Air Pollution Monitoring and Prediction System for Sustainable Metropolitan Cities using Iot

Nagaraja G S, Shreyas Srinath, Reshma Jakabal

Abstract: In today's world, the temperature of the environment is gradually rising. One of the main reasons for it is decreasing the quality of air mainly caused due to air pollution. There are many harmful substances present in the environment that is the cause of the declining quality of air. These pollutants get mixed along with the air and pollute the environment. The two air pollutants are considered here, CO2 and NO, to reduce air pollution, there is a need to know the number of pollutants, with the help of sensors in this experiment the level of pollutants are to be monitored and based on that a prediction mechanism is developed to determine the level of pollutants in the future. There are some machine learning concepts involved, K means clustering for classification of pollutants along with the S.V.M. (Support Vector Machine). With the successful prediction of the level of pollutants, the necessary counter measures can be adaopted.

Keywords: Air Monitoring, IoT, Machine Learning, Predictive Analysis, Support Vector Machine.

I. INTRODUCTION

Today every person wants to breathe in an atmosphere that contains fresh and pollution-free air. But it has become a major problem as people do not have a sufficient quantity of air to breathe because of the rising amount of pollution in the atmosphere. As a result, several countries in the world are suffering from various problems related to air pollution, both in developed and developing countries. Due to this, there are many health problems and environment affecting factors arising because of a large number of industries that release the pollutants along with the pollution caused by the vehicles, on a daily basis. These gaseous pollutants lead to serious health problems and harm to the environment. Air pollution is one of the major reasons for the contribution of global warming. The temperature of the planet, on average, is on the rise because of the increase in the toxicity of the air. One of the major causes of air pollution is the toxic gases that are released due to vehicular emission and mostly the factories that discharge harmful pollutants in the air without undergoing proper filtration checks. There are many

Nagaraja G.S*, Professor & Associate Dean, Dept of Computer Science & Engineering,R. V College of Engineering, ,Mysore Road, Bengaluru, Karnataka, India, nagarajags@rvce.edu.in,

Shreyas Srinath, Research Scholar, Dept of Computer Science & Engineering,R. V College of Engineering, Mysore Road,Bengaluru, Karnataka, India,shreyassrinath94@gmail.com

Reshma Jakabal, Assistant Professor, Dept of Computer Science & Engineering, RNS Institute of Technology, Bengaluru, Karnataka, India,reshma.jakabal@gmail.com

industries that do not take proper measures regarding the treatment of wastes before releasing them into the environment. These factories may be good for the growth of jobs for people or provide necessities to human beings but they can become a serious threat to mankind when it comes to pollution. The factory owners many times become irresponsible and forget to replace their air filters as a result, the toxic wastes are directly released into the environment leading to the degradation in the quality of the air. It is indicated clearly see that A.Q.I. (Air Quality Index) which is a method to scale how polluted is the air. Many cities who were once pollution-free and had good A.Q.I., has fallen into the hands of air pollution and led to degradation in their A.Q.I. To deal with this problem there is a need to first measure their quantity. There are many elements like CO2 (Carbon dioxide) and NO (Nitrogen monoxide) that enhance the level of pollution. If their quantity is known then they can be treated easily. There are many gases like CO2, SO2, NO, NO2, etc. Thus, the air pollution monitoring system has become a necessity, rather than a luxury in today's conditions. In the current state, the Internet of Things method to build the monitoring and prediction system.

The motive of the system is set with the requirements of each module. The first one is the detection and measurement of pollutants in the air using the IoT framework. The second one is storing the recorded data of pollutants in the cloud. The third one is to analyze the recorded data in the cloud and proper classifications are to be made and from that, the pollutant reduction method must be devised. The last objective is, based on the previously collected data to predict the future air quality using machine learning algorithms.

II. RELATED WORK

A study was conducted on the air pollution system that was connected to the internet. The working principle was that the sensors would collect the air samples and measure the values of gases and other toxic substances. There were several parameters detected by the sensors including benzene and Carbon-di-oxide. The quality of air was calculated based on the detected pollutants and displayed on the web-based system. An alarm was also connected that would be triggered if the value was above the safe level. Moisture and other air quality parameters were also detected in this system. In another system implemented in the literature, the Arduino microcontroller was used for processing purposes. The objective was to detect the quality of air and monitor it to measure the changes in the pollution levels. MQ 135 sensor was used in this experiment, that would detect the pollutants in the air and displayed on the screen.

