

A Smart Air Pollution Analytics Framework

Anindita Desarkar and Ajanta Das

Abstract Air pollution which is the worst environmental health risk across the world takes millions of lives every year both in developing and developed countries. These huge premature deaths happen due to long-term exposure to air pollutants as most of the cities do not meet the acceptable pollution level suggested by World Health Organization (WHO). So there is an urgent need to reduce the air pollution level across the globe. This paper proposes a state-of-the-art approach and proposes a layered air pollution reduction framework. The methodology of the proposed framework also suggests the action plans to reduce air pollution level with an innovative Rule Base and mining appropriate data from the huge dataset which is basically a data warehouse. It also discusses the expected outcome of the proposed framework beneficial to the citizens.

Keywords Predictive analysis · Air pollution · Action plan · Data mining · Data warehouse · Knowledge discovery

1 Introduction

Air pollution is basically contaminated air includes NO₂, SPM, photochemical oxidant, sulfur dioxide, carbon monoxide, and fine particulate matter are vulnerable to public health [1]. The major sources of air pollution include traffic sector, industrial domain, power plants, and fossil fuel burning. People who live in the polluted areas have increased risk of various heart and acute respiratory diseases, lung cancer, and other chronic problems which cause huge premature deaths, especially in developing countries [2, 3].

A. Desarkar · A. Das (✉)

Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Kolkata Campus, Kolkata 700107, India
e-mail: ajantadas@bitmesra.ac.in

A. Desarkar

e-mail: aninditadesarkar@gmail.com

The Central Pollution Control Board of India has taken up an initiative named NAQMP which stands for National Air Quality Monitoring Program. NAQMP takes into consideration 341 stations across 126 cities/towns and 4 union territories in India including Kolkata [4, 5]. In all of these locations, the amount of sulfur dioxide (SO₂), oxides of nitrogen (NO₂, etc.), suspended particulate matter (SPM), and respirable suspended particulate matter (RSPM/PM10) are regularly monitored. The monitoring is done continuously for 24 h in a frequency of twice a week, to have around 104 observations on a yearly basis. This monitoring is being carried out in association with SPCBs, National Environmental Engineering Research Institute (NEERI), Nagpur, etc. [6, 7].

The US Consulate General has set up an air pollution monitoring system in Park Street, Kolkata [8]. The data obtained from this monitoring system is displayed in raw concentrations only but this data is yet to be converted to an air quality index (AQI) [6, 9].

The major objective of this paper is to monitor air pollution level in Metropolitan cities and proposes a layered air pollution reducing framework. Novelty of this proposed air pollution monitoring framework is that it will collect various air pollution parameters from sample air and then it analyzes the level with respect to the AQI. This paper gives the emphasis on analytics and knowledge discovery part by which it will be able to forecast the forthcoming pollution level. Finally it suggests some action plan based on proposed Rule Base to reduce the pollution level if the prediction is higher than the threshold. The organization of rest of the paper as follows. Section 2 explains air pollution monitoring approach and methodology. Expected outcome is listed in Sect. 3, and Sect. 4 concludes the paper.

2 Air Pollution Monitoring Approach

According to Centre for Science and Environment (CSE), the level of air pollution in Kolkata far exceeds the permissible limit. Therefore the CSE recommends that Kolkata should take up emergency air pollution control measures as adopted in Delhi [9].

2.1 Proposed Air Pollution Framework

This section proposes a data analytics based layered framework toward monitoring and reducing air pollution in the metropolitan city. This proposed framework is presented in Fig. 1 consists of four layers and the layered-wise functionalities are described in the following.

