Investigations of Naive Bayes, Random Forest and Decision Tree Algorithm for building aeroponics system

Arunraja A
Department of Electronics and Communication Engineering
Christ (Deemed to be University)
Bengaluru, India.
e-mail: arunraja.a@christuniversity.in
Manjunatha Swamy C
Department of Computer Science and Engineering
Christ (Deemed to be University)
Bengaluru, India.
e-mail: manjuc_5@yahoo.com
Kiran B
Department of Computer Science and Engineering
ATME College of Engineering,
Mysuru India.
e-mail: ranjankiran03@gmail.com
Senthilnathan S
Department of Electronics and Communication Engineering
Christ (Deemed to be University)
Bengaluru, India.
e-mail: senthilnathan.s@christuniversity.in
D.Vinodha
Department of Computer Science and Engineering
Christ (Deemed to be University)
Bengaluru, India.
e-mail: vinodha.rajkumar@gmail.com

Abstract - Urban farming has becoming popular and most adopted recently as society has become more conscious about the food quality they consume and conservation of resources. The Internet of Things (IoT) and machine learning algorithms are employed to build and implement an automated aeroponics system. This system contains Arduino as a main processing unit and uses Node MCU for cloud storage. It gets inputs from sensors like ultrasonic sensor, light sensor, humidity sensor, temperature sensor and exhibit the output in a LCD unit. In response to the inputs, water pump, cooler fan and light acts automatically. The data are stored in cloud as a record for future use. Using machine learning algorithms Naive Bayes, Random Forest and Decision Tree the datasets from cloud are trained and tested to know the crop suitable for the given condition. According to the results, the system has the potential to save more labour and reducing water consumption than conventional methods.

Keywords- Aeroponics, Naive Bayes, Node MCU, Random Forest, Internet of Things and Decision Tree

I. INTRODUCTION

Agriculture is an important sector contributing to global economy and it supports livelihood. Several new systems, practices and technologies, and approaches have been implemented over time, driving the advancement of agriculture. One out of every three workers worldwide rely on agriculture for their livelihoods. It makes a substantial contribution to the economic development of underdeveloped countries, while also driving economic prosperity in developed countries. According to the findings of several studies, only 17% of the land is irrigated by global agriculture, by using 70% of fresh water that is available annually. Due to global warming, the overall amount of irrigated land is gradually shrinking so it leads to increase the food demand. To keep with the demands of the fast-growing population and urbanization, there must be a seventy percent increase in world food production. Thus Urban farming techniques are adopted. Aeroponic technology is one way to provide plants the perfect atmosphere. Equipment within the aeroponic environmental system is responsible for maintaining optimal levels of lighting, cooling, ventilation, heating, and air movement. The heating needs are determined by the specific temperature requirements of the growing plant. By incorporating a circulation fan into the setup, it become possible to keep the temperature in an aeroponic device at a constant level. Observing the colour of the leaves is essential in determining the appropriate level of lighting, emphasizing the importance of proper lighting. Plants grown in aeroponic systems can be watered automatically. Overwatering can have severe consequences for plant growth and survival, underscoring the importance of proper watering and nutrient provision. The irrigation system is regularly turned on and off by a control device and nutrients are supplied through water. The time a plant is actively growing.

Aeroponics is an alternative method for soilless cultivation in environments that control growth. The method known as aeroponics involves either spraying roots with hydroponic solutions that are floating in the air or growing plants in an atmosphere of mist or air without the use of soil. No soil nor aggregate media are used in it. In an aeroponic system, the plant roots are hung and nutrient-rich fluid is sprayed at predetermined intervals. This method delivers nutrients very accurately, and because roots have enough oxygen, growth could happen more quickly. Aeroponics technique can be done as a vertical farming that involves growing crops in layers that are piled vertically. Because of the growing population and a shortage of arable land, vertical farming may be employed to meet the world's rising food demands.

A Supervised learning Algorithm called Naïve Bayes that uses Bayes' theorem to classify data. It is commonly used for text categorization and requires a large training set to function effectively. Despite its simplicity, Naive Bayes is a highly effective classification algorithm that enables the creation of accurate machine learning models. The algorithm predicts the likelihood of an event occurring by using a probabilistic classifier.

