Vol. 8, No. 2, April 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Sensor-Driven Nutrient Monitoring Using a Two-Layer Machine Learning Model for Sugarcane Fertilization Recommendation

Fadiana[1], Didi Supriyadi[2,3], Daniel Yeri Kristiyanto[4,5,6], Isnaeni Nurul Agita[7]

[1,2,4]Information Systems Study Program, Telkom University, Purwokerto Campus, Purwokerto, Indonesia

[3,5]Center of Excellence for Sustainability Cities, Village, and Food Security, Research Institute for Intelligent Business and Sustainable Economy, Telkom University, Purwokerto, Central Java, Indonesia

[6]Department of Informatics, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia

[7]Agricultural Engineering Study Program, Jenderal Soedirman University, Purwokerto, Indonesia

Received:
September 26, 2025
Revised:
March 10, 2026
Accepted:
April 1, 2026
Published:
April 12, 2026

Corresponding Author:
Author Name*:
Didi Supriyadi
Email*:
didisupriyadi@telkomuniversity.ac.id

DOI:
10.63158/journalisi.v8i2.1547

Abstract. The growth of sugarcane requires optimal environmental conditions and the availability of balanced nutrients. However, fulfilling nutrition is a challenge because it requires targeted observation. The study proposes a machine learning-based decision support model using a predictive empirical approach to monitor nutrient needs and recommend fertilizer dosages. The proposed approach integrates field data with a two-layer modeling framework to support fertilization decision-making. The classification model predicts the status of nutrient adequacy, while the regression model estimates the level of fertilizer application. The target label (y) is generated through feature extraction using a rule-based empirical formula derived from the threshold of agronomic parameters. The nutrients analyzed included macronutrients (nitrogen, phosphorus, potassium) and micronutrients (iron, zinc, copper). Model development involves selecting the best-performing algorithm using recall for classification and RMSE and R^2 for regression. The results of the cross-validation showed that the Gradient Boosting algorithm achieved the most consistent performance, with a recall of 0.99 during training and >0.98 in holdout testing. The regression model also showed low RMSE and high R^2 values, especially for micronutrient estimation. The proposed model contributes to data-driven fertilization optimization.

Keywords: Decision Support Model, Fertilization Recommendation, Gradient Boosting, Sensor-based Monitoring, Sugarcane Nutrition

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Agriculture is one of the largest livelihood sectors in Indonesia [1]. The development of agriculture is an important indicator in the economic growth of a country [2]. According to Statistics Indonesia (BPS) in 2024, the agricultural sector contributed 12.53% to Indonesia's GDP in 2023 and increased to 13.78% in the second quarter of 2024. Several studies also highlight that the agricultural sector continues to make significant contributions to national economic output and regional development in Indonesia [3]. In the plantation subsector, sugarcane (Saccharum officinarum L.) is a strategic crop that plays an important role in sugar and ethanol production. Based on the recap of One Agricultural Data, the total national sugarcane production in 2024 will reach 2.46 million tons, with East Java Province as the largest contributor to sugarcane production in Indonesia [4]. East Java is the largest sugar-producing region in Indonesia, contributing approximately 40–47% of the national sugar production [5].

Sugarcane is a multi-year crop (ratoon crop) that has a relatively long growth period, which takes 8 to 12 months in a single planting period. Sugarcane growth is strongly influenced by physiological, biochemical and morphological variables that play an important role in the phytochemistry of plant complexes in the photosynthesis phase, cell division, cell wall arrangement, signaling molecules, secondary metabolites and plant protection against various biotic and abiotic stressors [9], [10], [7], [8]. In addition to environmental conditions, sugarcane requires the right source of nutrients, as it requires a variety of nutrients that include macro and micro elements to produce quality sugarcane with high sucrose levels and productivity [1], [9]. Sugarcane macronutrients include elements such as Nitrogen (N), Phosphorus (P) and Potassium (K). Nitrogen has a role to stimulate leaf and stem growth, Phosphorus promotes flowering growth and sugar formation, and Potassium increases sugar levels, drought and disease resistance [10]. Macronutrients are needed in large quantities, as they are highly influential in supporting vegetative growth and sugar formation [14]. Micronutrients also play an important role in the growth process of sugarcane, although the level of requirement is relatively small [15], [12], [16]. Micronutrient elements include Iron (Fe), Zinc (Zn) and Copper (Cu) [12]. Micronutrients play a role in the formation of chlorophyll, respiration, energy metabolism, hormone synthesis and the formation of *lignin* and oxidase enzymes [13],

[14]. These two elements require the right dosage of nutrients to maintain the stability of the sugarcane growth process [15].

Serious problems that farmers often face are the decline in productivity levels caused by drastic climate change, the decline of agricultural land, global warming, the decline in crop yields, a large amount of urbanization and a lack of knowledge about environmental factors [2], [9], [11]. The resilience of sugarcane is strongly influenced by the plant's ability to interact with biotic and abiotic factors, such as resistance to environmental changes (drought, high pH and extreme temperatures) as well as the plant's natural defenses against pest organisms [18], [16], [17]. Plant resistance can be increased through an adequate supply of nutrients, especially at the beginning of the planting period [18]. This study highlights the importance of adequate nutrient availability during the growth stage of sugarcane plants. Therefore, there is a need for an approach that can integrate environmental observations with data-driven modeling to support the monitoring of nutrient adequacy and assist in fertilizer decision-making. In this context, machine learning can be explored as a supporting tool to model complex relationships between environmental variables and nutrient conditions [19]. The dataset used in this study consists of real-time sensor readings collected from a 2m² demonstration plot planted with Bululawang sugarcane variety at the vegetative growth stage. The sensors employed include Nitrogen (N), Phosphorus (P), and Potassium (K) sensors to measure macronutrient content, as well as sensors for soil moisture, soil pH, temperature, and electrical conductivity to generate environmental indices correlated with both macronutrient and micronutrient conditions. Meanwhile, the data processing stage for modeling is conducted using a rule-based empirical formulation approach, in which the modeling process employs an empirical rule-based formulation to construct proxy target labels derived from agronomic thresholds and relationships among variables.

A decision support model was developed using a classification algorithm to categorize nutrient requirements and a regression algorithm to estimate the fertilizer dose based on the level of nutrient adequacy. This model was developed using the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework which is an international standard framework in guiding the data mining process and the development of machine learning models [20]. The development of this machine learning-based decision support model aims to support the management and monitoring of nutrients in sugarcane based

on sensor data [23], [24]. This model is in line with the Sustainable Development Goals (SDGs) section of the Zero Hunger section which emphasizes the importance of increasing agricultural productivity and sustainable food security [25], [26]. This research supports the achievement of Sustainable Development Goal (SDG) 2 – Zero Hunger, especially in targets 2.3 and 2.4 which focus on increasing agricultural productivity and sustainable food production practices.

Previous studies have extensively discussed nutrient monitoring in crops such as rice, corn, cereals, and even sugarcane. However, in general, existing research remains limited to monitoring plant macronutrients without explicitly addressing the elements of micronutrient [25], [26], [27]. A study by Islam et al. [28] also developing machine learning and IoT-based models for monitoring soil nutrients. However, the proposed approach focuses primarily on the process of collecting soil and plant nutrient data without comprehensively evaluating the level of nutrient adequacy in plants. On the other hand, a study by Asto Bonifacio et al. [29] develop a machine learning-based fertilizer dosing adjustment system using a combination of sensors and synthetic data. While this approach demonstrates the potential of integrating machine learning for fertilizer management, it may introduce dependencies on predicted outputs, which can influence the interpretation of the resulting recommendations. Based on the limitations in previous research, this study proposes a sensor-driven, empirical-label-assisted decision-support pipeline using a two-layered machine learning approach to support monitoring of sugarcane macronutrients and micronutrient adequacy and data-driven fertilization recommendations [27]. In this study, the proposed model is designed as a predictive machine learning framework for nutrient monitoring and fertilization recommendations that integrates sensor data and empirical rule-based labels. The model learns patterns from observations of real sensors, while the label is derived from an empirical formulation based on agronomic parameters. A two-layer approach is applied by determining the nutritional status through classification, then followed by the prediction of fertilizer dose based on the results of the classification [19]. The main contribution of this research lies in the development of a structured decision support pipeline that demonstrates how sensor data and empirical agronomic knowledge can be integrated within a two-layer machine learning framework. The proposed approach introduces an integrated framework that combines real-time sensor data acquisition, proxy label construction based on empirical formulas, and a two-layer machine learning pipeline

consisting of nutrient adequacy level classification followed by fertilization dose prediction. This research is the first step towards the development of more comprehensive agronomic modeling. However, there are limitations in this study regarding location, time and resources used. Therefore, further development and validation is indispensable to improve the generalization and implementation capabilities of the model in real practice.

