Machine Learning Based System for the Prediction and Analysis of Air Pollution

Dr.G.Sharada[1],
Head of the Department
Department of Information Technology
Malla Reddy College Of Engineering & Technology
Hyderabad, India.
gsharada8@gmail.com

Bolla Sai Laxmi[2],
Final Year B.Tech Student
Department of Information Technology
Malla Reddy College Of Engineering & Technology
Hyderabad, India.
sailaxmi1515@gmail.com

Bhavani Nikhil[3],
Final Year B.Tech Student
Department of Information Technology
Malla Reddy College Of Engineering & Technology
Hyderabad, India.
bhavaninikhil11@gmail.com

Dyagam Vamshi Krishna[4],
Final Year B.Tech Student
Department of Information Technology
Malla Reddy College Of Engineering & Technology
Hyderabad, India.
dyagamvamshi13@gmail.com

Abstract

Air pollution in contemporary society presents a significant and pressing issue. The accelerating growth across commercial, social, and economic domains has led to a corresponding escalation in pollutant concentrations across various regions globally, precipitating disruptions to human life. With the increasing concern over rising pollution levels, there is a growing need for effective tools to predict and manage air quality.

The primary objective of Air Quality Prediction using Machine Learning is to develop accurate and reliable models that can forecast Air Quality Index based on historical and real-time data. Machine Learning techniques can analyze vast datasets encompassing various environmental factors to forecast air quality levels. This project yields a user-friendly system that provides clear results, categorizing air quality as either good or bad based on the evaluated parameters. It could provide a comprehensive solution with the potential to make a meaningful impact on air quality management and public health.

Keywords- Air Quality Index, Machine Learning, Exploratory Analysis, Pollutant Concentration levels, Data Dynamics

I. Introduction

In contemporary society, energy use and its effects are unavoidable aspects of human activity. Human-made sources of air pollution encompass emissions from industries, vehicles, aircraft, the combustion of various fuels like straw, coal, and kerosene, as well as products like aerosol cans. These activities result in the continuous release of harmful pollutants such as carbon monoxide (CO), carbon dioxide (CO2), particulate matter (PM), nitrogen dioxide (NO2), Sulfur dioxide (SO2), ozone (O3), ammonia (NH3), among others, into our surroundings on a daily basis. The chemicals and particles found in air pollution have harmful effects on the health of humans, animals, and plants. They can trigger a range of serious illnesses in humans, including bronchitis, heart disease, pneumonia, and lung cancer. Additionally, poor air quality contributes to various environmental problems such as global warming, acid rain, reduced visibility, smog, aerosol formation, climate change, and premature deaths.

Scientists have come to understand that air pollution can harm historical landmarks. Emissions from vehicles, power plants, factories, and agriculture contribute to higher levels of greenhouse gases, which in turn disrupt climate conditions and hinder plant growth. These emissions also interfere with plant-soil interactions. Fluctuations in climate not only impact humans and animals but also agricultural productivity, leading to economic losses. The Air Quality Index (AQI) directly relates to public health, with higher AQI levels indicating greater risk to human health. Consequently, there's a growing need to predict AQI in advance, driving scientists to monitor and model air quality. Monitoring and forecasting AQI, particularly in urban areas with increasing motor and industrial activities, have become crucial yet challenging tasks.

Mostly, the air quality-based studies and research works target the developing countries, although the concentration of the most deadly pollutant like PM2.5 is found to be in multiple folds in developing countries. A few researchers endeavoured to undertake the study of air quality prediction for Indian cities.

After reviewing existing literature, there was a recognized need to address this gap by analysing and forecasting Air Quality Index (AQI) for India. Different models have been explored, including statistical, deterministic, physical, and Machine Learning (ML) models. Traditional methods relying on probability and statistics are often complex and less effective. ML-based AQI prediction models have demonstrated greater reliability and consistency. Achieving accurate and dependable predictions from extensive environmental data necessitates thorough analysis, a task that ML algorithms excel at handling efficiently. supervised ML algorithms are applied for environment protection issues. The present work investigates six years of air pollution data of the Indian cities and analyse twelve air pollutants and AQI. First, the dataset undergoes preprocessing and cleaning to ensure its quality. Then, data visualization techniques are applied to gain deeper insights and uncover underlying patterns and trends. Additionally, this study employs correlation coefficients alongside ML models to analyse relationships between variables and improve predictive accuracy. It identifies and handles data imbalances using a resampling technique. Four well-known ML models are employed alongside this resampling method. Machine learning algorithms learn from past experiences, adapt to changes, and become more efficient at their designated tasks over time. Therefore, ML techniques prove effective in building prediction models for forecasting air pollution. However, selecting the optimal ML technique depends on the specific problem, considering both ecological and environmental factors is crucial.

