Analysis of Mental Health During the Covid-19 Pandemic in Indonesia Using Twitter Data

Nurriasih Fatimah, Indra Budi, Aris Budi Santoso and
Prabu Kresna Putra

September 29, 2021

Analysis of Mental Health During the Covid-19 Pandemic in Indonesia using Twitter Data

Nurriasih Fatimah, Indra Budi, Aris Budi Santoso, Prabu Kresna Putra
Faculty of Computer Science
University of Indonesia
Jakarta, Indonesia
nurriasih.fatimah@ui.ac.id, indra@cs.ui.ac.id, aris.budi@ui.ac.id, prabu.kresna@ui.ac.id

***Abstract*— Covid-19, which has infected Indonesia, has had a significant impact on Indonesia in various sectors and has a direct psychological impact on the entire community, such as a fear attack, anxiety, stress, and depression. Not being able to meet friends, study and work from home, the existence of the PSBB policy, the large number of news and hoaxes about Covid-19, and worrying about being infected are some of the factors that can cause psychological problems. At this time, social media was helpful to get the latest information, share various content, tell stories, and express opinions or thoughts. This study will conduct a classification and analysis related to mental health during the pandemic using tweets shared by Indonesian users and then compare the algorithms, which are Naïve Bayes, SVM, Logistic Regression, and Random Forest. From the labeling process, 612 tweets indicate psychological problems, and 168 tweets indicate anxiety problems. This study succeeded in building two classification models to detect psychological problems and anxiety problems. Model 1 was built using the Naïve Bayes because Naïve Bayes algorithm has the highest results of all evaluations with 74.36% accuracy, 74.28% precision, 74.35% recall, and 74.30% f1-score. While model 2 was built using SVM algorithm because SVM has the highest score for accuracy with 76.42%, precision with 74.91%, and f1-score with 75.19%.**

Keywords*— *Machine Learning, Twitter, Mental Health, Covid-19

I. INTRODUCTION

Coronavirus disease-19 (Covid-19) is a respiratory tract infection caused by a new type of coronavirus, namely 2019 novel coronavirus (2019-nCov) [1], has infected more than 200 countries in the world, including Indonesia [2]. This virus has had a significant impact on Indonesia in various sectors and also has a direct impact on the physical, social, and psychological of entire people [3]. The fear attack, anxiety, stress, and depression [4], [5], [6] are a few examples of psychological impacts that can be felt by people due to the presence of Covid-19.

The Indonesian Psychiatrists Association (PDSKJI) surveyed 2,364 respondents from all over Indonesia to know the psychological problems that many Indonesians experience during the pandemic, such as anxiety disorders, depression, and trauma. The survey showed that 68% of respondents experienced anxiety disorders, 67% of respondents experienced depression, and 77% of respondents experienced psychological trauma [7]. Litbangkes (National Institute of Health Research and Development Indonesian Ministry of Health) data also records that 6.8% of Indonesia's 260 million people experience anxiety disorders during this pandemic [8], [9]. These psychological problems not only occur in adults but can also occur in children and adolescents. Not being able to meet friends, work and study from home, the existence of the PSBB policy, the large number of news and hoaxes about Covid-19, and worrying about being infected are some of the factors that can cause psychological problems [8], [10], [11].

At the time of the Covid-19 pandemic, social media was very helpful to get the latest information or news regarding the current conditions. However, information, news, and hoaxes that circulate a lot are often a burden on the minds of some people, especially for people who have psychological problems [12] so that this is one of the causes of psychological problems during this pandemic [13], [14].

Social media is not only used to view information or news, but users can also freely share various content, tell stories, and express their opinions or their thoughts. Many users share stories about what they experienced during this pandemic, such as experiences when they were infected, experiences when they had to go to school or work from home, and tell stories about mental health problems they experienced. One of the social media that is widely used by Indonesians to share content is Twitter. The evidence is Indonesia in the seventh position of the world's largest Twitter users with 13.2 million users, shown in Fig. 1 [15].

