RainReveal: Rainfall Prediction for Meteorological Department

A Project Report

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE WITH SPECIALIZATION IN

Internet of Things

Submitted by:

Ananya Singh (20BCS4585)
Siddharth Singh(20BCS6103)
Bhanu Yadav (20BCS6105)

Under the Supervision of:
Dr. Gaurav Soni

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

PUNJAB

May, 2023

CERTIFICATE

This is to certify that minor project entitled "**RainReveal: Rainfall Prediction for Meteorological Department"** in field Minor Project is a Bonafede work carried out in the sixth semester by "*ANANYA SINGH*, *BHANU YADAV, SIDDHARTH SINGH*" who carried out the project work under "*MR. GOURAV SONI"* supervision. In partial fulfilment for the award of the degree of Bachelor of Technology in Computer Science Engineering with specialization internet of things from during the academic year 2022
-2023.

(Signature of the HOD)
NAME & SIGNATURE

(Signature of the Supervisor)
NAME & SIGNATURE

ABSTRACT

This research paper focuses on the development and evaluation of an IoT-based rainfall prediction system. The system utilizes IoT sensors to collect data related to temperature, humidity, and atmospheric pressure, which are then processed using machine learning algorithms to predict rainfall patterns. The system's effectiveness was evaluated through extensive testing in real-world scenarios, and the results demonstrate its high accuracy and reliability in predicting rainfall.

Additionally, we discuss the potential applications of such a system in various industries, including agriculture, transportation, and emergency services. We also analyze the system's limitations and discuss potential areas for improvement.

Overall, the research highlights the potential of IoT-based systems in improving our ability to predict and mitigate the impact of natural disasters such as rainfall.

TABLE CONTENT

1. INTRODUCTION

The Internet of Things (IoT) is revolutionizing the way we interact with our environment by enabling the collection and analysis of large amounts of data from various sources. One such application is the prediction of rainfall, which can have a significant impact on various industries, including agriculture, transportation, and emergency services.

By utilizing IoT-based sensors and data analysis techniques, it is possible to predict rainfall with a high degree of accuracy, enabling timely decision-making and reducing the risk of damage to property and loss of life. In this research paper, we explore the use of IoT-based rainfall prediction models and evaluate their effectiveness in predicting rainfall patterns.We also analyze the potential benefits and limitations of such models, and discuss their implications for various industries and applications.

Traditional prediction techniques do not ensure much accuracy as most of the methods are based on a particular machine learning algorithm, which may not be the finest of all. Hence, to curb this problem, we have proposed to review the techniques for IoT-based rainfall prediction comparing machine learning algorithms namely Linear Regression, Decision Tree, Support Vector Machine and Random Forest to find out which one is the most algorithm for prediction. Finding out the most suitable algorithm for prediction can ensure accuracy in prediction making IoT based rainfall prediction more reliable and trustworthy.

1.1. Problem Definition

The problem being addressed by the IoT based rain fall prediction system is the difficulty in accurately predicting and monitoring rainfall in different regions. Traditional methods of measuring rainfall such as rain gauges and weather stations are often limited in their accuracy and scope. Additionally, with climate change and increasing variability in weather patterns, it is becoming even more challenging

Many of the times we come across situations where we check the weather or try to predict the chances of rainfall the next day but the other day, we get completely contrasting results. Traditional prediction techniques do not ensure much accuracy as most of the methods are based on a particular machine learning algorithm, which may not be the finest of all.

Hence, to curb this problem, we have proposed to review the techniques for IoT-based rainfall prediction comparing machine learning algorithms namely Linear Regression, Decision Tree, Support Vector Machine and Random Forest to find out which one is the most algorithm for prediction. Finding out the most suitable algorithm for prediction can ensure accuracy in prediction making IoT based rainfall prediction more reliable and trustworthy

1.2 Project Overview

Rainfall is one of the most critical natural phenomena that directly impacts human life, agriculture, and the environment. Accurate prediction of rainfall is crucial for disaster management, water resources management, and urban planning. Traditional methods of rainfall prediction rely on historical data, mathematical models, and satellite imagery. However, these methods have limitations in terms of accuracy, scalability, and cost. IoT-based rainfall prediction has emerged as a promising approach to overcome these limitations by leveraging the power of sensors, data analytics, and machine learning algorithms.

The IoT-based rainfall prediction system typically involves three stages: data acquisition, data processing, and prediction. The data acquisition stage involves the deployment of sensors in the field to capture rainfall data, such as rainfall intensity, duration, and frequency. The data processing stage involves the pre-processing of the raw data to remove noise, missing values, and outliers.

