

Summary report of the lead scoring project

Objective

X Education aimed to improve its lead conversion rate by identifying high-potential "hot leads" through a predictive model. The objective was to assign lead scores to prioritize sales efforts effectively.

Data:-

The dataset contained 9,240 rows with 37 attributes, including the target variable Converted. Key preprocessing steps included:

Missing Data Handling: Columns with excessive missing data (e.g., Lead Quality) were dropped, and other missing values were imputed using the mode (categorical) or median (numerical).

Addressing 'Select' Levels: Categorical variables with "Select" values were treated as null and imputed.

Outliers: Features like TotalVisits and Page Views Per Visit were capped at the 95th percentile to ensure data consistency.

Features were selected based on their correlation with the target variable and logistic regression coefficients:

Tags_Lost to EINS, Tags_Closed by Horizon, and Tags_Will revert after reading the email were the top predictors, reflecting significant behavioral insights.

Total Time Spent on Website had a strong positive correlation, indicating higher engagement leads to conversion.

Features were selected based on their correlation with the target variable and logistic regression coefficients:

Tags_Lost to EINS, Tags_Closed by Horizon, and Tags_Will revert after reading the email were the top predictors, reflecting significant behavioral insights.

Total Time Spent on Website had a strong positive correlation, indicating higher engagement leads to conversion.

Model Evaluation:-

The model demonstrated strong performance:

Accuracy: 90.8%

Precision: 89.2%

Recall: 86.7%

F1 Score: 87.9%

ROC-AUC: 96.4%

The high ROC-AUC score indicated excellent discriminatory power, while feature importance confirmed the relevance of behavioral tags and engagement metrics.

Business Applications:-

Aggressive Lead Conversion: During high-resource periods, lowering the probability threshold (e.g., to 0.3) enables more leads to be contacted, maximizing coverage.

Reducing Useless Calls: In low-activity periods, raising the threshold (e.g., to 0.8) focuses efforts on high-confidence leads, saving resources.

Actionable Insights: Tags like "Lost to EINS" and "Will revert after reading the email" enable targeted marketing and sales strategies.

Learnings:-

Data Handling Matters: Addressing missing values and cleaning categorical data improved model performance.

Behavioral Features Are Key: Tags and time spent on the website were critical for predicting conversion.

Threshold Adjustment Is Crucial: Dynamic probability thresholds aligned sales strategies with business goals.

Model Insights Drive Action: Logistic regression provided interpretable results, aiding decision-making.

This project showcased the power of predictive modeling to enhance lead management, boost conversion rates, and optimize resource allocation—all essential for X Education's growth.