II. LITERATURE REVIEW

Xiao Feng, Qi Li, Yajie Zhu, and colleagues (2015) proposed a new hybrid model for predicting daily average PM2.5 concentration[1]. This model integrates trajectory-based geographic modeling, wavelet transformation, MLP neural networks, meteorological forecasts, and pollutant predictors to improve PM2.5 forecasting accuracy.

In a separate study, M. S. Baawain and A. S. Al-Serihi et al. (2014) conducted research in Sohar, Oman, gathering data for daily predictions of various pollutants including CO, PM10, NO, NO2, NOx, SO2, H2S, and O3[2]. They utilized the Multi-layer Perceptron method with Back-Propagation for training prediction models. Their results demonstrated strong agreement between actual and predicted concentrations. Additionally, they investigated the sensitivity of the MLP model to variations in the epochs cycle, employing a trial-and-error approach to optimize adjustments.

RuiJun Yang; HaiLong Zhou; DanFeng Ding 2018 11th International Symposium on Air Quality Prediction Method in Urban Residential Area [3].Using SVM, Naïve Bayesian, and KNN classification algorithms, a robust mapping between housing prices and air quality in urban areas is established, demonstrating high accuracy in predicting air quality in Tianhe, Guangzhou, offering valuable guidance for buyers.

Prediction of Air Quality Index Based on Improved Neural Network by Wang Zhenghua, Tian Zhihui(2017)[4].This study employs an enhanced Back Propagation neural network, leveraging its nonlinear fitting capabilities for air quality index prediction, supplemented by genetic algorithms to address convergence issues. Applied to Xuchang city, the model achieves an average relative error of 22%, 80.44% accuracy rate, and 82.5% air quality accuracy, surpassing traditional BP networks, underscoring its efficacy and adaptability.

Prediction of Air Quality Index Based on LSTM by Yu Jiao, Zhifeng Wang, Yang Zhang Published in 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC) [5].This paper introduces an LSTM-based model for Air Quality Index (AQI) prediction, leveraging environmental data from factors like temperature, PM2.5, PM10, SO2, wind direction, NO2, CO, and O3, showcasing LSTM's efficacy in AQI prediction with analysis of prediction errors.

III. Dataset Description and Sample Data

The dataset includes air quality and AQI (air quality index) data from numerous stations in several Indian cities. The data are for the years 2016 through 2021. The original dataset included 10260 rows and 16 columns, which included all of the cities listed below. The cities are given below: Ahmedabad, Amritsar, Aizawl, Amaravati, Brajrajnagar, Bangalore, Bhopal, Chandigarh, Chennai, Delhi, Hyderabad. The attribute information is given below. 3.1. Date YYYY-MM-DD, City, PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, AQI, and AQI_Bucket.

The AQI_Bucket variable consists of Five categories: good, satisfactory, moderate, poor, and severe, representing different levels of air quality. The initial dataset has an imbalanced composition. using the Synthetic Minority Oversampling Technique (SMOTE) algorithm, which balances class distributions through oversampling. SMOTE increases the frequency of the minority class in the dataset, ensuring balance and enhancing algorithm performance while preventing overfitting. By interpolating between existing data points based on nearest neighbours, SMOTE generates synthetic data points that differ slightly from the originals, improving model robustness. Overall, SMOTE effectively tackles class imbalance, improving accuracy and reliability of results .SMOTE has the benefit of not producing duplicate data points but rather artificial data points that are marginally different from the actual data points.

IV. METHODOLOGY

The proposed methodology involves using three different algorithms to compare the Air Quality Index (AQI) values of four major cities in India: New Delhi, Bangalore, Kolkata, and Hyderabad. This comparison will be based on various parameters such as PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, and toluene levels. The objective is to determine the most accurate and efficient algorithm for predicting AQI values in these cities. By focusing on cities with high population densities, the study aims to provide valuable insights into pollution levels in major South Asian urban centers. To streamline the research process, additional cities have not been included to avoid excessive length in the research paper. Hence, the major cities of India have been chosen to analyse The pollution levels in various urban cities of India are significant due to their substantial contribution to environmental pollution. Among the existing algorithms utilized, there are three prominent ones:

Naive Bayes: This algorithm is based on Bayes' theorem and serves as a classifier method, Support Vector Machine (SVM): SVM is a supervised learning model used for both classification and regression tasks.

The proposed algorithms for comparison are as follows:

4.1 Synthetic Minority Oversampling Technique (SMOTE) Algorithm:

This technique generates synthetic samples for minority classes, effectively balancing imbalanced datasets and addressing overfitting issues associated with random oversampling.

4.2 Support Vector Regression (SVR):

SVR is a supervised learning method used for discrete value prediction. Similar to SVMs, SVR aims to find the most suitable line or hyperplane that best fits the data.

