International Journal of Engineering & Scientific Research
Vol. 5 Issue11, November 2017,
ISSN: 2347-6532 Impact Factor: 6.660
Journal Homepage: http://www.esrjournal.com Email: esrjeditor@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

AIR POLLUTION MONITORING ORGANIZATION CENTERED ON IOT: PREDICTIVE AND FORECASTING MODELING

Mala Malik
Assistant Professor, Department of Computer Science, Mata Ganga Khalsa College for Girls, Kottan, Ludhiana

ABSTRACT

Air pollution is individual of the most frequent problems we face today, and it affects us all. These days, the air in most cities is polluted, and as new contaminants have been introduced, the air has become even more dangerous. It's possible that Air Pollution is caused by both human and natural activity. We are introducing new pollutant types into the atmosphere, such as Sulphide oxides, carbon dioxide, nitrous oxides, chlorine fluorocarbons (CFC), lead, and mercury. Several real-time predictingschemes have been created recently utilizing the Internet of Things (IoT). One of the most difficult aspects of this technology is developing a prediction model that can reliably anticipate a wide range of space weather characteristics. The primary purpose of the article is to track and illustrate pollution statistics, as well as provide future projections. Todetermine the most accurate prediction model and calculating model for four distinct gases, three machine learning (ML) techniques were used.

Keywords: Internet of Things, Air Pollution, Machine Learning

1. INTRODUCTION

Air contamination is a broad term that includes a wide range of pollutants in the air, from gaseous pollutants to hazardous solid particles. Particles suspended in the air may include exhaust from automobiles, dust, or hazardous chemicals emitted by companies. Chronic ailments including asthma, cardiovascular disease, and brain and lung issues may be caused by air pollution's negative impact on human health. The quality of India's air has deteriorated dramatically in recent years, especially in major cities. In addition to well-known pollutants like Carbon Dioxide (CO2), the atmosphere has also been burdened with newer pollutants including Nitrogen Dioxide, Sulfur Dioxide, besides Carbon Monoxide.

Our health is jeopardized by most of the contaminants. CO, on the other hand, is far more dangerous. Silent Killer gets its name from the fact that it kills silently and fast. It replaces the oxygen molecules in the blood with its own, stingy brain and emotion of the oxygen they need to operate. If it's in air, it rapidly enters the bloodstream and causes indications including annoyance, illness, nausea, faintness, and disorientation. As the concentration raises, symptoms such as nausea, vomiting, and loss of consciousness may occur, as might brain damage or death if the exposure is prolonged.

According to the WHO, ambient air pollution refers to contaminants in the atmosphere that may be attributed to sources such as industry, the home, and motor vehicles. Particulate matter (PM) and gaseous air pollutants are examples of air pollutants. Human health is severely harmed by fine particulate matter generated by combustion of fossil oils in automobiles, power plants, and homes or from the burning of biomass. Fine particulate matter has been linked to 25 percent of all lung cancer fatalities, 8 percent of all COPD deaths, and 15 percent of all stroke deaths. Environmental factors such as air quality index (AQI) have a significant impact on how well our environment functions and how healthy it remains. Emissions of unmanageable amounts of hazardous gases and particulate matter (PM) into the atmosphere result in air pollution due to fast economic and industrial expansion, which is followed by a massive inflow of car industry. Impurities and gases in the air combine and may spread hazardous amounts both inside and outdoors, resulting in air pollution. In many parts of huge cities and industrial regions, the Internet of Things (IoT) offers an excellent option for collecting remotely monitored levels of dangerous gas emissions and filthy air index. For location, satellite navigation, and communication, ionospheric weather predictions are becoming more dependent on the capacity to anticipate space weather phenomena.

2. RESEARCH METHODOLOGY

2.1 Proposed method anddesign

Water and energy management, trash reduction, and an IoT air quality monitoring scheme are all included in the pollution control system. Such a platform gives users, particularly companies, insight into issues such as biochemical leaks, oil tumbles, incorrect disposal of harmful chemicals, while also enhancing environmental standards and laws in the process. This system addresses issues like:

- Common stage to track & manageAir qualityindex.
- Real time procedures view, audit event& findings.
- Trigger announcements for thresholdbreaks.
- Risk tourist attractions with thresholdbreak.

