Project Synopsis

On

AIR POLLUTION FORECASTING IN NCR

Submitted as a part of course curriculum for

Bachelor of Technology

In

Computer Science

Submitted by

Avika Tyagi (2000290120049)

Ankita Kushwaha (2000290120026)

Avi Chaudhary (2000290120048)

Under the supervision of

Prof. Abhishek Goyal

KIET Group of Institutions, Ghaziabad

Department of Computer Science

DR. A.P.J Abdul Kalam Technical University

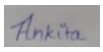
2022-2023

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature of Students





Name: Avi Chaudhary , Avika Tyagi , Ankita Kushwaha

Roll No.: 2000290120048, 2000290120049, 2000290120026
Date: 18-11-22

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the synopsis of the B.Tech Mini Project undertaken during B.Tech. Third Year. We owe a special debt of gratitude to Prof. ABHISHEK GOYAL , Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavours have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Ajay Kumar Shrivastava, Head of the Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature:
Date : 18-11-22
Name : Prof. Abhishek goyal

AIR POLLUTION FORECASTING IN NCR

INTRODUCTION

Air pollution is the world's largest environmental health threat and it is increasing day by day .The effective solution to this global issue is to implement machine learning algorithms to predict the AQI that can make people aware of the condition of the air of a certain region such that certain issued by the government for the improvement of the air quality in future. Because of new inventions there is a rapid increase in the development, serious population growth and increased number of vehicles will give rise to so many critical problems related to the environment such as acid rain, deforestation, air pollution, water pollution, emission of toxic materials and so on. To fulfil the needs of growing population there is the drastic increase in industrialization that may lead to the emission of harmful gases in the atmosphere from various industries that will cause the serious air pollution problem in urban areas throughout the world. This means that the air we or people breathe is not a clean air but it is polluted as so many harmful gases and particles are present in the air that adversely affect the human health. The quality of air degrades due to the pollution.

In most of the urban areas the air pollution becomes a serious concern. The people should know about the air they breathe. The National Ambient Air Monitoring Network generates the data that includes the concentration of various pollutants present in the air but this data is not easily understood by the common people.

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.

Problem Statement-

To estimate the air pollution level using past data-set and to predict the increasing pollution levels after a specific time

We all know that the air contamination is the world's biggest natural wellbeing danger and it is expanding day by day .The powerful answer for this worldwide issue is to execute AI calculations to anticipate the AQI that can make individuals mindful of the state of the demeanor of a specific locale to such an extent that specific gave by the public authority to improve the air quality in future. Due to new creations there is a quick expansion in the turn of events, serious populace development and expanded number of vehicles will lead to such countless basic issues connected with the climate like corrosive downpour, deforestation, air contamination, water contamination, discharge of harmful materials, etc. To satisfy the necessities of developing populace there is the radical expansion in industrialization that might prompt the emanation of destructive gases in the environment from different enterprises that will cause the serious air contamination issue in metropolitan regions all through the world. This implies that the air we or individuals inhale is definitely not a perfect air however it is dirtied as such countless hurtful gases and particles are available in the air that unfavourably influence the human wellbeing. The nature of air corrupts because of the contamination.

LITERATURE REVIEW

Anikender Kumar, Pramila Goyal et.al presented the study that forecasts the daily AQI value for the city Delhi, India using previous record of AQI and meteorological parameters with the help of Principal Component Regression (PCR) and Multiple Linear Regression Techniques. They perform the prediction of daily AQI of the year 2006 using previous records of the year 2000-2005 and different equations. After that this predicted value then compared with observed value of AQI of 2006 for the seasons summer, Monsoon, Post Monsoon and winter using Multiple Linear Regression Technique [4]. Principal Component Analysis is used to find the collinearity among the independent variables. The Principal components were used in Multiple Linear Regression to eliminate collinearity among the predictor variables and also reduce the number of predictors [4]. The Principal Component Regression gives the better performance for predicting the AQI in winter season than any other seasons. In this study only meteorological parameters were considered or used while forecasting the future AQI but they have not considered the ambient air pollutants that may cause the adverse health effects.

