Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3318~3324
e-ISSN: 2548-6861

The Application of Naïve Bayes Algorithm in Detecting Hoaxes on National News Portals in Indonesia

Cut Rifa Salsabil [1]*, Nurdin [2], Rizki Suwanda [3]***
* Departement of Informatics, Universitas Malikussaleh, Lhokseumawe, Indonesia
** Departement of Information Technology, Universitas Malikussaleh, Lhokseumawe, Indonesia
cut.220170123@mhs.unimal.ac.id [1], nurdin@unimal.ac.id [2], rizkisuwanda@unimal.ac.id [3]

Article Info

Article history:

Received 2026-06-11
Revised 2026-07-24
Accepted 2026-08-08

Keyword:

Disinformation,
Hoax Detection,
Multinomial Naïve Bayes,
Natural Language Processing,
TF-IDF.

ABSTRACT

The rapid advancement of information technology in Indonesia has led to a massive spread of digital disinformation, commonly known as an infodemic. The inability to filter inaccurate information manually necessitates a reliable, automated hoax detection system. This study aims to implement and evaluate the Multinomial Naïve Bayes algorithm combined with Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction to classify news articles as either factual or hoax. The research utilizes a dataset of 2,910 Indonesian news articles published in 2025, collected from verified national news portals and fact-checking websites. The text data underwent comprehensive preprocessing—including case folding, cleansing, stopword removal, and stemming—before being evaluated using 5-Fold Cross-Validation and an 80:20 data split. Experimental results demonstrate that the Naïve Bayes model achieves highly stable and competitive performance, recording an accuracy of 93.81%, a precision of 93.84%, a recall of 93.81%, an F1-Score of 93.82%, and a 5-Fold Cross-Validation F1-Score of 93.39%. Notably, the algorithm exhibited a significantly low False Negative rate, missing only 15 hoax documents out of 582 test samples. Furthermore, the trained model was successfully integrated into a real-time, web-based user interface using Streamlit. This practical implementation provides an accessible and efficient initial screening tool for the general public and journalists to assist in verifying news authenticity, thereby supporting efforts to mitigate the impact of digital hoaxes.

I. INTRODUCTION

The development of information technology in Indonesia has transformed online news portals into the main pillar of meeting people's information needs. However, this ease is followed by the phenomenon of "infodemic", where inaccurate information spreads faster than the verification of facts. In the context of this study, "hoaxes" encompass various forms of digital disinformation, including fully fabricated fake news, misleading content, satire, clickbait, and manipulated contexts. Based on data from a report from the Ministry of Communication and Information Technology (Kominfo), throughout 2024 alone, thousands of hoax issues have been identified spread across various digital platforms, dominated by political, health, and fraud issues. This is in line with the findings of researchers who stated that hoax communication interactions have become a serious threat to social welfare and the integrity of the digital information ecosystem in Indonesia [1][2][3].

The main problem faced today is the low ability to filter information at the end-user level. If left without a reliable automatic detection system, this hoax will trigger social polarization and distrust in public institutions. Legally, the Government of Indonesia has given serious attention through Law No. 19 of 2016 concerning Amendments to Law No. 11 of 2008 concerning ITE, especially Article 28 paragraph (1) which prohibits the spread of fake news that is detrimental to consumers. Although regulations are in place, law enforcement implementation is often reactive (after the

incident), so a preventive technology approach that is able to carry out automatic screening early is needed [4] [5] [6].

Theoretically, the disciplines of Natural Language Processing (NLP) and Machine Learning offer adaptive computational solutions through binary text classification methods. The implementation of NLP in systematically managing large text documents has been shown to be very effective in facilitating the grouping of complex information as well as improving the semantic understanding of a news text [7][8]. While recent advancements often utilize algorithms like Support Vector Machine (SVM), Logistic Regression, Random Forest, or Transformer-based models (such as IndoBERT) to capture deep semantic contexts, the Multinomial Naïve Bayes algorithm is selected as the primary focus of this study. Naive Bayes is a probability-based classification algorithm that works by calculating the likelihood that an article is included in the Hoax category or valid based on the distribution of words in it. The main advantages of Naive Bayes lie in its high computational efficiency, ease of implementation, and ability to establish a robust baseline with relatively fast training times. The relevance of this algorithm in handling the classification of textual data in Indonesia is still very high, with stable metric performance [8][12][13].

