Pertanika J. Sci. & Technol. 33 (6): 2833 - 2854 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Machine Learning Enabled IoT System for Agricultural Land Recommendation

Bayu Rima Aditya[1]*, Anindia Agusta Ken Nadila[1], Muhammad Qanit Al-Hijran[1], Muhammad Bintang Ramadhan[1] and Yudha Ginanjar[2]

[1]*School of Applied Science, Telkom University, Bandung 40257, Indonesia*
[2]*Rastek Inovasi Digital, Bandung 40257, Indonesia*

ABSTRACT

This study aims to develop an agricultural land recommendation system by integrating the Internet of Things (IoT) and machine learning (ML). IoT devices, including the JXBS-3001 soil sensor and Raspberry Pi Pico RP2040, collect real-time soil data, which is analyzed using the decision tree (DT) algorithm. The DT algorithm is chosen for its simplicity, efficiency, and interpretability over random forest (RF) and *k*-nearest neighbors (*k*-NN). It provides structured decision-making, faster training, and better handling of numerical data for parameters such as soil pH, nutrient content (NPK), moisture levels, and temperature. The findings show that the system provides accurate crop recommendations, helping farmers make informed decisions. The integration of IoT and ML enhances land assessment and optimizes agricultural productivity. Future improvements could include weather analysis and plant disease detection to further support decision-making.

Keywords: Agricultural, decision tree, Internet of Things, machine learning, recommendation system

ARTICLE INFO

Article history:
Received: 26 December 2024
Accepted: 06 June 2025
Published: 29 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.18

E-mail addresses:
bayu@tass.telkomuniversity.ac.id (Bayu Rima Aditya)
anindiaagustaken@student.telkomuniversity.ac.id (Anindia Agusta Ken Nadila)
muhammadqanital@student.telkomuniversity.ac.id (Muhammad Qanit Al-Hijran)
muhammadbr@student.telkomuniversity.ac.id (Muhammad Bintang Ramadhan)
yginanjar@gmail.com (Yudha Ginanjar)
* Corresponding author

INTRODUCTION

Agricultural management refers to the oversight of farming practices, which play a vital role in the agricultural industry by directly influencing the primary goal of agriculture: ensuring a sustainable food supply. According to the Food and Agriculture Organization of the United Nations (FAO) (2021), recent years have seen increasing demands for agricultural practices to deliver at least 70% of global food production annually. This requirement

e-ISSN: 2231-8526

is driven by the rapid growth of the global population, which is projected to reach approximately 9.6 billion by 2050 (FAO, 2021). Adding to the challenge, the availability of arable land continues to decline each year, further intensifying the need for effective agricultural management (Wang, 2022).

Indonesia is one of the countries with immense agricultural potential, supported by abundant natural resources and a favorable climate. However, the implementation of agricultural recommendation systems in farming activities remains underdeveloped (Iskandar, 2021). This is concerning, as limited knowledge among farmers about critical factors in agricultural decision-making, such as identifying the optimal planting season, evaluating land performance, or determining the most suitable crops for specific periods, continues to hinder productivity (Iskandar, 2021; Yundari, 2017). If left unaddressed, these challenges could lead to suboptimal agricultural practices, ultimately failing to meet Indonesia's growing demand for agricultural output. This condition indicates the importance of data-based decision-making to support farmers in choosing the most suitable crops (Priyadharshini et al., 2021; Ridoy et al., 2024).

To promote sustainability in Indonesia's agricultural industry, innovation is essential across various aspects of farming practices (Rajeswari et al., 2018). One key area for innovation is decision-making, which can be significantly enhanced by developing agricultural recommendation systems powered by ML (Singh et al., 2024). ML, a branch of artificial intelligence, focuses on developing methods that can learn from past data (Lestari et al., 2023). ML has emerged as a transformative tool across many industries, including agriculture, by providing advanced, data-driven solutions to boost efficiency and productivity (Musanase et al., 2023; Ridoy et al., 2024). In agriculture, ML leverages data from sensors and other sources to analyze land conditions, predict yields, and optimize resource use (Akintuyi, 2024). This technology enables farmers to make more accurate, data-informed decisions while mitigating risks associated with weather fluctuations and soil variability. By integrating ML, agriculture can evolve toward smarter, more sustainable methods, paving the way for a revolutionary shift in farming practices.

The main objective of this study is to develop a technology-driven agricultural land recommendation system. This system integrates IoT and ML to provide farmers with precise recommendations for selecting agricultural land suitable for their desired crops. IoT is a concept of integration that enables any object to send data through network modernization (Kiruthika & Karthika, 2023). In this study, the system acts as an intermediary between farmers and the system, enabling farmers to input desired parameters directly through the interface. These parameters are then transmitted to the ML system for further processing. Various ML techniques can be applied for recommendations, such as DT, Naive Bayes (NB), RF, rotation forest, *k*-NN, artificial neural networks (ANN), and support vector machine (SVM), among others. Each method has its unique strengths and limitations. DT, for instance, is chosen for its high accuracy and ability to trace each decision step taken by the algorithm. Additionally, DT can seamlessly integrate with database systems, offer strong accuracy, and identify unexpected data combinations (Sihombing & Arsani, 2021). It also effectively represents patterns, knowledge, or insights in a decision-tree format (Khoeri & Mulyana, 2021).

By integrating IoT and ML to facilitate farmers with the system and employing ML algorithms, specifically the DT method, the proposed system aims to conduct in-depth analyses of collected data. This process ultimately delivers tailored recommendations for the most suitable crop commodities for specific agricultural lands. Therefore, this study aims to design and develop an agricultural land recommendation system based on IoT and ML technologies.

RELATED WORKS

IoT-based land-monitoring systems have attracted significant attention in recent years, enabling data-driven decision-making in agriculture. IoT enables real-time data acquisition from diverse sensors, including moisture, pH, and nutrient sensors, to assess soil conditions (Aditya et al., 2024; Gaikwad et al., 2021; García et al., 2020; Ramson et al., 2021). However, existing IoT-based systems often lack intelligent decision-making capabilities, necessitating the integration of ML models to enhance agricultural recommendations.

Various ML techniques have been explored in crop recommendation, with studies utilizing RF, SVM, and ANN to improve predictive accuracy (Lokhande et al., 2022; Qiu et al., 2021; Vemulapalli et al., 2024). The study by Modi et al. (2021) highlighted the risks associated with improper crop selection, including reduced yields and severe issues, such as rising suicide rates among farmers. To mitigate these challenges, they proposed a crop recommendation system based on the SVM algorithm. This system analyzed soil parameters, including nitrogen (N), phosphorus (P), potassium (K), pH, moisture levels, rainfall, and temperature. Achieving an accuracy rate of 97%, the system provided farmers with precise recommendations for selecting the most suitable and productive crops, thereby reducing errors in crop selection and increasing agricultural output. The suggested approach does not focus on real-time processing.

