TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 23, No. 5, October 2025, pp. 1314~1322
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v23i5.25711

Lexicon-based comparison for suicide sentiment analysis on Twitter (X)

Munawar[1], Dwi Sartika[2], Fathinatul Husnah[2]
[1]Department of Computer Science, Faculty of Computer Science, Universitas Esa Unggul, Jakarta, Indonesia
[2]Department of Informatics, Faculty of Computer Science, Universitas Esa Unggul, Jakarta, Indonesia

Article Info

Article history:

Received Sep 10, 2023
Revised Mar 6, 2025
Accepted Aug 1, 2025

Keywords:

Indonesia sentiment
K-nearest neighbor
Naïve Bayes
Sentiment analysis
Suicide
Support vector machine
Valence aware dictionary and sentiment reasoner

ABSTRACT

Suicidal individuals frequently share their desires on social media. As a result, it was determined that a learning machine for early detection of suicide issues on social media was required. This study aims to examine Twitter (X) users' suicide-related sentiment expressions. The results of searching X for the keywords 'suicide', 'wish to die', and 'want to commit suicide' for 4 months yielded 5,535 tweets. Following the cleaning process, 2,425 tweets were collected. The findings of labeling with the lexicon-based valence aware dictionary and sentiment reasoner (VADER) and Indonesia sentiment (INSET) lexicon, which psychologists confirmed, revealed that VADER was more accurate (92.1%) than INSET (81.6%). Sentiment research reveals negative (86.4%), positive (11.1%), and neutral (2.5%) sentiment. Support vector machine (SVM), K-nearest neighbor (KNN), and Naïve Bayes modeling results show accuracy above 86%, with SVM having the best accuracy (87.65%). Because of its great accuracy, this model can be used to identify and analyze suspicious behavior relating to suicide on X. Further research is still required, despite the excellent identification of early indicators of suicide ideation from social media posts.

Corresponding Author:

Munawar
Department of Computer Science, Faculty of Computer Science, Universitas Esa Unggul
St. Arjuna Utara No 9, Duri Kepa, Kebun Jeruk, Jakarta 11510, Indonesia
Email: munawar@esaunggul.ac.id

1. INTRODUCTION

Suicide is a critical social issue, as approximately 727,000 people commit suicide each year worldwide [1], and it is the second leading cause of mortality in persons aged 10 to 34 years [2]. Suicide rates in Indonesia tend to rise year after year [3] and puts Indonesia in 15th place among nations with the highest suicide rates worldwide [4]. Suicide begins with the intention to commit suicide, followed by complaints of exhaustion, questions about the purpose of life, and entrusting people or valuable possessions to others [4] Most people with suicidal thoughts steer clear of medical care to avoid social stigma [5]. Ironically, many perpetrators update their status on social media before committing suicide as a kind of self-disclosure [6]. Teenagers and young adults can talk about suicidal thoughts on social media because it offers anonymity [7]. One of the most important steps in seeking help is to disclose suicidal thoughts [8]. Similar reactions can be elicited by other users when sharing suicidal thoughts on social media [9]. Even though copycat suicide is also influenced by the extensive media coverage of suicide cases [7].

Language usage and suicidal thoughts have been linked in several studies [5]. In the weeks preceding the suicide, a rise in melancholy tweets was discovered [6]. One important factor in identifying suicidal thoughts in X users has been demonstrated to be language framing [10]. The fact that the phrase 'want to commit suicide' is linked to a more severe suicidal intent than 'desire to die' reflects this. The amount of fear in posts containing suicidal thoughts was determined by O'Dea *et al*. [11]. Research shows that sentiment analysis for mental health in social media posts using natural language processing (NLP) is becoming increasingly common [12], [13]. Early indicators of suicidal ideation can be detected through sentiment analysis [5], which can also record user emotions [14]. However, neither psychiatrists nor psychologists endorse the studies as validators of the findings. Research on idioms frequently used in suicide-related social media posts is still deemed necessary.

In sentiment analysis, lexicon-based and machine learning-based approaches are two of the most popular methods. Each strategy has benefits and drawbacks of its own. The lexicon-based approach matches words in a text with a list of sentiment words to determine whether a text is positive, negative, or neutral. However, the machine learning method relies on a dataset labeled with a particular sentiment to train a model. Whereas the machine learning approach works best with large data, the lexicon approach works best with smaller data [15].