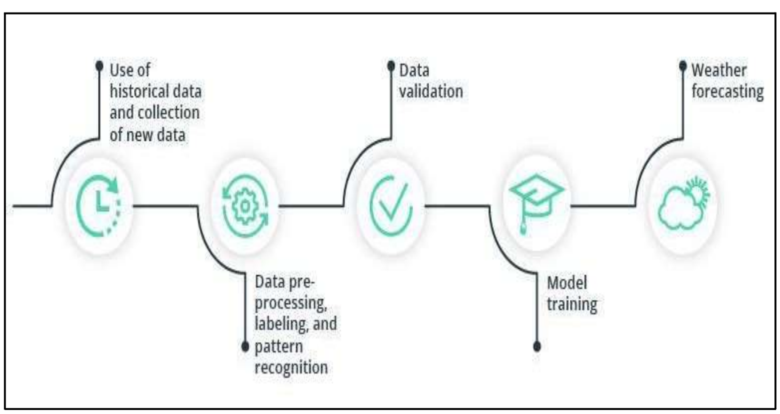

Fig 1. Schema of Machine Learning Model

The prediction stages involve the application of machine learning algorithms, statistical methods to predict rainfall on the processed data. has been introduced.

Several machine learning algorithms have been used for rainfall prediction, including decision trees, support vector machines, and artificial neural networks. Statistical methods such as regression analysis and time-series analysis have also been used. Artificial neural networks have shown promising results due to their ability to learn complex patterns in the data. However, these algorithms require large amounts of data and computational resources, which may pose a challenge in real-world scenarios.

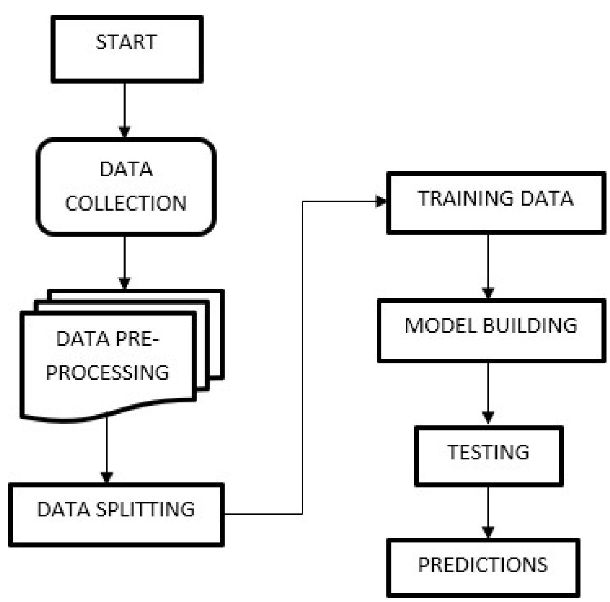

Fig 2. Schema of IoT based rainfall prediction system.

The traditional methods being used for weather prediction do not ensure much accuracy as most of the methods are based on a particular machine learning algorithm, which may not be the finest of all.

Hence, to curb this problem, we have proposed to review the techniques for IoT-based rainfall prediction comparing machine learning algorithms namely Linear Regression, Decision Tree, Support Vector Machine and Random Forest to find out which one is the most algorithm for prediction. Finding out the most suitable algorithm for prediction can ensure accuracy in prediction making IoT based rainfall prediction more reliable and trustworthy.

1.3. Hardware Specification

a) DHT11 Digital Temperature/Humidity Sensor

It is a module used for measuring temperature and humidity as shown In Fig. 6. It uses a capacitive humidity sensor and a thermistor in order to measure the Surrounding air's humidity and temperature.

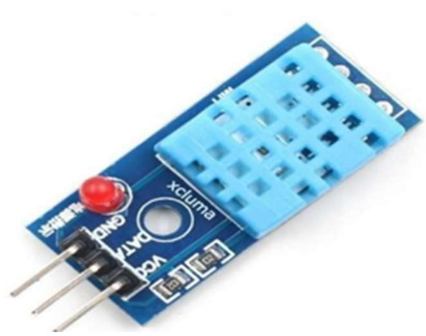

Fig. 3. DHT11 Module

b) V-Rain Drop Sensors

Raindrop sensors for VR (virtual reality) are used to simulate the sensation of raindrops falling on the user's body during VRexperiences. Raindrop sensors typically use haptic technology to simulate the feeling of raindrops, allowing users to feel the sensation of water droplets hitting their skin. Overall, raindrop sensors are a unique and innovative way to enhancethe immersive experience of virtual reality.

Fig. 4. V-Rain Drop Sensor

c) Wi-Fi Module (ESP8266)

Node MCU is an updated version of Arduino with inbuilt Wi-Fi chip as shown in Fig.5. It is cheaper than other modules performing the same function.

Fig. 5. NodeMCU Module

d) LDR Sensor

It is a device used for measuring the light density. It works on principle of photoconductivity. When LDR exposed in dark, its resistance decreases and when exposed in light its resistance increases as shown in Fig. 7.

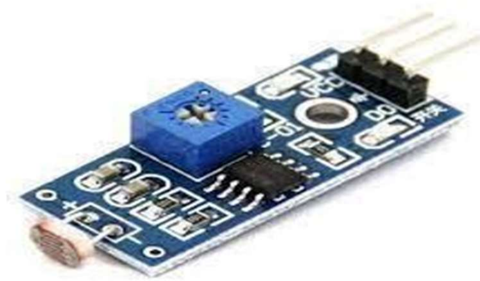

Fig.6. LDR Module

1.4. Software Specification

1. Rainfall Prediction Dataset (Kaggle)

This section provides a breakdown of the data sets used in the research. based on the sources of the data sets, availability, and geographical locations where the data sets were collected.

