Journal of Telecommunication Network (Jurnal Jaringan Telekomunikasi) Vol. 16, No.1 (2026)

Classification of Heavy Metal-Indicated Soil Types Using Decision Tree Algorithm and IoT-Based pH-Moisture Sensors

Dhea Adrika Zahro Asyhari[1], Hudiono[2*] and Hendro Darmono[3]

[1, 2, 3] Digital Telecommunication Network Study Program, Department of Electrical Engineering, State Polytechnic of Malang, 65141, Indonesia.
[1]dheaadrika17@gmail.com, [2]hudiono@polinema.ac.id, [3]hendro.darmono@polinema.ac.id

Abstract— **This study implements the Decision Tree algorithm for the classification of three soil types: Paddy Field, Well Excavation, and Lapindo Mud. The classification is performed based on two key parameters that influence soil characteristics: pH and Humidity. The dataset consists of 120 samples, partitioned using a ratio of 75% training data (90 samples) and 25% testing data (30 samples). The model testing results on the unseen data demonstrated very high performance, achieving an overall Accuracy of 93.33%. The analysis of performance metrics indicates that the model possesses strong discriminatory power across all classes. The Lapindo Mud class achieved perfect Recall (1.00), primarily driven by its unique alkaline pH (pH > 7.0) which acts as the root node separator. Furthermore, the Well Excavation class demonstrated perfect Precision (1.00), indicating absolute reliability in its predictions. The F1-Score for all classes exceeded 0.90. These results confirm that the Decision Tree Model is an efficient and effective method for identifying and classifying soil types based on pH and Humidity parameters.**

Keywords: Decision Tree, Heavy Metal, Internet of Things (IoT), pH Sensor, Soil Monitoring, Smart Agriculture, Soil Moisture Sensor, ESP32.

I. INTRODUCTION

Soil pollution by heavy metals has become a global environmental problem. Heavy metals such as copper have toxic properties that can damage soil fertility, develop a method for detecting heavy metals in soil using chemical sensors industrial, mining and chemical waste that is disposed of without proper management through accumulation in organisms. The mobility of heavy metals is greatly influenced to prevent further damage to the environment and health. Some become more easily absorbed by plants, causing structural damage Fertilizers containing heavy metals also increase the accumulation of toxic materials. and costs, such as Atomic Absorption Spectrophotometry (AA-Spectrophotometry), heavy metals such as copper, which can affect soil quality and yields The interaction between heavy metals, soil moisture, and pH is very necessary for large resources [1].

Therefore, more use of sensor technology can have a negative impact on plant and human health [1]. Metals major ecological losses and threatening human health if not addressed. in real-time in the field. Various studies have been conducted to increasingly pressing, especially in the agricultural sector. In Indonesia, agricultural activities 1 disrupt plant growth, and potentially contaminate the food chain Monitoring and detection of heavy metal content in soil is important both are the main causes of this contamination. In addition, the use of pesticides and by soil moisture and pH. In high humidity conditions, heavy metals Most conventional methods involve time-consuming laboratory analysis. which has high accuracy to low concentrations but requires in agricultural land [2][3]. Sludge -textured rice paddy soil is often exposed to plant tissue and metabolism [2]. Therefore, an understanding of agriculture.

Research shows that the accumulation of heavy metals in the soil. Therefore, this research focuses on the integration of sensor technology. (pH, soil moisture, and proximity) with the Decision Tree algorithm for detection heavy metals in agricultural soil. Infrared sensors will detect changes light spectrum associated with the presence of heavy metals, the pH sensor will measuring soil acidity which changes due to metal ions, and sensors. Conductivity will measure the ability of the soil to conduct electric current affected by heavy metal ions. By integrating IoT technology, the system. It is hoped that this will provide an efficient, fast and affordable solution for real-time monitoring of soil quality, enabling farmers to take data-driven decisions to maintain soil fertility and sustainability agriculture.

II. METHODOLOGY

This study employed an experimental research design aimed at developing and testing a smart monitoring and controlling system for utilizing household waste-derived biogas as an alternative energy source for a water heater. The system integrates multiple sensors, an ESP32 microcontroller, and machine learning (Naïve Bayes algorithm) to classify biogas quality and automate energy conversion through a biogas-powered generator.

A. Type of Research

The type of research used in this study is descriptive research. experiments with a systems-based research design

*Corresponding Author

approach. Research The experiment was chosen because it aimed to test the influence and effectiveness application of Internet of Things (IoT) technology and decision tree algorithms for monitoring and classifying heavy metal content in agriculture. Research This involves direct testing of the developed system, both fromboth hardware and software sides, to find out whether the system can functions according to the expected purpose, as shown in Fig. 1.

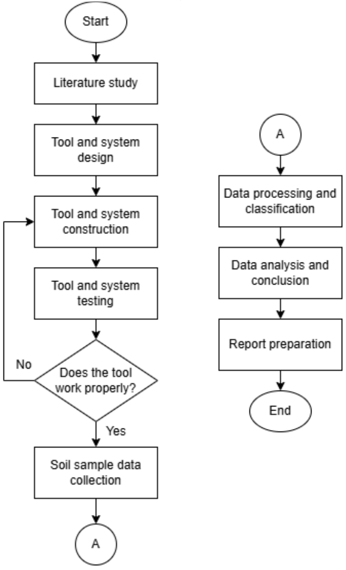

Fig. 1 Research Framework and Workflow

The research process begins with a literature review to gather foundational information for designing the tool and system. Based on this, an initial design is created and tested. If the tool does not function properly, improvements are made through a repeated cycle of design, testing, and repair until it works optimally. Once the tool operates as expected, the research moves to the data collection and analysis stage. Finally, a comprehensive report is prepared, covering the background, methodology, results, conclusions, and suggestions, marking the completion of the research.

