Lead Score Case Study

Lead Score Case Study for Education

Problem Statement :

X Education sells online courses to professionals in the sector. On numerous websites and search engines like Google, the firm advertises its courses.
When arriving at the website, these visitors may browse the courses, submit a form for the course, or watch some videos. These persons are categorized as leads when they fill out a form with their phone number or email address. Moreover, the organisation also gets leads through past referrals.
Once these leads are acquired, individuals from the sales team start making calls, composing emails, etc. Some leads are converted during this procedure, but most are not. At X Education, the normal lead conversion rate is roughly 30%.

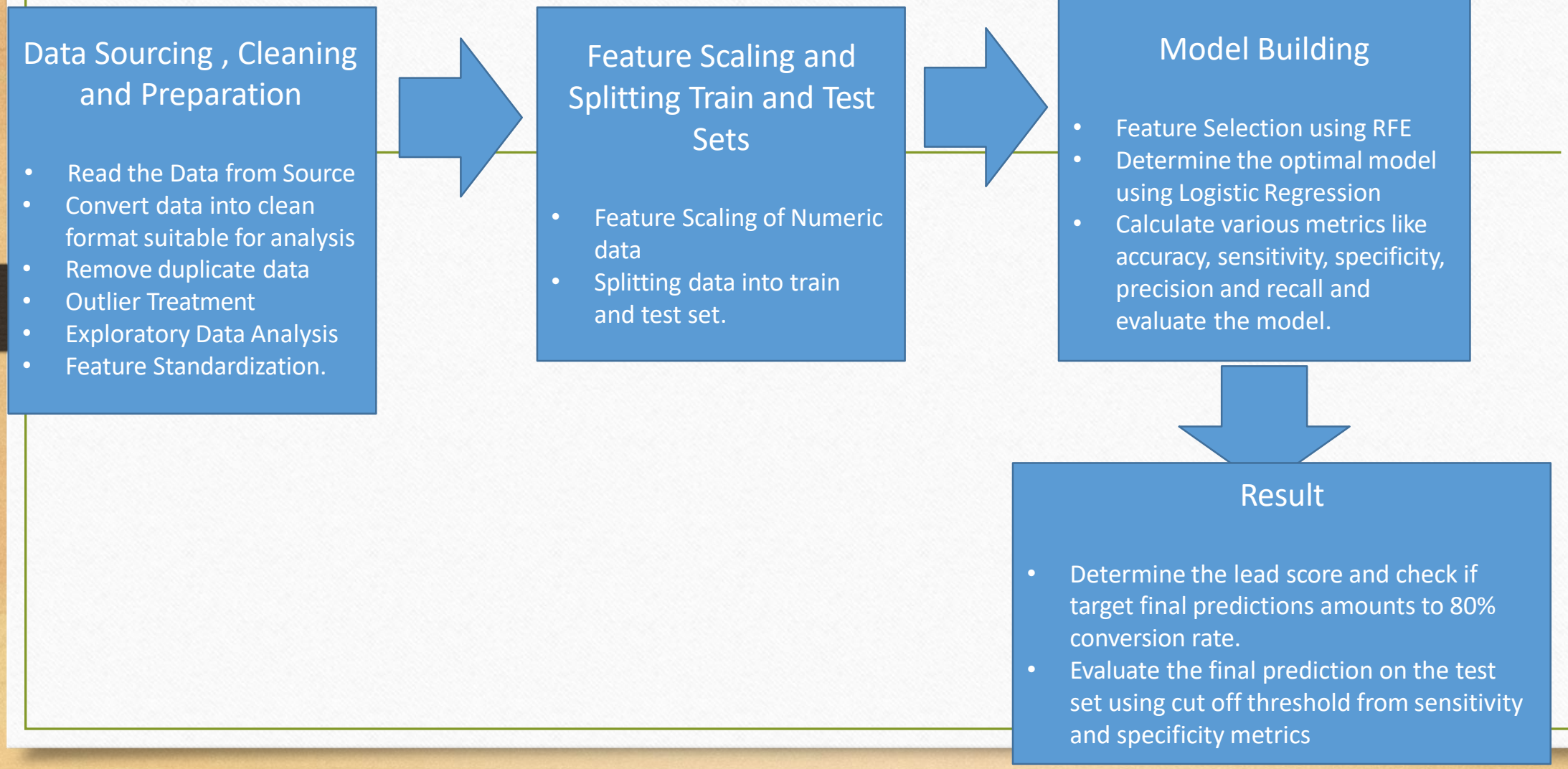
Business Goal:

- In order to choose the leads that have the best chance of becoming paying clients, or the most promising prospects, X Education requires assistance.
- The business needs a model where each lead is given a lead score, and leads with higher lead scores have a better chance of converting, while leads with lower lead scores have a lower chance of converting.
- The desired lead conversion rate has been estimated by the CEO to be in the range of 80%

Strategy

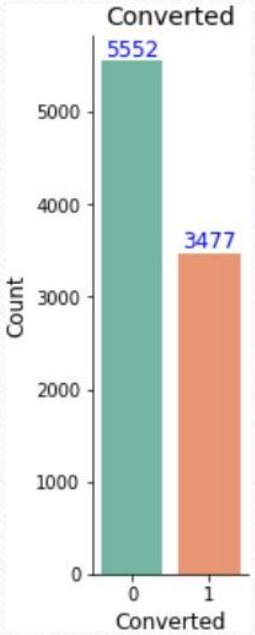
- Where to find the data for analysis Exploratory Data Analysis after data preparation and cleaning.
- Scaling of Feature
- dividing the dataset into a Train and Test dataset.
- Construction of a logistic regression model and determination of Lead Score.
- Assessing the model using several metrics, such as precision and recall or specificity and sensitivity.
- Using the most appropriate model in test data according to the sensitivity and specificity metrics.

