algorithm selection

	Number of Data Points	Non-linearity	Number of Features	Prediction Time	Training Time	Algorithm Transparency
Linear Regression	very low	very low	low	very low	high	very high
Support Vector Machine	low	low	low	very high	very high	medium
K-nearest Neighbors	medium	medium	medium	medium	very low	medium
Decision Tree	medium	medium	high	very low	medium	high
Random Forest	high	high	high	very low	high	high
Neural Network	very high	very high	very high	low	high	very low

3.5. Evaluation

In the evaluation, the models are compared to the goals defined in the first step and a comparison is covered with the current methods. For the prediction of values of attributes, the factors accuracy and prediction time are particularly important. In this step, the results are presented to the users and, if possible, tested in the real user environment. Depending on the type of algorithm, the decisions should be made as transparent as possible in order to increase the understanding and acceptance of the users. For example, the visualisation of decision trees or the determination of the feature importance offers a good possibility to check the plausibility of the results. As a result of this step, a decision has to be made whether and how the model should be used in the future.

4. Case Study: Prediction of Weights for Commercial Vehicles

The applicability and benefits of the methodical approach to predict values of product attributes for multi-variant product portfolios were evaluated together with a partner from the commercial vehicle industry using the example of vehicle weights. The expected benefits of using machine learning is to predict weights more accurately as well as faster as before and to automate the time-consuming and expensive rule definition process. Previously, the weights have been calculated based on expert rules formulated by a team of engineers over several months. With the rule-based expert system the company has an accuracy of about 97% and the calculation time is between one and two seconds in the real-world application.

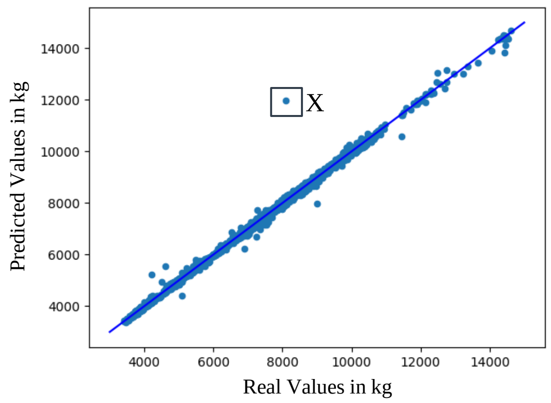
For the implementation a data set with 35 091 sold vehicle configurations was used. The vehicles were configured by the customer in a sales configurator and weighed after assembly. The product structure consisted of a total of 1 097 customer features, which were reduced to 886 features in the data preparation due to missing (> 95%) and constant values. After performing the one-hot encoding, different regression algorithms were applied to train and test the model. Since a large number of data samples and features are available and a high non-linearity is expected, the neural network, random forest and decision tree regression algorithms are particularly suitable in this case, as examined in (Table 3). However, linear regression, support vector machine and k-nearest neighbours were also taken into

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press

account to prove the factors for the selection of algorithms described before. The implementation was conducted with Python and the libraries scikit-learn as well as for the neural network Keras, which is a high-level API for TensorFlow 2. As the loss function, the squared error was used and the split between training and test data was chosen as 90% to 10% because of the high number of data samples. The results of applying the evaluation criteria to the predictions of the test data are shown in (Table 4). Linear regression and support vector machine provide inappropriate predictions as expected, because they cannot map the non-linear relationships between the different feature combinations and the vehicle weights. The other four algorithms deliver promising results. Above all, the neural network achieved with an MAE of 43 kg the best results. The data included vehicles from under 4 000 kg up to over 14 000 kg. This corresponds to a MAPE of just 0,6% and a prediction time of 1,206 x 10^{-4} s. This reduces the error compared to the current rule-based expert system by 80% (0,6% vs. 3%) and the calculation time by more than 99% (1,206 x 10^{-4} s vs. 1-2 s).
In (Figure 5) the predictions from the neural network are visualised. Almost all predictions are in line with the real measured values in the data. However, there are also some outliers. Especially the vehicle X has a predicted weight of 11 971 and a weighed value of 8 105 which corresponds to a deviation of 3 866 kg. Through an analysis of the configuration, it became apparent that there is an error in the original weighed data. The vehicle is a heavy-duty truck with four driven axles. Furthermore, an identical configuration ordered from the same customer with a subsequent ID could be identified, which had an actual weight of 11 995 kg.

