Identification of potential depression in social media posts

Munawar[1], Yulhendri[2]
[1]Department of Computer Science, Faculty of Computer Science, Esa Unggul University, Jakarta, Indonesia
[2]Department of Informatics, Faculty of Computer Science, Esa Unggul University, Jakarta, Indonesia

Article Info

Article history:

Received Dec 23, 2023
Revised Jan 31, 2025
Accepted Mar 15, 2025

Keywords:

Depression
Naïve Bayes
Random forest
Sentiment analysis
Social media
Support vector machine

ABSTRACT

The widespread use of social media to convey emotions (including depression) can be used to identify suspected depression in social media posts by examining the language that they have used on social media. This study aims to develop a system for detecting suspected depression in social media posts using sentiment analysis. This study collected data from X (Twitter) for three months using the keywords depression, mental health, and mental disorders. 1,502 data were generated due to the cleaning process of the 5,000 data collected. The findings of employing the validated by psychologist valence aware dictionary and sentiment reasoner (VADER) and Indonesian sentiment (InSet) lexicons demonstrate that VADER is more accurate (95.1%) than Inset (76.9%). The results of modeling with random forest, naive Bayes, and support vector machine (SVM) showed that random forest had the highest accuracy (83.3%), followed by naive Bayes (80.5%) and SVM (80.4%). Predicting social media data using lexicons and machine learning has limits that can be addressed by validation from clinical psychology. The frequency, timing, and idiom of posts on social media can reveal signs of depression. Depression seems to be best described by words like melancholy, stress, sadness, worthlessness, and depression.

Corresponding Author:

Munawar
Department of Computer Science, Faculty of Computer Science, Esa Unggul University
St. Arjuna Utara No 9, Jakarta Barat, Indonesia
Email: munawar@esaunggul.ac.id

1. INTRODUCTION

An important area of multidisciplinary research is analyzing social media posts related to psychiatric diseases like depression. Early detection of depressed symptoms through language use can help avert many negative outcomes, and can even assist in the timely implementation of effective treatment [1]. To find relief, persons with mental illnesses frequently disclose their conditions on social media [2]. Teenagers have been identified as the most frequent users of various social media networks among all users [3].

For many, social media has become the primary platform for expressing their feelings, experiences, and ideas [4], [5]. Monitoring public health [6], consumer segmentation [7], [8], product branding [9], and other areas have all benefited from an understanding of user expressions on social media. However, it is still limited in its application in mental health [10] because it provides unbiased use of language and behavior [11] regarding their feelings, including depression. Due to stigma, prejudice against specific individuals, and ignorance, those who suffer from depression and are reluctant to speak with clinic doctors may find that social media, with its choice of anonymity, provides a supportive environment [12]. Recognizing early indicators of mental disorders through social media posts can avert many adverse outcomes, such as self-harm, and suicide [13].

Depression is a medical illness characterized by emotions of sadness that can have a significant impact on a person's thoughts, actions, feelings, and mental health [14]. Depression, which is typified by losing joy over things that used to make fun of it, is a sickness that is becoming more and more prevalent these days [15]. Estimated 16 million Indonesians suffer from depression (6.1% of the population in 2019) [16]. Addiction to social media [17] and suicidal thoughts [18] are consequences of depression.

Social media posts can reveal whether or not someone is depressed, which has led to several studies on depression on social media [2]. Social media posts' expressions and emotions are related, and this relationship can be used to predict depression [10] by extracting emotion from social media posts [19]. The degree of depression in user posts on social media can be determined by the volatility of negative emotions in this context [20]. Because of this, sentiment analysis can be used to identify posts that are neutral, positive, or negative [21]. However, it can be argued that research on social media and depression is still in its infancy. Further research is required to better understand the connection between depression and social media posts [4], including what idioms can be used to describe depression [2], upload time [22], culture [23], and other signs of depression that can be found in social media posts.

This study aims to identify depression in X (Twitter) messages using sentiment analysis. Sentiment analysis of social media messages can help detect depression [24]. This study can be used to identify the idioms that best describe depression and to determine how posting frequency and time are related to depression. It is anticipated that the findings of this study will enhance the delivery of mental health care and assistance.

This paper is formatted as follows. A brief explanation of the research methodology will be provided in section 2. Findings, results, discussion, comparison with other research, and research limitations are outlined in section 3. Finally, conclusions are included at the end.

