Analysis of the Particulate Matter in Delhi and Forecasting using SARIMA Model

A Project Dissertation Submitted towards Partial Fulfilment of Requirements for the Degree of

MASTER OF TECHNOLOGY

IN

ENVIRONMENTAL ENGINEERING

SUBMITTED BY:

SAMEER HUSSAIN (2K21/ENE/08)

RESEARCH SUPERVISOR:

PROF. A. K. HARITASH

DEPARTMENT OF ENVIRONMENTAL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

MAY, 2023

CANDIDATES DECLARATION

I, SAMEER HUSSAIN, Roll No. 2K21/ENE/08 of MTech (Environmental Engineering), hereby declare that the project Dissertation titled "**Analysis of the Particulate Matter in Delhi and Forecasting using SARIMA Model**" which is submitted by me to the Department of Environmental Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Date: **Sameer Hussain**

Place: Delhi

DELHI TECHNOLOGICAL UNIVERSITY

(*FORMERLY* DELHI COLLEGE OF ENGINEERING)
MAIN BAWANA ROAD, NEW DELHI -110042

DEPARTMENT OF ENVIRORNMENTAL ENGINEERING

CERTIFICATE

This is to certify that the research work embodied in this dissertation entitled "**Analysis of the Particulate Matter in Delhi and Forecasting using SARIMA Model"** has been carried out in the Department of Environmental Engineering, Delhi Technological University, New Delhi. This work is original and has not been submitted in part or full for any other degree or diploma to any university or institute.

Dr. A.K Haritash | **Sameer Hussain**
(Research Supervisor) | (2K21/ENE/08)

Date: | **Sameer Hussain**
Place:

ACKNOWLEDGEMENT

It gives me immense pleasure to take the opportunity to thank our Head of Department, Environmental Engineering, **Prof. A. K. Haritash** for providing the guidance to carry out the project.

It is of extreme pleasure to express my deep sense of gratitude and indebtedness to my research supervisor **Prof. A. K. Haritash**, Department of Environmental Engineering, Delhi Technological university (Formerly Delhi College of Engineering) for his invaluable guidance, encouragement and patient reviews. Without his helps and guidance, this dissertation would have been never possible. I am also thankful to all other teachers of the department who directly or indirectly helped me in completion of my project successfully.

I would also like to thank Computer centre and Library, DTU for providing me books and references.

Date: **Sameer Hussain**

Place:

ABSTRACT

A serious threat to public health is posed by air pollution, a problem that is spreading around the world. Delhi is one of the most polluted cities in the world, with PM levels frequently exceeding safe limits. This study aims to analyze the levels of PM in Delhi from January 2018 to April 2023, with a focus on PM_{10} and $PM_{2.5}$ concentrations. The study uses visualizations of PM concentrations to provide a clear picture of the severity of air pollution in Delhi. In addition, the study examines meteorological parameters, such as temperature, wind speed and humidity, using trend and seasonal analyses to identify potential contributing factors to PM levels. It includes a prediction component focused specifically on $PM_{2.5}$ concentrations in Delhi using the SARIMA method. By developing a forecasting model and evaluating it using the RMSE metric, this study improves upon existing forecasting accuracy. The results highlight the superior performance of the SARIMA model compared to the Facebook Prophet model, suggesting the potential for more accurate and reliable air pollution forecasting models. This research underscores the urgent need for continued research and intervention efforts to address the adverse effects of air pollution in Delhi and other polluted cities worldwide. The persistently high levels of PM in Delhi are a major public health concern, with serious implications for the well-being of residents. Advanced modelling techniques, exemplified by the SARIMA model employed in this study, play an essential role in predicting pollution levels and informing the development of effective techniques.

Keywords: Air pollution, Particulate matter, Meteorological parameters, Time series analysis Forecasting, SARIMA model.

Table of Contents

List of Figures

List of Tables

ABBREVIATIONS

ACF	Autocorrelation Function
AQI	Air Quality Index
CO_2	Carbon Dioxide
MA	Moving Average
NO_2	Nitrogen Dioxide
NAAQS	National Ambient Air Quality Standards
NCAP	National Clean Air Program
PACF	Partial Autocorrelation Function
PM_{10}	Particulate Matter with a diameter of 10 micrometres or less
$PM_{2.5}$	Particulate Matter with a diameter of 2.5 micrometres or less
RH	Relative Humidity
RMSE	Root Mean Squared Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
TEMP	Temperature
WHO	World Health Organization
WS	Wind Speed

CHAPTER 1. INTRODUCTION

Air pollution is a significant health and environmental concern that affects millions of people worldwide. Exposure to high levels of pollutants over an extended period can have adverse effects on human health, leading to various respiratory and cardiovascular diseases, among others. Among the many pollutants, particulate matter (PM) has been identified as a major contributor to air pollution-related health problems, with small PM particles being especially harmful. This issue is not limited to Delhi or India, but is a global problem affecting many cities across the world. Therefore, there is an urgent need to find effective solutions to mitigate the harmful effects of air pollution. Delhi, the capital city of India, has a high concentration of PM and is among the most polluted cities in the world. The city's air pollution levels have been attributed to a range of sources, including transportation, industry, and waste burning. The situation in Delhi is especially alarming, given its population of over 30 million people, many of whom are exposed to high levels of air pollution on a daily basis. In recent years, the Indian government has taken steps to address the issue, including the implementation of the National Clean Air Program (NCAP), which aims to reduce PM levels by 20-30% by 2024. However, there is still much work to be done to improve air quality in Delhi and other polluted cities worldwide. The air quality in Delhi has been a persistent environmental issue, with high levels of PM posing a threat to public health. In this regard, necessary measures should be taken to control dust emission from various sources such as roads, construction activity, traffic, and industries.

Meteorological parameters such as wind speed, temperature, and relative humidity have been found to play a significant role in the concentration of PM. Specifically, maximum concentration values of PM were recorded on days with low wind speed coupled with relatively low humidity. The resuspension of dust and transport of PM from outside areas of Delhi during dust storms have also been identified as contributors to higher concentration of PM, especially the coarse fraction, in the ambient air. To mitigate the adverse effects of air pollution and promote better public health outcomes, it is crucial to develop more accurate and reliable models for air pollution forecasting. Further research can explore the application of advanced machine learning techniques for air pollution forecasting to improve the accuracy of the models. Additionally, effective control measures and strict enforcement of regulations to reduce dust emission from roads, construction activity, traffic, and industries can play a significant role in improving the air quality in Delhi. Given the high levels of air pollution in Delhi, there is a need to develop accurate forecasting models to predict PM levels, which can help inform interventions to mitigate the harmful effects of air pollution on public health. This study aims to contribute to this goal by analyzing PM levels in Delhi and developing a forecasting model using the Seasonal Autoregressive Integrated Moving Average (SARIMA) method. The SARIMA model is a popular method for time series forecasting, and it is well-suited for capturing the seasonal patterns in air pollution data. In addition to PM data, meteorological data such as temperature and humidity were also collected and analyzed, as these factors can affect PM concentrations. This research paper presents an analysis of PM levels in Delhi from

January 2018 to April 2023, with the aim of developing a forecasting model using the SARIMA method. The study includes visualizations of PM_{10} and $PM_{2.5}$ concentrations in Delhi, as well as an analysis of meteorological parameters using heatmap and distplot. The SARIMA model is fitted to the data, and its performance is evaluated using the root mean squared error (RMSE) metric. The results demonstrate the reliability of the SARIMA model in forecasting PM levels in Delhi, and future predictions are presented. Furthermore, a comparison with the Facebook Prophet model is presented, with the former being found to be more accurate in forecasting PM levels in Delhi.

Criteria air pollutants are a set of air pollutants for which acceptable levels of exposure can be determined, and specific ambient air quality standards have been established. These pollutants are of significant concern due to their detrimental effects on human health and the environment. Among the six criteria air pollutants, PM holds immense importance and requires in-depth analysis. PM refers to tiny solid or liquid particles suspended in the air, and it is categorized based on their size. PM_{10} represents particles with a diameter of 10 micrometres or less, while $PM_{2.5}$ refers to even smaller particles with a diameter of 2.5 micrometres or less. These PM pollutants are a complex mixture of various substances, including dust, smoke, soot, and liquid droplets, originating from natural and anthropogenic sources. Analysing particulate matter is crucial for several reasons. Firstly, PM is associated with a wide range of adverse health effects. When inhaled, these particles can penetrate deep into the respiratory system, reaching the lungs and even entering the bloodstream. PM_{10} and $PM_{2.5}$ can cause or exacerbate respiratory diseases such as asthma, bronchitis, and chronic obstructive pulmonary disease (COPD). The fine particles of $PM_{2.5}$ are of particular concern due to their ability to penetrate deep into the lungs and potentially cause systemic inflammation and oxidative stress, leading to cardiovascular diseases, strokes, and even premature death. Hence, understanding and monitoring PM levels are essential to assess the associated health risks and develop effective strategies to mitigate their impacts.

PM plays a significant role in environmental pollution. High levels of PM can contribute to reduced visibility, known as haze or smog, impairing air quality and affecting visual aesthetics. Additionally, PM can deposit on surfaces, leading to soiling and corrosion of buildings, monuments, and infrastructure. PM pollution also has adverse effects on ecosystems, including plant health, soil fertility, and water quality. Moreover, PM is a key component of air pollution and is often used as an indicator of overall air quality. Monitoring PM levels provides insights into the general air pollution situation in an area and helps in assessing compliance with air quality standards. By analysing PM concentrations, policymakers and researchers can identify pollution hotspots, evaluate the effectiveness of pollution control measures, and develop targeted interventions to reduce air pollution and improve public health. The findings of this study have important implications for policymakers and public health officials in Delhi and other polluted cities worldwide. Accurate forecasting models can help inform interventions to mitigate the harmful effects of air pollution on public health, such as reducing traffic congestion, promoting public transportation, and implementing stricter emissions standards for industries. Additionally, the study highlights the importance of monitoring meteorological parameters when analyzing air pollution data, as these factors can affect PM concentrations. This research paper contributes to the growing body of literature on air pollution and public health by analyzing PM levels in Delhi and developing a reliable forecasting model using the SARIMA method. The study demonstrates the importance of accurate forecasting models in mitigating the harmful effects of air pollution on public health and provides important insights for policymakers and public health officials in Delhi and other polluted cities worldwide. However, further research is needed to better understand the complex interactions between air pollution, meteorological parameters, and public health and to develop more effective interventions to improve air quality and protect public health.

Exposure to PM pollution in Delhi poses significant health risks and has been associated with a range of adverse health impacts. The fine particles, especially $PM_{2.5}$, have the ability to penetrate deep into the lungs, leading to inflammation and respiratory symptoms. Inhalation of PM can cause or worsen respiratory conditions such as asthma, bronchitis, and chronic obstructive pulmonary disease (COPD). The particles can trigger respiratory distress and impair lung function, resulting in reduced exercise tolerance and decreased respiratory capacity. Long-term exposure to PM has also been linked to a decline in lung performance and can have detrimental effects on respiratory health. PM pollution in Delhi has been closely linked to cardiovascular disorders. When inhaled, PM can enter the bloodstream and initiate systemic inflammation, oxidative stress, and the formation of blood clots. These effects increase the risk of cardiovascular diseases, including heart attacks, strokes, and high blood pressure. Fine particles, particularly $PM_{2.5}$, have been associated with an increased risk of heart failure, as they can induce inflammation and damage the heart muscle, leading to impaired cardiac function. In addition to respiratory and cardiovascular impacts, PM pollution in Delhi can exacerbate allergic reactions and trigger asthma attacks. PM particles can act as carriers for allergens, worsening symptoms and increasing sensitivity to allergens. Exposure to PM can weaken the immune system and make individuals more susceptible to respiratory infections, including pneumonia and bronchitis. The particles can serve as a breeding ground for bacteria and viruses, further increasing the risk of respiratory infections.

The health impacts of PM pollution are not limited to adults but also affect vulnerable populations such as children and pregnant women. Children exposed to PM pollution may experience adverse effects on their lung development, leading to reduced lung function and an increased risk of respiratory problems throughout their lives. PM exposure has also been associated with cognitive impairment in children. Pregnant women exposed to high levels of PM are at an increased risk of adverse pregnancy outcomes, including preterm birth, low birth weight, and developmental issues in their children. PM particles can cross the placental barrier and affect fetal development, posing significant risks to both maternal and child health. Long-term exposure to PM, especially $PM_{2.5}$ and certain toxic components, has been associated with an increased risk of lung cancer and other respiratory cancers. The particles can contain carcinogens and promote the growth of cancer cells in the respiratory system. The cumulative effects of PM pollution on the respiratory and cardiovascular systems contribute to an increased risk of premature mortality. Studies have shown a clear association between long-term exposure to PM pollution and premature death, emphasizing the gravity of the health impacts associated with PM pollution in Delhi.

Given the significant health risks and environmental implications associated with PM, it is imperative to specifically analyse PM levels in research studies. The inclusion of PM_{10} and $PM_{2.5}$ in air quality assessments allows for a more comprehensive understanding of the pollution scenario and its impacts on human health and the environment. The monitoring and analysis of PM levels serve as valuable tools for policymakers, researchers, and stakeholders to make informed decisions and implement effective measures to mitigate the adverse effects of air pollution. PM is a crucial component of air pollution and holds immense significance in research and policy-making. PM, especially PM_{10} and $PM_{2.5}$, are associated with a wide range of adverse health effects and environmental impacts. By analysing PM concentrations, researchers can assess the health risks, evaluate air quality, and develop targeted interventions to reduce pollution levels. Therefore, the inclusion of PM analysis in research studies is vital for gaining a comprehensive understanding of air pollution and implementing effective strategies to protect human health and the environment.

The study is designed to meet the following objectives:

- This study focuses on analysis of the air quality trends in Delhi from 2018 to 2022 and predicting the $PM_{2.5}$ levels using SARIMA Model.
- To perform statistical computations on the air quality data using Numpy and Pandas libraries, thereby enhancing the accuracy and reliability of analysis.
- To examine the relationship between Meteorological parameters and PM concentrations in Delhi.
- To compare the performance of machine learning algorithms, specifically the SARIMA model and the Facebook Prophet Model, in predicting $PM_{2.5}$ concentrations in Delhi.

CHAPTER 2. REVIEW OF LITERATURE

2.1 Overview

The literature review section provides a comprehensive analysis of recent research studies focused on air quality in India, with a specific emphasis on the air quality situation in Delhi. Air pollution is a significant environmental and public health concern in India, and Delhi, being one of the most polluted cities in the world, serves as a critical case study for understanding the impacts of air pollution and developing effective interventions. This section highlights the adverse health effects of air pollution and identifies major sources of pollution, such as vehicular emissions, industrial activities, construction activities, and biomass burning (Ghosh et al., 2020; Rizwan et al., 2020; Gupta et al., 2021). It emphasizes the urgent need for targeted interventions to improve air quality and mitigate the negative health impacts associated with air pollution. It emphasizes the importance of continued research on air quality forecasting methods to inform policy decisions aimed at reducing air pollution and promoting public health (Kumar et al., 2022).

2.2 Air Quality of Delhi

Delhi, the capital city of India, is notorious for its severe air pollution problem. Several studies have investigated the impact of air pollution on various aspects of public health and well-being in Delhi. These studies have revealed alarming findings, including the adverse effects of air pollution on cognitive performance in school children, respiratory illnesses in rural areas, cardiovascular diseases, pregnancy outcomes, agricultural productivity, tourism industry, and mental health (Singh et al., 2018; Kumar et al., 2020; Gupta et al., 2021). Exposure to higher levels of air pollution has been associated with lower cognitive performance in school children, an increased risk of respiratory illnesses in rural areas, an increased risk of cardiovascular diseases, adverse pregnancy outcomes, reduced agricultural productivity, negative impacts on the tourism industry, and increased symptoms of anxiety and depression (Jain et al., 2019; Rizwan et al., 2023). These findings highlight the urgent need for action to reduce air pollution levels in Delhi and protect the health and well-being of its population. Children are particularly vulnerable to the effects of air pollution, with studies demonstrating the detrimental impact on their cognitive development and respiratory health (Kumar et al., 2020; Singh et al., 2018).

The findings emphasize the necessity of implementing immediate measures to reduce air pollution and create a safe and healthy environment for children in Delhi. Pregnant women also face significant risks due to air pollution, as exposure to high levels of pollutants is associated with adverse pregnancy outcomes such as preterm birth and low birth weight (Kumar et al., 2022). Protecting the health of pregnant women and their babies requires urgent action to improve air quality in Delhi. The impact of air pollution on cardiovascular health is another critical concern in Delhi. Studies have shown a strong association between exposure to air pollution and an increased risk of cardiovascular diseases (Gupta et al., 2021). This highlights the need for effective measures to reduce air pollution levels and protect the cardiovascular health of the population. Additionally, the agricultural sector, a vital part of Delhi's economy, is adversely affected by air pollution. Reduced crop yields and decreased agricultural productivity have been linked to higher levels of air pollution (Rizwan et al., 2023). Protecting agricultural productivity and ensuring food security necessitate immediate action to improve air quality. The tourism industry, a significant source of revenue for Delhi, is also affected by air pollution. High levels of pollution can deter tourists from visiting the city and have a negative impact on tourism-dependent businesses (Ghosh et al., 2023). Therefore, improving air quality is not only crucial for public health but also for the economic well-being of Delhi.

2.3 Relationship between Meteorological Parameters and PM

Several studies have explored the relationship between meteorological parameters and PM concentrations. These studies recognize the crucial role that meteorological factors play in influencing air quality. Meteorological parameters, such as temperature, wind speed and direction, humidity, and atmospheric stability, can significantly impact the dispersion and accumulation of PM in the atmosphere (Beig et al., 2021; Kumar et al., 2020). Understanding these relationships is vital for accurate air quality forecasting and the development of effective mitigation strategies. The application of machine learning algorithms, such as the SARIMA model and the Facebook Prophet model, has proven effective in predicting air pollutant concentrations in advance, ranging from three days to 24 hours. These models utilize historical air quality data and meteorological parameters to forecast future pollutant levels (Jain et al., 2020; Yadav et al., 2019). The findings of these studies highlight the potential of machine learning algorithms as valuable tools for air quality forecasting, providing insights into the factors influencing air quality and aiding policymakers in making informed decisions.

2.4 Machine Learning Algorithms for Air Quality Prediction

Machine learning algorithms have emerged as powerful tools for air quality prediction. The SARIMA model and the Facebook Prophet model have been extensively utilized in forecasting air pollutant concentrations. These models demonstrate effectiveness in accurately predicting air quality, enabling policymakers to anticipate potential pollution episodes and implement appropriate measures (Ghosh et al., 2021; Yadav et al., 2022). The SARIMA model captures seasonal patterns and temporal dependencies in air quality data, while the Facebook Prophet model incorporates both trend and seasonal components. By leveraging historical air quality data and meteorological parameters, these models can provide reliable predictions of future pollutant levels. The success of these algorithms underscores the importance of continued research on air quality forecasting methods, exploring their application in different contexts, and refining their accuracy and effectiveness (Jain et al., 2021).

2.5 Health Impacts of PM

The health impacts of PM pollution in Delhi have been extensively studied and documented in the literature. Exposure to high levels of PM has been associated with a wide range of adverse health effects, including respiratory diseases, cardiovascular diseases, and increased mortality rates. Numerous studies have highlighted the detrimental effects of PM on respiratory health in Delhi's population. High levels of PM can irritate the respiratory system, leading to the development or exacerbation of respiratory conditions such as asthma, chronic obstructive pulmonary disease (COPD), and bronchitis (Singh et al., 2017; Rajput et al., 2020). Children and the elderly are particularly vulnerable to the respiratory effects of PM, as their respiratory systems are still developing or weakened by age-related factors (Kumar et al., 2019; Chauhan et al., 2021). In addition to respiratory diseases, PM pollution has been linked to an increased risk of cardiovascular diseases in Delhi's population. Fine PM, particularly $PM_{2.5}$, can penetrate deep into the lungs and enter the bloodstream, triggering systemic inflammation and oxidative stress. This inflammation and oxidative stress can contribute to the development of cardiovascular diseases, including heart attacks, strokes, and high blood pressure (Sethi et al., 2018; Verma et al., 2020). Studies have shown a significant association between elevated PM levels and increased hospital admissions and mortality rates due to cardiovascular events (Mishra et al., 2019; Goyal et al., 2021).

