Journal of
Information Systems Engineering
and Business Intelligence

Vol.8, No.2, October 2022
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Lexicon and Naive Bayes Algorithms to Detect Mental Health Situations from Twitter Data

Sheila Shevira[1)]*, I Made Agus Dwi Suarjaya[2)], Putu Wira Buana[3)]

[1)2)3)] Department of Information Technology, Faculty of Engineering, Udayana University, Indonesia
Jalan Raya Kampus Udayana, Badung
[1)]shevira@student.unud.ac.id, [2)]agussuarjaya@it.unud.ac.id, [3)]wbhuana@it.unud.ac.id

Abstract

Background: Twitter is a popular social media where users express emotions, thoughts, and opinions that cannot be channelled in the real world. They do this by tweeting short, concise, and clear messages. Since users often express themselves, Twitter data can detect mental health trends.
Objective: This study aims to detect suicidal messages through tweets written by users with mental health issues.
Methods: These tweets are analysed and classified using the lexicon-based and Naive Bayes algorithms to determine whether it contains suicidal messages.
Results: The classification results show that the 'normal' classification is predominant at 52.3% of the total 3,034,826 tweets, which indicates an increase from September to December 2021.
Conclusion: Most tweets are categorised as 'normal', therefore the mental health status appears secure. However, this finding needs to be re-examined in the future, especially in DKI Jakarta Province, which has the most cases of mental disorders. This study found that the Naive Bayes algorithm is more accurate (85.5%) than the lexicon-based algorithm. This can be improved in future studies by increasing performance at the pre-processing stage.

Keywords: Lexicon Based, Mental Disorder, Mental Health, Naïve Bayes, Twitter

Article history: Received 1 August 2022, first decision 24 August 2022, accepted 8 September 2022, available online 28 October 2022

I. INTRODUCTION

Mental health is a rising concern in modern life, with the cases of disorders becoming more prevalent globally. Most cases are found in Southeast Asia and America [1]. Data from Basic Health Research show that depression and anxiety cases saw a 6% increase from 2013 to 9.8% in 2018 and continue to increase every year, especially in 2020, due to the COVID-19 pandemic. The World Health Organization (WHO) claim that more than 500 million people worldwide are experiencing mental health disorders, with depression and anxiety being the most prevalent. Globally, 322 million people (4.4% of the population) suffer from depression, and about 264 million people (3.6% of the population) suffer from anxiety. Depression is a major contributor to suicide cases, reaching more than 800,000 suicide cases every year, or about one person every 40 seconds. The number of attempted suicides is estimated to be 20-25 times the number of deaths due to suicide. Mental health disorders can be treated, but people may not undergo proper treatments. The health system may not be sufficient to treat mental health problems, and the cost of treatment is high. Another problem is that individuals with mental health problems do not show their suffering and even deny that they are experiencing mental health problems.

The rapid development of information technology and the Internet has allowed people to express their emotions [2]. Twitter is a microblogging platform often used to write everything related to mind, moods, activities, social life, information, news, and others. Twitter has many data sources and provides API to assist with data retrieval [3]. Twitter data can be analysed using sentiment analysis. Through this, important information related to public opinions can be found. In this case, Big Data can help research various data structures in the high-volume Twitter data. Using big data, mental health situations can be analysed based on the tweets written by users [4].

Research on Twitter sentiment analysis has been done before to classify stress levels [5]. The method used is Improved K-Nearest Neighbour and Chi-square Feature Selection to improve system accuracy and eliminate less relevant features. The test was performed five times, resulting in the best value at a 50% feature ratio and k = 20, with

[*] Corresponding author

doi: http://dx.doi.org/10.20473/jisebi.8.2. 119-129

an average precision value of 70%, an average recall of 67.2%, an average accuracy of 83.3%, and an average f-measure of 66.3%. Almouzini et al. have also conducted machine learning research to analyse the sentiment of depression among Arabs using Twitter data [6]. Tweets from users are classified into two labels: depression and non-depression, and then build a predictive model based on supervised learning algorithms (Random Forest, Naïve Bayes, AdaBoostM1, and Liblinear). The results show that the Liblinear method is the most optimal, with an accuracy rate of 87.5%. Meanwhile, the method with the lowest accuracy rate is AdaBoostM1, with an accuracy of 55.2%.

