“ IoT and AI Based Monitoring System for Increasing the Yield in Sugarcane Production.”

Submitted in partial fulfillment of requirement of

University of Mumbai

For the Degree of

Bachelor of Engineering

Submitted By

Amaan Choudhary – 15

Azharsaied Qureshi– 62

Bhaskar Ananthraman – 23

Under Guidance of

Prof. Sonali Deshpande

Department of Computer Engineering

Smt. Indira Gandhi College of Engineering

Affiliated to University of Mumbai

(2020-2021)

Project Synopsis Approval Certificate

This Project synopsis entitled **AI and IoT Based Monitoring System for Increasing the Yield in Crop Production.**

By

Aman Choudhary – 15

Azharsaied Qureshi– 62

Bhaskar Ananthraman – 23

Is approved for the degree of **Bachelor of Engineering in Computer Engineering, Semester VII, University of Mumbai.**

Date:

Place: Ghansoli

Examiner 1 Examiner 2

Prof. Sonali Deshpande

Internal Guide

Prof. Sonali Deshpande

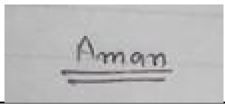
Head of Department

Dr. Sunil Chavan

Principal

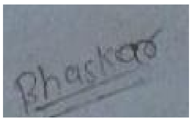
Declaration

We declare that this written submission represents our own ideas in our own words and where others ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any act/data/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.



Amaan Choudhary – 15

Azharsaied Qureshi– 62



Bhaskar Ananthraman – 23

ABSTRACT

Artificial Intelligence (AI) and Internet of things (IoT) based monitoring systems are in great demand and gives a precise extraction and analysis of data. In this system, the system is performed on a sugarcane plant to detect the most suitable conditions for plant growth. The philosophy behinds this work is to reduce the risks in agriculture and to promote smart farming practices. The effect of physical conditions like humidity, temperature, soil temperature and moisture and light intensity on the plant growth, is monitored using IoT based monitoring system. The data responsible for the plant growth is obtained using different sensors units like DHT11, LDR, and DS18B20, Soil Moisture sensors, single-board micro-controller and Application Programming Interfaces (APIs). The further analysis of the extracted parameters is done using different Machine Learning (ML) algorithms. This will help to reduce the farmer risk.

List of Abbreviations

- BE: Bachelor of Engineering
- DFD: Data Flow Diagrams
- ML: Machine Learning
- ANN: Artificial Neural network
- LSVM: Linear Support Vector Machine
- ADA Boost : Adaptive Boosting Classifier
- TS Cloud: ThingSpeak Cloud
- LDR: Light Intensity Sensor
- DHT11: Temperature Sensor
- API: Application Programming Interface

List of Figures

List of Tables

INDEX

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

Agriculture is the main source of our food and most people in India are dependent on agriculture as their prime source of income. According to the United Nations Food and Agricultural Organization, almost 800 million people are chronically hungry and 2 billion suffer micronutrient deficiencies. Further, another report from the State of Food Security and Nutrition in the World 2019 showed, 194.4 million people are undernourished in India. It is high time where the agriculture sector needs to be digitized and smart. Robert J. Mc Queen et al. (1995) [1], predicted the abilities of automatic process which could the combination of various systems. These systems could be applied in the farming sector for the welfare of humankind.

This era of 21st century focusses on automatic technologies like Internet of Things (IoT), Machine Learning (ML), and Data Science (DS). However, the application of these technologies in the agriculture domain is challenging. Because of regular variations in physical and chemical conditions of the surrounding, increasing the yield in the production of the crop is a real-time challenge and a problem needed to be solved. Hence, continuous monitoring and strict management is needed for agriculture parameters like humidity, temperature of the environment as well as of soil and light intensity, etc.

The major issues for the failure of crop production and low yield is lack of nutrients and suitable environmental conditions. Thus, the foremost motivation of this work is to provide an agricultural research solution. For this the ML algorithms is applied on the data generated by the technique of the Internet of Things. Here, we first create a dataset with the help of the Internet of things and the data collected is stored on the cloud. The data has features that consist of physical properties of surrounding like soil humidity, soil temperature and intensity of light. The target variable for our dataset is the rate of increase in height and width respectively which is the growth rate of the plant. After this the supervised machine learning algorithm are applied on the data which will give the best conditions for the plant for maximum growth.

1.1 PROBLEM STATEMENT

The project will carry out the analysis of the sensor's data related to sugarcanes plant from iot kit. The result of the project demonstrates the machine learning approach for predicting the suitable condition of plant growth.

1.2 OBJECTIVE

- Sugarcane sapling will be taken as a target plant for continuous monitoring of various physical parameters, through various sensors.
- The sensors are uniquely designed for measuring the targeted parameters i.e., humidity and temperature of the environment, soil moisture, soil temperature, and light intensity.
- To capture plants' height and width we used the concept of image processing and with the help of an open CV library. The extracted height and width of the plant are stored inform of a CSV file.
- All the environmental parameters will be collecting and storing in the cloud by using IoT.

1.3 SCOPE

- It can minimize the risk of farmer.
- It will help in increase the production of plant growth.
- It can be helpful to get the automatic system for farmer's and increase the level for artificial farming.

.

1.4 SYNOPSIS ORGANIZATION

In Chapter 2:

Planning and formulation of project is given, usage of spiral model and how we integrated and worked around the model.

In Chapter 3:

Shines light upon the requirements that are needed and analysis of system to uncover the additional requirements to the project.

In Chapter 4:

The system proposed is introduced which will tell the specification which of the project and will tell how the different modules of system will work, the flow of the project regarding data flow, control flow and other flow of the system.

In Chapter 5:

We will see the implementation of the algorithm of the project and process of model building.

In Chapter 6:

Conclusion and future scope of this project are mentioned.

CHAPTER 2

REVIEW OF LITERATURE

2. REVIEW OF LITERATURE

2.1 RESEARCH PAPER ANALYSIS

The below block diagram gives an overview of the system architect. Here, marigold sapling is taken as a target plant for continuous monitoring of various physical parameters, through various sensors. These sensors are uniquely designed for measuring the targeted parameters i.e., humidity and temperature of the environment, soil moisture, soil temperature, and light intensity.

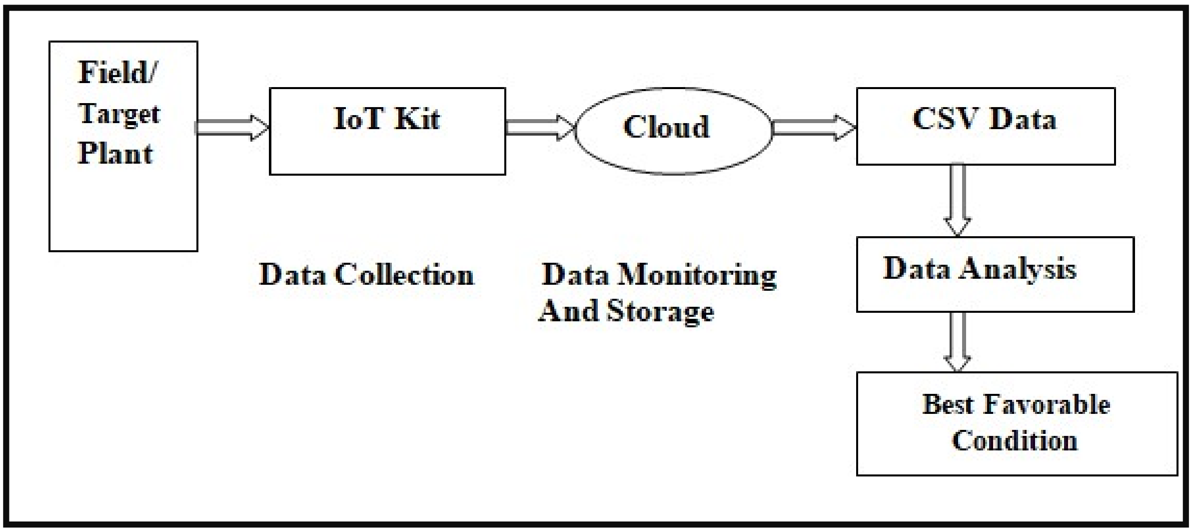


Fig 2.1.1: overview of the system architect

Arduino Integrated Development Environment (IDE) is used to write the embedded C codes in Arduino Module, compile and executes them. Arduino IDE is mainly used to give guidelines to Arduino [25]. Thingpeak platform is used to store and publish results online. It is an open-source IoT service application and API to store and get back data from various objects by using different HTTP and MQTT protocols [26]. NodeMCU is an IoT open-source platform that refers to the firmware, which is built on this development kit. It can work as an entire operating system using the Wi-Fi module of the NodeMCU. It uses the MQTT/HTTP protocol to send the data on the cloud. Arduino Uno is a single-board micro-controller developed by Arduino.cc, based on Microchip ATmega328P [27]. It has a boot-loader that allows uploading new codes without any external hardware.

The soil moisture sensor is used for measuring the volumetric content of water in the soil [8]. It is a parameter that determines the required amount of irrigation for the growth of the plant. This sensor module built around the LM393 comparator and consists of a variable resistor called 'potentiometer' for adjusting the sensitivity. DHT11 sensor module senses the relative humidity and temperature of the environment. Resistive type component is used to measure humidity, and the negative temperature coefficient type component is used to measure the temperature of the environment [19]. The percentage of water vapor present in the air is termed as Relative Humidity (RH) [9] which is an important parameter for a healthy plant growth.

As per the discussions, for measuring intensity, a voltage divider circuit is designed. Here, LDR is connected to 5 V supply via a 10K ohm resistor. The relationship between the resistance of LDR and lux [22] is shown below equation 1,

$$\text{lux} = (1.25\text{x } 107) \text{ x } R\text{-}1.4059 \qquad (1)$$

Where, R is the resistor of LDR Further, to measure the temperature of the soil a digital thermometer DS18B20 is used. The probe of the thermometer is inserted inside the soil for measuring the temperature. The unique characteristic of DS18B20 is direct to the digital temperature sensor [23]. The upper layer of earth 'Soil' is indispensable for plant life, support and nutrients, and water supply. It collects the heat during the daytime and releases heat during the night [24].

METHODOLOGY

A. Capturing the plant's height and width:
To capture plants' height and width we used the concept of image processing and with the help of open CV library. The extracted height and width of the plant is stored in form of CSV file. Meanwhile, all the environmental parameters are collected and stored in the cloud by using IoT. These values also extracted from a cloud in CSV form.

B. Preparing the Dataset:
We have got the target variable data that was the plant's height and width and then we found the rate of plant growth by using the formula in equation 2,

Height/Width Growth Rate =

$$\partial x/\partial t \qquad (2)$$

Where, ∂x changes in height or width. And ∂t is the difference in time during which the properties are measured.

C. Data Analysis and Cleaning:
We analysed and visualized the data and how valuable our features were and estimated a few predictions after features selection. Further, some of the extracted data was used for training the algorithm

D. Applying Algorithms:
After having the trained data, we applied supervised machine learning algorithms over it.

a) **Artificial Neural Network:**
An ANN is based on collection of connected unites or nodes called artificial neurons, which loosely model the neurons in a biological brain.

b) **Linear Support Vector Classifier:**
This model is essentially representing various classes in a multidimensional space hyperplane.

c) **Decision Tree:**
A decision tree is a classifier with feature marked internal nodes. Each node marked with possible output and leads to a lower-ranked decision node on various input features.

d) **Random Forest:**
This algorithm is used for classification as well as regression predictions. It creates different decision trees on data and gives predictions from them. Finally, provides the best solution to the problem for which it is employed.

e) **Adaptive Boosting Classifier:**
AdaBoost helps u combine multiple "week classifiers" into a single "strong classifier". The weak learners in AdaBoost are decision trees with a single split, called decision stumps. AdaBoost works by putting more weight on difficult to classify instances and less on those already handled well.

f) **Feed Forward algorithm:**
In this network, the information moves in only one direction -**Forward**- from the input nodes, through the hidden nodes. (if any) and to the output nodes.

g) **Backpropagation algorithm:**
The backpropagation algorithm calculates how much the final output values, o1 and o2, are affected by each of the weights. To do this, it calculates partial derivatives, going back from the error function to the neuron that carried a specific weight.

CHAPTER 3

PLANNING AND FORMULATION

3. PLANNING AND FORMULATION

3.1 SOFTWARE DEVELOPMENT MODEL

In order to effectively design and develop a cost-effective model spiral model was practiced. Spiral model is one of the most important Software Development Life Cycle models, which provides support for Risk Handling. In its diagrammatic representation, it looks like a spiral with many loops. The exact number of loops of the spiral is unknown and can vary from project to project. Each loop of the spiral is called a Phase of the software development process. The exact number of phases needed to develop the product can be varied by the project manager depending upon the project risks. As the project manager dynamically determines the number of phases, so the project manager has an important role to develop a product using spiral model.