Several studies have applied DT algorithms for agricultural recommendations due to their efficiency and interpretability (Khoeri & Mulyana, 2021; Sihombing & Arsani, 2021). While DT offers advantages in terms of fast training and explainability, its performance in handling complex feature interactions remains a challenge.

In a separate study by Islam et al. (2023), the authors present an IoT-based solution that integrates ML techniques to optimize crop production through real-time soil monitoring. The system uses various sensors to measure soil nutrients, moisture, temperature, and humidity, and sends the data to a server for analysis. Using ML algorithms, the system provides customized crop recommendations and fertilizer-use guidelines. However, this model cannot assess the percentage accuracy of its crop recommendations.

Another notable study by Rao et al. (2022) explored the impact of nutrient deficiencies and incorrect crop selection on agricultural outcomes. The research compared three ML algorithms — *k*-NN, DT, and RF classifier — using Gini and entropy criteria. The findings revealed that the RF classifier achieved the highest accuracy of 99.32%, while *k*-NN had the lowest accuracy of 97.04%. The DT algorithm performed moderately, with the Gini criterion yielding better results than entropy. However, this study still has limitations, particularly the lack of integration between crop prediction and IoT technology, which could enable real-time data collection to improve the accuracy and relevance of recommendations.

Recent advancements in agricultural technology, particularly the integration of IoT and ML, enable farmers to collect and analyze data in real time (Senapaty et al., 2023). In a separate study by Islam et al. (2023), the authors present an IoT-based solution that integrates ML techniques to optimize crop production through real-time soil monitoring. The system uses various sensors to measure soil nutrients, moisture, temperature, and humidity, and sends the data to a server for analysis. Using ML algorithms, the system provides customized crop recommendations and fertilizer-use guidelines. The suggested approach does not consider portability for farmers, as it is not designed in an integrated device that can be easily transported to various locations.

This study addresses the gap in the literature by integrating IoT and DT-based ML models to provide real-time, resource-efficient, and interpretable crop recommendations. Unlike previous works that rely on static datasets, this research incorporates real-time sensor data via IoT devices, enabling dynamic adjustments based on environmental conditions. Additionally, the proposed system optimizes DT hyperparameters to balance accuracy and computational efficiency, making it more suitable for deployment in low-power agricultural devices.

MATERIALS AND METHODS

IoT Devices and Sensors

This study uses three sensor types: a 3.5-inch thin-film transistor liquid-crystal display (TFT LCD) module (China), a Raspberry Pi Pico RP2040 (Raspberry Pi Foundation, United Kingdom), and a JXBS-3001 soil sensor (Weihai Jingxun Changtong Electronic Technology Co., Ltd., China). The circuit diagrams in Figures 1 and 2 present the configuration and specifications of the proposed IoT system components.

a. 3.5-inch TFT LCD

   TFT LCD is a type of display screen that uses thin-film transistors to control each pixel individually, resulting in sharper and more responsive images. In this system, a TFT

LCD displays the results of soil sensor data analysis. This component is a 3.5-inch (77.98 mm × 43.94 mm) touch display with a resolution of 320 × 240 pixels, offering sharp, detailed graphics for visual interfaces. It is equipped with a resistive touchscreen, allowing users to interact directly with the device. The display uses an ILI9341-16 controller that supports a 16-bit data interface. It is powered by the Universal TFT Library from Rinky-Dink Electronics, which simplifies programming across a variety of Arduino platforms. The display is compatible with the Arduino Mega and Arduino Due microcontrollers, both of which offer sufficient memory and I/O pins to handle the high-resolution data from the display. However, it does not support Arduino Uno with an 8-bit serial interface due to data transmission limitations.

b. Raspberry Pi Pico RP2040

The Raspberry Pi Pico RP2040 module serves as the primary microcontroller, facilitating communication between sensors and other system components. This printed circuit board (PCB) supports the SmartMatrix Arduino Library, ensuring optimal compatibility for a variety of visual applications. The Raspberry Pi Pico RP2040 delivers strong data processing capabilities, bolstered by Wireless Fidelity (Wi-Fi) connectivity. This enables the system not only to receive input from sensors but also to process the data using ML algorithms and transmit or receive information wirelessly. With good energy efficiency, this module also supports SmartMatrix HUB75 connectivity, specifically designed for red, green, and blue (RGB) matrix light-emitting diodes (LEDs). The pinouts available on this module enable high-precision, high-speed LED control, ensuring effective synchronization of visual elements on LED displays.

c. JXBS-3001 Soil Sensor

The JXBS-3001 soil pH sensor is an important component in the soil monitoring system, designed to integrate seamlessly with the Raspberry Pi Pico RP2040 via RS-485 communication. Powered by a 12-24V DC supply and utilizing the Modbus protocol, this sensor is capable of measuring soil parameters, including NPK, total dissolved solids (TDS), moisture, temperature, and pH, with a precision of ±0.3. The sensor's real-time data is transmitted to the Raspberry Pi Pico RP2040 for processing and displayed via an LED interface or a TFT screen. The integration of the JXBS-3001 soil sensor with the Raspberry Pi Pico RP2040 enables the system to provide optimal recommendations on crop land suitability. Supported by the microcontroller's rapid processing capabilities and wireless connectivity, the system ensures efficiency, stability, and reliability in field applications.

d. Direct current to direct current (DC-to-DC) step-up converter

The DC-to-DC step-up converter, also known as a boost converter, is a device that increases the input voltage to a higher output level. It is particularly useful when a device requires a higher voltage than the available power source, such as boosting low voltage from a battery to operate specific electronic components. The boost converter uses inductive charging to efficiently raise the voltage, making it an ideal choice for portable or renewable energy systems. In microcontroller-based sensor and control applications, the step-up converter plays a critical role in maintaining optimal performance by ensuring adequate voltage levels for various system components, even when operating on limited power sources.

e. 3.7 V battery

The battery serves as the primary or backup power source in electronic systems, providing stable and continuous energy to ensure smooth operations. In applications involving the Raspberry Pi Pico RP2040 or similar microcontrollers, the battery enables portable, independent operation, eliminating the need for a constant connection to an external power source. The battery type and capacity are tailored to meet the device's power requirements and the desired operational duration.

f. RS-485

RS-485 is a serial communication protocol widely used to connect devices that require long-distance communication in industrial networks or control systems. Its advantages include its ability to transmit data over long cable lengths with minimal interference and its support for multipoint communication. In systems with sensors and microcontrollers, RS-485 is often used to ensure stable, efficient device-to-device communication, even in challenging environments.

