MANTECH Publications
Journal of Artificial Intelligence, Machine Learning and Soft Computing
Volume 7, Issue 1, January-June 2022

Analysis of Air Quality in Urban Area using Machine Learning Approach

Sakshi S. Suroshe[1], S. V. Dharpal[2]
P.G. Student[1], Assistant Professor[2]
Department of Civil Engineering
Prof. Ram Meghe Institute of Technology and Research, Bandera SGB Amravati University, India
Email ID: *sakshs16@gmail.com[1]*

Abstract

Air Quality security has gotten one of the foremost fundamental exercises for the administration in numerous mechanical and concrete zones in today's world. The meteorological and traffic factors, consuming crude oil derivatives and mechanical parameters perform critical jobs in air contamination which make an adverse effect on living beings. With this expanding pollution on the earth, we also had different executing models that can record data about centralizations of air pollutants (SO2, NO2, etc.). The affidavit of those unsafe gases is noticeable all around; is influencing the character of individuals' lives, particularly in urban territories. Of late, numerous specialists began to study about this concern and mentioned several measures to manage these conditions with the assistance of the presidency and native people. Data Analytics is a leading approach as it includes natural detecting systems and sensor information accessible. Machine Learning strategies are utilized to predict the ratio with relation to other components present in the earth's atmosphere. Various regression models are used to predict the air quality and their relative effects.

Keywords: *Air Quality Index, Machine Learning, Air Pollutants, Adverse Effects, Parameters, Air Quality prediction, Air Pollution, Linear Regression, Regression Analysis.*

INTRODUCTION

The pollution is playing important role in our world as it affects human beings, animals, our planet and all living things. Pollution can lead to unstable climate change which can disrupt the ecosystem. As per a study 51% of the pollution is caused by industrial pollution, 27 % by vehicles, 17% by crop burning and 5% by other sources. The majority of the world's population continue to be exposed to levels of air pollution substantially above WHO Air Quality Guidelines and air pollution constitutes a major, and in many areas, increasing threat to public health (G. Shaddick, M. L. Thomas, et.al. 2020) agricultural productivity, transport system and our cultural assets (monuments and historical buildings).

The rising trends in population growth and the consequent effects on air quality are evident in the Indian scenario. As per WHO (2016) estimates, 10 out of the 20 most populated cities in the world are in India. Based on the concentrations of PM2.5 emissions, India was ranked the fifth most polluted country by WHO (2019), in which 21 among the top 30 polluted cities were in India (Dr. Bhola Ram Gurjar et.al. 2021).

There are many pollutants that are major factors in disease in humans. Among them, Particulate Matter (PM), particles of variable but very small diameter, penetrate the respiratory system via inhalation, causing respiratory and cardiovascular diseases, reproductive and central nervous system dysfunctions, and cancer.

Despite the fact that ozone in the stratosphere plays a protective role against ultraviolet irradiation, it is harmful when in high concentration at ground level, also affecting the respiratory and cardiovascular system. Furthermore, nitrogen oxide, sulfur dioxide, Volatile Organic Compounds (VOCs), dioxins, and polycyclic aromatic hydrocarbons (PAHs) are all considered air pollutants that are harmful to humans.(Ioannis Manisalidis, Elisavet Stavropoulou, et.al. (2020)).

The sources of emission vary from small unit of cigarettes to large volume of emission from motor engines of automobiles and industrial activities. Long and short term exposure to air suspended toxicants has a different toxicological impact on human including respiratory and cardiovascular diseases, neuropsychiatric complications, the eyes irritation, skin diseases, and long term chronic diseases such as cancer. Several

reports have revealed the direct association between exposure to the poor air quality and increasing rate of morbidity and mortality mostly due to cardiovascular and respiratory diseases. Air pollution is considered as the major environmental risk factor in the incidence and progression of some diseases such as asthma, lung cancer, ventricular hypertrophy, Alzheimer's and Parkinson's diseases, psychological complications, autism, retinopathy, fetal growth, and low birth weight **(Adel Ghorani-Azam, Bamdad Riahi-Zanjani, et.al. (2016).**

