IoT Based Air Quality Prediction with Pollution Monitoring and Data Analytics Using Machine Learning Approach

Dr. Krishna Kumar P R[1], Triveni N[2], Lohith C[3]

[1]Registrar, Department of ISE, Cambridge Institute of Technology, Bangalore, India
Email: *rana.krishnakumar[at]gmail.com*

[2]Assistant Professor, Department of ISE, Cambridge Institute of Technology, Bangalore, India
Email: *triveninagaraj33[at]gmail.com*

[3]Assistant Professor, Department of CSE, Gopalan College of Engineering and Management, Bangalore, India
Email: *lohithchandrashekhar[at]gmail.com*

Abstract: ***A Node MCU equipped with an ESP8266 WLAN connection and a MQ Series sensor are combined in an IoT - based air contaminant monitoring system to transmit sensor readings to the Ubidots cloud. Additional components of this study include a guaging model, which is simply a subset of prescient modelling, and a realistic AI model to predict the degree of air pollution. We will use our IoT device as a model to collect the data, and to extend our model, we used an authorised open - source dataset provided by the US Government. The paper's main objectives are to track, conceive of contaminated information, and determine it. To choose the optimal predictive model and a forecasting model for calculating the air quality index (AQI) of four different gases—Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), and Ozone—three machine learning (ML) calculations were specifically carried out (O3). Here, linear regression, arbitrary woods, and xgboost are used for ML computations, while an ARIMA model is used for time - series estimation. The exhibition measurements were based on Mean Outright Error and Root Mean Square Error (RMSE) (MAE). A NodeMCU equipped with an ESP8266 WLAN connection and a MQ Series sensor are combined in an IoT - based air pollution monitoring system to transmit sensor readings to the Ubicloud. Additional components of this study include a guaging model, which is simply a subset of prescient modelling, and a realistic AI model to predict the degree of air pollution. We will use our IoT device as a model to collect the data, and to extend our model, we used an authorised open - source dataset provided by the US Government. The key purposes of the suggested framework are screening, foreseeing, and predicting contaminated information.***

Keywords: Arima, NodeMCU, MQ - series sensors, IoT, predictive modelling, and machine learning

1. Introduction

Contamination has grown increasingly commonplace as industry and urbanisation have advanced quickly. When pollutants or poisonous compounds are present in the air, it compromises human health. The IOT Air Contamination Observing Framework's main objective is to highlight the fact that air pollution is now a growing problem. For a better future and a stable life for everyone, it is crucial to monitor and test the air quality. Web of things (IoT) is gradually becoming more well - known due to its versatility and low cost. With the urbanization and with the expansion in the vehicles on street the air conditions have significantly impacted. The harmful effects of contamination include mild hypersensitivity reactions like irritation of the throat, eyes, and nose as well as a few serious conditions including bronchitis, heart diseases, pneumonia, lung diseases, and aggravated asthma. Checking gives estimations of air toxin and sound contamination fixations, which can then be dissected deciphered and introduced. This data can then be material in numerous ways. Investigation of checking information permits us to evaluate how awful air contamination and sound contamination is from one day to another. The new progressions in implanted gadgets have prompted the use of remote organization advances in checking sensor information and air contamination. The purpose of this work is to develop predictions and a gauging model for certain air pollutants, such as CO, CO2, SO2, and temperature, which are thought to be extremely harmful. These models have excellent prescient limit, speculation power and have many applications.

An Internet of Things (IoT) - based air quality monitoring system includes a MQ Series sensor connected to a NodeMCU equipped with an ESP8266 WLAN connection to transmit sensor readings to a ThingSpeak cloud. A plausible AI model to predict the air pollution level and a gauging model, which is simply a subset of prescient displaying, are included in this work's further scope. This will use our IoT device as a model to collect the data, and for expanding our model, we used an authorised open source dataset provided by the US Government. The purpose of the document is primarily to scan, visualise, and anticipate contaminated information.

The following key borders are considered in the proposed system:

1) Carbon Dioxide (CO_2): CO_2 is an inert, odourless gas that cannot be burned. Additionally, it is considered to fall under the category of suffocation gases, which have the power to interfere with the tissues' ability to receive oxygen. Since it is one of the key building blocks of the photosynthesis process, which converts solar energy into chemical energy, carbon dioxide is a gas that is essential for life on Earth. Large - scale petroleum product consumption is mostly to blame for the growth

of the CO2 group. Plants grow swiftly as a result of this increase. The rapid growth of unwelcome plants leads to an increase in the use of synthetic chemicals to eradicate them.350 sections per million is the safe level of CO2 in the air (PPM).

