Lead Scoring Case Study

BY : SHAWN WILSON

Problem Statement

▶ X Education sells online courses to industry professionals.

▶ They get a lot of leads however the lead conversion rate is very poor. For example: Out of 100 leads they get, only 30 will be 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 communication with the potential leads rather than making calls to everyone.

Business Objective

- X education wants to know the most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

Solution Methodology

- Data cleaning and Manipulation
- Check for missing values
- Dropped columns if they contained a high amount of missing values and also dropped columns if they were irrelevant for the analysis.
- EDA
- Univariate Data Analysis: Value count, distribution of variables etc.
- Bivariate Data Analysis: Correlation Coefficients
- Feature scaling and Dummy Variables creation
- Classification technique: Logistic regression used for the model building and prediction
- Model Evaluation
- Conclusion

Data Manipulation

▶ Total Number of rows=37 , total number of columns= 9240

▶ Dropped columns that had a high amount of missing values such as Asymmetrique Activity Index', 'Asymmetrique Activity Score', 'Asymmetrique Profile Index', 'Asymmetrique Profile Score', 'Lead Quality', 'Tags', 'How did you hear about X Education'

▶ Dropped 'City' and 'Country' as they were not useful for the analysis

▶ Certain columns had the option "select" meaning the user did not give their input or data. We identified and checked every column for this and if the value count was high , we dropped that particular column

▶ Certain columns like Do Not Call, Search, Recieve More Updates About our Courses, Get Updates on DM Content etc has a very high number of No's as compared to Yes ,so we decided drop them as they won't help with our analysis

▶ We also dropped Null rows of columns that were significant for the analysis and 68% of the rows were retained.

Data Conversion

▶ Dummy variables were created for categorical or object type variables

▶ Scaled data using MinMaxScaler function for better analysis

Model Building

▶ Split the Data into Training and Test Sets

▶ Train-Test Split was performed in the ratio 70:30

▶ Used RFE for Feature Selection

▶ Checked For VIFs and the resulted values were in the appropriate range

▶ Removed Variables whose P- Value was grater than 0.5

▶ Overall accuracy of the model was 78%

ROC Curve
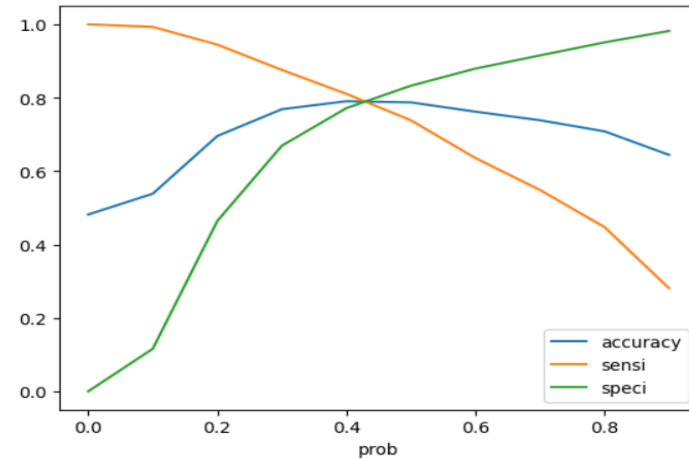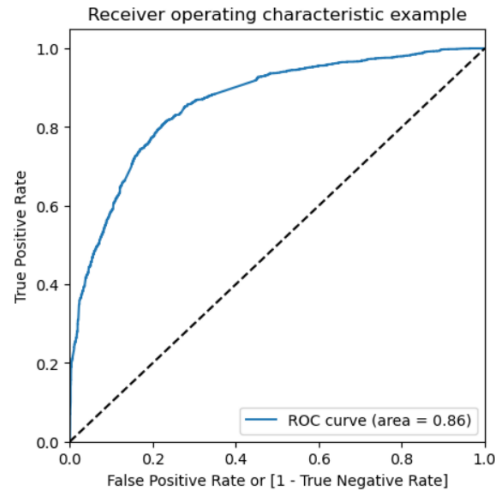
- Found the optimal cut off which was 0.42
- Checked for sensitivity and specificity
- Checked for overall accuracy of the model

Summary

▶ In the initial stage, there is a large number of leads generated, but only a small fraction of them actually become paying customers. To increase lead conversion, it is important to effectively nurture potential leads during the middle stage. This involves educating them about the product, maintaining constant communication, and building a relationship.

▶ To prioritize the most promising prospects among the generated leads, you should analyze variables such as 'TotalVisits,' 'Total Time Spent on Website,' and 'Page Views Per Visit.' These factors play a significant role in determining the likelihood of lead conversion.

▶ It is crucial to maintain a comprehensive list of leads to keep them informed about new courses, services, job opportunities, and future educational prospects. By closely monitoring each lead, you can personalize the information you send to them. Tailoring job offerings, information, and courses based on the specific interests of the leads will greatly enhance the chances of capturing them as potential customers.

▶ Give special attention to converted leads. Engage in question-answer sessions with them to gather the necessary information. Conduct further inquiries and schedule appointments to understand their intentions and mindset regarding joining online courses. This approach will help you gain valuable insights and tailor your offerings to meet their needs effectively.

Conclusion

▶ The variables 'TotalVisits,' 'Total Time Spent on Website,' and 'Page Views Per Visit' are the variables and factors that play a significant role in determining the likelihood of lead conversion

▶ Top 3 categorical/dummy variables to increase probability are Lead Source with elements google, Lead source with elements direct traffic, lead source with elements organic search

▶ Phone calls should be initiated with individuals in the following scenarios which will yield results:

▶ When they demonstrate a high level of engagement on the website, which can be achieved by making the website captivating and encouraging them to revisit.

▶ When they repeatedly return to the website, indicating sustained interest.

▶ When their most recent activity involves SMS or engaging in an Olark chat conversation.

▶ When they are identified as working professionals.