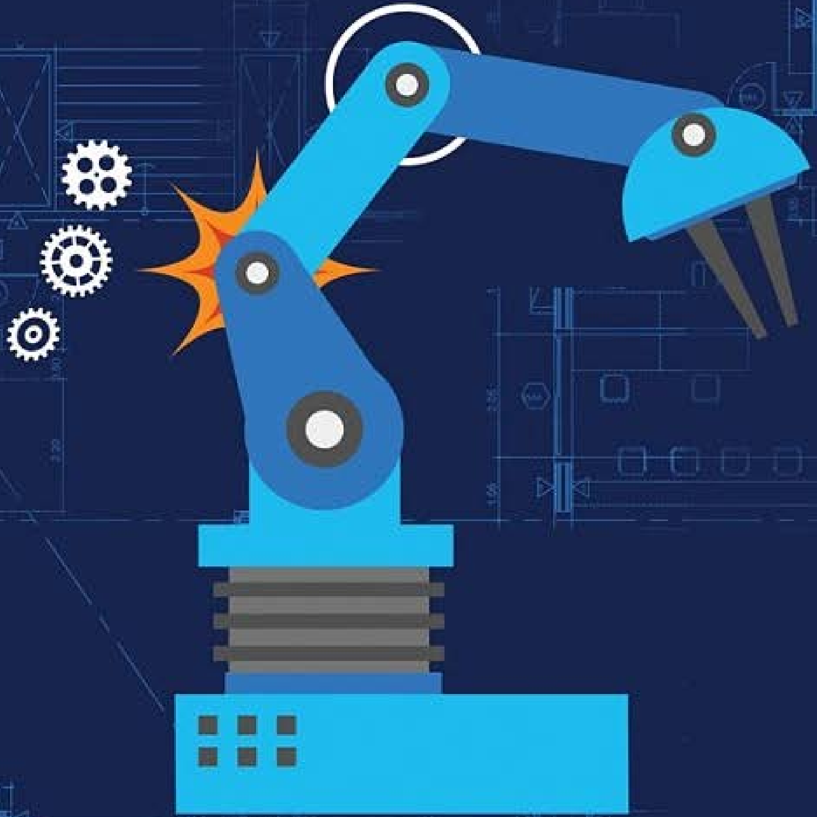
Finding balance between Corrective and Preventative Maintenance using Predictive Maintenance

Authors: Anant Jain, Viral Patel



Introduction: The promise of predictive maintenance

Corrective Maintenance

where parts are replaced as they fail

Preventative Maintenance

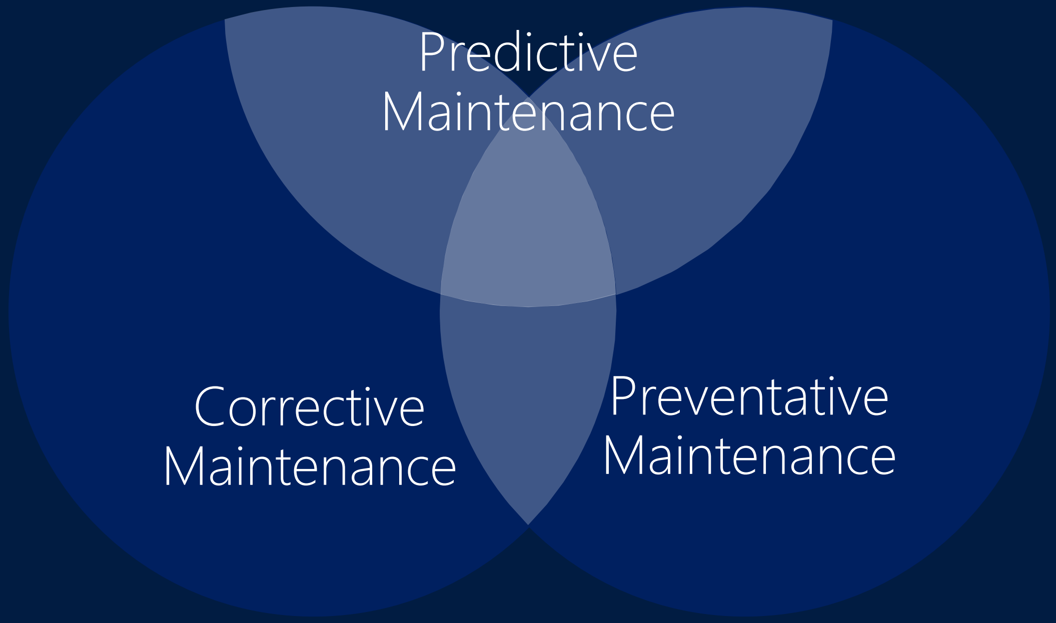
where a business may track or test component failures and determine a safe lifespan in which to replace that component before failure

