*Grenze International Journal of Engineering and Technology, June Issue*

# Weather-Driven Rainfall Estimation using Machine Learning

Dr S.Sai Kumar[1], Divya sri Malla[2], Lavanya Gampa[3], Varun Kodi[4], and Harsha vardhan Lagadapati[5]
[1-5]*Information Technology Department P.V.P Siddhartha Institute of Technology* Vijayawada, India
Email:saikumar@pvpsiddhartha.ac.in,19501a1258@pvpsiddhartha.ac.in,19501a1222@pvpsiddhartha.ac.in,
19501a1248@pvpsiddhartha.ac.in

***Abstract*— Predicting the amount of daily rainfall can be extremely useful for farmers and agricultural workers. This can improve agricultural productivity and ultimately help to secure food and water supplies, which is important for maintaining healthy communities. One popular approach is to use machine learning algorithms, such as artificial neural networks, to analyze historical data and identify patterns and relationships between different variables, such as temperature, humidity, and atmosphericpressure. By training these models on large amounts of data, researchers can develop highly accurate models that can predict rainfall with a high degree of accuracy.Changes in atmospheric circulation can be caused by a variety of factors, including natural variations in climate, human activities such as greenhouse gas emissions, and other environmental factors.For example, changes in the jet stream, which is a fast-moving, high-altitude air current, can alter the path and intensityof storms, leading to more or less rainfall in certain regions.One of the advantages of using machine learning algorithms for rainfall prediction is that they can handle large amounts of data and complex relationships between different variables, which can be difficult for humans to analyze manually. Additionally, these algorithms can be trained on historical data and continually updated with new data to improve their accuracy over time.**

***Index Terms*— Machine Learning, Dataset, Classification algorithms,RF,XGBoost**

## I. INTRODUCTION

Predicting rainfall is a complex task, and it can be challenging to forecast the amount and timing of precipitation accurately. One of such skilled and effective technologies is Machine Learning, Machine learning is a subset of artificial intelligence (AI) that involves training algorithms to learn patterns in data, and then using those patterns to make predictions or decisions about new data.Various Machine Learning algorithms such as Logistic Regression, Decision Tree, XGBoost, Random Forest are compared to find the most accurate model.it can be used to identify which variables are most strongly correlated with a target variable of interest. In this research a discussion and comparison of the existing classification techniques is made. To achieve this objective, a machine learning model can be trained using historical rainfall data and a set of relevant input parameters, such as temperature, humidity, wind speed, and atmospheric pressure. The parameters include date, location, min temperature, max temperature, humidity, wind direction, evaporation etc. These rainfall attributes are trained under fouralgorithms: Logistic regression, XGBoost, Decision Tree and Random Forest. Most efficient of these algorithms are RandomForest which give us the accuracy of approximately 88%.

*Grenze ID: 01.GIJET.9.2.334_1*

## II. Associated Work

Atmospheric moisture data can reveal patterns in humidity levels, and can help predict the likelihood of fog or dew formation. Precipitation data can be used to track rainfall and snowfall patterns, and can help identify trends in droughts or flooding. In our proposed model, the collected dataset is divided into sections that are useful to the machine learning model. After that, the dataset goes through data preprocessing, which involves passing the data through a process that replaces missing and error values in the dataset with mean values or the most frequently occurring value in that field. Our proposed model is taking the right approach by carefully preprocessing the dataset before feeding it into the machine learning algorithms. This will help to ensure that the model is able to learn effectively from the data and produce accurate predictions. Following the data preprocessing, the cleaned dataset is divided into two parts. During the training phase, the model learns from the input data and generates a hypothesis that maps the input data to the output data. After the model is trained, it is tested using the test set. The test set contains data that the model has never seen before, and it is used to evaluate the model's accuracy.

## II. Models

To compare the performance of the algorithms, you will need to evaluate their accuracy scores. This can be done by splitting your dataset into a training set and a testing set. You can then train each algorithm on the training set and evaluate their performance on the testing set. The accuracy score for each algorithm can be calculated by comparing the predicted rainfall amount with the actual rainfall amount in the testing set. Once you have the accuracy scores for each algorithm, you can compare them to determine which algorithm performed the best for predicting daily rainfall amount.

### *A. Logistic Regression*

It is a type of regression analysis that is particularly well-suited to situations where the dependent variable is binary, meaning it can take on only two possible values. In logistic regression, the dependent variable or outcome variable is binary, which means it takes on one of two possible values (e.g., 0 or 1). The independent variables or predictors can be either continuous or categorical. This logistic function is typically represented as an S-shaped curve known as a sigmoid function.

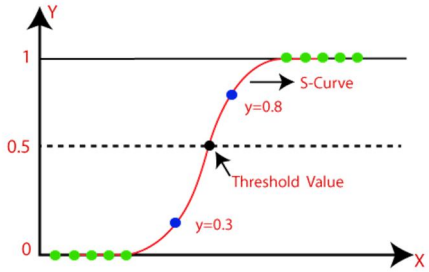


Fig: Logistic Regression curve

### *B. Random Forest*

Random forest is a machine learning algorithm that is used for both regression and classification tasks. It is a type of ensemble learning method that combines multiple decision trees to make predictions.Each decision tree in the forest is trained on a different subset of the data, and the prediction of all the trees are combined to make the final prediction

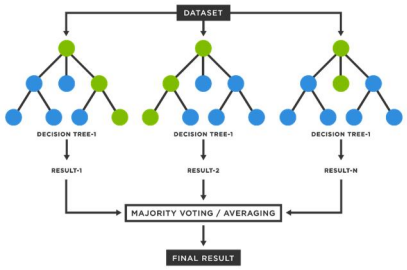


Fig :Random Forest algorithm

*C. Decision Tree Classifier*

A decision tree classifier is a type of machine learning algorithm that is used for classification tasks. It builds a model in the form of a tree structure, where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label.This algorithm helps in splitting the data into two or more related sets based on the most important indicators.

