Machine Learning for Predictive Maintenance

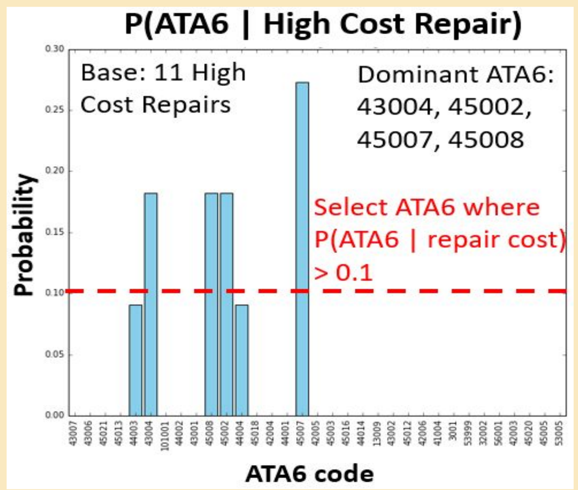
Authors: Julio Martinez (juliomz), Christianne Dennison (cdenn), Zhengyi Lian (zylian)
Machine Learning, Stanford University

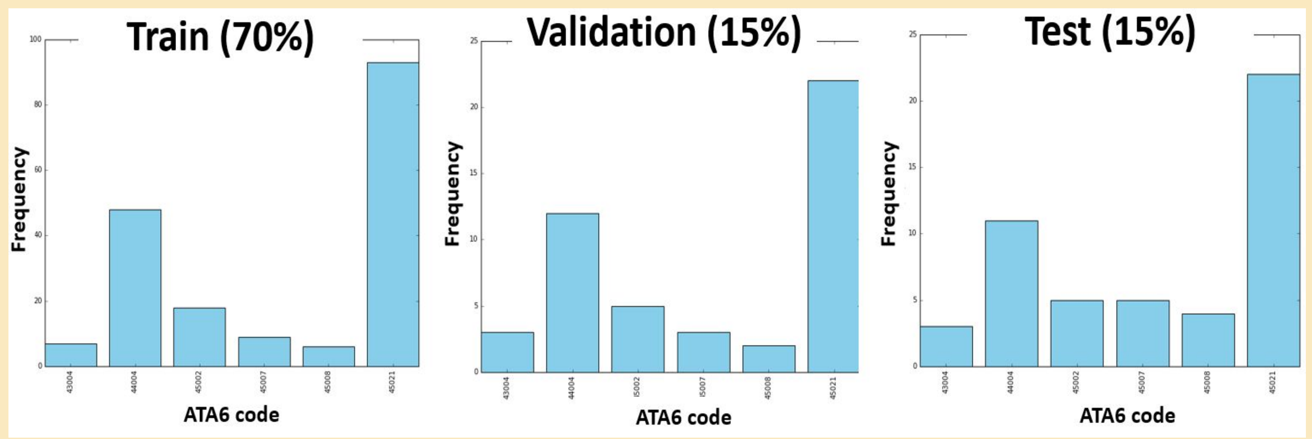
Motivation

PACCAR has made available a large dataset of sensor snapshots and repairs for commercial vehicles. Most ML for predictive maintenance involves using Random Forests [1][2][3] but the field is small.[4] **We take the problem a step further to predict a failure among a range of time windows** and determine the best model for predicting potential failure within a specific time range: Random Forest as a baseline, LSTM RNN, and GMM (Gaussian Mixture Models).

