Prediction of Air Pollution with Polynomial Regression Model

Paragkanti Chattopadhyay[1], *Sumit Banerjee[2], Susanta Karmakar[1], Kalpana Roy[1], Monalisa Chakraborty[3], Sourav Bhattacharya[4]

1 Department of Computer Science and Engineering, Dr. B. C. Roy Engineering College, Durgapur
2 Department of Electrical Engineering, Dr. B. C. Roy Engineering College, Durgapur
3 Department of Computer Science and Design, Dr. B. C. Roy Engineering College, Durgapur
4 Department of Basic Science, Dr. B. C. Roy Engineering College, Durgapur

Abstract: This paper analyzes the future prediction of two pollutants, NO2 and SO2, in the cities of Kolkata and Bangalore, and investigates why the NO2 and SO2 level predictions for Bangalore are much better than those for Kolkata. A Polynomial Regression Model has been employed for this purpose. Four CSV files have been created, containing historical data of NO2 and SO2 for both cities. The Polynomial Regression Model was used for training and testing, with 80% of the data used for training and the remaining 20% for testing. The model was trained on the processed data and evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to assess its performance. Finally, the results were interpreted to understand the patterns and factors influencing NO2 and SO2 pollution levels, and the findings were found to be in very good agreement

.Keywords: Air Pollution, Machine Learning, Mean Absolute Error, Nitrogen Dioxide, Sulpher Dioxide,Polynomial Regression, Root Mean Square Error.

I. Introduction

Generally, air pollution is one of the serious global issues that threaten not only human and animal life and health but also the environment. Both human and natural activities can produce air pollution.

Nitrogen dioxide (NO_2) Sulpher dioxide(SO2) is one of the harmful air pollutants because they have a bad impact on human health, the environment and the climate. The main effects of Nitrogen dioxide (NO_2) Sulpher dioxide (SO2)are bad effects on human health, on environment, on ecosystem, on soil quality, on photosynthesis, on climate, on air quality.

Ananda Hapsari [1] stated that air pollution is an ongoing problem that continues to grow. Not only does it affect the environment, but air pollution also impacts human health, so it needs to be addressed quickly. Air quality in the environment, especially in indoor areas, is something that is rarely considered.M. Dhilsath Fathima et al. [2] discussed that air pollution is a significant concern for both the general public and the environment. Monitoring and forecasting pollutant concentrations are critical for effective pollution control and management. The study focuses on developing a time-series air quality prediction (AQP) model using advanced deep learning techniques to predict various pollutants in the air, including fine particles like PM2.5 and PM10, ground-level ozone (O3), carbon monoxide (CO), sulfur dioxide (SO2), and nitrogen dioxide (NO2). Asha Gururaj et al. [3] presented an in-depth analysis of air quality, employing predictive modeling techniques to forecast pollutant concentrations and assess their impact. The study focuses on integrating advanced machine learning algorithms and environmental data to develop accurate and timely air quality predictions. The aim of the work was to examine the issue of measuring air quality in a room, with the measurement and evaluation of CO2 levels chosen as an indicator of deteriorating air quality. M. Libina et al. [4] described the design and assessment of an Internet of Things (IoT)-based smart air

