

Summary

Data Cleaning:

- ❖ Dropped columns with more than 40% null values.
- ❖ Imputed null values in categorical columns with the mode value.
- ❖ Imputed null values in numerical columns with the median.
- ❖ Checked for unique values and data distribution within each column.

EDA (Exploratory Data Analysis):

- ❖ Examined correlation between numeric columns using Heatmap and Pair plot.
- ❖ Conducted categorical analysis using count plot.

Data Preparation:

- ❖ Created dummy variables for categorical features.
- ❖ Split the data into Train and Test sets with a 70:30 ratio.
- ❖ Applied feature scaling using Standardization.

Model Building:

- ❖ Utilized Recursive Feature Elimination (RFE) to reduce variables from 94 to 20.
- ❖ Employed manual feature reduction by dropping variables with p-values > 0.05 .
- ❖ Built a total of 7 models before finalizing Model 8, which demonstrated stability with p-values < 0.05 and no multicollinearity ($VIF < 5$).
- ❖ Model 8, comprising 11 variables, was selected as the final model for prediction on both train and test sets.

Model Evaluation:

- ❖ Constructed a confusion matrix and determined a cutoff point of 0.42 based on accuracy, sensitivity, specificity, and precision-recall plots, yielding accuracy, specificity, and precision all above 85%.
- ❖ Achieved an accuracy of 89.6%, sensitivity of 85%, specificity of 92%, precision of 86%, and recall of 86%.

Predictions on Test Data:

- ❖ Applied scaling and made predictions using the final model on the test data.
- ❖ Evaluation metrics for both train and test sets were approximately 89.6%.
- ❖ Assigned lead scores based on predictions.
- ❖ Identified the top 3 features as Total Time Spent on Website, Tags Will revert after reading the email, and What is your current occupation Working Professional.

Recommendations:

- ❖ Prioritize making calls to leads associated with Tags Closed by Horizon, Tags Lost to EINS, Last Notable Activity Had a Phone Conversation, Lead Origin Lead Add Form, Last Notable Activity Email Bounced, and Last Notable Activity SMS Sent, as these have higher conversion probabilities.
- ❖ Target leads who are working professionals, as they exhibit higher conversion likelihood.
- ❖ Focus on leads from Lead Source_Welingak Website and those who spend more time on the website, as these also show higher conversion potential.