International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Predicting Humidity Levels by Applying Simple Regression Models in Greenhouses

Hrithvika Singh

Student, Grade XII
Email: *hrithvikasingh07[at]gmail.com*

Abstract: ***This research project aims to predict optimal humidity levels in greenhouses using machine learning models based on historical data collected on temperature and humidity. Through this project, we explore how to apply mathematical concepts such as statistics, linear regression, and algebra. The study compares the effectiveness of linear and polynomial regression models in predicting humidity levels. Sensors collect real time data, which is then processed using Python libraries. The results guide the automation of greenhouse systems, such as foggers, circulating fans, fan-pad system and sprinklers, ensuring precise environmental control. The integration with IoT technology enables remote monitoring and management, promoting resource efficiency and optimal plant growth conditions.***
website link: https://www.futuretechnurture.com

Keywords: machine learning models, green house, predicting humidity in green houses, regression model in green house operation, green house optimization

Purpose Statement

The purpose of this research study is to develop and evaluate regression models for predicting humidity levels in greenhouses, leveraging historical temperature and humidity data to optimize environmental control systems for improved plant growth and resource efficiency.

1. Introduction

Growing high quality plants with minimal resources is a challenging task for many. In a greenhouse, plant growth only occurs when the plant gets the right amount of sunlight, water, nutrients, and carbon dioxide for making its food by the process of photosynthesis.

In today's world, our resources are finite, which is why we need to conserve them and use them only when essential.

Hydroponics is soil-less farming and uses 90% less water than conventional farming as it recycles the water. All the essential nutrients needed for growth of a plant are added in the water.

This research demonstrates the potential of integrating machine learning models with IoT technology to enhance the efficiency of greenhouse operations, leading to better resource management and improved agricultural productivity.

Regression Analysis is used to predict and manage critical environmental parameters such as temperature and humidity. By analyzing historical data, the regression models can forecast future conditions, enabling proactive adjustments of environmental conditions inside the greenhouse. One of its main functions is to find the best-fit line from the data.

IOT sensors are used for collecting and monitoring data on temperature and humidity on a continuous basis.

2. Literature Review/Survey

In recent years, the advancements in technology particularly advancements in temperature controlled environments, hydroponic systems, machine learning, Internet of things (Iot) have produced promising solutions to enhance the efficiency and productivity of hydroponic greenhouses.

Hydroponics and Resource Efficiency
https://www.sciencedirect.com/science/article/pii/S2666154323002831#sec4
by Pathan, M.S., Kaur, S., and Singh, P. (2023). It discusses the significant role of IoT and smart sensors in modern agriculture, emphasizing their utility in measuring critical factors such as temperature, humidity, light intensity, and nutrient levels.

https://www.sciencedirect.com/science/article/pii/S0168169920302301
by Thomas van Klompenburg, Ayalew Kassahun, Cagatay Catal (2020). It talks about Crop yield prediction using machine learning.

Hypothesis
Controlling temperature and humidity in a hydroponic greenhouse using machine learning models. Further on using these machine learning models we aim to predict the data for temperature and humidity to maintain optimal growth of plants.

