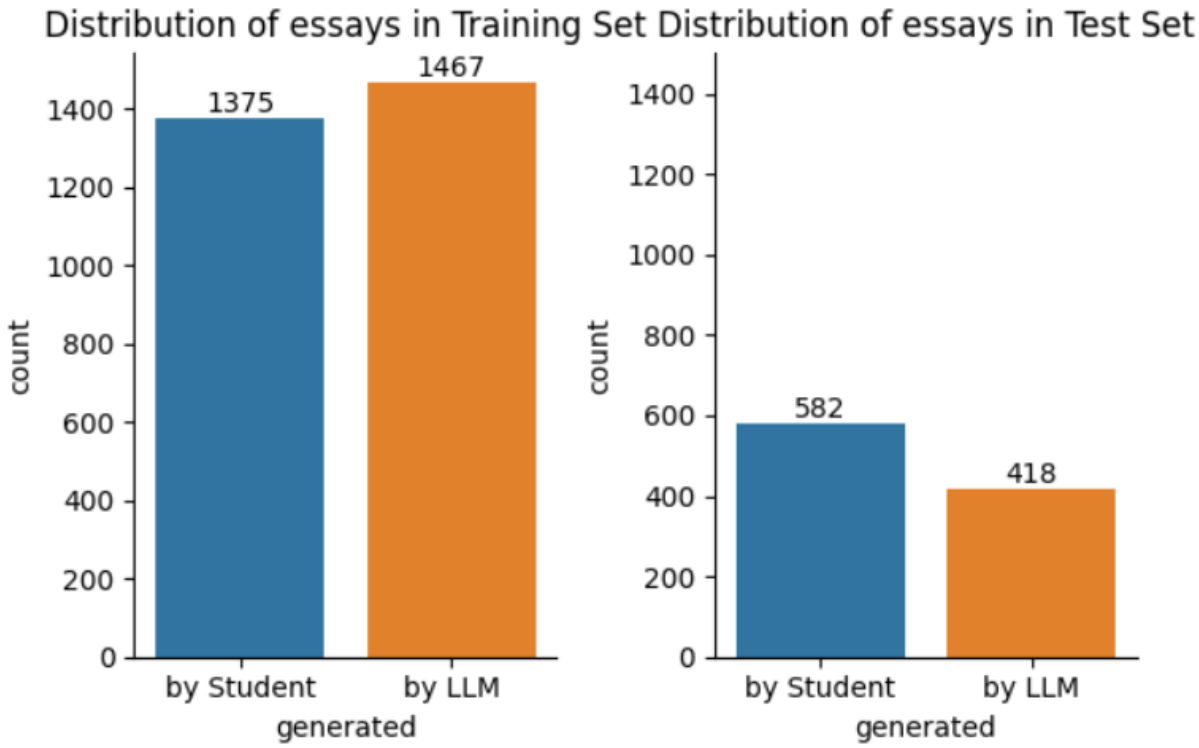
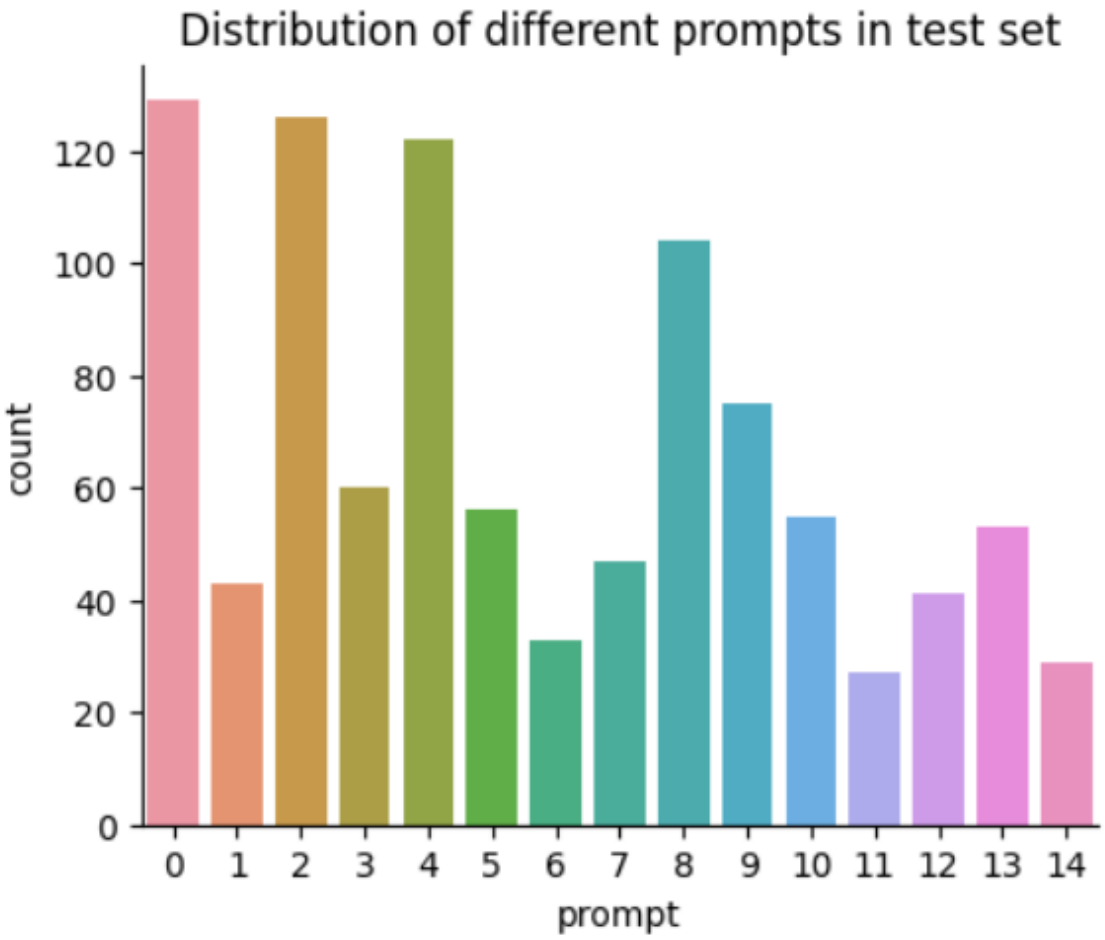
EXERCISE A3 PRACTICAL DATA SCIENCE