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7	159.8	458.3	2185.9	716.7
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.0	2957.4	156.7	236.1	1874.0	690.6
3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	571.0
4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7	25.4	344.7	2566.7	1.3	309.7	1624.9	630.8

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.200000	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.60000	3373.2000	136.3	560.300000	1696.3	980.300000
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.200000	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.50000	3520.7000	159.8	458.300000	2185.9	716.700000
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.000000	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.00000	2957.4000	156.7	236.100000	1874.0	690.600000
3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.000000	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.10000	3079.6000	24.1	506.900000	1977.6	571.000000
4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.300000	26.9	279.5	628.7	368.7	330.5	297.0	260.7	25.4	344.70000	2566.7000	1.3	309.700000	1624.9	630.800000
5	ANDAMAN & NICOBAR ISLANDS	1906	36.6	0.0	0.000000	0.0	556.1	733.3	247.7	320.5	164.3	267.8	128.9	79.20000	2534.4000	36.6	556.100000	1465.8	475.900000
6	ANDAMAN & NICOBAR ISLANDS	1907	110.7	0.0	113.300000	21.6	616.3	305.2	443.9	377.6	200.4	264.4	648.9	245.60000	3347.9000	110.7	751.200000	1327.1	1158.900000
7	ANDAMAN & NICOBAR ISLANDS	1908	20.9	85.1	0.000000	29.0	562.0	693.6	481.4	699.9	428.8	170.7	208.1	196.90000	3576.4000	106.0	591.000000	2303.7	575.700000
8	ANDAMAN & NICOBAR ISLANDS	1910	26.6	22.7	206.300000	89.3	224.5	472.7	264.3	337.4	626.6	208.2	267.3	153.50000	2899.4000	49.3	520.100000	1701.0	629.000000
9	ANDAMAN & NICOBAR ISLANDS	1911	0.0	8.4	0.000000	122.5	327.3	649.0	253.0	187.1	464.5	333.8	94.5	247.10000	2687.2000	8.4	449.800000	1553.6	675.400000
10	ANDAMAN & NICOBAR ISLANDS	1912	583.7	0.8	0.000000	21.9	140.7	549.8	468.9	370.3	386.2	318.7	117.2	2.30000	2960.5000	584.5	162.600000	1775.2	438.200000
11	ANDAMAN & NICOBAR ISLANDS	1913	84.8	0.5	1.300000	2.5	190.7	530.0	280.8	205.8	580.1	288.8	133.0	67.50000	2365.8000	85.3	194.500000	1596.7	489.300000
12	ANDAMAN & NICOBAR ISLANDS	1914	0.0	0.0	0.000000	37.7	298.8	383.3	792.8	520.5	310.8	139.8	184.4	289.70000	2957.8000	0.0	336.500000	2007.4	613.900000

