iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.53209

Air Contamination Forecast using Machine Learning

Venkanna Babu S[1], Mohamed Abdul Kader Jailani[2]
[1]Student, School of Computer Science and Applications, Reva University,
[2]Assistant Professor, Reva University, Bangalore-560064, India,

Abstract: ***Air Contamination is one of the most major problem faced by today's ecosphere. Environment becoming toxic due to contaminated cities and different contaminants have also been added in the air. Air Contamination is caused by both natural activities and human. Pollutants like Chlorofluorocarbon, Nitrogen Dioxide Carbon Dioxide, Sulphur Dioxide, Carbon Monoxide, Lead, and Mercury etc. are being added in the air due to human activities. In this paper data has been composed from government website. Linear Regression, Decision Tree Regression and Random Forest Regression were used to evaluate the level of CO. least error achieved by Random Forest Regression out of other algorithm chosen for evaluation of Air Pollution and hence resulted in supplementary accurateness.***
Keywords: Decision Tree Regression, Air Contamination, Sulphur Dioxide, Carbon Dioxide, Machine Learning Algorithm.

I. INTRODUCTION

Drastic diminution in Air quality is faced in most of the cities in India in the current years. Numerous newer contaminants like Sulphur Dioxide, Nitrogen Dioxide, and Carbon Monoxide has been added into the atmosphere Other than the normal contaminant like Carbon Dioxide. Most of the contaminants have detrimental effects on our well-being. But CO is more harmful. It takes life silently and rapidly. It straightaway penetrates into blood and substitutes the oxygen particles thus depriving the intellect and heart of required oxygen to function. If it is existing in the air,it rapidly enters into the blood triggering symptoms like nausea, headache, flu, etc. As the level upsurges, individual may get sickness, catalepsy and may consequence in harm of intellect or death.

There are numerous bases of fabricating CO like imperfect driving car, boiling of solid, boiling of charcoal, coal or wood, liquid and gaseous fuels etc. Carbon Monoxide also disturbs the capability of atmosphere to cleanse itself and accountable for pollution and lower atmospheric ozone. Almost all the large cities in India are contaminated by carbonmonoxide which spreads to as far as 10 kilometers fromthe shallow of the earth in the troposphere. Contamination from Vehicles is a chief contributor to high carbon monoxide stages. Though, investigators trust winds transmit the Carbon Monoxide shaped by biomass burning in Africa and Southeast Asian republics to the Indian subcontinent, Thus accumulation to the already high levels of the gas in the country's troposphere.As carbon monoxide due to vehicular contamination and other bases is released, it travels aloft. In monsoon as there is a portion of wind, Carbon Monoxide atoms extent heights of 10 kilometers as rapidly as two hours.

That is why in monsoon Carbon Monoxide values are the uppermost in higher layers of the troposphere. In winter, it is the opposite as there is not much wind and the Carbon Monoxide particles are closer to the Earth's surface

Absorption of Carbon Monoxide can be estimated in Parts per Million. For instance, 100 PPM of Carbon Monoxide implies that for each 9999,900 particles of air, there are 100 particles of Carbon Monoxide. Another way of estimating Carbon Monoxide concentration is Time Weighted Average. It measures individual's exposureto Carbon Monoxide over time.

In this paper a system established which trains a Machine Learning prototype using contamination data congregated from government sites and static sensors. The learnt model is used to evaluate the air contamination.

II. LITERATURE SURVEY

D.J. Briggs et al. [1] have developed a methodology that maps traffic related air pollution within GIS environment. They have considered NO2 in Huddersfield, Prague and Amsterdam. Theirs results show good predictions of pollutionlevels.

V. Singh et al. [2] have proposed a system that estimates and interpolates daily ozone concentrations. This approach is based on a technique called cokriging.

In [3], M Mead et al. have deployed the sensor nodes in static network in the Cambridge (UK) area and mobile network. They have provided the results for quantification of personal exposure.

P. Dutta et al. [4] have proposed system where individual can measure their personal exposure using participatory sensors then groups to summarize their members' exposure.