This study uses a hybrid approach of lexicon-based sentiment polarization and machine learning. Sentiment polarization is the practice of classifying emotions conveyed in text data into distinct polarities (usually positive, negative, and neutral), using lexicon-based with valence aware dictionary and sentiment reasoner (VADER), Indonesian sentiment (INSET), and validated by psychologists. INSET performs well as an Indonesian emotion lexicon in predicting brief written thoughts' negative and positive polarity [16]. VADER is a rule- and lexicon-based sentiment analysis tool precisely aligned with the sentiment expressed on social media [17]. The classification model based on machine learning uses the Naïve Bayes technique, K-nearest neighbor (KNN), and support vector machine (SVM). Naïve Bayes is a short algorithm based on Bayes' theorem for conditional probability. The Naïve Bayes algorithm assumes that all data is independent. The method is assumed to be capable of detecting the dependence on the training features [17]. A SVM is essential for an integrated and supervised classification strategy since the training process requires specific learning targets [18]. KNN is a method for classifying objects based on their proximity to the item. Despite its simplicity, KNN is highly good for categorization [19].

Suicide and online behavior research has been undertaken in various countries, including Japan [10], Taiwan [20], America [21], and Australia [22]. Similar studies employing Indonesian idioms are still regarded as extremely rare in Indonesia. Furthermore, because these idioms change over time, more research is required to fully understand them and use them to prevent suicidal thoughts as early as possible. Such expressions can only be obtained from social media platforms like X.

This study examines suicidal attitudes on the social media platform X in Indonesia. Understanding X users' feelings about suicide might provide significant insight into public perceptions, potential risk factors, and intervention opportunities. The purpose of this study is to contribute to broader efforts in mental health awareness and suicide prevention in Indonesia by analyzing the sentiment of suicide-related X messages.

2. METHOD

To establish a viable methodology to monitor suicide ideation, a methodical strategy for gathering, evaluating, and interpreting online discussions is necessary. The steps listed below will be taken:

a. A preliminary investigation was conducted to identify terms typically associated with suicide ideation. The process of finding keywords never ends. The more in-depth the study, the more likely it is that more terms related to suicidal thoughts will be found. Look over the keyword list and eliminate any too broad, too specific, or less relevant. Based on their usefulness, frequency of use, and potential impact on suicidal data, certain keywords are more important. The outcome of the preliminary investigation is a list of suggested crawlable keywords.
b. Crawling and scraping information from X using proposed keywords related to suicidal ideation.
c. Data preprocessing: data must be cleaned, folded into cases, tokenized, filtered, and stemmed before it can be analyzed. 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', 'is', and so on are deleted during filtering since they typically do not provide useful information to the study. Normalization is converting incomplete words, mistakes, and typing errors into words in the big indonesian dictionary (KBBI) or by Indonesian enhanced spelling (EYD). While lemmatization changes words to their base form (for example, 'better' becomes 'good'), stemming changes words to their origin form (for example, 'running' becomes 'run'). It helps to reduce word variations to their most basic form.
d. Sentiment polarization: to ascertain whether a sentiment is positive, negative, or neutral, VADER and INSET are used to polarize it.
e. Visualization and analysis.

f. Conduct sentiment analysis to determine the attitude of the content (positive, negative, or neutral). It provides insights into public beliefs about suicidal thoughts.
g. Word clouds are used to 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.
h. Then, using 10-fold cross-validation, feature extraction is performed using term frequency-inverse document frequency (TF-IDF) to evaluate the association between a word and a document and to assess the effectiveness of the model or algorithm. 10-fold cross-validation is a good modeling technique because its accuracy findings are less biased than other techniques [23].
i. The Naïve Bayes technique, KNN, and SVM are used in the classification model based on machine learning sentiment polarization.
j. The accuracy, precision, recall, and F1-score of the algorithm will be presented. The percentage of correct predictions based on the whole data is called accuracy. Precision is defined as the proportion of correct positive calculations to total positive calculations. The fraction of correct positive computations versus all correct positive data is referred to as recall. Finally, the F1-score contrasts precision and recall weighted averages.

Figure 1 depicts a summary of the stages of this research.