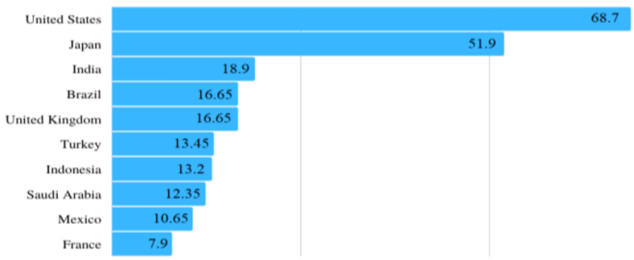

Fig. 1. The World's Largest Twitter Users

With a large number of users, Twitter can provide a source of information that can be use to conduct various studies, one of which is text classification, where content or tweets shared by Twitter users can be classified into various categories based on the research topic. This study will conduct classification and analysis related to mental health during the pandemic using tweet data shared by Twitter users to analyze the psychological conditions experienced by Indonesians during the pandemic so that it can help related organizations, such as the Indonesian Psychiatrists Association (PDSKJI), the Indonesian Psychological Association (HIMPSI), and the Ministry of Health in dealing with mental health in Indonesia, especially during this pandemic.

This study also aims to compare the algorithms used in this study, which are Naïve Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF). We choose NB, SVM, LR, and RF because many previous studies used those algorithms as the classifiers, especially study that analyze mental health. Study [16] compared 6 algorithms in their study, there are Decision Tree, NB, KNN, SVM, LR, and RF, and found that NB, SVM, LR, and RF have the best results with the same result, that is 90.3%. Also, in the study [17], SVM has the highest performance than NB, LR, and RF, while in the study [18], NB has the best result than SVM.

The paper is organized as follows. The study literature and the related works are discussed in Section II. Section III explains the stages of the research carried out. Section IV discusses the research result. Lastly, Section V discusses the conclusion and the suggestions for future works.

II. Literature Review

A. *Mental Health during the Covid-19 Pandemic*

Mental health is one of the problems faced when the Covid-19 pandemic hits [7]. Many people feel afraid and worried about being infected, stressed because they are required to face the new normal that must be carried out every day, and anxious about the news that appears both on the TV and the internet. The existence of Large-scale Social Restrictions (PSBB) regulations that are implemented in various regions in Indonesia also makes people have to be patient to take a vacation to relieve their fatigue from their routine or meet their families in their hometowns [19]. Apart from restricting people from traveling, PSBB also limits operating hours and the number of shops allowed to operate in a shopping centre [20], which causes many merchants to be unable to open their shops as usual. This, of course, has an impact on decreasing income for merchants. Currently, many companies have laid off their employees due to reduced activity at the company and to save on company operating costs, which has led to an increase in unemployment [21]. Economic problems that occur, as previously described, can also cause psychological problems for some people [22].

Many sufferers of Covid-19 also experience negative stigma, which can worsen the sufferer's condition not only physically but can also worsen the sufferer's condition psychologically, such as causing anxiety, stress, and even depression [23]. Doctors, nurses, or other health workers who help deal with Covid-19 have also experienced mental health problems. Many health workers experience stress, fatigue, and sleep problems due to overworking, the increasing number of confirmed cases, lack of contact with family, discrimination from residents around where they live, and fear of carrying and spreading the virus when returning home [24], [25]. Based on a survey conducted by the study [25], 1,563 health professionals showed symptoms of depression, anxiety, and sleep problems.

B. *Twitter*

Twitter is a social media that was founded by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams in March 2006 and officially launched on July 13, 2006 [26]. On Twitter, there are 1.3 billion accounts, 330 million monthly active users, and about 500 million tweets are sent every day [27]. The tweet is a content that is shared via Twitter, which can be text, image, or video. Initially, a tweet has a maximum length of 140 characters, then changed to 280 characters in 2017 [28]. Twitter is also included as microblogging because of the conciseness of the content it shares [29].

Apart from tweets, there are several features on Twitter, namely:

1. Username is a Twitter username that starts with the character @ and is followed by a username (@username).
2. Mention is used to call or mark other users when sending or replying to a tweet.
3. Retweet is used to send or share other users' tweets. This can indicate that a user supports or agrees with those tweets.
4. Like is used to like a tweet, either personal tweets or other users' tweets, and other users can see the tweets that are liked.
5. Bookmarks are used to tag a tweet and only can be seen by the account owner.
6. Hashtag is used to group a topic that is being discussed, using the character # (#topic).
7. Trending topic is a topic that is being discussed a lot by Twitter users. Usually comes from a hashtag, but can also come from a certain word.