K Hu et al. [5] have collected urban air pollution data withhigh spatial density by using many software applications and hardware devices. They have devised a web based tool and mobile app for the estimation and visualization of air pollution. Their system shows accurate exposure than thecurrent systems.

V. Sivaram et al. [6] developed a project that uses many mobile sensors attached to vehicle to measure the air pollution concentration. The collected data is uploaded to user's mobile. Afterwards pollution maps are created that show the exposurehistory and accordingly the route can be planned to reduce thefuture exposure.

K B Shaban et al. [7] developed a system that uses motes equipped with gaseous and meteorological sensors. Thesecommunicates to an intelligent sensing platform that comprises of various modules. Mainly four modules have been used for receiving data, preprocessing and converting thedata into meaningful information, predicting the pollutants and presenting the information through short message services, web portal and mobile app. They used three ML algorithm namely Support Vector Machine, MSP Model treesand Artificial Neural Networks.

D Hasenfratz et al. [8] have collected the measurements for more than a year through mobile sensor nodes. These nodes were installed on top of public transport vehicles in Zurich (Switzerland). From this obtained data, they developedregression models that create pollution maps with high resolution of 100m

Ke Hu et al. [11] have introduced a machine learning model that takes fixed station data and mobile sensor data andthen estimate the air pollution for any hour on any given day in Sydney city. They have used seven regression models and ten-fold cross validation.

Arnab Kumar Saha et al. [12] have used have used cloud based Air Pollution Monitoring Raspberry Pi controlled System. They measured Air Quality Index based on five criteria pollutants, such as particulate matter, ground level ozone, Sulphur Dioxide, Carbon Monoxide and Nitrogen Dioxide using Gas Detection Sensor or MQ135 Air Quality.

Kavitha B C et al. [13] have deployed various IoT sensors on the industrial floor to collect the data and implemented a pollution monitoring system

A Orun et al. [14] have used artificial intelligence technique such as Bayesian Networks to establish relation between traffic and traffic related air pollutants.

Nitin Sadashiv Desai and John Sahaya Rani Alex [15]have measured CO and CO2 level in the air with GPS by using pollution detection sensor and uploaded into Azure Cloud Services.

E Suganya and S Vijayashaarathi [16] have proposed a system that uses Mobile Ad Hoc Network routing algorithm and monitors the travelling vehicles by using number of sensors. The collected data is stored in cloud network toaccess the information about levels of pollution.

III. SYSTEM SCHEME AND OPERATION

The system follows a typical methodology as
Mentioned Below:

1) Feeding mobile sensed data and prerecording Data.
2) Preprocessing data
3) Giving training through machine learning model
4) Building final model.
5) Evaluating result.
6) Adopting machine learning model which gives Best accuracy.

A. Gathering Data

Data is composed from source:
Open Source Government Website(www.data.gov.in) [9].

B. Statistics Sampling

This part is comprised of two separate steps:

1) *Data Preprocessing:* This step involves cleaning rawdata by removing null values, removing unwanted Values, converting one feature into another,extracting other features from one feature etc.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

2) *Target Preparation:* This step involves convertingdata to a form that is easy to handle for the Machine Learning models. This is achieved by performingone of the many methods of target preparation such as Standardization (converts data to have mean 0 and same standard deviation), Min-MaxNormalization (converts data so they occur between any fixed interval) etc.

C. Data Training

This paper uses three Machine Learning models to predictAir pollution rates namely LR, DTR and RFR.

D. Valuation

K-fold cross authentication is a procedure to estimate the parameters. In this, the unique facts sample is arbitrarily divided into k equal sized clusters. Out of the k clusters, a first group is earmarked as the authentication data for testing the model and the remaining k-1 groups are used as training data sample. Every sample data is used in hold out set one time and then continual k-1 times to train the model. Particular assessment outcome can be formed by averaging the k outcomes from the folds.
The advantage of this method over repeated random sub sampling is that all data samples are used for both validation and training, and every data sample is used accurately once for validation. In this system, k=10 i.e. 10-fold cross-validation is used.