Numerous epidemiological studies have established the associations between the air pollutants and daily excess in mortality and morbidity. All the five different models were compared for their generalization and prediction abilities using statistical criteria parameters, viz. correlation coefficient, standard error of prediction (SEP), mean absolute error (MAE), root mean squared error (RMSE), bias, accuracy factor (Af), and Nash–Sutcliffe coefficient of efficiency (Ef) **(Kunwar P. Singh, Shikha Gupta, et.al. (2012).**

According to CPCB the AQI is calculated using 12 parameters (Air Pollutants) namely NO2 (Nitrogen Dioxide), SO2 (Sulfur Dioxide), CO (Carbon Monoxide), O3 (Ozone), PM10 (Particulate Matter having diameter 10 micron or less), PM2.5 (Particulate Matter having diameter 2.5 micron or less), NH3 (Ammonia), Pb (Lead), Ni (Nickel), As (Arsenic), Benzo(a)pyrene and Benzene . Most of the time AQI is based on the criteria pollutants (i.e. PM10, PM2.5, SO2, NO2, CO and O3) but while calculating the AQI using many pollutants from the list of 12 pollutants is more desirable. However, the selection of pollutants depends on the AQI objectives, averaging μperiod, Data Availability, Monitoring frequency and measurement methods. AQI can be defined as it is a numerical value that the governmental agencies used to measure the levels of air pollution in the atmosphere and communicate it with population. If AQI increases then large percentage of population is affected because it adversely affects the human health. As we know that AQI can be calculated by using the concentration of different air pollutants and finally we get the single numerical value as AQI. (Radhika M. Patil, Dr. H. T. Dinde et.al. (2020).

AIR QUALITY INDEX (AQI)

Commercial The Air Quality Index (AQI) is an environmental index which describes

the overall ambient air status and trend of a particular place based on specific standard. It is a tool that transforms the (weighted) values of individual air pollutants (parameters) into a single number or set of numbers (Rao, 1993). The overall ambient air quality of a specified area can be assessed in a better way and quantified in terms of AQI since it represents the cumulative effect of all the pollutants. AQI can also enable one to formulate the alternative policies for prevention of air pollution or to design control equipment which, for instance, will reduce the level of certain pollutants while increasing the levels of others. There are several methods and equations used for determining the AQI. However, here the below mentioned equation (Zlauddin and Siddiqui, 2006: Joshi and Semwal, 2011) has been used for computation of AQI value.

AQI = ¼ x (ISPM / SSPM + IRSPM / SRSPM + ISO2 / SSO2+ INOx / SNOx) x 100

Where ISPM, IRSPM, ISO2 and INOx = Individual values of suspended particulate matter, respirable particulate matter , sulphur dioxide and oxides of nitrogen respectively obtained on sampling.

SSPM, SRSPM, SSO2 and SNOx = standards of ambient air quality as prescribed by the Central Pollution Control Board of India (CPCB).

The higher the AQI value, greater is the level of air pollution and greater is the health risk. The AQI scale is divided into five categories as depicted in Table 2 It describes the range of air quality and its associated potential health effect (Panda B.K, Panda C.R et.al. (2012)).

