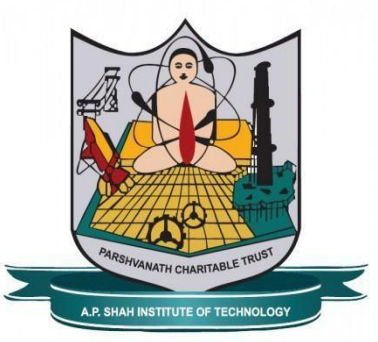
A
Mini-Project Report on

AIR QUALITY INDEX PREDICTION FOR INDIAN CITIES USING MACHINE LEARNING TECHNIQUES

Submitted in partial fulfillment of the requirements
for the degree of
BACHELOR OF ENGINEERING
IN
Computer Science & Engineering
(Artificial Intelligence & Machine Learning)

by
Nishant Hire (21106060)
Shipra Asthana (21106039)
Maviya Bubere (21106022)
Ayush Kargutkar (21106042)

Under the guidance of
Prof. Vijaya Bharathi J



Department of Computer Science & Engineering
(Artificial Intelligence & Machine Learning)
A. P. Shah Institute of Technology
G. B. Road, Kasarvadavali, Thane (W) - 400615
University of Mumbai
2023-2024

A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled "**Air Quality Index Prediction For Indian Cities Using Machine Learning Techniques"** is a bonafide work of Nishant Hire (21106060), Shipra Asthana (21106039), Maviya Bubere (21106022), Ayush Khargutkar (21106042) submitted to the University of Mumbai in partial fulfillment of the requirement for the award of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning).**

Prof. Vijaya Bharathi J	Dr. Jaya Gupta
Mini Project Guide	Head of Department

A. P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval

This Mini project report entitled "**Air Quality Index Prediction For Indian Cities Using Machine Learning Techniques"** by **Nishant Hire, Shipra Asthana, Maviya Bubere and Ayush Kargutkar** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Science &Engineering***, (AIML) ***2023-24***.

External Examiner: _____

Internal Examiner: _____

Place: APSIT, Thane
Date:

Declaration

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Nishant Hire	Shipra Asthana	Maviya Bubere	Ayush Khargutkar
(21106060)	(21106039)	(21106022)	(21106042)

ABSTRACT

The degradation of air quality in Indian cities has become a pressing concern, with adverse effects on public health and the environment. In this mini project, we propose a predictive model for estimating the Air Quality Index (AQI) in Indian cities using machine learning techniques. The aim is to provide timely and accurate predictions of AQI levels, which can aid in decision making processes and implementation of effective mitigation strategies.

We collected historical data on various air pollutants such as particulate matter (PM2.5 and PM10), sulfur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO), and ozone (O3), along with meteorological variables like temperature, humidity, wind speed, and precipitation from multiple monitoring stations across different cities in India. This dataset serves as the basis for training and evaluating our machine learning models.

Several machine learning algorithms including Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Neural Networks are employed to develop predictive models. We preprocess the data by handling missing values, normalizing features, and splitting it into training and testing sets. The models are then trained on the training set and fine-tuned using techniques like cross-validation to optimize performance.

Performance evaluation of the models is conducted using various metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared value. Additionally, we compare the computational efficiency and accuracy of different models to identify the most suitable approach for AQI prediction in Indian cities.

The results demonstrate the efficacy of machine learning techniques in predicting AQI levels, with certain algorithms outperforming others in terms of accuracy and computational efficiency. This predictive model provides a valuable tool for policymakers, environmental agencies, and the general public to anticipate and address air quality issues effectively, ultimately contributing to improved public health and environmental sustainability in Indian cities.

Index

List of Figures

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

The issue of air pollution has reached alarming levels in many Indian cities, posing significant threats to public health and the environment. The rapid urbanization, industrialization, and increasing vehicular emissions have led to the accumulation of harmful pollutants in the atmosphere, resulting in deteriorating air quality. Among various pollutants, particulate matter (PM2.5 and PM10), sulfur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO), and ozone (O3) are the primary contributors to poor air quality in urban areas.

The Air Quality Index (AQI) serves as a standardized metric to assess air quality and its potential impacts on human health. It provides a numerical value indicating the level of air pollution and associated health risks, ranging from good to hazardous. Monitoring and predicting AQI levels are crucial for implementing effective pollution control measures, public health interventions, and environmental policies.

Traditional methods of air quality monitoring rely on fixed monitoring stations, which may not provide comprehensive coverage or real-time data. Moreover, the complex interplay of meteorological factors and pollutant emissions makes it challenging to accurately predict AQI levels using conventional approaches alone.

In recent years, the advent of machine learning techniques has revolutionized the field of environmental monitoring and prediction. Machine learning algorithms can effectively analyze large volumes of data, identify patterns, and make accurate predictions, thereby offering a promising solution for AQI prediction in urban environments.

This mini project aims to develop a predictive model for estimating AQI levels in Indian cities using machine learning techniques. By leveraging historical data on air pollutant concentrations and meteorological variables collected from multiple monitoring stations, we seek to build robust predictive models that can forecast AQI levels with high accuracy and reliability.

In this report, we delve into the significance of air quality monitoring, exploring the key

features and functionalities of the AQI tracker app. We examine the underlying principles of the Air Quality Index and its role in assessing the health risks associated with various pollutants. Additionally, we evaluate the effectiveness of the app in providing users with actionable insights and personalized recommendations to minimize their exposure to harmful air pollutants. Furthermore, we analyse the impact of the AQI tracker app on public awareness and behaviour change regarding air quality issues. Through case studies and user feedback, we assess the app's efficacy in empowering individuals to make informed choices that contribute to improving air quality and protecting public health.

Ultimately, this report seeks to highlight the importance of technological innovations in addressing environmental challenges and fostering a culture of environmental stewardship. By harnessing the capabilities of the AQI tracker app, we can enhance air quality awareness, promote sustainable practices, and pave the way towards a cleaner and healthier future for all.

