

## **README File for Machine Learning Programming Assignment 2**

### **Project group**

- 1) Nikhil Lingutla (931-692-837)
- 2) Arwa Hamid (931-673-658)

#### **Folder /NaiveBayes/ nb\_bernoulli.m**

(Computes the Accuracy rate and confusion\_matrix using NaiveBayes bernoulli distribution on the provided 20 news groups dataset. It trains on the training data provided and checks for the accuracy on test dataset)

#### **Folder /NaiveBayes/ nb\_multinomial.m**

(Computes the Accuracy rate and confusion\_matrix using NaiveBayes Multinomial distribution on the provided 20 news groups dataset. It trains on the training data provided and checks for the accuracy on test dataset)

#### **Folder /NaiveBayes/ nb\_stop\_bernoulli.m**

(At first this file removes the stop words from the given vocabulary and then adjust the training data as required. Then the learned parameters are used to calculate the accuracy of classification on test dataset. This uses NaiveBayes bernoulli Distribution)

#### **Folder /NaiveBayes/ nb\_stop\_multinomial.m**

(At first this file removes the stop words from the given vocabulary and then adjust the training data as required. Then the learned parameters are used to calculate the accuracy of classification on test dataset. This uses NaiveBayes multinomial Distribution)

#### **Folder /NaiveBayes/ vocabulary.text**

(This contains the vocabulary of the given 20 news group dataset. This vocabulary is by using Bag of words model)

**Folder /NaiveBayes/  
stopwords**

(This file contains the various stop words collected from the internet to test our assumption that removal of stop words from vocabulary reduce the burden on the classifier training without much effect on the accuracy rate)