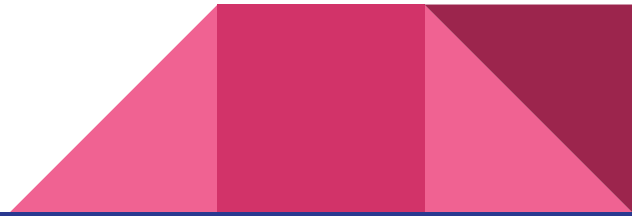
Covid-19 Misinformation Detection

GROUP-3

Richeng Zhang
Tengyue Chen
Zhanfu Yang

Overview

- ❏ Introduction
- ❏ Literature Review
- ❏ Research Question
- ❏ Exploratory Data Analysis
- ❏ Model Comparison
- ❏ Detection Model
- ❏ Result
- ❏ Model Improvement & Other Model
- ❏ Data Sources & Reference



Introduction

- The fake news about COVID19 vaccines
- Covid-19 vaccine is threatened by fake news
- This project tries to suggests a model for detecting fake news about the COVID-19 vaccine on social media.

Literature Review

- Fake news is considered false or misleading information spread in the media (Lazer et al., 2018)
- Fake news detection: content analysis.
- social network and neural network model
- Topic Modeling, XGBoost, LSTM, and BERT
- Challenge: Label

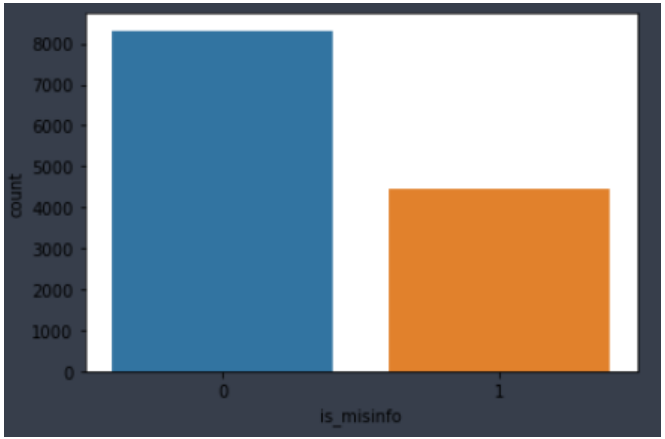
Example lable for COVID-19 vaccine misinformation
Abolish big pharma
antivaccine
Arrest Bill Gates
Between meandry doctor
Big pharma kills
Bill Gates Bio Terrorist
Bill gates evil
The COVID-19 vaccine causes infertility or miscarriages in women.
Natural COVID-19 immunity is better than immunity derived from a COVID-19 vaccine.

Research Question

- Building a model to detect fake news or misinformation about the COVID-19 vaccine on Twitter.
- Helping people to understand whether the tweets are fake COVID-19 vaccine information.