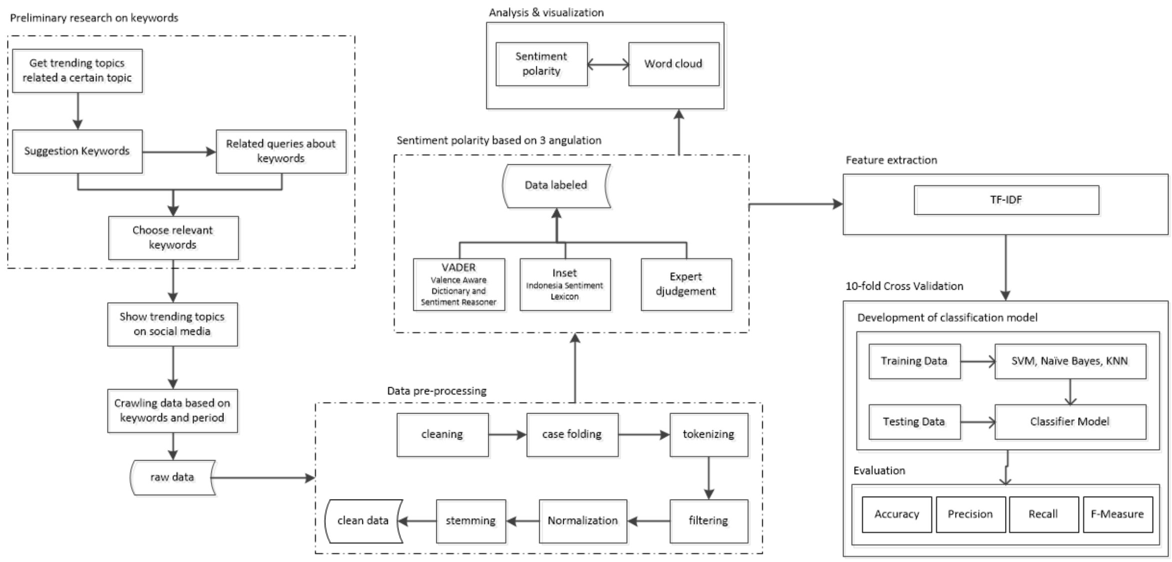


Figure 1. Synopsis of research stages

3. RESULTS AND DISCUSSION
3.1. Data collection

The lack of publicly available data on suicide in Indonesia is a barrier. However, with an increasing number of people releasing their feelings on social media, it can be used to collect data on the suicide issue. X is an excellent primary source for information about mental illnesses, including suicide [10], [18].

Finding keywords related to suicide is not a simple task. Search terms for suicide in Google Trends show ‘commit suicide,’ ‘how to commit suicide,’ and ‘suicide prevention’ [19], [20]. The use of the idioms ‘want to commit suicide’ and ‘want to die’ are posts that correlate with suicidal ideation [24]. However, wanting to commit suicide shows more seriousness of suicidal intentions [10].

Looking at the specific situation in Indonesia, the keywords selected for crawling X data include ‘suicide’, ‘wish to die’, and ‘want to suicide’. Data was collected for four (4) months, with a total of 5,535 tweets containing details of suicide (4634), wish to die (900), and want to suicide (100).

Data cleaning results 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 2.425 tweet data.

3.2. Data analysis

The clean data was then classified and confirmed by a psychologist using a lexicon based on INSET and VADER. The result are in Table 1. Some labeling variations exist between lexicon-based INSET, VADER, and experts concerning the suicide data acquired. VADER offered comparable analysis results of 2,233 data (92.1%) of the total clean data (2,425), while INSET had 1,979 data (81.6%). INSET and VADER mislabeled 464 data sets (19.1%). Table 2 shows some examples of discrepancies in labeling findings.

