Social-Politic in a Pandemics Indonesian Citizens' Twitter Conversation Analysis using Latent Dirichlet Allocation

Dian Sa'adillah Maylawati
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
diansm@uinsgd.ac.id
https://orcid.org/0000-0002-1193-3370

Muhammad Andi Septiadi
Department of Political Science
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
Septiadi.andi90@uinsgd.ac.id
https://orcid.org/0000-0001-8058-3928

Neng Nur Annisa
Magister of Sharia Economics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
annisa.nengnur93@gmail.com
https://orcid.org/0000-0003-1368-2365

Nunik Destria Arianti
Department of Informatics
Universitas Nusaputra
Sukabumi, Indonesia
nunik@nusaputra.ac.id

Riyan Naufal Hay's
Department of Information System
Universitas Serang Raya
Serang, Indonesia
riyan@unsera.ac.id
https://orcid.org/0000-0001-7849-0407

Zulmi Ramdani
Department of Psychology
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
zulmiramdani@uinsgd.ac.id
https://orcid.org/0000-0002-7052-4462

Abstract— During the COVID-19 pandemic, there has been an increase in communication through social media. However, this is not accompanied by the level of citizen discipline where the high spread of false issues causes problems in the social and political fields. This study aims to look at the views of the community during the COVID-19 pandemic regarding Social and Political. The research method uses Latent Dirichlet Allocation Analysis to extract topic cluster of political and social during COVID-19 pandemic in Indonesia. The results of this study indicate a public view that considers the COVID-19 pandemic to be a tool, interest, or political game, and social concerns have increased considerably during the COVID-19 pandemic. During COVID-19, the increase in public communication through social media has become a driving force for other social activities that can ease the burden on the community. Besides that, the community seems wise enough to choose which news is fake and what is true.

Keywords— COVID-19, LDA, Politic, Social

I. Introduction

As we all know, there are numerous social media platforms available, with Facebook and Twitter being the most popular [1]. These social media sites have made it simple to communicate to friends and family members [2]. People with different values become closer by expressing their information, interests, and knowledge. It is now very simple for anyone else to meet people who share their interests in order to learn and share valuable information.

Social media applications such as Twitter, Facebook, and Instagram are communication tools used by the Indonesian people [3]. Social media applications such as Twitter, Facebook, and Instagram are communication tools used by the Indonesian people [4]. This has resulted in creating new issues through social media that affect social life. The spread of these issues is not all true information. In social media, a lot of false information is created that affects the life of society and the state [5]. The importance of knowing people's views during the COVID-19 pandemic is one of the main things to be studied, especially in social and political matters. Due to the fact that during the COVID-19 pandemic, people's conditions are not in good condition, the emergence of fake issues will add to the government's work in solving COVID-19 problems.

Several analyzes of people's views have been carried out, such as in the study: by Chaerani et al. [6], which analyzes the mapping of micro, small and medium enterprises, and Rachman and Pramana [7] about sentiment analysis of the pros and cons of the COVID-19 vaccine and the research of Marleny and Mambang [8] on social media analysis in accelerating disaster and flood response. From these three studies, we focus on analyzing people's views during covid 19 from a social and political perspective where the analytical media we use are different from other studies. This study aims to determine the differences in political views that can affect the public's response to the COVID-19 pandemic, including the public's response to government policies during the COVID-19 pandemic.

II. Research method

This research uses a mixed-method between quantitative and qualitative. Quantitative is done during the development and evaluation of the modeling topic. While qualitative when interpreting and analyzing the results of model development. This research was conducted with a data science approach using the CRISP-DM data science methodology (Cross-Industry Standard Process for Data Mining) [9], [10]. The CRISP-DM Data Science methodology activities consist of business understanding, data understanding, data preparation, modeling, evaluation, deployment, and model management. Where text data will be grouped based on similarity in content or topics that are widely discussed. Furthermore, an in-depth analysis was carried out on the text data as a whole based on the topics that emerged from each cluster that was formed. The algorithms used to form these topic clusters are Latent Dirichlet Allocation (LDA).