Twitter also provides the opportunity for its users to conduct research using data from Twitter. So, Twitter provides an API facility that users can use to collect large amounts of tweets.

C. *Related Works*

Related works to mental health analysis on Twitter were used as a reference or basis for doing this research. Research [30] discusses the analysis on Twitter to determine a person's mental health. This study used Tweet data retrieved by the Twitter API with specific words or emotional keywords used by people with psychological problems. Those tweets are then manually labeled as positive or negative. Classification is carried out using two approaches. The first is unsupervised approach uses the MonkeyLearn API and Naïve Bayes algorithm, while the second is supervised approach uses SVM. This study also doing some analysis about tweeted source, day and time of users upload the tweet, word cloud, and users' location.

Research [17] discusses the analysis of mental health problems of social media users on Twitter. The data used were collected using the Twitter API with hashtags related to depression and anxiety as the searching keywords. They used positive, negative, and neutral as the labels. This study uses TF-IDF for feature extraction and the Naïve Bayes, SVM, Random Forest, and Regression as the algorithm. The results of this study indicate that the SVM algorithm has the best results when compared to the other three algorithms with 92.1% for the accuracy. This study also doing some analysis, there are word frequency and word cloud.

Research [16] conducted an analysis of depression in the Bengali Community on social media. Data collected from Facebook, Twitter, and other media. Feature extraction is done with TF-IDF and CountVectorizer, while the algorithms used are Decision Tree, Naïve Bayes, KNN, SVM, Logistic Regression, and Random Forest. The accuracy obtained is 90% for the average of the six algorithms, while the highest accuracy is 90.3% for Random Forest, SVM, Naïve Bayes, and Logistic Regression.

Research [18] discusses depression detection on Twitter by analyzing the emotions of users' tweets. Collected tweets will be labeled as neutral or negative. The classifier used is SVM and Naïve Bayes, and CountVectorizer for feature extraction. The result obtained from this study is Naïve Bayes has the best result than SVM.

III. METHODOLOGY

This section will explain the stages of the research carried out, as in Fig. 2 are data collection, data filtering, data labeling, data preprocessing, feature extraction, classification, evaluation, and analysis.

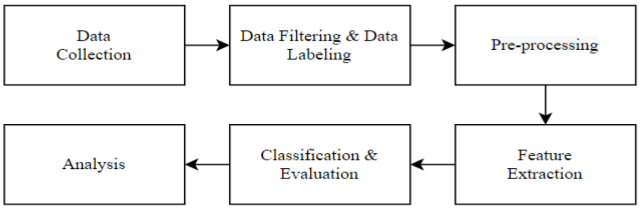

Fig. 2. Research Stages

A. Data Collection

The data was collected using a crawling technique using the Twitter API and tweepy library. The data that were collected are tweets in Bahasa Indonesia, with the search queries are queries related to psychological problems that can be felt by the community during this pandemic, there are anxiety, *cemas*, depression, *depresi*, stress, *stres*, and covid. In this search, query anxiety, depression, and stress were also used as search queries, because Indonesian Twitter users often wrote these words in English. Covid queries are also used to get tweets about psychological problems felt by the people due to the Covid-19 pandemic. Data were collected from 13 November to 31 December 2020 and obtained 6,746 data tweets.

B. Data Filtering dan Data Labeling

In the filtering process, duplicate tweets (i.e., retweeted by other users) and tweets that use local languages (i.e., Malay, Javanese, and Sundanese) are removed. There are 1170 tweets obtained from the filtering process, then be manually labeled in two stages. The first stage is to determine the tweets that indicate psychological problems. Then, tweets that indicate psychological problems from the first stage will be used for the second stage to see tweets that indicate anxiety problems. Each dataset has two classes, positive and negative. Positive is for tweets that indicate psychological problems and indicate anxiety problems, while negative is for tweets not indicate psychological problems and do not indicate anxiety problems. We proposed the second stage based on the result of the PDSKJI survey, in which is 68% of 2364 respondents experienced anxiety disorders, which is the highest result of the survey. Also, Litbangkes data records that 6.8% of Indonesia's experience anxiety disorders during the pandemic. The labeling process is done by the researcher and validated by a person who has a bachelor's degree in Psychology to ensure that the label is correct.