2. METHODS

This research develops a pipeline by integrating sensor data, empirical label construction and two-layer machine learning modeling. The stages carried out in this study include understanding business needs, understanding data, data preparation, modeling, evaluation and implications as shown in Figure 1. This framework is used to ensure that the model development process is carried out in a systematic and structured manner.

Figure 1. Stages of model research and development

Based on figure 1, this research flow consists of several main stages that are adjusted to the CRISP-DM framework, namely: (1) field setup begins with the preparation of tools and materials in the research which includes data, tools and plot demonstration for data collection; (2) sensor data collection, which is the process of collecting data on the demonstration field; (3) Data preprocessing and empirical label construction are carried out by extracting features using an empirical formula to produce a label (Y), the formation of this label is based on the threshold of agronomic parameters. This stage also includes the process of cleaning and preparing the data; (4) model training is carried out by cross-validation to train the model, as well as evaluate the best model performance; (5) Testing and interpretation of results is carried out by testing the model using test data and external data.

2.1 Study Area and Experimental Setup

The object of this study is to monitor the nutritional needs of sugarcane plants of the Bululawang (BL) variety which include macronutrients and micronutrients. The research process was carried out through the implementation of demonstration plot media with an area of 2x1 meters^2. The soil tends to be clay-like in texture and dark in color. Figure 2 shows the conditions of the demonstration land applied for data collection.

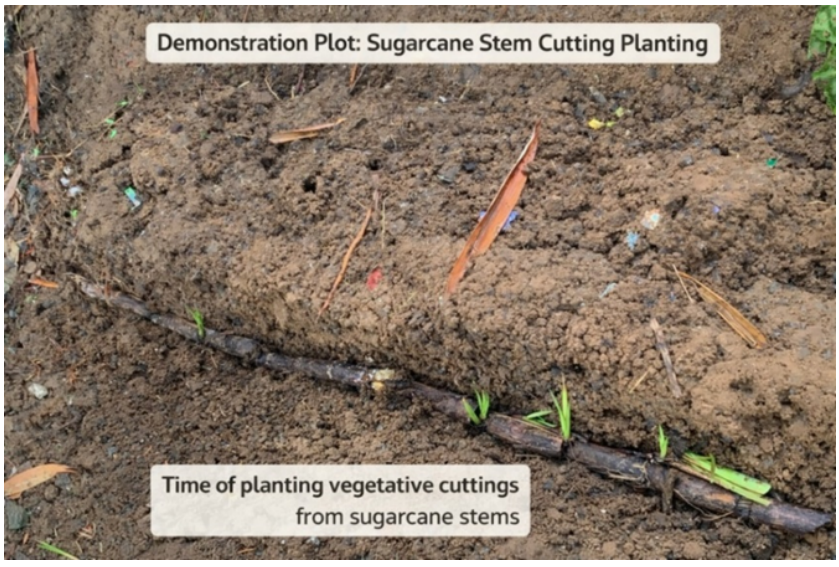

Figure 2. Demonstration plot for data collection (0 DAP)

The development of a machine learning prediction model was carried out by analyzing the optimal nutrient content for the growth of sugarcane plants in one planting period.

Based on the results of the agronomic literature review, optimal levels of macronutrients and micronutrients in sugarcane plants were obtained, as shown in Table 1.

Table 1. Range of macro and micro nutrient levels (mg/kg)

Element	Category		
	Low (mg/kg)	Medium (mg/kg)	High (mg/kg)
Nitrogen (N)	< 4	4-10	>10
Phosphorus (P)	< 3	3-9	>9
Potassium (K)	< 21	21-36	>36
Iron (Fe)	<9	9-18	>18
Zinc (Zn)	<0.6	0.6-1.2	>1.2
Copper (Cu)	<0.2	0.2-0.8	>0.8

Based on the reference provided in the book Analisis Kimia Tanah, Tanaman Air dan Pupuk [30], it is stated that the optimal macronutrient content for perennials is expressed as shown in Table 2. Meanwhile, micronutrients including Fe, Zn, Cu were mentioned in the study B. Vajanta et al. [31], referring to Diethyle Triamine Pentaacetic Acid (DTPA) data as shown in Table 1. Nutrient content is very important to determine the right fertilization projection based on nutritional adequacy conditions. The determination of fertilization schemes also requires proper scheduling. Previous research has stated that macronutrient and micronutrient fertilization tends to be carried out at the beginning of the planting phase [29]. The projection of macronutrient and micronutrient fertilization in this study is presented in Table 2.

Table 2. Fertilization projections in the sugarcane vegetative period

Element	Grant Scheme (Month)		
	0	1 -2	2.5 - 4
Nitrogen (N)	21.7gr	21.7gr	21.7gr
Phosphorus (P)	58gr	-	-
Potassium (K)	11.7gr	11.7gr	11.7gr
Iron (Fe)	6gr	conditional	conditional
Zinc (Zn)	6gr	conditional	conditional
Copper (Cu)	2gr	conditional	conditional

The fertilization scheme is determined based on the level of nutrient adequacy presented in Tables 2 and 3. The calculation of fertilizer application in each scheme is performed using formulas (1), (2), and (3), which are based on standard agronomic principles for estimating fertilizer requirements from plant nutrient demand and scaling them according to the land area [23].

$$U_{plot}\,(g) = U_{ha} \cdot 0.2 \tag{1}$$

Equation (1) is the result of adaptation to the mathematical formula used to calculate the estimated amount of fertilizer on a demonstration land of $2m^2$ with an average planting depth of 10 – 30 cm. The calculation is made based on the total fertilizer needed by sugarcane plants in one planting period in Table 3 presents the optimal amount of fertilizer application to sugarcane [33], [34].

Table 3. Total fertilizer needs for one sugarcane planting period

Element	Fertilizer	Total (Kg/ha^{-1})
N	Urea (46%N)	150
P	SP-36 (36%P_2O_5)	105
K	KCl (60%K_2O)	105
Fe	$FeSO_4$	30
Zn	$ZnSO_4$	30
Cu	$CuSO_4$	10

The fertilizer amounts presented in Table 3 are estimated based on standard agronomic principles, where fertilizer requirements are determined from nutrient adequacy adjusted to land area, as commonly adopted in fertilizer management and precision agriculture studies [26], [25].

$$Fertilizer(g) = \frac{U_{plot}(g)}{hara\ faction} \tag{2}$$

$$Dose(g) = Dose\ ha\ (kg) \cdot 1000 \cdot L \tag{3}$$

The calculation of the amount of fertilizer that can be applied is calculated using formulas (2) and (3). Formula (2) is used to calculate the dose of macronutrient fertilizer

in grams (gr) based on the plant nutrients and the fraction of nutrients in the fertilizer. Meanwhile, formula (3) is to calculate the dose of micronutrient fertilizer adjusted to the area of the demonstration land.

2.2 Data Acquisition and Construction

The data used in the machine learning modeling uses primary data from sensor readings in the demonstration field. The parameters produced from the sensor are presented in Table 4. related to the original data from the sensor readings.

Table 4. Sensor reading parameters

id	timestamp	nitrogen	phosporus	kalium	moisture	temperature	ph	conductivity	node
2	2025-05-x	55	45	60	15	25	5.5	60	1
6	2025-05-x	56	40	62	15	26	5.5	60	1
9	2025-05-x	26	42	60	15	27	5.5	60	2
10	2025-05-x	29	43	60	15	26	5.5	60	2

Table 4. Indicates the sensor reading parameters (original data). Sensor data was collected during a one-day observation period at high temporal resolution (approximately 8-15 seconds per reading) and generated 3,442 lines of data. Data collection was carried out in a single observation plot, so it represented a demonstration stage study and did not support extensive agronomic generalizations. The sensor reading results are processed at the data preprocessing stage to produce columns, parameters, indexes, and labels. The feature extraction process is carried out to produce micronutrient index columns using parameters (humidity, pH, temperature, conductivity). The micronutrient index serves to estimate the availability of micronutrient elements in the demonstration field. The resulting index columns are "index_Fe, index_Zn, index_Cu" using the empirical formulas (4), (5) and (6) constructed based on empirical relationships between soil parameters and micronutrient availability, as reported in previous studies such as Ivezić et al. [35].

$$Fe = 0.08 \cdot moist - 0.5 \cdot PH + 2 \quad (4)$$

$$Zn = 0.05 \cdot moist - 0.3 \cdot PH + 0.01 \cdot EC + 1.2 \quad (5)$$

$$Cu = 0.02 \cdot moist - 0.25 \cdot PH + 0.03 \cdot EC + 0.5 \quad (6)$$

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Equation (4) is an empirical model used to generate Fe index columns based on soil moisture and acidity level (pH) parameters. This index value is used as an indicator or relative estimate of the Fe content in the soil. The coefficients on the moisture variable show that the increase in soil moisture is related to the increased availability of the element Fe, making it easier to be absorbed by plants [35]. Meanwhile, the coefficients in the formula show that an increase in pH by one unit is related to a decrease in the value of the FE index of about 0.5. Higher pH conditions (above ~6.5) are known to reduce the availability of Fe and Zn elements [36]. Equation (5) and (6) are used to produce an estimated index of Zn and Cu availability in soil. The application of this formula is the same as the Fe formula, which shows that an increase in the conductivity value of 1 dS/m is related to an increase in the value of the Zn index of about 0.01 [34]. These three formulas are constructed previous studies Vladimir Ivezić's research [37], to generate index columns based on an empirical approach. The micronutrient data in this study It is not produced through direct chemical analysis, but is represented through an approach using empirical formulas.

