Improving Water Resource Management For Agriculture In Pune,Maharashtra:A Data Driven Approach to Rainfall Forecasting and Irrigation Planning

Manasi Phadatare[1], Shailesh Jadhav[2], Ilihas Patel,[2] Atharv Kulkarni[2] and Parth Kadav[2]

AISSMS College Of Engineering
Pune, Maharashtra ,India.

Abstract: Rainfall prediction and water supply management are essential for agriculture as they enable farmers to make informed decisions, optimize crop yields, mitigate risks, conserve resources, and contribute to food security and economic stability. This paper explores the methodology for predicting rainfall by analyzing key parameters, including maximum temperature, minimum temperature, maximum humidity, minimum humidity, evaporation, average wind speed, and wind direction of Pune district Maharashtra. Furthermore, it takes into account the water requirements of commonly cultivated crops and using the xyz algorithm amount for water that needed to supplied to crops is determined. This holistic analysis is conducted by using 3 major machine learning models XGBoost, Linear regression, Random forest and we found that XGBoost Algorithm have given highest accuracy of 90 percent among all three algorithms.

Keywords: Water management, Irrigation, Smart Agriculture, IoT Sensors, XGBoost, Linear regression, Random forest, Rainfall prediction.

1 INTRODUCTION

The role of rainfall prediction and water supply management is indispensable. As the impact of climate change continues to disrupt conventional weather patterns, the need for precise forecasting of rainfall and efficient water resource allocation becomes increasingly vital. This paper endeavours to explore the intricate interplay between rainfall prediction and water supply management in agriculture, recognizing their collective significance in optimizing crop yields, by addressing these challenges and advancing innovative strategies. The challenge of managing water supply to crops is a delicate balancing act, as both excess and deficiency can have detrimental consequences. Oversupply can lead to issues like salinity, while undersupply may result in reduced crop yields. Thus, estimating the right amount of water is paramount for enhancing agricultural productivity. Extensive research has identified that specific meteorological features like **temperature, humidity, evaporation, average wind speed, total rainfall** (see Table 1) as critical for precise rainfall forecasting in a given region. The data for these features is collected by sensors for 1 Acre land and this serves as the foundational input for our system. Our system leverages machine learning techniques to harness this data. It preprocesses the information and employs sophisticated algorithms to predict rainfall. Based on this predicted rainfall, our system effectively determine the amount of water to be supplied to crops by using xyz algorithms, ensuring that the right amount of water is delivered to optimize crop yields and mitigate water-related challenges. By exploring the nuanced dynamics of rainfall prediction, water supply management, and their vital role in agriculture, this paper aims to offer innovative solutions and insights that can positively impact the sustainability and productivity of our agricultural landscapes. We have only considered some common crops in this study (Table.3) . Relevant meteorological data, including temperature, humidity, evaporation, average wind speed, and total rainfall, is collected from sensors (see Fig.1,Fig 2,Fig 3) of Pune District Maharashtra. This data is stored in a dataset for further analysis. The system preprocesses the collected data, addressing missing values, outliers, and ensuring that all features are on a consistent scale. Any necessary encoding of categorical data is performed. Machine learning algorithms, including Random Forest Regressor, linear regression, and XGBoost, are employed for rainfall prediction. These models are trained on historical weather data .The system evaluates the accuracy of each machine learning model using suitable metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), depending on the nature of the rainfall prediction task. Based on this rainfall prediction, it calculates the water requirements for the selected crop type cultivated across 1 Acre Land. The system provides the user with recommendations regarding the amount of water required for proper irrigation and crop management, tailored to the specific crop type and the region's climatic conditions. By Using We found that XGBoost Algorithm have given the accuracy of 90.414 percent which is highest among all the three selected algorithm