Table 1. Data labelling

Label	INSET		VADER		Expert	
Positive	178	7.3%	190	7.8%	269	11.1%
Neutral	53	2.2%	6	0.2%	60	2.5%
Negative	2194	90.5%	2229	91.9%	2097	86.4%

Table 2. Example of difference in labeling findings

#	Clean Tweet	INSET	VADER	Experts
1	*bunuh diri sila vaksin* (suicide please vaccinate)	Negative	Negative	Neutral
2	*cerita bunuh diri pas sedang live streaming angkat kisah nyata* (an actual suicide story that was live-streamed)	Neutral	Negative	Negative
3	*bunuh diri masuk neraka nder percaya surga neraka hidup nunggu gilir* (commit suicide. Go to hell and stop believing in paradise. Hell will wait for you)	Negative	Negative	Positive
4	*waduch gabole kack orang mati hidup tetep semangat karen sudh masuk senin* (it's Monday, so keep your enthusiasm high even when you can't kick the dead)	Negative	Negative	Positive
5	*gua pengen mati muda tpi bunuh diri apajalah mati painless aib keluarga su* (although suicide is a painless death and a shame to the family, I still wish to die young)	Negative	Positive	Positive

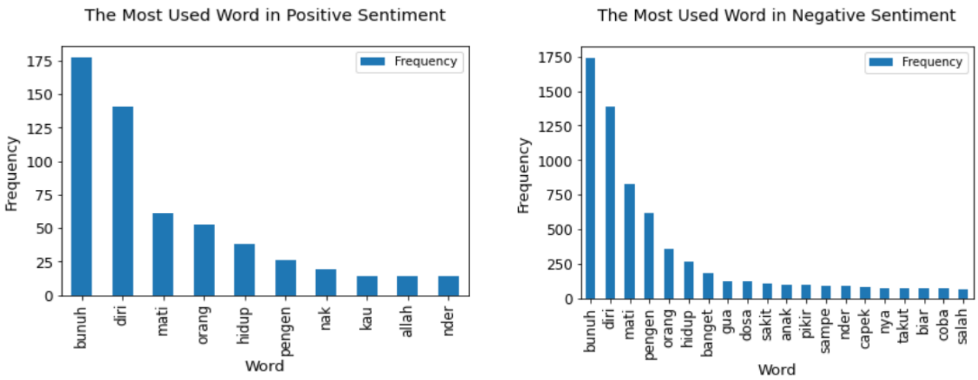
Figures 2 and 3 illustrate the terms that appear most frequently in positive and negative sentiment from Tweets respectively. Figure 2(a) positive word cloud and Figure 2(b) negative word cloud, Figure 3(a) positive and Figure 3(b) negative sentiments. Figures 3 and 4 show that while the words suicide are the same (equally prevalent in both positive and negative word clouds), the diction that goes with them differs. The phrasing accompanying it in the negative word cloud is wishing to die; however, in the positive word cloud, the diction accompanying it is Allah as an incentive to live.

(a) (b)

Figure 2. Word cloud positive and negative sentiment: (a) positive word cloud and (b) negative word cloud

3.3. Classification performance

The cleaned data is then modeled with Naïve Bayes, SVM, and KNN. The results show that SVM (87.65%) has the highest level of accuracy, followed by KNN (86.83%) and Naïve Bayes (86.21%). The high level of accuracy demonstrates that this approach can provide important insight into suicide on X. Figure 4 shows more comprehensive results from these three algorithms: Figure 4(a) SVM, Figure 4(b) KNN, and Figure 4(c) Naïve Bayes. Figure 5 depicts the model evaluation findings, where this matrix compares the classification results achieved by the model with the actual classification results.



(a) (b)

Figure 3. Words that often appear in (a) positive and (b) negative sentiments

(a)

	precision	recall	f1-score	support
negative	0.88	1.00	0.93	419
neutral	0.67	0.22	0.33	9
positive	0.88	0.12	0.21	58
accuracy			0.88	486
macro avg	0.81	0.45	0.49	486
weighted avg	0.87	0.88	0.84	486

(b)

	precision	recall	f1-score	support
negative	0.87	1.00	0.93	419
neutral	0.67	0.22	0.33	9
positive	0.80	0.07	0.13	58
accuracy			0.87	486
macro avg	0.78	0.43	0.46	486
weighted avg	0.86	0.87	0.82	486

(c)

	precision	recall	f1-score	support
negative	0.86	1.00	0.93	419
neutral	0.00	0.00	0.00	9
positive	0.00	0.00	0.00	58
accuracy			0.86	486
macro avg	0.29	0.33	0.31	486
weighted avg	0.74	0.86	0.80	486

