

ML ASSIGNMENT 3

YXD200003

PART 1: NAIVE BAYES ALGORITHM

STOP WORDS are the commonly used words in any language. When these words are removed from sentences, we are left with unique words which make the process of spam classification easier. Using an ideal stopwords filter, the accuracy of our classifier should either be at the same level or should increase by a substantial percentage. Accuracies may vary by using different stopword removal methods.

With Naïve Bayes Algorithm there is a slight increase in accuracy after removing the stop words.

RESULT:

Accuracy of Naive Bayes before removing stop words : 0.9476987447698745

Accuracy of Naive Bayes after removing stop words : 0.9518828451882845

```
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>python naive.py
Accuracy of Naive Bayes before removing stop words 0.9476987447698745
Accuracy of Naive Bayes after removing stop words 0.9518828451882845
(base) C:\Users\dell\Desktop\UTD\SEM_1\Machine Learning\assignment\Assignment III>
```

PART 2: LOGISTIC REGRESSION ALGORITHM

Learning Rate = 0.001

Lamda values = [0.2,0.1,0.05,0.01]

Iterations = 500, 1000

Lamda	Iterations	Accuracy with Stop Words	Accuracy without Stop Words
0.2	500	93.09623430962343	94.05146443514644
0.1	500	93.09623430962343	94.05146443514644
0.05	500	93.09623430962343	94.05146443514644
0.01	500	93.09623430962343	94.05146443514644
0.2	1000	93.46861924686192	94.14225941422595
0.1	1000	93.46861924686192	94.14225941422595
0.05	1000	93.46861924686192	94.14225941422595
0.01	1000	93.46861924686192	94.14225941422595

For this case, after removing stopwords there is a slight increase in accuracy when Logistic Regression Algorithm is used. For various lamda values accuracy was constant at 500 iterations and 1000 iterations. As iterations increases there is a slight increase in accuracy when stop words are removed

```
Anaconda Prompt (anaconda3)
Iteration no. 970
Iteration no. 971
Iteration no. 972
Iteration no. 973
Iteration no. 974
Iteration no. 975
Iteration no. 976
Iteration no. 977
Iteration no. 978
Iteration no. 979
Iteration no. 980
Iteration no. 981
Iteration no. 982
Iteration no. 983
Iteration no. 984
Iteration no. 985
Iteration no. 986
Iteration no. 987
Iteration no. 988
Iteration no. 989
Iteration no. 990
Iteration no. 991
Iteration no. 992
Iteration no. 993
Iteration no. 994
Iteration no. 995
Iteration no. 996
Iteration no. 997
Iteration no. 998
Iteration no. 999
LR Accuracy removing Stop Words : 94.14225941422595
lambda value is: 0.1
Iteration no. 0
Iteration no. 1
Iteration no. 2
Iteration no. 3
Iteration no. 4
Iteration no. 5
Iteration no. 6
Iteration no. 7
Iteration no. 8
Iteration no. 9
Iteration no. 10
Iteration no. 11
Iteration no. 12
Iteration no. 13
Iteration no. 14
Iteration no. 15
Iteration no. 16
Iteration no. 17
```