JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AIR QUALITY INDEX ANALYZER USING MACHINE LEARNING

Improving Air Quality Insights using Machine Learning

[1]Shazia Tazeen, [2]Samala Harshith Kumar, [3]Edla Yeshwanth
[1]Assistant Professor, [2]Student, [3]Student
[1]CSE Department,
[1]Vaagdevi Engineering College, Warangal, India

Abstract: One of the most important instruments for evaluating air quality based on pollutant concentrations is the Air Quality Index (AQI). This approach divides the air quality into six different categories, from "Good" to "Severe," each of which has a range of AQI values, accompanying symptoms, possible health risks, and suggested preventative measures. "Severe" denotes extremely dangerous pollution with major health hazards, whereas "Good" denotes great air quality with no symptoms at all. Knowing the AQI gives people the power to choose outdoor activities wisely and take suitable precautions to protect their health.

This system is an important resource for increasing awareness of and reducing the harmful effects of air pollution on human health. Additionally, the data is essentially split into the North and South regions. The random forest regressor demonstrated a great predictive performance with an accuracy score of 0.89 in the North, as seen by its Mean Squared Error of 1436.99 and R-squared value of 0.89. The random forest regressor demonstrated an accuracy score of 0.91 in the South, with a Mean Squared Error of 3020.29 and an R-squared value of 0.87.

Index Terms- **Air pollution, AQI, linear regression, prediction, and random forest regressor**

I. Introduction

A key and well-known instrument in the fields of environmental science and public health is the Air Quality Index, or AQI for short. In determining and conveying the quality of the air we breathe, it is essential. The Air Quality Index (AQI) is a thorough and user-friendly method of assessing the safety and cleanliness of the air in a specific area, providing important information about the possible health concerns related to air pollution.

The quality of air is a very vital issue which impacts not only with the human health but also the environment. The AQI is used to calculate the performances between the general people's everyday life and tangled scientific information on air pollution. This aids people, communities, and politicians make educated decisions regarding outdoor activities, health precautions, and pollution control techniques by reducing a wide range of pollutant measures into a single, understandable number or category.

The comprehensive tool known as Air Quality Index is basically used for the supervision of the quality of the air. This tool takes wide range of elements into an account such as gases, vapors present in water,vaporizers,atomizers and many more. The obscurity of air pollution which trunks from both natural and artificial (manmade)activities highlights the complications which results in reducing and controlling it. Controlling of air pollution plays a vital role which involves many parties with a variety of goals and interests.Not only that it is a place where science and public policy converge also plays a vital role.

Using formal consensus-building techniques becomes essential in this situation to guarantee transparent and repeatable decision-making procedures. Using a case study as an example, this paper explores different approaches to calculate the Air Quality Index (AQI) to evaluate the effects of air pollution. The Air Quality Index (AQI) is a crucial instrument for locating areas with high levels of air pollution and developing efficient control plans. It has been used extensively for many years in India and throughout the world.

A number of the current AQI systems account for synergistic effects by computing the mean ratios of pollutants over guideline levels for a specified time period, whereas earlier versions of the AQI primarily relied on exceeding health-based compliance standards set for a nation's ambient air over a given time frame. Different mean values, such as geometric, arithmetic, weighted, logarithmic, and breakpoint concentration, are used to categorize these AQIs. Simple Air Quality Index is basically castoff measure to assess the quality of air area wise and state wise.