B. Data Preprocessing

The data preprocessing stage is conducted to ensure the quality and validity of the raw data before it is integrated into the Decision Tree algorithm. This stage involves the following systematic steps: Sensor Calibration and Data Acquisition: Prior to data collection, the soil pH sensor undergoes a calibration phase using a linear regression method. This process is critical to minimize discrepancies between sensor readings and standard measuring instruments, resulting in an average error of only 5.8%. This ensures the consistency and reliability of the input values transmitted to the ESP32 microcontroller, which monitors both pH levels and soil moisture.

Data Validation and Filtering: The system performs an automated quality check on the data received from the sensors. Any invalid data points or reading errors caused by hardware instability are filtered out, and the system triggers a re-acquisition process until the data meets the predefined quality standards.

1. Dataset Construction and Labeling: A comprehensive dataset was constructed, comprising 120 samples. Each sample is labeled into one of three categories representing specific heavy metal (Iron/Fe) characteristics: Rice Field Soil, Well-Dug Soil, and Lapindo Mud. The classification is driven by two primary extracted features: pH value and moisture level.
2. Data Splitting: To facilitate model training and rigorous evaluation, the dataset is partitioned using a ratio of 75% for the training set (90 samples) and 25% for the testing set (30 samples). The training set is utilized to calculate Entropy and Information Gain to construct the decision tree structure, while the testing set is reserved exclusively for evaluating the model's accuracy and its ability to generalize to unseen data.

C. Data Collection and Analysis

A total of 120 samples were collected for this study. From this total, 90 samples were allocated as training data, derived from testing the three soil types (with each soil category providing a balanced data contribution). The remaining 30 samples were used as testing data for model validation.

The classification task is carried out using a decision tree algorithm, which is embedded in the ESP32 firmware to predict the heavy metal content in the soil based on the parameters read by the sensors. A decision tree is a tree structure, where each node represents a tested attribute. Each branch represents a distribution of test results, and leaf nodes represent specific class groups. The top-level node of a decision tree is the root, which is usually the attribute that has the greatest influence on a particular class. Decision trees generally employ a top-down search strategy for solutions. During the classification process, attribute values are tested by studying the path from the root node to the final leaf node, and then a new class is determined.

The mathematical basis of the Decision Tree algorithm is:

1. Entropy

Entropy measures the level of uncertainty (impurity) in a data set. If all data falls into one class, entropy = 0 (pure).

$$Entropy(S) = -\sum_{i=1}^{n} pi\,.\ \log_2(pi) \qquad (1)$$

Information:

- S: a subset of data
- pi: proportion of data from class i

- n: total number of classes

Example:

If in a dataset there are 4 “Low” data out of a total of 5, and 1 “Medium” data:

Entropy = −(0.8 log2 20.8+0.2 log2 20.2) = 0.722

2. Information Gain

Information Gain is used to select the best attribute as a split node.

$$Gain(S,A) = Entropy(S) - \sum_{v \epsilon Values(A)} \frac{|Sv|}{|S|} \,.\, Entropy(Sv) \qquad (2)$$

Information:

- S: main dataset
- A: evaluated attribute
- Sv: subset of data with value v on attribute A

Meaning:

The greater the Gain, the better attribute A is at separating the data.

3. Tree Development Process
    a. Calculate the total entropy of the dataset
    b. Calculate the information gain for each attribute
    c. Select the attribute with the highest gain → become the root node
    d. Divide the dataset based on that attribute's value
    e. Repeat steps 1–4 recursively until:
        - All data in the subset is homogeneous (entropy = 0)
        - No attributes remain.

The performance of the Decision Tree model was evaluated using a Confusion Matrix to derive quantitative insights into the system's classification capabilities. The analysis focuses on four key metrics: Accuracy, Precision, Recall, and F1-Score. The comparison between the algorithm's anticipated class and the actual class label derived from empirical observation such as flame tests or recognized quality ranges is represented by the confusion matrix. This matrix can be used to produce important performance metrics including accuracy, precision, and recall, which give information about how reliable the model is in practical settings.

TABLE 1
CONFUSION MATRIX

		TRUE	FALSE
Prediction Value	TRUE	(TP)	(FP)
	FALSE	(FN)	(TN)

In the confusion matrix table I there are True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). The four predictions have different meanings, where True Positive (TP) refers to instances where the predicted class matches the actual class. False Positive (FP) indicates instances where the predicted class is incorrect compared to the actual label. False Negative (FN) represents actual classes that were missed or wrongly predicted. True Negative (TN) is not explicitly shown in multi-class classification but is implicitly handled in per-class evaluation.

D. Model Training and Evaluation

The Decision Tree algorithm was chosen for the classification task due to its computational efficiency and effectiveness with continuous numerical data, making it ideal for implementation on an IoT device. The preprocessed dataset of 120 samples was split into two subsets, 75% for the training set (90 samples) to build the model, and the remaining 25% for the testing set (30 samples) to evaluate its performance on unseen data. The model's performance was quantitatively assessed using a confusion matrix, which provides a detailed summary of correct and incorrect predictions for each class. From this matrix, key performance metrics: Accuracy, Precision, Recall and F1 Score were used to measure the model's reliability and effectiveness in practice.

