JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. 9, No. 2, August 2026, pp. 123-131 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v9i2.11650 e-ISSN: 2656-1948

COMPARATIVE ANALYSIS OF SUPPORT VECTOR MACHINE AND RANDOM FOREST METHODS BASED ON RANDOMIZED SEARCH OPTIMIZATION IN HOAX NEWS CLASSIFICATION

Vincent Lawrence[1*], Frans Mikael Sinaga[2]

[1,2] Informatics, Faculty of Computer Science, Universitas Pelita Harapan, Medan, 20112, Indonesia
[*]Email: [1]03082210041@student.uph.edu, [2]frans.sinaga@lecturer.uph.edu

(Received: 31 March 2026, Revised: 09 April 2026, Accepted: 24 April 2026)

Abstract

In today's digital era, the spread of hoax news has increasingly escalated alongside the ease of access to information through social media and online news portals. This phenomenon has caused negative impacts such as public confusion, social conflict, and a decline in public trust toward information accuracy. Therefore, an effective classification method is needed to accurately detect hoax news. This study aims to analyze and compare the performance of the Support Vector Machine (SVM) and Random Forest (RF) algorithms in classifying hoax news, by applying the Randomized Search optimization technique to enhance model performance. The dataset used in this research was obtained from Kaggle, titled Indonesia Fact and Hoax Political News, consisting of news titles and narratives as input attributes, and hoax or factual labels as outputs. The results show that the SVM algorithm without optimization achieved an accuracy of 83.92%, which increased to 84.28% after optimization using Randomized Search. Meanwhile, the Random Forest algorithm without optimization achieved an accuracy of 85.93%, which increased to 86.05% after optimization. Based on these findings, it can be concluded that the application of Randomized Search successfully improved the accuracy, sensitivity, and stability of the classification models, with the Random Forest algorithm providing the best performance in detecting hoax news in this study.

Keywords: *Hoax News Classification, SVM, Random Forest, Randomized Search, Text Mining*

**Corresponding Author: Vincent Lawrence*

1. INTRODUCTION

In the rapidly evolving digital era, information can easily spread through various social media platforms and online news websites [1]. However, this ease of access to information also brings new challenges, namely the spread of hoax news [2]. A hoax is information that is engineered to conceal the actual facts. In other words, hoaxes can also be interpreted as an attempt to distort facts using information that appears convincing but cannot be verified for its truth [3]. Hoax news has impacts on society, including confusion, social conflict, and a decline in public trust [4]. Therefore, an effective method is needed to accurately classify hoax news so that it can be used as a preventive measure against the spread of misleading information in society.

In this context, text mining techniques are one of the approaches that can be used to analyze and detect hoax news. Several previous studies have examined the effectiveness of classification algorithms in detecting hoax news. For example, a study comparing the performance of the Naïve Bayes and Support Vector Machine (SVM) methods in detecting hoax news in online news in Indonesia showed that SVM achieved the highest accuracy in classifying hoax news [5]. Another study used Random Forest for hoax news classification and obtained fairly good results, indicating its effectiveness in detecting hoaxes even with imbalanced datasets [6].

However, these studies have not yet explored the use of hyperparameter optimization techniques to improve model performance [7][8]. In this regard, Randomized Search is one of the optimization methods that is quite efficient because it can randomly select parameter combinations within a wide parameter space, thereby accelerating the process of finding the best parameters compared to Grid Search [9]. Previous research supports this finding, showing that the use of Randomized Search in tuning SVM parameters is faster and tends to be

Table 1. Research Dataset

ID	Label	Date	Title	Narrative	Image File Name
71	1	17-Aug-20	*Pemakaian Masker Menyebabkan Penyakit Legionnaires*	A caller to a radio talk show recently shared that his wife was hospitalized n told she had COVID n ...	71.jpg
461	1	17-Jul-20	*Instruksi Gubernur Jateng tentang penilangan bagi yg tidak bermasker di muka umum Rp.150.000 menggu...*	*Yth.Seluruh Anggota Grup Sesuai Instruksi Gubernur Jawa Tengah Hasil Rapat Tim Gugus Tugas Covid 19 ...*	461.png

more efficient in finding near-optimal results compared to Grid Search [10].

Therefore, this study aims to conduct a comparative analysis of the performance of Support Vector Machine and Random Forest methods in hoax news classification by applying Randomized Search optimization to improve the accuracy and efficiency of the classification model.

2. RESEARCH METHOD

This section describes the research methodology used in this study to classify hoax news. The overall research stages are illustrated in Figure 1.

