Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 07 Issue 02
2024

*Research Article*
https://doi.org/10.56979/702/2024

# Evaluating Rainfall Prediction Models on Time Series Data from Bangladesh and Pakistan: A Comparative Approach

**Hira farman[1,2], Noman Hasany[2], Hamna Salman[3], Govarishankar[4], and Alisha Farman[1]**

[1]Department of Computer Science, IQRA University, Pakistan.
[2]Department of Software Engineering, Karachi Institute of Economics and Technology(KIET), Pakistan.
[3]Department of Computer Science, NED University of Engineering and Technology, Pakistan.
[4]Department of Computer Science, TIEST Constituent Institute of NED, Pakistan.
[*]Corresponding Author: Hira Farman. Email: hira.farman@iqra.edu.pk

**Received:** April 21, 2024 **Accepted:** August 12, 2024 **Published:** September 01, 2024

**Abstract:** Predicting rainfall is obviously a challenging task due to the multitude of factors and elements that impact climate conditions. Accurate rainfall forecasts are extremely important, especially for the agriculture industry, which depends heavily on timely and adequate rainfall for crop growth and result. The need for precise rainfall forecasts is further highlighted by the economic contribution that agriculture makes. Around the world, a variety of techniques have been employed to forecast rainfall patterns. This paper presents a comparative analysis of various rainfall prediction models, including statistical, machine learning on time series data and statistical approaches. We evaluate the performance of ARIMA, Random Forest, Linear Regression, Gradient Boosting and SVM models on diverse datasets from different geographical regions and climatic conditions. This study evaluates rainfall prediction using both machine learning and statistical model on distinct datasets from Bangladesh and Pakistan. The ARIMA (3, 2, and 1) model is applied to both datasets, demonstrating consistent performance on the Pakistan dataset with minimal differences between in-sample and out-of-sample results, indicating reliable forecasting ability. In contrast, the Bangladesh dataset shows a noticeable drop in performance from in-sample to out-of-sample, suggesting potential over fitting. Additionally, machine learning models such as Gradient Boosting (GB), Random Forest (RF), Support Vector Regressor (SVR), and Linear Regression (LR) are utilized. For the Bangladesh dataset, Gradient Boosting outperformed others with the lowest error values (MSE: 6435.975, RMSE: 80.225, MAE: 51.064) and the highest R² score (0.842). On the Pakistan dataset, Linear Regression produced the best results with the lowest MSE (270.138), RMSE (16.436), and MAE (11.572).Our findings highlight the strengths and limitations of each model, offering insights into their applicability for accurate rainfall forecasting.

**Keywords:** Rainfall; ARIMA; LR; GB; RF; SVR.

## 1. Introduction

Rainfall prediction is important for agricultural purposes, irrigation, and in preparation for disasters. There are several approaches that have been put forward given to improvement in computational techniques for modeling Rainfall. The objective of this study is to evaluate performance of at least three prediction models for different datasets. Rainfall prediction is a very significant factor in the management of the environment and the economy. Far reaching are the implications of precise rainfall predictions into different areas such as agriculture, disaster preparedness to mention but a few. Rainfall forecast accuracy [1] is therefore of paramount importance for countries such as India whose national revenue is mainly drawn from agriculture. The climate is volatile and hence, Statistical methods are not accurate in predicting rainfall. For the rainfall prediction tasks, the supervised machine learning methods, namely decision tree, Naïve Bayes, K-Nearest Neighbors, and Support Vector Machines are accurate [10]. It is important to predict the rainfall in the various ecological regions especially in countries such as Ghana where the climate is unreliable [3]. Since the climatic conditions differ over the different ecological regions, different machine learning classification techniques can be employed to forecast rainfall over these zones. Models using LSTM, Stacked LSTM and bidirectional LSTM Networks, XGBOOST and an ensemble of GBR, LSVR, and Ex-targeted trees were compared in the analysis of time-series data for the purpose of forecasting the volume of rainfall in one hour. . The major purpose of this study [5] is as follows: To investigate and assess the performance of the available machine learning techniques for modeling rainfall. . This study [11] used the meteorological and daily precipitation data of a forty-year period between 1980 and 2018 of the contiguous United States, and a model of three-dimensional convolutional neural network was also established. The objective of this research [16] is to design the rain prediction model using Machine Learning. This model extracted based on the dataset contains 2,391 records that data has collected from the website Bangladesh Jatiyo Tottho Batayon.

### 1.1. Importance of Rainfall Prediction

Specifically, the use of precise rainfall forecasts can help farmers [27] with the overall planning of planting, water supply, and other aspects so as to diminish certain crop production losses due to the unfavorable climate conditions. Thus, predicting droughts and floods is important in reducing risks, choosing crops, and using water-saving technologies among farmers. Precipitation forecasts also assists in the rational use of water by offering information on levels of reservoirs storage, groundwater, flood, and infrastructure. It leads to increased crop production, food security, and improved flood management practices which in turn increases community security and sustainable farming practices. Rainfall is an essential hydro climatic variable and is the only component of the hydrological cycle which has been potentially tapped and controlled to provide the desired quantity of rainfall for consumption. To establish the most valuable natural resource, which has a great impact on social and economic development, water should be awarded this honorable title. Fresh water supply is in most cases under pressure or there is a problem of balance between supply and demand in several parts of the world. This is because demand has increased, there is scarcity of water, or because water has been poorly utilized. Even so, the management of water and water planning is the only available strategy for lessening water stress or for closing the supply demand divide. Many efforts have been made across the globe to not only forecast the rainfall but also to identify that how the rainfall affects runoff in order to change the water availability so that water is available whenever required [18].

### 1.2. Traditional rainfall forecasting approaches

Previously, there has been a major focus on statistical methods and models for rainfall forecasting and less emphasis on other aspects such as the history of rainfall. Common approaches include. Empirical Models work off recorded rainfall data, and then extrapolates the development of patterns in this data to forecast future rates of rainfall. Techniques such as linear regression and time series analysis are often used. The NWP models have differential equations that provide computations of the atmospheric state and forecast weather conditions. Some prerequisites to the run include reasonable demands for computing resources and rather detailed preliminary weather conditions. Artificial Neural Networks (ANN). These types of networks are now called ANN's due to their rather marvelous characteristic of modelling non-linear phenomena. They work on large sets of data and are able to find intricate structures in the weather, which makes them ideal for use in rainfall prediction. Climatological Methods use historical rainfall data that has been accumulated over long periods of time to forecast rainfall. While they are easy to use, the results are sometimes inaccurate, because the averages do not consider short-term fluctuations These traditional methods though efficient to a certain degree are likely to exhibit one or the other of the following draws back; a large computational exercise, quite dependence on past records, and inability to incorporate with rapid and sudden change of weather. For this reason, with the aim of enhancing the precipitation forecasting accuracy and reliability, new approaches to machine learning, specifically the support vector machine (SVM) regression are under consideration[19,20,21,22].

### 1.3. Objectives

-To evaluate the performance of Statistical model ARIMA, and Machine learning model Random Forest, linear regression ,Gradient boosting and SVM models in predicting rainfall.

-In order to compare model accuracy, computational efficiency, and its stability with different datasets

-To provide recommendations according to comparative performance for the situation where you need to switch between forecasting scenarios.

The paper includes the following sections: study area, issues related to rainfall forecasting, a literature review And analysis of existing data on rainfall forecasts, a comparison of various literature sources that utilize rainfall data or weather forecasts, a proposed approach, dataset description and comparison, details of the algorithms Used, a comparative analysis of results, evaluation criteria, and a conclusion.

