1738850350

1738850350

Shri Vile Parle Kelavani Mandal

Document Details

Submission ID
trn:oid:::9832:81279438

Submission Date
Feb 6, 2025, 1:59 PM UTC

Download Date
Feb 6, 2025, 2:00 PM UTC

File Name
turn3-Tekrek_Riyantoro_Rev_02.docx

File Size
515.5 KB

9 Pages

2,864 Words

17,238 Characters

14% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Cited Text

Match Groups

37 Not Cited or Quoted 14%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

1 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

7% Internet sources
6% Publications
11% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **37** Not Cited or Quoted 14%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **1** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 7% Internet sources
- 6% Publications
- 11% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	islingtoncollege on 2025-01-16 — 1%
2	Submitted works	Kadir Has University on 2013-05-02 — <1%
3	Submitted works	Higher Education Commission Pakistan on 2024-03-13 — <1%
4	Submitted works	University of Abertay Dundee on 2024-08-11 — <1%
5	Internet	repository.uhamka.ac.id — <1%
6	Submitted works	University of Hertfordshire on 2023-01-09 — <1%
7	Publication	Ayat Droos, Awss Al-Mahadeen, Tasnim Al-Harasis, Rama Al-Attar, Mohammad Ab... — <1%
8	Internet	scitepress.org — <1%
9	Submitted works	Robert Kennedy College on 2024-10-23 — <1%
10	Submitted works	Universitas Kristen Satya Wacana on 2019-06-25 — <1%

11 Submitted works
University of Hertfordshire on 2023-03-13 <1%

12 Internet
garuda.kemdikbud.go.id <1%

13 Publication
Olga Krupková, Libor Dušek, Jiří Cuhorka, Graça Soares, Gabriela Kuchtová, Petr ... <1%

14 Internet
jurnal.bsi.ac.id <1%

15 Submitted works
uaq on 2024-11-17 <1%

16 Publication
Singh, Gurvinder. "Real-Time Quantum Computing Anomaly Detection Model on ... <1%

17 Publication
Tri Herdiawan Apandi, Castaka Agus Sugianto. "Algoritma Naive Bayes untuk Pre... <1%

18 Submitted works
University of East London on 2023-06-25 <1%

19 Submitted works
islingtoncollege on 2025-01-16 <1%

20 Internet
journal.umg.ac.id <1%

21 Publication
Arifur Rahman Rifath, Md Golam Muktadir, Mahmudul Hasan, Md Ashraful Islam.... <1%

22 Publication
Qurban A. Memon, Shakeel Ahmed Khoja. "Data Science - Theory, Analysis, and A... <1%

23 Submitted works
University of Greenwich on 2012-02-01 <1%

24 Submitted works
University of Greenwich on 2013-09-29 <1%

25 Submitted works
University of Salford on 2022-12-20 <1%

26 Internet
harbinengineeringjournal.com <1%

27 Internet
polgan.ac.id <1%

28 Internet
www.researchgate.net <1%

29 Submitted works
Imperial College of Science, Technology and Medicine on 2015-02-19 <1%

30 Publication
Kania Ardhani Putri, Wikky Fawwaz Al Maki. "Enhancing Pneumonia Disease Class... <1%

31 Submitted works
UIN Syarif Hidayatullah Jakarta on 2023-02-27 <1%

32 Submitted works
University of Glamorgan on 2023-10-24 <1%

33 Submitted works
Universitas Nasional on 2022-01-04 <1%

Analisis Sentimen Data Twitter Kabinet Indonesia Maju dengan Algoritma Naïve Bayes dan SVM

[1]Riyantoro, [2]Fauziah
[1, 2] Faculty of Communication and Information Technology, Nasional University
Pasar Minggu, Jakarta, 12520, Indonesia

