International Journal of Informatics and Information Systems ISSN 2579-7069
Vol. 5, No. 1, January 2022, pp. 56-68

COVID-19 Vaccination: A Retrospective Observation and Sentiment Analysis of the Twitter Social Media Platform in Indonesia

Andhika Rafi Hananto [1,*], Silvia Anggun Rahayu [2], and Taqwa Hariguna [3]

Universitas Amikom Purwokerto, Indonesia
andhikarh90@gmail.com; silviaanggunr@gmail.com; taqwa@amikomputwokerto.ac.id
* corresponding author

(Received: December 10, 2021 Revised: December 25, 2021 Accepted: January 8, 2022, Available online: January 29, 2022)

Abstract

Coronavirus (COVID-19) is a rapidly emerging and spreading infectious disease. To minimize the impact caused by the virus, it is necessary to have a vaccine. However, the existence of vaccinations for the Indonesian people has caused controversy so that it invites many people to give an opinion assessment, therefore people choose social media as a place to channel their opinions. In this study, a comparison was made with an observational infoveillance study by collecting data using a Python programming script (Python Software Foundation) to display posts related to the COVID-19 vaccine on Twitter as well as quantitative and qualitative analysis to identify trends and characterize the main themes discussed by twitter users on Twitter. Indonesia. Our research collects data through social media Twitter in the period August 2020 - March 2021. In this study we combine Retrospective Observation and Sentiment Analysis, with the aim of producing periodic timeline evaluations within a predetermined time frame. In this study author found that there was an interaction & increase in positive posts due to officially reported developments, on the other hand we were quite difficult to understand the factors behind the emergence of negative posts but we made a conclusion based on the results of sentiment analysis that most of the negative posts were caused by lack of information and understanding of vaccines and vaccines. the COVID-19 outbreak itself.

Keywords: Retrospective Observation; Sentiment Analysis; Twitter Opinion; Vaccine

1. Introduction

Coronavirus disease (COVID-19) is a rapidly emerging infectious disease caused by a new coronavirus called the severe acute respiratory syndrome (SARS) coronavirus. The COVID-19 outbreak began in late December 2019 in Wuhan, Hubei Province, China, with a group of patients with pneumonia of unknown origin and reporting exposure to seafood and live animal markets in the same city [1]. The World Health Organization (WHO) confirmed 41 cases and 1 death from the new coronavirus on January 12, 2020. Since this initial report, COVID-19 has spread rapidly in China and internationally, with WHO declaring a Public Health Emergency of International Concern (PHEIC) under the revised International Health Regulations on 30 January 2020 [2].

Since the PHEIC declaration [3], COVID-19 has spread to every continent except Antarctica, becoming a highly contagious global pandemic with continued community transmission. The severity of the COVID-19 outbreak, with approximately 1.8 million cases worldwide as of mid-April 2020, has far surpassed past coronavirus events such as the Middle East respiratory syndrome (MERS)-related coronavirus, which had 2494 cases as of November 2019, and the coronavirus. SARS 2003, which had more than 8000 cases and affected 26 countries. It is not known whether viral mutations will result in the annual pattern of reappearance seen in influenza strains.

Attempts to predict the epidemiological features (eg, prevalence, attack rate, replication or reproduction rate, morbidity, and mortality) of outbreaks to inform infection control and public health response are of great importance [4]. This can be a challenge during the early stages of an outbreak when there is a lack of sufficient information regarding the etiology of the disease, inadequate diagnostic and testing capabilities, and incomplete epidemiological

data regarding confirmed cases [5]. In the absence of such data, the use of information in electronic media such as social media conversations could allow a syndrome surveillance approach to characterize disease distribution and provide more accurate case counts more quickly [6].

This “notice” approach has been used to characterize a number of public health issues including topics related to mental health, substance abuse behavior, spread of foodborne illness, and monitoring of infectious disease outbreaks (e.g. pertussis, influenza, HIV/AIDS, dengue fever, West Nile, Zika virus, H1N1, and Ebola). In particular, the now ubiquitous nature of social media means that it represents an important “nontraditional” source for disease surveillance [7]. In particular, user-generated social media data can be mined to assess people's knowledge, attitudes and behavior towards disease, and can be very informative when cross-validated with traditional disease surveillance data [8]. Others have also used global social media platforms such as Twitter to examine the 2009 H1N1 pandemic, conduct content analysis, and identify key trends that may also correlate with outbreak incidence data.

Utilizing the info veillance approach, we conducted a retrospective observational study for the COVID-19 vaccine on one of Indonesia's largest social media platforms, Twitter. Twitter is a microblogging site and one of the most influential social media platforms in Indonesia [9]. According to its own press release as of August 2020, it has more than 486 million active users. Users can publish content such as messages on microblogs and share text, images, videos and music. Compared to Whatsapp, another popular social media platform in Indonesia, Twitter posts are generally more publicly visible; with Whatsapp, posts are generally more private and can only be seen by certain people selected by the user. As the platform is public, we tried to assess whether Twitter posts about the COVID-19 Vaccine could identify its development during the early stages of the outbreak and conducted a qualitative analysis of the themes related to the COVID-19 vaccine detected and discussed by users located in West Java.

2. Literature Review

2.1. Data Mining

Data Mining is a process for extracting (Mining) new knowledge and information from large amounts of data in the data warehouse, using artificial intelligence, statistics, machine learning and mathematics methods. The method used by the data mining will later identify and extract useful information from a database [10]. In addition, data mining is often referred to as a data mining process in very large amounts of data using statistical, mathematical methods, to utilizing the latest artificial intelligence technology which is expected to bridge communication between data and its use. The main purpose of data mining is to perform processing by utilizing the data in the database so as to obtain useful new information [11].