Sentiment analysis research with Naive Bayes and lexicon-based Features has been carried out previously involving different topics. Arifin Kurniawan et al. conducted an opinion analysis of film reviews using the KASKUS data by classifying reviews into a positive or negative opinion class [7]. The test results obtained values of accuracy, precision, recall, and f-measure of 0.8, 0.8, 0.8 and 0.8, respectively. Research conducted by Aulia and Amelia [8] aims to discover the stigma on mental health issues using the Naive Bayes method. The results show that positive sentiment was predominant.

Saputra et al. also researched Twitter sentiment analysis using lexicon and a tree structure to interpret a relationship between words in sentence formation by adding words to the lexicon sentiment [9]. This research was tested on data using three topics: the 2018 West Java gubernatorial election, the 2019 presidential election, and the COVID-19 pandemic. The three-test data show an unbalanced proportion, i.e., predominantly positive. The tree-based method resulted in an accuracy of 64.97% (an increase of 1.26%) in the 2018 West Java gubernatorial election data, 64.33% (an increase of 11.41%) in the 2019 presidential election data, and 66.24% (an increase of 7.61%) on the COVID-19 pandemic data.

This study analyses mental health trends based on Twitter data which aims to detect mental health disorders in the community as early as possible and assist in the prevention of mental health disorders. The analysis of the social-emotional tweets of the community using lexicon-based and Naive Bayes methods. Previous studies have produced fairly good accuracy using Naive Bayes. Meanwhile, the lexicon-based method, which directly compares opinion words with words in the dictionary, will be used as a comparison.

II. METHODS

Mental health analysis was developed using the lexicon and Naive Bayes methods proposed by Le et al. [10]. The following general description presents the research method used in analysing tweets of mental disorders. The research overview can be found in Fig. 1.

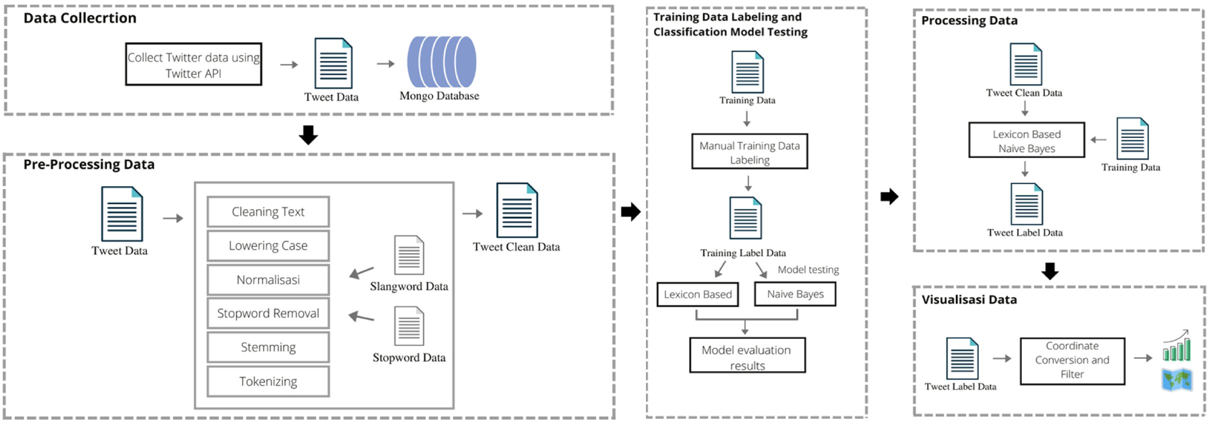

Fig. 1 Research overview

A. Data Collection

Data were retrieved from Twitter API and the Tweepy library from July 2021 to April 2022. The example of Twitter data collection can be seen in Table 1. Twitter data was collected by entering keywords ‘capek’, ‘stres’, ‘depresi’, ‘mau mati’, ‘tidak ada yang peduli’, obtained from interviews with a clinical psychologists.

The data collected amounted to 3.037.226 tweets, and all fields are stored in MongoDB in JSON format. This tweet data is still raw with symbols, numbers, URLs, hashtags, etc., so they need to be pre-processed to make all text uniform. A total of 2,400 data were taken to be used as training data and labelled manually.