Predictive Maintenance

Imagine if you could predict equipment failures before they happen, and systematically prevent them.

The objective: reducing downtime and decreasing maintenance

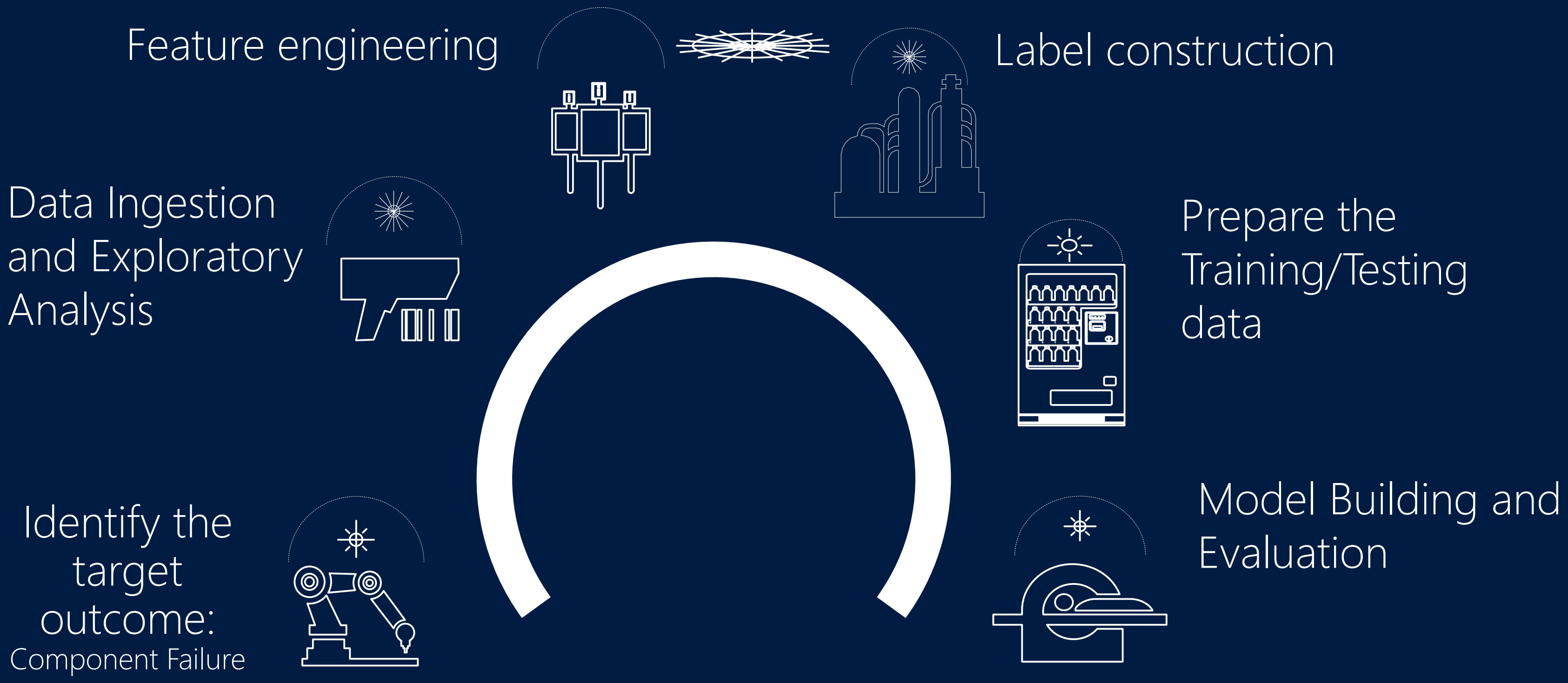
It involves using data to identify warning signs of potential problems, predict when equipment needs maintenance, and preemptively service that equipment before problems occur.



