Journal of Technology and Computer (JOTECHCOM)
Vol. 3, No. 1, February 2026, pp. 68~76
ISSN: 3048-0477 (Media Online)

Comparative Machine Learning Analysis for Sentiment Classification of Sumatra Disaster 2025

Nauval Alfarizi[1], Prima Lydia[2], Muhammad Syahputra Novelan[3], Adi Putra[4], Satria Sinurat[5]
[1,2,3,4,5]Department of Information Technology, Faculty of Pascasarjana, Universitas Pembangunan Panca Budi, Medan, Indonesia
[1]nauvalalfarizi026@gmail.com, [2]anpu3@gmail.com, [3]putranovelan@dosen.pancabudi.ac.id, [4]myreal.autobot@gmail.com, [5]satria.sinurat@yahoo.co.id

Article Info

Article history:

Received January 04, 2026
Revised February 02, 2026
Accepted February 11, 2026

Keywords:

Sentiment Analyst
Naive Bayes
Support Vector Machine
K-Nearest Neighbor
Natural Disaster

ABSTRACT

Indonesia is highly vulnerable to natural disasters due to its geological position, resulting in extensive disaster-related news coverage that shapes public sentiment. This study presents a comparative machine learning analysis for sentiment classification of online news related to natural disasters in Sumatra during December 2025. The dataset was collected through web scraping from two major Indonesian news portals, like CNN Indonesia and Detik, and categorized into three sentiment classes: negative, neutral, and positive. Sentiment classification was conducted using Naive Bayes, Support Vector Machine (SVM), and k-Nearest Neighbors (KNN) algorithms.
The results demonstrate that Naive Bayes achieved accuracy values of 0.57 on the CNN Indonesia dataset and 0.61 on the Detik dataset. However, its performance was highly biased toward the dominant negative class, as indicated by low macro-average F1-scores of (0.24) and (0.39). In contrast, SVM showed the most balanced performance by reducing class bias, achieving accuracies of (0.68) and (0.67) with macro-average F1-scores of (0.51) and (0.59), respectively. KNN demonstrated moderate performance, with accuracy values of 0.60 and 0.59, but remained less effective than SVM in handling minority sentiment classes.

Corresponding Author:

Nauval Alfarizi
University of Pembangunan Panca Budi
Email: nauvalalfarizi026@gmail.com

1. INTRODUCTION

Indonesia occupies a strategic geological position at the intersection of three major tectonic plates. This condition is characterized by a high frequency of natural hazards, including earthquakes, landslides, and related disasters. The increasing frequency of such events underscores the need for comprehensive studies to assess regional vulnerability and develop effective strategies for mitigating natural disaster risks[1]. Social media has emerged as an important channel for public news consumption. Moreover, short-term increases in informational news use on messenger platforms are associated with a reduced propensity for online political expression, petition signing, and donation activities [2].

The results suggest that greater attention to science-oriented news is associated with increased feelings of anxiety and anger toward the issue. In contrast, attention to political news is related to emotional responses such as fear and sadness or depression, alongside anxiety and anger. Furthermore, mediation analysis indicates that attention to science news indirectly enhances support for preventive policy initiatives via anxiety. In contrast, attention to political news indirectly strengthens support for punitive policy measures through emotions of anger and fear [3].

Content analysis is commonly used to examine how communities affected by crises utilize social media to disseminate real-time information, seek assistance, and provide mutual support. Public engagement in such contexts is often characterized by dynamic fluctuations in sentiment and discussion intensity in response to disaster-related events. Overall, these characteristics illustrate the complexity of social media use within crisis communication contexts [4]. Sentiment analysis is a computational approach used to identify and classify public opinions expressed in textual data, particularly from social media platforms, to capture positive and negative public responses toward specific issues or public figures.

Previous studies have highlighted the association between emotions conveyed in online news media and subsequent emotional expressions on social media platforms[5], [6]. Despite the effectiveness of sentiment analysis as an automated tool, extracting public sentiment from news-based textual data remains challenging due to contextual complexity, implicit meanings, and the limited scope of existing review studies in this domain, particularly when applied to large-scale crawling of news related to natural disasters [7].