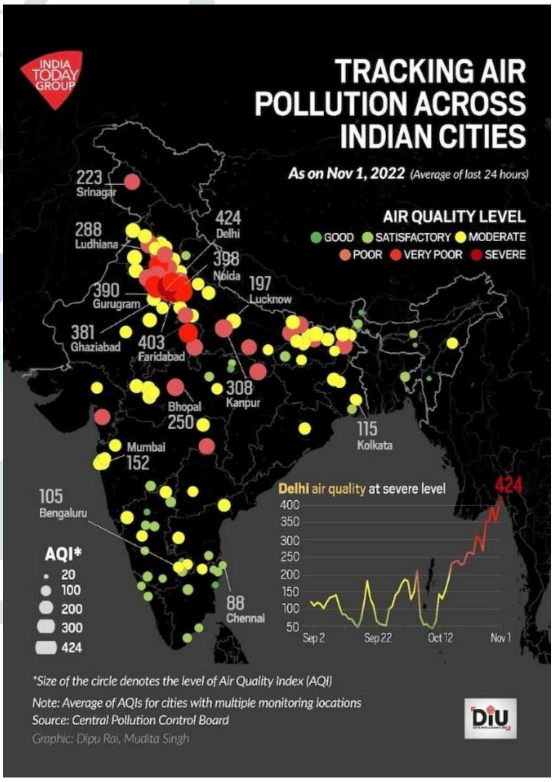
The Central Pollution Control Board's (CPCB)has a cutoff point which concentrates various methods of assessing AQI which has gained acceptance lately. According to this method evaluation of the AQI values depends on the individual pollutants and then indicating which has the highest value in order to get the overall AQI for that particular hour. Countries alike The United States Environmental Protection Agency (USEPA) and China have both adopted this system, which indicates the efficiency of managing air quality. It is important to note that the AQI considers only certain pollutants only into the consideration, but it overlooks other factors like temperature violations, pollen levels which can significantly affect both the human health and air quality. Recognizing these addition factors the scope of AQI plays an important role and hence there is a need to study it.

Presently these AQI systems recurrently taken into account coactively by performing calculations like the mean ratios of pollutants over standard levels for a precise time period, whereas the prior versions of AQI mainly focused on exceeding the health-based standards for outdoor quantity within a defined time period. We used to categorize AQIs which considers various mean values, such as arithmetic, geometric, logarithmic and weighted breakpoint concentration.

For assessing AQI, the Central Pollution Control Board's (CPCB) breakpoint concentration approach has gained popularity. With this method, the total AQI for a given hour is calculated by calculating the AQI values of each pollutant individually and then choosing the greatest value among them. Both China and the US Environmental Protection Agency (USEPA) have adopted this system because it works well for managing air quality. It's crucial to remember that the AQI only takes into account particular pollutants and ignores other elements that may have an impact on both human health and air quality, such as temperature or pollen levels.

II. METHODOLOGY

We carried out real-time air quality monitoring in a residential neighborhood to obtain insights into the fluctuations in air quality over time and to identify instances of increased air pollution in the chosen study area. Figure 1 shows the air pollution levels in the region as of November 1, 2022, according to an article published in India Today. Using the Environment S.A. CAAMS Analyzer (Continuous Ambient Air Monitoring Station), our research primarily assessed the ambient air quality. This high-tech equipment made it easier to quantify the quantities of tiny particulate matter, such as PM10 and PM2.5, as well as the amounts of Sulphur dioxide (SO2) and nitrogen dioxide (NO2).



The CAAMS fine particle meter measures PM10 and PM2.5 levels using the reliable Beta Attenuation Method. Every hour, a tiny C14 (Carbon-14 or Krypton 85) element emits high-energy electrons, often known as beta rays, continuously via a clean filter tape. The accurate detection of particulate matter concentrations is ensured by this technique. We used the UV fluorescence approach, which depends on SO2 molecules' light emission in response to ultraviolet (UV) radiation, to track SO2 levels.

With a chemiluminescence analyzer, we were able to quantify the amount of nitrogen oxides, specifically NO2, in the atmosphere. Our monitoring equipment was painstakingly calibrated using a traceable standard reference gas approach, guaranteeing the accuracy and dependability of the data gathered on air quality. Our comprehension of the environmental conditions in the study region has improved as a result of this thorough methodology, which gave us important insights into the temporal changes and episodic increases in air pollution.

It is becoming more and more important for society to priorities raising knowledge of daily air pollution levels in the modern world. An essential instrument in this effort is the Air Quality Index (AQI), which offers a way to evaluate and report on the general conditions and trends of air quality using predetermined criteria. The AQI index, is defined as the measurement of pollution and it is basically calculated in India using the Standard of CPCB. The result of this helps us to provide useful information on the present position of the quality of air around our surroundings. Furthermore, it is also indispensable in supporting the general population by understanding the level of pollution that is present in air which people breathe on a daily basis. This fact empowers individuals to make knowledgeable decisions about their activities which helps them to take measures which protects their health and well-being.

