

SUMMARY

1. 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. Few of the null values were replaced by fillna while for other negate was used as to not lose much data. Although they were later removed while making dummies.

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.

3.Dummy Variables:

- The dummy variables were created. 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. Later some of the variables were removed manually depending on 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 ROC curve was obtained.

7. Prediction:

- Prediction was done on the test data frame.