The architectural diagram illustrates a structured system integrating the JXBS-3001 soil sensor with the Raspberry Pi Pico RP2040, employing Wi-Fi communication to send data to an external application programming interface (API). The JXBS-3001 soil sensor measures soil parameters such as moisture, pH, and nutrients, transmitting data through the RS-485 protocol to the Raspberry Pi Pico RP2040, which serves as the system's data processing hub. RS-485 ensures reliable communication over long distances with minimal interference. The Raspberry Pi Pico RP2040 processes sensor data and directs it to two main outputs: first is real-time visualization on a 3.5-inch TFT LCD for immediate user feedback, and second is transmission via Wi-Fi using the Transmission Control Protocol/Internet Protocol (TCP/IP) to an external API for remote monitoring and integration with other applications. TCP/IP is a network protocol that standardizes communication on the internet and local networks. The system is powered by a 3.7 V battery connected to a DC-to-DC step-up converter to ensure a stable and sufficient voltage supply. For charging, there is a charger controller module that is connected to the recreational vehicle (RV) DC charger, managing the battery recharge to keep it charged and allowing the system to operate for extended periods without interruption. Overall, this integrated soil monitoring system provides real-time data visualization, remote data transmission, and optimized power efficiency with a rechargeable battery, making it ideal for field applications independent of constant external power sources.

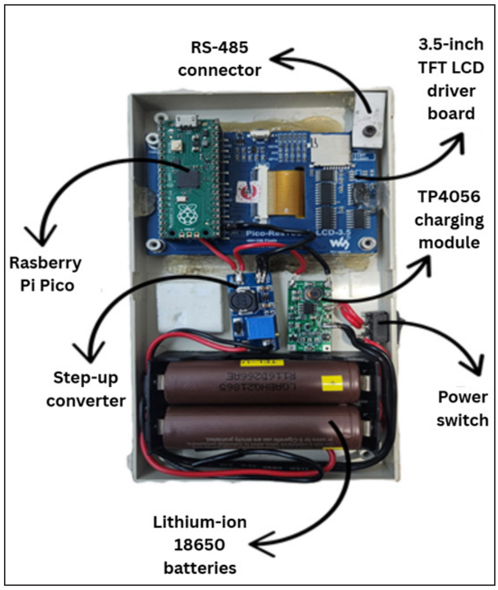

Figure 1. Block diagram of a portable device
Note. TFT LCD = Thin-film-transistor liquid-crystal display; RS-485 = Recommended standard 485

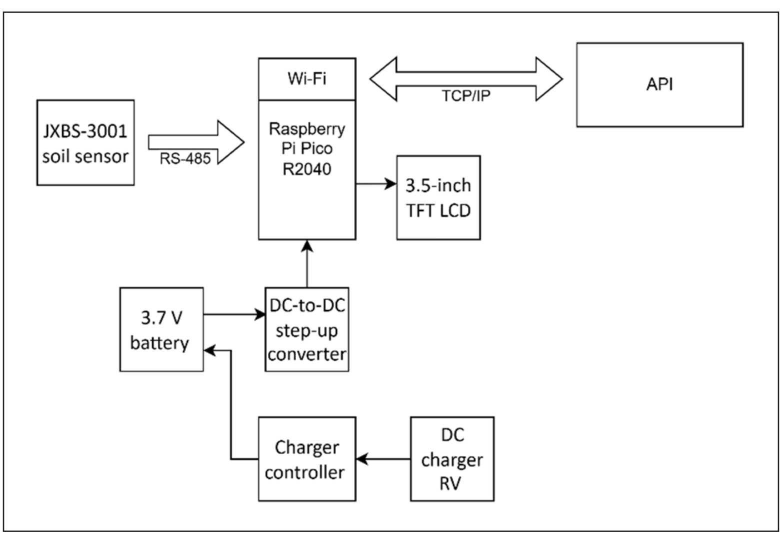

Figure 2. Schematic of an embedded system
Note. TFT LCD = Thin-film-transistor liquid-crystal display; TCP = Transmission Control Protocol; IP = Internet Protocol; API = Application programming interface; DC = Direct current

Data Transmission Process

In the described communication system, TCP/IP serves as the foundational network protocol, enabling data exchange between devices and servers via Hypertext Transfer Protocol (HTTP), a protocol used for data exchange over the internet. The process begins with the device sending a request to the server using the POST method. Data is structured in JavaScript Object Notation (JSON) format and included in the HTTP body. TCP ensures reliable delivery and correct sequencing of data, while IP handles addressing and routing across the network.

Once the server receives the data, it processes it based on the predefined application logic. The server then responds to the device in JSON format via HTTP. TCP connection ensures that the response data maintains integrity and proper sequencing, while IP manages addressing and routing to the correct destination on the network. The HTTP protocol governs the formatting and structuring of data communication, ensuring it can be understood and correctly processed.

The device then receives the server's response and displays the information to the user. This entire communication process, from sending a request to receiving a response, is supported by TCP/IP, which provides the basic mechanism for data transmission. Meanwhile, HTTP offers an application-level protocol to structure and format the transmitted data efficiently. This communication cycle highlights the integral implementation of TCP/IP in supporting efficient and structured HTTP-based communications.

ML Methods

a. DT algorithm classification report

The *croprekomendasi.csv* dataset has been analyzed and cleaned to improve data quality for crop recommendations based on environmental factors. The initial analysis showed no missing values, but several features, such as illumination (illum), rainfall (rain), and wind speed (wind), contained extreme values or uneven distributions. For instance, the illumination value reached 216, while 75% of the data had values below 9.86. Similarly, rainfall and wind speed had many zero values, as well as some extremely high values. To address this issue, the interquartile range (IQR) method was applied to detect and remove outliers. As a result, 998 data points (32.6%) were removed from the total 3,063 records, leaving 2,065 cleaned data points.

The DT is chosen over RF and *k*-NN due to its simplicity, efficiency, and ease of interpretation. Its clear tree structure allows for more transparent analysis than RF, which consists of multiple trees and is difficult to trace, and *k*-NN, which does not build an explicit model. In terms of efficiency, the DT trains are faster than RF, which requires multiple iterations to construct various DTs, and it also predicts faster than *k*-NN, which needs to compare each new data point with the entire training dataset. Additionally, the DT does not rely on distance metrics like *k*-NN, making it more suitable for data with numerical and categorical attributes that have different scales. This algorithm is also easier to optimize as it does not require extensive parameter tuning, such as determining the number of trees in RF or selecting the optimal *k* value in *k*-NN. Another advantage is its ability to handle hierarchical data, such as classifying crops based on environmental factors, which is difficult for k-NN, which relies solely on data proximity without understanding rule-based hierarchies. With these advantages, the DT becomes a more suitable choice for classifying sensor data and processed data.