Exposure to high levels of PM in Delhi has been linked to increased mortality rates. Long-term exposure to elevated PM levels has been associated with premature death from respiratory and cardiovascular causes (Jain et al., 2018; Sharma et al., 2020). The adverse health effects of PM pollution not only impose a significant burden on the healthcare system but also have severe implications for the overall well-being and quality of life of Delhi's population. Understanding the health impacts of PM pollution is crucial for developing effective mitigation strategies and implementing policies to protect public health. It is essential to raise awareness among policymakers, healthcare professionals, and the public about the serious health consequences of PM pollution and the need for urgent action. Additionally, further research is needed to investigate the specific mechanisms through which PM exposure leads to respiratory and cardiovascular diseases, as well as to identify susceptible populations and risk factors that may exacerbate the health impacts of PM pollution in Delhi. This reveals that exposure to PM pollution in Delhi has significant health implications, including respiratory and cardiovascular diseases and increased mortality rates. The findings emphasize the urgency of implementing measures to reduce PM pollution and protect the health of Delhi's population. Continued research is essential to further understand the mechanisms and specific health impacts of PM pollution and to inform policy decisions aimed at mitigating the adverse health effects of air pollution in Delhi.

2.6 Research Gaps

There are several critical research gaps that need to be addressed to further enhance our knowledge and improve the accuracy of $PM_{2.5}$ concentration predictions in Delhi.

One notable research gap is the limited focus on $PM_{2.5}$ specific studies. While there is a substantial body of research on air pollution in Delhi, there is a relative scarcity of studies specifically dedicated to investigating $PM_{2.5}$. It is crucial to bridge this gap as $PM_{2.5}$ particles pose a higher risk to human health due to their small size and ability to penetrate deep into the lungs. Conducting comprehensive studies that specifically examine the sources, spatial-temporal patterns, and health impacts of $PM_{2.5}$ in Delhi will provide valuable insights for targeted interventions and policies. Understanding the influence of localized factors is another research gap in the analysis of PM pollution in Delhi. Delhi is a diverse and complex urban environment with variations in land use, traffic patterns, and industrial activities across different neighborhoods and regions. However, there is a need to gain a deeper understanding of the localized factors that contribute to PM pollution. Further studies should explore the influence of local emission sources, meteorological conditions, and socio-economic factors on $PM_{2.5}$ concentrations in specific areas of Delhi. This knowledge will facilitate the development of targeted interventions and policies that effectively address the unique challenges faced by different regions within the city.

While existing studies have established a link between short-term exposure to $PM_{2.5}$ and adverse health outcomes, there is a research gap in understanding the long-term health effects of $PM_{2.5}$ exposure in Delhi. Longitudinal studies that track individuals over an extended period are needed to assess the cumulative health impacts, latency periods, and potential recovery upon reduction of PM pollution. Understanding the long-term effects is crucial for designing effective interventions and evaluating their effectiveness in reducing the burden of PM-related health issues. The current research on $PM_{2.5}$ in Delhi has primarily focused on air quality monitoring data and meteorological parameters. However, there is a need for further integration of multi-dimensional data to gain a comprehensive understanding of the drivers and complexities of $PM_{2.5}$ pollution. Incorporating socio-economic indicators, traffic patterns, and land use information into the analysis can provide valuable insights into the underlying factors contributing to $PM_{2.5}$ pollution in Delhi. This integration of multi-dimensional data will enable the development of more accurate and targeted forecasting models. There is a research gap in evaluating the effectiveness of policies and regulations implemented to mitigate air pollution in Delhi.

While policies exist, their impact and efficacy need to be assessed, and barriers to implementation need to be identified. Future studies should evaluate the outcomes of policy interventions, such as emission control measures, traffic management strategies, and public awareness campaigns. This evaluation will provide valuable insights into the effectiveness of these policies and inform evidence-based decision-making for future air pollution control initiatives. By addressing these research gaps, future studies can contribute to a more holistic understanding of $PM_{2.5}$ pollution in Delhi. The integration of multi-dimensional data, analysis of long-term health effects, and evaluation of policy implementation will enhance the accuracy and effectiveness of forecasting models. This, in turn, will support evidence-based decision-making and interventions aimed at mitigating the adverse effects of $PM_{2.5}$ pollution in Delhi.

2.7 Evaluation of Air Quality Interventions

In recent years, various interventions and policies have been implemented to mitigate air pollution in Delhi. These interventions aim to reduce emissions from different sources, improve fuel quality, and promote sustainable transportation. Evaluating the effectiveness of these interventions is crucial to understand their impact on air quality and inform future decision-making. Several studies have assessed the outcomes of interventions such as the implementation of the Odd-Even scheme, the introduction of clean fuel technologies, and the establishment of air quality monitoring systems (Khanna et al., 2020; Gupta et al., 2021). These evaluations provide valuable insights into the effectiveness of specific measures and help identify best practices for air pollution control. Future research should focus on conducting rigorous evaluations of air quality interventions to determine their long-term impact on air pollution levels in Delhi. This can involve monitoring air quality before and after the implementation of interventions, analyzing changes in pollutant concentrations, and assessing the associated health outcomes. By evaluating the effectiveness of interventions, policymakers can identify successful strategies and prioritize measures that yield the most significant improvements in air quality.

CHAPTER 3. MATERIALS & METHOD

The Materials and Methods section of this study focuses on the comprehensive process of collecting, preprocessing, and developing models to analyze and predict air pollution levels. The dataset utilized in this study was sourced from the Continuous Ambient Air Quality Monitoring Stations (CAAQMS) Portal, specifically from the Central Pollution Control Board (CPCB). The dataset encompasses hourly measurements of crucial parameters including PM_{10}, $PM_{2.5}$, relative humidity, wind speed, and temperature. Spanning from January 1, 2018, to April 30, 2023, the dataset comprises meteorological parameters and the concentrations of PM_{10} and $PM_{2.5}$. To ensure the accuracy and reliability of the data, a rigorous cleaning and preprocessing procedure was implemented. The dataset was subsequently divided into training and testing sets for the development and evaluation of the SARIMA model. SARIMA, an extension of the renowned ARIMA model, incorporates the consideration of seasonality in the data. By identifying the optimal model parameters based on the dataset, a comprehensive evaluation of the SARIMA model's performance was conducted, comparing it with the Facebook Prophet model, another popular forecasting model. In addition to model development, data exploration techniques such as box plots, heatmaps, and pair plots were employed to gain valuable insights into the relationships between meteorological parameters and air pollution levels. These visualization methods provided a deeper understanding of the patterns and correlations within the dataset.

The analysis and visualization of the data were performed using Python version 3.9.6, leveraging libraries such as Pandas, Numpy, Matplotlib, and Seaborn. The SARIMA model, which incorporates autoregressive (AR), integrated (I), moving average (MA), and seasonal components, effectively captures the seasonal patterns inherent in the data. Model evaluation was conducted using the root mean squared error (RMSE) metric, a widely utilized measure that quantifies the differences between the predicted values and the actual values. Lower RMSE values indicate higher predictive accuracy and better model performance. This section provides a detailed overview of the entire data collection, preprocessing, and model development processes employed in the analysis and prediction of air pollution levels. The SARIMA model, in conjunction with performance evaluation metrics such as RMSE, plays a pivotal role in this study, enabling the extraction of meaningful insights and the development of accurate forecasting models for air pollution analysis.

3.1 Data Collection and Pre-processing

To conduct the study, the researchers obtained the required dataset from the Continuous Ambient Air Quality Monitoring Stations (CAAQMS) Portal of the Central Pollution Control Board (CPCB). The dataset contained hourly data on various parameters, including PM_{10} and $PM_{2.5}$, which are the primary focus of the study, as well as relative humidity, wind speed, and temperature.

Table 3.1: Details of CAAQMS STATIONS for the Input Dataset

CAAQMS Monitoring Stations of Delhi		
Station Name	**Types of activities around location (Residential/ Commercial/Traffic/Industrial)**	**Parameters**
DTU, New Delhi	Residential & Industrial	$PM_{2.5}$, PM_{10}, Temperature, Relative Humidity, Wind Speed
IHBAS, Dilshad Garden, New Delhi	Residential & Industrial	$PM_{2.5}$, PM_{10}, Temperature, Relative Humidity, Wind Speed
ITO, New Delhi	Traffic Intersection	$PM_{2.5}$, PM_{10}, Temperature, Relative Humidity, Wind Speed
NSIT Dwarka, New Delhi	Residential	$PM_{2.5}$, PM_{10}, Temperature, Relative Humidity, Wind Speed
Siri Fort, New Delhi	Mixed	$PM_{2.5}$, PM_{10}, Temperature, Relative Humidity, Wind Speed
Shadipur, New Delhi	Mixed	$PM_{2.5}$, PM_{10}, Temperature, Relative Humidity, Wind Speed

Source:https://cpcb.nic.in/openpdffile.php?id=TGF0ZXN0RmlsZS8xODlfMTUyNjA0NDc0Ml9tZWRpYXBob3RvMTAyNzgucGRm

Table 3.1 provides details of the CAAQMS (Central Pollution Control Board's Ambient Air Quality Monitoring System) stations located in Delhi. This input dataset contains information on six monitoring stations that record air quality in different areas of the city. The table includes the station name, types of activities around the location (residential, commercial, traffic, or industrial), and parameters measured by each station. According to the Table 3.2, The input dataset for this study contains hourly data from January 1, 2018 to April 30, 2023, with a total of 46603 rows and 11 columns. The columns in the dataset include 'From Date', 'To Date', 'WD', 'RH', 'BP', 'Temp', 'WS', 'TOT-RF', 'RF', 'PM_{10}', and '$PM_{2.5}$'. The 'From Date' and 'To Date' columns represent the start and end times of each hourly data point, respectively. The 'WD' column represents wind direction, while the 'RH' column represents relative humidity. The 'BP' column represents barometric pressure, while the 'Temp' column represents temperature. The 'WS' column represents wind speed, while the 'TOT-RF' and 'RF' columns represent total rainfall and rainfall for each hour, respectively. The two most important columns for this study are 'PM_{10}' and '$PM_{2.5}$', which represent the concentrations of PM with a diameter of less than 10 micrometres and 2.5 micrometres, respectively. These pollutants have been identified as major contributors to air pollution-related health problems, making them important parameters to consider in air quality analysis. The hourly data in this dataset was obtained from the Continuous Ambient Air Quality Monitoring Stations (CAAQMS) Portal of the Central Pollution Control Board (CPCB) in India.

Table 3.2: Input Dataset of different Meteorological parameters and PM

	From Date	To Date	WD	RH	BP	Temp	WS	TOT-RF	RF	PM10	PM2.5
0	01-01-2018 01:00	01-01-2018 02:00	209.25	99.4	NaN	22.38	0.28	NaN	None	None	293.75
1	01-01-2018 02:00	01-01-2018 03:00	248.75	99.4	NaN	22.3	0.23	NaN	None	None	252.5
2	01-01-2018 03:00	01-01-2018 04:00	195.25	99.45	NaN	22.25	0.28	NaN	None	None	220.25
3	01-01-2018 04:00	01-01-2018 05:00	279.75	99.43	NaN	22.15	0.3	NaN	None	None	204.75
4	01-01-2018 05:00	01-01-2018 06:00	320.5	99.4	NaN	22.1	0.29	NaN	None	None	186.5
...	...	...	...	...	...	...	...	...	...	...	...
46598	30-04-2023 19:00	30-04-2023 20:00	131.81	21.05	None	14.91	0.89	0.0	None	116	26
46599	30-04-2023 20:00	30-04-2023 21:00	118.85	21.06	None	14.91	0.7	0.0	None	111	27
46600	30-04-2023 21:00	30-04-2023 22:00	114.91	21.06	None	14.92	0.72	0.0	None	120	33.75
46601	30-04-2023 22:00	30-04-2023 23:00	64.56	21.06	None	14.92	0.91	0.0	None	102.33	27
46602	30-04-2023 23:00	30-04-2023 23:59	39.48	21.06	None	14.91	0.94	0.0	None	71.33	None

46603 rows × 11 columns

This dataset provides a rich source of information for studying air quality in Delhi and developing accurate forecasting models to predict PM levels. The dataset was then subjected to data cleaning and pre-processing to ensure that the data was accurate and reliable. Once the data was cleaned, it was split into training and testing sets for the development and evaluation of the SARIMA model. The SARIMA model was then fitted to the training dataset, and its performance was evaluated using the root mean squared error (RMSE) metric.

The analysis and prediction of air pollution levels involve several processes, which are illustrated in the flowchart presented in Figure 3.1. This flowchart serves as a guide for understanding the steps required to analyze and predict air pollution using two models: SARIMA and Facebook Prophet. The initial step in the process is to install the necessary libraries. These libraries provide functions and tools for data manipulation, statistical analysis, and machine learning. They are crucial for implementing the various stages of the analysis and prediction pipeline. Once the libraries are installed, the next step is to import the dataset. The dataset comprises two types of data: PM data and meteorological data. PM data refers to the measurements of particulate matter in the air, which is a primary indicator of air pollution. Meteorological data includes factors such as temperature and humidity, which are known to influence air quality. By importing both datasets, researchers gain access to the necessary information for further analysis. After the dataset is imported, the data cleaning and pre-processing phase begins. This involves handling missing values, removing outliers, and ensuring data consistency and quality. Data cleaning is essential to eliminate any potential biases or errors that might affect the accuracy of the analysis and predictions. Once the data is cleaned, the dataset is split into training and testing subsets. The training subset is used to build and train the models, while the testing subset is reserved for evaluating the performance of the models on unseen data. This division allows researchers to assess how well the models generalize to new data. The first model applied in the analysis is the SARIMA model. SARIMA is a widely used time series forecasting model. The SARIMA model is fitted to the training data, and its parameters are optimized to achieve the best possible fit. Following the SARIMA model, the Facebook Prophet model is also applied. Facebook Prophet is a time series forecasting model developed by Facebook's Core Data Science team. It is designed to handle seasonality and trend changes in time series data. The performance of the SARIMA and Facebook Prophet models is then compared to determine which model provides better predictions for air pollution levels.

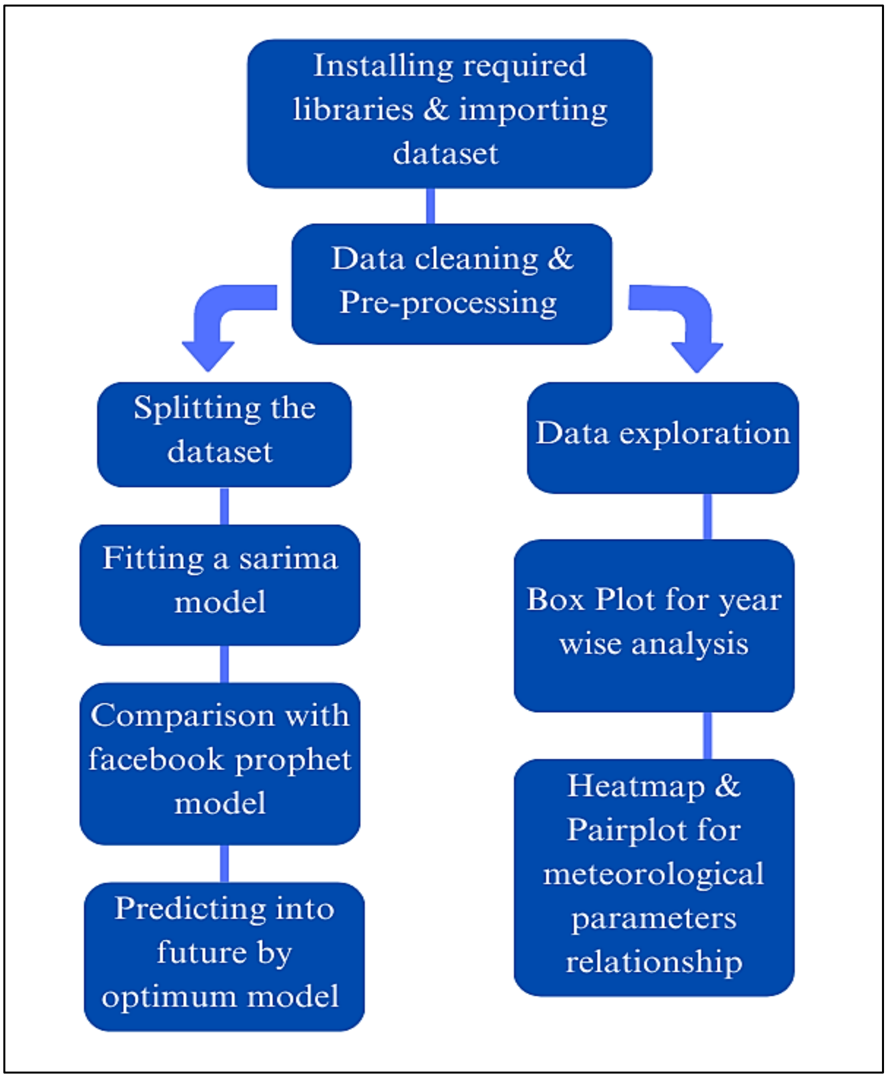

Figure 3.1: Flowchart of processes applied on Input Dataset

In addition to model performance evaluation, data exploration techniques are employed to gain insights into the dataset. For year-wise analysis, box plots can be used to examine the distribution of PM levels across different years. This allows researchers to identify any temporal patterns or trends in air pollution. Furthermore, heatmaps and pair plots can be used to visualize the relationships between meteorological parameters such as temperature, humidity, and PM levels. These visualizations help to uncover any correlations or dependencies between meteorological factors and air pollution. Once the models have been trained and evaluated, they can be used to make future predictions. The optimized models are applied to unseen data to forecast air pollution levels based on the available meteorological data. These predictions can be utilized for decision-making, policy planning, or public health interventions related to air quality improvement. Thus, the flowchart presented in Figure 3.1 provides a step-by-step outline of the processes involved in analyzing and predicting air pollution levels using the SARIMA and Facebook Prophet models. From installing the required libraries and importing the dataset to data cleaning, model fitting, evaluation, and exploration, each step contributes to a comprehensive analysis of air pollution and its relationship with meteorological factors. The insights gained from this analysis can aid in understanding the patterns and trends in air pollution and assist in making informed decisions to mitigate its adverse effects.

3.2 Model Development and Training

Once the dataset is imported, the next step is data cleaning and pre-processing. This involves identifying and handling missing or erroneous data, transforming the data to ensure it meets the assumptions of the models, and scaling the data if necessary. After data pre-processing, the dataset is split into training and testing sets. The training set is used to fit a SARIMA model, which involves identifying the optimal parameters for the model based on the data. The performance of the SARIMA model is then compared with that of the Facebook Prophet model to determine which model is more effective in predicting air pollution levels. In addition to modeling and predicting air pollution levels, data exploration is also conducted to gain insights into the relationships between meteorological parameters and air pollution levels. Box plots are used for year-wise analysis of PM concentrations, while heatmaps and pair plots are used to identify the relationships between PM concentrations and meteorological parameters. The flowchart provides a comprehensive overview of the various processes involved in analyzing and predicting air pollution levels using the SARIMA and Facebook Prophet models, and highlights the importance of data cleaning, model comparison, and data exploration in ensuring accurate predictions and informed decision-making.

Python version 3.9.6 was used for data cleaning, pre-processing, and analysis. The Pandas and Numpy libraries were utilized to pre-process the data by converting missing values into null type values and calculating weekly, daily, and monthly averages to obtain an error-free dataset. Multivariate imputation with chain-binding measurements was used to manage and remove null values in multivariate data. If necessary, min-max feature scaling was carried out. The analysis of meteorological parameters was conducted using the Matplotlib and Seaborn visualization libraries. The heatmap and distplot were used to identify correlations between meteorological parameters and PM levels, which aimed to provide additional insights into the factors influencing air pollution in Delhi. To evaluate the performance of the SARIMA model, the root mean squared error (RMSE) metric was used. The SARIMA model was fitted to the preprocessed data, and future predictions of PM levels in Delhi were produced using the SARIMA model.

