Project report

On

ANALYSIS OF AGRICULTURE CROP YIELD PREDICTION USING STATISTICAL TECHNIQUES

submitted
In partial fulfilment of the requirement for the degree of

MASTER OF SCIENCES

in

APPLIED STATISTICS AND DATA ANALYTICS
by

SUHANA RASHEED

(Reg No:SM22AS021)

(2022-2024)

under the supervision of

VISMAYA VINCENT

DEPARTMENT OF MATHEMATICS AND STATISTICS
ST TERESA'S COLLEGE (AUTONOMOUS)
ERNAKULAM, KOCHI-682011
APRIL 2024

ST TERESA'S COLLEGE(AUTONOMOUS) ERNAKULAM

CERTIFICATE

This is to certify that the project report entitled **ANALYSIS OF AGRICULTURE CROP YIELD PREDICTION USING STATISTICAL TECHNIQUES** submitted by **SUHANA RASHEED** in partial fulfilment of requirements for the award of the **DEGREE OF MASTER OF SCIENCES** in **APPLIED STATISTICS AND DATA ANALYTICS** is a record of the work carried out under my guidance and supervision at St Teresa's college autonomous Ernakulam affiliated to Mahatma Gandhi University, Kottayam .No part of this work has been submitted for any other degree elsewhere.

Date: 29/04/2024.
Place: Ernakulam

VISMAYA VINCENT
Assistant professor
Department of Mathematics and statistics
St Teresa's (Autonomous) College Ernakulam

ST. TERESA'S COLLEGE (AUTONOMOUS) DEPARTMENT OF MATHEMATICS AND STATISTICS ERNAKULAM

NISHA OOMMEN
Assistant Professor &HOD
Department of Mathematics and statistics
St Teresa's (Autonomous) College, Ernakulam

External Examiner

1.....
CHINU JOSEPH
29/4/2024.

2.....
LAKSHMI SURESH
29/04/2024.

DECLARATION

I here by declare that the project work entitled **"ANALYSIS OF AGRICULTURE CROP YIELD PREDICTION USING STATISTICAL TECHNIQUES"** being submitted in partial fulfilment for the award of the **DEGREE OF MASTER OF SCIENCES** in **APPLIED STATISTICS AND DATA ANALYTICS** is the original work carried outby me under the guidance of **VISMAYA VINCENT**, Assistant professor, department of mathematics and statistics. It has not formed the part of any other project work/internship submitted for award of any degree or diploma, either in this or any other University.

Place: Ernakulam
Date: 29/04/2024

SUHANA RASHEED
SM22AS021

ACKNOWLEDGEMENTS

A venture can't be completely by themselves.I take this opportunity to gratefully acknowledge various people who acted as guides along the way.

The success of any work requires the blessings of the Lord Almighty. I thank myGod for aiding me in my travel to success.

I would like to express my deep gratitude to **NISHA OOMMEN**, Head of the Department Mathematics and Statistics, for her valuable guidance and for the support she rendered on me.

I would also like to thank my guide **VISMAYA VINCENT**, Assistant Professor who revised my ideas and helped me to bring it to the research paper.

I would also like to express my gratitude to the staff of the St Teresa's Autonomous College Ernakulam for the technical support they gave.

This thanks giving cannot be complete without mentioning my friends and parents who gave the mental strength which I almost lost in between the journey.

Ernakulam

SUHANA RASHEED

Date: 29/04/2024

SM22AS021

ABSTRACT

In this paper, ideal time series model for forecasting the crop yield is empirically examined. Forecasting crop yield are crucial since they help to establish expectation for the future yield of the crop. Variables like temperature, rainfall, Ph, humidity, nitrogen, phosphorus, potassium values of seven crop are include in the dataset. In this paper analysis of rainfall and temperature is taken in account. All the variable has taken the monthly value so it is monthly data.

By using seasonal differencing make the data has stationary. The data show seasonal behaviour. Seasonal ARIMA and Holt-winters models are fitted for the timeseries data and the best timeseries models are selected using the RMSE and MSE. Forecast of the data sets are made using selected seasonal ARIMA and Holt-winters exponential smoothing model forecast value of agriculture crop yield are predicted till the year 2023 and reported in the study.

Keywords: Time series, seasonal ARIMA, Holts-winter exponential smoothing, rainfall, temperature and forecasting.

ST.TERESA'S COLLEGE (AUTONOMOUS) ERNAKULAM

Certificate of Plagiarism Check for Dissertation

Author Name	SUHANA RASHEED
Course of Study	M.Sc. Applied Statistics & Data Analytics
Name of Guide	Ms. VISMAYA VINCENT
Department	Post Graduate Mathematics & Statistics
Acceptable Maximum Limit	20%
Submitted By	library@teresas.ac.in
Paper Title	ANALYSIS OF AGRICULTURE CROP YIELDPREDICTIONUSING STATISTICAL TECHNIQUES
Similarity	5% AI 9%
Paper ID	1685815
Submission Date	2024-04-23 13:28:09

Signature of Student

Signature of Guide

Checked By
College Librarian

ST.TERESA'S COLLEGE LIBRARY ERNAKULAM

* This report has been generated by DrillBit Anti-Plagiarism Software

TABLE OF CONTENTS

Chapter-1
INTRODUCTION

Agriculture is the one of the important sectors in the country with respect to the economic and the sustainability. In today's world as the importance and the need of agriculture was increasing with the increasein population. The production of agriculture has to be improved in orderto have a sustainable balance.

The sustainable production of the crop mainly depends on various factors such as environment, climate, soil, water, etc. With the help of Time-series the overall harvest quality and accuracy can be improved. The farmers usually perform a prior crop and yield prediction based on their experience in that specificlocation. They don't have enough knowledge about the soil nutrient content like nitrogen, potassium, and phosphorus in the land, so they will only prefer the prior or neighborhood or widely used crop in that location for their land. Without the rotation of the crop and without adding adequate nutrients to the soil, crop production is reduced thereby decreasing the profit of the farmers.

India is the country where we witnessfarmer suicides most frequently due to this lower crop production and lower income. The other challenges faced by the farmers in selecting the appropriate crops to cultivate are the factors such as climate, soil type, and availability of resources. Crop selection is a critical task in agriculture for identifying the best crops to grow on a specific plot of land based on a variety of variables, including weather, soil nutrients, soil type, and previous crop yield. Time series have proven to be effective tools in crop yield recommendation.

In this paper, we are using classification technique for crop recommendation. It involve training a model on a dataset that contains temperature,humidity, rainfall and soil properties like pH

value, and soil nutrient content as the input parameters. In this paper take the only rainfall and temperature as variables. The model then learns and identifies the patterns in the dataset and predicts the crop that will probably produce the best yield in that specific land.

Chapter-2

REVIEW OF LITERATURE

This chapter presents a review of literature related to timeseries analysis of crop yield data which are based on the published information. A literature review is the writing process of summarizing, synthesizing and critiquing the literature found as a results of literature search.

Ghadge et al. (2018) Looking at the current situation faced by farmers In Maharashtra, we have observed that there is an increase in suicide rate over the years. The reasons behind this includes weather conditions, debt, family issues and frequent change in Indian government norms. Sometimes farmers are not aware about the crop which suits their soil quality, soil nutrients and soil composition. The work proposes to help farmers check the soil quality depending on the analysis done based on data mining approach.Thus the system focuses on checking the soil quality to predict the crop suitable for cultivation according to their soil type and maximize the crop yield with recommending appropriate fertilizer.

Champeneri (2018) highlights the adverse effects of climate change on India's agricultural crops, impacting their performance over the last two decades. Anticipating crop yield before harvest could aid policymakers and farmers in planning marketing and storage. This project aims to build an interactive prediction system prototype using a user-friendly web interface and the Random Forest algorithm. This algorithm, a potent supervised machine learning technique, constructs decision tree ensembles for reliable classification and regression. By addressing climatic challenges and leveraging data mining, this initiative offers solution for predicting crop yield and fostering agricultural economic growth in India.

Nigam et al. (2019) describe Agriculture is one of the major and the least paid occupation in India. Machine learning can bring a boom in the agriculture field by changing the income scenario through growing the optimum crop. This paper focuses on predicting the yield of the crop by applying various

machine learning techniques. The outcome of these techniques is compared on the basis of mean absolute error. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like temperature, rainfall, area, etc.

Medar et al. (2019) discuss that agriculture is the field which plays an important role in improving our countries economy. Agriculture is the one which gave birth to civilization. India is an agrarian country and its economy largely based upon crop productivity. Hence we can say that agriculture can be backbone of all business in our country. Selecting of every crop is very important in the agriculture planning. The selection of crops will depend upon the different parameters such as market price, production rate and the different government policies. Many changes are required in the agriculture field to improve changes in our Indian economy. We can improve agriculture by using machine learning techniques which are applied easily on farming sector. Along with all advances in the machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. The concept of this paper is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production.

Kale and Patil (2019) focus on India's agricultural significance, contributing around 23% to GDP and employing 59% of the workforce. Being the second-largest global agriculture producer, technology's role in enhancing yields is crucial. This research aims to predict crop yields using neural network regression. Diverse data encompassing crop cycles - summer, Kharif, rabi, autumn - sourced from an Indian government site is utilized. The dataset spanning 1998 to 2014 contains 240,000 records, with a focus on Maharashtra. Employing Python Pandas tools, the dataset is filtered and processed. A Multilayer Perceptron Neural Network is constructed; initially, RMSprop yields 45% accuracy, improved to 90% via layer adjustments, weight tuning, bias optimization, and changing to the Adam optimizer. This research pioneers a 3-layer ANN model for diverse crop yield predictions, utilizing the Rectified Linear activation unit (Relu) function.

Nishant et al. (2020) In India, we all know that Agriculture is the backbone of the country. This paper predicts the yield of almost all kinds of crops that are planted in India. This script makes novel by the usage of simple parameters like State, district, season, area and the user can predict the yield of the crop in which year he or she wants to. The paper uses advanced regression techniques like Kernel Ridge, Lasso and ENet algorithms to predict the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction.

Kavita and Mathur (2020) highlight India's agriculturally significant role in the economy and the challenge of feeding a growing population. Technology's potential to predict crop yields addresses this demand. Their study aims to forecast crop yield using machine learning techniques—Decision Tree, Linear Regression, Lasso Regression, and Ridge Regression—based on area, yield, production, and irrigation area. Decision Tree emerges as the most effective method.

Reddy and Kumar (2021) emphasize agriculture's vital role in India's economy, supporting over 50% of the population. Environmental variations and changing weather patterns pose significant risks to agriculture. Machine learning (ML) serves as a decision support tool for Crop Yield Prediction (CYP), aiding decisions on crop selection and cultivation strategies. This research presents a systematic review that extracts and synthesizes features utilized in CYP. Various artificial intelligence methods are employed to analyze crop yield prediction, yet Neural Networks exhibit limitations in reducing relative error and maintaining prediction efficiency. Supervised learning struggles with capturing nonlinear relationships between input and output variables, particularly in fruit grading or sorting. Numerous studies recommend agricultural development, aiming to enhance crop classification accuracy through models based on weather, crop diseases, and growth phases. This paper explores diverse ML techniques employed in crop yield estimation, providing a comprehensive analysis of their accuracy.

.Ranjani et al. (2021) emphasize agriculture's pivotal role in India's economy, sustaining over half the population. They apply machine learning techniques to predict crop production, considering parameters

like rainfall, crop type, and meteorological conditions. The versatile Random Forest algorithm, capable of classification and regression tasks, aids crop selection and minimizes yield losses amid environmental fluctuations. The paper recognizes weather and environmental challenges to agricultural sustainability. Machine learning's significance lies in its role as a decision-support tool for Crop Yield Prediction (CYP), aiding crop selection and cultivation strategies. The central aim of crop yield estimation is to enhance agricultural output, utilizing established models. Addressing a vital agricultural concern, yield prediction empowers farmers with insights for optimal crop decisions. The study's outcomes offer timely guidance to maximize agricultural yield, enabling informed decisions and future output forecasts.

Gupta et al. (2022) Food is one among the major necessities that is required for the survival of the humanity along with water, clothes. In Indian G.D.P agriculture plays a humongous role after which IT sector comes into play. The correct crop selection at the correct time of the year plays a vital role in getting good crop yield. In general, the methods used by farmers are based on the hype in the market and based on ancestral instincts which are proved to be not so effective.

Using the present system, we use the data considered which is preprocessed and based on this data certain models like Decision tree, Naïve Bayes, Support Vector Machine, Logistic Regression, Random Forest are trained, out of which Naïve Bayes is showing higher accuracy which helps in suggesting the crops. The crops suggested are based on the various factors like nitrogen, phosphorus, pH level, temperature and humidity. The data will be highly effective in helping the farmers to get good crop yield.

CHAPTER-3
BASIC DEFINITIONS

3.1 SARIMA

SARIMA (Seasonal Auto-Regressive Integrated Moving Average) is an extension of the **ARIMA (Autoregressive Integrated Moving Average)** model that incorporates ***seasonality*** in addition to the non-seasonal components. ARIMA models are widely used for time series analysis and forecasting, while SARIMA models are specifically designed to handle data with seasonal patterns.

SARIMA Model is represented as,

$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\text{Non-seasonal part of the model}} \quad \underbrace{(P, D, Q)_m}_{\text{Seasonal part of of the model}}$$

Where,

m = number of observations per year; **P** = Number of seasonal AR terms;

D = Number of seasonal differences; **Q** = Number of seasonal MA terms

We use uppercase notation for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model.

The seasonal component of SARIMA models adds the following three components:

1. **Seasonal Autoregressive (P)**: This component captures the relationship between the current value of the series and its past values, specifically at seasonal lags.

2. **Seasonal Integrated (D)**: Similar to the non-seasonal differencing, this component accounts for the differencing required to remove seasonality from the series.

3. **Seasonal Moving Average (Q)**: This component models the dependency between the current value and the residual errors of the previous predictions at seasonal lags.

3.2 HOLT'S WINTER EXPONENTIAL SMOOTHING

The Holt-Winters method is a very common time series forecasting procedure capable of including both trend and seasonality. The Holt-Winters method itself is a combination of 3 other much simpler components, all of which are smoothing methods:

- **Simple Exponential Smoothing (SES):** Simple exponential smoothing assumes that the time series has no change in level. Thus, it cannot be used with series that contain trend, seasonality, or both.

- **Holt's Exponential Smoothing (HES):** Holt's exponential smoothing is one step above simple exponential smoothing, as it allows the time series data to have a trend component. Holt's exponential smoothing is still incapable of cope with seasonal data.

- **Winter's Exponential Smoothing (WES):** Winter's exponential smoothing is an extension to Holt's exponential smoothing that finally allows for the inclusion of seasonality. Winter's exponential smoothing is what is referred to as the Holt-Winters method. The Holt-Winters method therefore is often referred to as triple exponential smoothing, as it is literally the combination of 3 smoothing methods built on top of each-other.

CHAPTER- 4

DATA AND METHODOLOGY

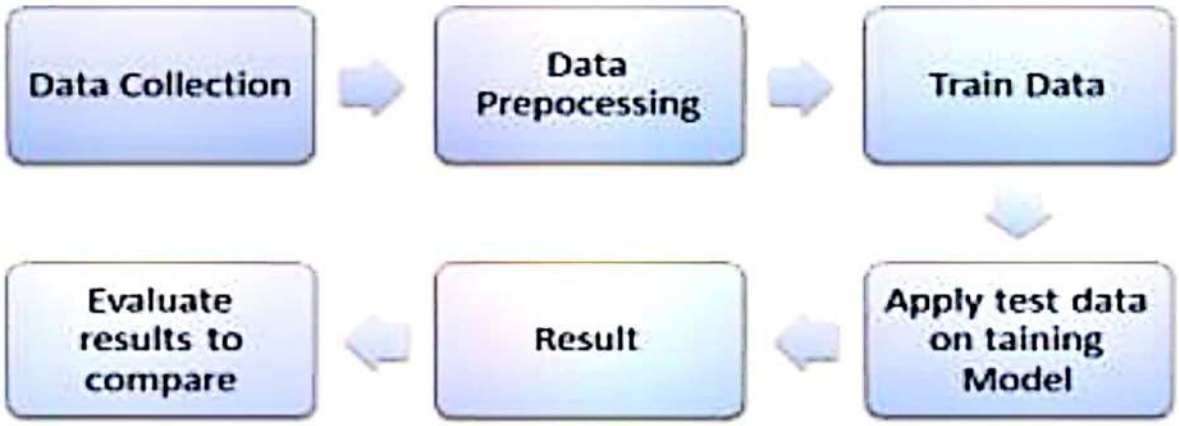


Fig 4.1 Outline of the work

4.1 DATA COLLECTION

The dataset used in this paper is extracted from Kaggle which provided the detailed environmental requirements for the 22 crops being cultivated in India. The dataset was created by augmenting datasets of rainfall, cli-mate, and fertilizer data available for India. It consists of 2201 entries,7attribute columns and 1 label column, which is well organized and ap- propriate for using both machine learning and deep learning algorithms.The data fields provided in this dataset are given below: N- the ratioof Nitrogen content P- the ratio of Phosphorous content K- the ratioof Potassium content Temperature- the temperature in degrees Celsius.
Humidity- relative humidity in %, pH value of the soil, rainfall of the areain mm.

4.2 DATA PRE-PROCESSING

In data preprocessing, Min Max Scaler from the Sklearn .preprocessing module is commonly used to scale features to a specific range, typically between 0 and 1. The process involves importing the necessary modules, instantiating the Min Max Scaler object, fitting the scaler on the training data to compute the minimum and maximum values, and transforming both the training and testing data using the learned scaling parameters. This scaling technique ensures that all features are on a similar scale, preventing dominance by features with larger values and enabling equal contribution from all features during model training.

