Towards a Machine Learning Algorithm for Predicting Truck Compressor Failures Using Logged Vehicle Data

By S. Nowaczyk, R. Prytz, T. Rögnvaldsson, S. Byttner

Outline

- System Description
- Problem Statement
- Problem Approach
- Algorithm Evaluation
- Conclusions
- Review

System Description

Air compressor provides operational power for

- Brake System
- Gearbox
- Suspension
- Several other systems

System Description

Loss of compressed air immediately halts the vehicle

Resulting costs include

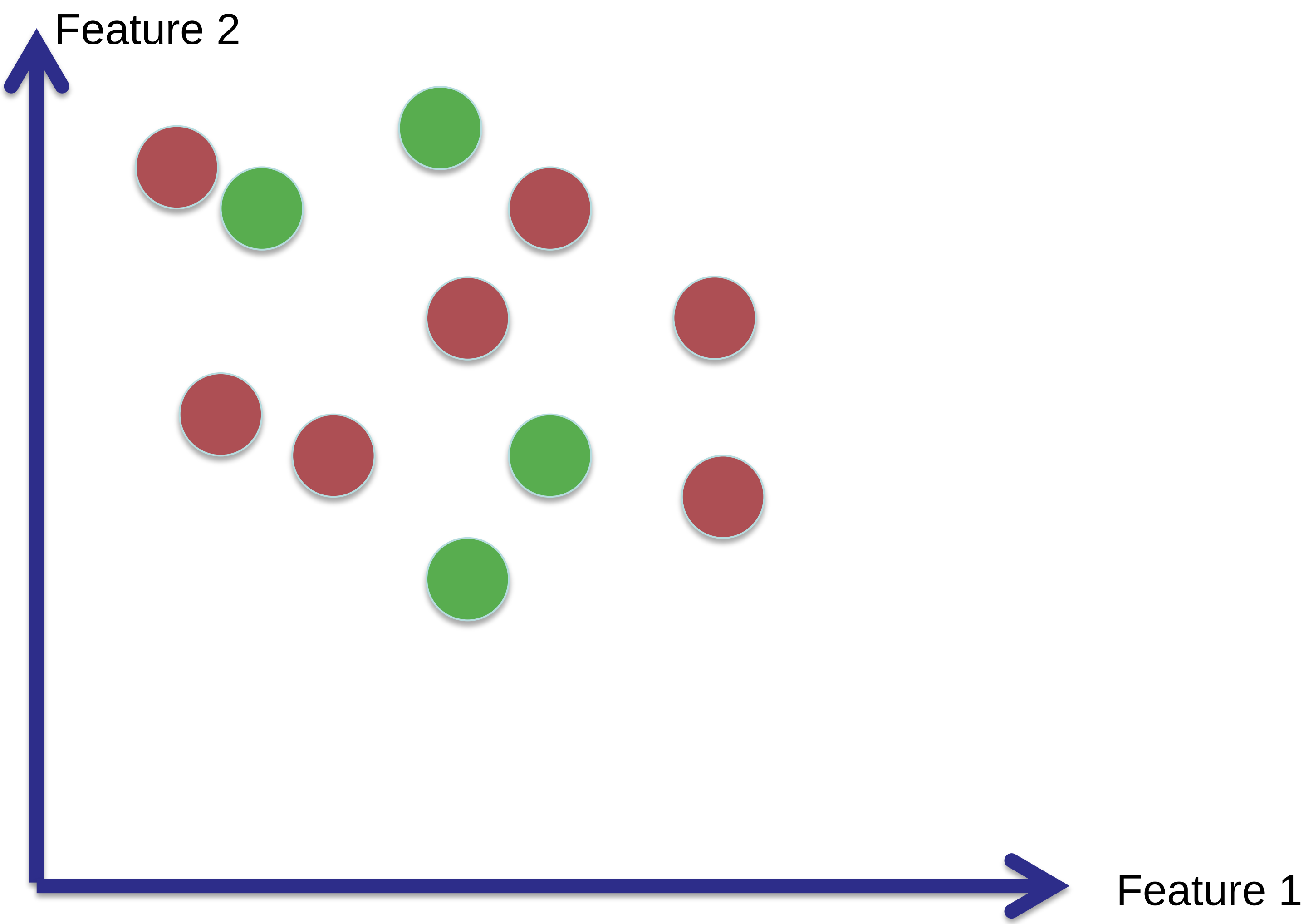
- Towing
- Disruption of garage workflow
- Repair costs
- Rent for replacement truck
- Possibly contract penalty and loss of reputation

Air compressor is a costly single point of failure!