**2. Literature Review**

Random Forest models, which belong to the family of ensemble learning, are known to perform well when dealing with high dimensionality and can identify non-linear associations as well. However, it can be less accurate in terms of temporal changes compared to time-series models. In this study [1] There have is large emphasis on accurate forecasting of rainfall as nations such as India which relies heavily on agriculture will benefit significantly from an accurate rainfall forecast. Owing to the uniqueness and constantly changing nature of weather conditions Statistical models present low accuracy for rainfall prediction. For the reasons of nonlinearity of rainfall data Artificial Neural Network is a better technique. State of the art and comparison of the various methods and algorithms used by the other researchers. This research is based on the development of [2] a new real-time rainfall prediction model for smart cities through the usage of a specifically developed machine learning fusion. The proposed framework employs four of the most adopted supervised ML algorithms, namely decision tree, Naïve Bayes, K-Nearest Neighbors, and Support Vector Machine. To provide better rainfall prediction, the concept of fuzzy logic is also included in the scheme to combine the forecast efficiency of the machine learning algorithms, termed as the fusion. The work in [3] therefore complements the application of different classification models for rainfall prediction across the diverse ecological regions in Ghana. This study [4] compared rainfall estimates using reduced complexity rainfall estimation models derived from teaching-learning-based ordinary Machine Learning algorithms and Deep Learning architectures appropriate for the downstream applications explored in this paper. Evaluation of LSTM, Stacked LSTM, Bidirectional LSTM Networks, XG Boosting, GBR + LSVR + Extra-trees was performed for the task of hourly rainfall volume prediction using time-series data. ARIMA models are the most common methods of time-series forecasting since they are easy to understand and implement. This model is useful when the data in question exhibits a strong temporal structure and the relationships within it are linear. The aim of this literature [5] is to review and analyze various machine learning models for rainfall prediction. First, the study seeks to find out the noticeable specific characteristics of the meteorological data which affects the accuracy of forecasts and to evaluate the performance of the models such as ARIMA and ANN under varying climate conditions. In line with this, this study aims at comparing these methods with an arguably simpler method in order to identify the relative accuracy and timeliness of the algorithm in giving out early warnings, which will in turn help the farmers in making the best decisions possible in so far as protecting the crops and improving the gains of agriculture are concerned. This survey paper [6] where deep learning is used in combination with meteorological data to forecast the rainfall. The papers are analyzed based on the methods of deep learning applied, geographical location of the study area, type of metric, software used for building the model and lastly the year of publication of the papers. Under the social media category, this paper explores the feasibility of employing Twitter as a research instrument in research concerning disasters. Consequently, this study shall use the most modern approaches in deep learning, machine learning and predicting of disasters [7]. Therefore, one of the goals of the project is to define all the potential types of data and their sources with reference to different professions and crisis management scenarios. The SHS was predicted based on the immediate and future precipitation using ARIMA modeling for the specific study area which is Klang River Basin in Selangor [8]. The study effort in [9] deals with the uni variate ARIMA model to forecast monthly rainfall and is based on the Khordha district in the Odisha state in India. There were no present rainfall records; thus, the scale was reported monthly for the years 1901 through 2002. The parameters of each fitted model were evaluated, employing the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), from which emerged the ARIMA (1, 2, 1) (1, 0, 1)12. The goal of this study is to determine different architectures of the neural networks utilized in the rainfall forecasting, to determine voids in the structures expected to exist while deploying the neural networks for various types of scales and outline the index of performance used on the various architectures of the neural networks. The aim of this study [10] is to improve the rainfall prediction by extending the application of RNN by the integration of SVR that has been optimized using the CPSO algorithm. Therefore, the performance of this herein proposed RSVRCPSO model is evaluated for the typhoon rainfall data of Northern Taiwan. The aim is to explain in what a manner the application of this complex composite approach can enhance the effectiveness and reliability of the rainfall estimation and thus can be beneficial in terms of both fundamental research in hydrology and applied fields like meteorology. This study [11] employed meteorological and daily precipitation data of the forty years between 1980 and 2018 of the contiguous United States and standardized a model of three-dimensional convolutional neural network for short term precipitation. One more research [12] is concerned with creation and validation of the combined model based on CNN and ridge regression for the investigation of seasonal changes of precipitation in China. The reason is that they rely on the spatial pictures of general circulation as predictors. The research done in the study [13] focuses on the capability of DL models in providing the precipitation expected in the next twelve hours by incorporating advanced methods in the machine learning domain and accurate high temporal precipitation data. As a repository for the investigation [14] that considers the application of the everyday deep learning approaches, it should be noted that the study is concerned with improving the geographical resolution of the precipitation data obtained with the use of the weather radar, satellites data and/or weather stations data. The aim of the present research [18] is to establish the rain prediction model using Machine Learning. This model is developed based on a 2,391 records dataset that data has collected from the website named Bangladesh Jatiyo Tottho Batayon.

## 3. Analysis of Different Approaches of Rainfall Prediction

In analyzing different approaches to rainfall prediction, various studies offer insights into the effectiveness of different methods. Table 1 presents a comparative analysis of these approaches, highlighting key aspects such as the country and region of study, the dataset utilized, the algorithm applied, and the accuracy of predictions.

**Table 1.** A comparison of works of literature that make use of rainfall or forecasts of the weather

| Ref | Country | Region | Dataset Description | Algorithm | Accuracy | Best accuracy | Prediction Type |
|---|---|---|---|---|---|---|---|
| [15] | Taiwan | Yilan River basin and stations | collected from Liwu station 2012 to 2018 Hourly rainfall data of 6 gauges, | KNN, Support Vector Regression (SVR), ,Fuzzy Inference Model | 1hour RMSE:0.07 CE :0.99, 2 hours RMSE: 0.15 CE:0.97 3 hours RMSE:0.25 CE:0.93 | SVR and Fuzzy Inference Model | hourly forecasting 1,2,3 hour |
| [16] | Bangladesh | Southwestern coastal region of Bangladesh. | no of times a region flooded in year (BARC) Bangladesh Agriculture Research Council | MLP, KNN ,RF, Genetic algorithm-(SVR) (GA;RBF;SVR) | MLP :0.967 KNN: 0.956 RF: 0.984 GA;RBF;SVR: 0.983 Optimized :0.987 | The optimized model given the good accuracy | Flood |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [17] | Bangladesh | information gathered every day from the Gazipur, Rangpur, and Barisal district stations | In the subsequent section, the rainfall data from 10 decades of the period of 2011 to 2020 is brought and compared with the rainfall data from the entire period of 1980 to 2020. Namely, described above dataset consists of 1,319 rows and 16 columns. | Decision trees(DT), Random forests(RF) (SVM), or neural networks (NN). | 2011-2020<br>LR( 0.8676 )<br>SVC (0.8088)<br>KNN(0.8235)<br>DTC (0.8088) | The Binary Logistic Regression (BLR) approach achieves the highest accuracy, with maximum precision and recall of 0.61 and 0.6667, respectively, as well as an accuracy rate of 0.8676. | Rainfall |
| [23] | Pakistan | Pakistan Lahore | 12 years of historical weather data (2005 to 2017)<br>*******<br>25,919 instances and 11 feature | (i)Decision tree(DT), (ii)Naïve Bayes(NB), (iii) K-nearest neighbors, and (iv)Support vector machines | Proposed Work accuracy= 0.94<br>Miss rate=6<br>SVM=0.92,<br>KNN= 0.93,<br>DT=0.91, (CART miss rate=19.7)<br>NB= 0.90 | KNN=93% | (Classification Task) Rainfall Prediction |
| [24] | Pakistan | Pakistan | annual average rainfall based on 65 | SFTS For comparison with | the accuracy ME =0.2046, | SFTS model was more effective | (Regression) the next ten years' monthly |

| | | | | | |
|---|---|---|---|---|---|
| | years for the period of 1951 to 2015 have been extracted from World Bank website from January to December.)= | ARIMA and exponential smoothing state space (ETS) models | RMSE=14.911,MAE = 10.969, MPE=145.791%, MAPE = 172.26%. | when compared to the conventional The ARIMA and ETS ones and the forecasts which are provided by SFTS models are also more accurate and reliable. | forecasts (2016-2025) were also collected with its corresponding 80% prediction interval. |