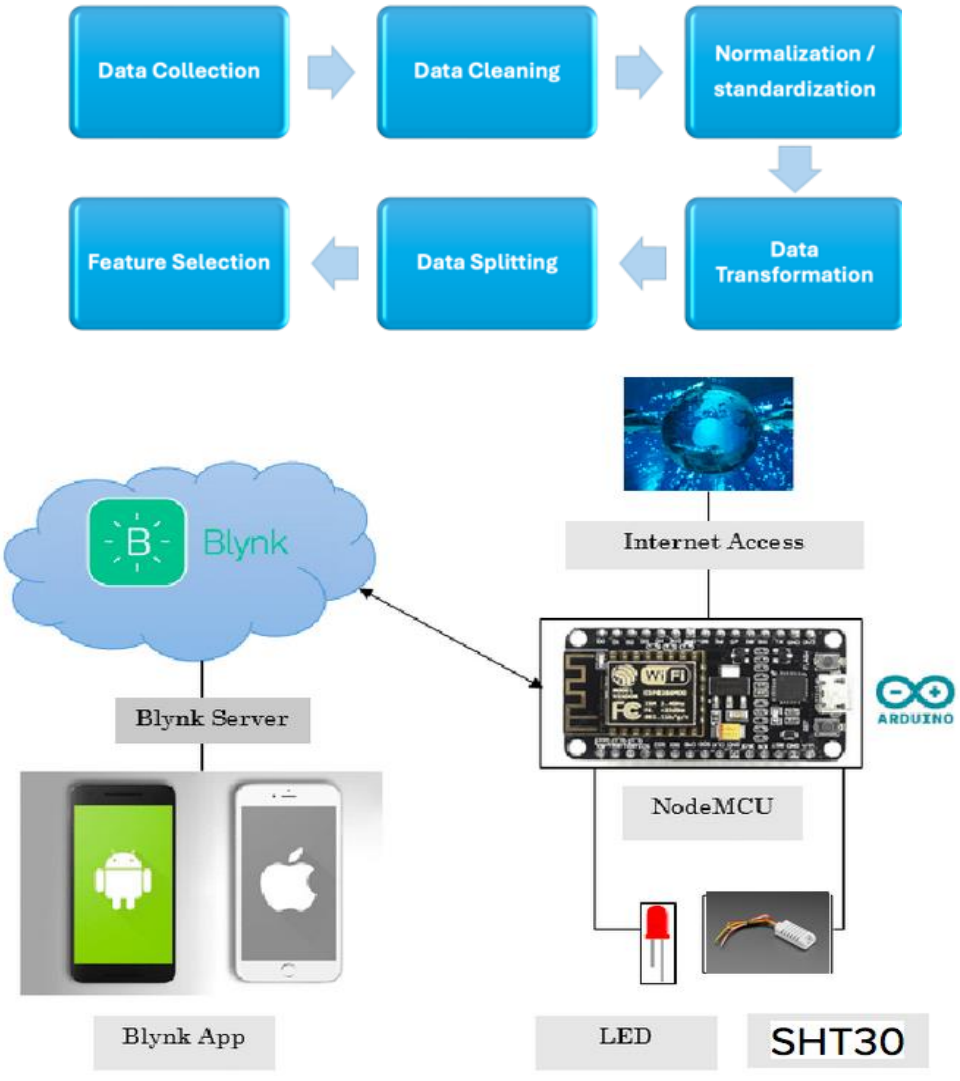
3. Methodology and Project Implementation

For improving the productivity of crops in the greenhouse, temperature and humidity are the most important factors. We use temperature and humidity sensors to collect data. For example, SHT-30(https://www.adafruit.com/product/5064) is one among the many sensors that can be used.

The data collected through the sensors can be processed by a microcontroller with a built-in wifi chip (https://store-usa.arduino.cc/products/nodemcu-esp8266). This microcontroller can also help to send the data collected to the cloud. Here, we are using the Blynk (https://blynk.io/) cloud provider to store the data in the cloud. From Blynk, we download this data and use it for further processing.

Once the data is available as a file, we apply python libraries such as Pandas to deal with empty cells, correct the data, and remove any duplicates.

Data preprocessing or Data cleaning techniques such as interpolation, deletion, or imputation handle missing data. Numerical features are normalized or standardized to prevent any single variable from dominating the model training process.



Source: Application of Wireless Internet in Networking using NodeMCU and Blynk App Scientific Figure on ResearchGate. https://www.researchgate.net/figure/How-Blynk-and-NodeMCU-works_fig1_346935386

The real time greenhouse data has been collected for the month of May 2024. The data is converted into a table format for easier graph creation. The collected sample data is represented below.