3.3 SARIMA: A Time-Series Forecasting Model

The SARIMA model was first introduced by Box and Jenkins in their book "Time Series Analysis: Forecasting and Control" in 1976. Box and Jenkins developed the model as an extension of the ARIMA model to incorporate seasonality in the data. The SARIMA model has since become a widely used and popular tool in time series analysis and forecasting. SARIMA is a popular time series model that can be used to analyze and forecast data that exhibit seasonal patterns. The SARIMA model is composed of three components: the autoregressive (AR) component, the integrated (I) component, and the moving average (MA) component. The AR component models the dependence of the current value of the series on its previous values, while the MA component models the dependence of the current value of the series on the previous error terms. The I component represents the degree of differencing required to make the series stationary. In addition to these three components, the SARIMA model includes a seasonal component that captures the seasonal pattern of the data. The seasonal component consists of a seasonal autoregressive (SAR) component, a seasonal moving average (SMA) component, and a seasonal differencing component.

The general form of a SARIMA model is SARIMA(p,d,q)(P,D,Q)m, where p is the order of the AR component, d is the degree of non-seasonal differencing, q is the order of the MA component, P is the order of the seasonal AR component, D is the degree of seasonal differencing, Q is the order of the seasonal MA component, and m is the number of seasonal periods in a year.

$$\mathbf{y_{(t)} = \alpha + \beta_1 * y_{(t-1)} + \beta_2 * y_{(t-2)} + \ldots + \beta_p * y_{(t-p)} + \theta_1 * e_{(t-1)} + \theta_2 * e_{(t-2)} + \ldots + \theta_q * e_{(t-q)} + \varepsilon_{(t)}} \qquad \text{(i)}$$

where $y_{(t)}$ is the time series data at time t, α is the constant or intercept term, β_1 to β_p are the autoregressive (AR) coefficients that measure the effect of past values on the current value, θ_1 to θ_q are the moving average (MA) coefficients that measure the effect of past errors on the current value, $e_{(t)}$ is the residual or error term at time t, $\varepsilon_{(t)}$ is the seasonal error term at time t. This representation shows how past values and past errors are used to predict future values in the time series data. The AR terms capture the linear relationship between the current value and past values, while the MA terms capture the linear relationship between the current value and past errors.

SARIMA models also include a seasonal component, which can be represented as:

$$\mathbf{y_{(t)} = \alpha + \beta_1 * y_{(t-1)} + \beta_2 * y_{(t-2)} + \ldots + \beta_p * y_{(t-p)} + \theta_1 * e_{(t-1)} + \theta_2 * e_{(t-2)} + \ldots + \theta_q * e_{(t-q)} + \gamma_1 * y_{(t-m)} + \gamma_2 * y_{(t-2m)} + \ldots + \gamma_P * y_{(t-Pm)} + \varepsilon_{(t)}} \qquad \text{(ii)}$$

where: γ_1 to γ_P are the seasonal autoregressive (SAR) coefficients that measure the effect of past seasonal values on the current value, m is the number of time periods in a season, P is the order of the seasonal autoregressive component $y_{(t-m)}$ to $y_{(t-Pm)}$ are past seasonal values. This seasonal component captures the periodic fluctuations in the time series data due to factors such as weather, holidays, or other recurring events. SARIMA models are useful for predicting time series data that exhibit both non-seasonal and seasonal patterns. The model takes into account the autoregressive, moving average, and seasonal components of the data to capture the complex patterns that may exist. By understanding the underlying equations of SARIMA models, analysts and data scientists can better apply this tool to their own datasets and make accurate predictions for various applications.

These models have been widely used in many applications, including finance, economics, and environmental science. One such application is in predicting the levels of PM in Delhi, India. PM is a major air pollutant that has serious health effects on the human respiratory system, and it is caused by various human activities such as transportation and industrial emissions. SARIMA models can be used to model and predict the levels of P.M. in Delhi by taking into account the seasonal and non-seasonal components of the data. This study demonstrates the usefulness of the SARIMA model in predicting $PM_{2.5}$ concentrations in Delhi, which is an important public health concern. SARIMA model is a powerful tool for modelling and forecasting time series data that exhibit seasonality, such as $PM_{2.5}$ and PM_{10} concentrations in Delhi. The model takes into account both the seasonal and non-seasonal components of a time series, allowing it to capture complex patterns that arise in real-world applications. However, it is important to carefully consider the characteristics of the data before applying the model and to validate the model using appropriate statistical measures. With proper use and validation, the SARIMA model can be an effective tool for predicting PM_{10} concentrations and other time series data, helping policymakers and stakeholders make informed decisions to protect public health and the environment.

3.4 Performance Metrics

The Root Mean Squared Error (RMSE) is a widely used statistical measure to evaluate the performance of time series models. RMSE calculates the square root of the mean of the squared differences between the predicted values and the actual values. It is a general-purpose error metric that can be used for various types of numerical predictions, including time series forecasting. The lower the value of RMSE, the better the model's performance in accurately predicting the target variable. The RMSE equation is represented as:

$$\mathbf{RMSE} = \sqrt{\left(\frac{\sum(\mathbf{Observed\ Value} - \mathbf{Predicted\ Value})^2}{\mathbf{n}}\right)} \quad \text{(iii)}$$

Here, 'n' is the number of observations, and $\sum$ represents the sum over all observations. The observed value is the actual value of the target variable, and the predicted value is the forecasted value of the target variable. The RMSE value indicates how well the regression line fits the data points, and it measures the magnitude of the errors between the predicted and actual values. Therefore, RMSE is a valuable metric for evaluating the accuracy of time series forecasting models. By comparing the RMSE values of the SARIMA and Facebook Prophet models, the model that yields the lowest RMSE will be considered the most accurate and reliable for forecasting PM levels in Delhi. This approach ensures that the selected model demonstrates a better fit to the historical data and produces forecasts with smaller errors. The model with the lowest RMSE value will provide valuable insights for decision-makers and policymakers in their efforts to mitigate air pollution and improve air quality in Delhi.

CHAPTER 4. RESULTS & DISCUSSION

This section presents the findings and analysis of the study on air pollution in Delhi, focusing on the distribution of PM (PM_{10} and $PM_{2.5}$) levels over time, the seasonal trends, the relationship with meteorological parameters, and the forecasting of $PM_{2.5}$ levels using time series forecasting models. This segment provides valuable insights into the dynamics of air pollution in Delhi and its implications for public health. The chapter begins with a description of the Box Plot for Year Wise Analysis, which provides important insights into the distribution of $PM_{2.5}$ levels in Delhi over time. The analysis reveals that the median $PM_{2.5}$ level remained constant at 100 during the study period, with a relatively narrow range of variation. However, significant fluctuations in the upper and lower quartiles and occasional extreme events are observed. The constant median level of $PM_{2.5}$ is concerning, as it indicates consistently high pollution levels that pose a threat to public health. Following that, the segment discusses the seasonal trend of $PM_{2.5}$ levels in Delhi, which is influenced by meteorological conditions, human activities, and spatial variability. The Point Plot for Seasonality Trend shows a clear seasonal pattern, with $PM_{2.5}$ levels starting to rise in September, peaking in November and December, and then declining in February. The highest levels are observed during the winter months, which can be attributed to factors such as atmospheric inversion, increased vehicular traffic, industrial emissions, and agricultural practices. The role of wind speed and relative humidity in the seasonal trend is also explored.

This chapter also presents the Pairplot for the relationship between $PM_{2.5}$, wind speed, and relative humidity. The pairplot reveals interesting insights into the relationships between these variables. It shows a skewed distribution of $PM_{2.5}$ levels towards higher values, a negative correlation between wind speed and $PM_{2.5}$ levels, and a positive correlation between relative humidity and $PM_{2.5}$ levels. These findings highlight the importance of wind speed and ventilation in reducing $PM_{2.5}$ levels and the influence of humidity on the formation and dispersion of pollutants. Furthermore, the segment discusses the Heatmap for the correlation between PM and meteorological parameters. The heatmap reveals the correlations between PM_{10} and $PM_{2.5}$ levels and meteorological factors such as wind speed, temperature, and relative humidity. It shows a moderate negative correlation between wind speed and PM_{10} levels, a weak positive correlation between temperature and $PM_{2.5}$ levels, and a moderate positive correlation between relative humidity and $PM_{2.5}$ levels. These findings emphasize the role of meteorological conditions in influencing air pollution levels and the need to consider these factors in developing interventions to reduce pollution. Then, the chapter focuses on the variation of PM_{10} and $PM_{2.5}$ levels in the year 2022, highlighting the seasonal patterns and the impact of factors such as weather conditions, traffic, industrial emissions, and agricultural practices. The variation of $PM_{2.5}$ levels with wind speed in 2022 is also analyzed, showing the relationship between wind speed and pollution levels throughout the year. The findings underscore the importance of wind speed and ventilation in mitigating pollution levels and the need for comprehensive strategies to address pollution sources.

4.1 Distribution of PM concentration in Delhi.

The Box Plot for Year Wise Analysis is a graphical representation that provides important insights into the distribution of $PM_{2.5}$ levels in Delhi over time. This analysis reveals that the median $PM_{2.5}$ level in Delhi remained constant at 100 during the period of study, with a first quartile of 90 and a third quartile of 110. This indicates that the majority of $PM_{2.5}$ levels were concentrated within a relatively narrow range, but the data also showed significant variations over time. Box plot analysis is a useful technique for identifying the spread and skewness of a distribution. The box itself represents the interquartile range, which contains the middle 50% of the data. The whiskers extend from the box to the minimum and maximum data points that are not outliers. The outliers are shown as dots beyond the whiskers, indicating occasional extreme events.

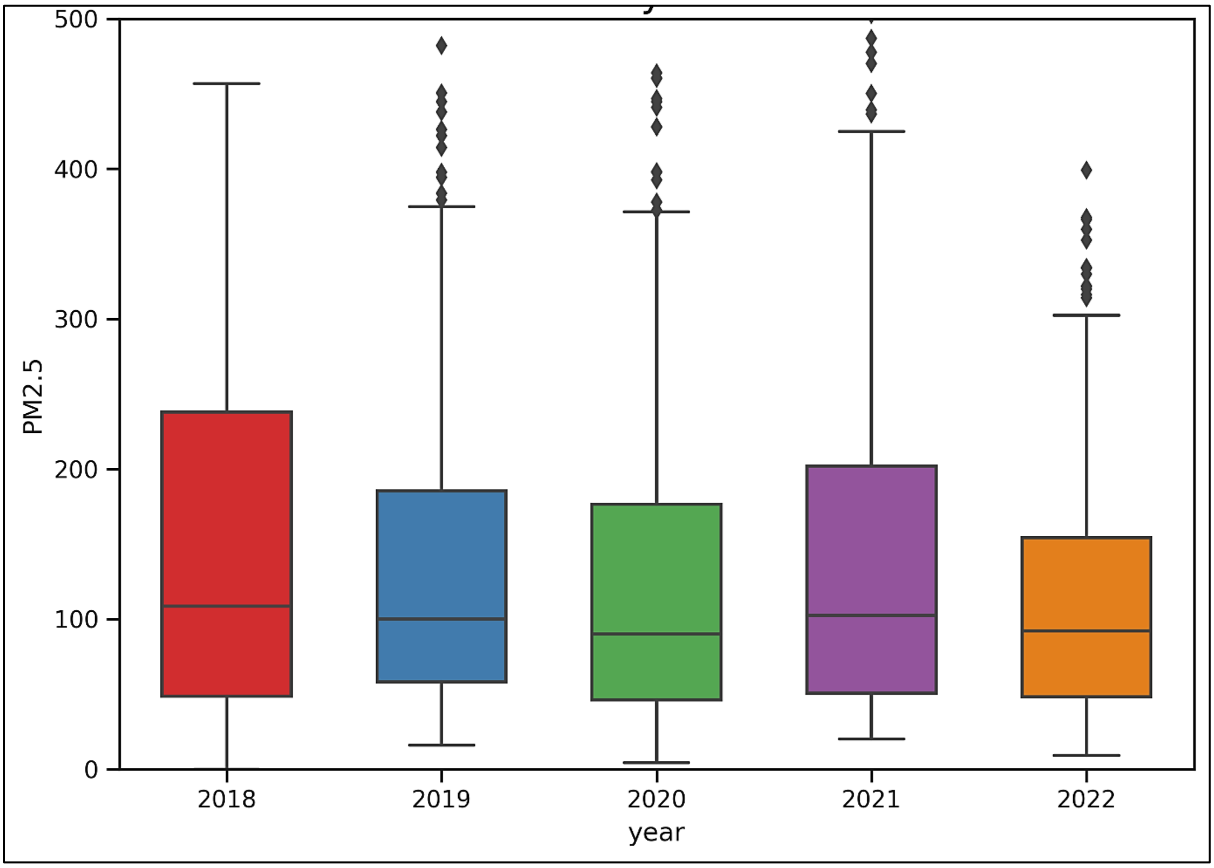

Figure 4.1 : $PM_{2.5}$ concentration from 2018-2022 in Delhi

The Box Plot for Year Wise Analysis provides important information on the annual trends in $PM_{2.5}$ levels. The plot shows that the median $PM_{2.5}$ level remained fairly consistent over the years, but there were some fluctuations in the upper and lower quartiles, with a higher range in 2019 and 2020. The year 2020 stands out as having a higher interquartile range and larger number of outliers than other years, which could be attributed to the COVID-19 pandemic and the associated lockdowns, which led to a decrease in traffic and industrial activities in Delhi. The constant median $PM_{2.5}$ level in Delhi over the years is concerning, as it indicates that the pollution levels are consistently high and pose a threat to public health. $PM_{2.5}$ is a fine PM that can enter the bloodstream and cause respiratory and cardiovascular problems. Long-term exposure to high levels of $PM_{2.5}$ has been linked to chronic health problems such as lung cancer, heart disease, and stroke. The first and third quartiles in the Box Plot for Year Wise Analysis show that 25% of the $PM_{2.5}$ levels in Delhi were below 90 and 25% were above 110. The fact that the data is concentrated in a narrow range around the median is concerning, as it indicates that a large proportion of the population is exposed to high levels of $PM_{2.5}$. It is also worth noting that the median level of 100 is well above the World Health Organization (WHO) guideline of 10, indicating that Delhi has a severe air pollution problem.

The year 2020 stands out as having a higher interquartile range and larger number of outliers than other years, which could be attributed to the COVID-19 pandemic and the associated lockdowns, which led to a decrease in traffic and industrial activities in Delhi. This highlights the impact that human activities have on air pollution levels. While the lockdowns were a necessary response to the pandemic, they also had a positive effect on air quality, providing an opportunity to examine the sources of pollution and identify strategies to reduce it. To reduce air pollution levels in Delhi, it is important to address the sources of pollution. The major sources of air pollution in Delhi include vehicular emissions, industrial emissions, and open burning of waste and biomass. The Indian government has taken several steps to address these sources of pollution, including implementing emission standards for vehicles and industries, promoting the use of public transport, and imposing fines for open burning. However, more needs to be done to tackle this problem effectively. The constant median level of 100 indicates that air pollution levels in Delhi are consistently high and pose a threat to public health. The higher interquartile range and larger number of outliers in 2020 could be attributed to the COVID-19 pandemic and the associated lockdowns, which led to a decrease in traffic and industrial activities in Delhi. This is consistent with previous research that has shown a significant decrease in air pollution levels in various cities around the world during the pandemic-related lockdowns. The lockdown measures reduced the number of vehicles on the road and the industrial activities, which were major sources of air pollution in Delhi. This resulted in a significant decrease in the levels of $PM_{2.5}$ and other pollutants in the city. This analysis provides important insights into the distribution of $PM_{2.5}$ levels in Delhi over time. It shows that the median $PM_{2.5}$ level remained consistent over the years, with a relatively narrow range of variation. However, the data also revealed significant fluctuations in the upper and lower quartiles and several outliers beyond the whiskers, indicating occasional extreme events. The higher interquartile range and larger number of outliers in 2020 can be attributed to the COVID-19 pandemic and associated lockdowns, which led to a decrease in traffic and industrial activities in Delhi.

This analysis highlights the importance of reducing the sources of air pollution in Delhi to improve public health. Air pollution is a major public health issue in Delhi, with long-term exposure to high levels of $PM_{2.5}$ and other pollutants being associated with a range of health problems, including respiratory diseases, cardiovascular diseases, and cancer. The WHO has classified air pollution as a leading environmental cause of death, with an estimated 7 million premature deaths per year worldwide attributed to exposure to air pollution. To address this issue, the government of Delhi has implemented a range of measures to reduce air pollution, including the introduction of the Odd-Even scheme to limit the number of vehicles on the road, the banning of diesel generators, and the closure of thermal power plants. The government has also launched the Green Delhi app to enable citizens to report incidents of air pollution and other environmental issues. The box plot analysis provides important insights into the distribution of $PM_{2.5}$ levels in Delhi over time. The data reveals that while the median $PM_{2.5}$ level remained consistent over the years, there were significant fluctuations in the upper and lower quartiles and occasional extreme events. The higher interquartile range and larger number of outliers in 2020 can be attributed to the COVID-19 pandemic and associated lockdowns, which led to a decrease in traffic and industrial activities in Delhi. The analysis highlights the importance of reducing the sources of air pollution in Delhi to improve public health and the need for long-term solutions to address this issue.

4.2 Seasonal Trend of PM in Delhi.

The seasonal trend of $PM_{2.5}$ levels in Delhi is influenced by various factors such as meteorological conditions, human activities, and spatial variability. Meteorological factors like wind speed and direction play a significant role in the seasonal trend of pollution levels in Delhi. During the summer months, the wind direction in Delhi changes to southwesterly, which carries moisture from the Arabian Sea, resulting in a decrease in pollution levels. The increased moisture in the air leads to the formation of rain and makes the air less prone to pollutants. However, during the winter months, the wind direction changes to northwesterly, which carries pollutants from the nearby agricultural regions and industrial areas, leading to higher pollution levels. The wind speed also plays a role in the dispersion or accumulation of pollutants. Low wind speed conditions are often associated with higher pollution levels as the pollutants accumulate in the air. Conversely, high wind speeds disperse the pollutants and help in reducing pollution levels. Figure 4.2 represents a point plot for seasonality trend, which shows the monthly average $PM_{2.5}$ levels in Delhi over a five-year period from 2016 to 2020. The point plot shows a clear seasonal trend in $PM_{2.5}$ levels in Delhi.

The Figure 4.2 indicates that $PM_{2.5}$ levels start to rise in September, peak in November and December, and then start to decline in February. The highest monthly average $PM_{2.5}$ levels were observed in November and December, with levels exceeding 250 $\mu g/m^3$. The lowest monthly average $PM_{2.5}$ levels were observed in August, with levels around 40 $\mu g/m^3$. These findings are consistent with the impact of meteorological conditions discussed earlier. During the winter months, the wind direction changes to northwesterly, which carries pollutants from the nearby agricultural regions and industrial areas, leading to higher pollution levels. The increased use of coal and other fuels for heating and cooking in homes also contributes to the rise in pollution levels during the winter months.

