

Report for the group of Wei Yu, You Peng and Mohsin Reza

Breakdown of our work:

Both analysis and presentation are done collaboratively by us to let everyone get involved.

1. A.The proportion of negative sentiment is 0.374122220872312; the proportion of neutral sentiment is 0.18738913862228163; the proportion of positive sentiment is 0.4383914469687766.

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

B.

C.

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

| Unnamed: 0 | | OriginalTweet | Sentiment | tokens | | url_removed |
|------------|-------|---|-----------|---|---|-------------|
| 0 | 0 | @MeNyrbie @Phil_Gahan @Chrisitv https://t.co/i... | 1 | [@, MeNyrbie, @, Phil_Gahan, @, Chrisitv, and,... | @MeNyrbie @Phil_Gahan @Chrisitv and and | |
| 1 | 1 | advice Talk to your neighbours family to excha... | 2 | [advice, Talk, to, your, neighbours, family, t... | advice Talk to your neighbours family to excha... | |
| 2 | 2 | Coronavirus Australia: Woolworths to give elde... | 2 | [Coronavirus, Australia, :, Woolworths, to, gi... | Coronavirus Australia: Woolworths to give elde... | |
| 3 | 3 | My food stock is not the only one which is emp... | 2 | [My, food, stock, is, not, the, only, one, whi... | My food stock is not the only one which is emp... | |
| 4 | 4 | Me, ready to go at supermarket during the #COV... | 0 | [Me, ,, ready, to, go, at, supermarket, during... | Me, ready to go at supermarket during the #COV... | |
| ... | ... | ... | ... | ... | ... | |
| 41150 | 41150 | Airline pilots offering to stock supermarket s... | 1 | [Airline, pilots, offering, to, stock, superma... | Airline pilots offering to stock supermarket s... | |
| 41151 | 41151 | Response to complaint not provided citing COVI... | 0 | [Response, to, complaint, not, provided, citin... | Response to complaint not provided citing COVI... | |
| 41152 | 41152 | You know itÄÄÄs getting tough when @KameronWi... | 2 | [You, know, itÄÄÄs, getting, tough, when, @, ... | You know itÄÄÄs getting tough when @KameronWi... | |
| 41153 | 41153 | Is it wrong that the smell of hand sanitizer i... | 1 | [Is, it, wrong, that, the, smell, of, hand, sa... | Is it wrong that the smell of hand sanitizer i... | |
| 41154 | 41154 | @TarttiCat Well new/used Rift S are going for ... | 0 | [@, TarttiCat, Well, new/used, Rift, S, are, g... | @TarttiCat Well new/used Rift S are going for ... | |

