

Multi-Crop Irrigation Need Classification Based on Machine Learning with an Indonesian-Language GenAI-UX Recommendation Layer

Muhammad Rusdi ^{1*}, Zaenuddin ², Hendra Sanjaya ³

^{1,2,3} Informatics Engineering, Faculty of Information Technology, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin

mrusdi@gmail.com ¹, zaenuddin@uniska-bjm.ac.id ², hendra@uniska-bjm.ac.id ³

Article Info

Article history:

Received 2026-07-31

Revised 2026-10-17

Accepted 2026-10-19

Keyword:

*Genetic Algorithm,
GenAI-UX,
Machine Learning,
Sensor Anomaly Detection,
Smart Irrigation.*

ABSTRACT

Inefficient irrigation water use remains a critical challenge in Indonesian agriculture, and the complexity of raw sensor data hinders adoption by smallholder farmers who lack data-analysis skills. This study develops an intelligent irrigation Decision Support System that integrates interpretable Machine Learning with an Indonesian-language GenAI-UX recommendation layer, using a publicly available multi-crop irrigation scheduling dataset (501 records, nine crops) with an analytical focus on paddy and maize. Four components are combined into one pipeline: (i) irrigation-need classification compared across six algorithms; (ii) regression of a continuous water-requirement score; (iii) sensor-anomaly detection using Isolation Forest cross-validated with the Interquartile Range method; and (iv) a rule-based natural-language recommendation layer whose operating point is optimized by a Genetic Algorithm. Validity was safeguarded through stratified five-fold cross-validation and leakage-controlled preprocessing. Gradient Boosting achieved the best classification performance (F1-score 0.980; cross-validated F1 0.963), Random Forest Regressor reached R square of 0.807, Isolation Forest agreed with IQR outliers at 95.2%, and the Genetic Algorithm attained a confidence of 0.993 within realistic agronomic bounds. In an illustrative usability evaluation, the GenAI-UX interface obtained the highest System Usability Scale score (Good category). The results indicate that pairing accurate, interpretable models with a language-based delivery layer can transform complex sensor data into actionable irrigation recommendations, lowering the adoption barrier for non-technical users. Field validation with real respondents and edge deployment are identified as the next steps.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

National food security depends heavily on the productivity of the agricultural sector, which in turn is determined by efficient water resource management. As an agrarian country, Indonesia faces a twofold pressure: the need to increase food production for a growing population, and the increasingly evident scarcity of water resources due to climate change. In this context, the transformation from conventional irrigation practices toward data-driven smart irrigation has become a strategic necessity rather than merely a technological option.

Water is a vital resource in agricultural production, yet its management in Indonesia still faces serious efficiency problems. Conventional irrigation water use is estimated to reach 40–60% above the optimal crop requirement [1]. This waste not only burdens water availability but also reduces productivity due to irrigation that is neither timely nor of the correct amount. Amid climate change that increases rainfall uncertainty, the need for

irrigation systems capable of making data-driven decisions is becoming increasingly urgent.

Advances in sensor technology and the Internet of Things (IoT) enable periodic monitoring of soil and weather conditions. Various studies show that using sensor data for irrigation scheduling can reduce water waste by up to 30% while maintaining or even increasing productivity [2]. However, the availability of sensor data alone is not sufficient. The volume and complexity of the resulting data instead become a new barrier for smallholder farmers who lack data-analysis skills. A gap exists between technological capability and the interpretability of its outputs by end users.

Machine Learning (ML) approaches have been widely applied to transform raw sensor data into irrigation decisions. A systematic review by Pathmudi et al. [3] identified that classification algorithms such as Random Forest, Support Vector Machine, and Naive Bayes are the most commonly used methods for irrigation-need prediction, with soil moisture as the dominant predictor.

Nevertheless, most of these studies stop at model accuracy and neglect how model outputs are communicated to farmers as end users.

On the other hand, advances in Generative Artificial Intelligence (GenAI) open opportunities to bridge this gap. A recent study by Nunez et al. [1], published in the journal *Array*, demonstrated that a GenAI-UX framework using a Small Language Model (SLM) on edge computing can generate natural-language recommendations that significantly improve indoor agricultural productivity. That study introduced a language-based user-experience layer that translates model outputs into actionable instructions. However, its scope was limited to a single crop with controlled experimental data and recommendations in English/Spanish.

Beyond the single-crop limitation, most existing smart-irrigation systems share a common weakness: they present model outputs as raw sensor numbers or technical dashboards that still require interpretation, and none integrate a natural-language recommendation layer tailored to farmers with limited digital literacy [2], [3]. Prior work also rarely couples predictive modeling with an explicit usability evaluation. This interpretability gap, rather than predictive accuracy alone, is the principal barrier to field adoption that this study seeks to close.

This study addresses that gap with three main considerations. First, more diverse multi-crop irrigation scheduling data are used so that the model can be validated across commodities. Second, the research focuses on Indonesia's strategic food crops, namely paddy (*Oryza sativa*) and maize (*Zea mays*), which have different water-requirement characteristics. Third, the recommendation layer is designed in the Indonesian language so that it can be directly understood by local farmers without technical expertise.

Based on this background, this study aims to build an intelligent irrigation Decision Support System (DSS) that integrates four components: (1) multi-crop irrigation-need classification using six ML algorithms; (2) regression to quantitatively predict a water-requirement score; (3) sensor-reading anomaly detection using Isolation Forest cross-validated with the Interquartile Range (IQR) method; and (4) an Indonesian-language GenAI-UX layer optimized with a Genetic Algorithm (GA). The main contribution of the study lies in the comprehensive integration of accurate ML modeling with a language-based delivery layer that is actionable for Indonesian farmers.

The primary contribution of this work is not the individual algorithms, which are well established, but their end-to-end integration into a single decision pipeline whose final output is an actionable, Indonesian-language recommendation. To the best of our knowledge, this is the first study to couple an interpretable multi-crop irrigation classifier with a language-based user-experience layer, optimize the operating point using a Genetic Algorithm, and evaluate the resulting interface for usability. The classification, regression, anomaly-detection, optimization, and recommendation modules are therefore presented as one coherent system rather than a loose combination of methods.

