COVID-19 Prediction via Vaccine Sentiment Analysis on Twitter

Jinyang Han
Georgia Institute of Technology
Atlanta, Georgia, USA
jhan411@gatech.edu

Jingjing Ye
Georgia Institute of Technology
Atlanta, Georgia, USA
jye312@gatech.edu

Qingqing Wu
Georgia Institute of Technology
Atlanta, Georgia, USA
qwu325@gatech.edu

ABSTRACT

As the COVID-19 pandemic spread globally, vaccines have been expected as the ultimate effective mechanism of defense. Issues related to vaccines receive lots of public attention. In this study, we plan to reveal public opinion towards COVID-19 vaccines with Twitter data and how such sentiment influences vaccination and cases/deaths in the US. Our results show the vital importance of vaccines in preventing the spread of pandemics and reducing the increase of cases and deaths. Besides, the public vaccination's willingness will be influenced significantly by sentiments. This conclusion provides the government with a clue that vaccine promotion is a useful defense method.

KEYWORDS

sentiment analysis, vaccination prediction, natural language processing, machine learning

ACM Reference Format:
Jinyang Han, Jingjing Ye, and Qingqing Wu. 2018. COVID-19 Prediction via Vaccine Sentiment Analysis on Twitter. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

There is no doubt that in the past two years, COVID-19 has subverted the previous lifestyle. But fortunately, the whole nation is on the path to being normal because of all adult vaccination programs released in March 2021. And the efficacy of Pfizer vaccine is 95% (95% credible interval, 90.3 - 97.6) for preventing COVID-19 after fully vaccinated, and the results of subgroups defined by sex, age, race and etc behave similarly[22].

As for social media platforms, the active users have skyrocketed in past decades because of the popularity of smartphones and comprehensive network coverage. For example, Facebook had 1.91 billion daily active users (DAU) on average for June 2020[12], while the DAU of Twitter increased 20% and reached 199 million in the first quarter of 2021[25].

People are increasingly inclined to use social media to record emotions and opinions, providing unprecedented rich resources for studying sentiment propagation and epidemic transmission [11] [26]. Moreover, during the COVID-19 outbreak period, people have increased the social media platforms usage and dependence[6] to remain connection. Considering the vital of vaccines and soaring social media usage, it will be useful for future vaccine campaigns and future epidemics' policy-making if public sentiment's influence on vaccination is discovered.

In this research, we plan to use the Twitter data from the Panacea Lab[3], combining data of the number of vaccine intake from the CDC to analyze the effect of public opinion towards vaccines and further predict whether they can affect the vaccination and confirmed cases/deaths.

Specifically, the problem can be formulated as: Given Twitter users' sentiment about vaccines, the detailed numbers of vaccination and cases/death, build a model to estimate the impact of public sentiment on vaccine intake and cases/deaths. Moreover, sentiment impact, vaccination and cases/deaths can be modeled for illustration among three large cities.

2 LITERATURE REVIEW

2.1 Social Media and Epidemics Forecast

Users from various places upload an abundant amount of raw data in form of text, photos, videos and audio on social media. Numerous studies suggest various means to apply these big data sets to the area of public health.

These studies usually focus on the spatio-temporal properties of social media users' sentiment to identify possible disease outbreaks [2] [10]. Using the distribution of total tweet volume, [5] detects a temporal lag of 6–27 days between the rises in the number of COVID-19 related tweets and officially reported deaths in various UK cities.

[24] reveals that the collective wisdom of the crowds at early stages of the pandemic can predict the extent of mortality reflecting the regional severity of the pandemic almost a month later, based on the intensity of initial COVID-19 related tweet attention at the beginning of the pandemic across Italian, Spanish, and United States regions.

2.2 Methods to Analyze Sentiment in Twitter

Sentiment analysis, emotion analysis, topic modeling, and other tools are implemented to explore public sentiment and emotions in Twitter. Lexicon based method, machine learning method or a mix of both methods are usually used in implementing sentiment analysis[10].

Lexicon based method is an unsupervised learning method, which does not require training data and only depends on the dictionary. Words are classified as positive or negative in the polarity lexicons. The occurrences of the terms in the text data are calculated and then transformed into sentiment indicators. This method highly depends on the quality of the lexical resources. The drawbacks are

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/1122445.1122456

also obvious, such as words can have different meanings based on the context, sentiment words may not express any sentiment. [1][7][18]

Machine learning method is a supervise learning method which requires training data. The most common used method in machine learning method is the SVM and Naive Bayes model. Naive Bayes is successful when applied on well-formed text corpus while support vector machine gives a good performance for low shape dataset ([15][10]). Using training data consists of Twitter messages with emoticons obtained through automated means, [14] shows that machine learning algorithms (Naive Bayes, Maximum Entropy, and SVM) have accuracy above 80% when trained with emoticon data.

[8] shows that lexicon and machine learning approaches are similar in accuracy, both achieving higher accuracy when classifying positive sentiment than negative sentiment. The combined approach demonstrates significantly improved performance in classifying positive sentiment.

