International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

Online News Hoax Detection Using Machine Learning Classification Algorithms

Chandra Kesuma *

Department of Computer Technology, Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, Purwokerto, Indonesia

Authors Scopus ID URL :
Chandra Kesuma[1*], https://www.scopus.com/authid/detail.uri?authorId=58289249700

Email address:
chandra.cka@bsi.ac.id

*Corresponding author : chandra.cka@bsi.ac.id

Received: November 15, 2025; **Accepted**: January 30,2026 ; **Published**: January 30,2026

Abstract: The rapid growth of digital media usage has significantly increased the spread of hoax news. Such information can lead to misinformation, social anxiety, and public misunderstanding. This study proposes an automatic detection approach for Indonesian-language hoax news using machine learning-based classification algorithms. A dataset consisting of 3,000 Indonesian news articles collected from social media platforms and online news portals was employed and validated using a fact-checking website (TurnBackHoax.id). The proposed method involves text preprocessing, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), and classification using Naive Bayes and Support Vector Machine (SVM) algorithms. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results indicate that the SVM algorithm achieves better performance than Naive Bayes in detecting hoax news. The findings demonstrate that machine learning-based classification can provide an effective solution for automatic hoax detection and can be further developed for practical implementation.

Keywords: Hoax detection, Text classification, Machine learning, TF-IDF, SVM, Naive Bayes.

1. Introduction

The rapid development of information and communication technology has transformed the way people access, produce, and disseminate information. The Internet and social media enable individuals to become both producers and consumers of information within a very short time.

This condition has a positive impact by accelerating information distribution, increasing transparency, and expanding access to various sources of knowledge. However, on the other hand, this convenience also gives rise to serious problems in the form of the widespread dissemination of false information or hoax news. Hoax news is defined as information that does not correspond to factual reality, whether created intentionally or unintentionally, and disseminated for certain purposes, such as influencing public opinion, causing panic, dividing society, or obtaining personal gain. In Indonesia, hoaxes are often related to political, health, disaster, and socio-cultural issues. The impact of hoax dissemination cannot be underestimated, as it can lead to poor decision-making, increased social conflict, decreased trust in official institutions, and the weakening of democratic quality.

The dissemination of hoax news in the form of misinformation exhibits distinctive linguistic characteristics and tends to spread faster than factual news, thereby potentially disrupting public understanding of important issues [1]. Although various fact-checking platforms such as TurnBackHoax.id are available, manual news verification processes still have limitations. The enormous volume of news, the rapid speed of information dissemination, and limited human resources make the clarification process unable to keep pace with the spread of hoaxes. Therefore, an automated solution is required to assist in performing initial filtering of news that potentially contains hoax content.

Machine learning approaches such as Naive Bayes and SVM based on TF-IDF representations have consistently become primary choices in the hoax news detection literature due to their ability to learn important word patterns in text [2]. By employing text classification techniques, a system can learn specific patterns from hoax and non-hoax news data and subsequently use them to automatically predict the class of new news articles. Various previous studies have shown that classification algorithms such as Naive Bayes, Support Vector Machine (SVM), Random Forest, as well as deep learning–based models achieve satisfactory performance in detecting hoax news.

A study reported that the use of Naive Bayes with tokenizing, stopword removal, and TF-IDF stages can identify hoaxes with competitive accuracy, indicating its relevance for text classification tasks [3]. Several studies in Indonesia have reported that Naive Bayes-based hoax news classification can achieve accuracy levels of around 80% on small- to medium-scale datasets. On the other hand, Support Vector Machine (SVM) is often reported to provide more stable and accurate performance, particularly on high-dimensional data such as TF-IDF representations. Nevertheless, differences in dataset characteristics, preprocessing techniques, and evaluation schemes result in varying outcomes across studies.

Based on these conditions, further research is still required to directly compare the performance of Naive Bayes and Support Vector Machine (SVM) algorithms on a larger Indonesian-language hoax news dataset with a clear validation process. In addition, many previous studies have focused mainly on accuracy, whereas other metrics such as precision, recall, and F1-score are also crucial, especially in the context of hoax detection, where misclassifying hoax news as valid news can lead to serious consequences.

