RESEARCH ARTICLE OPEN ACCESS

Air Quality Detection Using Machine Learning Techniques

Mrs R Jhansi Rani MCA [1], V Hari chandana [2]

[1] Asst. Professor, Department Computer Applications

[2] Student, Department of Computer Applications

[1], [2] Chadalawada Ramanamma Engineering College (Autonomous)

ABSTRACT

Over the past few decades, due to human activities, industrialization, and urbanization, air pollution has become a life-threatening factor in many countries around the world. Air, an essential natural resource, has been compromised in terms of quality by economic activities. Air pollution is a severe problem in areas where population density is high such as metropolitan cities. Various types of emissions caused by people's actions, such as transportation, power, and fuel use, are affecting air quality. Considerable research has been devoted to predicting instances of poor air quality, but most studies are limited by insufficient longitudinal data, making it difficult to account for seasonal and other factors. Several prediction models have been developed using an 11-year dataset collected by Taiwan's Environmental Protection Administration (EPA). We forecast the air quality by using machine learning to predict the air quality index of a given area. Air quality index is dataset for a standard measure used to indicate the pollutant (so2, no2, rspm, spm. etc.) levels over a period. The ML models like Decision tree and Random Forest Classifier has been implemented and compared to show the better accuracy.

Keywords: - Air Quality, Environment, Decision Tree, Learning Algorithm.

I. INTRODUCTION

The objective of this project is to show how sentimental analysis can help improve the user experience over a social network or system interface. The learning algorithm will learn what our emotions are from statistical data then perform sentiment analysis. Our main objective is also maintain accuracy in the final result. The main goal of such a sentiment analysis is to discover how the audience perceives the television show. The Twitter data that is collected will be classified into two categories; positive or negative. An analysis will then be performed on the classified data to investigate what percentage of the audience sample falls into each category. Particular emphasis is placed on evaluating different machine learning algorithms for the task of twitter sentiment analysis. 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 metropolitician 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. Examining and protecting air quality has become one of the most essential activities for the government in many industrial and urban areas today. The meteorological and traffic factors, burning of fossil fuels, and industrial parameters play significant roles in air pollution.With this increasing air pollution,Weare in need of implementing models which will record information about concentrations of air pollutants(so2,no2,etc).The deposition of this harmful gases in the air is affecting the quality of people's lives, especially in urban areas. Lately, many researchers began to use Big Data Analytics approach as there are environmental sensing networks and sensor data available. Air Quality Index(AQI), is used to measure the quality of air. Earlier classical methods such as probability, statistics were used to predict the quality of air, but those methods are very complex to predict the quality of air. Due to advancement of technology, now it is very easy to fetch the data about the pollutants of air using sensors and then store the data in files. Assessment of raw data to detect the pollutants needs vigorous analysis like ML models. Worldwide, air pollution is responsible for around 1.3 million deaths annually according to the World Health Organization (WHO). The depletion of air quality is

just one of harmful effects due to pollutants released into the air. Other detrimental consequences, such as acid rain, global warming, aerosol formation, and photochemical smog, have also increased over the last several decades. The recent rapid spread of COVID-19 has prompted many researchers to investigate underlying pollution-related conditions contributing to COVID-19 pandemics in countries. Several shreds of evidence have shown that air pollution is linked to significantly higher COVID-19 death rates, and patterns in COVID-19 death rates mimic patterns in both high population density and high PM2.5 exposure areas. All the above mentioned raises an urgent need to anticipate and plan for pollution fluctuations to help communities and individuals better mitigate the negative impact of air pollution. To do so, air quality evaluation plays a significant role in monitoring and controlling air pollution.

In the developing countries like India, the rapid increase in population and economic upswing in cities have lead to environmental problems such as air pollution, water pollution, noise pollution and many more. Air pollution has direct impact on humans health .There has been increased public awareness about the same in our country.Global warming, acid rains, increase in the number of asthma patients are some of the long-term consequences of air pollution. Précised air quality forecasting can reduce the effect of maximal pollution on the humans and biosphere as well. Hence, enhancing air quality forecasting is one of the prime targets for the society. Sulphur Dioxide is a gas. It is one of the major pollutants present in air.It is colourless and has a nasty, sharp smell.It combines easily with other chemicals to form harmful substances like sulphuric acid, sulphurous acid etc. Sulphur dioxide affects human health when it is breathed in. It irritates the nose, throat, and airways to cause coughing, wheezing, shortness of breath, or a tight feeling around the chest. The concentration of sulphur dioxide in the atmosphere can influence the habitat suitability for plant communities, as well as animal life. The proposed system is capable of predicting quality of air using ML models.

