Predictive Urban Air Quality Monitoring for Healthier Cities

Gresha Bhatia

Department of Computer Engineering, VESIT, Mumbai, Maharashtra, India, gresha.bhatia@ves.ac.in

KEYWORDS
Air quality, public health, pollution levels, SMOTE, Binning , predictive modeling.

ABSTRACT
The research work proposal outlines an innovative and crucial project for addressing urban air quality issues using advanced technological solutions. Establishing a real-time air quality monitoring system with integrated big data and machine learning will enable precise tracking and analysis of pollution levels. The research paper begins with the key aspects of air quality. It further focuses upon the need for utilizing advanced techniques to curb the issues due to poor air quality. A system is proposed to determine the air quality utilizing machine learning techniques, predicting the quality of air index. Evaluation measures are then undertaken for predicting the same using Random forest, Support vector and Catboost regression techniques.

1. Introduction

The air quality we intake forms one of the critical component that determines the health of all across the globe as well as environmental well-being,. This is particularly evident in urban areas where industrial activities, vehicular emissions, and other factors lead to elevated pollution levels. As cities continue to grow and develop, the need for effective air quality monitoring and management becomes increasingly urgent. Alarming effects over the air pollution are driving the demand for its efficient monitoring and management. Air quality control , itself, refers to the processes and regulations aimed at maintaining or improving the quality of air in a given environment. This involves reducing pollutants in the air, such as particulate matter (PM), volatile organic compounds (VOCs), sulphur dioxide (SO_2), nitrogen oxides (NO_x), carbon monoxide (CO), and ozone (O_3), which can affect human health and the environment. Thus, effective air quality control is crucial for mitigating environmental impacts , climate change and damage to ecosystems. The right amount of air quality is also needed to control human health. Exposure to poor air quality may contribute to breathing problems, cardiovascular conditions, and various other health concerns. This highlights the importance of accurate monitoring of the air quality in the environment. The advent of advanced technologies, such as big data analytics and machine learning, offers a transformative opportunity to address the challenge of monitoring air quality. This research paper centers around the development of a real-time air quality monitoring system tailored for urban environments focussing on integrating the cutting-edge technologies such as machine learning on improving the Air Quality Index (AQI). The next section specifies the key aspects of air control. This is then followed by the literature survey, proposed methodology. Results and evaluations are further elaborated. This is followed by the conclusion and future scope.

2. Key Aspects of Air Quality Control

More the value of AQI, the higher the level of air pollution. The effect of this would affect the health of a person drastically. Thus, controlling the quality of air has various dimensions to it as elaborated below .

Air Quality Monitoring: This implies the air pollution levels are tracked through the use of sensors and monitoring stations. With the help of this, areas with poor air quality as well as the sources of pollution can be determined .

Emission Control: This area focuses on building the Regulations and technologies in order to limit the release of harmful pollutants from industrial sources, vehicles, power plants, and other emission points. Common approaches include using scrubbers, filters, catalytic converters, and alternative fuels.

Air Quality Standards: This aspect refers to the legal limits set by the Government agencies on the amount of certain pollutants in the air, based on health risk assessments.

Public Health Policies: Depending on the factors of AQI and its impact on the health of people, policies or restrictions could be applied on certain activities during high-pollution periods.

Urban Planning and Green Spaces: Government and other agencies develop strategies to increase green spaces and reduce urban sprawl in order to reduce pollution levels.

Indoor Air Quality: In homes and buildings, air quality control involves managing ventilation, controlling humidity, and reducing indoor pollutants like mold, dust, and chemicals from cleaning products or building materials.

