To Predict Pollution Levels in various Industrial Areas

Abstract - **Examining and protecting air quality in this world has become one of the essential activities for every human in many industrial and urban areas today. The meteorological and traffic factors, burning of fossil fuels, and industrial parameters play significant roles in air pollution. With this increasing air pollution, we need to implement models that will record information about concentrations of air pollutants. The deposition of these harmful gases in the air is affecting the quality of people's lives by altering their health, especially in urban areas.Our solution helps by contributing a small part to the decrease in pollution by regulation of its levels.**

I. Introduction

Air of cities, especially in the developing parts of the world is turning into a serious environmental concern. The air pollution is because of a complex interaction of dispersion and emission of toxic pollutants from manufacturing Industries. Air pollution caused due to the introduction of dust particles, gases, and smoke into the atmosphere exceeds the air quality levels. Air pollutants are the precursor of photochemical smog and acid rain that causes the asthmatic problems leading into serious illness of lung cancer, depletes the stratospheric ozone, and contributes to global warming. In the present industrial economy era, air pollution is an unavoidable product that cannot be completely removed but stern actions can reduce it. Pollution can be reduced through collective as well as individual contributions. There are multiple sources of air pollution, which are industries, fossil fuels, agro waste, and vehicular emissions. Industrial processes upgradation, energy efficiency, agricultural waste burning control, and fuel conversion are important aspects to reducing pollutants which create the industrial air pollution. Mitigations are necessary to reduce the threat of air pollution using the various applicable technologies like CO2 sequestering, industrial energy efficiency, improving the combustion processes of the vehicular engines, and reducing the gas production from agricultural cultivations.

India is a rapidly developing economy with interrelated air quality, sustainable development, and climate change mitigation goals. There are unique challenges to achieving each of these goals as well as potential trade offs among them.

India has some of the worst air quality in the world, making its improvement a high priority for the Indian government as evidenced by the launch of the National Clean Air Programme in January 2019

Hundreds of millions of people in India are continually exposed to toxic air: they inhale, for example, a 24-hour average of up to 25 micrograms/cubic metre of air of the deadly, microscopic pollutant, PM 2.5—far above the World Health Organization's (WHO) limit of 10 micrograms/cubic metre. With long-term exposure, this particulate matter goes deep into the lungs and on to other organs and systems, gradually defeating the body's defence mechanism. Repeated exposure to toxic air causes cardiovascular and respiratory diseases, lung and other cancers, strokes, preterm birth, type-2 diabetes, and other illnesses.

Since February 2014, the government of India has been monitoring industrial emissions and effluents in rivers and lakes across the country. The monitoring is done through what is called the Online Continuous Emissions/Effluents Monitoring Systems (OCEMS). The 17 categories of industrial units that are required to have OCEMS—these include power plants; aluminium, zinc, copper plants; cement plants; distilleries; fertilisers; iron and steel plants; oil refineries; petrochemicals; and tanneries. The emissions monitored under the OCEMS regulations include particulate matter (PM), CO (carbon monoxide), NOx (nitrogen oxide), SO2 (sulfur dioxide), and fluoride.

In mid-March 2020, discussions in Parliament indicated that there were some 3,700 OCEMS installed in different industrial locations across the country. A month earlier, Parliament was informed by the Union government that the total number of targeted units was 4,245.

Meanwhile, there are only 234 continuous air pollution monitors (also known as CAAQMS – Continuous Ambient Air Quality Monitoring Systems) in the country as of October 2020; the data from these monitors serve as the basis for the AQI or national Air Quality Index. By this yardstick, it is apparent that the scale of monitoring of pollutants is bigger in the country's industrial sector. The OCEMS network is regulated by the same regulatory body, the Central Pollution Control Board (CPCB), and monitors similar parameters as those covered by the CAAQMS. However, in the OCEMS, the commissioning and operations of the monitoring systems is left to the same industries which are themselves being monitored for their emissions.

While the installation and operations of CEMS equipment is highly technical, the understanding of the data from the CEMS network does not need to be as complex. The nuance of this is best understood by identifying who are the people most affected by the pollution from these industrial units where the CEMS equipment is installed. Both CAAQMS and CEMS are designed to benefit the people who are living in the vicinity of that monitoring system.

