

## Summary :

The analysis is completed for X education and found 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 below steps are involved in it:

### 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 null values were changed to 'not provided' so as to not lose much data. Although they were later removed while making dummies. Since there were many from India and few from outside, elements were changed to 'India', 'Outside India' and 'not provided'.

### 2. EDA:

EDA was done to check the condition of our data. It was found that a lot of elements in categorical variables were irrelevant. The numeric values seems good and no outliers were found.

### 3. Dummy variables :

Dummy variables were created and later dummy with 'not provided' elements were removed. 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 :

RFE was done to attain the top 15 relevant variables. Rest of the variables were removed manually depending on VIF values and p-value.(The variables with  $VIF < 5$  and  $p\text{-value} < 0.05$  were kept ).

### 6. Model Evaluation :

Confusion matrix was created. The optimum cut off value was used to find the accuracy, sensitivity and specificity which came to be around 80% each.

### 7. Prediction:

Prediction was done on the test data and with optimum cutoff as 0.35 with accuracy, sensitivity and specificity of 80%.

#### 8. Precision-Recall:

The Precision-Recall method was also used to recheck and a cut off of 0.41 was found with precision around 73% and recall around 75% on test data.

It was found that the below variables are mattered most in potential buyers in descending order:

1. The total time spent on the website.
2. Total number of visits
3. When the lead source was: Google, direct traffic, Organic search, Welingak website.
4. When the last activity was: SMS, Olark chat conversation.
5. When the lead origin is Lead add format.
6. When the current occupation is as a working professional.

By having all these above points in mind, X Education can flourish as they have a very high chance to get almost all potential buyers to change their mind and buy their courses.