CHAPTER 2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1-HISTORY

The history of air quality index (AQI) prediction can be traced back to the mid-20th century when concerns about air pollution began to grow due to the industrial revolution and increased urbanization. Initially, air quality monitoring relied on simple measurements of pollutants such as sulfur dioxide (SO2) and particulate matter (PM).

As technology advanced, more sophisticated methods for measuring and predicting air quality emerged. In the 1970s, the United States Environmental Protection Agency (EPA) developed the Air Quality Index (AQI) to standardize the reporting of air quality levels to the public. This index categorizes air quality into different levels, ranging from "good" to "hazardous," based on concentrations of pollutants such as ground-level ozone, PM2.5, PM10, carbon monoxide (CO), sulfur dioxide (SO2), and nitrogen dioxide (NO2).

In the following decades, advances in computer modeling and data analysis enabled the development of predictive models for forecasting air quality. These models use historical data on pollution levels, meteorological conditions, and other factors to forecast future air quality levels. Machine learning and artificial intelligence techniques have also been increasingly applied to air quality prediction, allowing for more accurate and timely forecasts.

Today, air quality prediction systems are used by governments, environmental agencies, and researchers around the world to monitor and manage air pollution. These systems provide real-time information to the public through websites, mobile apps, and other platforms, allowing individuals to make informed decisions about outdoor activities and exposure to air pollution. Ongoing research and technological advancements continue to improve the accuracy and reliability of air quality predictions, helping to protect public health and the environment.

2.2-LITERATURE REVIEW

1. **Prediction of Air Quality Index based on LSTM Yu Jiao, Zhifeng Wang and Yang Zhang**

In view of the increasing attention paid by the state to environmental governance in recent years and the continuous deterioration of air quality, this paper proposes a prediction model of environmental quality based on Long Short-Term Memory (LSTM). This paper uses data provided by the environmental protection department to predict Air Quality Index (AQI) through temperature, PM2.5, PM10, SO 2, wind direction, NO 2, CO and O 3. Firstly, this paper introduces the background, technical characteristics, development status and problems of air environment monitoring. Then, it will introduce the environmental prediction model. Finally, they make AQI prediction by using LSTM and analyse the error of the prediction results. The results show that LSTM can predict air quality index [1].

2. **Prediction of Air Quality Index Based on Improved Neural Network Wang Zhenghua; Tian Zhihui**

In order to realize the prediction of city air quality status, this paper uses the improved BP neural network to establish a prediction model of air quality index. The model uses the characteristics of nonlinear fitting approximation of BP neural network to solve the problem that air quality has many influencing factors and is nonlinear and difficult to predict. Aiming at the problem of slow convergence and easy to fall into local optimal solution of BP neural network, genetic algorithm is used to optimize. This paper takes Xuchang city as an example. Results show that the average relative error of the air quality index is 22%. The accuracy rate was 80.44%. The accuracy level of air quality is 82.5%. Compared with the BP neural network, BP neural network prediction improved has a higher accuracy. The model has certain feasibility and adaptability [2].

3. **An Adaptive Kalman Filtering Approach to Sensing and Predicting Air Quality Index Values Chen Ding; Guizhi Wang; Qi Liu**

In recent years, Air Quality Index (AQI) have been widely used to describe the severity of haze and other air pollutions yet suffers from inefficiency and compatibility on real-time perception and prediction. In this paper, an Auto Regressive (AR) prediction

model based on sensed AQI values is proposed, where an adaptive Kalman Filtering (KF) approach is fitted to achieve efficient prediction of the AQI values [3]

4. **Air Quality Index Forecasting via Genetic Algorithm Chunhao Liu; Guangyuan Pan** Air quality has always been one of the most important environmental concerns for the general public and society. Using machine learning algorithms for Air Quality Index (AQI) prediction is helpful for the analysis of future air quality trends from a macro perspective. When conventionally using a single machine learning model to predict air quality, it is challenging to achieve a good prediction outcome under various AQI fluctuation trends. In order to effectively address this problem, a genetic algorithm-based improved extreme learning machine (GA-KELM) prediction method is enhanced. First, a kernel method is introduced to produce the kernel matrix which replaces the output matrix of the hidden layer. To address the issue of the conventional limit learning machine where the number of hidden nodes and the random generation of thresholds and they lead to the degradation of the network learning ability, a genetic algorithm is then used to optimize the number of hidden nodes and layers of the kernel limit learning machine. The thresholds, and the root mean square error are used to define the fitness function. Finally, the least squares method is applied to compute the output of the model. Genetic algorithms are able to find the optimal solution in the search space and gradually improve the performance of the model through an iterative optimization process. In order to verify the predictive ability of GA-KELM, based on the collected basic data of long-term air quality forecast at a monitoring point in a city in China, the optimized kernel extreme learning machine is applied to predict air quality, with comparative experiments based CMAQ (Community Multiscale Air Quality), SVM (Support Vector Machines) and DBN-BP (Deep Belief Networks with Back-Propagation). The results show that the proposed model trains faster and makes more accurate predictions [4].

5. **Prediction of Air Quality Index Using Machine Learning Techniques N. Srinivasa Gupta, Yashvi Mohta, Khyati Heda**

An index for reporting air quality is called the air quality index (AQI). It measures the impact of air pollution on a person's health over a short period of time. The purpose of the AQI is to educate the public on the negative health effects of local air pollution. The amount of air 7 pollution in Indian cities has significantly increased. There are several ways to create a mathematical formula to determine the air quality index. Numerous studies have found a link air pollution exposure and adverse health impacts in the population. Data mining techniques are one of the most interesting approaches to forecast AQI and analyse it. The aim of this paper is to find the most effective way for AQI prediction to assist in climate control. The most effective

