

This case study investigates lead conversion optimization for an online education organization that markets its courses through third-party search platforms. The CEO aims to increase the current lead conversion rate from 30% to 80%, focusing on identifying promising lead sources and optimizing the customer journey.

Analysis Approach

The study involved data cleaning and imputation to address outliers and missing values. A dataset was split into a 70:30 train-test ratio, followed by exploratory data analysis (EDA), feature selection, and model building. Correlations among parameters were analyzed to determine key factors influencing conversions. Logistic regression was employed as the modeling technique, iteratively refined for precision and predictive power.

Key Findings from EDA

1. Geographic Insights: India, particularly Maharashtra and Mumbai, contributed the most leads.
2. Lead Sources: Google was the primary source, followed by landing page submissions.
3. Audience Insights: Most leads were unemployed individuals motivated by better career prospects.
4. Unnecessary Variables: Columns like "Lead Profile" and "How did you hear about X Education" were dropped due to low utility or repetitive data.

Data Preparation

Several preprocessing steps were taken:

- Dropping irrelevant or null-value-heavy columns (e.g., "What matters most to you in choosing a course").
- Eliminating unique identifiers like Prospect IDs.
- Creating dummy variables for categorical data to enhance model compatibility.
- Reducing the dataset size from 9240 rows and 37 columns to 6373 rows and 75 columns after cleaning.

Model Building and Performance

Using Recursive Feature Elimination (RFE), five iterations of logistic regression were conducted to refine the model. Variables with high p-values or multicollinearity (as indicated by high VIF values) were systematically removed, yielding the final model. Notable removals included "Last Notable Activity_Had a Phone Conversation" and "Current Occupation_Housewife."

Final Model Results:

- Train Data: Accuracy 78.8%, Sensitivity 73.9%, Specificity 83.4%
- Test Data: Accuracy 78.4%, Sensitivity 77.9%, Specificity 78.9%

At a 0.44 probability cutoff, precision and recall were optimally balanced, ensuring robust predictions for potential conversions.

Recommendations

Based on the findings, the study suggests:

1. Prioritizing leads with higher visit counts and time spent on the site.
2. Monitoring lead origins, particularly focusing on high-conversion sources like Google and Welingak site submissions.
3. Enhancing landing page designs to encourage submissions.
4. Targeting marketing and HR management professionals for higher conversion rates.
5. Addressing the unemployed segment effectively while noting that working professionals yielded the most conversions.

By leveraging data-driven insights and refining lead-scoring models, the organization can significantly boost conversion rates. The study underscores the importance of tracking user behavior, optimizing touchpoints, and tailoring marketing efforts to align with high-value lead characteristics.