Lead score case study

Insights and Recommendations for X Education

Using Logistic regression

SUBMITTED BY

Prashant

Priyadharshni

Prashant Saraswat

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. They have process of form filling on their website after which the company that individual as a lead.
- 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%. Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot Leads.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Business Objective

- Lead X wants us to build a model to give every lead a lead score between 0 -100 . So that they can identify the Hot leads and increase their conversion rate as well.
- The CEO want to achieve a lead conversion rate of 80%.
- They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches.

Data Overview

- **Dataset**: Leads data from the past three years.
- **Total Records**: 9,240 leads.
- **Features**: 37 columns including lead origin, lead source, activity, and more.

Data Cleaning and Preparation

- Removed irrelevant columns.
- Handled missing values.
- Created dummy variables for categorical columns.

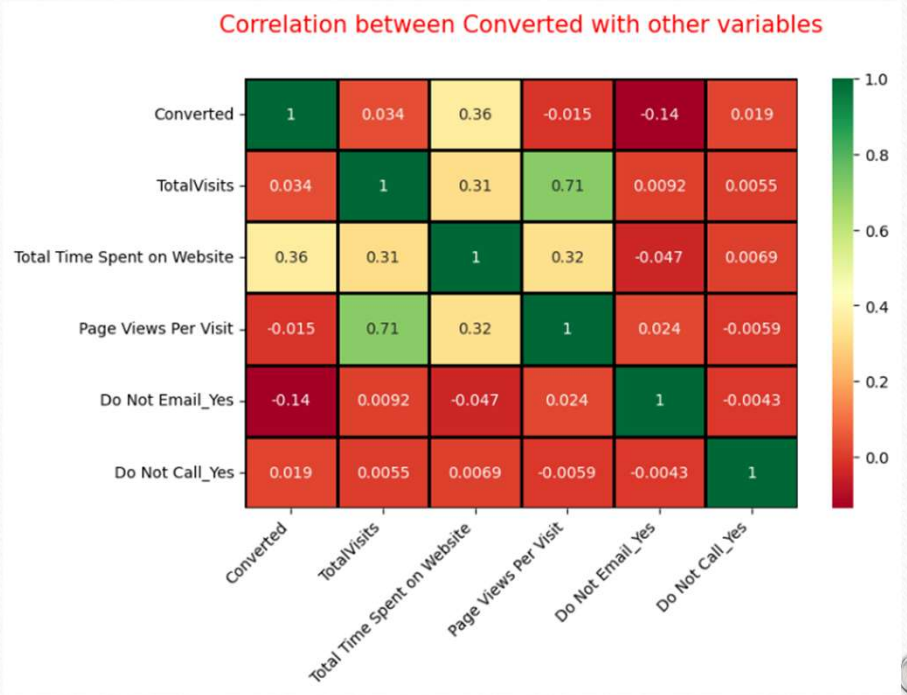
Exploratory Data Analysis

- Higher conversion rates for leads spending more time on the website.
- Significant lead sources: Welingak Website, Olark Chat.
- Important activities: SMS Sent, Email Opened.

Coorelation

- Target variable 'Converted' has positive correlation with 'Total Time Spent on Website'(0.36), 'TotalVists'(0.034) and'Do Not Call_Yes'(0.019).

- Target variable 'Converted' has negative correlation with 'Page Views Per Visit'(-0.015) and 'Do Not Email_Yes'(-0.014)



Feature Selection

- Total Time Spent on Website
- Lead Source_Welingak Website
- Last Activity_SMS Sent
- Current Occupation_Working Professional

