International Conference on Informatics Engineering, Science & Technology

Performance Evaluation of Naïve Bayes and SVM for Detecting Mental Health and Depression Sentiment in Social Media X

Agung Prayoga[1*], Rahma Wahdiniwaty[2], Dionisius Salvavictori Wanggur[1]

[1] Master of Information Systems, Universitas Komputer Indonesia, Bandung, Indonesia
[2] Universitas Komputer Indonesia, Bandung, Indonesia

[*]Corresponding E-mail: agung.75124001@mahasiswa.unikom.ac.id

Abstract. Highly effective strategies for early detection of mental health issues are urgently needed, given the increasing incidence of mental health issues among adolescents. This study investigates sentiment analysis as a tool for identifying depression in social media accounts. Selected reviews about depression will be used to compare two machine learning classifiers: Naive Bayes and Support Vector Machines (SVM). Data will be processed and vectorized using the TF-IDF technique, divided into 70% training and 30% testing. Algorithm performance will be evaluated using Recall, accuracy, precision, and F1 score. Support Vector Machines significantly outperformed the Naive Bayes model. SVM achieved 92.93% accuracy compared to 79.01% accuracy, 93.00% precision, and 92.93% recall. For Naive Bayes, the precision and recall were 80.71% and 80.71%. The determining factor for the superior machine learning performance of SVM is its ability to detect complex and non-linear linguistic aspects present in expressions of depression and mental health. These results demonstrate how advanced machine learning algorithms such as Support Vector Machines (SVMs) can form the basis for proactive and scalable mental health support systems.

Keywords: Sentiment Analyst, Naïve Bayes, SVM, Mental Health, Depression.

1. Introduction

Social media, such as platform X, has become a pervasive tool for people to share their thoughts, opinions, and feelings, including personal struggles with mental health and depression(Rahat et al., 2020). Mental health is a critical public health issue, defined by the World Health Organization (WHO) as "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community." Widespread conditions like anxiety and depression are becoming more prevalent, affecting millions worldwide. The WHO reports that over 280 million people suffer from depression, making it the second leading cause of illness globally. This growing crisis is exacerbated by social media's role as a platform for open, often unfiltered, expression, highlighting the urgent need to understand how these digital spaces reflect and influence public discourse on mental well-being. Many activities occur around us, and most people use social media to express their opinions and reviews. Social media site X is the most popular among other social media platforms (Wongkar & Angdresey, 2019). On social media X, people usually express their feelings about their mental health and depression problems. Mental health is a type of illness that affects a person's mental and psychological state. The most common are anxiety and depression (Gao et al., 2020; Shatte et al., 2019). Feeling sad and not interested are indications of depression. People often worry, don't think they matter, and lose interest (Abilkaiyrkyzy et al., 2024). It's also important to realize that mental diseases can make people do bad things, like hurt themselves or even kill themselves, which makes it a very serious issue (Hinduja et al., 2022). This is why it's important to talk about mental health concerns so that people don't do things that are bad for them or have long-term implications (Odja et al., 2024). More than 280 million people have depression, according to the World Health Organization (WHO). It is the second most common cause of illness in the world. Researchers saw that the number of people with serious depression and anxiety disorders grew by 28–26% over the world in the last year (Taquet et al., 2021). Tens of millions of additional people are experiencing depression and anxiety, in addition to the hundreds of millions already present.

Sentiment analysis will be used on public posts from the social media platform X in order to determine how online conversations about mental health impact public opinion. After gathering tweets with keywords associated with mental health and depression, the study will classify them into emotional groups, such as positive, negative, or other particular sentiments like sadness. This quantitative analysis will demonstrate the ways in which digital users express themselves, exposing patterns and trends that provide insight into how social media shapes public perceptions of this important issue.

2. Literature Review

2.1 X (Twitter).

X, formerly Twitter, is a social media platform that was officially launched on July 13, 2006. People can upload their short feelings (tweets) using mobile or web applications. With a character limit of 280, tweets are short and simple to read. Because of its special features, Twitter provides an almost limitless amount of text data for classification jobs(Khurniawan & Ruldeviyani, 2020). The platform is especially useful for text analysis and study because of its unique features, which include brevity, immediacy, and a wide range of themes. These characteristics set its messages apart from those of other social media (Fitri et al., 2019).

