IoT-Based Soil Moisture Monitoring And Soil Moisture Prediction Using Linear Regression (Case Study of Vinca Plants)

Kuindra Iriyanta [1]
Master of Information Technology,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: kuindra@utdi.ac.id

Bambang Purnomosidi Dwi Putranto
Master of Information Technology,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: bpdp@utdi.ac.id

Widyastuti Andriyani
Master of Information Technology,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: widya@utdi.ac.id

Soil moisture is something that becomes important. Indonesia as an agricultural country, most of the population has a profession as a farmer. In agriculture, one of the important parts is the water composition in the soil or soil moisture. One attempt to maintain soil moisture is to provide sufficient water intake to the soil. However, in practice, it is sometimes complicated for farmers to do proper irrigation of their agricultural land. This humidity condition will ultimately determine the success of vinca plant cultivators. The accuracy of giving water both in terms of time management and volume are two things which are an important focus of vinca crop growing. This system is designed using a humidity sensor which is used to measure the moisture composition contained in the soil, and an air temperature sensor. The NodeMCU ESP2866 micro controller acts as a link between Google spreadsheet sensors. NodeMCU ESP2866 will send humidity and temperature sensor reading data to Google spreadsheets using a RESTfull API which can connect one application to another. The sensor data is then saved to Google spreadsheet and processed using the linear regression method. The processing results will be displayed on the Google Data Studio dashboard. The output of this process is to provide information about soil moisture conditions, notification of soil moisture conditions if it is too dry or damp, thus the prevention of the death of vinca plants can be carried out. The benefit for users is that they can carry out periodic and real-time monitoring by simply using the Telegram instant messaging application, which is expected to reduce the risk of plant death due to drought or excessive watering.

KeyWords: *Internet of Things, Linear Regression, Google Data Studio, NodeMCU ESP2866*

This Article was:
submited: 19-06-23
accepted: 06-07-23
publish on: 14-07-23

How to Cite:
Kuindra I., et al, "IoT-Based Soil Moisture Monitoring And Soil Moisture Prediction Using Linear Regression (Case Study of Vinca Plants)", Journal of Intelligent Software Systems, Vol.2, No.1, 2023, pp.1-5, 10.26798/jiss.v2i1.929

1 Introduction

In agriculture, one of the succeeded determining factors is soil moisture. Soil moisture is the amount of water stored between the pores of the soil[1] This humidity will greatly affect the ability of plants to grow above the ground. High temperatures around the soil will accelerate evaporation which means it will reduce the amount of water in the soil pores. In an attempt to maintain soil moisture, one of the efforts is to provide sufficient water intake to the soil. However, not infrequently, this becomes a new problem if the water is not given correctly. For ornamental plants such as hanging vinca plants, soil moisture conditions are very important, because high humidity will trigger the root rot. Vinca (Catharanthus roseus) or Tapak Dara is an annual shrub. This plant has many benefits including as an ornamental plant. The various color patterns on the hybrid vinca plant are the main attraction for flower-growing enthusiasts in Indonesia[2]. Measurement, monitoring and prediction of soil moisture will be helpful for vinca flower growers to anticipate wilting or rotting of dead plants.

2 Research Methodology

Previous research still focuses on how to monitor and also respond to soil moisture conditions, such as doing automatic watering. However, there has been no research that discusses how to predict soil moisture based on air temperature. This prediction is significant, because using the predictions farmers can prevent and reduce the potential for mortality in vinca plants, so as to increase farmer productivity.

2.0.1 Data Collection. The data applied are air temperature and soil moisture data. This data is obtained using a sensor mounted on vinca plant.

2.0.2 System Flow. This system works as shown in Figure 1. In this image, data is taken using a soil moisture sensor and an air temperature sensor sent to NodeMCU. NodeMCU sends data to Google Drive with Google Sheet as the data logger.

[1]Corresponding Author.

