Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 5, No. 3, September 2026, pp. 791–799
e-ISSN: 2964-2922, p-ISSN: 2963-6191

Open access

Original Research

K-Nearest Neighbors Classification Analysis of Rice Farmer Productivity Using Harvest Data for Agricultural Evaluation in Sayur Matinggi District

Ahmad Fauzan Purba [a], Mulkan Azhari [b*]

[a,b] Department of Information Systems, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

*Corresponding Author: mulkan@umsu.ac.id

ABSTRACT

Rice productivity evaluation in Sayur Matinggi District has traditionally relied on descriptive comparisons of harvest records, which do not provide a consistent classification of productivity levels across villages. This study develops a web-based analytical system using the K-Nearest Neighbors (KNN) algorithm to classify rice farmer productivity from harvest data collected in 19 villages during 2023–2025. The original wide-format dataset was transformed into 57 village-year observations. Productivity labels were constructed objectively using the 33rd and 66th quantiles of the production-to-harvested-area ratio, producing Low, Medium, and High classes. A stratified random split allocated 45 observations to training data and 12 observations to testing data. Three numerical attributes were used in the final model: harvested area, rice production, and cropping index. Min-Max normalization was applied before Euclidean-distance computation and majority voting. Evaluation of K values from 1 to 10 showed the highest accuracy at K=4 and K=5, both reaching 41.67%; K=4 was selected as the simpler model. At K=4, five of twelve testing observations were classified correctly. The confusion matrix showed a tendency to over-predict the Medium class, while macro-average precision, recall, and F1-score were 44.44%, 38.33%, and 34.43%, respectively. Functional black-box testing across twelve scenarios produced results consistent with the expected outputs. The findings indicate that the KNN workflow was implemented correctly, while the limited predictive accuracy is primarily associated with small sample size and weak class separability in the available harvest attributes.

How to Cite:
A. F. Purba and M. Azhari, "K-Nearest Neighbors Classification Analysis of Rice Farmer Productivity Using Harvest Data for Agricultural Evaluation in Sayur Matinggi District," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 5, no. 3, pp. 791–799, 2026.
Doi: https://doi.org/10.69916/jkbti.v5i3.512

KEYWORDS

K-Nearest Neighbors
Rice Productivity
Data Mining
Agricultural Evaluation
Classification

ARTICLE INFO

Received: May 26, 2026
Accepted: September 16, 2026
Available Online: September 20, 2026

1 INTRODUCTION

Agriculture remains a strategic sector for ensuring global food security, sustaining rural employment, and maintaining regional economic stability. In Indonesia, rice occupies a central position within the national food system; thus, the accurate evaluation of rice productivity is critical for directing agricultural interventions and optimizing the allocation of production resources [1]. However, rice productivity is a multidimensional construct determined not merely by harvested area, but by a complex interplay of cropping intensity, farm management practices, input efficiency, and environmental adaptability [2]. Consequently, a large harvested area does not automatically translate to an efficient or highly productive farming unit, necessitating more nuanced evaluation

metrics beyond simple volumetric comparisons.

At the local level, such as in the Sayur Matinggi District of South Tapanuli Regency, agricultural evaluation traditionally relies on descriptive statistics and annual average reports. This conventional approach fails to capture the heterogeneity of micro-level production dynamics across different villages and years. For instance, harvest-sampling data from 19 villages in this district (2023–2025) reveal substantial disparities in production scale and a general downward trend in output. Descriptive averages obscure these localized variations, preventing agricultural policymakers from identifying specific village-year observations that require targeted assistance, thereby limiting the analytical and operational value of the available data [3].

Data mining and machine learning offer systematic mechanisms to extract actionable patterns from complex numerical records. In agricultural contexts, classification algorithms can transform raw harvest observations into structured, comparable productivity tiers [4]. Among these, the K-Nearest Neighbors (KNN) algorithm is highly regarded for its intuitive, instance-based computational structure, making it particularly suitable for numerical agricultural datasets [5]. Recent studies have successfully applied KNN for crop yield prediction and disease classification [6–8]. However, a critical review of the existing literature reveals a significant research gap. Most agricultural KNN applications rely on large, pre-labeled datasets or focus on macro-level yield predictions. There is a distinct scarcity of frameworks designed to handle small-scale, multi-year, village-level harvest data where ground-truth productivity labels do not inherently exist. Furthermore, many existing studies treat the classification process as a "black box," failing to integrate the algorithmic output into a transparent, explainable Decision Support System (DSS) that local agricultural officers can readily interpret and trust [9–11].