TABLE 1
TWITTER DATA COLLECTION RESULTS

#	Full Text Tweet
1	Capek pengen menghilang dari bumi :)
2	Intinya kalau sudah cinta tidak peduli apa yang terjadi, karena tidak ada kriteria khusus untuk mencintai atau dicintai oleh seseorang. Realistis banget lah endingnya. 😌 https://t.co/XZTz4GxM2y
3	capek pilek pagi pagi 😁
4	perempuan, anak pertama. keluarga taruh ekspektasi setinggi-tingginya ke gua. ketika gua gak bisa memenuhi ekspektasi mereka. mereka kecewa. gua kecewa. mereka mulai capek. gua mulai hancur.
5	Udah mulai capek sm idup yg gini aj

The 2.000 training data are labelled 0 to indicate mental disorders and 1 to indicate normalcy. The classification label is divided into 1.200 'normal' and 1.200 'indicated mental disorder'. The example of the training data is given in Table 2.

TABLE 2
TRAINING DATA

Tweet Training	Label
kaya rasa untung depresi selera musik highclass nct dream mau jakarta tolong cerah suhu	1
heesung makasih hadir dunia makasih semangat	1
untung orang tua saudara saling rangkul saling bantu kompak banget bantu pas sakit	1
syukur bangtan sumber bahagia semangat saat rasa depresi	1
hari dimana titik paling rendah hidup dimana rasa capek banget guna beban banyak masalah	0
harus korban capek mau curhat beban hidup pendam sendiri sesak	0
capek dewasa capek beban pikir hidup capek tumpu gak sanggup	0
kecewa capek cemburu sedih rasa sekarang	0

B. *Pre-processing Data*

Pre-processing aims to clean tweet data before the processing and analysis so that the data are eligible and accurate. Tweet data stored in MongoDB is recalled for pre-processing, such as cleaning, lowering case, normalisation, stop word removal, stemming, and tokenising, to make the data more structured and compatible [11].

TABLE 3
PRE-PROCESSING RESULTS

Pre-processing steps		Sample
Initial Text	:	Capek pengen menghilang dari bumi :)
After Text Cleaning	:	Capek pengen menghilang dari bumi
After Lowering Case	:	capek pengen menghilang dari bumi
After Normalization	:	capek ingin menghilang dari bumi
After Stop-word Removal	:	capek ingin menghilang bumi
After Stemming	:	capek ingin hilang bumi
After Tokenisation	:	'capek', 'ingin', 'hilang', 'bumi'

Text cleaning removes unnecessary characters from tweets, such as symbols, numbers, links, hashtags, etc. [12]. The tweets' punctuation marks, symbols, hashtags, etc., are deleted. Lowering case is the process of homogenising letters into lowercase [13]. All capital letters in tweets are converted into lowercase for more straightforward analysis. Normalisation is a process of improving word structure by changing spelling, or non-standard words into the standard form [14]; for example, the word 'pengen' is normalised to 'ingin'. The stopwords list was obtained from kaggle published by [15] with a total of 750 words. For the tweet on Table 6, word 'dari' is removed because it has a low information value. Stemming aims to change the word to its basic form [16]. The prefix 'meng' is removed from the word 'menghilang' so it becomes 'hilang'. Tokenisation is a process of splitting tweets into words. The sample of pre-processing results is given in Table 3. Table 4 shows the display of location data from all tweets that have successfully passed the location conversion process to be converted into provincial form as can be seen in Table 4.

TABLE 4
LOCATION CONVERSION RESULTS

Before Conversion	Mojokerto, Jawa Timur
After Conversion	Jawa Timur

C. *Training Data Labelling and Classification Model Testing*

The results of 2400 datasets were divided into test and training data and presented in a confusion matrix. The classification error is determined by comparing the actual label and the predicted label to obtain the accuracy value.

If the prediction class shows different results from the actual class label, then the classification results are erroneous. However, if the predicted class does not show the same result as the actual class label, the classification result is correct.