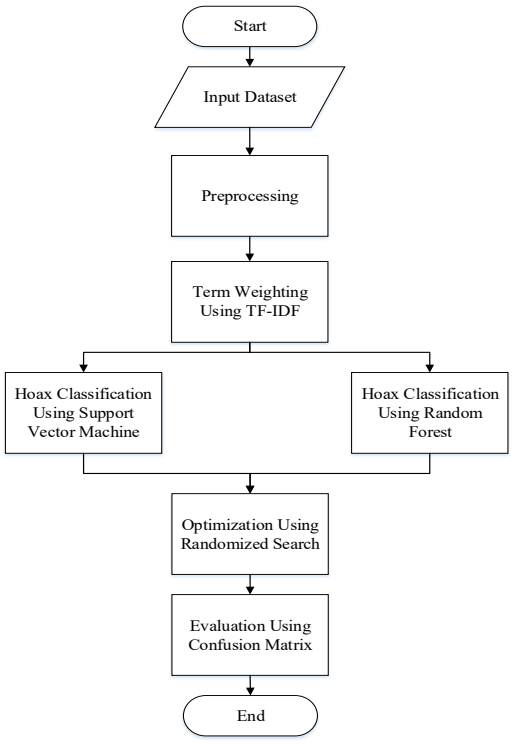

Figure 1. Research Method

2.1 Input Dataset

This stage involves the process of inputting the dataset to be used in this research. The dataset utilized in this study was obtained from Kaggle – Indonesia False News Dataset, available at https://www.kaggle.com/datasets/muhammadghazimuharam/indonesiafalsenews. Table 1 below presents sample attributes of the dataset used in this study.

From Table 1, it can be observed that the dataset contains several attributes; however, this study utilizes only the label, title, and narrative attributes. Subsequently, a preprocessing stage is conducted to prepare the data for the next phase of analysis.

2.2 Preprocessing

Preprocessing is conducted to ensure that the text data is clean and ready to be processed in machine learning models. In this study, the preprocessing stage consists of several steps, including lowercasing to convert all characters into lowercase, removing punctuation to eliminate symbols such as periods, commas, and question marks, tokenizing to split sentences into individual words, stopword removal to eliminate common words that do not carry significant meaning, and stemming to reduce words to their root forms [11].

2.3 Term Weighting Using TF-IDF

After preprocessing is performed on the news dataset, the next stage is to apply term weighting using TF-IDF so that hoax news classification can be carried out using Support Vector Machine and Random Forest. The TF-IDF method is one of the most popular techniques for determining the weight of each word. TF-IDF assigns weights to words in text documents, reflecting their level of importance within a document. The purpose of the TF-IDF method is to represent the significance of each term in a document for further classification processes. Data that has passed the preprocessing stage is then processed using term weighting with the TF-IDF method [12]. The stages in determining the weighting values in the TF-IDF method are as follows [13]:

- The calculation of Term Frequency (TF) in a text document assumes that each term has a proportional level of importance within the document. The TF formula is:

$$\mathrm{TF}(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (1)$$

where $f_{t,d}$ is the frequency of term ttt in document ddd, and $\sum_k f_{k,d}$ is the total number of all terms in document ddd.

- Furthermore, the rarer a term appears across documents, the higher its Inverse Document Frequency (IDF) value. The IDF formula is:

$$\mathrm{IDF(t)} = \log(\frac{N}{1+n_t}) \quad (2)$$

where N is the total number of documents, and n_t is the number of documents containing term t. The addition of 1 in the denominator is used to avoid division by zero.

- The IDF formula is: TF-IDF is the result of the multiplication between TF and IDF, formulated as:

$$\mathrm{TF-IDF(t,d)} = \mathrm{TF(t,d)xIDF(t)} \quad (3)$$

2.3 Hoax Classification

Hoax classification is performed using two algorithms, namely Support Vector Machine and Random Forest.

Support Vector Machine (SVM) is a supervised learning algorithm that works by finding the optimal hyperplane that separates data into different classes [7]. In this study, SVM is used to classify hoax and factual news based on TF-IDF feature representations. SVM is effective in handling high-dimensional data and is capable of producing good generalization performance. The model is trained using labeled data and optimized using Randomized Search to determine the best combination of hyperparameters, such as kernel type, C (regularization parameter), and gamma, in order to improve classification accuracy [14].

Random Forest is an ensemble learning method that consists of multiple decision trees built using random subsets of data and features. Each tree in the forest produces a classification result, and the final prediction is determined based on the majority vote of all trees. In this study, Random Forest is applied to classify hoax news using TF-IDF features. This method is known for its robustness against overfitting and its ability to handle imbalanced datasets. Similar to SVM, the Random Forest model is also optimized using Randomized Search to find the optimal hyperparameters, such as the number of trees (n_estimators), maximum depth, and minimum samples split, to enhance model performance [15].