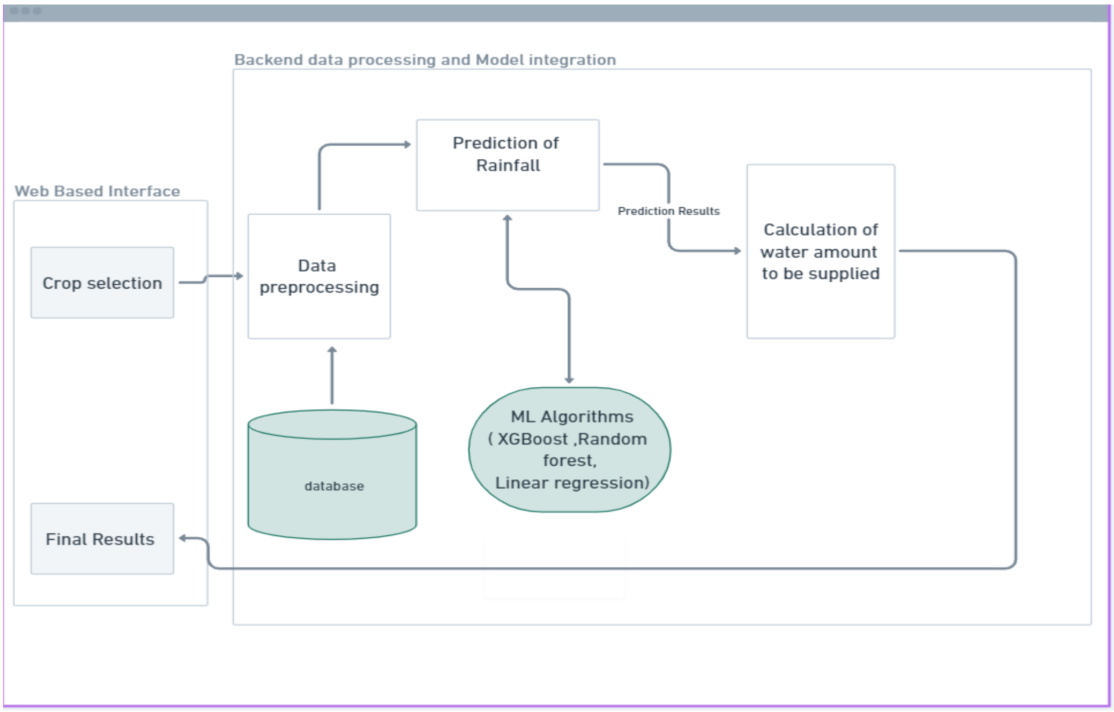


Fig.1 . System Architecture

2 LITERATURE REVIEW

(Atta-ur Rahman,2022) Using a machine learning fusion technique, this research suggests a unique real-time rainfall forecast system for smart cities. Four popular supervised machine learning techniques—decision trees, Naïve Bayes, K-nearest neighbors, and support vector machines—are used in the suggested framework. Fuzzy logic is a technique that is integrated into the framework to integrate the predicted accuracies of machine learning approaches, sometimes referred to as fusion, for effective rainfall prediction. Twelve years of historical weather data for the city of Lahore (from 2005 to 2017) are taken into account for the prediction. The dataset underwent pre-processing operations like cleaning and normalization prior to the classification procedure. 92.48% accuracy was attained.

(V.P.Tharun,,2018)With a case study of the Bijapur district in Karnataka State, India, the paper details the development of a system to offer information on rainfall characteristics and its prediction from historical data sets, as well as the selection of crops based on the forecast on a taluka basis. Rainfall is analyzed on a month-by-month, Nakshatra-by-month, week-by-week, and year-by-year basis. Agrometeorologists find great value in the developed method, which was created through the study of rainfall data gathered at the Agricultural Research Station in Bijapur during a 20-year period. The prediction algorithm is performing admirably, with an accuracy rate that ranges from 80% to 90%.