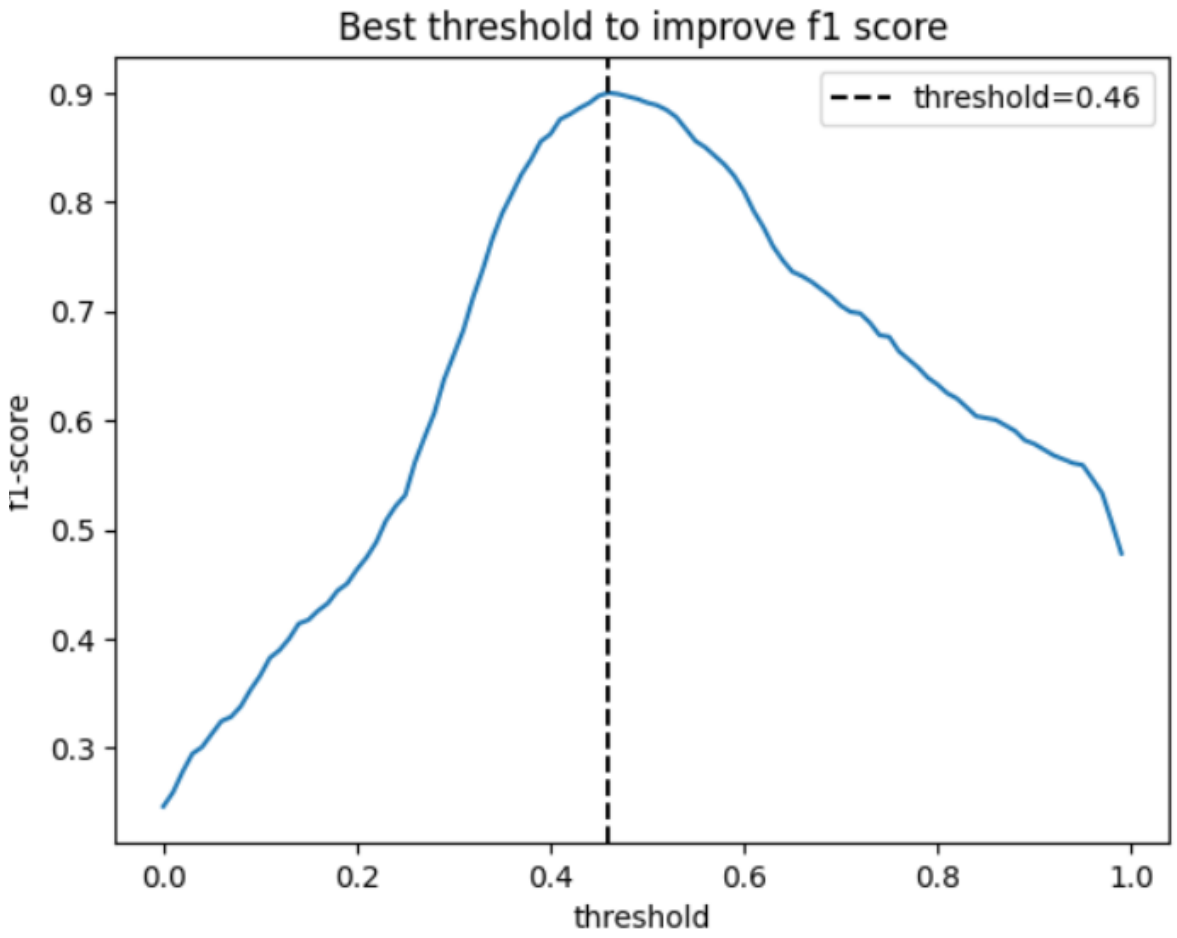
VOULGARAKIS DIONYSIOS





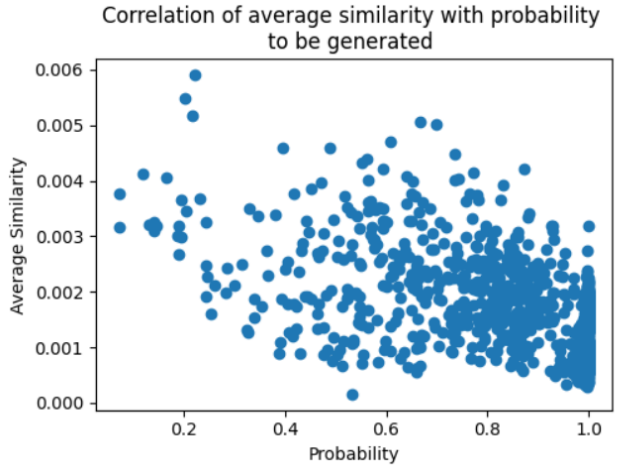

```
Fitting 5 folds for each of 7 candidates, totalling 35 fits
Classification report for MultinomialNB(alpha=0.5):
              precision    recall  f1-score   support

           0       0.99      1.00      0.99       605
           1       1.00      0.98      0.99       395

    accuracy                           0.99      1000
   macro avg       0.99      0.99      0.99      1000
weighted avg       0.99      0.99      0.99      1000
```

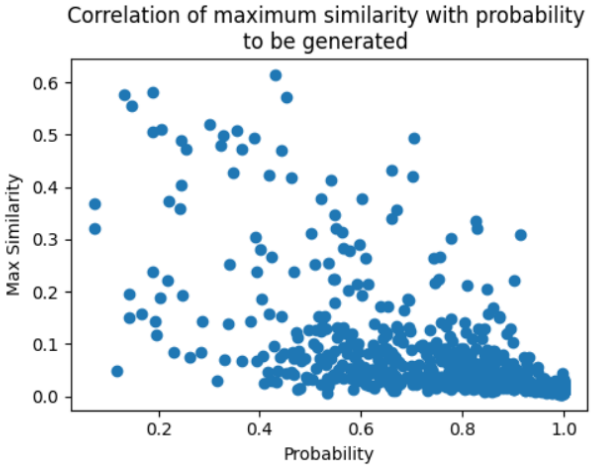
```
Fitting 5 folds for each of 7 candidates, totalling 35 fits
Classification report for MultinomialNB():
              precision    recall  f1-score   support

           0       0.85      0.99      0.92       582
           1       0.99      0.75      0.86       418

    accuracy                           0.89      1000
   macro avg       0.92      0.87      0.89      1000
weighted avg       0.91      0.89      0.89      1000
```