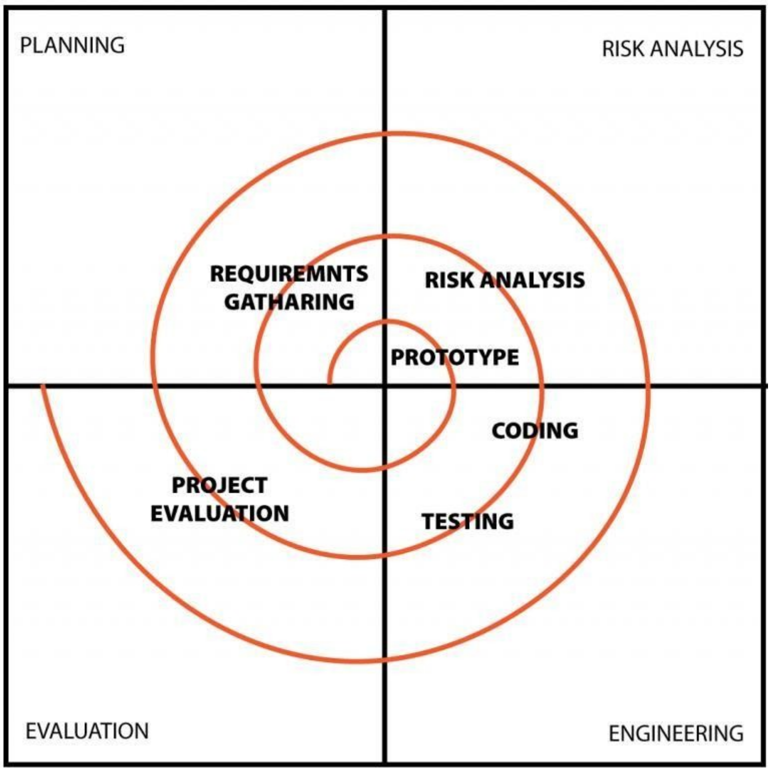


Figure 3.1.1: Spiral model

The Radius of the spiral at any point represents the expenses(cost) of the project so far, and the angular dimension represents the progress made so far in the current phase. Each phase of Spiral Model is divided into four quadrants as shown in the above figure. The functions of these four quadrants are discussed below-

1. **Objectives determination and identify alternative solutions:** Requirements are gathered from the customers and the objectives are identified, elaborated and analysed at the start of every phase. Then alternative solutions possible for the phase are proposed in this quadrant.
2. **Identify and resolve Risks:** During the second quadrant all the possible solutions are evaluated to select the best possible solution. Then the risks associated with that solution is identified and the risks are resolved using the best possible strategy. At the end of this quadrant, Prototype is built for the best possible solution.
3. **Develop next version of the Product:** During the third quadrant, the identified features are developed and verified through testing. At the end of the third quadrant, the next version of the software is available.
4. **Review and plan for the next Phase:** In the fourth quadrant, the Customers evaluate the so far developed version of the software. In the end, planning for the next phase is started.

Risk Handling in Spiral Model

A risk is any adverse situation that might affect the successful completion of a software project. The most important feature of the spiral model is handling these unknown risks after the project has started. Such risk resolutions are easier done by developing a prototype. The spiral model supports coping up with risks by providing the scope to build a prototype at every phase of the software development.

Advantages of Spiral Model:

- **Risk Handling:** The projects with many unknown risks that occur as the development proceeds, in that case, Spiral Model is the best development model to follow due to the risk analysis and risk handling at every phase.
- **Good for large projects:** It is recommended to use the Spiral Model in large and complex projects.
- **Flexibility in Requirements:** Change requests in the Requirements at later phase can be incorporated accurately by using this model.
- **Customer Satisfaction:** Customer can see the development of the product at the early phase of the software development and thus, they habituated with the system by using it before completion of the total product.

Disadvantages of Spiral Model:

- **Complex:** The Spiral Model is much more complex than other SDLC models.
- **Expensive:** Spiral Model is not suitable for small projects as it is expensive.
- **Too much dependable on Risk Analysis:** The successful completion of the project is very much dependent on Risk Analysis. Without very highly experienced expertise, it is going to be a failure to develop a project using this model.
- **Difficulty in time management:** As the number of phases is unknown at the start of the project, so time estimation is very difficult.

3.2 TIMELINE CHART

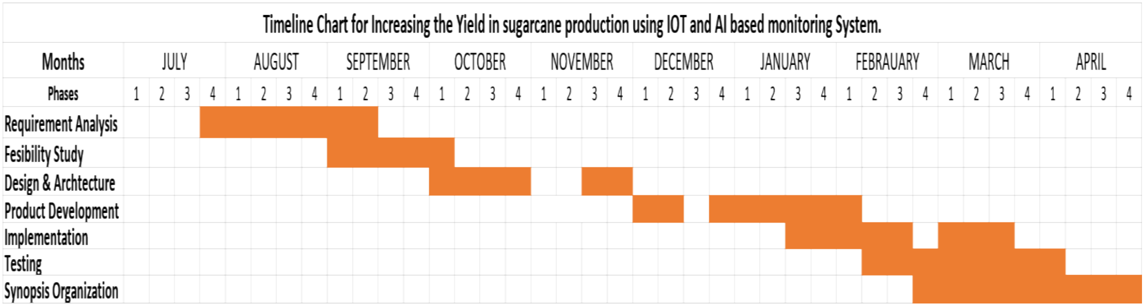


Table 3.2.1: Timeline chart

3.3 FEASIBILITY ANALYSIS

We have done a feasibility study which is an evaluation of the usefulness of a proposed project or system. A feasibility study aims to objectively and rationally uncover the strengths and weaknesses of an existing business or proposed venture, opportunities and threats present in the environment, the resources required to carry through, and ultimately the prospects for success. In its simplest terms, the two criteria to judge feasibility are cost required and value to be attained.

1. Technical Feasibility: We have evaluated the technical requirements and capacity of the systems required. We have also looked after the resources needed for the project.
2. Economic Feasibility: We have assessed a cost/ benefits analysis of the project, serving to verify the viability, cost, and benefits related to a project before monetary resources are allocated. The cost is comparable to the benefits of the project.
3. Legal Feasibility: All the data and resources of the project gathered is open source and is available to use for research, education, and non-profit use.
4. Operational Feasibility: The proposed project will use the data provided form GEE and help to identify the loss in the forest area in a given specific region.
5. Resource Feasibility: The data and system required for the development of the project will be provided by Google that will help to manage the required memory and processing power.

When these areas have all been examined, the feasibility analysis helps determine any constraints the proposed project might face, including:

• Internal Project Constraints, viz, Technical, Technology, Budget, Resource, etc.
• Internal Corporate Constraints, viz. Financial, Marketing, Export, etc.
• External Constraints, viz. Logistics, Environment, Laws, and Regulations, etc.

3.4 COST-BENEFITS ANALYSIS

Cost-benefits analysis (CBA), sometimes called benefit costs analysis (BCA), is a systematic approach to estimating the strengths and weaknesses of alternatives used to determine options which provide the best approach to achieving benefits while preserving savings (for example, in transactions, activities, and functional business requirements). A CBA may he used to compare completed or potential courses of actions, or to estimate (or evaluate) the value against the cost of a decision, project, or policy. It is commonly used in commercial transactions, business or policy decisions (particularly public policy), and project investment.

CHAPTER 4

SYSTEM ANALYSIS

4. SYSTEM ANALYSIS

4.1 HARDWARE REQUIREMENTS

- Iot kit (DHT11),
-LDR,
-DS18B20,
-Soil Moisture sensors,
-Single board micro-controllers
-Infrared sensor.

4.2 SOFTWARE REQUIREMENTS

1. Arduino Integrated Development Environment (IDE)
2. Python 3 (ML & AI)
3. Anaconda Navigator

4.3 FUNCTIONAL REQUIREMENT

In software engineering, a functional requirement defines a function of software system or its component. A function is described as a set of inputs, the behaviour, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioural requirements describing all the cases where the system uses functional requirements are captured in use cases. Functional requirements are supported by non-functional requirements, which impose constraints on the design or implementation. As defined in requirements specify results of a system. Functional requirements drive the application architecture of a system.

4.4 NON-FUNCTIONAL REQUIREMENT

In system engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of system, rather than specific behaviours. This should be contrasted with functional requirements is detailed in the system design. The plan for implementing non-functional requirements in detailed in the system architecture.

In general, functional requirements define what a system is supposed to do whereas non-functional requirements define how a system is supposed to be. Functional requirements are in the form of system shall do <requirement>, while non-functional requirements are in the form of system shall be <requirement>.

4.5 DATA FLOW DIAGRAMS(DFD)

Data flow diagrams is a graphical representation of flow of data through an information system. DFD can also be used for the visualization of data processing. It is common practise for designers to draw context-level DFD first which shows the interaction between the system and outside entities. This context-level DFD is then exploded to show more details of the system being modelled.

The components of data flow diagrams are:

- External Entities are outside of the system. Terminators represent where information comes from and where it goes. In designing a system, we have no idea about what these terminators do or how they do it.
- Processes modify the inputs in the process of generating the outputs.
- Data stores represents in the process where data comes to rest. A DFD does not say anything about the relative timing of the process, so a data store might be a place to accumulate data over year for the annual accounting process.

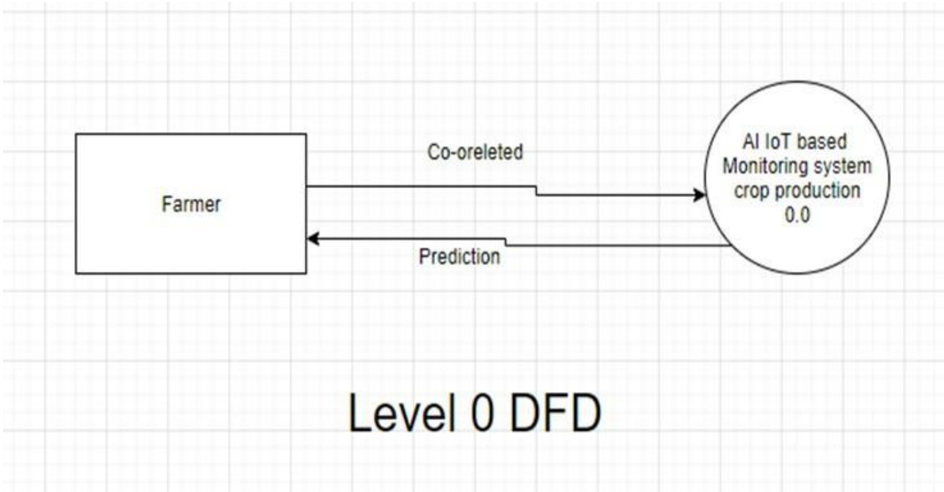


Figure 4.5.1: DATA FLOW DIAGRAM (LEVEL 0)

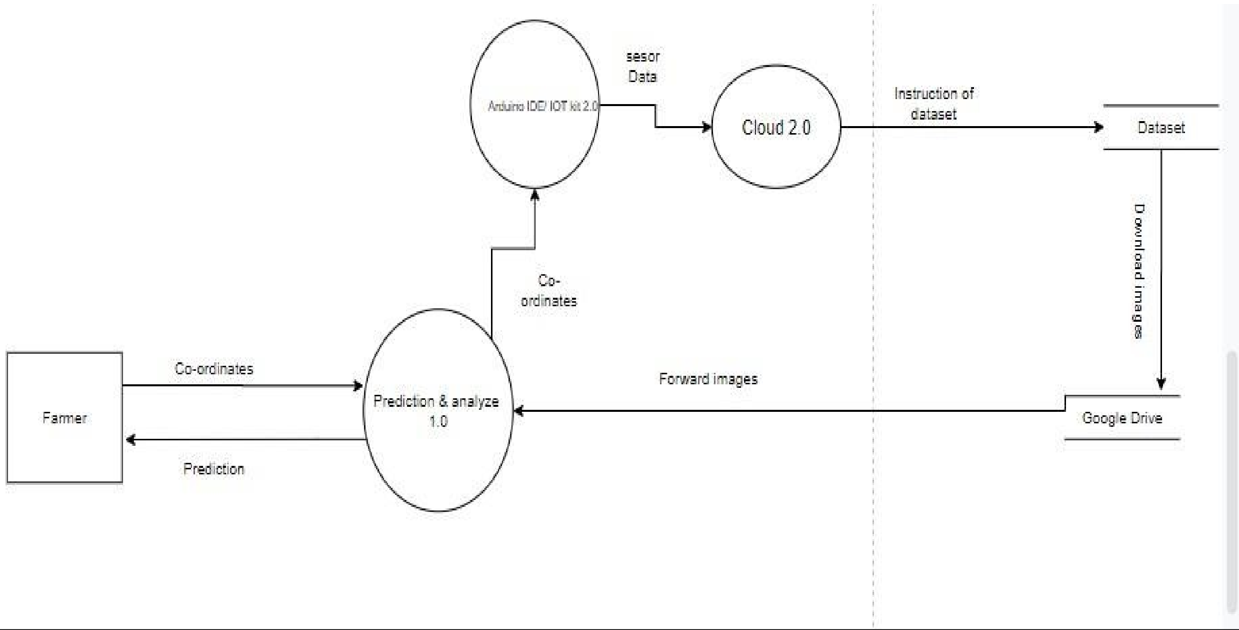


Figure 4.5.2: DATA FLOW DIAGRAM (LEVEL 1)

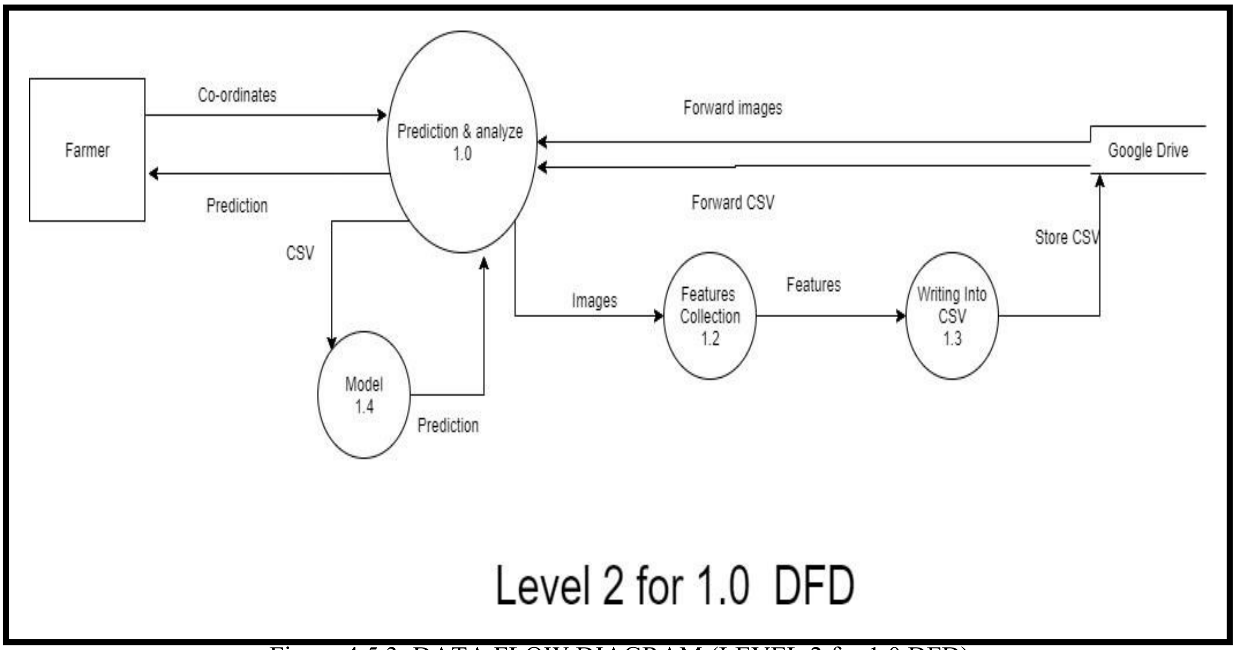


Figure 4.5.3: DATA FLOW DIAGRAM (LEVEL 2 for 1.0 DFD)

