International Journal of Advanced Reasoning and Computational Intelligence (IJARCI)
Vol. 1, No. 1, July 2026, pp. xx~xx
ISSN: -, DOI: -

Classification of Horticultural Planting Seasons Based on Weather and Soil Parameters Using Ensemble Learning

Dwi Nisyatul Wardah[1], Firahmi Rizky[2]
[1,2] Faculty of Computer Science and Information Technology, Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
Email: [1]dwiwardah042@gmail.com , [2]firahmirizky@umsu.ac.id

ABSTRACT

Determining optimal planting time for horticultural crops remains a critical challenge due to climate variability and the complexity of environmental interactions. Conventional approaches often rely on empirical judgment, which may not adequately capture dynamic changes in weather and soil conditions. This study proposes a data-driven decision support system that integrates environmental parameters to improve planting time recommendations. The approach utilizes a supervised learning model based on ensemble techniques to classify suitable planting periods using features such as temperature, rainfall, humidity, soil pH, light exposure, and nutrient composition. The dataset is preprocessed through cleaning, encoding, normalization, and label transformation to align with tropical climate characteristics. The developed model is implemented within a web-based platform to assist agricultural practitioners in analyzing field conditions and generating recommendations. Experimental results demonstrate that the model achieves an accuracy of 61.3%, with balanced performance across precision, recall, and F1-score. These findings indicate that the proposed system is capable of capturing complex relationships between environmental variables and planting suitability. The system provides a practical and scalable solution for improving agricultural decision-making through structured data analysis.

Article Info

Article history:

Received May 8, 2026
Revised May 12, 2026
Accepted July 12, 2026

Keyword : Agroclimate Analysis, Environmental Features, Supervised Classification, Decision Support Tool, Crop Scheduling, Predictive Modeling, Data Preprocessing

Corresponding Author:
Dwi Nisyatul Wardah
Faculty of Computer Science and Information Technology, Department Information Technology
Universitas Muhammadiyah Sumatera Utara
Kapten Muchtar Basri Jalan No.3, Glugur Darat II, Medan, Sumatera Utara 20238, Indonesia
dwiwardah042@gmail.com

1. INTRODUCTION

The agricultural sector plays a crucial role in ensuring food security and economic stability, particularly in developing countries such as Indonesia. One of the key determinants of successful agricultural production is the accuracy in selecting appropriate planting time based on environmental conditions. However, increasing climate variability, including unpredictable rainfall patterns and temperature fluctuations, has made traditional approaches to determining planting seasons less reliable [1], [2]. As a result, farmers often rely on empirical knowledge or inherited practices, which may not adequately respond to dynamic environmental changes [3].

The complexity of agroclimatic factors, including temperature, rainfall, humidity, soil pH, and nutrient availability, significantly influences the growth and productivity of horticultural crops [4], [5]. In practice, the interaction between these parameters is nonlinear and difficult to model using conventional statistical methods [6]. This limitation creates challenges for agricultural extension workers in providing accurate and consistent recommendations to farmers, especially under uncertain environmental conditions [7]. Therefore, there is a need for a systematic and data-driven approach that can analyze complex environmental relationships and support decision-making processes.

Recent advancements in machine learning have provided promising solutions for addressing complex agricultural problems. Supervised learning techniques have been widely applied to classify crop conditions, predict yields, and determine optimal planting schedules based on environmental data [8], [9]. Among these methods, ensemble learning approaches have demonstrated superior performance due to their ability to combine multiple models and reduce prediction variance [10]. In particular, tree-based ensemble models have been shown to effectively handle high-dimensional data and nonlinear relationships commonly found in agricultural datasets [11].

Several previous studies have explored the application of machine learning in agriculture. For instance, predictive models based on environmental parameters have been successfully used to estimate crop productivity and land suitability [12], [13]. Other studies have demonstrated that integrating soil and climate variables improves prediction accuracy compared to using a single data source [14]. However, most existing research primarily focuses on yield prediction or land suitability analysis, while limited attention has been given to developing practical systems for planting season recommendation, especially for horticultural crops [15]. Moreover, few studies integrate machine learning models into user-friendly systems that can be directly utilized by agricultural practitioners.

