LEAD SCORING

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. THE COMPANY REQUIRES YOU TO BUILD A MODEL WHEREIN YOU NEED TO ASSIGN A LEAD SCORE TO EACH OF THE LEADS SUCH THAT THE CUSTOMERS WITH A HIGHER LEAD SCORE HAVE A HIGHER CONVERSION CHANCE AND THE CUSTOMERS WITH A LOWER LEAD SCORE HAVE A LOWER CONVERSION CHANCE. THE GOAL IS TO FIND THE PROMISING LEADS.THE TARGET LEAD CONVERSION RATE HAS TO BE AROUND 80%.

DATA UNDERSTANDING

Leads dataset is provided with around 9000 data points. This dataset consists of various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc. which may or may not be useful in ultimately deciding whether a lead will be converted or not. The target variable, in this case, is the column ‘Converted’ which tells whether a past lead was converted or not wherein 1 means it was converted and 0 means it wasn’t converted.

APPROACH AND METHODOLOGY

- Null Values were calculated in the data.
- Columns having more than or equal to 40% null values were deleted.
- Correlation of variables with the TARGET variable was checked and the variables which had no relation with TARGET variable were deleted.
- Remaining variables with Missing values were then either imputed or removed.
- Numerical Variables were imputed with mean, median , mode which ever suited best.
- Categorical Variables were imputed with the most popular category

- Outliers were then identified using boxplots for the numerical variable.
- Some valid outliers which were there in the data were not deleted and binning/Capping was done for such variables.‘
- The data is now clean and ready for analysis.

METHODOLOGY

- Data Inspection
- Data Cleaning
- Exploratory Data Analysis
- Data Preparation
- Features Selection Using RFE
- Model Building
- Plotting the ROC Curve
- Model Evaluation
- Precision And Recall
- Making Predictions on the Test Data
- Inferences

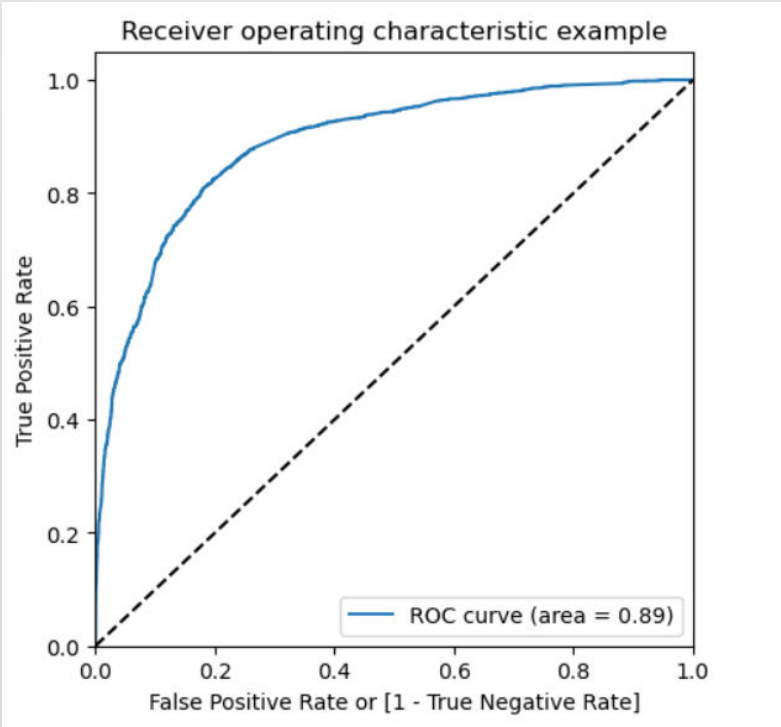
ANALYSIS
UNIVARIATE ANALYSIS

- LEAD ORIGIN was analysed and it was found that **API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.**
- **Google and Direct Traffic are generating the maximum number of Leads**
- **Reference has highest conversion rate.**
- **SMS sent has the highest conversion rate.**
- **Most of the leads have the Email opened as the last activity**
- **Working professional has highest conversion rates.**

FOLLOWING VARIABLES WERE DROPPED AS THEY HOLD NO SIGNIFICANCE FOR ANALYSIS

- 'Lead Number’
- 'Country’
- 'Search'
- 'Magazine’
- 'Newspaper Article’
- 'X Education Forums’
- 'Newspaper’
- 'Digital Advertisement’
- 'Through Recommendations'
- 'Receive More Updates About Our Courses‘
- ’Tags’
- 'Update me on Supply Chain Content’ ,'Get updates on DM Content',
- 'City', 'I agree to pay the amount through cheque’, 'A free copy of Mastering The Interview'

THE ROC CURVE



Since we have higher (0.89) area under the ROC curve , therefore our model is a good one.

CONCLUSION

- To improve overall lead conversion rate, we need to focus more on improving lead converion of API and Landing Page Submission origin and generate more leads from Lead Add Form.
- focus should be on improving lead converion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.
- Website should be made more engaging to make leads spend more time.

- Comparing the values obtained for Train & Test:
  - Train Data:
    - Accuracy : 81%
    - Sensitivity : 81.4%
    - Specificity : 81.2 %
  - Test Data:
    - Accuracy : 80%
    - Sensitivity : 80%
    - Specificity : 80%

Thus we have achieved our goal of getting a ballpark of the target lead conversion rate to be around 80% . 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 to get a higher lead conversion rate of 80%.