

The University of Texas at Dallas
Machine Learning – Prof. Anjum Chida
Assignment #02

Explanation:

By filtering out stop words the classification becomes more accurate as classification on unwanted/low quality features is avoided.

In most cases, these low quality features give negative results and at the same time may not be that negative in some cases where it doesn't affect classification just coincidentally.

Hence, classification is more accurately done in most of the cases when stop words are filtered out/ removed.

Naive Bayes		Logistic Regression {a = 0.75, b = 0.75, N = 100}	
Filter stopwords = No	Filter stopwords = Yes	Filter stopwords = No	Filter stopwords = Yes
95.46	92.22	79.07	85.56

a :- eta/ learning rate, b :- lambda, and N :- No of iterations

```

rahul:Assignment 02$ javac NaiveBayes.java
rahul:Assignment 02$ java NaiveBayes train test no
Naive Bayes Accuracy: 95.46436285097192
rahul:Assignment 02$ java NaiveBayes train test yes
After filtering Stop Words
Naive Bayes Accuracy: 92.22462203023758
rahul:Assignment 02$
rahul:Assignment 02$
rahul:Assignment 02$ javac LogisticRegression.java
rahul:Assignment 02$ java LogisticRegression train test no 0.75 0.75 100
Accuracy on Test data: 79.07949790794979
rahul:Assignment 02$ java LogisticRegression train test yes 0.75 0.75 100
After removing Stop Words
Accuracy on Test data: 85.56485355648536
rahul:Assignment 02$

```

For logistic regression, and upon filtering/ removing the stopwords, and N = 100:

a	b	Accuracy
0.5	0.5	82.63598326359832
0.9	0.5	83.89121338912135
0.5	0.9	85.98326359832636
1	0.75	86.40167364016736
0.75	1	82.00836820083683
0.9	0.9	87.44769874476988
0.5	0.9	84.93723849372385
0.05	0.9	80.9623430962343

```

rahul:Assignment 02$ javac LogisticRegression.java
rahul:Assignment 02$ java LogisticRegression train test yes 0.5 0.5 100
After removing Stop Words
Accuracy on Test data: 82.63598326359832
rahul:Assignment 02$ java LogisticRegression train test yes 0.9 0.5 100
After removing Stop Words
Accuracy on Test data: 83.89121338912135
rahul:Assignment 02$ java LogisticRegression train test yes 0.5 0.9 100
After removing Stop Words
Accuracy on Test data: 85.98326359832636
rahul:Assignment 02$ java LogisticRegression train test yes 1 0.75 100
After removing Stop Words
Accuracy on Test data: 86.40167364016736
rahul:Assignment 02$ java LogisticRegression train test yes 0.75 1 100
After removing Stop Words
Accuracy on Test data: 82.00836820083683
rahul:Assignment 02$ java LogisticRegression train test yes 0.9 0.9 100
After removing Stop Words
Accuracy on Test data: 87.44769874476988
rahul:Assignment 02$ java LogisticRegression train test yes 0.5 0.9 100
After removing Stop Words
Accuracy on Test data: 84.93723849372385
rahul:Assignment 02$ java LogisticRegression train test yes 0.05 0.9 100
After removing Stop Words
Accuracy on Test data: 80.9623430962343
rahul:Assignment 02$

```