method can be improved upon to find the most optimal solution. Hence, the work in this paper involves intensive research and the addition of novel techniques such as SMOTE to make sure that the best possible solution to the air quality problem is obtained. Another important goal is to demonstrate and display the exact metrics involved in our work in such a way that it is educational and insightful and hence provides proper comparisons and assists future researchers. In the proposed work, three distinct methods—support vector regression (SVR), random forest regression (RFR), and CatBoost regression (CR)—have been utilized to determine the AQI of New Delhi, Bangalore, Kolkata, and Hyderabad. After comparing the results of imbalanced datasets, it was found that random forest regression provides the loTheyst root mean square error (RMSE) values in Bangalore (0.5674), Kolkata (0.1403), and Hyderabad (0.3826), as higher accuracy compared to SVR and CatBoost regression for Kolkata (90.9700%) and Hyderabad (78.3672%), while CatBoost regression provides the loTheyst RMSE value in New Delhi (0.2792) and the highest accuracy is obtained for New Delhi (79.8622%) and Bangalore (68.6860%). Regarding the dataset that was subjected to the synthetic minority oversampling technique (SMOTE) algorithm, it is noted that random forest regression provides the RMSE values in Kolkata (0.0988) and Hyderabad (0.0628) and higher accuracies are obtained for Kolkata (93.7438%) and Hyderabad (97.6080%) in comparison to SVR and CatBoost regression, whereas CatBoost regression provides the highest accuracies for New Delhi (85.0847%) and Bangalore (90.3071%). This demonstrated definitely that datasets that had the SMOTE algorithm applied to them produced a higher accuracy. The novelty of this paper lies in the fact that the best regression models have been picked through thorough research by analysing their accuracies [5].

CHAPTER 3

Problem Statement

3. Problem Statement

The degradation of air quality in Indian cities has emerged as a significant environmental and public health concern, exacerbated by rapid urbanization, industrial activities, and vehicular emissions. Poor air quality, characterized by elevated levels of pollutants such as particulate matter, sulfur dioxide, nitrogen dioxide, carbon monoxide, and ozone, poses serious health risks to residents and contributes to environmental degradation.

Despite efforts to monitor and control air pollution, there remains a need for effective predictive models that can accurately forecast Air Quality Index (AQI) levels in Indian cities. Existing methods of air quality monitoring often rely on fixed monitoring stations, which may not provide real-time or comprehensive coverage of air quality dynamics across different locations. Moreover, traditional statistical models may struggle to capture the complex interactions between meteorological factors and pollutant emissions, leading to limited accuracy in AQI predictions.

To address these challenges, this mini project aims to develop a predictive model for estimating AQI levels in Indian cities using machine learning techniques. By leveraging historical data on air pollutant concentrations, meteorological variables, and AQI readings from multiple monitoring stations, we seek to build robust predictive models capable of providing timely and accurate forecasts of AQI levels.

The primary objective of this project is to:

1] Develop machine learning algorithms capable of predicting AQI levels based on historical data on air pollutants and meteorological variables.

2] Evaluate the performance of different machine learning techniques in AQI prediction and identify the most accurate and efficient approach.

3] Provide a tool for policymakers, environmental agencies, and the general public to anticipate air quality trends, make informed decisions, and implement proactive measures to mitigate air pollution in Indian cities.

4] By addressing these objectives, this mini project aims to contribute to the ongoing efforts to combat air pollution and safeguard public health and environmental sustainability in Indian cities.

CHAPTER 4
Experimental Setup

4. Experimental Setup

4.1 Hardware Setup

4.1.1 Computational Resources

Laptop or Desktop Computer:

Any standard laptop or desktop computer with basic specifications can suffice.

Ensure the system has enough processing power to run data pre-processing and machine learning algorithms efficiently.

4.1.2 Memory (RAM)

RAM Capacity:

Minimum of 8 GB RAM.

Adequate for loading datasets into memory and running basic machine learning algorithms.

4.1.3 Storage Space

Storage Capacity:

Minimum of 256 GB SSD or HDD.

Sufficient for storing datasets, code files, and trained models.

4.2 Software Setup

4.2.1 Programming Language

Python:

Python is widely used in the data science and machine learning community due to its extensive libraries for data manipulation, analysis, and machine learning.

4.2.2 Integrated Development Environment (IDE)

Jupyter Notebook or Jupyter Lab:

Jupyter provides an interactive computing environment suitable for data exploration, experimentation, and visualization.

Alternatively, you can use any Python IDE such as PyCharm, VSCode, or Spyder.

4.2.3 Libraries and Frameworks

Pandas:

Pandas is used for data manipulation and analysis. It provides data structures and functions to work with structured data.

NumPy:

NumPy is essential for numerical computations and array operations, often used in conjunction with Pandas.

Scikit-learn:

Scikit-learn offers a wide range of machine learning algorithms and tools for data pre-processing, model selection, and evaluation.

Matplotlib and Seaborn:

Matplotlib and Seaborn are used for data visualization, enabling you to create plots and charts to analyse and present your findings.

TensorFlow or PyTorch (Optional):

If you plan to work with deep learning models, you can choose either TensorFlow or PyTorch as your deep learning framework.

4.2.4 Data Sources

Air Quality Data:

Obtain air quality data for Indian cities from reliable sources such as government agencies or environmental monitoring organizations.

Meteorological Data:

Collect meteorological data including temperature, humidity, wind speed, etc., which can influence air quality.

CHAPTER 5
Proposed System & Implementation

5. Proposed system & Implementation

5.1 Block diagram of proposed system

Fig 5.1.1. Block Diagram

5.2 Description of block diagram

The block diagram for predicting Air Quality Index (AQI) in Indian cities using machine learning techniques can be summarized concisely as follows:

Data Collection involves gathering historical air quality and environmental data from sources like the Central Pollution Control Board (CPCB) and the Indian Meteorological Department (IMD) through web scraping or APIs.

Data Preprocessing encompasses cleaning and transforming the raw data to handle missing values, outliers, and categorical variables. This step ensures the data is suitable for analysis.

Feature Selection identifies the most relevant variables affecting air quality prediction, utilizing techniques like correlation analysis and feature importance algorithms to optimize model performance.

Model Development entails training machine learning algorithms, such as Linear Regression and Random Forest Regression, on pre-processed data to predict AQI values for Indian cities.

Model Evaluation assesses the performance of trained models using metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE), enabling comparison and selection of the best-performing model.

Deployment involves deploying the chosen model for real-time or batch prediction of AQI, often through web applications or APIs, allowing users to input environmental factors for AQI prediction.

