Crop Yield Prediction Using IoT and Machine Learning

Submitted as a partial fulfillment of Bachelor of Technology in Computer Science & Engineering
of
Maulana Abul Kalam Azad University of Technology
(Formerly known as West Bengal University of Technology)

Project Report

Submitted by

Name of Students	**University Roll No.**
SHAKYAJYOTI PYNE	**11600118029**
JIT SAHA	**11600118054**
SRIJA TALUKDAR	**11600118018**
ARKA BHOWMIK	**11600118064**
SHAYARI MONDAL	**11600118028**

Under the supervision of

Ms. Moumita Goswami
Assistant Professor, Dept. of Computer Science and Engineering

Department of Computer Science & Engineering,
MCKV Institute of Engineering
243, G.T. Road(N)
Liluah, Howrah - 711204

Department of Computer Science & Engineering
MCKV Institute of Engineering
243, G. T. Road (N), Liluah,
Howrah-711204

CERTIFICATE OF RECOMMENDATION

I hereby recommend that the thesis prepared under my supervision by **Shakyajyoti Pyne, Jit Saha, Srija Talukdar, Arka Bhowmik, Shayari Mondal** entitled **Crop Yield Prediction Using IoT and Machine learning** be accepted in partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Science & Engineering Department.

Mr. Avijit Bose
Assistant Professor & Head of the Department,
Computer Science & Engineering Department.
MCKV Institute of Engineering,Howrah

Ms. Moumita Goswami
Assistant Professor.
Computer Science & Engineering Department
MCKV Institute of Engineering, Howrah

MCKV Institute of Engineering
243, G. T. Road (N), Liluah,
Howrah-711204

Affiliated to

Maulana Abul Kalam Azad University of Technology
(Formerly known as West Bengal University of Technology)

CERTIFICATE

This is to certify that the project entitled **"Crop Yield Recommendation Using IoT and Machine Learning"** and submitted by

Name of students	University Roll No.
SHAKYAJYOTI PYNE	11600118029
JIT SAHA	11600118054
SRIJA TALUKDAR	11600118018
ARKA BHOWMIK	11600118064
SHAYARI MONDAL	11600118028

has been carried out under the guidance of myself following the rules and regulations of the degree of Bachelor of Technology in Computer Science & Engineering of **Maulana Abul Kalam Azad University of Technology (**Formerly West Bengal University of Technology).

____

(Signature of the project guide)
Ms. Moumita Goswami
Assistant Professor
Department of Computer Science and Engineering

Shakyajyoti Pyne
Jit Saha
Srija Talukdar
Arka Bhowmik
Shayari Mondal

CERTIFICATE OF APPROVAL

(B.Tech Degree in Computer Science & Engineering)

This project report is hereby approved as a creditable study of an engineering subject carried out and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is to be understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed and conclusion drawn therein but approve the project report only for the purpose for which it has been submitted

COMMITTEE ON FINAL EXAMINATION FOR EVALUATION OF PROJECT REPORT	1. ______________________
	2. ______________________
	3. ______________________
	4. ______________________
	5. ______________________

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to MCKV Institute of Engineering for supporting us throughout our B.Tech curriculum.

We would take the opportunity to express our deep sense of gratitude to Mr. Avijit Bose, Assistant Professor and Head of the Department, Computer Science and Engineering, MCKV Institute of Engineering for supporting us to make our project worth it.

We would like to express our sincere gratitude to our project guide, Ms.Moumita Goswami who gave us the golden opportunity to do this wonderful project on the topic **“Crop Yield Prediction Using IoT and Machine learning”** which has helped us in doing a lot of research and we came to know about so many new things.

We are also thankful to Mr.Abhisek Saha, the Project Coordinator for providing and clarifying the administrative formalities related to project proceedings.

We would also like to thank the Computer Science and Engineering Department of MCKV Institute of Engineering for their immense support, contributions and valuable instructions in conducting this project report successfully. Their guidance has helped us tremendously at all times in our research and writing of this project. Their immense knowledge, profound experience and professional expertise has enabled us to complete this project successfully. Without their support and guidance, this project would not have been possible. It helped us in increasing our knowledge and skills.

We thank all our other faculty members and technical assistants at MCKV Institute of Engineering for playing a significant role during the development of the project. Last but not the least; we thank all our friends for their cooperation and encouragement.

1. Jit Saha
2. Arka Bhowmik
3. Shakyajyoti Pyne
4. Srija Talukdar
5. Shayari Mondal

TABLE OF CONTENTS

Project Guide: Ms. Moumita Goswami

Researcher: Jit Saha, Arka Bhowmik, Srija Talukdar, Shakyajyoti Pyne, Shayari Mondal

Presentation title: Crop yield prediction using IoT and Machine Learning

Research Focus: Internet of Things (IoT)

College: MCKV Institute of Engineering

Presentation Type: PowerPoint Presentation and Public Speaking

ABSTRACT

Agriculture is the primary factor that is important for the survival and the economy in India. It is important to increase the productivity of agricultural and farming processes to improve crops and cost-effectiveness. Crop Yield Prediction using IoT and Machine Learning is a management concept focused on providing the agricultural industry with the infrastructure to leverage advanced technology including big data, cloud, and the internet of things (IoT) – for tracking, monitoring, automating and analyzing operations. Farming can be made more productive and accurate with the help of technologies like IoT. IoT can be used in different domains of farming or agriculture. The implementation area of IoT is wide and can be implemented in every field. The Internet of things (IoT) is a promising technology that provides efficient and reliable solutions for the modernization of several domains. IoT-based solutions are being developed to automatically maintain and monitor agricultural farms with minimal human involvement. Furthermore, IoT-based agriculture systems' connection with relevant technologies including cloud computing, machine learning, and analytics makes it more vibrant. In this paper, we have developed the system, that will help us to collect the relevant data from the different agricultural fields with the help of the sensors (i.e. Dht-11 sensor, Soil-moisture-water sensor, Rainfall sensor). Then this collected data is uploaded to a cloud server and the uploaded data is analyzed with the dataset available by using the KNN machine learning algorithm and then the suitable result is displayed on the user's device.

.

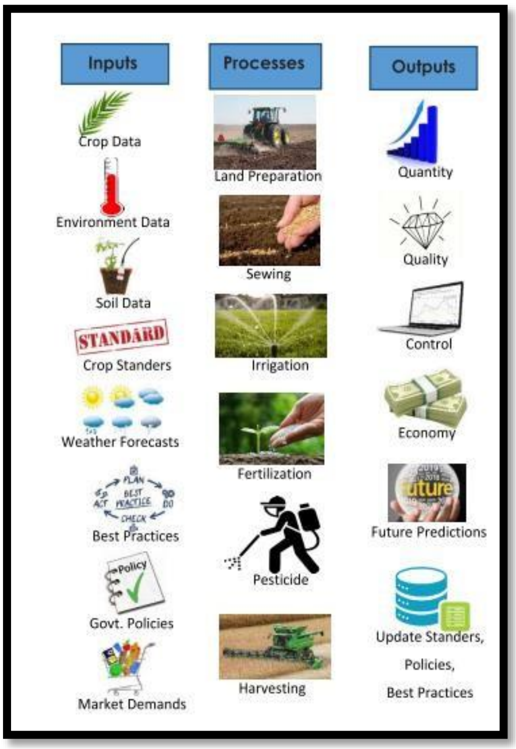
1. INTRODUCTION

To improve the agricultural yield with fewer resources and labor efforts, substantial innovations have been made throughout human history. Achieving maximum crop yield at minimum cost is one of the goals of agricultural production. Early detection and management of problems associated with crop yield indicators can help increase yield and subsequent profit [1]. Nevertheless, the high population rate never let the demand and supply match during all these times. According to the forecasted figures, in 2050, the world population is expected to touch 9.8 billion, an increase of approximately 25% from the current figure. Almost the entire mentioned rise of population is forecasted to occur among the developing countries.

On the other side, the trend of urbanization is forecasted to continue at an accelerated pace, with about 70% of the world's population predicted to be urban until 2050 (currently 49%).

Furthermore, income levels will be multiples of what they are now, which will drive the food demand further, especially in developing countries. As a result, these nations will be more careful about their diet and food quality; hence, consumer preferences can move from wheat and grains to legumes and, later, to meat. In order to feed this larger, more urban, and richer population, food production should double by 2050. Particularly, the current figure of 2.1 billion tons of annual cereal production should touch approximately 3 billion tons. Not only for food, but crop production is becoming equally critical for industry; indeed crops like cotton, rubber, and gum are playing important roles in the economy of many nations. Furthermore, the food-crops-based bioenergy market started to increase recently. Even before a decade, only the production of ethanol utilized 110 million tons of coarse grains (approximately 10% of the world production). Due to the rising utilization of food crops for biofuel production, bio-energy, and other industrial usages, food security is at stake. These demands are resulting in a further increase of the pressure on already scarce agricultural resources, resulting in immediate improvement in the agricultural yield. In this paper, we have developed an arduino device which will detect the necessary climatic conditions like - soil moisture, rainfall, humidity and temperature using sensors like DHT11 and Rainfall sensor, and will upload the readings to the Arduino ThinkSpeak cloud on one hand and will use the machine learning model which we have created using KNN algorithm, on the other hand to detect which crop is best suited for the particular land and will display that crop to the farmer in his mobile phone using bluetooth.

1.1 Major applications of IoT in Agriculture



By implementing the latest sensing and IoT technologies in agriculture practices, every aspect of traditional farming methods can be fundamentally changed. Currently, seamless integration of wireless sensors and the IoT in smart agriculture can raise agriculture to levels which were previously unimaginable. By following the practices of smart agriculture, IoT can help to improve the solutions of many traditional farming issues, like drought response, yield optimization, land suitability, irrigation, and pest control. While, major instances in which the advanced technologies are helping at various stages to enhance overall efficiency.

1.2 Benefits of Agriculture using IoT

- **Increased Production**

Optimized crop treatment such as accurate planting, watering, pesticide application and harvesting directly affects production rates.

- **Water Conservation**

Weather predictions and soil moisture sensors allow for water use only when and where needed.

