Lead Score Case Study

Submitted by
Soumya Khan
Sourit Chakraborty
Souvik Guchait

Lead Score Case Study for X Education

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%.

Problem Solving Methodology

Data Sourcing , Cleaning and Preparation

Read the Data from Source, Convert data into clean format suitable for analysis, Remove duplicate data, Outlier Treatment, Exploratory Data Analysis, Feature Standardization.

Feature Scaling and Splitting Train and Test Sets

Feature Scaling of Numeric data, Splitting data into train and test set.

Model Building

Feature Selection using RFE, Determine the optimal model using Logistic Regression, Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.

Result

Determine the lead score and check if target final predictions amounts to 80% conversion rate. Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics.

Exploratory Data Analysis

Fig 1- lead origin from ‘lead add form’ is highly convertable leads, more learner take the course.

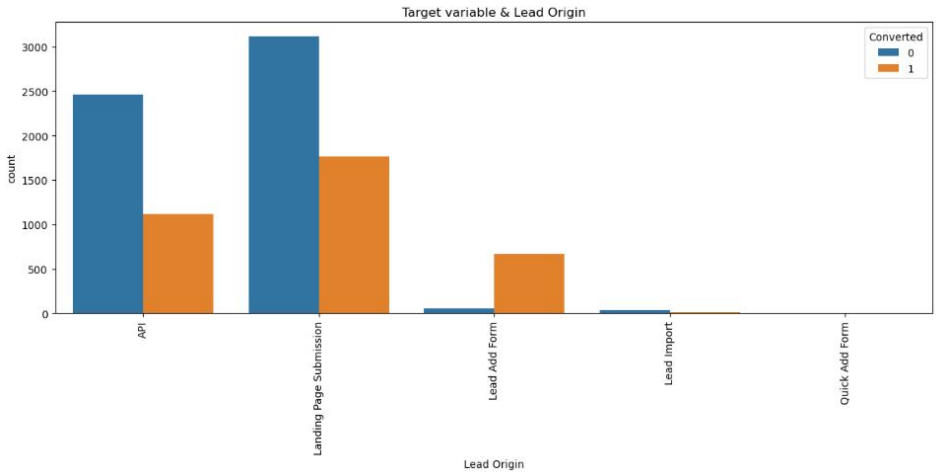


Fig-2: lead source ‘reference’, ‘welingak website’, and google search leads are highly convertable than others.

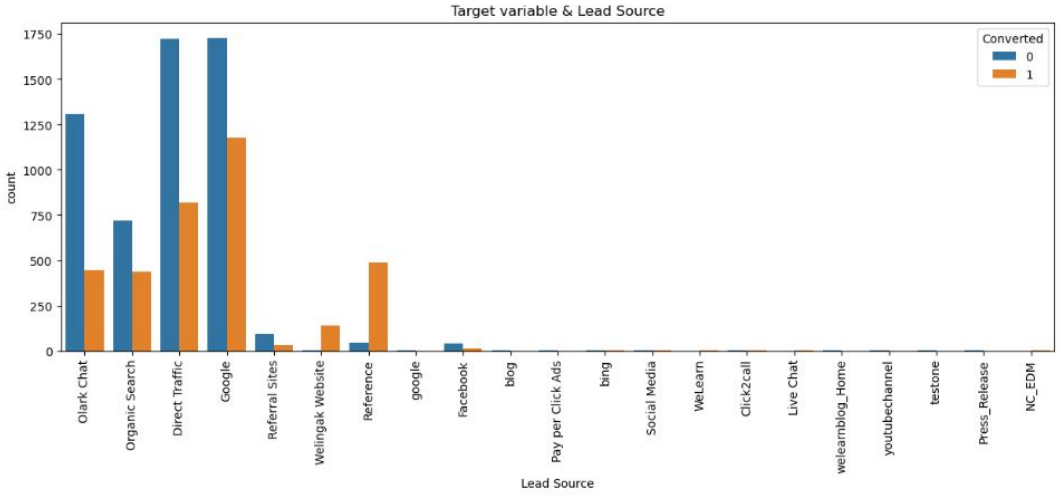
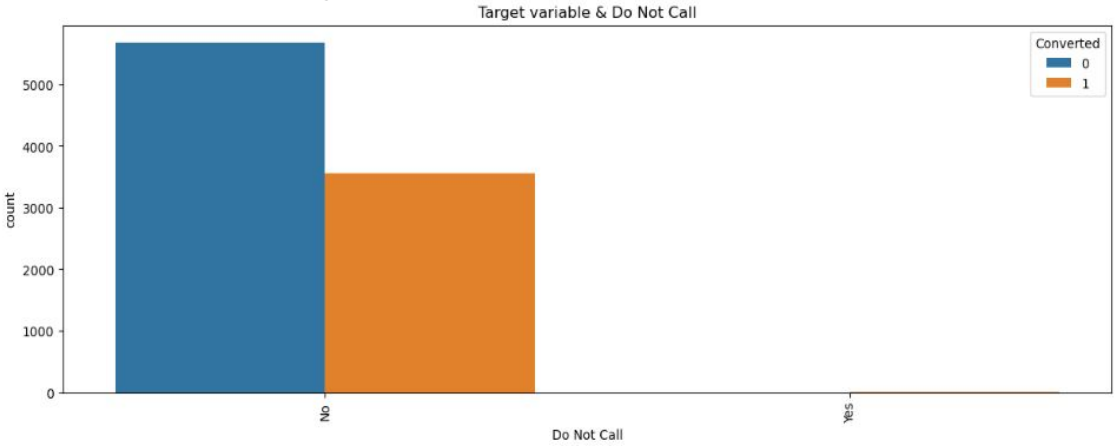


