

Smart farming based on IoT to predict conditions using machine learning

Mochammad Haldi Widiyanto, Yovanka Davincy Setiawan, Bryan Ghilchrist, Gerry Giovan

Computer Science Department, School of Computer Science, Bina Nusantara University, Bandung Campus, Jakarta, Indonesia

Article Info

Article history:

Received May 29, 2023

Revised Jan 22, 2024

Accepted Mar 21, 2024

Keywords:

Extreme gradient boosting

Internet of things

Machine learning

Predict

Smart farming

ABSTRACT

Smart farming is a type of technology that utilizes the internet of things (IoT) to provide information on agricultural and environmental conditions as well as perform automation. Some of these ecological conditions can be used and analyzed in machine learning (ML) data management. This study focuses on utilizing ML algorithms to find the best prediction; typically used methods include linear regression, decision tree (DT), random forest (RF), and extreme gradient boosting (XGBoost). In the application of smart farming, research on IoT and artificial intelligence (AI) is still uncommon since most IoT cannot make predictions like AI. Because basically, some IoT can't make predictions as AI does. In this Study, predictions were made by looking at the regression results in the form of root mean square error (RMSE) and absolute error. The results show a strong and weak correlation between features (positive or negative). The best prediction results are obtained by XGBoost when predicting temperature (RMSE 6.656 and absolute error 3.948) and (soil moisture 17.151 and absolute error 11.269). However, using different parameters (RMSE RF and absolute error DT) on RF and DT resulted in good and distinct results. Linear regression, on the other hand, produced unsatisfactory and poor result.

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



Corresponding Author:

Mochammad Haldi Widiyanto

Computer Science Department, School of Computer Science, Bina Nusantara University

Bandung Campus, Jakarta, Indonesia

Email: mochamad.widiyanto@binus.ac.id

1. INTRODUCTION

Soil is vital for plant life, but it may also be used by many other things, including people. In agriculture, the soil can be observed through various parameters, including moisture, pH, nutrients, and mineral content. Many signs could be discovered by focusing on these metrics, particularly soil moisture. For example, the health of the forest, how it might be damaged if a forest fire occurs, and how insects and other parasitic organisms are affected. These indicators prompted the necessity to monitor soil moisture measurement conditions, which are extensive and well-organized around the world [1].

The area in Indonesia is located at the equator, so it only has two seasons. Many mountain ranges enable the establishment of numerous plant species, which have a considerable impact on soil levels, particularly moisture, nutrients, temperature, and pH. To achieve the best results, these factors have a large influence on how the plant develops [2]. Agriculture is an extremely advanced and developed industry since it is inextricably linked to and influences the food industry. This, combined with the fact that soil is a vital component of agriculture, caused soil content studies to become more widespread, particularly in agricultural sectors. Because the Lembang area in Indonesia is mostly used for various plantation and agricultural activities, the author conducted smart farming study in this area [3].

The world of agriculture is in dire need of technology, especially research that utilizes the internet of things (IoT) [4], [5]. IoT contains sensing, and some IoT tools are helpful in getting information about soil, humidity, temperature, and pH. This is very helpful for farmers and workers in monitoring, automation, and recommendations. The utilization of technology that uses IoT is also known as smart farming. IoT-based smart farming has lately gained popularity because it can automatically monitor and maintain the agricultural sector by involving humans as objects, rather than subjects [6]. Not only that, but smart farming can also be combined with artificial intelligence (AI) technologies to increase maximum results [7]. Despite their numerous advantages, IoT tools [8] remain difficult to implement for rural farmers.

The results of IoT [9] sensing devices are raw data that can be processed to become a recommendation or even forecast data for the future of soil content. Machine learning (ML), which is one of the derivatives of AI, can help and even improve the quality of the harvest [10]. ML has many things that decrease human involvement or increase outcomes [10]. Often used methods are random forest (RF), linear regression, or extreme gradient boost (XGBoost). In addition, there is also the use of using deep learning (DL). The fundamental difference is that ML requires data to perform classification, while DL does not need it because it will do the clustering itself.

Abioye *et al.* [11] researching fresh water that affects the supply of nutrients and irrigation where plant growth is needed because it is used when there is a lack of rainfall. According to studies, plant activities require roughly 70% of available water; thus, responsible water consumption management is necessary. This Study investigates integrating different machine learning models (ML) that can provide optimal irrigation management decisions. Dubois *et al.* [7] makes agricultural decisions because it is an essential component in seeing the results in the future. In the science and context of intelligent agriculture, farmers need data from sensing devices embedded in crops, leveraging agronomic models to help. The research focuses on demonstrating the relationship between ML in solving problems as explained previously is because this method can maximize predictions accurately.