(Shaik Arshad Hussain,2022) The study assesses the precision and accuracy of rainfall prediction using Logistic Regression (LR) and innovative tree-specific XGBoost (XGB) machine learning methods. An innovative Tree Specific XGBoost classifier is used on a dataset of 145461 records from weather . A system that compares XGBoost and Losgistic Regression classifiers for rainfall prediction machine learning algorithms has been developed. Ten people in each group made up the sample size. Clinical analysis was used to determine the sample size, with alpha and beta values of 0.05 and 0.5, 95% confidence, 80% pretest power, and enrollment ratio. For both accuracy and precision, a significant value (p) of 0.019, or less than 0.05, was found. The classifiers' precision and accuracy were assessed and noted. The accuracy of the Logistic Regression classifier is 79.37%. The same is predicted by the XGboost classifier with an accuracy rate of 94.89% (Shaik Arshad Hussain,2022).

(Mohd Imran Khan,2020) Meteorological variable simulation using GCM is generally far superior to rainfall estimate, and meteorological variable observation data are few, if not nonexistent, in many places. By taking advantage of GCM's ability to simulate meteorological variables, the suggested strategy helps demonstrate how well the DL technique may improve the quality of rainfall forecast.As a result, about 85% of the data become the training set and the remaining 15% become the testing set.To calculate the difference between observed and anticipated values, RMSE is widely employed. It is always positive, and higher model performance is shown by a lower RMSE. During the deep MLP model's training phase, the average RMSE values vary from 5.89 to 19.92, average r values range from 0.4 to 0.72.

(Wanie M.Ridwan,2021) The reservoir may flood or become dry due to the unpredictable rains brought on by climate change. Several models and techniques were used in this study to forecast the rainfall data in Terengganu's Tasik Kenyir. The comparative study was carried out with an emphasis on creating and contrasting multiple Machine Learning (ML) models, assessing various scenarios and time horizons, and predicting rainfall with two alternative approaches. The average rainfall from ten stations around the study region was used as the basis for this research, and the station area and anticipated rainfall were weighted using the Thiessen polygon.The findings indicate a coefficient ranging from 0.5 to 0.9, with the daily (0.9739693), weekly (0.989461), 10-days (0.9894429), and monthly (0.99998085) scenarios showing the greatest values .

(Dr. Kesavulu Poola,2021)[Based on an analysis of the last 30 years of data (1987-2017) in the Vishakapattanam region, the XG-Boost Model is the best match to forecast the rainfall up to 3 to 5 years with 95% accuracy. According to the current article, the XG-Boost Model offers a reliable and satisfactory rainfall prediction for Vishakapattanam on a monthly basis. The monthly rainfall of Vishakapattanam is decreasing between 1987 and 2017.The XG-Boost algorithm's performance was assessed using data spanning from 1987 to 2017 by means of a graphical comparison between the observed and predicted data. The rainfall data that was seen and anticipated produced positive results for the XG-Boost algorithm.

(M T Anwar1,2021) We must create a precise rainfall prediction model in response to this circumstance in order to implement prescriptive measures. Previous studies on rainfall prediction have produced subpar results because they rely on models with inherent constraints. The goal of this work is to develop a multivariate rainfall forecast model with Extreme Gradient Boosting, which has proven to be the most effective method to date. The weather station's seven years of historical weather data served as the foundation for this model's construction. The outcome showed that, with a training RMSE of 2.7 mm and a testing MAE of 8.8 mm, the model is capable of generating reliable forecasts for daily rainfall.

Machine Learning Algorithms

2.1 XGBoost

XGBoost, is a high-performance machine learning algorithm renowned for its effectiveness in both classification and regression tasks. It excels by utilizing ensemble learning, combining the predictions of multiple decision tree models in a boosting framework. XGBoost stands out for its ability to handle large datasets, maintain predictive accuracy, and efficiently manage missing data. It incorporates regularization techniques, such as L1 and L2 regularization, to prevent overfitting and enhance model robustness. With its speed and versatility, XGBoost is a popular choice in data science and machine learning, making it an essential tool for predictive modeling and competitive machine learning challenges [5][8][9].