CHAPTER 5

SYSTEM DESIGN

5 SYSTEM DESIGN

5.1 BASIC FLOWCHART

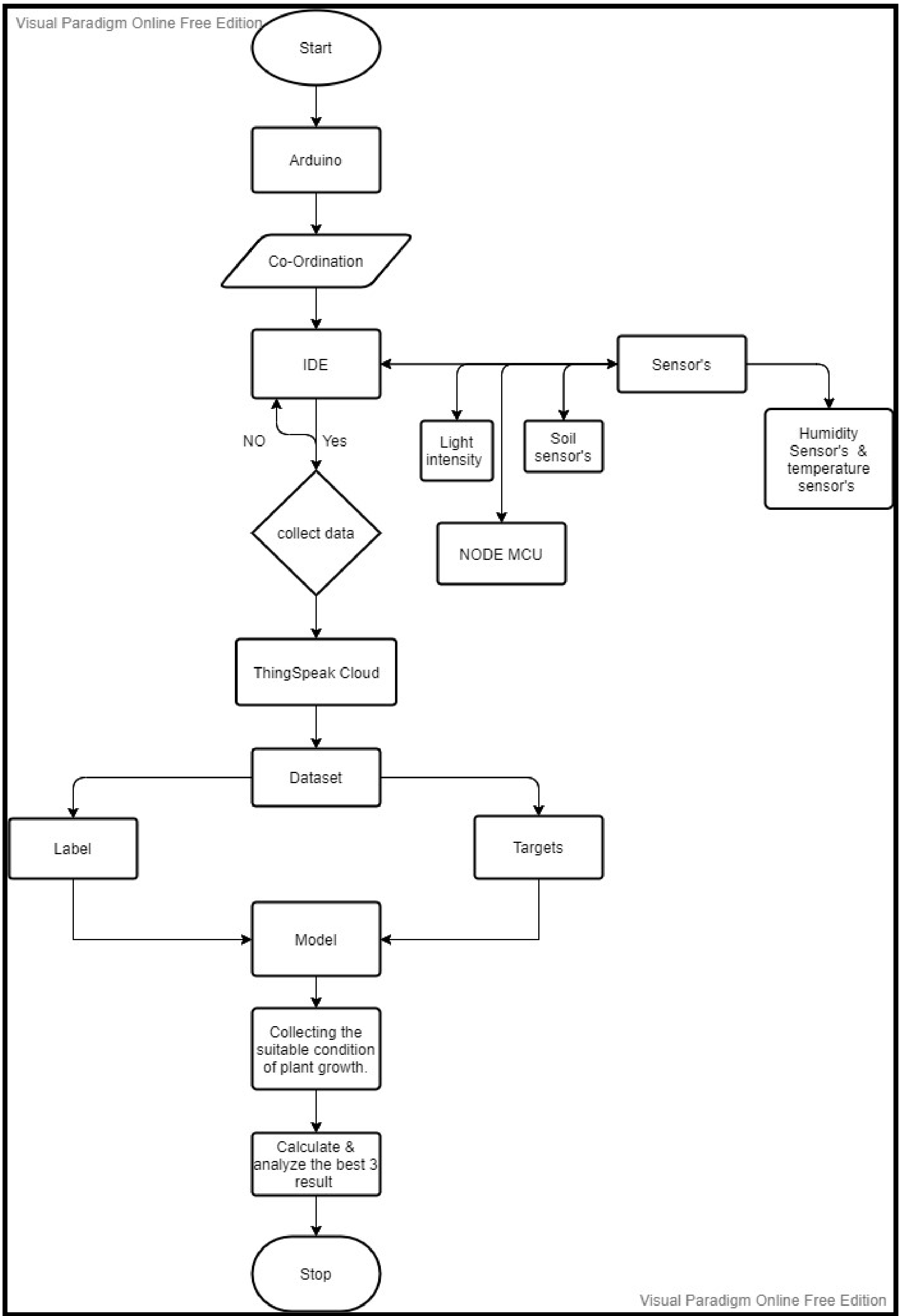


Figure 5.1: Basic Flowchart

5.2 USE CASE DIAGRAM

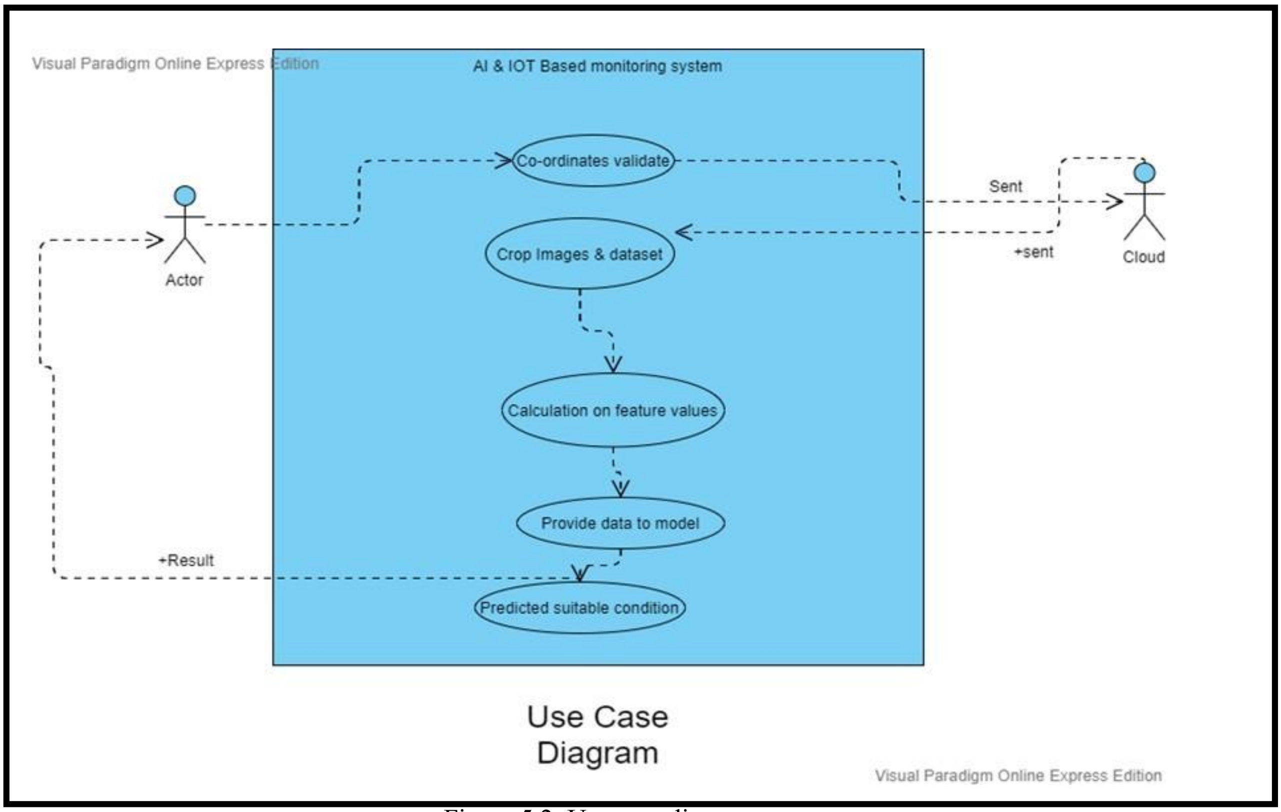


Figure 5.2: Use case diagram

5.3 ACTIVITY DIAGRAM

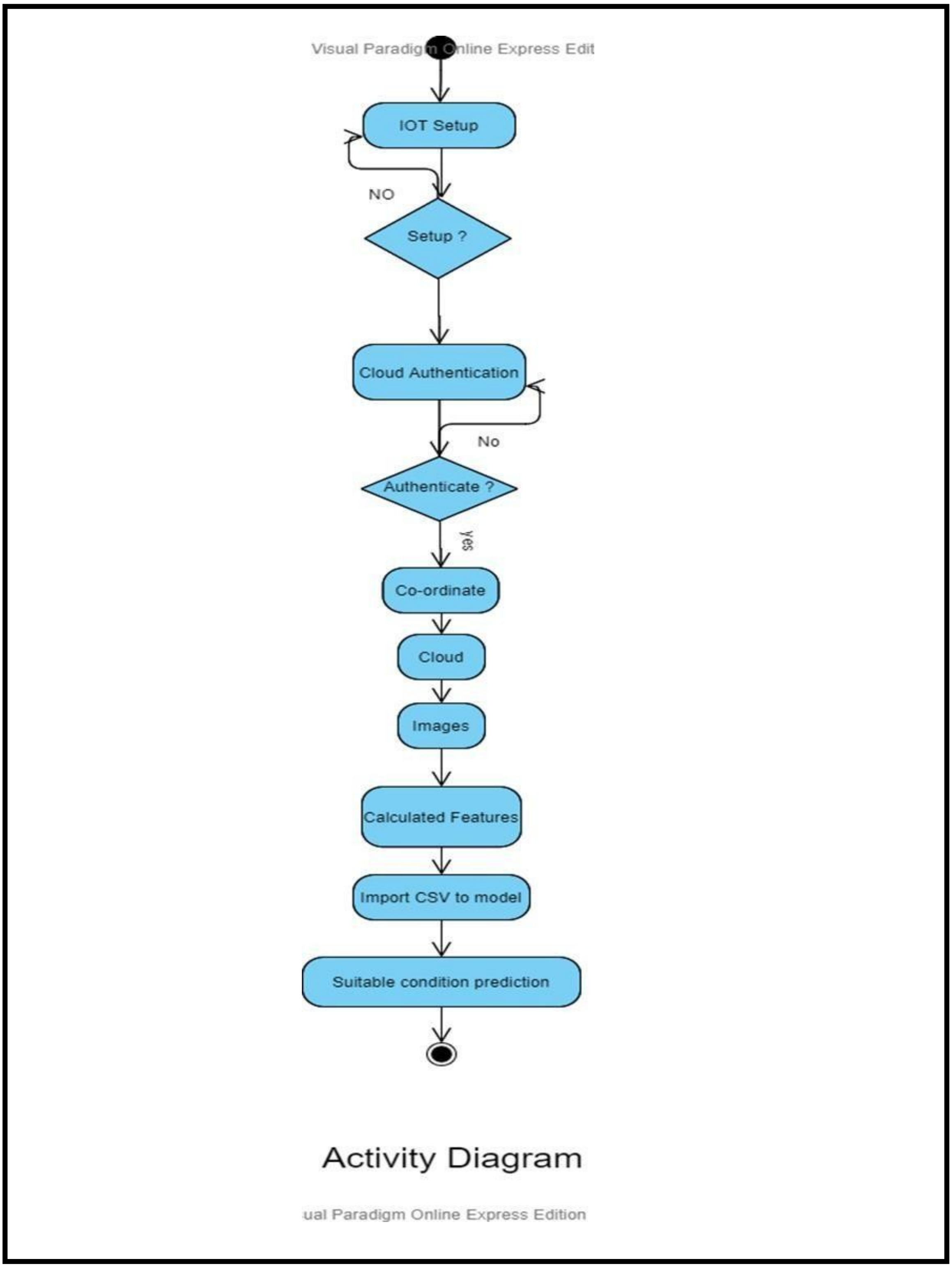


Figure 5.3: Activity diagram

5.4 SEQUENCE DIAGRAM

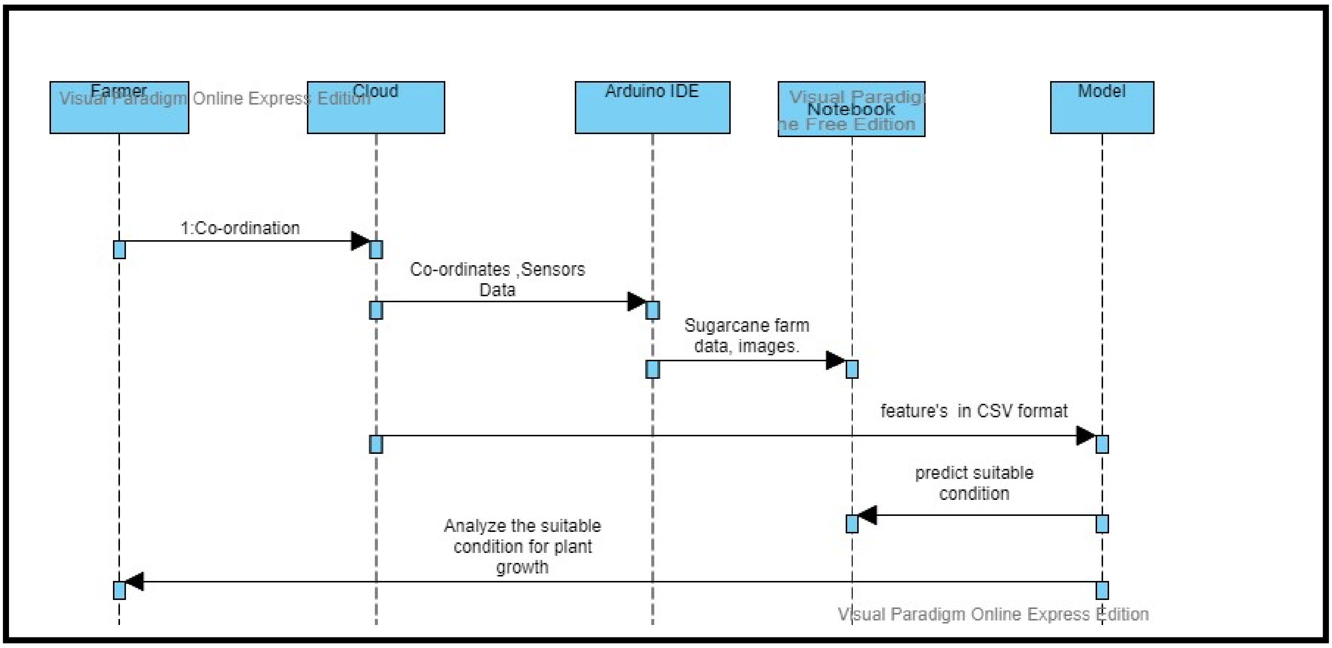


Figure 5.4.4: Sequence diagram

5.5 CLASS DIAGRAM

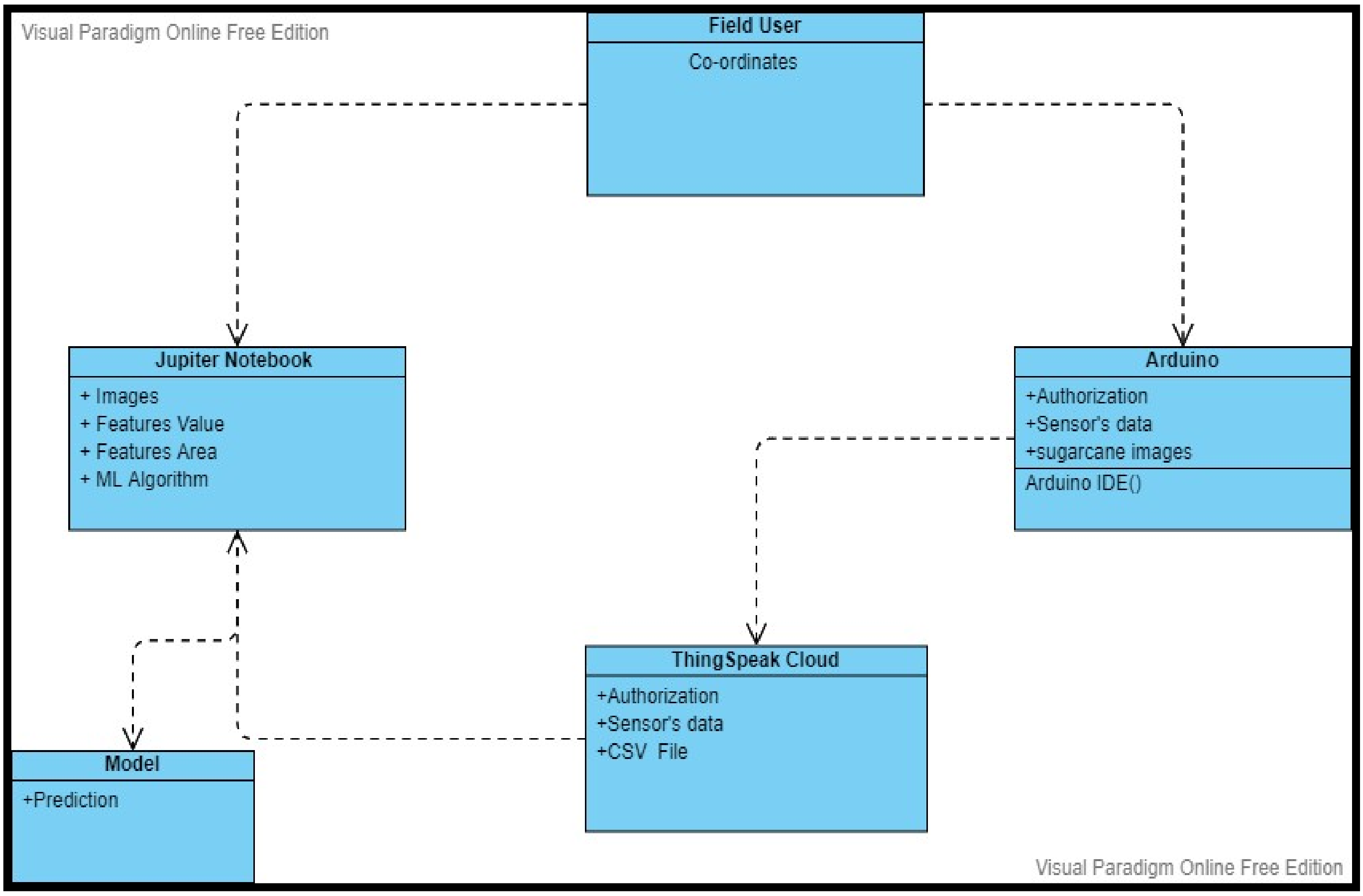


Figure 5.5.1: Class diagram

CHAPTER 6

ALGORITHM DEVELOPMENT

6 ALGORITHM DEVELOPMENT

6.1 FEATURE CALCULATION

A. COMPUTATIONAL ANALYSIS OF SENSOR PARAMETER

Here we include some basic analytic methods to calculate the Sensors parameters like Temperature and Humidity.LM35 Temperature sensor gives output voltage 10 mv for 1°C.this sensor output is connected to any analog pin of Arduino Uno. Uno converts analog voltage into digital using on chip ADC.

ADC reading=analog Read (A1); (1)
Voltage= ADC reading*5/ (1023); (2)
Temperature=Voltage*100; (3)
Relative Humidity = (density of water vapor / density of water vapor at saturation) x 100%. (4)

B. SENSOR DATA VALIDATION:

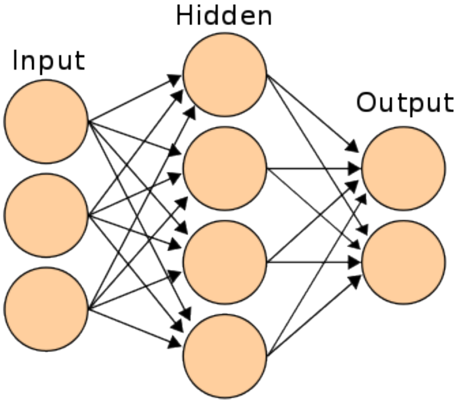
ALGORITHM:
Heuristic rule:
$\delta 1$ -minimum threshold
$\delta 2$ –maximum threshold
n –Total number of sensed values

Input: array of sensed data (x)
If (x[i] $\geq \delta 1$ and x[i] $\leq \delta 2$)
Then status [i] ← good
Else
Status [i] ←Not good
End if
End

Output: status for sensed data (x)
For i←0 ton do i←i+1

Formulation of Neural network:

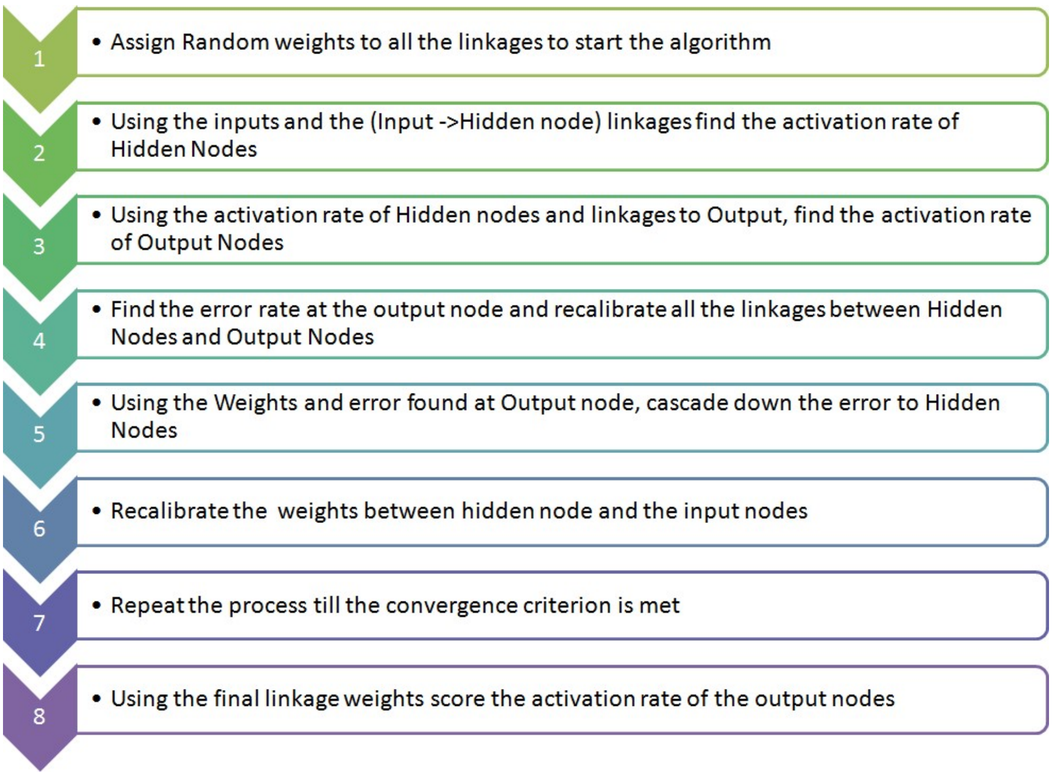
We will start with understanding formulation of a simple hidden layer neural network. A simple neural network can be represented as shown in the figure below:



The linkages between nodes are the most crucial finding in an ANN. We will get back to "how to find the weight of each linkage" after discussing the broad framework. The only known values in the above diagram are the inputs. Let's call the inputs as I1, I2 and I3, Hidden states as H1, H2, H3 and H4, Outputs as O1 and O2. The weights of the linkages can be denoted with following notation:

W (I1H1) is the weight of linkage between I1 and H1 nodes.

ANN algorithm:



Decision tree algorithm:

In general, Decision tree analysis is a predictive modelling tool that can be applied across many areas. Decision trees can be constructed by an algorithmic approach that can split the dataset in different ways based on different conditions. Decisions tress are the most powerful algorithms that falls under the category of supervised algorithms.

Implementing Decision Tree Algorithm: Gini Index

It is the name of the cost function that is used to evaluate the binary splits in the dataset and works with the categorial target variable “Success” or “Failure”.

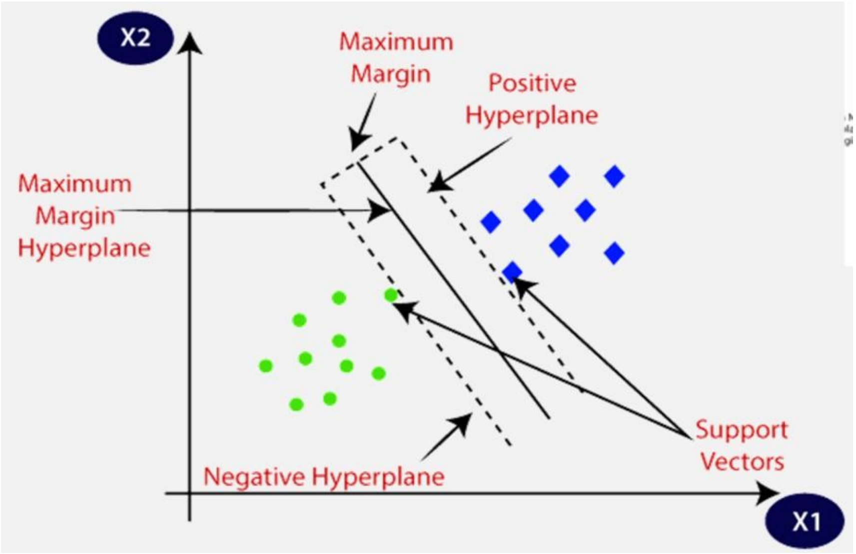
Higher the value of Gini index, higher the homogeneity. A perfect Gini index value is 0 and worst is 0.5 (for 2 class problem).

Split Creation:

- **Part1: Calculating Gini Score** − We have just discussed this part in the previous section.
- **Part2: Splitting a dataset** − It may be defined as separating a dataset into two lists of rows havingindex of an attribute and a split value of that attribute. After getting the two groups - right and left, from the dataset, we can calculate the value of split by using Gini score calculated in first part. Split value will decide in which group the attribute will reside.
- **Part3: Evaluating all splits** − Next part after finding Gini score and splitting dataset is the evaluation of all splits. For this purpose, first, we must check every value associated with each attribute as a candidate split. Then we need to find the best possible split by evaluating the cost of the split. The best split will be used as a node in the decision tree.

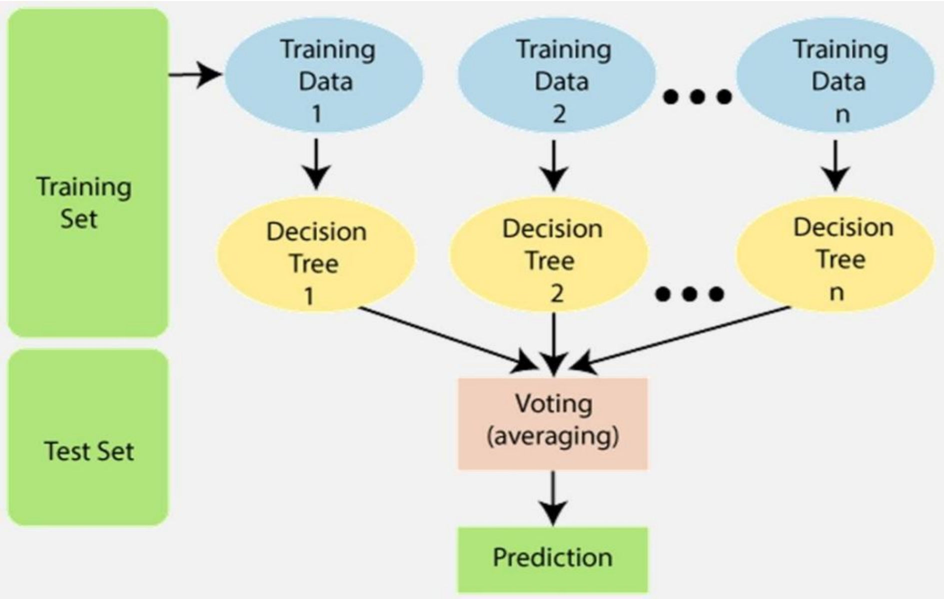
SVM:

“Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.
”



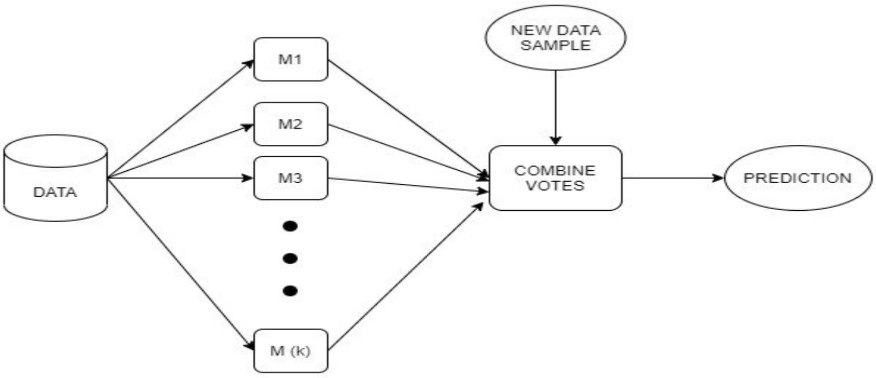
RANDOM FOREST:

"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."