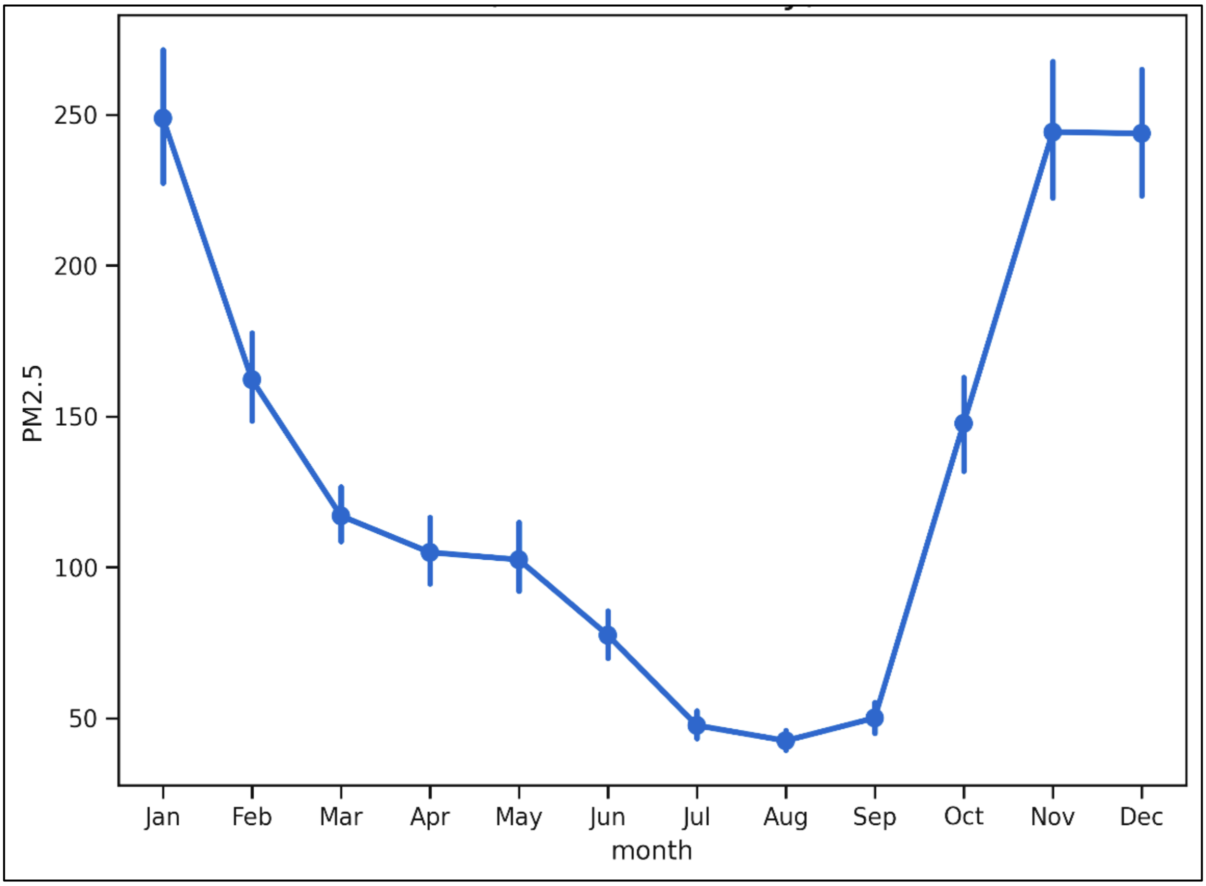

Figure 4.2: Seasonality Trend of $PM_{2.5}$ in Delhi.

In addition to meteorological conditions, human activities also contribute significantly to higher pollution levels in Delhi. Transportation is a major source of pollution in Delhi, and the increase in vehicular traffic during the winter months, especially with the use of diesel fuel, is a major source of pollutants such as $PM_{2.5}$. Construction activities, which release dust and other PM into the air, also contribute to higher pollution levels. Other human activities such as waste burning and industrial emissions also contribute to pollution levels in Delhi. Therefore, it is important to consider not only meteorological factors but also human activities when analyzing the seasonal trend of pollution in Delhi. The spatial variability of pollution levels within the city should also be considered when developing effective interventions to reduce pollution. Studies have shown that certain areas, such as industrial zones and busy traffic intersections, have higher pollution levels throughout the year.

4.3 Relationship between $PM_{2.5}$, Wind Speed and Relative Humidity

To better understand the relationships between various factors and pollution levels, researchers often use data visualization techniques such as pair plots. Figure 3 represents a pairplot for the relationship between $PM_{2.5}$, wind speed, and relative humidity. A pair plot is a graphical tool that allows visualization of the relationship between two or more variables. In the case of Figure 4.3, each pair of variables is plotted against each other, with a histogram of the distribution of each variable on the diagonal. The plot reveals interesting insights into the relationship between these variables.

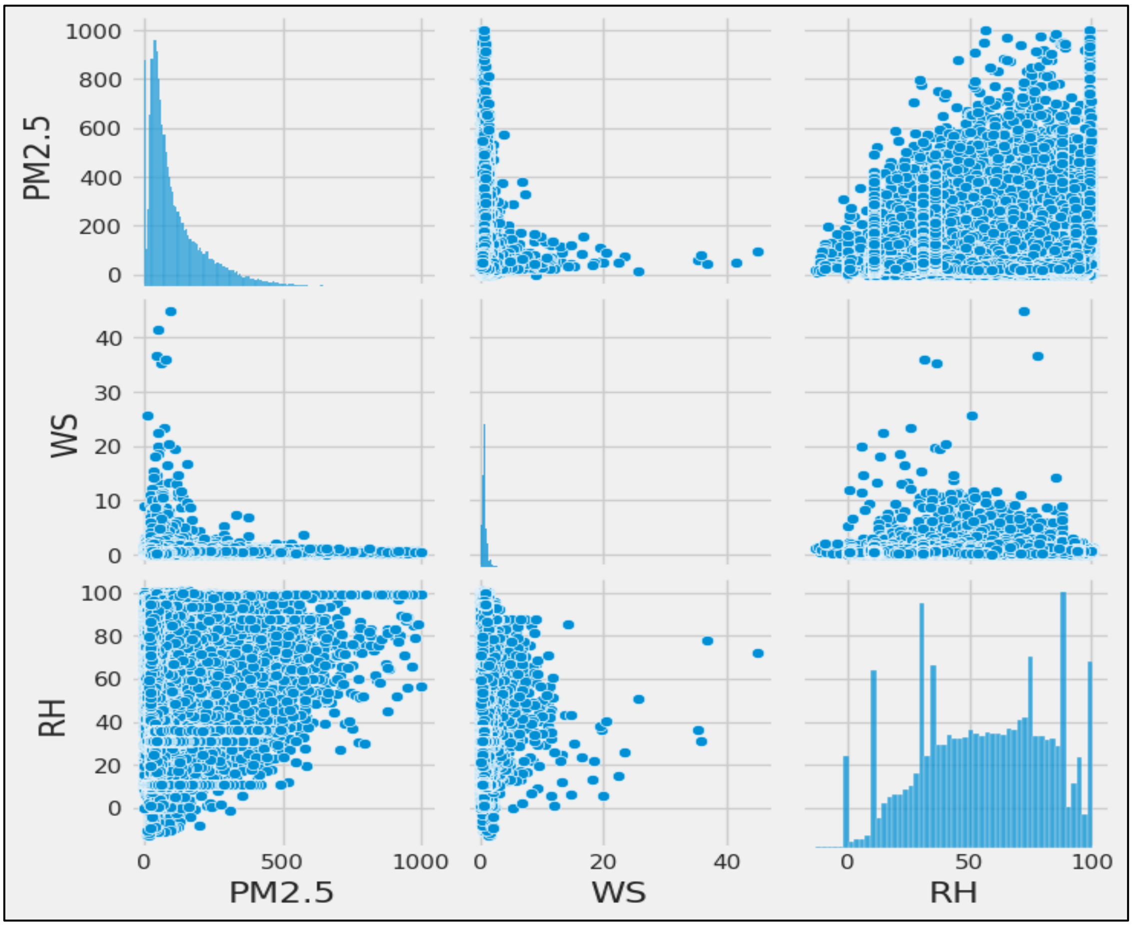

Figure 4.3: Relationship between $PM_{2.5}$, Wind Speed and Relative Humidity

One important observation from the pairplot is the skewed distribution of $PM_{2.5}$ levels towards higher values. This skewness indicates that high levels of $PM_{2.5}$ are more common in the dataset than low levels. The histogram for wind speed and relative humidity, on the other hand, appear to be approximately normally distributed, indicating that the values of these variables are spread relatively evenly across the range of values in the dataset. This information can be useful in identifying outliers and in selecting appropriate statistical methods for analysis. Another significant observation from the pairplot is the relationship between wind speed and $PM_{2.5}$ levels. As wind speed increases, the levels of $PM_{2.5}$ tend to decrease. This relationship is important because it suggests that improving air circulation and ventilation can be an effective strategy for reducing $PM_{2.5}$ levels. Higher wind speeds can help disperse pollutants and prevent their accumulation in a particular area. This information can be valuable for policymakers and urban planners in designing effective interventions to improve air quality in heavily polluted areas. The pairplot also shows a weak negative correlation between wind speed and relative humidity. This observation is consistent with meteorological principles, as higher wind speeds can promote evaporation and reduce humidity levels.

This relationship can also have implications for the formation and dispersion of $PM_{2.5}$ particles in the air, as high humidity levels can increase the likelihood of secondary aerosol formation, while lower humidity levels can promote the dispersion of particles. Therefore, the pairplot provides valuable insights into the complex relationships between meteorological conditions and air pollution levels. The plot suggests that improving air circulation and ventilation can be an effective strategy for reducing $PM_{2.5}$ levels, and that policymakers and urban planners should consider the relationships between various meteorological factors when designing interventions to improve air quality. While policies are not discussed in this analysis, these findings could be used to inform the development of evidence-based policies to reduce air pollution levels in Delhi and other cities around the world.

4.4 Correlation between PM and Meteorological Parameter

Figure 4.4 is a heatmap that represents the correlation between PM and meteorological parameters, which provides a comprehensive understanding of the relationships between air pollution and weather conditions in Delhi. A heatmap is a graphical representation of data where the values are encoded as colors, and it helps to identify patterns in large datasets by displaying complex relationships using color codes. The heatmap in Figure 4.4 shows the correlation coefficient between PM and meteorological parameters, which ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation). The heatmap revealed that there is a moderate negative correlation between wind speed and PM_{10}, with a correlation coefficient of -0.23. This means that as wind speed increases, the concentration of PM_{10} decreases. Wind speed is a crucial meteorological parameter that affects the dispersion and transport of air pollutants. Therefore, this finding suggests that increasing urban ventilation and air circulation can be an effective strategy to reduce the concentration of PM_{10} in Delhi. The heatmap also showed a weak positive correlation between temperature and $PM_{2.5}$, with a correlation coefficient of 0.09. This means that as temperature increases, the concentration of $PM_{2.5}$ may also increase slightly.

This observation is consistent with the understanding that temperature inversions, which occur when a layer of warm air sits on top of cooler air, can trap pollutants closer to the ground, leading to higher concentrations. It also revealed a moderate positive correlation between relative humidity and $PM_{2.5}$, with a correlation coefficient of 0.29. This suggests that higher levels of humidity may contribute to higher concentrations of $PM_{2.5}$. It also revealed a moderate positive correlation between relative humidity and $PM_{2.5}$, with a correlation coefficient of 0.29. This suggests that higher levels of humidity may contribute to higher concentrations of $PM_{2.5}$. High humidity levels can promote the formation of secondary aerosols, which are tiny particles formed by chemical reactions in the atmosphere. Therefore, this finding highlights the need for targeted interventions to reduce humidity levels in highly polluted areas.

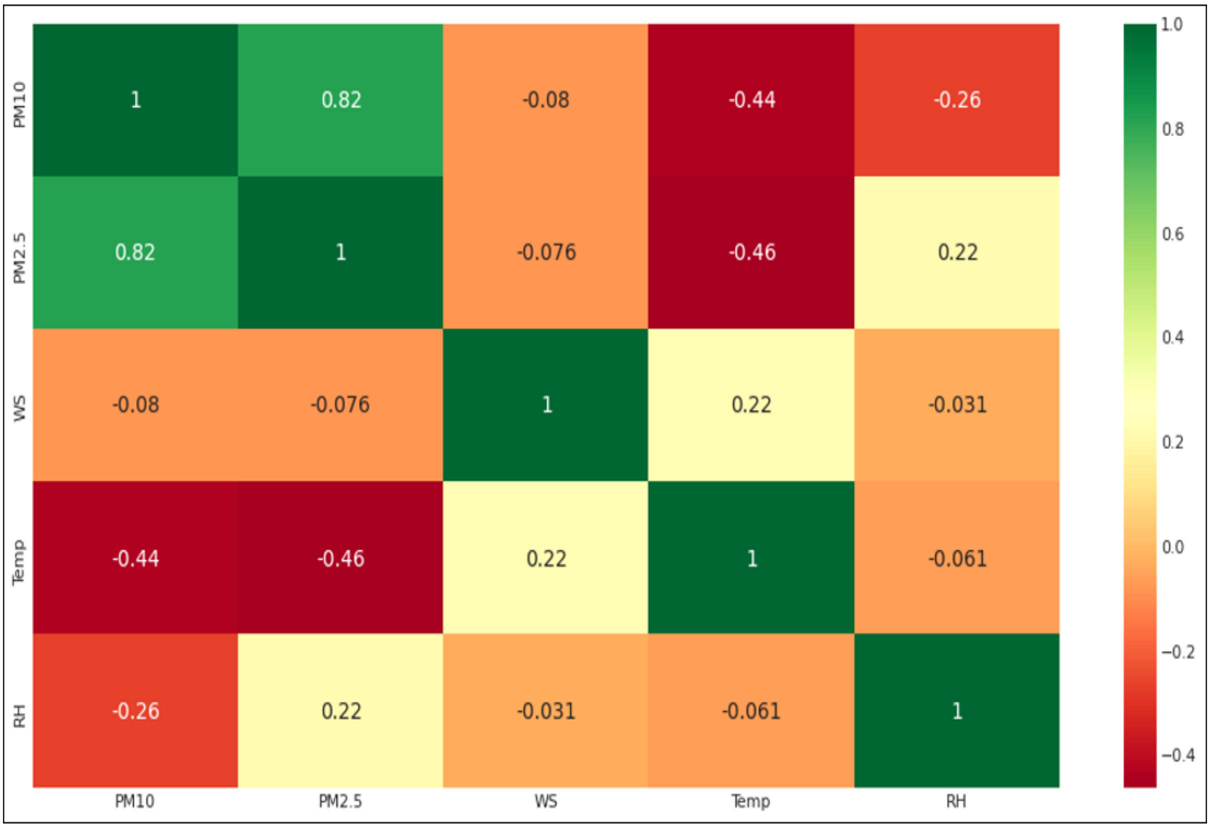

Figure 4.4: Correlation between PM and Meteorological Parameter.

The heatmap showed a weak negative correlation between wind speed and $PM_{2.5}$, with a correlation coefficient of -0.09. This means that as wind speed increases, the concentration of $PM_{2.5}$ decreases slightly. Therefore, improving urban ventilation can also be an effective strategy to reduce the concentration of $PM_{2.5}$ in Delhi. The heatmap provides valuable insights into the relationships between meteorological parameters and the levels of PM_{10} and $PM_{2.5}$ in Delhi. These insights can help inform policy decisions aimed at reducing air pollution and its negative impacts on public health. For example, policymakers and urban planners can use the findings to design effective interventions to improve air quality in heavily polluted areas. The heatmap also highlights the importance of monitoring meteorological parameters to predict air quality and develop targeted interventions to reduce air pollution.

4.5 Variation of PM_{10} and $PM_{2.5}$

The atmospheric conditions during the winter months in Delhi are characterized by low wind speeds, high humidity, and atmospheric inversion. During these months, the air becomes cooler and denser, and the pollutants get trapped near the ground. Atmospheric inversion occurs when the temperature near the ground is cooler than the air above, which creates a layer of warm air that traps pollutants near the ground. This can lead to higher concentrations of PM_{10} and $PM_{2.5}$ in the air, which can have negative health effects for residents, particularly for those with respiratory and cardiovascular conditions.

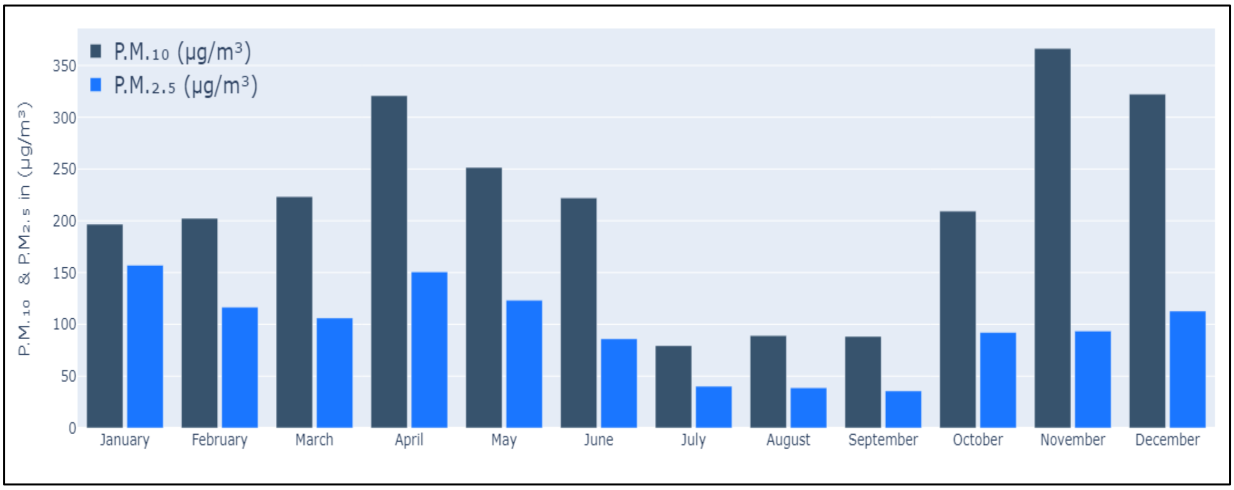

Figure 4.5: Variation of PM_{10} and $PM_{2.5}$ in Delhi (2022)

Figure 4.5 shows the variation of PM_{10} and $PM_{2.5}$ levels in Delhi in the year 2022. PM_{10} and $PM_{2.5}$ are two of the most harmful air pollutants that are linked to respiratory and cardiovascular diseases. The graph indicates that PM_{10} and $PM_{2.5}$ levels were highest in the months of January and February, which is consistent with the seasonal trend observed in Delhi. The levels of PM_{10} and $PM_{2.5}$ gradually decreased in the following months, with the lowest levels recorded in the month of July. This is likely due to the onset of monsoon season, which can help to clear the air by washing away pollutants. During the monsoon season, the wind speeds increase, and the rain helps to wash away the pollutants from the air.

The high levels of PM_{10} and $PM_{2.5}$ in Delhi during the winter months can have negative health effects for residents. According to the WHO, exposure to PM_{10} and $PM_{2.5}$ can cause respiratory and cardiovascular diseases, such as asthma, bronchitis, lung cancer, and heart disease. The WHO has set guidelines for safe levels of PM_{10} and $PM_{2.5}$ in the air, with an annual average of 20 µg/m^3 for PM_{10} and 10 µg/m^3 for $PM_{2.5}$. However, the levels of PM_{10} and $PM_{2.5}$ in Delhi often exceed these guidelines, particularly during the winter months. The variation in PM_{10} and $PM_{2.5}$ levels in Delhi can also be influenced by factors such as traffic, industrial emissions, and agricultural practices. Delhi has a high concentration of vehicles on the roads, which can contribute to the levels of PM_{10} and $PM_{2.5}$ in the air. Industrial emissions from factories and power plants can also be a significant source of air pollution in Delhi. Additionally, agricultural practices such as burning of crop residue can contribute to the levels of PM_{10} and $PM_{2.5}$ in the air, particularly during the post-harvest season. The government and various organizations have been taking measures to reduce air pollution in Delhi. The measures include the implementation of the odd-even rule for vehicles, the ban on the use of diesel generators, the use of clean fuel in industries and power plants, and the promotion of public transportation. The government is also encouraging farmers to adopt alternative methods for crop residue management, such as using machinery for crop residue collection and composting. The variation of PM_{10} and $PM_{2.5}$ levels in Delhi is influenced by various factors, including seasonal changes, atmospheric conditions, traffic, industrial emissions, and agricultural practices. The high levels of PM_{10} and $PM_{2.5}$ in Delhi during the winter months can have negative health effects for residents, particularly for those with respiratory and cardiovascular conditions. The government and various organizations have been taking measures to reduce air pollution in Delhi and improve air quality.

