

Summary

An education company named X Education sells online courses to industry professionals.

There are a lot of leads generated in the initial stage but only a few of them come out as paying customers. The potential leads need to be nurtured well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.

The dataset shared gave a lot of details such as people who frequently visit the site, spend a lot of time on the website, etc.

The following steps were performed on the data :-

- 1) Exploratory Data Analysis
- 2) Cleaning of data – handling missing/incorrect values, outliers, dropping columns/rows
- 3) Univariate & bivariate analysis
- 4) Creation of dummy variables for categorical data
- 5) Scaling of numerical variables to use in the model
- 6) Train-test split of the data at 70%-30%
- 7) Logistic regression model building – Using RFE to identify the top 15 variables to be used. Using VIF and p-value to remove the columns further from our model
- 8) Model evaluation – Creating a confusion matrix and then we create an ROC curve to find the cut off and report on the accuracy, sensitivity and specificity. Accuracy for the trained model came out to be 87.17%
- 9) Prediction – Prediction was done on the test data and the accuracy came out to be 87.71% which is quite good and above the target of 80% which was given to us

Some important inferences from our analysis are mentioned below.

The sales team should focus on variables such as :-

- 1) Total visits and total time spent on the website
- 2) What is the Lead source – Welingak website, Reference
- 3) What is the Lead origin – Lead add form
- 4) Is the individual a working professional, student or unemployed

Keeping these themes in mind and selectly targeting the correct individuals will lead to a high conversion percentage for X education.