IV. OUTCOMES AND CONVERSATION

The raw data obtained from both governmentwebsites are pre-processed into structured dataset. The below figures show the raw sample data obtained from sensors and government websites.


```
- <State id="Karnataka">
  - <City id="Bengaluru">
    - <Station id="BWSSB Kadabesanahalli, Bengaluru - CPCB" lastupdate="13-04-2018 07:00:00">
        <Pollutant_Index id="PM2.5" Min="NA" Max="NA" Avg="NA"/>
        <Pollutant_Index id="NO2" Min="NA" Max="NA" Avg="NA"/>
        <Pollutant_Index id="SO2" Min="NA" Max="NA" Avg="NA"/>
        <Pollutant_Index id="CO" Min="36" Max="52" Avg="44"/>
        <Pollutant_Index id="OZONE" Min="12" Max="17" Avg="14"/>
      </Station>
    - <Station id="City Railway Station, Bengaluru - KSPCB" lastupdate="13-04-2018 07:00:00">
        <Pollutant_Index id="PM10" Min="22" Max="142" Avg="108"/>
        <Pollutant_Index id="NO2" Min="24" Max="141" Avg="85"/>
        <Pollutant_Index id="SO2" Min="NA" Max="NA" Avg="NA"/>
        <Pollutant_Index id="CO" Min="26" Max="108" Avg="80"/>
      </Station>
    - <Station id="Peenya, Bengaluru - CPCB" lastupdate="13-04-2018 07:00:00">
        <Pollutant_Index id="PM2.5" Min="47" Max="98" Avg="51"/>
        <Pollutant_Index id="NO2" Min="1" Max="38" Avg="10"/>
        <Pollutant_Index id="NH3" Min="20" Max="22" Avg="22"/>
        <Pollutant_Index id="SO2" Min="5" Max="13" Avg="9"/>
        <Pollutant_Index id="CO" Min="28" Max="44" Avg="36"/>
      </Station>
    - <Station id="Sanegurava Halli, Bengaluru - KSPCB" lastupdate="13-04-2018 07:00:00">
        <Pollutant_Index id="PM10" Min="20" Max="89" Avg="59"/>
        <Pollutant_Index id="NO2" Min="14" Max="76" Avg="32"/>
        <Pollutant_Index id="SO2" Min="4" Max="6" Avg="4"/>
        <Pollutant_Index id="CO" Min="26" Max="64" Avg="40"/>
      </Station>
    </City>
  </State>
```


Fig. 1. Raw sample data obtained from government websites

The below figures show the pre-processed dataset obtainedfrom both government websites.

1	Date	Day	Time	CO
2	17-04-2018	Tuesday	09:00	14.00
3	17-04-2018	Tuesday	10:00	14.00
4	17-04-2018	Tuesday	11:00	14.00
5	17-04-2018	Tuesday	12:00	14.00
6	17-04-2018	Tuesday	13:00	14.09
7	17-04-2018	Tuesday	14:00	14.00
8	17-04-2018	Tuesday	15:00	14.11
9	17-04-2018	Tuesday	16:00	14.00

Fig. 2. Pre-processed dataset sample obtained from sensor

1	State	City	CO	Last Update Date	Last Update Time	Day
2	Karnataka	City Railway Station	55	13-04-2018	09:00:00	Friday
3	Karnataka	Peenya	35	13-04-2018	09:00:00	Friday
4	Karnataka	Sanegurava Halli	39	13-04-2018	09:00:00	Friday
5	Karnataka	City Railway Station	55	13-04-2018	10:00:00	Friday
6	Karnataka	Peenya	35	13-04-2018	10:00:00	Friday
7	Karnataka	Sanegurava Halli	41	13-04-2018	10:00:00	Friday
8	Karnataka	City Railway Station	56	13-04-2018	11:00:00	Friday
9	Karnataka	Peenya	36	13-04-2018	11:00:00	Friday
10	Karnataka	Sanegurava Halli	41	13-04-2018	11:00:00	Friday
11	Karnataka	City Railway Station	58	13-04-2018	12:00:00	Friday

Fig. 3. Pre-processed dataset sample obtained from government websites

The below figures present contrast between CO andsovereign variables like Time, City, Day.

