A MACHINE LEARNING APPROACH TO PREDICT AN AIR QUALITY INDEX

[1]A .Kiran Kumar, [2]Chandolu Bhargav Sankar, [3]Pettam Nikhil, [4]Kyadasi Lokesh
[1] Assistant Professor of Dept of CSE,CMR Technical Campus,Medchal, Hyderabad
[2,3,4] B.Tech Student,Dept of CSE, CMR Technical Campus, Hyderabad

Abstract: In this paper, the aim is to predict the air quality index of the areas that helps to find the concentration of the air pollutants in every hour. The air pollutants include O3, particulate matter (PM2.5,PM10), Sulphur dioxide (SO2), Carbon Monoxide (CO), Nitrogen oxide (NO2), Lead (Pb), Ammonia (NH3) etc. Machine Learning is one of the popular techniques which is used to train the modules on the big data by the usage of many machine learning algorithms. In this paper, we planned to redefine the model to predict the air pollution hour by hour, day by day, year by year and check the concentrations of air pollutants on the basis of meteorological data of the past days by preparing them over 24hr as the multitask learning problem (MTL). We plan to regularize the model by predicting the air quality index in every hour to be close to each other and compare it with different regularization of the multitask learning problem (MTL) including other approaches to compare the data and to predict the air quality. Our experiment is performed by using machine learning algorithms like Linear Regression, Lasso and Ridge Regression, Random forest Regression, Xgboost Regression are used to help in comparing a different data and make it efficient and accurately analyze the input data. With the help of these machine learning algorithms we can easily find out the concentration of air pollutants along with their air quality index.

Keywords — Machine Learning, Air Quality Index, Air Quality Prediction, Multitask Learnin g, Linear Regression, Lasso and Ridge Regression, Random forest Regression, Xgboost Regressi on.

I. INTRODUCTION

The increase in the concentration of air pollution is one of the major problems of the world. Now a days many people suffer from air borne related respiratory diseases. Mainly in the urban region or areas there are many factories which produces most of the harmful air pollutants like SO2, NO 2, CO etc, which causes the human beings and animals largely effected due to these pollutants. T he rise in the concentration of these harmful chemicals leads to increase in the air pollution so rap

idly. One of the main reasons of change in weather is due to air pollution. To predict the air pollu tion of the urban areas, helps in reducing the air pollution in the environment. The different pollu tion through different aspects in the environment can be observed and concentrated during differ ent period of time.

The negative impact of the air pollution is to expose the environment. The harmful air pollutants and many organic compounds along with the plastics and the metals present on our surroundings. Based on the previous studies of data from America Lung Association 10 parts per billion increas e in pollution due to ozone (O3) that result to the death of people. Mumbai and many other large cities of India has struggled with the air pollution as a result of industrialization and the urbaniza tion. Particle size is most dangerous and is the critical in finding the particle decomposition in th e human respiratory system. PM 2.5 is defined as the particle matter with the diameter of 2.5 mic rometer, has been increased to the matter and these particles can be deposited into the human rep ository system. Meteorological conditions, including regional meteorological are most dangerous in determining the air pollutant concentrations. According to the study, the ozone (O3) concentra tion was found to be most sensitive to the air pollution substance.

II. LITERATURE SURVEY

It's apparent that folks that labor in a company or concern are possible to be exposed to the hazar d of breathing damaging chemical compounds and air because of their constant acquaintance to p ollutants. Air pollution provides to the hazardous situation that creates negative influence on dwe lling items. It is likely an actual consideration for the whole earth. Contamination of the air is a g lobal trouble even for multi-national companies, government and the mass media. Making use of ordinary possession at a better fee than the character's capability to rebuild it can cause pollution of vegetation, air, and water. As opposed to the works done by people, there are other causes that result in the release of harmful toxins. Apart from quests made by men, natural calamities remini scent of many kinds may have an outcome in infecting the air. Technology has advanced in almo st all areas and movements of living organisms. In the current world everything is completed ma ing use of new science with a view to satisfy the demand of person, institution, manufacturer and so on. Internet of matters (IOT) is without doubt one of the predominant exchanging information traits within the last ten years. By means of this thought, it is feasible to attach numerous embedd ed objects that consume less power.

III. EXISTING SYSTEM

Prediction of air pollution was already a project in which most of them uses big data technique an

the data mining technique. But in those projects people used the concept of big data to get and ret rieve the data to know about the predicted air quality and the concentration levels of it's pollutant s. Like concentration of Ozone (O3), concentration of sulphur dioxide (SO2), Nitrogen oxide(NO 2).

