FREE
PPT TEMPLATES

INSERT THE TITLE
OF YOUR PRESENTATION HERE

LogoType™

Click to add title

Standard (4:3)

This PowerPoint Template has clean and neutral design that can be adapted to any content and meets various market segments. With this many slides you are able to make a complete PowerPoint Presentation that best suit your needs.

This PowerPoint Template has clean and neutral design that can be adapted to any content and meets various market segments. With this many slides you are able to make a complete PowerPoint Presentation that best suit your needs.

This PowerPoint Template has clean and neutral design that can be adapted to any content and meets various market segments. With this many slides you are able to make a complete PowerPoint Presentation that best suit your needs.

This PowerPoint Template has clean and neutral design that can be adapted to any content and meets various market segments. With this many slides you are able to make a complete PowerPoint Presentation that best suit your needs.

Click to add title

- Standard (4:3)
  - This PowerPoint Template has clean and neutral design that can be adapted to any content and meets various market segments. With this many slides you are able to make a complete PowerPoint Presentation that best suit your needs.
  - This PowerPoint Template has clean and neutral design that can be adapted to any content and meets various market segments. With this many slides you are able to make a complete PowerPoint Presentation that best suit your needs.
  - This PowerPoint Template has clean and neutral design that can be adapted to any content and meets various market segments. With this many slides you are able to make a complete PowerPoint Presentation that best suit your needs.
  - This PowerPoint Template has clean and neutral design that can be adapted to any content and meets various market segments.

Predictive Maintenance

Exploratory Data Analysis
Regression Modeling
Binary Classification
Multiclass Classification

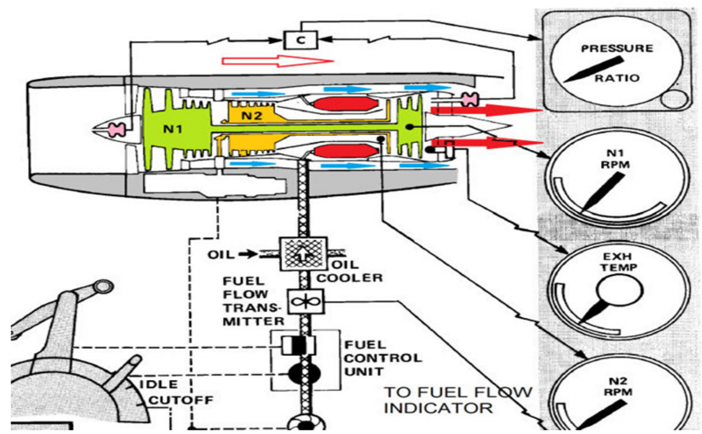
Sami Mustafa

The Problem

Airlines are interested in predicting engines failures in advance to enhance operations and reduce flight delays. Observing engine's health and condition through sensors and telemetry data is assumed to facilitate this type of maintenance by predicting Time-To-Failure (TTF) of in-service equipment. Using aircraft engine's sensors measurements, can we predict engine's TTF?

