

# Summary

## **Problem Statement**

An education company named X Education sells online courses to industry professionals. Now, although X Education gets a lot of leads, its lead conversion rate is very poor. We need to nurture the potential leads well in order to get a higher lead conversion.

The basic data provided gave us a 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. Few of the null values were changed to the value that had the maximum values so as to not lose much data.

### **2. EDA:**

EDA was done to check the if the columns are useful and how do they impact the conversion rate. Found some outliers. There were categorical variables which were irrelevant.

### **3. Dummy Variables:**

Created the dummy variables for categorical variables. 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 (The variables with  $VIF < 5$  and  $p\text{-value} < 0.05$  were kept).

### **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%, 65% and 89.6% respectively.

### **7. Prediction:**

Prediction was done on the test data frame and with an optimum cut off as 0.385 with accuracy, sensitivity and specificity of 80%.

### **8. Precision – Recall:**

This method was also used to recheck and a cut off of 0.43 was found with Precision around 72.3% and recall around 74.2% on the test data frame.

This analysis successfully identified key factors influencing lead conversion for the professional training course. By focusing on website engagement, targeting relevant user segments (unemployed, students, management interests), and optimizing marketing efforts high conversion area, we can potentially increase enrollment rates and drive business growth

The conversion rate can be higher if the X education team goes ahead with the given result.