

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.

Technical Approach:

1. Read and Cleaning of the Data:

The data was partially clean except for a few null values and the option select had to be replaced with a null value since it did not give us much information. Few of the null values were changed to 'not provided' 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. EDA

We performed Univariate and Bivariate analysis to identify the relationship between the data. Some of the modification and capping of the data was performed (outlier treatment).

3. Dummy Variable Creation and Test Train Split:

We performed the scaling of data using Standard scaler. The dummy variables were created, and original elements were removed.

4. 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 (The variables with $VIF < 5$ and $p\text{-value} < 0.05$ were kept). We then performed some iteration to identify the best model by removing variables and reached to a final model.

5. Evaluation of Model

A confusion matrix was made. Later, the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 85% and 92% respectively.

6. Prediction

Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 88% and 90% respectively.

7. Precision and Recall

This method was also used to recheck and a cut off was found to be 0.41 with Precision around 85% and recall around 88% on the test data frame.