[1]riyantoro2022@student.unas.ac.id, [2]fauziah@civitas.unas.ac.id

Abstrak

Twitter menjadi sarana penyebaran informasi yang banyak dimanfaatkan oleh masyarakat internet. Platform ini juga menjadi sumber data yang dapat digunakan untuk analisis sentimen dan pengambilan keputusan. Dalam konteks ini, analisis sentimen digunakan untuk secara otomatis mengkategorikan tweet pengguna ke dalam opini positif atau negatif. Kabinet Indonesia Maju adalah pemerintahan Indonesia yang dipimpin oleh Presiden Joko Widodo. Setelah para menteri dilantik, berbagai pendapat publik bermunculan mengenai kinerja dan tanggung jawab mereka. Salah satu cara untuk mengelompokkan opini yang tersebar di media sosial adalah dengan menerapkan analisis sentimen. Penelitian ini menggunakan dataset yang dikumpulkan melalui proses crawling di Twitter dengan kata kunci "Menteri Jokowi". Data yang diperoleh kemudian dianalisis menggunakan dua algoritma, yaitu Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM), guna membandingkan hasil cross-validation keduanya. Hasil analisis menunjukkan bahwa algoritma Naïve Bayes Classifier memiliki akurasi sebesar 91,70%, recall 91,69%, dan presisi 91,69%. Sementara itu, algoritma Support Vector Machine (SVM) mencapai akurasi 96,77%, recall 96,71%, dan presisi 96,71%. Berdasarkan hasil ini, dapat disimpulkan bahwa kedua algoritma dapat digunakan secara efektif dalam klasifikasi opini terkait menteri berdasarkan analisis sentimen, meskipun SVM menunjukkan akurasi yang lebih tinggi.

***Kata Kunci**: Naïve Bayes; Kabinet Presiden; Analisis Sentimen; Support Vector Machine; Twitter*

Abstract

Twitter has become a widely used platform for information dissemination among internet users. This platform also serves as a data source that can be utilized for sentiment analysis and decision-making. In this context, sentiment analysis is used to automatically categorize user tweets into positive or negative opinions. The Indonesia Maju Cabinet is the government of Indonesia led by President Joko Widodo. After the ministers were inaugurated, various public opinions emerged regarding their performance and responsibilities. One way to categorize the opinions circulating on social media is by applying sentiment analysis. This study uses a dataset collected through a crawling process on Twitter with the keyword "Menteri Jokowi" (Jokowi's Ministers). The obtained data was then analyzed using two algorithms: Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM), to compare their cross-validation results. The analysis results show that the Naïve Bayes Classifier algorithm achieved 91.70% accuracy, 91.69% recall, and 91.69% precision. Meanwhile, the Support Vector Machine (SVM) algorithm achieved 96.77% accuracy, 96.71% recall, and 96.71% precision. Based on these results, it can be concluded that both algorithms can be effectively used for classifying opinions about ministers through sentiment analysis, although SVM demonstrates higher accuracy.

***Keywords**: Naïve Bayes; Presidential Cabinet; Sentiment Analysis; Support Vector Machine; Twitter*

INTRODUCTION

Advancements in information and communication technology have significantly impacted the lifestyle changes of the global community. The internet has become a vast source of information, enabling anyone to access knowledge on a variety of topics. These changes demand regulations that function as guidelines and possess formal legal authority to ensure that no parties are disadvantaged in online activities. The Indonesian government responded to this need by enacting Law No. 11 of 2008 on Electronic Information and Transactions (ITE Law), which governs various aspects of electronic information (Judijanto & Khuan, 2024).

According to survey by We Are Social, a British media company collaborating with Hootsuite, Indonesians spend an average of 3 hours and 23 minutes daily on social media. Based on a report published on January 30, 2018, titled "Essential Insights Into Internet, Social Media, Mobile, and E-Commerce Use Around The World," out of Indonesia's total population of 265.4 million, 130 million are active social media users, with a penetration rate of 49 percent (Castellanos et al., 2021).

The Advanced Indonesia Cabinet (Kabinet Indonesia Maju), which is the governing cabinet of Indonesia under President Joko Widodo, consists of four coordinating ministers and 30 sectoral ministers. The cabinet was announced on October 23, 2019, and inaugurated through Presidential Decree No. 113/P of 2019 for the 2019-2024 term. With the inauguration of ministers in the cabinet, political figures who now serve as ministers have become the subject of various public opinions. Before the internet, people expressed their opinions, criticism, and suggestions through print media. However, with the rapid development of technology, many online platforms have emerged as avenues for people to voice their thoughts.