Overview :

- Identify all the potential sources of data: data from sensors and devices
- Use advanced analytics to quickly assess current conditions and recognize warning signs
- Format the problem as a multi-class classification problem : multiple components per machine
- Calculate probability that a machine goes down due to failure of a component within the next 7 days

Technical Workflow:



Classification models :

Random Forest Classifier

Decision trees are widely used since they are easy to interpret, handle categorical features, extend to the multiclass classification setting, do not require feature scaling, and are able to capture non-linearities and feature interactions. A random forest is an ensemble of decision trees. Random forests combine many decision trees in order to reduce the risk of overfitting.

Naive Bayes Classifier

Super simple, you're just doing a bunch of counts. If the NB conditional independence assumption actually holds, a Naive Bayes classifier will converge quicker than discriminative models, so you need less training data.

Gradient Boosting Classifier

Like other boosting methods, gradient boosting combines weak "learners" into a single strong learner in an iterative fashion, typically decision trees. GBDT training generally takes longer because of the fact that trees are built sequentially. However benchmark results have shown GBDT are better learners than Random Forests.

Neural Network Classifier

Neural networks offer a number of advantages, including requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training

Datasets:

Machines

Tracks a set of 1000 machines over the course of a single year (2015).

This data set includes information about each machine: Machine ID, model type and age (years in service).

Number of records: 1000

Errors

The error log contains non-breaking errors recorded while the machine is still operational.

These errors are not considered failures, though they may be predictive of a future failure event. The error datetime field is rounded to the closest hour.

Number of records: 11967

Maintenance

The maint. log contains both scheduled and unscheduled maint. records.

Scheduled maint. corresponds with regular inspection of components, unscheduled maint. may arise from mechanical failure or other performance degradations.

Number of records: 32592

Telemetry

The telemetry time-series data consists of voltage, rotation, pressure, and vibration sensor measurements collected from each machines in real time.

The data is averaged over an hour and stored in the telemetry logs for over the year 2015.

Number of records: 8.7 million

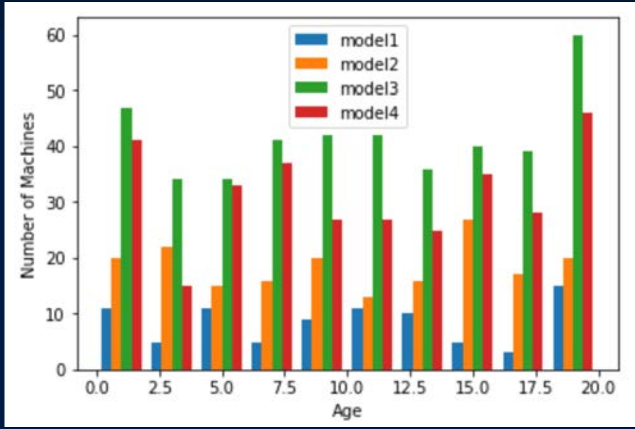
Failures

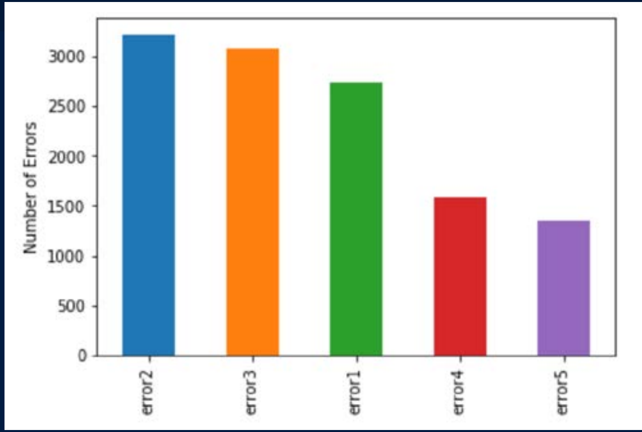
Failures correspond to component replacements within the maintenance log. Each record contains the Machine ID, component type, and replacement datetime.

