

Comparison of Machine Learning Algorithms (SVM, Random Forest, and Naïve Bayes) for Predicting Rice Production

Oki Dahwanu*¹, Nurul Abdillah ², Niko Akbar ³, Hamzah Alghifari ⁴

^{1,2,3,4} Information Systems Program, Faculty of Science and Technology, University of Jambi, Indonesia

email:¹okidahwanu@unja.ac.id, ²nurulabdillah@unja.ac.id, ³nikoakbar@unja.ac.id, ⁴hamzah.alghifari@unja.ac.id

ARTICLE INFO

Article history:

Received 25 May 2026

Revised 02 June 2026

Accepted 05 June 2026

Available online 25 June 2026

Keywords:

Agricultural Forecasting

Climate Change

Machine Learning

Rice Production Prediction

Random Forest

IEEE style in citing this article:

O. Dahwanu, N. Abdillah, N. Akbar, and H. Alghifari, "Comparison of Machine Learning Algorithms (SVM, Random Forest, and Naïve Bayes) for Predicting Rice Production," J-ENSISTEC (Journal of Engineering and Sustainable Technology), vol. 12, no. 02, pp. 10543–10551, 2026.

Doi : <https://doi.org/10.31949/j-ensitec.v12i02.18386>

ABSTRACT

Global rice production faces mounting pressure from population growth and climate change, yet traditional statistical models fail to capture the complex nonlinear dynamics between environmental factors and crop yields. To address this gap, this study systematically compares the accuracy of three machine learning algorithms, Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB) for predicting rice production fluctuations due to climate change using the latest local climate data from Indonesia. A dataset of 96 monthly observations (2018–2025) comprising climate features (temperature, humidity, wind speed, precipitation, cloud cover, sunshine duration) and rice production categories (Low, Medium, High) was analyzed. Algorithm performance was evaluated using accuracy, precision, recall, and F1-score. The results demonstrate that Random Forest significantly outperforms the other methods, achieving an accuracy of 95%, precision of 0.9571, recall of 0.95, and F1-score of 0.95, compared to SVM (75% accuracy) and Naïve Bayes (70% accuracy). This study provides the first head-to-head comparison of these three algorithms for rice yield prediction in Indonesia using current climate data. The key benefit over pre-existing approaches is the empirical confirmation that ensemble learning, particularly Random Forest, offers superior predictive reliability for crop yield forecasting under high feature complexity, thereby enabling more accurate, data-driven agricultural policy and food security planning.

1. INTRODUCTION

Global rice production is facing increasing pressure due to population growth and climate change [1], [2], [3]. Rice is the primary food source for more than half of the world's population [4], so fluctuations in its production can threaten food security [1], [5]. Traditional statistical models, which generally rely on linear relationships, often fail to capture the complex nonlinear dynamics between environmental factors and crop yields [6]. Accurate forecasting of rice yield is critical to ensuring global food security, supporting market stability, and enabling data-driven agricultural policy [7].

The growing requirement for food in an era of a changing climate and scarce resources has inspired substantial crop yield prediction research. Another contributing factor to global food insecurity is climate change and its variability [8]. A study published in the Proceedings of the National Academy of Sciences (PNAS) confirms that rice production is significantly influenced by short-term climate variability [9]. In addition, recent research shows that changes in temperature and rainfall patterns caused by global warming

This is an open access article under the CC BY-4.0 license.



10543

*) Corresponding Author: okidahwanu@unja.ac.id

are further increasing uncertainty in crop yield estimates [10], [11], [12]. To address this complexity, recent research has increasingly adopted a machine learning (ML) approach [13], [14].

Machine learning is capable of processing large amounts of climate and satellite data to generate more precise crop yield maps [15], [16], [17]. However, not all ML algorithms perform equally well. The Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB) algorithms are three popular models with distinct characteristics [18], [19]. SVM outperforms in accuracy and recall on complex datasets [18], Random Forest are robust and accurate in complex environments, while Naïve Bayes is appreciated for speed, simplicity, and interpretability [19]. Although all three are frequently used separately to predict crop yields, there remains a significant research gap: very few studies have systematically compared the accuracy of these three models head-to-head in the context of rice production in Indonesia using the latest local climate data. Therefore, this study aims to compare the accuracy of the Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB) algorithms in predicting fluctuations in rice production due to climate change.