Today, online platforms like Facebook, Twitter, Instagram, and others have become highly popular communication tools among internet users. Twitter, now more commonly referred to as X, is one of the social media platforms closely linked to user sentiments, as reflected in their tweets (Susanto et al., 2021; Syahputra et al., 2022). The growth of Twitter usage in recent years has been driven by its faster dissemination of information compared to digital or print news media (Barbosa & Feng, 2010). Consequently, researchers have developed systems to extract information from Twitter conversations, with one critical piece of information being public sentiment(Syah et al., 2023). Sentiment analysis of Twitter data has become important as it has the potential to influence evaluations of the performance of the Advanced Indonesia Cabinet for the 2019-2024 period. The objective of sentiment classification is to automatically group user opinions into positive or negative sentiments (Monika et al., 2022)

Sentiment analysis, or opinion mining, is the computational study of opinions, behaviors, and emotions regarding a specific entity. These entities may include individuals, events, or topics. Common classification techniques used in sentiment analysis include Naïve Bayes, Support Vector Machines (SVM), and K-Nearest Neighbor (KNN) (Liem et al., 2024; Madjid et al., 2023). Several studies have employed different sentiment analysis techniques on Twitter data, using a variety of algorithms to examine public sentiment on various topics. Pebrianto et al. (2022) conducted an analysis of public sentiment regarding Indonesian ministers by utilizing Support Vector Machine (SVM) and Naïve Bayes Classifier (NBC) algorithms (Pebrianto et al., 2022). Their study found that SVM performed slightly better, achieving higher accuracy, recall, and precision compared to Naïve Bayes. In a similar result, Ismail et al.(2023) applied Fine-Grained Sentiment Analysis to tweets related to the Indonesian presidential debates,

categorizing tweets into positive, negative, or neutral based on the dominance of positive or negative sentences (Ismail et al., 2023). Their results indicated that most tweets carried a positive sentiment. Joko Haryanto et al. (2018) focused on product reviews in Indonesian, using SVM with a Polynomial Kernel and integrating Query Expansion to improve classification accuracy (Joko Haryanto et al., 2018). This approach achieved an impressive accuracy rate of 96.25%, demonstrating the value of expanding the vocabulary in sentiment analysis. In another study, Winarno et al. (2023) explored sentiment analysis regarding the Jokowi administration's social welfare initiatives during the pandemic using a Lexicon-Based approach (Winarno et al., 2023). This study found that the majority of tweets were classified as neutral, with positive sentiment second and negative sentiment third. Finally, Muttaqin (2025) compared Naïve Bayes and SVM to analyze sentiment about the relocation of Indonesia's capital. The study concluded that SVM outperformed Naïve Bayes with a higher accuracy rate (Muttaqin et al., 2025)

Therefore, the objective of this study is to compare the performance of Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) in sentiment analysis of Indonesian political figures, specifically ministers, based on public perceptions of their performance. By leveraging Twitter as a source of real-time public opinion, this research aims to provide both theoretical and practical contributions to the field of sentiment analysis. Additionally, it seeks to evaluate the effectiveness of political leadership in the digital space by analyzing algorithmic performance within the context of Indonesian political sentiment.

METHOD

Conceptual Framework

This study provides both theoretical and practical insights, particularly for political analysts, policymakers, and researchers interested in real-time public opinion monitoring. Guided by the conceptual framework (Fig. 1), the research systematically evaluates the strengths and weaknesses of each sentiment analysis algorithm.