Exploratory Data Analysis

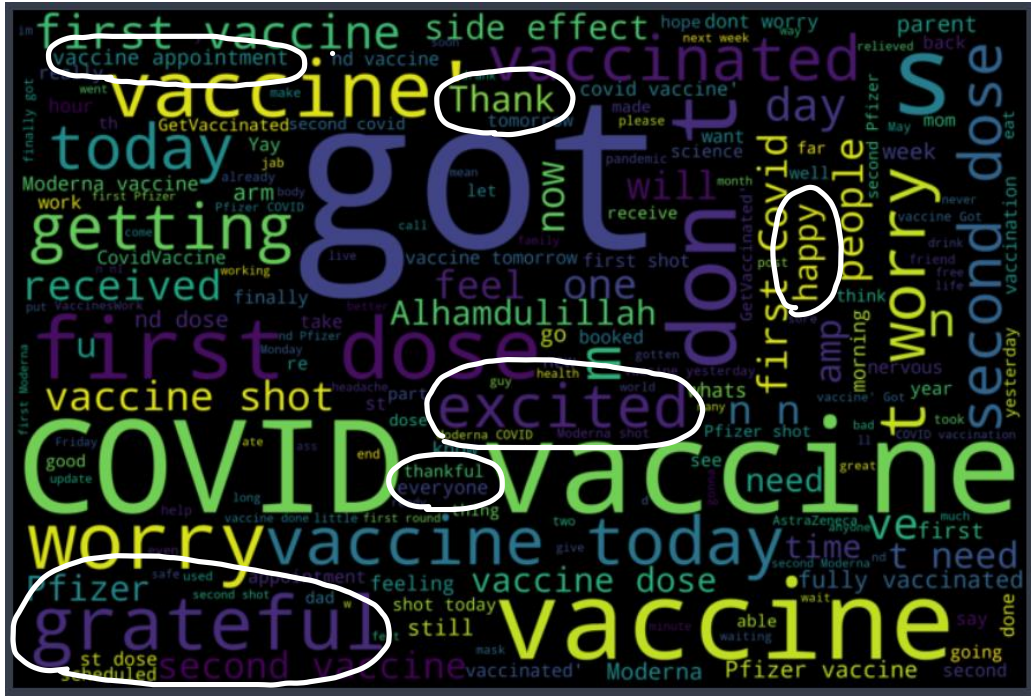
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 12751 entries, 0 to 2483
Data columns (total 2 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   is_misinfo    12751 non-null  int64
 1   text content  12751 non-null  object
dtypes: int64(1), object(1)
```



Is_Misinfo	Number
0	4436
1	8315

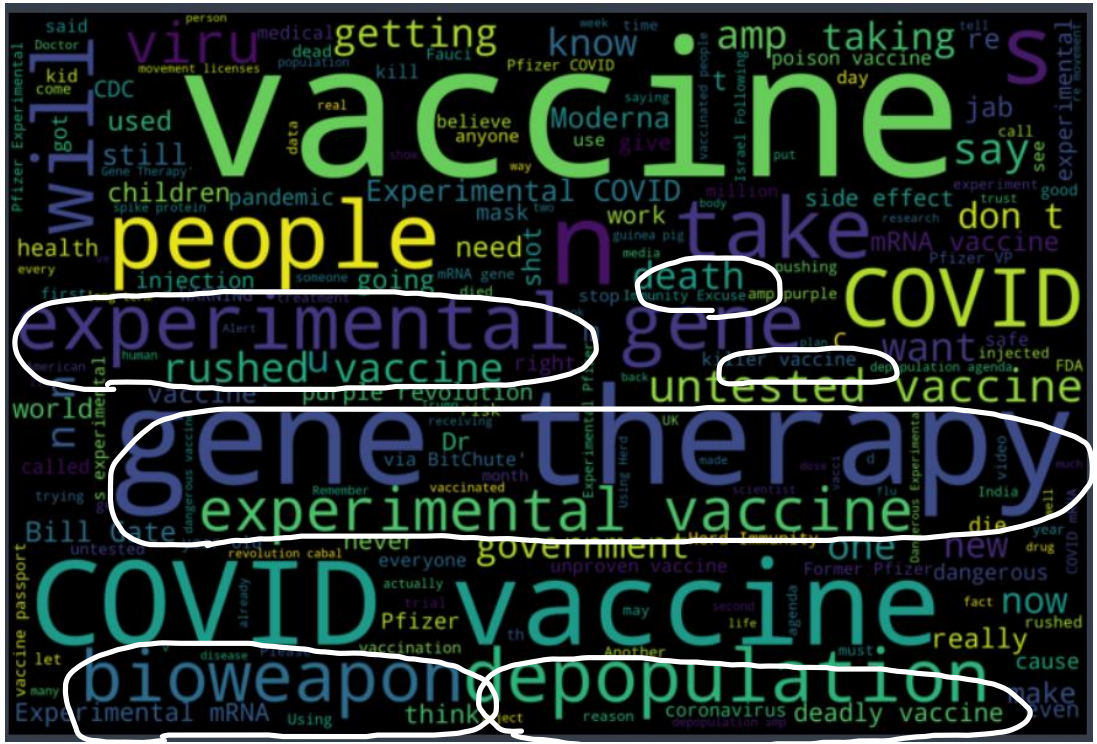
Word Cloud

Word Cloud of is_Misinfo = 0



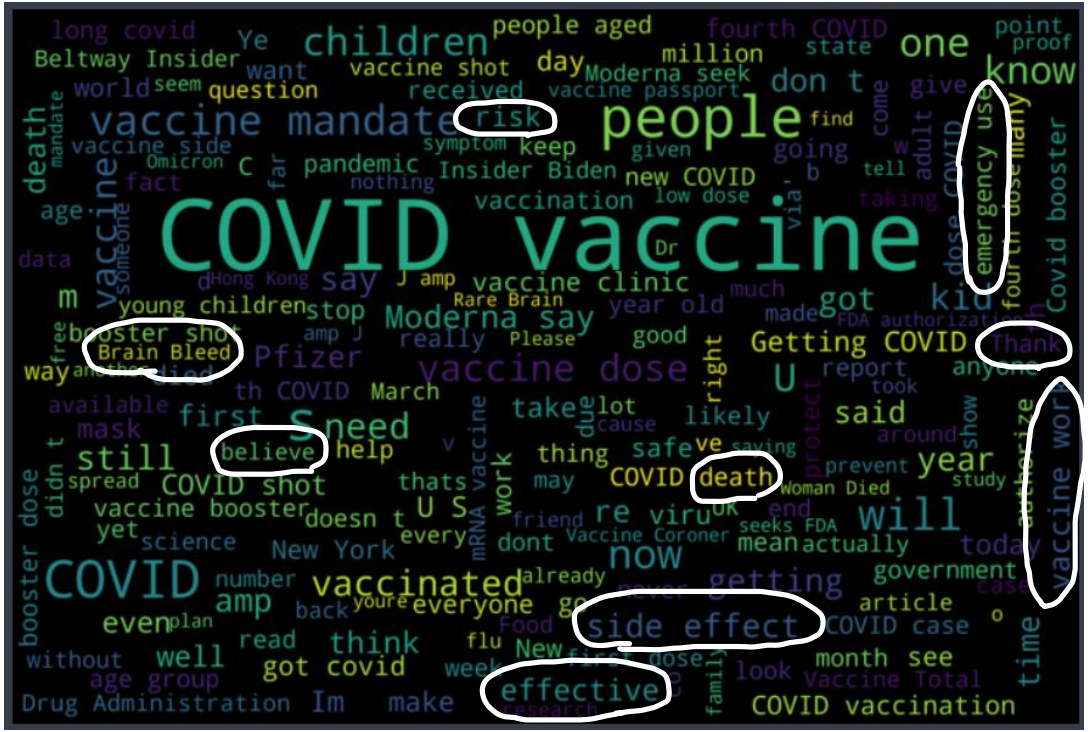
Word Cloud

Word Cloud of is_Misinfo = 1



Word Cloud

Word Cloud of Twitter Data



Model Comparison

K-means

Cosine Distance:

	precision	recall	f1-score	support
0	0.87	0.97	0.92	1627
1	0.94	0.75	0.83	924
accuracy			0.89	2551
macro avg	0.90	0.86	0.88	2551
weighted avg	0.89	0.89	0.89	2551

Naive Bayes

Naive Bayes:

	precision	recall	f1-score	support
0	0.96	0.96	0.96	1718
1	0.91	0.92	0.92	833
accuracy			0.94	2551
macro avg	0.94	0.94	0.94	2551
weighted avg	0.94	0.94	0.94	2551

Model Comparison

Logistic Regression

LogisticRegression()

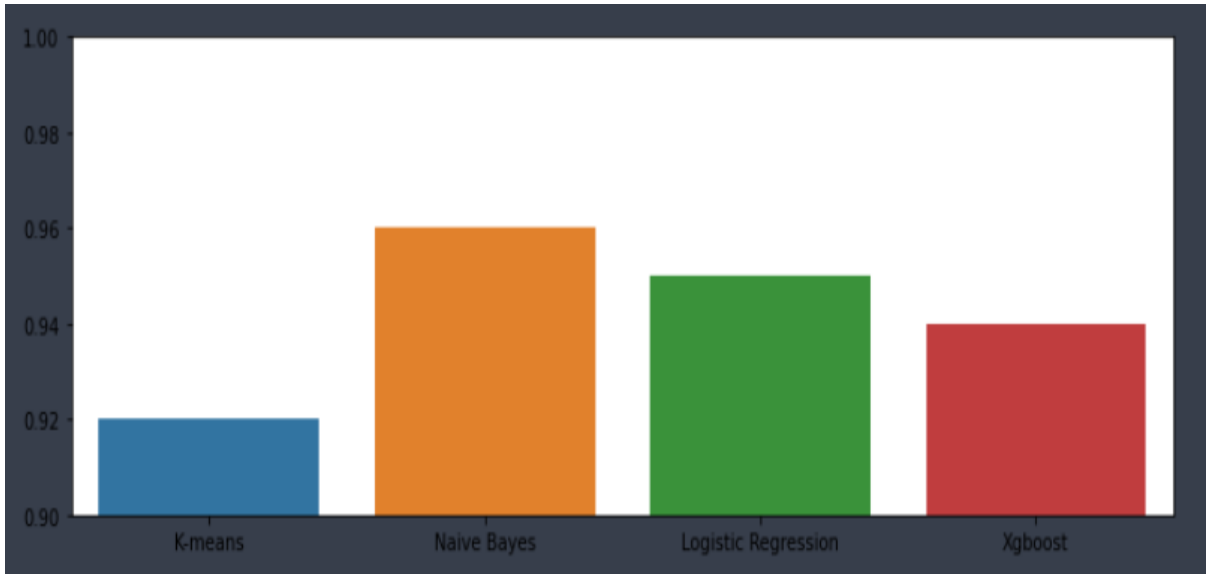
	precision	recall	f1-score	support
0	0.93	0.96	0.95	1691
1	0.92	0.87	0.89	860
accuracy			0.93	2551
macro avg	0.93	0.91	0.92	2551
weighted avg	0.93	0.93	0.93	2551

Xgboost

xgboost:

	precision	recall	f1-score	support
0	0.94	0.94	0.94	1691