Ishan et.al described the benefits of the Bidirectional Long - Short Memory[BiLSTM] method to forecast the severity of air pollution. The proposed technique achieved better prediction which models the long term, short term, and critical consequence of PM2.5 severity levels. In the proposed method prediction is made at 6h, 12h, 24h. The results obtained for 12h is consistent, but the result obtained for 6h, and 24h are not consistent [1].

Chao Zhang et.al proposed web service methodology to predict air quality. They provided service to the mobile device, the user to send photos of air pollution. The proposed method includes 2 modules

a) GPS location data to retrieve the assessment of the quality of the air from nearby air quality stations[2].

b) they have applied dictionary learning and convolution neural network on the photos uploaded by the user to predict the air quality[2].

The proposed methodology has less error rate compared to other algorithms such as PAPLE, DL, PCALL but this method has a disadvantage in learning stability due to this the results are less accurate.

Ruijun Yang et.al used the Bias network to find out the air quality and formed DAG from the data set of the town called as shanghai. The dataset is dived for the training and testing model. The disadvantage of this approach is they have not considered geographical and social environment characteristics, so the results may vary based on these factors[3].

TemeseganWalelignAyele et.al proposed an IoT based technique to obtain air quality data set. They have used Long Short-term Memory [LSTM] technique in-order to predict the air quality the proposed technique achieved better accuracy by reducing the time taken to train the model. But still, the accuracy can be improved by compared other techniques such as the Random forest method NadjetDjebbriet.al [5] proposed artificial based Regressive model which is nonlinear to predict 2 major air pollutants 2

Huixiang Liu et .al have taken two different cities Beijing and Italian city for the study purpose. They have forecasted the Air Quality Index (AQI) for the city Beijing and predicting the concentration of NOxin an Italian City depending on two different publicly available datasets. The first Dataset for the period of December 2013 to August 2018 having 1738 instances is made available from the Beijing Municipal Environmental Centre [5] which contains the fields like hourly averaged AQI and the concentrations of PM2.5, O3, SO2, PM10, and NO2 in Beijing. The second Dataset with 9358 instances is collected from Italian city for the period of March 2004 to February 2005. This dataset contains the attributes as Hourly averaged concentration of CO, Non methane Hydrocarbons, Benzene, NOx, NO2 [5]. But they focused majorly on NOxprediction as it is one of the important predictor for Air Quality evaluation. They used Support Vector Regression (SVR) and Random Forest Regression (RFR) techniques for AQI and NOx concentration prediction. SVR shows better performance in prediction of AQI while RFR gives the better performance in predicting the NOx concentration[8].

Ziyue Guan and Richard O. Sinnot used the various machine learning algorithms to predict the PM2.5 concentration. Data were collected from the official website of Environment Protection Agency (EPA) for the city Melbourne that contains PM2.5 air parameter and they have also collected the unofficial data from Airbeam which is the mobile device used to measure PM2.5 value [11]. The machine Learning Algorithms Artificial Neural Network (ANN), Linear Regression (LR) and Long Short Term Memory (LSTM) recurrent neural network were used

for the PM2.5 prediction but out of these algorithms LSTM gives the best performance ad predict the high PM2.5 value with reasonable Accuracy.

HeidarMaleki et.al predicted the hourly concentration values for the ambient air pollutants NO2, SO2, PM10, PM2.5, CO and O3 for the stations Naderi, Havashenasi, MohiteZist and Behdasht in Ahvaz, Iran which is the most polluted city in the world. They have also calculated and predicted Air Quality Index (AQI) and Air Quality Health Index (AQHI) for the four air quality monitoring stations in Ahvaz mentioned above. They used Artificial Neural Network (ANN) machine learning algorithm for the prediction of air pollutants concentration (hourly) and two air quality indices AQI and AQHI over the August 2009 to August 2010. Input to ANN algorithms involves the factors such as meteorological parameters, Air pollutants concentration, time and date[12].