The implementation is simple but effective to measure the required parameters and display it on the web-based system.

Once a study was conducted in explored the integration of hardware framework. Arduino microcontroller was connected to a Wi-Fi module (ESP 8266) that sends the values of the processor to any other Wi-Fi module or to be displayed on a web page.

Gas sensor MQ 135 was used to detect the pollutants and monitor their levels. A buzzer was connected to act as an alarm. The voltage output was calculated which was proportional to the level of pollutants detected by the sensor. LCD is used to display the level of pollution.

There are several types of air monitoring systems that are designed and implemented in the form of pollution monitoring, atmospheric dispersion systems, etc. The basic functionality requires a sensor, microcontroller, display or an alarm. These models require higher accuracy in order to sense the pollutants and microcontroller for faster processing. These systems can also be a part of a larger system that monitors the air quality across the city.

III. SYSTEM ARCHITECTURE

The Air pollution monitoring and prediction system consist of three modules, Gas measuring module, controlling and storing unit and prediction module.

The gas sensor attached system will be deployed in the polluted area then the sensors start sensing the different types of pollutants. They will produce the output in the form of analog values. These values are sent to the controller which will continue with the next process.

The purpose of this module is to set up the configuration parameters that will control the order assignment module. This can be fed by analog values from the gas sensors. And after converting these values will be sent to database unit for storing. The digital values on the LCD display which can get by converting analog values from the gas sensors. The functionality of this module is, the input is given to the microcontroller which will convert the analog values to digital values and make the display on the LCD screen, and send those values to database storage for further use.

In the third module, the processed data is stored in the cloud. And these data are used and applying some machine learning techniques to these data, which can predict future values. The processed data is stored in the cloud and machine learning algorithms technique. The set of processed data from the cloud will be taken and applying machine learning techniques and get predicted value of the air quality, and these predicted values on the data logging page.
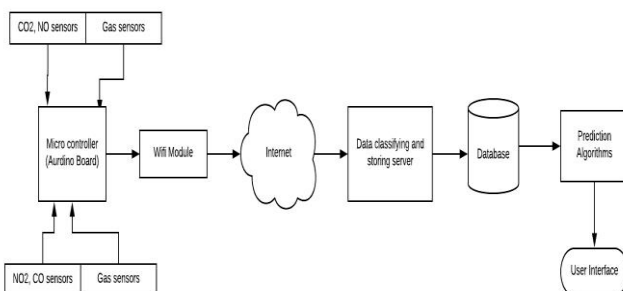

Figure 1: System Architecture

IV. K MEANS

K-Means originated from the concept of vector quantization. In the data mining field, it is very popular for cluster analysis. The objective of the K-means clustering method is that it uses different clusters to classify the observations. The mean value of the cluster is used to prototype the value and use the values for further processing. This algorithm has a loose relationship with the k- nearest neighbor classifier. It is a very famous machine learning technique used for classification but it is sometimes confused with the K Means because both of them have got a similar name.

A set of x values are considered in the form of observations. These observations could be the detected values, air pollutants in this case. Various sets (S) are used to cluster the values of observations. The sum of these values will be used to calculate the cluster of the observations and the mean is calculated using the following formula.

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{\mathbf{x}\in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg\min_{S} \sum_{i-1}^{k} |S_i| \operatorname{Var} S_i$$

where Ui represents the mean of S values. The value of deviations in the cluster is used to calculate the value of equivalence.

$$\arg\min_{S} \sum_{i=1}^{k} \frac{1}{2|S_i|} \sum_{\mathbf{x},\mathbf{y}\in S_i} \|\mathbf{x} - \mathbf{y}\|^2$$

The value of equivalence can be calculated using the following formula.

$$\sum_{\mathbf{x}\in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \sum_{\mathbf{x}\neq\mathbf{y}\in S_i} (\mathbf{x} - \boldsymbol{\mu}_i)(\boldsymbol{\mu}_i - \mathbf{y})$$

In this equation, the only constant is the value of variance. The sum of squared deviations is an important factor in this case. This value follows the total variance law in the functionality.

V. SUPPORT VECTOR MACHINE

Support Vector Machines in machine learning algorithms are supervised learning algorithms that analyze the data that has to be used for classification and regression. It was invented by Vladamir N. Vapnik and Alexey Ya Chervonenkis in 1963. When a set of training examples is to be given then each part belongs to one of the two categories. Support Vector Machine is a training algorithm, which basically creates a model that assigns new examples to one or the other category. S.V.M. builds a hyperplane or a collection of hyperplanes in a high or infinite-dimensional space.

