LEAD SCORE CASE STUDY

SUBMITTED BY:
PRIYANKA SARWAR
VISHWAJEET ANAND
NARENDRA SASANOOR

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 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%.

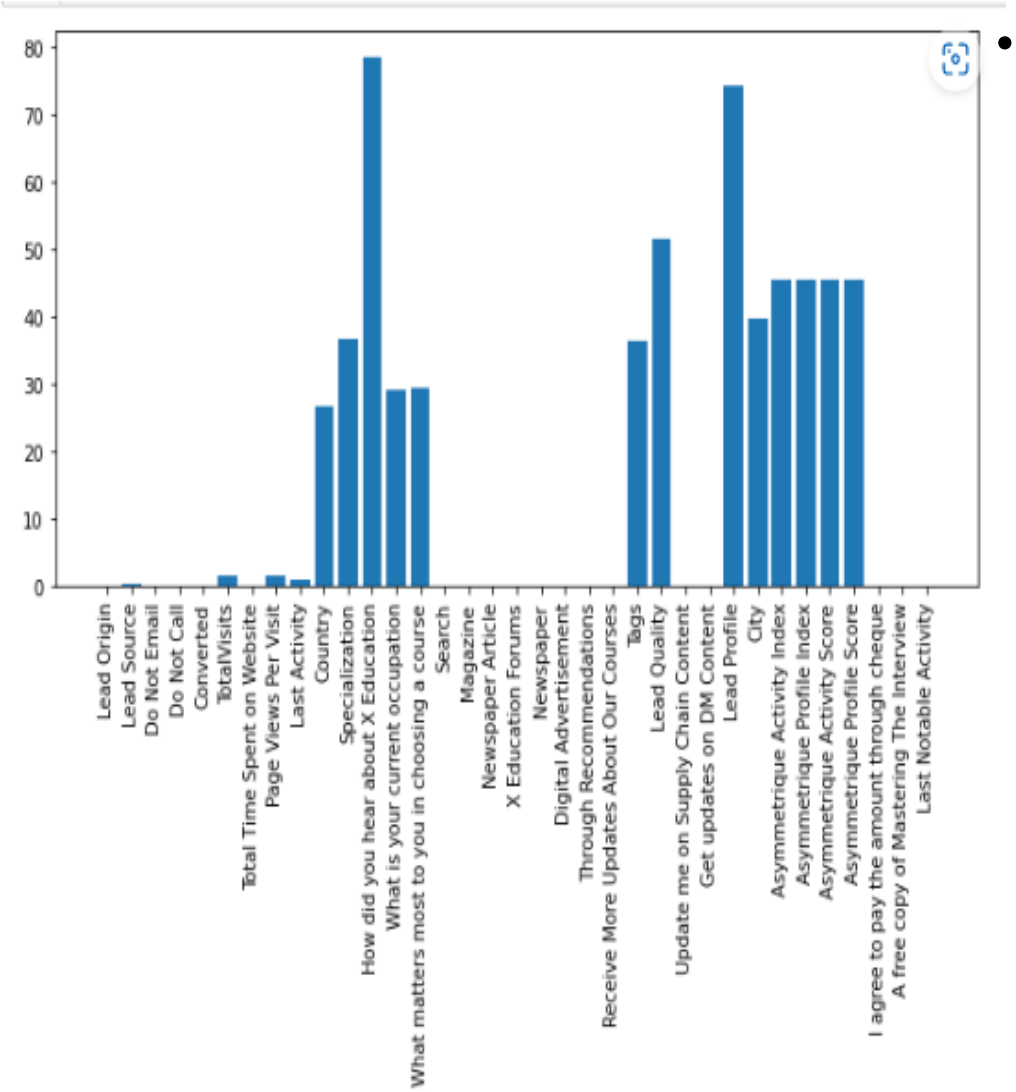
- **Business Goal**:

- 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

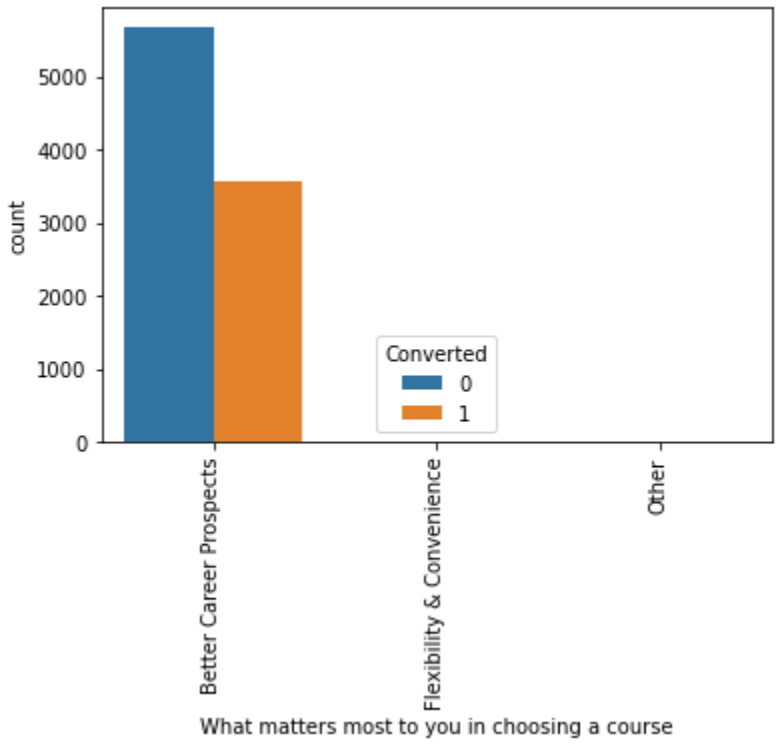
- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity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

Bar Plot for Missing Values



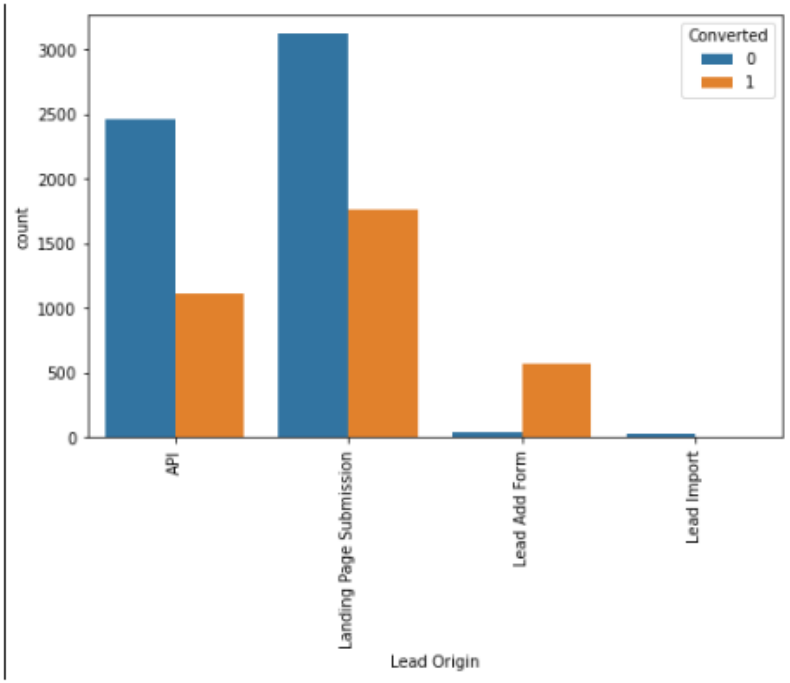
- The columns with highest missing values includes Lead Quality,Tags,Asymmetrique Activity Index,Asymmetrique Profile Index,Asymmetrique Activity Score etc.