Before the Naive Bayes algorithm can mathematically process news text, the document requires a feature extraction stage using the Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF works by assigning numerical weight to each word based on the frequency with which it appears in the document (TF) and its uniqueness throughout the corpus (IDF). It must be acknowledged, however, that TF-IDF is limited to word frequencies and cannot comprehend deep semantic contexts, inter-sentence relationships, or external factual knowledge. Consequently, the model primarily learns the differences in writing styles and linguistic patterns between standard news portals and fact-checking sites, acting as a probabilistic indicator of disinformation rather than a system capable of verifying absolute factual truth.

On a practical level, the effectiveness of fake news detection systems using the integration of probabilistic algorithms has been validated by various empirical studies. Research from Rahmadani et al. [15] proves that the Naive Bayes algorithm, when combined with the TF-IDF feature extraction method, is able to achieve a very high level of accuracy, which is 94.12%. The ability of Naive Bayes to handle the characteristics of local news in Indonesia is also strengthened by Rahutomo et al. [16] which shows that this algorithm has a very reliable performance for the characteristics of Indonesian datasets. The success of this classification is also greatly influenced by the accuracy of the application of pre-processing data scenarios, including the use of the text stemming method to reduce Indonesian suffix words to produce a cleaner and more precise representation of features [9][17][18][19].

Although the reliability of this model has been recognized, performance inconsistencies in some global tests indicate that the effectiveness of detection systems is highly dependent on data characteristics, preprocessing variations, and information retrieval time periods. The dynamic characteristics of Hoaxes trigger the need for comprehensive retesting using cutting-edge data, in order to find the most precise single detection solution amid the evolution of recent Hoax patterns [20][21]. Based on this urgency, this study aims to describe the effectiveness of the text data preprocessing stage (Case Folding, Cleaning, Tokenizing, Stopword Removal, Stemming) and TF-IDF feature extraction, as well as transparently measure the level of accuracy, precision, and recall generated by the Multinomial Naive Bayes algorithm in classifying Hoaxes on Indonesian national news portals.

In order to maintain the focus and objectivity of the research, the scope of the analysis is limited to Indonesian texts (article content) published throughout the period from January 1, 2025 to December 2025. Hoax data (label 0) is collected manually from TurnbackHoax.id fact-checking sites, while valid data (label 1) is taken from news articles published by verified national news portals such as Detik.com and Tempo.co. The model implementation was built using the Python programming language with the main libraries Scikit-learn, Pandas, and NLTK/Sastrawi. Through these restrictions, the research entitled "The Application of the Naive Bayes Algorithm in Detecting Hoaxes on National News Portals in Indonesia" is expected to make an empirical contribution and serve as a practical initial screening tool for the public and journalists to verify news authenticity, thereby supporting broader efforts to mitigate the impact of digital hoaxes.

II. METHOD

The research procedure begins with the collection of a national news dataset published throughout 2025, comprising a total of 2,910 documents. To ensure a balanced distribution and prevent majority class bias during the training phase, the dataset consists of 1,510 documents labeled as Hoax (Class 0) and 1,400 documents labeled as Factual (Class 1). The labeling process relies entirely on official validations: hoax data is manually collected from the *TurnbackHoax.id* fact-checking site, while factual data is sourced from verified national news portals such as *Detik.com* and *Antara News*. To maintain the objectivity of the references, no additional manual labeling intervention is applied.