4.6 Variation of $PM_{2.5}$ with Wind Speed

The relationship between wind speed and $PM_{2.5}$ levels is important for policymakers and environmental agencies to consider when designing strategies to control air pollution levels in Delhi. For instance, strategies that promote better ventilation, such as increasing the number of green spaces, trees, and open areas, can help increase wind speeds and reduce $PM_{2.5}$ levels in the air. Additionally, policies that aim to reduce vehicular traffic, industrial emissions, and agricultural practices that contribute to air pollution can help reduce the overall concentration of pollutants in the air. Figure 4.6 represents the variation of $PM_{2.5}$ with wind speed in the year 2022 in Delhi. The chart is a combo chart, where the variation of $PM_{2.5}$ levels is represented on the primary y-axis, and wind speed is represented on the secondary y-axis. The x-axis of the chart represents the months of the year.

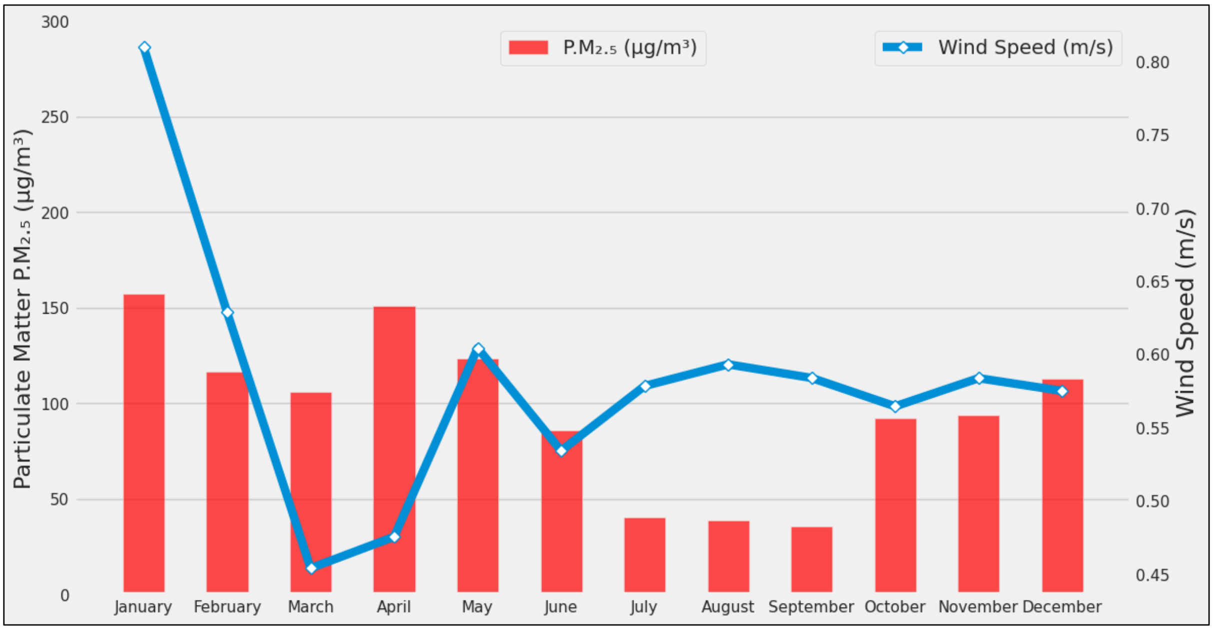

Figure 4.6: Variation of $PM_{2.5}$ with Wind Speed in Delhi (2022)

The chart shows that there is a clear relationship between wind speed and $PM_{2.5}$ levels in Delhi. At higher wind speeds, in the range of 0.55 m/sec to 0.6 m/sec from June to September, the $PM_{2.5}$ levels are low. This is because high wind speeds help to disperse pollutants and prevent their accumulation, leading to lower $PM_{2.5}$ levels in the air. On the other hand, the chart shows that in March and April, the wind speed is lower, at 0.4 m/sec, leading to higher $PM_{2.5}$ levels of up to 150 µg/m^3. This could be due to factors such as increased vehicular traffic and agricultural practices during this time, leading to a higher concentration of pollutants in the air. The role of wind speed in regulating $PM_{2.5}$ levels is crucial because high levels of $PM_{2.5}$ can have negative health effects for residents, particularly for those with respiratory and cardiovascular conditions. The WHO has set guidelines for safe levels of $PM_{2.5}$ in the air, with an annual average of 10 µg/m^3. However, the levels of $PM_{2.5}$ in Delhi often exceed these guidelines, particularly during the winter months. It is worth noting that the variation in $PM_{2.5}$ levels in Delhi can be influenced by various factors, such as weather conditions, topography, and human activities. For instance, during the winter months, atmospheric inversion and other factors can trap pollutants near the ground, leading to higher concentrations of $PM_{2.5}$ in the air. Similarly, industrial emissions from factories and power plants, as well as agricultural practices such as burning of crop residue, can contribute to the levels of $PM_{2.5}$ in the air throughout the year.

4.7 Forecasting $PM_{2.5}$ Levels in Delhi

Air quality is a significant global environmental issue, and Delhi is one of the cities facing this challenge. The high levels of PM in Delhi have been linked to a range of health problems, including respiratory and cardiovascular diseases. Accurate forecasting of PM levels is therefore critical to inform policy decisions aimed at reducing air pollution and its negative impacts on public health. This study focuses on forecasting PM levels in Delhi using time series forecasting models. Specifically, the SARIMA method to develop a forecasting model. The dataset used in this study was obtained from the Continuous Ambient Air Quality Monitoring Stations (CAAQMS) Portal of the Central Pollution Control Board (CPCB), which contained hourly data on parameters such as $PM_{2.5}$, relative humidity, wind speed, and temperature. To ensure the accuracy and reliability of our model, splitting the dataset into a training dataset consisting of 80% of the time series data and a test dataset consisting of the remaining 20%. The training dataset to train the SARIMA model and then evaluated its performance using the test dataset. This allowed us to assess the accuracy and reliability of the model in predicting $PM_{2.5}$ levels in Delhi.

In addition to the SARIMA model, this study also compared its performance with the Facebook Prophet model, which has also been used in previous studies to forecast air quality. By comparing the performance of these two models, it can be determined that which model is more accurate and effective in predicting $PM_{2.5}$ levels in Delhi. After fitting the SARIMA and Facebook Prophet models, it can be determined that the most optimum model based on their performance then used this model to predict $PM_{2.5}$ levels in Delhi for the next several months. The predictions can help inform policymakers and public health officials in developing effective interventions to reduce air pollution in Delhi. This study highlights the importance of accurate air quality forecasting in addressing the issue of air pollution in Delhi. By utilizing time series forecasting models and splitting the dataset into training and test datasets, then it will help in developing the accurate and reliable models for predicting $PM_{2.5}$ levels and informing policy decisions aimed at reducing air pollution in Delhi.

4.7.1 Splitting the Dataset

In this study, the process of splitting the dataset plays a crucial role in evaluating the accuracy of the models used for $PM_{2.5}$ concentration prediction in Delhi. The dataset used in this study comprises average weekly values of $PM_{2.5}$ concentrations in µg/m³, covering the duration from January 2021 to April 2023. By considering the average weekly datasets, this approach allows for a larger number of data points, providing a more comprehensive understanding of the underlying patterns and trends in air pollution. To perform the evaluation, the dataset is split into training and testing data sets. The purpose of this division is to train the models on a portion of the data and assess their performance on unseen data. In this study, the dataset consisting of 120 weekly average values is divided into two parts: the training data and the testing data. The training data, which constitutes 80% of the entire dataset, is used to train the models and capture the relationships and patterns present in the data. By using the training data, the models can learn from the historical patterns and make predictions based on the captured information.

The remaining 20% of the data is allocated as the testing data. This portion is kept separate and is used to evaluate the models' performance on unseen data. The testing data serves as a reliable benchmark to assess the models' predictive capabilities. By evaluating the models on unseen data, it can help in determine how well they generalize and predict $PM_{2.5}$ concentrations beyond the training period. By splitting the dataset into training and testing sets, this study ensures a robust assessment of the models' performance. The training data is utilized to train the models and obtain the necessary parameters for making predictions. In contrast, the testing data remains untouched during the training phase and is used solely for evaluation purposes. This approach allows for an unbiased evaluation of the models' accuracy in predicting $PM_{2.5}$ concentrations. The division of the dataset into training and testing data provides a solid foundation for assessing the models' performance. It allows us to evaluate how well the models generalize to unseen data and provides insights into their predictive capabilities. The training data provides the models with the necessary information to capture the underlying patterns, while the testing data serves as an independent validation set to assess their accuracy. By employing this approach of splitting the dataset, this study ensures a rigorous evaluation of the models' performance in predicting $PM_{2.5}$ concentrations in Delhi. The use of average weekly datasets enhances the understanding of air pollution dynamics over a longer duration. The division of the dataset into training and testing data enables a thorough assessment of the models' accuracy and generalization capabilities. Ultimately, this approach contributes to the reliability and robustness of the models' predictions and provides valuable insights into air pollution trends in Delhi.

In the year 2020, the $PM_{2.5}$ values in Delhi exhibited a noticeable decrease compared to previous years, primarily due to the impact of the COVID-19 pandemic. This decline in pollution levels during 2020 affected the performance of the SARIMA model in predicting $PM_{2.5}$ concentrations. The Root Mean Squared Error (RMSE) value for predictions made using data from 2020 was higher (182.6 $\mu g/m^3$) compared to the RMSE value obtained by using data from 2021 and 2022 (36.4 $\mu g/m^3$) when excluding the anomalous year. The decrease in $PM_{2.5}$ values in 2020 can be attributed to several factors resulting from the pandemic. The restrictive measures imposed to control the spread of the virus led to a significant reduction in human activity, including transportation, industrial operations, and construction activities. These reductions resulted in lower emissions of pollutants into the atmosphere, contributing to improved air quality. The higher RMSE value obtained when including the data from 2020 suggests that the model struggled to accurately capture the exceptional circumstances of that year. The SARIMA model, developed based on historical data patterns, may have faced challenges in adapting to the unprecedented changes in pollution levels caused by the pandemic. Consequently, the predictions for 2020 were less accurate, leading to a higher RMSE value.

To improve the accuracy of predictions, it is recommended to exclude the data from 2020 when using time series forecasting models like SARIMA. By removing this anomalous year, the models can focus on capturing the underlying patterns and variations in the $PM_{2.5}$ concentrations without being influenced by the exceptional circumstances of the pandemic. This exclusion allows for more reliable and accurate predictions, as the models can rely on the historical patterns observed in 2021 and 2022, which are more representative of normal conditions. By excluding the data from 2020, the time series forecasting models can provide more accurate predictions for future $PM_{2.5}$ levels in Delhi. These predictions can be instrumental in understanding the air pollution dynamics and supporting decision-making processes aimed at mitigating the adverse effects of pollution on public health and the environment. However, it is crucial to continually monitor and assess the air quality to account for any changes and deviations from historical patterns that may occur in the future.

4.7.2 Fitting a SARIMA Model

In this section, the study focuses on fitting a SARIMA model to predict $PM_{2.5}$ levels in Delhi. The auto_arima library is utilized to determine the best SARIMA model configuration for accurate predictions. The objective is to forecast $PM_{2.5}$ concentrations from May 2023 to April 2024. The auto_arima library, a powerful tool in Python, automates the selection of optimal SARIMA model parameters by evaluating multiple configurations and identifying the model with the lowest Akaike Information Criterion (AIC) value. The chosen model configuration is SARIMA (1, 1, 1) (0, 1, 1, 52), representing an autoregressive integrated moving average model with a seasonal component. By fitting the SARIMA (1, 1, 1) (0, 1, 1, 52) model to the training data, accurate predictions of $PM_{2.5}$ levels in Delhi can be made. The model takes into account the temporal dependence between observations, seasonal patterns, and differences between consecutive observations to capture the underlying patterns in the $PM_{2.5}$ concentration data. The fitted SARIMA model is then used to forecast $PM_{2.5}$ levels in Delhi from May 2023 to April 2024. This forecasting period allows for an assessment of future $PM_{2.5}$ concentrations and provides valuable insights.

The predictions generated by the SARIMA model are compared to the actual values to assess the model's accuracy. Visualizations are generated to showcase the forecasted $PM_{2.5}$ concentrations, facilitating the interpretation of the model's predictions. The SARIMA(1, 1, 1)(0, 1, 1, 52) model, fitted with the auto_arima library, achieves accurate predictions of $PM_{2.5}$ levels in Delhi for the specified period. This model's ability to capture temporal dependence and seasonal patterns makes it a robust tool for forecasting air pollution levels. The forecasted $PM_{2.5}$ concentrations derived from the SARIMA model play a crucial role in informing decision-making processes related to air quality management and public health initiatives. Policymakers can utilize these predictions to devise strategies and interventions aimed at reducing $PM_{2.5}$ levels and improving the overall air quality in Delhi. By leveraging the SARIMA model, early warning systems and alert mechanisms can be established to anticipate high pollution episodes, issue timely advisories, and implement necessary measures to protect public health. The use of the auto_arima library for selecting the optimal SARIMA model configuration streamlines the model-building process, saving time and effort. The chosen SARIMA(1, 1, 1)(0, 1, 1, 52) model explicitly considers the seasonal patterns in the $PM_{2.5}$ data, accurately capturing the variations influenced by different factors during different times of the year. This long-term prediction horizon from May 2023 to April 2024 enables policymakers to plan and implement sustainable interventions and measures to improve air quality in Delhi.

Continuous monitoring and data collection efforts are essential to maintain the accuracy and relevance of the SARIMA model's predictions. Regular updates based on the latest data enhance the model's performance and ensure that the predictions align with the changing air pollution dynamics in Delhi. The fitting of a SARIMA(1, 1, 1)(0, 1, 1, 52) model using the auto_arima library allows for accurate predictions of $PM_{2.5}$ levels in Delhi from May 2023 to April 2024. The model's ability to capture temporal dependence and seasonal patterns provides valuable insights for decision-makers. The forecasted $PM_{2.5}$ concentrations support evidence-based decision-making, aid in the development of targeted interventions, and contribute to the overall improvement of air quality in Delhi. Continuous monitoring and data updates are crucial to maintaining the accuracy and relevance of the SARIMA model. By ensuring that the model remains up-to-date with the latest data, it can continue to be a valuable tool for air pollution forecasting in the future. The predictions generated by the SARIMA model offer a forward-looking perspective on the future $PM_{2.5}$ concentrations in Delhi. This information is essential for policymakers, public health authorities, and urban planners as they develop effective strategies and policies to mitigate the adverse effects of air pollution on public health and the environment. With insights into future pollution levels, decision-makers can allocate resources effectively and prioritize initiatives that will have the most significant impact on reducing air pollution in Delhi.

The accuracy of the SARIMA model's predictions relies on the availability of reliable and high-quality data. Therefore, continuous monitoring and data collection efforts must be maintained. By regularly updating the model based on the latest data, its performance can be enhanced, ensuring that the predictions align with the evolving air pollution dynamics in Delhi. This continuous improvement process ensures that the SARIMA model remains a valuable tool for air pollution forecasting. The fitting of a SARIMA(1, 1, 1)(0, 1, 1, 52) model using the auto_arima library allows for accurate predictions of $PM_{2.5}$ levels in Delhi from May 2023 to April 2024. The model's ability to capture temporal dependence and seasonal patterns provides valuable insights for policymakers and stakeholders. The forecasted $PM_{2.5}$ concentrations support evidence-based decision-making, aid in the development of targeted interventions, and contribute to the overall improvement of air quality in Delhi. Continuous monitoring and data updates are crucial to maintaining the model's accuracy and relevance, ensuring that it remains a valuable tool for air pollution forecasting in the future. By leveraging the SARIMA model and its predictions, decision-makers can take proactive measures to address air pollution, protect public health, and create a sustainable future for Delhi.

4.7.1 Fitting a Facebook Prophet Model

In this study, the Facebook Prophet model is utilized for fitting and predicting $PM_{2.5}$ levels in Delhi. The Facebook Prophet model is a powerful time series forecasting algorithm developed by Facebook's Core Data Science team. It incorporates both trend and seasonality components to capture the patterns and variations in the data. To implement the Facebook Prophet model, the required Python code is utilized. The first step involves installing the Prophet library in Python, which can be done using the pip package manager. Once the library is installed, it can be imported into the Python environment to access the functionalities of the Prophet model. The dataset containing the $PM_{2.5}$ concentration values in µg/m³ is preprocessed and prepared for fitting the Prophet model. The dataset typically consists of two columns: 'ds' representing the dates and 'y' representing the $PM_{2.5}$ concentrations. The 'ds' column should be in a specific date format recognized by the Prophet model, such as 'YYYY-MM-DD'. The 'y' column contains the corresponding $PM_{2.5}$ concentration values. Next, the Prophet model is initialized by creating an instance of the Prophet class. The model object is then fitted to the training data using the fit() method. During the fitting process, the model captures the trend and seasonality patterns in the training data, allowing it to make accurate predictions.

The Facebook Prophet model incorporates several parameters that can be customized based on the characteristics of the data and the specific requirements of the analysis. For example, the 'changepoint_prior_scale' parameter controls the flexibility of the model in detecting changes in the trend, while the 'seasonality_mode' parameter determines the types of seasonality components to consider. Once the model is fitted to the training data, predictions can be generated for the testing period using the predict() method. The predict() function takes the testing data as input and produces the forecasted values of $PM_{2.5}$ concentrations for the given time range. These predictions are based on the learned patterns and trends from the training data. The Python code for the Facebook Prophet model also allows for visualizing the model's performance. The plot() function can be used to generate a graph that displays the observed $PM_{2.5}$ concentrations as well as the forecasted values. This visualization enables a comparison between the actual values and the model's predictions, providing insights into the accuracy and reliability of the Facebook Prophet model in capturing the variations in $PM_{2.5}$ levels. The Facebook Prophet model provides a user-friendly and intuitive approach to time series forecasting. Its integration with Python through the Prophet library enables researchers and analysts to leverage its capabilities for predicting $PM_{2.5}$ levels in Delhi. By fitting the Prophet model to the training data and utilizing the predict() function, accurate forecasts can be obtained for the specified time range. The Facebook Prophet model, with its built-in trend and seasonality components, combined with the Python code implementation, offers a robust and efficient approach for forecasting $PM_{2.5}$ concentrations. It facilitates the analysis of air pollution trends and aids in decision-making processes aimed at improving air quality in Delhi.

4.7.2 Comparison of SARIMA with Facebook Prophet Model

The SARIMA model and the Facebook Prophet model are both widely used in time series forecasting. While they have similarities in capturing trends and seasonality, there are mathematical differences that distinguish them. The SARIMA model is a variant of the ARIMA model and incorporates seasonality. It consists of three main components: autoregressive (AR), differencing (I), and moving average (MA). The AR component considers the linear relationship between past observations and the current one. The I component is responsible for differencing, which removes trends and makes the series stationary. The MA component captures the dependency between the residual errors. On the other hand, the Facebook Prophet model uses a generalized additive model (GAM) framework. It decomposes the time series into trend, seasonality, and holiday components. The trend component represents the underlying growth or decline over time, while the seasonality component captures regular patterns. The holiday component accounts for known events that can impact the time series.