The backend service will store the accumulated data in the database. The stored data includes both raw and processed sensor data. A popular machine-learning technique for classification tasks is the DT algorithm, which is based on predictive modeling. In this context, the algorithm partitions a dataset into groups based on specific characteristics relevant to the target categories. The DT starts with a root node and branches into smaller nodes, where each branch represents a decision or classification based on attributes. To produce leaf nodes that reflect the most appropriate classification choices, the primary objective is to optimize homogeneity within each data subset at each branch.

Choosing the most informative attribute to divide the data into smaller subsets is the first step in the classification process. This attribute is selected based on criteria such as information gain (IG) or the Gini index, which measure how well it can classify the data. For instance, soil pH or moisture might be critical attributes for predicting suitable crops, as they have significant discriminative strength for separating plant types based on their needs.

The DT divides the data iteratively during the learning process, with each new node representing a classification decision based on the tested attribute's value. Once the tree is fully developed, new data can be classified by following a series of binary decisions that guide the data from the root node to the leaf nodes, where a final classification or recommendation is provided.

b. DT ML algorithm for recommendations

To make crop recommendations, the DT algorithm divides the dataset into smaller subsets based on relevant characteristics, such as soil pH, nutrient content (NPK), moisture levels, and temperature. The algorithm maps the relationships between these parameters and the crops best suited to the conditions, using a DT structure derived from a series of if-then-else conditions. Each node in the tree represents a test on a specific feature (e.g., soil pH), the branches show the test results, and the final leaves offer the crop recommendation.

The tree-building process involves techniques that either maximize IG or minimize the Gini index (GI). At each step, the algorithm selects the most significant attribute to split the data and create a new node. The IG is calculated using an entropy formula:

$$Entrophy\ (S) = -\sum_{i=1}^{n} P_i\ P_i$$

In this context, P_i represents the proportion of the class i within subset S. The IG is then calculated as follows:

$$IG(T, A) = Entropy\ (T) - \sum_{v \in Values\ \ (A)} \frac{|T_v|}{|T|} Entrophy\ (T_v)$$

where T is the entire dataset, A is the attribute being tested, and T_v is the subset of data where attribute A has value v. A higher IG indicates that the attribute is more effective in dividing the data, making it a higher-priority candidate for a DT node.

Once the DT is constructed, new data (e.g., soil sensor data) is analyzed by starting at the root node and traversing branches based on the conditions the data meet until a leaf node is reached that provides a crop recommendation. This algorithm effectively recommends the best crop varieties based on agricultural conditions, including soil pH, moisture levels, and temperature. Its strength lies in its ability to handle multidimensional data and produce decisions that farmers can understand and use with ease.

Proposed Framework

The proposed framework integrates IoT sensors, data transmission, and ML to enhance decision-making processes in agriculture. The system employs various soil sensors (pH, NPK, TDS, moisture, and temperature) installed in agricultural fields to measure soil conditions in real-time. The proposed framework is illustrated in Figure 3. These sensors are connected to portable devices used by farmers, which serve as data-collection hubs. The portable devices are equipped with Raspberry Pi Pico RP2040, responsible for gathering data from sensors and transmitting it to remote servers via the Internet using TCP/IP and HTTP protocols.

ML algorithms are used to process the data when it arrives at the server. These algorithms analyze soil parameters and generate personalized crop recommendations tailored to the current soil conditions. Appropriate crop varieties, ideal planting times, and accurate fertilizing techniques are among the suggestions. After that, the analysis results are then sent back to the farmer's portable device, where they may see the data and suggestions right on the screen.

Additionally, the system provides continuous soil condition monitoring, enabling farmers to make data-driven decisions that enhance productivity and efficiency. To accommodate an increasing number of users and agricultural lands, the framework is designed with efficient server management, including optimized database handling, caching mechanisms, and load balancing strategies to maintain performance as demand grows. By combining IoT and ML technologies with a well-structured server-based infrastructure, the proposed method aims to optimize land use and maximize agricultural yields with precision and reliability.

