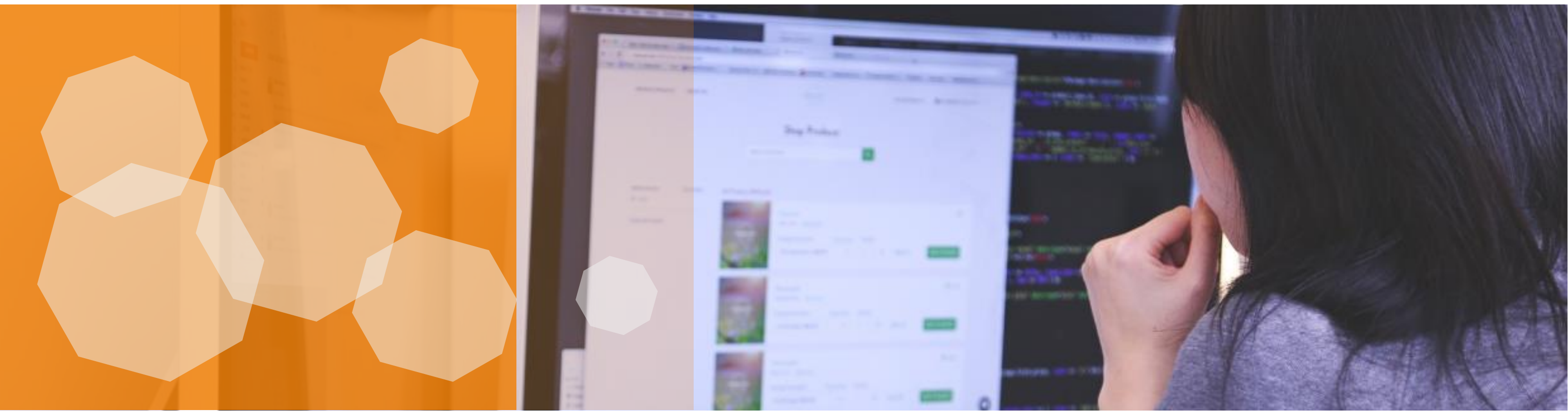

Prediction of Test Bench Times for Varying Automotive Configurations through Machine Learning

Project Goal

Before vehicles leave the factory, automobile manufacturers have to ensure their safety and reliability. To do this they conduct tests on test benches. There are massive amounts of different car configurations. These configurations result in different test durations on the test benches. In the automotive industry, test benches are critical bottleneck resources where simultaneous assignments and idling must be avoided at all costs.

The aim of the project was therefore to develop a model that precisely predicts the test time a vehicle will spend on the test bench depending on its configuration, so that scheduling of test bench allocations can be optimized.

Provided Data

The training and test data set each consisted of 4,200 vehicle configurations and 400 features. The training data set also included the test times of each vehicle configuration. Test times were missing in the test data set.

Applied Methods

As the data sets contained a relatively high number of features compared to the amount of vehicle configurations, the 400 features needed to be grouped and reduced, so that the same prediction could be made with fewer features. Therefore, the data analysis technique called Dimensionality Reduction, with the best predictive result for this case needed to be selected. To do so, two methods were examined in detail.

The first was the Multiple Correspondence Analysis (MCA), usually used on categorical data, in order to analyze the structure in the data records and calculate the significance of each feature towards test bench times. Through application of this algorithm it was possible to reduce the 400 features to the 14 most important ones with little loss of information. The second Dimension Reduction technique tested was the Principal Component Analysis (PCA), which was only able to reduce the number of features to a total of 24. As the MCA performed noticeably better, it was used as the basis for the Machine Learning model that was developed subsequently.

Using the reduced data set different machine learning algorithms were compared, including different regression algorithms. Cross validation was used as validation technique of the tested models on the training data. As test times were known for this data, this can be used to confirm prediction accuracy of the respective models. In the end, the Gradient Boosting Regressor was chosen as it provided the best results. After parameter tuning of the model, it was ready to make predictions on the original test data set.

Challenges

The first challenge was that relatively few data points were available for the comparatively large number of features, making a Dimension Reduction necessary to determine the features with high impact on the resulting test bench times.

The second challenge was the choice of the most performant regression algorithm as algorithms applicable for the provided data are similar in principle.

Project Outcome

The Machine Learning Model that has been developed made it possible to predict test bench times for all vehicle configurations. This allows test cycles to be run more efficiently in the future to better optimize both allocation and timing.

Through the use of the previously mentioned dimensionality reduction, it has also been possible to identify the most essential features influencing test time duration of a vehicle.