In this context, sentiment analysis is typically carried out using two main approaches: lexicon-based and machine learning-based methods. The lexicon-based approach relies on predefined sentiment dictionaries to assign polarity values and categorize text as positive, negative, or neutral. In contrast, machine learning-based methods employ data-driven techniques that enable models to learn sentiment patterns automatically[8], [9].

Previous studies have shown that machine learning-based sentiment analysis methods, particularly Naïve Bayes, Support Vector Machine (SVM), and k-Nearest Neighbor (KNN), are effective in text classification tasks. These algorithms have been used in various domains, including online user reviews and public opinion analysis, and have demonstrated strong performance in identifying sentiment patterns. [10], [11].

One of the main advantages of the Naive Bayes algorithm is its efficiency in handling textual data and its consistent performance, even when faced with high variability in the dataset. This effectiveness arises from Naive Bayes's reliance on probabilistic word distributions, which allows for more accurate sentiment classification. In contrast, K-Nearest Neighbors (KNN) relies on distance-based computations between vector representations, which can diminish its effectiveness when working with diverse text data [12].

Sentiment analysis is a branch of web content mining that utilizes machine learning techniques to extract and interpret user opinions from online using text data. By enabling the analysis of unstructured information, machine learning is essential for assessing user satisfaction, identifying emerging trends, and categorizing sentiments [13], [14]. This study employs a comparative machine learning approach using Naive Bayes and Support Vector Machine (SVM) algorithms to classify sentiment in web-scraped online news articles about natural disasters in Sumatra for the year 2025.

Previous studies highlight that k-Nearest Neighbor is a widely used and fundamental method in text mining and classification tasks. However, its performance is highly influenced by data dimensionality and computational complexity, motivating the development of optimized and alternative machine learning approaches [15].

2. METHOD

This research employed a comparative based on machine learning analysis, with the primary feature being the classification of sentiment analysis through web scraping. The data collection began by scraping trusted regional Indonesian news portals such as https://www.cnnindonesia.com/ and https://www.detik.com/. The scraping time was carried out from early to late December 2025, as this research is based on real-time data. To test the comparative machine learning, we do implemented a comparison using algorithms such as Naive Bayes, Support Vector Machine (SVM), and k-Nearest Neighbor (KNN), with almost procedural machine learning concepts, including extraction, cleaning, preprocessing, training, and performance evaluation.

2.1 Web Scrapping

Integrating web scraping as a tool is necessary in independent research, primarily due to the limited availability of datasets, which are often expensive and difficult to obtain, and their credibility is questionable. To overcome this problem, individuals can use a method that utilizes scraping or pulling data from a web-based source [16].

2.2 Sentiment Analysis

Sentiment analysis in social networks is an expanding area of research that has garnered significant interest beyond academia, with its diverse applications and the accessibility of data from social media platforms, such as Twitter, Facebook, and others. This type of analysis serves various purposes, including providing social insights and informing the research purpose [17].

Various machine learning methods can be utilized to classify sentiment in web-scraped online news data. These algorithms are well-suited for handling large volumes of textual data by effectively learning sentiment patterns from extracted features.

2.3 Naive Bayes

Naive Bayes assumes that the presence or absence of a unique characteristic of a class is unrelated to the attributes of other classes. Each feature has an independent class label. The Naive Bayes classifier in this study calculates the mean probability and standard deviation when classifying continuous features. [18].

2.4 Support Vector Machine

SVM (Support Vector Machine) used to train sentiment trends. SVM applies a classification method that works by determining the line of opinion formed through clusters in negative and positive forms [19].

2.5 K-Nearest Neighbor

K-Nearest Neighbor (KNN) regression used to predict the values of the target variable by examining the proximity between a test instance and its nearest training samples K-Nearest Neighbor (KNN) itself is part of a machine learning model that is very helpful in finding a solution to a problem [20].

2.6 Model Performance

The performance of the proposed models in this study is evaluated using widely accepted classification confusion metrics [21]. To evaluate the performance of the supervised classifiers, this study employs precision, recall, and F1-score metrics based on the confusion matrix. (1) Accuracy represents the ratio of correct predictions to the total number of predictions made by the classification model.