Table 4. Results case study

	Linear Regression	Support Vector Machine	K-nearest Neighbors	Decision Tree	Random Forest	Neural Network
Expected Suitability	very low	low	medium	high	very high	very high
MSE	2,347 x 10^{23}	1540507	28751	17018	13456	9477
MAE	4,35 x 10^{10}	745	81	59	53	43
MAPE	5,48 x 10^{10} %	11,74%	1,10%	0,72%	0,81%	0,6%
R2	2,552 x 10^{-5}	0,405	0,989	0,993	0,995	0,996
Training Time	283 s	4776 s	0,362 s	34 s	96 s	200 s
Prediction Time	7,806 x 10^{-5} s	0,358 s	4,61 x 10^{-3} s	2,814 x 10^{-5} s	3,628 x 10^{-5} s	1,206 x 10^{-4} s



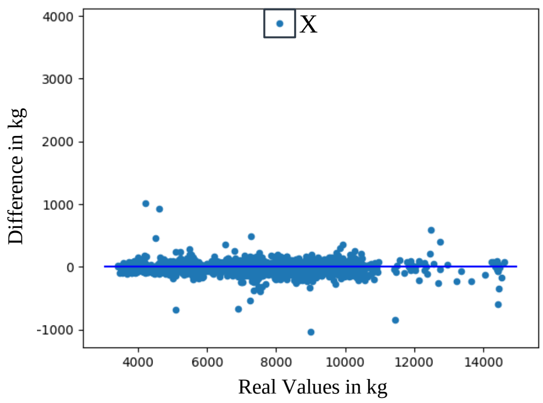


Figure 5. Visualisation of the Predictions from the Neural Network

5. Discussion

The methodical approach enables the prediction of product attribute values of multi-variant product portfolios based on customer feature configurations with machine learning and thereby answers the first research question raised at the beginning. Furthermore, the second research question is answered. Machine learning enables in this application the early prediction of the actual values of product attributes during the configuration process and reduces the effort to a minimum compared to designing rule-based

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press

expert systems. Moreover, the generated machine learning models have both, higher prediction accuracy and speed. However, machine learning also has disadvantages. For example, although some machine learning models are comprehensible, they are not as simple as rules written by humans. In addition, machine learning models are based on statistical considerations. This makes it more difficult compared to rule-based systems to take outliers in the product portfolio into account. The developed approach can be applied to continuous or categorical and technical or non-technical product attributes. The applicability and success were confirmed by the implementation at an industrial partner from the commercial vehicle sector. As the results are superior to the established processes, the deployment of a software tool and the introduction into the business processes are currently being worked on with the industrial partner for the practical, daily application. However, the data and the underlying product model of only one industry partner has been analysed so far. Moreover, only continuous and technical attributes were predicted with a regression. An application to categorical and non-linear product attributes is in progress.

6. Conclusion and Outlook

Industrial companies offer their customers an almost endless number of product variants to satisfy their needs in the best possible way. Although an identical feature configuration was not built before, the attribute values must be determined already during the configuration process. This paper introduces a new methodical approach to predict values of product attributes from combinations of customer features by using machine learning. First, the objectives are defined and an understanding of the underlying product structure, as well as the data structure, is created. The data is then transformed and encoded. An application-specific evaluation of the regression and classification algorithms enables their target-oriented selection. The developed approach was applied to a real-world data set from the commercial vehicle industry and its success was confirmed. As attribute values, the vehicle weights were predicted. The best results were achieved with a neural network. Compared to rule-based expert systems, the effort, time, and accuracy of attribute prediction could be improved compared to rule-based expert systems. Future research activities include the application to categorical and non-technical product attributes with classification and the implementation as well as evaluation with the products of further companies. Furthermore, the consideration of the deployment phase and the related integration in a software tool is ongoing.