2.3 Vaccine Sentiment in Twitter

There have been several works analyzing Twitter datasets to reveal vaccine sentiment. [16] analyzes tweets with location information in the US and reveals raising public confidence in vaccines in most states with increasing positive sentiment and decreasing negative sentiment. Besides, critical social/international events (such as clinical trials from Moderna or Pfizer), announcements of political leaders and authorities (such as Donald Trump tweeting "Great News on Vaccines!") and vaccine-adverse conspiracy (such as claim related to Bill Gate that the pandemic is a cover for his plan to implant trackable microchips made by Microsoft) may have potential impacts on public opinion towards COVID-19 vaccines[16][21].

People still take a positive attitude towards vaccination instead of some adversarial effects of some of the vaccines and the emotion of trust regards to vaccine dominates the discussion continually[23][19].

Further, there exists some geospatial difference in public opinion on vaccines since negative sentiments and emotions are more obvious in some states. [16]

However, limited studies have researched the impact of social media such as Twitter on public vaccination behavior using empirical data. Our research hopes to fill this gap to explore whether and how vaccine sentiment on Twitter influences vaccination rate as well as the epidemic.

2.4 Machine Learning Model

Supervised deep learning models will be suitable in this context. More recently, machine learning models have drawn attention and have established themselves as serious contenders to classical statistical models in the forecasting community[4]. Multilayer Perceptron will be used as the predict model.

Assuming adequate data and computing resources, if a strong theoretical understanding of the problem is available, a full numerical model is perhaps the most desirable solution. However, in general, as the complexity of a problem increases, the theoretical understanding decreases (due to ill-defined interactions between systems) and statistical approaches are required. Recently, the use of neural networks, and in particular the multilayer perceptron, has been shown to be effective alternatives to more traditional statistical technique[13].

3 MATERIAL AND METHODOLOGY

3.1 Data Acquisition

Tweets dataset is from Panacea Lab includes tweets with keywords containing "COVID-19" or "vaccine" from December 1, 2020 to October 31, 2021. In this work, we choose the clean file (without re-tweets) for research.

Additionally, content, language as well as location are all stored stratified under metadata "tweet_id". These information is most relevant to our research since we plan to extract sentiment information using content analysis. "Hydrator" will be used to recover the full tweet content and geolocation data based on the tweet IDs.

Hydrating all tweets are very time-consuming, thus we pick a random sample. Specifically, we screened out the English-only posting tweets and randomly pick 10% of the IDs to form the twitter dataset. The Number of daily final selected tweets is plotted in Figure 1. The sample size fluctuates between 5625 and 48006, with an average of 14558. Finally, the library Carmen is used to extract geolocation for tweets. Given a tweet, Carmen will uses both coordinates and other information in a tweet to make geolocation decisions, which greatly increases the number of geolocated tweets over what Twitter provides[9].

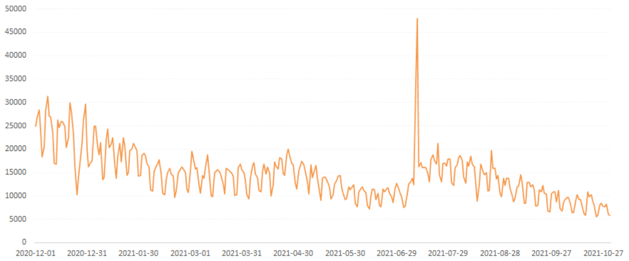


Figure 1: Number of Final Selected Tweets On Each Day

The Figure 2 shows the proportion of tweets samples with geolocation, in which 'Original' represents the percent of geolocated tweets provided by Twitter and 'Carmen' represents that the percent of geolocated tweets provided by Carmen. As can be seen, on average 50.46% of the tweets are tagged with geolocation, which is in contrast with 0.06% in the original Twitter.

After the completion of "Hydrating", the extraction of sentiment features would be done. In order to explore the relationship between the tweets propagation and vaccination more representatively, we choose "retweet counts", "favorites counts" of each tweet, and "followers counts", "friends counts", "user favorites" of each user as our targeted features.

Vaccination and cases/death datasets are from CDC. CDC provides thorough data about vaccines (Trends in Number of COVID-19 Vaccinations) and cases/death (Trends in Number of COVID-19 Cases and Deaths) in the US by nation and by states. Further, we choose three representative cities: New York, Los Angeles, Phoenix as case studies. These three cities are with the top 5 population in

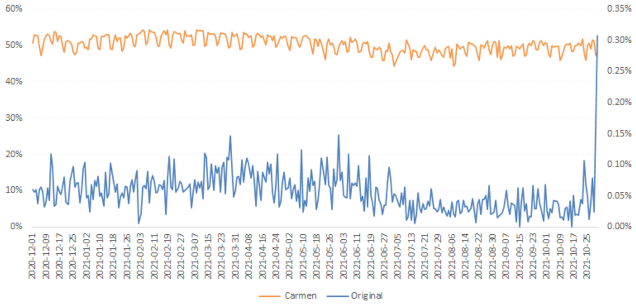


Figure 2: The Proportion of Tweets with Geolocation

the United States according to the census led by the Census Bureau on 2019.

The cities' cases/death and vaccination number by adding the data of the counties. The cases/deaths data by counties is collected from pandemic tracking data system contributed by The New York Times and the county-level vaccination data is obtained from CDC(Vaccinations in the United States, County).

3.2 Sentiment Analysis

As illustrated in the literature review, lexicon method and machine learning method are implemented to analyze text sentiment. In this study, three methods are used to examine the fraction of tweets with negative, neutral and positive sentiments related to vaccine.

