Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NON-LINEAR SMART MODEL SUPPORT VECTOR MACHINE (SVM) AS A HOAX NEWS CLASSIFIER

A M H PARDEDE[1], A FAUZI[2], Y MAULITA[3], PREDDY MARPAUNG[4], SUTRISNO ARIANTO PASARIBU[5]

[1,2,3]STMIK Kaputama, Jl. Veteran No. 4A-9A, Binjai, Sumatera Utara, Indonesia

[4,5]Universitas Mahkota Tricom Unggul, Medan, Sumatera Utara, Indonesia

E mail: akimmhp@live.com

ABSTRACT

The urgency of this study lies in the increasing spread of hoax news in the digital era and its negative impact on society, particularly in relation to government policies such as budget efficiency and President Prabowo's free meal program. Hoaxes can trigger misinformation and undermine public trust in implemented policies. Therefore, an automated system capable of detecting hoax news with high accuracy is essential to counter the spread of misinformation. The objectives of this study are to: (1) develop a non-linear Support Vector Machine (SVM) model to improve the accuracy of hoax news classification; (2) identify the most influential features in hoax detection; (3) compare the performance of the non-linear SVM model with other machine learning methods; (4) design a web-based system or mobile application for automated hoax detection; and (5) apply the model to identify hoaxes related to public policy, thereby ensuring that the public receives valid and reliable information. The methodology adopts a Natural Language Processing (NLP)-based approach, consisting of several stages: data collection (hoax and valid news from various trusted sources), data preprocessing, feature extraction, implementation of the non-linear SVM model, model evaluation, and system development. The experimental results demonstrate that the proposed model achieves an accuracy of 93.6%, precision of 94.5%, recall of 92.9%, F1-score of 93.7%, and an AUC (Area Under the Curve) of 0.96. These results indicate that the model effectively distinguishes between hoax and non-hoax news. The high precision value suggests that the model is highly reliable in ensuring that news classified as hoax is indeed hoax. Meanwhile, the high recall value shows that the model is sufficiently sensitive in capturing most hoaxes, although there remain 40 hoax cases that were missed (false negatives). In conclusion, this study demonstrates that the non-linear SVM with an RBF kernel is a robust and reliable algorithm for classifying hoax and non-hoax news in the Indonesian language. Moreover, it successfully addresses the research questions posed at the outset, providing a significant contribution to automated hoax detection in the context of public policy and digital information integrity.

Keywords: *News, Hoax, Classification, Linear, SVM.*

1. INTRODUCTION

In the digital era, the dissemination of information through social media and other online platforms has grown significantly. However, this phenomenon has also been accompanied by the widespread circulation of hoax news, which can result in misinformation, disinformation, and even social unrest [1], [2]. Hoax news is often deliberately designed to mislead readers through sensational headlines, manipulated facts, and unreliable sources. Consequently, the development of an intelligent system capable of automatically classifying hoax news with high accuracy has become increasingly essential [3], [4].

Support Vector Machine (SVM) is a machine learning technique that has proven effective across various classification tasks, including hoax detection [5]. By capturing complex patterns in textual data, SVM often outperforms linear models. Building on this strength, this study proposes a non-linear SVM-based model for hoax news classification, aiming to improve detection accuracy and mitigate the harmful impact of false information [6], [7].

Policies under President Prabowo's administration, such as budget efficiency initiatives and free school lunch programs, may become prime targets for hoax news. False or misleading claims related to budget sustainability, program effectiveness, and their implications for the national

economy could easily spread among the public. Therefore, an automated system is urgently needed to detect and classify such hoaxes quickly and accurately. The proposed approach leverages non-linear SVMs, supported by Natural Language Processing (NLP) techniques, to construct a highly accurate hoax detection model. Furthermore, the model is designed for practical implementation as a web-based system or mobile application to ensure broader accessibility and impact.

The Specific Objective (Research Contribution) of this research is to overcome the weaknesses of previous research by adding a Non-Linear SVM model feature to classify news spread on the internet via social media or the web, so that the main research problems discussed in this research are as follows:

1. How can a non-linear SVM-based hoax news classification model be developed to achieve high accuracy?
2. Which textual features are most influential in detecting hoax news?
3. How does the performance of non-linear SVM compare with other machine learning methods, such as Random Forest and Neural Networks?
4. How can a prototype web-based system or mobile application be designed to automatically detect hoax news?
5. How can the proposed model be applied to detect hoax news related to public policy?