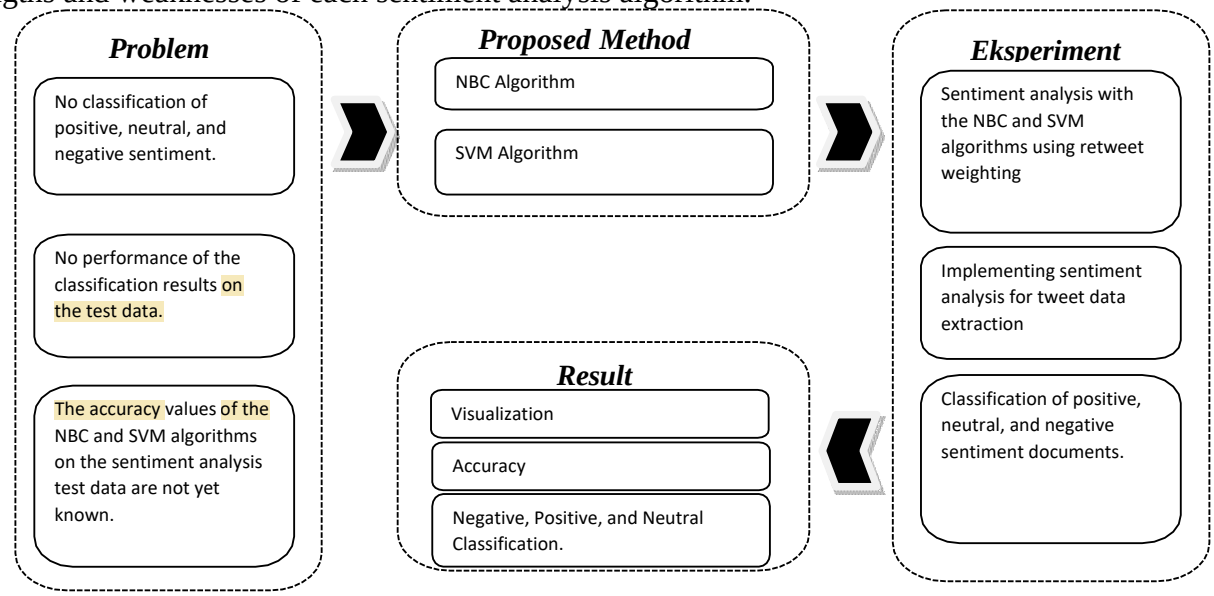

Figure 1. Conceptual Framework

Data Collection

This study collected 29,670 tweets using the Twitter API integrated with R Studio, employing keywords such as "#menterijokowi," "kabinet Indonesia maju," and "menteri jokowi" to capture public sentiment on the Kabinet Indonesia Maju (2019–2024). The API facilitated efficient web crawling and data extraction based on predefined criteria, ensuring relevance to the topic.

Preprocessing Steps

The preprocessing step involves removing unnecessary symbols, converting text to lowercase, replacing slang with standard words, reducing words to their root form, splitting text into tokens, and eliminating common stopwords to enhance analysis.

Weighting and Sentiment Classification

The weighting and sentiment classification step involves assigning weights to tweets based on polarity and subjectivity using a lexicon-based approach, followed by categorizing them into sentiment classes (positive, negative, neutral) based on the calculated weights.

Machine Learning Model Selection

1. **Naïve Bayes Classifier (NBC)**

Naïve Bayes is based on Bayes' Theorem, which expressed as: (Eq.1)

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad \text{..... (1)}$$

Explanation:

P(A|B): posterior probability.
P(B|A): likelihood probability
P(A): The probability of event A.
P(B): The probability of event B (evidence).

2. **Support Vector Machines (SVM)**

In the context of SVM, a hyperplane can be defined as: (Eq. 2)

$$w \cdot x + b = 0 \quad \text{..... (2)}$$

Where:
w: is the weight vector.
x: is the feature vector of the data.
b: is the bias or offset of the hyperplane.

The goal of SVM is to find the values of w and b that maximize the margin, which is calculated as: (Eq. 3)

$$\text{Margin} = \frac{2}{\|w\|} \quad \text{.....(3)}$$

To maximize the margin, SVM minimizes ‖w‖ (the norm of the weight vector) with the constraint that all data points must be correctly classified: (Eq. 4)

$$y_i(w \cdot x_i + b) \geq 1 \quad \text{untuk semua } i \qquad \text{.....(4)}$$

Where yi is the class label (for example, +1 for the positive class and -1 for the negative class).

RESULT AND DISCUSSION

The study identifies key patterns in the dataset, such as the dominance of certain keywords like "Prabowo," "Papua," and "Radikal," which indicate public discussions on political and security issues. The sentiment distribution shows that 60% of the tweets are classified as positive, 30% as negative, and 10% as neutral (Fig. 2).

Figure 2. Word Cloud Dataset

The results show that SVM outperforms NBC in terms of accuracy, precision, recall, and F1-score. SVM achieves an accuracy of 96.71%, while NBC achieves an accuracy of 91.69% (Table 1).

Table 1. Comparison between NBC and SVM Algorithms

	NBC	SVM
Precision	91.69 %	96.71 %
Accuracy	91.70 %	96.77 %
Recall Rate	91.69 %	96.71 %
F1 Score	91.69 %	96.71 %

The confusion matrix for SVM shows fewer false positives and false negatives compared to NBC, indicating that SVM is better at separating the classes (Fig. 3).

