

Summary

This analysis is done for X Education and to find 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:

1. Cleaning data:

The data was partially clean except for a few null values and the option select had to be replaced with a null value since it did not give us much information. Few of the null values were changed to 'other' so as to not lose much data. Although they were later removed while making dummies. Since there were many from India and few from outside, the null elements were changed to 'India'.

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 seem 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 used the MinMaxScaler.

4. Train-Test split:

The split was done at 70% and 30% 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 on the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 91% each.

7. Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.2 with accuracy, sensitivity and specificity of 91%.

8. Precision-Recall:

This method was also used to recheck and a cut-off of 0.2 was found with precision around 93.32% and recall around 85.15% on the test data frame.

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

- Do not Email
- Last notable activity was SMS sent.
- When the lead source was:
 - Wellingak website
- When the tags were:
 - Busy
 - Closed by Horizon
 - Lost to EINS
 - Ringing
 - Will revert after reading email
 - Switched Off
- When the lead origin is lead add form and lead import.
- When the lead quality was:
 - Not sure
 - Worst
- When the current occupation was:
 - Unemployed

Keeping the above mentioned points in the mind the X education can increase all the potential buyers to change their mind and buy their courses.