E. System Architecture and Material

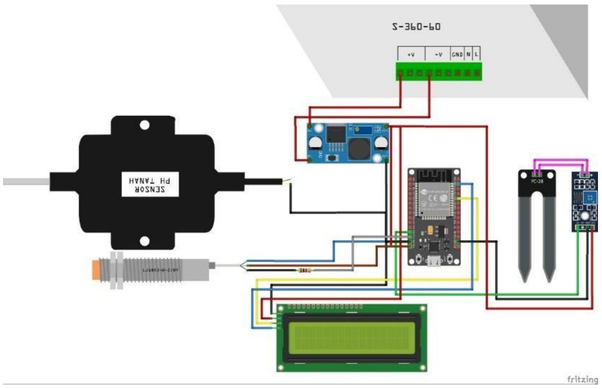

Fig. 2. Architecture System

a. ESP32 Microcontroller

The ESP32 is a feature-rich microcontroller with integrated Wi-Fi and Bluetooth connectivity for a wide range of applications, as shown in Fig. 2. It is an open-source microcontroller based on an input/output board. The ESP32 has 36 General Purpose Input/Output (GPIO), 14 of which are connected to an Analog to Digital Converter (ADC), allowing for the connection of an analog sensor. When a 3.3Volt voltage is connected to one of these ADC pins, the read value is 4095. The ESP32's required supply voltage is 2.2Volt to 3.6Volt. The ESP32 has a 12-bit ADC and an 8-bit DAC. The ESP32 has pins for I2C, SPI, and serial (Rx/Tx) communication. All pins on the ESP32 except the input-only pins (GPIO 34, GPIO 35, GPIO 36, and GPIO 39) can output PWM [14].

b. Soil Moisture YL69 Sensor

The YL-69 sensor, according to Prasetyo et al (2015), is a module that includes an LM393 IC, which functions as a comparator for offsets lower than 5mV, and has high stability and precision. The level of detection sensitivity can be adjusted by turning the potentiometer on the processing module. To get more accurate detection, we can use a microcontroller or Arduino with an analog output (connection to the ADC pin or analog input on the microcontroller) that will display the humidity value on a scale of 0 V (to GND) to Vcc (power supply voltage). This module can be operated with a power source between 3.3 volts and 5 volts [15]-[18]

c. Soil pH Sensor

This sensor works by detecting the ADC value obtained from the soil and then converting it to the soil pH. The ADC conversion is obtained using the equation:

$$y = -0.0099x + 16.396$$

The equation above shows how to convert the sensor's ADC value, where x is the sensor's ADC value and y is the converted value.

d. Stepdown LM2596

The LM2596 is a type of DC-DC (stepdown) voltage regulator that can convert a higher DC input voltage to a lower, more stable DC output voltage. The LM2596 is also compatible with microcontrollers and can be connected to various types of sensors and electronic devices that require a stable and efficient DC voltage. This module has 4 pins, 2 on the left and 2 on the right for input and output current.

e. Power Supply 12V 10A

This hardware comes in several power levels, one of which is the 12V power supply. The 12V power supply, which serves as the power source for all computer hardware, works by converting AC current to DC current.

The primary function of a power supply is to supply electrical current to computer components and hardware using DC (direct current). The current entering the power supply is AC (alternating current), which is then converted to DC (direct current).

f. Firebase Realtime Database

Firebase is a BaaS (Backend as a Service) currently owned by Google. Firebase itself is a solution offered by Google to simplify the work of mobile app developers. With Firebase, app developers can focus on developing their apps without having to invest significant effort in backend functions.

The Firebase Realtime Database itself is a NoSQL cloud-based database that synchronizes data across all clients in real time and provides offline functionality. Input data is stored in the Realtime database as JSON. All connected and sharing clients will automatically receive updates with the latest data. [16]

g. Visual Studio Code

Visual Studio Code (VS Code) is a cross-platform source code editor developed by Microsoft, specifically designed for software development with extensive programming language support. VS Code has an Electron-based architecture, allowing applications to run on Windows, macOS, and Linux with lightweight performance.

F. Flowchart System

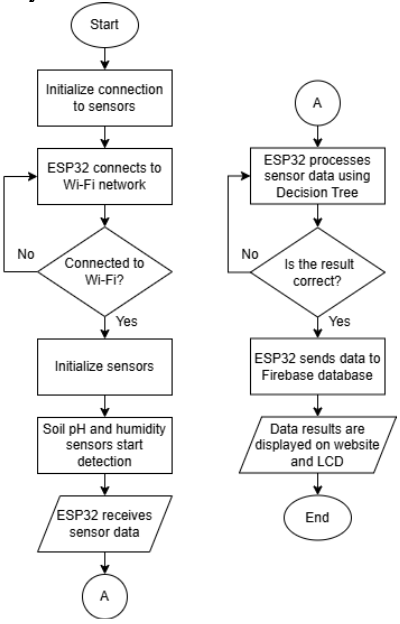

Fig. 3. Flowchart System

This Figure 3 illustrates the workflow of a heavy metal detection system in soil. The process begins with system initialization, where the ESP32 first establishes a connection to the sensors. Afterward, the ESP32 attempts to connect to a Wi-Fi network to ensure an internet connection is available. If the ESP32 is unable to connect to the Wi-Fi network, the system will continue trying until a connection is established.

After successfully connecting to the Wi-Fi network, the next step is to initialize the sensors, including the soil pH sensor and moisture sensor. These two sensors will begin working to detect pH and moisture levels in the soil, and send the resulting data to the ESP32. The ESP32 then processes the received data using a decision tree algorithm, which allows the system to classify or make decisions based on the values detected by the sensors.

After data processing, the system checks whether the results meet certain parameters or criteria. If the results are not, the system may repeat the data collection and processing process. However, if the results are, the ESP32 will send the data to the Firebase database. The final step of this process is data presentation, where the processing results are displayed via the monitoring website and LCD.

III. RESULTS AND DISCUSSION

The system's performance was evaluated through a series of tests focusing on hardware reliability, energy conversion efficiency, and the accuracy of the machine learning model.