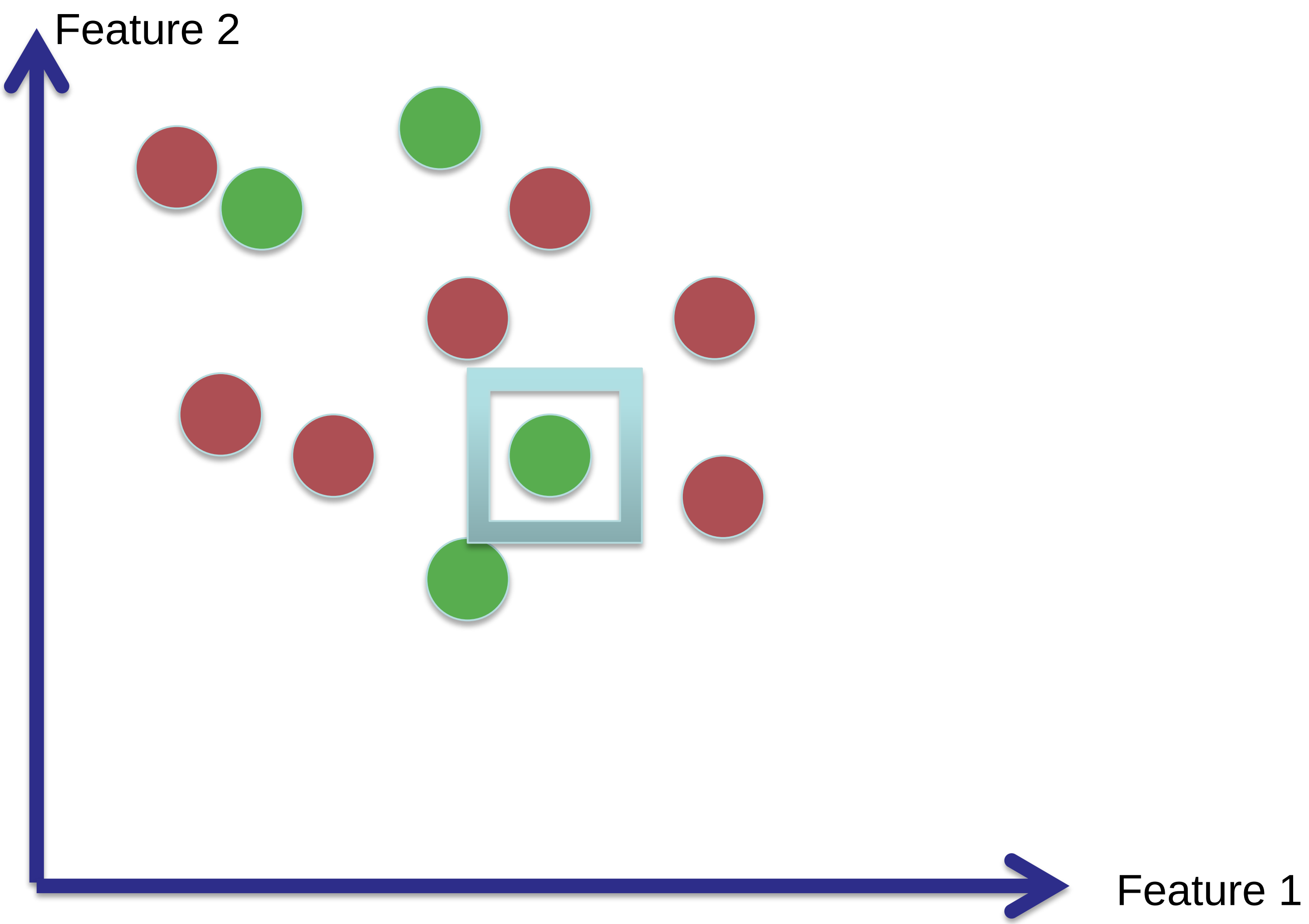
Problem Statement

- Approx. 5 per cent of Trucks have air compressors fail during lifetime (180 out of 4000 in database)
- High costs of compressor failure make prediction desirable
- Traditional approaches don‘t suffice
  - Previous papers used KNN, C5.0 and Random Forest