According to WHO air quality guidelines they offer a criteria in which they give pollution levels which are detrimental not only to human health but also have given the evaluation of the side effects of air pollution. as the table illustrates.

Pollutants	Time Weighted Average	Standard limits as per WHO guidelines (μg/m^3)
Particulate matter (PM) – 2.5	Annual mean	10
	24 hours mean	25
Particulate matter (PM) – 10	Annual mean	20
	24 hours mean	50
Ozone (O_3)	8 hour mean	100
Nitrogen dioxide (NO_2)	Annual mean	40
	1 hour mean	200
Sulphur dioxide (SO_2)	24 hours mean	20
	10 minute mean	500

Fig 1:Air Quality Guidelines as per WHO

AQI (CPCB,1994)	AQI (Malaysia,1999)	AQI (Wt. Avg.)	AQI (ORAQI)	AQI (USEPA, 2006)	AQI (CPCB, 2014)
(AQI<100) Good	(AQI<10) Very Clean	(0≥AQI≤0.5) Acceptable	(0≥AQI≤25) Clean	(up to 50) Good	(0-50) Good
(AQI>100) Harmful	(10≥AQI<25) Clean	(0.51≥AQI≤1) Unacceptable	(26≥AQI≤50) Light Air Pollution	(51-100) Moderate	(51-100) Satisfactory
	(25≥AQI<50) Fairly Clean	(1.01≥AQI≤2) Alert	(51≥AQI≤75) Moderately Polluted	(101-150) Unhealthy for sensitive Groups	(101-200) Moderately polluted
	(50≥AQI<75) Moderately Polluted	(AQI≥2.01) Significant harmful	(76≥AQI≤100) Heavy Air Pollution	(151-200) Unhealthy	(201-300) Poor
	(75≥AQI<100) Polluted		(AQI>100) Severe Air Polluted	(201-300) Very Unhealthy	(301-400) Very Poor
	(100≥AQI<125) Highly Polluted			(301-500) Hazardous	(401-500) Severe
	(AQI≥125)				

Table : Classification of AQI used for Comparative Study

Fig 2: Classification of AQI for Comparative Study

Method 1:

The AQI helps us to determine the concentrations of pollutants which is set against the standards based on some sample arithmetic mean calculations. Various pollutants like PM10, PM2.5, NO2, and SO2 will be calculated by scaling average values by 100 gives you a final AQI index value. As per our analysis Kaushik et al.'s study (2006), he has calculated AQI values by using a predetermined rating scale. The AQI is calculated for each and every distinct pollutant using the given formula. This technique provides a consistent and systematic way for the evaluation of quality of air, and also helps in making assessments and examination of pollution levels simpler.

$$AQI = \left(\frac{C}{C_S}\right) * 100$$

AQI is defines as follows

• The letter "C" is represented by pollutants which are observed value of air quality parameters, such as PM10, PM2.5, NO2, and SO2.

• The letter "Cs" is corresponding to the CPCB 2009 guidelines for residential areas.

Method 2:

The Oak Ridge National Air Quality Index (ORNAQI) have given a specific formula for calculate the complete AQI which ensures standard and consistent approach for the calculation of AQI.

$$AQI = \left[39.02 \sum \frac{c}{c_s}\right]^{0.967}$$

The study was carried out Bhuyan et al.'s 2010s, the Air Quality Index (AQI) was then assessed and compared with the relative overall Air Quality Index (ORAQI) values. This comparison has led the light on understanding the relationship between the measured AQI and also the overall air quality status which is determined by ORAQI values.

Method 3:

Accumulated Index Calculation:

Accumulated Index = AQI/I = Max (I1, I2… In).