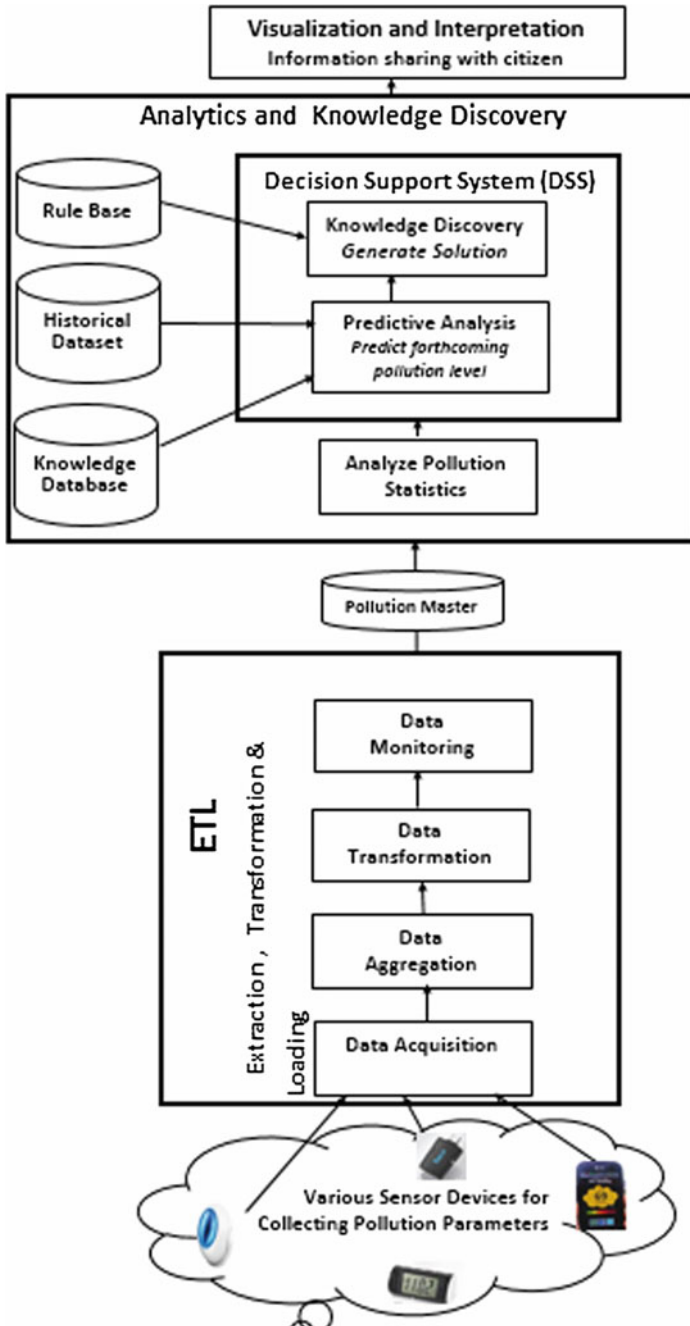


Fig. 1 Proposed air pollution monitoring framework

2.1.1 Source of Data

It includes various types of sensor devices

- Sensors, cameras, and other devices should be placed in various parts of the city for sample air collection.
- Various parameters for measuring air pollution like SO₂, NO₂, RSPM, and SPM will be calculated from the sample dataset.

2.1.2 Extraction, Transformation, and Loading (ETL)

- Sample data generated from various air samples would be aggregated for further analysis.
- The aggregated data should be transformed in proper format based on the monitoring requirement.
- After successful monitoring, the result dataset will be loaded into the pollution master database.

2.1.3 Analytics and Knowledge Discovery

- Specific analysis will be carried out from the pollution master database and statistical result can be interpreted.
- Predictive Analysis or *Analytics* will be performed on the output dataset to forecast the pollution level of the forthcoming days. Historical dataset and knowledge database both will be treated as two main sources of input generation for better prediction. Historical dataset would contain the pollution history and corresponding incidents occurred, as example, *how many accidents happened in a particular pollution level due to low visibility, how the pollution level affected the city life*, etc. Knowledge database will store information regarding the forthcoming events which may affect the pollution level of the city, like *information about the general holiday list*—assuming that city will encounter less traffic in the holidays which leads to lesser pollution level.
- The next part is *knowledge discovery* which would be performed based on the prediction generated in the previous level and with the help of *Rule Base*. Rule Base is basically a set of predefined rules along with their application criteria based on the pollution level. For example, Rule 1 will be applied if pollution parameters exceed a certain threshold value. Sample Rule Base is discussed in later section.

2.1.4 Visualization and Interpretation

- The last layer will represent the result after taking proper actions of the government authorities and city Management.
- The result must reflect the highly polluted region and laws or rules to be followed for those specific regions.
- It also suggests the safer region or less polluted region so that citizens can avail those routes to reach their destinations.

