

Summary

To uncover strategies to attract more business people to their courses, X Education conducted this investigation. We learned a lot from the fundamental statistics 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:
Except for a few null values, the data was mostly clean. However, the option chosen had to be changed to a null value because it provided little useful information. To avoid losing too much information, some of the null values were converted to 'not given'. Nevertheless, they were eventually taken out while manufacturing dummies. The components were altered to "India," "Outside India," and "not given" because there were more people from India than from other countries.
2. EDA:
To quickly assess the state of our data, an EDA was performed. It was discovered that several of the category variables' components were unnecessary. No outliers were discovered, and the numerical values appear to be satisfactory.
3. Dummy Variables:
After the dummy variables were made, those that had the phrase "not given" were later deleted. We unutilized MinMaxScaler to scale numerical numbers.
4. Train-Test split:
For train and test data, the split was done at 70% and 30%, 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 came to be approximately 80% for each.
7. Prediction:
On the test data frame, the prediction was made with an ideal cut-off of 0.35 and 80% accuracy, sensitivity, and specificity.
8. Precision — Recall:
On the test data frame, a cut-off of 0.3 was discovered using this method, with a precision of roughly 68% and a recall of almost 85%.

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

1. The total time spent on the Website.
2. Reason of course selection
3. When their current occupation is as a working professional.
4. When the lead source was:
 - a. Google
 - b. Direct traffic
 - c. Organic search
 - d. Welinger website
5. When the last activity was:
 - a. SMS
 - b. Olark chat conversation
6. When the lead origin is Lead add format.

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.