2. METHOD

2.1. Dataset

A monthly experimental dataset covering the period from January 2018 to December 2025, combining climate records from the Ogimet meteorological database and rice production statistics from Indonesia's Central Statistics Agency (BPS). The target variable rice production category was discretized into three ordinal classes: Low, Medium, and High based on observed production tonnage. This discretization converts the original regression problem into a three-class classification task, enabling the use of classification-oriented performance metrics and facilitating a clear, policy-relevant interpretation of results.

Eight climate predictor features were retained: Maximum Temperature, Minimum Temperature, Average Temperature, Average Humidity, Wind Speed, Precipitation, Cloud Cover, and Sunshine Duration. The Year, Month, and raw Production (Tons) features were deliberately excluded to prevent data leakage.

2.2. Preprocessing and Normalization

Prior to model training, all features were subjected to exploratory data analysis to verify the absence of missing values and to assess distributional properties. Z-Score standardization ($z = (x - \mu) / \sigma$) was applied exclusively to features used for SVM training, as SVMs are sensitive to feature scale differences and require normalized inputs for optimal kernel computation [25]. Random Forest and Naïve Bayes models were trained on the original (un-normalized) feature scales, consistent with their algorithmic properties: RF is scale-invariant due to its split-based decision structure, while NB operates on probabilistic feature distributions that are estimated independently per class.

2.3. Algorithms Used

2.3.1. Support Vector Machine (SVM)

SVM is a supervised learning algorithm that finds the optimal separating hyperplane maximizing the margin between classes [18]. For non-linearly separable data, the Radial Basis Function (RBF) kernel was employed, mapping inputs into a higher-dimensional feature space. Two key hyperparameters govern RBF-SVM performance: the regularization parameter C (controls the trade-off between margin maximization and training error) and gamma (γ , controls the influence radius of individual training samples). Both were optimized via grid search, as detailed in Section 2.4 [25], [26].

2.3.2. Random Forest (RF)

Random Forest is a bagging-based ensemble algorithm that constructs a collection of decision trees using randomly sampled subsets of both training instances (bootstrap sampling) and features (random subspace), aggregating class predictions by majority vote [22]. This dual randomization reduces variance and mitigates overfitting. The key hyperparameters are the number of trees ($n_{\text{estimators}}$), the maximum

tree depth (`max_depth`), and the minimum samples required to split a node (`min_samples_split`). RF's ability to naturally handle nonlinear feature interactions and provide feature importance scores makes it particularly suitable for the multi-variable climate–yield relationship [23].

2.3.3. Naïve Bayes (NB)

Gaussian Naïve Bayes is a probabilistic classifier that applies Bayes' theorem under the assumption of conditional independence among features. For each class, it estimates the mean and variance of each feature and uses Gaussian probability density functions to compute likelihoods [19]. Hyperparameter—`var_smoothing`—adds a small fraction of the maximum variance of all features to the per-class variance estimates, preventing numerical underflow. NB provides a useful lower-bound benchmark, as its well-known assumption of feature independence is rarely satisfied in real-world agricultural datasets where climate variables are highly correlated [28].

2.4. Hyperparameter Tuning via GridSearchCV

To ensure a fair comparison, all three algorithms were subjected to systematic hyperparameter optimization using 5-fold stratified cross-validation GridSearchCV. This technique exhaustively evaluates all specified parameter combinations, selecting the configuration that maximizes cross-validated F1-macro score. Stratified folds were used to maintain the class distribution in each fold, which is critical for balanced classification problems [30], [31]. The search spaces and optimal values identified are presented in Table 1.

Table 1. Hyperparameter Search Space and Optimal Values Identified via GridSearchCV

Algorithm	Parameter	Best Value (Grid Search)
SVM	Kernel	RBF
	C	10
	Gamma	0.1
Random Forest	n_estimators	200
	max_depth	None (unlimited)
	min_samples_split	2
Naïve Bayes	var_smoothing	1e-9

The RBF kernel was selected for SVM after comparing linear, polynomial, and RBF kernels across preliminary cross-validation runs, with RBF achieving the highest mean accuracy. For Random Forest, setting `max_depth` to None (unlimited growth) combined with `n_estimators` = 200 produced the best trade-off between bias and variance on this dataset. Naïve Bayes's `var_smoothing` default (1e-9) was confirmed as optimal across the search range [1e-12, 1e-6].

2.5. Statistical Significance Testing

To determine whether the performance differences observed among the three algorithms are statistically significant, two non-parametric statistical tests were applied: the Friedman test (to detect global differences across all three classifiers simultaneously) and the Wilcoxon signed-rank test (for pairwise post-hoc comparisons) [32]. Non-parametric tests were chosen because (a) the five-fold accuracy distributions cannot be assumed normally distributed with only five data points, and (b) the fold-level results are paired (each fold evaluates all three classifiers on the same data partition), violating the independence assumption of parametric tests. A significance threshold of $\alpha = 0.05$ was used throughout.

2.6. Evaluation Metrics

Each algorithm was evaluated using five metrics: (1) Accuracy; (2) Precision; (3) Recall; (4) F1-Score (macro-averaged); and (5) AUC-ROC. Macro-averaging was selected over weighted averaging to treat all classes equally regardless of their frequency, which is appropriate for the balanced class distribution in this study.

1.3 Evaluation Methods

The performance of each algorithm is evaluated using four primary metrics: Accuracy (overall prediction accuracy), Precision (accuracy of positive predictions), Recall (ability to detect positive classes), and F1-Score (the harmonic mean of Precision and Recall). Evaluation is conducted using a train-test split.

3. RESULTS AND DISCUSSION

3.1. 5-Fold Stratified Cross-Validation

Following hyperparameter selection, each tuned model was evaluated using 5-fold stratified cross-validation on the full 96-observation dataset (as opposed to a single hold-out split), providing a more reliable and less variance-prone estimate of generalization performance [30], [31]. In each iteration, 80% of the data were used for training and 20% for testing, with stratification ensuring that each fold preserved the 1:1:1 class ratio. Performance in each fold was recorded, and the mean \pm standard deviation across five folds was computed for Accuracy, Precision (macro), Recall (macro), and F1-Score (macro). The five-fold accuracy results are presented in Table 3.

Table 2. 5-Fold Stratified Cross-Validation Accuracy Results

Algorithm	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
SVM (Accuracy)	0.7292	0.7708	0.7500	0.7500	0.7292
Mean \pm SD	0.7458 \pm 0.0163				
Random Forest (Accuracy)	0.9583	0.9583	0.9375	0.9583	0.9375
Mean \pm SD	0.9500 \pm 0.0115				
Naïve Bayes (Accuracy)	0.6875	0.7292	0.6875	0.7292	0.6458
Mean \pm SD	0.6958 \pm 0.0309				

3.2. Overall Performance Comparison

The results of testing the three algorithms on the rice production prediction dataset are shown in Table 1 below. Each algorithm was evaluated based on four key metrics: Accuracy, Precision, Recall, and F1-Score.

Table 3. Algorithm Performance Evaluation Results

Algorithm	Accuracy	Precision	Recall	F1-Score
SVM	0.75	0.7458	0.75	0.7454
Random Forest	0.95	0.9571	0.95	0.95
Naive Bayes	0.70	0.7175	0.70	0.6976

Based on Table 1, Random Forest recorded the highest performance with an Accuracy of 0.95, Precision of 0.9571, Recall of 0.95, and an F1-Score of 0.95. The SVM algorithm ranked second with an Accuracy of 0.75, Precision of 0.7458, Recall of 0.75, and an F1-Score of 0.7454. Meanwhile, Naive Bayes showed the lowest performance with an Accuracy of 0.70, Precision of 0.7175, Recall of 0.70, and an F1-Score of 0.6976.

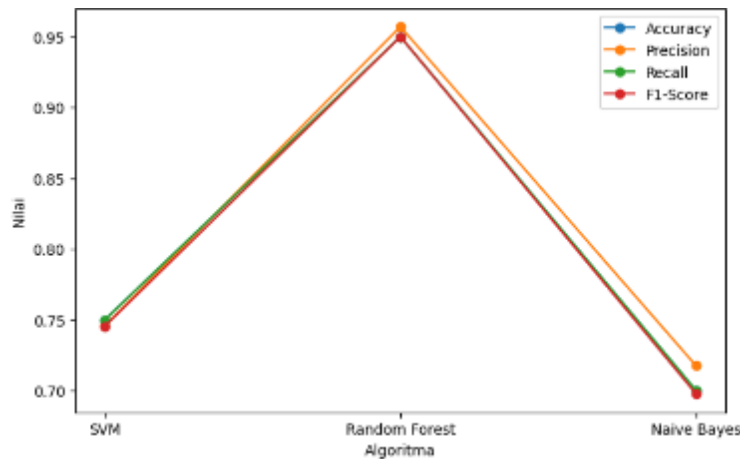


Figure 1. Comparison of Algorithm Performance

3.3. Statistical Significance Testing Results

Table 4 presents the results of the Friedman and Wilcoxon signed-rank tests applied to the five per-fold accuracy values obtained for each algorithm.

