Societal Transformation: AI and Big Data Journal
Vol: 2(1): 64-93, 2024
DOI:10.20547/aibd.242104

Innovative Time Series Forecasting Methods and Models for Rainfall Forecasting to Boost Agricultural Business in Pakistan

Hira Farman[1,2], Noman Hasany[2]

Abstract: The challenge of predicting rainfall is undoubtedly a difficult one because there are an extensive number of different factors and elements that influence the conditions of the climate. It is of extreme significance to have accurate rainfall forecasts, particularly for the agricultural sector, which is highly reliant on timely and sufficient rainfall for the growth and yield of crops. The contribution that agriculture makes to the economy is another factor that highlights the need of accurate rainfall forecasts. There are many different ways that have been utilized all around the world in order to forecast the patterns of rainfall. These methods include classical statistical methods that are frequently used to advanced machine learning techniques. Thus, in this work, two quite opposite methods of creating a model are described. Such approaches comprise of statistical method which employs the Autoregressive Integrated Moving Average (ARIMA) alongside with the vector auto regression (VAR) model. The Box-Jenkins method provided the foundation on which the ARIMA and VAR modelling procedures employed in this study were built on. These techniques consisted of four primary stages: fitting or estimating a model, making predictions, assessing the model's quality, and evaluating the model's parameters. The purpose of these stages was to assess the performance of the prediction process and for this a mean annual rainfall data for each year from 1901 to 2016 of country Pakistan was used.. The models were trained using data spanning 116 years, which included the mean annual rainfall. When applied to the data, Time series forecasting methodologies are utilized in order to produce the optimization coefficients and the regression coefficients, correspondingly. Based on the findings of the study. The ARIMA model doesn't fit the data as well as the VAR model does. This is evident from the fact that the RMSE, MAE, MAPE, AIC/BIC, POCID, and R^2 numbers are smaller and stronger, respectively. In this case, the VAR model is better at predicting how much rain will fall each month and throughout the year. The government, researchers, and private individuals should take into consideration these univariate and multivariate models when making plans for the future in order to increase agricultural commodity production and prevent damages caused by excessive precipitation.

[1]Department of Computer Science, Karachi Institute of Economics and Technology, Karachi, Pakistan.
[2]Department of Computer Science, Iqra University, Karachi, Pakistan

Keywords: Time Series, Rainfall, Vector Auto Regression (VAR), and ARIMA (Autoregressive Integrated Moving Average)

1. Introduction

Rainfall is one among the various phenomena in the weather system which cannot be predicted with a large degree of accuracy [1]. Indeed, it can be said that almost all business industries are related to this projection and in particular the agriculture industry is considerably significant as the industry has a huge contribution to the national economy. It is for this reason that one can note with a considerable interest that a number of attempts have been initiated with the help of a variety of methodologies with the objective of predicting its behavioral pattern as far as the international context is concerned. Thus, it remains very valuable to be ready for a disaster which means that the accurate forecast of the amount of rainfall is the indefeasible essential there is the identification of the study [6]: it implies that the objective of the study is to follow the efficiency of many different time series models in the prediction of rainfall. For instance, the Auto-Regressive Integrated Moving Average Model which is widely known as the ARIMA Model are exceptional tools that can be used with a view of coming up with forecasts for the short term. Therefore, the aim of this study is to clearly determine a situation where there is no possible data that could have preceded this result. Regarding

The efficiency of this model in addressing the forecasting scenarios, and, in particular, the long-term projections which this present case can feed to it for better result formulation having incorporated the other range of inputs that were not included in the current case such as historical rainfall, the specific preferred model is the RNN which is an abbreviation of recurrent neural network. This current investigation aims at wanting to establish how this feature manifests in the rainfall by comparing the two methods which include the VAR method and the ARIMA model [15-17]. This is done with the intention of making a comparison between the two techniques of ARIMA and VAR in the view to establish whether any relationship between the two exists.

The overall aim of this experimental study is to design and evaluate a new the hybrid rainfall prediction model that combines fuzzy models and artificial neural networks and SARIMA. The idea is to amplify the weaknesses of the SARIMA models that are caused by linearity of the model and dependency on the large amount of data. By combining of these methodologies, this study will seek to enhance the performance of seasonal time series forecast especially where they might not be enhanced enough by normal SARIMA models due to high rates of technological change and unforecastable events [30]. This study [31] reveals that the ARIMA model then provides a better forecast of rainfall with less mean absolute error. Such results that obtained can help to predict the monsoon rainfall in the subsequent years.

In this empirical research work, multivariate vector auto regression model (VAR) and uni variate autoregressive integrated moving average (ARIMA) models were used. The most commonly used method for describing temporal dependencies for a given set of time series, which all take place in a single point in time, is ARIMA which includes autoregressive work (AR), differencing between (I), and moving average (MA).VAR is a multivariate model with more than one-time variable that builds on the univariate autoregressive model that is connected to other time series. On this, its main strength or primary application is that the model deals with non-stationary data through differencing, which is beneficial for series with patterns and seasonality. As compared with ARIMA, the use of VAR models takes into consideration the interdependencies between a number of time series variables and thus provides a better understanding of how these variables influence or react to each other.

1.1 Importance of Rainfall for Agricultural Business Department

It's crucial to be able to precisely and accurately predict rain for a number of reasons. It helps handle water resources by making sure that irrigation systems and reservoirs have enough water. Accurate precipitation prediction also helps farmers plan activities like planting and gathering, which lowers the chance that crops will fail because of bad weather condition. Improving the accuracy of rainfall forecasts can help Pakistan country, where agriculture makes up a big part of the(Gross Domestic Product)GDP, get more crops and use its water resources more efficiently. Being able to correctly interpret when it will rain can completely change and advantage the farming industry in Pakistan. If farmers can trust the weather forecasts, they can plan their planting date's better, use less water for irrigation, and keep crop losses to a minimum. Precisely predicting when it will rain can also help with planning how to get rid of pests and diseases, since some weather conditions can make these problems worse. Better predictions of rain can help farmers better manage their crops, get higher yields, and make more money, which is good for the growth and stability of Pakistan's agricultural industry as a whole. Figure 1 illustrate why rainfall prediction is so vital for farming in Pakistan:

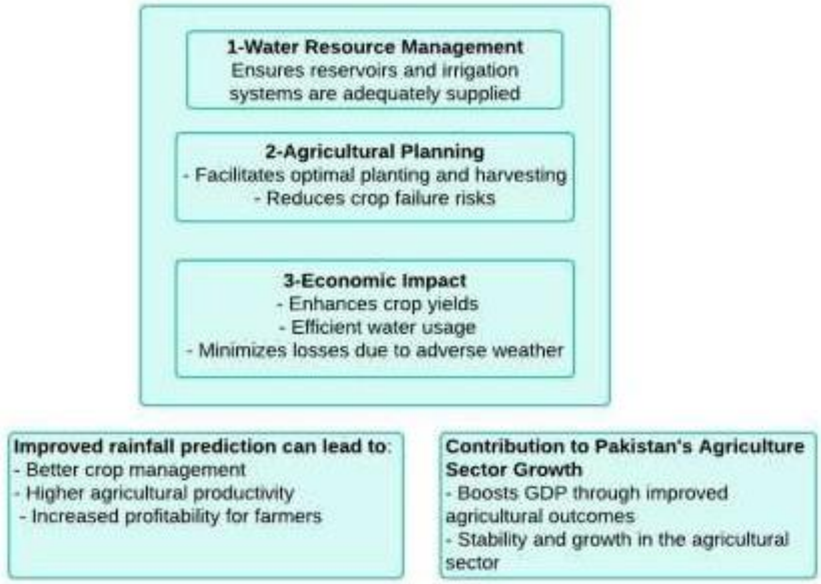

Fig 1: Importance of rainfall for agriculture in Pakistan:

Predictions of terrible rain can have measure effects on many people life , Infrastructure and business, especially in places like Pakistan where farming and managing water resources are very important. These effects can be roughly put into four groups: economic, environmental, social, and farming. Figure 2 shows some of the results.

