

Summary

This analysis has been done for X Education and to find various ways to get more leads (potential learners) to join their courses. The basic data provided provides 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 steps were followed:

1. Cleaning data:

The data cleaning was done by dropping unnecessary columns except for a few null values and the option 'select' had to be dropped since it did not give us much information. Instead non-null values were taken for some of the columns to take care of missing or null values and not lose much data. Although they were later removed while making dummies.

2. EDA:

A quick EDA was done to check the condition of data. The numeric values seems good and no outliers were found.

3. Dummy Variables:

The dummy variables were created and later the dummies with original name of the columns were removed. For numeric values we used the MinMaxScaler.

4. Train-Test split:

The split was done at 80% and 20% for train and test data respectively.

5. 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).

6. Model Evaluation:

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 between 75% - 80%.

7. Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.42 with accuracy, sensitivity and specificity between 77% - 80%.

8. Precision – Recall:

This method was also used to recheck and a cut-off: 0.44 was found with Precision around 74% and recall around 80% on the test data frame.

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

S.No	Features	VIF
1	What is your current occupation_Unemployed	2.85
2	Total Time Spent on Website	2.00
3	Total Visits	1.58
4	Last Activity_SMS Sent	1.51
5	Lead Source_Olark Chat	1.34
6	Lead Source_Reference	1.12
7	Do Not Email_Yes	1.08
8	Lead Source_Welingak Website	1.07
9	What is your current occupation_Student	1.07
10	Last Activity_Had a Phone Conversation	1.01

X Education can flourish as they have a very high chance to get almost all the potential buyers to converted to purchasing the courses and join the institute for their growth.