3.2.1 Naive Bayes. Naive Bayes is a simple model which works well on text categorization[20]. A Naive Bayes model is trained in this study. Naive Bayes model assumes words position does not matter and relies on a very simple representation of document, that is bag of words. Documents are represented by feature and naive Bayes model assumes the feature probabilities are independent given the class.

Class $c*$ is assigned to tweet d, that is

$$\begin{aligned} c* &= arg \max_c P_{NB}(c|d) \\ &= arg \max_c \frac{P(c)(d|c)}{P(d)} \\ &= arg \max_c P(d|c)P(c) \\ &= arg \max_c P(x_1, \ldots, x_m|c)P(c) \\ &= arg \max_c p(c) \prod_i P(x_i|c), \end{aligned}$$

where x_i represents a feature and there are m features. The parameters can be estimated through maximum likelihood estimate and Laplace (add-1) smoothing is utilized for unseen features, that is

$$\begin{aligned} P(x_i|c) &= \frac{count(x_i, c) + 1}{\sum_{x_j \in V}(count(x_j, c) + 1)} \\ &= \frac{count(x_i, c) + 1}{\sum_{x_j \in V} count(x_j, c) + |V|}. \end{aligned}$$

This study implements two Naive Bayes models trained on two different datasets to classify sentiment in tweets.

- TextBlob
  Users can determine the opinion or emotion that a text holds, and the sentiment function of this software offers users a polarity and subjectivity value after analysis. The polarity value ranges from -1 to 1, where -1 indicates it is a negative statement. TextBlob's default sentiment analysis is trained on customer reviews hand-tagged with values for polarity and subjectivity. Another option in TextBlob is NaiveBayesAnalyzer, which is trained on movie reviews associated with positive or negative rating scores.
- Classifier Trained on the Sentiment140 dataset
  Sentiment140[14] dataset is widely used in analysing the sentiment in tweets. The dataset contains about 1.6 million tweets collected through keyword search and annotated automatically by detecting emoticons. Tweets are determined to have positive, neutral, or negative sentiment.
  A naive Bayes classifier is trained on this dataset and implemented to classify the sentiment in tweets related to vaccine as either positive or negative.

3.2.2 VADER. VADER (Valence Aware Dictionary and sentiment Reasoner) is a lexicon and rule-based feeling analysis instrument that is explicitly sensitive to suppositions communicated in web-based media [17]. It is trained by asking and paying people to score a very big list of words. VADER utilizes a mix of lexical highlights that are, for the most part, marked by their semantic direction as one or the other positive or negative. Thus, VADER not only tells about the Polarity score yet, in addition, it tells us concerning how positive or negative a conclusion is.

3.3 Machine Learning Model

MLPs are feed forward neural networks which are typically composed of several layers of nodes with unidirectional connections, often trained by back propagation[27]. The learning process of MLP network is based on the data samples composed of the N-dimensional input vector x and the M-dimensional desired output vector d, called destination. By processing the input vector x, the MLP produces the output signal vector $y(x, w)$ where w is the vector of adapted weights. The error signal produced actuates a control mechanism of the learning algorithm. The corrective adjustments are designed to make the output signal $y_k(k = 1, 2, , M)$ to the desired response dk in a step by step manner.

Gradient algorithm to get minimization. Adaptation of weights is performed step by step in gradient algorithm. $p(k$ is the direction of minimization in kth step, is the learning coefficient, and w is the adaptation coefficien

The learning algorithm of MLP is based on the minimization of the error function defined on the learning set (x_i, d_i) for $i = 1, 2, , N$ using the Euclidean norm:

$$E(w) = \frac{1}{2} \sum_{i=1}^{N} \|y(x_i, w) - d_i\|^2 .$$

Adaptation of weights is performed step by step

$$w(k + 1) = w(k) + \gamma p(k),$$

where $p(k)$ is the direction of minimization in kth step,γ is the learning coefficient, and w is the adaptation coefficient.

Most effective is the Levenberg–Marquard algorithm for medium size networks and conjugate gradient for large size networks.

Levenberg–Marquard algorithm

Least square formulation of learning problem is exploited:

$$E(w) = \frac{1}{2}\sum_{i=1}^{M}(y_i(w) - d_i)^2 .$$

Solved by using second order method of Newton type:

$$p(k) = -G(k)^{-1}g(k),$$

where $g(k) = \frac{\partial E}{\partial w(k)}$ is the gradient of error function Eq. $G(k)$ is the Hessian approximation, determined by applying the Jacobian matrix $J(k)$:

$$G(k) = J(k)^T J(k) + v.$$

In this equation the Jacobian matrix J is equal

$$J = \frac{\partial e}{\partial w}e = [y_i(w) - d_i, \ldots, y_M(w) - d_M]^T .$$

Conjugate gradient

Direction $p(k)$ is evaluated according to the formula.

$$p(k) = -g(k) + \beta p(k-1),$$

where the conjugate coefficient β is usually determined according to the Polak-Ribiere rule:

$$p(k) = \frac{g(k)^T(g(k) - g(k-1))}{g(k-1)^T g(k-1)}.$$

4 EXPERIMENTS AND RESULTS

4.1 Multilayer Perceptron

As shown in Figure 3, the network layer and parameters are shown as follows:

- Split dataset: test size is 0.3.
- MLP: Set three hidden layer and use Rectified Linear Unit (ReLU) as the activation function.
- Dataloader: batchsize: 64
- Optimizer: Use Adam as the optimize and use MSELoss to calculate loss.
- Learning rate: 0.01.
- Epoch: 1000.