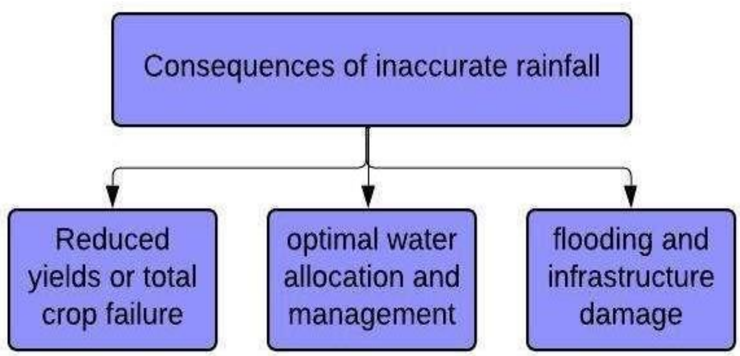

Fig 2: Consequences of Rainfall

1.3 Types of Rainfall

Rainfall can be classified into several types [18-19] based on its intensity and duration, including light rain, moderate rain, heavy rain, and thunderstorms. Each type has different implications for agricultural activities. Light and moderate rain may be beneficial for crops, while heavy rain and thunderstorms can cause damage. Understanding the types of rainfall and their potential impact is essential for effective agricultural planning and risk management. Fig 3 shows some types of rainfall and their impact

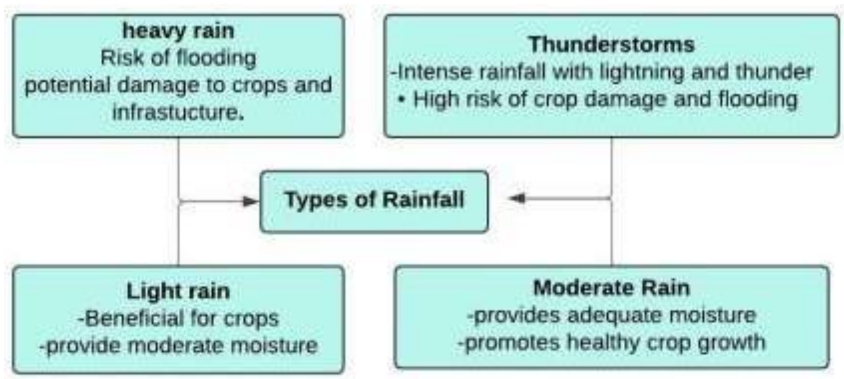

Fig 3: Types of Rainfall

2. Literature Review

This [1] study aim was to establish a reliable method of forecasting rainfall in the coastal area of the Indian sub-continent of using the ARIMA model. Forecasting is one of the significant and interesting professions of the twenty-first century particularly forecasting the rainy days. Climate and precipitation, of course, are generally very complex – and, often, non-linear – quantities that require detailed computer, mathematics and simulations for Global Projection. It does, however, mean that precipitation is temporally and spatially correlated and that is why it does so. Recently, flood disasters related to rainfall increase in frequency and therefore weather prediction is a more and more important branching. Equally often, the necessary data corresponded with the accessible meteorological data in a linear manner that was the aim of the researcher [4]. In the context of this study, the sample case of discussion is Twitter as a social media platform applicable for disaster-related research. As for the primary research methods of this study, the following tools are considered [5]: the specifics of deep and machine learning as well as the high technologies of disaster prediction. The second objective is to classify the concepts of various types of data and their source in relation to the given professions and crises. Regarding the short term and long term forecast of the rainfall, the suitable technique to model was ARIMA for the mentioned area of study which was Klang River Basin, Selangor [14]. According to the formulated problem of the study [15], the use of ARIMA in the prediction of the monthly rainfall of the Khordha district in the state of Odisha, India is proposed. Since there were no records for the recent years, the records of the rainfall were taken on monthly basis to fill the gaps between the years 1901 and 2002. The analysis of the present investigations was then advanced further to compare the AIC and BIC, to arrive at the appropriate model choice; as a result it has also been found out that the proper fitting model for the present occasions is the ARIMA (1, 2, 1) $(1, 0, 1)^{12}$ model. In this work [16], the auto-regressive integrated moving average (ARIMA) model is generated encompassing the seasonal property in rainfall of INRB besides the non-seasonal component. To build the monthly precipitation forecast model used in the paper, the rainfall gathered from the whole basin over the period of June 1951 – May 2005 is incorporated. The validation of the model is done yearly for a total of fourteen years starting in June 2005 and ending in May 2019. As highlighted earlier, the yield time series with the appropriate forecasting model ARIMA (2, 0, 0) $(4, 1, 4)^{12}$ is the hypothesis that gives the best fit as depicted from the study outcomes. This study [17] employed ARIMA models for some Louisiana State stations with the integrated component of order one, and ARMA models for some of the stations without the integrated component. The average of Monthly rainfall for the two locations was predicted using the models constructed above. Whereas, the frequency analysis is helpful for recognizing the rainfall behavior during long time periods, though in the outcome of the numerical researches, mostly the conclusions are limited to the maximum values [20-23]. However, other authors have also examined the characteristics of precipitation and its behavior globally through various methods [24][25][26]. Another issue in the research in the natural resources is the availability of data. To obtain historical data for analysis is relatively easy only in a few countries. Such confounding implies that it becomes very challenging for academics to make trustworthy inferences about decade or season trends as a result. In the context of the present work [28], LT and CT are used for constructing rainfall model through VAR technique As revealed in. For the climatology data, the Darmaga Bogor station was used, the climatology period which was between the years 2006 to 2017. Also, t-2 air temperature, t-1 moisture air, and t-1 moisture air influence the rainfall variables within the VAR model of the t-month (2). For the purpose of modeling the next time period, the following form of VAR model (2) can be used. The value of 42. MAPE was found to be equal to 18 and it was found that the most appropriate model to describe the studied time series was the VAR (2).The identification of wind direction, which was done circular data, and the forecast of rainfall by employing VAR modelling are two major innovations of this study. This study aims at making a forecast in respect of the Geographical variation of Rain fall that will happen in Assam and Meghalaya in the subsequent five years [29]. Therefore, to achieve the aim of the current research, a new MHF model will be proposed and tested. This model will integrate two different computer intelligence approaches, SARIMA and artificial neural networks, and fuzzy models. As a target of this undertaking, it is aimed at eradicating the constraints of SARIMA models because of its linearity and dependency on vast historical data. The goal of this work is to increase the efficiency of Seasonal Time Series forecasting in conditions

where there are a great many methods to use but fewer examples when SARIMA by itself would
be enough, because of unpredictability and fast-paced development in technology [30]. As stated in this study [31], the current error of the model related to the ARIMA model is somewhat lower, so the model developed can be employed to predict monsoon rainfall in the future years. Hybrid model [33] also design for rainfall prediction tasks.

3. Comparative Analysis

More detailed and structured example of a comparative analysis table 1 based on actual research papers, focusing on time series modeling and forecasting:

Table 1: comparative analysis with past work

ref	Model	metrics	Country
14	ARIMA ARIMA (1,0,3) ARIMA (1,0,2)	R^2 (Coefficient of Determination)	Klang River Basin in Selangor
15	ARIMA ARIMA (1, 2, 1) $(1, 0, 1)^{12}$	Bayes information criterion (BIC) and an Akaike information criterion (AIC) are two examples of information criteria.	Indian state of Odisha's Khordha district
17	ARMA ((0,0,1) $(0,0,2)^{12}$)	RMSE and MAE ,AIC AND BIC	Louisiana
27	LSTM (Long Short-Term Memory)	Various measures like loss, root mean squared error (RMSE), mean absolute error(MAE) and root mean squared logarithmic error were applied to measure the performance of the models.	Australia
28	VAR model (2)	Mean Absolute Percentage Error (MAPE)	Bogor, Indonesia
31	ARIMA	Two metrics used being mean Squared Error (MSE) and the other being Root Mean Squared Error (RMSE).	Tamilnadu.

A variety of models have been investigated in previous research on rainfall prediction in order to address the intricacies of environmental variability and the vital significance of precise forecasts for disaster planning, water resource management, and agriculture. For example, because ARIMA models can capture seasonality and temporal dependence in rainfall patterns, they are frequently used. Research like the ones conducted in the Khordha area of Odisha and the Klang River Basin in Selangor have used ARIMA variations such as (1,0,3) and (1,2,1)(1,0,1), respectively, highlighting the model's effectiveness in particular regional contexts. Similar to this, the application of ARMA models in Louisiana demonstrated how well they could capture transient variations and support local decision-making. Recent, studies, conducted in Australia [27], shows that, methods such as Long Short-Term Memory or LSTM networks can be useful in reaching high level of forecast accuracy when it comes to paying attention to long-term relations and non-linearities, in the rainfall data.In the same context, the case of Bogor, Indonesia has made it clear how good the VAR models are in the prediction of fluctuation in rainfall and the overall efficiency of the VARs in all its multivariate modeling. Together, these studies demonstrate the various approaches and their advantages in tackling the difficulties associated with rainfall prediction in various climatic and geographical contexts, thereby advancing global efforts to improve agricultural planning, water management plans, and disaster risk reduction.

4. Methodology

4.1 Dataset Description

The collection called "Pakistan Dataset 1901 to 2016" has historical rainfall data for the whole country of Pakistan from 1901 to 2016. "Rainfall – (MM)," "Year," and "Month" are three separate features in this dataset. In millimeters (mm), the "Rainfall - (MM)" attribute shows how much rain fell in a certain month. The "Year" feature lists the year that the rainfall was recorded, and the "Month" column lists the month that the rainfall was recorded. There are 1392 rows in the dataset, and all of them are not null, which means there are no missing numbers. The properties have the following data types: "Rainfall (MM)" is a decimal number, "Year" is a number, and "Month" is an alphabetical character data type, which can be seen in fig 4.

