



Lead Scoring Case Study

ANKIT KUMAR SINGH

INTRODUCTION

- In this case study we will go through various steps involved in this case study and will have a depth understanding of how to approach such problems
- We will go through the lead scoring dataset have a basic understanding of the terms used in it , understand the problem statement and apply EDA and model building so that we can achieve the required results

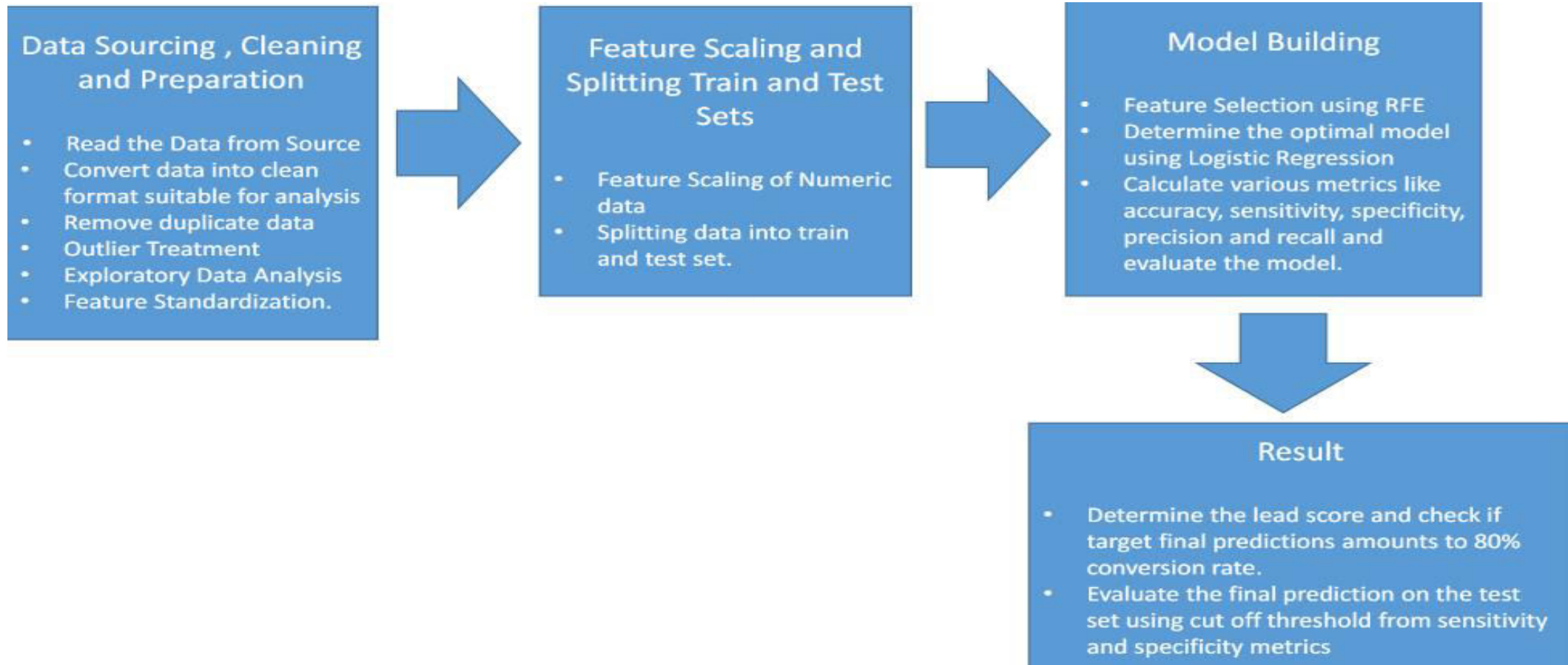
Problem Statement

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

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%.

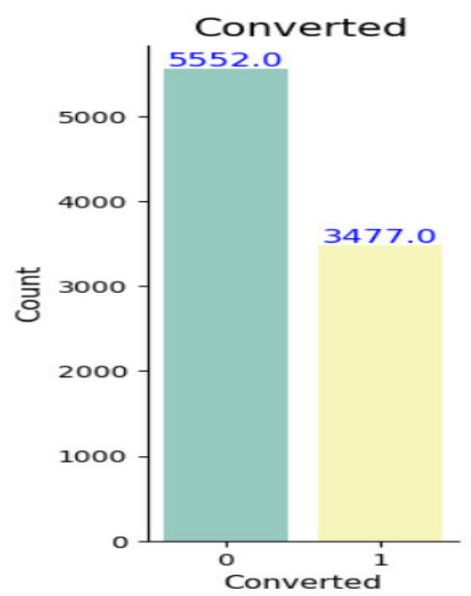
X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a 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%.

