TECHNICAL ANSWERS FOR REAL WORLD PROBLEMS (ECE1901)

FALL SEM – 2023-2024

PROJECT REPORT
ON

"MONITORING AND CONTROLLING OF VARIOUS PARAMETERS IN SMART GREENHOUSE SYSTEM & YIELD PREDICTION USING ML"

UNDER THE GUIDANCE OF
DR. RAJESH KUMAR V
ASSISTANT PROFESSOR SR.
DEPARTMENT OF COMMUNICATION ENGINEERING
SENSE, VIT, VELLORE – 632014

BY

GAURAV BANSAL	20BEC0147
DARSH GUPTA	20BEC0563
ANIRUDH KOGANTI	20BEC0152
SIDDHARTH BHATTACHARYA	20BEC0785
AKASH IYER	20BEC0046
ADITYA SINGH	20BEC0206
ANIMESH AGARWAL	20BEC0309

SLOT: TD1

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Abstract

With the increasing demand for sustainable agriculture practices and the need for efficient resource utilization, smart greenhouse systems have emerged as a promising solution. This project focuses on the design, implementation, and optimization of a Smart Greenhouse System that integrates advanced monitoring and control mechanisms to create an environment conducive to optimal plant growth. Additionally, the project explores the application of Machine Learning (ML) techniques for yield prediction, enhancing the precision and efficiency of agricultural production.

The Smart Greenhouse System is equipped with a network of sensors to monitor and collect real-time data on various environmental parameters such as temperature, humidity, soil moisture, light intensity, and CO2 levels. The collected data is processed and analysed to ensure the creation of an ideal and controlled environment for plant growth. Automated control mechanisms, driven by this data, are employed to regulate factors like irrigation, ventilation, and shading, optimizing resource usage and minimizing waste.

The ML component of the project focuses on predicting crop yield based on historical and real-time data. Various ML algorithms are explored and implemented to analyse the correlation between environmental parameters and crop productivity. By leveraging these models, the system aims to provide accurate predictions of crop yield, enabling farmers to make informed decisions regarding crop management, resource allocation, and harvest planning.

The integration of monitoring, control, and ML-based yield prediction not only enhances the efficiency of the greenhouse system but also contributes to sustainable agriculture by minimizing resource usage and maximizing yield. The project aims to demonstrate the feasibility and effectiveness of combining smart technologies and machine learning in agricultural practices, paving the way for future innovations in precision farming and sustainable food production.

Test outcomes showed that the dependability of the framework in spreading data straightforwardly to the agriculturists could be picked up astoundingly in different conditions:

- We propose the use of microcontrollers and sensors to minimize this tedious task of monitoring and maintaining condition inside the greenhouse.
- We will first use sensors for measuring the condition of the greenhouse.
- The collected data will be given to the microcontrollers which will decide whether to take action to improve the condition.
- We would also save the data and actions taken by the system in cloud for the user to view later.
- We would also like to implement a phone app which will inform the user in case of any changes made in the green house.

Introduction:

Agriculture, being the backbone of human sustenance, is continually evolving to meet the demands of a growing global population. In this context, precision agriculture has emerged as a key strategy to enhance productivity while minimizing resource utilization. The advent of smart technologies, particularly in the form of Smart Greenhouse Systems, has opened new avenues for sustainable and efficient farming practices. This project delves into the development of a Smart Greenhouse System, focusing on the meticulous monitoring and control of various environmental parameters, coupled with the integration of Machine Learning (ML) for precise yield prediction.

Traditional agricultural practices are increasingly being complemented by technological advancements to address challenges such as climate change, resource scarcity, and the need for optimized crop management. Smart Greenhouse Systems represent a significant leap in this direction, providing a controlled environment that fosters ideal conditions for plant growth. The incorporation of sensors for monitoring crucial parameters, including temperature, humidity, soil moisture, light intensity, and CO2 levels, forms the foundation of our project. Real-time data acquisition from these sensors facilitates a nuanced understanding of the greenhouse environment.

Moreover, the project extends beyond mere monitoring, incorporating automated control mechanisms. By leveraging the collected data, the system dynamically regulates environmental factors such as irrigation, ventilation, and shading. This not only ensures the creation of an optimized growth environment but also minimizes resource wastage, aligning with the principles of sustainable agriculture.

In addition to monitoring and control, the project explores the integration of Machine Learning techniques for yield prediction. By analyzing historical and real-time data, ML models are developed to discern patterns and correlations between environmental parameters and crop productivity. This predictive capability empowers farmers with insights into potential yields, enabling them to make informed decisions regarding crop management strategies, resource allocation, and harvest planning.

The synergy of smart technologies and machine learning in agriculture holds immense promise for revolutionizing conventional farming practices. The proposed Smart Greenhouse System, with its comprehensive approach to monitoring, control, and ML-driven yield prediction, aims to contribute to the advancement of precision agriculture. Through this project, we seek to demonstrate the feasibility and efficacy of these integrated technologies in fostering sustainable, resource-efficient, and high-yield agricultural practices.

Literature Survey:

Sr. No.	Title	Journal Title	Publisher	Year	Research Gap
1	Secured IoT Based Smart Greenhouse System with Image Inspection	Advanced Computing and Communication System	IEEE	2020	The research gap in this project lies in the lack of consideration for potential inaccuracies in image processing caused by factors like leaf movement and varying camera positions within the automated greenhouse system. Addressing these challenges to ensure the reliability of image processing is an area requiring further research.
2	Smart Greenhouse Monitoring System Using Internet of Things and Artificial Intelligence	Wireless Personal Communications	Springer	2022	The project's research gap is the necessity for a thorough evaluation of the system's real-world feasibility and adaptability. While the project underscores the advantages of an automated smart greenhouse incorporating ANFIS and IoT technology, it lacks concrete data on the system's performance in a large-scale agricultural context. Subsequent research should prioritize assessing the system's effectiveness, economic viability, and security in a practical environment to ascertain its potential for widespread adoption in agriculture.
3	A CNN-RNN Framework for Crop Yield Prediction	Computational Genomics	Frontiers	2019	The research gap in this paper is the need for further exploration of the CNN-RNN model's applicability beyond corn and soybean crops in the United States. While the study demonstrates the model's effectiveness for yield prediction in a specific region and crop types, there is a gap in understanding how well this approach generalizes to other crops and diverse geographic regions. Future research should investigate the model's performance with various crop types and across different global agricultural settings.

