Improvement_of_Machine_Learning_Algorithms_with.pdf

by dppm7 uir

Improvement of Machine Learning Algorithms with Hyperparameter Tuning on Various Datasets

1st Akmar Efendi
department of Informatics Engineering
Universitas Islam Riau
Pekanbaru, Indonesia
akmarefendi@eng.uir.ac.id

2nd Iskandar Fitri
Department of Information Technology
Universitas Putra Indonesia YPTK
Padang, Indonesia
if@upiyptk.ac.id

3rd Gunadi Widi Nurcahyo
Department of Information Technology
Universitas Putra Indonesia YPTK
Padang, Indonesia
gunadiwidi@yahoo.co.id

***Abstract*—In the digital era with a data explosion, classification techniques have become a primary aspect of machine learning, especially in Supervised Learning methods. These techniques allow computers to learn from existing data and apply their knowledge to classify new data based on patterns found in the training data. Although algorithms such as Support Vector Machine (SVM) and Naïve Bayes are reliable in many cases, they are not always optimal due to data complexity. This study evaluates the performance of various models and applies optimization techniques to enhance model performance across different datasets. The study uses three different datasets: academic data from the Faculty of Engineering at Universitas Islam Riau (UIR), tweet data from the social media platform X, and diabetes disease data from Kaggle. Each model is tested with a 70:30 data split, employing techniques such as SMOTE, Hyperparameter Optimization with Optuna, and XGBoost to improve model performance. The combination of SMOTE with SVM or GNB shows significant improvement in accuracy, precision, recall, and F1-Score when optimization techniques are applied. For instance, the use of SMOTE, SVM, and Optuna achieves 100% accuracy on academic data, 97% on Twitter data, and 80% on diabetes data. Similarly, the combination of SMOTE, GNB, and XGBoost provides significant improvement. This study concludes that the application of optimization techniques like Optuna and its integration with algorithms such as XGBoost significantly enhance the performance of classification models across various datasets. This opens up opportunities for the development of more advanced and effective classification models in the future and makes a significant contribution to understanding the use of classification algorithms in various practical applications.**

Keywords*—Machine Learning, SVM, GNB, Optuna, XGBoost

I. Introduction

In the digital era filled with an explosion of data, classification techniques have become one of the main aspects of Machine Learning, especially in Supervised Learning methods [1] These techniques allow computers to learn from existing data and then apply their knowledge to distinguish and classify new data based on patterns found in the training data [2]. The main goal is to empower programs to distinguish and classify new observations based on patterns hidden in the existing training data [3]. In the classification process, algorithms learn from past experiences and then apply their insights to classify new data into predefined categories [4]

Although classification algorithms such as Support Vector Machine (SVM) and Naïve Bayes have proven their reliability in many cases, they do not always provide optimal performance in all contexts.

Previous research, such as that conducted by [5], [6], highlights the variations in the performance of SVM and Naïve Bayes, depending on the type of data used. For example, when faced with the PeduliLindungi application dataset on the Google Play Store, SVM provided an accuracy of 80.5%, while Naïve Bayes achieved 76.9%. Then, when these algorithms were applied to the Twitter dataset, SVM accuracy increased to 87% and Naïve Bayes reached 83% [7]. Additionally, both have been used to predict cable TV payments, where Naïve Bayes achieved an impressive accuracy of 96%, while SVM only reached 66% [8].

However, fluctuations in this accuracy are often caused by the complexity of the data faced [9]. Therefore, this research aims to explore various models of Naïve Bayes and SVM algorithms, as well as apply techniques such as SMOTE [10], Hyperparameter Optuna [11], and XGBoost [12] to increase model performance. This study utilizes three different datasets: academic data from the Faculty of Engineering, Islamic University of Riau (UIR), tweet data from social media X, and diabetes disease data from Kaggle. Each model is tested using a 70:30 data split.

The main goal of this research is to develop effective classification models to improve accuracy across various types of datasets. By combining careful experiments with the application of state-of-the-art techniques in Machine Learning, this study aims to provide a deeper understanding of the performance of classification algorithms in diverse data contexts. Through this research, it is also expected to pave the way for the development of more sophisticated and effective classification models in the future, and also make a significant contribution to the understanding of the use of classification algorithms in various practical applications.

979-8-3503-7384-4/24/$31.00 ©2024 IEEE

II. Method

In picture 1, the developed model for testing can be seen.

