LEAD SCORE PREDICTION MODEL

—

By,
Apurv Dhingra,
Anwin Aby George &
Anupama Shrivastav

SUMMARY REPORT

We embarked on this assignment with the primary goal of helping X Education reach its target of enhancing its lead conversion rate from the existing 30% to approximately 80%. The objective was to identify 'Hot Leads' with a higher likelihood of conversion by developing a lead scoring model, ultimately aiming to achieve a conversion rate of around 80%.

Our journey began with data importation. We imported the essential data analysis libraries and loaded the dataset, laying the foundation for subsequent analysis. Recognizing the importance of data quality, we performed data cleaning, eliminating columns with more than 25% missing values, addressing null values in specific columns, and removing unnecessary columns with little information.

Univariate analysis provided insights into individual variable distributions, setting the stage for our subsequent analyses. Bivariate analysis allowed us to investigate the relationships between variables and lead conversion, helping us identify key factors influencing conversion.

Preparing the data for modeling was pivotal. Categorical variables were effectively handled through dummy encoding, making them compatible with machine learning models. We divided the data into training and testing sets, ensuring model robustness, and standardized numerical features to eliminate biases due to varying variable magnitudes.

Model building was a significant phase. We constructed a logistic regression model using all independent variables, well-suited for binary classification tasks like lead conversion. Feature selection using Recursive Feature Elimination (RFE) improved the model's performance, helping us identify the most impactful features. Furthermore, we removed irrelevant features based on p-values and VIF values.

Model evaluation was critical. We assessed the model's performance with a random cutoff point of 0.5, calculating various metrics, including accuracy, sensitivity, specificity, and ROC curves. We determined an optimal cutoff point of 0.4 to strike a balance between sensitivity and specificity. On the test data, the model exhibited an accuracy of 80.5%, a sensitivity of 77.53%, and a specificity of 82.31%.

In summary, the lead scoring model developed in this assignment equips X Education with a strategic advantage. It enables the prioritization of leads with a higher likelihood of conversion, enhancing lead management effectiveness. This aligns seamlessly with X Education's overarching goal of achieving a CEO's target conversion rate of approximately 80%. The assignment's comprehensive focus on data analysis, modeling, and fine-tuning empowers X Education to make data-driven decisions, propelling lead conversion rate enhancement.