Greenhouse gas estimation

Now a days air pollution is one of the big challenge's facing in our life. The main sources of gree n house effect emissions are caused by Transportation. The two approaches are used here to estim ate the greenhouse effect emissions which are bottom up and top down model approach's.This mo dels was proposed by IPCC. In top down approach the total CO2 level is calculated from total fue l consumption in the urban areas. These carbon levels are transmitted into the transportation secto rs. Whereas in bottom up approach each economic fuel level's are calculated then only the carbon emission of the fuel is calculated.

Matern function based extended fractional Kalman filtering

This technique was used to predict the air pollution emission. Unpredictably in data involves the measurement error or the instrument precision related error. Hence, the input data are only observ ed and filtering techniques are suffered from few obstacles.

FCM-HMM Clustering and TS Fuzzy Inference Algorithm

The multi model framing system is based on FCM-HMM clustering technique and the TS Fuzzy algorithm model is used to predict the air pollution index. Here, the data used were meteorologic al and dependencies between the hidden state cannot be expressed.

IV. PROPOSED SYSTEM

In proposed system, the Linear regression, Lasso and Ridge regression, Random Forest regressio n, Xgboost regression algorithms are used for the prediction of air pollution of next day.
The working of the proposed system is as follows:

1. Fetching Pollution Data:

Fetching some components like SO2, NO2, PM10, PM2.5, CO, Ozone, Pb, NH_3 etc. Parameters used are Average Temperature (^oC), Maximum Temperature (^oC),Minimum Temperature(^oC), A verage relative humidity (%), Total rainfall and / or snow melt (mm), Average Visibility (Km), A verage wind speed (Km/h), Maximum sustainable wind speed (Km/h) etc. These parameters are r epresented as T, TM, Tm, H, PP, VV, V, VM etc. Each parameters will be used for analysis of ai

r pollution in the city areas.

2. Preprocessing:

The data we get from different sources may contain inconsistent data, missing values and repeate d data. To get proper prediction result, the dataset must be cleaned, missing values must be taken care of either by deleting or by filling with mean values or some other method. Also, redundant d ata must be removed or eliminated so as to avoid biasing of the results. Some dataset may have s ome outlier or extreme values which also have to be removed to get good prediction accuracy. C lassification and clustering algorithms and other data mining methods will work well only if all t his preprocessing is done on the data.

3. Building the classification model:

I. Predicting the excellent air is an obstacle, choice of tree algorithm prediction mannequin is ef fective on account that of the following factors: It presents higher outcome in classifications dif ficulty.

II. First we have to divide the dataset into training and testing set. The predicting model is firstly trained with the training dataset. Later it will be tested with the testing dataset. Otherwise k-fol d cross validation can also be used.

III. After testing the model, the accuracy of the model is estimated by using different parameters.

4. Training on Existing data:

Existing data along with their air quality parameter will be provided in training dataset. The outp ut of the training dataset will be air quality index. Data will be provided in bar graph in the form of SO2, NO2, PM10, PM2.5, CO, NH_3, Pb, Ozone.

The algorithms that have been used in this project are:

a. Linear Regression

b. Lasso and Ridge Regression

c. Random Forest Regression

d. XGBoosting Regression

e. Adaptive Boosting Regression

f. Support Vector Machine

V. RESULTS

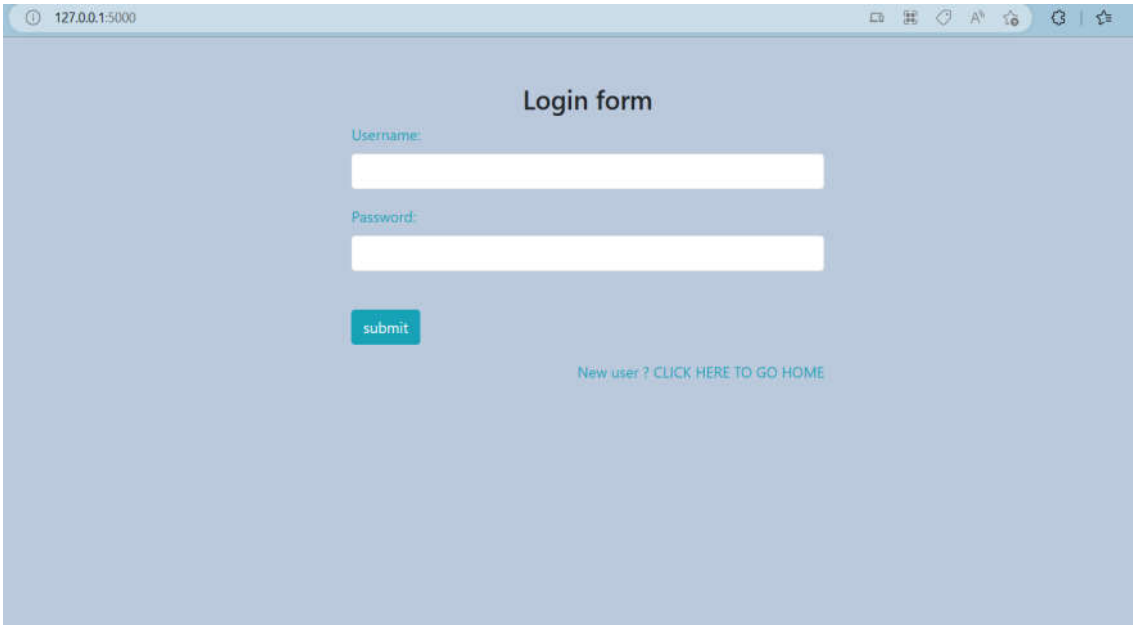

Figure 5.0: Login Screen

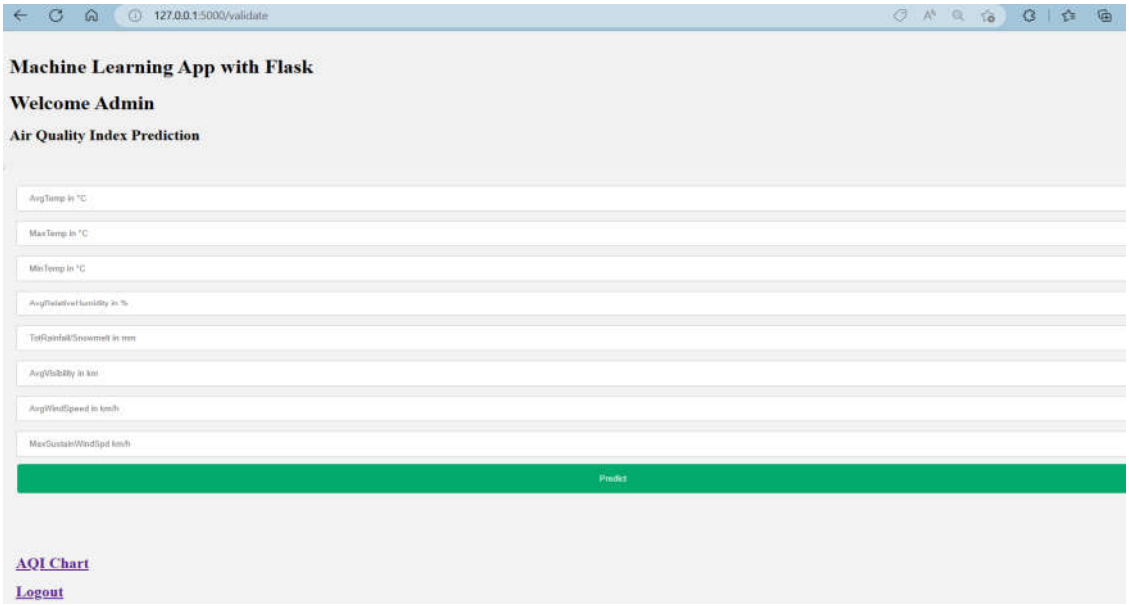

Figure 5.1: ML Web Application with Flask Interface