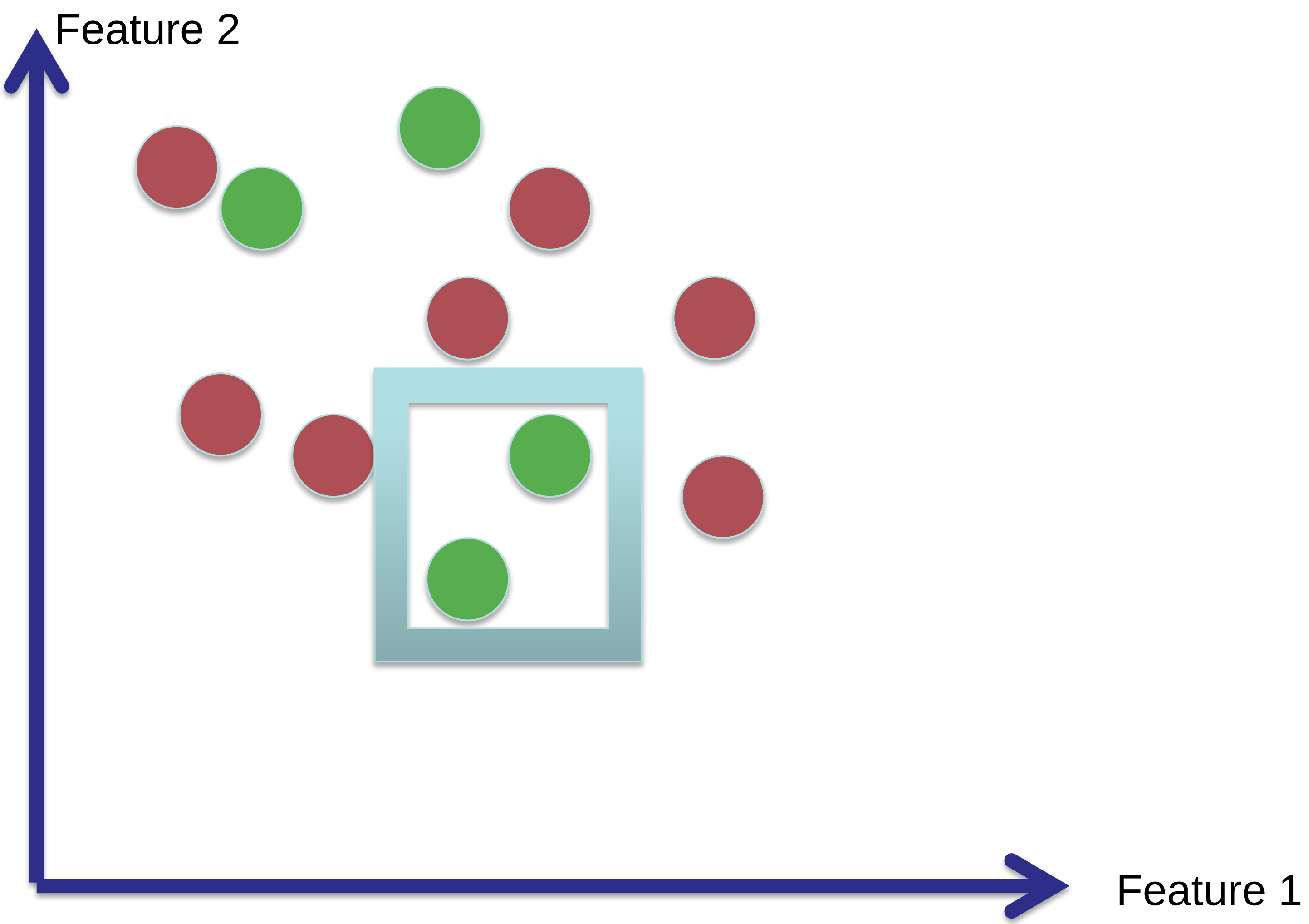
Problem Approach - Rule Induction



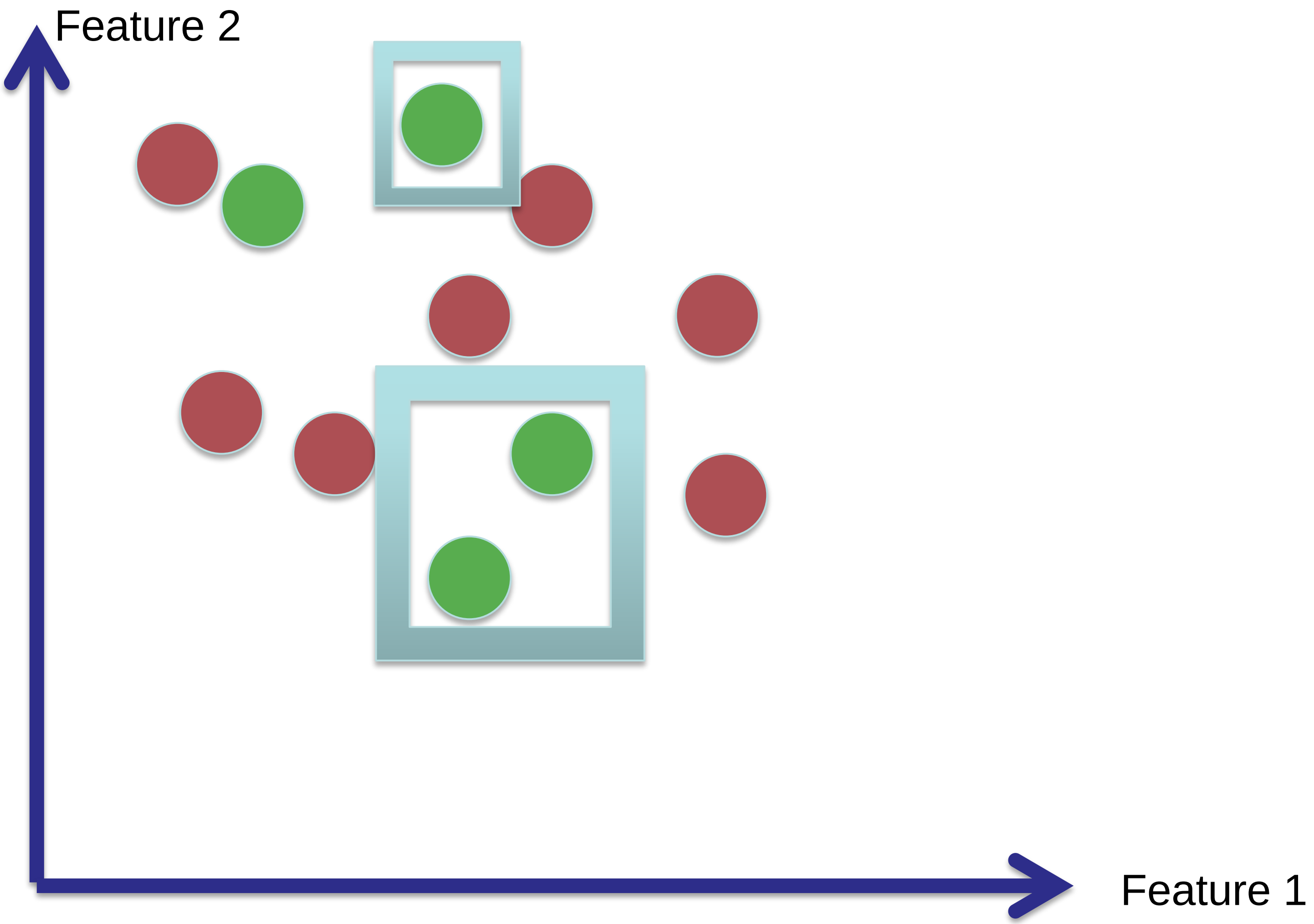
Problem Approach - Rule Induction



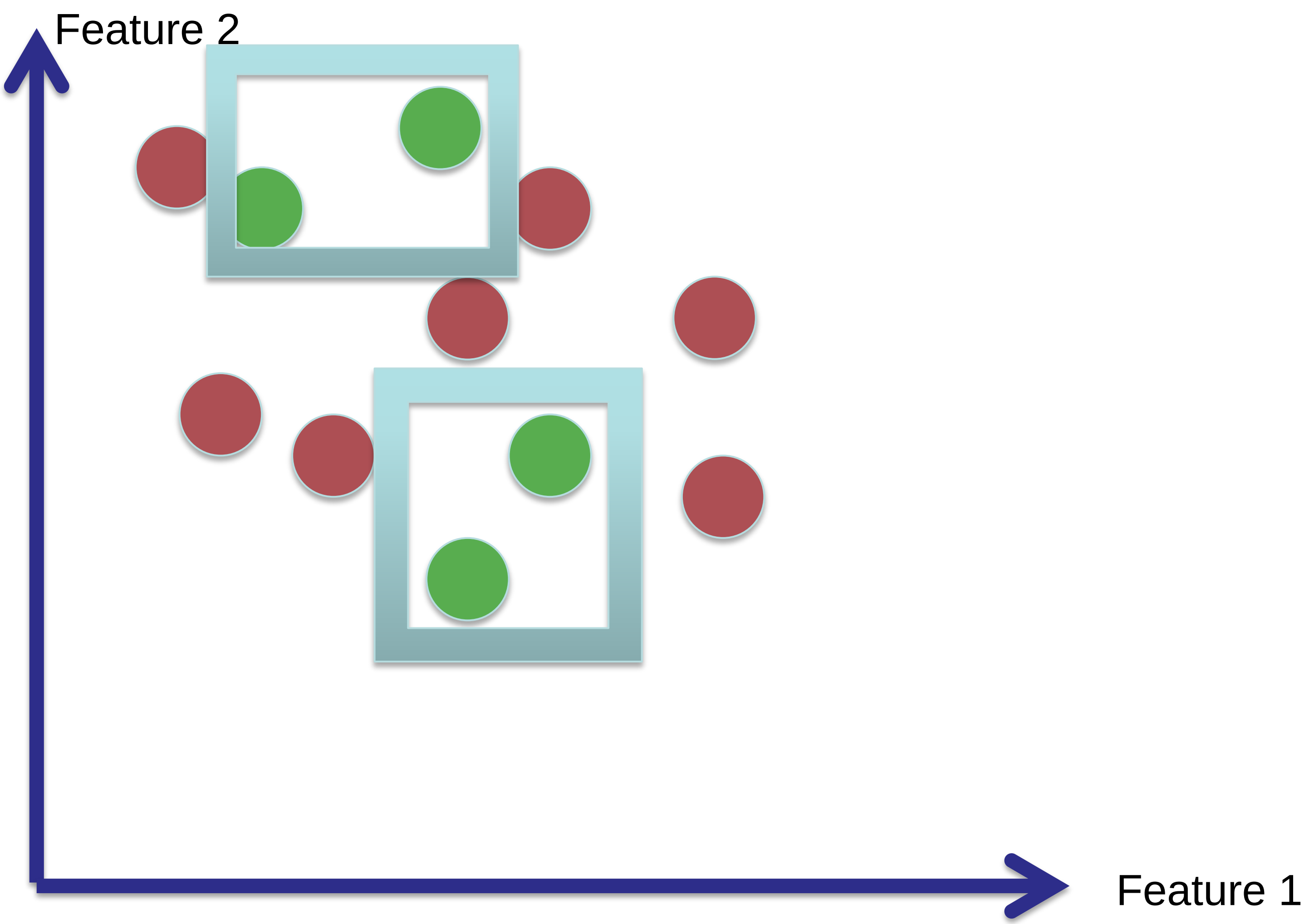
Problem Approach - Rule Induction



Problem Approach - Rule Induction



Problem Approach - Rule Induction



Problem Approach

