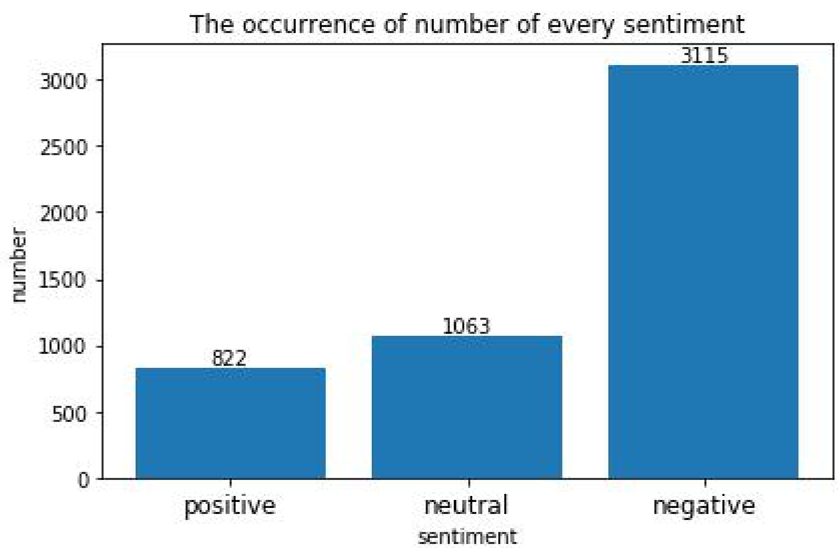
9414 Assignment2

Z5239803 Zhengyang Jia

Question 1:

We use the function pyplot from matplotlib to get the following chart, which shows the distribution intuitively.



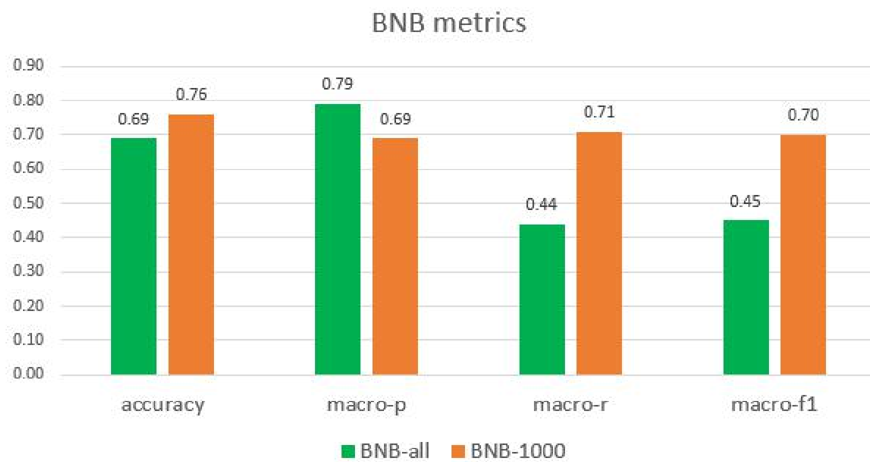
From the chart we know that positive and neutral comments are relatively fewer than negative comments. The positive data has the lowest frequency, while the negative data has the highest frequency.

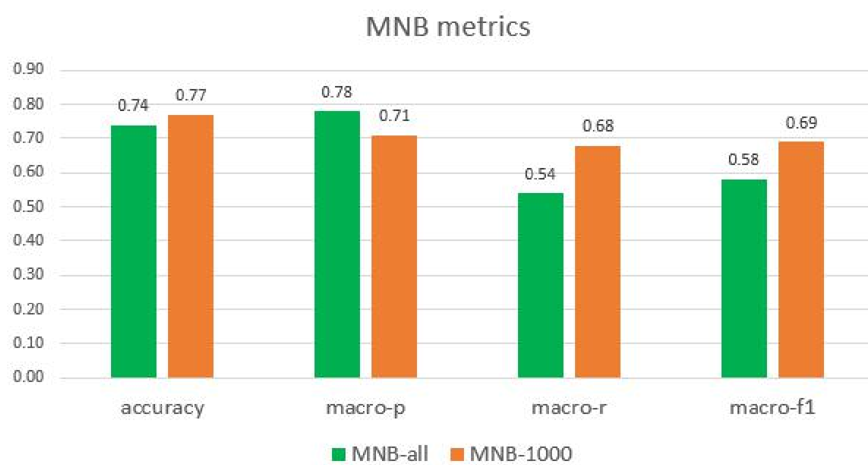
Question 2:

According to the definition of micro, we have:

micro-precision= micro-recall= micro-f1 = accuracy.