According to experts [12-14], the purpose of data mining is to extract and identify data for certain information related to a large database or big data. In simple terms, data mining is the mining or discovery of new information by looking for certain patterns or rules from a very large amount of data [15]. Data mining is also referred to as a series of processes to explore added value in the form of knowledge that has not been known manually from a data set [16]. Data mining, also known as knowledge discovery in database (KDD). KDD is an activity that includes collecting, using historical data to find regularities, patterns or relationships in large data sets [17]. Data mining is defined as the process of finding patterns in data. This process is automatic or often semi-automatic. The pattern found must be meaningful and the pattern provides benefits, usually economic benefits. Large amounts of data are needed [18]. Characteristics of data mining as follows:

- Data mining is related to the discovery of something hidden and certain data patterns that were not previously known.
- Data mining usually uses very large data. Usually big data is used to make the results more reliable.
- Data mining is useful for making critical decisions, especially in strategy.

Based on some of these understandings, it can be concluded that data mining is a technique of extracting valuable information that is hidden or hidden in a very large data collection (database) so that an interesting pattern is found that was previously unknown. The word mining itself means the effort to get a little value from a large number of

basic materials. Because of that data mining actually has long roots from fields of science such as artificial intelligence (artificial intelligence), machine learning, statistics and databases. Several methods that are often mentioned in the data mining literature include clustering, classification, association rules mining, neural networks, genetic algorithms and others [19].

2.2. Sentiment Analysis

Sentiment Analysis is a technique to find out public opinion on a particular object obtained from a collection of data. Sentiment analysis itself or also commonly referred to as opinion mining is one of the 16 parts of text mining. Sentiment Analysis is a field of study that discusses people's opinions, sentiments, evaluations, behavior and emotions towards an entity such as products, services, organizations, individuals, problems, topics, events and their attributes.

With the growth of increasingly sophisticated information technology, it influences changes in the way humans communicate with each other. With the growth of increasingly sophisticated information technology, this is accompanied by the use of social media which is widely used by the general public to communicate or express their opinions. One of the social media that is widely used by the public is Twitter which is used by the community to share what is felt by its users. Through posting on Twitter, people can share and get information about anything. By utilizing data from social media Twitter, an analysis of public opinion and opinion on vaccines can be carried out through Sentiment Analysis.

According to Hannum [16] sentiment analysis or opinion mining refers to a broad field of natural language processing, computational linguistics and text mining which has the aim of analyzing opinions, sentiments, evaluations, attitudes, judgments and emotions of a person whether the speaker or writer relates to a topic, specific product, service, organization, individual or activity. In addition, the journal written by Hannum [16] explains that sentiment analysis is a field of study that analyzes a person's opinions, sentiments, evaluations, attitudes, and emotions from written language. Sentiment analysis is one of the most active research areas in natural language processing and is also widely studied in data mining, web mining, and text mining. Apart from computer science, sentiment analysis is also useful in other fields, such as management science and social science [20]. Sentiment analysis also uses algorithms to process and classify the built data. There are many algorithms that can be used in sentiment analysis research. There are ten best algorithms commonly used, including C4.5, The K-Means, Support Vector Machine, Apriori, Maximum Entropy PageRank, AdaBoost, k-nearest neighbor, Naive Bayes, CART.

From some of the opinions above, it can be concluded that sentiment analysis is a processor that can identify an opinion, opinion, and emotion from a text by classifying it into negative, positive, and neutral classes. Incorrectly Take data through user posts or tweets and the system will automatically retrieve data through posts. One application of sentiment analysis is carried out on Twitter, namely the system will automatically include or tweet from the user and the system will perform a classification to assess whether the post contains neutral, positive sentences, or negative. so that with the results obtained based on these classifications, users can know and evaluate a particular topic and can make a decision.

2.2. Retrospective Observation

A retrospective study looks back and examines exposure to a suspected risk or protective factor in relation to outcomes defined at the start of the study. Many valuable case-control studies, such as the Becken investigation [21], of breast cancer risk factors, are retrospective investigations. Most sources of error due to confounding and bias are more common in retrospective studies than prospective studies. For this reason, retrospective investigations are often criticized [22]. However, if the desired results are not general, the size of the prospective investigation required to estimate the relative risk is often too large to be feasible. In a retrospective study, the odds ratio provides an estimate of the relative risk. Care must be taken to avoid sources of bias and confounding in retrospective studies.

3. Methodology

This observational infoveillance study was conducted in two phases: data collection using an automated Python (Python Software Foundation) programming script to compile posts related to the COVID-19 vaccine on Twitter, and quantitative and qualitative analysis to identify trends and characterize the main themes discussed by Indonesian users. All information collected from this research is from the public domain, and this research does not involve any interaction with users. Unspecified user information has been removed from the study results. There were 115,299 posts collected during the study period, with an average of 2956 Twitter posts per day. There is a high degree of variation in the number of posts depending on the date of collection, the highest number of posts (13,740) was collected on September 17, 2020.

