

This analysis was conducted for X Education to explore ways to attract more industry professionals to enroll in their courses. The initial dataset provided insights into how potential customers interact with the website, including their visit behavior, time spent on the site, referral sources, and conversion rates.

Steps Followed:

1. Data Cleaning:
 - The dataset was mostly clean, except for some null values. The “option select” entries were replaced with null values since they held no meaningful information.
 - Some null values were labeled as “not provided” to retain as much data as possible. However, these were eventually excluded during the creation of dummy variables.
 - User location data was categorized into “India,” “Outside India,” and “not provided” for simplification.
2. Exploratory Data Analysis (EDA):
 - A preliminary analysis was conducted to assess the dataset.
 - Many categorical variable entries were found irrelevant and cleaned.
 - Numeric variables were in good condition, with no outliers detected.
3. Dummy Variables:
 - Dummy variables were created for categorical data. Entries with “not provided” were removed.
 - MinMaxScaler was applied to scale numeric values.
4. Train-Test Split:
 - The data was split into training (70%) and testing (30%) datasets.
5. Model Building:
 - Recursive Feature Elimination (RFE) was used to identify the top 15 significant variables.
 - Additional variables were manually removed based on Variance Inflation Factor ($VIF < 5$) and p-values (< 0.05).
6. Model Evaluation:
 - A confusion matrix was created to evaluate model performance.
 - The optimum cut-off value was determined using the ROC curve, achieving approximately 80% accuracy, sensitivity, and specificity.
7. Prediction:
 - Predictions were made on the test dataset. Using an optimum cut-off of 0.35, the accuracy, sensitivity, and specificity remained at 80%.
8. Precision-Recall Analysis:
 - This method was used to validate results. An alternative cut-off of 0.41 was identified, achieving 73% precision and 75% recall on the test dataset.

Key Insights:

The following variables were identified as the most influential in determining potential buyers (in descending order of importance):

1. Total time spent on the website.
2. Total number of visits.
3. Lead sources such as:
 - Google
 - Direct traffic
 - Organic search
 - Welingak website
4. Last activity:
 - SMS interaction
 - Olark chat conversations
5. Lead origin from the “Lead Add Form.”
6. Current occupation as a working professional.

Conclusion:

By focusing on these key variables, X Education has a strong opportunity to convert potential buyers into enrolled students. This strategy can significantly enhance their ability to target industry professionals effectively and drive enrollment growth.