4.3 MATERIALS AND METHODS

4.3.1 DATASET OVERVIEW

The dataset used in this paper is extracted from website and provide detailed environments for 2 crops cultivated in India.

1.Rice

2.Maize

It consists of 132 entries with 2 attribute columns (Temperature and Rainfall) and 1 label column as crop.

4.3.2 TOOLS FOR ANALYSIS

- SARIMA (Seasonal Autoregressive integrated moving average.)
- Holts-winters exponential smoothing model.

4.3.3 TOOLS FOR COMPARISON

- RMSE (Root Mean Square Error)
- MSE (Mean Square Error)

CHAPTER- 5
ANALYSIS

This chapter discusses a comparative study of time series modelling and forecasting of monthly rainfall and temperature that affect the agricultural crop yield using SARIMA and Holt's Winter forecasting procedure.

Python programming language is used in the model fitting procedure.

5.1 OBJECTIVES

To predict each crop that suitable for particular seasons

- To model and forecast crop yield in India using SARIMA models.
- To Forecast crop yield using holt-winters exponential smoothing technique.

5.2 ANALYSIS USING SARIMA MODEL

5.2.1 RICE PREDICTION ACCORDING TO TEMPERATURE

5.2.1.1 Plot the data

Firstly, display the data in the excel then plot this data to a graph. using python code. plotting the data helps to determine numerical information in a graphical format, to identify the trend and relationship within the data. Plotting the graph helps in gaining insights, making data-driven decisions.

Fig 5.1 depicts the time series plot of monthly rainfall data. It is visible that there is seasonality in the data.

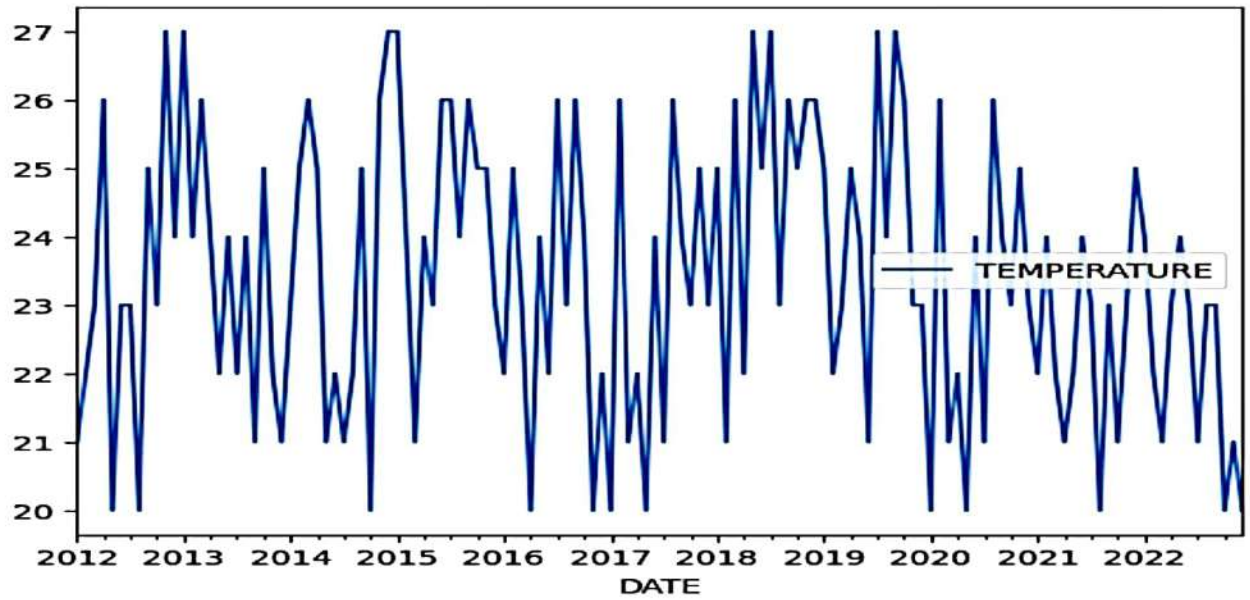


Fig 5.1 Time series plot

5.2.1.2 DECOMPOSITION GRAPH

It is a graph that visualize the time series dataset, it include components such as trend, seasonality and residuals.

Three main components are,

1.Trend: it explains the long-term movement or direction of the data, also represent overall growth or decline over time.

2. Seasonality: it shows where the data is fluctuating at a regular interval such as daily, weekly, monthly or yearly.

3.Residuals: The random fluctuations or irregularities in the trend seasonality of the data.

The decomposition graph that helps to analyze the each component separately also give an individual contribution to the behaviours of time series data.

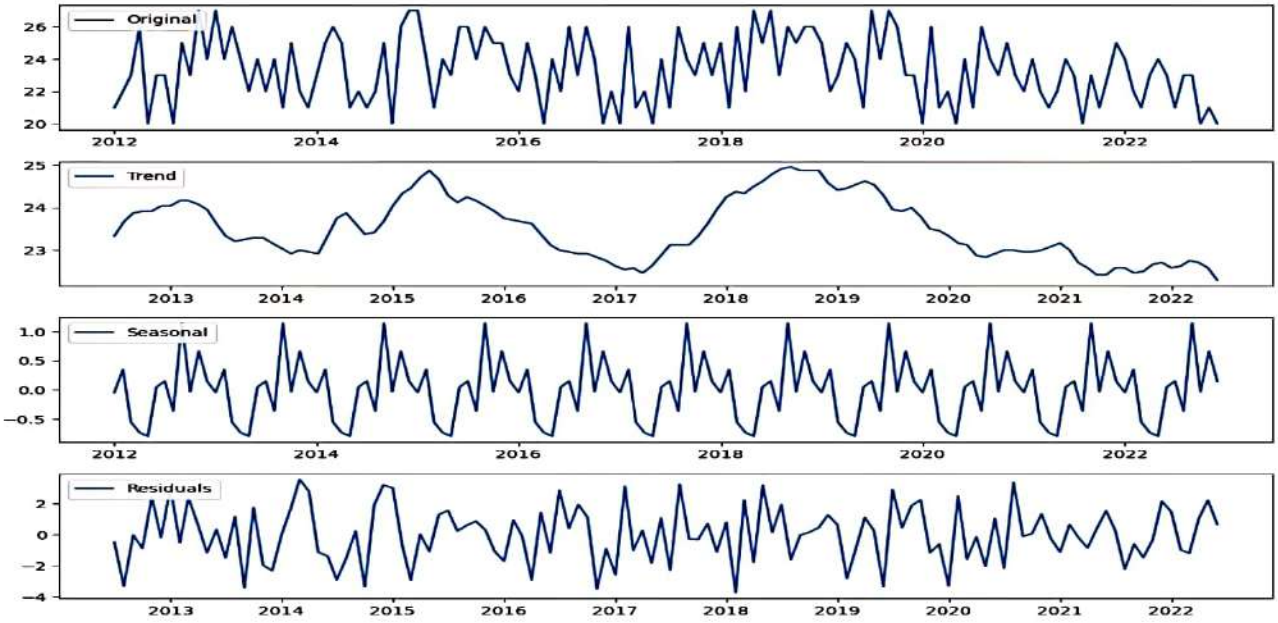
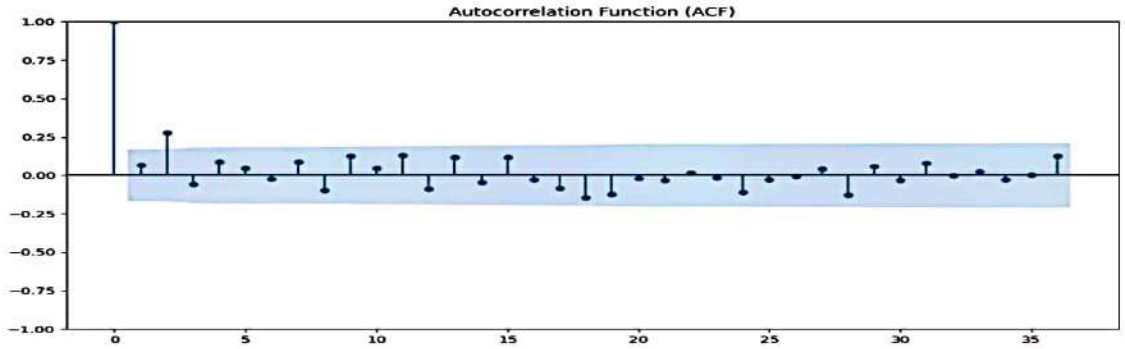


Fig **5.2** Decomposition graph

5.2.1.3 AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTION

Next step in analysing time series is to examine the autocorrelation function (ACF)and partial Autocorrelation function(PACF) . check the stationarity, plot the ACF and PACF values.

ACF and PACF plot is given below



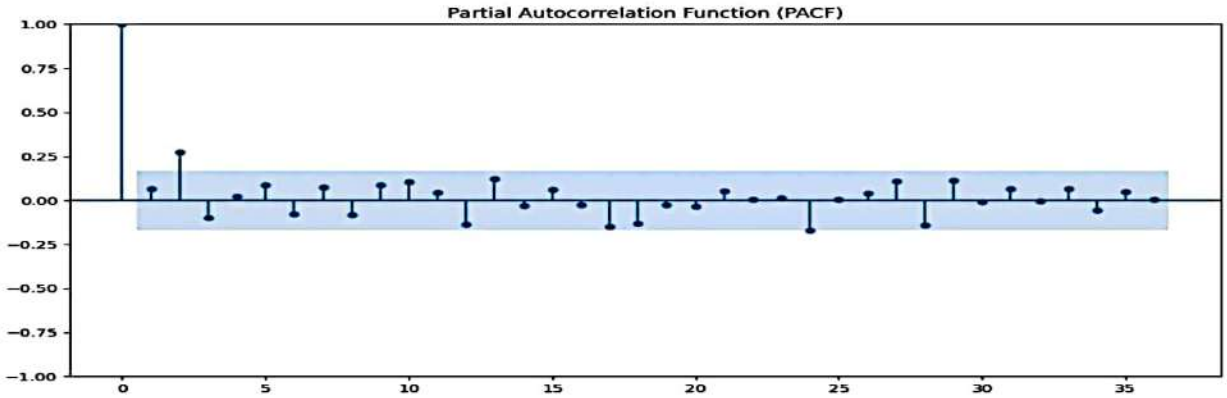


Fig 5.3 ACF and PACF

5.2.1.4 STATIONARITY USING AUGMENTED DICKEY-FULLER TEST

To test the time series data for stationarity using ADF test, follows a hypothesis testing approach. The null hypothesis H0 is given by,

H0: The data is non stationary.

The alternative hypothesis is given by

H1: The data is stationary

The outcome achieved, table

ADF test statistic	-5.661259567975618
Lag order	35
p-value	9.368446816502934e-07

Table- 5.1

The ADF test given the p-value is 9.368446816502934e-07, so fail to accept H0 and hence can conclude that the data is stationary.

5.2.1.5 SARIMA MODEL FOR RICE PREDICTION ACCORDING TEMPERATURE

Next step is to choose the best model for forecasting. Choose the best model from all possible model according to Akaike Information Criterion (AIC). The model the lowest AIC value is the best model. Thus,the possible time series models along with their corresponding AIC statistic,

p- The number of lags in the observation.

d-Order of differencing.

q- order of the moving average.

Values obtained through AIC can be used for forecasting. This process helps to save time and improve the accuracy of the forecasting models.

Sl.NO	**MODEL ARIMA (p,d,q) x (P,D,Q)**	**AIC**
1	ARIMA(0,1,0)x(0,1,0,12)12	474.39076347269656
2	ARIMA(0,1,1)x(1,0,0,12)12	510.1524582867008
3	ARIMA(0,1,1)x(1,0,1,12)12	501.53283163895856
4	ARIMA(0,1,1)x(1,1,0,12)12	520.05695387242743334

Table- 5.2

VALUE OBTAINED THROUGH AIC:

(0, 1, 0) x (0, 1, 0, 12)12 - AIC: 474.39076347269656

5.2.1.6 MODEL EVALUATION

Model evaluation includes assessing the errors or residuals that show difference between observed and predicted values from the model.

1.Standardized Residuals: It divides the standard deviations, it helps to determine how many standard deviations each residuals is from the expected values. A good model range from -2 to 2.

2.Histogram: it visually represents the distribution of the data. Histogram includes bell-shaped curve.it indicate that data is normally distributed.

3.Normal Q-Q plot (Quantile-Quantile plot): it compares the quantiles of the normal distribution. if the residuals satisfy the normal distribution Q-Q plot is in straight line otherwise it show some deviation.

4. Correlogram: it is auto-correlation plot that shows the correlation of residuals at different lags.

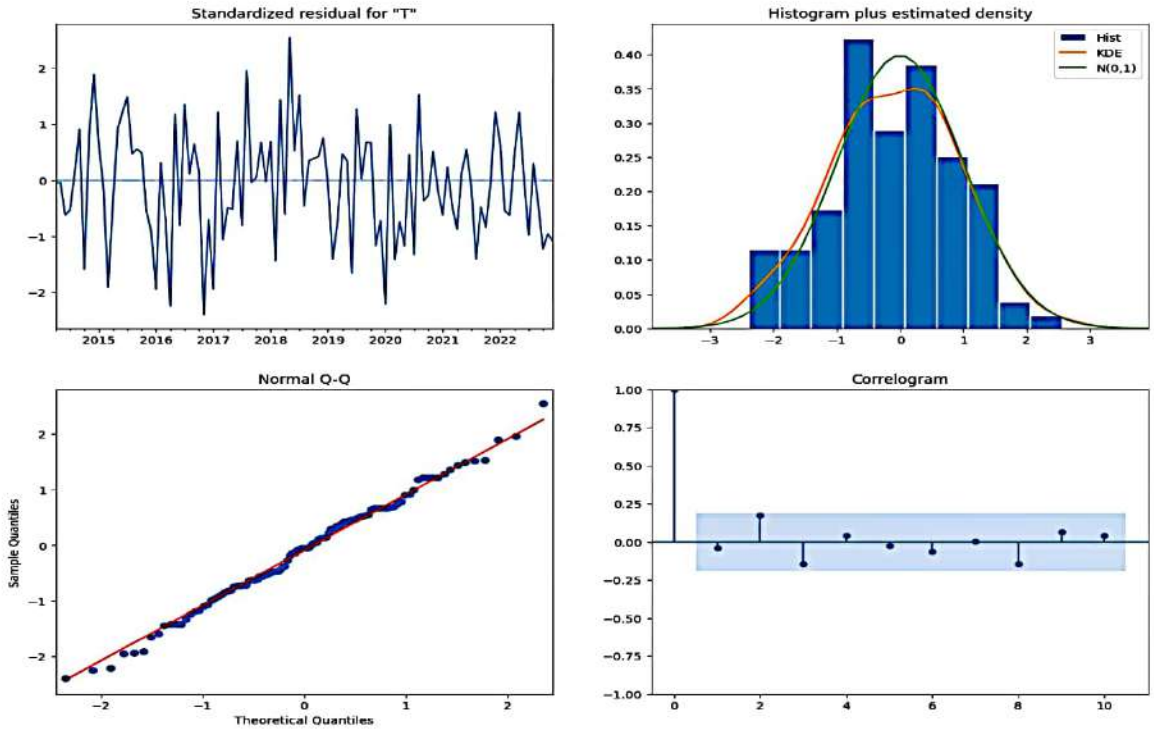


Fig 5.4 Model evaluation

5.2.1.7 ACTUAL V/S PREDICTED TEMPERATURE

Actual v/s predicted graph contains two lines, one represents actual values and other represents the predicted values. This type of graphs are commonly used in regression model. If two line comes close to each other gives better the model's performance in accurately predicting the variable of interest.

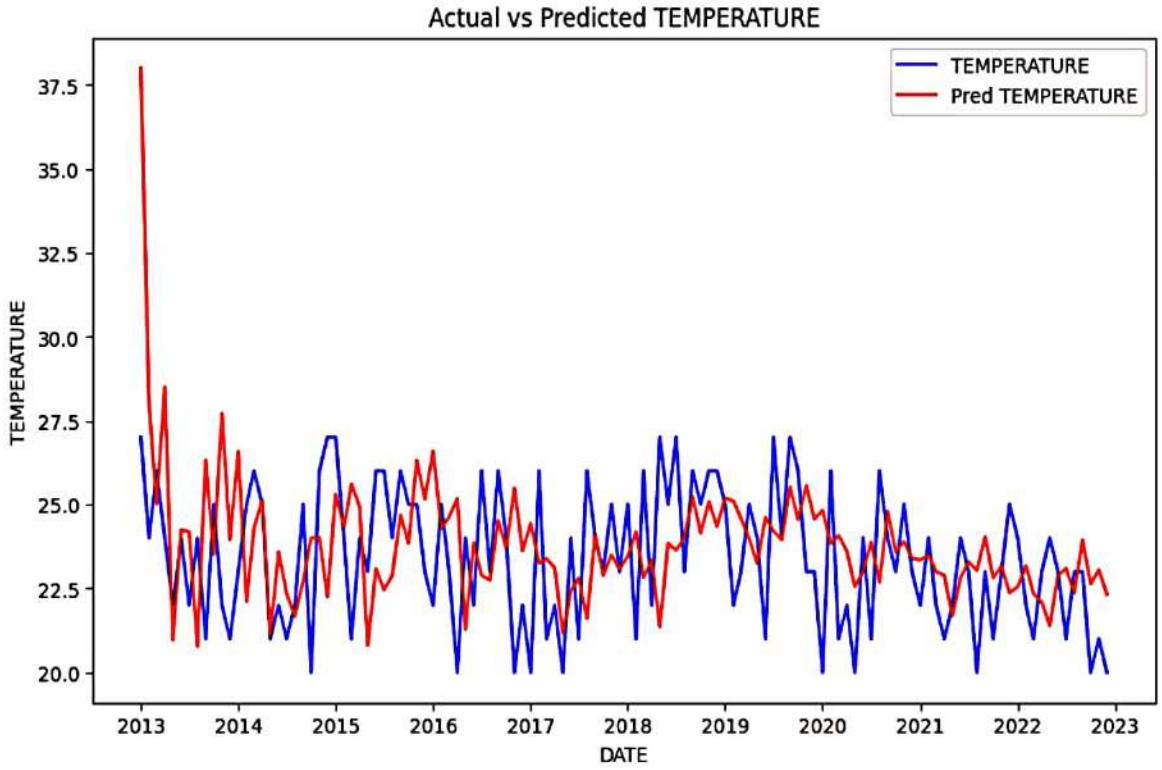


Fig 5.5 Actual v/s predicted temperature

5.2.1.8 FORCASTED VALUES

Forecasted values is an estimate of the future value based on the historical data,trend and predictive models . it helps to determine what happens in the future values and it give approximate similar type of results by

referring the previous data. Forecasting is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depend on the quality of the data, and the uncertainity related with the future events.

