International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 12 | Dec 2020 www.irjet.net p-ISSN: 2395-0072

DETERMINATION OF AIR QUALITY MONITORING & PREDICTION

Purvesh Patil[1], Parth Modi[2], Ninad Kamble[3]

[1-3]*Information Technology Engineering, Padmabhushan Vasantdada Patil Pratishthan College of Engineering, Maharashtra, India*

Abstract - *Air pollution and its prevention are constant scientific challenges during last decades. However, they still remain huge global problems. Affecting human's respiratory and cardiovascular system, they are cause or increased mortality and increased risk for diseases for the population. Many efforts from both local and state government are done in order to understand and predict air quality index aiming improved public health. The levels of air pollutants in the environment are increasing manifold. As a result, a lot of improvement is still required in this field for prediction analysis. With incomplete data parameters and their significance (priority), most of the methods fail to predict the pollution levels significantly. The regulation of air pollutant levels is rapidly becoming one of the most important tasks for the governments of developing countries, especially China. We compare simple machine learning algorithms. The air pollution dataset contains four major pollutants (Air Quality Index(AQI), Nitrogen Dioxide(NO2), Sulphur Dioxide(SO2), Suspended Particulate Matter (SPM) and Respirable suspended particulate matter or (RSPM)) The results are promising and it was proven that implementation of these algorithms could be very efficient in predicting air quality index.*

***Key Words*: SVR, LR, DTR, SPM, RSPM, Air Quality Prediction, Air Quality Analysis, Air Quality Index.**

1. INTRODUCTION

Air is one of the most essential natural resources for the existence and survival of the entire life on this planet. All forms of life including plants and animals depend on air for their basic survival. Thus, all living organisms need good quality of air which is free of harmful gases to continue their life. According to the world's worst polluted places by Blacksmith Institute in 2008, two of the worst pollution problems in the world are urban air quality and indoor air pollution. Various methods like 'Climatology' (based on the assumption that the past is a good predictor of the future) have been used for air quality forecasting. These approaches are usually used to predict exceeding limits from specific thresholds, not ambient concentrations. As a result, a lot of Improvement is still required in this field for prediction analysis. With incomplete data parameters and their significance (priority), most of the methods fail to predict the pollution levels significantly.

The regulation of air pollutant levels is rapidly becoming one of the most important tasks for the governments of developing countries. This paper is one scientific contribution towards this challenge. We compare simple machine learning algorithms. The air pollution dataset contains four major pollutants (Air Quality Index(AQI), Nitrogen Dioxide(NO2), Sulphur Dioxide(SO2), Suspended Particulate Matter (SPM) and Respirable suspended particulate matter or (RSPM)). The results are promising and it was proven that implementation of these algorithms could be very efficient in predicting air quality index.

1.1 Literature Survey

Unadulterated air is a mix of various gases, for instance, Nitrogen, oxygen, argon, carbon dioxide, and small proportions of various gasses in a settled degree. In the event that the synthesis of air is balanced by some other technique; it is known as air contamination which prompts terrible effect on human wellbeing, their condition and other living life forms. It prompts troublesome medical problems. Clean air is thought to be a fundamental, essential of human wellbeing and prosperity. However, air pollution continues to be a basic hazard to prosperity around the globe (WHO 2011).

A system based on neural networks and data mining used in Order to predict the hour, eight hours and 24 hours ozone Levels in advance. The networks performance is evaluated using the root mean square error, the mean absolute error and the correlation coefficient. The result showed worse forecast for eight hours in advance. Data mining for prediction of air pollution is applied in which proposed two methods of feature selection, genetic algorithm and linear method. The selected feature sets take part in prediction of the atmospheric pollutants PM10, SO2, NO2 and O3. This paper discusses the numerical aspects of the air pollution prediction problem, concentrating on the methods of data mining used for building the most accurate model of prediction. Artificial Neural Network model and Data mining techniques were used for the analysis and prediction of ambient air quality for Tamil Nadu. The pattern obtained from these models could serve as an important reference for the Government policy makers in devising future air pollution standard policies. The application of correlation between predicted variables and features is not a good practice as it had resulted in the lots of prediction errors in some of the above papers. The techniques used were not shown the good accuracy.

1.2 Comparative Analysis

Table - 1: Analysis of previous papers

Parameter	Air Quality Analysis and Prediction	Identification of regions and probable health risk due to air pollution
Algorithm used	Linear regression and support vector regression	K-means algorithm
Accuracy	High	Low
Time for processing the data and Graphs	Medium	High
Parameters Covers	Displays the data in time format	Displays data in region wise

2. METHODOLOGY

There are two primary phases in the system: 1. Training phase: The system is trained by using the data in the data set and fits a model (line/curve) based on the algorithm chosen accordingly. 2. Testing phase: the system is provided with the inputs and is tested for its working. The accuracy is checked. And therefore, the data that is used to train the model or test it, has to be appropriate. The system is designed to detect and predict AQI level and hence appropriate algorithms must be used to do the two different tasks. Before the algorithms are selected for further use, different algorithms were compared for its accuracy. The well-suited one for the task was chosen.

3. PROBLEM DEFINITION

Air pollution is rapidly increasing due to various human activities and is the introduction into atmosphere of chemicals, particulates or biological materials that cause discomfort, disease or death of humans, damage other living organisms such as food crops, or damage natural environment or built environment. Indeed air pollution is one of the important environmental problems in metro-politician 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, next year air pollution count to avoid these problems.