2. METHOD

To establish a viable method to monitor depression in social media, a methodical strategy for gathering, evaluating, and interpreting online discussions is necessary. The actions to be taken are as follows:

- Crawling and scraping information from X (Twitter) using keywords depression, mental health, and mental disorders.
- Data preprocessing: data must be cleaned, folded into cases, tokenized, filtered, and stemmed before being analysed. Cleaning the data to remove any unnecessary characters, symbols, and non-text objects. Case folding changes all text to lowercase to facilitate uniform comparisons. To tokenize a text means to divide it into tokens or individual words. Common terms (stop words) such as 'and', 'the', and 'is' are deleted during filtering since they typically do not provide helpful information to the study. Normalization converts incomplete words, mistakes, and typing errors into words in the big Indonesian dictionary (KBBI) or by Indonesian enhanced spelling (EYD). Stemming returns words to their original form ('running' becomes 'run'), whereas lemmatization returns words to their base form ('better' becomes 'good', for example). Reducing word variants to their most basic form can be helpful.
- Sentiment polarization: sentiment polarization is the practice of classifying emotions conveyed in text data into distinct polarities (usually positive, negative, and neutral), using lexicon-based machine learning with valence aware dictionary and sentiment reasoner (VADER), Indonesian sentiment (InSet), and validated by a psychologist. Inset performs well as an Indonesian emotion lexicon in predicting brief written thoughts' negative and positive polarity [25]. VADER is a rule- and lexicon-based sentiment analysis tool that aligns precisely with the sentiment expressed on social media [26], [27].
- Visualization and analysis: i) conduct sentiment analysis to determine the attitude of the content (positive, negative, or neutral). It provides insights into public beliefs about depression thoughts and ii) word clouds graphically represent the terms that appear the most frequently in the data. It concisely reviews the most discussed topics, including any positive or negative background.
- Next, feature extraction is carried out using term frequency-inverse document frequency (TF-IDF) to examine the relationship between a word and a document and to gauge the efficacy of the model or algorithm using 10-fold cross-validation. 10-fold cross-validation is a good modelling technique because its accuracy findings are less biased than other techniques [28].
- The naive Bayes technique, random forest, and support vector machine (SVM) are used in the classification model. Naive Bayes is a short algorithm based on Bayes' theorem for conditional probability. The naive Bayes algorithm assumes that all data is independent. The method is assumed to be capable of detecting the dependence on the training features [24]. A SVM is essential for an integrated and supervised classification strategy since the training process requires specific learning targets [29]. Random forests are a method of averaging numerous decision trees to reduce variation [30].

– The algorithm's accuracy, precision, recall, and f1-score will be presented. The percentage of correct predictions based on the whole data is called accuracy. Precision is the proportion of correct positive calculations to total positive calculations. Recall is the fraction of correct positive computations versus all correct positive data. Finally, the f1-score contrasts precision and recall weighted averages.

3. RESULTS AND DISCUSSION
3.1. Findings

X (Twitter) data was collected for three months using depression, mental health, and mental disorders; 5,000 data were retrieved. Results of data cleaning which include case folding (converting data to lowercase and cleaning text), tokenizing (cutting strings based on the constituent words), normalization (correction of incomplete words and typing errors adjusted for EYD–enhanced spelling), removing stop words, stemming (retrieving basic words) and eliminating duplication. Finally, we collected a total of 1,502 tweet data.

Independent manual labeling or unsupervised grouping might lead to oversimplification and bias in results [13] and, therefore ineffective in the therapeutic setting [31]. The labeling procedure used the VADER and Inset lexicons to avoid bias and was confirmed by a psychologist. Table 1 shows the comparison results of this sentiment analysis. Based on psychological validation, the accuracy of sentiment analysis labeling with VADER (95.1%) outperforms InSet (76.9%).

Table 1. VADER, InSet, and psychologist labeling results

Sentiment	VADER (%)	InSet (%)	Psychologist (%)
Positive	31.2	10.0	31.1
Negative	67.5	85.8	62.7
Neutral	1.3	4.2	6.2

The two leading terms on the positive word cloud are mental health and depression, which demonstrate mental health following depression. Meanwhile, the leading term in the negative word cloud is depression, followed by less dominant supporting words such as mental and stress. This demonstrates that posts indicating depression can only use the word depression or the phrase depression connected with mental disease or stress. Figure 1 depicts both positive and negative word clouds. Figure 2 shows words commonly used to represent negative and positive emotions.