The results of feature extraction using empirical Equation (4), (5) and (6) produce columns for the micronutrient elements 'index_fe', 'index_Zn', 'index_Cu', which are used in the label formation process in the micronutrient element classification and regression model. For macronutrients, no input index is formed, as the data is obtained directly through the NPK sensor shown in table 4. The target label (y) is generated using a rule-based empirical formula that is proxy without validation of laboratory measurements or independent field observations. Meanwhile, the modeling process applies a strict surveillance concept, where the input variables used are based on sensor data without involving empirical extraction features, thus preventing direct replication during model training. This approach aims to minimize the risk of data leakage, allowing models to study the relationship between environmental conditions, nutrient status, and fertilizer dose prediction through data-driven learning. The machine learning model serves as an approximation layer to the construction label, so that the model does not fully represent actual nutritional conditions in the field, but rather studies the representation of nutrients derived from sensor data and empirical formulations. In addition, the resulting model does not yet represent a ground-truth-based nutrition model, as the extracted and labeled features derived from empirical formulations (rather than direct sensor

readings) are not validated through laboratory measurements. Thus, the results of this modeling reflect a proxy-based approach to nutritional conditions.

2.3 Exploration and Preparation Data

This exploration stage is able to show the characteristics of the data used. The heatmap correlation in Figure 3. shows a high positive correlation relationship in the features of temperature, moisture and conductivity. In addition, the Fe, Zn, Cu indices also show a very strong correlation. This reflects the similarity of the forming factors in each index. The labels $FeSO_4$, $ZnSO_4$, and $CuSO_4$ also showed a strong correlation. All of these columns are applied to the model by applying feature selection. Before modeling, the dataset is prepared to clear the data from missing value, inconsistency and outliers. The outlier handling here is carried out using the Interquartile Range (IQR) approach with a rule of 1.5 x IQR and limiting the extreme value by capping or winsorizing method to the bottom-top limit.

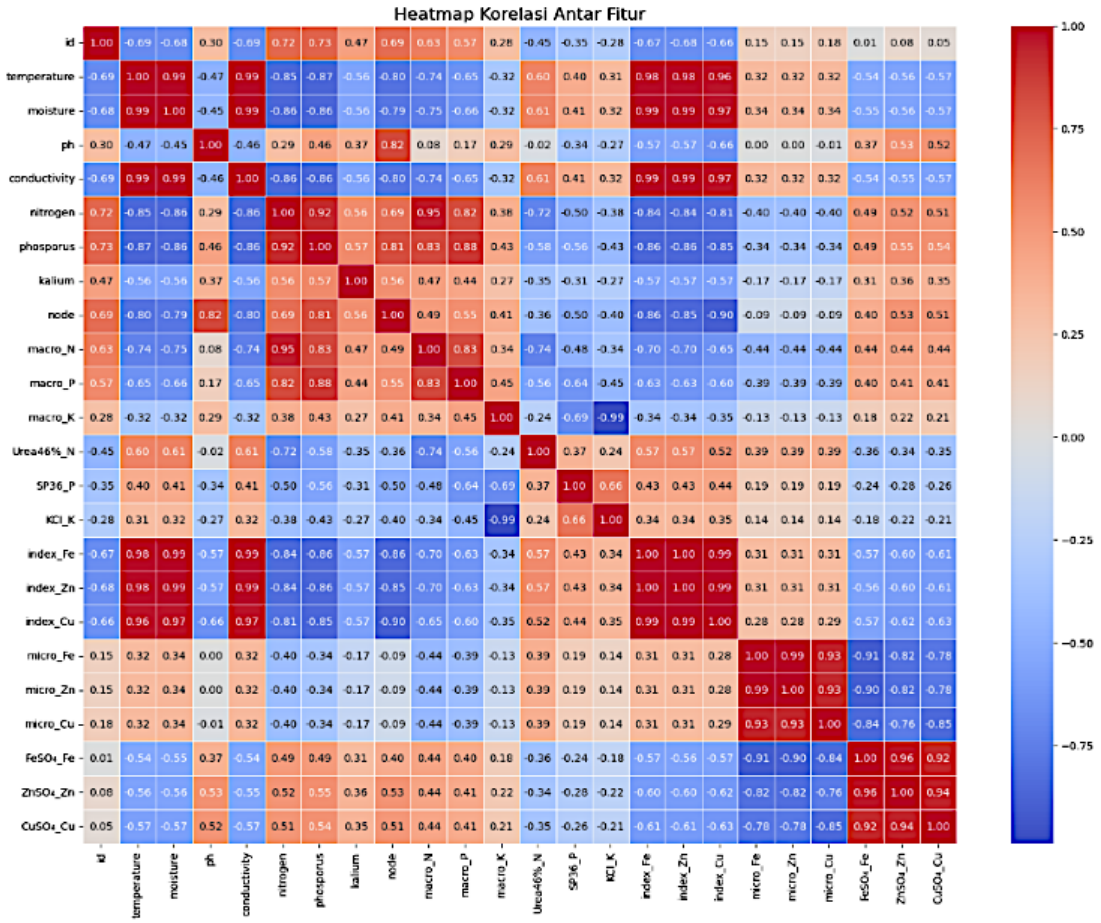

Figure 3. Heatmap Correlation between features in a dataset

2.4 Modeling

The modeling stage is carried out by defining the input and label features to produce the classification labels 'macro_N', 'macro_P', 'macro_K' and 'micro_Fe', 'micro_Zn', 'micro_Cu'. In the modeling process, the input features are derived from the reading column of the original sensor. Therefore, the variables used in the construction of the empirical label are excluded from the input features during the training and testing phases. This design is used to reduce the potential for data leakage and maintain separation between label construction and input features. Table 5. presents a list of features applied to classification and regression prediction models.

Table 5. Features for the development models

Columns Feature	Input (x)	Label (y)	Exception Feature
Classifier Model	*phosporus*, *kalium*, moisture, ph, temperature, conductivity	macro_N	nitrogen
	nitrogen, *kalium*, moisture, ph, temperature, conductivity	macro_P	*phosporus*
	phosporus, nitrogen, moisture, ph, temperature, conductivity	macro_K	*kalium*
	phosporus, nitrogen, *kalium*, moisture, ph, temperature, conductivity	micro_Fe	index_Fe
	phosporus, nitrogen, *kalium*, moisture, ph, temperature, conductivity	micro_Zn	index_Zn
	phosporus, nitrogen, *kalium*, moisture, ph, temperature, conductivity	micro_Cu	index_Cu
Regressor Model	*phosporus*, *kalium*, moisture, ph, temperature, conductivity	Urea46%_N	nitrogen, macro_N
	nitrogen, *kalium*, moisture, ph, temperature, conductivity	SP36_P	*phosporus*, macro_P
	phosporus, nitrogen, moisture, ph, temperature, conductivity	KCl_K	*kalium*, macro_K
	phosporus, nitrogen, *kalium*, moisture, ph, temperature, conductivity	$FeSO_4$_Fe	index_Fe, micro_Fe
	phosporus, nitrogen, *kalium*, moisture, ph, temperature, conductivity	$ZnSO_4$_Zn	index_Zn, micro_Zn
	phosporus, nitrogen, *kalium*, moisture, ph, temperature, conductivity	$CuSO_4$_Cu	index_Cu, micro_Cu

The dataset for modeling is 80% used for training and 20% for testing. The total rows in the dataset are 3,442 rows. Based on this division, around 2,754 rows are used for training and 688 lines for testing models. This data sharing is applied to classification and regression models. The modeling process is carried out by comparing several algorithms to identify the sensitivity and ability of the model to produce predictions.