2.4 Optimized Using Randomized Search

After performing classification using the Support Vector Machine (SVM) and Random Forest (RF) algorithms, the next stage is to conduct hyperparameter optimization to obtain the best combination of parameters that yields the highest accuracy on the training data. The optimization process is carried out using Randomized Search, which is a method that randomly searches combinations of parameters within a specified number of iterations.

The parameters optimized in the SVM algorithm include C, kernel, and gamma. The parameter C acts as a regularization parameter that controls the balance between maximizing the margin and minimizing classification errors. Smaller values of C result in a wider margin but allow more misclassification (underfitting), while larger values aim to classify all data correctly but may lead to overfitting. The kernel function is used to map data into a higher-dimensional space, with several types applied in this study, including linear, radial basis function (RBF), polynomial (poly), and sigmoid, each suitable for different data patterns. Meanwhile, gamma controls the influence range of a data point, particularly in non-linear kernels such as RBF and polynomial; higher values focus more on nearby points, while lower values provide a broader influence. The optimization process is carried out using Randomized Search, where a number of parameter combinations are randomly selected from the defined distributions within a specified number of iterations. For each combination, the SVM model is trained using the training data and evaluated using k-fold cross-validation. The performance of each combination is measured using evaluation metrics such as mean accuracy, and the combination with the highest validation accuracy is selected as the best parameter configuration [16].

The parameters optimized in the Random Forest algorithm include n_estimators, max_depth, min_samples_split, min_samples_leaf, and max_features. The parameter n_estimators determines the number of decision trees in the forest, where a higher number generally improves prediction stability but increases computational time. The max_depth parameter controls the maximum depth of each tree; smaller values help prevent overfitting, while larger values allow the model to capture more complex patterns. The min_samples_split defines the minimum number of samples required to split an internal node, where larger values produce simpler trees. Similarly, min_samples_leaf specifies the minimum number of samples in each leaf node, which helps reduce model complexity and improve generalization. The max_features parameter determines the number of features considered at each split, with options such as sqrt, log2, or using all features. The optimization process is performed using Randomized Search by randomly selecting combinations of these parameters over a predefined number of iterations. Each combination is evaluated using k-fold cross-validation, and its performance is measured using metrics such as mean accuracy. The parameter combination that yields the highest validation performance is selected as the optimal configuration [17].

2.5 Evaluation using Confusion Matrix

The evaluation stage in this study is carried out using a Confusion Matrix to measure the performance of the classification models. The Confusion Matrix is used to compare the predicted labels generated by the model with the actual labels in the dataset, which consist of two classes: hoax and factual news [18].

The dataset is divided into training and testing data with a ratio of 80:20, where the training data is

used to build the model and the testing data is used to evaluate its performance. The evaluation process is applied to both Support Vector Machine (SVM) and Random Forest (RF) algorithms, before and after applying Randomized Search optimization.

Based on the Confusion Matrix, several evaluation metrics are calculated, including accuracy, precision, recall, and misclassification error. Accuracy measures the overall correctness of the model, precision indicates the proportion of correctly predicted positive instances, recall measures the model's ability to identify all relevant instances, and misclassification error represents the proportion of incorrect predictions [19].

These evaluation metrics are used to assess and compare the performance of each classification model in detecting hoax news.

3. RESULT AND DISCUSSION

In this subsection, the results of testing and evaluation of the developed system are presented. The presentation focuses on analyzing the model's performance in the hoax news classification process. The testing is conducted using two algorithms, namely Support Vector Machine and Random Forest with a Randomized Search optimization approach. The evaluation aims to assess the performance of each algorithm based on relevant evaluation metrics.

3.1 Results

Testing was conducted on the Support Vector Machine and Random Forest algorithms using a Randomized Search optimization approach for hoax news classification. The total dataset consisted of 4,231 posts, with an 80:20 split, where 80% (3,385 posts) were used as training data and 20% (846 posts) were used as testing data. The evaluation using a Confusion Matrix aims to assess the performance of the classification models (Support Vector Machine and Random Forest). The Confusion Matrix is constructed based on the prediction results and the actual labels, as shown in Table 2.1, where the labels consist of two classes: hoax and fact. Testing was first carried out on the Support Vector Machine algorithm. Figure 2 shows the Confusion Matrix plot of the Support Vector Machine algorithm.