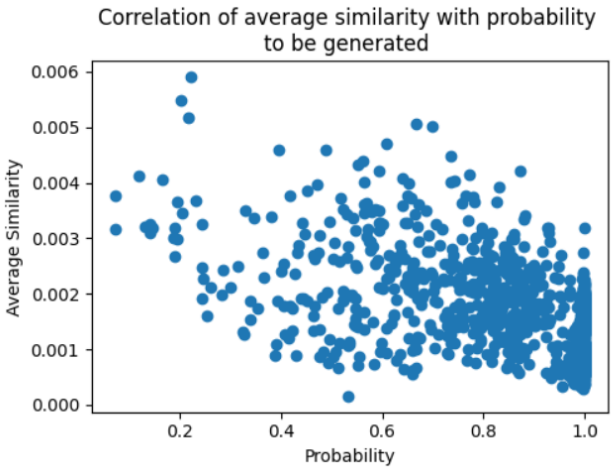


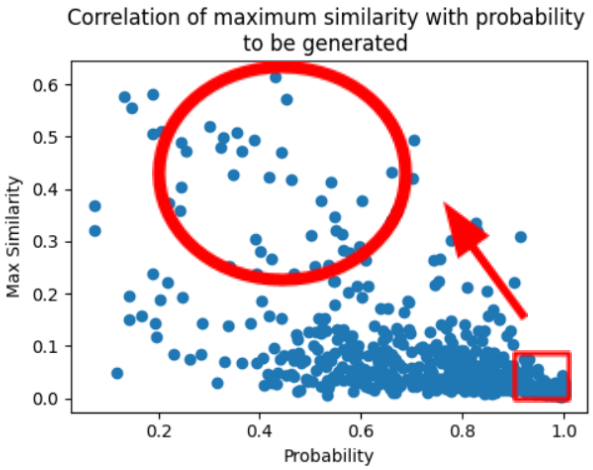
Classification report for MultinomialNB():

	precision	recall	f1-score	support
0	0.87	0.98	0.92	582
1	0.97	0.79	0.87	418
accuracy			0.90	1000
macro avg	0.92	0.89	0.90	1000
weighted avg	0.91	0.90	0.90	1000