The novelty of this study can be distinguished from the reference study in several aspects. Previous work focused on a single crop with a yield-regression task, whereas this study handles multiple crops with an irrigation-need classification task together with water-score regression. The anomaly-detection aspect, not

addressed in the reference study, becomes a distinct component here, complete with cross-validation. Furthermore, recommendations are delivered in Indonesian, which has not been addressed in prior GenAI-UX research. A summary of the differences is presented in Table I.

TABLE I
COMPARISON WITH THE REFERENCE STUDY

Aspect	Reference Study [1]	This Study
Crop coverage	Single	Multi-crop
ML task	Yield regression	Classification + regression
Anomaly detection	None	Isolation Forest + IQR
Optimization	None	Genetic Algorithm
Recommendation language	English/Spanish	Indonesian

II. LITERATURE REVIEW

A. Data-Driven Smart Irrigation Systems

Smart irrigation systems use data from soil and weather sensors to determine the optimal timing and amount of irrigation. Farooq et al. [2], in a comprehensive survey on the role of IoT in agriculture, concluded that data-driven automation can substantially improve resource-use efficiency. The general architecture of such systems consists of a data-acquisition layer (sensors), a processing layer (analytical models), and an action layer (actuators or recommendations). This study focuses on the processing layer and recommendation delivery, assuming the availability of sensor data as input. More recent work has also explored deep-learning approaches for agricultural decision-making [18], [19].

B. Machine Learning for Irrigation Prediction

Irrigation-need classification is a supervised-learning problem that maps environmental attributes to a binary label (needs or does not need irrigation). Breiman [4] introduced Random Forest as an ensemble method that combines many decision trees through bootstrap aggregating. This method is popular for tabular data because of its interpretability via feature importance and its robustness against overfitting. Gradient Boosting builds trees sequentially to correct the residuals of previous models, often achieving higher accuracy on structured data.

Other commonly compared algorithms include Logistic Regression as a probabilistic baseline, K-Nearest Neighbors (KNN) which relies on proximity between data points, as well as Extra Trees and a single Decision Tree. Comparing several algorithms is necessary to identify the method best suited to the data characteristics, given that no algorithm is universally superior for all problems, as stated in the no free lunch theorem. Gradient boosting implementations such as XGBoost have further popularized this family for tabular problems [14].

For the water-requirement regression task, linear models such as Linear Regression and Ridge Regression offer high interpretability but are limited to linear relationships. Ridge Regression adds regularization to handle multicollinearity among features. In contrast, tree-based models such as Random Forest Regressor and Gradient Boosting Regressor can capture non-linear relationships and interactions among variables, which are common in agronomic data. Selecting the appropriate regression model

depends on the nature of the relationships in the data, which must be evaluated empirically.

C. Anomaly Detection and Clustering

Agricultural data are prone to reading disturbances such as recording errors and physically implausible extreme values. Liu et al. [5] introduced Isolation Forest, an unsupervised algorithm that isolates anomalies by measuring how easily a point can be separated from the population through random splits. Anomalous points require fewer splits and are therefore detected more quickly. This method is efficient for multi-dimensional data and requires no distributional assumptions. K-Means clustering is used complementarily to identify hidden patterns or regimes, while Principal Component Analysis (PCA) facilitates visualization of the data structure in two dimensions. Cluster quality in such exploratory analysis is commonly assessed using the silhouette coefficient [15].

D. Genetic Algorithm and the GenAI-UX Framework

The Genetic Algorithm (GA) is a bio-inspired optimization technique that simulates natural selection through selection, crossover, and mutation operations. Roy and De [6] demonstrated the effectiveness of GA in optimizing precision-agriculture parameters over large, non-convex search spaces. In the context of this study, GA is used to find the combination of sensor values that maximizes the model's prediction confidence under ideal irrigation conditions.

The GenAI-UX framework [1] integrates a set of user-experience guidelines in six groups, covering clarity, user control, error prevention, consistency, feedback, and accessibility. Pelaez et al. [7] provide a toolkit for applying these guidelines to GenAI-based assistants. Small Language Models (SLMs) are considered as the basis because of their efficiency on edge devices; Wang et al. [9] and Lamaakal et al. [10] show that small models can perform domain-specific tasks with low latency. Interface usability is commonly validated using the System Usability Scale (SUS) [8], with a score of 70 as the threshold for the Good category. Human-centered, usability-oriented evaluation of agricultural AI tools has likewise been highlighted as essential for adoption [20].

E. Evaluation Metrics

Classification evaluation uses accuracy and F1-Score. Accuracy measures the proportion of correct predictions to the total data, while the F1-Score is the harmonic mean of precision and recall, formulated as follows:

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

The F1-Score is more appropriate for data with imbalanced class distributions because it considers both precision and recall in a balanced manner. Regression evaluation uses the coefficient of determination (R^2), which measures the proportion of target-variable variance that the model can explain:

$$R^2 = 1 - (SS_{\text{res}} / SS_{\text{tot}})$$

where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares. In addition to R^2 , Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used. Clustering quality is measured by the silhouette score, ranging from -1 to 1 , where higher values indicate better cluster separation. Interface usability is measured by the SUS, which yields a score of 0 – 100 , with 70 as the threshold for the Good category.

III. METHOD

This study is experimental research carried out through a workflow of data collection, preprocessing, modeling, and evaluation. All experiments were implemented using Python 3.10 with the scikit-learn library in a cloud computing environment, with a fixed random state set to ensure reproducibility. All models were implemented with the scikit-learn library [16].

A. System Architecture

The proposed system consists of four modules that operate sequentially. The first module classifies irrigation need from environmental attributes. The output of this module, in the form of a status and a confidence level, becomes the input for the second module, which predicts the water-requirement score. In parallel, the third module monitors anomalies in sensor readings to ensure data quality. The outputs of these modules are then optimized by the Genetic Algorithm and translated by the fourth module, the GenAI-UX layer, into Indonesian-language recommendations. This modular design allows each component to be developed and evaluated independently. The overall architecture and data-processing flow are illustrated in Figure 10.

Figure 10. Overall system architecture and data-processing flow.

