Lead Score Case Study Summary

Steps taken:

1. Loading and Analysing the Data

2. Data Manipulation and Cleaning

Dropped the columns which had more than 40% of null values. Imputing null values with mode or 'Others' where applicable. Outliers were identified and treated.

3. EDA

4. Dummy Variable Creation

Dummy variables were created for categorical columns.

5. Train-Test Split

Data was split into train set and test set by 70% and 30% respectively.

6. Scaling the Data

StandardScaler was used to scale numerical variables.

7. Model Building

We started with RFE while selecting top 20 features. Later features with high p-values and high VIF values were removed.

8. Model Evaluation

Model was evaluated using confusion matrix. And after finding optimal cut-off point (using ROC curve) accuracy was 92%, sensitivity was 91.69% and specificity was 92.22%.

9. Precision and Recall

Precision was 87.86% and Recall was 91.69%.

10. Predictions on Test Set

According to the model built, test data set gave accuracy of 92.10%, sensitivity of 90.89% and specificity of 92.84%.