Acknowledgement

This research work is part of "FORCuDE@BEV - Bavarian research association for customized digital engineering for bavarian SME's" and funded by the "Bayerische Forschungsstiftung (BFS)". We direct special thanks to the Bayerische Forschungsstiftung (BFS) for financial support of the whole research project.

References

Agard, B. and Kusiak, A. (2004), “Data-mining-based methodology for the design of product families”, International Journal of Production Research, Vol. 42 No. 15, pp. 2955–2969.

Backhaus, K., Erichson, B., Plinke, W. and Weiber, R. (2016), Multivariate Analysemethoden, Springer Berlin Heidelberg, Berlin, Heidelberg, available at:https://doi.org/10.1007/978-3-662-46076-4.

Bertoni, A., Larsson, T., Larsson, J. and Elfsberg, J. (2017), “Mining data to design value: A demonstrator in early design”, Proceedings of the International Conference on Engineering Design, ICED, Vol. 7 No. DS87-7, pp. 21–29.

Boyarkin, G.N., Revina, I. V. and Shevelyova, O.G. (2019), “Forecasting the Price of Products Using Neural Networks”, IOP Conference Series: Materials Science and Engineering, Vol. 582 No. 1, available at:https://doi.org/10.1088/1757-899X/582/1/012005.

Braun, F. (2021), “Application of algorithm-based validation tools for the validation of complex, multi-variant products”.

Burkov, A. (2019), The Hundred-Page Machine Learning Book, Vol. 1, Andriy Burkov Canada.

ElMaraghy, H., Schuh, G., Elmaraghy, W., Piller, F., Schönsleben, P., Tseng, M. and Bernard, A. (2013), “Product variety management”, CIRP Annals - Manufacturing Technology, Vol. 62 No. 2, pp. 629–652.

Ester, M. and Sander, J. (2000), Knowledge Discovery in Databases, Knowledge Discovery in Databases, Springer Berlin Heidelberg, Berlin, Heidelberg, available at:https://doi.org/10.1007/978-3-642-58331-5.