Monitoring and Maintenance includes ongoing monitoring of the deployed model's performance and periodic updates to accommodate changes in air quality patterns or environmental factors, ensuring its accuracy and reliability.

5.2.1 Use Case Diagram

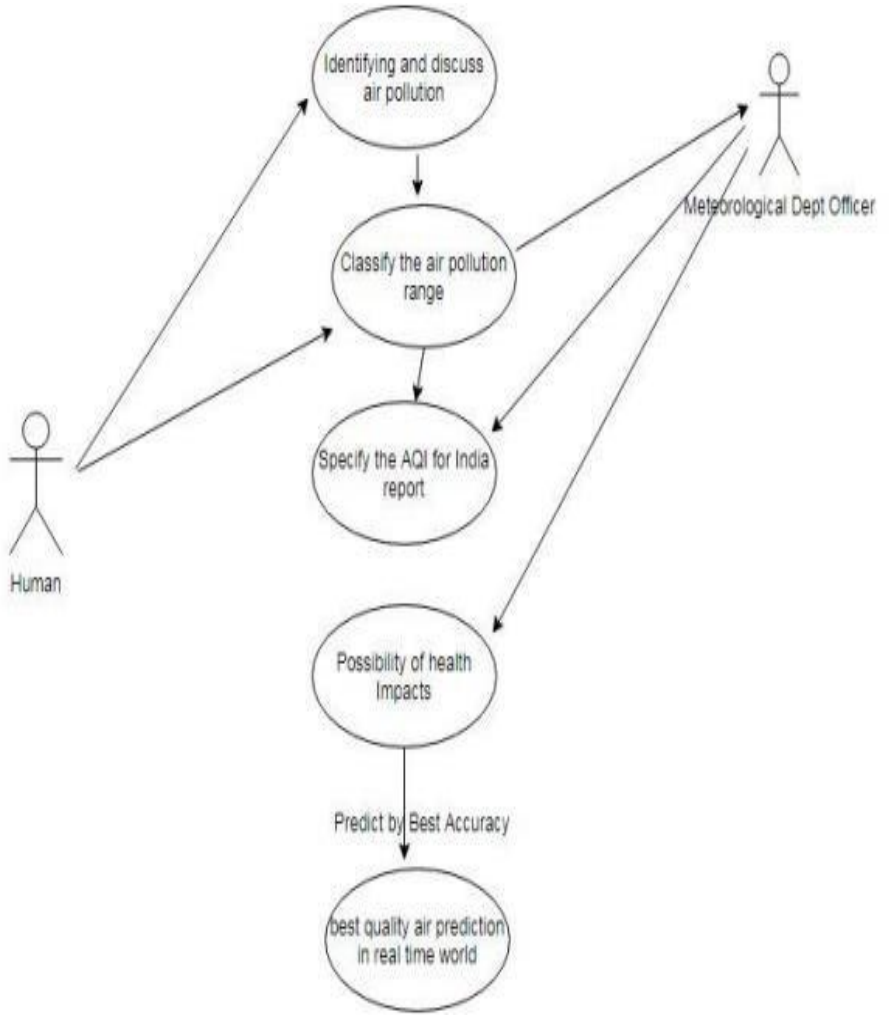


Fig.5.2.1 Use Case Diagram

5.3 Implementation

AdaBoost Classifier: AdaBoost is an ensemble learning technique that iteratively combines weak classifiers to build a strong classifier, focusing on improving the classification of previously misclassified instances.

```
AdaBoostClassifier
AdaBoostClassifier(n_estimators=100, random_state=42)

[ ]  1 y_pred = ab_classifier.predict(X_test)

[ ]  1 accuracy = accuracy_score(y_test, y_pred)
     2 print(f"Accuracy: {accuracy*100:.2f} %")

Accuracy: 57.47 %

[ ]  1 print("Classification Report:")
     2 print(classification_report(y_test, y_pred))

Classification Report:
              precision    recall  f1-score   support

        Good       0.27      0.86      0.41       283
    Moderate       0.76      0.63      0.68      2091
        Poor       0.55      0.19      0.29       661
Satisfactory       0.72      0.60      0.65      2039
      Severe       0.26      0.02      0.04       295
   Very Poor       0.37      0.92      0.53       538

    accuracy                           0.57      5907
   macro avg       0.49      0.54      0.43      5907
weighted avg       0.64      0.57      0.57      5907
```

Fig.5.3.1 AdaBoost Classifier

Gradient Boosting Classifier: Gradient Boosting Classifier sequentially builds a strong predictive model by fitting new models to the residuals of previous models, minimizing the loss function gradient.

