OPTIMIZING MACHINE LEARNING PERFORMANCE WITH THE NAIVE BAYES CLASSIFIER PROCESS IN SMART FARMING

Made Yosfin Saptura[1], I Wayan Santiyasa[2]

[1,2]Informatika, Universitas Udayana

email: yosfinsaputra2002@gmail.com[1], santiyasa@unud.ac.id[2]

Abstract

Indonesia is a country that relies heavily on the agricultural and plantation sectors to meet its needs for food and industrial raw materials. But farmers face challenges such as falling commodity prices and the negative impact of global warming, which has resulted in widespread drought. As a result, competition for water resources between the agricultural, industrial and household sectors is getting tighter, making it increasingly difficult for farmers to guarantee water supplies. The phenomenon of global warming has caused challenges in the current era. In Bali, although there is a method called "Subak" to manage rice field irrigation systems, it has not been fully implemented. To overcome this problem, a solution is needed that can automate water distribution based on soil moisture levels, temperature, light and air humidity. It uses machine learning techniques specifically using Naive Bayes Classifier to make real-time decisions regarding crop irrigation. The aim of this research is to increase the efficiency and effectiveness of crop irrigation in agriculture while reducing the impact of warming. The results of testing the scenario with orchid plants obtained an accuracy of around 80% and with general plants obtained around 80% which was tested every time 5 data were collected. Testing with a total of 84 training data and 26 test data. From the test results, an accuracy of 92.30769% was obtained.

Keywords : Agriculture, Global Warming, Machine Learning, Naive Bayes Classifier

INTRODUCTION

Indonesia is known as an agricultural country, this is because most of its population works in the agricultural and plantation sectors. In it, farmers are the main actors in the agricultural sector who play an important role in realizing food security. Agriculture is an important sector in the future development of Indonesian society, especially as a provider of food and employment opportunities [1]. Household food needs and raw materials for industrial needs can be met properly. However, farmers are often faced with various complex problems such as decreasing prices of goods, poor product quality, and so on. Added to this is global warming, where there is an increase in global temperature which causes drought everywhere. According to BMKG, in 2022 Indonesia will rank 13th hottest with an anomaly of 0.2 ^{O}C [2]. Global warming has a very broad impact on the agricultural sector. In the future, the water discharge intended for agriculture will become smaller due to competition over the use of water sources by industry and households.

The occurrence of global warming has caused water sources to become smaller and even dry, making it difficult for farmers to obtain water supplies for agricultural land. This forces farmers to be able to organize and share in the irrigation of agricultural land.

Therefore, we need a system that helps monitor the influence of temperature from global warming, soil moisture, and light from the conditions around the plants and take precautions by watering the plants according to the levels needed to maximize existing resources. IoT is a system that connects various devices such as sensors, actuators, microcontroller, and others via the internet to perform certain tasks for users [3]. In this case, it is also necessary to use machine learning to assist in processing previous data to be processed based on various conditions to produce decisions (decision making) for watering plants and the amount of water needed. Apart from that, machine learning is needed to process real-time data based on data obtained directly in the area around the plant so that the method to be implemented is simple but with good accuracy.

This research designs a smart farming system to water plants using a Solenoid Valve,

implementing a smart faucet based on soil moisture, air temperature and light, where the data is processed using machine learning, namely the Naive Bayes Classifier, to carry out processing and produce decisions in real time. With this research, it is hoped that it can help in watering gardens and agriculture effectively and efficiently.

Some of the research that became a reference in the development of this research was research from Zusrotun et al in 2022 regarding "Sentimen Analisis Belajar Online Di Twitter Menggunakan Naïve Bayes" [4]. The Naive Bayes Classifier algorithm was proven to be an accurate algorithm because it produced an accuracy value of 74.08%. To confirm the results of this research, testing was also carried out with K-Fold Cross Validation with k of 15 which resulted in an accuracy value of 76.39%. Research from Saputra and Hakim in 2022 entitled "Implementasi Algoritma Gaussian Naive Bayes Classifier Untuk Prediksi Potensi Tsunami Berbasis Mikrokontroler" [5]. Calculating the potential for a tsunami and applying the microcontroller-based Gaussian Naive Bayes Classifier algorithm with the classification of "Tsunami Potential" and "No Tsunami Potential" has an accuracy of 96%. Research from Pratama et al in 2018 entitled "Implementasi Algoritma Naive Bayes Menggunakan Arduino Uno untuk Otomatisasi Lampu Ruangan Berdasarkan Kebiasaan dari Penghuni Rumah" [6]. From several tests, a percentage of accuracy was produced for the system using the Naive Bayes method with 40 training data and 13 testing data, an accuracy of 84.61% for the first person and 92.30% for the second person. Average processing time of 0.25 seconds from 13 tests. Then research from Ginantri et al in 2022 entitled "Analisis Sentimen Ulasan Villa Di Ubud Menggunakan Metode Naïve Bayes, Decision Tree, Dan K-NN" [7]. The Naive Bayes method after using SMOTE up-sampling got 530 positive, 232 neutral and 104 negative sentiment predictions. The performance confusion matrix obtained is, accuracy 86.66%, precision 88.81%, recall 86.66%, and overall performance 87.38%.