2) Sulfur Dioxide (SO2) - Sulfur Dioxide is a dry gas that may be identified by its distinct odour and flavour. Similar to CO2, it is mostly caused by the consumption of non - renewable energy sources and contemporary cycles. High fixations may cause breathing problems, especially in sensitive groups, much like asthmatics. This intensifies corrosive downpours. The average sulphur dioxide convergences shouldn't exceed the 350 g/m3 limit more than nine times annually, and they shouldn't in any way exceed the 570 g/m3 standard. As a 24 - hour average, the general air quality standard for sulphur dioxide is 120 g/m3.
3) Carbon monoxide (CO) - Our faculties are unable to distinguish CO since it lacks an aroma, flavour, or variety. This suggests that dangerous centralizations of the gas can form within, and that until they cause illness, individuals have no meaningful means of knowing when this is happening. Additionally, when people pass out, the side symptoms are similar to this season's sickness, which may cause victims to ignore the first signs of CO injury. CO is produced whenever a substance consumes.
4) Smoke – Approximately 1 million people worldwide, primarily from agricultural countries, have an inclination to smoke tobacco. According to a 2007 estimate, 4.9 million people consistently passed away as a result of smoking.
5) Thermosphere and humidity The estimation of temperature has an impact on our fundamental skills and is important for human health. By measuring temperature and comparing temperature variations between the past and the present, especially since the current transition using environment information, nursery influence may be noticed. Stickiness is a kind of gas that protects humans from the sun's UV rays and helps to slow down global warming, making the climate on the planet a pleasant place to live. However, when moisture increases, Earth's radiance also rises, making our lives more stressful. For diverse capacity and food handling offices, moistness is essential. When the surroundings is exposed to water, outright wetness increases from nearly nothing to roughly 30 g (1.1 oz) per cubic metre.

Due to factors that might affect human health, such as businesses, urbanisation, population growth, and automobile use, the level of pollution is rising swiftly. The most important problem facing every nation, whether it is developed or developing, is air pollution. Medical problems have been growing more quickly, especially in urban regions of non - industrialized countries where industrialization and an increase in the number of cars has caused a shipment of vaporous poisons to arrive. The harmful effects of contamination include mild hypersensitivity reactions like irritation of the throat, eyes, and nose as well as a few serious conditions including bronchitis, heart conditions, pneumonia, lung diseases, and aggravated asthma.

The issue must be addressed to save human existence and to control perilous air contamination that is expanding step by step.

Developing urbanization and number of business urban areas construct it an interest to have a definite worry of the environmental elements. Depleting to remain perception ceaselessly bound destinations like enterprises, occupied traffic lights, towns at risk to eroding high co2 fixation and so forth. Because of expanding grouping of hurtful gases in climate all living organic entities experience the ill effects of serious sicknesses connected with breath and lungs.

With the assistance of checking air contaminations with continuous information, can take preventive measure which slowly help in controlling the air contamination. In this project, we will create an IOT - based air contamination checking framework. Using the web, we will monitor the air quality and set off an alert when it drops below a certain point, which denotes when there is an adequate amount of harmful gases like CO2, smoke, CO, SO2, and temperature in the air. It will display the air quality in PPM on the LCD and on the page so that we can easily screen it.

2. Related Work

E. Gambi and colleagues describe Excellent air quality is anticipated to keep up with great illnesses in the living environment. Involving areas might be detected as having unsafe gases and contaminated air. Using a boundary like the IAQI estimate provided by the suggested framework is a fruitful way to learn about the quality of the air. The proposed framework includes an ESP8266 to interface with the IOT stage and provide data about contaminations to the faculty. The method for identifying the climate's toxins helps one take better care of themselves.

According to M. Molinara et al., this research provides a prepared Web of Things solution for differentiating and grouping contaminants in the indoor air observation system. It is dependent on a scaled - down, low - power coordinated architecture that has the ability to both identify and handle situations. The managing stage is mostly based on AI methods, implanted in a low power and low asset small regulator unit, for grouping reasons, while the detecting stage is built up of a sensor show on which electrical impedance estimations are carried out by a microprocessor, dubbed SENSIPLUS.

