

Lead Scoring Case Study Summary

Problem Statement:

X Education sells online courses to industry professionals and seeks assistance in identifying the most promising leads – those most likely to convert into paying customers. The company requires a model that assigns a lead score to each lead, indicating higher conversion chances for leads with higher scores and lower conversion chances for those with lower scores. The CEO has specified a target lead conversion rate of approximately 80%.

Solution Summary:

Step 1: Reading and Understanding Data

- First, we read and analyzed the dataset to gain a comprehensive understanding.

Step 2: Data Cleaning

- We addressed variables with a high percentage of NULL values by dropping them
- Imputed missing values for the columns with a low percentage of NULL values
- We have filled the null values with 'Not Shared'
- We also check for outliers and handle them by capping

Step 3: Data Analysis

- Next, we conducted Exploratory Data Analysis (EDA) to understand the dataset better and shared our insights.
- During this process, we have performed the Univariate, Bivariate, and Multivariate Analysis.

Step 4: Creating Dummy Variables

- Binary variables are converted into 0's or 1's.
- Categorical variables were transformed into dummy variables.

Step 5: Train-Test Split

- The dataset was divided into training and testing sets using a 70-30% ratio.

Step 6: Feature Rescaling

- Numerical variables were scaled using Min-Max Scaling.

Step 7: Feature Selection using RFE

- Recursive Feature Elimination (RFE) was employed to select the top 30 most important features.
- Through iterative analysis of P-values, significant variables were retained, while insignificant ones were dropped.
- Ultimately, we identified 21 significant variables with acceptable Variance Inflation Factors (VIF).
- Assuming a threshold of 0.5 for binary classification, we derived Confusion Metrics and evaluated the model's Accuracy, Sensitivity, and Specificity.

Step 8: Plotting the ROC Curve

- We plotted the ROC curve, achieving an area under the curve of 90%, indicating strong model performance.

Step 9: Finding the Optimal Cutoff Point

- We determined the optimal probability cutoff point by analyzing accuracy, sensitivity, and specificity values across different probability thresholds.
- The cutoff point was identified as 0.36, resulting in approximately 82% correct predictions by the model.
- Sensitivity on train set 82%
- Specificity on train set 81%

Step 10: Computing Precision and Recall Metrics

- Precision and Recall metrics were computed, resulting in values of 76% and 78% respectively on the training dataset.
- A cutoff value of approximately 0.41 was chosen based on the Precision-Recall tradeoff.

Step 11: Making Predictions on Test Set

- Finally, we applied our model to the test dataset, predicting conversion probabilities. The model achieved an accuracy of 81%
- Sensitivity on test set 77%
- Specificity on test set 83%.
- Precision on test set 73%
- Recall on test set 77%

Recommendations:

- To improve the potential lead conversion rate, X-Education should focus on the following key features that contribute to higher conversion rates:

1. Lead Source: Welingak Website

- Leads who discovered the course through the 'Welingak Website' have a higher conversion rate. The company should focus on this website to attract more potential leads.

2. Lead Add Form (from Lead Origin)

- Leads who engaged through the 'Lead Add Form' show a higher conversion rate. The company should prioritize this form to capture more leads with a higher chance of conversion.

3. Working Professional (from Current Occupation)

- Leads who are 'Working Professionals' exhibit a higher conversion rate. The company should target this demographic to generate more leads.

4. Phone Conversation or SMS Sent (from Last Notable Activity/Last Activity)

- Leads whose last notable activity was a 'Phone Conversation' or 'SMS Sent' are more likely to convert. The company should consider these activities as indicators of potential leads and focus efforts accordingly.