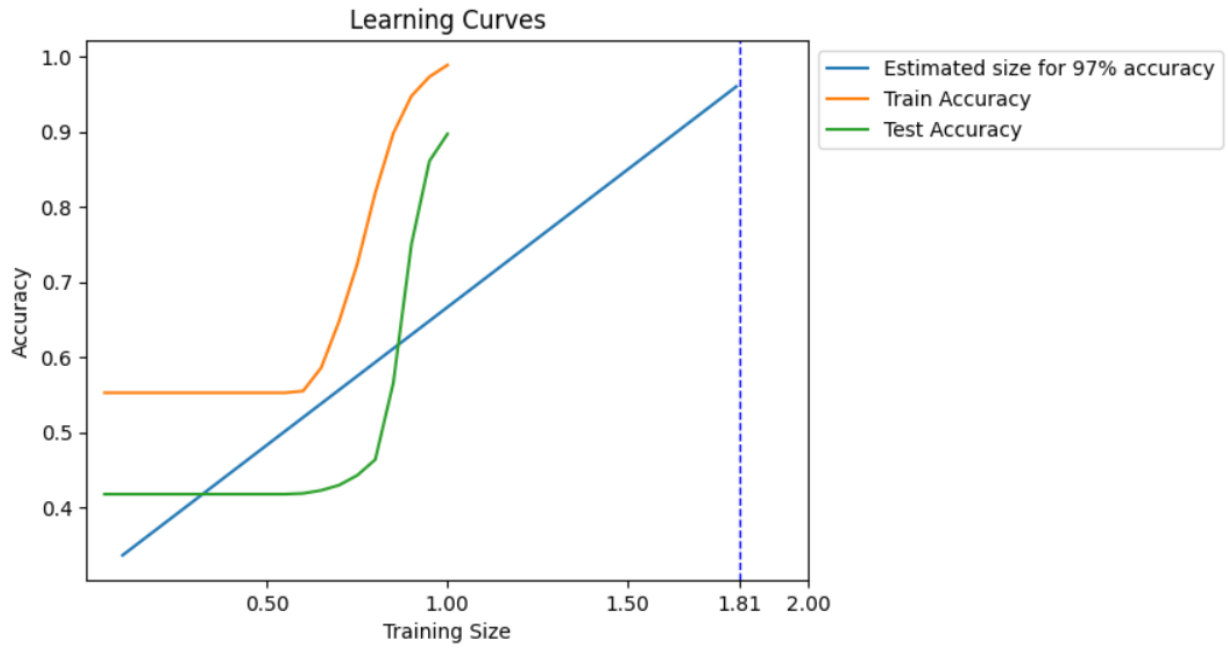


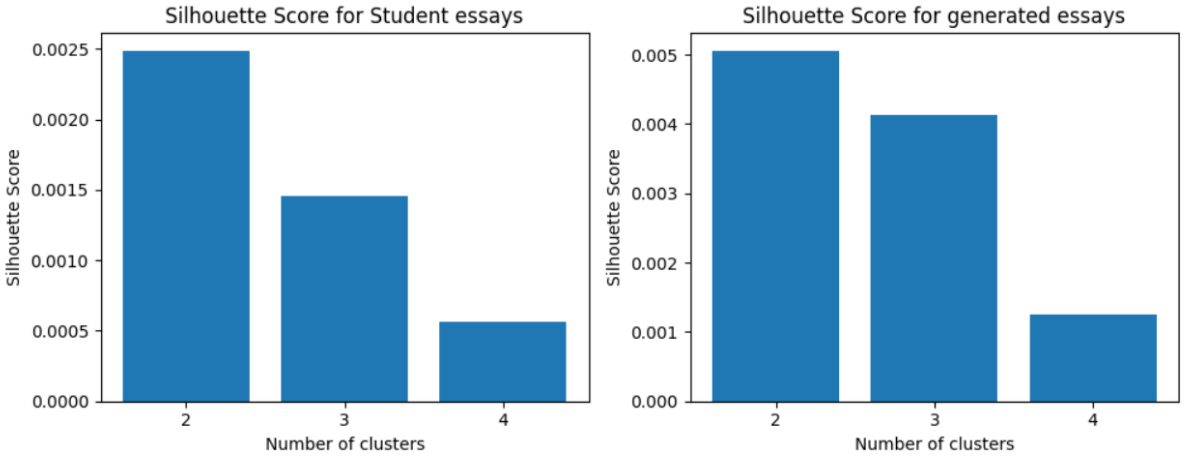


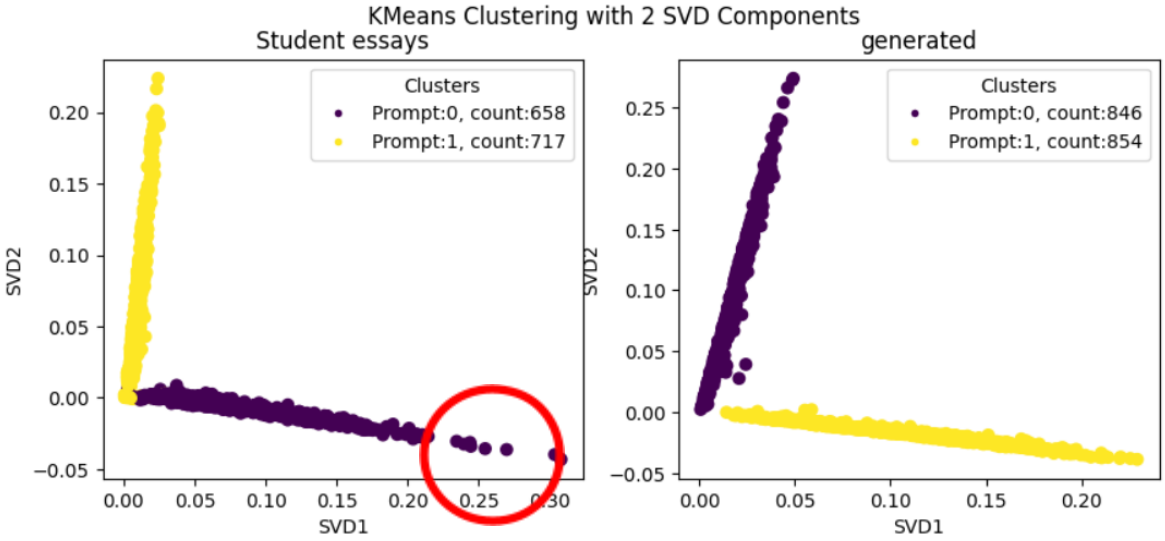


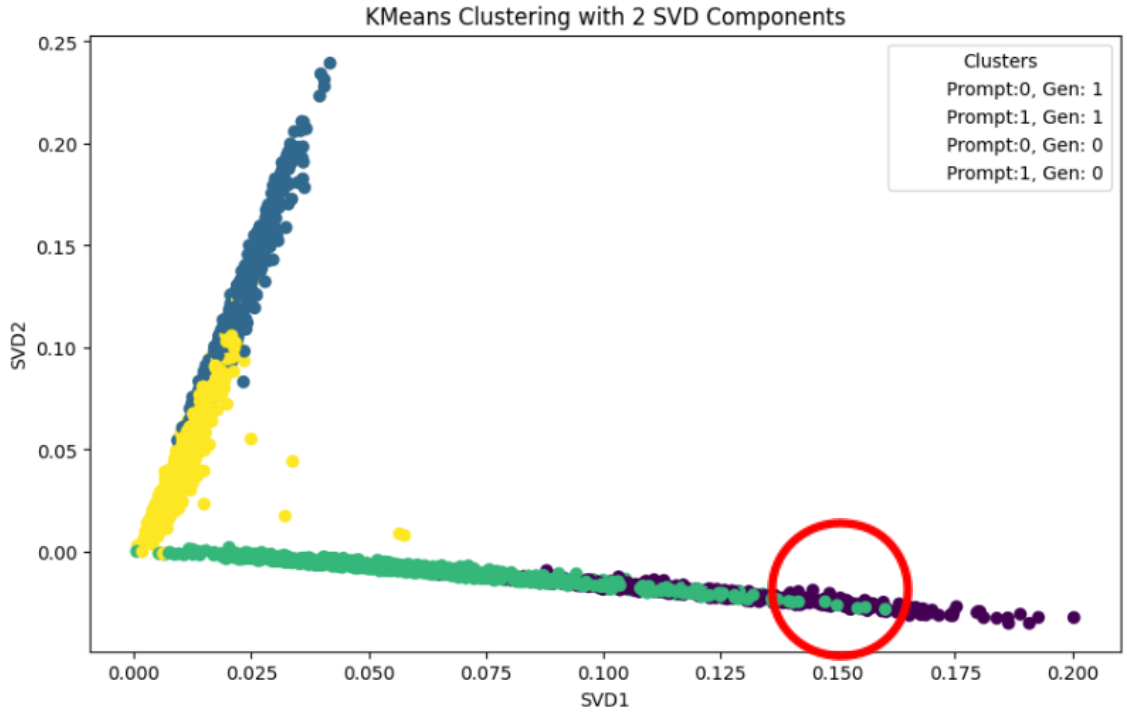



```
Fitting 5 folds for each of 6 candidates, totalling 30 fits
Classification report for MultinomialNB(alpha=0.5):
              precision    recall  f1-score   support

           0       0.87      0.97      0.92       582
           1       0.95      0.80      0.87       418

    accuracy                           0.90      1000
   macro avg       0.91      0.88      0.89      1000
weighted avg       0.90      0.90      0.90      1000
```


Learning Curves
Estimated size for 97% accuracy
Train Accuracy
Test Accuracy
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.50
1.00
1.50
1.81
2.00
Training Size


Silhouette Score for Student essays
Silhouette Score
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
2
3
4
Number of clusters
Silhouette Score for generated essays
Silhouette Score
0.005
0.004
0.003
0.002
0.001
0.000
2
3
4
Number of clusters


KMeans Clustering with 2 SVD Components
Student essays
generated
Clusters
Prompt:0, count:658
Prompt:1, count:717
Prompt:0, count:846
Prompt:1, count:854
SVD1
SVD2


KMeans Clustering with 2 SVD Components
Clusters
Prompt:0, Gen: 1
Prompt:1, Gen: 1
Prompt:0, Gen: 0
Prompt:1, Gen: 0
SVD1
SVD2

Classification report for MultinomialNB(alpha=0.5):

	precision	recall	f1-score	support
0	0.91	0.92	0.92	582
1	0.89	0.88	0.88	418
accuracy			0.90	1000
macro avg	0.90	0.90	0.90	1000
weighted avg	0.90	0.90	0.90	1000