1. Abstract

Air pollution is still a major environmental and public health concern, it especially in developing nations like India causes problems. Using machine learning methods to find and evaluate important pollutants and evaluate the impact of the COVID-19 lockdown on air quality, this report investigates years of air pollution data from Three hundred and forty-one Indian cities. The results show that temporary lockdown policies greatly reduced pollutants, so improving the air quality.

With Random Forest regression surpassing all others, this work forecasts Air Quality Index (AQI) values using machine learning models. Three pollution levels—SO2, NO2, and PM10—as further classifications for the AQI values help to explain variations in air quality throughout different areas. This report underlines the need of long-term actions to maintain the air cleaner as well as the need of data-driven approaches to grasp trends in air pollution. Findings show that long-term solutions are needed to preserve better air quality even if temporary lockdowns greatly lowered pollution. This report underlines the need of data-driven approaches in tackling air pollution and stresses the part of technology in forming next environmental policies.

2. Problem Statement

Air pollution is becoming an issue of high concern in the recent years. Especially in India which is a developing country there is significant increase in the amount of air pollutants and thereby the Aqi also increases. As economic and industrial centers, cities suffer disproportionately high degrees of air pollution, daily exposing millions of people to dangerous airborne toxins. Extended exposure to contaminated air has adverse effects that lead to respiratory diseases including asthma and chronic obstructive pulmonary disease (COPD), cardiovascular problems, and other systemic health conditions.

There are primarily three pollutants which greatly affect the air quality namely: sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and particulate matter (PM10). SO_2 is mainly released from burning of fuels like coal and oil. It can cause severe respiratory problems and lung diseases which are harmful. NO2 is mainly released from vehicles and is a serious pollutant. It also gives rise to many other pollutants

such as ozone. PM10 is a very hazardous pollutant as it is directly inhalable and can also the lungs causing severe diseases and can also cause issues related to heart health which is a major concern.

This causes many long-term health related issues which remain unaddressed in many densely populated cities of India which is a matter of great concern. To provide solution to this problem firstly analyze the relationship between pollutants and their effect on air quality. There has to be an alarm for all the environment protection organizations and legislatives to give importance to this issue and work upon it.

3. Introduction

Air pollution has become a major issue globally. It slowly destroys the environment and also leads to severe health related issues. India which is a developing country is facing increasing air pollution and bad air quality due to various forces such as overpopulation, industrialization and vehicle emission. Air pollution has severe damages in mainly every developing country. There are cases of over 1800 deaths per day due to respiratory problems caused by air pollution.

The majority of the air pollution is caused by vehicle emission and burning of fossil fuels. Vehicles emit harmful gases, such as NO_2, SO_2, and PM10, which get easily mixed in the air causing degrading impact on human health. These pollutants also have great impact on monuments and agriculture leading to bad food quality which in turns leads to losses. People who already have respiratory problems such as old aged people and young children are the ones who are affected by these the most.

Due to the devastating effects of air pollution effective measures have to be taken and policies have to be formed to control this problem. However, existing steps and policies have failed to address the impact of long-term air pollution especially in densely populated areas with huge traffic jams.

The goal is to analyse the impact of air pollution on the basis of various metrics mainly focusing on respiratory and heart problems. Further this study aims to analyse and fill the gaps using machine learning methods. The goal is to develop and implement a predictive model which can accurately predict the AQI from past data. This report uses insights from various past literature sources which have been a source of guidance for the prediction model of the AQI.

4. Literature Review

S. No	References	Link	Description
1	**Pope et al. (2019)**	https://www.nejm.org/doi/full/10.1056/NEJMsa0805646	Linear regression, Correlation, Random Forest Dataset is from WHO Air Quality Database, Variables are pollutant levels (PM2.5, PM10)
2	**Brok et (2020)**	https://ieeexplore.ieee.org/document/9316 128	Support Vector Machines, Time-series analysis, Dataset is from regional air quality and public health statistics variables are Long-term pollutant expo- sure (NOx, O_3).
3	**Kalyan Khatri (2023)**	https://www.researchgate.net/publication/372406390	Support Vector machines, Random Forest regression Dataset is from The Central Pollution Control Board variables are PM2.5, NO, NH3, Xylene, and AQI.
4	**Ashima Tyagi (2020)**	https://www.researchgate.net/publication/349802100_	Linear regression, Correlation, Random Forest. Dataset is Open AQ from Kaggle, variables are pollution, ground-level ozone, carbon monoxide, Sulphur oxides, nitrogen oxides.

5	**Sarita Jiyal (2020)**	https://ieeexplore.ieee.org/document/9315831	Air Pollution Prediction, Metrics: industrialization, overpopulation, traffic control, electric vehicles, and pollution assessment.
6	**Alka Pant (2023)**	https://www.semanticscholar.org/reader/d37144492386b4b2b726fc78553e702fd131db1f	ARMA, Time Series Analysis, Air Quality Index (AQI) Prediction. Dataset is from five years of monthly AQI data from Dehradun, Uttarakhand.
7	**Chenchen Li (2021)**	https://ieeexplore.ieee.org/document/9708644	Random Forest Regression (RFR), Gradient Boosting Regression (GBR), Dataset has PM2.5, PM10, SO2, NO2, CO, and O3 as primary pollutants, with AQI predicted variable.
8	**Humaib Nasir(2016)**	https://www.researchgate.net/publication/311499664_Review_of_Air_Quality_Monitoring_Case_Study_of_India	National Air Quality Index (NAQI), Air Pollution Analysis, Air Quality Monitoring (AQM). Dataset is from past NAQI from various Indian cities. It analyses pollutants PM2.5, SO2, NO2 PM10, and other harmful gases.
9	**GK Kang (2018)**	https://ieeexplore.ieee.org/document/9315831.	Analyzes air pollution trends using machine learning and big data techniques. The dataset is from historical air quality data. It contains pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3.

10	**T Madan (2020)**	https://ieeexplore.ieee.org/abstract/document/9362912	Predicting air quality and analyzing its impact on mainly lung and heart diseases. Dataset has Air Quality Index (AQI) levels, and pollutants like CO_2, NO_2, and CO from industrial wastage, vehicles, and fuel burning.
11	**J Wang (2021)**	https://link.springer.com/article/10.1007/s00521-020-05535-w	CT-LSTM model for AQI prediction, combining a chi-square test with an LSTM network. It is trained on historical air quality data. It outperforms SVR, MLP, BP neural networks and Simple RNN while achieving accuracy of 93.7% in AQI prediction.
12	**W Wang (2021)**	https://www.sciencedirect.com/science/article/abs/pii/S221067072030785X	TS-LSTME model for 24/7 AQI prediction, gathering correlations and pollutants data. It is applied in China's Jing-Jin-Ji region. It outperforms MLR, SVR, LSTM, and LSTME models, achieving high accuracy
13	**F Ramos (2020)**	https://www.mdpi.com/2076-3417/10/7/2401	This reviews 41 papers on air pollution prediction using machine learning in smart cities. The results show PM2.5 as the primary pollutant, and 24/7 predictions. Integrating

			data improves accuracy, with open data use increasing.
14	**P Bhalgat (2019)**	https://www.researchgate.net/profile/Sachin-Bhoite/publication/335911816	This study reviews air pollution prediction using machine learning in smart cities. Advanced models are replacing simpler ones, with PM2.5 as the primary focus and 24/7 predictions and forecasts most common.
15	**D Seng (2021)**	https://www.sciencedirect.com/science/article/pii/S1110016820306438	LSTM- model for air quality prediction in Beijing. The datasets contain pollutants such as PM2.5, PM10, SO2, NO2, O3 and other secondary pollutants.