## 4. Methodology

The methodology for rainfall prediction involves utilizing two distinct datasets from Bangladesh and Pakistan. The process begins with the collection of meteorological data, which includes various factors such as temperature, year, month and rainfall records. This data is then divided into two separate datasets: one for Bangladesh and one for Pakistan. This process begins with the collection of a comprehensive historical weather dataset, which spans from 1901 to 2016 from Pakistan and 1901 to 2019 from Bangladesh rainfall. To predict future rainfall, both machine learning and statistical approaches are employed. The machine learning models used include Random Forest, Support Vector Regressor (SVR), Linear Regression, and Gradient Boosting, while the statistical approach primarily involves the ARIMA model. Each model is applied to both datasets to forecast rainfall, and their performance is evaluated using several metrics, including RMSE, MAE, MSE, MAPE, and the correlation coefficient. Finally, the computational efficiency and robustness of these models are assessed to determine their suitability for predicting rainfall in each region. This comprehensive methodology allows for a thorough evaluation of different predictive approaches across diverse datasets as shown in figure 1.

### 4.1. Dataset Description

In this study two datasets are analyzed and compared shown in figure 2 and 3 one is related to Bangladesh weather and other related to Pakistan The first dataset Bangladesh dataset contain 4 attribute like year, month, Temperature and rainfall and 1429 observations. The second dataset Pakistan dataset contain 3 attribute like year, month, and rainfall and 1393 observations. Table 2 shown description.

### 4.2. Model Description

In this study for rainfall prediction utilized both machine learning model (Random Forest ,Linear Regression ,Gradient Boosting ,Support Vector Regression)and statistical model ARIMA.

#### *4.2.1. Random Forest*

Random Forest is a type of ensemble learning where several decision trees are trained to make the prediction shown in fig 4. It can be applied for time series forecasting by inclusion of lagged variables and other features based on time. It processes multiple attributes simultaneously and is efficient on big data sets. In many cases, with the help of ensemble it reduces over fitting and Random forest is best for datasets which contain non linearity, where model interpretability is required. For regression Final prediction shown in equation 1.

$$\dot{y}=\frac{1}{B}\sum_{b=1}^{B} T_b\ (x) \qquad (1)$$

Where,

-Dataset Training $\{(x_i\ ,y_i\ )\ \}_{i=1}^{n}$

-count of trees B.

-$T_b$ recursively split tree

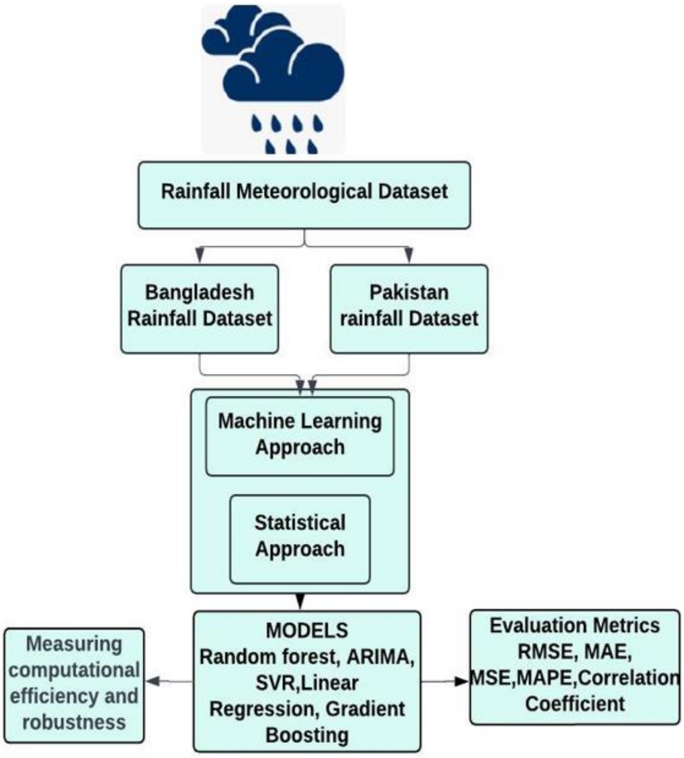
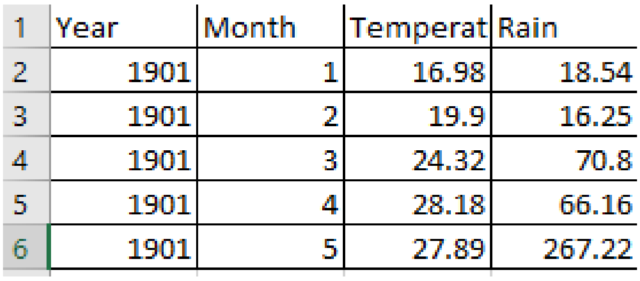
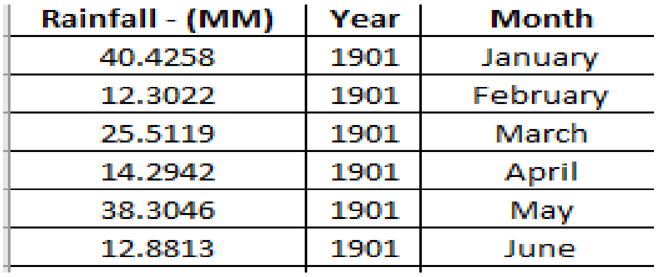

Rainfall Meteorological Dataset
Bangladesh Rainfall Dataset
Pakistan rainfall Dataset
Machine Learning Approach
Statistical Approach
MODELS Random forest, ARIMA, SVR,Linear Regression, Gradient Boosting
Measuring computational efficiency and robustness
Evaluation Metrics RMSE, MAE, MSE,MAPE,Correlation Coefficient

**Figure 1.** Proposed Methodology of work



| 1 | Year | Month | Temperat | Rain |
|---|---|---|---|---|
| 2 | 1901 | 1 | 16.98 | 18.54 |
| 3 | 1901 | 2 | 19.9 | 16.25 |
| 4 | 1901 | 3 | 24.32 | 70.8 |
| 5 | 1901 | 4 | 28.18 | 66.16 |
| 6 | 1901 | 5 | 27.89 | 267.22 |

**Figure 2.** Bangladesh Dataset Representation



| Rainfall - (MM) | Year | Month |
|---|---|---|
| 40.4258 | 1901 | January |
| 12.3022 | 1901 | February |
| 25.5119 | 1901 | March |
| 14.2942 | 1901 | April |
| 38.3046 | 1901 | May |
| 12.8813 | 1901 | June |

**Figure 3.** Pakistan Dataset Representation

**Table 2.** Datasets Description

| Dataset | Features | No of feature's | No of rows | unit | Description |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| **First Dataset Bangladesh Weather** | Year, Month, Temperature, Rain | 4 | 1429 | Year, Month, °C (Temperature), mm (Rainfall) | Captures monthly weather data, including temperature and rainfall. This dataset tracks the monthly temperature and rainfall in Bangladesh. |
| **Second Dataset Pakistan** | Rainfall (MM), Year, Month | 3 | 1393 | mm (Rainfall), Year, Month | Records monthly rainfall data for different months in Pakistan.<br>This dataset includes the monthly rainfall data for Pakistan, showing the total rainfall per month in a specific year. |