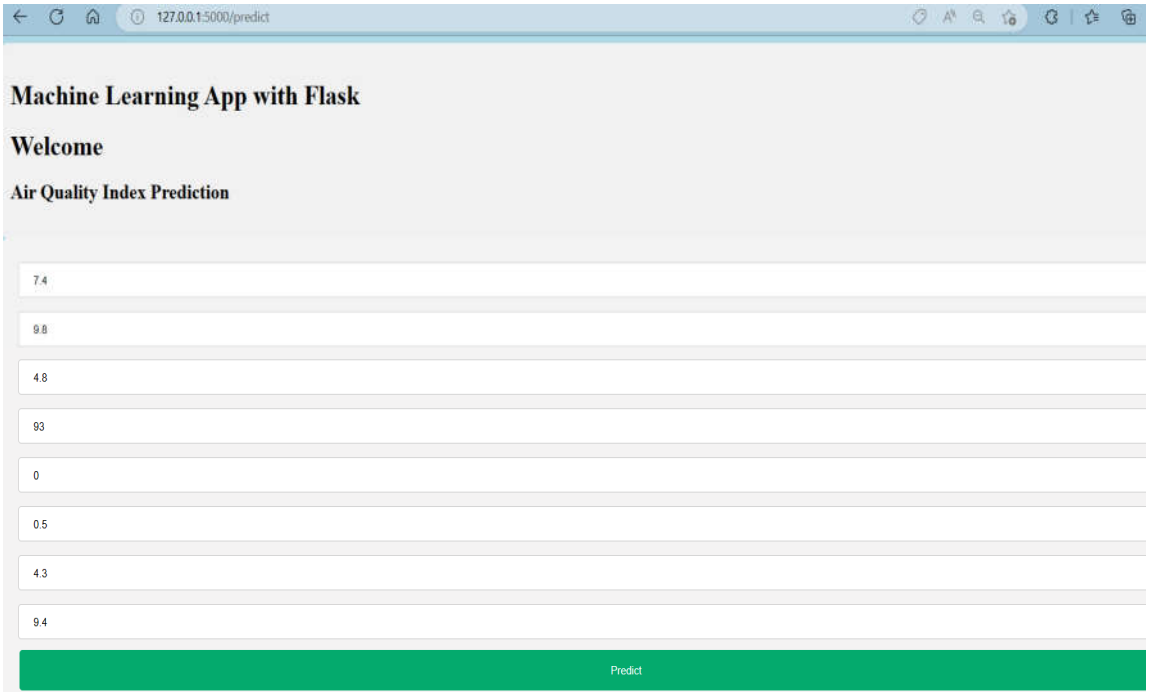

Figure 5.2: Inputing the parameters for predicting AQI value

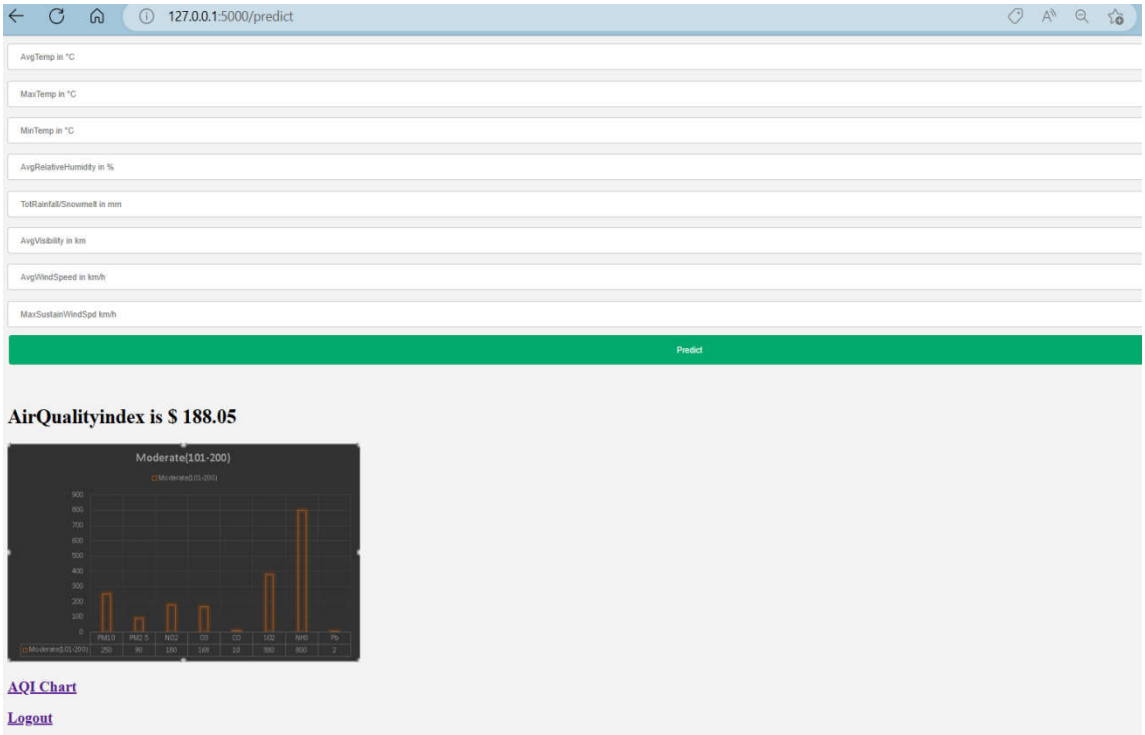

Figure 5.3: Obtaining AQI value and its Bar Graph Representation as output

Air Quality Index (AQI) Values	Levels of Health Concern	Colors
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon

Figure 5.4: Air Quality Index Chart

AQI Category (Range)	PM_{10} 24-hr	$PM_{2.5}$ 24-hr	NO_2 24-hr	O_3 8-hr	CO 8-hr (mg/m^3)	SO_2 24-hr	NH_3 24-hr	Pb 24-hr
Good (0-50)	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200	0-0.5
Satisfactory (51-100)	51-100	31-60	41-80	51-100	1.1-2.0	41-80	201-400	0.6 –1.0
Moderate (101-200)	101-250	61-90	81-180	101-168	2.1- 10	81-380	401-800	1.1-2.0
Poor (201-300)	251-350	91-120	181-280	169-208	10.1-17	381-800	801-1200	2.1-3.0
Very poor (301-400)	351-430	121-250	281-400	209-748*	17.1-34	801-1600	1201-1800	3.1-3.5
Severe (401-500)	430 +	250+	400+	748+*	34+	1600+	1800+	3.5+

Figure 5.5: Breakpoints for AQI Scale 0-500 units

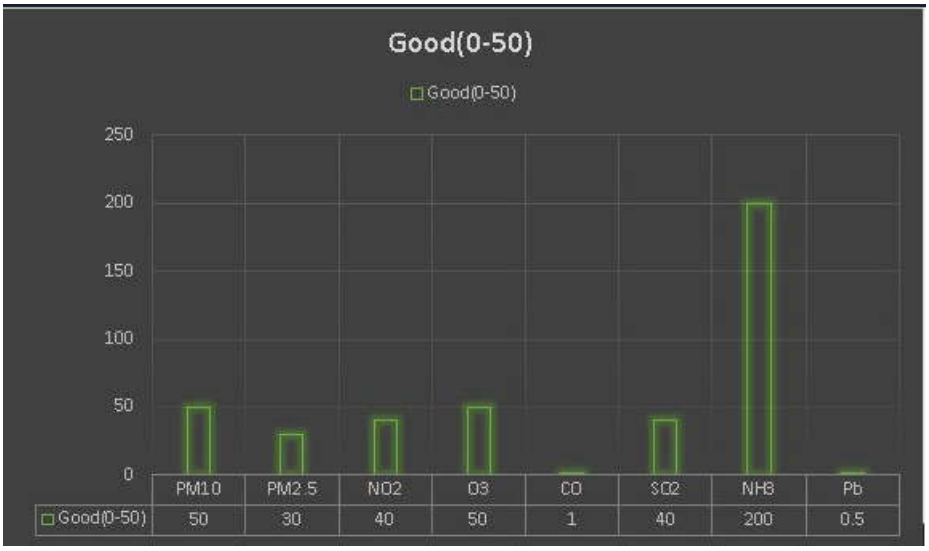

Figure 5.6: The Bar Graph Representation for Good AQI value

Figure 5.7: The Bar Graph Representation for Satisfactory AQI Value

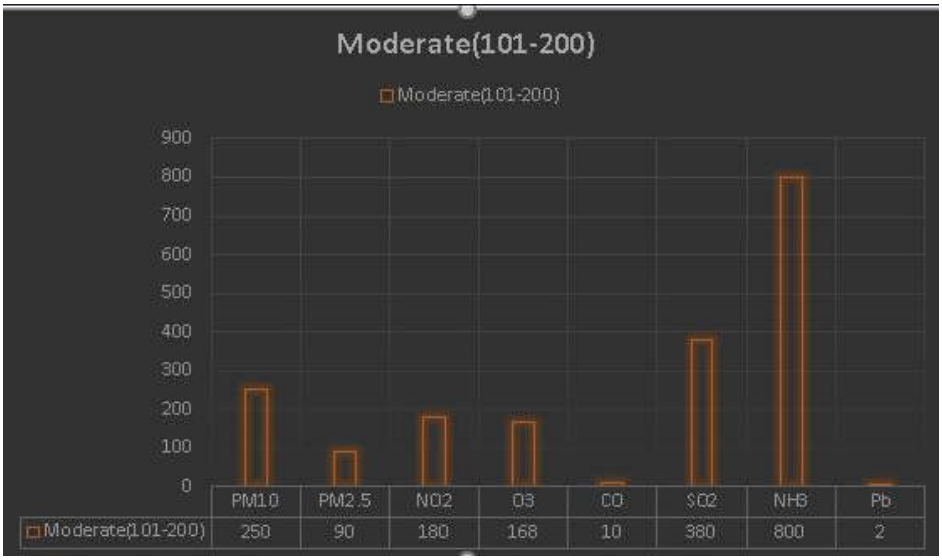

Figure 5.8: The Bar Graph Representation for Moderate AQI Value

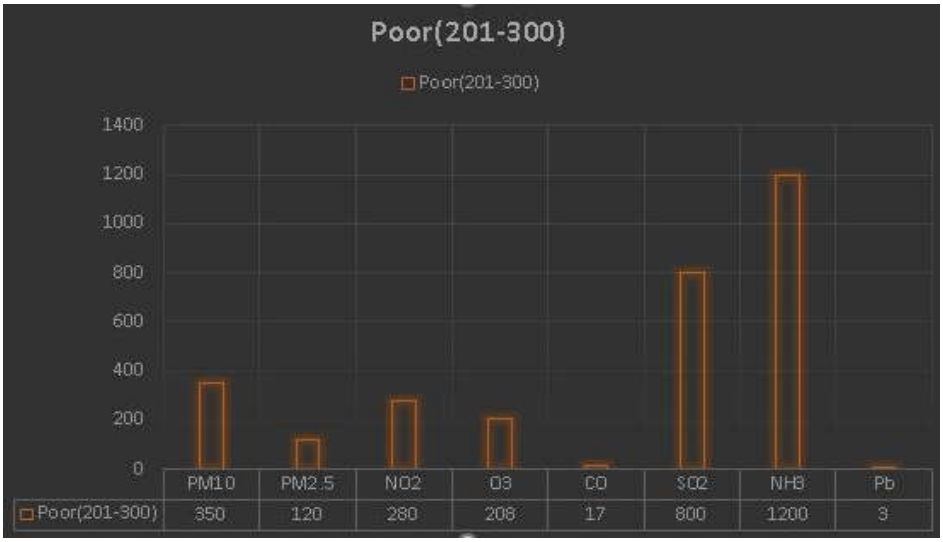

Figure 5.9: The Bar Graph Representation for Poor AQI Value

CONCLUSION

With this model we can forecast the AQI and alert the respected regions of the country also it is a progressive learning model and it is capable of tracing back to the particular locations needed atte ntion is provided by the Bar Graph representation of various polluted gases of every possible regi ons. The air quality information utilized in this paper originates from the india air quality checkin g and investigation stage, and incorporates the every normal day of fine particulate issue (PM2.5), inhalable particulate issue (PM10), ozone (O3), CO, SO_2, NO_2 fixation, Ammonia (NH_3) and Lea d (Pb). The essential perspectives that should be viewed as with regards to gauging of the poison focus are its different sources alongside the components that impact its fixation.