4	Crop Yield Maximization Using an IoT-Based Smart Decision	AI-Driven Intelligent Sensor Networks	Hindawi	2022	The research gap in this study is the need to expand the applicability of the Smart Crop Selection (SCS) model beyond its specific region (Bahawalpur, Pakistan) and limited crop variety. Further research should aim to test the SCS model in diverse global agricultural contexts, consider more crop types, address data quality issues, improve result interpretability, and adapt to changing climate conditions. This would enhance the model's usefulness and reliability for farmers worldwide.
5	IoT Based Automated Greenhouse Monitoring System	Intelligent Computing and Control Systems	IEEE	2018	The research gap in this project is the limited integration of real-time data analytics and predictive capabilities. While the project addresses the challenges of manual greenhouse monitoring through IoT technology, it primarily focuses on real-time control and data display. A significant research gap is the lack of advanced data analytics and predictive modelling that could provide insights into future conditions and optimize greenhouse operations for maximum plant growth and yield. Future research should explore the potential for incorporating machine learning or other predictive techniques to enhance greenhouse management and further improve agricultural productivity.
6	IoT based smart greenhouse	Humanitarian Technology Conference	IEEE	2016	The research gap in this work is the need for a comprehensive evaluation of the scalability, integration, and sustainability of modern agricultural technologies. While the project introduces a smart greenhouse model with automated farming tasks, several gaps include the lack of focus on holistic system integration, advanced data analytics, sustainability assessment, scalability to different farm sizes, and user-friendly accessibility. Future research should aim to address these gaps to better understand and optimize the impact of modern technology in agriculture for improved productivity and resource efficiency.

7	IoT Based Smart Greenhouse Framework and Control Strategies for Sustainable Agriculture	-	IEEE Access	2022	The study's research gaps are centred on addressing challenges and security concerns related to IoT-based smart greenhouse farming. While the paper explores IoT technology in agriculture, it doesn't provide specific solutions for overcoming these challenges and ensuring greenhouse system security. Future research should concentrate on devising effective strategies to tackle these issues, unlocking the full potential of IoT in smart farming.
8	Internet of Things Based Smart Greenhouse: Remote Monitoring and Automatic Control	DEStech Transactions on Environment Energy and Earth Science	ResearchGate	2019	This paper's research gap lies in the requirement for advanced data analytics and predictive features in the proposed automatic greenhouse control system. While the system effectively manages climate conditions and stores data, it doesn't offer future-focused optimization and insights. Subsequent research should delve into machine learning to improve the system's capabilities and aid farmers in making informed decisions.
9	Greenhouse monitoring and control system based on wireless Sensor Network	Computing, Control, Networking, Electronics and Embedded Systems Engineering	IEEE	2015	The papers discuss the use of wireless sensor networks (WSN) in a greenhouse environment to monitor parameters like temperature, humidity, moisture, light, and CO_2. However, they do not address the issue of crop yield prediction or the integration of machine learning for optimizing agricultural practices. The project, on the other hand, focuses on crop yield prediction using machine learning, which is a significant research gap not covered in those papers.
10	Green house by using IOT and cloud computing	Recent Trends in Electronics, Information & Communication Technology	IEEE	2016	The paper discusses the various techniques available for precision agriculture to monitor and control the environment for the growth of crops. They highlight the challenges related to the unequal distribution of rainwater and the need for irrigation methods suitable for different weather conditions, soil types, and crop varieties. Greenhouses are proposed as a solution to these issues. However, these abstracts do not address the specific topic

					of crop yield prediction or the application of machine learning for optimizing agricultural practices. The project, on the other hand, focuses on developing accurate machine-learning models for crop yield prediction in smart greenhouses, which represents a research gap not covered in these abstracts.
11	IoT-based Smart Greenhouse with Disease Prediction using Deep Learning	International Journal of Advanced Computer Science and Applications, (IJACSA)	-	2021	The paper discusses the impact of rapid industrialization and urbanization on agricultural land and productivity worldwide. They emphasize the increasing demand for chemical-free organic vegetables and the growing popularity of greenhouses, particularly in extreme weather countries, to address these challenges. These abstracts introduce IoT-based Smart Greenhouse systems with features like monitoring, alerting, cloud storage, automation, and disease prediction. While they address greenhouse management and disease identification, they do not explicitly mention the aspect of crop yield prediction, which is a research gap covered in their project. Their project focuses on leveraging machine learning for accurate crop yield prediction in smart greenhouses, a topic not explicitly mentioned in these abstracts.
12	Smart Agriculture: IoT-based Greenhouse Monitoring System	Computers, Communications & Control	ResearchGate	2022	The paper discusses the concept of "smart agriculture," emphasizing its reliance on advanced technologies such as big data, cloud computing, and the Internet of Things (IoT) to enhance agricultural management. They mention the use of software and sensors for precision agriculture but do not delve into the specific aspect of crop yield prediction or the application of machine learning for optimizing agricultural practices. The projects, on the other hand, focus on developing accurate and scalable machine learning models for crop yield prediction, addressing research gaps not covered in this paper.

13	A Novel Approach for Monitoring of Smart Greenhouse and Flowerpot Parameters and Detection of Plant Growth with Sensors	-	MDPI	2022	The paper emphasizes the importance of utilizing smart technologies to address agricultural challenges globally and locally, particularly in the context of smart greenhouses. They describe the construction and programming of smart greenhouse prototypes to create optimal soil and climatic conditions for growing agricultural products. While they discuss the positive impact of the smart greenhouses on plant development and water savings, they do not explicitly mention crop yield prediction or the application of machine learning for optimizing agricultural practices. Their projects, in contrast, focus on developing machine learning models for crop yield prediction in smart greenhouses, addressing research gaps not covered in these abstracts.
14	Internet of Things Approaches for Monitoring and Control of Smart Greenhouses in Industry	-	MDPI	2022	The authors discuss the application of IoT in smart greenhouses for agriculture, emphasizing the importance of technology in addressing climate change and resource shortages. They focus on real-time monitoring and data collection and processing to efficiently control indoor parameters in greenhouses. While they provide insights into the benefits and opportunities of IoT in agriculture, their abstract does not specifically address the aspect of crop yield prediction or the application of machine learning for optimizing agricultural practices. Their projects, on the other hand, aim to develop machine learning models for accurate crop yield prediction in smart greenhouses, filling research gaps not covered in these abstracts.