1. Sensor Performance Evaluation

Sensor testing was conducted to determine the accuracy level of the data readings from each sensor used. The method employed in this evaluation involves a comparison between the sensor reading values and a previously calibrated reference measuring instrument. The error percentage (%) is calculated using the following Equation 3.1:

$$Error\ Relatif\ (\%) = \left|\frac{Sensor\ Values - Reference\ Values}{Reference\ Values}\right| \times 100\% \quad (3)$$

To calculate the average relative error (error percentage) from multiple sensor measurements, the following Equation 3.2 can be used:

$$Mean\ Error\ Relatif\ (\%) = \frac{1}{n}\sum_{i=0}^{n}\left(\left|\frac{Sensor_i - Reference_i}{Reference_i}\right| \times 100\%\right) \quad (4)$$

a. Results of pH Measurement Accuracy Testing

This section presents the data obtained from the pH sensor testing compared to the reference meter. Table II presents the comparison data between the soil pH sensor and a pH meter, following linear regression analysis across 10 trials for each sample. The results show an average error percentage of 1.33%, indicating that the calibration process significantly improved the accuracy compared to the pre-calibration state.

TABLE II
ACCURACY TESTING RESULT OF THE SOIL PH SENSOR

No.	ph ADC	pH Sensor	pH Meter	Difference	Error (%)
1	1102	5	5.5	1.1	20
2	1001	6.8	6.5	0.3	4.6
3	899	7.2	7.5	0.3	4
4	951	7	7	0	0
5	1053	6.6	6.5	0.6	10
6	982	6.8	6.7	0.1	1.5
7	1023	6.7	6.3	0.4	6.4
8	1079	6.3	5.7	0.6	10.5
9	968	6.9	6.8	0.1	1.5
10	991	6.6	6.6	0	0
Mean Error Relatif (%)					**1.33**

b. Results of Soil Moisture Measurement Accuracy Testing

Table III presents the data from the soil moisture sensor testing compared to a standard soil moisture meter. Based on the 10 test trials conducted, an average error percentage of 1.25% was obtained.

TABLE III
ACCURACY TESTING RESULT OF THE SOIL MOISTURE SENSOR

No	Soil Condition	YL-69 (%)	Three-Way Meter (%)	Difference	Error
1	Unwatered	20	21	1	5.00
2	Watered 1x	38	38	0	0
3	Watered 2x	50	50	0	0
4	Watered 3x	67	67	0	0
5	Watered 4x	74	75	1	1.35
6	Watered 5x	89	90	1	1.12
Mean Error Relatif (%)					1.25

2. Device Testing Results on Paddy Field Soil

Table IV displays the test results for paddy field soil. This testing includes pH data and moisture levels.

TABLE IV
DEVICE TESTING RESULT ON PADDY FIELD SOIL

No.	pH	Humidity (%)	Heavy Metal Content (Fe(mg/kg))
1	5.0	25	13.143
2	5.1	25	13.143
3	5.2	25	13.143
4	5.3	25	13.143
5	5.4	25	13.143
6	5.5	25	13.143
7	5.6	25	13.143
8	5.7	25	13.143
9	5.8	25	13.143
10	5.9	25	13.143
11	5.0	45	13.143
12	5.1	45	13.143
13	5.2	45	13.143

No.	pH	Humidity (%)	Heavy Metal Content (Fe(mg/kg))
14	5.3	45	13.143
15	5.4	45	13.143
16	5.5	45	13.143
17	5.6	45	13.143
18	5.7	45	13.143
19	5.8	45	13.143
20	5.9	45	13.143
21	6.0	65	13.143
22	6.3	65	13.143
23	6.6	65	13.143
24	6.9	65	13.143
25	6.1	85	13.143
26	6.4	85	13.143
27	6.7	85	13.143
28	6.2	95	13.143
29	6.5	95	13.143
30	6.8	95	13.143

3. Device Testing Results on Well-Dug Soil

Table V displays the test results for well-dug soil. This testing includes pH data and moisture levels.

TABLE V
DEVICE TESTING RESULT ON WELL-DUG SOIL

No.	pH	Humidity (%)	Heavy Metal Content (Fe(mg/kg))
1	6.0	25	26.924
2	6.1	25	26.924
3	6.2	25	26.924
4	6.3	25	26.924
5	6.4	25	26.924
6	6.5	25	26.924
7	6.6	25	26.924
8	6.7	25	26.924
9	6.8	25	26.924
10	7.0	25	26.924
11	6.0	45	26.924
12	6.1	45	26.924
13	6.2	45	26.924
14	6.3	45	26.924
15	6.4	45	26.924
16	6.5	45	26.924
17	6.6	45	26.924
18	6.7	45	26.924
19	6.9	45	26.924
20	7.0	45	26.924
21	5.0	65	26.924
22	5.2	65	26.924
23	5.4	65	26.924
24	5.5	65	26.924
25	5.7	65	26.924
26	5.8	65	26.924
27	5.0	85	26.924
28	5.3	85	26.924
29	5.6	85	26.924
30	5.9	85	26.924

4. Device Testing Results on Lapindo Mud

Table VI displays the test results for Lapindo mud. This testing includes pH data and moisture levels.

TABLE VI
DEVICE TESTING RESULT ON LAPINDO MUD

No.	pH	Humidity (%)	Heavy Metal Content (Fe(mg/kg))
1	7.1	25	16.435
2	7.4	25	16.435
3	7.7	25	16.435
4	7.9	25	16.435
5	8.0	25	16.435

No.	pH	Humidity (%)	Heavy Metal Content (Fe(mg/kg))
6	8.3	25	16.435
7	8.6	25	16.435
8	7.1	45	16.435
9	7.4	45	16.435
10	7.7	45	16.435
11	8.0	45	16.435
12	8.3	45	16.435
13	8.7	45	16.435
14	7.2	45	16.435
15	7.5	45	16.435
16	7.8	45	16.435
17	8.1	45	16.435
18	7.2	65	16.435
19	7.5	65	16.435
20	7.8	65	16.435
21	8.1	65	16.435
22	8.4	65	16.435
23	8.8	65	16.435
24	7.2	85	16.435
25	7.5	85	16.435
26	7.8	85	16.435
27	8.1	85	16.435
28	8.4	85	16.435
29	8.9	85	16.435
30	9.0	95	16.435

5. Decision Tree Structure

Figure 4 illustrates the resulting decision tree, which consists of the Root Node, Node 2, and Node 3. This tree was derived from the calculation of 90 training data points, comprising 30 samples of paddy field soil, 30 samples of well-dug soil, and 30 samples of Lapindo mud. This decision tree serves as the foundational rules for the Decision Tree algorithm implemented in this system.