	Time	Temperature	Humidity
0	07/05/24 07:54:00 AM	30.790000	84.520500
1	07/05/24 07:53:00 AM	30.700667	85.057667
2	07/05/24 07:52:00 AM	30.608167	85.221833
3	07/05/24 07:51:00 AM	30.546833	85.189167
4	07/05/24 07:50:00 AM	30.402000	85.344167
5	07/05/24 07:49:00 AM	30.342333	85.295500
6	07/05/24 07:48:00 AM	30.286000	85.306333
7	07/05/24 07:47:00 AM	30.234500	85.628000
8	07/05/24 07:46:00 AM	30.198667	85.987000
9	07/05/24 07:45:00 AM	30.234333	86.329667
10	07/05/24 07:44:00 AM	30.180333	86.271833

Arduino: Sensor Code

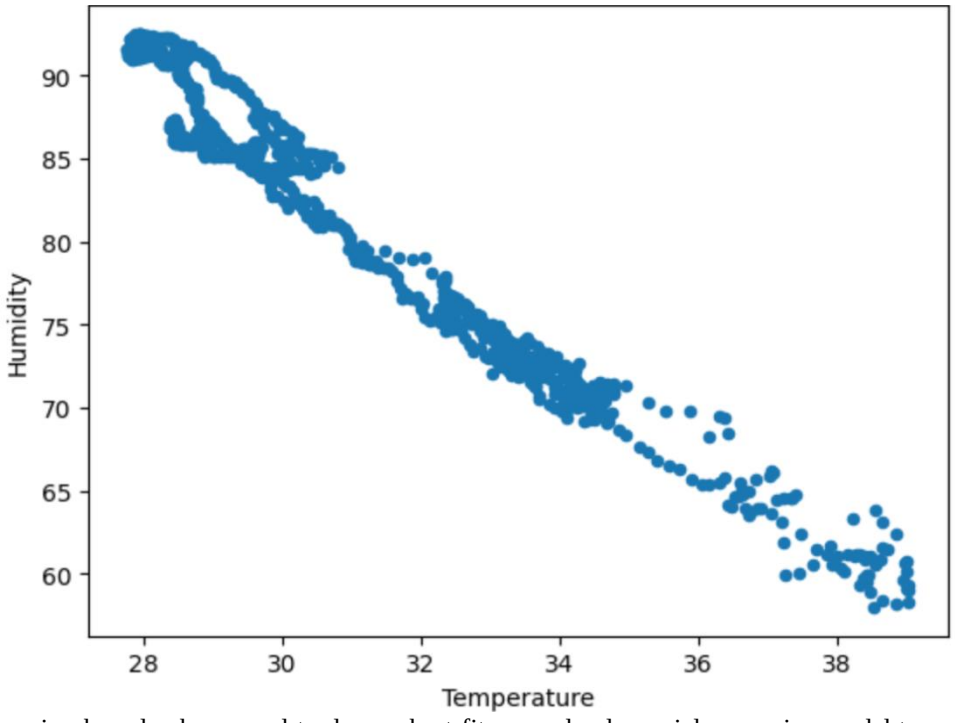
Please find a screenshot here. Full code can be accessed from Github link shared in References section

```
void setup()
{

  Serial.begin(115200);
  delay(100);

  BlynkEdgent.begin();
  Wire.begin(D6,D5);
  sht.begin();
  timer.setInterval(10000, sendTemp);

}

void sendTemp()
{
  sht.read();
  t=sht.getTemperature();
  h=sht.getHumidity();

  Serial.print("\t");
  Serial.print(t);
  Serial.print("\t");
  Serial.println(h);
  Blynk.virtualWrite(V10,t);
  Blynk.virtualWrite(V11,h);
}

void loop() {
  BlynkEdgent.run();
  timer.run();
}
```

We assign values to the independent variable (temperature) and the dependent variable (humidity) just as in linear equations.

Pandas library has been used to plot the data (scatter plot). After reviewing the scatter plot, linear and polynomial regression has been applied. Python libraries are again used to find out the slope of the linear line, the x and y intercepts,the correlation coefficient, the strength of the linear relationship and the standard error.

