SMART AGRICULTURE

A BTP Report

By

K Sreenivasulu Reddy, S20190020217
Donda Vishal, 20190020210

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRI CITY

Date 18/12/2022

Final Report

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRI CITY

CANDIDATE’S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled “**SMART AGRICULTURE**” in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology SriCity, is an authentic record of my own work carried out during the time period from January 2022 to December 2022 under the supervision of Dr. Priyanka Dwivedi, Indian Institute of Information Technology Sri City, India.

I have not submitted the matter presented in this report for the award of any other degree of this or any other institute.

Signature of the student with date

K Sreenivasulu Reddy, 18/12/2021 TU

Donda Vishal, 12/12/2021

_ _

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of BTP Supervisor with date

ABSTRACT

Climate change (CC) clearly impacts food production, but risks on climatic suitability of agricultural areas for vegetable crops, their pests and associated natural enemies are largely explored. Farmers cultivating different crops need to adapt to an increase in the potential for outbreaks of pests favored by CC and disruption of biological control, yet, no attempt has been made to simultaneously evaluate CC effects on a crop-pest-natural enemy system for rice or any other crop. Here, we modelled the suitability of areas equipped with irrigation facilities (AEI) in 2080 for rice, tomato. Model results have a significant relationship with growth rates for the three species measured in outdoor experiments.

Precision agriculture is in trend nowadays. Precision agriculture is a modern farming technique that uses the data of soil characteristics, soil types, crop yield data, weather conditions and suggests the farmers with the most optimal crop to grow in their farms for maximum yield and profit. This technique can reduce the crop failures and will help the farmers to take informed decision about their farming strategy. In order to mitigate the agrarian crisis in the current status quo, there is a need for better recommendation systems to alleviate the crisis by helping the farmers to make an informed decision before starting the cultivation of crops. From this project we hope to say the chance of growing the crop at different locations in future, suggest a crop for high yield, good profit.

Contents

1. INTRODUCTION 5

2. LITERATURE SURVEY 7

3. CONTRIBUTIONS 8

4. METHODOLOGY 9

5. RESULTS 18

6. WORK DONE FOR FINAL EVALUATION 28

7. CONCLUSION & FUTURE WORK 28

8. CHALLENGES FACED 29

9. ACKNOWLEDGMENTS 30

10. REFERENCES 31

1. INTRODUCTION.

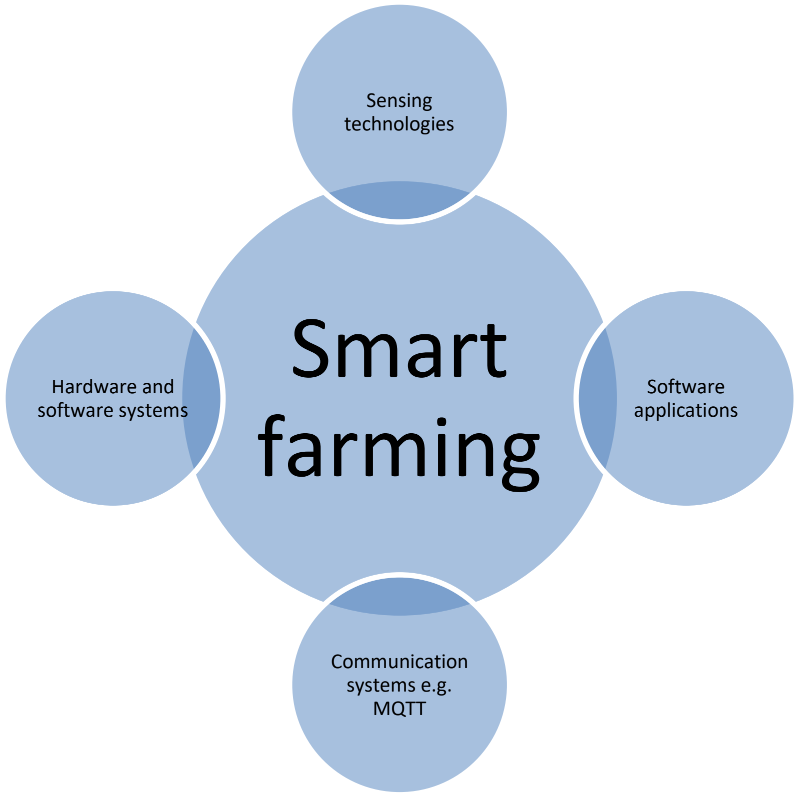
Agriculture is the practice of cultivating planta and livestock. Agriculture is the main source is raw materials. There are almost 12 types of agriculture: Arabic farming, Shifting cultivation, Commercial grain agriculture, etc. Word Agriculture is adapted from the Latin word ager 'field' and cultura 'growing'. Though the farmers have grown the same crops for centuries, the ever-changing weather conditions, soil fertility, pests and diseases etc. are affecting the yield of the crop.

Some Factors affecting agriculture are:

- Moisture levels are high importance to yields; thus, plants will not grow and develop with inadequate soil moisture.
- Humidity is important to make photosynthesis possible.
- Evaporating water can evacuate a lot of plant head and is an efficient way of cooling for a plant.
- Soil pH is important because it influences several soil factors affecting plant growth, such as

    1. Soil Bacteria
    2. Nutrient leaching
    3. Toxic elements
    4. Nutrient availability
    5. Soil structure

Throughout the years the farmers have grown the same crops for centuries. But to the ever-changing weather conditions (also future), soil fertility, pests and diseases, etc. factors are affecting the suitability and yield of the crop which gives rise to our project SMART AGRICULTURE. Smart farming is a management concept focused on providing the agricultural industry with infrastructure to leverage advanced technology – including big data, the cloud and the internet of things (IoT) – for tracking, monitoring, automating and analyzing operations.


Sensing technologies
Hardware and software systems
Smart farming
Software applications
Communication systems e.g. MQTT

2. LITERATURE SURVEY

- Climate-smart Agriculture is an approach for transforming and reorienting agricultural systems to support food security under the new realities of climate change. Widespread changes in rainfall and temperature patterns threaten agricultural production and increase the vulnerability of people dependent on agriculture for their livelihoods. Climate change disrupts food markets, posing population-wide risks to food supply. Threats can be reduced by increasing the adaptive capacity of farmers as well increasing resilience and resource use efficiency in agricultural production systems.

- Climate-smart agriculture (CSA) is an integrated approach to managing landscapes—cropland, livestock, forests and fisheries—that addresses the interlinked challenges of food security and accelerating climate change. CSA aims to simultaneously achieve three outcomes:
  - **Increased productivity:** Produce more and better food to improve nutrition security and boost incomes, especially of 75 percent of the world's poor who live in rural areas and mainly rely on agriculture for their livelihoods.
  - **Enhanced resilience:** Reduce vulnerability to drought, pests, diseases and other climate-related risks and shocks; and improve capacity to adapt and grow in the face of longer-term stresses like shortened seasons and erratic weather patterns.
  - **Reduced emissions:** Pursue lower emissions for each calorie or kilo of food produced, avoid deforestation from agriculture and identify ways to absorb carbon out of the atmosphere.

- Some of the work regarding this has been done earlier like automatic watering to plants, impacts of climate change on tomato, A suite of pest's attack tomato, including the two-spotted spider mite (Meck et al 2013, Gigon et al 2016), one of the world's most notorious pests (Vacante 2015), with a host range of more than 200 cultivated crops (Migeon and Dorkeld 2006–2017). Control costs exceed $400 million annually in pesticide sales alone (van Leeuwen et al 2015). The pest has developed resistance to close to 100 chemical active ingredients, more than any other plant-feeding arthropod (Michigan State University 2019). The sequencing of its genome illuminated a vast array of unique detoxification genes that contribute to its impressive ability to develop resistance (van Leeuwen et al 2010, Grbić et al 2011). How the crops are getting affected by pests, etc. Predicting the suitable areas to grow tomato for high yield. Also using some dataset predicting the crop as result which gives high yield.

3. CONTRIBUTIONS

- **K Sreenivasulu Reddy** (S20190020217) – Code for Sensors, Hardware Integration, Future Climate Forecasting (FCF), Estimating Crop Growing Locations, **Training and Testing data, Prediction**.
- **Donda Vishal**(S20190020210) – MQTT Code, Data for FCF, clipped data to INDIAN subcontinent, Encapsulation Design, 3D Printing, **Data Cleaning, Correlation Analysis, SVM**.