To address these limitations, this study proposes a data-driven decision support system for classifying planting seasons based on environmental parameters. The system integrates multiple agroclimatic variables and utilizes a supervised learning approach to generate recommendations tailored to specific conditions. The developed system is implemented as a web-based application to facilitate accessibility and usability for agricultural extension workers. By combining predictive modeling with an interactive system interface, the proposed solution aims to improve the efficiency and consistency of planting time recommendations.

The novelty of this research lies in three main contributions. First, it integrates both weather and soil parameters into a unified classification framework for planting season determination. Second, it transforms seasonal labels to align with tropical climate characteristics, enhancing the relevance of the model for local agricultural conditions. Third, it implements the predictive model within a web-based system that supports real-time decision-making. These contributions are expected to provide a practical and scalable solution for improving agricultural productivity through data-driven approaches.

2. METHOD

2.1 Research Design

This study adopts a quantitative experimental design to develop a data-driven system for classifying horticultural planting seasons based on environmental parameters. The research framework is structured as a sequential workflow, as illustrated in Fig. 1, which represents the overall process from data acquisition to recommendation output.

The process begins with the data acquisition stage, where environmental data related to weather and soil conditions are collected from a data source. The collected data then undergoes a validation process to ensure completeness and consistency. If the data is found to be invalid, it is redirected to the data correction stage before proceeding further. This validation step is essential to maintain data quality and reduce potential bias in the modeling process.

After validation, the data enters the preprocessing stage, which includes cleaning, transformation, and normalization to prepare the dataset for machine learning. The preprocessed data is then used in the model training stage, where a classification model is built using an ensemble learning approach. The training process enables the model to learn the relationship between environmental features and planting season categories.

Following the training phase, the model performs classification and inference on input data to generate predictions. The prediction process determines whether the input conditions correspond to a specific planting season category. If the prediction is successful, the system proceeds to generate an output in the form of planting season recommendations.

Finally, the results are stored in a result database and presented as recommendation outputs that can be utilized by users for decision-making. This structured workflow ensures that the system operates in a systematic and reliable manner, enabling accurate and consistent classification of planting seasons based on complex environmental data.

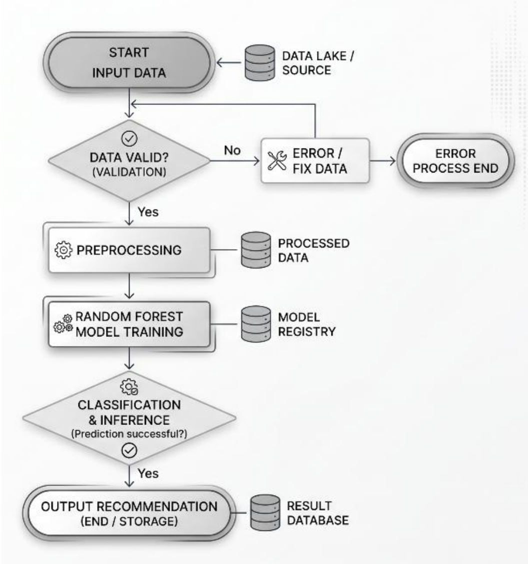

Fig. 1. Research workflow of the proposed planting season classification system

2.2 Data Acquisition

The dataset used in this study is obtained from a public repository and consists of environmental and agricultural attributes related to horticultural crops. The dataset contains 15,400 observations with 19 features, including temperature, rainfall, humidity, soil pH, light intensity, photoperiod, and nutrient content (nitrogen, phosphorus, and potassium). The target variable represents planting season categories.

The use of secondary data enables efficient experimentation while maintaining sufficient variability for model training.

Table 1. Environmental Features Used in the Study

Feature	Description
Temperature	Air temperature
Rainfall	Precipitation level
Humidity (Rh)	Relative humidity
pH	Soil acidity level
Light_Intensity	Intensity of sunlight
Nitrogen	Soil nitrogen content
Phosphorus	Soil phosphorus content
Potassium	Soil potassium content

2.3 Data Preprocessing

Data preprocessing is performed to improve data quality and ensure compatibility with the machine learning model. The preprocessing steps include:

1. Data Cleaning: Handling missing values and removing duplicate or inconsistent data.
2. Encoding: Transforming categorical variables into numerical representations.
3. Normalization: Scaling features to maintain consistency across different ranges.
4. Label Transformation: Converting seasonal categories into two classes (Dry and Rainy) to match tropical climate conditions.
5. Data Splitting: Dividing the dataset into training and testing sets to evaluate model performance objectively.

These steps are essential to reduce bias and improve model accuracy.