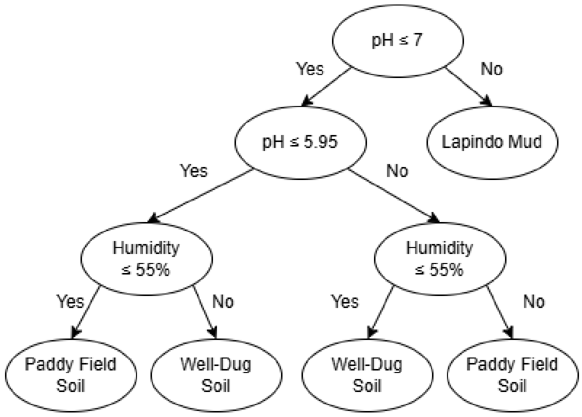

Fig. 4. Decision Tree Structure

6. Decision Tree Model Testing Results

The Decision Tree model utilizes a ratio of 75% training data and 25% testing data. From a total of 120 data points, 90 were used to train the model, while 30 were used to evaluate the performance on data that the model had not encountered during the training process, as shown in Table VII.

TABLE VII
DECISION TREE MODEL TESTING DATA

No.	pH	Humidity (%)	Model Prediction Result
1	5.0	41	Paddy Field Soil
2	5.3	33	Paddy Field Soil
3	5.6	54	Paddy Field Soil
4	5.8	48	Paddy Field Soil
5	5.9	55	Paddy Field Soil
6	6.1	67	Paddy Field Soil
7	6.4	82	Paddy Field Soil
8	6.7	71	Paddy Field Soil
9	6.9	93	Paddy Field Soil
10	6.8	60	Paddy Field Soil
11	5.1	79	Well-Dug Soil
12	5.4	42	Paddy Field Soil
13	5.5	81	Well-Dug Soil
14	5.7	81	Well-Dug Soil
15	5.9	85	Well-Dug Soil
16	6.0	28	Well-Dug Soil
17	6.3	49	Well-Dug Soil

No.	pH	Humidity (%)	Model Prediction Result
18	6.5	37	Well-Dug Soil
19	6.6	55	Well-Dug Soil
20	7.0	50	Well-Dug Soil
21	7.1	28	Lapindo Mud
22	7.5	45	Lapindo Mud
23	7.8	68	Lapindo Mud
24	8.0	31	Lapindo Mud
25	8.3	59	Lapindo Mud
26	8.5	74	Lapindo Mud
27	8.9	88	Lapindo Mud
28	9.0	95	Lapindo Mud
29	7.2	50	Lapindo Mud
30	7.6	70	Lapindo Mud

The testing data, consisting of 30 samples, was then used to evaluate the performance of the trained Decision Tree model. The evaluation was conducted by comparing the actual labels with the model's predicted results. The following is the Confusion Matrix of the model's prediction results. The model performance results are presented in Figure 5.

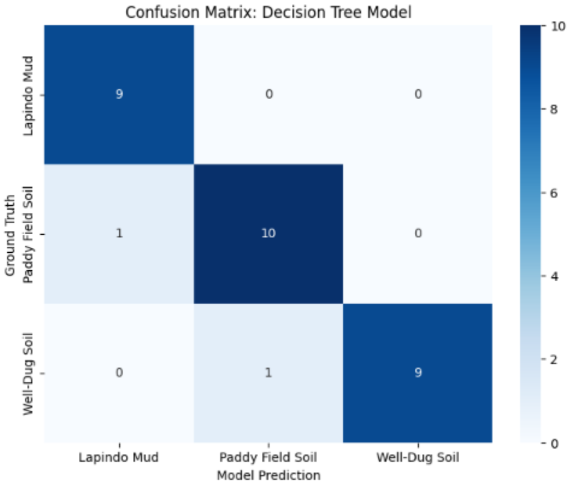

Fig. 5. Confusion Matrix Result

Based on Figure 5, the results of the confusion matrix calculation for accuracy, precision, recall, and F1-score from the data split using a 75:25 ratio are shown in Table VIII.

TABLE VIII
MODEL EVALUATION RESULT

The Result of Split Data 75:25				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
Paddy Field	0.91	0.91	0.91	11
Well-Dug Soil	1.00	0.90	0.95	10
Lapindo Mud	0.90	1.00	0.95	9
Accuration			**0.93** **93%**	**30**

Based on the confusion matrix and the model's prediction results presented in Figure 5 and Table 8, a performance evaluation can be conducted for the three prediction outcomes: “Paddy Field,” “Well-Dug Soil,” and “Lapindo Mud.” This evaluation process refers to key metrics such as Precision, Recall, and F1-Score. The detailed calculations are as follows:

- "Paddy Field" Class

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{10}{10 + 1} = 0.91$$

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{10}{10 + 1} = 0.91$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1\ Score = 2 \times \frac{0.91 \times 0.91}{0.91 + 0.91} = 0.91$$

- "Well-Dug Soil" Class

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{9}{9 + 0} = 1.00$$

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{9}{9 + 1} = 0.90$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1\ Score = 2 \times \frac{1.00 \times 0.90}{1.00 + 0.90} = 0.95$$

- "Lapindo Mud" Class

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{9}{9 + 1} = 0.90$$

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{9}{9 + 0} = 1.00$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1\ Score = 2 \times \frac{0.90 \times 1.00}{0.90 + 1.00} = 0.95$$

- Model Accuracy

Total data = 30

Total correct predictions (TP) = 9+9+10 = 28

$$Akurasi = \frac{Total\ TP}{Total\ Data} \times 100\%$$

$$Akurasi = \frac{28}{30} \times 100\% = 93.33\%$$

A. Discuccion

This section discusses the results of the tests conducted and provides an analysis of the obtained data to evaluate the system's performance.