To prevent data leakage—especially considering that multiple articles might cover the same events or topics—the data is split using a stratified random sampling method. This approach ensures that the distribution of hoax and factual classes remains proportionally consistent when partitioned into 80% training data (2,328 documents) to build the model, and 20% testing data (582 documents) for evaluation. Furthermore, the model is evaluated using a 5-Fold Stratified Cross-Validation method to ensure robust generalization stability [4][5][6]. Finally, a performance evaluation is

conducted on the implemented model architect ure using a confusion matrix framework.

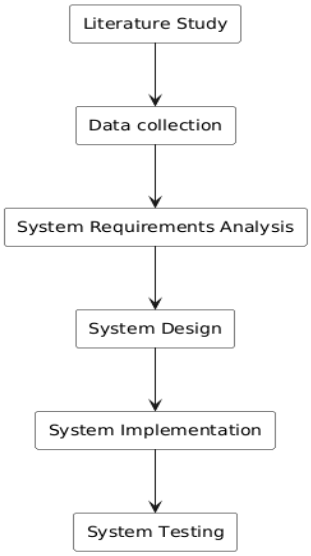

Figure 1. Research Stages

A. Naïve Bayes Scheme

This system diagram illustrates the workflow from textual data input of national news, text preprocessing steps, feature extraction using TF-IDF, to displaying the predicted final class of the news article.

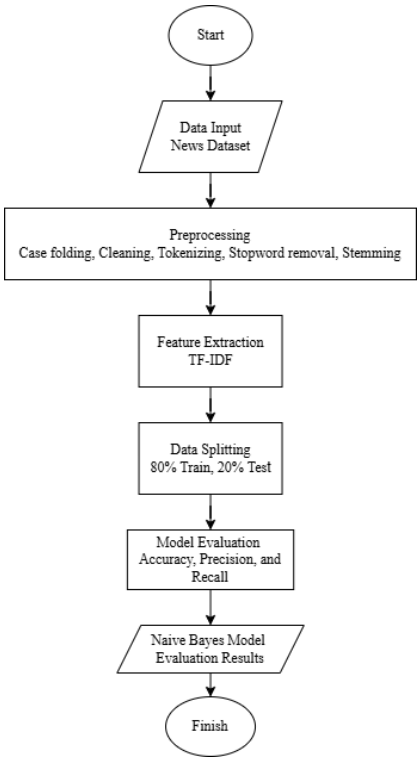

Figure 2. System Scheme

B. Naïve Bayes Classifier Algorithm

Natural Language Processing (NLP) is a field of computer science and artificial intelligence used for processing and managing human language tasks systematically, especially when handling a large volume of unstructured textual documents [7][10]. To allow text data to be processed efficiently by machine learning algorithms, the raw text must go through text preprocessing stages to eliminate noise and standardize the data structure [7][15]. Based on established methodologies, the text preprocessing consists of several key stages:

1. Case Folding: Converts all uppercase letters into lowercase and removes special characters, numbers, or punctuation marks that lack contextual meaning [9].
2. Tokenizing: Partitions text paragraphs into smaller units called tokens or individual words to analyze individual word frequencies [18].
3. Stopword Removal: Filters out common words that appear frequently but do not carry significant informative value for classification, such as conjunctions and prepositions [17].
4. Stemming: Reduces inflected words to their root form by stripping prefixes, suffixes, and confixes, which is highly vital for handling the complex morphology of the Indonesian language [22].

Optimizing these text cleaning phases has been empirically proven to stabilize model generalization and significantly increase the final accuracy of textual classification [8][23]. Furthermore, machine learning classifiers cannot process raw string text directly, meaning text must be transformed into numerical vectors [14]. Term Frequency-Inverse Document Frequency (TF-IDF) is an optimal feature extraction method that assigns weights based on word importance across the corpus [14][15]. In this study, the TF-IDF configuration is set to retain a maximum of 8,000 vocabularies at the unigram level. This specific feature selection strategy successfully isolates informative keywords from highly provocative, emotional, and non-linear hoax structures while discarding uninformative linguistic noise [15].