Count Plot for what matters most to you in choosing a course



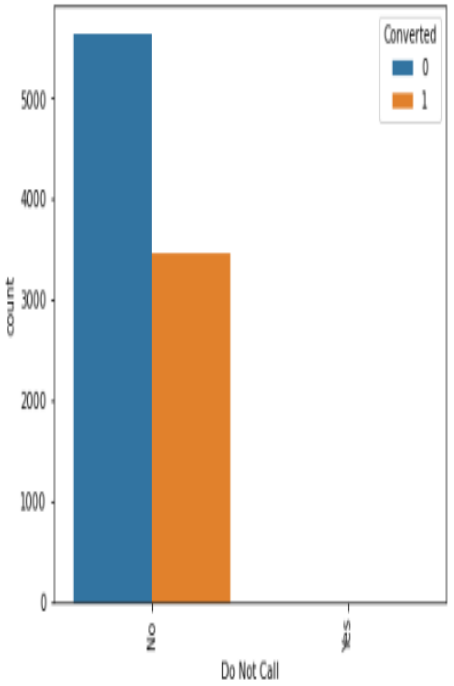
- From the count plot of what matters most to you in choosing in a course most of the converted have selected better carrier prospects

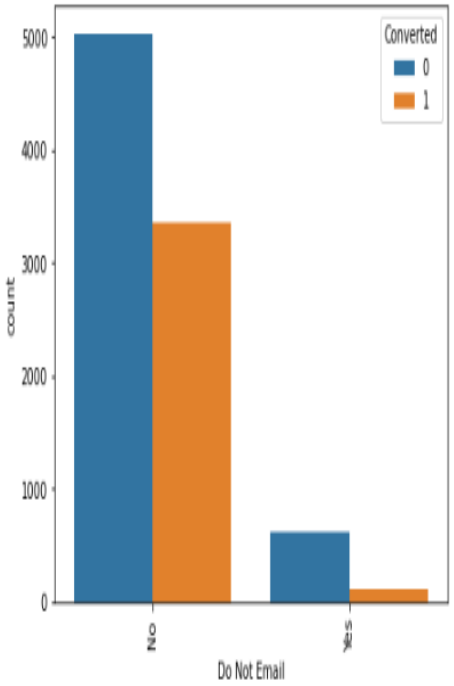
Count Plot For Lead Origin



- In Lead Origin, Maximum Conversation Happened From Landing Page Submission
- In lead origin the second most converts are identify from API

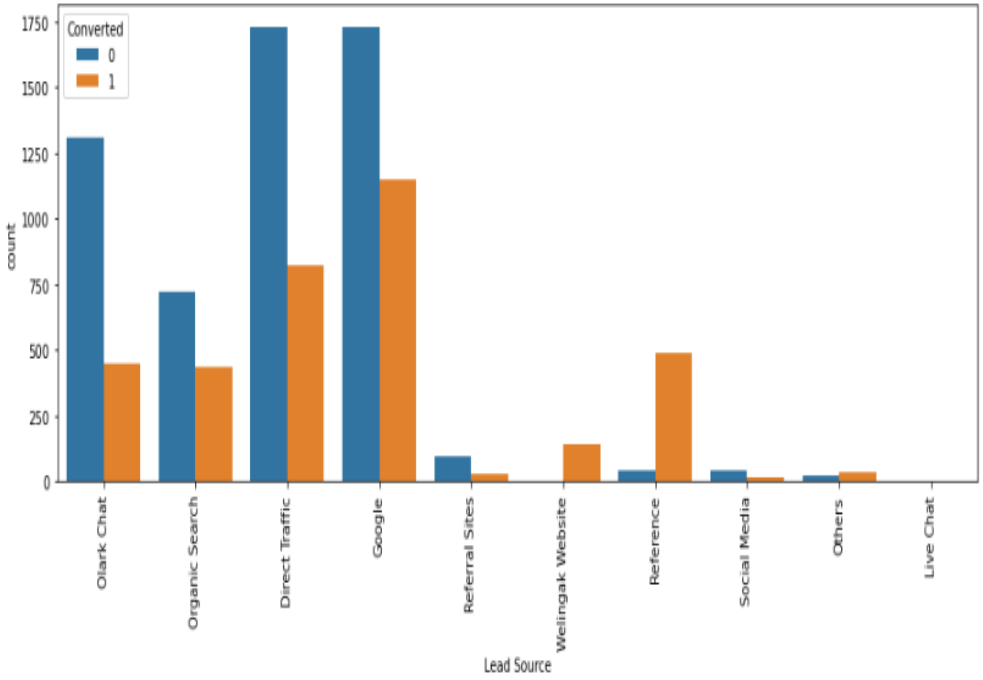
Major Conversion Has Happened From Emails Sent And Calls Made





