

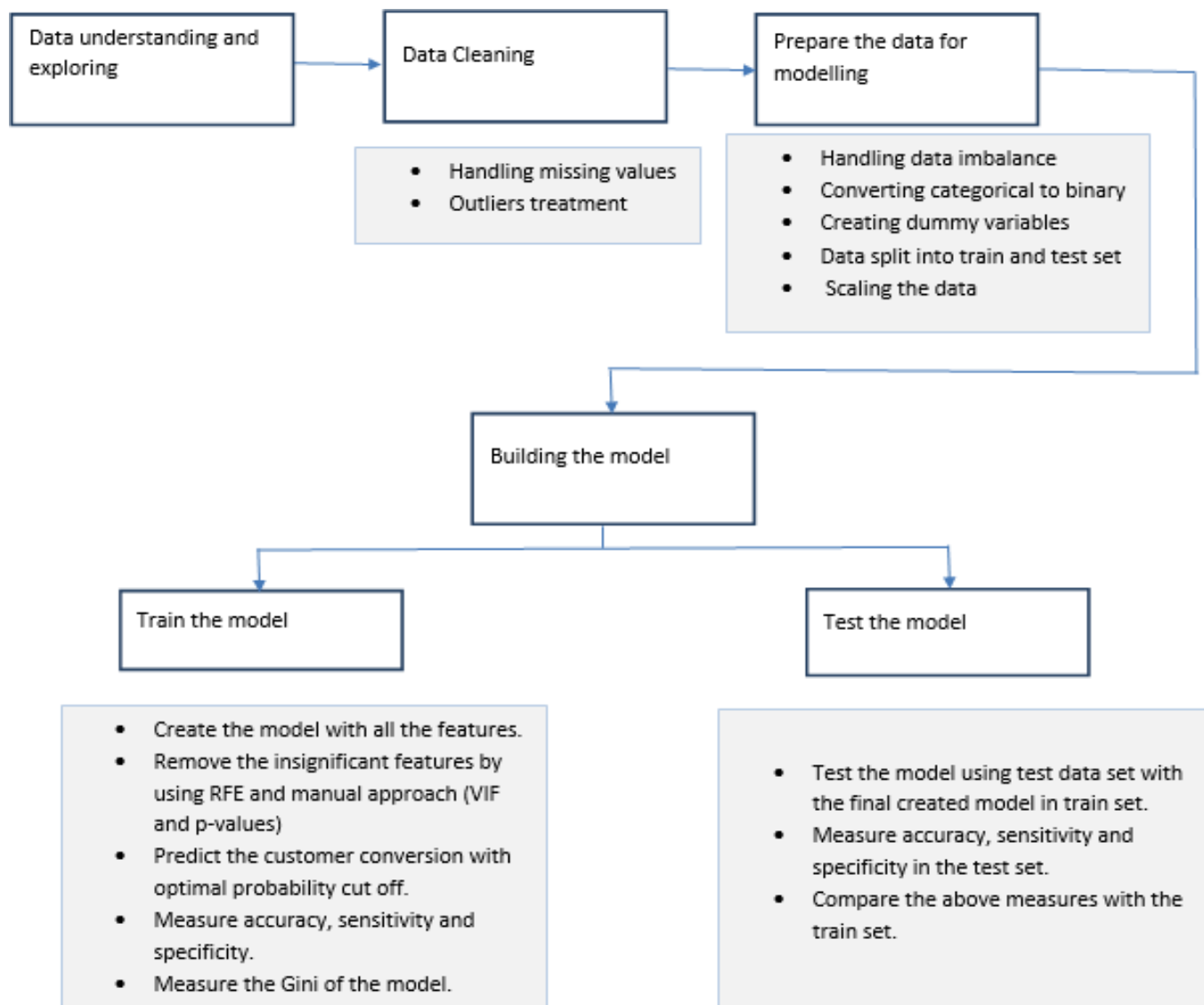
Summary Report

Problem statement: -

Identify the set of leads of X Education so that the lead conversion rate should go up a sales team of the company focus more on communication with the potential leads rather than making calls to every customer.

Analysis approach: -

Flow diagram of step by step approach.

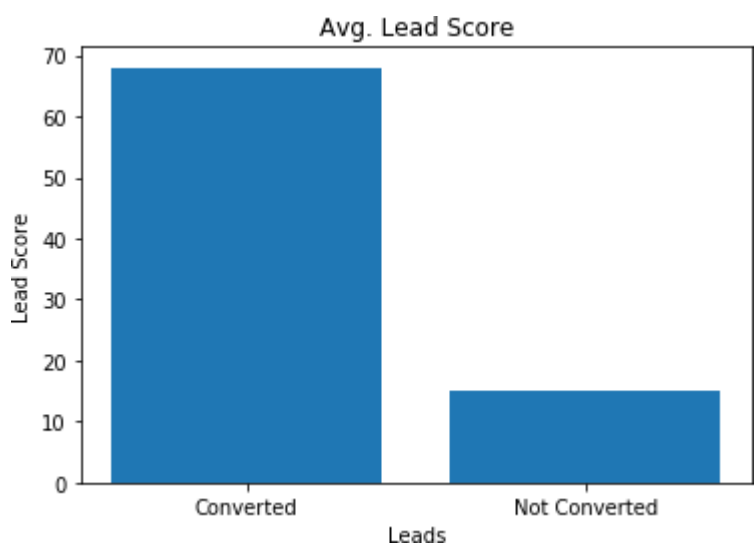


Model outcome: -

Optimum probability cut off: - 0.35

Lead score: -

The lead score, calculated from a model, predicts customer conversion. A score above 35 indicates likely conversion. Higher scores suggest higher conversion chances. Converted leads average a score of 68, while non-converted leads average 15.