Bayes' theorem is used to calculate the probability, it is a crucial factor in Bayes' theorem. The following formula is used to represent Bayes theorem

$$P(A\backslash B)=P(B\backslash A)P(A)/P(B) \quad (1)$$

Where, *The posterior probability, or P(A|B),* represents the probability that a given hypothesis (A) will occur. While the Likelihood Probability *P(B|A)* measures the probability that the hypothesis is true based on the available evidence. The prior probability (P(A))is the probability of the hypothesis before considering any new evidence P(B) stands for the Marginal Probability. The Naive Bayes classifier involves the following steps: 1. From the supplied dataset the frequency tables are build. 2. The probability table is created by calculating the odds of the provided attributes.

3. Then by using Bayes theorem the posterior probability is determined. The Random Forest Algorithm is a widely used supervised machine learning technique for regression and classification problems. The problem-solving ability and accuracy of a Random Forest Algorithm improve, as more trees are added to the algorithm. The Random Forest classifier employs multiple decision trees on various subsets of the input data to improve the predictive accuracy of a dataset. This technique is based on the concept of ensemble learning, which involves integrating several classifiers to solve a complex problem and it enhance the model's performance.

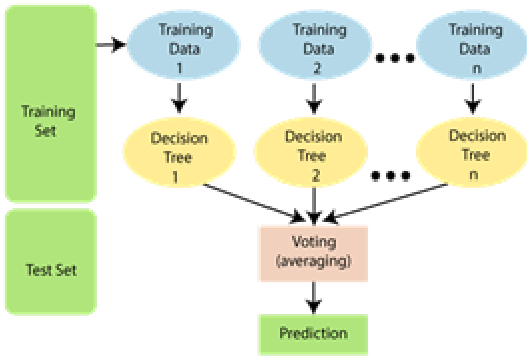


Figure. 1. Working of Random Forest Algorithm

The following steps of the Random Forest Algorithm's functioning are described:

Step 1: From a given training set or data collection choose the random samples.
Step 2: The algorithm will build a decision tree, For each training set of data.
Step 3: By using an average of the decision tree the voting will be conducted.

Step 4: At last, the final prediction result is found based on the maximum votes received.

Decision tree algorithm is a supervised learning approach, mainly used for classification and regression applications to solve the problem. It is organized hierarchically. It has a root node, internal nodes, leaf nodes and branches. Decision trees are a machine learning technique that resembles human decision-making, making them easy to comprehend. They exhibit a tree-like structure, making the logic behind-making transparent. The algorithm starts from the root node and then moves forward to predict the class of the tree in the given dataset. It follows the branches and compares the root attribute values with the attributes of the record (real dataset) to jump to the next node. The algorithm repeats this process by checking the attributes values with other sub-nodes before moving to next node until it reaches the tree's leaf node.

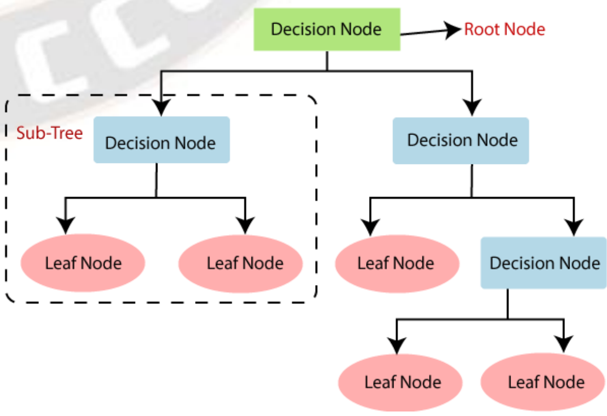


Figure. 2. Working of Decision Tree Algorithm

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

II. LITERATURE REVIEW

Issac K Joy1 and Christy V Vazhappilly (2022) proposed the design and construction of an aeroponics system using the Internet of Things (IoT) for online and automated monitoring. Aeroponics provides a significant reduction in water usage with higher yields when compared to hydroponics or traditional farming. In the tank a pH sensor is used to maintain the solution's pH level. The temperature and humidity levels in this chamber are constantly monitored by the AM1011A sensor. The liquid level sensor is used to liquidate the tank level using a peristaltic pump. With the At mega 328P microcontroller and ESP 32 Wi-Fi module, all of these operations are accurately timed and transmitted to the ThingSpeak Platform and Blynk App for user engagement.

Adache Paul Et al.,(2022) proposed a system that is independent automated structure for indoor small farming in urban areas does not require soil or manual watering. This mechanization will stimulate the provision of fresh product in a set quantity to those who aren't even actively engaged in gardening, and it will also allow for the monitoring of system status using a web service that provides parameter data. They came to the conclusion that the system would become more effective and exact by adding a pH sensor, which would also measure the nutrient solution's quality.