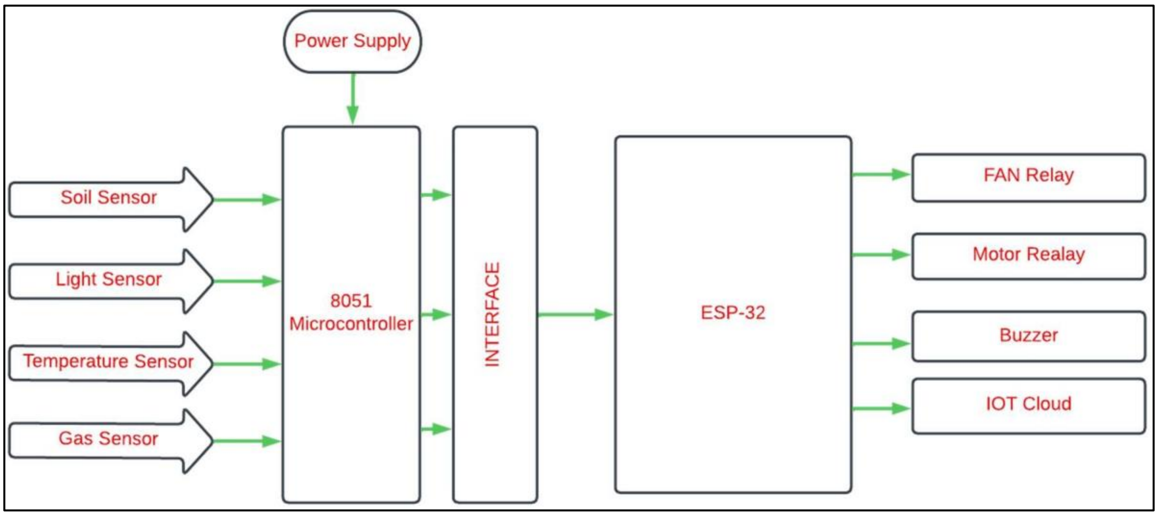
15	An Intelligent Monitoring and Controlling of Greenhouse: Deployment of Wireless Sensor Networks and Internet-of-Things	-	Europe PMC	2018	The authors highlight the limitations of traditional farming methods, which lack precision in providing the right number of fertilizers and setting specific climate parameters. They propose an architecture involving embedded systems and wireless sensor networks to monitor various climatic parameters in a greenhouse, including humidity, temperature, carbon dioxide levels, acidity, soil moisture, and light intensity. While they focus on monitoring and controlling these parameters, their abstract does not explicitly address the aspect of crop yield prediction or the application of machine learning for optimizing agricultural practices. Their projects, however, aim to develop machine learning models for accurate crop yield prediction in smart greenhouses, addressing research gaps not covered in these abstracts.

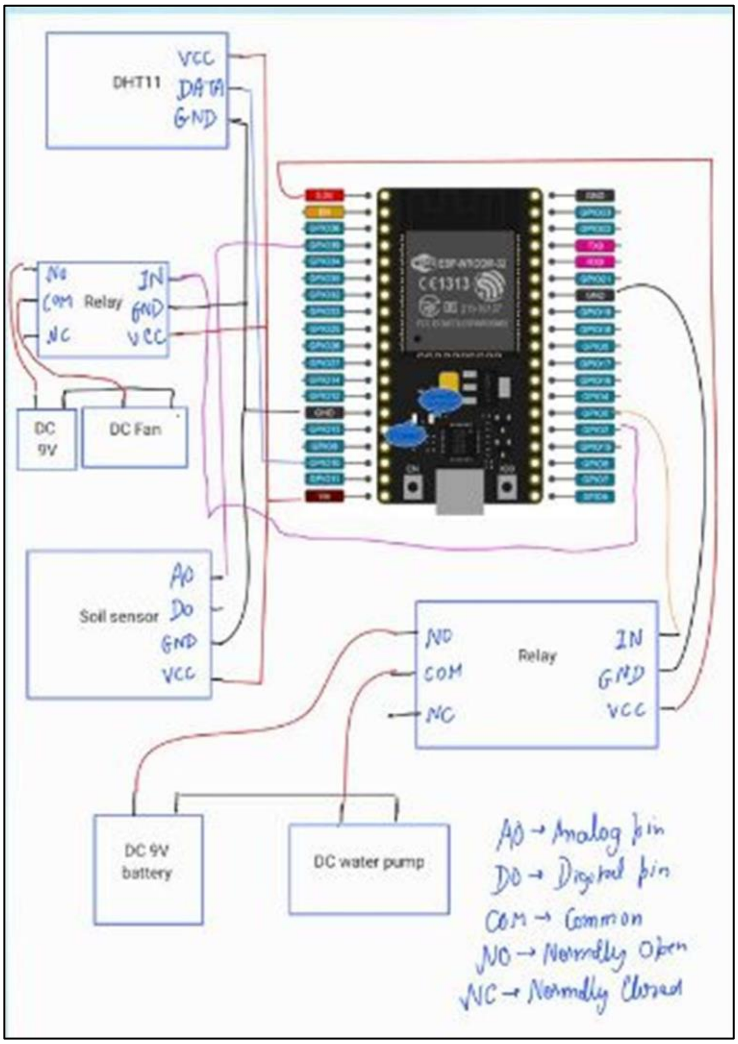
Working Principle

The system operates on a sophisticated integration of sensor technology, control mechanisms, and machine learning algorithms. The working principle of the project can be broken down into several key components:

1. **Sensor Network:**
    - The project employs a network of sensors strategically placed throughout the greenhouse to monitor various environmental parameters critical for plant growth. These parameters include temperature, humidity, soil moisture, light intensity.
2. **Data Acquisition:**
    - The sensors continuously collect real-time data on the monitored parameters. This data is then fed into the central control system, creating a comprehensive dataset that reflects the current environmental conditions within the greenhouse.
3. **Control Mechanisms:**
    - An automated control system interprets the data received from the sensors to make informed decisions about regulating the greenhouse environment. Control mechanisms are implemented to adjust factors such as irrigation, ventilation, and shading. For example, if the temperature exceeds a certain threshold, the system might trigger ventilation or cooling mechanisms to maintain an optimal temperature.
4. **Smart Greenhouse Environment:**
    - The combination of real-time monitoring and automated control mechanisms creates an environment within the greenhouse that is finely tuned to the specific needs of the plants. This ensures that plants receive optimal conditions for growth, leading to increased efficiency in resource utilization and minimizing waste.
5. **Machine Learning for Yield Prediction:**
    - Concurrently, historical and real-time data collected by the sensors are used to train machine learning models. These models are designed to identify patterns and correlations between environmental parameters and crop yield. Various machine learning algorithms, such as regression or neural networks, are employed to create predictive models.
6. **Yield Prediction and Decision Support:**
    - The trained machine learning models predict crop yield based on the current and historical environmental data. This predictive capability provides farmers with valuable insights into expected crop performance. Farmers can use this information for decision-making related to crop management practices, resource allocation, and harvest planning.
7. **Continuous Optimization:**
    - The system operates in a feedback loop, continuously monitoring, analyzing, and adjusting parameters based on the evolving conditions. This iterative process ensures that the greenhouse environment is consistently optimized for the best possible crop growth and yield.