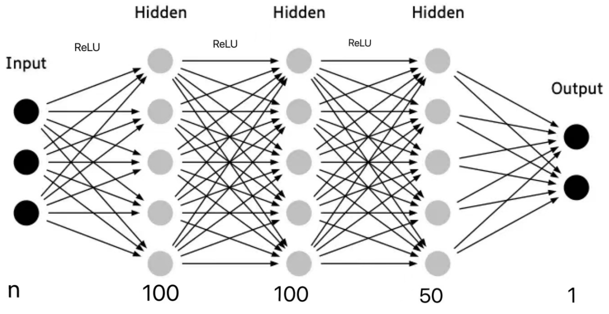


Figure 3: MLP

Table 1: Correlation of Vaccine Sentiment of Three Methods

correlation	sentiment140	textblob	vader	n
sentiment140	1	-0.3873	0.4196	0.4711
textblob	-0.3873	1	-0.0785	-0.2751
vader	0.4196	-0.0785	1	0.1405
n	0.4711	-0.2751	0.1405	1

4.2 Twitter Vaccine Sentiment and Vaccination

4.2.1 Vaccine Sentiment. The vaccine sentiment is based on tweets that are assumed to be posted by users in the United States, of which the geolocation are estimated by Carmen. Figure 4 shows an example of daily vaccine sentiment index from 20210411 to 20210611. In the figure, line "n" is the number of tweets related to vaccine in the random sample in each day.

The vaccine sentiment is estimated by the proportion of positive sentiment in vaccine tweets , that is

$$S_t = \frac{n_{p,t}}{n_t},$$

where S_t is vaccine sentiment on day t, $n_{p,t}$ is the number of tweets with positive sentiment on day t, and n_t is the number of tweets related to vaccine on day t.

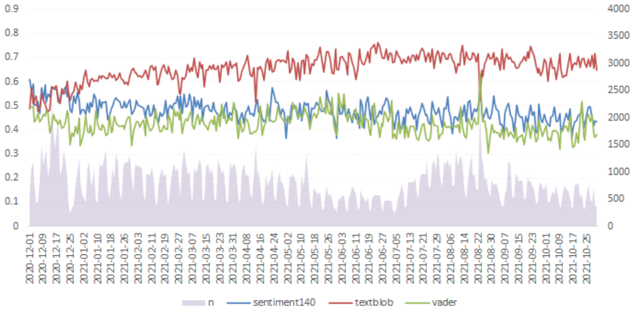


Figure 4: Daily Sentiment Index

The correlation of vaccine sentiment index estimated via three methods is shown in Table 1. The levels of sentiment estimated by sentiment140 and VADER show consistency with correlation of 0.4196. Also, the number of vaccine tweets can also reflect vaccine sentiment to some extent, which can be explained from the aspect of public attention.

As sentiment140 and VADER show significant correlation and textblob exist negative correlation with other estimates, it is quite likely that sentiment140 and VADER are more reasonable and believable than textblob in vaccine sentiment estimate.

Since the sample tweets are selected randomly, to avoid bias from the samples, we randomly picked 5% tweets from the total datasets and compare the sentiment estimate with that of 10% sample. As in Table 2, the average value of the sentiment values are close to each other and there exists high correlation between the estimated. Thus, our sample of 10% can be representative and the results based these set of data are robust.

Table 2: Vaccine Sentiment Comparison of Two Samples

feature	5%	10%	Pearson	Spearman
sentiment140	0.4805	0.4798	0.7296	0.7060
textblob	0.6588	0.6553	0.8608	0.8067
vader	0.4254	0.4240	0.7329	0.7232
n	394.5493	791.0448	0.9907	0.9893

Table 3: Statistical Data of Features

feature	retweet	favorite	followers	friends	user favorite
mean	2.82	10.37	87941.68	2007.45	24765.38
median	2.69	9.66	86321.76	2031.85	24527.52

4.2.2 Twitter Feature. Table 3 shows the features' statistical data of whole research period. The correlation of new cases & new death shows with retweets & favorites is shown in Figure 5 and 6. From these, we can see that this summer, the pandemic has slowed down significantly. Although the discussion on Twitter has become lower, it is still maintained at a high level as before.

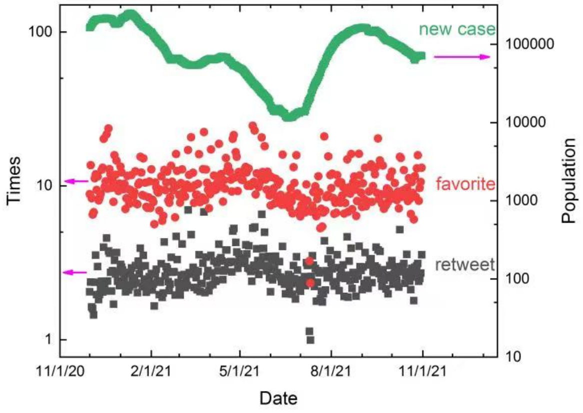


Figure 5: The Correlation among Retweet, Favorite and New Cases

4.2.3 Twitter and Vaccination.

- Correlation Analysis From the correlation analysis in Figure 7a, cumulative vaccination in the U.S. is negatively related to vaccine sentiment, which may be caused by the monotonically increasing vaccination rate. However, the daily vaccination is positively related to vaccine sentiment. Beside, the feature 'followers' is also positively related to the daily vaccination, which implies that tweets posted by more influential users can cause more daily vaccination.
- Predication Displayed in Figure 8, the prediction value fits in well with the vaccination data.