4.3 Random Forest Regression (RFR) Algorithm:

RFR is a commonly used supervised machine learning technique for both classification and regression tasks. It constructs decision trees based on various samples and utilizes averaging for regression and voting for classification.

4.4 Decision Tree Algorithm:

Decision tree algorithms analyze complex environmental datasets, selecting relevant features and providing interpretable predictions of air pollution levels. They consider factors such as meteorological data and historical pollution records.

Data Analysis:

Step 1: Dataset Selection

The process involves carefully selecting a dataset from Kaggle that meets specific requirements related to air quality indices (AQI) in urban areas of India. The dataset should contain comprehensive information on pollutant levels and AQI measurements for various cities.

Step 2: Data Preprocessing

Data preprocessing encompasses several tasks aimed at preparing the dataset for analysis. This includes cleaning the dataset by removing null values and irrelevant entries. Specifically, data for major Indian cities such as Ahmedabad, Aizawl, Amaravati, etc., are extracted due to their significance in representing different pollution levels across various urban areas. The cleaning process involves utilizing Microsoft Excel software to filter out unnecessary and erroneous data points.

Step 3: Applying the SMOTE Algorithm

The Synthetic Minority Oversampling Technique (SMOTE) is employed to address class imbalances within the AQI_Bucket values. This technique generates synthetic samples for minority classes, ensuring a balanced representation of different AQI levels across the dataset. Manual iterations may be required for cities like Delhi, Bangalore, Kolkata and Hyderabad to achieve optimal balance.

Step 4: Splitting of the Dataset

The dataset is divided into training and test subsets using an 80:20 ratio. This split allows for model training using the majority of the data while reserving a portion for testing and validation purposes. Random sampling is commonly utilized for this purpose, ensuring representative subsets for both training and testing.

Step 5: Training the Dataset

Training the dataset involves using a portion of the data (typically 80%) to train machine learning models. During training, the models learn from the input data and adjust their parameters to optimize performance. Various algorithms and techniques may be employed for model training, depending on the specific objectives of the analysis.

Step 6: Testing the Dataset

The remaining portion of the dataset (typically 20%) is reserved for testing the trained models. This allows for the evaluation of model performance on unseen data, providing insights into their generalization capabilities. Similar to training, random sampling is used to ensure representative testing subsets.

Step 7: Feature Scaling

Feature scaling is performed to normalize the range of features in the dataset, making it more flexible and consistent for model training. The Standard Scaler from the Scikit-Learn library is commonly utilized for this purpose. It standardizes features by removing the mean and scaling to unit variance, ensuring that each feature contributes equally to the analysis.

Step 8: Applying Machine Learning Techniques

After feature scaling, various machine learning algorithms such as K-Nearest Neighbour, Decision Trees, Random Forest Regression, and Support Vector Regression are applied to forecast AQI levels for different cities. Each algorithm is evaluated for its accuracy and effectiveness in predicting AQI values, with comparisons made to identify the most suitable algorithm for each city.

Step 9: Applying ML Technique - Random Forest Regression

Random Forest Regression is a supervised machine learning algorithm specifically used for regression problems. It works by constructing multiple decision trees based on random subsets of the data and aggregating their predictions to produce a final output. Random Forest Regression is known for its ability to handle large datasets effectively and provide accurate predictions that are easy to interpret.

Step 10: Applying ML Technique - Support Vector Regression (SVR)

Support Vector Regression (SVR) is a supervised machine learning algorithm specifically designed for regression problems. Unlike classification tasks, where the goal is to predict discrete labels, SVR aims to predict continuous values. The core principle of SVR is to find the optimal hyperplane or line that best fits the data points in a high-dimensional space. This hyperplane is determined by maximizing the margin, or distance, between the data points and the hyperplane. The SVR algorithm seeks to minimize the error between the predicted values and the actual values while still staying within a predefined margin of tolerance. The flexibility of SVR allows practitioners to adjust the margin of tolerance, enabling them to control the trade-off between model complexity and accuracy.

Step 11: Applying ML Technique - Decision Tree

Decision tree algorithms are powerful tools for analyzing complex environmental datasets and making predictions based on a series of decision rules. In the context of air pollution analysis, decision trees are used to select relevant features or variables that influence air quality and make interpretable predictions of pollution levels. These decision rules are derived from factors such as meteorological data (e.g., temperature, humidity, wind speed) and historical pollution records. Decision trees recursively partition the dataset into subsets based on the values of different features, ultimately leading to a tree-like structure where each leaf node represents a prediction. By examining the decision path from the root node to a particular leaf, one can gain insights into the factors driving air pollution levels and make informed decisions for mitigation strategies.