Problem Solving Approach

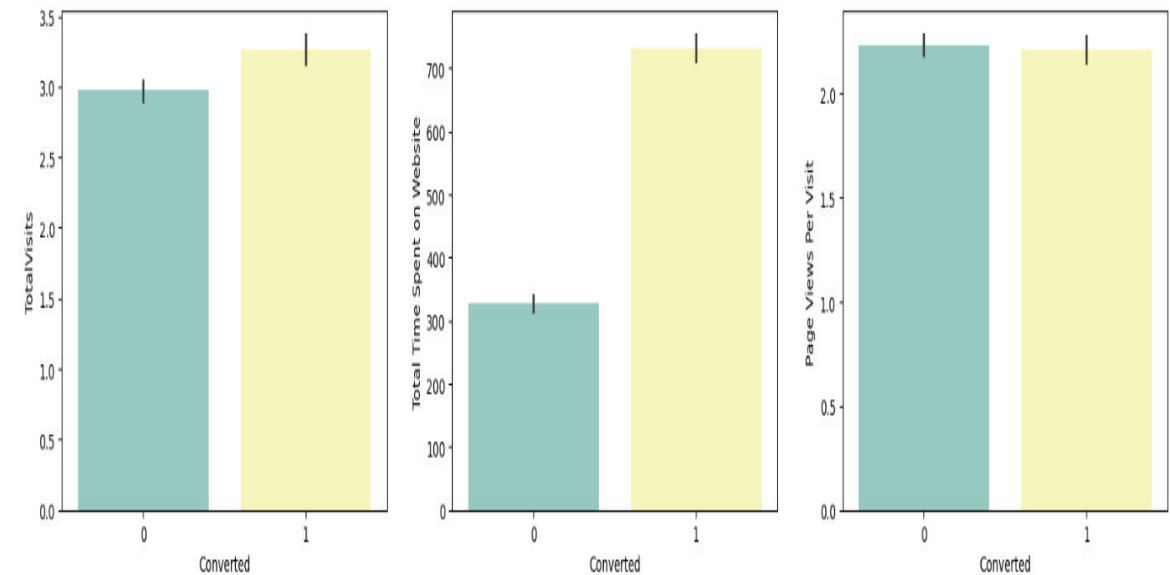


Exploratory Data Analysis

CONVERSION RATE IS AROUND 39%[I](#)