DATE	PREDICTED VALUE
2023-01-01	21.870013
2023-02-01	22.095774
2023-03-01	21.368502
2023-04-01	21.459411
2023-05-01	20.823048
2023-06-01	21.823047
2023-07-01	21.823047
2023-08-01	21.550320
2023-09-01	22.913956
2023-10-01	21.550320
2023-11-01	22.277593
2023-12-01	21.732136

Table-5.3

5.2.1.19 FORCASTED GRAPH

Forecasted graph displays the visualization of values over time . it combines with the historical data and the predictive values to estimate the future trends or patterns. Forecasting graphs is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depend on the quality of the data, and the uncertainity related with the future events.

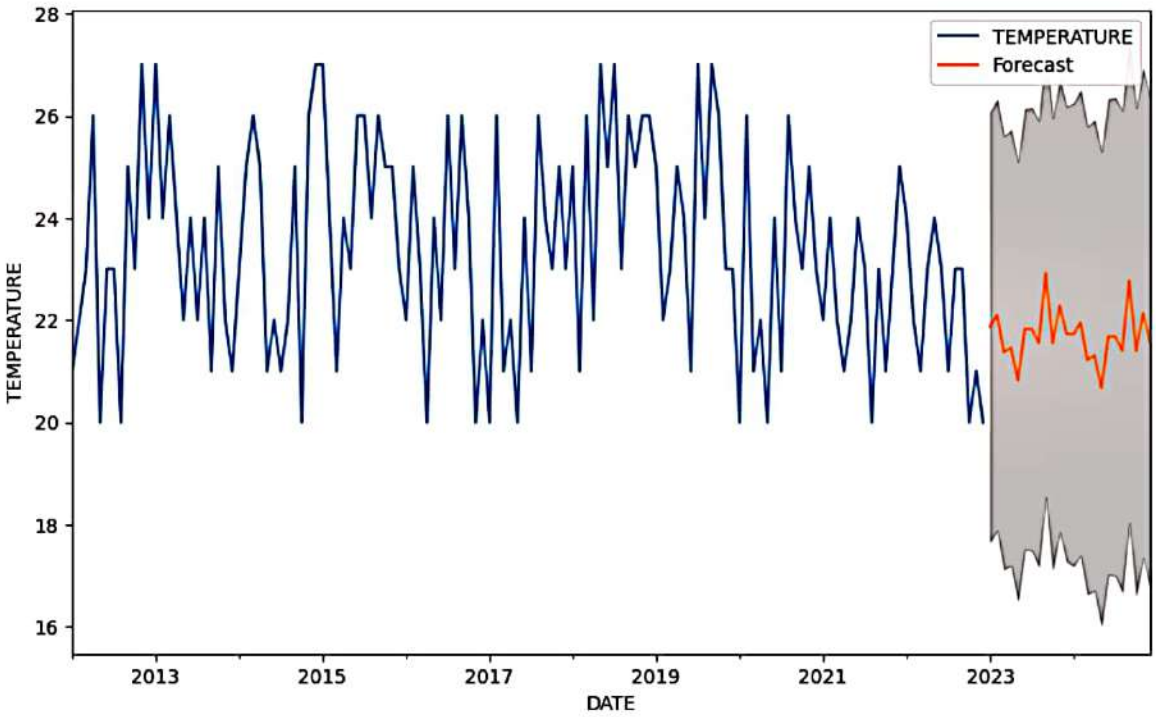


Fig 5.6 Forecasted graph

5.2.2 RICE PREDICTION ACCORDING TO RAINFALL

5.2.2.1 Plot the data

Firstly, display the data in the excel then plot this data to a graph. using python code. plotting the data helps to determine numerical information in a graphical format, to identify the trend and relationship within the data. Plotting the graph helps in gaining insights, making data-driven decisions.

Fig 5.6 depicts the time series plot of monthly rainfall data. It is visible that there is seasonality in the data.

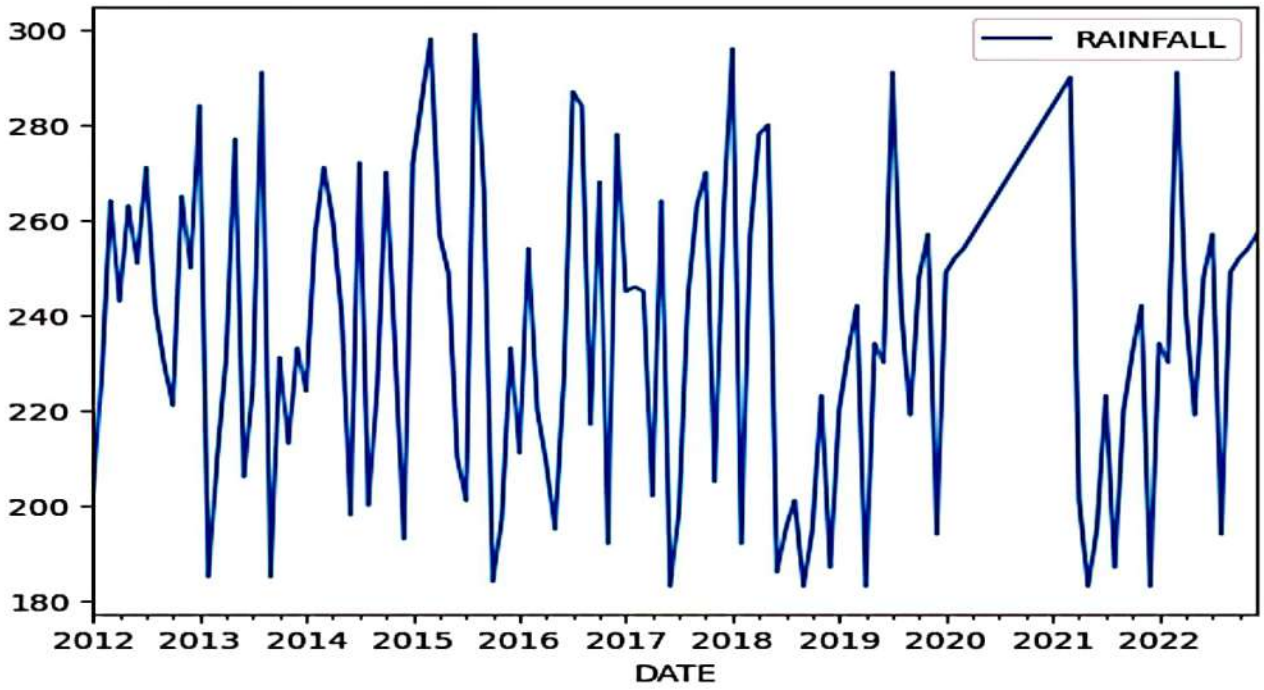


Fig 5.7 Time series plot

5.2.2.2 DECOMPOSITION GRAPH

It is a graph that visualize the time series dataset, it include components such as trend seasonality and residuals,

Three main components are,

1.Trend: it explains the long-term movement or direction of the data, also represent overall growth or decline over time.

2. Seasonality: it shows where the data is fluctuating at a regular interval such as daily, weekly, monthly or yearly.

3.Residuals: The random fluctuations or irregularities in the trend seasonality of the data.

The decomposition graph that helps to analyze the each component separately also give an individual contribution to the behaviour of time series data.

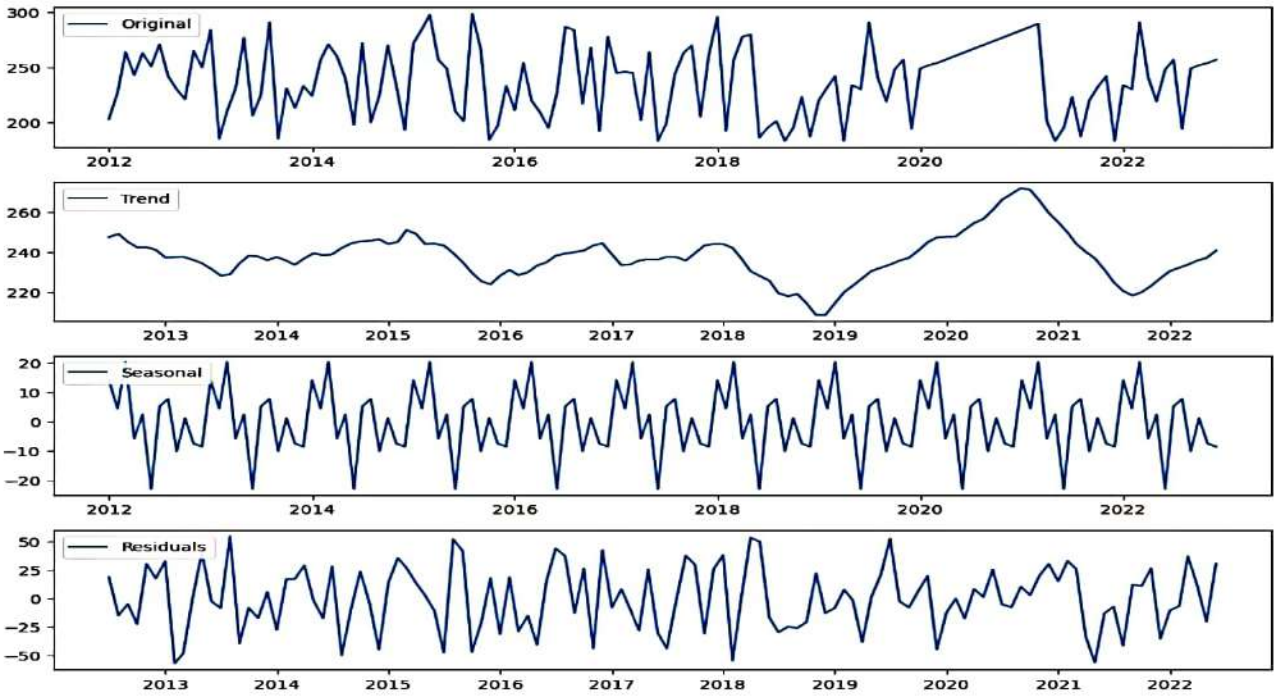


Fig 5.8 Decomposition graph

5.2.2.4 AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTION

Next step in analysing time series is to examine the autocorrelation function (ACF)and partial Autocorrelation function(PACF) . check the stationarity, plot the ACF and PACF values.

ACF and PACF plot is given below

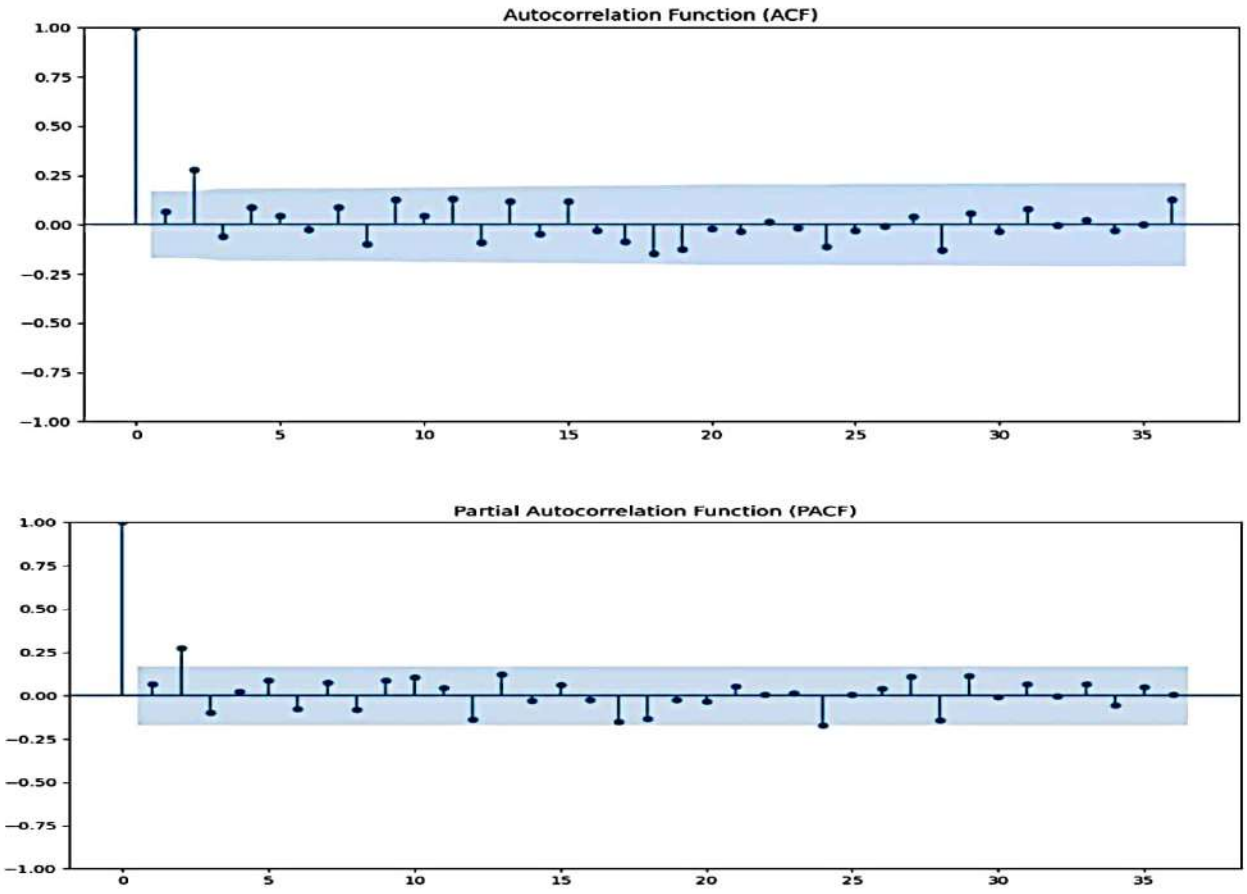


Fig 5.9 ACF and PACF

5.2.2.5 STATIONARITY USING AUGMENTED DICKEY-FULLER TEST

To test the time series data for stationarity using ADF test, follows a hypothesis testing approach. The null hypothesis H0 is given by,

H0: The data is non stationary.

The alternative hypothesis is given by

H1: The data is stationary

The outcome achieved, table

ADF test statistic	-5.661259567975618
Lag order	35
p-value	9.69946816502934e-07

Table-5.4

The ADF test given the p-value is 9.69946816502934e-07, so fail to accept H0 and hence can conclude that the data is stationary.

5.2.2.6 SARIMA MODEL FOR RICE PREDICTION ACCORDING TO RAINFALL

Next step is to choose the best model for forecasting. Choose the best model from all possible model according to Akaike Information Criterion (AIC). The model the lowest AIC value is the best model. Thus,the possible time series models along with their corresponding AIC statistic.

It is a statistical technique to automatically select the optimal parameters(p,d,q) for an ARIMA model.

p- The number of lags in the observation.

d-Order of differencing.

q- order of the moving average.

Values obtained through AUTO-ARIMA can be used for forecasting. This process helps to save time and improve the accuracy of the forecasting models

Sl.NO	MODEL ARIMA (p,d,q) x (P,D,Q)	AIC
1	ARIMA(1,1,1)x(1,0,0,12)12	1168.5924219575577
2	ARIMA(1,1,1)x(1,0,1,12)12	1162.0997682099037
3	ARIMA(1,1,1)x(1,1,0,12)12	1091.2967049123436
4	ARIMA(1,1,0)x(1,1,0,12)12	1057.5124579450933

Table-5.5

VALUE OBTAINED THROUGH AIC (1, 1, 0) x (1, 1, 0, 12)12 - AIC:1057.5124579450933

5.2.2.7 MODEL EVALUATION

Model evaluation includes assessing the errors or residuals that show difference between observed and predicted values from the model.

1.Standardized Residuals: It divides the standard deviation.it helps to determine how many standard deviations each residuals is from the expected values. A good model range from -2 to 2.

2.Histogram: it visually represents the distribution of the data histogram includes bell-shaped curve.it indicate that data is normally distributed.

3.Normal Q-Q plot (Quantile-Quantile plot): it compares the quantiles of the normal distribution. if the residuals satisfy the normal distribution Q-Q plot is in straight line otherwise it show some deviation.