1	0.88	0.87	0.88	860
accuracy			0.92	2551
macro avg	0.91	0.91	0.91	2551
weighted avg	0.92	0.92	0.92	2551

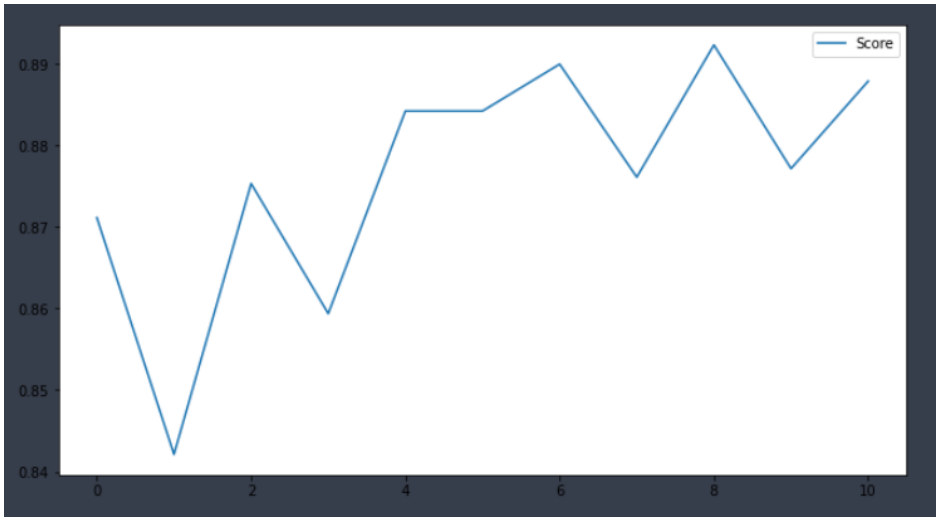
Model Comparison



Detection Model

TF_IDF & KNN Model Training

	Score	k values
0	0.871145	1
1	0.842133	2
2	0.875327	3
3	0.859383	4
4	0.884213	5
5	0.884213	6
6	0.889963	7
7	0.876111	8
8	0.892316	9
9	0.877156	10
10	0.887872	11



Detection Model

	a	Label
0	those blood pressure medication have have deca...	0
1	kettering health doctor work build trust close...	0
2	now the time get boost household member covid ...	0
3	watch dudu sherarami director public health en...	0
4	partnership with the cayuga county health depa...	0
...	...	...
2998	why miralax laxative the covid vaccine have sp...	0
2999	icymi florida republican rep byron donalds joi...	0
3000	file this under shit ive be say since hence th...	0
3001	the news moderna now have billion sign deal fo...	0
3002	injection objection guy hatchard discuss polis...	0

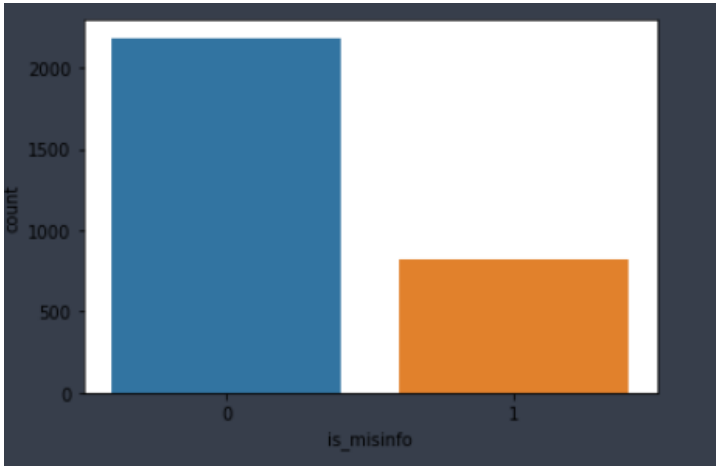
3000 rows × 2 columns

	a	Label
0	those blood pressure medication have have deca...	0
1	kettering health doctor work build trust close...	0
2	now the time get boost household member covid ...	0
3	watch dudu sherarami director public health en...	1
4	partnership with the cayuga county health depa...	0
