JIKO (Jurnal Informatika dan Komputter) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. 8, No. 3, Desember 2025, pp. 252-258 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v8i3.10866 e-ISSN: 2656-1948

PERFORMANCE ANALYSIS OF MACHINE LEARNING MODEL COMBINATION FOR SPACESHIP TITANIC CLASSIFICATION USING VOTING CLASSIFIER

Haria Wirawan[1*], Robet[2], Jackri Hendrik [3]

[1,2,3] Informatics Engineering, STMIK TIME, Medan, 20212, Indonesia
Email: [*1]hariawirawan65@gmail.com, [2]robertdetime@gmail.com, [3]jackri.hendrik@gmail.com

(Received: 16 October 2025, Revised: 27 October 2025, Accepted: 24 November 2025)

Abstract

The Spaceship Titanic dataset is fictional yet complex and challenging, featuring a mix of numerical and categorical features and missing values. This study aims to evaluate the performance of three machine learning model scenarios for classifying passenger status as “Transported” or “not”. The three scenarios implemented include linear-like models, a combination of the Top 5 Diverse models, and tree-based/ensemble models, each using a voting classifier approach. The voting model is employed because it can combine the strengths of multiple algorithms to reduce bias and variance, thus improving overall prediction accuracy and stability. The voting mechanism aggregates predictions from several base classifiers using two strategies: hard voting, which selects the majority class, and soft voting, which averages the predicted probabilities across models. The dataset was obtained from Kaggle and processed through several stages: data preprocessing, data splitting, model training, and evaluation. The evaluation results show that the tree-based/ensemble scenario achieved the highest accuracy of 90.38%, followed by the Top 5 Diverse model combination at 87.31% and the Linear-like model at 76.51%. visualization using the confusion matrix, ROC Curve, and Feature importance analysis further supports the claim that ensemble models are superior at detecting complex classification patterns. These findings suggest that tree-based ensemble models provide the most optimal approach for classification tasks on a dataset like Spaceship Titanic.

Keywords: *Machine Learning, Spaceship Titanic, Classification, Ensemble, Voting Classifier*

**Corresponding Author: Haria Wirawan*

1. INTRODUCTION

The Spaceship Titanic is a unique fictional data entity, inspired by a combination of two films: the 1997 Titanic and the Star Wars series. The Spaceship Titanic represents an intergalactic spacecraft that transports thousands of passengers from various home planets, yielding a complex, diverse dataset. During the journey, a mysterious event occurs that causes some passengers to be "transported" to an unknown dimension without any logical explanation or clear evidence.

Artificial intelligence (AI) and machine learning (ML) technologies have advanced significantly across various data sectors [1][2][3]. Machine learning algorithms such as Decision Tree, Random Forest, Logistic Regression, ExtraTreesClassifier, and Naïve Bayes are widely applied in disease prediction tasks because of their ability to learn patterns from data and make accurate classifications on unseen data [4][5]. The use of methods such as Random Forest, XGBoost, and SHAP allows for the interpretation of Feature Importance in data with class imbalance [6][7][8]. These techniques are relevant for understanding key predictors in an imbalanced dataset. Their ability to process, analyze, and extract patterns from data has made them essential tools in solving classification problems. Each classification algorithm has distinct strengths and weaknesses, making algorithm selection crucial in supervised learning tasks. Evaluation metrics such as accuracy, precision, and recall are essential for comparing model performance [9][10].

The classification problem in this fictional dataset presents not only a technical challenge but also an opportunity to examine how machine learning can generalize to complex data, including mixed types (both numerical and categorical) and missing values, which demand careful preprocessing and imputation strategies. Addressing these issues effectively allows for better generalization and prediction capability of the model. To solve this, combining different models and dividing them into several scenarios offers a practical solution[11][12].

This study aims to analyze the performance of machine learning models across three combination scenarios: linear-like models, a combination of the Top 5 Diverse models, and tree-based/ensemble models. Ensemble learning is a machine learning paradigm where multiple base learners are strategically generated and combined to solve a particular classification problem [13]. This technique improves the model's overall performance by reducing variance and bias [14][15]. Performance evaluation was conducted by measuring the accuracy of each combined model using a voting classifier and by generating confusion matrices for all three scenarios. The voting classifier technique combines multiple classification algorithms to improve prediction accuracy [16]. Demonstrated that combining various models through Ensemble Bagging and Ensemble Voting improved classification performance in email datasets. Their study confirmed that ensemble techniques could enhance model stability and reduce sensitivity to class imbalance compared to single-model approaches. Inspired by this, the present study also applies the Voting Classifier mechanism to evaluate how different model groups, linear-like, diverse, and tree-based, affect classification performance on the Spaceship Titanic dataset [17].

This research contributes by applying two ensemble approaches, hard voting and soft voting, that combine diverse models to enhance classification performance. Unlike previous studies on Ensemble Voting Classifier-based Machine Learning Models for Phishing Detection, which focused on optimizing model weights or balancing strategies in voting mechanisms, this study emphasizes the structural comparison of three distinct model groups, linear-like, diverse, and tree-based ensembles, within the same dataset [7]. The research gap lies in analyzing how each voting strategy (hard vs. soft) behaves across different model compositions, rather than improving a single ensemble configuration. This approach provides a more comprehensive understanding of model synergy and its effect on classification accuracy for complex datasets such as Spaceship Titanic.

It can be implemented using hard or soft voting, depending on whether the majority class or the average probability is used for final decision making [18][19]. It is also essential to consider that in many real-world machine learning problems, the presence distributions can complicate the classification process. Effective preprocessing, including missing value imputation, encoding, and feature scaling, is fundamental to ensuring the reliability of model training. The structure of each classification scenario plays a crucial role in final model performance. By identifying which combination of scenarios yields the best results, this research is expected to provide practical recommendations for selecting classification strategies for similar datasets.

2. RESEARCH METHOD

2.1 Research Stage

The Spaceship Titanic dataset consists of 8.693 row and 14 columns. Key attributes in this dataset include PassengerId, HomePlanet, CryoSleep, Cabin, Destination, Age, VIP, RoomService, FoodCourt, ShoppingMall, Spa, VRDeck, Name, and Transported. The primary classification target column is Transported (True/False). The initial stage of this study is data collection.

After data collection, preprocessing was conducted, including removing empty records, imputing missing values, encoding categorical features, and normalizing numerical features. These preprocessing steps are crucial in preparing clean data for modeling and ensuring generalization across unseen datasets [20]. The next step was to split the dataset into 80% for training and 20% for testing. Subsequently, multiple models were grouped into three scenarios, each built using a voting classifier approach [21]. The model performances were then compared using accuracy metrics. Finally, results were further analyzed through a Confusion matrix, ROC curve visualization, and feature importance analysis.

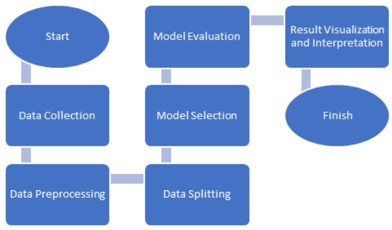

Figure 1. Stages of Research Methods

2.2 Modeling and Classification Techniques

The selection of classification methods was carried out by comparing three combinations of machine learning models across the following three different scenarios :

a. Linear-like: Logistic Regression, LinearSVC, SGDClassifier, RidgeClassifier, Perceptron.
b. Combination of the Top 5 Diverse Models: Logistic Regression, SVC, ExtraTreesClassifier, BaggingClassifier, KNeighborsClassifier.
c. Tree-Based/Ensemble: ExtraTreesClassifier, BaggingClassifier, AdaBoostClassifier, DecisionTreeClassifier, ExtraTreeClassifier.

These scenarios were combined using the Voting Classifier technique, with hard and soft voting methods, where the final prediction is determined by the accuracy of all models within each scenario [16][22]. Accuracy is one of the most widely used evaluation metrics for assessing classification model performance in machine learning. It measures the proportion of correct predictions, both positive and negative, relative to the total number of predictions made by the model.

Table 1. Model Types in Each Scenario

Scenario	Type of Model Combination	Number of Models
Linear-like	Logistic Regression, LinearSVC, SGDClassifier, RidgeClassifier, Perceptron.	5
Combination of the Top 5 Diverse Models	Logistic Regression, SVC, ExtraTreesClassifier, BaggingClassifier, KNeighborsClassifier.	5
Tree-based/Ensemble	ExtraTreesClassifier, BaggingClassifier, AdaBoostClassifier, DecisionTreeClassifier, ExtraTreeClassifier.	5

2.3 Model Performance Evaluation

To evaluate the model's performance in classifying passenger status in the Spaceship Titanic dataset, several evaluation methods were employed: accuracy score, voting classifier (soft and hard voting), ROC Curve with AUC, and feature importance interpretation.

1. Accuracy

Accuracy measures the proportion of correct predictions. It is defined as:

$$\mathrm{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \times 100\% \quad (1)$$

Accuracy is a primary metric for comparing classification model performance across scenarios, as it indicates how well the model recognizes patterns and generates correct predictions overall. The evaluation was carried out using the scikit-learn library, which included a confusion matrix and a classification report to reinforce the classification results of the best-performing model.

2. Voting Classifier

To describe the workflow of the proposed voting model in more detail, the process begins with the output predictions from several base classifiers such as Logistic Regression, SVC, Extra Trees, and AdaBoost. Each classifier produces either a class label (for hard voting) or a probability value (for soft voting). These predictions are then combined in an aggregation layer, where the final classification decision is made. The flow of this process can be illustrated as follows:

1. Input dataset - Preprocessing - Model training for each base classifier.
2. Each model generates its individual prediction output.
3. Voting mechanism:
   a. Hard voting: selects the majority class among model outputs.
   b. Soft voting: averages the predicted probabilities and selects the class with the highest mean probability.
4. Final output - Predicted class label “Transported” or “Not Transported.”

Each model scenario is implemented using a voting classifier, an ensemble method that combines multiple base classifiers to enhance prediction performance.

a. Soft voting

This method selects the final class based on the majority class predicted by all models :

$$\mathrm{P_{ensemble}}(\mathrm{C_k}) = \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{P_i}\,(\mathrm{C_k}) \quad (2)$$

b. Hard voting

This method averages the predicted class probabilities from all models and selects the class with the highest average probability :

$$\mathrm{P_{ensemble}} = \mathrm{mode}(\mathrm{C_1}, \mathrm{C_2}, \dots, \mathrm{C_n}) \quad (3)$$

3. ROC Curve

The ROC (Receiver Operating Characteristic) Curve is used to assess a classifier's diagnostic performance. It plots:

$$\mathrm{P_{ensemble}} = \mathrm{mode}(\mathrm{C_1}, \mathrm{C_2}, \dots, \mathrm{C_n}) \quad (4)$$

4. Feature Importance

Feature importance is a quantitative measure of how much each input variable contributes to the model’s prediction. To interpret which input features most influence the classification result, feature importance is calculated for applicable models. For tree-based models (and models that support it), the importance of a feature is calculated using:

$$FI(f) = \sum_{t \in t_f} \frac{N_t}{N} \cdot \triangle_i\,(t) \quad (5)$$

Higher importance values indicate that the feature plays a more significant role in improving the model's predictive performance.

3. RESULT AND DISCUSSION

3.1 Data Collection

The dataset used in this study was obtained from Kaggle under the title 'Spaceship Titanic' and consisted of 8,693 rows. These features include demographic information (such as age and home planet), cabin details, and entertainment and accommodation expenses on board the spacecraft.

Table 2. Description of the Spaceship Titanic Dataset

Column	Data Type	Number of Nulls	Number of Unique Values
PassengerId	object	0	2
HomePlanet	object	0	2
CyroSleep	bool	0	2
Cabin	object	0	2
Destination	object	0	2
Age	int64	0	2
VIP	bool	0	2
RoomService	float64	0	2
FoodCourt	float64	0	2
ShoppingMall	float64	0	2
Spa	float64	0	2
VRDeck	float64	0	2
Name	object	0	2
Transported	bool	0	2

3.2 Data Preprocessing

To evaluate the model's performance in classifying passenger status in the Spaceship Titanic dataset, several evaluation methods were employed: accuracy score, voting classifier (soft and hard voting), ROC Curve with AUC, and feature importance interpretation.

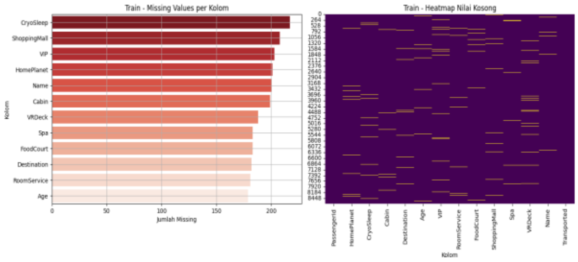

Figure 2. Missing Values and Heatmap – Train

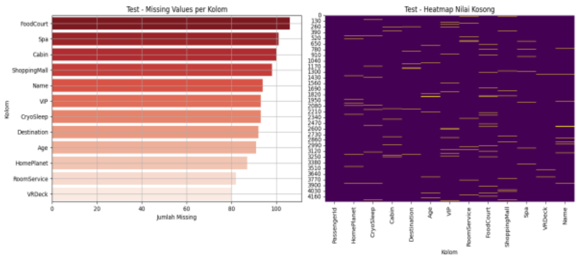

Figure 3. Missing Values and Heatmap – Test

3.3 Data Splitting

After data preprocessing, the data will first be evaluated in its unsplit form, depicted in Figure 4.

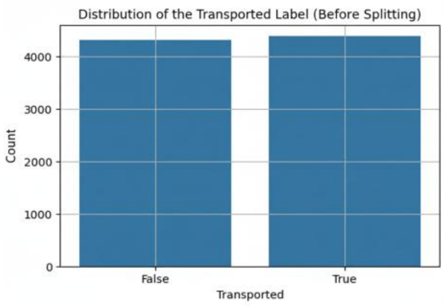

Figure 4. Distribution of the Transported (Before Splitting)

The dataset was then split into training (80%) and test (20%).

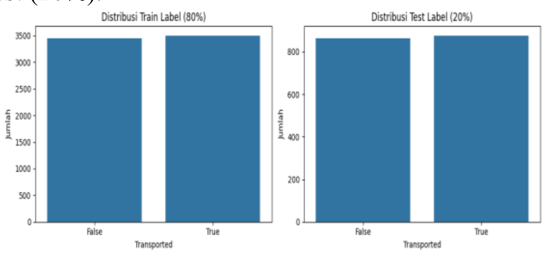

Figure 5. Distribution of the Transported (After Splitting)

3.4 Model Selection

After the data is split, the models are grouped into three combination scenarios as follows :

1. Linear-like: Logistic Regression, LinearSVC, SGDClassifier, RidgeClassifier, Perceptron
2. Combination of the Top 5 Diverse Models: Logistic Regression, SVC, ExtraTreesClassifier, BaggingClassifier, KNeighborsClassifier
3. Tree-based/ensemble: ExtraTreesClassifier, BaggingClassifier, AdaBoostClassifier, DecisionTreeClassifier, ExtraTreeClassifier

Each scenario is implemented using the Voting Classifier approach, incorporating both hard and soft voting methods, with classification decisions based on the majority vote across all models.

3.5 Model Evaluation

The performance evaluation is conducted using the accuracy metric, calculated from predictions made on the test data. The evaluation results are presented in Table 3 below:

Table 3. Accuracy Results of the Three Scenarios

Scenario	Type of Model Combination	Type of Voting	Accuracy
Combination of the Top 5 Diverse Models	Combination of the Top 5 Diverse Models	Soft Voting	87.31%
Linear-like	Linear-like	Hard Voting	76.51%
Tree-based	Ensemble	Soft Voting	**90.38%**

Scenarios 1 and 3 use soft voting, while Scenario 2 uses hard voting because it consists of linear models, some of which do not support probability prediction (predict_proba), which is required for soft voting. The tree-based/ensemble scenario produced the best result, achieving an accuracy of 90.38%.

3.6 Result Visualization and Interpretation

After evaluating the models, the next step is to visualize the three scenarios using the confusion matrix, ROC curve, and feature importance. The results will indicate which scenario performs the best.

1. Confusion Matrix

The confusion matrix is used to visualize the differences in results among the three scenarios.

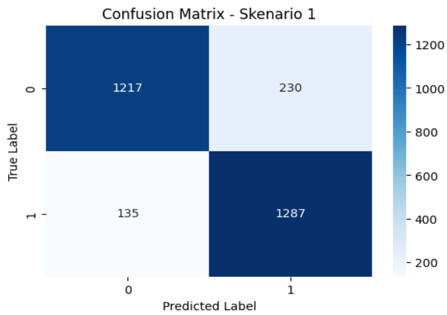

Figure 6. Confusion Matrix - Scenario 1

The results of Scenario 1 show that the model performs well in identifying both positive and negative classes, as indicated by the high number of True Positives (TP) and True Negatives (TN). In contrast, the number of False Positives (FP) and False Negatives (FN) remains relatively low. Although the accuracy is high, the model demonstrates a fairly balanced performance.

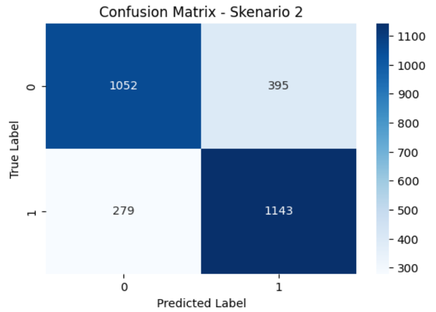

Figure 7. Confusion Matrix - Scenario 2

The results of Scenario 2 show a significant increase in False Positives (FP) and False Negatives (FN), indicating that the model frequently makes incorrect predictions and has lower precision. This results in lower overall performance than in Scenario 1.

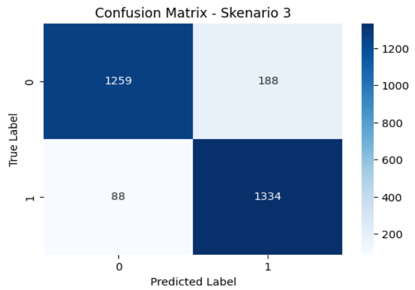

Figure 8. Confusion Matrix - Scenario 3

The results of Scenario 3 indicate that it outperforms the previous two scenarios, with the highest number of True Positives (TP) and True Negatives (TN), and the lowest number of False Positives (FP) and False Negatives (FN). This reflects excellent accuracy, precision, and recall, demonstrating that the tree-based ensemble model generalizes better than the other two scenarios.

2. ROC Curve (Receiver Operating Characteristic Curve)

The method involves visualizing the three scenarios using the ROC (Receiver Operating Characteristic) curve.

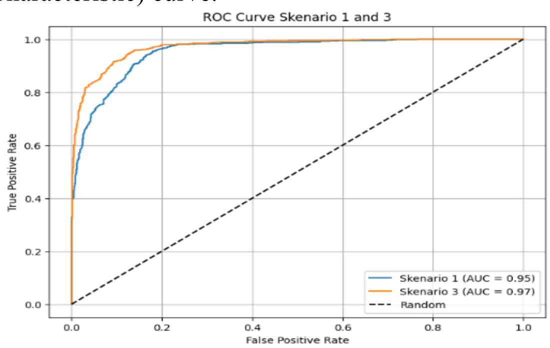

Figure 9. ROC Curve - Scenarios 1 and 3

The ROC Curve results show that Scenario 3 achieved an AUC of 0.97, higher than Scenario 1's 0.95. This indicates that the model in Scenario 3 is better at distinguishing between positive and negative classes. Scenario 2 is not included in the ROC visualization because it consists of a Combination of the Top 5 Diverse models, such as LinearSVC, RidgeClassifier, Perceptron, and Logistic Regression. In Scenario 2, only a few models achieved high accuracy, while others performed poorly, leading to inconsistent probability scores. For this reason, Scenario 2 is also excluded from the following method, which is Feature Importance.

3. Feature Importance

In this study, feature importance is used to identify the most influential passenger characteristics that affect classification results. For example, higher importance values of CryoSleep, Spa, and VRDeck indicate that these variables play a critical role in determining whether a passenger was "Transported." This analysis helps to interpret the model's decision process and validate which attributes are truly impactful within the dataset. The following step visualizes Scenarios 1 and 3 using feature importance.

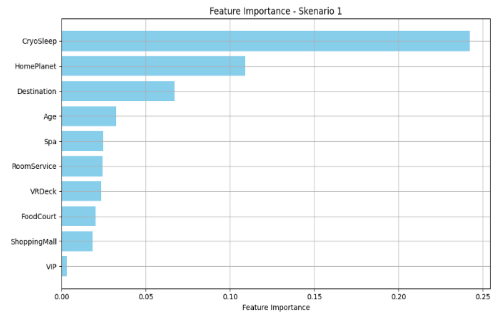

Figure 10. Feature Importance - Scenario 1

In Scenario 1, which uses linear-like models, the most dominant features significantly influence the classification results. However, features such as HomePlanet, Destination, and Age also make significant contributions, indicating that passengers' origin, destination, and age are relevant to the likelihood of being "Transported".

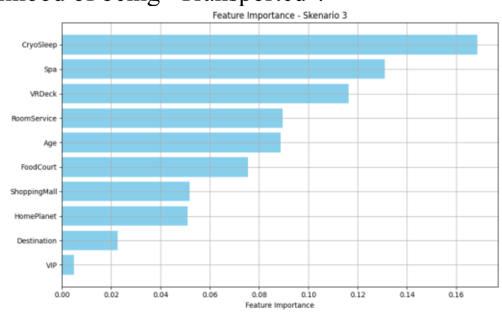

Figure 11. Feature Importance - Scenario 3

Scenario 3 shows that the CryoSleep feature exerts a greater influence. However, features such as Spa, VRDeck, and RoomService carry more weight than in Scenario 1, as tree-based models can capture complex patterns and interactions among features. Both scenarios agree that CryoSleep is a key feature, but tree-based models emphasize the passenger's service usage, while linear models emphasize the journey's origin and destination.

After presenting all the visualizations and interpretations, the model in Scenario 3 is proven to deliver the best overall performance. The confusion matrix indicates high, balanced classification accuracy across classes; the ROC Curve shows the highest AUC, reflecting optimal detection capability; and the feature importance highlights the most relevant variables. This makes the tree-based ensemble approach highly recommended for use on the Spaceship Titanic dataset, which is complex, contains a mix of numerical and categorical features, and includes missing values.

The research findings show that the soft voting mechanism consistently produced better accuracy than hard voting. This is because soft voting accounts for each classifier's probabilistic confidence, leading to smoother, more reliable decision boundaries. In contrast, hard voting may overlook the confidence score and count the majority class, thereby reducing sensitivity to borderline cases.

Compared with previous research, which also reported that soft voting improves stability and accuracy in ensemble-based classification tasks. This confirms that combining probabilistic outputs from diverse models can yield more consistent predictions, especially on datasets with mixed numerical and categorical features. Overall, the proposed model demonstrates that tree-based soft voting ensembles are not only accurate (90.38%) but also robust in generalizing across feature distributions. These results reinforce the effectiveness of ensemble learning methods and can serve as a reference for future studies that apply hybrid model combinations to complex classification problems.

4. CONCLUSION

This study evaluated three combinations of machine learning models on the Spaceship Titanic dataset: linear-like, Top 5 Diverse, and tree-based/ensemble. The classification process used a voting classifier and included data preprocessing, data splitting, model training, and evaluation. The accuracy results show that the tree-based/ensemble achieved the highest performance with a score of 90.38%, followed by the Combination of the Top 5 Diverse at 87.31%, and the linear-like at 76.51%.

The confusion matrix visualization shows that the ensemble model achieves the most balanced results in recognizing both target classes. Meanwhile, the other two scenarios still show relatively high misclassification in one of the classes. The ROC Curve reinforces this finding, with a higher AUC for Scenario 3, indicating greater model sensitivity. Additionally, the feature importance analysis reveals that CryoSleep is the dominant feature across all scenarios, with differing preferences for additional features between linear and ensemble models.

Based on the overall evaluation, the tree-based/ensemble model is the optimal approach for classification on the Spaceship Titanic dataset. This model excels not only in accuracy but also in prediction stability and the ability to capture complex patterns from both numerical and categorical features. These findings are expected to serve as a reference for selecting ensemble-based classification strategies for similar datasets in the future.

5. REFERENCE

[1] A. Rahman and S. S. Prasetiyowati, "Performance Analysis of the Hybrid Voting Method on the Classification of the Number of Cases of Dengue Fever," *Int. J. Inf. Commun. Technol.*, vol. 8, no. 1, pp. 10–19, 2022, doi: 10.21108/ijoict.v8i1.614.

[2] A. Gumilar, S. S. Prasetiyowati, and Y. Sibaroni, "Performance Analysis of Hybrid Machine Learning Methods on Imbalanced Data (Rainfall Classification)," *J. RESTI*, vol. 6, no. 3, pp. 481–490, 2022, doi: 10.29207/resti.v6i3.4142.

[3] A. J. Barid, Hadiyanto, and A. Wibowo, "Optimization of the algorithms use ensemble and synthetic minority oversampling technique for air quality classification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 3, pp. 1632–1640, 2024, doi: 10.11591/ijeecs.v33.i3.pp1632-1640.

[4] R. Mardianto, Stefanie Quinevera, and S. Rochimah, "Perbandingan Metode Random Forest, Convolutional Neural Network, dan Support Vector Machine Untuk Klasifikasi Jenis Mangga," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 63–71, 2024, doi: 10.52158/jacost.v5i1.742.

[5] D. N. Cholis and N. Ulinnuha, "An Ensemble Voting Approach for Dropout Student Classification Using Decision Tree C4.5, K-Nearest Neighbor and Backpropagation," *Indones. J. Artif. Intell. Data Min.*, vol. 6, no. 1, p. 107, 2023, doi: 10.24014/ijaidm.v6i1.23412.

[6] F. T. Kurniati, D. H. Manongga, E. Sediyono, S. Y. J. Prasetyo, and R. R. Huizen, "Object Classification Model Using Ensemble Learning with Gray-Level Co-Occurrence Matrix and Histogram Extraction," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 3, pp. 793–801, 2023, doi: 10.26555/jiteki.v9i3.26683.

[7] Enas Mohammed Hussien Saeed, "An Ensemble Voting Classifier based on Machine Learning Models for Phishing Detection," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 12, no. 1, pp. 15–27, 2025, doi: 10.32628/ijsrset251211.

[8] S. Yousefi and M. Poornajaf, "Comparison of result of machine learning algorithms in predicting heart disease," *Front. Heal. Informatics*, vol. 12, 2023, doi: 10.30699/fhi.v12i0.402.

[9] S. Putri Aulia, B. Rahmat, and A. Junaidi, "Enhancing Heart Disease Prediction through SMOTE-ENN Balancing and RFECV Feature Selection," *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 3, pp. 1968–1973, 2025, doi:

10.59934/jaiea.v4i3.1057.

[10] S. Tomar, D. Dembla, and Y. Chaba, "Analysis and Enhancement of Prediction of Cardiovascular Disease Diagnosis using Machine Learning Models SVM, SGD, and XGBoost," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 4, pp. 469–479, 2024, doi: 10.14569/IJACSA.2024.0150449.

[11] K. Cao *et al.*, "Prediction of cardiovascular disease based on multiple feature selection and improved PSO-XGBoost model," *Sci. Rep.*, vol. 15, no. 1, pp. 1–12, 2025, doi: 10.1038/s41598-025-96520-7.

[12] G. Alwakid, F. Ul Haq, N. Tariq, M. Humayun, M. Shaheen, and M. Alsadun, "Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective," *BMC Cardiovasc. Disord.*, vol. 25, no. 1, 2025, doi: 10.1186/s12872-025-04550-w.

[13] S. Akinola, R. Leelakrishna, and V. Varadarajan, "Enhancing cardiovascular disease prediction: A hybrid machine learning approach integrating oversampling and adaptive boosting techniques," *AIMS Med. Sci.*, vol. 11, no. 2, pp. 58–71, 2024, doi: 10.3934/medsci.2024005.

[14] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," *Processes*, vol. 11, no. 4, 2023, doi: 10.3390/pr11041210.

[15] D. Kurniadi, A. I. Pertiwi, and A. Mulyani, "Multi-Algorithm-Based Ensemble Voting Classifier and SMOTE Method for He ar t Disease Classification," vol. 14, no. 2, pp. 145–153, 2025.

[16] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Appl. Soft Comput.*, vol. 133, pp. 1–37, 2023, doi: 10.1016/j.asoc.2022.109924.

[17] A. Helmut and D. T. Murdiansyah, "Multiclass Email Classification by Using Ensemble Bagging and Ensemble Voting," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, pp. 144–149, 2023, doi: 10.33387/jiko.v6i2.6394.

[18] N. Agustina and C. N. Ihsan, "Pendekatan Ensemble untuk Analisis Sentimen Covid19 Menggunakan Pengklasifikasi Soft Voting," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 263–270, 2023, doi: 10.25126/jtiik.20231026215.

[19] I. P. Adi Pratama, E. S. Jullev Atmadji, D. A. Purnamasar, and E. Faizal, "Evaluating the Performance of Voting Classifier in Multiclass Classification of Dry Bean Varieties," *Indones. J. Data Sci.*, vol. 5, no. 1, pp. 23–29, 2024, doi: 10.56705/ijodas.v5i1.124.

[20] A. Munandar, W. Maulana Baihaqi, and A. Nurhopipah, "A Soft Voting Ensemble Classifier to Improve Survival Rate Predictions of Cardiovascular Heart Failure Patients," *Ilk. J. Ilm.*, vol. 15, no. 2, pp. 344–352, 2023, doi: 10.33096/ilkom.v15i2.1632.344-352.

[21] Octavian, A. Badruzzaman, M. Y. Ridho, and B. D. Trisedya, "Enhancing Weighted Averaging for CNN Model Ensemble in Plant Diseases Image Classification," *J. RESTI*, vol. 8, no. 2, pp. 272–279, 2024, doi: 10.29207/resti.v8i2.5669.

[22] A. S. Kirono and Y. Nataliani, "Perbandingan Algoritma Machine Learning dalam Analisis Penyebab Penyakit Gagal Jantung," *J. Edukasi dan Penelit. Inform.*, vol. 10, no. 2, p. 296, 2024, doi: 10.26418/jp.v10i2.78369.