Assignment 2

The program implements Naive Bayes classification as well as Logistic regression classification with L2 Regularization to classify messages as either Spam or Ham.

Program Name : assignment2.java

The Program Contains two parts:
1. Naive Bayes classification
2. Logistic Regression with default lambda value = 0.001 and number of iterations = 500

No command line arguments required

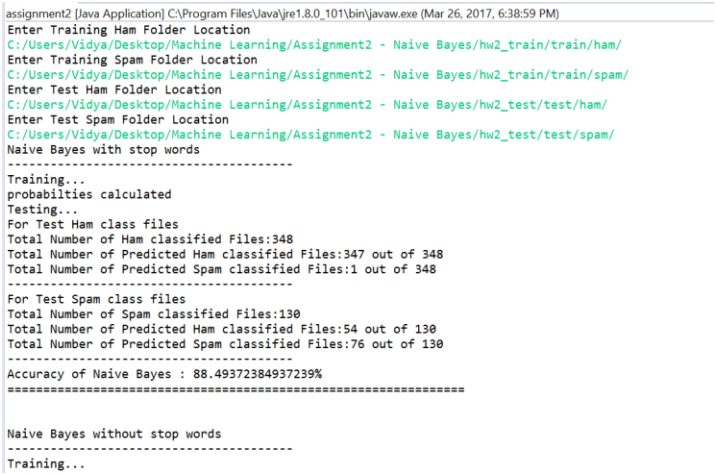
Instructions to execute the program

- compile assignment2.java

The program will prompt you to enter folder locations for training and test folders, for example

```
Enter Training Ham Folder Location
C:/Users/Vidya/Desktop/Machine Learning/Assignment2 - Naive Bayes/hw2_train/train/ham/
```

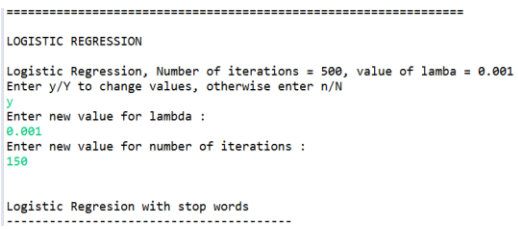
The user will be prompted to enter all folder locations. The Naïve Bayes classifier with the use of stop words and without the use of stop words will execute following this.



```
assignment2 [Java Application] C:\Program Files\Java\jre1.8.0_101\bin\javaw.exe (Mar 26, 2017, 6:38:59 PM)
Enter Training Ham Folder Location
C:/Users/Vidya/Desktop/Machine Learning/Assignment2 - Naive Bayes/hw2_train/train/ham/
Enter Training Spam Folder Location
C:/Users/Vidya/Desktop/Machine Learning/Assignment2 - Naive Bayes/hw2_train/train/spam/
Enter Test Ham Folder Location
C:/Users/Vidya/Desktop/Machine Learning/Assignment2 - Naive Bayes/hw2_test/test/ham/
Enter Test Spam Folder Location
C:/Users/Vidya/Desktop/Machine Learning/Assignment2 - Naive Bayes/hw2_test/test/spam/
Naive Bayes with stop words
----------------------------------------
Training...
probabilties calculated
Testing...
For Test Ham class files
Total Number of Ham classified Files:348
Total Number of Predicted Ham classified Files:347 out of 348
Total Number of Predicted Spam classified Files:1 out of 348
----------------------------------------
For Test Spam class files
Total Number of Spam classified Files:130
Total Number of Predicted Ham classified Files:54 out of 130
Total Number of Predicted Spam classified Files:76 out of 130
----------------------------------------
Accuracy of Naive Bayes : 88.4937238493723%
============================================================


Naive Bayes without stop words
----------------------------------------
Training...
```

For the Second part of the program, default lambda value is 0.001 and default number of iterations is 500. The user is then asked, if these values needs to be changed.



```
============================================================

LOGISTIC REGRESSION

Logistic Regression, Number of iterations = 500, value of lamba = 0.001
Enter y/Y to change values, otherwise enter n/N
y
Enter new value for lambda :
0.001
Enter new value for number of iterations :
150


Logistic Regresion with stop words
----------------------------------------
```

For the specified lambda and number of iterations value, the logistic regression executes with the use of stop words and then without the use of stop words.