

Summary

This study for X Education aims to increase the enrollment of industry professionals in their courses. By analyzing user data, insights were drawn regarding site visits, duration, referral sources, and conversion rates.

Methodology

1. Data Cleaning:

- The data was mostly clean, but some null values and uninformative options required adjustment. Null values were relabeled as 'not provided' and subsequently omitted during dummy variable creation. Geographical data was simplified to 'India', 'Outside India', and 'not provided'.

2. Exploratory Data Analysis (EDA):

- Initial EDA revealed irrelevant categories in the categorical variables, while numerical data appeared valid without outliers.

3. Dummy Variables:

- Dummy variables were created, and those with 'not provided' elements were excluded. Numeric values were standardized using MinMaxScaler.

4. Train-Test Split:

- Data was divided into 70% training and 30% testing sets.

5. Model Building:

- Recursive Feature Elimination (RFE) identified the top 15 variables. Further refinement was based on Variance Inflation Factor (VIF) and p-values, retaining variables with $VIF < 5$ and $p\text{-value} < 0.05$.

6. Model Evaluation:

- A confusion matrix was constructed, and the ROC curve helped determine the optimal cut-off value, yielding

approximately 80% accuracy, sensitivity, and specificity.

7. Prediction:

- Predictions on the test data used an optimal cut-off of 0.35, maintaining 80% accuracy, sensitivity, and specificity.

8. Precision-Recall Analysis:

- A cut-off of 0.41 was determined, resulting in 73% precision and 75% recall on the test data.

Key Findings

The most influential factors for potential buyers included:

1. Total time spent on the website.
2. Number of visits.
3. Lead sources such as:
 - Google
 - Direct traffic
 - Organic search
 - Welingak website
4. Last activity such as:
 - SMS
 - Olark chat conversation
5. Lead origin from Lead add format.
6. Current occupation as a working professional.

These insights suggest X Education can significantly increase conversions by focusing on these key variables, effectively targeting almost all potential buyers.