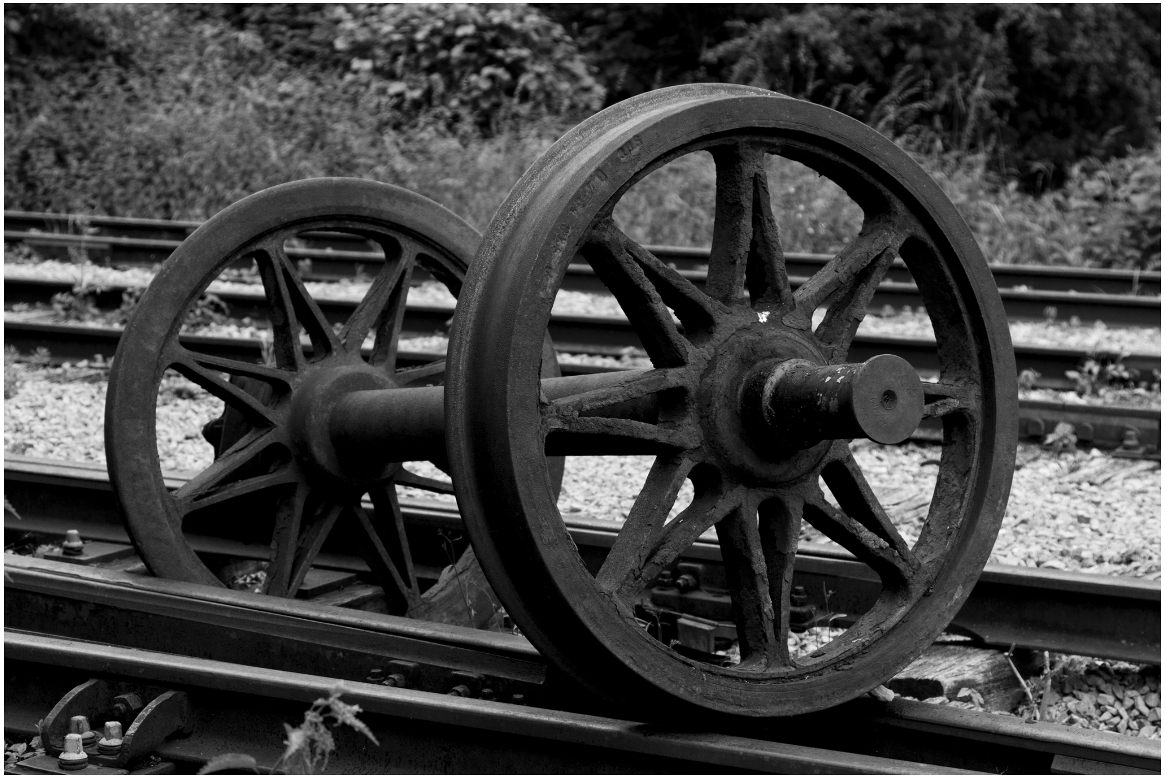
Predictive Maintenance:
Learning to Predict Train Wheel Failures