5. Research Gap

- The research is mainly conducted on urban cities while the pollution in the rural areas is immense due to crop and plastic burning which is largely ignored.
- There is a lack of knowledge and research on the changing climate factors which also play a huge role in air pollution.
- There is still not enough research on the long-term impact of air pollution policies relating to vehicle and industrial pollution which is a matter of concern.
- Air pollution factors for every city are different from other city so the generalized approach to figure out pollutants and solving their issues remain unorganized.
- The use of eco-friendly material electrical vehicles and the deforestation is not implemented to a standard that will make a major change. To make a major change strict guidelines must be implemented.

6. Research objectives

- Air pollution levels in Indian cities are rising as India is one of the fastest developing countries in the world. Through analysis there is a finding that SO2, NO2, PM10, O3, PM2.5 are the major pollutants that are present in the air which makes the air harmful. There are various sources of this pollution like vehicle emission, crop burning, waste led into the environment by industries. Pollution levels are different in different seasons in winter there is our serious smog cases in major cities while in summer there is spike in ozone these are harmful for humans.
- To predict AQI based on the past data gathered by the Central Pollution Control Board using random forest regression. With the analysis and impact of the major pollutants that contribute to air pollution. This model can tell us about the future that what would be the air conditions. There has also been a comparison done among models and the most reliable model have been chosen after research. This prediction can help in finding the best possible way to reduce air pollution and formulate policies for its control.
- There is an urgent need for implementing effective monitoring of the air quality. There need to be real time tracking networks for proper tracking. Checking the quality and reliability of the current monitoring networks will help in identifying the gaps that are present in the current system. There are severe health diseases from long term air pollution mainly related to respiratory problems. Analyzing these impacts of air pollution is crucial as and only then we will be able to provide solutions for it. Implementing strict policies and new technological solution such as electric vehicles can be a great initiative to combat this serious problem of air pollution.
- Combating air pollution requires strict legislative decisions made by the government. There also should be a severe fine if the public does not abide by these rules and still burns crops or cause pollution. There should be proper utilization of the monitoring and tracking systems to seek the place where there is pollution and take measures for it. They should encourage the use of eco-friendly products, afforestation and the use of electric vehicles so there would be lesser pollution.

7. Research Methodology

The first step of the project was to go through various literature reviews which are mentioned in the report. The literature review is very important step to as to gather existing information about the project and see its working in the real world.

After the reading of the literature review comes the process of data collection. This dataset was taken from the Central Pollution Control Board of India website containing the pollutants such as SO2, NO2, PM10 and AQI of three hundred and forty-one cities of India

After the collection of data the it is then set to predict AQI using the random forest regression model after ample research. This prediction tells us that how important is to predict the pollution of the future so that their appropriate combat measures can be set-up. There can be formation of new rules and policies based on the findings of this prediction model which will then help to reduce pollution. This prediction model is the first step of the path which can be developed into a huge model of monitoring and providing measures of the problem.

8.Dataset Studied

The dataset used in the model is collected from Central Pollution Control Board of India website. This dataset talks about the quantity of the various pollutants that are present in air of various cities.

There are six columns each representing some data.

- State / Union Territory: - This informs about the state of the city.
- City: - This shows the city which data is taken out.
- SO2: - This shows the amount of SO2 present in the air in that particular city.
- NO2: - This shows the amount of NO2 present in the air in that particular city
- PM10: - This shows the amount of PM10 present in the air in that particular city.
- AQI: - This shows the Air Quality Index of the city.

The dataset tells us about the air quality of three hundred and forty-one cities all across India by determining the quantity of the pollutant in that respective city. It has pollutants such as sulphur dioxide (SO_2), nitrogen dioxide (NO_2), and particulate matter (PM_{10}), Air Quality Index (AQI), which tells us about the air quality of that particular city. By using this dataset, we are easily analysing and compare the air quality of different cities located at different geographical locations and lay importance on where the pollution is the most.

This dataset lays the framework on which the legislative agencies and environmentalists and policy makers study upon and try to find the solutions for these problems that are constantly arising in cities. They analyse the trends in which the pollution data is following and sought to find the reason for such trend which may be a geographical reason or seasonal reason. Through this dataset there is an aim to identify the diseases associated with these pollutants and how to reduce them mainly breathing related diseases. It can also help in formulating of policies relating to the pollutants like we get to know that PM10 is high in a certain city which means that there is a major traffic congestion issue and thus more vehicle emissions.

This dataset is of utmost use and importance for the those who want to develop a prediction model based on this dataset. This dataset contains past data through which anyone can create a predictive model based of it for the future so that they can analyse and control air pollution and assist in identifying the issues at hand and provide appropriate solution for them. The dataset also helps in making decisions about the transformation sector and how to control pollution in that sector and also identify the areas in which industries illegally emit harmful hazardous gases into the atmosphere. Thus this dataset will help in identifying all the major pollutants and their area so that the pollution can be restricted.

Dataset Link: (Click Here)