...	...	...
2998	why miralax laxative the covid vaccine have sp...	0
2999	icymi florida republican rep byron donalds joi...	0
3000	file this under shit ive be say since hence th...	1
3001	the news moderna now have billion sign deal fo...	0
3002	injection objection guy hatchard discuss polis...	0

3000 rows × 2 columns

Result

Is_Misinfo	Number
0	2183
1	817



Result

Naive Bayes

Naive Bayes:

	precision	recall	f1-score	support
0	0.96	0.96	0.96	1718
1	0.91	0.92	0.92	833
accuracy			0.94	2551
macro avg	0.94	0.94	0.94	2551
weighted avg	0.94	0.94	0.94	2551

Naive Bayes:

	precision	recall	f1-score	support
0	0.93	0.94	0.93	2089
1	0.87	0.86	0.86	1062
accuracy			0.91	3151
macro avg	0.90	0.90	0.90	3151
weighted avg	0.91	0.91	0.91	3151

Result

Logistic Regression

LogisticRegression()

	precision	recall	f1-score	support
0	0.93	0.96	0.95	1691
1	0.92	0.87	0.89	860
accuracy			0.93	2551
macro avg	0.93	0.91	0.92	2551
weighted avg	0.93	0.93	0.93	2551

LogisticRegression()

	precision	recall	f1-score	support
0	0.91	0.96	0.94	2089
1	0.92	0.82	0.87	1062
accuracy			0.91	3151
macro avg	0.92	0.89	0.90	3151
weighted avg	0.92	0.91	0.91	3151

Result

Xgboost

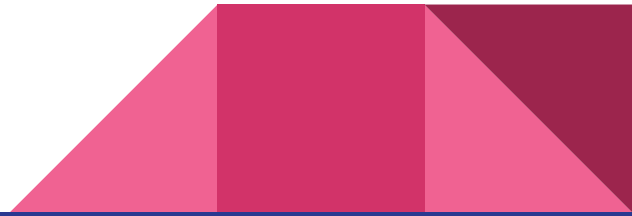
xgboost:

	precision	recall	f1-score	support
0	0.94	0.94	0.94	1691
1	0.88	0.87	0.88	860
accuracy			0.92	2551