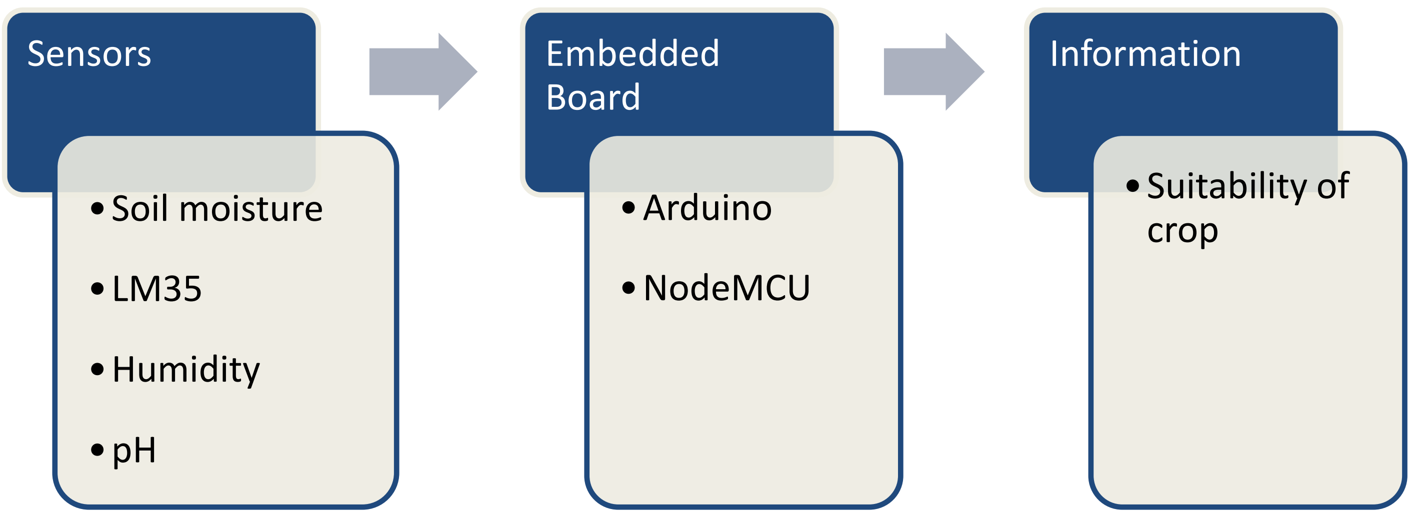
Problem Statement:

Our Aim is to Collect data from field to maintain continuous monitoring on the field, Auto Irrigation to the plants, Climate Weather Forecasting, Chance of crop growth at future climatic conditions, **Crop Prediction for good yield production**.

4. METHODOLOGY

In this project, we want to collect data from the crop location, for that we

- Plant Sensors like soil moisture, humidity, temperature, pH, etc. in the field and collect the data of various factors.
- From the collected sensors data suitable crop is suggested for better yield



- Planted sensors like LM-35, humidity, Soil Moisture, pH in the field.
- Interconnected these sensors with embedded boards like Arduino and NodeMCU.
- Used MQTT protocol to transfer information to cloud.

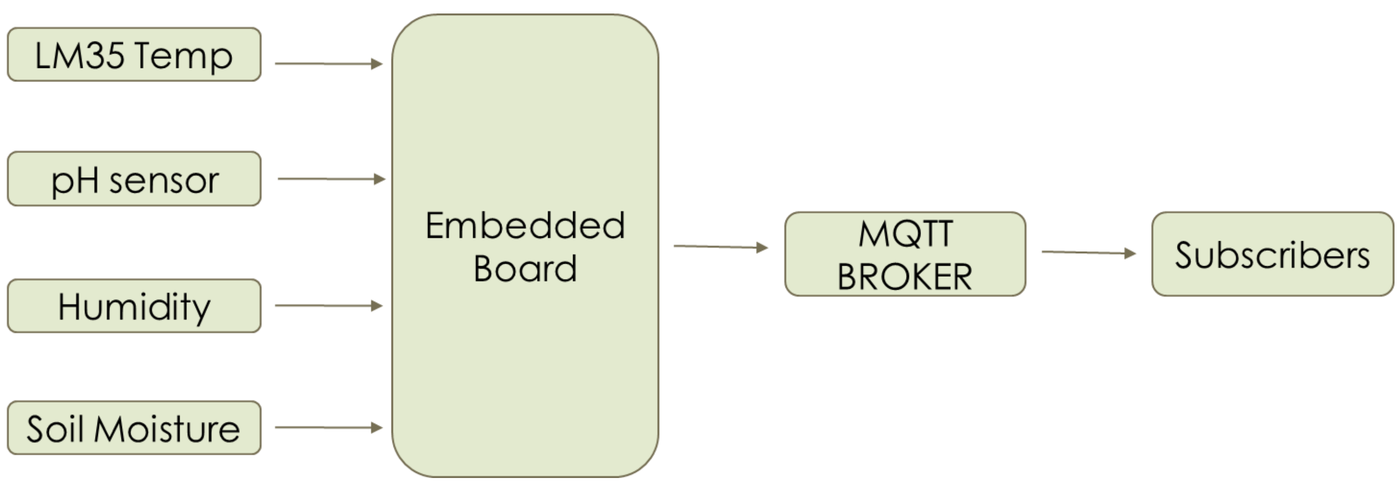
MQTT relies on the TCP protocol for data transmission. A variant, MQTT-SN, is used over other transports such as UDP or Bluetooth. The

MQTT protocol defines two types of network entities: a message broker and a number of clients. An MQTT broker is a server that receives all messages from the clients and then routes the messages to the appropriate destination clients. An MQTT client is any device (from a micro controller up to a fully-fledged server) that runs an MQTT library and connects to an MQTT broker over a network.

Information is organized in a hierarchy of topics. When a publisher has a new item of data to distribute, it sends a control message with the data to the connected broker. The broker then distributes the information to any clients that have subscribed to that topic. The publisher does not need to have any data on the number or locations of subscribers, and subscribers, in turn, do not have to be configured with any data about the publishers.

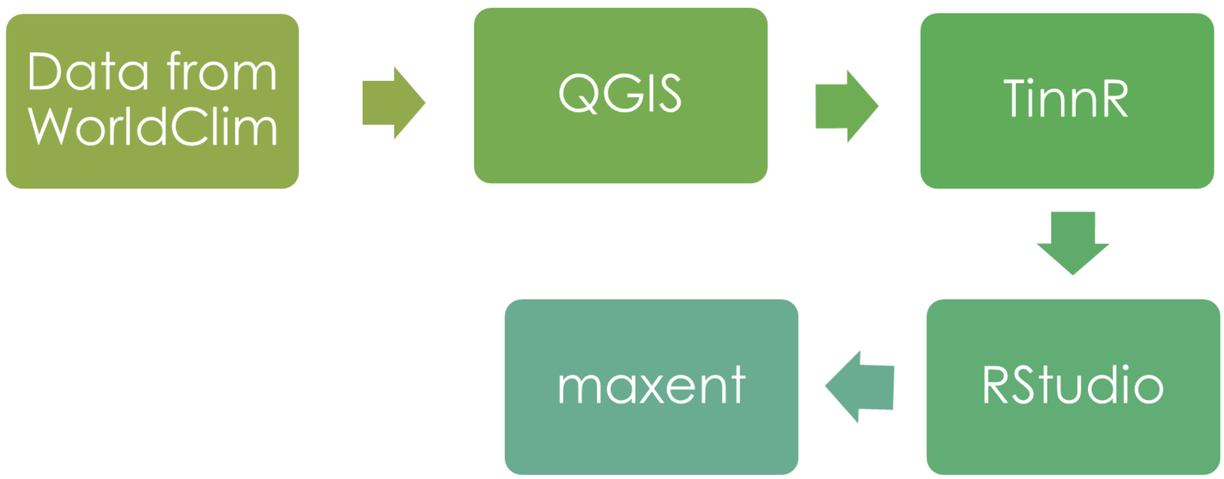
- Fixed a threshold value for soil moisture.
- Depending upon the moisture level of the soil, pump motor is used for auto irrigation.

Block Diagram:



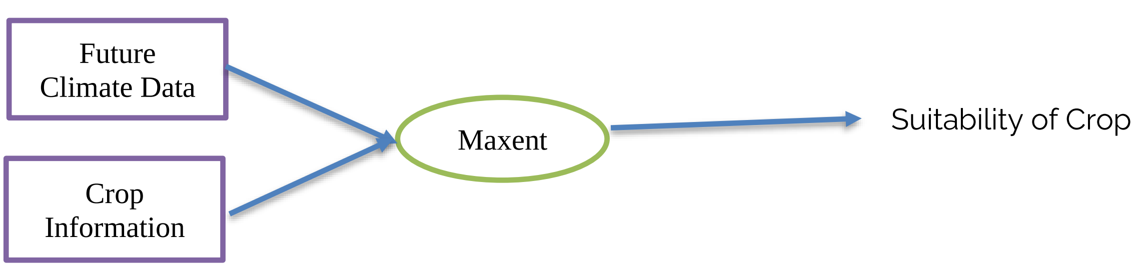
- Future climatic information in tiff format is downloaded from WorldClim.
- Clipped data to Indian subcontinent.
- Analyzed future climatic conditions.