Plan of work

OBJECTIVES

The main objective of these Networks is to record the concentration levels of atmospheric pollutants in order to define air quality levels and establish action plans if high levels of contamination are detected.

Other objectives are

: • Locating contamination problem areas and understanding their spacetime changes

. • Complying with atmospheric air protection legislation

• Obtaining the necessary information to define Action Plans as stipulated by European directives or other international regulations if alert thresholds are breached.

• Informing citizens regarding local air quality status.

ALGORITHM USED

The Air pollutants information is retrieved from the sensors which are processed in a unified schema and stored as a dataset. This dataset is preprocessed with different functionalities such as normalization, attribute selection and discretization. Once the dataset is ready, it is split into training dataset and test dataset. And further Supervised Machine Learning Algorithms are applied on the training dataset. The obtained results are matched with the testing dataset and results are analyzed. Describes the architecture of the proposed model. The air pollution prediction using Linear Regression and Multi-Layer Perceptron

Linear Regression is used to predict the real values using continuous variables. It is used in many areas such as Economics, Finance, Biology, etc.

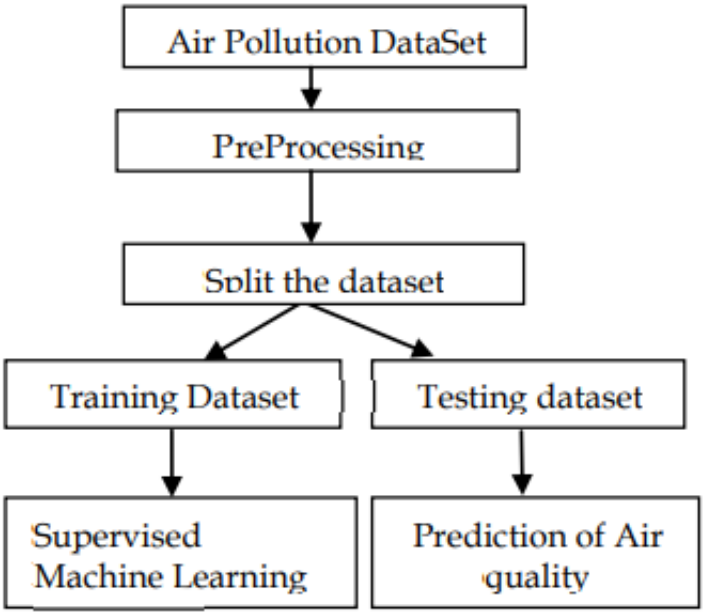


fig.1, Architecture of the proposed model

Assumptions in Linear Regression:

• Linear Relationship: The relationship between independent and dependent variables should be linear. Using the scattered graph, the linearity is tested.

• Little or No Multi Co-linearity between the dependent variable and independent variable.

• Little or No Auto-Correlation: Auto co-relation is the replica of delay of its function. In linear regression, it is assumed that there is no autocorrelation.

• Dependent Residual errors: It occurs when there is too much dependency among the independent variables.

The equation of the line is given by Yp = Xi(a + b), Where Yp is predicted variable, Xi is independent variable, a is slope and b is the intercept. The error E is,

$$E = \sum_{k=1}^{m} (AP - PO)^2$$

which is the square of the difference between actual output to the predicted output, where AP is actual output and PO is predicted output. T.