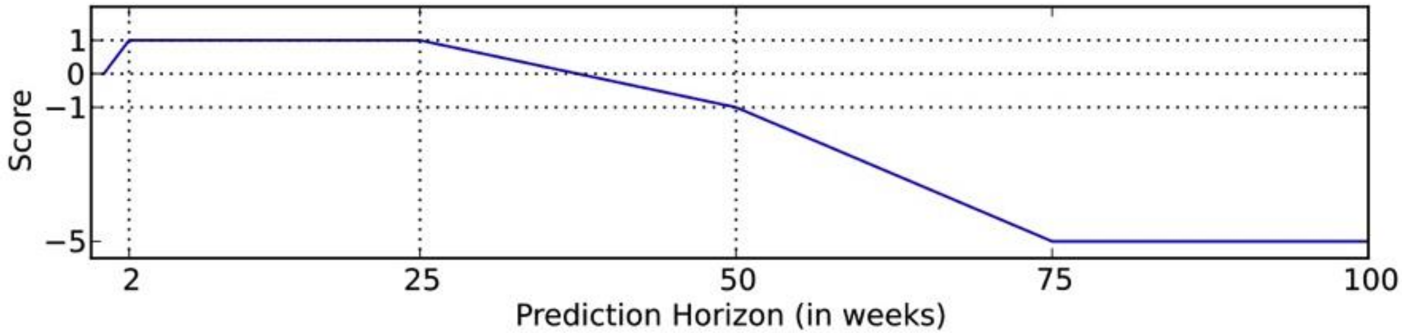
- Data
  - 3 Volvo Truck Service databases containing approximately 4000 Trucks
  - Each readout consists of hundreds of parameters
- Algorithm
  - Rule Induction with Relaxed Prediction Horizon (RRP)

Problem Approach – Novel ideas

- There is no suitable fixed prediction horizon for all vehicles
- Data sample readouts of single truck are not independent
- Cost function can be formulated as monetary cost saving function

Problem Approach – Relaxed Prediction Horizon

- There is no suitable fixed prediction horizon for all vehicles
- Preferred prediction Horizon between 2 and 25 weeks
- Penalize rules that result in too early warnings