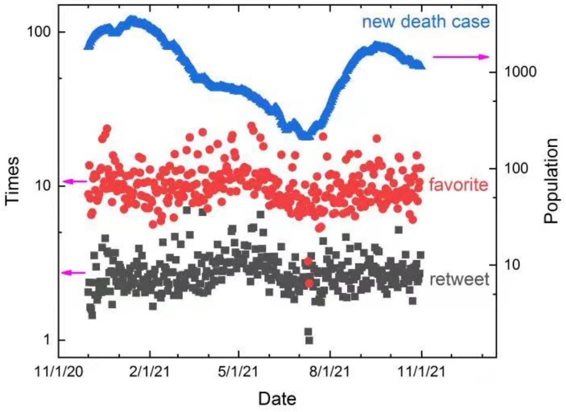


Figure 6: The Correlation among Retweet, Favorite and New Deaths

Table 4: Loss of Prediction Model for Vaccination

Types	Percentage	Sentiment140	textblob	VADER
Cumulative	10%	$2.38*10^{15}$	$2.01*10^{15}$	$2.66*10^{15}$
	5%	$2.74*10^{15}$	$2.15*10^{15}$	$2.89*10^{15}$
Daily	10%	$1.25*10^{11}$	$1.68*10^{11}$	$1.75*10^{11}$
	5%	$1.30*10^{11}$	$2.01*10^{11}$	$8.02*10^{10}$

Table 5: Loss of Prediction Model for Cases/Deaths

Type	vaccine VS. cases	vaccine VS. death
Loss	209213223.5022	8627.0933

Further, loss data in Table 4 shows that Sentiment140 is more useful to reflect vaccine sentiment in Twitter with overall lower prediction loss, especially for daily vaccination.

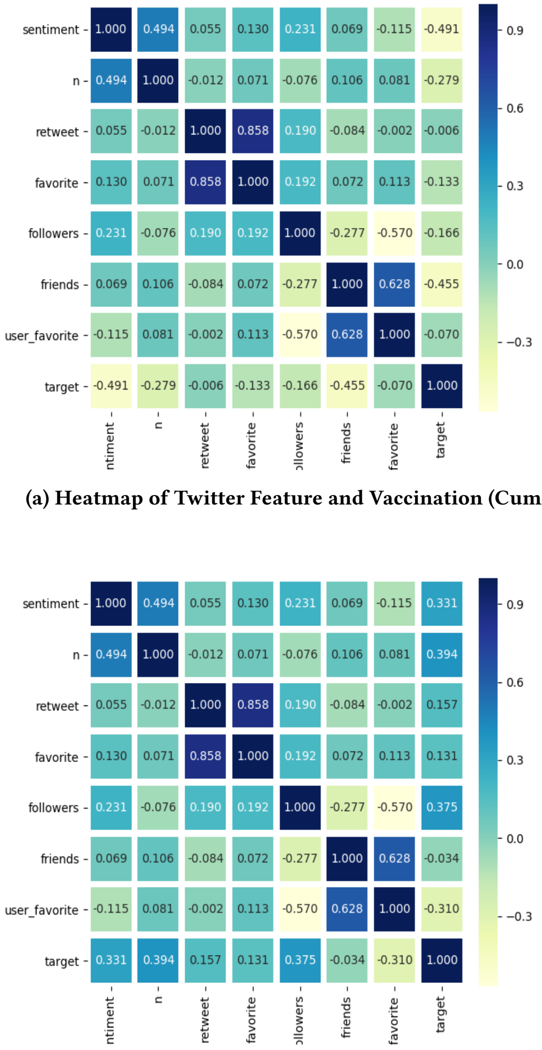
4.3 Vaccination and Cases/Deaths

- Correlation Analysis As shown in Figure 9, the cumulative number of people receiving 1 or more doses is negatively related to cases and deaths. Besides, the negative correlation is more significant for deaths than cases. It illustrates that vaccination plays more role in preventing deaths.
- As in Figure 10, the prediction value based on vaccination fits in well with cases and deaths.

4.4 Cases Studies

Besides the national analysis, we also use the data of three large cities to take the cases studies.

- Correlation Analysis Results for Los Angeles are shown in the report and the same trend can be observed in New York and Phoenix.



