APS Failure Causation Prediction

in Scania Trucks

Abstract:

Introduction:

APS (Air Pressure System) is an essential system in heavy vehicles that manages breaking system mostly. Normally in heavy trucks, we see different mechanical issues in various components causing component failures. With maintenance, the chance of getting component failure decreases. But it costs money for regular maintenance. Also APS system is connected to other systems and failure in APS also impacts other connected systems. With failure type (Failed component as a part of APS system or not) prediction, we can save cost and time.

Business Problem Statement:

So the failed components can be part of APS system. It will be helpful if we can differentiate the failed component and isolate from the APS system so that it wont take much maintenance of the full APS system and it wont affect other systems connected to APS. With the daily usage data that can be collected from a truck during its operation at failure time, we can categorize whether the failure occurred in APS system or not after checking in service centers.

Now we can have data related to different components at different conditions and a binary value for the failure type (Related to APS or not). With these data, if we can predict whether the failure would be due to APS or not, it will be helpful and benificial for the manufacturer by isolating the failed component from APS system of the truck.

Business Constraint:

1. Low Latency: The model should predict result with low latency to avoid future failure in APS system and avoid high maintenance cost.
2. Cost of Misclassification: Wrong prediction would increase the maintenance cost as well as affect the owner due to future failure in APS system. Cost should be high for false negative prediction.

ML Problem Statement:

The above problem can be converted to a binary classification machine learning problem, where given sensor information, the model needs to predict whether the failed component is a part of APS system or not. Here positive class refers that the failure occurred due to a component of APS system and negative class refers that the APS system is not responsible for the failure.

Dataset Overview:

Total there are 171 columns out of which 170 are the collected features and one column contains the target value (‘Pos’, ‘Neg’). The feature names are not disclosed for proprietary reasons. All these features are numeric features. We have train and test dataset where in train set, we have 60000 data points out of which 1000 are positive and 5900 are negative which shows it’s a highly imbalanced data. While building models, we need to consider class imbalance. In test set, we have 16000 data points. So we have enough data to train and test.

Evaluation Metric:

The model performance will be calculated using confusion matrix, false positives and false negatives. We will penalize more to the false negative than than false positive. Because heavy truck maintenance are costly and if the model predicts that the failed component is from APS system, but its not, then if the manufacturer work on that component, it wont affect much and it will be considered as an extra/additional maintenance operation. But if the model predicts that the failed component isn’t from the APS system, but it actually is, then the manufacturer will miss the maintenance operation and failed to isolate that component. It will result more failures in that component and the APS system, which will cost much more than the previous case.

So we can set a penalty of 10 for cost1 (false positive) which refers to cost of unnecessary maintenance checks to be done. Similarly we can set a penalty of 500 for cost2 (false negative) which refers to cost of missing a faulty truck, which may cause a breakdown.