According to Vishal Choudhary et al., the Metropolitan air pollution is a serious problem in many urban areas all over the world. A drawn - out receptivity to airborne particles is harmful to one's health. Government agencies and companies use pricey, EPA - approved Air Quality Checking Frameworks (AQMS) using expensive, specialised equipment. In any event, the air quality data provided by organisations is area - based and is based on postponed midpoints. The AirQ stage is a clever and useful solution for air quality monitoring. The AirQ device provides continual and location - specific air quality information while being small and inexpensive. It supports wireless technology advancements as WiFi, 4G, Bluetooth, LoRa, and others. To aid information executives and the execution of cunning information examination computations, they have established a cosmology - based backend. The AirQ Dashboard and mobile application provide real - time alerts and information on the air quality.

The model that Rohan Kumar Jha used to demonstrate here combines Arduino UNO programming and hardware with the Gas sensor MQ135, MQ7, and dust sensor GP2Y1010AU0F, which aids in recognising chemicals like NO2, CO, and PM2.5 while conservatively estimating their sum. The Air Quality is also monitored by this investigation work using an IOT investigation stage called ThingSpeak using web - connected equipment via the Wi - Fi module ESP8266. Additionally, it can synchronise the ongoing data into our mobile application, which was created specifically for this purpose using Android Studio. Finally, the circuit displays the PPM values as well as the level of gas air quality on an Android application using data from Things.

Managing air pollution poses a crucial natural challenge under bright metropolitan circumstances, according to S. Ameer et al. Constantly updating pollution data enables local experts to clearly analyse and make decisions about the city's current traffic situation. The Web of Things - based sensors' organisation has significantly altered the factors involved in predicting air quality. A closer examination of these techniques is anticipated to have a better understanding of their handling time for diverse datasets. Current exploration has used several AI instruments for contamination expectation. Using four high level relapse techniques, we performed contamination expectation in this research and offer a comparable report to determine the optimum model for accurately forecasting air quality with regard to information size and handling time. We have done contamination assessment using a variety of datasets and guided tests using Apache Flash. These relapse models have been examined using the Mean Outright Error (MAE) and Root Mean Square Error (RMSE) as assessment criteria.

The suggested study effort by J. Huang et al. aims to address the problems of correspondence geographical plan, exactness of the Nature of Administration (QoS) levels assessment, detection throughput, and streamlining of power use. The suggested IoT - based air quality monitoring system has been deployed in the accepted work in indoor and outdoor locations to measure air quality parameters as PM10, PM2.5, carbon monoxide, temperature, and stickiness. The suggested structure is also tested at a variety of administration levels at both indoor and outdoor locations.

Since people typically spend more than 80% of their time inside, indoor air quality (IAQ) monitoring has attracted growing attention with the rapid expansion of industrialization and urbanisation in contemporary society, according to L. Zhao et al. To address the requirements of a broad range of scenarios, a smart IAQ finder (IAQD) coordinated with various correspondence interfaces has been conceived, built, modified, communicated, and tested. The IAQD provides accessible estimations of the IAQ data, including temperature, moistness, CO2, residue, and formaldehyde. Modern Web of Things (IoT) innovations enable the IAQD to be applied to wired communications, short - range remote communications, and remote transmission to the cloud. The IAQD is coordinated with Modbus, LoRa, WiFi, general bundle radio assistance, and NB - IoT correspondence interfaces. The intended programming in the cloud enables users to track the indoor air quality (IAQ) of their residence, place of business, or other locations.

BhavikaBathiya and colleagues discuss how air pollution has become a challenging problem in all megacities. It is crucial to constantly monitor the state of the environment, however data on pollution obtained from permanent stations is insufficient for a precise assessment of the spray contamination level of the air. The spatiotemporal target of the obtained information can substantially be expanded by portable estimation devices. Tragically, when compared to stationary sensors, the quality of results from portable, inexpensive sensors is generally worse. Because of this, it's critical to evaluate the many aspects of testing frameworks according to the characteristics of the portable sensors used. The hour of contamination detection is treated as an irregular variable in this paper's technique. As far as we know, we may quickly draw a conclusion that the components of the checking framework together have the combined appropriation capacity of the contamination identification time.

3. Proposed Methodology

The article is primarily intended for screening; consider the information on contamination and its anticipation. The information examination approach is specifically used to identify the most accurate predictive model and a determining model for calculating AQI of four different gases: Temperature, Sulfur Dioxide (SO2), Carbon Dioxide (CO2), and Carbon Monoxide (CO).