The flow chart below helps visualize the implementation specifics. (fig. 1),

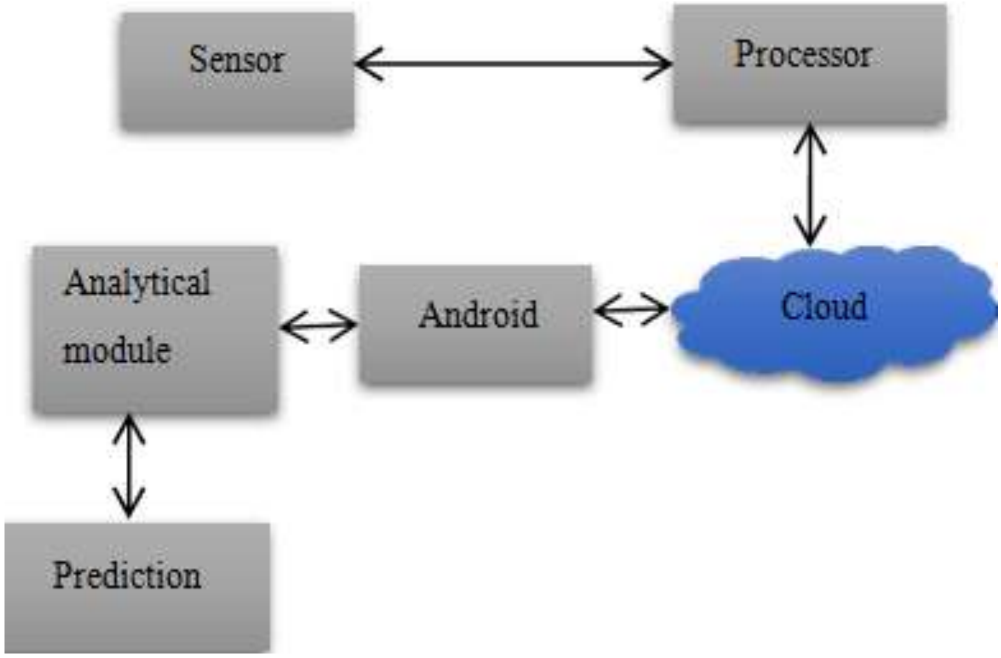

Fig. 1: System Overview

2.2 IOT DEVICE PROTOTYPE

By using the advantages of OSI layered architecture, IoT mostly succeeds in linking smart devices (implanted hardware gadgets) to Itb. It proposes a collection of Air Quality Monitoring Sensor Bits for this study, which is used to estimate the meeting of air contaminations that may be seen anywhere. As a result, all of the Air Sensors have been furnished with schemeobtainability and are linked to Itb, forming a global network of affiliated items with a small implanted stage. The Node MCU, an open source development board containing ESP8266-12E chips, has been mostly used. The MQ-2 Gas Sensor is used to get an idea of how much gas is in the atmosphere. In order to get IoT-based information, this sensor data would be captured and delivered to the Thing Speak cloud through IoT. Fig. 2 depicts the suggested framework design for the project.