Wordcloud positive sentiment Wordcloud negative sentiment

Figure 1. Positive and negative word cloud

After cleaning the data, naive Bayes, random forest, and SVM are used to model it. The findings revealed that random forest has the highest accuracy (83.33%), followed by naive Bayes (80.5%) and SVM (80.4%). Table 2 contains the complete computation results. Meanwhile, the results of evaluating the model with the confusion matrix are explained in greater detail in Figures 3(a) to 3(c), where the model is tested for how many positive predictions it makes in conditions that should be negative (false negative) and vice versa (false positive). Confusion matrices assess a method's effectiveness following categorization [32].

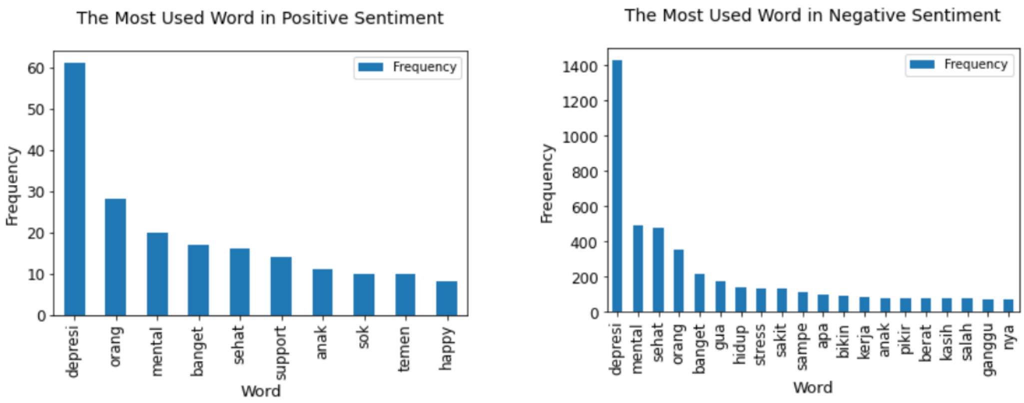
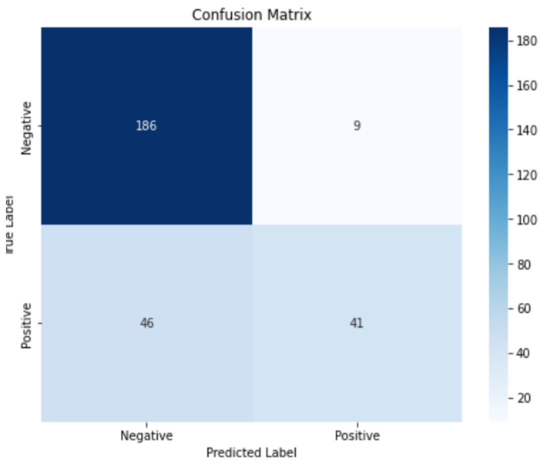


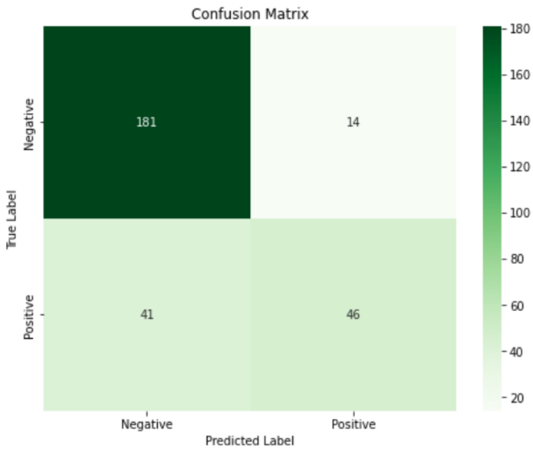
Figure 2. The most used word in positive or negative sentiment

