Technische Hochschule Brandenburg
University of Applied Sciences
Fachbereich Informatik und Medien

Interpretable Machine Learning in Procurement: Analysis and Modeling of Sourcing Decisions using Decision Trees

Franziska Gohlicke

Bachelor Thesis • Studiengang Informatik • Fachbereich Informatik und Medien • 09.09.2025

Objective

The objective of the thesis is to provide some groundwork for automazing Sourcing Decisions (SD) with the possibility of human intervention. The thesis assess if it is possible to model the SD using the available corporate data of the Robert Bosch GmbH and machine learning (ML) algorithms.
The model should be interpretable, so that it is possible to understand the models‘ structure and predictions. The models trained are classification Decision Trees (DT).

Methodology

The approach taken corresponds to standard Data Science procedures, as shown in Fig. 1.

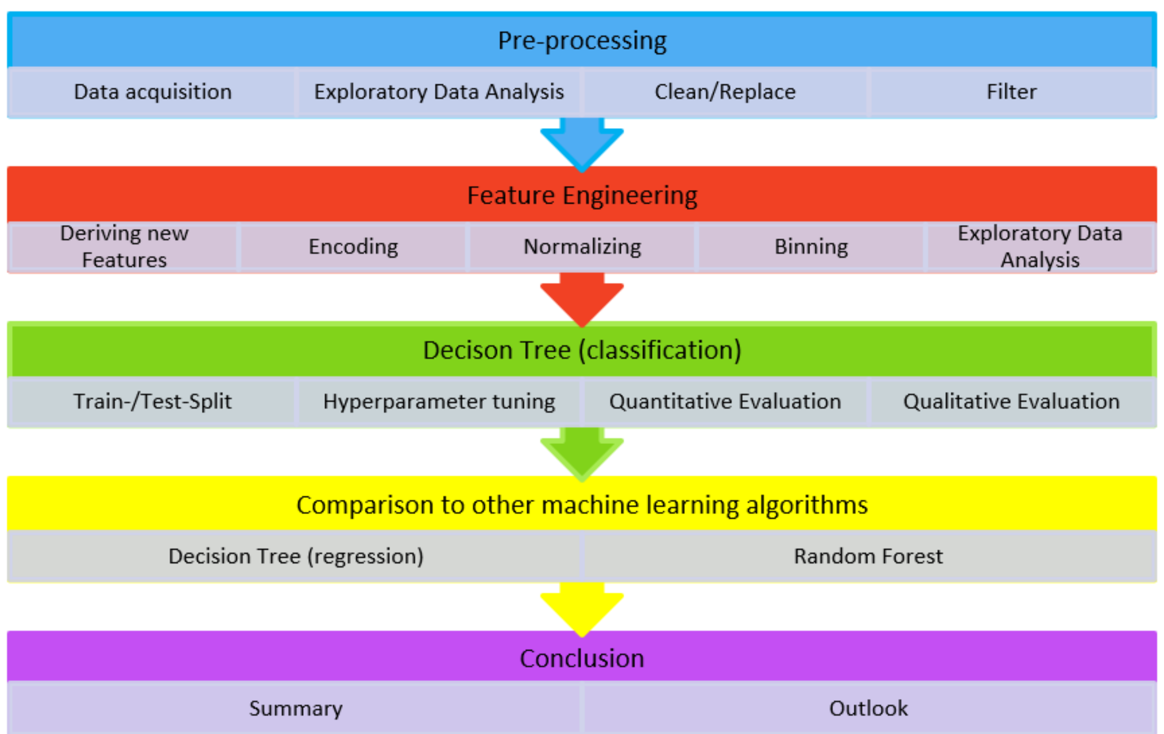

Fig. 1: Overview of the steps taken.

The raw data was first analyzed to get an understanding of the structure and quality. Problems, such as missing or erroneous values were uncovered and handled. After that, new features were derived from the existing ones. Then the categorical features were encoded to become numeric features, and the numeric features were normalized to fall between -1 and 1. The target feature were linearly binned, so that 5 five classes were created which the model later predicts.
Different setups for the models were tested and compared, to find the optimal setups. Those classification DTs were also compared to regression DTs and Random Forest (RF) models, to assess if the general approach if the best, or if other ML models are better suited for this problem.