2.2 Text Mining

Text mining is the process of finding hidden information in written content. Text mining finds information that isn't obvious. But the storage that was given does not have it as information. So, information that comes from storage is not the same as hidden knowledge (Shabat & Abbas, 2020). Text mining is a growing field that combines things like Information Retrieval (IR), Information Extraction (IE), mining information from text and databases, visualization techniques, and technologies like deep learning, machine learning, statistics, computational linguistics, natural language processing (NLP), and corpus (Sakthi Vel, 2021).

2.3 Naïve Bayes

Many people use the Naive Bayes algorithm for analysis because it is easy and fast to use. This algorithm approach works very well with large data sets. This is why the Naive Bayes machine learning algorithm is often used for things like sorting text, identifying spam, and understanding people's feelings (Wang et al., 2023)(Ren et al., 2022)(Qorib et al., 2023). In sorting various things in statistics and probability, Thomas Bayes successfully developed the Naive Bayes machine learning (Putra et al., 2022). The NB model is an easy-to-use supervised machine learning algorithm. Naive Bayes can use information from the past to make predictions. This method can obtain prior probabilities for a given class label, examine its features, and then find the conditional probability of the target outcome (Peretz et al., 2024).

2.4 Support Vector Machine (SVM)

The SVM algorithm is a type of supervised learning, meaning the model requires labeled data to learn (Rahardi et al., 2022; Zahoor & Rohilla, 2020). Over the past few decades, computer science research has used a lot of the basic ideas underpinning the SVM method. These ideas include kernels, hyperplane margins, and other notions that help. Support Vector Machines are based on the idea of linear classifiers. Kernel methods have improved linear classifiers such that they can tackle nonlinear problems in higher-dimensional workspaces. This has brought attention to the theoretical and practical benefits of SVMs and made people more interested in research on pattern recognition. SVM machine learning models are valuable tools for solving real-world problems today, and they usually function better than alternative methods, such as artificial neural networks (Meliani et al., 2022).

2.5 Related Research

Sentiment analysis on social media for mental health has been the subject of numerous previous studies. Using a sentiment analysis technique, Yang and Luo investigated the community's mental health in social media during the COVID-19 epidemic. Their study identified signs of depression with an accuracy of 85% with a data set of 50,000 social media posts. The results of this research indicate that during the lockdown, there was a significant increase in negative sentiments about mental health (Cai et al., 2023).

Haris (Herdiansyah et al., 2022) Previous research showed that mental health disorders continue to be an issue related to changing social phenomena. Cultural stigma often leads to individuals with mental disorders being considered unproductive or problematic. Data from 5,537 tweets containing words like "emotion," "panic," and "stress" were categorized into three groups positive, neutral, and negative. This information suggests that Twitter is a safe place for people to discuss their feelings of mental illness. Using sentiment analysis techniques and various machine learning models such as k-Nearest Neighbors (k-NN), Random Forest, and Neural Networks.

Kasimirus (Odja et al., 2024) observed mental health on Reddit. The collected data covered the time before and after the epidemic. The Random Forest algorithm performed best in detecting depression, with an accuracy rate of 80.6% for recall, precision, and F1 score. The Neural Network model had the fewest errors of all algorithms (0.06496), although it was slightly less accurate (79%). On the other hand, the k-NN model made the most errors and was the least accurate. This study demonstrates the importance of building better models to detect and address problems early. This paper also discusses how machine learning and sentiment analysis can detect mental health issues on social media.

3. Method

3.1 Proposed Work

Data collection was conducted using the official API from X as a tool for word discovery with an emphasis on important keywords. Data collected from tweets on X will undergo a text preparation process, which includes cleaning, tokenization, normalization, and removal of unnecessary words. Machine learning algorithms are trained on the pre-processed dataset. After the data is processed, the model's performance is tested using assessment metrics such as accuracy, precision, recall, and F1 score. The goal of this method is to find broad and useful sentiment patterns in the X data. This section details the steps taken to implement the methodology in this study. Figure 1 shows the workflow of sentiment analysis.