Table 2. Classification test result

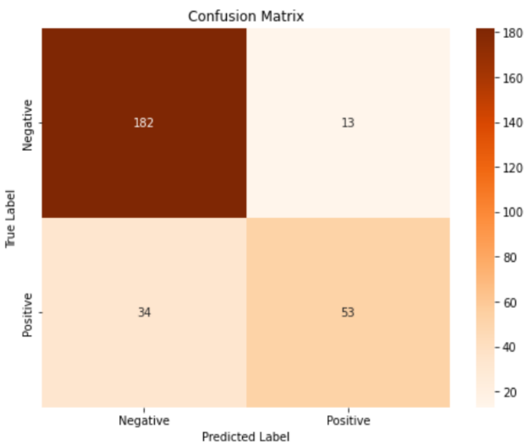
Metric	Naive Bayes (%)	SVM (%)	Random forest (%)
Accuracy	80.5	80.4	83.33
Precision	80.74	80.03	83.04
Recall	80.5	80.49	83.33
F1-score	78.71	79.33	82.62



(a)



(b)



(c)

Figure 3. Confusion matrix of (a) naïve Bayes, (b) SVM, and (c) random forest

3.2. Discussions

Social media has evolved into a media for people to share and express themselves. The writing style and word choice in social media posts can reflect whether or not someone is depressed [4]. There are numerous ways to identify depression on social media [33].

Figures 1 and 2 show that a person's depressive inclinations frequently surface in their posts via phrases with depressive connotations, such as depression, stress, or illness. Individuals suffering from depression frequently use language that conveys negative feelings (such as fear or crying) and juxtaposes the phrases life, death, and religion. Posts indicating depression use phrases connected with depression (such as stress or anxiety), words that represent suffering (e.g., pain, sadness, or medication), or words that terminate action (e.g., before or again).

Examining user messages progressively over a predetermined amount of time is one method of identifying depression on social media [33]. Several findings can be drawn from the data gathered, including the following ones:

- Repeated posts sent at odd times, like in the middle of the night, may indicate potential depression. This aligns with the research conducted in [22] showing that excessive usage of social media is linked to depression.
- Emotional terms like insomnia, hopelessness, and generally depressing or extremely negative messages. This is in line with research on anxiety, sadness, and sleeplessness by Bard [34].
- Posts mentioning physical health issues could be linked to depression. The study on the connection between physical ailments and depression is supported by these findings [35].
- Suicidal ideas or intents expressed in posts should be regarded seriously as a sign of depression and should be dealt with right away [26].
- Depression may be indicated by lonesome posts [22].
- Depression is more common on certain days of the week when there are more posts than on other days [36].

It is not appropriate to utilize the symptoms above as a single diagnosis, even though they can allude to various emotional issues. To ensure a more precise assessment, get assistance from a mental health specialist. Sentiment analysis, which identifies and classifies the emotions represented in the text as positive, negative, or neutral sentiment, is another method for spotting depression in social media posts [37]. A higher frequency of negative phrases and sentiment expressions associated with melancholy, or hopelessness may be a symptom of possible depression in the context of depression identification.

The sentiment analysis results suggest that posts indicating the probability of depression have a more dominating negative sentiment than posts unrelated to depression. Critical phrases like despair, stress, and disease also occur more frequently in these posts. Depressed people frequently focus on themself [38]. It is evident in their social media posts, with many utilizing first-person pronouns like myself. Depression can produce cognitive distortions that impact a person's emotions [39]. It is apparent in their social media posts that they use absolute phrases such as really, always, and never.

The usage of a lexicon is quite beneficial in sentiment analysis. Lexicon VADER outperforms InSet (76.9%) in accuracy (95.1%). However, these findings must be replicated in additional scenarios to provide the best Indonesian language sentiment analysis suggestions.

The results of modelling with random forest, naive Bayes, and SVM revealed that random forest had the highest accuracy (83.3%), followed by naive Bayes (80.5%) and SVM (80.4%). This high accuracy demonstrates that this model can detect depression in social media posts. However, it is believed that more research utilizing the association rule mining technique is required to find words that frequently occur associated with depression so that early identification of depression can be carried out.

3.3. Comparison with another research

This research's findings must be compared with those of other studies to be useful. To decide what has to be improved going forward, the benefits and drawbacks of each must be compared. For an extended explanation, go to Table 3.

3.4. Research limitations

It may perform better to combine textual and visual elements than to use just one mode. Adding images, and emoticons to posts could provide a more realistic view of user activity. Social media data does not provide information about its users' real-world behaviour, culture, and social standing. Therefore, it is essential to combine social media conduct with real-world behaviour to create the ideal environment for researching, learning, and gaining insightful knowledge about the illness of depression.