3. Literature Survey

The quality of the air we breathe affects the health of the people that are habitats particularly in urban areas where industrial activities, vehicular emissions, and other factors can lead to elevated pollution levels. As cities continue to grow and develop, the need for effective air quality monitoring and management becomes increasingly urgent. application of Machine learning methods for anticipating air quality issues in urban regions and its impact has been researched through a literature survey . This survey consisted of referring to a number of research papers and newspaper articles.

a. Research papers :-

Research paper 1, focussed on measuring the impact of air pollution on individual health through air quality index (AQI) parameters. This was done by utilizing data mining techniques on cities like New Delhi, Bangalore, Kolkata, and Hyderabad. Soubhik Mahanta et al in their paper performed regression analysis to determine the air quality by using meteorological as well as pollution related data across cities [2]. N. Srinivasa et al worked to determine the AQI predictions through Support Vector Regression (SVR), Random Forest Regression (RFR), and CatBoost Regression (CR) on the cities mentioned earlier. They further optimized the values using SMOTE.[3]. In paper [4], Hongna He1 et al mentioned that the prediction is a problem that has multiple variables, is nonlinear in nature and forms a time-series problem. This is further applied to a circular neural network consisting of Long Short Time Memory (LSTM) to predict Air Quality Index (AQI). K. Kumar et al in paper [5], investigated the use of machine learning techniques on six years of air pollution data from 23 Indian cities. Manuel Méndez et al, in their paper evaluated a deep learning model for determining AQI. They further considered [6] geographical distribution, predicted values, predictor variables for determining the AQI. In paper [7], the researchers predicted air pollution using K Means algorithm, Multinomial Logistic Regression and Decision tree using R programming. In paper [8], parameters such as concentration of nitrogen oxide in the air and its exposure were used to test the potential cause of numerous respiratory, circulatory, nervous diseases and environmental pollution. Shengdong Du et al [9] elaborated on the use of 1-D Convolutional Neural Networks (CNNs) and Bi-directional Long Short-term Memory (Bi-LSTM) networks are applied to identify local trend features and spatial correlations in the nonlinear, dynamic characteristics of multivariate air quality data. In paper [10] , air quality forecasting models as well as real time monitoring tools and techniques based on historical data were elaborated. A linear regression algorithm was applied to predict air quality, utilizing AQI data generated through IoT setups. The model's performance was evaluated using MAE, MSE, RMSE, and MAPE metrics. In paper [11] were determined to assess the model's performance . A back propagation neural network (BPN) algorithm was considered to determine the quality of air quality in the spatial dimension mentioned by the authors in paper [12]. Weather stations and environment monitoring data was used to predict the AQI through ML-based neural network technique [13]. Anticipation of normal air quality levels were done using computational insight techniques and its root mean square error (RSME) was determined as mentioned in paper [14]. In the study [15], the authors elaborated on using machine-learning techniques considering input variables such as pressure, temperature, initial gas to condensate ratio. To add further, multilayer perceptron machines(LS SVM) along with least squares support vectors were used to evaluate the predictions made. O. Hasbeh et al in their paper [16], elaborated on the multi layer perceptron method (MLP) to determine the accuracy of prediction. Error terms RMSE and R-SQUARE, were further evaluated for the test dataset considered.

b. Newspaper articles :-

1. Blanket of Smog covers Delhi, Air quality remains "severe", Published on 8th Nov 2024, NDTV.

A thick layer of smog engulfed Delhi city . It further indicated that the air quality index (AQI) had moved into the severe category as seen in figure 1.

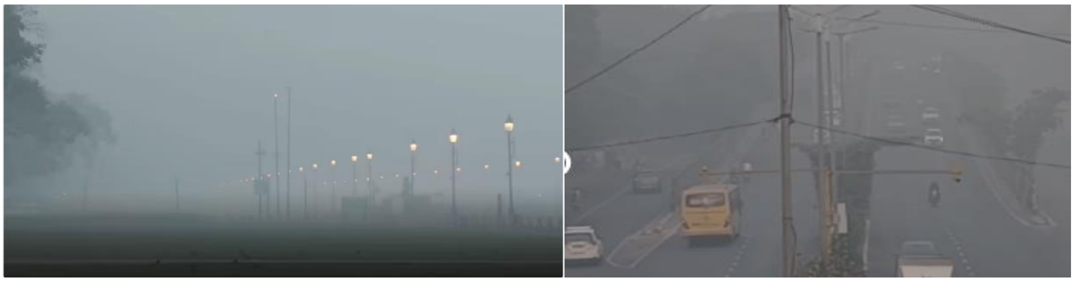

Figure 1 :- Smog in NewDelhi

Source :- https://www.ndtv.com/video/delhi-pollution-blanket-of-smog-covers-delhi-air-quality-remains-severe-857506