In the labeling process, a tweet must be view carefully because a tweet that contains the words stress, depression, or anxiety can be a tweet that does not indicate psychological problems, which the example can be seen in Table I. In a tweet, some words or emotions can indicate a person's psychological state [31], [32] even though many Twitter users clearly express that they are experiencing psychological problems, such as:

1) *capek, lelah* (tired) and *bosan* (bored) with the daily routine or work (e.g., *Capek tau, kerja di Rumah Sakit pas covid begini, stress* (you know, it's very tiring working at the hospital during this covid, stress)).
2) *capek, lelah* (tired), *bosan* (bored), or *putus asa* (hopelessness) of life can refer to a tweet that express a sense of depression [18] (e.g., *Aku sudah sampai di tahap yang gak tahu mau buat apa dengan hidup ini* (I've reached the stage where I don't know what to do with life)).
3) excessive of *cemas, gelisah* (anxious), *takut* (fear), or *parno* (paranoid) can refer to a tweet that express a sense of anxiety [33], [34] (e.g., *Pas dengar berita tentang mutasi baru covid -19 naik kegelisahan aku tuh. Tiba-tiba takut dengan segala hal* (When I heard the news about the new covid -19 mutation, my anxiety increased. Suddenly afraid of everything)).

TABLE I. EXAMPLE OF NORMAL TWEET

Types	Tweet
a sarcasm to insinuate or insult someone	*Bos besar stress berat tidak bisa ngatasi covid, FPI di bubarkan agar konsen ngatasi covid* (Big boss so stressed that he can't handle Covid, FPI is disbanded to concentrate on handling Covid)
a quote	*"Gejala yang lain adalah seperti nyeri sendi, nyeri otot, itu juga bisa muncul…," terang Agus* ("Other symptoms, such as joint pain, muscle pain, can also appear…," said Agus)
tweet to encourage someone who is affected by Covid-19	*Dengar teman kena covid, semangat kawan! Jangan merasa malu, jangan stress! Kamu pasti bisa pulih* (Hear that my friend affected by covid, cheer up, friend! Don't feel ashamed, don't stress! You can definitely recover)
Twitter users responding to new symptoms of Covid-19	*Ini gejala covid apa stress? Sama aja* (Is this the symptom of Covid or the symptom of stress? there is no difference)

C. Data Preprocessing

This stage is the most basic stage of text classification, used to clean the tweets so that they become structured words. At this stage several processes are carried out, such as:

- data cleansing, is done to remove unnecessary characters, such as numbers, punctuation marks, and special characters.
- case folding, changing all letters in a sentence into lowercase.
- tokenizing, breaking a sentence into tokens.
- stopword removal, delete the general words and the words that have no meaning. We used Indonesian stopwords list from the Sastrawi library.

- stemming, changing a word into its basic form. In this study, we used the Sastrawi library to conduct the stemming process.

D. *Feature Extraction*

Extraction of the features to be used is done using CountVectorizer. CountVectorizer works by calculating the frequency features contained in the document [35]. The purpose of using CountVectorizer is to help convert text into vector [36], where the vector has several dimensions [37]. Research [36] compared several feature extraction techniques, including CountVectorizer, TF-IDF, and Word2Vec, and found that the technique using CountVectorizer had the best performance.

E. *Classification and Evaluation*

This stage will classify the tweet data that has been labeled. The data then divided into 80% to be used as training data and 20% to be used as testing data. There are 936 training data and 234 testing data. Classification is carried out using Naïve Bayes (NB), SVM, Logistic Regression (LR), and Random Forest (RF). The feature that will use is the CountVectorizer Unigram. The evaluation of this study is measuring accuracy, recall, precision, and f1-score. The evaluation results obtained from the three classification algorithms are then compared to determine the most effective algorithm.

