INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 2)

Available online at: www.ijariit.com

Prediction of urban air pollution by a machine learning method

Saniya Afshan
saniyaafshan.kh@gmail.com
Kammavari Sangha Institute of Technology, Bengaluru, Karnataka

Nithin S
nithselv@gmail.com
Kammavari Sangha Institute of Technology, Bengaluru, Karnataka

Venkata Sai Nishanth
cvsainishanth@gmail.com
Kammavari Sangha Institute of Technology, Bengaluru, Karnataka

S Nithin
nithinnaidu03@gmail.com
Kammavari Sangha Institute of Technology, Bengaluru, Karnataka

Aditya Pai H
adityapaih2007y@gmail.com
Kammavari Sangha Institute of Technology, Bengaluru, Karnataka

ABSTRACT

In addition to the impact of rapid population growth and urbanization, the level of pollution in the cities is largely modulated by meteorological factors. A compilation of meteorological data in many cities is available in Indian Government websites. They provide the concentration of several parameters such as PM2.5, NO2, CO, etc. With the help of this website data, we use Data Mining Prediction technique Collaborative Filtering and Machine learning to predict the next day data and compare with the real data. This fundamental research intends to be a preliminary step in the development of a web-based platform to alert the inhabitants of urban areas about the risk to human health, with potential future application in other urban areas.

To make the real-time scenario Service Orientated Web Architecture is used in this system.

Keywords: *Collaborative Filtering, Data Mining, Machine learning, Web-based application.*

1. INTRODUCTION

The present world's population of 7.4 billion people (Fig. 1) is expected to cross 9.7 billion over the next 30 years, with most of the growth occurring in urban areas of the less developed parts of the world as shown in Fig.2. The effects of the rapid growth of the world's population are reflected in the overuse and scarcity of natural resources, deforestation, climate change and especially environmental pollution.[1]

According to the latest urban air quality statistics, 98% of cities in low and middle-income countries with more than 100,000 inhabitants do not meet the World Health Organization (WHO) air quality guidelines. A recent global atmospheric chemistry model study estimated that annually 4.5 million premature deaths are linked to outdoor air pollution as shown in Fig. 3, mainly due to fine particulate matter (PM2.5). Over the last decade, studies have shown that exposure to fine particulate air pollutants has adverse effects on cardiopulmonary health.[2].

Population size	Date	Time to add next billion
1 billion	1800	NA
2 billion	1927	127
3 billion	1960	43
4 billion	1974	14
5 billion	1987	13
6 billion	1999	12
7 billion	2011	12

Fig. 1 Population count against the year and number of years to increase one billion

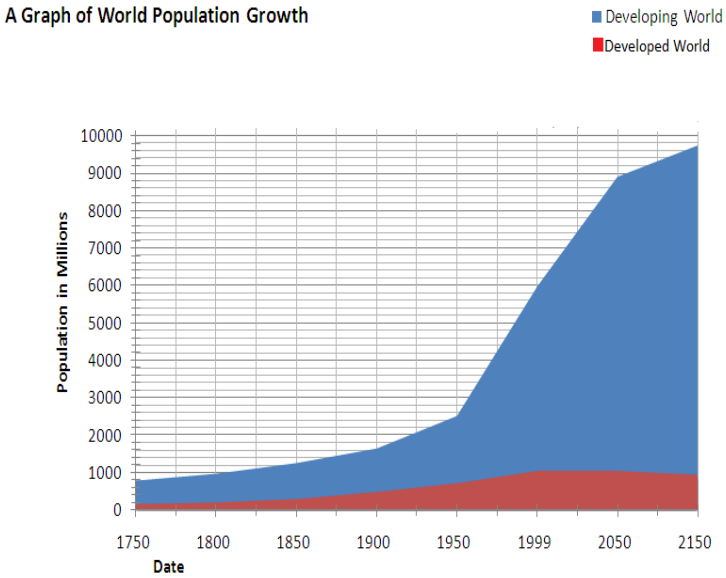

Fig. 2.Population growth in developing and developed countries.

