# ASSIGNMENT 3

NAÏVE BAYES:

Before filtering stopwords :

Number of files: 463

Number of spam files: 123

Number of ham files: 340

Accuracy 0.9219214600635702

After filtering stopwords :

Number of files: 463

Number of spam files: 123

Number of ham files: 340

Accuracy = 0.923

After the stopwords are filtered, the accuracy of the files is slightly increased, this indicates that the stopwords are not common and stopwords influence the accuracy.

LOGISTIC REGRESSION :

Before filtering stopwords:

Number of Iterations:5

Accuracy: 0.8368200836820083

Number of Iterations:10

Accuracy: 0.8263598326359832

Iterations = 500

Lambda = 0.1

Accuracy = 0.92

Iterations = 500

Lambda = 0.01

Accuracy = 0.88

Iterations = 500

Lambda = 0.001

Accuracy = 0.83

After filtering stopwords:

Iterations = 5

Accuracy: 0.8305439330543933

Iterations = 500

Lambda = 0.1

Accuracy = 0.89

Iterations = 500

Lambda = 0.01

Accuracy = 0.79

Iterations = 500

Lambda = 0.001

Accuracy = 0.887

As we can see lambda and removal of the stopwords improves the accuracy level of the logistic regression.