Future Climatic Forecasting:



- WorldClim is a well-known database which has a set of global climate layers (climate grids) with a spatial resolution of 1 sq km.
- Collected climatic variables data @1970-2000 and @2081-2100.
- Used Qgis platform which helps in visualizing the data.
- Eliminated all countries other than India and converted it into single band raster format using R programming.
- Uploaded the crop information and future climate data to maxent which as result gives the plot of suitability of crop at different locations.

Block Diagram:

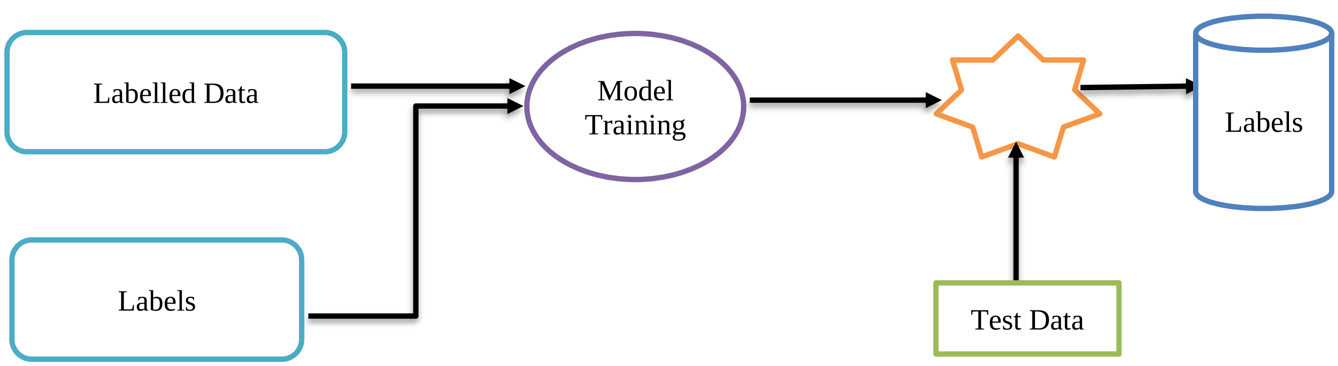


- Collected the data set which contains attributes like temperature, pH, Humidity, rainfall information, N, P, K (which are primarily responsible for the growth of plant).

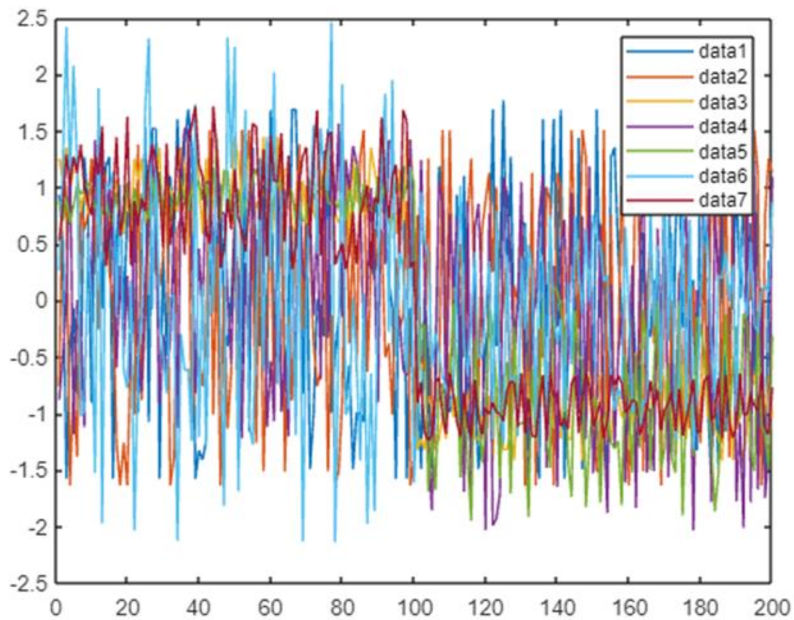
	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

- Crops are mentioned as the labels (classes).
- Performed Data Cleaning.
- Checked whether any outliers present in the data.
- Trained data using some supervised ML algorithms like SVM, Decision Tree, Logistic regression, XGBoost.

Block Diagram:



- Normalized data



- Five Number Summary

	N	P	K	temperature	humidity	ph	rainfall
count	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000
mean	50.551818	53.362727	48.149091	25.616244	71.481779	6.469480	103.463655
std	36.917334	32.985883	50.647931	5.063749	22.263812	0.773938	54.958389
min	0.000000	5.000000	5.000000	8.825675	14.258040	3.504752	20.211267
25%	21.000000	28.000000	20.000000	22.769375	60.261953	5.971693	64.551686
50%	37.000000	51.000000	32.000000	25.598693	80.473146	6.425045	94.867624
75%	84.250000	68.000000	49.000000	28.561654	89.948771	6.923643	124.267508
max	140.000000	145.000000	205.000000	43.675493	99.981876	9.935091	298.560117

Checking if any outliers present in temperature

Q1 = 22.77

Q2 = 25.6

Q3 = 28.56

Min = 8.825

Max = 43.675

IQR = Q3-Q1

= 5.79

Outlier: {Q1 - 1.5 * IQR, Q3 + 1.5 * IQR}

= {5.79,37.245}

So, no outliers present.

- Correlation between two attributes, correlation helps which two attributes are strongly related and affects the growth of the crop simultaneously.
- Trained dataset using SVM, Logistic Regression, Decision Tree, XGBoost Algorithms.
- Compared accuracy of all classifiers.
- Selected a classifier which gives better output with less error rate.

SVM Classifier:

Support Vector Machine or SVM is one of the most popular Supervised machine learning algorithms, Goal of SVM is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in future. This best decision boundary line is called hyperplane.

Logistic Regression:

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.

Decision Tree:

Decision Tree is a Supervised learning technique. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node. For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree.

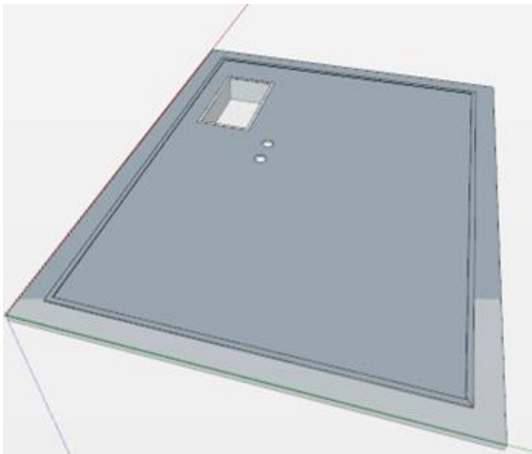
XGBoost:

XGBoost stands for Extreme Gradient Boosting.

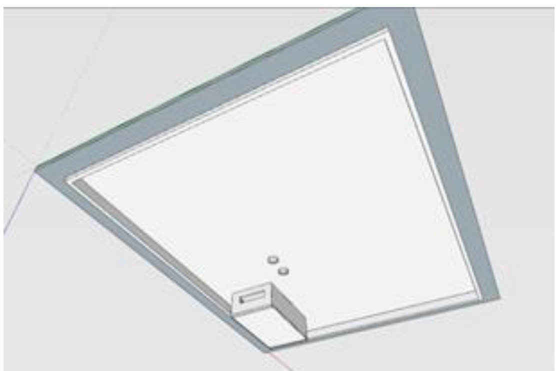
In this algorithm, decision trees are created in sequential form. Weights play an important role in XGBoost. Weights are assigned to all the independent variables which are then fed into the decision tree which predicts results. The weight of variables predicted wrong by the tree is increased and these variables are then fed to the second decision tree. These individual classifiers/predictors then ensemble to give a strong and more precise model. It can work on regression, classification, ranking, and user-defined prediction problems.

- **Encapsulation of PCB**
  - Simulator: SketchUp Pro 2022.
  - SKP -> STL -> Gcode file.
  - 3D filament (PLA PRO+) material is used.
  - Color – Daisy White, Diameter – 1.75mm.
  - Dimensions of box: 11.5cm x 15cm x 5mm.