(a) Heatmap of Twitter Feature and Vaccination (Cumulative)

(b) Heatmap of Twitter Feature and Vaccination (Daily)

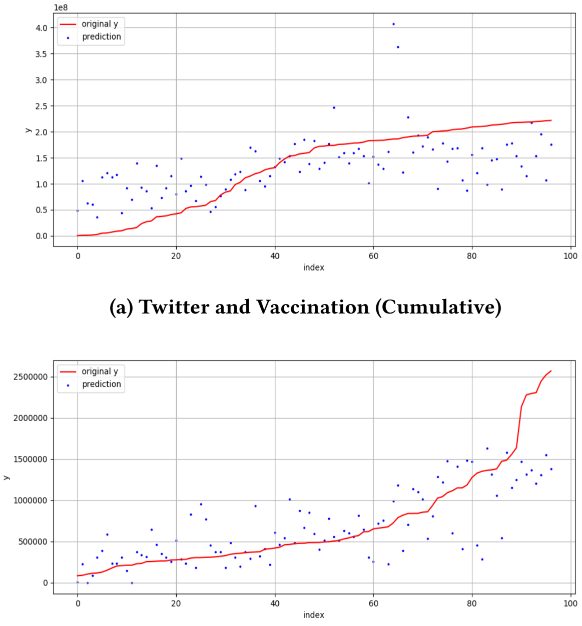
Figure 7: Heatmap of Twitter Feature and Vaccination

From Figure 11, we can see the vaccinations of all age-ranges people all have negative relationships with cases and deaths in Los Angeles. This negative correlation is increasingly significant with increasing age. It indicates that the vaccination coverage in aged people is of more significance to control the pandemic.

Furthermore, the negative correlation is more significant for deaths than cases. It implies that the vaccination is more effective in reducing the death rate than preventing infection.

- Since the number of tweets in specific cities is limited, the tweets location in each state of the city are filtered to extract vaccine sentiment. Figure 12 shows that only models based on vaccine sentiment can not predict vaccination well. Compared with results in chapter 4.1.3, others features in Twitter can improve the power of the models to a large extent.

  As in Figure 13, the model based on vaccination can predict the cases and deaths in Los Angeles well.



(a) Twitter and Vaccination (Cumulative)

(b) Twitter and Vaccination (Daily)

Figure 8: Twitter and Vaccination (Sentiment140)

Table 6: Loss Data of Prediction Model for Cases/Deaths in Three Cities

Type	vaccine VS. cases	vaccine VS. death
Los Angeles	417785357.0844	123795.5258
New York	363116316.1600	119730.9026
Phoenix	390530621.7244	125920.2472

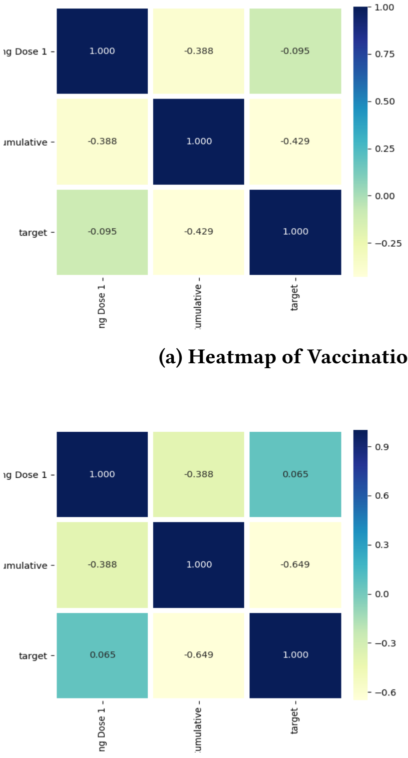
5 CONCLUSION AND DISCUSSION

Firstly, vaccine sentiment extracted from Twitter can positively predict daily vaccination.

As for the effect of vaccination, the vaccination can positively predict both cases and deaths, which implies that improving vaccination rate is a reasonable way to tackle COVID-19. Further, vaccine is more effective in preventing death than reducing the risk of infection.

Results from cases study in three large cities provide additional information. Increasing vaccination rate in aged people is of more importance in controlling the pandemic both in term of cases and deaths.

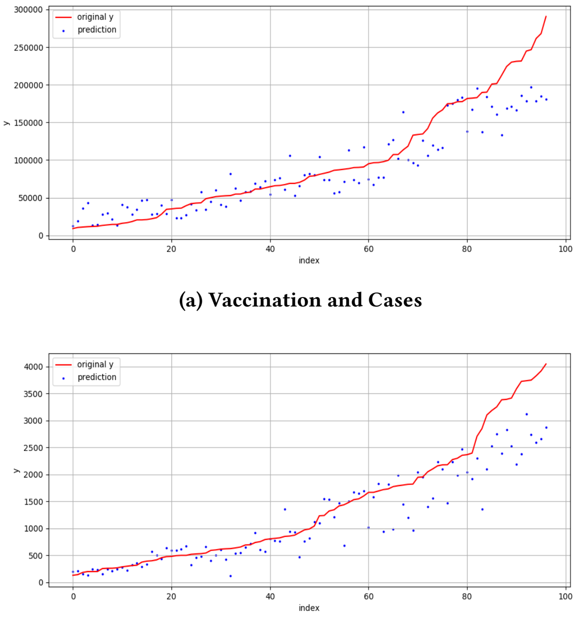
From our research result, we can see for either national or regional levels, vaccinations have a significant effect in preventing



(a) Heatmap of Vaccination and Cases

(b) Heatmap of Vaccination and Deaths

Figure 9: Heatmap of Vaccination and Cases/Deaths



(a) Vaccination and Cases

(b) Vaccination and Deaths

Figure 10: Vaccination and Cases/Deaths

the spread of pandemics and reducing the increase of cases and deaths. And the public sentiment also has an effect on the vaccination's willingness. Therefore, the government should use all feasible methods (e.g. make full use of social media) to dispel public doubts and encourage vaccination to increase the vaccination rate, In this way, it is possible to prevent and control the epidemic effectively.