Figure 4. Model testing results: (a) SVM, (b) KNN, and (c) Naïve Bayes

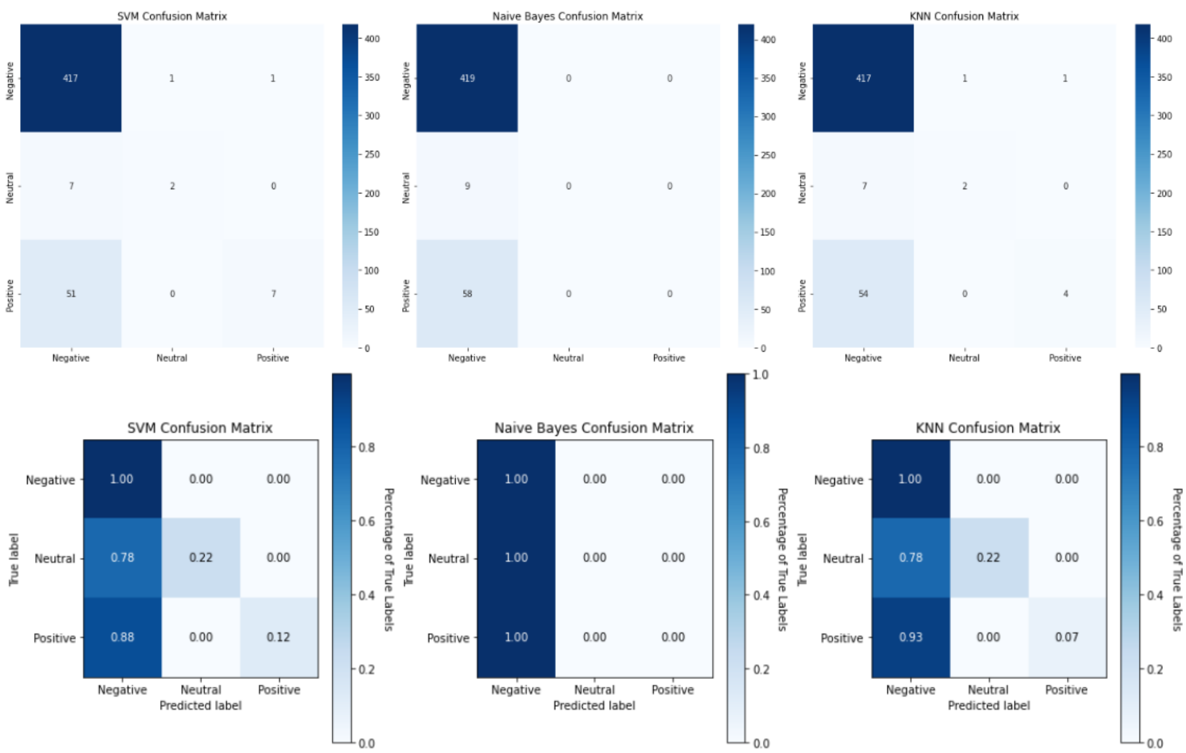


Figure 5. Model evaluation results

3.4. Discussion

Suicide prevention requires early detection and action. Although search engines like Google Trends can assist in mapping suicide trends in a given location [19], [25], they cannot identify who has suicidal intent. The increasingly widespread expression of emotion on social media, particularly X, can aid in the early detection of suicide, particularly among youth. As time passes, idioms about suicide, particularly among youth, evolve. Using the keywords ‘suicide’, ‘wish to die’, and ‘want to commit suicide’ is highly effective in collecting suicide data from X.

The use of a lexicon greatly aids labeling. It is merely that because it is related to suicide, a psychologist must validate the sentences acquired. VADER is more accurate in the suicide context than in the INSET. Further research on other issues is required to obtain recommendations for employing a more comprehensive Indonesian lexicon that may be used widely.

According to the sentiment analysis results, the gathered tweet content included 86.4% negative, 11.1% positive, and 2.5% neutral content. It demonstrates the seriousness of the suicidal content. Positive content, which accounts for only 11.1%, indicates that other X users encourage content uploaders to remain passionate about life and remember Allah. Given the harmful impact of openness in the actual world, it is argued that decision-makers should make a more serious effort to build channels to support persons who have suicidal thoughts.

