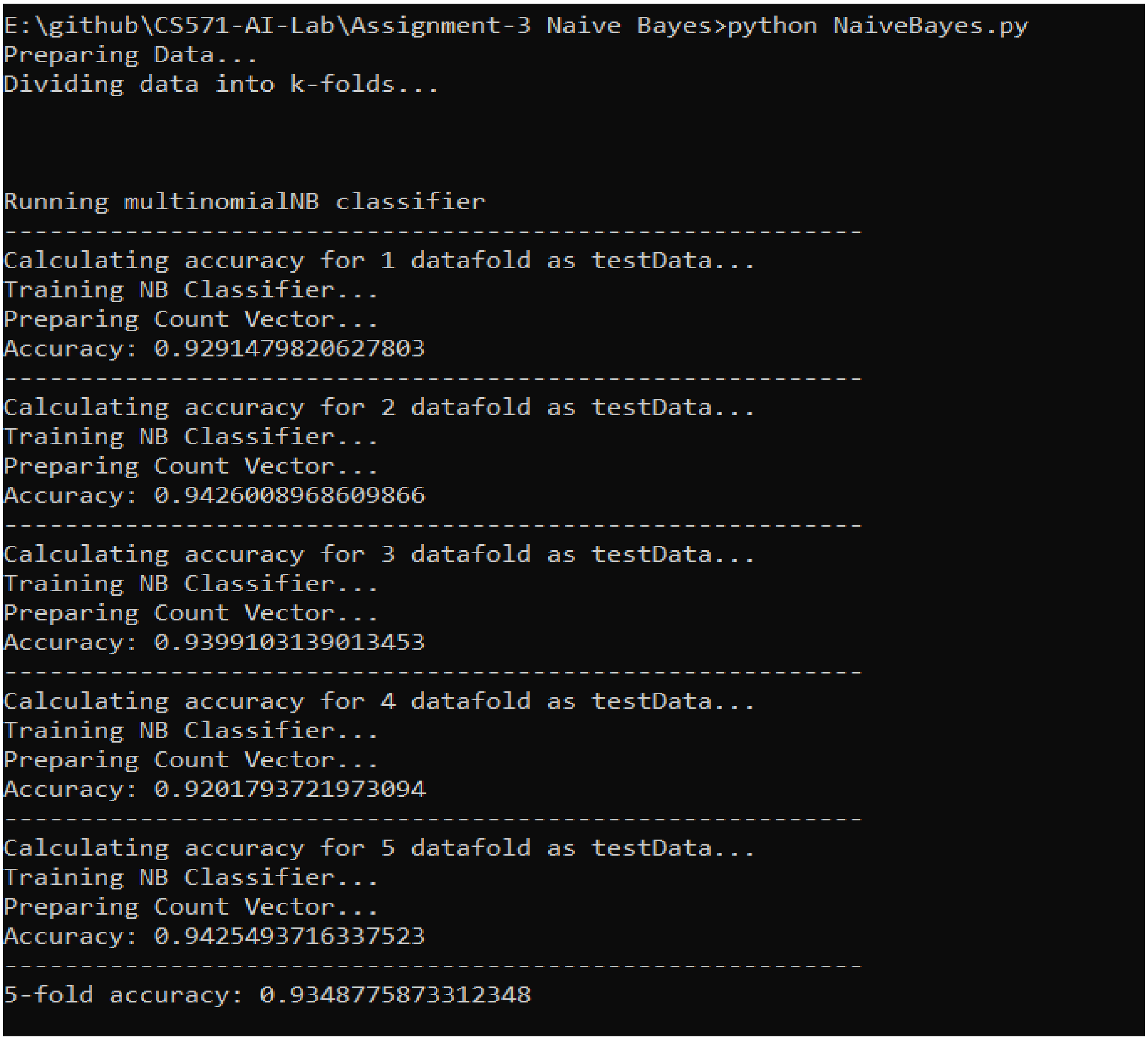
CS561 - ARTIFICIAL INTELLIGENCE LAB
ASSIGNMENT-3: Naive Bayes Classifier

GroupID : 1801cs12_1801cs16_1801cs22
Date : 14/09/2021

1801CS12 : Bablu Kumar
1801CS16 : Mangesh Chandrawanshi
1801CS22 : Hrishabh Raj

Result for Multinomial Naive Bayes Classifier :

