

This analysis is carried out for X Education in an effort to attract more industry professionals to their courses. We learned a lot from the fundamental data on how potential customers use the site, how long they stay there, how they got there, and the conversion rate.

The following are the steps used:

1. **Data Cleaning:**

Most of the data was clean, except for a few null values, and the option 'Select' had to be changed to a null value because it didn't provide much information. To avoid losing too much data, mean/median were used to replace a small number of the null values. Dummies for categorical columns were produced.

2. **Exploratory Data Analysis:**

A quick EDA was conducted to assess the state of our data and to graph/visualize it. Many of the category variables' components were discovered to be unnecessary and were removed. For numerical columns, outliers treatment was carried out.

b) Dummy Variables: The dummy variables were designed for categorical columns. We utilised the MinMaxScaler to scale numerical numbers.

3. **Test-Train Split and Model Building:**

For train and test data, the split was done at 70% and 30%, respectively. The top 15 pertinent variables were then determined using RFE. Later, the analysis was conducted using variables having a VIF score of 5 or lower and a p-value less than equal to 0.05.

4. **Model Evaluation:**

Confusion matrix for the model was created. Later, a decision was made regarding the best cutoff value (using the ROC curve) based on the accuracy, sensitivity, and specificity values. Accuracy on the training set was 81%, and accuracy on the test set was 80% when the model was later assessed for the same threshold.

5. **Precision and Recall value for test dataset:**

Recall being around 80% and precision being around 71% on the test data frame.

It was found that the variables that mattered the most in the potential lead conversion are as follows:

1. The total time spend on the Website.
2. Do not Email.
3. When the lead source was:

- a. Olark Chat
- b. Welingak website
- 4. When the last activity was:
 - a. Converted to Lead
 - b. Olark chat conversation
- 5. When the lead origin is Lead add format.
- 6. When their current occupation is as a working professional to name few.

With these in mind, X Education can succeed since they have a very good probability of convincing practically all prospective students to change their minds and enroll in their courses.