2.2 Random Forest

Random Forest is a versatile ensemble learning technique widely applied in machine learning for tasks such as classification and regression. It consists of multiple decision trees, each created on bootstrapped subsets of the training data with randomly selected features for splitting. This diversity mitigates overfitting. During prediction, the Random Forest combines the results from these trees, yielding robust and accurate predictions, particularly when dealing with complex or noisy datasets. It excels in handling missing data, provides insights into feature importance, and finds practical use in domains requiring both accuracy and interpretability[11][12].

2.3 Linear Regression

Linear regression is a fundamental statistical and machine learning technique used for modelling the relationship between a dependent variable and one or more independent variables. It assumes a linear, proportional relationship, where changes in the independent variables result in a linear change in the dependent variable. The goal of linear regression is to find the best-fitting line (or hyperplane in multiple dimensions) that minimizes the sum of squared differences between the observed data points and the predicted values. This linear model is characterized by coefficients representing the slope and intercept of the line, allowing for prediction and inference. Linear regression is widely employed for tasks like prediction, trend analysis, and understanding the impact of one variable on another, making it a foundational tool in data analysis and modelling [13][15][16].

3 METHODOLOGY

Our system, (see Fig.1), features a web-based user interface that empowers users to select a specific crop. This selection is pivotal as different crops exhibit varying water requirements and responses to environmental conditions. The system compiles real-time data from Database ,this database consists of the data of the weather climatic conditions of past one year of Pune District which is collected by IOT sensors[1]. Database consists of atmospheric parameters such as temperature, humidity, average wind speed, wind direction, evaporation rates, and total rainfall (see Table.1). To facilitate analysis, data pre-processing is done in which null values are replaced by average value of that specific column, categorical variable wind Direction is preprocessed using Label-Encoder. Further, Standardization of dataset is done using Standard Scaler Library. Subsequently, The dataset is split into 80 percent for training and 20 percent for testing, machine learning model, is trained on training data . On the other hand The system extracts the selected crop-specific water requirements from the database itself (see Table 3) . It's important to clarify that, in this study, we have focused solely on the general water requirements for specific crops,(see Table 2,Table 3). After training , Current weather conditions like temperature, humidity, average wind speed, wind direction, evaporation rates is given input to model and Rainfall for that specific day is predicted .With this information at hand, the system computes the necessary water quantity in liters that need to supplied to 1 Acre land of selected crop by using proposed algorithm . The result is Displayed to the screen. We have Used Three major algorithms in this research Linear Regression ,XGBoost, Random Forest Regressor and afterwards we found the XGBoost have given the highest accuracy of 90.414 percent among all three (see Table 4 ,Fig 3,Fig 4) .

PROPOSED ALGORITHM

Abbreviations :
X : Rainfall (mm) predicted by model
R : Water Required by Crop (mm)
S : Water To Be Supplied (litres)
Eto : Reference evapotranspiration (mm/day)
P : Predicted Rainfall Over 1 Acre Land (litres)

Input : X
Output : S

Step 1: Calculate the Amount of Predicted Rainfall in Liters in One acre Land
P = X*0.001*4046.86*1000
Step 2 : calculate the kc of selected crop The kc depend on crop stage
Step 3: calculate the Eto of grass using penman Monteith algorithm
Step 4: calculate the Eto of crop = kc * Eto of grass
Step 5 : R = max(0, ETo - A)
Step 6 : S = R- S , if S > 0 then , supply S litres of water

Penman Monteith Algorithm :

Input : Air_temperature ,Wind_speed ,Vapor_pressure ,Relative_humidity ,Net_radiation,Soil_heat_flux ,Albedo