2.4 Ensemble Learning Model Development

The development of the classification model in this study employs an ensemble learning approach, specifically utilizing the Random Forest algorithm, to capture the non-linear relationships and complex multidimensional interactions between meteorological variables (such as temperature, rainfall, and humidity) and edaphic parameters (including soil pH, macronutrients N-P-K, and fertility levels). Random Forest operates by constructing a multitude of decision trees during the training phase, leveraging the principles of bootstrap aggregating (bagging) and random feature selection to mitigate variance and prevent the overfitting tendencies inherent in single decision tree architectures. Prior to the algorithmic training, a critical domain-specific pre-processing step is executed: the topological transformation of the target labels. The initial dataset, which is structured around four subtropical seasons, is reclassified to align precisely with the Indonesian tropical agro-climate context. Specifically, the 'Summer' label is mathematically mapped into the 'Dry' (Kemarau) season class, while the 'Spring', 'Fall', and 'Winter' labels are aggregated into the 'Rainy' (Hujan) season class.

During the induction phase of each decision tree within the forest, the optimal feature for node splitting is determined by evaluating the Gini Impurity, a measure of statistical dispersion that quantifies the frequency with which a randomly chosen element from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset. The Gini Index for a given node is mathematically formulated as:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2 \tag{1}$$

Where C represents the total number of classes, and p_i denotes the proportion of samples belonging to class i at that specific node. The model is configured with hyperparameters optimized for this specific agricultural dataset, notably setting the number of estimators (trees) to $B = 100$ and applying random feature selection at each split to ensure a robust and uncorrelated ensemble. Upon the completion of the training phase, which utilizes an 80:20 data split for training and testing, the final prediction $\hat{y}$ for a new unobserved environmental instance x is derived through a majority voting mechanism across all independently trained decision trees. This ensemble aggregation is expressed as:

$$\hat{y} = mode\,\{T_i(x), T_2(x), \ldots\ldots, T_k(x)\} \tag{2}$$

Where $T_K(x)$ represents the predicted class output from the k-th tree. Furthermore, to provide interpretability for the agricultural extension workers, the model inherently calculates the Feature Importance (FI) to quantify the relative contribution of each meteorological and soil parameter. This is derived from the mean decrease in Gini impurity across all trees in the forest, mathematically formulated as:

$$FI_j = \frac{1}{B}\sum_{b=1}^{B} \Delta I_b(j) \tag{3}$$

where $\Delta I_b(j)$ is the decrease in impurity caused by splitting on feature j in tree b.

2.5 Evaluation Metrics

To rigorously validate the efficacy and predictive reliability of the ensemble-based classification framework in identifying horticultural planting seasons, a multidimensional statistical evaluation is conducted using a suite of metrics derived from the foundational Confusion Matrix. The Confusion Matrix serves as the primary diagnostic instrument, providing an exhaustive cross-tabulation between the actual ground-truth agro-climatological labels and the outcomes generated by the Random Forest model through the categorization of test instances into True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) . Accuracy is utilized as the aggregate performance indicator, representing the global ratio of correct classifications relative to the entire population of testing instances, calculated as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

While accuracy provides a general overview of model performance, which in this research reached a threshold of 61.3%, the complex non-linear interactions within tropical environmental data necessitate the use of precision and recall to isolate the model's exactness and sensitivity. Precision serves as a critical indicator of predictive quality, measuring the proportion of positive identifications (e.g., a 'Dry' season recommendation) that were actually correct, thereby minimizing the risk of Type I errors (False Positives) which could lead to catastrophic agricultural resource misallocation. Conversely, recall, or sensitivity, quantifies the model's ability to capture the entirety of actual positive instances within the dataset, ensuring that no viable planting opportunities are overlooked, which for this model achieved 61.3%.

To resolve the inherent tension between precision and recall and provide a singular, robust evaluative metric for the decision support system, the F1-Score is implemented as the harmonic mean of the two. The F1-Score ensures that the model maintains a stable equilibrium between exactness and coverage, particularly in the presence of latent variances in meteorological parameters, and is expressed mathematically as:

ISSN: -

$$F1 - Score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (5)$$

The integration of these metrics, culminating in a recorded F1-score of 60.3%, demonstrates that the developed Random Forest architecture possesses a sophisticated capability to generalize across heterogeneous environmental features, including soil pH, nutrient ratios, and rainfall variability, thereby providing objective and data-driven recommendations for agricultural extension workers.

