

Analysis for X Education

Done By: Ravneet Singh, Rhenius Rajkumar, Rushabh Daithankar

Objective:

The goal of this analysis is to identify strategies to attract more industry professionals to X Education's courses. The dataset provided insights into customer interactions, such as site visits, time spent, traffic sources, and conversion rates.

Steps Undertaken

1. Data Cleaning
 - a. Addressed null values and replaced the 'select' option with null to improve clarity.
 - b. Some null values were labeled as "not provided" to preserve data during the cleaning phase. Consolidated location data into three categories: "India," "Outside India," and "Not Provided."
2. Exploratory Data Analysis (EDA)
 - a. Analyzed categorical variables to remove irrelevant elements.
 - b. Verified that numeric data had no significant outliers.
3. Feature Engineering:
 - a. Created dummy variables for categorical data, excluding those labeled "not provided."
 - b. Applied MinMaxScaler to normalize numeric values.
4. Train-Test Split
 - a. Data was divided into 70% training and 30% testing sets.
5. Model Development
 - a. Used Recursive Feature Elimination (RFE) to select the top 15 features.
 - b. Further refined variables by removing those with a VIF > 5 or p-value > 0.05.
6. Model Evaluation
 - a. Utilized a confusion matrix and ROC curve to determine an optimal cut-off value, achieving approximately 80% accuracy, sensitivity, and specificity.
7. Prediction and Validation
 - a. On test data, an optimal cut-off of 0.35 provided 82% accuracy, sensitivity, and specificity.
 - b. Precision-recall analysis identified a cut-off of 0.41, yielding precision of 78% and recall of 77%.

Key Insights

The analysis revealed that the following factors most influence customer conversions:

1. Total Time Spent on Website
2. Lead Source_Google
3. Last Notable Activity_Modified
4. Lead Source_Direct Traffic
5. Last Notable Activity_Email Opened
6. What is your current occupation_not provided
7. Lead Origin_Lead Add Form
8. Lead Source_Organic Search
9. Last Activity_Olark Chat Conversation
10. Lead Source_Welingak Website
11. What is your current occupation_Working Professor

- 12. Do Not Email _ Yes
- 13. Lead Source _ Referral Sites
- 14. Last Notable Activity _ Email Link Clicked