

Summary

The analysis was conducted for X Education with the aim of increasing the enrollment of industry professionals in their courses. The initial data provided valuable insights into the website behavior of potential customers, including their visit patterns, time spent on the site, referral sources, and conversion rates.

The following steps were followed:

1. **Data Cleaning:** The data was cleaned by addressing null values and replacing irrelevant options with null or "not provided" values. Certain categorical variables were simplified to "India," "Outside India," or "not provided."
2. **Exploratory Data Analysis (EDA):** A preliminary EDA was performed to assess the data's condition. Categorical variables were examined for relevance, and no outliers were detected in the numeric values.
3. **Dummy Variables:** Dummy variables were created, and dummies with "not provided" values were removed. Numeric values were scaled using MinMaxScaler.
4. **Train-Test Split:** The data was divided into training and testing sets, with a split of 70% and 30% respectively.
5. **Model Building:** Feature selection using Recursive Feature Elimination (RFE) was applied to identify the top 15 relevant variables. Additional variables were manually removed based on VIF (Variance Inflation Factor) and p-value criteria (keeping variables with $VIF < 5$ and $p\text{-value} < 0.05$).
6. **Model Evaluation:** A confusion matrix was generated, and the optimal cutoff value was determined using the ROC curve. This resulted in an accuracy, sensitivity, and specificity of approximately 80% each.
7. **Prediction:** The model was used to make predictions on the test dataset using an optimal cutoff of 0.35, yielding an accuracy, sensitivity, and specificity of 80%.
8. **Precision-Recall:** Precision-Recall analysis was also performed, identifying a cutoff value of 0.41. This yielded a precision of around 73% and a recall of around 76% on the test dataset.

Overall, these steps helped analyze the data and build a predictive model to target industry professionals effectively and improve enrollment rates for X Education's courses.

Based on the analysis conducted, it has been identified that certain variables have a significant impact on potential buyers and their likelihood to change their minds and purchase courses from X Education. These variables, in descending order of importance, are:

Total time spent on the Website: The amount of time potential buyers spend on the X Education website has a strong influence on their decision-making process.

Total number of visits: The frequency of visits to the website also plays a crucial role in engaging potential buyers and increasing the chances of conversion.

Lead source: The source from which the potential buyers found X Education is an important factor. The following lead sources were found to be influential: a. Google b. Direct traffic c. Organic search d. Welingak website

Last activity: The last activity the potential buyers engaged in before making a decision is also significant. The following activities were found to be influential: a. SMS b. Olark chat conversation

Lead origin: When the lead originates from the Lead add format, it has a positive impact on the potential buyer's decision-making process.

Current occupation: Potential buyers who are working professionals are more likely to change their minds and make a purchase.

Considering the importance of these variables, X Education has a high chance of success in converting potential buyers into customers. By focusing on optimizing the website experience, attracting leads from various sources, engaging in effective communication through SMS and chat conversations, targeting lead add formats, and tailoring courses to appeal to working professionals, X Education can maximize its potential to convert potential buyers into satisfied customers.