4. Correlogram: it is a auto-correlation plot that shows the correlation of residuals at different lags.

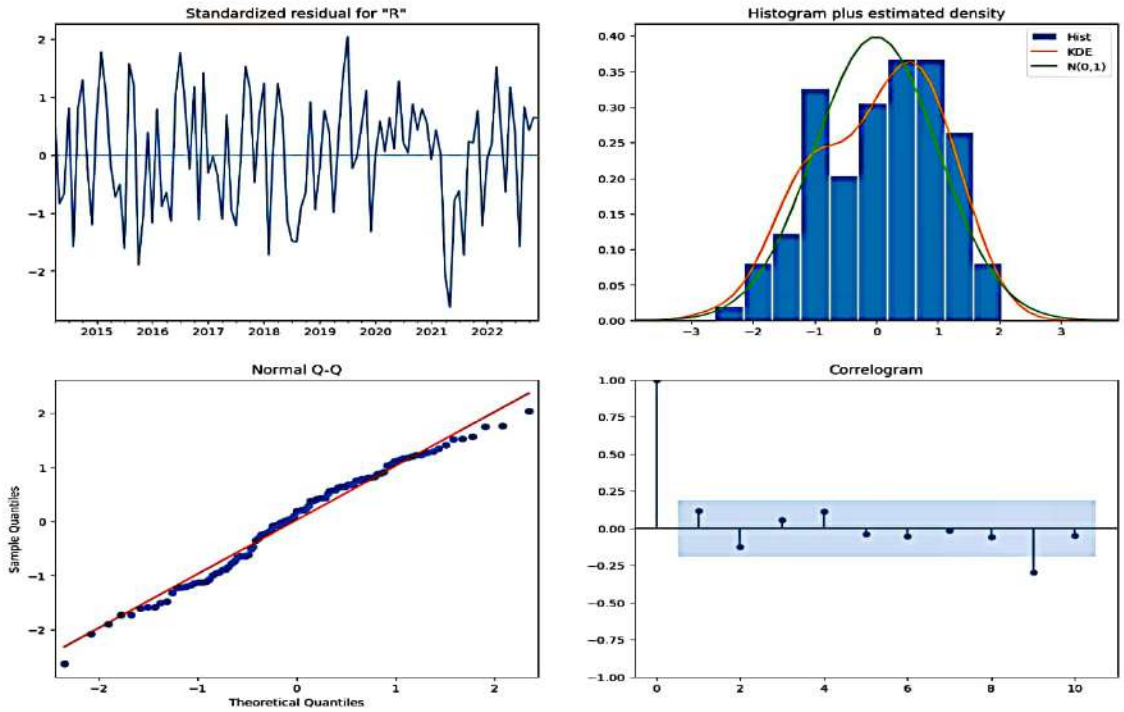


Fig 5.10 Model evaluation

5.2.2.8 ACTUAL V/S PREDICTED TEMPERATURE

Actual v/s predicted graph contains two lines, one represents actual values and other represents the predicted values. This type of graphs are commonly used in regression model. If two line comes close to each other gives better the model's performance in accurately predicting the variable of interest.

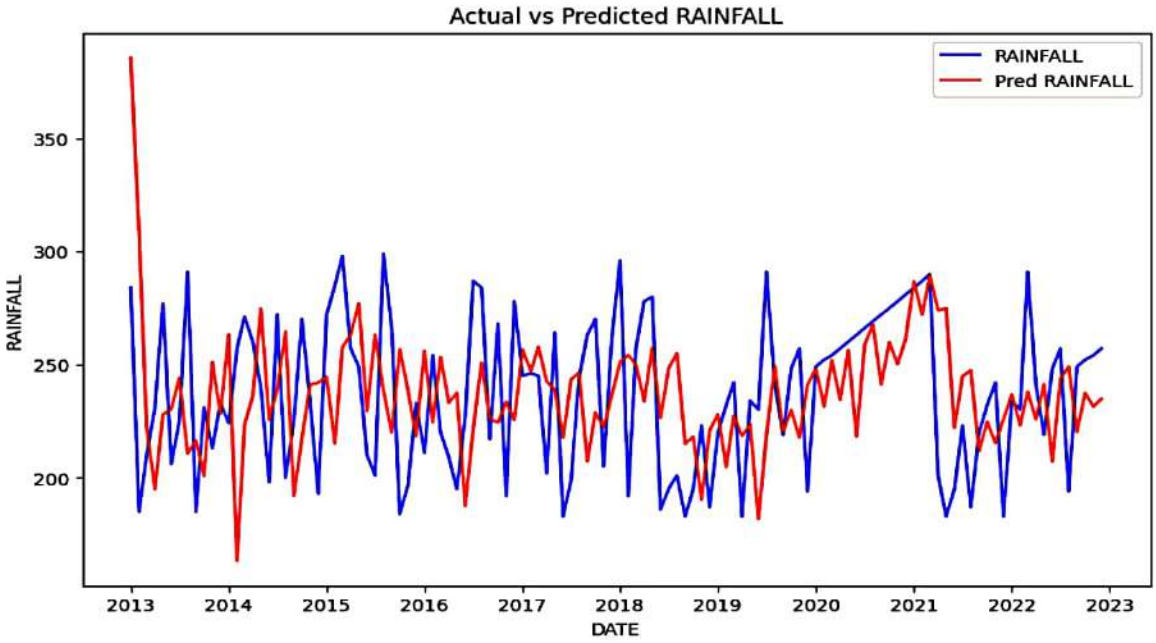


Fig 5.11 Actual v/s predicted temperature

5.2.2.9 FORECASTED VALUES

Forecasted values is an estimate of the future value based on the historical data,trend and predictive models . it helps to determine what happens in the future values and it give approximate similar type of results by referring the previous data. Forecasting is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depend on the quality of the data, and the uncertainity related with the future events.

DATE	PREDICTED VALUE
2023-01-01	261.873169
2023-02-01	250.009179
2023-03-01	267.735724
2023-04-01	242.284345
2023-05-01	251.554324
2023-06-01	227.285011
2023-07-01	253.645389
2023-08-01	250.464438
2023-09-01	239.283835
2023-10-01	249.918563
2023-11-01	242.010513
2023-12-01	241.186597

Table-5.6

5.2.2.10 FORCASTED GRAPH

Forecasted graph displays the visualization of values over time.it combines with the historical data and the predictive values to estimate the future trends or patterns. Forecasting graphs is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depend on the quality of the data, and the uncertainity related with the future events.

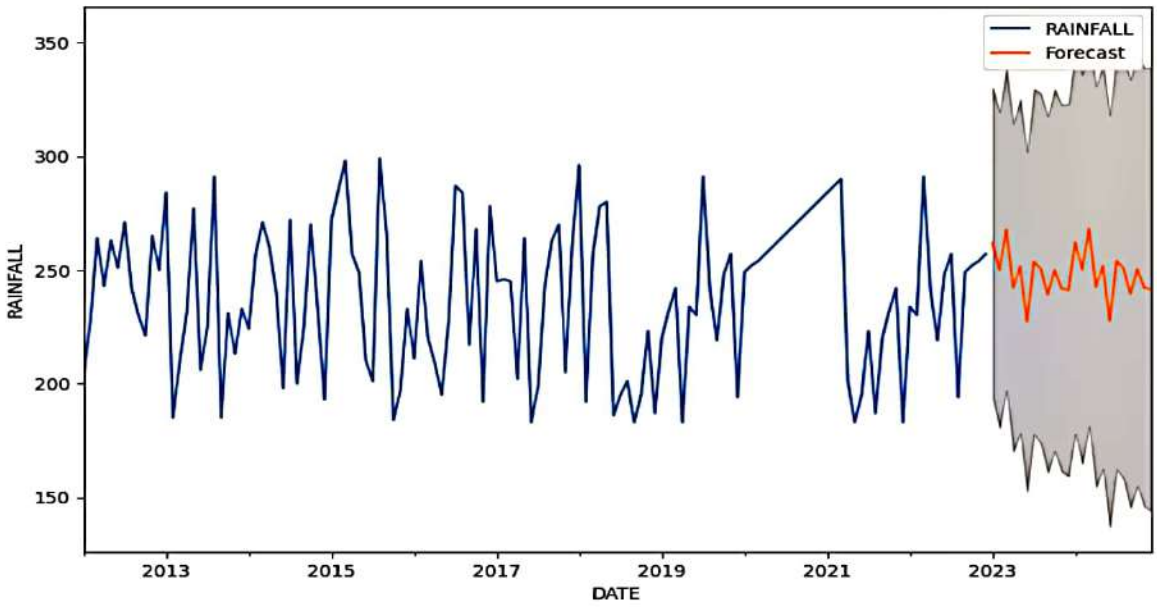


Fig 5.12 Forecasted graph

5.2.3 MAIZE PREDICTION ACCORDING TO TEMPERATURE

5.2.3.1 Plot the data

Firstly, display the data in the excel then plot this data to a graph. using python code. plotting the data helps to determine numerical information in a graphical format, to identify the trend and relationship within the data. Plotting the graph helps in gaining insights, making data-driven decisions.

Fig 5.15 depicts the time series plot of monthly rainfall data. It is visible that there is seasonality in the data.

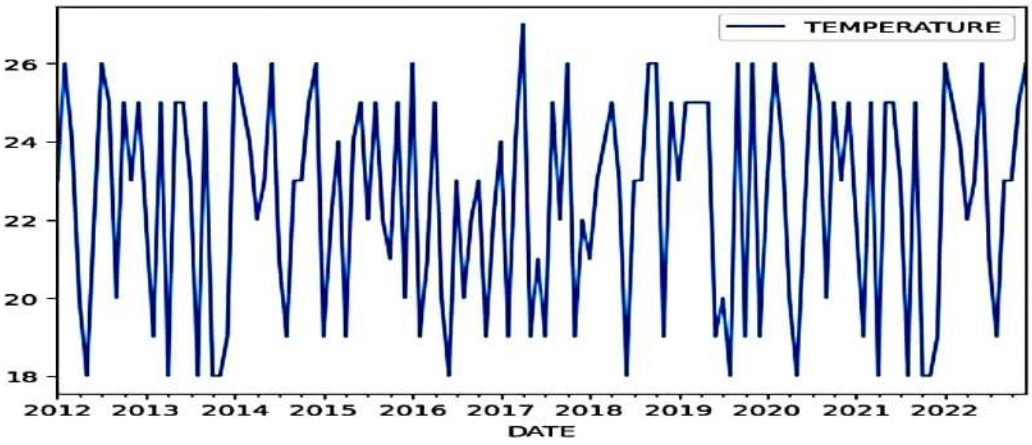


Fig 5.13 :Time series plot

5.2.3.2 DECOMPOSITION GRAPH

It is a graph that visualize the time series dataset, it include components such as trend seasonality and residuals,

Three main components are,

1.Trend: it explains the long-term movement or direction of the data, also represent overall growth or decline over time.

2. Seasonality: it shows where the data is fluctuating at a regular interval such as daily ,weekly, monthly or yearly.

3.Residuals: The random fluctuations or irregularities in the trend seasonality of the data.

The decomposition graph that helps to analyze the each component separately also give an individual contribution to the behaviour of time series data.

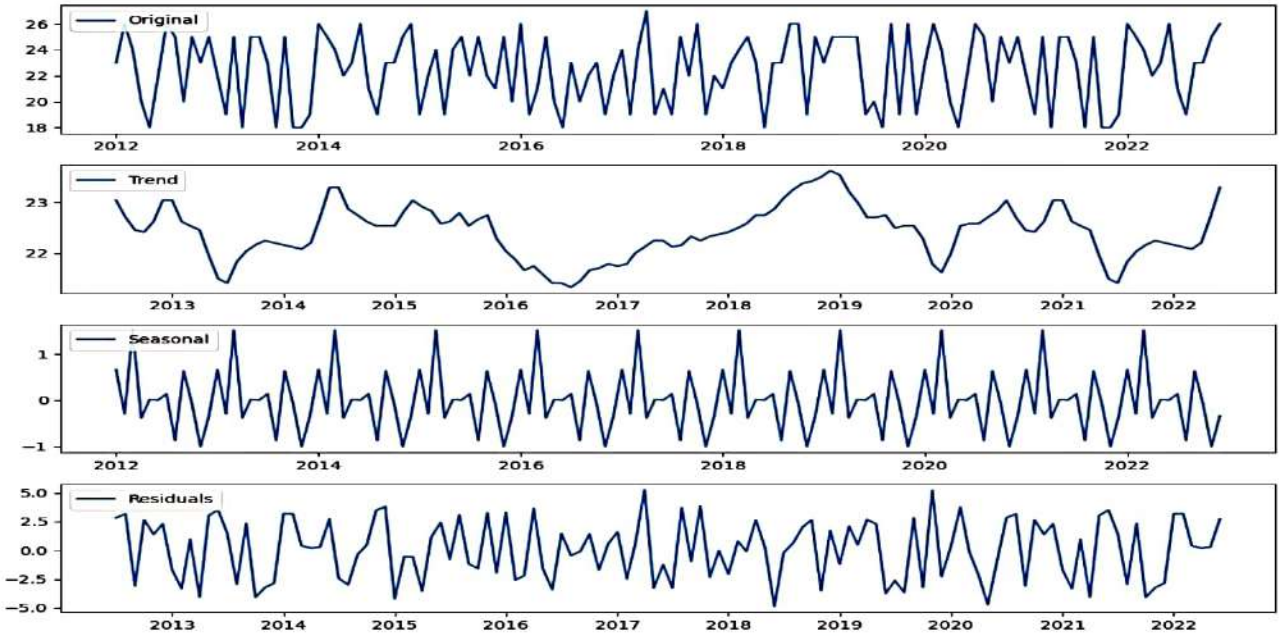
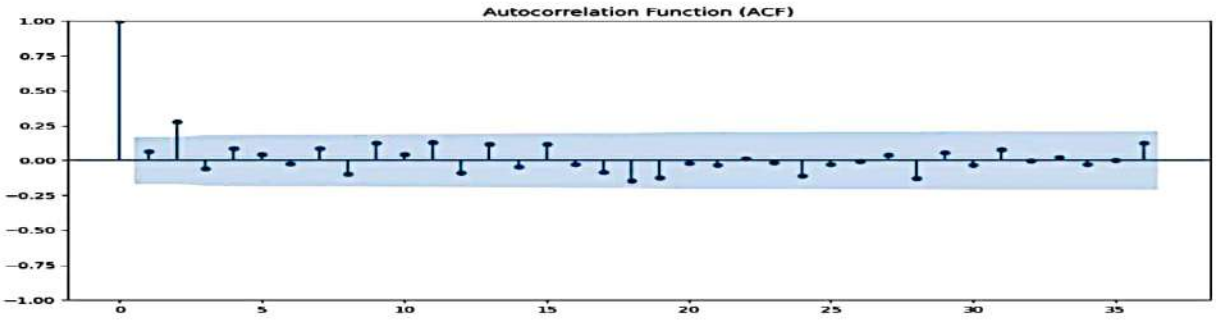


Fig 5.14 Decomposition graph

5.2.3.3 AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTION

Next step in analysing time series is to examine the autocorrelation function (ACF)and partial Autocorrelation function (PACF) . check the stationarity, plot the ACF and PACF values.

ACF and PACF plot is given below



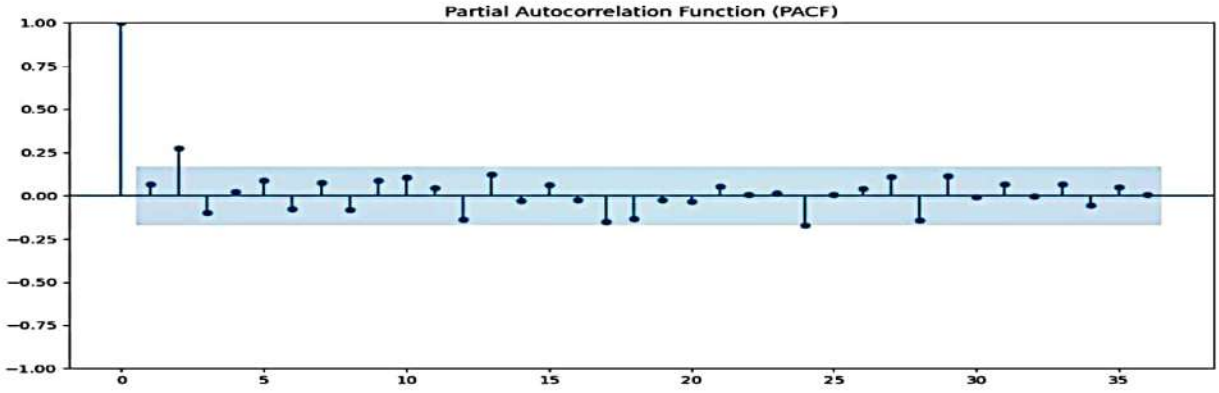


Fig 5.15 ACF and PACF

5.2.3.4 STATIONARITY USING AUGMENTED DICKEY-FULLER TEST

To test the time series data for stationarity using ADF test, follows a hypothesis testing approach. The null hypothesis H0 is given by,

H0: The data is non stationary.

The alternative hypothesis is given by

H1: The data is stationary

The outcome achieved, table

ADF test statistic	-5.56759567975618
Lag order	35
p-value	9.653446816502934e-07

Table-5.7

The ADF test given the p-value is 9.653446816502934e-07so fail to accept H0 and hence can conclude that the data is stationary.

5.2.3.5 SARIMA MODEL FOR MAIZE PREDICTION ACCORDING TEMPERATURE

Next step is to choose the best model for forecasting.Choose the best model from all possible model according to Akaike Information Criterion(AIC).The model the lowest AIC value is the best model. Thus the possible time series models along with their corresponding AIC statistic.

5.2.3.6 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

It is a statistical technique to automatically select the optimal parameters(p,d,q) for an ARIMA model.

p- The number of lag in the observation.

d-Order of differencing.

q- order of the moving average.