Reddy DJ and Kumar MR (2021) proposed a structured review to extracts and synthesize the features for predicting crop yield. Artificial intelligence techniques have been developed for analyzing crop yield prediction (CYP), but the Neural Network method has limitations in reducing relative error and decreasing prediction accuracy. Supervised learning approaches for selecting, grading or sorting fruits also face challenges due to the non-linear relationship between input and output variables. Several studies have proposed different methods to establish effective and accurate models for classifying the crop also includes crop yield estimation, it is done on the basics of crop disease, weather, classification of crops based on the growing phase, and other factors.

Shahhosseini M et al., (2021) suggested that by giving the input as simulation crop model variables (APSIM) to ML models can reduce the yield forecast RMSE (Root Mean Squared Error) from 7 to 20%. The most crucial APSIM variables for ML predictions, according to additional research into the partial inclusion of APSIM characteristics in ML prediction models, are those linked to soil moisture, followed by parameters related to crops and phenology. The mean water table depth over the growth season and the mean simulated APSIM drought stress have been found to be the two most significant APSIM inputs to ML, according to the feature importance measure. This suggests that in order for ML models to give better yield estimates, additional hydrological inputs are necessary in addition to meteorological data.

Iwan Fitrianto Rahmad et al., (2020) a suggestion was made to build an automatic control system for aeroponics. This system's main processing unit is a latte panda, and the interface may be changed to fit the needs of the grown plant. It outputs data to the LCD screen module after receiving it from temperature, light, and humidity sensors. With the help of simulation software, the system results are simulated. The evaluation of the experimental results of the aeroponics control system is then based on the surrounding environment. An aeroponic mechanism is used to grow the intended lecture. To develop an automatic control system in aeroponics the green plant was chose. Thus the results show the system is capable of reducing labor expenses and it increase the economic value of the final product.

III. BLOCK DIAGRAM

The system that been suggested comprises of a pair of microcontrollers, namely the Arduino UNO and the Node MCU. The former is used for sensing the parameters, displaying the outputs and controls the aeroponic system. Node MCU is used for cloud storage. The IoT sensors consisting of LM35 temperature sensor, Ultrasonic sensor, LDR sensor, Moisture sensor reads the input and displays the output on a 16x2 LCD display. Water pump, light and cooler fan are included for controlling the suitable environment for plant growth.

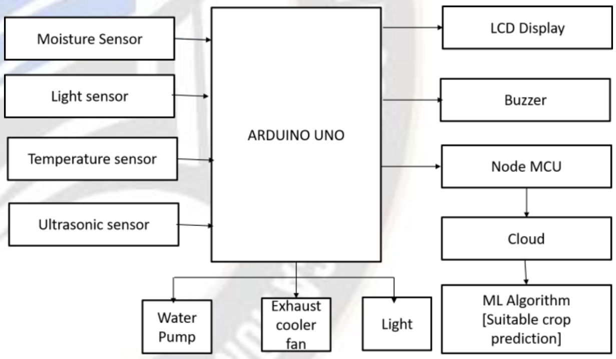


Figure. 3. Block Diagram

Methodology And Working

The actions in this system are all controlled by an ATmega328P. Depending on the program that is uploaded, it will assign the appropriate timing and responsibilities to the relevant components. On the basis of a microcontroller signal, relays are used to supply power to output devices like pumps, fans, and lights. The system measures air temperature using a temperature sensor. The cooling fan is turned on to raise the temperature to the normal range if it is outside of it. When the temperature range exceeds 30°C, cooling fans turn on and remain on until it does. Using relay devices, circuits are opened and closed. Submersible pump will be triggered if the moisture sensor value is less than the threshold value, and it will run by spraying water with necessary amount of nutrients. The input sensor data is sent to the ThingSpeak platform using the ESP 32 Wi-Fi module as an intermediary. Thinkspeak data is then accessed by using the read API key or read URL of the channel. After scrapping the data from thinkspeak the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

prediction algorithms such as Random Forest, Naïve Bayes, Decision Tree are used for crop recommendation of aeroponic system by considering the parameters such as temperature, humidity, phosphorous, nitrogen, and potassium.

The prediction algorithm involves the following process,

- Data collection
- Data Pre-Processing
- Training data and Test data
- Model Creation
- Model Prediction

Data collection is a Gather data while taking into account factors such as temperature, humidity, potassium, nitrogen, and phosphorus, whereas previous research considered only finite number of factors that impact the prediction of crop yield. The two parameters that have the biggest effects on crop yield prediction are temperature and nutrients.