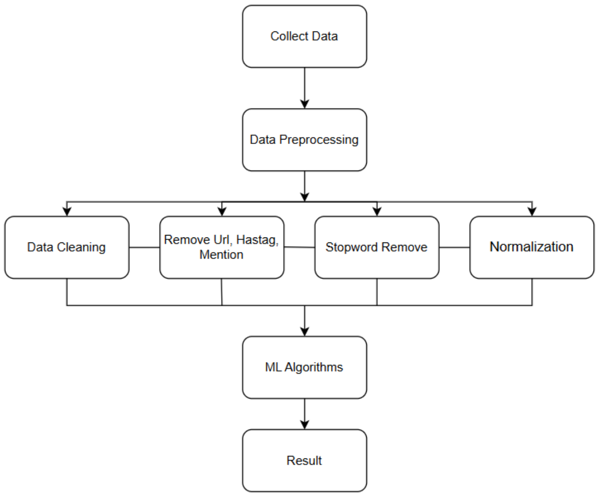

Figure 1. Workflow Of Sentiment Analysis

3.2 Collect Data

Data collection by crawling using the official API X as a word discovery tool with an emphasis on important keywords in the time span of January 2023 to October 2024 is to explore various perspectives and reviews on depression and mental health. The goal is to explore various perspectives and reviews on depression and mental health. 3,900 tweets have been obtained from the collection, consisting of reviews throughout the collection period covering perspectives on depression and mental health. These tweets will then be analyzed to extract information and draw conclusions for research purposes. Figure 2 shows data taken from X.

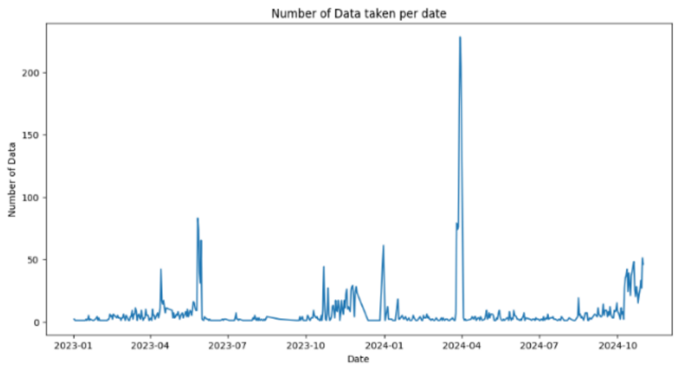

Figure 2. Data Taken from X

3.3 Remove URL, Hashtag, and Mention

Eliminating "@" and the URLs should be our first step. Once more, use "@" to indicate a username. Something is tagged with it. Therefore, any words that begin with "@" could be eliminated. The hashtag "#" appears in Twitter messaging. The hashtag is what makes a topic popular. There might be a slugger in the hashtag. #mentalhealth is one example. So, we opt to keep the whole word after the hashtag but leave off the "#" sign.

3.4 Normalization

Normalization is a procedure method in sentiment analysis, used to help clean, standardize, and prepare text to produce more precise and reliable sentiment analysis results.

Text preprocessing is the beginning in this process, used to transform the data into a format that matches the standards. Punctuation and some letters make the text easier to read and understand, although case folding reduces all text to lowercase and removes excessive capitalization.

3.5 Stopword Remove

Removing insignificant stopwords is an important part of getting data ready. Words like "the," "and," and "is" are examples of stopwords. They don't add much to the analysis. The Python Natural Language Tool Kit (NLTK) library makes this easier by giving you a list of the stopwords in 16 languages. When you take out the stopwords from the phrase "I like to eat, so I eat," the words "like," "eat," and "eat" are still there. This phase makes it easier to simplify content, which makes sentiment or text analysis more focused and useful..