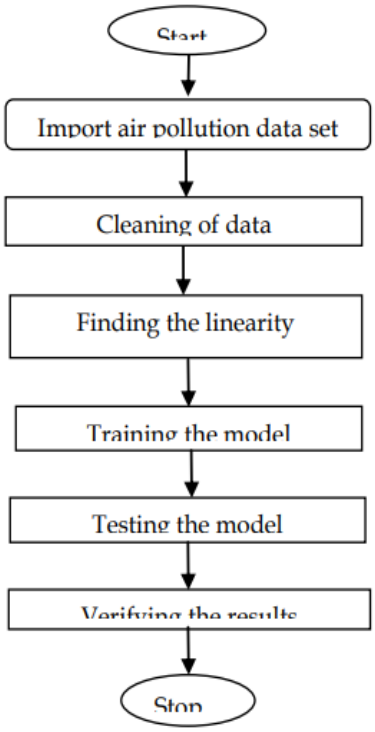


fig.2. flow chart of Linear Regression

Research methodology

The graphs in fig.3. illustrate how meteorological factors, daily industrial waste gas emissions, and AQI varied in Zhangdian District from 1st January 2017 to 31th December 2019. It can be seen that the period with the largest variations in AQI was from December to March, as there are multiple peaks during this time. The minimum value of AQI during these three months was 13 while the maximum AQI value was 313. Between June and August, there were also significant variations in the AQI. In the other months, the range of change was relatively low, with the AQI remaining around 90. The relative humidity fluctuated greatly from February to May, with an average of 46.2%. However, between July and August, the relative humidity only varied slightly, with an average value of 73.7%. The average temperature in February was the lowest, then from March to August it rose slowly, while from August to

October it gradually decreased. The wind scale was relatively stable, although in March and April the wind scale was more erratic. In the other months, the wind scale was generally category 1 or 2. Visibility varied greatly throughout the study period. The average value was about 12.5 km, while the maximum value was 29.5 km and the minimum value was 1.0 km. The average pressure in July was the lowest with an average of 98.7 kPa then from August to December it rose slowly, while from December to July it gradually decreased. Total sunshine intensity varied greatly throughout the study period. The average value was about 15.85 J/m_2, while the maximum value was 28.59 J/m_2 and the minimum value was 0.66 J/m_2. The precipitation fluctuated greatly from February to May, with an average of 46.2%. Finally, the average daily emissions of industrial waste gas over the whole study period were 153 million cubic meters, while the maximum value was 270 million cubic meters and the minimum value was 100 million cubic meters. Average daily emissions of industrial waste gas in 2019 were 33 million cubic meters and 85 million cubic meters more than in 2018 and 2017, respectively, but the AQI annual average in 2019 was lower than both 2018 and 2017. Because Shandong Province implemented several air pollutant emission standards (“Emission standard of air pollutants for building materials industry”, Effective January 1, 2019) (“Emission standard of air pollutants for industrial furnace and kiln”, Effective June 1, 2019) in 2019, stricter pollutant emission concentration limits were implemented.

