

**A**

**PROJECT**

**ON**

**Time Series and Machine learning Approach for Weather Forecasting**

**Submitted to**

**MIT ADT UNIVERSITY**

**SCHOOL OF ENGINEERING & SCIENCE**

**DEPARTMENT OF APPLIED SCIENCE & HUMANITIES**

In partial fulfilment of the requirements for the award of the degree of

**MASTER OF SCIENCE**

**IN**

**APPLIED STATISTICS (DATA SCIENCE)**

Submitted by

**Mr. Suraj Ramchandra Jagtap MITU22MSDS0012**

**Mr. Rohit Mahendra Dixit MITU22MSDS0028**

**Under the Guidance of**

**Dr. Ashok Kumar**

**December, 2023**



**SCHOOL OF ENGINEERING & SCIENCE**

**DEPARTMENT OF APPLIED SCIENCE & HUMANITIES**

**RAJBAUG, LONI KALBHOR, PUNE-412201**

**CERTIFICATE**

**This is to certify that the Capstone Mini Project-22MSDS321 entitled**

**Time Series and Machine learning Approach for Weather Forecasting**

Submitted by

**Mr. Suraj Ramchandra Jagtap MITU22MSDS0012**

**Mr. Rohit Mahendra Dixit MITU22MSDS0028**

is a bonafide work carried out by them, under the supervision of guide and co-guide, it is submitted towards the partial fulfilment of the requirement of MIT Art, Design and Technology University, Pune for the award of the Master of Science in Applied Statistics (Data Science).

|  |  |
| --- | --- |
| Dr. Ashok Kumar  Guide | Prof. Dr. Satish Jadhav  Co-guide |
| Dr. Pratibha Jadhav  Program Co-ordinator | |

|  |  |
| --- | --- |
| Prof. Dr. Haribhau Bhapkar  HOD,  Applied Sciences & Humanities | Prof. Dr. Virendra Shete,  Director,  MIT School of Engineering & Science |

External Examiner

Place: Pune

Date:

**DECLARATION**

We hereby declare that the Project entitled “**Time Series and Machine learning Approach for Weather Forecasting”** submitted towards the partial fulfilment of the requirement of MIT-ADT University, Pune for the award the Master of Science in Applied Statistics (Data Science) of the is a record of bonafide work carried out by us under the supervision of Guide Name and Co-guide name, Department of Applied Science & Humanities, MIT School of Engineering & Science, Pune.

We further declare that the work reported in this report has been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Pune Student Names and Sign

………………..

(Suraj R. Jagtap)

………………..

(Rohit M. Dixit)

**ACKNOWLEDGEMENT**

I would like to thank all teaching and non-teaching staff and all my colleagues. I would like to express my profound gratitude towards guide Dr. Ashok Kumar and co-guide Prof. Satish Jadhav for his valuable guidance for completion of this project. I am also thankful to Dr. Pratibha Jadhav, Rohit Raskar Sir, for their timely suggestions and encouragements.

I would like to thank Prof. Dr. Haribhau Bhapkar, Head of Department Applied Science & Humanities, MIT Art, Design and Technology University, Pune (India), for providing me the necessary facilities.

**ABSTRACT**

Accurate temperature forecasting is crucial for various applications, including weather prediction, agriculture planning, and energy management.

This project aims to compare two time series forecasting models namely SARIMA, exponential smoothing, and random ARIMA model, for predicting daily temperature in Pune, Maharashtra, India, using historical data from 2018 to 2022.

The SARIMA model, a statistical method based on autoregressive and moving average components, is well-suited for capturing seasonal patterns and trends in time series data. The SARIMA model is commonly used for forecasting time series data.

Exponential smoothing, on the other hand, is a simpler and more computationally efficient approach that weights recent observations more heavily than older ones, making it suitable for short-term forecasting.

**Data Collection and Preprocessing:** We collected and cleaned a vast dataset of Pune weather dataset (1 January 2018 to 11 March 2022).

Data preprocessing is playing a crucial role in time series analysis. It involves tasks like handling missing values, removing outliers, scaling features, and encoding categorical variables. By preprocessing the data, we ensure that data is in right format and ready for training our models.

**Fitting time series model:** We fit the time series models like SARIMA model and Exponential Smoothing Model for forecasting the temperature.

SARIMA is a powerful tool and it is used for modelling and forecasting time series data, especially when dealing with seasonality and trend components.

Exponential Smoothing is a popular time series forecasting method used to make predictions about future data points based on historical data. It is particularly well-suited for time series data with trends and seasonality.

The ARIMA model, which stands for Autoregressive Integrated Moving Average, is a powerful statistical method for analyzing and forecasting time series data. It's particularly adept at handling data that exhibits trends, seasonality, and autocorrelation.

**CONTENTS**

Title Page ……………………………………………………………………… 1

Certificate …………………………………………………………………….. 2

Declaration …………………………………………………………………… 3

Acknowledgement …………………………………………………………. 4

Abstract ………………………………………………………………………… 5-6

Chapter-1 Introduction …………………………………………………………………. 8-9

Chapter-2 Literature Reviews …………………………………………………………. 10

Chapter-3 Objective- ……………………………………………………………………… 11

Chapter-4 Methodology- ………………………………………………………………… 12-36

4.1 Data Description ……………………………………………………. 14-15

4.2 Data Preprocessing …………………………………………………. 16-17

4.3 Data Visualization ………………………………………………….. 18-23

4.4 Analysis Using Software ………………………………………….. 24

4.5 Statistical Modelling ……………………………………………….. 25-36

Chapter-5 Conclusion and Discussion …………………………………………….. 37

Chapter-6 Limitation and Future Scope …………………………………………… 38-39

Chapter-7 Reference ………………………………………………………………………. 40

Annexures: Plagiarism Report …………………………………………. 41

**Chapter-1**

**Introduction**

Pune, located in state of Maharashtra, India. Pune is known for its dynamic and changeable environment. Pune has four different seasons like hot and dry summer, monsoon season with heavy rainfall, spring and a chilly winter. Temperature during the summer frequently increase, resulting in burns and dry conditions. Summer season start from March to May. In summer months the temperature increases up to 38 degrees Celsius. The monsoon season occurs from June to September. The monsoon produces heavy rain that change the weather. During the monsoon the monthly the average Rainfall of Pune is 61.8 mm per month. In monsoon months humidity levels rises and affecting the human comfort and climate condition.

In Spring season there is beautiful weather and comfortable temperature. Spring season start from October to January. In winter season Pune has lower temperature 10 degrees Celsius which provides a relief from the summer heat. Pune’s weather changes due to seasonal changes in wind speed, atmospheric pressure, humidity and rainfall and other climatic factors.

Pune's temperature patterns significantly impact various aspects of daily life, including agriculture, urban planning, and public health. The distinct wet and dry seasons influence agricultural practices, with farmers adapting their crop choices and irrigation strategies based on the prevailing temperature and rainfall patterns.