Problem Approach –
Relational Readouts

TECHNISCHE UNIVERSITÄT DARMSTADT

- Truck readout data exhibits relations
  - Independency assumption loses information!
- Easy solution:
  - Generate Rules based on final readout before failure
  - Generalize to match *trucks*

Problem Approach – Monetized cost function

TECHNISCHE UNIVERSITÄT DARMSTADT

- How much can be saved by correct classification?
- What is the cost of misclassification?

$$C_{Save} = TruePositives \times (C_U - C_s) - FalsePositives \times C_p$$

C_{Save}	Cost savings
C_U	Cost of unplanned breakdown (here: 2.5 x C_s)
C_s	Cost of scheduled replacement
C_p	Not explained

Algorithm Evaluation – F-score

$$F = (1 + \beta^2)\frac{\text{precision} * \text{recall}}{\beta^2 * \text{precision} + \text{recall}}$$

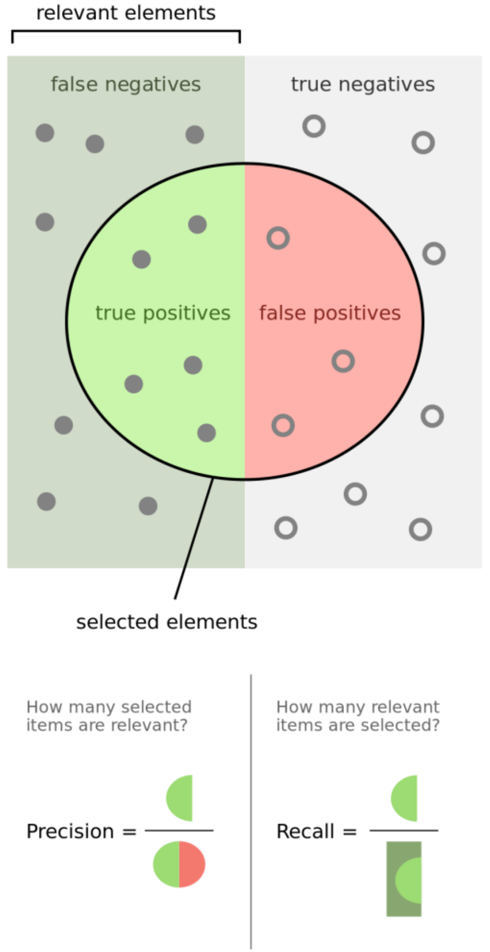
Beta: weighting factor between precision and recall

Precision: ratio of selected true positives to false positives

Recall: ratio of selected true positives to all positive examples

Algorithm Evaluation – F-score

$$F = (1+\beta^2)\frac{\text{precision} * \text{recall}}{\beta^2 * \text{precision} + \text{recall}}$$