D. Data Processing

The data processing classifies tweets using a system based on calculations as in Equations (1) to (3) [17].

$$P(d) = P(c)\ x\ P(c)\ x \ldots x\ P(t_n|c) \tag{1}$$

P(c|d) is the probability of the document in class c, P(c) is the prior probability of class c, and P(tn|c) is the probability of the n-word in class c. Class c prior probability can be calculated using the formula in equation (2).

$$P(c) = \frac{N_c}{N} \tag{2}$$

Equation (2) calculates the probability by dividing the number of documents by class c (Nc) and the number of all documents (N).

$$P(c) = \frac{T_{ct}+1}{\sum \quad t' \in VT_{ct}+B'} \tag{3}$$

Equation (3) is the formula for calculating Laplace Smoothing in determining the conditional probability of term t in class c documents. Tct is the number of occurrences of term t in class c documents, the number of frequencies for all terms in class c, and B is the number of unique words.

The results of data processing using the Naïve Bayes method are carried out to provide a classification label on the dataset based on the training data that has been labelled manually, using the source code in Algorithm 1.

```
def klasifikasi(tweet):
    predict_MNB = model_MNB.predict(tweet)
    klasifikasi=[]
    label=[]
    if predict_MNB == 1:
        klasifikasi = "Normal"
        label = "1"
    if predict_MNB == 0:
        klasifikasi = "Terindikasi gangguan mental"
        label = "0"
    return klasifikasi, label
```

Algorithm 1. Naïve Bayes Predict

Positive and negative probabilities were used to obtain classification results for each data. If normal probability > mental disorder probability, then the tweet is labelled 0. If normal probability < mental disorder probability, then the tweet is labelled 1.

E. Data Visualisation

The visualisation displays the analysis results using the Tableau tools in charts and maps [18].

III. RESULTS

The results obtained in this study are based on Twitter data taken from July 2021 to April 2022, with a total of 3,034,826 tweets. The three scenarios of the testing and training data are shown in the Table 5.

TABLE 5
METHOD ACCURACY

Method	Scenario	Training: Testing	Accuracy	Precision	Recall	F1-Score
Lexicon Based	I	70: 30	78.6%	77.2%	80.5%	78.8%
	II	80: 20	79.5%	80%	79.7%	79.8%
	III	90:10	81.2%	80.8%	82.6%	81.7%
Naïve Bayes	I	70: 30	83.4%	83.7%	80.4%	82%
	II	80: 20	84.5%	86.3%	82.2%	84.2%
	III	90:10	85.5%	87.2%	82.8%	84.9%

The best results are from Scenario III, which uses training and testing of 90:10. The accuracy value was 81.2% for the lexicon-based method and 85.5% for the Naïve Bayes method. These results indicate that the Naïve Bayes method generates more accurate results for classification than the lexicon-based method. The mental health analysis data are visualised in bar charts and maps, as shown in Fig. 2.

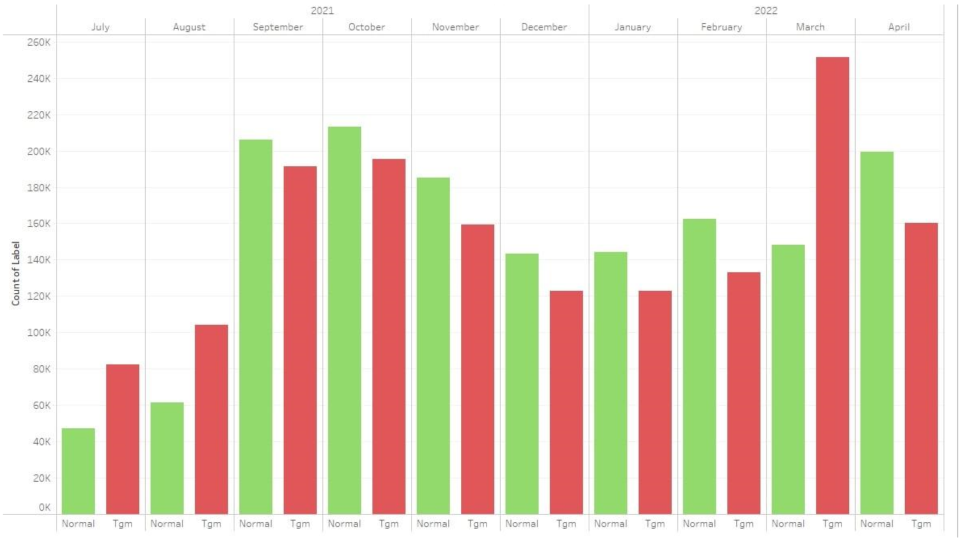

Fig. 2 Bar chart classification results each month

The label classification 'mental disorder indication' were more predominant in July and August 2021. From September to December 2021, the tweet data labelled 'mental disorder indication' decreased, and the 'normal' data began to increase, with a peak in October 2021, 213,367 tweets. However, in 2022, tweets indicating mental disorders rose again, hitting a peak in March, with as many as 251,329 tweets. The classification results are illustrated using the line chart and it can be seen in Fig. 3.

