

BOĞAZIÇI UNIVERSITY

CMPE 493 - Information Retrieval

Spring 2020

Assignment II

Movie Review Classification
with Naive Bayes

Mahmut Uzunpostalcı

1 Results

1.1 Multinomial Naive Bayes

		Actual	
		Positive	Negative
Prediction	Positive	TP 239	FP 51
	Negative	FN 61	TN 249

Figure 1: Contingency table of multinomial naive bayes model

Macro-Averaged Scores	Precision	0.8137
	Recall	0.8133
	F1 Score	0.8133
Micro-Averaged Scores	Precision	0.8133
	Recall	0.8133
	F1 Score	0.8133

1.2 Binary Naive Bayes

		Actual	
		Positive	Negative
Prediction	Positive	TP 230	FP 40
	Negative	FN 70	TN 260

Figure 2: Contingency table of binary naive bayes model

Macro-Averaged Scores	Precision	0.8199
	Recall	0.8167
	F1 Score	0.8162
Micro-Averaged Scores	Precision	0.8167
	Recall	0.8167
	F1 Score	0.8167

1.3 Bernoulli Naive Bayes

		Actual	
		Positive	Negative
Prediction	Positive	TP 204	FP 24
	Negative	FN 96	TN 276

Figure 3: Contingency table of bernoulli naive bayes model

Macro-Averaged Scores	Precision	0.8183
	Recall	0.8000
	F1 Score	0.7971
Micro-Averaged Scores	Precision	0.8000
	Recall	0.8000
	F1 Score	0.8000

2 Approximate Randomization Tests

P values of randomization tests between micro-averaged F-scores:

Multinomial & Binary	0.8871
Multinomial & Bernoulli	0.4815
Binary & Bernoulli	0.2407

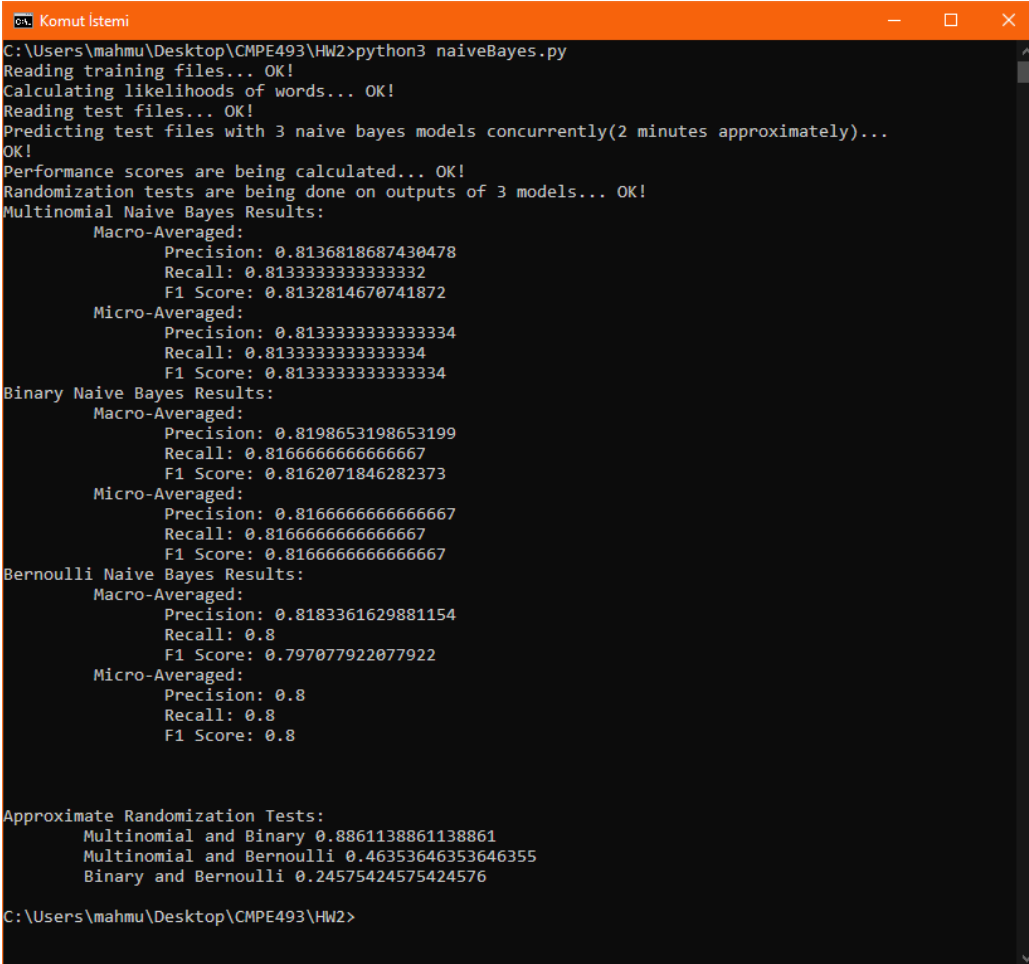
3 Discussion

Results of the approximate randomization tests indicate that we can not reject the null hypothesis. This is not surprising as the models are not different from each other that much. For example the only difference between multinomial and binary naive bayes is there aren't any duplicate words in a document for binary naive bayes. Similar to binary naive bayes, bernoulli naive bayes also disregards duplicate words in document.

According to p values of approximate randomization test and the contingency tables, bernoulli is the most different from other two models. One reason of this maybe that the model does take into consideration the probability of not being in the document when predicting. Other interesting thing about bernoulli naive bayes that I couldn't find a reason, is it is more likely to predict reviews as negative with respect to other two models.

Overall the results of the models are not that different from each other even though they are different models. I think the reason is that all three models are probabilistic models and they do not consider locations of words while training.

4 Screenshot



```
C:\Users\mahmu\Desktop\CMPE493\HW2>python3 naiveBayes.py
Reading training files... OK!
Calculating likelihoods of words... OK!
Reading test files... OK!
Predicting test files with 3 naive bayes models concurrently(2 minutes approximately)...
OK!
Performance scores are being calculated... OK!
Randomization tests are being done on outputs of 3 models... OK!
Multinomial Naive Bayes Results:
    Macro-Averaged:
        Precision: 0.8136818687430478
        Recall: 0.8133333333333332
        F1 Score: 0.8132814670741872
    Micro-Averaged:
        Precision: 0.8133333333333334
        Recall: 0.8133333333333334
        F1 Score: 0.8133333333333334
Binary Naive Bayes Results:
    Macro-Averaged:
        Precision: 0.8198653198653199
        Recall: 0.8166666666666667
        F1 Score: 0.8162071846282373
    Micro-Averaged:
        Precision: 0.8166666666666667
        Recall: 0.8166666666666667
        F1 Score: 0.8166666666666667
Bernoulli Naive Bayes Results:
    Macro-Averaged:
        Precision: 0.8183361629881154
        Recall: 0.8
        F1 Score: 0.797077922077922
    Micro-Averaged:
        Precision: 0.8
        Recall: 0.8
        F1 Score: 0.8

Approximate Randomization Tests:
    Multinomial and Binary 0.8861138861138861
    Multinomial and Bernoulli 0.46353646353646355
    Binary and Bernoulli 0.24575424575424576

C:\Users\mahmu\Desktop\CMPE493\HW2>
```

Figure 4: Screenshot of program running