II. RELATEDWORKS

Supervised learning is built to make prediction, given an unforeseen input instance. A supervised learning algorithm takes a known set of input dataset and its known responses to the data (output) to learn the regression/classification model. An algorithm is used to learn the dataset and train it to generate the model for prediction of rainfall for the response to new data or test data. Supervised learning uses classification algorithms and regression techniques to develop predictive models. Methods have performance limitations because of wide range of variations in data and amount of data is limited. Issue involved in rainfall classification is choosing the required sampling recess of Observation-Forecasting of rainfall, which is dependent upon the sampling interval of input data. Less accuracy. B. Lakshmi Sravya , A.S. MahaLakshmi , D.Balaji Bhavya Swarupini, B.V. Sai Jaswanth proposed A Deep Learning based Air Quality Prediction. Industries are the major means of air pollutants. Air pollution in the form of carbon dioxide and methane raisesthe earth's temperature, the less gasoline we burn, the better we do to reduce air pollution and harmful effects of climate change. Especially at metropolitan cities, the change in the temperature combined with harmful chemicals may lead to dangeroussigns of air pollution. Quality of air prediction techniques has a major importance in the current learning world. Many machine learning algorithms done a lot of research in identifying the air quality index. Applying deep learning models on these data can showgreat difference in predicting the quality of air. We proposed an LSTM based deep learning technique in evaluating hourly based encompassing air quality. The proposed results outperformed the existing model results through predicting RMSE value.

Heni Patel, Swarndeep Saket proposed Air Pollution Prediction System for Smart City using Data Mining Technique: A Survey**.** Air pollution is one of the major hazards among the environmental pollution. As each living organism needs fresh and good quality air for every second. None of the living things can survive without such air. But because of automobiles, agricultural activities, factories and industries, mining activities, burning of fossil fuels our air is getting polluted. These activities spread sulphur dioxide, nitrogen dioxide, carbon monoxide, particulate matter pollutants in our air which is harmful for all living organism. The air we breathe every moment causes several health issues. So we need a good system that predicts such pollutions and is helpful in better

environment. It leads us to look for advance techniques for predicting the air pollution. So here we are predicting air pollution for our smart city using data mining technique. In our model we are using multivariate multistep Time Series data mining technique using random forest algorithm. Our system takes past and current data and applies them to our model to predict air pollution. This model reduce the complexity and improves the effectiveness and practicability and can provide more reliable and accurate decision for environmental protection departments for smart city.

M.Gayathri, R. Shankar and S. Duraisamy proposed Air Pollution Prediction using Data Mining Technique. Air pollution is one ofthemajor hazards among the environmental pollution. As each living organism needs fresh and good quality air for every second? None of the living things can survive without such air. But because of automobiles, agricultural activities, factories and industries, mining activities, burning of fossil fuels our air is getting polluted. These activities spread sulphur dioxide, nitrogen dioxide, carbon monoxide, particulate matter pollutants in our air which is harmful for all living organism. The air we breathe every moment causes several health issues. So we need a good system that predicts such pollutions and is helpful in better environment. So here we are predicting air pollution for our city using data mining technique. In our model we are using data mining technique c4.5decision tree algorithm. Our system takes past and current data and applies them to our model to predict air pollution. This model reduces the complexity and improves the effectiveness and practicability and can provide more reliable and accurate decision for environmental city.

Kalash Agarwal, Yatender Singh, Jasmendra Singh, Abhishek Goyal proposed Air Pollution Predication Using Machine Learning. In the populated and developing countries, governments consider the regulation of air as a major task. Monitoring air quality is a necessary exercise in the meteorological and movement factors, stubble burning and open construction practice these factors contribute a lot in air pollution. So forecast air quality index using a machine learning model to predict air quality index for NCR(national capital region). The values of major pollutants like SO2, PM2.5, CO, PM10, NO2, and O3.in recent years machine learning in most emerging technology for predicting on historical data with 99.99% of accuracy. we implemented different classification and regression techniques like Linear Regression, multiple linear regression, KNN, Random Forest Regression, Decision Tree Regression, Support Vector Regression, Artificial Neural Networks. To make more accurate our prediction use Mean square error, mean absolute error and R square. To prognosticating air quality index of NCR(national capital region) in different aspects of like stubble farming, Motor vehicle emission, and open construction practice which result in the air quality of NCR.