The analytical procedure began from information cleaning and processing, incomplete records, de tailed evaluation and in the end, model constructing and evaluation. The first rate accuracy on pub lic test dataset is having good parameter values by the way of accuracy with classification record. This application can help India meteorological division in predicting the way forward for nice air and its reputation and it will depend on that they are able to take motion.

FUTURE ENHANCEMENT

a) India meteorological department wants to automate the detecting of air quality is weather good or not from eligibility process (which is real time).

b) To automate this process by showing the predicted result in web application or in a desktop ap plication.

c) To optimize the work to implement in Artificial Intelligence environment.

REFERENCES

1) Hakob GRIGORYAN, “Prediction of Air quality Method Based on Support Vector Machine (SVM) and Independent Component Analysis (ICA)”, DSJ 2016.
2) Nallakaruppan, M.K., and Harun Surejllango. “Location Aware Climate Sensing and Real Ti me Data Analysis.” Computing and Communication Technologies (WCCCT), 2017 World C ongress on. IEEE, 2017.
3) Kumar, Anikender and P.Goyal, “Forecasting of daily air quality index in Delhi”, Science of the Total Environment 409, no. 24(2011): 5517-5523.
4) T.chiwewe and J.Ditsela “Machine learning based estimation of ozone using spatio-temporal data from air quality monitoring stations,” presented at 2016 IEEE 14th.
5) E Kalapanidas and N Avouris “Applying machine learning techniques in air quality predictio n,” in proc. ACAI, vol 99, September 2017.