Chunsheng Yang, Sylvain Létourneau

TECHNISCHE UNIVERSITÄT DARMSTADT

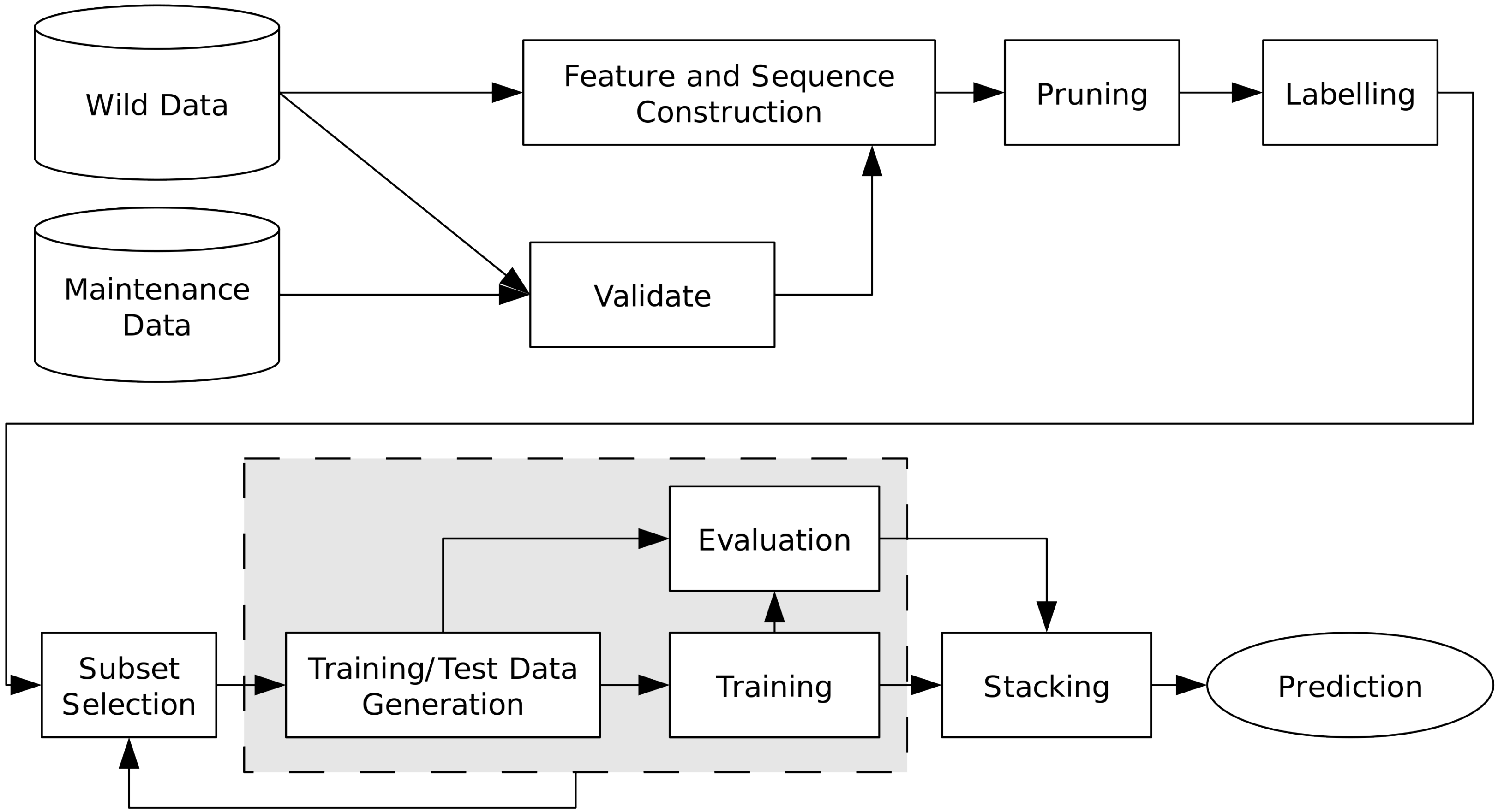


https://flic.kr/p/cjnA6E

Outline

- Process Overview
- Datasets
- Current State and Target
- Data Processing
- Training, Testing, Evaluation
- Stacking
- Summary and Conclusion

Process Overview



Data Sets: WILD Data

TECHNISCHE UNIVERSITÄT DARMSTADT

- 17 months, 804 cars, 200,808 observations → 0.5 obs/d/car
- distribution?
- readily available

"Railway-defect-detectors-sensors" by Marcus Wong Wongm - Own work. Licensed under GFDL via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Railway-defect-detectors-sensors.jpg

Attribute	Description
Date/time	Date and time of measurement
SiteName	The WILD site's name
Dir	Direction of the train (S, N, E, W)
Speed	Average speed of train
CarID	The ID of the car
NomLoad	Nominal load of the car
L01	Impact for wheel on left side of axle 1
R01	Impact for wheel on right side of axle 1
L02	Impact for wheel on left side of axle 2
R02	Impact for wheel on right side of axle 2
...	...
L12	Impact for wheel on left side of axle 12
R12	Impact for wheel on right side of axle 12

Current State and Target

TECHNISCHE UNIVERSITÄT DARMSTADT

- 75 % wheel failures discovered during visual inspection
- Impact Event > 140 kips → replace wheel immediately (< 25 %), threshold based
  - delay
  - disruption of schedule, reduced throughput
  - cost
- some cases: wheel actually fails

- **Better: Predict impact event and replace wheel before the impact occurs.**

Data Sets: Maintenance Data

- 20,700 records related to the fleet
- small fraction related to wheel failures:
  - tread shelled, flange defects, out-of-round
- validate impact events (succeeded by a replacement)
- 2 % of axles in the dataset fail (wheel lifetime?)