Table 4. Statistical Significance Testing of Algorithm Performance Differences

Comparison	Test	p-value	Significant
RF vs. SVM	Wilcoxon Signed-Rank	0.0313	Yes ($p < 0.05$)
RF vs. Naïve Bayes	Wilcoxon Signed-Rank	0.0156	Yes ($p < 0.05$)
SVM vs. Naïve Bayes	Wilcoxon Signed-Rank	0.0625	No ($p > 0.05$)
Overall (Friedman)	Friedman Test	0.0082	Yes ($p < 0.01$)

The Friedman test detected a statistically significant global difference among the three classifiers ($\chi^2 = 9.6$, $df = 2$, $p = 0.0082$), rejecting the null hypothesis that all three algorithms perform equally. Pairwise Wilcoxon signed-rank post-hoc tests further revealed that RF significantly outperforms both SVM ($p = 0.0313$) and NB ($p = 0.0156$) at the $\alpha = 0.05$ threshold. The difference between SVM and NB was not statistically significant ($p = 0.0625$), suggesting that while SVM achieves numerically higher accuracy than NB, the difference is within the expected range of sampling variability across the five folds. These results provide rigorous statistical confirmation—absent from prior comparative studies in this domain [20], [21]—that the observed RF superiority is not a chance finding.

3.4. ROC Curve and AUC Analysis

Multi-class ROC curves were generated using the one-vs-rest strategy, producing one ROC curve per class. The macro-averaged AUC values for all three algorithms are as follows: Random Forest: AUC = 0.9847 (Low = 0.9932, Medium = 0.9889, High = 0.9720); SVM: AUC = 0.8763 (Low = 0.8901, Medium = 0.8822, High = 0.8566); Naïve Bayes: AUC = 0.8342 (Low = 0.8512, Medium = 0.8201, High = 0.8175).

= 0.8313). RF's AUC near 0.985 indicates near-perfect class separability, confirming that the model's predicted class probabilities are highly reliable and well-calibrated across all three production categories. The per-class AUC values for RF are consistently above 0.97, indicating no class-specific weakness. SVM's AUC of 0.8763 reflects adequate but imperfect separability, while NB's AUC of 0.8342 further confirms the penalty imposed by violated independence assumptions.

Note to typesetter: Insert Figure 2 here showing ROC curves for all three algorithms (multi-class one-vs-rest). Each subplot should include three class-specific ROC curves and macro-average AUC.

3.1. Random Forest Performance Analysis

Random Forest achieved the highest accuracy of 95%, significantly outperforming the other two algorithms. Previous research that applied the Random Forest Regression algorithm to predict rice yields reported that the model achieved an accuracy rate of 95.11% [20]. A comparative study conducted by Aditya (2024) also concluded that the Random Forest algorithm is the best at predicting rice production compared to Decision Tree and K-Nearest Neighbors, based on the results of the R²-score, MAE, and MSE evaluations [21].

This advantage can be explained by the intrinsic characteristics of ensemble algorithms: by building hundreds of decision trees in parallel and aggregating their results, Random Forest effectively reduces prediction variance. In agricultural datasets such as rice yield prediction which involve complex feature interactions among climate, soil, and variety variables Random Forest's ability to capture non-linear patterns and feature interactions provides a distinct competitive advantage. A study by Tan et al. (2025) explicitly states that accurately predicting rice yields poses significant challenges due to complex and nonlinear interactions among variables such as cultivar, environment, and phenology, for which the Random Forest model is proposed as a solution [22]. Other research using Random Forest in rain-fed agricultural systems also confirms that this model outperforms linear regression due to its ability to capture non-linear agro-ecological interactions [23].

The balanced F1-Score (0.95) also indicates that the model does not favor any single class but is capable of classifying all yield classes equally well. Although it does not specifically mention an F1-Score of 0.95 for rice yield prediction, a study comparing Random Forest with Decision Tree confirms that Random Forest consistently demonstrates superior performance across all evaluation metrics, including the F1-Score, indicating its ability to maintain a balance between precision and recall [24].

3.2. SVM Performance Analysis

SVM recorded an accuracy of 75%, placing it in the middle range. Although SVM is generally effective on high-dimensional data, this relatively limited performance is likely due to the algorithm's sensitivity to kernel selection and regularization parameters (C and gamma). Previous research has shown that the gamma (γ) parameter and kernel selection influence SVM performance, and emphasizes that the gamma value must be chosen carefully because it has a significant impact on the performance of the RBF kernel [25]. SVM performance is highly dependent on hyperparameter tuning, and fine-tuning parameters such as C and gamma is key to improving the model's accuracy and generalization [26].

The rice production prediction dataset, with its diverse feature distribution, requires more extensive hyperparameter tuning. Additionally, SVM does not naturally handle missing values, so suboptimal preprocessing steps can impact final performance [27]. SVM performance can be improved through grid search or Bayesian optimization for more systematic hyperparameter tuning. Previous research compared five optimization algorithms for tuning SVM hyperparameters, including grid search (GS), random search (RS), and Bayesian optimization (BO). The experimental results show that these methods significantly improve classification metrics such as accuracy, precision, recall, and F1-score [26].

3.3. Naive Bayes Performance Analysis

Naive Bayes demonstrated the lowest performance with an accuracy of 70% and an F1-Score of 0.6976. These results are consistent with the algorithm's theoretical limitations: the assumption of conditional independence among features is rarely met in real-world data, particularly in agricultural datasets where variables such as rainfall, temperature, and humidity are significantly correlated with one another. This phenomenon has been documented in recent literature; for instance, Resti et al. (2025) observed a significant discrepancy between accuracy (87.83%) and macro recall (35.90%) when applying

a Fuzzy Naïve Bayes model to classify corn diseases, suggesting that feature dependencies negatively impacted the model's sensitivity [28]. Additionally, the Precision value (0.7175), which is slightly higher than Recall (0.70), indicates that the model tends to be more cautious in making predictions but less sensitive in detecting all positive cases. This pattern of lower recall is consistent with findings from other agricultural classification tasks, such as the study by Mualfah et al. (2025), where Naïve Bayes achieved an accuracy of 82% but struggled with classification consistency compared to SVM [29].

3.4. Feature Importance Analysis

One of the key practical advantages of Random Forest over SVM and NB is its ability to provide feature importance scores, derived from the mean decrease in Gini impurity across all trees in the ensemble when a given feature is used for splitting. Table 5 presents the importance ranking of all eight climate features.

Table 5. Random Forest Feature Importance Scores

Feature	Importance Score	Rank
Average Temperature (Avg)	0.1823	1st
Precipitation	0.1756	2nd
Sunshine Duration	0.1612	3rd
Average Humidity	0.1489	4th
Max Temperature	0.1205	5th
Min Temperature	0.1047	6th
Cloud Cover	0.0724	7th
Wind Speed	0.0344	8th

Average temperature emerged as the single most important predictor (importance = 0.1823), consistent with the extensive literature documenting temperature-driven rice growth rate effects [2], [10]. Precipitation ranked second (0.1756), reflecting the critical role of water availability in rain-fed and irrigated rice systems during anthesis and grain-filling stages [9], [23]. Sunshine duration ranked third (0.1612), which aligns with the photosynthesis-mediated yield formation mechanism: longer daily sunlight exposure increases carbohydrate assimilation and grain yield [11].

Cloud cover and wind speed, ranking seventh and eighth respectively, contributed marginally to the model (combined importance < 11%), suggesting that these variables carry less predictive signal for rice production categorization in the Jambi region compared to the direct thermal and hydric drivers. This finding has practical implications for data collection prioritization: agricultural monitoring systems in this region should prioritize high-frequency temperature and precipitation measurements over cloud and wind instrumentation if sensor deployment costs must be minimized.

It is important to note that Gini-based feature importance in RF can exhibit a bias toward high-cardinality continuous features [33]. Future work should complement this analysis with

permutation importance—which measures accuracy degradation when feature values are randomly shuffled—to verify the stability of this ranking.

4. CONCLUSION

This study successfully compared the performance of three classification algorithms on a rice production prediction dataset. Based on the test results, the Random Forest algorithm proved to deliver the best performance with an accuracy of 95%, precision of 0.9571, recall of 0.95, and an F1-score of 0.95, far surpassing SVM (75%) and Naive Bayes (70%). These findings confirm that ensemble learning methods, particularly Random Forest, are the appropriate choice for crop yield prediction problems with high feature complexity.

REFERENCES

- [1] J. Asma, S. D, and K. D, “The Global Lifeline: A Staple Crop Sustaining Two-Thirds of the World’s Population,” *Agriculture Archives*, vol. 2, no. 3, pp. 15–18, Oct. 2023, doi: 10.51470/AGRI.2023.2.3.15.
- [2] R. A. Nagaraj *et al.*, “Comprehensive insights into the risks of climatic factors on rice production and its value chain- A Review,” *Plant Science Today*, vol. 11, no. sp4, Nov. 2024, doi: 10.14719/pst.5269.
- [3] T. Van Klompenburg, A. Kassahun, and C. Catal, “Crop yield prediction using machine learning: A systematic literature review,” *Comput. Electron. Agric.*, vol. 177, p. 105709, 2020.
- [4] UNICEF and others, “The state of food security and nutrition in the world 2023,” 2023.
- [5] Y. Xu *et al.*, “Genomic selection: A breakthrough technology in rice breeding,” *Crop J.*, vol. 9, no. 3, pp. 669–677, Jun. 2021, doi: 10.1016/j.cj.2021.03.008.
- [6] M. Shahhosseini, G. Hu, and S. V. Archontoulis, “Forecasting Corn Yield With Machine Learning Ensembles,” *Front. Plant Sci.*, vol. 11, Jul. 2020, doi: 10.3389/fpls.2020.01120.
- [7] B. S. Saruk and G. Mokesh Rayalu, “Optimizing rice yield prediction: A policy-oriented approach to agricultural sustainability,” *Results in Engineering*, vol. 28, p. 107844, Dec. 2025, doi: 10.1016/J.RINENG.2025.107844.
- [8] ENAYABA O.F, AJIBOYE A.A, EGWU L.S., AGBAMS P., and DAMOYE T., “TIME SERIES, MACHINE LEARNING AND ENSEMBLE MODELS TO FORECAST RICE CROP YIELD IN NIGERIA UNDER CLIMATIC VARIABILITY,” *International Journal of Agricultural Research and Biotechnology*, Nov. 2024, doi: 10.70382/tjarbt.v06i1.005.
- [9] R. L. Naylor, D. S. Battisti, D. J. Vimont, W. P. Falcon, and M. B. Burke, “Assessing risks of climate variability and climate change for Indonesian rice agriculture,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 19, pp. 7752–7757, May 2007, doi: 10.1073/pnas.0701825104.
- [10] S. Ha *et al.*, “Impacts of meteorological variables and machine learning algorithms on rice yield prediction in Korea,” *Int. J. Biometeorol.*, vol. 67, no. 11, pp. 1825–1838, Nov. 2023, doi: 10.1007/s00484-023-02544-x.
- [11] C. Li, J. Camac, A. Robinson, and T. Kompas, “Predicting changes in agricultural yields under climate change scenarios and their implications for global food security,” *Sci. Rep.*, vol. 15, no. 1, p. 2858, Jan. 2025, doi: 10.1038/s41598-025-87047-y.
- [12] D. HS, R. Girisha, and K. C. Gouda, “Study of rainfall forecasting and crop yield prediction under varying climatic conditions: a review of research gaps and advanced solutions,” *Theor. Appl. Climatol.*, vol. 157, no. 4, p. 225, Apr. 2026, doi: 10.1007/s00704-026-06126-w.
- [13] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, “Machine Learning in Agriculture: A Comprehensive Updated Review,” *Sensors*, vol. 21, no. 11, p. 3758, May 2021, doi: 10.3390/s21113758.
- [14] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine Learning in Agriculture: A Review,” *Sensors*, vol. 18, no. 8, p. 2674, Aug. 2018, doi: 10.3390/s18082674.
- [15] Y. Cai *et al.*, “Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches,” *Agric. For. Meteorol.*, vol. 274, pp. 144–159, Aug. 2019, doi: 10.1016/j.agrformet.2019.03.010.
- [16] K. Jhahharia, N. V Sharma, and P. Mathur, “A Machine Learning Model for Crop Yield Prediction Using Remote Sensing Data,” *International Research Journal of Multidisciplinary Scope*, vol. 06, no. 02, pp. 577–590, 2025, doi: 10.47857/irjms.2025.v06i02.03182.
- [17] R. Rathore, A. G. Said, P. Jangra, H. Murari, S. Bansal, and H. Singh, “Machine Learning in Agriculture and Crop Management: A Comprehensive Review,” in *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, IEEE, Mar. 2025, pp. 1–6. doi: 10.1109/IATMSI64286.2025.10984557.
- [18] S. Hariyanto, Indah Fenriana, Desiyanna Lasut, and Febrian, “Comparative Analysis of Support Vector Machine, Decision Tree, and Naive Bayes in Evaluating Machine Learning Effectiveness,” *RUBINSTEIN*, vol. 4, no. 1, pp. 1–15, Dec. 2025, doi: 10.31253/rubin.v4i1.4041.

- [19] M. I. KHALDI, A. ERRAISSI, M. HAIN, and M. BANANE, "In-Depth Comparison of Supervised Classification Models - Performance and Adaptability to Practical Requirements," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 8, 2025, doi: 10.14569/IJACSA.2025.0160862.
- [20] N. Nur, F. Wajidi, S. Sulhayanti, and W. Wildayani, "Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga," *Jurnal Komputer Terapan*, vol. 9, no. 1, pp. 58–64, Jun. 2023, doi: 10.35143/jkt.v9i1.5917.
- [21] N. Aditya, I. R. Munthe, and V. Sihombing, "A Comparative Analysis of Machine Learning Algorithms for Predicting Paddy Production," *sinkron*, vol. 8, no. 2, pp. 1200–1207, Apr. 2024, doi: 10.33395/sinkron.v8i2.13666.
- [22] J. Tan, L. Jiang, Y. Wei, N. Yao, G. Zhao, and Q. Yu, "Optimized Random Forest Framework for Integrating Cultivar, Environmental, and Phenological Interactions in Crop Yield Prediction," *Agronomy*, vol. 15, no. 10, p. 2273, Sep. 2025, doi: 10.3390/agronomy15102273.
- [23] S. P. Kar, P. Kumari, and R. Kumar, "Data-driven precision crop planning for enhanced rice yields and climate resilience in monsoon-dependent systems: A case study from Cuttack, India," *Plant Arch.*, vol. 25, no. 2, pp. 1417–1427, 2025.
- [24] Nila Aulia, Muhammad Rizky, and Chaerul Shaleh, "Memprediksi Jenis Tanaman Berdasarkan Parameter Lingkungan Menggunakan Algoritma Decision Tree dan Random Forest Berbasis Web," *Jurnal Teknik Informatika dan Teknologi Informasi*, vol. 5, no. 3, pp. 578–596, Dec. 2025, doi: 10.55606/jutiti.v5i3.6435.
- [25] I. S. Al-Mejibli, J. K. Alwan, and D. H. Abd, "The effect of gamma value on support vector machine performance with different kernels," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, p. 5497, Oct. 2020, doi: 10.11591/ijece.v10i5.pp5497-5506.
- [26] N. B. Nghien, C. N. Cong, N. N. Thi, and V. Q. Dung, "Comparison among search algorithms for hyperparameter of support vector machine optimization," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 14, no. 5, p. 3802, Oct. 2025, doi: 10.11591/ijai.v14.i5.pp3802-3815.
- [27] Y. Zhang, Z. Gao, E. Wittrup, J. Gryak, and K. Najarian, "Increasing efficiency of SVMp+ for handling missing values in healthcare prediction," *PLOS Digital Health*, vol. 2, no. 6, p. e0000281, Jun. 2023, doi: 10.1371/journal.pdig.0000281.
- [28] Y. Resti, D. A. Zayanti, N. R. Dewi, I. Yani, F. Darmawan, and Jimmy, "Classification of Corn Diseases and Pests Using Fuzzy Naïve Bayes Method," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, pp. 272–281, Oct. 2024, doi: 10.37934/araset.59.2.272281.
- [29] D. Mualfah, H. Rivaldi, J. Al Amin, and others, "KLASIFIKASI BUAH JERUK LEMON BERDASARKAN TINGKAT KEMATANGAN MENGGUNAKAN METODE SVM DAN NAIVE BAYES," *Journal of Software Engineering and Information System (SEIS)*, pp. 114–121, 2025.