41155 rows × 5 columns

| Unnamed: 0 | | OriginalTweet | Sentiment | tokens | | url_removed | lowercase_tokens | tokens_no_punct | stemmed_tokens |
|------------|------|---|-----------|---|---|---|---|---|---|
| 0 | 0 | TRENDING: New Yorkers encounter empty supermar... | 0 | [TRENDING, :, New, Yorkers, encounter, empty, su... | TRENDING: New Yorkers encounter empty supermar... | [trending, :, new, yorkers, encounter, empty, su... | [trending, :, new, yorkers, encounter, empty, su... | [trend, :, new, yorkers, encount, empt, superma... | [trend, :, new, yorkers, encount, empt, superma... |
| 1 | 1 | When i couldn't find hand sanitizer at find Me... | 2 | [When, i, couldn't, find, hand, sanitizer, a... | When i couldn't find hand sanitizer at find Me... | [when, i, couldn't, find, hand, sanitizer, a... | [when, i, couldn't, find, hand, sanitizer, a... | [when, i, couldn't, find, hand, sanitiz, at, fi... | [when, i, couldn't, find, hand, sanitiz, at, fi... |
| 2 | 2 | Find out how you can protect yourself and love... | 2 | [Find, out, how, you, can, protect, yourself, ... | Find out how you can protect yourself and love... | [find, out, how, you, can, protect, yourself, ... | [find, out, how, you, can, protect, yourself, ... | [find, out, how, you, can, protect, yourself, ... | [find, out, how, you, can, protect, yourself, ... |
| 3 | 3 | #Panic buying hits #NewYork City as anxious sh... | 0 | [#, Panic, buying, hits, #, NewYork, City, as... | #Panic buying hits #NewYork City as anxious sh... | [#, panic, buying, hits, #, newyork, city, as... | [#, panic, buying, hits, #, newyork, city, as... | [#, panic, buying, hits, #, newyork, city, as, a... | [#, panic, buy, hit, newyork, city, as, anxio... |
| 4 | 4 | #toiletpaper #dunnypaper #coronavirus #coronav... | 1 | [#, toiletpaper, #, dunnypaper, #, coronavirus... | #toiletpaper #dunnypaper #coronavirus #coronav... | [#, toiletpaper, #, dunnypaper, #, coronavirus... | [#, toiletpaper, #, dunnypaper, #, coronavirus... | [#, toiletpaper, #, dunnypaper, #, coronavirus, ... | [#, toiletpap, #, dunnypap, #, coronavirus, com... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 3793 | 3/93 | Meanwhile in A Supermarket in Israel -- People... | 2 | [Meanwhile, in, A, Supermarket, in, israel, ... | Meanwhile in A Supermarket in Israel -- People... | [meanwhile, in, a, supermarket, in, israel, ... | [meanwhile, in, a, supermarket, in, israel, ... | [meanwhile, in, a, supermarket, in, israel, ... | [meanwhi, in, a, supermarket, in, israel, , p... |
| 3794 | 3/94 | Did you panic buy a lot of non-perishable item... | 0 | [Did, you, panic, buy, a, lot, of, non-perisha... | Did you panic buy a lot of non-perishable item... | [did, you, panic, buy, a, lot, of, non-perisha... | [did, you, panic, buy, a, lot, of, nonperishab... | [did, you, panic, buy, a, lot, of, nonperishab... | [did, you, panic, buy, a, lot, of, nonperish... |
| 3795 | 3/95 | Asst Prof of Economics @conceas was on @NBCPh... | 1 | [Asst, Prof, of, Economics, @, conceas, was, o... | Asst Prof of Economics @conceas was on @NBCPh... | [asst, prof, of, economics, @, conceas, was, o... | [asst, prof, of, economics, @, conceas, was, o... | [asst, prof, of, economics, @, conceas, was, on... | [asst, prof, of, econom, @, conceas, was, on, , rb... |
| 3796 | 3/96 | Gov need to do somethings instead of biar je r... | 0 | [Gov, need, to, do, somethings, instead, of, b... | Gov need to do somethings instead of biar je r... | [gov, need, to, do, somethings, instead, of, b... | [gov, need, to, do, somethings, instead, of, b... | [gov, need, to, do, somethings, instead, of, b... | [gov, need, to, do, someth, instead, of, biar... |
| 3797 | 3/97 | i and @ForestandPaper members are committed to... | 2 | [i, and, @, ForestandPaper, members, are, comm... | i and @ForestandPaper members are committed to... | [i, and, @, forestandpaper, members, are, comm... | [i, and, @, forestandpaper, members, are, comm... | [i, and, @, forestandpaper, members, are, commi... | [i, and, @, forestandpap, member, are, committ... |

3/98 rows × 10 columns

0 [@, menyrbie, @, phil_gahan, @, chrisitv, and,...
1 [advice, talk, to, your, neighbours, family, t...
2 [coronavirus, australia, :, woolworths, to, gi...
3 [my, food, stock, is, not, the, only, one, whi...
4 [me, ,, ready, to, go, at, supermarket, during...
D. Name: lowercase_tokens, dtype: object

```

0      [, menyrbie, , phil_gahan, , chrisitv, and, and]
1      [advice, talk, to, your, neighbours, family, t...
2      [coronavirus, australia, , woolworths, to, giv...
3      [my, food, stock, is, not, the, only, one, whi...
4      [me, , ready, to, go, at, supermarket, during,...

```

Name: tokens_no_punct, dtype: object

```

0      [trending, :, new, yorkers, encounter, empty, ...
1      [when, i, could, n't, find, hand, sanitizer, a...
2      [find, out, how, you, can, protect, yourself, ...
3      [#, panic, buying, hits, #, newyork, city, as,...
4      [#, toiletpaper, #, dunnypaper, #, coronavirus...

```

Name: lowercase_tokens, dtype: object

```

0      [trending, , new, yorkers, encounter, empty, s...
1      [when, i, could, nt, find, hand, sanitizer, at...
2      [find, out, how, you, can, protect, yourself, ...
3      [, panic, buying, hits, , newyork, city, as, a...
4      [, toiletpaper, , dunnypaper, , coronavirus, ,...

```

Name: tokens_no_punct, dtype: object

A scenario I might want to keep some forms of punctuation is when punctuation plays an important role. For example, when people just type “?” without any word text to convey their feeling of surprise.