Solution Approach

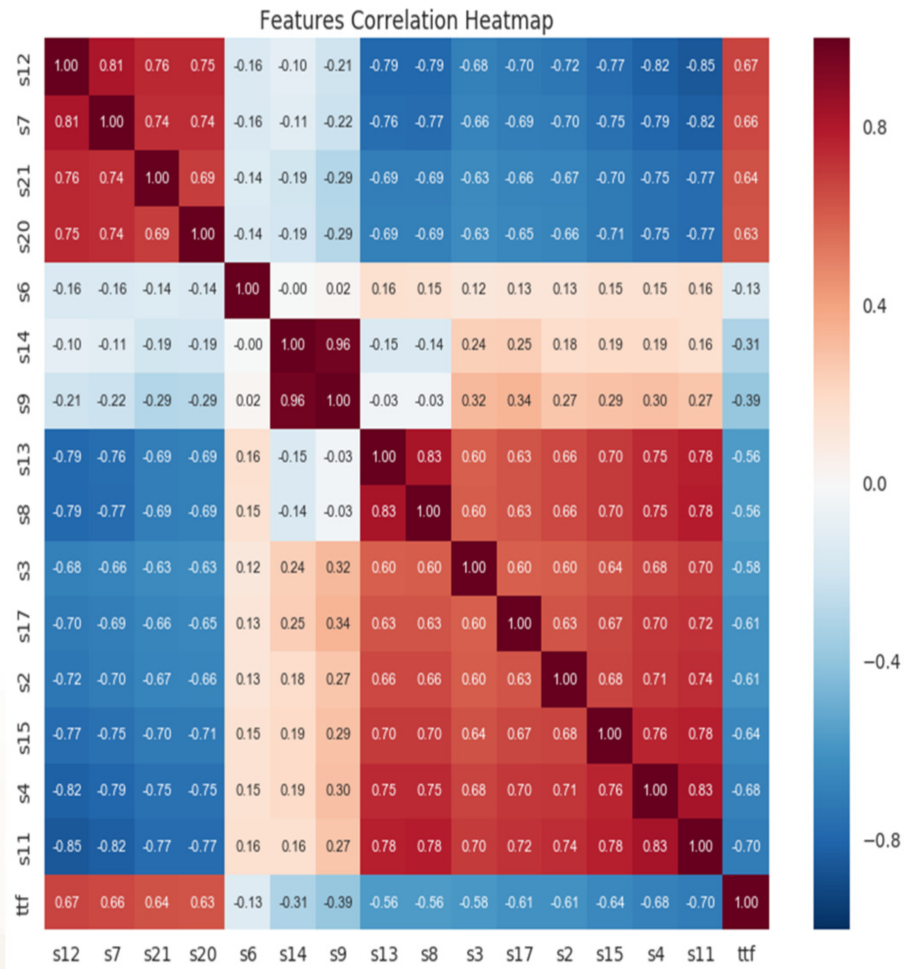
By exploring aircraft engine’s sensor values over time, machine learning algorithm can learn the relationship between sensor values and changes in sensor values to the historical failures in order to predict failures in the future.