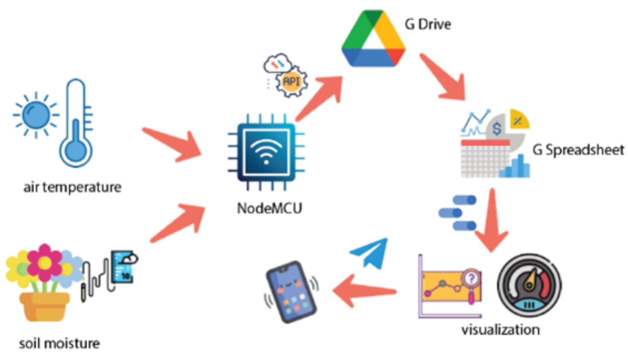

Fig. 1 System Schema

The data which has been obtained and stored in the Google Sheet is then processed using a simple linear regression analysis method to predict soil moisture. The results of the analysis are then visualized with Google Data Studio. Moreover, it provides notifications for vinca farmers when soil moisture is beyond the normal limits. The normal limit for soil moisture is below 45% will be considered dry[3].

Table 1 Humidity Limit

Dry	Wet	Very Wet
45%	46 – 75 %	76%

3 Result and Discussion

3.1 Building the Hardware. The DHT22 sensor as an air temperature sensor is used to read air temperature, connected to pin D7 as a digital input. Meanwhile, the soil moisture sensor is connected to pin A0 as an analog input to read soil moisture. This sensor will later be immersed into the ground so that it can read the soil conditions.

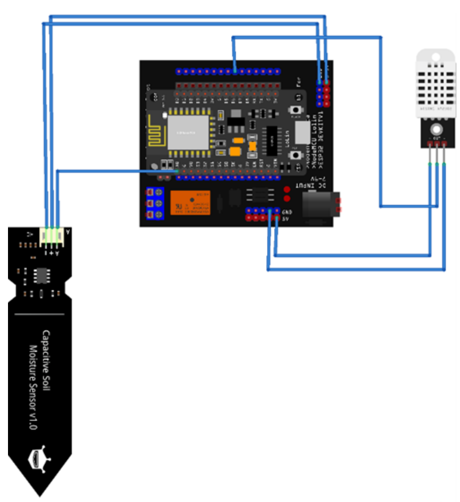

Fig. 2 Hardware Circuit Schema

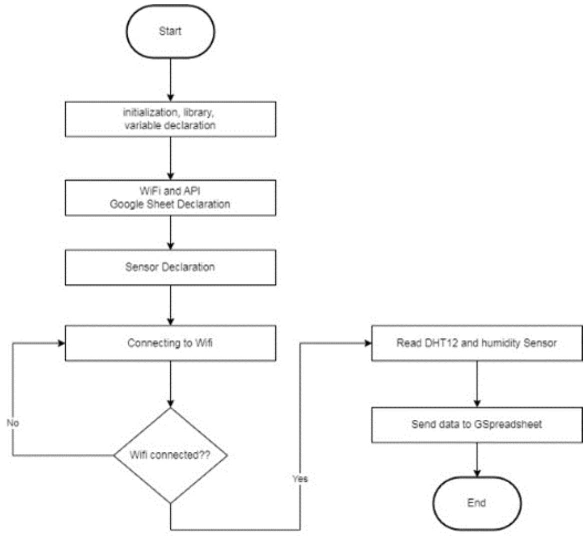

Fig. 3 System Flow

The main program will convert the sensor value then send it to the Google Sheet periodically via an available internet connection. The program is used to obtain data, such as air temperature and soil moisture data. The data obtained is 501 data with data collection carried out from 07.00 to 17.00 with a range of retrieval every 1 minute. The time was chosen because that time is when the air temperature continues to increase. Samples for data collection were sourced from vinca plants in pots which were placed and exposed to direct sunlight. The vinca plant is 4 months old. The morning before measurement, the plants were given sufficient water with a sign that the soil was wet but not awash.

Table 2 Data Collection Results

No	Date	Time	Air Temperature (°C) (X)	Soil Moisture (%RH) (Y)	Condition
1	12 August 2022	7:00:00	26,60	78	Very Wet
2	12 August 2022	7:01:12	26,60	78	Very Wet
3	12 August 2022	7:02:24	26,60	78	Very Wet
4	12 August 2022	7:03:36	26,60	78	Very Wet
5	12 August 2022	7:04:48	26,70	78	Very Wet