2.5 Model Evaluation and Performance Metrics

Model evaluation is carried out to determine the quality and performance of each model. This study conducted a comparative test of the ensemble learning algorithm based on classification with an evaluation formula (7)-(10) and regression with an evaluation Equation (11)-(14).

$$Precision = \frac{TP}{TP+FP} \tag{7}$$

Measuring the accuracy of predictions with positive true values from all positive prediction results [38], [39].

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

Recall is the main metric in the evaluation of the developed model. Recall measures the model's ability to predict the true level or sensitivity of positive [38].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

Evaluation metrics in a classification algorithm to measure the value of the proportion of the "true" prediction to the overall value of the resulting prediction [38].

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \tag{10}$$

The F-1 score serves to summarize the overall results of precision and recall calculations by making harmonic averages [38].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{11}$$

The result of the square root of the MSE value. RMSE serves to present errors in the original unit of value [40].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_1)^2 \quad (12)$$

Calculate the square mean of the difference between the actual value and the predicted value [41].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_1| \quad (13)$$

MAE is used to measure the average absolute difference between the predicted value generated and the actual value without considering positive or negative values [40].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (14)$$

A metric to measure the proportion of variance in dependent and independent variables. An *R-squared* value close to 1 indicates better model capability [41].

3. RESULTS AND DISCUSSION

3.1 Exploration & Preprocessing Results

The processing of the dataset using empirical formulas yielded target columns for the nutrient requirement classification model and regression models for fertilization prediction based on nutrient adequacy.

	id	temperature	moisture	ph	conductivity	nitrogen	phosporus	kalium	node	macro_N	macro_P	macro_K	Urea46%_N	SP36_P	KCl_K	index_Fe	index_Zn	index_Cu	micro_Fe	micro_Zn	micro_Cu	$FeSO_4$_Fe	$ZnSO_4$_Zn	$CuSO_4$_Cu
0	2.0	15.0	15.0	5.0	60.0	55.0	45.0	60.0	2.5	1.0	1.0	1.0	0.0	0.0	0.0	1.2	0.8	-0.20	0.0	0.0	0.0	0.0052	0.0000	0.004
1	6.0	15.0	15.0	5.0	60.0	56.0	40.0	62.0	2.5	1.0	1.0	1.0	0.0	0.0	0.0	1.2	0.8	-0.20	0.0	0.0	0.0	0.0052	0.0000	0.004
2	9.0	14.0	15.0	5.0	60.0	57.0	42.0	60.0	2.5	1.0	1.0	1.0	0.0	0.0	0.0	1.2	0.8	-0.20	0.0	0.0	0.0	0.0052	0.0000	0.004
3	10.0	15.0	15.0	5.0	60.0	56.0	43.0	60.0	2.5	1.0	1.0	1.0	0.0	0.0	0.0	1.2	0.8	-0.20	0.0	0.0	0.0	0.0052	0.0000	0.004
4	12.0	15.0	15.0	5.0	60.0	58.0	40.0	61.0	2.5	1.0	1.0	1.0	0.0	0.0	0.0	1.2	0.8	-0.20	0.0	0.0	0.0	0.0052	0.0000	0.004
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3437	3517.0	15.0	13.3	6.2	55.3	57.6	47.6	99.6	5.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	-0.78	1.0	1.0	1.0	0.0000	0.0000	0.000
3438	3518.0	15.5	14.7	6.2	59.6	59.0	45.8	58.7	5.0	1.0	1.0	1.0	0.0	0.0	0.0	0.1	0.1	-0.75	1.0	1.0	1.0	0.0000	0.0000	0.000
3439	3519.0	15.0	12.1	6.2	68.0	57.6	45.8	163.8	5.0	1.0	1.0	1.0	0.0	0.0	0.0	-0.1	-0.1	-0.81	0.0	0.0	1.0	0.0061	0.0066	0.000
3440	3520.0	15.1	15.4	6.2	57.7	59.3	46.7	98.7	5.0	1.0	1.0	1.0	0.0	0.0	0.0	0.1	0.1	-0.74	1.0	1.0	1.0	0.0000	0.0000	0.000
3441	3521.0	15.1	16.4	6.2	58.5	59.4	47.6	99.8	5.0	1.0	1.0	1.0	0.0	0.0	0.0	0.2	0.2	-0.72	1.0	1.0	1.0	0.0000	0.0000	0.000

3442 rows × 24 columns

Figure 4. Preprocessing dataset for model development

The amount of data used for modeling is 3,442 rows and 24 columns shown in Figure 4. The results of the dataset preprocessing result in the 'macro_N', 'macro_P', 'macro_K' column as the macronutrient classification model label. The 'micro_Fe,' 'micro_Zn', 'micro_Cu' column as the label of the micronutrient classification model. The columns 'Urea46%_N', 'SP36_P', 'KCl_K' as the macronutrient regression label and '$FeSO_4$_Fe', '$ZnSO_4$_Zn', '$CuSO_4$_Cu' as the micronutrient regression label. Meanwhile, the number of unique values for each input and label feature is presented in Table 6.

Table 6. Exploration of the unique value of input and label features

No.	Feature/Label	Unique Value
1	temperature	43
2	moisture	109
3	ph	16
4	conductivity	180
5	nitrogen	203
6	*phosporus*	124
7	*kalium*	1477
8	index_Fe	39
9	index_Zn	28
10	index_Cu	88
11	macro_N	2
12	macro_P	2
13	macro_K	2
14	micro_Fe	2
15	micro_Zn	2
16	micro_Cu	2
17	Urea46%_N	101
18	SP36_P	5
19	KCl_K	9
20	$FeSO_4$_Fe	23
21	$ZnSO_4$_Zn	34
22	$CuSO_4$_Cu	26

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The modeling dataset consists of 24 columns shown in Figure 4 In the modeling practice, feature selection is applied to select relevant input features in the model prediction. Meanwhile, the number of unique values in each column is shown in Table 6 with labels that provide an overview of the characteristics of the data distribution and the types of variables used in the modeling. The classification label indicates the sum of the unique values = 2 which indicates a binary output.

```
target_biner = ['macro_N', 'macro_K', 'macro_P',
'micro_Fe', 'micro_Zn', 'micro_Cu']
target_biner = [c for c in target_biner if c in dt.columns]
target_cont = ['Urea46%_N', 'SP36_P', 'KCl_K','FeSO4_Fe', 'ZnSO4_Zn', 'CuSO4_Cu']
target_cont = [c for c in target_cont if c in dt.columns]
auto_low_var = [
        c for c in numeric_cols
        if (dt[c].nunique() <= 2) and (c not in target_cont)
]
exclude_cols = list(set(target_biner + auto_low_var))
```

Figure 5. Outlier handling by excluding labels

The preprocessing stage consists of the process of handling outlier values in the dataset. Figure 5 shows a piece of code to handle outliers by excluding labels that aim to avoid data leakage and target distortion. In addition, labels need to be maintained to maintain the validity of predictions and interpretation of results. Meanwhile, outlier handling is carried out using the capping method to control the extreme values of the sensor's real-time reading data without removing important information and changing the data distribution structure.

```
lower = Q1 - 1.5 * IQR
  upper = Q3 + 1.5 * IQR
  dt[col] = dt[col].clip(lower=lower, upper=upper)
```

Figure 6. Code snippets to limit outlier values with upper and lower limits

Figure 6 shows the implementation of the capping method aimed at limiting the extreme values detected using IQR. This process controls the value of the sensor readings that are too extreme without erasing the original data.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