2.7 Text Preprocessing

Data written in a combination of foreign languages is often found in both digital and physical forms; therefore, it is necessary to easily demonstrate how to process text with various modular forms. Furthermore, although more work is needed to validate these results with multiple configurations, datasets, and classification techniques, the results quickly show the great potential of applying more comprehensive text pre-processing techniques when using supervised machine learning techniques to build text classification models, such as typo correction [22].

2.8 Research Framework

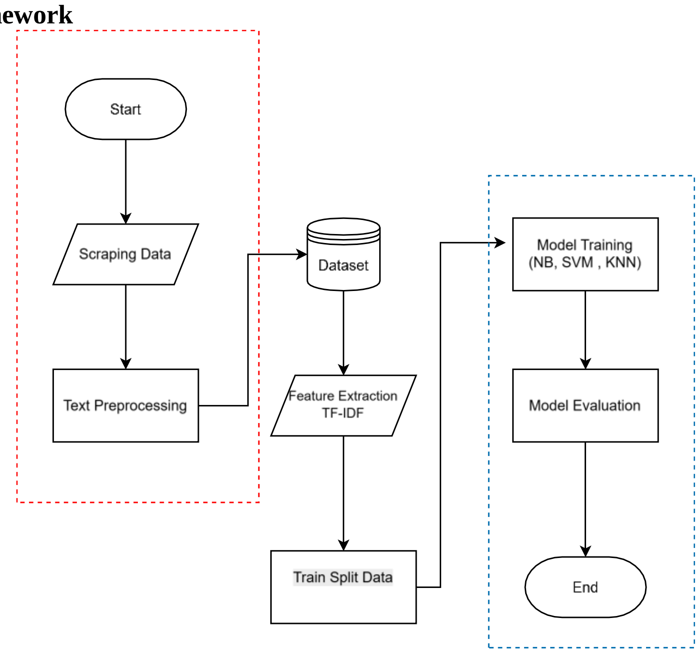


Figure 1. Research Framework Pipeline

This study is divided into two segments. Before starting the comparative analysis using the supervised machine learning algorithm to determine the segmentation and performance of each algorithm, it is necessary to ensure the dataset is relevant because later it will not only be in the form of the performance of the existing framework, but also show the results of segmentation and interpretation in visual form.

3. RESULTS AND DISCUSSION

3.1 Dataset Summary

In this study, the author searched for data based on news tags for the period of December 2025. The selected news tags were affiliated with or correlated with trending news tags in Indonesia. The selected tags were filtered based on subheading tags found on national news portal websites such as CNN Indonesia and Detik. The obtained data was scraped using Python to be processed into sentiment and to compare algorithm performance in comparative machine learning. performed a series of data searches on popular Indonesian national media channel platforms such as CNN Indonesia with the domain linkage https://www.cnnindonesia.com/ for pre-preparation to screen datasets that are suitable for analysis and also to get maximum results, experimenters only focused on specific tags such as https://www.cnnindonesia.com/tag/banjir-sumatra and for the next media channel platform Detik as a comparative value comparison with the domain https://www.detik.com/ with specific subdomains such as https://www.detik.com/tag/bencana-alam.

3.2 Evaluating The Perfomance of naive bayes

As illustrated in Table 1, the Naive Bayes classifier achieved an accuracy of 57% based on the data. The data underwent a Train-Test Split with an 80:20 ratio, which is a standard benchmark for initiating the process. The training model demonstrated strong recall for the negative class, indicating its effectiveness in identifying negative sentiment in disaster-related news from Sumatra in 2025.

