IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-11, Nov 2019 E-ISSN: 2347-2693

Analyses the Pollution Data and Suggest Measure to Reduce Pollution

Khan Salman Tasawar[1*], Khan Moin Mohd Umar[2], Piyush Kishore Vaidya[3], Shakila Shaikh[4]

[1,2,3,4]Computer engineering dept. RCOE, Rizvi College of Engineering, Bandra(W), Mumbai, India

**Corresponding Author: Khansalman09472@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i11.148153 | Available online at: www.ijcseonline.org

Accepted: 15/Nov/2019, Published: 30/Nov/2019

Abstract: Today pollution has been one amongst the foremost issues to cope with for any country. In South Asia it's hierarchical because the sixth most dangerous killer. One doesn't very understand the harmful impact of a pollution if he/she has not full- fledged within the initial place in 2016, a World Health Organisation (WHO) study found that fourteen of the twenty world's most contaminated cities belonged to Asian nation. Kanpur, in province, emerged because the town with the best PM2.5 level, standing at 173 (17 times more than the limit set for safety). Air pollution doesn't acknowledge geographical boundaries. even as contaminated air from rural areas travels into cities, cities too contribute towards rural pollution. Thus, it's crucial f or anti- pollution efforts to be coordinated across completely different levels. Urban-rural and inter-state responses square measure integral to crafting undefeated solutions. luckily, the govt of Asian nation (GoI) has well-versed the pollution epidemic with a nation-wide programme. this is often possible to own terribly positive impact on the health of all voters, particularly town dwellers. The Air Quality Life Index indicates that if national standards with relation to air quality square measure met, expectancy would go up by 2 years.

Keywords: *Dangerous killer, harmful, World health organisation(WHO), Government of India, pollution epidemic.*

I. INTRODUCTION

India is the 15 of the most contaminated urban areas on the planet as indicated by the World Health Organization (WHO) and Greenpeace are gracious the subcontinent. Air contamination affecting our life on everyday premise. The air quality is more regrettable by discharge from a loss to vitality plant. It's can have outcome enduring on the nation wellbeing. Over a million Indian kicked the bucket as indicated by the worldwide weight of sickness study however Indian

Government has reliably kept up there is no immediate co-connection between air contamination and passing. There is sufficient information accessible that air contamination kills science must proceed however activity ought not anticipate the aftereffect of logical examinations since that may take 10-20 years and at that point, they would have lost a large number of individuals pointlessly to this threat. Air contamination jumped at the chance to the third of all lung malignancy demise in India. The breathing is the most significant practice that we're doing each minute and that our wellspring of life. So the demeanour of we breathing if that is poor that certainly is a critical benefactor of any respiratory sickness.

India is one of the world's quickest developing economies and despite the fact that it's development has hinder plans to be worth 5 trillion dollar throughout the following couple of years yet as indicated by the world Bank the nation lost over 8% of its GDP in 2013 in light of the fact that the air contamination and most recent research from the Indian Statistical foundation shows that decreasing contamination would enable the nation to increase billion. Air contamination ruled the features in the nation in couple of years currently in any event, moving in the major ideological groups to incorporate it in their proclamations. The administration is making a push for clean vitality and electric vehicles among other measure. Study show that almost 700 million Indian are presented to contaminated air there is brief period to lose it.

The worldwide weight of ailment (GBD) appraisals, recorded outside air contamination among the main 10 wellbeing dangers in India. The investigation assessed 695,000 unexpected losses and loss of 18.2 million solid life years because of outside PM2.5 and ozone contamination (IHME, 2013). Among the wellbeing danger elements examined, open air contamination was positioned fifth in mortality and seventh in by and large wellbeing trouble in India. Family unit (indoor) air contamination from consuming of strong energizes was liable for an extra one million unexpected losses. A considerable increment was seen in the instances of ischemic coronary illness (which can prompt cardiovascular failures), cerebrovascular ailment (which can prompt strokes), ceaseless obstructive

aspiratory maladies, lower respiratory diseases, and malignant growths (in trachea, lungs, and bronchitis). A few different examinations have evaluated untimely death rates because of open air PM contamination for a few Indian urban communities, utilizing comparable strategies.