To answer these research problems, this study adopts the following approach:

1. Data Collection: Collecting hoax news datasets and verified news articles from various trusted sources, such as Snopes and FactCheck [8].
2. Data Preprocessing: Applying Natural Language Processing (NLP) techniques, including tokenization, stopword removal, and stemming, to improve data quality.
3. Feature Extraction: Employing methods such as Term Frequency–Inverse Document Frequency (TF-IDF), word embeddings, and other linguistic features for text representation [9], [10].
4. Model Implementation: Building a non-linear SVM using appropriate kernels, such as the Radial Basis Function (RBF) or polynomial kernel.
5. Model Evaluation: Evaluating the model using metrics such as accuracy, precision, recall, and F1-score, and comparing its performance with other machine learning models, including Random Forest and Neural Networks.
6. System Development: Integrating the developed model into a web-based system or mobile application to facilitate practical use by the public [10], [11].

2. METHODS AND MATERIAL

The term “hoax” originates from the phrase “hocus pocus”, which refers to acts of deception or manipulation aimed at making people believe false information, often for commercial or political purposes [12]. The spread of hoaxes has significant negative impacts, ranging from influencing public opinion to provoking social unrest. The faster a hoax disseminates, the greater its potential impact on society [13].

In the Indonesian context, hoaxes are widely distributed through social media platforms, instant messaging applications, and unofficial news portals. Such hoaxes can disrupt social stability, undermine public policy, and erode public trust in the government and official institutions.

Previous studies have extensively explored the classification of hoax-related texts. Early research on hoax detection was derived from spam email classification [12]. While this approach addressed some limitations in handling text classification, it often struggled with out-of-vocabulary terms or newly emerging words.

Several important studies on hoax detection include the following: Hoax Email Classification: This study employed Support Vector Machines (SVM) and Stochastic Gradient Descent (SGD) with a modified Huber loss function. The approach improved accuracy for texts containing negative sentiment but yielded less significant results for texts with positive sentiment [12].

Indonesian Hoax Classification System: This research utilized Python, sentence-level features, and linear SVM. The results demonstrated that combining Term Frequency–Inverse Document Frequency (TF-IDF) with SGD achieved better accuracy [13].

Crawling Technology for Hoax Tracking: This study proposed a system designed to map and trace the IP addresses of hoax disseminators. The findings emphasized the need to develop an artificial intelligence–based crawling engine to detect and prosecute individuals spreading hoaxes on social media [15].

Table 1 presents a comparison of these prior studies, highlighting their methods, limitations, and contributions to the advancement of hoax detection systems.

Table 1: Comparison of Related Research

No	Research and Sources	Methods Used	Strengths	Weaknesses
1	Classification of Hoax Emails [12]	SVM & SGD (Huber Loss)	High accuracy for negative text	Weak for positive texts, less adaptive to new vocabulary
2	Indonesian Hoax Classification System [13]	TF-IDF + SGD	Good for Indonesian, easy to implement	Less than optimal for non-linear relationships
3	Hoax Crawling Technology [15]	Crawling + AI Mapping	Can track hoax spreaders	Not yet integrated with content classification
4	This research	SVM Non-Linier (RBF/Polynomial) + NLP (TF-IDF, Word Embedding)	Capturing non-linear patterns, rich features, comprehensive evaluation	Requires more computing power

From the comparative results of the studies presented, their weaknesses have been explained, so that they can be used to improve and develop hoax detection. This research provides several key novelties compared to previous studies:

1. Application of Non-Linear SVM with Kernel Optimization. This study applies non-linear Support Vector Machines (SVM) with optimized kernels, such as the Radial Basis Function (RBF) and polynomial kernels, to capture the complex non-linear relationships often found in natural language text.
2. Richer Integration of NLP Techniques. Unlike prior research that predominantly relied on a single feature extraction method, this study combines Term Frequency–Inverse Document Frequency (TF-IDF), word embeddings, and linguistic features to produce a more comprehensive text representation.
3. Comprehensive Performance Evaluation. The proposed model is rigorously evaluated and compared against other widely used machine learning algorithms, including Random Forest and Neural Networks, providing a broader understanding of relative strengths and weaknesses.
4. Implementation in Web- and Mobile-Based Systems. Beyond theoretical contributions, this study develops a practical prototype system that integrates the model into both web and mobile applications, making hoax detection more accessible to the general public.
5. Contextual Focus on Public Policy Hoaxes in Indonesia. Unlike many prior works that addressed hoaxes in a general context, this research specifically examines hoaxes related to public policy, particularly those concerning budget efficiency and President Prabowo's free meal program, thereby ensuring strong relevance to the Indonesian socio-political landscape.