However, the data on industrial emissions is almost impossible to access for the public. Each industrial unit's operator has access to the data as well as the Central and state pollution control boards. The stated aim of OCEMS is for monitoring and self-regulation. But here is where the CPCB appears to sidestep transparency of OCEMS data.

Types of industries contributing to the higher pollution levels-

Agriculture

Agricultural activities produce emissions, which have the potential to pollute the environment. Ammonia (NH_3) and nitrous oxide (N_2O) are the key pollutants released from agricultural activities. The other agricultural emissions include methane emissions from enteric fermentation processes, nitrogen excretions from animal manure, such as CH_4, N_2O, and NH_3, methane emissions from wetlands, and nitrogen emissions from agricultural soils (N_2O, NO_X, and NH_3) due to the addition of fertilisers and other residues to the soil (Gurjar, Aardenne, Lelieveld. et al 2004). Agricultural processes, such as 'slash and burn' are prime reasons for photochemical smog resulting from the smoke generated during the process. Crop residue burning is another process that results in toxic pollutant emissions. This is how neighbouring cities of Delhi contribute to the agricultural pollution load. This is an example of how external sources contribute to the menace of air pollution in the city (Nagpure, Gurjar, Kumar, et al. 2016).

Power Plants

The contribution of power plants to air emissions in India is both immense and worrisome. The thermal power plants manufacture around 74% of the total power generated in India (Gurjar, Ravindra, and Nagpure 2016). According to The Energy and Resources Institute (TERI), the emissions of SO_2, NO_X, and PM increased over 50 times from 1947 to 1997. Thermal power plants are the main sources of SO_2 and TSP emissions (Gurjar, Aardenne, Lelieveld, et al. 2004), thereby contributing significantly to the emission inventories. In Delhi, power plants contributed 68% of SO_2 emissions and 80% of PM_{10} concentrations over a period from 1990 to 2000 (Gurjar, Aardenne, Lelieveld, et al. 2004). Thus, there is an urgent need to adopt alternative power sources including green and renewable resources for meeting the energy requirements.

Waste Treatment and Biomass Burning

In India, about 80% of municipal solid waste (MSW) is still discarded into open dumping yards and landfills, which leads to various GHG emissions apart from the issues of foul odour and poor water quality at nearby localities. The lack of proper treatment of MSW and biomass burning has been responsible for causing air pollution in urban cities. In Delhi alone, around 5300 tonnes of PM_{10} and 7550 tonnes of $PM_{2.5}$ are generated every year from the burning of garbage and other MSW (Nagpure, Gurjar, Kumar, et al. 2016).

Methane (CH_4) is the major pollutant released from landfills and wastewater treatment plants. Ammonia (NH_3) is another by-product, which is released from the composting process. The open burning of wastes, including plastic, produces toxic and carcinogenic emissions, which are a grave concern (Gurjar, Aardenne, Lelieveld, et al. 2004).

Domestic Sector

Households are identified as a major contributor of air pollution in India. The emissions from fossil fuel burning, stoves or generators come under this sector, thereby affecting the overall air quality. Domestic energy is powered by fuels, such as cooking gas, kerosene, wood, crop wastes or cow dung cakes (Gurjar, Aardenne, Lelieveld, et al. 2004).

Though liquefied petroleum gas (LPG) is used as an alternative source of cooking fuel in most urban cities, the major share of India's rural population continues to rely on cow dung cakes, biomass, charcoal or wood as the primary fuel for cooking and other energy purposes and demands. These lead to severe implications on air quality, especially the indoor air quality, which could directly affect human health. According to HEI (2019), about 60% of India's population was exposed to household pollution in 2017.

Construction and Demolition Waste

Another major source of air pollution in India is waste, which is an outcome of construction and demolition activities. Guttikunda and Goel (2013) inferred from their study that around 10,750 tonnes of construction waste is generated in Delhi every year. Even after the construction phase, these buildings have the potential to be the major contributors of GHG emissions. Nowadays, the increasing interest in green building technologies and the application of green infrastructure and materials during construction could tackle this issue to a large extent, thereby preserving our biodiversity and maintaining cleaner air quality.

II. Related Work

Air Quality Prediction Through Regression Model by A.Aarthi, P.Gayathri, N.R.Gomathi , S.Kalaiselvi , Dr.V.Gomathi

This research paper was close to the best thing we referred to while operating on our dataset.

Our Analysis precedes the latter in terms of the analytical and mathematical approach but far succeeds the the latter in terms of graphical approach and visualization.This makes it easier to understand for any individual and operate the dataset to fulfill our problem statement.