The Naïve Bayes Classifier is a supervised machine learning algorithm based on Bayes' Theorem that computes probability scores to categorize documents into target classes [11][13]. It is widely used as a benchmark text classification standard due to its exceptional computational efficiency and ability to handle high-dimensional text feature spaces rapidly [9][10][11]. A key structural assumption of Naïve Bayes is feature independence, where each word is treated independently given the class label [11]. The core mathematics of the posterior probability prediction is formulated as follows [9][11]:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Where C represents the target class (Hoax or Valid) and X denotes the document feature vector. In practical applications, the efficiency of this probabilistic framework has been widely validated across diverse datasets [1][12][19][20][24][25]. Rather than computing complex geometric boundaries, Naïve

Bayes selects the class maximizing the posterior probability, yielding highly robust generalization even when managing dynamically evolving news topics and social media trends [2][3][21].

The performance evaluation of the Multinomial Naïve Bayes model aims to assess how well the system detects and classifies hoaxes within national news articles. Four key metrics are used to measure the system's performance: Recall, Precision, F1 Score, and Accuracy [4][11][12].

a. Recall evaluates the system's ability to retrieve all relevant true positive instances from the given dataset.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

b. Precision evaluates the accuracy of the system in providing relevant, true positive responses.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

c. The F1 Score is a metric that combines precision and recall into a single value, calculated as the harmonic mean of the two.

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} \quad (4)$$

d. Accuracy is a metric used to evaluate how often the system makes correct predictions overall, both positive and negative, compared to the actual values.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

TABLE I
CONFUSION MATRIX

Confusion Matrix		Actual	
		TRUE	FALSE
Predicted	TRUE	TP (True Positif)	FP(False Positif)
	FALSE	FN (False Negatif)	TN (True Negatif)

This table provides a framework for evaluating the performance of a detection model using metrics such as precision, recall, f1 score, and accuracy, based on the combination of actual and predicted values.

III. RESULTS AND DISCUSSION

Detecting the spread of hoax news in real-time is a complex challenge influenced by various factors, such as variations in language patterns, provocative narrative structures, and the large volume of daily information on digital news portals. This detection process requires a system that is able to recognize the difference between valid information and disinformation with a high degree of accuracy. The application of the Multinomial Naïve Bayes algorithm is particularly relevant because of its ability to process and classify large-dimensional textual data quickly and efficiently. In this study, the model acts as a probabilistic indicator to detect language patterns associated with hoaxes, rather than an absolute factual verifier.

A. Data Description

The data used to train the Naïve Bayes model was obtained from articles published by various national news portals (such as Detik.com and Antara News) as well as fact-checking sites (Turnbackhoax.id) throughout 2025. The dataset consists of 2,910 documents divided into two categories: 1,400 documents labeled Fact (Valid) and 1,510 documents labeled Hoax.

TABLE II
SAMPLE DATASET

No.	News Title Excerpt	Class
1	Polisi Bakal Gelar Olah TKP Kebakaran Pasar Taman Puring	Fact
2	KemenPAN-RB & Kadin Kerja Sama Mudahkan Usaha Lewat...	Fact
3	3 PMI Ilegal Terjebak di Kamboja, Kena Pecat Usai Perang...	Fact
4	Sejumlah Pesawat Melintasi Langit Iran di Tengah Peperangan	Fact
5	Arahan Prabowo ke Guru Sekolah Rakyat: Buat Siswa Gembira	Fact
6	[SALAH] Pesan Berantai Sakit Difteri karena Kencing Tikus	Hoax
7	[SALAH] Presiden Prabowo Temui Pendemo dan Meminta Maaf	Hoax
8	[SALAH] Anies Terima Penghargaan dari Universitas Wageningen	Hoax
9	[PENIPUAN] Pembukaan Pendaftaran PPG Prajabatan Tahun 2025	Hoax
10	[SALAH] Video Banjir Setinggi 4 Meter Menenggelamkan...	Hoax

This dataset includes various news texts, highlighting differences in diction, sentiment, and sentence structure. Some texts show a formal and neutral news structure (Facts), while other texts use provocative and hyperbolic language (Hoaxes). Before being processed by the model, all text in the dataset underwent a comprehensive text preprocessing stage and was extracted into numerical vectors using TF-IDF weighting to improve the model's generalization capabilities.