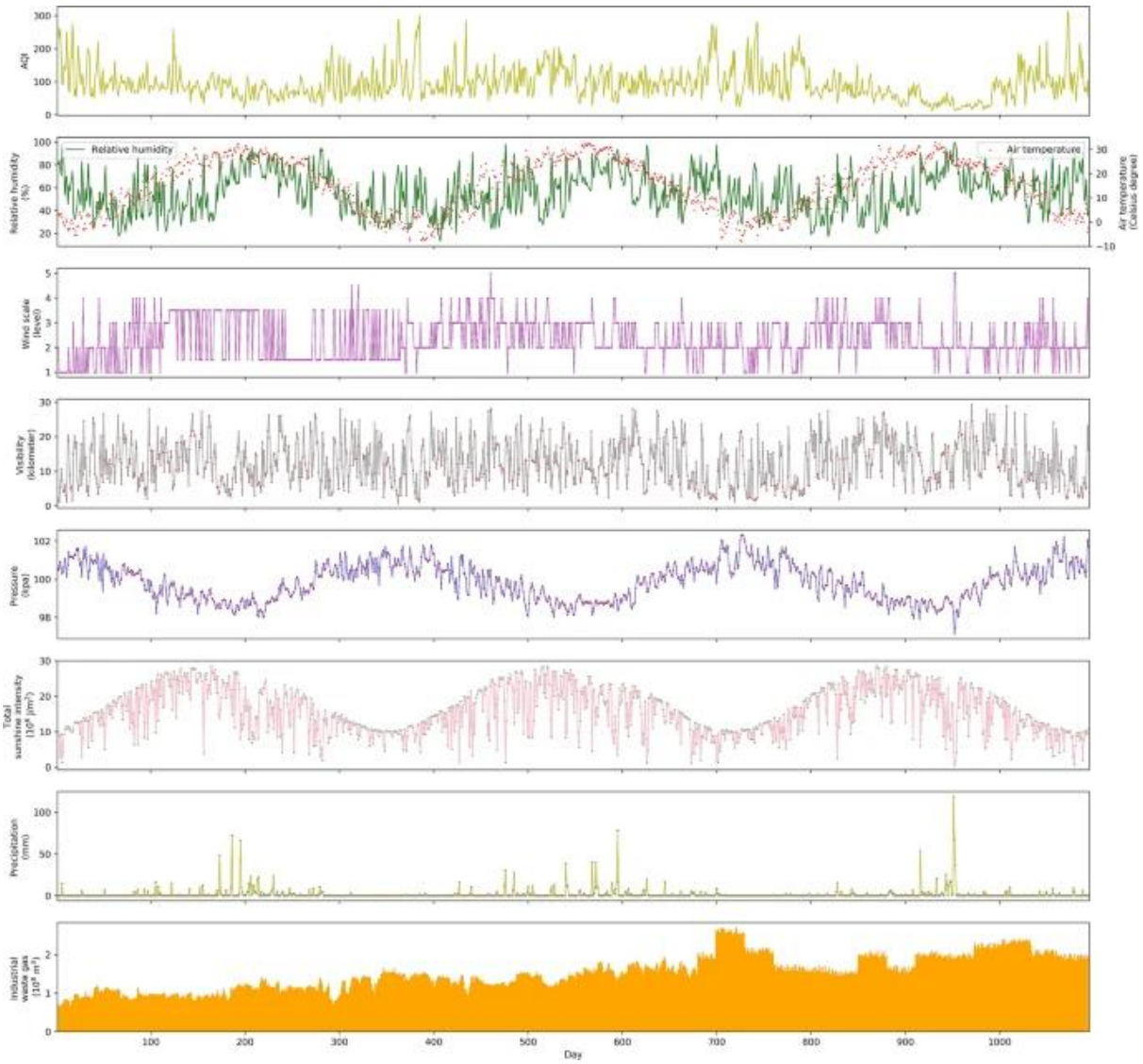


fig.3. Trends of meteorological factors, industrial waste gas emissions, and AQI.