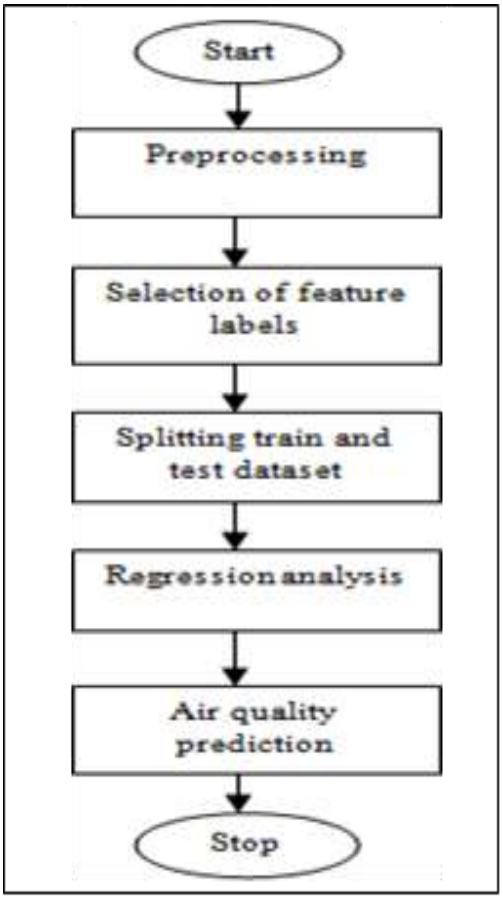
Table: 1

AQI Value	**Remarks**	**Health Concern**
00 – 25	Clean air(CA)	None/minimal health effect
26 – 50	Light Air	Possible respiratory or cardiac effect for most sensitive group
51 – 75	Moderate Air Pollution(LAP)	Increasing symptoms of respiratory and cardiovascular illness
76- 100	Heavy Air Pollution(HAP)	Aggravation of heart and lung diseases
>100	Severe Air Pollution(SAP)	Serious aggravation of heart and lung diseases Risk of death in children

METHOD

Many methods have been applied in the literature 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. The processed data sets are analyzed through different regression analysis techniques for accurate results. Regression analysis is the form of a predictive modeling technique that investigates the relationship between a dependent and independent variable.

This technique is used for forecasting or predicting, time series modeling, and finding the causal effect relationship between the variables. Regression analysis is a method of analyzing and modeling data. There are different kinds of regression techniques available to make predictions namely Linear regression, Support vector regression, Decision tree regression and Lasso regression following Figure represents the flow diagram of the system. The diagram represents the step by step process, from data preprocessing to air quality prediction.



Regression Analysis

A Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables first, regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Second, in some situations regression analysis can be used to infer causal relationships between the independent and dependent variables.

Linear Regression Method

The processed data sets are used to create a function to plot the training and validation data for the different models such as linear regression, Support vector regression, Decision tree regression, and Lasso regression .Linear regression is a basic and best-used type of predictive analysis. Linear regression is used to examine two things; namely, it checks whether a set of predictor variables is doing a good job in predicting an outcome (dependent) variable And checks which variables, in particular, are the significant predictors of the outcome variable.

$Y=c+b*x$

Multivariate Linear Regression

Multiple linear regressions is the most common form of linear regression analysis. As a predictive analysis, the multiple linear regressions are used to explain the relationship between one continuous dependent variable and two or more independent variables.

The general regression equation is,

$y1 = a0 + a1x1 + a2x2 +. . .+anxn$

Where a1; a2; a3; . . . an are the coefficients

Where y1 = Air quality value x1, x2..,x3 = Meteorological Parameters.

Air Quality Prediction

Air quality prediction Air quality is predicted using the R squared value. R square determines the proportion of variance in the dependent variable of the system that can be explained by the independent variable. It is a statistical measure in a regression model. It is also called a coefficient of determination. The predicted R square values indicate how well a regression model predicts responses for the given observations. R square value generally lies between -1 to +1. In this project, R square value for training and test dataset is calculated using four different regression models. Here in this project R square value of the training dataset is always greater than the test data.

If the R square value is near to 1, then the regression model is better for than dataset.

$$R^2 = 1 - \frac{\text{Explained Variation}}{\text{Total Variation}}$$

Root Mean Square Error is the standard deviation (SD) of the prediction errors. Residuals are the measure of how far from the regression line data points are; RMSE tells you how the data is concentrated around the best fit line or a measure of how the residuals are spread out. It is commonly used in forecasting, climatology, and regression analysis to verify experimental results.

$$RMSE = \sqrt{(f-o)^2}$$

Where f = forecasts (unknown results)
and o = observed values (known results).

CONCLUDING REMARK

Based on the literature reviewed from various sources following conclusions appears to be justified:

- There are large numbers of people exposed to harmful levels of air pollution. Although precise quantification of the outcomes of specific policies is difficult, coupling the evidence for effective interventions with global, regional and local trends in air pollution can provide essential information for the evidence base that is key in informing and monitoring future policies. There have been major advances in methods that expand the knowledge base about impacts of air pollution on health, from evidence on the health effects, modelling levels of air pollution and quantification of health impacts.

- There is a continuing need for further research, collaboration and sharing of good practice between scientists and international organisations, for example the WHO and the World Meteorological Organization, to improve modelling of global air pollution and the assessment of its impact on health. This will include developing models that address specific questions, including for example the effects of transboundary air pollution and desert dust, and to produce tools that provide policy makers with the ability to assess the effects of interventions and to accurately predict the potential effects of proposed policies.