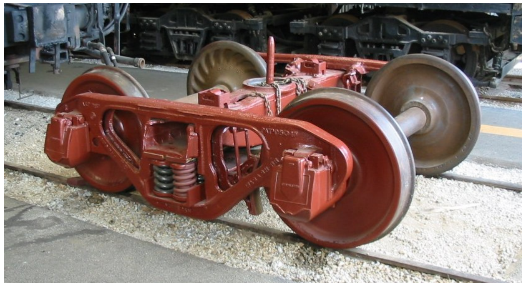
http://interfacejournal.com/archives/599

Attribute	Description
Date/time	Date & time of wheel replacement
JobCode	Repair work completed
CarID	Car Id on the train
WheelID	Wheel position on the car
WhyMadeCode	Reason for repair
Description	Job description
Cost	Cost of repair

Feature Construction

- per car → per axle
- relation to other axle on same truck
- Moving average over the recent instances

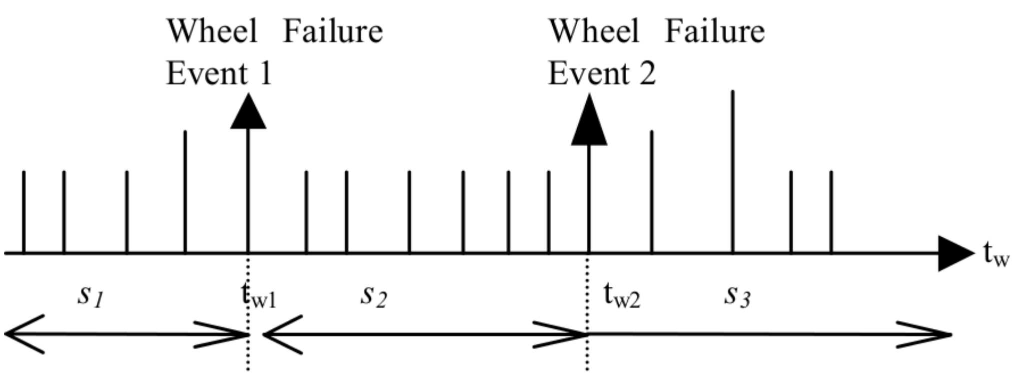
Attribute	Description
Date/time	Date and time of measurement
SiteName	The WILD site's name
Dir	Direction of the train (S, N, E, W)
Speed	Average speed of train
CarID	The ID of the car
NomLoad	Nominal load of the car
AxleNum	Axle number in the car
MaxMeas	Maximal WILD impact measured among the two wheels on the axle
MaxMeasPos	Position that a wheel has higher impact (Right, Left)
MinMeas	Minimal WILD impact measured among the two wheels on the axle
DiffMeas	Difference between impacts from the two wheels on the axle
AvgMeas	Average of two wheels' impact measurements
AvgMeasOAST	Average of impacts from other axle on same truck
AvgMeasWithOASC	Average of all four wheels on the same truck
DiffAvgMeasOAST	Difference between AvgMeasOAST and AvgMeas
SequenceID	Sequence ID for each wheel failure event
WithinSeqID	Instance ID in a sequence
TimeToFailure	The time to wheel failure



Sequence Construction & Pruning

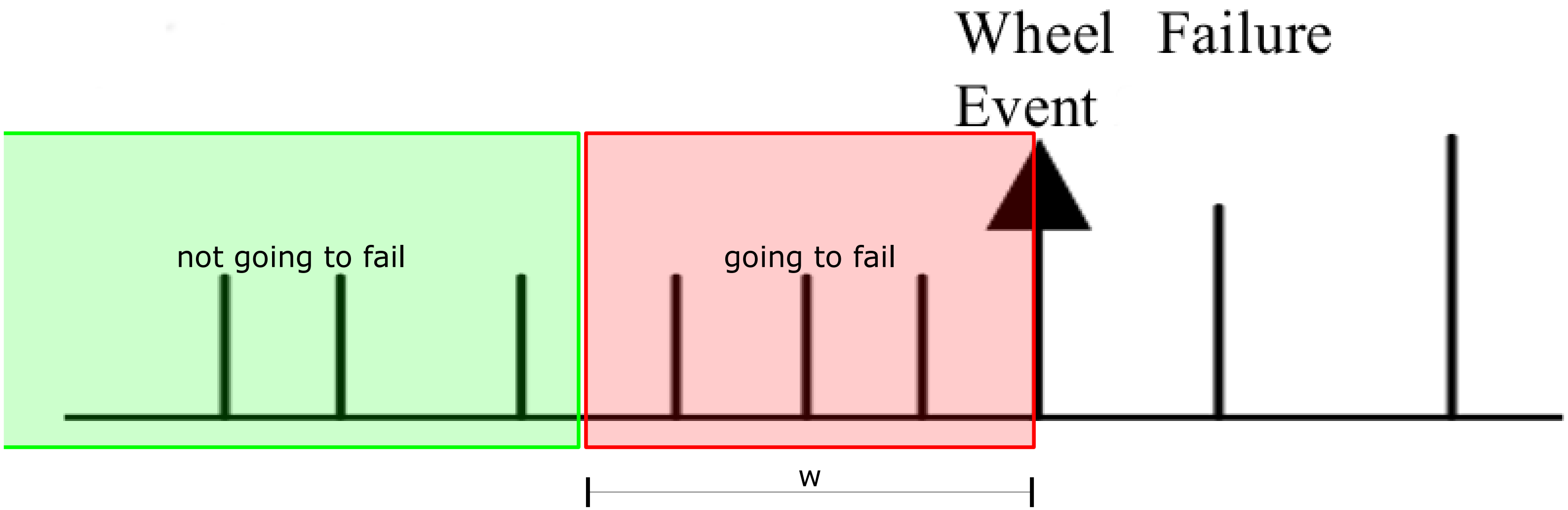
- 9,906 distinct sequences
- only sequences that lead to a failure are considered → 210 sequences
- limit sequence to 150 instances (300 days on avg.)