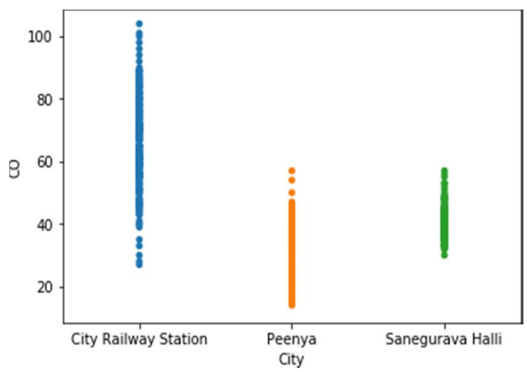


Fig. 4. City vs CO levels for data gained from Government website.

Table I.Root Mean Square Fault Attained InEvery Model

Machine Learning Model	*Fault*	*Accurateness*
Decision Tree Regression	11.6576	88.3424
Linear Regression	13.8275	86.1725
Random Forest Regression	9.7527	90.2473

V. ASSESING OUTCOME

Arnab Kumar Saha et al. [12] have used have used cloud based Air Pollution Monitoring Raspberry Pi controlled System. They measured Air Quality Index based on five criteria pollutants, such as particulate matter, ground level ozone, Sulphur Dioxide, Carbon Monoxide and Nitrogen Dioxide using Gas Detection Sensor or MQ135 Air Quality.

But their results don't show Carbon Monoxide level and also accuracy has not been computed. In our work we used MQ7 sensor and government site for measuring the CO level and computed the accuracy using three machine learning algorithms giving the maximum accuracy as 90.2%.

E Suganya and S Vijayashaarathi [16] have proposed a system that monitors the level of NO2, Humidity, Temperature, Carbon Monoxide by using NO2 sensor, Humidity Sensor, Temperature Sensor and Carbon Monoxide Sensor respectively but in the results Carbon Monoxide level and accuracy is not shown. In our results we have shown both Carbon Monoxide level and accuracy.

Nitin Sadashiv Desai and John Sahaya Rani Alex [15] have not mentioned any algorithm on which the pollution metrics were predicted. Only Azure Machine Learning Service is used. Accuracy component is missing. In our system we have used different Machine Learning algorithms to compute the accuracy.

Kavitha B C et al. [13] have used the sensors to sense the level of various gases in the industrial floor. They just monitored the amount of pollution but accuracy is missing. Our system shows both accuracy and CO level.

A Orun et al. [14] have established relation between traffic and traffic related air pollutants namely, SO_2, NO_2 and CO. They have developed a Bayesian predictive model using Bayesian classifier with classification accuracies 85%, 78% and 81% respectively for the above-mentioned pollutants. Our work considers only one main pollutant namely CO and assesses the CO level using three Machine Learning Algorithms and it gives the accuracy as 90.2%.

Table II. Contrast of performance of projected methodology with numerous approaches

Novelists	Contaminants	Dataset Or Sensor Used	Technology Or Algorithm Used	Accurateness
Suganya et al.	NO2, Humidity, Temperature, CO	Humidity Sensor, NO2 Sensor, Temperature Sensor, CO Sensor	MANET	-
Arnab Kumar Saha et al.	Ground Level Ozone, Particulate Matter, CO, SO_2 and NO_2	LM393, MQ135, DHT11	Raspberry – pi/Calculates Air Quality Index	-
A Orun et al.	SO_2, NO_2 and CO, Temp, Air Pressure etc.	Dataset	Bayesian Network	SO_2- 85% NO_2-78% CO-81%
Nitin Sadashiv Desai et al.	CO and CO_2	MQ7, MQ11	Beagle Bone Black	-
Kavitha B C et al.	CO, LPG, Methane, Butane	MQ135/6/7, DHT11	Raspberry- pi/IoT Shield	-
Our System	CO	MQ7 and dataset	Three ML Algortihms- LR, DTR and RFR	CO- 90.2%

VI. CONCLUSION

In this paper, machine learning based system is used to estimate dense air pollution using historical data from government monitoring sites. Decision Tree Regression, Random Forest Regression and Linear Regression were used, and performances estimation were compared. We Random Forest Regression selected as the machine learning algorithm because of its low error index.