I will use accuracy to represent micro-precision, micro-recall, and micro-f1 in the following questions.





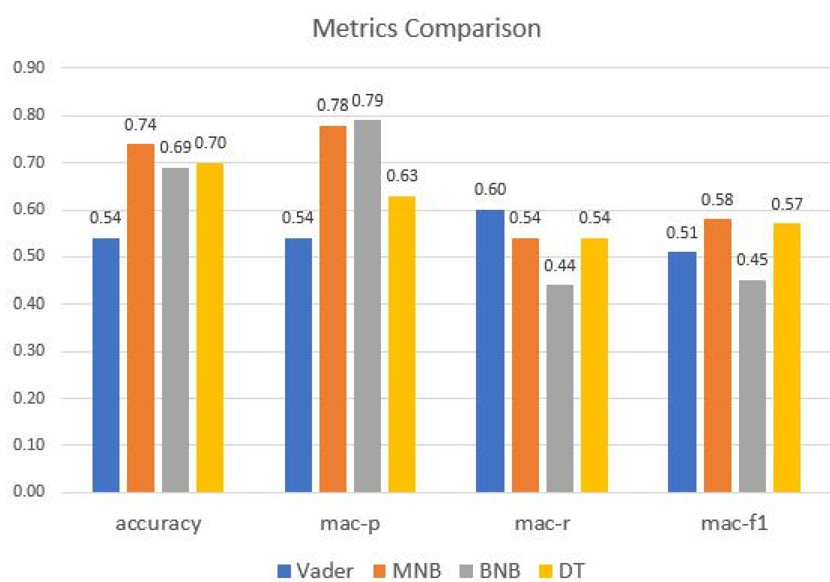
From the two groups of comparison, we can find that for micro-metric (accuracy), group(a)-all is lower than group(b)-1000. This applies to both BNB and MNB.

For macro-metrics, we find macro-precision for group(a) is higher than group(b), while for macro-recall and macro-f1, group(a) is lower than group(b). These properties also apply to both BNB and MNB.

These may indicate that micro focuses on minority while macro focuses on majority.

We can find something else: The distinction between (a)-all and (b)-1000 in macro-r and macro-f1 is much larger in BNB than MNB.

Question 3:

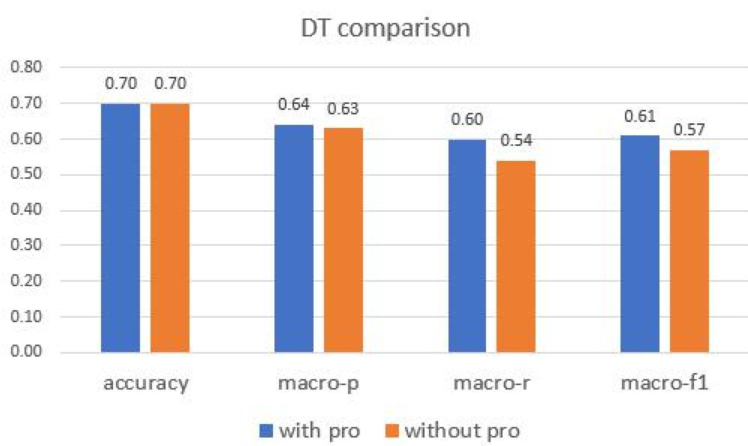


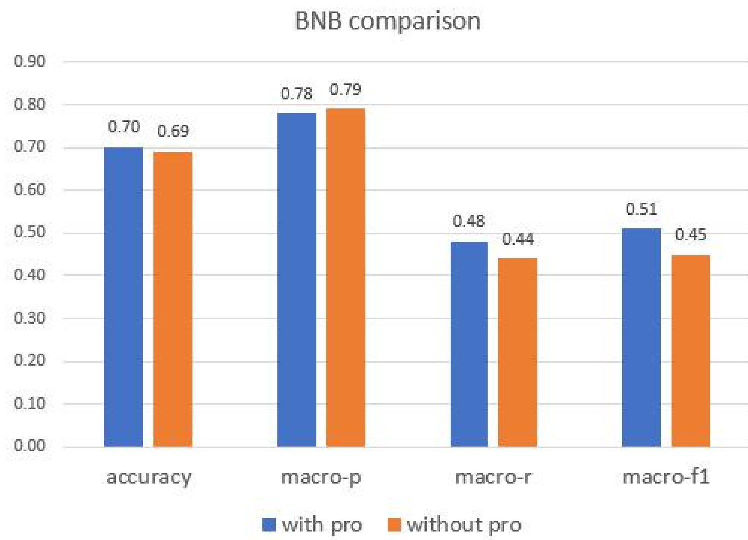
In the diagram containing these metrics obtained from results, we can find that the accuracy of Vader is the lowest, while MNB accuracy is higher than others. For macro-precision, MNB and BNB have relatively higher values than the other two.

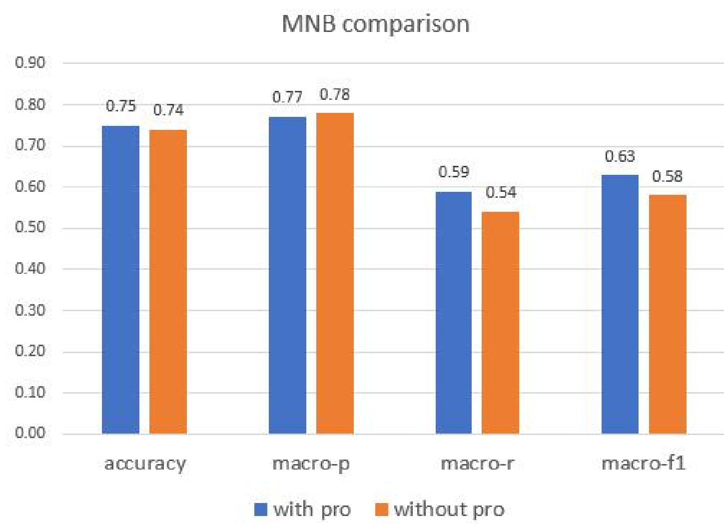
For macro-recall, Vader has the highest value. For macro-f1, Vader ranks in the middle.

From the result we know though Vader performs poor in micro and macro-precision, it can be used to improve macro-recall.

Question 4:







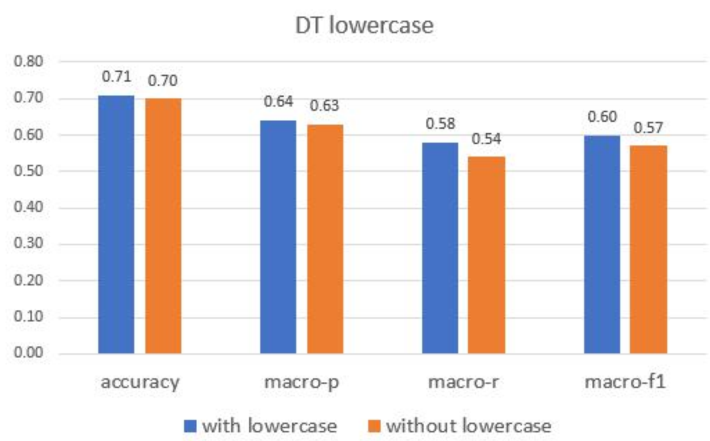
From the charts, we know after applying preprocessing, the overall metrics grew higher.

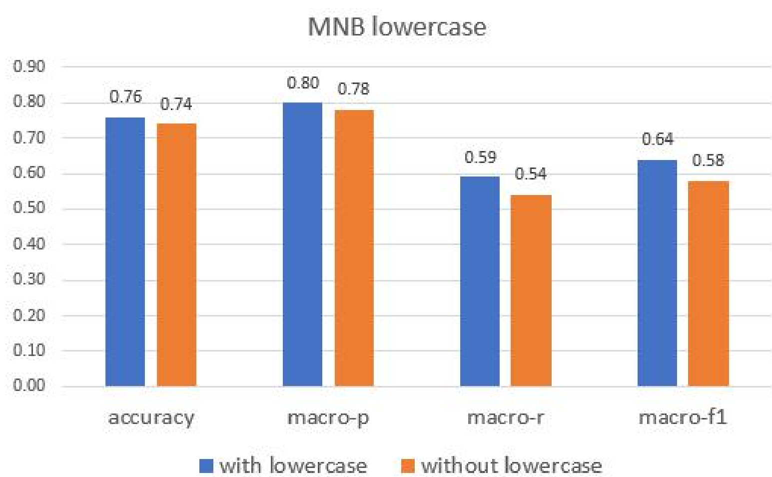
For DT, the metrics got better after preprocessing.

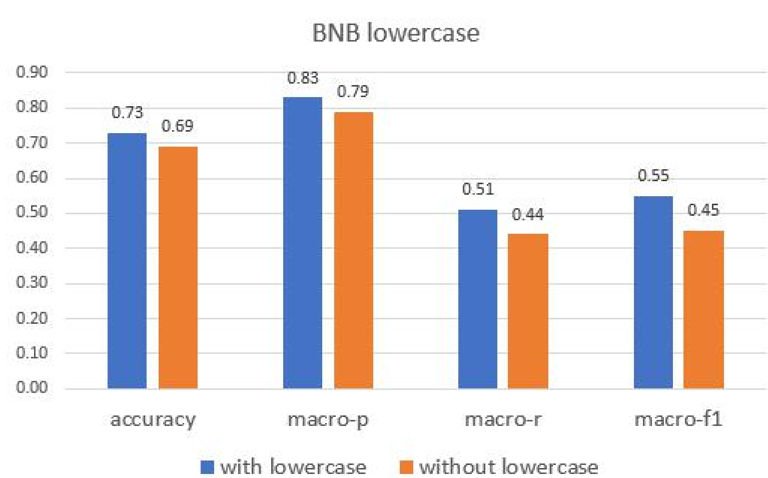
For BNB and MNB, after preprocessing, only macro-precision got lower, while all other metrics became larger.

The result indicates that after removing stop word and stemming, these three models function better generally.

Question 5:


DT lowercase
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.71
0.70
0.64
0.63
0.58
0.54
0.60
0.57
accuracy
macro-p
macro-r
macro-f1
with lowercase
without lowercase


MNB lowercase
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.76
0.74
0.80
0.78
0.59
0.54
0.64
0.58
accuracy
macro-p
macro-r
macro-f1
with lowercase
without lowercase


BNB lowercase
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.73
0.69
0.83
0.79
0.51
0.44
0.55
0.45
accuracy
macro-p
macro-r
macro-f1
with lowercase
without lowercase

From the charts, we know after lowercase conversion, all the metrics grew higher for three models: DT, MNB, BNB.

The result shows that lowercase conversion can improve performance for all three models.

Lowercase conversion could be a great optimization for sentimental assessment.

Question 6:

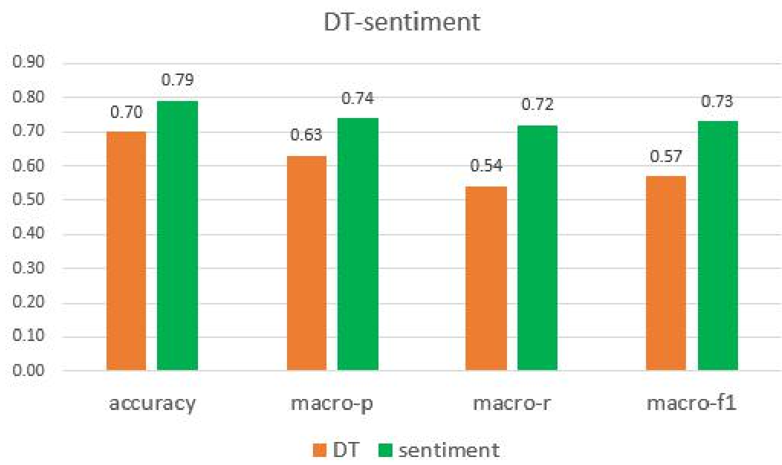
After comparing those above potential optimizations, the relatively better model has been found, which contains the following properties:

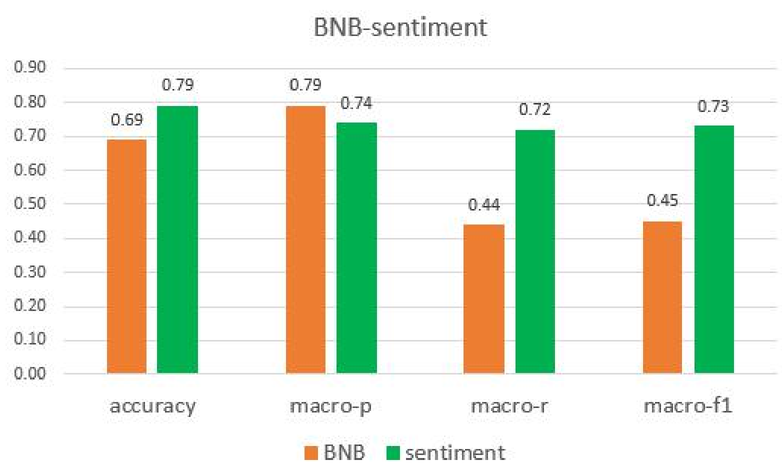
1. MNB **2.** Using most frequent 1000 words rather than the whole vocabulary