Step 12: AQI Prediction

Machine learning techniques play a crucial role in predicting the Air Quality Index (AQI) by leveraging historical pollution data and relevant environmental variables. These techniques utilize trained models, such as Support Vector Regression (SVR) and Decision Trees, to make predictions of AQI values based on input features. The predicted AQI values provide valuable insights into the current and future air quality conditions, enabling stakeholders to take proactive measures to mitigate pollution and protect public health. Through rigorous analysis and model evaluation, machine learning algorithms contribute to accurate and reliable AQI predictions, facilitating informed decision-making for environmental management and policy implementation.

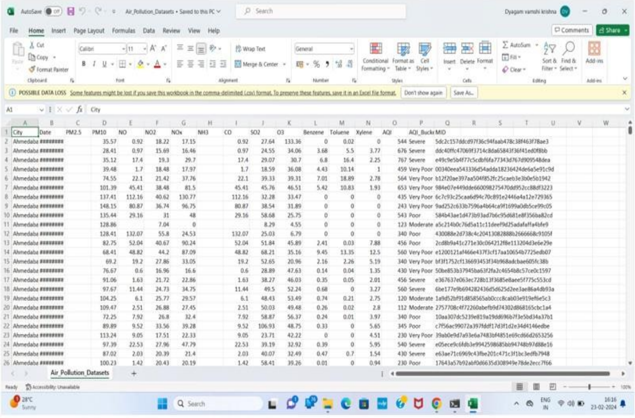
V. RESULT AND ANALYSIS

The dataset shown in the fig 1 consists of 10260 rows and 16 columns, which included all of the cities listed below. The cities are given below: Ahmedabad, Amritsar, Aizawl, Amaravati, Brajrajnagar, Bangalore, Bhopal, Chandigarh, Chennai, Delhi, Hyderabad.

The attribute information is given below.

Date YYYY-MM-DD, City, PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, AQI, and AQI_Bucket.

The AQI_Bucket variable consists of Five categories: good, satisfactory, moderate, poor, and severe, representing different levels of air quality.



POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

City

City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Buck	MID
Ahmedaba	########		35.57	0.92	18.22	17.15		0.92	27.64	133.36	0	0.02	0	544	Severe	5dc2c157ddcd97f36c94faab478c38f463f78ae3
Ahmedaba	########		28.41	0.97	15.69	16.46		0.97	24.55	34.06	3.68	5.5	3.77	676	Severe	ddc40ffc47069f3714c8da65843f36f41ed0f8bb
Ahmedaba	########		35.12	17.4	19.3	29.7		17.4	29.07	30.7	6.8	16.4	2.25	767	Severe	e49c9e5b4f77c5cdbf6fa77343d767d909548dea
Ahmedaba	########		39.48	1.7	18.48	17.97		1.7	18.59	36.08	4.43	10.14	1	459	Very Poor	00340eea543336d54adda18236424de6a5e91c9d
Ahmedaba	########		74.55	22.1	21.42	37.76		22.1	39.33	39.31	7.01	18.89	2.78	564	Very Poor	b12f20ae397aa504f852fc25caeb3e3b0e5b1942
Ahmedaba	########		101.39	45.41	38.48	81.5		45.41	45.76	46.51	5.42	10.83	1.93	653	Very Poor	984e07e449dde660098275470dd952cc88df3223
Ahmedaba	########		137.41	112.16	40.62	130.77		112.16	32.28	33.47	0	0	0	435	Very Poor	6c7c93c25caa6d94c70c891e2446a4a12e729365
Ahmedaba	########		148.15	80.87	36.74	96.75		80.87	38.54	31.89	0	0	0	243	Very Poor	9ad252c633b7596a4b64ca9f1699a0db5ce99c05
Ahmedaba	########		135.44	29.16	31	48		29.16	58.68	25.75	0	0	0	543	Poor	584b43ae1d473b93ad7b6c95d681e8f356ba82cd
Ahmedaba	########		128.86		7.04	0			8.29	4.55	0	0	0	123	Moderate	a5c214b0c76d5a11c11deef9d25adafaffa4bfe9
Ahmedaba	########		128.41	132.07	55.8	24.53		132.07	25.03	6.79	0	0	0	340	Poor	430088e2d738c4c20413082888b2666668c9105f
Ahmedaba	########		82.75	52.04	40.67	90.24		52.04	51.84	45.89	2.41	0.03	7.88	456	Poor	2cd8b9a41c271e30c064212f8e113204d3e6e29e
Ahmedaba	########		68.41	48.82	44.2	87.09		48.82	68.21	35.16	9.45	13.35	12.5	560	Very Poor	e1200121af466e437f3cf17aa10654b7725edb07
Ahmedaba	########		69.2	19.2	27.86	33.05		19.2	52.65	20.96	2.16	2.26	5.19	340	Very Poor	bf3f1752cf136693453f34b968adcbae605fc38b
Ahmedaba	########		76.67	0.6	16.96	16.6		0.6	28.89	47.63	0.14	0.04	1.35	430	Very Poor	50be853b37945ba63f2fa2c4654b8c57ce0c1597
Ahmedaba	########		91.06	1.63	21.72	22.86		1.63	38.27	46.03	0.35	0.05	2.01	456	Severe	e367637e063ec728b13f3685e8aee5f775c553cd
Ahmedaba	########		97.67	11.44	24.73	34.75		11.44	49.5	52.24	0.68	0	3.27	560	Severe	6be177e9b694282436d5d625d2ee3ae86a4db93a
Ahmedaba	########		104.25	6.1	25.77	29.57		6.1	48.43	53.49	0.74	0.21	2.75	120	Moderate	1a9d52bf91d858565ab0ccc8cab03e919ef6e5c3
Ahmedaba	########		109.47	2.51	26.88	27.45		2.51	50.03	49.48	0.26	0.02	2.8	112	Moderate	2757708c4f72260abefb9d74302d868165cbc1a4
Ahmedaba	########		72.25	7.92	26.8	32.4		7.92	58.87	56.37	0.24	0.01	3.97	340	Poor	10aa307dc5239e819a19dd696b7f3e5bd34a37b1
Ahmedaba	########		89.89	9.52	33.56	39.28		9.52	106.93	48.75	0.33	0	5.65	345	Poor	c7f56ac99072a397fddf17d3f1d2e34d4146edbe
Ahmedaba	########		113.24	9.05	17.51	22.33		9.05	23.71	42.22	0	0	4.51	230	Very Poor	39ab0e9d7a93e6a7483bf4851e69cd66d2653256
Ahmedaba	########		97.39	22.53	27.96	47.79		22.53	39.19	32.92	0.39	0	5.95	540	Severe	e05ece9c6fdb3e99425986 85bb94748b97d88e16
Ahmedaba	########		87.02	2.03	20.39	21.4		2.03	40.07	32.49	0.47	0.7	1.54	430	Severe	e63ae71c6969c43fbe201c471c3f1bc3edfb7948
Ahmedaba	########		100.23	1.42	20.43	20.19		1.42	58.41	39.26	0.01	0	0.94	230	Poor	17643a57b92abf0d6635d308949e78de2ecc7f66