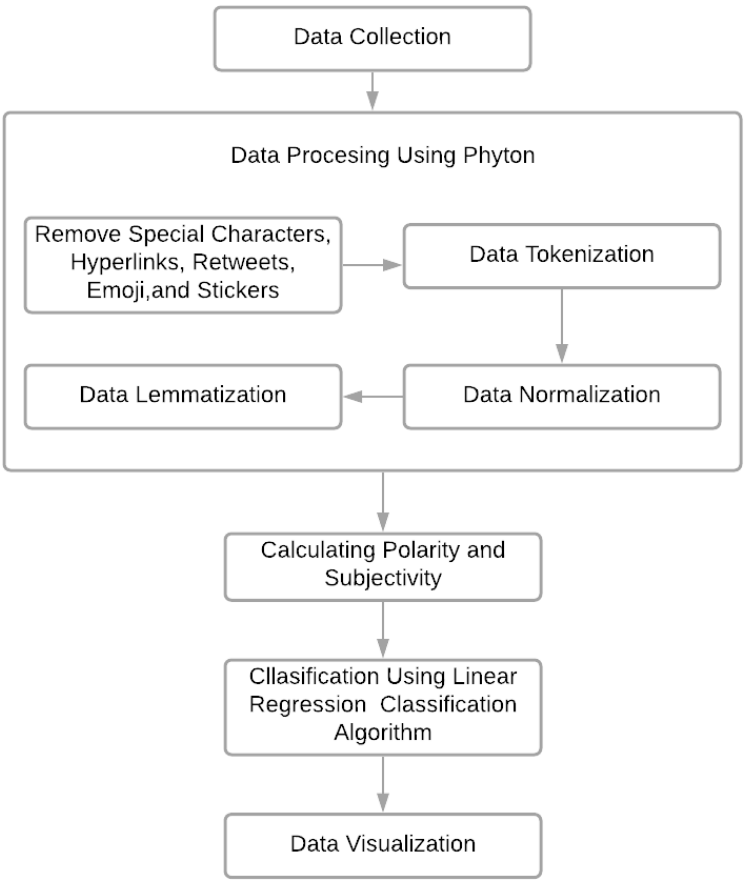

Figure. 1. System model

In this paper, a system is proposed to analyze Twitter tweets about the COVID-19 vaccine and indicate positive or negative sentiment in the text using python and a linear regression algorithm. To implement the system, the first tweet data was retrieved from Twitter by standard Twitter search, and the retrieved data was saved in CSV file format. This data needs to be processed first because it contains special characters, hyperlinks, retweets, emojis, and stickers. Python is used to preprocess data and makes it suitable for implementing linear regression algorithms. After removing the special characters, the data is tokenized. For the next process, normalization and lemmatization of data are carried out. Once the data is processed using Python, they are computed. Finally, data visualization was performed on the classified data and further analyzed for comparison. The complete system diagram is illustrated in Figure 1.

3.1. Data Collection

Work on implementing the proposed system was carried out on text data from special Twitter tweets related to the COVID-19 vaccine. To extract tweets, a Twitter developer account is mandatory. There the Twitter key/API credential will be stored in a variable and then create an authentication object. Finally, the access token and access token secret are assigned and authenticated to Twitter. Figure 2 shows a process flow diagram.

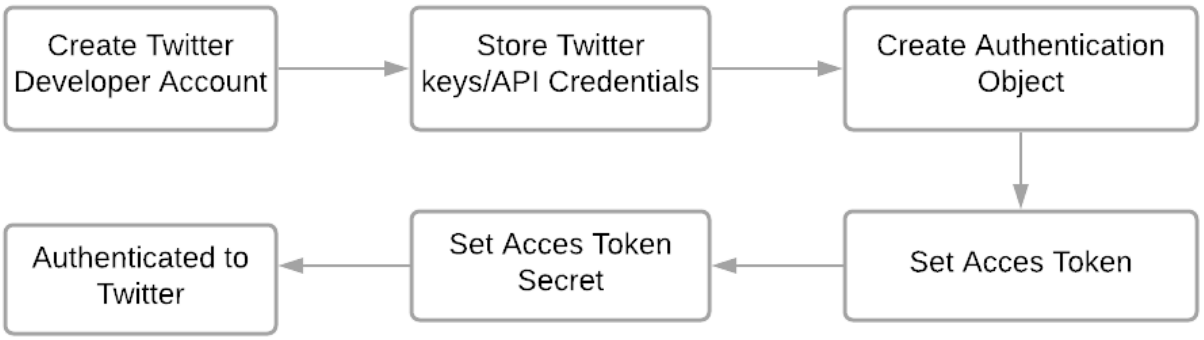

Figure. 2. Twitter authentication

To extract COVID-19 vaccine tweets from Twitter, tweets using vaccine keywords, namely “coronavirus”, “vaccine”, “COVID”, “pandemic” were targeted. Five thousand (5,000) tweets for each hashtag taken. These keywords were selected based on manual searches via the platform's public search function to detect the basis of user conversations related to vaccines for our systematic data collection process. From tweets, only text data is extracted and saved in CSV format as a dataset. Programming script written in Python programming language to extract posts on Twitter in Indonesian from users who self-reported their location in the province of west java. A Python script was implemented to continuously collect filtered data for keywords related to the COVID-19 vaccine from August 1, 2020, to March 31, 2021.