4. Results and Discussion

4.1 Data Set

The purpose of this study was to investigate different viewpoints on depression and health by creating a mental health dataset from 3,900 tweets gathered by web crawling on platform X between January 2023 and October 2024. We observed generated by users evaluations including both positive and negative perspectives of depression and mental health. To get results, we examined these reviews. The Data Distribution of the Review Dataset is given in Figure. 3

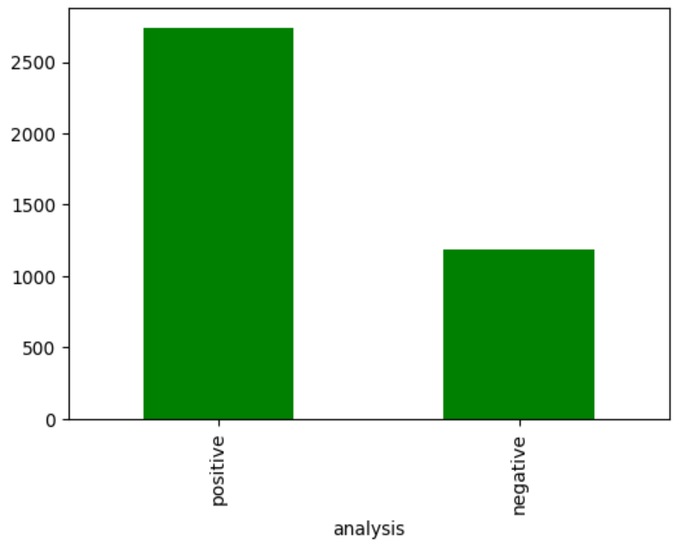

Figure 3. Data Distribution of the Review

4.2 WordCloud

The WordCloud visualisation of the most commonly occurring terminology in the review dataset is displayed in Figure 4. While the smaller words appear less frequently, the larger terms in the WordCloud show a higher frequency of occurrence. This image shows that respondents commonly use terms like "mental", "health", "depression," and "disorder," suggesting that these aspects are the primary emphasis of the reviews that were submitted. The primary emphasis or themes that are commonly discussed in the reviewed data analysis are summarized in this WordCloud. Figure 4. Shows the WordCloud – Vocabulary from Reviews

Figure 4. WordCloud – Vocabulary from Reviews

4.3 Confusion Matrix

In this part, two machine learning models are assessed using a variety of experiments. Key assessment metrics based on the confusion matrix, such as accuracy, recall, precision, and F-measure, are used to gauge classification efficacy.

TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) are the categories into which the confusion matrix separates the data. TP indicates true positive predictions, FP for incorrect positive predictions, TN for accurate negative predictions, and FN for incorrect negative predictions. Tables 1 and 2 display the confusion matrices for the Naive Bayes models and Support Vector Machine.

Table 1. Confusion Matrix Naive Bayes

NB	***Positive***	***Negatif***
Positive	TP = 566	FP = 157
Negative	FN = 8	TN = 55

Table 2. Confusion Matrix SVM

SVM Linear	***Positive***	***Negatif***
Positive	TP = 557	FP = 38
Negative	FN = 17	TN = 174

Using Metrics for Evaluation The equation shows that Precision is the prediction ratio among the total positive instance, based on TP, FP, TN, and FN.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

The equation shows that accuracy is the ratio of all the specimens in the class to the total number of samples.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

The equation that starts the scheme correctly sorts the advantageous classes from all the other classes and shows recall.

$$\text{Recall} = \frac{\text{TP}}{TP + FN}$$

F1 score looks at both accuracy and recall at the same time, however it uses the harmonic mean instead of the arithmetic mean, as shown in the equation.

$$\text{F1Score} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}}$$

4.4 Naïve Bayes and SVM

Before conducting the experiment, the dataset was divided with a 70% proportion for the training data, and the random state was set to 42. Two algorithms, namely Support Vector Machine (SVM) and Naive Bayes, are used to predict the outcome. In SVM, a linear kernel is used through SVC, while Naive Bayes applies the Naive Bayes Multinomial algorithm. Based on the analysis, the SVM algorithm produces a higher accuracy rate compared to Naive Bayes. The classification report results from both algorithms are presented in Table 3.