1. Discussion of Hardware Testing Data

a. Soil pH Sensor

The accuracy of the soil pH sensor was tested against a pH meter as a reference instrument. The test results indicate that prior to calibration, the average sensor error reached 10.59%. After applying calibration via linear regression using ADC values and actual pH levels, the error rate dropped significantly to 1.33%. This demonstrates that the calibration process is effective in enhancing the sensor's reading accuracy relative to actual soil conditions. The pH sensor is highly sensitive to the aqueous solution within the soil, making ADC data correction crucial for achieving accurate readings. Figure 6 illustrates the comparison graph of the soil pH sensor values before and after calibration.

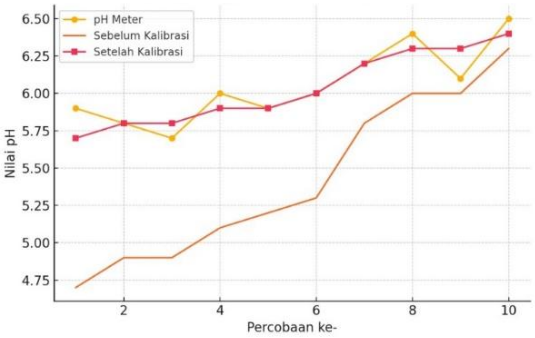

Fig. 6. Soil pH Sensor Accuration Graph

b. Soil Moisture Sensor

Figure 7 illustrates the comparison of soil moisture readings between the YL-69 sensor and the Three-Way Meter across six different moisture conditions, ranging from unwatered soil to soil watered five times. The horizontal axis represents the sequence of moisture conditions, while the vertical axis shows the soil moisture values in percentage (%). Both the blue line for the YL-69 sensor and the orange line for the Three-Way Meter exhibit a consistent upward pattern as the amount of watering increases, demonstrating that more frequent watering leads to higher moisture levels. The lines for both devices appear very close, even identical at several points, indicating that the YL-69 sensor readings have excellent accuracy compared to the reference tool. The small discrepancy between the two results in an average error of only 1.25%, leading to the conclusion that the YL-69 sensor is highly reliable for detecting soil moisture.

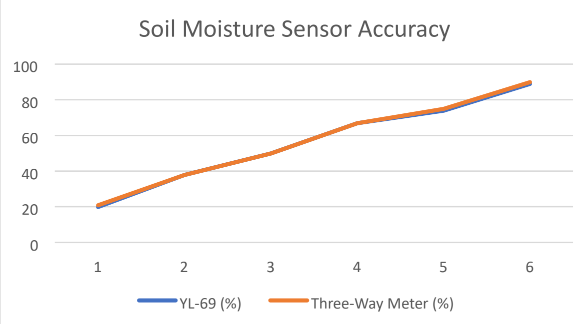

Fig. 7. Soil Moisture Sensor Accuration Graph

2. Discussion of Testing Data on Paddy Field Soil

Testing of the Paddy Field soil samples revealed clear characteristics in the relationship between pH and moisture, validating how the Decision Tree model classifies this soil type. In general, pH values for Paddy Field soil range from 5.0 to 6.9, falling within the acidic to neutral category. This consistently sub-7.0 pH value serves as a crucial hallmark, as it strictly distinguishes Paddy Field samples from Lapindo Mud samples, which possess an alkaline pH (pH > 7.0) and serve as the split boundary at the model's root node.

Further graphical analysis indicates a positive correlation between pH and moisture. At lower pH levels (approximately 5.0 to 5.9), moisture tends to be low (25% to 45%). Conversely, as the pH rises toward the neutral threshold (6.0 and above), moisture experiences a significant surge, peaking at up to 95%. This pattern directly justifies the effectiveness of the rules in the Decision Tree's sub-branches. The model relies on this combined characteristic: it classifies a sample as Paddy Field when low pH is combined with low moisture (pH ≤ 5.95 and moisture ≤ 55%), as well as when moderate pH is combined with high moisture (pH > 5.95 and moisture > 55%). The high moisture levels at higher pH points are the determining factor that prevents these samples from being misclassified as Well-Dug Soil, thereby ensuring high model accuracy. The graph for the Paddy Field testing data is shown in Figure 8.

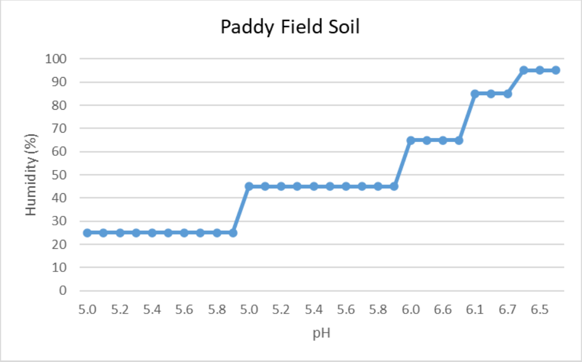

Fig. 8. Classification Prediction Graph for Paddy Field Soil

3. Discussion of Testing Data on Well-Dug Soil

Testing of the Well-Dug Soil samples shows a unique distribution pattern of pH and moisture, providing validation for the classification rules within the Decision Tree model. In general, pH values for Well-Dug Soil have a broad range, from 5.0 to 7.0, causing it to overlap with the Paddy Field soil range. Similar to Paddy Field soil, this pH value of $\le$ 7.0 ensures that Well-Dug Soil samples pass through Node 1 (Lapindo Mud) and are then classified based on moisture. Graphical analysis shows that the moisture factor becomes the crucial differentiator. When the pH is within the low to weakly neutral range, two prominent moisture segments emerge:

- Low Moisture Segment: At initial and intermediate pH levels, the moisture levels remain low, ranging from 25% to 45%.
- High Moisture Segment: Along with changes in pH, the moisture increases significantly, reaching a peak of up to 85%.