Attribute	Description
Date/time	Date and time of measurement
SiteName	The WILD site's name
Dir	Direction of the train (S, N, E, W)
Speed	Average speed of train
CarID	The ID of the car
NomLoad	Nominal load of the car
AxleNum	Axle number in the car
MaxMeas	Maximal WILD impact measured among the two wheels on the axle
MaxMeasPos	Position that a wheel has higher impact (Right, Left)
MinMeas	Minimal WILD impact measured among the two wheels on the axle
DiffMeas	Difference between impacts from the two wheels on the axle
AvgMeas	Average of two wheels' impact measurements
AvgMeasOAST	Average of impacts from other axle on same truck
AvgMeasWithOASC	Average of all four wheels on the same truck
DiffAvgMeasOAST	Difference between AvgMeasOAST and AvgMeas
SequenceID	Sequence ID for each wheel failure event
WithinSeqID	Instance ID in a sequence
TimeToFailure	The time to wheel failure



Labelling

- binary classification
- unsupervised (experts not feasible), automatic
- Assumption: abnormal state observable before failure
- window trade-off: enough time for maintenance action, not replacing components too early (w = 20 days)



Attribute Subset Selection

TECHNISCHE UNIVERSITÄT DARMSTADT

- "[…] We evaluated various subsets of attributes […]."
- "We kept the four most promising subsets […]."
- Which sets were tested?
- Which attributes?
- Size of the attribute sets?
- Evaluation score of the sets?

Attribute	Description
Date/time	Date and time of measurement
SiteName	The WILD site's name
Dir	Direction of the train (S, N, E, W)
Speed	Average speed of train
CarID	The ID of the car
NomLoad	Nominal load of the car
AxleNum	Axle number in the car
MaxMeas	Maximal WILD impact measured among the two wheels on the axle
MaxMeasPos	Position that a wheel has higher impact (Right, Left)
MinMeas	Minimal WILD impact measured among the two wheels on the axle
DiffMeas	Difference between impacts from the two wheels on the axle
AvgMeas	Average of two wheels' impact measurements
AvgMeasOAST	Average of impacts from other axle on same truck
AvgMeasWithOASC	Average of all four wheels on the same truck
DiffAvgMeasOAST	Difference between AvgMeasOAST and AvgMeas
SequenceID	Sequence ID for each wheel failure event
WithinSeqID	Instance ID in a sequence
TimeToFailure	The time to wheel failure

Training/Test Data Generation

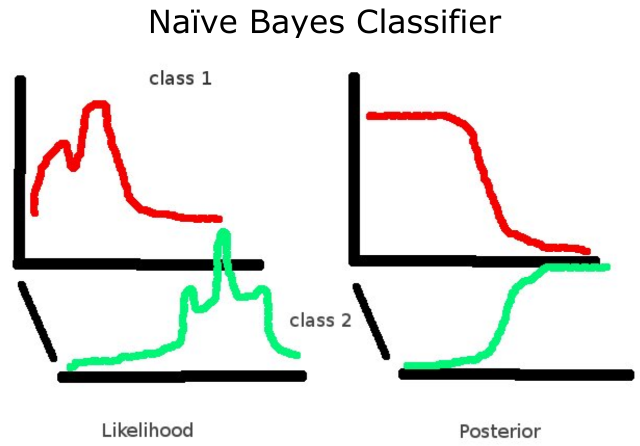
- Sequences must not be torn apart
- Instances within sequence are not independent (same failure event)
- Random sampling not applicable
- manual assignment