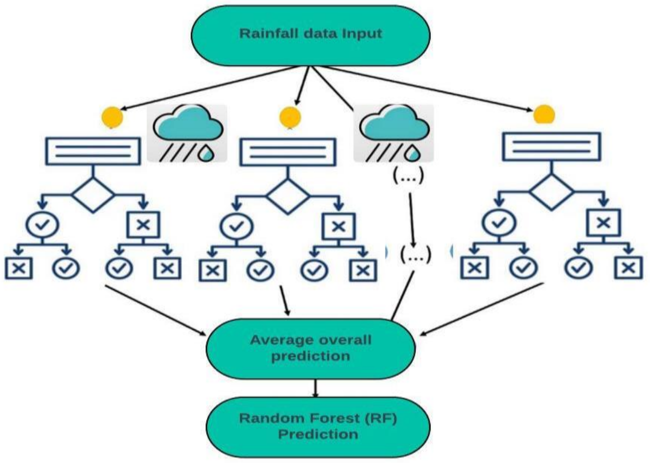


**Figure 4.** Representation of Random forest

### *4.2.2. Gradient Boosting Machines (GBM)*

Most machine learning algorithms, including gradient boosting machines, XGBoost, or Light GBM, can be used for a time series by including lag features, rolling means or variances and other temporal features. This algorithm is suitable for large database because it take less amount of time to search through a large database. When tuned it is Robust to over fitting. Indeed, it is successfully applied to Rainfall data with non-linear relationships and it has intricate feature interconnections. Gradient Boosting is an ensemble technique where decision trees are combined sequentially shown in figure 5. Each tree corrects the errors of its predecessors by focusing on residuals (the difference between predicted and actual values).The final model is the sum of all the weak learners shown in equation 2

$$F(x)= F_0(x) + \sum_{m=1}^{M} \partial_m \; h_m(x) \qquad (2)$$

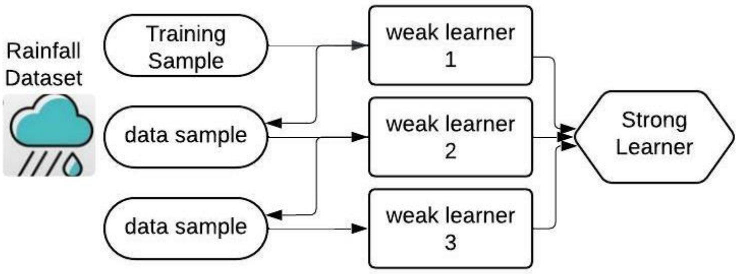


**Figure 5.** Representation of Gradient Boosting Algorithm

*4.2.3. Support Vector Regression (SVR)*

SVR is an analogy of Support Vector Machines which is used in regression problems. It is fine when used to analyze small to midsize data sets, and is able to incorporate interactions. It is, however, effective in high dimensional space and its create a hyper plane segregates data shown in equation 2 mathematically and figure 6 its representation. Covariance is good for non-linear data and it is quite resistant to outliers. This is well suited for Rainfall data that exhibit complicated and non-linear trends and where data is limited.The mathematical description of support vector machine is shown in eq (2).

$\{X_i, Y_i\}_{i=1}^{n}$ , Where $X_i \in \mathrm{R}^d$, $Y_i \in \{-1, +1\}$

Hyper plane:

$$w \cdot x + b = 0 \tag{3}$$

-Some Constraints:

$Y_i \ (w \cdot X_i + b) \geq 1, \qquad \forall_i \in \{1, 2, \ldots, n\}$

B ∈ R is the bias term and w ∈ $\mathrm{R}^d$ is the vector form.This constraint ensures that:

If $Y_i$=+1 then w·$X_i$b≥1, If $Y_i$ =−1 then w. $X_i$+b≤−1.

In other words, all the positive examples are found in one side of the hyper plane (or on the boundary surface) and all the negative examples lie in the other side of the hyper plane (or on the boundary surface).

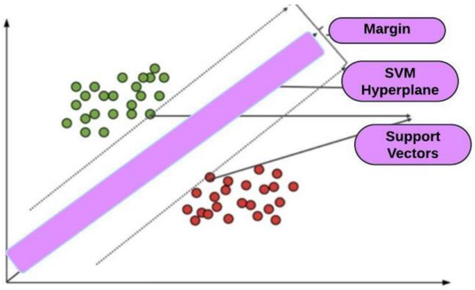


**Figure 6.** Representation of Support Vector Regressor

*4.2.4. Linear Regression for Time Series Rainfall Data*

Linear Regression is a basic type of supervised learning, which helps to understand model. It refers to the procedure of predicting the extent of the connection between an event known as the dependent variable and one or more factors known as independent variables. As much as this Linear Regression can be used in the context of time series forecasting under the following considerations. Another type of regression that could also be useful in the given time series data set is the Linear Regression and the technique used generally for this type of data set is working with the feature lag such as by using the past dates of the dependent variable we use rainfall. Coefficients in the model are quite useful in establishing the degree of relatedness between the given predictors and the target variable. Regression is friendly in terms of computational complexity and hence can be useful for large data sets or where immediate results are required shown in figure 7 . This model is widely employed in time series forecasting as a benchmark. Indeed, if Linear Regression turns out to perform well, it sends a strong signal that more complex model may not be needed. When the relationship between past and future rainfall is not complex and very close to being a straight line, Linear Regression can be useful for short-term prediction. The single and multiple attributes equation shown 3 and 4.

For a single independent variable x, the model is:

$$y=\beta_0+\beta_1 x+\epsilon \tag{4}$$

When there are multiple independent variables $x_1, x_2,\ldots, x_n$ , the model extends to:

$$y=\beta_0+\beta_1 x_1+\beta_2\ x_2+\cdots+\beta_n Xn+\epsilon \tag{5}$$

Where:

- y is the control variable or the dependent variable which has the predicted value.
- identified as the independent variable: x.
- $\beta_0$ is the y-intercept, or the point at which the regression line intersects the y-axis.
- $\beta_1$ is the coefficient, which is the slope of the line that measures the change in y for a one-unit change in x.
- $\epsilon$ is the error which is the difference between actual observation and the predicted values.

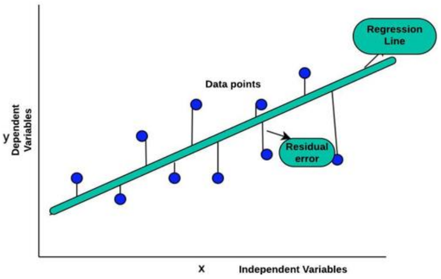


**Figure 7.** Representation of Linear Regression

*4.2.5. Time Series Forecasting*

When a variable or data has only one dimension or when the data comprises of only one variable that changes with time then the ARIMA model [25, 26,34] is used. From the above described times series data, the often used type is the ARIMA type, which is an abbreviation for Auto Regressive Integrated Moving Average. ARIMA has certain specifiable parameters which are mostly depicted by letters p, d and q shown in table 3.

