

Summary

This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The basic data provided gave us a lot of information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate.

The following are the steps used:

1. Data Cleaning: The data was partially clean except for a few null values. Few of the null values were changed to 'not provided' so as to not lose much data. Although they were later removed while making dummies. Since there were many from India and few from outside, the elements were changed to 'India', 'Outside India' and 'not provided'.

2. Exploratory Data Analysis: EDA done and observe. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seem good and no major outliers were found in the variables.

3. Dummy Variables Creating: The dummy variables were created and then drop out the necessary variables for train-test split. For numeric values we used the MinMaxScaler.

4. Train-Test split: The split was done at 70% and 30% for train and test data respectively.

5. Model Building: Firstly, RFE was done to attain the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value.

6. Building Model For Evaluation: A confusion matrix was made. Later on the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 96% each.

7. Prediction: Prediction was done on the test data frame and with an optimum cut off as 0.3 with accuracy, sensitivity and specificity of 96%.

8. Precision – Recall: This method was also used to recheck and a cut off of 0.41 was found with Precision around 73% and recall around 75% on the test data frame.

9. Conclusion from the Analysis: It was found that the variables that are the most of potential buyers are:

- The lead source was:
 - Google
 - Direct traffic
 - Organic search
 - Welingak website
- The last activity from observation are:
 - SMS
 - Olark Chat conversation
- The total time spend on the Website.
- Total number of visits.