Rahman *et al.* [12] in his research on statistics, agriculture makes a significant contribution to mushroom farming in the market. Therefore, the popularity of mushroom cultivation is needed. Farmers, especially in remote areas, typically still employ traditional methods to monitor crucial factors in fungal growth, such as temperature, humidity, and pH conditions. As a result, the focus of this research is on using ML and IoT architecture to construct smart mushroom farming with exceptional results. A study conducted trials on ML technology has been adopted to classify fungi using ML models such as linear regression (LR), decision tree (DT), k-nearest neighbour (KNN), naïve bayes (NB), support vector machine (SVM), and RF. The highest accuracy gained with the ensemble model is 100%. Widiyanto *et al.* [5] is a previous study that is the basis of this study. In a previous study, the author conducted a survey to collect data in mountainous areas. The research results focus on generating data utilizing IoT tools. Next, the root mean squared error (RMSE) error measurement was carried out by comparing the results from IoT with the actual value, but not yet utilizing ML models. According to several studies, few have applied original data from Indonesia's unique regions, especially West Java. Because the nature of the data from temperature, pH, and humidity varies from country to country, by using ML, the author can forecast some of these features to help farmers at the forefront.

This research contributes to a comparative model of several ML methods that can be assessed on the RMSE results and absolute error, to search for the best results in soil condition forecasting for farmers. In this study, several algorithms will be used to perform comparisons, such as DT [13], [14], RF [15], [16], LR [17], [18], and XGBoost [19], [20]. By using this algorithm, it can be seen which performance produces the best predictions. It is hoped that rural farmers can use it with data taken from IoT devices on a secondary basis (data retrieval has been carried out for several months). After understanding the background of why ML is needed in forecasting, the next chapter will discuss theory (chapter 2), system design (chapter 3), results (chapter 4), and conclusions (chapter 5). It is hoped that this research can be used for further research or other industries.

2. RESEARCH METHOD

2.1. Internet of things

This technology is a system for connecting computers digital and mechanical devices, which connects subjects, objects, and even liaisons between individuals with a unique design for sending data and can click human-to-human even on computer-to-human. The connection between the internet is that things in IoT are like connecting with humans and computers. Many sectors have utilized IoT in daily life by proliferating intelligent applications and services that use AI. The application of AI techniques requires centralized data processing and collection. It allows it to be carried out realistically on any application scheme due to the highly scalable nature of IoT on the network [21]. In this study, IoT is used in retrieving data that is processed and retrieved in real-time in the Indonesian West Java Region.

2.2. Smart farming

Smart farming stems from the idea that food shortages and rapid population growth are major global obstacles. Advanced technologies such as IoT and mobile internet usually support smart farming. IoT is one of the important components in smart systems because it communicates between devices and sensors in carrying out fundamental tasks. Smart farming technology can be used for important functions such as seeding, harvesting, irrigation, weed detection, livestock applications, and pest spraying. Through the help of technologies such as IoT, big data (BD), AI, ML [22], and DL [23].

All of the previous significantly impact smart farming as it can deliver the entire supply chain, especially in producing essential crops such as in (for Indonesians). All components are considered in increasing the variety and amount of data captured in the IoT, and the results of the collected data greatly affect the modelling process's performance of the ML algorithm [24]. The system can see the flow between software and hardware components [25]. This technology has also become one of the parameters of the success of developed countries in developing food security. According to the author, using ML is more effective if it already has secondary data than ML is used in this research.

2.3. Machine learning

AI has derivatives of techniques that can be applied to computers to do the same thing as human behaviour or in human decision-making to complete complex tasks independently or with little human involvement [26]. Therefore, this relates to various other problems because intelligence requires reasoning, knowledge representation, planning, learning, communication, and perception to refer to different methods and tools [27]. However, the scheme has faced several obstacles due to the unique nature of humans, who always struggle to explain all knowledge in a complex manner [28].

ML, on the other hand, can overcome these obstacles; ML can improve program performance by taking prior experience and performance measures [29]. Therefore, ML can automate the task of building analytical models that are cognitive in nature in performing language or object detection because ML can implement programs that can learn from training data. ML can be applied well, especially when the task is related to data with many features such as regression, classification, and clustering. By learning from previous experiences, ML can help produce reliable and repeatable decisions [30]. This study will use several research algorithms using ML, such as RF, LR, DT, and XGBoost.