II. LITERATURE REVIEW

Particulates can be defined as dispersed matter that exists in the condensed phase (either solid or liquid) in which the individual units have sizes that range from 0.005 nm to about 500 fim. The size limits are somewhat arbitrary but are meant to indicate that particulate species can be as small as a cluster of several molecules or they can be as large as visible dust kernels. Particulate matter is extremely varied as to its chemical and physical properties. Very fine particulates behave almost like a gas or vapor; they are subject to Brownian motion, follow fluid streamlines, and are capable of coagulation and condensation. Larger particulates are more characteristic of solid matter; they are strongly influenced by gravity and seldom coalesce or condense.

The chemical behaviour of particulates can be determined by either the composition of the individual particles or by the types of gases that can adsorb on the surfaces of the particles. In some cases, the combination of particles and adsorbed gases produces a synergistic chemical effect more powerful than that of the individual components. Particulates are usually characterized as primary or secondary. Primary particulates are those produced as a direct result of the chemical or physical processes peculiar to any emission source. Secondary particulates, on the other hand, are those that are produced as a result of chemical reactions that occur in the atmosphere

These particulates can have sizes that range from molecular clusters having diameters on the order of 0.005 /tin to particles with diameters as large as several microns. Field studies of several urban aerosols13" 15 have shown that the highest concentrations of secondary particulates are usually found in the range 0.01 to 1.0 /an. The data also indicate this range can be divided into two subranges, one having mean diameters less than 0.05 /an and the other containing diameters between 0.05 and 1.0 /an. The smaller particles result directly from photochemical reactions, and the larger particles in turn, are produced from the coagulation or condensation of the photochemically generated particles. Both sets of particles vary with sunlight intensity, ozone concentration, etc.

The principal factors governing these size distributions, of course, are the respective rates of particulate formation and removal. The smallest particles have lifetimes on the order of several minutes before they coagulate with larger particles, but they are constantly being created in the daylight hours. The larger particles (i.e., those on the order of 0.01 to 0.1 /an) last about 12 hours16' 17 before they eventually coagulate or condense to form even larger particles. The overall life cycle of secondary particulates is difficult to assess; estimates range from a week to 40 days.18 The eventual loss mechanisms are rainout or coagulation with larger primary particles which eventually settle from the atmosphere.

The principal ingredients for forming secondary particulates are sunlight (and to some extent cosmic rays or terrestrial radiation) and chemicals such as SO2, NH3, NO2, H2O, and hydrocarbons. These chemicals are present in the atmosphere from both natural and man-made sources. Manmade sources include combustion systems, vehicle emissions, and industrial processes. Robinson and Robbins19 estimate that about 8% of the total global aerosol is provided by anthropogenic secondary particulates. Natural sources for secondary particulates include the oceans, volcanoes, geysers, as well as the vapor from trees, plants and decaying organic matter. About 56% of the total atmospheric aerosol is provided by secondary particulates resulting from natural sources

III. METHODOLOGY

The methodology we will use in next version of our project is open source software framework MongoDB framework and with the help of Naïve Bayes Algorithm. The Naïve Bayes algorithm is a classification technique which is based on Bayes Theorem. Naive Bayes model is easy to build and particularly useful for very large datasets [2]. Bayes theorem provides a way of calculating future probability P(c/x) from P(c), P(x) and P(x/c).Look at the equation below:

$$\mathrm{P(c|x)} = \frac{\mathrm{P(x|c)} * \mathrm{P(c)}}{P(x)} \quad (1)$$

Where, P(c/x) is future probability of class(c, target).P(c) is the prior probability of the class.P(x/c) is the likelihood which is the probability of predictor given class.P(x) is the prior probability ofpredictor.Similarly equation for predicting Pollution data of India from, is shown below:
P

$$\mathrm{P(SO2|Total)} = \frac{P(Total|SO2) * P(SO2)}{P(Total)} \quad (2)$$

$$\mathrm{P(NO2|Total)} = \frac{P(Total|NO2) * P(NO2)}{P(Total)} \quad (3)$$

After getting future probabilities of all the parameters if the probabilities of those parameters are greater than or equal to 70% then chances of India getting affected by SO2, NO2etc is most likely India pollution data. If probabilities of those parameters are lesser than or equal from 69% to 50% then India from pollution are likely to be not affected by pollution.