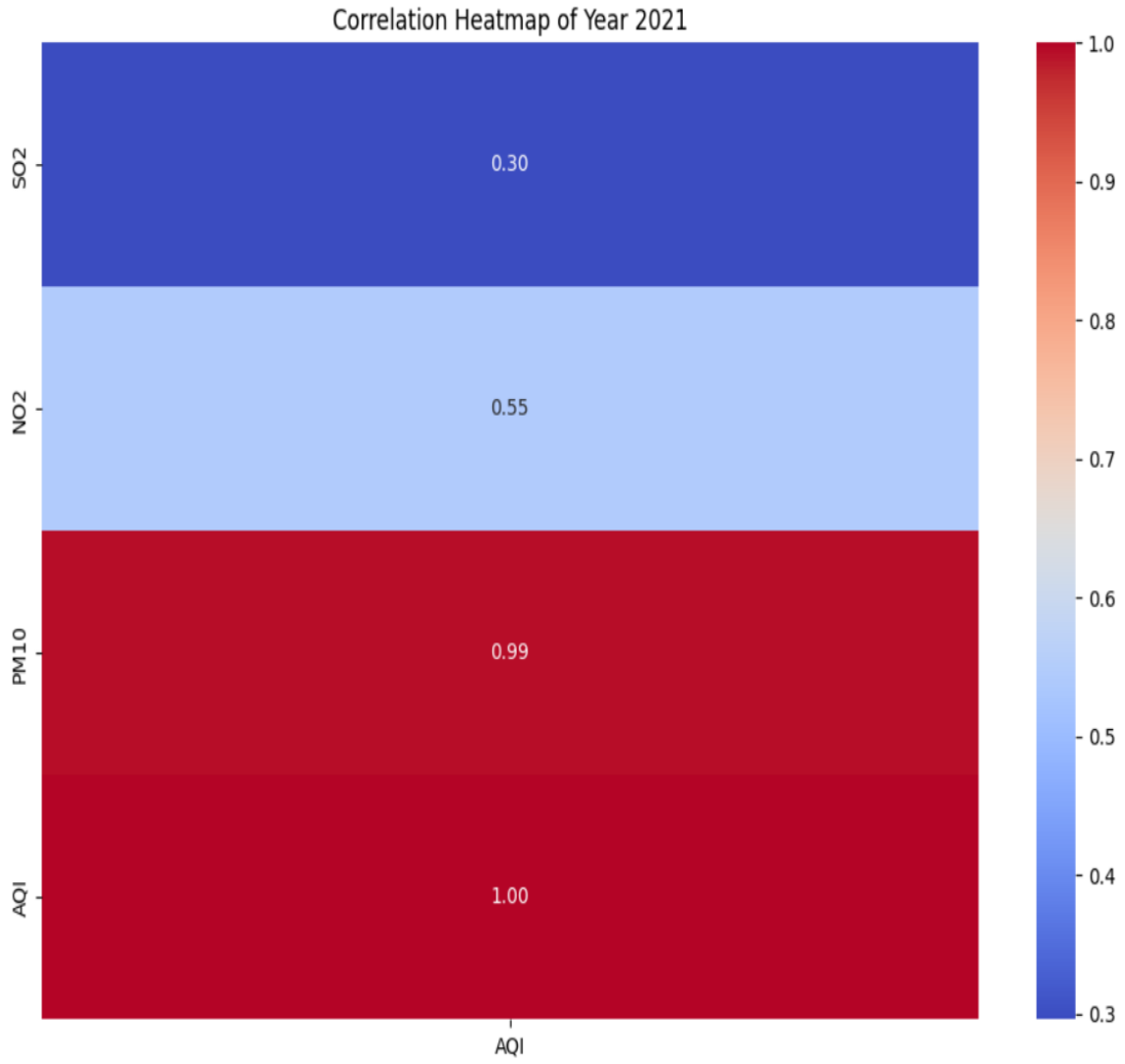

Fig 1: Correlation Heatmap of Year 2021

The above is the correlation heatmap for the year 2021 which tells us about the relation between air pollutants and the Air Quality Index (AQI). The highest correlation is between PM10 and AQI which is (0.99) this indicate that PM10 have the major impact on the air quality index. NO2 has a correlation of (0.55) which is moderate in its comparison with PM10 which tells us that it has lesser impact on AQI than PM10. SO2 has the lowest correlation of (0.30) which signifies that it has the least amount of impact on air pollution in comparison with the other two. The color scheme of the heatmap on the right emphasizes that dark red color means a strong correlation while blue and dark blue shows weak correlation. This shows that PM10 plays the major role in the contribution of air pollution and thus its emission needs to be controlled.

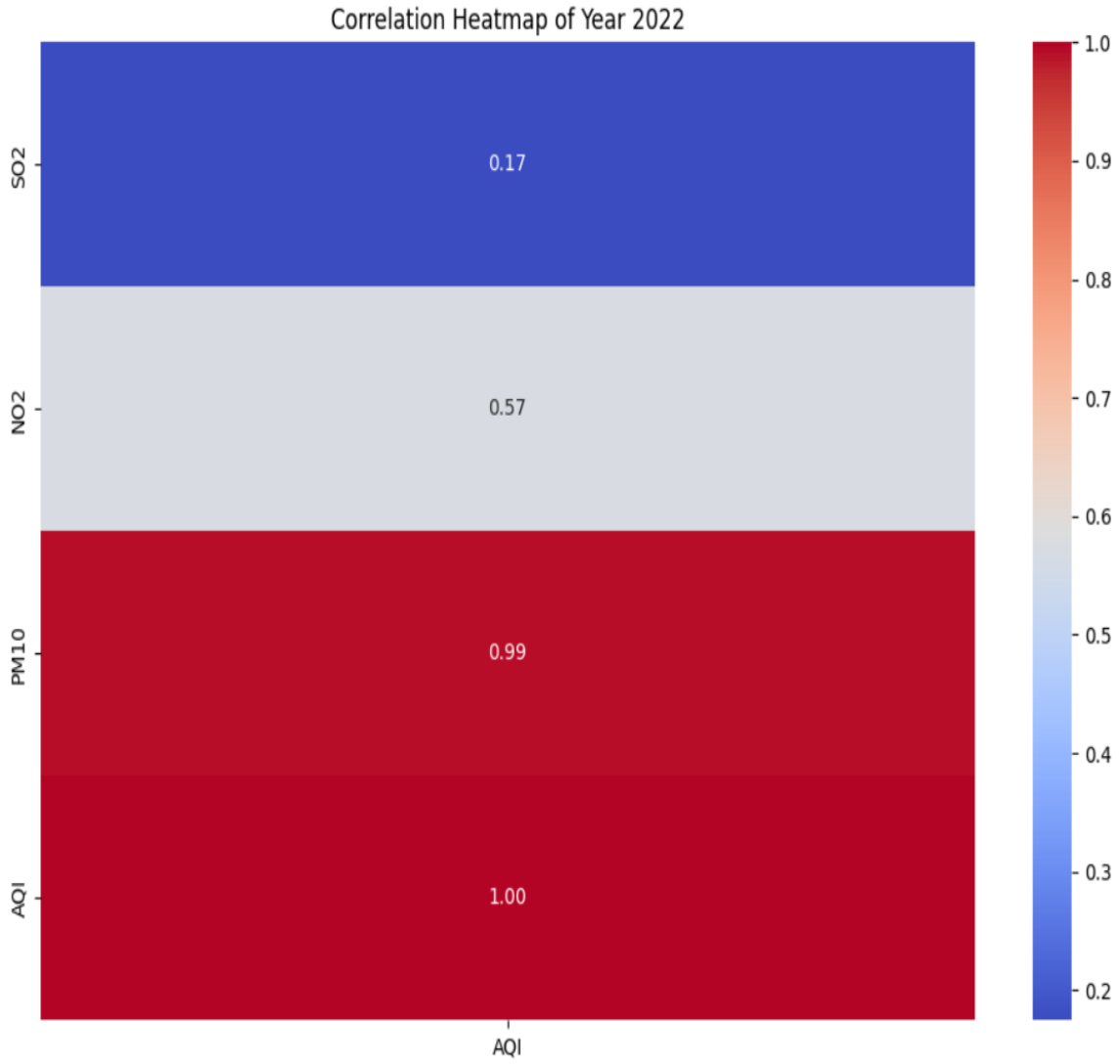

Fig 2: Correlation Heatmap of Year 2022

The above is the correlation heatmap for the year 2022 which tells us about the relation between air pollutants and the Air Quality Index (AQI). The highest correlation is between PM10 and AQI which is (0.99) this indicate that PM10 have the major impact on the air quality index. NO2 has a correlation of (0.57) which is moderate in its comparison with PM10 which tells us that it has lesser impact on AQI than PM10. SO2 has the lowest correlation of (0.17) which signifies that it has the least amount of impact on air pollution in comparison with the other two. The color scheme of the heatmap emphasizes that dark red color means a strong correlation while blue and dark blue shows weak correlation. This shows that PM10 plays the major role in the contribution of air pollution and thus its emission needs to be controlled.