Tabel 1. Classification Performance of naive bayes (CNN Indonesia)

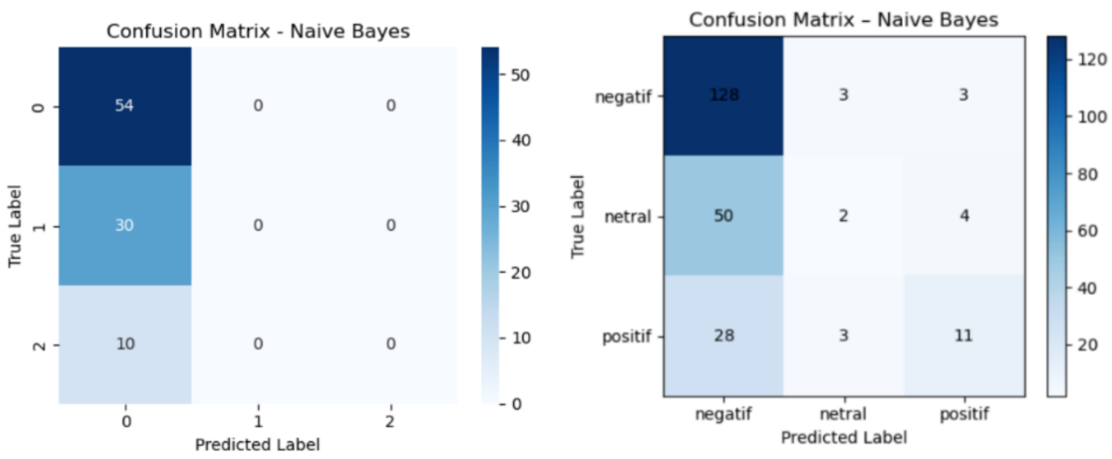
Kelas	Precision	Recall	F1-Score	Support
Negatif	0.57	1.00	0.73	54
Netral	0.00	0.00	0.00	30
Positif	0.00	0.00	0.00	10
Accuracy			0.57	94
Macro Avg	0.19	0.33	0.24	94
Weighted Avg	0.33	0.57	0.42	94

The Recall indicator shows that the data is 100% successful in capturing all existing negative data, with an accuracy level of 0.57% for the negative class. This imbalance in performance is reflected in the low macro-average F1-score (0.24), suggesting that the model does not perform equally across all sentiment classes. The results indicate the presence of class imbalance in the dataset, leading the Naive Bayes classifier to favor the dominant negative class while struggling to classify minority classes correctly. This suggests a class imbalance issue and a tendency of the model to favor the dominant negative class.

Tabel 2. Classification Performance of naive bayes (Detik)

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.62	0.96	0.75	134
Netral	0.25	0.04	0.06	56
Positif	0.61	0.26	0.37	42
Accuracy			0.61	232
Macro Avg	0.49	0.42	0.39	232
Weighted Avg	0.53	0.61	0.52	232

Table 2 shows an improvement in classification performance on the Detik dataset, achieving an overall accuracy of 61%. Unlike the CNN Indonesia dataset, the Detik data shows a more balanced sentiment distribution, enabling the Naive Bayes classifier to distinguish between negative, neutral, and positive classes more effectively. Consequently, both macro-average and weighted-average F1-scores increased compared to the CNN Indonesia dataset. With the neutral and positive values higher than the previous dataset (0.25), (0,61), it means the model is training well. With the performance as we know, shows the highest Recall remains in the negative class (0.96) performance, with a neutral bias (0.04) for F1-Score, which is also similar, where the bias occurs in the neutral class with a value weight (0.06) inversely proportional to the superior negative class and followed by the positive class as number 2. The accuracy level increases compared to the existing dataset on CNN Indonesia, primarily due to the increased data sampling.



(a) (b)

Figure 2. Confusion Matrix Naive Bayes CNN Indonesia (a) Detik (b)

Based on the results conducted at points (a) and (b), there is a problem of highly biased data imbalance, which can be seen in Figure 2. Figure 2 also explicitly explains that the trained model tends to move towards the majority model, namely the negative one. In the dataset that can be seen in point (a) the model failed to classify models that have positive and neutral labels due to an extreme bias towards one of the models, namely negative, so that even though the model accuracy can be seen in table 1 is (0.57), the recall value for the related model is still 0 for the positive and neutral class labels. This is in contrast to the confusion matrix results obtained with point (b), which are slightly better than point (a).

3.3 Evaluating The Perfomance of Support vector machines (SVM)

Table 3 shows that SVM has begun to outperform previous algorithms, such as Naive Bayes. This is evidenced by its relatively stable precision level, but the lowest positive labeling value is (0.33).

