

Lead Scoring Case Study – 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. Although they were later removed while making dummies. Since there were many from India and few from outside, the elements were changed to "India", "Outside India" and "not provided".

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 we 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.

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 0.35 with accuracy, sensitivity and specificity of 80%.

8. Precision-Recall:

This method was also used to recheck and a cut off of 0.41 was found with Precision around 73.24% and recall around 76.61% on the test data frame.

There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well in order to get a higher lead conversion.

First, sort out the best prospects from the leads you have generated. 'TotalVisits', 'Total Time Spent on Website', 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted.

Then, keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads. Focus on converted leads. Hold question-answer sessions with leads to extract the right information you need about them. Make further inquiries and appointments with the leads to determine their intention to join online courses.