The projected model performs better for outsized training set. But it proceeds with more time to train model. Use of mobile sensors can help to shape system to generate air contamination map. Assessment accuratenesscan be augmented in future by presenting climatological influences such as wind rapidity and climate in the system further more testing[11] and ai[13] can be used to improve the system.

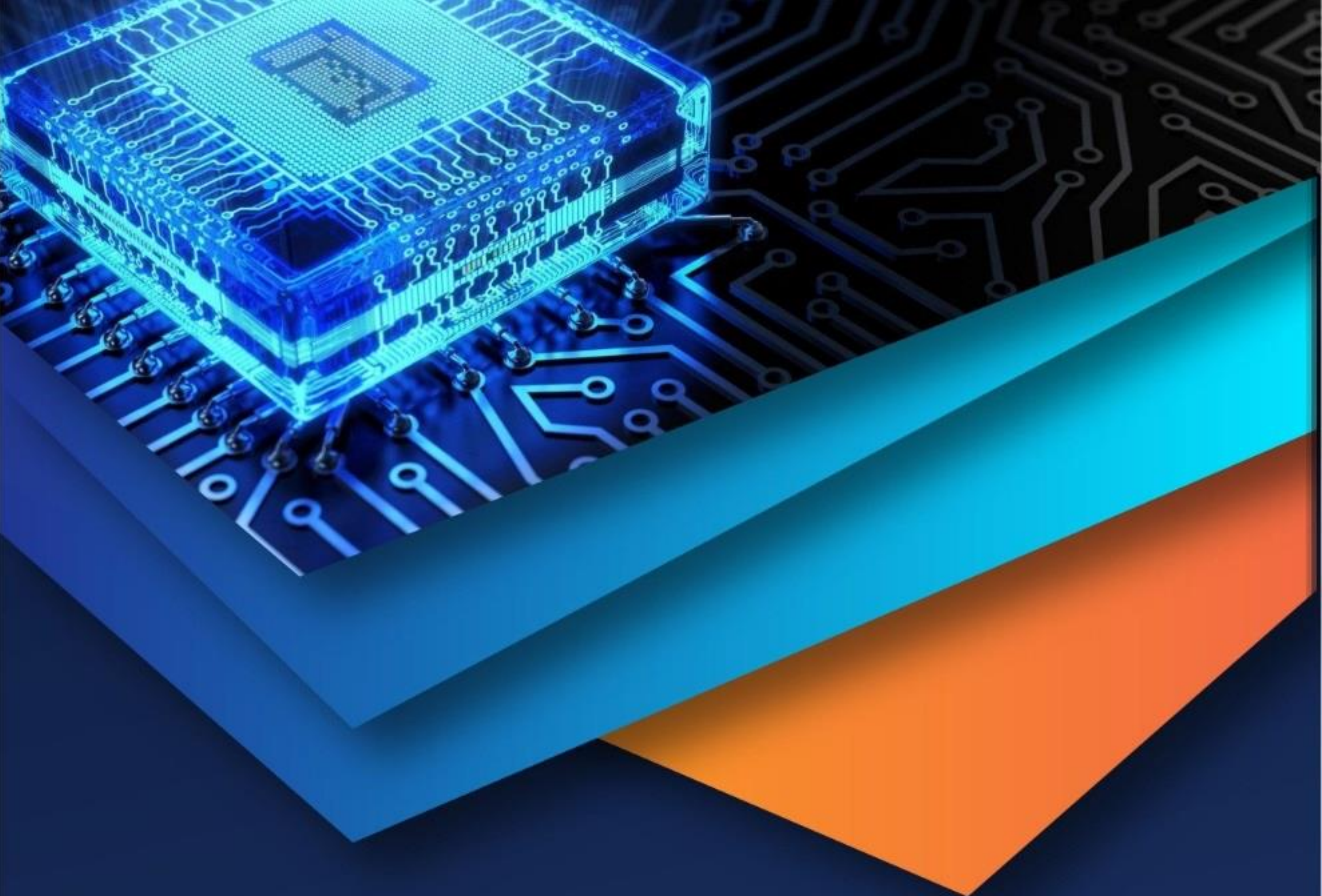
REFERENCES

[1] D. J. Briggs et al, "Mapping urban air pollution using GIS: a regression-based approach," International Journal of Geographical Information Science, vol. 11, no. 7, pp. 699–718, 1997.
[2] V. Singh et al., "A cokriging based approach to reconstruct air pollution maps, processing measurement station concentrations and deterministic model simulations," Environmental Modelling and Software, vol. 26, no. 6, pp. 778 – 786, 2011.
[3] M. Mead et al., "The Use of Electrochemical Sensors for Monitoring Urban Air Quality in Low-Cost, High-Density Networks," Atmospheric Environment, vol. 70, pp. 186–203, May 2013.
[4] P. Dutta et al., "Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors," in Proc. SenSys Demonstration, Berkeley, CA, USA, Nov 2009.
[5] K. Hu et al., "Design and evaluation of a metropolitan air pollution sensing system," IEEE Sensors Journal, vol. 16, no. 5, pp. 1448–1459, March 2016.
[6] V. Sivaraman et al., "Hazewatch: A participatory sensor system for monitoring air pollution in sydney," in Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 38th Conference on, Oct 2013, pp. 56–64.
[7] K. B. Shaban et al., "Urban air pollution monitoring system with forecasting models," IEEE Sensors Journal, vol. 16, no. 8, pp. 2598– 2606, April 2016.
[8] D. Hasenfratz et al., "Pushing the spatio-temporal resolution limit of urban air pollution maps," in Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on, March 2014, pp. 69–77.
[9] Open Source Government website: https://data.gov.in/
[10] CO information website: https://www.detectcarbonmonoxide.com/co-health-risks/