2. Delhi's air quality stays 'poor' for third straight day, some areas hit 'severe' levels, Economic Times, 16th Oct 24

Delhi's air quality was identified into 'poor' category and then moved to the 'very poor' and 'severe' categories as identified by some monitoring stations indicated in figure 2

Figure 2 :-Poor air quality index

Source :- https://economictimes.indiatimes.com/news/india/delhis-air-quality-stays-poor-for-third-straight-day-some-areas-hit-severe-levels/articleshow/114291459.cms?from=mdr

A review of the literature indicates a need for applying machine learning techniques to predict the Air Quality Index (AQI) as part of efforts to tackle global air pollution. Methods like Support Vector Regression, Random Forest Regressor, CatBoost Regressor, and LSTM neural networks must be fine-tuned and implemented to accurately forecast AQI. The papers also provide a comprehensive overview of state-of-the-art machine learning techniques for air quality forecasting, aiding researchers and practitioners in understanding the strengths and limitations of different models. This research aims to provide insights into the dynamics of air pollution, facilitating the design of effective mitigation policies and strategies, and contributing to the improvement of air quality and overall well-being of the population.

4. Challenges Faced in Determining Air Quality

Determining air quality involved various challenges namely:-

1. Complexity of Pollutants

The primary pollutants contributing to air pollution include particulate matter (PM), nitrogen oxides (NOx), sulfur dioxide (SO2), carbon monoxide (CO), volatile organic compounds (VOCs), and ozone (O3). Each pollutant has different sources, behaviors, and health impacts. Monitoring all pollutants simultaneously requires sophisticated and varied technologies.

2. Spatial Variability

Air quality can vary significantly over short distances due to factors like traffic, industrial zones, and geographic features. Fixed monitoring stations may not capture these variations, and deploying enough sensors to monitor small-scale differences is often expensive and logistically challenging.

3. Temporal Variability

Air quality fluctuates over time due to daily human activities (e.g., traffic peaks), seasonal changes, and weather conditions. To obtain accurate data, continuous monitoring is necessary, but maintaining and operating equipment 24/7 can be resource-intensive.

4. Cost of Equipment and Maintenance

High-precision air quality monitoring equipment, such as continuous emission monitors and sophisticated sensors, are expensive to purchase, install, and maintain. Smaller, lower-cost sensors are more accessible but may lack accuracy and reliability compared to larger, more sophisticated systems. 5. 5. Meteorological Factors

Weather conditions like wind speed, humidity, and temperature play a significant role in pollutant dispersion and concentration. High winds can disperse pollutants, while temperature inversions can trap them close to the ground. Accounting for these factors requires integrated meteorological data, making analysis more complex.

6. Data Interpretation

The sheer volume of data generated by modern air quality monitoring systems can be overwhelming. Turning raw data into meaningful information for policy makers and the public requires advanced data processing techniques, statistical modeling, and sometimes artificial intelligence (AI).

7. Indoor Air Quality

A significant amount of human exposure to air pollution occurs indoors, where monitoring is even more complex. Different sources of pollution (e.g., cooking, heating, building materials) exist indoors, and measuring this alongside outdoor air pollution is a significant challenge.

8. Global and Local Disparities

In developing countries or rural areas, monitoring infrastructure may be lacking or non-existent. As a result, there is often a gap in data from less populated or economically disadvantaged regions, leading to uneven air quality assessments on a global scale

9. Regulatory Differences

Different countries and regions have varying air quality standards and thresholds. This can make cross-border comparisons difficult and complicate global efforts to address air pollution comprehensively

10. Public Awareness and Engagement

Even with available air quality data, engaging the public and ensuring understanding of risks is a challenge. Without clear communication, people may not alter their behavior or advocate for improved policies.

Overcoming these challenges requires investment in technology, infrastructure, and education, as well as collaboration between governments, industries, and communities.

