

Lead Score Case Study Summary

This analysis is done for X Education to find potential leads to increase the conversion rate i.e., 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:

Cleaning data:

The data was partially clean except for a few null values and the columns with more than 3000 null values are removed. While the other rows with null values are deleted retaining the columns. The variables City, Country, Prospect ID, and Lead number are removed as they are not much used in the analysis. All the columns are analysed for values in them and the variables with 'select' or if only one value has major percentage to decrease the bias in the model.

EDA:

A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seemed good and no outliers were found.

Dummy Variables:

The dummy variables were created and later few dummies were removed as required.

Train-Test split:

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

Scaling:

The numeric variables are scaled using MinMaxScaler to make them uniformly distributed with mean 0 and standard deviation as 1.

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-value >0.05 are removed).

Model Evaluation:

A confusion matrix was generated, and the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 78% each.

Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.42 with accuracy, sensitivity, and specificity of 78%.

Precision – Recall:

This method was also used to recheck and a cut-off of 0.44 was found with Precision around 78% and recall around 77% on the test data.

It was found that the variables that mattered the most in the potential buyers are (In descending order):

- Total Visits
- Total Time Spent on Website
- Lead Origin
 - Lead Add form
- Lead Source
 - Olark Chat
 - Welingak Website
- Current occupation as student or unemployed

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.