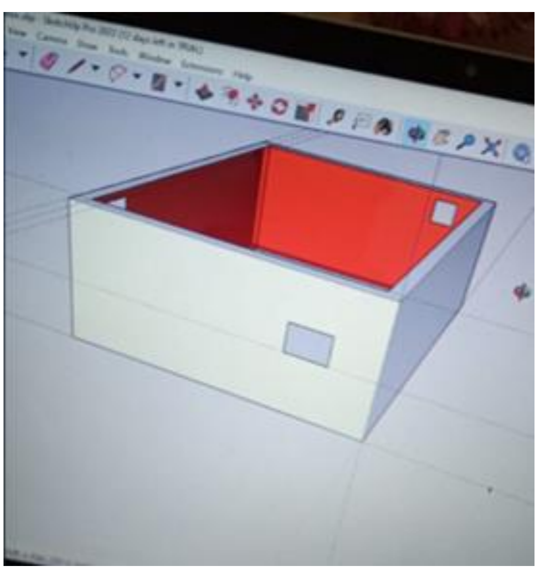
1. Top Layer(top-view)

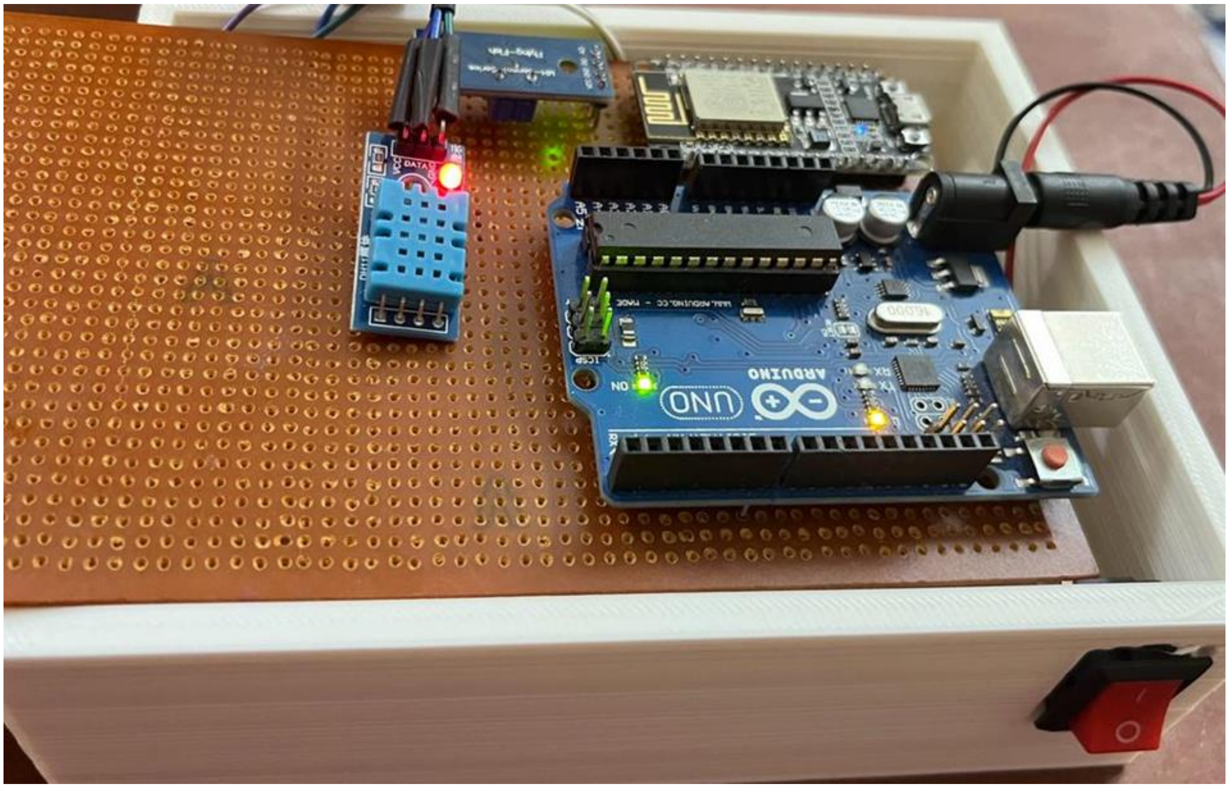


2. Top Layer(bottom-view)

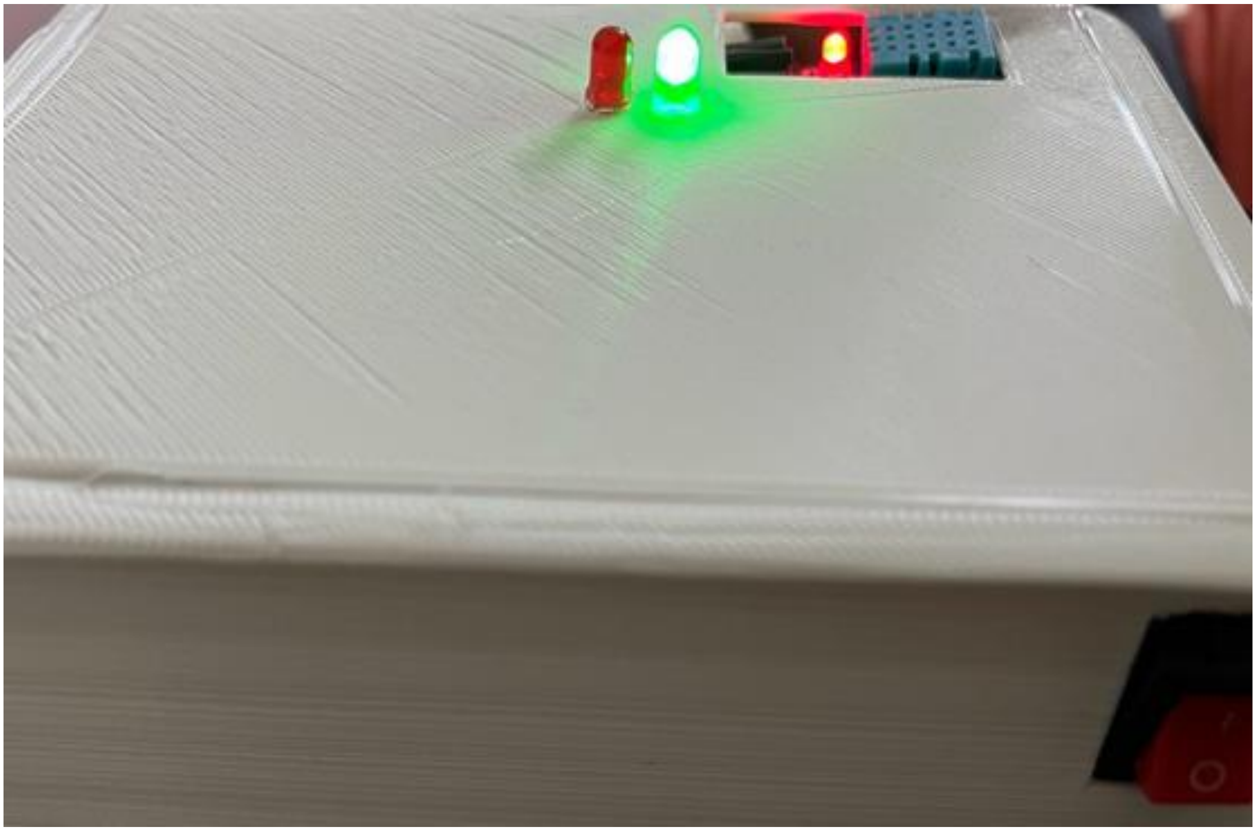


3. Box





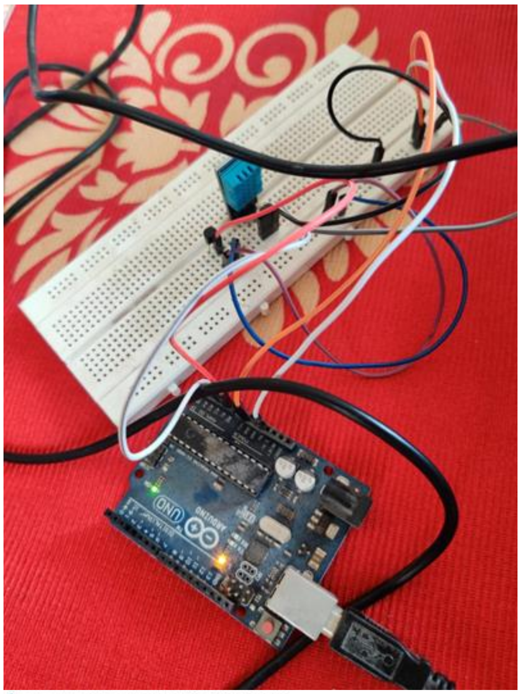
Sensors Integration on PCB

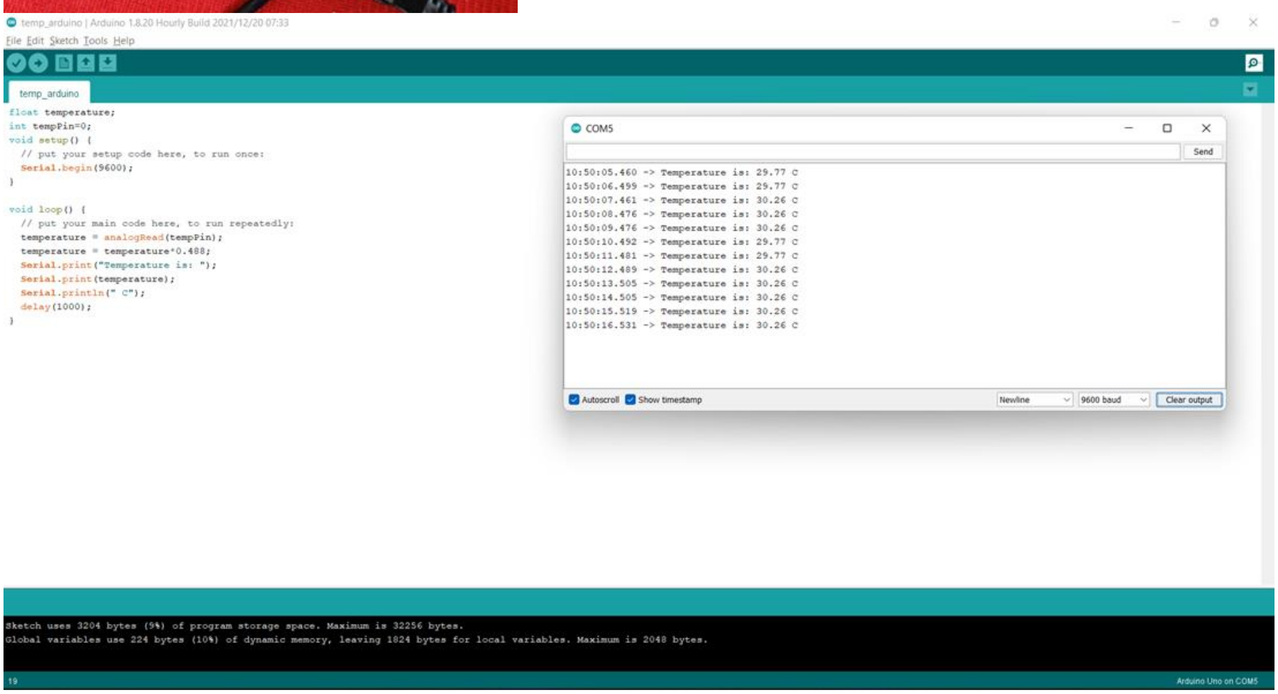


Complete Box

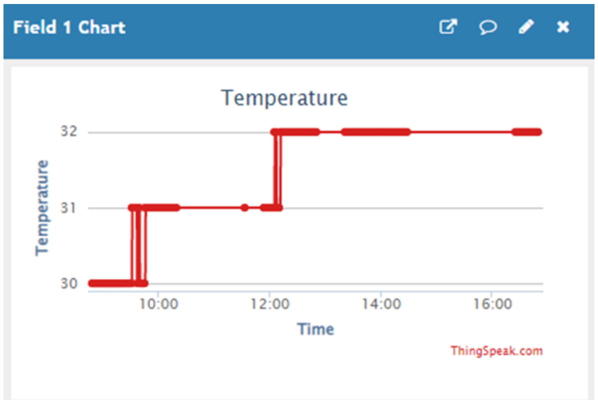
5. RESULTS

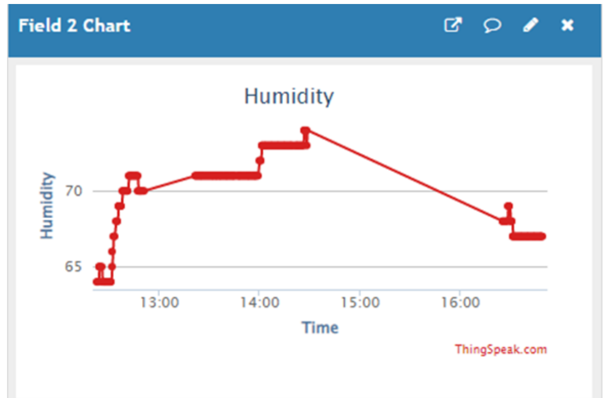
- Results @Phase-1





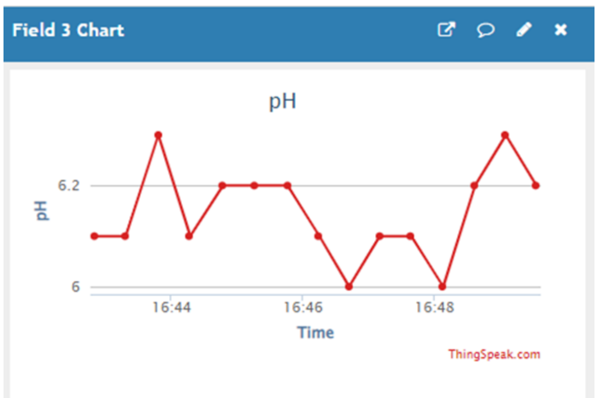
- Results @Phase-2

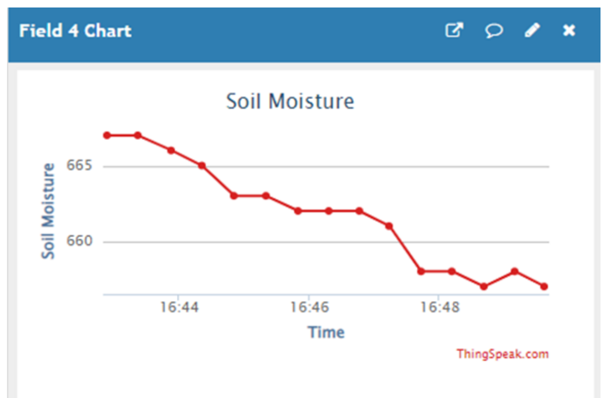




Topic-1

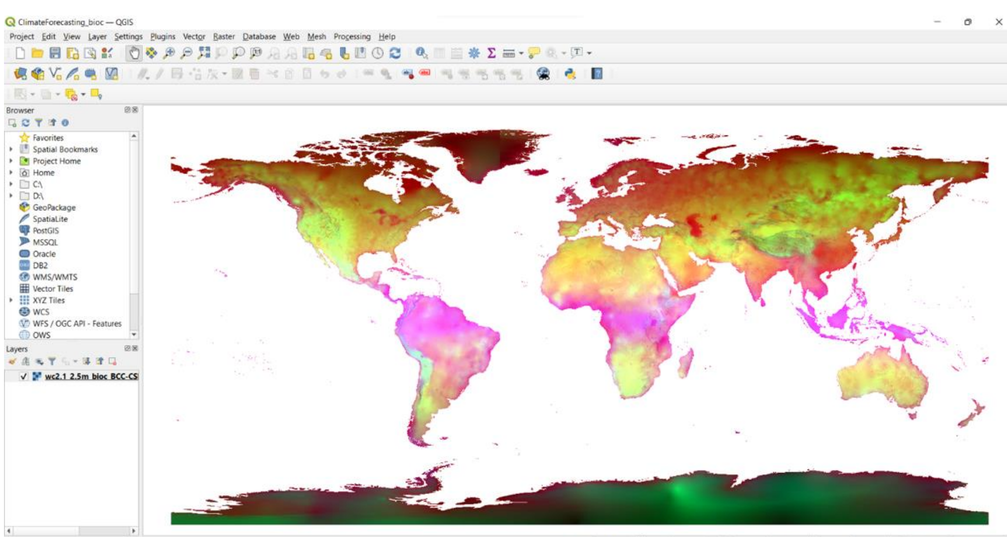
Topic-2





Topic-3

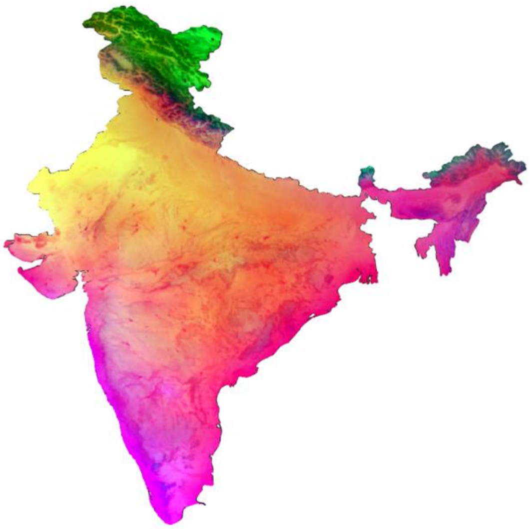
Topic-4



BIOC

Clipped To India

Climatic Variables



BIOC

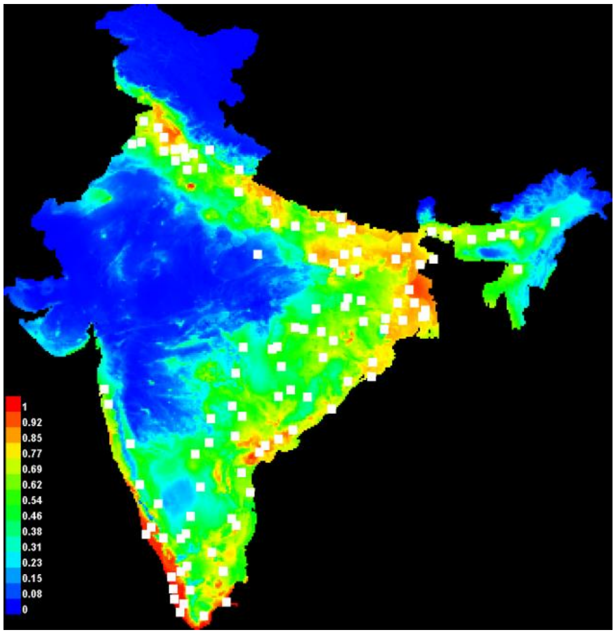
PREC

TMAX

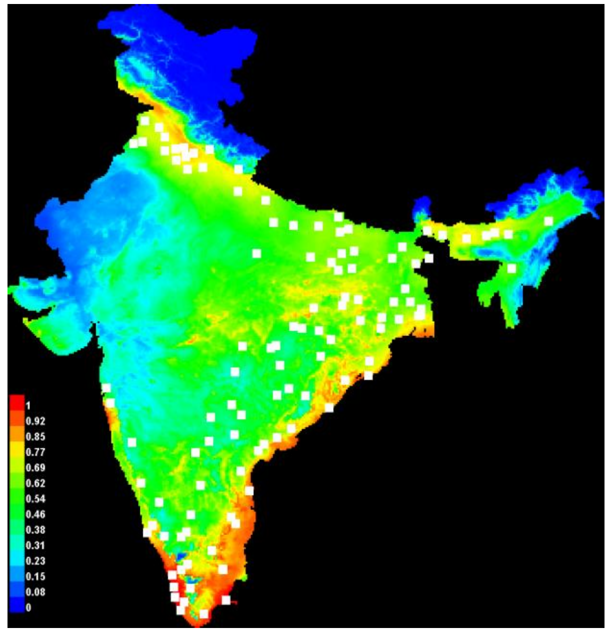
TMIN

Results @Phase-3 & Phase-4

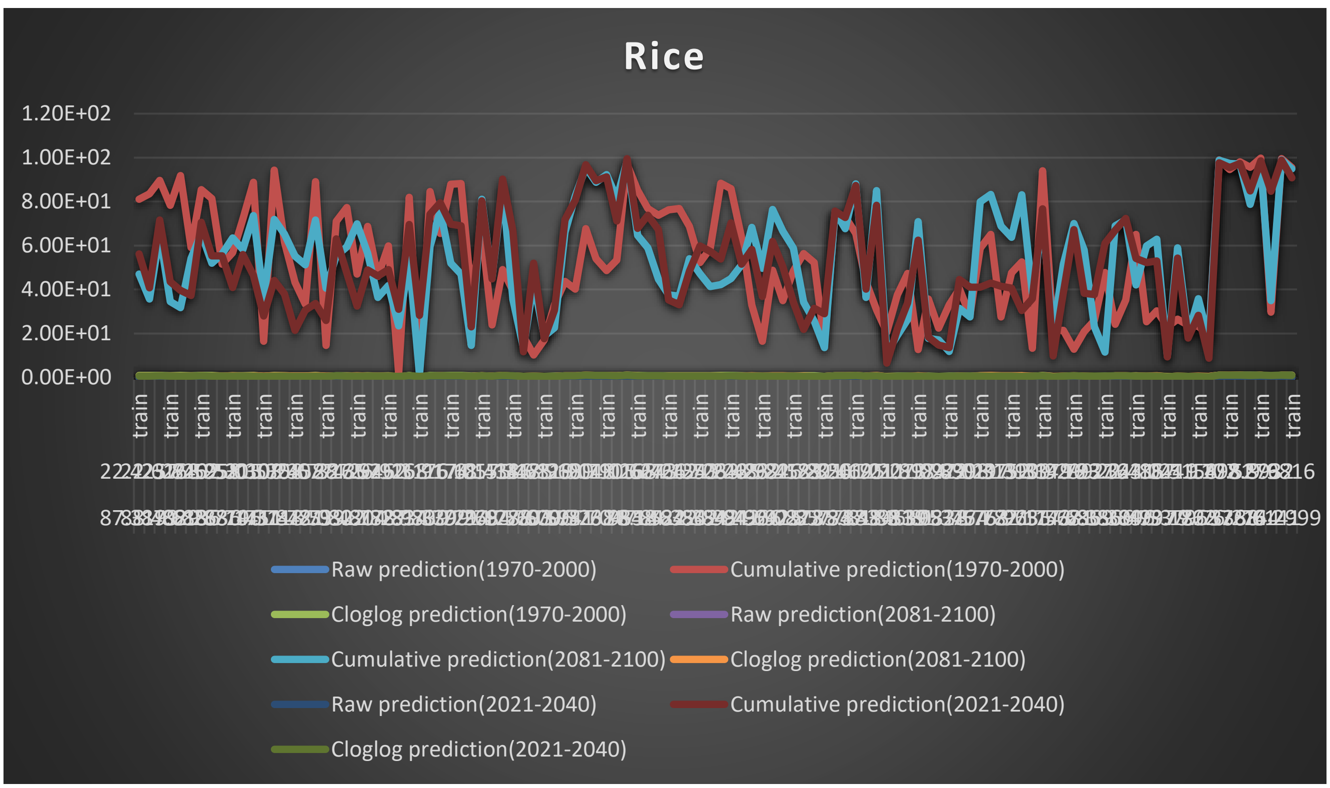
