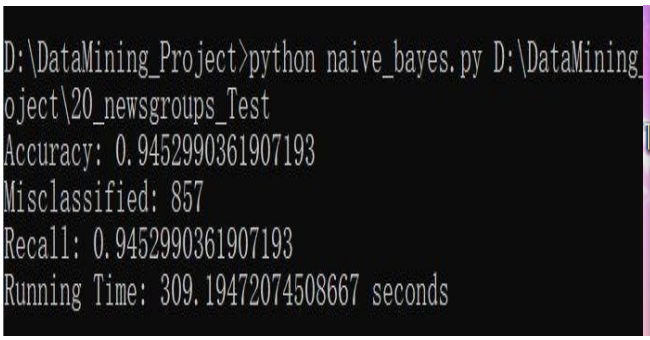
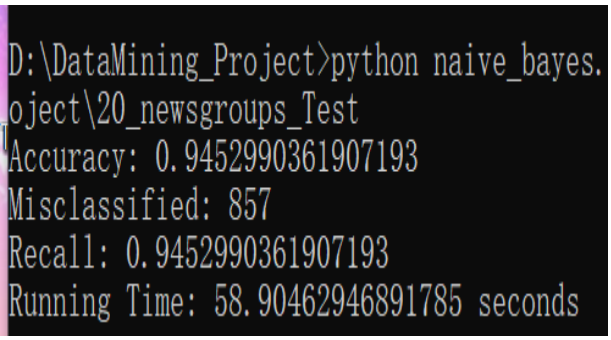
Part a:

Naïve Bayes

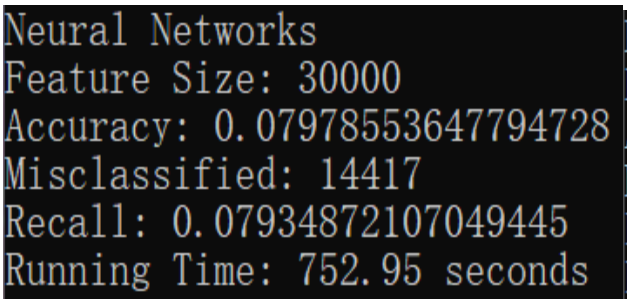


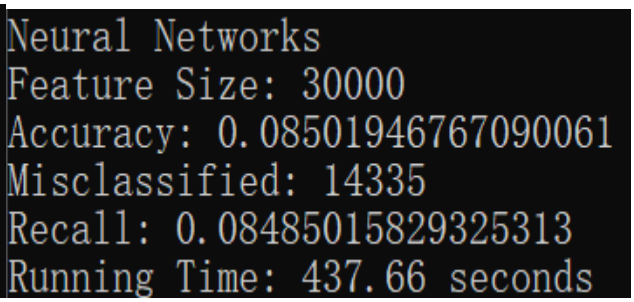


Neural Networks

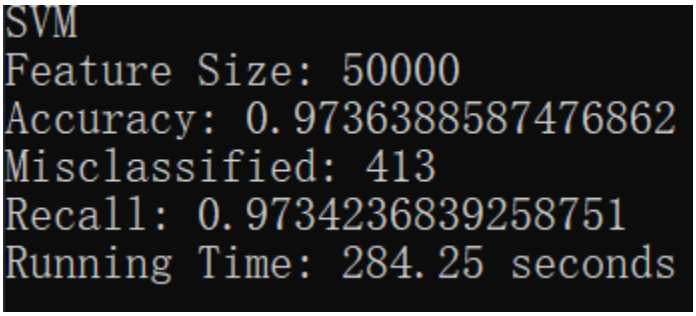
Left: 50 layers with 24 neurons

Right: 10 layers with 4 neurons





SVM



Part b:

Use 50,000 as feature size for SVM, 30,000 for Neural Network since it will run 20 mins if did not change.

Part c:

For Neural Networks I used "**sklearn**" and "**tensorflow**", SVM only use "**sklearn**"

Part d:

Dictionary size	Accuracy	Number of Misclassified	Recall	Running Time
70000	0.9462	843	0.9462	61.069 seconds

50000	0.9385	963	0.9385	59.689 seconds
30000	0.9184	1278	0.9184	61.271 seconds
10000	0.8337	2606	0.8337	58.206 seconds

```
Task7 part d
Dictionary Size: 70000
Accuracy: 0.9461926341992724
Number of Misclassified: 843
Recall: 0.9461926341992724
Running Time: 61.069172620773315 seconds

Dictionary Size: 50000
Accuracy: 0.9385332226973894
Number of Misclassified: 963
Recall: 0.9385332226973894
Running Time: 59.68855357170105 seconds

Dictionary Size: 30000
Accuracy: 0.9184272675049467
Number of Misclassified: 1278
Recall: 0.9184272675049467
Running Time: 61.270870208740234 seconds

Dictionary Size: 10000
Accuracy: 0.8336631135507755
Number of Misclassified: 2606
Recall: 0.8336631135507755
Running Time: 58.20561099052429 seconds
```

Part e:

When the dictionary size decreases the accuracy decreases. The reason can cause this issue I believe is information loss. When the feature size is reduced it causes information loss, there is a big chance that lost information could be the words that are critical for distinguish. Also, small dictionary size could lead to underfitting, and it will cause the model to lack the capacity to learn the complexity of the data.

Part f:

NN:

Left: 50 layers with 24 neurons

```
Neural Networks
Feature Size: 10000
Accuracy: 0.08112593349077679
Misclassified: 14396
Recall: 0.08086373308965242
Running Time: 750.42 seconds
```

Right: 10 layers with 4 neurons

```
Neural Networks
Feature Size: 10000
Accuracy: 0.07123252696751133
Misclassified: 14551
Recall: 0.07130388719799782
Running Time: 369.89 seconds
```

SVM:

```
SVM
Feature Size: 50000
Accuracy: 0.9736388587476862
Misclassified: 413
Recall: 0.9734236839258751
Running Time: 284.25 seconds

SVM
Feature Size: 10000
Accuracy: 0.9735112018893215
Misclassified: 415
Recall: 0.9733027344460673
Running Time: 237.93 seconds
```