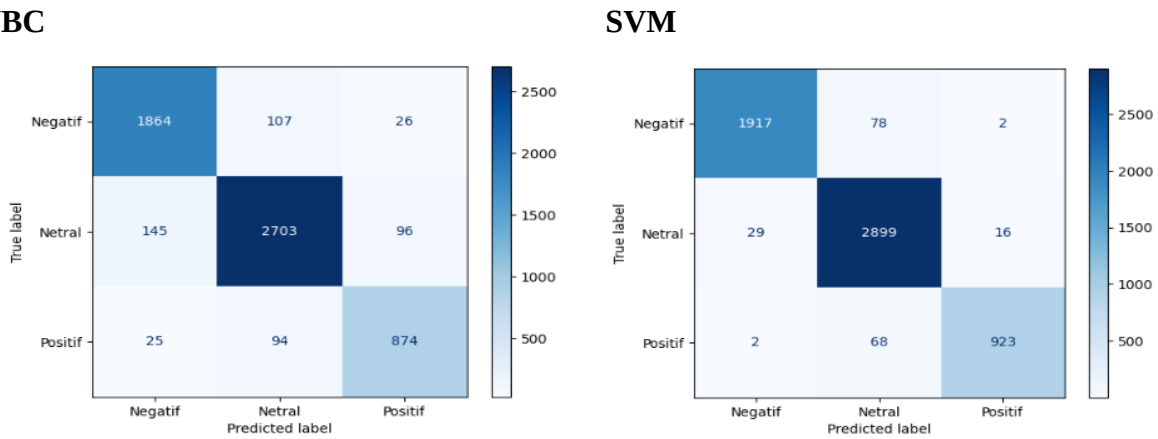

Figure .3 Confusion matrix comparison for NBC and SVM Algorithms

The analysis results indicate that the SVM algorithm outperforms the NBC algorithm in terms of accuracy and class separation. Several key performance metrics highlight the advantages of SVM over NBC. SVM achieved a precision of 96.71%, while NBC obtained 91.69%. This demonstrates that SVM is more effective in correctly predicting the target class in this dataset compared to NBC. The accuracy of SVM reached 96.77%, whereas NBC obtained 91.70%. This suggests that SVM is more effective in identifying tweets accurately, with fewer false positives than NBC as also mentioned by previous study (Sugitomo et al., 2021). As a result, SVM's positive predictions are more reliable. SVM achieved a recall rate of 96.71%, while NBC scored 91.69%. This indicates that the SVM algorithm is superior in detecting all instances within a class, meaning fewer positive classifications were missed. It shows that SVM has a greater capacity to capture variations in the data. SVM obtained an F1 score of 96.71%, compared to 91.69% for NBC. This highlights that SVM provides a better balance between precision and recall than NBC, as the F1 score is the harmonic mean of both.

The superiority of SVM over NBC can be attributed to several factors. NBC is a simple linear model that assumes independence between features (words), which is often not true in textual data where features can be interdependent. This limitation prevents NBC from capturing more complex patterns in the data. SVM is more capable of managing large datasets by searching for a hyperplane that maximizes the separation between classes. This makes SVM more effective at distinguishing overlapping or difficult-to-separate data points. While NBC is effective in many cases, it struggles when class boundaries are non-linear or complex. SVM,

utilizing kernel tricks, maps features into higher-dimensional spaces, allowing it to handle non-linear decision boundaries more effectively. NBC is more affected by ambiguous or irrelevant tweets, as well as imbalanced datasets, which can impact its accuracy. In contrast, SVM is more resilient to noisy and imbalanced data. NBC assumes that features are independent, which is rarely the case in real-world text data. Ignoring feature dependencies can degrade model performance. SVM, on the other hand, does not rely on feature independence, making it better suited to identifying complex relationships between features. Based on this comparison, it is evident that SVM consistently delivers superior results compared to NBC, making it a more effective choice for sentiment classification in this dataset.

However, there are several limitations to this study. Firstly, the dataset is restricted to tweets, which may not represent the diversity of public sentiment across different platforms or demographics. Additionally, the reliance on basic machine learning models like NBC and SVM might not capture the complexity of sentiment expressed in social media posts as effectively as more advanced techniques. Further improvements can be made by incorporating more advanced models, such as deep learning techniques like Long Short-Term Memory (LSTM) networks and Bidirectional Encoder Representations from Transformers (BERT), which have shown to provide higher accuracy in understanding context and nuances in language. Expanding the scope of the analysis to include other forms of media and using more sophisticated models could enhance the robustness and accuracy of sentiment analysis in political contexts.

CONCLUSION

The study concludes that SVM is more effective than NBC for sentiment analysis tasks, especially when dealing with complex and non-linear data. The preprocessing steps, including filtering, case folding, and stemming, are crucial for improving the accuracy of the models. Further improvements can be made by incorporating more advanced techniques such as deep learning models like LSTM and BERT. This journal article offering valuable insights into the application of NLP techniques for sentiment analysis on social media data.