REFERENCES

[1] Sanjida Akter and Muhammad Tareq Aziz. 2016. Sentiment analysis on facebook group using lexicon based approach. In *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*. IEEE, 1–4.

[2] Kashif Ali, Hai Dong, Athman Bouguettaya, Abdelkarim Erradi, and Rachid Hadjidj. 2017. Sentiment analysis as a service: a social media based sentiment analysis framework. In *2017 IEEE International Conference on Web Services (ICWS)*. IEEE, 660–667.

[3] Juan M. Banda, Ramya Tekumalla, Guanyu Wang, Jingyuan Yu, Tuo Liu, Yuning Ding, Ekaterina Artemova, Elena Tutubalina, and Gerardo Chowell. 2021. A Large-Scale COVID-19 Twitter Chatter Dataset for Open Scientific Research—An International Collaboration. *Epidemiologia* 2 (8 2021). Issue 3. https://doi.org/10.3390/epidemiologia2030024

[4] Bontempi, Gianluca Ben Taieb, Souhaib Le Borgne, and Yann. 2013. Machine Learning Strategies for Time Series Forecasting. (2013), 62–77. https://doi.org/10.1007/978-3-642-36318-4_3

[5] I Cheng, Johannes Heyl, Nisha Lad, Gabriel Facini, and Zara Grout. 2021. Evaluation of Twitter data for an emerging crisis: an application to the first wave of COVID-19 in the UK. *Scientific Reports* 11, 1 (2021), 1–13.

[6] Wong AHo SOlusanya OAntonini MLyness D. 2021. The use of social media and online communications in times of pandemic COVID-19. , 255-260 pages. Issue 3. https://doi.org/10.1177/1751143720966280

[7] Bijoyan Das and Sarit Chakraborty. 2018. An improved text sentiment classification model using TF-IDF and next word negation. *arXiv preprint arXiv:1806.06407* (2018).

[8] Chedia Dhaoui, Cynthia M Webster, and Lay Peng Tan. 2017. Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing* (2017).

[9] Mark Dredze, Michael J Paul, Shane Bergsma, and Hieu Tran. 2013. Carmen: A Twitter Geolocation System with Applications to Public Health. In *AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*.

[10] Zulfadzli Drus and Haliyana Khalid. 2019. Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science* 161 (2019), 707–714.

[11] Erhu Du, Eddie Chen, Ji Liu, and Chunmiao Zheng. 2021. How do social media and individual behaviors affect epidemic transmission and control? *Science of the Total Environment* 761 (3 2021). https://doi.org/10.1016/j.scitotenv.2020.144114

[12] Facebook. 2021. Facebook Reports Second Quarter 2021 Results. (2021). https://s21.q4cdn.com/399680738/files/doc_news/Facebook-Reports-Second-Quarter-2021-Results-2021.pdf

[13] M.W Gardner and S.R Dorling. 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment* 32, 14 (1998), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0

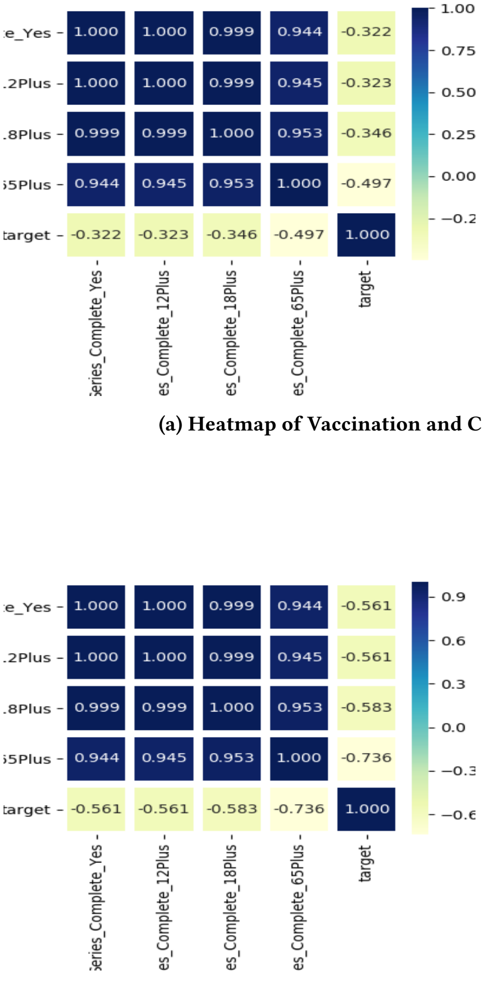
[14] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford* 1, 12 (2009), 2009.

[15] Anees Ul Hassan, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq, and Sungyoung Lee. 2017. Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 138–140.

[16] Tao Hu, Siqin Wang, Wei Luo, Mengxi Zhang, Xiao Huang, Yingwei Yan, Regina Liu, Kelly Ly, Viraj Kacker, Bing She, and Zhenlong Li. 2021. Revealing public opinion towards COVID-19 vaccines with Twitter Data in the United States: a spatiotemporal perspective. *medRxiv* (2021). https://doi.org/10.1101/2021.06.02.21258233

[17] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8.