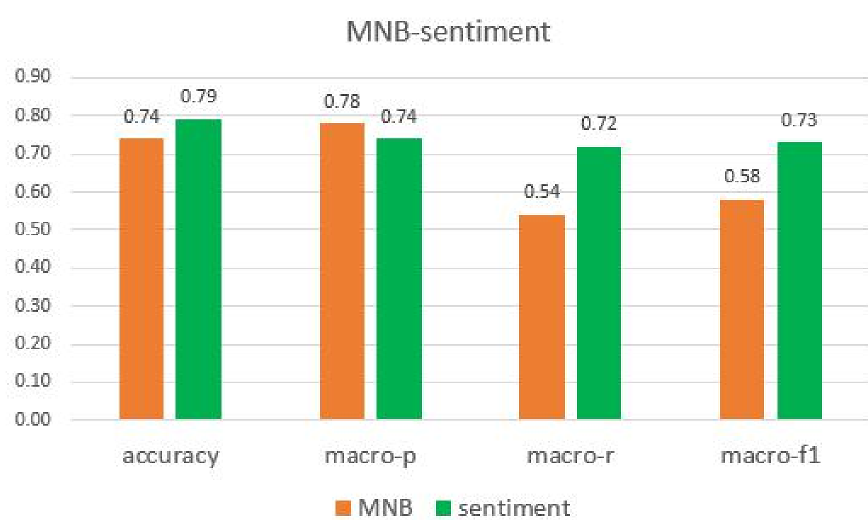
3. Using stemming without applying NLTK stop word

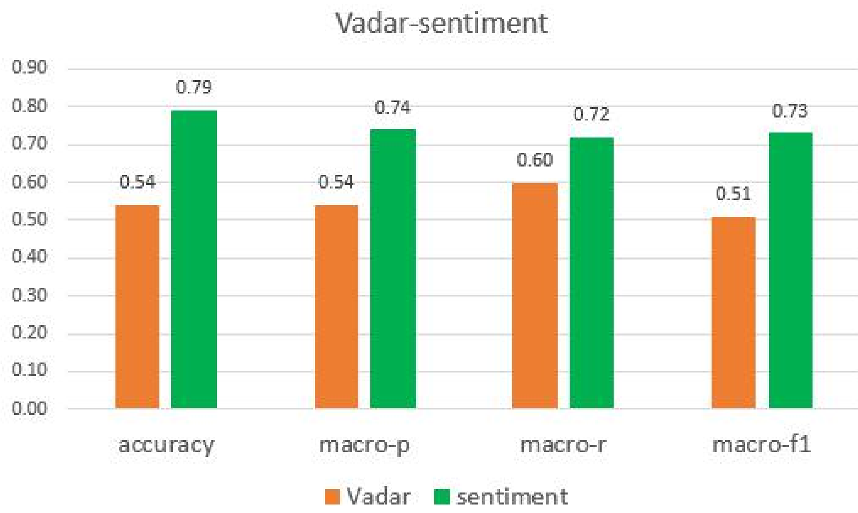
4. Turning all letters into lowercase

The following charts are the comparisons:


DT-sentiment
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.70
0.79
0.63
0.74
0.54
0.72
0.57
0.73
accuracy
macro-p
macro-r
macro-f1
DT
sentiment


BNB-sentiment
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.69
0.79
0.79
0.74
0.44
0.72
0.45
0.73
accuracy
macro-p
macro-r
macro-f1
BNB
sentiment


MNB-sentiment
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.74
0.79
0.78
0.74
0.54
0.72
0.58
0.73
accuracy
macro-p
macro-r
macro-f1
MNB
sentiment



From the charts we can see the relatively better model sentiment exceeds others in almost every metrics except in macro-precision. Nevertheless, no other model is found functioning better the model sentiment.

Thus, I chose this model sentiment to be the best optimization.