- **Real-Time Data and Production Insight**

Farmers can visualize production levels, soil moisture, sunlight intensity and more in real time and remotely to accelerate decision making process.

- **Lowered Operation Costs**

Automating processes in planting, treatment and harvesting can reduce resource consumption, human error and overall cost.

- **Increased Quality of Production**

Analyzing production quality and results in correlation to treatment can teach farmers to adjust processes to increase quality of the product.

- **Accurate Farm and Field Evaluation**

Accurately tracking production rates by field over time allows for detailed predicting of future crop yield and value of a farm.

- **Improved Livestock Farming**

Sensors and machines can be used to detect reproduction and health events earlier in animals. Geofencing location tracking can also improve livestock monitoring and management.

- **Reduced Environmental Footprint**

All conservation efforts such as water usage and increased production per land unit directly affect the environmental footprint positively.

- **Remote Monitoring**

Local and commercial farmers can monitor multiple fields in multiple locations around the globe from an internet connection. Decisions can be made in real-time and from anywhere.

- **Equipment Monitoring**

Farming equipment can be monitored and maintained according to production rates, labor effectiveness and failure prediction.

1.3 Shortfalls of Agriculture using IoT

- The smart agriculture needs availability of internet continuously. Rural part of most of the developing countries do not fulfil this requirement. Moreover the internet connection is slower.
- The smart farming-based equipment require farmers to understand and learn the use of technology. This is major challenge in adopting smart agriculture farming at large scale across the countries.

1.4 Objective

- The goal of smart agriculture research is to ground a decision-making support system for farm management.
- To recommend crop based on the soil as well as climatic conditions.

- To make a system that optimises and examines how high-tech farming can aid the production output as well as focuses on the preservation of resources
- By providing them with the benefits of technological advancements, smart agriculture aims to reduce the heavy workload of the farm workers, hence improving their quality of life.
- Smart farming deems it necessary to address the issues of population growth, climate change and labour that has gained a lot of technological attention, from planting and watering of crops to health and harvesting.

2. COMPONENTS USED IN THE PROPOSED PRODUCT

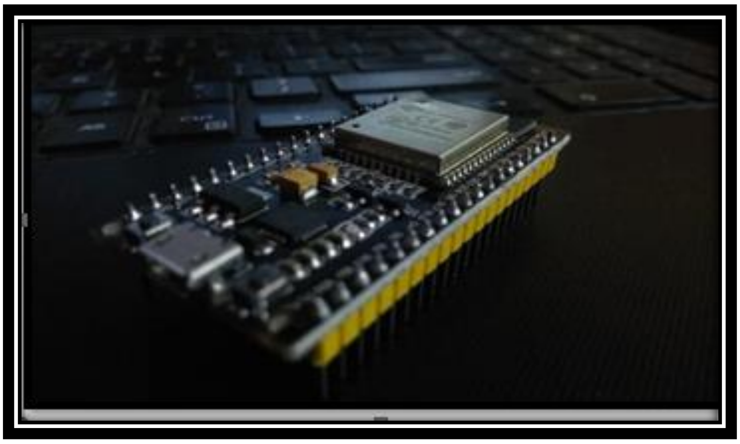
2.1 Definition of our Product

This device monitors the farm or greenhouse and based upon the readings of different kind of sensors like temperature, humidity, soil moisture, UV, IR, soil nutrients and gives different types of messages to the farmer about the present conditions so that the farmer can take quick action. The quick actions taken by the farmers will help them increase the productivity in their farming and proper use of natural resources will be done, which will make our product environment friendly also. Our product will increase the quantity and quality of the crops by properly monitoring the various present conditions. It is an IoT device with the concept of “Plug and Sense”. Live data for different parameters can be seen on Laptop and Smart Phones.

2.2 Different Components Used

Serial Number	Name of the Component
01	ESP32s Node MCU
02	Breadboard
03	DHT11 Temperature and Humidity Sensor
04	Jumper Wire
05	Rainfall Sensor
06	Soil-moisture-Water Sensor

2.1.1 ESP32s Node MCU

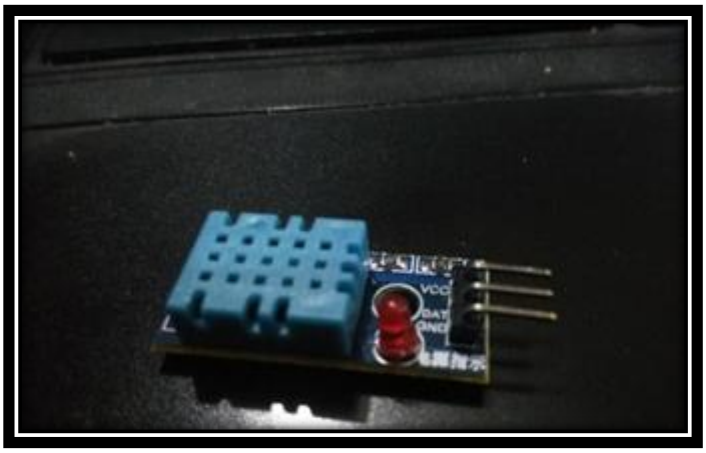


The core of this module is the ESP32 chip, which is scalable and adaptive. Two CPU cores can be individually controlled. The clock frequency is adjustable from 80 MHz to 240 MHz and supports RTOS. It is a general- purpose Wi-Fi+BT+BLE MCU module.ESP-WROOM-32s.The module integrates traditional Bluetooth, Bluetooth low energy and Wi-Fi. Wide range of uses: Wi-Fi supports a wide range of communication connections, as well as direct connection to the Internet via a router; Bluetooth allows users to connect to a mobile phone or broadcast a BLE Beacon for signal detection. The module supports data rates up to 150 Mbps and antenna output power of 20 dBm for maximum wireless communication. As a result, this module has industry-leading specifications and performs well in terms of high integration, wireless transmission distance, power consumption, and network connectivity.

2.2.2 Breadboard

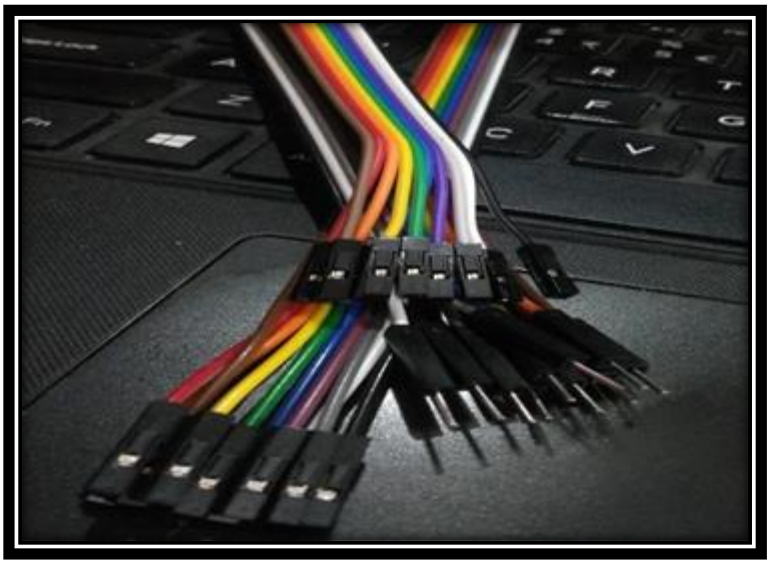
A breadboard is a rectangular plastic board with a bunch of tiny holes in it. These holes let you easily insert electronic components to prototype (meaning to build and test an early version of) an electronic circuit, like this one with a battery, switch, resistor, and an LED.

2.2.3 DHT11 Temperature and Humidity Sensor



The DHT11 is a basic, ultralow- cost digital temperature and humidity sensor. It uses a capacitive humidity sensor and a thermistor to measure the surrounding air, and spits out a digital signal on the data pin (no analog input pins needed).

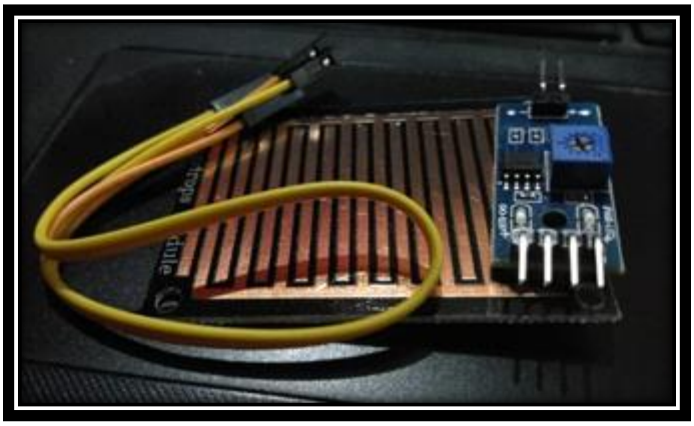
2.2.4 Jumper Wire



A jump wire (also known as jumper, jumper wire, jumper cable, DuPont wire or cable) is an electrical wire, or group of them in a cable, with a connector or pin at each end (or sometimes without them – simply "tinned"), which is normally used to interconnect the components of a breadboard or other prototype or test circuit, internally or with other equipment or components, without soldering.

Individual jump wires are fitted by inserting their "end connectors" into the slots provided in a breadboard, the header connector of a circuit board, or a piece of test equipment

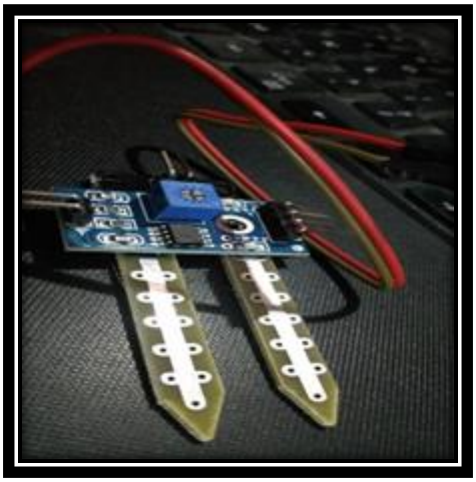
2.2.5 Rainfall Sensor



A rain sensor is one kind of switching device which is used to detect the rainfall. It works like a switch and the working principle of this sensor is, whenever there is rain, the switch will be normally closed. This sensor includes four pins which include the following.