Green House Data Representation Using Scatter Plot



Polynomial regression has also been used to draw a best fit curved line. We have compared the linear regression model and polynomial regression model to see which will give us a much more efficient graph with the least error.

The past humidity data initially collected by the sensors is compared with current sensor data to determine the error between predicted and actual data.

Based on this error, a Python program is written to handle unexpected values or outputs. Python libraries such as Pandas, NumPy, SciPy, Matplotlib, and Seaborn are used. Pandas is used for data cleaning, manipulation, and analysis; NumPy supports efficient array computation and mathematical functions; SciPy provides additional functionality for scientific computing; Matplotlib creates visualizations; and Seaborn simplifies complex visualizations.

Physiology of Plants

In the greenhouse, transpiration by plants causes moisture to build up on the leaf surface, increasing the humidity. Due to this phenomenon and other thermal variables, it is necessary to maintain optimal humidity and temperature levels. The greenhouse has a fan pad system, foggers, sprinklers, and a ceiling net to block heat from the sun. The fan pulls air into the greenhouse, passing it through the pad system, which has water running through it to cool the incoming air. This cool and humid air replaces the hot air previously in the greenhouse.

To further control humidity, foggers fixed below the net are used to increase humidity or cool the temperature. The net above the foggers blocks 50 per cent of the sunlight and can be removed on cloudy days to allow more light in or put it back on sunny days to regulate light and heat.

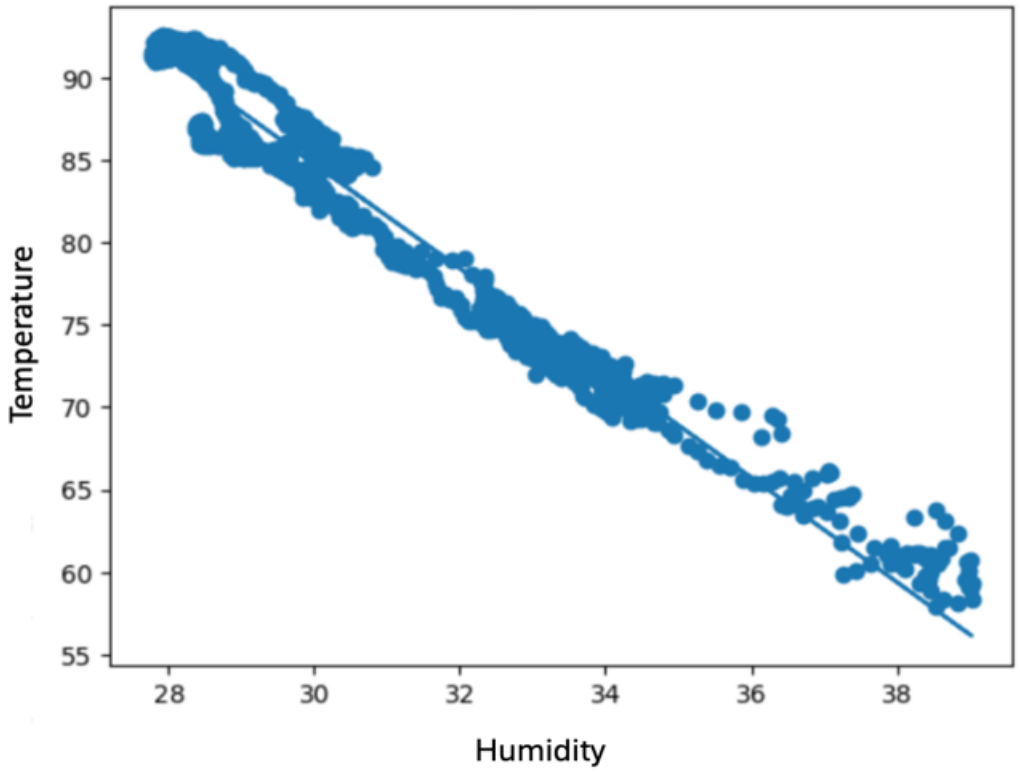
The Python program is integrated with the Blynk app, allowing remote control of the fan pad system, foggers, and sprinklers based on specific conditions set in the program. This integration enables remote operation through the Blynk app, ensuring efficient and automated environmental control in the greenhouse.