Figure 1 - Dataset

Figure 1: Dataset

Service Provider Login Page

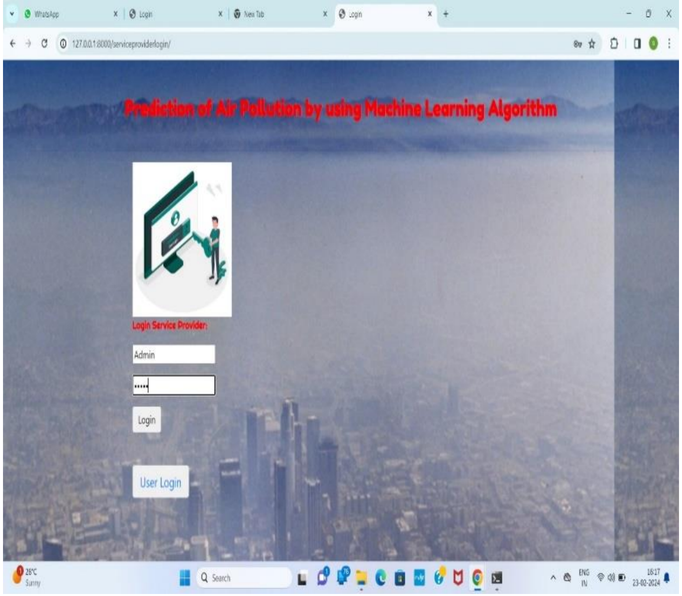
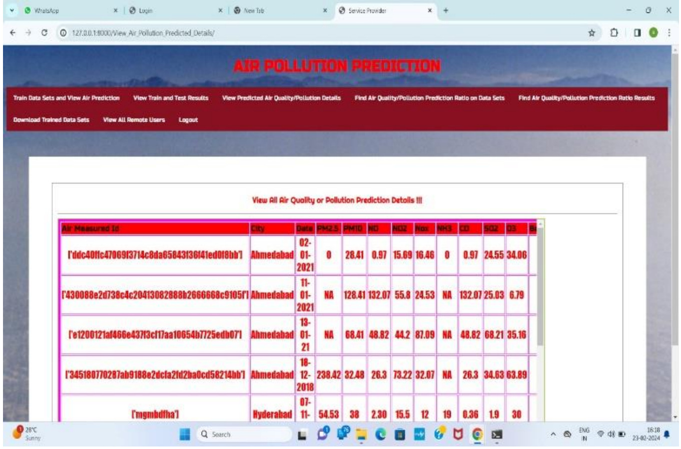


Figure 2: Login Page For The Service provider

In Figure 2 ,This interface serves as the login page for the service provider, requiring the Admin to input their Username password into the designated field. Upon completing the password entry, the user must click on the "Login" button to initiate the authentication process and gain access to the system or platform.


