supper&supper

Predictive Maintenance - Predicting Failures in Pneumatic Systems of Trucks using Sensor Data

Project Goal

The pneumatic system is a sensitive centerpiece of every truck. Being responsible to generate pressurized air, that is utilized for functions such as braking and gear changes, a defect can cause dramatic and cost -intensive consequences. This also results in high maintenance costs for the pneumatic system with the purpose to prevent those defects.

A lot of sensors register the current conditions for numerous components in the trucks. The aim of the project was to utilize this data in a machine learning model to prevent breakdowns, caused by the pneumatic system. The model was supposed to predict, if a failure of the pneumatic system will happen, to initialize the correct maintenance measures in time. With this approach unnecessary breakdowns and maintenance operations can be avoided. An important factor for the success of this project is a good prediction for failures as well as a insignificant small number of false alarms. With a successful implementation a high amount of cost savings can be realized.

Provided Data

The data provided in this project consisted of two datasets (training and test set). The training data set contained sensor data of 60,000 trucks, including 1,000 trucks that have been labelled with a defect of the pneumatic system. The test data set contained sensor data of 16,000 trucks, where a defect of the pneumatic system had to be predicted. For each truck 171 features, received by the sensors, were recorded.

Challenges

First of all the datasets contained many missing values in their features. A decision about handling these values has to be taken.

Different types of wrong predictions have different weightings. Predicting an pneumatic system not to be broken, when it actually is (“False-Negatives”) can lead to a breakdown on the road. When an pneumatic system is predicted as broken and it is not (“False-Positives”) costs arise by applying unnecessary check ups. Comparing the costs produced by a breakdown to the costs of an unnecessary check up the breakdown costs are way more expensive, so the algorithm had to prioritize while it was trained.

The proportions between positive (broken pneumatic system) and negative (working pneumatic system) label (1,000 : 59,000) were unbalanced. To achieve a high precision on the True-Positives (we predict the pneumatic system is broken and it is actually broken) and to reduce the False-Negatives (we predict the pneumatic system is working and it is actually broken) we needed to balance the training data and to adapt the algorithm.

Applied Methods

Missing values were replaced by the column mean. Like this the lines with missing values did not distort the model. As the values of the features had a wide range and include very high numbers, a normalization method was performed. This method transforms all values into numbers between -1 and +1. This supported the performance of the applied models. To prevent negative effects caused by the dissimilar distribution of positive and negative labels, bootstrapping and class weightings were applied to the data. Like this the high importance for failure detections was incorporated.

In the process of model selection several algorithms were evaluated according to their predictive performance. This was achieved by finetuning the model parameters with grid search methods. The models considered were Logistic Regression, Random Forest Classifier, Support Vector Classifier, Quadratic Discriminant Analysis and Neural Networks.

The Support Vector Classifier received a very good overall precision of 98.9%. Even though it causes almost no false alarms, it missed a substantial number of failures - which is the critical factor. Like this it was not suitable for a practical application. The best performing algorithm to decrease the number of missed detections was the Quadratic Discriminant Analysis. Even though the overall precision was only at 96.1%, it was able to detect 92.3% of all failures and only had a small amount of false alarms (3.8%).

Project Outcome

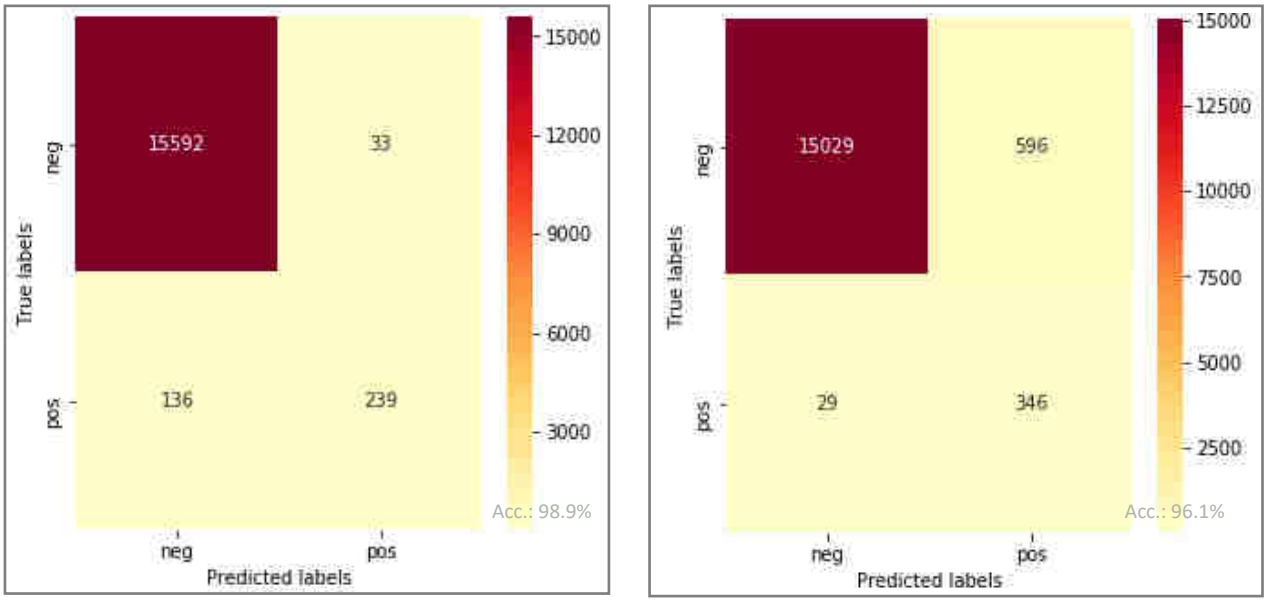
The trained model reached a good precision to predict the failures (92.3%) and had a decent number of false alarms (3.8%). Since the false alarms are not too harmful the results are highly acceptable and provide a good model, that is ready to use. It is not intended to replace the mechanics with their technical knowledge, but it is meant to assist them and discover cases that would probably have caused breakdowns, if left undiscovered.

Further applications

The model can be extended with a process control framework to detect anomalies in real time to get automated warnings for specific sensors. This approach can then be implemented as a rapid alert system in any vehicle.

As the provided dataset was rather small, more training data can be used to increase the prediction precision. Adding more sensors and additional data could also help to improve the model.

Predictions for other truck components are a possible scenario as well. An adapted version of the used sensor data set can be used to accomplish that.



Prediction outcomes for the Support Vector Classifier (left) and the Quadratic Discriminant Analysis (right)