

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.

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. Dropped columns with more than 25% Data missing.

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 seems 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.

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 80% each.

7. Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.37 with accuracy, sensitivity and specificity around 80%.

8. Precision – Recall:

This method was also used to recheck and a cut off of 0.41 was found with Precision around 70% and recall around 78% on the test data frame.

Recommendation:

To improve the potential lead conversion rate X-Education will have to mainly focus on below TOP 3 features responsible for good conversion rate are :-

1) Lead Origin_Lead Add Form:

- We should focus on more budget in terms of advertising, etc. to attract more leads.

2) Last Notable Activity_SMS Sent :

- We can provide discounts for providing references that convert to lead to encourage more references.

3) Last Activity_other_act:

- We should communicate over the phone call