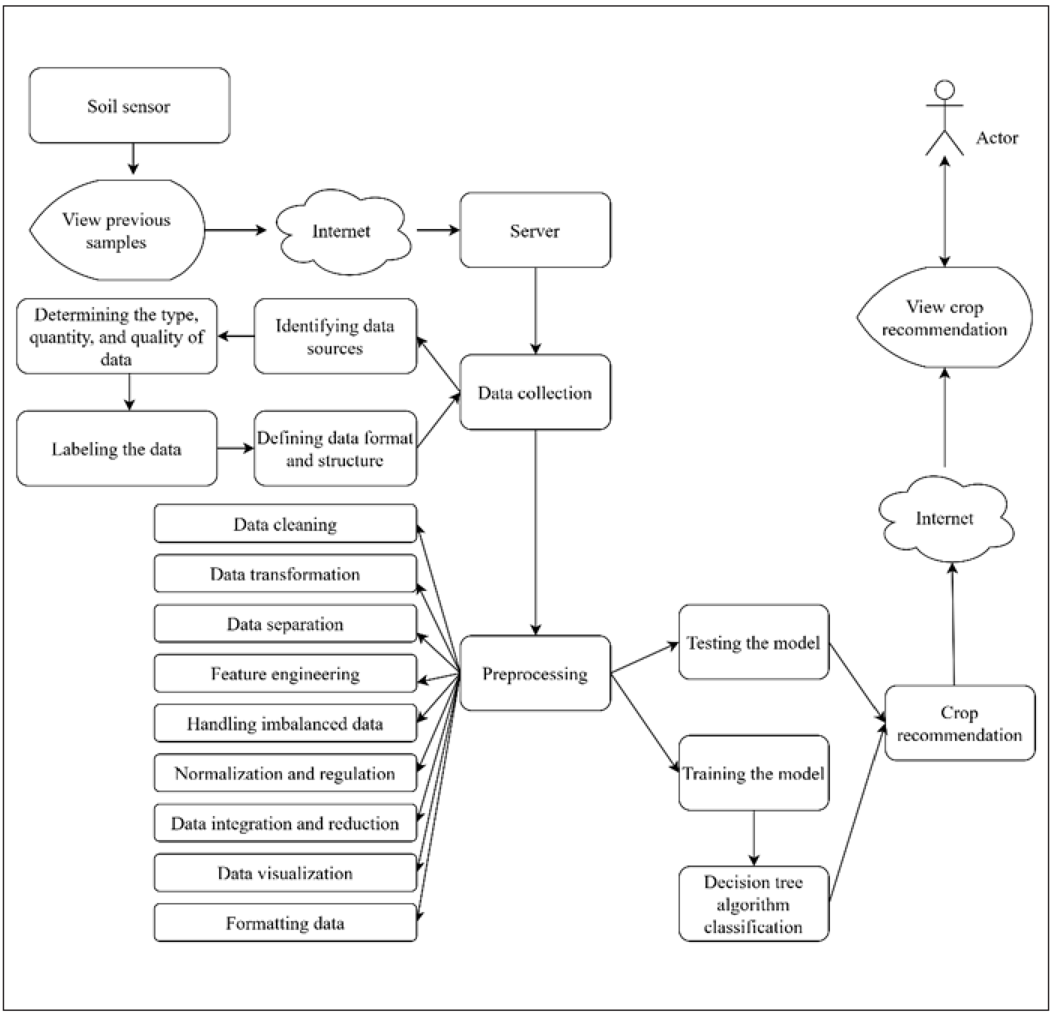

Figure 3. Framework for intelligent crop recommendation

RESULTS

Experimental Setup

The system comprises a single sensor that measures critical soil parameters, such as pH, moisture, and nutrient levels. This sensor connects to a microcontroller unit (MCU) via serial input/output (I/O), which manages data acquisition from the sensor. The MCU serves as the primary processing hub, handling the collection, processing, and transmission of sensor data to other system components. Figure 4 illustrates the system block diagram of the proposed system.

A display module is incorporated to provide real-time visualization of the sensor data, enabling users to monitor soil conditions instantly. Additionally, a Wi-Fi module facilitates wireless data transmission, linking the device to a server or cloud system via TCP/IP and HTTP protocols. The server can further analyze the transmitted data using ML algorithms to generate recommendations for crops best suited to the current soil conditions.

The system also includes a storage component to store data locally for historical records or in-depth analysis. A power supply ensures all components remain operational, supporting uninterrupted functionality. This experimental setup is designed to continuously collect soil data, enabling ML-driven analyses that deliver actionable crop recommendations that can be directly accessed via the display.

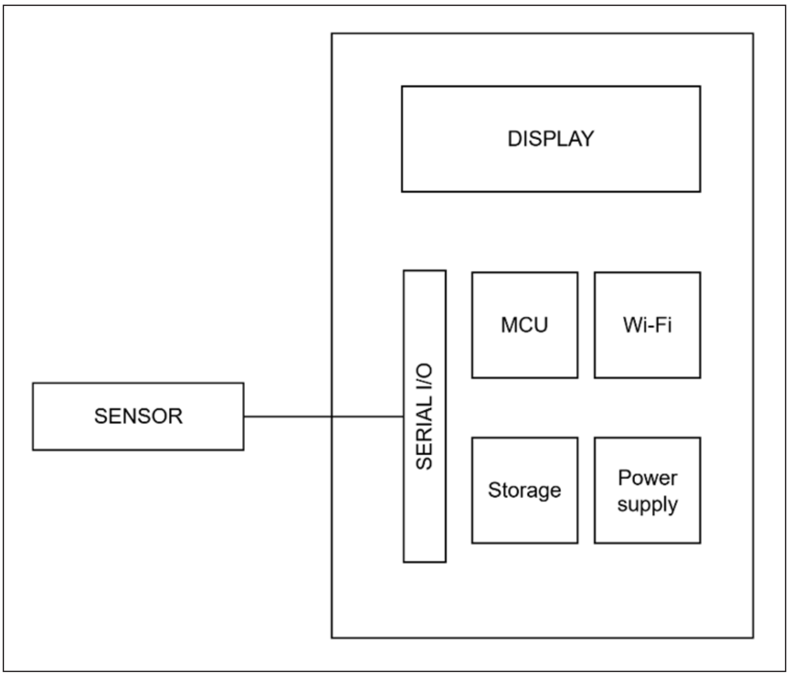

Figure 4. System block diagram
Note. SERIAL I/O = Serial input/output; MCU = Microcontroller unit; Wi-Fi = Wireless Fidelity

Data Collection and Preprocessing

The collected data originates from real-time environmental and soil monitoring, serving as input for training, validating, and testing the ML model. High-quality, relevant data is crucial to developing an accurate and effective model. The data collection process follows these steps:

1. Identifying data sources
2. Determining the type, quantity, and quality of data
3. Labeling the data
4. Defining data format and structure

Data preprocessing is an essential step in evaluating the model's performance and ensuring its reliability. Therefore, data cleaning, data transformation, data separation, feature engineering, handling imbalanced data, normalization, and regularization, data integration and reduction, data visualization, and organizing data into a format suitable for training and testing the model are carried out.

Summary Statistics of a Feature in ML

Tables 1 and 2 present a statistical summary of a specific feature within the ML dataset. They include the minimum, maximum, and average values of all sensor parameters observed for the crops under study, specifically corn and shallots.

Table 1
Statistical summary of corn

Parameter	Min	Mean	Max
N (mg/kg)	0.00	6.32	17.00
P (mg/kg)	29.40	59.96	85.64
K (mg/kg)	23.00	52.38	78.18
Soil temperature (°C)	9.84	26.11	38.63
Humidity (%)	46.62	86.21	99.90
pH	22.80	55.13	69.50
Conductivity (μS/cm)	97.20	213.9	267.0
Soil humidity (%)	12.88	25.44	46.17
Temperature (°C)	18.30	26.09	35.60
TDS (ppm)	48.40	106.70	133.18
Salinity (ppt)	53.20	117.20	146.30

Note. Min = Minimum; Max = Maximum; N = Nitrogen; P = Phosphorus; K = Potassium; pH = Potential of hydrogen; TDS = Total dissolved solids

Table 2
Statistical summary of shallots

Parameter	Min	Mean	Max
N (mg/kg)	0.00	1.53	69.20
P (mg/kg)	5.03	32.73	206.70
K (mg/kg)	2.70	24.90	200.10
Soil temperature (°C)	16.81	24.84	35.70
Humidity (%)	42.51	80.01	99.90
pH	29.83	48.65	90.00
Conductivity (µS/cm)	38.06	157.53	518.11
Soil humidity (%)	0.58	9.68	33.43
Temperature (°C)	18.30	26.09	35.60
TDS (ppm)	19.22	78.53	258.80
Salinity (ppt)	21.27	86.17	284.50

Note. Min = Minimum; Max = Maximum; N = Nitrogen; P = Phosphorus; K = Potassium; pH = Potential of hydrogen; TDS = Total dissolved solids

DT Algorithm Classification Report

Figure 5 illustrates the classification report generated by the DT algorithm. The table evaluates each crop, providing metrics such as precision, recall, F1 score, and support. Precision represents the proportion of correctly identified positive cases, while recall measures the proportion of actual positive instances that were accurately classified. The F1 score is a harmonic mean of precision and recall, offering a balanced metric for performance evaluation. Support indicates the number of samples used to calculate each metric. Additionally, the report includes macro averages and weighted averages to reflect performance across multiple classes.

```
DecisionTrees's Accuracy is:  98.36779107725789
              precision    recall  f1-score   support

           0       0.99      0.98      0.99       540
           1       0.97      0.99      0.98       379

    accuracy                           0.98       919
   macro avg       0.98      0.98      0.98       919
weighted avg       0.98      0.98      0.98       919
```