IV.DESIGN AND ANALYSIS

This topic includes various information and architecture diagram of our project our project measuring and analysis world bank data. It explains the working model of the project. Each block of the diagram is explained in detailed regarding the work it is implementing. It includes blocks like database, collection and document.

Database: In simple words, it can be called the physical container for data. Each of the databases has its own set of files on the file system with multiple databases existing on a single MongoDB server.

Collection: A group of database documents can be called a collection. The RDBMS equivalent to a collection is a table. The entire collection exists within a single database. There are no schemas when it comes to collections. Inside the collection, various documents can have varied fields, but mostly the documents within a collection are meant for the same purpose or for serving the same end goal.

Document: A set of key–value pairs can be designated as a document. Documents are associated with dynamic schemas. The benefit of having dynamic schemas is that a document in a single collection does not have to possess the same structure or fields. Also, the common fields in a collection's document can have varied types of data. The figure showing is given below:

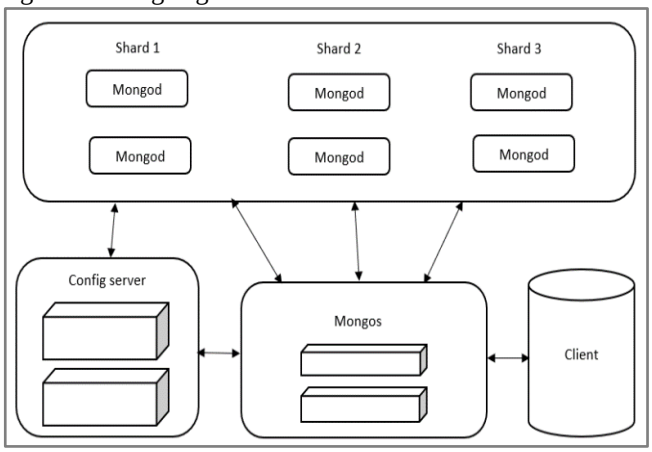

Fig 4.1:Architecture of MongoDB NoSQL Database

MongoDB is a NoSQL database, which enables building and running applications using modern development techniques. There may be data stored online as well as offline. The online data can be managed with MongoDB or RDBMS. For offline data, there is Hadoop or Enterprise Data Warehouse (EWD). You may have some applications like analytical tools, BI apps, mobile apps and CRM etc. Then there is data management activities with applications and infrastructure which consists of OS & virtualization, storage and network. The figure showing is given below:

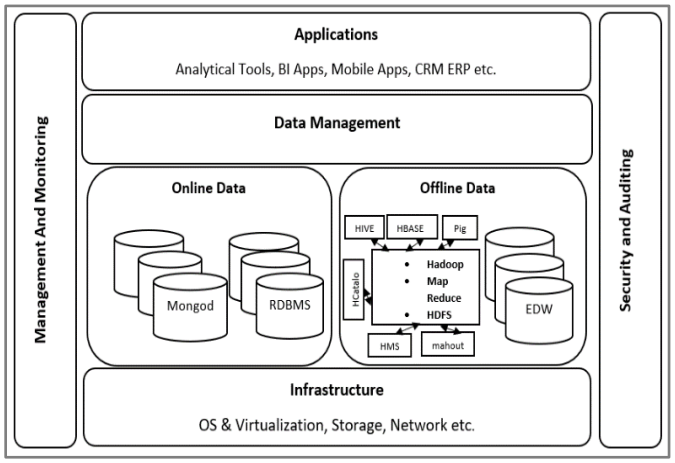

Fig 4.2: MongoDB High Level Architecture

Data Modelling in MongoDB-

The data in MongoDB has a flexible schema. Unlike in SQL databases, where you must have a table's schema declared before inserting data, MongoDB's collections do not enforce document structure. This sort of flexibility is what makes MongoDB so powerful.