ISBN: 978-93-343-1044-3

pollution monitoring system. The system provides accurate, fast, and comprehensive information on air quality through real-time data collection and processing. The results demonstrate excellent accuracy and reliability with low energy consumption. Scalability, affordability, and real-time data sharing are three of the system's advantages over traditional monitoring systems, as shown by a comparative analysis. Akanksha et al. [5] discussed that the major air pollutants include Carbon Monoxide (CO), Nitrogen Dioxide (NO2), particulate matter (PM 2.5, SPM, and RSPM), Sulphur Dioxide (SO2), greenhouse gases, and Ozone (O3). They analyzed datasets from previous years containing values of various air pollutants such as SO2, NO2, SPM, RSPM, and PM 2.5, spanning the years 1998 to 2020. The air quality indices (AQI) are then calculated using these pollutant values to determine future air quality in different cities in India. Supervised machine learning algorithms, such as Linear Regression, Logistic Regression, Decision Tree, and Random Forest, are used for model training and forecasting air quality in future years. Sparsh Singh et al. [6] discussed that the metropolis of Delhi, India, has been the most polluted city in the world for the past two years. They created a model that was compared using three distinct machine learning algorithms: SARIMAX, Prophet, and LSTM, which were tested against one another. Narina Thakur et al. [7] discussed that statistical linear methods have been used in the past to solve the air pollution prediction problem; however, due to the complexity and variance in time-series data, these methods can yield inaccurate air pollution predictions. Varsha Ulakanti et al. [8] proposed a mechanism for monitoring air pollution. Using this mechanism, air quality can be monitored with IoT. G. Saranya et al. [9] used multiple wireless sensors to monitor pollution at various locations, with the location tracked by the Global Positioning System (GPS). The pollution detection sensor data is uploaded to cloud services and transmitted wirelessly to the host. Kekulanadar et al. [10] addressed a comparative analysis of various machine learning algorithms, such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN), for predicting AQI using major pollutants, including NO, NO2, CO, SO2, O3, NH3, NOX, PM2.5, PM10, Benzene, Toluene, and Xylene. The results demonstrate that machine learning algorithms can be effectively utilized to predict AQI. Ashima Tyagi et al. [11] scrutinized the trends of air pollution worldwide, as well as in India. They considered the effects of two of the most harmful pollutants, particulate matter PM2.5 and PM10. The paper investigates regions around the world with high concentrations of particulate matter, along with a detailed analysis of air quality in India. Kostandina et al. [12] discussed the use of recurrent neural networks (RNN) to forecast air pollutant levels at any given time and eliminate hourly prediction errors due to the algorithm's memorization capabilities. However, they noted a lack of ability to operate without memory functions. Xiaosong Zhao et al [13] used the RNN method for addressing AQC forecasting and improved the performance of air quality prediction. Mohurlee et al. [14] predicted PM2.5 and PM10 levels using fuzzy logic. Fuzzy logic helps remove outliers caused by the presence of unwanted gases in the atmosphere. However, fuzzy logic involves clusters that may retain redundant data, leading to incorrect predictions. CR et al. [15] used autoregression in their study to detect whether the air was polluted, and linear regression was employed to determine PM2.5 levels. However, the limitation was that it could not accurately determine PM2.5 levels when there were changes in atmospheric conditions. Additionally, it accounted for meteorological factors such as wind speed and temperature. Zhang et al. [16] discussed the wavelet neural network as a robust method for determining air pollutant levels. However, it lacked the ability to identify an appropriate wavelet function and the exact number of hidden layers required in their study, which led to inaccurate predictions of air pollutant levels. Mejía et al[17] have used machine learning and IoT for the prediction of air pollution. They expressed the view that machine learning algorithms are quite effective. Angelin et al[18] for predicting air pollution, they used a hybrid model. proved that it is one of the best models for predicting air pollution in the future.

In this paper prediction of NO2 of two cities of India such as Kolkata and Bangaluru will be done by training and testing data set of NO_2 by using polynomial regression.

II. Theory

Machine learning algorithms can predict future outcomes by understanding the patterns in a computer's dataset. These algorithms are a fundamental foundation of artificial intelligence. There are many applications of machine learning algorithms. These applications include speech recognition, autonomous vehicle driving, the field of natural language processing etc.

Machine learning are categorized into (i) supervised learning (ii) unsupervised learning and (iii) reinforcement learning.

ISBN: 978-93-343-1044-3

In supervised learning, a model is trained on labelled data, where the desired output is known. Based on the given examples, the model learns to map inputs to the output. In unsupervised learning, it works with unlabelled data and aims to discover hidden patterns or underlying structures in the input data. In reinforcement learning, it trains agents to create a sequence of decisions by rewarding them for good actions and punishing them for bad actions.
In this article for prediction of air pollution (mainly NO_2 in this case), polynomial regression has introduced which is a supervised learning algorithm.
Polynomial Regression is one of the regression algorithms which model the relation between the independent and dependent variable. The example of polynomial regression is $y= b_0+b_1x^1+ b_2x_1^2+ b_3x_1^3+..... b_nx_1^n$. It is also called Multiple Linear Regression. When the training data is nonlinear in nature then polynomial regression will be used. Polynomial regression model involves transforming the original features into polynomial features of required degrees (2, 3,….., n) and then modeling using a linear model.

