INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1199)
Available online at: https://www.ijariit.com

Automated air pollution monitoring and forecasting system using Machine Learning

Vinayak Pragada
vinayakpragada7@gmail.com
Vidyalankar Institute of Technology, Mumbai, Maharashtra

Romil Keniya
romilkeniya97@gmail.com
Vidyalankar Institute of Technology, Mumbai, Maharashtra

Abhijeet Chauhan
abhijeetchauhan431@vit.edu.in
Vidyalankar Institute of Technology, Mumbai, Maharashtra

Beena Ballal
beena.ballal@vit.edu.in
Vidyalankar Institute of Technology, Mumbai, Maharashtra

ABSTRACT

Various Air Quality Monitoring systems have been recently developed and installed in some of the cities such as Mumbai, Bangalore, Delhi, Hyderabad etc. But the actual results do not satisfy enough. Several issues in the existing system were the number of gases detected were very few. There was a need for such a system that monitors more amount of gases in less amount of time. Time was also an issue regarding the monitoring of the gases. Some gases consumed an enormous amount of time, so it became crucial to decrease it and make a more efficient system. The proposed system is automated air pollution monitoring and forecasting system using machine learning which uses an Arduino Mega, MQ gas sensors module for its development along with some other basic components like16x2LCD, Buzzer, Potentiometer. MQ135 gas sensor detects gases like NH3, NOx, alcohol, benzene, smoke, CO2, SO2 etc. which are the main reason for degrading air quality. Wi-Fi module is used to connect our system to the Cloud so that all the readings and the data can be transferred to desired server for the forecasting purpose. A trained model is expected to use ARIMA algorithm for best prediction purpose. This model is then applied to the upcoming data and forecasting is done. Various existing system were learned which are developed till date, it was found that the air quality numerical model such as WRF-Chem, community multi-scale air quality model (CMAQ), CAMx, NAQPMS were used. The drawbacks in these models were found like, Source list was not updated in time for WRF-Chem, Detailed information about the source of the pollutants and other variables are generally not known. After knowing all the facts from the study, this system is expected to use a Scikit machine learning tool for forecasting. This model is expected to detect harmful gases in chemistry labs, washrooms, private organizations, public places etc. The proposed model monitors the upcoming data in no time and if the value goes beyond the threshold value it activates the buzzer and the server is notified. The future scope of the system can be extended by integrating it with more sensors to detect more harmful gases precisely.

Keywords— *Machine Learning, Arduino, Time Series Forecasting, Gas Sensors*

1. INTRODUCTION

Air Pollution all over the world is increasing day by day. It is happening due to rapid increase in vehicles, improper disposal of harmful gases from the industries, rapid increase in population, meteorological factors such as global warming etc. Therefore, harmful gases such as Carbon monoxide, Nitrogen oxides, Ground-level ozone, Particle Pollution (particulate matter), Sulphur oxides, and Lead etc. gets released into the air which results in air quality degradation. Specially in the urban areas, air pollution is a real-life problem. In the urban areas, the increased number of petrol and diesel vehicles and the presence of industrial areas at the outskirts of the major cities are the main causes of air pollution. The problem is seriously intensified in the metropolitan cities. The governments all around the world are taking every measure in their capacity. Hence there is a need of a system which can not only monitor the air pollution but also forecast it, so that population gets informed about the air quality and can take preventive and required measures so as to deal with the upcoming situations.

Models such as WRF-chem and Community multi-scale air quality has been developed to overcome the needs of modern pollution problems. Primary pollutants which are monitored by these models are PM10, PM2.5, SO2, NO2, CO and O3. But these models deals with certain issues such as high cost, less efficiency, etc. Source list in WRF-chem model is not updated in time due to which it results in poor prediction performance. Also, these model does not have detailed information about the pollutants source and other variables. These are the major drawback of these models. Various other models has been developed to deal with air pollution problems. A model named

MM5 is used to predict the pollution levels from meteorological conditions which include wind, temperature and air density. This model uses various data mining classifiers such a trees. J48, rules.ZeroR, rules.OneR, bayes.Naïve Bayes, lazy.IBk and functions.SMO. This model makes use of many classifiers for comparison which takes a lot of time to predict. And it is not so efficient because this approach is only used for particulate matter, wind velocity, wind direction and precipitations.

To overcome all the drawbacks of previous models a system is designed and developed using various sensors like MQ series gas sensor with machine learning algorithms. This system developed in this project is based on Arduino Mega. The Arduino board connects with port of a device to send data to it. These sensors are integrated together to make an efficient system. The sensor data is also displayed on a character LCD. To forecast the air pollution condition ARIMA machine learning algorithm is used along with the data gathered.