Data Pre-Processing is an important process, it means in our data the important changes like cleaning, transforming and integrating the data are made before giving it to the algorithm. The main purpose is to increase the quality of data.

Training data and Test data

1. For the purpose of selecting a model, we split our dataset into train and test.
2. Here, the data are split into a 3:1 ratio.
3. 70 percent of the data is selected from training dataset, and 30 percent of data from testing dataset.
4. The split process is being performed using the train-test-split model.
5. The results are xtrain, xtest and ytrain ,ytest.

Model Creation

- Incorporate context-based learning into your organization.
- Examine the data and select the algorithm type.
- Make the dataset ready and clean.
- Split the available dataset, then carry out cross validation.
- Implement the model.

Model Prediction

One can apply convolution neural network and multi-layer perceptron algorithm to predict the accuracy by its specifications

Random Forest is well known due to its scalability, ability to handle high-dimensional data and accuracy. Random Forest offers several advantages when compared to other machine learning techniques. It can handle missing data, noisy data, and unbalanced data, and it is less prone to overfitting than other decision tree-based algorithms. In addition to its versatility, Random Forest can offer insights into the importance of different features in predicting the target variable through feature importance measures. These characteristics make Random Forest a potent and popular algorithm in data science and machine learning.

Decision Trees are used for their interpretability, simplicity and ability to handle both numerical and categorical data. Based on the feature values the Decision Tree algorithm iteratively divides the dataset into smaller subsets for functioning. During training, the algorithm learns to predict the target variable by selecting the best option to split the data, based on some criteria like Gini impurity or information gain. Comparing with other machine learning algorithms Decision Tree has several advantages. It is quite easy to visualize and interpret, making it a popular choice for decision-making and data analysis. Decision Tree can also handle both categorical and numerical data, and it is less prone to overfitting than other algorithms, especially when the tree is pruned or regularized.

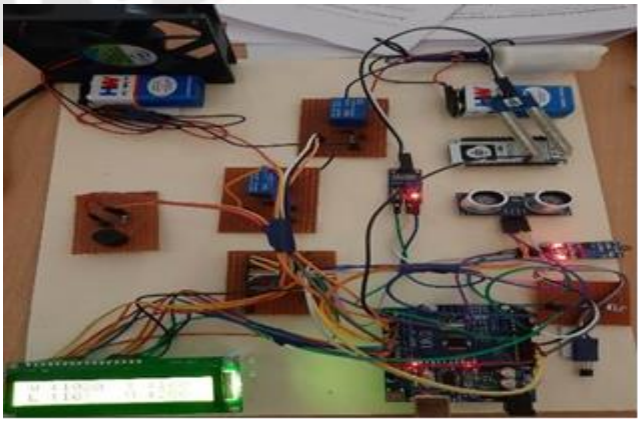
Naive Bayes is one of the efficient and powerful algorithm, so it is widely used in machine learning and data science, especially for text-based classification tasks. However, its assumption of feature independence may not hold in all cases, which can lead to suboptimal performance in some situations.

Compared with other machine learning algorithms Naive Bayes has several advantages. It is simple, fast, and requires a small amount of training data. Naive Bayes can also handle high-dimensional data and is less prone to overfitting than other algorithms.

These algorithms are compared and the algorithm with high accuracy in prediction is used for crop recommendation. In smart aeroponics using IoT and ML Random Forest, Decision Tree and Naive Bayes with accuracies of 83.090%,98.909% and 99.454% is obtained. Then Flask is used to build up web-application for crop recommendation. Python has an API called Flask. The Jinja2 template engine and the WSGI tools form the foundation of Flask.

IV. Result

Using IoT sensors and Machine learning algorithms, the environmental factors of crop are identified and the suitable crop to be cultivated is predicted using algorithms. The output values are displayed on a LCD and stored in Think speak. The cloud data are utilized in machine learning. Three different algorithms are compared based on its accuracy. Random Forest, Decision Tree and Naive Bayes algorithms with accuracies of 99.090%,90.0% and 99.090% are obtained. Naïve Bayes has highest accuracy, hence it is used in web development to predict the suitable crop .