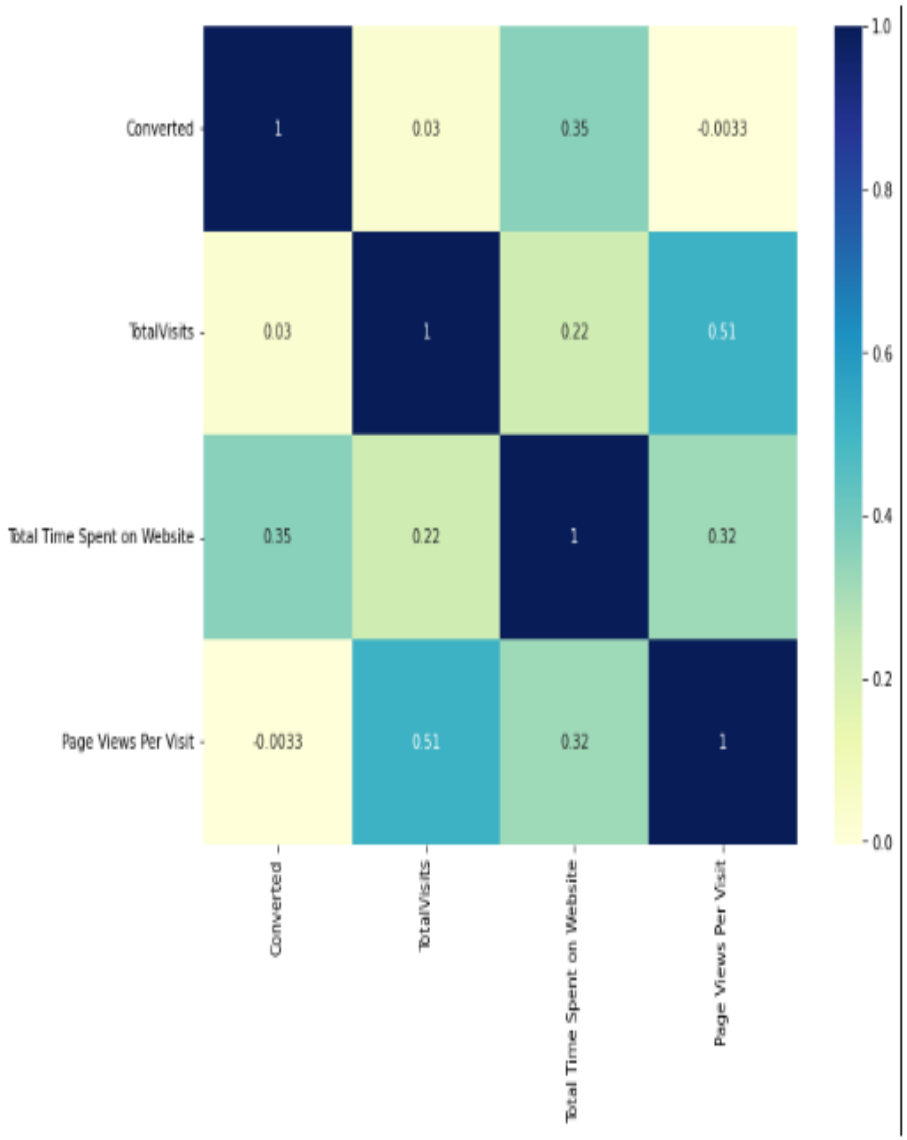
- Most of those converted have selected ‘No’ for do not Call
- Most of those converted have selected ‘No’ for do not Email