Model Building

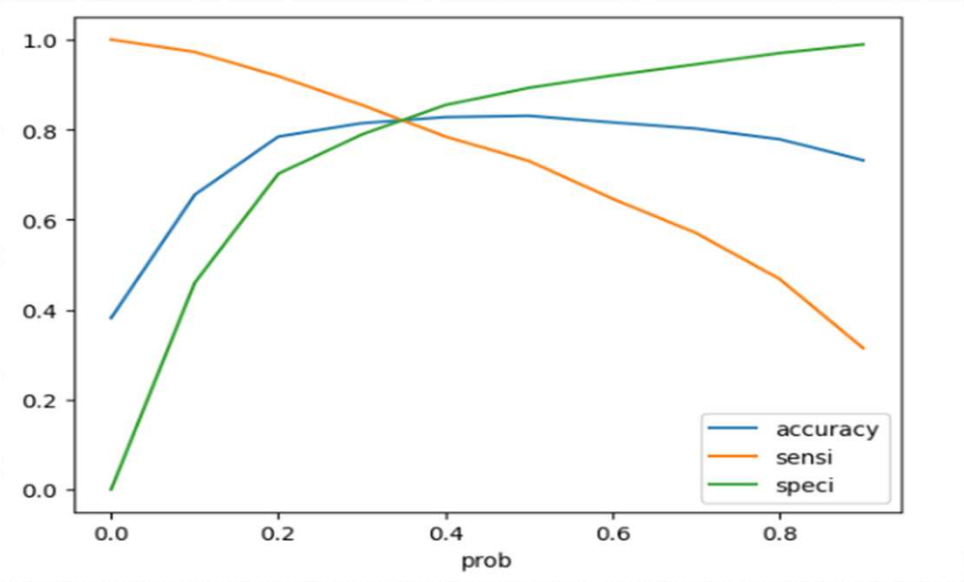
- **Model Used**: Logistic Regression
- **Feature Selection**: Recursive Feature Elimination (RFE)
- **Selected Features**: 20 most significant features based on RFE.

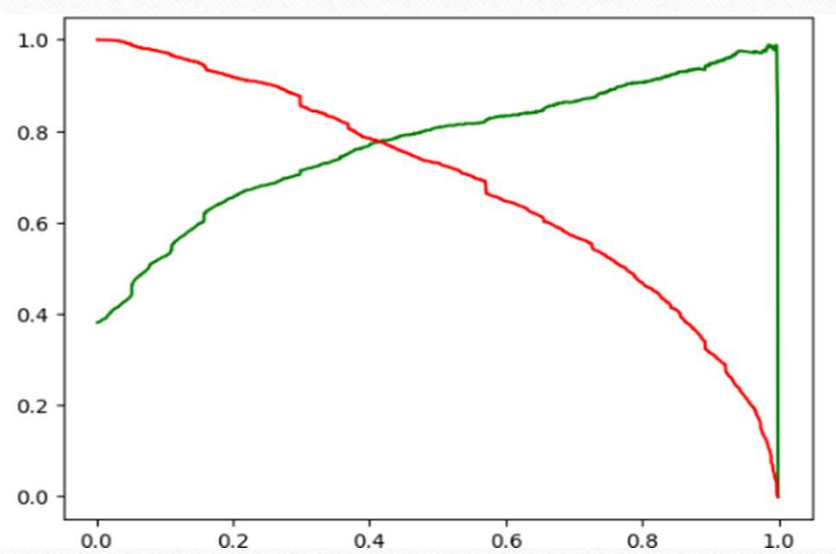
Model Evaluation

- **Metrics**:
  - Accuracy
  - Precision
  - Recall
  - F1 Score

- **Results**: Achieved high accuracy and precision in predicting lead conversions.

From the ROC curve above, 0.37 is the optimum point to take it as a cutoff probability.





Observation

Comparing the values obtained for Train & Test:

Train Data:
- *Accuracy : 82.3 %*
- *Sensitivity : 80.5 %*
- *Specificity : 83.5 %*

Test Data:
- *Accuracy : 82.7 %*
- *Sensitivity : 79.7 %*
- *Specificity : 84.7 %*

Final Features List

Lead Origin_Lead Add Form
Lead Profile_Lateral Student
Lead Source_Welingak Website
Current Occupation_Working Professional
Last Activity_Had a Phone Conversation
Recent Activity_Unreachable
Last Activity_SMS Sent
Lead Profile_Potential Lead
Lead Source_Olark Chat
Total Time Spent on Website
Last Activity_Email Opened
Recent Activity_Olark Chat Conversation
Recent Activity_Modified
Course Selection Criteria_Other
Specialization_Hospitality Management
Do Not Email_Yes
const
Lead Profile_Student of SomeSchool

Recommendations

- The company should focus its outbound calling efforts on leads originating from "Welingak Websites"
- Prioritize calling to more people as People who had conversation with employee about courses are more likely to convert.
- The company should prioritize calls to leads who have spent more time on the website.
- Prioritize calls to working professional leads, lead with profile as Potential Leads and Lateral Student.
- Develop personalized scripts based on lead information (e.g., industry, company size, pain points) to increase engagement.
- The company should contact the leads whose most recent activity was sending an SMS, as they are more likely to convert.