The high frequency of specific terms in tweets users send indicates that these words can spark suicidal thoughts. Suicidal intent can be shown by using suicide and desiring to die in a negative context. Using suicide with life and Allah, on the other hand, can convey motivation to stay alive no matter how challenging the problems are. On the other hand, this indicator could imply suicide ideation from public view, allowing for early intervention.

There are several methods for spotting suicidal warning signs on social media. From the collected data, some conclusions can be made, including the following:

- A high degree of suicidal intent is indicated by posts that utilize the terms ‘suicide’ and ‘want to die’. This can serve as an early warning sign for suicide attempt prevention [26].
- Another early sign of suicidal thoughts is the usage of the word ‘capek’ (tired) which conveys sentiments of being a burden (for instance, as a result of illness) [8].
- Persistent stress or depression might exacerbate suicidal thoughts [27].
- Suicidal thoughts can be a precursor to depression and feelings of guilt and sin. In the presence of hopelessness, these suicidal thoughts will become suicidal intentions [28].

These results are merely preliminary indications. It has to be investigated further to serve as a precursor of suicide intent in social media posts. This is significant since about 6% of people who have suicidal thoughts end their lives [29].

The modeling results with SVM, KNN, and Naïve Bayes all show accuracy above 86%, with SVM having the most remarkable accuracy (87.65%). This high accuracy demonstrates that this model can identify or assess suspicious behavior related to suicide issues to carry out prevention and early intervention. However, more efforts are required to address the issue of suicide using an association rule mining technique to uncover words that frequently appear together about suicide. It will considerably aid in creating more effective and relevant preventative and intervention techniques.

3.5. Comparison with other research

For the results of this study to be helpful, they must be compared to those of other investigations. The advantages and disadvantages of each must be weighed to determine what has to be improved moving forward. Refer to Table 3 for a more detailed explanation.

Table 3. Comparison with others’ research

Activities	Proposed research	Machine classification for suicide ideation [12]	Detecting and analyzing suicidal ideation [5]
Labeling technique	Lexicon based (VADER and INSET) and validated by a psychologist	Bag of word	Word embedding Word2Vec
Classification model	SVM, KNN, Naïve Bayes	SVM, random forest, decision tree, Naïve Bayes, Prism	XGBoost CNN-BiLSTM
Result	**SVM (87.65%)**, KNN (86.83%), Naïve Bayes (86.21%)	SVM (79.25%), random forest (83.33%), decision tree (76.25%), Naïve Bayes (79.25%), **Prism (91.6%)**	**CNN-BiLSTM (95%)** XGBoost (91%)
Depression identification in social media post	- Sentiment analysis (positive, neutral, negative) - A psychologist’s confirmation of suicidal ideation post	Suicide, non-Suicide, flippant	Suicidal and non-suicidal

4. CONCLUSION

The keywords ‘suicide’, ‘wish to die’, and ‘want to commit suicide’ reflect X users’ passionate outbursts concerning suicide. This keyword is critical to the data crawling process. It verifies prior studies on the right keywords for suicide, specifically ‘want to commit suicide’ and ‘want to die’, though with varying idioms depending on the language employed.

A lexicon-based tagging procedure can aid in determining sentiment analysis results. According to the psychologist’s validation data, VADER delivered an accuracy of 92.1% in cases of suicide, compared to the INSET (81.6%). These findings must be replicated in other circumstances before they can be generalized.

The phrases ‘suicide’ and ‘want to die’ have negative sentiment connotations. In contrast, suicide combined with life and Allah conveys more positive thoughts to remind those considering suicide. Unfortunately, the high proportion of negative feelings (86.4%) against positive sentiments (11.1%) implies that a few X users are warning and inspiring content uploaders with suicide intentions.

The high accuracy of all models (>86%) demonstrates that the models can be used to detect suicidal behavior. However, more study is needed to uncover phrases that frequently appear together to suicide utilizing an association rule mining approach to aid in establishing preventative and intervention measures connected to suicidal intentions from content uploaders.