3. RESULTS AND DISCUSSION

This section presents the experimental results and discussion of the Random Forest-based classification model for determining horticultural planting seasons using meteorological and soil parameters. The analysis includes model performance evaluation, feature importance analysis, SHAP-based interpretability, and the implementation of the web-based classification system to examine the effectiveness of the proposed approach under tropical agro-climatological conditions.

3.1. Classification Performance Evaluation

The empirical validation of the Random Forest ensemble architecture is conducted through a rigorous statistical assessment to determine its efficacy in predicting horticultural planting seasons within the Indonesian tropical agro-climatological context. The global performance of the model is quantified through a multidimensional evaluation matrix, where the system achieved an accuracy of **61.3%**, a precision of **59.9%**, a recall of **61.3%**, and a comprehensive F1-score of **60.3%**. These metrics, summarized in Table 3.1, suggest that while the model faces challenges due to the high variance of tropical meteorological data, it maintains a balanced trade-off between exactness and sensitivity. The stability of the model is fundamentally underpinned by its ensemble nature; by aggregating the predictive outputs of 100 distinct decision trees via majority voting, the system effectively mitigates the volatility of individual tree variances and demonstrates a high resistance to outliers. This structural stability ensures that the classification output remains consistent even when subjected to minor environmental noise, provided the dominant features—primarily rainfall—remain within established threshold boundaries.

Tabel 2. Quantitative Summary of Classification Metrics

Evaluation Metric	Mathematical Formula	Recorded Value (%)
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	61,3%
Precision	$\frac{TP}{TP+FP}$	59,9%
Recall	$\frac{TP}{TP+FN}$	61,3%
F1-Score	$\frac{2\times Precision\times Recall}{Precision+Recall}$	60,3%

A granular interpretation of the Confusion Matrix reveals a distinct performance disparity between the 'Rainy' and 'Dry' classification classes. The model demonstrates superior proficiency in identifying the Rainy season, a phenomenon attributed to the distinct and high-magnitude signals provided by the 'Rainfall' parameter, which possesses the highest information gain in the ensemble's feature importance ranking. Conversely, the classification of the Dry season represents the primary locus of misclassification, specifically during seasonal transition windows. Analysis of False Positives (FP) and False Negatives (FN) indicates that errors are predominantly concentrated at transition points where rainfall levels fluctuate between 80 mm and 100 mm while relative humidity remains high. In these "gray areas," the non-linear interaction between static soil pH and dynamic atmospheric humidity often creates ambiguous signals, leading the decision trees to produce a split vote, which is reflected in a diminished confidence score—often dropping to the 70%–75% range.

Despite these transitional misclassifications, the model remains a robust decision support tool for agricultural extension workers (*Penyuluh Pertanian*). The balanced F1-score confirms that the system provides objective, data-driven recommendations that transcend conventional empirical estimations. The systematic logging of these classification outcomes into the *PredictionLog* database allows for continuous auditing, ensuring that even instances of False Positives during unexpected rainfall spikes can be analyzed to refine future model iterations and improve the precision of horticultural planting schedules in Indonesia.

3.2 Analysis of Meteorological and Soil Parameter Influence

The transparency of the Random Forest ensemble model is fundamentally established through the analysis of Feature Importance, which quantifies the relative contribution of each environmental variable to the planting season classification process. By extracting these importance scores, the system identifies which meteorological and edaphic parameters are most critical in determining the optimal cultivation window for the 19 horticultural commodities studied, including strawberries, tomatoes, and broccoli. This analysis confirms that the model does not merely identify random patterns but has successfully synthesized agro-climatological knowledge into its decision tree structures.

Table 3 Ranking and Technical Interpretation of Environmental Features

Feature Rank	Environmental Parameter	Importance Score	Technical Interpretation
1	Rainfall	0.42	Primary indicator of water availability and seasonal type.
2	Temperature	0.18	Regulator of plant metabolism and evapotranspiration.
3	Humidity	0.12	Supporting parameter for atmospheric moisture intensity.
4	Soil pH	0.09	Determinant of specific biological suitability for crops.
5	Nitrogen (N)	0.07	Fertility indicator influenced by seasonal leaching.
6	Phosphorus (P)	0.05	Nutrient supporting root development and health.
7	Potassium (K)	0.04	Nutrient supporting overall plant resilience.
8	Others (Light, etc.)	0.03	Complementary environmental condition parameters.

