

Summary

This analysis was conducted for X Education to identify strategies for attracting more industry professionals to enrol in their courses. The foundational data provided ample information regarding potential customers' site visits, duration of stay, referral sources, and conversion rates. The following steps were employed:

1. Data Cleaning:

The data was largely clean, with a few null values present. The option 'select' was substituted with a null value due to its lack of informational value. Some null values were changed to 'not provided' to preserve data integrity. These were later excluded during the creation of dummy variables. Geographic elements were categorized into 'India', 'Outside India', and 'not provided'.

2. Exploratory Data Analysis (EDA):

A preliminary EDA was performed to assess the condition of the data. Many elements in the categorical variables were found to be irrelevant. The numeric values appeared satisfactory, with no outliers detected.

3. Dummy Variables:

Dummy variables were created, and those with 'not provided' elements were subsequently removed. For numeric values, the MinMaxScaler was applied.

4. Train-Test Split:

The dataset was split into 70% for training and 30% for testing.

5. Model Building:

Recursive Feature Elimination (RFE) was used to identify the top 15 relevant variables. Subsequently, other variables were manually removed based on Variance Inflation Factor (VIF) values and p-values (variables with $VIF < 5$ and $p\text{-value} < 0.05$ were retained).

6. Model Evaluation:

A confusion matrix was constructed. The optimal cut-off value, determined using the ROC curve, was then utilized to calculate accuracy, sensitivity, and specificity, each approximately 80%.

7. Prediction:

Predictions were made on the test dataset, employing an optimal cut-off value of 0.42, resulting in an accuracy, sensitivity, and specificity of 80%.

8. Precision-Recall:

This method was also employed for verification, identifying a cut-off value of 0.44, with a precision of around 78% and recall of approximately 77.18% on the test dataset. Key factors influencing potential buyers were identified as:

- Total Time Spent on Website.
- Total Number of Visits.
- Lead Source:
 - Olark Chat
 - Wellingak Website
 - Last Activity:
 - SMS
 - Olark Chat Conversation
 - Lead Origin:
 - Lead Add Form
 - Current Occupation:
 - Working Professionals
 - Student
 - Unemployed
 - Other

By focusing on these points, X Education can effectively encourage potential buyers to consider enrolling in their courses.