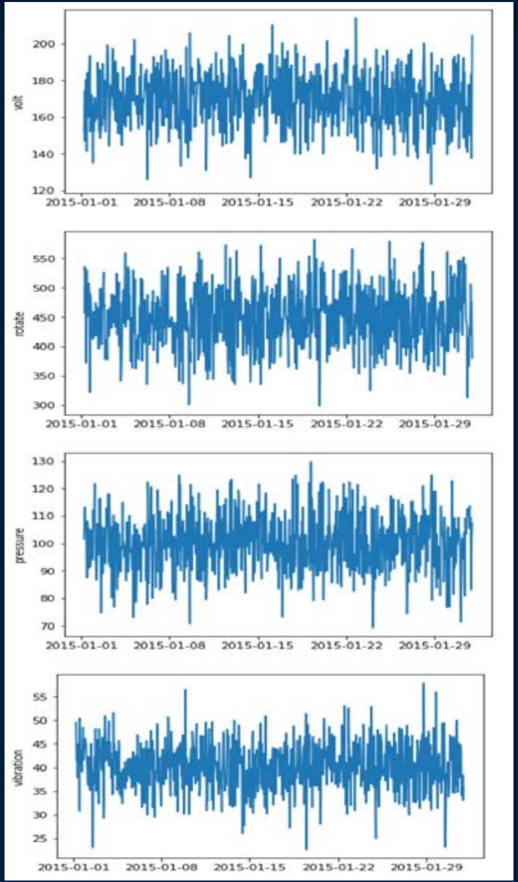
These records will be used to create the machine learning labels we will be trying to predict.

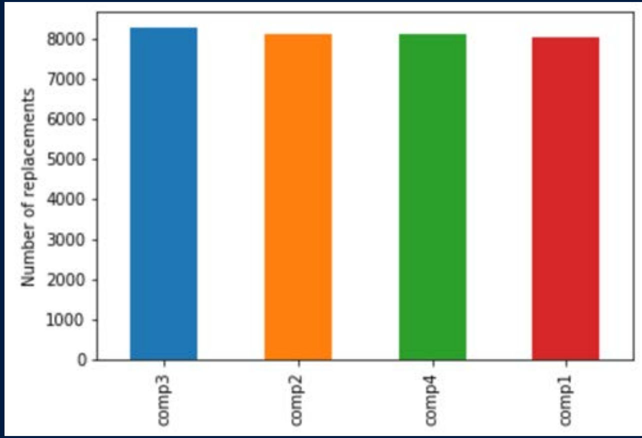
Number of records: 6726

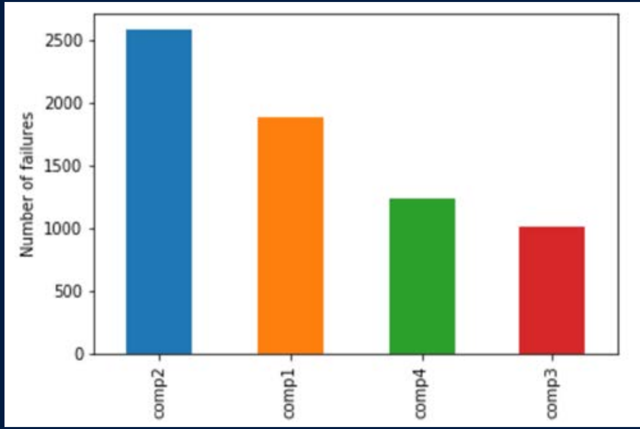
Experimental results: Exploratory Analysis











Experimental results: Feature engineering

	machineID	voltmean_12h	rotatemean_12h	pressuremean_12h	vibrationmean_12h	voltsd_12h	rotatesd_12h	pressuresd_12h	vibrationsd_12h	
count	729000.000000	729000.000000	729000.000000	729000.000000	729000.000000	729000.000000	729000.000000	729000.000000	729000.000000	7
mean	500.500000	170.759128	446.609585	100.836497	40.349084	14.707291	49.127996	9.854067	4.915801	
std	288.675188	5.712698	21.216195	5.232821	2.292131	3.193583	10.710131	2.219156	1.082040	
min	1.000000	148.918494	240.555739	86.499636	33.282624	3.132666	10.366925	2.231350	1.069356	
25%	250.750000	167.228630	438.433805	98.141429	39.074348	12.470419	41.643277	8.325662	4.163477	
50%	500.500000	170.244885	449.054520	100.145988	40.077281	14.575053	48.683220	9.735461	4.865620	
75%	750.250000	173.387910	459.156515	102.233761	41.123582	16.797384	56.090431	11.234719	5.609491	
max	1000.000000	225.106859	516.419586	166.739169	63.894733	35.667974	118.183889	30.070135	13.129524	