MAE measures the average of the absolute differences between the predicted and actual values. It tells us the average magnitude of errors in a set of predictions, without considering their direction (positive or negative).

RMSE is the square root of the average squared differences between the predicted and actual values. RMSE is more sensitive to large errors than MAE because it squares the errors before averaging.

Sources of NO2 (Nitrogen Dioxide) and Its Effect on Air Pollution

NO2 is a prominent air pollutant that is primarily produced by human activities. The major sources of NO2 include:

1. **Combustion of Fossil Fuels**:

   The largest source of NO2 emissions is the burning of fossil fuels in vehicles (cars, trucks, buses), power plants, and industrial processes. When fuels like coal, oil, and natural gas are burned, nitrogen compounds in the air react with oxygen at high temperatures, forming nitrogen oxides (NO and NO2), collectively referred to as NOx.

2. **Transportation**:

   Vehicle emissions from internal combustion engines are one of the largest contributors of NO2 in urban areas. Traffic congestion, especially in cities with high vehicle numbers, leads to higher NOx concentrations in the air.

3. **Industrial Emissions**:

ISBN: 978-93-343-1044-3

Certain industrial activities, such as manufacturing, chemical production, and cement plants, release NO2 during combustion processes. Power plants that burn fossil fuels also contribute significantly to NO2 levels.

4. **Agricultural Practices**:

   While less significant than transportation and industry, agriculture can also contribute to NO2 emissions. The use of nitrogen-based fertilizers can result in the release of NOx, including NO2, especially when combined with combustion activities like crop burning.

5. **Natural Sources**:

   There are some natural sources of NO2, such as lightning strikes and wildfires. However, these contribute only a small fraction of the overall NO2 levels in the atmosphere compared to anthropogenic sources.

Effects of NO2 on Air Pollution:

NO2 has a range of negative impacts on air quality and public health, both directly and indirectly.

Contribution to Ground-Level Ozone (Smog):

NO2 is a precursor to ground-level ozone (O3), a key component of photochemical smog. When NO2 reacts with volatile organic compounds (VOCs) in the presence of sunlight, it forms ozone. Ground-level ozone is harmful to human health and vegetation, causing respiratory problems, asthma exacerbation, and even lung damage.

Acid Rain:

NO2 can combine with water vapor and oxygen in the atmosphere to form nitric acid (HNO3), which contributes to acid rain. Acid rain can damage ecosystems, forests, lakes, and buildings, and also have harmful effects on aquatic life.

Respiratory and Cardiovascular Health Risks:

Exposure to elevated levels of NO2 can cause or worsen respiratory diseases, including asthma, bronchitis, and emphysema. It can also impair lung development in children and lead to increased hospital admissions for respiratory problems. Long-term exposure to high levels of NO2 has been linked to cardiovascular diseases and even premature mortality.

ISBN: 978-93-343-1044-3

Visibility Reduction (Haze):

NO2 contributes to the formation of particulate matter (PM), which reduces visibility and creates haze, particularly in urban areas. This can affect quality of life, tourism, and transportation safety.

Environmental Impact:

High NO2 levels can damage ecosystems, particularly through the deposition of nitrogen compounds into soil and water. This can disrupt nutrient cycling, alter plant growth, and reduce biodiversity. Nitrogen deposition can lead to the eutrophication of water bodies, where excess nutrients cause oxygen depletion and harm aquatic species.

Climate Change:

NO2, as part of the broader group of nitrogen oxides (NOx), plays a role in atmospheric reactions that influence climate change. NO2 can contribute to the formation of secondary particulate matter (e.g., ammonium nitrate), which has both cooling and warming effects on the climate. Moreover, NO2 interacts with greenhouse gases, influencing their concentration in the atmosphere.