Train/Test Split

The processed the data was split into a training and a testing set. The training set was used to train the models, and the testing set was used to evaluate the models’ performance on new unseen data. The resulting class distributions is very similar to the entire dataset, as shown in Fig. 2. This means that the subsets are representative of the entire data.

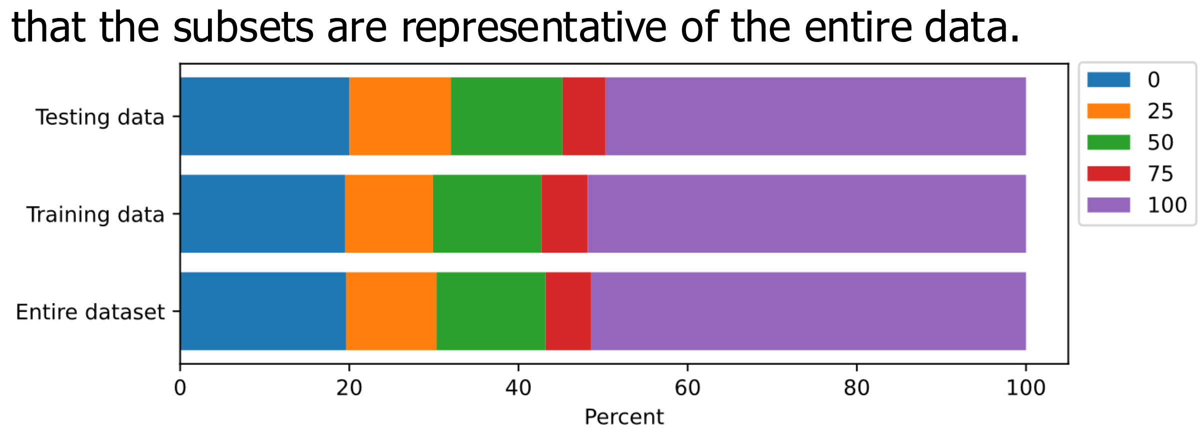

Fig. 2: Class distributions of the training, testing and entire dataset.

Hyperparameter tuning

Using gridsearch, multiple setups were tested and their accuracies averaged to see, which splitting criterion performs better. Fig. 3 shows that gini works the best.

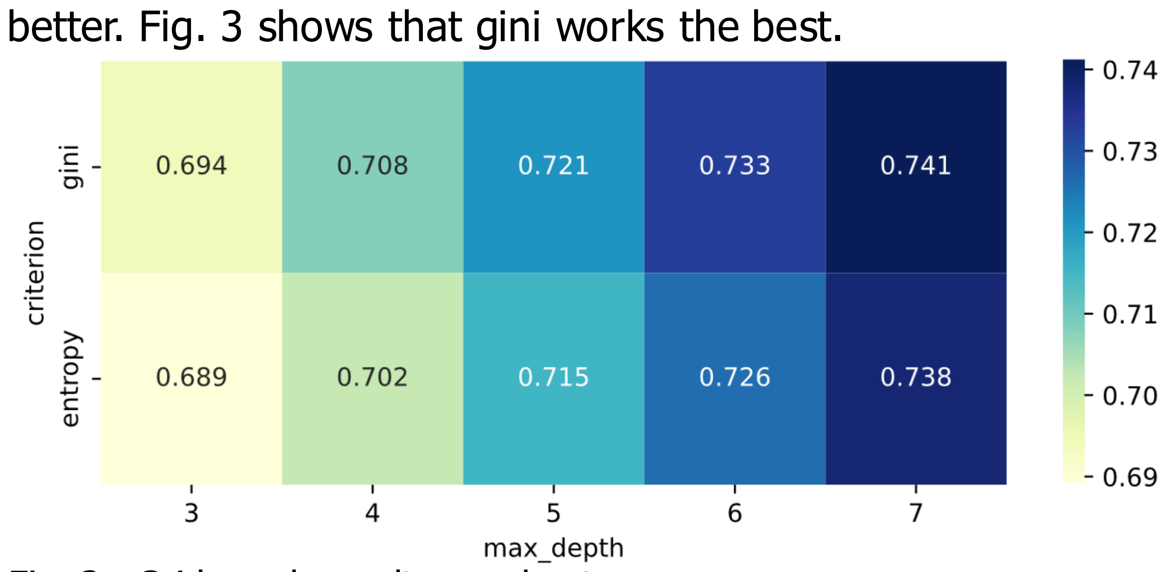

Fig. 3: Gridsearch results as a heatmap.

