IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006048 |

Air Pollution Prediction Using Machine Learning

A.Sudharshan Reddy[1], Ch.Vasantha Gopi[2], K.Naveen Kumar[3] , E.Lakshman Sai[4],J.Bhanu Prakash[5]

Assistant Professor, Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[1]

B.Tech Student, Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[2,3,4,5]

ABSTRACT: Prediction of pollution is an increasingly important problem.Nowadays, air pollution has reached critical levels and the air pollution level in many major cities has crossed the air quality index value as set by the government. It has a major impact on the health of the human. The air quality monitoring system measures various air pollutants in various locations to maintain good air quality. It is the burning issue in the present scenario. Air is contaminated by the arrival of dangerous gases into the climate from the industries, vehicular emissions, etc...With the advancement in technology of machine learning, it is now possible to predict the pollutants based on the past data. In this paper we are introducing a device that can continue that can take present pollutants and with the help of past pollutants, we are running an algorithm based on the machine learning to predict the future data of pollutants. The sensed data is saved inside the Excel sheet for further evaluation. These sensors are used on the Arduino Uno platform to collect the pollutant data.

KEYWORDS: Airpollutants, AQI, Machinelearning, Accuracy.

I. INTRODUCTION

Air Quality Index- Nowadays pollution levels are increasing due to the PM2.5 gases which affect the heart functionalities, lung cancer and other respiratory and breathing problems. The long term damage to the liver, kidney, brain, nerve and other organs in the human body system is affected by air pollution.

Air tracking manner to measure ambient ranges of air pollutants inside the air. Monitoring has become a major job as air pollution has been increasing day by day. Continuous monitoring of air pollution at a place gives us the levels of pollution in that area. From the information obtained by the device gives us information about the source and intensity ofthe pollutants in that area. Using that information we can take measures or make efforts to reduce the pollution level so that we can breathe in a good quality of air.

The major cases of air pollution are Ozone (O3),Nitrogendioxide(NO2),CarbonMonoxide(CO),Sulphurdi oxide(SO2),Particularmatter(PM). These gases are cannot been seen or noticed which are produced from burning of fossil fuels, wood burnings, industrial boilers and from the explosion of volcano. They may cause the affects in humans and are the main reason for causing cancer, birth defects and breathing-related problems.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006048 |

$$I = \frac{I_{high} - I_{low}}{C_{high} - C_{low}}(C - C_{low}) + I_{low}$$

where:

I = the (Air Quality) index,
C = the pollutant concentration,
C_{low} = the concentration breakpoint that is $\leq C$,
C_{high} = the concentration breakpoint that is $\geq C$,
I_{low} = the index breakpoint corresponding to C_{low},
I_{high} = the index breakpoint corresponding to C_{high}.

Environmental protection agency breakpoint table[1]

AQI Category, Pollutants and Health Breakpoints

AQI Category (Range)	PM_{10} (24hr)	$PM_{2.5}$ (24hr)	NO_2 (24hr)	O_3 (8hr)	CO (8hr)	SO_2 (24hr)	NH_3 (24hr)
Good (0–50)	0–50	0–30	0–40	0–50	0–1.0	0–40	0–200
Satisfactory (51–100)	51–100	31–60	41–80	51–100	1.1–2.0	41–80	201–400
Moderately polluted (101–200)	101–250	61–90	81–180	101–168	2.1–10	81–380	401–800
Poor (201–300)	251–350	91–120	181–280	169–208	10–17	381–800	801–1200
Very poor (301–400)	351–430	121–250	281–400	209–748	17–34	801–1600	1200–1800
Severe (401–500)	430+	250+	400+	748+	34+	1600+	1800+

II.PROPOSED METHODOLOGY

The air quality values are predicted using three binary machine learning algorithms are presented in [4]. In this error analysis is done with GLM, SVM and Bayes methods. The accuracy of the simple machine learning methods are compared in [5] and variation in the accuracy is presented with different sizes and data divisions. The data set of air quality consists of pollutant data of CO, O3, NO2, SO2, PM10, and PM2.5. For the better air quality prediction, we must co-relate the pollutant data with meteorological data [Temperature, Wind Speed, Humidity, Wind direction]. Neural network method provides better accuracy compared to others. The air pollution prediction is implemented in [6]based on different norm regularization and optimization algorithms as a machine .

Learning tools.

For pollution estimation or prediction, linear regression algorithms are suitable and for forecasting the pollution levels neural network methods and SVM based methods are preferred[7].

The air quality index is predicted by using machine learning algorithms for the detection of PM2.5 level using logistic regression [8].There are applications that show the constant PM2.5 levels, while some show the forecast of a specific day. This framework abuses AI models to recognize and forecast PM2.5 levels dependent on an informational collection consisting of meteorological conditions in a particular city. The data set [9] used in detection of PM2.5level consists of Temperature, Wind speed, Dew point, Pressure, PM2.5, Concentration (ug/m^3).

A.Gaps Identified in the Literature

In those papers, they only implemented the prediction of PM2.5. In this project they want to implement prediction of all the pollutants [CO, O3, NO2, SO2, PM2.5, PM10] with the help of meteorological data for better prediction.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006048 |

III.ARRANGEMENT OF THE PROPOSED SYSTEM

A. *Flow Chart* The proposed technique is represented within the under block diagram as proven in Fig. 1.

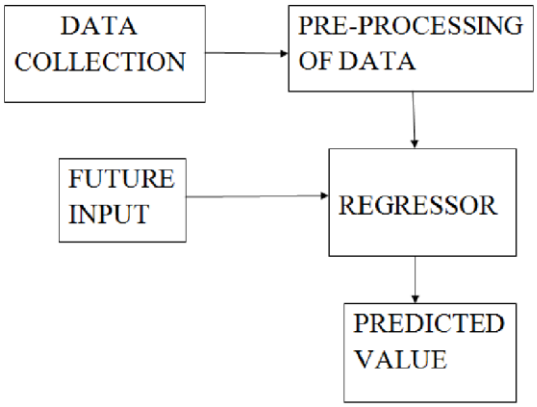


Fig1: System design

IV.PROPOSEDMETHODOLOGY

A. *USING MACHINE LEARNING MODELS*

Linear regression:
Linear Regression [10] is nothing but an algorithm based on the machine learning are depends on supervised learning which performs a regression task. Depending on independent variables linear regression gives a target prediction value which is most likely used for finding the relationship among variables and forecasting. Depending on the connection among the established and the independent variables, different regression models differ, they are being considered and List of independent variables used.
y = mx+c
Decision Tree:
The Regression on the Decision Tree [11] is both a nonlinear and non-continuous construct. It represents a function that takes an attribute values vector as input, and returns a decision.
Decision tree falls within the Supervised Learning group. It can be used to solve regression as well as classification problems. By conducting a series of operations a decision tree makes a decision

Random forest:
Natural forest is a method of bagging, and not of boosting. The trees are running in parallel in random woods. There is no contact among those trees while the trees are being installed.
It performs by constructing a multitude of decision trees during training time and outputting the class which is the particular trees ' class mode (classification) or average prediction(regression).
a random forest[12] is a metaestimator (i.e., it combines the outcome of many predictions) that aggregates many decision trees, with some useful improvements.The number of functions at each node that can be split on is limited to a certain percentagof the total (known as the hyper parameter).

Support vector Regressor:
Support vector machines can also be used as a regression method in ml. SVR gives us the flexibility to define how much error is acceptable in our model and will find an appropriate line to fit the data.
In this algorithm we preprocess the data. It means we do the same process two times in order to get the best fit.

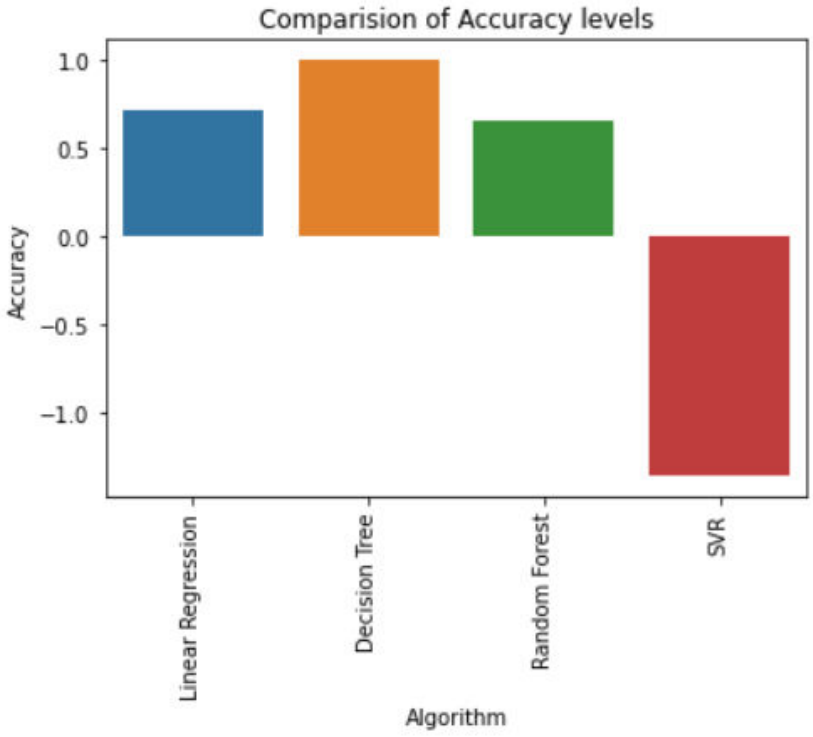
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006048 |

V.RESULTS AND DISCUSSIONS

We implemented the different machine learning algorithms in Python using Jupyter notebook. The following plot shows that all the features that are considered for the prediction are correlated and thus can be considered to train the model.



VI.CONCLUSION

We predict the air quality index by using different machine learning algorithms like.As it is a regression model, we have taken four algorithms to predict air pollution. The model which gives highest score is the best model. From our results we got decision tree value as best .So we conclude Decision tree is the best compared to other algorithms.

REFERENCES

1. K. B. Shaban, A. Kadri, and E. Rezk, “Urban Air Pollution Monitoring System With Forecasting Models” IEEE Sensors Journal, vol. 16, no. 8, pp. 2598–2606, Apr. 2016
2. Li S., and Shue L., "Data mining to aid policy making in air pollution management," Expert Systems with Applications, vol. 27, pp. 331-340, 2004.
3. Gu, Ke, JunfeiQiao, and Weisi Lin. "Recurrent Air Quality Predictor Based on Meteorology and Pollution Related Factors." IEEE Transactions on Industrial Informatics (2018).
4. García Nieto, P.J., Sánchez Lasheras, F., García Gonzalo, E. et al. “Estimation of PM10 concentration from air quality data in the vicinity of the major steel works site in the metropolitan area using machine learning techniques ”StochEnvironResRiskassess(2018),https://link.springer.com/article/10.1007/s00477- 018-1565-6.
5. Hu, Ke, Ashfaqur Rahman, Hari Bhrugubanda, and Vijay Sivaraman. "Hazeest: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors." IEEE Sensors Journal 17, no. 11 (2017): 3517-3525.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details