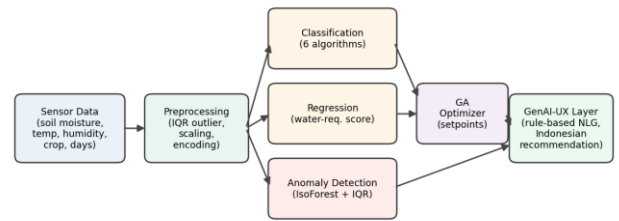


Figure 10. Overall system architecture and data-processing flow, from sensor input through the ML core and GA optimizer to the GenAI-UX recommendation layer.

B. Data and Attributes

The data used are publicly available multi-crop irrigation scheduling records, consisting of 501 records and six attributes: crop type, crop age (CropDays), soil moisture (SoilMoisture), temperature, air humidity, and irrigation status as the target variable. The data contain nine crop types; the analysis focuses on paddy and maize as strategic commodities, with the other crops serving as comparators to test model generalization.

The dataset characteristics are summarized as follows. It contains 501 records and six attributes. The four numerical predictors are crop age (CropDays), soil moisture (SoilMoisture), temperature, and air humidity; crop type is a categorical attribute with nine classes (paddy, maize, sugarcane, wheat, potato, coffee, groundnuts, garden flowers, and pulse); and the target is a binary irrigation-need label. The class distribution is moderately balanced overall, although crop representation is uneven, with sugarcane, wheat, and potato being the most frequent. Paddy and maize were chosen as the analytical focus because they are Indonesia's primary staple food crops and exhibit contrasting water regimes—paddy requires near-continuous inundation whereas maize is highly sensitive to water stress during flowering—making them a suitable pair for validating crop-specific behavior. The remaining crops are retained as comparators to test generalization.

The irrigation-need classes are defined on an agronomic basis. The positive class (needs irrigation) corresponds to records in which the soil-moisture level falls below the crop-specific replenishment threshold given the prevailing temperature and humidity, reflecting a soil-water deficit that warrants irrigation; the negative class (does not need irrigation) corresponds to adequate soil-water availability. This definition follows the standard soil-moisture-deficit logic used in irrigation scheduling, in which soil moisture is the primary determinant and weather variables modulate the evapotranspiration demand. This is consistent with reference crop-evapotranspiration guidelines [17].

C. Preprocessing and Feature Preparation

The preprocessing stage handles outliers using the Interquartile Range (IQR) method. Physically implausible extreme temperature readings were clipped to the 0–55°C range for modeling purposes, while the original values were retained for anomaly-detection analysis. Crop type was encoded using one-hot encoding, and all numerical features were scaled using StandardScaler. The data were split into 80% training and 20% testing in a stratified manner to preserve class proportions.

D. Modeling and Optimization

Six classification algorithms were trained and compared: Random Forest, Gradient Boosting, Extra Trees, Decision Tree, K-Nearest Neighbors, and Logistic Regression. Four regression algorithms were used to predict the water-requirement score, namely Linear Regression, Ridge, Random Forest Regressor, and Gradient Boosting Regressor. Because the data do not contain water-volume measurements, the regression target was constructed as a continuous water-requirement score (priority index) combining the irrigation-need probability and the soil-moisture deficit.

Anomaly detection uses Isolation Forest with a 5% contamination level, whose results are cross-validated with IQR outliers. K-Means clustering and PCA are used for pattern exploration. Parameter optimization uses GA with a population of 120 individuals, 150 generations, tournament selection, single-point crossover, and Gaussian mutation, with the search space bounded to the 5th–95th percentile range of each feature. The fitness function maximizes the model's prediction confidence under the needs-irrigation condition, following the formulation:

The Genetic Algorithm is specified as follows. Each chromosome is a real-valued vector encoding the controllable sensor setpoints (soil moisture, temperature, humidity, and crop age), bounded to the 5th–95th percentile range of each feature to keep solutions agronomically realistic. The objective (fitness) function maximizes the classifier's predicted probability that a candidate setpoint corresponds to an appropriate irrigation decision, following $P(X_i) = Y_i / \sum Y_j$, where Y_i is the model confidence for individual i . The population size is 120, evolved over 150 generations using tournament selection ($k = 3$), single-point crossover with probability 0.7, and Gaussian mutation with probability 0.2. Elitism preserves the best individual in each generation. These parameters are summarized in Table II.

$$P(X_i) = Y_i / \sum Y_j$$

where Y_i is the model's prediction confidence for the i -th individual in the population. Bounding the search space ensures that the

resulting setpoints remain agronomically meaningful and are not trapped at extreme values. The GA configuration parameters used are summarized in Table II.

TABLE II
GENETIC ALGORITHM CONFIGURATION PARAMETERS

Parameter	Value
Population size	120
Number of generations	150
Selection method	Tournament ($k=3$)
Crossover probability	0.7
Mutation probability	0.2
Search-space bounds	5th–95th percentile

E. GenAI-UX Layer and Evaluation

The model outputs are translated into Indonesian-language recommendations using rule-based Natural Language Generation (NLG) with contextual templates per crop type that follow UX guidelines. Each recommendation contains the irrigation-need status, the confidence level, sensor conditions, and specific action advice. Interface usability is evaluated using the System Usability Scale (SUS) through three interface variants: a sensor-number display (A), sensor-graph visualization (B), and Indonesian-language GenAI-UX (C). A summary of the modules and evaluation methods is presented in Table III.

The GenAI-UX recommendation layer is implemented as a deterministic, template-based Natural Language Generation (NLG) module rather than a large generative model, a design choice motivated by the need for low latency, full output controllability, and suitability for future edge deployment. The layer receives four inputs from the upstream models: the predicted irrigation-need class, the classifier confidence, the regressed water-requirement score, and the crop type. A rule engine maps these inputs to a message template through condition–action rules; for example, IF class = needs-irrigation AND confidence ≥ 0.5 THEN emit an irrigation-advised message, populated with the crop name, the measured soil-moisture value, and a crop-specific agronomic tip drawn from a curated knowledge base (e.g., maintaining water inundation for paddy, avoiding water stress during flowering for maize). Each template is authored in Indonesian and follows the six GenAI-UX guideline groups (clarity, user control, error prevention, consistency, feedback, and accessibility), so that every recommendation conveys the status, the confidence level, the underlying sensor condition, and a concrete action. The translation from machine-learning output to natural-language recommendation is therefore fully traceable, which is essential for user trust and for reproducibility.