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.200000	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.60000	3373.2000	136.3	560.300000	1696.3	980.300000
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.200000	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.50000	3520.7000	159.8	458.300000	2185.9	716.700000
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.000000	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.00000	2957.4000	156.7	236.100000	1874.0	690.600000
3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.000000	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.10000	3079.6000	24.1	506.900000	1977.6	571.000000
4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.300000	26.9	279.5	628.7	368.7	330.5	297.0	260.7	25.4	344.70000	2566.7000	1.3	309.700000	1624.9	630.800000
5	ANDAMAN & NICOBAR ISLANDS	1906	36.6	0.0	0.000000	0.0	556.1	733.3	247.7	320.5	164.3	267.8	128.9	79.20000	2534.4000	36.6	556.100000	1465.8	475.900000
6	ANDAMAN & NICOBAR ISLANDS	1907	110.7	0.0	113.300000	21.6	616.3	305.2	443.9	377.6	200.4	264.4	648.9	245.60000	3347.9000	110.7	751.200000	1327.1	1158.900000
7	ANDAMAN & NICOBAR ISLANDS	1908	20.9	85.1	0.000000	29.0	562.0	693.6	481.4	699.9	428.8	170.7	208.1	196.90000	3576.4000	106.0	591.000000	2303.7	575.700000
8	ANDAMAN & NICOBAR ISLANDS	1910	26.6	22.7	206.300000	89.3	224.5	472.7	264.3	337.4	626.6	208.2	267.3	153.50000	2899.4000	49.3	520.100000	1701.0	629.000000
9	ANDAMAN & NICOBAR ISLANDS	1911	0.0	8.4	0.000000	122.5	327.3	649.0	253.0	187.1	464.5	333.8	94.5	247.10000	2687.2000	8.4	449.800000	1553.6	675.400000
10	ANDAMAN & NICOBAR ISLANDS	1912	583.7	0.8	0.000000	21.9	140.7	549.8	468.9	370.3	386.2	318.7	117.2	2.30000	2960.5000	584.5	162.600000	1775.2	438.200000
11	ANDAMAN & NICOBAR ISLANDS	1913	84.8	0.5	1.300000	2.5	190.7	530.0	280.8	205.8	580.1	288.8	133.0	67.50000	2365.8000	85.3	194.500000	1596.7	489.300000
12	ANDAMAN & NICOBAR ISLANDS	1914	0.0	0.0	0.000000	37.7	298.8	383.3	792.8	520.5	310.8	139.8	184.4	289.70000	2957.8000	0.0	336.500000	2007.4	613.900000
13	ANDAMAN & NICOBAR ISLANDS	1915	45.0	56.7	33.300000	40.9	170.2	334.7	269.0	317.2	429.8	468.1	258.4	318.00000	2741.3000	101.7	244.400000	1350.7	1044.500000
14	ANDAMAN & NICOBAR ISLANDS	1916	0.0	0.0	0.000000	0.5	487.4	450.1	317.3	425.0	561.2	369.7	192.6	133.70000	2937.5000	0.0	487.900000	1753.6	696.000000
15	ANDAMAN & NICOBAR ISLANDS	1917	8.0	3.6	112.000000	4.5	295.9	301.1	394.8	437.4	471.8	238.1	108.3	236.90000	2612.4000	11.6	412.400000	1605.1	583.300000
16	ANDAMAN & NICOBAR ISLANDS	1918	77.4	6.9	11.400000	10.7	729.3	710.8	200.9	455.4	303.3	227.0	366.9	175.00000	3275.0000	84.3	751.400000	1670.4	768.900000
17	ANDAMAN & NICOBAR ISLANDS	1919	10.2	18.0	0.000000	35.5	283.9	542.5	246.5	259.8	170.7	186.2	340.4	258.40000	2352.1000	28.2	319.400000	1219.5	785.000000
18	ANDAMAN & NICOBAR ISLANDS	1920	122.3	7.4	3.100000	13.0	237.4	546.9	294.4	467.4	505.4	397.5	262.9	85.50000	2943.2000	129.7	253.500000	1814.1	745.900000
19	ANDAMAN & NICOBAR ISLANDS	1921	13.2	3.1	0.000000	37.5	351.2	282.7	487.1	330.0	581.2	360.7	118.2	41.50000	2606.4000	16.3	388.700000	1681.0	520.400000
20	ANDAMAN & NICOBAR ISLANDS	1922	245.3	34.3	15.600000	323.1	289.7	506.1	425.8	307.4	511.7	162.0	541.0	192.20000	3554.2000	279.6	628.400000	1751.0	895.200000
21	ANDAMAN & NICOBAR ISLANDS	1923	79.5	0.0	27.359197	91.3	293.5	808.4	636.9	182.2	560.5	131.9	197.4	70.60000	1411.0089	79.5	155.901753	2188.0	399.900000
22	ANDAMAN & NICOBAR ISLANDS	1924	28.7	0.0	14.800000	89.7	191.2	261.2	493.3	290.9	251.2	331.1	378.6	18.87058	1411.0089	28.7	295.700000	1296.6	154.100487
23	ANDAMAN & NICOBAR ISLANDS	1925	36.6	0.0	8.600000	50.4	282.2	663.8	241.8	278.2	201.9	249.5	271.5	196.00000	2480.5000	36.6	341.200000	1385.7	717.000000
24	ANDAMAN & NICOBAR ISLANDS	1926	122.1	0.0	0.000000	0.5	198.4	370.0	195.3	523.7	719.3	443.8	148.4	560.70000	3282.2000	122.1	198.900000	1808.3	1152.900000

2. Google Collab

Google Collab is a free cloud-based platform that provides users with a Jupiter Notebook environment for Python programming. It offers easy access to powerful computing resources, including GPUs and TPUs, to run machine learning models and data analysis tasks.

Additionally, Collab allows users to collaborate with others in real-time, making it a valuable tool for team projects or classroom settings. Overall, Google Collab is a versatile and convenient tool for data scientists, researchers, and educators alike

2. LITERATURE SURVEY

2.1 Existing System

There are several existing systems of rainfall prediction, ranging from traditional methods to modern technological systems. Here are a few examples:

1. **Traditional forecasting methods:** These methods are based on local knowledge, experience, and observation of weather patterns. They include indicators such as the behavior of animals and plants, cloud formations, wind direction, and humidity.

2. **Statistical methods:** These methods use historical rainfall data to analyze trends and patterns and forecast future rainfall. These methods include regression analysis, time-series analysis, and probability-based models.

3. **Numerical weather prediction (NWP):** This method uses computer models and mathematical equations to simulate atmospheric conditions and predict future weather patterns.

4. **Remote sensing:** This method uses satellite imagery to detect and track weather patterns, such as cloud formations and precipitation, and predict rainfall.

5. **Artificial Intelligence (AI):** This method uses machine learning algorithms to analyses large datasets of weather data, including historical data and real-time data.

These are just a few examples of existing systems of rainfall prediction, and new techniques are being developed and refined allthe time.

2.2 Proposed System

In this, we present the theory on rainfall prediction through IOT. In this proposed block diagram several sensors (LDR sensor, Humidity sensor, temp sensor, Rain Sensor) are connected to our controller. The controller is accessing the sensor values, processing them and displaying them on LCD Display and uploading data over the web server. All sensors interface with microcontrollers.

Updated readings are sent into a Wi-Fi module that translates the data into a graphical and statistical manner. Improving the accuracy of machine learning techniques on weather forecasting has been the primary concern of many researchers over decades.