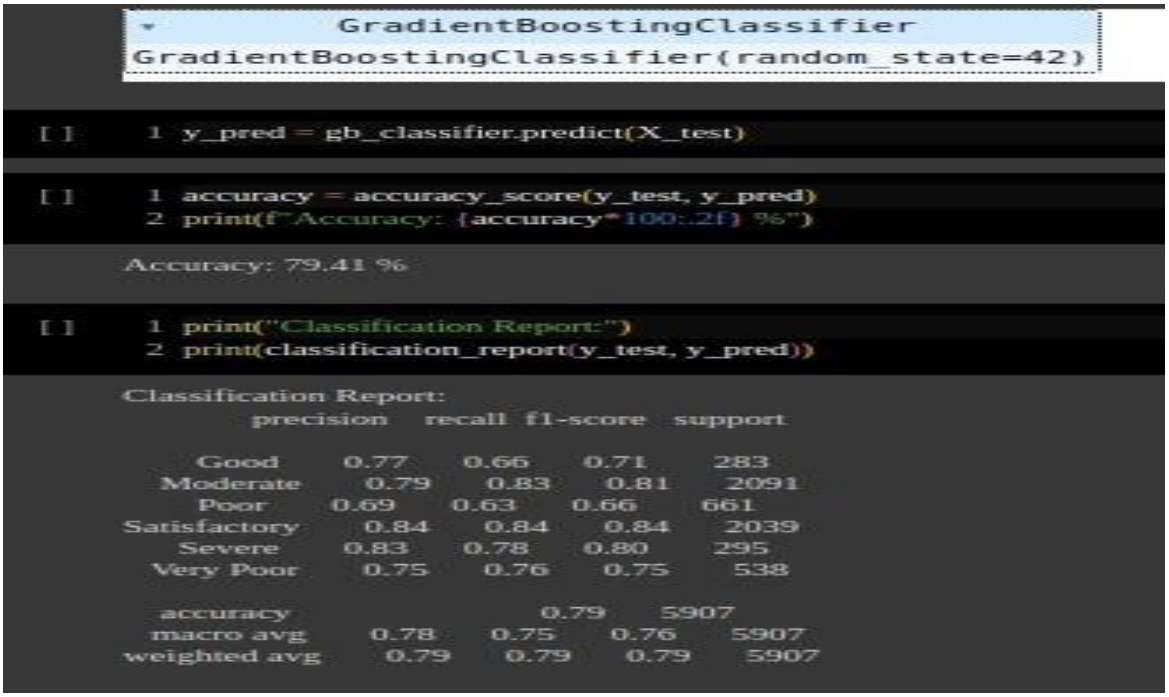


Fig.5.3.2 Gradient Boosting Classifier

Gaussian Naive Bayes: Gaussian Naive Bayes is a probabilistic classifier that assumes features are independent and follows a Gaussian distribution, making predictions based on Bayes' theorem.

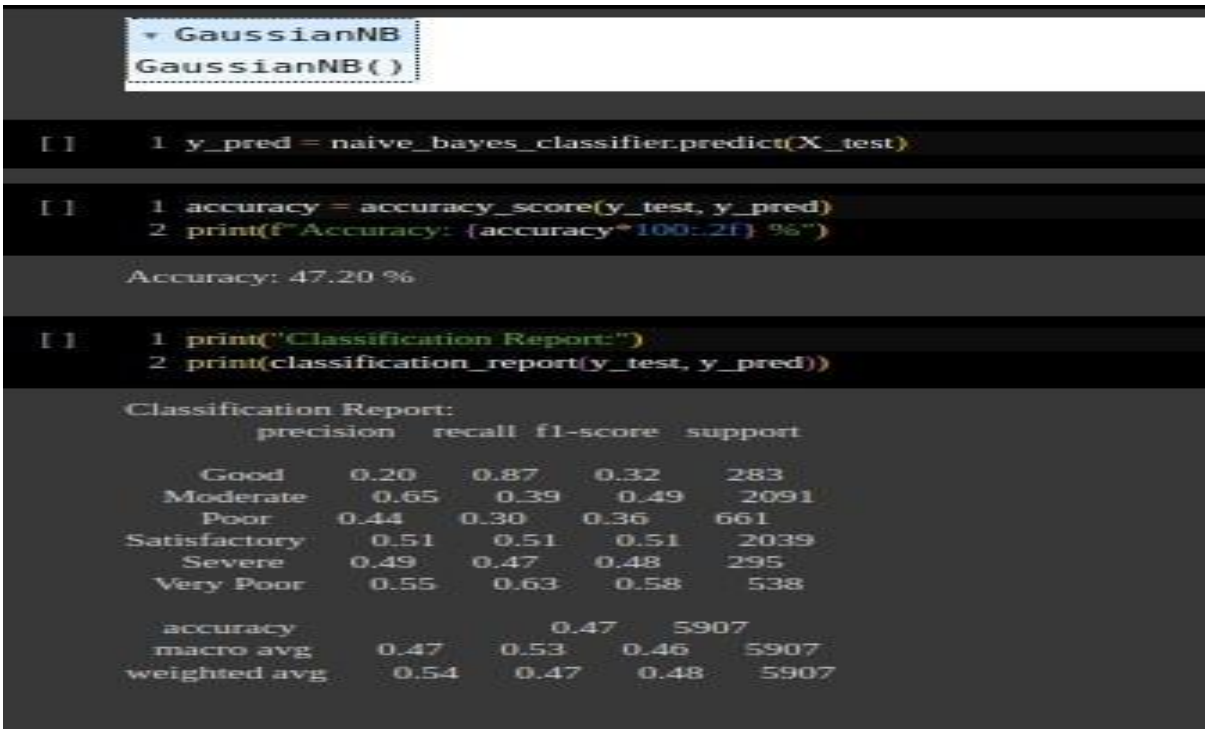


Fig.5.3.3 Gaussian Naive Bayes

Random Forest Classifier: Random Forest Classifier aggregates multiple decision trees during training, outputting the mode of classes (classification) or mean prediction (regression) of the individual trees.