Table. 1. Tweets Sample

Username	Account Name	Tweets	Date	Reply	Likes
@Ferdinandyeah	Ferdinand Hutapea	Belakangan ini saya perhatikan jalanan jakarta mulai ramai lancar bahkan padet. Tampaknya jakarta sudah mulai terbiasa dengan kondisi covid dan aktivitas sudah meningkat menuju normal. Semoga vaksin semakin banyak agar ekonomi kembali bergerak normal	2021-03-13	34	351
@Bagussakti	Bagus Sakti Nugroho	Hingga hari ini, 5.5 juta orang lebih di indonesia sudah menerima vaksin covid. Rata2 100rb perhari, rate harian saat ini 200rb orang sehari. Semoga bisa naik ke kisaran 300rb orang per hari. Lumayan. Jakarta bisa Clear	2021-01-22	12	122
kompastv	KOMPAS TV	Presiden @jokowi sudah menjalani penyuntikan vaksin COVID-19 dari Sinovac pada Rabu (13/1). Penyuntikan vaksin	2021-01-13	22	508

		COVID-19 terhadap Presiden Jokowi dilaksanakan sekitar pukul 09.30 WIB di istana Merdeka , Jakarta Pusat. https://t.co/YxsOjF6pTd https://t.co/IJ3p9IWBzy			
detik.com	detikcom	#Foto Presiden Jokowi telah disuntik vaksin COVID-19 Sinovac . Begini momen saat vaksinasi yang yang dilakukan di istana Negara, Jakarta tersebut: #Jokowi #VaksinCorona Baca beritanya di https://t.co/EporxOIzmO Foto:Agus Suparto/Setpres https://t.co/IcjUQ8oSCj	2021-01-13	6	206

Referring to the previous example, we can see that positive sentiments are expressed in tweets discussing the halalness of vaccines, vaccines as an effort, and the safety of vaccinations in the city of Jakarta. Moments later, a tweet with neutral sentiment was sent by a news account, and they only mentioned the vaccination program in Jakarta. Twitter users expressed their dissatisfaction with the vaccine through surveys and public complaints about the socialization of the vaccination program in Jakarta, according to tweets with negative sentiments. Posts are filtered for geographic location, thus identifying users specifically in the province of West Java. Posts with geographic locations outside these provinces were not included in this study, as the aim is to focus on the vaccine timeline in these regions. Posts are collected from all types of accounts including personal accounts, media accounts and government accounts. The technical limitations of the Twitter platform and the Python script used to collect the data limit our data collection to a maximum of 2000 posts per hour. However, during the 936 hours of data collection, we did not reach this limit for every hour of collection.

3.2. Pre-Processing

To make the data suitable for the application of machine learning algorithms, the raw data must go through a preprocessing stage. Python is used in this system for data preprocessing. At this stage, the text data is first converted to lowercase. From this form, all stop words are omitted and contractions are replaced. The stop word list is defined in the python nltk library used in this process and to replace the contraction, a special function is created to complete the task. To avoid hassle, a spell check is performed to correct misspelled words. One of the most important steps in the preprocessing in this work was replacing the emojis with the expressions they represent in plain English. Furthermore, special characters, URLs, and HTML tags are removed from the text. Finally, tokenization, normalization, and lemmatization are performed on text data before moving to Object Identification. Figure 3 shows the process flow.

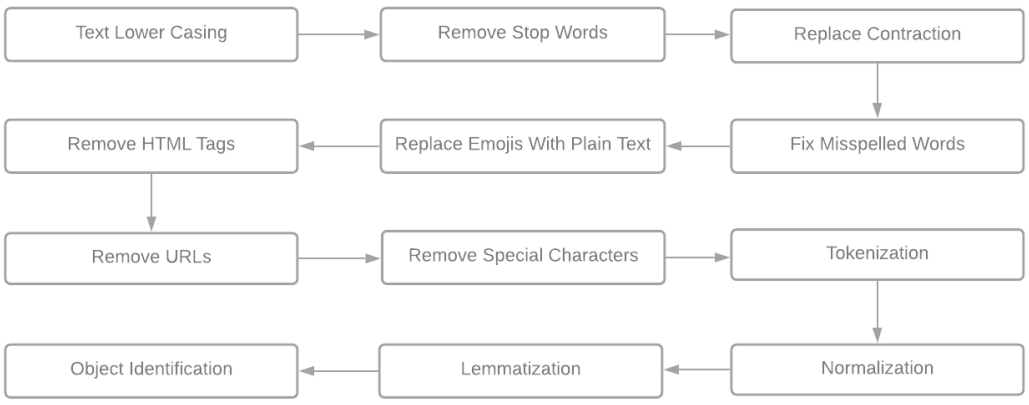

Figure. 3. Data pre-processing

Tokenization, Normalization, and Lemmatization are the three main functions in natural language processing for pre-processing text before classification.

1) Tokenization: Tokenization refers to the separation of a text document into smaller units. Each unit is called a token. In this work, each word is converted into a token.
2) Normalization: Normalization of text is changing the unusual text into its standard form. Sometimes, people write words in unusual forms to express themselves. This text needs to be converted into the correct form and correct spelling.
3) Lemmatization: A word can have different forms based on tense, gender and comparative adjectives, the basic form of each word is called lemma and the process of changing any word to its basic form is called lemmatization.
4) Object Identification: The last step of preprocessing is object identification. This function takes each column of data and checks if the column is empty. If the column is empty, it sets the value 0 and the other sets the value 1 and stores it in the new identification column.