	Training-Set	Test-Set
Instances	22,083	9,337
Sequences	145	65

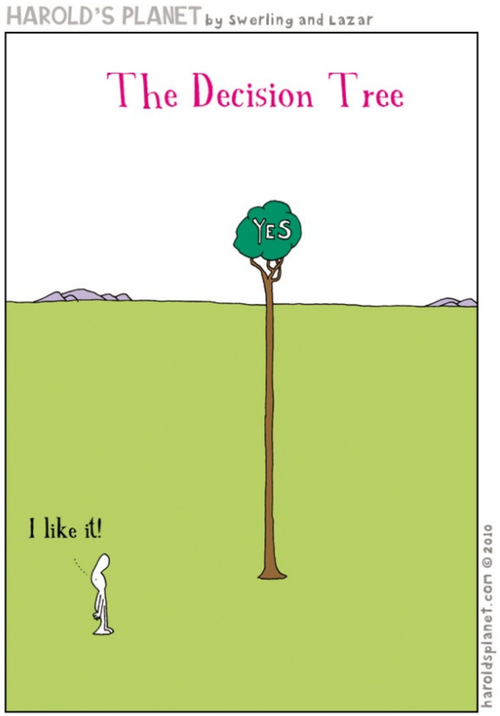
Model Building

TECHNISCHE UNIVERSITÄT DARMSTADT

- Simpler models preferred that can easily be explained to layman
- Each with and w/o costs (ratio 2:1)