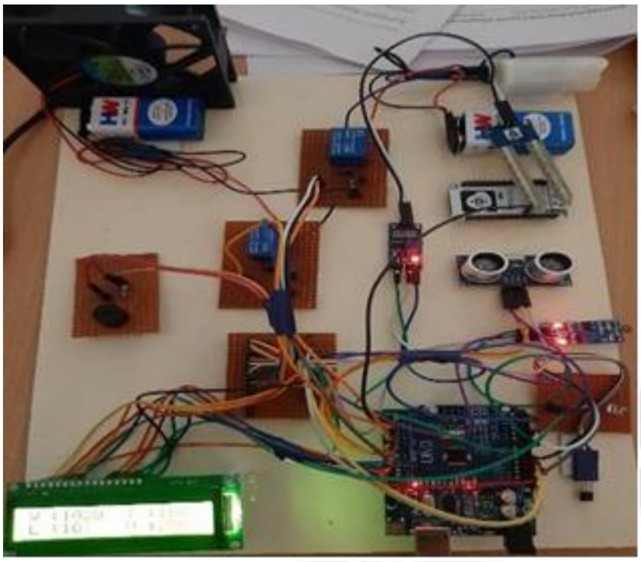


Figure. 4 Hardware implementation

Figure 4 represents the hardware setup with temperature sensor, LDR sensor, Moisture sensor, ultrasonic sensor connected to microcontroller. Relays control the cooler fan, light, water pump. Microcontroller receives the input from the sensors and provides the output to the relay. Figure 5 shows the output of temperature sensor, LDR sensor, Moisture sensor, Ultrasonic sensor displayed in LCD.

Figure. 5 Sensor Output displayed in LCD

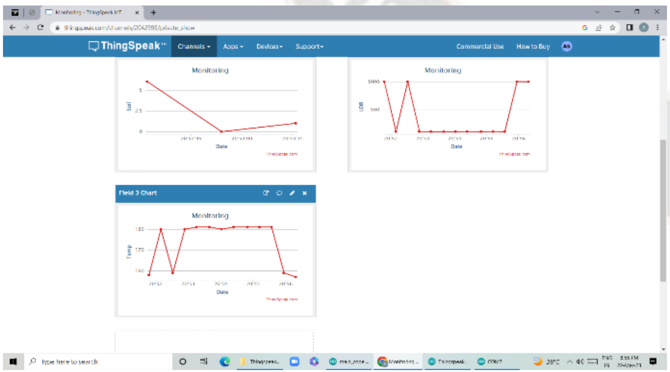


Figure.6 Graphical Representation of data in Thingspeak

Figure 6 represents the data collected from the sensors which are transferred to the cloud using NODE MCU. The data collected from the sensor are represented in a graphical format that shows the value of sensor, date and time.

Table 1 Comparing the accuracy and error rate of Decision tree, Random Forest, Naïve Bayes algorithm

ALGORITHMS	DECISION TREE		RANDOM FOREST		NAÏVE BAYES	
DATASET	TRAINING DATASET	TESTING DATASET	TRAINING DATASET	TESTING DATASET	TRAINING DATASET	TESTING DATASET
ACCURACY	88.1	90.0	99.6	99.0	99.8	99.0
ERROR RATE	10	11.9	0.4	1	0.2	1

According to Table 1, a method for measuring a classification using machine learning algorithm's accuracy is how frequently it classifies a data point correctly. The error rate is determined by dividing the total number of incorrect predictions on the test set by the total number of prediction models on the test set. Here, the Random Forest and Naive bayes algorithm have the highest accuracy.

Table 2 Comparing the precision, recall, f1-score, support of Decision tree, Random Forest, Naïve Bayes algorithm

ALGORITHMS	DECISION TREE	RANDOM FOREST	NAÏVE BAYES
PRECISION	0.86	0.99	0.99
RECALL	0.90	0.99	0.99
F1-SCORE	0.87	0.99	0.99
SUPPORT	440	440	440

From the Table 2 it is observed that the following metrics such as Precision, Recall, F1 Score, Support were analyzed and the results are tabulated.