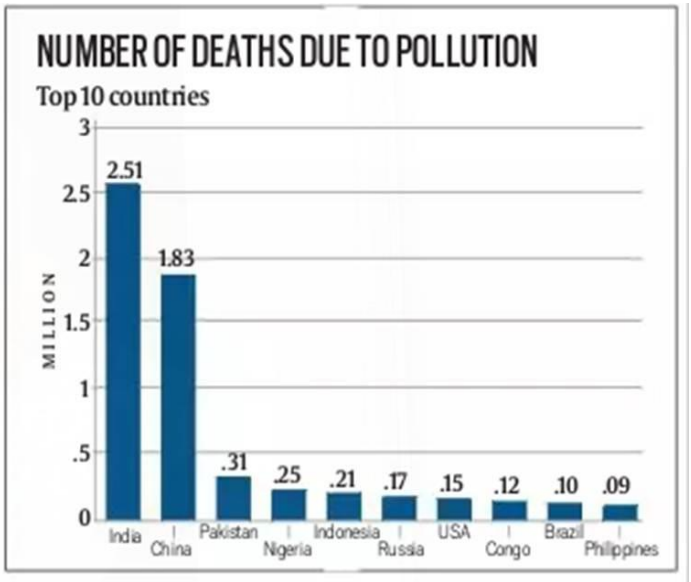

Fig. 3 Number of premature deaths due to pollution

Over the past years, pollution in Delhi has increased rapidly. This led to study and research in this area. Data mining was used to analyze the existing trends in air pollution in Delhi and make predictions about the future. The data mining techniques used are linear regression and multilayer perceptron.

Machine learning (ML) algorithms are investigated to build more accurate forecasting models. Machine ML algorithms such as support vector machines, M5P model trees, and artificial neural networks (ANN) were used.
Collaborative filtering along with Machine Learning can be combined to obtain a model that will provide highly accurate predictions.

Collaborative filtering proves to be more advantageous as content-based filtering is built around the attributes of a given object, while collaborative filtering relies on the behavior of many users and is more flexible.[5]

These predicted values can be used to assess a completely hypothetical situation before it occurs, such as deciding if an industry is eligible to be built in a place or changing the infrastructure of an existing building. The power of this is that the problem can be solved before it even exists.

2. TYPES OF AIR POLLUTANTS

There are various pollutants in the air, among which the ones majorly affecting pollution are [2] :-

I. Carbon monoxide: Carbon monoxide is a colorless, odorless gas. It is formed when fuels containing carbon are burnt in low-oxygen conditions.
Carbon monoxide reacts with other pollutants in the air to form potentially harmful ground-level ozone. It does not have any significant environmental effects at a global level. Inhalation of carbon monoxide at high concentrations can be fatal because it prevents the transport of oxygen in the blood around the body.

II. Sulphur dioxide: It is invisible and has a nasty, sharp smell. It is formed mostly by burning of fossil fuels particularly from power stations, converting wood pulp to paper, production of sulphuric acid, incineration of refused products and smelting.. Sulphur dioxide when inhaled irritates the nose, throat, and airways to cause coughing, wheezing, shortness of breath, or a tight feeling in the chest. The effects of sulphur dioxide are felt very quickly and most people would feel the worst symptoms in 10 or 15 minutes after inhaling it. Those most at risk of developing problems if they are exposed to sulphur dioxide are people with asthma or similar conditions.
III. Nitrogen Dioxide: Its presence in air contributes to the formation and modification of other air pollutants, such as ozone and particulate matter, and to acid rain. Studies on human populations indicate that long-term exposure to NO_2 levels may decrease lung function and increase the risk of respiratory symptoms such as acute bronchitis and cough and phlegm, particularly in children and also affect mortality. People with asthma and children, in general, are considered to be more vulnerable to NO_2 exposure.

IV. Particulate matter (PM): It is a mixture of solid particles and liquid droplets found in the air. Some particles, such as dust, dirt, soot, or smoke, are large or dark enough to be seen with the naked eye. Others are so small they can only be detected using an electron microscope.
Particle pollution includes:

PM_{10} : inhalable particles, with diameters that are generally 10 micrometers and smaller.