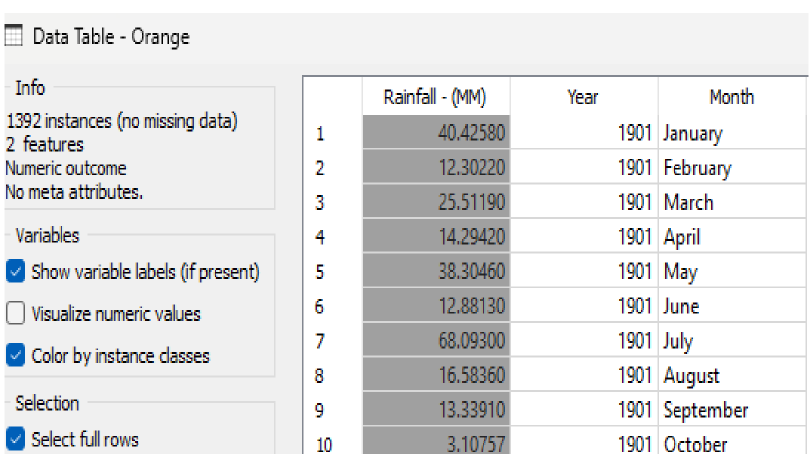

Fig4: Representation of dataset rows and features

4.2 Month-wise Rainfall Prediction Insights

Climatic conditions of seasonal precipitation make it easier to subdivide the year into period sections with expected measures of precipitation to make forecaster weather plans. The mains sources of water are not expected to be much during the dry months of May, June, October and November which are thus the best times to observe water conservation and drought readiness. July and August are the months characterized by highly intensive and frequent rainfall, what means high probability of flood occurrence; that underlines the crucial importance of accurate meteorological predictions in the frames of the problem of disaster prevention. Rainfall distribution of the periods: January, February, March, April and December is considered to be moderate, therefore these periods are suitable for effective organization of agricultural activities and water resources management. Specifically, fluctuations of temperatures during two periods: in May – June and in September – October signal significant changes in the weather. It is also in this respect that accurate forecasts are necessary in order to be able to predict transitions from dry to wet and from wet to dry situations. Through this systematic approach toward understanding fluctuations in rainfall through the change of seasons, it is imperative to employ models for specific forecasting within industries and or organizations whose operations depend on weather and climate, with a view of enhancing preparedness in terms of climate behaviors within different time horizons.

The monthly rainfall distribution in Pakistan from 1901 to 2016 is depicted in Figure 4 of the chart, with a focus on both the yearly rainfall intensity and frequency. The y-axis displays the number of rainfall events, and the x-axis shows the months. The color scale, ranging from blue (low rainfall) to yellow (high rainfall), provides a visual representation of rainfall intensity in millimeters. The data reveals distinct seasonal patterns. January to April sees moderate rainfall, with March experiencing a slight peak. Rainfall sharply decreases in May and reaches its lowest in June, marking the start of a dry period. The most significant rainfall occurs in July and August, evidenced by high frequencies and intensities, including periods associated with flooding. This is the monsoon season, critical for water resource planning and flood risk management. Rainfall declines in September but remains relatively higher than in earlier months. October and November are notably dry, while December experiences a moderate resurgence in rainfall. These insights into monthly rainfall patterns are essential for developing predictive models, supporting agricultural planning, water resource management, and preparing for extreme weather events like floods.

Fig 4. Representation of Rain occurrence Month wise

4.3 Overall Explanation of Year-wise Rainfall Prediction (1901-2016)

The spiralogram visualization provides an overview of year-wise rainfall distribution in country Pakistan from 1901 to 2016. The chart represents the data in a circular format, with each segment corresponding to a specific time period. The inner circle begins with the earliest years (pre-1902), and the outer segments represent later years up to 2000. The bin width is set at 2.5 years, offering a detailed breakdown of rainfall patterns over short intervals. The color coding indicates the mean rainfall values for each period, ranging from blue (22 to 25 mm) to yellow (32 ranges 35 mm). The presence of alternating blue and yellow segments suggests variability in annual rainfall patterns, with certain periods experiencing higher rainfall (yellow) and others lower (blue).Key insights from this visualization include.

-In Early 20th Century: Before 1925, there were years with more and less rain, which shows that early history data was not always stable.

-In Mid 20th Century: From 1947.5 to 1975, there is a clearer change between higher and lower rainfall, which shows that the weather was changing during this time.

-In Late 20th century: The years before 2000 also followed a similar trend, with big changes between years with more and less rain.

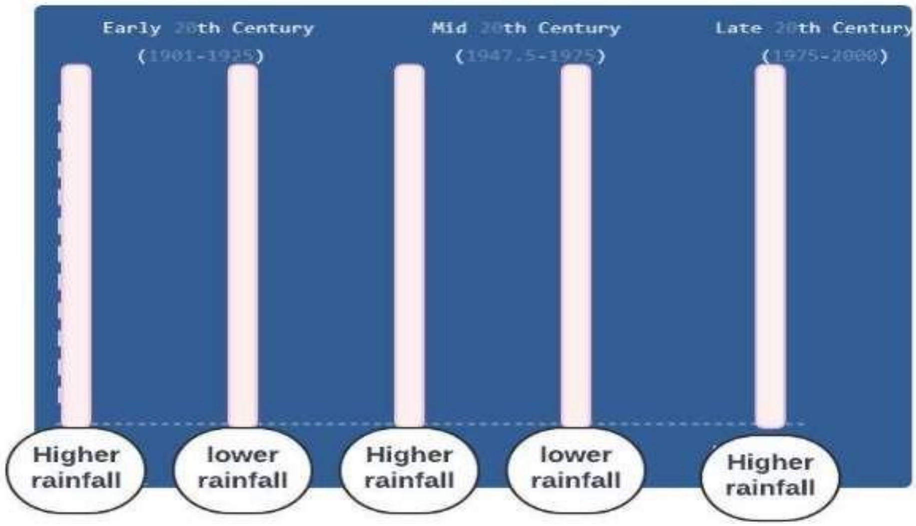

Fig 5: Representation of Rainfall according to century

In order to establish the variation in rainfall that has occurred in Pakistan, the data of the rainy season generally throughout the century depicted in fig 5 is divided year-wise. Such information is of great importance for analysis of changes occurring within the climate for years and improvement of the models for the calculation of the further rainfall. These patterns may provide the basis for water resources, agricultural and natural disasters activities and plans.

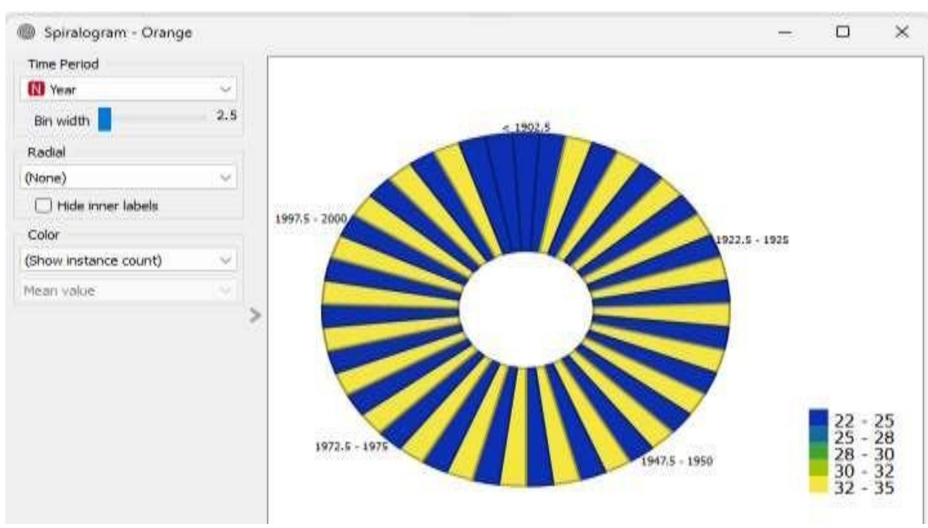

Fig 6: Representation of Rain occurrence year wise

Figure 6 shows a spiral gram that shows how much rain fall in Pakistan each year from 1901 to 2016. Its main points show important information about how the weather has changed over the past 100 years. There are periods of higher (yellow segments) and lower (blue segments) precipitation amounts shown on the chart, which shows that rainfall happens in cycles. There are clear changes, especially around the middle of the 20th century and as got closer to 2000, which shows that rainfall in the area changes over time. The fact that this changes over time shows how many complicated climate factors affect Pakistan's rain trends. These kinds of findings are very important for figuring out long-term changes in the climate and can help shape plans for long-term water management, farming, and being ready for disasters in the specific area.