Table 3. Accuracy of Naïve Bayes Algorithm and SVM

	Naïve Bayes	***SVM***
Accuracy	79.01	93.00
Precision	80.71	92.93
Recall	79.00	93.00
F1 Measure	74.52	92.88

The results comparison of the two applying methods is shown in Table 3. SVM has an accuracy of 93.00, while Naïve Bayes has an accuracy of 79.01. SVM's precision and recall values are 92.93 and 93.00, while Naïve Bayes's are 80.71 and 79.00.

The comparison of the two algorithms is now displayed in Figure 5. In this experiment, SVM provides almost 93.00% accuracy, while Naïve Bayes yields 79.01% accuracy.

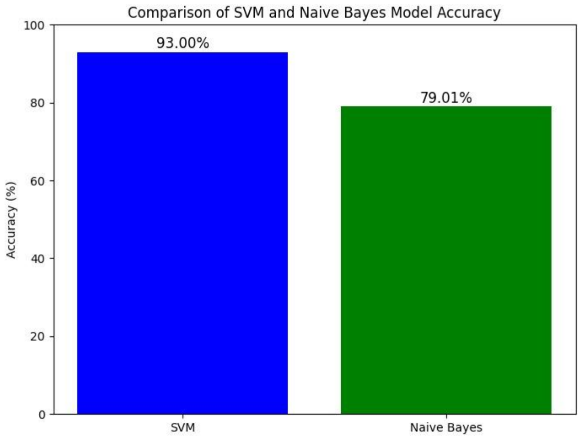

Figure 5. Accuracy Comparison of Both Algorithms

4.5 Discussion

An overview of the findings and discussion 3,900 tweets pertaining to mental health and depression that were gathered from the X platform were examined in this study. Naïve Bayes (NB) and Support Vector Machine (SVM) were the two classification models whose performance was to be compared.

Words like "mental," "health," and "depression" are the most commonly used terms, as shown by the WordCloud visualization, confirming that the dataset is in fact pertinent to the study issue. Accuracy, precision, recall, and F1-Measure—standard performance metrics that were computed from their confusion matrices—were used to evaluate the model. A critical element in the context of health, the confusion matrix itself is significant since it not only displays the overall accuracy number but also highlights the model's errors (for example, the number of missed or incorrectly classified tweets concerning depression that are labeled as positive)(Jannah & Kusnawi, 2024).

SVM exceeds Naïve Bayes (79% accuracy) by a wide margin, according to the data. This benefit is a result of the two algorithms' basic differences rather than an accident. Although this is frequently not the case in natural language—for instance, the words "not" and "good" have various meanings when combined—Naive Bayes makes the "naive" assumption that every word in a sentence is independent of every other word. SVM, on the other hand, is a discriminative model that can handle complex and high-dimensional text input since it looks for the best separation border (hyperplane) to separate sentiment classes. (Bhardwaj et al., 2024; Liem et al., 2024).

This result implies that precise SVM models can be an effective tool for social media mental health monitoring. A 93% accuracy rate suggests that this algorithm may be trusted to identify legitimate sentiment patterns and support public health initiatives. The study's shortcomings, however, are its exclusive focus on a single platform and the difficulties in deciphering linguistic subtleties like irony or sarcasm. Thus, more sophisticated deep learning models, like BERT, which are intended to better comprehend context and word relationships, could be investigated in future research in order to potentially provide deeper analysis and higher accuracy (Huang, 2023).

5. Conclusion

The results of the analysis using data from Twitter combined with machine learning using two algorithms, namely naive Bayes and Support Vector Machine, concluded the importance of conducting sentiment analysis on mental health and depression. When viewed from the use of technology, this study aims to assess how well machine learning algorithms can understand people's feelings about this issue. This is a step in helping prevent and can also provide support for decision-making in the health sector. Some important things done in the analysis stage include: preparation of the tweet discussion data that has been collected. Then the data is used to train and explore two machine learning algorithms. Then, they continued with the next stage by classifying feelings into two groups, namely positive and negative.