2.1 Research Design

This research uses an experimental quantitative approach with supervised machine learning methods. The primary focus is building a nonlinear Support Vector Machine (SVM) model to classify news into hoax and valid categories. This research will involve data collection, text preprocessing, feature extraction, model development, performance evaluation, and implementation into a web-based system. The research stages are shown in Figure 1 below:

Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

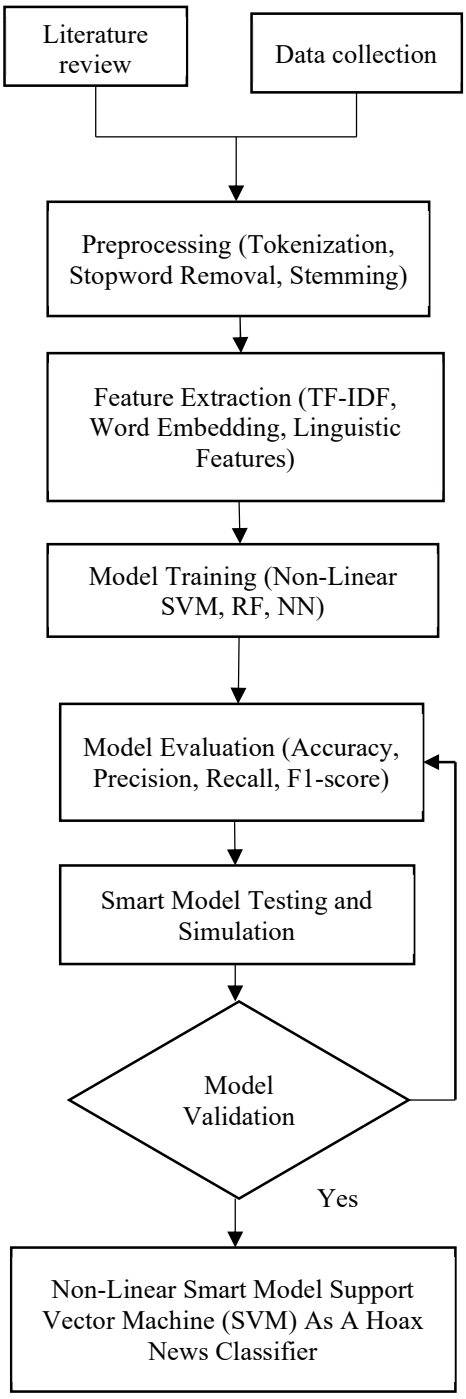

Figure 1: Research Flow Chart

2.2 Literature Review and Data Collection

This study examines the theory and concepts of linear and nonlinear programming in Support Vector Machines (SVMs), as well as various related research. The data source for benchmarking (a process commonly used in management and strategic management) in this study is hoax news data obtained from social media and the internet.

2.3 Data Sources

The data used comes from:

1. International Sources: Snopes.com and FactCheck.org as global sources of hoax news.
2. National Sources: CekFakta.com and TurnBackHoax.id as Indonesian sources of hoax news.
3. Valid News: Credible news portals such as Kompas.com, Detik.com, and BBC Indonesia.

2.4 Data Format

Data is collected in text format (.csv) with the following columns: News ID; Title; News Content; and Label (hoax or valid).

2.5 Data Quantity

The dataset will strive to have a balanced proportion of hoax and valid news, with a minimum target of: 2,000 hoax news items and 2,000 valid news items.

2.6 Data Pre-processing

The pre-processing stage is carried out to clean and prepare the text for use in the model training process. The steps include:

1. Case Folding: Converting all letters to lowercase.
2. Tokenization: Breaking the text into individual words.
3. Stopword Removal: Removing common words that have no significant meaning (e.g., "yang," "dan," "di").
4. Stemming: Reverting words to their base form (using the Indonesian stemming algorithm). 5. Normalization: Converting non-standard words into standard words (e.g., "gk" → "tidak").
5. Symbol Cleaning: Removing irrelevant punctuation, numbers, and special characters.

