



Are you going to wear a mask?

By Xinpeng Shan(Joy), Rui Miao(Amy),
and Shiyuan Zhou(Eric)

Motivation



Ayesha ManzoorWattoo @ayeshamwattoo · Mar 29

...

Its the only way to keep yourself and your loved ones safe [#WearAMask](#) 😊
[#WashYourHands](#) 🧼 frequently, [#SocialDistancing](#)



David Silvey @DavidSilveyTX · Mar 29

...

Take off the mask.

Go back to work.

Open the economy.

Free the people.

[#Texas](#) [#TakeOffYourMask](#)



Research Questions

Can we infer whether someone is supporting wearing a mask or not by analyzing the NLP of tweets related to masks?



What is the proportion of people who support wearing a mask?



Does individual who refuse to wear masks also refuse to take vaccine? (We assume that there are group of people refuse both)



Data Background

MASK ON!



MASK OFF!

#maskon 350 obs.

3000 obs. **#wearamask**

#wearmask 400 obs.

Attitude:1

#nomasks 400 obs.

#TakeOffYourMask

3000 obs.

#maskfree
40 obs.

Attitude:-1

(We labeled our data depending on the tags of the tweets)

Limitation of Data



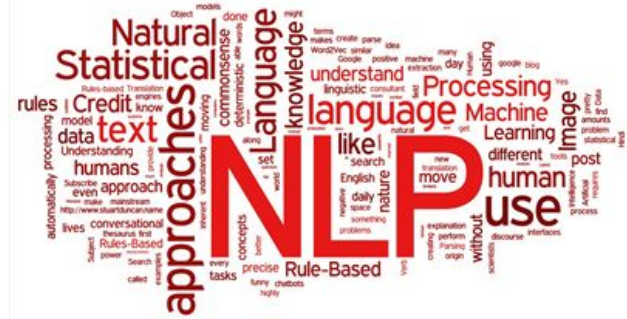
- Relative small size
- Tags may not cover all opinions of the population → overrepresented
- 7-days restriction
- Incorrect collected data

-Removed Duplicate Tweets

-Tokenization

-Removed URLs, tags, punctuations, and stop-words

-Lemmatization & Stemming



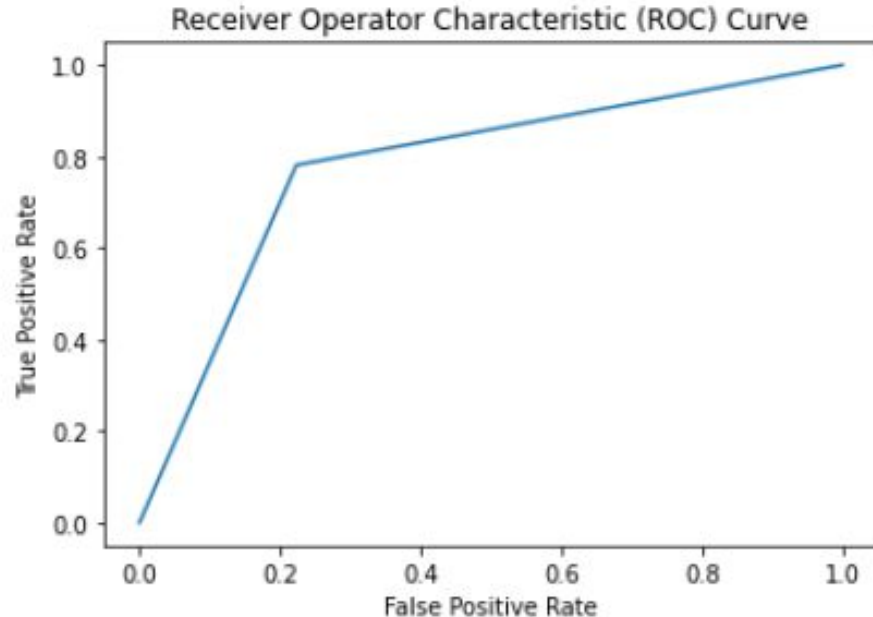
Our Machine Learning Model



- NLP (Nature Language Processing)
- Based on the **Distributional Hypothesis**, measure the similarities by **vectorizing** token collections using TF-IDF and fit our training data to a **Naive Bayes Model** to predict future data.
- Supervised model (with attitudes as labels)

Accuracy of Prediction - Mask Model

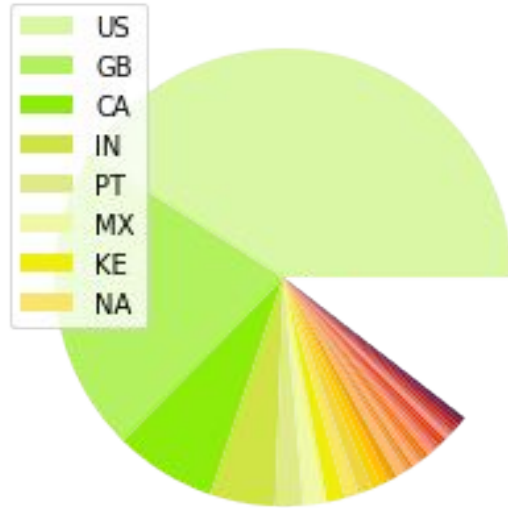
Test accuracy with simple Naive Bayes: 0.7791754018169113