2.3.1. Linear regression

There are many regression models. This analysis is useful in estimating the variable's value as the dependent example 'y' with its effect on the independent variable 'x' [31]. However, this study only focuses on linear regression. This algorithm is a model with the condition that the variable must be single-independent. Linear regression has the (1):

$$Y = a + bX \quad (1)$$

where Y =the dependent variable; X =the explanatory variable; a =the intercept; b =the slope of the line.

The (1) is a simple formula for performing LR. This algorithm can distinguish the effect between these variables. However, this algorithm is only used as a simple predictive measurement, so the results are unlikely to be good for diverse data [32].

2.3.2. Random forest

RF [15] algorithm has a tree for making decisions that can be interpreted with a parametric model. Done to integrate DT analysis, prediction models like this can be said to be more comprehensive to conclude. RF regression is a non-parametric regression algorithm derived from a tree.

2.3.3. Decision tree

This algorithm is one of the ML that is often used because it is a popular classifier. Because this algorithm model is easy to explain, one of them can perform very satisfactorily. DT is widely used because of the increasing need to use ML models. This model also has many derivatives, so many say that DT is one of the bases of several models [33].

2.3.4. XGBoost

This algorithm, usually called XGBoost is a boost in the decision method [34]. This algorithm is an implementation of the gradient adder engine (GBM). This algorithm can be used for several classification and regression problems. Data researchers very much need this algorithm because it has a very high computational speed when viewed in core computing [35].

2.3.5. Matrix correlation

There are numerous matrices to observe; however, in this study, the author focuses on the correlation matrix, where a matrix has a correlation coefficient with values located in the interval [-1,1]. A correlation coefficient is part of a value to see how closely the relationship between variables is with other variables. The set of coefficients is presented in a correlation matrix [36]. The correlation formula itself is found in (2) [37]:

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} \quad (2)$$

which r =correlation coefficient; x =data x ; \bar{x} =data average x ; y =data y ; \bar{y} =data average y .

The (2) for each correlation between variables will be mapped in a heat map to show the relationship's size. Correlation analysis is usually used in statistical measures that can be used in depth to see different study situations from an efficient identification of relationships between other attributes of a dataset obtained from IoT tools (see Figure 1) [38].

Data has a positive or strong positive correlation if it continuously increases in the positive direction and vice versa for negative and strongly negative correlations. On the contrary, if the data is always random, it will be said to be uncorrelated. However, if the correlation results form a hill, it can be said to have a non-linear correlation.

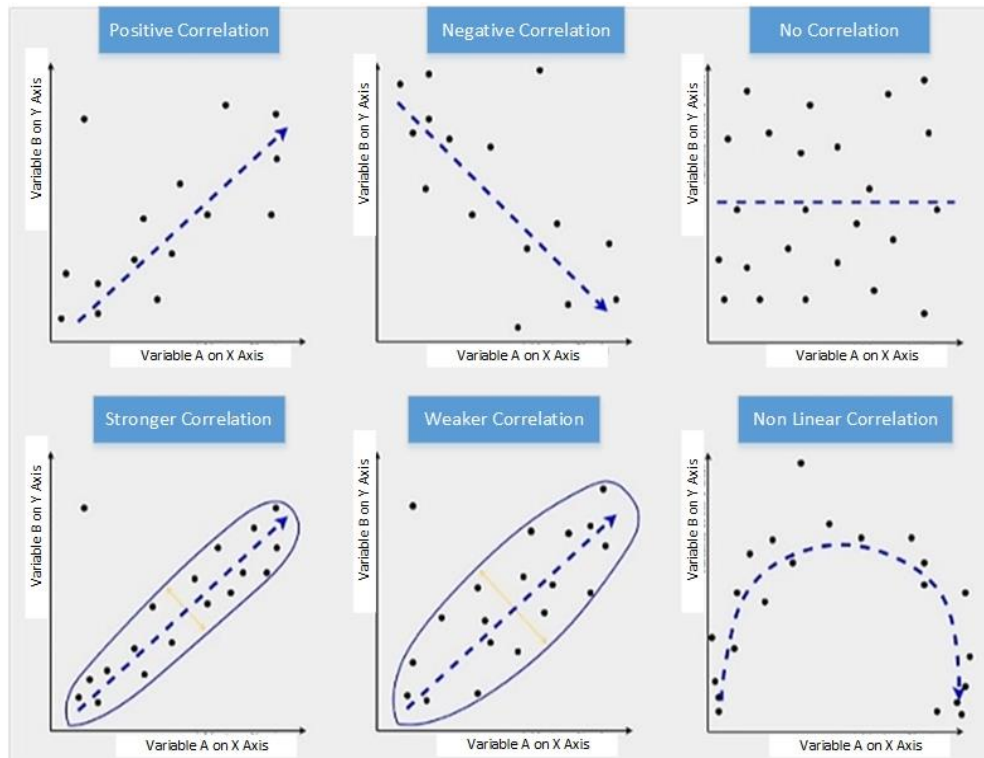


Figure 1. Type of correlation [38]