Fig-3: Sales team should not wast time on the leads which already ask not to send email to them.

Fig-1: Sales team should not west time on the leads which ask not to call them again.



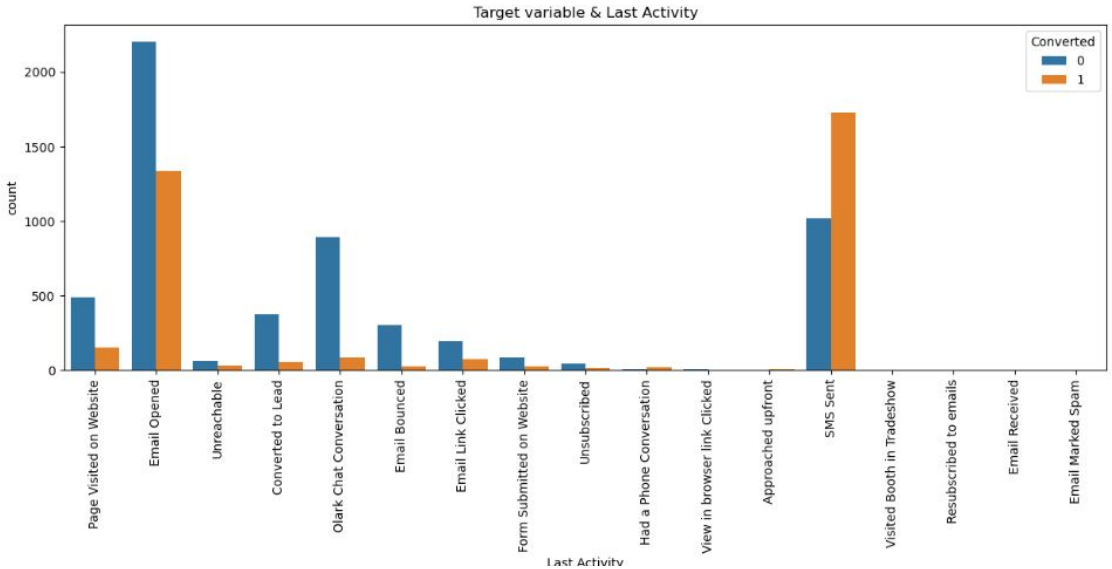


Fig-2: Working professional are more keen to take up the course as well as unemployed people too. students are less likely to convert into sell.