- ROC Curve for measuring the accuracy of the model of the attitudes to wearing masks
- The accuracy is about **0.7792**.
- **Moderate**

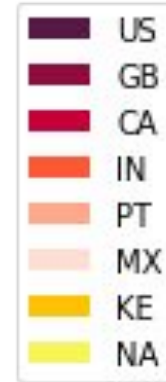
Locations of people support #maskon and #maskoff

#Maskon



- 1. USA 2. India

#MaskOff



- 1. USA 2. UK

There are too many missing location info for each tweets. Thus, the graph may be overrepresented.

#1 Can we infer whether someone is supporting wearing a mask or not by analyzing the NLP of tweets related to masks?

Predicting major media's attitude towards mask

-Picking 3 famous medias:



-Searching by word 'mask'

```
search_with_filter = "mask" + " from:@ABC"
```

-Concatenate the dataset and predict with mask model

```
full_X_new = np.concatenate([X, new_X])
```

-More 1 than -1, the media has a **positive** attitude

More -1 than 1, the media has a **negative** attitude

Equal number of -1 and 1, the media has a **neutral** attitude

Predicting major media's attitude towards mask

Prediction Result

-ABC news



Positive

-NBC news



Positive

-CNN news



Positive

#2 What is the proportion of people who support wearing a mask?

Proportion of People who Support Wearing a Masks

-Extract 2000 tweets mentioned both mask & covid

```
search_words = '(covid.*mask|mask.*covid)|(covid.*mask.*covid)'
```

-Concatenate with mask dataset

```
full_X = np.concatenate([X, part2_X])
```

-Data processing & Predict attitude by mask model

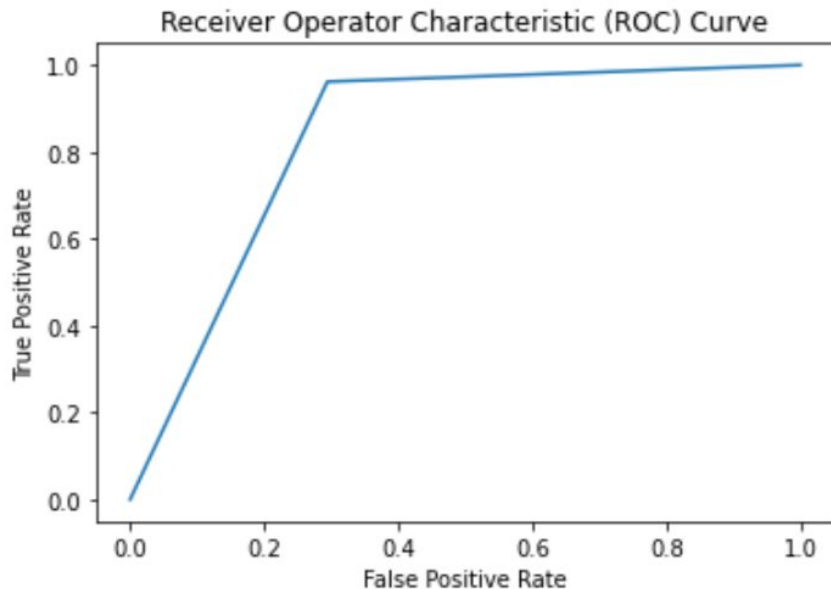
-Reporting the proportion of 1s in column 'attitude'

```
np.sum(part2_pred == 1)/2000
```

-The proportion is 56.75%

#3 Does individual who refuse to wear masks also refuse to take vaccine?

Accuracy of Prediction - Vaccine Model



- ROC curve for measuring the accuracy of the model of **the attitudes to vaccinate**
- The accuracy is about **0.8683**.
- **Close to High**