3.5 Proposed Methodology

The first step in predicting rain is to collect data in this experimental study dataset collected from Kaggle website. The Pakistan Meteorological Department(PMD) data has 1392 entries and 3 features rainfall, year, month in order to predict rainfall accurately and the utilized model's job is to guess how much rain will fall each month in the future (in millimeters). This includes preprocessing the data to make sure it is consistent, looking for missing numbers, and making sure the order of events is correct. In this study after analyzing the data through orange tool no missing values found data is built in clean.

Time series forecasting two models used for this work the first one is ARIMA which is univariate and the other one is VAR which is multivariate .ARIMA is a one-variable model, so when data is gathered, it is mostly focused on the "Rainfall (MM)" column. The right factors (p, d, and q) are found after the data's stationarity is checked and differencing is used if necessary. The ARIMA model is then put in place using orange tool, and residue analysis checks to make sure that the leftovers are just random noise. The fitting model is used to guess how much rain will fall each month in the future. The goal of the VAR model is to figure out how rainfall, month, and year affect each other and to predict future numbers based on these interactions. As part of getting the data ready for this experiment, differencing is used to make sure that the series for Rainfall (MM), Month, and Year are consistent and stationary. Utilized performance metric AIC, BIC, or HQIC to choose the lag order, which is the number of time periods the model uses to guess what the future values will be. Some diagnostic checks are residual analysis to make sure there is no white noise and stability checks to make sure the model is stable. The fitted VAR model is then used to guess what the Rainfall (MM) amount, on the basis of Month, and Year numbers will be in the future. Figure 7 shows that both models stress the importance of data uniformity, stationarity checks, and diagnostic validation to make sure that predictions are correct and dependable.

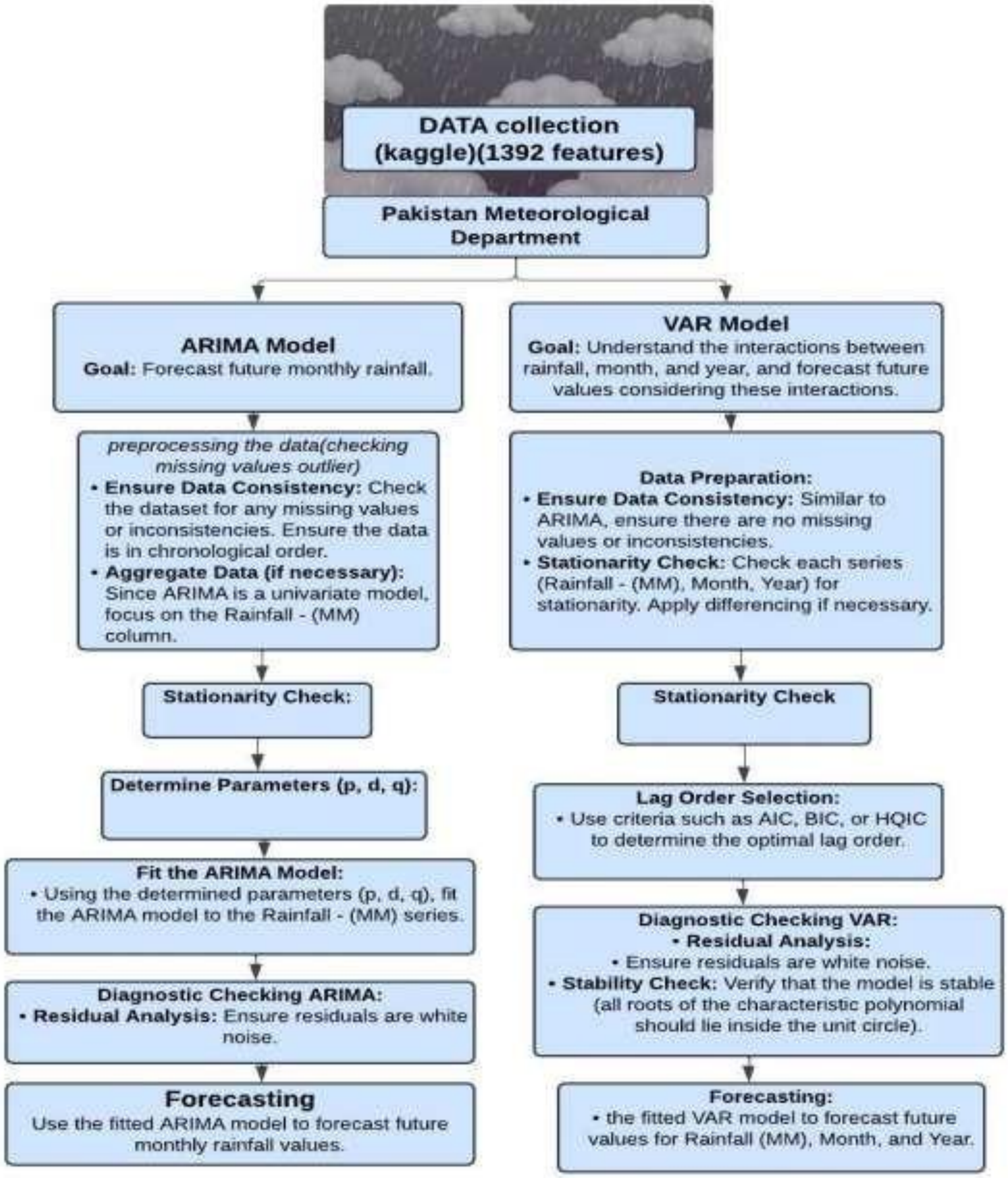

Fig 7: Proposed model for rainfall prediction using ARIMA and VAR

4.6 Model Discussion

A time-series model comprises successive data points projected at a predetermined interval of time. It includes techniques that try to infer or forecast something about a time series, or grasp the underlying idea of the data points in the time series. When utilizing time-series analysis to forecast data, a significant model is used to project future outcomes based on known historical results. Time series analysis models divided into two category univariate and multi variants Figure 8 displays the time series models and approach. In this study utilized one univariate model ARIMA (Autoregressive Integrated Moving Average) and one multi variate model VAR(Vector Auto regression) for rainfall forecasting by incorporating metrological data. The other models are ARCH and LSTM used for time series forecasting.

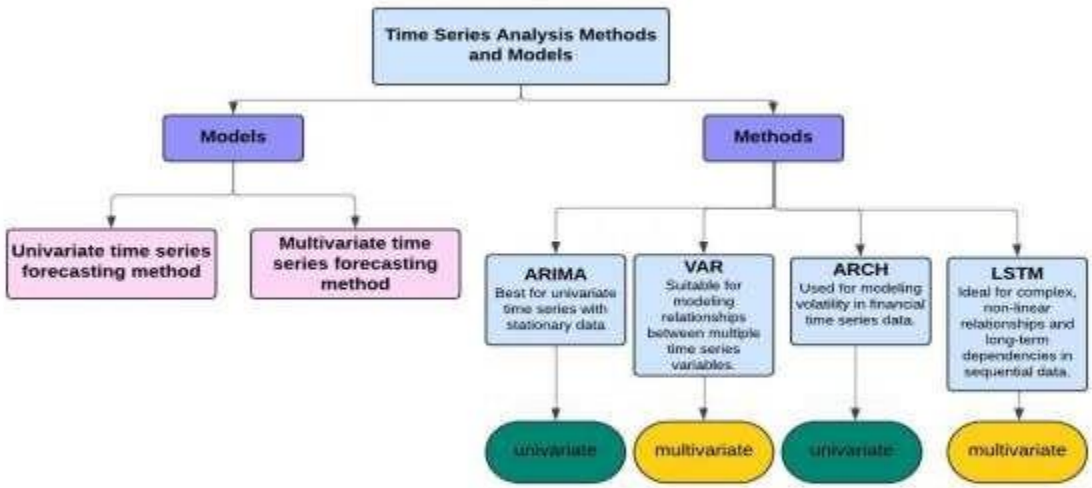

Fig 8. Time series models and methods used in study

4.6.1. Arima Model

When dealing with data that are univariate or those data which consist of one single variable that varies with time then the ARIMA model [1,5,8] is applied. Arising from the above time series data, the model that is frequently used is the ARIMA, which is an abbreviation for Auto Regressive Integrated Moving Average. ARIMA model has several key parameters which are usually represented by a p, d and q symbols.