FUNDING INFORMATION

Authors state no funding involved

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	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	
Dwi Sartika		✓			✓			✓		✓	✓	✓		✓
Fathinatul Husnah	✓		✓			✓	✓		✓	✓	✓		✓	

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

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] “Suicide,” *World Health Organization*, 2025. https://www.who.int/news-room/fact-sheets/detail/suicide (accessed Aug. 20, 2025).
[2] A. Z. Ivey-Stephenson *et al.*, “Suicidal Ideation and Behaviors Among High School Students - Youth Risk Behavior Survey, United States, 2019,” *MMWR supplements*, vol. 69, no. 1, pp. 47–55, Aug. 2020, doi: 10.15585/mmwr.su6901a6.
[3] S. Onie *et al.*, “Indonesia’s first suicide statistics profile: an analysis of suicide and attempt rates, underreporting, geographic distribution, gender, method, and rurality,” *The Lancet Regional Health - Southeast Asia*, vol. 22, 2024, doi: 10.1016/j.lansea.2024.100368.
[4] D. Setiyawati, N. Puspakesuma, W. N. Jatmika, and E. Colucci, “Indonesian Stakeholders’ Perspectives on Warning Signs and Beliefs about Suicide,” *Behavioral Sciences*, vol. 14, no. 4, p. 295, Apr. 2024, doi: 10.3390/bs14040295.
[5] T. H. H. Aldhyani, S. N. Alsubari, A. S. Alshebami, H. Alkahtani, and Z. A. T. Ahmed, “Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 19, p. 12635, Oct. 2022, doi: 10.3390/ijerph191912635.
[6] G. Coppersmith, K. Ngo, R. Leary, and A. Wood, “Exploratory analysis of social media prior to a suicide attempt,” in *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical*