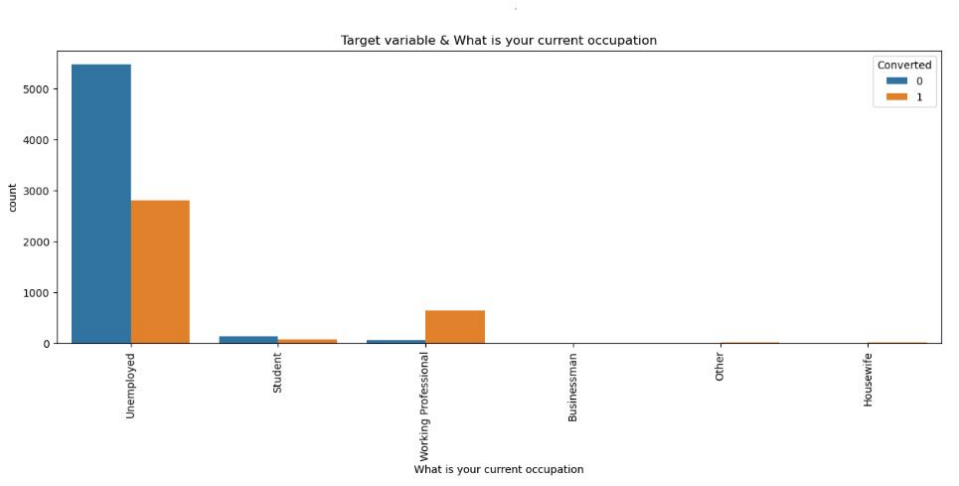


Fig-3: Last activity Email opened and sms sent state that lead is more active to know about the course.

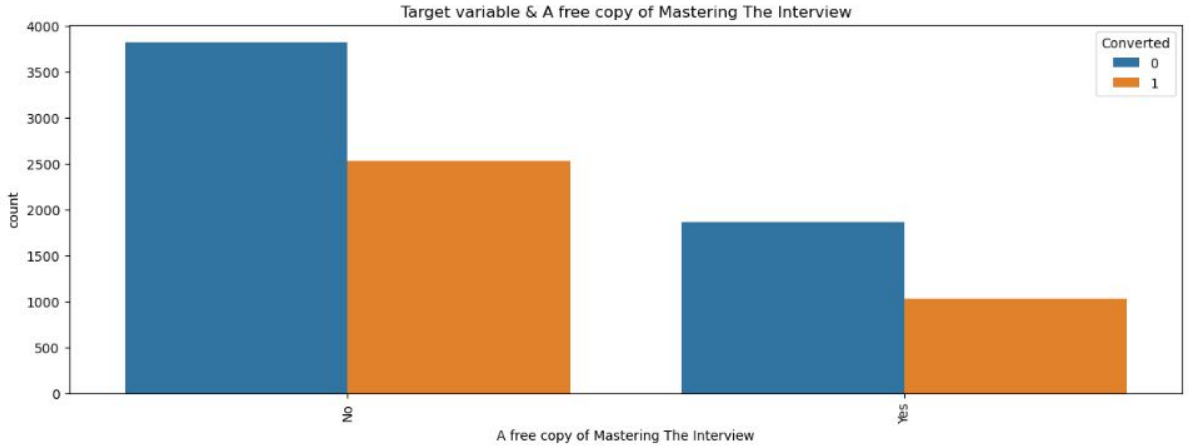


Fig-1: A free copy of mastering interview is really not much impacting the conversion part

