

## AIR QUALITY PREDICTION USING BIG DATA

Swastik A Bhomkar<sup>\*1</sup>, Vasant Chatter<sup>\*2</sup>, Akshay K<sup>\*3</sup>, Pratiksha Asnotkar<sup>\*4</sup>, Sudhakara B<sup>\*5</sup>

<sup>\*1,2,3,4</sup>B.E(Student) Department of Computer Science and Engineering,  
Srinivas Institute of Technology, Mangalore, Karnataka, India

<sup>\*5</sup>Assistant Professor, Department of Computer Science and Engineering,  
Srinivas Institute of Technology, Mangalore, Karnataka, India

### ABSTRACT

Having recent air or respiratory pure healthy air has become extinct during this generation. Pollution is one amongst the most important and heavy challenges facing our cities these days. All types of life together with plants and animals rely upon air for his or her basic survival. Thus, all living organisms would like sensible quality of air that is freed from harmful gases to continue their life. Writing by The Hindu says that pollution in Bengaluru to travel up by 74 percent by 2030. It'll be primarily owing to vehicle exhaust, construction dirt, and on-road dirt within the town. In this paper, we tend to use Big Data analytics technique to predict pollution and quality prediction model as there's accessibility of meteoric information and sensor information. This may facilitate the society to stay track of pollution and awareness on the pollution happening.

**KEYWORDS:** Air Pollution, big data, Map Reduce, Vehicle Count

### I. INTRODUCTION

According to the study by Urban emissions along with researchers from the middle for Study of Science, Technology and Policy (CSTEP), the calculable PM10 (that is, stuff of but ten microns in size) pollution could increase by seventy four percent by 2030, LED primarily by vehicle exhaust, construction dirt, and on - road dirt. Researchers established an urban emission inventory for larger Bengaluru that coated sources like transport (over seventy 100000 vehicles currently), diesel generator sets, industries, brick kilns, urban road dirt, and open waste burning, among others. For 2015, the town emitted an calculable 31,300 tons of PM2.5 and 67,100 tons of PM10, states the study printed recently within the journal, part Pollution analysis. The increasing population, its cars and industries are polluting all the air at an alarming rate. Pollution will cause semi-permanent and short health effects. It's found that the aged and young youngsters are lot of stricken by pollution. During this work we are going to be grouping datasets of traffic and pollution from numerous websites like Kaggle and filter and scale back those datasets using hadoop with Map Reduce programming, at the moment we tend to can merge each data and obtain a trained model and so by using java server pages we tend to develop a UI and take current real time traffic data of a route from traffic sensors or live traffic websites and therefore predict the pollution.

### II. CONNECTED WORK

In the past few years, several researches and aspects were taken under consideration using Air quality. It is one amongst the foremost alarming considerations for India these days. Addressing this concern, within the past decades, several researchers have spent many time on finding out and developing completely different models and ways in air quality analysis and analysis. One amongst the most comes that projected was Am Agent based Traffic Regulation System for Air quality control [1] to mix the advantages of agent technology with each machine learning and massive information tools. An Artificial Neural Networks (ANN) model and also the Dijkstra algorithm was used for air quality prediction and also the least polluted path finding within the road network. All information processing tasks were performed over a Hadoop based framework: HBase and Map Reduce. In Air Quality Prediction: huge information and Machine Learning Approaches [2] Investigates numerous big-data and machine learning primarily based techniques for air quality prediction. In [3] pollution prediction through Internet of things technology and Big data analytics that examines the likelihood to create a fusion between the two new ideas within the context of predicting pollution that happens once harmful substances; like NO<sub>2</sub>, SO<sub>2</sub>, CO and O<sub>3</sub> were introduced into Earth's atmosphere. In [4] says that pollution in Bengaluru to travel up by seventy four percent by 2030. In [5] period pollution prediction model supported Spatiotemporal huge information Introduces a mixture of an extended immediate memory (LSTM) unit for statistic information and a Neural Network model for different pollution impact factors like weather to make a hybrid prediction model. This model is easy in design however still brings sensible prediction ability.

### III. PROJECTED SYSTEM

These are following steps projected:

**Step 1:** Datasets Collected are keep within the CSV file, this file consists of pollution information that's of pollutants index in their individual columns.

**Step 2:** A Map Reduce Program is written in Eclipse Platform and this code extracts the desired columns from the inserted dataset. The program is of split, scan and write the given input. S

**Step 3:** As told in Step one, equally for the road and traffic information has been done.

**Step 4:** Taking Hash Mapping construct under consideration each the cleansed information is created mapped into file that shows on that date and time specific vehicle with their count and are expected and this can be the trained model of the project.

**Step 6:** Frontend is created of JSP and CSS to create the framework look higher. Server here used is Tomcat. Foremost a user can login at the moment he's asked wherever he wants the air quality.

**Step 7:** User when coming into the situation wherever he needs air quality of that individual space it'll evoke the real time vehicle count, and for that it's suggested to enter the Traffic of that individual time from the traffic analysis web site of involved neck of the woods.

