Summary

The following are the steps followed:

1. Cleaning up dataset:

2. EDA (Univariate and Multivariate Analysis:

3. Created Dummy Variables for all categorical variables:

4. Train-Test split:

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

5. Model Building:

6. Model 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 80% each.

7. Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 80%.

8. Precision – Recall:

Its observed that the maximum time spend on the website, total number of visits, the lead source (like Google, Direct traffic, Organic search, Welingak website), last activity (*viz.,* SMS, Olark chat conversation), lead origin and when their current occupation is as a working professional are the key factors to predict the course buyers.

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.