Figure 5. Classification report for a decision tree model
Note. avg = Average

Accuracy of Crop Recommendation Models

Table 3
Accuracy of crop recommendation

No.	Model	Accuracy
1	Random forest	0.99
2	SVM	0.90
3	Naïve Bayes	0.91
4	Decision tree	0.98
5	XGBoost	0.99

Note. SVM = Support vector machine

The XGBoost, known for its ensemble approach to combining DT models, and the RF, which similarly aggregates multiple DTs, generally outperform standalone DT algorithms. These ensemble methods leverage the collective insights of multiple trees, enhancing their overall predictive accuracy. In contrast, individual DTs may underutilize some features during training. The significance of each attribute in these models accounts for all features except those excluded during training, which can limit the performance of a single DT.

The high accuracy achieved by the crop recommendation model reflects the rigorous optimization techniques applied throughout its development. Careful feature selection ensured that only the most relevant soil and environmental parameters were included, with redundant or less impactful features removed based on correlation analysis. This step allowed the model to focus on the most meaningful data, improving its predictive performance. Furthermore, extensive hyperparameter tuning using techniques such as grid search and Bayesian optimization was conducted to improve the model's learning.

In addition, the ensemble nature of XGBoost and RF enabled the model to capture complex relationships in the data, reducing bias and variance and improving generalization. The dataset itself was carefully curated, cleaned, and preprocessed to ensure high data quality, minimizing the risk of noise and inconsistencies. To further validate the model's robustness, cross-validation techniques were employed, ensuring that the accuracy remained consistent across different subsets of data.

By integrating these advanced methodologies, the crop recommendation model effectively leverages the strengths of ensemble learning and rigorous data processing, leading to its impressive accuracy. Table 3 presents the detailed accuracy results of the model.

The DT algorithm was chosen as the main model in this system not only because of its high accuracy, but also because it has a low absolute error value. This low error is significant in crop recommendations, where a small error can impact the decision taken by the farmer, such as the selection of the wrong crop. In addition, DT offers a good level of interpretability, which allows the system to provide clear and easy-to-understand recommendations. The DT structure used in this model makes the results explainable in a more transparent way. Therefore, although ensemble methods may be considered in the future to improve the accuracy of the system, DT remains the top choice due to its balance between good accuracy and easier interpretability.

Dimensionality Reduction with Variable Feature Issue

In this context, XGBoost and DT models demonstrate minimal errors, largely due to their ability to handle non-linear relationships and effectively manage data complexity. XGBoost employs iterative boosting techniques to refine model performance, while DTs adaptively learn the structural intricacies of data. DTs are particularly effective in capturing non-linear relationships between features and target variables. By adaptively partitioning the feature space, the DT can modify the model for training data, resulting in accurate predictions. Table 4 presents the absolute error of the crop recommendation model.

Table 4
Absolute error of crop recommendation

No.	Model	Absolute error
1	Random forest	0.009793
2	SVM	0.097933
3	Naïve Bayes	0.091404
4	Decision tree	0.013058
5	XGBoost	0.008705

Note. SVM = Support vector machine

Crop Recommendation Results

Figure 6 illustrates the crop recommendation results obtained after applying the DT ML algorithm. In the results, crops are represented with codes: code 0 indicates shallots, while code 1 represents corn. This classification identifies the most suitable crop for a given soil condition.

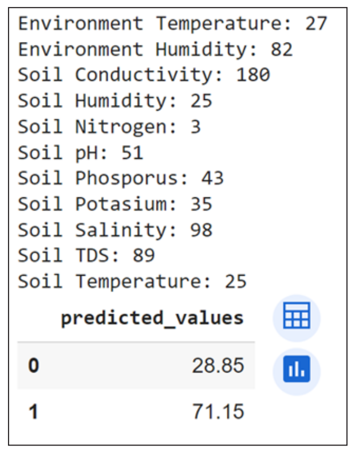

Figure 6. Soil sensor data and predicted values

IoT-ML Enabled Agriculture Platform

This study introduces a crop recommendation model for corn and shallots, based on an ML architecture with various parameter configurations. The DT algorithm was employed to process and manage soil parameter data. The model's performance was evaluated using multiple metrics, including accuracy, loss, precision, recall, F1 score, and confusion matrix. Experimental results demonstrated that the DT algorithm provides highly accurate predictions for crop-to-land suitability. This is evidenced by the high accuracy achieved on test data (98%), supporting its potential for future applications. In summary, compared to conventional techniques, the ML-based crop and soil suitability recommendation model with the DT model produces better crop suitability prediction outcomes. A preview of the suggestion tool's user interface (UI) is displayed in Figures 7-10.

Figure 7. Set sampling parameter interface

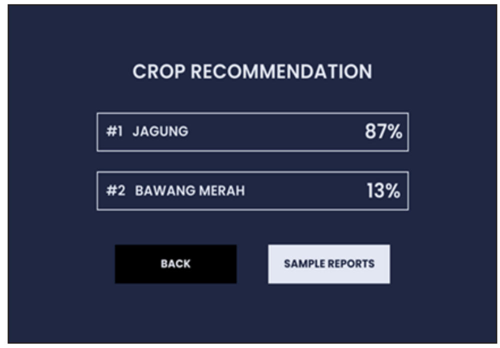

Figure 8. Crop recommendation interface

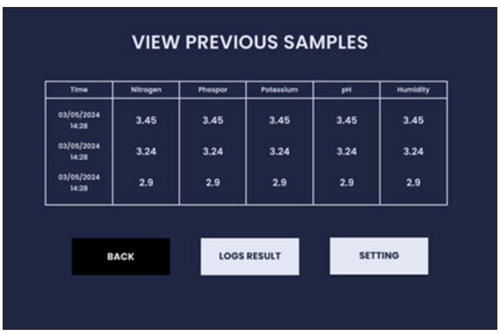

Figure 9. View sample interface

Figure 10. Proposed device

DISCUSSION

This study suggests a method based on ML and the IoT to collect data on soil parameters and suggest appropriate crops depending on the information gathered. The JXBS-3001 soil sensor, which measures soil characteristics like pH, moisture, and nutrient concentrations (NPK and TDS), is one of the main components. A Raspberry Pi Pico RP2040 serves as the data collection hub, connecting to the sensor via the RS-485 protocol. The TCP/IP and HTTP protocols are used to send sensor data to a server, enabling remote monitoring and real-time data processing. A portable device with a 3.5-inch TFT LCD screen provides farmers with analyzed soil data, facilitating data-driven decision-making in the field.