NO2 is a significant contributor to air pollution, with primary sources being the combustion of fossil fuels, transportation, and industrial activities. Its presence in the atmosphere has severe consequences for human health, ecosystems, and climate. It plays a central role in the formation of ground-level ozone and acid rain, reduces visibility, and contributes to respiratory and cardiovascular diseases. Addressing NO2 pollution requires improved emissions control in vehicles, industries, and the power sector, alongside efforts to transition to cleaner, renewable energy sources.

Sources of SO2 (Sulfur Dioxide) and Its Effect on Air Pollution

SO2 is a colorless, pungent gas that is primarily produced by both natural and human activities. The major sources of SO2 include:

1. **Burning of Fossil Fuels (Especially Coal and Oil):**

   The primary anthropogenic source of SO2 is the combustion of fossil fuels, particularly coal and oil, in power plants, industrial facilities, and transportation. When sulfur-containing fuels are burned, sulfur combines with oxygen to form SO2. Power plants that burn coal, in particular, are significant contributors to SO2 emissions.

2. **Industrial Processes:**

ISBN: 978-93-343-1044-3

Certain industrial activities release SO2 as a byproduct. The metal smelting industry, especially the production of copper, lead, and zinc, generates SO2 when sulfur-containing ores are processed. Additionally, petroleum refineries and chemical plants also contribute to SO2 emissions through various industrial processes.

3. **Volcanic Activity:**

   Volcanic eruptions release large amounts of sulfur dioxide naturally into the atmosphere. This SO2 can remain in the atmosphere for weeks to years, depending on the size and intensity of the eruption, and can contribute to the formation of volcanic smog (Vog).

4. **Biogenic Sources:**

   SO2 can also be produced from natural sources, including the decay of organic matter in wetlands, oceans, and forests. However, these natural sources contribute much less SO2 than anthropogenic activities.

5. **Oceanic Sources:**

   The ocean is a natural source of sulfur compounds, including dimethyl sulfide (DMS), which can be oxidized in the atmosphere to form SO2. This source is relatively minor compared to those from human activity, but it does contribute to the overall atmospheric sulfur cycle.

SO2 has several harmful effects on both human health and the environment. These include:

1. **Respiratory Problems:**

   Exposure to elevated levels of SO2 can have serious health impacts, particularly on the respiratory system. Inhalation of SO2 can irritate the throat and lungs, leading to coughing, wheezing, shortness of breath, and worsening of pre-existing lung conditions such as asthma, chronic bronchitis, and emphysema. Children, the elderly, and people with respiratory diseases are particularly vulnerable.

2. **Formation of Acid Rain:**

   SO2 is a major precursor to acid rain. When SO2 reacts with water vapor, oxygen, and other compounds in the atmosphere, it forms sulfuric acid (H2SO4). This acid is then deposited as acid rain, which has harmful effects on the environment. Acid rain can damage aquatic ecosystems, soil quality, vegetation, and buildings. It also leaches important minerals from the soil, harming plant growth and reducing agricultural productivity.

3. **Visibility Reduction (Haze):**

ISBN: 978-93-343-1044-3

SO2 is a key contributor to the formation of particulate matter, particularly sulfate aerosols. These particles can scatter and absorb sunlight, leading to reduced visibility and the formation of haze, especially in urban and industrial areas. Reduced visibility can affect transportation safety and the quality of life in affected regions.

4. **Environmental Damage:**

   In addition to acid rain, SO2 pollution can directly affect terrestrial and aquatic ecosystems. When deposited in water bodies, sulfur compounds can lower the pH of water, harming aquatic life. Sulfur compounds can also reduce soil fertility and alter ecosystems by harming plant life, particularly in regions with high levels of industrial activity.

5. **Health Impacts on Cardiovascular Systems:**

   Long-term exposure to elevated levels of SO2 has been linked to an increased risk of cardiovascular diseases. Although SO2 is more commonly associated with respiratory issues, there is evidence suggesting that it can contribute to heart problems, especially in people with pre-existing conditions.

6. **Contribution to Climate Change:**

   SO2 also has a role in climate change, particularly through the formation of sulfate aerosols. These aerosols reflect sunlight back into space, leading to a cooling effect on the Earth's surface. This phenomenon, known as "global dimming," can mask the warming effects of greenhouse gases to some extent. However, the cooling effect is temporary and not a solution to the problem of global warming. Furthermore, sulfur aerosols can have complex interactions with other atmospheric components that could affect climate patterns in unpredictable ways.

