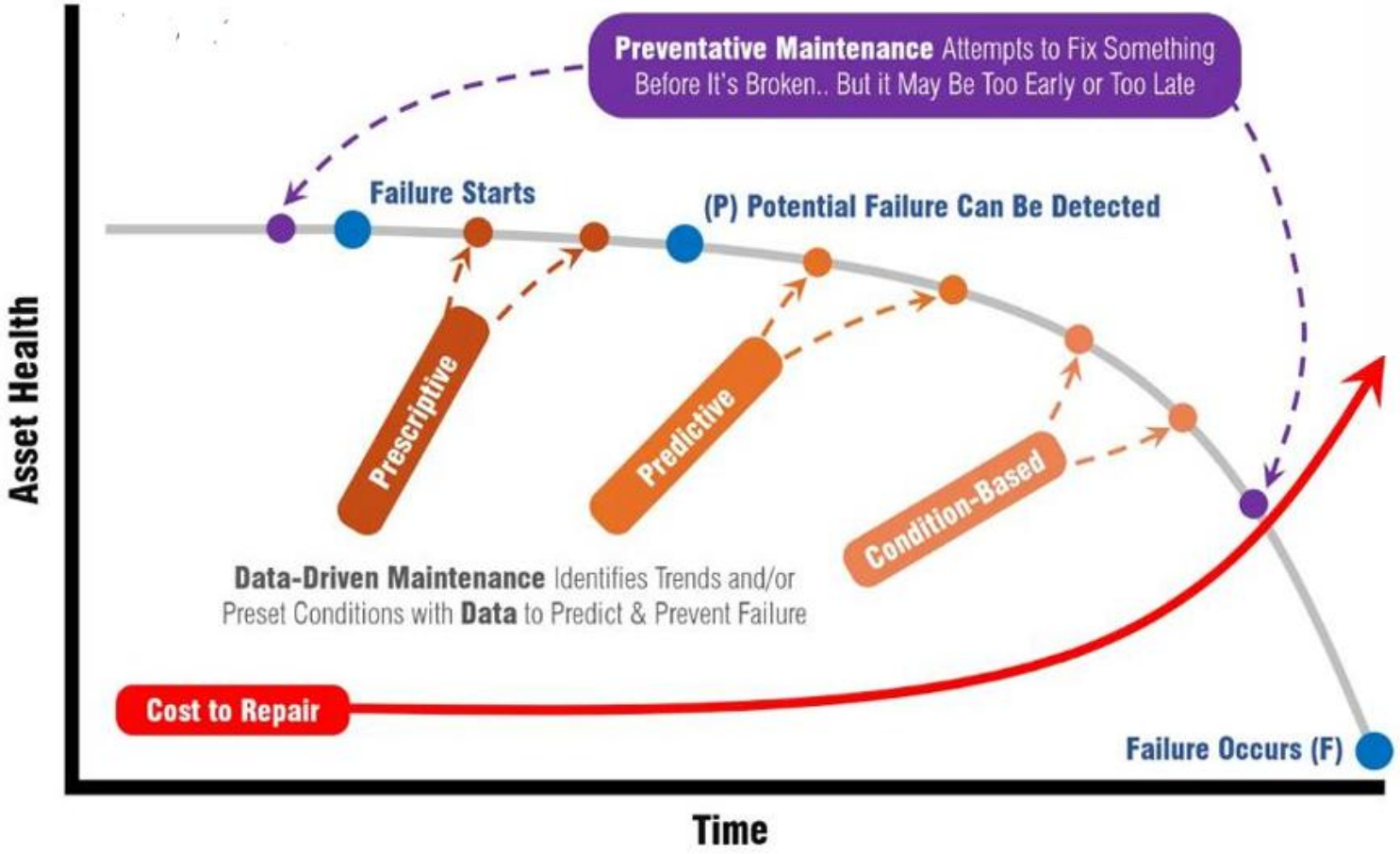
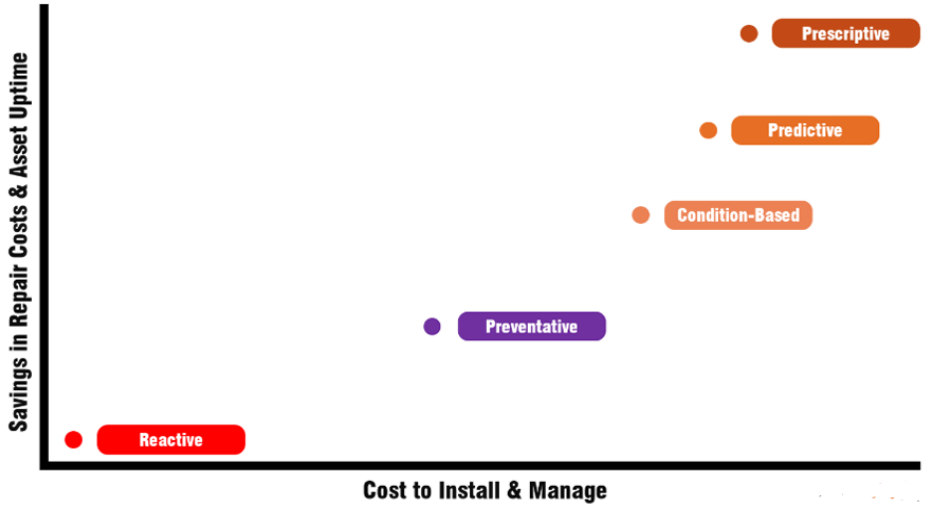
Device Failure Analysis