E. We use Porter stemmer,

After stemming:

```

0      [, menyrbi, , phil_gahan, , chrisitv, and, and]
1      [advic, talk, to, your, neighbour, famili, to,...
2      [coronaviru, australia, , woolworth, to, give,...

```

Name: stemmed_tokens, dtype: object

\

After stemming:

```

0      [trend, , new, yorker, encount, empti, superma...
1      [when, i, could, nt, find, hand, sanit, at, fr...
2      [find, out, how, you, can, protect, yourself, ...

```

Name: stemmed_tokens, dtype: object

```

After removal:
['Unnamed: 0', 'OriginalTweet', 'Sentiment', 'stemmed_tokens']
0      [menyrbi, phil_gahan, chrisitv, and, and]
1      [advic, talk, to, your, neighbour, famili, to,...
2      [coronaviru, australia, woolworth, to, give, e...
3      [my, food, stock, is, not, the, onli, one, whi...
4      [me, readi, to, go, at, supermarket, dure, the...
Name: tokens, dtype: object
Current Columns:
['Unnamed: 0', 'OriginalTweet', 'Sentiment', 'tokens']

```

```

After removal:
['Unnamed: 0', 'OriginalTweet', 'Sentiment', 'stemmed_tokens']
Current Columns:
['Unnamed: 0', 'OriginalTweet', 'Sentiment', 'tokens']

```

```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.

```

F.

```

41150    [airlin, pilot, offer, stock, supermarket, she...
41151    [respons, complaint, not, provid, cite, covidl...
41152    [know, itââ, get, tough, when, kameronwild, ra...
41153    [wrong, smell, hand, sanit, start, turn, coron...
41154    [tartiiicat, well, newus, rift, s, go, 70000, a...
Name: tokens, dtype: object
.

```

```

3793    [meanwhil, supermarket, israel, peopl, danc, s...
3794    [panic, buy, lot, nonperish, item, echo, need,...
3795    [asst, prof, econom, cconc, wa, nbcphiladelphi...
3796    [gov, need, someth, instead, biar, je, rakyat,...
3797    [forestandpap, member, commit, safeti, employe...
Name: tokens, dtype: object

```

G.

The length of the vocabulary is 1000.

H.

Train Accuracy: 0.68234722390961

Test Accuracy: 0.42338072669826227

I. No, because ROC curve is not suitable for labels involvin
g multiple classes

J. Train Accuracy: 0.6654841453043373, which is a bit lower than the train accuracy using count vectors. Test Accuracy: 0.42627698788836227, which is a bit higher than the accuracy using count vectors. They are about the same.

K. Train Accuracy: 0.6654841453043373; Test Accuracy: 0.42627698788836227, lower than the accuracy with stemming.

Bonus:

Naive bayes is a generative model since it is based on the joint probability, $p(x, y)$, of the inputs x and the label y , and make their predictions by using Bayes rules to calculate $p(y | x)$, and then picking the most likely label y .

2.



| | tweet | followers |
|---|---|-----------|
| 0 | Kid at heart, Yes 😊 | 145 |
| 1 | im trolling bro i dont care what so ever cool... | 394 |
| 2 | Veintiocho. \nHagan stream y Shazam.\nl vote f... | 133 |
| 3 | No problem | 160 |
| 4 | if having 6 drops is your reasoning for why ... | 448 |

| | tweet | followers | labels |
|---|---|-----------|--------|
| 0 | Kid at heart, Yes 😊 | 145 | 1 |
| 1 | im trolling bro i dont care what so ever cool... | 394 | 1 |
| 2 | Veintiocho. \nHagan stream y Shazam.\nl vote f... | 133 | 1 |
| 3 | No problem | 160 | 1 |
| 4 | if having 6 drops is your reasoning for why ... | 448 | 1 |

```

289 686 25
<bound method Series.mean of 0      145
1      394
2      133
3      160
4      448
...
995      3
996     181
997    1095
998     393
999     114
Name: followers, Length: 1000, dtype: int64>
\

```

```

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
0      [Kid, at, heart, ,, Yes😊]
1      [im, trolling, bro, i, dont, care, what, so, e...
2      [Veintiocho, ., Hagan, stream, y, Shazam, ., I...
3      [No, problem]
4      [if, having, 6, drops, is, your, reasoning, fo...
Name: tokens, dtype: object

```

```

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
0      [Kid, at, heart, ,, Yes😊]
1      [im, trolling, bro, i, dont, care, what, so, e...
2      [Veintiocho, ., Hagan, stream, y, Shazam, ., I...
3      [No, problem]
4      [if, having, 6, drops, is, your, reasoning, fo...
Name: tokens, dtype: object

