

# Air Quality Monitoring System

G. Vijay Kumar <sup>a,1</sup>, K. Kiran Kumar <sup>a</sup>, Bravara Manohar <sup>a</sup>, Chintha Venkatesh <sup>a</sup>, Shaik Chand Basha <sup>a</sup>, Shaik Javeed Akthar <sup>a</sup>

<sup>a</sup>Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur-522502, India

**Abstract.** Safeguarding and checking of air quality has becoming a most important practice in various modern and metropolitan zones now a days. The behaviour of air is unfavourably influenced by various types of contamination that we get by the transportation in production of power, natural resources utilizes and so for the emission of different poisonous gases is making a major danger for the personal life in various urban communities. The contamination of air is getting expanded day by day. Our main task is to provide a effective air contamination prediction observing techniques by which we can gather the data about infectious or poisonous gases present in every zone and gives us an idea of air contamination in each zone. Therefore, air quality measuring has become amongst important technique. The shape of air is influenced by multi-dimensional elements contains of area, time and unsure factors. As of now, various specialists get started to use the very large information investigation methodology because of various ways in large information applications and using the resources of nature detecting organisations and information sensed by sensor. The examination researches different various information and air quality evaluation methods using artificial intelligence methods. Moreover, it identifies and predict the future explanation needs.

**Keywords.** Air Quality Evaluation, Air Quality Prediction, Data Mining, Random Forest, Decision Tree.

## 1. Introduction

Air is the very important resources that we get from the nature. By inhaling this air various living organisms survive their life on this planet. All the living organism's life in all forms it is directly attached on the air for survival. Living organisms need sufficient amount of fresh air from emission of poisonous gases on this planet. Based on a survey in the past decades amount of air pollution has moved above 20% by the three analyses. Global warming is one most severe problem of the environment. In the atmosphere CO<sub>2</sub> level and greenhouse effect increasing day to day. These drastic changes increase radiation and the sunlight makes the planet hotter, this change in atmosphere is called greenhouse effect, it causes to rise in earth's mean temperature. Air contamination is one the major concerning issue in smart cities.

---

<sup>1</sup> G Vijay Kumar, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation Greenfields, Vaddeswaram, Guntur-522502  
Email: gvijay\_73@kluniversity.in.

Emission of poisonous gases in to the environment or releasing of poisonous gases in to the atmosphere causes direct or indirect problems to many living organisms and other species of the eco system. Contamination of air has major severe problem of air by various ways of emission contaminants in to the atmosphere. The substances which get pollute air and also leads to serious health issues and environment threat problems [1].

Smoke venting from vehicles is the reason for air contamination. The petroleum products burning are generous wellspring of sulphur dioxide. The raised degrees of sulphur dioxide from power plant and autos are the astounding impact of corrosive downpour. This downpour brings about consumption of metals and disintegration of structures. The essential presence of environments is through Ozone layer [2]. It is liable for safeguarding environments from poisonous bright (UV) beams. The uplifted environmental chlorofluorocarbons cause contamination in atmosphere. All in all, are responsible for arrangement off a few imperfections in birth, respiratory issues, heart issue, malignancy and nerve harm. By Air pollution, every year seven million people are dying worldwide. The data with WHO shows that 9 out of 10 people breathe air that exceeds WHO guideline limits containing high levels of pollutants or dust particles, with low and middle-income countries are suffering from the greater exposures. WHO is supporting countries to help on air pollution?

## 2. Related Work

Many researches are conducted in this field to improve the quality of air and to measure the amount of fresh air by using artificial intelligence and data mining. The problem of Quality of Air contamination is a big issue globally due to increase of industries and urbanization. Pollution levels around the globe are increasing day by day which is alarming. Different types of Air Quality system are proposed and are designed for monitoring pollution [3] levels at different levels. The major problem of the system's present for monitoring is that they require continuous power supply and Wi-Fi connectivity to work. The expansion of smart cities requires intelligent devices with rapid connectivity [4]. The usage and power of Wi-Fi is increasing, which can connect different physical device to the Internet. IOT provides access of internet to the object-flow of information. Earlier, WSN (Wireless Sensor Network) and RFID (Radio Frequency Identification) are used as targeted technologies in IOT. There is a possibility of finding regular patterns [5, 6] in air pollution system. There are many designs on surveillance of the issue through WSN. IOT enabled sensor networks were deployed where transmission node composed of sensing entity and connects to receiver node in wireless mode. The sensors which are fixed indifferent regions in different locations gathers data and next the data is transmitted to station where status report is generated [7].

### 3. Problem Statement

The main objective in this paper is, we are going to measure of infectious or poisonous gaseous particles that obtained in environment and as well we have to predict the coming days atmosphere conditions utilizing data mining & AI procedures with iot. By using machine learning& data mining techniques for future prediction is possible and we can predict the climatic conditions. Data mining, ML these are the present booming technologies across globe for the coming up generation.