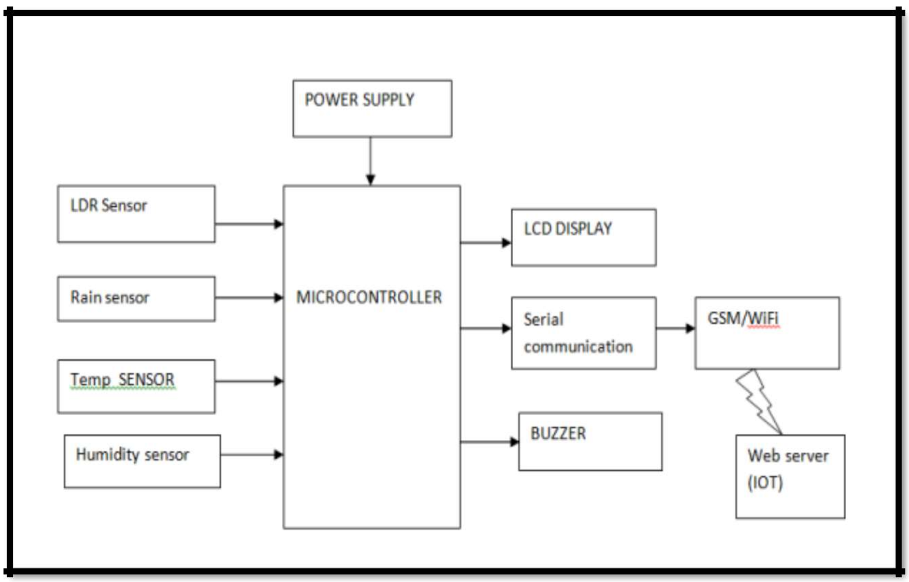

Fig 7 : Flowchart

Here will apply certain machine learning algorithms in the data set and compare the accuracy and other analysis terms. Some of techniques we used are regression, decision tree, supported vector machine, random forest.

3.PROBLEM FORMULATION

Objective: The objective of the problem is to predict rainfall accurately and in a timely manner using IoT sensors and data analysis techniques.

Several machine learning algorithms have been used for rainfall prediction, including decision trees, support vector machines, and artificial neural networks. Statistical methods such as regression analysis and time-series analysis have also been used. Artificial neural networks have shown promising results due to their ability to learn complex patterns in the data. However, these algorithms require large amounts of data and computational resources, which may pose a challenge in real-world scenarios.

Traditional prediction techniques do not ensure much accuracy as most of the methods are based on a particular machine learning algorithm, which may not be the finest of all.

Hence, to curb this problem, we have proposed to review the techniques for IoT-based rainfall prediction comparing machine learning algorithms namely Linear Regression, Decision Tree, Support Vector Machine and Random Forest to find out which one is the most algorithm for prediction. Finding out the most suitable algorithm for prediction can ensure accuracy in prediction making IoT based rainfall prediction more reliable and trustworthy.

.

4.RESEARCH OBJECTIVES

The objective of IoT-based rainfall prediction is to accurately and timely predict rainfall using IoT sensors and data analysis techniques. The prediction of rainfall is important for various applications such as agriculture, water resource management, flood control, and disaster preparedness.

By using IoT sensors to collect data on weather conditions, such as rainfall, humidity, and temperature, and applying data analysis techniques such as machine learning, the system can predict rainfall patterns and alert stakeholders in advance. This can help farmers make informed decisions on when to plant crops and irrigate, water resource managers to plan for water availability, and emergency services to prepare for potential floods or other weather-related disasters.

In summary, the objective of IoT-based rainfall prediction is to provide accurate and timely information on rainfall patterns to aid in decision-making and disaster preparedness, ultimately improving the overall quality of life and safety of communities.

5.METHODOLOGY

The methodology for IoT-based rainfall prediction can be divided into the following steps:

1. *Data Collection:* Collecting data from IoT sensors such as weather stations, rainfall gauges, and soil moisture sensors is the first step in the process. This data is collected at regular intervals and transmitted to a central server for further analysis.

2. *Data Preprocessing:* Once the data is collected, it needs to be preprocessed to remove any noise or inconsistencies. This involves cleaning the data, filling in any missing values, and converting the data into a suitable format for analysis

3. *Feature Selection:* After the data is preprocessed, relevant features need to be selected that have a high correlation with rainfall. This can be done using techniques such as correlation analysis, principal component analysis (PCA), and feature importance analysis.

4. *Machine Learning Models:* Various machine learning models such as Decision Trees, Random Forests, Neural Networks, Support Vector Machines, and Gradient Boosting can be used to develop predictive models based on the selected features.

5. *Model Training and Testing:* The selected machine learning model is then trained and tested using the collected data. The model is trained on a portion of the data and tested on the remaining data to evaluate its accuracy.

6. *Model Deployment:* Once the model is trained and tested, it is deployed into the IoT system. The deployed model will receive input data from the IoT sensors and provide real-time rainfall predictions.
7. *Model Maintenance and Improvement:* The final step involves maintaining and improving the deployed model. This involves monitoring the system's performance and retraining the model with new data to improve its accuracy

8. Linear Regression: Train a linear regression model on the training data. Linear regression is a statistical method that uses a linear approach to model the relationship between the input variables and the output variable. In this case, the input variables would be the relevant weather parameters and the output variable would be the rainfall measurement.