AIR POLLUTION PREDICTION
Train Data Sets and View Air Prediction
View Train and Test Results
View Predicted Air Quality/Pollution Details
Find Air Quality/Pollution Prediction Ratio on Data Sets
Find Air Quality/Pollution Prediction Ratio Results
Download Trained Data Sets
View All Remote Users
Logout

View All Air Quality or Pollution Prediction Details !!!

Air Measured Id	City	Date	PM2.5	PM10	NO	NO2	Nox	NH3	CO	SO2	O3
['ddc40ffc47069f3714c8da65843f36f41ed0f8bb']	Ahmedabad	02-01-2021	0	28.41	0.97	15.69	16.46	0	0.97	24.55	34.06
['430088e2d738c4c20413082888b2666668c9105f']	Ahmedabad	11-01-2021	NA	128.41	132.07	55.8	24.53	NA	132.07	25.03	6.79
['e1200121af466e437f3cf17aa10654b7725edb07']	Ahmedabad	13-01-21	NA	68.41	48.82	44.2	87.09	NA	48.82	68.21	35.16
['345180770287ab9188e2dcfa2fd2ba0cd58214bb']	Ahmedabad	18-12-2018	238.42	32.48	26.3	73.22	32.07	NA	26.3	34.63	63.89
['mgmbdfha']	Hyderabad	07-11-	54.53	38	2.30	15.5	12	19	0.36	1.9	30

Figure 3 : View Predictions of Remote Users by Service Provider

In Figure 3 , After logging in, the administrator can see all the information related to pollution predictions made by users. This includes things like past predictions, current predictions, how accurate the predictions were, and any other details about the prediction process. Basically, they can see everything about how pollution is being predicted by all users of the system.

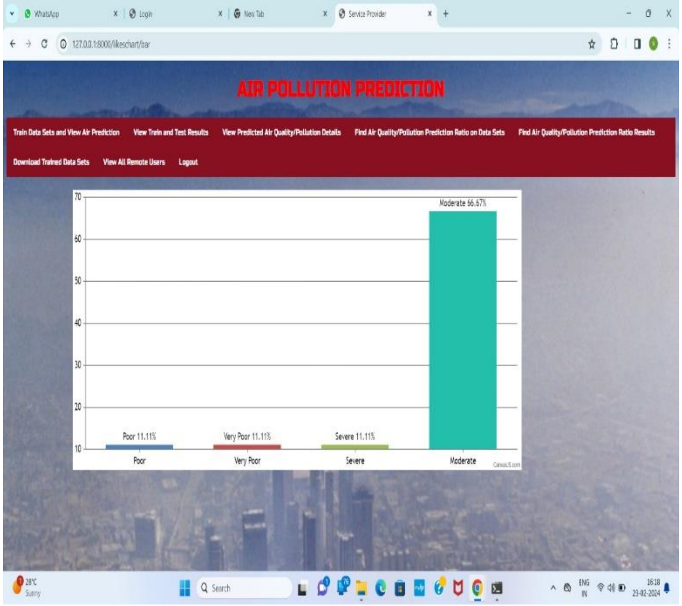
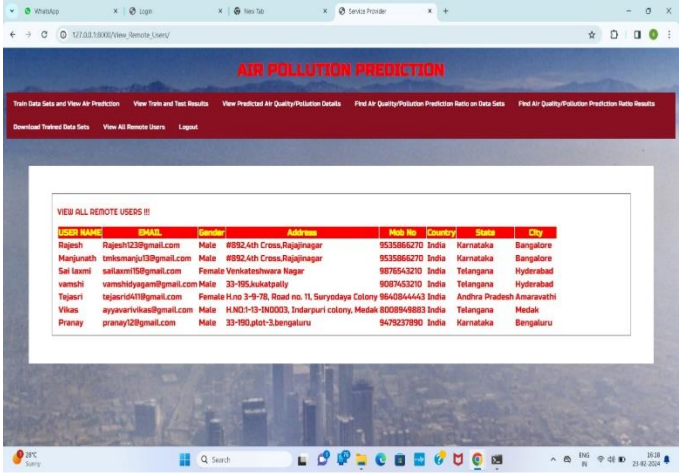


Figure 4: Air Prediction Bargraph Of Users

Figure 4 illustrates a bar graph depicting the distribution of air pollution predictions among registered users, categorized into the following levels: poor, moderate, severe, and very poor