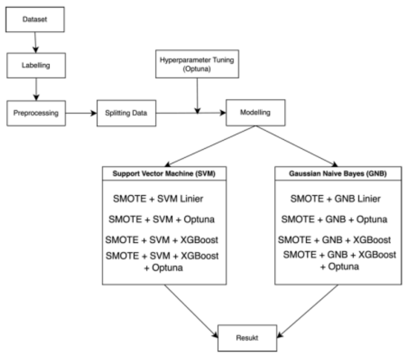

Fig. 1. Model Development

A. Dataset

The dataset in this study consists of 2 datasets. The first dataset is academic data from the Faculty of Engineering, Islamic University of Riau in 2016. Dataset 1 comprises 1282 records of students who have completed their thesis courses. Dataset 2 is sourced from social media X and includes 10,001 records. Dataset 3 is obtained from diabetes disease data on Kaggle, containing 768 records.

B. Labelling

Labeling data in dataset 1 involves 4 categories: DropOut (DO), Graduated On Time, Graduated Late, and Stopped. The labeling process follows these rules:

1. A student is labeled as DropOut (DO) if they exceed the specified time limit of 14 semesters or if they commit a serious violation.
2. A student is labeled as Graduated On Time if they graduate in semester 7, 8, or 9.
3. A student is labeled as Graduated Late if they graduate after 9 semesters.
4. A student is labeled as Stopped if they do not register for a thesis or if they discontinue their studies in a specific semester.

Dataset 2 consists of Twitter data categorized into 3 labels: positive, negative, and neutral sentiments. Dataset 3 is divided into 2 labels: diabetes and non-diabetes. In the labeling process of this study, each label is not balance, so data balancing is performed using the SMOTE method. SMOTE (Synthetic Minority Over-sampling Technique) is an oversampling technique where new minority class instances are generated to match the number of majority class instances [13].

C. Preprocessing

Dataset 1 and 3, after being labeled, the data standardization is taking place using the Standard Scaler. The Standard Scaler is applied to ensure data has a consistent scale and range. Additionally, the Standard Scaler has another benefit which is more stable handling of outliers compared to some other normalization methods. Outliers can heavily influence other normalization methods, but the Standard Scaler relies on mean and standard deviation, which are less affected by extreme values. Meanwhile, Dataset 2 uses data cleaning, case folding, text normalization, tokenizin, filtering, and stemming [14].

D. Model

Then this research conducted several test with various models, table 1 represent the model that constructed in this research

TABLE I. Experiments conducted in this research

No	Researcher	Trial	Accuracy
1	[15]	Support Vector Machine (SVM) Algorithm	85.40%
2	[16]	SVM + Hyperparameter (Random Search)	63.81%
3	[17]	SVM + XGBoost	79%
4	-	SVM + XGBoost + Hyperparameter	-
5	[18]	Gaussian Naive Bayes (GNB) Algorithm	96%
6	[19]	GNB + Hyperparameter (Genetic Algorithm)	93.2%
7	[12]	GNB + XGBoost	81.55%
8	-	GNB + XGBoost + Hyperparameter	-

Table 1 presents the experimental results from some studies applying various combinations of classification algorithms and optimization techniques on different datasets. From the table, it can be seen that research using the Gaussian Naive Bayes (GNB) algorithm achieved the highest accuracy at 96%. However, there is variation in accuracy results among other studies. For instance, using the Support Vector Machine (SVM) algorithm with hyperparameter optimization (Random Search) only achieved an accuracy of 63.81%, which is significantly lower compared to GNB. This finding highlights the importance of selecting appropriate classification algorithms and optimization techniques to achieve optimal accuracy in data classification tasks. However, there are 2 models that have not been found from previous research that combine this, namely SVM + XGBoost + Hyperparameter and GNB + XGBoost + Hyperparameter. This research will use Optuna as the hyperparameter used in this research

Optuna is a library that can automate the process of parameter tuning. Not only that automate this process, but Optuna also includes effective search algorithms that make the search process more efficient [11].