9. Random Forest: Train a random forest model on the training data. Random forest is a machine learning algorithm that uses an ensemble of decision trees to model the relationship between the input variables and the output variable.

10. Support Vector Machine: SVM is a powerful supervised algorithm that works best on smaller datasets but on complex ones. Support Vector Machine, abbreviated as SVM can be used for both regression and classification tasks, but generally, they work best in classification

problems.

11. Model Evaluation: Evaluate the performance of both models using the testing data. Use metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared to measure the accuracy of the models.

12. Model Comparison: Compare the performance of both models to determine which one is better for rainfall prediction.

13. Deployment: Once the best model is selected, deploy it to the IoT devices to predict rainfall in real-time.

In summary, the methodology for IoT-based rainfall prediction involves data collection, pre-processing, feature selection, machine learning modeling, model training and testing, model deployment, and model maintenance and improvement. By following these steps, an accurate and reliable IoT-based rainfall prediction system can be developed.

6.RESULTS AND ANALYSIS

Fig. 8. Visualization using Heatmap

Heatmap is representation of data graphically using colors to visualize the values of the matrix. In this particular map, the darker the color, the value is of that much significance. It shows the data that can have significant impact on the trained model. This heatmap helps us visualize the data in a better way, so that we can drop all the unnecessary data variables.

In this, to represent more common values or higher activities brighter colors basically red colors are used and to represent less common or activity values, darker colors are preferred.

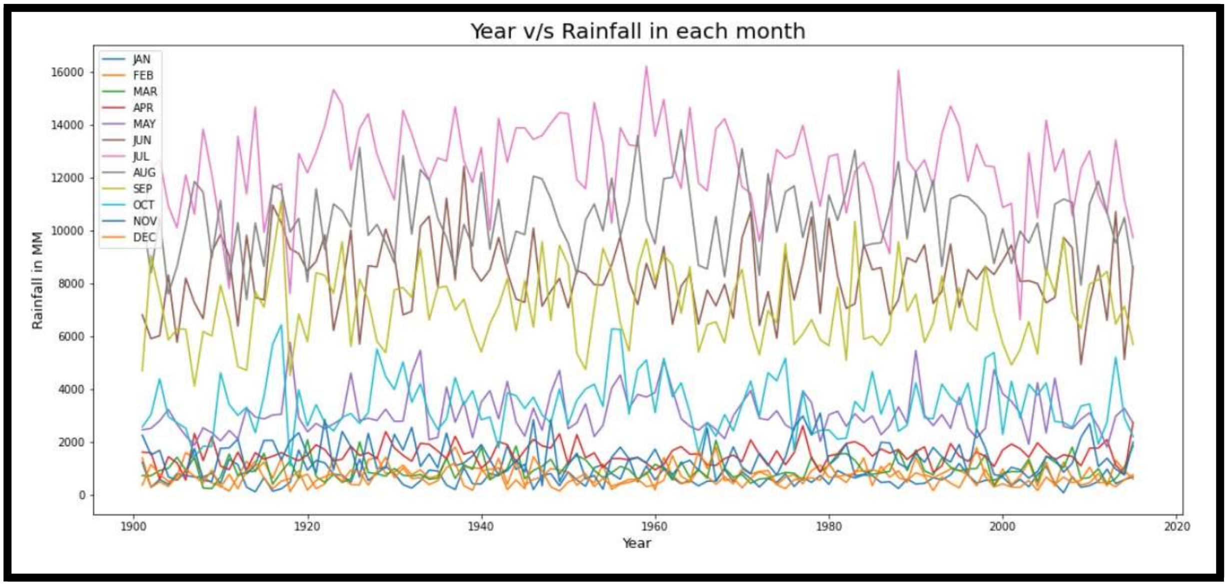

Fig.9. Year vs Rainfall in each year

Line plots are also known as line charts or time-series plots, depending on the type of data being visualized. They are typically used to display quantitative data, such as numerical values or measurements, and can be used to compare multiple data sets or track changes in a single data set over time.

The above figure shows the amount of rainfall according to Years v/s Rainfall in each month. Different colors used in the plot depicts the twelve months in a year. Further, the plot depicts how much amount of rainfall took place in millimeters in a specific month of the year.