ML in device failure prediction

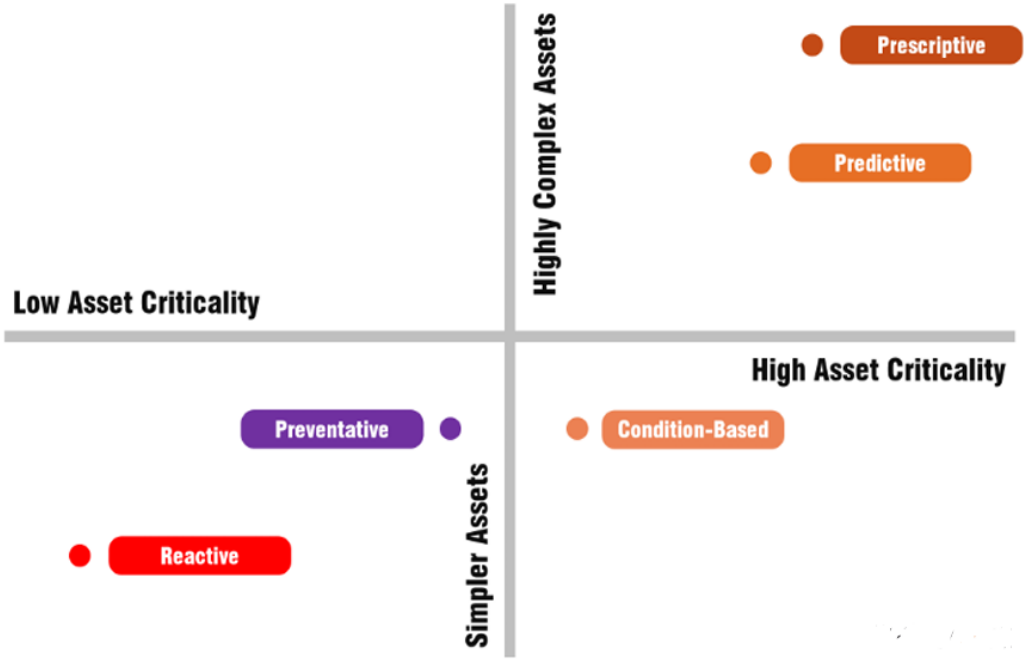
- Machine Learning (ML) allows for predictions from vast amount of data -> *increased reliability*
- ML allows engineers to take proactive steps by providing early warning signs of impending failures/malfunctions -> *proactive maintenance*
- ML allows savings in time and money by reducing downtime and increasing productivity -> *improved efficiency*

Predictive maintenance offers more cost savings as compared to preventive maintenance