Table 2: Representation of Arima Model Components parameter

ARIMA Univariate Model (Focus on a single time series (rainfall).					
Components	**Usage**	**Parameters**	**complexity**	**stationarity**	**Model fit pattern**
Time series AR (Autoregressiv	Univariate ARIMA is used in	p	low	Ensure stationarity and	Fit the model and forecast future rainfall

e):	forecasting single time series data set; it is ideal for time series data that is self correlated in that one event affects future events.			determine p, d, q	values.
Time Series I (Integrated):		d			
Time Serues MA(Moving Average):		q			

A. Parameter Description

Typically, an ARIMA model parameter is written as ARIMA (p, d, q), where:
p: (p represents the order of the autoregressive component).This variable represents the total number of lag observations that are incorporated into the model. It is sometimes referred to as the lag order

d: (d is the difference in degree between the two.)The frequency with which the time series is rendered stable by dividing the raw observations is denoted by the alphabet d.

q:(q symbolizes the order of the moving average component)The window size of the moving an average, which is also referred to as averaging order in some instances.

B. Components of ARIMA

1. 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 1

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

2. Integrated component (I)

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 2

$$Y`_t=Y_t-Y_{t-1} \qquad (2)$$

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

3. Component of Moving Average (MA)
In this section, we use the relationship between an observation and a residual error from a moving average model applied to lagged observations and an observation. In mathematics, equation 3 shows how to write an MA model of order q (MA(q)).

$$Y_t=\alpha+\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\cdots+\theta_q\epsilon_{t-q} \quad (3)$$

where:

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

4.6.2. VAR Model

This model, called Vector Autoregressive (VAR), takes multivariate time series data and integrates it to the single autoregressive model (AR). A VAR model lets it model several time-dependent variables at the same time. This means that each variable can be a linear function of both its own values and the prior values in the system as well as the prior values of the other variables in the system. This is why VAR is so useful for predicting multivariate time series data, which has many factors that depend on each other. For a VAR model with k time series variables, shown in equation 4, VAR(p), which stands for the model's order, looks like this:

$$Y_t==c+A_1Y_{t-1}+A_2Y_{t-2}+\cdots+A_PY_{t-p}+\epsilon_t \quad (4)$$

where:
$A_1, A_2, \ldots, A_P$ are vectors of coefficients, k * k matrices of coefficients, ϵ_t, is a vector of error terms (white noise), and $Y_t$is a vector of the k time series variables at time t.

Table 3: Representation of VAR Model Components and parameter

VAR multivariate time series Model (Consider multiple time series (rainfall, month, and year).)					
Components	**Usage**	**Parameters**	**complexity**	**stationarity**	**Model fit pattern**
Auto Regression:	VAR is suitable for datasets where multiple time series influence each	Lag Order (p):	VAR is more complex due to its consideration of interactions between multiple variables.	Ensure stationarity and select the optimal lag order.	In auto regression Fit the model and predict future values, considering the interactions

	other. It is often used in economics and finance where various indicators interact				between variables.

5. Result Discussion

Both the ARIMA (Autoregressive Integrated Moving Average) and VAR (Vector Auto Regression) were used in this study. They are both time series forecasting models, but they are not the same in terms of how difficult they are or how they can be used.

5.1 Result ARIMA Model

Automated time series forecasting uses the ARIMA approach widely, where A stands for autoregressive, I for integrated, and MA for moving average. Annual rainfall an example of an ARIMA model can be predicted following the direction given by figure 9.

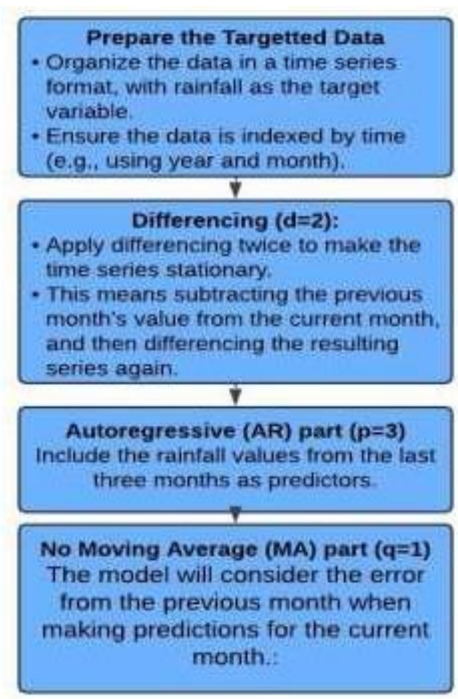

Fig 9: Arima Model fitted component values with explanation

Parameters Interpretation:

There are 3 parameters and there interpretation for experiment is:

- Since, p=3: This means the model includes three lagged terms in the autoregressive part. The model uses the rainfall data from the past three months.
- Since, d=2: This means the time series is difference twice to achieve stationarity. This means subtracting the previous month's value from the current month, and then differencing the resulting series again.
- Since, q=1: This means there are no moving average terms. (The model uses not just the past rainfall data of previous months (that's the autoregressive part), but also the past errors to predict future rainfall.) The model will consider the error from the previous month when making predictions for the current month.

In this study Fig 10 illustrates the findings of orange data mining and machine learning computer program in testing an ARIMA (3, 2,1) model for predicting the yearly rainfall. Two distinct methods are used to evaluate the ARIMA (3,2,1) model: in-sample evaluation and cross-validation: basic. The values of p, d, and q in the model are set up as p=3, implying that it uses the last three time steps in autocorrelation, d=2 to make the series stationary, and q=1 that it considers the previous period's error. The dataset is split into twelve parts and twelve-way cross-validation is applied on the cross-validation part. Each of the parts is used once for validation and the rest for training the model to measure the model's performance on unseen data. The results from cross-validation this cross-validated ARIMA (3,2,1) model are as follows: The R^2 (Coefficient of Determination) is -. 427 approximately, mean Absolute Error or MAE of 16. 6, a BIC of 12528, an RMSE of 24. 55 with the total average error of 0. 728. R-squared and adjusted R-squared are indicating how much additional data is fitted and predicted by the model when applying the model to a bigger dataset. While the in-sample evaluation, on the other hand, provides information about conventional efficacy of the model and its correspondence with the training set of data. As for the examination of performance outcome using RMSE, it was determined to be at 24. 3, MAE of 13. 8, MAPE of 0. 716, POCID of 41. 5%, R^2 of -0. This result is based on AIC of 198, AIC of 12822, and BIC of 12848, the in-sample estimate of the ARIMA (3,2,1) model is given below.

The slightly reduced error metrics in the in-sample evaluation show a better fit to the training data compared to the cross-validated findings, even if the negative R^2 values in both evaluations suggest that the model might not be the best fit for the rainfall data as it performs not good than a simple mean model. All things considered, comparing these evaluations helps understand how the model works and whether it is over fitting, which helps guide future model improvements for better prediction accuracy.

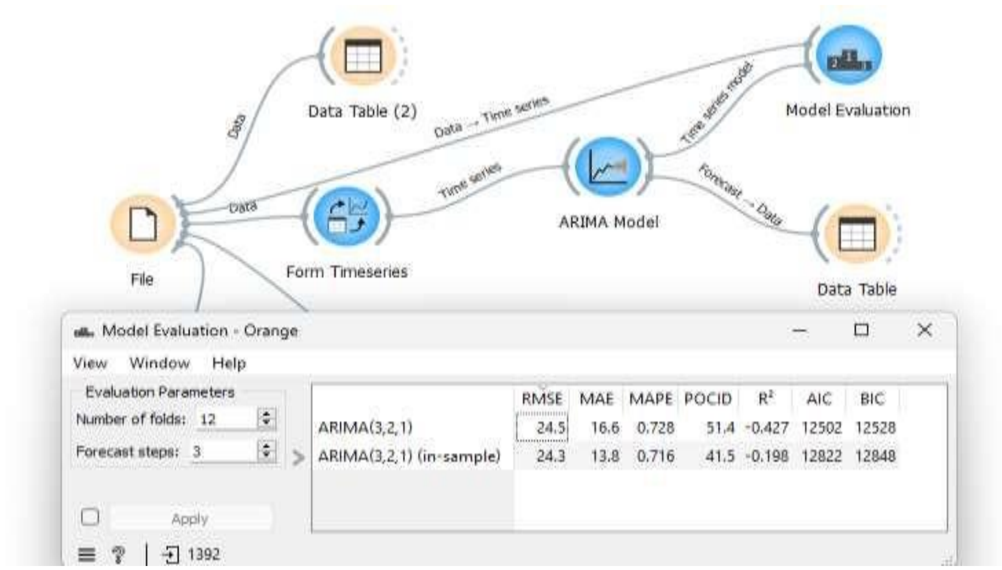

Fig 10. Result representation by using ARIMA Model

Fig 11 given line chart, with the Y-axis denoting rainfall in millimeters (MM) and the X-axis denoting time steps, illustrates the pattern of rainfall over ten time steps. The line labeled "Rainfall - (MM) (95% CI, High)" indicates an upward trend in rainfall, suggesting an increase over the observed period. This label implies that the line represents the upper bound of a 95% confidence interval for the predicted rainfall values, providing a high-end estimate of expected rainfall over time. By fitting this model, we can make predictions about future monthly rainfall based on past data.