DOI: 10.52362/ijiems.v5i1.2287

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

This study aims to develop and evaluate Indonesian-language hoax news detection models using machine learning approaches with Naive Bayes and Support Vector Machine (SVM) algorithms. The dataset consists of 3,000 news articles collected from social media platforms and online news portals and validated through a fact-checking website. Feature representation is performed using the TF-IDF method, while model evaluation employs accuracy, precision, recall, and F1-score metrics. The main contributions of this study are: (1) providing a comparative performance analysis of two popular algorithms on a medium-scale dataset, (2) presenting comprehensive evaluation results using multiple metrics, and (3) offering insights into the potential development of automatic hoax detection systems that can be integrated into digital platforms. It is expected that the results of this study can serve as a reference for system developers and future researchers in developing more accurate and reliable hoax detection solutions.

2. Literature Review

Hoax Detection and Research Trends

Hoax news detection is an important research area in natural language processing (NLP) and machine learning due to its impact on public opinion and social stability. Early studies largely emphasized the use of text mining techniques to identify hoaxes through linguistic features and word patterns commonly found in fake news. One study in Indonesia employed the Naive Bayes algorithm with TF-IDF weighting and achieved an accuracy of 82% on a 600-article Indonesian-language dataset, demonstrating the effectiveness of a simple yet consistent approach for hoax text processing [4]. In addition, research applying Naive Bayes to Indonesian-language tweets using TF-IDF also indicated that this method has potential for hoax detection, although it must be combined with thorough preprocessing steps to obtain optimal results [5].

Beyond the Naive Bayes approach, other algorithms such as SVM have been widely used because of their capability to handle high-dimensional data, which is common in TF-IDF text representations. A hoax detection study on Platform X using SVM reported that the model was able to classify hoaxes with accuracy above 80%, precision above 80%, and strong recall values, indicating SVM's ability to learn complex word patterns from large datasets [6]. Subsequent studies have also compared several classical machine learning algorithms for hoax news detection tasks, such as Logistic Regression, Random Forest, SVM, and Naive Bayes, and concluded that SVM often outperforms others in detecting news validity on Indonesian-language datasets, both in terms of accuracy and other evaluation metrics [7]. Another study that employed Logistic Regression on a large-scale dataset of more than 7,000 articles showed that the combination of TF-IDF and this algorithm can achieve accuracy up to 94.98%, indicating that classical models with appropriate feature representations remain relevant in modern hoax detection research [8].

Machine Learning and NLP Techniques for Hoax Detection

DOI: 10.52362/ijiems.v5i1.2287

E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

Globally, the implementation of machine learning techniques for fake news detection has developed rapidly. Various approaches include combinations of classical NLP and deep learning models, as well as hybrid methods that demonstrate significant improvements in classification performance. For example, a study combining NLP and machine learning on an English-language dataset showed that the use of algorithms such as Naive Bayes, Logistic Regression, and SVM with TF-IDF feature extraction can achieve accuracy close to 95%, highlighting the benefits of more refined word processing in text classification [9]. Furthermore, a major trend from 2020 to 2025 is the adoption of Transformer-based models (such as BERT and IndoBERT) for hoax and misinformation detection tasks. Recent studies indicate that hybrid models integrating contemporary NLP architectures with Bayesian optimization and recurrent units can achieve accuracy approaching 99.7%, pointing to a new direction in automatic hoax detection with richer language representation capabilities [10]. Another study utilizing IndoBERT on a large-scale dataset developed a robust Indonesian-language hoax detection system with tens of thousands of data samples, indicating that representations based on pre-trained language models are becoming an increasingly attractive alternative for fake news classification. The integration of Transformer-based models is also reflected in other research comparing the performance of IndoBERT and mBERT for political hoax detection, showing that locally pre-trained language models can provide performance advantages in specific linguistic contexts.

Benefits and Limitations

Classical machine learning methods such as Naive Bayes and SVM offer several advantages, including fast training time, interpretability, and relatively low computational requirements compared to deep learning models. However, their limitations become apparent when dealing with complex linguistic variations, ambiguous semantic contexts, and irony that frequently appear in online news. On the other hand, deep learning and Transformer-based models provide richer and more contextual feature representations but require much larger datasets and substantial computational resources. Both approaches demonstrate complementary research trends, in which classical methods remain relevant for medium- to large-scale hoax detection systems with limited resources, while modern models achieve higher performance in the context of large and complex datasets.