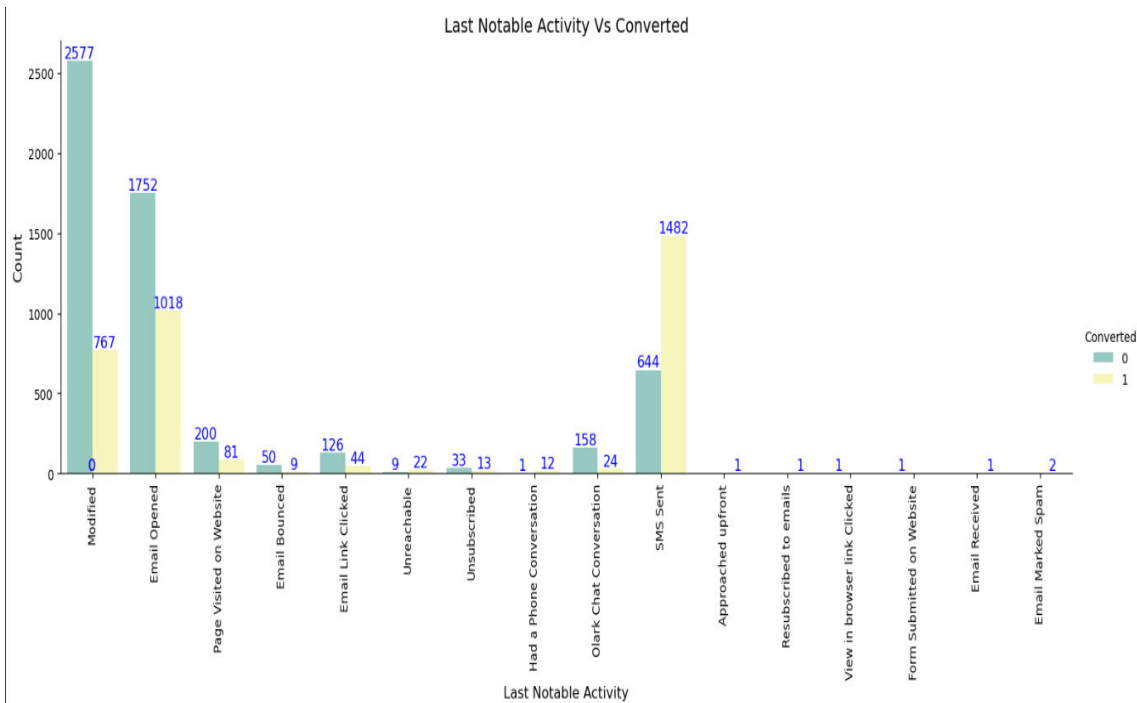


CONVERSION RATE ARE HIGH FOR TOTAL VISITS,
TOTAL TIME SPENT ON WEBSITE AND PAGE VIEWS
PER VISIT

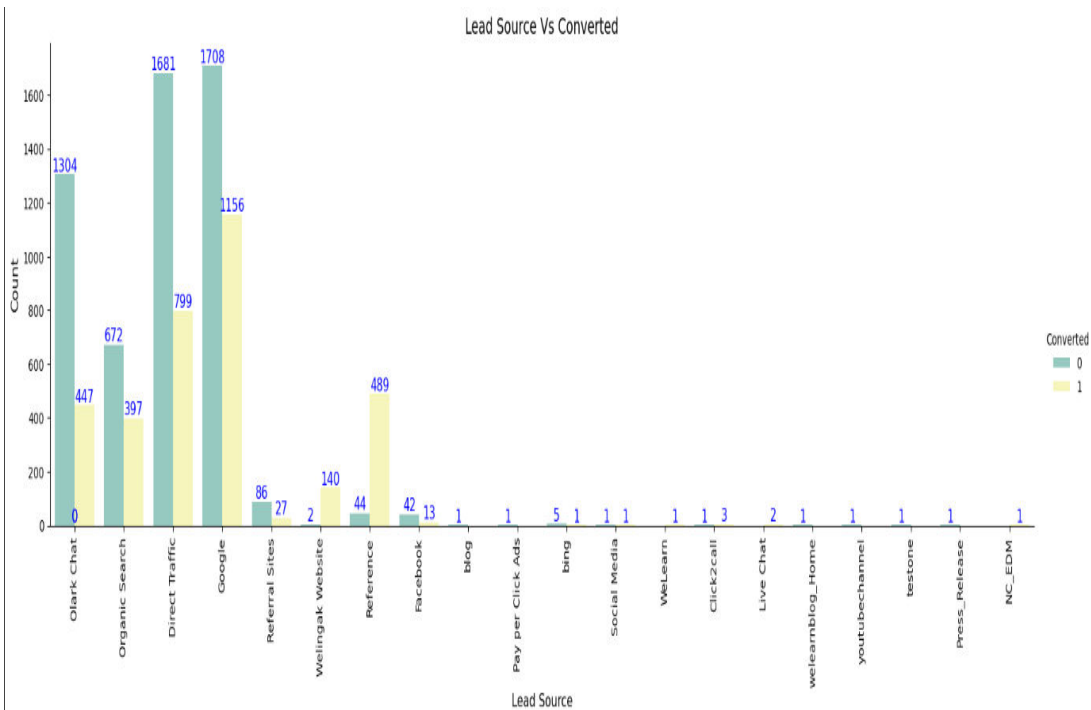


Exploratory Data Analysis

CONVERSION IS HIGH FOR 'SMS SENT'

