HSBC

The problem statement given is a Classification Problem.

1) Analysis:
   a) Calculating the number of rows and columns
   b) Describing the dataframe
   c) Checking the datatypes
   d) Checking the null values
   e) Checking the duplicate values
   f) Checking the value counts and performing label encoding
   g) Dropping the unique columns
   h) Grouping the dataset on the basis of “V6”
   i) Removing the outliers in “V4”
   j) Checking the class imbalance

2)Training the model
   a) Logistic Regression
   b) Oversampling technique(SMOTE: using knn)
   c) Decision tree classifier
   d) Gausssian NB
   e) AdaBoosterClaassifier
   f) GradientBoostingClassifier
   g) RandomForestClassifier
   h) SVC

3)Testing the model using logistic regression