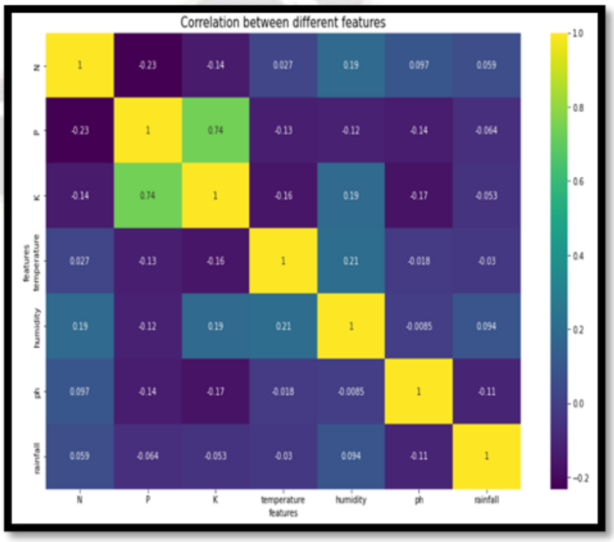


Figure 7 Correlation Matrix

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Figure 7 represents the Correlation matrix between different features such as n,p,k ,temperature, humidity, pH..

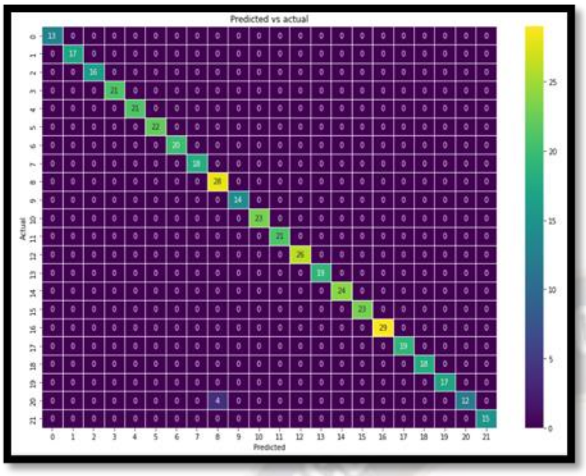


Figure 8 Confusion Matrix for Naïve Bayes algorithm

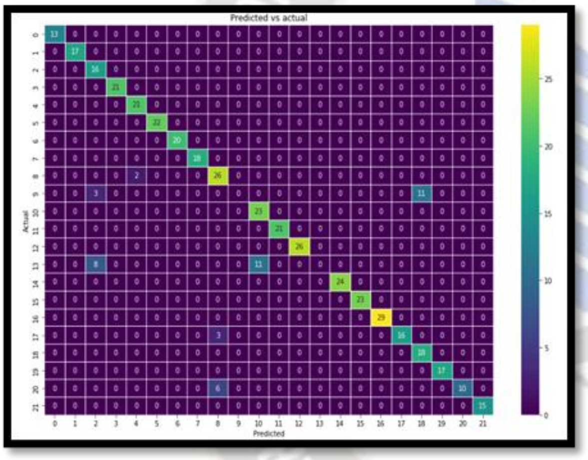


Figure 9 Confusion Matrix for Decision tree algorithm

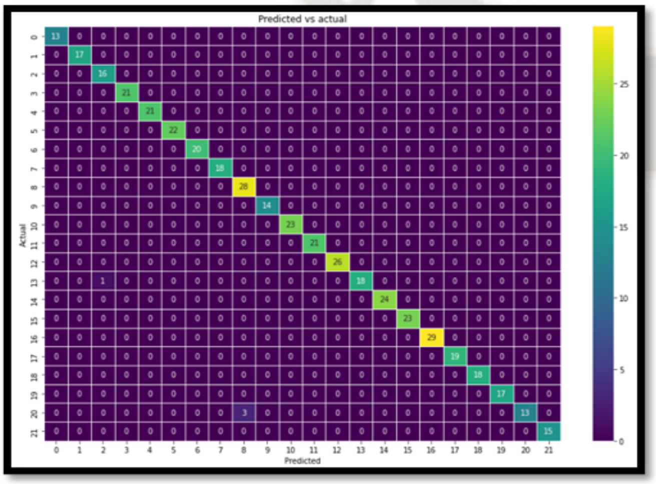


Figure 10 Confusion Matrix for Random Forest algorithm

The Confusion Matrix which gives the True Positive, True Negative, False Positive and False Negative for the Naïve bayes, Random Forest and decision tree model are shown in the figures 8, 9 and 10 where the values along the diagonal are True Positives.