3. Methods

This study employs a quantitative approach with an experimental method to evaluate the performance of machine learning algorithms in detecting Indonesian-language hoax news. The research stages include dataset collection, text preprocessing, feature extraction, classification model training, and performance evaluation using statistical metrics. The general workflow of the research method can be illustrated as follows:

DOI: 10.52362/ijiems.v5i1.2287

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

Data Collection → Text Preprocessing → Feature Extraction → Model Training → Model Evaluation

Dataset
The dataset used in this study consists of 3,000 Indonesian-language news articles divided into two classes: hoax news and non-hoax news. The data were obtained through a scraping process from social media platforms and online news portals. Each data instance was subsequently validated using the fact-checking website TurnBackHoax.id to ensure the correctness of class labels. The dataset composition is balanced, with each class containing 1,500 samples. Class balancing is intended to prevent model bias toward a particular class and to improve the reliability of the evaluation results.

The dataset is divided into two subsets:
80% training data (2,400 news articles)
20% testing data (600 news articles)

This data split aims to provide sufficient data for model training while reserving independent data to assess the model's generalization capability.

Text Preprocessing
Text preprocessing aims to clean and normalize textual data so that it can be more effectively processed by machine learning algorithms. The preprocessing stages applied in this study include:
Case Folding : All characters in the text are converted to lowercase to avoid differences in meaning caused by capitalization.
Tokenization : The text is segmented into word units (tokens) based on whitespace and punctuation marks.
Stopword Removal : Common words that do not carry significant meaning for classification (e.g., "and," "the," "in," "to") are removed to reduce noise.
Indonesian Stemming : Each word is reduced to its root form using an Indonesian stemming algorithm (e.g., Nazief-Adriani or Sastrawi), so that variations of affixes can be standardized.
These preprocessing steps aim to reduce data complexity, decrease feature dimensionality, and improve the quality of text representation.

Feature Extraction Using TF-IDF
After preprocessing, the text is transformed into a numerical representation using the Term Frequency–Inverse Document Frequency (TF-IDF) method. TF-IDF measures the importance of a term in a document relative to the entire document collection (corpus).

DOI: 10.52362/ijiems.v5i1.2287

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

Mathematically, TF-IDF is defined as:

$$TFIDF(t,d) = TF(t,d) \times IDF(t)$$

$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$

Descriptions:

TF(t, d) = frequency of occurrence of term t in document d
df(t) = number of documents containing term t
N = total number of documents

The TF-IDF method assigns low weights to terms that frequently appear across many documents and high weights to terms that rarely appear but are informative. This representation is well suited for high-dimensional text classification tasks.

Classification Algorithms
This study employs two classification algorithms:

1) Naive Bayes

Naive Bayes is a probabilistic algorithm based on Bayes' Theorem with the assumption of feature independence. The probability that a document belongs to a particular class is calculated as:

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

Naive Bayes is known for its advantages in computational efficiency, low resource requirements, and good performance on large-scale text datasets.

2) Support Vector Machine (SVM)

SVM is a classification algorithm that seeks an optimal hyperplane to separate data from different classes with the maximum margin. For data that are not linearly separable, SVM can

DOI: 10.52362/ijiems.v5i1.2287

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

employ kernel functions. The strengths of SVM lie in its ability to handle high-dimensional data and its robustness against overfitting.

Model Training Process
The models are trained using the training data (80% of the dataset) with TF-IDF features as input. The default parameter settings of each algorithm are applied to maintain objectivity in the comparison. After training, the models are used to predict the class labels of the testing data.

Model Performance Evaluation
The model performance is evaluated using a confusion matrix and the following metric:

Accuracy

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

Precision

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

Recall

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

F1 Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Descriptions:
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

DOI: 10.52362/ijiems.v5i1.2287

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

The use of multiple evaluation metrics aims to provide a more comprehensive view of model performance, particularly in the context of hoax detection, where misclassification can have significant consequences.

4. Result And Discussion

Dataset Splitting and Testing Scenarios

The dataset consists of 3,000 news articles divided into two classes: hoax and non-hoax. The data distribution is presented in Table 1.