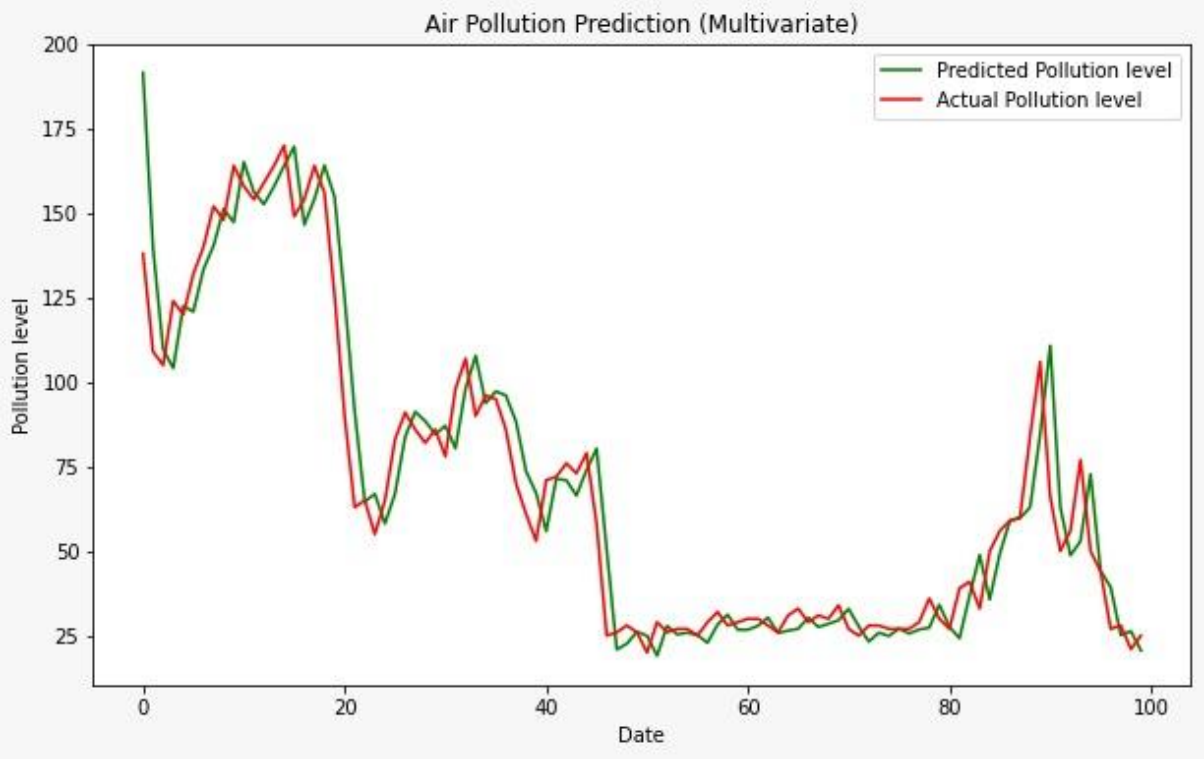


fig.4. graph between actual vs predicted values

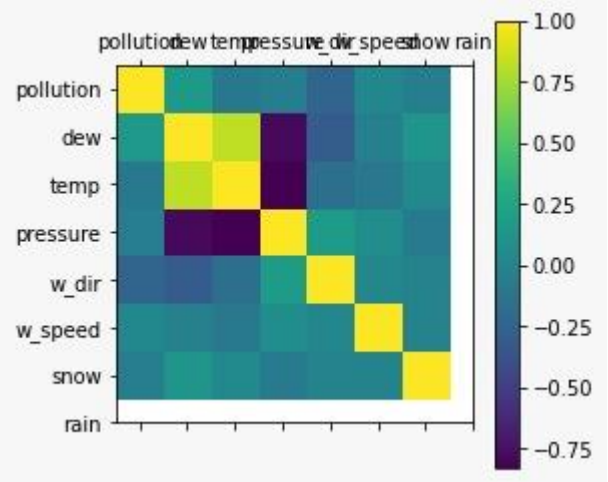


fig.5. relationship between different features

Conclusion

Throughout this project, several models which can predict Pm2.5 levels and classify them into different pollution bands were experimented and their performance was successfully evaluated. The exploratory data analysis and feature engineering methods implemented for the prediction models revealed interesting correlations between weather and pollution data. We obtained several notable outcomes from the predictive models that are worth being discussed.

Future scope of work

Combining the above model with a real time API would help in predicting the pollution bandsalmost in real time. Further research can be done on Geographic Information System (GIS) on Air Pollution which can help make decisions to geography based on human thinking patterns. It can also organize geographic data, to select and use specific task or project like checking Air quality at location. We can also improve our project by working with live API capability. This readily available API service also allows the predictions of AQI in real time.

Reference

[1]. Verma, Ishan, Rahul Ahuja, HardikMeisheri, andLipikaDey. ”Air pollutant severity rediction using Bi-directional LSTM Network.” In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 651-654. IEEE, 2018.

[2]. Figures Zhang, Chao, Baoxian Liu, Junchi Yan, Jinghai Yan, Lingjun Li, Dawei Zhang, XiaoguangRui, and RongfangBie. ”Hybrid Measurement of Air Quality as a RH w.r.t tin oxide RH w.r.t C6H6 Mobile Service: An Image Based Approach.” In 2017 IEEE International Conference on Web Services (ICWS), pp. 853- 856. IEEE,2017.