LDA forms two processes [11]: first, a document is made up of themes, and second, a topic is made up of tokens or words. In LDA, this topic generates words using a probability distribution. Documents are the probability density (or distribution) of subjects, and topics are the probability density (or distribution) of words. LDA processes include: (1) mapping a collection of text documents in the form of a document word matrix. The rows represent the number of papers, while the columns represent the unique words in each document. Term Frequency and Inverse Document Frequency (TF-IDF) weighting can be used to find these particular words

[12]–[14]; and (2) Enter the vector space computation process from the LDA algorithm, as shown in Figure 2. The procedure begins with the calculation of Dirichlet parameters, the distribution of topics in the document, the distribution of words in topics, the distribution of selected words, the calculation of topic distribution based on words, and the formation of topics and frequency of topics for each document.

III. RESULT AND DISCUSSION

A. Business Understanding

For instance, Twitter is a social media communication platform that is rich in sources of information related to public opinion on hot issues being discussed. People are so much more able to express themselves on social media. Of course, during the COVID-19 pandemic, various opinions, pros and cons, hoaxes, and other important information were distributed via social media. Differences in beliefs about religion, politics, gender, social causes, and health during the COVID-19 pandemic are compelling to evaluate.

Social media is a source of energy communication platform. Natural Language Processing's social media analysis technology can be used to find important information on social media. There ought to be a mapping of the topics most discussed by the public on social media instead of counting the messages conveyed on social media one by one. Topic modeling techniques with a clustering approach would be used to obtain an overall community view.

B. Data Understanding

This study uses data taken from Twitter social media, with several data collection scenarios according to the needs and objectives of the research, including:

1. Data is collected from Twitter from January 1, 2021, to July 31, 2021. This timeframe is considered when COVID-19 began to enter Indonesia until the period of the Emergency Community Activity Restrictions (PPKM).
2. The processed Twitter data is the tweet text grouped based on the research objectives. The Twitter data does not consider the user's age or gender, which Twitter's Application Programming Interface (API) does not provide access to. In addition, the data are grouped by four-time series for Political and Social views:
   a) The period January to March 2020 at the beginning of the COVID-19 pandemic.
   b) The period from March to May 2020 during the COVID-19 pandemic increased.
   c) The period of June-December 2020 is the new normal period (new habits), where people begin to adapt to COVID-19.
   d) Period January to July 2021 during the PPKM period until the COVID-19 emergency.
3. Twitter data is taken using keywords that are grouped according to the research objectives, among others :
   a. COVID-19 Group and Politics
      -COVID-19/corona and the government
      -COVID-19/corona and the central government
      -COVID-19/corona and local government
      -COVID-19/corona and DPR
      -COVID-19/corona and political parties
      -COVID-19/corona and politicians
      -COVID-19/corona and politics

      -COVID-19/corona and board members
      -COVID-19/corona and the president
      -COVID-19/corona and vice president
      -COVID-19/corona and the minister of health
      -COVID-19/corona and the minister of religion
      -COVID-19/corona and social minister
      -COVID-19/corona and Jokowi
      -COVID-19/corona and Maruf Amin
      -COVID-19/corona and Prabowo
      -COVID-19/corona and tourism
      -COVID-19/corona and the economy
      -COVID-19/corona and PSBB
      -COVID-19/corona and lockdown
      -COVID-19/corona and PPKM
      -COVID-19/corona and Emergency PPKM
      -COVID-19/corona and emergency brakes
      -COVID-19/corona and quarantine
      -COVID-19/corona and regional quarantine
      -COVID-19/corona and going home
      -COVID-19/corona and tadpoles
      -COVID-19/corona and corruption
      -COVID-19/corona and foreign debt
      -COVID-19/corona and health workers
      -COVID-19/corona and prioritize health workers
      -COVID-19/corona and doctors
      -COVID-19/corona and nurses
   b. COVID-19 and Social Group
      -COVID-19/corona and going home
      -COVID-19/corona and celebrations
      -COVID-19/corona and invitations
      -COVID-19/corona and villages
      -COVID-19/corona and cities
      -COVID-19/corona and alms
      -COVID-19/corona and giving
      -COVID-19/corona and help
      -COVID-19/corona and social assistance
      -COVID-19/corona and compensation
      -COVID-19/corona and food
      -COVID-19/corona and groceries
      -COVID-19/corona and rich
      -COVID-19/corona and poor
      -COVID-19/corona and work
      -COVID-19/corona and unemployment
      -COVID-19/corona and village government
4. The Twitter data collection process resulted in 966,605 data, with the distribution of the data as shown in Table 1.