This paper mainly dealt with arithmetic description on how regression analysis(linear) could be done on the dataset

Our method of using code to show that makes it easier to understand regression and to implement it.

Results for other types of analysis couldn't even come close to our speculations/predictions of what we wanted the outcome as.Regression being the widely and most use algorithm proved to be useful to us in such a manner

III. Dataset and Features

In the proposed system, the air quality dataset is downloaded, which is available in CSV format. The comma separated value data format can easily be processed and analyzed fast using a computer and the data utilized for various purposes. It is imported to the project by using a panda package available in google colab and Jupyter software. The dataset contains 15 important attributes that help in air quality prediction. Initially, the dataset is preprocessed with suitable techniques to remove the inconsistent and missing valued data, and the needed features from the dataset are selected for better results.

The air quality dataset for this project is collected from the UCI repository. The dataset is available in CSV format. It is downloaded and imported to the project by mentioning the location of a downloaded dataset using the panda package available in Jupyter and Colab. The dataset contains data of average hourly responses of different elements in the air for nearly one year from 2016 to 2018. Dataset consists of 2115 rows and 20 columns

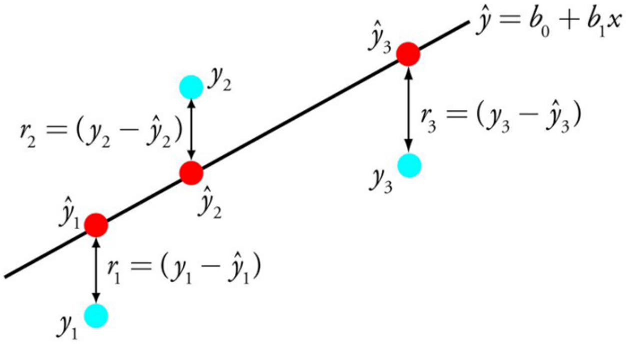
Column	Description
State	Name of the state
City	Name of the city
Location	Location in the city where the recording was taken
{Pollutant} Monitoring days / year	Number of days that pollutant's recordings were taken that year.
{Pollutant} Minimum	Minimum value of that pollutant that year
{Pollutant} Maximum	Maximum value of that pollutant that year
{Pollutant} Annual Avg	Average value of that pollutant that year

In the Dataset pollutant gasses of SO2,NO2,PM10,PM25 (PM is Particulate matter,10,15 is the size) have been mentioned of various regions of states of the Indian subcontinent.On this matter pre-processing has been done to prepare the data for analysis.

Architecture Methods and Learning Algorithms

Linear Regression

- Linear regression is perhaps one of the most well-known and well-understood algorithms in statistics and machine learning
- Predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explainability. We will borrow, reuse and steal algorithms from many different fields, including statistics and use them towards these ends.
- The representation of linear regression is an equation that describes a line that best fits the relationship between the input variables (x) and the output variables (y), by finding specific weightings for the input variables called coefficients (B).



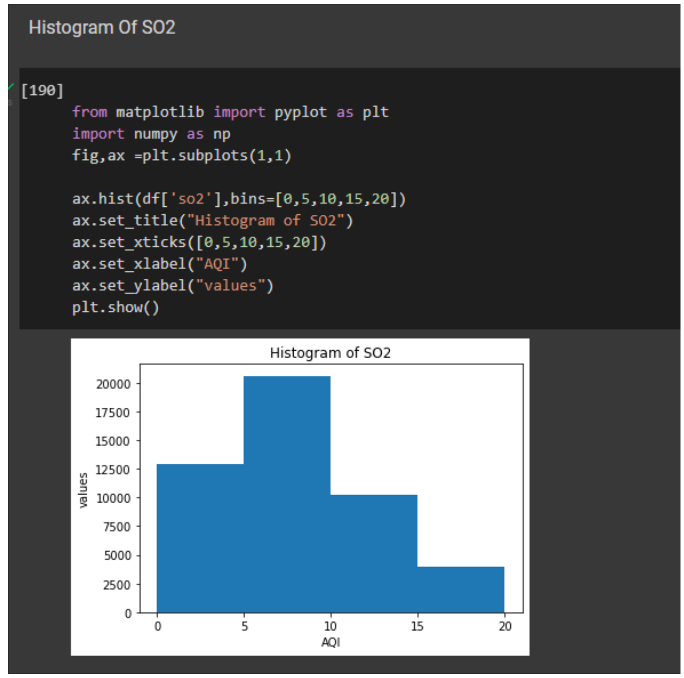
For example: y = B0 + B1 * x

- We will predict y given the input x and the goal of the linear regression learning algorithm is to find the values for the coefficients B0 and B1.
- Different techniques can be used to learn the linear regression model from data, such as a linear algebra solution for ordinary least squares and gradient descent optimization.