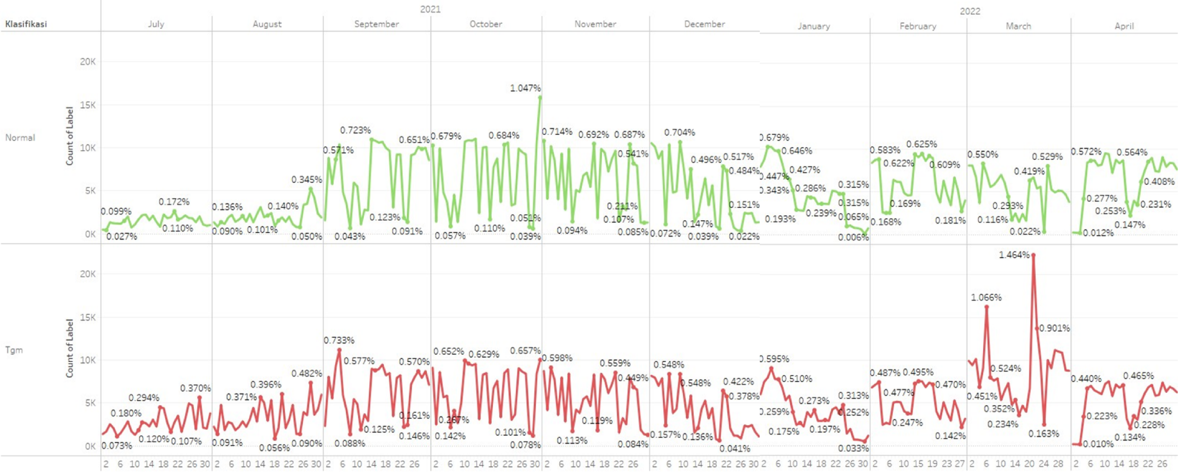

Fig. 3. Line chart classification results

Tweets labelled normal hit a peak on 31 October 2021 at 1.047%, which equals 15,832 tweets. The trending topics widely discussed as of that date were the costumes worn by K-pop artists on Halloween, and also the live streaming of the badminton match played by Kevin Sanjaya/Marcus Gideon at the French Open 2021. In 2022 tweets indicating mental disorders hit a peak on 21 March 2022 at 1.464%.