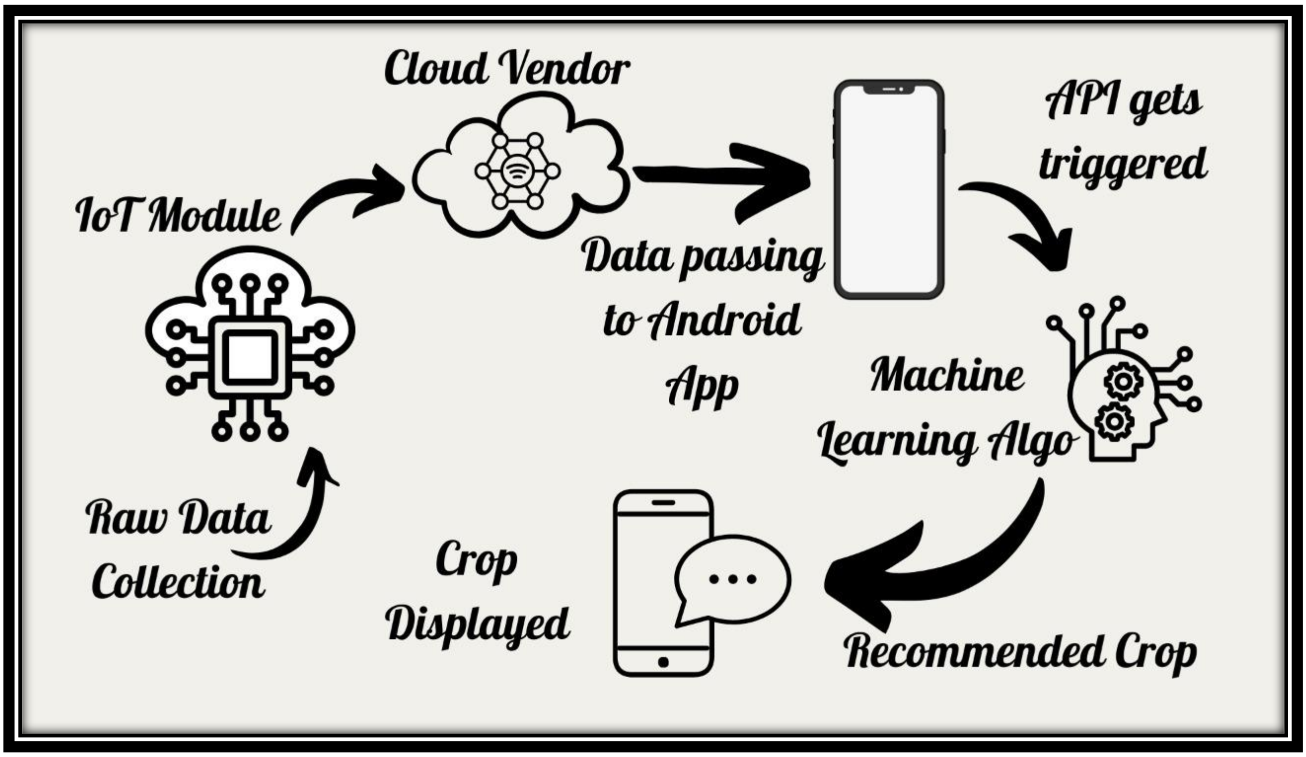
- Pin1 (VCC): It is a 5V DC pin
- Pin2 (GND): it is a GND (ground) pin
- Pin3 (DO): It is a low/ high output pin
- Pin4 (AO): It is an analog output pin

2.2.6 Soil-Moisture-Water Sensor

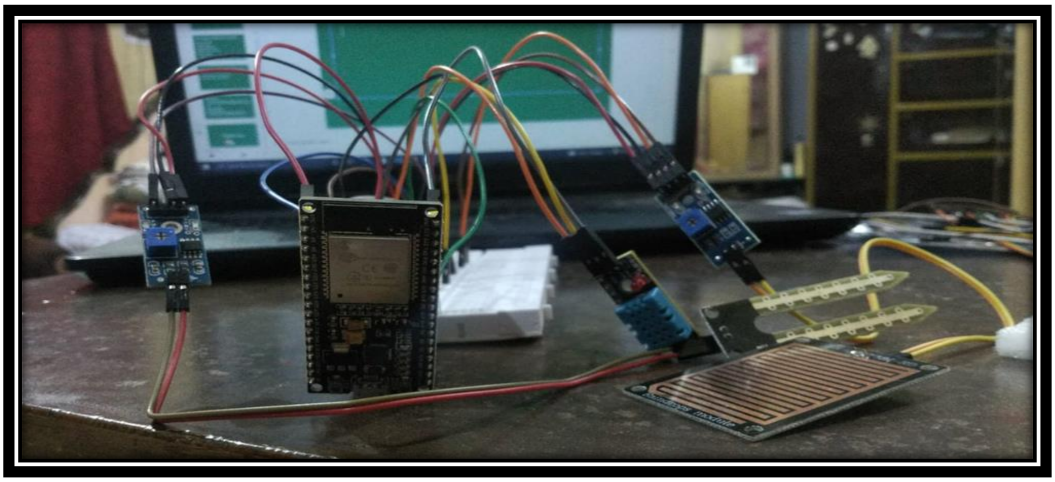


Soil moisture sensors typically refer to sensors that estimate volumetric water content. Another class of sensors measure another property of moisture in soils called water potential; these sensors are usually referred to as soil water potential sensors and include tensiometers and gypsum blocks.

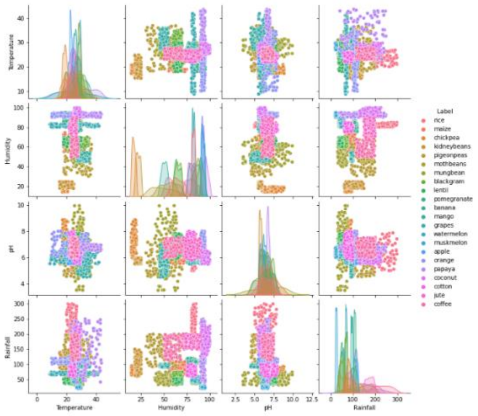
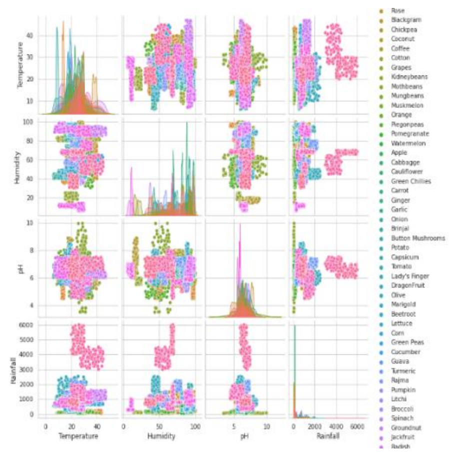
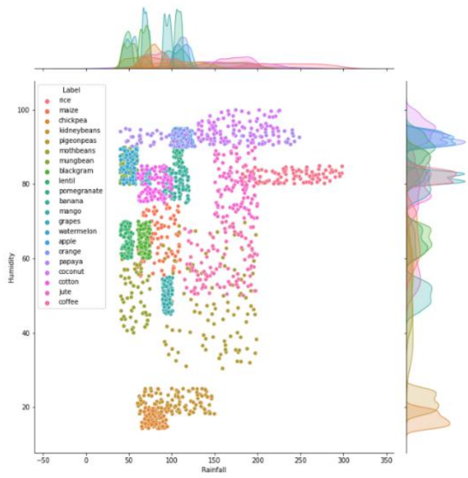
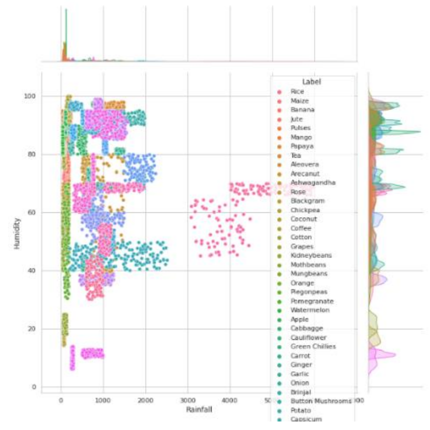
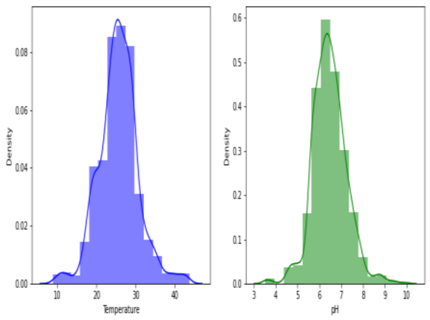
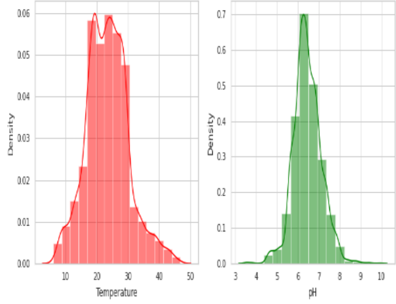
2.3 Workflow Model