From the data above, the information obtained as follows:

$$n = 501$$

$$\sum x.y = 898861, 09$$

$$\sum x = 15360, 90$$

$$\sum y = 29650, 2$$

$$\sum x^2 = 472814, 26$$

$$(\sum X)^2 = 235957248, 8$$

$$\bar{x} = 30, 66047904$$

$$\bar{y} = 59, 18203593$$

$$\sum x-\overline{x}.y-\overline{y}=-10228,24568862$$

$$\sum (x-\overline{x})^2=1841,70748503$$

$$\sum (y-\overline{y})^2=108064,43832336$$

3.2 Simple Linear Regression Analysis. Linear Regression Analysis is an approach to model the relationship between the dependent variable and one or more independent variables. If there is 1 independent variable, it is called simple linear regression, whereas if it has more than 1 independent variable it is called multiple linear regressions[4]. The study involved 2 variables; they are air temperature as the independent variable and soil moisture as the dependent variable. Simple linear regression prediction is used to predict soil moisture conditions based on air temperature. This method is carried out by forming a regression equation so that it can simulate predicting soil moisture conditions with existing data. The linear regression equation is as follows.

$$y=\beta_0+\beta_1 x \tag{1}$$

To determine the relationship of these variables, the Correlation Coefficient equation is applied, one of which is the Pearson correlation coefficient. Pearson correlation produces a correlation coefficient that serves to measure the strength of the linear relationship between two variables[5].

$$[H]r=\frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2(y-\overline{y})^2}} \tag{2}$$

Table 3 Deep Relation

No	Score	Relation
1	0 – 0,2	Indicated that the relationship was very weak,
2	0,2 – 0,4	Indicated that was weak relationship
3	0,4 – 0,7	Stated that the relationship was quite strong,
4	0,7 – 0, 9	Expresses a strong relationship
5	0,9 - 1	Indicated a very strong relationship.

The value of the Pearson correlation coefficient is obtained as follows:

$$r=\frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2(y-\overline{y})^2}}$$

$$r=\frac{(-10228,24568862)}{\sqrt{1841,70748503}\sqrt{108064,43832336}}$$

$$r=\frac{(-10228,24568862)}{(42,91511954)(328,7315597)}$$

$$r=\frac{-10228,24568862}{14107,55418}$$

$$r=-0,7250190613$$

By a value of r = -0.7250190613 , the soil moisture variable is affected by air temperature as 72.5% and the rest is influenced by other variables.
β_0 is a parameter hypothesis test that aims to test whether the constant has a significant role in the regression model which is created. While β_1 aims to test whether the predictor variable has a large effect on the response variable.

$$\beta_0=\frac{\sum Y\sum X^2-\sum X\sum XY}{n\sum X^2-(\sum X)^2}$$

Table 4 Test Results With Real Conditions

No	Soil Moisture (% RH) (Y)	Condition	Measurement Result (% RH)	Condition Prediction
1	75,70	Very Wet	75,69	Very Wet
2	75,70	Very Wet	75,69	Very Wet
3	75,50	Very Wet	75,49	Very Wet
4	75,50	Very Wet	75,49	Very Wet
5	75,50	Very Wet	75,49	Very Wet
6	75,50	Very Wet	75,49	Very Wet
7	75,50	Very Wet	75,49	Very Wet

$$\beta_0=\frac{(29650,2)(472814,26)-(15360,90)(898861,09)}{(501)(472814,26)-(235957248,8)}$$

$$\beta_0=\frac{(14019037372)-(13807315317)}{(501)(472814,26)-(235957248,8)}$$

$$\beta_0=\frac{(211722054,5)}{(236879944,3)-(235957248,8)}$$

$$\beta_0=\frac{(211722054,5)}{(922695,45)}$$

$$\beta_0=229,4603864$$

$$\beta_1=\frac{n\sum XY-\sum X\sum Y}{\sum X^2-(\sum X)^2}$$

$$\beta_1=\frac{(501)(624323,94)-(10850,00)(21427,8)}{(501)(472814,26)-(235957248,8)}$$

$$\beta_1=\frac{(450329406,1)-(455453757,2)}{(236879944,3)-(235957248,8)}$$

$$\beta_1=\frac{(-5124351,09)}{(922695,45)}$$

$$\beta_1=-5,553675473$$

By obtaining the value β_0=229,4603864 and the value β_1=-5,553675473, the regression equation is obtained as follows:

$$y=\beta_0+\beta_1 x$$

$$y=229,4603864+(-5,553675473)x$$

After measuring by means of a sensor, the measurement results are contrasted with conventional measurements, using a soil moisture meter on the market.

