Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 165 (2019) 322–327

Procedia Computer Science

www.elsevier.com/locate/procedia

INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019

Low Cost IoT Based Air Quality Monitoring Setup Using Arduino and MQ Series Sensors With Dataset Analysis

KinneraBharath Kumar Sai*, Subhaditya Mukherjee, Parveen Sultana H

Vellore Institute of Technology, Vellore, 632014, India

Abstract

ThispapertalksaboutcosteffectiveArduinobasedAirQualityMonitoringsetupusingMQseriessensorswhicharequitesuitable to install in both indoor and outdoor provided that properly calibrated before installing. There are many MQ series sensors out of which MQ135 and MQ7 have been considered here as MQ135 is able to detect ammonia, carbon dioxide, alcohol or even smokeandMQ7helpstocalculateCarbonMonoxidealoneandthesetwosensorsarequitesuitablefortheapplicationconsidered.

Government of India has taken enough measures already to minimize the air being polluted. The whole setup can be made as acompact device with low cost and can be used as a carry-in device such that awareness is brought among the people of how's the airqualitylevelofthe areasurroundedbythepersoneitherindoororoutdoor.Adverseeffectsofairpollutionleadstorespiratory problems, skin diseases etc., Moreover, the data collected by these sensors will be pushed to the cloud on back end, say here ThingspeakischosenandthearemanyopensourceIoTsupportingplatforms.Attheend,dataanalysiswasdoneonthedataset collected from the setup which is installed at various places across the VIT University, Vellore. This analysis helps in deeper understanding of the air quality status such that people will be aware of what will happen if the same air quality continues for a longtime.

Keywords: IoT; MQ135 Sensor; MQ7 Sensor; Thingspeak; Machine Learning; PPM (parts per million);

1. Introduction

Air pollution is not new to everyone. Just like any other pollutions, air pollution has serious effects on human health. It is quite important to be aware of this pollution level in our surroundings. With serious demand in vehicles andtransportthesedays,knowinglyorunknowinglythereisagradualincreaseinthelevelofpollutionwhichresults in respiratory and skin diseases. Not only the case of vehicles, but also with the deforestation, the air quality index is worsening day to day. Just like, the temperature of a specified location can be known, the air quality index can also beknownwiththehelpofenablerInternetofThings[4].Affordablesensorshavebeenconsideredhereandcalibrated

* *Corresponding author. Tel.:+91-8977777267*
E-mail address: yourfriend9014@gmail.com

10.1016/j.procs.2020.01.043

properly to bring the accurate values. As IoT helps to project the values to the cloud and the values can be read anywherefromonline.Insteadofsimplyfetchingthevaluesfromthesensorsplacedatalocationandcalibratedwith respect to the location installed, Data analysis is made which gives a clear idea of the status and conditions of the pollution data fetched so far. The reason to go for data analysis is to know that if the same pollution continues for a long time without any proper measures taken[5]. The units for the Air Pollution are taken PPM (parts per million) here. The raw data fetched from the sensors is properly converted to PPM in the Arduino code and necessary help is taken from sensor datasheets. The proposed work makes use of affordable development kit boards which is reliable andcost-effective[9].Propercalibrationsofthesensorsusedandestablishingpropermathematicalbackgroundwhile convertingtherawdatatothePPMunitsismaintained.Attheend,resultswillbepushedtothemobileapplicationor to a computer about the air quality index[3].

2. RelatedWork

Before understanding the setup, it is important to know the levels of air quality which affects human beings.[4]. The table 1 mentioned below give as a clear insight that level 0-50PPM and 51-100PPM are good to humans. Anyof above these ranges are totally harmful topeople.

Table 1. Air Quality Index

Range(PPM)	Status
0-50	Good
51-100	Moderate
100-150	Unhealthy for sensitivegroups
151-200	Unhealthy
201-300	VeryUnhealthy
301-500	Hazardous

The paper cited at [1] has used LCD to show the values of Air quality being measured. If the paper is based on IOT, our usual intension is to show the sensors values converting to PPM(parts per million) on cloud or web but they showed it on LCD module which incurs additional cost. The paper cited at [2] has not properly calibrated the sensors because the readings shown are 300 PPM. As per the table 1 mentioned, 300 PPM is deadly dangerous. The paper cited at [3], also projected wrong PPM values and its clear indication that they haven't calibrated the sensor as mentioned in the Procedure section. The paper cited at [4] has used two sensors where the heater element inside the sensors will draw more current which arduino can't deliver enough and hence it needs an additional battery and they didn't use it so obviously the output is unpredicted and the values they claimed were notcorrect[8].

3. Procedure

TheimportantstepistoconverttheSensorrawdataoutputintoPPM(Partspermillion).Iftheseequationsincluded in arduino code, then final output will be in PPMonly[6].