macro avg	0.91	0.91	0.91	2551
weighted avg	0.92	0.92	0.92	2551

xgboost:

	precision	recall	f1-score	support
0	0.88	0.96	0.92	2089
1	0.91	0.74	0.82	1062
accuracy			0.89	3151
macro avg	0.89	0.85	0.87	3151
weighted avg	0.89	0.89	0.88	3151

Model Improvement & Other Model



Data Source & Reference

Data Source: https://github.com/SakibShahriar95/ANTiVax

References:

Lazer, David M., et al. “The Science of Fake News.” Science, vol. 359, no. 6380, 2018, pp. 1094–1096., https://doi.org/10.1126/science.aao2998.

“Munich Security Conference.” World Health Organization, World Health Organization, https://www.who.int/director-general/speeches/detail/munich-security-conference.

Marco-Franco, Julio Emilio, et al. “Covid-19, Fake News, and Vaccines: Should Regulation Be Implemented?” International Journal of Environmental Research and Public Health, vol. 18, no. 2, 2021, p. 744., https://doi.org/10.3390/ijerph18020744.

Allcott, Hunt, and Matthew Gentzkow. “Social Media and Fake News in the 2016 Election.” JSTOR, 1 Apr. 2017, http://www.jstor.org/stable/10.2307/44235006?refreqid=searchgateway. Accessed 24 Mar. 2022.

Botei, Mircea. “Misinformation with Fake News.” Bulletin of the Transilvania University of Braşov, Series VII: Social Sciences and Law, no. 2, Oct. 2017, pp. 133–140.

THANK YOU