E. Performance Evaluation

The performance evaluation in this study involves assessing various classification models such as SVM and GNB, as well as combinations of optimization techniques like Optuna and XGBoost, across three distinct datasets. Metrics including accuracy, precision, recall, and F1-Score are employed to comprehensively evaluate model effectiveness. Accuracy measures the overall correctness of predictions relative to the total number made. Precision gauges the accuracy of positive predictions, indicating how many correctly predicted positives exist among all positive predictions. Recall assesses the completeness of positive predictions by determining how many actual positive instances were correctly identified by the model. The F1-Score, being the harmonic mean of precision and recall, provides a balanced measure that considers both metrics. This evaluation framework aims to understand each model's capability to handle diverse datasets and identifies optimal combinations of optimization techniques to enhance performance. Additionally, analyzing prediction errors helps uncover scenarios where models struggle and identifies underlying data patterns contributing to these errors. Such insights contribute to refining classification models for future applications across varied data contexts.

III. RESULT AND DISCUSSION

The results of the research are as follows:

TABLE II. SVM DATA AKADEMIK

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + SVM Linier	91%	91%	91%	91%
SMOTE + SVM + Optuna	100%	100%	100%	100%
SMOTE + SVM + XGBoost	95%	96%	95%	95%
SMOTE + SVM + Optuna + XGBoost	100%	100%	100%	100%

Table 2 presents the evaluation results of various classification models applied to academic data using the Support Vector Machine (SVM) algorithm. The finding shows that using the SMOTE technique alongside linear SVM achieves an accuracy of 91%, with corresponding precision, recall, and F1-Score all at the same level. Furthermore, combining SMOTE with SVM and using the optimization algorithm Optuna results in perfect accuracy of 100%, along with precision, recall, and F1-Score also reaching 100%. Moreover, employing SMOTE with SVM and XGBoost yields an accuracy of 95%, with precision at 96%, recall at 95%, and F1-Score at 95%. Finally, utilizing a combination of SMOTE, SVM, Optuna, and XGBoost also achieves perfect accuracy of 100%, with optimal precision, recall, and F1-Score rates. From the table, it can be concluded that employing optimization techniques such as Optuna and combining them with the XGBoost algorithm significantly enhances the performance of classification models on this academic dataset.

TABLE III. SVM DATA TWITTER

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + SVM Linier	94%	94%	94%	94%
SMOTE + SVM + Optuna	96%	96%	96%	96%
SMOTE + SVM + XGBoost	93%	93%	93%	93%
SMOTE + SVM + Optuna + XGBoost	97%	97%	97%	97%

Table 3 presents the evaluation results of various classification models applied to data from the Twitter platform using the Support Vector Machine (SVM) algorithm. The results show that using the SMOTE technique alongside linear SVM achieves an accuracy of 94%, with precision, recall, and F1-Score all at the same level. Furthermore, combining SMOTE with SVM and employing the Optuna optimization algorithm increases the accuracy to 96%, with optimal precision, recall, and F1-Score. Next, employing SMOTE with SVM and the XGBoost algorithm results in an accuracy of 93%, with consistent precision, recall, and F1-Score rates. Finally, using a combination of SMOTE, SVM, Optuna, and XGBoost achieves the highest accuracy of 97%, with optimal precision, recall, and F1-Score rates. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm significantly enhances the performance of classification models on Twitter data.

TABLE IV. SVM DATA DIABETES

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + SVM Linier	78%	78%	78%	78%
SMOTE + SVM + Optuna	80%	80%	80%	80%
SMOTE + SVM + XGBoost	73%	73%	73%	73%
SMOTE + SVM + Optuna + XGBoost	77%	78%	77%	77%

Table 4 presents the evaluation results of various classification models applied to diabetes-related data using the Support Vector Machine (SVM) algorithm. The results indicate that using the SMOTE technique alongside linear SVM achieves an accuracy of 78%, with precision, recall, and F1-Score all at the same level. Furthermore, combining SMOTE with SVM and employing the Optuna optimization algorithm increases the accuracy to 80%, with optimal precision, recall, and F1-Score. However, using SMOTE with SVM and the XGBoost algorithm results in slightly lower accuracy at 73%, with consistent precision, recall, and F1-Score rates. Finally, using a combination of SMOTE, SVM, Optuna, and XGBoost achieves an accuracy of 77%, with relatively stable precision, recall, and F1-Score rates. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm improves the performance of classification models on diabetes data, although the accuracy levels vary depending on the combination used.