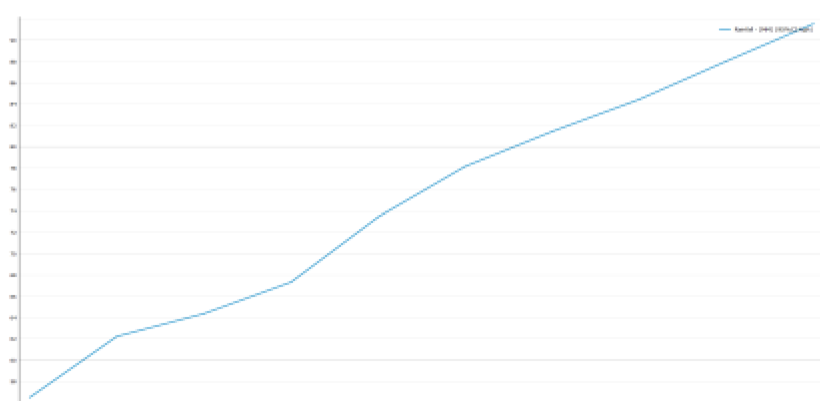

Fig 11. Representation of ARIMA Result

5.2 Result VAR Model

For this study orange data mining tool used model implemented through orange tool. The result presented in figure 12 is of VAR model, more specifically Vector Auto Regression, which was applied to a given time series data set. Lastly, it demonstrated how accurate the model is in identifying out of sample and in sample. The performance of the model can be rated rather high since the RMSE is 19. 4 and an MAE of 12. 7 for predictions outside samples over a forecast horizon of three time steps. This implication will mean that the level of error is acceptable and moderate. The MAPE is 0. 593, so the signs indicating the direction of time series changes are given by 62% of the model. Approximate 7% of the time and the POCID is 62. 7% which inform that it correctly obtains the absolute percentage error 59%. 3% of the time. The low coefficient of determination, R^2 with the value of 0. It is clear from the analysis in 030 that these prognostications contain much less.

However, the performance of the model evaluated on the in-sample data results in RMSE equals to 21. 3 and an MAE of 12. 3, which proved to be slightly higher than the MAPE of 0. 608, it can be inferred that the percentage errors of the in-sample forecasts are slightly higher than those of the out-of-sample forecasts. The POCID for the in-sample predictions is 0.50. 1%, meaning that out-of-sample identification is less accurate than out-of-sample concerning the identification of the direction of changes. This test yielded the in-sample R2 of 0. Again, 085 means that the model is able to explain more of the training data variability than the testing data. The AIC and BIC both differ from the above, its slightly lower value for in-sample which is 5. 682 for AIC and 5. 716 for BIC as compared to 5. 697 and 5. 732 in out of sample cases, hence the current model is somewhat better in fitting the training data. Therefore, according to the evaluation, the efficiency of applying the VAR model both on in-sample and on out-of-sample data is the same; however, other components of the efficiency of in-sample data are over-estimated, particularly regarding the accuracy of predictions.

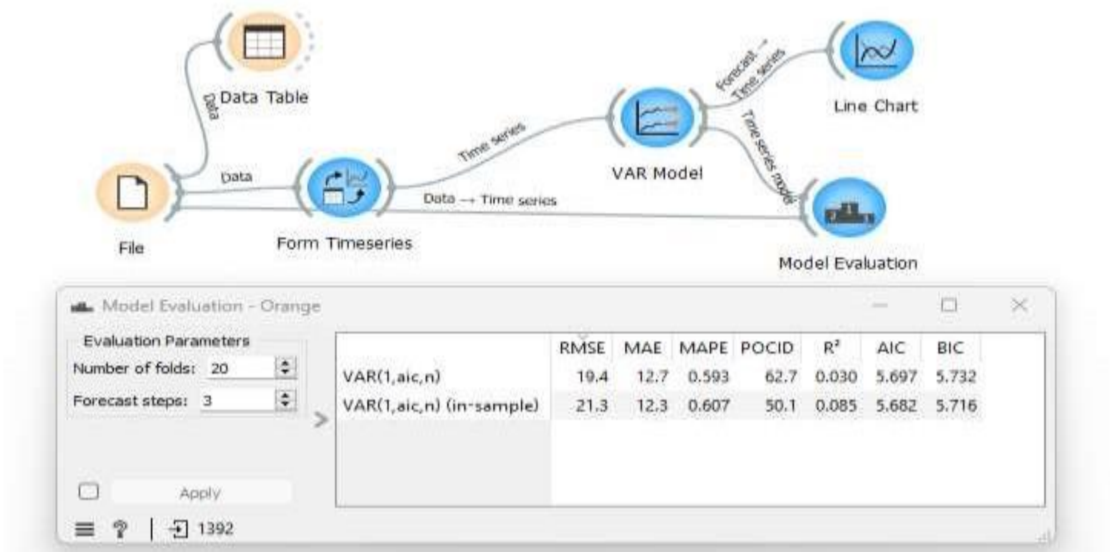

Fig 12. Result representation by using VAR Model

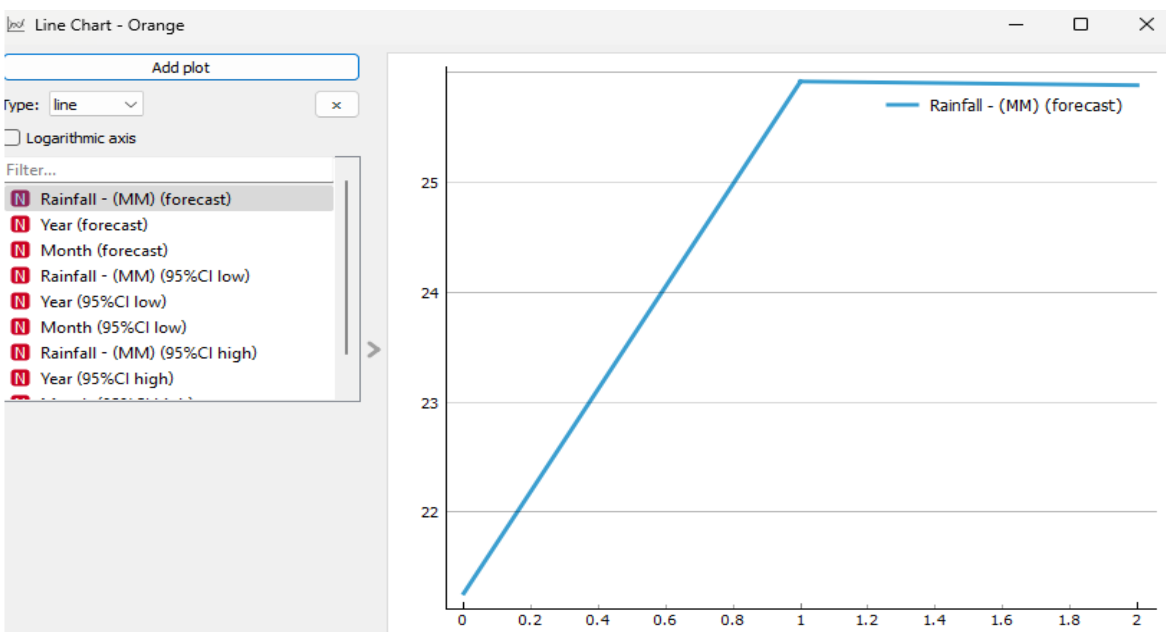

Fig 13. Result representation line chart by using VAR Model

The line chart in the fig 13 depicts the forecasted rainfall measured in millimeters over a specified forecast horizon. Initially, the forecast shows a significant increase in rainfall from approximately 22 mm at the starting point (time step 0) to slightly above 25 mm by time step 1. Following this sharp rise, the forecasted rainfall values remain constant at the higher level for the remainder of the forecast period shown (up to time step 2). This indicates an expected stabilization in rainfall at this elevated level after the initial increase.

5.2.1 Optimizing AR Order by Criteria

In the concept of VAR which stands for Vector Auto Regression, “optimizing AR order by” is concerned with the process of choosing the right lag order for the model. Selecting the correct lag order is important since it contains a direct impact on the capacity of the model to estimate all essential dynamics and relations in the data more accurately. The figure offer many criteria for deciding the optimal AR order and each of these has a different way of achieving a better model fit as well as model complexity.

Different model selection optimization criteria are described in Table 4. While the BIC is better for simpler models to prevent over fitting, particularly in big datasets, the AIC is favored for guaranteeing a decent fit without severely penalizing complexity. In order to create a balanced approach, the HQ criterion balances the AIC and BIC show in fig 13 also shown different optimized AR order by Hannan Quinn(HQ), Final prediction error (FPE) and average of the above.

