

SUMMARY

We have analysed the dataset of **X Education** and 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.

As per the old flow the conversion rate was **39%**

The following are the steps used for increasing the conversion rate:

1. Cleaning data:

The dataset was partially clean except null value. We have removed null values and columns as follows.

- columns having null value more than **30%** is removed.
- The Rows having Maximum Numbers of **SELECT** and **NO** are removed
- We have removed null values for modelling.

2. EDA:

A quick EDA was done to check the condition of our data. It was found that

- Lot of elements in the categorical variables were irrelevant.
- The numeric values seems good and some outliers were found.

3. Dummy Variables:

- The dummy variables were created by adding Specialization
- We have removed dummies with Select Elementa
- 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.
- The rest of the variables were removed manually depending as below
 - VIF more than 5 were removed
 - P value more than 0.5 were removed

6. Model Evaluation:

A confusion matrix was made. Later the optimum cut off value (using ROC curve where **Area under ROC curve is 0.86**) was used to find as below

- Accuracy = 79%
- Sensitivity = 78.4%
- Specificity = 79.6%

7. Prediction:

Prediction was done on the test data frame and with an **optimum cut off as 0.42** with data as below

- Accuracy = 79%
- Sensitivity = 78.4%
- Specificity = 79.6%

8. Precision – Recall:

This method was also used to recheck and a **cut off of 0.44** was found with below data:

- Precision : 77%
- Recall : 80%

As per the analysis below are the top 3 columns with variable which need to focus for increasing the conversion up to **79%**

1. Total Time Spent on Website
2. Total Visits on Website
3. What is your current Occupation
 - Working Profession
 - Student