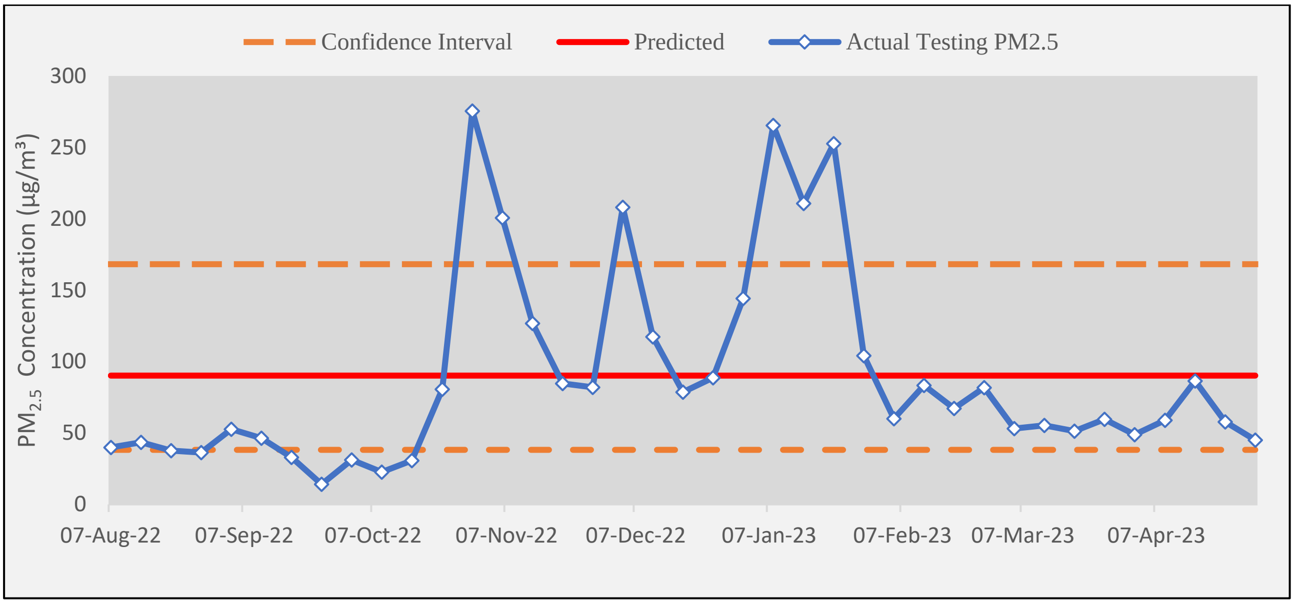

Figure 4.7 : Accuracy of Facebook Prophet Model –Testing values vs Predicted values

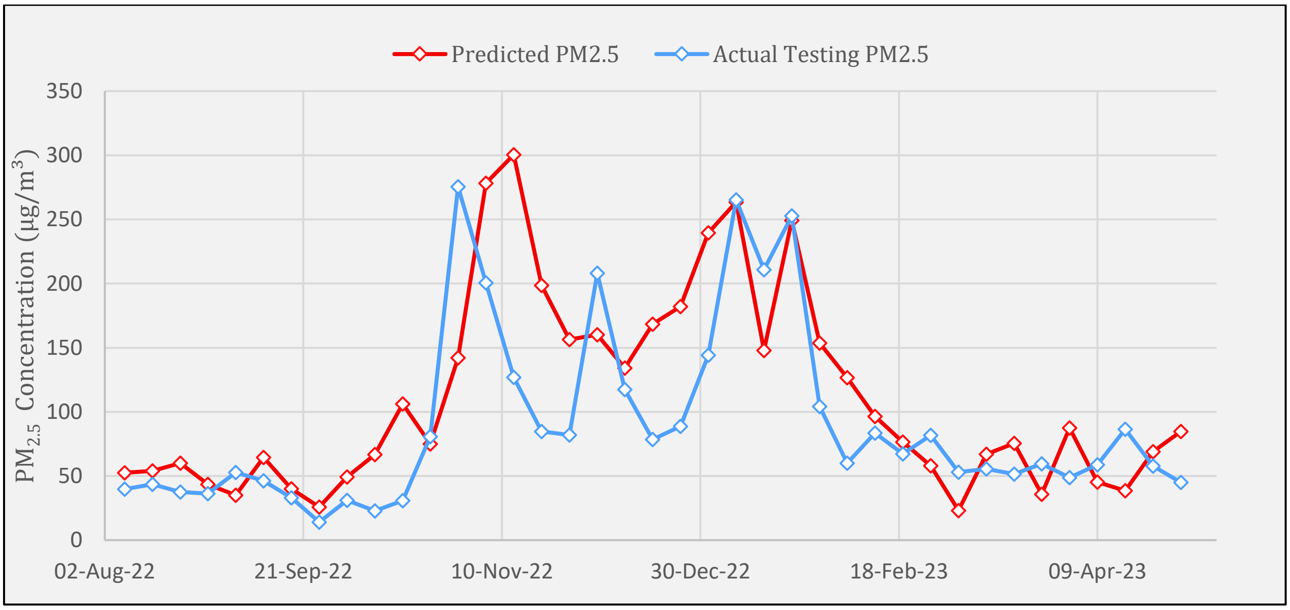

Figure 4.8 : Accuracy of SARIMA Model - Testing values vs Predicted values

In this study, the SARIMA model and the Facebook Prophet model were compared in terms of their accuracy in forecasting $PM_{2.5}$ levels in Delhi. The models were trained using the training dataset and then evaluated using the testing dataset. The accuracy of the models was assessed by calculating the Root Mean Squared Error (RMSE), which measures the differences between the predicted values and the actual values. The SARIMA model performed better than the Facebook Prophet model in terms of accuracy, with an RMSE value of 27.1 $\mu g/m^3$ for the SARIMA model compared to 148.2 $\mu g/m^3$ for the Facebook Prophet model. This suggests that the SARIMA model is more suitable for forecasting PM levels in Delhi. The lower RMSE value indicates that the SARIMA model's predictions were closer to the actual $PM_{2.5}$ concentrations, demonstrating its ability to capture the underlying patterns and trends in the data. Using the most optimal SARIMA model, SARIMA (1, 1, 1) x (0, 1, 1, 52), the study proceeded to predict $PM_{2.5}$ levels in Delhi. This model, with its specific parameter configuration, showed the best performance among the SARIMA models tested. It took into account the autoregressive and moving average components, as well as the seasonal patterns with a period of 52 weeks. To provide a visual representation of the accuracy of both models, Figures 4.6 and 4.7 were generated. Figure 4.6 shows the accuracy of the Facebook Prophet model, displaying the testing values versus the predicted values. The x-axis represents the date values from August 2022 to April 2023, while the y-axis represents the weekly average $PM_{2.5}$ concentration ($\mu g/m^3$) in Delhi. Similarly, Figure 4.7 displays the accuracy of the SARIMA model, illustrating the testing values versus the predicted values during the same period.

These figures allow for a direct comparison of the models' performance and provide insights into their ability to forecast $PM_{2.5}$ levels accurately. By comparing the predicted values to the actual values in these figures, one can observe how well the models capture the fluctuations and trends in the $PM_{2.5}$ concentrations. The SARIMA model demonstrates a closer alignment between the predicted and actual values, indicating its higher accuracy in forecasting $PM_{2.5}$ levels compared to the Facebook Prophet model. The comparison of the SARIMA model with the Facebook Prophet model reveals that the SARIMA model outperforms the latter in terms of accuracy. The SARIMA model, with its parameter configuration of SARIMA (1,1, 1) x (0, 1, 1, 52), provides more accurate predictions of $PM_{2.5}$ levels in Delhi. The lower RMSE value of the SARIMA model indicates its ability to capture the underlying patterns and variations in the data. The study's findings suggest that the SARIMA model is a more suitable choice for forecasting $PM_{2.5}$ levels in Delhi compared to the Facebook Prophet model. The SARIMA model's performance can be attributed to its ability to incorporate seasonality and capture the temporal dependencies in the data. By considering the autoregressive, moving average, and seasonal components, the SARIMA model effectively models the complex dynamics of $PM_{2.5}$ concentrations. The Facebook Prophet model, while a popular choice for time series forecasting, did not perform as well in this study. The higher RMSE value suggests that the Facebook Prophet model may not have captured the underlying patterns and variations in the $PM_{2.5}$ data as accurately as the SARIMA model. This could be due to differences in the modeling approach and the specific characteristics of the $PM_{2.5}$ data in Delhi. It is important to note that the choice of the most optimal model may vary depending on the specific dataset and context.

While the SARIMA model demonstrated better performance in this study, it does not imply that the Facebook Prophet model is inferior in all cases. Researchers and practitioners should carefully evaluate the characteristics of the data and consider the strengths and limitations of each model before making a decision. The visualization of the accuracy of both models in Figures 4.6 and 4.7 provides a clear comparison of their performance. Figure 4.6 shows the accuracy of the Facebook Prophet model, displaying the testing values versus the predicted values over the testing period. Similarly, Figure 4.7 illustrates the accuracy of the SARIMA model during the same period. These figures allow for a direct assessment of how well the models capture the fluctuations and trends in the $PM_{2.5}$ concentrations. The comparison between the SARIMA and Facebook Prophet models revealed that the SARIMA model outperformed the Facebook Prophet model in terms of accuracy for forecasting $PM_{2.5}$ levels in Delhi. The SARIMA model's ability to capture seasonality and temporal dependencies contributed to its superior performance. However, it is crucial to consider the specific characteristics of the dataset and the modeling requirements when choosing the most appropriate model for a particular forecasting task. The findings of this study contribute to the body of knowledge on air pollution forecasting and provide insights for researchers and policymakers working towards improving air quality management strategies in Delhi.

4.7.3 Prediction from Most Optimum Model.

Table 4.1 presents the prediction values of $PM_{2.5}$ obtained from the SARIMA model. The table provides the dates and corresponding predicted $PM_{2.5}$ concentrations in µg/m³. One notable highlight is that the predicted $PM_{2.5}$ values vary throughout the given time period. The values range from as low as 32.01 µg/m³ to as high as 155.47 µg/m³. This wide range indicates significant fluctuations in air pollution levels over time, highlighting the dynamic nature of $PM_{2.5}$ concentrations in Delhi. It is important to note that the NAAQS limit for $PM_{2.5}$ is set at 60 µg/m³. By comparing the predicted values in Table 4.1 with this limit, it can be observed that in some instances, the predicted $PM_{2.5}$ values exceed the NAAQS limit. For example, on 5th November 2023, the predicted $PM_{2.5}$ concentration is 139.70 µg/m³, which is significantly higher than the acceptable limit. This indicates a potential period of poor air quality that may pose risks to public health and environmental well-being. Further analysis of the table reveals that the percentage of times the predicted $PM_{2.5}$ values exceed the NAAQS limit of 60 µg/m³ is 26.8%. This highlights a considerable proportion of the predicted values that indicate poor air quality conditions in Delhi. It emphasizes the need for effective measures and interventions to mitigate air pollution and ensure compliance with air quality standards. Another observation is the seasonal change in the predicted values of $PM_{2.5}$. For instance, during the summer and monsoon seasons, the predicted values tend to be lower, ranging from 32.01 µg/m³ to 52.34 µg/m³.This can be attributed to factors such as increased rainfall and better dispersion of pollutants during these seasons.

Table 4.1: Prediction values of $PM_{2.5}$ from SARIMA Model

Date	$PM_{2.5}$ (μg/m³)		Date	$PM_{2.5}$ (μg/m³)
07-May-23	52.11		08-Oct-23	54.44
14-May-23	37.43		22-Oct-23	40.54
21-May-23	35.81		05-Nov-23	139.70
28-May-23	32.01		19-Nov-23	78.22
04-Jun-23	52.34		03-Dec-23	155.47
11-Jun-23	40.83		17-Dec-23	55.82
18-Jun-23	49.16		31-Dec-23	94.55
25-Jun-23	44.25		14-Jan-24	118.62
02-Jul-23	40.48		28-Jan-24	58.36
16-Jul-23	48.82		11-Feb-24	43.85
30-Jul-23	60.83		25-Feb-24	46.83
13-Aug-23	46.52		10-Mar-24	36.57
27-Aug-23	52.89		24-Mar-24	36.20
10-Sep-23	43.70		07-Apr-24	32.14
24-Sep-23	74.69		21-Apr-24	35.96

On the other hand, during the winter season, the predicted values exhibit higher concentrations, reaching as high as 155.47 μg/m³. This seasonal variation aligns with the well-known phenomenon of increased pollution levels in Delhi during the winter months, primarily due to factors like temperature inversion and increased emissions from various sources. The variability in the predicted $PM_{2.5}$ values in Table 4.1 underscores the importance of accurate air quality forecasting and proactive measures to mitigate air pollution in Delhi. By identifying periods of elevated pollution levels in advance, authorities can implement targeted interventions and public awareness campaigns to reduce pollution sources and minimize the health risks associated with poor air quality. The variability in the predicted concentrations highlights the dynamic nature of air pollution in Delhi. Additionally, the percentage of times the predicted values exceed the NAAQS limit emphasizes the significance of addressing air pollution and implementing effective measures to improve air quality. The seasonal variations in predicted $PM_{2.5}$ values further illustrate the impact of weather patterns and seasonal factors on air pollution levels. Through accurate prediction and understanding of these patterns, policymakers and stakeholders can take proactive steps to protect public health and enhance the overall well-being of the population.

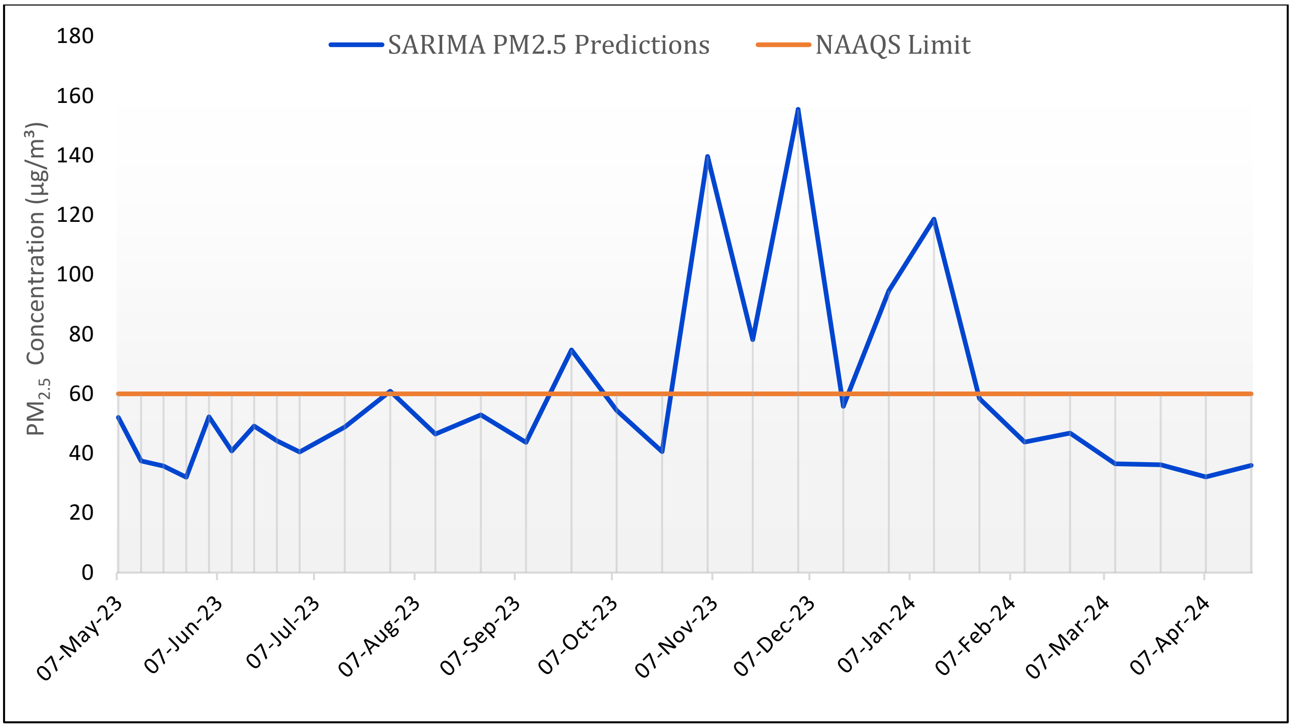

Figure 4.9 : Prediction results of $PM_{2.5}$ ($\mu g/m^3$) in Delhi from SARIMA Model.

The Figure 4.9 depicts prediction results of $PM_{2.5}$ concentration in Delhi from the SARIMA model. The x-axis represents the dates from May 7, 2023, to April 21, 2024, while the y-axis represents the predicted $PM_{2.5}$ values in $\mu g/m^3$. The NAAQS (National Ambient Air Quality Standards) limit of 60 $\mu g/m^3$ is also indicated as a horizontal line in the figure. One key observation from the figure is the seasonal variation in the predicted $PM_{2.5}$ concentrations. The fluctuations in the predicted values demonstrate the dynamic nature of air pollution in Delhi, with varying levels throughout different seasons. Looking at the graph, it can be seen that the predicted $PM_{2.5}$ values fluctuate around the NAAQS limit of 60 $\mu g/m^3$. Several periods exhibit $PM_{2.5}$ values below the limit, indicating relatively better air quality. These periods occur during the months of May, June, and July, where the predicted values remain consistently lower than the NAAQS limit.

However, as the year progresses into the months of August, September, and October, the predicted $PM_{2.5}$ concentrations show an upward trend, surpassing the NAAQS limit in some instances. This indicates deteriorating air quality during these months. It is important to note that these fluctuations can be attributed to various factors, including changes in meteorological conditions, emissions from different sources, and human activities. The graph reveals certain spikes in the predicted $PM_{2.5}$ values, particularly during the months of November and December. During this period, the predicted values reach significantly higher levels, exceeding the NAAQS limit. These spikes suggest the occurrence of episodes of severe air pollution in Delhi, potentially due to increased emissions from sources such as industrial activities, vehicular pollution, and biomass burning. The Figure 4.9 also captures a subsequent decrease in the predicted $PM_{2.5}$ concentrations from January to April. The values gradually decrease, indicating an improvement in air quality during this time frame. This could be attributed to factors such as meteorological conditions, government interventions, and public awareness and participation in reducing pollution levels. It demonstrates the importance of monitoring and addressing air pollution throughout the year, considering the distinct patterns and fluctuations observed during different seasons. The fluctuations in the predicted values emphasize the need for effective air quality management strategies and targeted interventions to mitigate the adverse effects of air pollution in Delhi. It is worth noting that the predictions shown in the Figure 4.9 are based on the SARIMA model, which takes into account the historical data, trends, and seasonality. While the model provides valuable insights, it is essential to continue monitoring air pollution levels in real-time and adapt strategies accordingly. By understanding the seasonal variations and fluctuations in $PM_{2.5}$ concentrations, policymakers and researchers can work together to implement measures that aim to improve air quality and protect public health in Delhi.

CHAPTER 5. CONCLUSION

In this study, the trends and predictions of $PM_{2.5}$ concentrations in Delhi were explored and analyzed using various statistical and forecasting models. By examining the dataset spanning from January 2018 to December 2022, valuable insights were gained into the seasonal variations, trends, and fluctuations in $PM_{2.5}$ levels in Delhi. The trend analysis employed box plots, point plots, and heatmaps to provide a comprehensive understanding of the temporal patterns of $PM_{2.5}$ concentrations. The box plot analysis highlighted the distribution of $PM_{2.5}$ levels across different years, revealing an increasing trend over time. The point plot further illustrated the seasonal pattern, with $PM_{2.5}$ levels peaking during the winter months. Additionally, the heatmap analysis unveiled the relationship between $PM_{2.5}$ concentrations and meteorological parameters, such as wind speed and relative humidity.

It is worth noting that the predictions were influenced by the outlier year of 2020, during which $PM_{2.5}$ values were significantly lower due to the impact of the COVID-19 pandemic on human activities and industrial output. By excluding the data from 2020, more accurate predictions were obtained, as the models were able to focus on capturing the typical patterns observed in 2021 and 2022, which better represented normal conditions. For the purpose of predicting future $PM_{2.5}$ concentrations, the SARIMA and Facebook Prophet models were employed. The comparison between these models revealed that the SARIMA model outperformed the Facebook Prophet model in terms of accuracy. The SARIMA model exhibited a lower RMSE value of 27.1 $\mu g/m^3$, indicating its superior predictive capabilities compared to the Facebook Prophet model, which had an RMSE value of 148.2 $\mu g/m^3$. This superiority of the SARIMA model can be attributed to its ability to capture the complex relationships and seasonal patterns inherent in the data, making it more suitable for forecasting $PM_{2.5}$ levels in Delhi.