### 4. Methodology

Air quality prediction system contains various data collected with huge amount of monitoring devices by the EPA, Control agencies or offices. AQS contains geological, about informative to each controlling station, and data quality control. This data is very useful to evaluate quality of air, non-compliance, assess regional and as well as national plan implementation in urban areas, perform modelling for required analysis & other type of qualities of air management functions. According to the "CLEAN AIR ACT", where AQS information is helpful to reports. Random forest algorithm uses technique where several decision trees constructed from subset of data and a summary of predictions. Use of this method is used to determine air quality for sensing system. The city data generated in sensing includes weather report, road data information, such real-time conditions. The use of Chinese standard that AQI is measured from six atmospheric gas levels, namely nitrogen dioxide (NO<sub>2</sub>), sulphur dioxide (SO<sub>2</sub>) and suspended small particles which have an aerodynamic diameter is less than 10 µm, which is to be not greater than PM<sub>10</sub> suspended particle.

#### 4.1 Data Mining and Machine Learning Techniques

AI& ML are the fields where a lot of immense rises in usage in last year. The technology of Artificial Intelligence in which machine makes decision on its own, rather than running most effective with the aid of using orders given by programmer as conventional programming works, progressively began out influencing all components of our life. Starting from early-stage, start-up companies and finishing to big companies, for all of them, Artificial intelligence & it's part machine learning is becoming the key areas. Learning technique in where machine which implements artificial intelligence gathers data from sensors in an environment and learns& decide show to act. One of the reasons why we have opted ML to predict the quality of air index, was this ability of adapting of machine learning algorithms.

Data mining is a method of extracting information and finding patterns in data sets and methods using database & machine learning systems. Data mining is a computer science and statistics interdisciplinary subfield with the aim of extracting useful data sets and transforming them into a human-readable structure.

## 4.2 Data Mining Algorithms

**Decision Tree Algorithm:** The idea of decision tree algorithms is to divide data into classifications so that an algorithm can predict where new data points will land. The values of independent & dependent variables are used for these classifications.

**Random Forest Algorithm:** Random Forest is process having multiple trees, a forest of trees. Those trees can be of equal type or the forest can be made of merging of trees (algorithms). Here are few interesting further metaphorical thoughts that describe how the forest acts (decides).

## 4.3 Air Quality Evaluation

Air quality prediction is most used & important monitoring technique and evaluating quality of air. Influence for supply of gas characteristics is its adaptability for specified purposes. Several air pollutants are called Air pollutants. These contaminations can affect the health & endanger your health. Current standard the pollutants are:

- Particulate matter (PM)
- Sulphur Dioxide (SO<sub>2</sub>).
- Ozone (O<sub>3</sub>), Nitrogen Dioxide (NO<sub>2</sub>)
- Lead (Pb)
- Carbon Monoxide (CO)

AQPS stores the data of air pollution received from EPA, state, local agency from thousands of air sensors. AQPS also stores the descriptive weather data Information about each station and quality of data Assurance/quality information control. AQPS data are used for sensing and to assist in meeting/not meeting standards Designate and evaluate the national implementation plan non-arrival areas, modelling for permitting the analysis of review and other air quality prediction management. Productivity management system This information is also used for report preparation for Congress because "Clean Air Act" regulations.

## 4.4 Air Quality Standards

The Office of Air Quality Plans and Standards (OAQPS) is in charge of overseeing EPA programmes aimed at improving air quality in the following regions. The current qualities are not acceptable, & prevent contamination in areas where air is polluting more. To achieve this, OAQPS established NAAQS Every standard pollutant. There are two types Standard-Elementary and Middle School.

- Main criteria: they can prevent bad health effect;
- Secondary criteria: prevent welfare Impacts, such as injures the vegetables, injures to crops, injures the building for preventing welfare impacts.

Since other contamination has other effects, the standards of NAAQS are not same. Some particles have some standards differently for longer-term average time & shorter-term average time. The standards for shorter-term are developed to prevent shorter-term effects of health, while longer-term Established standards for the prevention of

chronic health effect. Since some other pollutants will have effects differently, the standards for NAAQS are different as well. These standards or regulations designed to prevent shorter-term impact on health while establishing longer-term regulations or standards are to protect from health effects such as chronic.