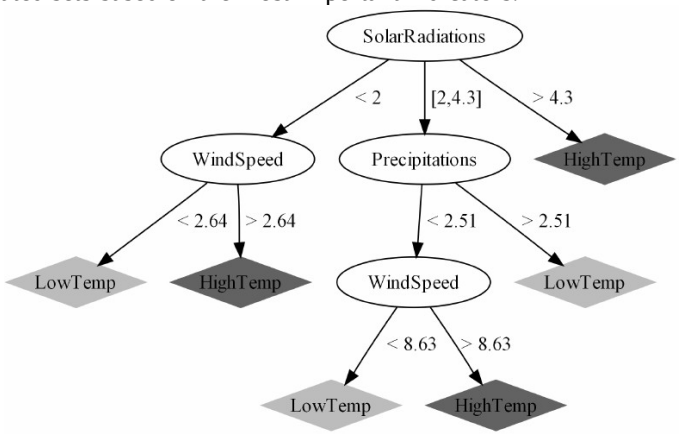


Fig: Decision Tree Classifier

III. XGBOOST GRADIENT DESCENT

XGBoost (Extreme Gradient Boosting) is a popular machine learning algorithm that is used for both regression and classification problems. Gradient boosting is an ensemble learning technique that combines multiple weak models (such as decision trees) to create a stronger overall model.

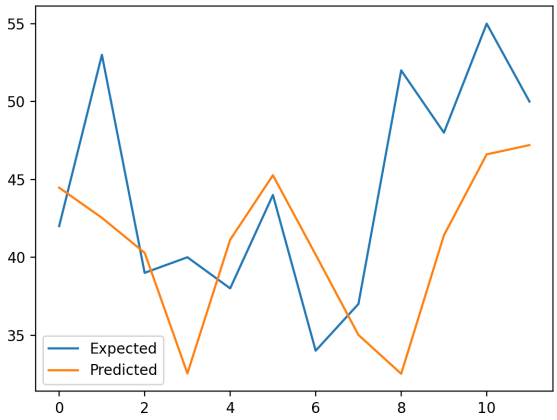


Fig: XGBoost for time series forecasting

*A. Evaluation*

Accuracy: accuracy is the ratio of the number of correct predictions to the total number of predictions made by the model.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Precision: It is the number of correct positive correct results divided by the number of positive results predicted by the classifiers.

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

Recall:It is the number of correct positive correct results divided by the number of positive results and number of negative results predicted by the classifiers.

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

## IV. INFERENCE

The focus of this research paper is to compare the performance of different machine learning algorithms for predicting rainfall and determine which algorithm is the most accurate.Accurate predictions of rainfall can have significant impacts on various industries, including agriculture, hydrology, and energy.Once the model is trained on the training set, it is tested on the testing set to evaluate its accuracy. The accuracy score represents the percentage of correct predictions made by the model on the testing set.

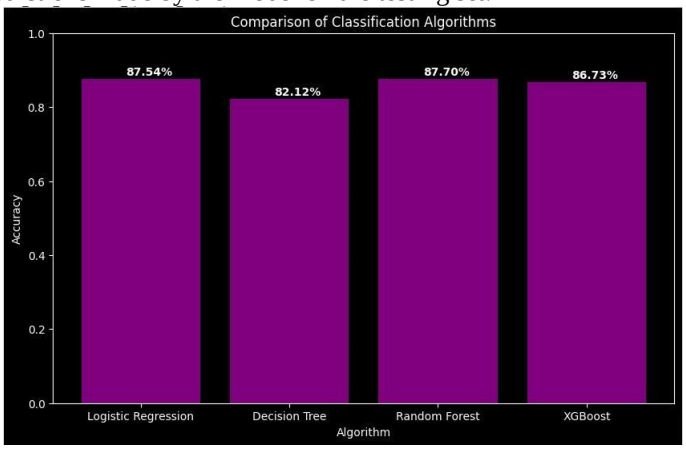


Fig: Comparision of classification algorithms

## REFERENCES


[1] Kumar Abhishek. Abhay Kumar, Rajeev Ranjan, Sarthak Kumar," A Rainfall Prediction Model using Artificial Neural Network", 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC2012), pp. 82-87, 2012

[2] G. Geetha and R. S. Selvaraj, “Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model,” Int. J. of Eng. Sci. and Technology,vol. 3, no. 1, pp. 211 213, 2011.

[3] Zahoor Jan, Muhammad Abrar, Shariq Bashir and Anwar M Mirza, "Seasonal to interannual climate prediction using data mining KNN technique",International Multi-Topic Conference, pp. 40-51, 2008.

[4] Elia Georgiana Petre, "A decision tree for weather prediction", Seria Matematica - Informatica] – Fizic, no. 1, pp. 77-82, 2009.

[5] Gupta D, Ghose U. A Comparative Study of Classification Algorithms for Forecasting Rainfall. IEEE. 2015./

[6] ajeevan, M., Pai, D. S., Anil Kumar, R. & Lal, B. New statistical models for long-range forecasting of southwest monsoon rainfall over India. Clim. Dyn. 28,813–828 (2007).

[7] Mishra, V., Smoliak, B. V., Lettenmaier, D. P. & Wallace, J. M. A prominent pattern of year-to-year variability in Indian Summer Monsoon Rainfall. Proc.Natl Acad. Sci. USA 109, 7213–7217 (2012).

[8] Thirumalai, C., Harsha, K. S., Deepak, M. L., & Krishna, K. C. (2017). Heuristic prediction of rainfall using machine learning techinque