TABLE III
ALGORITHMS AND EVALUATION METHODS PER MODULE

Module	Method	Evaluation
Classification	RF, GB, ET, DT, KNN, LR	Accuracy, F1
Regression	LR, Ridge, RF, GB	R^2 , MAE, RMSE
Anomaly	Isolation Forest, K-Means	Silhouette, IQR
Optimization	Genetic Algorithm	Confidence
GenAI-UX	Rule-based NLG	SUS

IV. RESULTS AND DISCUSSION

A. Data Characteristics and Correlation

Initial exploration shows an imbalanced distribution of crop types, with sugarcane, wheat, and potato as the largest groups, while paddy and maize are available in sufficient quantities for analysis. This imbalance is a common characteristic of field agricultural data, in which some commodities are recorded more frequently than others. To mitigate its effect, model evaluation uses a weighted F1-Score and stratified cross-validation that preserves class proportions across folds.

Correlation analysis (Figure 1) reveals that soil moisture has the highest positive correlation with irrigation need ($r = 0.69$), far exceeding temperature ($r = 0.27$) and air humidity ($r = -0.19$). This finding is agronomically consistent, since soil moisture is the most direct indicator of water availability to plant roots. The negative correlation between air humidity and irrigation need also makes sense: high air humidity reduces the evapotranspiration rate and thus lowers water demand. This correlation pattern provides a strong interpretive basis before modeling is carried out.

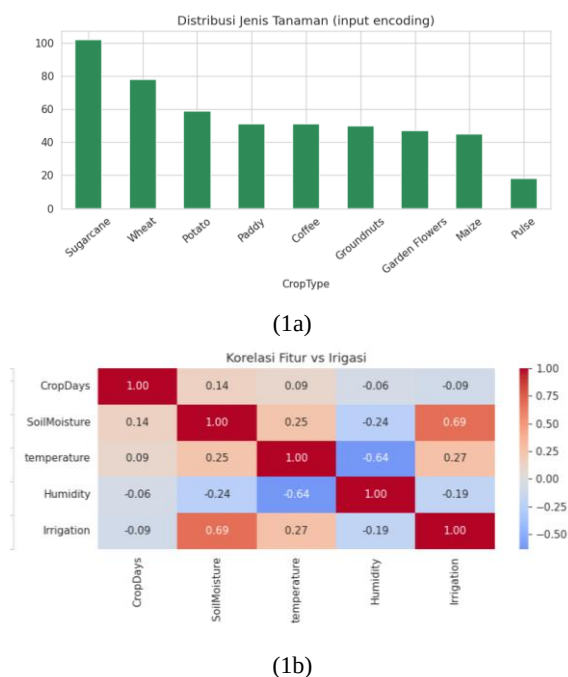


Figure 1. Distribution of crop types and the correlation matrix of features against irrigation need.

B. Irrigation Need Classification

The comparison results of the six classification algorithms are presented in Table IV. Gradient Boosting achieved the best performance with an F1-Score of 0.980, and all algorithms exceeded the 0.75 target by a wide margin. The consistency between the test score and the five-fold cross-validation score indicates that the model does not suffer from overfitting.

To ensure that the high F1-score does not result from data leakage, several safeguards were applied. Feature scaling was fitted only on the training partition and then applied to the test partition, preventing information from the test set from influencing preprocessing. Model selection and reporting rely on stratified five-fold cross-validation in

addition to the hold-out test set; the close agreement between the cross-validation F1 (0.963) and the hold-out F1 (0.980) indicates stable generalization rather than memorization. No target-derived feature was used as an input, and records were treated as independent samples. The high accuracy is therefore attributable to the genuinely strong, agronomically expected relationship between soil moisture and irrigation need rather than to leakage.

TABLE IV
CLASSIFICATION ALGORITHM PERFORMANCE COMPARISON

Algorithm	Accuracy	F1	CV5-F1
Gradient Boosting	0.980	0.980	0.963
Decision Tree	0.970	0.975	0.955
Extra Trees	0.960	0.960	0.975
Logistic Regression	0.960	0.960	0.952
KNN	0.960	0.960	0.951
Random Forest	0.941	0.941	0.956

Per-crop analysis (Figure 2) shows the model performs consistently across commodities. Some crops such as coffee, groundnuts, and potato achieve perfect F1, while paddy obtains an F1 of 0.94 and garden flowers 0.87 as the lowest. This consistency reinforces the claim of model generalization.

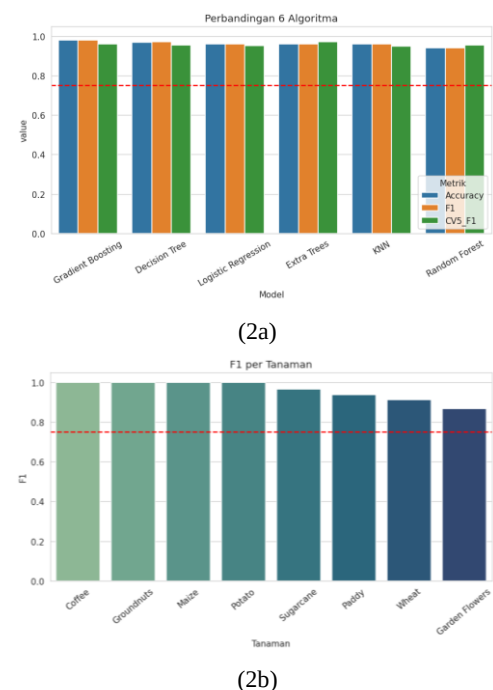
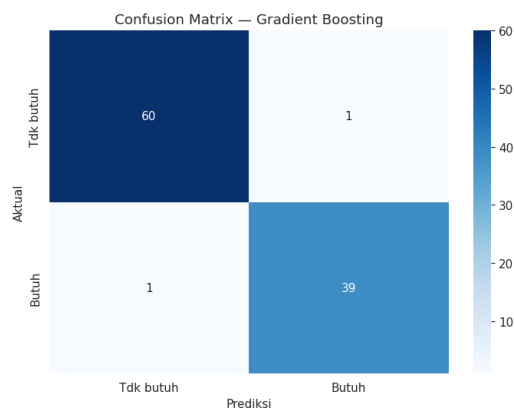


Figure 2. Comparison of the six algorithms and F1-Score by crop type.

