

## Summary

This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The 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:**

We dropped the variables that had high percentage of NULL values in them. This step also included imputing the missing values as and where required with median values in case of numerical variables and creation of new classification variables in case of categorical variables. The outliers were identified and removed.

### 2. **EDA:**

Then we started with the Exploratory Data Analysis of the data set to get a feel of how the data is oriented. In this step, there were around 3 variables that were identified to have only one value in all rows. These variables were dropped. The numeric values seems good and no outliers were found.

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

The dummy variables were created and later on the dummies with 'not provided' elements were removed.

### 4. **Train-Test split:**

The next step was to divide the data set into test and train sections with a proportion of 80-20% values.

### 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.

### 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 79% each.

### **7. Prediction:**

Prediction was done on the test data frame and with an optimum cut off as 0.45 with accuracy, sensitivity and specificity of 80%.

### **8. Precision – Recall:**

This method was also used to recheck and a cut off of 0.44 was found with Precision around 78% and recall around 79% on the test data frame.

There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion. First, sort out the best prospects from the leads you have generated. 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted. Then, You must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads. A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects. Focus on converted leads. Hold question-answer sessions with leads to extract the right information you need about them. Make further inquiries and appointments with the leads to determine their intention and mentality to join online courses.