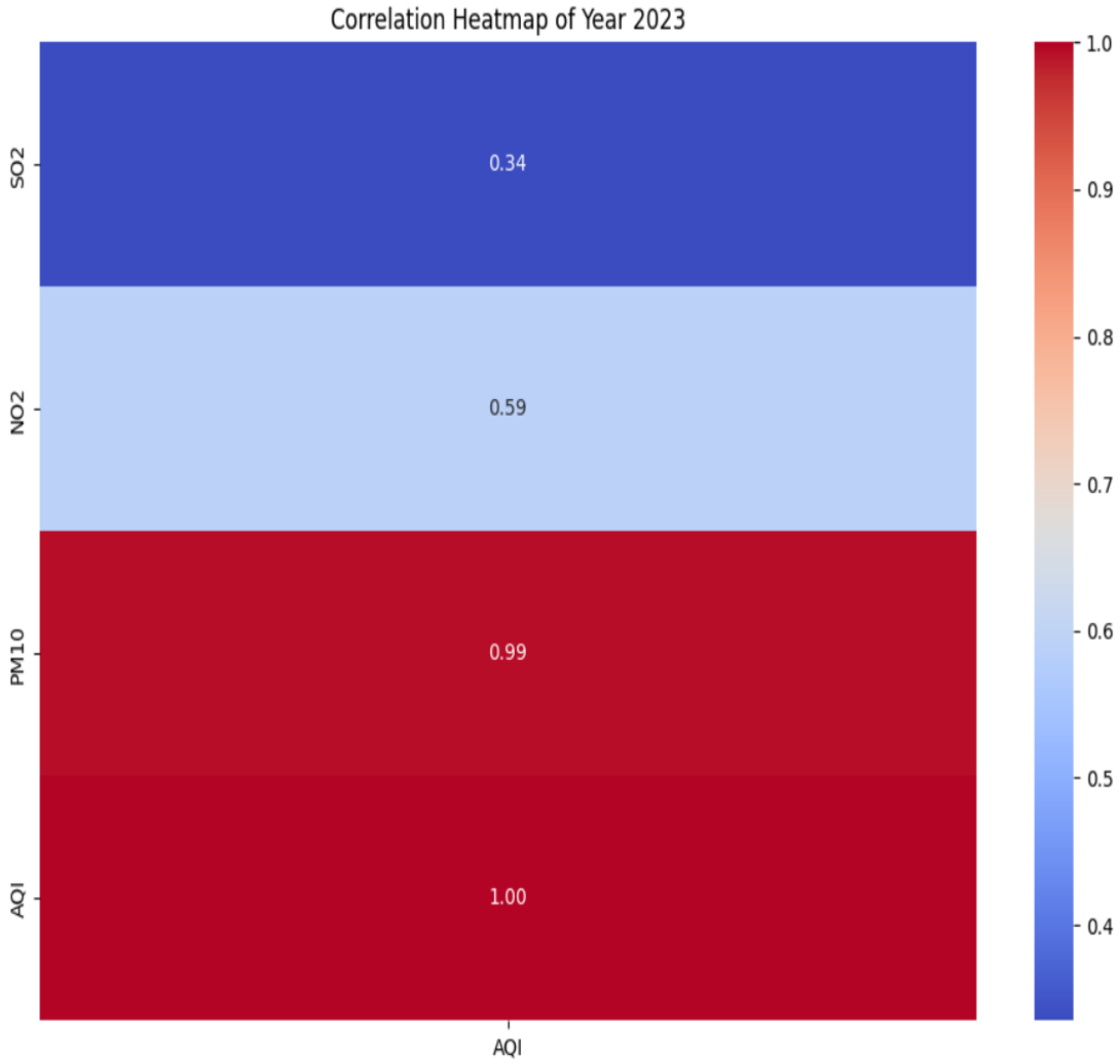

Fig 3: Correlation Heatmap of the Year 2023

The above is the correlation heatmap for the year 2023 which tells us about the relation between air pollutants and the Air Quality Index (AQI). The highest correlation is between PM10 and AQI which is (0.99) this indicate that PM10 have the major impact on the air quality index. NO2 has a correlation of (0.59) which is moderate in its comparison with PM10 which tells us that it has lesser impact on AQI than PM10. SO2 has the lowest correlation of (0.34) which signifies that it has the least amount of impact on air pollution in comparison with the other two. The color scheme of the heatmap emphasizes that dark red color means a strong correlation while blue and dark blue shows weak correlation. This shows that PM10 plays the major role in the contribution of air pollution and thus its emission needs to be controlled.