Culture influences depression expression and also determines attitudes toward mental health, mental disease, and mental illness treatment. Social media may be useful for studying and researching people from diverse cultural backgrounds. Future research could focus on how individuals with depression from various cultural backgrounds behave online and whether there are any cross-cultural variations in textual, visual, user behaviour, and other aspects.

Table 3. Comparison with others' research

Criteria	Proposed research (%)	Depression detection with sentiment analysis [24] (%)	Depression detection using natural language processing and sentiment analysis [21] (%)
Labeling technique	Lexicon based (VADER and Inset) and validated by a psychologist	Bag of word	Bag of word
Classification model	Naïve Bayes, SVM, random forest	Naïve Bayes, SVM	Recurrent neural network (RNN), Convolutional neural network (CNN)
Result	Naïve Bayes (80.5), SVM (80.4), random forest (83.3)	SVM (72), naïve Bayes (72.9)	RNN (80.99), CNN (77.40)
Depression identification in social media post	– Sentiment analysis (positive, neutral, and negative) – Validation by a psychologist in recognizing possible depression posts based on social media usage and mental health issues	Sentiment analysis (positive, neutral, and negative)	Sentiment analysis (depress and non-depress)

4. CONCLUSION

Nowadays, practically everyone utilizes social media to express their thoughts, opinions, and emotions. Understanding their writing style and diction in social media posts can indicate whether they have depressed tendencies. A lexicon can be used for automatic labelling. This may speed up classification. Psychologists' validation results demonstrate that the VADER lexicon's accuracy (95.1%) is significantly higher than that of the InSet lexicon (76.9%). Those who suffer from depression use first-person singular pronouns more frequently on social media than those who do not. People who are depressed often employ diction associated with bad sentiments and emotions, frequently discuss issues such as life, death, and religion, and use absolute language in their social media posts. The high accuracy of all models (>80%) demonstrates that the models can be used to detect depression in social media posts. To help develop depression-related preventive and intervention strategies, more research is necessary to identify words that frequently co-occur with depression using an association rule mining approach and to gauge the severity of depression from social media posts.

FUNDING INFORMATION

We thank Esa Unggul University for funding and supporting our work, especially the Master of Computer Science Program.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Munawar	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓
Yulhendri		✓				✓				✓	✓			✓
Mohammad Fachriza	✓		✓	✓		✓	✓	✓	✓	✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] S. Bucci, M. Schwannauer, and N. Berry, "The digital revolution and its impact on mental health care," *Psychology and Psychotherapy: Theory, Research and Practice*, vol. 92, no. 2, pp. 277–297, 2019, doi: 10.1111/papt.12222.

[2] S. Hemtanon, S. Aekwarangkoon, and N. Kittiphattanabawon, "Proactive depression detection from Facebook text and behavior data," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 5, pp. 5027–5035, 2022, doi: 10.11591/ijece.v12i5.pp5027-5035.

[3] M.-V. K. and M. M. Friese C, Yang J, "Longitudinal and incremental relation of cybervictimization to negative self-cognitions and depressive symptoms in young adolescents," *Physiology and Behavior*, vol. 46, no. 2, pp. 248–256, 2019, doi: 10.1007/s10802-015-0123-7.

[4] N. Z. Zulkarnain, H. Basiron, and N. Abdullah, "Writing style and word usage in detecting depression in social media: a review," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 1, pp. 124–135, 2020.

[5] O. V. Prikhodko *et al.*, "Ways of expressing emotions in social networks: essential features, problems and features of manifestation in internet communication," *Online Journal of Communication and Media Technologies*, vol. 10, no. 2, 2020, doi: 10.29333/ojcmt/7931.

[6] S. Kanchan and A. Gaidhane, "Social media role and its impact on public health: a narrative review," *Cureus*, vol. 15, 2022, pp. 1–10, 2023, doi: 10.7759/cureus.33737.

[7] I. Pawełoszek, "Customer segmentation based on activity monitoring applications for the recommendation system," *Procedia Computer Science*, vol. 192, pp. 4751–4761, 2021, doi: 10.1016/j.procs.2021.09.253.

[8] J. Serrano-Malebran, C. Vidal-Silva, and I. Veas-González, "Social media marketing as a segmentation tool," *Sustainability*, vol. 15, no. 2, pp. 1–16, 2023, doi: 10.3390/su15021151.