AdaBoost algorithm :

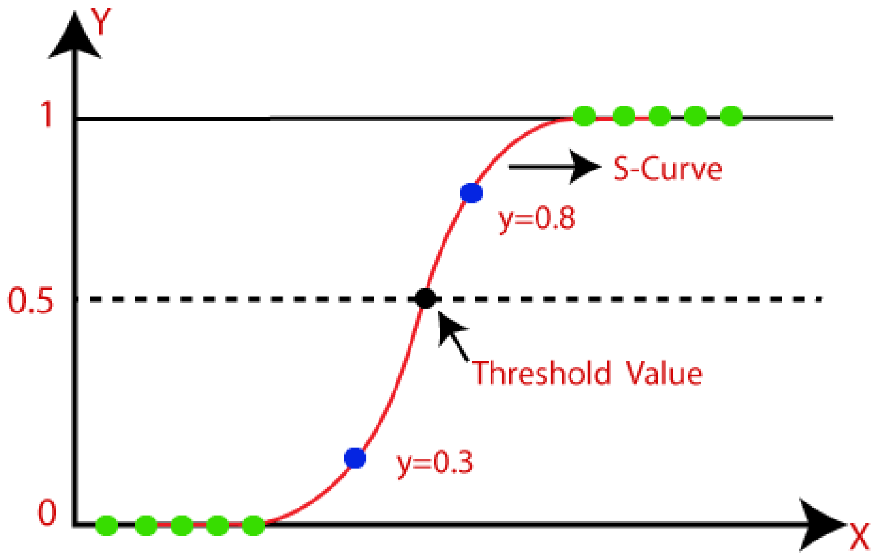
"AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique that is used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights to incorrectly classified instances."



Logistic regression:

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes.

In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).
Mathematically, a logistic regression model predicts P(Y=1) as a function of X.
It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.



Naïve Bayes:

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
It is mainly used in text classification that includes a high-dimensional training dataset.
Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

Likelihood — Class Prior Probability

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

Posterior Probability — Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

CHAPTER 7

IMPLEMENTATION

7. IMPLEMENTATION

```
In [1]: import random
        plantwidth = []
        plantheight= []
        phvalue=[]
        temp=[]
        humid=[]#in percentange
        moist=[]#in percentange
        growth=[]#growth
        for i in range(1,10001):
            n = random.randint(14,40)
            m = random.randint(15,120)
            l=random.randint(3,10)
            j=random.randint(17,40)
            k=random.randint(40,110)
            z=random.randint(10,45)
            plantwidth.append(n)
            plantheight.append(m)
            phvalue.append(l)
            temp.append(j)
            humid.append(k)
            moist.append(z)
```

Data set Creation Block

This block is use to initialize or create the data set. Here imported the random value and then initialize the list of plant width, plant height, PH value, temperature, humidity, moisture, growth after that we parsing the value and append the value in each list.

```
]: x=dff_new.drop(['possible','plantheight','plantwidth','phvalue','Unnamed: 0'],axis=1)
   y=dff_new['possible']
```

```
]: from sklearn.ensemble import AdaBoostClassifier
   from sklearn.tree import DecisionTreeClassifier

   import pandas as pd
   import numpy as np
   from sklearn.model_selection import train_test_split
   from sklearn.metrics import confusion_matrix
   import pickle
   from sklearn.preprocessing import LabelEncoder
```

```
]: x_train,x_test , y_train, y_test = train_test_split(x, y,test_size=0.5)
```

Importing Libraries

In this block we have initialized or imported the all the library which are going to use for to train the model. The goal of ensemble methods is to combine the predictions of several base estimators built with a given learning algorithm in order to improve generalization / robustness over a single estimator

```
: from sklearn import metrics
  from sklearn.tree import DecisionTreeClassifier
  from sklearn.model_selection import train_test_split
  from sklearn.metrics import roc_curve,roc_auc_score,f1_score
  from sklearn.ensemble import RandomForestClassifier
  from sklearn.metrics import confusion_matrix
  from sklearn.metrics import accuracy_score
  from sklearn.metrics import roc_curve
  from sklearn.metrics import f1_score
  from sklearn.metrics import precision_score,recall_score
  from sklearn.metrics import roc_auc_score,roc_curve
```

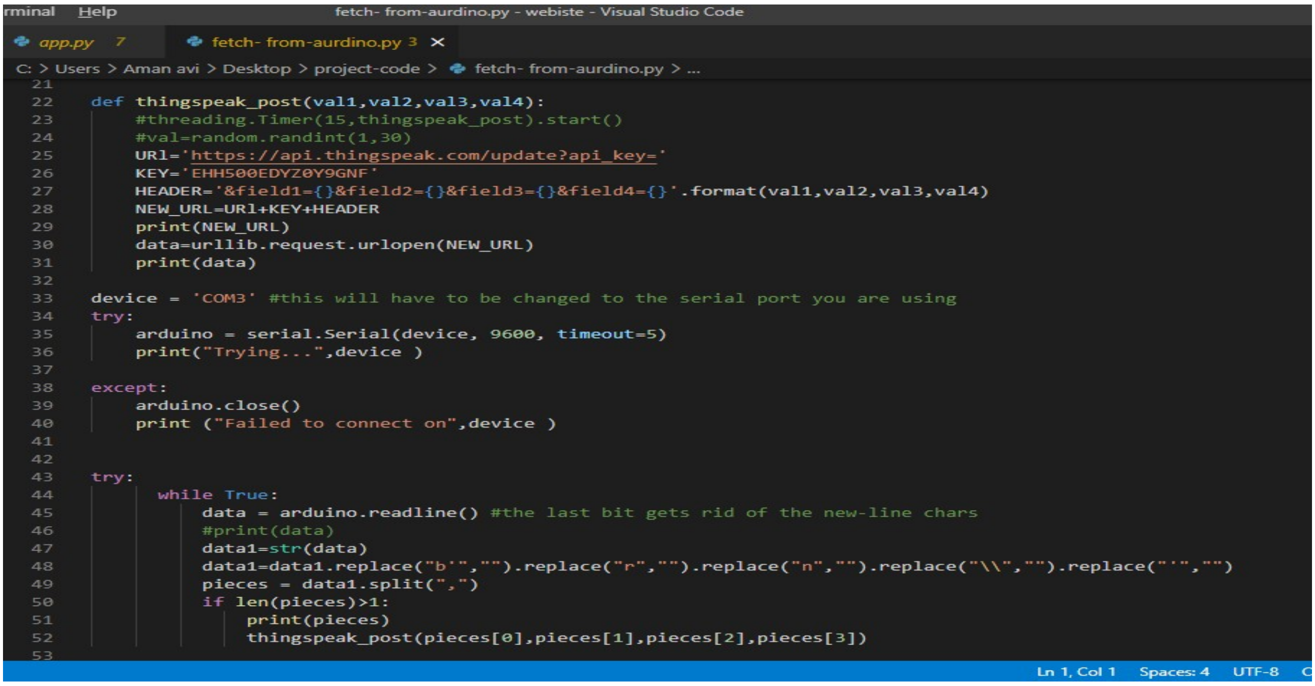
.A confusion matrix is a technique for summarizing the performance of a classification algorithm. In this block we are importing all the require ML library which is useful for to train the Machine learning model.

```
In [34]: dt=DecisionTreeClassifier()
         dt.fit(x_train,y_train)
         Y_pred_dt = dt.predict(x_test)
         score_dt = round(accuracy_score(Y_pred_dt,y_test)*100,2)

         print("The accuracy score achieved using Decision Tree is: "+str(score_dt)+" %")

         The accuracy score achieved using Decision Tree is: 100.0 %

In [35]: lr_cm=confusion_matrix(y_test,Y_pred_dt)
         plt.title("Decision Tree")
         sns.heatmap(lr_cm,annot=True,cmap="Blues",fmt="d",cbar=False)

Out[35]: <AxesSubplot:title={'center':'Decision Tree'}>
```


```
In [36]: import pickle
         with open('dt.sav','wb') as f:
             pickle.dump(dt,f)
```

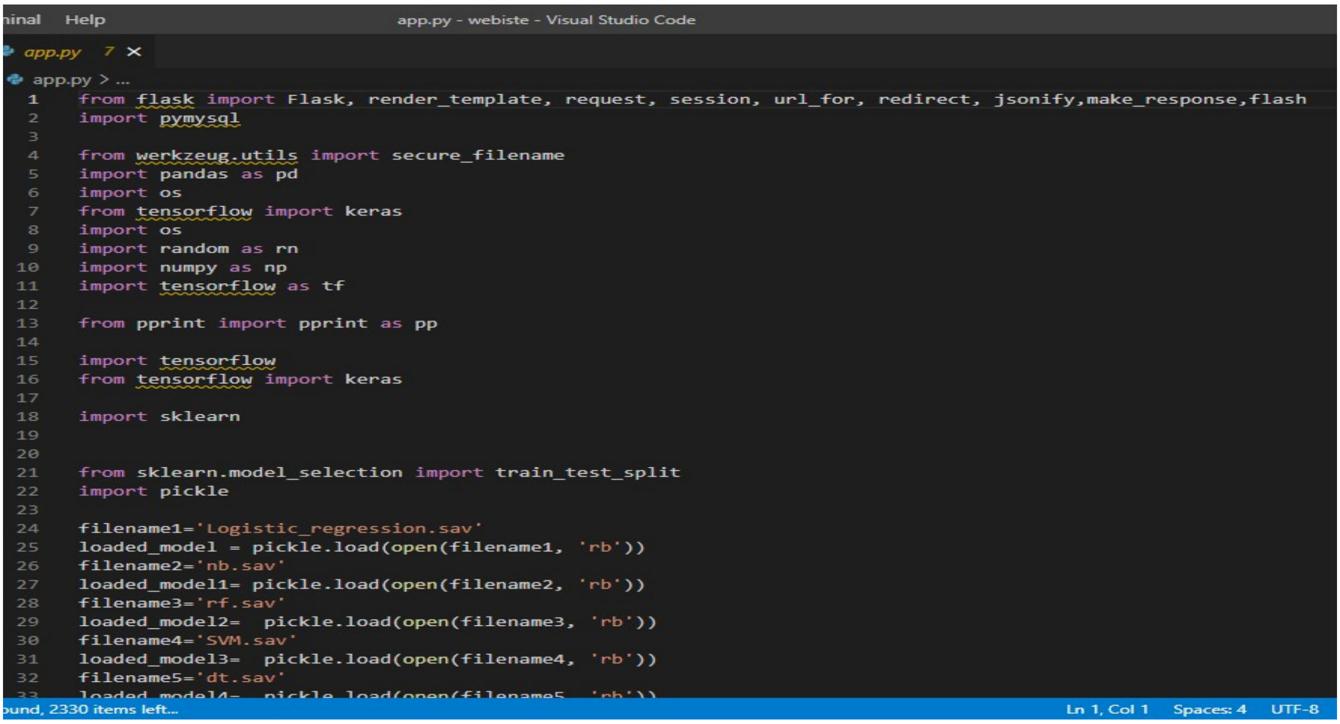
Conversion of pickle files

In fitting we have to generalize the similar data on which we had divided into 2 part test and train. And find the total accuracy of Decision tree algorithm . after that we have displayed the confusion matrix as well. And then converted this result in a pickle format.



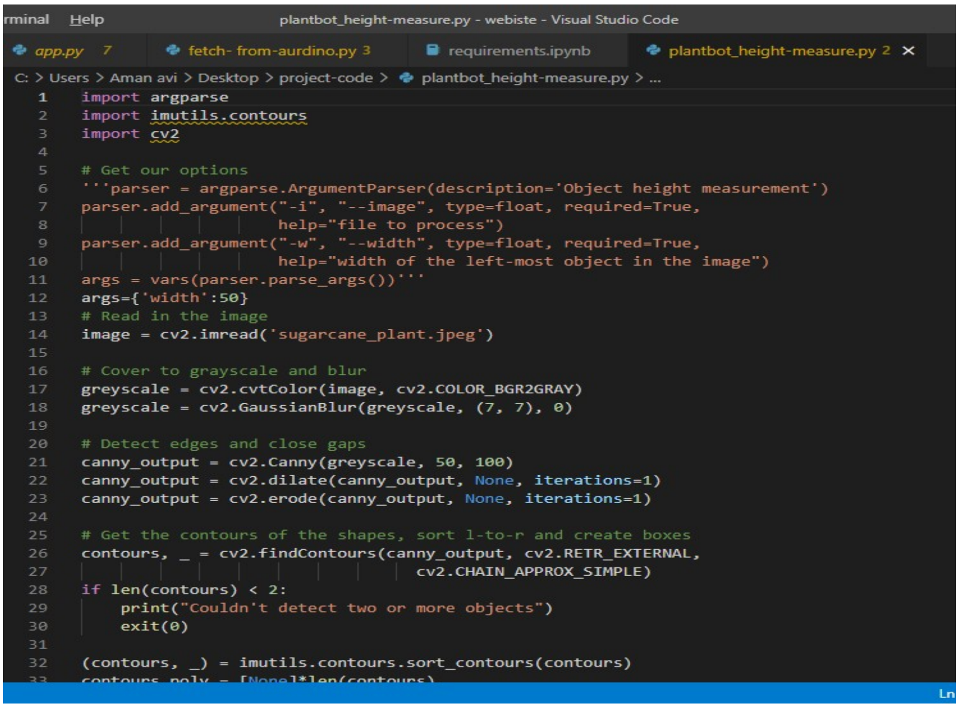
Measure the sensor's Input

in this block we had fetching or catching the data from arduino which is microcontroller is connected to the various sensor & that sensor data send to the cloud. We are using thingspeak cloud for processing the sensor data. To send the data on ThingSpeak cloud here we are using API Key to end the data to the cloud.



