International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072

# Recognition of Future Air Quality Index Using Artificial Neural Network

**Ruchi Raturi[1], Dr. J.R. Prasad[2]**

[1]*Student, Department of Computer Engineering, SCOE, Vadgaon, Pune, Maharashtra*
[2]*Professor, Department of Computer Engineering, SCOE, Vadgaon, Pune, Maharashtra*

---

**Abstract**- *Due to a major increase in pollution day by day so it is required to predict pollution of the next date, next months, next year. Using some previous air related data. Air pollution is rapidly increasing due to various human activities, and it is the introduction into the atmosphere of chemicals, particulates, or biological materials that cause loss of human lives, and it also harms the natural environment. Indeed, air pollution is one of the important environmental Problems in metropolitan and industrial cities. So it's very important to predict pollution and avoid these problems.*

*Air pollution prediction using data mining is one of the most interesting and challenging tasks. Many systems are designed to support air pollution data storing, inventory management and generation of simple statistics. Some systems use decision support systems, but they are largely limited. They can answer simple queries like "What is the maximum limit of air pollution", "which area has a maximum pollution" However, they cannot answer complex queries like "Predict next month air pollution count.", "Given me, tomorrows pollution details" this type of prediction techniques are used in this system.*

***Key Words:*** **Air quality prediction, SO2, NO2, Ozone, RSPM, Model Generation, Multilayer Perceptron(MLP), Artificial Neural Network(ANN), Linear Regression**.

## 1. INTRODUCTION

Air pollution is rapidly increasing due to various human activities, and the occurrence of particulates, chemicals or biological resources into the environment that cause unexpected, humans death, or disease, damage source of revenue, or spoil the natural environment. In reality, pollution content in the air is most vital environmental problems in developed and urban cities. So its very important to predict pollution and avoid these problems. Air pollution calculation is one amongst the demanding tasks and we give the prediction techniques used to give next day, next month air pollution count to avoid the problems.

The environment is affected in terms of global climate change and adverse effects on plants and ecosystems due to urbanization in recent years. Vehicles are a significant source of emissions into the atmosphere. The commonly occurred air pollutants are CO, NO, NO2, PM 10, O3, SO2 and several organic compounds. These air pollutant cause hazardous effects on the ecological system of a human being such as a disease, discomfort or death to humans, damage to other living organisms like food harvest, or spoil the natural environment. Hence there is need to monitor air pollution. Many decision support systems are designed for monitoring the data but they are largely limited.

### 1.1 Problem Statement

Air pollution is rapidly increasing due to various human activities, and it is the introduction into the atmosphere of chemicals, particulates, or biological materials that cause discomfort, disease, or death to humans, damage other living organisms such as food crops, or damage the natural environment or built environment. Indeed, air pollution is one of the important environmental problems in metropolitan and industrial cities. So it's very important to predict pollution and avoid these problems.

Air pollution prediction using data mining is one of the most interesting and challenging tasks and we give the prediction techniques used to give next day, next month air pollution count to avoid the problems.

## 2. RELATED WORK

Suganya E and Vijayashaarathi [1] proposed a system for the moving vehicles is monitor the NO2, Humidity, Temperature, CO levels of air contamination by using the NO2 sensor, Humidity sensor, Temperature sensor, CO sensor. MANET (Mobile Ad Hoc Network) routing algorithm is used which has nearly 28 mobile nodes(Vehicles) provide a coverage area of 300meters around the city.

Krzysztof SIWEK and Stanislaw OSOWSKI [2] analyze and compare two methods of feature selection. One applies the genetic algorithm, and the second the linear method of stepwise fit. On the basis of such analysis, it is possible to select the most important features influencing the prediction. As a mathematical tool for final prediction, they apply the neural networks.

Niharika et al [4] give a different type of numerical as well as statistical tools for the prediction and analysis of air quality, but Artificial Neural Network is considered to be an excellent predictive and data analysis tool for Air quality forecasting. The paper gives an inclusive review of existing air quality forecasting techniques through soft computing.

KRZYSZTOF SIWEK et al. [5] applies a genetic algorithm and a linear method of stepwise fit [2] and depending on that two sets of the most predictive features are selected. Two approaches to prediction are compared i.e. features selected are directly applied to the random forest (RF), [5]which forms an ensemble of decision trees. and intermediate predictors built on the basis of neural networks.

Shweta et al.[6] used the data mining techniques like linear regression and multilayer perceptron. Trends of various air

pollutants like sulfur dioxide(SO2), nitrogen dioxide (NO2), particulate matter (PM), carbon monoxide (CO), ozone (O3) is analyzed.

Lijun et al [7] used the K-means clustering and Logistic model and forecasted the carbon emissions in 30 provinces and autonomous regions in China from 2014 to 2023. K-means cluster analysis method was used to divide the carbon emission into 5 types. Finally, the Logistic model of carbon emissions growth was built, to predict the carbon emissions.