[9] R. T. Rust, W. Rand, M. H. Huang, A. T. Stephen, G. Brooks, and T. Chabuk, "Real-time brand reputation tracking using social media," *Journal of Marketing*, vol. 85, no. 4, pp. 21–43, 2021, doi: 10.1177/0022242921995173.

[10] X. Chen, M. D. Sykora, T. W. Jackson, and S. Elayan, "What about mood swings: identifying depression on Twitter with temporal measures of emotions," *WWW '18 Companion: The 2018 Web Conference Companion*, vol. 2, pp. 1653–1660, 2018, doi: 10.1145/3184558.3191624.

[11] I. Herrera-Peco, I. Fernández-Quijano, and C. Ruiz-Núñez, "The role of social media as a resource for mental health care," *European Journal of Investigation in Health, Psychology and Education*, vol. 13, no. 6, pp. 1026–1028, 2023, doi: 10.3390/ejihpe13060078.

[12] A. Aldkheel and L. Zhou, "Depression detection on social media: a classification framework and research challenges and opportunities," *Journal of Healthcare Informatics Research*, vol. 8, no. 1, pp. 88–120, 2024, doi: 10.1007/s41666-023-00152-3.

[13] M. Kabir *et al.*, "DEPTWEET: a typology for social media texts to detect depression severities," *Computers in Human Behavior*, vol. 139, pp. 1–17, 2023, doi: 10.1016/j.chb.2022.107503.

[14] J. E. Rondón Bernard, "Depression: a review of its definition," *MOJ Addiction Medicine & Therapy*, vol. 5, no. 1, pp. 5–7, 2018, doi: 10.15406/mojamt.2018.05.00082.

[15] D. Moreno-Agostino, Y. T. Wu, C. Daskalopoulou, M. T. Hasan, M. Huisman, and M. Prina, "Global trends in the prevalence and incidence of depression:a systematic review and meta-analysis," *Journal of Affective Disorders*, vol. 281, 2020, pp. 235–243, 2021, doi: 10.1016/j.jad.2020.12.035.

[16] Y. S. Handajani, E. Schröder-Butterfill, E. Hogervorst, Y. Turana, and A. Hengky, "Depression among older adults in Indonesia: prevalence, role of chronic conditions and other associated factors," *Clinical Practice & Epidemiology in Mental Health*, vol. 18, no. 1, pp. 1–10, 2022, doi: 10.2174/17450179-v18-e2207010.

[17] M. A. A. Mamun and M. D. Griffiths, "The association between Facebook addiction and depression: a pilot survey study among Bangladeshi students," *Psychiatry Research*, vol. 271, pp. 628–633, 2019, doi: 10.1016/j.psychres.2018.12.039.

[18] N. Wang *et al.*, "Learning models for suicide prediction from social media posts," *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology*, pp. 87–92, 2021, doi: 10.18653/v1/2021.clpsych-1.9.

[19] D. Xiao, R. Xia, and J. Yu, "Emotion cause extraction on social media without human annotation," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 1455–1468, 2023, doi: 10.18653/v1/2023.findings-acl.94.

[20] E. M. Seabrook, M. L. Kern, B. D. Fulcher, and N. S. Rickard, "Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates," *Journal of Medical Internet Research*, vol. 20, no. 5, 2018, doi: 10.2196/jmir.9267.

[21] Z. Lim Yam, Z. Hayati Abdullah, and N. Filzah Mohd Radzuan, "Depression detection based on Twitter using NLP and sentiment analysis," *Applied Mathematics and Computational Intelligence*, vol. 11, no. 1, 2022.

[22] B. Kuzmanović and M. Zogović, "The influence of social media on insurgency (in Serbian: *Upotreba društvenih medija u pobunjeničke i protivpobunjeničke svrhe*)," *Bezbednost, Beograd*, vol. 62, no. 1, pp. 138–156, 2020, doi: 10.5937/bezbednost2001138k.

[23] K. Loveys, J. Torrez, A. Fine, G. Moriarty, and G. Coppersmith, "Cross-cultural differences in language markers of depression online," *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pp. 78–87, 2018, doi: 10.18653/v1/w18-0608.