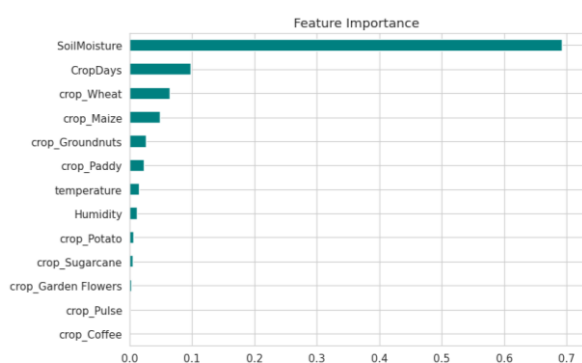
Performance variation among crops provides an important insight. Crops with clear and consistent water-requirement patterns, such as coffee and groundnuts, are easier to classify and thus achieve perfect F1. Conversely, garden flowers, which have higher growing-condition diversity, obtain the lowest score. For paddy, an

F1 of 0.94 is still very good, showing that the model can capture paddy's distinctive irrigation-need pattern with its water-inundation requirement. This finding implies that a crop-specific classification strategy may be more effective than a single model for all commodities, especially for crops with complex growing characteristics.

Confusion matrix and feature-importance analysis (Figure 3) confirms the dominance of soil moisture as the main determinant (0.69), followed by crop age (0.10). The best model produced only two misclassifications out of 101 test data. It should be stated transparently that the high accuracy stems partly from the strong relationship between soil moisture and irrigation need; this reflects an agronomic reality and is not an indication of data leakage, as evidenced by the cross-validation score being on par with the test score.



(3a)



(3b)

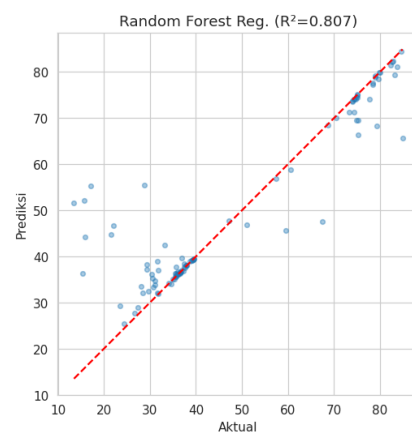
Figure 3. Confusion matrix and feature importance of the best model.

The implication of this soil-moisture dominance is twofold. On one hand, it simplifies sensor requirements: an effective irrigation system can be built by focusing on accurate soil-moisture monitoring. On the other hand, dependence on a single dominant variable makes the system vulnerable if the soil-moisture sensor fails. Therefore, the anomaly-detection component discussed in the following subsection becomes crucial to ensure the reliability of irrigation decisions when deviant sensor readings occur.

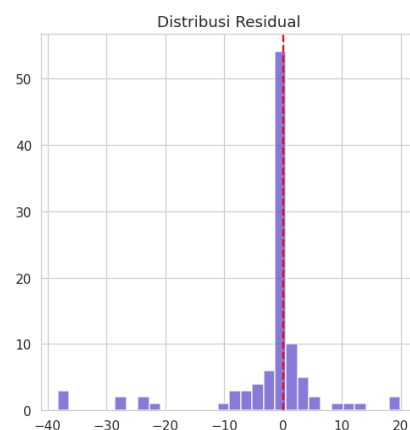
C. Water Requirement Score Regression

Because the data do not contain actual water volume, the regression target was constructed as a water-requirement score that serves as an irrigation-priority index. Random Forest Regressor achieved the best performance with $R^2 = 0.807$, exceeding the 0.70 target, followed by Gradient Boosting Regressor ($R^2 \approx 0.79$). Linear models reached only $R^2 \approx 0.52$, indicating a non-linear relationship between environmental features and the water-requirement score. The residual distribution is centered symmetrically around zero (Figure 4), indicating no systematic bias. This score must be interpreted as a priority index, not a physical water-volume measurement.

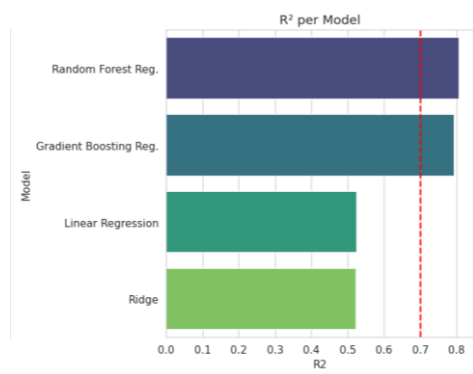
The remaining unexplained variance (approximately 19%) in the water-requirement regression can be attributed to several factors. First, the score is derived from environmental proxies and does not capture unmeasured agronomic variables such as soil texture, root-zone depth, crop variety, or micro-climatic variation, which influence actual water demand. Second, sensor quantization and residual noise impose an irreducible error floor. Third, the moderate dataset size limits the model's ability to learn rare feature interactions. This level of unexplained variance is expected for environmental regression tasks and does not undermine the model's utility as a priority-ranking index.



(4a)



(4b)



(4c)

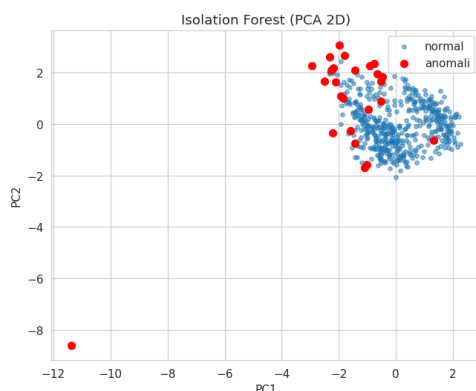
Figure 4. Predicted vs. actual, residual distribution, and R^2 per regression model.