Tabel 3. Classification Performance of SVM (CNN Indonesia)

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.68	0.87	0.76	54
Netral	0.73	0.53	0.62	30
Positif	0.33	0.10	0.15	10
Accuracy			0.68	94
Macro Avg	0.58	0.50	0.51	94
Weighted Avg	0.66	0.68	0.65	94

The Support Vector Machine (SVM) algorithm successfully produced a more balanced F1-score value for each sentiment class compared to the previous algorithm. This indicates that SVM can reduce bias towards the majority class and begins to exhibit fairer performance on the neutral and positive courses. However, some noise is present in the positive label, resulting in the lowest performance. The Sentiment Analysis results performed labeling far more than expected, with (0.73) for an accuracy level and successfully began to learn a better sentiment pattern with an accuracy level of (0.68).

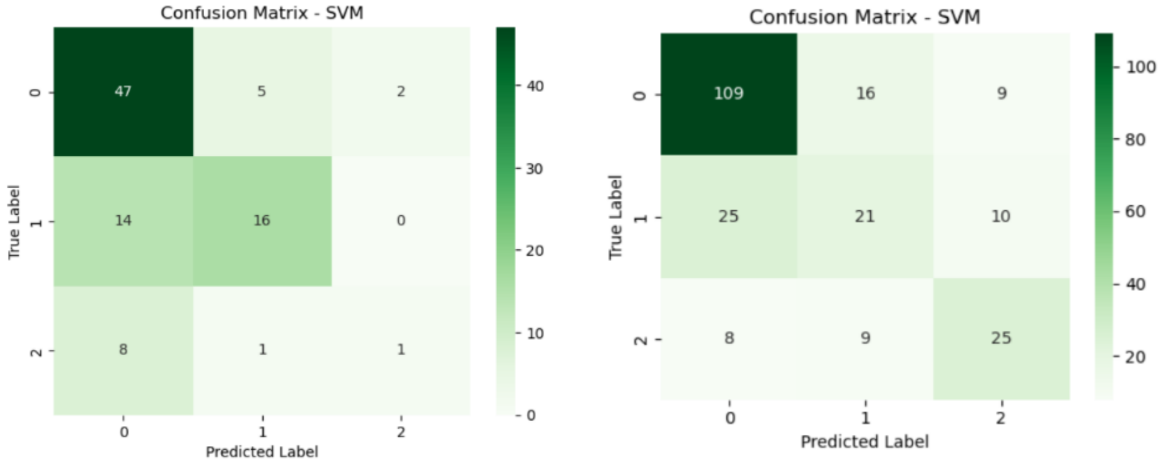
The results in the table below compare the Support Vector Machine (SVM) algorithm, which yields stable accuracy results, with those in Table 4.

Tabel 4. Classification Performance of SVM (Detik)

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.77	0.81	0.79	134
Netral	0.46	0.38	0.41	56

Kelas	Precision	Recall	F1-Score	Support
Positif	0.57	0.60	0.58	42
Accuracy			0.67	232
Macro Avg	0.60	0.59	0.59	232
Weighted Avg	0.66	0.67	0.66	232

With a high accuracy of 0.67, SVM can recognize patterns in the dataset obtained from Table 4 more reliably and provides fairly good results with an average of 0.60.



(a) (b)

Figure 3. Confusion Matrix Support Vector Machine (SVM) CNN Indonesia (a) Detik (b)

With CNN Indonesia dataset, this algorithm performed best in the negative class. This is evident from the high number of correct predictions in the negative class diagonal cells, indicating that most disaster news was correctly classified as negative. However, performance in the neutral and positive classes was still not perfomed. This pattern reflects an imbalance at tabel (3) and table (4) in the class distribution, where negative news dominates over other sentiments. In contrast to CNN Indonesia, the confusion matrix on the Detik dataset shows a more balanced distribution of predictions. The number of correct predictions in each class increased, especially in the neutral and positive classes.

3.4 Evaluating The Perfomance of k-Nearest Neighbors (KNN)

Table 5 explains that the overall accuracy of the K-Nearest Neighbors algorithm model is (0.60), where the values in this study, compared to the naive Bayes algorithm, are slightly higher. As before, the negative class value remains dominant and outperforms both the positive and negative classes.