**Table 3.** Representation of Arima Model Components parameter

| **ARIMA Univariate Model (Focus on a single time series (Rainfall).** | | | | | |
|---|---|---|---|---|---|
| **Components** | **Usage** | **Parameters** | **Description** | **stationarity** | **Model fit pattern** |
| AR (Autoregressive)<br>I (Integrated)<br>MA (Moving Average) | When forecasting a single time series data collection, univariate ARIMA is utilised; this technique works best with time series data that is self-correlated, meaning that one occurrence influences subsequent events. | P for Time series autoregressive<br>d for time series Integrated<br>q fir time series Moving Average | -Helps capture the persistence of the time series data(AR)<br>-Differencing helps stabilize the mean(I)<br>-Models past forecast errors as part of future predictions(MA) | Ensure stationarity and determine p, d, q | -Build the model and predict future rainfall values<br>-Identify appropriate differencing order<br>-Estimate the impact of past errors on current values |

A. Parameter Description

It is denoted as ARIMA(p,d,q), where p, d and q are parameters of an ARIMA model as follows: Typically, an ARIMA model parameter is written as ARIMA (p, d, q), where:

-p: Where p is the order of the autoregressive component of the model. This variable is the number of lag observations that is included in the model that is used in summary form here. It is sometimes called the lag order This involves the determination of the number of observations that can either precede or follow the independent variable whereby in most cases it is denoted as p.

-d: The frequency at which the time series is made more stable by breaking down the raw observations is described by the symbol 'd', and the degree of this difference is also given by the same alphabet 'd'.

-q: q stands for the order of the moving average component It is also known as the averaging order sometimes whereby the- window size of the moving average is in question.

B. Components of ARIMA

-Autoregressive (AR): The dependency between an observation and several lag observations (prior values) is used in this component. An AR model of order $p$ (AR(p)) can be expressed mathematically shown in equation 5

$$t=\alpha+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p\ y_{t-p}\ +\epsilon \quad (6)$$

Integrated component (I): This is to make the time series stationary, this aspect entails differencing the data points (i.e., removing trends and seasonality). Mathematically, the differentiation procedure is expressed as in equation 6

$$Y`_t=Y_t-Y_{t-1} \quad (7)$$

$Y`_t$ is the differenced series in this case d. For a distinction structure of d time repeated

C. Component of Moving Average (MA)

In this section, we use the relationship that holds between an observation and a residual error in a moving average model when used with respect to lagged observations and a particular observation. In mathematics, equation 7 represents one of the methods by which an autoregressive moving average model of the order q (MA(q)) can be expressed.

$$Y_t=\alpha+\epsilon t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\cdots+\theta_q\epsilon_{t-q} \quad (8)$$

Where:

- $Y_t$ is the dependent value at time t,
- $\alpha$ is a constant value,
- $\theta_1$, $\theta_2$,…, $\theta_q$ are the parameters of the model,
- $\epsilon_t$, $\epsilon_{t-1}$,,…, $\epsilon_{t-q}$,are the error terms (white noise)

4.3. Evaluation Metrics

The following evaluation performance metrics are used in this study for checking the algorithm performance.

*4.3.1. Mean* Absolute *Error (MAE)*

This is commonly used as the measure of model performance by taking an average of absolute difference between the predicted or modeled value and the actual observation value shown in equation 8. MAE comes up with an easy to understand figure of the average error magnitude. The minimum score is desirable for better model performance shown in equation 8.

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|yi-\hat{y}i| \quad (9)$$

*4.3.2. Root Mean Squared Error (RMSE)*

The square root of the mean of the square of the residuals – predicted values minus the actual values. RMSE is even more sensitive to outliers than MAE, because it gives penalties proportional to the squared values of the difference. It is used mostly for the evaluation of the performance of the models shown in equation 9.

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi-\hat{y}i)^2} \quad (10)$$

*4.3.3. Mean Absolute Percentage Error (MAPE)*

It is the average computed from the absolute percent differences between the predicted and actual values shown in equation 10. MAPE expresses error in terms of percentage and therefore is more relativity in size of error than the other measure of error. But if the actual values are small numbers, then it may be misleading to use relative changes measurements.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{yi-\hat{y}i}{yi}\right| \times 100\% \tag{11}$$

*4.3.4. Correlation Coefficient (R )*

An estimate indicating the co-linearity of two variables and especially the direction and intensity of any linear association between the predicted and actual values. l is a measure of the strength and direction of the relationship between two variables; a value of 1 shows a perfect positive linear relationship; -1 shows a perfect negative linear relationship; and 0 shows no linear relationships shown in equation 11.

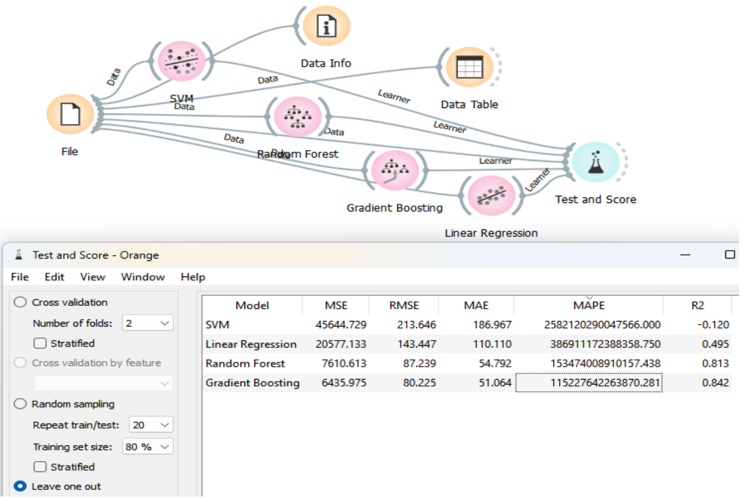
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi-\hat{y}i)^2}{\sum_{i=1}^{n}(yi-\bar{y})^2} \tag{12}$$

## 5.Results

The results are calculated by performing implementation on different dataset thr bangladesh and pakistan

### 5.1. Performance on Bangladesh Weather Dataset

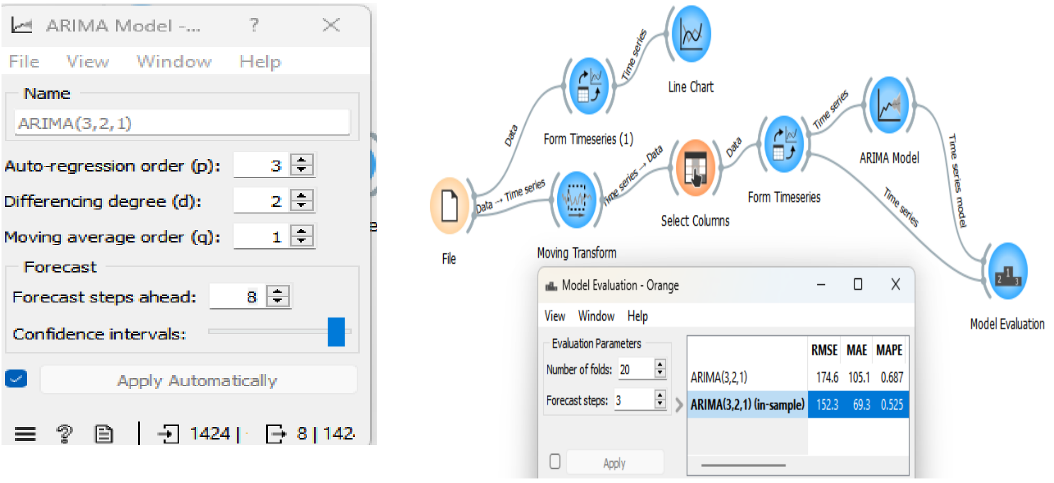
The figure 8 showcases a machine learning analysis workflow using the Orange data mining tool, focusing on predicting rainfall (in milliliters) based on a meteorological dataset from Bangladesh stored in CSV format. Four models support vector machine for regression SVM, Linear Regression, Random Forest, and Gradient Boosting—are trained and evaluated. The evaluation process uses cross-validation with 2 folds and stratified sampling, alongside a random sampling method that splits the dataset into 80% training and 20% testing data. Key metrics like MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and R² (coefficient of determination) are used to measure performance. The Gradient Boosting model outperforms the others with the lowest error values (MSE: 6435.975, RMSE: 80.225, MAE: 51.064) and the highest R² score (0.842), indicating its superior ability to capture the complex relationships in the rainfall data. The Random Forest model also shows strong performance but slightly lags behind Gradient Boosting. In contrast, the SVM model performs poorly with high error metrics and an R² value of -0.120, suggesting that it struggles to capture the underlying patterns in this dataset. The dataset likely includes features such as historical rainfall in mm, temperature, year, month, and other meteorological factors that influence rainfall prediction. The Gradient Boosting model's effectiveness in handling complex, non-linear interactions and feature importance makes it the best choice in this scenario, especially given the intricate and dynamic nature of weather-related data. This is a specific type of cross-validation utilized in experiment almost all of the data as the training set and leave out only one data point for testing. This process is repeated such that each data point in the dataset gets to be the test set exactly once.



