**Read Me**

**To compile the attached java files:**

javac filename.java

**To run the file:**

java filename.java <path of training file> <path of test file>

**For Naïve Bayes:**

java NaiveBayes C:/eclipse/ML Assign 2/src/training.txt C:/eclipse/ML Assign 2/src/test.txt

**For Logistic Regression:**

java LogisticRegression C:/eclipse/ML Assign 2/src/training.txt C:/eclipse/ML Assign 2/src/test.txt

**Accuracy**

**The accuracy for Naïve Bayes was found to be:**

91% for Ham and 80% for Spam - with stop words

89.4% for Ham and 78% for Spam - without stop words

**The accuracy for Logistic Regression was found to be:**

88% for Ham and 79% for Spam - with stop words

86.8% for Ham and 78.6% for Spam - without stop words

Accuracy generally increases when stop words are removed (i.e., not considered). This is because stop words can occur equally in spam and ham emails. They cannot be used as distinguishing words for determining whether an email is spam or ham. This is why the accuracy improves both in Logistic Regression and Naïve Bayes when stop words are removed. But, while implementing this program it was found that accuracy decreased when stop words are removed.