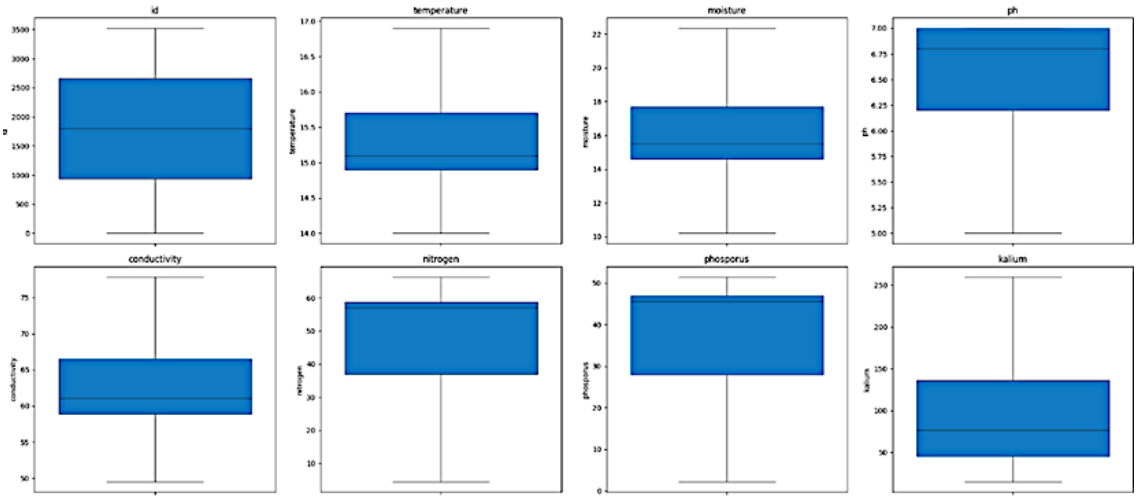

Figure 7. Visualization of the input feature boxplot after outlier handling

The results of the outlier identification show that the input columns 'temperature', 'humidity', 'ph', 'conductivity', and 'nitrogen' have multiple outliers. The handling process is carried out by limiting the upper and lower limits as shown in Figure 6. The results of outlier handling using capping are shown in Figure 7, where the distribution of data becomes more controlled with the reduction of extreme values, while the distribution and median pattern of each variable are maintained.

3.2 Modeling: Training and Testing

The modeling process is carried out in 2 stages, namely classification and regression modeling. The input features for each model are presented in Table 5. The modeling process involves comparisons with several ensemble learning models such as Random Forest, Extra Trees, Bagging, Adaboost, Gradient Boosting, LightGBM, and XGBoost. The results of the overall training of the model showed that Gradient Boosting was the model with the best performance as evidenced by the high level of sensitivity in the classification model and low RMSE for regression.

Table 7 shows the results of the evaluation on the classification and regression model in the training process. Based on the results of the training and model testing carried out, it is known that the Gradient Boosting model is the model with the most consistent performance, when compared to several other algorithms. The results of the training using cross-validation (CV) showed a high recall rate, which was 0.97-0.99, which shows the model's ability to recognize data patterns. The results of cross-validation training on

the regression model showed the RMSE value = 0.0002-2.35 and the R2 value = 0.59 – 0.99. The results of the evaluation of these two metrics reflect the ability of regression models to model the relationship between input variables and predicted target values.

Table 7. Results of classification and regression model training

Model	Train-Clf (CV)				Train-Reg (CV)			
	Recall	Prec	F1	Acc	RMSE	R^2	MSE	MAE
GB-Nitrogen	0.990664	0.994954	0.992803	0.988379	2.351664	0.593970	5.530326	0.851311
GB-*phosporus*	0.991364	0.975617	0.983382	0.971818	1.120193	0.857468	1.254833	0.242310
GB-*kalium*	0.976816	0.983147	0.979939	0.961941	0.294076	0.983355	0.086481	0.012569
GB-Fe	0.997143	0.999286	0.998211	0.998547	0.000301	0.989876	9.080744e-08	0.000072
GB-Zn	0.998571	0.999286	0.998928	0.999128	0.000227	0.996238	5.162323e-08	0.000068
GB-Cu	0.997857	0.999288	0.998569	0.998837	0.000283	0.997471	7.994555e-08	0.000046

Table 8. Results of model testing with the holdout method

Model	Test-Holdout-Clf				Test-Holdout-Reg			
	Recall	Prec	F1	Acc	RMSE	R^2	MSE	MAE
GB-Nitrogen	0.9946 ↑	0.9928 ≈	0.9937 ↑	0.9898 ↑	1.961717 ↓	0.714207 ↑	3.848334 ↓	0.817467 ↓
GB-phosporus	0.9879 ↓	0.9812 ↑	0.9845 ↑	0.9739 ↑	0.589479 ↓	0.969100 ↓	0.347486 ↓	0.148748 ↓
GB-*kalium*	0.9848 ↑	0.9833 ≈	0.9840 ↑	0.9695 ↑	1.633121 ↑	0.511879 ↑	2.667083 ↓	0.538235 ↑
GB-Fe	0.9929 ↓	0.9964 ≈	0.9946 ≈	0.9956 ≈	0.000427 ≈	0.979578 ≈	0.000000 ↑	0.000172 ≈
GB-Zn	1.0000 ↑	1.0000 ↑	1.0000 ↑	1.0000 ↑	0.000424 ≈	0.986864 ≈	0.000000 ↓	0.000134 ≈
GB-Cu	0.9964≈	1.0000 ↑	0.9982 ≈	0.9985 ≈	0.000347 ≈	0.996154 ≈	0.000000 ↓	0.000069 ≈

Table 8 presents the results of the test with the holdout method showing that most models are able to maintain relatively consistent performance between training and testing. This consistency indicates the model's level of stability to the distribution of data,

particularly in the context of the proxies-based labels used in this study. The nitrogen, potassium and zinc classification model shows high performance on the test data. This indicates that the model does not experience a significant performance degradation on the same data share. In the regression model, the elements nitrogen and phosphorus also showed relatively stable performance with relatively low RMSE values and high R^2 values. This reflects the model's ability to represent the relationship between input features and labels based on empirical constructions, derived from the agronomic rules in this study. However, the potassium regression model experienced a decline in performance influenced by a narrower distribution of test data, higher variability, and a more complex relationship between environmental parameters and potassium content than other nutrient elements. In addition, potassium has more dynamic agronomic characteristics mainly because it is influenced by interactions between other elements, making the relationship between sensor variables and labels more difficult to model accurately. In general, the results of these holdout tests interpret the model's internal stability indication of the observed data distribution, rather than as evidence of the model's broad agronomic generalization capabilities.

3.3 Interpretation of Model Merging Results

This section explains the interpretation of research results in developing machine learning-based decision support models. The results of the evaluation and comparison test have been carried out through the results of training and testing in Tables 8 and 9. A visualization of the results between the actual value and the prediction is shown in Figure 8.

Figure 8 presents a visualization of the predicted value results to the actual value based on the test data. The process continues by integrating classification and regression models, then storing and retesting with new data. Most prediction points are near the diagonal line, which indicates that the model's predictions are generally consistent with the actual values. However, some deviations can still be observed, especially in the potassium fertilization model, which shows a slightly lower level of prediction accuracy compared to other nutrient elements. This is influenced by the uneven distribution of potassium data, higher value variability and the complex relationship between environmental parameters and potassium content. Nonetheless, the results of follow-up

testing showed that the model was still able to follow the general patterns in the data, albeit with a greater degree of deviation than the other elements.