$PM_{2.5}$: fine inhalable particles, with diameters that are generally 2.5 micrometers and smaller..
Particulate matter contains microscopic solids or liquid droplets that are so small that they can be inhaled and cause serious health problems. Particles less than 10 micrometers in diameter pose the greatest problems because they can get deep into your lungs, and some may even get into your bloodstream.

3. EXISTING SYSTEM

In order to forecast the pollution, a data mining approach was followed. To implement this solution, the well-known machine learning workbench “Weka” was used. For each site, the data were composed of 2223 instances and 4 attributes. The instances correspond to the individual days when measurements were taken, each of which was used when running the machine learning algorithms (2223 days are approximately 6 years). The first 3 attributes are the features used to build the model. These parameters are level of precipitation, wind speed, and wind direction. Precipitation and wind speed are continuous variables whereas wind direction is a discrete variable (North, West, South, and East). The last attribute is the class attribute that we are interested in predicting. This attribute is the concentration of PM2.5 and is divided into 2 classes: concentrations above 15µg/m3 vs. concentrations below 15µg/m3. Here “Weka” tool is used which is more time and/or memory inefficient, and there is much less flexibility in terms of which optimization functions or numerical approximations are used.[1]

4. METHODOLOGY

The overall process can be explained in four steps- collection of data, prediction process, display of results, and notification to users.

This can be explained in the flowchart given in figure 4.

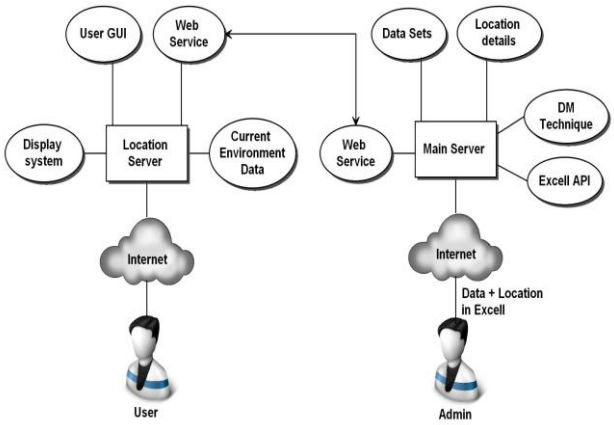

Fig. 4. System architecture

A) Collection of Data
This can be explained in following ways:

1) SITE DESCRIPTION:-

According to a survey by UN State of the World Population report in 2007, by 2030, 40.76% of country's population is expected to reside in urban areas. The statistic shows the degree of urbanization in India from 2006 to 2016 and details the percentage of the entire population, living in urban areas.

In 2016, approximately 33.14 percent of the total population in India lived in cities. India lies on the Indian Plate, the northern portion of the Indo-Australian Plate, whose continental crust forms the Indian subcontinent. The country is situated north of the equator between 8°44' to 37°6' north latitude and 68°7' to 97°25' east longitude. It is the seventh-largest country in the world, with a total area of 3,287,263 square kilometers (1,269,219 sq mi).India measures 3,214 km (1,997 mi) from north to south and 2,933 km (1,822 mi) from east to west. It has a land frontier of 15,106.7 km (9,387 mi) and a coastline of 7,516.6 km (4,671 mi). India has 40 cities with more than a million people, 397 cities with between 100,000 and 1 million people, and 2500 cities with between 10,000 and 100,000 people. The largest city in India is Mumbai, with a population of 12,691,836 people.[4]

2) Dataset collection:-

The data for the prediction is collected from CPBC. The Central Pollution Control Board (CPCB), is a statutory organization, was constituted in September 1974 under the Water (Prevention and Control of Pollution) Act, 1974.Further, CPCB was entrusted with the powers and functions under the Air (Prevention and Control of Pollution) Act, 1981.Principal Functions of the CPCB, as spelt out in the Water (Prevention and Control of Pollution) Act, 1974, and the Air (Prevention and Control of Pollution) Act, 1981,

(i) to promote cleanliness of streams and wells in different areas of the States by prevention, control and abatement of water pollution.
(ii) to improve the quality of air and to prevent, control or abate air pollution in the country.