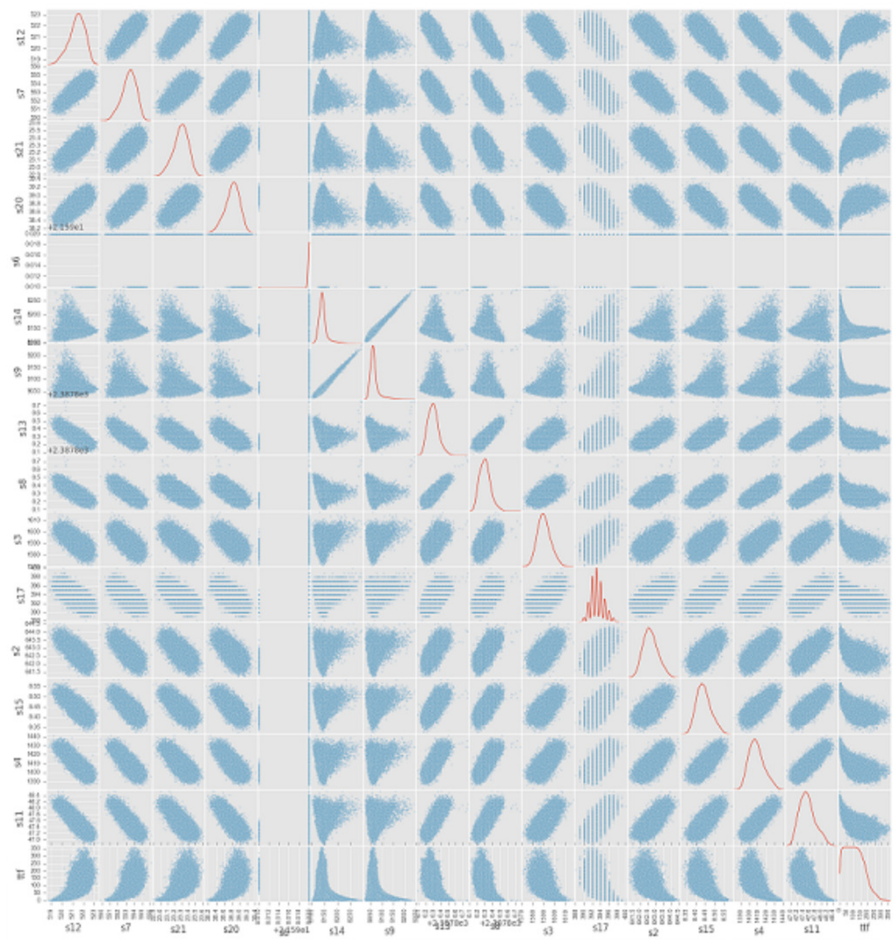


Data

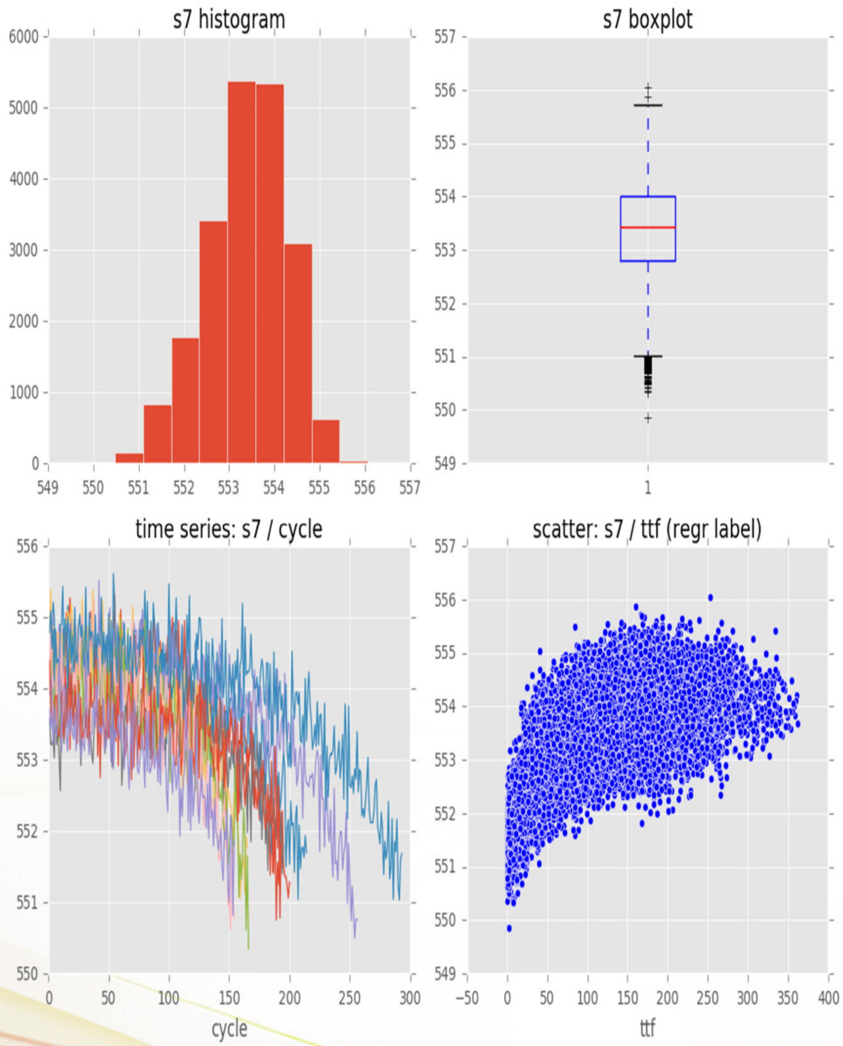
Text files contain simulated aircraft engine run-to-failure events, operational settings, and 21 sensors measurements are provided by Microsoft. It is assumed that the engine progressing degradation pattern is reflected in its sensor measurements.

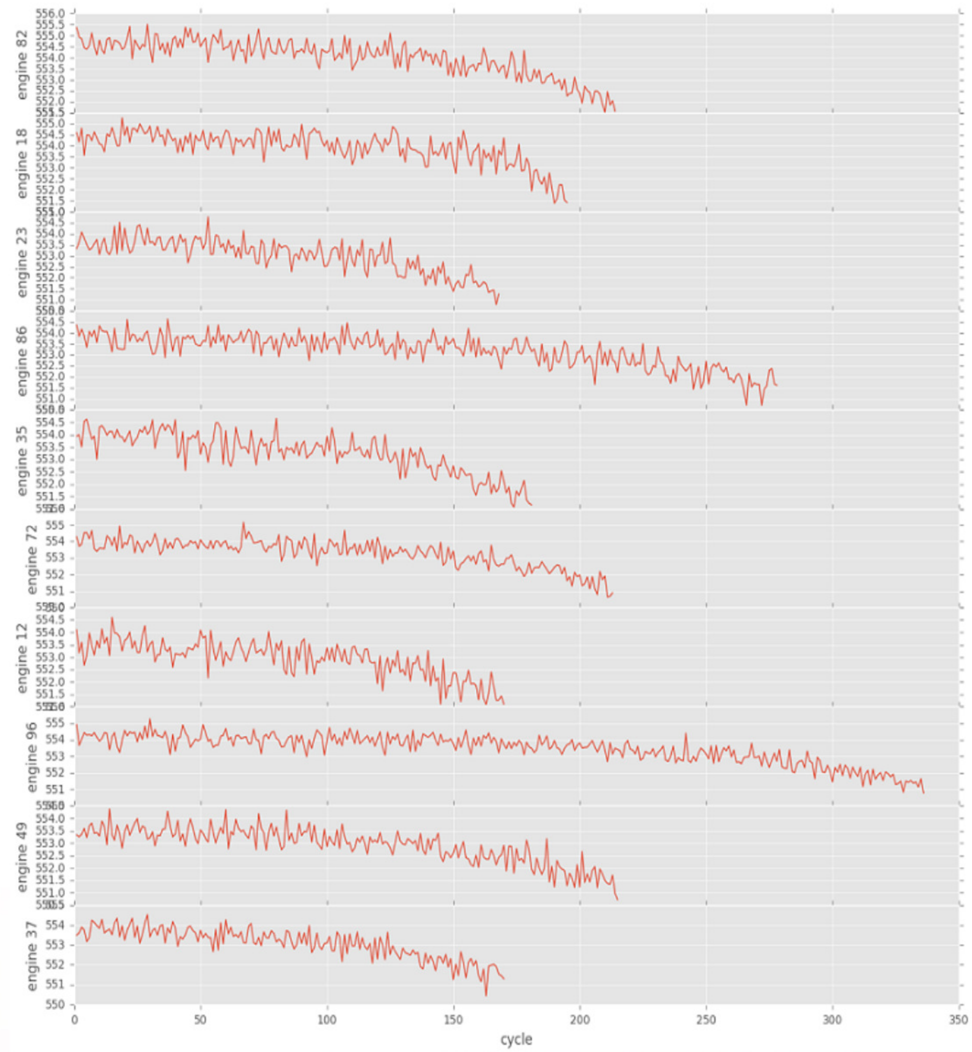
Exploratory Data Analysis





Exploratory Data Analysis (S7)





Regression Modeling

Used to predict the number remaining cycles before engine failure

Machine learning algorithms used:

- ✓ Linear Regression
- ✓ LASSO Regression
- ✓ Ridge Regression
- ✓ Decision Tree Regression
- ✓ Polynomial Regression
- ✓ Random Forest Regression

Regression metrics calculated:

- ✓ R-squared (R^2)
- ✓ Root Mean Squared Error (RMSE)
- ✓ Mean Absolute Error
- ✓ Explained Variance

Regression Modeling Results

	Linear	LASSO	Ridge	Decision Tree	Polynomial	Random Forest
Root Mean Squared Error	32.04	31.97	31.97	32.10	29.68	28.63
Mean Absolute Error	25.59	25.55	25.54	24.32	22.38	23.17
R^2	0.41	0.41	0.41	0.40	0.49	0.53
Explained Variance	0.67	0.67	0.67	0.63	0.65	0.77

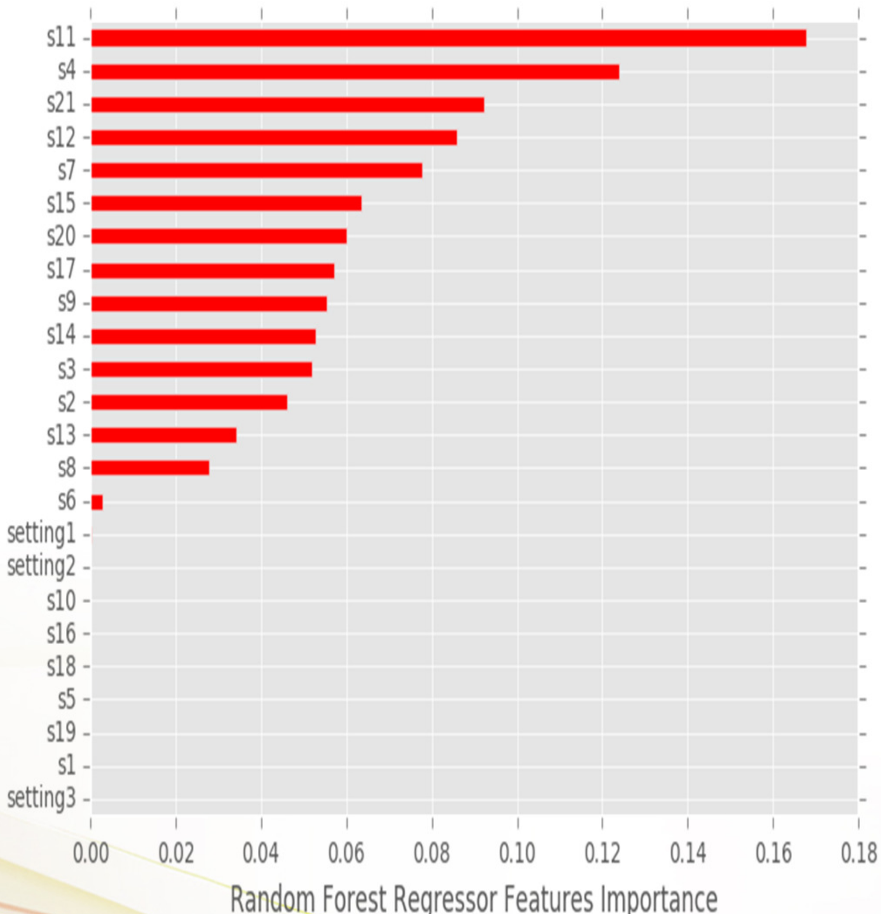
Random Forest Regressor is likely to perform better than other models

