

Final-Summary

To draw more business people to their courses, X Education is implementing this analysis. The basic information on how prospective clients use the site, how long they remain there, how they got there, and the conversion rate taught us a lot.

The procedures are as follows:

1. Cleaning data:

The data was mainly clean, with the notable exception of a few null values. However, because the option wouldn't provide much helpful information, it had to be switched to a null value. Some of the null values were converted to "not provided" in order to prevent losing too much data; however, these were eventually removed while making dummies. There were many persons from India and few from elsewhere, therefore the elements were changed to "India," "Outside India," and "not provided."

2. EDA (Exploratory Data Analysis):

A short EDA was carried out to evaluate the status of our data. It was determined that some of the components of the descriptive analysis were superfluous. No inconsistencies were found, and the numerical values are correct.

3. Dummy Variables:

The dummy variables were created, and afterwards those that had the word "not provided" were removed. We scaled numerical values using the MinMax-Scaler.

4. Train-Test split:

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

5. Model Building:

RFE first identified the top 20 relevant parameters. The subsequent step was to manually remove the remaining variables based on the VIF values and p-value (the variables with VIF 5 and p-value 0.05 were retained).

6. Model Evaluation:

It resulted in a confusing matrix. Later, using the ROC curve, the accuracy, sensitivity, and specificity were calculated, and they all came out to be about 80% each.

7. Prediction:

Using an ideal threshold of 0.35, predictions on the test data frame attained 80% accuracy, sensitivity, and specificity.

8. Precision – Recall:

Using this method, a cut off of 0.41 was found for the test data frame, with a precision and recall of about 72% and 80%, respectively.

According to study, the following individuals had the most effects on potential buyers (in descending order):

1. TotalVisits
2. Lead Source_Google
3. Lead Source_Direct Traffic
4. What is your current occupation_Unemployed
5. Total Time Spent on Website

6. Last Notable Activity_Modified
7. Lead Source_Organic Search
8. Last Activity_Olark Chat Conversation
9. Last Notable Activity_Email Opened
10. Lead Origin_Lead Add Form
11. Last Notable Activity_Olark Chat Conversation
12. What is your current occupation_Working Professional
13. Lead Source_Welingak Website
14. Do Not Email_Yes
15. Last Notable Activity_Page Visited on Website
16. Lead Source_Referral Sites
17. What is your current occupation_Student
18. Last Notable Activity_Email Link Clicked

Keeping these in mind, X Education has a strong chance of being successful because they have a very good chance of persuading almost all potential students to alter their thoughts and enroll in their courses.