Lead Score Case Study

Submitted by :

ABHISHEK KUMAR

Problem Statement :

- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

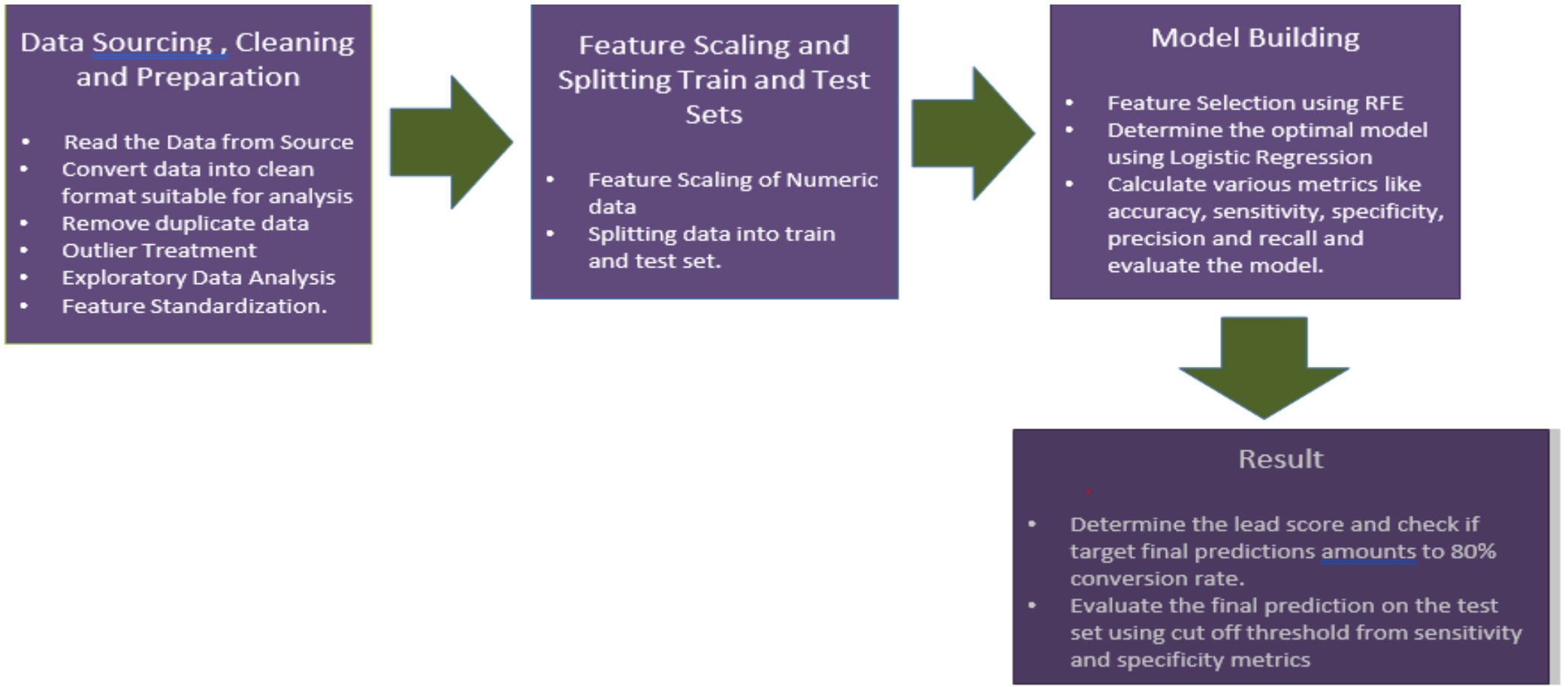
Business Goal:

- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Strategy

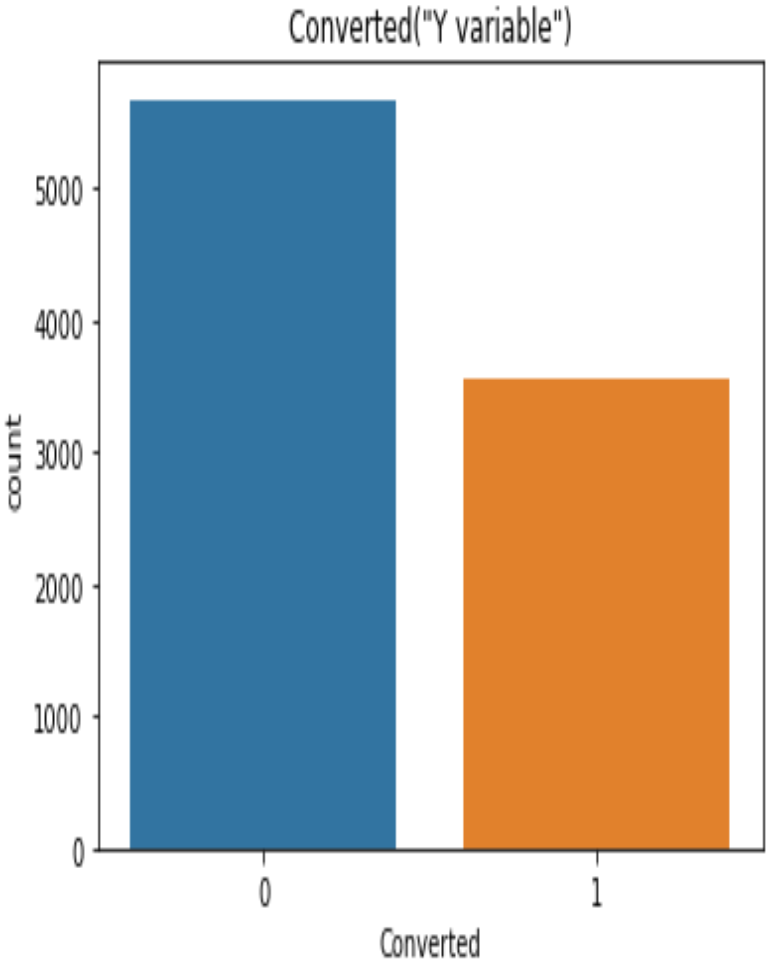
- Pre data analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