- Comparison of Rice Crop



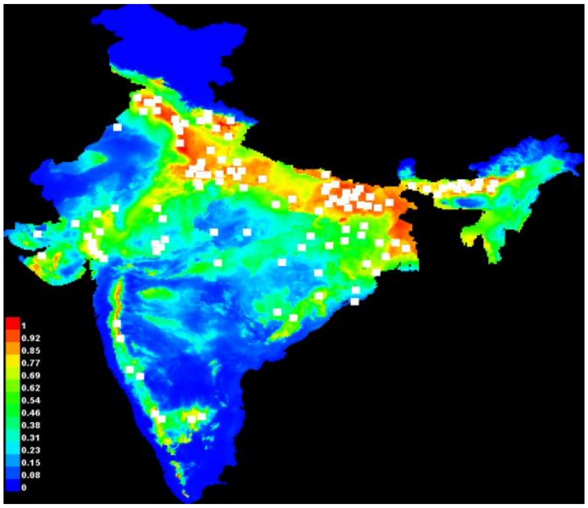
1970-2000



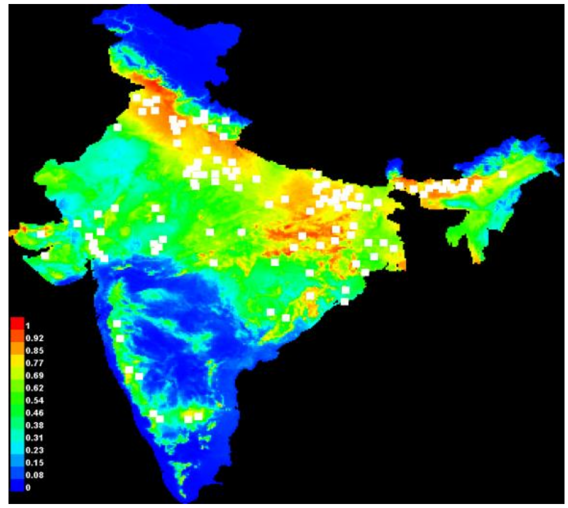
2081-2100



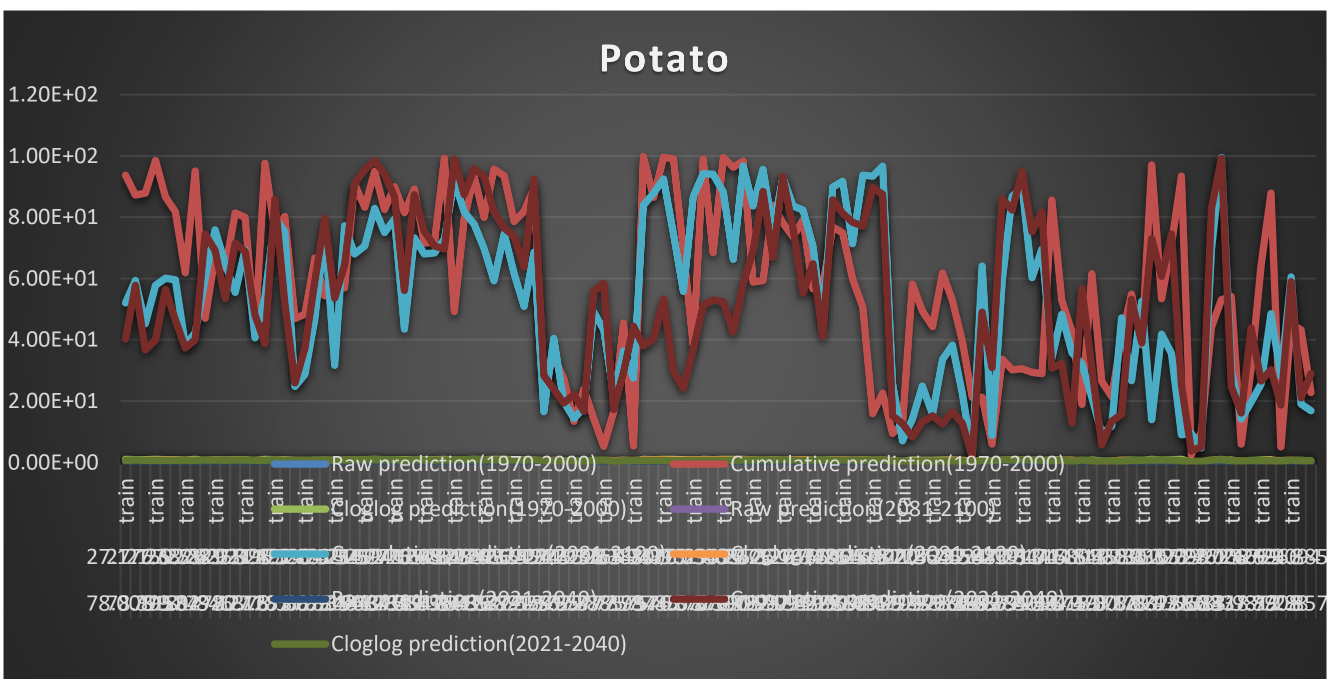
- Comparison of Potato Crop



1970-2000

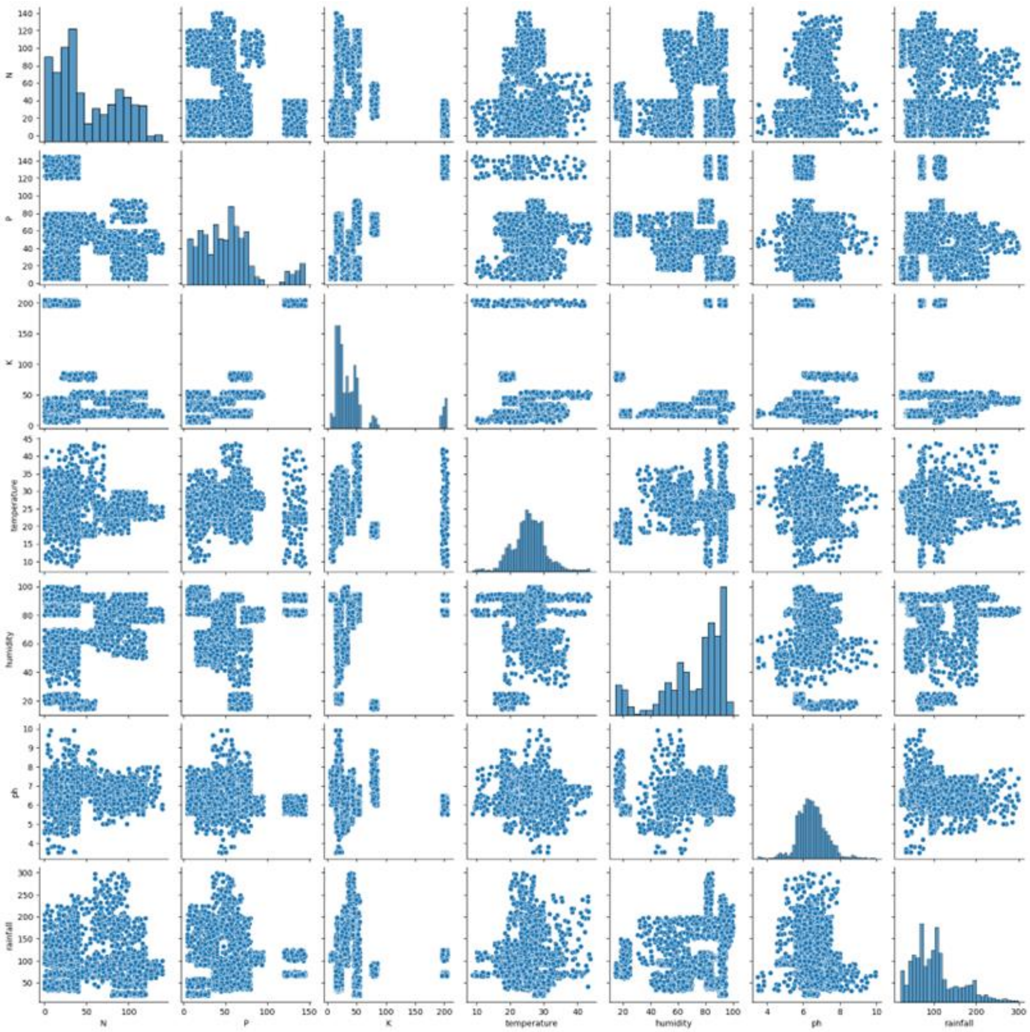


2081-2100



○ **Pair Plot of the data**

A pair plot allows us to see both distributions of single variables and relationships between two variables.



- **SVM**

SVM's Accuracy is: 0.10681818181818181

	precision	recall	f1-score	support
apple	1.00	0.23	0.38	13
banana	1.00	0.24	0.38	17
blackgram	1.00	0.19	0.32	16
chickpea	1.00	0.05	0.09	21
coconut	1.00	0.05	0.09	21
coffee	0.00	0.00	0.00	22
cotton	1.00	0.05	0.10	20
grapes	1.00	0.06	0.11	18
jute	1.00	0.07	0.13	28
kidneybeans	0.03	1.00	0.07	14
lentil	0.00	0.00	0.00	23
maize	0.00	0.00	0.00	21
mango	0.00	0.00	0.00	26
mothbeans	0.00	0.00	0.00	19
mungbean	1.00	0.12	0.22	24
muskmelon	1.00	0.30	0.47	23