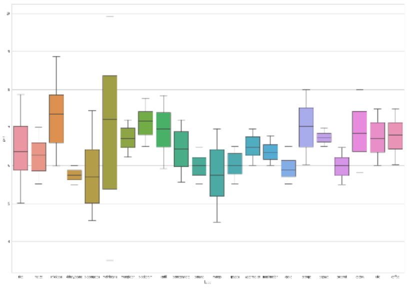
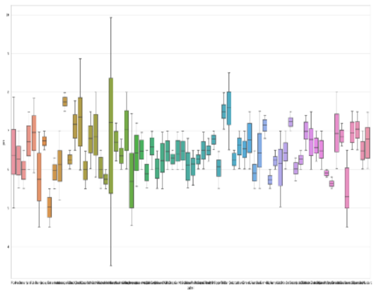
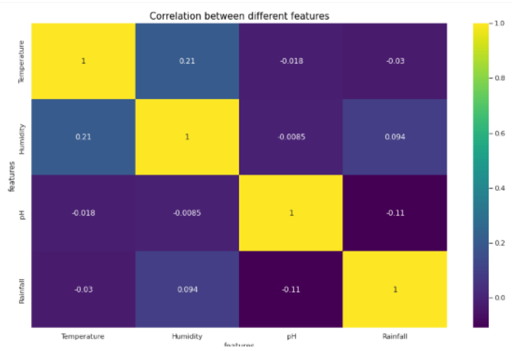
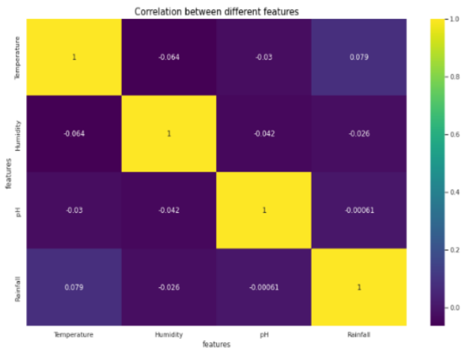


This device monitors the farm or greenhouse and is based upon the readings of different kinds of sensors like temperature, humidity, soil moisture, and gives different types of messages to the farmer about the present conditions so that the farmer can take proper decisions in selecting appropriate crops. Next, the Classification algorithm is used for crop prediction. In the workflow diagram, as given above, the raw data, that is, the climatic conditions like soil moisture, humidity, temperature and rainfall are collected from the fields by the IoT module using the sensors including rainfall and DHT11 sensors and the data is then uploaded to the Arduino and the data is then passed onto the ML model created using KNN algorithm after which the data is analyzed and the recommended crop is shown to the farmer in the mobile phones via push message.



3. <u>DATA COLLECTION:</u>

Category	Old Dataset	New Dataset
Size	The dataset consists of 2200 rows and 5 columns	The dataset consists of 7000 rows and 5 columns
Number of Crops	It consists of 22 different crops under 100 different conditions.	It consists of 70 different crops under 100 different conditions.
Pair plot of Dataset		
Joint plot of Dataset		
Distance plot of Dataset		

<table>
<tr>
<td>Box plot of Dataset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Columns in the Dataset</td>
<td>Temperature, Rainfall, Humidity, pH and Label</td>
<td>Temperature, Rainfall, Humidity, pH and Label</td>
</tr>
<tr>
<td>Heatmap of Dataset</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Crops in the Dataset</td>
<td>

```
df['Label'].unique()
```

```
array(['rice', 'maize', 'chickpea', 'kidneybeans', 'pigeonpeas',
       'mothbeans', 'mungbean', 'blackgram', 'lentil', 'pomegranate',
       'banana', 'mango', 'grapes', 'watermelon', 'muskmelon', 'apple',
       'orange', 'papaya', 'coconut', 'cotton', 'jute', 'coffee'],
      dtype=object)
```

</td>
<td>

```
df['Label'].unique()
```

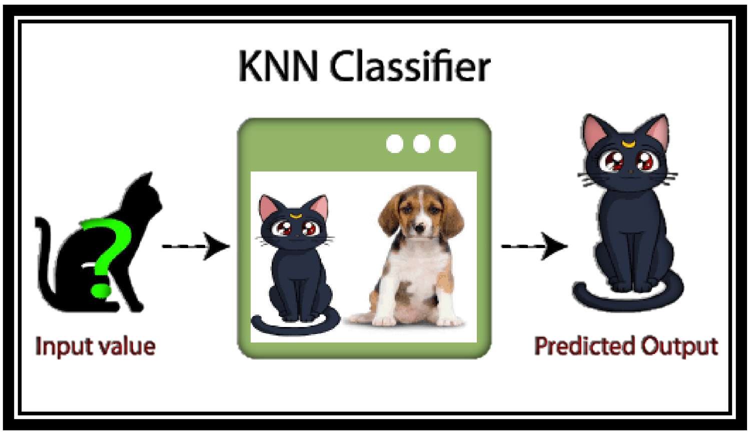
```
array(['Rice', 'Maize', 'Banana', 'Jute', 'Pulses', 'Mango', 'Papaya',
       'Tea', 'Aleovera', 'Arecanut', 'Ashwagandha', 'Rose', 'Blackgram',
       'Chickpea', 'Coconut', 'Coffee', 'Cotton', 'Grapes', 'Kidneybeans',
       'Mothbeans', 'Mungbeans', 'Muskmelon', 'Orange', 'Piegonpeas',
       'Pomegranate', 'Watermelon', 'Apple', 'Cabbagge', 'Cauliflower',
       'Green Chillies', 'Carrot', 'Ginger', 'Garlic', 'Onion', 'Brinjal',
       'Button Mushrooms', 'Potato', 'Capsicum', 'Tomato',
       "Lady's Finger", 'DragonFruit', 'Olive', 'Marigold', 'Beetroot',
       'Lettuce', 'Corn', 'Green Peas', 'Cucumber', 'Guava', 'Turmeric',
       'Rajma', 'Pumpkin', 'Litchi', 'Broccoli', 'Spinach', 'Groundnut',
       'Jackfruit', 'Radish', 'Chinese Cabbage', 'Drumstick', 'Soybean',
       'Sweet Potato', 'Poppy Seeds', 'Coriander', 'Walnuts',
       'Cashewnuts', 'French Beans', 'Sugarcane', 'Bajra', 'Mustard'],
      dtype=object)
```

</td>
</tr>
<tr>
<td>Description of the Dataset</td>
<td>
<table>
<tr><th></th><th>Temperature</th><th>Humidity</th><th>pH</th><th>Rainfall</th></tr>
<tr><td>count</td><td>2200.000000</td><td>2200.000000</td><td>2200.000000</td><td>2200.000000</td></tr>
<tr><td>mean</td><td>25.616244</td><td>71.481779</td><td>6.469480</td><td>103.463655</td></tr>
<tr><td>std</td><td>5.063749</td><td>22.263812</td><td>0.773938</td><td>54.958389</td></tr>
<tr><td>min</td><td>8.825675</td><td>14.258040</td><td>3.504752</td><td>20.211267</td></tr>
<tr><td>25%</td><td>22.769375</td><td>60.261953</td><td>5.971693</td><td>64.551686</td></tr>
<tr><td>50%</td><td>25.598693</td><td>80.473146</td><td>6.425045</td><td>94.867624</td></tr>
<tr><td>75%</td><td>28.561654</td><td>89.948771</td><td>6.923643</td><td>124.267508</td></tr>
<tr><td>max</td><td>43.675493</td><td>99.981876</td><td>9.935091</td><td>298.560117</td></tr>
</table>
</td>
<td>
<table>
<tr><th></th><th>Temperature</th><th>Humidity</th><th>pH</th><th>Rainfall</th></tr>
<tr><td>count</td><td>7000.000000</td><td>7000.000000</td><td>7000.000000</td><td>7000.000000</td></tr>
<tr><td>mean</td><td>23.493807</td><td>71.318934</td><td>6.454138</td><td>751.476229</td></tr>
<tr><td>std</td><td>6.755644</td><td>22.289074</td><td>0.667650</td><td>825.472175</td></tr>
<tr><td>min</td><td>6.105382</td><td>6.029440</td><td>3.504752</td><td>20.211267</td></tr>
<tr><td>25%</td><td>18.948438</td><td>58.055782</td><td>6.049754</td><td>123.595859</td></tr>
<tr><td>50%</td><td>23.344269</td><td>77.238068</td><td>6.406537</td><td>644.904867</td></tr>
<tr><td>75%</td><td>27.686253</td><td>90.259315</td><td>6.859321</td><td>1051.049795</td></tr>
<tr><td>max</td><td>46.791488</td><td>99.981876</td><td>9.935091</td><td>5989.995521</td></tr>
</table>
</td>
</tr>
</table>

4. ALGORITHM USED:

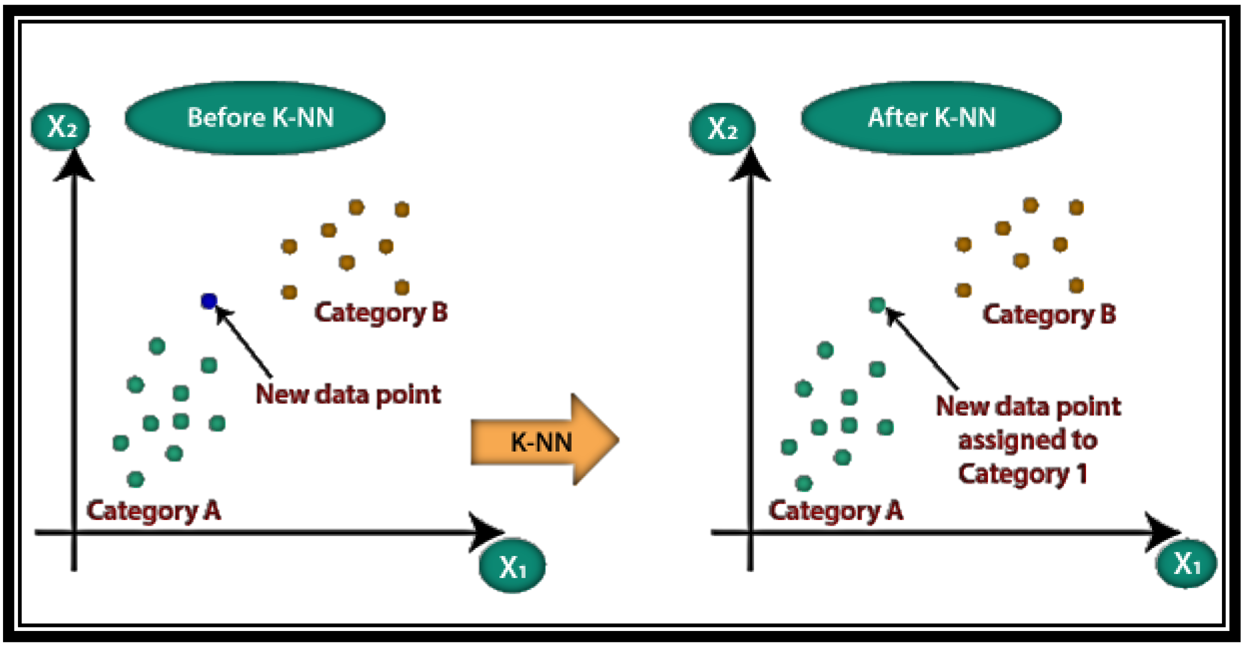
4.1 K-nearest neighbors (KNN):

- K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.
- It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.
- Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.