Materials Used

Category	Item
Devices	Temperature Sensors
	Humidity Sensors
	Foggers
	Sprinklers
	Ceiling Net
	Node MCU/Microcontroller
	Wifi Modules
	Blynk Cloud
Python Libraries	Pandas
	NumPy
	SicPy
	Scikit-Learn
	Matplotlib
	Seaborn
Machine Learning Models	Linear Regression, Polynomial Regression
Statistical Methods	Various Statistical Methods
Applications	Blynk App

4. Data Summary

Best-Fit Line (drawn using Linear Regression)



```
def predict_humidity(temperature, slope, intercept):

    return slope * temperature + intercept


x_value = 30.1033
predicted_humidity = predict_humidity(x_value, slope, intercept)
print(f"Predicted humidity for temperature {x_value} is {predicted_humidity}")
```

```
Predicted humidity for temperature 30.1033 is 84.42525591686942
```

```
def predict_humidity(temperature, slope, intercept):

    return slope * temperature + intercept


x_value = 32
predicted_humidity = predict_humidity(x_value, slope, intercept)
print(f"Predicted humidity for temperature {x_value} is {predicted_humidity}")
```

```
Predicted humidity for temperature 32 is 78.41594271548536
```

```
def predict_humidity(temperature, slope, intercept):

    return slope * temperature + intercept


x_value = 34
predicted_humidity = predict_humidity(x_value, slope, intercept)
print(f"Predicted humidity for temperature {x_value} is {predicted_humidity}")
```

```
Predicted humidity for temperature 34 is 72.07934420081877
```

```
x_value = 36
predicted_humidity = myfunc(x_value)
print(f"Predicted humidity for temperature {x_value} is {predicted_humidity}")
```

```
Predicted humidity for temperature 36 is 65.74274568615218
```

Analysis of Linear Regression Model

Using the equation of the best line we have predicted that value of humidity when temperature is given. We will now compare the predicted values with the real data (given below) found by the sensors and calculate the average error.

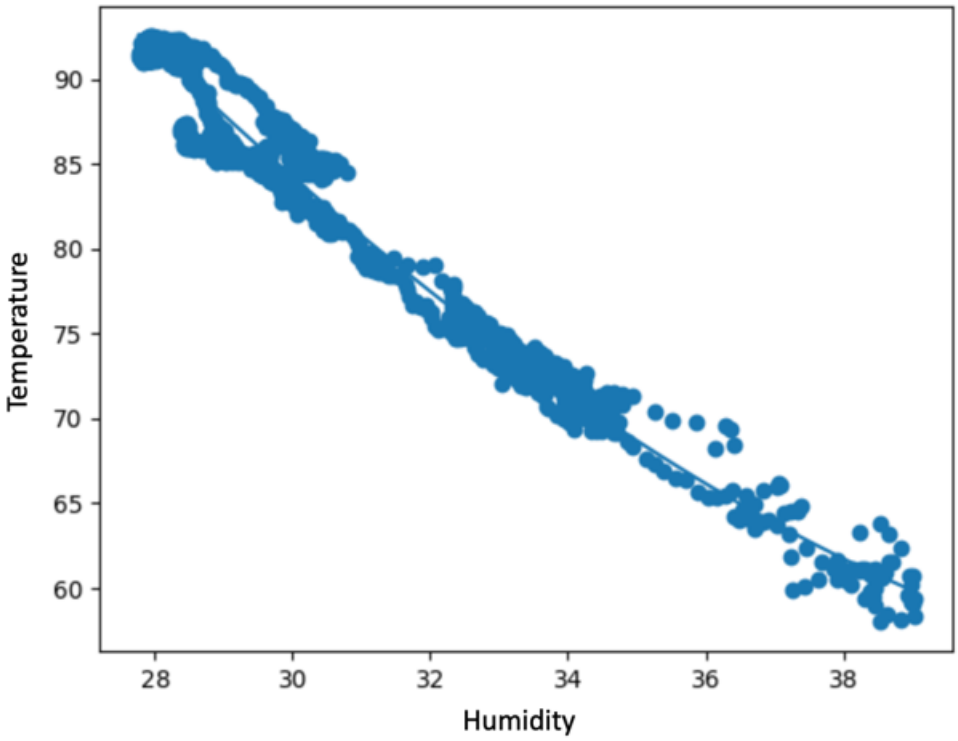
Temperature	Humidity
30.06083333	65.2035
32.0545	59.202
34.1945	53.6746666666667
36.04316666	65.2035