Tabel 5. Classification Performance of KNN (CNN Indonesia)

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.61	0.93	0.74	54
Netral	0.60	0.20	0.30	30
Positif	0.00	0.00	0.00	10
Accuracy			0.60	94
Macro Avg	0.40	0.38	0.35	94
Weighted Avg	0.54	0.60	0.52	94

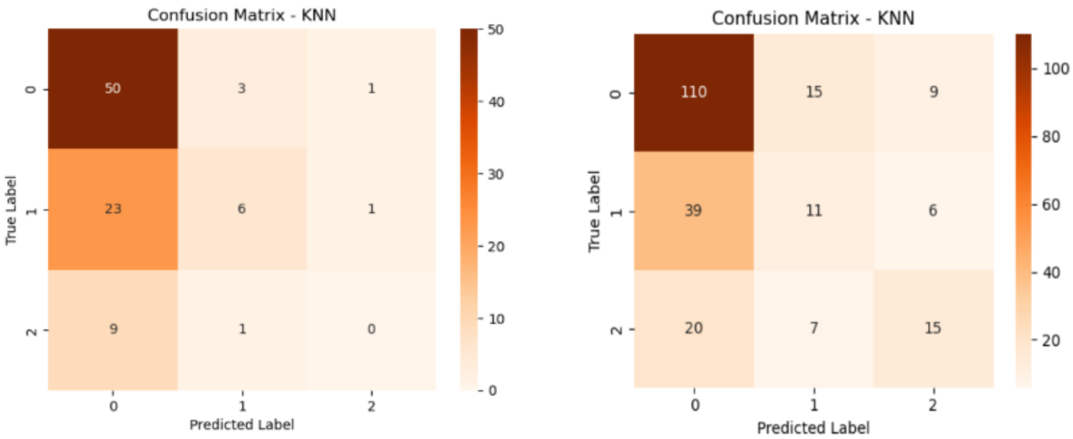
The discrepancy between the macro-average F1-score (0.35) and the weighted-average F1-score (0.52) indicates the presence of class imbalance, where the model performance is dominated by the majority negative class.

Tabel 6. Classification Performance of KNN (Detik)

Class	Precision	Recall	F1-Score	Support
Negative	0.65	0.82	0.73	134
Neutral	0.33	0.20	0.25	56

Class	Precision	Recall	F1-Score	Support
Positive	0.50	0.36	0.42	42
Accuracy			0.59	232
Macro Avg	0.49	0.46	0.46	232
Weighted Avg	0.55	0.59	0.55	232

From this table 6 it is concluded that the performance of the KNN Algorithm is much more stable in training many sample datasets compared to the results carried out in table 5. This can be proven by the large difference between the precision in the neutral and positive classes which is much more varied such as (0.33), (0.50) compared to (0.60), (0.00). However, this remains a new analysis which focuses on how to keep the data value stable. Furthermore, the existing accuracy (0.59) looks consistent with the weighted average which only has small noise.



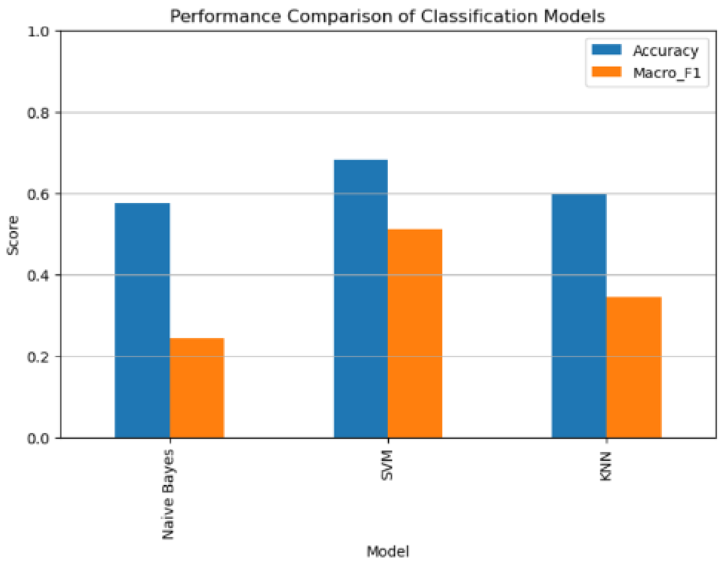
(a) (b)