Problem solving methodology

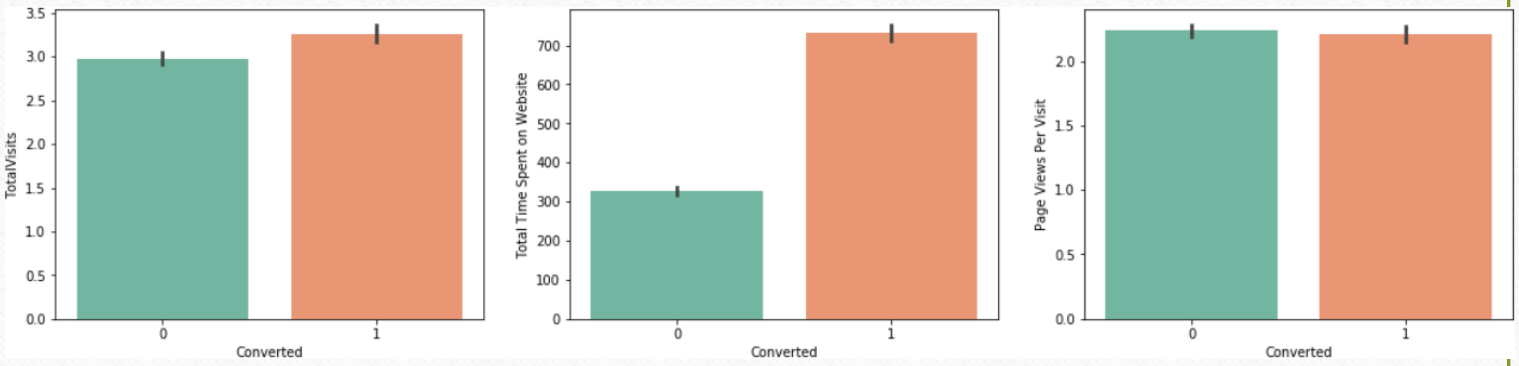


Exploratory Data Analysis

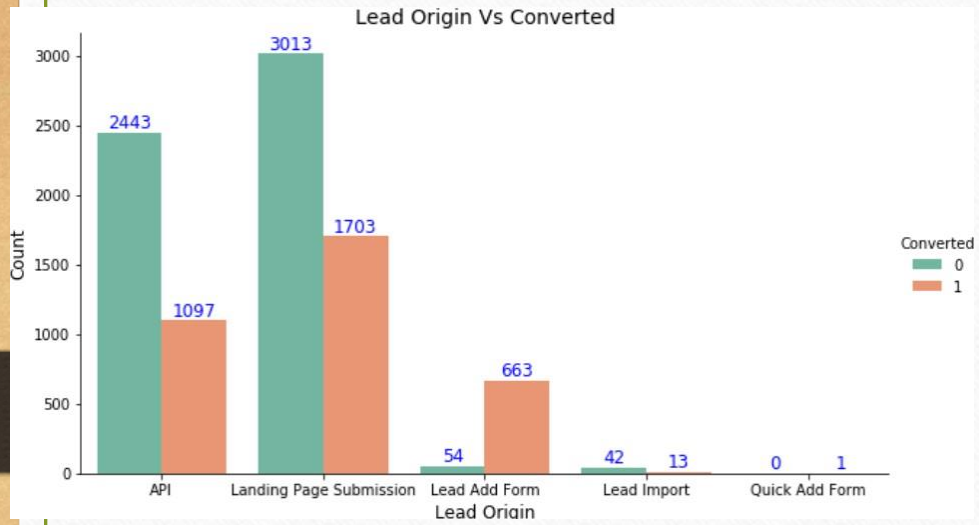
We have around 39% Conversion rate in Total



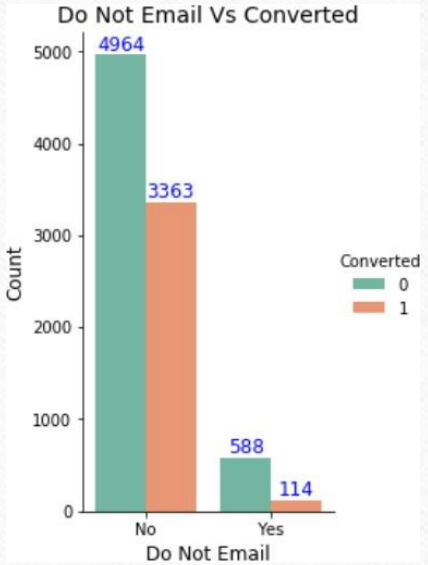
The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit

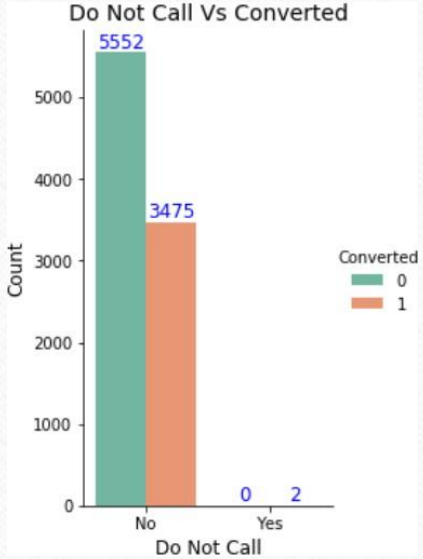


In Lead Origin, maximum conversion happened from Landing Page Submission

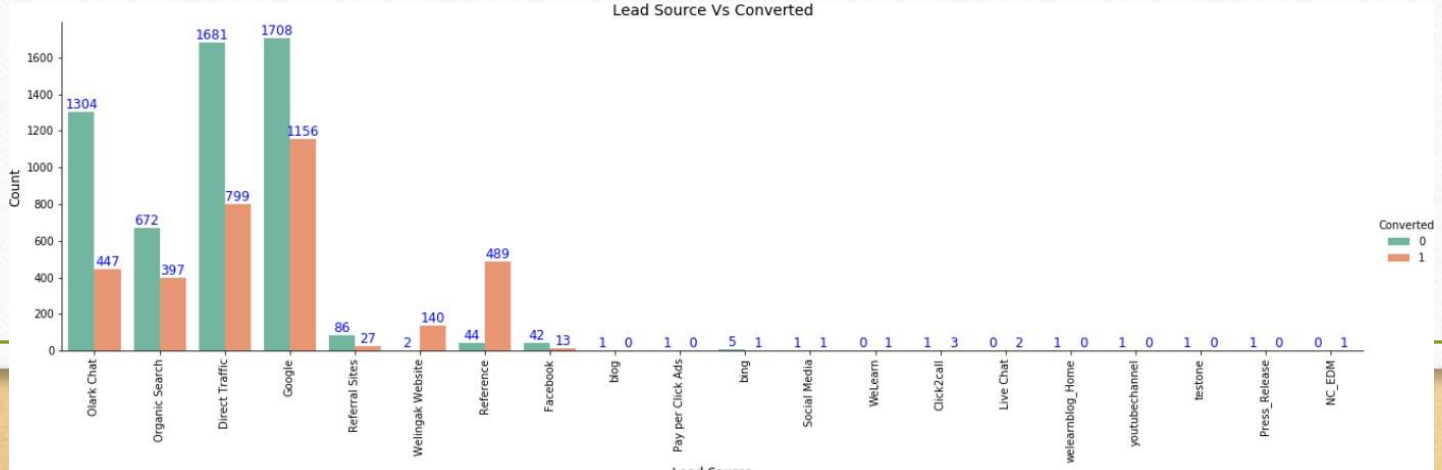


Major conversion has happened from Emails sent and Calls made

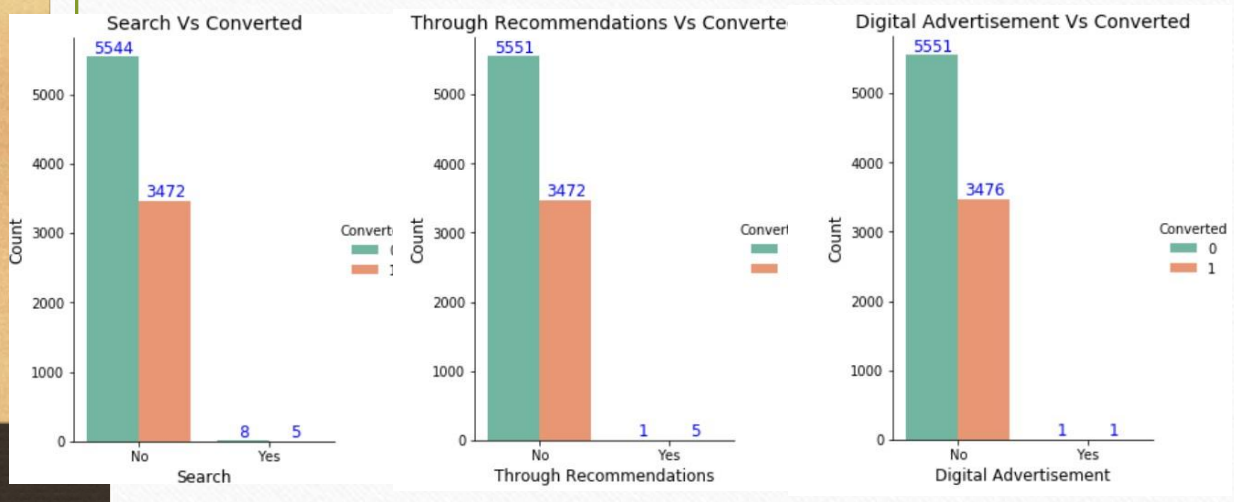




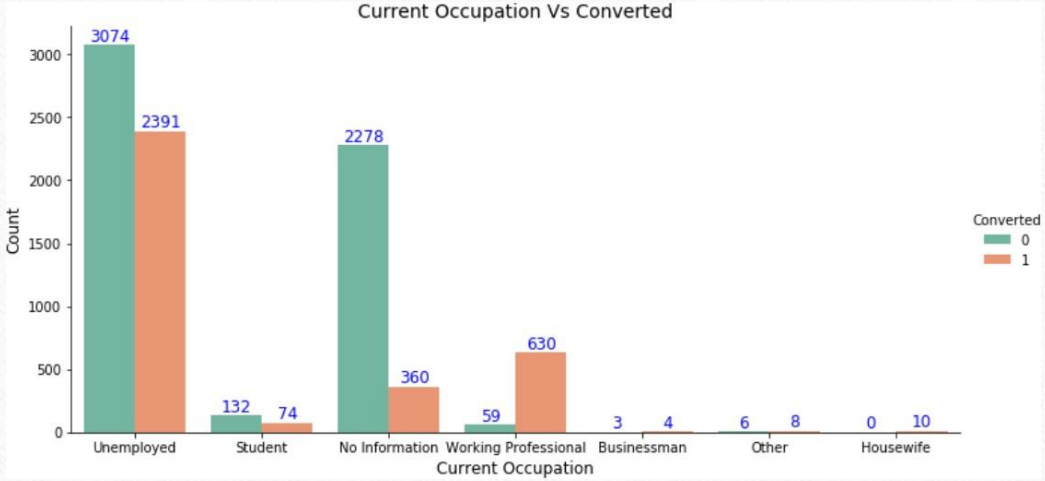