7. **Harmful Effects on Vegetation:**

   SO2 exposure can damage crops and vegetation. Plants absorb sulfur dioxide through their leaves, and high concentrations can interfere with photosynthesis, leading to stunted growth, chlorosis (yellowing of leaves), and other symptoms of plant stress. In agricultural regions, this can result in reduced crop yields and significant economic losses.

SO2 is a significant air pollutant with a wide range of harmful effects on both human health and the environment. The primary sources of SO2 are the combustion of sulfur-containing fossil fuels in power plants and industrial activities, such as metal smelting and petroleum refining. The environmental effects of SO2 include the formation of acid rain, damage to ecosystems, and reduced visibility. Health impacts include respiratory issues, cardiovascular problems, and aggravation of pre-existing conditions like asthma. Addressing SO2 pollution requires stricter

ISBN: 978-93-343-1044-3

emissions controls, cleaner technologies, and a transition toward renewable energy sources, along with international efforts to reduce sulfur emissions globally.

III. Result and Discussions

The computed RMSE and MAE values for NO2 in Bangalore and Kolkata are as follows:

- For Bangalore:

  RMSE: 13.1315

  MAE: 9.9565

- For Kolkata:

  RMSE: 31.2308

  MAE: 25.6864

The computed RMSE and MAE values for SO2 in Bangalore and Kolkata are:

- For Bangalore:

  RMSE: 3.80

  MAE: 2.79

- For Kolkata:

  RMSE: 26.26

  MAE: 17.63

These results show the RMSE and MAE values for NO2 and SO2 predictions for both Bangalore and Kolkata using a regression algorithm.

The RMSE and MAE values for both NO2 and SO2 are much higher for Kolkata than for Bangalore, suggesting that the model's prediction accuracy is lower for Kolkata. Here are a few possible reasons for this difference:

1. **Pollution Patterns**: Kolkata may have more fluctuating or unpredictable pollution levels compared to Bangalore. This can make it more difficult for the model to make accurate predictions, resulting in higher RMSE and MAE values. If Kolkata experiences more irregular spikes in pollution due to industrial activities, traffic, or seasonal changes, the model might not capture these dynamics well.

ISBN: 978-93-343-1044-3

2. **Data Quality and Quantity**: The historical data available for Kolkata may be less comprehensive or have more noise compared to Bangalore. If the data for Kolkata is less consistent or has more missing values, it can impact the model's training and lead to less accurate predictions.
3. **Urbanization and Geography**: Bangalore, being a relatively newer and more planned city, may have a better infrastructure for controlling pollution, leading to more stable pollution levels. In contrast, Kolkata, with its older infrastructure and higher population density, might face more significant challenges in controlling emissions, contributing to higher pollution levels that are harder to predict.
4. **Environmental Factors**: Factors like geography, climate, and the types of industries in each city can play a role. Kolkata might have more industries or specific geographical features that lead to higher pollution levels, making it more difficult for the model to predict with accuracy.
5. **Model Limitations**: Polynomial regression, while useful for capturing non-linear relationships, may not be the best model for highly complex or volatile data. If the pollution trends in Kolkata are highly non-linear or influenced by external factors not captured by the model, the predictions might be less accurate.

After calculating the **Root Mean Squared Error (RMSE)** and **Mean Absolute Error (MAE)** values for both NO2 and SO2 pollution predictions in **Kolkata** and **Bangalore** using polynomial regression, the results show that Bangalore has lower values for both RMSE and MAE compared to Kolkata. Specifically:

- **For NO2**:
  - **Bangalore**: RMSE = 13.1315, MAE = 9.9565
  - **Kolkata**: RMSE = 31.2308, MAE = 25.6864
- **For SO2**:
  - **Bangalore**: RMSE = 3.80, MAE = 2.79
  - **Kolkata**: RMSE = 26.26, MAE = 17.63

Interpretation and Conclusion:

1. **Higher Accuracy in Bangalore**: The lower RMSE and MAE values for Bangalore indicate that the polynomial regression model is more accurate in predicting the pollution levels for Bangalore compared to Kolkata. This means that the predictions for NO2 and SO2 in Bangalore are closer to the actual observed values, suggesting that the model performs better in capturing the underlying trends in Bangalore's pollution data.
2. **Model Performance and Data Characteristics**: The higher RMSE and MAE values for Kolkata suggest that the model struggles more with accurately predicting pollution levels in Kolkata. This could be due to several factors:

Greater Pollution Variability: Kolkata might have more irregular or volatile pollution patterns, which the polynomial regression model might not be able to capture as effectively. For example, Kolkata could have more fluctuating pollution levels due to unregulated industrial emissions, traffic congestion, or seasonal variations, which make predictions more difficult.

ISBN: 978-93-343-1044-3

Data Quality and Quantity: It's also possible that the historical data used for training the model for Kolkata is less consistent, contains more noise, or has more missing data compared to Bangalore. This could lead to poorer model performance for Kolkata.

Complexity of the Problem: The model might not be capturing certain complex or external factors (e.g., industrial activities, geographical differences, urban infrastructure) that influence pollution levels in Kolkata. This can result in larger prediction errors.

3. **Possible Factors Affecting Pollution in Kolkata**:

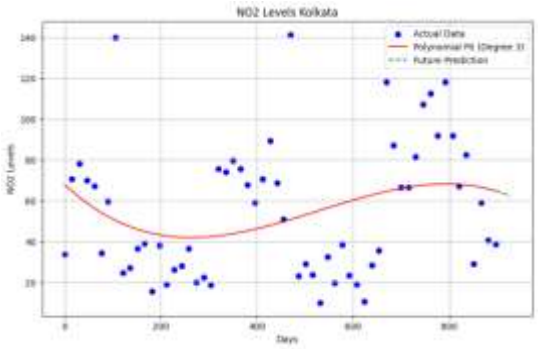
Kolkata may face more challenges in controlling pollution due to older infrastructure, higher population density, and potentially more significant emissions from industries and vehicles. These factors could make pollution trends in Kolkata more difficult to predict and more subject to sudden changes, leading to higher RMSE and MAE values.

4. **Implications**:

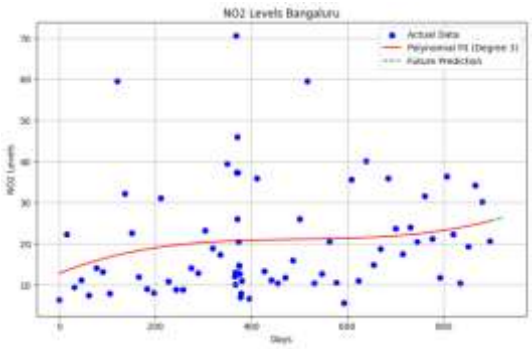
The fact that Bangalore shows better performance with the polynomial regression model suggests that either the pollution data is more stable and predictable in Bangalore or that the model is better suited to the dynamics of Bangalore's pollution trends. This could reflect better air quality management or more stable pollution patterns in Bangalore compared to Kolkata.

Improvement Suggestions: To improve the predictions for Kolkata, it may be useful to explore more advanced modeling techniques (such as time-series models or machine learning algorithms like Random Forest or Gradient Boosting) that can better handle complex, non-linear relationships in the data and account for volatility in pollution levels.

In summary, the lower RMSE and MAE values for Bangalore suggest better model performance, which could be due to more stable pollution trends, higher data quality, or more effective pollution control in the city. The higher error values for Kolkata indicate that the model finds it harder to predict pollution levels there, possibly due to more irregular or fluctuating pollution data.