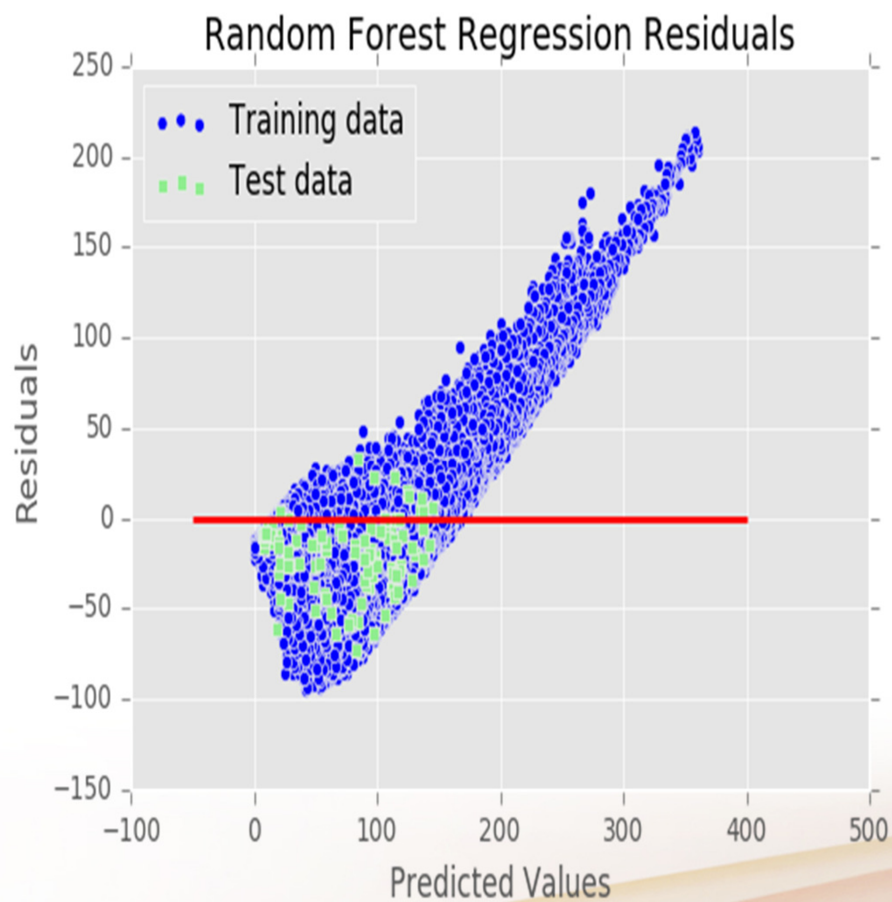
Random Forest Regressor predictions vs. actual for first 10 samples:

	0	1	2	3	4	5	6	7	8	9	10
Actual	112.00	98.00	69.00	82.00	91.00	93.00	91.00	95.00	111.00	96.00	97.00
Prediction	151.58	119.27	74.42	96.47	112.59	130.28	128.11	100.69	116.12	127.37	74.37
Difference	-39.58	-21.27	-5.42	-14.47	-21.59	-37.28	-37.11	-5.69	-5.12	-31.37	22.63

Regression Modeling

Random Forest Regressor: Feature Importance and Residuals





Binary Classification

Used to predict if the engine will fail within specific cycles window or not

Machine learning algorithms used:

- ✓ Logistic Regression
- ✓ Decision Trees
- ✓ Support Vector Machines
- ✓ Linear Support Vector
- ✓ K Nearest Neighbors
- ✓ Gaussian Naive Bayes
- ✓ Random Forests

Classification metrics calculated:

- ✓ Area Under the Curve ROC (AUC ROC)
- ✓ Precision
- ✓ Recall
- ✓ F1 Score
- ✓ Accuracy