TABLE I.   DISTRIBUTION OF TWITTER DATASET

Type of Data	Group of Data	Amount of Data
Main Dataset	Covid and Politics	654,313
	Covid and Social	312,292
Total Data		966,605

C. Data Preparation

The intermediate data preparation activity process is conducting a review and pre-processing of the text data. Pre-processing text data is very important because, at this stage, the data is prepared, cleaned, and selected based on the need to maintain the quality of the input data [15]–[17]. Activities carried out in the pre-processing of this text include: (1) labeling the data as a sentiment analysis requirement, the data

is labeled with positive, negative, and neutral labels; (2) case folding, namely the process of uniforming the size of the text into lowercase letters, because in computing using the Python language, capital letters and lowercase letters are distinguished (case sensitive); (3) cleaning the text of unused characters (regular expressions) and redundant characters, such as the use of excessive punctuation; (4) changing Indonesian slang words into their standard words; (5) changing emoticons that will affect the results of sentiment analysis; (6) change the abbreviation into the standard word; (7) eliminating stopwords or unimportant words that have no meaning, for example the conjunctions "which", "at", "is", "that is", and so on; (8) stemming, namely the process of changing words that have affixes into their basic forms, such as "eating" to "eating"; and (9) tokenizing is the process of cutting words which will then be represented in a structured form such as bag of words and 2-grams, and the frequency of occurrence of words is calculated using Term Frequency and Inverse Document Frequency (TF-IDF) [13], [14].

D. Modeling

COVID-19 Group and Politics

In the COVID-19 and Politics data group, the best clusters formed were 5 clusters, with the best CV (Coherence Value) value being 0.506. Figure 1 shows a CV graph that determines the best number of clusters that can be formed. Figure 2 shows five clusters formed from the LDA algorithm with three far apart clusters, which means that each cluster has unrelated similarities between clusters. Each of these clusters has a clearly different topic of conversation. In contrast, the other two clusters overlap each other, which means that there are similar topics of conversation between the two clusters.

Table 2 shows a word cloud that visualizes the most discussed topics in each cluster in the form of words and word pairs.
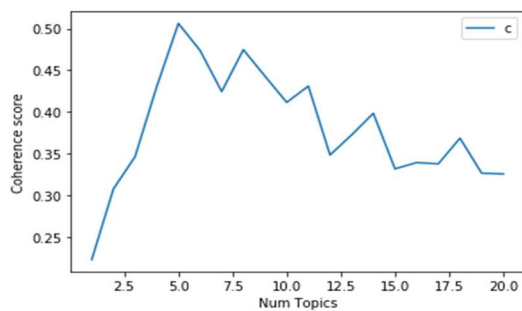

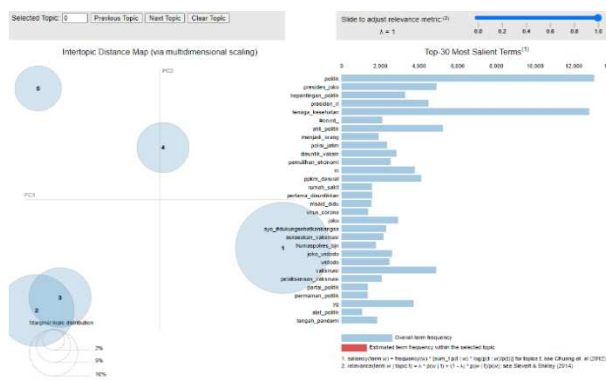Fig. 1. CV for COVID-19 and Political group