Values obtained through AUTO-ARIMA can be used for forecasting. This process helps to save time and improve the accuracy of the forecasting models

Sl.NO	MODEL ARIMA (p,d,q) x (P,D,Q)	AIC
1	ARIMA(0,1.0)x(0,1,0,12)12	532.8180737151155
2	ARIMA(0,0,1)x(1,0,0,12)12	685.4534480109485
3	ARIMA(0,0,1)x(1 0,1,12)12	610.3070975522243
4	ARIMA(0 0,1)x(1 1,0,12)12	567.2591377387527

Table-5.8

VALUE OBTAINED THROUGH AUTOARIMA:

ARIMA (0, 1, 0) x (0, 1, 0, 12)12 - AIC:532.8180737151155

5.2.3.7 MODEL EVALUATION

Model evaluation includes assessing the errors or residuals that show difference between observed and predicted values from the model.

1.Standardized Residuals: It divides the standard deviation,it helps to determine how many standard deviations each residuals is from the expected values. A good model range from -2 to 2.

2.Histogram: it visually represents the distribution of the data.histogram includes bell-shaped curve.it indicate that data is normally distributed.

3.Normal Q-Q plot (Quantile-Quantile plot): it compares the quantiles of the normal distribution. if the residuals satisfy the normal distribution Q-Q plot is in straight line otherwise it show some deviation.

4. Correlogram: it is a auto-correlation plot that shows the correlation of residuals at different lags.

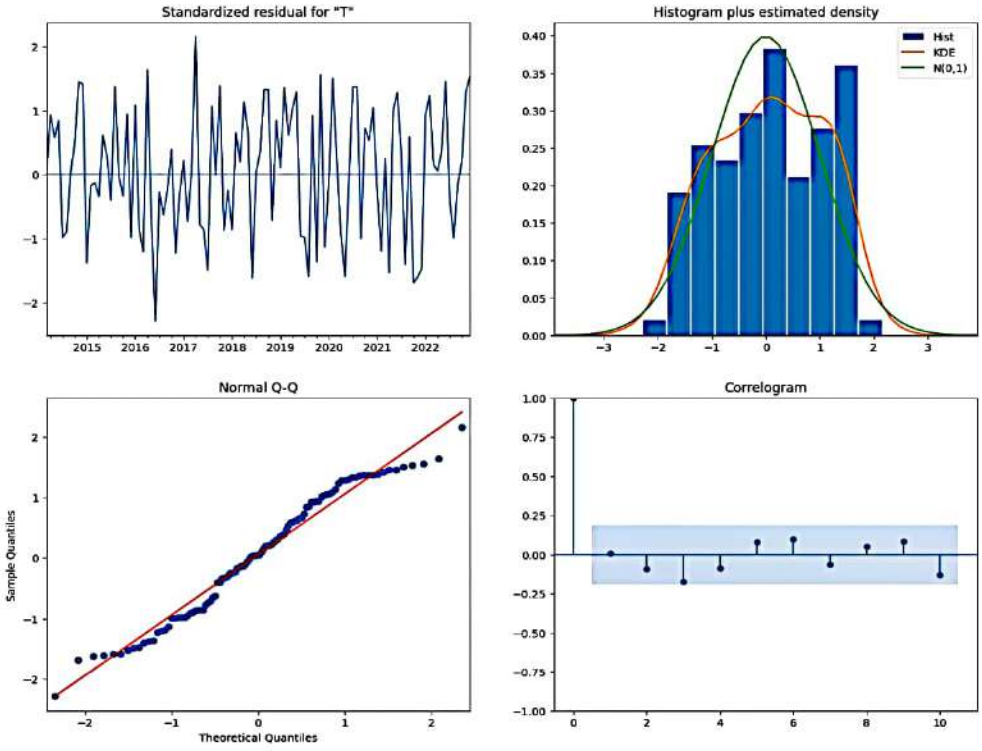


Fig 5.16 Model evaluation

5.2.3.8 ACTUAL V/S PREDICTED RAINFALL

Actual v/s predicted graph contains two lines, one represents actual values and other represents the predicted values. This type of graphs are commonly used in regression model. If two line comes close to each other gives better the model's performance in accurately predicting the variable of interest.

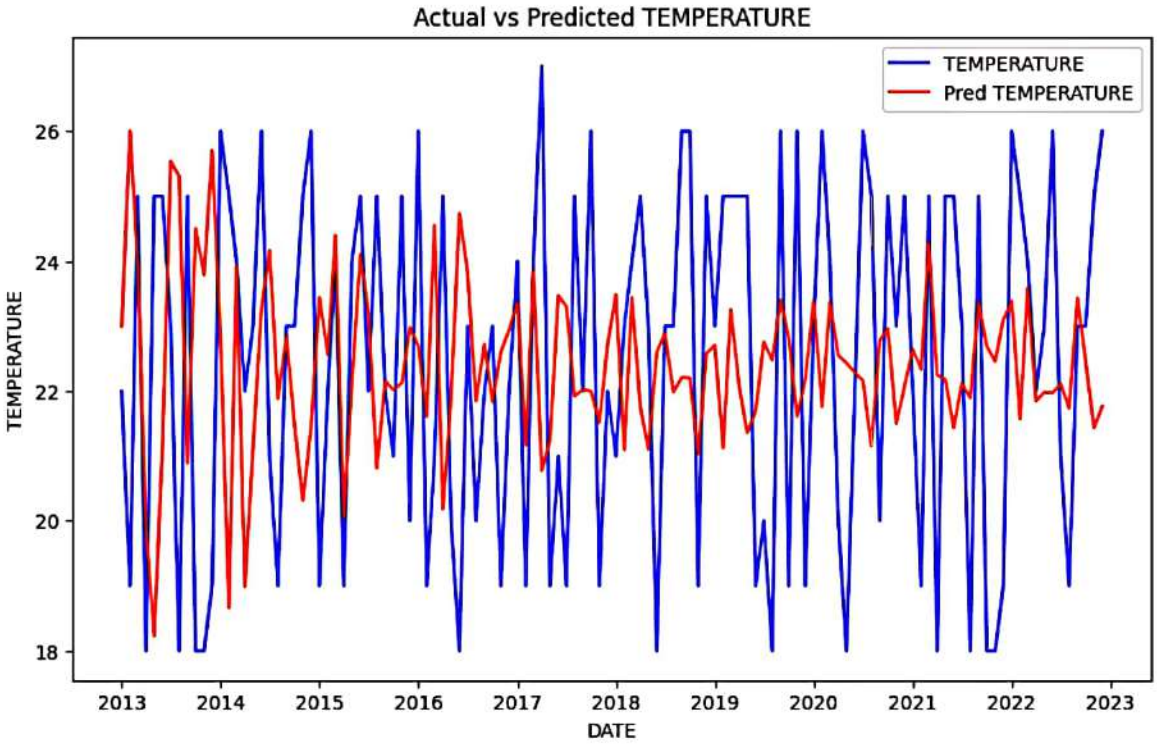


Fig 5.17 Actual v/s predicted temperature

5.2.3.9 FORCASTED VALUES

Forecasted values is an estimate of the future value based on the historical data,trend and predictive models . it helps to determine what happens in the future values and it give approximate similar type of results by referring the previous data. Forecasting is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depends on the quality of the data, and the

uncertainity related with the future events.

DATE	PREDICTED VALUE
2023-01-01	22.685905
2023-02-01	22.203551
2023-03-01	23.999994
2023-04-01	23.999994
2023-05-01	22.090904
2023-06-01	22.454540
2023-07-01	22.454540
2023-08-01	21.363631
2023-09-01	23.090903
2023-10-01	22.454540
2023-11-01	21.818177
2023-12-01	22.545448

Table-5.9

5.2.3.10 FORCASTED GRAPH

Forecasted graph displays the visualization of values over time. it combines with the historical data and the predictive values to estimate the future trends or patterns. Forecasting graphs is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depends on the quality of the data, and the uncertainity related with the future events

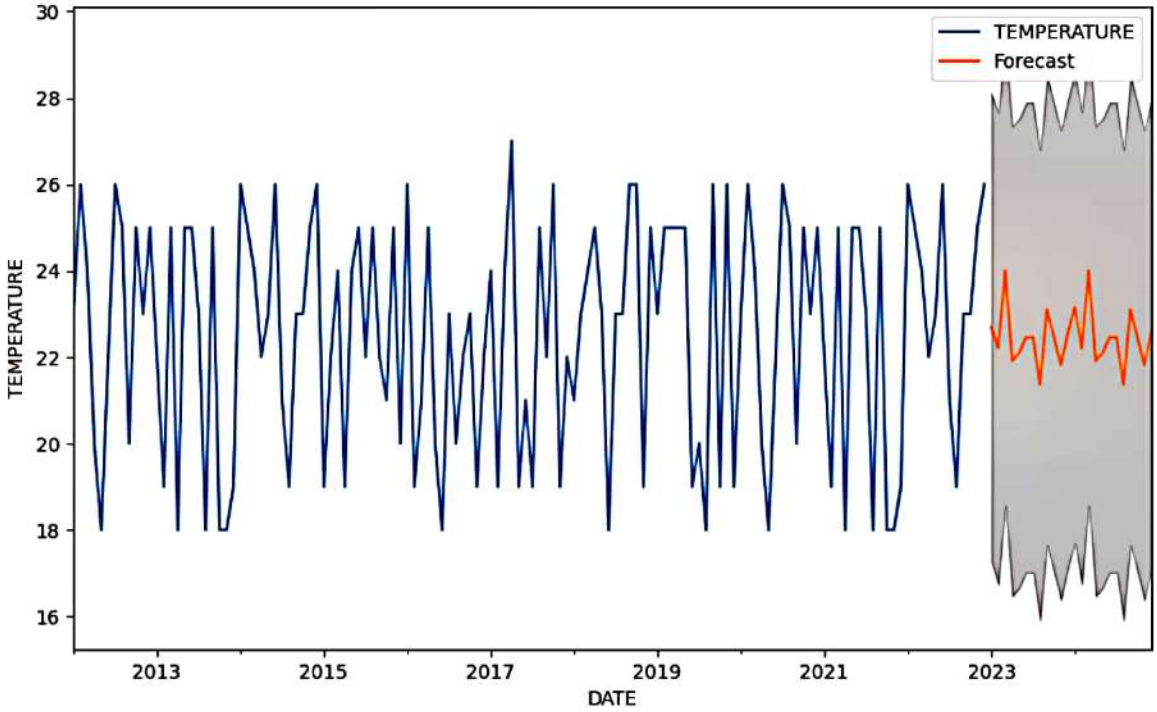


Fig 5.18 Forecasted graph

5.2.4 MAIZE PREDICTION ACCORDING TO RAINFALL

5.2.4 1 Plot the data

Firstly, display the data in the excel then plot this data to a graph. using python code. plotting the data helps to determine numerical information in a graphical format, to identify the trend and relationship within the data. Plotting the graph helps in gaining insights, making data-driven decisions.

Fig 5.22 depicts the time series plot of monthly rainfall data. It is visible that there is seasonality in the data.

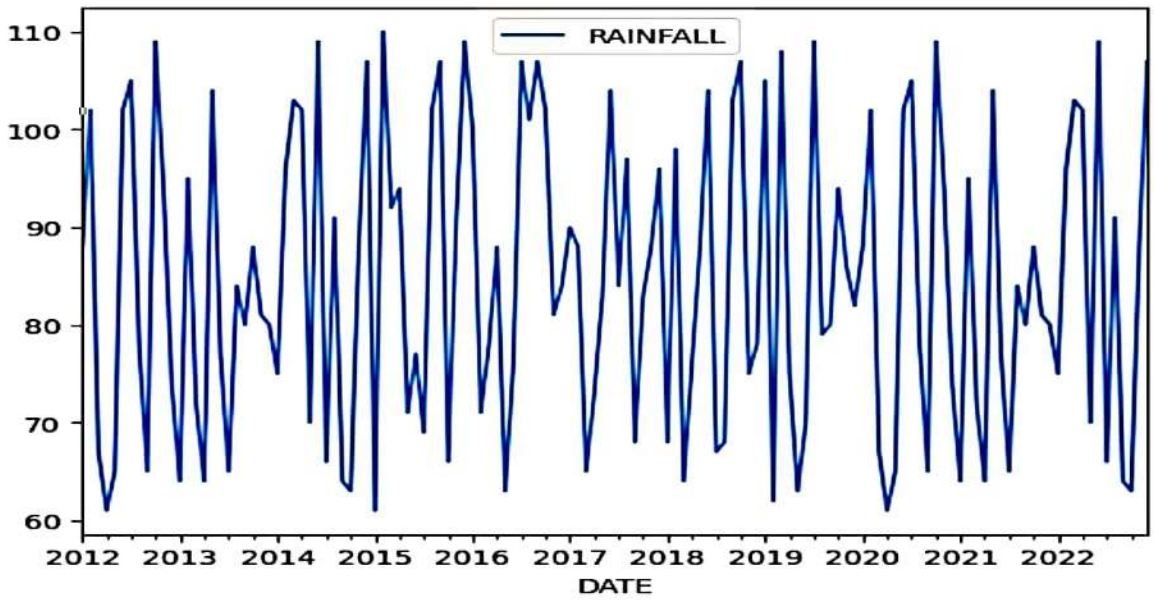


Fig 5.19 :Time series plot

5.2.4 2 DECOMPOSITION GRAPH

It is a graph that visualize the time series dataset, it include components such as trend seasonality and residuals,

Three main components are,

1.Trend: it explains the long-term movement or direction of the data, also represent overall growth or decline over time.

2. Seasonality: it shows where the data is fluctuating at a regular interval such as daily , weekly, monthly or yearly.

3.Residuals: The random fluctuations or irregularities in the trend seasonality of the data.

The decomposition graph that helps to analyze the each component separately also give an individual contribution to the behaviour of time series data.

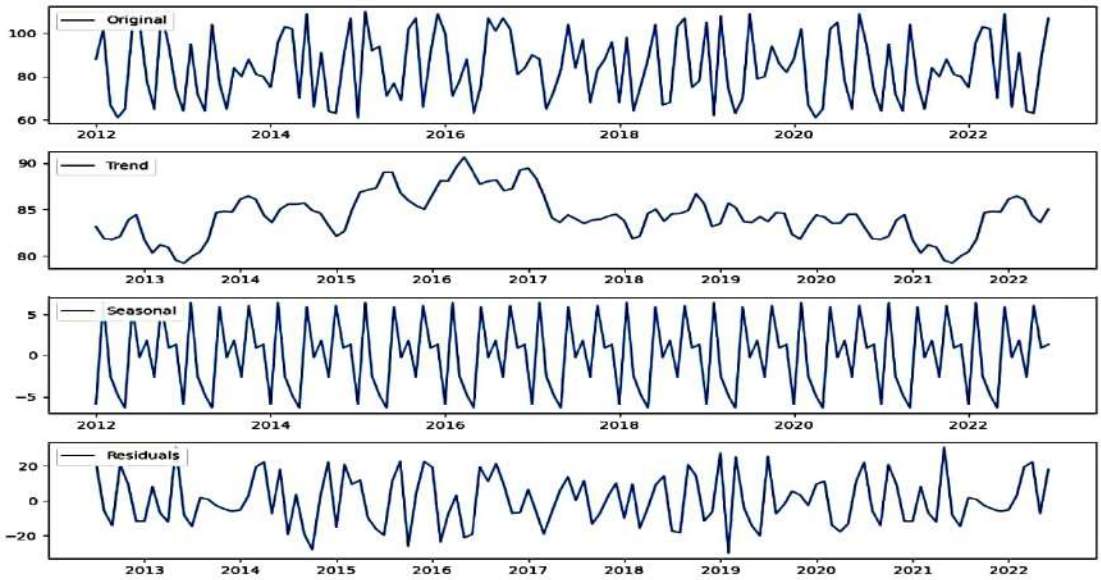
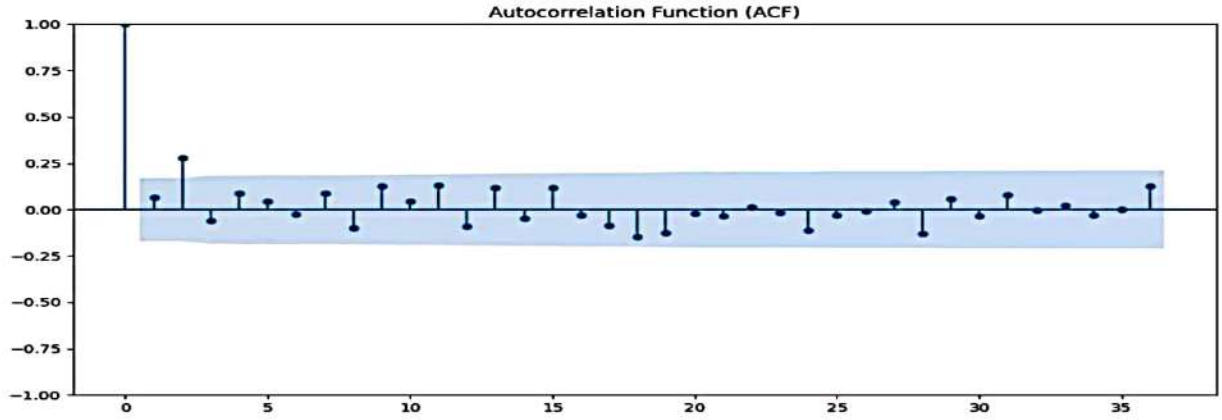


Fig 5.20 Decomposition graph

5.2.4 3 AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTION

Next step in analysing time series is to examine the autocorrelation function (ACF)and partial Autocorrelation function(PACF) . check the stationarity, plot the ACF and PACF values.