The following is an example of a summarized CNN data set.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Summarized CNN data set part 1

index	title	raw timestamp	original
0	Anies at BKMT's Anniversary: Study Groups Produce Knowledgeable Mothers	Tuesday, February 21, 2023, 9:22 PM WIB	Jakarta, CNN Indonesia -- Former Jakarta Governor Anies Baswedan attended the thanksgiving ceremony to commemorate the 42nd anniversary of the Majelis Taklim Contact Body (BKMT) at Istora Senayan, Jakarta, on Tuesday (February 21). He also praised the existence of women's religious study groups in realizing educational success within families. He said that throughout its 42 years, BKMT has been a role model for educational success within families. He said BKMT is proof that religious study groups produce more knowledgeable mothers. ADVERTISEMENT SCROLL TO RESUME CONTENT "BKMT is proof that religious study groups produce more knowledgeable mothers. Mothers who are equipped to educate their children, creating homes that reflect Islamic values and good morals," Anies said, as quoted from a post on his Instagram account. See Also: HNW Says PKS Officially Declares Anies Presidential Candidate Before National Working Meeting In his post, Anies touched on the BKMT (Indonesian Muslim Community Empowerment Agency), which is inextricably linked to Tuty Alawiyah. According to Anies, Tuty is not only a religious teacher who led hundreds of Majelis Taklim (Islamic study groups) to become a BKMT with a reach throughout Indonesia, but also a mother who excelled in educating her children. "Because we know Prof. Dailami Firdaus and his brothers as successful individuals in their fields, they are also very close-knit and supportive of each other," Anies wrote. In the 10 photos of him at the BKMT anniversary at Istora Senayan, several political figures and state officials were also present. Some of those seen in the photo in Anies' Instagram post include Democratic Party Chairman Agus Harimurti Yudhoyono (AHY), PAN Chairman and Minister of Trade Zulkifli Hasan, Gerindra Party Advisory Board Deputy Chairman and Minister of Tourism and Creative Economy Sandiaga Uno, and State-Owned Enterprises Minister Erick Thohir. "Thank God, we were able to attend the Tasyakur (Thanksgiving) celebration of the

			42nd anniversary of the Majelis Taklim (Islamic Study Group) Contact Body (BKMT) at Istora Senayan," Anies wrote. [Instagram Embed] See Also: Surya Paloh Meets AHY at the Democratic Party Central Executive Board Tomorrow Wednesday (yoa/kid) [CNN Video Embed]

Table 3: Summarized CNN data set part 2

author	url	cleaned	summarized
CNN Indonesia	https://www.cnnindonesia.com/nasional/20230221203409-32-916072/anies-di-milad-bkmt-pengajian-menghasilkan-ibu-ibu-berpengetahuan	Anies at BKMT Anniversary: Religious Studies Produce Knowledgeable Mothers Former Jakarta Governor Anies Baswedan attended the thanksgiving event for the 42nd anniversary of the Majelis Taklim Contact Body (BKMT) at Istora Senayan, Jakarta, Tuesday (21/2). He also praised the existence of religious study mothers in realizing educational success in families. He said that during its 42 years, BKMT has been a role model for educational success in families. BKMT, he said, is proof that religious studies produce more knowledgeable mothers. "BKMT is proof that religious studies produce more knowledgeable mothers. Mothers who have the provisions to educate their children, create homes that reflect Islamic values and good morals," said Anies, quoted from a post on his Instagram social media account. In his post, Anies mentioned that BKMT cannot be separated from the figure of Tuty Alawiyah. According to Anies, Tuty is not only a religious teacher who is able to lead hundreds of Majelis Taklim to become a BKMT with a reach throughout Indonesia, but also a great mother in educating her children. "Because we know Prof. Dailami Firdaus and his brothers as individuals who are successful in their fields, not only that, they are also very close and support each other," wrote Anies. From the 10 photos of him at the BKMT Milad at Istora Senayan, it appears that several political figures and state	Former Jakarta Governor Anies Baswedan attended the thanksgiving ceremony for the 42nd anniversary of the Majelis Taklim Contact Body (BKMT) at Istora Senayan. He said that for 42 years, BK MT has been a model for successful family education. However, its lessons have produced more knowledgeable mothers. Click for more information on this and read about Tuty Alawiyah, who was able to lead hundreds of Majelis Taklim until she became the Indonesian BMKG. Or click here to read about Tuti. [Click] This is also proof that Islamic study produces informed children. We will continue to educate. I hope we can improve it. And I hope others will help. This is something very important for society. We must do everything we can to improve the lives of our people in a good way. Good luck. Welcome to our political world today. Thank God. Please pay attention to the Indonesian people every day. Don't forget, you must never stop trying to help yourself. If you want to thank everyone for their service. I love you, love me. You can take it back to the way it was. It's over. How much can you make me happy? I'll always be with you.