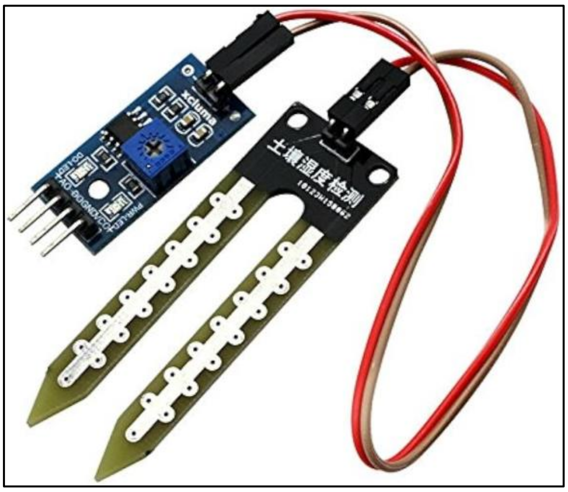
Block Diagram



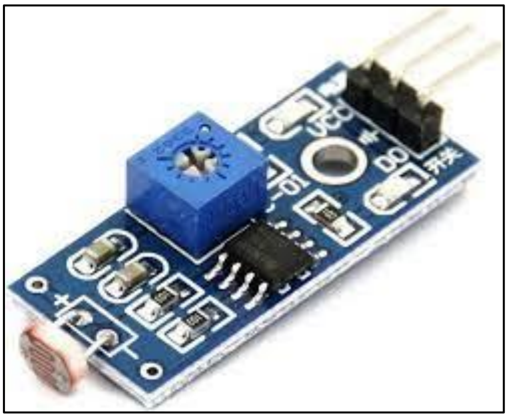


Explanation of Parts:

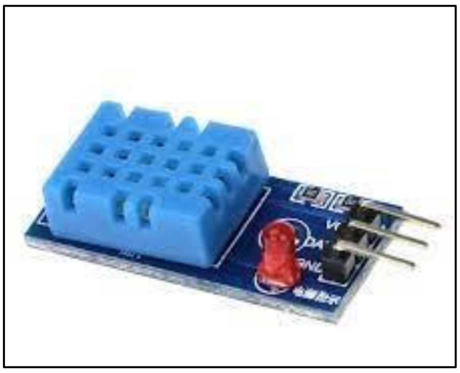
1. **Soil Sensor:** A soil sensor is a vital agricultural tool designed to measure and monitor key properties of soil. These sensors assess parameters such as moisture content, temperature, electrical conductivity, and pH levels, offering crucial insights into soil health. Installed at various depths within the soil, these sensors provide real-time data, enabling farmers to make informed decisions about irrigation, fertilization, and overall crop management. Advanced soil sensors often feature wireless data transmission, facilitating remote monitoring and analysis. By promoting precision agriculture, soil sensors contribute to resource-efficient farming, enhanced crop yields, and sustainable soil management practices in a rapidly evolving agricultural landscape.



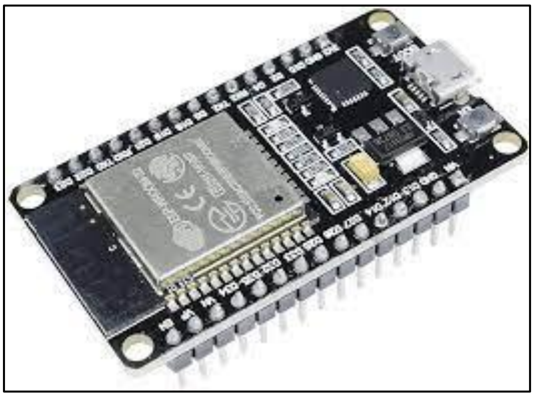
2. **Light Sensor:** A light sensor, also known as a photodetector, is a compact electronic device designed to measure the intensity of light in its environment. This sensor utilizes photodiodes or phototransistors to convert light energy into an electrical signal, allowing for precise detection of light levels. Commonly employed in various applications, including ambient light sensing in electronic devices, outdoor lighting control systems, and industrial automation, light sensors play a crucial role in optimizing energy efficiency. By providing real-time data on light conditions, these sensors enable automatic adjustments in lighting systems, contributing to energy conservation and creating environments that balance illumination requirements with sustainability considerations.



3. **DHT (Humidity and Temperature) Sensor:** A DHT sensor is a compact and versatile device commonly used in electronics for measuring temperature and humidity levels in the surrounding environment. Named after its two key parameters, the DHT sensor combines a humidity sensor and a thermistor. Employing a digital signal output, it provides accurate and reliable data on temperature and humidity, crucial for various applications such as climate control systems, weather stations, and IoT devices. DHT sensors are known for their ease of use, cost-effectiveness, and compatibility with microcontrollers, making them popular choices for projects requiring environmental monitoring. Their ability to deliver precise data contributes to improved decision-making in areas ranging from agriculture to home automation.



4. **ESP32:** The ESP32 is a powerful and widely used microcontroller and system-on-a-chip (SoC) that belongs to the ESP family developed by Espressif Systems. It integrates a dual-core processor, Wi-Fi and Bluetooth connectivity, numerous GPIO pins, and various peripherals, making it a versatile choice for IoT (Internet of Things) applications and embedded systems. The ESP32's capabilities, low power consumption, and ease of programming make it a popular platform for developing smart devices, home automation systems, and connected projects. Its compatibility with the Arduino IDE and support for multiple programming languages contribute to its widespread adoption in the maker and development communities.



5. **Relay Motor:** A relay motor control system is a fundamental component in industrial and automation settings, facilitating the efficient operation of electric motors. Relays serve as electromagnetic switches, controlling the electrical current that activates the motor. When a low-power signal triggers the relay, it closes or opens the higher-power circuit connected to the motor, allowing for precise control of motor functions. This setup is commonly employed in various applications, including manufacturing machinery, robotics, and home automation. The relay motor system enhances safety, provides remote control capabilities, and contributes to the overall efficiency of electromechanical systems by enabling the seamless integration of motors into diverse automated processes.

6. **DC Pump:** A DC water pump, harnessing the power of direct current, is a compact and efficient device designed for pumping water in diverse applications. These pumps utilize the simplicity and controllability of DC power, making them suitable for tasks such as water circulation in solar water heaters, garden irrigation, and recreational vehicles. With varying flow rates and pressure capabilities, DC water pumps offer flexibility in addressing specific water pumping needs. Their compact size, energy efficiency, and ease of integration with renewable energy sources make them particularly advantageous in off-grid or remote locations. DC water pumps play a crucial role in providing reliable and sustainable water management solutions.

ML Implementation:

1. Random Forest Implementation: Random Forest is a powerful machine learning algorithm known for its versatility and accuracy. Falling under the ensemble learning category, it constructs multiple decision trees during training and outputs the mode or mean prediction of the individual trees for classification tasks or regression, respectively. Each tree is trained on a random subset of the training data and features, introducing diversity and reducing overfitting. Random Forest excels in handling large datasets, high-dimensional features, and is robust against outliers. Its widespread application spans from finance to healthcare, owing to its capability to provide reliable predictions and insights across diverse domains.