IOT Gadget:

The model was planned utilizing an Arduino Uno microcontroller, Wi - Fi module ESP8266, CO sensor, CO2 sensor, SO2 sensor, temperature sensor, GSM and a 16 by 2 fluid precious stone presentation (LCD) Screen.

The framework outline methodology was arranged into Five (5) layers as displayed. The main layer was the natural boundaries which are gotten by estimation. The subsequent layer was the investigation of the qualities and highlights of the sensors. The direction, detection, estimation, setting of the edge valve, frequency of response, time, and space made up the third layer. Information gathered from sensors made up the fourth layer. The all - encompassing knowledge climate made up the fifth layer. When the sensor was controlled by the microcontroller, it collected data and used the Wi - Fi module to send it to the internet for analysis. Clients had the option to screen estimated boundaries on their cell phones.

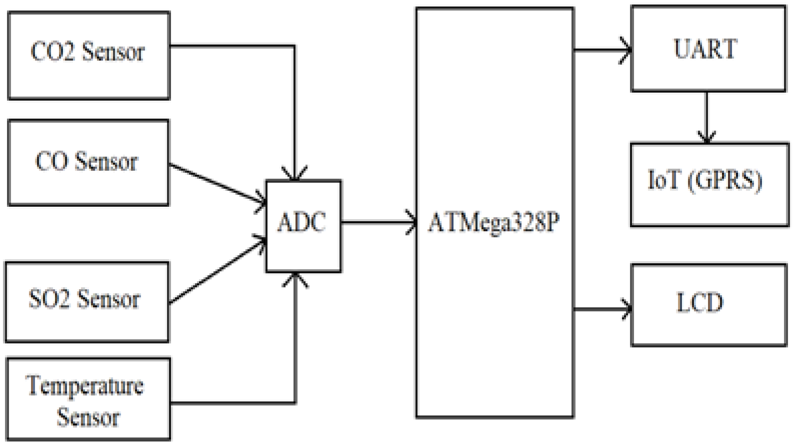

Figure 1: Block Diagram

MG811 CO2 Sensor

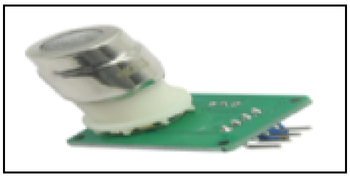

Figure 2: CO_2 Sensor

MG811 CO_2 sensor estimates vaporous carbon dioxide levels by observing how much infrared (IR) radiation consumed via carbon dioxide atoms.

This kind of sensor is utilized in checking indoor air quality, landfills, process control, and controlled climate cultivation.

MQ7 Gas Sensor

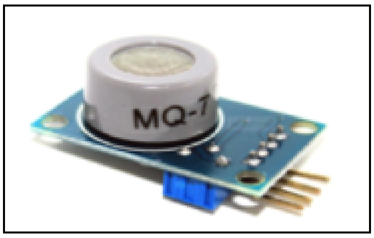

Figure 3: MQ7 Sensor

MQ7 Gas sensor is predominantly used to distinguish Carbon Monoxide.
It can identify Carbon Monoxide Gas in the scope of 20 PPM to 2000 PPM in the air.
4 pins - Vcc, G, A0, D0

DGS SO2 Sensor

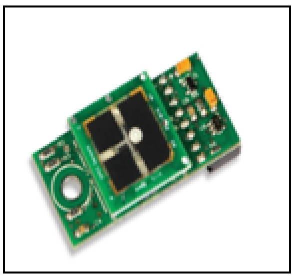

Figure 4: SO_2 Sensor

SPEC Sensors currently offers a simple method for adding gas detecting to the Web of Things. Consolidating our Screen - Printed Electro - Synthetic sensor innovation.

The cutting edge gadgets and calculations, empowers simple incorporation of little, lightweight, superior execution, super low power utilization gas detecting into remote, versatile, and organized arrangements.