2.2 Framework Methodology

Major source of emission of carbon monoxide is transportation system. In order to control the air pollution, it needs to be monitored and vehicles responsible for polluting should be identified. Air quality index (AQI) plays major role in measuring or monitoring air pollution. It is a number provided by the government agencies to the citizen for their proper awareness. This number must act as a threshold value and should be displayed to the public who are traveling the region.

In India, the six categories of AQI are: *Good, Satisfactory, Moderately Polluted, Poor, Very Poor, and Severe*, depending upon the air quality. Associate health impacts (are numbered accordingly, like Impact 1, etc.) related to these AQI is presented in Table 1.

Suggested steps or plan of action to monitor air quality as well as pollution reduction are as follows.

Table 1 Categories of AQI and its associated health impacts [2, 10]

AQI	Associated health impacts
Good (0–50)	Impact 1: Safe
Satisfactory (51–100)	Impact 2: Slight uneasiness in breathing for sensitive people
Moderately polluted (101–200)	Impact 3: Uneasiness to people with heart disease, children and older adults
Poor (201–300)	Impact 4: Uneasiness to all kinds of people having heart disease
Very poor (301–400)	Impact 4 Impact 5: Respiratory illness could be found to the people on prolonged exposure. More vulnerable in people with lung and heart diseases
Severe (401–500)	Impact 5 Impact 6: May cause respiratory impact even on healthy people. The health impacts may be experienced even during light physical activity

2.2.1 Collection of Pollution Parameters from Sample Air

Different Air Pollution measurement parameters will be calculated from the samples collected through various sensors placed across the city. Collection of pollution is the main input to the framework. These ambient concentrations of the pollutants are used to measure the air quality index. Air pollutants can be SO₂, NO₂, RSPM, and SPM. Indices for single or individual pollutants should be collected and then summed up for generating the total air quality index. This process is iterative to standardize the value for the specific region at specific instant of time. The calculated value then needs to be stored in an “air pollution data file” (presented in Table 2) according to the location and time, respectively.

2.2.2 Analysis of Pollution Statistics

The air sample should be analyzed to find whether the present condition is vulnerable for city life. The mapping is developed between AQI and pollution percent violation based on the AQI impact on city life. Then a report can be prepared and minimal attributes for this analysis report are displayed in Table 3.

2.2.3 Predictive Analysis of the Forthcoming Pollution Level

This decision support system will predict the pollution level for forthcoming days by analyzing the pollution statistics collected from air sample along with historical dataset and knowledge database. As Example:

- Historical dataset (as presented in Table 4) includes various pollution history and their corresponding phenomenon, like 5 accidents occurred in a day due to low visibility if pollution percent violation is greater than 20. So it captures the event of creating smog in the air due to air pollution which effects visibility and hampers the safety of city life.

Table 2 Air pollution data file

Location	Date_Time	SO ₂ ()	NO ₂ ()	RSPM ()	SPM ()	AQI
----------	-----------	--------------------	--------------------	---------	--------	-----

Table 3 Sample statistics analysis report

Location	AQI	Weekly average	Standard deviation	Valid monitoring days	Percent violation
----------	-----	----------------	--------------------	-----------------------	-------------------

Table 4 Sample structure of historical dataset

Location	Pollution parameters	Period	Pollution percent violation	Overall pollution percent violation
----------	----------------------	--------	-----------------------------	-------------------------------------

Table 5 Sample structure of knowledge database

Location	Date	Major incident	Holiday
----------	------	----------------	---------

Table 6 Pollution metrics related rule set

Pollution percent violation	Selection of rule
<20%	1. Suggestive measure 1 2. Suggestive measure 2
≥ 20 and $\leq 40\%$	1. Rule 3 and (rule 4 or rule 5) 2. Suggestive measure 1 3. Suggestive measure 2
>40 and $\leq 60\%$	1. Rule 2
$>60\%$	1. Rule 1

Table 7 Sample Rule Base

S. No.	Rule name
Rule 1	No private vehicle should pass touching the region. Emergency situation would be excused
Rule 2	Private vehicles carrying four people can only pass through. Emergency situation would be excused
Rule 3	Only senior citizens would be allowed to pass through. Emergency situation would be excused
Rule 4	Only odd numbered private vehicles should pass the place. Emergency situation would be excused
Rule 5	Only even numbered private vehicles should pass the place. Emergency situation would be excused
Suggestive measure 1	Shared service cab providers would get extra benefit in tax
Suggestive measure 2	Solar energy enabled vehicles will receive special discount on paying road tax

- Knowledge database (as presented in Table 5) includes various events like the incoming holidays, other events like any large people gathering events which affect traffic situation.

