

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 were the steps that we followed:

1. Cleaning the data:

The data was almost clean for the most fields but a majority of the fields had missing values.

We replaced those nulls with 'Not Provided' for most of the fields and later dropped them after not finding any relevant insight.

2. EDA:

We did EDA and found some of the categorical variables were not relevant and we decided to drop them.

3. Dummy Variables:

The dummy variables were created and later the dummies with 'not provided' elements were removed. For numeric values we used the standard scaler.

4. Train-Test split:

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

5. Model Building:

Firstly, we did RFE to attain the top 15 relevant variables. At a later point rest of the variables were removed manually depending on the VIF values and p-value ($VIF < 5$ and $p\text{-value} < 0.05$).

6. Model Evaluation:

A confusion matrix was created. Later, the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be approximately 79% each.

7. Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.35 with Accuracy as 80%, sensitivity as 83% and specificity as 77%.

8. Precision – Recall:

This method was also used to recheck and a cut off of 0.4 was found with Precision around 68% and recall around 83% on the test data frame.

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

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was:
 - b. Direct traffic
 - c. Organic search
 - d. Welingak website
4. When the last activity was:
 - a. SMS
 - b. email bounced
5. When the lead origin is Lead add form.
6. When their current occupation is as a working professional.

Noting the above insights, X Education can use the results to market their products to the target audience in a more planned way. This will definitely help them to expand their business.