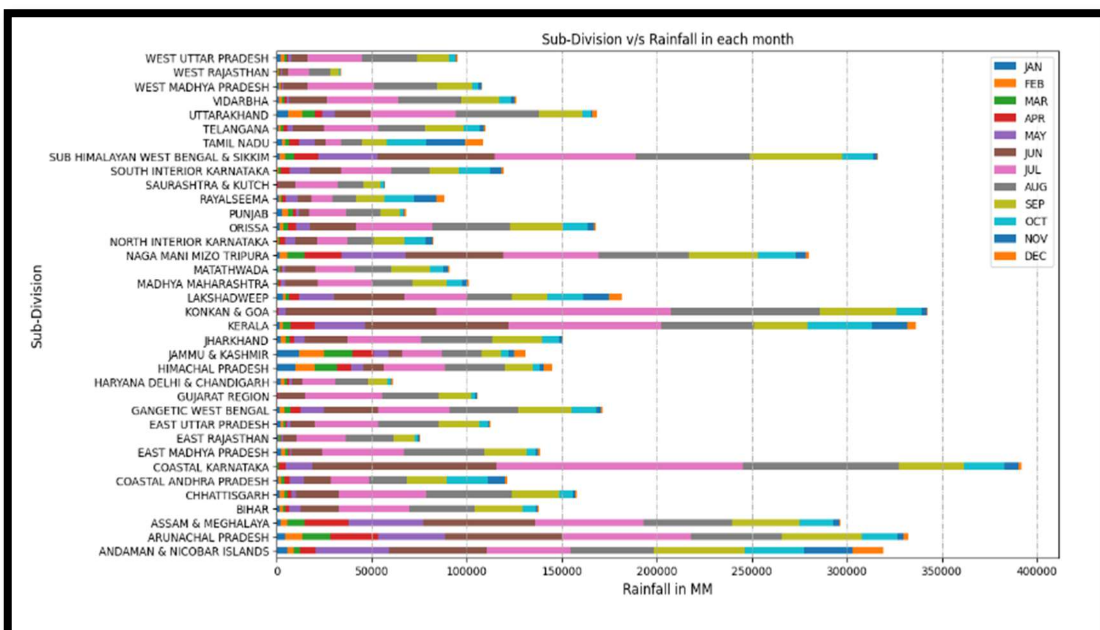

Fig.10. Year vs Rainfall in each year

The above plot depicts the amount of rainfall that took place in the several months of the year in a particular state. “Sub Himalayan West bengal and Sikkim” and “Coastal Karnatka” are the regions with maximum rainfall over the year. Moreover, “West Rajasthan” and “Saurashtra and Kutch” only received rainfall spread over the region of approximate (00-53000) MM in a year.

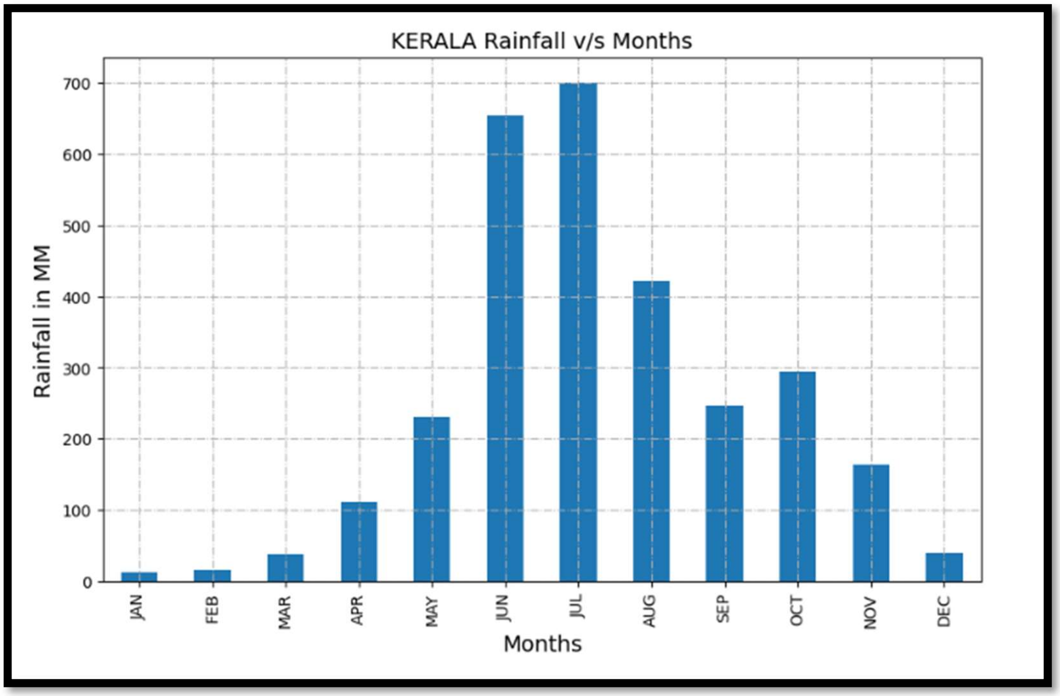

Fig. 11. Visualisation using Bar graph in Kerela

The above figure shows the rainfall in millimeters in specific month over the years for the state of Kerela specifically.It is a bar plot which is plotted to get an visual idea of the maximum and minimum rainfall in Kerela.

The above plot visualizes and shows the month of July seen the maximum rainfall over the years in Kerela while minimum amount of rainfall took place in the month of January.

Linear Regression Model

```
-------Test Data--------
MAE: 102.15453952323351
MSE: 16248.327064937163
RMSE: 127.46892587974986
Explained Variance Score: 0.027561502102628865 2

-------Train Data--------
MAE: 95.80008074338575
MSE: 14657.15403937771
RMSE: 121.06673382633939

-----Training Accuracy-------
4.3999999999999995
-----Testing Accuracy--------
2.5
```

Fig.13. Linear Regression Model

The algorithm on implementation over the Test data has given Explained Variance Score of 0.027. The score depicts the Dispersion of errors of the given dataset. The algorithm stands. out with the Mean absolute error of 102.15, Mean Squared Error (MSE) of 162248.32 and Root Mean square value of 127.46.