orange	1.00	0.03	0.07	29
papaya	1.00	0.05	0.10	19
pigeonpeas	0.00	0.00	0.00	18
pomegranate	1.00	0.12	0.21	17
rice	0.50	0.06	0.11	16
watermelon	1.00	0.13	0.24	15
accuracy			0.11	440
macro avg	0.66	0.13	0.14	440
weighted avg	0.66	0.11	0.13	440

Accuracy: 10.6%

- Decision Tree

DecisionTrees's Accuracy is: 90.0

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	13
banana	1.00	1.00	1.00	17
blackgram	0.59	1.00	0.74	16
chickpea	1.00	1.00	1.00	21
coconut	0.91	1.00	0.95	21
coffee	1.00	1.00	1.00	22
cotton	1.00	1.00	1.00	20
grapes	1.00	1.00	1.00	18
jute	0.74	0.93	0.83	28
kidneybeans	0.00	0.00	0.00	14
lentil	0.68	1.00	0.81	23
maize	1.00	1.00	1.00	21
mango	1.00	1.00	1.00	26
mothbeans	0.00	0.00	0.00	19
mungbean	1.00	1.00	1.00	24
muskmelon	1.00	1.00	1.00	23
orange	1.00	1.00	1.00	29
papaya	1.00	0.84	0.91	19
pigeonpeas	0.62	1.00	0.77	18
pomegranate	1.00	1.00	1.00	17
rice	1.00	0.62	0.77	16