2.3.6. Performance

The regression results usually used several approaches, and this study's authors have several approaches. Uncertainty is used by the method or observation is used to see the results of the comparison between observers and the model, so the RMSE approach is applied [39] and absolute error using (3) and (4):

$$\sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (3)$$

$$(\Delta x) = |x_i - x| \quad (4)$$

in (3) and (4) is one way to find the performance of the regression where n is the data, and i is the amount of data available. After studying the theory used in this research, the next chapter will explain the system's design to be formed.

3. SYSTEM DESIGN

In this section, the author will discuss several designs used in conducting this research. According to the author, these designs are essential in explaining to readers how this study works. Therefore, the author will explain how the research works through the results: i) data shape, ii) data correlation matrix, and iii) ML design.

3.1. Data shape

Data is retrieved using IoT devices. The author can create ML applications combined with IoT to make accurate predictions in predicting temperature and soil moisture. Analogous results were obtained through garden temperature, soil moisture, light resistance, and air humidity. An example of the data form is shown in Table 1 [5]: i) wemos D1 R2 (ESP8266), ii) capacitive soil moisture sensor, iii) light dependent resistor (LDR) photoresistor sensor, iv) temperature and humidity sensor (DHT22), v) modem Wi-Fi router, and vi) power supply unit 5V/10A (PSU). Table 1 shows the results obtained by utilizing IoT sensing devices. The data will be correlated, which will then be used to see the prediction performance of several ML with temperature and soil moisture predictions.

3.2. Matrix correlation and machine learning design

As previously explained, this matrix helps see the relationship between several features in the datasheet. For its use, it utilizes Rapidminer (student version). The design is divided into 2 parts: i) discussing correlation matrix design and ii) discussing ML design. The design is shown in Figure 2. In Figure 2(a) the data uses secondary data, which is processed by data normalization, then using (2), the correlation results are displayed. In the design of processing Table 1 data, several schemes are used, as shown in Figure 2(b). Thus, the author proposes the design with several parameters, such as: i) split data using automatic sampling, ii) a regression label is placed on temperature and soil moisture, and iii) performance on RMSE and mean absolute error (MAE).

Table 1. Research datasheet [5]

Entry id	Temperature	Soil moisture (%)	Light intensity resistance (Ω)	Humidity (%)
1	34	100	1024	72
2	33	58	1024	63
3	33	57	1024	70
4	28	57	1024	63
5	27	56	1024	62
6	27	50	1024	61
7	27	44	1024	53
8	29	53	1024	53
9	34	51	1024	69
10	33	53	1024	51

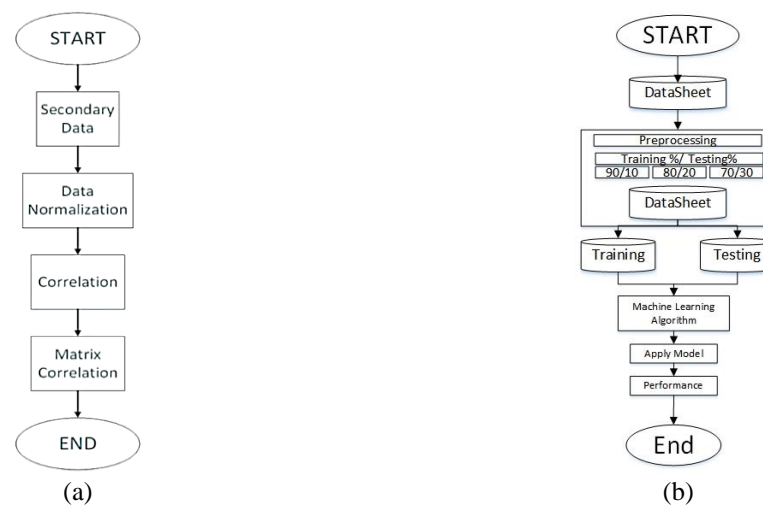


Figure 2. Flowchart design of (a) matrix correlation design and (b) ML design

4. RESULT

In this section, the author will discuss the results of this study. These results are a series of combined ML and IoT that make predictions on data sets. Those can be displayed in the points: i) matrix correlation result test, ii) regresion temperature test, and iii) regresi soil moisture test.