When modelling data in Mongo, keep the following things in mind-

1. What are the needs of the application – Look at the business needs of the application and see what data and the type of data needed for the application. Based on this, ensure that the structure of the document is decided accordingly.
2. What are data retrieval patterns – If you foresee a heavy query usage then consider the use of indexes in your data model to improve the efficiency of queries.
3. Are frequent insert's, updates and removals happening in the database – Reconsider the use of indexes or incorporate sharding if required in your data modelling design to improve the efficiency of your overall MongoDB environment.

Key Components of MongoDB Architecture-

Below are a few of the common terms used in MongoDB

1. **_id** – This is a field required in every MongoDB document. The _id field represents a unique value in the MongoDB document. The _id field is like the document's primary key. If you create a new document without an _id field, MongoDB will automatically create the field. So for example, if we see the example

of the above customer table, Mongo DB will add a 24 digit unique identifier to each document in the collection.

2. **Collection** – This is a grouping of MongoDB documents. A collection is the equivalent of a table which is created in any other RDMS such as Oracle or MS SQL. A collection exists within a single database. As seen from the introduction collections don't enforce any sort of structure.
3. **Cursor** – This is a pointer to the result set of a query. Clients can iterate through a cursor to retrieve results.
4. **Database** – This is a container for collections like in RDMS wherein it is a container for tables. Each database gets its own set of files on the file system. A MongoDB server can store multiple databases.
5. **Document** - A record in a MongoDB collection is basically called a document. The document, in turn, will consist of field name and values.
6. **Field** - A name-value pair in a document. A document has zero or more fields. Fields are analogous to columns in relational databases.The following diagram shows an example of Fields with Key value pairs. So in the example below Years and 1991 is one of the key value pair's defined in the document.
7. **JSON** – This is known as JavaScript Object Notation. This is a human-readable, plain text format for expressing structured data. JSON is currently supported in many programming languages.

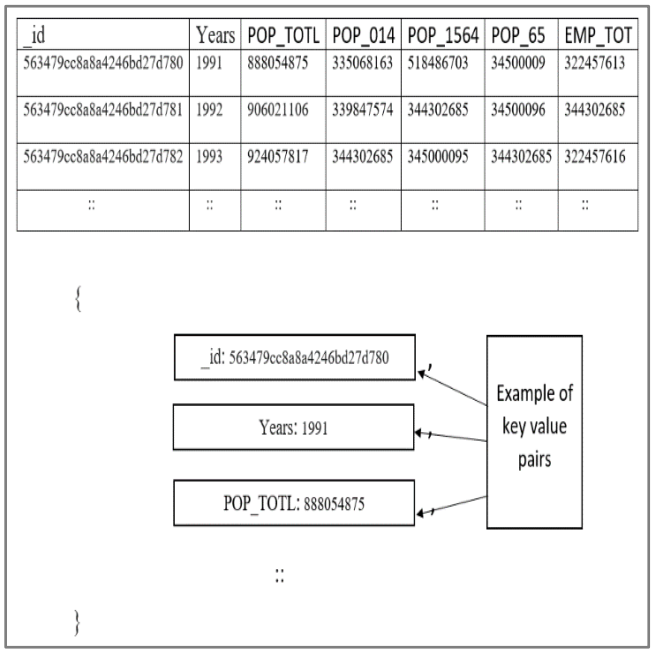

Fig 4.3:MongoDB High Level Architecture

V. RESULT AND ANALYSIS

The analysis and prediction of data of pollution in India using Studio3T is done successfully in the first version of project. MongoDB provides an engine that executes data flows on studio 3T. To expressing data flows it includes a language, MongoDB. It includes operators for many of the traditional data operations and it has the ability for users to develop their own functions for reading, processing, and writing data. Studio 3T of import the Data file name is "PollutionData" and makes the collection name is "pollutionview". The filename of "PollutionData" is 1990 to 2015 years in the prediction result and analysis in the execution of MongoDB.