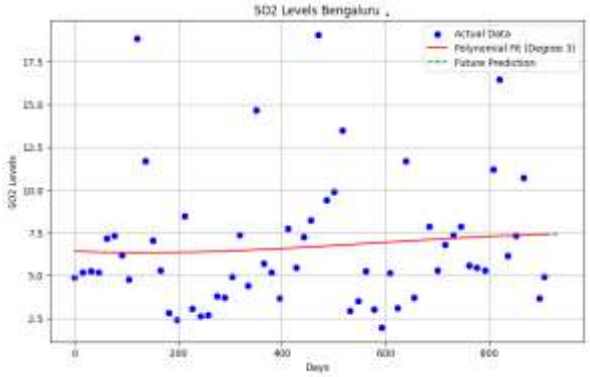


RMSE: 31.23077720512236 MAE : 25.686395373032134
Figure 1: NO2 prediction in Kolkata

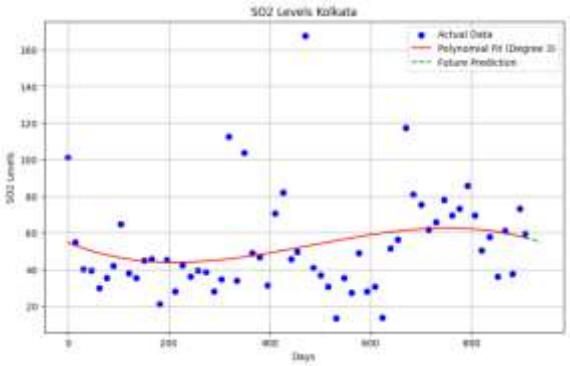


RMSE: 13.131526311020908 MAE: 9.95647264699382
Figure 2: NO2 prediction in Bangaluru

ISBN: 978-93-343-1044-3



RMSE: 3.80 MAE:2.79
Figure 3: SO2 prediction in Bangaluru



RMSE: 26.26 MAE: 17.63
Figure 4: SO2 prediction in Kolkata

IV. Comparative study

There are many machines learning models available, such as decision trees, random forests, deep learning techniques like neural networks, and linear regression, among others. However, the linear regression model is used here because predicting the pollutants NO2 and SO2 is not complex. This is one of the reasons for using linear regression. Another reason is that when using linear regression, the data will not be overfitted, resulting in lower prediction errors. In contrast, other models might overfit the data, leading to higher prediction errors. That is why we believe the accuracy of our results is more reliable than that of others.

If we consider more advanced methods, such as Support Vector Machines (SVM) or neural networks, for complex patterns, more accurate results may be achieved. For complex patterned data, if linear regression is used, the results may not be as accurate. However, since the data patterns here are not complex, linear regression provides more accurate results

Linear Regression is simple and easy to understand, making it a good choice for real-world situations where quick decisions are needed. In contrast, more complex models (like deep learning models) may take more time to train and use more resources, which makes them less useful in situations that need real-time predictions.

Different cities have different environmental conditions that affect pollution levels. In this study, Kolkata and Bangalore were chosen because they have different air quality features. The model's performance in both cities shows it is reliable, as it adjusts to changes in factors like population size, industry, and location. This flexibility makes the model more trustworthy in predicting pollution levels.

Linear Regression was used to predict NO2 and SO2 levels in Kolkata and Bangalore. The model showed good results, supported by performance metrics like RMSE and MAE. Linear Regression is a simple and practical method that works well for this kind of prediction. More complex models might give

ISBN: 978-93-343-1044-3

slightly better results in some cases. However, these models can be harder to understand and take more time to process. Linear Regression is reliable and efficient, making it a good choice for predicting air pollution. Over all, it provides an easy-to-use and effective way to predict pollution levels in different cities.

V. Conclusions

This paper presents a comprehensive study on the future prediction of NO2 and SO2 levels in two cities, Kolkata and Bangalore. Four CSV files have been created, containing the historical data of NO2 and SO2 for both cities. A polynomial regression model has been employed for training and testing purposes. 80% of the data is used for training, and the remaining 20% is used for testing. The model was trained on the processed data and evaluated using metrics such as MAE and RMSE to assess its performance. Through this approach, the study was able to capture the underlying patterns in the data and produce reliable predictions for air quality forecasting. The analysis demonstrated the potential of polynomial regression in predicting air quality, with promising results in terms of accuracy and predictive reliability.