- Linear regression has been around for more than 200 years and has been extensively studied. Some good rules of thumb when using this technique are to remove variables that are very similar (correlated) and to remove noise from your data, if possible

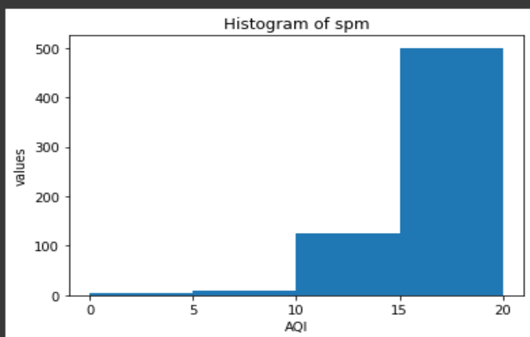
Experiments and Inferences

Redundancy and Correlation Analysis



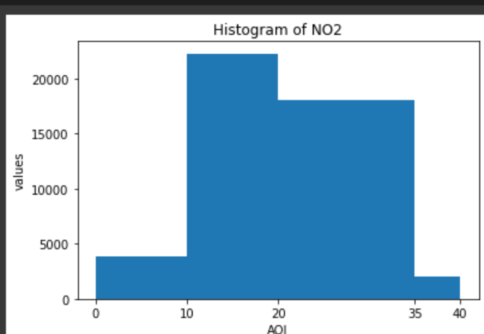
Histogram of spm

```
[192] from matplotlib import pyplot as plt
      import numpy as np
      fig,ax =plt.subplots(1,1)

      ax.hist(df['spm'],bins=[0,5,10,15,20])
      ax.set_title("Histogram of spm")
      ax.set_xticks([0,5,10,15,20])
      ax.set_xlabel("AQI")
      ax.set_ylabel("values")
      plt.show()
```



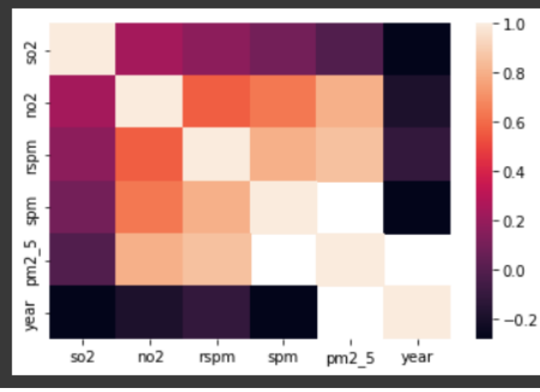
Histogram of NO2

```
[191] # Histogram for NO2
      from matplotlib import pyplot as plt
      import numpy as np
      fig,ax =plt.subplots(1,1)

      ax.hist(df['no2'],bins=[0,10,20,35,40])
      ax.set_title("Histogram of NO2")
      ax.set_xticks([0,10,20,35,40])
      ax.set_xlabel("AQI")
      ax.set_ylabel("values")
      plt.show()
```



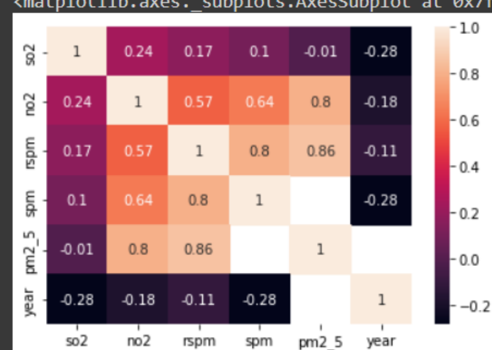
Heatmap

```
sns.heatmap(df.corr(method='pearson'))
plt.show()
plt.savefig("heatmap_pearson.png")
plt.clf()
plt.close()
```

