7th International Conference on Computer Science and Computational Intelligence 2022

Their post tell the truth: Detecting social media users mental health issues with sentiment analysis

Haris Herdiansyah[a], Rusdianto Roestam[a], Richard Kuhon[a], Adhi Setyo Santoso[a] *

[a]President Univesity, Jababeka Education Park, Jl. Ki Hajar Dewantara, Bekasi West Java 17530, Indonesia

Abstract

Mental health disorders remain a problem that always appears throughout the ages because its originators are related to everyday social phenomena that are always changing. One of the serious obstacles is cultural factors that view people with mental health disorders as people who cannot function fully, need to be avoided, have problems, and are given a negative social stigma. On the other hand, people with mental health disorders need a comfortable space where they can express their emotions and thoughts. Social media such as Twitter is one of the potential cathartic media for people with mental health disorders. This study aims to identify mental health disorders through words or tweet narration with the keywords "emotions", "hallucinations", "panic", "mental illness", "stress", and "fear". 5537 clean tweet data were collected from the Indonesian population containing these keywords using rapidminer sentiment analysis which were categorized into three categories; Positive, Negative, and Neutral. The results of the analysis are strengthened by searching randomly sampled text tweets. As a result, it is proven that social media Twitter is effective in identifying symptoms of mental health disorders, and Twitter is considered a safe and comfortable cathartic medium for people with mental health disorders.

1. Introduction

The issue of mental health is an issue that will never subside because it is not only related to mental disorders suffered. Mental health is closely related to the stigma of society who have not fully regarded people with mental disorders as people who need help. In the context of Indonesian culture, mental disorders are still considered a shameful thing, a disgrace of the family, and interfere with self-esteem and good name in the society.

---

* Corresponding author.
  E-mail address: adhi.setyo@president.ac.id

Cultures and perspectives that still view negatively towards people with mental disorders are still widely prevailing in various countries in Asia. For example, in Arab culture, Arabs view people with mental disorders as a disorder that interferes with social relations, giving shame to their family, and the sufferer of a mental disorder is perceived as troubled individuals [1]. This is also still felt in many other cultures, especially in Asian nations. With conditions that are less socially acceptable, everyone with mental disorder is not easy to express themselves. There are certain psychological barriers and worries if the state of his mental disorder is known to others, then he will be viewed as bad and negative. Once the stigma is accepted, it will be difficult to let go. Therefore, people will relatively avoid this negative stigma.

Nevertheless, people with mental health disorders need space to express their emotions. They will look for spaces that are considered safe to express their emotions without the threat of a negative social stigma [8]. In this digital era, people can choose a variety of media to express themselves in cyberspace, especially using social media, because of several considerations, including; 1) cyberspace does not present itself directly in front of the public, 2) expression can take various forms ranging from videos, writings, images, and others, 3) even if there is criticism or innuendo, it is asynchronous in nature and does not directly attack the personal self, 4) can be filtered against comments that are not pleasant, 5) can control anyone who sees the video or writing what we upload [2]. Uploads or expressions of emotions of people with mental disorders through social media are considered more appropriate, efficient, and more accurate to be used as a database than traditional methods such as self-report surveys or interviews [1] .

Detecting Arabic Depressed Users from Twitter Data" using sentiment analysis as a tool to detect depressive disorders in Arab society on twitter platform [1]. The results of their research show that sentiment analysis is suitable for mapping the emotional expression of people with depressive disorders where the local culture is antipasti or not allowing people with depression disorders to express their emotions more freely. Based on that research, this study has a similar condition where in the context of Indonesian culture, people with mental disorders are relatively more difficult to express their emotions due to the existence of negative pressure and social-labeling towards people with mental disorders. For this reason, sentiment analysis will be used in this study to capture and identify the emotional expression of people with mental health disorders in the Indonesian context.

Sentiment analysis is one of the most important research trends in the world of health (and/or mental health) that can identify disease characteristics through textual analysis of social media [1, 9]. The general purpose of this study is to provide examples of real populations who use social media platform (Twitter) in Indonesia to gain knowledge about a person's psychological and mental health conditions. This can improve identification and early intervention so that treatment or can be started earlier; potential outcomes can be improved, and the stigma associated with mental illness can be minimized.

The stages in this study will be divided into two sessions, namely; First, researchers identified a collection of data (tweets) from Indonesian people using Bahasa Indonesia, by detecting it with keywords that we have classified that are related to the condition of mental health disorders. When a tweet contains words containing these keywords appear, the we classify it into two categories, namely Positive (having a mental health condition) or Negative (has no mental health disorder condition). Second, researchers conducted deeper identification to compare in more detail the behaviors (represented by narratives/words in respondents' tweets) to make more contrasting comparisons between Positive and Negative groups.