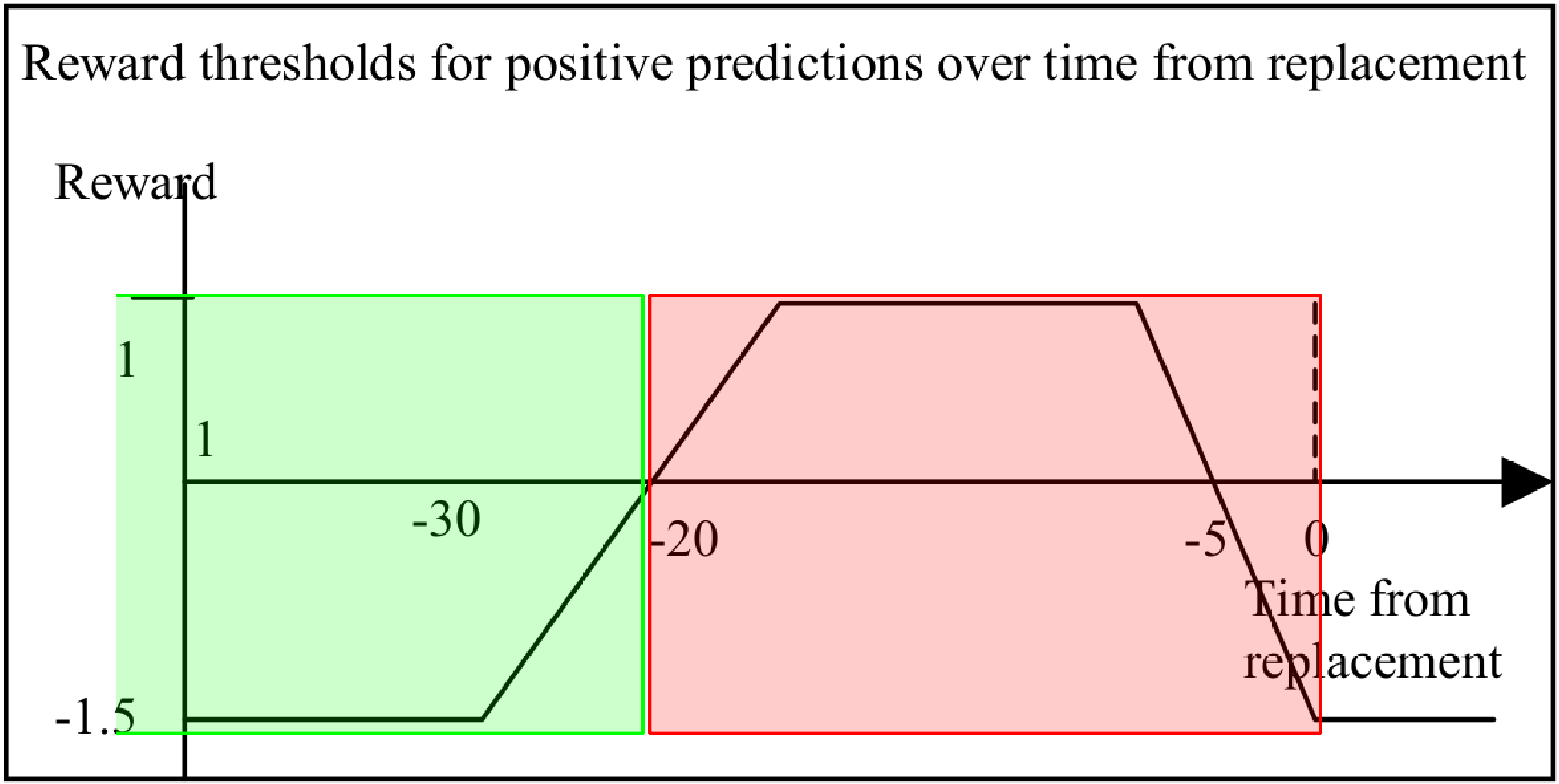
http://i.stack.imgur.com/RjlwS.jpg



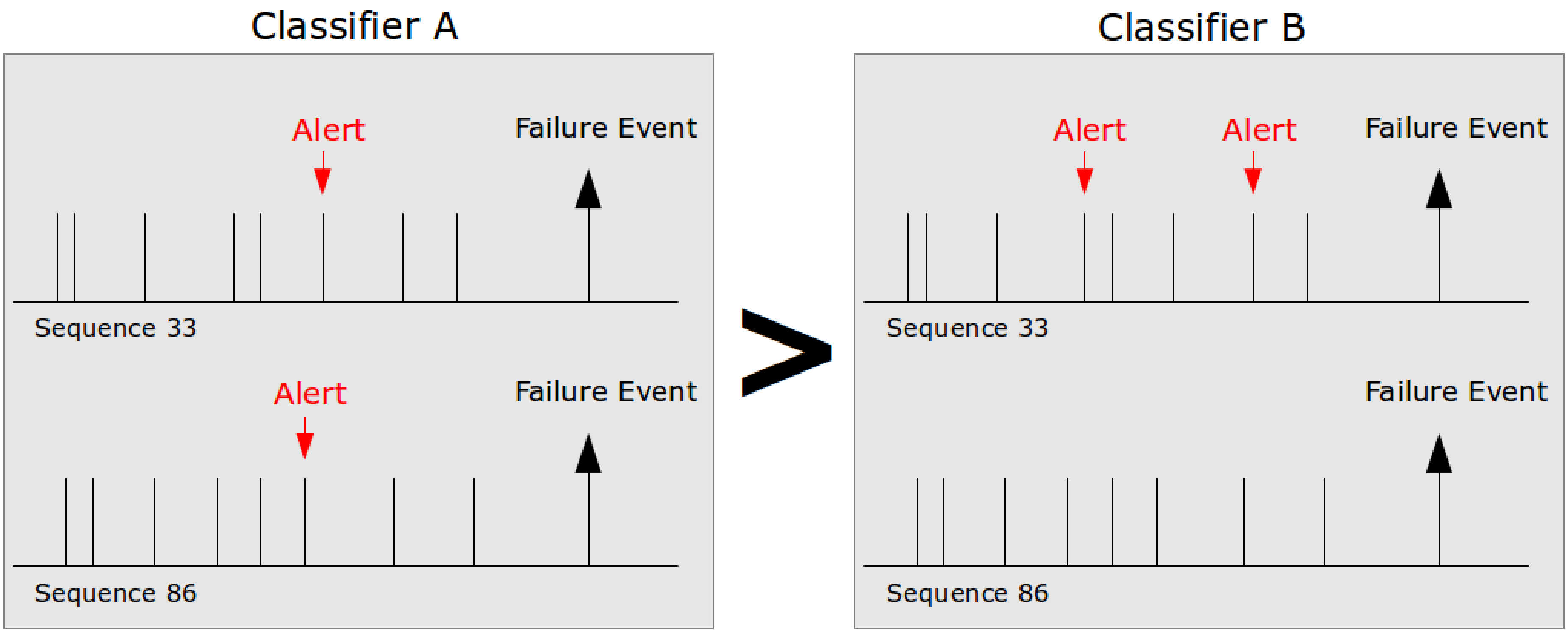
Evaluation

- Two properties of evaluation desired:
  1. Timeliness
  2. Coverage of sequences is better than coverage of instances
- This excludes traditional metrics like ROC, accuracy, precision,…

Evaluation – Timeliness



Evaluation – Distribution (1)



Evaluation – Distribution (2)

