Credit: Scania AB

Minimizing Costs using Predictive Maintenance

—

Sebastian R.
2021

Overview

Scania, a heavy haulage truck manufacturer, wants to reduce costs by minimizing the failure rate of air pressure systems in the field. The goal is predicting a failure before it occurs, and performing cost effective preventative maintenance, rather than expensive rescue and repair in the field. The compressed air supply is a critical component of any cargo transport vehicle, as it is responsible for brakes, transmission gear changes, and suspension, and by performing selective, preventative maintenance could save a great deal of both money and time.

Goals

1. Predict an imminent failure of the Air Pressure System from sensor data.
2. Reduce costs by both selective, preventative maintenance to avoid a failure in the field, and avoiding unnecessary loss of man hours.

Specifications

Scania released a set of 76000 data points. Each data point consists of 170 features of operational data taken from various sensors, but data is completely anonymized for proprietary reasons. There are 1000 confirmed failures in the data specifically in the Air Pressure System (APS) of a given truck.

Data Structure

The data is all in numeric format, except for the 'class' feature, being either 'pos' representing failure, or 'neg'. Due to the data being sanitized, feature selection in this case becomes a challenge. Several columns have only 20% of the data complete.

Column	Missing
br_000	49264
bq_000	48722
bp_000	47740
bo_000	46333

Metrics

I. Classification

Scania wants the results of the prediction model presented in terms of costs in dollars using the following metric:

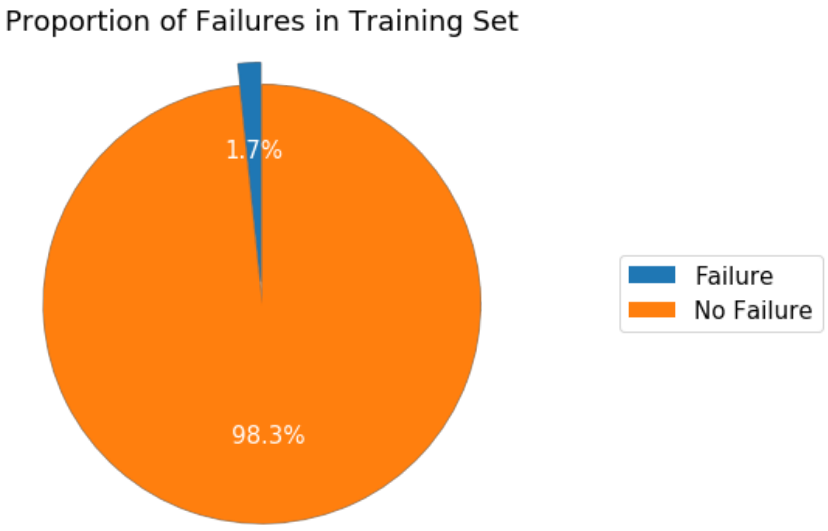
- ➔ Trucks incorrectly predicted to have an APS failure will cost $10, representing unnecessary maintenance. (Cost 1)
- ➔ A missed prediction on a faulty truck will cost $500, representing a possible failure in the field. (Cost 2)

II. Results as Costs

The predictions are presented directly as costs, and not just an accuracy score. Faulty trucks with imminent failures that are incorrectly predicted as perfectly functional have a cost rating 50 times higher than trucks that receive maintenance that isn't required.

Total Cost = (Number of Cost 1 * 10) + (Number of Cost 2 * 500)