Calculation of the pollutant deliberation is given by using multiple linear regressions (MLR) and multilayer perceptron neural networks (MLP NN). A comparatively recent way, extreme learning machine (ELM),[8] as well accustomed to conquer the limitation of linear methods as well as the large computational demand of MLP NN.

Zhu and Peng [9] shown that urbanization had distinct positive effects on carbon emissions. Yi et al. [10] analyzed the relationships among the carbon footprint and other elements, including population, level of economic development, industrial structure, and energy structure.

Fan et al. [11] use the STIRPAT representation to understand the effect of inhabitants, wealth, and technology on total CO2 emissions of countries at different income levels from 1975 to 2000. The result indicated that the effect of population on CO2 emissions is especially great in low income per capita countries.

## 3. PROPOSED SYSTEM

The figure shows the architecture diagram of the proposed system. This system will use the Linear regression and Multilayer Perceptron (ANN)Protocol for prediction of the pollution of next day.The working of the proposed system is as follows:

**1. Fetching Pollution Data :**

Get data from https://data.gov.in/ this website. Fetching some parameters like SO2, NO2, PM, Ozone, Air Quality and so on. Parameters used are "SO2","NO2","RSPM (Respirable Suspended Particulate Matter) /PM10 (Particulate Matter)", "Particulate Matter /PM 2.5". Each parameter will be used for analysis of pollution in the city.

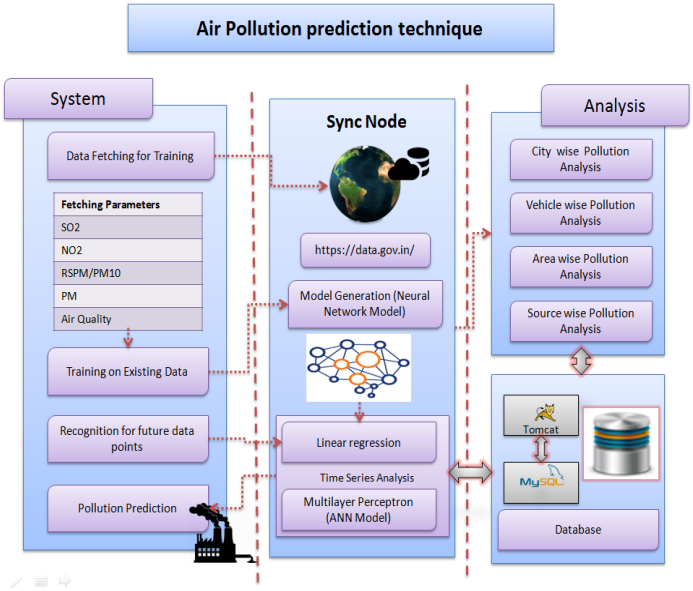


Fig-1: Diagram for Proposed System

a. SO2 is mostly formed by the burning of fossil fuel

b. no2 is produced during burning of fuel under high temperature

c. RSPM - pm2.5 - pm10 - is a complex pollutant and consists of a variety of components.

d. O3 - is formed when a chemical reaction of volatile organic compounds and no occurs in sunlight.

**2. Training on Existing Data:**

Existing data parameters along with their air quality parameter will be provided in training set. The output of the training set will be AIR Quality index. Data has to be provided in time series in the form of ( SO2, NO2, RSPM, PM, air quality - day wise).

**3. Model Generation for Multilayer Perceptron**

Get historical data from data.gov.in website and train this data using multilayer perceptron (MLP) is an artificial neural network (ANN) model that plots set of input data into a set of appropriate outputs. And generate a model for future use.

**4. Recognition for future data points using Time Series Analysis**

Using MLP(ANN) model for recognition for future data points. Based on the trained model, new values will be provided to get the next day AIR quality index prediction. Also, you can predict next month next year wise air quality prediction using this MLP model.

**5. City wise Pollution Analysis**

Display city wise Air quality index prediction. We store all historical data into the database at the time of fetching data and prediction of an air quality index. And using this data we provide city wise analysis graphs.

**6. Vehicle wise Pollution Analysis**

Finding correlation of vehicle wise stats Its very important because you don't know which vehicle is most affected, so this vehicle wise pollution analysis is done in this system.

**7. Area wise Pollution Analysis**

The Area Wise (Colleges, Hospital ..., and so on), State Wise, Country wise would help to take precautions for the people leaving in a particular region to avoid the problems.

## 4. ALGORITHM USED

### A. Linear Regression

Linear regression is a numerical way that consent to review and revise interaction among two continuous variables.

Linear regression is a linear approach to modeling the relationship between a scalar dependent variable y and one or more explanatory variables denoted X.

The case of one descriptive inconsistent value is called simple linear regression. For further than one variable, the process is called multiple linear regression.