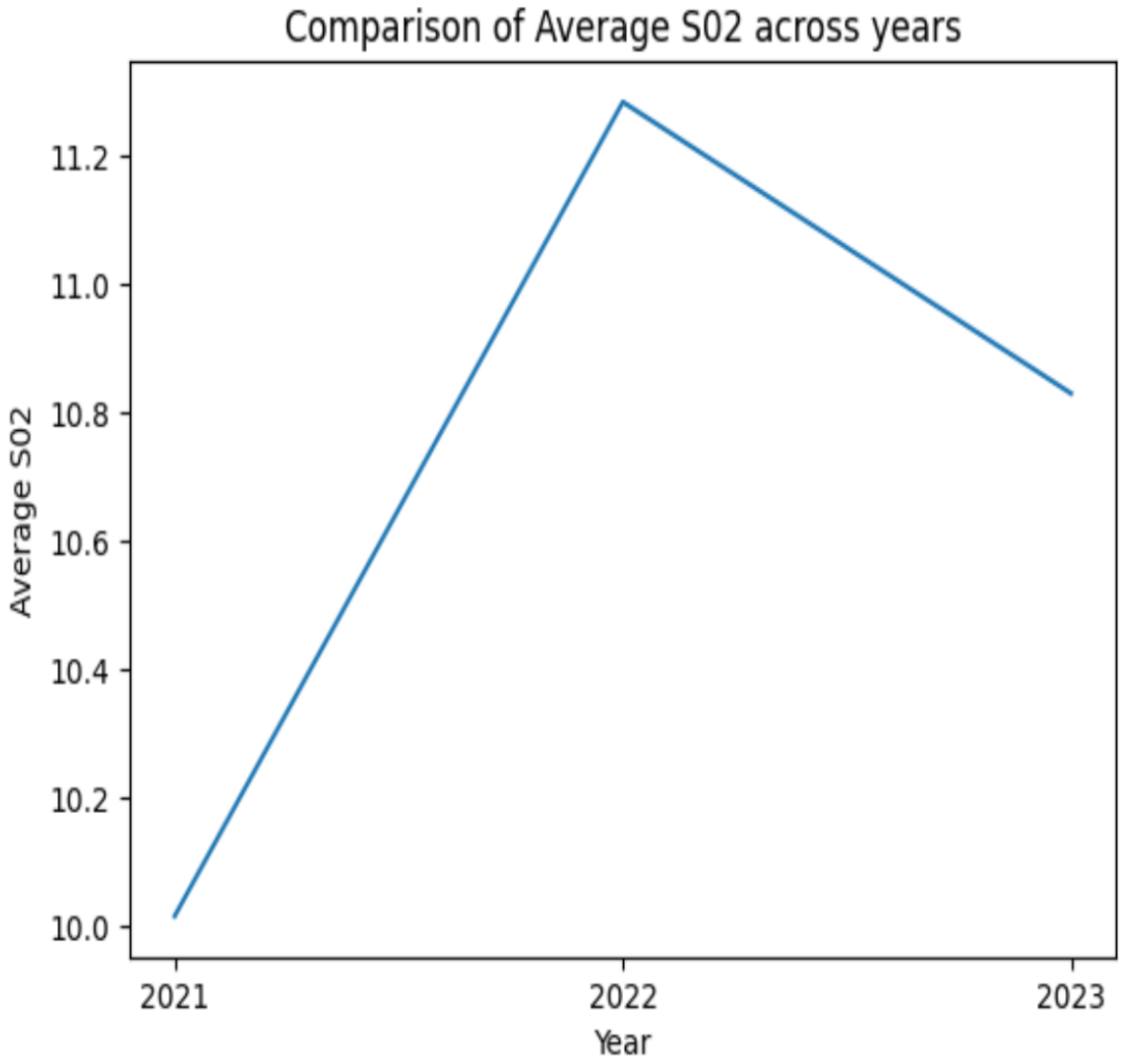

Fig 4: Comparison of SO2 across Years

The above graph shows the comparison of the pollutant SO2 through 2021-2023. The x-axis of this graph shows the years while the y-axis of this graph shows the average SO2 present in the air. The graph shows that the average SO2 levels increased from the year 2021 to 2022 at which it reached the maximum. After 2022 the levels declined in 2023 while still remaining higher than the year 2021. This tells us that the peak was in 2022 after which there was a sudden decline which tells us that there could be various factors for this decline such as environmental factors, and various other factors.

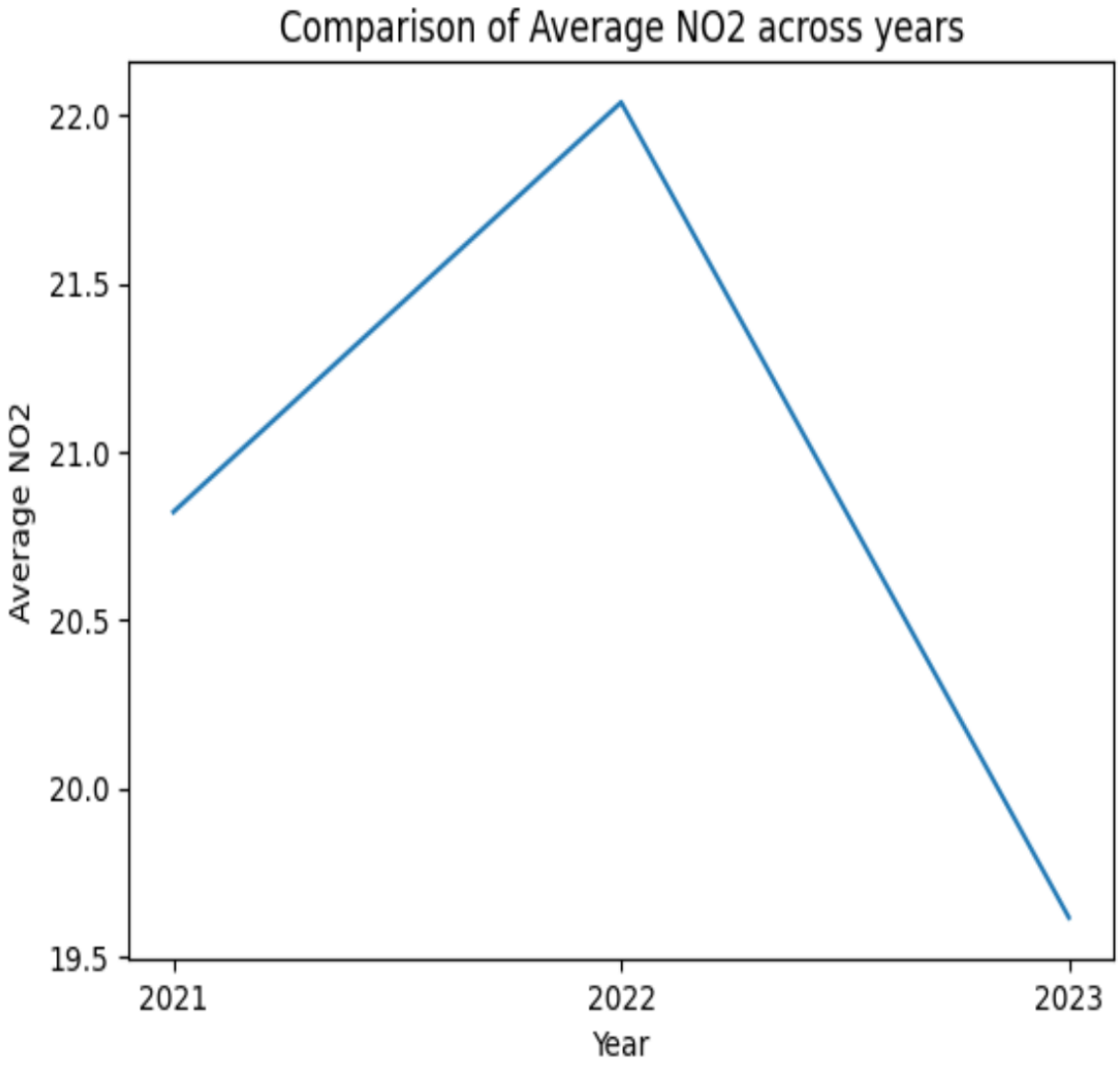

Fig 5: Comparison of NO2 across Years

The above graph shows the comparison of the pollutant NO2 through 2021-2023. The x-axis of this graph shows the years while the y-axis of this graph shows the average NO2 present in the air. The graph shows that the average NO2 levels increased from the year 2021 to 2022 at which it reached the maximum. After 2022 the levels declined in 2023 lower than the year 2021. This tells us that the peak was in 2022 to its lowest in 2023 after which there was a sudden decline which tells us that there could be various factors for this decline such as environmental factors, and various other factors.