Heni Patel, Swarndeep Saket proposed Air Pollution Prediction System for Smart city using Data Mining Technique**.** Each living organism needs fresh and good quality air for every second. None of the living things can survive withoutsuch air. But today air pollution becomes one of the major hazards. Because of automobiles, agricultural activities, factories and industries, mining activities, burning of fossil fuels our air is getting polluted. These activities spread sulphur dioxide(SO2), nitrogen dioxide(NO2), carbon monoxide(CO), Ozone(O3), particulate matter(PM10&PM2.5) pollutants in our air which is harmful for all living organism. The air we breathe every moment causes several health issues. So we need a good system that predicts such pollutions and is helpful to make our environment better. It leads us to look for advance techniques for predicting such pollutants. So here we are predicting air pollution for our smart city using data mining technique. In our model we are using multivariate multistep Time Series data mining technique using random forest algorithm. Our system takes time series data ofthese pollutants. Also takes data of temperature, wind speed & direction and applies them to our model to predict air pollution. This model reduces the complexity and improves the effectiveness and practicability and can provide more reliable and accurate decision for environmental protection departments for smart city.

III. PROPOSED SYSTEM ARCHITECTURE

Earlier techniques such as Probability, Statistics etc. were used to predict the quality of air, but those methods are very complex to predictthe

Machine Learning (ML) is the better approach to predict the air quality. With the need to predict air relative humidity by considering various parameters such as CO, Tin oxide, nonmetallic hydrocarbons, Benzene, Titanium, NO, Tungsten, Indium oxide, Temperature etc, In this system the air quality of an environment can be predicted with the help of machine learning algorithm like RF and decision tree based on previous weather details. 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.

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 ML algorithms.

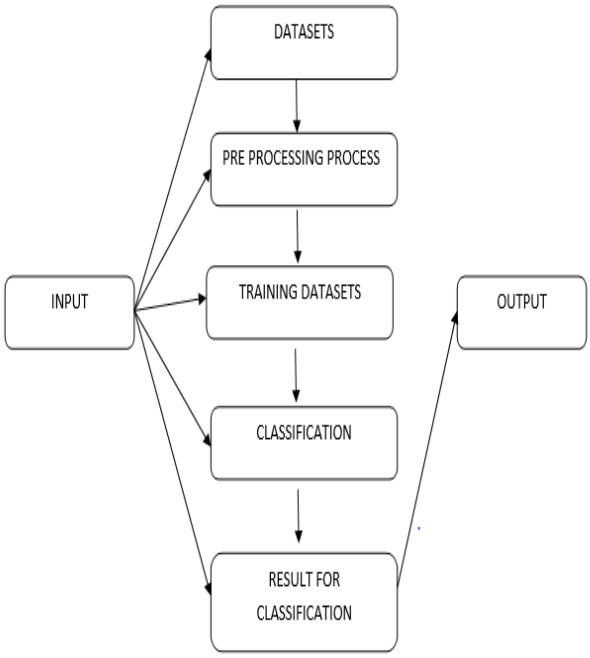

Fig.1 Proposed methodology

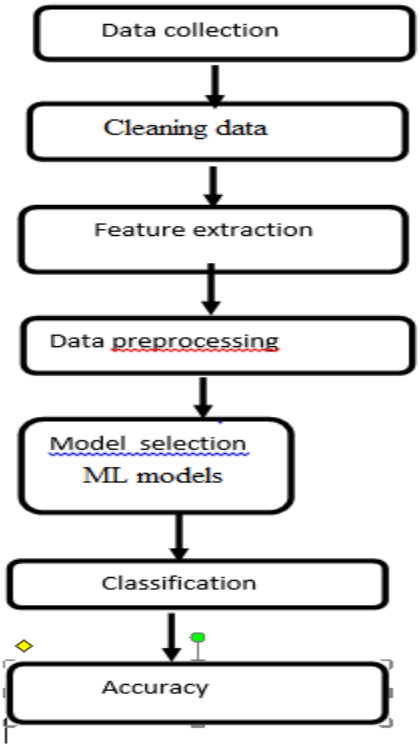

Fig.2 Process flow of proposed methodology

IV. RESULTS AND DISCUSSION

The output screens obtained after running and executing the system are shown from Fig.3 to Fig. 7

