

Summary

Analysis is done for X Education to find ways to have more industry professionals to join their platform for courses. Data was given about how potential customers visit their site, how much time they spend etc.

The following steps were performed during the analysis:

A. Data Cleaning:

There were a few null values although lot of data was partially clean, some null values were changed to 'NP' as not to lose much data. Many customers were from India and very few were from outside India.

B. Exploratory Data Analysis (EDA):

EDA was done to check the data where it was found that in the numeric values no outliers were found and the categorical data seemed irrelevant.

C. Creation of Dummy Variables:

Dummy variables were created where 'NP' elements were removed and for the numeric values MinMaxScaler was used.

D. Train-Test Split:

For train and test data split was done at 30% and 70% respectively.

E. Building the Model:

RFE was used to find top 15 relevant variables. Rest of the variables were manually removed taking into consideration the VIF values and p-value.

F. Evaluating the Model:

Optimum cutoff value using ROC curve was used to find the Accuracy, sensitivity and specificity which was around 80%.

G. Prediction:

Test data frame was used for prediction and the optimum cutoff value as Accuracy of 80.27%, sensitivity of 70.85% and specificity of 88.73%.

Variables that mattered the most in the potential customers were:

1. Total number of visits
2. Total time they spent on the website
3. When the last activity was SMS
4. When the Lead origin or Lead add form
5. When the current occupation is:
 - a. Student
 - b. Unemployed
 - c. Working professional

X Education can increase the number of potential buyers to buy their courses by implying and focusing on the above variables.