Error in Humidity
Where O is observed Humidity
P is predicted Humidity

1. Subtract the observed value from the predicted value:
$$P - O = 84.42 - 65.2 = 19.22$$
2. Square the result:
$$(P - O)^2 = (19.22)^2 = 369.0884$$
3. Since there is only one value, divide by 1:
$$\frac{369.0884}{1} = 369.0884$$
4. Take the square root of the result:
$$\sqrt{369.0884} = 19.22$$

So, the RMSE for the humidity values is 19.22. ●

Best fit-Curve (drawn using polynomial regression)




```
[ ]  x = weather_df['Temperature']
     y = weather_df['Humidity']
     mymodel = np.poly1d(np.polyfit(x, y, 3))


     new_temperatures = [30]

     predicted_humidities = mymodel(new_temperatures)

     for temp, humidity in zip(new_temperatures, predicted_humidities):
         print(f"Predicted humidity for temperature {temp} is {humidity}")
```

```
Predicted humidity for temperature 30 is 84.24130251214126
```

```
[ ] x = weather_df['Temperature']
    y = weather_df['Humidity']
    mymodel = np.poly1d(np.polyfit(x, y, 3))


    new_temperatures = [32]

    predicted_humidities = mymodel(new_temperatures)

    for temp, humidity in zip(new_temperatures, predicted_humidities):
        print(f"Predicted humidity for temperature {temp} is {humidity}")
```

```
Predicted humidity for temperature 32 is 77.47090033298167
```

```
[ ] x = weather_df['Temperature']
    y = weather_df['Humidity']
    mymodel = np.poly1d(np.polyfit(x, y, 3))


    new_temperatures = [34]

    predicted_humidities = mymodel(new_temperatures)

    for temp, humidity in zip(new_temperatures, predicted_humidities):
        print(f"Predicted humidity for temperature {temp} is {humidity}")
```

```
Predicted humidity for temperature 34 is 71.41510832351884
```

```
[ ] x = weather_df['Temperature']
    y = weather_df['Humidity']
    mymodel = np.poly1d(np.polyfit(x, y, 3))


    new_temperatures = [36]

    predicted_humidities = mymodel(new_temperatures)

    for temp, humidity in zip(new_temperatures, predicted_humidities):
        print(f"Predicted humidity for temperature {temp} is {humidity}")
```

```
Predicted humidity for temperature 36 is 66.14122747798882
```

Analysis Of Polynomial Regression Model

Using the equation of the best line we have predicted that value of humidity when temperature is given. We will now compare the predicted values with the real data (given below) found by the sensors and calculate the average error.

Temperature	Humidity
30.06083333	65.2035
32.0545	59.202
34.1945	53.6746666666667
36.04316666	65.2035

Error in Humidity
Where O is observed Humidity
P is predicted Humidity

1. Subtract the observed value from the predicted value:

$$P - O = 84.24 - 65.2 = 19.04$$

2. Square the result:

$$(P - O)^2 = (19.04)^2 = 362.5216$$

3. Since there is only one value, divide by 1:

$$\frac{362.5216}{1} = 362.5216$$

4. Take the square root of the result:

$$\sqrt{362.5216} = 19.03$$

So, the RMSE for the humidity values is 19.03.