3.3. Calculating polarity

Basically, sentiment analysis depends on polarity and subjectivity. Subjectivity contains facts, opinions and desires. For this python library, call TextBlob. To process these tasks like Sentiment Analysis, the python library TextBlob provides an API. From the polarity and subjectivity data, the mean, median, minimum mean, maximum mean were calculated for each vaccine. Maximum average polarity is calculated per 10 tweets. Below are the equations used in the calculations.

$$mean, \bar{x} = \frac{\Sigma x}{n} \quad (1)$$

$$median = \frac{n+1}{2} \quad (2)$$

$$Min\ Average = \frac{(n-1)\ min+max}{n} \quad (3)$$

$$Max\ Average = \frac{min + (n-1)max}{n} \quad (4)$$

3.3. Data Classification

Linear regression is a probabilistic model. Linear regression is used to find the probability of event=Success and event=Failure. Linear regression is used when the dependent variable is binary (0/1, Positive/Negative). Here the value of Y ranges from 0 to 1 and can be represented by the following equation:

$$odss = \frac{p}{(1-p)} = \frac{Positive\ Sentiment\ Chance}{Negative\ Sentiment\ Chance} \quad (5)$$

$$In(odds) = in(p/(1 - p)) \quad (6)$$

$$Logit(p) = in(p/(1 - p)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k \quad (7)$$

Above, p is the probability of the presence of the desired characteristic. To work with the binomial distribution (dependent variable), we need to choose the link function that is most suitable for this distribution and, here it is the logit function. In the above equation, the parameters are chosen to maximize the probability of observing sample values rather than minimizing the number of squared errors (as in ordinary regression).

The logit function is simply the average function of the response variable Y that we use as the response instead of Y itself. All that means when Y is categorical, we use the logit of Y as the response in our regression equation, not just Y:

$$in(p/(1 - p)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k \quad (8)$$

The logit function is the log of the probability that Y is equal to one of the categories. For mathematical simplicity, we will assume that Y has only two categories and relate them as 0 and 1 or positive and negative.

4. Result and Discussion

All information collected from this research is from the public domain, and this research does not involve any interaction with users. Unspecified user information has been removed from the study results. There were 115,299 posts collected during the study period, with an average of 2956 Twitter posts per day. There is a high degree of variation in the number of posts depending on the date of collection, the highest number of posts (13,740) was collected on September 17, 2020.

Linear regression shows a positive relationship between Twitter posts and officially reported developments in West Java Province, with approximately 35 positive posts per 50 daily social media posts. The official number of cases in Indonesia does not include West Java Province. These results indicate that there is a statistically significant positive relationship with Twitter posting and vaccination developments within the province of West Java, and that the process of administering the vaccine is larger than that observed throughout Indonesia excluding West Java. Social media posts from official government accounts are also a factor in the increase in postings in West Java, but not only in West Java and even throughout Indonesia. This result may imply that posts from official government accounts provide a 37% increase in the number of daily posts in the number of posts for all of Indonesia but provide a fairly high increase in local posts (Javanese). Therefore, it is possible that the number of Twitter posts related to vaccines is reactive to how governments act and respond to vaccinations while understanding vaccination conditions for the wider region.

A longitudinal trend visualization found that this association was generally undisturbed, except for dramatically fewer posts on October 8 and January 25. The decline for January 25 coincided with the first wave of vaccinations across Indonesia, but this may not explain the decline observed on January 28 (see Figure 4). Further observations in our qualitative analysis identify certain events and news stories during the study period that might also affect the number of users' posts on a given observation date and are described further in the next section.

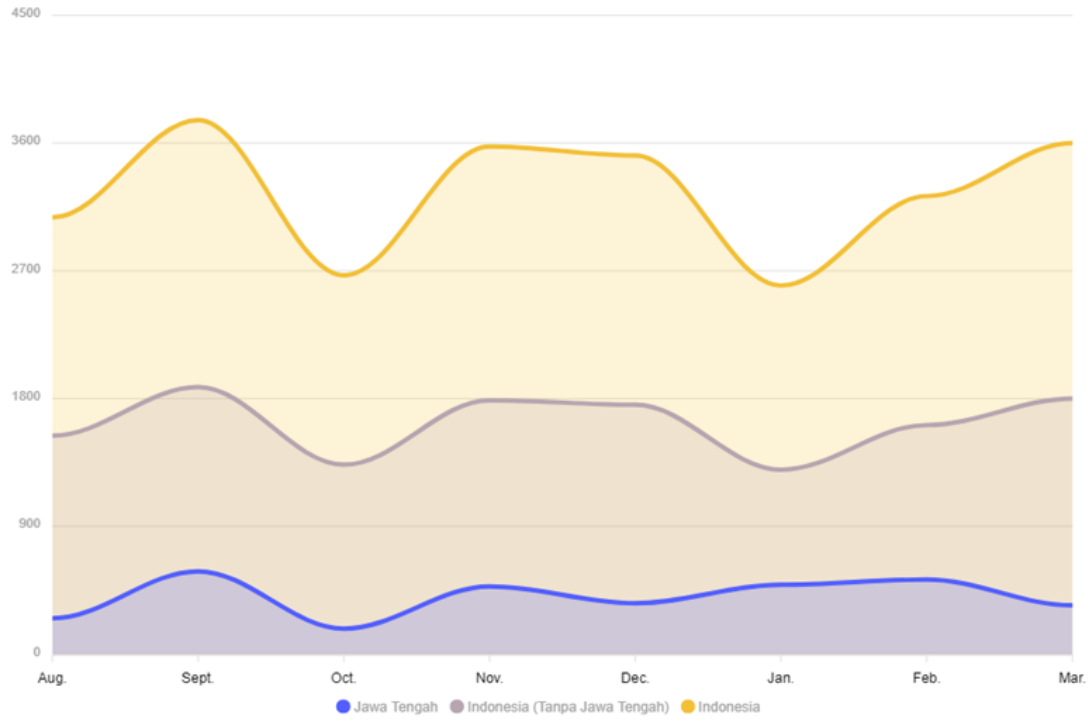

Figure. 4. Longitudinal trend of vaccine postings (August 2020 - March 2021)

A total of 12,113 posts from the first 48 days of our study (1 August 2020 to 31 March 2021) that detected vaccine-related posts were manually coded. Qualitative analysis revealed that certain common terms or language in the aggregated collection of Twitter posts reflected the parent classification of the initial Vaccine theme including “coronavirus”, “COVID”, “pandemic”. The distribution of these terms varied over the study period, with "vaccine" returning the largest proportion of Twitter posts during the initial vaccine period through March 31, 2021, after which the term "vaccine" began to generate more mentions of Twitter posts.