watermelon	1.00	1.00	1.00	15
accuracy			0.90	440
macro avg	0.84	0.88	0.85	440
weighted avg	0.86	0.90	0.87	440

Accuracy: 90%

- Logistic Regression

Logistic Regression's Accuracy is: 0.9522727272727273

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	13
banana	1.00	1.00	1.00	17
blackgram	0.86	0.75	0.80	16
chickpea	1.00	1.00	1.00	21
coconut	1.00	1.00	1.00	21
coffee	1.00	1.00	1.00	22
cotton	0.86	0.90	0.88	20
grapes	1.00	1.00	1.00	18
jute	0.84	0.93	0.88	28
kidneybeans	1.00	1.00	1.00	14
lentil	0.88	1.00	0.94	23
maize	0.90	0.86	0.88	21
mango	0.96	1.00	0.98	26
mothbeans	0.84	0.84	0.84	19
mungbean	1.00	0.96	0.98	24
muskmelon	1.00	1.00	1.00	23
orange	1.00	1.00	1.00	29
papaya	1.00	0.95	0.97	19
pigeonpeas	1.00	1.00	1.00	18
pomegranate	1.00	1.00	1.00	17
rice	0.85	0.69	0.76	16
watermelon	1.00	1.00	1.00	15
accuracy			0.95	440
macro avg	0.95	0.95	0.95	440
weighted avg	0.95	0.95	0.95	440