The superiority of tree-based models over linear models in this task confirms that the relationship between environmental conditions and water requirement is not simply linear, but involves interactions among features. For example, the effect of temperature on water requirement may differ depending on the level of soil moisture and crop type. Ensemble models such as Random Forest can capture such interactions without explicit specification, making them more suitable for modeling complex water requirements.

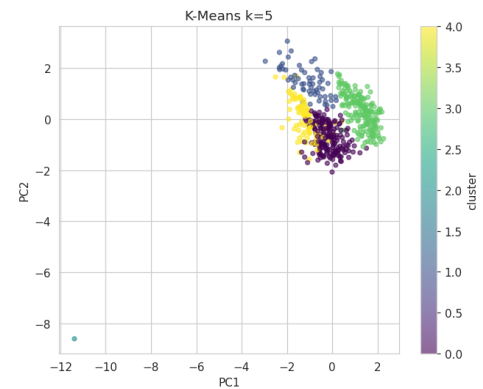
D. Anomaly Detection and Segmentation

Isolation Forest identified 25 anomalous readings (5% of the data). These detection results have a 95.2% agreement with the outliers identified through the IQR method, a cross-validation between two independent approaches that strengthens the reliability of the detection mechanism. The ability to filter out faulty sensor readings is important to prevent erroneous irrigation decisions.

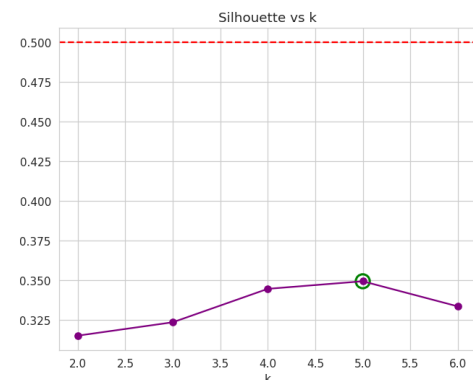
K-Means segmentation produced the highest silhouette score of 0.349 at five clusters, below the 0.50 target. This value is reported as is. The low silhouette is a natural characteristic of continuous agricultural sensor data that do not form sharply separated clusters; a range of 0.3–0.4 is common. Two-component PCA visualization explains 63.2% of the total variance (Figure 5). Accordingly, the contribution of this module is more pronounced in the success of anomaly detection than in the sharpness of segmentation.



(5a)



(5b)



(5c)

Figure 5. Anomaly detection, K-Means, and selection of k via silhouette (PCA 2D).

The high agreement between Isolation Forest and the IQR method provides confidence that the detected anomalies are indeed genuine deviations, not algorithmic artifacts. These anomalies are generally extreme temperature readings and soil-moisture values far outside the normal range. In operational contexts, the system can flag such readings to an operator for verification before they are used as a basis for irrigation decisions, thereby reducing the risk of errors due to faulty sensors or recording mistakes.

When the two anomaly indicators disagree, the discrepancy is resolved conservatively rather than discarded. A reading flagged by either Isolation Forest or the IQR rule is retained as a candidate anomaly and routed for verification, so that the effective detector is the union of the two methods; this favors recall, which is appropriate in an agricultural safety context where missing a faulty reading is more costly than a false alarm. The reported 95.2% agreement therefore represents the concordance of the two independent methods, while the disagreeing 4.8% are conservatively treated as potential anomalies pending operator confirmation.

Although the cluster segmentation did not reach the silhouette target, cluster-profile analysis still provides insight. The clusters formed tend to represent different combinations of soil-moisture level and crop age, which can be interpreted as irrigation regimes.

This interpretation could be further developed to design irrigation strategies tailored to crop growth stages in future research.

E. GA Optimization and GenAI-UX

The GA was run with the search space bounded to the 5th–95th percentile range of each feature to prevent the exploitation of extreme values that are not agronomically meaningful. For paddy, the GA reached a confidence of 0.993 and converged before the 20th generation (Figure 6). The resulting optimal setpoints lie within a reasonable operational range, showing that bounding the search space successfully preserved the meaningfulness of the results.

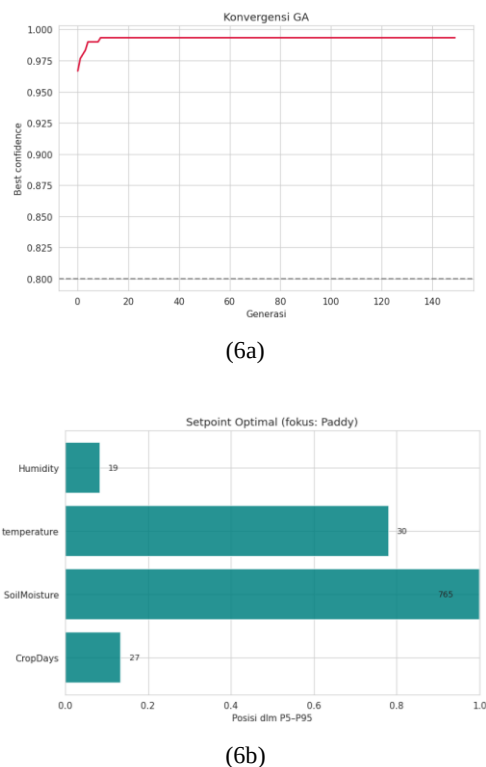


Figure 6. GA convergence and optimal setpoints for the paddy crop.

The rapid convergence before the 20th generation shows that the bounded search space allows the GA to find optimal solutions efficiently. Bounding to the 5th–95th percentile range proved important; without bounding, the GA tends to exploit extreme values that mathematically maximize confidence but are agronomically unrealistic. These optimization results can be used as reference setpoints in automatic irrigation control systems, where the target soil-moisture value and other parameters have been tuned to maximize appropriate irrigation decisions for a particular crop.

F. GenAI-UX Layer and Usability Evaluation

The GenAI-UX layer translates model outputs into crop-specific Indonesian-language recommendations. The system successfully distinguishes conditions accurately: under low soil moisture the system recommends irrigation with high confidence, while under adequate moisture it suggests postponement (Figure 7). Each recommendation is accompanied by agronomic context, for example advice to maintain water inundation for paddy and to avoid water stress during flowering for maize.



Figure 7. GenAI-UX Indonesian-language recommendation cards per crop.