When it comes to reducing prediction mistakes in future forecasting, the FPE is perfect. An average method takes into account several factors to select a more balanced model.

Table 4: Optimizing criteria description

AIC	**BIC**	**HQ**	**FPE**	**Average**
(Akaike Information Criterion	(Bayesian Information Criterion):	(Hannan-Quinn Criterion):	(Final Prediction Error):	
Preferred to ensure a good fit without overly penalizing complexity. AIC may perform better because it	This parameter aim for a simpler model that avoids over fitting, especially in large datasets. BIC optimization	Useful when seeking a balance between AIC and BIC.	Ideal when focusing on minimizing prediction errors in future forecasting	This parameter is Good if want a more balanced approach that considers multiple aspects.

penalizes complexity less severely.	parameter also often yields better results by favoring simpler models due to its stronger penalty on the number of parameters.

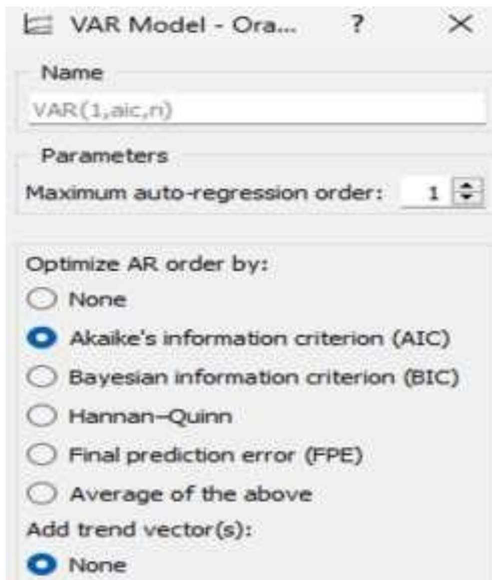

Fig 13: *Optimizing AR Order by Criteria*

5.3 Result comparison table

The table 5 highlights the performance metrics for both uni-variate and multi variate models, showing that the time series forecasting VAR (vector autoregressive)model generally outperforms or better than the time series forecasting ARIMA model for the obtained dataset and evaluation criteria.

Table 5: Result comparison ARIMA and VAR

Metric	VAR Model	ARIMA Model
RMSE	19.4	24.5
MAE	12.7	16.6
MAPE	0.593	0.728
POCID	62.7%	51.4%
R^2	0.030	-0.427

AIC	5.697	12502
BIC	5.732	12528

The VAR model shows better accuracy and model fit than the ARIMA model based on lower RMSE, MAE, MAPE, and AIC,BIC values, as well as a higher POCID and positive R^2. Therefore the VAR model is more suitable for monthly and yearly rainfall prediction in this experimental study.

By comparing our findings of time series forecasting ARIMA and VAR model, it was observed that the multivariate VAR model provides better accuracy in forecasting the rainfall compared to the ARIMA model. The model can deal with way more interacting time series features and therefore it can capture many more intricate patterns and dependency relations in data which are the reasons why the current VAR model does better. The results indicate that the proposed VAR model has higher R2 values and lower RMSE, MAE and MAPE when compared to ARIMA model. For the purpose of validating the model's ability to predict future rainfalls, these measurements reveal that VAR has better fit and provides a better insight than that from the previous model. While applied to univariate series, the univariate ARIMA model was slightly less capable than VAR in capturing the multi-dimensional structures involved in rainfall patterns. The enhanced accuracy of the latter is, therefore, its ability to take additional parameters, features and to model the interactions between them.

5.3.1 Limitations ARIMA and VAR model

As for the methodology employed in this empirical study, the following techniques were used: VAR and ARIMA. To capture temporal dependencies for a set of time series which all occurs in a single time point, the most basic and useful time series forecasting model is the ARIMA which combines autoregressive term (AR), differencing term (I), and moving average term (MA). Var is a model with more than one time variable and is more of an extension of the multivariate autoregressive model that is related to other time series. Its fundamental feature is data differencing, which makes it possible to use it for the analysis of non-stationary data or series with such characteristics as trends and seasonality factors. As has been noted earlier, however, VAR models are not as focused on individual time series as ARIMA, yet it is indeed possible to describe not just the correlation between multiple parameters of the given system but also how these parameters may interact – which can give more data about how one or another factor may affect the others..

However, VAR models, they handle several time series at once, have their own complexities. Especially when it comes to computation of large systems with a numerous number of parameters, these approaches require much data and computational resources. While the VAR models are efficient when it comes to capturing complex patterns and interdependencies among the series, the same models assume that interactions between the variables are linear, which could result in failure to capture drastic changes to the values or non-linear processes. Furthermore, the considered specification of VAR models raises the likelihood of model complexity thereby elevating the probability of overfitting which is a case where a model perfectly performs well in an old sample but poorly in new circumstances. This is even more so especially where data is scarce as the scholarly papers analyzed in this study have pointed out.

5.3.2 Practical Implementation in Real-World Agricultural Settings

There are several critical steps when implementing univariate ARIMA and multivariate VAR models in real-world agricultural contexts, as follows: In the first instance, the most important parameters to measure are rainfall and its frequency together with other relevant parameters such as wind, temperature maximum, minimum and humidity. When they are coupled with real time meteorological data, these provide optimum forecast models with a solid grounding. ARIMA and VAR models can then be integrated to agricultural decision support systems. Given that the VAR model includes multiple climatic factors and recognizes complex relationships, it excels at projection over the long term, while the ARIMA model is suitable for short-term forecasts when linear trends are apparent. These projections are very useful since they contain data required in crop forecasting, irrigation regimes and insurance strategies.

For instance rainfall, if predicted accurate and well, will assist the irrigators to properly harness water resources hence enabling them to expand the benefits of rainfall to boost crop yield. Another important aspect of the implementation practice is to enhance the understanding of the agricultural stakeholders about how to use and apply the model outputs. The approaches make the advises feasible to be implemented by farmers in agricultural operations by making them easy to apply and giving consistent suggestions.

6. Evaluation Criteria

Following are the evaluation criteria mathematically used in this study.

6.1. Root mean square error (RMSE)

Finds the average error size between the values that were noticed and those that were expected. Before combining, the mistakes are squared, which makes the bigger mistakes in equation 5 more important.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (yi - \hat{y}i)^2} \tag{5}$$

6.2. Mean absolute Error (MAE)

Performance metric MAE find the average error size between the observed and predicted anticipated values without taking the direction of the errors into account. The mean of the absolute errors is what it is shown in equation 6.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |yi - \hat{y}i| \tag{6}$$

6.3. Mean Absolute Percentage Error (MAPE)

Performance metric MAPE Calculates or find the mean absolute percentage error in between the observed or actual and predicted or projected values. Accuracy is expressed as a percentage shown in equation 7.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{yi - \hat{y}i}{yi} \right| \times 100\% \tag{7}$$

6.4. (Percentage of Correct Direction (POCID)

Calculates the proportion of the time that the expected direction—either up or down—matches the true direction shown in equation 8.

$$POCID = \frac{1}{n} \sum_{i=1}^{n} \{1, \quad if\ (yi - \hat{y}i)(yi + 1 - \hat{y}i + 1) > 0\ 0, \quad otherwise \times 100 \tag{8}$$

6.5. Coefficient of Determination R²

R² or the coefficient of determination indicates how much of the variation in the dependent variable as a result of variations in the independent factors. From 0 to where it cannot even make a guess to 1 where it can finish the guess with accuracy. This is mathematically given in equation 9 above

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (yi - \hat{y}i)^2}{\sum_{i=1}^{n} (yi - \tilde{y})^2} \tag{9}$$

6.6 Akaike Information Criterion (AIC)

It can serve as a model selection criterion. It tests the model fit and at the same time incorporates the penalty on the basis of complexity to avoid over fitting. Mathematically formulated by: where, ∅ represents the rate of return on the capital assets of the bank.

$$AIC=2k - 2 \ln (L) \tag{10}$$

Where,

L is the model's randomness, and k is the model's total amount of factors.

6.7 Bayesian Information Criterion (BIC)

In accordance with AIC, but penalized more heavily for larger model sizes. Complexity and model fit are balanced. Mathematically in equation 11

.

$$BIC = k \ln(n) - 2\ln(L) \tag{11}$$

Where,

the variables n, k, and L represent the number of observations, parameters, and likelihood, respectively, in the model.

7. Conclusion

Pakistan's agriculture industry depends heavily on accurate rainfall forecasting since prompt forecasts have a big impact on crop yields and financial results. This study examines how well two well-known time series forecasting techniques, ARIMA and VAR, predict patterns of rainfall using 116 years' worth of mean annual rainfall data from 1901 to 2016.