Features of the final model: -

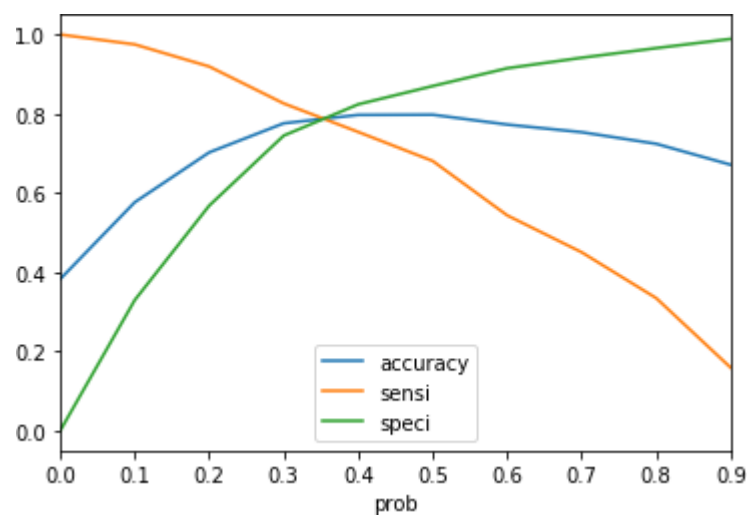
Below are the significant features of the final model arranged in the descending order impact on the basis of their coefficients.

Features	Coefficient	Impacts
Lead Source_Welingak Website	4.5256	Positive
Lead Source_Reference	3.1268	Positive
Last Activity_Olark Chat Conversatio	n-2.8298	Negative
Last Activity_Converted to Lead	-2.4484	Negative
Last Activity_Form Submitted on We	bs-1it.e9713	Negative
Last Activity_Email Bounced	-1.8597	Negative
Last Activity_Email Link Clicked	-1.6500	Negative
Last Activity_Page Visited on Website	-1.5718	Negative
Do Not Email	-1.3403	Negative
Lead Source_Direct Traffic	-1.3241	Negative
Lead Source_Referral Sites	-1.2794	Negative

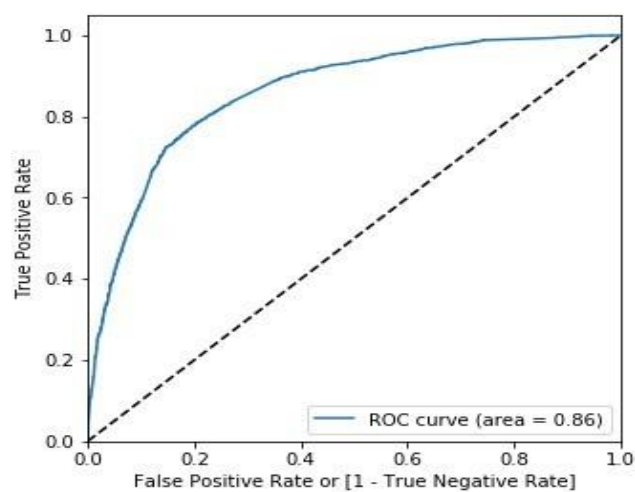
Total Time Spent on Website	1.1267	Positive
Last Activity_Email Opened	-1.1030	Negative
Lead Source_Organic Search	-1.0773	Negative
Lead Source_Google	-0.9545	Negative

Important measures of the model: -

Measures	Train set	Test set
Accuracy	0.79	0.79
Sensitivity	0.79	0.79
Specificity	0.78	0.78



Gini of the model: -



Conclusion: -

The model has good accuracy, sensitivity and specificity.

Recommendation for higher conversion rate: -

Leads likely to convert:

- Lead score exceeds 68.
- Over 12 hours spent on the website.
- Lead sources are Welingak Website and Reference.

Leads unlikely to convert:

- Customers chose the 'Do not email' option.
- Lead score is below 15.
- Less than 5 hours spent on the website.
- Lead sources include Direct Traffic, Referral Sites, Organic Search, and Google.
- Last customer activity includes 'Olark chat conversation', 'page visited on website', 'Email bounced', 'Form submitted on the website', or 'Email link clicked'.

OBSERVATIONS:

Key points for data preparation and model building:

- Treat missing values and eliminate outliers in the data.
- Address significant data imbalance in features by dropping or merging imbalanced values.
- Ensure all features are on the same scale.
- Avoid multicollinearity among variables.
- Determine the optimal probability cut-off for a balance between Sensitivity and Specificity, ensuring good Accuracy.
- The model should exhibit strong performance on the test set in terms of Sensitivity, Specificity, and Accuracy.