The predicted value of 84.24 has less error (RMSE = 19.03) compared to the predicted value of 84.42 (RMSE = 19.22).
Cause of error in both the cases
1) Extreme temperature and humidity values
2) Fault with sensors

Action to be taken based on Predicted Value
This means that in order to maintain the high humidity levels, our model will instruct the green house workers to close the overhead curtain and start foggers. If the prediction would have been on the lower side, we would have opened the overhead curtains and turned off the water in the fan-pad system.

Temperature Humidity: Greenhouse Data
(Graphical Representation)

Blynk Automated Control Panel

5. Result

Using our model, one can work with the predicted data and real data, find errors, and change the temperature or humidity accordingly. Our model ensures that the crops are grown up to their utmost potential as best growth conditions are maintained. It saves water, conserves energy, and uses a limited amount of electricity, helps hydroponic farmers manage a suitable environment to grow crops in and allows them to manage and adjust the environment with ease.

Integration with the Blynk app and IoT technology allows for remote monitoring and controls the greenhouse environment. Farmers can access real-time data and manage the system from anywhere in the world, providing greater flexibility and convenience.

6. Conclusion

This research study successfully integrates machine learning models with automated systems to optimize humidity levels in greenhouses. By leveraging historical data and IoT technology, the model provides accurate predictions for environmental control, enhancing resource efficiency and agricultural productivity.

The innovative use of sensor technology and data analytics allows for real-time monitoring and precise control of the greenhouse environment. This not only helps in detecting and correcting errors promptly but also ensures consistent and optimal conditions for plant growth. The model's ability to work with both predicted and real-time data further enhances its reliability and effectiveness.

The project emphasizes resource efficiency by significantly reducing water usage through hydroponic methods and optimizes energy consumption with automated environmental controls. The integration with the Blynk app and IoT technology provides farmers with the convenience of remote monitoring and control, enabling them to manage their greenhouse operations from anywhere in the world.

Overall, the project combines advanced technologies to improve agricultural practices, conserve resources, increase productivity, and pave a path for sustainable agriculture. Future work should focus on incorporating additional environmental factors and refining machine learning algorithms to further improve predictive accuracy.

7. Future Scope

While significant progress has been made, there are still gaps in the research that need to be addressed. Future studies should focus on developing more advanced machine learning algorithms tailored for specific crops and environmental conditions. There is also a need for improved sensor accuracy and reliability to enhance data quality. This literature review highlights the advancements in hydroponics, sensor technology, machine learning, and IoT that contribute to the development of automated greenhouses. By building on this existing knowledge, the current research aims to increase crop productivity while conserving resources, demonstrating the potential for sustainable and efficient agricultural practices.

Despite its efficiency in water usage and potential for higher yields,hydroponic farming faces several challenges. These challenges include maintaining precise control over environmental factors, dealing with the system failures, ensuring continuous and accurate monitoring, and managing the nutrient delivery systems. Since we are replacing soil as the growing medium, we need to use right mix of essential nutrients. Additionally, hydroponic systems often require significant initial investment and technical expertise to operate effectively.

8. Discussion

Increase in global population, high food demand have intensified optimization of agriculture production. To counter overuse of natural resources we are seeing a surge in greenhouse numbers with smart automation.

It paves the way for future research focused on both the production and consumption of greenhouses.

If this project were to be repeated, several improvements could be made to enhance its effectiveness and outcomes. These improvements can be categorized into technological enhancements, methodological refinements, and expanded research scope.

Integrating renewable energy sources, such as solar panels, could further reduce the greenhouse's reliance on external power sources.

Experimenting with more sophisticated machine learning algorithms, such as deep learning or ensemble methods, could improve the predictive accuracy of the models.mental variables and crop productivity.

Irrigation and fertilizers account for highest energy usage.

Using AI and machine learning efficiently decreases energy consumption in greenhouse farming and increases the yield.

Future research should focus on refining these technologies, expanding their applications, and integrating renewable energy sources to further reduce the environmental impact of agriculture.

In the future we can use more sophisticated machine learning models other than linear regression, specifically ones that provide more accuracy. This way, our model can be much more efficient.