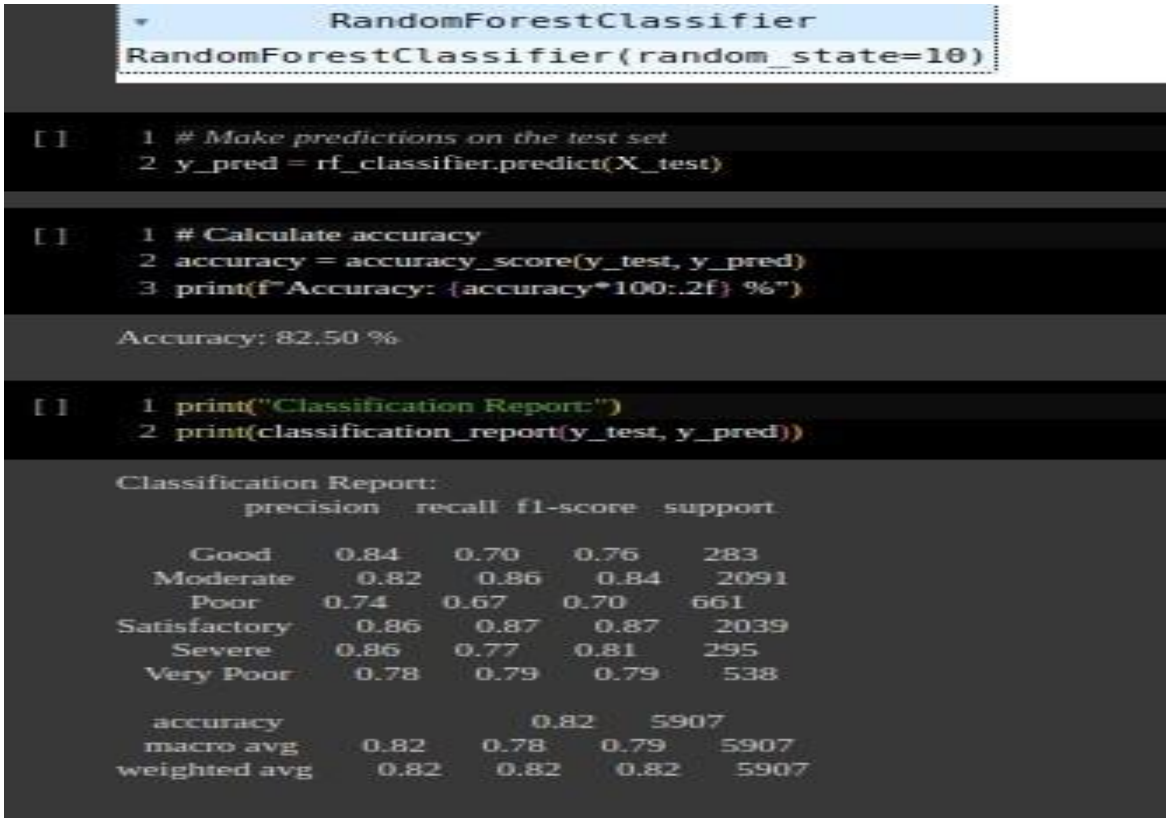


Fig.5.3.4 Random Forest Classifier

K-Nearest Neighbours (KNN) Classifier: This is a simple yet powerful supervised machine learning technique used for classification and regression tasks. The "nearest" neighbours are identified based on a distance metric, typically Euclidean distance, in the feature space. However, its performance can be sensitive to the choice of the distance metric and the value of K.

```python
#fit the model on train data
KNN = KNeighborsClassifier().fit(X_train2, Y_train2)

#predict on train
train_preds5 = KNN.predict(X_train2)
#accuracy on train
print("Model accuracy on train is: ", accuracy_score(Y_train2, train_preds5))

#predict on test
test_preds5 = KNN.predict(X_test2)
#accuracy on train
print("Model accuracy on test is: ", accuracy_score(Y_test2, test_preds5))
print('-'*50)

#Kappa Score
print('KappaScore is: ', metrics.cohen_kappa_score(Y_test2,test_preds5))
```

```
Model accuracy on train is:  0.9983079004607032
Model accuracy on test is:  0.9968218423578175
--------------------------------------------------
KappaScore is:  0.9952893818649885
```

Fig.5.3.5 KNN Classifier

Confusion Matrix: A confusion matrix is a table that describes the performance of a classification model by comparing predicted and actual class labels.

