## To run the code

1. Naïve Bayes
    - Normal run- python NaiveBayes.py ../data/imdb1
    - 'Filter Stop Words' run- python NaiveBayes.py -f ../data/imdb1
    - 'Boolean Naïve Bayes' run- python NaiveBayes.py -b ../data/imdb1
2. Perceptron
    - Python Perceptron.py ../data/imdb1 <ITERATION_COUNT>

## Summary of Results

Following is the summary of results obtained. Detailed results are included in last section

| Algorithm | Average Accuracy |
|---|---|
| Naïve Bayes- normal run | 0.816500 |
| Naïve Bayes, filter stop words | 0.811000 |
| Boolean Naïve Bayes | 0.814500 |
| Perceptron, iterations=1 | 0. 713000 |
| Perceptron, iterations=1, filter stop words | 0.704500 |
| Perceptron, iterations=10 | 0.822500 |
| Perceptron, iterations=10, filter stop words | 0.806000 |
| Perceptron, iterations=50 | 0.831000 |
| Perceptron, iterations=50, filter stop words | 0.818000 |
| Perceptron, iterations=100 | 0.838000 |
| Perceptron, iterations=100, filter stop words | 0.832500 |
| Perceptron, iterations=1000 | 0.825500 |
| Perceptron, iterations=1000, filter stop words | 0.824000 |

## Analysis of Results

- Naïve Bayes, despite having the strong assumption of independence of words, provides comparable results to perceptron
- Filtering of stop words doesn't affect the results much in case of Naïve Bayes
- In perceptron algorithm, increase in no. of iterations significantly improves the accuracy initially but saturates later
- Run time tremendously with increase in no. of iterations (>100)

## **Limitations**-

- Perceptron approach takes long time to complete as we increase the number of iterations for very little increase in accuracy. This can be fixed in 2 ways-

  i) We can limit the number of iterations, as the increase in accuracy is very low. For eg, increasing iterations from 100 to 1000 increases accuracy from just 82 to 83%

  ii) We can reduce the number of perceptron units by lowering the number of features.

## **Detailed Results-**

Naïve Bayes – normal run

[INFO] Fold 0 Accuracy: 0.765000
[INFO] Fold 1 Accuracy: 0.850000
[INFO] Fold 2 Accuracy: 0.835000
[INFO] Fold 3 Accuracy: 0.825000
[INFO] Fold 4 Accuracy: 0.815000
[INFO] Fold 5 Accuracy: 0.820000
[INFO] Fold 6 Accuracy: 0.835000
[INFO] Fold 7 Accuracy: 0.825000
[INFO] Fold 8 Accuracy: 0.755000
[INFO] Fold 9 Accuracy: 0.840000
[INFO] Accuracy: 0.816500

Naïve Bayes, FILTER STOP WORDS= TRUE

[INFO] Fold 0 Accuracy: 0.765000
[INFO] Fold 1 Accuracy: 0.825000
[INFO] Fold 2 Accuracy: 0.815000
[INFO] Fold 3 Accuracy: 0.830000
[INFO] Fold 4 Accuracy: 0.795000
[INFO] Fold 5 Accuracy: 0.830000
[INFO] Fold 6 Accuracy: 0.835000
[INFO] Fold 7 Accuracy: 0.835000
[INFO] Fold 8 Accuracy: 0.760000
[INFO] Fold 9 Accuracy: 0.820000
[INFO] Accuracy: 0.811000

Naïve Bayes, BOOLEAN_NB = TRUE

[INFO] Fold 0 Accuracy: 0.795000
[INFO] Fold 1 Accuracy: 0.820000
[INFO] Fold 2 Accuracy: 0.850000
[INFO] Fold 3 Accuracy: 0.810000
[INFO] Fold 4 Accuracy: 0.820000
[INFO] Fold 5 Accuracy: 0.805000
[INFO] Fold 6 Accuracy: 0.815000
[INFO] Fold 7 Accuracy: 0.835000
[INFO] Fold 8 Accuracy: 0.745000
[INFO] Fold 9 Accuracy: 0.850000
[INFO] Accuracy: 0.814500

Perceptron, No. of iterations= 1, Filter Stop Words= True

[INFO] Fold 0 Accuracy: 0.680000
[INFO] Fold 1 Accuracy: 0.765000
[INFO] Fold 2 Accuracy: 0.710000
[INFO] Fold 3 Accuracy: 0.705000
[INFO] Fold 4 Accuracy: 0.680000
[INFO] Fold 5 Accuracy: 0.705000
[INFO] Fold 6 Accuracy: 0.695000
[INFO] Fold 7 Accuracy: 0.655000
[INFO] Fold 8 Accuracy: 0.745000
[INFO] Fold 9 Accuracy: 0.705000
[INFO] Accuracy: 0.704500

Perceptron, No. of iterations= 1, Filter Stop Words= False

[INFO] Fold 0 Accuracy: 0.715000
[INFO] Fold 1 Accuracy: 0.715000
[INFO] Fold 2 Accuracy: 0.720000
[INFO] Fold 3 Accuracy: 0.735000
[INFO] Fold 4 Accuracy: 0.715000
[INFO] Fold 5 Accuracy: 0.720000
[INFO] Fold 6 Accuracy: 0.730000
[INFO] Fold 7 Accuracy: 0.640000
[INFO] Fold 8 Accuracy: 0.730000
[INFO] Fold 9 Accuracy: 0.710000
[INFO] Accuracy: 0.713000

Perceptron, No. of iterations= 10, Filter Stop Words= True

[INFO] Fold 0 Accuracy: 0.750000
[INFO] Fold 1 Accuracy: 0.825000
[INFO] Fold 2 Accuracy: 0.805000
[INFO] Fold 3 Accuracy: 0.820000
[INFO] Fold 4 Accuracy: 0.830000
[INFO] Fold 5 Accuracy: 0.780000
[INFO] Fold 6 Accuracy: 0.815000
[INFO] Fold 7 Accuracy: 0.800000
[INFO] Fold 8 Accuracy: 0.785000
[INFO] Fold 9 Accuracy: 0.850000
[INFO] Accuracy: 0.806000

Perceptron, No. of iterations= 10, Filter Stop Words= False

[INFO] Fold 0 Accuracy: 0.845000
[INFO] Fold 1 Accuracy: 0.820000
[INFO] Fold 2 Accuracy: 0.815000
[INFO] Fold 3 Accuracy: 0.820000
[INFO] Fold 4 Accuracy: 0.820000
[INFO] Fold 5 Accuracy: 0.805000
[INFO] Fold 6 Accuracy: 0.865000
[INFO] Fold 7 Accuracy: 0.770000
[INFO] Fold 8 Accuracy: 0.795000
[INFO] Fold 9 Accuracy: 0.870000
[INFO] Accuracy: 0.822500

Perceptron, No. of iterations= 50, Filter Stop Words= True

[INFO] Fold 0 Accuracy: 0.810000
[INFO] Fold 1 Accuracy: 0.845000
[INFO] Fold 2 Accuracy: 0.815000
[INFO] Fold 3 Accuracy: 0.830000
[INFO] Fold 4 Accuracy: 0.825000
[INFO] Fold 5 Accuracy: 0.815000
[INFO] Fold 6 Accuracy: 0.830000
[INFO] Fold 7 Accuracy: 0.800000
[INFO] Fold 8 Accuracy: 0.775000
[INFO] Fold 9 Accuracy: 0.835000
[INFO] Accuracy: 0.818000

Perceptron, No. of iterations= 50, Filter Stop Words= False

[INFO] Fold 0 Accuracy: 0.800000
[INFO] Fold 1 Accuracy: 0.845000
[INFO] Fold 2 Accuracy: 0.845000
[INFO] Fold 3 Accuracy: 0.855000
[INFO] Fold 4 Accuracy: 0.810000
[INFO] Fold 5 Accuracy: 0.845000
[INFO] Fold 6 Accuracy: 0.860000
[INFO] Fold 7 Accuracy: 0.790000
[INFO] Fold 8 Accuracy: 0.810000
[INFO] Fold 9 Accuracy: 0.850000
[INFO] Accuracy: 0.831000

Perceptron, No. of iterations= 100, Filter Stop Words= True

[INFO] Fold 0 Accuracy: 0.795000
[INFO] Fold 1 Accuracy: 0.850000
[INFO] Fold 2 Accuracy: 0.810000
[INFO] Fold 3 Accuracy: 0.850000
[INFO] Fold 4 Accuracy: 0.840000
[INFO] Fold 5 Accuracy: 0.805000
[INFO] Fold 6 Accuracy: 0.840000
[INFO] Fold 7 Accuracy: 0.840000
[INFO] Fold 8 Accuracy: 0.805000
[INFO] Fold 9 Accuracy: 0.890000
[INFO] Accuracy: 0.832500

Perceptron, No. of iterations= 100, Filter Stop Words= False

[INFO] Fold 0 Accuracy: 0.795000
[INFO] Fold 1 Accuracy: 0.850000
[INFO] Fold 2 Accuracy: 0.845000
[INFO] Fold 3 Accuracy: 0.845000
[INFO] Fold 4 Accuracy: 0.835000
[INFO] Fold 5 Accuracy: 0.855000
[INFO] Fold 6 Accuracy: 0.870000
[INFO] Fold 7 Accuracy: 0.815000
[INFO] Fold 8 Accuracy: 0.820000
[INFO] Fold 9 Accuracy: 0.850000
[INFO] Accuracy: 0.838000

Perceptron, No. of iterations= 1000, Filter Stop Words= True

[INFO] Fold 0 Accuracy: 0.795000
[INFO] Fold 1 Accuracy: 0.830000
[INFO] Fold 2 Accuracy: 0.820000
[INFO] Fold 3 Accuracy: 0.855000
[INFO] Fold 4 Accuracy: 0.860000
[INFO] Fold 5 Accuracy: 0.840000
[INFO] Fold 6 Accuracy: 0.815000
[INFO] Fold 7 Accuracy: 0.800000
[INFO] Fold 8 Accuracy: 0.765000
[INFO] Fold 9 Accuracy: 0.875000
[INFO] Accuracy: 0.825500

Perceptron, No. of iterations= 1000, Filter Stop Words= False

[INFO] Fold 0 Accuracy: 0.805000
[INFO] Fold 1 Accuracy: 0.835000
[INFO] Fold 2 Accuracy: 0.825000
[INFO] Fold 3 Accuracy: 0.805000
[INFO] Fold 4 Accuracy: 0.845000
[INFO] Fold 5 Accuracy: 0.835000
[INFO] Fold 6 Accuracy: 0.795000
[INFO] Fold 7 Accuracy: 0.800000
[INFO] Fold 8 Accuracy: 0.815000
[INFO] Fold 9 Accuracy: 0.880000
[INFO] Accuracy: 0.824000