Machine Learning Assignment 2 Report

NAIVE BAYES : WITHOUT STOPWORDS

Accuracy on ham files : 96.26436781609196

Accuracy on spam Files : 90.76923076923077

Total Accuracy : 94.76987447698745

NAIVE BAYES: WITH STOPWORDS

Accuracy on ham files : 96.55172413793103

Accuracy on spam files : 88.46153846153845

Total Accuracy : 94.35146443514645

MCAP LOGISTIC REGRESSION:WITHOUT STOPWORDS

Regularization Parameter : 2

Number of Iterations : 100

Learning rate : 0.01

Accuracy on ham files : 95.33527696793003

Accuracy on spam Files : 84.44444444444444

Total Accuracy : 92.25941422594143

MCAP LOGISTIC REGRESSION: WITH STOPWORDS

Regularization Parameter : 2

Number of Iterations : 100

Learning rate : 0.01

Accuracy on ham files : 88.61111111111111

Accuracy on spam Files : 75.42372881355932

Total Accuracy : 85.35564853556485

There is a slight decrease in accuracy due to the fact that stopwords typically remove such things as "a, an, the, it". It might be throwing away useful words which might useful for spam filter as ham/spam classification task is more abstract . By removing certain words, we could be throwing valuable information.