4.2 Why do we need a K-NN Algorithm?

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:

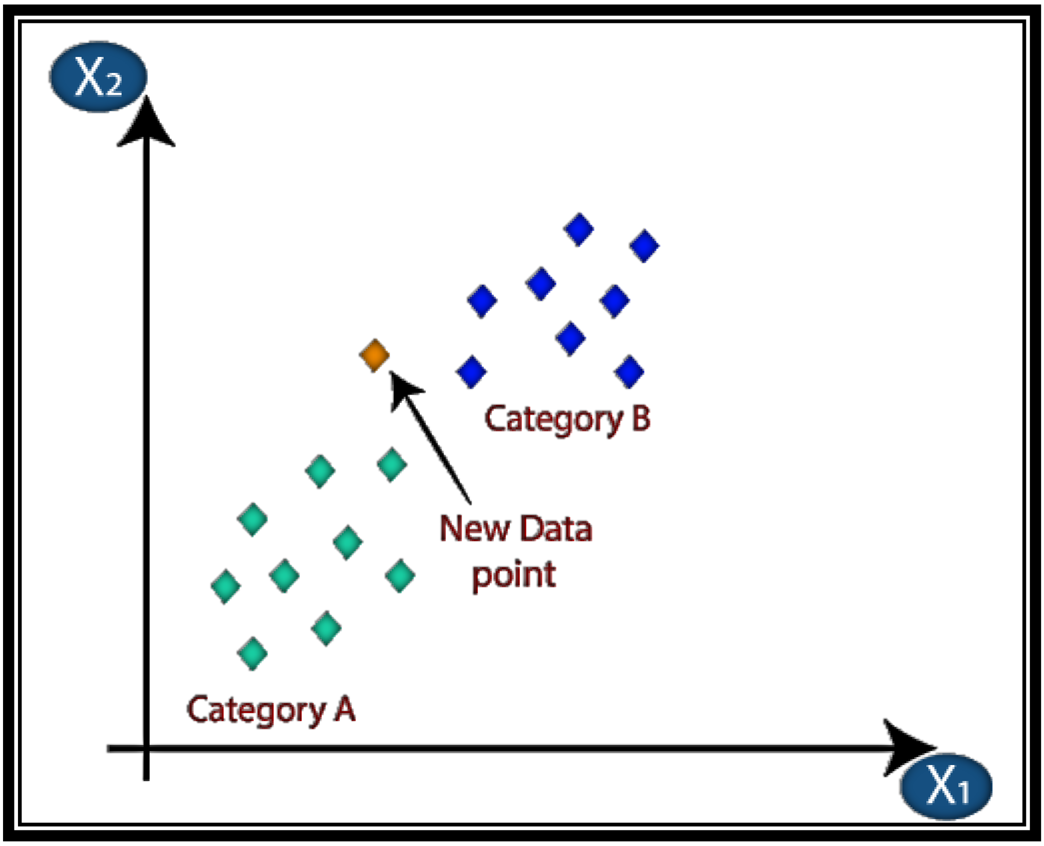


4.3 How does K-NN work?

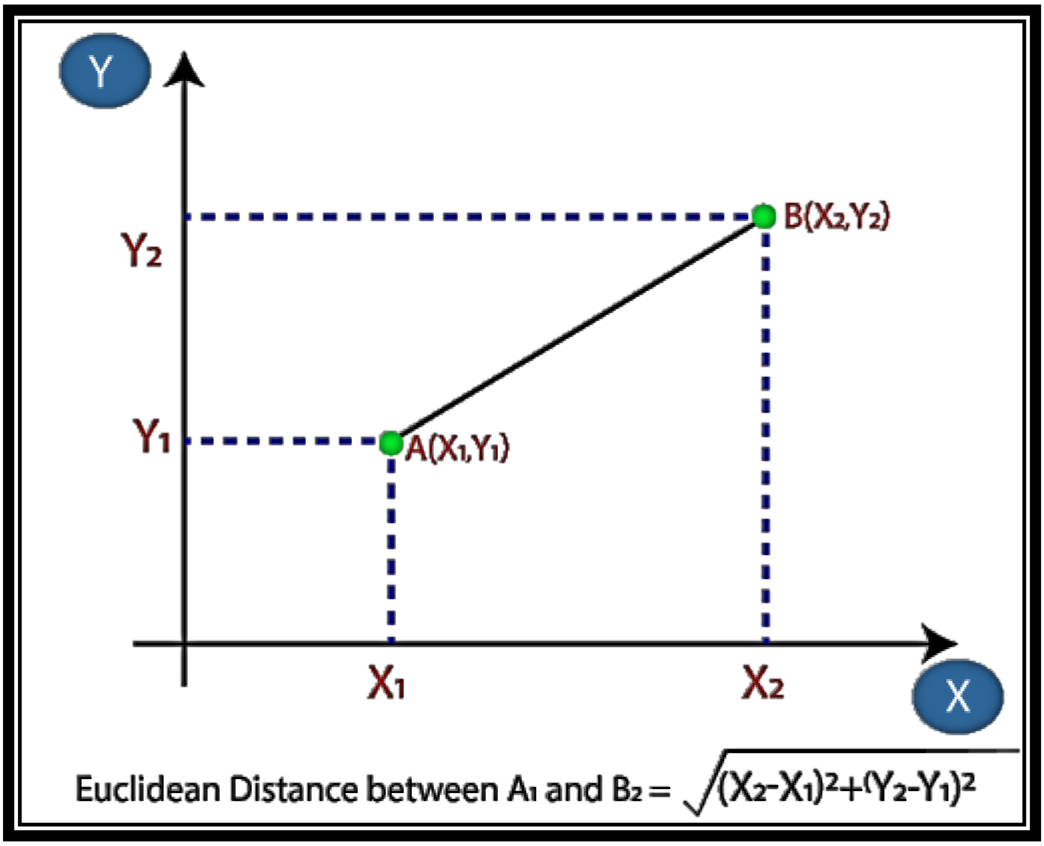
The K-NN working can be explained on the basis of the below algorithm:

- Step-1: Select the number K of the neighbors
- Step-2: Calculate the Euclidean distance of K number of neighbors
- Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
- Step-4: Among these k neighbors, count the number of the data points in each category.
- Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
- Step-6: Our model is ready.

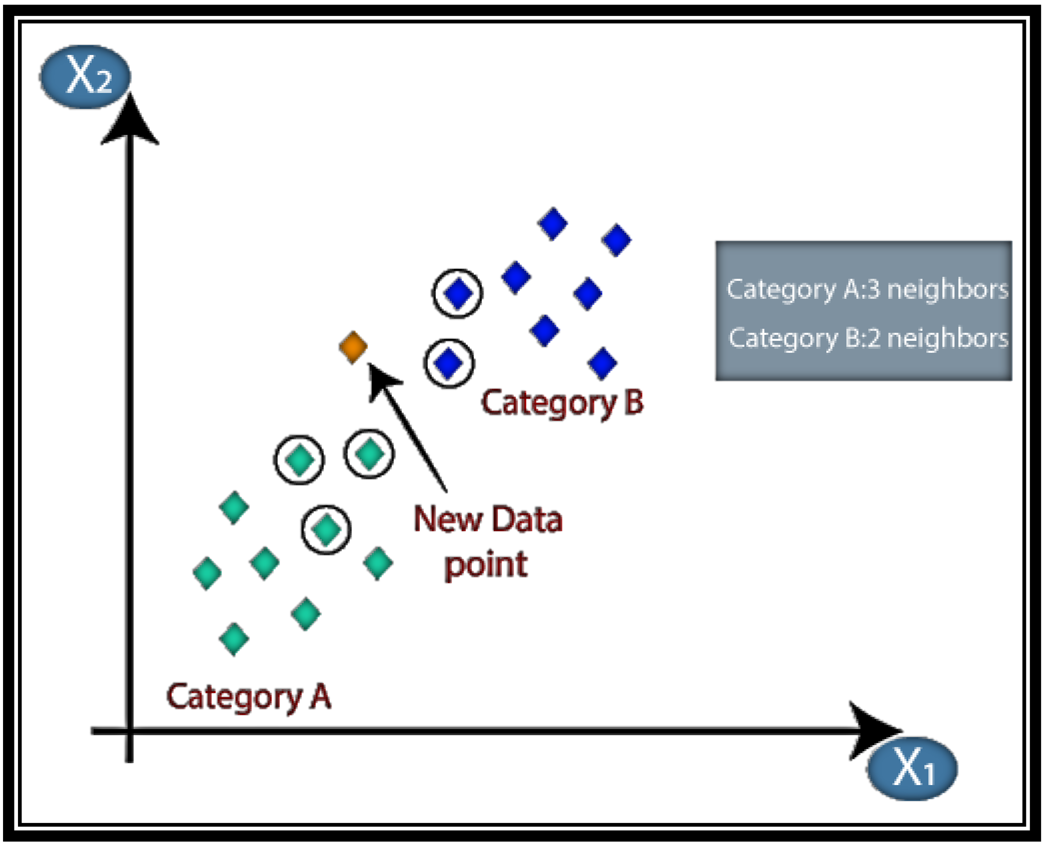
Suppose we have a new data point and we need to put it in the required category. Consider the below image:



- Firstly, we will choose the number of neighbors, so we will choose the k=5.
- Next, we will calculate the Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as:



- By calculating the Euclidean distance we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in category B. Consider the below image:



- As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.