Flask API

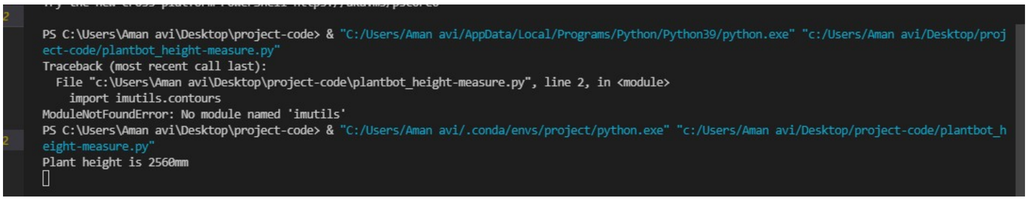
Flask is a web framework for Python, meaning that it provides functionality for building web applications, including managing HTTP requests and rendering templates. In this section, we will create a basic Flask application. This is a basic terminology of flask API . so definitely we are using Flask to deploy or run our GUI Part . and after that we are going to attach our model in the flask block and run on localhost/ 8000 address port.



```
rminal   Help        plantbot_height-measure.py - webiste - Visual Studio Code
app.py 7   fetch- from-aurdino.py 3   requirements.ipynb   plantbot_height-measure.py 2 ×
C: > Users > Aman avi > Desktop > project-code > plantbot_height-measure.py > ...
import argparse
import imutils.contours
import cv2

# Get our options
'''parser = argparse.ArgumentParser(description='Object height measurement')
parser.add_argument("-i", "--image", type=float, required=True,
                    help="file to process")
parser.add_argument("-w", "--width", type=float, required=True,
                    help="width of the left-most object in the image")
args = vars(parser.parse_args())'''
args={'width':50}
# Read in the image
image = cv2.imread('sugarcane_plant.jpeg')

# Cover to grayscale and blur
greyscale = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
greyscale = cv2.GaussianBlur(greyscale, (7, 7), 0)

# Detect edges and close gaps
canny_output = cv2.Canny(greyscale, 50, 100)
canny_output = cv2.dilate(canny_output, None, iterations=1)
canny_output = cv2.erode(canny_output, None, iterations=1)

# Get the contours of the shapes, sort l-to-r and create boxes
contours, _ = cv2.findContours(canny_output, cv2.RETR_EXTERNAL,
                               cv2.CHAIN_APPROX_SIMPLE)
if len(contours) < 2:
    print("Couldn't detect two or more objects")
    exit(0)

(contours, _) = imutils.contours.sort_contours(contours)
contours_poly = [None]*len(contours)
```

Height Measures

In this block we have to go for to find the height of sugarcane plant. Importing the Open - CV library , argparse , imutils and then intialize the width of args . in ths we have to give image path of sugarcane plant , after that converting the image into greyscle image, remove the blurriness of the image and remove the close gap in between in image by canny operation.



```
PS C:\Users\Aman avi\Desktop\project-code> & "C:/Users/Aman avi/AppData/Local/Programs/Python/Python39/python.exe" "c:/Users/Aman avi/Desktop/project-code/plantbot_height-measure.py"
Traceback (most recent call last):
  File "c:\Users\Aman avi\Desktop\project-code\plantbot_height-measure.py", line 2, in <module>
    import imutils.contours
ModuleNotFoundError: No module named 'imutils'
PS C:\Users\Aman avi\Desktop\project-code> & "C:/Users/Aman avi/.conda/envs/project/python.exe" "c:/Users/Aman avi/Desktop/project-code/plantbot_height-measure.py"
Plant height is 2560mm
```

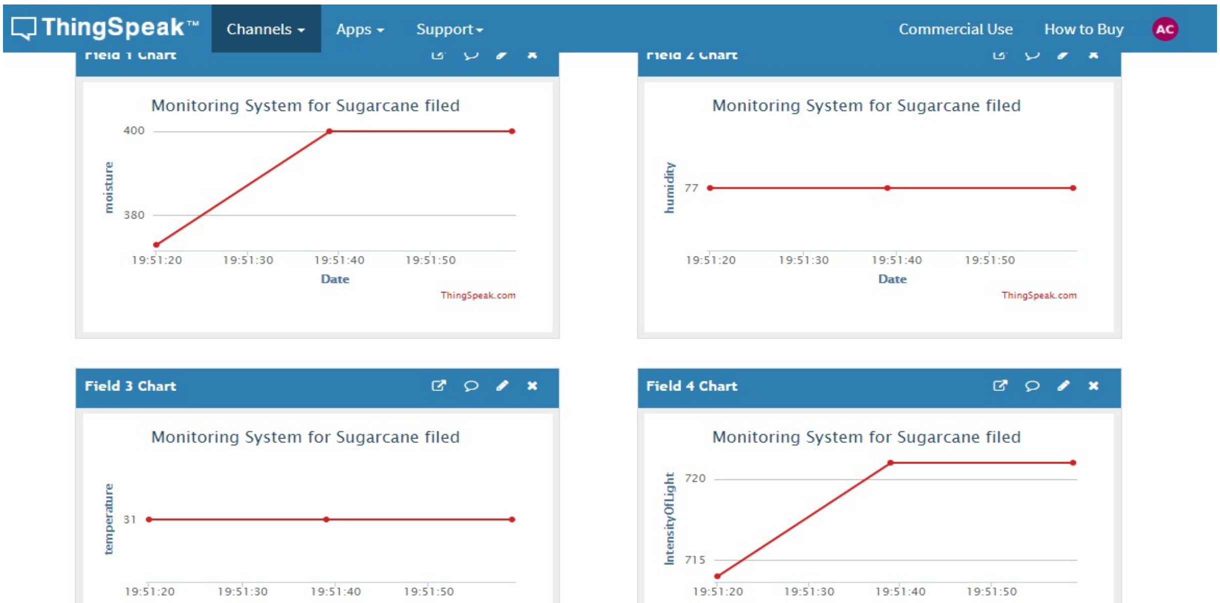
CCHAPTER 8

RESULT AND DISCUSSION

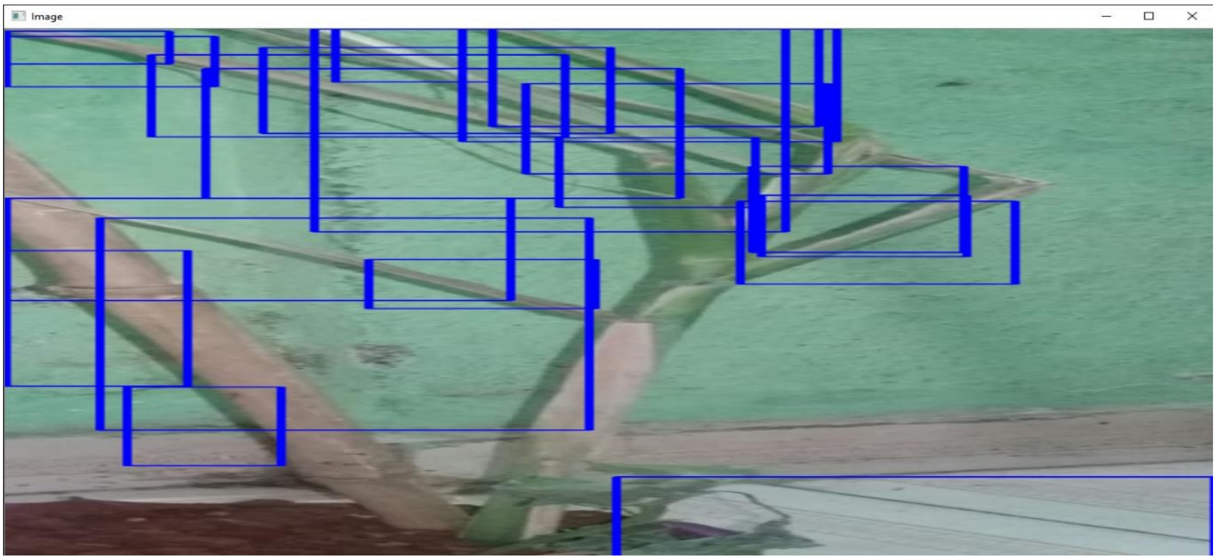
8. RESULT AND DISCUSSION

1. Objective Achieved

- All the sensor's data successfully send to the cloud.
- For each sensor's data we have assigned one particular filed.
- We can analyze the data with the help of graph on ThingSpeak cloud.
- We have successfully created the dataset and train the model with the help of machine learning algorithm.
- On GUI user can register itself , login the account and then access the services.
- We can also see here the accuracy of ML algorithm.
- By using this model we can easily predict the sugarcane yield prediction using some parameters like temperature, humidity, moisture, LDR , plant height and width .



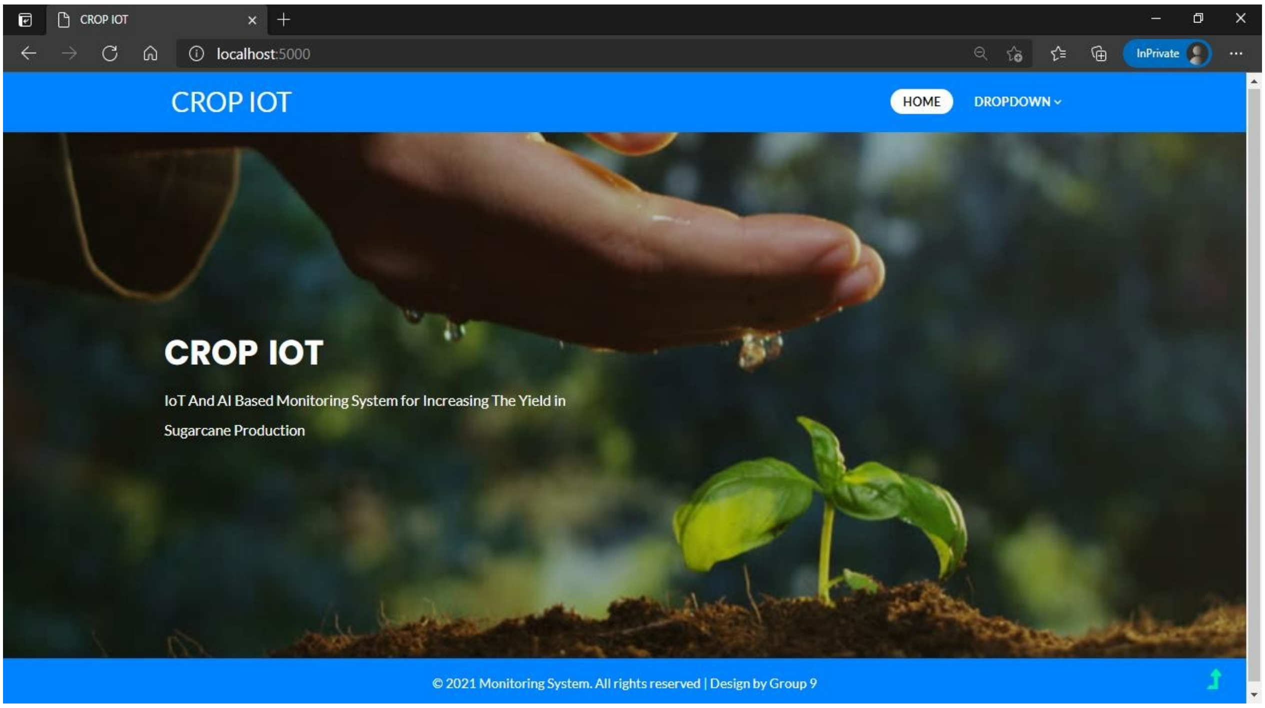
Sensor's Output on ThingSpeak Cloud



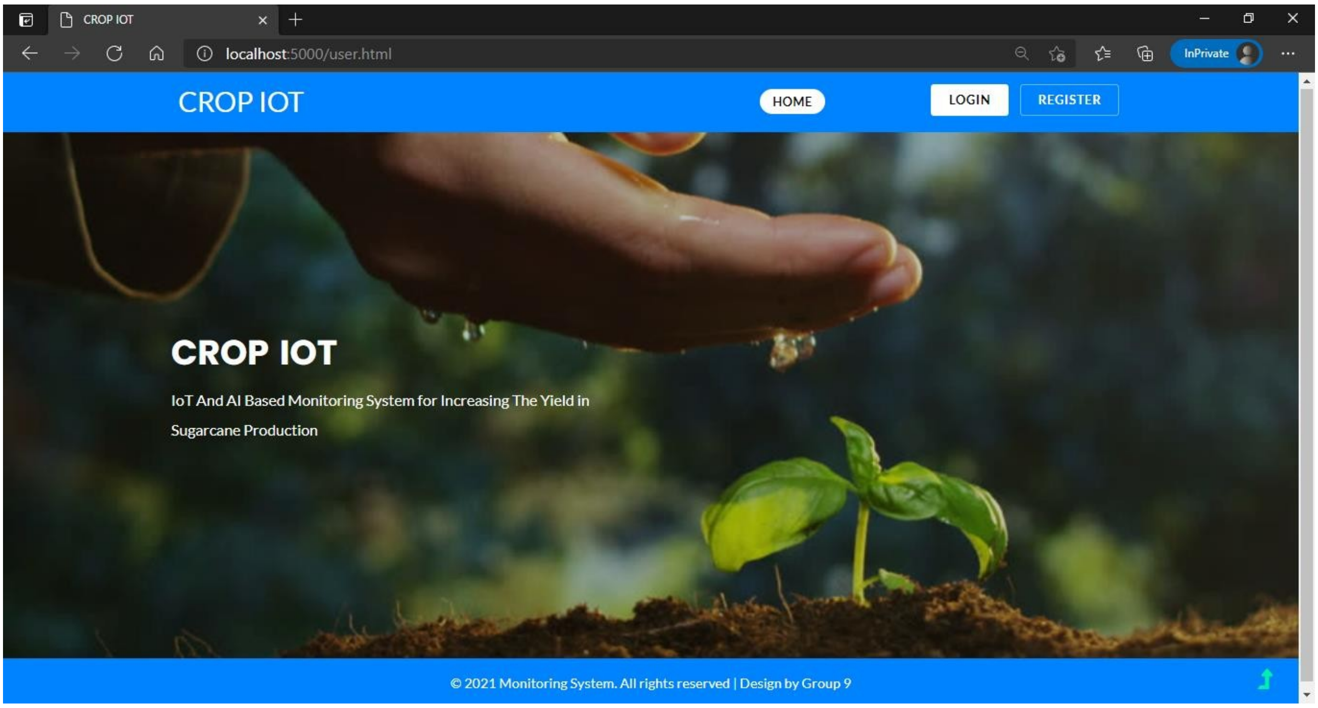
Height Measure using Open-CV

8.2.GUI Design:

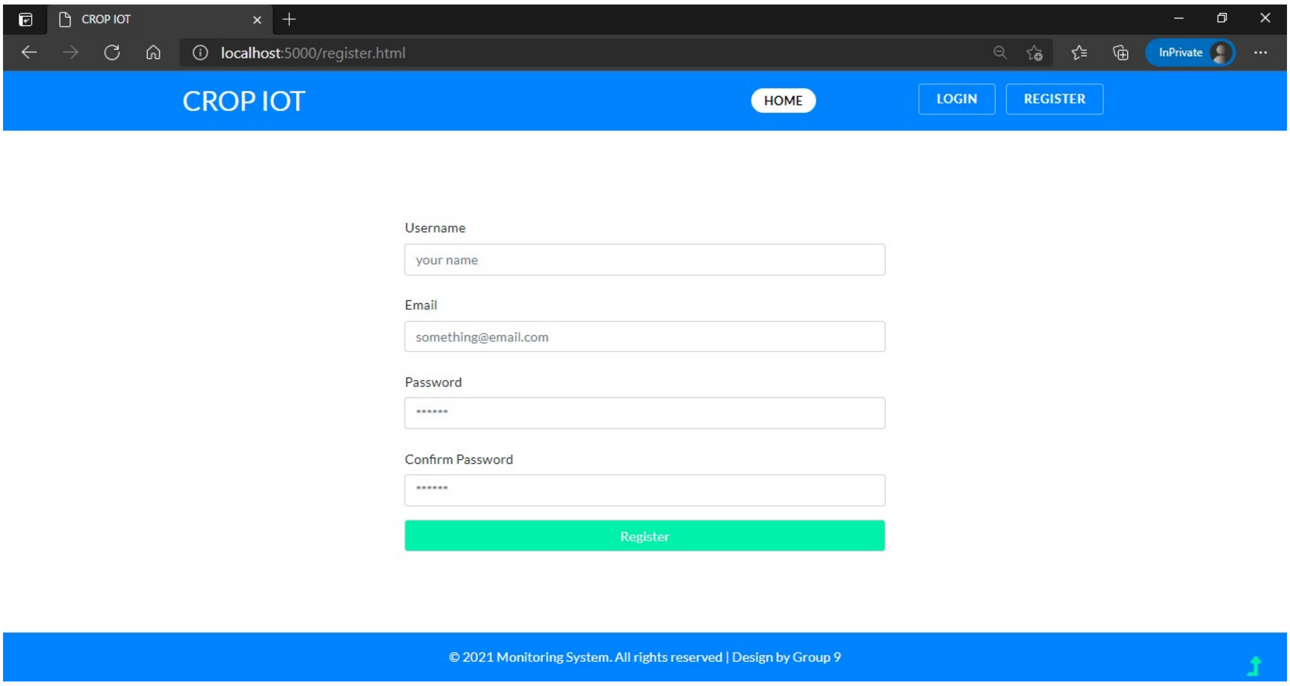
1. Home Page:



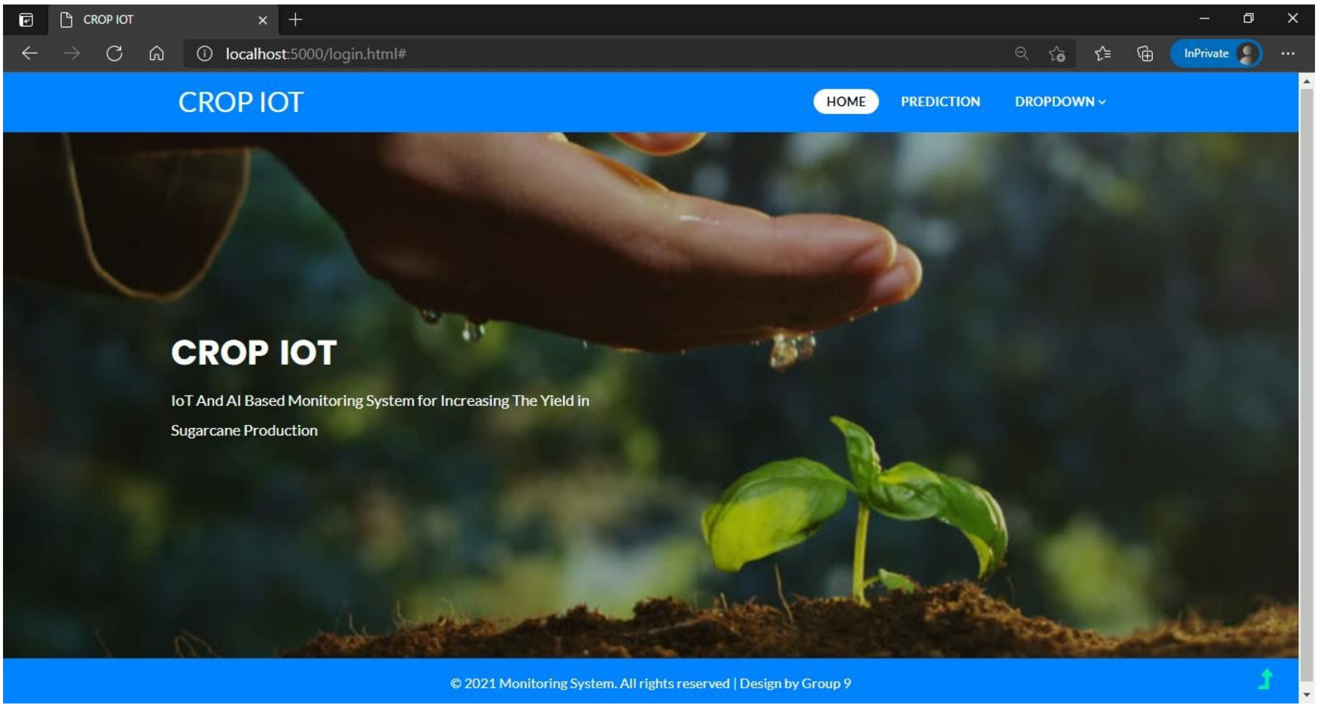
2. Login page:

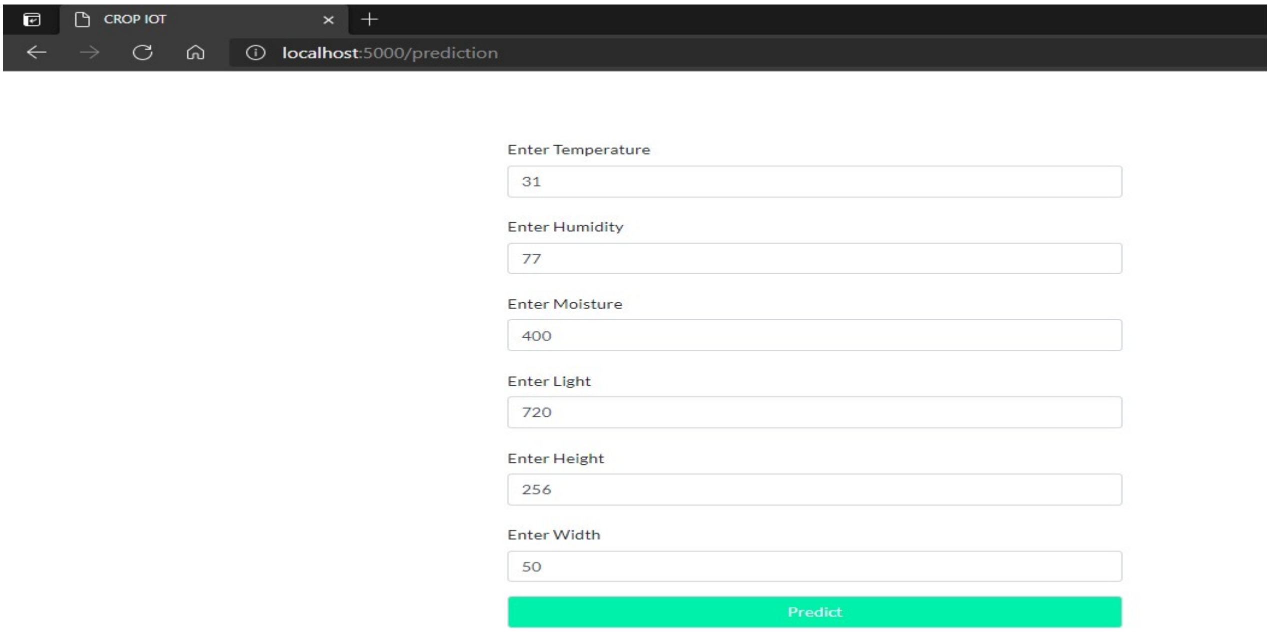


3. Register Page:

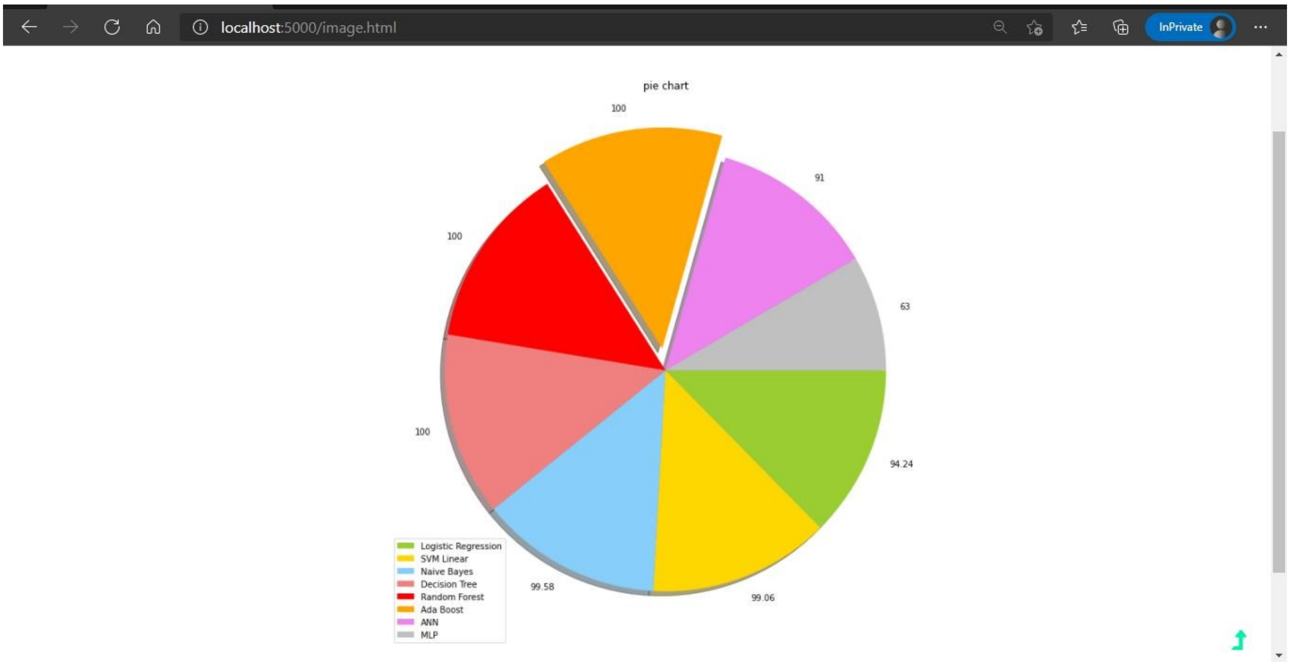


4. Prediction page:





5. Accuracy of ML Model:



CONCLUSION

We have to find the best favorable conditions for the marigold plant by data visualization and later on, this trained model can be used to detect the rate of growth of sugarcane plant just supplying it the environmental physical conditions. Not only we can detect the best favourable conditions for particular crops like millet, chilies, tea, coffee etc., but also we could prepare a dataset using IoT which could tell us the best suitable plant for a particular soil and climatic condition.

7.1 Future Scope

- The future work is trying to improve the topology structure to make all nodes communicate with each other, also to improve the stability of wireless sensors in communication by better software and hardware design.
- Data analytic according to area and can be used for future prediction references.
- Clean Solar Energy can be used to operate the whole farm.
- Damage cause by predictors also can be reduced using different devices.

REFERENCES:

1. Richa Singh, Sarthak Srivastava, Rajan Mishra, "AI and IoT Based Monitoring System for Increasing the Yield in Crop Production ". 2020 International Conference on Electrical and Electronics Engineering (ICE3-2020).
2. **https://ieeexplore.ieee.org/document/9122894**
3. https://www.researchgate.net/publication/343934843_AI_and_IoT_Based_Monitoring_System_for_Increasing_the_Yield_in_Crop_Production
4. Anju Upadhyay. Sanjay Kumar Maurya, "Protecting the Agriculture field by IoT Application", 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARK).
5. Kasara Sai Pratyush Reddy, Y. Mohana Roopa, Kovvada Rajeev L.N., Narra Sai Nandan, "IoT based Smart Agriculture Using Machine Learning", 2020 Second International Conference on Inventive Research in Computing Application (ICIRCA).
6. M.K. Gayatri, J. Jayasakthi and G.S. Anandhamala, "Providing Smart Agriculture Solutions to Farmers for Better Yielding Using IoT", IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development.
7. Sourabh Patil, Kunal Goswami and Kuldip Singh, "Analysis of Crop Selection on Basis of Dynamic Environmental Factors and Live Market Condition using Techniques of Machine Learning and IOT" in , Ashish Ramdasi Computer Engineering, Sinhgad Academy of Engineering.
8. R. Minerva, A. Biru and D. Rotondi, "Towards a definition of the Internet of Things (IoT)", IEEE Internet Initiative, vol. 1, 2015.
9. J.P. Singh, M.P. Singh, Rakesh Kumar and Prabhat Kumar, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", International Journal on Engineering Technology, May 2015.
10. N. Putjaika, S. Phusae, A. Chen-Im, P. Phunchongharn and K. Akkarajitsakul, "A control system in an intelligent farming by using arduino technology", 2016 Fifth ICT International Student Project Conference (ICT-ISPC), pp. 53-56, 2016.
11. Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic and Marimuthu Palaniswami, "Internet of Things (IoT): A vision architectural elements and future directions" in Future Generation Computer Systems, ELSEVIER, vol. 29, pp. 1645-1660, 2013.
12. S. Djodiltachoumy, "A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, March 2017.
13. Siti Khairunniza-Bejo, Samihah Mustaffha and Wan Ishak Wan Ismail, "Application of Artificial Neural Network in Predicting", Journal of Food Science and Engineering, January 2014.

ACKNOWLEDGEMENT

It is pleasant task to express gratitude to all those who contributed in many ways to this project and made it an unforgettable experience for us. First of all, we would like to thank our project guide **prof**. **Sonali Deshpande &** project coordinator **Prof. Deepti Vijay Chandran** . This work would not have been possible without her guidance, support and encouragement.

We are highly indebted to Dr Sunil Chavan, Principal, **Smt. Indira Gandhi College of Engineering**,

Prof. Sonali Deshpande, Head of the Department, **Computer Department** and also all the faculty members.

We are thankful to lord for keeping us energized that every end seems to be new beginning.

SURVEY PAPER

AI and IoT Based Monitoring System for Increasing the Yield in Crop Production & Security: A Survey

Amaan Choudhary [#1]Azharsaied Qureshi[#2], Bhaskar Ananthraman [#3], Prof. Sonali Deshpande[*4]
[#]Student, Dept. of Computer Engineering, Smt. Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India
[*]Professor, Dept. of Computer Engineering, Smt. Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India
[1] 2017c16@sigce.edu.in, [2]2016c40@sigce.edu.in,
[3] 2017c08@sigce.edu.in, [4]sonali.deshpande@sigce.edu.in

***Abstract* - Artificial Intelligence (AI) and Internet of things (IoT) based monitoring systems are in great demand and gives a precise extraction and analysis of data. In this system, the system is performed on a sugarcane plant to detect the most suitable conditions for plant growth. The philosophy behinds this work is to reduce the risks in agriculture and to promote smart farming practices. The effect of physical conditions like humidity, temperature, soil temperature and moisture and light intensity on the plant growth, is monitored using IoT based monitoring system. The data responsible for the plant growth is obtained using different sensors units like DHT11, LDR, and DS18B20, Soil Moisture sensors, single-board microcontrollers and Application Programming Interfaces (APIs). The further analysis of the extracted parameters is done using different Machine Learning (ML) algorithms. This system will also help to protect the crop from the cattle. In India, land owned by the farmer is minimal in size. Therefore, the safety of the crop field is essential. This system will be uses passive infrared sensors to detect the motion of cattle in the fields. This system will provides smart irrigation system which predicts the water requirement for a crop, using machine learning algorithm.**

Keywords- Agriculture, Smart farming, KNN Prediction, Node MCU, IOT, Sensors, Machine Learning, Analytics.

I.INTRODUCTION

Agriculture is the main source of our food and most people in India are dependent on agriculture as their prime source of income. According to the United Nations Food and Agricultural Organization, almost 800 million people are chronically hungry and 2 billion suffer micronutrient deficiencies. Further, another report from the State of Food Security and Nutrition in the World 2019 showed, 194.4 million people are undernourished in India. It is high time where the agriculture sector needs to be digitized and smart. Robert J. Mc Queen et al. (1995) [1], predicted the abilities of automatic process which could the combination of various systems. These systems could be applied in the farming sector for the welfare of humankind.

This era of 21st century focusses on automatic technologies like Internet of Things (IoT), Machine Learning (ML), and Data Science (DS). However, the application of these technologies in the agriculture domain is challenging. Because of regular variations in physical and chemical conditions of the surrounding, increasing the yield in the production of the crop is a real-time challenge and a problem needed to be solved. Hence, continuous monitoring and strict management is needed for agriculture parameters like humidity, temperature of the environment as well as of soil and light intensity, etc.

The major issues for the failure of crop production and low yield is lack of nutrients and suitable environmental conditions. Thus, the foremost motivation of this work is to provide an agricultural research solution. For this the ML algorithms is applied on the data generated by the technique of the Internet of Things. Here, we first create a dataset with the help of the Internet of things and the data collected is stored on the cloud. The data has features that consist of physical properties of surrounding like soil humidity, soil temperature and intensity of light. The target variable for our dataset is the rate of increase in height and width respectively which is the growth rate of the plant. After this the supervised machine learning algorithm are applied on the data which will give the best conditions for the plant for maximum growth.

II. LITERATURE SURVEY

IoT solutions are focused on helping farmers close the supply demand gap, by ensuring high yields, profitability, and protection of the environment. The approach of using IoT technology to ensure optimum application of resources to achieve high crop yields and reduce operational costs is called precision agriculture. IoT in agriculture technologies comprise specialized equipment, wireless connectivity, software and IT services.BI Intelligence survey expects that the adoption of **IoT devices in the agriculture industry will reach 75 million in 2020**, growing 20% annually. At the same time, **the global smart agriculture market size is expected to triple by 2025**, reaching $15.3 billion (compared to being slightly over $5 billion back in 2016).

Smart farming based on **IoT technologies enables growers and farmers to reduce waste and enhance productivity** ranging from the quantity of fertilizer utilized to the number of journeys the farm vehicles have made, and enabling efficient utilization of resources such as water, electricity, etc. IoT smart farming solutions is a system that is built for monitoring the crop field with the help of sensors (light, humidity, temperature, soil moisture, crop health, etc.) and automating the irrigation system. **The farmers can monitor the field conditions from anywhere.** They can also select between manual and automated options for taking necessary actions based on this data. For example, if the soil moisture level decreases, the farmer can deploy sensors to start the irrigation. Smart farming is highly efficient when compared with the conventional approach.

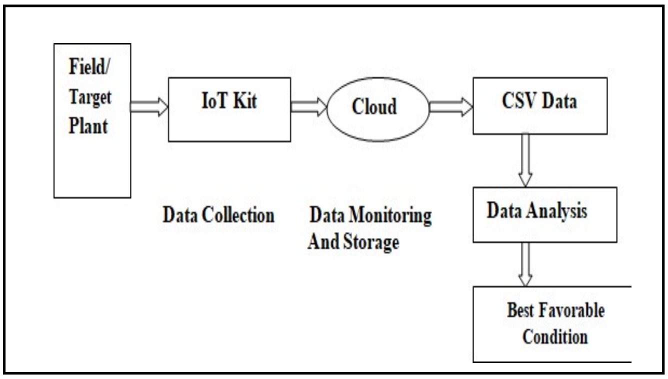
IoT have the potential to transform agriculture in many aspects and these are the main ones. **Data collected by smart agriculture sensors,** in this approach of farm management, a key component are sensors, control systems, robotics, autonomous vehicles, automated hardware, variable rate technology, motion detectors, button camera, and wearable devices. This data can be used to track the state of the business in general as well as staff performance, equipment efficiency. The ability to foresee the output of production allows to plan for better product distribution.

Agricultural Drones Ground-based and aerial-based drones are being used in agriculture in order to enhance various agricultural practices: crop health assessment, irrigation, crop monitoring, crop spraying, planting, and soil and field analysis. **Livestock tracking and geofencing** Farm owners can utilize wireless IoT applications to collect data regarding the location, well-being, and health of their cattle. This information helps to prevent the spread of disease and also lowers labour costs.

Smart Greenhouses a smart greenhouse designed with the help of IoT intelligently monitors as well as controls the climate, eliminating the need for manual intervention. **Predictive analytics for smart farming** Crop predication plays a key role, it helps the farmer to decide future plan regarding the production of the crop, its storage, marketing techniques and risk management. To predict production rate of the crop artificial network use information collected by sensors from the farm. This information includes parameters such as soil, temperature, pressure, rainfall, and humidity. The farmers can get an accurate soil data either by the dashboard or a customized mobile application. **Farmers have started to realize that the IoT is a driving force for increasing agricultural production in a cost-effective way.** Because the market is still developing, there is still ample opportunity for businesses willing to join in.

II. METHODOLOGY

The main aim of the suggested system is to take various sensor readings from the soil and predict the type of crop that is the most suitable to grow for that particular type of soil. First, let us understand the architecture of our system which will help us to understand our concept very much in-depth.



1. System Architecture

A. Capturing the plant's height and width:
To capture plants' height and width we used the concept of image processing and with the help of open CV library. The extracted height and width of the plant is stored in form of CSV file. Meanwhile, all the environmental parameters are collected and stored in the cloud by using IoT. These values also extracted from a cloud in CSV form.

B. Preparing the Dataset:
We have got the target variable data that was the plant's height and width and then we found the rate of plant growth by using the formula in equation 2,

Height/Width Growth Rate=$\partial x/\partial t$ (2)

Where, ∂x changes in height or width. And ∂t is the difference in time during which the properties are measured.

C. Data Analysis and Cleaning:
We analyzed and visualized the data and how valuable our features were and estimated a few predictions after features selection. Further, some of the extracted data was used for training the algorithm

D. Applying Algorithms:
After having the trained data, we applied supervised machine learning algorithms over it.

a) Artificial Neural Network:
An ANN is based on collection of connected unites or nodes called artificial neurons, which loosely model the neurons in a biological brain.

b) Linear Support Vector Classifier:
This model is essentially representing various classes in a multidimensional space hyperplane.

c) Decision Tree:
A decision tree is a classifier with feature marked internal nodes. Each node marked with possible output and leads to a lower-ranked decision node on various input features.

d) Random Forest:
This algorithm is used for classification as well as regression predictions. It creates different decision trees on data and gives predictions from them. Finally, provides the best solution to the problem for which it is employed.

e) Adaptive Boosting Classifier:
AdaBoost helps u combine multiple "week classifiers" into a single "strong classifier". The weak learners in AdaBoost are decision trees with a single split, called decision stumps. AdaBoost works by putting more weight on difficult to classify instances and less on those already handled well.

f) Feed Forward algorithm:
In this network, the information moves in only one direction - **Forward**- from the input nodes, through the hidden nodes. (If any) and to the output nodes.

g) Backpropagation algorithm:
The Backpropagation algorithm

calculates how much the final output values, o1 and o2, are affected by each of the weights. To do this, it calculates partial derivatives, going back from the error function to the neuron that carried a specific weight.

E. Hardware Components:

a) Soil Moisture Sensor

Soil moisture sensors measure the volumetric water content in soil. Since the direct gravimetric measurement of free soil moisture requires removing, drying, and weighing of a sample, soil moisture sensors measure the volumetric water content indirectly by using some other property of the soil, such as electrical resistance, dielectric constant, or interaction with neutrons, as a proxy for the moisture content.

b) LDR sensors

An LDR is a component that has a (variable) resistance that changes with the light intensity that falls upon it. This allows them to be used in light sensing circuits. Light Dependent Resistors (LDR) are also called photoresistors. They are made of high resistance semiconductor material.

c) DHT11

The DHT11 is a commonly used Temperature and humidity sensor. The sensor comes with a dedicated NTC to measure temperature and an 8-bit microcontroller to output the values of temperature and humidity as serial data. The sensor is also factory calibrated and hence easy to interface with other microcontrollers.

d) Node MCU

Node MCU is a low-cost open source IoT platform. It initially included firmware which runs on the ESP8266 Wi-Fi SoC from Espress if Systems, and hardware which was based on the ESP-12 module. Later, support for the ESP32 32-bit MCU was added.

IV. CONCLUSION

- We have to find the best favorable conditions for the marigold plant by data visualization and later on, this trained model can be used to detect the rate of growth of sugarcane plant just supplying it the environmental physical conditions.
- This technique has a great scope in the field of artificial farming.
- Not only we can detect the best favorable conditions for particular crops like millet, chilies, tea, coffee etc., but also we could prepare a dataset using IoT which could tell us the best suitable plant for a particular soil and climatic condition.

V. REFERENCES

1. Richa Singh, Sarthak Srivastava, Rajan Mishra, "AI and IoT Based Monitoring System for Increasing the Yield in Crop Production ". 2020 International Conference on Electrical and Electronics Engineering (ICE3-2020).
2. Anju Upadhyay. Sanjay Kumar Maurya, "Protecting the Agriculture field by IoT Application", 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARK).
3. Kasara Sai Pratyush Reddy, Y. Mohana Roopa, Kovvada Rajeev L.N., Narra Sai Nandan, "IoT based Smart Agriculture Using Machine Learning", 2020 Second International Conference on Inventive Research in Computing Application (ICIRCA).
4. https://ieeexplore.ieee.org/document/908723
5. https://ieeexplore.ieee.org/document/91

CERTIFICATES

PROF. SONALI DESHPANDE

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Prof. Sonali Deshpande

In recognition the publication of the manuscript entitled

AI and IoT based Monitoring System for Increasing the Yield in Crop Production & Security: A Survey

published in our Journal Volume 7 Issue 12 December 2020

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net

1.) **AMAN CHOUDHARY**

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Amaan Choudhary

In recognition the publication of the manuscript entitled

AI and IoT based Monitoring System for Increasing the Yield in Crop Production & Security: A Survey

published in our Journal Volume 7 Issue 12 December 2020

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net

2.) <u>AZHARSAIED QURESHI</u>

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Azharsaied Qureshi

In recognition the publication of the manuscript entitled

AI and IoT based Monitoring System for Increasing the Yield in Crop Production & Security: A Survey

published in our Journal Volume 7 Issue 12 December 2020

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net

3.)Bhaskar Ananthraman

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Bhaskar Ananthraman

In recognition the publication of the manuscript entitled

AI and IoT based Monitoring System for Increasing the Yield in Crop Production & Security: A Survey

published in our Journal Volume 7 Issue 12 December 2020

Impact Factor : 7.529

www.irjet.net

Editor in Chief

E-mail : editor@irjet.net