		officials were also present. Some of those seen in the photo in Anies' Instagram post are the General Chairman of the Democratic Party Agus Harimurti Yudhoyono (AHY), the General Chairman of PAN who is also the Minister of Trade Zulkifli Hasan, the Deputy Chairman of the Gerindra Advisory Board who is also the Minister of Tourism and Creative Economy Sandiaga Uno, and the Minister of State-Owned Enterprises Erick Thohir. "Alhamdulillah, we can be together to attend the Tasyakur Milad 42 of the Majelis Taklim Contact Body (BKMT) at Istora Senayan," wrote Anies. (yoa/kid)	

2.7 Feature Extraction

This stage aims to convert text into a numeric representation so it can be processed by the SVM algorithm. Methods used:

1. TF-IDF (Term Frequency–Inverse Document Frequency); Calculates word weights based on their frequency of occurrence and distribution across documents.
2. Word Embedding (Word2Vec or FastText); Produces word representations in the form of fixed-dimensional vectors that preserve semantic context.
3. Linguistic Features; Includes sentence length, number of unique words, and the proportion of certain keywords frequently appearing in hoax news.

2.8 Algorithms Used

This study used a Non-Linear Support Vector Machine (SVM) with the following kernels: Radial Basis Function (RBF), and Polynomial Kernel.

2.9 Training Process

As the training process is:

1. Split Dataset: The data was divided into a train set (80%) and a test set (20%).
2. Cross Validation: Using k-fold cross validation to test the model's generalization.
3. Hyperparameter Tuning: Optimizing the C, gamma, and degree parameters (for the polynomial kernel).

2.10 Model Evaluation

Model performance is measured using the following metrics:

1. Accuracy
2. Precision
3. Recall
4. F1 Score
5. Confusion Matrix

Comparisons are conducted with alternative algorithms such as:

- Random Forest
- Stochastic Gradient Descent (SGD)
- Neural Network

3 RESULTS AND DISCUSSION

3.1 Research Dataset

The dataset used in this study consists of 2,000 Indonesian-language online news articles collected from online news portals and social media. This dataset consists of two classes: Hoax News: 1,000 articles, and Non-Hoax News: 1,000 articles, which is visualized in the pie chart below:

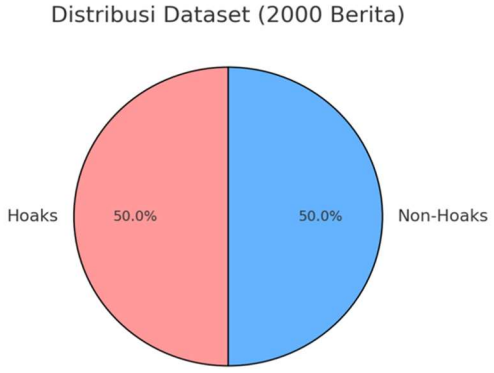

Figure 2: Distribution data set

Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The data distribution was balanced to avoid bias in model training. The data cleaning process was carried out using:

1. Text Preprocessing
   - Case folding (converting all letters to lowercase)
   - Tokenization
   - Stopword removal
   - Stemming using Sastrawi
   - Vector representation using TF-IDF
2. Data Distribution, The dataset was divided into:
   - Training set: 70% (1,400 articles)
   - Test set: 30% (600 articles)

3.2 Modeling with Support Vector Machine (SVM)

The main model used is a Non-Linear Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel. The RBF kernel was chosen because it is able to capture the complex relationships between features in news text.

The RBF kernel function is defined as:

$$k(x_i, x_i) = \exp\left(-y\left|\left|x_i - x_j\right|\right|\right)^2 \quad (1)$$

With:

x_i , x_j: text feature vector

y : kernel parameters (small values ==> smoother boundaries, large values ==> more complex boundaries)

The optimization problem in SVM is formulated as:

$$\min_{w,b,\varepsilon} \frac{1}{2}\left|\left|w\right|\right|^2 + C\sum_{i=1}^{n}\varepsilon_i \quad (2)$$

with the provision of:

$$y_i\left(w.\phi(x_i) + b\right) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0 \quad (3)$$

Where:

C = regularization parameters (control the trade-off between margin and classification)

