

## GEOGRAPHICAL AIR POLLUTION PREDICTION

Mrs. ROJARAMANI

Assistant Professor

Department Of CSE

[roja.adapa@tkrec.ac.in](mailto:roja.adapa@tkrec.ac.in)

Teegala Krishna Reddy Engineering College,  
Hyderabad

N Soumitha

Department Of CSE

[soumitha.nagabandhi@gmail.com](mailto:soumitha.nagabandhi@gmail.com)

Teegala Krishna Reddy  
Engineering College  
Hyderabad

Korra Jeevan

Department Of CSE

[jeevankorra91@gmail.com](mailto:jeevankorra91@gmail.com)

Teegala Krishna Reddy  
Engineering College  
Hyderabad

MD.Nawaz

Department Of CSE

[itsmenawaz7799@gmail.com](mailto:itsmenawaz7799@gmail.com)

Teegala Krishna Reddy  
Engineering College  
Hyderabad

### ABSTRACT:

Air quality has been an enormous health concern in recent decades as the place has become further industrialized and more and more citizens have begun driving automobiles. Random Forest Algorithm successfully bring about better accuracy and true positive rate with 92.49 accuracy. In our proposed model we are using Random forest Algorithm. This is a supervised learning algorithm which can be used for both regression and classification problems. Our system takes input of PM2.5, PM10, NO, NO<sub>2</sub>, NO<sub>x</sub>, NH<sub>3</sub>, CO, SO<sub>2</sub>, O<sub>3</sub>, Benzene, Toluene, Xylene and then inferring Air Quality Index (AQI). System produces range of pollution in the city such as severe, Moderate, poor, very poor, satisfactory and ARIMA model predict future values based on past values. Furthermore, there is a greater likelihood of motivating changes in both individual behaviour and public policy, as people want air quality.

### INTRODUCTION

Air pollution is dangerous for human health and should be decrease fast in urban and rural areas so it is necessary to predict the quality of air accurately. There are many types of pollution like water pollution, air pollution, soil pollution etc, but most important among these is air pollution which should be controlled immediately as humans inhale oxygen through air. There are various causes of air pollution. Outdoor air pollution caused by industries, factories, vehicles and Indoor air pollution is caused if air inside the house is contaminated by smokes, chemicals, smell. Two types of Pollutants that is causing air pollution are Primary Pollutants and Secondary Pollutants. Primary Pollutants include: - Carbon dioxide (CO<sub>2</sub>): Carbon dioxide is playing an important role in causing air pollution. It is also named as Greenhouse gas. Global warming a major concern caused by increase in carbon dioxide in air. CO<sub>2</sub> is exhale by Human. CO<sub>2</sub> is also released by burning of fossil fuels. Sulphur oxide (SO<sub>x</sub>): Sulphur dioxide (SO<sub>2</sub>) released by burning coal and petroleum. It is released by various industries. When react with Catalyst (NO<sub>2</sub>),

results in  $H_2SO_4$  causing acid rains that forms the major cause of Air Pollution. Nitrogen oxide (NOX): Most commonly Nitrogen dioxide (NO<sub>2</sub>) that is caused by thunderstorm, rise in temperature. Carbon monoxide (CO): -Carbon monoxide is caused by burning of coal and wood. It is released by Vehicles .It is odorless, colorless, toxic gas. It forms a smog in air and thus a primary pollutant in air pollution. Toxic metals –Example are Lead and Mercury Chlorofluorocarbons (CFC): -Chlorofluorocarbons released by air conditioners, refrigerators which react with other gases and damage the Ozone Layer. So in Proposed system it takes input of PM<sub>2.5</sub> ,PM<sub>10</sub> ,NO ,No<sub>2</sub>, NO<sub>x</sub>, NH<sub>3</sub>, CO, SO<sub>2</sub>,O<sub>3</sub>,Benzene,Toluene,Xylene and then inferring Air Quality Index (AQI).System produces range of pollution in the city such as severe, Moderate, poor, very poor, satisfactory for user awareness.

While many steps have been taken to reduce air pollution at the national and state level, air pollution in Indian cities is getting worse every day. Keeping in mind the serious situation all Indian cities are, air pollution has taken over the entire country.

## 1.2 PROBLEM DEFINITION

Observing and retaining high standard air has become the crucial challenge in metropolitan areas which have more industries, companies and population. As there is a rise in population, there is an increase in the transportation, usage of electricity and fuels. There is a lot of waste dumped in the land which we are well aware of. The air is also highly contaminated which causes a more serious threat to all kinds of living organisms in the earth. This gives rise to the need for monitoring and assessing the quality of air

and accordingly the government should be given alert to take necessary actions.