4.1. Matrix correlation result test

In this section, the results of data correlation will be shown. Data correlation shows the relationship between features with other features, and feature relationships can be strongly positive or strongly negative. The results are shown in Table 2, and the heatmap will be shown in Figure 3.

Table 2 shows that temperature has a strong negative correlation with soil moisture, then it is weakly negative on humidity, while intensity resistance has a weak positive correlation. Distinguishing strong and weak is seen on the heatmap in Figure 3. If the image gets darker blue it is negative, however a dark red image is positive. Table 2 also reveals that soil moisture has a strong negative relationship with temperature, a weak negative relationship with light intensity resistance, and a weak positive relationship with humidity, as illustrated in Figure 3. Therefore, it can be said that the correlation matrix test shows a correlation between features that are useful in determining the next performance.

Table 2. Matrix correlation result

Parameter	Temperature (°C)	Soil moisture (%)	Light intensity resistance (Ω)	Humidity (%)
Temperature (°C)	1	-0.43452	0.136054	-0.25063
Soil moisture (%)	-0.43452	1	-0.28637	0.31639
Light intensity resistance (Ω)	0.136054	-0.28637	1	-0.44699
Humidity (%)	-0.25063	0.31639	-0.44699	1

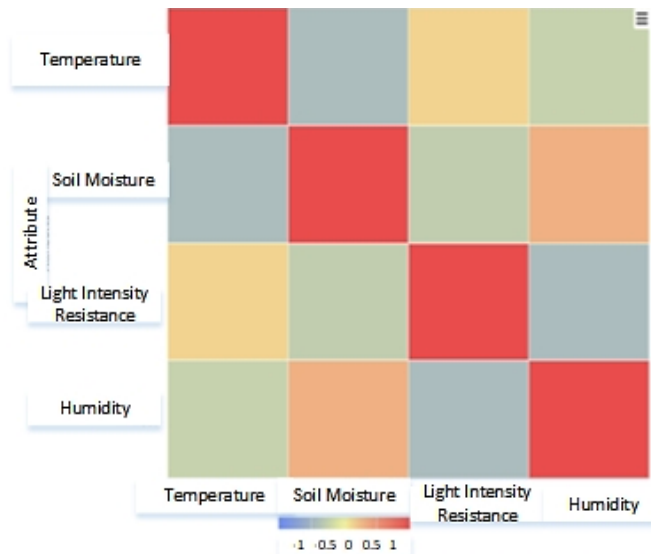


Figure 3. Heatmap correlation result

4.2. Temperature regresion test

In this section, the author will test the performance prediction on temperature data to measure performance based on several ML approaches and the changed parameters according to the amount of testing data and training data. Tables 3 and 4 will explain how the prediction results are divided into 2 parts:

- for temperature.
- for soil moisture prediction results.

Tables 3(a) and 4(a) show good performance results in the XGBoost algorithm with the best RMSE at 6.656 and absolute error at 3.948. This is very reasonable to guide because this algorithm is one of the best-boosting algorithms and shows if XGBoost can work in a state of correlation between data that not all features are strong. The results show that the RF outperforms the DT, with an RMSE of 7.013. The absolute error, however, is bigger than the DT. This indicates that the performance data for each algorithm has the opposite result, or it can be said that some algorithms are better at different performance approaches as well.

The worst algorithm for making predictions is linear regression with very large RMSE results and absolute error.

Tables 3(b) and 4(b) prove that the RMSE and absolute error results for each predicted feature show the opposite. The best algorithm is still XGBoost, with an RMSE of 17.151 and an absolute error of 11.269, far from the prediction for temperature. The uncertain nature of soil moisture data features and other correlation factors evidences this. The same thing happened to the RF with an RMSE of 17.209, which was better than the DT and had a higher absolute error than the DT. This still proves that the regression test can produce different results for the algorithm. Poor results are also shown in linear regression and other tests. This shows that LR is not suitable if used in predictions if the data is unpredictable or data does not have a robust correlation with other features.