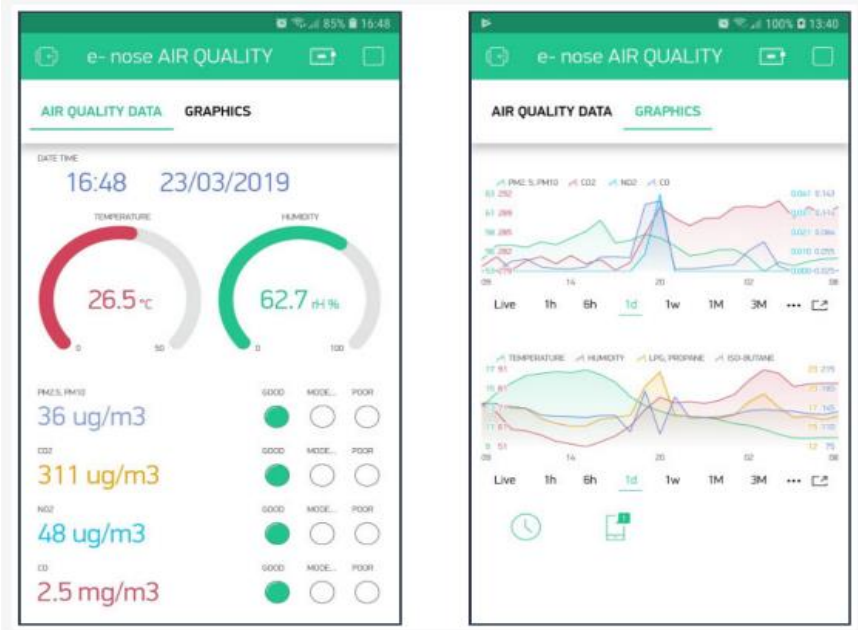

Fig.3 Mobile User Interface

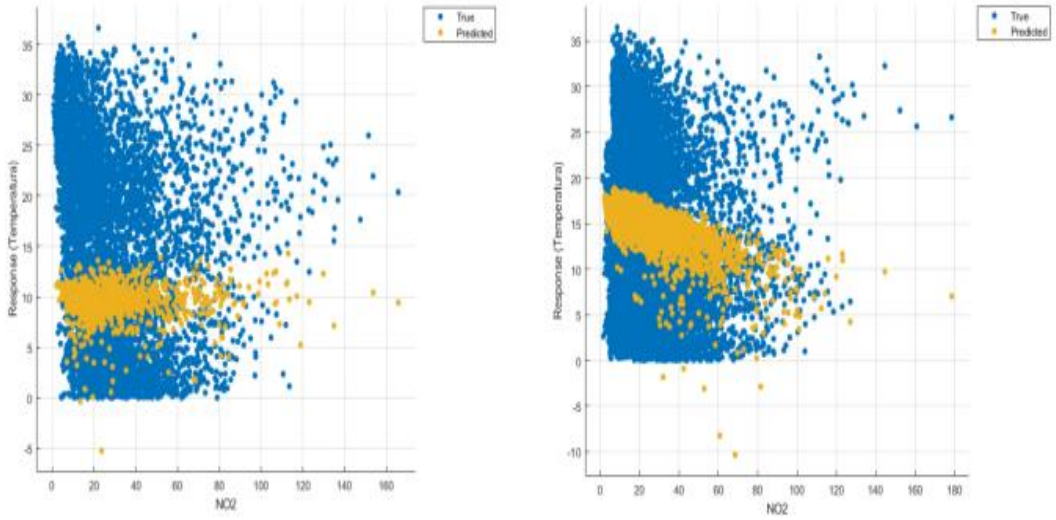

Fig. 4 Prediction considering only NO2 (a) March 2019–February 2020, (b) March 2020– February 2021.