ACF and PACF plot is given below



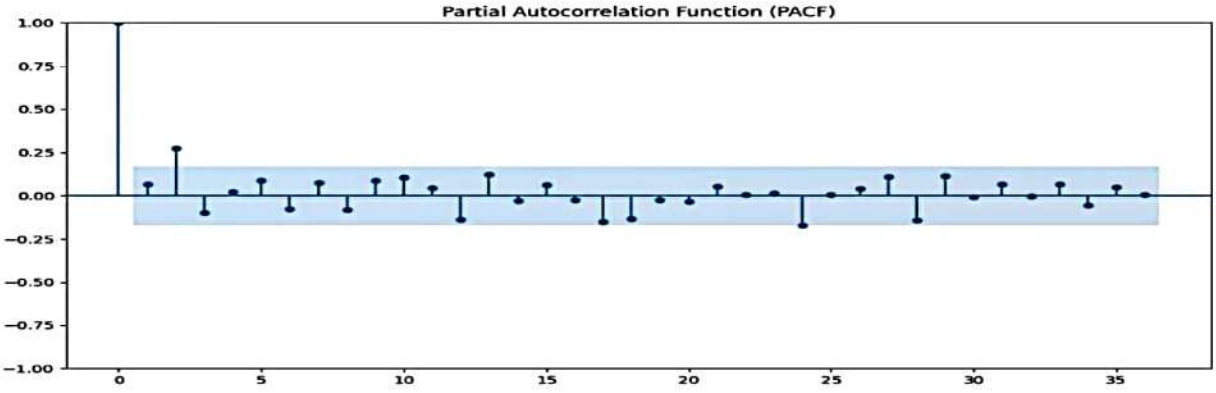


Fig 5.21 ACF and PACF

5.2.4 4 STATIONARITY USING AUGMENTED DICKEY-FULLER TEST

To test the time series data for stationarity using ADF test, follows a hypothesis testing approach. The null hypothesis H0 is given by,

H0: The data is non stationary.

The alternative hypothesis is given by

H1: The data is stationary

The outcome achieved, table

ADF test statistic	-5.661259567975618
Lag order	35
p-value	8.754446816502934e-07

Table-5.10

The ADF test given the p-value is 8.754446816502934e-07, so fail to accept H0 and hence can conclude that the data is stationary.

5.2.4 5 SARIMA MODEL FOR RICE PREDICTION ACCORDING TO RAINFALL

Next step is to choose the best model for forecasting. Choose the best model from all possible model according to Akaike Information Criterion(AIC). The model the lowest AIC value is the best model .Thus, the possible time series models along with their corresponding AIC.

5.2.4 6 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

It is a statistical technique to automatically select the optimal parameters(p,d,q) for an ARIMA model.

p- The number of lags in the observation.

d-Order of differencing.

q- order of the moving average.

Values obtained through AUTO-ARIMA can be used for forecasting. This process helps to save time and improve the accuracy of the forecasting models

Sl.NO	MODEL ARIMA (p,d,q) x (P,D,Q)	AIC
1	ARIMA(1,1,0)x(0,1,0,12)12	904.9710704562509
2	ARIMA(1,1,1)x(1,0,0,12)12	989.4518227175948
3	ARIMA(1,1,1)x(1,0,1,12)12	983.2531674597292
4	ARIMA(1,1,1)x(1,1,0,12)12	948.9037615250819

Table-5.11

VALUE OBTAINED THROUGH AUTOARIMA:

ARIMA (1, 1, 0) x (0, 1, 0, 12)12 - AIC:904.9710704562509

5.2.4 7 MODEL EVALUATION

Model evaluation includes assessing the errors or residuals that show difference between observed and predicted values from the model.

1.Standardized Residuals: It divides the standard deviation,it helps to determine how many standard deviations each residuals is from the expected values. A good model range from -2 to 2.

2.Histogram: it visually represents the distribution of the data.histogram includes bell-shaped curve.it indicate that data is normally distributed.

3.Normal Q-Q plot (Quantile-Quantile plot): it compares the quantiles of the normal distribution. if the residuals satisfy the normal distribution Q-Q plot is in straight line otherwise it show some deviation.

4. Correlogram: it is a auto-correlation plot that shows the correlation of residuals at different lags.

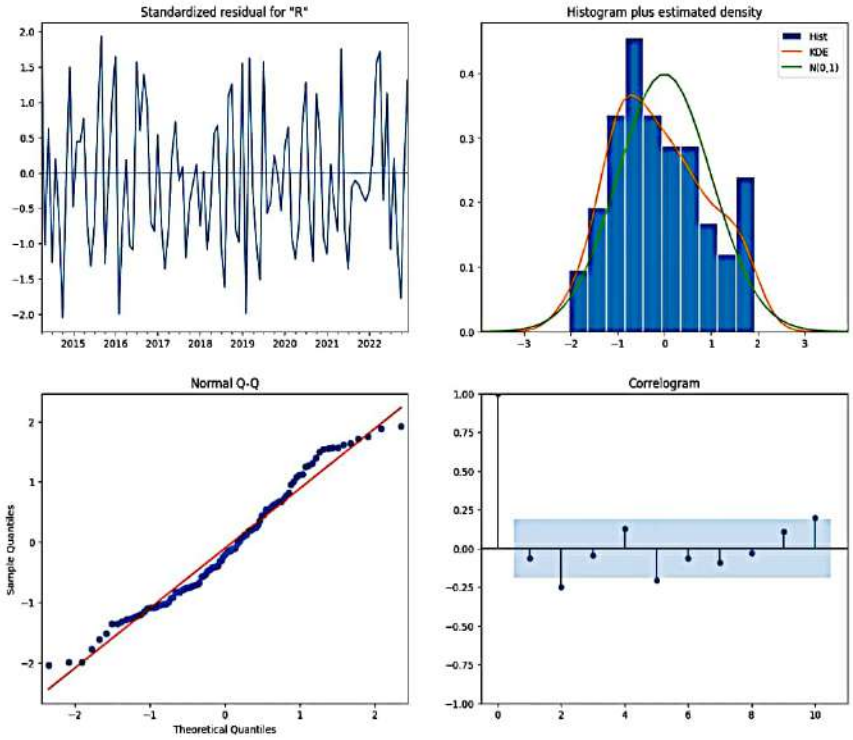


Fig 5.22 Model evaluation

5.2.4 8 ACTUAL V/S PREDICTED RAINFALL

Actual v/s predicted graph contains two lines, one represents actual values and other represents the predicted values. This type of graphs are commonly used in regression model. If two line comes close to each other gives better the model's performance in accurately predicting the variable of interest.

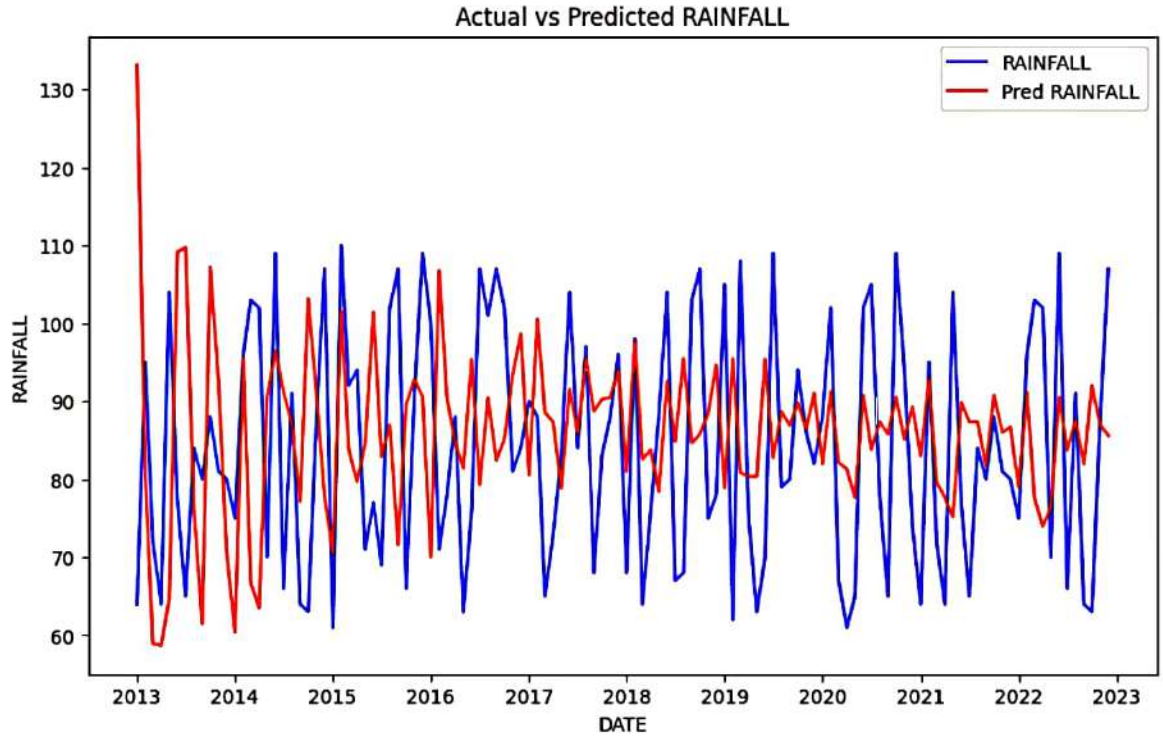


Fig 5.23 Actual v/s predicted rainfall

5.2.3.9 FORCASTED VALUES

Forecasted values is an estimate of the future value based on the historical data,trend and predictive models . it helps to determine what happens in the future values and it give approximate similar type of results by referring the previous data. Forecasting is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depends on the quality of the data, and the uncertainity related with the future events.

DATE	PREDICTED VALUE
2023-01-01	77.485980
2023-02-01	92.173348
2023-03-01	80.761016
2023-04-01	77.951513
2023-05-01	76.678408
2023-06-01	91.314361
2023-07-01	82.314777
2023-08-01	86.405411
2023-09-01	80.042060
2023-10-01	88.132687
2023-11-01	85.860160
2023-12-01	88.040958

Table-5.12

5.2.3.10 FORCASTED GRAPH

Forecasted graph displays the visualization of values over time. it combines with the historical data and the predictive values to estimate the future trends or patterns. Forecasting graphs is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depends on the quality of the data, and the uncertainity related with the future events

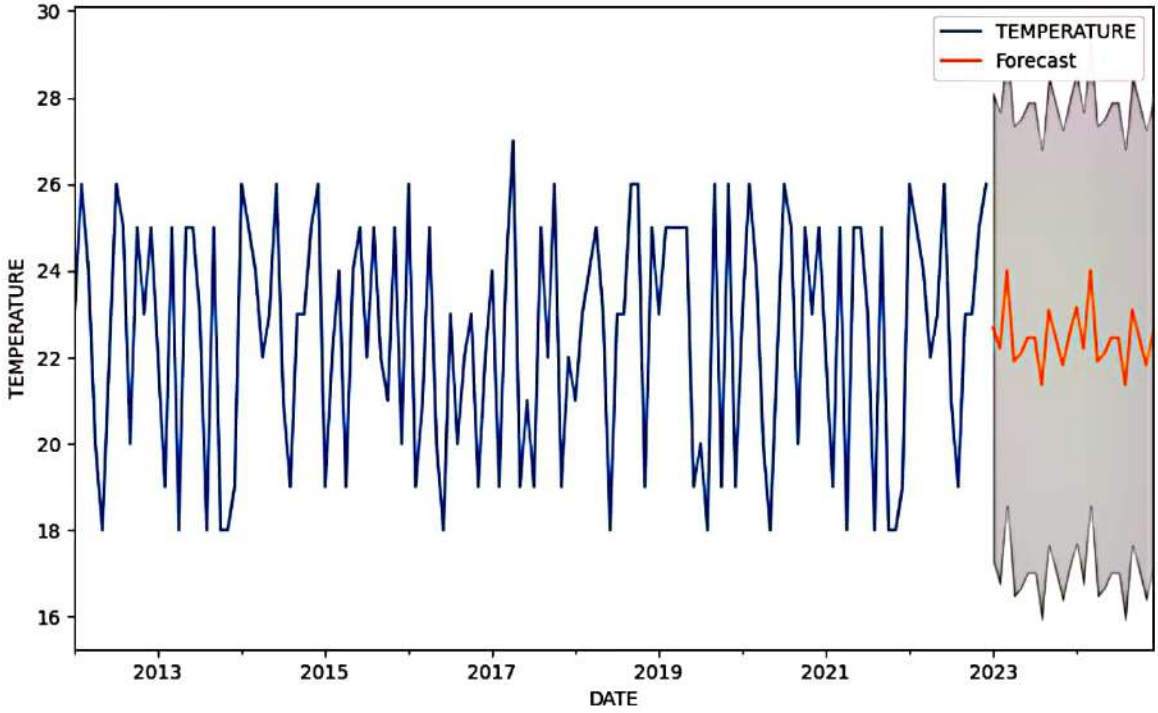


Fig 5.24 Forecasted graph

5.3 ANALYSIS USING HOLTS WINTER EXPONENTIAL SMOOTHING

5.3.1 RICE PREDICTION ACCORDING TO TEMPERATURE

5.3.1.1 PREDICTED GRAPH

Predicted graph displays the visualization of values over time.it combines with the historical data and the predictive values to estimate the future trends or patterns. Forecasting graphs is usually used in the field of weather prediction, finance, economics etc.The accuracy of the forecasted values depend on the quality of the data, and the uncertainity related with the future events

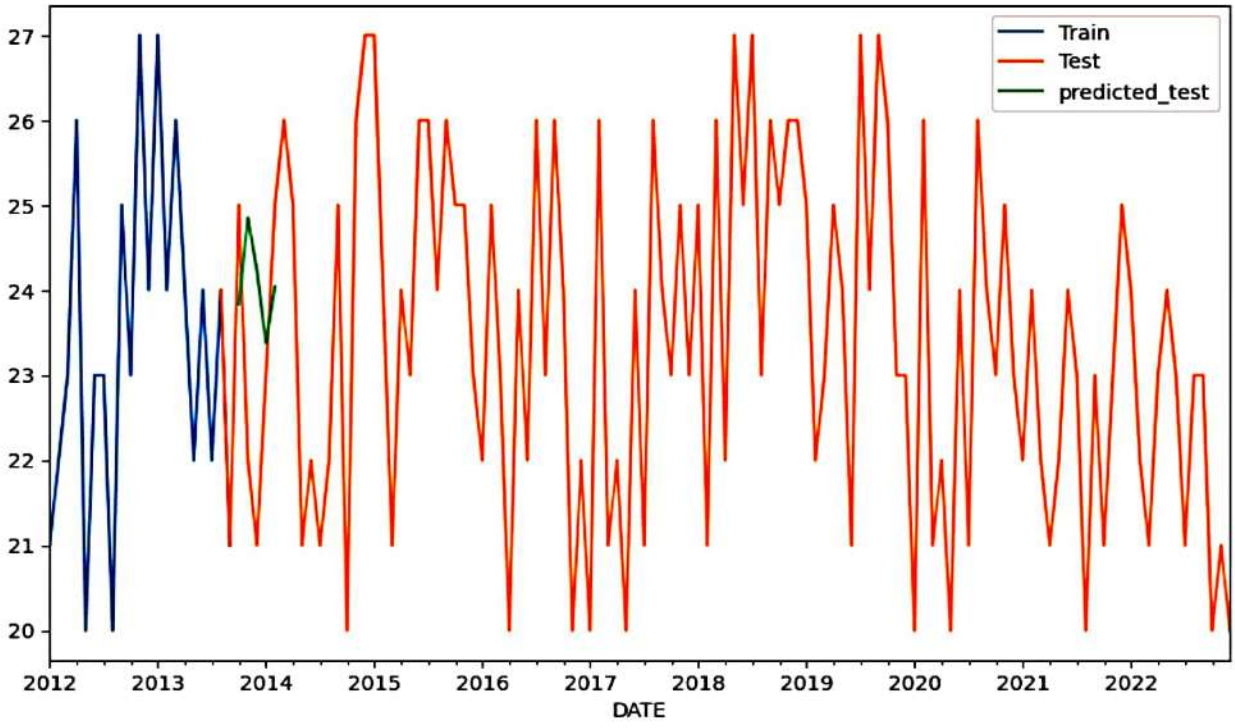


Fig 5.25 Predicted graph

5.3.1.2 DECOMPOSITION GRAPH

It is a graph that visualize the time series dataset, it include components such as trend seasonality and residuals,

Three main components are,

1.Trend: it explains the long-term movement or direction of the data, also represent overall growth or decline over time.

2. Seasonality: it shows where the data is fluctuating at a regular interval such as daily , weekly, monthly or yearly.

3.Residuals: The random fluctuations or irregularities in the trend seasonality of the data.

The decomposition graph that helps to analyze the each component separately also give an individual contribution to the behaviour of time series data.