2. Support Vector Machines (SVM): Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification and regression tasks. It works by finding a hyperplane in a high-dimensional space that best separates data points into distinct classes. SVM is particularly effective in scenarios with complex decision boundaries, leveraging support vectors to define the optimal separation. It excels in both linear and non-linear classification tasks, using kernel functions to transform data into higher-dimensional spaces. SVM is widely employed in image classification, text categorization, and bioinformatics due to its ability to handle diverse and intricate datasets. Its robust performance and versatility make SVM a valuable tool in various domains.

3. Linear Regression: Linear Regression is a fundamental statistical and machine learning technique used for predicting the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables and calculates the best-fit line through the data using the method of least squares. The goal is to minimize the difference between the predicted and actual values. Widely applied in fields such as economics, finance, and science, linear regression serves as a foundational tool for forecasting and trend analysis. Its simplicity, interpretability, and efficiency make it a go-to method for modelling and understanding linear relationships in diverse datasets.

5. LightGBM: LightGBM, or Light Gradient Boosting Machine, is a high-performance and efficient gradient boosting framework for machine learning tasks. Developed by Microsoft, it is designed to handle large datasets and offers advantages in terms of speed and memory usage. LightGBM employs a tree-based learning approach, using gradient-based optimization methods to construct decision trees sequentially. Noteworthy for its ability to handle categorical features and distributed training, LightGBM is commonly applied in tasks such as classification, regression, and ranking. Its versatility and scalability have made it a popular choice in various domains, contributing to its widespread adoption in both research and industry.

IMPLEMENTATION:

- We want to predict the yield percentage of the crops before we grow them according to the environmental parameters.
- This would allow us to choose the best parameters to grow the crops and also give us an idea of the yield amount that we would be getting during the time of harvesting.
- This would not only increase the quality of the crops that are grown but also give the farmers an approximate of the revenue that can be generated by the crops.
- We have also implemented an algorithm that will tell us which is the best crop that should be grown in a particular area based on the average environmental parameters present there.
- Our hardware can currently monitor 3 parameters. We can easily increase them to collect more parameters that are required in the machine learning algorithm since we have already developed the prototype. In simple terms, we can take advantage of reusability and expand for prediction analysis using any ML algorithm, we first need to define independent and dependent variables. In the dataset that we are using we have first calculated yield as PRODUCTION/AREA.
- This yield is made the dependent variable while all the other parameters will be made independent variables.
- Yield would depend on the defined independent variables.
- We have then used Data preprocessing methods to remove null values and outliers from the dataset. We have also created dummy variables to handle the columns that are given string values.
- We are then splitting the dataset for training and testing the ML algorithms.
- We have trained different ML algorithms using Training dataset. Using this model and the Testing dataset we then predict the yield values.
- The accuracy of the model is calculated using these predicted values and the Testing yield values.

Comparison:

- To determine the accuracy values for the given model in yield prediction, we should typically follow these steps:

- Data Preparation: Ensure that the dataset is properly prepared, and features are appropriately selected and preprocessed.

- Model Selection: Choose the appropriate machine learning models for your problem. SVM, XGBoost, and Naive Bayes are all valid choices, but their suitability depends on the characteristics of our data.

- Training and Validation: Split your dataset into training and validation sets. Train each model on the training set and evaluate its performance on the validation set.

- Hyperparameter Tuning: Fine-tune the hyperparameters of each model to optimize their performance. This process may involve using techniques like grid search or randomized search.

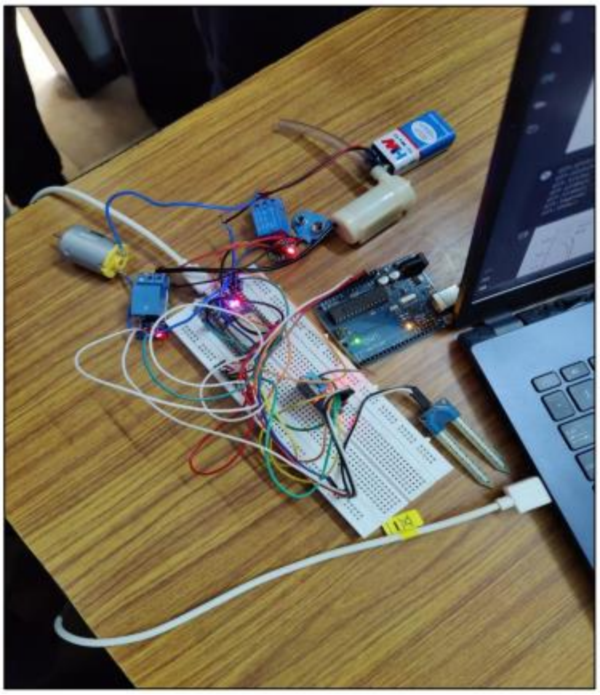
- Evaluation: Evaluate the models using appropriate evaluation metrics. This could include accuracy, precision, recall, F1 score, or other relevant metrics.

- Comparison: Compare the performance of SVM, XGBoost, and Naive Bayes to determine which model performs best for your specific problem.

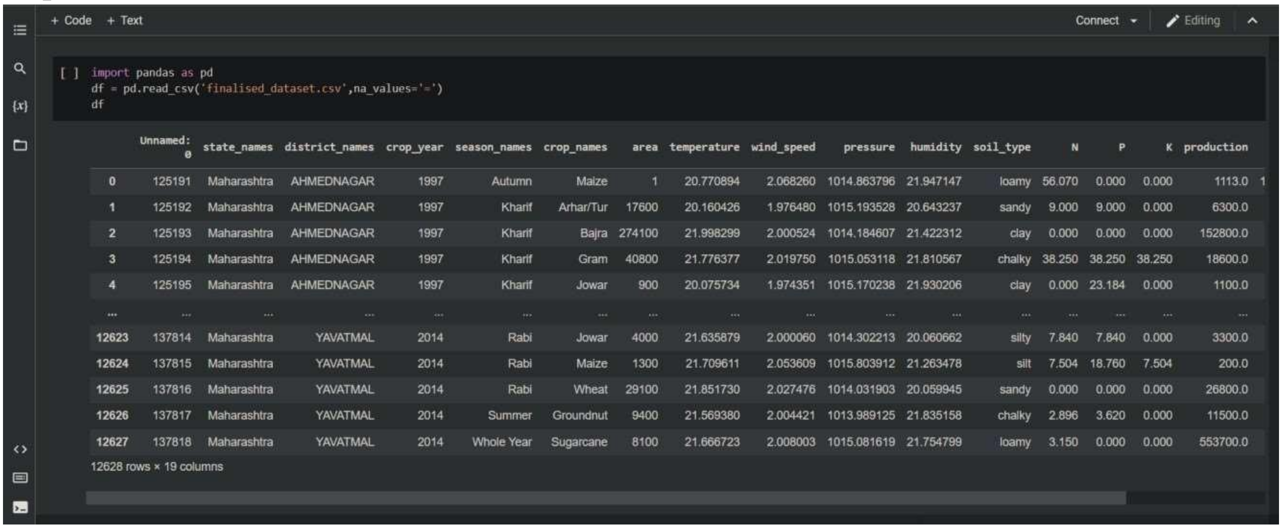
Based on the accuracy metrics that are extrapolated from machine learning models like SVM, XGB and Naïve Bayes being deployed, we can compare and conclude that SVM has higher accuracy (around 95%) followed by XGB (around 90%) and lastly Naive bayes (around 80%).