3.1. Equations

Beforefindingthe$\omega(R_0)$value,thesensorreadingsgives$\beta(R_s)$infreshairanditistonotethatthescaleconsidered in figure 1 is a logarithmic scale[8].

The following derivations can be calculated from Figure 1

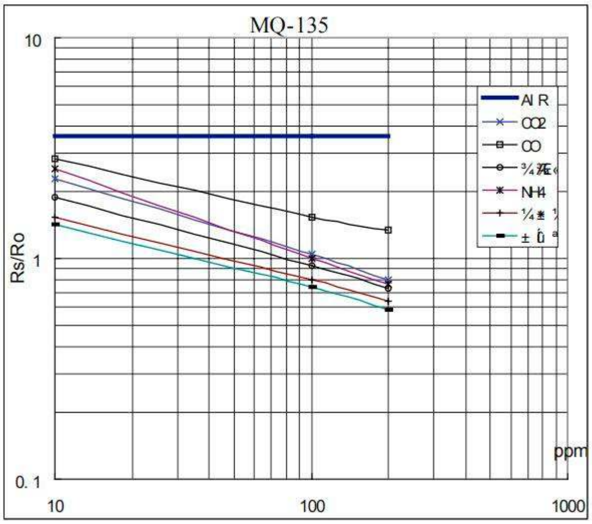
$$\mu = \lambda\chi+\tau \tag{1}$$

The above equation 1 should be considered as below equation 2 as the figure 1 is on logarithmic scale

$$log_{10}\mu=\lambda*log_{10}\chi+\tau \tag{2}$$

Fig. 1. R_s/R_0 Vs PPM

Inordertocalculatetheλ,anytwopointscanbechosenfromthegraphsho



wninfigure1.Here,CO_2line(10,2.2) and (200,0.8) are considered. This graph is taken from MQ135 datasheet and it is generated by subjecting the sensor exclusive to various PPM levels of the gases. Now, λ can be written as[8]

$$\lambda = \frac{log(y2) - log(y1)}{log(x2) - log(x1)} \tag{3}$$

To obtain the equation 4, apply logarithmic quotient rule to the equation 3,

$$\lambda = \log\left(\frac{y2}{y1}\right) \tag{4}$$

Now passing the values (10,2.2) and (200,0.8) to equation 4

$$\lambda \Rightarrow \frac{log(0.8/2.2)}{log(200/10)} \tag{5}$$

$$\lambda = -0.3376 \tag{6}$$

After finding the slope λ, y-intercept τ is calculated by considering the above data

$$log(\mu) = \lambda * log(\chi) + \tau \tag{7}$$

$$\tau \Rightarrow log(0.8) - (-0.3376) * log(200) \tag{8}$$

$$\tau \Rightarrow 0.6799 \tag{9}$$

After finding λ and χ, gas concentration can be found for any ratio with the following formula:

$$log(\chi) = \frac{log(\mu) - \tau}{\lambda} \tag{10}$$

χ can be obtained from above equation 10 by applying inverse logarithmic function, hence

$$\chi = 10^{\frac{log\mu - \tau}{\lambda}} \tag{11}$$

Usingequation11,ourfinalintentionistoconverttherawdatagivenbythesensorintoPPMandthisequationwill be given in arduinounocode.

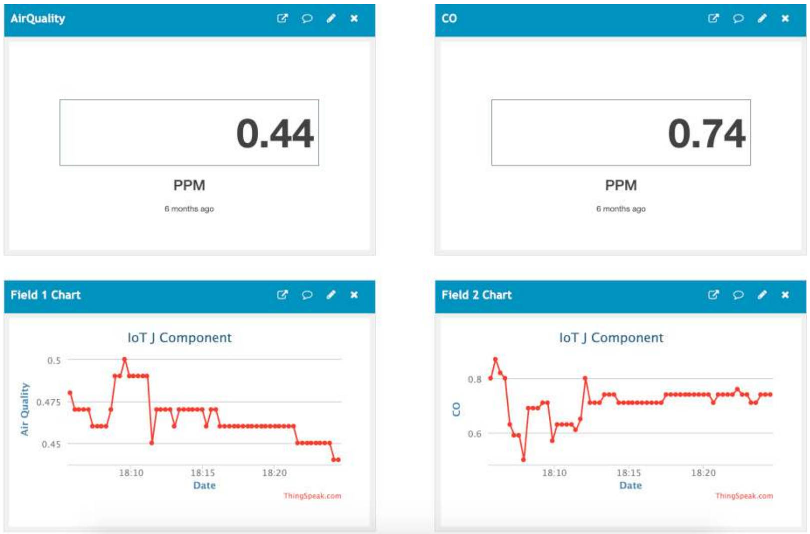


Fig. 2. Output on Thingspeak

4. Results

ArduinoUnodoesn'thaveWi-ficapability.So,ESP-01isusedtoenablewififortheArduinoUno.Thiswillpush the data output from Arduino Uno to Thingspeak and it will be auto visualized with built-in tools which is shown in figure2.Airqualitylevelgraphgoeshighindaytimeduetovehiclemovementsandaveragevalue0.15PPMshownin figure2isabsolutelysafe[4].Whileplottingtheresults,DHT11Temperaturecumhumiditysensorisalsoconsidered.