○ Cross validation
Number of folds: 2
□ Stratified
○ Cross validation by feature
○ Random sampling
Repeat train/test: 20
Training set size: 80 %
□ Stratified
● Leave one out

| Model | MSE | RMSE | MAE | MAPE | R2 |
|---|---|---|---|---|---|
| SVM | 45644.729 | 213.646 | 186.967 | 2582120290047566.000 | -0.120 |
| Linear Regression | 20577.133 | 143.447 | 110.110 | 386911172388358.750 | 0.495 |
| Random Forest | 7610.613 | 87.239 | 54.792 | 153474008910157.438 | 0.813 |
| Gradient Boosting | 6435.975 | 80.225 | 51.064 | 115227642263870.281 | 0.842 |

**Figure 8.** Result on Bangladesh dataset using Machine Leaning model



Name: ARIMA(3,2,1)

Auto-regression order (p): 3

Differencing degree (d): 2

Moving average order (q): 1

Forecast

Forecast steps ahead: 8

Confidence intervals:

☑ Apply Automatically

1424 | 8 | 142

Evaluation Parameters

Number of folds: 20

Forecast steps: 3

|  | RMSE | MAE | MAPE |
|---|---|---|---|
| ARIMA(3,2,1) | 174.6 | 105.1 | 0.687 |
| ARIMA(3,2,1) (in-sample) | 152.3 | 69.3 | 0.525 |

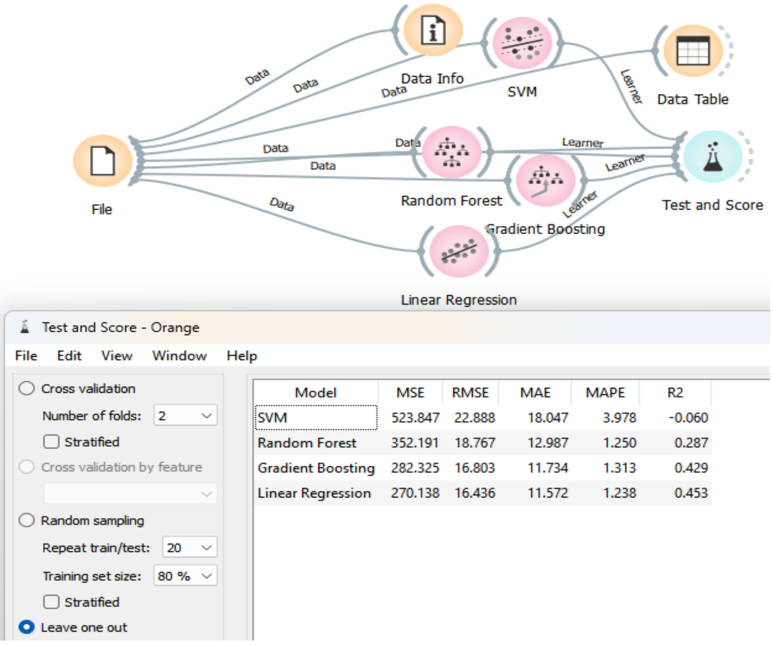
**Figure 9.** Result on Bangladesh dataset using ARIMA Statistical model

The figure 9 on Bangladesh dataset it shows the application of an ARIMA (3,2,1) model on a dataset from Bangladesh to forecast time series data. The ARIMA model parameters are defined as follows: 3 autoregressive terms (p=3), 2 differencing operations (d=2), and 1 moving average term (q=1). The model evaluation results compare the performance in both in-sample and out-of-sample (forecast) contexts. The in-sample results, which assess how well the model fits the training data, show better performance with lower error metrics: RMSE (152.3), MAE (69.3), and MAPE (0.525). On the other hand, the out-of-sample results, which measure the model's predictive power on unseen data, yield slightly higher errors: RMSE (174.6), MAE (105.1), and MAPE (0.687). This comparison indicates that while the model performs reasonably well on the training data, its forecasting accuracy diminishes slightly when predicting future values

5.2. Performance on Dataset 2 Pakistan (target monthly)

The figure 10 displays the performance results of different machine learning models—SVM, Random Forest, Gradient Boosting, and Linear Regression using the Orange data mining tool on a dataset focused on rainfall prediction, likely from a meteorological dataset. The models are evaluated using 2-fold cross-validation with stratified sampling to maintain the proportional distribution of data across folds. Additionally, random sampling is employed, where the dataset is split into 80% training and 20% testing data, repeated 20 times for robust evaluation. The "Leave one out" technique is also applied, where a single instance is left out at each iteration while the model is trained on the remaining data, making it especially suitable for small datasets. This comprehensive evaluation setup helps in accurately assessing each model's performance across different scenarios. The results indicate that Linear Regression and Gradient Boosting perform best with lower error metrics and higher R² values, while SVM struggles with higher errors and a negative R², suggesting it is less suitable for this particular dataset.Among the models evaluated, Linear Regression produced the best results. It exhibited the lowest Mean Squared Error (MSE) of 270.138, the lowest Root Mean Squared Error (RMSE) of 16.436, and the lowest Mean Absolute Error (MAE) of 11.572. Additionally, it achieved a relatively high R² value of 0.453, indicating that it explains a good portion of the variance in the rainfall prediction data. This result suggests that Linear Regression is the most effective model for this particular dataset, outperforming the other models in terms of accuracy and predictive power.