Cross Prediction - Mask to Vaccine

`full_X = np.concatenate([X, X2])` Concatenating two dataset

-Data processing

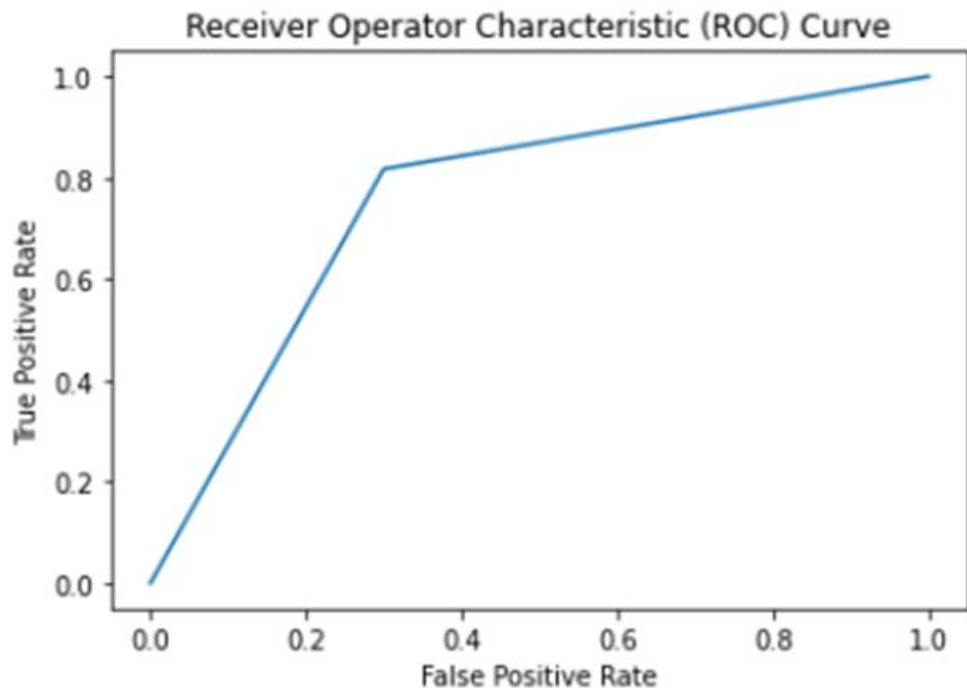
`nb.fit(X, y)` Model fitting with mask dataset

`pred = nb.predict(X2)` Using tokens of vaccine to predict attitudes

`accuracy_score(pred,y2)` Accuracy Score

`fpr, tpr, thresholds = roc_curve(y2,pred, pos_label = 1)` Plotting ROC curve

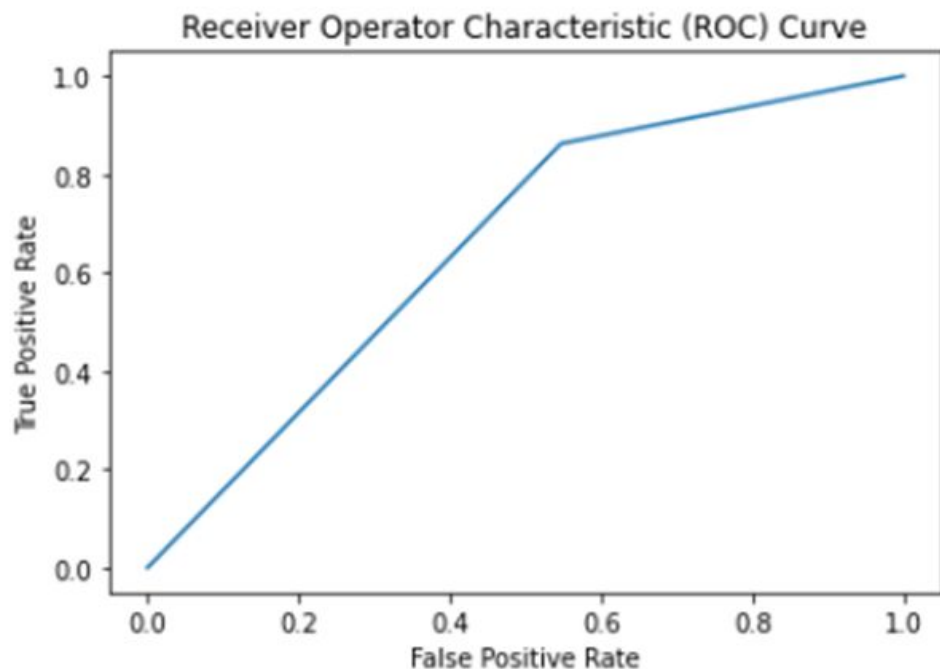
Cross Prediction - Mask to Vaccine



- ROC Curve for measuring the accuracy of **using masks'** model to predict attitude towards vaccines

- The accuracy is about **0.776036**

Cross Prediction – Vaccine to Mask



- ROC Curve for measuring the accuracy of **using vaccines' model to predict attitude towards masks**

- The accuracy is about **0.666760**

Conclusion

-#1 Attitude Prediction

Medias tend to have positive attitudes towards masks.

-#2 Proportion Result

56.75% Support wearing mask

-#3 Cross Prediction

Predictive to some extent

Conclusion

Model	Mask	Vaccine	Mask predict Vaccine	Vaccine predict Mask
Accuracy	0.779175	0.868343	0.776036	0.666760
Strength	Moderate	Close to high	Moderate	Weak (not predictive)

Strengths

- Use both **Lemmatization** and **Stemming**
- No duplicate tweets
- Remove the tags from tweets

Weaknesses for Model

- Data limitations mention before
- Dependence on Twitter data.
- **Randomness** of Tweets.
- Prediction only by wording

Discussion



Vaccine & Mask debate

Discussion



Mask Mandate

Discussion



Thank you