

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. Data Cleaning:

Except for a few null values, the data was mostly clean. However, the option selected had to be changed to a null value because it provided little useful information. To avoid losing too much data, a few 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 outside.

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 figures are accurate, and no outliers were discovered.

3. Dummy Variables:

After the dummy variables were made, those that had the phrase "not provided" were later deleted. We utilized the MinMaxScaler to scale numerical values.

4. Train-Test split:

For train and test data, the split was done at 70% and 30%, respectively.

5. Building the model:

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 $p\text{-value} < 0.05$ were kept).

6. Evaluation of the model:

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. Prediction – Recall:

On the test data frame, a cut off of 0.41 was discovered using this method, with precision around 73% and recall around 75%.

According to research, the following factors affected potential purchasers the most (in descending order):

- The total time spend on the Website
- Total number of visits
- When the lead source was:
 - a. Google
 - b. Direct traffic
 - c. Organic search
 - d. Welingak website
- When the last activity was:
 - a. SMS
 - b. Olark chat conversation
- When the lead origin is Lead add format
- 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.

X-----X-----X-----X