1.db.pollutionview.find({},{"type":1,"_id":0}).sort({"type ":1})

After storing the dataset in MongoDB we have use Studio 3T find the total pollution in residential type ,industrial type and other type where the 1991 to 2016 years. After firing a query we have stored the result in some specific folder. The result of the query is shown in following graph:

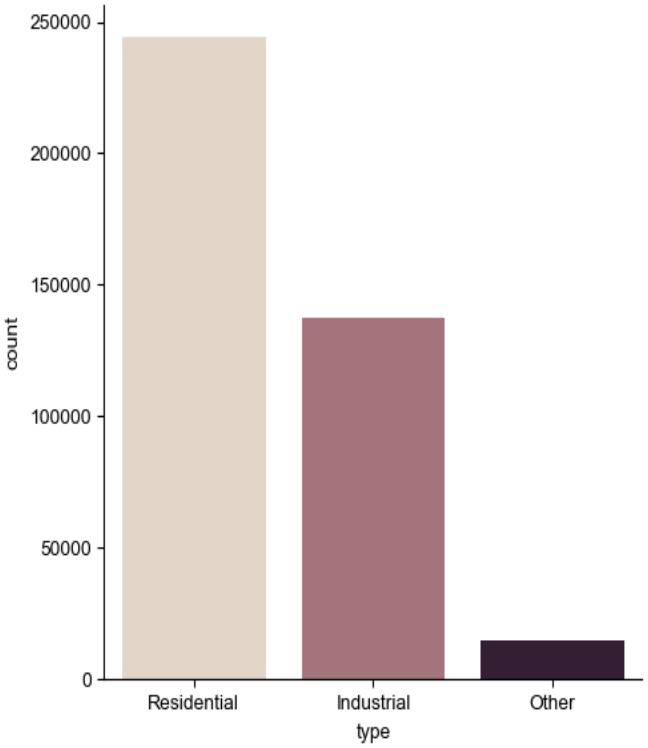

Fig 5.1:Analysis of Pollution Parameter

2.db.pollutionview.find({$and:[{"sampling_date":{$gte:1990}},{"so2":{$lte:45}}]})

After firing the query above Figure is showing the result of all SO2 concentration in all states in India which are populated. The result of the query is shown in following graph:

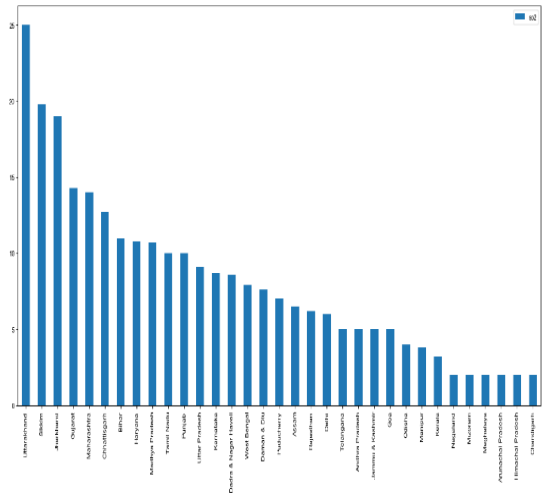

Fig 4.2: Analysis of Total so2 pollution in different cities in India.

3.db.pollutionview.find({$and:[{"sampling_date":{$gte:1990}},{"no2":{$lte:55}}]})
After firing the query above Figure is showing the result of all NO2 concentration in all states in India which are populated. The result of the query is shown in following graph:

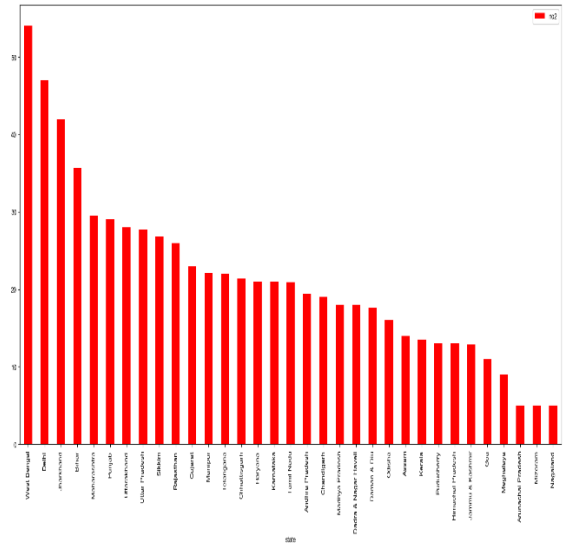

Fig 5.3: Analysis of total NO2 pollution in different cities in India.