B. Naïve Bayes Model Evaluation Results

This section presents the results of the final evaluation process for the Multinomial Naïve Bayes model. The model was tested on a validation dataset covering 20% of the total dataset (582 news documents) and further validated using a 5-Fold Stratified Cross-Validation method to evaluate its robustness against data variance.

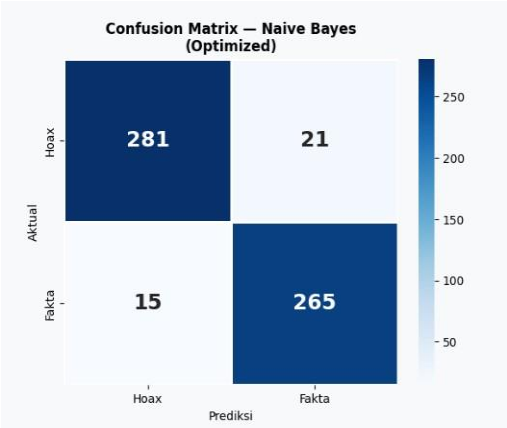

Figure 3. Confusion Matrix Naive Bayes

The Confusion Matrix illustrates how well the model classifies documents into predefined categories. Out of 582 test documents, the model successfully classified 265 documents as True Positive (Hoax detected Hoax) and 281 documents as True Negative (Fact detected Fact). Misclassifications were suppressed to a very low number: 21 False Positives and 15 False Negatives.

TABLE III
MODEL EVALUATION RESULTS ON TESTING DATA

Class / Metrics	Overall
Number of Test Data	582
Precision (P)	93,84%
Recall (R)	93,81%
F1-Score	93,82%
Accuracy	93,81%

Based on Table III, the model achieved an overall prediction accuracy of 93.81%. The stability of this performance is confirmed by the 5-Fold Cross-Validation F1-Score, which remains highly competitive at 93.39%. The most crucial advantage of the Naïve Bayes model lies in the low number of detection leaks (False Negatives). In the context of disinformation filtering, passing hoax news and treating it as a valid fact is a highly dangerous scenario. The strong Recall value (93.81%) and the low False Negative count prove that this probability-based model acts resiliently.

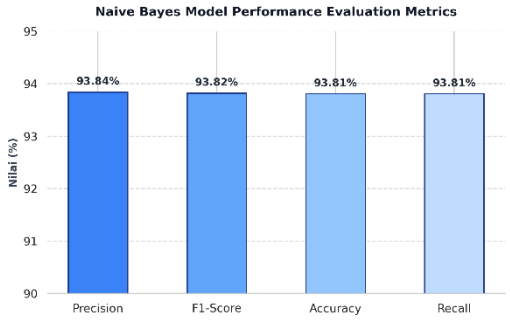

Figure 4. Naïve Bayes Model Evaluation Results Graph

The graph in Figure 6 visualizes the model's performance based on key evaluation metrics. The balance of harmonic values on the F1-Score graph (93.82%) confirms the stability of the model in handling news-grade distribution. The most crucial advantage of the Naïve Bayes model lies in the low number of detection leaks (False Negative). In the context of disinformation filtering, passing hoax news and treating it as valid facts is the most dangerous scenario. The high value of Recall and the lack of False Negative prove that this probability-based model acts very resilient and efficient to be implemented as a daily information protection system.