F. *Analysis*

We do some analysis, which are common words, word clouds, the frequency of time for users to upload tweets, and the users' location. For the frequency of time for users to upload tweets, we categorize the time into 4 periods, morning (03.00 – 08.59), afternoon (09.00 – 14.59), evening (15.00 – 20.59), and night (21.00 – 02.59). For the user's location, we analyze location from users' profiles using NLP-ID library's POS (Part-Of-Speech) tagger [38] to extracts the users' location because users often write words besides place names. Also, extraction is carried out with phrases POS tagger that can tokenize single or multi-words because there are location names in Indonesia that consist of more than one word, e.g., DKI Jakarta and *Jawa Tengah*. We select words and phrases that have NNP and NP tags. The NNP tag is a tag for proper nouns, such as place names (e.g., Jakarta and Indonesia), while the NP tag is a tag for noun phrases (e.g., *Jakarta Barat* and *Sumatera Utara*). The analysis was also done by built word clouds and graphs using matplotlib.

IV. Result and Discussion

This section will explain the results of this study and their discussion. Table II shows the results of the labeling process, for Model 1 there are 612 tweets indicate psychological problems, while 558 tweets did not indicate psychological problems. Also, for Model 2, there are 168 tweets indicate anxiety problems, while 444 tweets did not indicate anxiety problems.

Table III shows the comparison of the evaluation results. The Naïve Bayes (NB) algorithm has the highest score for all evaluations for Model 1 with 74.36% accuracy, 74.28% precision, 74.35% recall, and 74.30% f1-score, while the lowest score for all evaluations are measured by SVM algorithm with 71.37% accuracy, 71.54% precision, 71.36% recall, and 71.42% f1-score. For Model 2, SVM has the highest score for accuracy with 76.42%, precision with 74.91%, and f1-score with 75.19%, while the highest score for precision is LR with 81.75%. NB, LR, and RF has the same score for accuracy, 75.61%, which are the lowest score. The lowest precision score is measured by NB and RF with 73.83%. Also, NB, LR, and RF have the lowest score for recall with 75.60% and for the lowest f1-score is measured by LR with 67,74%.

TABLE II. Label Distribution

Label	Total
First Stage or Model 1	
Indicate psychological problems	612
Not indicate psychological problems	558
Second Stage or Model 2	
Indicate anxiety problem	168
Not indicate anxiety problem	444

TABLE III. The Evaluation Result

Evaluation	NB	SVM	LR	RF
Model 1				
Accuracy	74.36%	71.37%	73.50%	71.79%
Precision	74.28%	71.54%	73.81%	71.92%
Recall	74.35%	71.36%	73.50%	71.79%
F1-Score	74.30%	71.42%	73.57%	71.84%
Model 2				
Accuracy	75.61%	76.42%	75.61%	75.61%
Precision	73.83%	74.91%	81.75%	73.83%
Recall	75.60%	76.42%	75.60%	75.60%
F1-Score	71.11%	75.19%	67.74%	71.11%

TABLE IV. Examples of Prediction Results

Sentences	Prediction
Model 1 (psychological problems)	
Stress banget sekarang, sekolah dirumah, tugas numpuk, gak bisa keluar (I'm really stressed right now, school at home, too many assignments, can't go out)	POS
Semenjak covid aku didiagnosa psikiater kena depresi minor (Since this covid pandemic, I was diagnosed by a psychiatrist with minor depression)	POS
Kalau kena covid tidak boleh stress biar cepat sembuh (If you infected with covid, don't be stress so you can get well really soon)	NEG
Model 2 (anxiety problems)	
Anxiety aku kambuh kalau lihat berita tentang covid (My anxiety relapses when I see news about covid)	POS
Kalau keluar rumah bawaannya takut, takut pulang bawa virus (I was afraid to leave the house, afraid if I back home with the virus)	POS
Enggak cemas sih cumin lagi banyak pikiran saja (I'm not worried, just have a lot of things in my mind)	NEG

The models that have been built are then tested by doing some prediction tests using text input. Table IV shows examples of prediction results of both models. Fig. 3 shows the common words without stopwords from the dataset. Covid has the highest frequency with appears 1,274 times.