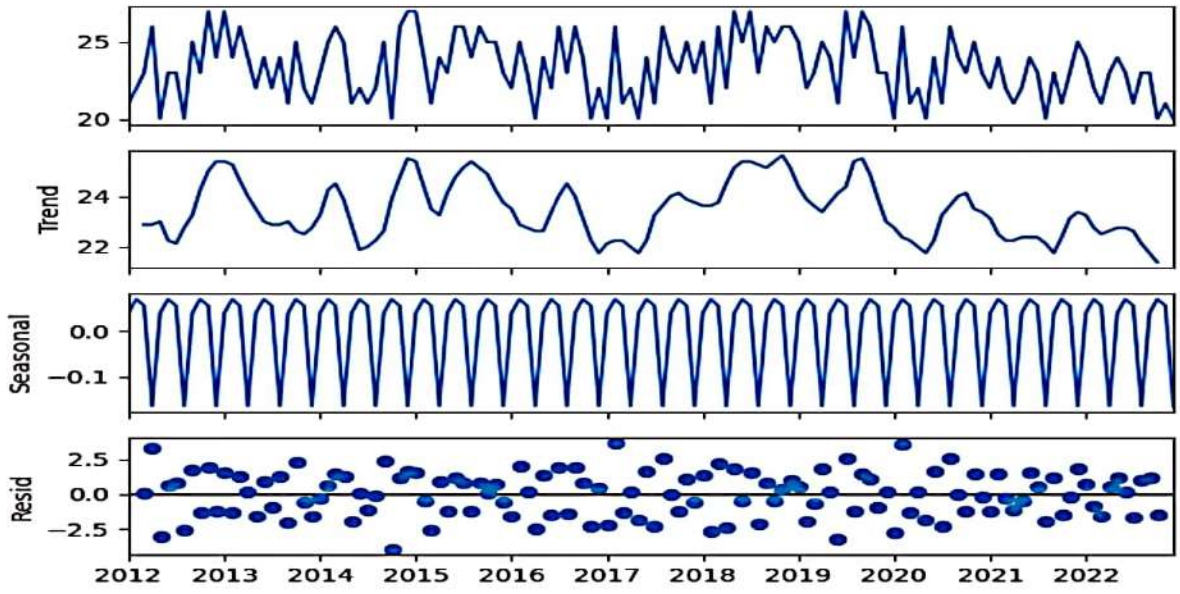


Fig 5.26 Decomposition graph

5.3.1.3 FORCASTED VALUES

Forecasted values is an estimate of the future value based on the historical data, trend and predictive models. it helps to determine what happens in the future values and it give approximate similar type of results by referring the previous data. Forecasting is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depends on the quality of the data, and the uncertainity related with the future event.

DATE	PREDICTED VALUE
2023-01-01	22.276040
2023-02-01	22.259587
2023-03-01	22.032974
2023-04-01	22.032974
2023-05-01	22.197920
2023-06-01	22.241631
2023-07-01	22.225190

2023-08-01	21.998914
2023-09-01	22.163592
2023-10-01	22.207222
2023-11-01	21.818177
2023-12-01	21.964854

Table-5.13

5.3.2.4 EXPONENTIAL SMOOTHING GRAPH

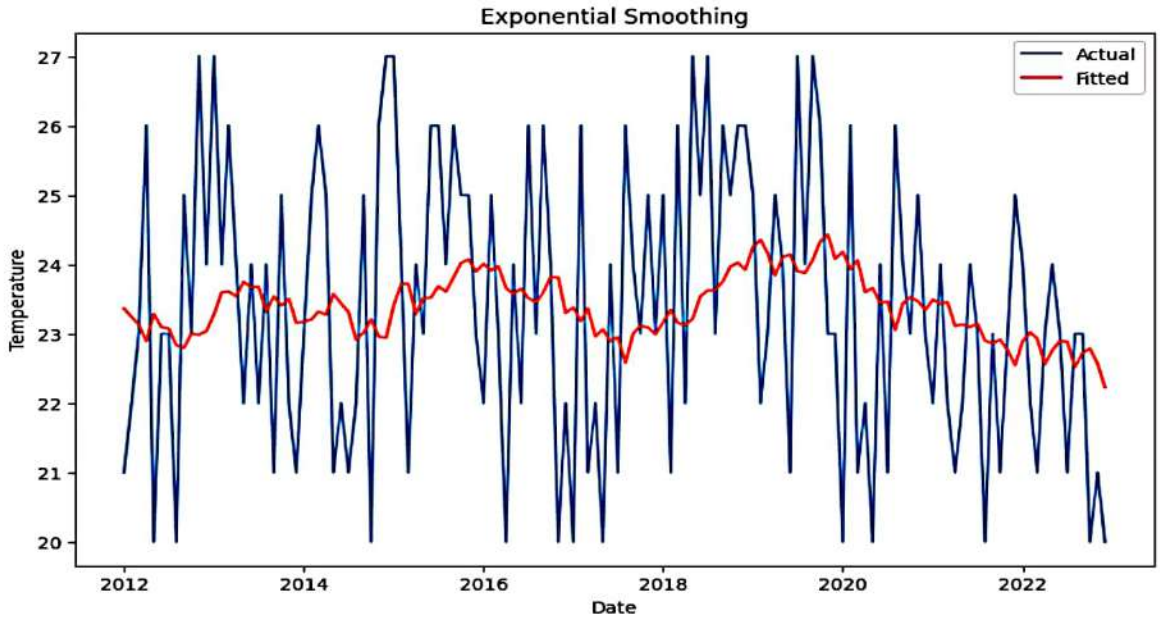


Fig 5.27 Exponential smoothing graph

5.3.2 RICE PREDICTION ACCORDING TO RAINFALL

5.3.2.1 PREDICTED GRAPH

Predicted graph displays the visualization of values over time.it combines with the historical data and the predictive values to estimate the future trends or patterns. Forecasting graphs is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depend on the quality of the data, and the uncertainity related with the future events

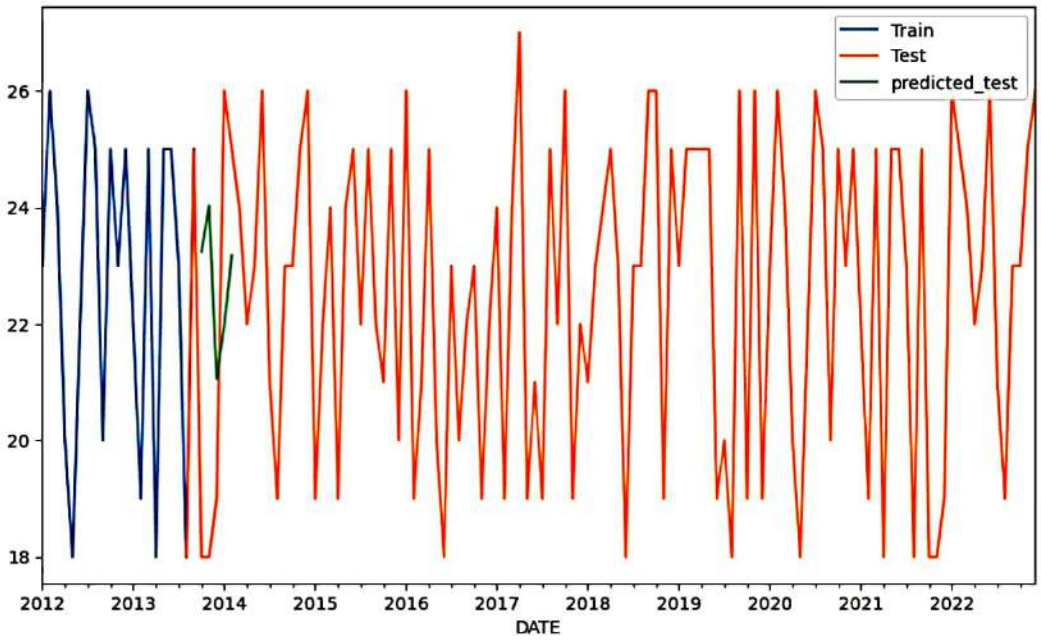


Fig 5.28 Predicted graph

5.3.2.2 DECOMPOSITION GRAPH

It is a graph that visualize the time series dataset, it include components such as trend seasonality and residuals,

Three main components are,

1.Trend: it explains the long-term movement or direction of the data, also represent overall growth or decline over time.

2. Seasonality: it shows where the data is fluctuating at a regular interval such as daily , weekly, monthly or yearly.

3.Residuals: The random fluctuations or irregularities in the trend seasonality of the data.

The decomposition graph that helps to analyze the each component separately also give an individual contribution to the behaviour of time series data.

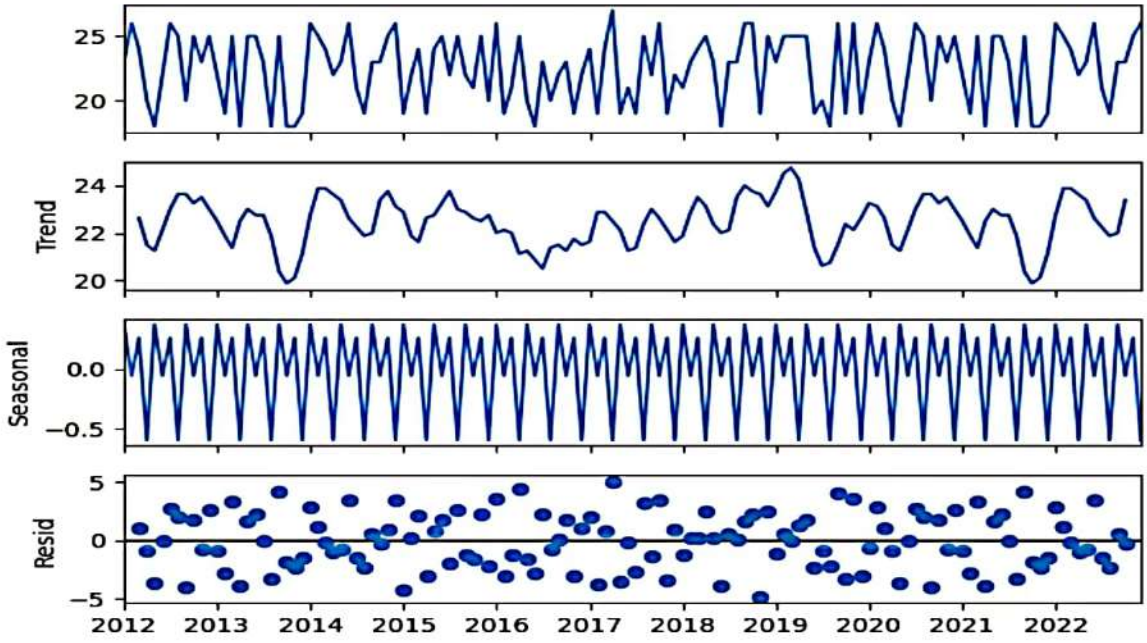


Fig 5.29 Decomposition graph

5.3.2.3 FORCASTED VALUES

Forecasted values is an estimate of the future value based on the historical data, trend and predictive models. it helps to determine what happens in the future values and it give approximate similar type of results by referring the previous data. Forecasting is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depends on the quality of the data, and the uncertainity related with the future event.

DATE	PREDICTED VALUE
2023-01-01	78.65664
2023-02-01	90.428935
2023-03-01	83.013879
2023-04-01	84.083859
2023-05-01	78.729995
2023-06-01	90.411292
2023-07-01	82.997681
2023-08-01	84.067452
2023-09-01	78.714632
2023-10-01	90.393648

2023-11-01	82.981484
2023-12-01	84.051045

Table-5.14

5.3.2.4 EXPONENTIAL SMOOTHING GRAPH

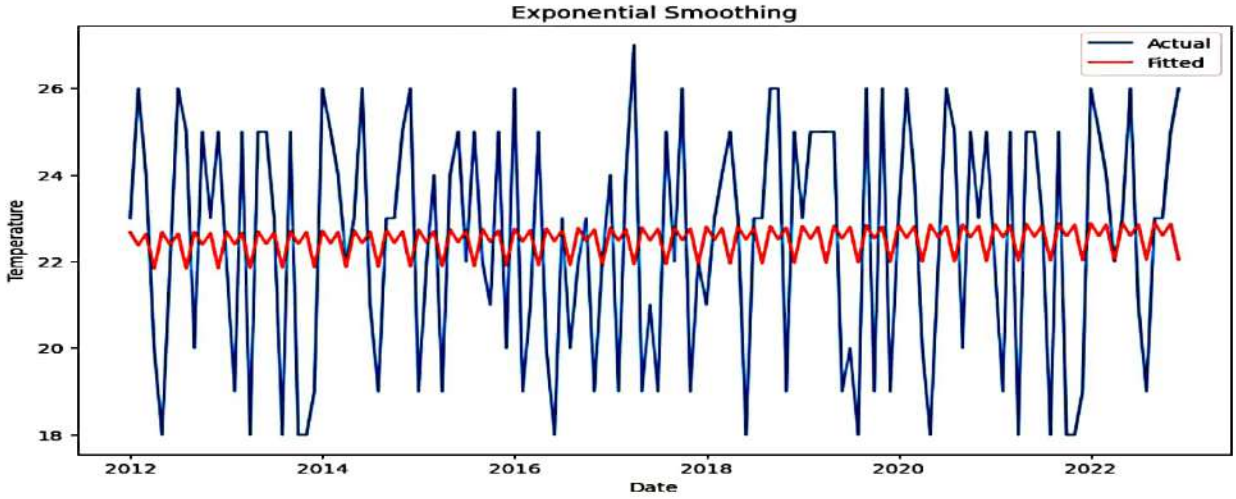


Fig 5.30 Exponential smoothing graph

5.3.3 MAIZE PREDICTION ACCORDING TO TEMPERATURE

5.3.3.1 PREDICTED GRAPH

Predicted graph displays the visualization of values over time.it combines with the historical data and the predictive values to estimate the future trends or patterns. Forecasting graphs is usually used in the field of weather prediction, finance, economics etc.The accuracy of the forecasted values depend on the quality of the data, and the uncertainity related with the future events

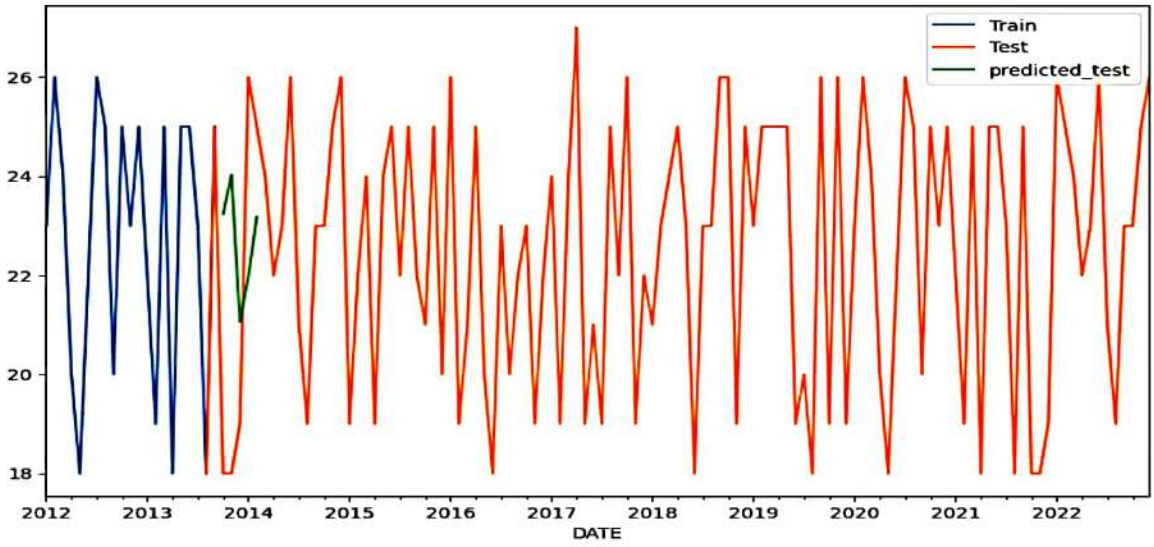


Fig 5.31 Predicted graph

5.3.3.2 DECOMPOSITION GRAPH

It is a graph that visualize the time series dataset, it include components such as trend seasonality and residuals,

Three main components are,

1.Trend: it explains the long-term movement or direction of the data, also represent overall growth or decline over time.

2. Seasonality: it shows where the data is fluctuating at a regular interval such as daily , weekly, monthly or yearly.

3.Residuals: The random fluctuations or irregularities in the trend seasonality of the data.

The decomposition graph that helps to analyze the each component separately also give an individual contribution to the behaviour of time series data.

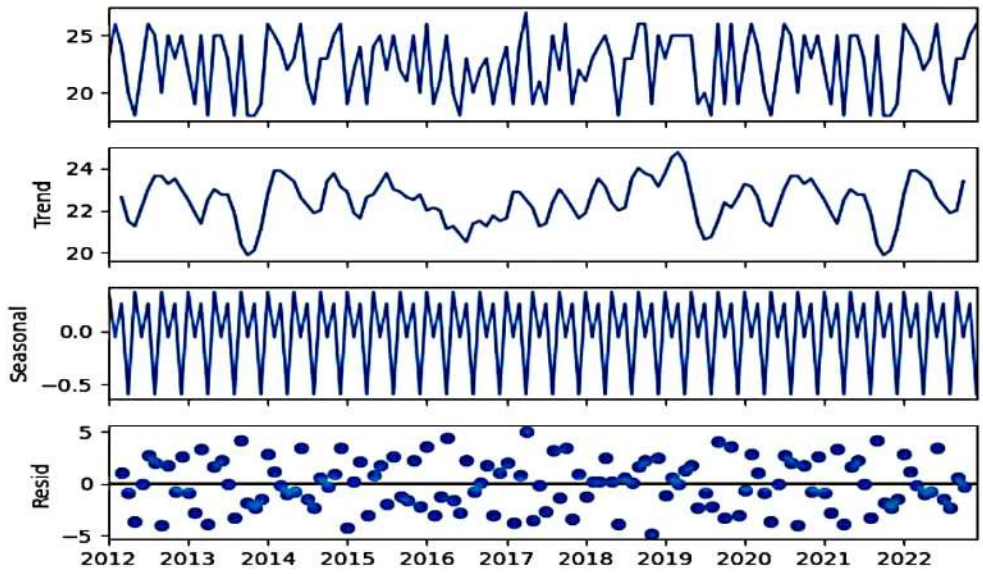


Fig 5.32 Decomposition graph

5.3.3.3FORCASTED VALUES

Forecasted values is an estimate of the future value based on the historical data, trend and predictive models . it helps to determine what happens in the future values and it give approximate similar type of results by referring the previous data. Forecasting is usually used in the field of weather prediction, finance, economics etc. The accuracy of the forecasted values depends on the quality of the data, and the uncertainity related with the future events.