To address these gaps, this study introduces a novel methodological and practical framework for micro-level agricultural evaluation. The primary novelty of this research lies in two aspects: (1) the implementation of a dynamic, quantile-based ground-truth labeling mechanism that objectively constructs Low, Medium, and High productivity classes from small-sample, multi-year data without relying on arbitrary static thresholds; and (2) the end-to-end integration of the KNN algorithm into an explainable, web-based Decision Support System (SIKLAPAD). Unlike conventional studies that stop at algorithmic accuracy, this framework provides methodological transparency by embedding calculation traces and black-box validated interfaces, allowing non-technical stakeholders to understand the spatial proximity and voting logic behind each classification [12].

Therefore, this study analyzes rice farmer productivity in Sayur Matinggi District using the KNN algorithm. The research encompasses the transformation of wide-format harvest data into longitudinal village-year observations, objective label construction using the 33rd and 66th quantiles, and stratified data partitioning. The KNN model is optimized through Min-Max normalization, Euclidean distance computation, and rigorous K-value evaluation [13]. The model's performance is comprehensively assessed using accuracy, confusion matrix, precision, recall, and F1-score. Finally, the analytical workflow is deployed and validated through functional black-box testing within the SIKLAPAD web system, ensuring that the algorithmic outputs are both scientifically robust and practically applicable for local agricultural evaluation [14, 15].

2 RESEARCH METHODS

2.1 Data Source and Research Variables

The dataset consists of harvest-sampling records from 19 villages in Sayur Matinggi District for the 2023–2025 period. The source data initially used a wide format in which each village occupied one row and annual production values were stored in separate year columns. The data were transformed into a long format so that every row represented one village-year observation. This transformation produced 57 records.

The initial research plan considered harvested area, production, cropping index, seed, and fertilizer. In the final implementation, only harvested area, rice production, and cropping index were used as KNN features because seed and fertilizer data were not available in the harvest-sampling source. This distinction is important because the implemented model must reflect the attributes actually observed in the dataset rather than variables that were only planned conceptually.

Table 1. Example of transformed harvest data

Year	Village	Harvested area (Ha)	Production (ton)	Cropping index
2023	Aek Badak Jae	102	1487.16	2.7
2024	Aek Badak Jae	102	1432.08	2.7
2025	Aek Badak Jae	102	1321.92	2.7
2023	Sipange Godang	41	553.50	2.5
2024	Sipange Godang	41	533.00	2.5
2025	Sipange Godang	41	492.00	2.5

2.2 Ground-Truth Productivity Categories

Because KNN is a supervised-learning algorithm, the training and testing observations require known reference labels. The source harvest records did not contain productivity classes; therefore, the study generated ground-truth labels from the productivity ratio. For each observation, the ratio was calculated as production divided by harvested area [16]:

$$R = \frac{P}{L} \tag{1}$$

The 57 ratio values were ordered and divided using the 33rd and 66th percentiles. Observations at or below the 33rd percentile were labelled Low; observations above the 33rd percentile and at or below the 66th percentile were labelled Medium; and observations above the 66th percentile were labelled High. Quantile-based thresholds were selected to avoid manually imposed cut-off values and to adapt the categories to the empirical distribution of the available dataset.

2.3 Training and Testing Split

An initial year-based split was considered but rejected because most observations from the same year exhibited similar productivity ratios, causing the testing set to concentrate in one category. The final procedure therefore used stratified random sampling with an 80:20 proportion so that Low, Medium, and High categories remained represented in both subsets. The resulting split contained 45 training records and 12 testing records [17–19].

Table 2. Stratified training and testing distribution

Category	Training	Testing	Total
Low	18	5	23
Medium	16	4	20
High	11	3	14
Total	**45**	**12**	**57**

2.4 Min-Max Normalization and KNN Classification

KNN depends directly on distances between observations. Because harvested area, production, and cropping index have different numerical ranges, Min-Max normalization was applied to place all three features on the interval from zero to one. Without scaling, variables with large numeric magnitudes could dominate the distance calculation [20]:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{2}$$

After normalization, the distance from each testing observation to every training observation was calculated using Euclidean Distance. This metric measures geometric proximity in multidimensional numerical space and is commonly used with KNN classification [21]:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \tag{3}$$

All training observations were sorted from the smallest distance to the largest. The first K observations were selected as the nearest neighbors, and the predicted category was determined through majority voting. When voting produced a tie, the category associated with the closest neighbor was selected. The value of K was evaluated from 1 through 10 because inappropriate values can increase sensitivity to local noise or blur boundaries between classes [22].

