

Lead Scoring Case Study

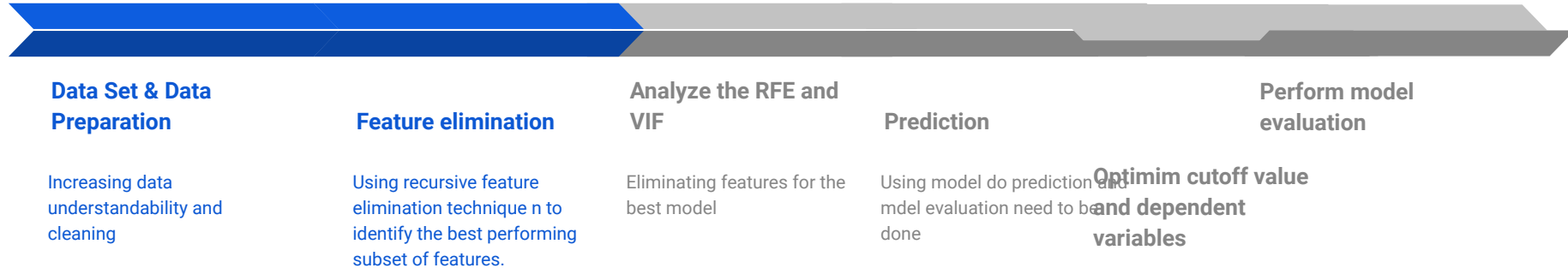
Hitha Mv
Anandu Mv

Business Objective

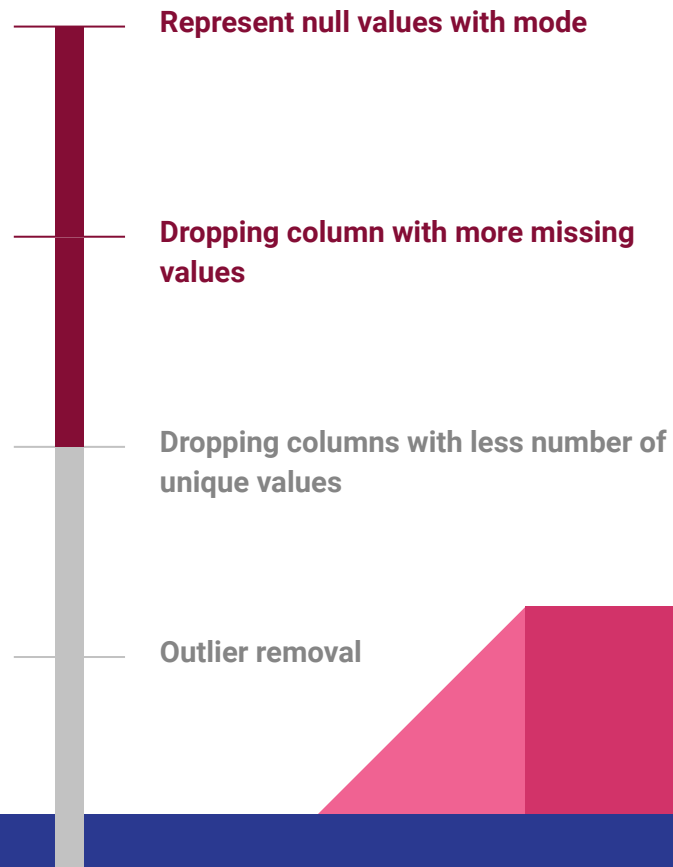
- To understand X Education to select the most promising leads, i.e. promising customers.
- To develop a model to assign a lead score value between 0 and 100 to each of the customer and can be used by the company to target potential customers.
- Model should be able to adjust with companys future requirement

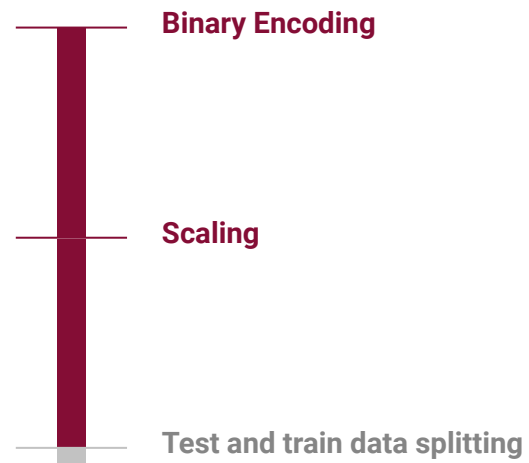


Problem Solving Methodology



Data Set & Data Preparation





Feature Elimination

Recursive feature elimination

```
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression()

from sklearn.feature_selection import RFE
rfe = RFE(logreg, 20)          # running RFE with 15 variables as output
rfe = rfe.fit(X_train, y_train)
```