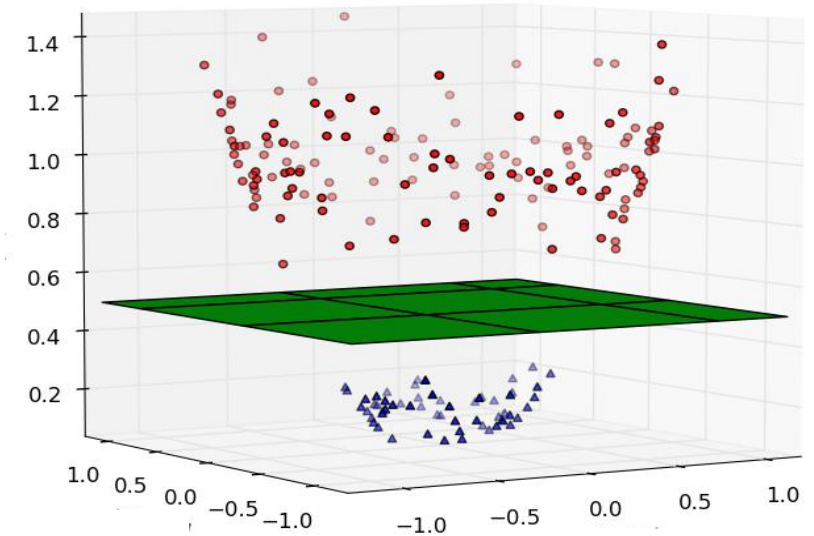


Which asset type needs what kind of maintenance ?