Support Vector Machines (SVMs) appear to excel at analyzing sentiment about mental health and depression, with an accuracy of around 93%, significantly higher than Naïve Bayes' 79%. However, this difference suggests that SVMs may be better at capturing the subtle and complex patterns hidden in social media posts about mental health.

These results can have a very significant impact in creating a machine learning-based system that can reveal mental health problems early.SVM's exceptional accuracy demonstrates that it can be a helpful tool for corporations and mental health providers. We can swiftly and successfully assist those who are depressed or suffering from emotional problems by accurately tracking the moods of public users on social media. To increase accuracy and the capacity to integrate scenarios in real time, more research on multi-class sentiment classification is advised. This study should make use of a variety of social media sites and combine deep learning techniques.

.

References

Abilkaiyrkyzy, A., Laamarti, F., Hamdi, M., & Saddik, A. El. (2024). Dialogue System for Early Mental Illness Detection: Toward a Digital Twin Solution. *IEEE Access*, *12*(January), 2007–2024. https://doi.org/10.1109/ACCESS.2023.3348783

Bhardwaj, A., Yadav, T., & Chaudhary, R. (2024). Predicting Hotel Booking Cancellations using Machine Learning Techniques. *2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024*, 1–6. https://doi.org/10.1109/ICCCNT61001.2024.10725148

Cai, M., Liu, P., Xu, C., & Luo, H. (2023). Editorial: Understanding the impact of social media on public mental health and crisis management during the COVID-19 pandemic. In *Frontiers in psychology* (Vol. 14, p. 1304586). https://doi.org/10.3389/fpsyg.2023.1304586

Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, *161*, 765–772. https://doi.org/10.1016/j.procs.2019.11.181

Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., Wang, Y., Fu, H., & Dai, J. (2020). Mental health problems and social media exposure during COVID-19 outbreak. *PLoS ONE*, *15*(4), 1–10. https://doi.org/10.1371/journal.pone.0231924

Herdiansyah, H., Roestam, R., Kuhon, R., & Santoso, A. S. (2022). Their post tell the truth: Detecting social media users mental health issues with sentiment analysis. *Procedia Computer Science*, *216*(2022), 691–697. https://doi.org/10.1016/j.procs.2022.12.185

Hinduja, S., Afrin, M., Mistry, S., & Krishna, A. (2022). Machine learning-based proactive social-sensor service for mental health monitoring using twitter data. *International Journal of Information Management Data Insights*, *2*(2), 100113. https://doi.org/10.1016/j.jjimei.2022.100113

Huang, K. (2023). *Recent Deep Learning Techniques for Sentiment Analysis*. https://doi.org/10.1109/ICNETIC59568.2023.00047

Jannah, N. Z. B., & Kusnawi, K. (2024). Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces. *Sinkron*, *8*(2), 727–733. https://doi.org/10.33395/sinkron.v8i2.13559

Khurniawan, F. S., & Ruldeviyani, Y. (2020). Twitter Sentiment Analysis: Case Study on the Revision of the Indonesia's Corruption Eradication Commission (KPK) Law 2019. *2020 International Conference on Data Science and Its Applications, ICoDSA 2020*. https://doi.org/10.1109/ICoDSA50139.2020.9212851

Li, Z., Li, R., & Jin, G. (2020). Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary. *IEEE Access*, *8*, 75073–75084. https://doi.org/10.1109/ACCESS.2020.2986582

Liem, S., Setiawan, T., & Pribadi, M. R. (2024). Perbandingan Algoritma SVM dan Naïve Bayes Berbasis SMOTE dalam Analisis Sentimen Komentar Tiktok pada Produk Skincare. *Applied Information Technology and Computer Science (AICOMS)*, *3*(1), 28–32. https://doi.org/10.58466/aicoms.v3i1.1523

Liu, H., Chen, X., & Liu, X. (2022). A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis. *IEEE Access*, *10*, 32280–32289. https://doi.org/10.1109/ACCESS.2022.3160172

Meliani, B. P., Pratiwi, O. N., & Andreswari, R. (2022). Comparison of Support Vector Machine and Random Forest Algorithms in Sentiment Analysis on Covid-19 Vaccination on Twitter Using Vader and Textblob Labelling. *Proceedings of the International Conference on Applied Science and Technology on Social Science 2022 (ICAST-SS 2022)*, 620–626. https://doi.org/10.2991/978-2-494069-83-1_108