C. Error Analysis and System Limitations

Despite the high accuracy, the occurrence of 21 False Positives (factual news predicted as hoaxes) and 15 False Negatives (hoaxes predicted as factual) necessitates an error analysis. The primary limitation of the TF-IDF and Naïve Bayes combination lies in its inability to comprehend deep semantic contexts, inter-sentence relationships, or external factual knowledge. TF-IDF solely relies on the lexical frequency of words.

Consequently, the model demonstrated a thematic bias when processing named entities, particularly technology and social media platforms (e.g., WhatsApp, Facebook, Meta, Instagram). Because these terms appear predominantly in hoax narratives within the training data, the model heavily weighted them toward the Hoax class. When valid national news articles objectively reported on these platforms, the system occasionally misclassified them as hoaxes (False Positives). This underscores that the algorithmic model primarily learns the differences in writing styles rather than verifying the actual truth of the content.

D. System Implementation

In this section, the implementation of a real-time hoax detection system using the trained Naïve Bayes model is discussed. These systems allow users, such as the general public or journalists, to manually enter news text or provide a link (URL) of an article for analysis. Once the text or URL is entered, the system performs web scraping (if using a link), processes the text, and detects it using the Naïve Bayes model. The detection results are clearly displayed in the form of a class label (Hoax or Fact) along with the probability or confidence level of the model (confidence score).

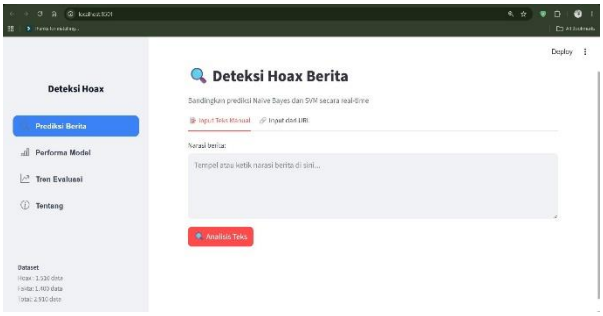

Figure 5. Home Page View (News Prediction)

The main page of this detection system is designed (using the Streamlit library in Python) to be simple and informative. Users can quickly perform text analysis and view the results side by side. Information about news classes is displayed in the form of text and prominent indicator colors on the screen, making it easy for users to judge the veracity of information without any special technical expertise. The implementation of this system aims to facilitate more efficient hoax mitigation.

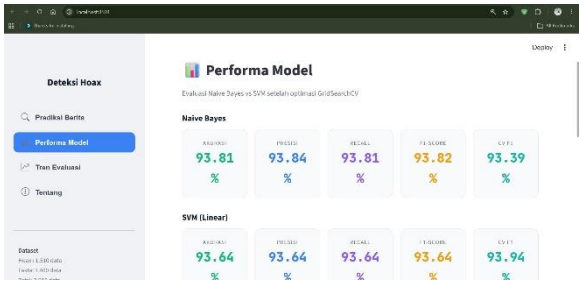

Figure 6. Model Performance Page View

The model performance page (Dashboard) provides a concise and clear overview of the status and reliability of the system. This page displays important information such as statistical metric values (Accuracy, Precision, Recall, F1-Score), metric visualization bar charts, and confusion matrix interactivity. Overall, this page helps users or developers to monitor the real-time performance of the system transparently.

IV. CONCLUSION

This study has successfully implemented and evaluated the Multinomial Naïve Bayes algorithm combined with Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction for hoax detection on Indonesian national news portals. Utilizing a balanced dataset of 2,910 news articles from 2025, the experimental results demonstrate highly competitive and stable performance. The evaluation yielded an accuracy of 93.81%, a precision of 93.84%, a recall of 93.81%, an F1-Score of 93.82%, and a 5-Fold Stratified Cross-Validation F1-Score of 93.39%.

A crucial finding in this research is the algorithm's exceptional ability to minimize False Negatives, recording only 15 missed hoax documents out of 582 test samples. In the domain of information filtering, a low false-negative rate is highly significant, as it ensures that the system rarely validates dangerous disinformation as factual news. However, based on the error analysis, it must be acknowledged that this TF-IDF and Naïve Bayes approach acts primarily as a probabilistic detector of disinformation writing styles and language patterns, rather than an absolute verifier of factual truth. Therefore, the conclusion that this model supports national information security must be interpreted proportionally; it serves as a robust early warning screening tool rather than a definitive solution to the infodemic.

Furthermore, the successful integration of this classification model into a real-time, web-based user interface using Streamlit provides a practical software engineering contribution. It serves as an accessible initial screening tool for the general public and journalists to assist in the daily news verification process. For future development, it is highly recommended to explore the integration of deep learning architectures and transformer models, such as IndoBERT, to overcome the limitations of TF-IDF in capturing deeper semantic contexts and inter-sentence relationships in complex hoax narratives. Additionally, evaluating the model's capabilities on an external validation dataset is necessary to ensure that the achieved performance remains consistent and robust across diverse timeframes and unstructured multi-platform sources.

BIBLIOGRAPHY

[1] Muhammad Salim Albana, Alif Dava Mahesa, Indriani Putri, and Noerma Kurnia Fajarwati, "Interaksi Komunikasi Hoax Di Media Sosial Serta Antisipasinya," *SABER J. Tek. Inform. Sains dan Ilmu Komun.*, vol. 2, no. 2, pp. 34–39, 2024, doi: 10.59841/saber.v2i2.958.

[2] M. Solikhah, "the Effect of Machine Learning Algorithms on Hoax Detection on Social Media: Implications for National Information Security," *J. Artif. Intell. Res.*, vol. 1, no. 1, pp. 11–20, 2025, doi: 10.64910/jouair.v1i1.7.

[3] A. Sandu, L. A. Cotfas, C. Delcea, C. Ioanăș, M. S. Florescu, and M. Orzan, "Machine Learning and Deep Learning Applications in Disinformation Detection: A Bibliometric Assessment," *Electron.*, vol. 13, no. 22, 2024, doi: 10.3390/electronics13224352.

[4] P. S. More, A. Jadhav, and S. Yadav, "Detection of Fake News using Machine Learning," *Ijarcce*, vol. 10, no. 4, pp. 485–490, 2021, doi: 10.17148/ijarcce.2021.10484.

[5] S. Pandey, S. Prabhakaran, N. V. S. Reddy, and D. Acharya, "Fake News Detection from Online media using Machine learning Classifiers," *J. Phys. Conf. Ser.*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012027.

[6] H. F. Villela, F. Corrêa, J. S. de A. N. Ribeiro, A. Rabelo, and D. B. F. Carvalho, "Fake news detection: a systematic literature review of machine learning algorithms and datasets," *J. Interact. Syst.*, vol. 14, no. 1, pp. 47–58, 2023, doi: 10.5753/jis.2023.3020.

[7] A. Badawi, "the Effectiveness of Natural Language Processing (Nlp) As a Processing Solution and Semantic Improvement," *Int. J. Econ. Technol. Soc. Sci.*, vol. 2, no. 1, pp. 36–44, 2021, doi: 10.53695/injects.v2i1.194.

[8] A. Khaidar, S. Muliana, and P. Studi Magister Teknologi Informasi, "Sentiment Analysis of Instagram Comments on the BPS Province X Account Using the Naive Bayes Algorithm Based on Machine Learning," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 3, pp. 1231–1237, 2025, doi: 10.30811/jaise.v5i3.7815.

[9] E. Y. Hidayat and M. A. Rizqi, "Klasifikasi Dokumen Berita Menggunakan Algoritma Enhanced Confix Stripping Stemmer dan Naïve Bayes Classifier," *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 90–99, 2020, doi: 10.25077/teknosi.v6i2.2020.90-99.