## 1.3 SCOPE OF THE PROJECT

Air quality has been an enormous health concern in recent decades as the place has become further industrialized and more and more of its citizens have begun driving automobiles. The occurrence of air pollution takes place in ways like, Release and generation of pollutants from their source. Carry of pollutants in the atmosphere. Penetrating and negatively impacting human health and ecosystem. PM 2.5 is especially dangerous because it can pass through the human body's natural filters and enter the lungs. Health concerns related to PM 2.5 include heart and lung disease, asthma, bronchitis, and other respiratory problems. Existing systems use Machine learning algorithms such as Linear Regression, KNN, Stochastic Gradient Descent, Decision Tree. Although performance of all models was comparable, Random Forest Algorithm successfully bring about better accuracy and true positive rate with 92.49 accuracy. So in our proposed model we are using Random forest Algorithm. This is a supervised learning algorithm which can be used for both regression and classification problems. Our system takes input of PM<sub>2.5</sub> ,PM<sub>10</sub> ,NO ,No<sub>2</sub>, NO<sub>x</sub>, NH<sub>3</sub>, CO, SO<sub>2</sub>,O<sub>3</sub>,Benzene,Toluene,Xylene and then inferring Air Quality Index (AQI).System produces range of pollution in the city such as severe, Moderate, poor, very poor, satisfactory and ARIMA model predict future values based on past values. Accurate forecasting helps people plan ahead, decreasing the effects on health and the costs associated. People should be aware of variations in the quality of the air they breathe, the effect of pollutants on

health as well as concentrations likely to cause adverse effects and actions to curtail pollution. Furthermore, there is a greater likelihood of motivating changes in both individual behaviour and public policy, as people want air quality.

## II. LITERATURE SURVEY

One of our era's greatest scourges is air pollution, on account not only of its impact on climate change but also its impact on public and individual health due to increasing morbidity and mortality. There are many pollutants that are major factors in disease in humans. Among them, Particulate Matter (PM), particles of variable but very small diameter, penetrate the respiratory system via inhalation, causing respiratory and cardiovascular diseases, reproductive and central nervous system dysfunctions, and cancer. Despite the fact that ozone in the stratosphere plays a protective role against ultraviolet irradiation, it is harmful when in high concentration at ground level, also affecting the respiratory and cardiovascular system. Furthermore, nitrogen oxide, sulfur dioxide, Volatile Organic Compounds (VOCs), dioxins, and polycyclic aromatic hydrocarbons (PAHs) are all considered air pollutants that are harmful to humans.

The capital city of India, New Delhi has been recognized as the most polluted city in the world. World Health Organization (WHO) has reported this as per their findings in 2014. It is indeed an alarming issue for the health of our future generations in India. We are aware of the ill effects of environmental pollutants and toxicants on health status of human as well as other living organisms and the environment. The most vulnerable to the toxic effects of the pollutants are children and old people. Some significant measures

should be taken and some strict laws should be made to prevent environmental pollutions in the major cities of India. Deadly diseases like cancer and asthma etc. are increasing in Indian responsible for such increasing incidences of diseases.

## III. SYSTEM ANALYSIS

### 3.1 EXISTING SYSTEM

In first approach monitoring of real-time Air Quality Monitoring and another is developing statistical models using historical data and some existing systems use Machine learning algorithms such as Linear Regression, Stochastic Gradient Descent, Decision Tree. These models are used to predict the air quality index of a given area.

### LIMITATIONS OF EXISTING SYSTEM

- These models fail to predict accurate AQI values when the dataset is large.
- A small change in the data can lead to a large change in structure and this causes instability.

### 3.2 PROPOSED SYSTEM

Proposed system takes input of PM<sub>2.5</sub>, PM<sub>10</sub>, NO, NO<sub>2</sub>, NO<sub>x</sub>, NH<sub>3</sub>, CO, SO<sub>2</sub>, O<sub>3</sub>, Benzene, Toluene, Xylene and then inferring Air Quality Index (AQI). System produces range of pollution in the city such as severe, Moderate, poor, very poor, satisfactory for user awareness.

### MERITS OF PROPOSED SYSTEM

- Random Forest Algorithm successfully bring true positive rate with 92.49 accuracy with large dataset.
- Random forest is used for both Regression and classification. It can automatically handle missing values.

- Autoregressive integrated moving average model predict future values based on past values.

### 3.3 ALGORITHM

#### RANDOM FOREST

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

The below diagram explains the working of the Random Forest algorithm:

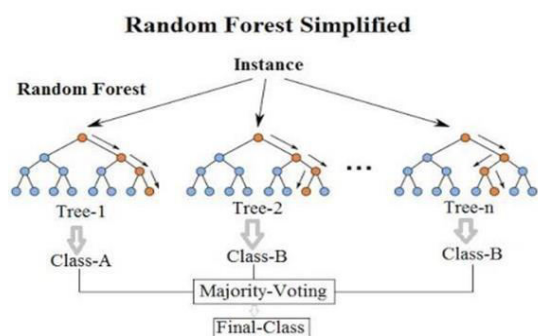


Fig :Random Forest

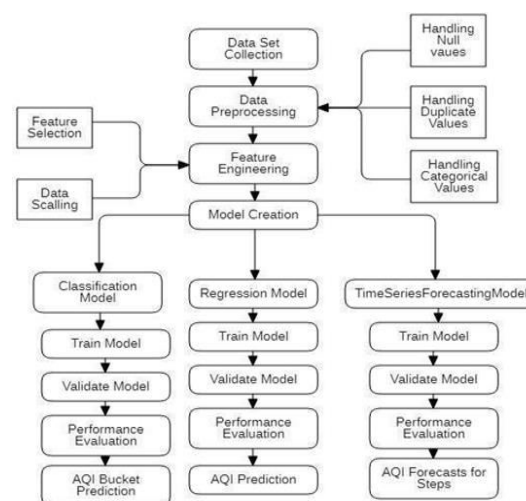
#### Assumptions for Random Forest:

Since the random forest combines multiple trees to predict the class of the

dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

### IV. SYSTEM ARCHITECTURE



#### 4.1 MODULES

##### Data Collection:

Data is composed from a different source and optimized for data sets. The data is used to calculate the AQI value AQI bucket and to predict the future pollution. Several abstract online outlets, like Kaggle, Google weather forestation.

##### Data Cleansing:

Preprocessing the data is considered as significant step in machine learning phase. It involves adding the missing values, the

correct set of data and extracting the functionality. Data set form is important to the process of analysis.

### Regression Model:

The Regression is done by algorithm called Random Forest it is used for both Classification and regression problems. The outcome of the Regression model is AQI value.

### Classification Model:

The classification is done by algorithm called Random Forest it is used for both Classification and regression problems. The outcome of the Classification model is AQI value.

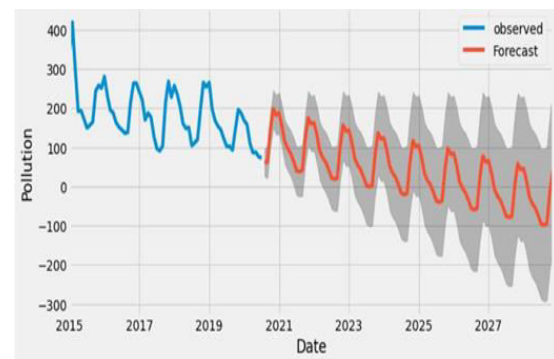
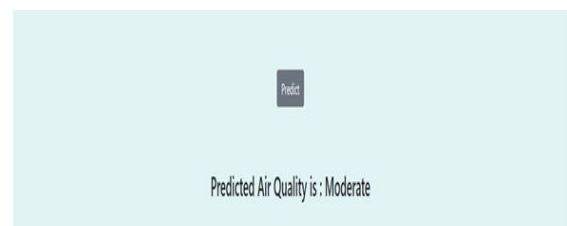
### Timeseries Forecasting Model:

ARIMA is a timeseries Forecasting Model used for predicting the future values of air pollution based on the past available data.

## V. INPUT AND OUTPUT SCREENS

### INPUT

### OUTPUT



## VI. CONCLUSION

Geographical air pollution prediction is extremely useful as air quality has been an enormous health concern in recent decades as the place has become further industrialized and more and more of its citizens have begun driving automobiles. PM 2.5 is especially dangerous because it can pass through the human body's natural filters and enter the lungs. Health concerns related to PM 2.5 include heart and lung disease, asthma, bronchitis, and other respiratory problems. System produces range of pollution in the city such as severe, Moderate, poor, very poor, satisfactory and ARIMA model predict future values based on past values. Accurate forecasting helps

people plan ahead, decreasing the effects on health and the costs associated. People should be aware of variations in the quality of the air they breathe, the effect of pollutants on health as well as concentrations likely to cause adverse effects and actions to curtail pollution. Furthermore, there is a greater likelihood of motivating changes in both individual behaviour and public policy, as people want air quality.

grained air quality based on big data,” in Proc. the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2267-2276, August 10, 2015.

└

## BIBLIOGRAPHY

1. V. M. Niharika and P. S. Rao, “A survey on air quality forecasting techniques,” *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 103-107, 2014.
2. NAAQS Table. (2015). [Online]. Available: <https://www.epa.gov/criteria-air-pollutants/naaqs-table>
3. E. Kalapanidas and N. Avouris, “Applying machine learning techniques in air quality prediction,” in *Proc. ACAI*, vol. 99, September 2017.
4. Questioning smart urbanism: Is data-driven governance a panacea? (November 2, 2015).
5. D. J. Nowak, D. E. Crane, and J. C. Stevens, “Air pollution removal by urban trees and shrubs in the United States,” *Urban Forestry & Urban Greening*, vol. 4, no. 3, pp. 115-123, 2014.
6. T. Chiwewe and J. Ditsela, “Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations,” presented at 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), IEEE, 2016.
7. Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, “Forecasting fine-