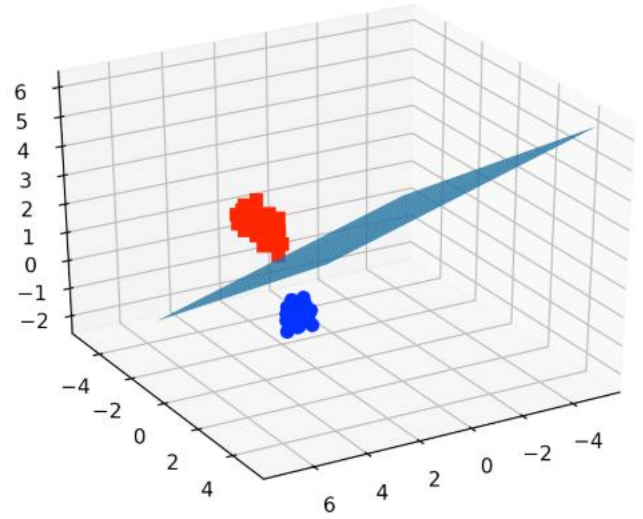


Methodology

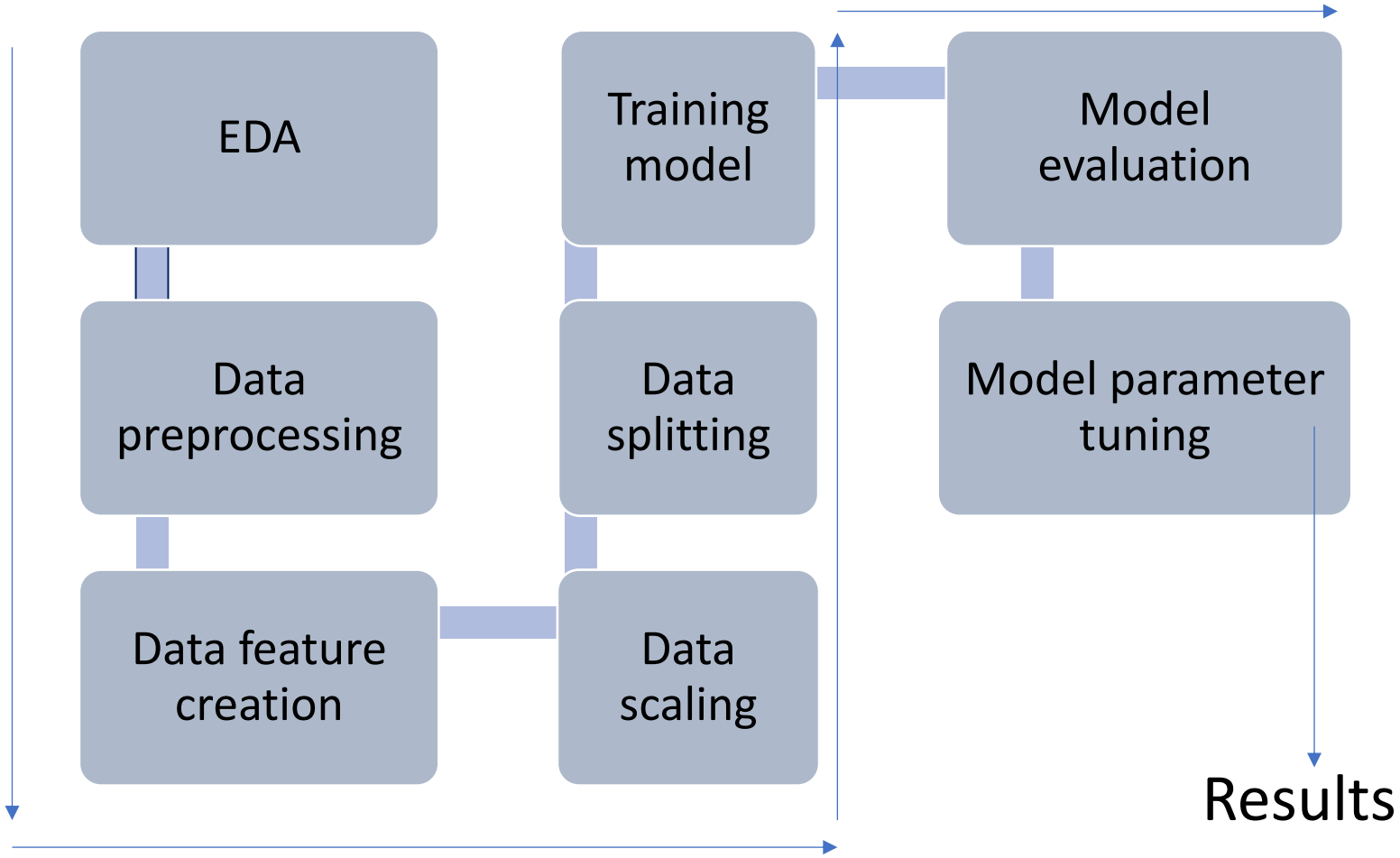
- Supervised learning: classification
- Model training data is labelled and the data provided is of 1 year (2015)
- Response class in data is binary, target or response is termed ‘malfunction’: 0 is non-failure and 1 is failure
- Logistic Regression is explored to predict malfunction/failure using data of a fleet of devices/products



#	Column	Non-Null Count	Dtype
0	date	124494 non-null	object
1	product	124494 non-null	object
2	malfunction	124494 non-null	int64
3	feature1	124494 non-null	int64
4	feature2	124494 non-null	int64
5	feature3	124494 non-null	int64
6	feature4	124494 non-null	int64
7	feature5	124494 non-null	int64
8	feature6	124494 non-null	int64
9	feature7	124494 non-null	int64
10	feature8	124494 non-null	int64
11	feature9	124494 non-null	int64



Approach


EDA
Data preprocessing
Data feature creation
Data splitting
Data scaling
Training model
Model evaluation
Model parameter tuning
Results

Challenges

- Data understanding, especially when there is imbalance
- Dataset is a mix of qualitative and quantitative data
- Minimizing FPs and FNs
- Choosing the right metric to validate the model for business value-addition

Discussion

- Very high model accuracy score may mean either too few observations or training examples or too high regularization of features

- Feature Engineering is essential when we have data with large number of features with proper understanding of each from a domain or SM expert, and how they can impact device failure.

solver	penalty	multinomial multiclass
'lbfgs'	'l2', None	yes
'liblinear'	'l1', 'l2'	no
'newton-cg'	'l2', None	yes
'newton-cholesky'	'l2', None	no
'sag'	'l2', None	yes
'saga'	'elasticnet', 'l1', 'l2', None	yes