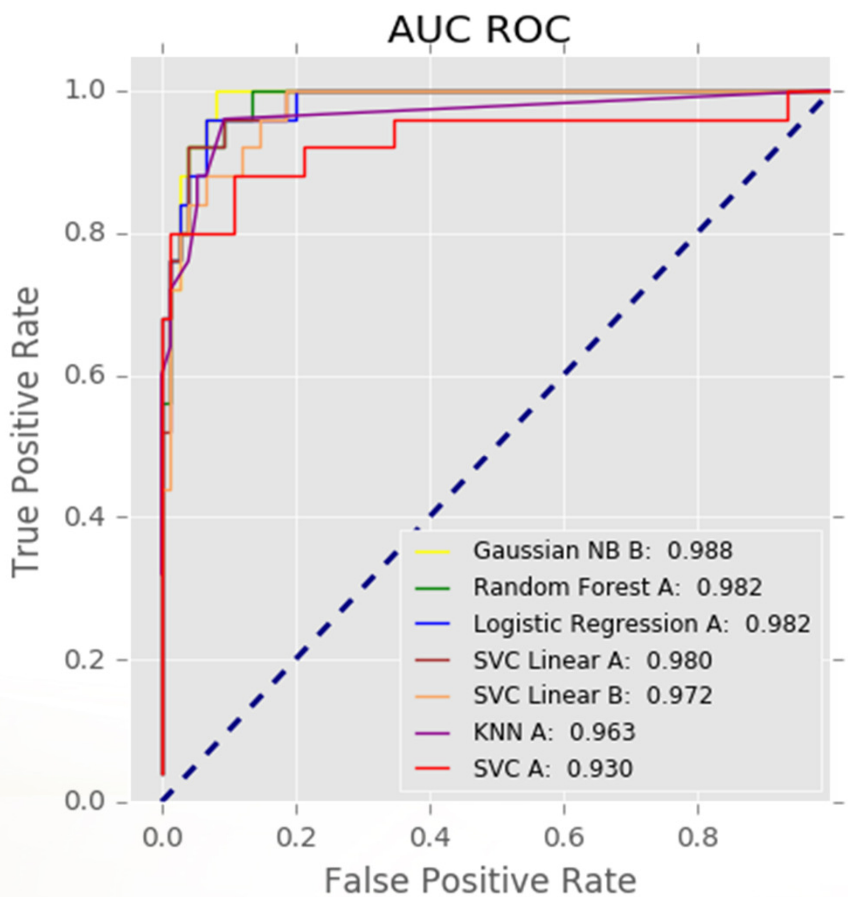
Binary Classification Results

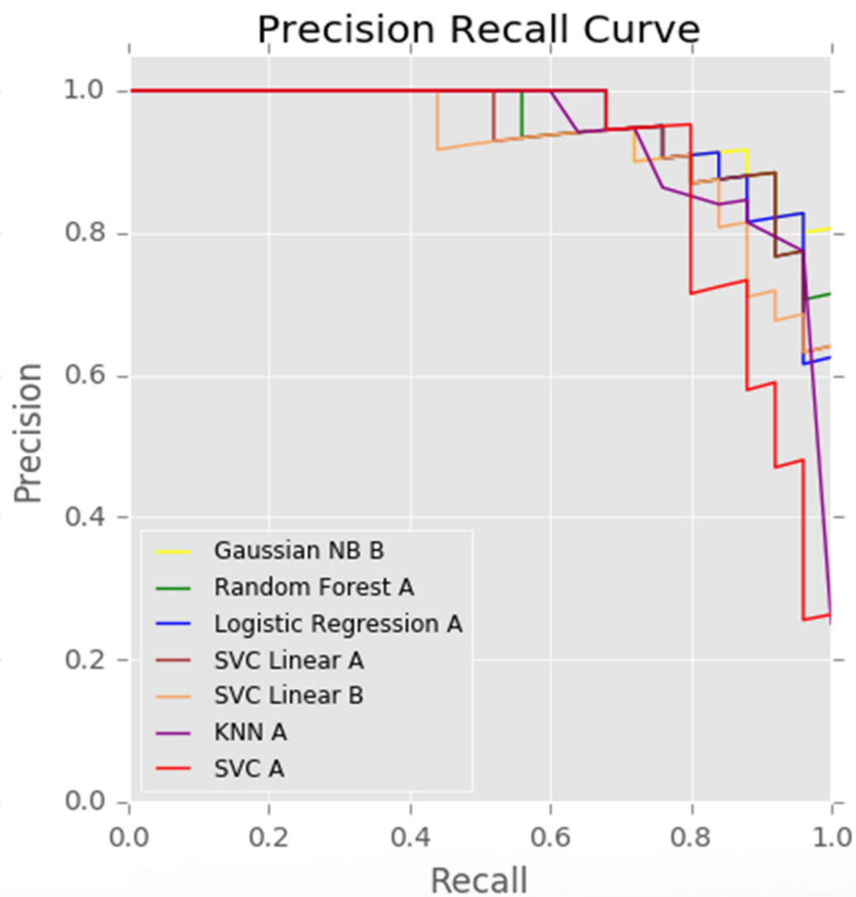
	Logistic Regression B	Logistic Regression A	Decision Tree B	Decision Tree A	Random Forest B	Random Forest A	SVC B	SVC A	SVC Linear B	SVC Linear A	KNN B	KNN A	Gaussian NB B	Gaussian NB A
ROC AUC	0.9803	0.9819	0.9451	0.9629	0.9803	0.9824	0.8917	0.9301	0.9717	0.9797	0.9352	0.9635	0.9877	0.9805
Precision	0.9333	1.0000	0.9333	0.9474	0.9444	0.9444	0.9444	0.9474	1.0000	0.4310	0.9444	0.9474	0.8276	0.8276
F1 Score	0.7000	0.8095	0.7000	0.8182	0.7907	0.7907	0.7907	0.8182	0.5714	0.6024	0.7907	0.8182	0.8889	0.8889
Accuracy	0.8800	0.9200	0.8800	0.9200	0.9100	0.9100	0.9100	0.9200	0.8500	0.6700	0.9100	0.9200	0.9400	0.9400
Recall	0.5600	0.6800	0.5600	0.7200	0.6800	0.6800	0.6800	0.7200	0.4000	1.0000	0.6800	0.7200	0.9600	0.9600