4. PROPOSED SYSTEM

As the existing system only predict the quality of air country wise which is not sufficient to understand the air quality impact in depth level. One drawback in existing system is that it cannot predict air quality in sub-regions, where each sub region can have different air pollutant level than other. To overcome this we use Air Quality Analysis and Prediction system. In this system we fetch the air pollution data. Once data is fetched data is trained according to environment. This data is use to generate patterns in later phase. Region wise air quality analysis is performed and prediction of future air quality is determined.

Step 1: Extraction of historical dataset.
Step 2: Data pre-processing and normalization.
Step 3: Divide dataset in 70:30 ratio.
Step 4: Perform Feature selection on the dataset features.
Step 5: Train and test using different regression algorithms.

4.1 Linear Regression

Linear Regression attempt to model the relationship between two variables by fitting a linear equation to observed data. The other is considered to be dependent variable. For Example: A modeler might want to relate weights of individuals to their heights using a linear regression model.

A linear regression line has an equation of the form:

$$\mathbf{Y = A + Bx},$$

Where X is the explanatory variable and Y is the dependent variable.

The slope of the line is 'b' and 'a' is the intercept. It is next step after correlation. It is used when we want to predict the value of a variable based on the value of an another variable.

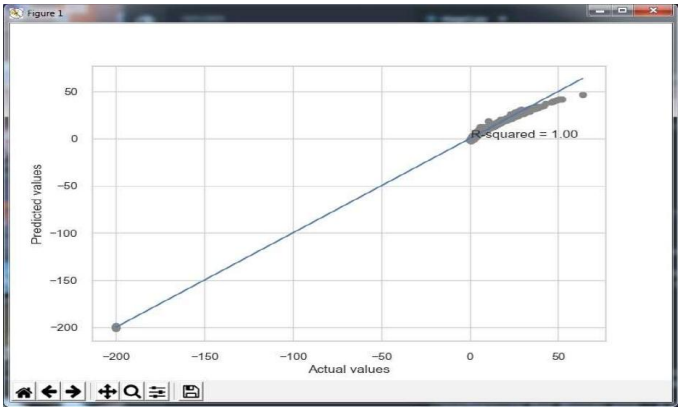

Fig - 1: Linear Regression

4.2 SVR (Support Vector Regression)

SVR is similar to LR in that the equation of the line is

$$Y = Wx + b$$

In SVR, this straight line is referred to as hyperplane. The data points on either of the hyperplane that are closest to the hyperplane are called support vectors which are used to plot the boundary line. SVR tries to fit the best line within a threshold value (distance between hyperplane and boundary line).

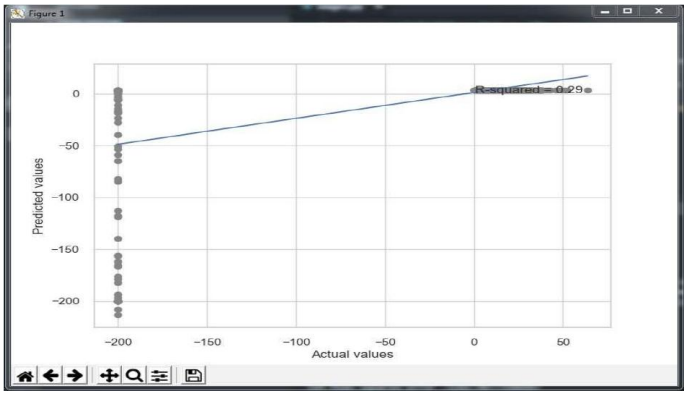

Fig - 2: SVR (Support Vector Regression)

4.3 DTR (Decision Tree Regression)

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

5. FUTURE SCOPE

In future this system may bind with real time sensors and gives live predictions about AQI level.
Also we may add large historical data of AQI index with more independent variable which can help to improve the results of this system.

6. CONCLUSIONS

The regulation of air pollutant levels is rapidly becoming one of the most important tasks. It is important that people know what the level of pollution in their surroundings is and takes a step towards fighting against it. The results show that machine learning models (logistic regression and auto regression) can be efficiently used to detect the quality of air and predict the level of AQI in the future.

The proposed system will help common people as well as those in the meteorological department to detect and predict pollution levels and take the necessary action in accordance with that. Also, this will help people establish a data source for small localities which are usually left out in comparison to the large cities. The paper compares four different algorithms for machine learning : SVR, LR, DTR and Lasso.

REFERENCES

[1] Sai Sabitha, Bonny Paulose, Ritu Punhani "Identification of Regions and Probable Health Risks. Due To Air Pollution Using K-Means clustering Techniques"; (IEEE 2018).

[2] Shweta Taneja, Dr.Nidhi Sharma "Predicting Trends in Air Pollution in Delhi using Data Mining"; (IEEE 2016).

[3] Ranjana Waman Gore, Deepa S. Deshpande "An Approach for Classification of Health Risks Based on Air Quality Levels"; (IEEE 2017).

[4] Sachit Mahajan, Ling-Jyh Chen, Tzu-Chieh Tsai: An Empirical Study of PM2.5 Forecasting Using neural network. IEEE Smart World Congress, At San Francisco, USA [2017].