ACKNOWLEDGEMENT

The authors would like to acknowledge Fitri Kurniasari, Ph.D for technical support.

DAFTAR PUSTAKA

Barbosa, L., & Feng, J. (2010). *Robust Sentiment Detection on Twitter from Biased and Noisy Data*. http://twittersentiment.appspot.com/

Castellanos, E., Xie, H., & Brenner, P. (2021). Global News Sentiment Analysis. In Z. Yang & E. von Briesen (Eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas* (pp. 121–139). Springer International Publishing.

Ismail, I., Nugroho, M. A., & Sulistiyowati, R. (2023). Sentiment Analysis Netizens on Social Media Twitter Against Indonesian Presidential Candidates in 2024 Using Naive Bayes Classifier Algorithm. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *7*(3), 1611. https://doi.org/10.30865/mib.v7i3.6536

Joko Haryanto, D., Muflikhah, L., & Ali Fauzi, M. (2018). *Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion* (Vol. 2, Issue 9). http://j-ptiik.ub.ac.id

Judijanto, L., & Khuan, H. (2024). Juridical Analysis of Law Number 11 of 2008 on Electronic Information and Transactions (ITE) and its Impact on Creative Economy Development in Indonesia Article Info ABSTRACT. In *West Science Law and Human Rights* (Vol. 2, Issue 04).

Liem, S., Setiawan, T., & Rizky, P. M. (2024). Comparison of SVM and Naïve Bayes Algorithms in Sentiment Analysis of TikTok Comments on Skincare Products. *Applied Information Technology and Computer Science (AICOMS)*, *3*(1), 28–32. https://doi.org/10.58466/aicoms.v3i1.1523

Madjid, M. F., Ratnawati, D. E., & Rahayudi, B. (2023). Sentiment Analysis on App Reviews Using Support Vector Machine and Naïve Bayes Classification. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, *7*(1), 556–562. https://doi.org/10.33395/sinkron.v8i1.12161

Monika, P., Kulkarni, C., Harish Kumar, N., Shruthi, S., & Vani, V. (2022). Machine learning approaches for sentiment analysis. *International Journal of Health Sciences*, 1286–1300. https://doi.org/10.53730/ijhs.v6ns4.6119

Muttaqin, A. F., Sunge, A. S., & Zy, A. T. (2025). Comparative Analysis of Earthquake Prediction with SVM, Naïve Bayes, and K-Means Models: Comparative Analysis of Earthquake Prediction with SVM, Naïve Bayes, and K-Means Models. *Journal of Computer Networks, Architecture and High Performance Computing*, *7*(1), 104–119. https://doi.org/10.47709/cnahpc.v7i1.5085

Pebrianto, R., Nugraha, S. N., Latif, A., & Firdaus, M. R. (2022). *ANALISIS SENTIMEN TWITTER TERHADAP MENTERI INDONESIA DENGAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES*. *17*, 1–12.

Sugitomo, J. C., Kevin, N., Jannatri, N., & Suhartono, D. (2021). Sentiment Analysis using SVM and Naïve Bayes Classifiers on Restaurant Review Dataset. *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, *1*, 100–108. https://doi.org/10.1109/ICCSAI53272.2021.9609776

Susanto, A., Maula, M. A., Mulyono, I. U. W., & Sarker, M. K. (2021). Sentiment Analysis on Indonesia Twitter Data Using Naïve Bayes and K-Means Method. *Journal of Applied Intelligent System*, *6*(1), 40–45. https://doi.org/10.33633/jais.v6i1.4465

Syah, R. I., Hoiriyah, H., & Walid, M. (2023). ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TERHADAP APLIKASI M-HEALTH PEDULI LINDUNGI DENGAN METODE LEXICON BASED DAN NAÏVE BAYES. *Indonesian Journal of Business Intelligence (IJUBI)*, *6*(1). https://doi.org/10.21927/ijubi.v6i1.3275

Syahputra, R., Yanris, G. J., & Irmayani, D. (2022). SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, *6*(2), 671–678. https://doi.org/10.33395/sinkron.v7i2.11430

Winarno, Kurniabudi, & Hendri. (2023). *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM) Analisis Sentimen Pada Twitter Mengenai Pemerintahan Jokowi Terkait Kesejahteraan Masyarakat Menggunakan Lexicon Based*. *3*(2). https://doi.org/10.33998/jakakom.v3i2