Urban planners consider temperature patterns when designing buildings and infrastructure, incorporating measures to mitigate the effects of extreme heat and ensure thermal comfort for residents. Additionally, public health officials monitor temperature trends to anticipate and address potential heat-related health issues.

Humidity and Rainfall

Pune's humidity levels vary throughout the year, with the highest humidity during the monsoon season (around 80%). During the summer, humidity levels are lower, averaging around 40%.

Rainfall is concentrated during the monsoon season, with an average annual rainfall of around 660 mm (26 inches). The city experiences very little rainfall during the rest of the year.

Wind

Pune experiences moderate wind speeds throughout the year, with an average wind speed of around 10 km/h (6 mph). The wind direction varies depending on the season. During the summer, the prevailing wind direction is from the southwest, while during the winter, it is from the northeast.

**Data Overview:**

The Pune weather dataset is historical and large dataset that contains various factors which effects on weather. This dataset is from 1 January 2018 to 11 March 2022. Dataset provides the detailed insights into daily(hourly) weather pattern and trends.

The dataset includes information about average temperature, rainfall in mm, atmospheric pressure in pascal, Humidity and wind speed in Kmph. Overall, the Pune weather dataset shows that the which factors affects the weather condition.

**Chapter-2**

**Literature Reviews**

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is a statistical model that is used for analysing and forecast the time series data. Statisticians George Box and Gwilym Jenkins developed the model in 1970’s. SARIMA model is popular tool for forecasting the time series data with seasonality.

1. A study by Aarati G Gangshetty, et al. (2021) apply a SARIMA model to forecast temperature in Pune using data from 2009 to 2020. The authors evaluated the model's performance using mean absolute error (MAE) and root mean squared error (RMSE)

Exponential Smoothing was developed by two researchers, Robert Goodell Brown and Charls C. Holt in 1950’s. Exponential Smoothing model is widely used in variety of applications such as sales forecasting, stock prize forecasting. They are also used in quality control and signal processing. It is powerful tool for time series forecasting model.

1. Exponential smoothing use to forecast the annual temperature of India. Use data from the India Meteorological Department (IMD) for the period from 1901 to 2021.
2. "Monthly Temperature Prediction Based on ARIMA Model: A Case Study in Dibrugarh Station of Assam, India" by Goswami, K., Hazarika, J., & Patowary, A. N. (2017) likely explores the application of ARIMA modeling to predict monthly temperatures in the Dibrugarh Station of Assam, India. Unfortunately, I don't have direct access to the content of the paper, but I can provide a general overview based on the information available in the title.

**Chapter-3**

**Objectives**

Our goal in doing this project is to analyse and the forecast the temperature of Pune city, using time series modelling and machine learning techniques.

The specific objectives are:

* **To Study the trend pattern of weather dataset and identify seasonal patterns and variations.**
* **To fit appropriate time series.**
* **Find the forecast series of trend and seasonality.**
* **Compare the performance of time series models.**

**Chapter-4**

**Methodology**

**Data Collection:** The historical data on Pune Weather (1 Jan 2018 – 11 Mar 2022). This data is collected from the Kaggle.com.

**Data Pre-processing:** An exploratory data analysis (EDA) was conducted to gain meaningful insights from the datasets. This included:

* Understanding the Dataset: Examining the structure, size, and data types of the features.

* Handling Missing Data: Identify the missing values in the dataset.

**Data Visualization:** After cleaning the data, it is important to visualize the data. This step can be done by using Python libraries such as Matplotlib, Seaborn, Plotly, and other libraries.

**Models:** In this project for the forecasting the temperature, use the three forecasting time series models SARIMA (Seasonal Autoregressive Integrated Moving Average) and Exponential Smoothing Model.

**Model Comparison:** The performance of the three forecasting models, SARIMA, exponential smoothing, and ARIMA, was evaluated using metrics like mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE).

**Conclusion:** The results of this project suggest that the exponential smoothing model is the best model for forecasting the temperature in Pune, Maharashtra, India. This is likely because the exponential smoothing model is a simple model that is well-suited for short-term forecasting.

**4.1 Data Description: -**

Data is taken from the Kaggle website. This is large dataset contains 36745 rows and 25 columns.

Here is short overview of Data:

* Source: Kaggle.com
* Link: <https://www.kaggle.com/datasets/dipakdeshmukh/weather-data-of-punemaharashtraindia2008-2022>
* Weather Data of Pune, Maharashtra, India. (2018-2022).



There are some important variables in dataset are:

**Date/Time**: This column typically records the date and time of the weather observation.

**Temperature**: It includes columns such as "Temperature," which records the air temperature at the time of the observation. It has sub-columns like maximum temperature (Max Temp) and minimum temperature (Min Temp) for the day.

**Humidity**: This column indicates the level of humidity in the air, represented as a percentage.

**Precipitation**: Precipitation columns might include data on rainfall, snowfall, or other forms of precipitation. Common columns include "Precipitation”, “Rainfall, and “Snowfall”.

**Pressure**: Atmospheric pressure is often included, and it can be measured in various units like millibars(mb).

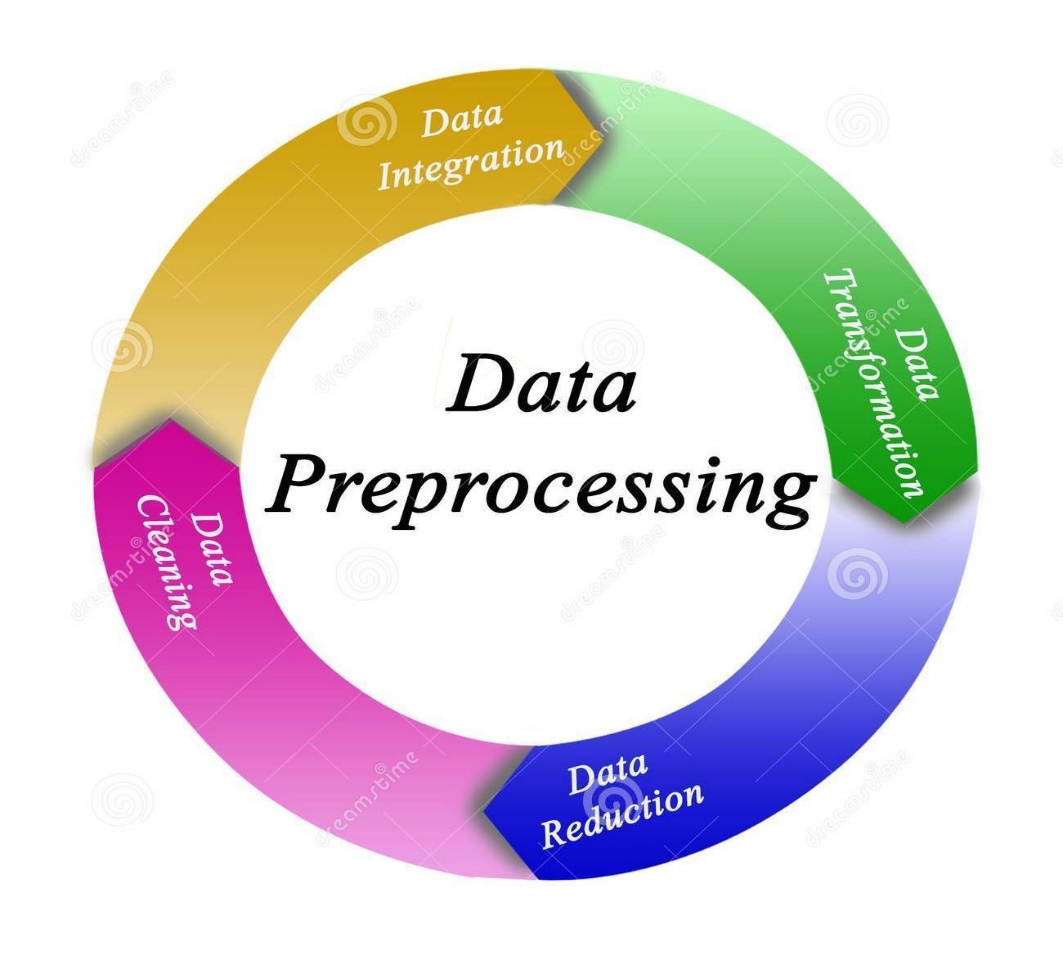
**Weather Conditions**: This column might describe the overall weather conditions, such as "Clear," "Cloudy," "Rainy," "Snowy," etc.

