Lead Scoring Case Study

By,
Priti Pratikshya Pradhan and Gayathri Hari

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%.

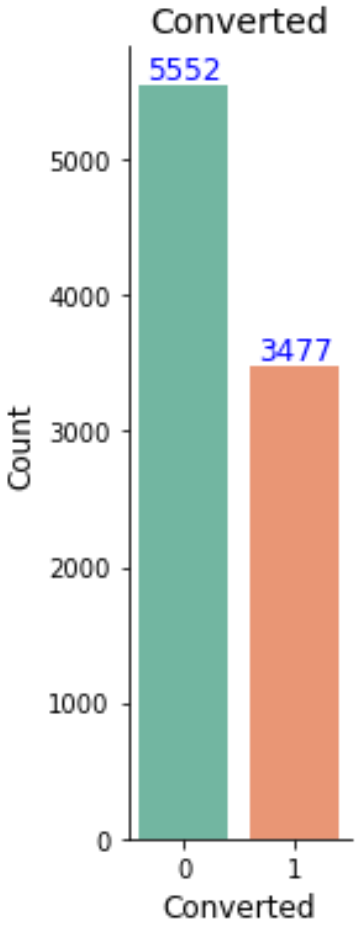
Approach

- 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%.
- Dataset has column ‘converted’ which helps in identifying which customer is successfully converted for the online courses after the calls.
- There are various factors which are mentioned as part of different columns which indicates the customer data at various stages.
- Exploratory data analysis is done on the given dataset and Logistic Regression Model is run on the same to identify main factors.
- Various metrics are calculated and one finalized to satisfy the business problem.

Important Steps performed

- Exploratory data Analysis and Data Preparation:-
  - EDA: Data is checked for null value, variance in information and outlier.
  - Data is prepared for logistic regression model by replacing character values into numerical format wherever possible.
  - Dummy variables are created for categorical columns.
  - Many variables which are created with 0% or less variance in information are dropped.

Overall conversion rate (around 39%)

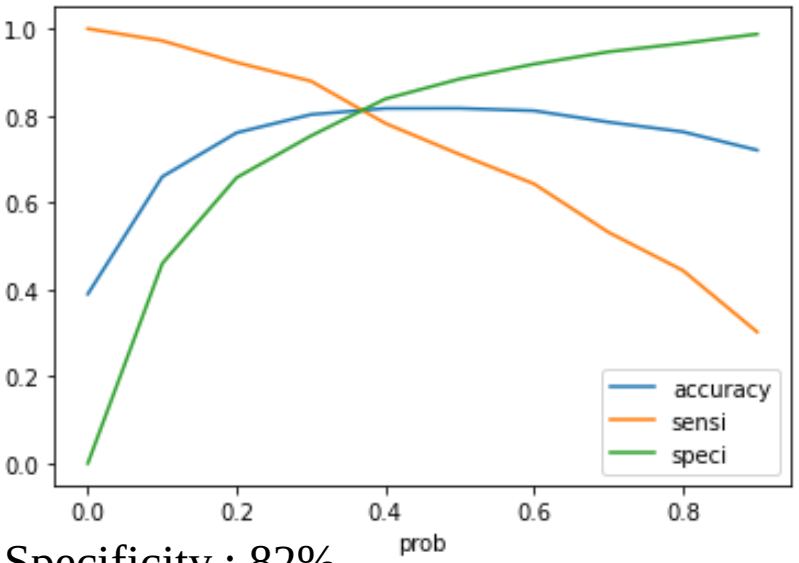


Variables Impacting the Conversion Rate

- 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



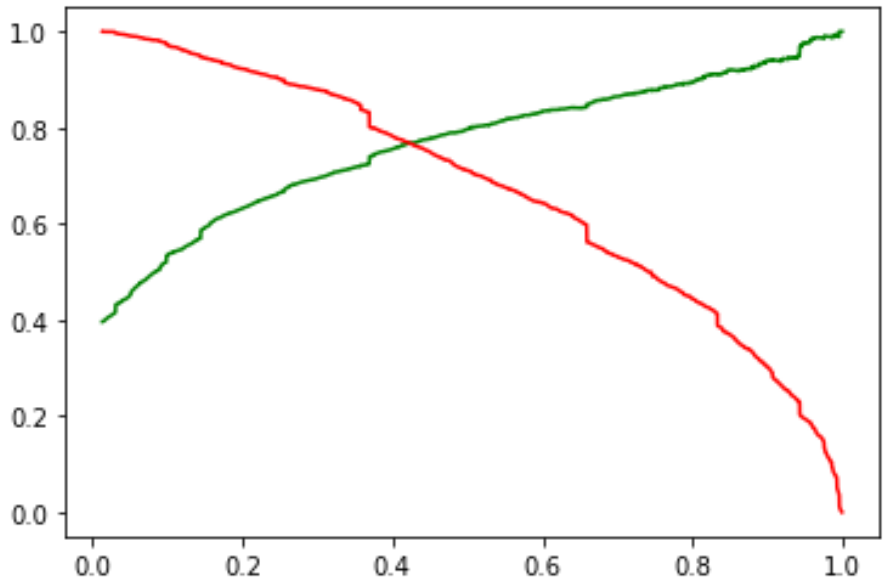
Specificity : 82%

Sensitivity: 80%

Accuracy: 81%

Model Evaluation-Precision and Recall on Train Dataset

- Precision vs Recall is plotted on graph to find optimal cut-off. Cut-off: 0.4



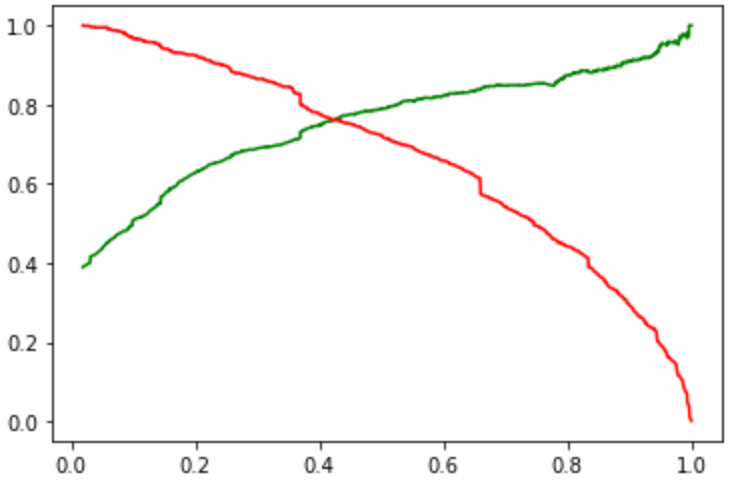
Accuracy, Confusion Metrics, Sensitivity, Specificity, on final prediction on Test set

Specificity : 82.23%

Sensitivity: 80%

Accuracy: 81.39%

- Precision and Recall metrics for the test set



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%, 80% and 82% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is around 80%
- Hence overall this model seems to be good.

Inferences

- Based on the analysis and regression built, following are the three most important factors to be targeted for a higher conversion rate:
  - Total Time Spent on Website
  - Lead Add Form (from Lead Origin)
  - Had a Phone Conversation (from Last Notable Activity)
- So the concentration should be increased on leads that:
  - Spend more time on the platform
  - Fill out a form
  - Have a phone conversation with the team
  - Are working professionals

THANK YOU!