Output : ETo_grass

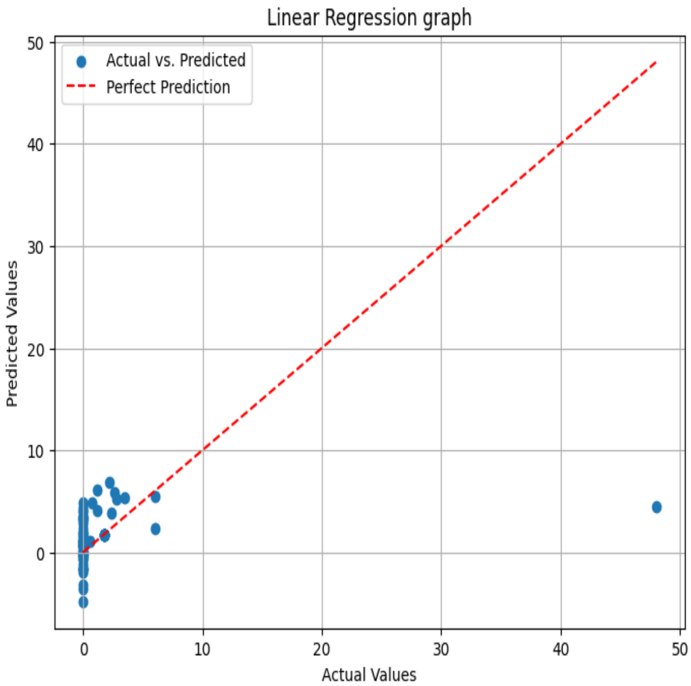
```
Algo_Eto_grass (Air_temperature, Wind_speed, Vapor_pressure, Net_radiation,
Soil_heat_flux, Albedo)
{
    specific_heat = 0.001013 # MJ/kg/°C
    psychometric_constant = 0.067 # kPa/°C
    latent_heat_vaporization = 2.45 # MJ/kg
    Stefan_Boltzmann_constant = 0.000000004903 # MJ/m²/day/K⁴


    delta = (4098 * (0.6108 * np.exp((17.27 * Air_temperature) / (Air_temperature +
237.3)))) / np.power((Air_temperature + 237.3), 2)
```

```
    gamma = psychometric_constant

    # Calculate reference evapotranspiration (ETo) using Penman-Monteith equation
    ETo = (0.408 * delta * Net_radiation + gamma * (900 / (Air_temperature + 273)) * Wind_speed * (Vapor_pressure / (Air_temperature + 273))) / (delta + gamma * (1 + 0.34 * Wind_speed))

    return Eto

}
```

ETo_grass = **Algo_Eto_grass** (Air_temperature, Wind_speed, Vapor_pressure-Net_radiation, Soil_heat_flux, Albedo)