This can be efficiently used for various calculations such as classification and also regression which can be the parameters used in the detection of outliers. A good separation is achieved if the value calculated is separated from the training point with a large deviation. It is considered that more is the amount of margin then less is the generalization error of the classifier.

VI. SPECIFICATION OF THE SYSTEM

A. *Arduino IDE Software*

Arduino IDE is a software application that enables C-program to be compiled and upload to the Arduino board. It is programmed to work with different types of microcontrollers.

Once after compiling and uploading the code necessary action is performed. In this paper, Arduino hardware acts as a microcontroller that receives the values from the sensors and processes them accordingly before sending them back to the actuators for a specific function.

B. *Android Studio*

Android Studio is a platform that enables the users to build their own apps across various platforms and the work accordingly to their function. The app is designed to prevent minor accidents and send the same information to the person's emergency contacts with the location.

C. *Cloud storage*

Cloud is used to store the data sent from the gas sensors in a tabulated manner and can access the same data from the cloud and apply the algorithms to predict the future values and predicted values are stored and return to the cloud.

Process of the system

The current system is developed to prevent and detect the pollution level in the metropolitan cities. Since the real-time data are taken from the gas sensors, all three modules are executed simultaneously. The user needs to login to access the webpage. The system will provide the percentage of different gases present in the air are logged on the Data logger with future prediction.

Gas sensor called MQ-135 which reads the percentage of different gases present in the air. The gas sensors here are the Carbon Monoxide and Nitrogen Monoxide. The sensors detect their values when they are exposed to a certain environment. This can be performed in heavily polluted areas to somewhat less polluted areas .These values are sent to the cloud using a wi-fi module. In the cloud, these values are tabulated using classification algorithms. From these recent tabulated values, 3 to 4 sets of values are taken and trained for predictive machine learning algorithms. By this technique, will get the future air quality. This predicted value is displayed on the webpage. Users can then access the data from there and then can be used later.

D. *Arduino Microcontroller*

The ATmega328 is a microcontroller chip mounted on the Arduino UNO. It acts as the microcontroller and central part of the system. Wi-Fi module and gas sensors are connected. The obtained data from the different gas sensors is processed and send to the cloud using a Wi-Fi module that is connected to the controller. If more data sets are required then again gas sensors are turned on and start reading the pollutant gas level. For more precession, Data read is done from sensors continuously to get real-time values.Hence, the air pollution monitoring and prediction system rely on the Arduino UNO microcontroller.
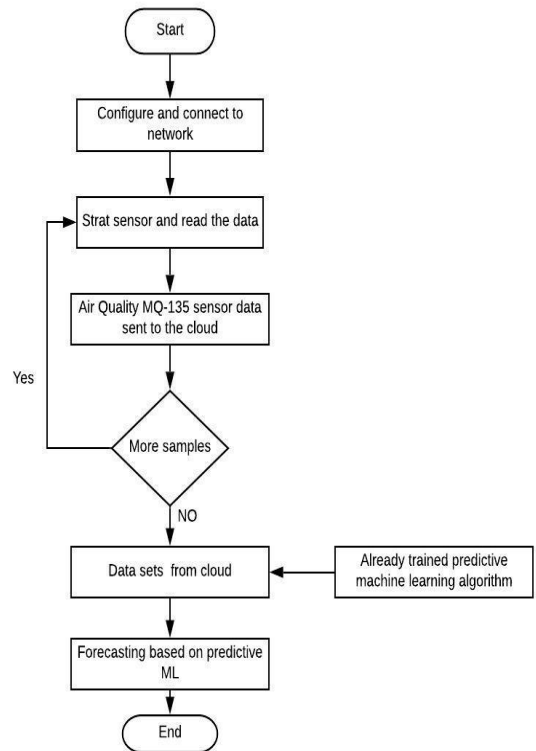

Figure2: Flow chart of the system

E. *. MQ-135 Gas sensor*

This sensor is very easy to use and can use in almost all conditions. This sensor will detect the gases present in the air of concentrations anywhere between 20 to 2000 ppm (parts per million). This sensor is made of some sensing element that is very sensitive and has a very quick response time. CO2 and NO sensors are used to detect the specific gases present in the air. After reading the values they are processed in controller i.e. conversion of analog values to digital value takes place in the controller.