Temperature Sensor (DS18B20)

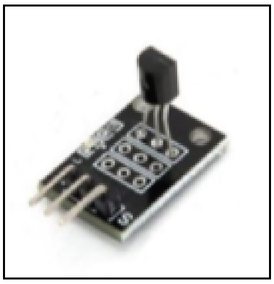

Figure 5: Temperature Sensor

1) Model → DS18B20
2) Advanced waterproof sensor
3) 3 Pins
4) VCC, Gnd, Information
5) Gathers temp information from climate utilizing I2C transport (8bit)

Modeling for prediction and forecast:
We aim to predict accurately groups of CO_2, SO_2, CO, and temperature. We try to develop a predictive model using the verified data from the gas sensors and their AQI value that displays the graphical depiction of the AQI value by one - stride ahead supposition and dynamic figure produced by ARIMA calculation.

The most popular method for developing forecast models essentially controls three parts:

- **Information Pre - handling:** The first stage in creating an expectation model is information pre - handling, which involves cleaning the data, adding missing characteristics, eliminating exceptions, and organising the data in a way that is suitable for AI calculations.

- **Highlight Designing:** Elements, such as day, month, and time, are among the key factors that increase prediction accuracy.
- **Creating a Determining Model:** A model is developed to predict the future, for instance using oblique information in light of verifiable information.

One of the important direct models in time plan decision during the past thirty years is autoregressive coordinated moving average (ARIMA). Recent experiments with artificial neural networks (ANNs) in prediction reveal that ANNs may be a potential alternative to conventional straight techniques. Regarding the similarity in choosing models, ARIMA models and ANNs are continuously distinguished and fall into various categories. The suggested Framework's design explains how to gather information for testing and preparation

4. Implementation

These experiences and findings from our research are drawn from the context of the work that we did. The AQI values projected for a single area are represented graphically.

```
In [42]: y_forecasted = pred.predicted_mean
         print(y_forecasted)
         y_truth = data['2018-10-23':]
         print(y_truth)

Date
2018-10-23    120.858173
2018-10-26    126.556226
2018-11-16    120.279137
2018-11-19    121.331396
2018-11-22    107.507738
2018-11-26    118.753330
2018-11-28    130.979785
2018-11-29    122.336238
2019-02-01    117.026381
2019-01-13    115.190405
2019-01-21    110.621976
2019-01-23    129.627457
2019-01-28    105.187684
2019-01-30    108.815296
dtype: float64
            AQI
Date
2018-10-23  131
2018-10-26  117
2018-11-16  118
2018-11-19   87
2018-11-22  111
2018-11-26  141
2018-11-28  123
2018-11-29  109
2019-02-01  104
2019-01-13   94
2019-01-21  137
2019-01-23   85
2019-01-28   90
```

Figure 6: Comparison of Actual and Predicted Data

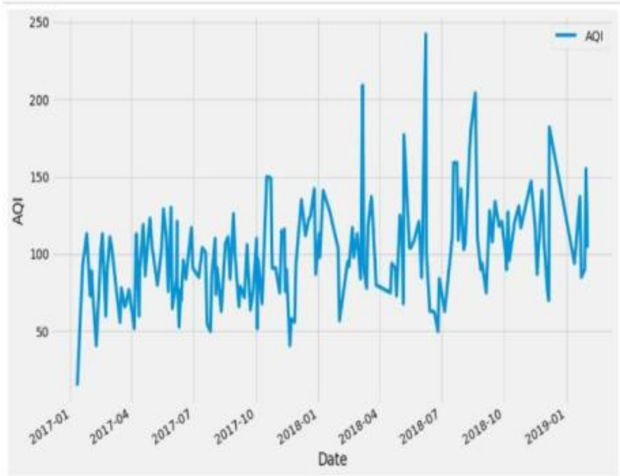

Figure 7: Analyzing the Result of Dataset

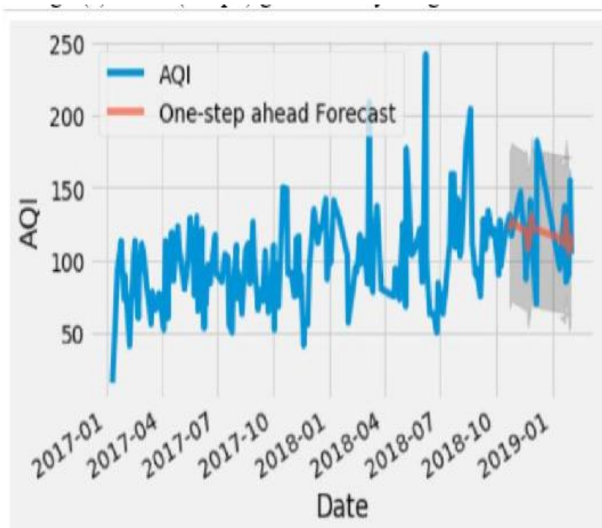

Figure 8: Results Evaluation Graph produced using the one-step- ahead forecasting technique