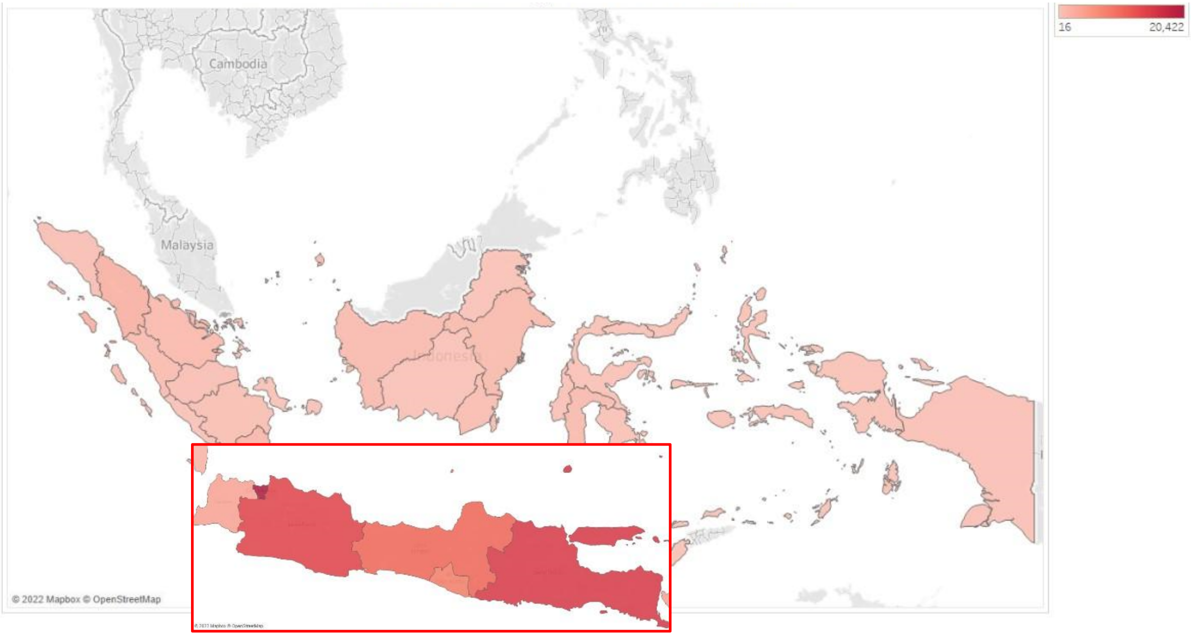

Fig. 4 Maps classification of "Indicated mental disorder"

The colour gradient on the maps in Fig. 4 shows the rank of the classification results by province, with Java provinces being the densest. The darker the colour in the circle in a province, the more the number of tweets indicating mental disorders. Jakarta emerged as the province with the most 1-labelled tweets, totalling 37,884.

IV. Discussion

This research aims to identify mental health trends using Twitter data. The captured tweets were filtered by Indonesian language and locations. The classification results show that the 'normal' label was dominant, which = increased from September to December 2021. This coincided with the decreasing number of COVID-19 cases. The Special Region of Jakarta showed the highest number of tweets with a 'mental disorder' label. The final percentage obtained was 52.3% for tweets classified as normal, and 47.7% for tweets classified as mental disorders.

The best classification accuracy from the three scenarios is the Naive Bayes method at 85.5%. Meanwhile, the best accuracy of the lexicon-based is 81.2%. This is in line with the accuracy values obtained by Chrismanto and Lukito [19], namely 78.49% for the Naive Bayes method and 77.25% for the SVM method. Research conducted by Saputra et al. [9] using a Lexicon-based tree structure obtained the best accuracy of 66.24%. Almouzini et al. [6] also calculated the error using a different method, so accuracy of 75.6% was obtained for the Naive Bayes method and 83% for the Random Forest method. The Naive Bayes method in this study shows the best accuracy because the classification using Naive Bayes is based on training data. By contrast, the lexicon-based method is only limited to comparing word dictionaries. If the words are not in the lexicon dictionary, then the classification algorithm will miss them. Le [20] mentioned that combining Naive Bayes and lexicon feature extraction methods can generate better results.

V. Conclusions

The conclusion that can be drawn is the classification using Naive Bayes generates better results than the lexicon-based approach, with the highest accuracy of 85.5%. Based on 3,034,826 tweets on Indonesian-language tweets in Indonesia, it was found that the 'normal' classification is slightly dominant, especially increasing from September to December 2021, with a final percentage ratio of 52.3% and 47.7%. This shows that there is a slight increase in the number of tweets labelled "normal". Tweets classified as 'mental disorders' are mostly found in Jakarta. The limitation of this study is that it only compares two methods. Future research can compare more methods, as well as increase accuracy by improving data quality at the data pre-processing stage.

Author Contributions: *Sheila Shevira*: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Visualization, Data Curation, Writing-Original Draft. *I Made Agus Dwi Suarjaya*: Supervision, Methodology,

Validation, Writing – Review & Editing. *Putu Wira Buana*: Supervision, Validation, Formal Analysis, Writing – Review & Editing.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] S. Naveed *et al.*, “Prevalence of Common Mental Disorders in South Asia: A Systematic Review and Meta-Regression Analysis,” *Front. Psychiatry*, vol. 11, no. September, pp. 1–8, 2020, doi: 10.3389/fpsyt.2020.573150.

[2] R. Yunita, “Aktivitas Pengungkapan Diri Remaja Putri Melalui Sosial Media Twitter,” *J. Komun.*, vol. 10, no. 1, pp. 26–32, 2019, doi: 10.31294/jkom.v10i1.5073.

[3] B. Nurfadhila, “Analisis Sentimen Untuk Mengukur Tingkat Indikasi Depresi Pada Twitter Menggunakan Text Mining,” no. 1, 2018.