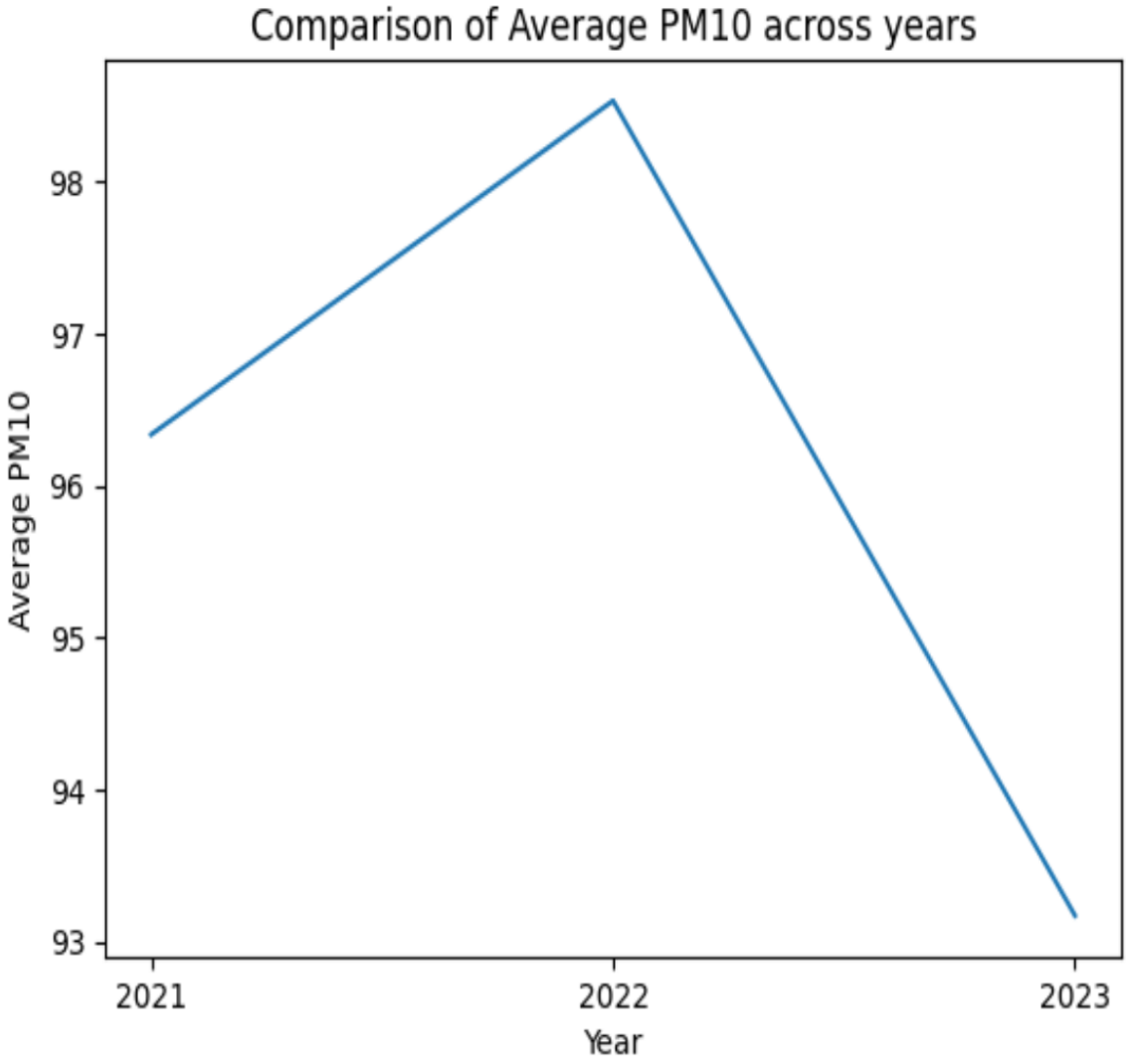

Fig 6: Comparison of PM10 across Years

The above graph shows the comparison of the pollutant PM10 through 2021-2023. The x-axis of this graph shows the years while the y-axis of this graph shows the average PM10 present in the air. The graph shows that the average PM10 levels increased from the year 2021 to 2022 at which it reached the maximum. After 2022 the levels declined in 2023 reaching lower than the year 2021. This tells us that the peak was in 2022 after which there was a sudden decline which tells us that there could be various factors for this decline such as environmental factors, vehicle emissions and various other factors.

Air Quality Index		
AQI Category and Color	**Index Value**	**Description of Air Quality**
Good Green	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Moderate Yellow	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups Orange	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Unhealthy Red	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy Purple	201 to 300	Health alert: The risk of health effects is increased for everyone.
Hazardous Maroon	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

Fig 7: AQI Table

The above image is of a AQI table which is divided into different categories of different colour with an index value starting from Good (0-50) Colour green to Hazardous (301-Higher) Colour maroon.

9.Impact On Health

Air pollution causes a huge amount of damage to human lives. It affects huge amount of people all over the world. The exposure to hazardous pollutants like sulfur dioxide (SO_2), nitrogen dioxide (NO_2) and particulate matter (PM10) can lead to breathing issues, heart problems and it can even lead to the weakness of immune systems. It also has short-term effects such irritation in the eyes and breathing problems for people who have asthma.

- **SO2:** Sulfur dioxide (SO_2) is a hazardous air pollutant mainly produced by industries and vehicle emissions. Short-term effects can cause eye irritation, choking of respiratory tract and shortness of breath. People already having asthma and respiratory problems are mainly on the radar to get affected by this pollutant. Long-term effects can lead to severe lung diseases, immense respiratory problems and also leads to risks of cardiovascular issues. SO2 is also a primary reason of Acid rain which degrades monuments and environment.

- **NO2:** Nitrogen dioxide (NO_2) is also a major pollutant that is emitted for vehicular emissions and burning of fuels such as petrol and diesel. It is the pollutant which causes smog in urban cities. Short term effects are such as inflammation, lung infection and coughing. Long-term effects play a major role in developing asthma and heart diseases in people. It can further degrade air quality by forming ground -level ozone which is very much hazardous to health.

- **PM10:** PM10 refers to the particulate matter which have diameter of 10mm or less and which can easily penetrate the respiratory tract of the human being. They are formed by burning of industrial waste, vehicle emissions and wildfires. Short-term effects are throat irritation, asthma and bronchitis. Long-term effects cause an increased risk of lung failure and it can also lead to heart attacks and strokes which are very dangerous for humans. This particulate matter also carries hazardous things which if enter the bloodstream can lead to damage of various organs.

- **AQI:** The Air Quality Index (AQI) is a universal measure of calculating the quality of air and the amount of threat each level poses. It is calculated by various pollutants such as SO2, NO2, PM10, PM2.5 and ozone. The higher the Aqi the worse the air quality of that place is and the more dangerous the threats posed by that air. High AQI levels leads to various problems such as inflammation, breathing difficulty, lung infection and many more. Long-term effects lead to asthma, bronchitis, severe cardiovascular diseases, lung failure and in severe cases it can also lead to heath attacks causing premature death of that person. People sensitive to it such as children, elder people must protect themselves from these pollutants as they are at the greater risk and will be the first ones to get affected.

10.Results

A detailed analysis of air pollution data was taken place using various methods like data visualizations, correlation heatmaps, line plots and AQI tables. These visual images provided crucial information about the data and the relationship of data such as the effects of pollutants such as SO2, NO2, PM10 on AQI.

One of the most significant visualizations was of the correlation heatmaps of different years. That shows that the correlation between AQI and PM10 is the greatest (0.99) in comparison with other pollutants. This shows that PM10 is the most significant factor while calculating AQI and it has the most impact. Next is the line chart depicting the annual comparison of the pollutants over the years. In which the peak was observed in the year 2022 while having a decline in the year 2023. There could be various factors for this decline such as environmental factors, and various other factors.

