

Lead Score Summary

This is to identify the Lead score for the potential customers for the X Education. The data include 9000 records with 37 variables. Some of the variables are;

- Total visits to the website
- Total time spent on the website
- Last activity
- Tags
- Lead Source

Steps followed:

1. EDA and Data Cleaning
 - a. Data included 9000 records and 37 variables
 - b. Missing value
 - i. Variables with missing value greater than 40% are dropped from the dataset
 - ii. Others are impute using mode
 - c. Categorical variables like Lead Sources, Specialization, Occupation, Tags are collapsed to meaningful groups
 - d. Numerical variable like Total Time Spent on Website and Page Views Per Visit included outlier. Removed top and bottom data to remove the outlier
 - e. Dropped columns which included imbalanced distribution of values like Do Not Call
2. Data Preparation
 - a. Created dummy variables for categorical variables
 - b. The split was done at 70% and 30% for train and test data respectively
 - c. Scaling of numeric data using standard scaler
3. Modelling
 - a. RFE used to attain the top 15 relevant variables
 - b. Built logistic regression model.
 - c. Variables were removed manually depending on the VIF values and p-value and redone the Model
 - d. Model 10 is selected as final model where all variables have $VIF < 5$ and $p\text{-value} < 0$
 - e. Model predictions was first done at 0.5 cut off
 - f. Based on ROC curve, the optimum cut off value is ~ 0.4 is cut-off
 - g. Comparison of Metrics based on Optimal cut-off is ~ 0.4 ;
 - i. Train Data:
 1. Accuracy : 90.39%
 2. Sensitivity : 88.22%
 3. Specificity : 91.7%
 4. Recall : 86.12%
 5. Precision : 89.38%
 - ii. Test Data:
 1. Accuracy : 90.39%
 2. Sensitivity : 88.22%
 3. Specificity : 91.7%

4. Recall : 86.12%
5. Precision : 89.38%

Recommendation:

Company should make calls to the leads with below category;

- Tags with Closed by Horizon
- Tags with Will revert after reading the email
- Leads who are Working Professional, Student and Unemployed
- Lead Source from Welingak Website

Company should not make calls to the leads with below category;

- Lead Origin with Landing Page Submission
- Specialization Not Specified
- Lead opted for Do Not Email

Hot Leads:

There are 708 leads which can be contacted and have a high chance of getting converted.