**4.2 Data Pre-processing:**

Data preprocessing is playing a crucial role in time series analysis. It involves tasks like handling missing values, removing outliers, scaling features, and encoding categorical variables. By preprocessing the data, we ensure that data is in right format and ready for training our models. Data preprocessing is a primary step in machine learning where we prepare our data before training our models. It involves various techniques to clean, transform, and enhance the quality of our data.

Some common steps in data preprocessing includes:

1. Data Cleaning.
2. Data Integration.
3. Data Transformation.
4. Data Reduction.

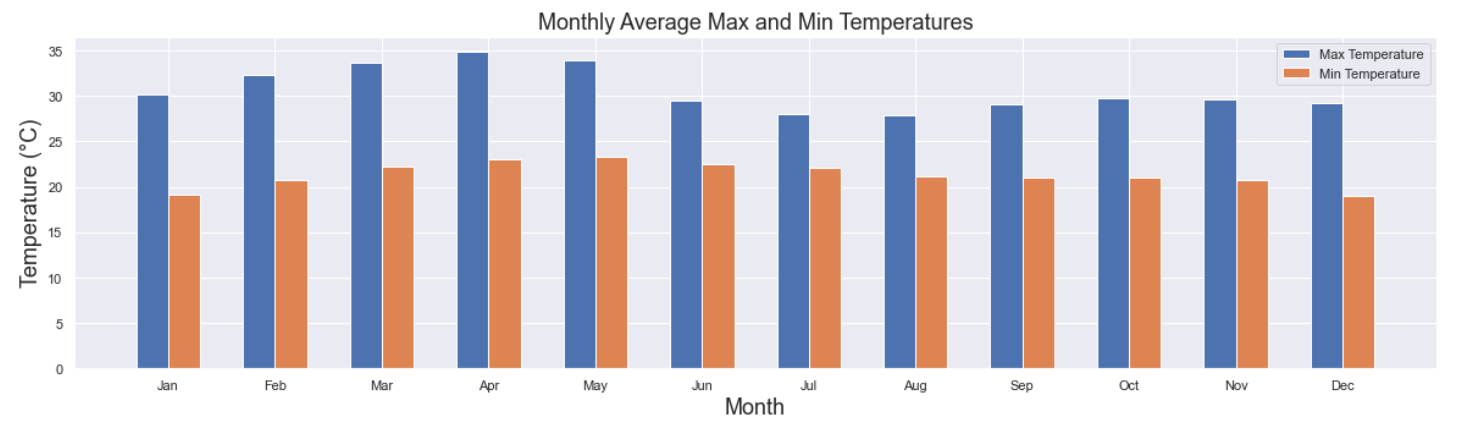


One important task in data preprocessing is handling missing values. But there are no missing values present in this dataset.

Data preprocessing helps us in achieving better model performance and more accurate predictions. It ensures that our data is in the right format, free from inconsistencies, and ready to be fed into our machine learning models.