On the server, data from the sensors is processed using ML models, notably the DT algorithm, to recommend crops suited to the soil conditions. This algorithm functions as a classification tool, utilizing the measured soil parameters as inputs. It categorizes data into groups based on decision rules generated from the training dataset. The classification process helps identify the most suitable crop for specific soil conditions by analyzing soil pH, temperature, and moisture levels. Using IG and entropy, the algorithm selects the most effective attributes to split the data, resulting in accurate and reliable recommendations.

The primary advantage of this system lies in its use of the DT algorithm for crop recommendation. This algorithm enables the system to develop models that handle non-linear data, allowing precise predictions of the most suitable crops under complex soil conditions. Furthermore, the system's ability to process large volumes of data through a server integrated with ML models ensures efficiency in decision-making, a feat that traditional methods often struggle to achieve.

The integration of IoT and ML within this framework also highlights the benefits of automation and remote monitoring in agriculture. This system not only reduces farmers' workload but also enhances land management accuracy, particularly in selecting optimal crops. With real-time data collection, farmers can make informed decisions based on algorithm-driven analyses while continuously monitoring soil and crop conditions through a portable device equipped with a visual display.

Additionally, utilizing the server as a processing platform supports system scalability, allowing ML models to improve as more data is collected from sensors. This scalability opens the door to further development, such as integrating plant disease analysis modules or improving yield predictions. By employing this technology, the system can significantly increase agricultural productivity while promoting environmental sustainability by optimizing fertilization and irrigation using accurate data.

To ensure that the developed system can be used effectively by farmers, usability testing was conducted by considering three main aspects: efficiency, effectiveness, and system usability scale (SUS). The test results showed that the system achieved an efficiency and effectiveness level of 100%, indicating that users can complete tasks without obstacles and in an efficient time. The SUS score of 81 indicates a very positive user assessment, which classifies the system into the "excellent" category. The overall average of the three indicators reached 93%, confirming that the user interface design, designed with farmers' needs in mind, has resulted in an optimal usage experience in the agricultural context.

Finally, collaboration between researchers, agricultural practitioners, and technology developers can strengthen the implementation of these solutions in the field. A participatory approach to research and development will ensure that the resulting systems are relevant and adaptable to the real needs of farmers, which, in turn, will increase the adoption of technology within the agricultural sector. Through these steps, future research could make significant contributions to the development of more efficient and sustainable agricultural practices.

CONCLUSION

In this study, a combination of an IoT-based system and ML was successfully developed to collect soil parameter data and provide suitable crop recommendations. The use of the JXBS-3001 soil sensor and the Raspberry Pi Pico RP2040 device allowed for the accurate collection of soil condition data, which was then processed using a DT algorithm to generate data-driven recommendations.

The primary strength of this system lies in its ability to manage complex data and provide smarter solutions for agricultural decision-making. Additionally, the integration of IoT and ML not only enhances land management efficiency but also enables better real-time monitoring of soil and crop conditions.

Through this approach, farmers can make more informed decisions regarding crop selection and farming practices, ultimately contributing to increased productivity and sustainability in the agricultural sector. This research also opens up opportunities for further development, including the integration of weather analysis modules and plant disease detection, which could further strengthen the system in addressing the challenges of modern agriculture. Overall, the results of this study highlight the great potential of revolutionizing agricultural practices with a more technology- and data-driven approach.

FUTURE RESEARCH

In future work, scalability testing of the system is needed to evaluate its performance in the face of increasing data loads and higher user demands. These tests include the system's ability to handle large data volumes, the reliability of data transmission on networks with limited bandwidth, and the server's capacity to handle multiple simultaneous requests. All these tests aim to ensure the system can function efficiently and reliably on a larger scale and under varied field conditions.

Overall, the proposed system demonstrates significant potential in revolutionizing modern agriculture by introducing a smarter and more efficient approach. The integration of IoT with ML algorithms offers a comprehensive solution that enhances crop productivity, optimizes resource utilization, and reduces uncertainty in land management. Future work could include further development to support weather-based farming, plant growth pattern recognition, and more advanced pest detection systems.

This study outlines several future research directions that could expand the effectiveness of the proposed system. First, the development of more advanced ML models, such as ensemble algorithms or deep learning, could be applied to improve the accuracy of crop recommendation predictions. This approach would allow the system to capture more complex patterns in the data collected from sensors, while also enhancing the model's resilience to data variability.

Second, there might be further advantages to incorporating weather monitoring technology into the system. Farmers may gain a more comprehensive understanding of the environmental factors that could impact crop development by utilizing weather prediction modules. This would enable timelier, data-driven decision-making regarding planting, fertilization, and irrigation.

Furthermore, research could focus on developing early detection systems for plant diseases. By utilizing visual data or imagery collected from drones or cameras, pattern recognition algorithms could be employed to detect early signs of plant diseases, allowing for prompt and efficient intervention. This would improve crop resilience and reduce yield losses due to pest and disease outbreaks.

Additionally, more studies should examine environmental sustainability in relation to better farming methods. Research on the application of technology that facilitates precision farming methods could be combined to optimize land management while reducing water and fertilizer consumption. By doing this, the system could support initiatives to preserve ecosystem balance in addition to increasing productivity.

Although the proposed system has the potential to revolutionize modern agriculture, it is important to note that field validation has yet to be conducted. Future research will include comprehensive field trials to assess the system's performance under actual farming conditions. In terms of technical limitations, the current prototype has been designed for use in general outdoor environments but is not fully optimized for harsh field conditions such as heavy rain or extreme weather. In addition, one major challenge is the system's reliance on Wi-Fi or cellular networks for data transmission. In remote rural areas where connectivity is limited or unavailable, the system cannot transmit sensor data in real time, so crop recommendations cannot be determined. Overcoming this limitation will be a major focus of future system upgrades, potentially through the integration of local storage or offline processing capabilities.

ACKNOWLEDGEMENTS

This research activity is supported through Riset dan Inovasi untuk Indonesia Maju (RIIM) Kompetisi funding from the Indonesia Endowment Fund for Education Agency, the Ministry of Finance of the Republic of Indonesia, and the National Research and Innovation Agency of Indonesia (BRIN), according to contract numbers 128/IV/KS/11/2022 and 374/SAM4/PPM/2022. We also thank Telkom University, Indonesia, and Unit Pelaksana Teknis (UPT) Pertanian Wilayah 5 Kabupaten Garut, Indonesia, for supporting this research activity.