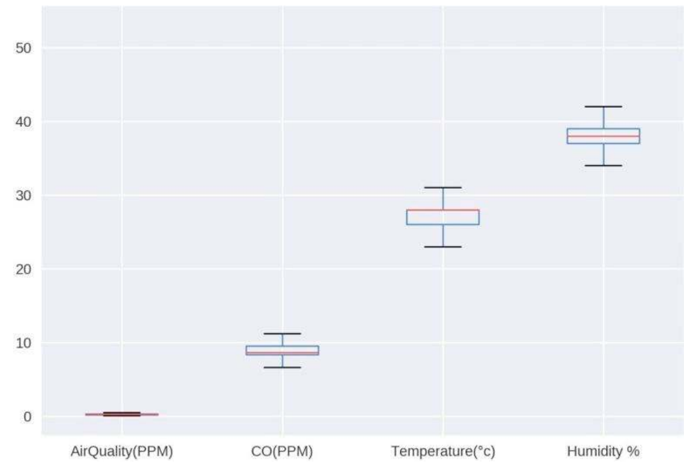


Fig. 3. Box Plot

Themedianofvaluespresentinthedatasetcanbeidentifiedfromtheplotshowninfigure3[10].Thisisthenused for further observation. It is observed that the values of air quality is between 0 to 10 PPM at the region where the experiment is carried out which found to be safe from Table 1. The values of CO are around 9PPM which is a little highbutpeoplewillnotsufferfromanyadverseeffects.Thetemperaturealsohasvaluebetween20-30andhumidity percentage is around40%.

From Histogram plot shown in figure 4, maximum and minimum values over a period for every data point is obtained. Most of the Air Quality values are between 0 and 1, which indicate not very high values of pollution. For

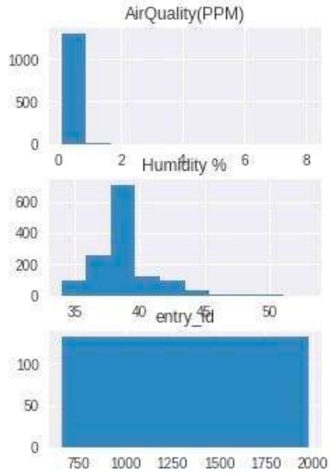


Fig. 4. Histogram of the dataset

thevaluesofCarbonMonoxide,mostofthevaluesarebetween7.5whichalsoindicatesmoderatevaluesofCO.The entry id Histogram can be ignored as it is of no significance. The temperature values are also mostly between 28 and 30 degrees which indicated room temperature. As for humidity %, most of the values seem to be between 35 and 40 whichseemstobelittlehigh.Fromtheplotshowninfig5,thedatacanbeeasilyviewedforeverydatapointover