F. *Wi-Fi Module*

There are several types of Wi-Fi modules available. However, the requirement was to have a low-cost module that would support the Arduino microcontroller platform and runs on a voltage less than 5 volts. ESP 8266 was selected for the reason that satisfies all the programming requirements and also has TCP/IP stack for internet connectivity.

G. *Buzzer*

The alarming system was used in the experiment that should be triggered if the programmed value reaches above or below a threshold level. A buzzer is favorable because it produces an audio signal alarming the users of any deviation from the threshold values

H. *LCD Display*

Liquid crystal display is a basic display that uses 16x2 values to be displayed as programmed. It is used to indicate the percentage of air pollutants present in the air.

VII. RESULTS

The sensors successfully monitored the readings of the quantity of the pollutants in the air in the place tested and those values were fed to the system. In doing so, the software part did successful work by predicting the number of pollutants in the future. The predicted values have been cross-checked and they were authentic.

The test was conducted over a month of time and it came out to be that the predicted values for the next few days came exactly similar to the actual values of pollutants when compared after few days.
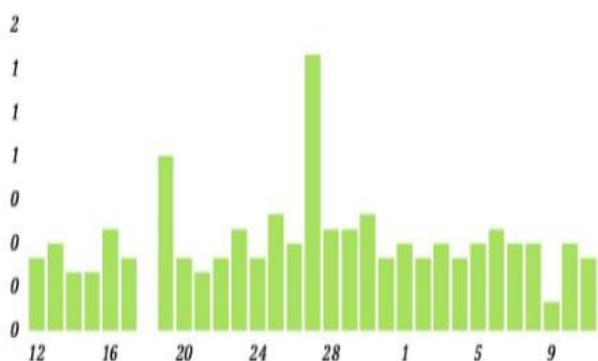

Figure 3: Pollutant level of CO

The figure 3 shows the predicted values of the CO pollutant in the atmosphere, tested in terms of a microgram per metre cube.
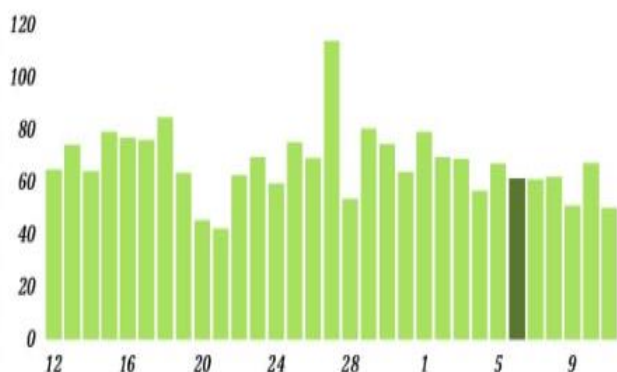

Figure 4: Pollutant level of NO

The figure 4 states the level of NO predicted for the next few days and then the same activity was repeated for over a month of time and it was crosschecked, it came out to be that the predicted values were similar to the actual values.

The model performs well in all the conditions ranging from heavily polluted areas to low polluted areas. The only thing to be worried about is that the wi-fi module stays connected and is working in order to acquire real-time data.
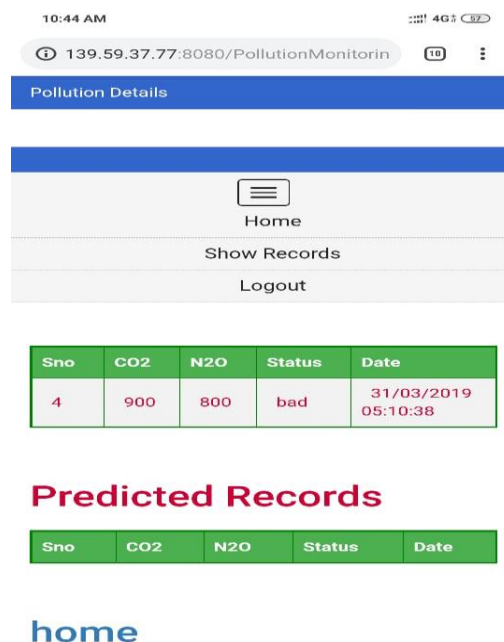

Figure 5: Data Logging Page

The figure 5 shows the Datalogging page when the model started working, where it also stores the predicted values in their records apart from predicting the level of pollutants in the future.

VIII. CONCLUSION AND FUTURE ENHANCEMENTS

This paper introduces an air pollution monitoring and prediction system using IoT. The experimentation involved the detection of various parameters that harm air quality. The values are detected using a sensor. The microcontroller calculates the values and triggers the alarm if the value is beyond the threshold value.