3.3 Testing with Real Conditions. After measuring by way of a sensor, the measurement results are compared with conventional measurements, namely using a soil moisture meter on the market. The result is that the results using the sensor are 0.1 points higher.

3.4 Testing with Confusion Matrix. Confusion matrix is a method applied to calculate accuracy on data mining concepts.

Table 5 Prediction and Actual Result Test

No	Time	Air Temperature (°C) (X)	Soil Moisture (%RH) (Y)	Condition	Moisture Prediction (HR)	Condition Prediction
1	8:30:00	28,22	75,70	Very Wet	72,74	Wet
2	8:31:12	28,32	75,70	Very Wet	72,18	Wet
3	8:32:24	28,32	75,50	Very Wet	72,18	Wet
4	8:33:36	28,32	75,50	Very Wet	72,18	Wet
5	8:34:48	28,32	75,50	Very Wet	72,18	Wet
6	8:36:00	28,32	75,50	Very Wet	72,18	Wet
7	8:37:12	28,32	75,50	Very Wet	72,18	Wet
8	8:38:24	28,32	75,30	Very Wet	72,18	Wet

Evaluation with the confusion matrix produces values for accuracy, precision and recalls[6] Confusion matrix is a method applied to calculate accuracy on data mining concepts. Evaluation with the confusion matrix produces values for accuracy, precision and recalls[6]. To check the accuracy, the incoming data is tested and then predicted using the existing regression equation. These results are then compared with the actual situation obtained using conventional measuring instruments. The test took a sample of 102 data. The data is taken at certain hours, between 08.30 - 09.00, 11.30 - 12.30 and 16.00 - 16.30.

The testing data set for the confusion matrix can be formed in a matrix table. The matrix table is a 3 class matrix, because there are dry, wet and very wet classes. Data obtained for the True Positive value are 62 data, namely humidity conditions which are predicted to be wet and true have a wet truth value, or humidity conditions are predicted to be dry and true have a dry truth value. Meanwhile, there are 40 False Positive data, namely 26 predicted values which are wet but in fact dry and 14 predicted values which are wet but in fact very wet.

$$Accuracy = \frac{TP}{TotalData} 100\%$$

$$Accuracy = \frac{62}{102} 100\%$$

$$Accuracy = 60,7$$

Table 6 Confusion Matrix

		PREDICTION Dry	Wet	Very Wet
ACTUAL	Dry	0	26	0
	Wet	0	62	0
	Very Wet	0	14	TP

Table 7 Recall

	Dry	Wet	Very Wet
TP	0	62	0
FN	0	40	0
Recall	0	0,6078431373	0
All Recall		0,20	

Table 8 Precision

	Dry	Wet	Very Wet
TP	0	62	0
FN	26	0	14
Precision	0	1	0
All Precision		0,33	

MAD is used to measure forecasting error in the similar units of measurement as the original data.

$$MAD = \frac{\sum |y_i - \hat{y_i}|}{n}$$

$$MAD = \frac{120,64}{102} = 1,182791926$$

The MAD value is 1.182791926, meaning that this number is quite small with 102 data used as test data. Mean Absolute Percentage Error (MAPE) is a measure of relative error, in addition to Mean Absolute Deviation (MAD).

$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y_i}}{\hat{y_i}} \right| .100\%$$

$$MAPE = \frac{-3,969216744}{102}.100$$

MAPE values can be interpreted or categorized into 4 categories, they are <10% = very accurate, 10-20% = good, 20-50% = fair, >50% = inaccurate[7].

3.5 Comparison with Agricultural Standards. The accuracy value of 60.75 does not impact too much on vinca plants. This 60.7% means that out of 100 times data readings; there is an inaccuracy of 39.3%. This inaccuracy results in misinformation regarding actual soil moisture conditions. The impact on plants is that the plants experience drought. The effect of maximum temperature on plants is that the plant tissue will wither if the temperature reaches 45°C to 55°C for 2 hours[8]. While soil moisture below 40% for 24 hours will damage tissue and below 60% will cause a decrease in plant growth rate[9].