5. Proposed Work

This paper aims to propose an air quality monitoring system for urban areas by harnessing the power of big data from air quality sensors and applying cutting-edge machine learning algorithms. By continuously collecting and analyzing air quality data, the system aims to provide up-to-the-minute updates on pollution levels, enhancing proactive measures, public awareness and enabling residents and authorities to make informed decisions. The various modules comprise of as mentioned in figure 3:

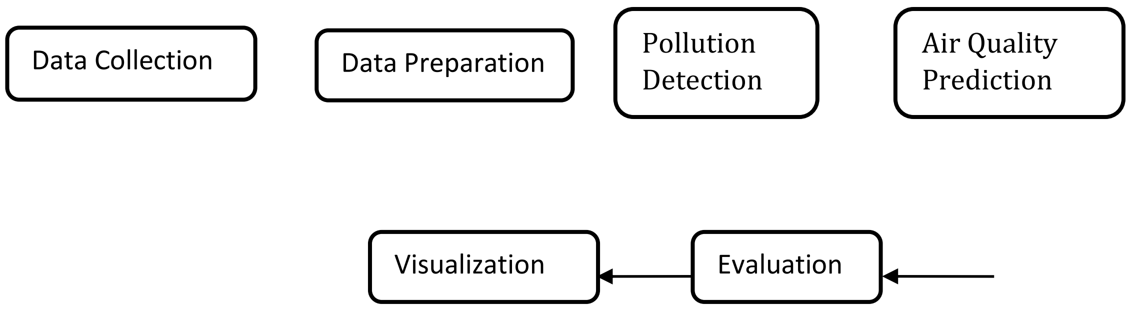

Figure 3 : Block representation of AQI system

Data Collection :-The key aspects in this system include collecting data from various sources such as the meteorological department , satellite images, weather station data, the state whose quality check needs to be done, temperature , humidity , pollutant concentrations in each area.

Data preparation :- Once the data is captured, it has to be brought to a standard format, missing values has to be handled through interpolation or deletion techniques, outlier detection and removal as well as feature extraction and scaling.

Pollution detection :- Based on the pollutants (impurities) in the air , the quality if air is determined. This is done by comparing it with the threshold values set by the meteorological department. The output thus obtained is classified into various levels namely "Good," "Moderate," or "Unhealthy for Sensitive Groups."

Prediction:- This is the core of the system, utilizing machine learning models to predict future air quality levels. It comprises of :

i. Model Training: Historical air quality data, along with temperature, wind speed, humidity., traffic or industrial emissions are used to train the various models considered.

ii. Model Selection: Machine learning models such as Linear Regression, Support Vector Regression (SVR), Random Forest and Deep Learning Models using LSTM, back propagation neural network are evaluated to determine the best model to capture the relationship between pollutants ,determining temporal dependencies and making robust, accurate predictions.

Visualization :- Comprehensive report and visualization techniques are then incorporated to to determine the current AQI levels, pollutant concentrations over a period of time.

6. Experimental Setup

Inorder to take timely decisions and planning interventions, accurate air pollution forecasts are essential. These proactive actions are expected to significantly improve both public health and environmental conditions, especially in urban environments. The experimental setup was carried out for cities such as Kolkata, New Delhi, Hyderabad, and Bangalore, with analysis done for each. The machine learning module requires a minimum configuration of an x64 machine with 8GB of RAM, a 2.3 GHz Intel i7 processor, at least 512 MB of physical storage, and the use of standard visualization and sklearn libraries. Pandas DataFrame was used for handling missing data, and VS Code was chosen as the IDE. The training process was carried out with a 7:3 ratio, consisting of 4400 training values and 2000 test values. Linear Regression, Neural Network Regression, Decision Forest were then applied on the considered dataset .

Data input having features such as PM2.5 , PM10, oxides of the Nitrogen are considered as input with respect to the area under consideration.

In order to determine the AQI levels, various algorithms were implemented. At the initial step, the data was preprocessed. To resolve the issue of class imbalance in the dataset, SMOTE and Binning strategies were implemented.