4.4 How to select the value of K in the K-NN Algorithm?

Below are some points to remember while selecting the value of K in the K-NN algorithm:

- There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them. The most preferred value for K is 5.
- A very low value for K such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model.
- Large values for K are good, but it may find some difficulties.

4.5 Advantages of KNN Algorithm:

- It is simple to implement.
- It is robust to the noisy training data
- It can be more effective if the training data is large.

4.6 Disadvantages of KNN Algorithm:

- Always needs to determine the value of K which may be complex some time.
- The computation cost is high because of calculating the distance between the data points for all the training samples.

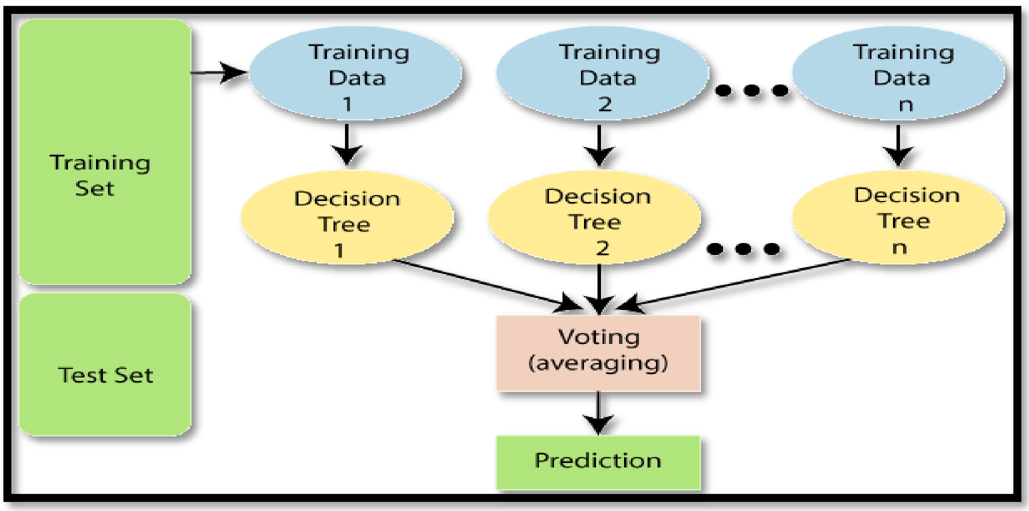
5. RANDOM FOREST

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

The below diagram explains the working of the Random Forest algorithm:



5.1 Assumptions for Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random Forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

5.2 Why use Random Forest?

Below are some points that explain why we should use the Random Forest algorithm:

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

5.3 How does Random Forest algorithm work?

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

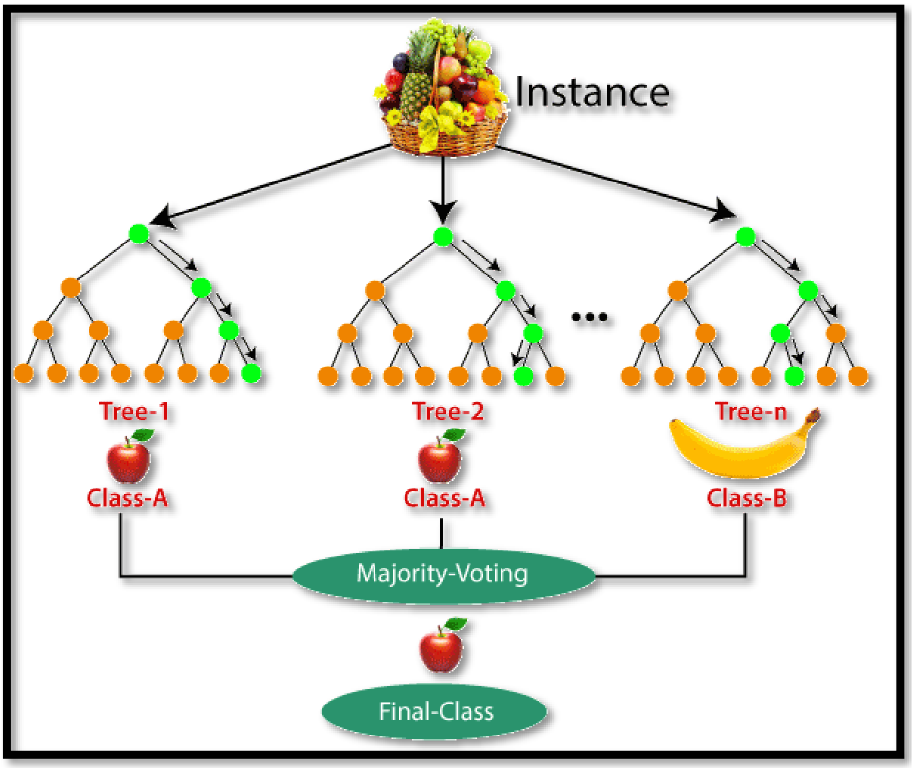
Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

The working of the algorithm can be better understood by the below example:

Example: Suppose there is a dataset that contains multiple fruit images. So, this dataset is given to the Random forest classifier. The dataset is divided into subsets and given to each decision tree. During the training phase, each decision tree produces a prediction result, and when a new data point occurs, then based on the majority of results, the Random Forest classifier predicts the final decision. Consider the below image:



5.4 Applications of Random Forest

There are mainly four sectors where Random forest mostly used:

1. Banking: Banking sector mostly uses this algorithm for the identification of loan risk.
2. Medicine: With the help of this algorithm, disease trends and risks of the disease can be identified.
3. Land Use: We can identify the areas of similar land use by this algorithm.
4. Marketing: Marketing trends can be identified using this algorithm.

5.5 Advantages of Random Forest

- Random Forest is capable of performing both Classification and Regression tasks.
- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

5.6 Disadvantages of Random Forest

- Although random forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.

6. IMPLEMENTATION CHALLENGES OF IOT IN AGRICULTURE

In this section the challenges of deploying IoT in the farm land is classified based on the technological challenging in the farmland. When incorporating recent technology such as IoT, Big data and Cloud computing in farming regions where only simple internet connectivity is a challenge. There are certain challenges the farmer need to be aware of it if they have decided to invest on smart farming. This section described the five main challenges that the agriculture sector faces in implementing IoT technology as follows.

i) Interoperability standard issues

In an agricultural environment deploying IoT technologies where information exchange should take place between all the interconnected IoT devices. Those devices are communicating via a common protocol, with required connectivity and standards. For this reason the most challenging fact is Interoperability. It is difficult to integrate heterogeneous data coming from different sensors (like moisture sensor, soil sensors, temperature sensor etc.) We need to spend a great deal of time and money to create standardized common protocols to all IOT devices.

ii) Storage Issues for massive agricultural data

Large volume of data generated from different IoT devices (like sensors, camera, weather stations etc.) connected in a farmland. For storing such huge volume of data, a large repository is needed. Storing data in the database is not sufficient to handle such a huge volume of data. Technologies like cloud computing and fog computing would do the better performance for storage and latency concern.

iii) Connectivity Issues

Providing connectivity with the data exchange between a IoT device and database or cloud is a major issue since making wireless connectivity is too difficult. The basic problem is poor Internet Connectivity in Farms. Most farms are situated in remote areas where access to the internet may not be good enough for quick transmission speed. Furthermore, communication lines may be disturbed by crops, bad weather, and other physical barriers. The 5G technologies which use space-based Internet might be the solution for this in future.

iv) Hardware and software implementation challenges and their maintenance

For developing an Agri based solution we have to select the right hardware and software tools and techniques. The quality of sensors, better storage techniques to store data and powerful data analytical tools. Because the results will be based on the accuracy of the data and its reliability. Since the sensors placed for monitoring the field might get affected by the animals, high wind, rain etc., Hardware maintenance is a challenging task in the field of agriculture. Thus, we require to make ensure hardware is durable and maintainable. Otherwise we need to replace sensors more frequently that leads to extra expenses.

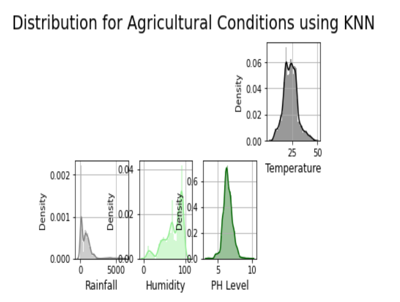
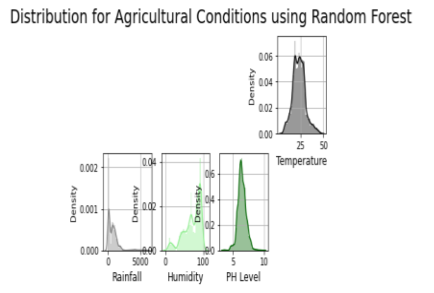
v) Security Issues

When implementing IoT on farmers are in need of having knowledge of security concept and it set up security policies. Smart farming and IoT technology imply working with huge volume of data from different sensors which increases the number of potential security loopholes that perpetrators can use for data theft and hacking attacks. Data security in agriculture is an unaware and challenging task. Many farms use drones that transmit data to farm devices. These devices connected to the Internet but have little to no security protections like sharing passwords or remote access authentications.

7. EXPERIMENTAL RESULTS

7.1 Comparative study using New Database

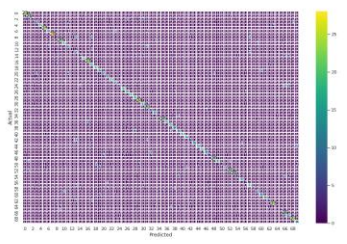
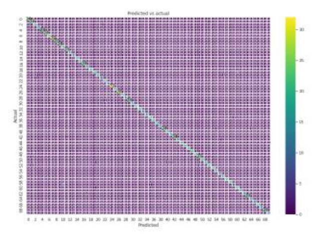
The prediction accuracy of the model using KNN algorithm accounts to 84.35% and prediction accuracy of the Random Forest algorithm is 96.5% on the above used dataset. Hence Random Forest is used for the deployment of the final model.