Results and Discussion:

Hardware



Import our data from csv files

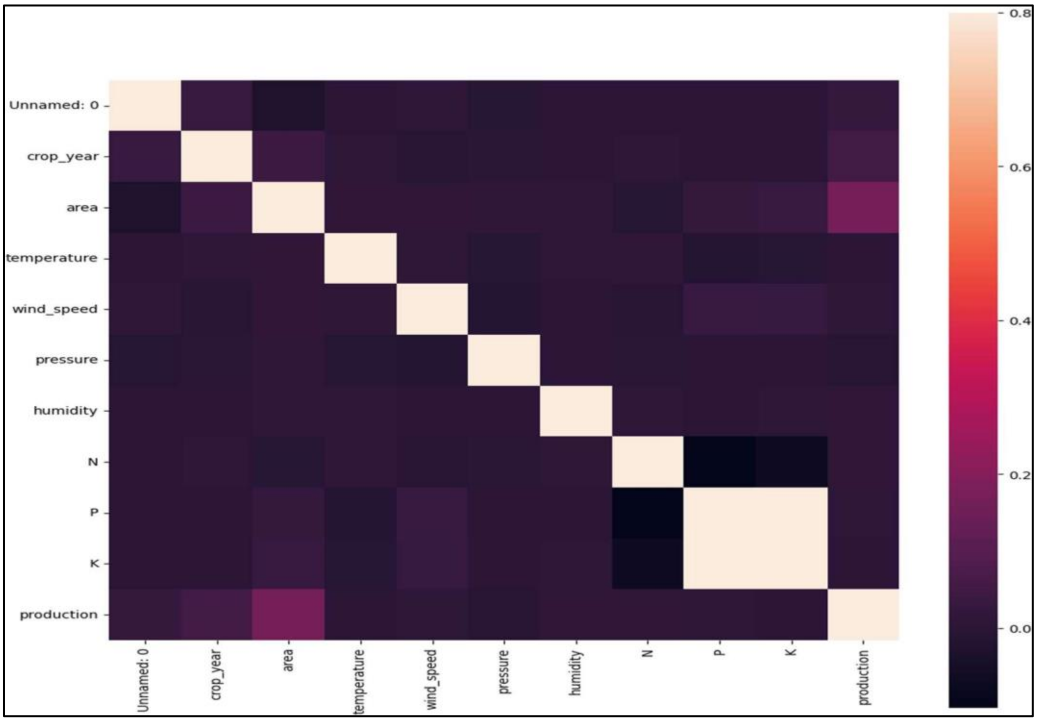


```
import pandas as pd
df = pd.read_csv('finalised_dataset.csv',na_values='=')
df
```

	Unnamed: 0	state_names	district_names	crop_year	season_names	crop_names	area	temperature	wind_speed	pressure	humidity	soil_type	N	P	K	production
0	125191	Maharashtra	AHMEDNAGAR	1997	Autumn	Maize	1	20.770894	2.068260	1014.863796	21.947147	loamy	56.070	0.000	0.000	1113.0
1	125192	Maharashtra	AHMEDNAGAR	1997	Kharif	Arhar/Tur	17600	20.160426	1.976480	1015.193528	20.643237	sandy	9.000	9.000	0.000	6300.0
2	125193	Maharashtra	AHMEDNAGAR	1997	Kharif	Bajra	274100	21.998299	2.000524	1014.184607	21.422312	clay	0.000	0.000	0.000	152800.0
3	125194	Maharashtra	AHMEDNAGAR	1997	Kharif	Gram	40800	21.776377	2.019750	1015.053118	21.810567	chalky	38.250	38.250	38.250	18600.0
4	125195	Maharashtra	AHMEDNAGAR	1997	Kharif	Jowar	900	20.075734	1.974351	1015.170238	21.930206	clay	0.000	23.184	0.000	1100.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
12623	137814	Maharashtra	YAVATMAL	2014	Rabi	Jowar	4000	21.635879	2.000060	1014.302213	20.060662	silty	7.840	7.840	0.000	3300.0
12624	137815	Maharashtra	YAVATMAL	2014	Rabi	Maize	1300	21.709611	2.053609	1015.803912	21.263478	silt	7.504	18.760	7.504	200.0
12625	137816	Maharashtra	YAVATMAL	2014	Rabi	Wheat	29100	21.851730	2.027476	1014.031903	20.059945	sandy	0.000	0.000	0.000	26800.0
12626	137817	Maharashtra	YAVATMAL	2014	Summer	Groundnut	9400	21.569380	2.004421	1013.989125	21.835158	chalky	2.896	3.620	0.000	11500.0
12627	137818	Maharashtra	YAVATMAL	2014	Whole Year	Sugarcane	8100	21.666723	2.008003	1015.081619	21.754799	loamy	3.150	0.000	0.000	553700.0