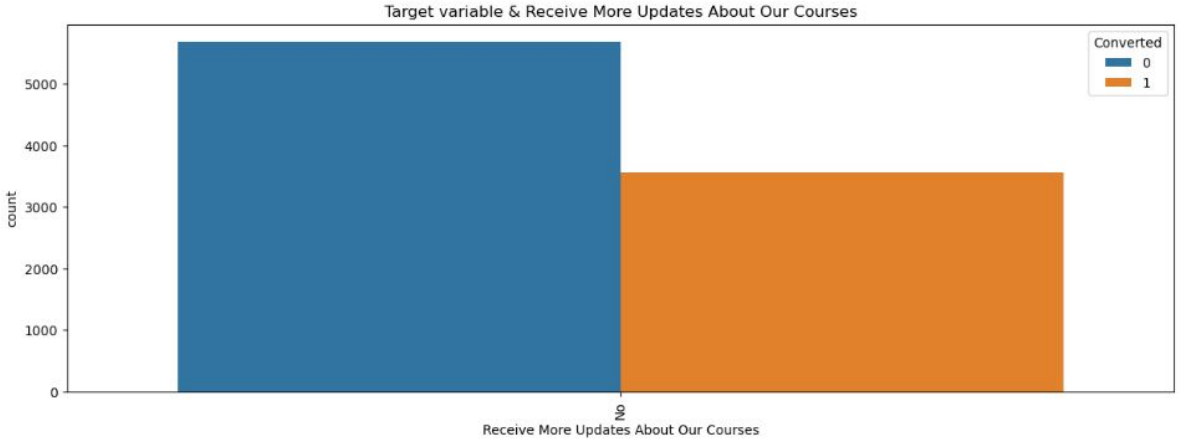
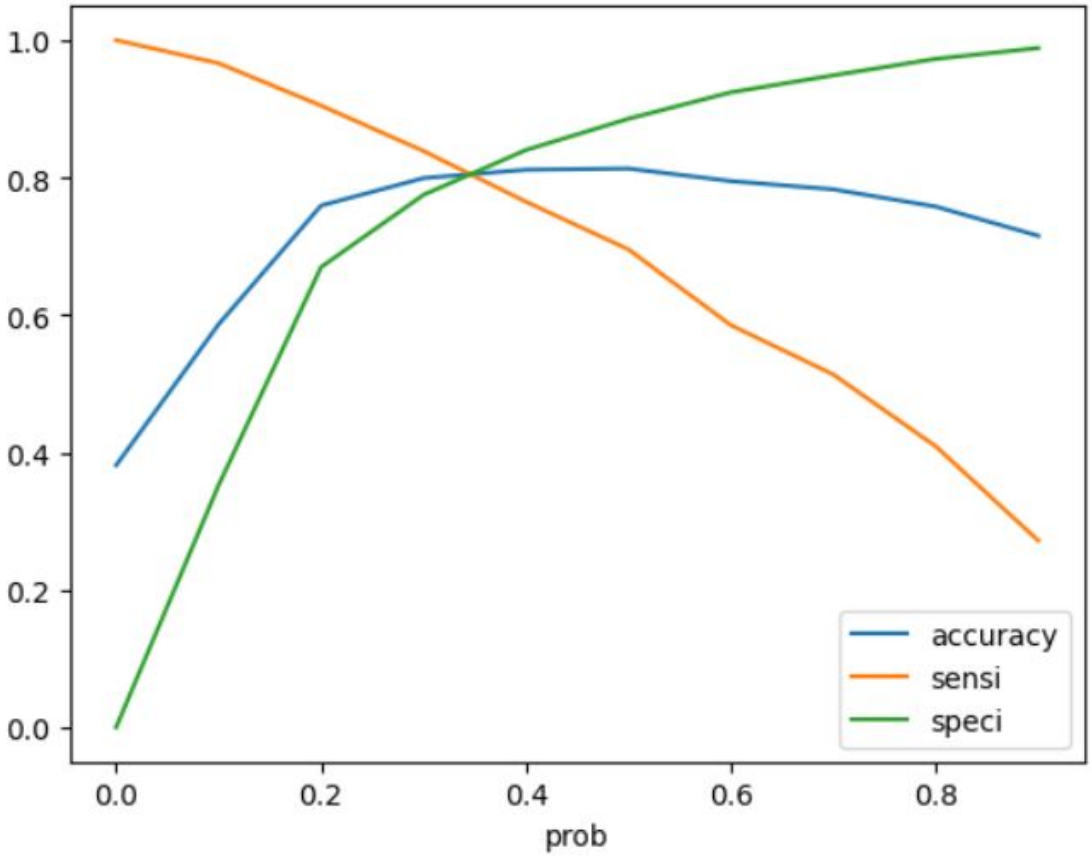


Fig-2: Sales team should push for giving more updates about the course

Variables Impacting Convertions

- Total Visits
- Do not email
- Do not call
- 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 –Working Professional
- Last Notable Activity –Had a Phone Conversation