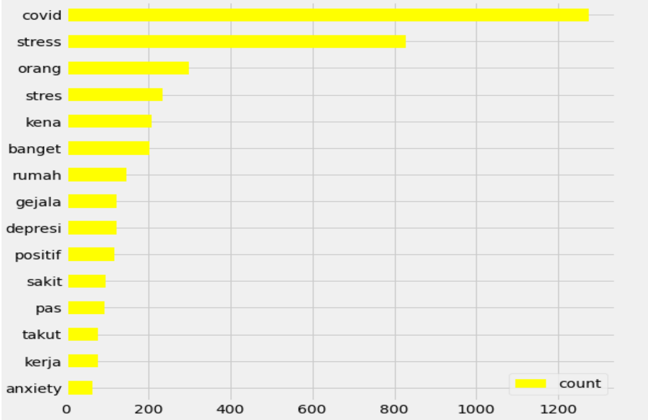

Fig. 3. Common Words in Dataset (Without Stopwords)

The word cloud of all tweets can be seen in Fig. 4. There are 500 words that users frequently mentioned in their tweets. The most mentioned words will have a large size, while the rarely mentioned words have a smaller font size.

Fig. 4 Word cloud from Dataset

Analysis of the frequency of time for users to upload tweets can be seen in Fig. 5. Users more often upload tweets during the afternoon, which there is a possibility of an interaction between users.

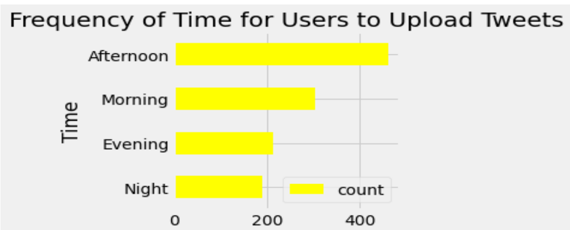

Fig. 5 The Frequency of Time for Users to Upload Tweets

Fig. 6 shows 15 most locations displayed by the users on their profile. Indonesia got the first position of the location most written by the users. In the results of the analysis, Jakarta and DKI Jakarta are included in different calculations because they have different tags. Jakarta has an NNP tag, while DKI Jakarta has an NP tag.

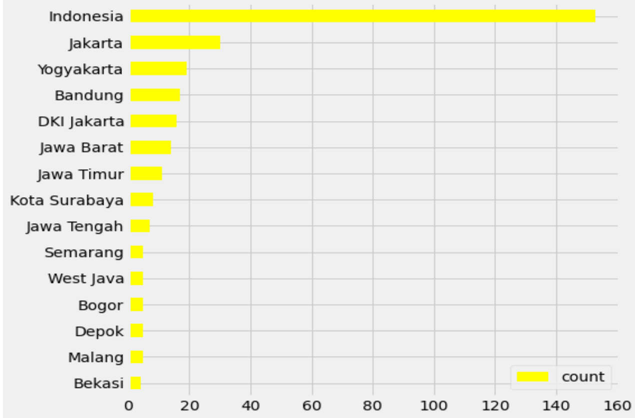

Fig. 6 Users' Location

V. Conclusion and Recommendations

Based on the result of labeling process, model 1 has 612 tweets indicate psychological problems and model 2 has 168 tweets indicate anxiety problems experienced by Indonesian Twitter users. This study succeeded in building two classification models to detect psychological problems and anxiety problems in Indonesian communities using Twitter data. Model 1 was built using the Naïve Bayes algorithm because that algorithm is the most effective algorithm among the other algorithm used in this study based on the evaluation results. The Naïve Bayes algorithm has the highest results of all evaluations with 74.36% accuracy, 74.28% precision, 74.35% recall, and 74.30% f1-score. While model 2 was built using SVM algorithm because SVM has the highest score for accuracy with 76.42%, precision with 74.91%, and f1-score with 75.19%.

Based on the analysis result of the frequency of time for users to upload tweets, it was found that the time most often used by users to upload tweets was during the afternoon, which is during that time relevant organizations can interact more with Twitter users by sharing information about mental health or open private consultation sessions for users who had psychological problems. Relevant organizations can also conduct further surveys by using more correspondents in every city, especially in the largest cities in Indonesia to see mental health is in that city or organize mental health awareness online seminars.

Due to the limited time for data collection, the data obtained is relatively small. Improving the search query for searching the tweets can be done to get more numerous tweets. Inviting professionals for doing the labeling process or validating the label can also be done to get more accurate labels and improve the accuracy score. Indonesian Twitter users also often mix informal, slang, and harsh words into one tweet. So, we suggest using Indonesian slang or harsh words dictionary to eliminate or normalize those words contained in a tweet. For future works, these four things can be a consideration for improvement. This study only analyzes general psychological problems during the pandemic, in the future, we suggest analyzing various emotions, such as fear, trust, anger, disgust, etc.

