

Summary

This analysis was done for X Education to figure out how to attract more professionals from the industry to enrol in their courses. The initial data provided us with a lot of useful information about how potential customers visit the site, how long they stay there, how they find the site, and the rate at which they actually sign up for courses.

Here's what we did:

Cleaning the data	The data was mostly clean, but there were a few missing values that needed to be addressed. We replaced the "select" option with a null value since it didn't give us useful information. Some null values were changed to "not provided" to retain data. Later, they were removed while creating dummy variables. We categorized locations into "India", "Outside India", and "not provided"
Exploratory Data Analysis (EDA)	We quickly checked the condition of our data. We found that some elements in categorical variables were irrelevant, but the numeric values looked good with no outliers.
Creating Dummy Variables	We made dummy variables and removed those with "not provided" elements. For numeric values, we used the Min Max Scaler
Train-Test Split	We split the data into 70% for training and 30% for testing.
Building the Model	We used Recursive Feature Elimination (RFE) to select the top 15 relevant variables. Then, we manually removed the rest based on VIF values and p-values (keeping variables with $VIF < 5$ and $p\text{-value} < 0.05$).
Evaluating the Model	We created a confusion matrix and used the ROC curve to find the optimal cut-off value, resulting in an accuracy, sensitivity, and specificity of around 80%.
Making Predictions	We made predictions on the test data with an optimal cut-off of 0.35, achieving an accuracy, sensitivity, and specificity of 80%.
Precision-Recall	We used this method to double-check and found a cut-off of 0.41, resulting in a precision of around 73% and a recall of around 75% on the test data.

We found that the most important variables for potential buyers are:

- Total time spent on the website
- Total number of visits
- Lead source, especially if it's Google, Direct traffic, Organic search, or the Welingak website
- Last activity, particularly if it's SMS or an Olark chat conversation.
- Lead origin as Lead add format
- Current occupation as a working professional
- By focusing on these factors, X Education can greatly increase their chances of convincing potential buyers to enroll in their courses