Air Quality Monitoring System Using Linear Regression and Machine Learning.

Air Quality Prediction

Vasudeva M B [1], Pradeep K [2], Dr. Chetana Prakash [3], Dr. Nirmala C R [4]

[1,2,3,4] *Computer Science and Engineering, BIET, Davanagere, India.*

---***---

Abstract— Air is an essential natural resource, has been compromised in terms of quality by economic activities. The aim of this paper is to investigate a data a set of air pollutants record for India meteorological sector using machine learning technique. Air quality index is used to measure the Air quality and to implement the different classification of regression technique to measure the air pollutant with major air quality index. The techniques is then evaluated using Mean absolute error and which show that Linear Regression is best suited for predicting the air quality in New Delhi. Monitoring air pollution is of increasing concern today. People are suffering from health problems as a result of prolonged exposure to polluted environments. By using python built a machine learning algorithm and developed to detect the air pollution model.

Keywords— Air Quality Index, Pollutant, Linear Regression, Pollution Prediction. **(key words)**

I. INTRODUCTION

Improved lifestyle has considerable the air pollution in urban areas of the regulation of the pollution in New Delhi the capital city of India is one of the most populated cities in the world. In the recent days has pointed out that the concentration of pollutants in Delhi is much higher than the limits of the area so have to control the air pollution with the help of Air Quality Index with forecast disclose the air quality trend with the machine learning field in which programs are achieved efferently. To overcome this situation of the much air pollution to declare the healthy life without air pollution is the air quality index with this have to define air pollutant and collect the data from the meteorological department with the help of some advisor to get the result of the air pollution to predict the air quality to form the better life lead in the New Delhi. Local air quality affects the health problems and it is slow poison it will kill human in slow manner so to defeat the air quality have to manage air quality and issue health advisors. By this we can lead a good life with no problems in the health issue. So to improve our life style with the harmless air pollution quality to monitor the air quality predictor to reduce the pollution by controlling the pollutants in the air. The air that should be good pollutant to live the better life.

II. EASE OF USE

A. Air Quality Index (AQI)

The air quality index is used to measure the air quality to predict the air whether the air is good or not for the healthy issue to detect the air pollutant and showing the changes in the temperature by measuring air pollutant with the road traffic and nitrogen dioxide and by using simple tool called diffuse tube. The diffuse tube is used to absorb a specific pollutant from the air and no power supply is needed and monitoring the air pollutant to decrease the health issues from the air quality index.

B. Monitoring Air Pollution

The monitoring air pollution by monitoring the devices to acquire the air to which the air quality is good. By accessing the levels of pollution in air quality index it is more affect to the health so this monitors to the better air quality to survive the life with the healthy way to defuse the life in the better way. By this monitoring system our health issue will better than un-monitoring periods to detect those controls to monitor the air pollutants with right way to simulate the pollutant in which the controlled system used to de-monitor the air which is affecting the health with badly and by correct monitor can survive the life in a good manner.

III. EXISTING SYSTEM

- Statistical and numeral models are used in the existing system.

- By the air quality index with the categories of different index to monitor whether the air is good, satisfactory, moderate, poor, very poor, severe based. The data is displayed on a monthly, weekly or daily basis.

- It measures only the half of the air quality and tested sample had accuracy between 65% to 75%

IV.   PROPOSED SYSTEM

Accuracy plays an important role in prediction by the system to predict the air which is very harmful to the health purpose to survive the life without good air is impossible in order to effuse the monitoring system to investigate the data set to the air pollutant to index from the comparing various performance of the parameters to obtain the most accurate results.

V.   OBJECTIVES

- To Detects the levels of PM2.5 based on given atmospheric values.

- To predicts the levels of PM2.5 for a particular data.

- To detect whether a data sample is either polluted or not polluted.

VI.   LITERATURE SUMMARY

A. Abbreviations and Acronyms

AQI: Air Quality Index.

GRAP: Graded Response Action Plan.

NCR: National Capital Region.

AQ-EWS: Air Quality Early Warning System.

IITM: Indian Institute of Tropical Meteorology.

NCMRWF: National Centre for Medium-Range Weather

Forecasting.

B. Equations

1) Linear Regression

Three common evaluation metrics for regression problems:

- Mean Absolute Error (MAE) is the mean of the absolute value of the errors:

$$\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i|$$

- Mean Squared Error (MSE) is the mean of the squared errors, MSE "punishes" larger errors, which tends to be useful in the real world:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

- Root Mean Squared Error (RMSE) is the square root of the mean of the squared errors, RMSE is interpretable in the "y" units:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)$$

C. Authors and Affiliations

Author Name: Huixiang Liu

Year: 2018

Remarks: They have forecasted the Air Quality Index (AQI) for the city Beijing and predicting the concentration of Noxin. They used to support vector regression and Random Forest Regression for AQI prediction.

Author Name: Aditya C R

Year: 2019

Remarks: Employed the machine algorithms to detect and forecast the PM2.5 concentration level on the basis of dataset containing atmospheric concentration level for a particular date.

Author Name: Nidhi Sharma

Year:2020

Remarks: Had gone trough the detailed data analysis of air pollutants from 2010-2020 and also proposed critical observation of 2017-2018 air pollutant trend in Delhi, India. They have predicted the future trends of various pollutants as Sulfur Dioxide(SO2), Nitrogen Dioxide(NO2), Suspended Particulate Matter(PM), Ozone(O3), Carbon Monoxide(CO) and Benezene.