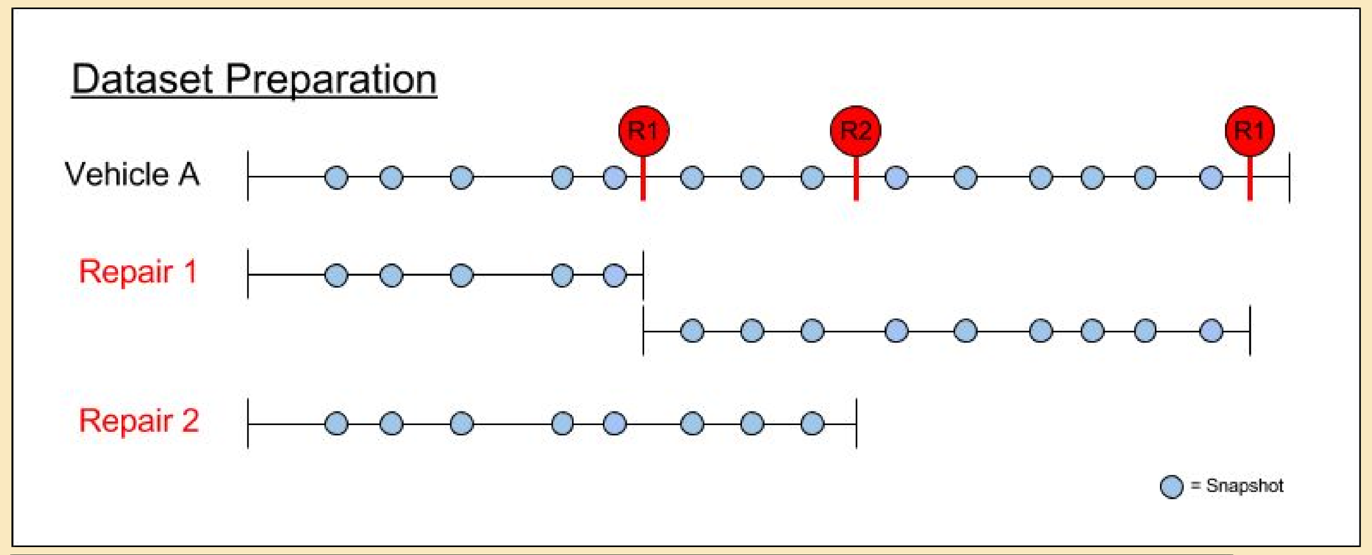
Pre-processing

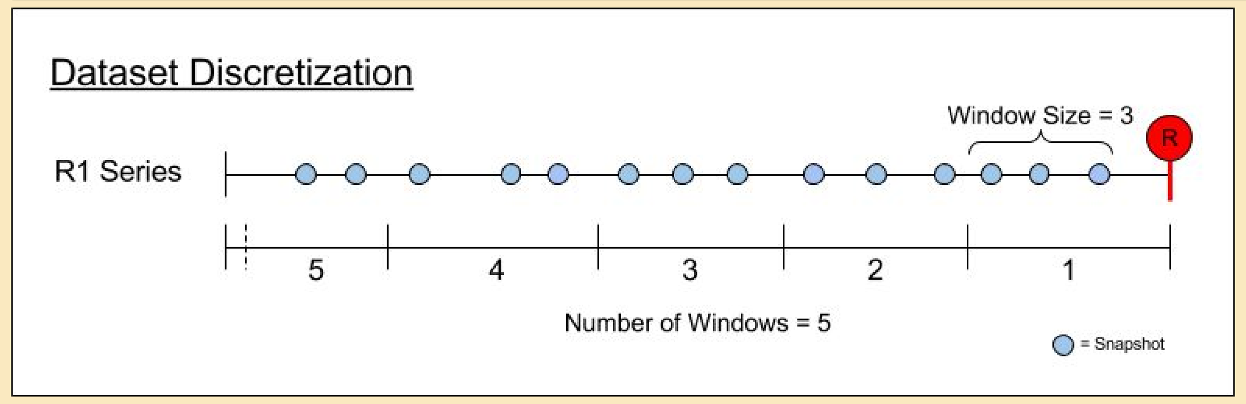
- Data cleaning (dedupe, impute, encode, whiten/shift)
- Select ATA6 codes to predict (freq. in higher cost repairs)
- Split into train, validation and test sets (systematic sampling)





- Map veh. ID, ATA6 code and time-window before a repair with that ATA6 to relevant sequences of snapshots





Typical dataset vehicles

Dataset

Vehicle Snapshot Data:

- 1015071 snapshots of 63 raw features collected from 797 vehicles during their trips.