**Step 8:** On the premise of vehicles count it'll predict the air pollution of the actual space

**Step 9:** Finally, displaying the Results of pollution through bar chart that whether it is high or low.

### IV. FLOWCHART

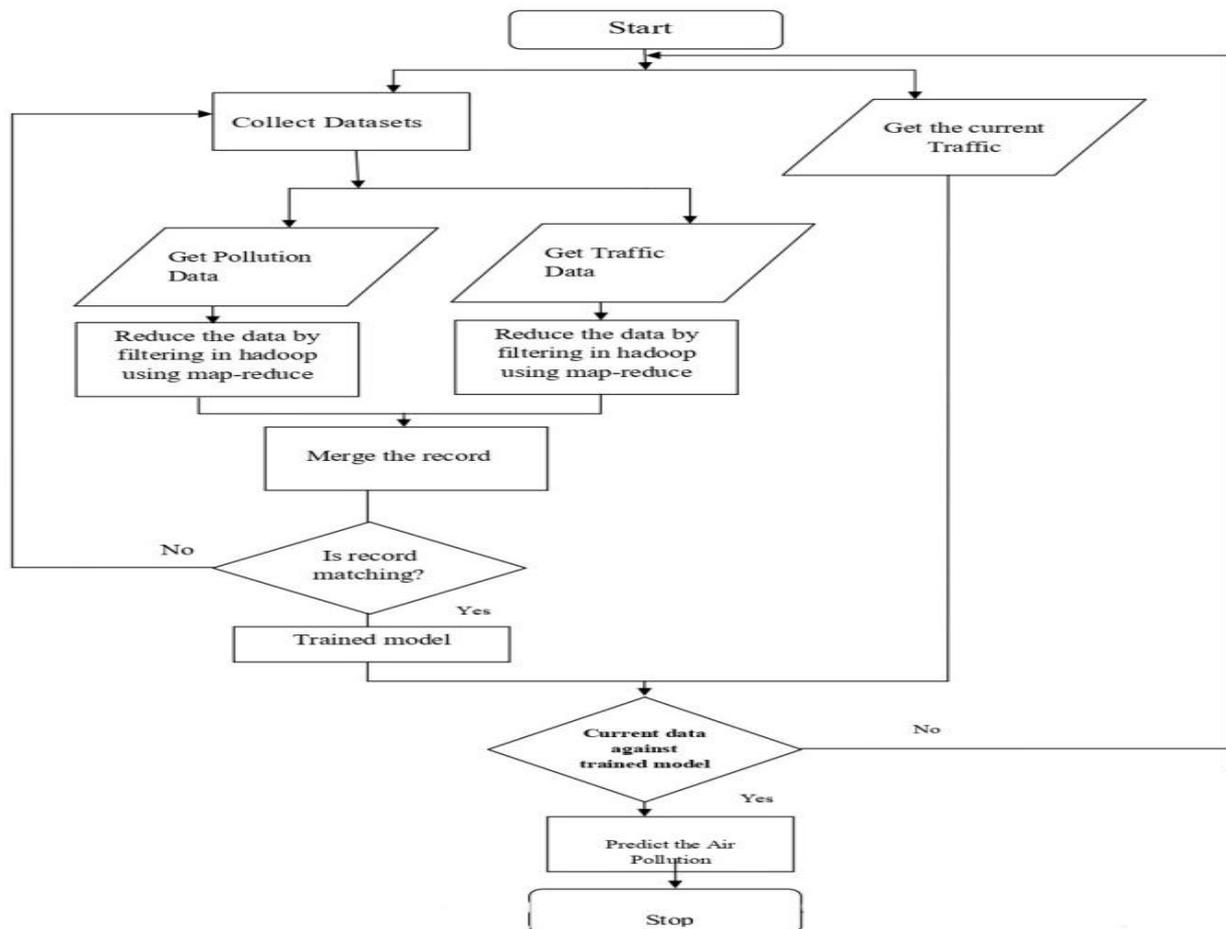


Figure 1: Shows the flowchart of the proposed system of air pollution prediction.

## V. AIR QUALITY EVALUATION

Air quality analysis is a very important task to monitor and manage pollution. The characteristics of air provide have an effect on its suitability for a selected use. Pollutants will injure health, hurt the atmosphere and cause property injury.

Pollutants are:

- 1) CO (CO)
- 2) Lead (Pb)
- 3) Nitrogen dioxide (NO<sub>2</sub>)
- 4) Ozone gas (O<sub>3</sub>)
- 5) Particulate matter (PM)
- 6) Sulfur dioxide (SO<sub>2</sub>).

In our work as a result to point out how smart the air pollution is we've got set some grades if the ultimate average results range from zero to fifty its Excellent and then on as shown in table 1 below

**Table 1:** Air pollution grades

0-50	Excellent
51-100	Good
101-150	Lightly Polluted
151-200	Moderately Polluted
201-300	Heavily Polluted
300+	Severely Polluted

## VI. CONCLUSION

With the Big data technology, Real time air quality monitor and analysis is fascinating for future good cities. Here during this work air quality analysis has been done considering the vehicle count based on traffic data and historical data. We tend to propose a real time pollution prediction model with historical and real time data. This model is straightforward in design however still brings smart prediction ability. In the future, we'll try and realize a lot of correct prediction models and be able to predict pollution values for any points of interest within the observation town. We'll additionally compare the prediction model using sensory device data to alternative hourly collected pollution datasets.

## VII. REFERENCES

- [1] El Menon, D. and Prasad, A. M. (2013). Development of Building Systems using Glass Fibre Reinforced Gypsum (GFRG) Panels, The Masterbuilder.
- [2] Gaganjot Kaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu, and Gang Xie, "Air Quality Prediction: Big Data and Machine Learning Approaches" International Journal of Environmental Science and Development, 2018.9.1.1066
- [3] Safae Sossi Alaoui; Brahim Aksasse; Yousef Farhaoui "Air pollution prediction through internet of things technology and big data analytics" International Journal of Computational Intelligence Studies, 2019 Vol.8No.3, pp.177 – 191
- [4] Staff Reporter "Air pollution in Bengaluru to go up by 74% by 2030: Study" thehinduarticle26064689
- [5] V. Duc Le, Sang Kyun, Cha "Real-time air pollution prediction modal based on spatiotemporal Big Data" in 2018/04/05