Accuracy: 95.2%

- XGBoost

XGBoost's Accuracy is: 0.9931818181818182

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	13
banana	1.00	1.00	1.00	17
blackgram	1.00	1.00	1.00	16
chickpea	1.00	1.00	1.00	21
coconut	1.00	1.00	1.00	21
coffee	0.96	1.00	0.98	22
cotton	1.00	1.00	1.00	20
grapes	1.00	1.00	1.00	18
jute	1.00	0.93	0.96	28
kidneybeans	1.00	1.00	1.00	14
lentil	0.96	1.00	0.98	23
maize	1.00	1.00	1.00	21
mango	1.00	1.00	1.00	26
mothbeans	1.00	0.95	0.97	19
mungbean	1.00	1.00	1.00	24
muskmelon	1.00	1.00	1.00	23
orange	1.00	1.00	1.00	29
papaya	1.00	1.00	1.00	19
pigeonpeas	1.00	1.00	1.00	18
pomegranate	1.00	1.00	1.00	17
rice	0.94	1.00	0.97	16
watermelon	1.00	1.00	1.00	15

Accuracy: 99.32%

Precision, Recall and f1 score:

Precision:

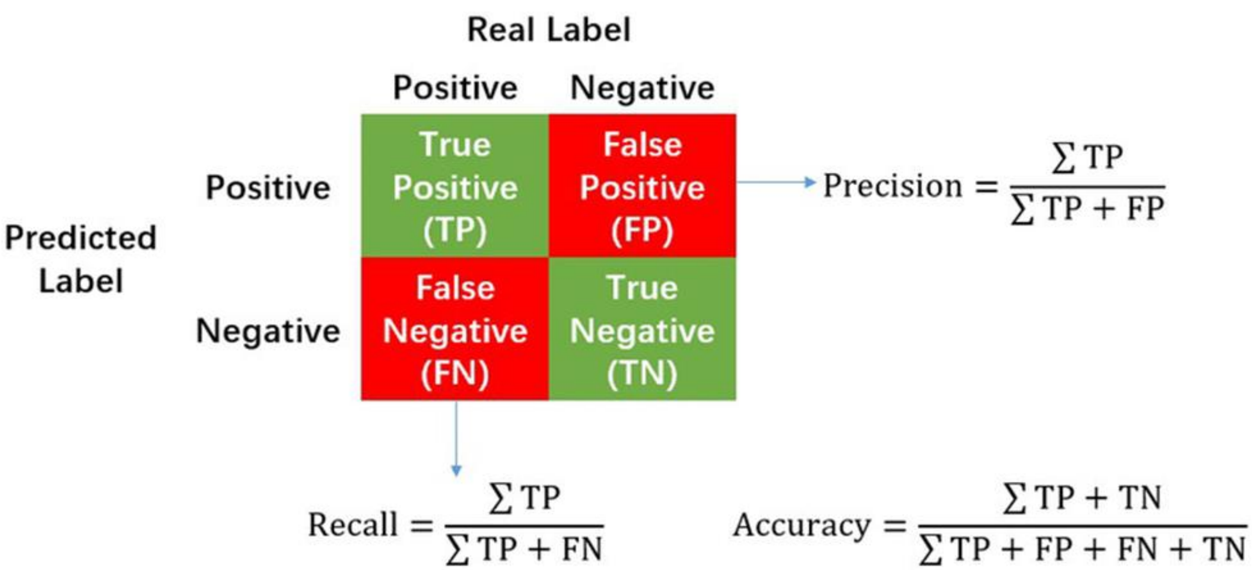
Precision tells us about the quality of true positive predictions. Precision is the number of correct positive predictions per total number of positive predictions.

Recall:

Recall is the ratio between the numbers of positive samples correctly classified as Positive to the total number of Positive samples.

F1 Score:

It is the harmonic mean of precision and recall. It takes both false positive and false negative into account. Therefore, it performs well on imbalanced dataset.



- **Crop Prediction:**

```
data = np.array([[90, 55, 44, 23, 82.3, 7.8, 263.9]])
prediction = RF.predict(data)
print(prediction)
```

- ['rice']

```
data = np.array([[55,66, 22,30.9, 68.79, 7.7, 66.91]])
prediction = RF.predict(data)
print(prediction)
```

- ['blackgram']

```
data = np.array([[24, 33, 35, 29.6, 54.3, 5.8, 100.9]])
prediction = RF.predict(data)
print(prediction)
```

- ['mango']

6. WORK DONE FOR FINAL EVALUATION

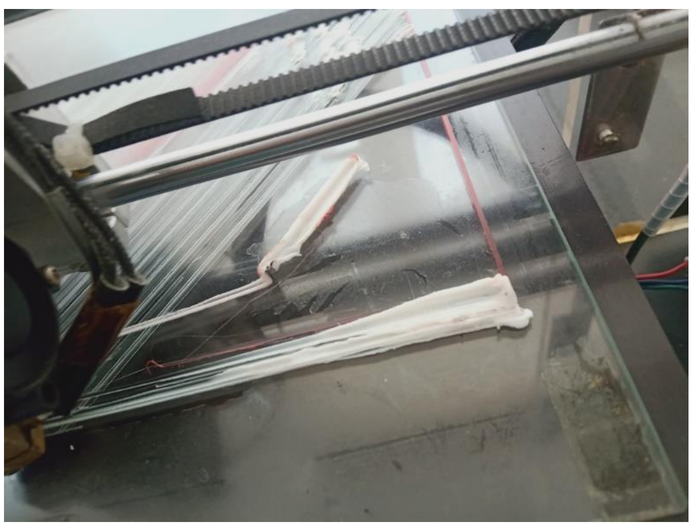
- Collected dataset which contain the information of different crops and their respective attributes information like amount of rainfall, degree of temperature, etc.
- Performed Data Cleaning to check whether any outliers present in the data or not.
- No outliers present in the data.
- Removing outliers is must because if outliers present then there is a chance of bias towards some sample cases.
- Trained data using supervised ML algorithms like SVM, Logistic Regression, Decision Tree, XGBoost.
- Compared training accuracy of different algorithm.
- Predicted the crop by giving some truly related values to the attributes, the one which has less error rate to predict was selected such that test accuracy will be great enough.

7.CONCLUSION & FUTURE WORK

- Southern states of India are highly suitable for cultivating rice crop even in future (2081-2100).
- Some of the middle part of India is also suitable for growing rice and getting some good amount of yield in future while at present those land is not suitable for rice cultivation.
- Even in future southern states are not suitable for potato crop cultivation.
- North-East states of India are highly suitable for cultivating potato.
- Temperature and humidity have significant positive correlation value, both factors affect the crop growth and yield of the crop.
- XGBoost algorithm predicts the crop with less error rate compared to logistic regression, decision tree, support vector machine algorithms.
- One of the future works is to build a fully functional web page which takes attribute values as input and sending a notification to farmer suggesting crop which gives best yield.
- Continuously monitoring the diseases in crop with the help of drones (image processing), helping the crop to grow healthily.

8. CHALLENGES FACED

- For designing encapsulation, needed to run multiple times to test the rigidity and regular standards of the prototype.
- Integrating the components on the PCB.
- Fixing the print bed to the equal levels uniformly over all four sides was afflicting and time consuming.
- Time Precision was not estimated correctly:
    - Expected time shown is 6.15 hrs but it took more than 12 hrs to print the complete box.
    - Need to observe the corrective Ness of the printing over complete time.



Uneven Printing

10. ACKNOWLEDGMENTS

We would like to thank our mentor, Dr. Priyanka Dwivedi for guiding us in this project and providing inputs and suggestions during evaluations and working. All team members have contributed equally, sharing the workload since the beginning.

11. REFERENCES

[1] Soni, Dipa & Makwana, Ashwin. (2017). A SURVEY ON MQTT: A PROTOCOL OF INTERNET OF THINGS(IOT). ResearchGate.

[2] Das Bhowmik, Rajarshi & Suchetana, Bihu & Lu, Mengqian. (2019). Shower effect of a rainfall onset on the heat accumulated during a preceding dry spell. Scientific Reports. 9. 10.1038/s41598-019-43437-7.

[3] Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology.

[4] Gogtay, Nithya J., and Urmila M. Thatte. "Principles of correlation analysis." Journal of the Association of Physicians of India 65.3 (2017): 78-81.

[5] Phillips, Steven J. "A brief tutorial on Maxent." AT&T Research 190.4 (2005): 231-259.

[6] Thingspeak as MQTT Broker - link

[7] WorldClim_Future_Climate_Data link

[8] https://statisticsbyjim.com/basics/five-number-summary/link