**4.3 Data Visualization:**

1. **Bar Graph of monthly average maximum and minimum temperature.**

****

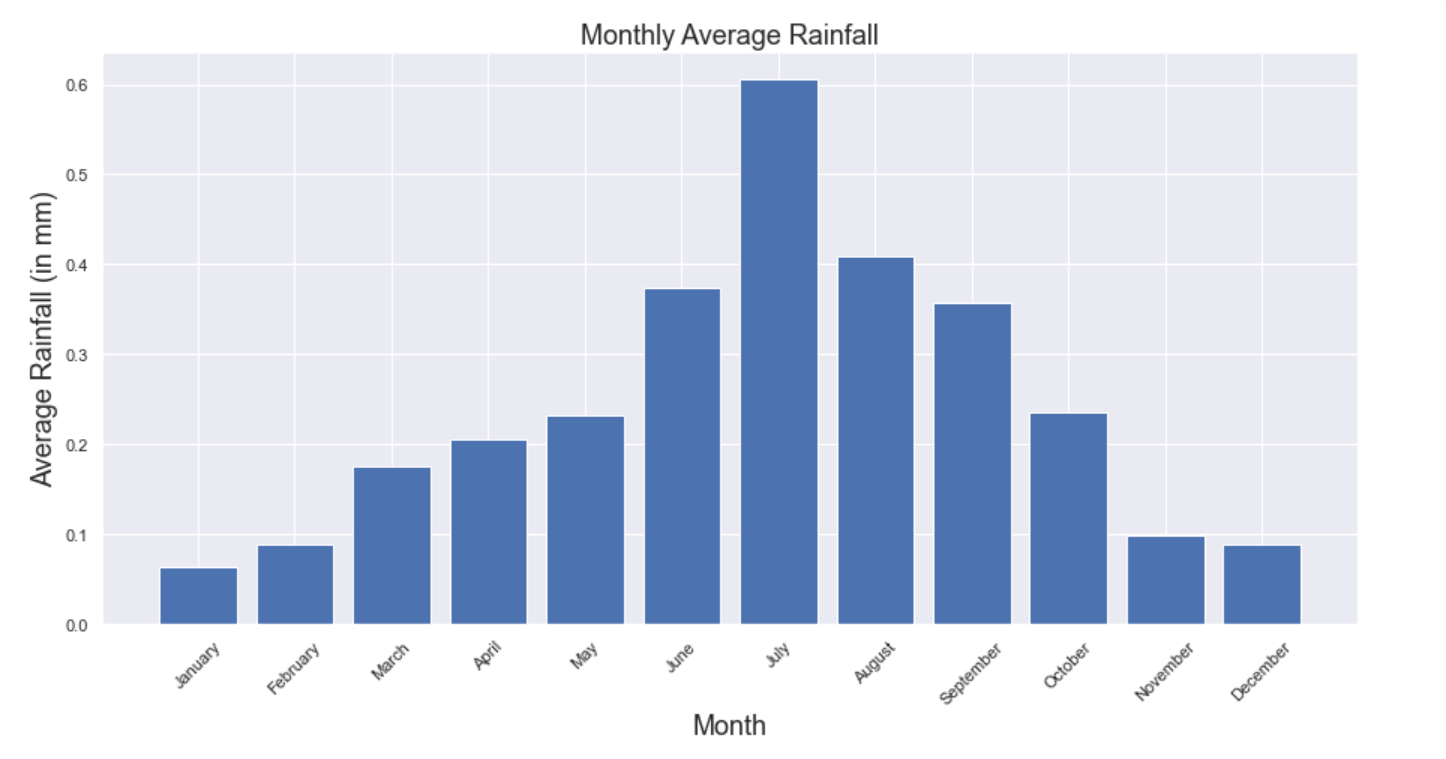
**Fig.1**

**Interpretation:**

This figure shows the monthly average maximum and minimum temperature in Pune. The blue bar represents the maximum temperatures and the orange bar represent minimum temperatures.

The figure shows that the average temperature in Pune varies throughout the year the highest temperature is seen in April and May. The average maximum temperature reaches around 35°C. The lowest temperatures are seen in December and January. The average minimum temperature drops to around 10°C.

1. **Bar Plot of Monthly Average Rainfall.**

****

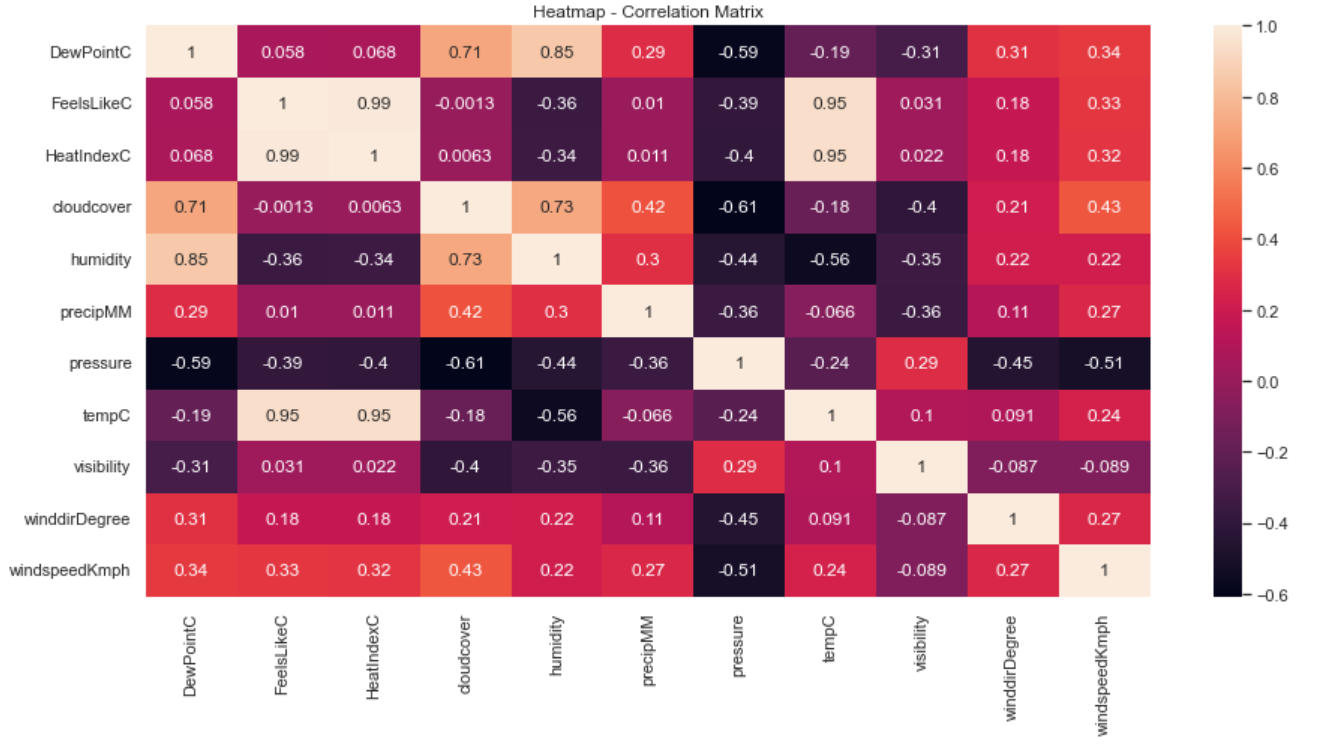
**Fig.2**

**Interpretation:**

The figure shows the monthly average rainfall in Pune. The figure indicates that the average rainfall is highest rainfall accrues in July, which receives an average of 0.6 mm/h. The lowest average rainfall occurs in January, which receives an average less than 0.1 mm/h.

The figure shows a strong seasonal pattern in the rainfall in Pune. Overall Pune city receives the moderate amount of rainfall throughout the year.

1. **Correlation Heatmap.**

****

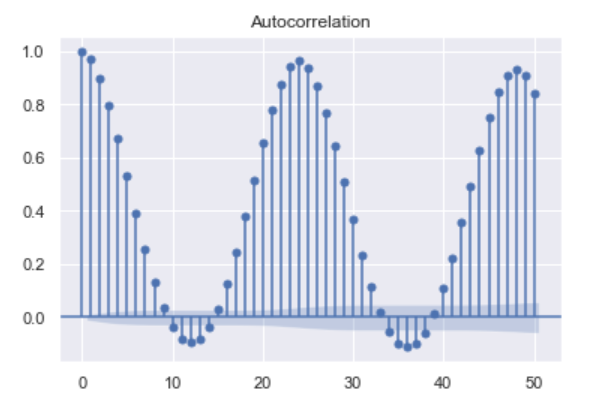
**Fig.3**

**Interpretation:**

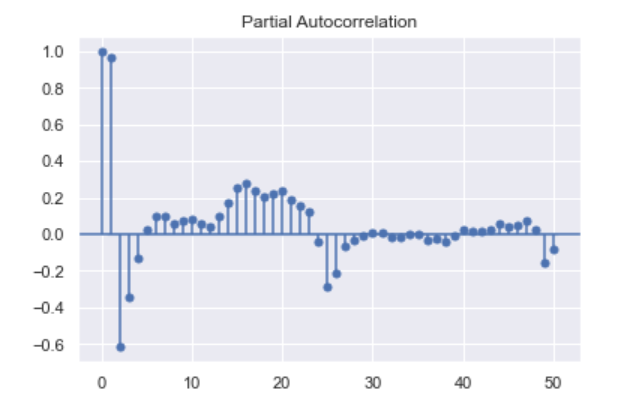
The plot shows the correlation matrix of weather data collected in Pune. The correlation matrix shows the linear relationship between the different variables.