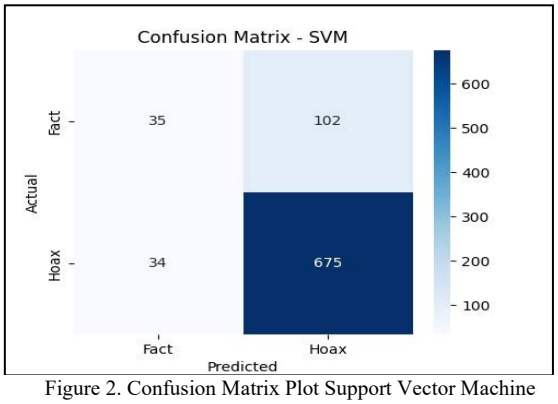

Figure 2. Confusion Matrix Plot Support Vector Machine Algorithm

The performance testing of the Support Vector Machine classification model based on Figure 2 is as follows:

$$\text{Accuracy} = \frac{675+35}{675+35+102+34} = \frac{710}{846} \text{ x } 100\% = 83.92\%$$

$$\text{Precision} = \frac{675}{675+102}\ x\ 100\% = 86.87\%$$

$$\text{Recall} = \frac{675}{675+34}\ x\ 100\% = 95.20\%$$

$$\text{Misclassification Error} = \frac{102+34}{846} \text{ x } 100\% = 16.08\%$$

Based on the results of the Confusion Matrix test on the Support Vector Machine algorithm, the accuracy value obtained was 83.92%, precision was 86.87%, recall was 95.20%, and misclassification error was 16.08%.

Furthermore, to improve the performance of the Support Vector Machine algorithm, testing was carried out by adding Randomized Search optimization. In the optimization stage using Randomized Search, several important parameters in the Support Vector Machine (SVM) algorithm were tested to find the best combination that could improve model performance. The first parameter is C, which is a regularization parameter that functions to control the balance between classification errors in training data and the separation margin between classes. A large C value makes the model attempt to classify all training data correctly, but risks overfitting, while a C value that is too small can produce a wider margin but increases classification errors. The second parameter is gamma, which is used in non-linear kernels such as RBF, and determines how much influence one data item has on another. A high gamma value makes the decision boundary more complex, while a value that is too small causes the model to be too simple. In addition, kernel parameters are also optimized to determine the type of kernel function that best fits the data distribution, such as linear, polynomial (poly), radial basis function (rbf), or sigmoid. By randomly testing the combination of the three parameters, the optimal parameter configuration was obtained so that the SVM model can produce better and more stable classification performance. Figure 3 shows the structure of the Confusion Matrix Plot of the Support Vector Machine algorithm based on Randomized Search optimization.

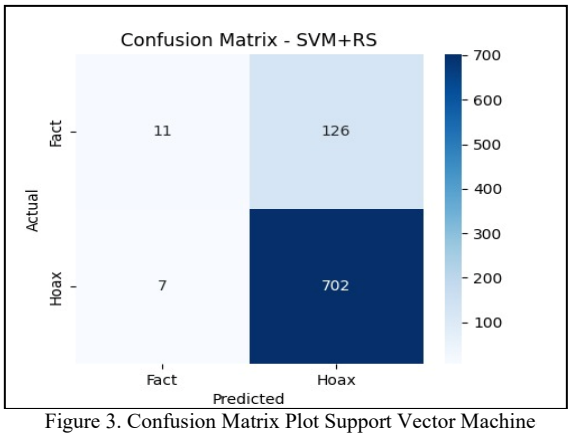

Figure 3. Confusion Matrix Plot Support Vector Machine Algorithm Based on Randomized Search Optimization

Performance testing of the Support Vector Machine algorithm classification model based on Randomized Search optimization based on Figure 3 includes:

$$\text{Accuracy} = \frac{702+11}{702+11+126+7}\ x\ 100\% = 84.28\%$$

$$\text{Precision} = \frac{702}{702+126}\ x\ 100\% = 84.78\%$$

$$\text{Recall} = \frac{702}{702+7}\ x\ 100\ \% = 99.01\%$$

$$\text{Misclassification Error} = \frac{126+7}{846}\ x\ 100\% = 15.72\%$$

Based on the results of the Confusion Matrix test on the Support Vector Machine algorithm based on Randomized Search optimization, the accuracy value was 84.28%, precision 84.78%, recall 99.01%, and misclassification error 15.72%.

Next, testing was carried out on the Random Forest algorithm. Figure 4 shows the Confusion Matrix Plot structure of the Random Forest algorithm.