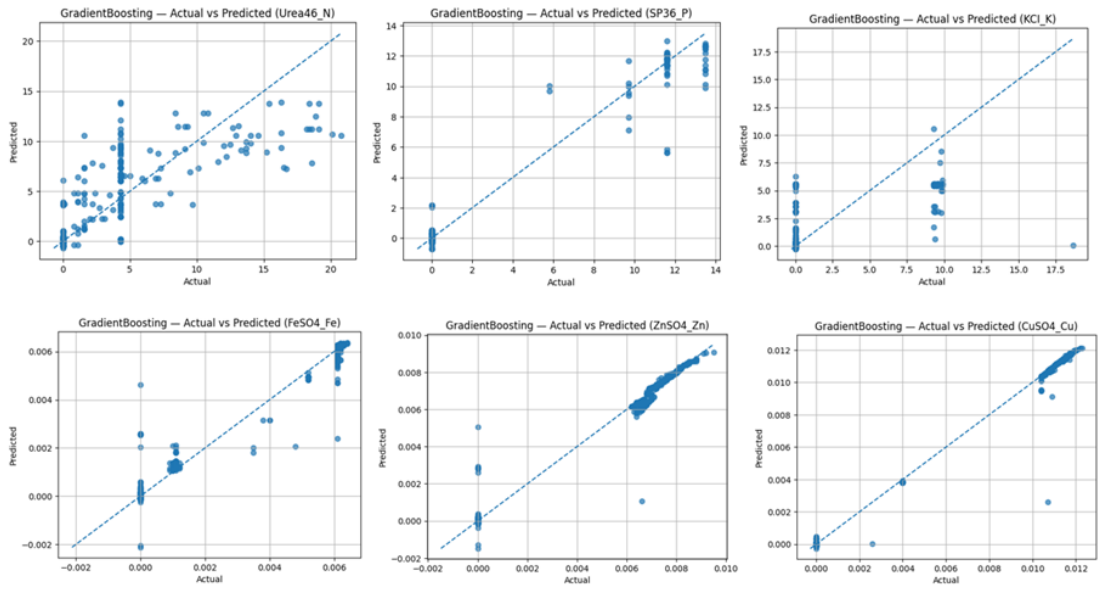

Figure 8. Visualization of predicted vs actual values on regression labels

Table 9 Present the results of the model testing against new input data (VA = Actual Value, VP = Predicted Value) that show the behavior of the two-layer pipeline in generating fertilization recommendations based on nutrient adequacy status. The results of tests with external data showed that the model was able to follow the resulting label pattern, although some predictive value discrepancies were still found. Fertilization recommendations are closely related to the nutritional adequacy status produced by the classification model. Fertilization recommendations are only given at the condition of nutritional adequacy status (0 = necessary), while at the condition (1 = sufficient) there are no fertilization recommendations that need to be done. Based on the test results in Table 10, there are still several errors in the prediction value. In this condition, the auto-decision mechanism acts as a consistency rule in a two-layer pipeline, where the results of fertilization recommendations are readjusted based on the status of nutrient adequacy classification. This aims to maintain consistency between the results of the classification and the fertilization recommendations. However, this mechanism also has limitations and has the potential to reduce the visibility of prediction errors in fertilization recommendation models. This indicates that the reported outcomes should

be interpreted as the behavior of an integrated decision-support system rather than a purely predictive regression model.

Table 9. Results of predicting the status of nutritional needs (actual and predicted) and recommended fertilization dosage results from combining classification and regression models

Sample	Act-N	Pred-N	Urea 46%	Act-P	Pred-P	SP36	Act-K	Pred-K	KCl
Temp = 27, Moist = 44.47, ph = 5.31, conductivity = 1.08, nitrogen = 2.66, *phosporus* = 0.37, *kalium* = 2.48	Va = 0	Vp = 0	7.132	Va = 0	Vp = 0	12.387	Va = 1	Vp = 1	0 (sufficient)
Temp = 31.61, Moist = 25.58, ph = 6.24, conductivity = 0.91, nitrogen = 2.06, *phosporus* = 1.7, *kalium* = 2.72	Va = 1	Vp = 1	0 (sufficient)	Va = 0	Vp = 0	12.001	Va = 1	Vp = 1	0 (sufficient)
Temp = 29.86, Moist = 31.69, ph = 5.09, conductivity = 1.74, nitrogen = 1.33, *phosporus* = 0.99, *kalium* = 1.16	Va = 0	Vp = 0	3.482	Va = 0	Vp = 0	10.550	Va = 1	Vp = 1	0 (sufficient)
Temp = 28.79, Moist = 34.65, ph = 7.27, conductivity = 1.04, nitrogen = 0.66, *phosporus*	Va = 1	Vp = 1	0 (sufficient)	Va = 0	Vp = 0	0.7806	Va = 1	Vp = 1	0 (sufficient)

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Sample	Act-N	Pred-N	Urea 46%	Act-P	Pred-P	SP36	Act-K	Pred-K	KCl
= 1.42, *kalium* = 0.6									
Sample	**Act-Fe**	**Pred-Fe**	**$FeSO_4$**	**Act-Zn**	**Pred-Zn**	**$ZnSO_4$**	**Act-Cu**	**Pred-Cu**	**$CuSO_4$**
Temp = 27, Moist = 44.47, ph = 5.31, conductivity = 1.08, nitrogen = 2.66, *phosporus* = 0.37, *kalium* = 2.48	Va = 1	Vp = 1	0 (sufficient)	Va = 1	Vp = 1	0 (sufficient)	Va = 1	Vp = 1	0 (sufficient)
Temp = 26.43, Moist = 32.18, ph = 5.22, conductivity = 1.66, nitrogen = 1.81, *phosporus* = 2.07, *kalium* = 1.7	Va = 0	Vp = 0	0.0032	Va = 0	Vp = 1	0 (detect sufficient)	Va = 0	Vp = 1	0.00145 (auto decision)
Temp = 27.46, Moist = 47.37, ph = 5.11, conductivity = 1.04, nitrogen = 0.56, *phosporus* = 2.26, *kalium* = 1.28	Va = 0	Vp = 1	0.0032 (auto decision)	Va = 0	Vp = 1	0 (detect sufficient)	Va = 0	Vp = 1	0.00145 (auto decision)

3.4 Discussion

This section presents an in-depth discussion related to research results, implications, scientific contributions and the usefulness of research results for related parties.

1) Sustainability Implications and Projections

This research succeeded in developing a machine learning-based decision support model using a predictive empirical approach. The datasets used in the machine learning process

consist of real-time data collected from a single demonstration field. The resulting parameters include environmental variables used to indicate the availability of macronutrients and micronutrients in the soil. Feature extraction is performed to create target labels and micronutrient index parameters using a rule-based approach based on predefined thresholds. This threshold is applied to the target labels of the classification and regression models, which are derived based on agronomic relationships [30], [31]. Although the model relies on empirical formulations, it does not aim to replace rule-based systems; instead, it incorporates empirical rules within a hybrid decision-support framework. The machine learning process employs proxy-based input features to limit direct dependencies between variables used in label construction and predictor variables. This design allows the model to study the relationships between environmental parameters in a more general way, similar to other machine learning applications that model complex prediction problems in different domains [42], [43]. However, it must be recognized that model performance can still be influenced by the empirical formulations used in label construction, so the high evaluation results cannot yet be fully interpreted as real agronomic accuracy. In addition, there is a possibility that the dependency of the model's learning results may occur, where the model has the potential to represent part of the rule structure used in label formation, so that the performance obtained needs to be adjusted to the proxy label, rather than as a direct representation of field conditions. The selection process of machine learning algorithms was carried out by performance comparison to select the best algorithm for the two-layer pipeline in this study. The results show that The Gradient Boosting model shows the best performance of all models across evaluation metrics. Therefore, it was chosen to predict the status of nutrient adequacy and fertilizer recommendations for sugarcane. Gradient Boosting has a strong ability to handle non-linear relationships, combine multiple weak learners, and manage complex feature interactions, making it suitable for the characteristics of datasets on those that tend to be medium and heterogeneous [44], [45]. The dataset used in this study consisted of real-time sensor readings, which tended to be heterogeneous. In addition, the input variables used in the modeling process include environmental parameters (pH, humidity, temperature, and conductivity) that have a complex relationship with soil nutrient conditions. Gradient Boosting ensemble learning mechanism gradually minimizes prediction errors, thus producing relatively stable predictions in the context of the data used. The results of the classification and regression model training using the cross-validation (CV) scheme in Table 7 show stable

performance based on the evaluation metrics. Overall, the classification model achieved a recall value above 0.97, while precision, F1 score, and accuracy showed relatively stable values in identifying nutritional adequacy status. The regression model also showed stable performance with relatively low RMSE and MAE values. However, with the exception of the potassium regression (KCl_K) model, which shows that the test evaluation value is lower than the performance in the training process. This can be influenced by the narrower distribution of potassium data and the higher variability in the target values shown in figure 9. This condition makes the relationship between environmental parameters and potassium content more difficult to accurately model.

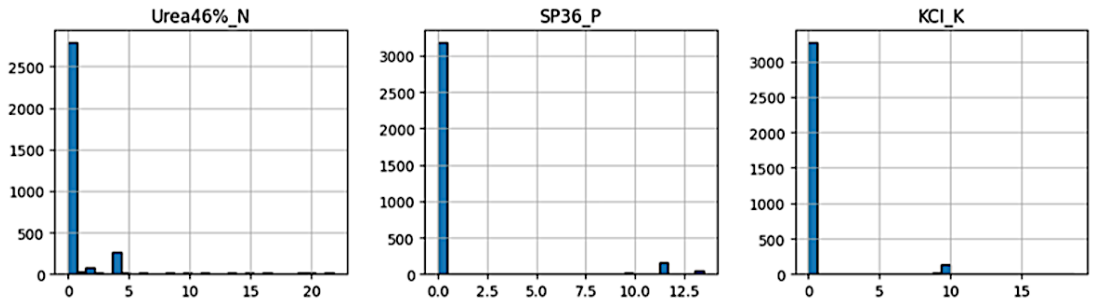

Figure 9. Data distribution conditions on macronutrient regression labels

The results of the evaluation of the micronutrient model show relatively high performance, which can be influenced by heterogeneous data sets and lower variability of target values. In general, micronutrient datasets tend to have lower variability compared to macronutrients. The difference in evaluation results between the training process (Table 7) and the holdout testing (Table 8) was relatively small, suggesting that the model showed no significant indication of overfitting in the observed data. Table 9 presents the test results using a new (external) dataset on a two-layer model. The combined prediction results of the classification and regression models are able to reflect the pattern of nutritional needs according to the labels used. Some prediction errors in macronutrients and micronutrients can be addressed through auto-decision mechanisms in classification and regression models in a two-layer pipeline, thus helping to reduce the potential for fertilization recommendation errors.