Algorithm Evaluation – F-score over dataset size

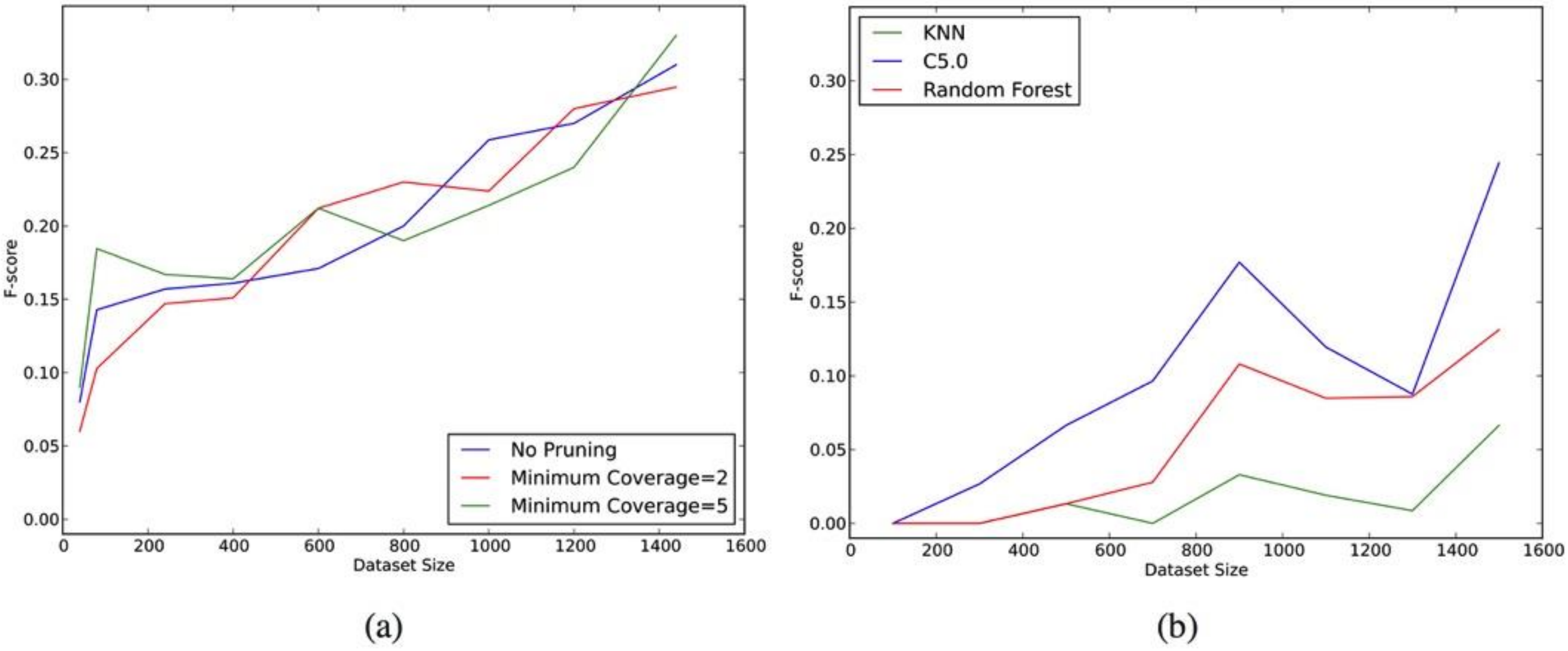


Figure 2. Impact of dataset size on $F_{0.5}$-score, comparing (a) the RRP algorithm against (b) results obtained by three popular methods, as described in an earlier work