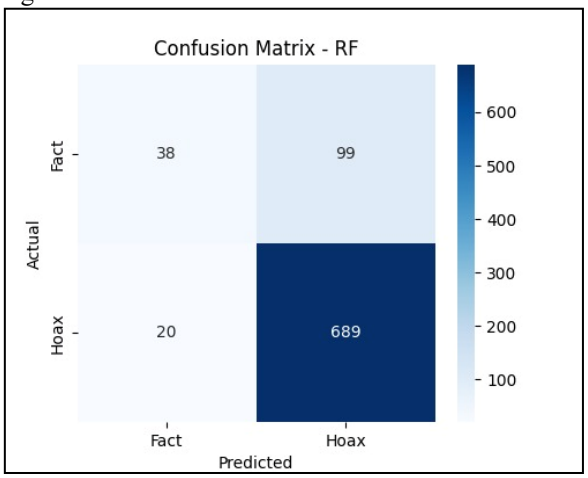

Figure 4. Confusion Matrix Plot Random Forest Algorithm

The performance testing of the Random Forest classification model based on Figure 4 is as follows

$$\text{Accuracy} = \frac{689+38}{689+38+99+20}\ x\ 100\% = 85.93\%$$

$$\text{Precision} = \frac{689}{689+99}\ x\ 100\% = 87.44\%$$

$$\text{Recall} = \frac{689}{689+20}\ x\ 100\% = 97.18\%$$

$$\text{Misclassification Error} = \frac{99+20}{846}\ x\ 100\% = 14.07\%$$

Based on the results of the Confusion Matrix test on the Random Forest algorithm, the accuracy value obtained was 83.92%, precision was 86.87%, recall was 95.20%, and misclassification error was 16.08%.

Next, to improve the performance of the Random Forest algorithm, testing was carried out by adding Randomized Search optimization. In the optimization stage using Randomized Search, several important parameters in the Random Forest (SVM) algorithm were tested to find the best combination that could improve model performance. The first parameter optimized was n_estimators, which is the number of decision trees used in the ensemble; the more trees used, the accuracy generally increases, but the computation time also increases. The second parameter is max_depth, which sets the maximum depth of each tree; too large a value can lead to overfitting, while too small a value can make the model less able to capture complex patterns in the data. Next, the min_samples_split and min_samples_leaf parameters are used to control the minimum number of samples needed to split a node and the minimum number of samples per leaf of the tree, both of which play a crucial role in avoiding the formation of overly complex trees. Other parameters such as max_features were also tested to determine the number of features considered at each node split, thus influencing the balance between model bias and variance. By randomly combining these parameters, the optimization process becomes more efficient than a full grid search, as only a subset of parameter combinations are tested. Through this approach, the Random Forest model is expected to achieve more optimal classification performance with a good level of generalization to new data. Figure 5 shows the structure of the Confusion Matrix Plot of the Random Forest algorithm. based on Randomized Search optimization.

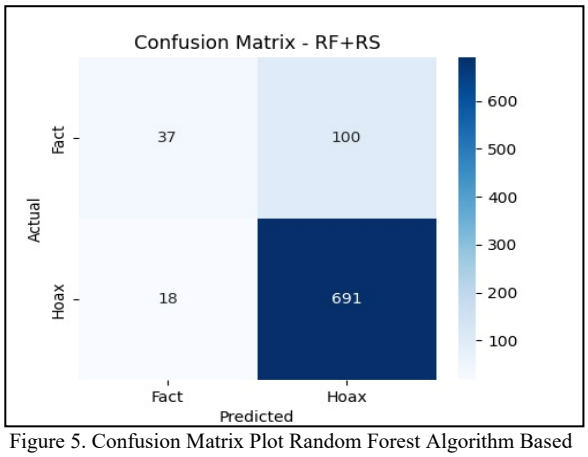

Figure 5. Confusion Matrix Plot Random Forest Algorithm Based on Randomized Search Optimization

Performance testing of the Support Vector Machine algorithm classification model based on Randomized Search optimization based on Figure 3 includes:

$$\text{Accuracy} = \frac{691+37}{691+37+100+18}\ x\ 100\%\ 86.05\%$$

$$\text{Precision} = \frac{691}{691+100}\ x\ 100\%\ 87.36\%$$

$$\text{Recall} = \frac{691}{691+18}\ x\ 100\% = 97.46\%$$

$$\text{Misclassification Error} = \frac{100+18}{846}\ x\ 100\% = 13.95\%$$

Based on the results of the Confusion Matrix test on the Random Forest algorithm based on Randomized Search optimization, the accuracy value was 86.05%, precision 87.36%, recall 97.46%, and misclassification error 13.95%.

A summary of the results of testing the algorithm's performance in classifying hoax news is shown in Table 2 and Figure 6.

both algorithms. In general, both SVM and Random Forest demonstrate good classification capabilities in detecting hoax news, although their performance varies after hyperparameter tuning.

For the Support Vector Machine (SVM) algorithm, the initial accuracy was 83.92%, with a precision of 86.87%, a recall of 95.20%, and a misclassification error rate of 16.08%. These results indicate that the model can recognize most of the positive data effectively. However, after hyperparameter optimization using Randomized Search, the model's accuracy increased to 84.28%, and recall improved significantly to 99.01%. Although precision decreased slightly to 84.78%, the misclassification rate also dropped to 15.72%. This indicates that the optimized model has a better ability to detect hoax news overall (high recall), albeit with a slight trade-off in precision when classifying true news versus hoax news. This improvement demonstrates that selecting more appropriate parameters can increase the model's sensitivity to hoax news patterns in the dataset.

