

## **Summary – Lead Scoring Case Study**

The case study is for X Education. Students and industry professionals can take courses at X Education. The goal of the case study is to improve the cold calling/sales strategy to customers in order to increase the number of people who buy the courses. To accomplish this goal, we used the dataset provided, which contains information about the consumer (such as time spent on the website, most recent notable action, and so on), and ran a series of analyses to create a model that can forecast convertible leads. The stages involved in doing the analysis and developing the model are outlined below.

- 1. Data Reading:** Reading the data and understanding the columns and values.
- 2. Cleaning Data:** The data provided was partly cleaned. Although some of the data in the columns were required to be cleaned and organized. The 'select' values in the column were changed to NULL as 'select' did not specify anything about the column. The country column was changed to 'India', 'Outside India' and 'Not Provided'. Some of the columns having low percentage of NULL values, the null values were imputed as 'Not Provided'.
- 3. Exploratory Data Analysis:** EDA was performed on the data to analyse the data. We found that many of the categories in categorical columns are not significant for the model. No outliers were found in the data.
- 4. Data Preparation:** We created the dummy variables and removed the dummy variables with the syntax: 'ColumnName\_NotProvided', as 'not provided' categories were not needed.
- 5. Train-Test Split:** Train Test split was done on the data, with, 70% train and 30% test data.
- 6. Building Model:** The model was built and feature selection was done by RFE. Top 15 columns were taken as output. The columns were dropped on the basis of VIF( $VIF > 5$ ) and p-values ( $p\text{-value} > 0.05$ ).
- 7. Model Evaluation:** We made the confusion matrix, and found out the optimal cut-off value using the ROC curve. The accuracy, sensitivity, specificity were found out.
- 8. Making Prediction:** Prediction was made on the test data, with the optimal cut-off value of 0.35 and accuracy, sensitivity, specificity of approx. 80%.
- 9. Precision-Recall:** With the current cut-off value of 0.35, Precision was found to be around 78% and Recall was found to be around 70%.

**It was found that the variables that mattered the most in the potential buyers are (In descending order) :**

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was:
  - a. Google
  - b. Direct traffic
  - c. Organic search
  - d. Welingak website
4. When the last activity was: a. SMS b. Olark chat conversation
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional.

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.