Author Name: Archonotoula Chaloulakou

Year:2020

Remarks: The research have been implemented Artificial Neural Network(ANN) and Multiple Linear Regression (MLR) Algorithms to forecast the PM10. Comparison between ANN and MLR was also done in this study that indicates ANN is better in performance than MLR.

Author Name: R Gunashekharn

Year:2020

Remarks: The main objective of this study is to monitor the air quality salem sadeshwari college, Tamil nadu area for the period of April 2019 to March 2020 this has been shown that this area has to serious pollution issues related to the pollutants as Sulfur Dioxide, Oxides of Nitrogen.

D. Figures and Tables

Table 1: Air quality index table

AQI	ASSOCIATED HEALTH IMPACTS
GOOD (0-50)	**Minimal Impact**
SATISFACTORY (51-100)	**May cause minor breathing discomfort to sensitive people.**
MODERATELY POLLUTED **(101-200)**	**Under the problem with the breathing discomfort to people with lung disease such as asthma and discomfort to people with heart disease for children's and adults.**
POOR (201-300)	**Cause breathing discomfort to people on prolonged exposure and discomfort to people with heart disease.**
VERY POOR (301-400)	**illness on prolonged exposure.**
SEVERE (401-500)	**Highly impact on healthy people. Serious health impacts on people with lung/heart disease.**

The above table shows the Air Quality Index associated with the health issue impacts from the dataset collection with the health issues of the air quality prediction is either good, satisfactory, moderately polluted, poor, very poor, severe.

VII.   METHODOLOGY

Air Pollutant defines the pollutant and stores the past data from the meteorological department and compare the behavior of the data transformation from the splitting data from the data frame with training the model by comparing ML algorithm and preprocessing the selection of the feature label and test the dataset using regression analysis of the air quality prediction. The proposed system will detect the air quality and defines the pollutant in the air. By the use of past data available in the meteorological department the data fetched will be transformed and the model will be created. To that model the pollutant data will be sent to check for the result of the air quality to detect whether it is good or not.
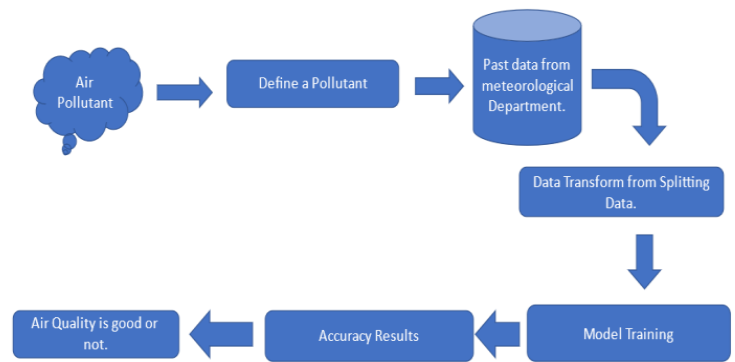

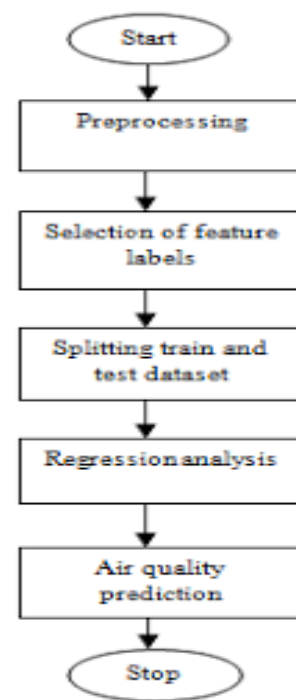Fig 1: Design Architecture of Methodology Diagram

Fig 2: Data flow diagram.

By visualizing the PM2.5 Data to seems to be some repetitions in the dataset after index 700, so let's check for duplicates in the dataset and remove them if the duplicates rows to except the first occurrence based on the all columns of fine particulate matter is in air pollutant that is a concern for people's health and when levels in air are high tiny particles in the air that reduce visibility and cause the air to appear and elevated.

REFERENCES

- World Health Organization.   Air Pollution. Available online: http://www.who.int/health-topic/airpollution# tab=tab_1/ (accessed on 13 March 2020).

- Rocca, J. Ensemble Methods: Bagging, Boosting and Stacking Available online: http://towardsdatascience.com/ ensemble-methods-bagging-boosting-and-stacking-c9214a10a205 (accessed on 23 April 2019)

- Taiwan's Environmental Protection Administration. Taiwan Air Quality Monitoring Network. Available online: http://taqm.epa.gov.tw/taqm/en/b0201.aspx (accessed on 13 March 2020).

CONCLUSION

The aim of this project is to investigate a data a set of air pollutants record for India meteorological sector using machine learning technique. To reduce the air quality with the good pollutants and reduce the health issues of human and controls the monitoring system to the air quality prediction with the help of air quality index can control and know the situation of the people's health issues by calculating the air quality system monitoring to declare the good air or not. It also indicates that the prediction performance varies over the different regions with the datasets of the value of prediction level. The successful prediction will maximize the benefits of the customer algorithm used will predict air quality compared the accuracy using machine learning algorithms. The aim of the paper or project is to investigate a data a set of air pollutants record for India meteorologic department using machine learning technique with the accurate result.