Problem solving methodology

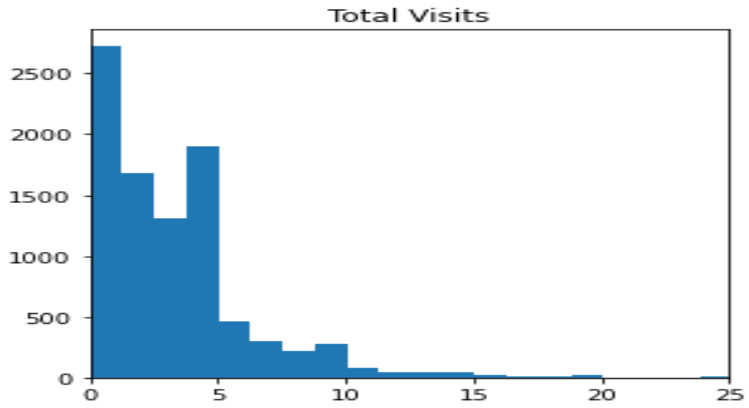


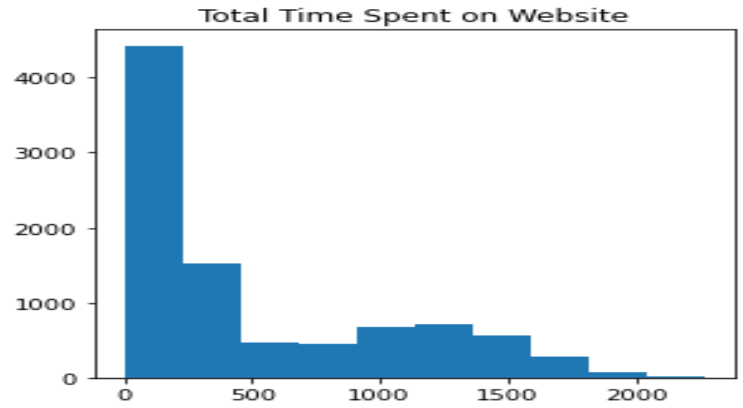
Exploratory Data Analysis

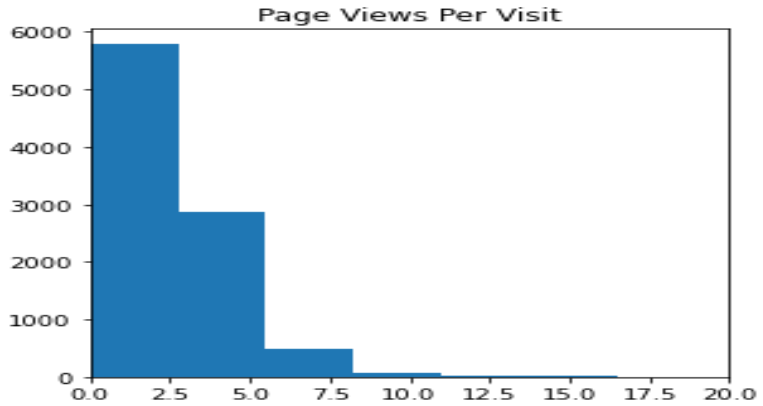