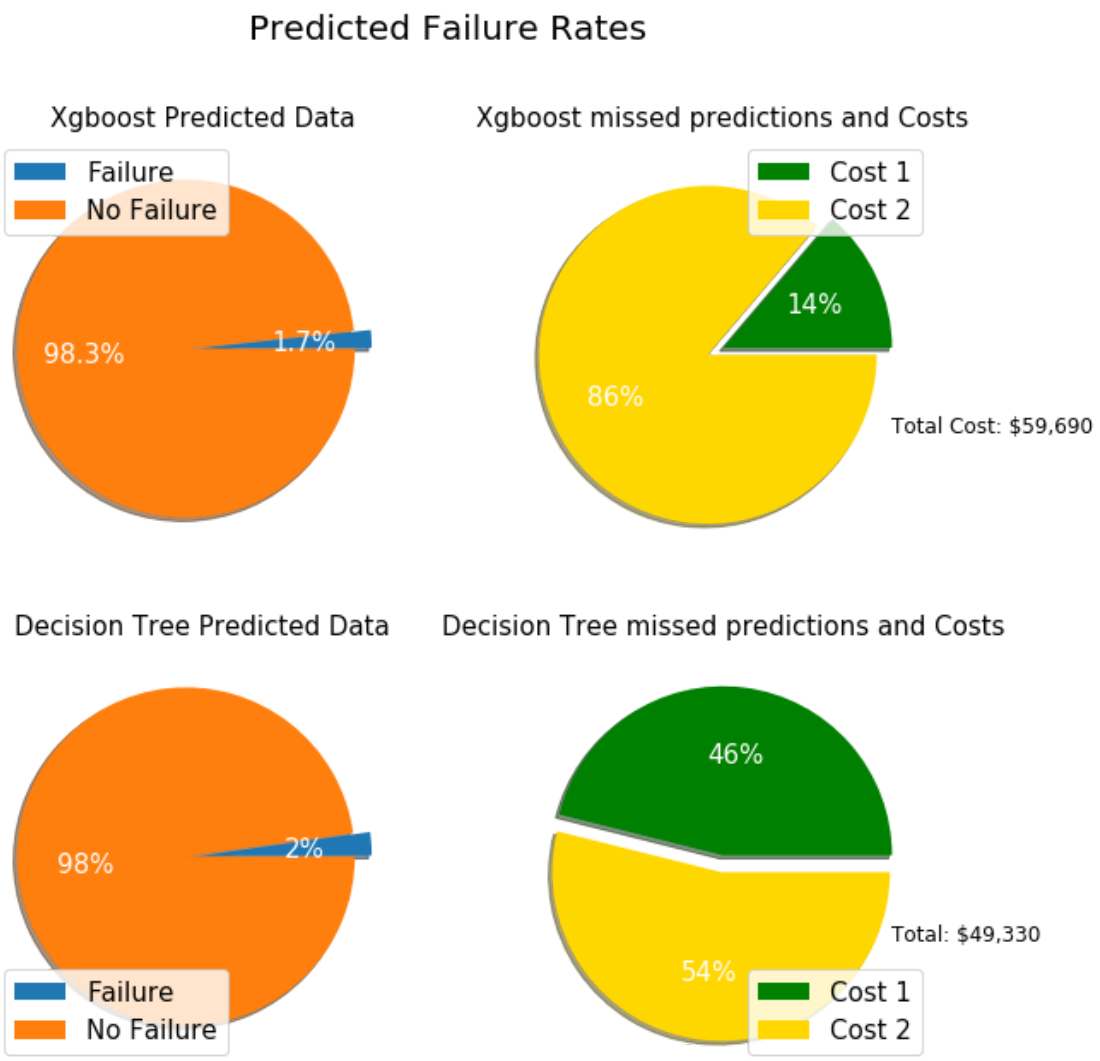
The data the model will be trained on contains less than 2% failures



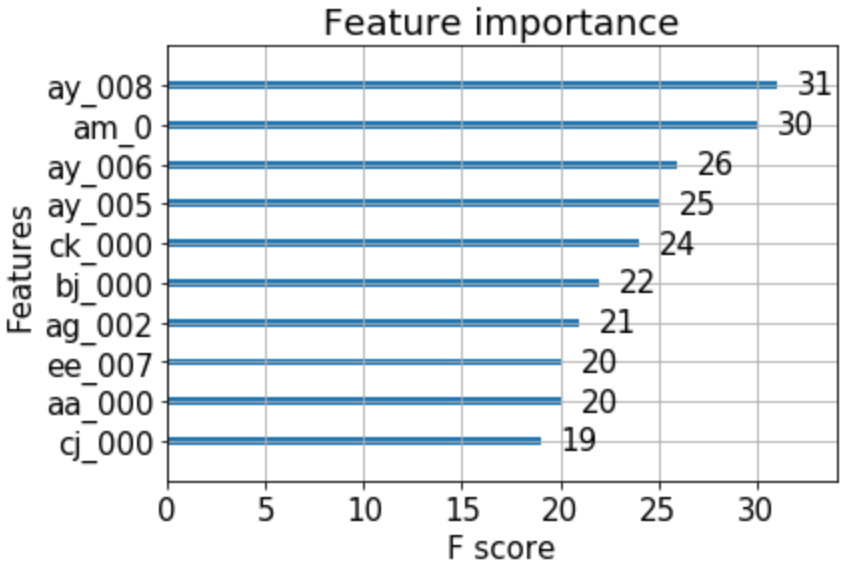
Results

The test dataset contains 375 confirmed cases of failures. Using the manufacturer's assessment guidelines, that will cost $187,500.

