Lead Scoring Case Study Summary

Prasanna R

- The given lead data consisted of 9240 rows × 37 columns
- Checked for null values in all columns
- Dropped columns having null values >40%
- Replaced 'Select' values in all the columns with 'Null'
- Imputed other columns with mode for categorical and median for continuous columns
- Performed EDA, and later dropped columns with least significance
- Mapped 'yes'/'no' with 1/0
- Created dummy variables for categorical columns
- Train-test split was performed
- Scaling was performed for numerical data columns
- Feature selection was performed using RFE method
- Built the model using statsmodel library and GLM method
- Removed few columns based on P values and VIF values
- Checked for accuracy, specificity, sensitivity
- ROC curve was plotting to check area under the curve
- Optimal cut-off point was found to make the probability above that point as 1 and make that user as potential candidate
- Test Data predictions:
- Accuracy : 80%
- Sensitivity : 81%
- Specificity: 80%