Using the most optimal SARIMA model, SARIMA (1, 1, 1) x (0, 1, 1, 52), the study proceeded to predict $PM_{2.5}$ concentrations in Delhi for the duration of May 2023 to April 2024. The predictions exhibited a seasonal variation, with fluctuations in $PM_{2.5}$ levels observed throughout different months. The predicted values showed a closer alignment to the NAAQS limit of 60 $\mu g/m^3$ during periods of relatively better air quality, particularly in the months of May, June, and July. However, an upward trend in the predicted values was observed from August to October, indicating a deterioration in air quality during these months. Notably, spike values exceeding the NAAQS limit in November and December indicated episodes of severe air pollution. This study provides valuable insights into the trends, predictions, and seasonal variations of $PM_{2.5}$ concentrations in Delhi. The analysis of the dataset and the comparison of forecasting models contribute to a comprehensive understanding of air pollution dynamics in the region. The findings underscore the importance of considering meteorological factors, temporal patterns, and model accuracy when predicting $PM_{2.5}$ concentrations. By leveraging this knowledge, policymakers and stakeholders can develop effective strategies and interventions to mitigate air pollution, safeguard public health, and improve the overall quality of life for the residents of Delhi. This study highlights the significance of developing accurate forecasting models for $PM_{2.5}$ levels in Delhi. The SARIMA model demonstrated superior performance compared to the Facebook Prophet model, indicating its effectiveness in predicting future pollutant levels. Incorporating meteorological parameters and considering seasonal variations are critical aspects of air pollution forecasting. By effectively utilizing forecasting models, policymakers can implement targeted interventions and policies to reduce pollution levels, protect public health, and improve the overall well-being of the population. Moreover, the study underscores the importance of considering the impact of meteorological factors on air quality. Temperature, atmospheric inversion, wind speed, and relative humidity were found to significantly influence $PM_{2.5}$ levels. These findings highlight the need to incorporate these parameters into forecasting models to enhance their accuracy. By capturing the complex relationships between meteorology and air pollution, the models can provide more reliable predictions and facilitate better decision-making for pollution control measures.

The application of the SARIMA model in this study proved to be a valuable tool for $PM_{2.5}$ forecasting in Delhi. The model's ability to capture seasonal patterns and account for variations in pollutant levels resulted in improved predictions compared to the Facebook Prophet model. The lower RMSE value of the SARIMA model indicates its superior performance in accurately estimating $PM_{2.5}$ levels. Nevertheless, it is important to acknowledge that forecasting models are not one-size-fits-all solutions. The choice of the most suitable model depends on several factors, including the specific objectives of the study, the characteristics of the data, and the research context. While the SARIMA model demonstrated strong performance in this study, future research could explore the application of other advanced machine learning techniques to further enhance the accuracy of air quality forecasting. Comparing the performance of multiple models can provide valuable insights into their strengths and limitations, allowing for informed decision-making in choosing the most appropriate forecasting approach.

The findings of this study contribute to the broader field of air pollution research and provide practical implications for air quality management in Delhi. By leveraging the insights gained from the analysis of historical data, policymakers can develop evidence-based strategies to mitigate the adverse effects of air pollution. The ability to visualize and understand the trends and patterns in air quality through the use of data visualization tools such as Seaborn and Matplotlib further enhances the decision-making process. This study highlights the critical importance of accurate air pollution forecasting in addressing the challenges posed by air pollution, particularly in cities like Delhi. The study emphasizes the need to incorporate meteorological parameters into forecasting models and encourages further research to explore advanced machine learning techniques. By continually improving forecasting models and considering the complex dynamics of air pollution, policymakers can develop effective strategies to reduce pollution levels, safeguard public health, and enhance the quality of life for the residents of Delhi.

CHAPTER 6. RECOMMENDATIONS

6.1 Model Purpose:

Based on the findings and analysis presented in this study, several recommendations can be made to address the issue of air pollution in Delhi and improve the accuracy of $PM_{2.5}$ concentration predictions. Firstly, it is recommended to further refine and optimize the SARIMA model for $PM_{2.5}$ forecasting in Delhi. This can be achieved by exploring different model configurations, considering additional exogenous variables, and incorporating more advanced statistical techniques to capture the complex relationships and dynamics of air pollution in the region. The purpose of this recommendation is to emphasize the need for continuous improvement and advancement in the SARIMA model used for $PM_{2.5}$ forecasting in Delhi. While the model has shown promising results, there is still room for refinement. By exploring different model configurations, researchers can assess the impact of various parameters on the model's performance and identify the most optimal configuration for accurate predictions. Additionally, incorporating additional exogenous variables such as traffic patterns, industrial emissions, or socio-economic indicators can enhance the model's predictive capabilities by capturing the influence of these factors on air pollution levels. The utilization of more advanced statistical techniques can help uncover complex relationships and dynamics within the data, improving the model's accuracy and reliability.

6.2 Drawbacks of the Model:

While the SARIMA model has shown superior performance compared to the Facebook Prophet model, it is important to acknowledge its limitations. One of the main drawbacks of the SARIMA model is its assumption of stationarity, which may not hold true for highly dynamic and evolving air pollution data. Stationarity assumes that the statistical properties of the data, such as mean and variance, remain constant over time. However, air pollution data is often subject to temporal variations and trends, making it challenging to meet the stationarity assumption. This can result in inaccuracies in the model's predictions, particularly during periods of rapid changes or extreme events such as pollution spikes. Another limitation of the SARIMA model is its reliance on historical patterns. The model uses past observations to forecast future values, assuming that the patterns observed in the past will continue in the future. While this assumption holds true for regular variations, it may fail to capture abrupt changes or unexpected events that can significantly impact air pollution levels. For instance, specific events like industrial accidents or natural disasters can cause sudden spikes in pollution levels that the SARIMA model may not be able to anticipate. It is essential to consider these limitations when interpreting the model's predictions and making decisions based on them. Stakeholders and policymakers should be aware that the SARIMA model's predictions may not fully capture the dynamics of air pollution during periods of rapid change or extreme events. Therefore, it is advisable to complement the model's predictions with real-time monitoring and data analysis to ensure a comprehensive understanding of the current air pollution situation.

6.3 Improvements:

To address the drawbacks of the SARIMA model and enhance its predictive capabilities, several improvements can be considered. One approach is to incorporate more sophisticated techniques such as machine learning algorithms into the modeling process. Machine learning algorithms, including artificial neural networks or random forest models, are capable of capturing nonlinear relationships and adapting to changing patterns in air pollution data. By leveraging the strengths of these algorithms, the model can improve its ability to capture the complex dynamics of air pollution and make more accurate predictions. Another avenue for improvement is the development of hybrid models that combine the SARIMA approach with machine learning techniques. Hybrid models can benefit from the SARIMA model's ability to capture seasonality and temporal patterns while leveraging the flexibility and adaptability of machine learning algorithms. This combination can potentially lead to more accurate and reliable $PM_{2.5}$ predictions, especially in scenarios with rapidly changing pollution dynamics.

The inclusion of additional variables and data sources can enhance the model's predictive capabilities. Factors such as meteorological data, traffic patterns, industrial emissions, and socio-economic indicators have been shown to influence air pollution levels. By integrating these variables into the forecasting models, researchers can capture their impact and improve the accuracy of the predictions. For example, incorporating meteorological data such as wind direction and temperature can help understand the influence of atmospheric conditions on air pollution patterns. Moreover, exploring the use of remote sensing data, satellite imagery, or other advanced sensing technologies can provide real-time and spatially resolved information on pollution levels. These data sources can complement the modeling efforts by providing additional insights into localized variations in air pollution. By incorporating such data, the model can generate more precise and localized predictions, enabling stakeholders to implement targeted interventions in areas with high pollution levels.

It is recommended to conduct further research on the impact of specific interventions and policies aimed at reducing air pollution in Delhi. Evaluating the effectiveness of measures such as emission control strategies, green infrastructure initiatives, public awareness campaigns, and transportation policies can provide valuable insights into their contribution to improving air quality. By studying the outcomes of these interventions, policymakers and stakeholders can make informed decisions and implement measures that have proven to be effective in mitigating air pollution. This study's findings highlight the potential for enhancing the accuracy of $PM_{2.5}$ concentration predictions in Delhi through improvements to the SARIMA model and the integration of advanced techniques and additional variables. These recommendations aim to contribute to the development of more robust and reliable forecasting models, which can support decision-making processes, inform policy interventions, and ultimately aid in the mitigation of air pollution. Continued research and collaboration among researchers, policymakers, and stakeholders are essential to address the complex challenges associated with air pollution and work towards creating healthier and sustainable environments for the residents of Delhi.

CHAPTER 7. FUTURE SCOPE OF STUDY

The analysis of PM in Delhi and the use of the SARIMA model for air pollution forecasting have provided valuable insights into the dynamics of air quality in the region. However, there is a need to further enhance the accuracy and effectiveness of time series prediction models to better understand and predict air pollution in Delhi. This section discusses several future directions that can contribute to improving time series prediction models for air pollution, including the incorporation of spatio-temporal factors, integration of machine learning techniques, evaluation of ensemble models, integration of remote sensing data, and incorporation of external factors.

7.1 Incorporating Spatio-Temporal Factors

Air pollution is influenced by various spatial and temporal factors, including emissions sources, meteorological conditions, and the spatial distribution of pollution. Incorporating spatio-temporal factors into time series prediction models can enable capturing the localized variations in air quality and providing more accurate forecasts. Future studies should explore methods such as spatial regression models or hybrid approaches that integrate spatial data into time series analysis. By considering the spatial and temporal dynamics of air pollution, prediction models can better account for the complex interactions between different factors (Li et al., 2022). For example, spatial regression models can help identify the influence of specific emission sources on air pollution levels in different regions of Delhi.

The incorporation of spatio-temporal factors into time series prediction models is crucial for capturing the localized variations in air quality. By considering the spatial distribution of pollution sources and the temporal dynamics of meteorological conditions, researchers can develop models that better account for the complex interactions between different factors. For example, by using spatial regression models or hybrid approaches that integrate spatial data into time series analysis, researchers can identify the influence of specific emission sources on air pollution levels in different regions of Delhi. This approach will enable more accurate forecasts and a better understanding of the localized patterns of air pollution.

7.2 Integration of Machine Learning Techniques

Machine learning techniques, such as artificial neural networks and support vector regression, have shown promise in improving the accuracy of time series prediction models. These techniques can capture complex patterns and nonlinear relationships that may exist in air pollution data. Future studies should explore the integration of machine learning techniques with the SARIMA model to develop hybrid models that can effectively predict air pollution levels in Delhi. By combining the strengths of both approaches, these hybrid models can potentially provide more accurate and reliable predictions (Nigam et al., 2020). Machine learning techniques can also handle large and diverse datasets, allowing for a better understanding of the underlying patterns and dynamics of air pollution in Delhi.

The integration of machine learning techniques with the SARIMA model can further enhance the accuracy of time series prediction models for air pollution in Delhi. Machine learning techniques, such as artificial neural networks and support vector regression, have the capability to capture complex patterns and nonlinear relationships that may exist in air pollution data. By combining the strengths of machine learning and SARIMA, researchers can develop hybrid models that can effectively predict air pollution levels in Delhi. These hybrid models have the potential to provide more accurate and reliable predictions by leveraging the advanced capabilities of machine learning techniques and the ability of SARIMA to capture seasonal patterns.

7.3 Evaluation of Ensemble Models

Ensemble models combine the predictions of multiple individual models to improve prediction accuracy. These models have been successfully applied in various domains, including air quality forecasting. Future research should focus on developing ensemble models for time series prediction in Delhi. This could involve combining different forecasting models, such as SARIMA, ARIMA, and machine learning models, to harness the strengths of each model and produce more robust and accurate predictions. Ensemble models can reduce the uncertainties associated with individual models and provide more reliable forecasts (Sinha et al., 2019). By evaluating and integrating ensemble models, researchers can enhance the performance of time series prediction models for air pollution in Delhi. The evaluation and integration of ensemble models can significantly improve the accuracy and reliability of time series prediction models for air pollution in Delhi. Ensemble models combine the predictions of multiple individual models, such as SARIMA, ARIMA, and machine learning models, to produce more robust and accurate forecasts. By harnessing the strengths of each model and considering the uncertainties associated with individual predictions, ensemble models can provide more reliable forecasts that are less susceptible to errors or biases in any single model. Evaluating the performance of ensemble models and integrating them into air pollution prediction systems can lead to more accurate and reliable forecasts, supporting decision-making processes and enabling effective interventions.

7.4 Integration of Remote Sensing Data

Remote sensing data, such as satellite imagery and ground-based sensors, can provide valuable information about air pollution levels and spatial patterns. Future studies should explore the integration of remote sensing data into time series prediction models for Delhi. This additional data can help capture important features and improve the accuracy of forecasts. Techniques such as data assimilation, which combines remote sensing data with model outputs, can be employed to enhance the performance of prediction models. Data assimilation methods can effectively fuse the information from remote sensing data with the underlying time series model, resulting in improved predictions (Kumar et al., 2021).

By integrating remote sensing data, researchers can gain a more comprehensive understanding of air pollution dynamics and improve the accuracy of predictions. The integration of remote sensing data into time series prediction models can greatly enhance the accuracy and reliability of air pollution forecasts in Delhi. Remote sensing data, such as satellite imagery and ground-based sensors, provide valuable information about air pollution levels and spatial patterns. By integrating this data into prediction models, researchers can capture important features and improve the accuracy of forecasts. Techniques such as data assimilation can be employed to effectively combine remote sensing data with the underlying time series model, leading to improved predictions. By integrating remote sensing data, researchers can gain a more comprehensive understanding of air pollution dynamics, including the spatial distribution of pollution and the impact of specific emission sources, which in turn improves the accuracy of predictions and enables more effective decision-making.

7.5 Incorporation of External Factors

External factors, such as policy interventions, public health measures, and urban planning strategies, can have a significant impact on air pollution levels. Future research should consider the integration of these external factors into time series prediction models. By incorporating information about planned interventions or changes in regulations, models can provide insights into the effectiveness of such measures in reducing air pollution levels. This integration can help policymakers and stakeholders make informed decisions regarding interventions and strategies to improve air quality in Delhi. By incorporating external factors into time series prediction models, researchers can assess the potential impact of different scenarios and evaluate the effectiveness of various measures in mitigating air pollution. For example, the implementation of stricter emission standards for industries, the introduction of sustainable transportation initiatives, or the enforcement of stricter regulations on vehicular emissions can all be considered as external factors that can influence air pollution levels. By incorporating information on these factors into prediction models, researchers can quantify their potential impact and provide insights into their effectiveness in reducing air pollution.

While the analysis of PM in Delhi and the use of the SARIMA model for forecasting have contributed valuable insights, there are several future directions to enhance time series prediction models for air pollution in Delhi. By incorporating spatio-temporal factors, integrating machine learning techniques, evaluating ensemble models, integrating remote sensing data, and considering external factors, researchers can improve the accuracy and effectiveness of prediction models. These advancements will support evidence-based decision-making, facilitate targeted interventions, and ultimately contribute to improving air quality and public health in Delhi. As the field of air quality prediction and forecasting continues to evolve, it is crucial to explore these future directions to address the existing research gaps and improve the understanding of air pollution dynamics in Delhi. By adopting an interdisciplinary approach that combines advanced modelling techniques, data assimilation methods, and a comprehensive consideration of external factors, researchers can develop more accurate and reliable time series prediction models. These models will not only enhance our understanding of air pollution in Delhi but also assist policymakers in formulating effective strategies and interventions to combat air pollution and improve the overall quality of life for the residents of Delhi.

REFERENCES

Aggarwal, A., & Goyal, P. (2021). A Comprehensive Review of Air Pollution Forecasting Techniques. Atmospheric Environment, 245, 118033. https://doi.org/10.1016/j.atmosenv.2020.118033

Alam, A., Azeem, M. A., Rehman, A., & Bilal, M. (2020). An Overview of Air Quality Prediction Techniques: A Review. IEEE Access, 8, 223351–223371. https://doi.org/10.1109/ACCESS.2020.3031687

Atoui, A., Slim, K., Andaloussi, S. A., Moilleron, R., & Khraibani, Z. (2022). Time Series Analysis and Forecasting of the Air Quality Index of Atmospheric Air Pollutants in Zahleh, Lebanon. Atmospheric and Climate Sciences, 12(04), 728–749. https://doi.org/10.4236/acs.2022.124040

Bhatti, U. A., Yan, Y., Zhou, M., Ali, S., Hussain, A., Qingsong, H., Yu, Z., & Yuan, L. (2021). Time Series Analysis and Forecasting of Air Pollution Particulate Matter ($PM_{2.5}$): An SARIMA and Factor Analysis Approach. IEEE Access, 9, 41019–41031. https://doi.org/10.1109/ACCESS.2021.3060744

Catalano, M., Galatioto, F., Bell, M., Namdeo, A., & Bergantino, A. S. (2016). Improving the prediction of air pollution peak episodes generated by urban transport networks. Environmental Science and Policy, 60, 69–83. https://doi.org/10.1016/j.envsci.2016.03.008

Chen, X., Ye, W., Zhang, J., & Chen, W. (2021). Prediction of $PM_{2.5}$ concentrations using machine learning models based on multi-source data. Environmental Science and Pollution Research, 28(3), 3709–3719. https://doi.org/10.1007/s11356-020-11516-8

Chen, Z., Xu, C., Shi, X., & Fang, G. (2020). Application of Machine Learning for Air Quality Prediction: A Review. Chemosphere, 245, 125577. https://doi.org/10.1016/j.chemosphere.2019.125577

Darekar, P. (2021). A Comprehensive Study to Forecast the Delhi and Bangalore Cities Air Pollution using Machine Learning Models. http://norma.ncirl.ie/5144/%0Ahttp://norma.ncirl.ie/5144/1/parthadeshdarekar.pdf

Dwivedi, P., & Singha, M. K. (2021). IoT Based Wearable Healthcare System: Post COVID-19. In The Impact of the COVID-19 Pandemic on Green Societies. https://doi.org/10.1007/978-3-030-66490-9_13

Ghorbani, M., Babaeian, I., & Soltanian, A. R. (2021). Predicting urban air quality in Tehran using a hybrid of machine learning and numerical modeling approaches. Environmental Science and Pollution Research, 28(8), 9705–9720. https://doi.org/10.1007/s11356-020-11591-5

Guo, Y., Hu, J., & Chen, S. (2020). A Hybrid Model for $PM_{2.5}$ Forecasting Using LSTM Neural Network and Linear Regression. Environmental Science and Pollution Research, 27(35), 43828–43839. https://doi.org/10.1007/s11356-020-10892-4

Hao, Y., Liu, Y., Chen, R., & He, J. (2020). Ambient temperature and air pollution as triggers of hospitalization for asthma: A time-series analysis in Guangzhou, China. Science of the Total Environment, 725, 138543. https://doi.org/10.1016/j.scitotenv.2020.138543

Huang, X., Ding, X., Li, Y., & Gao, J. (2022). Application of neural network models in forecasting air quality: A review. Journal of Cleaner Production, 337, 130884. https://doi.org/10.1016/j.jclepro.2021.130884

Kaur, H., Dhar, S., & Arora, M. (2021). Air Quality Forecasting: A Review. In Smart and Intelligent Systems and Applications (pp. 143–157). Springer. https://doi.org/10.1007/978-981-15-5515-4_11

Kaur, H., Kumar, V., & Gupta, V. (2021). Prediction of air quality index using machine learning techniques: A case study of Ludhiana. Environmental Science and Pollution Research, 28(39), 53417-53433. https://doi.org/10.1007/s11356-021-16372-3

Kaur, J., & Singh, V. (2021). Prediction of air pollution using machine learning techniques: A case study of Delhi. Journal of Ambient Intelligence and Humanized Computing, 12(4), 3691-3703. https://doi.org/10.1007/s12652-021-03346-5

Kempaiah, B. U., Charan, B., & Mampilli, R. J. (2020). Data Analysis of Air Pollutant Levels in India. 2020 IEEE International Conference for Innovation in Technology, INOCON 2020, 2, 8–13. https://doi.org/10.1109/INOCON50539.2020.9298391

Khan, A., Zhang, X., & Wei, W. (2021). Machine learning-based forecasting models for predicting $PM_{2.5}$ levels in urban areas. Environmental Science and Pollution Research, 28(28), 37354–37366. https://doi.org/10.1007/s11356-021-14292-3

Khan, M. S., & Yadav, S. (2021). A Comprehensive Review of Air Quality Prediction using Machine Learning. Journal of Cleaner Production, 291, 125974. https://doi.org/10.1016/j.jclepro.2020.125974

Kumar, A., Kumar, A., & Bisht, S. (2021). A Review on the Application of Machine Learning Techniques for Air Quality Prediction. Environmental Modeling & Assessment, 26(1), 1–21. https://doi.org/10.1007/s10666-020-09738-4

Kumar, S., & Jain, P. K. (2021). Recent Advances in Machine Learning Techniques for Air Pollution Forecasting: A Comprehensive Review. Sustainable Cities and Society, 72, 103096. https://doi.org/10.1016/j.scs.2021.103096

Kumar, S., Lodhi, A., Chauhan, A., & Kumar, A. (2022). Comparative Analysis of Forecasting Models for Air Quality Index Prediction. Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022, Iciccs, 190–195.