4.5 Air Quality Analysis

Today, air quality solutions or designs are gaining attention and getting more interest. With help of "data mining", we can design the air system with strong dynamics and strong spatial scalability. It is heterogeneous in behaviour. These models get data from various sources, such as satellites, public Institutions, sensors, etc. Advancement in satellite sensors provide new dataset for monitoring and shrivelling urban zones air pollution scales. Articles published in accordance with Chicago policy Comment [5], in contrast with traditional data dependent on data sets Sample or rough summary, "big data" is huge large size, fast speed, and various types. Since the year 2000, the amount of data exploded because of the speed at which technologies is being developed and implemented, Networking, information processing, & storage od data is also part of the digital infrastructure. Can generate big data from targeted, and voluntary resources. The difference between availability the required resolution. Key for doing big data the feasible method is to use, standardize and integrate massive data. Figure 1 we are discussing some important research Using data-driven research around the world the method of predicting air quality in the following paragraphs.

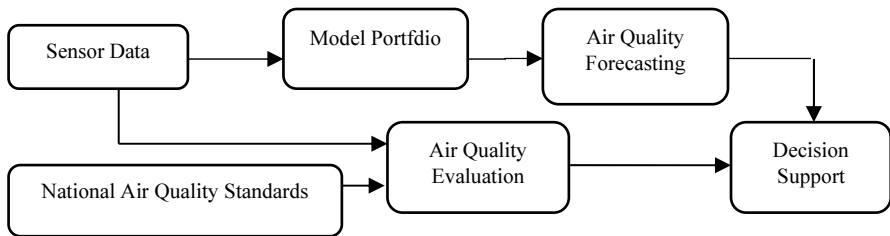


Figure 1. Block Diagram

The following are the steps to follow. Here the data will get imported first, after importing the data we have to train the model. Then accuracy functions have to be constructed. After that the model should be graphically represented. Evaluate the model which is previously represented graphically for getting the results. Input- Training the Data (D) to the model, Method Used-Here is Classification method and produce the result as a Decision tree. Initially, we extract the training data set (D). Let us assume I as input variable to model&T as the targeted variable. Figure 2 initially, the root nodes which have data of whole training set or may contain the data is constructed.

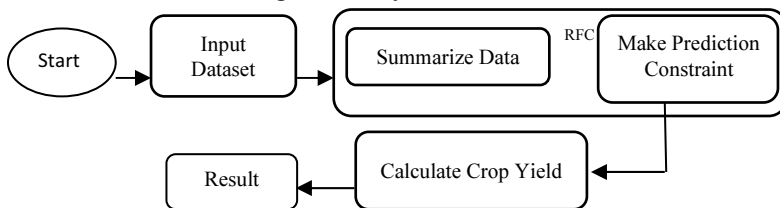


Figure 2. Schematic Diagram

Computing the Gini ratio for all the inputs  $I$  by the help of formula. Let us assume, the Gini ratio for total data set as  $G(d)$  and the attribute for Gini ratio as  $GA(d)$ . Now the formula is:

$$G(d) = 1 + \sum_{i=0}^m pi^2$$

Now the Gini variable is calculated as:

$GA(d) = (d1/d)G(d1) + (d2/d)G(d2)$ , where  $pi$  is referred as the relative frequency of class  $d$ . Let us assume  $X$  as splitting variable with min. index  $GA(d)$  among  $I$ . Splitting the  $d$  variable w.r.t.  $X$ . Now create a child node for current node for every subset and we have to pass these instances to the node. We have to repeat everything until the  $GA(d)$  is minimum. All the instances or values now belong to same class if the leaf or end node is less.

## 5. Conclusion

In this paper we find a proposed method of solution to standard air quality prediction techniques using big data & ML algorithms. we can monitor air quality and evaluate the future prediction. We can get real time quality of air and evaluate it with aid of traditional air prediction techniques, big data technology, and machine learning techniques. This paper reports on our latest literature review, which discusses and compares work quality assessment using big data analytics, ML approaches, models, and technology. Finally, some observations on future research issues, needs and challenges.

## References

- [1] Jerry ZeyuGao, GangXie, Sen Chiao, Gaganjot Kaur Kang, ShengqiangLu. Air Quality Prediction: Big Data and Machine Learning Approaches, 1 January 2018.
- [2] J. Ditsela and T. Chiwewe, "Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations", IEEE, 2016.
- [3] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.
- [4] P. S. Rao and V. M. Niharika, "A survey on air quality forecasting techniques", International Journal of Computer Science and Information Technologies, vol. 5, 2014.
- [5] G. Vijay Kumar, M. Sreedevi, NVS Pavan Kumar "Mining Regular Patterns in Transactional Databases using Vertical Format" International Journal of Advanced Research in Computer Science, Volume (2), Issue (5) 2011.
- [6] G. Vijay Kumar, V. Valli Kumari, "Sliding Window Technique to Mine Regular Frequent Patterns in Data Streams using Vertical Format", 2012 IEEE International Conference on Computational Intelligence and Computing Research, 2012.