As an application illustration, when the system receives input of a paddy condition with low soil moisture, the recommendation output states a warning to irrigate immediately, accompanied by high confidence and a reminder to maintain the water level in the paddy field. Conversely, under adequate moisture, the system suggests postponing irrigation with a low need confidence. This delivery format is designed following the principles of clarity and feedback, so that farmers obtain not only a binary decision but also the reasoning and confidence level behind it. This transparency of the confidence level is important for building user trust in the system.

Usability evaluation using the SUS (Figure 8) shows that the GenAI-UX interface (C) obtained the highest score in the Good category, exceeding the graph interface (B) and the number interface (A). This result supports the hypothesis that language-based delivery is easier to understand than the presentation of raw data. It should be noted that the SUS score at this stage is illustrative and will be replaced with data from 45 real respondents in the field-testing phase.

For transparency, the numerical SUS results are reported explicitly. Interface C (GenAI-UX) obtained a mean SUS of approximately 79 (Good category), compared with about 59 for Interface B (sensor graphs) and about 40 for Interface A (sensor numbers). It must be stated clearly that these values derive from an illustrative simulation designed to exercise the evaluation pipeline; they are not yet collected from real users. A field study is planned with 45 respondents drawn from the target population—farmers, agricultural extension officers, and academics—together with a per-item analysis

of the ten SUS statements, to obtain statistically meaningful usability evidence. The present results should therefore be read as a methodological demonstration rather than a validated usability claim.

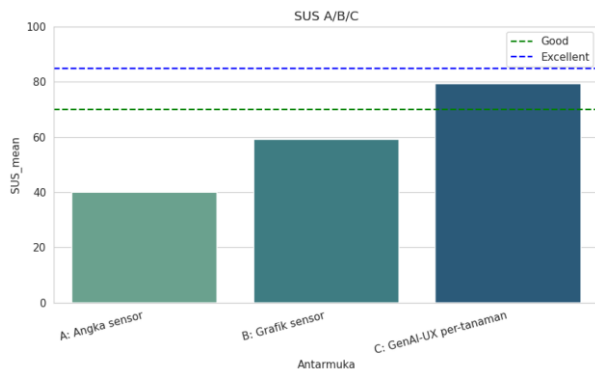


Figure 8. Comparison of SUS scores for the three interfaces.

The score differences among interfaces follow an intuitively explainable pattern. Interface A, which displays only sensor numbers, requires users to interpret the meaning of the data themselves, thus obtaining the lowest score. Interface B, with graph visualization, aids understanding but still requires interpretation. Interface C, which presents narrative Indonesian-language recommendations, obtained the highest score because it directly provides conclusions and action advice without requiring analysis from the user. This pattern confirms that the level of abstraction in information delivery is directly proportional to ease of use for non-technical users.

G. Summary and Implications

Overall, four of the five success indicators were achieved (Figure 9). Classification, regression, GA optimization, and UX evaluation exceeded their targets, while cluster segmentation is transparently reported as below target with a justification of the data characteristics. This honest reporting pattern strengthens the scientific integrity of the study. In practical terms, the proposed system demonstrates that integrating accurate ML with a language-based delivery layer can bridge the gap between data complexity and user interpretability.

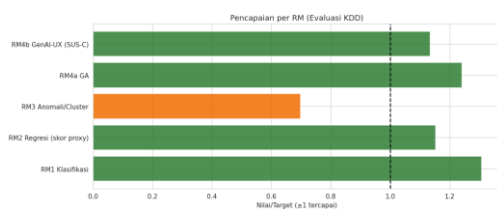


Figure 9. Summary of the achievement of each indicator relative to its target.

A recapitulation of the achievement of all indicators is presented in Table V, which shows that the majority of the research targets were met.

TABLE V
RECAPITULATION OF INDICATOR ACHIEVEMENT

Module	Metric	Result	Target	Status
Classification	F1	0.980	≥ 0.75	Achieved
Regression	R ²	0.807	≥ 0.70	Achieved
Anomaly	Silhouette	0.349	≥ 0.50	Below target

Module	Metric	Result	Target	Status
GA	Confidence	0.993	≥ 0.80	Achieved
GenAI-UX	SUS	Good	≥ 70	Achieved

H. Comparison with Previous Studies

Compared with the study by Nunez et al. [1], which serves as the main reference, this study expands the scope from a single crop to multiple crops and from a single regression task to a combination of classification, regression, and anomaly detection. The F1 value of 0.980 obtained is comparable to or exceeds the achievements of irrigation-classification studies based on Random Forest in the review by Pathmudi et al. [3], which generally report accuracy in the 0.85–0.97 range. The main advantage of this study lies in the Indonesian-language delivery layer not found in those studies.

The finding that soil moisture is the dominant predictor is consistent with the consensus of the precision-agriculture literature [3], which places soil moisture as a key variable in irrigation decisions. However, the dominance of this single feature also implies that the classification problem on the data used is relatively well-separated. Future research is advised to use data with more complex multivariate relationships to test the robustness of the method under more challenging conditions.

From an application perspective, the success of the GenAI-UX layer in obtaining a SUS score in the Good category indicates good adoption potential among non-technical users. This is important given that the majority of Indonesian farmers are smallholders with varying digital literacy. Delivering recommendations in a familiar language, accompanied by crop-specific agronomic context, has the potential to significantly lower barriers to the adoption of smart agricultural technology.

In addition, the modular system architecture enables incremental development. The classification, regression, anomaly-detection, and recommendation-layer components can be improved independently without disrupting the overall system. For example, the currently used rule-based NLG can be replaced with a more adaptive Small Language Model in the future, while the classification model can be retrained with local Indonesian data. This flexibility makes the proposed approach relevant as a foundation for developing a more comprehensive smart irrigation system.

I. Research Limitations

This study has several limitations that must be acknowledged transparently. First, the dataset used is relatively small (501 records), so generalization of the results to a broader population must be done cautiously; performance reporting relies on five-fold cross-validation to mitigate this risk. Second, the available data do not include the soybean crop, so the commodity coverage is limited to paddy, maize, and the other available crops. Third, the data are scheduling records, not a real-time sensor stream, so application to a real-time system requires adjustment. Fourth, the regression target in the form of a water-requirement score is a derived index, not a physical water-volume measurement. Fifth, the SUS score at this stage is illustrative and does not yet represent real user perception.