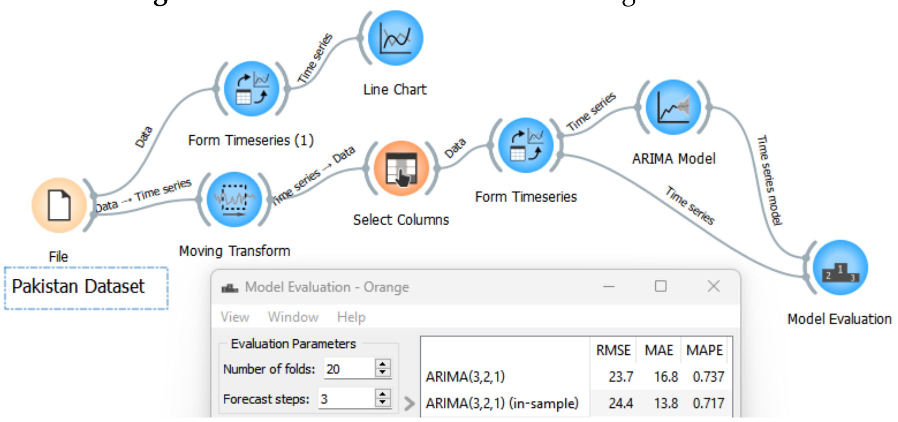
The figure 11 presents the results of applying an ARIMA (3,2,1) model to a Pakistan dataset, focusing on both in-sample and out-of-sample performance. The ARIMA model, configured with 3 autoregressive terms (p=3), 2 differencing operations (d=2), and 1 moving average term (q=1), is used to analyze the time series data. The out-of-sample evaluation, which measures the model's predictive ability on new, unseen data, shows an RMSE of 23.7, MAE of 16.8, and MAPE of 0.737. The in-sample evaluation, which assesses how well the model fits the training data, yields an RMSE of 24.4, MAE of 13.8, and MAPE of 0.717. These results indicate that the model's performance is relatively consistent between the training data and the unseen data, with slight differences in error metrics. This suggests that the ARIMA model is performing reasonably well for forecasting on this dataset.



○ Cross validation
Number of folds: 2
□ Stratified
○ Cross validation by feature
○ Random sampling
Repeat train/test: 20
Training set size: 80 %
□ Stratified
● Leave one out

| Model | MSE | RMSE | MAE | MAPE | R2 |
|---|---|---|---|---|---|
| SVM | 523.847 | 22.888 | 18.047 | 3.978 | -0.060 |
| Random Forest | 352.191 | 18.767 | 12.987 | 1.250 | 0.287 |
| Gradient Boosting | 282.325 | 16.803 | 11.734 | 1.313 | 0.429 |
| Linear Regression | 270.138 | 16.436 | 11.572 | 1.238 | 0.453 |

**Figure 10.** Result on Pakistan dataset using Arima Model



Evaluation Parameters
Number of folds: 20
Forecast steps: 3

| | RMSE | MAE | MAPE |
|---|---|---|---|
| ARIMA(3,2,1) | 23.7 | 16.8 | 0.737 |
| ARIMA(3,2,1) (in-sample) | 24.4 | 13.8 | 0.717 |

**Figure 11.** Result on Pakistan dataset using Arima model

### 5.3. Discussion on Result

For dataset 1 Bangladesh gradient boosting algorithm work well and for dataset 2 Pakistan Linear regression algorithm work well according to evaluation criteria. This paper analyzes the results of applying the ARIMA models to time series data originating from Bangladesh and Pakistan. For the Pakistan data, the ARIMA(3,2,1) model shows good in and out-sample fit again signifying a good forecasting model by ARIMA models. In contrast to the Afghanistan dataset, the Bangladesh dataset presents a clearer picture of deterioration in the model's performance from in-sample to out-of-sample values implying that the model has perhaps over fit. Furthermore, machine learning models were applied to both the datasets. For Bangladesh, Gradient Boosting outperformed others with the lowest error values (MSE: (Mean Absolute Error, RMSE, R-squared 6435. 975, 80. 225, 51. 064) and highest $R^2$ score of 0. 842 which helped in modeling more complex relations in rainfall data set. When forecasting in the Pakistan dataset, the best performance of Linear Regression was shown with the least MSE of 270. 138, RMSE of 16. 436, and the smallest MAE of 11. 572.

## 6. Model Accuracy and Robustness

### 6.1. ARIMA Model Accuracy and Robustness

ARIMA is highly effective for time-series data that exhibit consistent linear trends and seasonal patterns. It can model the relationship between past and future values, which makes it a strong choice for data with predictable patterns. However, ARIMA struggles with capturing complex, non-linear relationships within the data. It requires the data to be stationary or requires differencing to achieve stationarity, which can sometimes lead to loss of important information.

### 6.2. Random Forest Accuracy and Robustness

Random Forest is known for its ability to handle diverse features and non-linear relationships in the data. It works well across various types of datasets and is particularly effective when there are many variables that interact in complex ways. While Random Forest is robust in terms of handling noise and over fitting, it may not be the best choice for time-series data, as it doesn't inherently capture temporal dependencies. Additionally, it can sometimes be less interpretable compared to linear models.

### 6.3. Linear Regression

Linear Regression is easy to understand and provide good results when the relationship between independent and dependent variable is direct. Nonetheless, it is less effective for curvilinear interactions or those cases where the data involves cross-interactions. However, it cannot model situations where the relationship between the independent and dependent variable is other than linear thus making it rigid while, it is accurate in apprehending situations that do not require complex analysis because of large data sets.

### 6.4. Gradient Boosting

Gradient Boosting is highly accurate for various types of data, particularly when there are complex, non-linear relationships to be captured. It builds an ensemble of weak learners, typically decision trees, to create a strong predictive model. Gradient Boosting is robust in terms of handling different data distributions and is less prone to over fitting than some other models. However, it can be computationally expensive and may require extensive hyper parameter tuning.

### 6.5. Support Vector Regressor

SVR is effective for both linear and non-linear data through the use of different kernel functions. It works well when the relationship between variables is not strictly linear and when the dataset is relatively small. While SVR is robust in handling small to medium-sized datasets, it can be sensitive to the choice of kernel and parameters. It also tends to be computationally intensive, especially with large datasets, and may struggle with noisy data.

## **7. Computational efficiency and robustness across datasets**.

Thus, Linear Regression and ARIMA are the least time-consuming models in terms of their computational time and are recommended for use when working with big data or if the data processing time is critical. But they are not as sound when operating across different data sets – Linear Regression fails on anything non-linear, ARIMA only works with regularly reoccurring, linear time series information. Thus, Gradient Boosting and Support Vector Regressor (SVR) as well as Random Forest (RF) are more intricate in terms of computational complexity. As for computational efficiency, GB is slightly slower than other methods, yet it is one of the most stable algorithms in terms of datasets, and it actively works with non-linear correlations. SVR is good for both linear and non-linear values but the computation time increases while dealing with a number of values. Random Forest can be considered balanced in terms of computational complexity and performance; works with high order interactions and noise; it is generally insensitive to model order, making it suitable for various datasets, albeit, it is not suitable for dynamics. In general, the decision on which model to choose depends on certain properties of the utilized dataset as well as the tolerance of time for the execution of the algorithm and the level of noise in the given data.

## **8. Conclusion**

For the Pakistan dataset, the ARIMA model is more stable because the in-sample and out-of-sample errors are relatively close. A relatively high in to out of sample performance decline in case of the Bangladesh dataset is observed from the model suggesting the model might be more overfitting. Overall, the present study shows that the ARIMA model performs favourably on the Pakistan dataset especially in out of sample forecast. Among the machine learning models in the Bangladesh dataset, The Gradient Boosting model has the least error values in terms of MSE (6435.975), RMSE (80.225) and MAE (51.064) while having the highest $R^2$ score of 0.842 meaning that this Gradient Boosting has the best performance in the dataset and the capability to fit the rainfall data. In the models fitted from the Pakistani data, the best performance was realized by the Linear Regression model. This model showed the best performance because of least values of Mean Squared Error (MSE) which was 270.138, Root Mean Squared Error (RMSE) of 16.436 and Mean Absolute Error (MAE) of 11.572.

## 9. Future Work

-To more precisely capture geographical dependencies in rainfall data, future research should concentrate on integrating spatial data and sophisticated spatial modelling tools.

-To further improve prediction accuracy, research hybrid models that incorporate deep learning techniques with ARIMA or VAR to capture intricate temporal and geographical patterns for rainfall forecasting.