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models.

The smallest square waning line intended for the set of n data points is specified by,

y = ax + b

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j}{\sum_{i=1}^{n}(x_i^2) - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}$$

$$= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\mathrm{Cov}[x,y]}{\mathrm{Var}[x]} = r_{xy}\frac{s_y}{s_x},$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x},$$

### B. Multilayer Perceptron (ANN)

A multilayer perceptron (MLP) is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. An MLP contains numerous coating of nodes in a focussed graph, with every layer completely linked to the subsequent one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP make use of a supervised learning method described as back propagation for training the network. MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable. Activation Function:

While a multilayer perceptron has a linear activation function in the entire neurons, that is, a linear function that maps the weighted inputs to the output of each neuron, then it is easily proved with linear algebra that any number of layers can be reduced to the standard two-layer input output model.

$$y(v_i) = \tanh(v_i) \quad \text{and} \quad y(v_i) = (1 + e^{-v_i})^{-1},$$

In which the former function is a hyperbolic tangent which ranges from -1 to 1, and the latter, the logistic function, is similar in shape but ranges from 0 to 1. Here is the output of the node (neuron) and is the weighted sum of the input synapses.

a. Layers

The multilayer perceptron consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes and is thus considered a deep neural network

b. Learning through back propagation

Learning to take place in the perception by altering correlation weights subsequent to every part of data is processed, based on the amount of error in the output compared to the expected result. This is an example of supervised learning and is carried out through back propagation, a generalization of the least mean squares algorithm in the linear perception.

- We signify the fault in output node $j$ in the $n$th data direct by

$$e_j(n) = d_j(n) - y_j(n),$$

- Where $d$ the target is rate and $y$ is the value produced by the perceptron. We then make corrections to the weights of the nodes based on those corrections which minimize the error in the entire output, given by

$$\mathcal{E}(n) = \frac{1}{2}\sum_{j} e_j^2(n).$$

Using gradient descent, we find our change in each weight to be

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} y_i(n)$$

Where $y_i$ is the output of the previous neuron and $\eta$ is the learning rate.

- The derivative to be calculated depends on the induced local field $v_j$, which itself varies. It is simple to show that for an output node this plagiaristic can be simplify to

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = e_j(n)\phi'(v_j(n))$$

where $\phi'$ is the derivative of the activation function described above. Conventionally, the input layer is layer 0, and when we talk of an N layer network we mean there are N layers of weights and N non-input layers of dealing units

## 5. CONCLUSIONS

In this scheme, we are receiving data by Multilayer Perceptron (MLP) which is an Artificial Neural Network (ANN) to identify the future data points using Time Series Analysis. Based on the trained model, new values will be provided to get the upcoming day Air Quality Index prediction. Also, we can predict subsequently month, subsequently year wise air quality prediction.

The system helps to predict next date pollution details based on basic parameters and analysing pollution details and forecast future pollution. Our system gives solution for air pollution prediction. However, it also provides relevant information about the areas people are actually interested in.

## REFERENCES

[1] Suganya E, Vijayashaarathi S,"Smart Vehicle Monitoring System for Air Pollution Detection using Wsn",International Conference on Communication and Signal Processing. April 6-8, 2016, India..

[2] Krzysztof SIWEK,Stanislaw OSOWSKI, "Data mining methods for prediction of air pollution".

[3] Sandeep Kaur,Dr. Sheetal Kalra"Disease Prediction using Hybrid K-means and Support Vector Machine",2016 IEEE.

[4] Niharika, Venkatadri M,"A survey on Air Quality forecasting Techniques",International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 103-107.

[5] KRZYSZTOF SIWEK a, STANISŁAW OSOWSKI"DATA MINING METHODS FOR PREDICTION OF AIR POLLUTION",Int. J. Appl. Math. Comput. Sci., 2016, Vol. 26, No. 2, 467–478. DOI: 10.1515/amcs-2016-0033.

[6] Shweta Taneja1,Dr. Nidhi Sharma2"Predicting Trends in Air Pollution in Delhi using Data Mining",2016 IEEE.

[7] Lijun Ma, Kangqing Lin, Meijiao Guan"The prediction of carbon emission in all provinces of China with the K-means cluster based Logistic model".

[8] Huiping Peng"Air Quality Prediction by Machine Learning Methods".

[9] Zhu, Q., Peng, X., 2012. The impacts of population change on carbon emissions in China during 1978-2008. Environ. Impact Assess. Rev. 36, 1-8.

[10] Yi, B., Chen, G., Lu, L., Yuan, P., 2011. Research and development of carbon footprint analysis in Hunan Province. Energy Proc. 5, 1210-1217.

[11] Fan, L., Liu, L., Wu, G., et al., 2006. Analyzing impact factors of CO2 emissions using the STIRPAT model. Environ.Impact Assess. Rev. 26 (4), 377-395.