Fig. 2. LDA result for COVID-19 and political group

TABLE II. VISUALIZATION OF THE WORD CLOUD FOR COVID-19 AND POLITICAL GROUPS

COVID-19 and Social Group

The best clusters formed in the COVID-19 and social data groups were 2 clusters with the best CV value of 0.528. Figure 3 shows a CV graph that determines the best number of clusters that can be formed. Figure 4 shows two clusters formed from the LDA algorithm with two clusters far apart, which means that each cluster has unrelated similarities between clusters. Each of these clusters has a clearly different topic of conversation. Table 3 shows a word cloud that visualizes the most discussed topics in each cluster in the form of words and word pairs.
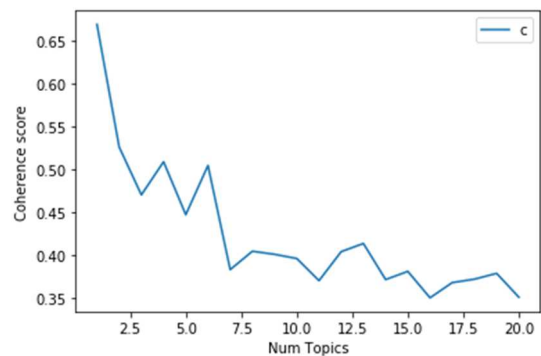

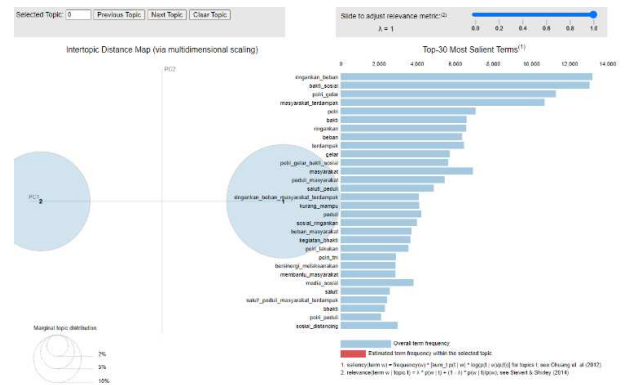Fig. 3. CV for COVID-19 and Social group

Fig. 4. LDA result for COVID-19 and social group

TABLE III.   VISUALIZATION OF THE WORD CLOUD FOR COVID-19 AND SOCIAL GROUPS

COVID-19 and Politics Analysis

The best number of clusters that can be formed is 5 clusters. In clusters 1, 2, and 3, there are indications of the public's view of COVID-19 being used as a tool, interest, or political game. There are many possibilities, from the health facilities, patient handling, vaccinations, and health workers, to political issues. In clusters 2 and 3, which are interrelated, there are interesting keywords between msaid_didu and politics during the COVID-19 pandemic.

This msaid_didu can be in the form of an account that is often mentioned or keywords that are often called so that they often appear together. It turned out that during the COVID-19 pandemic, news related to Muhammad Said Didu was a fairly viral figure. Many of his statements are related to the issue of preparing to move the capital city of a new country which is still ongoing amidst the efforts of the government and all parties to deal with the COVID-19 outbreak. Said Didu said that this shows that the government does not prioritize the welfare of the general public and only attaches importance to legacy (This article was published on Kompas.com with the title "Had Trending on Twitter, This is Said Didu's Track Record" [18].

In the third cluster, "health workers" appeared quite a lot because there were a lot of discussions related to meeting the needs of health workers in dealing with COVID-19. Such as the need to accelerate the certification of health workers. This intensive health workforce is also the main discussion. Many health workers resigned ("They resigned because the workload was heavy and the incentives for handling the pandemic had not been disbursed"). Intensive disbursement for health workers is slow and is associated with plans to move to the capital city, which should not be a priority during the COVID-19 pandemic. Many tweets come from the public and party politicians who expect a delay in laying the groundwork for the new capital city, prioritizing health workers' interests, and focusing on COVID-19 vaccinations.

The 4th and 5th clusters of society focus on President Jokowi's steps in implementing vaccination policies, PPKM, and economic recovery. In the 5th cluster, the word "president Joko" appeared due to the number of tweets discussing President Jokowi's policies regarding PPKM, including the Java-Bali Emergency PPKM and the national vaccination policy.

The word "political game" appears quite often in the 2nd cluster. This word comes from the public's view of the issue that COVID-19 is just a political game, for example: "Finally, many people think that covid is just a political game.... Don't always blame the people, people just imitate/follow what high state officials say ", "They say that covid is just a political game.", "I believe that covid is just a political game", "In fact, some believe that the hoax is that Covid 19 is just a political game. If it is true that there are parties who deliberately spread the virus, it does not mean that the virus

is not around us and there is no need to ignore it either. It's like watching a movie where a pharmaceutical company cheats so they can buy the vaccine Big spending to contain COVID-19.", "I'm one of those people who believe that Covid exists but I also believe that Covid is a world political game and the media can be profitable. certain "party."