Several important themes were identified in our inductive content analysis of Twitter posts, including four master classifications with discussion of expectations of government action, comments on government action, public reactions to vaccine controls and responses and criticism and suggestions. The prevailing theme during the vaccination period that changes based on the availability of new information is that there is a lot of false information, which creates uncertainty in the administration of the vaccine. This initial period of uncertainty was followed by disclosure of information about government strategies or actions by the Indonesian National Health Commission and other official government and academic sources. When vaccines begin to develop, the prevalence rate and the very high number of cases tend to be the cause of the government's hesitation in establishing an initial strategy for administering the vaccine.

The presence of the parent classification also changes from time to time. For example, at the beginning of the spread of the vaccine, many users discussed why the government prioritized medical personnel, knowing that the level of COVID-19 in Indonesia, especially West Java, was already quite dangerous. From January 13 to February 27, posts about the beginning of the spread of vaccinations also increased with most of the conversation discussing new types of vaccines and viruses. Of course, there are also many posts that are fake, this is enough to be responded very firmly not only by official government accounts but also by the general public. Detection of posts related to comments on government actions was relatively consistent over the period assessed. However, every time the government issues a new regulation and mandate, there is always a spike in discussion about government actions in dealing with vaccinations. Related, at the start of the vaccine, there were several posts where users expressed their reactions and personal concerns about the vaccination.

The more specific sub themes regarding knowledge, attitudes, and responses of users to vaccines change as more information becomes available about vaccinations which are underlined. In particular, the terms “confirmed number”, “suspected symptom”, “case death”, “vaccine type”, “vaccine administration”, “public health surveillance”, and “quarantine” became more common as vaccines began to develop. Along with this change in terminology, we also observe wide variations in user reactions due to information from government sources being disseminated. This includes posts that convey protective behavior (e.g., cleaning hands, avoiding crowds, wearing a medical mask in public), while others convey information and opinions that are likely to be misleading. New topics of user feedback began to emerge towards the end of the vaccination deployment period, including criticism of the health ministry's response and government uncertainty regarding news of quarantines, travel restrictions, and the new normal project.

Most importantly, the nature of the content and the volume of posts are likely to be driven by a combination of government information releases and news events. For example, November 30, 2020 has the second highest posting volume, which is in accordance with confirmation from an official source (Kemenkes RI) about giving priority to vaccines for medical personnel detected in West Java. There was also an increase in posting of vaccines on Twitter on December 5, 2020, possibly fueled by news on December 4, 2020 that laboratory tests had found side effects from certain types of vaccines and the growing controversy over how the vaccine would be distributed to the public. In addition, on December 13, 2020, the Minister of Health announced their response to the symptoms discussed. This prompted a second increase in the overall number of posts on 18 December 2020 due to a lack of detail from the information provided. Finally, on January 27, 2021, the vaccination timeline and schedule was confirmed, this resulted in a large number of conversations from media accounts and Twitter users, resulting in the largest increase in postings observed during this time period.

Most of the veillance info studies have analyzed data from English-language social media platforms such as the microblogging site Facebook or Google data, but only a few have examined foreign-language platforms. This study sought to identify, characterize, and assess the potential relationship between Indonesian social media conversations in West Java and the interactions of these social media users on vaccine development within the timeframe of our study. We also seek to understand how user perceptions change as additional information becomes available from government and media sources as a vaccine develops while also trying to identify the parent classification of user-generated themes that emerged as the vaccine was accelerated.

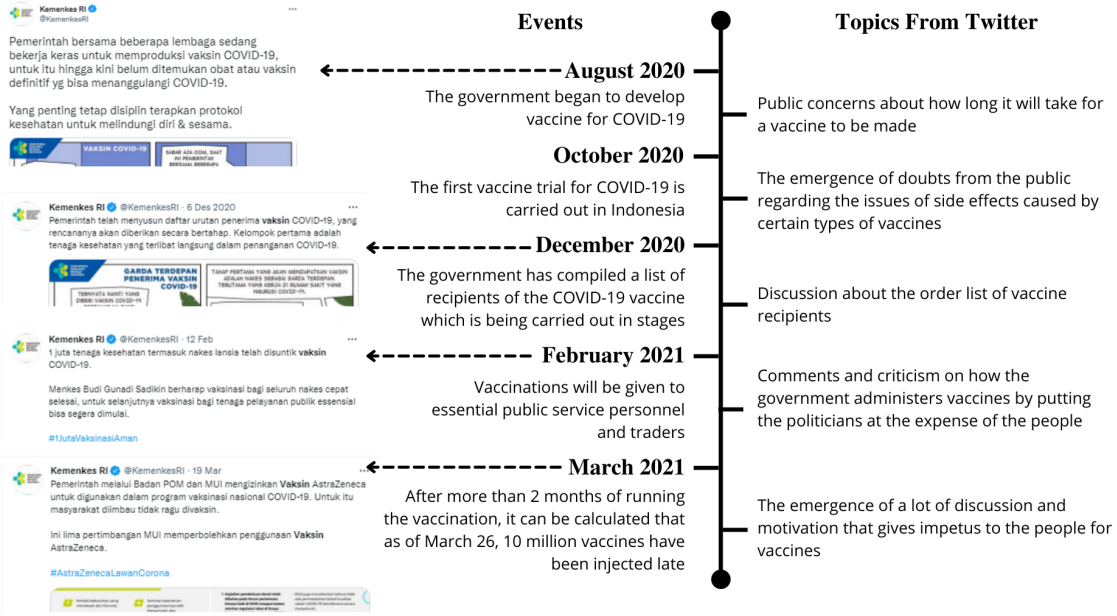

Figure. 5. Timeline of COVID-19 vaccination, Themes detected in twitter posts, and Government posts that support vaccine development