Air Quality Monitoring is an important part of the air quality management. The National Air Monitoring Programme (NAMP) has been established with objectives to determine the present air quality status and trends and to control and regulate pollution from industries and another source to meet the air quality standards. It also provides background air quality data needed for industrial siting and towns planning.

Besides this, CPCB has an automatic monitoring station. At this station Respirable Suspended Particulate Matter (RSPM), Carbon Monoxide (CO), Ozone (O3), Sulphur Dioxide (SO2), Nitrogen Dioxide (NO2) and Suspended Particulate Matter (SPM) are being monitored regularly. This information on Air Quality at ITO is updated every week. This updated data is taken as a dataset for our machine learning which is given to the main server by the admin through the internet in the form of excel sheets.

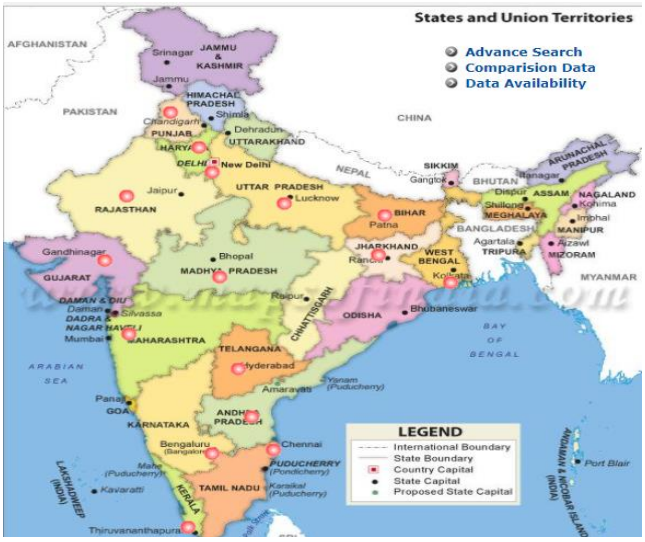

Fig.5. Pollutant monitoring locations in urban cities in Central pollution control board India

Parameters	Date	Time	Concentration	Unit	Concentration (previous 24 Hours)
PM2.5	15 Mar 2018	09:45:00	30.4	ug/m3	55.04
NO	15 Mar 2018	09:45:00	3.61	ug/m3	26.02
NO2	15 Mar 2018	09:45:00	15.15	ug/m3	31.79
NOx	15 Mar 2018	09:45:00	9.64	ppb	38
SO2	15 Mar 2018	09:45:00	4.69	ug/m3	63.86
CO	15 Mar 2018	09:45:00	0.57	mg/m3	1.04
Ozone	15 Mar 2018	09:45:00	39.55	ug/m3	30.45
Benzene	15 Mar 2018	09:45:00	0	ug/m3	0.87
Toluene	15 Mar 2018	09:45:00	1.15	ug/m3	0.38
Eth-Benzene	15 Mar 2018	09:45:00	0	ug/m3	11.89
MP-Xylene	15 Mar 2018	09:45:00	0	ug/m3	0.73
O Xylene	15 Mar 2018	09:45:00	0	ug/m3	0.13
RH	15 Mar 2018	09:45:00	68.38	%	42.94
WS	15 Mar 2018	09:45:00	0.94	m/s	-
WD	15 Mar 2018	09:45:00	93.97	deg	-
VWS	15 Mar 2018	09:45:00	0.01	degree	-
SR	15 Mar 2018	09:45:00	613.19	W/mt2	140.14
BP	15 Mar 2018	09:45:00	1012.8	mmHg	764.88

Fig. 6. Data set to be uploaded

B) Perform Prediction:-

The admin uploads the data of the particular location in the form of excel sheets. In the main server, the prediction is performed using the datasets provided, the location details, and the Data Mining technique. The algorithm we have used for prediction is Collaborative filtering.
(write about collaborative)

This process of prediction can be expressed as a flowchart in figure 7.