* FeelslikeC and Heat Index have a high positive correlation (0.99).
* Pressure and Cloud cover have High negative correlation (–0.61).

1. **Autocorrelation Function and Partial Autocorrelation Function.**



**Fig.4**

****

**Fig.5**

**Interpretation:**

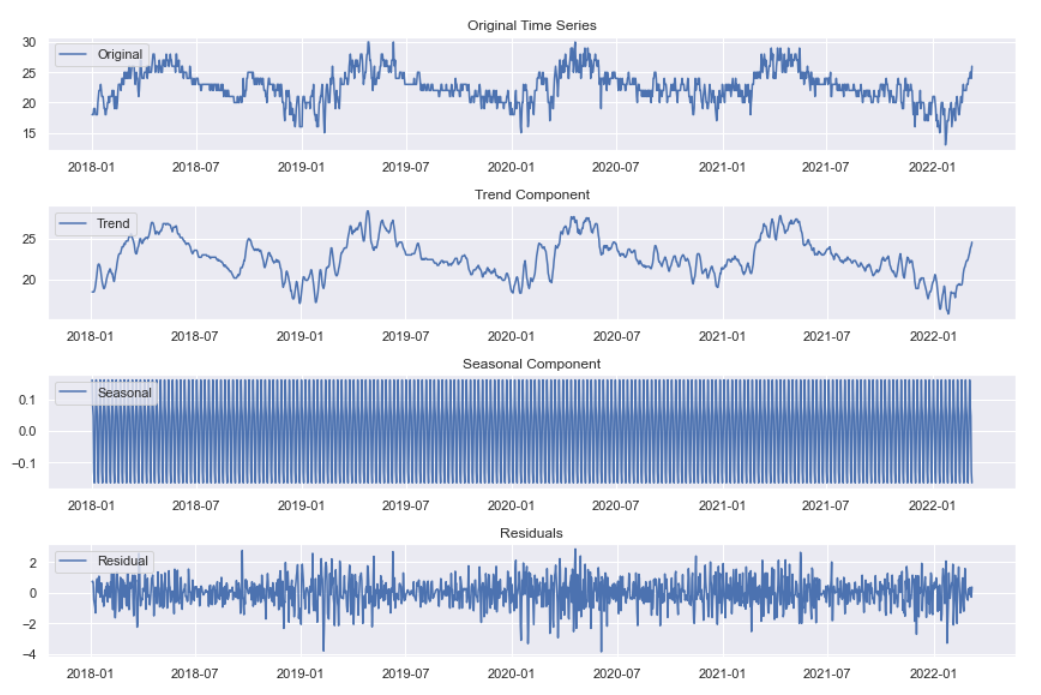
The plot shows the Autocorrelation and Partial autocorrelation of daily temperature data from Pune.

**ACF plot:** The ACF plot shows that the temperature data is highly autocorrelated at previous lags. This means that the temperature on any given day is strongly correlated with the temperature on the previous days.

**The PACF:** The PACF plot shows that there is a significant positive partial autocorrelation at lag 1, meaning that the current value of the time series is positively correlated with the value at the previous time step, even after accounting for the autocorrelation at lag 0. This further confirms the presence of a trend in the time series.

Overall, the ACF and PACF plots suggest that the time series is non-stationary, meaning that its mean and variance are not constant over time. This is due to the presence of a trend in the time series. A seasonal ARIMA model can be used to model the data and forecast future values.

**Seasonal Decomposition**

****

**Fig.-6:**

**Interpretation:** The trend component shows whether there is a significant long-term trend in the temperature data.

The seasonal plot shows that the temperature has a clear seasonal pattern, with higher temperatures in the summer and lower temperatures in the winter.

Residuals shows the residuals or errors after removing the trend and seasonal components.

**4.4 Analysis Using Software**

**Data Preparation:**

Data preprocessing is a primary step in machine learning where we prepare our data before training our models. It involves various techniques to clean, transform, and enhance the quality of our data.

Data may be organized and manipulated using tools like pandas in Python.

**Data Visualization:**

Bar graphs, correlation heatmaps, and time series plots are created to visually understand the patterns in the data.

Matplotlib, Seaborn, or ggplot2 may be used for creating visualizations.

Various software tools can be used for data visualization, including:

* Matplotlib: A popular Python library for creating 2D plots and charts.
* Seaborn: A Python library built on top of Matplotlib, providing a higher-level interface for creating complex visualizations.

**Software Usage:**

Python is likely used for data manipulation, visualization, and modeling. Libraries such as pandas, Matplotlib, Seaborn, statsmodels, and scikit-learn are commonly used.

R could be an alternative, especially for statistical modeling and visualization.

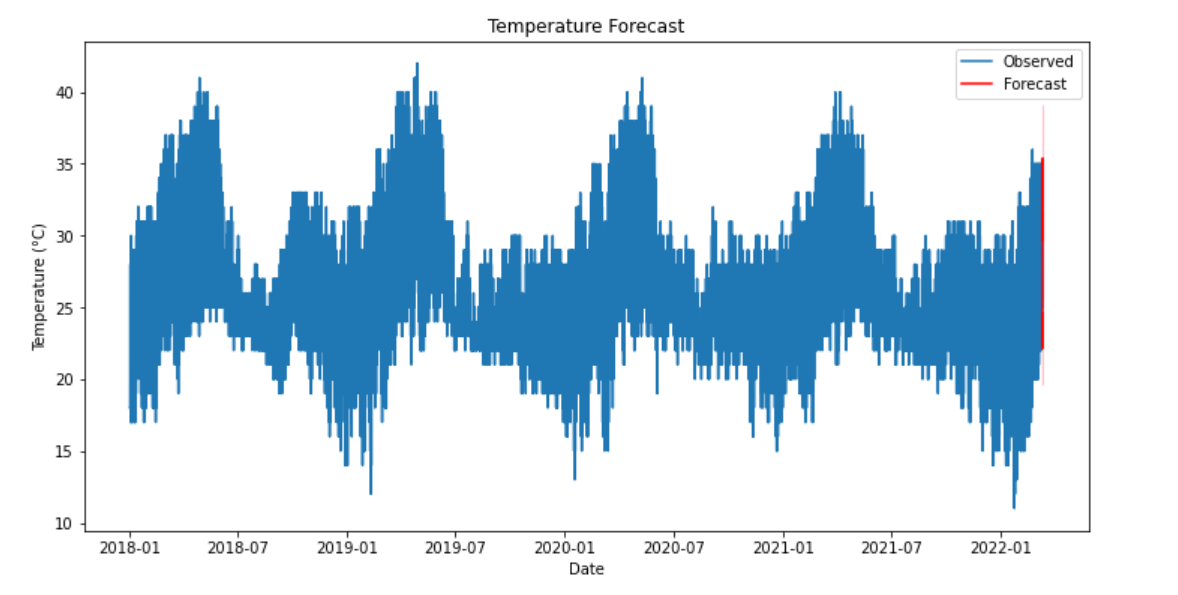
Jupyter Notebooks or R Markdown documents may be utilized to document and present the analysis.