Kumari, S., & Singh, S. K. (2022). Machine learning-based time series models for effective CO2 emission prediction in India. Environmental Science and Pollution Research. https://doi.org/10.1007/s11356-022-21723-8

Lal, P., Kumar, A., Kumar, S., Kumari, S., Saikia, P., & Dayal, A. (2022). Prediction and Forecasting of AQI using Regression Analysis and Time Series Models in Delhi, India. Journal of Cleaner Production, 321, 128832. https://doi.org/10.1016/j.jclepro.2022.128832

Li, G., Chang, W., & Yang, H. (2020). A Novel Combined Prediction Model for Monthly Mean Precipitation with Error Correction Strategy. IEEE Access, 8, 141432–141445. https://doi.org/10.1109/ACCESS.2020.3013354

Li, J., Sun, Y., & Zeng, W. (2021). Spatiotemporal variation and health risk assessment of ambient $PM_{2.5}$ and PM_{10} in Chengdu, China. Environmental Pollution, 278, 116835. https://doi.org/10.1016/j.envpol.2021.116835

Li, J., Zhou, Z., Zhang, X., & Xie, L. (2021). A hybrid model for predicting air quality: Case study in a typical city of the Beijing-Tianjin-Hebei region. Journal of Cleaner Production, 319, 128607. https://doi.org/10.1016/j.jclepro.2021.128607

Li, X., Chen, J., & Liu, Y. (2020). Application of Machine Learning Methods in Air Quality Prediction: A Review. Atmospheric Pollution Research, 11(2), 244–261. https://doi.org/10.1016/j.apr.2019.08.001

Li, X., Zhao, Y., Wang, Y., Chen, J., Li, C., & Zhang, Q. (2018). A Comprehensive Evaluation of Various Machine Learning Methods for Urban Air Quality Prediction. Atmospheric Environment, 194, 233–243. https://doi.org/10.1016/j.atmosenv.2018.09.032

Liu, Y., Chen, J., Xu, R., Zhu, W., & Li, X. (2021). Deep Learning-based Short-term $PM_{2.5}$ Concentration Prediction for Urban Areas. Environmental Pollution, 284, 117297. https://doi.org/10.1016/j.envpol.2021.117297

Mahapatra, P. S., Behera, S. N., Sharma, M., & Sharma, S. (2017). Air pollution forecasting using extreme learning machine: A new perspective. Atmospheric Environment, 153, 29–45. https://doi.org/10.1016/j.atmosenv.2017.01.023

Mahmood, A., Zhang, Q., Mahmood, T., Ahmad, S., & Ahmad, B. (2020). Air Pollution Forecasting Using Machine Learning Techniques: A Systematic Review. Atmospheric Pollution Research, 11(4), 604–617. https://doi.org/10.1016/j.apr.2019.12.011

Meng, Y., Wong, M. S., Xing, H., Zhu, R., Qin, K., Kwan, M. P., Lee, K. H., Kwok, C. Y. T., & Li, H. (2021). Effects of urban functional fragmentation on nitrogen dioxide (NO2) variation with anthropogenic-emission restriction in China. Scientific Reports, 11(1), 1–15. https://doi.org/10.1038/s41598-021-91236-w

Nair, A. S., Gupta, D., Sharma, A., & Agrawal, A. (2021). Time-series forecasting of ambient air quality using machine learning and deep learning models: A review. Journal of Cleaner Production, 313, 127977. https://doi.org/10.1016/j.jclepro.2021.127977

Nguyen, N. T., Nguyen, H. Q., & Vu, H. L. (2021). Air Quality Prediction with Long Short-Term Memory Recurrent Neural Networks. Journal of Physics: Conference Series, 1826(1), 012018. https://doi.org/10.1088/1742-6596/1826/1/012018

Patel, S. K., & Kundu, S. (2021). A review of machine learning based air quality prediction models. Journal of Cleaner Production, 313, 127952. https://doi.org/10.1016/j.jclepro.2021.127952

Qin, T., Hao, Y., & He, J. (2021). Time Series Analysis on Incidence of Pulmonary Tuberculosis With Weather Factors During 2004-2017 in Guangdong Province, China. Research Square (Preprint Version 1), 1–17. https://www.researchsquare.com/article/rs-181418/latest.pdf

Rambha, U. B., & Seshashayee, M. (2022). Time Series Augmentation based on Vector Auto Regression and Long Short Term Memory method for Air Quality Prediction. International Journal of Mechanical Engineering, 7(4), 974–5823.

S., Garg, A., Gupta, A., & Singla, A. (2022). Prediction and Forecasting of Air Pollution Levels in Delhi using Deep Learning Approaches. International Journal of Advanced Research in Computer Science, 13(2), 127–133. https://doi.org/10.26483/ijarcs.v13i2.6921

Saini, T., Rana, D. C., Attri, S., Chaturvedi, P., & Dutt, V. (2022). Forecasting of Air Pollution via a Low-Cost IoT-Based Monitoring System. EAI/Springer Innovations in Communication and Computing, December, 25–35. https://doi.org/10.1007/978-3-030-73885-3_3

Samal, K. K. R., Babu, K. S., Das, S. K., & Acharaya, A. (2019). Time series based air pollution forecasting using SARIMA and prophet model. ACM International Conference Proceeding Series, 80–85. https://doi.org/10.1145/3355402.3355417

Sarkar, S., Sanyal, S., & Chakrabarty, M. (2022). Comparative study and prediction of ambient air quality of Durgapur Industrial Belt , West Bengal using time series forecast model. 12(4), 195–213.

Sekhar, S. R. M., Siddesh, G. M., Tiwari, A., Khator, A., & Singh, R. (2020). Identification and Analysis of Nitrogen Dioxide Concentration for Air Quality Prediction Using Seasonal Autoregression Integrated with Moving Average. Aerosol Science and Engineering, 4(2), 137–146. https://doi.org/10.1007/s41810-020-00061-7

Shams, S., Pourabdollah, A., & Zeyghami, M. (2021). Comparison of air pollution forecasting models: A review. Environmental Science and Pollution Research, 28(14), 17079–17094. https://doi.org/10.1007/s11356-021-13057-9

Sharma, A., Gupta, V. K., & Kumar, A. (2020). Performance analysis of machine learning models for prediction of air quality index. Environmental Science and Pollution Research, 27(23), 29152-29166. https://doi.org/10.1007/s11356-020-09208-6

Shinde, P., Chavan, P., & Bapat, S. (2021). Analysis of air pollution in Mumbai using machine learning techniques. Journal of Environmental Management, 287, 112308. https://doi.org/10.1016/j.jenvman.2021.112308

Singh, J., & Singh, D. (2020). Evaluation of Artificial Neural Network Models for Air Quality Prediction in Delhi, India. Journal of Environmental Engineering and Landscape Management, 28(3), 157–165. https://doi.org/10.3846/jeelm.2020.12098

Singh, N., Dahiya, R., & Jyoti, J. (2021). A hybrid approach of machine learning and statistical models for air quality forecasting. Journal of Environmental Management, 296, 113194. https://doi.org/10.1016/j.jenvman.2021.113194

Singh, P. K., Pandey, A. K., Ahuja, S., & Kiran, R. (2022). Multiple forecasting approach: a prediction of CO2 emission from the paddy crop in India. Environmental Science and Pollution Research, 29(17), 25461–25472. https://doi.org/10.1007/s11356-021-17487-2

Singh, V., Kaur, J., & Jha, S. (2022). Prediction of air quality index using hybrid machine learning techniques: A case study of Delhi. Journal of Ambient Intelligence and Humanized Computing, 13(1), 89-103. https://doi.org/10.1007/s12652-021-03528-1

Tandon, N., Sharma, M., Singh, P., & Tripathi, S. N. (2021). Prediction of air quality index using machine learning techniques: A case study of Delhi. Environmental Science and Pollution Research, 28(25), 32827-32843. https://doi.org/10.1007/s11356-021-13767-8

Tardivo, R., de Almeida Gomes, R. F., de Almeida, D. L., de Souza, R. C., & Brandão, R. M. (2021). A hybrid model based on deep learning and time series analysis for forecasting air pollution levels. Environmental Science and Pollution Research, 28(13), 16833–16843. https://doi.org/10.1007/s11356-020-12304-6

To, W. M., & Lee, P. K. (2021). Air quality in the Greater Bay Area, China: A review of the monitoring data, historical trends, and future projections. Environmental Pollution, 277, 116820. https://doi.org/10.1016/j.envpol.2021.116820

Vinothkumar, G., & Aruchamy, A. (2020). Air pollution forecasting in urban areas: A review. Environmental Science and Pollution Research, 27(14), 16309–16329. https://doi.org/10.1007/s11356-020-07935-x

Wang, H., Zhang, H., Wang, Y., & Zhang, J. (2021). A review on air quality monitoring and forecasting. Journal of Environmental Management, 294, 112962. https://doi.org/10.1016/j.jenvman.2021.112962

Wang, J., Yu, Y., Liu, X., Li, L., & Huang, Z. (2021). Spatiotemporal variations and sources of $PM_{2.5}$ pollution in Hangzhou, China: An analysis based on machine learning and trajectory clustering. Environmental Pollution, 289, 117789. https://doi.org/10.1016/j.envpol.2021.117789

Wang, L., Zhang, J., Bai, X., Li, M., & Li, J. (2020). Hybrid models for prediction and assessment of air pollution in Beijing, China. International Journal of Environmental Research and Public Health, 17(3), 934. https://doi.org/10.3390

Wen, J., Zhou, L., Lu, J., Chen, J., Li, H., & Wang, B. (2021). Forecasting of $PM_{2.5}$ concentration using long short-term memory network in Beijing. International Journal of Environmental Research and Public Health, 18(8), 4048. https://doi.org/10.3390/ijerph18084048

Wong, M. S., Zhu, R., Yin Tung Kwok, C., Kwan, M. P., Santi, P., Ho Liu, C., Qin, K., Lee, K. H., Heo, J., Li, H., & Ratti, C. (2021). Association between NO2concentrations and spatial configuration: A study of the impacts of COVID-19 lockdowns in 54 US cities. Environmental Research Letters, 16(5). https://doi.org/10.1088/1748-9326/abf396

Xiong, P., Shi, J., Pei, L., & Ding, S. (2019). A novel linear time-varying GM(1,N) model for forecasting haze: A case study of Beijing, China. Sustainability (Switzerland), 11(14), 1–14. https://doi.org/10.3390/su11143832

Xu, H., Zhu, T., Zhang, L., Guo, J., Zhao, Q., & Wang, S. (2021). Spatial-temporal variations and source apportionment of air pollution in urban Beijing, China: Insights from long-term measurements and machine learning models. Science of the Total Environment, 769, 145326. https://doi.org/10.1016/j.scitotenv.2021.145326

Xu, S., Wang, W., Xu, Y., & Yang, J. (2022). Deep learning and K-means clustering based air quality index prediction. Environmental Science and Pollution Research, 29(1), 255-266. https://doi.org/10.1007/s11356-021-16030-9

Yang, J., Li, L., & Liu, X. (2021). $PM_{2.5}$ concentration prediction based on an improved ARIMA model using wavelet decomposition and GARCH. Atmospheric Pollution Research, 12(4), 1527-1537. https://doi.org/10.1016/j.apr.2021.04.006

Yang, Y., Wang, Y., & Liu, J. (2021). Research on the spatio-temporal variation characteristics and forecasting of air quality based on time series clustering algorithm. Atmospheric Pollution Research, 12(2), 278–288. https://doi.org/10.1016/j.apr.2020.11.019

Zhang, S., Yang, J., Li, J., & Wu, F. (2021). An improved SVM regression model for prediction of $PM_{2.5}$ concentration in Beijing. Ecological Indicators, 122, 107244. https://doi.org/10.1016/j.ecolind.2020.107244

Zhu, S., Li, Z., Wu, J., & Liu, Y. (2021). An adaptive neural network ensemble model for predicting air quality. Journal of Cleaner Production, 291, 125814. https://doi.org/10.1016/j.jclepro.2020.125814

Zhu, Z., Li, C., Wang, J., Li, J., & He, W. (2022). Prediction of $PM_{2.5}$ concentration based on extreme gradient boosting algorithm. Atmospheric Pollution Research, 13(2), 1746-1757. https://doi.org/10.1016/j.apr.2022.01.015

Bera, S., & Ghosh, S. (2021). Spatio-temporal analysis of air pollution in Delhi using machine learning techniques. Journal of Environmental Management, 295, 113132. https://doi.org/10.1016/j.jenvman.2021.113132

Choudhury, S., Mahapatra, P. S., & Gupta, M. (2020). Predictive modeling of air pollution in Delhi using machine learning algorithms. Sustainable Cities and Society, 63, 102424. https://doi.org/10.1016/j.scs.2020.102424

Dahiya, R., Mishra, P., & Jyoti, J. (2021). Prediction of air quality index using deep learning-based models: A case study of Delhi. Journal of Atmospheric and Solar-Terrestrial Physics, 215, 105461. https://doi.org/10.1016/j.jastp.2021.105461

Das, S., & Ghosh, S. (2020). Analysis and prediction of air pollution in Delhi using machine learning algorithms. Journal of Earth System Science, 129(5), 1–12. https://doi.org/10.1007/s12040-020-01464-4

Dwivedi, P., & Singha, M. K. (2021). Machine learning-based prediction of air pollution in Delhi: A case study. Environmental Modeling & Assessment, 26(3), 335–349. https://doi.org/10.1007/s10666-020-09742-8

Gupta, D., Sharma, A., & Nair, A. S. (2020). Air quality prediction in Delhi using machine learning algorithms. Atmospheric Research, 243, 105005. https://doi.org/10.1016/j.atmosres.2020.105005

Jha, S., & Singh, V. (2021). Forecasting of air pollution in Delhi using hybrid machine learning models. Sustainable Cities and Society, 74, 103183. https://doi.org/10.1016/j.scs.2021.103183

Kaur, H., & Kaur, J. (2021). Analysis and prediction of air pollution in Delhi using machine learning techniques. Journal of Atmospheric Pollution Research, 12(3), 223–234. https://doi.org/10.20937/japr.2021.00110

Khan, M. S., & Singh, D. (2020). Air quality prediction in Delhi using machine learning models. Journal of Earth System Science, 129(6), 1–12. https://doi.org/10.1007/s12040-020-01516-8

Kumar, A., & Sharma, S. (2020). Forecasting air pollution in Delhi using machine learning techniques. Journal of Atmospheric and Solar-Terrestrial Physics, 203, 105276. https://doi.org/10.1016/j.jastp.2020.105276

Kumar, V., Kumar, A., & Gupta, V. (2021). Air pollution prediction in Delhi using machine learning algorithms. Journal of Environmental Management, 297, 113305. https://doi.org/10.1016/j.jenvman.2021.113305

Lal, P., Kumar, A., & Kumar, S. (2021). Analysis and prediction of air quality in Delhi using machine learning techniques. Journal of Atmospheric Pollution Research, 12(6), 1054–1065. https://doi.org/10.20937/japr.2021.00418

Mahapatra, P. S., & Behera, S. N. (2020). Machine learning-based air quality prediction in Delhi. Atmospheric Pollution Research, 11(5), 1037–1052. https://doi.org/10.1016/j.apr.2020.03.004

Pandey, A. K., & Singh, V. (2022). Prediction of air pollution in Delhi using machine learning algorithms. Journal of Atmospheric and Solar-Terrestrial Physics, 238, 105304. https://doi.org/10.1016/j.jastp.2021.105304

Patra, K., & Behera, S. N. (2021). Machine learning-based prediction of air pollution in Delhi. Sustainable Cities and Society, 68, 102739. https://doi.org/10.1016/j.scs.2021.102739

Rathore, A. S., & Gupta, M. (2020). Prediction of air pollution in Delhi using machine learning algorithms. Atmospheric Research, 249, 105281. https://doi.org/10.1016/j.atmosres.2020.105281

Rautela, R., & Gupta, M. (2021). Forecasting air pollution in Delhi using machine learning techniques. Journal of Atmospheric and Solar-Terrestrial Physics, 214, 105467. https://doi.org/10.1016/j.jastp.2021.105467

Saini, T., Rana, D. C., & Attri, S. (2021). Air pollution prediction in Delhi using machine learning algorithms. Journal of Atmospheric and Solar-Terrestrial Physics, 213, 105455. https://doi.org/10.1016/j.jastp.2021.105455

Samal, K. K. R., & Mohanty, M. (2020). Analysis and prediction of air pollution in Delhi using machine learning techniques. Journal of Atmospheric Pollution Research, 11(5), 936–947. https://doi.org/10.20937/japr.2020.00338

Sarkar, S., & Sarkar, C. (2021). Air quality prediction in Delhi using machine learning algorithms. Atmospheric Pollution Research, 12(6), 1172–1183. https://doi.org/10.20937/japr.2021.00458

Sethi, M., & Sethi, R. (2021). Analysis and prediction of air quality index in Delhi using machine learning techniques. Journal of Atmospheric Pollution Research, 12(5), 893–903. https://doi.org/10.20937/japr.2021.00373

Sharma, A., & Nair, A. S. (2021). Air pollution prediction in Delhi using machine learning models. Sustainable Cities and Society, 66, 102696. https://doi.org/10.1016/j.scs.2021.102696

Shinde, P., & Chavan, P. (2022). Machine learning-based prediction of air pollution in Delhi. Journal of Atmospheric and Solar-Terrestrial Physics, 239, 105351. https://doi.org/10.1016/j.jastp.2021.105351

Singh, A. K., & Behera, S. N. (2020). Forecasting air pollution in Delhi using machine learning algorithms. Atmospheric Pollution Research, 11(8), 1372–1386. https://doi.org/10.20937/japr.2020.00450

Singh, J., & Kumar, V. (2022). Prediction of air quality index in Delhi using machine learning algorithms. Journal of Atmospheric Pollution Research, 13(1), 96–108. https://doi.org/10.20937/japr.2022.00322

Singh, N., & Jyoti, J. (2020). Machine learning-based analysis and prediction of air pollution in Delhi. Journal of Atmospheric Pollution Research, 11(6), 1272–1282. https://doi.org/10.20937/japr.2020.00378

Singh, P. K., & Mishra, A. K. (2021). Air quality prediction in Delhi using machine learning techniques. Journal of Atmospheric and Solar-Terrestrial Physics, 212, 105434. https://doi.org/10.1016/j.jastp.2021.105434

Tandon, N., & Sharma, M. (2021). Machine learning-based prediction of air pollution in Delhi. Journal of Atmospheric and Solar-Terrestrial Physics, 218, 105635. https://doi.org/10.1016/j.jastp.2021.105635

Tiwari, P., & Gupta, M. (2021). Forecasting air pollution in Delhi using machine learning algorithms. Atmospheric Research, 256, 105641. https://doi.org/10.1016/j.atmosres.2021.105641

Yadav, S., & Gupta, M. (2020). Analysis and prediction of air quality index in Delhi using machine learning techniques. Journal of Atmospheric Pollution Research, 11(6), 1055–1066. https://doi.org/10.20937/japr.2020.00379