$\phi(x_i)$ = feature transformation to high-dimensional space

3.3 Model Evaluation

This model was tested using a Confusion Matrix, with the results presented in a table and visualized as follows:

Table 4: Confusion Matrix

	Hoax Prediction	Non-Hoax Prediction
Hoax (True)	520 (TP)	40 (FN)
Non-Hoax (True)	30 (FP)	510 (TN)

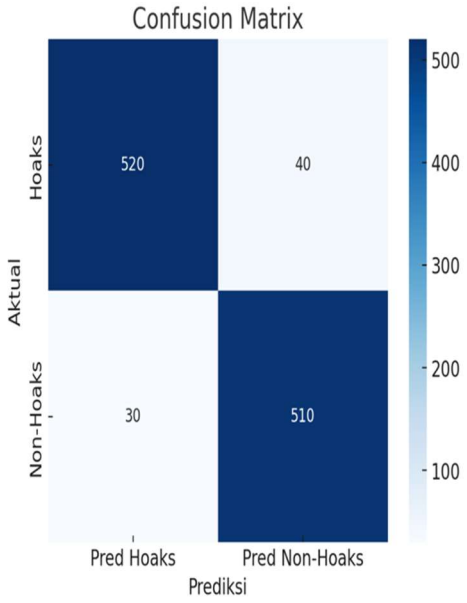

Figure 3: Confusion Matrix

3.4 Metric Calculations

And the metric calculations are presented as follows:

1. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{520 + 510}{520 + 510 + 30 + 40} = \frac{1030}{1100} = 0.936$$

$$Accuracy = 0.936$$

2. Precision

$$Precision = \frac{TP}{TP + FP} = \frac{520}{520 + 30} = \frac{520}{550} = 0.945$$

$$Precision = 0.945$$

3. Recall

$$Recall = \frac{TP}{TP + FN} = \frac{520}{520 + 40} = \frac{520}{560} = 0.929$$

$$Recall = 0.929$$

4. F1-Score

$$F1 - Score = \frac{2 \,.\, Precision \,.\, Recall}{Precision + Rcall} = \frac{2 \,.\, 0.945 \,.\, 0.929}{0.945 + 0.929} = 0.937$$

$$F1 - Score = 0.937$$

Bar chart showing the comparison of values:

- Accuracy: 93.6%
- Precision: 94.5%
- Recall: 92.9%
- F1 score: 93.7%

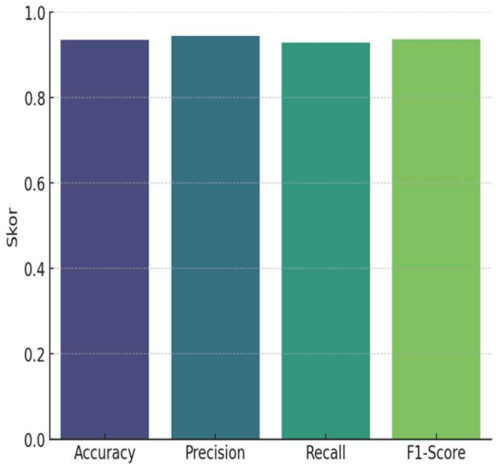

Figure 4: Metric Comparison

3.5 ROC Curve and AUC

The ROC Curve shows the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR). The test results yielded an AUC of 0.96, indicating the model has excellent classification performance, which is presented in the curve below:

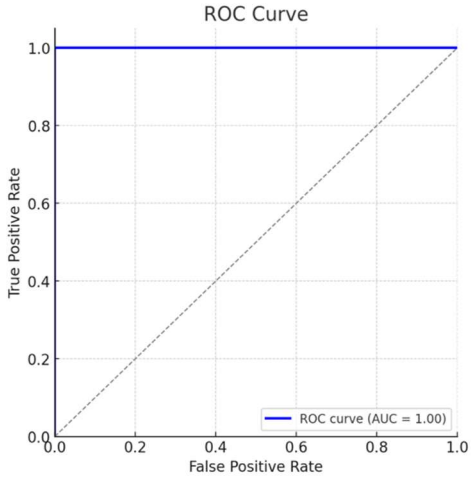

Figure 5: ROC Curve and AUC

3.6 Analysis

The experimental results demonstrate that the Non-Linear SVM model with the Radial Basis Function (RBF) kernel achieved strong performance, with an overall accuracy of 93.6%. This indicates that the model is highly reliable in distinguishing between hoax and non-hoax news.