Meanwhile, the Random Forest (RF) algorithm exhibited more stable performance and generally outperformed SVM. The Random Forest model without optimization achieved 85.93% accuracy, 87.44% precision, and 87.44% recall, with a misclassification error rate of 14.07%. After optimization using Randomized Search, the model's performance increased slightly to 86.05% accuracy, 87.36% precision, and 87.36% recall, while the misclassification error decreased to 13.95%. This improvement shows that parameter tuning can

Table 2. Summary of Algorithm Testing Results

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	Misclassification Error (%)
Support Vector Machine (SVM)	83.92	86.87	95.20	16.08
SVM + Randomized Search	84.28	84.78	99.01	15.72
Random Forest (RF)	85.93	87.44	97.18	14.07
RF + Randomized Search	86.05	87.36	97.46	13.95

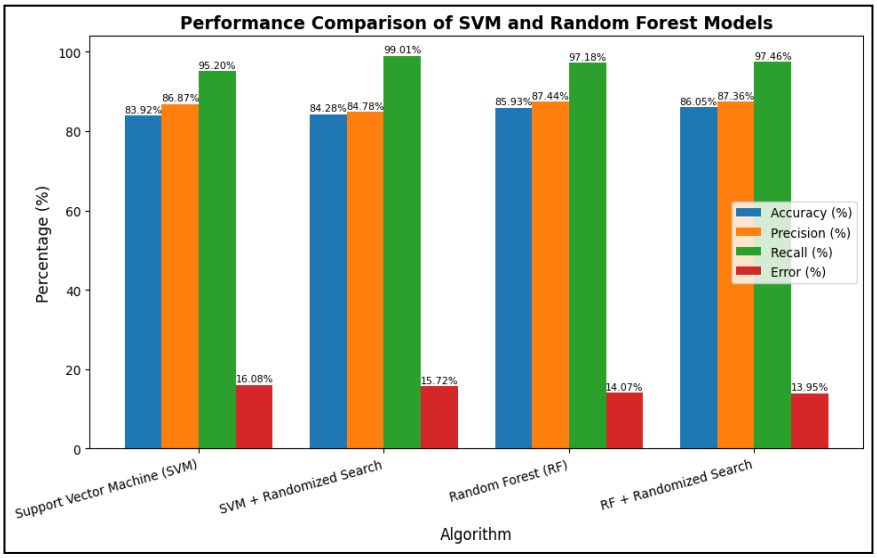

Figure 6. Bar Graph Summary of Algorithm Testing Results

3.2 Discussion

Based on the test results of the classification models, namely Support Vector Machine (SVM) and Random Forest (RF), using Randomized Search optimization, it can be observed that the optimization process significantly improves the performance of

enhance the model's generalization ability without sacrificing the balance between precision and recall. Random Forest has proven to be more consistent in providing balanced results, due to its nature of combining multiple decision trees to reduce variance and prevent overfitting.

Several previous studies have emphasized the importance of hyperparameter optimization techniques in improving model performance. For instance, a study by Rahmadayana et al. (2021) applied Randomized Search optimization on SVM for sentiment analysis and reported that tuning hyperparameters combined with appropriate preprocessing significantly improved model performance, particularly in terms of F1-score [16]. In another study, Fajri et al. (2023) demonstrated that both Random Search and Grid Search are effective in identifying optimal parameter combinations for SVM, leading to improved classification performance across multiple datasets [10]. Furthermore, research by Tripathy et al. (2025) showed that hyperparameter optimization using search-based methods can increase SVM accuracy from 98.08% to 99.12%, highlighting the significant impact of proper tuning on model performance [20]. However, when applied to the hoax news classification case in this study, the improvement achieved through Randomized Search tends to be relatively small. This indicates that although optimization methods are generally effective in enhancing performance, their impact depends on the initial model performance and dataset characteristics. When the baseline model already performs well, optimization tends to yield only marginal improvements, which is consistent with the findings of previous studies.

Overall, when comparing the models, Random Forest with Randomized Search optimization provided the best results among the four tested models, both in terms of accuracy and the balance between precision and recall. On the other hand, SVM with Randomized Search achieved the highest recall, indicating that this model is very sensitive in detecting hoax news, although with a slight reduction in precision. Therefore, the choice of the best model depends on the priorities of the classification system—whether it is more important to detect as many hoax news items as possible (high recall) or to produce more accurate predictions with minimal errors (high precision).