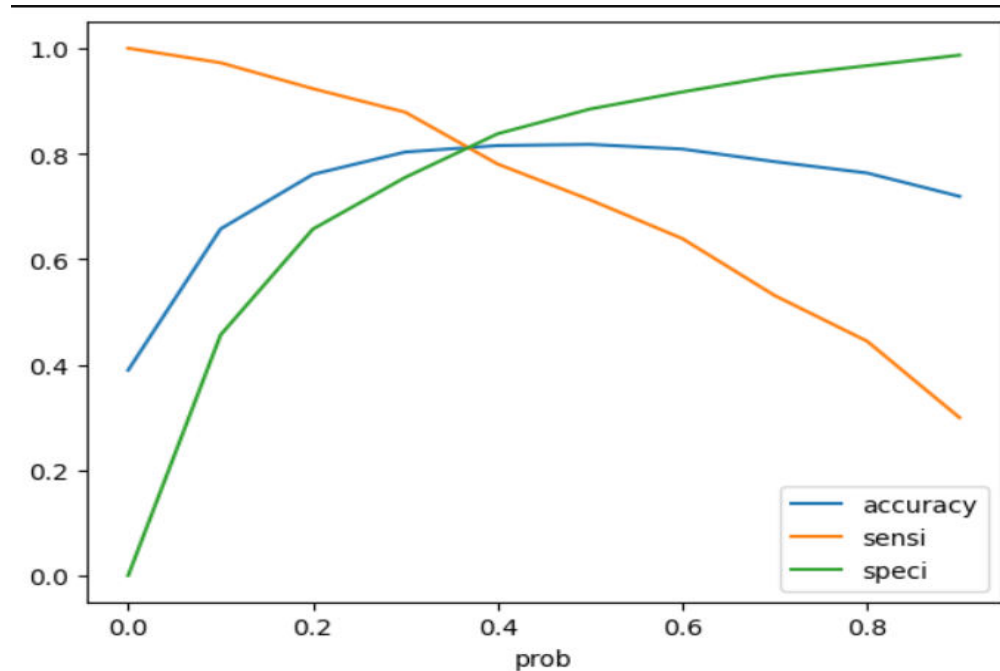


MAXIMUM CONVERSION ID FROM GOOGLE



Model Evaluation on Train Data Set

THE GRAPH DEPICTS OPTIMAL CUTOFF OF 0.37 BASED ON ACCURACY, SENSITIVITY AND SPECIFICITY



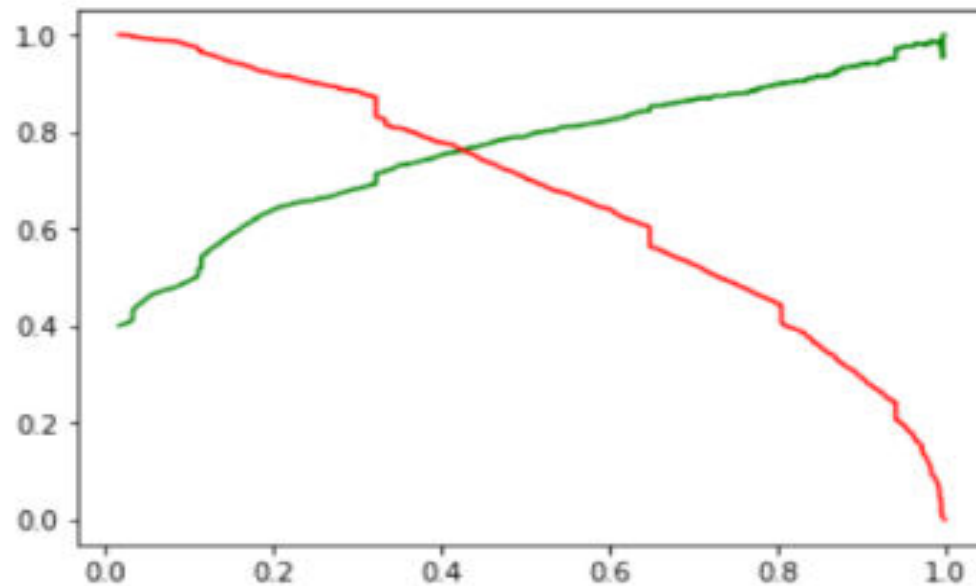
CONFUSION MATRIX

3161	697
974	1965

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

Model Evaluation on Train Data Set

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



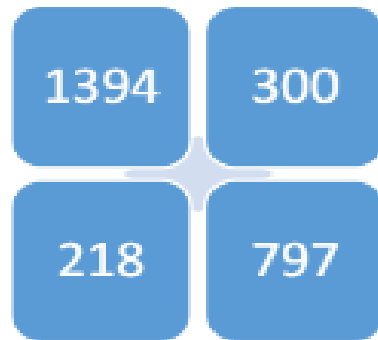
Confusion Matrix

3397	461
725	1737

- Precision - 79 %
- Recall - 71 %

Model Evaluation on Train Data Set

Confusion Matrix

A diagram of a 2x2 confusion matrix. It consists of four blue rounded squares arranged in a 2x2 grid. The top-left square contains the number 1394, the top-right square contains 300, the bottom-left square contains 218, and the bottom-right square contains 797. A small grey star is positioned at the center where the four squares meet.

1394	300
218	797

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

CONCLUSION

Conclusion :

- ❑ We have verified both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off which is 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 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.