Some models that we can use to do so are:

Time Series Analysis Models can be used to analyze data collected or recorded at specific time intervals.

Clustering Algorithms could be used to identify different environmental conditions and optimize the control systems accordingly.

Neural Networks, used to represent complex patterns and relationships in data such as non linear relationships.

Gradient Boosting Machines (GBM) can be used to improve the predictive models by combining weak models to form a strong predictive model.

Acknowledgements
I would like to acknowledge the help and support of my father who is a technocrat and an urban farmer. He not only helped me understand the mathematical concepts involved in analyzing the problem and finding suitable solutions but also helped with the understanding of how a greenhouse operates. He helped me understand the finer nuances of the Blynk app and how we could use it to understand the behavior of various parameters.

I would also like to acknowledge the support provided by my teacher Ms.Safia Sheik who helped me understand how to apply concepts of linear regression, machine learning models and helped me analyze models that would suit my research most appropriately. Together we analyzed a lot of data.

Lot of my understanding came from the Research Paper written by Saumya Singh and Charen Vishwa highlighting the working of smart greenhouse using machine learning.

Other reference articles also helped me understand the concept and take it forward.

References

[1] https://www.irjet.net/archives/V10/i10/IRJET-V10I10110.pdf
[2] Engineering, vol. 2017, Article ID 9324035, 25 pages, 2017. https://doi.org/10.1155/2017/9324035
[3] Humanitarian Technology Conference (GHTC), 2014 IEEE, Oct 2014, pp. 325– 332
[4] https://www.sciencedirect.com/science/article/abs/pii/S0168169917314710
[5] https://ieeexplore.ieee.org/document/8784034
[6] D. S. Kim, T. H. Shin, and J. S. Park, "A security framework in rfid multi-domain system," in Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on, April 2007, pp. 1227–1234.
[7] https://www.sciencedirect.com/science/article/abs/pii/S0168169917308803
[8] https://www.mdpi.com/1424-8220/18/8/2674
[9] https://www.researchgate.net/figure/How-Blynk-and-NodeMCU-works_fig1_346935386

Project Related Links

Github Repository
https://github.com/Hrithvikasingh07/futuretechnurture

Python program:
https://github.com/Hrithvikasingh07/futuretechnurture/blob/main/Greenhouse_version_2.ipynb

Arduino sensor program:
https://github.com/Hrithvikasingh07/futuretechnurture/blob/main/arduino-sensor.ino

Temperature-Humidity data file:
https://github.com/Hrithvikasingh07/futuretechnurture/blob/main/humiditytemperaturedata.csv

Course taken on Hydroponics
https://www.udemy.com/course/future-farms-hydroponics-course/

This course teaches us about the basic physiology of hydroponically grown plants, explaining the factors needed for plant growth, specifically the nutrients to be poured, pros and cons of hydroponics and more.

Course taken on Linear Regression
https://www.coursera.org/learn/regression-models

This course is carried on by a professor from John hopkins university. In the course, I have paid special concentration to the topics of linear regression, multivariable regression, polynomial regression, variance, error, standard deviation and more.

Author Profile

Hrithvika Singh is a student of 12th grade. Her interest and research led to the data which revealed that to reduce overutilization of natural resources, a large number of greenhouses are turning to automation globally, to meet the growing demand for food. However as the cost of automation is extremely high, in a developing country like India, farmers are not able to afford fully automated farms. Also cheap labour encourages farmers to undertake certain work manually. Most greenhouses are successfully not able to optimize production because they use manual interventions to control parameters like temperature and humidity. Also, the data could be more predictive hence call to action is sometimes delayed, affecting productivity. Author would like to deepen her research and develop a system that could monitor and predict the internal environment of naturally ventilated greenhouses and successfully predict when the vent should be opened and closed so that optimal temperature is maintained at all times, leading to optimization in production. She has gathered data during summer break at her parent's farm. As the data involves continuous variables like temperature and humidity, author used linear regression and polynomial regression to understand the relationship between dependent and independent variables.