

## **Summary**

The analysis is done for X-Education and to find ways to get more learners to join their courses. The basic data provided gave us a lot of information about how the potential customers visit the website, explore, and spend time in this website.

The following steps used:

### **1. Clearing data:**

The data was partially clean except for few null values and the option select had to be replaced with null values. The columns which have more than 30% of null values we drop the columns, and rest we treat the null values to minimize the loss of data.

### **2. EDA:**

EDA has 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 for the categorical variables.

### **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 variables. Later the rest of the variables were removed manually depending on the VIF values and P-value (VIF should be  $<5$  and p-value should be  $<0.05$ )

### **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 test data frame and with an optimum cut off as 0.4 with accuracy, sensitivity and specificity of 81%.

### **8. Precision-Recall:**

This method was used to recheck and a cutoff of 0.41 was found with Precision-recall around 74% and recall around 76% on the test data frame.

It was found out that the variables that mattered the most in the potential buyers are:

1. The total time spent on website.
2. Total number of visits.
3. When the Lead source was:
  - a) Olark chat
  - b) Wellingak Website
4. When the last activity was:
  - a) SMS

- b) Olark Chat Conversation.
- 5. When the lead origin is lead add form.
- 6. When the current occupation was:
  - a) Working professional
  - b) Student
  - c) Unemployed.
  - d) Other.

Considering the other mentioned points in mind the X-education can increase all the potential buyers to change their mind and buy their courses.