References

[1] A. E. Gorbalenya *et al.*, "Severe acute respiratory syndrome-related coronavirus: The species and its viruses–a statement of the Coronavirus Study Group," 2020.

[2] World Health Organization, "COVID-19 weekly epidemiological update, 22 December 2020," 2020.

[3] H. D. Windarwati, W. Oktaviana, I. Mukarromah, N. A. L. Ati, A. F. Rizzal, and A. D. Sulaksono, "In the middle of the COVID-19 outbreak: Early practical guidelines for psychosocial aspects of COVID-19 in East Java, Indonesia," *Psychiatry Res.*, vol. 293, p. 113395, 2020, doi: https://doi.org/10.1016/j.psychres.2020.113395.

[4] M. Akat and K. Karataş, "Psychological Effects of COVID-19 Pandemic on Society and Its Reflections on Education.," *Electron. Turkish Stud.*, vol. 15, no. 4, 2020.

[5] K. Kontoangelos, M. Economou, and C. Papageorgiou, "Mental health effects of COVID-19 pandemia: a review of clinical and psychological traits," *Psychiatry Investig.*, vol. 17, no. 6, p. 491, 2020.

[6] I. Ifdil, R. P. Fadli, K. Suranata, N. Zola, and Z. Ardi, "Online mental health services in Indonesia during the COVID-19 outbreak," *Asian J. Psychiatr.*, 2020.

[7] S. Winurini, "'Permasalahan Kesehatan Mental Akibat Pandemi Covid-19," *Info Singkat, Kaji. Singk. terhadap Isu Aktual dan Strateg.*, pp. 13–18, 2020.

[8] D. Ramdhani, "Pasien Gangguan Cemas di RSJ Jabar Meningkat Selama Pandemi," *Kompas.com*. 2020, Accessed: Jan. 02, 2021. [Online]. Available: https://bandung.kompas.com/read/2020/10/08/09265741/pasien-gangguan-cemas-di-rsj-jabar-meningkat-selama-pandemi.

[9] The Jakarta ost, "Mental health services disrupted in 93% of countries during COVID-19 pandemic: WHO - National - The Jakarta Post," *The Jakarta Post*. 2020, [Online]. Available: https://www.thejakartapost.com/news/2020/10/08/mental-health-services-disrupted-in-93-of-countries-during-covid-19-pandemic-who.html.

[10] J. Xiong *et al.*, "Impact of COVID-19 pandemic on mental health in the general population: A systematic review," *J. Affect. Disord.*, vol. 277, pp. 55–64, 2020, doi: https://doi.org/10.1016/j.jad.2020.08.001.

[11] S. Mukhtar and W. Rana, "COVID-19 and individuals with mental illness in psychiatric facilities," *Psychiatry Res.*, 2020.

[12] World Health Organization, "Mental health and psychosocial considerations during the COVID-19 outbreak, 18 March 2020," World Health Organization, Geneva PP - Geneva. [Online]. Available: https://apps.who.int/iris/handle/10665/331490.

[13] A. Casero-Ripolles, "Impact of Covid-19 on the media system. Communicative and democratic consequences of news consumption during the outbreak," *El Prof. la Inf.*, vol. 29, no. 2, 2020, doi: http://dx.doi.org/10.3145/epi.2020.mar.23.

[14] S. T. Ahmed, "Managing news overload (MNO): The COVID-19 infodemic," *Inf.*, vol. 11, no. 8, 2020, doi: 10.3390/INFO11080375.

[15] J. Clement, "Countries with most Twitter users 2020." Statista, 2020.

[16] D. B. Victor, J. Kawsher, M. S. Labib, and S. Latif, "Machine Learning Techniques for Depression Analysis on Social Media-Case Study on Bengali Community," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 1118–1126.

[17] M. Faryal, M. Iqbal, and H. Tahreem, "Mental Health Diseases Analysis on Twitter using Machine Learning," *iKSP J. Comput. Sci. Eng.*, vol. 1, no. 2, pp. 16–25, 2021.

[18] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017, pp. 858–862.

[19] World Health Organization, "Mental health & COVID-19," 2020. [Online]. Available: https://www.who.int/teams/mental-health-and-substance-use/covid-19.