[11] Udupa, Pradeep. "An Integrated Methodology for Testing Source Code by Using Multi Constraint Reduction, Test Suite Prioritization and Prioritized Parallelization", IJET, vol. 10, no 2, pp. 221–225.
[12] Ke Hu et al., "HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation from Fixed and Mobile Sensors" DOI 10.1109/JSEN.2017.2690975, IEEE Sensors Journal
[13] Udupa, Pradeep. "Application of artificial intelligence for university information system." Engineering Applications of Artificial Intelligence 114 (2022): 105038.
[14] Arnab Kumar Saha et al., "A Raspberry Pi Controlled Cloud Based Air and Sound Pollution Monitoring System with Temperature and Humidity Sensing", 2018 IEEE
[15] Kavitha B C et al., "IOT Based Pollution Monitoring System Using Raspberry-Pi", 2018, Vol 118 International Journal of Pure and Applied Mathematics
[16] A. Orun et al., "Use of Bayesian Inference Method to Model Vehicular Air Poillution in Loacal Urban Areas", Transportation Research Pard D63 (2018) , pp. 236-243, Elsevier
[17] Nitin Sadashiv Desai and John Sahaya Rani Alex, "IoT Based Air Pollution Monitoring and Predictor System on Beagle Bone Black", 2017 International Conference on Nextgen Electronic Technologies, IEEE
[18] E Suganya and S Vijayashaarathi et al., "Smart Vehicle Monitoring System for Air Pollution Detection using WSN", 2016 International Conference on Communication and Signal Processing, IEEE

AUTHORS PROFILE

Venkanna Babu S is pursuing Master of Computer Applications, School of CSA, REVA University, Bengaluru, India. His area of interest includes machine learning . E-mail: venkannababusunkavalli@gmail.com

Mohamed Abdul Kader Jailani N M.Tech, MCA is working as Assistant Professor in the school of CSA, REVA UNIVERSITY, Bengaluru, India. He passed both TN-SET and NET examinations. He is interested in doing research in image processing, machine and deep learning.
Email: jailani.msa@gmail.com



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)