

Summary Report

X Education faces a challenge with a low lead conversion rate of approximately 30%. The CEO has set a target of 80%, and to achieve this, a model was developed to assign a lead score, prioritizing leads with a higher likelihood of conversion.

Data Cleaning:

Columns with over 40% missing values were dropped. Categorical columns were analyzed to decide on actions such as imputation, creating new categories, or dropping columns that added no value. Numerical data was imputed with the mode, and columns with only one unique response were removed. Outliers, invalid data, and low-frequency values were addressed.

Exploratory Data Analysis (EDA):

The data revealed a 38.5% lead conversion rate. Univariate and bivariate analyses were conducted on key variables like 'Lead Origin' and 'Current Occupation,' revealing valuable insights. Time spent on the website positively impacted lead conversion.

Data Preparation:

Categorical variables were one-hot encoded, and the data was split into a 70:30 train-test ratio. Features were scaled, and highly correlated columns were dropped.

Model Building:

RFE was used to reduce variables from 48 to 15. A manual feature reduction process followed, resulting in a final model with 12 variables, showing no multicollinearity. The model was validated and used for predictions.

Model Evaluation:

A confusion matrix helped determine an optimal cut-off point of 0.345, providing accuracy, specificity, and precision around 80%. However, precision-recall metrics were slightly lower at 75%. The sensitivity-specificity view was chosen to align with the CEO's goal. The lead score was then assigned to the data.

Predictions on Test Data:

The model was applied to the test set, showing metrics close to 80%. The top predictive features were 'Lead Source_Welingak Website,' 'Lead Source_Reference,' and 'Current Occupation_Working Professional.'

Recommendations:

The company should increase the budget for advertising on the Welingak Website, offer incentives for references that convert, and aggressively target working professionals due to their high conversion rate and financial capability.