Kara Capps 03-29-2022
927000068

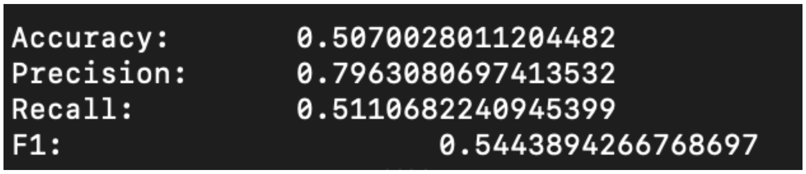
Programming Assignment 2

1. Classifier Results

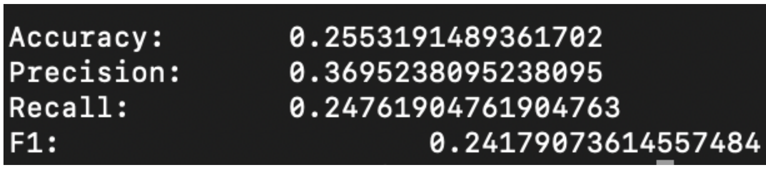
NAIVE BAYES CLASSIFIER

Naive Bayes Train

Training scores: Validation scores:



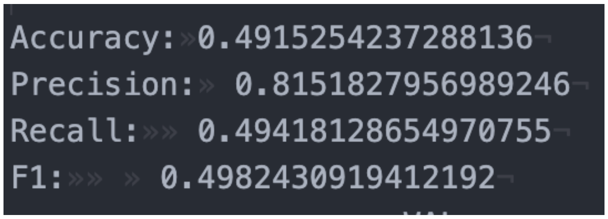
```
Accuracy:        0.5070028011204482
Precision:       0.7963080697413532
Recall:          0.5110682240945399
F1:                      0.5443894266768697
```



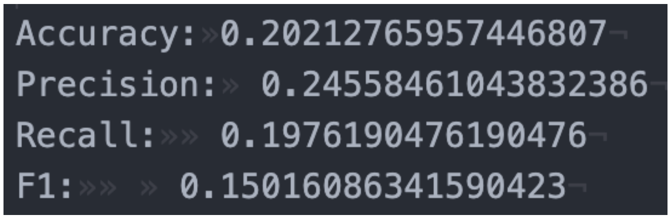
```
Accuracy:        0.2553191489361702
Precision:       0.3695238095238095
Recall:          0.24761904761904763
F1:                      0.24179073614557484
```

Naive Bayes Train Half

Training scores: Validation scores:



```
Accuracy:»0.4915254237288136¬
Precision:» 0.8151827956989246¬
Recall:»» 0.49418128654970755¬
F1:»» » 0.4982430919412192¬
```

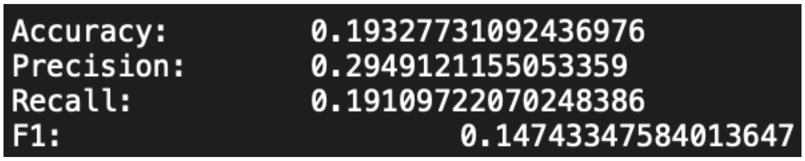


```
Accuracy:»0.20212765957446807¬
Precision:» 0.24558461043832386¬
Recall:»» 0.1976190476190476¬
F1:»» » 0.15016086341590423¬
```

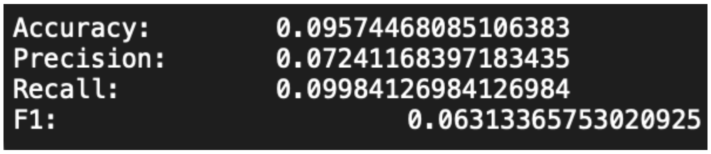
KNN CLASSIFIER

knn train

Training scores: Validation scores



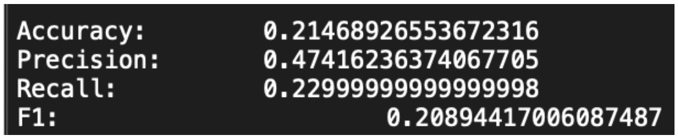
```
Accuracy:        0.19327731092436976
Precision:       0.2949121155053359
Recall:          0.19109722070248386
F1:                      0.14743347584013647
```



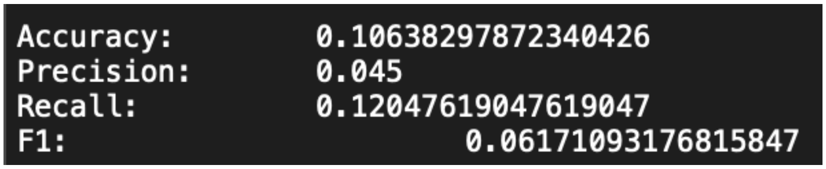
```
Accuracy:        0.09574468085106383
Precision:       0.07241168397183435
Recall:          0.09984126984126984
F1:                      0.06313365753020925
```

knn train half

Training scores: Validation scores:



```
Accuracy:        0.21468926553672316
Precision:       0.47416236374067705
Recall:          0.22999999999999998
F1:                      0.20894417006087487
```



```
Accuracy:        0.10638297872340426
Precision:       0.045
Recall:          0.12047619047619047
F1:                      0.06171093176815847
```

rocchio

2.The validation scores for Naive Bays are lower than the training scores, meaning that the model has a high bias.

For knn, the Training scores are also higher than the validation scores, but the difference is not as drastic as Naive Bayes, meaning that Ann’s bias is lower than Naive Bayes’ bias.

Naive Bayes’ scored higher than knn for every metric.

3. The time complexity of Naive-Bayes train() was O(n^2) because of the two for loops that loop through the documents and then each word in each document. The predict is O(n).

The time complexity of knn train() is O(n^2) because of looping through the document set and then each word inside the document. The time complexity of knn predict() is O(n) because of the for loop that loops through all of the indexes in the bag of words vector.

The time complexity of Rocchio train() is O(n^2) because we loop through the document set and then each word inside the document.