12628 rows × 19 columns

Correlation HeatMap

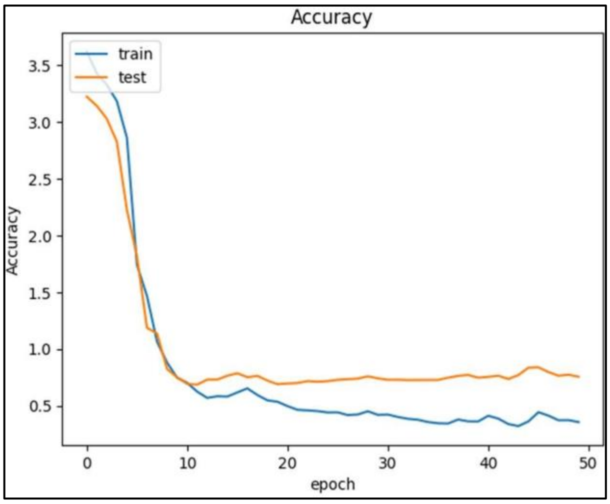


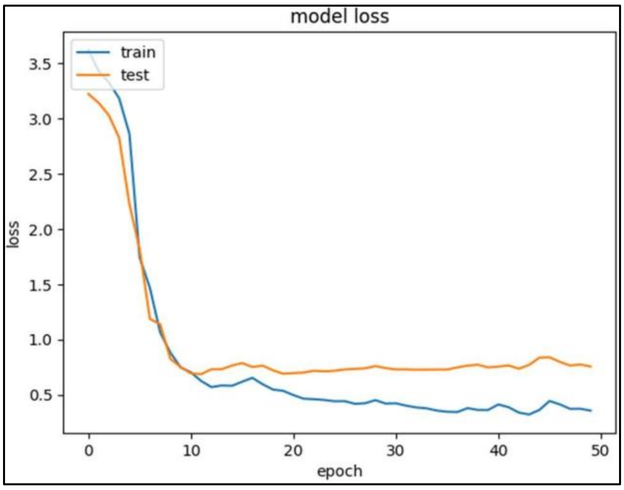
Get the Yield using Random Forest Regression

```
from sklearn.ensemble import RandomForestRegressor
regr = RandomForestRegressor(max_depth=2, random_state=0, n_estimators=100)
regr.fit(a_train, b_train)
b_pred = regr.predict(a_test)

from sklearn.metrics import mean_squared_error as mse
from sklearn.metrics import mean_absolute_error as mae
from sklearn.metrics import r2_score
print('MAE =', mae(b_pred, b_test))
print('Yield =', r2_score(b_pred, b_test))
print('Yield in percentage =', r2_score(b_pred, b_test)*100)
```

```
MAE = 0.8953650873829122
Yield = 0.9589614680509005
Yield in percentage = 95.89614680509005
```

Get the Yield using Support Vector Machines (SVM)

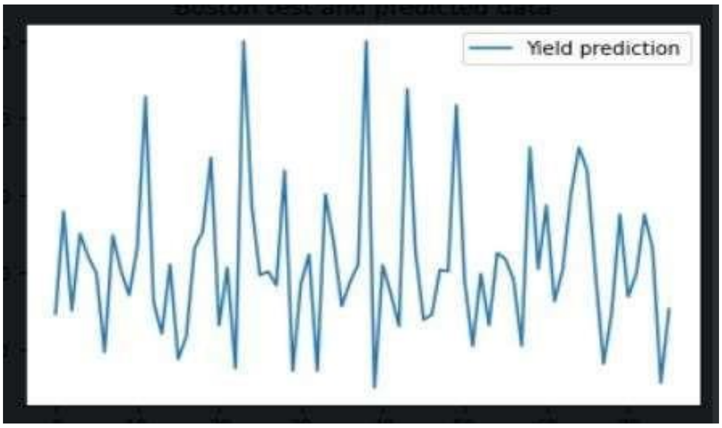





```
from sklearn.svm import SVR
regressorpoly=SVR(kernel='poly',epsilon=1.0)
regressorpoly.fit(a_train,b_train)
pred=regressorpoly.predict(a_test)
print("MAE=", regressorpoly.score(a_test,b_test))
print('Yield=', r2_score(b_test,b_pred))

MAE= 0.6312488835754394
Yield= 0.9598856437260073
```

Yield Prediction Using XGBRegression

```
[152] from xgboost import XGBRegressor
      from sklearn.metrics import mean_absolute_error
      XGBModel = XGBRegressor()
      XGBModel.fit(a_train,b_train , verbose=False)

      # Get the mean absolute error on the validation data :
      XGBpredictions = XGBModel.predict(a_test)
      MAE = mean_absolute_error(b_test , XGBpredictions)
      print('XGBoost validation MAE = ',MAE)
      print(r2_score(b_test , XGBpredictions))
      print("Yield  after apply XGboost :",(r2_score(b_test , XGBpredictions))*100)
      XGBpredictions

      XGBoost validation MAE =  0.6670485576475581
      0.9654928330252374
      Yield  after apply XGboost : 96.54928330252373
      array([1.0828081 , 0.64165634, 0.80906236, ..., 1.2118115 , 0.9264982 ,
             0.68652374], dtype=float32)
```