The above image is of a AQI table which is divided into different categories of different colour with an index value starting from Good (0-50) Colour green to Hazardous (301-Higher) Colour maroon. This table is essential to spread knowledge about air pollution and determine the risk associated with the level of the index. It is especially for sensitive people like children and elderly people or people having breathing issues.

The R-Squared (R^2) value of 0.9913 showed an exceptionally strong correlation, indicating that 99.13% of the variance in AQI can be determined by these pollutants. The results show that pollutants like SO2, NO2, PM10 play a crucial role in contribution to air pollution and must be controlled. There was strong correlation between PM10 and AQI which signifies that it is the most hazardous pollutant that contributes to air pollution controlling and reducing it can help in making the AQI levels low and good air quality. SO2 and NO2 have moderate correlation with AQI which signifies that it has affect on the AQI but this effect is less compared to PM10.

The peak in 2022 in pollution levels could be due to multiple factors such being the first year after the Corona virus lockdown. So many people got back to their work after very long and there would be sudden rise in the traffic levels and the emissions drawn out of them. The decline in 2023 may be due to various reasons such as effective

pollution control, environmental changes, favourable weather changes, rainfall and there could be many other reasons.

After through research on who is the best model random forest regression model emerged as the most accurate model in comparison to other models. This ensures that random forest models' ability to handle complex datasets and give prediction upon them smoothly. Linear Regression model did not perform well and struggled to operate the dataset leading to low accuracy and bad results. Support Vector Machine (SVM) displayed a good performance but it required additional information to carry out predictions more effectively. Neural Networks worked well but has a huge challenge of immense data preprocessing and huge number of resources required to carry out operations. Random forest proved to be the best model among these other models with results showing great accuracy and high performance. In the future there can be use of deep machine learning models to ensure more better accuracy and long-term predictions. The results show that there is a urgent need to control PM10 emissions as it is the most decisive factor in maintaining air quality and also have the most impact on AQI. Efforts should also be made to reduce sources of SO2 and NO2 as they also have an effect on the air quality but the primary focus should be on the reduction of PM10 by reducing vehicular emissions and developing strict policies which everyone have to follow.

It is found out that in India the pollution in cities is mainly due to vehicle emissions and industrial waste. Random forest regression model ensured high accuracy in predicting AQI of the cities which can help in the effective policy formulation and monitoring of the air quality. The current monitoring systems have huge gaps which can be resolved by deploying real time-based sensors. Analysis of health impact demonstrated that there is strong link between the respiratory and cardiovascular diseases in India relating to bad air quality. Strict rules and policies formed by the government on vehicle norms, waste management, illegal burning of crops must be made only then the air quality would be able to come under control. There should also be encouragement of the usage of eco-friendly products and in the future, there can be many technological solutions which will help in combating this huge issue of air pollution at hand.

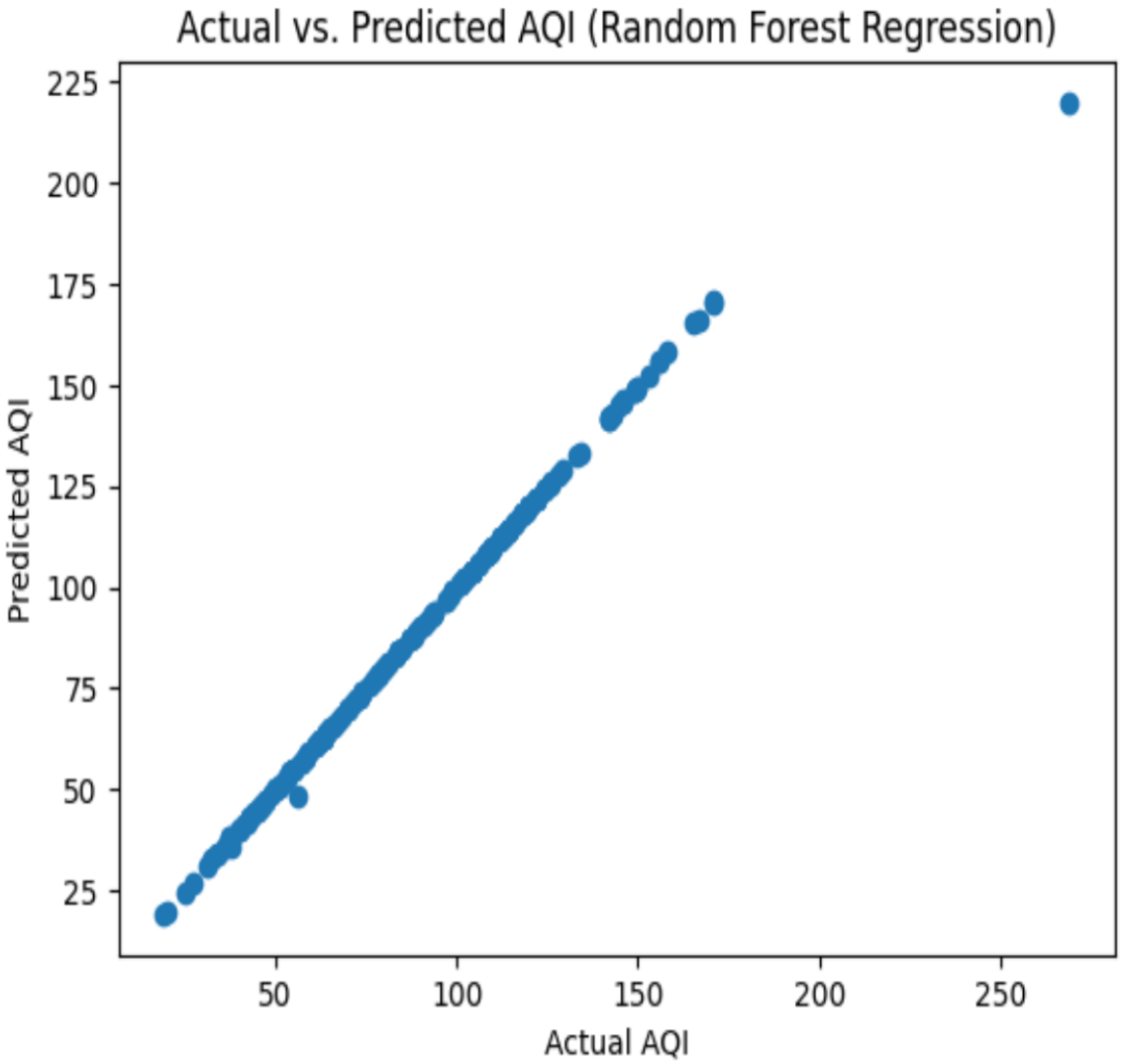

Fig 8: Prediction of AQI

The above picture displays a scatter plot between the predicted value of AQI and the actual value of AQI using random forest regression model. The points in the plot are in a diagonal line which indicate that there is very little error and the predictions are mostly accurate. The plot displays the accuracy of the model by comparing the actual and predicted AQI. There is a strong alignment on the diagonal line which ensures that the model is well trained to predict the data with accuracy. The model has learned the steps to carry out AQI prediction well and performs well in most of the cases. But there may be some outliers or some extreme places in which the points are very different from the actual ones.