METHODS

A. Machine Learning

Machine learning can be defined as the application of computers and mathematical algorithms adopted by means of learning derived from data used in solving problems [8]. Machine learning is machine learning where the machine will be given a certain algorithm so that the machine can learn and run automatically. The use of machine learning helps in automating

technology so that human activities can be made easier and completed with the help of machines. The machine learning applied is the Naive Bayes Classifier algorithm which will be applied to the smart faucet using datasets from the capacitive soil moisture sensor, light sensor and DHT11 temperature sensor so that the data will be processed and decisions will be made (Decision Making) based on the data.

Naive Bayes Classifier is a simple probability classification algorithm that calculates several opportunities by adding up the frequencies and combinations of values from an existing dataset. Naive Bayes theorem equation:

$$P(c|x) = \frac{P(x|c)\,P(c)}{P(x)}$$

By using Bayes' theorem, we can find the probability of x occurring if c has occurred [9]. This opportunity can be found using the Bayes theorem formula above.
Note:
x　　　: Data for the class is not yet known
c　　　: Hypothesis data from a special class
$P(c|x)$: Probability of hypothesis c based on condition x (posterior probability)
$P(c)$　: Probability of hypothesis c based on condition x (prior probability)
$P(x|c)$: Probability of hypothesis x based on condition c (likelihood)
$P(x)$　: Probability of hypothesis x

Advantages of the Naive Bayes Classifier algorithm:
1. Simple and easy to implement
2. Doesn't require a lot of training data
3. Handle discrete and continuous data
4. Fast and can be used to make predictions in real time

B. Preprocessing

In this research, data was collected and identified for subsequent preprocessing, namely cleaning the data and replacing (missing values). After going through the data preprocessing stage, the data is ready to be used and then processed to the next stage [10]. Data cleaning (to remove inconsistent data noise) Data integration (where fragmented data sources can be combined). Data preprocessing is the stage for carrying out an initial process in data processing. At this stage the data will be processed with the aim of avoiding disturbing data (noise) or inconsistent data [11]. Preprocessing is used to improve machine learning performance and assist in the

application of the Naive Bayes method so that the resulting accuracy increases.

In general, building data preprocessing consists of five main tasks, namely data cleaning, reduction, scaling, transformation, and partitioning [12]. The following is an explanation regarding each type.

1. Data Cleaning. The goal is to improve data quality by eliminating imputed values and removing outliers. In a dataset, if there is still a lot of data missing or empty, then data cleaning is carried out so that the data used is complete data [13].
2. Data Reduction. Reduction is used to reduce the dimensionality of data and therefore, reduce the associated computational costs.
3. Scaling. It aims to transform the original data into the same range for predictive modeling
4. Transformation. It is used to structure the original data into a format suitable for various data mining algorithms.
5. Partition. It aims to divide the entire data set into different groups based on the operating characteristics of the building.

C. Smart Faucet

Smart faucet is an automatic faucet where this technology makes it possible to activate/deactivate the faucet based on a designed algorithm. The use of smart faucets helps in the process of automating the use of faucets based on algorithms designed in the system. This smart faucet uses a 12V ½ inch Solenoid Valve DC as an electric faucet which is powered by DC current. This faucet will also be connected to a 5V Relay as a switch and controlled using a microcontroller.

This system uses several parameters as data to be processed, including soil moisture, temperature, humidity and light. Soil moisture is the amount of water stored between all the soil pores. Soil moisture is caused by several factors, including evaporation through the soil surface, transpiration and percolation. Soil moisture has an important role on the land surface as a system for studying regional water cycles, climate change, agricultural irrigation management, and monitoring environmental conditions. Soil moisture is used for water resource management, warning or early signs of drought, irrigation scheduling, and weather forecasting. Apart from that, air temperature also affects the condition of the plants. Air temperature and humidity affect the rate of water and nutrient absorption, transpiration, and photosynthesis. So monitoring temperature helps in increasing the rate of plant growth and development.

Meanwhile, light intensity affects plant development, photosynthesis and the plant transpiration process. However, excess light will cause parts of the plant to dry out and become damaged due to excess light. Light and temperature are the main factors that influence the plant photosynthesis process. High and low values of temperature and light intensity usually exert adverse effects on plant photosynthesis, thereby greatly reducing plant yield [14]. In addition, light and temperature do not independently influence the photosynthesis and growth properties of plants, but have complementary and interesting interrelationships. Changes in pH, temperature, amount of water, and changes in light intensity greatly affect the quality of plants, especially vegetables [15].

The following is an analysis and system design of Smart Faucet, among others.