COVID-19 and Social Analysis

In this Covid and social theme, two ideal clusters are formed where the two clusters are quite far away, which means they have different characteristics. In the first cluster, the words that appeared the most were "social service," "lighten the burden," and "police title," which turned out to be interrelated. The National Police is quite present in publishing COVID-19 social service activities (the National Police organize many social service activities). During PPKM level 4 the National Police Chief instructed all his staff to accelerate the distribution of social assistance to relieve the community affected by COVID-19. This social service is a form of Polri's concern for the community during the pandemic. Police are ready to help distribute social assistance to residents affected by COVID-19. The National Police also implements the distribution of social services with a door-to-door system.

Furthermore, Police are also quite viral in socializing vaccination to increase national vaccination. The National Police also intensively urges not to go home to minimize the spread of COVID-19, the National Police conduct isolation at many points, and many home comers turn back. Calls to comply with health protocols are also often spread by the Police. The word "lighten the burden" that often appears also comes from social service activities to ease the burden on people affected by COVID-19.

In the second cluster, the word "social distancing" appeared the most, apart from the community and the government's appeal to keep their distance and stay away from crowds. However, it turns out that social distancing has also emerged because many people have not obeyed the rules of social distancing, and public protests against unscrupulous people's representatives who do not comply with the health protocol, such as: "The government does not think about the people at all, both from a social and humanitarian perspective. Like they don't know social distancing, for example after the demonstration they don't know what's wrong with covid 19", "For a moment, three months ago the government and the Indonesian House of Representatives were active in campaigning for social distancing. Still, they created a new cluster of COVID-19 spread by impressing the law without realizing it. Controversy", "We failed to focus. Where are the crowds who refused to gather together yesterday to reject people who are not social distancing, who wear masks on their chins? Where are you? :((«, «People were told to do social distancing. You instead told people to do social distancing. demonstration.", "The government just wastes social distancing, *asu*." There are also quite extreme public opinions such as Social distancing, quarantine, self-isolation, and other rules are nothing more than hidden prisons."

The word "social media" arose because news about covid was widely conveyed on social media, both news on the development of COVID-19 to hoax news about COVID-19, such as: "The government can also apply empathy communication ethics in the COVID-19 pandemic. to reach

the public, including by utilizing social media.", "Let's use social media wisely so as not to be deceived by hoaxes circulating!", "Continue to monitor our social media channels to watch the video! Be positive in using social media", "Active Role Doctors and Researchers on Social Media Help Prevent the Spread of COVID-19 Hoaxes", "During a pandemic like today, everyone is looking for information about the development of the coronavirus, don't let us easily believe the confusing news about covid 19 on social media.", " A public figure, as well as a YouTuber, is again reaping controversy on social media regarding COVID-19", "Recently, social media has been busy with photos of the bodies of COVID-19 patients wrapped in plastic wrap. plastic. Healthy friends, have you seen the photos?", "Social Media." Even social media is also a tool for Pilkada, such as: "Optimizing Social Media in Pilkada during the COVID-19 Pandemic" to Regency/City KPUs throughout Central Java.", "The campaign was changed indirectly by using social media. And the issues that will be discussed are only 2 COVID" "Due to the COVID-19, communication, and socialization about the Regional Head Election of the Regent and Deputy Regent can only be done through Social Media."

IV. Conclusion

From a political point of view, there are interesting issues that many people talk about through Twitter. There is a public view that considers the COVID-19 pandemic a tool, interest, or political game. This includes socializing the importance of maintaining health protocols during the COVID-19 pandemic in election campaign activities. In this data group, many talks about meeting the needs of health workers. In addition to the lack of health workers, the fulfillment of the needs of these health workers is felt to be slow, such as intensive care that is too late to find. This is related to the plan to move to the new capital city so that the interests of health workers are neglected. Then, regarding policies for vaccination, PPKM, and economic recovery, the community has discussed various pros and cons. During the COVID-19 pandemic, various social activities emerged to relieve people affected by COVID-19. Many social service activities were also held. It turns out that the Police are the ones who most exist to inform their social service activities on social media. The National Police Chief even instructed all his staff to distribute social assistance for residents affected by COVID-19 directly. Social awareness has increased considerably during this COVID-19 pandemic. People help each other ease the burden on people affected by COVID-19, remind each other to limit crowding activities, and maintain distance. However, on the other hand, hoax news related to COVID-19 has emerged through social media, which is considered disturbing to the public. However, social awareness makes people wiser and smarter in dealing with hoax news. Further research can use social media analysis technology with other methods such as sentiment analysis, topic modeling, segmentation of social media users, opinion mining and others to uncover various phenomena in society.

