LEAD SCORING CASE STUDY

LOGISTIC REGRESSION

BATCH- DS C63

SURODEEP DEY
JASDEEP SINGH CHAWLA
ROHIT SHARMA

Problem Statistics

X Education sells online courses to industry professionals. The company markets its courses on several websites and search en

gin es like

Google. Once people land on the website, they might browse the courses, fill out a form for a course, or watch some videos. W hen they

provide their email address or phone number through a form, they are classified as leads. Additionally, the company acquires lea ds through

past referrals. Once leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this pro cess, some

leads get converted, while most do not. The typical lead conversion rate at X Education is around 30%.

OBJECTIVE

X Education needs assistance in selecting the most promising leads

those most likely to convert into paying customers. The company requires

a model that assigns a lead score to each lead, indicating a higher likelihood of conversion for leads with higher scores (ho t l eads) and a lower

likelihood for those with lower scores (cold leads). The CEO has specified a target lead conversion rate of approximately 80%

STEPS INVOLVED

- Data Sourcing
- Data loading and cleaning
- Exploratory Data Analysis
- Data Preparation and dummy column generation.
- Train test split.
- Feature scaling.
- Model Building including RFE approach, calculating the VIF etc.
- Model Evaluation on Train sets.
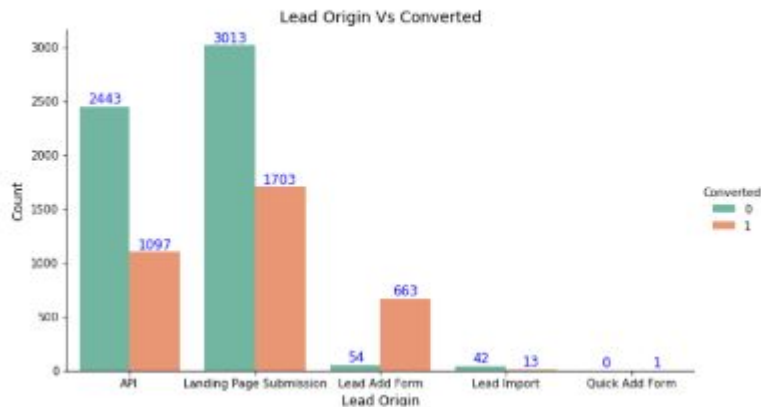- Model Evaluation on Test sets.

EXPLORATORY DATA ANALYSIS

We have 39% conversion rate in total. The conversion rates are high for total visits, Total time spent on website and Page views per visit.
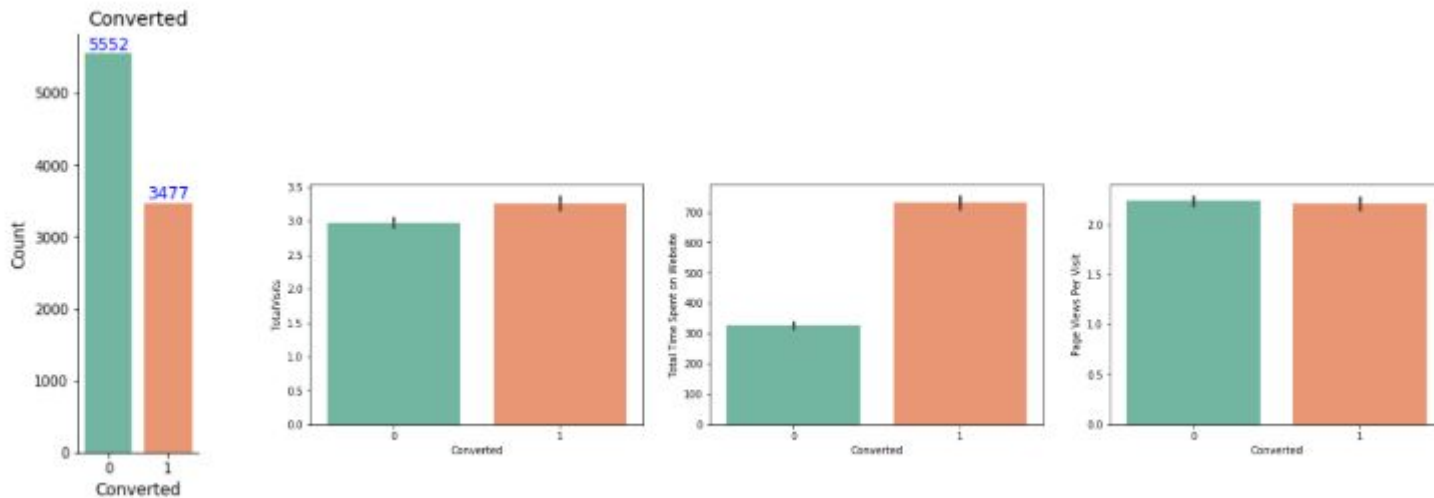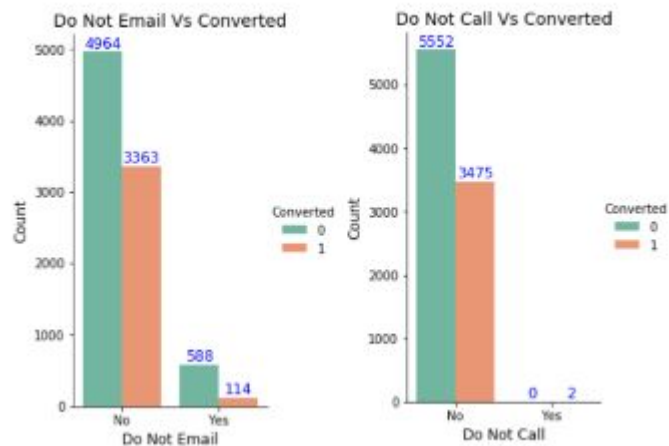
CONTD...