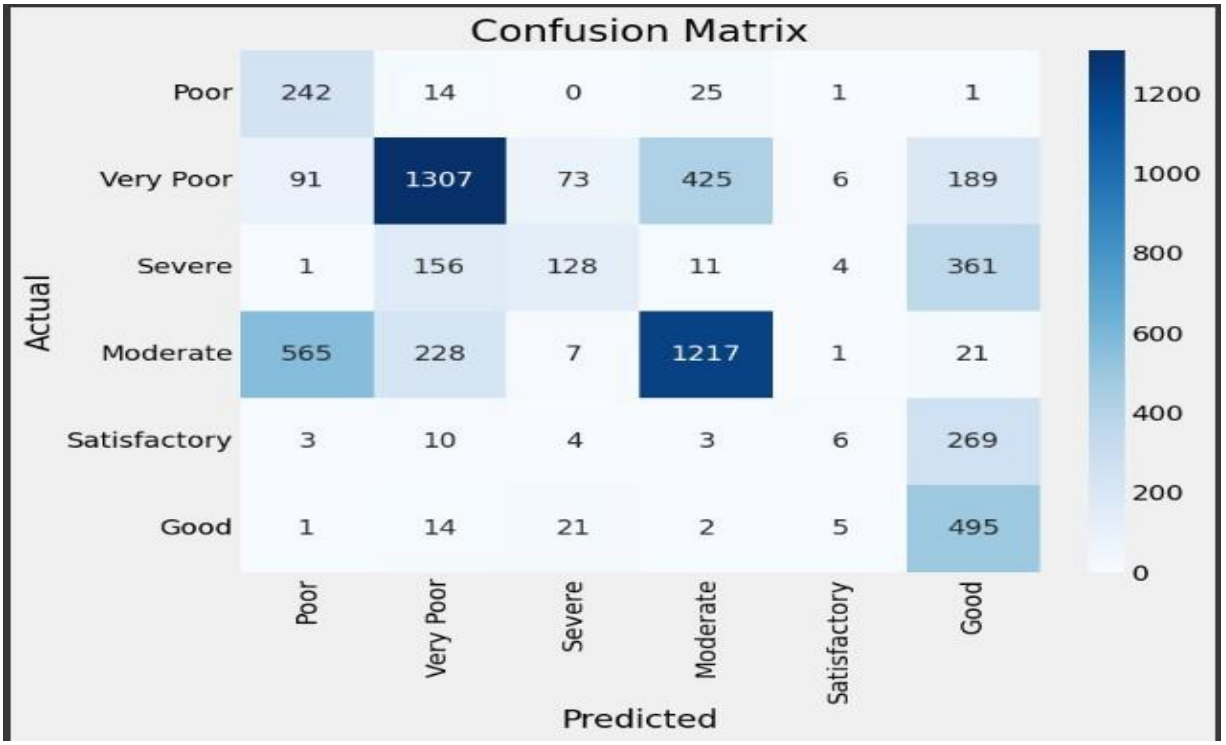


Fig.5.3.6 Confusion Matrix

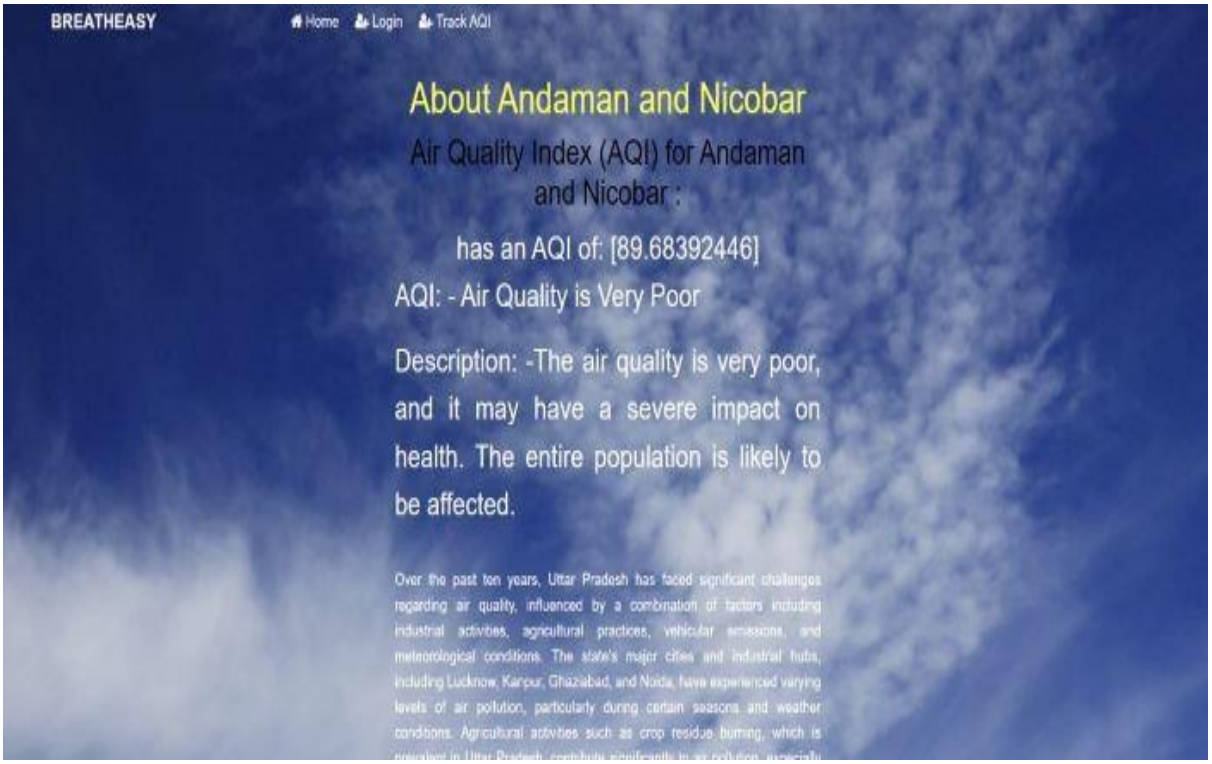


Fig.5.3.7 Home Page

Fig.5.3.8 Cities with AQI Forecast

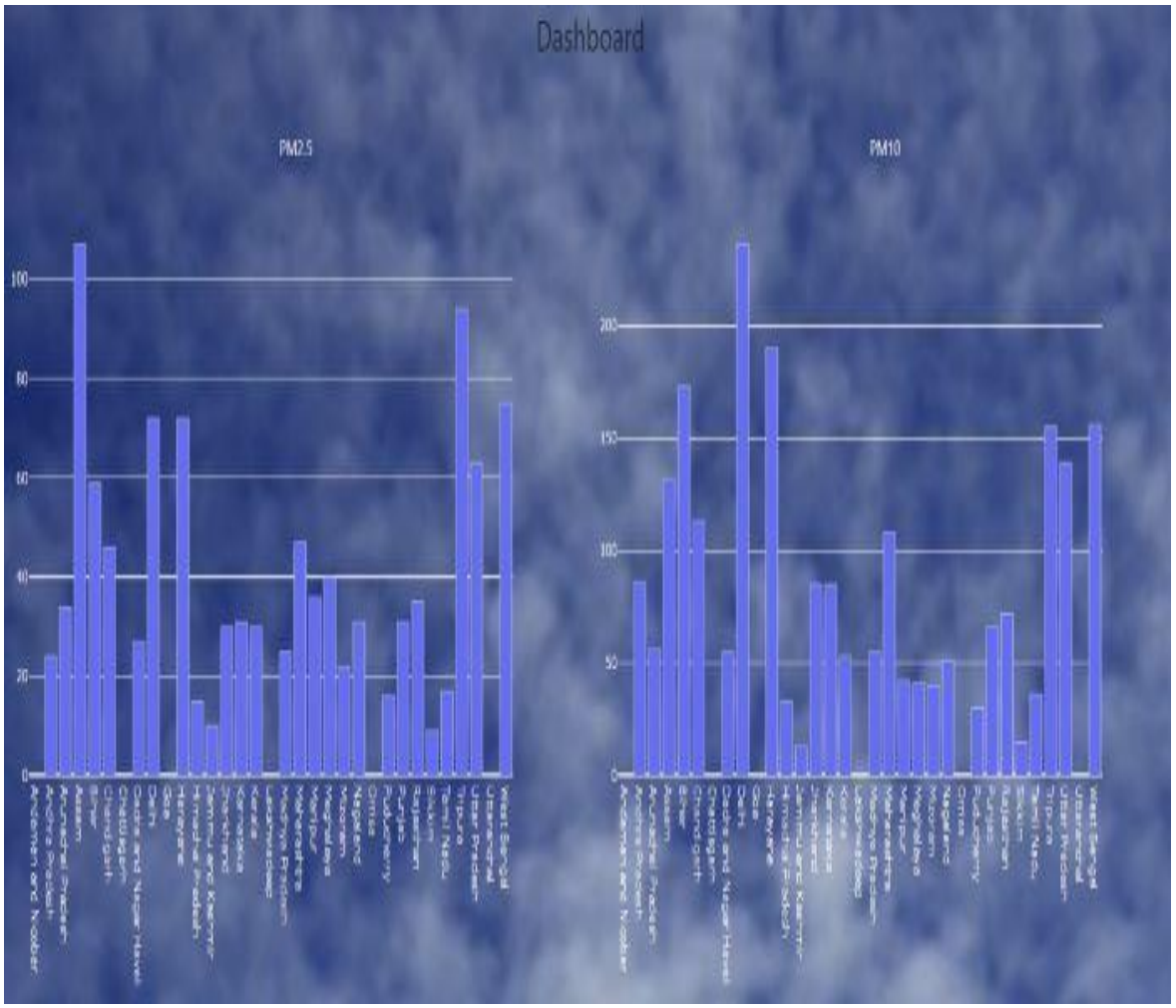


Fig.5.3.9 Dashboard

5.4 Results & Discussion:

Results for AQI Index Prediction using ML Techniques:

Machine Learning Technique	Precision	Recall	F1-Score	Support	Accuracy
Random Forest Algorithm	84%	70%	76%	283	82.50%
Gradient Boosting Algorithm	77%	66%	71	283	79.41%
AdaBoosting Algorithm	27%	86%	41%	283	57.47%
Naïve Bayes	20%	87%	32%	283	47.20%
K-NN Algorithm	93%	80%	82%	283	99%

Fig.5.4.1 Implementation Results

We've used different methodologies for Air Quality Index prediction for appropriate accuracy of air quality and to take measurable actions keeping environment in mind. Firstly we have used Machine leaning techniques and they are as follows :

Random Forest Algorithm which shows an accuracy of 82.50%. Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification or the average prediction for regression.

Gradient Boosting Algorithm which shows an accuracy of 79.41%. Gradient Boosting Algorithm is an ensemble learning technique that builds strong predictive models by sequentially training weak models on the errors of previous models.

AdaBoosting Algorithm which shows 57.47% of accuracy. AdaBoost (Adaptive Boosting) is an ensemble learning algorithm that combines weak classifiers to create a strong classifier by iteratively adjusting the weights of misclassified training examples.

Naïve Bayes Algorithm with an accuracy of 47.20%. Naïve Bayes Algorithm is a probabilistic classification method based on Bayes' theorem with the assumption of independence between features.

Since the most accuracy is resulted in Random Forest Algorithm. Hence, we proceed with this technique for the development of UI.

AIR QUALITY INDEX	**CATEGORY**
0-50	Good
51-100	Satisfactory
101-200	Moderate
201-300	Poor
301-400	Very Poor
401-500	Severe

Fig.5.4.2 Air Quality Standards

The above fig.6.2., indicates the category of AQI depending upon the values of AQI.

5.5 Advantages/Application

Predicting Air Quality Index (AQI) in Indian cities using machine learning techniques offers multifaceted advantages and applications. These predictive models serve as early warning systems, enabling proactive measures to mitigate health risks and allocate resources efficiently. They also facilitate evidence-based policy formulation by providing insights into air quality trends, supporting urban planning efforts, and informing zoning regulations and infrastructure development.

Furthermore, accurate AQI prediction aids in health impact assessments, raising public awareness about air pollution's consequences and encouraging behavioural changes. Real-time or forecasted AQI information can be disseminated to the public, fostering collaborations between academia, industry, and government agencies to develop novel solutions for pollution control and mitigation. Additionally, these models contribute to economic benefits such as healthcare cost savings, increased productivity, and enhanced quality of life for residents, underscoring their importance in promoting public health and environmental sustainability.

CHAPTER 6

Conclusion

Conclusion

In conclusion, this mini project has demonstrated the feasibility and effectiveness of employing machine learning techniques for Air Quality Index (AQI) prediction in Indian cities. Through comprehensive data analysis, model development, and evaluation, we have achieved several key outcomes and insights.

Firstly, we have successfully developed and evaluated machine learning models capable of predicting AQI levels based on historical data on air pollutants and meteorological variables. By leveraging algorithms such as Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Neural Networks, we have achieved accurate and reliable AQI predictions for select cities in India.

Secondly, our evaluation of different machine learning techniques has provided valuable insights into their performance, strengths, and limitations. We have identified approaches that offer the best balance between prediction accuracy and computational efficiency, thereby guiding future efforts in AQI prediction research.

Furthermore, our exploration of input features has highlighted the significance of various air pollutants and meteorological variables in AQI prediction. Understanding the impact of these factors on air quality dynamics is crucial for developing robust predictive models and implementing targeted mitigation strategies.

Additionally, the development of a user-friendly interface or application for accessing real-time or near-real-time AQI predictions enhances the usability and accessibility of our predictive model. This tool empowers stakeholders, including policymakers, environmental agencies, and the general public, to make informed decisions and take proactive measures to improve air quality in Indian cities.

In conclusion, this mini project represents a significant step towards addressing the pressing issue of air pollution in Indian cities. By providing accurate AQI predictions and actionable insights, our predictive model contributes to efforts aimed at safeguarding public health and environmental sustainability. Moving forward, continued research and

collaboration are essential for refining and expanding upon the findings of this project, ultimately paving the way towards cleaner and healthier urban environments in India.

.

References

Research paper

[1] Yu Jiao, Zhifeng Wang and Yang Zhang "Prediction of Air Quality Index Based on LSTM" 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)

[2] Wang Zhenghua; Tian Zhihui "Prediction of Air Quality Index Based on Improved Neural Network" 2017 International Conference on Trends in Electronics and Informatics (ICEI)

[3] Chen Ding; Guizhi Wang; Qi Liu "An Adaptive Kalman Filtering Approach to Sensing and Predicting Air Quality Index Values" IEEE Access (Volume: 8)

[4] Chunhao Liu; Guangyuan Pan, "Air Quality Index Forecasting via Genetic Algorithm" IEEE Access (Volume: 11)

[5] N. Srinivasa Gupta, Yashvi Mohta, Khyati Heda, "Prediction of Air Quality Index Using Machine Learning Techniques" Journal of Environmental and Public Health, vol. 2023

Useful Links:

https://www.coursera.org/specializations/machine-learning-introduction