Based on our analysis on Figure 5, there appears to be a positive correlation between the number of vaccine-related twitter posts from the province of West Java and the official information from the government that was consistently provided during this early stage of vaccination. This effect size is larger than what is observed for the whole of Indonesia excluding West Java Province. However, any potential predictive value of using social media data as proxies for world public health surveillance statistics requires more rigor and an additional layer of data to confirm possible associations, especially in the context of user reactions to news events such as those discussed in this and other studies. Despite these limitations, the qualitative analysis characterizes the early stages of vaccination as having different levels of uncertainty for the Indonesian public regarding the risks posed by the vaccine. When information emerged about this type of vaccine, users expressed new concerns, which also led to changes in knowledge, attitudes and behavior among Indonesian social media users, some positive and some conveying information that tended to be negative.

In response to public uncertainty, the Indonesian government issued a series of announcements on its "official" Twitter account in an effort to classify vaccine characteristics as they become known, including official information to the public on August 28, 2020, about how the vaccination process goes and a subsequent announcement on October 28, 2020 regarding vaccine simulations. This event proves that social media is used as a vaccine communication tool by governments, media and users (who repost content) and causes the dissemination and reaction of users to information about vaccines whose trajectory will be global.

Public perceptions of vaccines changed during the study period, with initial conversations focusing on public concerns about the manufacture of vaccines that would later be given to the public, and then leading to discussions on the issue of side effects caused by certain types of vaccines. Overall, we observed wide variation in users' reactions to information, with some users expressing a willingness to engage in positive behaviors and other users downplaying risks and engaging in behaviors that could make matters worse. Importantly, these observed attitudes and behaviors occurred before the Indonesian government announced the administration of the vaccine. After the announcement of this vaccine action, there has been a sizable increase in the volume of Indonesian Twitter posts that we upload.

While the mixed quantitative and qualitative results of this study are primarily exploratory in nature, they provide important insights into the changes in knowledge, attitudes and behavior of Indonesian social media users that are at the center of what is now a rapidly growing pandemic that has impacted all aspects of society. global. More research is needed to better understand the effectiveness of communication strategies regarding vaccines, particularly in the context of how information is understood, shared, and acted upon by users in the face of uncertainty and changing information. In particular, we need to better understand how social media platforms can influence public risk perceptions, their trust and credibility with various sources of information, and, ultimately, how they change real-world behavior that can impact control measures put in place to reduce outside events. normal.

Initial reports suggest that the social media platform Twitter is struggling with the volume of vaccine information and user-generated content flooding their platform, some of which is helpful and accurate and some of which is rumours and misinformation. In fact, the social media platform Twitter has announced steps to better ensure access to credible and accurate information about vaccines, although whether this platform fulfills this task remains an open question. social media can act as a positive tool to promote global health goals, particularly in the context of health emergencies, will be tested by vaccines, along with their usefulness as a modern approach to public health surveillance.

5. Conclusion

Several limitations apply to this investigation. To begin with, our data collection is limited to one Indonesian social media site and a certain period of time. As a result, it cannot be used for all Vaccination-related social media exchanges between Twitter users. Due to the difficulty of obtaining data on this platform, we do not analyze Indonesian private communication programs (e.g. Whatsapp). Future research should examine a wider variety of conversational data across multiple platforms and leverage sentiment analysis and machine learning to aid in the classification of larger discussion volumes. Second, our data collection begins with the first report on vaccine

procurement. Sometimes during this period of time, determining the elements that influence the number of positive and negative posts can be challenging. Due to initial terminology issues, Twitter users may have used other terms to describe vaccination scenarios that were not included in this study. Third, simple linear regression revealed a positive relationship between vaccine-related Twitter posts and the total number of vaccine-related tweets over the study period. Furthermore, it is not known whether the predictive trendlines we found will persist or whether the trendlines can be transferred to other countries or communities where vaccination is practiced. It is most likely that only under certain conditions this correlation will occur, namely the lack of knowledge and socialization of the early stages of giving vaccines, new viruses with high transmission and sustainable community outreach, and high social media engagement involving conversations about vaccination. Further, as previously stated, it is highly likely that post volume is related to users' reactions to government news and announcements. Finally, due to censorship in Indonesia, posts may have been deleted prior to data collection. In fact, some of the detected messages including related comments have been deleted and cannot be retrieved.

As for suggestions that can be given for the further development of this research, it is hoped that further research can collect better data and in a longer period of time. So that the results obtained are more accurate and have a greater impact on further research. In further research, it is better if the processed data is processed and normalized so as not to produce a sentiment that is too broad. So that the resulting level of accuracy is higher. In subsequent research, not only in sentiment analysis on the topic of the COVID-19 vaccine on Twitter, but in all sentiment analysis in the future in this study, we found that the most important part of sentiment analysis in the future is not so much related to improving the accuracy of the algorithm, but lies in the area determining where the research can link sentiment with behavior.

References

[1] S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos, and S. Sharif, “An analysis of COVID-19 vaccine sentiments and opinions on Twitter,” Int. J. Infect. Dis., vol. 108, pp. 256–262, 2021, doi: 10.1016/j.ijid.2021.05.059.