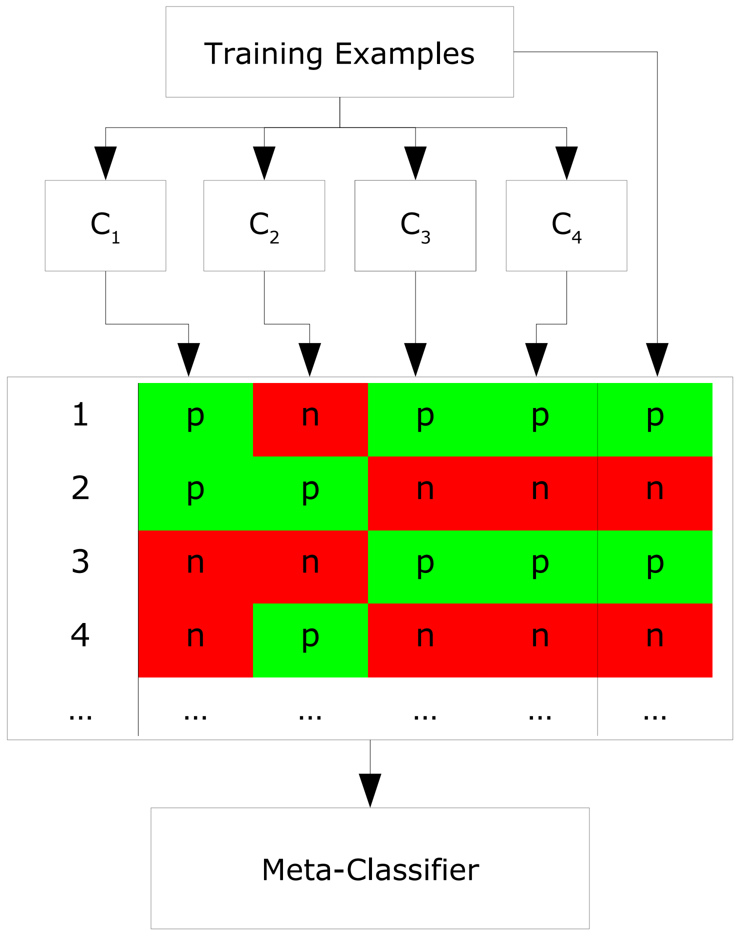
$$score = \left(\frac{\# Detected}{\# Cases} \right)^{Sign} \sum_{i=1}^{p} sc_i$$

- From -9500 to 1200 for the given scenario
- Missed sequences can be compensated with more true positives
- Sign = -1 & #Detected = 0 → score = 0

Evaluation – Results

Model Name	*Version of Algorithms*	*Model Score*	*False Positive Rate*	*Problem Detection Rate*
m_{11}	**Decision Trees**	**315.58**	**0.11**	**0.97**
m_{13}	Decision Trees with costMatrix	290.86	0.04	0.95
m_{14}	Naïve Bayes with costMatrix	198.69	0.12	0.97
m_{12}	Naïve Bayes	164.51	0.13	0.97
m_{21}	**Decision Trees**	**295.29**	**0.10**	**0.97**
m_{23}	Decision Trees with costMatrix	290.48	0.06	0.95
m_{24}	Naïve Bayes with costMatrix	188.81	0.15	0.97
m_{22}	Naïve Bayes	155.21	0.15	0.97
m_{34}	**Naïve Bayes with costMatrix**	**290.45**	**0.14**	**0.97**
m_{32}	Naïve Bayes	273.39	0.16	0.98
m_{33}	Decision Trees with costMatrix	161.51	0.12	0.91
m_{31}	Decision Trees	138.08	0.15	0.94
m_{44}	**Naïve Bayes with costMatrix**	**382.42**	**0.15**	**0.98**
m_{43}	Decision Trees with costMatrix	362.34	0.13	0.92
m_{42}	Naïve Bayes	349.60	0.16	0.98
m_{41}	Decision Trees	160.20	0.13	0.82