3.6 Data Notification and Visualisation. Notification or early warning mechanisms are carried out using the Telegram platform. Notifications are sent through the Telegram Channel periodically. The notifications are sent every 30 minutes. Meanwhile, if there are abnormal conditions, such as the soil condition is too dry or wet; a warning will be sent in every minute. One tool which can be applied to visualize data is Google Data Studio. Google Data Studio is a data visualization program designed as an easy-to-use tool for representing complex data sets in an attractive and clear way[10]. This data visualization is real-time and web-based so it can be accessed from anywhere.

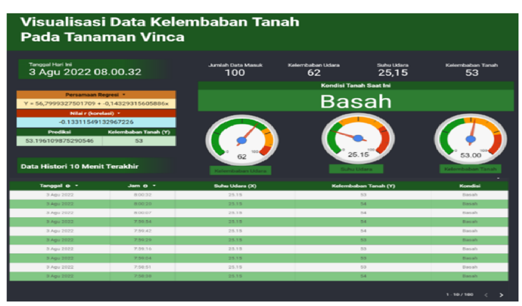

Fig. 4 Data Visualization

4 Conclusion

The process of monitoring air temperature and soil moisture in vinca plants using the predictive analytics method can indeed monitor in real time and also predict soil moisture with an accuracy rate of 60.7% in the confusion matrix test. Prediction of soil moisture using simple linear regression method obtains an accuracy rate of 60.7%. This value illustrates that the use of the simple linear regression method is not suitable in this case. Then an accuracy of 60.7% in its application does not really have an impact on the survival of vinca plants. However, this system can provide information on soil moisture and air temperature in real time via the dashboard or telegram notifications. The real time notifications are also given when the soil conditions are too dry or too wet, and notifications are sent via Telegram.

References

[1] Karyati, Putri, O., R., , and Syafrudin, M., 2018, "Soil Temperature and Humidity at Post Mining Revegetation in PT Adimitra Baratama Nusantara, East Kalimantan Province," Agrifor, **17**(1), pp. 103–114.

[2] Rahmi, H., Nurhaeni, S., and Muharam, 2020, "Pengaruh Berbagai Jenis Zat Pengatur Tumbuh dan Asal Stek Batang Terhadap Pertumbuhan Vegetatif Bibit Tanaman Tapak Dara (Catharanthus roseus (L.) G. Don)," Jurnal Agrotek Indonesia, **2**(5), p. 47.

[3] Husdi, 2018, "MONITORING KELEMBABAN TANAH PERTANIAN MENGGUNAKAN SOIL MOISTURE SENSOR FC-28 DAN ARDUINO UNO," Jurnal Ilmiah, **10**, pp. 237–243.

[4] Kalaian, S. A. and Kasim, R. M., 2017, *Predictive Analytics in Decision Management: Concepts, Methodologies, Tools, and Applications*, IGI Global, Hershey, Pennsylvania, USA.

[5] Safitri., W. R., 2014, "Analisis Korelasi Pearson Dalam Menentukan Hubungan Antara Kejadian Demam Berdarah Dengue Dengan Kepadatan Penduduk Di Kota Surabaya Pada Tahun 2012-2014," Jurnal Kesehatan Masyarakat, **1**(3), pp. 1–9.

[6] Han, J. and Kamber, M., 2006, *Data Mining-Concepts and Techniques (2nd edition)*, Morgan Kaufmann, Burlington, Massachusetts, United States.

[7] Nabillah, I. and Ranggadara, I., 2020, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," JOINS (Journal of Information System), **5**(2), pp. 250–255.

[8] Jaisyurahman, U., Wirnas, D., Trikoesoemaningtyas, and Purnamawati, H., 2020, "Dampak Suhu Tinggi terhadap Pertumbuhan dan Hasil Tanaman Padi," Jurnal Agronomi Indonesia (Indonesian Journal of Agronomy), **47**(3), pp. 248–254.

[9] Sukarman, S., Darwati, I., and Rusmin, D., 2000, "Karakter Morfologi Dan Fisiologi Tapak Dara (Vinca Rosea L.) Pada Beberapa Cekaman Air," Jurnal Penelitian Tanaman Industri, **6**(2), p. 50.

[10] Snipes, G., 2018, "Google Data Studio [Product Review]," Journal of Librarianship and Scholarly Communication, **6**, pp. 1–4.