References

[1] J. Isom, M. Walsh, and J. M. Gardner, "Social media and pathology: where are we now and why does it matter?," *Adv. Anat. Pathol.*, vol. 24, no. 5, pp. 294–303, 2017.

[2] M. Anderson and J. Jiang, "Teens, social media & technology 2018," *Pew Res. Cent.*, vol. 31, no. 2018, pp. 1673–1689, 2018.

[3] A. K. P. Nasution, "Integrasi media sosial dalam pembelajaran generasi z," *J. Teknol. Inf. Dan Pendidik.*, vol. 13, no. 1, pp. 80–86, 2020.

[4] R. Watrianthos *et al.*, *Belajar dari Covid-19: Perspektif teknologi dan pertanian*. Yayasan Kita Menulis, 2020.

[5] C. Nasution, "Kajian Hukum Model Literasi Media Dalam Menganalisa Informasi Berita Palsu (Hoax) Pada Media Sosial," *J. Huk. Responsif*, vol. 7, no. 2, pp. 157–170, 2020.

[6] D. Chaerani, M. N. Talytha, T. Perdana, E. Rusyaman, and N. Gusriani, "Pemetaan Usaha Mikro Kecil Menengah (UMKM) Pada Masa Pandemi Covid-19 Menggunakan Analisis Media Sosial Dalam Upaya Peningkatan Pendapatan," *Dharmakarya*, vol. 9, no. 4, pp. 275–282, 2020.

[7] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Indones. Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020.

[8] F. D. Marleny and M. Mambang, "Sosial Media Analisis Berbasis NLP Untuk Mempercepat Tanggap Bencana Banjir," *Temat. J. Teknol. Inf. Komun.*, vol. 9, no. 1, pp. 1–7, 2022.

[9] P. Chapman *et al.*, "Cross Industry Standard Process for Data Mining 1.0," *Step-by-step Data Min. Guid.*, 2000.

[10] P. Chapman *et al.*, "The CRISP-DM user guide," in *4th CRISP-DM SIG Workshop in Brussels in March*, 1999, vol. 1999.

[11] N. Seth, "Part 2: Topic Modeling and Latent Dirichlet Allocation (LDA) using Gensim and Sklearn," *Analytics Vidhya*, 2021. https://www.analyticsvidhya.com/blog/2021/06/part-2-topic-modeling-and-latent-dirichlet-allocation-lda-using-gensim-and-sklearn/ (accessed Oct. 09, 2021).

[12] S. Andayani and A. Ryansyah, "Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen," *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 1, no. 1, pp. 53–62, 2017.

[13] M. Chahal, "Measuring Similarity between Documents Using TF-IDF Cosine Similarity Function," *Int. J. Res. Publ. Semin.*, vol. 9, no. 1, pp. 53–57, 2018.

[14] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach," in *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, 2016, doi: 10.1109/ICEEOT.2016.7754750.

[15] D. S. Maylawati, H. Aulawi, and M. A. Ramdhani, "Flexibility of Indonesian text pre-processing library," *Indones. J. Electr. Eng. Comput. Sci.*, 2019, doi: 10.11591/ijeecs.v13.i1.pp420-426.

[16] S. Kannan *et al.*, "Preprocessing Techniques for Text Mining," *Int. J. Comput. Sci. Commun. Networks*, 2015.

[17] S. Vijayarani, J. Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining - An Overview," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. 1, pp. 7–16, 2015, [Online]. Available: http://www.ijcscn.com/Documents/Volumes/vol5issue1/ijcscn20150 50102.pdf.

[18] Kompas.com, "Sempat trending di Twitter, ini Rekam Jejak Said Didu," *kompas.com*, 2020. .

.