8

4 RESULT



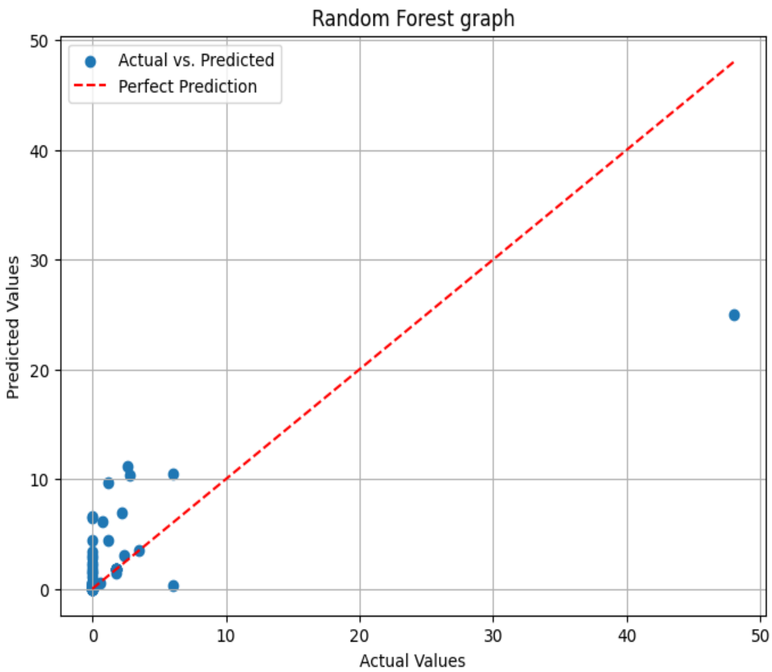
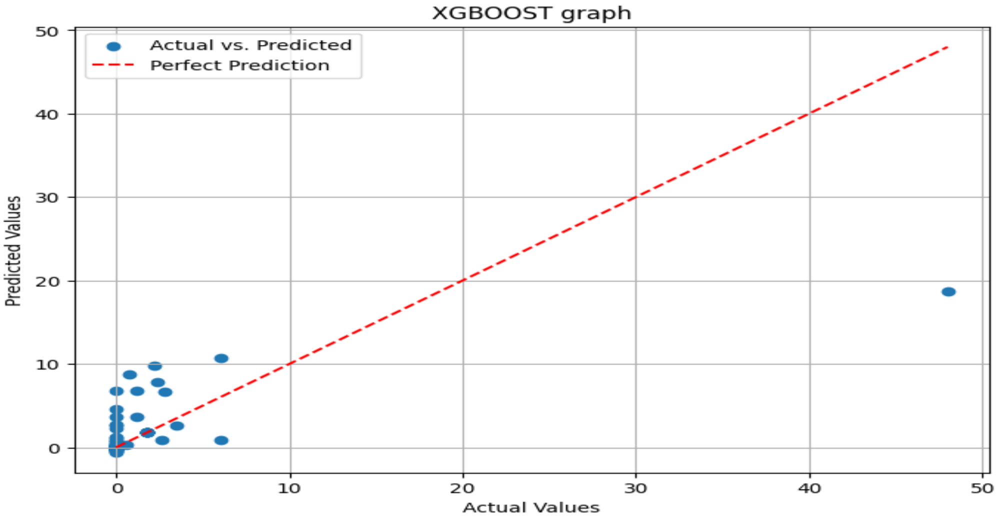


Fig 1 . Linear Regression

Fig 2 .Random Forest

Fig 3 . XGBOOST



9

Features	Type	Scale
Temperature	Numerical	Degree Celsius
Maximum Humidity	Numerical	Percentage
Minimum Humidity	Numerical	Percentage
Evaporation	Numerical	Millimeter/day
Average Wind Speed	Numerical	Kilometer per Hour
Wind Direction	Numerical	Cardinal Direction
Rainfall	Numerical	Millimeters

Table 1: Features of Dataset used for Rainfall prediction.

Crop	Water requirement (mm)
Wheat	450 – 650
Maize	500 – 800
Tomato	400-600
Soybean	450 – 700
Onion	350 – 550

Table 2 General Crop Water Requirements

Crop	Soil Type	Water Requirements for Temperature (15-25 ℃) in mm/day	Water Requirements for Temperature (more than 25 ℃) in mm/day	Durations(days)	Depth (in cm)	Depth(feet)
onion	Black	3 to 4	5 to 6	70-90	10-30	0.6
wheat	Clay loam	6 to 7	8 to 9	60-150	0-20	0.3
tomato	red loam soil	6 to 7	8 to 9	90-110	25-50	0.9
maize	sandy loam	6 to 7	8 to 9	100-120	15-25	0.6
Soyabean	Sandy loam	6 to 7	8 to 9	90-140	15-25	0.6

Table 3. Selected Crop water Requirements based on soil , depth and temperature

Algorithms	MAE	MSE	RMSE
Random Forest	1.58	13.77	3.71
Linear Regression	2.23	30.86	5.55
XGBoost	1.49	16.72	4.08

