LEAD SCORING CASE STUDY

By : Narendar M

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
- Certaincolumns 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

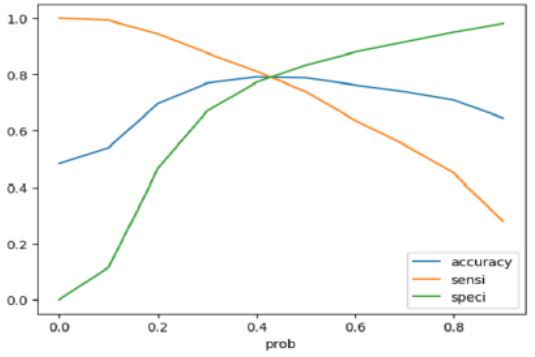
Scaled data using MinMaxScaler function for better analysis

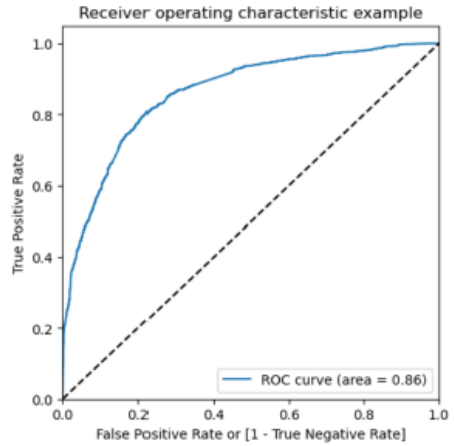
MODEL BUILDING

- 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

- During the initial stage, numerous leads are generated, but only a few convert into paying customers. To improve lead conversion, it is vital to nurture potential leads in the middle stage. This includes educating them about the product, maintaining regular communication, and fostering a strong relationship.
- To identify the most promising prospects among the generated leads, it is essential to analyze variables like 'TotalVisits,' 'Total Time Spent on Website,' and 'Page Views Per Visit.' These factors greatly influence the likelihood of lead conversion.
- Maintaining a comprehensive list of leads is vital for keeping them informed about new courses, services, job opportunities, and future educational prospects. Personalizing the information and tailoring job offerings, information, and courses based on their specific interests greatly increases the likelihood of capturing them as potential customers.
- Focus on converted leads. Engage in Q&A sessions to gather information, conduct inquiries, and schedule appointments to understand their intentions and mindset about joining online courses. This approach provides valuable insights to tailor offerings and meet their needs effectively..

CONCLUSION

- Factors such as 'TotalVisits,' 'Total Time Spent on Website,' and 'Page Views Per Visit' significantly influence lead conversion.
- To increase the likelihood of conversion, prioritize leads from sources like Google, direct traffic, and organic search.
- Initiate phone calls in the following scenarios:
- High website engagement and encouraging revisits
- Repeated visits indicating sustained interest
- Recent SMS activity or engagement in Olark chat
- Identification as a working professional.