Table 3. RMSE performance ML (a) temperature (amount of training data %/ total testing data %) and (b) soil moisture (amount of training data %/ total testing data %)

(a)				(b)			
Parameter	90%/10%	80%/20%	70%/30%	Parameter	90%/10%	80%/20%	70%/30%
LR (RMSE)	9.784	9.824	9.837	LR (RMSE)	19.210	19.456	19.654
DT (RMSE)	7.575	7.744	7.947	DT (RMSE)	18.374	18.584	19.383
RF (RMSE)	7.013	7.244	7.325	RF (RMSE)	17.209	17.345	17.940
XGBoost (RMSE)	6.656	6.765	6.889	XGBoost (RMSE)	17.151	17.334	17.993

Table 4. Absolute error performance performance ML (a) temperature (amount of training data %/ total testing data %) and (b) soil moisture (amount of training data %/ total testing data %)

(a)				(b)			
Parameter	90%/10%	80%/20%	70%/30%	Parameter	90%/10%	80%/20%	70%/30%
LR (absolute error)	8.008	8.017	8.095	LR (absolute error)	16.066	15.674	15.730
DT (absolute error)	4.235	4.405	5.370	DT (absolute error)	11.477	11.617	11.853
RF (absolute error)	5.057	5.133	5.382	RF (absolute error)	11.578	11.713	12.144
XGBoost (absolute error)	3.948	4.061	4.099	XGBoost (absolute error)	11.269	11.486	11.774

5. CONCLUSION

This work focuses on utilizing some ML in smart farming, and the resulting data in the form of temperature, soil moisture, light intensity resistance, and humidity. All features are generated from farm IoT devices. These features generated an abundance of data, which was then predicted using AI, specifically the AI branch known as ML. Several ML algorithms help prediction, such as linear regression, DT, RF, and XGBoost. What is tested in this work is the correlation between features in determining feature relationships and prediction tests in the form of RMSE and absolute error. The results show that XGBoost is very good at making predictions on this work with the temperature feature, the RMSE is 6.656, and the absolute error is 3.498. There is a uniqueness when comparing RMSE, and absolute error in RF and DT, where the RF is better when testing RMSE and the DT is better when trying absolute error. In the second test, when the prediction is placed on the soil moisture feature, the XGBoost algorithm is still better, with only the value of RMSE and absolute error being more significant. This is due to the nature and type of data on various soil moisture features. The last result also shows that linear regression is the worst in both tests. This is very reasonable because LR is not sensitive to data that is not highly correlated.

REFERENCES




- [1] E. Ayres, A. Colliander, M. H. Cosh, J. A. Roberti, S. Simkin, and M. A. Genazzio, "Validation of SMAP soil moisture at terrestrial national ecological observatory network (NEON) sites show potential for soil moisture retrieval in forested areas," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10903–10918, 2021, doi: 10.1109/JSTARS.2021.3121206.
- [2] F. Vincent *et al.*, "L-band microwave satellite data and model simulations over the dry chaco to estimate soil moisture, soil temperature, vegetation, and soil salinity," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6598–6614, 2022, doi: 10.1109/JSTARS.2022.3193636.
- [3] G. Patrizi, A. Bartolini, L. Ciani, V. Gallo, P. Sommella, and M. Carratu, "A virtual soil moisture sensor for smart farming using deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022, doi: 10.1109/TIM.2022.3196446.
- [4] T. M. Bandara, W. Mudiyansele, and M. Raza, "Smart farm and monitoring system for measuring the environmental condition using wireless sensor network - IoT technology in farming," in *2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA)*, Nov. 2020, pp. 1–7. doi: 10.1109/CITISIA50690.2020.9371830.
- [5] M. H. Widiyanto, B. Ghilchrist, G. Giovan, R. K. Widyasari, and Y. D. Setiawan, "Development of internet of things-based instrument monitoring application for smart farming," in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, Oct. 2022, pp. 1–6. doi: 10.1109/ICORIS56080.2022.10031470.