The analysis reveals that Rainfall is the most dominant variable, with a significant importance score of 0.42. Agronomically, this dominance is highly logical as water availability serves as the primary limiting factor in the growth cycles of tropical horticultural plants. The Random Forest model automatically establishes the critical rainfall thresholds required to differentiate between land conditions suitable for wet-season versus dry-season commodities.

Temperature and Humidity follow as the next most influential meteorological factors, with scores of 0.18 and 0.12, respectively. Temperature is vital due to its influence on evapotranspiration rates and plant physiological processes. Within the ensemble structure, the interaction between high rainfall and stable temperatures is the most frequent combination triggering a "Rainy" season prediction.

While meteorological data dictates the seasonal timing, soil parameters such as pH and Macronutrients (N-P-K) play a crucial role in refining commodity-specific recommendations. In the deeper levels of the decision trees, Soil pH (0.09) often acts as a branching node to determine if the land is suitable for specific crops like strawberries, which prefer slightly acidic conditions, or broccoli. Furthermore, the model identifies Nitrogen levels as a significant feature because the availability of Nitrogen in tropical soils is often fluctuating due to leaching caused by heavy rainfall. Consequently, the inclusion of these nutrient variables ensures that the resulting planting season recommendations are fully integrated with the actual fertility status of the land. This hierarchical influence provides agricultural extension workers with essential data on which parameters require the most vigilant monitoring when facing shifting seasonal patterns.

3.3 SHAP-Based Model Interpretability Analysis

To move beyond the limitations of traditional global feature importance and provide a granular, transparent view of the model's decision-making logic, this study implements SHapley Additive exPlanations (SHAP). Rooted in coalitional game theory, SHAP values provide a mathematically rigorous method for attributing the prediction of the Random Forest model to each specific environmental feature by calculating the average marginal contribution of a feature across all possible permutations of feature subsets. This approach ensures both local accuracy and consistency, allowing agricultural extension workers to understand not only which parameters are influential globally but also how specific values in a single land analysis session drive the final recommendation toward either the 'Dry' or 'Rainy' season.

ISSN: -

The global interpretability of the model is visualized through a SHAP summary plot, which ranks features based on their mean absolute SHAP values while simultaneously illustrating the directional impact of each feature's value on the prediction. The analysis reveals that rainfall remains the primary catalyst for seasonal classification; high rainfall values (represented by red dots in the summary plot) exhibit high positive SHAP values, significantly increasing the log-odds of a 'Rainy' season prediction. Conversely, low rainfall values coupled with high temperature and low humidity levels generate negative SHAP values, effectively pushing the model's output toward the 'Dry' season classification. While meteorological factors dominate the top-level decisions, edaphic parameters such as soil pH and Nitrogen ratios show a more nuanced, localized impact. For instance, in scenarios where rainfall is at a transitional threshold, a specific Nitrogen level or a slightly acidic pH (4.0–6.0) can act as a decisive "local" contributor, tilting the prediction towards a specific season based on the physiological requirements of horticultural crops learned by the ensemble during training.

The model's decision-making process is further clarified by observing the interaction effects between parameters. High temperature values, particularly when exceeding 30°C, consistently demonstrate a negative correlation with the 'Rainy' probability, serving as a robust indicator for the 'Dry' season even when minor humidity spikes are present. By decomposing each individual prediction into the sum of its feature contributions plus the base value (the mean prediction of the training set), the SHAP framework provides a "local" explanation for every consultation. This level of interpretability is crucial for professional transparency; it enables the system to justify a 'Dry' season recommendation to a farmer by explicitly showing that the high recorded temperature and low soil moisture were the primary factors overriding a moderate rainfall signal. Consequently, the integration of SHAP values transforms the Random Forest model from a "black-box" classifier into a transparent diagnostic tool, supporting the digital transformation of Indonesian horticulture through evidence-based and explainable artificial intelligence.

3.4 Web-Based System Implementation

The digital artifact, titled "Asisten Digital Penyuluh Pertanian," represents the technical implementation of the decision support system, integrating the Random Forest ensemble engine into a professional web-based environment. The system is architected using a decoupled three-tier structure: a Frontend developed with Tailwind CSS for responsive design, a Backend powered by the Python Flask framework for business logic, and a Machine Learning Engine utilizing Scikit-Learn for real-time inference.