A. Basic Design of Machine Learning Based Smart Faucet

Along with the rapid development of technology, new innovations are needed to optimize agricultural land irrigation and plant watering mechanisms. The proposed irrigation innovation is an innovation equipped with automatic valve opening and closing technology which is controlled using a microcontroller connected to a capacitive soil moisture sensor, photodiode, and DHT11 temperature sensor based on the level of soil moisture and air temperature required by the plants. The microcontroller automatically opens and closes the valve based on decision making from the machine learning method applied to classify and predict the soil moisture and temperature data received. That way, this smart faucet system can water based on the condition of the plants and the surrounding environment, whether the plants need water or not.

This research will design a smart faucet device that is placed in the plant area that you want to monitor and water. Apart from that, this system will be connected to a monitoring web server as an interface to monitor data movement and display watering prediction results.

Analysis of system requirements where this research requires several hardware and software for system design, namely:

1. Hardware Requirements
   1. Microcontroller: ESP32
   2. Capacitive Soil Moisture Sensor as a soil moisture sensor
   3. DHT11 as a temperature and humidity sensor

4. Photodiode as a light sensor
5. 5V relay as switch
6. Socket DC 12V Female Connecter 2.1mm
7. 12V 2.1mm adapter
8. Jumper Cables and Breadboard
9. Solenoid Valve as a valve for opening and closing water channels
2. Software Requirements
   1. Arduino IDE
   2. XAMPP
   3. Visual Studio Code
   4. Brave

In general, this research consists of four stages where these stages are the result of modifications to the R&D (Research & Development) method, shown in Figure 1.
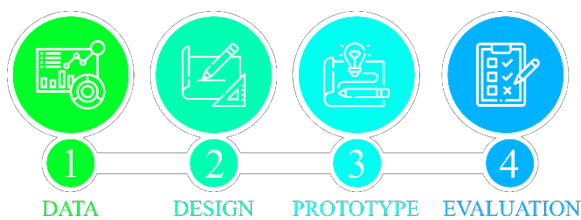

Figure 1. Flow Diagram Stages

B. Dataset Collection and Analysis Stage

Datasets are collected based on soil moisture levels, light intensity, air humidity, and air temperature using sensors installed in areas near the plants. These parameters are used based on previous research as a determinant of whether to water or not. Apart from that, these parameters also determine the environmental conditions in the area near the plant, the condition of the soil, the level of dryness of the air, the intensity of the bright light the plant is currently facing which influences the growth and development conditions of the plant so that knowing these 4 parameters will help in making decisions about watering or No. The next stage, the dataset obtained will be trained using the Naive Bayes Classifier method to obtain an algorithm for decision making regarding faucets.

The ADC value is a value obtained from the results of soil moisture sensor testing on 3 types of soil conditions [16]. The sensor will provide a value from 0 – 1023 as the soil moisture level. Because this research uses an ESP32 as a microcontroller, the ADC range value ranges from 0 – 4095. To convert analog data into digital data, an ADC conversion formula with 12 bit resolution is needed for the ESP32. ESP32 as microcontroller is a functional computer system which contains a processor core, memory and input output equipment as the brain in processing actuators, sensors and others [17]. In this percentage, the data for the humid part ranges

from 70 - 80% where this data adjusts to the surrounding environmental conditions. The temperature level adjusts to the average BPS data for Bali Province in 2022, where the lowest temperature reaches 20.4 ^{O}C and the highest temperature reaches 38 ^{O}C [18]. For medium values, the air temperature ranges from 26 - 32 ^{O}C. For the lux value of light intensity, the percentage value conversion is between 0 and 100% [19].

For the measurement results of orchid plants, we used some data from research by Najikh, et al (2018) which explains the needs of orchid plants in relation to conditions of light intensity, temperature and air humidity and soil moisture. Orchid plants require between 60 – 80% humidity. Humidity in that place also needs to be maintained. Apart from that, during the day, the ideal temperature for orchid plants is between 27 – 30 degrees Celsius [20]. The light intensity adjusts to the need to maintain soil temperature and moisture conditions. For this research, we used a type of orchid plant whose planting medium was soil. Normal conditions for soil moisture are 40% - 60% [21].

The Naive Bayes Classifier method will be applied using discrete data that has been classified at the previous level, then using the following equation:

Table 1. Example Dataset Table

c_1	$\cdots$	c_n	X
…	…	…	yes/no

$$P(x|c_i, \ldots, c_n) = \frac{P(x)\ \prod_{i=1}^{n} P(c_i|x)}{\prod_{i=1}^{n} P(c_i)}$$

Where, c_i = data class or feature, for example soil moisture and temperature
x = data from the class is not yet known.

For data where the numerator does not change, the predictor prior can be removed as follows:

$$P(X|c_i, \ldots, c_n) \propto P(x) \prod_{i=1}^{n} P(c_i|x)$$

The results of the Naive Bayes method will provide a probability with 2 values, namely On and Off of the tap on the smart faucet. After that, the result of this probability is based on $P(On|c)$ comparison with $P(Off|c)$.

From this comparison a decision will be taken where the higher probability will determine the decision making made by the system.

C. Preprocessing

Preprocessing is the process of preparing and cleaning raw data before further processing in data analysis or modeling. This process is carried out to make the dataset produce better values when entered into the machine learning model later. The preprocessing process in this system will carry out data cleaning to avoid the probability value of the dataset and the resulting class getting a value of 0. If the probability of the resulting feature is 0, it will affect the prediction comparison value later between the watering output labels "yes" and "no".

The data cleaning process begins by removing classes for features whose probability value is 0. After this data cleaning is carried out, the class that has a probability value of 0 disappears and this makes the comparison of the predicted output values better and less imbalanced.

Table 2. Class of Each Features

humidity	score (%)	temperature	score (OC)	light	score (%)	soilMoisture	score (%)
dry	<50	high	>32	high	<40	dry	>=40
normal	50-80	normal	26-32	normal	40-60	wet	<40
wet	>80	low	<26	low	>60		
dryO	<60	highO	>30	highO	<40	dryO	>=40
normalO	60-70	normalO	27-30	normalO	40-60	wetO	<40
wetO	>70	lowO	<27	lowO	>60		

Each parameter/feature has a range of values for each class. Humidity consists of 3 classes, namely "dry", "normal" and "wet" classes with the value range for each class implemented as in the table above. The other features are the same, having a wide range of values implemented in the system for the General Plants section.

The Orchid Plant section also has its own class name so that applying the class in Naïve Bayes mode can determine which class is for Orchid Plants or General Plants. In the table above, the letter "O" is added for classes such as dryO, normalO, and others to make it easier to write the model on dataset as a class for orchids. In the Naive Bayes method, each class for features is made unique to assist in the application of training data so that classes for general plants and orchid plants are differentiated. The number of Orchid Plant classes is the same as General Plants, but the class range values are different.

D. Naive Bayes Implementation

The implementation of Naïve Bayes Machine Learning is carried out on a web monitoring system. When the data sent by the microcontroller has entered the web server, Naïve Bayes processing will begin. Naïve Bayes has several processes to get the final probability. Here are several parts to the Naïve Bayes process.

1. Data Training

Training data is data that is used as a guide or reference for the naive Bayes model to determine the prediction output results. The training data implemented in this naive Bayes model is as follows.

Table 3. Data Training

Humidity	Temperature	Light	Soil Moisture	Output
normal	high	high	dry	yes
dry	high	high	dry	yes
dry	normal	high	dry	yes
…	…	…	…	…
wetO	lowO	lowO	wetO	no

The class of each feature is adjusted to previous research and also adapts to real sensor conditions with tested data. The number of training data combinations used was 84 data. The training data above for general plant and special plant data is combined. After that, the probability calculation for each variable will be carried out for the Naive Bayes method.

2. Class Probability

Class probability calculations are carried out by counting the number of each class output that has a value of yes and no. From the results of the implemented training data, the class probability results are as follows.

Table 4. Class Probability

Output	Value
yes	0,607142857
no	0,392857143

Class probability is obtained by, for example, dividing the number of output classes by the total number of classes. For output value no, have the same way with that.

3. Feature Probability

To get the feature probability or likelihood, this can be done using, for example, the Humidity feature, for the dry class with the output value yes. The calculation is the number of dry class Humidity features whose output is yes divided by the number of Humidity features whose output is yes. For an output with a value of no, to get the probability the method is the same as for a yes output. The following are the results for the Humidity feature probability.

The probability of other features such as Temperature, Light and Soil Moisture is the same as the Humidity feature.

Table 5. Humidity Feature Probability

Humidity	yes	no
dry	0,17647	0,09091
normal	0,21569	0,21212
…	…	…
normalO	0,21569	0,21212
wetO	0,09804	0,21212

4. Classification

This classification is the process of implementing the Naive Bayes model on testing data. Several probabilities that have been obtained will be used to predict the output value on the specified testing data. The following is an example of applying classification to testing data.

Table 6. Data Testing Classification

Humidity	Temperature	Light	Soil Moisture	Output	Prediction	Prediction Yes	Prediction No
wet	normal	high	dry	yes	Yes	0,0007269	0,0001043
normal	normal	normal	wet	no	No	0,0003998	0,0016696
…	…	…	…	…	…	…	…
wetO	lowO	highO	wetO	yes	No	0,0001346	0,0009739

To get the predicted value, it is done by comparing the predicted output yes with the output no. To find the predicted value of the output yes, you can do this by multiplying the probability value of the feature with the output value yes in the testing data, for example wet, normal, high, dry, and output yes probability. So it produces a value of 0.0007269 as in the table.

To predict the output no, the method is almost the same but using the probability feature for the output no. After that, compare the values between the predicted output yes and no. The highest value will be the predicted value from the Naive Bayes model.

E. Technical Design of Tool Stage

The preparation of the technical design will be used as a guide in manufacturing to produce appropriate dimensions. The running system flow is shown in Figure 2 with the following description. This Smart Faucet has a system workflow that starts from collecting soil moisture data using a soil moisture sensor, air temperature and light.

Soil moisture sensors will be embedded in the soil adjacent to the plants, temperature and light sensors will be installed in the control box. These sensors will send data to the microcontroller. After the data is sent, the data will be sent to the server to store the data in the database and will be processed using the Naive Bayes model to produce predictions. The results of this prediction will be sent to the microcontroller to produce action on the Solenoid Valve.
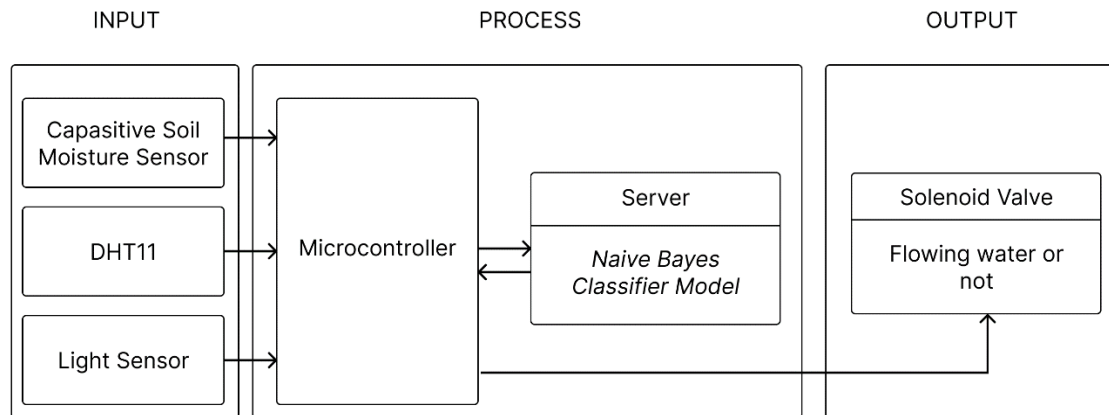
Figure 2. Smart Faucet System Overview

F. Manufacturing and Deployment Stages

Hardware design uses components in requirements analysis where the wiring design is adjusted to the system circuit being developed. All components that have been manufactured are combined according to the system design as shown below. This is the wiring or wiring circuit design of the smart faucet system which will be implemented later.

At the top is a solenoid valve which is connected to a female DC socket and also to a relay as a switch to disconnect or connect the electric flow from the socket to the solenoid valve.

Then on the far right of the image there is a light sensor connected to the ESP32 as a microcontroller on the smart faucet. On the left, there is an RGB LED as a tool indicator, then DHT11 as a temperature and humidity sensor. The sensor will also be connected to the ESP32 to transmit data.
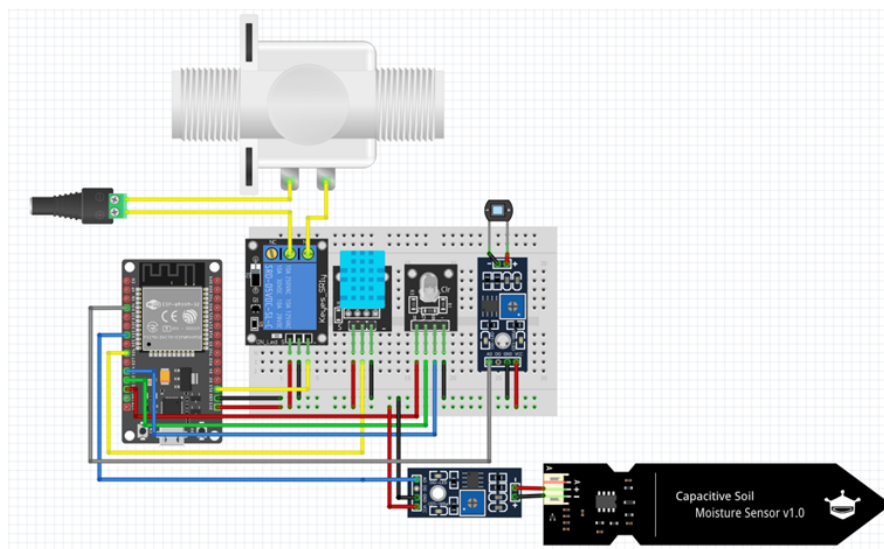

Figure 3. Smart Faucet System Wiring Series

The next left part is a relay and finally the leftmost part is the ESP32 microcontroller. At the bottom of the breadboard, there is a soil moisture sensor (capacitive soil moisture sensor) to send soil moisture data to the tool. The power source here will be divided into 2, namely the power source for the ESP32 using a 5V electric current and for the solenoid valve via a socket using a 12V electric current.

G. Evaluation Functionality Testing Stage

This stage will test the smart faucet system with conditions that occur in real time. This test will be carried out to test the level of accuracy and delay of the system. The level of accuracy of the watering system is measured by comparing the results between the dataset and the results from the system to determine whether the results are appropriate or not. The accuracy of the Naïve

Bayes model will be measured by comparing the calculation results of the system with manual calculations. System accuracy assessment can be carried out using reference data from trusted agencies or using sample data that has been tested [5].