2.5 Model and System Evaluation

The study evaluated K values from 1 to 10 by repeatedly classifying the testing set and calculating accuracy for each value. The highest accuracy was used as the principal selection criterion, with the smaller K preferred when two values produced the same maximum result. Additional model evaluation used a confusion matrix

together with class-wise precision, recall, and F1-score so that the pattern of classification errors could be examined beyond overall accuracy [23]:

$$\text{Accuracy} = \frac{\text{Correct predictions}}{\text{Total testing observations}} \times 100\% \tag{4}$$

The web system was implemented using PHP Native, MySQL-compatible data access, Tailwind CSS, JavaScript, jQuery, Chart.js, DataTables, and SweetAlert2. Database design details are intentionally omitted from this journal article because the analytical workflow, model behavior, and implemented user-facing results constitute the primary research focus. Functional validation used black-box testing on authentication, data management, import, data splitting, K evaluation, classification, calculation details, reporting, and security validation.

3 RESULTS AND DISCUSSION

3.1 Implemented Classification Interface

The final web implementation provides an administrator workflow for managing productivity data, preparing training and testing subsets, selecting K, executing KNN classification, reviewing detailed calculations, examining evaluation results, and generating reports.

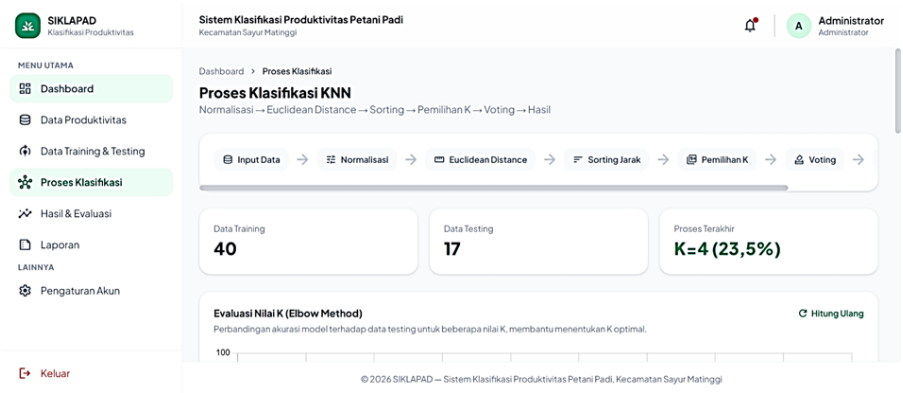

Figure 1. Classification process page after execution with K=4

Figure 1 presents the main KNN classification page after the process has been executed. The interface summarizes the number of training and testing observations and records the selected K value. It also provides a visual evaluation area for comparing candidate K values. This page operationalizes the methodological sequence from data preparation to model execution and gives the administrator a single control point for rerunning the classification when the data split or K value changes.

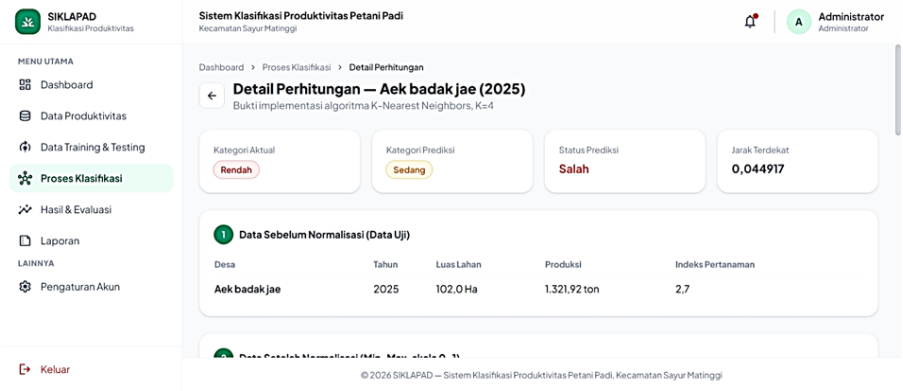

Figure 2. Detailed KNN calculation page for an individual testing observation

Figure 2 displays the calculation trace for one testing observation. The page shows the actual category, predicted category, prediction status, and nearest-neighbor distance, followed by the attributes used in the calculation. Its principal function is methodological transparency: the user can inspect how a specific decision was produced instead of receiving only a final class label. This traceability is particularly important when the model produces an incorrect classification because the nearest observations and voting pattern can be examined directly.