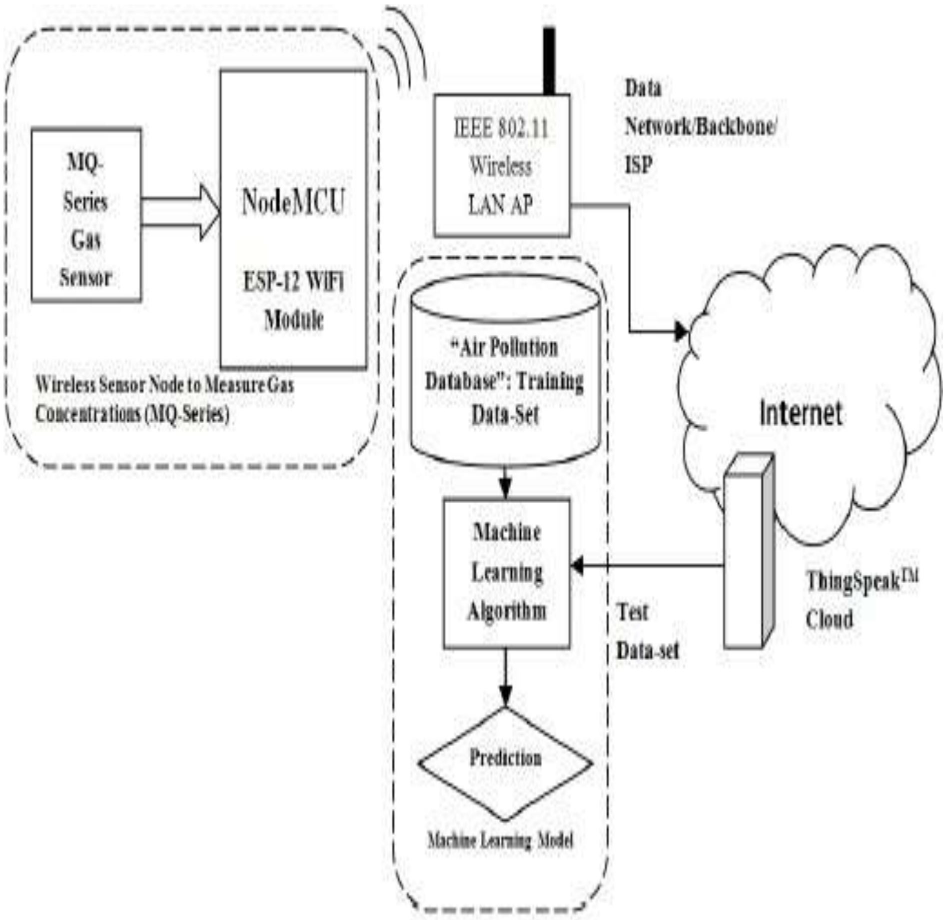

Fig. 2:Block diagram of proposed IoT created Air Pollution Monitoring Scheme

3. ANALYSIS

ML mainly controls computational strategies that increase the execution of automatingthe verification of experience-based learning improvements. AI's goal is to let machines take in more and more knowledge from more complicated arrangements of information while still handling fundamental challenges with growing savviness. The contamination anticipating model may be used in the same way a typical climate gauging is done for the next day in order for people to take precautionary measures. O3, SO2, NO2, and CO concentrations should be predicted with precision. Our model was tested using a US Environmental Protection Agency-approved contamination dataset that includes contamination data for 46 US states over the last 15 years. Most of the time, the forecast models are structured in three steps:

1. **Statistics Pre-preparing:**The first stage in building an expectation perfect is information pre-handling, which involves cleaning data, filling in missing characteristics, removing anomalies, and then organizing the data such that it is suitable for Machine Learning computations.

2. **Feature Engineering:**Features such as day, time of day, and so on are important since they increase the forecast's accuracy.

3. **Building Forecasting Model:**To foretell the future, a model is developed based on hidden knowledge that relies on external verification**.**

This information is prepared by putting target factors in a safe place, picking out the most important parts, and then using those parts throughout the computations. Cross-approval and assessment are used for the testing component of the produced models. RMSE and MAE were used as cross validation metrics to evaluate the model's robustness and scalability (MAE). The replicas are brought to life and their separate execution is judged in this piece, resulting in a skewed view of the model's performance. A forecast's mean absolute error (MAE) estimates how big a folly it is to make a prediction without considering the ramifications (1).

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - \hat{y}_j\right| \qquad (1)$$

All things careful, best highlights and prediction model determinationbe used for the hidden info for the actualization of the model. When the weather changes unexpectedly, the process of preparing, testing, and shipping will need to be repeated on a more frequent basis. Model execution may be improved by doing this iterative method over and over again. However, only Random Forest and XGBoost proved to be superior when tested on a large number of data sets. With XGBoost you may do "Outrageous Gradient Boosting" in yet another way. "Tendency-Boosting" is presented by Friedman as an approximate Greedy Function: A Gradient-Boosting Mechanism. Slope boosting is now one of most well-known methods for effectively showing seemingly insurmountable datasets. Angle boosting may be quickly and easily done using XGBoost. Because of this unique approach, XGBoost may be quite effective. In order to improve speed and efficiency, XGBoost makes use of slope-supported decision trees. A well-known AI method is the Gradient Boosting Model (GBM) in light of the decision tree. For relapse and grouping concerns, slope boosting is an AI (ML) method that generates a prediction model from a collection of flimsy forecast models based on a selection of trees. GBM uses a phase-aware design model in the same way as other boosting techniques. In order to sum things up even further, a self-assertive differentiable misery capacity is used, which allows for simplification.