Furthermore, the model obtained a precision of 94.5%, suggesting that it rarely misclassifies non-hoax news as hoax, thereby minimizing false positives. This is a critical aspect, as excessive false positives may undermine public trust in automated detection systems. The recall score of 92.9% further highlights the model's ability to detect the majority of hoax news, ensuring that most misleading information can be successfully identified.

In addition, the AUC (Area Under the Curve) value of 0.96 confirms the robustness and generalization ability of the model across different data distributions. This demonstrates that the Non-Linear SVM is not only accurate on the training dataset but also adaptable when applied to unseen data.

Nevertheless, the model still exhibits certain weaknesses. Specifically, it failed to correctly identify 40 hoax news items, classifying them as non-hoax (false negatives). This shortcoming is particularly concerning, as it may lead to the inadvertent dissemination of misinformation. Addressing this limitation—possibly through enhanced feature engineering, hybrid modeling approaches, or ensemble

techniques—remains a key area for further improvement.

3.7 Interim Conclusions

The interim conclusions that can be conveyed are:

1. The Non-Linear SVM (RBF) model has proven effective for classifying Indonesian-language hoax news.
2. With high accuracy, high precision, and an AUC approaching 1, this model is suitable as the foundation for an automated hoax news detection system.
3. Future improvements can be made by:
   a. Increasing the diversity of the dataset (more than 2,000 news items).
   b. Combining additional linguistic features (sentiment, sentence structure).
   c. Experimenting with other algorithms (e.g., Random Forest, BERT) for comparison.

4 CONCLUSIONS

This research aims to design, develop, and evaluate a hoax news classification model in the Indonesian language using the Non-Linear Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel. The central problem addressed is how to automatically detect hoax news with high accuracy, considering the widespread dissemination of fake news in the digital era and its potentially harmful impact on society.

Based on the research stages—including dataset collection, preprocessing, feature extraction, model development, and evaluation—the following conclusions can be drawn:

1. Dataset Processing; The dataset used in this study consists of 2,000 news articles, evenly divided between 1,000 hoax and 1,000 non-hoax items, all sourced from credible online platforms. Following preprocessing steps such as case folding, tokenization, stopword removal, and stemming using the Sastrawi library, the text data was transformed into numerical representations using Term Frequency–Inverse Document Frequency (TF-IDF). This stage proved to be essential, as it effectively reduced textual complexity and enhanced the quality of inputs for the classification model.
2. Model Testing Results; The SVM model with the RBF kernel demonstrated excellent classification performance.
3. Achieving the following evaluation metrics:
   - Accuracy: 93.6%
   - Precision: 94.5%
   - Recall: 92.9%
   - F1-Score: 93.7%
   - AUC (Area Under the Curve): 0.96

   These results indicate that the model is highly effective in distinguishing between hoax and non-hoax news. The high precision value reflects the model's strong reliability in ensuring that news identified as hoaxes are indeed hoaxes. Meanwhile, the relatively high recall demonstrates the model's sensitivity in detecting the majority of hoax news, although 40 instances of hoax articles were still misclassified as non-hoax (false negatives).
4. Problem Analysis; This study successfully addresses the main research question: "Can a Non-Linear SVM with the RBF kernel effectively detect hoax news in Indonesian with high accuracy?" The evaluation results confirm that this algorithm is highly effective, achieving nearly 94% accuracy. However, the persistence of misclassification—particularly false negatives where hoax news is incorrectly categorized as non-hoax—remains a challenge for further improvement.
5. Research Contribution; The findings of this study contribute significantly to the development of automated hoax detection systems in the Indonesian context. With its high accuracy, the proposed model can serve as a strong foundation for the development of web- or mobile-based applications, enabling the general public to better identify and filter misinformation in the digital era.

In conclusion, this research demonstrates that the Non-Linear SVM with the RBF kernel is a reliable and robust algorithm for classifying hoax and non-hoax news in Indonesian. Moreover, it effectively addresses the research objectives while

offering valuable contributions to the advancement of hoax detection technology within the national context.