- The suggested model of LSTM can efficiently decrease the error rates and

enhanced the forecast of air pollution, although there is still space for enhancement to overcome the levels of PM form the atmosphere.

- The residents of rural areas are seldom aware of the harmful effects of air-borne pollutants and their consequence to human health. Public awareness programmes should be initiated by the government in every city, both rural and urban, highlighting the importance of managing air pollution at source and the various control measures that could be adopted to reduce pollutant emissions. Such initiatives could significantly reduce the activities, such as open burning of wastes, crop burning, use of biomass as a fuel for cooking and burning of plastic and rubber materials during winters. A holistic approach incorporating all of the mentioned measures could be beneficial to attain cleaner air quality in Indian cities and guarantee a healthier place to inhabit.

- The industrialization of societies is necessary to develop, but a long-term health problem and ecological impacts of such growth should always be considered prior to imposing a large financial burden on the societies. Therefore, it is suggested to adopt a balance between economic development and air pollution by legislating policies to control all activities resulting in air pollution

- Standardization of motor engines and manufacturing engines with low fuel consumption is another strategy to reduce the level of air pollutants Improving public transportation systems by using more subways (metro), trams, and electrical bus routes. Reducing the costs for the people who are using such systems is an optimal solution for lowering air pollution.

- Standardization of vehicle's fuel as much as possible and also finding a new source of energy for motor engines has attracted great attention.

- Imposing penalties for polluting industries and implementing low tax policy for clean technologies

- Continuous monitoring of air quality, designing and developing tools to identify the pollutants, finding the origin of the particles, and the use of particulate filter for diesel engines and other nonroad cars are other suggested

practical approaches to reduce air pollution

- Extensive media campaign to increase public awareness about air quality, environmental, and public health issues.

- Regression analysis techniques are used to predict the concentration of Carbon monoxide C.O. in the environment. The shortterm exposure of Carbon monoxide causes headaches, dizziness, vomiting, nausea, irritation of the airways, coughing, and difficulty in breathing and long term exposure causes irregular heartbeat, nonfatal heart attack, and even death due to lung disease.

- Further studies will be conducted by collecting the air Pollution data & meteorological data of Amravati Town for the study of model analysis approach which will give the best suitable regression model for Prediction of Air Quality in Amravati Town.

ACKNOWLEDGMENT

The authors are very much thankful to Principal Dr. A. P. Bodkhe and Dr. P. S. Pajgade, Head, Department of Civil Engineering, PRMIT&R, Badnera, Amravati for their fullest cooperation. The authors are also thankful to Prof. Sachin Dharpal for timely guidance. Lastly thanking the college librarian to permit us for collecting the Research papers online as well as manually.

REFERENCES

1. G. Shaddick, M.L. Thomas, et al.,(2020)," Half the World's Population are Exposed to increasing Air Pollution", npj Climate and Atmospheric Science, pp 1-5.

2. Ioannis Manisalid is et al., (2020)," Environmental and Health Impacts of Air Pollution: A Review", Frontiers in Public Health, pp 1-13.

3. Dr. Bhola Ram Gurjar, (2021), "Air Pollution in India: Major Issues and Challenges", pp 1-23.

4. Adel Ghorani-Azam et al., (2016)," Effects of Air Pollution on Human Health and Practical Measures for Prevention in Iran", Published by Wolters Kluwer - Medknow, Journal of Research in Medical Sciences, pp 1-12.

5. Kunwar P. Singh, Shikha Gupta et al.,(2012)," Linear and nonlinear modelling approaches for urban Air Quality Prediction", Science of the Total Environment 426, pp 244-255.

6. Panda B.K. and Panda C.R., (2012), " Estimation of ambient air quality status in Kalinga Nagar industrial complex in the district of Jajpur of Odisha", International Journal of Environmental Sciences, Vol. 3, No 2 , pp 767-775.

7. Radhika M. Patil, Dr. H. T. Dinde et.al., (2020), " A Literature Review on Prediction of Air Quality Index and Forecasting Ambient Air Pollutants using Machine Learning Algorithms ", pp 1-5.

8. Aarthi, P. Gayathri, N. r. Gomathi, et.al., (2020), " Air Quality Prediction Through Regression Model", International Journal of Scientific & Technology Research Vol 9, pp 923-928.