Major Conversion In The Lead Source Is From Google



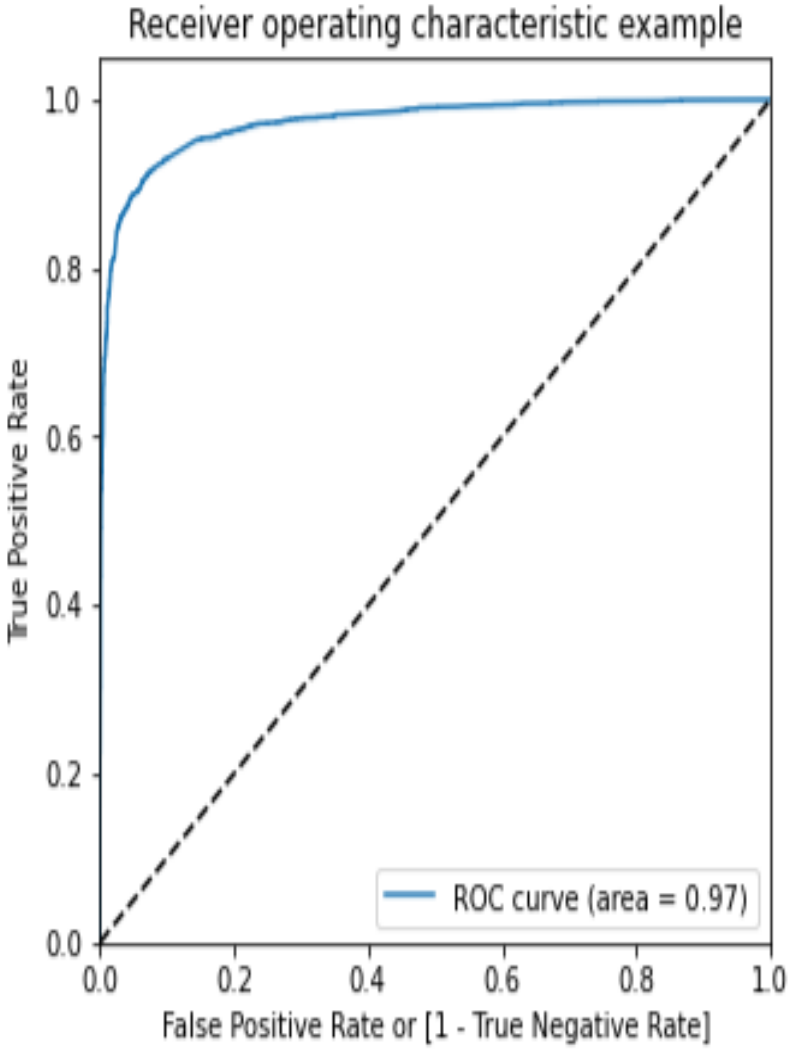
- Major conversion in the lead source is from Google followed by direct traffic, organic search, olark chat and references

HEATMAP



- Page views for visit and total visit have a correlation of 0.51.
- Page views per visit total time spent on website as correlation of 0.32.
- Total time spent on website and converted have a correlation of 0.35.
- Total time spent on website and total visit have a correlation of 0.22