**4.5 Statistical Modelling**

**Fitting the SARIMA model.**

SARIMA is a powerful tool and it is used for modelling and forecasting time series data, especially when dealing with seasonality and trend components.

Performs time series analysis on temperature data using a SARIMA model. It checks for stationarity, determines the model's orders, fits the SARIMA model to the data, and then forecasts future temperature values. The results are visualized in plots for easy interpretation.



**Fig.7**

**Model:**

SARIMA(p, d, q)(P, D, Q, s)

SARIMA(1,1,1)(1,1,1,24)

This formula indicates that SARIMA model includes an autoregressive component (ARIMA) and a seasonal component (SARIMA) to capture both the short-term and long-term patterns in the temperature time series data.

**y\_t = ϕ\_1 y\_{t-1} + ϵ\_t - θ\_1 ϵ\_{t-1} + Φ\_ {24} y\_{t-24} + Θ\_ {24} ϵ\_{t-24}**

**where:**

y\_t is the value of the time series at time t.

ϕ\_1 is the autoregressive parameter of order 1.

ϵ\_t is the white noise error.

θ\_1 is the moving average parameter of order 1.

Φ\_ {24} is the autoregressive parameter of order 24 (seasonal).

Θ\_ {24} is the moving average parameter of order 24 (seasonal).

Using the SARIMA model, Forecast the Temperature of the next 2 days (48 Hours).

**Table No.1**

|  |  |
| --- | --- |
| 2022-03-12 | 2022-03-13 |
| 2022-03-12 00:00:00 24  2022-03-12 01:00:00 24  2022-03-12 02:00:00 23  2022-03-12 03:00:00 23  2022-03-12 04:00:00 22  2022-03-12 05:00:00 22  2022-03-12 06:00:00 22  2022-03-12 07:00:00 22  2022-03-12 08:00:00 26  2022-03-12 09:00:00 28  2022-03-12 10:00:00 30  2022-03-12 11:00:00 33  2022-03-12 12:00:00 34  2022-03-12 13:00:00 34  2022-03-12 14:00:00 35  2022-03-12 15:00:00 35  2022-03-12 16:00:00 34  2022-03-12 17:00:00 33  2022-03-12 18:00:00 31  2022-03-12 19:00:00 29  2022-03-12 20:00:00 28  2022-03-12 21:00:00 26  2022-03-12 22:00:00 25  2022-03-12 23:00:00 25 | 2022-03-13 00:00:00 24  2022-03-13 01:00:00 24  2022-03-13 02:00:00 23  2022-03-13 03:00:00 23  2022-03-13 04:00:00 22  2022-03-13 05:00:00 22  2022-03-13 06:00:00 22  2022-03-13 07:00:00 22  2022-03-13 08:00:00 26  2022-03-13 09:00:00 28  2022-03-13 10:00:00 31  2022-03-13 11:00:00 33  2022-03-13 12:00:00 34  2022-03-13 13:00:00 35  2022-03-13 14:00:00 35  2022-03-13 15:00:00 35  2022-03-13 16:00:00 34  2022-03-13 17:00:00 33  2022-03-13 18:00:00 31  2022-03-13 19:00:00 29  2022-03-13 20:00:00 28  2022-03-13 21:00:00 26  2022-03-13 22:00:00 25  2022-03-13 23:00:00 25 |

**Interpretation:**

The SARIMA model is fitted to the temperature time series and the resulting model is used to forecast the temperature for the next 48 hours. The forecast is shown in the above Fig.7.

After fitting the SARIMA model and making predictions, plot the original data and the forecast. Here's the conclusion of the plot:

1. The blue line in the plot represents the actual temperature data over the time period covered by the dataset.
2. The orange line represents the forecasted temperatures generated by the SARIMA (Seasonal Autoregressive Integrated Moving Average) model.

Using the SARIMA model, Forecast the Temperature of the next 2 days (48 Hours) are shown in Table No.1.

**Fitting the Exponential Smoothing model:**

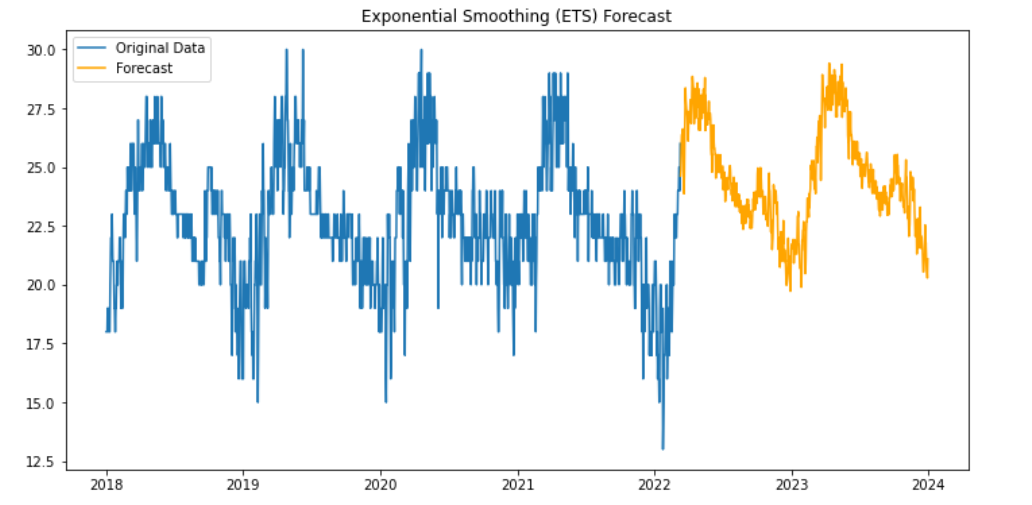
Exponential Smoothing is a popular time series forecasting method used to make predictions about future data points based on historical data. It is particularly well-suited for time series data with trends and seasonality.

It provides a way to make forecasts based on weighted averages of past observations, with more recent data points receiving higher weights. Exponential smoothing can be adapted to various scenarios by adjusting its components.

Types of Exponential Smoothing:

1. Double Exponential Smoothing (DES).
2. Triple Exponential Smoothing (TES).

* Temperature Forecast Graph

****

**Fig.8**

**Model:**

The Exponential Smoothing (ETS) model with trend and seasonal components, as fitted to the temperature data can be written in the following mathematical formula:

**y\_t = F\_t + S\_t + e\_t**

**where:**

yt is the observed temperature at time t.