4.db.pollutionview.find({$and:[{"sampling_date":{$gte:1990}},{"rspm":{$lte:300}}]})
After firing the query above Figure is showing the result of all Respirable suspended particulate matter or **RSPM** in all states in India which are populated. The result of the query is shown in following graph:

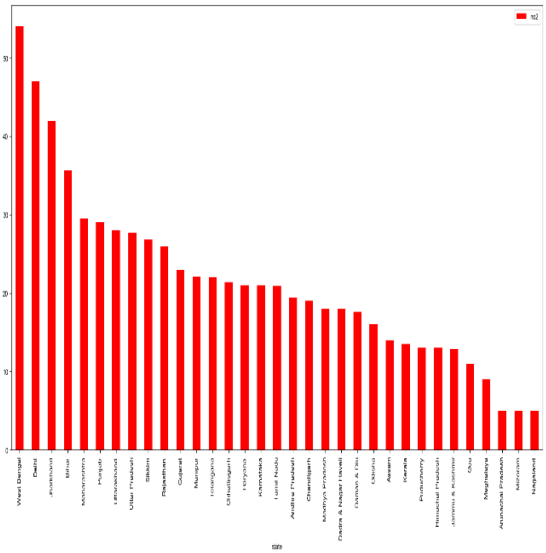

Fig *5.4*: Analysis of total *RSPM pollution in different cities in India*

5.db.pollutionview.find({$and:[{"sampling_date":{$gte:1990}},{"spm":{$lte:650}}]})
After firing the query above Figure is showing the result of all Suspended particulate matter or **SPM** in all states in India which are populated. The result of the query is shown in following graph:

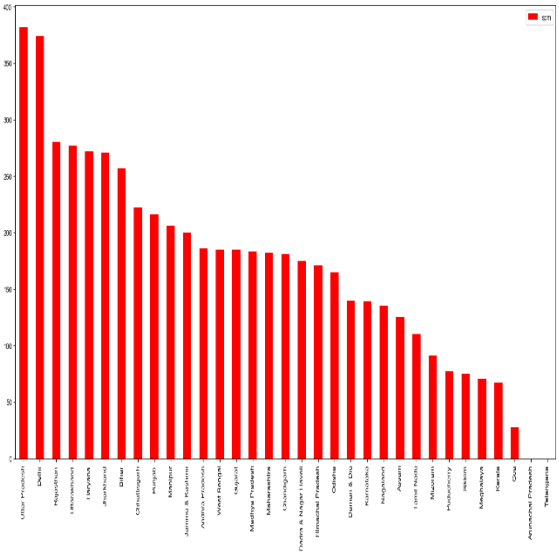

Fig 5.5: Analysis of total *SPM pollution in different cities in India*

6..db.pollutionview.find({$or:[{"sampling_date":{$gte:1990}},{"pm":""}]})
After firing the query above Figure is showing the result of all Particulate matter or **PM** in all states in India which are

populated. The result of the query is shown in following graph:

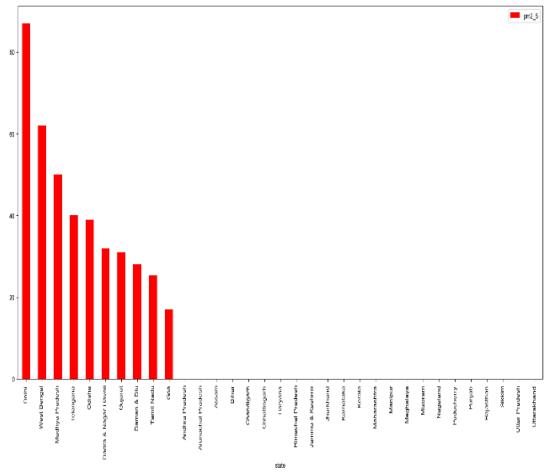

Fig5.6: Analysis of total *PM pollution in different cities in India*

VI. CONCLUSION

Thus, we have successfully found from the above analysis we see that the majorly affected states in India by air pollution belong to the northern region. States like Delhi, Punjab, Uttar Pradesh, Haryana are heavily polluted and require immediate action. Thus, we have successfully found different pollutant agent in from a given data set in different cities of India. we see that the majorly affected states in India by air pollution belong to the northern region. States like Delhi, Punjab, Uttar Pradesh, Haryana are heavily polluted. This is very useful method that can be used by Pollution Control Board. Future work of this project is to analysis the different types pollution condition and to predict and to work with other classification algorithm to become more accurate in analyzing.

REFERENCES

[1]. Tiwari, R. Sam and S. Shaikh, "Analysis and prediction of churn customers for telecommunication industry," *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, 2017,pp.218222.doi:10.1109/ISMAC.2017.8058343

[2]. S. Navadia, P. Yadav, J. Thomas and S. Shaikh, "Weather prediction: A novel approach for measuring and analyzing weather data," *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, 2017, pp. 414-417.
doi: 10.1109/I-SMAC.2017.8058382

[3]. S. Shaikh, S. Rathi and P. Janrao, "IRuSL: Image Recommendation Using Semantic Link," *2016 8th International Conference on Computational Intelligence and Communication Networks (CICN)*, Tehri, 2016, pp.305-308.doi: 10.1109/CICN.2016.66

[4]. S. Shaikh, S. Rathi and P. Janrao, "Recommendation System in E-Commerce Websites: A Graph Based Approached," *2017 IEEE 7th International Advance Computing Conference (IACC)*, Hyderabad, 2017, pp. 931-934.doi: 10.1109/IACC.2017.0189

[5]. Aarnio Pet al. (2005). The concentrations and composition of and exposure to fine particles (PM2.5) in the Helsinki subway system. *Atmospheric Environment*, 39(28):5059–5066.

[6]. Abbey DEet al. (1999). Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *American Journal of Respiratory and Critical Care Medicine*, 159(2):373–382.

[7]. Almeida-Silva M, Almeida SM, Freitas MC, Pio CA, Nunes T, Cardoso J. Impact of Sahara dust transport on Cape Verde atmospheric element particles. J Toxicol Environ Health A. 2013;76(4-5):240–51.

[8] Alnawaiseh NA, Hashim JH, Md Isa Z. Relationship between vehicle count and particulate air pollution in Amman, Jordan. Asia Pac J Public Health. 2012;27(2):NP1742–51.

[9]..Alolayan MA, Brown KW, Evans JS, Bouhamra WS, Koutrakis P. Source apportionment of fine particles in Kuwait City. Sci Total Environ. 2013;448:14–25.

[10]. Amador-Muñoz O, Bazan-Torija S, Villa-Ferreira SA, Villalobos-Pietrini R, Bravo-Cabrera JL, Munive-Colín Z, et al. Opposing seasonal trends for polycyclic aromatic hydrocarbons and PM_{10}: health risk and sources in southwest Mexico City. Atmos Res. 2013;122:199–212.

Authors Profile

Mr. Salman Khan currently pursuing Bachelor of Computer Engineering from University of Mumbai, India.He has done a extra courses of Python in college. His main interest is in Computer Hardware. He is currently pursing an ethical hacking course in udemy.

Mr. Moin Mohd Khan currently pursuing Bachelor of Computer Engineering from University of Mumbai, India. He is android developer and a web development consultant. His main interest is in gaming and game development.

Mr. Piyush Vaidya currently pursuing Bachelor of Computer Engineering from University of Mumbai, India.He has done a course in C/C++ and got ISO certificate.His main interest is in Music Production.