ASSIGNMENT 2 REPORT
(NAIVE BAYES AND LOGISTIC REGRESSION)
by
PRAVEEN RAMANI(pxr170005)

NAIVE BAYES (*without stopwords elimination)*
spam accuracy without stopwords elimination: 97.6923076923077
Ham accuracy without stopwords elimination: 93.67816091954023
Net Accuracy: 94.76987447698745

NAIVE BAYES (*with stopwords elimination)*
spam accuracy with stopwords elimination: 97.6923076923077
Ham accuracy with stopwords elimination: 94.25287356321839
Net Accuracy: 95.18828451882845

LOGISTIC REGRESSION *(without stopwords elimination)*
iterations = 100
regularization_parameter = 1
learning_rate = 0.01
Spam Accuracy without stop words elimination: 84.21052631578947
Ham Accuracy without stop words elimination: 94.78260869565217
Net Accuracy: 91.84100418410041

LOGISTIC REGRESSION *(with stopwords elimination)*
iterations = 100
regularization_parameter = 1
learning_rate = 0.01
Spam Accuracy with stop words elimination: 86.82170542635659
Ham Accuracy with stop words elimination: 94.84240687679083
Net Accuracy: 92.67782426778243

LOGISTIC REGRESSION *(without stopwords elimination)*
iterations = 100
regularization_parameter = 0.1
learning_rate = 0.01
Spam Accuracy without stop words elimination: 83.84615384615385
Ham Accuracy without stop words elimination: 93.96551724137932
Net Accuracy: 91.21338912133892

LOGISTIC REGRESSION *(with stopwords elimination)*
iterations = 100
regularization_parameter = 0.1
learning_rate = 0.01
Spam Accuracy with stop words elimination: 88.09523809523809
Ham Accuracy with stop words elimination: 94.60227272727273
Net Accuracy: 92.88702928870293

There is an increase in accuracy when stop words are removed. This might be because we eliminate common words, which occur frequently in all documents and so affects the classification just by occurring more frequent or less frequent in certain documents. By eliminating these words the algorithm focuses on less common words, thereby increasing the accuracy.