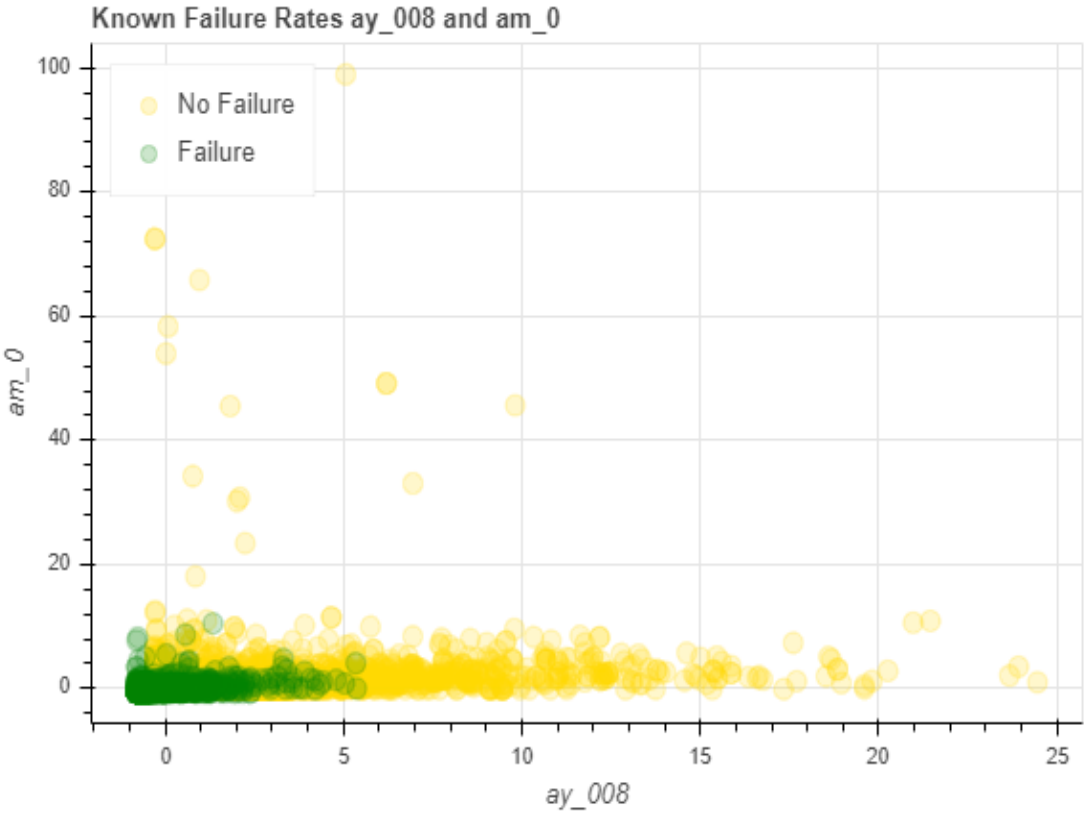
Using two different classifier algorithms, we see an immediate and substantial reduction in cost to $59,690. Running the data again using a slightly different classifier, the cost drops further to $49,330, and the split between Cost1 and Cost2 is much more evenly split.



The top features used by the classifier are shown below.



Plotting the classifier's two highest rated features shows an association. The failures occurred while both ay_008 and am_0 are in the bottom 20% of their recorded ranges.



Conclusions

The objective of this challenge was to reduce costs by predicting failures before they occurred in the field. By selectively performing maintenance a company could avoid the expense of dealing with a stranded truck and cargo.

Due to the deliberately undesignated nature of the data, feature selection and data cleaning were minimal.

Using a classifier algorithm to identify trucks that are in danger of failing resulted in cost reduction of 74%, even with the absolute minimum of practical insight into the data itself.

GitHub link to Jupyter Notebook