10.Literature Comparison

Comparing accuracy of the model with few papers is very important step and the comparison is listed below

Paper	Author	Accuracy	My accuracy (Test)
Fine-particulate air pollution and life expectancy	**Pope et al**	88%	99.13%
Implementing Big Data Analytics for Predicting City-Level Air Quality Index (AQI)	**Kalyan Khatri**	99.75%	99.13%
Scrutinizing Patterns of Air Pollution in India. 5-9.	**Ashima Tyagi**	95.34%	99.13%
Review of Air Quality Monitoring	**Humaib Nasir**	92.9%	99.13%
Research on Air Quality Prediction Based on Machine Learning	**Chenchen Li**	98.73%	99.13%

11.Limitations of the Study

While we developed a prediction model but there are certain limitations that are not under the control and must be looked upon.

- Ignorance of External Factors: Air quality is greatly affected by external factors such as environmental factors, irregular weather conditions, unknown pollutant emissions and many more factors. The model is based on past air quality data but there must be some cases of unpredictable events for which the predicted values won't match to the actual values.
- Quality of the data: The accuracy of any models mainly depends on the quality and accuracy of its dataset and its completeness. If the data have missing or incorrect values the prediction is not accurate.
- Generalisation of the prediction model: Random Forest regression model performed good in the prediction but it can often generalize a number of things for different cities which remains out of hand.
- Limitations of power: Random Forest is model which is highly powerful in nature but can cause problems while dealing with huge datasets thus requiring a huge amount of system power though which the model can process and the prediction can be taken place.

12.Conclusion

This report tells us about the urgent need to combat air pollution as it has a great amount of impact on our everyday lives which in long-term can cause several severe respiratory diseases. The ongoing development of the country, overpopulation and the emission from vehicles leads to bad air quality. These causes many problems namely heart and lung related health hazards. Major cities in India like Delhi, Kolkata, Mumbai are in high alert from pollutants such as NO2, PM10, and SO2, making rise for many diseases and can also lead to premature deaths. There is also cases of severe smog on roads leading to accidents.

With this model there is a finding that PM10 is a major pollutant which contributes to air pollution and high air quality index. By reduction of PM10 from the air we can make a great step in reducing air pollution. This can be done by adoption of electric

vehicles and quality checks of vehicles regularly. The findings show that there are these key pollutants such as Sulfur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Particulate Matter (PM10) and other harmful gases across different cities, which are leading to the rise of the pollution in the air. The prediction model states that there could be a further rise of pollution in some cities and the situation will become out of hand if adequate pollution control measures and policies are not formed.

The findings of this report tell us that there is an urgent need to control this large issue of air pollution because the situation will only become worse in the future. So, to prevent it from becoming disastrous there need to be formulation of strict policies and initiatives taken up by government such as plantation of trees making the country greener. The public should also become more careful and should think about protecting their environment from air pollution.

13.References

1. Pope, C. A., Ezzati, M., & Dockery, D. W. (2019). Fine-particulate air pollution and life expectancy. *New England Journal of Medi- cine*, 360(4), 376-386.Retrieved from: https://www.nejm.org/doi/full/10.1056/NEJMsa0805646
2. Brook, R. D., Rajagopalan, S., & Pope, C. A. (2020). Particulate matter air pollution and cardiovascular disease*: An update to the scientific statement from the American Heart Association.* Circulation, 121(21), 2331-2378.Retrieved from: https://www.ahajournals.org/doi/10.1161/CIR.0b013e3181dbece1
3. Khatry, Kalyan & Adhikari, Prajwal. (2023). *Implementing Big Data Analytics for Predicting City-Level Air Quality Index (AQI) and Examining the Impact of COVID-19 using Comprehensive PM and Gas Emission Data.* Retrieved from: https://www.researchgate.net/publication/372406390
4. Tyagi, Ashima & Kharb, Latika & Chahal, Deepak. (2020). *Scruti- nizing Patterns of Air Pollution in India*. 5-9. 10.1109/ICACCCN51052.2020.9362990. https://www.researchgate.net/publication/349802100_
5. Jiyal, Sarita**.** (2020). *Air Quality Prediction using Machine Learning Techniques.* IEEE Xplore. DOI: 10.1109/ICCCNT49239.2020.9315831.

6. Pant, Alka. (2023). *Air Pollution Prediction and Analysis Using Machine Learning Techniques.* Semantic Scholar. Available at: https://www.semanticscholar.org/reader/d37144492386b4b2b726fc78553e702fd131db1f.
7. Li, Chenchen. (2021). *Air Quality Prediction Using Machine Learning Approaches.*IEEE,Xplore. Available at: https://ieeexplore.ieee.org/document/9708644.
8. Nasir, Humaib. (2016). *Review of Air Quality Monitoring: Case Study of India.* ResearchGate. Available at: https://www.researchgate.net/publication/311499664_Review_of_Air_Quality_Monitoring_Case_Study_of_India.
9. GK Kang, Gaganjot Kaur., Gao, Jerry Zeyu., Chiao, Sen., Lu, Shengqiang., & Xie, Gang. (2017). *Air Quality Prediction: Big Data and Machine Learning Approaches.* San Jose State University & Taiyuan University of Technology. Available at: https://ieeexplore.ieee.org/document/9315831.
10. T Madan. (2020). *Predicting Air Quality Using Machine Learning Algorithms.* IEEE Xplore. Available at: https://ieeexplore.ieee.org/abstract/document/9362912.
11. Wang, Jing yang., Li, Jia zheng., Wang, Xiao xiao., Wang, Jue., & Huang, Min. (2020). *Air Quality Prediction Using CT-LSTM.* Neural Computing and Applications,33,4779–4792. Available at: https://link.springer.com/article/10.1007/s00521-020-05354-x.
12. Wang, Mao, Wenjing., Weilin., Jiao, Limin., Zhao, Suli., & Liu, Anbao. (2020). *Modeling Air Quality Prediction Using a Deep Learning Approach: Method Optimization and Evaluation.* Sustainable Cities and Society. Available at: https://doi.org/10.1016/j.scs.2020.102567.
13. Iskandaryan, Ditsuhi., Ramos, Francisco., & Trilles, Sergio. (2020). *Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review.* Applied Sciences, 10(7), 2401. Available at: https://doi.org/10.3390/app10072401.
14. Bhalgat, Pooja., Bhoite, Sachin Bhimrao., & Pitare, Sejal. (2019). *Air Quality Prediction using Machine Learning Algorithms.* International Journal of Computer Applications Technology and Research, 8(9). Available at:

https://doi.org/10.7753/IJCATR0809.1006.

15. Seng, Dewen., Zhang, Qiyan., Zhang, Xuefeng., Chen, Guangsen., & Chen, Xiyuan. (2020). *Spatiotemporal Prediction of Air Quality Based on LSTM Neural Network.* Alexandria Engineering Journal. Available at: https://doi.org/10.1016/j.aej.2020.12.009.