Figure 4. Confusion Matrix k-Nearest Neighbors (KNN) CNN Indonesia (a) Detik (b)

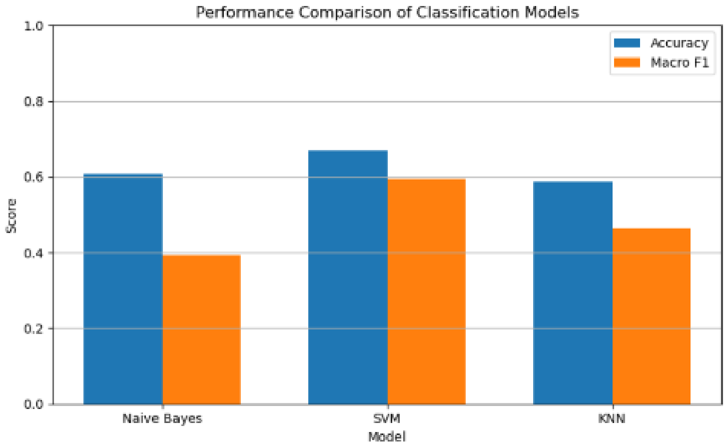
KNN can already allocate forecasts to all classes (not just guessing Negative). Although it can detect Both Positive and Neutral classes, the number of prediction errors (false negatives) is still quite large when compared to SVM.

3.5 Result of Performance

Overall, the results demonstrate that SVM consistently outperforms Naive Bayes and KNN in both datasets, achieving higher accuracy and more balanced classification performance.



(a)



(b)

Figure 5. Performance Comparisonn Algorithms CNN Indonesia (a) Detik (b)

KNN shows stable but moderate performance, while Naive Bayes exhibits sensitivity to class imbalance, particularly on the CNN Indonesia and Detik dataset.

4. CONCLUSION

In this study, it can be figured that the Support Vector Machines algorithm significantly outperforms other algorithms, such as Naive Bayes and k-Nearest Neighbors, in conducting text-based sentiment analysis, as evidenced by the stable performance of the Support Vector Machines algorithm in tokenization, training data/models, and evaluation results. This is evidenced by the results obtained in Tables 3 and 4, which contain data with an accuracy level of 67-68%. Furthermore, Naive Bayes occupies the last position in the performance test, as evidenced by Figure No. x, and position two is held by the Support Vector Machines algorithm.

ACKNOWLEDGEMENTS

The author would like to express heartfelt gratitude for the support and guidance received during the completion of this research. Special thanks go to Assoc. Prof. Muhammad Syahputra Novelan, S.Kom., M.Kom., for his invaluable supervision, insightful guidance, and constructive feedback. The author also appreciates the encouragement and assistance provided by colleagues and everyone involved in this study.

REFERENCES

[1] A. Baldah, A. V. Duarisah, and R. A. Maulana, "Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode K-Means," J. Ilm. Inform. Glob., vol. 14, no. 2, pp. 31–36, 2023, doi: 10.36982/jiig.v14i2.3186.

[2] S. J. Kim, "The role of social media news usage and platforms in civic and political engagement: Focusing on types of usage and platforms," Comput. Human Behav., vol. 138, no. September 2022, 2023, doi: 10.1016/j.chb.2022.107475.

[3] S. Kim and Y. C. Kim, "Attention to news media, emotional responses, and policy preferences about public health crisis: The case of fine dust pollution in south korea," Int. J. Environ. Res. Public Health, vol. 18, no. 24, 2021, doi: 10.3390/ijerph182413325.

[4] D. Erokhin and N. Komendantova, "Social media data for disaster risk management and research," Int. J. Disaster Risk Reduct., vol. 114, no. October, p. 104980, 2024, doi: 10.1016/j.ijdrr.2024.104980.