Stacking – Explanation



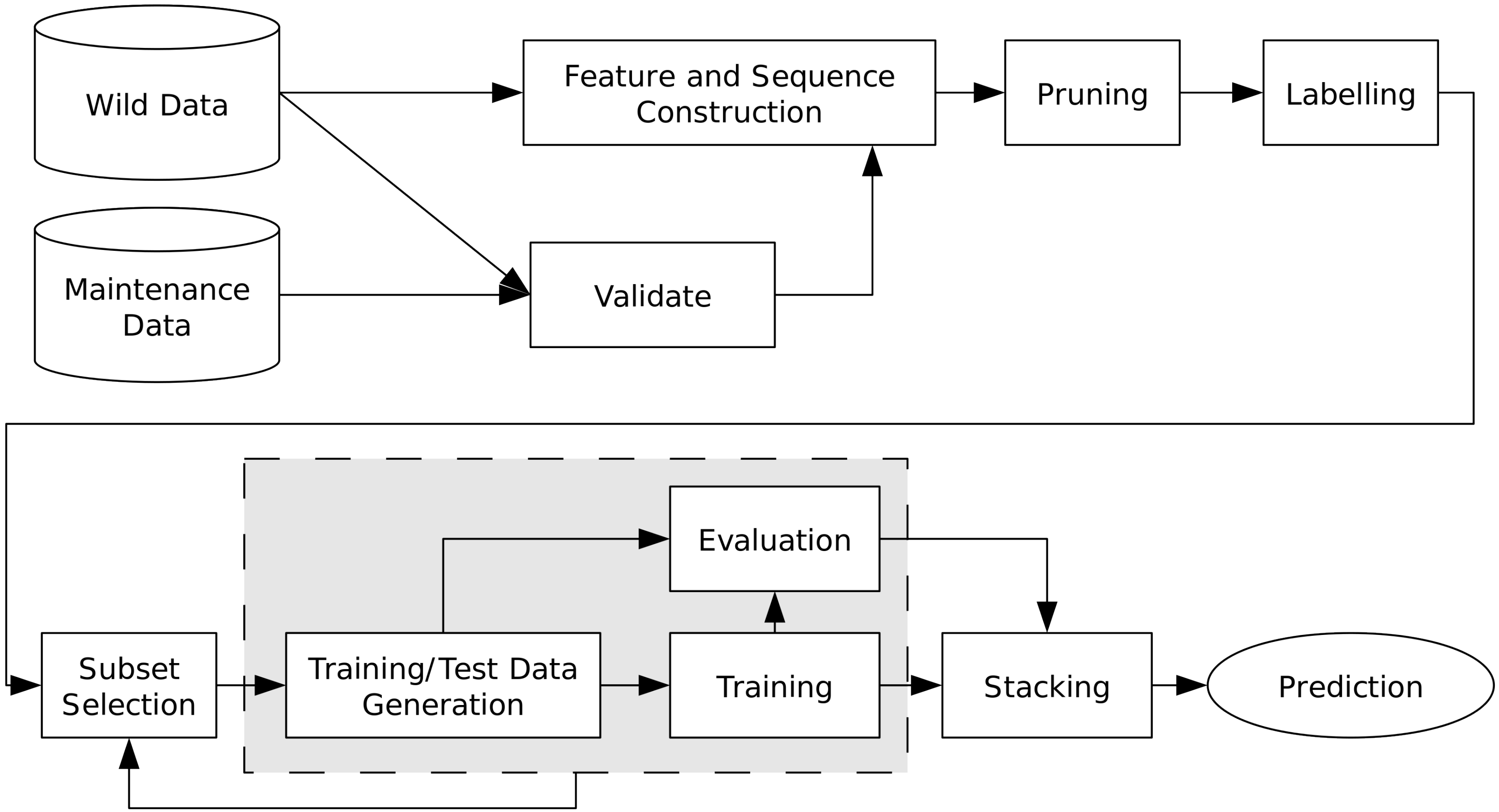
Stacking – Result

- Build meta-model out of predictions and confidence factors

Meta-Model Name	*Version of Algorithms*	*Model Score*	*False Positive Rate*	*Problem Detection Rate*
m_1^c	**Decision Trees**	**698.49**	**0.08**	**0.97**
m_2^c	Decision Trees with costMatrix	650.94	0.08	0.97
m_3^c	Naïve Bayes with costMatrix	643.35	0.12	0.98
m_4^c	Naïve Bayes	622.67	0.13	0.98

Model Name	*Version of Algorithms*	*Model Score*	*False Positive Rate*	*Problem Detection Rate*
$\mathbf{m_{11}}$	**Decision Trees**	**315.58**	**0.11**	**0.97**
m_{13}	Decision Trees with costMatrix	290.86	0.04	0.95
m_{14}	Naïve Bayes with costMatrix	198.69	0.12	0.97
m_{12}	Naïve Bayes	164.51	0.13	0.97
$\mathbf{m_{21}}$	**Decision Trees**	**295.29**	**0.10**	**0.97**
m_{23}	Decision Trees with costMatrix	290.48	0.06	0.95
m_{24}	Naïve Bayes with costMatrix	188.81	0.15	0.97
m_{22}	Naïve Bayes	155.21	0.15	0.97
$\mathbf{m_{34}}$	**Naïve Bayes with costMatrix**	**290.45**	**0.14**	**0.97**
m_{32}	Naïve Bayes	273.39	0.16	0.98
m_{33}	Decision Trees with costMatrix	161.51	0.12	0.91
m_{31}	Decision Trees	138.08	0.15	0.94
$\mathbf{m_{44}}$	**Naïve Bayes with costMatrix**	**382.42**	**0.15**	**0.98**
m_{43}	Decision Trees with costMatrix	362.34	0.13	0.92
m_{42}	Naïve Bayes	349.60	0.16	0.98
m_{41}	Decision Trees	160.20	0.13	0.82

Summary


Wild Data
Maintenance Data
Feature and Sequence Construction
Validate
Pruning
Labelling
Subset Selection
Training/Test Data Generation
Evaluation
Training
Stacking
Prediction

Conclusion

- generic workflow
- real-time wheel-monitoring
- Cannot predict failures not preceded by a 140 kips event.
- Does it generalize to other types of cars?
- A bit more detail of various parts.
- interesting insight

The End

Thank you for your attention!