

Summary

X Education's lead conversion rate is low at 30%. The company needs a model to assign a lead score to each lead, aiming for an 80% conversion rate set by the CEO.

Data Cleaning:

- Dropped columns with more than 70% null values.
- Reviewed categorical columns: dropped if imputation caused skew, created new categories (e.g., "others"), imputed high-frequency values, and eliminated non-contributory columns.
- Imputed numerical categorical data with the mode and removed columns with only one unique response.
- Treated outliers, corrected invalid data, grouped low-frequency values, and mapped binary categorical values.

EDA:

- Found only 38.5% of leads converted.
- Univariate analysis showed key variables like 'Lead Origin', 'Current Occupation', and 'Lead Source' significantly impacted conversion.
- Noted positive impact of time spent on the website on lead conversion.

Data Preparation:

- Created dummy features using one-hot encoding for categorical variables.
- Split data into training and testing sets (70:30 ratio).
- Applied feature scaling using standardization.
- Dropped several highly correlated columns.

Model Building:

- Reduced variables to 20 using RFE.
- Dropped variables with p-value > 0.05 .
- Built 10 models, finalizing on Model 10, which was stable with p-values < 0.05 and no multicollinearity ($VIF < 5$).
- Selected log10 as the final model with 12 variables for predictions.

Model Evaluation:

- Created a confusion matrix and chose a cutoff point of 0.345 based on accuracy, sensitivity, and specificity. This cutoff yielded around 80% for accuracy, specificity, and precision, with precision-recall metrics around 75-80%.
- To meet the 80% conversion target, used the sensitivity-specificity view for the optimal cutoff.
- Assigned lead scores to the training data using 0.345 as the cutoff.

Making Predictions on Test Data:

- Scaled and predicted using the final model.
- Achieved evaluation metrics close to 80% for both train and test sets.
- Top 3 features:
 - Lead Source: Welingak Website
 - Lead Source: Reference
 - Current Occupation: Working Professional

Recommendations:

- Prioritize calling leads from "Welingak Websites" and "Reference" sources.
- Prioritize calling leads who are "working professionals".
- Prioritize calling leads who spent "more time on the websites".
- Prioritize calling leads from the lead source "Olark Chat".
- Prioritize calling leads whose last activity was "SMS Sent".
- Avoid calling leads whose last activity was "Olark Chat Conversation".
- Avoid calling leads from "Landing Page Submission".
- Avoid calling leads whose specialization is "Others".
- Avoid calling leads who selected "Do not Email" as "yes".

By following these recommendations, X Education can focus on leads with the highest potential for conversion, aiming to achieve the CEO's target conversion rate of 80%.