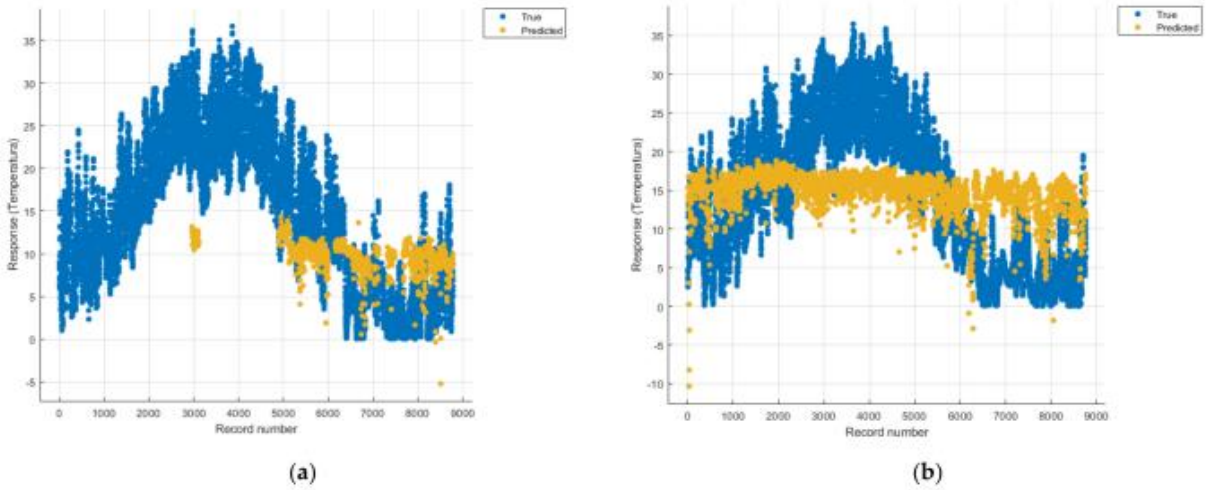

Fig.5 Prediction considering all pollutants (a) March 2019–February 2020, (b) March 2020– February 2021.

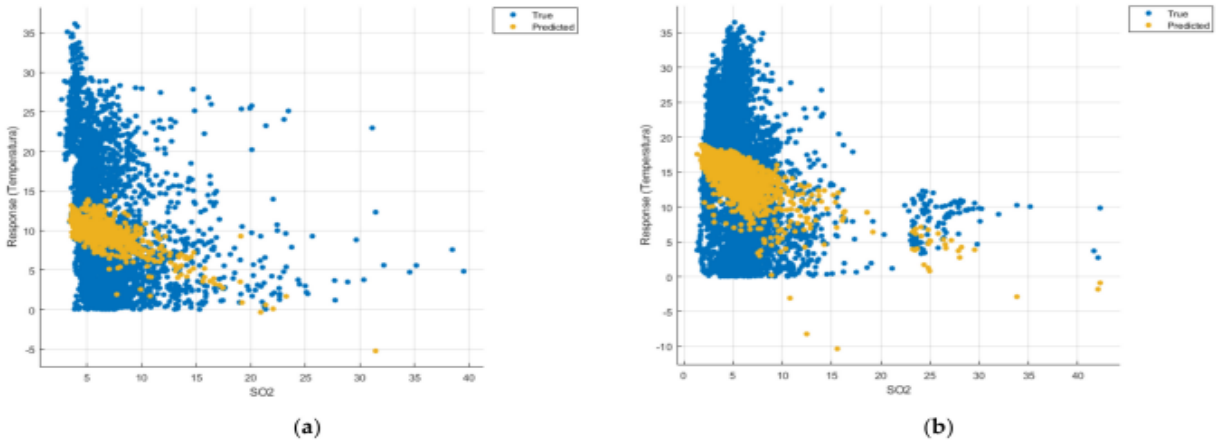

Fig. 6 Prediction considering only SO2—Linear Regression. (a) March 2019–February 2020, (b) March 2020–February 2021.