A: **A**fter Feature Engineering **B**: **B**efore Feature Engineering

- ✓ Naïve Bayes and Random Forests scoring best results for AUC are likely to perform better than others
- ✓ Most of the binary classifiers showed better performance metrics with addition of new features
- ✓ Linear SVC has totally different scoring metrics before and after FE, switching between Precision and Recall

Binary Classification Results





Naïve Bayes and Random Forests have the best curves for AUC ROC and Precision-Recall

Expected Profit

Rank	Profit	Model	Queue	Threshold	TP	FP	TN	FN	TPR	FPR	TNR	FNR
0	19.00	Gaussian NB B	0.31	0.09	25	0	69	6	1.00	0.08	0.92	1.00
1	18.69	Logistic Regression B	0.28	0.11	24	1	71	4	0.96	0.05	0.95	0.99
2	18.69	Gaussian NB A	0.28	0.97	24	1	71	4	0.96	0.05	0.95	0.99
3	17.70	Logistic Regression A	0.29	0.06	24	1	70	5	0.96	0.07	0.93	0.99
4	17.35	Random Forest A	0.26	0.18	23	2	72	3	0.92	0.04	0.96	0.97
5	17.35	SVC Linear A	0.26	0.78	23	2	72	3	0.92	0.04	0.96	0.97
6	17.00	Random Forest B	0.33	0.10	25	0	67	8	1.00	0.11	0.89	1.00
7	15.72	KNN A	0.31	0.08	24	1	68	7	0.96	0.09	0.91	0.99
8	13.05	SVC Linear B	0.27	(0.59)	22	3	70	5	0.88	0.07	0.93	0.96
9	12.16	SVC A	0.21	(0.23)	20	5	74	1	0.80	0.01	0.99	0.94
10	12.08	SVC B	0.28	(0.94)	22	3	69	6	0.88	0.08	0.92	0.96
11	10.70	KNN B	0.26	0.31	21	4	70	5	0.84	0.07	0.93	0.95
12	10.14	Decision Tree A	0.30	0.18	22	3	67	8	0.88	0.11	0.89	0.96
13	7.82	Decision Tree B	0.29	0.08	21	4	67	8	0.84	0.11	0.89	0.94

Algorithms have been ranked based on expected profit/engine calculations (unconstrained queue):

Expected Profit = Prop(+ve) x [TPR x benefit(TP) + FNR x cost(FN)] + Prob(-ve) x [TNR x benefit(TN) + FPR x cost(FP)]

100 samples:
prob_positive = 0.25, prop_negative = 0.75, TPb = $300, TNb = $0, FPc = $-100, FNc = $-200

Multiclass Classification

Used to predict in which cycles window will the engine fail

xx.

Next Steps

XX

xx.