In Lead Origin, maximum conversion happened from Landing Page Submission

CONTD....

Major conversion has happened from Emails sent and Calls made

VARIABLES IMPACTING CONVERSION RATES

- 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 –Welingkar 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 Conversat.

MODEL EVALUATION – SENSITIVITY AND SPECIFICITY

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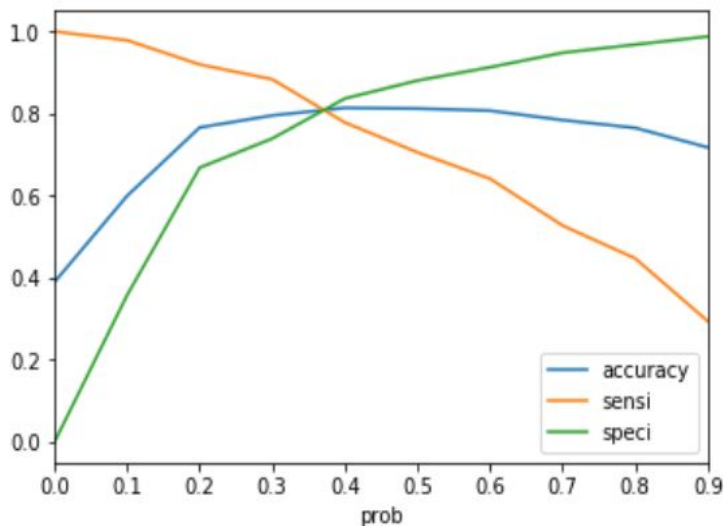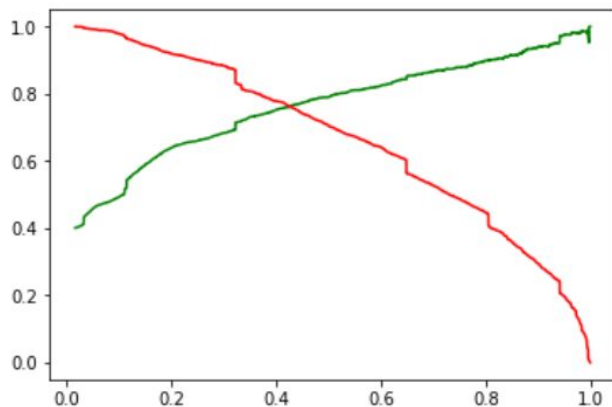

- Accuracy - 81%
- Sensitivity - 80 %
- Specificity - 82 %
- False Positive Rate - 18 %
- Positive Predictive Value - 74 %
- Positive Predictive Value – 86%

MODEL EVALUATION-PRECISION AND RECALL ON TRAIN DATASET

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

- Precision - 79 %
- Recall - 71 %

MODEL EVALUATION –SENSITIVITY AND SPECIFICITY ON TEST DATASET

Confusion Matrix

1394	300
218	797

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

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 respectivevalues 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

Total time spent on website

Lead Add Form from Lead Origin

Had a Phone Conversation from Last Notable Activity

Hence overall this model seems to be good.