

SUMMARY

This analysis is done for X education and to find ways to get more industry professionals to enroll in their courses.

The following steps were used:

1.Cleaning the Data

The data was partially clean except for a few null values and the option select had to be replaced with null values since it did not give much information. Few of the null values were changed to 'not provided' so as to not lose much of data. Although they were later removed while making dummies.

2.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 were good and no outliers were found.

3.DUMMY VARIABLES

The dummy variables were created and later on the dummies with 'not provided' elements were removed. For numeric values we use MinMaxScaler.

4.Train-Test Split

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

5.MODEL BUILDING

The RFE was done to attain the top 15 relevant variables. Later the rest of variables were removed manually depending on the VIF values and p-value.

6.MODEL EVALUATION

A confusion matrix was made. Later on the optimum cut off value (using ROC value) was used to find 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%.

It was found that variables that mattered the most in the potential buyers are:

1. The total time spent on the website.
2. Total number of visits.
3. When the lead source was:
 1. Google
 2. Direct traffic
 3. Organic research
 4. Welingak Website
4. When the last activity was:
 - SMS
 - Olark chat conversation
5. When the lead origin is Lead add format.

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 minds and buy their courses