[3]. Yang, Ruijun, Feng Yan, and Nan Zhao. ”Urban air quality based on Bayesian network.” In 2017 IEEE 9th Fig. 10. RH w.r.t NO RH w.r.t NO2

[4]. Anikender Kumar, PramilaGoyal, “Forecasting of air quality in Delhi using principal component regression technique”, Atmospheric Pollution Research, 2 (2011) 436-444.

[5]. https://www.aqi.in/blog/aqi/

[6].https://www.researchgate.net/profile/Shovan_Sahu/pub lication/315725810/figure/tbl1/AS:668795018440728 @1536464566616/Breakpoints-of-different-pollutantsin-IND-AQI-CPCB-2014.png

[7]. https://app.cpcbccr.com/ccr_docs/FINALREPORT_AQI_.pdf

[8].Huixiang Liu, Qing Li, Dongbing Yu, Yu Gu, “Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms”, Applied Sciences, ISSN 2076-3417; CODEN: ASPCC7, 2019, 9, 4069; doi:10.3390/app9194069.

[9]. PoojaBhalgat, SejalPitale, Sachin Bhoite, “Air Quality Prediction using Machine Learning Algorithms”, International Journal of Computer Applications Technology and Research Volume 8–Issue 09, 367- 370, 2019, ISSN:-2319–8656.

[10]. https://www.lung.org/clean-air/outdoors/air-qualityindex

[11]. Ziyue Guan and Richard O. Sinnot, “Prediction of Air Pollution through Machine Learning on the cloud”, IEEE/ACM5th International Conference on Big Data Computing Applications and Technologies (BDCAT), 978-1-5386-5502-3/18/$31.00 ©2018 IEEE DOI 10.1109/BDCAT.2018.00015.

[12]. Heidar Malek, Armin Sorooshian, Gholamreza Goudarzi, Zeynab Baboli, Yaser Tahmasebi Birgani, Mojtaba Rahmati, “Air pollution prediction by using an artifcial neural network model”, Clean Technologies and Environmental Policy, (2019) 21:1341–1352.

[13]. Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu, “Detection and Prediction of Air Pollution using Machine Learning Models”, International Journal of Engineering Trends and Technology (IJETT) – volume 59 Issue 4 – May 2018

[14]. https://www.iqair.com/us/india

[15]. Pallavi Pant, Raj M. Lal, Sarath K. Guttikunda, Armistead G. Russell, Ajay S. Nagpure, AnuRamaswami, Richard E. Peltie, “Monitoring particulate matter in India: recent trends and future outlook”, Air Quality, Atmosphere & Health, 2018.

[16]. M. Bahattin CELIK, Ibrahim KADI, “The Relation Between Meteorological Factors and Pollutants Concentrations in Karabuk City”, G.U. Journal of Science, 20(4): 87-95 (2007).

[17]. Nidhi Sharma , ShwetaTaneja , VaishaliSagar , Arshita Bhatt, “Forecasting air pollution load in Delhi using data analysis tools”, ScienceDirect, 132 (2018) 1077– 1085.

[18]. Mohamed Shakir, N. Rakesh, “Investigation on Air Pollutant Data Sets using Data Mining Tool”, IEEE Xplore Part Number:CFP18OZV-ART; ISBN:978-1- 5386-1442-6.

[19]. KazemNaddafi, Mohammad SadeghHassanvand, MasudYunesian, , FatemehMomeniha, RaminNabizadeh, SasanFaridi, Akbar Gholampour, “Health impact assessment of air pollution in megacity of Tehran, Iran”, IRANIAN JOURNAL OF ENVIRONMENTAL HEALTH SCIENCE & ENGINEERING, 2012, 9:28
[20]. YusefOmidiKhaniabadi, GholamrezaGoudarzi, Seyed Mohammad Daryanoosh, Alessandro Borgini, Andrea Tittarelli, Alessandra De Marco, “Exposure to PM10, NO2, and O3 and impacts on human health”, Environ SciPollut Res, 2016.