Table 4 : Performance evaluation of the Algorithms

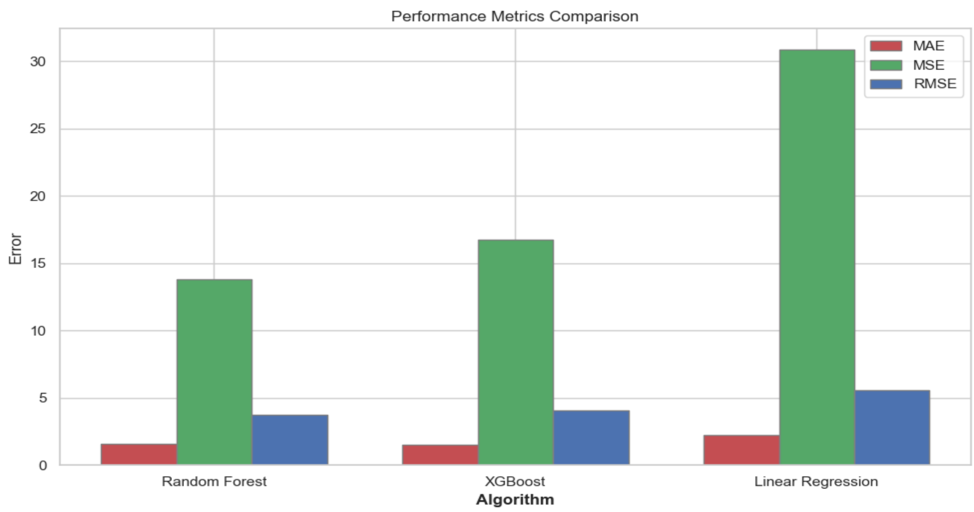


Fig 4 . Error Evaluation Graph of Algorithms

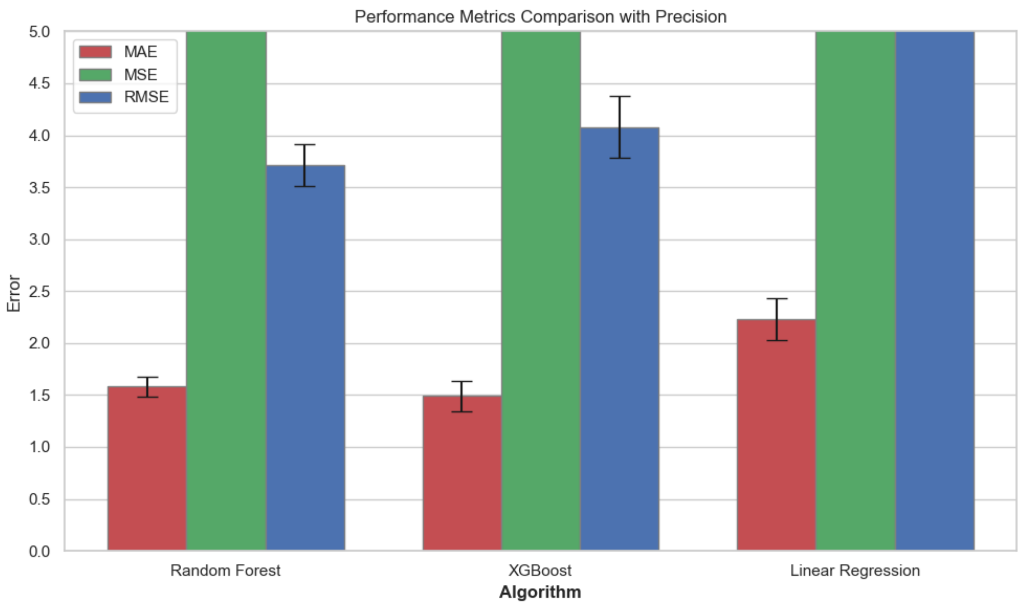


Fig 5. Precise Error Evaluation Graph of Algorithms

The performance evaluation of the algorithms (see Table 4) shown us that mean absolute error of XGBoost algorithm is 1.49 which is less as compared to other Algorithms. While Random Forest have given less Mean Squared Error and Root Mean Squared Error with value of 13.77 and 3.17 respectively.