Major conversion in the lead source is from Google



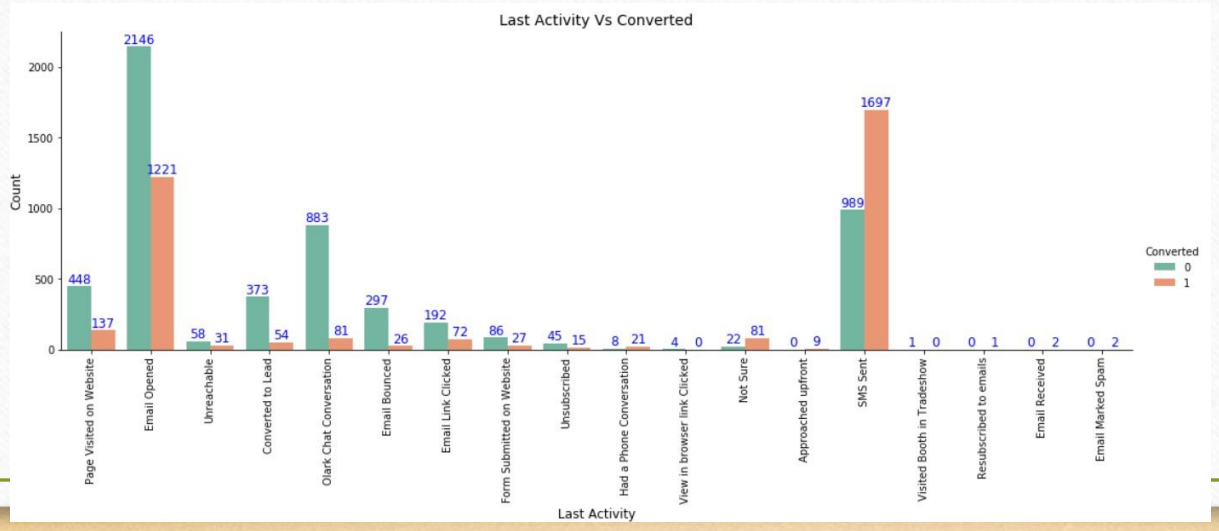
Not much impact on conversion rates through Search, digital advertisements and through recommendations



More conversion happened with people who are unemployed



Last Activity value of SMS Sent' had more conversion.