<table>
<tr><th>Category</th><th>K-NEAREST NEIGHBORS</th><th>RANDOM FOREST</th></tr>
<tr><td>Distribution of Agricultural conditions</td><td>
</td><td>
</td></tr>
<tr><td>Accuracy</td><td>Accuracy of K-Nearest Neighbors is 84.35%.</td><td>Accuracy of Random Forest is 96.5%.</td></tr>
<tr><td>Cross Validation Score</td><td><pre>score = cross_val_score(knn,features,target,cv=5)
print('Cross validation score: ',score)

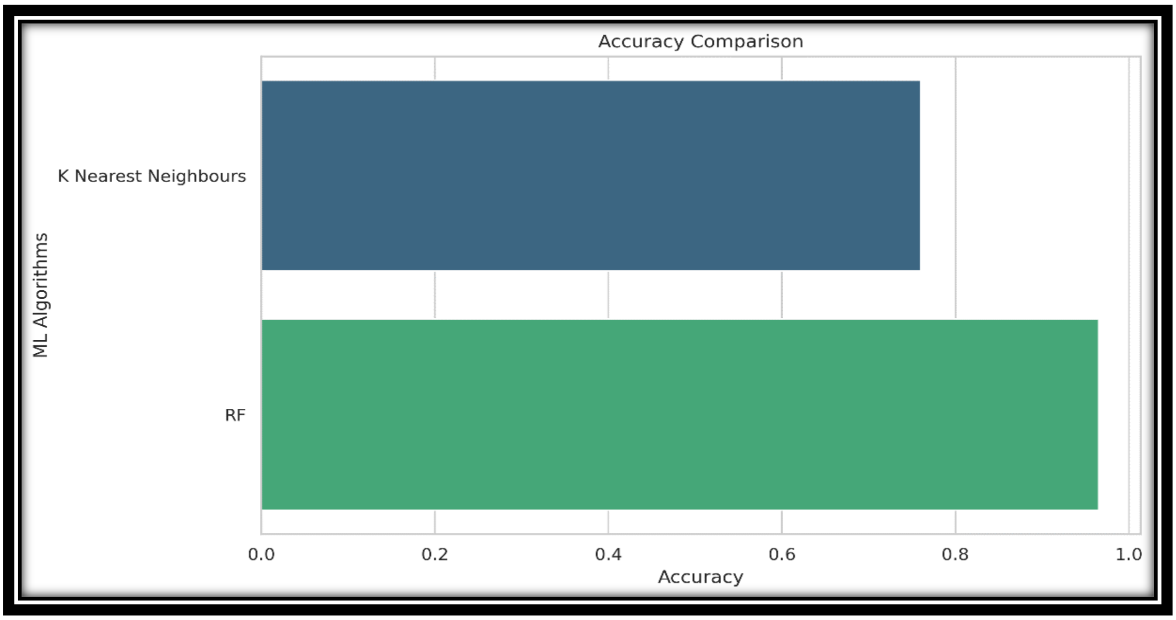
Cross validation score: [0.74642857 0.75285714 0.74928571 0.76357143 0.75857143]</pre></td><td><pre>score = cross_val_score(RF,features,target,cv=5)
print('Cross validation score: ',score)

Cross validation score: [0.96357143 0.96357143 0.96214286 0.955   0.95571429]</pre></td></tr>
<tr><td>Train-Test Accuracy</td><td><pre>knn_train_accuracy = knn_1.score(x_train,y_train)
print("knn_train_accuracy = ",knn_1.score(x_train,y_train))
#Print Test Accuracy
knn_test_accuracy = knn_1.score(x_test,y_test)
print("knn_test_accuracy = ",knn_1.score(x_test,y_test))

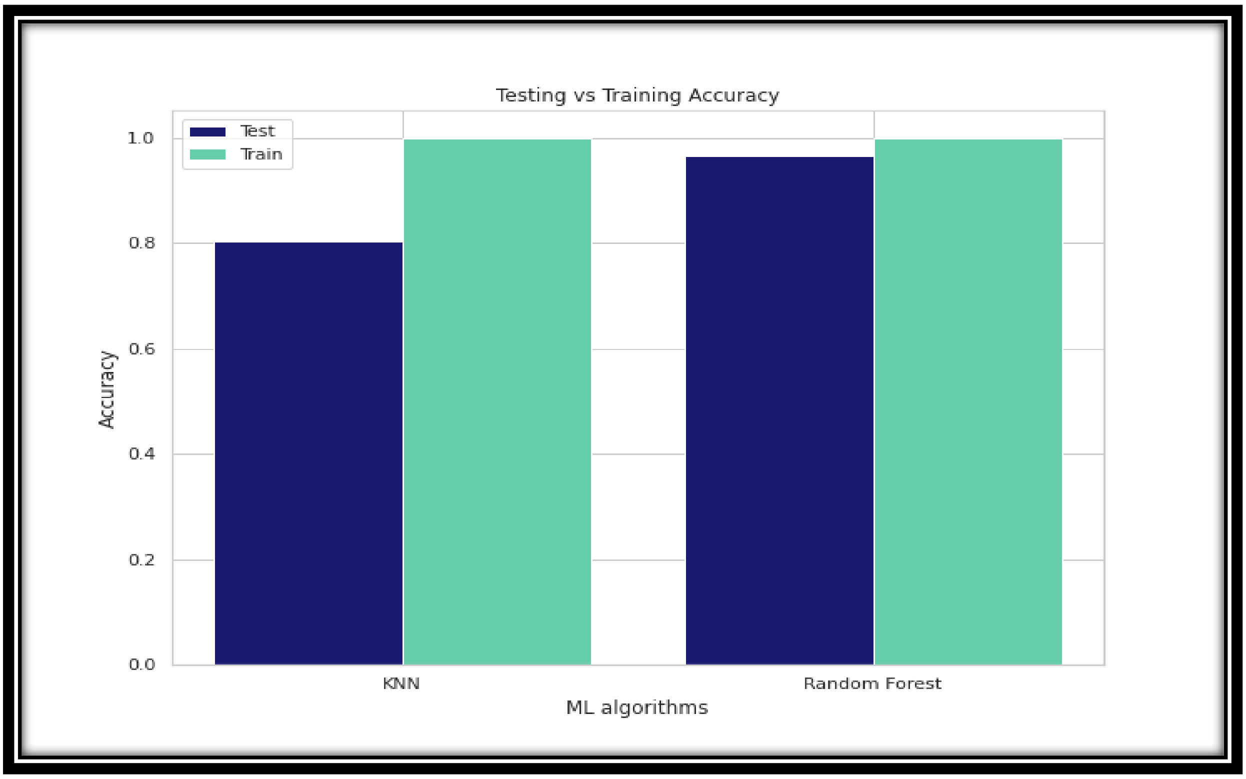
knn_train_accuracy = 1.0
knn_test_accuracy = 0.8435714285714285</pre></td><td><pre>rf_train_accuracy = RF.score(x_train,y_train)
print("Training accuracy = ",RF.score(x_train,y_train))
#Print Test Accuracy
rf_test_accuracy = RF.score(x_test,y_test)
print("Testing accuracy = ",RF.score(x_test,y_test))

Training accuracy = 0.9996428571428572
Testing accuracy = 0.965</pre></td></tr>
<tr><td>Confusion Matrix</td><td></td><td></td></tr>
</table>

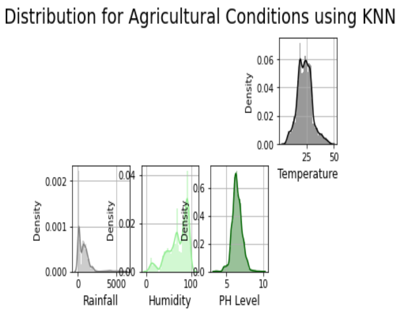
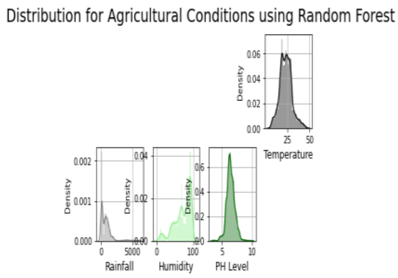
7.1.1 Accuracy Graphs of the New Dataset using KNN and Random Forest



7.1.2 Train vs Testing Accuracy Graphs of the New Dataset using KNN and Random Forest



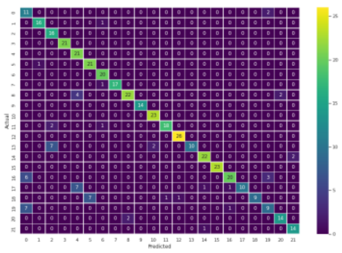
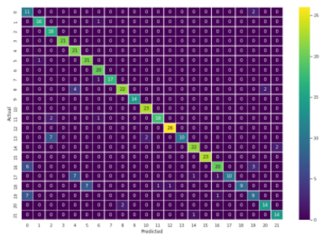
7.2 Comparative study using Old Database

<table>
<tr><th>Category</th><th>K-NEAREST NEIGHBORS</th><th>RANDOM FOREST</th></tr>
<tr><td>Distribution of Agricultural conditions</td><td></td><td></td></tr>
<tr><td>Accuracy</td><td>Accuracy of K-Nearest Neighbors is 85.68%.</td><td>Accuracy of Random Forest is 97.7%.</td></tr>
<tr><td>Cross Validation Score</td><td><code>score = cross_val_score(knn,features,target,cv=5)</code>
<code>print('Cross validation score: ',score)</code>