Ft is the trend component at time t.

St is the seasonal component at time t.

et is the error term at time t.

The trend component is modeled using a simple exponential smoothing equation with smoothing parameter **α=0.7786512:**

**F\_t = 0.7786512 \* y\_t + (1 - 0.7786512) \* F\_{t-1}**

The seasonal component is modeled using a seasonal exponential smoothing equation with smoothing parameter **γ=5.952e−10:**

**S\_t = 5.952e-10 \* (y\_t - F\_t) + (1 - 5.952e-10) \* S\_{t-365}**

The error term is assumed to be white noise.

The overall forecast for the temperature at time t is then calculated as the sum of the trend and seasonal components:

**\Hat{y}\_t = F\_t + S\_t**

Using the Exponential Smoothing model, Forecast the Temperature of the next 2 days (48 Hours).

**Table No.2**

|  |  |
| --- | --- |
| 2022-03-12 | 2022-03-13 |
| 2022-03-12 00:00:00 24.216335  2022-03-12 01:00:00 23.930055  2022-03-12 02:00:00 23.145503  2022-03-12 03:00:00 22.522041  2022-03-12 04:00:00 22.374394  2022-03-12 05:00:00 21.947741  2022-03-12 06:00:00 21.058343  2022-03-12 07:00:00 21.958772  2022-03-12 08:00:00 25.659012  2022-03-12 09:00:00 28.009900  2022-03-12 10:00:00 30.459140  2022-03-12 11:00:00 32.782232  2022-03-12 12:00:00 34.045847  2022-03-12 13:00:00 34.829583  2022-03-12 14:00:00 35.460214  2022-03-12 15:00:00 35.634063  2022-03-12 16:00:00 35.044983  2022-03-12 17:00:00 34.182908  2022-03-12 18:00:00 32.427268  2022-03-12 19:00:00 30.821995  2022-03-12 20:00:00 29.154468  2022-03-12 21:00:00 26.961755  2022-03-12 22:00:00 25.822303  2022-03-12 23:00:00 25.089666 | 2022-03-13 00:00:00 24.230857  2022-03-13 01:00:00 23.944577  2022-03-13 02:00:00 23.160025  2022-03-13 03:00:00 22.536563  2022-03-13 04:00:00 22.388916  2022-03-13 05:00:00 21.962263  2022-03-13 06:00:00 21.072865  2022-03-13 07:00:00 21.973294  2022-03-13 08:00:00 25.673534  2022-03-13 09:00:00 28.024421  2022-03-13 10:00:00 30.473662  2022-03-13 11:00:00 32.796754  2022-03-13 12:00:00 34.060369  2022-03-13 13:00:00 34.844105  2022-03-13 14:00:00 35.474736  2022-03-13 15:00:00 35.648585  2022-03-13 16:00:00 35.059505  2022-03-13 17:00:00 34.197430  2022-03-13 18:00:00 32.441790  2022-03-13 19:00:00 30.836517  2022-03-13 20:00:00 29.168989  2022-03-13 21:00:00 26.976276  2022-03-13 22:00:00 25.836824  2022-03-13 23:00:00 25.104187 |

**Interpretation:**

After fitting the ETS model and making predictions, plot the original data and the forecast. Here's the conclusion of the plot:

1. Original Data (Blue Line): The blue line in the plot represents the actual temperature data over the time period covered by the dataset.
2. Forecast (Orange Line): The orange line represents the forecasted temperatures generated by the Exponential Smoothing (ETS) model. The model uses historical temperature data and the specified ETS configuration (additive trend and additive seasonal components) to make predictions about future temperatures.

**Conclusion Model Summary:**

The Exponential Smoothing model with additive trend and seasonal components was fitted to the temperature data. The model was able to capture the overall trend and seasonal patterns in the data. This means that the model gave more weight to recent observations when forecasting the trend and seasonal components.

Overall, the ETS model provides a good fit to the temperature data. The model can be used to forecast future values of the temperature data with a reasonable degree of accuracy.

**Fitting the ARIMA (Autoregressive Integrated Moving Average)**

ARIMA, which stands for Autoregressive Integrated Moving Average, is a popular time series forecasting model widely used in various fields such as finance, economics, and environmental science. It is a powerful tool for predicting future values based on past observations.

* Temperature Forecast Graph

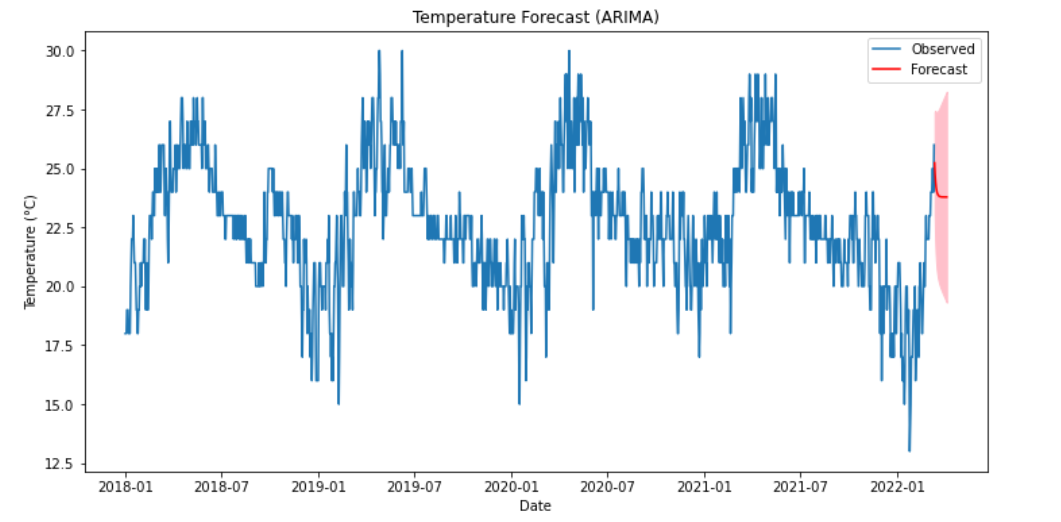


Fig.9

yt = c + ϕ₁yt-₁ + ϕ₂yt-₂ + ... + ϕpyt-p + θ₁εt-₁ + θ₂εt-₂ + ... + θqεt-q + εt

where:

yt is the current value of the time series at time t

c is the constant term

ϕ₁, ϕ₂, ..., ϕp are the autoregressive coefficients

θ₁, θ₂, ..., θq are the moving average coefficients

εt is the error term at time t

**Interpretation:**

The plot shows the predicted average temperature for each month from 2018 to 2022. The blue line shows the observed temperatures, and the red line shows the forecasted temperatures. The shaded area around the red line shows the prediction interval, which is a range of values within which the actual temperature is likely to fall.the ARIMA model seems to have captured the seasonality of the temperature data well, with the forecasted temperatures following the same pattern as the observed temperatures. The model also appears to be accurate in predicting the overall trend of the temperatures, although there are some small deviations between the forecasted and observed temperatures.