The confusion matrix indicates that a large portion of instances are classified as hoax, which can be attributed to several underlying factors related to the dataset and model behavior. One primary cause is the potential imbalance or distribution tendency within the dataset, where certain patterns or features associated with hoax news are more dominant and easier for the model to learn compared to factual news. In addition, the use of text representation methods such as TF-IDF may emphasize distinctive terms that frequently appear in hoax-related content, leading the model to be more sensitive toward detecting hoax patterns. This is further reinforced by the optimization process, particularly in SVM, where the significant increase in recall suggests that the model prioritizes minimizing false negatives (i.e., missing hoax news), even at the expense of slightly higher false positives. From a practical perspective, this behavior is still acceptable in hoax detection systems, as it is generally more critical to correctly identify as many hoax instances as possible rather than risking undetected misinformation. Therefore, the tendency of the model to classify more instances as hoax reflects its sensitivity to the target class and is influenced by both dataset characteristics and the optimization objective.

The results of this study indicate that the application of hyperparameter optimization using Randomized Search on Support Vector Machine (SVM) and Random Forest (RF) leads to relatively limited performance improvements, particularly in terms of accuracy. This can be explained by the concept of diminishing returns, where models with strong baseline performance tend to show only marginal gains after further optimization. In this context, small improvements are considered reasonable, as the models are already near their optimal performance. Random Forest demonstrates strong stability and generalization even with default parameters due to its ensemble mechanism, while SVM shows a more noticeable improvement in recall, indicating higher sensitivity in detecting specific classes, although this does not always significantly improve overall accuracy. These findings are consistent with previous research on water quality classification using the Water Quality and Potability dataset, where Random Search optimization on Random Forest only increased accuracy slightly from 84.12% to 84.38%, despite higher computational cost [17].

Furthermore, regarding the reviewer's suggestion to incorporate feature selection, experimental results in this study show that reducing features actually decreases model performance, indicating that all features contribute meaningfully to the classification process. This suggests that feature selection does not always guarantee performance improvement, especially in datasets with rich and complex feature representations. Therefore, the use of Randomized Search remains relevant, provided that the trade-off between performance improvement and computational cost is carefully considered. In conclusion, hyperparameter optimization using Randomized Search is effective in improving machine learning model performance [21], while ensemble methods such as Random Forest offer advantages in stability and consistent accuracy, making them suitable for developing reliable hoax news classification systems.

4. CONCLUSION

Based on the study results, the Support Vector Machine (SVM) without optimization achieved 83.92% accuracy, 86.87% precision, 95.20% recall, and 16.08% misclassification error, while SVM optimized with Randomized Search improved performance to 84.28% accuracy, 99.01% recall, and 15.72% misclassification error, with a slight decrease in precision to 84.78%. Meanwhile, Random Forest (RF) without optimization reached 85.93% accuracy, 87.44% precision, 97.18% recall, and 14.07% misclassification error, and RF optimized with Randomized Search achieved the best performance with 86.05% accuracy, 87.36% precision, 97.46% recall, and 13.95% misclassification error. These findings demonstrate that hyperparameter optimization is effective and capable of improving model sensitivity, accuracy, and stability in hoax news detection. For future research, it is recommended to explore alternative optimization methods such as Grid Search and metaheuristic approaches (e.g., Genetic Algorithm, Ant Colony Optimization) to investigate the potential for achieving more significant performance improvements. Additionally, expanding the dataset from multiple platforms and incorporating additional features or deep learning approaches such as BERT may further enhance model performance.

5. REFERENCES

[1] S. Rahmatullah and R. E. Dwi Yuliati, "Media Sosial Sebagai Sumber Berita Altenatif," *J. Stud. Jurnalistik*, vol. 4, no. 2, pp. 47–54, 2022, doi: 10.15408/jsj.v4i2.28966.

[2] Ubaidillah, Taufik, and Suharto, "Analisis Dampak Media Sosial Dalam Menyebarkan Informasi Berita Kampus oleh Humas UIN Datokarama Palu," *J. UIN Datokarama Palu*, vol. 3, no. 1, pp. 48–62, 2024.

[3] A. R. Hanum *et al.*, "Analisis Kinerja Algoritma Klasifikasi Teks BERT Dalam Mendeteksi Berita Hoaks," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 3, pp. 537–546, 2024, doi: 10.25126/jtiik938093.

[4] Y. D. Butar, "Analisis Penyebaran Hoax Di Media Sosial Dan Dampaknya Terhadap Masyarakat Masyarakat," *JPBB J. Pendidikan, Bhs. dan Budaya*, vol. 3, no. 2, pp. 252–258, 2024.