Obiedat, R., Al-Darras, D., Alzaghoul, E., & Harfoushi, O. (2021). Arabic Aspect-Based Sentiment Analysis: A Systematic Literature Review. *IEEE Access*, *9*, 152628–152645. https://doi.org/10.1109/ACCESS.2021.3127140

Odja, K. D., Widiarta, J., Purwanto, E. S., & Ario, M. K. (2024). ScienceDirect ScienceDirect Mental illness detection using sentiment analysis in social media Mental illness detection using sentiment analysis in social media. *Procedia Computer Science*, *245*, 971–978. https://doi.org/10.1016/j.procs.2024.10.325

Peretz, O., Koren, M., & Koren, O. (2024). Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. *Engineering Applications of Artificial Intelligence*, *136*(PB), 108972. https://doi.org/10.1016/j.engappai.2024.108972

Putra, Y. H., Wahdiniwaty, R., Rahman, N. A., & Zukarnain, Z. A. (2022). *PENENTUAN PENULARAN COVID-19 BERDASARKAN GEJALA KANDIDAT COVID-19 MEMANFAATKAN ALGORITMA NAÏVE BAYES*. *12*(1), 406–412.

Qorib, M., Oladunni, T., Denis, M., Ososanya, E., & Cotae, P. (2023). Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Systems with Applications*, *212*(August 2022), 118715. https://doi.org/10.1016/j.eswa.2022.118715

Rahardi, M., Aminuddin, A., Abdulloh, F. F., & Nugroho, R. A. (2022). Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia. *International Journal of Advanced Computer Science and Applications*, *13*(6), 534–539. https://doi.org/10.14569/IJACSA.2022.0130665

Rahat, A. M., Kahir, A., & Masum, A. K. M. (2020). Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. *Proceedings of the 2019 8th International Conference on System Modeling and Advancement in Research Trends, SMART 2019*, 266–270. https://doi.org/10.1109/SMART46866.2019.9117512

Ren, Q., Zhang, H., Zhang, D., Zhao, X., Yan, L., Rui, J., Zeng, F., & Zhu, X. (2022). A framework of active learning and semi-supervised learning for lithology identification based on improved naive Bayes. *Expert Systems with Applications*, *202*, 117278. https://doi.org/https://doi.org/10.1016/j.eswa.2022.117278

Sakthi Vel, S. (2021). Pre-Processing techniques of Text Mining using Computational Linguistics and Python Libraries. *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, 879–884. https://doi.org/10.1109/ICAIS50930.2021.9395924

Shabat, H. A., & Abbas, N. A. (2020). Independent Component Analysis Based on Natural Gradient Algorithm for Text Mining. *Proceedings of 2020 1st Information Technology to Enhance E-Learning and Other Application Conference, IT-ELA 2020*, 72–76. https://doi.org/10.1109/IT-ELA50150.2020.9253072

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(9), 1426–1448. https://doi.org/10.1017/S0033291719000151

Taquet, M., Holmes, E. A., & Harrison, P. J. (2021). Depression and anxiety disorders during the COVID-19 pandemic: knowns and unknowns. *The Lancet*, *398*(10312), 1665–1666. https://doi.org/10.1016/S0140-6736(21)02221-2

Wang, S., Ren, J., & Bai, R. (2023). A semi-supervised adaptive discriminative discretization method improving discrimination power of regularized naive Bayes. *Expert Systems with Applications*, *225*(April), 120094. https://doi.org/10.1016/j.eswa.2023.120094

Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*, 1–5. https://doi.org/10.1109/ICIC47613.2019.8985884

Zahoor, S., & Rohilla, R. (2020). Twitter Sentiment Analysis Using Machine Learning Algorithms: A Case Study. *Proceedings - 2020 International Conference on Advances in Computing, Communication and Materials, ICACCM 2020*, 194–199. https://doi.org/10.1109/ICACCM50413.2020.9213011