Using the ARIMA model, We Forecast the temperature of the next 48 days.

**Table No.3**

|  |
| --- |
|  |
| 2022-03-12 25.240448  2022-03-13 24.741840  2022-03-14 24.414530  2022-03-15 24.199667  2022-03-16 24.058620  2022-03-17 23.966030  2022-03-18 23.905249  2022-03-19 23.865350  2022-03-20 23.839158  2022-03-21 23.821964  2022-03-22 23.810678  2022-03-23 23.803268  2022-03-24 23.798405  2022-03-25 23.795212  2022-03-26 23.793116  2022-03-27 23.791740  2022-03-28 23.790837  2022-03-29 23.790244  2022-03-30 23.789855  2022-03-31 23.789599  2022-04-01 23.789432  2022-04-02 23.789321  2022-04-03 23.789249  2022-04-04 23.789202 |

**Comparison performance of Time Series models.**

1. SARIMA Model
2. Exponential Smoothing Model
3. ARIMA model

Compare the performance of SARIMA model, Exponential Smoothing model and ARIMA on the basis of the AIC (Akaike Information Criterion) and the BIC (**Bayes Information Criterion).**

**The AIC and BIC values for the SARIMA model**

* **AIC: 65674.056**
* **BIC: 65716.611**

**The AIC and BIC values for the Exponential Smoothing model**

* **AIC: 689.845**
* **BIC: 2757.972**

**The AIC and BIC values for the ARIMA model**

* **AIC: 4643.686**
* **BIC: 4659.685**

**The AIC and BIC values are both measures of goodness of fit for a model. The lower the AIC or BIC value, the better the model. In this case, the Exponential Smoothing model has the lowest AIC and BIC values, which means that it is the best model for this data.**

**On AIC and BIC values of the SARIMA model, Exponential Smoothing model and ARIMA model it is seen that the Exponential Smoothing model is best model for the forecasting the temperature.** **It is a simple and effective model that is easy to understand and implement, and it produces accurate forecasts.**

**Also Compare the performance of SARIMA model, Exponential Smoothing model and ARIMA model on the basis of the MSE (Mean Square Error) and MAE (Mean Absolute Error) and RMSE (Root Mean Square Error).**

**The MSE, RMSE and MAE of SARIMA model**

* **MSE: 38.85**
* **RMSE: 6.23**
* **MAE: 5.99**

**The MSE, RMSE and MAE of the Exponential Smoothing model**

* **MSE: 5.17**
* **RMSE: 2.27**
* **MAE: 1.98**

**The MSE, RMSE and MAE of the ARIMA model**

* **MSE: 8.98**
* **RMSE: 2.99**
* **MAE: 2.25**

The MSE is the mean squared error, which is the average of the squared differences between the predicted values and the actual values. The RMSE is the root mean squared error, which is the square root of the MSE. The MAE is the mean absolute error, which is the average of the absolute differences between the predicted values and the actual values.

In conclusion, the Exponential Smoothing model is the best model for this data based on all three metrics (MSE, RMSE, and MAE). It is a simple and effective model that is easy to understand and implement, and it produces accurate forecasts.

**Chapter-5**

**Conclusion and Discussion**

* **The analysis of the weather data revealed distinct seasonal patterns and variations. The average maximum temperature reaches its highest point in April and May, while the average minimum temperature drops to its lowest point in December and January. The rainfall pattern also exhibits a seasonal trend, with the highest rainfall occurring in July and the lowest rainfall occurring in January.**
* **The SARIMA, Exponential Smoothing and ARIMA models were successfully fitted to the temperature data. The SARIMA model and Exponential model, with its ability to capture both trend and seasonal components.**
* **Both the SARIMA and Exponential Smoothing models provided forecasts for the temperature series. The SARIMA model's forecasts exhibited a clearer trend and seasonal pattern, while the Exponential Smoothing model's forecasts were smoother and more closely aligned with the actual data points. Also, ARIMA model provided forecasts for the temperature series.**
* In conclusion, the Exponential Smoothing model is the best model for this data based on all three metrics (MSE, RMSE, and MAE). It is a simple and effective model that is easy to understand and implement, and it produces accurate forecasts.

**Chapter-6**

**Limitation and Future Scope**

**Limitations**

In the fitting of the SARIMA model for forecasting the temperature the memory error occurs in python. The error indicates that there is insufficient memory available to accommodate the require computations.

Because of this error the SARIMA model does not forecast the temperature of the next one year the model only forecast the temperature of the 2 days (48 Hours).

In the ARIMA, it gives the all the values of the forecasted temperature of one year. The forecasted temperature values are more than 1000 because of this in this report there are only 24 day forecasted values are mention.

**Future Scope**

Temperature is a key component of weather forecasts, which are used by individuals, businesses, and governments to make informed decisions about activities that are affected by the weather, such as travel, agriculture, and energy consumption.

Agriculture: Temperature forecasts are used by farmers to plan their crops and make decisions about irrigation and pest control. Accurate temperature forecasts can help farmers to increase yields and reduce losses.

Energy: Temperature forecasts are used by energy companies to plan for demand and supply.

Transportation: Temperature forecasts are used by transportation companies to plan for road conditions and air travel. Accurate temperature forecasts can help transportation companies to avoid delays and cancellations.

Public health: Temperature forecasts are used by public health officials to warn people about extreme heat and cold events. Accurate temperature forecasts can help to reduce the number of heat-related and cold-related deaths.

In addition to these specific applications, temperature forecasting is also used in a variety of other industries, such as construction, insurance, and retail. As technology continues to advance, temperature forecasting is becoming increasingly accurate and reliable, which is leading to new and innovative applications.

**Chapter-7**

**Reference**

Gangshetty, A., & Kaur, G. (2021). Time Series Prediction of Temperature in Pune Using Seasonal ARIMA Model (No. 6701). EasyChair.

Monica, C., Kalaivani, S., & Jayalalitha, G. (2018). Forecasting the annual temperature of India by Using Exponential Smoothing. Journal of Advanced Research in Dynamical and Control Systems, 6, 972-976.

Goswami, K., Hazarika, J., & Patowary, A. N. (2017). MONTHLY TEMPERATURE PREDICTION BASED ON ARIMA MODEL: A CASE STUDY IN DIBRUGARH STATION OF ASSAM, INDIA. International Journal of Advanced Research in Computer Science, 8(8).

Dubey, A. K., Kumar, A., García-Díaz, V., Sharma, A. K., & Kanhaiya, K. (2021). Study and analysis of SARIMA and LSTM in forecasting time series data. Sustainable Energy Technologies and Assessments, 47, 101474.

<https://www.kaggle.com/datasets/dipakdeshmukh/weather-data-of-punemaharashtraindia2008-2022>