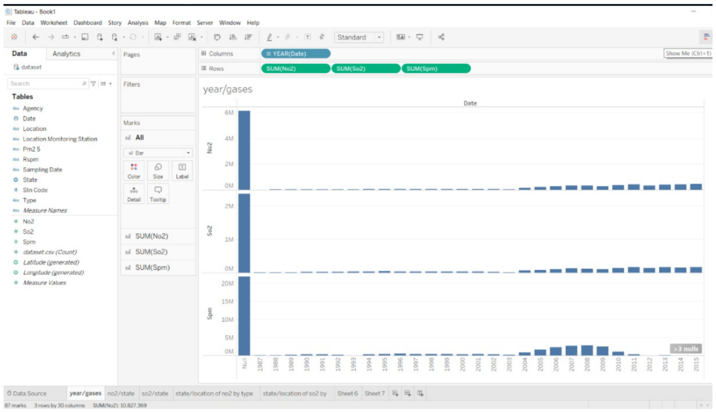


Correlation Matrix

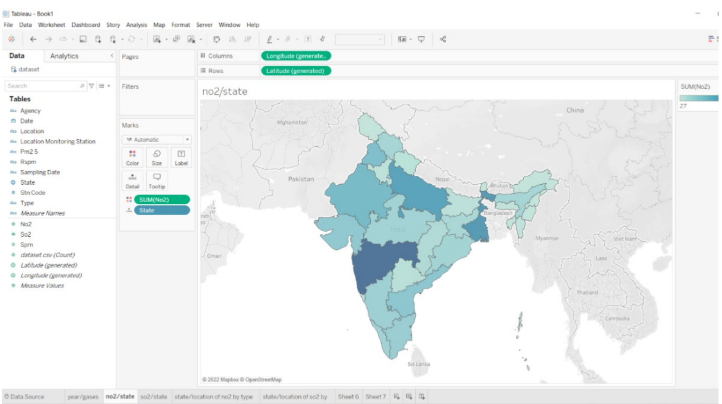
```
[198] correlation_matrix=df.corr().round(2)
      sns.heatmap(data=correlation_matrix,annot=True)

<matplotlib.axes._subplots.AxesSubplot at 0x7f81b8d186d0>
```



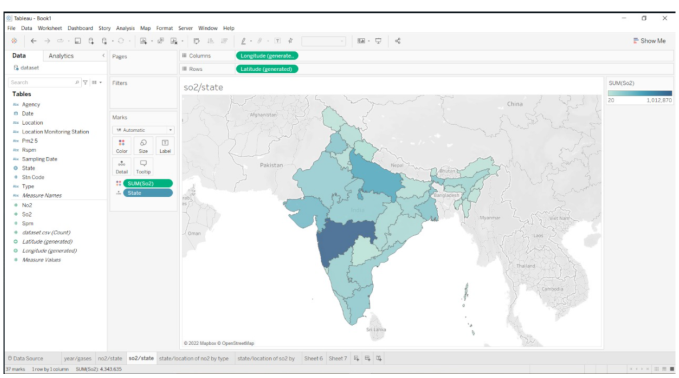
Graphs Made Using Tableau



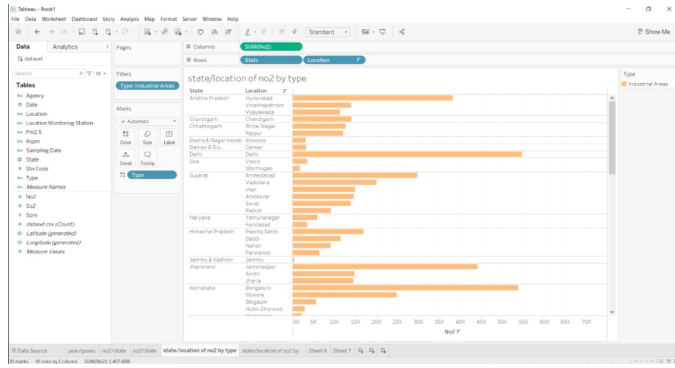
The above graph represents the absolute cumulative amount of gasses in th entire Indian subcontinent



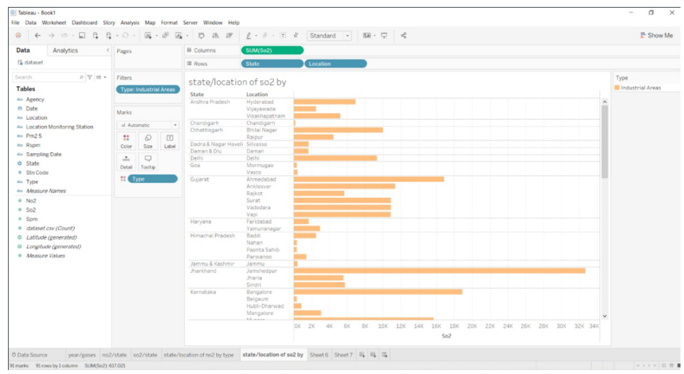
The above graph represents the gross amount of pollution(NO_2) in various states. Darker shade indicates more pollution.