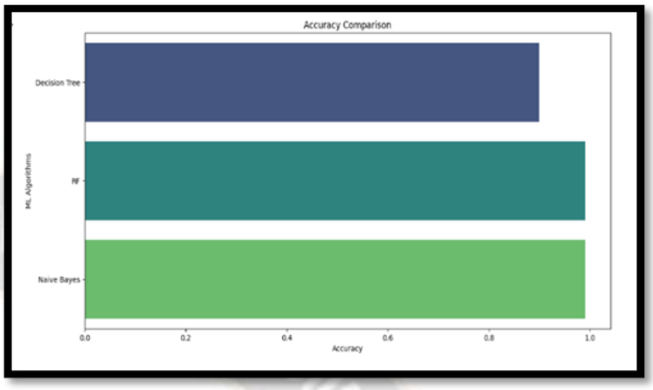


Figure 11 Accuracy comparison

Figure 11 shows the comparison between the accuracies of three different algorithms where Naïve bayes and random forest algorithms have greater accuracies than Decision tree algorithm.

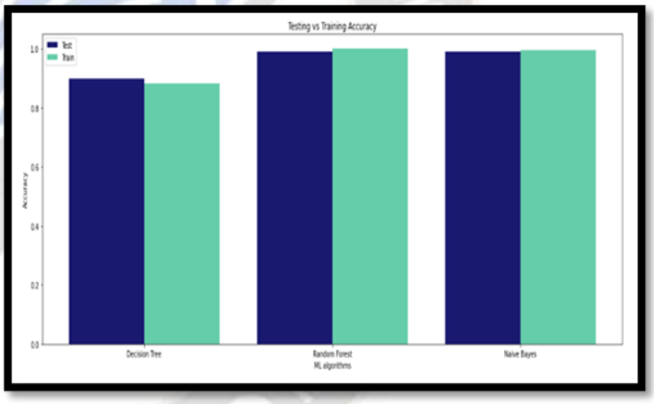


Figure 12 Training vs Testing accuracy

Figure 12 shows the Training and Testing accuracies of Naïve Bayes, Random Forest and Decision algorithms are compared and resulted as a bar chart.

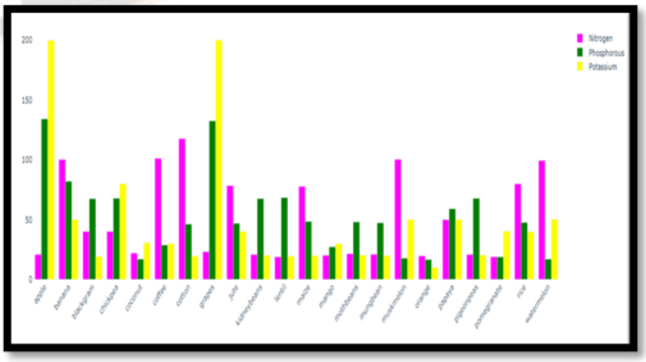


Figure 13 N-P-K Value comparison between crops

The bar graph in figure 13 depicts the N-P-K requirement for various crops. Apple and grapes require higher potassium and phosphorous than others. Cotton requires greater nitrogen level compared to other crops.

Figure.14 Crop prediction page

Figure 14 depicts the web page developed using flask web frameworks for crop suggestion for aeroponics system based on the current environmental factors such as Temperature, Humidity, Dry level, Light compact, Nitrogen, Phosphorous, and Potassium content.

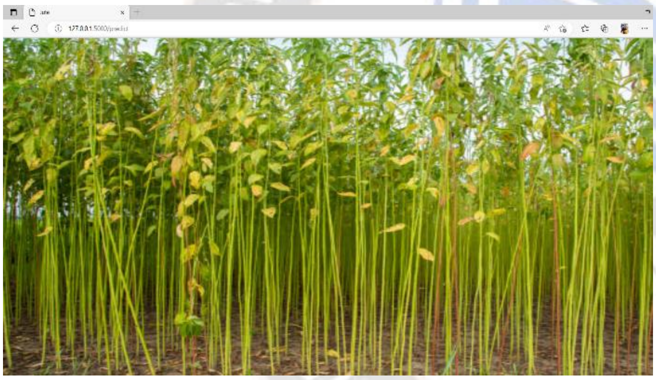


Figure 15 displays the web page that predicts suitable crop based on sensors data and shows the image of the crop.

V. Advantages & Disadvantages

A. Advantages

- Productive and Sustainable
- Air exposure promotes healthy and rapid plant growth
- Minimal usage of land
- Efficient to control the growth environment
- Saving labor
- Less water consumption

B. Disadvantages

- Proper expertise and training are necessary
- Maintenance process is complex.
- High initial capital
- Power outages

VI. Conclusion

Designing an IoT-based aeroponics system for sensing and managing three crucial plant development factors, namely temperature, light, and humidity, has been done step-by-step. The platforms Thing Speak is used to record and effectively visualize all sensor data. After scrapping the data from think speak, the training and testing of prediction algorithms like Random Forest, Naive Bayes and Decision Tree. In this system Random Forest, Decision Tree and Naive Bayes algorithms with accuracies of 99.090%,90.0% and 99.090% are obtained. Then Flask is used to build up web-application for crop recommendation. This system has overcome shortcomings by using less water, being less complicated, and enabling a flexible and ideal shape to sustain an aeroponic condition.