Our results show that when it comes to rainfall prediction, the Box-Jenkins-based ARIMA model performs better than the VAR model. The univariate time series forecasting model known as ARIMA is a reliable tool in the rainfall prediction since it can capture the temporal dependencies in the univariate time series data. Nonetheless, in this case, it yielded comparatively lower results than ARIMA, which although is effective for dealing with multiple related time series. While in the future, the work may be extended to cover many more stations. The rainfall levels that are produced are supposed to help the farmers in selecting their crop. This study increases the accuracy of developing an agricultural calendar activity , provides data on policy recommendations for food and disaster security, and helps mitigate the effects of climate change through improved rainfall prediction.

8. Future Work Recommendations:

1. Future research should focus on incorporating spatial data and advanced spatial modeling techniques to capture spatial dependencies in rainfall data more accurately.
2. Investigating hybrid models that combine ARIMA or VAR with deep learning approaches could further enhance prediction accuracy by capturing complex temporal and spatial patterns for rainfall forecasting.

3. In order to localize models to the specific farming conditions and climatic factors, undertake research appropriate to a specific area. Different regions may have different rain patterns and also interact with climates in distinct ways; this in a way calls for different methods of forecasting for every place.

Reference

[1] A. Geetha and G. M. Nasira, "Time-series modeling and forecasting: Modeling of rainfall prediction using ARIMA model," *Int. J. Soc. Syst. Sci.*, vol. 8, no. 4, pp. 361–372, 2016.

[2] "Time Series Model Evaluation," *Orange Data Mining*, accessed Mar. 2025.

[3] A. Rajab et al., "Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh," *Water*, vol. 15, no. 22, p. 3970, 2023.

[4] C. Song and X. Chen, "Performance comparison of machine learning models for annual precipitation prediction using different decomposition methods," *Remote Sens.*, vol. 13, no. 5, p. 1018, 2021.

[5] H. F. A. Baig, N. Islam, and A. Alim, "Exploring Machine Learning and Deep Learning Approaches for Disaster Prediction and Management: A Survey of Different Approaches for Disaster prediction and management through tweets," *Pak. J. Eng. Technol. Sci.*, vol. 11, no. 1, pp. 45–73, 2023.

[6] N. Ponnamperuma and L. Rajapakse, "Comparison of time series forecast models for rainfall and drought prediction," in *Proc. 2021 Moratuwa Eng. Res. Conf. (MERCon)*, pp. 626–631, IEEE, Jul. 2021.

[7] A. Rusyana, N. Tatsara, R. Balqis, and S. Rahmi, "Application of clustering and VARIMA for rainfall prediction," in *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 796, no. 1, p. 012063, Mar. 2020.

[8] T. Dimri, S. Ahmad, and M. Sharif, "Time series analysis of climate variables using seasonal ARIMA approach," *J. Earth Syst. Sci.*, vol. 129, pp. 1–16, 2020.

[9] M. Sidiq, "Forecasting rainfall with time series model," in *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 407, no. 1, p. 012154, 2018.

[10] H. Farman and S. Ahmed, "Novel framework for efficient detection of QRS morphology for the cardiac arrhythmia classification," *J. Comput. Biomed. Inform.*, vol. 5, no. 2, pp. 12–20, 2023.

[11] H. F. A. Baig, N. Islam, and A. Alim, "Exploring Machine Learning and Deep Learning Approaches for Disaster Prediction and Management: A Survey of Different Approaches for Disaster prediction and management through tweets," *Pak. J. Eng. Technol. Sci.*, vol. 11, no. 1, pp. 45–73, 2023.

[12] A. H. Khan et al., "Spatial Correlation Module for Classification of Multi-Label Ocular Diseases Using Color Fundus Images," *Comput. Mater. Continua*, vol. 76, no. 1, 2023.

[13] M. A. Mahmood, H. Malik, A. H. Khan, M. Adnan, and M. I. A. Khan, "Neural Network-Based Prediction of Potential Ribonucleic Acid Aptamers to Target Protein," *J. Comput. Biomed. Inform.*, vol. 4, no. 1, pp. 21–36, 2022.

[14] M. M. H. Khan, M. R. U. Mustafa, M. S. Hossain, S. Shams, and A. D. Julius, "Short-Term and Long-Term Rainfall Forecasting Using ARIMA Model," *Int. J. Environ. Sci. Dev.*, vol. 14, no. 5, pp. 292–298, 2023.

[15] S. Swain, S. Nandi, and P. Patel, "Development of an ARIMA model for monthly rainfall forecasting over Khordha district, Odisha, India," in *Recent Findings in Intelligent Computing Techniques*, Springer, pp. 325–331, 2018.

[16] S. Swain, S. K. Mishra, A. Pandey, and D. Dayal, "A stochastic model-based monthly rainfall prediction over a large river basin," in *Sustainability of Water Resources: Impacts and Management*, Cham: Springer Int. Publ., pp. 133–144, 2022.

[17] Y. A. Twumasi et al., "Time series analysis on selected rainfall stations data in Louisiana using ARIMA approach," *Open J. Stat.*, vol. 11, no. 5, pp. 655–672, 2021.

[18] D. Dolšak, N. Bezak, and M. Šraj, "Temporal characteristics of rainfall events under three climate types in Slovenia," *J. Hydrol.*, vol. 541, pp. 1395–1405, 2016.

[19] D. Peñarrocha, M. J. Estrela, and M. Millán, "Classification of daily rainfall patterns in a Mediterranean area with extreme intensity levels: The Valencia region," *Int. J. Climatol.*, vol. 22, no. 6, pp. 677–695, 2002.

[20] M. Faridzad et al., "Rainfall Frequency Analysis for Ungauged Regions Using Remotely Sensed Precipitation Information," *J. Hydrol.*, vol. 563, pp. 123–142, 2018.

[21] H. Li et al., "Non-Stationary Frequency Analysis of Annual Extreme Rainfall Volume and Intensity Using Archimedean Copulas: A Case Study in Eastern China," *J. Hydrol.*, vol. 571, pp. 114–131, 2019.

[22] M. S. Rahman and A. R. M. T. Islam, "Are Precipitation Concentration and Intensity Changing in Bangladesh Overtimes? Analysis of the Possible Causes of Changes in Precipitation Systems," *Sci. Total Environ.*, vol. 690, pp. 370–387, 2019.

[23] N. Vivekanandan, "Evaluation of Estimators of Probability Distributions for Frequency Analysis of Rainfall and River Flow Data," *Evaluation*, vol. 4, pp. 69–75, 2020.

[24] X. Gong and M. B. Richman, "On the Application of Cluster Analysis to Growing Season Precipitation Data in North America East of the Rockies," *J. Climate*, vol. 8, pp. 897–931, 1995.

[25] M. F. Salviano, J. D. Groppo, and G. Q. Pellegrino, "Trends Analysis of Precipitation and Temperature Data in Brazil," *Rev. Bras. Meteorol.*, vol. 31, pp. 64–73, 2016.

[26] X. Tan, T. Y. Gan, and D. Shao, "Wavelet Analysis of Precipitation Extremes over Canadian Ecoregions and Teleconnections to Large-Scale Climate Anomalies," *J. Geophys. Res.: Atmos.*, vol. 121, pp. 14469–14486, 2016.

[27] A. Y. Barrera-Animas et al., "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting," *Mach. Learn. Appl.*, vol. 7, p. 100204, 2022.

[28] T. Rosita and R. D. Estuningsih, "Vector Autoregressive (VAR) for Rainfall Prediction," *Int. J. Eng. Manag. Res.*, vol. 8, no. 2, pp. 96–102, 2018.

[29] S. Bora and A. Hazarika, "Rainfall time series forecasting using ARIMA model," in *Proc. 2023 Int. Conf. Artif. Intell. Appl. (ICAIA)*, IEEE, Apr. 2023.

[30] M. Khashei, M. Bijari, and S. R. Hejazi, "Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting," *Soft Comput.*, vol. 16, pp. 1091–1105, 2012.

[31] U. Ashwini, K. Kalaivani, K. Ulagapriya, and A. Saritha, "Time series analysis based Tamilnadu monsoon rainfall prediction using seasonal ARIMA," in *Proc. 2021 6th Int. Conf. Inventive Comput. Technol. (ICICT)*, IEEE, Jan. 2021, pp. 1293–1297.

[32] A. S. Kumar et al., "Rainfall prediction using machine learning," in *Advancements in Climate and Smart Environment Technology*, IGI Global, 2024, pp. 100–113.

[33] S. Q. Dotse, I. Larbi, A. M. Limantol, and L. C. De Silva, "A review of the application of hybrid machine learning models to improve rainfall prediction," *Model. Earth Syst. Environ.*, vol. 10, no. 1, pp. 19–44, 2024.