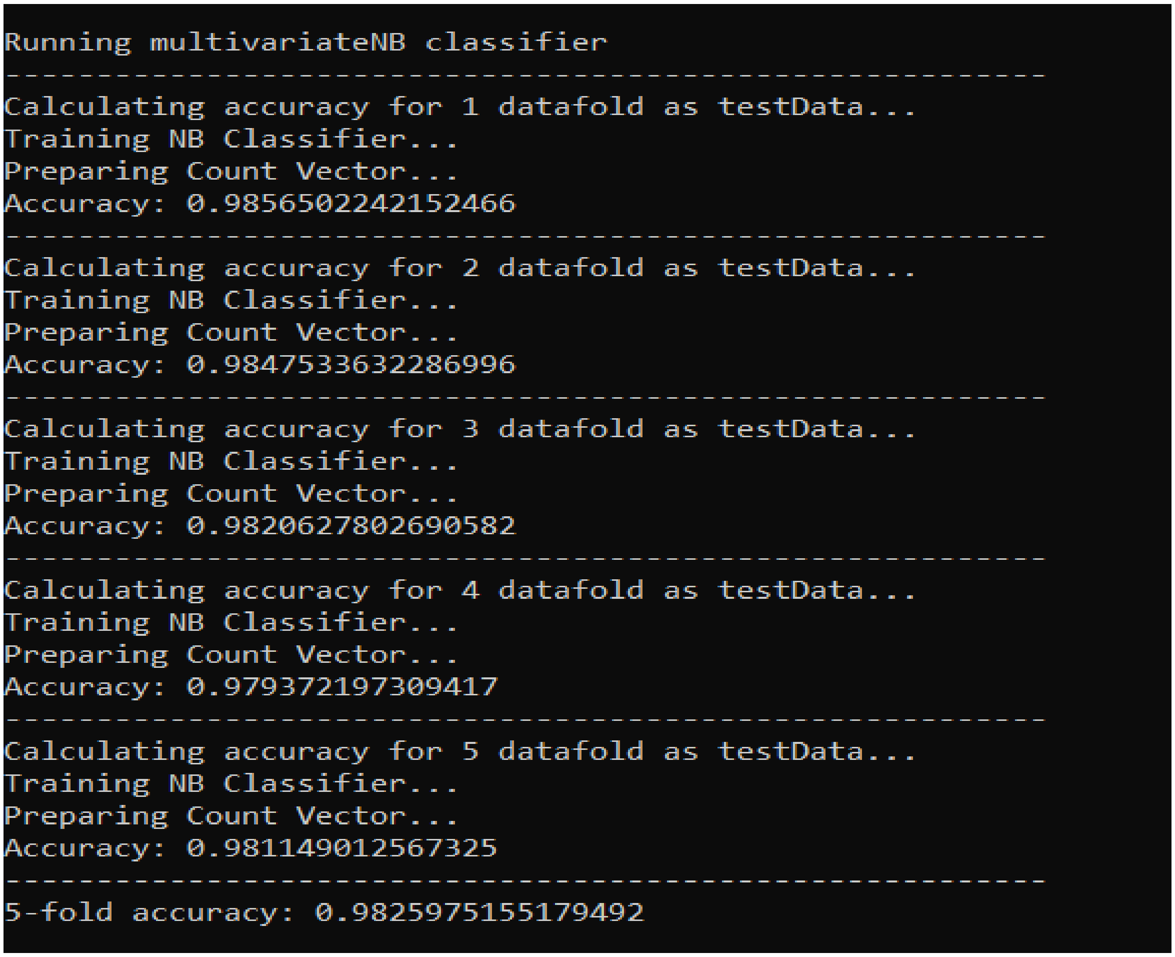


```
E:\github\CS571-AI-Lab\Assignment-3 Naive Bayes>python NaiveBayes.py
Preparing Data...
Dividing data into k-folds...



Running multinomialNB classifier
---------------------------------------------------------
Calculating accuracy for 1 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.9291479820627803
---------------------------------------------------------
Calculating accuracy for 2 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.9426008968609866
---------------------------------------------------------
Calculating accuracy for 3 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.9399103139013453
---------------------------------------------------------
Calculating accuracy for 4 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.9201793721973094
---------------------------------------------------------
Calculating accuracy for 5 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.9425493716337523
---------------------------------------------------------
5-fold accuracy: 0.9348775873312348
```

Result for Multivariate Naive Bayes Classifier :



```
Running multivariateNB classifier
---------------------------------------------------------
Calculating accuracy for 1 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.9856502242152466
---------------------------------------------------------
Calculating accuracy for 2 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.9847533632286996
---------------------------------------------------------
Calculating accuracy for 3 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.9820627802690582
---------------------------------------------------------
Calculating accuracy for 4 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.979372197309417
---------------------------------------------------------
Calculating accuracy for 5 datafold as testData...
Training NB Classifier...
Preparing Count Vector...
Accuracy: 0.981149012567325
---------------------------------------------------------
5-fold accuracy: 0.9825975155179492
```

Comparisons :

	Multinomial NB	**Multivariate NB**
5-fold Accuracy	**~94%**	**~98%**
Time of Execution	**less**	**more**

In case of the multivariate NB classifier, we iterate over every word present in the training data therefore it leads to better results in spam detection as some words are spam identifiers in most cases (like CONGRATULATIONS, FREE, HURRY UP, etc) but time of execution is large.

In the case of the multinomial NB classifier, probabilities of words present in the test sample are only considered so the search space is reduced which results in quicker execution but affects accuracy to some extent.

Note : For execution and program structure details, refer to README.md