**Binary Classification – Assignment Summary**

**Step 1: Inspect the Data**

Importing the data, basic inspection of data, checking for various data types, statistical summary, invalid values etc was done and it was discovered that all columns of the data had numeric values. A closer look on the statistical summary inferred there were a few outliers and none of the columns except the target variable could be treated as categorical values.

**Step 2 : EDA**

Since there was no necessity to handle missing values and outliers. Check for multicollinearity was done. Based on the result a few features were dropped.

None of the features had a high correlation with the target variable.

The data has light class imbalance. Did not choose to add samples to handle imbalance.

**Step 3: Data Preparation**

The features and target variables were separated, Features were scaled to standardize.

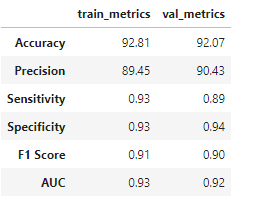
**Step 4: Model Building & Model Selection**

Started with a simple generalized linear model and noticed a few features were not significant.

Used recursive feature elimination method to select appropriate features, Variable inflation factor >2 was eliminated.

Appropriate threshold was selected based on comparison of various metrics for multiple arbitrary cutoffs.

The model gave very good results both on train and validation sets as shown below:



**Since the model gave good accuracy and was able to generalize well for both train and validation data sets. No other models were built to compare.**

**Step 5: Output**

Predictions made on the test data was output into a csv file ‘**output.csv’**

**Note: Observations have been commented in the notebook wherever necessary. All code for EDA, model building, model performance, generating output has been included in a single notebook.**