The proposed two-layer approach contributes to the development of a sensor data-driven machine learning approach by integrating classification and regression models

into a single pipeline. In addition, the classification model provides clear indicators that consist of two classes, namely "0 = need" and "1 = sufficient". This representation simplifies the interpretation of nutrient adequacy when the value of environmental data is obtained from soil sensor readings, as presented in this study. The fertilizer recommendations produced by the model are adjusted according to the fertilization projections shown in Table 2. This adjustment is based on a relatively limited demonstration area, so the estimated fertilizer dose is relatively lower. The two-layer model is also able to detect inconsistencies in prediction results with an auto-decision mechanism, thus allowing consistency between the classification results and fertilization recommendations as shown in Table 9.

The proposed two-layer model uses relatively complex input features and represents the relationships between environmental parameters in the context of the data used. Unlike rule-based systems that apply deterministic decisions directly based on threshold values, the two-layer approach allows the model to represent a more flexible relationship between environmental parameters, so that the resulting recommendations can adjust to the variations in sensor data. However, this model still has limitations because the datasets used in this study have a limited temporal range and spatial coverage, resulting in a short-term representation of data. Further validation over a longer timeframe and testing in larger agricultural areas is needed to support the long-term sustainability of the model. The findings of this study also show the potential for further development through integration with IoT devices, dashboard systems, and supporting applications, thus allowing fertilization recommendation information to be delivered in a more informative and user-friendly manner. The practical implementation of this study shows early potential as a sensor-based decision-support system in plant nutrition management, but real implementation requires further validation before it is widely implemented.

2) Scientific Contributions

The proposed model in this study makes a methodological contribution to the application of machine learning for nutrient monitoring, especially in sugarcane plants based on real-time data. Data preprocessing using agronomic empirical formulation provides a representation of the availability of nutrients in the demonstration land. The integration of classification and regression models in macronutrient and micronutrient adequacy

monitoring can bridge gaps in previous studies that generally focused on single nutrient monitoring. The two-layered approach proposed in this study combines machine learning (Gradient Boosting) with the construction of index and label features based on empirical rules, thus allowing the model to represent a non-linear relationship between environmental parameters and soil nutrient conditions. This approach provides an initial framework for the development of a fertilization recommendation system tailored to environmental conditions through direct data collection on demonstration land.

4. CONCLUSION

The results of this study show that a two-layer pipeline that integrates classification and regression models using Gradient Boosting is able to reproduce patterns of nutrient adequacy status and fertilization recommendations based on sensor data on demonstration land conditions. This approach demonstrates the model's ability to reproduce the pattern of relationships between macronutrients and micronutrients in a single pipeline. However, this study has several limitations, including limited spatial coverage, relatively short data collection period, and label construction based on empirical formulations. Therefore, further validation and generalization are needed on various environmental conditions and time periods. Wider land cover and a more comprehensive agronomic evaluation are still needed to assess the consistency and reliability of the model.

ACKNOWLEDGMENT

The authors would like to acknowledge the Research, Community Service, and Collaboration Unit of Telkom University for the financial support and facilitation provided during this research.

REFERENCES

[1] H. F. Rochmah, R. P. Mumpuni, and D. E. Ramadhani, "Integrasi Pembibitan Tanaman Tebu Tunas Tunggal dan Budi Daya Ikan Lele (Integrated Farming of Single Bud Sugar Cane and Catfish)," *J. Pertan. Indones.*, vol. 26, no. 4, pp. 591–596, 2021, doi: 10.18343/jipi.26.4.591.

[2] L. Syed, "Smart Agriculture using Ensemble Machine Learning Techniques in IoT Environment," *Procedia Comput. Sci.*, vol. 235, pp. 2269–2278, 2024, doi: 10.1016/j.procs.2024.04.215.

[3] R. Rosyadi, S. Darma, D. C. Darma, and M. Ló, "What Driving Gross Domestic Product of Agriculture? Lessons from Indonesia (2014-2021)," *Int. J. Sustain. Dev. Plan.*, vol. 18, no. 3, pp. 683–692, 2023, doi: 10.18280/ijsdp.180304.

[4] N. Dyah, D. Suprayitno, A. Zairina, and Y. Quarta, "Spatial Analysis of Sugarcane Production to Support National Sugar Self-Sufficiency: A Case Study of East Java Province," *J. Agric. Socioecon. Bus.*, vol. 8, no. 2, pp. 180–190, 2025.

[5] D. B. Santoso, E. Y. Arvianti, and B. Santosa, "Analysis of The Efficiency of The Use of Production Factors And Income in Sugarcane Farming (Case Study of Independent Sugarcane Farming in Kedungjajang Sub-District, Lumajang District)," *Int. J. Technol. Agric. Nat.*, vol. 2, no. 2, pp. 71–84, 2025, doi: 10.64515/irdhijtans.v2i2.806.

[6] H. Takaragawa and M. Wakayama, "Responses of leaf gas exchange and metabolites to drought stress in different organs of sugarcane and its closely related species Erianthus arundinaceus," *Planta*, vol. 260, Art. no. 90, 2024, doi: 10.1007/s00425-024-04508-w.

[7] R. Kumar *et al.*, "Drought and salinity stresses induced physio-biochemical changes in sugarcane : an overview of tolerance mechanism and mitigating approaches," *Front. Plant Sci.*, vol. 14, Art. no. 1225234, 2023, doi: 10.3389/fpls.2023.1225234.

[8] F. Mehdi, X. Liu, Z. Riaz, and U. Javed, "Expression of sucrose metabolizing enzymes in different sugarcane varieties under progressive heat stress," *Front. Plant Sci.*, vol. 14, Art. no.1269521, 2023, doi: 10.3389/fpls.2023.1269521.

[9] G. N. Ahmed and S. Kamalakkannan, "Micronutrient Classification in IoT based Agriculture using Machine Learning (ML) Algorithm," *Proc. 4th Int. Conf. Smart Syst. Inven. Technol.*, pp. 1–9, 2022, doi: 10.1109/ICSSIT53264.2022.9716293.

[10] M. Manzoor *et al.*, "Ecological Zones and Irrigation Sources Amidst Environmental Stressors," *Plants*, vol. 12, no. 20, Art. no. 3526, 2023, doi: 10.3390/plants12203526.

[11] Q. De Castro and A. P. Coelho, "Fertilizer source and application method influence sugarcane production and nutritional status," *Front. Plant Sci.*, vol. 14, Art. no. 1099589, 2023, doi: 10.3389/fpls.2023.1099589.

[12] F. Mehdi *et al.*, "Factors affecting the production of sugarcane yield and sucrose accumulation: suggested potential biological solutions," *Front. Plant Sci.*, vol. 15, Art. no. 1374228, 2024, doi: 10.3389/fpls.2024.1374228.

[13] M. O. A. Galal and A. R. A. Moursy, "Impact of Micronutrients Spraying on Yield and Quality of Sugar Cane in heavy textured soil at Upper Egypt," *J. Sohag Agriscience*, vol. 9, no. 1, pp. 36–54, 2024, doi: 10.21608/jsasj.2024.352295.

[14] D. H. Sary and M. E. A. El-aziz, "Synthesis and characterization of nano-micronutrient fertilizer and its effect on nutrient availability and maize (Zea Mays L.) productivity in calcareous soils," *Sci. Rep.*, vol. 15, Art. no. 11273, 2025, doi: 10.1038/s41598-025-11273-7.

[15] B. A. Jyothi, T. Sujatha, and K. S. C. Bose, "Assessment of Optimum Dose and Source of Sulphur for Getting Improved Yield, Quality and Nutrient Uptake in Sugarcane," *Biol. Forum – An Int. J.*, vol. 13, no. 2, pp. 633–636, 2021.

[16] P. Justice, "Drought stress tolerance mechanisms and breeding effort in sugarcane : A review of progress and constraints in South Africa," *Plant Stress*, vol. 2, Art. no. 100027, 2021, doi: 10.1016/j.stress.2021.100027.