Table 1. Dataset Distribution

News Class	Number of Samples	Percentage
Hoaks	1.500	50%
Non-Hoaks	1.500	50%
Total	**3.000**	**100%**

The dataset is split using an 80% training and 20% testing scheme, resulting in:
Training data: 2,400 news articles
Testing data: 600 news articles
This approach is chosen to ensure that the model receives sufficient training data while maintaining reliable evaluation quality.

Examples of Preprocessing and Feature Extraction Results

The following is an example of text transformation at the preprocessing stage:

Original Text:
"Pemerintah RESMI menyatakan vaksin menyebabkan penyakit serius!!!"
After Preprocessing:
pemerintah resmi nyata vaksin sebab penyakit serius

The preprocessed text is then converted into numerical vectors using TF-IDF. An example of TF-IDF representation is shown in Table 2.
Table 2. Sample TF-IDF Values

Kata	Dokumen 1	Dokumen 2	Dokumen 3
pemerintah	0,214	0,000	0,187
vaksin	0,321	0,298	0,000
penyakit	0,187	0,000	0,245

DOI: 10.52362/ijiems.v5i1.2287

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), P-ISSN: 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

resmi	0,256	0,000	0,000

TF-IDF reduces the influence of common terms and emphasizes discriminative words for hoax classification.

Confusion Matrix of Classification Results

To provide a more detailed overview, the evaluation is also analyzed using confusion matrices.

1) Naive Bayes

Table 3. Naive Bayes Confusion Matrix

	Prediksi Hoaks	Prediksi Non-Hoaks
Hoaks	238	62
Non-Hoaks	54	246

Naive Bayes still produces a relatively significant number of misclassifications, particularly for hoax news that is classified as non-hoax.

2) Support Vector Machine (SVM)

Table 4. SVM Confusion Matrix

	Prediksi Hoaks	Prediksi Non-Hoaks
Hoaks	264	36
Non-Hoaks	30	270

SVM shows a lower number of errors compared to Naive Bayes, especially in terms of false negatives (hoax news classified as non-hoax).

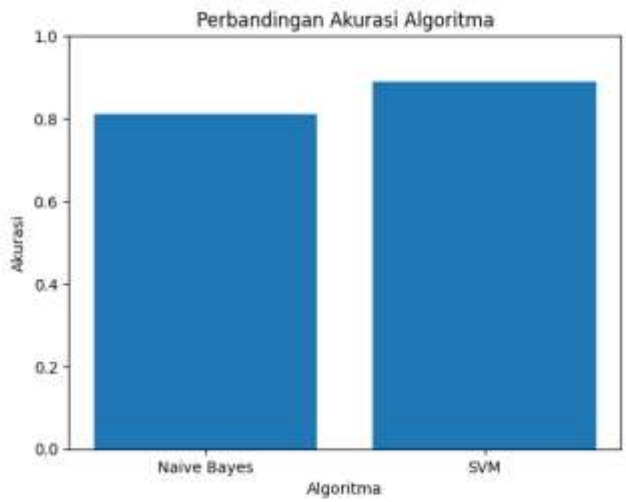
Table 5. Comparison of Algorithm Performance

Algoritma	Akurasi	Presisi	Recall	F1-Score
Naive Bayes	0,81	0,82	0,79	0,80
SVM	0,89	0,90	0,88	0,89

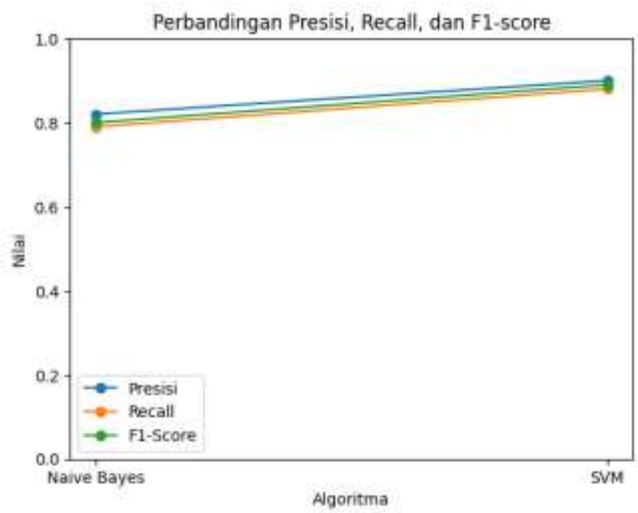
Based on Table 5, SVM outperforms Naive Bayes across all evaluation metrics.