[2] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, “Sentiment analysis and classification of Indian farmers’ protest using twitter data,” Int. J. Inf. Manag. Data Insights, vol. 1, no. 2, p. 100019, 2021, doi: 10.1016/j.jjimei.2021.100019.

[3] D. Thorpe Huerta, J. B. Hawkins, J. S. Brownstein, and Y. Hswen, “Exploring discussions of health and risk and public sentiment in Massachusetts during COVID-19 pandemic mandate implementation: A Twitter analysis,” SSM - Popul. Heal., vol. 15, no. February, 2021, doi: 10.1016/j.ssmph.2021.100851.

[4] P. Sharma and A. K. Sharma, “Experimental investigation of automated system for twitter sentiment analysis to predict the public emotions using machine learning algorithms,” Mater. Today Proc., no. xxxx, 2020, doi: 10.1016/j.matpr.2020.09.351.

[5] D. Antonakaki, P. Fragopoulou, and S. Ioannidis, “A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks,” Expert Syst. Appl., vol. 164, no. September 2020, p. 114006, 2021, doi: 10.1016/j.eswa.2020.114006.

[6] T. Bettuzzi et al., “Efficacy and safety of treatments in cutaneous polyarteritis nodosa: A French observational retrospective study,” J. Am. Acad. Dermatol., pp. 1–7, 2021, doi: 10.1016/j.jaad.2021.06.872.

[7] X. Zhang et al., “Risk factors for prolonged intensive care unit stays in patients after cardiac surgery with cardiopulmonary bypass: A retrospective observational study,” Int. J. Nurs. Sci., vol. 8, no. 4, pp. 388–393, 2021, doi: 10.1016/j.ijnss.2021.09.002.

[8] K. Ekelöf, O. Andersson, A. Holmén, K. Thomas, and G. Almquist Tangen, "Depressive symptoms postpartum is associated with physical activity level the year prior to giving birth – A retrospective observational study," Sex. Reprod. Healthc., vol. 29, no. July, 2021, doi: 10.1016/j.srhc.2021.100645.

[9] Z. Zvizdic et al., "The predictors of perforated appendicitis in the pediatric emergency department: A retrospective observational cohort study," Am. J. Emerg. Med., vol. 49, pp. 249–252, 2021, doi: 10.1016/j.ajem.2021.06.028.

[10] A. P. Nunes, J. D. Seeger, A. Stewart, A. Gupta, and T. McGraw, "Cardiovascular Outcome Risks in Patients With Erectile Dysfunction Co-Prescribed a Phosphodiesterase Type 5 Inhibitor (PDE5i) and a Nitrate: A Retrospective Observational Study Using Electronic Health Record Data in the United States," J. Sex. Med., vol. 18, no. 9, pp. 1511–1523, 2021, doi: 10.1016/j.jsxm.2021.06.010.

[11] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[12] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," J. Informetr., vol. 3, no. 2, pp. 143–157, 2009, doi: 10.1016/j.joi.2009.01.003.

[13] R. Feldman, "Techniques and applications for sentiment analysis," Commun. ACM, vol. 56, no. 4, pp. 82–89, 2013, doi: 10.1145/2436256.2436274.

[14] A. Joshi, P. Bhattacharyya, and S. Ahire, Sentiment Resources: Lexicons and Datasets. 2017.

[15] C. Udanor and C. C. Anyanwu, "Combating the challenges of social media hate speech in a polarized society: A Twitter ego lexalytics approach," Data Technol. Appl., vol. 53, no. 4, pp. 501–527, 2019, doi: 10.1108/DTA-01-2019-0007.

[16] C. Hannum, K. Y. Arslanli, and A. F. Kalay, "Spatial analysis of Twitter sentiment and district-level housing prices," J. Eur. Real Estate Res., vol. 12, no. 2, pp. 173–189, 2019, doi: 10.1108/JERER-08-2018-0036.

[17] S. Becken, A. R. Alaei, and Y. Wang, "Benefits and pitfalls of using tweets to assess destination sentiment," J. Hosp. Tour. Technol., vol. 11, no. 1, pp. 19–34, 2020, doi: 10.1108/JHTT-09-2017-0090.

[18] K. Fiok, W. Karwowski, E. Gutierrez, and M. Wilamowski, "Analysis of sentiment in tweets addressed to a single domain-specific Twitter account: Comparison of model performance and explainability of predictions," Expert Syst. Appl., vol. 186, no. July 2020, p. 115771, 2021, doi: 10.1016/j.eswa.2021.115771.

[19] S. Bashir et al., "Twitter chirps for Syrian people: Sentiment analysis of tweets related to Syria Chemical Attack," Int. J. Disaster Risk Reduct., vol. 62, no. May, p. 102397, 2021, doi: 10.1016/j.ijdrr.2021.102397.

[20] Y. Ding, R. Korolov, W. (Al) Wallace, and X. (Cara) Wang, "How are sentiments on autonomous vehicles influenced? An analysis using Twitter feeds," Transp. Res. Part C Emerg. Technol., vol. 131, no. July, p. 103356, 2021, doi: 10.1016/j.trc.2021.103356.

[21] S. Liu and J. Liu, "Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis," Vaccine, vol. 39, no. 39, pp. 5499–5505, 2021, doi: 10.1016/j.vaccine.2021.08.058.

[22] M. M. Rahman et al., "Socioeconomic factors analysis for COVID-19 US reopening sentiment with Twitter and census data," Heliyon, vol. 7, no. 2, p. e06200, 2021, doi: 10.1016/j.heliyon.2021.e06200.