This is the most popular way of calculating the Accumulated Index

IV. RESULT

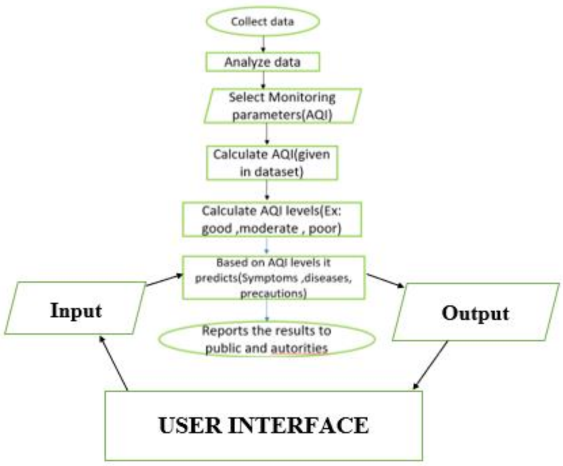


Fig3: Flowchart of AQI

The Air Quality Index (AQI) is a vital instrument for evaluating and comprehending the quality of the air we breathe. Clean and healthy air is a basic requirement for human well-being. Air quality is divided into six different buckets by the AQI system, each of which is defined by a certain range of AQI values. These classifications, which vary from "Good" to "Severe," include vital details regarding the concentrations of pollutants in the atmosphere and how they could affect human health. This introduction highlights the significance of AQI awareness in preserving public health by examining the AQI classification, the related illnesses and symptoms, and the suggested measures at each level. Making educated decisions regarding outdoor activities and proactively reducing risks requires an understanding of the AQI.

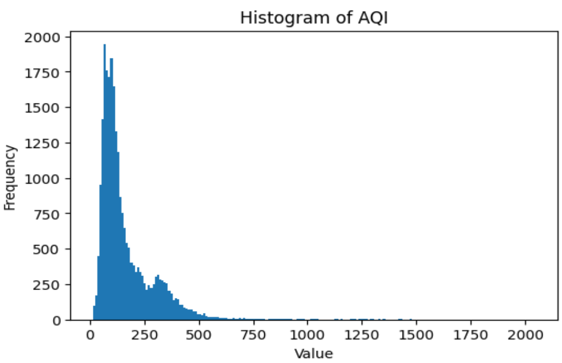


Fig 4: Histogram of AQI

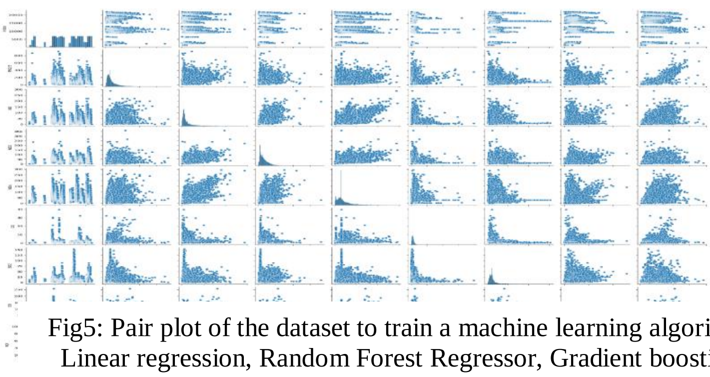
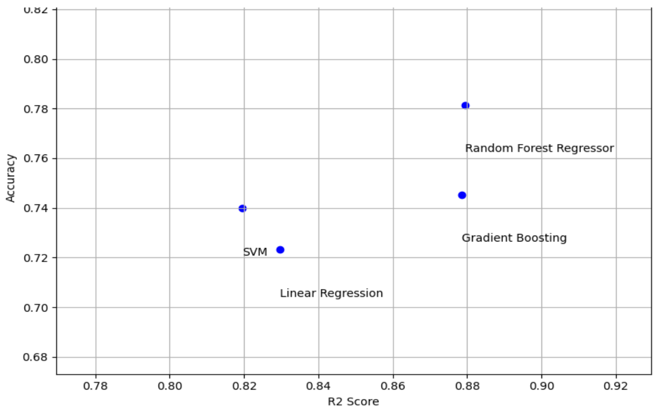


Fig5: Pair plot of the dataset to train a machine learning algorithm like Linear regression, Random Forest Regressor, Gradient boosting and Support Vector machine.