DATE	PREDICTED VALUE
2023-01-01	22.5754444
2023-02-01	22.604444
2023-03-01	22.917364
2023-04-01	22.056189
2023-05-01	22.611494
2023-06-01	22.885528
2023-07-01	22.063067
2023-08-01	22.924510
2023-09-01	22.618545
2023-10-01	22.892664
2023-11-01	22.892665
2023-12-01	22.77666

Table-5.13

5.3.3.4 EXPONENTIAL SMOOTHING GRAPH

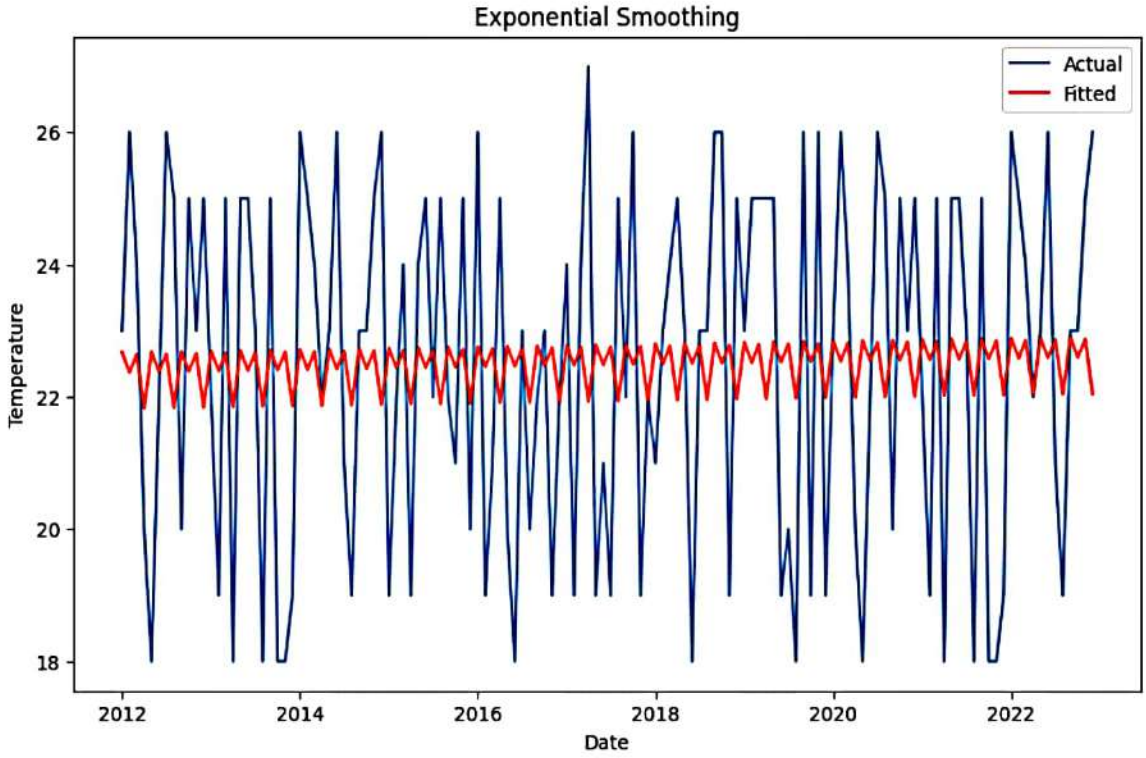


Fig 5.33 Exponential smoothing graph

5.3.4 MAIZE PREDICTION ACCORDING TO RAINFALL

5.3.4.1 PREDICTED GRAPH

Predicted graph displays the visualization of values over time.it combines with the historical data and the predictive values to estimate the future trends or patterns. Forecasting graphs is usually used in the field of weather prediction, finance, economics etc.The accuracy of the forecasted values depend on the quality of the data, and the uncertainity related with the future events

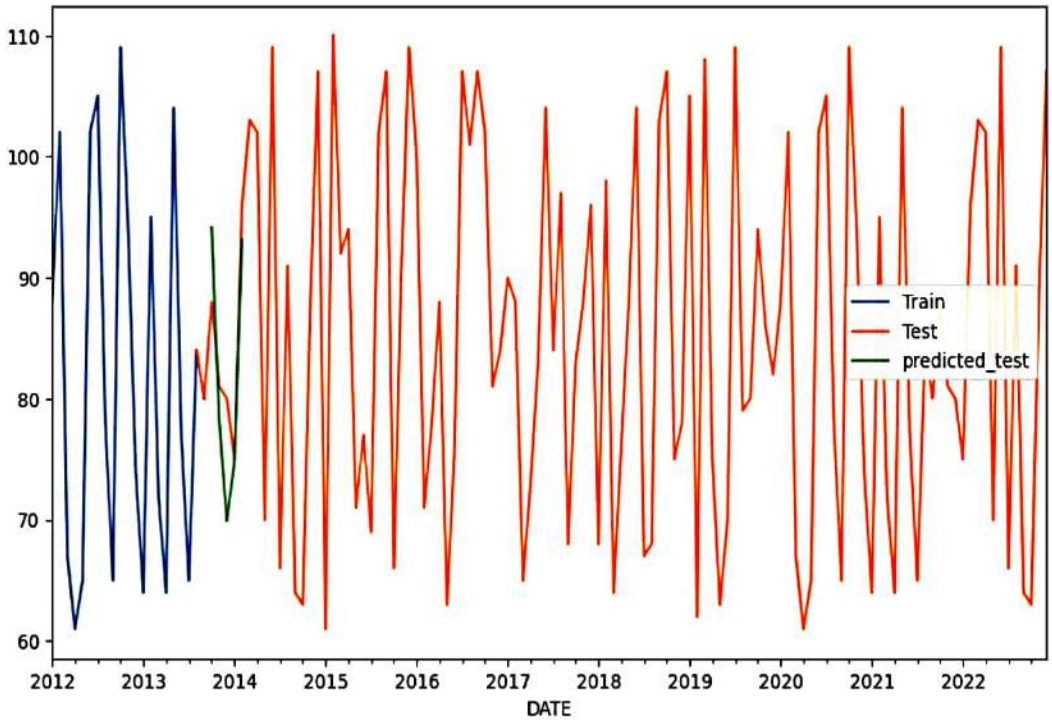


Fig 5.34 Predicted graph

5.3.4.2 DECOMPOSITION GRAPH

It is a graph that visualize the time series dataset, it include components such as trend seasonality and residuals,

Three main components are,

1.Trend: it explains the long-term movement or direction of the data, also represent overall growth or decline over time.

2. Seasonality: it shows where the data is fluctuating at a regular interval such as daily , weekly, monthly or yearly.

3.Residuals: The random fluctuations or irregularities in the trend seasonality of the data.

The decomposition graph that helps to analyze the each component separately also give an individual contribution to the behaviour of time series data.

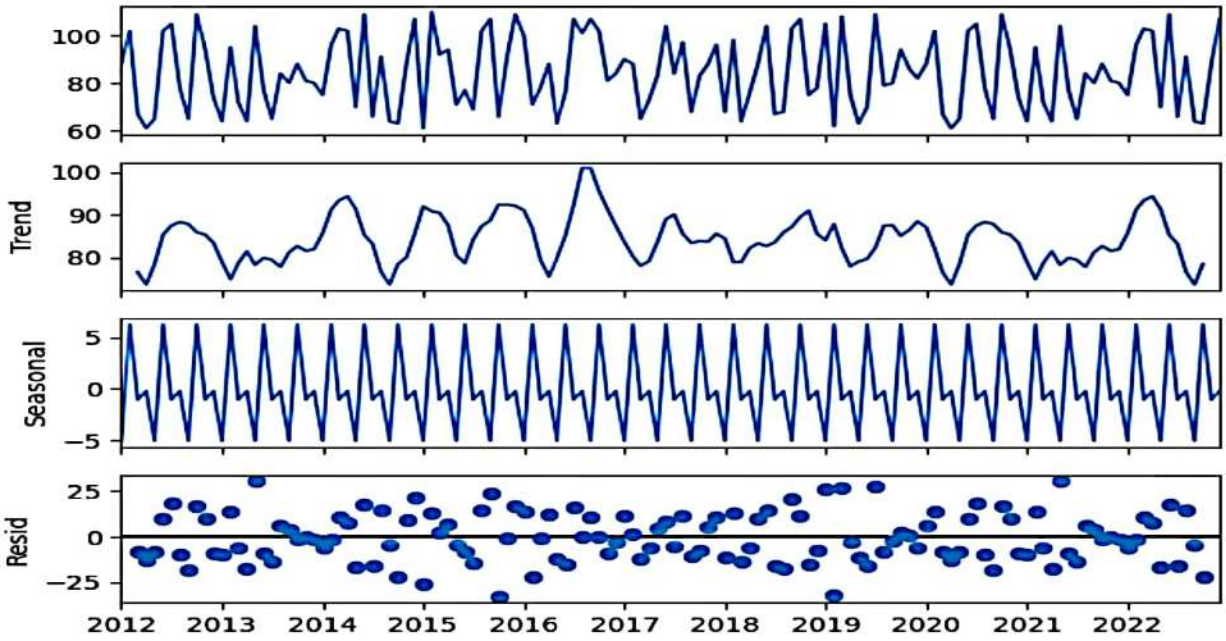


Fig 5.35 Decomposition graph

5.3.4.3 FORCASTED VALUES

DATE	PREDICTED VALUE
2023-01-01	78.745359
2023-02-01	90.428935
2023-03-01	83.013879
2023-04-01	84.083859
2023-05-01	78.729995
2023-06-01	90.411292
2023-07-01	82.997681
2023-08-01	84.067452
2023-09-01	78.714632
2023-10-01	90.393648
2023-11-01	82.981484
2023-12-01	84.051045

Table-5.1

5.3.4.4 EXPONENTIAL SMOOTHING GRAPH

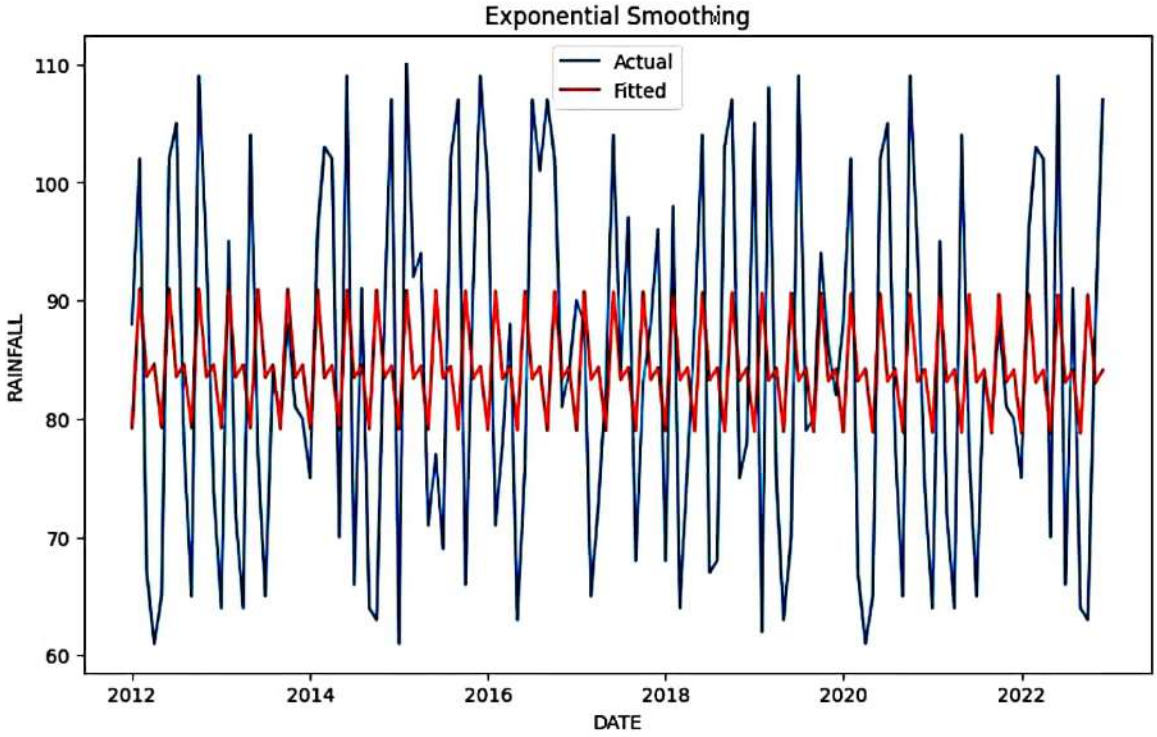


Fig 5.36 exponential smoothing graph

COMPARISON USING MSE AND RMSE VALUES

RICE PREDICTION ACCORDING TO RAINFALL;

SARIMA

MSE=682.1624

RMSE=26.11824

HOLT'S WINTER EXPONENTIAL SMOOTHING

MSE=648.0435

RMSE=25.4567

Conclusion- SARIMA is best model for Rice Prediction According to Rainfall

MAIZE PREDICTION ACCORDING TO TEMPERATURE;

SARIMA

MSE=9.33510

RMSE=3.05534

HOLT'S WINTER EXPONENTIAL SMOOTHING

MSE=3.8978

RMSE=1.9743

Conclusion- Holt's Winter Exponential Smoothing is best model for Maize Prediction According to Temperature

MAIZE PREDICTION ACCORDING TO RAINFALL;

SARIMA

MSE=244.0796

RMSE=15.62305

HOLT'S WINTER EXPONENTIAL SMOOTHING

MSE=283.0314

RMSE=16.82354

Conclusion- SARIMA is best model for Maize Prediction According to Rainfall

CHAPTER- 7

CONCLUSION AND FUTURE WORKS

In conclusion, the crop recommendation system developed in this has demonstrated its effectiveness in providing accurate and reliable recommendations for crop selection based on essential parameters suchas temperature, humidity, pH value of soil, rainfall, and N, P, K values.The system holds significant potential for revolutionizing agricultural practices by assisting farmers in making informed decisions and optimizing resource allocation.

Overall, the crop recommendation system has the potential to revolutionized the agricultural sector by improving crop productivity, optimizing resource utilization, and reducing environmental impact. Continued re- search and development in this field will contribute to the advancementof precision agriculture and ultimately benefit farmers and the overall sustainability of the agricultural industry.

7.1 FINAL OUTCOME AND FUTURE SCOPE

RESULTS

SARIMA model is more accurate than Holts winters exponential smoothing because it give more accurate predicted value and predicted graph. Here we got that rice prediction according to temperature Holt's Winter Exponential Smoothing is best model. Rice prediction according to rainfall SARIMA is the best model. Maize prediction according to temperature Holt's Winter Exponential Smoothing is best model and Maize prediction according to rainfall SARIMA is the best model.We know that Holt's Winter Exponential Smoothing have an limitation so we doesn't taken that Holt's Winter Exponential Smoothing has best model .overall SARIMA is the best model.

FUTURE SCOPE

The developed crop recommendation system has the potential to revolutionize agriculture in India by providing accurate crop selection recommendations based on various parameters.

Future developments if a crop get correct amount of temperature and rainfall help to increase the crop yield.

Creating a mobile application would enhance accessibility for farmers allowing them to easily access the system and receive real-time recommendations based on their location and weather conditions.

REFERENCES

- Champaneri, M., Chachpara, D., Chandvidkar, C., & Rathod, M. (2016). Crop yield prediction using machine learning. *Technology*, *9*(38).
- Ghadge, R., Kulkarni, J., More, P., Nene, S., & Priya, R. L. (2018). Prediction of crop yield using machine learning. *Int. Res. J. Eng. Technol.(IRJET)*, *5*, 2237-2239.. Kale, S. S., & Patil, P. S. (2019, December). A machine learning approach to predict crop yield and success rate. In *2019 IEEE Pune Section International Conference (PuneCon)* (pp. 1-5). IEEE.
- Gupta, M., Krishna, B. S., Kavyashree, B., Narapureddy, H. R., Surapaneni, N., & Varma, K. (2022, April). Crop Yield Prediction Techniques Using Machine Learning Algorithms. In *2022 8th International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-7). IEEE
- Kavita, M., & Mathur, P. (2020, October). Crop yield estimation in India using machine learning. In *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)* (pp. 220-224). IEEE.
- Medar, R., Rajpurohit, V. S., & Shweta, S. (2019, March). Crop yield prediction using machine learning techniques. In *2019 IEEE 5th international conference for convergence in technology (I2CT)* (pp. 1-5). IEEE.
- Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. (2019, November). Crop yield prediction using machine learning algorithms. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). IEEE
- Nishant, P. S., Venkat, P. S., Avinash, B. L., & Jabber, B. (2020, June). Crop yield prediction based on Indian agriculture using machine learning. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-4). IEEE.
- Ranjani, J., Kalaiselvi, V. K. G., Sheela, A., & Janaki, G. (2021, December). Crop yield prediction using machine learning algorithm. In *2021 4th international conference on computing and communications technologies (ICCCT)* (pp. 611-616). IEEE.

Reddy, D. J., & Kumar, M. R. (2021, May). Crop yield prediction using machine learning

algorithm.In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1466-1470). IEEE

Website: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset