

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 potential customers visit the site, the time they spend there, how they reach the site and the conversion rate.

The following are the steps used:

1.cleaning data:

The data was partially clean except there were null values and the option select had to be replaced with null value since it did not give us much information. We drop the columns which has more than 30% null values and other null values we replaced with median and most frequent values.

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. We did the univariate, bivariate analysis and did the outlier analysis. We found the upper range outliers and we did outliers treatment.

3.dummy variables:

The dummy variables were created and later dropped the repeated columns. We used the StandardScaler

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 and p-value

6.Model Evaluation:

A confusion matrix was made. Later on the optimum cut off value was used to find the accuracy, sensitivity and specificity which came to be around 80%

7.Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity around 80%

8.Precision Recall:

This method was also used to recheck and a cut off of 0.35 was found with precision around 75% and recall around 76% on the test data frame.

To improve the potential lead conversion rate x-education will have to mainly focus important features responsible for good conversion rate are:

- Lead_Source_welingak Website : as conversion rate is higher for those leads who got to know about course from 'Welingak Website' so company can focus on this website to get more number of potential leads.
- Lead Origin_Lead Add Form: leads who have engaged through 'Lead Add Form' having higher conversion rate so company can focus on it to get more number of leads cause have a higher chances of getting converted.
- What is your current Occupation_Working Professional: the lead whose occupation is 'Working Professional' having higher lead conversion rate, company should focus on working professionals and try to get more number of leads.