failure
none
none
none
none
none

4h	...	error1count	error2count	error3count	error4count	error5count	comp1	comp2	comp3	comp4	age
00	...	729000.000000	729000.000000	729000.000000	729000.000000	729000.000000	728000.000000	728000.000000	728000.000000	728000.000000	729000.000000
42	...	0.007494	0.008826	0.008421	0.004364	0.003690	53.257865	52.820198	51.466714	52.957393	9.862000
95	...	0.086480	0.093721	0.091545	0.066017	0.060633	59.152141	59.317405	57.241221	59.517514	6.089746
21	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.250000	0.250000	0.250000	0.000000
23	...	0.000000	0.000000	0.000000	0.000000	0.000000	13.250000	12.750000	12.750000	12.750000	5.000000
61	...	0.000000	0.000000	0.000000	0.000000	0.000000	32.250000	31.750000	30.750000	31.750000	10.000000
11	...	0.000000	0.000000	0.000000	0.000000	0.000000	69.250000	68.250000	67.250000	68.250000	15.000000
54	...	2.000000	2.000000	2.000000	2.000000	1.000000	438.750000	433.750000	398.750000	452.750000	20.000000

```
none     678103
comp2     19845
comp1     15068
comp4      8955
comp3      7389
Name: failure, dt
```

Experimental results: Model Building

Random Forest Classifier	Naive Bayes Classifier	Gradient Boosting Classifier	Neural Network Classifier

Random Forest Classifier

predicted_failure failure	comp1	comp2	comp3	comp4	none
comp1	1341	26	1	9	2556
comp2	18	2018	5	6	3189
comp3	8	18	583	4	1264
comp4	14	20	2	946	1385
none	61	91	10	32	172472

Naive Bayes Classifier

predicted_failure failure	comp1	comp2	comp3	comp4	none
comp1	1129	93	100	75	2536
comp2	94	1664	135	87	3256
comp3	21	32	780	9	1035
comp4	31	69	50	863	1354
none	2745	3653	4737	2787	158744

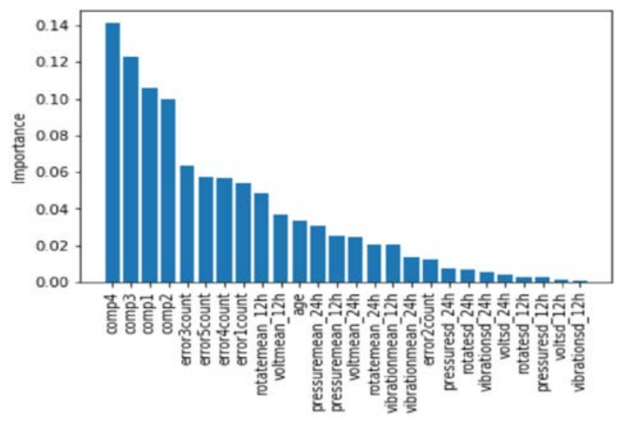
Gradient Boosting Classifier

predicted_failure failure	comp1	comp2	comp3	comp4	none
comp1	1530	52	11	37	2303
comp2	35	2107	19	23	3052
comp3	10	20	737	8	1102
comp4	12	24	6	1038	1287
none	184	175	68	96	172143

Neural Network Classifier

predicted_failure failure	none
comp1	3933
comp2	5236
comp3	1877
comp4	2367
none	172666

Model/ Measure	Random Forest Classifier	Naive Bayes Classifierssifier	Gradient Boosting Classifier	Neural Network Classifier
Training Time	2min 18s	951 ms	13min 2s	3min 14s
Accuracy	0.953144	0.876939	0.954191	NA
Precision	0.937656	0.231597	0.874031	NA
Recall	0.368017	0.351589	0.411371	NA
F1	0.528575	0.279248	0.559438	NA



Q&A