We have around 37% Conversion rate in Total



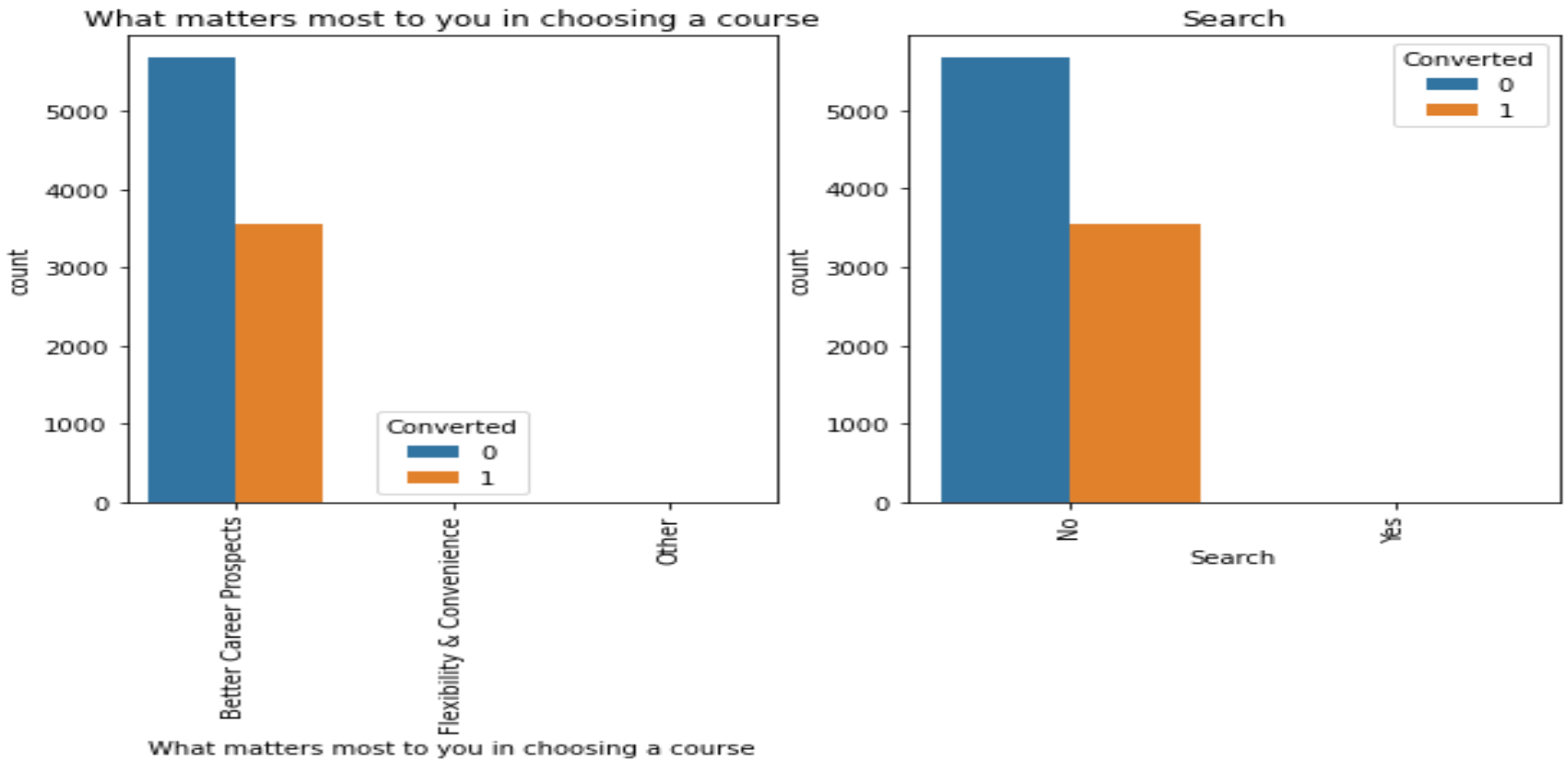
The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit







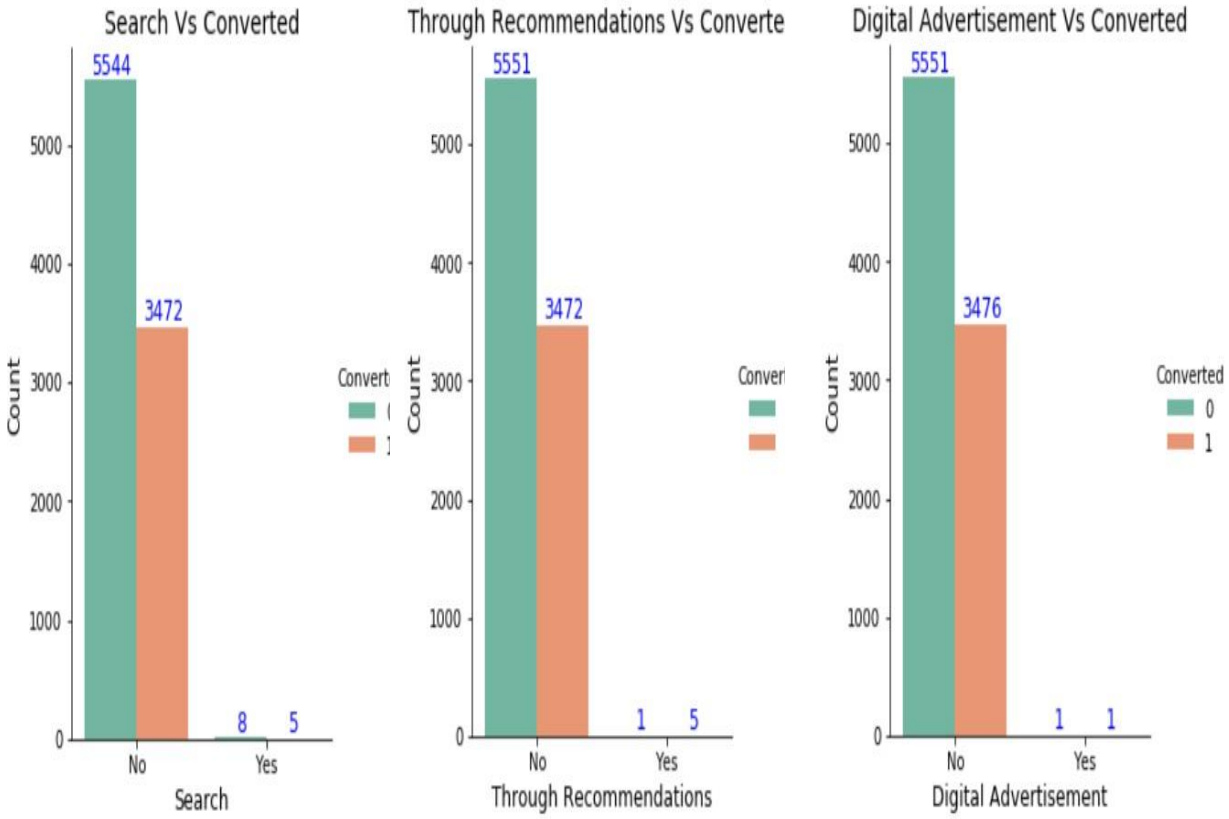
Exploratory Data Analysis

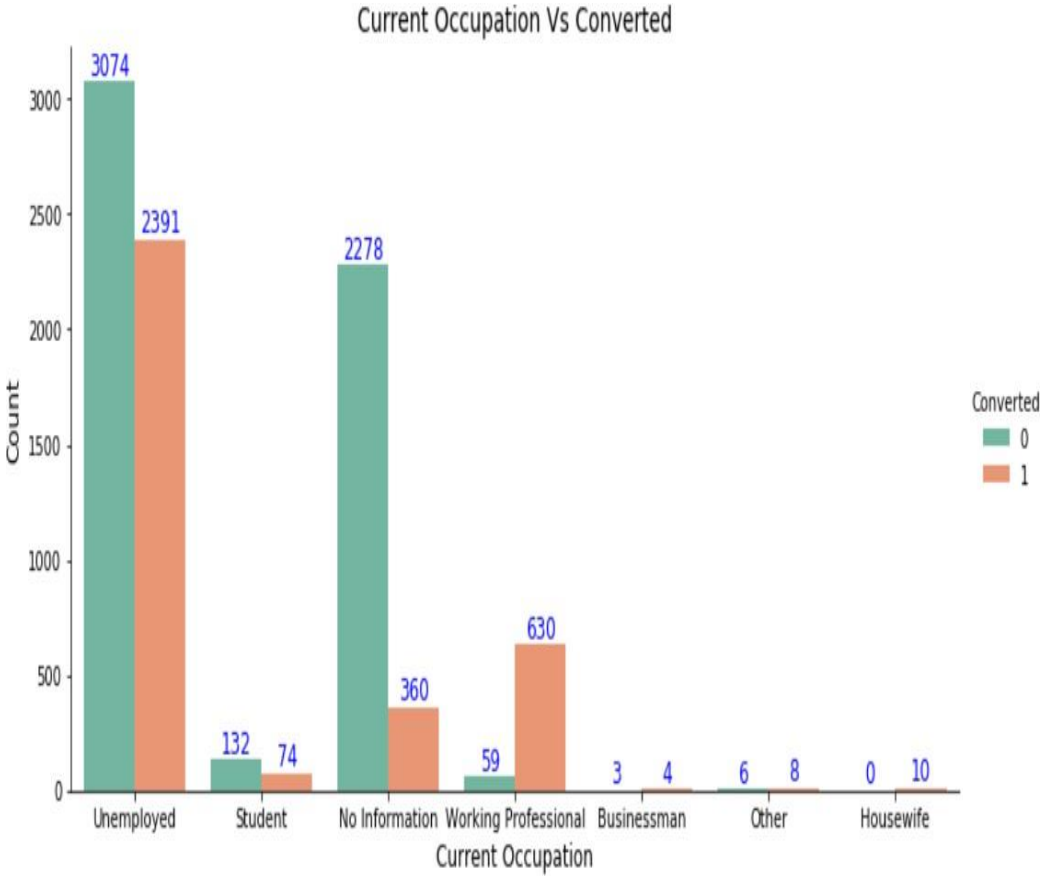


Better career prospect is the main reason for choosing this course

Not much impact on conversion rates through Search, digitaladvertisements and through recommendations

More conversion happened with people who are unemployed





Variables Impacting the Conversion Rate

- Total Time Spent on Website1.
- Lead Origin_Landing Page Submission
- Specialization_Management Specializations
- Last Notable Activity_Modified
- Last Activity_Olark Chat Conversation1
- Lead Source_Olark Chat1.
- Last Notable Activity_Email Opened1
- Lead Origin_Lead Add Form1.
- Specialization_Industry Specializations1.
- Last Notable Activity_Olark Chat Conversation1
- Lead Source_Welingak Website1.
- Last Activity_Converted to Lead1.
- What is your current occupation_Working Profession
- Last Notable Activity_Page Visited on Website
- Last Notable Activity_Email Link Clicked1.

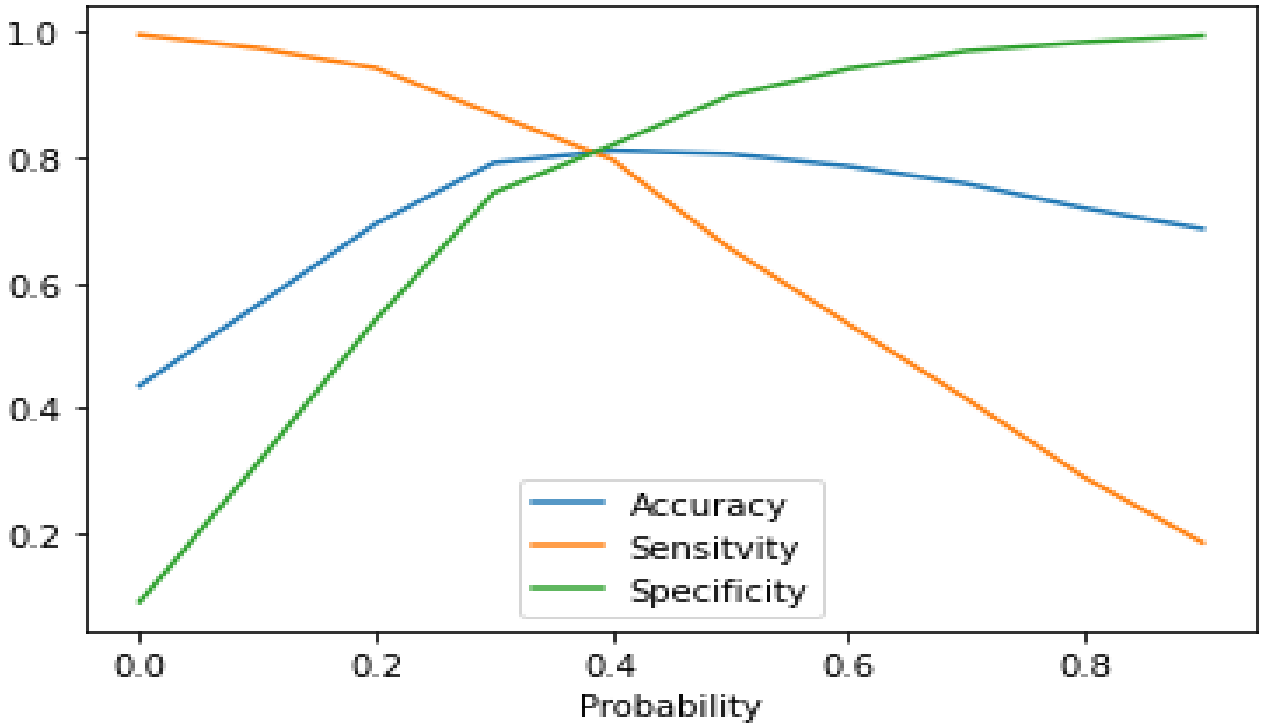
Model Evaluation - Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.4 based on Accuracy, Sensitivity and Specificity

Specificity - 89 %

Accuracy - 80.8%

Sensitivity - 7 5 %

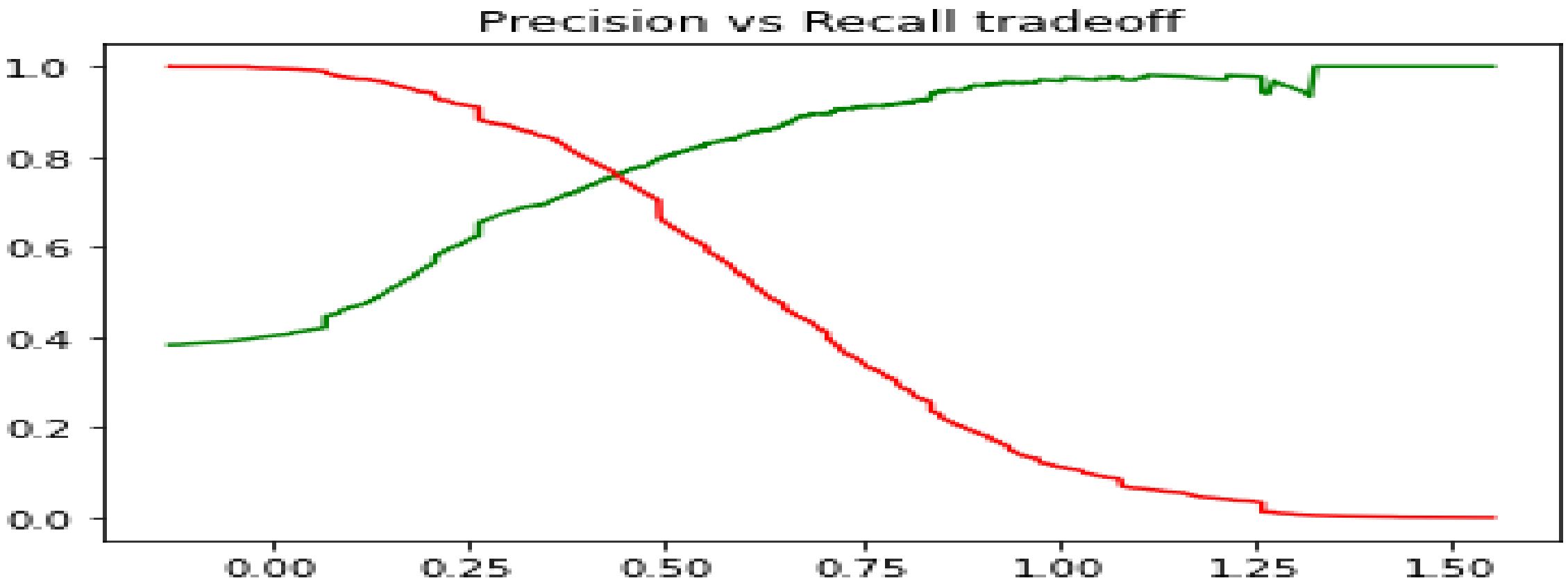


Confusion Matrix

False predictive value - 12 %

Positive Predictive Value - 74 %

Model Evaluation- Precision and Recall on Train Dataset



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

Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction. –
- Accuracy, Sensitivity and Specificity values of test set are around 81%, 79% and 82% which are approximately closer to the respective values
- calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 80% (in train set) and 79% in test set
- The top 3 variables that contribute for lead getting converted in the model are

  1.Last Notable Activity_Had a Phone Conversation

  2.Lead Origin_Lead Add Form and

  3.What is your current occupation_Working Professional

  4.Last Activity_SMS Sent
- Hence overall this model seems to be good.