This pattern directly validates the two most critical rules within the model. Well-Dug Soil tends to be classified when the pH is in the medium zone (> 5.95) and moisture is relatively low (≤ 55%), or when the pH is in the low zone (≤ 5.95) but moisture is high (> 55%). This indicates that although pH does not completely separate Well-Dug Soil from Paddy Field soil, it is this specific combination of pH and moisture that the Decision Tree relies on to differentiate Well-Dug Soil. The graph for the Well-Dug Soil testing data is shown in Figure 9.

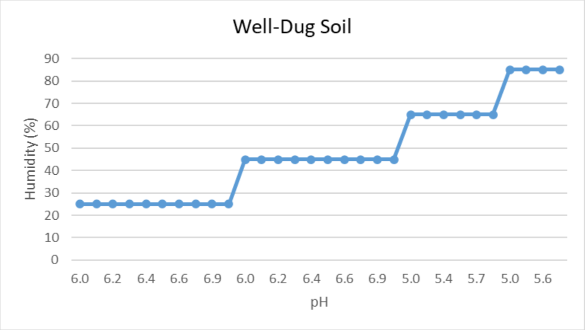

Fig. 9. Classification Prediction Graph for Well-Dug Soil

4. Discussion of Testing Data on Lapindo Mud

Testing of the Lapindo Mud samples reveals the most distinct characteristics compared to other soil types, particularly regarding pH values. Consistently, the pH values for Lapindo Mud fall within the alkaline range, starting from 7.1 to 8.9. This feature serves as the most dominant classification determinant. The Decision Tree model utilizes this alkaline pH trait directly at Node 1 (the root node), where all samples with pH > 7.0 are immediately classified as Lapindo Mud. Consequently, the Lapindo Mud classification achieves a 100% Recall and very high Precision, as these samples are almost never mispredicted. On the graph, although moisture values also vary from low (25%) to very high (95%), this moisture variation is irrelevant to the classification process for Lapindo Mud. Regardless of whether the moisture is at a low level (e.g., 25% at pH = 7.1) or a high level (e.g., 95% at pH = 8.9), as long as the pH value exceeds the 7.0 threshold, the resulting classification is inevitably Lapindo Mud. Therefore, for the model, the single feature of pH > 7.0 is sufficient to confirm the soil's identity, making the model highly efficient in separating this alkaline soil type from other acidic or neutral soil types. The graph for the Lapindo Mud testing data is shown in Figure 10.

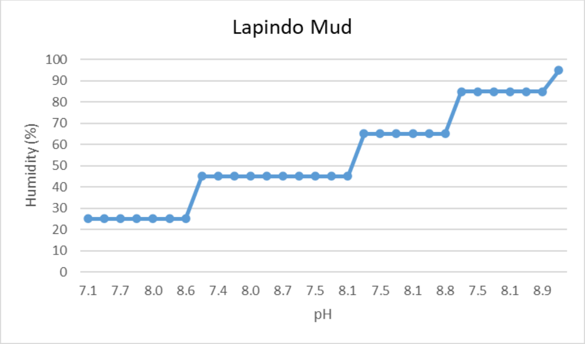

Fig. 10. Classification Prediction Graph for Paddy Field Soil

5. Discussion of Decision Tree Model Accuracy Testing Data

This discussion focuses on the performance evaluation of the Decision Tree model, which was tested using 30 testing data samples (75:25 ratio). The test results indicate that the model achieved a very high overall Accuracy of 93.33%, with 28 out of 30 predictions correctly classified. A more in-depth analysis through Precision, Recall, and F1-Score metrics reveals robust and balanced model performance across all classes. The following is the performance evaluation based on soil class.

a. Paddy Soil Class

The Paddy Field class demonstrates a balanced performance, with uniform Precision, Recall, and F1-Score values of 0.91. This balance reflects the model's equal capability in correctly predicting this class (Precision) as well as identifying all samples that should belong to this class (Recall). Similar to the Well-Dug Soil class, there was one False Negative case (one missed Paddy Field sample), indicating that Paddy Field classification relies more heavily on the combination of pH and moisture at a sensitive threshold (55%), making it more susceptible to overlapping with other classes.

b. Well-Dug Soil Class

The Well-Dug Soil class also demonstrates very strong performance, evidenced by a perfect Precision (1.00) and an equally high F1-Score (0.95). A Precision value of 1.00 indicates that every sample predicted as Well-Dug Soil by the model was correct, showing an absolute level of prediction reliability for this class. However, a Recall of 0.90 confirms that there was one False Negative case (one actual Well-Dug Soil sample that was missed and misclassified into another class), which represents one of the two total errors in this test.

c. Lapindo Mud Class

The Lapindo Mud class demonstrates superior performance with perfect Recall (1.00) and the highest F1-Score (0.95). A Recall value of 1.00 indicates that the model successfully identified all 9 actual Lapindo Mud samples. This dominant performance is due to the distinct alkaline pH characteristic (pH > 7.0) serving as the separating root node, ensuring these samples are isolated from other soil types at the beginning of the classification process. Although there was one False Positive case (one sample from another class was incorrectly predicted as Lapindo Mud), the strength of this pH feature makes the classification for this class highly reliable.

