

SUMMARY & RECOMMENDATION

SUMMARY:

This analysis is done for X Education to find ways to get more industry professionals to join their courses. From the problem statement provided gave us lot of information about how the potential customer visit the site, the time they spend there, the source from which they reached the site and conversation rate. Towards the start there are lot of leads generated in the initial stage but only few (~30%) of them are converted to paying customers.

In the middle stage, the focus was on the potential leads like educating the leads about the product, constantly communicating etc. so that more leads can get converted as students. From the analysis we have identified that 'TotalVisits', 'Total Time Spent on Website', 'Leads Origin' which contribute most towards the probability of a lead getting converted as students. To come to conclusion and recommendation.

We performed below steps:

1. Data Cleaning: 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.

2. Exploratory Data Analysis: An EDA was performed to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. So, in later steps we have removed the same. The numeric values seem good after a few tweaks, and no outliers were found.

3. Dummy Variables: The dummy variables were created for categorical variables. And for numeric values we used the Min-Max Scaler for scaling.

4. Train- Test Split: We split the data into training & testing data in 70:30 ratio.

5. In Model Building RFE was done to obtain 15 variables b. Based on the VIF (< 0.05) other variables were eliminated and obtained a total of 9 variables

6. Model Evaluation:

a. Confusion matrix was used.

b. ROC was used to find accuracy, sensitivity and specificity

c. Precision – Recall: This method was also used to recheck and a cut off

RECOMMENDATION:

From the analysis we have identified that 'Total Visits', 'Total Time Spent on Website', 'Leads Origin' which contribute most. After the entire exercise it is observed below are top 3 hot leads

- Total Time Spent on Website category
- Lead Origin category
- Total number of visits.

There is high probability of conversation of leads with the last activity is SMS.

When the lead origin is Lead add format.

Time spent on website is also key factor to be considered as the members are interested and have high conversion probability.