[4] P. Noviyanti, A. Deolika, S. Hartinah, C. A. Haris, T. Maryana, and N. D. Sari, “Perbandingan Query Response Time pada Model Query View dan Cross Product,” *e-Jurnal JUSITI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 7–2, no. 2, pp. 131–141, 2018, doi: 10.36774/jusiti.v7i2.248.

[5] M. I. Maulana and A. A. Soebroto, “Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan Metode Improved k-Nearest Neighbor dan Seleksi Fitur Chi- square,” vol. 3, no. 7, pp. 6662–6669, 2019.

[6] S. Almouzini, M. Khemakhem, and A. Alageel, “Detecting Arabic Depressed Users from Twitter Data,” *Procedia Comput. Sci.*, vol. 163, pp. 257–265, 2019, doi: 10.1016/j.procs.2019.12.107.

[7] Arifin Kurniawan, Indriati Indriati, and Sigit Adinugroho, “Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 9, pp. 8335–8342, 2019.

[8] K. Aulia and L. Amelia, “Analisis Sentimen Twitter Pada Isu Mental Health Dengan Algoritma Klasifikasi Naive Bayes,” vol. 6, no. 2, pp. 60–65, 2020.

[9] F. T. Saputra, Y. Nurhadryani, S. H. Wijaya, and D. Defina, “Analisis Sentimen Bahasa Indonesia pada Twitter Menggunakan Struktur Tree Berbasis Leksikon,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 135, 2021, doi: 10.25126/jtiik.0814133.

[10] P. M. Mathapati, A. S. Shahapurkar, and K. D. Hanabaratti, “Sentiment Analysis using Naïve bayes Algorithm,” *Int. J. Comput. Sci. Eng.*, vol. 5, no. 7, pp. 75–77, 2017, doi: 10.26438/ijcse/v5i7.7577.

[11] N. Putu, A. Widiari, I. M. Agus, D. Suarjaya, and D. P. Githa, “Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali,” vol. 8, no. 2, pp. 137–145, 2020.

[12] D. Sebastian and K. A. Nugraha, “Text normalization for Indonesian abbreviated word using crowdsourcing method,” *2019 Int. Conf. Inf. Commun. Technol. ICOIACT 2019*, pp. 529–532, 2019, doi: 10.1109/ICOIACT46704.2019.8938463.

[13] P. Nambisan, Z. Luo, A. Kapoor, T. B. Patrick, and R. A. Cisler, “Social Media, Big Data, and Public Health Informatics: Ruminating behavior of depression revealed through Twitter,” *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 2015-March, no. March, pp. 2906–2913, 2015, doi: 10.1109/HICSS.2015.351.

[14] G. N. Aulia and E. Patriya, “Implementasi Lexicon Based Dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019,” *J. Ilm. Inform. Komput.*, vol. 24, no. 2, pp. 140–153, 2019, doi: 10.35760/ik.2019.v24i2.2369.

[15] F. Z. Tala, “A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia,” *M.Sc. Thesis, Append. D*, vol. pp, pp. 39–46, 2003.

[16] D. P. Andita Dwiyoga Tahitoe, “Implementasi Modifikasi Enhanced Confix Stripping Stemmer Untuk Bahasa Indonesia Dengan Metode Corpus Based Stemming,” *J. Ilm.*, pp. 1–15, 2010.

[17] Bustami, “Penerapan Algoritma Naive Bayes Untuk Nasabah Asuransi,” *J. Inform.*, vol. 8, no. 1, pp. 884–898, 2014.

[18] A. Kesumawati, “Visualisasi Data dengan Tableau (1),” *medium.com*, 2018. https://medium.com/@ayundyahkesumawati/visualisasi-data-dengan-tableau-8f1ff7eea464.

[19] A. R. Chrismanto and Y. Lukito, “Identifikasi Komentar Spam Pada Instagram,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. 3, p. 219, 2017, doi: 10.24843/lkjiti.2017.v08.i03.p08.

[20] C. C. Le, P. W. C. Prasad, A. Alsadoon, L. Pham, and A. Elchouemi, “Text classification: Naïve bayes classifier with sentiment Lexicon,” *IAENG Int. J. Comput. Sci.*, vol. 46, no. 2, pp. 141–148, 2019.