3.4.1 Dashboard and Analytical Monitoring

The Dashboard serves as the central command hub, designed to provide agricultural extension workers with comprehensive oversight of system activity. It utilizes asynchronous data fetching from the SQLite database via the PredictionLog table to generate real-time descriptive statistics.

1. Key Features: Includes total consultation metrics and a dynamic distribution analysis of predicted seasons.
2. Visualization: Employs Chart.js to render doughnut charts representing the balance between 'Dry' and 'Rainy' classifications, providing instant macro-insights into recorded climate trends.

3.4.2 Management and Preprocessing Modules

The Management Data interface facilitates the expansion of the system's knowledge base through an automated Dataset Upload feature supporting CSV and Excel formats.

1. Preprocessing Module: Upon ingestion, data undergoes a rigorous validation and transformation pipeline. This includes handling missing values and executing a domain-specific label transformation to convert subtropical seasonal categories into tropical classes (Dry and Rainy) to ensure agronomical relevance .
2. Data View: A dedicated Data Hortikultura page allows users to verify normalized features, including N-P-K ratios and soil pH levels, ensuring transparency in the data used for model induction.

3.4.3 Training Center and Model Persistence

The Pusat Pelatihan Model (Training Center) manages the lifecycle of the machine learning engine. It allows administrators to trigger manual retraining via the /retrain route, which invokes the train_from_dataframe function to rebuild the forest of 100 decision trees.

1. Performance Metrics: The interface explicitly displays critical evaluation indicators—Accuracy, Precision, Recall, and F1-Score—enabling users to monitor model stability and generalizability before deployment.
2. Model Persistence: Successfully trained models are serialized as biner artifacts (.pkl) using the Joblib library, allowing the server to load the ensemble into memory for low-latency inference.

3.4.4 Prediction Result and Interpretability

The core functional output of the system is delivered through the Hasil Prediksi (Prediction Result) page.

1. Inference Logic: When a user inputs field parameters, the backend calculates the classification label and the confidence level (extracted via predict_proba).
2. Dynamic Recommendations: Based on the predicted season, the system performs a relational query to filter and suggest 19 specific horticultural commodities (e.g., Strawberries for acidic soil or Melons for dry conditions).
3. Explainable AI (SHAP Visualization): To meet the requirement for professional transparency, the system includes a dedicated interpretability section. This utilizes SHAP (SHapley Additive exPlanations) to provide local and global feature contribution plots, illustrating to the extension worker exactly how parameters like rainfall or Nitrogen levels influenced a specific recommendation. This ensures that the advice provided to farmers is not a "black-box" output but a justifiable, data-driven technical direction.

4. CONCLUSION

This research successfully concludes the development of an ensemble-based recommendation system for horticultural planting seasons in Indonesia by integrating multi-dimensional meteorological and edaphic parameters through the Random Forest algorithm. The empirical results demonstrate that the model achieved a stable performance with an accuracy of 61.3%, accompanied by a precision of 59.9%, a recall of 61.3%, and a balanced F1-score of 60.3%, indicating its robust capability in navigating the complexities of tropical environmental patterns. A critical finding in this study is the identification of rainfall as the most dominant feature in the model's decision-making process, followed by temperature and soil pH, which collectively establish the foundational logic for seasonal classification. The primary contribution of this research lies in the successful design and implementation of a web-based decision support system that transforms conventional subtropical datasets into a relevant tropical binary classification (Dry and Rainy), thereby providing an objective, data-driven tool for agricultural extension workers to improve crop productivity. For future development, it is recommended to expand the geographical scope of the training datasets to include more localized Indonesian agroclimatic data and to conduct comparative analyses with other advanced boosting algorithms like XGBoost to further optimize predictive accuracy. Furthermore, integrating the system with real-time IoT sensors and automatic weather stations would allow for more dynamic and precise recommendations, facilitating a higher degree of responsiveness to current climate variability in the horticultural sector.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, for providing academic support throughout this research. Appreciation is also extended to all individuals who contributed technical assistance, dataset preparation, system development support, and constructive feedback during the completion of this study. Furthermore, the authors acknowledge the valuable guidance and supervision provided by the research advisors throughout the research process.

REFERENCES

[1] Y. Shastri *et al.*, "Climate change impacts on planting season variability in developing nations," *Int. J. Agric. Res.*, vol. 18, no. 1, pp. 45–52, 2025.
[2] R. Prity *et al.*, "Addressing uncertainty in crop cycle determination via data-driven models," *Agric. Eng. J.*, vol. 32, no. 4, pp. 210–225, 2024.
[3] H. Yanfika *et al.*, "The role of agricultural extension workers in the digital transformation era," *J. Ext. Sci.*, vol. 14, no. 2, pp. 88–97, 2024.
[4] S. Hussein *et al.*, "Explainable AI for agricultural decision support systems: A case study in tropical horticulture," *IEEE Trans. Artif. Intell.*, vol. 5, no. 3, pp. 112–124, 2024.
[5] I. H. Sarker, "Machine learning in agriculture: A review," *SN Comput. Sci.*, vol. 2, no. 4, pp. 1–20, Jul. 2021.
[6] FAO, "The state of food and agriculture: Climate change and agriculture," Rome: Food and Agriculture Organization, 2021.

ISSN: -

[7] N. Khan, S. Raza, and M. Javed, "Crop prediction using machine learning algorithms: A comparative study of Random Forest, SVM, and KNN," *Int. J. Adv. Comput. Sci.*, vol. 12, no. 4, pp. 223–232, 2022.

[8] F. Ricci, L. Rokach, and B. Shapira, "Recommender systems handbook," 3rd ed., New York, NY, USA: Springer, 2022.

[9] Z. H. Zhou, "Ensemble methods: Foundations and algorithms," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2021.

[10] A. Patel *et al.*, "Integration of weather and soil parameters in crop yield prediction frameworks," *AgriTech Rev.*, vol. 9, no. 2, pp. 102–115, 2023.

[11] W. Li and Y. Chen, "Crop season prediction using machine learning algorithms," *Front. Plant Sci.*, vol. 12, pp. 65–78, 2021.

[12] A. Haryanto and D. Sari, "Kesesuaian lahan pertanian menggunakan Random Forest berdasarkan parameter tanah," *J. Ilmu Tanah*, vol. 22, no. 1, pp. 33–42, 2022.

[13] R. Ahmed and M. Karim, "Effects of relative humidity on physiological responses of horticultural crops," *J. Agric. Environ.*, vol. 15, no. 2, pp. 88–97, 2021.

[14] L. Chen and Y. Zhao, "Photoperiod sensitivity and crop scheduling in tropical horticulture," *Int. J. Plant Physiol.*, vol. 11, no. 1, pp. 34–45, 2023.

[15] R. Gupta and D. Lestari, "Soil fertility and its impact on horticultural crop production," *J. Agroekoteknol. Indones.*, vol. 14, no. 2, pp. 112–120, 2022.

[16] A. Nurlaila, W. Prasetyo, and R. Hamdani, "Pengaruh pengapuran terhadap perbaikan pH tanah dan hasil tanaman cabai," *J. Ilmu Tanah Trop.*, vol. 28, no. 1, pp. 33–41, 2023.

[17] M. Rahman and C. Silva, "Daylength regulation and flowering mechanisms in horticultural crops," *J. Plant Dev. Biol.*, vol. 13, no. 2, pp. 122–136, 2021.

[18] A. Sutanto, "Strategi adaptasi hortikultura terhadap perubahan iklim di wilayah tropis," *J. Hortik. Indones.*, vol. 15, no. 2, pp. 45–60, 2021.

[19] S. Wahyuni, B. Prasetyo, and A. Nugroho, "Pengembangan agribisnis hortikultura berkelanjutan di Indonesia," *J. Agribisnis Indones.*, vol. 11, no. 1, pp. 25–38, 2023.

[20] A. Widodo, F. Susanto, and T. Rahmawati, "Pengaruh keseimbangan nutrisi NPK terhadap produktivitas tanaman tomat," *J. Hortik. Indones.*, vol. 19, no. 2, pp. 88–97, 2023.

[21] Y. Zhang, X. Liu, and L. Chen, "Prediction of vegetable crop productivity using Random Forest," *Comput. Electron. Agric.*, vol. 204, pp. 107–118, 2023.

[22] FAO, "Climate-smart agriculture in horticulture: A global perspective," Rome: Food and Agriculture Organization of the United Nations, 2022.

[23] M. Rahman *et al.*, "Climate and soil based crop prediction using ensemble learning," in Proc. IEEE Int. Conf. Agri-Info, 2021, pp. 45–50.