Assignment Report

**Naïve Bayes with Filtering**

Ham accuracy = 91.66666666666666

Spam accuracy = 72.3076923076923

Overall Accuracy = 86.40167364016736

**Naïve Bayes without Filtering**

Ham accuracy = 89.36781609195403

Spam accuracy = 76.15384615384615

**Why does accuracy improve?**

When we remove the stop words from the vocabulary we remove their contribution to the classification of the email as spam or ham. The probability of an email to be classified as spam or ham is solely dependent on the words which are not such commonly occurring words in both categories of email. So the accuracy increases with filtering the stop words.