TABLE V. GNB DATA ACADEMIC

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + GNB	90%	90%	90%	90%
SMOTE + GNB + Optuna	91%	91%	91%	91%
SMOTE + GNB + XGBoost	95%	96%	95%	95%
SMOTE + GNB + Optuna + XGBoost	91%	91%	91%	91%

Table 5 shows the evaluation results of various classification models applied to academic data using the

Gaussian Naive Bayes (GNB) algorithm. From the table, it is observed that using the SMOTE technique alongside GNB achieves an accuracy of 90%, with precision, recall, and F1-Score all at the same level. When combined with the Optuna optimization algorithm, the accuracy slightly increases to 91%, maintaining optimal precision, recall, and F1-Score. Furthermore, employing SMOTE with GNB and the XGBoost algorithm yields the highest accuracy of 95%, with precision at 96% and recall and F1-Score at 95%. However, using a combination of SMOTE, GNB, Optuna, and XGBoost maintains accuracy at 91%, with consistent precision, recall, and F1-Score rates. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm significantly enhances the performance of classification models on academic data. The accuracy levels vary depending on the combination used, highlighting the importance of selecting appropriate techniques for specific datasets.

TABLE VI. GNB DATA TWITTER

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + GNB	68%	74%	69%	67%
SMOTE + GNB + Optuna	50%	77%	50%	44%
SMOTE + GNB + XGBoost	81%	85%	81%	80%
SMOTE + GNB + Optuna + XGBoost	54%	78%	54%	50%

Table 6 depicts the evaluation results of several classification models applied to Twitter data using the Gaussian Naive Bayes (GNB) algorithm. The results show that using the SMOTE technique alongside GNB achieves an accuracy of 68%, with precision at 74%, recall at 69%, and F1-Score at 67%. However, when this combination is supplemented with the Optuna optimization algorithm, the accuracy decreases to 50%, with precision at 77%, recall at 50%, and F1-Score at 44%. Furthermore, employing SMOTE with GNB and the XGBoost algorithm improves accuracy to 81%, with precision at 85%, recall at 81%, and F1-Score at 80%. However, using a combination of SMOTE, GNB, Optuna, and XGBoost results in lower accuracy at 54%, with precision at 78%, recall at 54%, and F1-Score at 50%. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm can enhance the performance of classification models on Twitter data, although there is variation in accuracy levels depending on the combination used.

TABLE VII. GNB DATA DIABETES

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + GNB	75%	75%	75%	75%
SMOTE + GNB + Optuna	73%	74%	73%	73%
SMOTE + GNB + XGBoost	73%	75%	73%	73%
SMOTE + GNB + Optuna + XGBoost	80%	81%	80%	80%

Table 7 presents the evaluation results of several classification models applied to diabetes-related data using the Gaussian Naive Bayes (GNB) algorithm. From the table, it can be seen that using the SMOTE technique alongside GNB achieves an accuracy of 75%, with precision, recall, and F1-Score all at the same level. However, when this combination is supplemented with the Optuna optimization algorithm, the accuracy slightly decreases to 73%, with precision, recall, and F1-Score remaining stable. Similarly, using SMOTE with GNB and the XGBoost algorithm yields similar results, with accuracy and other evaluation metrics remaining relatively stable around 73%. However, the combination of SMOTE, GNB, Optuna, and XGBoost improves accuracy to 80%, with optimal precision, recall, and F1-Score. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm can enhance the performance of classification models on diabetes data, although there is variation in accuracy levels depending on the combination used.

Based on the evaluation results from those tables, it can be concluded that the use of optimization techniques such as Optuna and integration with the XGBoost algorithm consistently improves the performance of the classification model, regardless of the nature and type of data used.

To give more comprehensive analysis, this study include several key aspects. Firstly, performance comparison evaluations of various classification models such as SVM and GNB, as well as combinations of optimization techniques like Optuna and XGBoost, were conducted on three different datasets. Model performance was measured using metrics such as accuracy, precision, recall, and F1-Score to identify which models performed best in the context of each dataset. The findings indicate that the model combining SMOTE, SVM, and Optuna achieved the best results with 100% accuracy on the academic dataset. On the Twitter dataset, the combination of SMOTE, SVM, and Optuna+XGBoost achieved the highest accuracy of 97%, while on the diabetes dataset, the combination of SMOTE, GNB, and Optuna+XGBoost achieved 80% accuracy.

