Executive Summary – Lead Scoring Case Study

Steps followed for the Case Study analysis

- Step 1: Importing Data

- Step 2: Inspecting the Data frame

- Step 3: Data Preparation: To do an initial data pre-processing

  - ➢ Handling missing values & imputing them

  - ➢ Handling columns with imbalanced data

  - ➢ Handling outliers

- Step 4: Data Visualizations

- Step 5: Data Preparation – II

  - ➢ Converting binary variables (Yes/No) to 0/1

  - ➢ For categorical variables with multiple levels - create dummy features

  - ➢ Test-Train Split

  - ➢ Feature Scaling

  - ➢ Looking at Correlations

- Step 6: Model Building

  - ➢ Feature Selection Using RFE

  - ➢ Assessing the model with Stats Models

  - ➢ Evaluate model accuracy, confusion matrix & other metrics

  - ➢ Plotting the ROC Curve

  - ➢ Finding Optimal Cutoff Point

- Step 7: Making predictions on the test set - Assign Lead Score & Hot Leads

Model Results

- ➢ A Logistic Regression Model has been built and trained by the team to achieve 82% lead conversion rate and it achieves the same accuracy for test data as well.

- ➢ A lead score between 0 and 100 is assigned to each of the leads which can be used by the company to target potential leads.

- ➢ The most potential leads, also known as 'Hot Leads' can be predicted by the model created.

- As per the final model, the following features have Positive Coefficients:

  - Total Time Spent on Website

  - Lead Source_Olark Chat

  - Lead Source_Reference

  - Lead Source_Welingak Website

  - What is your current occupation_Working Professional

  - Last Activity_Email Opened

  - Last Activity_SMS Sent

  - Last Activity_Unreachable

- As per the final model, the following features have Negative Coefficients:

  - Lead Source_Direct Traffic

  - Last Activity_Converted to Lead

  - Last Activity_Email Bounced

  - Last Activity_Olark Chat Conversation

Model Evaluation

- Model Evaluation Metrics on train data:

  - Accuracy - 77.92

  - Sensitivity - 84.23

  - Specificity - 71.83

  - Precision - 74.25

  - Recall - 84.23

- Model Evaluation Metrics on test data:

  - Accuracy - 77.69

  - Sensitivity - 82.25

  - Specificity - 73.58

  - Precision - 73.72

  - Recall - 82.25

- Receiver Operating Characteristic curve:

  - The ROC curve looks and the area under the curve is 0.86.

Recommendations

- ➢ The company should target the leads who spent "more time on the Website" since they are more likely to get converted.

- ➢ The company should make calls to the leads coming from the lead sources – "Welingak Websites, Reference & Olark Chat" as these are more likely to get converted.

- ➢ The company should reach out to the leads who are "Working Professionals" since they have more chances of conversions.

- ➢ The company should make calls to the leads who have last activity as "Email Opened, SMS Sent & Unreachable".

- ➢ The company should NOT make any calls to the leads coming from "Direct Traffic" since it has very less chances of conversion.

- ➢ The company should NOT make calls to the leads who have last activity as "Converted to Lead, Email Bounced & Olark Chat Conversation".