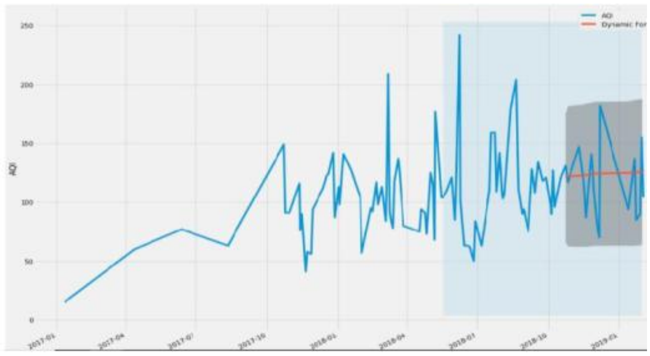

Figure 9: Graph of results produced by the Dynamic Forecast Model

5. Conclusion

A fundamental problem that directly affects human prosperity is air quality. Data on air quality are gathered remotely from measuring devices that are equipped with a variety of vaporous sensors as well as meteorological ones. Utilizing intelligent machine to machine stage, this data is checked and exploited as a component of anticipated obsessive evaluations of defilements. The next step is to prepare from the gathered data using ML - based calculations to create the assessment models. ARIMA works reasonably well as an estimating model, which may be used to provide daily figures that closely resemble standard weather gauging.

References

[1] S. Ali, T. Glass, B. Parr, J. Potgieter and F. Alam, "Low Cost Sensor With IoTLoRaWAN Connectivity and Machine Learning - Based Calibration for Air Pollution Monitoring, " in IEEE Transactions on Instrumentation and Measurement, vol.70, pp.1 - 11, 2021, Art no.5500511, doi: 10.1109/TIM.2020.3034109.

[2] E. Gambi, G. Temperini, R. Galassi, L. Senigagliesi and A. De Santi’s, "ADL Recognition Through Machine Learning Algorithms on IoT Air Quality SensorDataset, "in IEEE Sensors Journal, vol.20, no.22, pp.13562 - 13570, 15Nov.15, 2020, doi: 10.1109/JSEN.2020.3005642.

[3] M. Molinara, M. Ferdinandi, G. Cerro, L. Ferrigno and E. Massera, "An End to EndIndoor Air Monitoring System Based on Machine Learning and SENSIPLUS Platform, "in IEEE Access, vol.8, pp.72204 - 72215, 2020, doi: 10.1109/ACCESS.2020.2987756.

[4] Vishal Choudhary, Jun HaoTeh, Victoria Beltran, Hock Beng Lim, "AirQ: A SmartIoT Platform for Air Quality Monitoring", in 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC).

[5] Rohan Kumar Jha, "Air Quality Sensing and Reporting System Using IoT" in 2020Second International Conference on Inventive Research in Computing Applications (ICIRCA).

[6] S. Ameeretal., "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities, " in IEEE Access, vol.7, pp.128325 - 128338, 2019, doi: 10.1109/ACCESS.2019.2925082.

[7] J. Huang et al., "A Crowdsource - Based Sensing System for Monitoring Fine - GrainedAir Quality in Urban Environments, " in IEEE Internet of Things Journal, vol.6, no.2, pp.3240 - 3247, April 2019, doi: 10.1109/JIOT.2018.2881240.

[8] L. Zhao, W. Wu and S. Li, "Design and Implementation of an IoT - Based Indoor Air Quality Detector with Multiple Communication Interfaces, "in IEEE Internet of Things Journal, vol.6, no.6, pp.9621 - 9632, Dec.2019, doi: 10.1109/JIOT.2019.2930191.

[9] Ayaskanta Mishra “Air pollution monitoring system based on IOT: Forecasting and predictive model using Machine learning”. International Conference on Applied Electromagnetics, Signal Processing and Communication 2018.

[10] Bhavika Bathiya; Sanjay Srivastava, Biswajit Mishra; “Air Pollution Monitoring Using Wireless Sensor Network, ” IEEE International WIE Conference on Electrical and Computer Engineering, 2016