For finding the optimal depth, the accuracy for the training and the testing data for different depths are analyzed. The results shown in Fig. 4 shows the gap between the accuracies widening from depth 8 and forward, indicating overfitting. Because of this, the max. depth of 7 seems to be optimal.

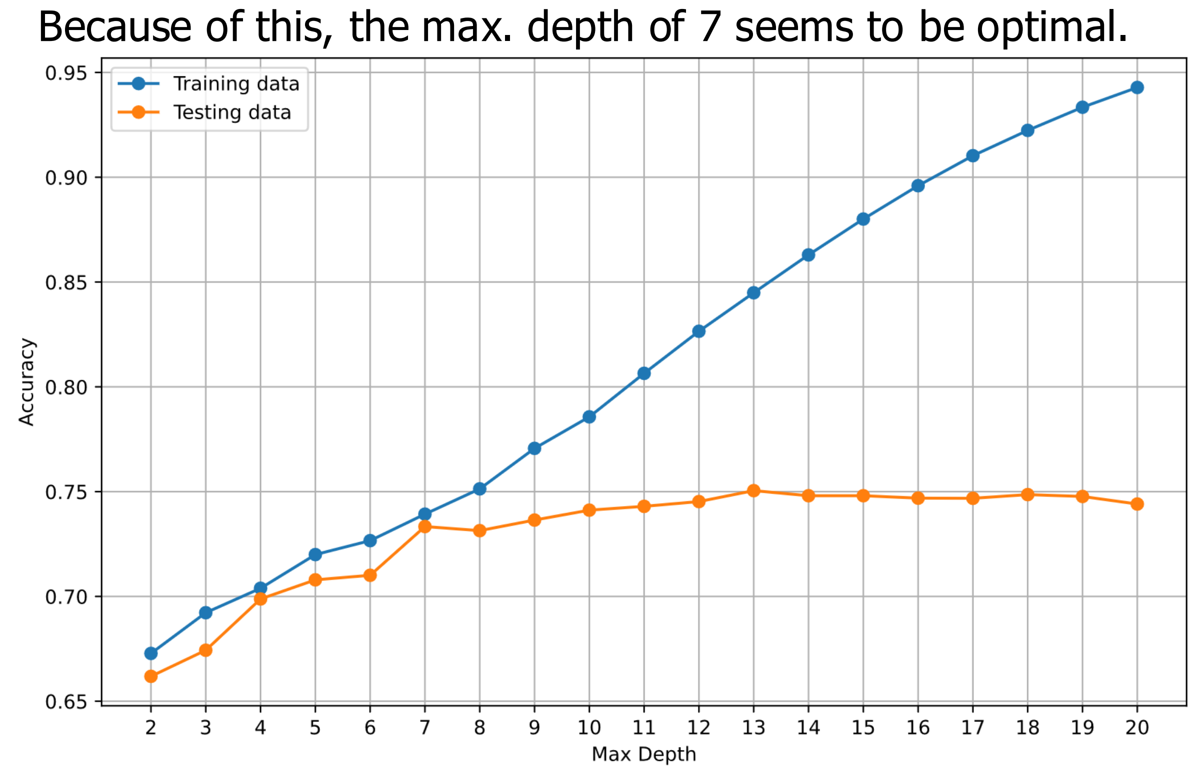

Fig.4: Accuracy vs max. depth for training and testing data

This optimal model was assessed by domain experts, who agreed that the model has potential to be used for normal cases. However, there were some aspects for special cases missing that are either not in the data or were in the data but the model does not make use of.

Comparison to other ML algorithms

Additionally, regression DTs and RF models were trained to assess, if they work better than classification DTs. The regression DTs and RFs predictions were slightly more accurate, but the improvements were not significant enough to outweigh the simplification of the problem and the interpretability of the classification DTs.

Conclusion

This thesis proved that it is indeed possible to model the SD using the available data and ML. The model can handle normal cases, but for special cases, it does not work well.

Betreuung Prof. Dr. Emanuel Kitzelmann• Technische Hochschule Brandenburg • Marc-Alexander Frey • Robert Bosch GmbH