This is very simple and having quick response time compared to existing monitoring systems. This system extended to make an analysis of real-time data. Using this system, Air pollution value can be monitored from anywhere with an internet connection. The values detected are accurate and the monitoring system can be extended to detect several other air pollutants. This system can be improved using equipment of higher accuracy.

REFERENCES

1. Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao. "A survey on security and privacy issues in internet-of-things," IEEE Internet of Things Journal, vol. 4, no. 5, pp. 1250–1258, Oct. 2017.
2. K. Ding and P. Jiang. "Rfid-based production data analysis in an IoT-enabled smart job-shop," IEEE/CAA Journal of Automatica Sinica, vol. 5, no. 1, pp. 128–138, Jan. 2018.
3. Anne H Ngu, Mario Gutierrez, Vangelis Metsis, Surya Nepal, and Quan Z Sheng. "Iot middleware: A survey on issues and enabling technologies," IEEE Internet of Things Journal, vol. 4, no. 1, pp. 1–20, 2017
4. V. Gokul and S. Tadepalli. "Implementation of a wifi based plug and sense device for dedicatedair pollution monitoring using iot," in 2016 Online International Conference on Green Engineering and Technologies (IC-GET), pp. 1–7, Nov. 2016.

5. Nikheel A. Chourasia, Surekha P. Washimkar," ZigBee Based Wireless Air Pollution Monitoring" International Conference on Computing and Control Engineering (ICCCE 2012), 12 & 13 April, 2012

6. Nikheel A. Chourasia, Surekha P. Washimkar," ZigBee Based Wireless Air Pollution Monitoring" International Conference on Computing and Control Engineering (ICCCE 2012), 12 & 13 April, 2012

7. Riteeka Nayak, Malaya Ranjan Panigrahy , Vivek Kumar Rai and T Appa Rao:IOT based air pollution monitoring system Vol-3, Issue-4, 2017

8. Palaghat Yaswanth Sai: An IoT Based Automated Noise and Air Pollution Monitoring System Vol. 6, Issue 3, March 2017

9. .Kriegel, Hans-Peter; Schubert, Erich; Zimek, Arthur (2016). "The (black) art of runtime evaluation: Are we comparing algorithms or implementations?". Knowledge and Information Systems. 52 (2): 341–378.

10. MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press.

11. Lloyd, S. P. (1957). "Least square quantization in PCM". Bell Telephone Laboratories Paper. Published in journal much later: Lloyd., S. P. (1982).

12. Arthur, D.; Manthey, B.; Roeglin, H. (2009). "k-means has polynomial smoothed complexity". Proceedings of the 50th Symposium on Foundations of Computer Science (FOCS).

13. Garey, M.; Johnson, D.; Witsenhausen, H. (1982-03-01). "The complexity of the generalized Lloyd - Max problem (Corresp.)". IEEE Transactions on Information Theory. 28 (2): 255–256

14. Inaba, M.; Katoh, N.; Imai, H. (1994). Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering. Proceedings of 10th ACM Symposium on Computational Geometry. pp. 332–339

15. D., Manning, Christopher (2008). Introduction to information retrieval. Raghavan, Prabhakar., Schütze, Hinrich. New York: Cambridge University Press.

16. Arthur, David; Vassilvitskii, Sergei (2006-01-01). How Slow is the k-means Method?. Proceedings of the Twenty-second Annual Symposium on Computational Geometry. SCG '06. New York, NY, USA: ACM. pp. 144–153.

17. Ben-Hur, Asa; Horn, David; Siegelmann, Hava; and Vapnik, Vladimir N.; "Support vector clustering"; (2001); Journal of Machine Learning Research, 2: 125–137.

18. Press, William H.; Teukolsky, Saul A.; Vetterling, William T.; Flannery, Brian P. (2007). "Section 16.5. Support Vector Machines". Numerical Recipes: The Art of Scientific Computing (3rd ed.). New York: Cambridge University Press. ISBN 978-0-521-88068-8. Archived from the original on 2011-08-11.

19. Pradhan, Sameer S., et al. "Shallow semantic parsing using support vector machines." Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004. 2004.

20. Barghout, Lauren. "Spatial-Taxon Information Granules as Used in Iterative Fuzzy-Decision-Making for Image Segmentation". Granular Computing and Decision-Making. Springer International Publishing, 2015. 285–318.