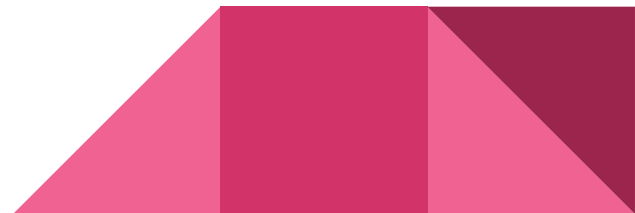
Predicting Probability

Predicted
probabilities

In the given
sce

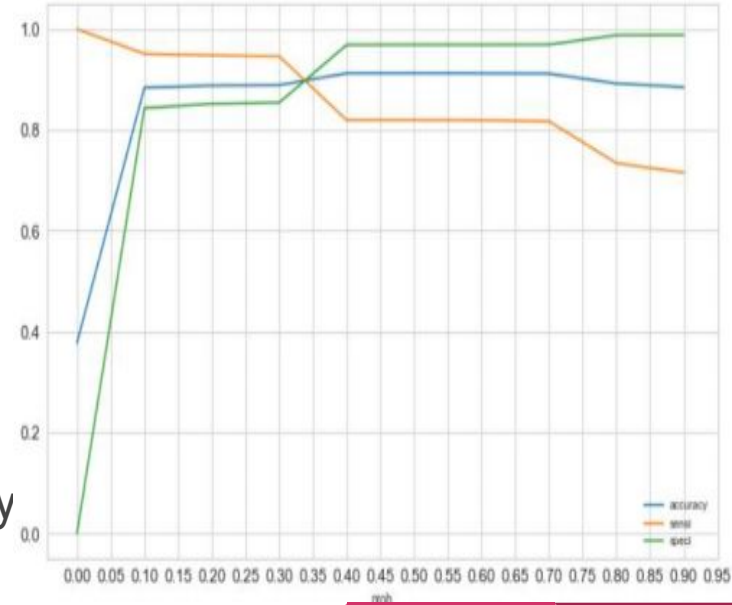
	Converted	Converted_prob	Prospect ID
0	0	0.001180	7551
1	1	0.923196	8564
2	0	0.027128	4354
3	1	0.997553	112
4	1	0.604931	7270

	Converted	Converted_prob	Prospect ID	predicted
0	0	0.001180	7551	0
1	1	0.923196	8564	1
2	0	0.027128	4354	0
3	1	0.997553	112	1
4	1	0.604931	7270	1



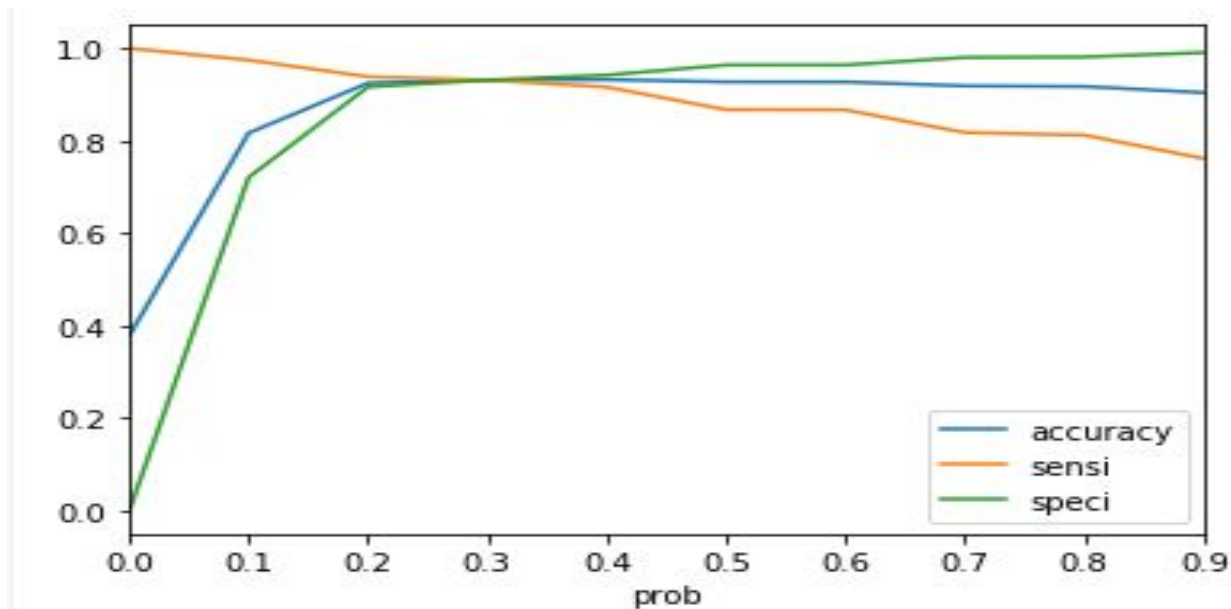
Finding optimal probability curve

- The accuracy sensitivity and specificity was calculated for various values of probability threshold and plotted in the graph to the right.
- From the curve above, 0.33 is found to be the optimum point for cutoff probability.
- At this threshold value, all the 3 metrics - accuracy sensitivity and specificity was found to be well above 80% which is a well acceptable value.



Plotting ROC

Receiver Operating Characteristics (ROC) Curve



Making predictions on data set

The final model on the train dataset is used to make predictions for the test dataset

	Prospect ID	Converted	Converted_prob	final_predicted
0	6233	0	0.009222	0
1	339	0	0.001551	0
2	2519	1	0.994833	1
3	2410	0	0.104106	0
4	1534	0	0.179586	0



Lead score calculation

- The train and test dataset is concatenated to get the entire list of leads available.
- Higher the lead score, higher is the probability of a lead getting converted and vice versa,

	Prospect ID	Converted	Converted_prob	final_predicted
0	6233	0	0.009222	0
1	339	0	0.001551	0
2	2519	1	0.994833	1
3	2410	0	0.104106	0
4	1534	0	0.179586	0