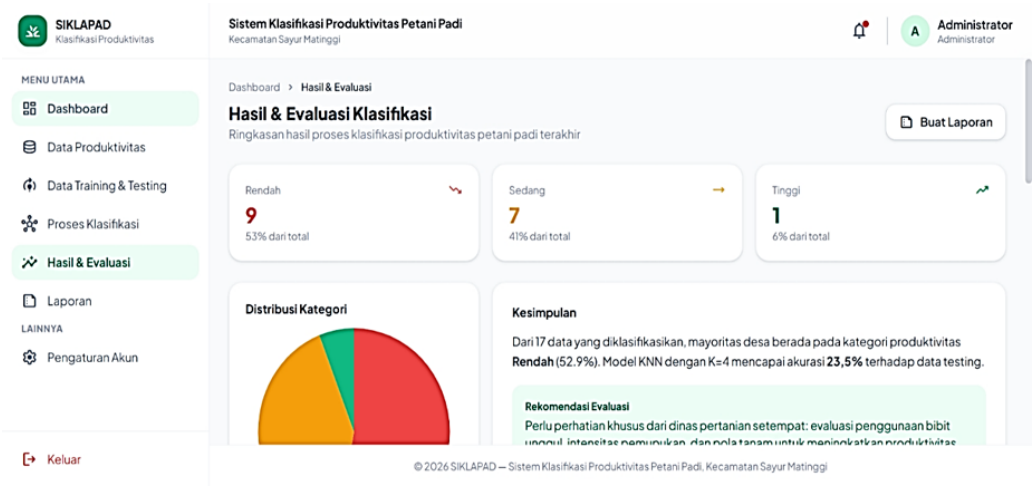

Figure 3. Classification result and evaluation dashboard

Figure 3 summarizes the distribution of predicted productivity categories and provides an automatically generated conclusion and evaluation recommendation. The visualization reveals the dominance of the Medium prediction class in the final test run. The result page therefore serves two purposes: it communicates the classification outcome to non-technical users and highlights patterns that require further interpretation, especially when class predictions are concentrated disproportionately in one category.

Figure 4. Classification report history and export interface

Figure 4 presents the reporting module that stores historical classification runs together with K values, training size, testing size, and achieved accuracy. The interface also provides print/PDF and spreadsheet export options. Recording multiple runs is useful for comparing model behavior after changes in the training-testing split or K selection and prevents the evaluation from being reduced to a single undocumented experiment.

3.2 Selection of K

The selection of the parameter K is a critical determinant of KNN model behavior, as it directly governs the trade-off between model sensitivity to local noise (high variance at low K) and the smoothing of class boundaries (high bias at high K) [7]. An overly small K causes the classifier to be dominated by the nearest single neighbor, making predictions vulnerable to outliers and measurement error. Conversely, an excessively large K dilutes the influence of genuinely similar training instances, causing the decision boundary to collapse toward the majority class distribution. To identify an appropriate neighborhood size for the harvest dataset, the study systematically evaluated K values ranging from 1 to 10. For each candidate K, the trained model classified the 12 stratified testing observations, and the resulting accuracy was recorded. Table 3 summarizes the classification outcomes across the evaluated K values.

Table 3. Accuracy evaluation for K=1 to K=10

K	Correct predictions	Accuracy (%)
1	2	16.67
2	2	16.67
3	4	33.33
4	**5**	**41.67**
5	5	41.67
6	2	16.67
7	4	33.33
8	3	25.00
9	3	25.00
10	3	25.00

The results reveal a distinctly non-monotonic accuracy pattern across the evaluated K values, which is consistent with the behavior of instance-based classifiers operating on small, overlapping datasets. The highest accuracy of 41.67

It is important to interpret the 41.67

3.3 Classification Results at K=4

The final $K = 4$ experiment classified 12 testing observations. Five observations matched their ground-truth labels, resulting in 41.67

Table 4. Final classification of the testing set at K=4

Village	Year	Actual	Predicted	Nearest distance	Status
Aek Badak Jae	2024	Medium	High	0.022459	Incorrect
Bange	2025	Low	Medium	0.004893	Incorrect
Huta Pardomuan	2023	High	Medium	0.032367	Incorrect
Jm. Baringin	2025	Low	Medium	0.028950	Incorrect
Jm. Baringin	2023	Medium	Medium	0.014475	Correct
Kel. Sayurmatinggi	2025	Low	Medium	0.074422	Incorrect
Lumban Huayan	2023	High	Medium	0.025321	Incorrect
Silaiya	2025	Low	Low	0.000391	Correct
Sipange Godang	2024	Medium	Medium	0.016718	Correct
Sipange Godang	2023	Medium	Medium	0.025076	Correct
Sipange Siunjam	2023	High	Medium	0.004404	Incorrect
Somanggal Parmonangan	2025	Low	Low	0.034789	Correct

The classification results at $K = 4$ reveal a pronounced asymmetry between the distribution of actual and predicted classes. Of the twelve testing observations, five were correctly classified (41.67

A class-level breakdown further elucidates the model's limitations. The High productivity class was the most severely affected: all three High observations (Huta Pardomuan 2023, Lumban Huayan 2023, and Sipange Siunjam 2023) were incorrectly predicted as Medium, yielding a recall of 0

3.4 Confusion Matrix and Class-Level Metrics

To obtain a more diagnostic view of the KNN model's behavior, the confusion matrix was constructed for the $K = 4$ configuration, cross-tabulating actual productivity labels against predicted labels.

Table 5. Confusion matrix for K=4

Actual / Predicted	Low	Medium	High	Total
Low	2	3	0	5
Medium	0	3	1	4
High	0	3	0	3
Total	**2**	**9**	**1**	**12**

The confusion matrix reveals a strongly asymmetric prediction pattern. Of the twelve testing observations, the classifier assigned nine predictions to the Medium category, three to the Low category, and only one to the High category. The Medium productivity ratio acts as an attractor zone, absorbing observations that would otherwise be classified as Low or High.

The class-level performance metrics derived from the confusion matrix provide a more detailed assessment, as summarized in Table 6.

Table 6. Precision, recall, and F1-score by category

Category	TP	FP	FN	Precision (%)	Recall (%)	F1-score (%)
Low	2	0	3	100.00	40.00	57.14
Medium	3	6	1	33.33	75.00	46.15
High	0	1	3	0.00	0.00	0.00
Macro average	—	—	—	**44.44**	**38.33**	**34.43**

The Low productivity class exhibited the highest precision at 100

3.5 Functional Testing

Black-box testing covered twelve positive and negative scenarios. All scenarios produced system responses that matched expected behavior.

Table 7. Summary of black-box testing

No.	Test scenario	Observed result	Conclusion
1	Valid administrator login	Redirected to dashboard and session created	Passed
2	Incorrect password	Error message displayed	Passed
3	Protected page without login	Redirected to login page	Passed
4	Add valid productivity data	Data stored and category recalculated	Passed
5	Invalid productivity input	Request rejected with validation message	Passed
6	Delete productivity data	Record removed successfully	Passed
7	Import without a file	Import rejected with error message	Passed
8	K exceeds training size	Classification request rejected	Passed
9	Invalid CSRF token	HTTP 403 response returned	Passed
10	Export to Excel	Spreadsheet response generated	Passed
11	Open principal pages after login	All pages returned HTTP 200	Passed
12	Execute K=4 classification	41.67% accuracy and 12 results stored	Passed

The black-box testing results confirm that the SIKLAPAD system operates consistently with its functional specifications across all twelve evaluated scenarios, demonstrating robust authentication, data management, classification logic, and reporting dimensions.

3.6 Discussion

The findings of this study demonstrate that the KNN algorithm can be successfully operationalized within a web-based decision support system (SIKLAPAD) for micro-level agricultural evaluation. However, the observed predictive performance, peaking at 41.67

The pronounced bias toward the Medium class and the complete failure to identify High-productivity observations (0

Despite modest accuracy, this study introduces significant methodological novelty: the implementation of a dynamic, quantile-based ground-truth labeling mechanism coupled with an explainable user interface. Unlike traditional agricultural evaluations relying on static productivity thresholds [1, 3], the 33rd and 66th quantile approach ensures adaptive classifications. By exposing the confusion matrix, class-level metrics, and individual calculation traces, SIKLAPAD empowers agricultural officers to interpret predictions with appropriate epistemic caution, preventing the misallocation of agricultural support resources.

Future research should address sample size limitations, temporal decline confounding effects, and expand feature sets by integrating high-resolution agronomic and environmental parameters.

4 CONCLUSION

This study successfully developed and validated a web-based rice productivity classification system (SIKLAPAD) for 19 villages in Sayur Matinggi District, utilizing harvest data from 2023 to 2025. The analytical workflow transformed wide-format records into 57 longitudinal village-year observations, generated objective productivity labels using the 33rd and 66th quantiles, and applied a stratified 80:20 data split. The KNN implementation, incorporating Min-Max normalization and Euclidean distance, achieved its optimal performance at $K = 4$ with an accuracy of 41.67

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Universitas Muhammadiyah Sumatera Utara for supporting this research.

DECLARATIONS

Author Contributions:

Ahmad Fauzan Purba: Conceptualization, Methodology, Software, Writing–original draft.

Mulkan Azhari: Data curation, Validation, Visualization, Supervision, Writing–review & editing.

Funding:

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest:

The authors declare no conflict of interest regarding the publication of this paper.

Data Availability:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AI Use Statement:
During the preparation of this work, the authors used the Qwen large language model in order to improve the language, readability, and academic phrasing of the manuscript. After using this tool, the authors thoroughly reviewed, edited, and verified all technical content, methodological descriptions, and analytical interpretations, and take full responsibility for the final content of the published article.

Additional Information:
No additional information is available for this paper.

REFERENCES

[1] A. Leiansyah and A. Putra, "Evaluasi Potensi Sektor Pertanian Provinsi Jambi dalam Perspektif Kontribusi dan Produktivitas," *Prosiding Diseminasi Ilmiah UM Jambi*, 2024.

[2] N. N. Khasanah and E. Y. A. Gunanto, "Pengaruh Luas Panen Padi, Produktivitas Lahan, Pertumbuhan Harga Beras dan Jumlah Penduduk terhadap Ketersediaan Beras di Indonesia tahun 1990–2022," *Diponegoro J. Econ.*, vol. 13, no. 2, pp. 67–79, Jun. 2024, doi: 10.14710/djoe.44900.

[3] D. A. Saputri, "Pemanfaatan Data Mining dalam Pengolahan Data Hasil Panen untuk Mendukung Pengambilan Keputusan Pertanian," *J. Sist. Inf. dan Komputasi*, vol. 8, no. 1, pp. 22–31, 2025.

[4] K. Mustaqim, N. Riza, and M. R. Hidayatullah, "Perbandingan Metode Regresi Linear Dan K-Nearest Neighbor (KNN) Dalam Memprediksi Produksi Tanaman Padi Di Pulau Sumatera," *Data Sci. Indones.*, vol. 4, no. 2, pp. 60–74, 2024.

[5] B. Y. Yana, Y. V. Via, and A. L. Nurlaili, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Daun Padi Menggunakan Ekstraksi HOG," *J. Algoritm.*, vol. 6, no. 1, pp. 110–120, 2025, doi: 10.35957/algoritme.v6i1.12154.

[6] H. N. Wismadi et al., "Klasifikasi Varietas Biji Kismis dengan Artificial Neural Network," *J. Optimasi Tek. Ind.*, vol. 5, no. 1, pp. 8–13, 2023.

[7] A. S. Putra, A. Irianti, and D. A. Heri, "Identifikasi Hama Dan Penyakit Pada Tanaman Kacang Tanah menggunakan Case Based Reasoning dan Algoritma K-Nearest Neighbor," *Konf. Nas. Ilmu Komput.*, pp. 32–38, 2021.

[8] T. Supianti, A. A. R. Ashril, L. P. I. Kharisma, and Khairunnazri, "Sistem Pakar Diagnosa Penyakit Bawang Merah Menggunakan Metode Forward Chainning," *Tek. Teknol. Inf. dan Multimed.*, vol. 1, no. 2, pp. 7–13, 2021, doi: 10.46764/teknimedia.v1i2.18.

[9] S. S. Sitanggang and A. S. Manalu, "Implementation of Additive Weighting in Providing Bidik Misi Scholarships at the Politeknik Bisnis Indonesia," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 2, no. 1, pp. 171–179, 2020, doi: 10.47709/cnahpc.v2i1.949.

[10] A. Prasetio, N. Mulyani, and F. M. Yuma, "Metode SAW dalam Penentuan Pemberian Kredit Calon Konsumen pada PT. Interyasa Mitra Mandiri," *J-Com (Journal Comput.)*, vol. 1, no. 1, pp. 65–72, 2021, doi: 10.33330/j-com.v1i1.1090.

[11] S. T. Ananda, A. Susano, and E. A. R. Pinahayu, "Design of A Decision Support System for Employee Recruitment Using the Simple Additive Weighting (SAW) Method Based on Java," *Int. J. Law Soc. Sci. Manag.*, vol. 1, no. 4, 2024, doi: 10.69726/ijlssm.v1i4.56.

[12] J. Olaniyan, D. Olaniyan, I. C. Obagbuwa, B. M. Esiefarienrhe, and M. Odighi, "Transformative Transparent Hybrid Deep Learning Framework for Accurate Cataract Detection," *Applied Sci.*, pp. 1–20, 2024.

[13] T. Bagaskara and R. Aulia, "A Hybrid YOLOv5-SqueezeNet Framework with Crop-Union Context Verification for Real-Time Motorcycle Rider Identification," *J. Comput. Technol.*, vol. 4, no. 1, pp. 34–42, 2026.

[14] S. Mehdipour, S. A. Mirroshandel, and S. A. Tabatabaei, "Vision transformers in precision agriculture: A comprehensive survey," *Intell. Syst. with Appl.*, vol. 29, p. 200617, 2026, doi: 10.1016/j.iswa.2025.200617.

[15] B. M. Abuhayi and A. A. Mossa, "Coffee disease classification using Convolutional Neural Network based on feature concatenation," *Informatics Med. Unlocked*, vol. 39, p. 101245, 2023, doi: 10.1016/j.imu.2023.101245.

[16] J. Guo, X. Guo, Q. Yi, H. Hao, and Y. Zhao, "Retinal vessel segmentation driven by structure prior tokens," *Med. Phys.*, vol. 52, no. 8, p. e18018, Aug. 2025, doi: 10.1002/mp.18018.

[17] M. T. Rahmawita, R. Riswandi, I. Maita, and Z. Zarnelly, "Analisis Kepuasan Mahasiswa dengan Metode EUCS dalam Penggunaan SIAFY Fakultas Tarbiyah dan Keguruan," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 8, no. 2, p. 201, Aug. 2022, doi: 10.24014/rmsi.v8i2.18487.

[18] M. Ghahremani, H. A. Otieno, M. Kazreen, and S. Shiaeles, "Machine Learning-Based Loan Approval Automation: Enhancing Efficiency, Accuracy and Fairness in Credit Decision-Making," *Expert Syst.*, vol. 43, no. 7, p. e70307, 2026, doi: 10.1111/exsy.70307.

[19] B. Imran, Zaeniah, Sriasih, S. Erniwati, and Salman, "Data Mining Using a Support Vector Machine, Decision Tree, Logistic Regression and Random Forest for..." *J. INFOKUM*, vol. 10, no. 2, pp. 792–802, 2022.

[20] P. P. Allorerung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *Prosiding Seminar Nasional*, 2024.

[21] M. A. Aprihartha, T. N. Alam, and M. Husniyadi, "Perbandingan Metrik Euclidean dan Metrik Manhattan untuk K-Nearest Neighbors dalam Klasifikasi Kismis," *J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 21–30, 2024.

[22] W. Nurazijah, R. Kurniawan, and Y. A. Wijaya, "Analisis Dampak Nilai K Optimal Terhadap Akurasi Pada Data Balita Puskesmas Cipaku," *JIKA (Jurnal Inform.)*, vol. 8, no. 2, pp. 197–203, 2024.

[23] D. N. Aini, B. Oktavianti, M. J. Husain, D. A. Sabillah, and T. Rizaldi, "Seleksi Fitur untuk Prediksi Hasil Produksi Agrikultur pada Algoritma K-Nearest Neighbor (KNN)," *J. Sist. Komput. dan Inform.*, vol. 3, no. 2, pp. 140–145, 2022.