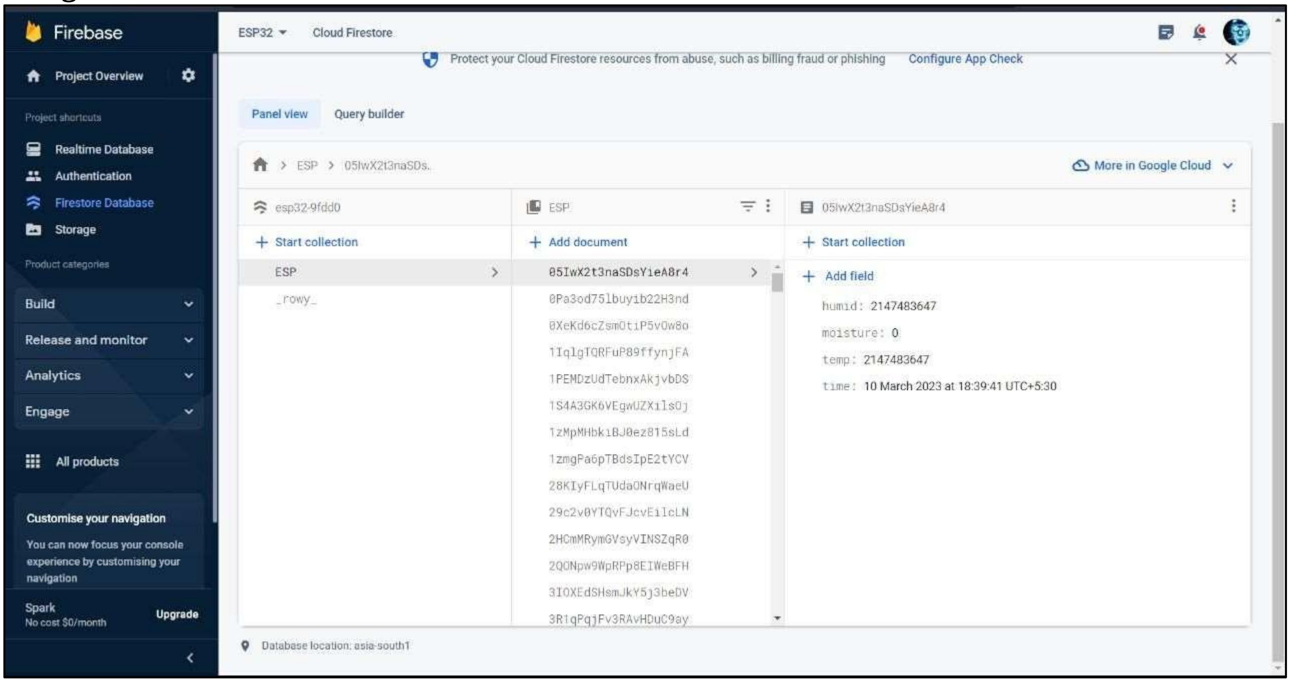
Yield Prediction Using LightGBM

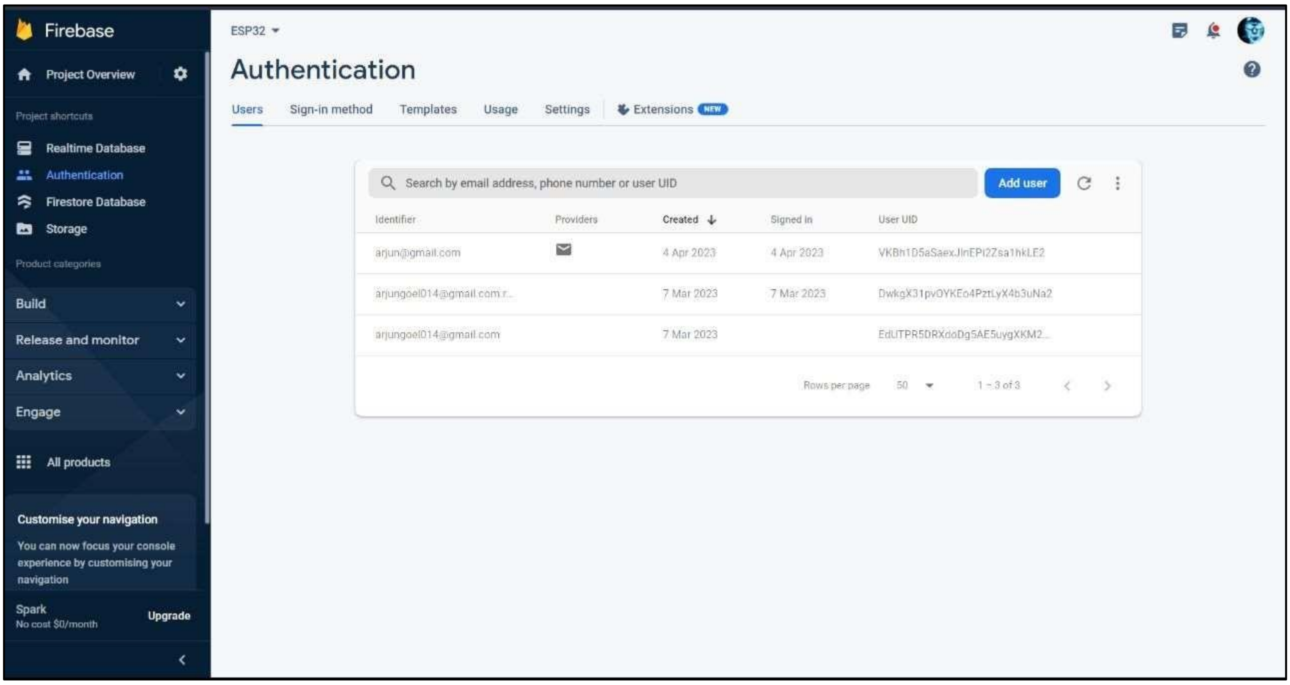
```
[221] import lightgbm as lbm
      model=lbm.LGBMClassifier()
      model.fit(a_train,b_train)
      b_pred=model.predict(a_test)
      from sklearn.metrics import accuracy_score
      score=accuracy_score(b_pred,b_test)
      print("Score: ",score)

      Score:  0.8407079646017699
```

Parameter Based Prediction of crop

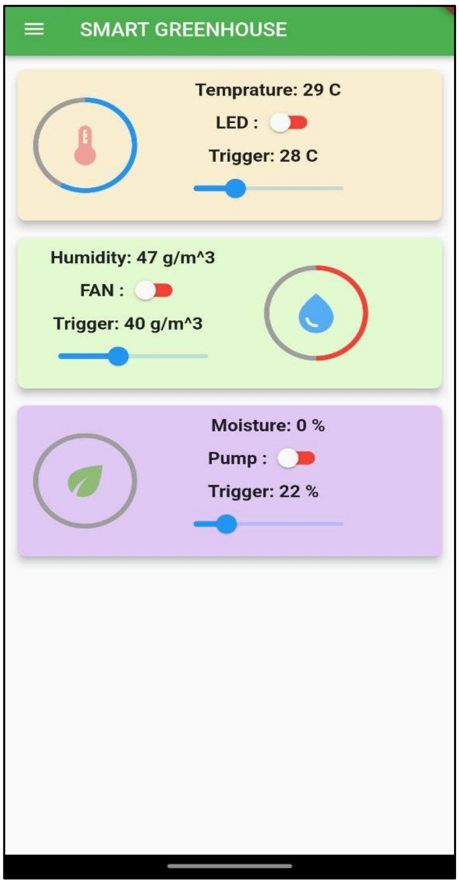
```
[181] model.predict([[ 20.933285, 2.011367, 1015.406196, 20.815299, 0.465, 0.465, 0.465, 182.653061]])

      array(['Cotton(lint)'], dtype=object)
```

Google Firebase





Application Interface: Build Using Flutter



Inference:

Crop yield prediction using machine learning involves leveraging the trained model to make predictions on new, unseen data. The model takes input data, such as current weather conditions, soil characteristics, and crop growth stages, and applies the learned patterns to generate crop yield predictions. These predictions can provide valuable insights to farmers and agricultural stakeholders, helping them make informed decisions about crop management practices, resource allocation, and marketing strategies to optimize crop yields and improve agricultural productivity.

Conclusion:

In conclusion, the project "Monitoring and Controlling of Various Parameters in Smart Greenhouse System & Yield Prediction Using ML" represents a significant stride towards sustainable and precision agriculture. Through the integration of advanced technologies, this initiative aims to address the challenges faced by traditional farming methods and provides a comprehensive solution for optimizing resource usage while maximizing crop yield.

The implementation of a Smart Greenhouse System, with its intricate network of sensors, has enabled real-time monitoring of key environmental parameters critical for plant growth. This meticulous monitoring serves as the foundation for an automated control system, dynamically adjusting factors such as irrigation, ventilation, and shading to create an environment tailored to the specific needs of the crops. The result is a more efficient and resource-conscious approach to agriculture that aligns with the principles of sustainability.

Furthermore, the incorporation of Machine Learning adds a layer of intelligence to the system, allowing for accurate yield predictions based on historical and real-time data analysis. This predictive capability empowers farmers with valuable insights into crop performance, enabling them to make informed decisions on crop management practices and resource allocation. The fusion of monitoring, control, and ML-driven predictions not only enhances the overall efficiency of agricultural operations but also contributes to the long-term viability of farming practices.

The project underscores the potential of technology in transforming traditional agriculture into a more resilient, adaptive, and productive system. The success of this endeavor signifies a step forward in harnessing the power of data, automation, and machine learning to meet the demands of a growing global population while minimizing the environmental impact of agriculture.

As we move forward, it is crucial to continue refining and expanding these technologies, fostering collaborations between agricultural experts, technologists, and data scientists. The insights gained from this project pave the way for future innovations in precision farming, smart agriculture, and sustainable food production. Through ongoing research and development, we can build upon these foundations to create a more resilient and sustainable future for global agriculture.