- [6] S. Lee, H. Ahn, J. Seo, Y. Chung, D. Park, and S. Pan, "Practical monitoring of undergrown pigs for IoT-based large-scale smart farm," *IEEE Access*, vol. 7, pp. 173796–173810, 2019, doi: 10.1109/ACCESS.2019.2955761.
- [7] A. Dubois, F. Teytaud, and S. Verel, "Short term soil moisture forecasts for potato crop farming: a machine learning approach," *Computers and Electronics in Agriculture*, vol. 180, p. 105902, Jan. 2021, doi: 10.1016/j.compag.2020.105902.
- [8] J. V. Y. Martnez, A. F. Skarmeta, M. A. Zamora-Izquierdo, and A. P. Ramallo-Gonzalez, "IoT-based data management for smart agriculture," in *2020 Second International Conference on Embedded & Distributed Systems (EDiS)*, Nov. 2020, pp. 41–46. doi: 10.1109/EDiS49545.2020.9296443.
- [9] M. H. Widiyanto, A. Ramadhan, A. Trisetyarso, and E. Abdurachman, "Energy saving on IoT using LoRa: a systematic literature review," *International Journal of Reconfigurable and Embedded Systems (IJRES)*, vol. 11, no. 1, p. 25, Mar. 2022, doi: 10.11591/ijres.v11.i1.pp25-33.
- [10] M. H. Widiyanto, M. I. Ardiansyah, H. I. Pohan, and D. R. Hermanus, "A systematic review of current trends in artificial intelligence for smart farming to enhance crop yield," *Journal of Robotics and Control (JRC)*, vol. 3, no. 3, pp. 269–278, May 2022, doi: 10.18196/jrc.v3i3.13760.
- [11] E. A. Abioye *et al.*, "Precision irrigation management using machine learning and digital farming solutions," *AgriEngineering*, vol. 4, no. 1, pp. 70–103, Feb. 2022, doi: 10.3390/agriengineering4010006.
- [12] H. Rahman *et al.*, "IoT enabled mushroom farm automation with machine learning to classify toxic mushrooms in Bangladesh," *Journal of Agriculture and Food Research*, vol. 7, p. 100267, Mar. 2022, doi: 10.1016/j.jafr.2021.100267.
- [13] S. Alex, K. J. Dhanaraj, and P. P. Deepthi, "Private and energy-efficient decision tree-based disease detection for resource-constrained medical users in mobile healthcare network," *IEEE Access*, vol. 10, pp. 17098–17112, 2022, doi: 10.1109/ACCESS.2022.3149771.
- [14] Q. Yin, J. Cheng, F. Zhang, Y. Zhou, L. Shao, and W. Hong, "Interpretable POLSAR image classification based on adaptive-dimension feature space decision tree," *IEEE Access*, vol. 8, pp. 173826–173837, 2020, doi: 10.1109/ACCESS.2020.3023134.
- [15] L. Dong *et al.*, "Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique—subtropical area for example," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 113–128, 2020, doi: 10.1109/JSTARS.2019.2953234.
- [16] R. Zhang *et al.*, "A comparison of gaofen-2 and sentinel-2 imagery for mapping mangrove forests using object-oriented analysis and random forest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4185–4193, 2021, doi: 10.1109/JSTARS.2021.3070810.
- [17] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley and Sons, 2021.
- [18] D. van den Bergh *et al.*, "A tutorial on bayesian multi-model linear regression with BAS and JASP," *Behavior Research Methods*, vol. 53, no. 6, pp. 2351–2371, Apr. 2021, doi: 10.3758/s13428-021-01552-2.
- [19] S. K. Patel *et al.*, "Encoding and tuning of THz metasurface-based refractive index sensor with behavior prediction using XGBoost regressor," *IEEE Access*, vol. 10, pp. 24797–24814, 2022, doi: 10.1109/ACCESS.2022.3154386.
- [20] W. Xue and T. Wu, "Active learning-based XGBoost for cyber physical system against generic AC false data injection attacks," *IEEE Access*, vol. 8, pp. 144575–144584, 2020, doi: 10.1109/ACCESS.2020.3014644.
- [21] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for internet of things: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1622–1658, 2021, doi: 10.1109/COMST.2021.3075439.
- [22] M. Alam, M. S. Alam, M. Roman, M. Tufail, M. U. Khan, and M. T. Khan, "Real-time machine-learning based crop/weed detection and classification for variable-rate spraying in precision agriculture," in *2020 7th International Conference on Electrical and Electronics Engineering (ICEEE)*, Apr. 2020, pp. 273–280. doi: 10.1109/ICEEE49618.2020.9102505.
- [23] E. S. Mohamed, A. A. Belal, S. K. Abd-Elmabod, M. A. El-Shirbeny, A. Gad, and M. B. Zahran, "Smart farming for improving agricultural management," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 3, pp. 971–981, Dec. 2021, doi: 10.1016/j.ejrs.2021.08.007.
- [24] R. Alfred, J. H. Obid, C. P.-Y. Chin, H. Haviluddin, and Y. Lim, "Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks," *IEEE Access*, vol. 9, pp. 50358–50380, 2021, doi: 10.1109/ACCESS.2021.3069449.
- [25] K. Dineva and T. Atanasova, "Cloud data-driven intelligent monitoring system for interactive smart farming," *Sensors*, vol. 22, no. 17, p. 6566, Aug. 2022, doi: 10.3390/s22176566.
- [26] S. J. Russell and P. Norvig, *Artificial intelligence a modern approach*, 3rd ed., London: Pearson, 2010.
- [27] S. H. Chen, A. J. Jakeman, and J. P. Norton, "Artificial intelligence techniques: an introduction to their use for modelling environmental systems," *Mathematics and Computers in Simulation*, vol. 78, no. 2–3, pp. 379–400, Jul. 2008, doi: 10.1016/j.matcom.2008.01.028.
- [28] E. Brynjolfsson and A. Mcfee, *The business of artificial intelligence*. Harvard Business Review, 2017.
- [29] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.
- [30] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/s12525-021-00475-2.
- [31] D. Maulud and A. M. Abdulazeez, "A Review on linear regression comprehensive in machine learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.
- [32] M. S. Acharya, A. Armaan, and A. S. Antony, "A comparison of regression models for prediction of graduate admissions," in *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, Feb. 2019, pp. 1–5. doi: 10.1109/ICCIDS.2019.8862140.
- [33] L. Canete-Sifuentes, R. Monroy, and M. A. Medina-Perez, "A review and experimental comparison of multivariate decision trees," *IEEE Access*, vol. 9, pp. 110451–110479, 2021, doi: 10.1109/ACCESS.2021.3102239.
- [34] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [35] A. I. A. Osman, A. N. Ahmed, M. F. Chow, Y. F. Huang, and A. El-Shafie, "Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1545–1556, Jun. 2021, doi: 10.1016/j.asej.2020.11.011.
- [36] J. Schmeel, "Matrices with applications in statistics," *Technometrics*, vol. 27, no. 1, pp. 88–89, Feb. 1985, doi: 10.1080/00401706.1985.10488021.
- [37] E. Saccenti, M. H. W. B. Hendriks, and A. K. Smilde, "Corruption of the pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models," *Scientific Reports*, vol. 10, no. 1, p. 438, Jan. 2020, doi: 10.1038/s41598-019-57247-4.