[24] Hemanthkumar M and Latha A, "Depression detection with sentiment analysis of tweets," *International Research Journal of Engineering and Technology*, vol. 6, no. 5, pp. 1197–1201, 2019.

[25] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: evaluation of a word list for Indonesian sentiment analysis in microblogs," *2017 International Conference on Asian Language Processing (IALP),* Singapore, 2017, pp. 391-394, doi: 10.1109/IALP.2017.8300625.

[26] V. Leiva and A. Freire, "Towards suicide prevention: early detection of depression on social media," *Internet Science*, pp. 428–436, 2017, doi: 10.1007/978-3-319-70284-1_34.

[27] W. Gata and A. Bayhaqy, "Analysis sentiment about islamophobia when Christchurch attack on social media," *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 18, no. 4, pp. 1819–1827, 2020, doi: 10.12928/TELKOMNIKA.V18I4.14179.

[28] D. Berrar, "Bayes' Theorem and naive Bayes classifier," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, pp. 403–412, 2018, doi: 10.1016/B978-0-12-809633-8.20473-1.

[29] M. Ahmad, S. Aftab, M. S. Bashir, and N. Hameed, “Sentiment analysis using SVM: a systematic literature review,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, pp. 182–188, 2018, doi: 10.14569/IJACSA.2018.090226.

[30] S. Saifullah, Y. Fauziyah, and A. S. Aribowo, “Comparison of machine learning for sentiment analysis in detecting anxiety based on social media data,” *Jurnal Informatika*, vol. 15, no. 1, p. 45, 2021, doi: 10.26555/jifo.v15i1.a20111.

[31] S. K. Ernala *et al.*, “Methodological gaps in predicting mental health states from social media: triangulating diagnostic signals,” *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2019, doi: 10.1145/3290605.3300364.

[32] I. Düntsch and G. Gediga, “Confusion matrices and rough set data analysis,” *Journal of Physics: Conference Series*, vol. 1229, no. 1, 2019, doi: 10.1088/1742-6596/1229/1/012055.

[33] M. Conway and D. O. Connor, “Social media, big data, and mental health: current advances and ethical implications,” *Current Opinion in Psychology*, vol. 9, pp. 77–82, 2016, doi: 10.1016/j.copsyc.2016.01.004.Social.

[34] H. A. Bard, C. O’Driscoll, C. B. Miller, A. L. Henry, J. Cape, and C. A. Espie, “Insomnia, depression, and anxiety symptoms interact and individually impact functioning: a network and relative importance analysis in the context of insomnia,” *Sleep Medicine*, vol. 101, pp. 505–514, 2023, doi: 10.1016/j.sleep.2022.12.005.

[35] P. Frank *et al.*, “Association between depression and physical conditions requiring hospitalization,” *JAMA Psychiatry*, vol. 80, no. 7, pp. 690–699, 2023, doi: 10.1001/jamapsychiatry.2023.0777.

[36] M. De Choudhury, “Role of social media in tackling challenges in mental health,” *SAM '13: Proceedings of the 2nd International Workshop on Socially-Aware Multimedia*, pp. 49–52, 2013, doi: 10.1145/2509916.2509921.

[37] H. Almeida, A. Briand, and M. J. Meurs, “Detecting early risk of depression from social media user-generated content,” *CEUR Workshop Proceedings*, vol. 1866, 2017.

[38] D. Mowery *et al.*, “Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study,” *Journal of Medical Internet Research*, vol. 19, no. 2, pp. 1–17, 2017, doi: 10.2196/jmir.6895.

[39] M. Al-Mosaiwi and T. Johnstone, “In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation,” *Clinical Psychological Science*, vol. 6, no. 4, pp. 529–542, 2018, doi: 10.1177/2167702617747074.

BIOGRAPHIES OF AUTHORS

Munawar holds a Ph.D. in computer science from Universiti Teknologi Malaysia. He is an associate professor at the Department of Computer Science at Esa Unggul University. His research interests include database management systems, data warehouses, data science, legal text mining, blockchain, halal, and software engineering. He can be contacted at email: munawar@esaunggul.ac.id.

Yulhendri holds a master in informatics from Bandung Institute of Technology (ITB). He is an assistant professor at the Department of Computer Science at Esa Unggul University. His research interests include information system, machine learning and deep learning, data science, generative AI, AI for education, and information technology for education. He can be contacted at email: yulhendri@esaunggul.ac.id.