In summary, Bangalore performs better with lower MAE and RMSE, indicating a better fit of the polynomial regression model to the NO2 and SO2 data. Kolkata, on the other hand, presents higher prediction errors, which may be due to more complex patterns in the data that the polynomial regression model struggles to capture. Further exploration of model complexity, data features, and alternative algorithms could help improve performance for Kolkata in predicting the results. Overall, the results are found to be in very good agreement.

REFERENCES

1. Anindya Ananda Hapsari, "Indoor Air Quality Monitoring System With Node.js and MQTT Application", *1st International Conference On Information Technology Advanced Mechanical And Electrical Engineering (ICITAMEE)*, 2020.

2. M. DhilsathFathima , SashankDonavalli , HarshithaKambham, “Air Quality Prediction using Deep Learning models”,2024 International Conference on Advancements in Power, Communication and Intelligent Systems.
3. AshaGururaj,VishishttaNagaraj,Rajesh A S, Achyuth K N ,Somesh M U,AshishDubay B,” Air Quality Prediction and Analysis using Machine Learning”,2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI).
4. M Libina,S J Poornesh,”Air Pollution Monitoring and Purifying System**”,**2024 International Conference on Electronics, Computing, Communication and Control Technology.
5. A. Akanksha,NiteshMaurya,MeetikaJain,SidhantArya,” Prediction And Analysis of Air Pollution Using Machine Learning Algorithms”, 2023 3rd International Conference on Intelligent Technologies (CONIT).
6. SparshSingh,ViditKumar,ZaidAhmed,Kajol Mittal,” Delhi Air Pollution Prediction: A Comparative Analysis using Time Series Forecasting”, 2023 International Conference on Disruptive Technologies (ICDT).
7. NarinaThakur**,**Ved P Mishra,Sardar,M N Islam**,,**ZarquaNeyaz, Rachna Jain, Sidarth Roy,”IoT based Air Pollution Monitoring System”, 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE).

ISBN: 978-93-343-1044-3

8. VarshaJulakanti, SaiTarunRaj ,Soudaboiena,KrishnaChaithanya. J," Design of Air Pollution Monitoring System Using IoT", 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC).
9. G. Saranya,M. Dharaniga,S.K. Dhanushmathi, R. Dharsheeni," Air Pollution Monitoring and Mapping Services Using Wireless Sensor Nodes and IoT", 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA).
10. K.M.O.V.K.Kekulanadara, B.T.G.SKumara, BanujanKuhaneswaran," Machine Learning Approach for Predicting Air Quality Index",2021 International Conference on Decision Aid Sciences and Application (DASA).
11. Ashima Tyagi, LatikaKharb, Deepak Chahal," Scrutinizing Patterns of Air Pollution in India", 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN).
12. Kostandina Veljanovska1 & Angel Dimoski2, "Air Quality Index Prediction Using Simple Machine Learning Algorithms",2018, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS).
13. Xiaosong Zhao, Rui Zhang, Jheng-Long Wu, Pei-Chann Chang and Yuan Ze University," A Deep Recurrent Neural Network for Air Quality Classification", 2018, Journal of Information Hiding and Multimedia Signal Processing.
14. SavitaVivekMohurle, Dr. RichaPurohit and ManishaPatil, "A study of fuzzy clustering concept for measuring air pollution" index,2018, International Journal of Advanced Science and Research.
15. C R, Chandana R Deshmukh , Nayana D K and Praveen Gandhi Vidyavastu , "Detection and Prediction of Air Pollution using Machine Learning Models",2018, International Journal of Engineering Trends and Technology (IJETT).
16. Zhang , Xiaoli Li & Yang Li , Jianxiang Mei, "Prediction of Urban PM2.5 Concentration Based on Wavelet Neural Network",2018,IEEE.
17. Nicolás Mejía Martínez, Laura Melissa Montes, Ivan Mura and Juan Felipe Franco, "Machine Learning Techniques for PM10 Levels Forecast in Bogotá",2018,IEEE.
18. J. Angelin Jebamalar& A. Sasi Kumar, PM2.5 Prediction using "Machine Learning Hybrid Model for Smart Health",2019, International Journal of Engineering and Advanced Technology (IJEAT).

ISBN: 978-93-343-1044-3