TANUSHRI SINGH
TTS150030
CS 6375.001 - MACHINE LEARNING
ASSIGNMENT 2
INSTRUCTOR - ANJUM CHIDA

Sample Output:-
Arguments -> python logisticRegression.py .01 300 True
Results ->
********** -------- **********
The Ratio of Success For Spam Emails -> 83.0769%
The Ratio of Success For Ham Emails -> 97.4138%
The Ratio of Success For All Emails -> 93.5146%
********** -------- **********

Arguments -> python logisticRegression.py .01 55 True
Results ->
********** -------- **********
The Ratio of Success For Spam Emails -> 74.6154%
The Ratio of Success For Ham Emails -> 97.1264%
The Ratio of Success For All Emails -> 91.0042%
********** -------- **********

Arguments -> python logisticRegression.py .015 300 False
Results ->
********** -------- **********
The Ratio of Success For Spam Emails -> 84.6154%
The Ratio of Success For Ham Emails -> 94.5402%
The Ratio of Success For All Emails -> 91.8410%
********** -------- **********

Arguments -> python logisticRegression.py .015 55 False
Results ->
********** -------- **********
The Ratio of Success For Spam Emails -> 83.0769%
The Ratio of Success For Ham Emails -> 94.8276%
The Ratio of Success For All Emails -> 91.6318%
********** -------- **********

Arguments -> python logisticRegression.py .02 110 True
Results ->

********** _______ **********
The Ratio of Success For Spam Emails -> 80.0000%
The Ratio of Success For Ham Emails -> 96.5517%
The Ratio of Success For All Emails -> 92.0502%
********** _______ **********

Arguments -> python logisticRegression.py .02 110 False
Results ->
********** _______ **********
The Ratio of Success For Spam Emails -> 81.5385%
The Ratio of Success For Ham Emails -> 94.5402%
The Ratio of Success For All Emails -> 91.0042%
********** _______ **********

Arguments -> python naiveBayes.py True
********** _______ **********
The Ratio of Success For Spam Emails -> 98.4615%
The Ratio of Success For Ham Emails -> 94.5402%
The Ratio of Success For All Emails -> 95.6067%
********** _______ **********

Arguments -> python naiveBayes.py False
********** _______ **********
The Ratio of Success For Spam Emails -> 98.4615%
The Ratio of Success For Ham Emails -> 94.8276%
The Ratio of Success For All Emails -> 95.8159%
********** _______ **********

Explanation -> It is clear that using the StopWords file benefits in the success ratio when it comes to Logistic Regression. However, when these stop words are used with regards to Naive Bayes, it does not make too vast of a difference. In fact it slightly reduces the overall success for all emails, even though it results in slightly better success rates for the Ham emails. Learning rate is hardcoded to .001 in Logistic Regression and the lambda value, iterations and whether or not stop words are used can be controlled by input user. These input values affect how the weights are calculated in Logistic Regression. Another observation is how higher iterations results in higher total success rates but it does take a long time to run. Both Logistic Regression and Naive Bayes first train on a set of data then it tests against the other set. Overall there does not seem to be traces of overfitting in either of the two approaches.