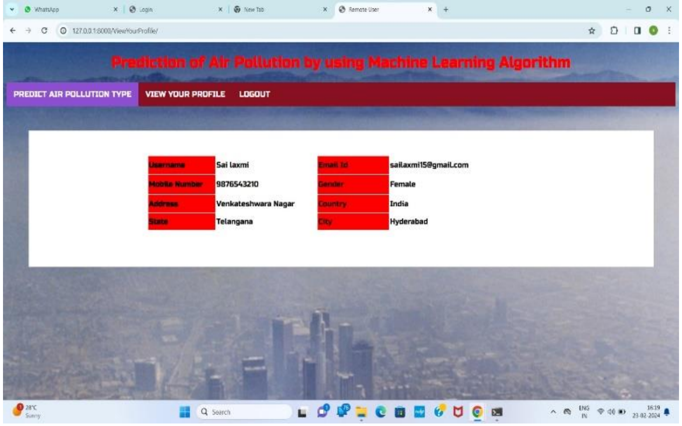

AIR POLLUTION PREDICTION
Train Data Sets and View Air Prediction
View Train and Test Results
View Predicted Air Quality/Pollution Details
Find Air Quality/Pollution Prediction Ratio on Data Sets
Find Air Quality/Pollution Prediction Ratio Results
Download Trained Data Sets
View All Remote Users
Logout

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Gender	Address	Mob No	Country	State	City
Rajesh	Rajesh123@gmail.com	Male	#892,4th Cross,Rajajinagar	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	Male	#892,4th Cross,Rajajinagar	9535866270	India	Karnataka	Bangalore
Sai laxmi	sailaxmi15@gmail.com	Female	Venkateshwara Nagar	9876543210	India	Telangana	Hyderabad
vamshi	vamshidyagam@gmail.com	Male	33-195,kukatpally	9087453210	India	Telangana	Hyderabad
Tejasri	tejasrid411@gmail.com	Female	H.no 3-9-78, Road no. 11, Suryodaya Colony	9640844443	India	Andhra Pradesh	Amaravathi
Vikas	ayyavarivikas@gmail.com	Male	H.NO:1-13-IN0003, Indarpuri colony, Medak	8008949883	India	Telangana	Medak
Pranay	pranay12@gmail.com	Male	33-190,plot-3,bengaluru	9479237890	India	Karnataka	Bengaluru

Figure 5: View All registered Remote Users

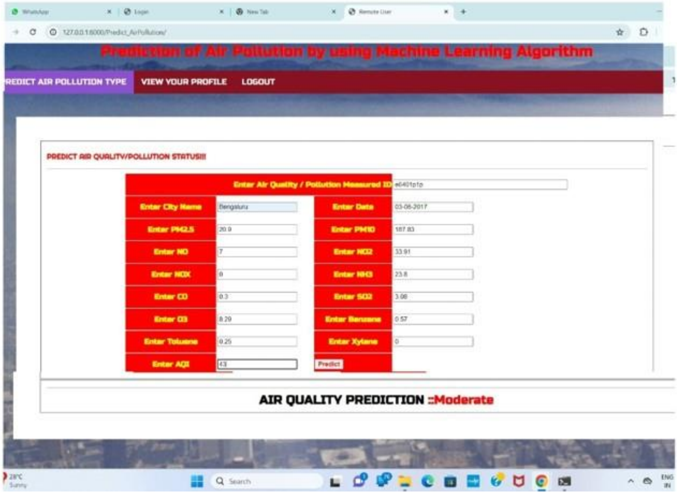
Figure 5 depicts the administrative interface designed for user management, facilitating the oversight of all registered users within the system. This interface specifically enables the administrator to access and review the details of users who are registered to utilize the air pollution prediction functionalities.


Prediction of Air Pollution by using Machine Learning Algorithm
PREDICT AIR POLLUTION TYPE
VIEW YOUR PROFILE
LOGOUT

Username	Sai laxmi	Email Id	sailaxmi15@gmail.com
Mobile Number	9876543210	Gender	Female
Address	Venkateshwara Nagar	Country	India
State	Telangana	City	Hyderabad

Figure 6: Users Profile With in their Respective Accounts

Figure 6 presents the user interface module dedicated to displaying and managing user profiles within their respective accounts.


WhatsApp
Login
New Tab
Remote User
127.0.0.1:8000/Predict_AirPollution/
Prediction of Air Pollution by using Machine Learning Algorithm
REDICT AIR POLLUTION TYPE
VIEW YOUR PROFILE
LOGOUT
Predict
28°C Sunny
Search
ENG IN

PREDICT AIR QUALITY/POLLUTION STATUS!!!

