Naive Bayes Classifier (Assignment 5)
by Daniel McDonough
CS534 12/6/18

Testing Email Body...
True Labels [0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0]
Self Pred labels [0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Naive Bayes from Scratch f-measure: 0.5714285714285715 (k=1)

SciPy Pred Labels [0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1]
Naive Bayes from SciPy f-measure: 0.6666666666666666

Testing Email Subject...
True Labels [0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0]
Self Pred labels [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1]
Naive Bayes from Scratch f-measure: 0.0 (k=1)

SciPy Pred Labels [0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1]
Naive Bayes from SciPy f-measure: 0.6666666666666666

From my implementation of the Naive Bayes Classifier, we can see that using the body is a better
method of determining spam messages by the better f-score.

However, when using the SciPy implementation, we can see that both are equally good to determine if
a message is spam or not. This is due to them updating the known bag of words on each test.