[17] A. A. Muliasari and R. Trilaksono, "Insidensi Hama dan Penyakit Utama Tebu (Saccharum officinarum L) di PT PG Rajawali II Jatitujuh Majalengka," *J. Sains Terap.*, vol. 10, no. 1, pp. 40–52, 2020, doi: 10.29244/jst.10.1.40.

[18] H. Martín-Cardoso, "Impact of Nutrient Stress on Plant Disease Resistance," *Int. J. Mol. Sci.*, vol. 26, no. 4, Art. no. 1780, 2025, doi: 10.3390/ijms26041780.

[19] J. Padmapriya and T. Sasilatha, "Engineering Applications of Artificial Intelligence Deep learning based multi-labelled soil classification and empirical estimation toward sustainable agriculture," *Eng. Appl. Artif. Intell.*, vol. 119, Art. no. 105690, 2023, doi: 10.1016/j.engappai.2022.105690.

[20] C. Schröer, F. Kruse, and J. Marx, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[21] P. P. Ruwanpathirana *et al.*, "Evaluation of Sugarcane Crop Growth Monitoring Using Vegetation Indices Derived from RGB-Based UAV Images and Machine Learning Models," *Agronomy*, vol. 14, no. 9, Art. no. 2059, 2024, doi: 10.3390/agronomy14092059.

[22] J. Sun, Z. Li, F. Li, Y. Shen, Y. Qian, and T. Li, "EF yolov8s: A Human-Computer Collaborative Sugarcane Disease Detection Model in Complex Environment," *Agronomy*, vol. 14, no. 9, Art. no. 2099, 2024, doi: 10.3390/agronomy14092099.

[23] B. Cai, F. Shi, B. A. Geremew, A. K. Addis, M. C. Abate, and W. Dessie, "Precision agriculture techniques for optimizing chemical fertilizer use and environmental sustainability: a systematic review," *Front. Agron.*, vol. 7, Art. no. 1665444, 2025, doi: 10.3389/fagro.2025.1665444.

[24] J. Botero-Valencia *et al.*, "Machine Learning in Sustainable Agriculture : Systematic Review and Research Perspectives," *Agriculture*, vol. 15, no. 4, Art. no. 4377, 2025, doi: 10.3390/agriculture15040377.

[25] C. Musanase, A. Vodacek, D. Hanyurwimfura, and A. Uwitonze, "Data-Driven Analysis and Machine Learning-Based Crop and Fertilizer Recommendation System for Revolutionizing Farming Practices," *Agriculture*, vol. 13, no. 11, Art. no. 2141, 2023, doi: 10.3390/agriculture13112141.

[26] J. Gao *et al.*, "A Fertilization Decision Model for Maize, Rice, and Soybean Based on Machine Learning and Swarm Intelligent Search Algorithms," *Agronomy*, vol. 13, no. 5, Art. no. 1400, 2023, doi: 10.3390/agronomy13051400.

[27] S. Rani and A. K. Mishra, "Machine learning - based optimal crop selection system in smart agriculture," *Sci. Rep.*, vol. 13, Art. no. 15997, 2023, doi: 10.1038/s41598-023-42356-y.

[28] R. Islam, K. Oliullah, M. Kabir, M. Alom, and M. F. Mridha, "Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation," *J. Agric. Food Res.*, vol. 14, Art. no. 100880, 2023, doi: 10.1016/j.jafr.2023.100880.

[29] R. F. Asto Bonifacio, "An Intelligent Fertilizer Dosing System Using a Random Forest Model for Precision Agriculture," *J. Robot. Control*, vol. 6, no. 3, pp. 1197–1207, 2025, doi: 10.18196/jrc.v6i3.26421.

[30] Evianti *et al.*, *Analisis Kimia Tanah, Tanaman Air dan Pupuk*, 3rd ed. Purwokerto: Kementerian Pertanian Republik Indonesia, 2023.

[31] B. Vajantha, K. V Nagamadhuri, P. Prasunarani, and T. M. Hemalatha, "Mapping of DTPA Extractable Micronutrients in Sugarcane Soils of Sudhalagunta Sugar Factory, Andhra Pradesh, India – A GIS Approach," *Int. J. Plant Soil Sci.*, vol. 34, no. 21, pp. 755–762, 2022, doi: 10.9734/IJPSS/2022/v34i2131329.

[32] Y. Li *et al.*, "Delaying First Fertilization Time Improved Yield and N Utilization Efficiency on Direct Seeding Rice," *J. Plant Growth Regul.*, vol. 43, no. 2, pp. 423–433, 2024, doi: 10.1007/s00344-023-11091-z.

[33] H. W. Atmojo, Machmudi, F. Nursadi, and A. R. Puspitasari, "Pengaruh Pemupukan Anorganik pada Budidaya Tanaman Tebu (Saccharum officinarum L.) Varietas PSKA 942 di Pusat Penelitian Perkebunan Gula Indonesia," *Indones. Sugar Res. J.*, vol. 4, no. 2, pp. 13–23, 2024, doi: 10.54256/isrj.v4i1.120.

[34] M. H. Siddiqui, "Balanced Use of Zn, Cu, Fe, and B Improves the Yield and Sucrose Contents of Sugarcane Juice Cultivated in Sandy Clay Loam Soil," *Agronomy*, vol. 12, no. 3, Art. no. 696, 2022, doi: 10.3390/agronomy12030696.

[35] V. Ivezić, Å. R. Almås, B. R. Singh, and Z. Lončarić, "Prediction of trace metal concentrations (Cd, Cu, Fe, Mn and Zn) in wheat grain from unpolluted agricultural soils," *Acta Agric. Scand.*, vol. 63, no. 4, pp. 360–369, 2013, doi: 10.1080/09064710.2013.785586.

[36] K. M. Magomere, S. Nchimbi-msolla, and G. M. Tryphone, "Effects of soil pH levels on iron and zinc concentrations of common bean (Phaseolus vulgaris L.) genotypes," *African J. Agric. Res.*, vol. 20, pp. 510–516, 2024, doi: 10.5897/AJAR2022.15924.

[37] V. Kuzmanovski and D. E. Larsen, "Optimization of arable land use towards meat-free and climate-smart agriculture : A case study in food self-sufficiency of Vietnam," *Proc. 2019 IEEE Int. Conf. Big Data (Big Data)*, pp. 5140–5148, 2019, doi: 10.1109/BigData47090.2019.9006264.

[38] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, Art. no. 56706, 2024, doi: 10.1038/s41598-024-56706-x.

[39] S. Sathyanarayanan and B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, vol. 27, no. 4S, pp. 4023–4031, 2024, doi: 10.53555/AJBR.v27i4S.4345.

[40] T. O. Hodson, "Root-Mean-Square Error (RMSE) Or Mean Absolute Error (MAE): When To Use Them Or Not," *Geosci. Model Dev.*, vol. 15, no. 2, pp. 5481–5487, 2022, doi: 10.5194/gmd-15-5481-2022.

[41] D. Chicco, M. J. Warrens, and G. Jurman, "The Coefficient Of Determination R-Squared Is More Informative Than SMAPE, MAE, MAPE, MSE And RMSE In Regression Analysis Evaluation," *PeerJ Comput. Sci.*, vol. 7, Art. no. 623, 2021, doi: 10.7717/peerj-cs.623.

[42] L. Yang *et al.*, "Radar Composite Reflectivity Reconstruction Based on FY-4A Using Deep Learning," *Sensors*, vol. 23, no. 1, Art. no. 81, 2023, doi: 10.3390/s23010081.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[43] X. Yang, Y. Li, X. Yu, H. Tan, J. Yuan, and M. Zhu, "Regional/Single Station Zenith Tropospheric Delay Combination Prediction Model Based on Radial Basis Function Neural Network and Improved Long Short-Term Memory," *Atmosphere (Basel)*, vol. 14, no. 2, Art. no. 303, 2023, doi: 10.3390/atmos14020303.

[44] N. Orsini, A. Moore, and A. Wolk, "Interaction Analysis Based on Shapley Values and Extreme Gradient Boosting: A Realistic Simulation and Application to a Large Epidemiological Prospective Study," *Front. Nutr.*, vol. 9, Art. no. 871768, 2022, doi: 10.3389/fnut.2022.871768.

[45] A. D. Lainder and R. D. Wolfinger, "Forecasting With Gradient Boosted Trees: Augmentation, Tuning, And Cross-Validation Strategies: Winning Solution To The M5 Uncertainty Competition," *Int. J. Forecast.*, vol. 38, no. 4, pp. 1426–1433, 2022, doi: 10.1016/j.ijforecast.2021.12.003.