The article is structured with the following composition: section 2 presents an overview of mental health and the main concepts related to sentiment analysis. Section 3 is about the research methods used in this research. Section 4 presents the results of data analysis. Section 5 contains the conclusion of the research results and recommendations.

2. Background and related work

2.1 Mental health overview

Mental health is defined as a prosperous condition where every individual understands and realizes their potential, can go through life crises normally, can work productively, and can make a positive contribution to their communities and society at large [6]. Mental health leads to a fully-functioning person where each individual is able to express his potential to achieve life goals, gain happiness (happiness), and develop healthy social relationships. Having healthy mental health means that individuals have full awareness of their abilities. Individuals can adapt and manage daily

stress, are able to think positively and are able to contribute to the wider community and society. Conversely, in disturbed mental health conditions, individuals will have difficulty adapting to all the challenges they face, think negatively and pessimistically about their abilities, and are unable to achieve their life goals [5, 10].

Every day we are often faced with various stressors ranging from mild to severe. Although everyone has the ability to deal with stressors, these abilities differ from one person to another. In addition, the intensity of stressors also varies, some are easy to solve but some are difficult. If a person faces a stressor that he considers difficult, and the condition lasts for a certain period of time, then the condition begins to interfere with his activities and social relationships, that's when mental health disorders appear. Thus, everyone has the potential to experience mental health disorders [4, 9].

Disrupted mental health conditions require a cathartic space for emotional expression and relieve stressor pressure. However, the public space does not necessarily provide comfort for people with mental health disorders to express their emotions because in certain cultures (such as in Indonesia) people with mental disorders are still correlated with negative stigma, are still considered a disgrace or a socially embarrassing condition. In the midst of the need for people with mental health disorders to express themselves, there is a space that is quite potential and relatively safe from negative stigma, namely the social media space. For this reason, everyone who begins to experience mental health disorders, takes advantage of this social media space as the chosen cathartic forum.

2.2 Sentiment Analysis

Sentiment analysis (SA), also known as opinion mining, is a process that applies natural language processing (NLP) text analysis and machine learning to identify the sentiment of the text as positive, negative or neutral [1, 3]. So basically this analysis technique is development of NLP theory. By analyzing people's opinions, SA uncovers people's sentiments towards different topics, products, services, events, issues, etc [8]. An opinion is composed of five parts: an object (i.e., the target of the opinion), feature (i.e., the object's attribute), sentiment/opinion orientation (i.e., positive, negative, neutral), opinion holder (i.e., person/organization that expresses the opinion), and time (i.e., time opinion was expressed) [7].

There are two basic approaches in studies that use sentiment analysis: the supervised approach and the unsupervised approach. The supervised approach is typically machine learning and the unsupervised approach is the lexicon-based approach. The supervised approach uses corpus data with labels (positive/negative) to train a set of classifiers [1]. The classifiers used are machine learning algorithms such as support vector machine (SVM), naive Bayes (NB), artificial neural network (ANN), and k-nearest neighbor (KNN) (Boudad et al and Shoukry et al, in Almouzini, 2019) . On the other hand, the unsupervised approach or the lexicon-based approach determines the orientation (polarity) of each word entry as a numerical value. Positive and negative sentimental words use both positive and negative values, with a value of zero to refer to neutral words [2].

2.3 Related work

This section presents a review of previous research related to the detection of mental health disorders on social media. With this review, a more comprehensive picture of the implementation of this research will be obtained. The main characteristics of this study will be summarized in Table 1.

Two strategies used by researchers in collecting data in detecting the presence or absence of mental health disorders from respondents are divided into two ways; first, detecting mental health disorders through social media, in this case is tweet data obtained via twitter from respondents, namely the Indonesian people, and secondly, to identify more deeply from the tweet data whether the respondent in question has explicitly described himself as an individual with mental health disorders.

The two methods above were adopted by researchers from a previous study entitled "Detecting Arabic Depressed Users from Twitter Data" conducted by Salma Almouzini, Maher khemakhema, Asem Alageel which was published at the 16th International Learning & Technology Conference 2019. Although there are differences in terms of respondents, there are similarities in conditions. Arab community considers people with depression and mental disorders culturally to get a negative social stigma [1]. This condition is similar to the conditions that occur in Indonesian culture where individuals who have mental health disorders will be ostracized, considered a disgrace, and get a negative stigma. Another research is a study conducted by Park et al with the title "Depressive Moods of Users

Portrayed in Twitter" which identifies depressive moods through tweeters [11]. In addition, another research is a study conducted by Choudhury et al with the title "Social Media as a Measurement Tool of Depression in Populations." This study uses social media as a measuring tool in recognizing symptoms of depression in the population [12].

Other studies as references for this is the research conducted by Coppersmith et al with the title "Quantifying Mental Health Signals in Twitter" and research conducted by Coppersmith et al with the title "Measuring Post Traumatic Stress Disorder in Twitter". The two studies conducted by Coppersmith et al above use social media twitter which is seen as very efficient in identifying the expressions of people experiencing mental health disorders [9, 10]. Based on these considerations, this study also uses social media Twitter as a means to track and identify the presence or absence of mental health disorders of the respondents through the tweets they write. The following is table 1. Summary of the previous studies.

Table 1. Summary of the previous studies.

Researchers	Year	Source	N(Users)	N(Post)	Basis of Data Collection	Performance Measures
Park, M et al	2012	Twitter	69-participants	21103 tweets	Keyword: depression	NA
Choudhury, et al	2013	Twitter	489-participant	69,514 tweets	Twitter user's tweets	Accuracy=73% Precision =0.826
Coppersmith et al	(2014a)	Twitter	21,866 -participants	NA	Twitter user's tweets	Precision Bipolar=0.48 Depression=0.64 PTSD=0.67 SAD=0.42
Coppersmith et al	(2014b)	Twitter	NA	2.5 million tweets	Twitter user's tweets	Accuracy =81%

3. Methodology

3.1. Data collection

Twitter is the choice to get data which in this case is set at 10,000 twitter data. Rapidminer provides an API for pulling data from Twitter. Duplicate tweeter data such as blank text, character "@", character "?" and RT (re-tweet) and other duplicate data will be deleted using the filters available in the Rapidminer application. After the process of deleting unnecessary data above, the total data becomes 5537 twitter data (clean data) which is divided into several keywords; "emotions", "hallucinations", "panic", "mental illness", "stress", and "fear". All data resulting from this process then goes to the next stage, which is given a label through a labeling process which is done manually.

3.2 Manually labeling process

All clean data resulting from the above process must be labeled based on the following three criteria; positive (there is an indication of a mental health disorder), negative (there is no indication of a mental health disorder), and neutral (the presence or absence of a mental health disorder cannot be identified). If the tweet contains words that describe conditions related to mental health, such as; consulting a doctor, taking medication, experiencing stress, experiencing depression, and words that describe a mental disorder, it will be labeled as positive mental health disorders. If there are no words, even though the tweet contains the keyword mental illness, it will be given a negative label as having mental health disorders. In addition, if there are words that mean positive emotions such as happiness, joy, excitement,

and others, they will also be given a negative label for experiencing mental health disorders. Then, if there are keywords outside the positive or negative label, they will be labeled neutral.

3.3 Cleaning and Pre-Processing Data

Using the AffectiveTweets package, we selected the TweetToSparseFeatureVector filter which has many options for implementing data cleaning and pre-processing. The data cleaning process involves deleting words that are not related to the meaning of the sentence in the tweet. From these results, we get an overview of the number of tweets containing statements of mental health disorders.

4. Result

The following is the composition of the data divided by keywords; emotions, hallucinations, panic, paranoia, mental illness, stress, and fear, with a total data of N = 5537 tweets. In this data by keyword table, the highest to lowest percentages of keywords related to mental health disorders are "panic" (29%), "emotion" (23%), "stress" (23%), "fear" (23%), "hallucination" (1%), and "mental illness" (1%). From the data it can be compared that the highest percentage is in the keyword "Panic". This indicates that people with mental disorders generally feel panic disorder, so panic words appear dominant in respondent's tweets. Panic is a symbol of the psychological condition of people with mental disorders that respondents want to show to get public attention, ask for help, or maybe just to be expressed in social media. The second percentage order is the keywords "emotion", "stress", and "fear". These three keywords are the intensity of the severity of mental health disorders under "panic". Respondents expressed the words emotion, stress, and fear to show that they are mentally disturbed and want to be known by the public. The third percentage order is the keywords "hallucination" and "mental illness". People with hallucinatory disorders, and people who are already mentally disturbed are no longer trying to convey their message and condition to the public because the severity of their mental disorders has reached a high level which is actually getting more and more urgent to get professional help such as psychologists and psychiatrists. If hallucinations have appeared, then the individual concerned has less and less interacting with his social media. The composition of the data can be seen in Figure 1 and Table 2 below.
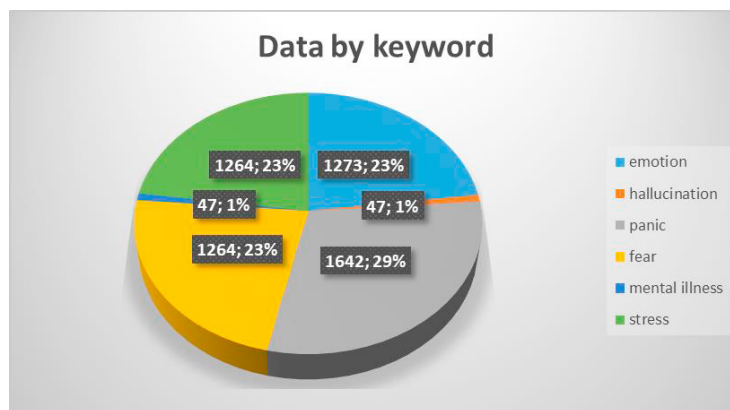

Fig. 1. Keywords Composition Data

Table 2. The composition of the Positive, Negative, or Neutral labels for each keyword.

Keyword	Positive	Negative	Neutral
Panic	791	400	450
Emotion	777	320	175

Keyword	Positive	Negative	Neutral
Stress	759	319	185
Fear	886	287	90
Hallucination	356	121	98
Mental Illness	343	141	100

Understanding mental illnesses and the warning indications of these conditions We modeled the development of linguistic indicators through time and demonstrated the significance of monitoring the language of the afflicted participants from the disease's earliest stages. Although negative sentiment is a frequent technique in neuro-linguistic programming, we argue that treating distinct emotions independently and relating them to unique cognitive types of individuals, as expressed in language indicators, is more significant. Emotions are especially important for mental health. We found different patterns that distinguish healthy users from those who have or are developing a mental disease via the computational analysis of these multi-dimensional linguistic properties across time.

5. Conclusion

The conclusion of this research, it has been proven that the use of social media space, especially Twitter as a place for people with mental health disorders to express their emotions, or as a means of self-recognition that they feel they have a mental health disorder. The most-searched-for terms included panic, emotion, stress, fear, hallucination, and mental illness. From this study, researchers built a predictive model to predict whether user tweets were able to describe subjects with mental health disorders or not based on keywords that were strongly correlated with mental health disorders. In this case, more contextual issues can be explored by leveraging the vast volume of data (Big Data) and innovative data streams produced by Twitter users' behavior on social networks.

Acknowledgement

References

[1] Almouzini. Salma, Khemakhem. Salma, and Alageel. Asem (2019). "Detecting Arabic Depressed Users from Twitter Data." 16th International Learning & Technology Conference 2019. Procedia Computer Science 163 (2019) 257–265

[2] Choudhury, Munmun De, Scott Counts, and Eric Horvitz. (2013) "Social Media as a Measurement Tool of Depression in Populations." In Proceedings of the 5th Annual ACM Web Science Conference, 47–56.

[3] Zolezzi, Monica, Maha Alamri, Shahd Shaar, and Daniel Rainkie. (2018) "Stigma Associated with Mental Illness and Its Treatment in the Arab Culture: A Systematic Review." International Journal of Social Psychiatry 64 (6): 597–609

[4] Hungerford. C, Hodgson. D, Clancy. R, Monisse. M, Bostwick. R, Jones. T., (2016)., Mental Health Care: An Introduction for Health Professionals in Australia. Melbourne: John Wiley & Sons.

[5] Levin. B.L, Hennessy. K.D, Petrila. J., (2010)., Mental Health Services: A Public Health Perspective. Oxford: Oxford University Press

[6] Rogers. A, Pilgrim. D, (2021)., A Sociology of Mental Health and Illness 6e. London: McGraw Hill

[7] Boudad, Naaima, Rdouan Faizi, Rachid Oulad Haj Thami, and Raddouane Chiheb. (2017) "Sentiment Analysis in Arabic: A Review of the Literature." Ain Shams Engineering Journal.

[8] Abdulla, Nawaf a, Nizar a Ahmed, Mohammed a Shehab, and Mahmoud Al-ayyoub. (2013) "Arabic Sentiment Analysis: Lexicon-Based and Corpus-Based." Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT'13) 6 (12): 1–6.

[9] Coppersmith, Glen, Mark Dredze, and Craig Harman. (2014) "Quantifying Mental Health Signals in Twitter." In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 51–60.

[10] Coppersmith, Glen a, Craig T Harman, and Mark H Dredze. (2014) "Measuring Post Traumatic Stress Disorder in Twitter." In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media Measuring, 579–82

[11] Park, M, C Cha, and M Cha. (2012) "Depressive Moods of Users Portrayed in Twitter." In Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD), 1–8.

[12] Choudhury, Munmun De, Scott Counts, and Eric Horvitz. (2013) "Social Media as a Measurement Tool of Depression in Populations." In Proceedings of the 5th Annual ACM Web Science Conference, 47–56

[13] Nadeem, Moin, Mike Horn, Glen Coppersmith, and Sandip Sen. (2016) "Identifying Depression on Twitter."