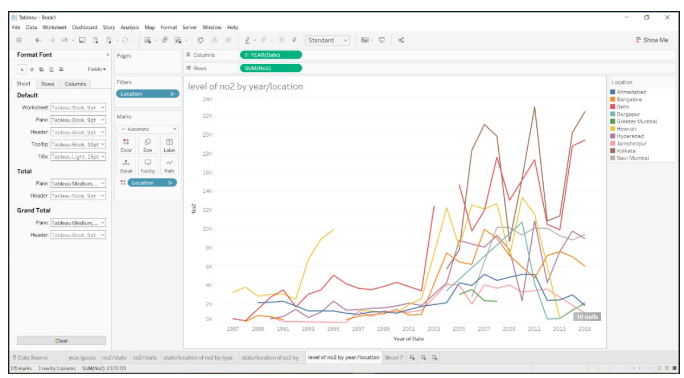
The above graph represents the gross amount of pollution(SO_2) in various states. Darker shade indicates more pollution.



The above graph represents the sum of AQI (NO_2) of various cities in various states of India.(Values are in larger numbers as it is the absolute sum of AQI but still are approximately very close to the real averages when divided by 1000).



The above graph represents the sum of AQI (SO_2) of various cities in various states of India.(Values are in larger number as it is the absolute sum of AQI but still are approximately very close to the real averages when divided by 1000).



The above graph represents the total level of NO2   in various major cities in India by year

III. Result

The main aim of this paper is to propose a cheap and feasible automated attendance system using the iris as the biometric for enrolment and identification. The attendance system proposed by this paper is web based. The designed prototype showed a lot of promise and presents an alternate and more accurate method of taking attendance that is quite spoof proof and relatively cheap to implement

VI. Conclusion

Thus with the help and aid of The Linear Regression Algorithm we were able to perform operations on the dataset which helped us to achieve the goal that our problem statement posed in the first place.Thus with the help of our analysis the process of detection of higher levels of pollution has become much easier and it would contribute greatly to safeguard the environment as well as the human race.

VII. Future Work

Our project if taken to a larger scale will be very helpful due to the fact that ever-growing pollution problems will pose a major threat to humans and all other living organisms soon in the future.If an analysis of this type gets implemented on a larger scale it would be easier to cutback on the sources of major air pollution and will help us all.

VIII. References

1. Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu "Detection and Prediction of Air Pollution using Machine Learning Models." (IJETT) – Volume 59 Issue 4 – May 2018 Baldasano J.M, Akita Y "Large scale Air pollution estimation method combining land use regression modeling in geostatistical framework," Environmental Science & Technology, vol. 48, no. 8, 2014
2. GnanaSoundari.A Mtech, (Ph.D.), Mrs. J.GnanaJeslin M.E, (Ph.D.), Akshaya A.C. "Indian Air Quality Prediction And Analysis Using Machine Learning." ISSN 0973-4562 Volume 14,Number 11, 2017
3. McCollister G.M. and Wilson K.R. (2008), "Linear regression model for forecasting daily maxima and hourly concentrations of air pollutants," Atmospheric Environment.
4. Rao, S.T., and Zurbenko, I.G.(2014)."Detecting and Tracking Changes in Air Quality using regression analysis". J. Air Waste Manage. Assoc. 44: 1089–1092.
5. RuchiRaturi, Dr. J.R. Prasad "Recognition Of Future Air Quality Index Using Regression and Artificial Neural Network" IRJET .e-ISSN: 2395- 0056 p-ISSN: 2395-0072 Volume: 05 Issue: 03 Mar-2018
6. Y. Zheng and et al., "Predicting Fine-Grained Air Quality Based on Linear regression," KDD '15, May 2011 Proceeding of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining