CS777 – Assignment 5

Gaurav Tungare

------------------------------------------------------------------------------------------------------------------------------------------

**Task 1**

**-------Logistic Regression total Read time in sec : 140.73113536834717**

**------Logistic Regression total Read time in sec : 141.7999882698059**

**----------- Logistic Regression total Model testing and training time in sec : 1166.768543958664**

**tp= 375**

**fp= 1**

**fn= 2**

**tn= 18346**

**-----Logistic Regression F1 Score----**

**F1 : 0.9960159362549801**

**----------- Confusion matrix Logistic Regression calcuation time in sec : 161.73884677886963**

**---- Logistic Regression total time in sec : 1470.3074851036072**

Graphical user interface, text, application, Word, email

Description automatically generated

Graphical user interface, application, table, Word

Description automatically generated

Graphical user interface, application, table

Description automatically generated

**Task 2**

**-------SVN total Read time in sec : 143.17851042747498**

**----------- SVN total Model testing and training time in sec : 2099.4157168865204**

**tp= 0**

**fp= 0**

**fn= 377**

**tn= 18347**

**----------- Confusion matrix SVN calcuation time in sec : 196.13077855110168**

**-----SVN F1 Score----**

**F1 : 0.0**

**---- total time SVN in sec : 2438.7251422405243**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application, table

Description automatically generated

A picture containing timeline

Description automatically generated

**Task 3**

**Task3 - Logistic Regression**

**-------Logistic Regression total Read time in sec : 136.02258205413818**

**----------- Logistic Regression total Model testing and training time in sec : 1009.0990524291992**

**tp= 0**

**fp= 50**

**fn= 377**

**tn= 18297**

**-----Logistic Regression F1 Score----**

**F1 : 0.0**

**----------- Confusion matrix Logistic Regression calcuation time in sec : 687.1757960319519**

**---- Logistic Regression total time in sec : 1832.2975931167603**

Graphical user interface, text, application, email

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Graphical user interface, application, table

Description automatically generated

**Task3 - SVM**

**-------SVN total Read time in sec : 138.08575630187988**

**----------- SVN total Model testing and training time in sec : 976.6605532169342**

**tp= 0**

**fp= 0**

**fn= 377**

**tn= 18347**

**----------- Confusion matrix SVN calcuation time in sec : 290.30811977386475**

**-----SVN F1 Score----**

**F1 : 0.0**

**---- total time SVN in sec : 1405.054556131363**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application, table

Description automatically generated

Graphical user interface, table

Description automatically generated

**Discuss if your dimension reduction approach is applicable on very large data sets**

**So what I have observed is that the there is a drop in the F1 score after dimension reduction, the reason I believe is because of imbalance nature of the data set which is given**

**Ideally per the industry if we choose the dimension reduction/selection properly then we should get better result but in our example this is not the case**