

Lead Scoring Case Study Summary

The Lead Scoring Case Study for X Education aimed to improve the company's lead conversion rate from 30% to 80% by identifying high-priority leads ("Hot Leads") through machine learning techniques. The objective was to optimize the sales team's efforts by focusing on leads most likely to convert.

Data Understanding

The dataset included features such as Lead Source, Specialization, and Total Time Spent on Website, with "Converted" as the target variable indicating whether a lead became a paying customer. Understanding the data was crucial to identifying the variables that could impact lead conversion.

Exploratory Data Analysis (EDA)

Exploratory analysis revealed several missing values in columns like Lead Source and Specialization. Categorical features such as "Select" and "Not Specified" were treated as null values. Relationships between variables and conversions were explored through visualizations, revealing patterns that influenced the modeling process.

Data Preprocessing & Feature Engineering

We performed One-Hot Encoding for categorical features such as Lead Source and Specialization, converting them into binary variables. Numerical features like Total Time Spent on Website were scaled using Min-Max normalization. Features that contributed little to no variability, like Newspaper and Search, were dropped from the model. We selected relevant features using Recursive Feature Elimination (RFE) and removed features that showed multicollinearity.

Model Building

Logistic regression, suitable for binary classification, was used to build the predictive model. We employed RFE for feature selection and built the model on selected features, including Lead Source, Total Time Spent on Website, and Specialization.

Model Evaluation

The model was evaluated using metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. The ROC curve was plotted, and the AUC score was calculated to assess model performance. The final model achieved an accuracy of 81.71%, a precision of 74.37%, and an F1-Score of 78.44% at an optimal cutoff point of 0.50.

Lead Scoring

Lead scores were assigned based on conversion probabilities from 0 to 100. Leads were categorized into three groups: 0-50 (Low probability), 50-79 (Moderate probability), and 80-100 (High probability). This scoring system enables the sales team to prioritize leads more likely to convert, improving efficiency and conversion rates.

Insights & Recommendations

Key insights:

1. **Lead Source:** Google and direct traffic resulted in higher conversions.

2. **Total Time Spent on Website:** More time spent on the website strongly correlates with higher conversion rates.
3. **Specializations:** Marketing Management and Finance Management leads have better conversion rates.

Recommendations:

1. Prioritize leads with scores between 80-100 for faster follow-ups.
2. Focus marketing resources on Google and direct traffic sources.
3. Tailor offerings to high-conversion specializations like Marketing and Finance Management.

Conclusion

This assignment provided a comprehensive understanding of building and evaluating a lead scoring model, from data preprocessing to deriving actionable insights. The model's insights will help X Education prioritize leads, improve conversions, and ultimately boost revenue.