IV. CONCLUSION

An IoT-based monitoring system for heavy metal indications in soil using the ESP32 microcontroller, pH sensor, and moisture sensor has been successfully implemented and functions in real-time. Sensor data acquisition, classification results, and device connection status are accurately displayed on the monitoring dashboard website. The Decision Tree (ID3) algorithm was successfully implemented into the ESP32 firmware to classify soil types into Lapindo Mud, Paddy Field, and Well-Dug Soil. The decision logic based on split points of pH ≤ 7.0, pH ≤ 5.95, and Moisture ≤ 55% has proven to be optimal. Accuracy testing of the Decision Tree model against 30 valid test samples has proven effective in classifying the three soil types. Although two prediction errors occurred (1 missed Well-Dug Soil and 1 missed Paddy Field), the model's strength lies in its ability to identify dominant characteristics (Alkaline pH for Lapindo Mud) and integrate pH and Moisture features at high levels of overlap (Well-Dug and Paddy Field Soil). An accuracy rate of 93.33% and F1-Scores mostly above 0.90 indicate that the developed model is optimal and reliable for this classification case.

REFERENCES

[1] R. Novandi, ”Remediasi tanah tercemar logam timbal (Pb) menggunakan tanaman bayam cabut (Amaranthus tricolor L.),” Jurnal Teknologi Lingkungan Lahan Basah, vol. 2(1), 2014.

[2] P. Oktavian, M. Anas, I. N. Sudiana, och J. Safaani, ”Studi Kajian Literatur: Pengaruh Keberadaan Logam Berat Terhadap Tingkat Kesuburan Tanah di Indonesia.,” Einstein's: Research Journal of Applied Physics, vol. 2(1), pp. 20-23, 2024.

[3] R. Burchan, F. A. Soelistianto och K. Koesmarijanto, ”Telemonitoring Kadar EC pH dan Suhu Penyehatan Tanah Tercemar Menggunakan Sensor Konduktivitas Berbasis Web,” Journal of Telecommunication Network (Jurnal Jaringan Telekomunikasi), vol. 10(4), pp. 202-206, 2020.

[4] C. O. Handayani, S. Sukarjo och T. Dewi, ”Distribusi Logam Berat Pb, Cd, Cr, Ni dan Risiko Kesehatan Akibat Paparan Logam Berat Melalui Saluran Pencernaan di Lahan Sawah Sekitar Kawasan Industri Kabupaten Bandung.,” Jurnal Tanah dan Iklim, vol. 46(1), pp. 47-59, 2022.

[5] N. W. A. Ulandari, ”Implementasi Metode ID3 Untuk Mendeteksi Penyakit Tanaman Manggis Berbasis

Android.," Pros. Semin. Nas. Pendidik. Tek. Inform, pp. 217-222, 2018.

[6] S. C. Hung, C. C. Lu och Y. T. Wu, "An investigation on design and characterization of a highly selective LED optical sensor for copper ions in aqueous solutions.," Sensors, vol. 21(4), p. 1099, 2021.

[7] P. Destinugrainy, S. Cambaba och I. N. Surya, "Analisis Unsur Hara Karbon Organik dan Nitrogen Pada Tanah Sawah di Kecamatan Seko, Kabupaten Luwu Utara. Cokroaminoto," Journal of Biological Science, vol. 2(1), pp. 12-16, 2020.

[8] D. Sriwahyuni och A. Afdal, "Identifikasi pH, TDS, Konduktivitas Listrik, Kandungan Timbal (Pb), Tembaga (Cu), dan Kadmium (Cd) pada Bak Reservoir PDAM Kota Padang Panjang," Jurnal Fisika Unand, vol. 10(4), pp. 504-510, 2021.

[9] H. Marcos och H. Muzaki, "Monitoring Suhu Udara Dan Kelembaban Tanah Pada Budidaya Tanaman Pepaya.," Jurnal Teknologi dan Sistem Tertanam, vol. 3(2), 2022.

[10] W. Komarawidjaja, "Paparan limbah cair industri mengandung logam berat pada lahan sawah di Desa Jelegong, Kecamatan Rancaekek, Kabupaten Bandung.," Jurnal Teknologi Lingkungan, vol. 18(2), pp. 173-181, 2017.

[11] A. P. Kusuma, R. N. Hasanah och H. S. Dachlan, "DSS untuk menganalisis ph kesuburan tanah menggunakan metode Single Linkage.," Jurnal EECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems), vol. 8(1), pp. 61-66, 2014.

[12] Z. Zainudin och R. Kesumaningwati, "Pengaruh Eco Enzyme Terhadap Kandungan Logam Berat Lahan Bekas Tambang Batubara (Effect Of Eco Enzyme On Heavy Metal Content In Ex-Coal Mining Land)," Ziraa'ah Majalah Ilmiah Pertanian, vol. 47(2), p. pp. 154–161, 2022.

[13] G. H. Sandi och Y. Fatma, "Pemanfaatan Teknologi Internet of Things (Iot) Pada Bidang Pertanian.," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7(1), pp. 1-5, 2023.

[14] N. Cameron, "Electronics Projects with the ESP8266 and ESP32: Building Web Pages, Applications, and WiFi Enabled Devices," Edinburgh, UK: Apress, 2020.

[15] F. B. A. Putra och L. Hayat, "Rancang Bangun Miniatur Sistem Parkir Cerdas Bertingkat Berbasis Internet of Things Menggunakan ESP32.," Jurnal Riset Rekayasa Elektro, vol. 3(1), pp. 23-34, 2021.

[16] R. C. N. Santi och S. Eniyati, "Implementasi Statistik dengan Database Mysql.," Dinamik, vol. 20(2), 2015.

[17] A. H. Nasrullah, "Implementasi algoritma Decision Tree untuk klasifikasi produk laris.," Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar, vol. 7(2), pp. 45-51, 2021.

[18] M. Kurniawan och B. Kurniawan, "Implementasi Pemrograman Python Menggunakan Visual Studio Code.," JIK: Jurnal Informatika dan Komputer, vol. 11(2), pp. 1-9, 2020.