For this problem statement we have selected **XGBoost** algorithm since it has low MAE, which directly reflects the low average absolute error.

5 CONCLUSION

The proposed design addresses crucial challenges in modern farming.
Firstly, it provides a means to accurately measure and determine the amount of precipitation that is truly beneficial for agricultural crops. This knowledge enables farmers to

manage their water resources more efficiently, ensuring that they use water effectively and avoid over-irrigation or under-irrigation.
Secondly, by understanding the precise water requirements of different types of crops, this technology allows for tailored irrigation practices. Farmers can provide crops with the right amount of water at the right time, promoting healthier growth and maximizing yield while conserving water resources.
Thirdly, the integration of rain sensors into agriculture is expected to revolutionize water supply management for crops in specific regions. By collecting data on local precipitation patterns, these sensors can predict water availability and assist in planning irrigation strategies accordingly. This advancement is likely to expand the areas that can be effectively monitored, reducing water wastage and contributing to sustainable and more productive agriculture. Overall, this innovation promises to bring about significant positive changes in the field of irrigation and crop management.

6 REFERENCES

[1] Atta-ur Rahman, S. A. (2022). Rainfall Prediction System Using Machine Learning Fusion for Smart Cities. MDPI Sensors. *MDPI Sensors*.

[2] Chkeir Sandy, A. A. (2023). Nowcasting extreme rain and extreme wind speed with machine learning techniques applied to different input datasets. *Atmospheric Research, Volume 282, article id. 106548*.

[3] Dr. Dayanand. G. Savakar (2016). Rainfall Prediction based on Rainfall Statistical Data. *International Journal on Recent and Innovation Trends in Computing and*, 270-275.

[4] Fahim Jawad, T. U. (2016). Analysis of Optimum Crop Cultivation Using FuzzySystem. 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS).
.

[5] Shaik Arshad Hussain1 , Terrance Frederick Fernandez . Early prediction of rainfall using xgboost algorithm and logistic Regression 2022.

[6] Mohd Imran Khan, R. M. (2020). Hybrid Deep Learning Approach for Multi-Step-Ahead Daily Rainfall Prediction Using GCM Simulations. IEEE , 52774 - 52784.
.

[7] Wanie M.Ridwan, M. S. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. Ain Shams Engineering Journal, 1651-1663

[8] Dr. Kesavulu Poola.Prediction of rainfall by using extreme gradient boost (XG boost) in Vishakapattanam area, Andhra Pradesh.

[9] M T Anwar . Rainfall prediction using Extreme Gradient Boosting Annual Conference on Science and Technology 2020 Journal of Physics.

[10] V. P. Tharun, R. P. (2018). Prediction of Rainfall Using Data Mining Techniques. *Second International Conference on Inventive Communication and Computational Technologies* (pp. 1507-1512). Coimbatore: IEEE.

[11] Prediction using Regression Model J.Refonaa, M. Lakshmi, Raza Abbas, Mohammad Raziullha International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S3, July 2019.

[12] Rainfall Prediction using Regression Model J.Refonaa, M. Lakshmi, Raza Abbas, Mohammad Raziullha International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S3, July 2019.

[13] Ouma, Yashon "Dam Water Level Prediction Using Vector AutoRegression, Random Forest Regression and MLP-ANN Models Based on Land-Use and Climate Factors." *Sustainability* (2022).

[14] Tolani, Manoj "Analysis & Estimation of Soil for Crop Prediction using Decision Tree and Random Forest Regression Methods." *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)* (2022): 752-755.

[15] Roshani "Analyzing trend and forecast of rainfall and temperature in Valmiki Tiger Reserve, India, using non-parametric test and random forest machine learning algorithm." *Acta Geophysica* 71 (2022): 531 - 552.

[16] Lotfirad, Morteza et al. "Drought monitoring and prediction using SPI, SPEI, and random forest model in various climates of Iran." *Journal of Water and Climate Change* (2021).

.