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996), “From data mining to knowledge discovery in databases”, AI Magazine, Vol. 17 No. 3, pp. 37–53.
Géron, A. (2017), Hands-on Machine Learning with Scikit-Learn and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems, O’Reilly Media.
Greisel, M., Kissel, M., Spinola, B. and Kreimeyer, M. (2013), “Design for adaptability in multi-variant product families”, Proceedings of the International Conference on Engineering Design, ICED, Vol. 4 DS75-04 No. August, pp. 179–188.
Haug, A., Hvam, L. and Mortensen, N.H. (2012), “Definition and evaluation of product configurator development strategies”, Computers in Industry, Elsevier B.V., Vol. 63 No. 5, pp. 471–481.
Hochdörffer, J., Laule, C. and Lanza, G. (2018), “Product variety management using data-mining methods - Reducing planning complexity by applying clustering analysis on product portfolios”, IEEE International Conference on Industrial Engineering and Engineering Management, Vol. 2017-Decem, pp. 593–597.
Hu, S.J., Zhu, X., Wang, H. and Koren, Y. (2008), “Product variety and manufacturing complexity in assembly systems and supply chains”, CIRP Annals - Manufacturing Technology, Vol. 57 No. 1, pp. 45–48.
Jonas, H. (2013), Eine Methode Zur Strategischen Planung Modularer Produktprogramme, Technische Universität Hamburg-Harburg.
Kreimeyer, M., Baumberger, C., Deubzer, F. and Ziethen, D. (2016), “An integrated product information model for variant design in commercial vehicle development”, Proceedings of International Design Conference, DESIGN, Vol. DS 84 No. 1, pp. 707–716.
Kusiak, A., Smith, M.R. and Song, Z. (2007), “Planning product configurations based on sales data”, IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, Vol. 37 No. 4, pp. 602–609.
Ma, J. and Kim, H.M. (2014), “Continuous preference trend mining for optimal product design with multiple profit cycles”, Journal of Mechanical Design, Transactions of the ASME, Vol. 136 No. 6, pp. 1–14.
Ma, J. and Kim, H.M. (2016), “Product family architecture design with predictive, data-driven product family design method”, Research in Engineering Design, Springer London, Vol. 27 No. 1, pp. 5–21.
Ma, J., Kwak, M. and Kim, H.M. (2014), “Demand trend mining for predictive life cycle design”, Journal of Cleaner Production, Elsevier Ltd, Vol. 68, pp. 189–199.
Mehlstäubl, J., Braun, F. and Paetzold, K. (2021a), “Data Mining in Product Portfolio and Variety Management – Literature Review on Use Cases and Research Potentials”, 2021 IEEE Technology & Engineering Management Conference-Europe (TEMSCON-EUR), pp. 442–447.
Mehlstäubl, J., Nicklas, S., Gerschütz, B., Sprogies, N., Schleich, B., Lohner, T., Wartzack, S., et al. (2021b), “Voraussetzungen für den Einsatz datengetriebener Methoden in der Produktentwicklung”, Proceedings of the 32nd Symposium Design for X (DFX2021).
Murphy, K.P. (2012), Machine Learning: A Probabilistic Perspective, MIT press, available at:https://doi.org/10.1109/pes.2005.1489456.
Ponn, J. (2016), “Systematisierung des Lösungsraums”, Handbuch Produktentwicklung, Carl Hanser Verlag GmbH & Co. KG München, pp. 715–742.
Schuh, G., Riesener, M. and Jank, M.-H. (2018), “Managing Customized and Profitable Product Portfolios Using Advanced Analytics”, Customization 4.0, pp. 203–216.
Sutton, R.S. and Barto, A.G. (2018), Reinforcement Learning: An Introduction, MIT press.
Tucker, C. and Kim, H.M. (2011a), “Predicting emerging product design trend by mining publicly available customer review data”, ICED 11 - 18th International Conference on Engineering Design - Impacting Society Through Engineering Design, Vol. 6 No. August, pp. 43–52.
Tucker, C.S. (2014), “Quantifying the relevance of product feature classification in product family design”, Advances in Product Family and Product Platform Design: Methods and Applications, pp. 147–177.
Tucker, C.S. and Kim, H.M. (2009), “Data-driven decision tree classification for product portfolio design optimization”, Journal of Computing and Information Science in Engineering, Vol. 9 No. 4, pp. 1–14.
Tucker, C.S. and Kim, H.M. (2011b), “Trend mining for predictive product design”, Journal of Mechanical Design, Transactions of the ASME, Vol. 133 No. 11, pp. 1–11.
Weber, C. (2005), “CPM/PDD–an extended theoretical approach to modelling products and product development processes”, Proceedings of the 2nd German-Israeli Symposium on Advances in Methods and Systems for Development of Products and Processes, Vol. 6, Fraunhofer-IRB-Verlag Stuttgart.
Weiss, S.M. and Kulikowski, C.A. (1991), Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems, Morgan Kaufmann Publishers Inc.
Wirth, R. and Hipp, J. (2000), “CRISP-DM : Towards a Standard Process Model for Data Mining”, Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, Vol. 1, Springer-Verlag London, UK, pp. 29–39.

https://doi.org/10.1017/pds.2022.168 Published online by Cambridge University Press