Model Evaluation -Sensitivity and Specificity on Train Data Set



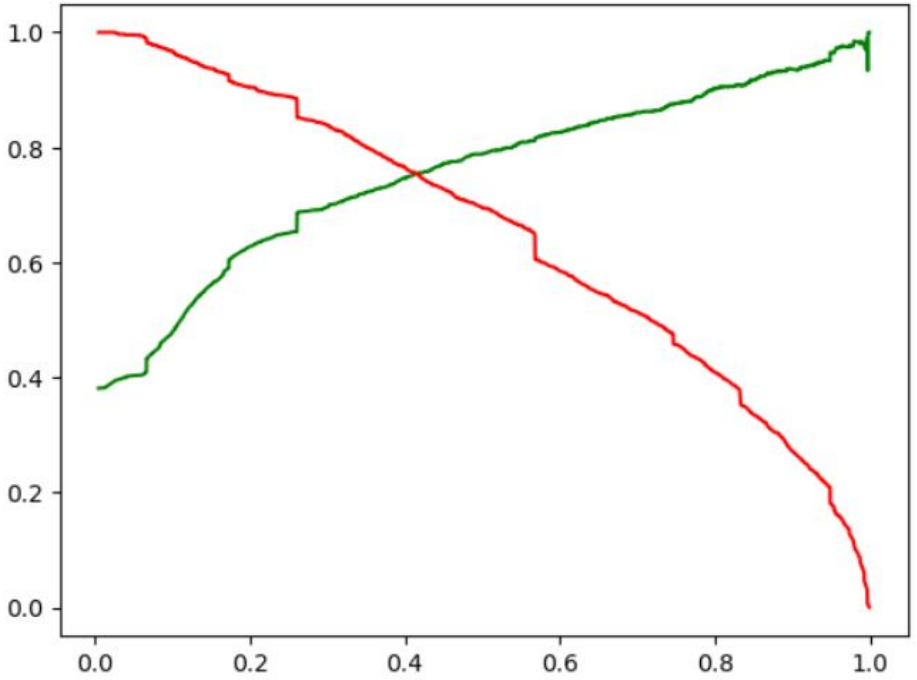
The graph shows a optimum cuttoff point 0.4 or 0.37

Confusion Matrix

3362	640
580	1886

- Accuracy: 81.13%
- Sensitivity: 76.48%
- Specificity: 84%
- False Positive rate: 25.33%
- Positive Predictive Value: 84%

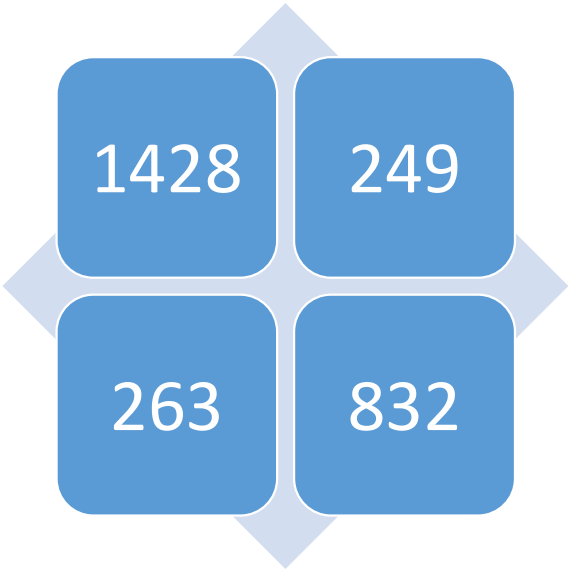
Model Evaluation-Precision and Recall on Train Dataset



- Precision- 74.67%
- Recall- 76.48%

Model Evaluation –Sensitivity and Specificity on Test Dataset

Confusion Matrix



- Accurecy- 81.52%
- Specificity- 85.15%
- Sencitivity- 75.98%
- Precision- 76.96%
- Recall- 75.98%

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 81.13%, 76.48% and 84% 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 81.13% (in train set) and 81.52% in test set
- □ The top 3 variables that contribute for lead getting converted in the model are
  a) Total time spent on website
  b) Lead Add Form from Lead Origin
  c) Had a Phone Conversation from Last Notable Activity
- □ Hence overall this model seems to be good.