Boğazıçı University

CMPE 493 - Information Retrieval

Spring 2020

Assignment II

Movie Review Classification with Naive Bayes

Mahmut Uzunpostalcı

1 Results

1.1 Multinomial Naive Bayes

Figure 1: Contingency table of multinaomial naive bayes model

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

1.2   Binary Naive Bayes

Figure 2: Contingency table of binary naive bayes model

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

1.3 Bernoulli Naive Bayes

Figure 3: Contingency table of bernoulli naive bayes model

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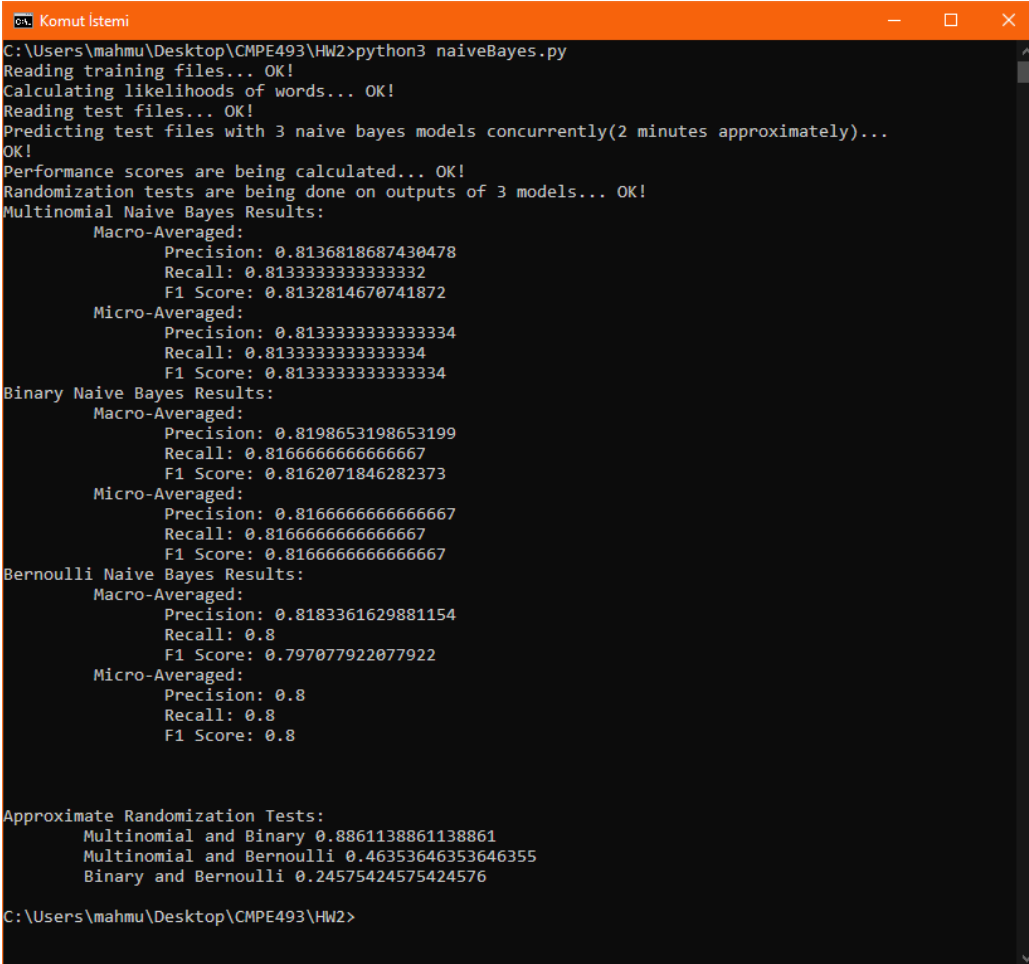
3   Discussion

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

Figure 4: Screenshot of program running