RESULTS AND DISCUSSION

When creating a web monitoring system, HTML, CSS, PHP and Javascript are combined into one unit as shown in the following image.
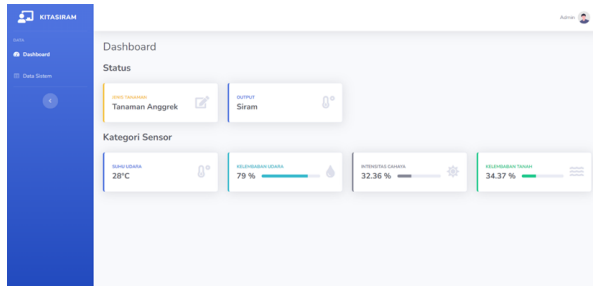

Figure 4. Web Dashboard View

In figure 4 is a display of the Web Monitoring Dashboard section which displays data which is divided into 2, namely Status and Sensor Category. The Status section contains the types of plants that are currently determined, of which there are 2 types, namely General Plants and Orchid Plants. This section can be changed according to the plants currently monitored by the system. Then the Sensor Category section contains sensor data on the watering system, namely air temperature, air humidity, light intensity and soil moisture. This section displays the latest sensor data sent to the database.

Apart from that, in Web Monitoring there is a System Data section which contains detailed information on data that has been saved to the database. This section contains tables that include time, data on sensor types, status of watering decision results, and plant type. In this section there is also a delete data feature to reduce the amount of historical data in the database. The following is a display of the System Data section.
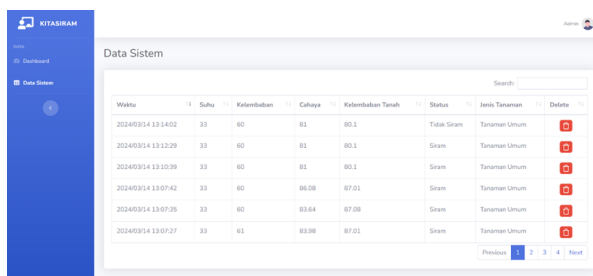

Figure 5. System Data View

In Figure 6 below is the hardware design, there is several hardware such as ESP32, Light sensor, DHT11, Solenoid Valve, Relay, Soil Humidity sensor, and cables for electrical media. These devices have the function of detecting air temperature, air humidity, light intensity and soil moisture.
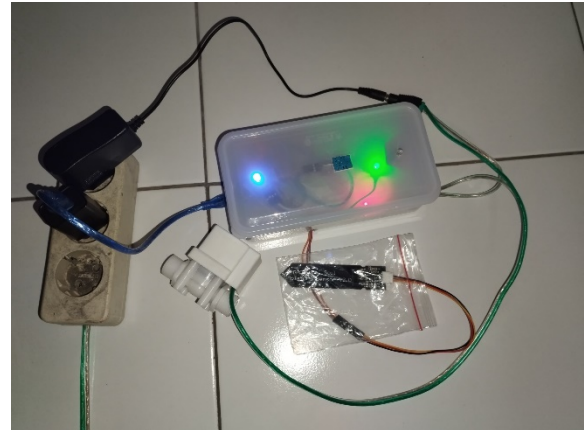

Figure 6. Hardware View

Figure 7. Hardware Implementation

Apart from that, the ESP32 as a microcontroller is used to control sensors in retrieving data and processing data to be sent to the server via WiFi. Apart from monitoring data, this system also plays an important role in the communication process with the server to send machine learning results which will later be received by the microcontroller to be processed as a watering action. Each data processing result can be seen on the local server database.

To ensure the system can function properly, a system test is carried out which consists of several parts, the following is the test part.

A. Data Delivery Testing

This data transmission test was carried out by testing the automatic watering tool. The test is carried out by checking that when the tool is turned on, it can send sensor data to the

database or not. And also check whether the server is able to send a response back to the tool. If the data is successfully sent from the device to the server, checking the success of data delivery can be seen via the server database. Then the success of sending the response from the server to the database can be checked on the Serial Monitor on the Arduino IDE application connected to the tool. The following are the test results for sending data from the tool to the server.

Figure 8. Data Delivery Display on Web Monitoring

The results of sending data from the tool to the server can be checked as in Figure 8 where the tool sent data to the server at 2024/03/15 07:07:54 to 2024/03/15 07:08:28 with a total of 3 data transfers. The results of successful delivery are recorded in the database and displayed in the web monitoring System Data section. Then the results of sending responses from the server to the tool can be seen as follows.

In figure 9 below is a display of the results of sending data to the Arduino IDE serial monitor connected to the tool. It can be seen that if the tool successfully sends data to the server, it will give a payload response in the form of "flush" or "not flush".

Figure 9. Data Sending Display on the Tool

In the image above, it can be seen that the tool successfully sent data with a print result of "Data successfully sent" and received a response of "no flush" 3 times and also. The httpCode display functions to display whether the tool is connected to the server or not, which indicates the tool can communicate and send data. The indication of the device's connection to the server is marked with an RGB LED.

B. Naïve Bayes Machine Learning Testing

Testing was carried out on the Naïve Bayes model which had been preprocessed using data cleaning. After carrying out this process, 100 training data and testing data were taken randomly. The resulting dataset for training data was 84 data and testing data was 26 data. The following are the results of testing the accuracy of the model.

Table 7. Machine Learning Testing and Manual Calculations

Humidity	Temperature	Light	SoilMoisture	Output	System	Manual
wet	normal	high	high	yes	yes	Yes
normal	low	high	high	yes	yes	Yes
wet	low	high	high	yes	yes	Yes
normal	high	normal	high	yes	yes	Yes
high	high	normal	high	yes	yes	Yes
high	normal	normal	high	yes	yes	Yes
high	normal	normal	wet	no	no	No
normal	normal	normal	wet	no	no	No
wet	normal	normal	wet	no	no	No
normal	low	normal	wet	no	no	No
wet	low	normal	wet	no	no	No
wet	low	low	high	no	yes	Yes

Humidity	Temperature	Light	SoilMoisture	Output	System	Manual
normal	high	high	wet	yes	yes	Yes
highO	normalO	normalO	highO	yes	yes	Yes
normalO	normalO	normalO	highO	yes	yes	Yes
wetO	normalO	normalO	highO	yes	yes	Yes
normalO	lowO	normalO	highO	yes	yes	Yes
wetO	lowO	normalO	highO	yes	yes	Yes
normalO	highO	lowO	highO	yes	yes	Yes
highO	highO	lowO	highO	yes	yes	Yes
wetO	normalO	highO	wetO	no	no	No
normalO	lowO	highO	wetO	no	no	No
wetO	lowO	highO	wetO	no	no	No
normalO	highO	normalO	wetO	yes	yes	Yes
normalO	highO	lowO	wetO	no	yes	Yes
highO	highO	lowO	wetO	yes	yes	Yes

Judging from the table above, the accuracy value from comparing the number of system calculations, manual calculations, and output from testing data produced is 24 / 26 x 100% = 92.30769%. The influence of preprocessing causes an increase in the accuracy value of the applied Naïve Bayes model.

C. System Accuracy Testing

System accuracy testing is carried out using direct data. There are 10 sample data used which are divided into 2, namely 5 general plants and 5 orchid plants. The following is a table of test results.

Table 8. Orchid Plant System Accuracy Testing

Sensor				Expected Status	System Results	Information
Temperature	Humidity	Light	Soil Moisture			
28	79	32,36	34,37	No	No	Correct
36	55	34,24	56,24	Yes	Yes	Correct
49	27	39,56	52,7	Yes	Yes	Correct
49	28	39,05	54,7	Yes	Yes	Correct
55	21	36,04	32,5	Yes	Yes	Correct

Table 9. General Plant System Accuracy Testing

Sensor				Expected Status	System Results	Information
Temperature	Humidity	Light	Soil Moisture			
35	55	33,94	33,58	Yes	Yes	Correct
35	57	34,16	35,97	Yes	Yes	Correct
39	51	37,07	34,02	Yes	Yes	Correct
49	30	37,9	76,58	Yes	Yes	Correct
54	23	37,83	33,55	Yes	Yes	Correct

From the results table 8, it is concluded that the system runs well in providing results for each manual calculation that has been carried out. The accuracy obtained is 5 / 5 x 100% = 100%. To test system accuracy on general plants, see the following table.

Table 9 displays the results of general plant accuracy testing, which based on the information shows the results are in accordance with the manual calculations carried out. From the results shown, the system runs well in processing data with an accuracy of 5 / 5 x 100% = 100%.

CONCLUSION

Based on all stages of research that have been carried out in the development of a machine learning-based smart faucet monitoring and watering system, the following conclusions can be drawn.

1. Monitoring of the watering system is carried out based on several sensors, namely temperature sensors, air humidity, light intensity and soil moisture. The performance of the system functionality runs quite well according to the tests that have been carried out previously.

2. The results of the test scenario with orchid plants obtained an accuracy of around 100% and with general plants obtained around 100% which were tested each time 5 data were collected.

3. Tests to increase accuracy have been carried out by eliminating classes in the Naïve Bayes model whose probability is 0. There are 2 classes removed, including the SoilMoisture feature, the normal and normalO classes. Then testing and changes were carried out on the dataset so that the total for training data was 84 data and testing data was 26 data. From the test results, the accuracy obtained was around 92.30769%.

REFERENCES

[1] Cahyaningrum, Y.D. 2020. Sektor Pertanian: Harapan Pemulihan?. URL : https://feb.ub.ac.id/sektor-pertanian-harapan-pemulihan/. Diakses tgl. 25 Februari 2023.