ROC Curve Plot

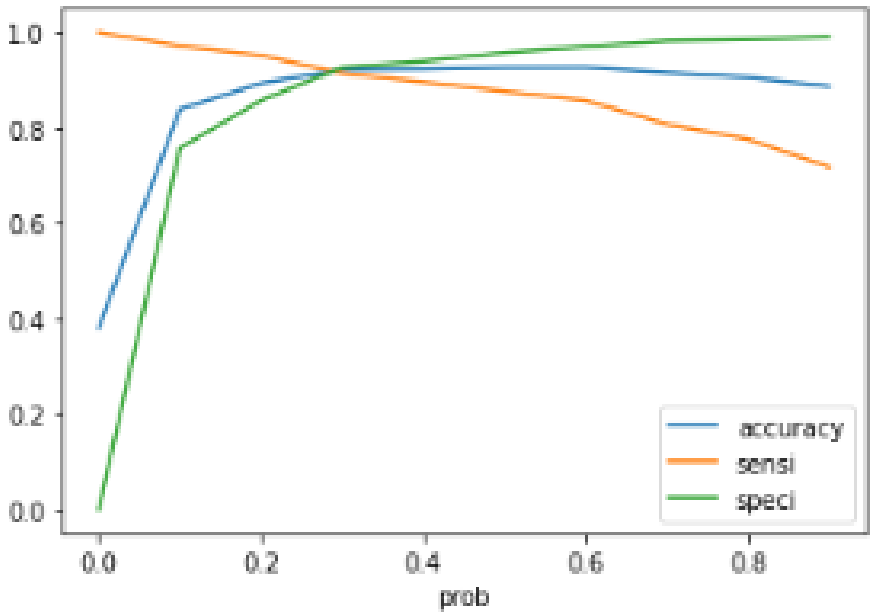


- The ROC curve has an area equals to 0.97

Variables Impacting the Conversion Rate

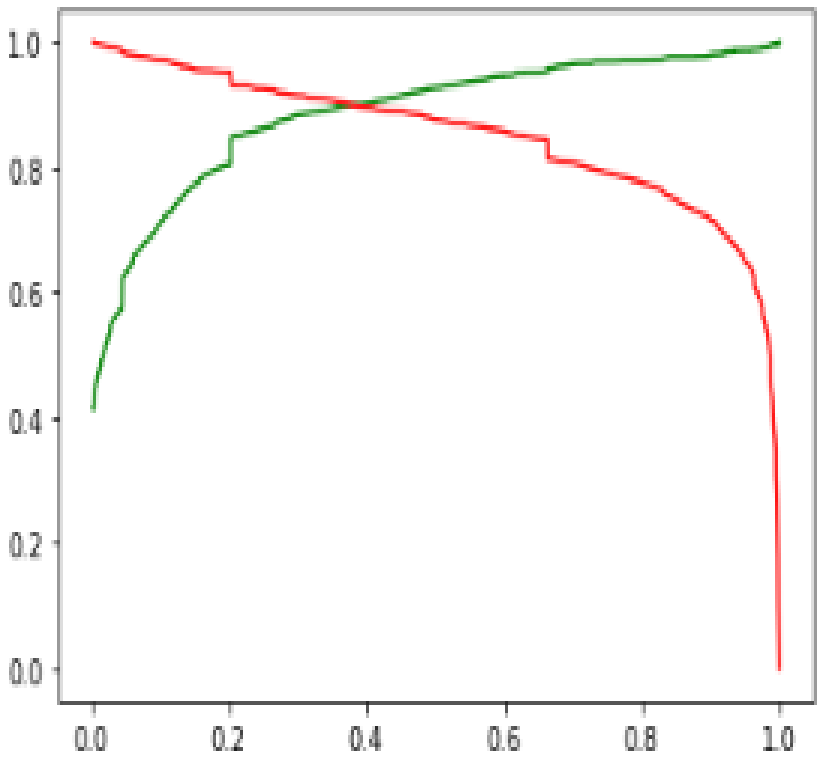
- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent
- Current Occupation – No Information
- Current Occupation – Working Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - Unreachable

The Graph Between Accuracy, Sensitivity, Specificity



- The Graph Depicts An Optimal Cut Off Of 0.35 Based On Accuracy, Sensitivity, Specificity

Graph Between Accuracy & Sensitivity



- The Graph Depicts An Optimal Cut Off Of 0.35 Based On Accuracy, Sensitivity, Specificity

Final Observation Of Text Data And Train Data

- **Observation:**
- After running the model on the Test Data these are the figures we obtain:
- Accuracy : 92.78%
- Sensitivity : 91.98%
- Specificity : 93.26%
- **Final Observation:**
- Let us compare the values obtained for Train & Test:
- **Train Data:**
- Accuracy : 92.29%
- Sensitivity : 91.70%
- Specificity : 92.66%
- **Test Data:**
- Accuracy : 92.78%
- Sensitivity : 91.98%
- Specificity : 93.26%
- The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model

Summary

- The recall score obtained for the train data set was 91.70 & the recall score obtained for test data set was 91.98
- The dummy variable of Tags_Closed by Horizzon with coefficient of 7.3423, dummy variable of Tags_Lost to EINS with coefficient of 5.9447 , dummy variable of Tags_Will revert after reading the email with coefficient of 4.6423 , contributes most towards leads getting converted
- The dummy variable of Tags, Tags_Will revert after reading the email, which has a coefficient of 4.6423 , dummy variable of Lead Source, Lead Source_Welingak Website which has a coefficient of ~ 4 , dummy variable of Lead Origin, Lead Origin_Lead Add Form which has a coefficient of 1.0754 are the categorical variables that should be focused the most by X-education in order to increase the probability of lead conversion
- One possible approach would be to segment the potential leads into different groups based on their predicted lead score of conversion and target the group with the highest predicted lead score first.
- Another approach would be to prioritize Leads with the highest scores for follow-up.