[5] J. Homepage, A. S. Dimas, P. Sinaga, and A. S. Aji, "Public Sentiment Analysis Towards Mayor Teddy Indra Wijaya with Logistic Regression Approach Analisis Sentimen Publik Terhadap Mayor Teddy Indra Wijaya dengan Pendekatan Logistic Regression," vol. 5, no. 1, pp. 222–231, 2025.

[6] E. Mayor, M. Miché, and R. Lieb, "Associations between emotions expressed in internet news and subsequent emotional content on twitter," Heliyon, vol. 8, no. 12, 2022, doi: 10.1016/j.heliyon.2022.e12133.

[7] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A

systematic literature review," J. King Saud Univ. - Comput. Inf. Sci., vol. 36, no. 4, p. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.
[8] C. Baden et al., "Public Opinion Negotiations in a Digital Media Ecosystem: A Conceptual Framework," Int. J. Public Opin. Res., vol. 37, no. 3, 2025, doi: 10.1093/ijpor/edaf049.
[9] B. B. Budisusetyo, W. A. Dhitya, A. Nugroho, and G. Pangestu, "Sentiment analysis methods recommendation: A review of aibased techniques on social media analysis," Procedia Comput. Sci., vol. 245, no. C, pp. 1120–1128, 2024, doi: 10.1016/j.procs.2024.10.341.
[10] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 2, pp. 375–384, 2024, doi: 10.57152/malcom.v4i2.1206.
[11] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," J. KomtekInfo, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
[12] M. Syahputra Novelan and M. Iqbal, "Comparative Analysis of the Performance of K-Nearest Neighbor (K-NN) and Naive Bayes Algorithms on User Satisfaction Levels of the Tokopedia Application," 2025.
[13] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," Inf. Process. Manag., vol. 58, no. 5, p. 102656, 2021, doi: 10.1016/j.ipm.2021.102656.
[14] Y. A. Hafiz and E. Sudarmilah, "Implementasi Web Scraping Pada Portal Berita Online Pendahuluan Informasi semakin pesat dan penting seiring dengan semakin canggihnya era digital (Fathir et al ., 2023). Portal berita online adalah salah satu sumber informasi yang dapat memberikan manfa," J. Ilmu Komput., vol. 12, no. 1, pp. 55–60, 2023.
[15] N. Ukey, Z. Yang, B. Li, G. Zhang, Y. Hu, and W. Zhang, "Survey on Exact kNN Queries over High-Dimensional Data Space," Sensors, vol. 23, no. 2, pp. 1–44, 2023, doi: 10.3390/s23020629.
[16] J. Y. Guyt, H. Datta, and J. Boegershausen, "Unlocking the Potential of Web Data for Retailing Research," J. Retail., vol. 100, no. 1, pp. 130–147, 2024, doi: 10.1016/j.jretai.2024.02.002.
[17] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," Expert Syst. Appl., vol. 223, no. March, 2023, doi: 10.1016/j.eswa.2023.119862.
[18] A. Rafdi, H. Mawengkang, and S. Efendi, "Sentiment Analysis Using Naive Bayes Algorithm with Feature Selection Particle Swarm Optimization (PSO) and Genetic Algorithm," Int. J. Adv. Data Inf. Syst., vol. 2, no. 2, 2021, doi: 10.25008/ijadis.v2i2.1224.
[19] A. Widodo, B. A. Herlambang, and R. Renaldy, "Optimizing Support Vector Machine (SVM) for Sentiment Analysis of Blu by BCA Reviews with Chi-Square," J. Appl. Informatics Comput., vol. 9, no. 5, pp. 2588–2597, 2025, doi: 10.30871/jaic.v9i5.10541.
[20] M. Yusuf et al., "Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia," vol. 2, pp. 11–16, 2021.
[21] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," J. Informatics Web Eng., vol. 3, no. 1, pp. 67–75, 2024, doi: 10.33093/jiwe.2024.3.1.5.
[22] R. Setiabudi, N. M. S. Iswari, and A. Rusli, "Enhancing text classification performance by preprocessing misspelled words in Indonesian language," Telkomnika (Telecommunication Comput. Electron. Control., vol. 19, no. 4, pp. 1234–1241, 2021, doi: 10.12928/TELKOMNIKA.v19i4.20369.