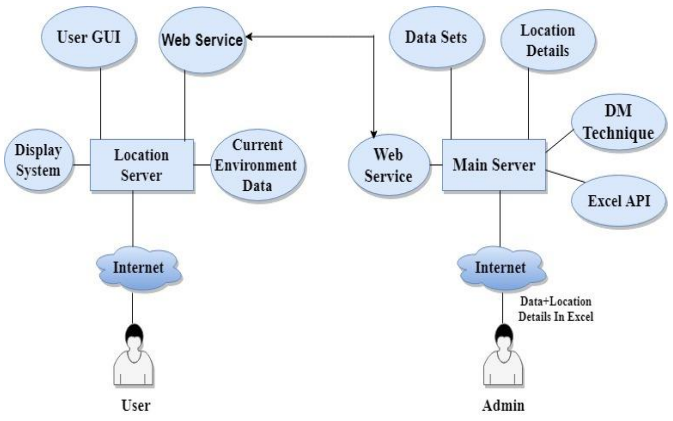

Fig. 7. Prediction process data flow diagram

The user requests for a particular location's air pollutants details. That requested location's pollution data is presented to the user. If the user requests for prediction of next few days, the selected locations dataset is sent to the main server. When the main server receives the dataset, it uses Data mining Techniques and performs prediction. Consider N number of pollutants. Let C=1 for the first pollutant. Using collaborative filtering the prediction for C^{th} parameter is done and stored in the database. The value of C is incremented to the next pollutant. This process iterates until all the pollutants prediction is complete.[3]

C) Display the Predicted Data:-

After the prediction of all the pollutants is done and stored in the database, the results are then displayed to the user.

D) Notification:-

The predicted data can be analyzed and if the pollution level goes above a limiting threshold, the registered users of that location are sent a notification. This notification can be used to take preventive measures. If this notification through email or SMS is sent to government pollution control authority, preventive rules can be implemented so keep the area clean and pollution free.

5. CONCLUSION

The objective of our work is not only to bring awareness but also to minimize pollution through preventive measures and ensure that vehicles are emitting the pollutants within the range of regular pollution check. This can lead to a pollution free area. To achieve this, collaborative filtering algorithm has played a major role by making automatic and accurate predictions based on previous trends of pollutant levels and database in the server.
If predicted values of concentration of the pollutant levels go above a threshold in a particular area, the registered users of that location are sent a notification which describes the accurate pollutant levels in that area.

6. REFERENCES

[1] R. Zalakeviciute, V. Diaz, and Y. Rybarczyk,"Machine Learning Approach to Forecasting Urban Pollution", IEEE Ecuador Technical Chapters Meeting (ETCM), 2016
[2] Shweta Taneja, Dr. Nidhi Sharma, Kettun Oberoi, Yash Navoria, "Predicting Trends in Air Pollution in Delhi using Data Mining", Bhagwan Parshuram Institute of Technology Delhi, 2016.
[3] Khaled Bashir Shaban , Abdullah Kadri , Eman Rezk, "Urban Air Pollution Monitoring System With Forecasting Models", IEEE Sensors Journal (Volume: 16, Issue: 8, April15, 2016)
[4]Airpollutionlevelsrisinginmanyoftheworld'spoorestcities,(online)available:http://www.who.int/mediacentre/news/releases/2016 /air pollution-rising/.
[5] R. Zalakeviciute, V. Diaz, and Y. Rybarczyk, "PM2.5 pollution management in Andean metropolis", in preparation.
[6] C.A. Pope, C. A., and D.W. Dockery, "Health effects of fine particulate matter air pollution: lines that connect", Journal of the Air & Waste Management Association, 2006.
[7] Ke Hu, Ashfaqur Rahman, Hari Bhrugubanda, and Vijay Sivaraman, "HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation from Fixed and Mobile Sensors" in Local Computer Networks Workshops (LCN Workshops),2016.
[8] Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, "The contribution of outdoor air pollution sources to premature mortality on a global scale", Nature, 2015.