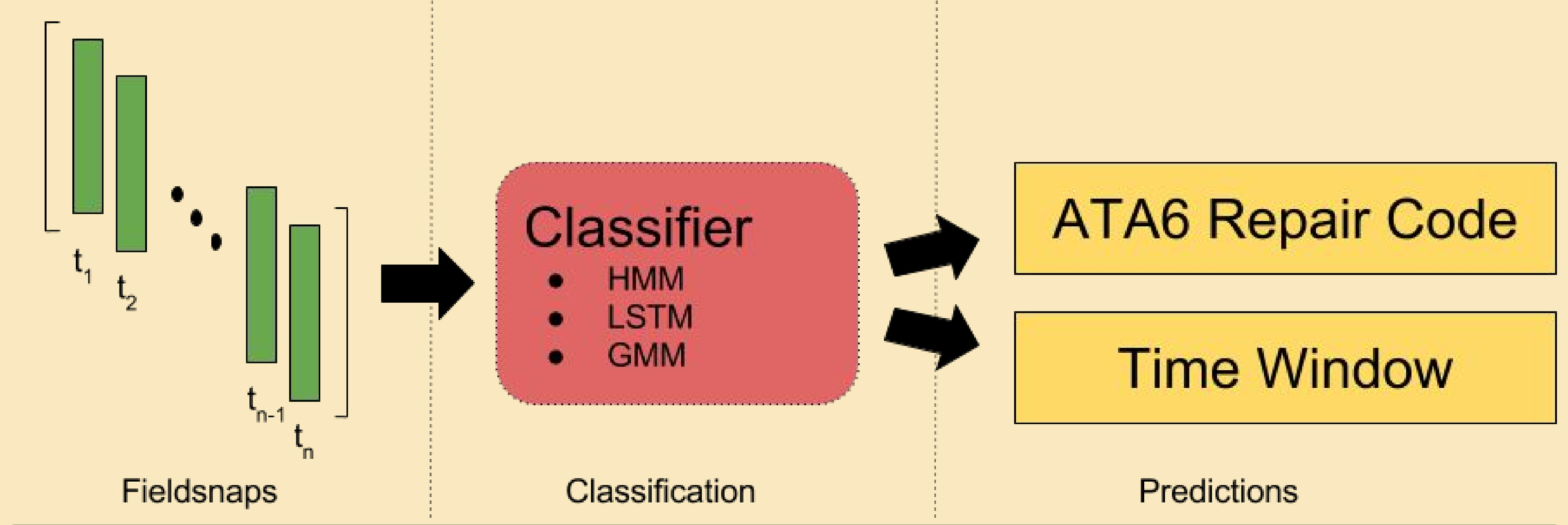
Veh Ref ID	Event DateTime	Event Type Description	Acc Pedal Position	Ambient Air Temp	Barometric Press	Brake Switch	Bus Utilization	Cat Intake Gas Temp	Cat Outlet Gas Temp	...	Altitude	Engine Start Ambient	Engine Start Coolant	Latitude	Longitude
254.0	2016-04-03 18:00:44	trip_start	0.0	38.59	101.5	0.0	49.0	211.78	214.28	...	48.0	37.77	86.66	37.5	-121.01

Vehicle Repair Claims Data:

