Summary

This analysis is carried out for X Education in an effort to attract more business 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. Cleaning data:

Most of the data was clean, save for a few null values, and the option pick had to be changed to a null value because it didn't provide much information. To avoid losing too much data, only a small number of the null values were changed to 'not provided'. Nevertheless, they were later taken out while manufacturing dummies. The elements were altered to "India," "Outside India," and "not provided" because there were a lot of people from India and a small number from elsewhere.

2. EDA:

To quickly assess the state of our data, an EDA was performed. It was discovered that several of the categorical variables' components were unnecessary. The numerical numbers appear accurate, and no outliers were discovered.

3. Dummy Variables:

The fake variables were made, then later the fakes with the 'not provided' bits were taken away. The MinMaxScaler was utilised for numerical values.

4. Train-Test split:

The split was done at 70% and 30% for train and test data respectively.

5. Model Building:

First, the top 15 pertinent factors were determined by RFE. Later, based on the VIF values and p-value, the remaining variables were manually deleted (the variables with VIF 5 and p-value 0.05 were retained).

6. Model Evaluation:

A matrix of confusion was created. Later, the accuracy, sensitivity, and specificity were determined using the ROC curve, and they all came to be about 80% each.

7. Prediction:

On the test data frame, predictions were made using an optimal cutoff of 0.35 and had 80% accuracy, sensitivity, and specificity.

8. Precision – Recall:

On the test data frame, this procedure was also utilised to recheck, and a cut off of 0.41 was discovered with precision and recall averaging about 73% and 75%, respectively.

It was found that the variables that mattered the most in the potential buyers are (In descending order)

1. The total time spend on the Website.

2. Total number of visits.

3. When the lead source was: a. Google b. Direct traffic c. Organic search d. Welingak website

4. When the last activity was: a. SMS b. Olark chat conversation

5. When the lead origin is Lead add format.

6. When their current occupation is as a working professional

With these in mind, X Education can succeed since they have a very good probability of persuading nearly all prospective customers to change their minds and purchase their courses.