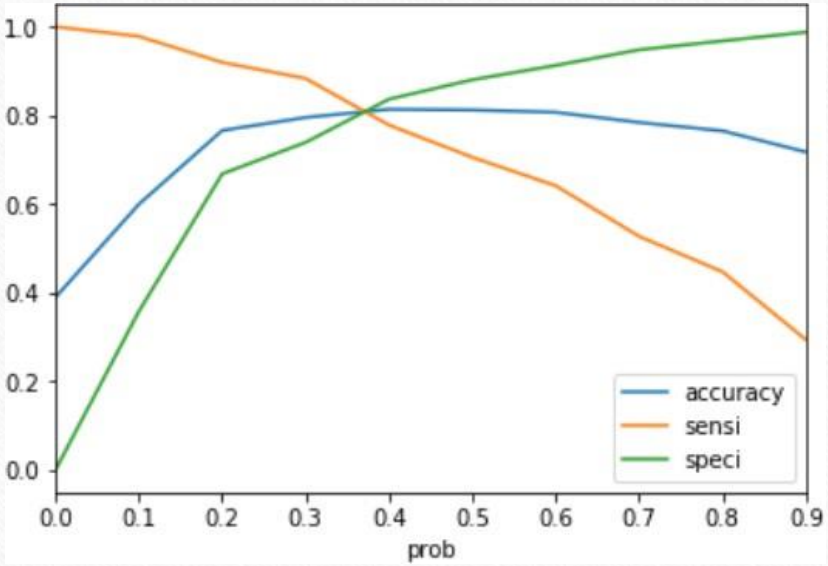


Variables Impacting the Conversion Rate

- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent
- Current Occupation – No Information
- Current Occupation – Working Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - Unreachable

Model Evaluation - Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.37 based on Accuracy, Sensitivity and Specificity



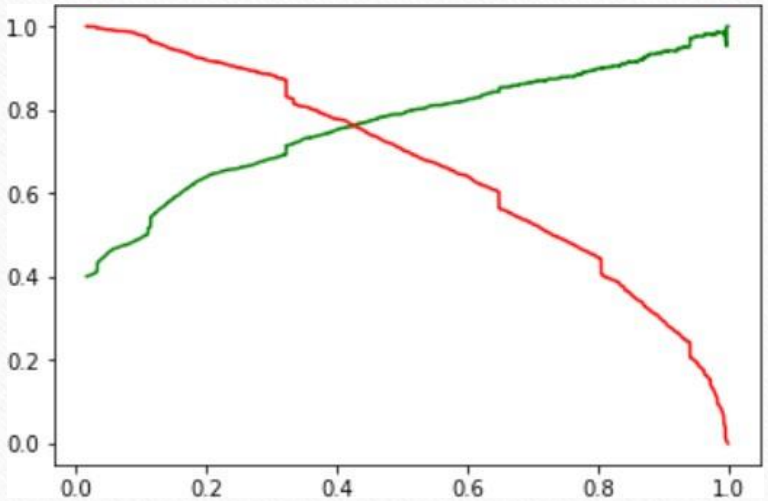
Confusion Matrix

3161	697
974	1965

- Accuracy - 81%
- Sensitivity - 80 %
- Specificity - 82 %
- False Positive Rate - 18 %
- Positive Predictive Value - 74 %
- Positive Predictive Value – 86%

Model Evaluation- Precision and Recall on Train Dataset

The graph depicts an optimal cut off of 0.42 based on Precision and Recall

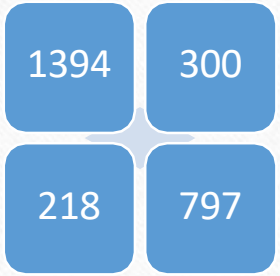


Confusion Matrix

3397	461
725	1737

- Precision - 79 %
- Recall - 71 %

Model Evaluation – Sensitivity and Specificity on Test Dataset



- Accuracy - 81 %
- Sensitivity - 79 %
- Specificity - 82 %

Conclusion

- While we examined both Sensitivity-Specificity and Precision and Recall Metrics, we decided on the ideal cut off based on Sensitivity and Specificity for generating the final prediction. –
- The test set's values for accuracy, sensitivity, and specificity are around 81%, 79%, and 82%, respectively, which are somewhat closer to the corresponding figures determined using the training set.
- As determined by the lead score, the conversion rate on the final predicted model is approximately 80% (in the train set) and 79% (in the test set).
- In the model, the top 3 factors that influence how many leads are converted are
- Total time on site Lead Add Form from Lead Origin Phone Call from Last Noteworthy Activity
- Hence overall this model seems to be good.