[2] BMKG. 2023. Anomali Suhu Udara Rata-Rata Bulan Januari 2023. URL : https://www.bmkg.go.id/iklim/?p=ekstrem-perubahan-iklim. Diakses tgl 19 Februari 2023.

[3] Fajar, M., Chandra, F.B., and Arfandy, H. 2023. Applying Use Case 2.0 Approach To The Development Of Iot-Based Rainfall Monitoring System. Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI. 12(2): 244-253.

[4] Zusrotun, O.P., Murti, A.C., and Fiati, R. 2022. Sentimen Analisis Belajar Online Di Twitter Menggunakan Naïve Bayes. Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI. 11(3): 310-320.

[5] Saputra, D.I. and Hakim, D.L. 2022. Implementasi Algoritma Gaussian Naive Bayes Classifier Untuk Prediksi Potensi Tsunami Berbasis Mikrokontroler. Journal of Electrical Engineering and Information Technology. 2 (20): 122-138.

[6] Pratama, S.M., Kurniawan, W. and Fitriyah, H. 2018. Implementasi Algoritma Naive Bayes Menggunakan Arduino Uno untuk Otomatisasi Lampu Ruangan Berdasarkan Kebiasaan dari Penghuni Rumah. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. 2 (9): 2485-2490.

[7] Ginantri, N.L.W.S.R., Yanti, C.P., Prasetya, G.D., Sarasvananda, I.B.D., and Wiguna, I.K.A.G. 2022. Analisis Sentimen Ulasan Villa Di Ubud Menggunakan Metode Naïve Bayes, Decision Tree, Dan K-NN. Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI. 11(3): 205-216.

[8] Roihan, A., Sunarya, P.A. and Rafika, A.S. 2020. Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper. IJCIT (Indonesian Journal on Computer and Information Technology). 5 (1):75-82.

[9] Pradana, L.E. and Ruldeviyani, Y. 2023. Sentiment Analysis of Nanovest Investment Application Using Naive Bayes Algorithm. Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI. 12(2): 283-293.

[10] Romli, I. and Dewi, R.F.P. 2021. Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit ISPA. Indonesian Journal of Business Intelligence. 4 (1):10-15.

[11] Alghifari, F. and Juardi, D. 2021. Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. Jurnal Ilmiah Informatika. 9 (2):75-81.

[12] Fan, C., Chen, M., Wang, X., Wang, J., and Huang, B. 2021. A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. Frontiers in Energy Research. 9: 1-17.

[13] Subawa, I.G.B. 2019. Prediksi Kelulusan Mahasiswa Menggunakan Teorema Teorema Bayes (Studi Kasus: Universitas Pendidikan Ganesha). Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI. 8(3): 227-236.

[14] Zhou, J., Li, P., and Wang, J. 2022. Effects of Light Intensity and Temperature on the Photosynthesis Characteristics and Yield of Lettuce. Holticulturae. 8(178): 1-1.

[15] Ridwan, K.M.S.M. 2021. Penerapan IoT Dalam 62 Sistem Otomatisasi Kontrol Suhu, Kelembaban, dan Tingkat Keasaman Hidroponik. J. Tek. Pertan. Lampung, 10(4): 481–487, doi: 10.23960/jtep-l.v10i4.481-487.

[16] Rahardjo, P. 2022. Sistem Penyiraman Otomatis Menggunakan Sensor Kelembaban Tanah Berbasis Mikrokontroler Arduino Mega 2560 Pada Tanaman Mangga Harum Manis Buleleng Bali. Majalah Ilmiah Teknologi Elektro. 21 (1):31-34.

[17] Prayasa, I.W.O., Abiyasa, A.P., and Kristyawan, I.P.A. 2023. Ceramic Firing Temperature Trajectory Monitoring System On Iot-Based Gas Furnace. Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI. 12(3): 398-405.

[18] BPS, 2022. Pengamatan Unsur Iklim di Stasiun Geofisika Denpasar, 2022. URL : https://bali.bps.go.id/statictable/2022/07/13/183/pengamatan-unsur-iklim-di-stasiun-geofisika-denpasar-2021.html. Diakses tgl. 12 April 2023.

[19] Hidayatullah, P., Orisa, M., and Mahmudi, A. 2022. Rancang Bangun Sistem Monitoring dan Kontrol Tanaman Hidroponik Berbasis Internet Of Things (IOT). Jurnal Mahasiswa Teknik Informatika. 6 (2):1200-1207.

[20] Najikh, R.A., Hanafi, M.H., and Kurniawan, W. 2018. Monitoring Kelembaban, Suhu, Intensitas Cahaya Pada Tanaman Anggrek Menggunakan ESP8266 Dan Arduino Nano. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. 2(11): 4607-4612.

[21] Mas, F.A., Suciyati, S.W., Pauzi, G.A., and Junaidi. 2022. Smart Greenhouse Monitoring With Soil Temperature and Humidity Control on Internet of Things (IoT) Based Orchid Plants. Journal of Energy, Material, and Instrumentation Technology. 3 (3):79-88.