[5] R. Astuti and A. T. Sipahutar, "Perbandingan Klasifikasi Berita Hoax Politik Pada Media Sosial X Menggunakan Algoritma Naive Bayes dan Random Forest," *J. Ilmu Komputer, Sist. Inf. dan Teknol. Inf.*, vol. 2, no. 1, pp. 61–67, 2025.

[6] D. Setyadin, R. H. Winasis, and G. Triyono, "Deteksi Berita Palsu menggunakan Algoritma Random Forest," *Sist. J. Sist. Inf.*, vol. 14, no. 3, pp. 1142–1153, 2025.

[7] S. Angelina, S. D. Siregar, A. Ridwan, and L. Francolim, "Comparative Analysis of Ensemble Classification Models and Support Vector Machines in Measuring Stress Levels Based on EEG Signals," *JIKO (Jurnal Inform. dan Komputer)*, vol. 9, no. 1, 2026.

[8] Andi, R. O. Ong, Thamrin, and Roseline, "Ensemble Algoritma Random Forest Classifier dan AdaBoost Dalam Perancangan Aplikasi Penerjemah Bahasa Isyarat," *J. TIMES*, vol. 14, no. 1, pp. 51–59, 2025.

[9] A. D. Rachmatsyah, T. Sugihartono, and K. Irfan, "Perbandingan Teknik Optimasi Grid Search dan Randomized Search Dalam Meningkatkan Akurasi Metode Klasifikasi SVM Pada Sentimen Ulasan Pengguna Aplikasi JKN Mobile," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, no. 1, pp. 13–22, 2024, doi: 10.36080/skanika.v8i1.3328.

[10] M. Fajri and A. Primajaya, "Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 14–19, 2023, doi: 10.30871/jaic.v7i1.5004.

[11] Andi, Thamrin, A. Susanto, E. Wijaya, and D. Djohan, "Analysis of the random forest and grid search algorithms in early detection of diabetes mellitus disease," *J. Mantik*, vol. 7, no. 2, pp. 1117–1124, 2023, doi: 10.35335/mantik.v7i2.3981.

[12] I. S. Wibowo, A. Witanti, and I. Susilawati, "Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF," *J. Tek. Inform. dan Sist. Inf.*, vol. 11, no. 1, pp. 99–111, 2024, [Online]. Available: http://jurnal.mdp.ac.id.

[13] M. D. Afandi, A. Homaidi, A. Ghofur, and A. Zubairi, "Penerapan Information Retrieval dalam Sistem Analisis Kemiripan Proposal Skripsi menggunakan Cosine Similarity," *Swabumi*, vol. 12, no. 1, pp. 39–46, 2024.

[14] I. N. P. Trisna, I. M. W. J. Putra, and W. O. Vihikan, "Classifying Indonesian Hoax News Titles with SVM, XGBoost, and BiLSTM," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 19, no. 4, pp. 1–10, 2025, doi: 10.22146/ijccs.xxxx.

[15] T. Tambunan, M. Yohanna, and A. P. Silalahi, "Penerapan Metode Random Forest Dalam Mendeteksi Berita Hoax," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 7, no. 2, pp. 301–306, 2023.

[16] F. Rahmadayana and Y. Sibaroni, "Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 936–942, 2021.

[17] A. F. Firmansyah, B. Rahmat, and M. M. A. Haromainy, "Optimization of the Random

Forest Algorithm Using Random Search for Potable Water Quality Classification," *J. Artif. Intell. Eng. Appl.*, vol. 5, no. 1, pp. 19–23, 2025.

[18] A. Andi, C. Juliandy, and D. David, "Clustering Analysis of Tweets About COVID-19 Using the K-Means Algorithm," *Sinkron*, vol. 8, no. 1, pp. 543–533, 2023, doi: 10.33395/sinkron.v8i1.12145.

[19] Mubarak, L. Tanti, and R. Rosnelly, "Perbandingan Algoritma Decision Tree dan Naive Bayes Pada Analisis Sentimen Masyarakat Terhadap Pejabat Pertamina Pasca Kasus Pertamax Oplosan," *J. Minfo Polgan*, vol. 15, no. 1, pp. 179–188, 2026.

[20] S. S. Tripathy and B. Behera, "Hyperparameter Tuning-Based Optimized Performance Analysis of Machine Learning Algorithms for Network Intrusion Detection," *Int. J. Netw. Secur. Appl.*, vol. 17, no. 5/6, pp. 1–26, 2025.

[21] M. Sholeh, U. Lestari, and D. Andayati, "Hyperparameter Optimization Using Grid Search and Random Search to Improve the Performance of Prediction Models with Decision Trees," *J. Ris. Multidisiplin dan Inov. Teknol.*, vol. 3, no. 03, pp. 453–464, 2025.