[20] R. A. Umasugi, "Mal di Jakarta Kemungkinan Ditutup Saat PSBB Kembali Diterapkan," *Kompas.com*. 2020, Accessed: Jan. 12, 2021. [Online]. Available: https://megapolitan.kompas.com/read/2020/09/10/13441011/mal-di-jakarta-kemungkinan-ditutup-saat-psbb-kembali-diterapkan.

[21] CNN Indonesia, "Rincian Angka Pengangguran di Setiap Daerah per Agustus 2020," *CNN Indonesia*. 2020, Accessed: Jan. 10, 2021. [Online]. Available: https://www.cnnindonesia.com/ekonomi/20201106093653-532-566598/rincian-angka-pengangguran-di-setiap-daerah-per-agustus-2020.

[22] A. Bacigalupe and A. Escolar-Pujolar, "The impact of economic crises on social inequalities in health: what do we know so far?," *Int. J. Equity Health*, vol. 13, no. 1, p. 52, 2014.

[23] W. Sulistiadi, S. R. Slamet, and N. Harmani, "Handling of Public Stigma on COVID-19 in Indonesian Society," *Kesmas J. Kesehat. Masy. Nas. (National Public Heal. Journal)*, 2020.

[24] D. Lu and J. Bouey, "Public Mental health crisis during COVID-19 pandemic, China," *Emerg. Infect. Dis.*, vol. 26, no. 7, 2020.

[25] A. Pragholapati, "MENTAL HEALTH IN PANDEMIC COVID-19," *Available SSRN*, vol. 3596311, 2020.

[26] V. Ahuja and M. Shakeel, "Twitter presence of jet airways-deriving customer insights using netnography and wordclouds," *Procedia Comput. Sci.*, vol. 122, pp. 17–24, 2017.

[27] M. Ahlgren, "50+ Twitter Statistics & Facts For 2021 You Should Know About," *Website Hosting Rating*. 2020, Accessed: Jan. 02, 2021. [Online]. Available: https://www.websitehostingrating.com/twitter-statistics/.

[28] A. Rosen, "Tweeting Made Easier," *Twitter Blog*. 2017, Accessed: Jan. 02, 2021. [Online]. Available: https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html.

[29] A. M. Kaplan and M. Haenlein, "The early bird catches the news: Nine things you should know about micro-blogging," *Bus. Horiz.*, vol. 54, no. 2, pp. 105–113, 2011.

[30] C. L. Rakshitha and S. Gowrishankar, "Machine Learning based Analysis of Twitter Data to Determine a Person's Mental Health Intuitive Wellbeing," *Int. J. Appl. Eng. Res.*, vol. 13, no. 21, pp. 14956–14963, 2018.

[31] D. Gruda and S. Hasan, "Feeling anxious? Perceiving anxiety in tweets using machine learning," *Comput. Human Behav.*, vol. 98, pp. 245–255, 2019.

[32] A. Li, D. Jiao, and T. Zhu, "Detecting depression stigma on social media: A linguistic analysis," *J. Affect. Disord.*, vol. 232, pp. 358–362, 2018.

[33] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," *Curr. Opin. Behav. Sci.*, vol. 18, pp. 43–49, 2017.

[34] P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, and J. Boyd-Graber, "Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter," in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015, pp. 99–107.

[35] S. C. Eshan and M. S. Hasan, "An application of machine learning to detect abusive bengali text," in *2017 20th International Conference of Computer and Information Technology (ICCIT)*, 2017, pp. 1–6.

[36] C.-G. Cuşmuliuc, L.-G. Coca, and A. Iftene, "Early Detection of Signs of Anorexia in Social Media," in *5th Proceedings of the Conference on Mathematical Foundations of Informatics*, 2019, pp. 3–6.

[37] M. G. Hussain, M. R. Hasan, M. Rahman, J. Protim, and S. Al Hasan, "Detection of Bangla Fake News using MNB and SVM Classifier," in *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, 2020, pp. 81–85.

[38] Z. Juwantara, F. Eddy, D. H. Sasmita, T. N. Khoolish, and B. Aldiyansyah, "Kumparan NLP Library." Zenodo, Feb. 2021, doi: 10.5281/zenodo.4556870.