Algorithm Evaluation –
Maintenance cost savings over fault frequency

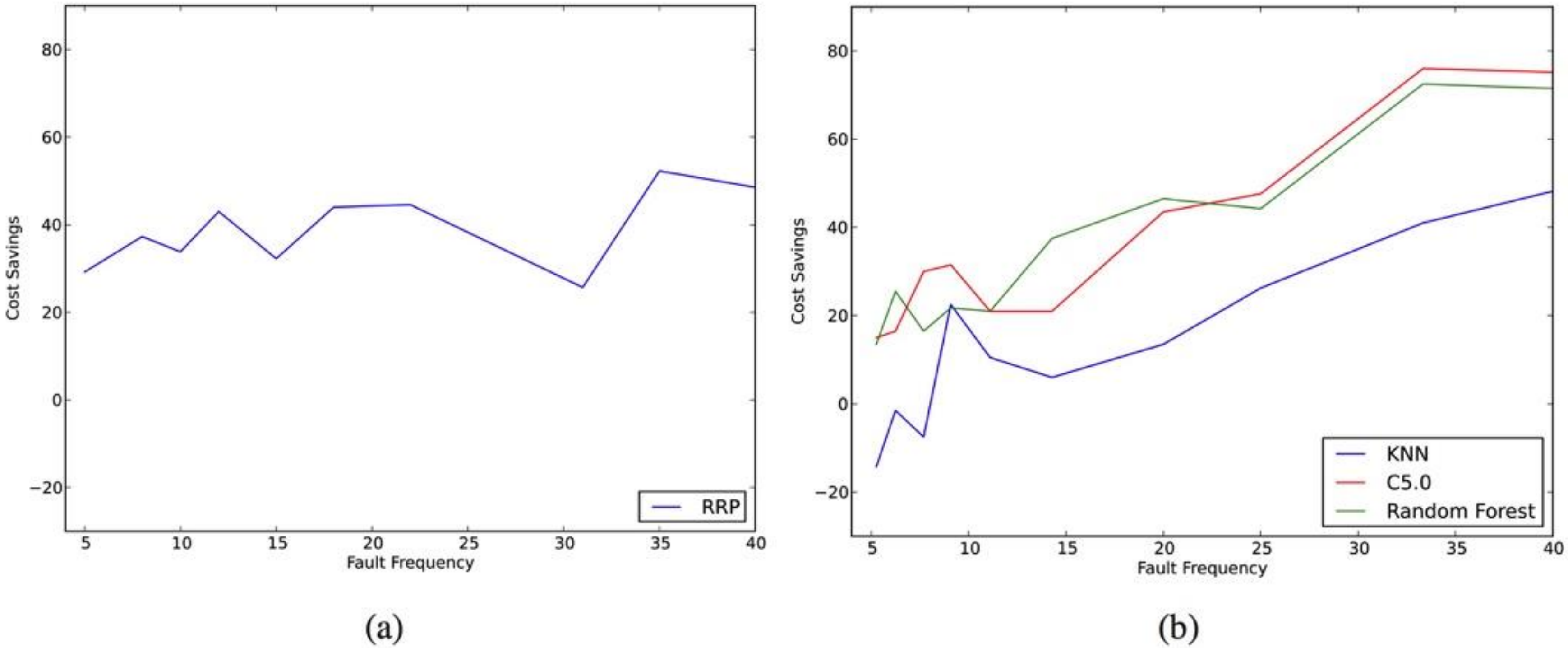


Figure 4. Maintenance cost savings, for different fault frequencies.

Algorithm Evaluation –
Classification errors over fault frequency

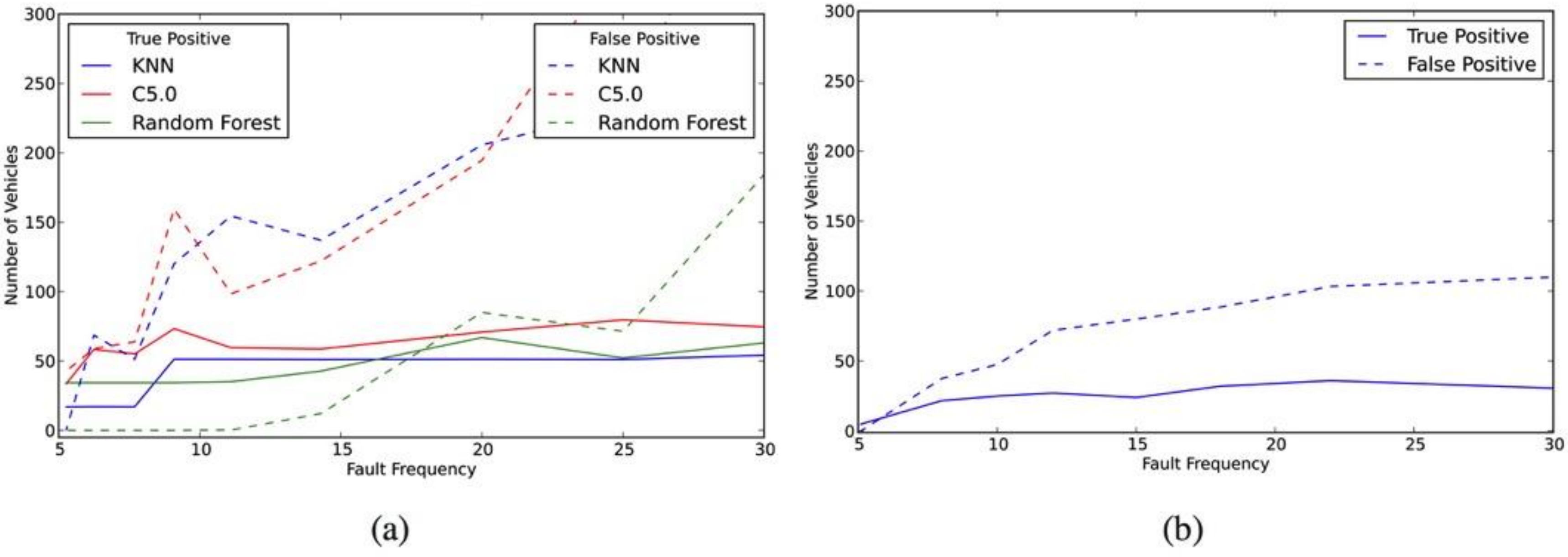


Figure 5. True Positives and True Negatives, trained on data with varying fault frequency and tested on data with 5% of faulty trucks

Conclusions

Key properties of new approach:

- Conservative classifier
  - Less mistakes on healthy trucks
  - Issues fewer warnings overall
- Relaxed Prediction Horizon

Conclusions – Future work

- Analysis of further components
- Automatic parameter selection
- Adaption to wear and tear parts
- „Time to Repair“ prediction

Review

- Cost function lacks parts of information
- It is not exactly clear which data is used and if/how it is preprocessed
- Minor inconsistencies in text
- It is not exactly clear how the independency assumption is circumvented
- Authors acknowledge this paper is intermediate step
- Overall this paper provides good insight to a novel adaptation of well-known algorithm

Thanks for your attention!

Picture sources:

Precision and recall graphic:

https://upload.wikimedia.org/wikipedia/commons/2/26/Precisionrecall.svg, created by Walber