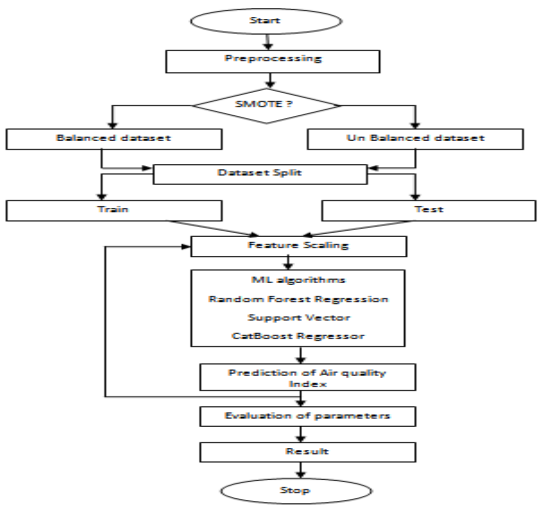

Figure 4 : - Proposed algorithms

Various algorithms as mentioned in figure 4 were then worked upon as mentioned below. Random Forest Regression, an ensemble learning technique was utilized to identify the prominent factors to determine the AQI. A number of decision trees were developed during the training phase. Pruning on them were further done in order to provide outcomes that are more accurate . Hyper parameter tuning was then performed to improve the model's accuracy. The model was further tested on various evaluation parameters like Mean Squared Error (MSE) and R-square error(RMSE). A trained Random Forest Regressor is employed to forecast AQI values based on current environmental factors, capturing air quality changes. The Support Vector Regressor (SVR) algorithm proves to be the most suitable for regression tasks due to its ability to handle non-linear relationships, fit data within defined tolerance margins, and adjust to diverse data distributions. This makes SVR a powerful tool for precise AQI monitoring, aiding effective air quality management in urban areas. The CatBoost regression model incorporates parameters specific to categorical values. By training them on relevant datasets, accurate predictions of AQI values could be obtained.

Data from various sources and Kaggle for New Delhi, Kolkata, Bengaluru, and Hyderabad cities were systematically gathered , analyzed and visualized to present the amount of pollutants in an intuitive manner through Tableau. Utilizing Air Quality Programmatic API, dynamically retrieves real-time pollution data from the cities mentioned above. Based on the city considered, the website fetches the current AQI readings along with additional pollution-related metrics.

7. Results

Based on the dataset of various cities, on a respective date, the AQI obtained as shown in figure 5 gives an imbalanced set. On applying pre-processing SMOTE algorithm balanced class distribution is obtained as represented in figure 6.

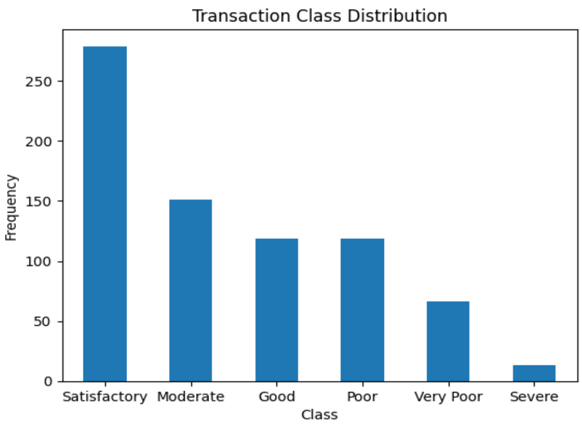

Figure 5 : Imbalanced class distribution

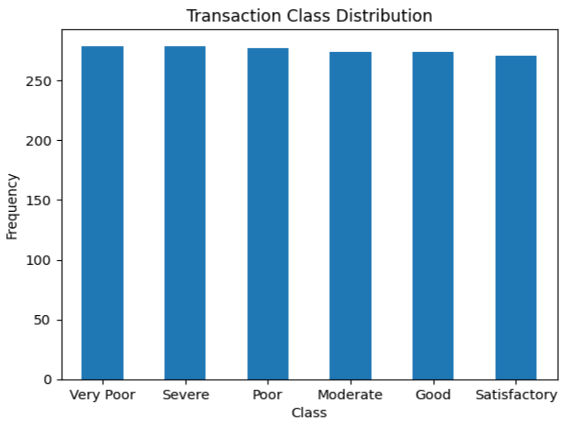

Figure 6: Balanced class distribution using SMOTE

The dataset comprises diverse pollution indicators, including nitrogen oxides (NO, NO2) and sulphur compounds (SO, SO2), PM2, PM10,Benzene, Toluene were considered. Each entry in the dataset included the AQI value for a particular day, categorized into distinct AQI Buckets such as Poor, Satisfactory, Moderate or Good.

The AQI parameter for Kolkata city was obtained to be very poor. This was due to the higher amount of pollutants namely Parameterized matter 2.5, PM 10, NO, benzene and Toluene as indicated by table 1,2,3 below

Table 1:- Determination of AQI and its bucket for the given set of parameters.

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	AQI	AQI_Bucket
0	Kolkata	2018-11-13	183.674773	294.994310	111.391660	60.165020	164.892990	23.309276	1.625968	15.104298	29.741835	14.063900	25.738073	332.007664	Very Poor
1	Kolkata	2018-11-13	182.602597	295.346357	111.188346	60.298546	166.367512	23.441374	1.635240	15.144647	29.407338	13.963749	25.948123	331.693104	Very Poor
2	Kolkata	2018-11-13	182.977015	294.709747	110.994541	60.950291	166.052102	23.712466	1.614929	15.191690	29.956972	13.690267	26.299087	330.320870	Very Poor
3	Kolkata	2018-11-13	183.455779	293.099319	110.202293	59.214776	165.778791	23.810407	1.618048	15.097887	29.752792	14.142652	26.878472	331.488682	Very Poor
4	Kolkata	2018-11-13	181.743876	294.624160	110.487551	59.784879	164.875157	23.330946	1.643663	15.179412	29.887768	13.884121	25.986428	331.263962	Very Poor

Over a number of years, the air quality over the region of Kolkata improved as represented by table 2.

Table 2:- Determination of AQI and its bucket for the given set of parameters for the year 2020

742	Kolkata	2020-06-27	7.890000	24.730000	5.240000	9.250000	14.520000	8.390000	0.350000	5.730000	23.140000	1.680000	11.310000	37.000000	Good
743	Kolkata	2020-06-28	10.580000	25.560000	5.940000	12.630000	18.620000	6.920000	0.380000	5.920000	27.760000	1.320000	11.070000	43.000000	Good
744	Kolkata	2020-06-29	14.530000	32.400000	5.420000	15.980000	21.450000	7.590000	0.450000	7.010000	30.640000	3.180000	9.780000	48.000000	Good
745	Kolkata	2020-06-30	14.000000	35.850000	6.250000	12.290000	18.120000	9.640000	0.370000	5.660000	24.590000	2.460000	10.750000	47.000000	Good
746	Kolkata	2020-07-01	10.530000	31.580000	6.250000	10.010000	16.290000	12.870000	0.300000	4.320000	15.520000	1.820000	10.250000	45.000000	Good

Table 3:- Determination of AQI and its bucket for the given set of parameters for New Delhi