- [38] S. Kumar and I. Chong, "Correlation analysis to identify the effective data in machine learning: prediction of depressive disorder and emotion states," *International Journal of Environmental Research and Public Health*, vol. 15, no. 12, p. 2907, Dec. 2018, doi: 10.3390/ijerph15122907.
- [39] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)," *Geoscientific model development discussions*, vol. 7, no. 1, pp. 1525–1534, 2014.

BIOGRAPHIES OF AUTHORS






Mochammad Haldi Widiyanto    5 years teaching at Bina Nusantara University, Bandung. He has a bachelor's degree in telecommunications engineering and has worked as a consultant at a ministry in Indonesia. Next, study for a master's degree in the field of electro-communication. Currently pursuing further studies and getting a doctoral candidate in computer science at Bina Nusantara University. He can be contacted at email: mochamad.widiyanto@binus.ac.id.






Yovanka Davincy Setiawan    is an alumnus of the Faculty of Computer Science, Department of Informatics Engineering, Bina Nusantara University located in Bandung. He has written a paper on "Development of IoTs-based instrument monitoring application for smart farming using solar panels as energy source" which was published at IJRES 2023, which was published in IEEE and presented at ICORIS 2023, and "website and mobile application design with e-commerce and hydroponic digital marketing for micro, small and medium enterprises (MSMEs) in the city of Bandung" presented at the ICCD 2021 conference. He can be contacted at email: yovanka.setiawan@binus.ac.id.



Bryan Ghilchrist    is an alumnus of Faculty of Computer Science, Department of Informatics Engineering, Bina Nusantara University located in Bandung. He has written a paper on "Development of IoTs-based instrument monitoring application for smart farming using solar panels as energy source" which was published at IJRES 2023, which was published in IEEE and presented at ICORIS 2023, and "website and mobile application design with e-commerce and hydroponic digital marketing for micro, small and medium enterprises (MSMEs) in the city of Bandung" presented at the ICCD 2021 conference. He can be contacted at email: bryan.ghilchrist@binus.ac.id.



Gerry Giovan    is an alumnus and has obtained a bachelor's degree in Faculty of Computer Science, Department of Informatics Engineering at Bina Nusantara University, Bandung. His final assignment is an article on the application of IoT in the agricultural sector which was presented at the 2022 International Conference on Cybernetics and Intelligent Systems (ICORIS). He can be contacted at email: gerry.giovan@binus.ac.id.