REFERENCES

Aditya, B. R., Hernawati, E., Gunawan, T., & Aji, P. (2024). Design an agricultural soil and environment monitoring system based on IoT. *Pertanika Journal of Science and Technology*, *32*(6), 2575–2589. https://doi.org/10.47836/pjst.32.6.09

Akintuyi, O. B. (2024). Adaptive AI in precision agriculture: A review: Investigating the use of self-learning algorithms in optimizing farm operations based on real-time data. *Research Journal of Multidisciplinary Studies*, *7*(2), 16–30. https://doi.org/10.53022/oarjms.2024.7.2.0023

Food and Agriculture Organization of the United Nations. (2021). *The state of food and agriculture 2021: Making agrifood systems more resilient to shocks and stresses*. FAO. https://doi.org/10.4060/cb4476en

Gaikwad, S. V., Vibhute, A. D., Kale, K. V., & Mehrotra, S. C. (2021). An innovative IoT based system for precision farming. *Computers and Electronics in Agriculture*, *187*, 106291. https://doi.org/10.1016/j.compag.2021.106291

García, L., Parra, L., Jimenez, J. M., Lloret, J., & Lorenz, P. (2020). IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture. *Sensors*, *20*(4), 1042. https://doi.org/10.3390/s20041042

Iskandar, A. F. (2021). *Kerabatani: Ubiquitous smart farming platform dilengkapi sistem rekomendasi untuk meningkatkan produktivitas pertanian* [Kerabatani: Ubiquitous smart farming platform equipped with a recommendation system to increase agricultural productivity]. [Unpublished Master's thesis]. Universitas Negeri Malang.

Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. F. (2023). Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. *Journal of Agriculture and Food Research*, *14*, 100880. https://doi.org/10.1016/j.jafr.2023.100880

Khoeri, I., & Mulyana, D. I. (2021). Implementasi machine learning dengan Decision Tree Algoritma C4.5 dalam penerimaan karyawan baru pada Pt. Gitareksa Dinamika Jakarta [Implementation of machine learning with Decision Tree Algorithm C4.5 in new employee recruitment at PT. Gitareksa Dinamika Jakarta]. *Jurnal Sosial dan Teknologi*, *1*(7), 615–623. https://doi.org/10.59188/jurnalsostech.v1i7.126

Kiruthika, S., & Karthika, D. (2023). IoT-based professional crop recommendation system using a weight-based long-term memory approach. *Measurement: Sensors*, *27*, 100722. https://doi.org/10.1016/j.measen.2023.100722

Lestari, I., Akbar, M., & Intan, B. (2023). Perbandingan algoritma *machine learning* untuk klasifikasi *amenorrhea* [Comparison of machine learning algorithms for amenorrhea classification]. *Journal of Computer and Information Systems Ampera*, *4*(1), 32–43. https://doi.org/10.51519/journalcisa.v4i1.371

Lokhande, D. B., Chavan, T., & Patil, D. P. (2022). Cloud computing information loading and security issues review. *International Journal of Research and Analytical Reviews*, *9*(4), 520–526.

Modi, D., Sutagundar, A. V., Yalavigi, V., & Aravatagimath, A. (2021). Crop recommendation using machine learning algorithm. In *5th International Conference on Information Systems and Computer Networks* (pp. 1–5). IEEE. https://doi.org/10.1109/ISCON52037.2021.9702392

Musanase, C., Vodacek, A., Hanyurwimfura, D., Uwitonze, A., & Kabandana, I. (2023). Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. *Agriculture*, *13*(11), 2141. https://doi.org/10.3390/agriculture13112141

Priyadharshini, A., Chakraborty, S., Kumar, A., & Pooniwala, O. R. (2021). Intelligent crop recommendation system using machine learning. In *5th International Conference on Computing Methodologies and Communication* (pp. 843-848). IEEE. https://doi.org/10.1109/ICCMC51019.2021.9418375

Qiu, Z., Ma, F., Li, Z., Xu, X., Ge, H., & Du, C. (2021). Estimation of nitrogen nutrition index in rice from UAV RGB images coupled with machine learning algorithms. *Computers and Electronics in Agriculture*, *189*, 106421. https://doi.org/10.1016/j.compag.2021.106421

Rajeswari, S., Suthendran, K., & Rajakumar, K. (2018). A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. *International Journal of Pure and Applied Mathematics*, *118*(8), 365–370.

Ramson, S. R. J., León-Salas, W. D., Brecheisen, Z., Foster, E. J., Johnston, C. T., Schulze, D. G., Filley, T., Rahimi, R., Soto, M. J. C. V., Bolivar, J. A. L., & Malaga, M. P. (2021). A self-powered, real-time, LoRaWAN IoT-based soil health monitoring system. *IEEE Internet of Things Journal*, *8*(11), 9278–9293. https://doi.org/10.1109/JIOT.2021.3056586

Rao, M. S., Singh, A., Reddy, N. V. S., & Acharya, D. U. (2022). Crop prediction using machine learning. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012033). IOP Publishing. https://doi.org/10.1088/1742-6596/2161/1/012033

Ridoy, A. A. I., Siddique, M. A. I., & Joyti, O. (2024). A machine learning-driven crop recommendation system with IoT integration. In *6th International Conference on Electrical Engineering and Information and Communication Technology* (pp. 812–817). IEEE. https://doi.org/10.1109/ICEEICT62016.2024.10534479

Senapaty, M. K., Ray, A., & Padhy, N. (2023). IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture. *Computers*, *12*(3), 61. https://doi.org/10.3390/computers12030061

Sihombing, P. R., & Arsani, A. M. (2021). Comparison of machine learning methods in classifying poverty in Indonesia in 2018. *Jurnal Teknik Informatika*, *2*(1), 51–56. https://doi.org/10.20884/1.jutif.2021.2.1.52

Singh, P., Kaur, M., & Bajaj, R. (2024). An IoT-enabled crop recommendation system utilizing MQTT for efficient data transmission to AI/ML model. In *7th International Conference on Circuit Power and Computing Technologies* (pp. 315–320). IEEE. https://doi.org/10.1109/ICCPCT61902.2024.10672704

Vemulapalli, S., Sri, M. S., Varshitha, P., Kumar, P. P., & Vinay, T. (2024). An experimental analysis of machine learning techniques for crop recommendation. *Nigerian Journal of Technology*, *43*(2), 301–308. https://doi.org/10.4314/njt.v43i2.13

Wang, X. (2022). Managing land carrying capacity: Key to achieving sustainable production systems for food security. *Land*, *11*(4), 484. https://doi.org/10.3390/land11040484

Yundari, V. S. (2017). *Sistem rekomendasi potensi lahan pertanian menggunakan metode* Simple Additive Weighting *(SAW)* [Agricultural land potential recommendation system using the Simple Additive Weighting (SAW) method] [Master's thesis, Universitas Pendidikan Indonesia]. UPI Repository. http://repository.upi.edu/29867/1/S_KOM_1000161_Title.pdf