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	AQI	AQI_Bucket
10229	Delhi	########	313.22	607.98	69.16	36.39	110.59	33.85	15.2	9.25	41.68	14.36	24.86	472	Severe
10230	Delhi	########	186.18	269.55	62.09	32.87	88.14	31.83	9.54	6.65	29.97	10.55	20.09	454	Severe
10231	Delhi	########	87.18	131.9	25.73	30.31	47.95	69.55	10.61	2.65	19.71	3.91	10.23	143	Moderate
10232	Delhi	########	151.84	241.84	25.01	36.91	48.62	130.36	11.54	4.63	25.36	4.26	9.71	319	Very Poor
10233	Delhi	########	146.6	219.13	14.01	34.92	38.25	122.88	9.2	3.33	23.2	2.8	6.21	325	Very Poor
10234	Delhi	########	149.58	252.1	17.21	37.84	42.46	134.97	9.44	3.66	26.83	3.63	7.35	318	Very Poor
10235	Delhi	########	217.87	376.51	26.99	40.15	52.41	134.82	9.78	5.82	28.96	4.93	9.42	353	Very Poor
10236	Delhi	########	229.9	360.95	23.34	43.16	51.21	138.13	11.01	3.31	30.51	5.8	11.4	383	Very Poor
10237	Delhi	########	201.66	397.43	19.18	38.56	45.6	140.6	11.09	3.48	32.94	5.25	11.12	375	Very Poor
10238	Delhi	########	221.02	361.74	24.79	46.39	55.19	134.06	9.7	5.91	34.12	4.87	9.44	376	Very Poor
10239	Delhi	########	205.41	393.2	28.46	47.29	57.88	131.1	10.98	5.54	50.37	5.93	10.59	379	Very Poor
10240	Delhi	########	212.41	345.63	24.77	44.71	54.66	148.51	9.3	5.17	40.08	6.2	10.68	375	Very Poor
10241	Delhi	########	197.61	301.04	34.28	45.88	62.95	145.48	11.52	5.95	39.54	5.18	9.5	366	Very Poor
10242	Delhi	########	164.39	227.38	20.78	41.17	48.09	166.7	10.32	4.85	42.05	4.57	10.72	353	Very Poor
10243	Delhi	########	166.19	283.93	47.91	56.7	79.73	124.64	10.81	6.94	43.25	6.98	20.55	340	Very Poor
10244	Delhi	########	174.98	309.99	42.33	49.41	70.6	136.4	11.46	4.69	34.7	7.75	15.83	356	Very Poor
10245	Delhi	########	196.46	356.2	42.18	53.04	80.18	125.49	8.87	5.13	25.77	9.92	17.59	360	Very Poor
10246	Delhi	########	201.51	359.75	45.51	53.77	75.83	120.45	8.99	5.43	23.22	9	20.95	370	Very Poor
10247	Delhi	########	183.35	305.13	27.32	39.88	55.67	148.49	9.02	4.21	25.34	5.57	12.93	362	Very Poor
10248	Delhi	########	165.63	257.04	16.47	34.23	37.68	146.31	8.03	4.11	22.69	3.42	6.76	340	Very Poor
10249	Delhi	########	159.54	235.27	12.27	22.56	33.02	63.23	9.01	7.05	18.56	4.1	8.26	338	Very Poor
10250	Delhi	########	143.68	183.89	14.75	21.82	34.88	39.65	10.58	8.64	6.94	4.15	9.54	332	Very Poor
10251	Delhi	########	128.73	136.07	11.99	17.35	29.46	39.98	8.74	6.77	10.91	2.51	7.64	254	Poor
10252	Delhi	########	164.98	200.02	10.38	15.77	25.49	38.77	9.39	8.89	9.35	2.6	5.5	324	Very Poor
10253	Delhi	########	148.59	203.81	11.82	20.83	28.29	45.62	11.44	5.96	16.65	3.37	6.98	333	Very Poor
10254	Delhi	########	129.34	137.86	12.46	25.66	29.48	60.96	10.19	6.78	18.18	3	5.87	292	Poor
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

8. Evaluation of the System

The proposed system is evaluated by comparing the various algorithms used to predict the AQI for the various cities. Table 4, 5 below represents the evaluation measures .

Table 4 : Evaluation Measures for New Delhi

Evaluation Measure	Random Forest	Support Vector Regressor	CatBoost Regressor
R2 Score	0.9832345	0.855926	0.983394
RMSE	22.337685	65.45278	22.22828
MAE	11.915853	50.20274	12.98784
MSE	498.52552	4284.066	494.0945

Table 5 : Evaluation Measures for Kolkata

Evaluation Measure	Random Forest	Support Vector Regressor	CatBoost Regressor
R2 Score	0.9943719	0.923445	0.9919151
RMSE	10.315504	38.04498	12.363749
MAE	5.1162061	29.26691	6.8284592
MSE	106.40962	1447.420	152.86230

From the above table, it is observed that CatBoost regressor has a better output compared to other algorithms.

9. Conclusion

Leveraging the power of CatBoost on our AQI monitoring website opens doors to exciting possibilities beyond basic prediction. By integrating predicted AQI values with weather forecasts and historical data, we can create a sophisticated air quality information hub. Users won't just see a single number - they'll have access to a dynamic picture of how air quality might evolve throughout the day, week, or even longer. Imagine being able to plan outdoor exercise routines around predicted good air quality days or proactively reschedule picnics based on potential spikes in pollution. Furthermore, the website can become an interactive platform. Real-time sensor data can be fed back into the CatBoost model, creating a self-learning loop. This allows the model to constantly adapt to changing environmental conditions, ensuring the predictions remain razor-sharp. Ultimately, this user-centric approach empowers individuals to make informed health decisions. By providing a clear understanding of air quality trends, the website becomes a valuable tool for safeguarding public health and promoting a more proactive approach to well-being.