-Do research relevant to a particular place in order to localize models to the particular farming conditions and environmental considerations. Varied regions may interact with climates differently and have varied patterns of precipitation; hence, different forecasting techniques are required for each location.

### References

1. Parmar, A., Mistree, K., & Sompura, M. (2017, March). Machine learning techniques for rainfall prediction: A review. In International conference on innovations in information embedded and communication systems (Vol. 3).
2. Hussein, Eslam A., Mehrdad Ghaziasgar, Christopher Thron, Mattia Vaccari, and Yahlieel Jafta. "Rainfall prediction using machine learning models: literature survey." Artificial Intelligence for Data Science in Theory and Practice (2022): 75-108.
3. Appiah-Badu, N. K. A., Missah, Y. M., Amekudzi, L. K., Ussiph, N., Frimpong, T., & Ahene, E. (2021). Rainfall prediction using machine learning algorithms for the various ecological zones of Ghana. IEEE Access, 10, 5069-5082.
4. Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Machine Learning with Applications, 7, 100204.
5. Basha, C. Z., Bhavana, N., Bhavya, P., & Sowmya, V. (2020, July). Rainfall prediction using machine learning & deep learning techniques. In 2020 international conference on electronics and sustainable communication systems (ICESC) (pp. 92-97). IEEE.
6. Hussain, J., & Zoremsanga, C. (2021, November). A survey of rainfall prediction using deep learning. In 2021 3rd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE) (pp. 1-10). IEEE.
7. Baig, H. F. A., Islam, N., & Alim, A. (2023). Exploring Machine Learning and Deep Learning Approaches for Disaster Prediction and Management: A Survey of Different Approaches: A Survey of Different Approaches for Disaster prediction and management through tweets. Pakistan Journal of Engineering, Technology & Science, 11(1), 45-73.
8. Khan, M. M. H., Mustafa, M. R. U., Hossain, M. S., Shams, S., & Julius, A. D. (2023). Short-Term and Long-Term Rainfall Forecasting Using ARIMA Model. International Journal of Environmental Science and Development, 14(5), 292-298.
9. Swain, S., Nandi, S., & Patel, P. (2018). Development of an ARIMA model for monthly rainfall forecasting over Khordha district, Odisha, India. In Recent findings in intelligent computing techniques (pp. 325-331). Springer, Singapore.
10. Hong, W. C. (2008). Rainfall forecasting by technological machine learning models. Applied Mathematics and Computation, 200(1), 41-57.
11. Chen, G., & Wang, W. C. (2022). Short-term precipitation prediction for contiguous United States using deep learning. Geophysical Research Letters, 49(8), e2022GL097904.
12. Jin, W., Luo, Y., Wu, T., Huang, X., Xue, W., & Yu, C. (2022). Deep learning for seasonal precipitation prediction over China. Journal of Meteorological Research, 36(2), 271-281.
13. Espeholt, L., Agrawal, S., Sønderby, C., Kumar, M., Heek, J., Bromberg, C., ... & Kalchbrenner, N. (2022). Deep learning for twelve hour precipitation forecasts. Nature communications, 13(1), 1-10.
14. Kumar, B., Atey, K., Singh, B. B., Chattopadhyay, R., Acharya, N., Singh, M., ... & Rao, S. A. (2023). On the modern deep learning approaches for precipitation downscaling. Earth Science Informatics, 16(2), 1459-1472.
15. Dinh Ty Nguyen , andShien-Tsung Chen ,"Real-Time Probabilistic Flood Forecasting Using Multiple Machine Learning Methods", Water 2020, 12(3), 787; https://doi.org/10.3390/w12030787
16. Nushrat Jahan Ria; Jannatul Ferdous Ani; Mirajul Islam; Abu Kaisar Mohammad Masum,"Standardization Of Rainfall Prediction In Bangladesh Using Machine Learning Approach", 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)
17. Miah Mohammad Asif Syeed; Maisha Farzana; Ishadie Namir; Ipshita Ishrar; Meherin Hossain Nushra; Tanvir Rahman,"Flood Prediction Using Machine Learning Models", 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)
18. Y. Tikhamarine, D. Souag-Gamane, A. N. Ahmed, S. S. Sammen, O.Kisi, Y. F. Huang, and A. El-Shafie, "Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization," Journal of Hydrology, vol. 589, p. 125133, 2020.
19. Pour, S. H., Abd Wahab, A. K., & Shahid, S. (2020). Physical-empirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia. Atmospheric Research, 233, 104720.
20. He, S., Raghavan, S. V., Nguyen, N. S., & Liong, S. Y. (2013). Ensemble rainfall forecasting with numerical weather prediction and radar-based nowcasting models. Hydrological Processes, 27(11), 1560-1571.
21. Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., & Lane, J. E. (2009). Precipitation: Measurement, remote sensing, climatology and modeling. Atmospheric Research, 94(4), 512-533.

22. Souto, Y. M., Porto, F., Moura, A. M., & Bezerra, E. (2018, July). A spatiotemporal ensemble approach to rainfall forecasting. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
23. Rahman, A. U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., ... & Mosavi, A. (2022). Rainfall prediction system using machine learning fusion for smart cities. Sensors, 22(9), 3504.
24. 2Yasmeen, F., & Hameed, S. (2018). Forecasting of rainfall in pakistan via sliced functional times series (SFTS). World Environment, 8(1), 1-14.
25. Dimri, T., Ahmad, S., & Sharif, M. (2020). Time series analysis of climate variables using seasonal ARIMA approach. Journal of Earth System Science, 129, 1-16.
26. Geetha, A., & Nasira, G. M. (2016). Time-series modeling and forecasting: Modeling of rainfall prediction using ARIMA model. International Journal of Society Systems Science, 8(4), 361-372.
27. Yudianto, M. R. A., Agustin, T., James, R. M., Rahma, F. I., Rahim, A., & Utami, E. (2021). Rainfall forecasting to recommend crops varieties using moving average and naive bayes methods. International Journal of Modern Education and Computer Science, 12(3), 23.
28. Farman, H., Khan, A. W., Ahmed, S., Khan, D., Imran, M., & Bajaj, P. (2024). An Analysis of Supervised Machine Learning Techniques for Churn Forecasting and Component Identification in the Telecom Sector. Journal of Computing & Biomedical Informatics, 7(01), 264-280.
29. Farman, H., & Ahmed, S. (2023). Novel framework for efficient detection of QRS morphology for the cardiac arrhythmia classification. Journal of Computing & Biomedical Informatics, 5(02), 12-20.
30. Farman, H., Islam, N., Ali, S. A., Khan, D., Khan, H. A., Ahmed, M., & Farman, A. (2024). Advancing Rainfall Prediction in Pakistan: A Fusion of Machine Learning and Time Series Forecasting Models. International Journal of Emerging Engineering and Technology, 3(1), 17-24.
31. Farman, H., Ahmed, S., Imran, M., Noureen, Z., & Ahmed, M. (2023). Deep Learning Based Bird Species Identification and Classification Using Images. Journal of Computing & Biomedical Informatics, 6(01), 79-96.
32. Bakr, M. A., Amjad, U., Raza, A., Khurram, M., Farman, H., & Ali, A. COVID-19 Diagnosis Using Transfer Learning on X-ray Images.
33. Kumar, R., Farman, H., Kumar, N., Abro, S. A., Farman, A., & Umer, M. Evaluation of Learning management system using Data Mining techniques.
34. Dabral, P. P., & Murry, M. Z. (2017). Modelling and forecasting of rainfall time series using SARIMA. Environmental Processes, 4(2), 399-419.