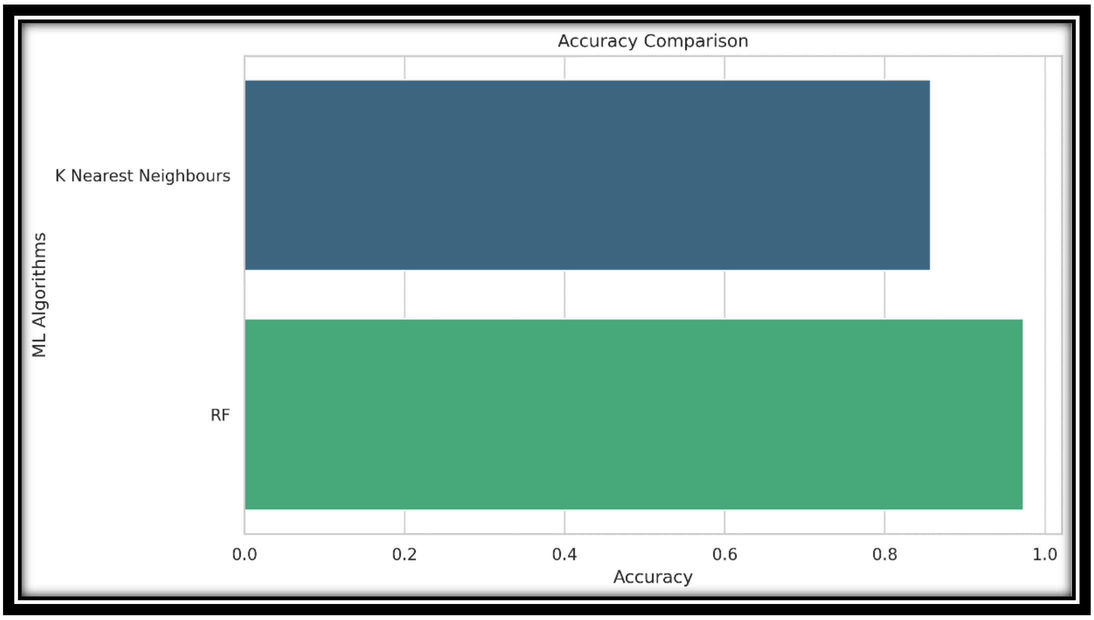
<code>Cross validation score: [0.86363636 0.85227273 0.87272727 0.85227273 0.87954545]</code></td><td><code>score = cross_val_score(RF,features,target,cv=5)</code>
<code>print('Cross validation score: ',score)</code>

<code>Cross validation score: [0.96363636 0.96818182 0.96136364 0.95454545 0.95454545]</code></td></tr>
<tr><td>Train-Test Accuracy</td><td><code>knn_train_accuracy = knn_1.score(x_train,y_train)</code>
<code>print("knn_train_accuracy = ",knn_1.score(x_train,y_train))</code>
<code>#Print Test Accuracy</code>
<code>knn_test_accuracy = knn_1.score(x_test,y_test)</code>
<code>print("knn_test_accuracy = ",knn_1.score(x_test,y_test))</code>

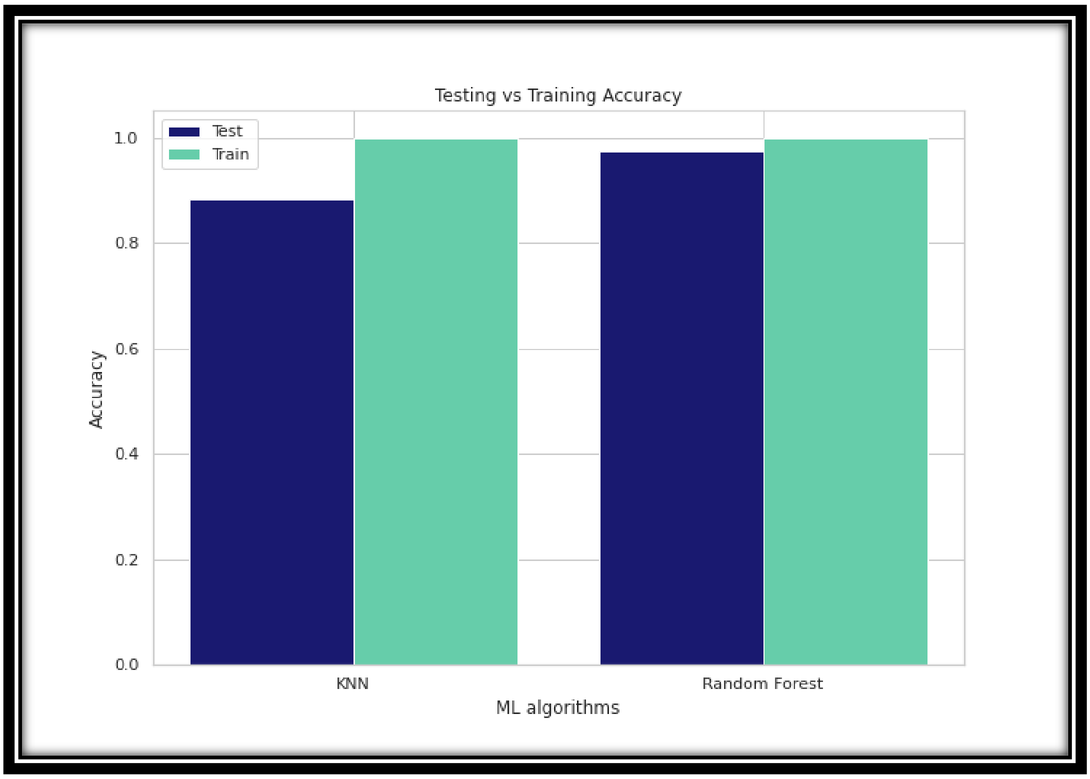
<code>knn_train_accuracy = 1.0</code>
<code>knn_test_accuracy = 0.884090909090909</code></td><td><code>#Print Train Accuracy</code>
<code>rf_train_accuracy = RF.score(x_train,y_train)</code>
<code>print("Training accuracy = ",RF.score(x_train,y_train))</code>
<code>#Print Test Accuracy</code>
<code>rf_test_accuracy = RF.score(x_test,y_test)</code>
<code>print("Testing accuracy = ",RF.score(x_test,y_test))</code>

<code>Training accuracy = 0.9994318181818181</code>
<code>Testing accuracy = 0.9727272727272728</code></td></tr>
<tr><td>Confusion Matrix</td><td></td><td></td></tr>
</table>

7.2.1 Accuracy Graphs of the Old Dataset using KNN and Random Forest



7.2.2 Train vs Testing Accuracy Graphs of the Old Dataset using KNN and Random Forest



8. CONCLUSION AND FUTURE SCOPE

India is a nation in which agriculture plays a prime role. In the prosperity of the farmers, the nation prospers. Thus our work would help farmers in sowing the right seed based on soil requirements to increase productivity and acquire profit out of such a technique. Thus the farmer can plant the right crop increasing his yield and also increasing the overall productivity of the nation. Our future work is aimed at an improved data set with a large number of attributes and also implements yield prediction.

Overall, the project has achieved its objectives, our IoT device is successfully working and the KNN algorithm is giving appropriate results accordingly. Thus the device can be used to predict the best fit crop to be grown in the fields.

The future scopes of our project include-

a) Fertilizer recommendation with the quantity measures.

b) Acknowledge the farmers about various loans and policies which may help them to earn extra profit.

c) Implementation of various cost-effective technological measures like push message notification using bluetooth technology over our current module which would lead to no mandatory need for smart phones.

d) Implementing other possible machine learning algorithms to check our model's efficiency and increase the accuracy if possible.

9. REFERENCES

[1] Miss.Snehal S.Dahikar (EXTC), Dr.Sandeep V.Rode (EXTC), "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach" *International Journal Of Innovative Research In Electrical, Electronics, Instrumentation and Control Engineering (Vol.2, Issue 1, January 2014)*

[2] Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H., Dimantha M.A.C, Pabasara M.G.P, "Crop Recommendation System", *International Journal of Computer Applications (0975 – 8887) Volume 175– No. 22, October 2020.*

[3] Junyan Ma, Xingshe Zhou and Zhigang Li, "Connecting Agriculture to the Internet of Things through Sensor Networks", *Internet of Things (iThings/CPSCom) 2011 International Conference on and 4th International Conference on Cyber Physical and Social Computing*, pp. 184-187, Oct 2011.

[4] S.Pudumalar, E.Ramanujam, R.Harine Rajashreeń, C.Kavyań, T.Kiruthikań, J.Nishań. , "Crop

Recommendation System for Precision Agriculture", *2016 IEEE Eighth International Conference on Advanced Computing (ICoAC)*

[5] Jirapond Muangprathub,Nathaphon Boonnam et al ,Computers and electronics in agriculture, computers and electronics in agriculture original papers IoT and agriculture data analysis for smart farm, volume 156, January 2019, pages 467-474

[6] Panel. Mohanraja Kirthika Ashokumarb and J. Narenc, Procedia Computer Science Field Monitoring and Automation Using IOT in Agriculture Domain, Procedia Computer Science Volume 93, 2016, Pages 931-939

[7] IoT in smart farming (Visited on 5th July, 2019)

[8] ESP32 Document (visited on 5th July, 2019)

[9] IoT application diagram (visited on 7th of July,2019)

[10] BezaNegashGetu, Hussain A. Attia, “Automatic Control of Agricultural Pumps Based on Soil Moisture Sensing”, IEEE conference publication,2015.

[11] G. M.K., J. J. and A. M. G.S, 2015. Providing Smart Agriculture Solutions to Farmers for better yielding using IOT. IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR), 2015.

[12] G. Parameswaran, K.Sivaprasath, - Arduino Based Smart Drip Irrigation System Using Internet of Things,DOI 10.4010/2016.1348 ,ISSN 2321 3361 © 2016 IJESC.

[13] Guido Fastellini and Calogero Schillaci, “Precision farming and IOT case studies across the world,” Agricultural Internet of Things and Decision Support for Precision Smart Farming, pp: 331-4152020, doi: 10.1016/B978-0-12-818373-1.00007-X

[14] Hemlata Channe1, Sukhesh Kothari2, Dipali Kadam3, “Multidisciplinary Model for Smart Agriculture using Internet-of-Things (IOT), Sensors, Cloud-Computing, Mobile- Computing & Big-Data Analysis”, International Journal of Int.J.Computer Technology &Applications,Vol 6 (3),May-June 2015.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[15] I. Marcu, G. Suciu, C. Bălăceanu, A. Drăgulinescu and M. A. Dobrea, "IOT Solution for Plant Monitoring in Smart Agriculture," 2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME), Cluj-Napoca, Romania, 2019, pp. 194- 197.