Reality, CLPsych 2016 at the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2016, Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, pp. 106–117, doi: 10.18653/v1/w16-0311.
[7] J. Robinson *et al.*, “Social media and suicide prevention: A systematic review,” *Early Intervention in Psychiatry*, vol. 10, no. 2, pp. 103–121, Apr. 2016, doi: 10.1111/eip.12229.
[8] P. Davies, M. Veresova, E. Bailey, S. Rice, and J. Robinson, “Young people’s disclosure of suicidal thoughts and behavior: A scoping review,” *Journal of Affective Disorders Reports*, vol. 16, p. 100764, Apr. 2024, doi: 10.1016/j.jadr.2024.100764.
[9] R. C. Oconnor, S. Rasmussen, and K. Hawton, “Adolescent self-harm: A school-based study in Northern Ireland,” *Journal of Affective Disorders*, vol. 159, pp. 46–52, Apr. 2014, doi: 10.1016/j.jad.2014.02.015.
[10] H. Sueki, “The association of suicide-related Twitter use with suicidal behaviour: A cross-sectional study of young internet users in Japan,” *Journal of Affective Disorders*, vol. 170, pp. 155–160, Jan. 2015, doi: 10.1016/j.jad.2014.08.047.
[11] B. O’Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, “Detecting suicidality on twitter,” *Internet Interventions*, vol. 2, no. 2, pp. 183–188, May 2015, doi: 10.1016/j.invent.2015.03.005.
[12] M. Manisha, A. Kodali, and V. Srilakshmi, “Machine classification for suicide ideation detection on twitter,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 4154–4160, Oct. 2019, doi: 10.35940/ijitee.L3655.1081219.
[13] S. Pachouly, G. Raut, K. Bute, R. Tambe, and S. Bhavsar, “Depression Detection on Social Media Network (Twitter) using Sentiment Analysis,” in *International Research Journal of Engineering and Technology*, 2021, pp. 1834–1839.
[14] A. Giachanou and F. Crestani, “Like it or not: A survey of Twitter sentiment analysis methods,” *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–41, Jun. 2017, doi: 10.1145/2938640.
[15] H. S. Hota, D. K. Sharma, and N. Verma, “Lexicon-based sentiment analysis using Twitter data,” in *Data Science for COVID-19 Volume 1: Computational Perspectives*, Elsevier, 2021, pp. 275–295, doi: 10.1016/B978-0-12-824536-1.00015-0.
[16] Munawar and Yulhendri, “Identification of potential depression in social media posts,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 14, no. 3, pp. 2096–2103, 2025, doi: 10.11591/ijai.v14.i3.pp2096-2103.
[17] Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, “Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification,” in *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, IEEE, Nov. 2017, pp. 294–298. doi: 10.1109/ICITISEE.2017.8285514.
[18] R. Sawhney, P. Manchanda, R. Singh, and S. Aggarwal, “A Computational approach to feature extraction for identification of suicidal ideation in tweets,” in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Student Research Workshop*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2018, pp. 91–98, doi: 10.18653/v1/p18-3013.
[19] P. Solano *et al.*, “A Google-based approach for monitoring suicide risk,” *Psychiatry Research*, vol. 246, pp. 581–586, Dec. 2016, doi: 10.1016/j.psychres.2016.10.030.
[20] A. Yang, “P-1449 - Association of internet search trends with suicide death in Taiwan,” *European Psychiatry*, vol. 27, p. 1, Jan. 2012, doi: 10.1016/s0924-9338(12)75616-x.
[21] P. W. C. Wong *et al.*, “Accessing suicide-related information on the internet: A retrospective observational study of search behavior,” *Journal of Medical Internet Research*, vol. 15, no. 1, p. e3, Jan. 2013, doi: 10.2196/jmir.2181.
[22] K. Mok, A. F. Jorm, and J. Pirkis, “The perceived impact of suicide-related internet use: A survey of young Australians who have gone online for suicide-related reasons,” *Digital Health*, vol. 2, Jan. 2016, doi: 10.1177/2055207616629862.
[23] D. Berrar, “Bayes’ theorem and naive bayes classifier,” in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, Elsevier, 2018, pp. 403–412, doi: 10.1016/B978-0-12-809633-8.20473-1.
[24] A. Pourmand, J. Roberson, A. Caggiula, N. Monsalve, M. Rahimi, and V. Torres-Llenza, “Social Media and Suicide: A Review of Technology-Based Epidemiology and Risk Assessment,” *Telemedicine and e-Health*, vol. 25, no. 10, pp. 880–888, Oct. 2019, doi: 10.1089/tmj.2018.0203.
[25] J. F. Gunn and D. Lester, “Using google searches on the internet to monitor suicidal behavior,” *Journal of Affective Disorders*, vol. 148, no. 2–3, pp. 411–412, Jun. 2013, doi: 10.1016/j.jad.2012.11.004.
[26] M. T. Tull, Z. T. DeMoss, M. D. Anestis, J. M. Lavender, M. J. McDermott, and K. L. Gratz, “Examining associations between suicidal desire, implicit fearlessness about death, and lifetime frequency of suicide attempts,” *Suicide and Life-Threatening Behavior*, vol. 52, no. 6, pp. 1110–1120, Dec. 2022, doi: 10.1111/sltb.12905.
[27] J. A. Kumar, T. E. Trueman, and A. K. Abinesh, “Suicidal risk identification in social media,” *Procedia CIRP*, vol. 189, pp. 368–373, 2021, doi: 10.1016/j.procs.2021.05.106.
[28] M. B. Tonkuş, B. B. Çalışkan, and E. Alagöz, “The relationship between suicide and hopelessness in young adults aged 18-30: A systematic review,” *Journal of Psychiatric Nursing*, vol. 13, no. 3, pp. 253–262, 2022, doi: 10.14744/phd.2022.76993.
[29] K. S. Lim *et al.*, “Global lifetime and 12-month prevalence of suicidal behavior, deliberate self-harm and non-suicidal self-injury in children and adolescents between 1989 and 2018: A meta-analysis,” *International Journal of Environmental Research and Public Health*, vol. 16, no. 22, p. 4581, Nov. 2019, doi: 10.3390/ijerph16224581.

BIOGRAPHIES OF AUTHORS

Munawar holds a Ph.D. in Computer Science from Universiti Teknologi Malaysia. Currently, he is an Associate Professor at the Department of Computer Science - Esa Unggul University. His research interests include data analytics, blockchain technologies, halal, and software engineering. He can be contacted at email: munawar@esaunggul.ac.id.

Dwi Sartika holds a Master of Informatics Engineering from Binus University. She is a lecturer at the Department of Informatics at Esa Unggul University. Her research interests include data warehouses, data science, blockchain, and software engineering. She can be contacted at email: dwi.sartika@esaunggul.ac.id.

Fathinatul Husnah holds a Bachelor's in Information Technology from Esa Unggul University. Currently, she is focused on a career in design. With a background in computer science, she has research interests in data mining and text mining. She can be contacted at email: fathinatulhusnah12@gmail.com.