2.2.4 Designing Sample Rule Base

The sample table, Table 6, describes rule name and their application criteria based on the parameter “pollution percent violation” value. Table 7 describes few sample rules which can be applied depending on the pollution forecast for the incoming days. New rules can be added or removed depending on the requirement. It is suggested that adequate laws and corresponding measures should be applied so that rules can be successfully implemented across the city.

3 Expected Outcome

Expected benefits from the proposed air pollution framework are presented in the following:

- (a) Citizens will lead life in a better healthy environment with clean air.
- (b) Risk of diseases will be reduced with reduction of SO₂ and NO₂ in air. It will ensure relief from unnecessary smog, acid rain, and lung diseases.
- (c) The amount of carbon emissions and also the fatalities caused due to air pollution will be reduced.
- (d) Number of accidents will be reduced with increasing the visibility in road by reducing the pollution level accordingly.
- (e) India will become a cleaner planet by its Swachh Bharat Abhiyan (Clean India Mission) campaign by 2019.

4 Conclusion

Air pollution monitoring study reveals that concentration of air pollutants is increasing in the air due to various reasons where growing number of vehicles dominate the other sources. Several initiatives have been taken to tackle the issue effectively and efficiently by the government authorities as well as various NGOs which includes both policy reformation and technological innovation [11]. This paper proposes a smart framework for metropolitan cities toward reducing air pollution. It presents a novel Rule Base for better healthy environment. Finally it also lists the expected outcome for the study.

References

1. "Air Quality Assessment" developed by Central Pollution Control Board, Retrieved from <http://www.cpcb.nic.in/15-44.pdf> on 17-06-2016.
2. Jana Spiroska, Md. Asif Rahman and Saptarshi Pal, "Air Pollution in Kolkata: An Analysis of Current Status and Interrelation between Different Factors", SEEU Review. Volume 8, Issue 1, Pages 182–214, ISSN (Print) 1409–7001, DOI: [10.2478/v10306-012-0012-7](https://doi.org/10.2478/v10306-012-0012-7), February 2013
3. Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., & Karr, C. (2015). A cohort study of traffic-related air pollution impacts on birth outcomes (Doctoral dissertation, University of British Columbia).
4. "Citizens Report – Air Quality and Mobility in Kolkata" developed by Center for Science and Environment, New Delhi 2011 Retrieved from <http://www.cseindia.org/userfiles/Kolkata%20Report.pdf> on 17-06-2016.
5. A report on "National Air Quality Index (AQI)" launched by the Environment Minister under initiative 'Swachh Bharat' on 17th Oct-2014 Retrieved from <http://pib.nic.in/newsite/PrintRelease.aspx?relid=110654> on 17-06-2016

6. Mansi Shah, "Waiting for Health Care: A Survey of a Public Hospital in Kolkata". Accessed from http://ccs.in/internship_papers/2008/Waiting-for-Healthcare-A-survey-of-a-public-hospital-in-Kolkata-Mansi.pdf on 22nd February 2016.
7. Pethuru Raj and Anupama C. Raman, (2015) "Intelligent Cities: Enabling Tools and Technology", CRC Press. (c) 2015
8. "Air Pollution level in Kolkata is much higher than permissible limit: CSE" Retrieved from http://articles.economictimes.indiatimes.com/2016-01-31/news/70222631_1_air-quality-anumita-roychowdhury-pm-2-5-levels on 17-06-2016
9. SuklaBhaduri, "Vehicular Growth and Air Quality at Major Traffic Intersection Points in Kolkata City, An Efficient Intervention Strategies", The SIJ Transactions on Advances in Space Research & Earth Exploration (ASREE), Vol. 1, No. 1, September-October 2013
10. "Air Quality Data" Retrieved from <http://kolkata.usconsulate.gov/airqualitydata.html> on 17-06-2016.
11. Schnelle Jr, K. B., Dunn, R. F., & Ternes, M. E. (2015). Air pollution control technology handbook. CRC press.