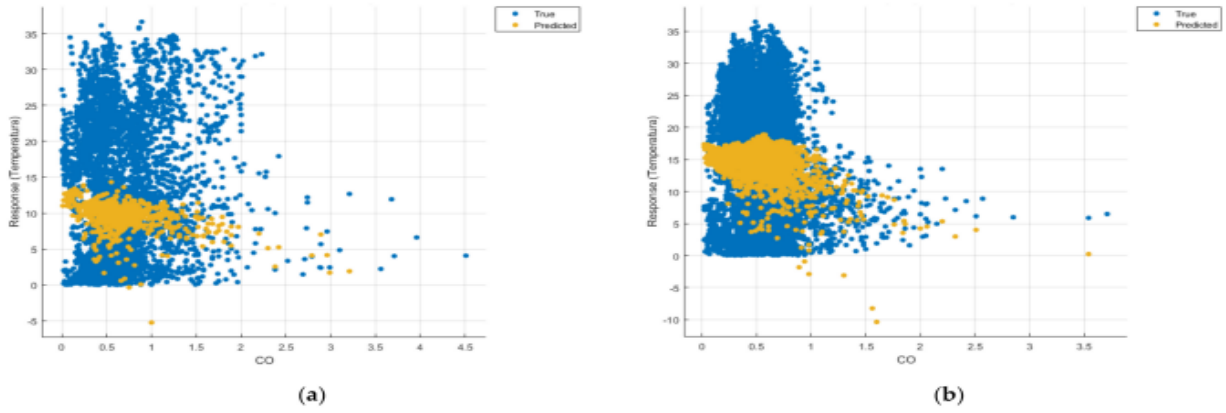

Fig. 7. Prediction considering only CO—Linear Regression. (a) March 2019–February 2020, (b) March 2020–February 2021.

V. FUTURE SCOPE AND CONCLUSION

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 agenda of our work is not only to bring awareness but also to minimize pollution through proper measures and ensure that the vehicles are emitting the pollutants within the range of regular pollution check. This can lead to a pollution free region in the area. 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.

REFERENCES

[1] Shweta Taneja, Dr. Nidhi Sharma, Kettun Oberoi, Yash Navoria ,"Predicting Trends in Air

Pollution in Delhi using Data Mining", IEEE(2016)

[2] Nidhi Sharmaa, Shweta Tanejab*, Vaishali Sagarc, Arshita Bhattd, "Forecasting air pollution load in Delhi using data analysis tools.", Elseviere (ICCIDS 2018)

[3] KRZYSZTOF SIWEK, STANISŁAW OSOWSKI," Data mining methods for prediction of Air Pollution", amcs(2016)

[4] Mansi Yadav, Suruchi Jain and K. R. Seeja," Prediction of Air Quality Using Time Series Data Mining", Springer (2019)

[5] Manisha Bisht and K.R. Seeja," Air Pollution Prediction Using Extreme Learning Machine: A Case Study on Delhi.", Springer(2018)

[6] Khaled Bashir Shaban, Senior Member, IEEE, Abdullah Kadri, Member, IEEE, and Eman Rezk," Air Pollution Monitoring System With Forecasting Models.", IEEE(2016)

[7] Khaled Bashir Shaban, Abdullah Kadri, Eman Rezk, "Urban Air Pollution Monitoring System With Forecasting Models",IEEE SENSORS JOURNAL, VOL. 16, NO. 8, APRIL 15, 2016

[8] Forecasting Criteria Air Pollutants Using Data Driven Approaches; An Indian Case Study Tikhe Shruti, Dr. Mrs. Khare , Dr. Londhe ,IOSR-JESTFT (Mar. - Apr. 2013)

[9] Forecasting Criteria Air Pollutants Using Data Driven Approaches; An Indian Case Study Tikhe Shruti, Dr. Mrs. Khare , Dr. Londhe,IOSR-JESTFT (Mar. - Apr. 2013)

[10] Air Quality Forecasting Methods, https://www.airvisual.com/air-pollutioninformation/research/air-quality-forecast-methods

[11] Multivariate Multistep Time series Forecasting model for Air Pollution. https://machinelearningmastery.com/how-to-developmachine-learning-models-for-multivariate-multi-stepair-pollution-time-series-forecasting/K. Elissa, "Title of paper if known," unpublished.

[12] Yi-Ting Tsai, Yu-Ren,Zeng, Yue-Shan Chang, "Air pollution forecasting using RNN with LSTM", IEEE(2018)