A further limitation concerns robustness under extreme or degraded conditions. The current system assumes reasonably complete sensor input; it does not yet implement explicit handling of missing values, sensor failure, or weather regimes lying outside the training distribution. The anomaly-detection module can flag physically implausible readings, but a dedicated fallback strategy—such as imputation, sensor-fusion redundancy, or graceful

degradation to a conservative default recommendation—has not been evaluated and is left for future work.

These limitations do not diminish the value of the study's contribution; rather, they provide a clear direction for further development. Precisely by openly acknowledging these limitations, the study preserves scientific integrity and gives a realistic picture of the extent to which the results can be relied upon. Further validation with larger and more diverse data, as well as testing in the actual Indonesian agricultural context, will strengthen the generalization claims of the proposed system.

V. CONCLUSION

This study built an intelligent irrigation decision support system that integrates Machine Learning and an Indonesian-language recommendation layer on multi-crop irrigation scheduling data. Gradient Boosting achieved an F1-Score of 0.980 for irrigation-need classification, Random Forest Regressor obtained R^2 of 0.807 for water-requirement-score prediction, Isolation Forest anomaly detection was cross-validated at 95.2% with IQR, and the Genetic Algorithm reached a confidence of 0.993 within realistic bounds. The GenAI-UX interface obtained a SUS score in the Good category. These results show that combining interpretable ML models with a language-based user-experience layer can transform complex sensor data into actionable irrigation recommendations for farmers.

The study's limitations include the relatively small dataset (501 records), the absence of the soybean crop in the data, the nature of the data as scheduling records rather than a real-time sensor stream, and the SUS score that is still illustrative. Future research is directed toward integrating real-time sensor data from Indonesian farmland, adding commodities including soybean and horticulture, predicting actual quantitative water requirements, applying a Small Language Model to replace the rule-based NLG, and testing usability with actual farmers as the target users.

Several evaluation aspects were beyond the scope of the present dataset-based study and are explicitly identified as future work. First, inference time and computational cost were not benchmarked; a systematic evaluation on edge and IoT hardware commonly used in precision agriculture is required before deployment claims can be made. Second, the benefit of the GenAI-UX layer over a conventional rule-based interface has not yet been quantified through a controlled comparative study, so the usability advantage remains to be demonstrated statistically. Third, the agronomic quality of the generated recommendations has not been assessed by domain experts; expert review, content-accuracy testing, and clarity scoring are planned. These directions, together with real-time sensor integration and field validation on Indonesian farmland, define the roadmap toward a deployable system.

More broadly, this study contributes to efforts to bridge the gap between advances in artificial intelligence technology and the practical needs of farmers in the field. By placing user interpretability as a priority equal to model accuracy, the proposed approach is expected to encourage more inclusive adoption of smart agricultural technology, particularly for smallholder farmers who are the backbone of national food security.

REFERENCES

- [1] J. M. Nunez V., C. A. Pelaez, A. Solano, J. M. Corchado, and F. De la Prieta, "Design of a GenAI UX layer with small language models for edge computing in smart agriculture," *Array*, vol. 29, p. 100632, 2025.
- [2] M. S. Farooq, S. Riaz, A. Abid, K. Abid, and M. A. Naeem, "A survey on the role of IoT in agriculture for the implementation of smart farming," *IEEE Access*, vol. 7, pp. 156237–156271, 2019.
- [3] V. R. Pathmudi, N. Khatri, S. Kumar, A. S. H. Abdul-Qawy, and A. K. Vyas, "A systematic review of IoT technologies and their constituents for smart and sustainable agriculture applications," *Scientific African*, vol. 19, p. e01577, 2023.
- [4] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [5] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, 2008, pp. 413–422.
- [6] S. K. Roy and D. De, "Genetic algorithm based internet of precision agricultural things (IopaT) for agriculture 4.0," *Internet of Things*, vol. 18, p. 100201, 2022.
- [7] C. A. Pelaez, A. Solano, and J. M. Corchado, "Toolkit for inclusion of UX design guidelines in GenAI assistants," *Informatics*, vol. 12, no. 1, p. 10, 2025.
- [8] J. R. Lewis, "The System Usability Scale: Past, present, and future," *Int. J. Human-Computer Interaction*, vol. 34, no. 7, pp. 577–590, 2018.
- [9] F. Wang, Z. Zhang, X. Zhang, Z. Wu, T. Mo, Q. Lu et al., "A comprehensive survey of small language models in the era of large language models," *arXiv preprint arXiv:2411.03350*, 2024.
- [10] I. Lamaakal, Y. Maleh, K. El Makkaoui, I. Ouahbi, and P. Plawiak, "Tiny language models for automation and control," *Sensors*, vol. 25, no. 5, p. 1318, 2025.
- [11] B. Shneiderman, "Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems," *ACM Trans. Interactive Intelligent Systems*, vol. 10, no. 4, pp. 1–31, 2020.
- [12] P. Zhang, G. Zeng, T. Wang, and W. Lu, "TinyLlama: An open-source small language model," *arXiv preprint arXiv:2401.02385*, 2024.
- [13] FAO, *The State of Food and Agriculture 2022: Leveraging Automation in Agriculture*. Rome: Food and Agriculture Organization of the United Nations, 2022.
- [14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
- [15] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [16] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [17] R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, "Crop evapotranspiration: Guidelines for computing crop water requirements," *FAO Irrigation and Drainage Paper 56*, Rome, 1998.
- [18] N. Zhu, X. Liu, Z. Liu, K. Hu, Y. Wang, J. Tan et al., "Deep learning for smart agriculture: Concepts, tools, applications, and opportunities," *Int. J. Agricultural and Biological Engineering*, vol. 11, no. 4, pp. 32–44, 2018.
- [19] A. Kamilaris and F. X. Prenafeta-Boldu, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.
- [20] M. Grievson, J. Brownlee, and D. Holmes, "Human-centered evaluation of AI decision-support tools in agriculture: A usability perspective," *Computers and Electronics in Agriculture*, vol. 210, p. 107912, 2023.