DOI: 10.52362/ijiems.v5i1.2287

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94



The bar chart shows that the Support Vector Machine (SVM) algorithm achieves the highest accuracy of 0.89, while Naive Bayes attains an accuracy of 0.81. This indicates that SVM is more effective in distinguishing between hoax and non-hoax news in the Indonesian-language dataset used.



The line chart illustrates that SVM outperforms Naive Bayes across all evaluation metrics. The high recall value of SVM (0.88) indicates its strong capability to detect the majority of hoax news, thereby minimizing the risk of hoaxes being misclassified as valid news.

Naive Bayes offers advantages in terms of speed and model simplicity, making it suitable for systems with limited resources. However, its assumption of feature independence makes it less

DOI: 10.52362/ijiems.v5i1.2287

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

optimal for handling the complex structures of natural language. In contrast, Support Vector Machine (SVM) can perform optimally on high-dimensional data such as TF-IDF representations. SVM's ability to determine the optimal hyperplane allows it to be more robust against noise and common word variations frequently found in hoax news. The high recall value of SVM is particularly important in the context of hoax detection, as misclassifying hoaxes as valid news can have serious societal consequences. The findings of this study are consistent with previous research, which indicates that SVM is one of the most stable and accurate algorithms for Indonesian-language text classification tasks.

To strengthen the scientific contribution, the results of this study are compared with several related studies conducted over the past five years.

Table 6. Comparison with Previous Studies

Peneliti & Tahun	Dataset	Metode	Akurasi
Sudrajat et al., 2025	600 berita	Naive Bayes	0,82
Lazuardi et al., 2023	1.200 berita	Passive Aggressive	0,85
Tandiano & Jollyta, 2025	2.000 berita	SVM	0,87
Fernandes & Shita, 2025	2.500 berita	SVM + RF	0,88
Penelitian ini	**3.000 berita**	**SVM + TF-IDF**	**0,89**

Based on Table 6, this study demonstrates competitive performance, even surpassing some previous studies. This is influenced by:

1. A larger and balanced dataset.
2. A comprehensive text preprocessing pipeline (case folding, stopword removal, stemming).
3. The effective use of TF-IDF in representing text features.

The improvement in accuracy, although not drastic, indicates that classical approaches such as SVM remain relevant and stable for Indonesian-language hoax detection.

Research Novelty

This study confirms that the combination of TF-IDF and SVM remains superior on large-scale Indonesian-language hoax datasets, while also providing a comprehensive evaluation analysis using confusion matrices and cross-study comparisons.

5. Conclusions And Suggestions

This study demonstrates that text-based machine learning can detect Indonesian-language hoax news with a relatively high level of accuracy. Among the two algorithms tested, Support Vector

DOI: 10.52362/ijiems.v5i1.2287

E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

Machine (SVM) delivered the best performance on this dataset. This implementation can serve as a foundation for developing automatic hoax detection systems in real-world applications.

Implications and Future Development
These results open opportunities for the development of real-time hoax detection systems in:
1. Social media platforms
2. News portals
3. Digital literacy educational applications

Further development can be pursued through:
1. Incorporation of deep learning algorithms (e.g., LSTM, BERT)
2. Semantic-based feature analysis
3. Larger datasets with more diverse topics

