

Summary:

The analysis is done on X Education company which sells online courses to industry professionals. The aim of this case study is to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The following are the steps implemented for the analysis and understanding of the data

1.Understanding and Cleaning the data:

After understanding the basic structure of the data, number of null values in the data were found. If any column has more than 50% of the data as null, then the column was dropped. If the missing values are less in number, then they are imputed.

2.EDA:

Exploratory Data Analysis was done on the data including univariate and bivariate analysis which was helpful in visualizing the data and finding any outliers present in the data

3.Dummy Variables:

Dummy variables were created for categorical variables by converting the data into binary variables. StandardScaler was used here

4.Splitting of the data in to train and test sets:

The data was split into train set and test set in 70% and 30% proportion.

5.Model Building:

Recursive Feature Elimination (RFE) and Variance Inflation Factor were used in retaining the most relevant variables in the data. In this case the number of variables are 14

6.Model Evaluation:

An ROC curve was made and the cut off was found to be 0.89. A confusion matrix was made and accuracy, specificity and sensitivity were calculated.

7.Prediction:

With the cut off as 0.35 the precision is 79% and recall is 70.9%

8.Precision-Recall:

With the cut off as 0.41 the precision is 76.1% and recall is 76.6%

The following are the relevant variables for the data set in order;

1. Last Activity_Other_Activity
2. Lead Source_Welingak Website
3. Last Notable Activity_Olark Chat Conversation
4. Do Not Email
5. What is your current occupation_Working Professional?

6. Last Activity_Page Visited on Website
7. Total Time Spent on Website
8. Lead Source_Reference
9. Last Notable Activity_Modified
10. Lead Source_Olark Chat
11. Last Activity_SMS Sent
12. Last Activity_Email Opened
13. Specialization_Others
14. Lead Origin_Landing Page Submission