Enter Air Quality / Pollution Measured ID	e6401ptp		
Enter City Name	Bengaluru	Enter Date	03-06-2017
Enter PM2.5	20.9	Enter PM10	107.83
Enter NO	7	Enter NO2	33.91
Enter NOX	8	Enter NH3	23.8
Enter CO	0.3	Enter SO2	3.06
Enter O3	8.29	Enter Benzene	0.57
Enter Toluene	0.25	Enter Xylene	0
Enter AQI	43	Predict	

AIR QUALITY PREDICTION ::Moderate

Figure 7: Prediction Of Air Pollution

Figure 7 showcases the user interface page designed to facilitate the querying of air pollution type based on user-input values for various attributes. These attributes include Date (formatted as YYYY-MM-DD), City, and specific pollutant concentrations such as PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, and Toluene. Additionally, the user can input the Air Quality Index (AQI) and AQI Bucket.

The AQI_Bucket variable categorizes air quality into five distinct categories: good, satisfactory, moderate, poor, and severe, representing different levels of air pollution severity. Based on the provided values for these attributes, the system predicts the corresponding category, indicating the prevailing air quality condition.

VI. CONCLUSION

In conclusion, the process of predicting air quality begins with meticulous data cleaning and processing, ensuring the integrity and quality of the dataset. This is followed by exploratory analysis to gain insights into the data and identify patterns or correlations between different attributes. Subsequently, models are constructed and evaluated to predict air quality based on given attributes, with the decision tree method often demonstrating superior performance in classification tasks.The integration of proposed statistical models into deterministic air quality models can significantly enhance predictive capabilities, particularly in capturing spatial scenarios. While time series models excel in forecasting future trends, their limitations in reproducing spatial scenarios underscore the importance of integrating statistical models to provide comprehensive predictions. Research efforts in this domain primarily focus on forecasting Air Quality Index (AQI) and pollutant concentration levels to offer an accurate representation of air quality. Various machine learning algorithms, including KNN, Decision Trees, Support Vector Machine, and Logistic Regression, are widely employed for AQI prediction and pollutant concentration forecasting, each offering unique advantages depending on the nature of the data. Future studies may benefit from incorporating additional meteorological parameters and air contaminants to further enhance predictive models, providing a more holistic understanding of air quality dynamics. Leveraging real-time data analysis through cloud computing can improve model performance

by adapting to revolving data dynamics and facilitating timely decisionmaking.Furthermore, combining multiple machine learning algorithms and processing large datasets can lead to more accurate predictions, underscoring the potential for enhancing air quality forecasting through advanced analytical techniques.

VII. REFERENCES

[1] A. GnanaSoundariMtech, and J. GnanaJeslin M.E., along with Akshaya A.C., authored a research paper titled "Indian Air Quality Prediction And Analysis Using Machine Learning," which was published in the International Journal of Applied Engineering Research in 2019 (Volume 14, Number 11) https://www.researchgate.net/publication/335911816_Air_Quality_Prediction_using_Machine_Learning_Algorithms

[2] Aditya C R, Chandana R Deshmukh, Nayana D K, and Praveen Gandhi Vidyavastu authored a paper titled "Detection and Prediction of Air Pollution using Machine Learning Models," which was published in the International Journal of Engineering Trends and Technology (IJETT), Volume 59, Issue 4, in May 2018.

[3] Ni, X.Y., Huang, H., and Du, W.P., authored a paper titled "Relevance analysis and short-term prediction of PM 2.5 concentrations in Beijing based on multi-source data," which appeared in Atmos. Environ. 2017, Volume 150, Pages 146-161.

[4] V. M. Niharika and P. S. Rao authored a paper titled "A survey on air quality forecasting techniques," published in the International Journal of Computer Science and Information Technologies, Volume 5, Issue 1, Pages 103-107, in 2014.

[5] The National Ambient Air Quality Standards (NAAQS) Table, last updated in 2015, can be accessed online at: https://www.epa.gov/criteria-air-pollutants/naaqs-table

[6] E. Kalapanidas and N. Avouris presented a paper titled "Applying machine learning techniques in air quality prediction" at the Proceedings of the Artificial Intelligence

Applications and Innovations conference (ACAI), Volume 99, in September 2017.

[7] Suhasini V. Kottur and Dr. S. S. Mantha presented a paper titled "An Integrated Model Using Artificial Neural Network and Kriging For Forecasting Air Pollutants Using Meteorological Data" in the International Journal of Advanced Research in Computer and Communication Engineering, Volume 4, Issue 1, January 2015.

[8] Ruchi Raturi and Dr. J.R. Prasad contributed a paper titled "Recognition Of Future Air Quality Index Using Artificial Neural Network" published in the International Research Journal of Engineering and Technology (IRJET), Volume 05, Issue 03, March 2018.

[9] Tom M. Mitchell (1997). Machine Learning. McGraw-Hill International Editions.

[10] Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. The MIT Press.

[11] Marsland, S. (2015). Machine Learning: An Algorithmic Perspective (2nd ed.). Chapman and Hall/CRC.