References

[1] Alghamdi, J dkk. (2023). “A comprehensive survey on machine learning approaches for fake news detection”. Multimedia Tools and Applications. https://link.springer.com/content/pdf/10.1007/s11042-023-17470-8.pdf
[2] Alshuwaier, FA & Alsulaiman, FA. (2025). Fake News Detection Using Machine Learning and Deep Learning Algorithms: A Comprehensive Review and Future Perspectives. MDPI Vol. 14 Issue 9 https://www.mdpi.com/2073-431X/14/9/394
[3] Haq, MZ dkk. (2024). “Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial”. Jurnal Minfo Polgan Vol.13 No.1 (2024). https://jurnal.polgan.ac.id/index.php/jmp/article/view/13937
[4] A. Sudrajat, R. R. Wulandari, and E. Syafwan. (2022). “Indonesian Language Hoax News Classification Based on Naïve Bayes,” *Journal of Applied Intelligent System*, vol. 7, no. 1, May 2022. DOI: 10.33633/jais.v7i1.5985. [Online]. https://publikasi.dinus.ac.id/jais/article/view/5985
[5] R. Embun Safira and A. Nurlayli. (2025). “Comparative analysis of Indonesian news validity detection accuracy using machine learning,” *Journal of Engineering and Applied Technology*, vol. 4, no. 1, 2025. DOI: 10.21831/jeatech.v4i1.58791
[6] M. D. Desriansyah, I. Utna Sari, and Z. Zulfahmi. (2024)). “Analisis Efektivitas Algoritma Machine Learning dalam Deteksi Hoaks: Pada Berita Digital Berbahasa Indonesia,” *Jurnal Sistem Informasi dan Informatika (JISKA)*, vol. 3, no. 1, 2024. DOI: 10.47233/jiska.v3i1.2024
[7] R. Rakhmat Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, and F. Alzami. (2022). “Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v0i0.xxx
Volume 5 , Issue 1 , January 2026, pp. 82-94

untuk Klasifikasi Berita Hoax pada Berita Online Indonesia," *Jurnal Masyarakat Informatika*, vol. 13, no. 2, Nov. 2022. DOI: 10.14710/jmasif.13.2.47983

[8] A. Kartika Dewi, N. F. Rahmadani, R. Syahputri, L. R. Nasution, and M. Furqon. (2025). "Deteksi Berita Hoax Pada Platform X Menggunakan Pendekatan Text Mining dan Algoritma Machine Learning," *Data Sciences Indonesia*, vol. 5, no. 1, Jul. 2025. DOI: 10.47709/dsi.v5i1.6011. [Online]. https://jurnal.itscience.org/index.php/dsi/article/view/6011

[9] I. Indra, A. U. Hamdani, S. Setiawati, Z. D. Mentari, and M. H. Purnomo. (2024). "Comparison of K-NN, SVM, and Random Forest Algorithm for Detecting Hoax on Indonesian Election 2024," *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 13, no. 1, Mar. 2024. DOI: 10.23887/janapati.v13i1.76079. [Online]. https://ejournal.undiksha.ac.id/index.php/janapati/article/view/76079

[10] M. Y. Ridho and E. Yulianti. (2024). "From Text to Truth: Leveraging IndoBERT and Machine Learning Models for Hoax Detection in Indonesian News," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 3, Sep. 2024. DOI: 10.26555/jiteki.v10i3.29450. [Online]. https://journal.uad.ac.id/index.php/JITEKI/article/view/29450

[11] A. Sudrajat, R. R. Wulandari, & E. Syafwan. (2025). *Indonesian Language Hoax News Classification Based on Naïve Bayes*, JAIS, 2025.

[12] V. Prisscilya & A. S. Girsang. (2024). *Classification of Indonesia False News Detection Using Bertopic and IndoBERT*, JIST, 2024.

[13] M. D. Desriansyah et al.. (2024).*Analisis Efektivitas Algoritma Machine Learning dalam Deteksi Hoaks*, JISKA, 2024.

[14] A. Kartika Dewi et al. (2025). *Deteksi Berita Hoax Pada Platform X…*, Data Sciences Indonesia, 2025.

[15] M. F. Lazuardi et al. (2023). *Hoax News Detection Using Passive Aggressive Classifier…*, Jurnal Teknik Informatika, 2023.

[16] T. A. Roshinta et al. (2023). *Sistem Deteksi Berita Hoax Berbahasa Indonesia Bidang Kesehatan*, REMIK, 2023.

[17] T. Faujar Mustafa & H. Alfianti. (2025). *Klasifikasi Berita Palsu Berbahasa Indonesia…*, JSIT, 2025.

[18] D. Samudra A. Suryama & S. Jatmiko. (2025). *Development of an Indonesian Hoax Detection System…*, J-INTECH, 2025.

[19] A. Halim Tandiano & D. Jollyta. (2025). *Classification of Fake News… Using SVM*, JURTEKSI, 2025.

[20] R. P. Fernandes & R. T. Shita. (2025). *Penerapan Metode SVM dan Random Forest…*, Ticom, 2025.

DOI: 10.52362/ijiems.v5i1.2287