10. Future Scope

Improved Data Integration: Integrating air quality data with other environmental datasets (e.g., weather, traffic) will lead to a more comprehensive understanding of pollution dynamics.

Machine Learning Applications: Machine learning can further enhance air quality forecasting, allowing for more accurate predictions and earlier warnings.

Real-time Decision Support Systems: Combining air quality data with real-time traffic management, energy production, and industrial emission control systems could lead to dynamic pollution mitigation strategies.

References

[1] Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative AnalysisN. Srinivasa Gupta,1Yashvi Mohta,2Khyati Heda,2Raahil Armaan,2B. Valarmathi,2and G. Arulkumaran, Hindawi Journal of Environmental and Public Health, Volume 2023, Article ID 4916267

[2] A time series forecasting based multi-criteria methodology for air quality prediction, Raquel Espinosa, José Palma, Fernando Jiménez , Joanna Kamińska , Guido Sciavicco, Estrella Lucena-Sánchez.

[3] Deep Air Quality Forecasting Using Hybrid Deep Learning Framework, IEEE Transactions on Knowledge and Data Engineering (Volume: 33, Issue: 6, 01 June 2021)

[4] Urban Air Quality Prediction Using Regression Analysis, Soubhik Mahanta; T. Ramakrishnudu; Rajat Raj Jha; Niraj Tailor, TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)

[5] Air Quality Prediction using Machine Learning Algorithms –A Review, Tanisha Madan; Shrddha Sagar; Deepali Virmani, 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)

[6] A. G. Soundari, J. Gnana, and A. C. Akshaya, “Indian air quality prediction and analysis using machine learning,” International Journal of Applied Engineering Research, vol. 14, p. 11, 2019.

[7] H. Liu, Q. Li, D. Yu, and Y. Gu, “Air quality index and air pollutant concentration prediction based on machine learning algorithms,” Applied Sciences, vol. 9, p. 4069, 2019.

[8] M. Bansal, “Air quality index prediction of Delhi using LSTM,” Int. J. Emerg. Trends Technol. Comput. Sci, vol. 8, pp. 59–68, 2019.

[9] G. Mani, J. K. Viswanadhapalli, and A. A. Stonie, “Prediction and forecasting of air quality index in Chennai using regression and ARIMA time series models,” Journal of Engineering.

[10] N. Srinivasa Gupta , Yashvi Mohta, Khyati Heda, Raahil Armaan, Valarmathi and G. Arulkumaran, “Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis” Journal of Environmental and Public Health, Hindawi Journal of Environmental and Public Health Volume 2023, Article ID 4916267, 26 pages https://doi.org/10.1155/2023/4916267

[11] R. Kumar, P. Kumar, and Y. Kumar, “Time series data prediction using IoT and machine learning technique,”Procedia Computer Science, vol. 167, no. 2020, pp. 373–381,2020

[12] X. Zhao, M. Song, A. Liu, Y. Wang, T. Wang, and J Cao,“Data-Driven temporal-spatial model for the prediction ofAQI in nanjing,” Journal of Artificial Intelligence and SoftComputing Research, vol. 10, no. 4, pp. 255–270, 2020.

[13] Zhou, S. De, G. Ewa, C. Perera, and K. Moessner, “Data-Driven air quality characterization for urban environments: a case study,” IEEE Access, vol. 6, Article ID 77996, 2018.

[14] Hajek and V. Olej, “Predicting common air quality index -the case of Czech microregions,” Aerosol and Air QualityResearch, vol. 15, no. 2, pp. 544–555, 2015

[15] A. R. Behesht Abad, S. Mousavi, N. Mohamadian et al.,“Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs,” Journal of Natural Gas Science and Engineering,vol. 95, Article ID 104210, 2021

[16] O. Hasbeh, M. Ahmadi Alvar, K. Y. Aghdam, H. Ghorbani,N. Mohamadian, and J. Moghadasi, “Hybrid computing models to predict oil formation volume factor using multi-layer perceptron algorithm,” Journal of Petroleum and MiningEngineering, vol. 23, no. 1, pp. 17–30, 2021