Fig 6: Descriptive Analysis of South Region



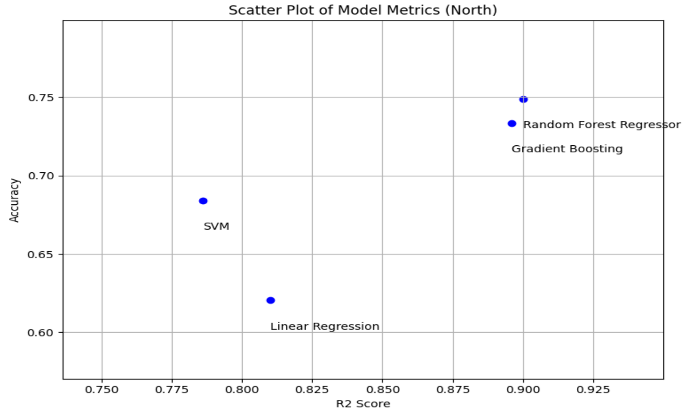
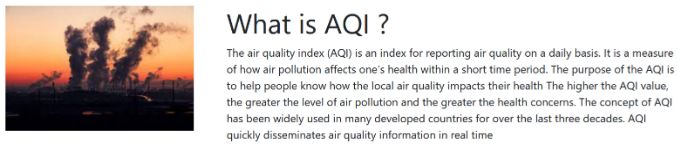


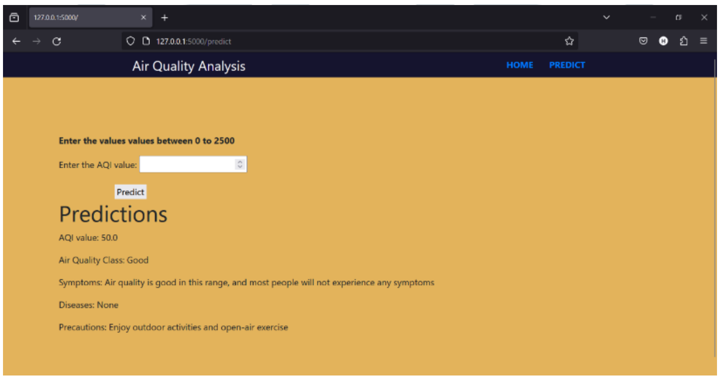
Fig 6: Descriptive Analysis of South Region

Fig 7: Descriptive Analysis of North Region

V. USER INTERFACE

This user interface interacts with the user and give the output for the user like Aqi_bucket, Symptoms, Diseases, Precautions.





Screen 2 where we enter the value and it gives the predictions and the precautions.

VI. CONCLUSION

To sum up, the suggested Air Quality Index (AQI) prediction and management system offers a comprehensive approach to the pressing problems of community involvement, public health protection, and air quality monitoring. Through the integration of sophisticated predictive modelling with intuitive user interfaces, the system enables people to take control of their health, make educated decisions, and actively engage in the management of air quality. Key project elements, such as feature selection, data preprocessing, model training, and user interface development, combine to provide a comprehensive framework that facilitates the delivery of real-time AQI forecasts and related health information.

This initiative marks a major advancement in the field of environmental science and technology, contributing to the creation of healthier living conditions and cleaner air. The system can improve public health, environmental preservation, and international cooperation in addressing air pollution problems with further study, innovation, and community involvement.

VII. ACKNOWLEDGMENT

This is our dedicated work of our research assistants in machine learning which have significantly contributed to the originality and depth of our findings. Our project is a evidence to the collective effort and steadfast commitment of all these individuals and institutions to our groundbreaking research.

REFERENCES

[1] https://link.springer.com/article/10.1007/s10661-005-9161-x
[2] https://www.researchgate.net/publication/282222215_A_Review_on_Air_Quality_Indexing_System
[3] https://www.indianjournals.com/ijor.aspx?target=ijor:zijmr&volume=4&issue=10&article=009