Scania Truck Air Pressure System Failures

Dataset Used in Intelligent Data Analysis Competition (IDA2016) (source Kaggle/UCI ML)

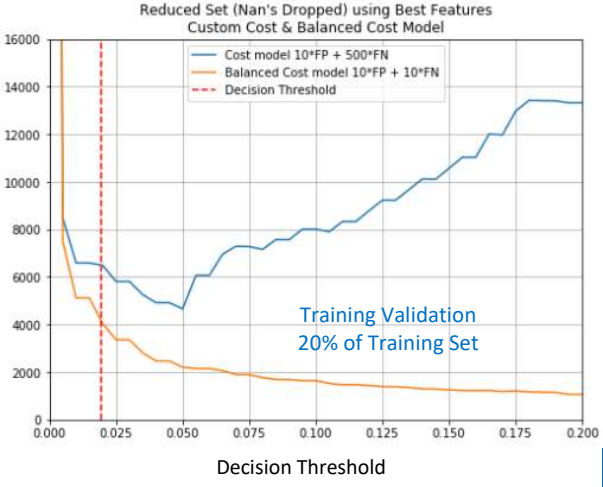
Objective
Minimize on road service cost and extra inspection cost

False Negative cost is 50 x higher than False Positive

Missed Service cost: 500
Extra Inspection cost: 10

60000 sample Training set
16000 sample Test set

Training



Train model with class weights

Weighted decision threshold to account for unbalanced costs (vs. using 0.5)

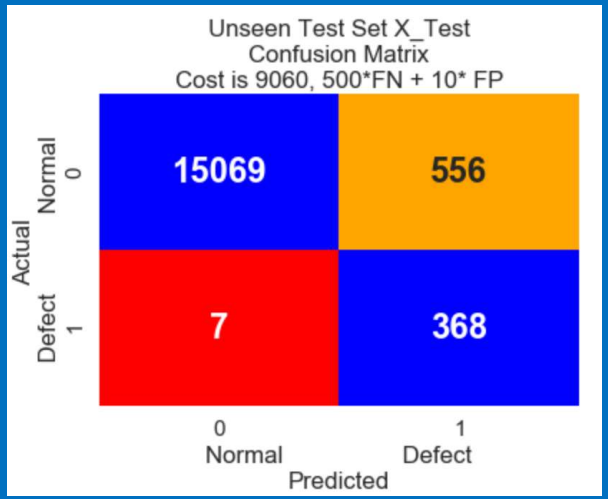
DTc = (FP+FN)/(FP'+FN')* 0.5
= (10+10)/(10+500) * 0.5 -> 0.0196

Vs. True Pos / All Test = 0.0234

Weighted & balanced cost models shown

Fix decision threshold before modelling

Results



False Negative Rate: 1.9%
False Positive Rate: 3.6%

98% of trucks needing service are identified

1.5 False Positives per True Positive

Total Cost: 500*7 + 10*556 = 9060

Lowest Cost at IDA2016: 9920

Modelled Prepared Feb 2020, Slide Prepared 03/21 by *Dave Buckley*