[18] Muhammad Taimoor Khan, Mehr Durrani, Armughan Ali, Irum Inayat, Shehzad Khalid, and Kamran Habib Khan. 2016. Sentiment analysis and the complex



(a) Heatmap of Vaccination and Cases

(b) Heatmap of Vaccination and Deaths

Figure 11: Heatmap of Vaccination and Cases/Deaths (Los Angeles)

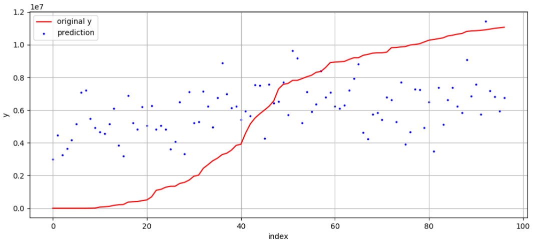
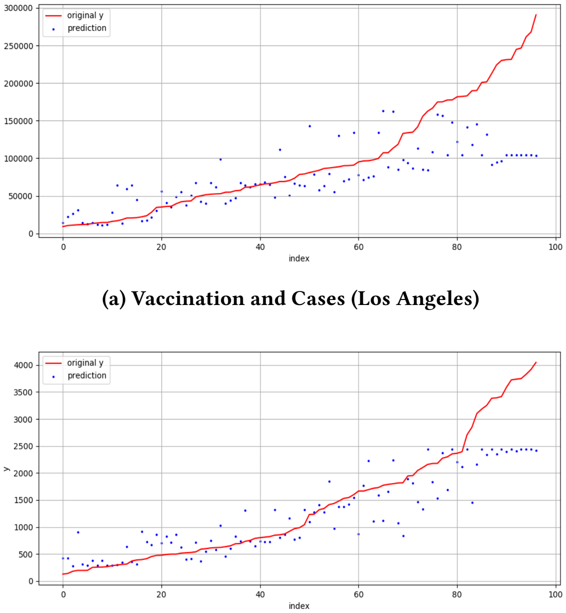


Figure 12: Sentiment and Vaccination



(a) Vaccination and Cases (Los Angeles)

(b) Vaccination and Deaths (Los Angeles)

Figure 13: Vaccination and Cases/Deaths (Los Angeles)

natural language. *Complex Adaptive Systems Modeling* 4, 1 (2016), 1–19.

[19] Joanne Chen Lyu, Eileen Le Han, and Garving K Luli. 2021. COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *Journal of Medical Internet Research* 23 (6 2021). Issue 6. https://doi.org/10.2196/24435

[20] Christopher Manning and Hinrich Schutze. 1999. *Foundations of statistical natural language processing.* MIT press.

[21] Pooria Taghizadeh Naderi, Ali Asgary, Jude Kong, Jianhong Wu, and Fattaneh Taghiyareh. 2021. COVID-19 Vaccine Hesitancy and Information Diffusion: An Agent-based Modeling Approach. (9 2021).

[22] Fernando P. Polack, Stephen J. Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L. Perez, Gonzalo Pérez Marc, Edson D. Moreira, Cristiano Zerbini, Ruth Bailey, Kena A. Swanson, Satrajit Roychoudhury, Kenneth Koury, Ping Li, Warren V. Kalina, David Cooper, Robert W. Frenck, Laura L. Hammitt, Özlem Türeci, Haylene Nell, Axel Schaefer, Serhat Ünal, Dina B. Tresnan, Susan Mather, Philip R. Dormitzer, Uğur Şahin, Kathrin U. Jansen, and William C. Gruber. 2020. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine* 383, 27 (2020), 2603–2615. https://doi.org/10.1056/NEJMoa2034577 arXiv:https://doi.org/10.1056/NEJMoa2034577 PMID: 33301246.

[23] Naw Safrin Sattar and Shaikh Arifuzzaman. 2021. COVID-19 Vaccination Awareness and Aftermath: Public Sentiment Analysis on Twitter Data and Vaccinated Population Prediction in the USA. *Applied Sciences* 11 (6 2021). Issue 13. https://doi.org/10.3390/app11136128

[24] Jeremy Turiel, Delmiro Fernandez-Reyes, and Tomaso Aste. 2021. Wisdom of crowds detects covid-19 severity ahead of officially available data. *Scientific Reports* 11, 1 (2021), 1–9.

[25] Twitter. 2021. Q1 2021 Letter to Shareholders. (2021). https://s22.q4cdn.com/826641620/files/doc_financials/2021/q1/Q1'21-Shareholder-Letter.pdf

[26] Samira Yousefinaghani, Rozita Dara, Zvonimir Poljak, Theresa M. Bernardo, and Shayan Sharif. 2019. The Assessment of Twitter's Potential for Outbreak Detection: Avian Influenza Case Study. *Scientific Reports* 9 (12 2019). Issue 1. https://doi.org/10.1038/s41598-019-54388-4

[27] E.A. Zanaty. 2012. Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification. *Egyptian Informatics Journal* 13, 3 (2012), 177–183. https://doi.org/10.1016/j.eij.2012.08.002