```

```

0      [kid, at, heart, ,, yes😊]
1      [im, trolling, bro, i, dont, care, what, so, e...
2      [veintiocho, ., hagan, stream, y, shazam, ., i...
3      [no, problem]
4      [if, having, 6, drops, is, your, reasoning, fo...
Name: lowercase_tokens, dtype: object

```

```

0          [kid, at, heart, , yes]
1  [im, trolling, bro, i, dont, care, what, so, e...
2  [veintiocho, , hagan, stream, y, shazam, , i, ...
3          [no, problem]
4  [if, having, 6, drops, is, your, reasoning, fo...
Name: tokens_no_punct, dtype: object

```

```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

```

```

0          [kid, heart, , yes]
1  [im, trolling, bro, dont, care, so, ever, cool...
2  [veintiocho, , hagan, stream, y, shazam, , vot...
3          [no, problem]
4  [6, drops, reasoning, why, isn, , t, top, 10, ...
Name: tokens_no_sw, dtype: object

```

Before stemming:

```

0          [kid, heart, , yes]
1  [im, trolling, bro, dont, care, so, ever, cool...
2  [veintiocho, , hagan, stream, y, shazam, , vot...
Name: tokens_no_sw, dtype: object

```

After stemming:

```

0          [kid, heart, , ye]
1  [im, troll, bro, dont, care, so, ever, cool, g...
2  [veintiocho, , hagan, stream, y, shazam, , vot...
Name: stemmed_tokens, dtype: object

```

```

[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Package wordnet is already up-to-date!

```

```

0          [kid, heart, , yes]
1  [im, trolling, bro, dont, care, so, ever, cool...
2  [veintiocho, , hagan, stream, y, shazam, , vot...
3          [no, problem]
4  [6, drop, reasoning, why, isn, , t, top, 10, ...
Name: lem_tokens, dtype: object

```

After removal:

```

['tweet', 'followers', 'labels', 'lem_tokens']
0          [kid, heart, yes]
1  [im, trolling, bro, dont, care, so, ever, cool...
2  [veintiocho, hagan, stream, y, shazam, vote, h...
3          [no, problem]
4  [6, drop, reasoning, why, isn, t, top, 10, the...
Name: tokens, dtype: object

```

Current Columns:

```

['tweet', 'followers', 'labels', 'tokens']

```



```

Test accuracy with simple Naive Bayes: 0.69
      precision    recall  f1-score   support

     0         0.00      0.00      0.00         60
     1         0.69      1.00      0.82        138
     2         0.00      0.00      0.00          2

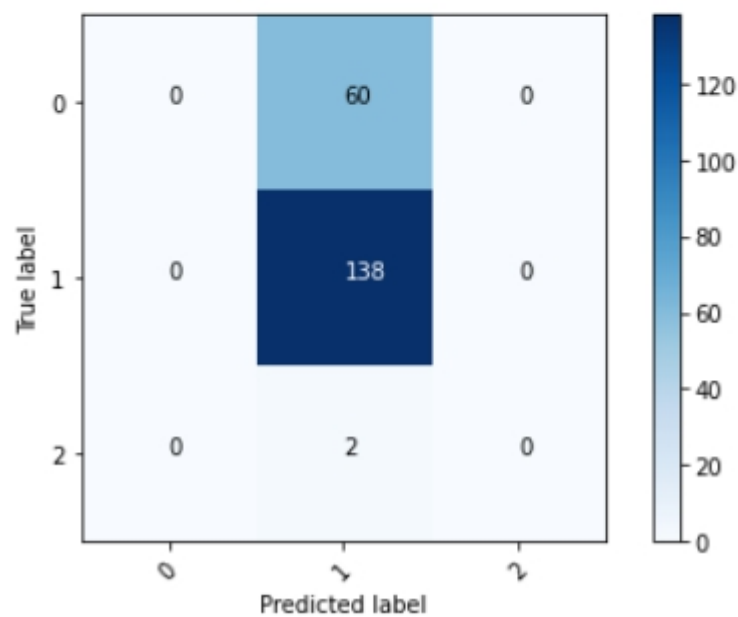
 accuracy          0.69         200
 macro avg         0.23      0.33      0.27         200
 weighted avg         0.48      0.69      0.56         200

```

```

/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:136:
 _warn_prf(average, modifier, msg_start, len(result))

```



The csv file can be downloaded at the end of the ipynb file.