VII. Future Work

One does not need to get their hands filthy, remove grass, uproot plants, or plant anything other than seedlings, aeroponics is sometimes referred to as "white-collar agriculture" and "farming of the future." Aeroponics is the best tool for agriculture in urban spaces. This method can be used to turn building rooftops into farms for growing vegetables. This could make it possible for communities to have more ecologically responsible structures and healthier food farming. In future, the system can be upgraded by adding additional features such as cameras for live monitoring, pH sensor etc.

IX. References

[1] Issac K Joy, Christy V Vazhappilly , "Smart Aeroponics Using IoT", International Research Journal of Engineering and Technology (IRJET) ,2022,Volume 09,Issue 08.

[2] Adache Paul, Dr. Aminat Ajibola, Dr. Emmanuel, "Internet of Things Hydropnics Agriculture Using Web/Mobile Application", International Research Journal of Engineering and Technology (IRJET), 2022, Volume 09, issue 02.

[3] Reddy DJ, Kumar MR. Crop yield prediction using machine learning algorithm. In2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) 2021 May 6 (pp. 1466-1470). IEEE.

[4] Shahhosseini M, Hu G, Huber I, Archontoulis SV. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific reports. 2021 Jan 15;11(1):1-5.

[5] Iwan Fitrianto Rahmad, Lili Tanti , Ratih Puspasari , Evri Ekadiansyah , Vidi Agung Fragastia, "Automatic Monitoring and Control System in Aeroponic Plant Agriculture "The 8th International Conference on Cyber and IT Service Management, 2020.

[6] Lischeid G, Webber H, Sommer M, Nendel C, Ewert F. Machine learning in crop yield modelling: A powerful tool, but no surrogate for science. Agricultural and Forest Meteorology. 2022 Jan 15;312:108698.

[7] Kang Y, Ozdogan M, Zhu X, Ye Z, Hain C, Anderson M. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. Environmental Research Letters. 2020 May 19;15(6):064005.
[8] Crane-Droesch A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environmental Research Letters. 2018 Oct 26;13(11):114003.
[9] W. Tao, L. Zhao, G. Wang, and R. Liang, "Review of the internet of things communication technologies in smart agriculture and challenges," Computers and Electronics in Agriculture, vol. 189, article 106352, 2021.
[10] M. O. H. D. B. K. Putra and N. Nopriadi, "IOT based SMART agriculture using fuzzy logic," Computer and Science Industrial Engineering (COMASIE), vol. 6, no. 2, pp. 52–61, 2022.
[11] E. Nigussie, T. Olwal, G. Musumba, T. Tegegne, A. Lemma, and F. Mekuria, "IoT-based irrigation management for smallholder farmers in rural sub-Saharan Africa," Procedia Computer Science..
[12] H. Pang, Z. Zheng, T. Zhen, and A. Sharma, "Smart Farming," International Journal of Agricultural and Environmental Information Systems (IJAEIS), vol. 12, no. 1, pp. 55–67, 2021.
[13] F. Sabrina, S. Sohail, F. Farid, S. Jahan, F. Ahamed, and S. Gordon, "An interpretable artificial intelligence based smart agriculture system," Computers, Materials & Continua, vol. 72, no. 2, pp. 3777–3797, 2022.
[14] V. Bhatnagar and R. C. Poonia, ""sustainable development in agriculture: past and present scenario of Indian agriculture." research anthology on strategies for achieving agricultural sustainability," IGI Global, pp. 1342–1364, 2022.
[15] Arunraja A & Jayanthy S 2022, 'Novel super junction technique used in AlGaN/GaN HEMT for high power applications'. Materials Research Express, vol. 9, no. 7, Article : 075901, 2022.
[16] N. G. Rezk, E. E.-D. Hemdan, A.-F. Attia, A. El-Sayed, and M. A. El-Rashidy, "An efficient IoT based smart farming system using machine learning algorithms," Multimedia Tools and Applications, vol. 80, no. 1, pp. 773–797, 2021.