[10] M. Suhendri and Y. Afrilia, "SISTEMASI: Jurnal Sistem Informasi Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (Nbc)," vol. 10, pp. 268–279, 2021, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[11] O. Peretz, M. Koren, and O. Koren, "Naive Bayes classifier – An ensemble procedure for recall and precision enrichment," *Eng. Appl. Artif. Intell.*, vol. 136, no. PB, p. 108972, 2024, doi: 10.1016/j.engappai.2024.108972.

[12] A. F. Watratan, A. P. B, and D. Moeis, "Implementation of the Naive Bayes Algorithm to Predict the Spread of Covid-19 in Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, 2020, doi: https://doi.org/10.52158/jacost.v1i1.9.

[13] S. Nasyira and L. Rosnita, "Comparison of K-Nearest Neighbors Method and Naïve Bayes Method in Classifying the Quality of Oil Palm Seed Varieties," vol. 10, no. 2, pp. 413–425, 2025, doi:

10.31572/inotera.Vol10.Iss2.2025.ID539.

[14] H. D. Abubakar and M. Umar, "Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec," *SLU J. Sci. Technol.*, vol. 4, no. 1&2, pp. 27–33, 2022, doi: 10.56471/slujst.v4i.266.

[15] R. Rahmadani, A. Rahim, and R. Rudiman, "Analisis Sentimen Ulasan 'Ojol the Game' Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.

[16] F. Rahutomo, I. Y. R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia," *J. Penelit. Komun. Dan Opini Publik*, vol. 23, no. 1, 2019, doi: 10.33299/jpkop.23.1.1805.

[17] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J. Big Data*, vol. 8, no. 1, pp. 1–16, 2021, doi: 10.1186/s40537-021-00413-1.

[18] N. V. Pusean, N. Charibaldi, and B. Santosa, "Comparison of Scenario Pre-processing Performance on Support Vector Machine and Naïve Bayes Algorithms for Sentiment Analysis," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 8, no. 1, pp. 57–63, 2023, doi: 10.25139/inform.v8i1.5667.

[19] M. N. Raza, "Sistem Deteksi Berita Hoax Menggunakan Algoritma Naïve Bayes Dan Random Forest Pada Machine Learning," *Pondasi J. Appl. Sci. Eng.*, vol. 1, no. 2, pp. 43–57, 2024, [Online]. Available: https://journal.alshobar.or.id/index.php/pondasi/article/view/221

[20] Febriyanty Nur Elyta, "Deteksi Berita Hoax Dari Media Online Indonesia Menggunakan Algoritma Naive Bayes dan Support Vector Machine," pp. 1–128, 2023.

[21] J. S. Wibowo, E. N. Wahyudi, and H. Listiyono, "Performance Comparison of SVM, Naive Bayes, and Random Forest Models in Fake News Classification," *Eng. Technol. J.*, vol. 09, no. 08, pp. 4799–4804, 2024, doi: 10.47191/etj/v9i08.27.

[22] A. P. Wibawa, F. A. Dwiyanto, I. A. E. Zaeni, R. K. Nurrohman, and A. Afandi, "Stemming javanese affix words using nazief and adriani modifications," *J. Inform.*, vol. 14, no. 1, p. 36, 2020, doi: 10.26555/jifo.v14i1.a17106.

[23] A. Alaiya, N. Nurdin, and C. Agusniar, "Sentiment Analysis of E-Commerce Product Reviews on Tokopedia Using Support Vector Machine," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2869–2878, 2025, doi: 10.30871/jaic.v9i5.10977.

[24] F. Nurwanda and J. R. Rizkiani, "Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine pada Analisis Sentimen Twitter Topik Lifestyle," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 21, pp. 314–323, 2023, [Online]. Available: https://doi.org/10.5281/zenodo.10077023

[25] I. A. Ropikoh, R. Abdulhakim, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19," *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 64–73, 2021, doi: 10.30871/jaic.v5i1.3167.