Autoregressive integrated moving normal (ARIMA) is one of the best straight models in the last three decades in time plan decision. Artificial neural networks (ANNs) have been shown to be a potential alternative to traditional straight procedures in recent research practices in forecasting. When it comes to the preponderance of gauging models, ARIMA and ANN models are regularly distinct and combined. An informative index for preparation and testing can be shown in Figure 3, where the suggested framework design can be seen. Cross approval processes are included in the suggested design, which includes preprocessing and highlight extraction.

4. RESULT

After pre-processing the data and extracting the features, we've spoken about how well the various models perform. The first half of this section focuses on the data collected by our device, while the second portion extends our model to the open-source dataset. Here are some of the key findings and conclusions drawn from the collected data. There are three separate areas in Kolkata, India where pollution data has been gathered in Table 1.

Table 1 Air Pollution Connection Details

	created_at	entry_id	Chadni Chowk	Gariahat	Dharmatala
0	26-10-2017 07:59	167601.0	2.562637	3.015531	2.758462
1	26-10-2017 08:00	167602.0	2.565861	3.016337	2.761685
2	26-10-2017 08:00	167603.0	2.569890	3.019560	2.764103
3	26-10-2017 08:01	167604.0	2.573114	3.019560	2.766520
4	26-10-2017 08:02	167605.0	2.576337	3.019560	2.768132

XGBoost and Random Forest both include error evaluation tables in Table 2, whereas Linear Regression does not. As previously mentioned, the factors taken into account for error evaluation are the Mean Average Error and the Root Mean Square Error.

Table 2 Machine Learning Error Evaluation

Error Evaluation Table		
A. Linear Regression:		
Gas Name	MAE	RMSE
NO_2	0.352	0.426
O_3	0.241	0.286
SO_2	~0	~0
CO	~0	~0
B. For XGBoost:		
Gas Name	MAE	RMSE
NO_2	0.002	0.005
O_3	0.001	0.002
SO_2	~0	~0
CO	~0	~0
C. For Random Forest:		
Gas Name	MAE	RMSE
NO_2	0.005	0.023
O_3	0.010	0.0
SO_2	~0	~0
CO	0.010	0.072

In addition to traditional prediction models, we also use a time-series approach. We see that all of the models do quite well; no model is particularly terrible or outperforms any other. Though linear regression may be the weakest of the bunch, the difference between it and the rest of them is rather tiny. Other models are almost indistinguishable.

5. CONCLUSION

There is no doubt that poor air quality has a negative influence on human well-being. Meteorological sensors and a variety of vaporous sensors are used to gather air quality data remotely from inspecting bits. This information is gathered and utilized to create precise estimates of pollution using a clever machine-to-machine organizing technique. While the ZigBeeplusGPS-based air pollution observing framework uses on-going air pollution remote checking, the WSN-based air pollution checking outline uses ZigBee to transfer poison data to the pc for further analysis. The next step is to use machine learning (ML) to build estimation models using the data that has been collected. Our conclusion is that XGBoost may be used for prediction because of its approach that is level-wise, and it aids in the development of models that have low bias and low variance.

REFERENCE

1) Alirezaie, M. and Loutfi, A. (2015). Reasoning for sensor data interpretation: An application to air quality monitoring. J. Ambient Intell. Smart Environ. 7: 579–597 https://doi.org/10.3233/AIS-150323

2) M. Karar et al., Localization and tracking of aortic valve prosthesis in 2D fluoroscopic image sequences, in SPIE Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling, Lake Buena Vista, FL, USA, 2009, vol. 7261, pp. 72611Q-8: SPIE. (2009).

3) Pratima Gupta and Ranjit Kumar, ShalendraPratap Singh and Ashok Jangid, “A study on monitoring of air quality and modeling of pollution control”, IEEE conference, pp 1-4,2016

4) 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. 186203, May 2013.

5) Jones, W.D., Foster, G.P. and Putinas, J.M. (1987). The catalytic activation and functionalization of carbonhydrogen bonds. Aldimine formation by the insertion of isonitriles into aromatic carbon-hydrogen bonds. J. Am. Chem. Soc. 109: 5047–5048. https://doi.org/10.1021/ja00 250a060

6) Kim, J.Y., Chu, C.H. and Shin, S.-M.J.I.S.J. (2014). ISSAQ: An integrated sensing systems for real-time indoor air quality monitoring. IEEE Sensors J. 14: 4230–4244. https://doi.org/10.1109/JSEN.2014.2359832

Table 1 has data related to Oct 2017. Paper is to be published in August 2017. Please change that.