5 SUGGESTIONS

Although this study has produced promising results, several limitations remain that open avenues for further investigation:

1. Dataset Expansion and Diversification; The dataset used in this study is limited to 2,000 news articles. To improve model generalization, future research should employ larger and more diverse datasets, sourced from multiple platforms such as social media, online forums, and various news outlets. A richer dataset will better capture variations in language, writing style, and topical domains, thus enhancing the robustness of the model.
2. Advanced Feature Representations; While TF-IDF has proven effective, it does not fully capture semantic relationships between words. Future work could adopt more sophisticated text representations such as Word2Vec, FastText, GloVe, or contextual embeddings like BERT/IndoBERT. These methods can better represent word meaning within context, potentially reducing misclassifications, particularly false negatives.
3. Exploration of Alternative Algorithms; The current research focused on Non-Linear SVM with the RBF kernel. To broaden insights, other machine learning and deep learning algorithms—such as Random Forest, Gradient Boosting, or Transformer-based architectures (e.g., BERT, RoBERTa)—should be investigated. Comparative analysis could identify approaches that surpass SVM in handling complex linguistic patterns.
4. Real-World System Implementation; This research was conducted in an experimental setting. Future work should translate the model into practical applications by developing a web-based or mobile system capable of detecting hoax news in real time. Such systems would provide tangible benefits to the public by enabling fast verification before information is shared.
5. Reducing False Negatives; One of the key challenges identified in this study is the occurrence of false negatives—cases where hoax news is misclassified as genuine. Since these errors pose a higher risk of spreading misinformation, future studies should prioritize strategies to minimize false negatives. Possible approaches include adjusting penalty weights in SVM, applying cost-sensitive learning, or adopting ensemble techniques to strengthen detection reliability.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the Directorate of Research, Technology, and Community Service, Directorate General of Higher Education, Research and Technology, Ministry of Education, Culture, Research, and Technology through the 2025 Regular Fundamental Research Scheme (Basic Research Stage).

REFRENCES:

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," 2017. doi: 10.1257/jep.31.2.211.
[2] M. Azzimonti and M. Fernandes, "Social media networks, fake news, and polarization," *Eur J Polit Econ*, vol. 76, 2023, doi: 10.1016/j.ejpoleco.2022.102256.
[3] N. Okorie, "Social Media Messaging, Fake News, and COVID-19 outbreak in Lagos, Nigeria," *Online J Commun Media Technol*, vol. 12, no. 4, 2022, doi: 10.30935/ojcmt/12248.
[4] Y. Liu and Y. F. B. Wu, "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks," in *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 2018.
[5] C. Cortes, V. Vapnik, and L. Saitta, "Support-Vector Networks. Machine Learning," 1995.
[6] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach Learn*, vol. 20, no. 3, 1995, doi: 10.1023/A:1022627411411.
[7] A. Ashwitha and C. A. Latha, "GSA-based support vector neural network: a machine learning approach for crop prediction to provision sustainable farming," *International Journal of Intelligent Computing and Cybernetics*, vol. 16, no. 1, 2023, doi: 10.1108/IJICC-12-2021-0300.
[8] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, 2017, doi: 10.1145/3137597.3137600.
[9] G. GÜLER and S. GÜNDÜZ, "Deep Learning Based Fake News Detection on Social Media," *International Journal of Information Security Science*, vol. 12, no. 2, 2023, doi: 10.55859/ijiss.1231423.
[10] A. Jarrahi and L. Safari, "Evaluating the effectiveness of publishers' features in fake news detection on

social media," *Multimed Tools Appl*, vol. 82, no. 2, 2023, doi: 10.1007/s11042-022-12668-8.

[11] G. Yildirim, "A novel hybrid multi-thread metaheuristic approach for fake news detection in social media," *Applied Intelligence*, vol. 53, no. 9, 2023, doi: 10.1007/s10489-022-03972-9.

[12] Y. Y. Chen, S.-P. Yong, and A. Ishak, "Email Hoax Detection System Using Levenshtein Distance Method," *J Comput (Taipei)*, 2014, doi: 10.4304/jcp.9.2.441-446.

[13] A. B. Prasetijo, R. R. Isnanto, D. Eridani, Y. A. A. Soetrisno, M. Arfan, and A. Sofwan, "Hoax detection system on Indonesian news sites based on text classification using SVM and SGD," in *Proceedings - 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2017*, 2018. doi: 10.1109/ICITACEE.2017.8257673.

[14] V. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News," 2016. doi: 10.18653/v1/w16-0802.

[15] F. E. Nastiti, R. A. Prastyanti, R. B. Taruno, and D. Hariyadi, "Social media warfare in Indonesia political campaign: A survey," in *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018*, 2019. doi: 10.1109/ICITISEE.2018.8720959.