- 713 repairs from 558 vehicles

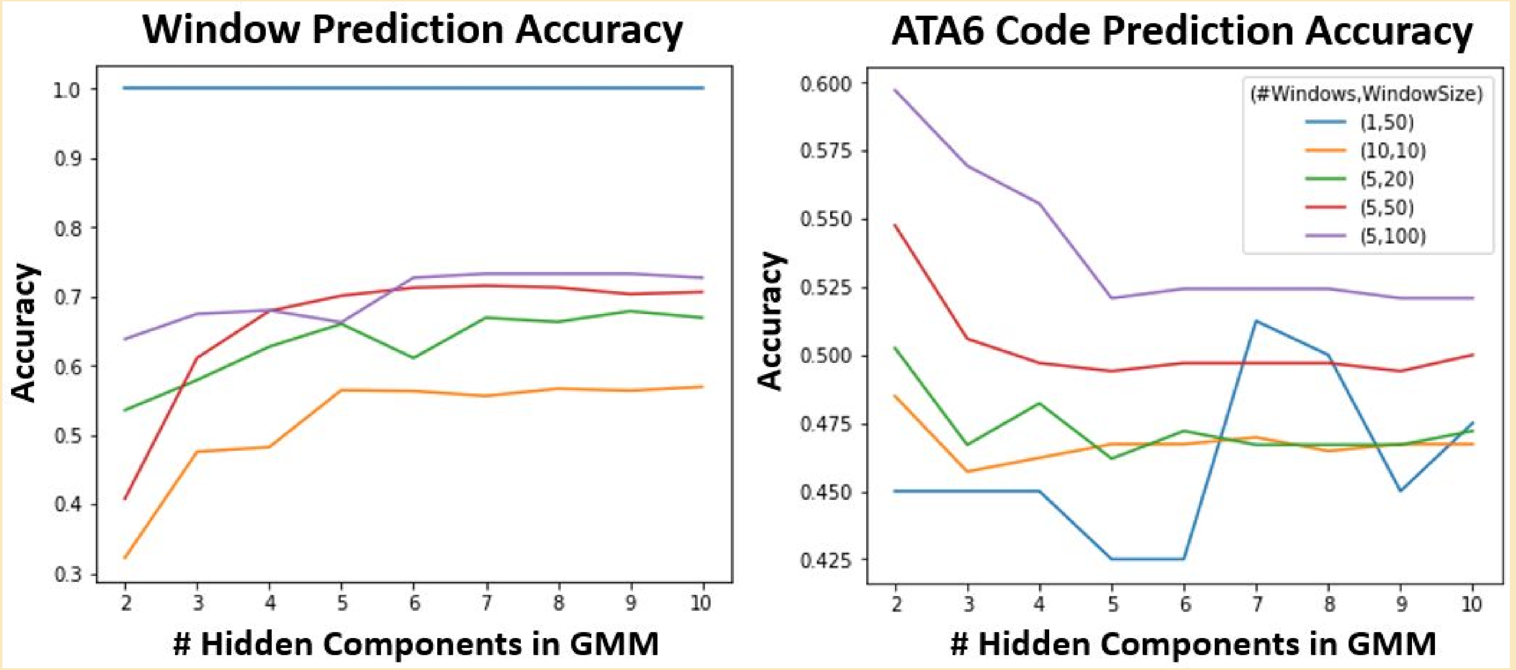
Chassis Reference Number	Model Vehicle	Build_Dt	Dlvry_Dt	In Service Date	Miles	Rpr_Dt	ATA3	ATA3Desc	ATA6	ATA6Desc	ATA9	ATA9Desc	Fail Type	Repair Cost
194.0	T880	1/22/2016	2/12/2016	2/12/2016	10	2016-02-03	43	EXHAUST SYSTEM (043)	43007	DIESEL EXHAUST FLUID (007)	43007021	FLUID - DEF, DEF SYSTEM (007)	MISCELLANEOUS	very low

Method & Experiments



General classification pipeline showing how a sequence of sensor snapshots are fed into a learned classifier (i.e. HMM, LSTM, GMM) for comparison. The prediction is the most likely Repair Code for that sequence along with the most likely number of time steps (time window) until that repair is needed.

- From initial tests: best accuracy for smaller number of windows but larger windows
- Each model was hyperparameter tuned
  - Random Forests: # of trees (70)
  - GMM: # Hidden Gaussian components (2)
  - LSTM: LR (0.01), Reg. (0.01), Hidden Units (139)



Results

Model	Val Acc.	Test Acc.
Random Forest (10,10)	48.8	
Random Forest (5,20)	51.2	
Random Forest (5,50)	55.2	
Random Forest (5,100)	59.3	
Random Forest (1,50)	72.3	69.9
GMM (10,10)	40.4	
GMM (5,20)	51.9	
GMM (5,50)	47.8	
GMM (5,100)	61.8	
GMM (1,50)	72.5	70.3
LSTM RNN (10,10)	44.8	
LSTM RNN (5,20)	47.7	39.4
LSTM RNN (5,50)	40.5	
LSTM RNN (5,100)	47.2	
LSTM RNN (1,50)	37.5	

GMM is best model at 70.3% test accuracy

Conclusions

The problem of window and failure type is a formidable problem with much room to explore. **These results on experimenting with other models help direct future research for predictive maintenance.** With more time, more data could be collected and our models could be validated in the field. Predictive maintenance offers our transportation, communication, and other infrastructure systems savings in downtime and repair as our population and connectivity continues to grow.

References

[1] R. Prytz, S. Nowaczyk, T. Rognvaldsson, and S. Byttner. Analysis of truck compressor failures based on logged vehicle data. Las Vegas, Nevada, USA. 9th International Conference on Data Mining.
[2] S. Ramanujam. The data science behind predictive maintenance for connected cars. 2016.
[3] M. K. Erik Frisk and E. Larsson. Data-driven lead-acid battery prognostics using random survival forests. Annual Conference of the Prognostics and Health Management Society 2014.
[4] R. Prytz. Machine learning methods for vehicle predictive maintenance using off-board and on-board data. Halmstad University Dissertations, 9, 09 2014.