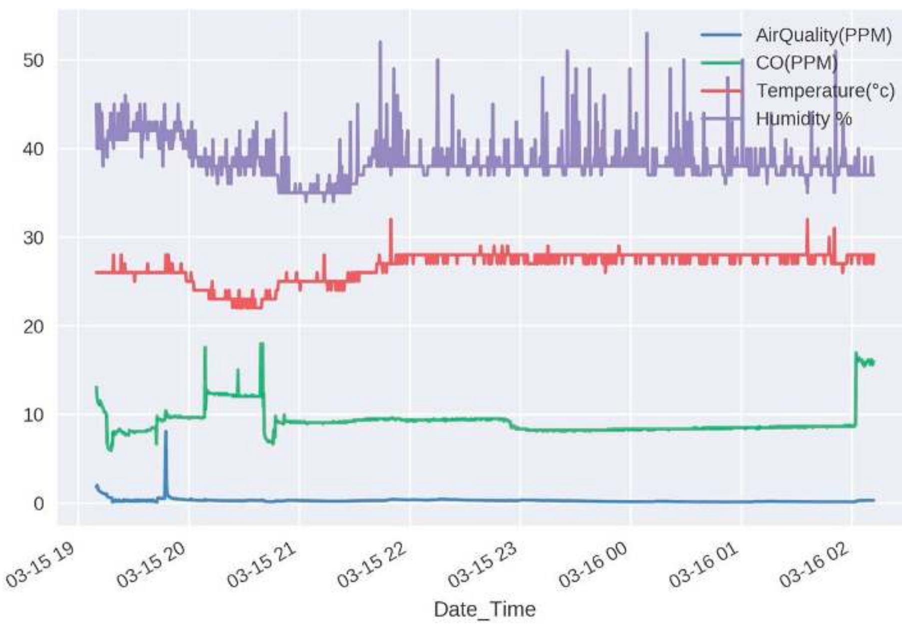


Fig. 5. Line Plot of the data

a period of time. This is done after outliers are removed so it becomes easy to draw conclusions from this data. It is foundthatAirQualityvalueswerehigheronsomedaysbutdroppeddownandvariedalotinitially[5].Thiscoincided withtherootmeansquareerrorforeverydatapointisasfollowsthedippingofTemperaturevalueswhichisamajor point to note. The CO values did not change much but considerably varied with variation of Temperature which is also a point to be noted. As for the Humidity values, they do not seem to have a lot of relation with the rest of the variables.

- rmse value for AirQuality(PPM) is0.08274985657405932
- rmse value for CO(PPM) is2.8055052798658817
- rmse value for Temperature(c) is1.4761492661822175
- rmse value for Humidity is1.706479462754771

This indicates a relatively good performance of the model, especially for the Air Quality Values. For this analysis, Jupyter Notebook hosted on Google Co-Laboratory is used. A GPU was used to train the model and reduce computation time. The data was collected by combination of four sensors kept in and around the University for a few days. Continuous data was collected before saving it to a storage device over the cloud. For any dataset, pre-processing is the most important step[6]. The first step was to parse the timestamp generated from the device and make it into a formatusablebythemodel.Thiswasdoneusingthefollowingcode.Thesecondstepwastoremoveoutliersfromthe dataandfillmissingvalues.Thiswasdonebyreplacingthemissingvaluewithavaluefromthesamecolumnwhich wouldensurethatthevaluesdidnotgetskewedcompletely.Thethirdwastoremoveoutliers.Fromdatavisualization, it was seen that there was a recurring value -999 which could be the sensor output when the sensor was switched offand these redundant values wereremoved[7].

5. Conclusion

So, using low cost arduino setup, accurate PPM values can be projected using the equations shown in 3.1 section. While training the model the 'Vector Auto Regression model' was found to be the best choice to train this model. For VAR(1), each variable is a linear function of lag1 values and so on. Such a model in general implies that every variable depends on every other variable and thus the VAR model in the end can be written as a series of individual models.TheVARmodelcanalsobeestimatedbyestimatingeachequationseparately.Manymodelsweretestedand evenanLongShortTermMemoryModel(LSTM)wasconsideredanditwasfoundthattheVARgavethebestresults in the form of root mean squareerror.

References

[1] Liu, S., Xia, C., & Zhao, Z. (2016, October). A low-power real-time air quality monitoring system using lpwan based on lora. In 2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT) (pp. 379-381).IEEE

[2] Desai, N. S., & Alex, J. S. R. (2017, March). IoT based air pollution monitoring and predictor system on Beagle bone black. In 2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2) (pp. 367-370).IEEE.

[3] Balestrini, M., Diez, T., Marshall, P., Gluhak, A., & Rogers, Y.(2015). IoT community technologies: leaving users to their own devices or orchestration of engagement?. EAI Endorsed Transactions on Internet of Things,1(1).

[4] Doukas, C., &Maglogiannis, I. (2012, July). Bringing IoT and cloud computing towards pervasive healthcare. In 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (pp. 922-926).IEEE.

[5] Girish, S. V., Prakash, R., & Ganesh, A. B. (2016). Real-time remote monitoring of indoor air quality using internet of things (IoT) and GSM connectivity.InArtificialintelligenceandevolutionarycomputationsinengineeringsystems(pp.527-533).Springer,NewDelhi.

[6] Mehta, Y., Pai, M. M., Mallissery, S., & Singh, S. (2016, March). Cloud enabled air quality detection, analysis and prediction-a smart city applicationforsmarthealth.In20163rdMECInternationalConferenceonBigDataandSmartCity(ICBDSC)(pp.1-7).IEEE.

[7] Saikumar, C. V., Reji, M., &Kishoreraja, P. C. (2017). IOT Based Air Quality Monitoring System. International Journal of Pure and Applied Mathematics, 117(9),53-57.

[8] Systems,J.(2016,May09).UnderstandingaGasSensor.Retrievedfromhttps://jayconsystems.com/blog/understanding-a-gas-sensor

[9] Kang,J.,&Hwang,K.I.(2016).Acomprehensivereal-timeindoorair-qualitylevelindicator.Sustainability,8(9),881.

[10] Kumar, P., Martani, C., Morawska, L., Norford, L., Choudhary, R., Bell, M., & Leach, M. (2016). Indoor air quality and energy management through real-time sensing in commercial buildings. Energy and Buildings, 111,145-153.