Random Forest Model

```
-------Test Data--------
MAE: 37.86719737995522
MSE: 3296.013536058157
RMSE: 57.410918265240774

-------Train Data--------
MAE: 28.418216663016555
MSE: 2016.8954858814905
RMSE: 44.909859562032594

[44] print("-----------Training Accuracy-----------")
print(round(random_forest_model.score(X_train,y_train),3)*100)
print("-----------Testing Accuracy-----------")
print(round(random_forest_model.score(X_test,y_test),3)*100)

-----------Training Accuracy-----------
86.8
-----------Testing Accuracy-----------
80.2
```

Fig.14. Random Forest Model

The score depicts the Dispersion of errors of the given dataset. The algorithm stands. out with the Mean absolute error of 37.86, Mean Squared Error (MSE) of 3296.013 and Root Mean square value of 57.410.

Supported vector Machine

```
-------Test Data--------
MAE: 97.98188405797102
MSE: 26265.097826086956
RMSE: 162.0651036654312

-------Train Data--------
MAE: 89.10054347826087
MSE: 23271.352355072464
RMSE: 152.54950788210516
```

```
52] print("-----------Training Accuracy------------")
    print(round(svm_regr.score(X_train,y_train),3)*100)
    print("-----------Testing Accuracy------------")
    print(round(svm_regr.score(X_test,y_test),3)*100)
```

```
-----------Training Accuracy------------
16.5
-----------Testing Accuracy------------
15.9
```

Fig.15. Support Vector Machine

The score depicts the Dispersion of errors of the given dataset. The algorithm stands. out with the Mean absolute error of 97.98, Mean Squared Error (MSE) of 23271.35 and Root Mean square value of 152.54

7.CONCLUSION AND FUTURE SCOPE

The research that is conducted in this paper shows that using IoT sensor like DHT-11, V-rain-LDR etc we can help in improvement in rainfall prediction by including Machine Learning algorithms. Th study conducted shows that using different models of Machine Learning such as Linear Regression, Support Vector machine and Random Forest Model we can improve the existing rainfall prediction methods. The study shows that Random Forest is best suited for implementation of this type of system as it gives less error in prediction and the highest accuracy.

The rainfall prediction can further be improved with use of better sensors and more accurate models There are many different challenges when it comes to rainfall prediction such as data scalability, security etc. When taking into consideration all these factors with the dataset at hand it shows that for this type of prediction Random Forest is most suited.

8.REFERENCES

1) S. G. Patil and S. S. More, “A linear regression-based approach for rainfall prediction using IoT sensors". in IEEE International Student conference (2020).

2) "IoT-Based Rainfall Prediction Using Linear Regression" by S. S. Rathod and A. S. Kale. (2019) in IEEE International Congress,pp. 557–564.

3) Mr. G.G. Rathod, Omkar Hase, Yash Nehe, Vilas Chaudhari, “Rainfall Prediction through IOT And Artificial Intelligence” International Research Journal of Engineering and Technology (IRJET), vol. 19, pp. 182–189, 2012.

4) J. Shivang, S. S. Sridhar, “Weather prediction for Indian location using Machine learning,” International Journal of Pure and Applied Mathematics, vol. 118, no. 22 pp.1945-1949, 2018M. Conti, E. S. Kumar, C. Lal, and S. Raj, “A survey on security and privacy issues of bitcoin,” IEEE Communications Surveys & Tutorials, Vol. 20, no. 4, pp. 3416–3452, 2018.

5) S. S. Bhatkande1, R. G. Hubballi2, “Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques.” Belgaum India: International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no.5, pp. 483-48, 2016.

6) Y. Radhika, and M. Shashi, “Atmospheric temperature prediction using support vector machines.” International Journal of Computer Theory and Engineering 1.1, vol. 1, no. 1, pp.1793-8201, 2009.

7) D. Chauhan, J. Thakur, “Data mining techniques for Weather Prediction:” International Journal of Computer Science Trends and Technology (IJCST), vol. 6, issue 3, pp.249-254, 2018.

8) S.S. Gadhiya, B. V. Wakode, P. N. Chatur, “Analysis Of Temperature And Humidity Data For Future Value Prediction” International Journal Of Computer Science And Information Technologies , vol. 3, no.1 pp.3012-3014, 2012.

9) G. J. Sawale, S. R. Gupta, “Use of Artificial Neural Network in Data Mining For Weather Forecasting”, International Journal Of Computer Science And Applications vol. 6, no.2, pp.383-387, 2013.

10) A. Gautam, G. Verma, S. Qamar, S. Shekhar, “Vehicle Pollution Monitoring, Control and Challan System Using MQ2 Sensor Based on Internet of Things”, Wireless Personal Communications (An International Journal of Springer with impact factor of 1.2), November 2019.

5.1 F. Olaiya, Adesesan Barnabas Adeyemo, “Application of Data Mining Techniques in Weather Prediction and Climate Change Studies”, IJCSNS International Journal of Computer Science and Network Security, vol.17, no.6, pp 22-25, 2017 .