The influence of optimization techniques was also deeply analyzed. Optuna was utilized for hyperparameter optimization, significantly increase the performance of the classification models by helping to find the best parameter sets that maximize accuracy and other evaluation metrics. Additionally, the integration of XGBoost, known for its ability to handle imbalanced data and improve model performance, also showed significant performance improvements in several datasets.

Error analysis was conducted to understand prediction errors in the model, including false positives and false negatives. This analysis helps in comprehending specific conditions where the model fails and identifying patterns or data characteristics that lead to these errors. It is crucial for improving models in the future.

Performance variation based on datasets was also analyzed. The classification models showed the best performance on the academic dataset, achieving the highest accuracy of 100%. Despite the Twitter dataset being more complex and diverse, the optimized models also demonstrated very good performance, indicating that the optimization techniques used are effective in handling varied social media text data. On the diabetes dataset, model performance showed greater variation, with the best model achieving 80% accuracy. This suggests that medical data may require specific approaches and more optimization to achieve optimal performance.

With this much deeper approach, this research not only provides performance evaluation results but also a comprehensive analysis of how and why certain models outperform others. It offers further insights into research directions in the future as well.

IV. CONCLUSION

Overall, the evaluation results from the tables shows the relative performance of various classification models in processing different datasets. Meanwhile, the conclusions provide an overview of the use of optimization techniques and recommendations for future research.

The evaluation results provide concrete information about the accuracy, precision, recall, and F1-Score from each classification model on the datasets used. This helps assess the relative performance of each model and algorithm combination with clear evaluation metrics.

On the other hand, the conclusions provide a summary of these findings and explore further implications. This includes recommendations for future research, such as the development of more advanced optimization techniques or exploring models on more diverse datasets. Conclusions may also include a brief reflection on the key findings and their relevance in wider research contexts.

REFERENCES

[1] R. Pugliese, S. Regondi, and R. Marini, “Machine learning-based approach: Global trends, research directions, and regulatory standpoints,” *Data Science and Management*, vol. 4. KeAi Communications Co., pp. 19–29, Dec. 01, 2021. doi: 10.1016/j.dsm.2021.12.002.

[2] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, pp. 685–695, 2021, doi: 10.1007/s12525-021-00475-2.

[3] I. Nurcahyani and J. W. Lee, “Role of machine learning in resource allocation strategy over vehicular networks: A survey,” *Sensors*, vol. 21, no. 19. MDPI, Oct. 01, 2021. doi: 10.3390/s21196542.

[4] J. S. Harrison, M. A. Josefy, M. Kalm, R. Krause, and C. S. Joseph Harrison, “Using supervised machine learning to scale human-coded data: A method and dataset in the board leadership context,” *Strat Mgmt J*, vol. 44, pp. 1780–1802, 2023, doi: 10.5281/zenodo.7304697.

[5] A. Salma and W. Silfianti, “Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors,” *International Research Journal of Advanced Engineering and Science*, vol. 6, no. 4, pp. 158–162, 2021.

[6] P. T. Putra, A. Anggrawan, and H. Hairani, “Comparison of Machine Learning Methods for Classifying User Satisfaction Opinions of the PeduliLindungi Application,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 3, pp. 431–442, Jun. 2023, doi: 10.30812/matrik.v22i3.2860.

[7] L. D. Utami and S. Masripah, “Comparation Of Classification Algorithm On Sentiment Analysis Of Online Learning Reviews And Distance Education,” *Jurnal Techno Nusa Mandiri*, vol. 18, no. 2, pp. 101–110, 2021, doi: 10.33480/techno.v18i2.2715.

[8] M. E. Lasulika, “Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Accuracy Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel,” *ILKOM Jurnal Ilmiah*, vol. 11, no. 1, pp. 11–16, 2019, doi: 10.33096/ilkom.v11i1.408.11-16.

[9] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput Sci*, vol. 2, no. 3, pp. 1–21, May 2021, doi: 10.1007/s42979-021-00592-x.

[10] M. K. Anam *et al.*, “Sentiment Analysis for Online Learning using The Lexicon-Based Method and The Support Vector Machine Algorithm,” *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, pp. 290–302, 2023, doi: 10.33096/ilkom.v15i2.1590.290-302.

[11] W. A. G. Kodri and S. Hadianti, “Optimization of The Machine Learning Approach using Optuna in Heart Disease Prediction,” *Journal Medical Informatics Technology*, vol. 1, no. 3, pp. 59–64, Sep. 2023, doi: 10.37034/medinftech.v1i3.15.

[12] T. R. Mahesh, V. Vinoth Kumar, V. Muthukumaran, H. K. Shashikala, B. Swapna, and S. Guluwadi, “Performance Analysis of XGBoost Ensemble Methods for Survivability with the Classification of Breast Cancer,” *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/4649510.

[13] N. Matondang and N. Surantha, “Effects of oversampling SMOTE in the classification of hypertensive dataset,” *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 4, pp. 432–437, 2020, doi: 10.25046/AJ050451.

[14] A. P. Joshi and B. V. Patel, “Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process,” *Oriental journal of computer science and technology*, vol. 13, no. 0203, pp. 78–81, Jan. 2021, doi: 10.13005/ojcst13.0203.03.

[15] R. Haque, N. Islam, M. Tasneem, and A. K. Das, “Multi-class sentiment classification on Bengali social media comments using machine learning,” *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 21–35, Jun. 2023, doi: 10.1016/j.ijcce.2023.01.001.

[16] H. S. Kang, Y. S. Choi, J. S. Yu, S. W. Jin, J. M. Lee, and Y. J. Kim, “Hyperparameter Tuning of OC-SVM for Industrial Gas Turbine Anomaly Detection,” *Energies (Basel)*, vol. 15, no. 22, pp. 1–12, Nov. 2022, doi: 10.3390/en15228757.

[17] R. Rahmaddeni, M. K. Anam, Y. Irawan, S. Susanti, and M. Jamaris, “Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination,” *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, pp. 32–38, Apr. 2022, doi: 10.33096/ilkom.v14i1.1090.32-38.

[18] M. Vedaraj, C. S. Anita, A. Muralidhar, V. Lavanya, K. Balasaranya, and P. Jagadeesan, “Early Prediction of Lung Cancer Using Gaussian Naive Bayes Classification Algorithm,” *International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 11, no. 6s, pp. 838–848, 2023, [Online]. Available: www.ijisae.org

[19] I. Firman Ashari, M. Cahyo Untoro, E. Maharani Sutowo, D. Salsabila, and D. Athiyah Zhabiyan, “Hyperparameter Tuning Feature Selection with Genetic Algorithm and Gaussian Naïve Bayes for Diabetes Disease Prediction,” *Jurnal Telematika*, vol. 17, no. 1, pp. 17–23, 2022, doi: 10.61769/jurtel.v17i1.488.

Improvement_of_Machine_Learning_Algorithms_with.pdf

ORIGINALITY REPORT

20%	16%	15%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.bright-journal.org Internet Source	5%
2	dokumen.pub Internet Source	1%
3	joiv.org Internet Source	1%
4	Abd Mizwar A. Rahim, Theopilus Bayu Sasongko. "Identify the Condition of Corn Plants Using Gray Level Co-occurrence Matrix and Bacpropagation", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2025 Publication	1%
5	Muliadi Muliadi, Agus Perdana Windarto, Solikhun Solikhun, Putrama Alkhairi. "A revolutionary convolutional neural network architecture for more accurate lung cancer classification", IAES International Journal of Artificial Intelligence (IJ-AI), 2025 Publication	1%
6	ijsrem.com Internet Source	1%
7	www.mdpi.com Internet Source	1%
8	www.aconf.org Internet Source	1%
9	Apri Siswanto, Akmar Efendi, Evizal Abdul Kadir. "Fingerprint Authentication In Smart	1%

Home Environment Based On Embedded System", 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2022
Publication

10	Nha Tran, Hung Nguyen, Dat Ly, Khanh Ngo, Hien D. Nguyen. "Advancing Violence Detection with Graph-Based Skeleton Motion Analysis", SN Computer Science, 2025 Publication	1 %
11	Akmar Efendi, Furizal, Apri Siswanto, Mursyidah. "Exam Reporting System (Case Study: Faculty Of Engineering, Universitas Islam Riau)", 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2022 Publication	1 %
12	Boby Pranata, Susanti. "Support Vector Machine untuk Sentiment Analysis Bakal Calon Presiden Republik Indonesia 2024", The Indonesian Journal of Computer Science, 2023 Publication	1 %
13	Ali Khumaidi, Herry Wahyono, Risanto Darmawan, Harry Dwiyana Kartika, Nuke L. Chusna, Muhammad Kaisar Fauzy. "RFM-AR Model for Customer Segmentation using K-Means Algorithm", E3S Web of Conferences, 2023 Publication	<1 %
14	Agung Ramadhanu, Jufriadif Na'am, Gunadi Widi Nurcahyo, Yuhandri. "Implementation of the Affine Segmentation Point Method and Image Blending Techniques in Creating New Songket Motifs", 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2022	<1 %

Publication

15	Abdulrahman, Amer Abdulfattah Mohamad. "Enhancing Cyber Threat Detection in Tweets Using Diffusion Models and Convolutional Neural Networks.", Rochester Institute of Technology Publication	<1 %
16	docplayer.info Internet Source	<1 %
17	papers.academic-conferences.org Internet Source	<1 %
18	Abulwafa Muhammad, Sarjon Defit, Gunadi Widi Nur Cahyo. "Comparison of Machine Learning Text Classification for Intent Sentiment Analysis", 2023 IEEE 7th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2023 Publication	<1 %
19	You-Jeng Chang, Ying-Lei Lin, Ping-Feng Pai. "Support Vector Machines with Hyperparameter Optimization Frameworks for Classifying Mobile Phone Prices in Multi-Class", Electronics, 2025 Publication	<1 %
20	hk.aconf.org Internet Source	<1 %
21	iccs.ac.in Internet Source	<1 %
22	Evizal Abdul Kadir, Apri Siswanto, Abdul Syukur. "Performance analysis of wireless LAN 802.11n standard for e-Learning", 2016 4th International Conference on Information and Communication Technology (ICoICT), 2016 Publication	<1 %

23	Faddilla Aulia Dara, Irfan Pratama. "Sentiment Analysis of the TikTok Tokopedia Seller Center Application Using Support Vector Machine (SVM) and Naive Bayes Algorithms", International Journal Software Engineering and Computer Science (IJSECS), 2025 Publication	<1%
24	Sreekumari S, Rajni Bhalla, Gursharan Singh. "Feature Selection and Model Evaluation for Heart Disease Prediction Using Ensemble Methods", Procedia Computer Science, 2025 Publication	<1%
25	Thompson, Chelsie C.. "A Machine Learning Approach to Evaluate Ransomware and Data Breach in Higher Education Institutions", The George Washington University, 2024 Publication	<1%
26	Vellingiri J, Kalaivanan K, Gopinath M P, Gobinath C, Prabhakar Rontala Subramaniam, Sarathkumar Rangarajan. "Strategies for classifying water quality in the Cauvery River using a federated learning technique", International Journal of Cognitive Computing in Engineering, 2023 Publication	<1%
27	jurnal.polibatam.ac.id Internet Source	<1%
28	jurnal.sar.ac.id Internet Source	<1%
29	repository.canterbury.ac.uk Internet Source	<1%
30	www.jait.us Internet Source	<1%
31	www.saiie.co.za Internet Source	<1%

32	Abdul Rachman Manga, Muhammad Acqmal Fadhilla Latief, Andi Widya Mufila Gaffar, Huzain Azis, Ramdan Satra, Yulita Salim. "Hyperparameter Tuning of Identity Block Uses an Imbalance Dataset with Hyperband Method", 2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM), 2024 Publication	<1 %
33	Akella Subrahmanya Narasimha Raju, K. Venkatesh, B. Padmaja, CH. N. Santhosh Kumar et al. "Exploring vision transformers and XGBoost as deep learning ensembles for transforming carcinoma recognition", Scientific Reports, 2024 Publication	<1 %
34	jtiulm.ti.ft.ulm.ac.id Internet Source	<1 %
35	repository.smuc.edu.et Internet Source	<1 %
36	theses.gla.ac.uk Internet Source	<1 %
37	www.ijisae.org Internet Source	<1 %
38	ro.ecu.edu.au Internet Source	<1 %
39	Agustia Kuspita Aryanti, Rudy Herteno, Fatma Indriani, Radityo Adi Nugroho, Muliadi Muliadi. "Implementation of Copeland Method on Wrapper-Based Feature Selection Using Random Forest For Software Defect Prediction", Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2025 Publication	<1 %

40	Mohamed Lahby, Al-Sakib Khan Pathan, Yassine Maleh. "Combatting Cyberbullying in Digital Media with Artificial Intelligence", CRC Press, 2023 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		