2. EXISTING SYSTEM

Various Air Pollution monitoring devices have been developed lately, however the result is still disappointing since all kinds of devices have its drawbacks.
Existing systems are:
1. WRF Chem Model
2. Community Multi-scale air quality model (CMAQ).
3. Mesoscale Model (MM5) Machine Learning Algorithms that are being used for the prediction purpose in above models are Decision-trees, J48, rules.ZeroR, rules.OneR, Bayes.Naïve Bayes, lazy.IBk etc Still the results are not upto expectations.

3. DRAWBACKS

- Source list is not updated in time for WRF-Chem which results in poor prediction performance.
- Detailed information about the source of the pollutants and other variables in these models are generally not known.
- Time taken by the prediction model is very high.
- Sensors used in the system are not very efficient.
- Issues regarding to the end to end security and privacy needs to be addressed.
- Requirement of support for scalable data analysis system is required.

4. IMPORTANCE OF THE PROJECT

To provide an effective solution for air pollution, a system is developed to monitor & forecast the air quality. This can be helpful for the environment activists to pay close attention to the environment changes and pave a path to work towards improving the existing conditions based on the readings and calculations performed on them.

The graphical representations will help the users visualize the circumstances better and the prediction will be made much more appropriate with the amount of data provided later. More the data feeded into the system , the more accuracy of the prediction which will ultimately increase with time. This project will also be a cheaper alternative to resolve the issue than the existing systems.

5. IMPLEMENTATION

PHASE 1: (Planning, Analysis, Design, Coding)

- Planning: Conceptualizing and designing the architecture for proposed solution. Gathering all the required information. Choosing the best components in all aspects.
- Analysis: Critical study, analysis and review of feasibility for proposed solution. Selecting the components such as Arduino MEGA, MQ gas sensors, easy and portable LCD.
- Design: Designing the flow chart & Algorithm for the proposed system.
- Coding: Using different Machine Learning Algorithm such as neural network will be able to take different dataset and predict the result.

PHASE 2: (Integration, Testing, Deployment)

- Integration: Integrating of different proposed modules that sense different gases present in air.
- Testing: Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution.
- Deployment: Give the completed prototype to the authorized person for evaluation purpose and beta testing.

6. HARDWARE

6.1 Arduino Mega

Arduino MEGA is one of the most popular prototyping boards. It is small in size and packed with rich features. The board comes with a built-in Arduino bootloader. It is an Atmega 328 based controller board which has 14 GPIO pins, 6 PWM pins, 6 Analog inputs and on board UART, SPI and TWI interfaces. It is used in the project to get the data from the sensors, do the necesssary calculations and push it to the serial output port of the connected device.

6.2 MQ Series Gas Sensors

The MQ series of gas sensors use a small heater inside with an electro-chemical sensor. The MQ-135 gas sensor senses the gases like ammonia nitrogen, oxygen, alcohols, aromatic compounds, sulfide and smoke. The operating voltage of this gas sensor is from 2.5V to 5.0V. MQ-135 gas sensor can be implementation to detect the smoke, benzene, steam and other harmful gases. MQ2 gas sensor module is useful for gas leakage detection (home and industry). It is suitable for detecting H2, LPG, CH4, CO, Alcohol, Smoke or Propane. Gas Sensor(MQ9) module is useful for gas leakage detection (in home and industry). It is suitable for detecting LPG, CO, CH4. Due to their high sensitivity and fast response time, measurement can be taken as soon as possible.

6.3 LCD Module

This I2C 16x2 Arduino LCD Screen is using an I2C communication interface. It means it only needs 4 pins for the LCD display: VCC, GND, SDA, SCL. It will saves at least 4 digital / analog pins on Arduino. All connector are standard XH2.54 (Breadboard type). You can connect with jumper wire directly. It is used to display the values of all the sensors updated every 2seconds.

6.4 Buzzer

A 5V Active Alarm Buzzer Module for Arduino is an audio signaling device, which may be mechanical, electromechanical, or piezoelectric. Just like what you are viewing now, it is 5V DC Electronic Part Active Buzzer Module. Using top quality material, it is durable in use. Buzzer will be activated if any of the sensor values cross the threshold value and will be active until it reaches to potentially accepted value.

7. WORKING OF A HARDWARE:

The project will have three purposes. First it uses MQ2 gas sensor to detect smoke, secondly MQ9 sensor is used to detect any combustable gas and finally MQ135 gas sensor will be used to detect the overall gas quality. The sensors will be calibrated

based on the R0 values with respect to their data sheet values and the values thrown by each sensors will be fetch on different analog pins of Arduino Mega. These values will be presented on the LCD. It will also be monitored individually for enabling the buzzer when the values cross a certain threshold value. And these values will also be given to a port of the laptop where the python code will fetch these values. Then these values will be sorted and sent to the website where it will be not only displayed on the website continuously. There will also be a prediction model which will be trained based on the real values of 135 sensors. As you pass the dates of future in the model from the website , it will return the predicted value of the 135 gas sensor on that date.

8. SOFTWARE

8.1 Arduino IDE

To Program Arduino MEGA we need IDE. Arduino IDE contains a text editor for writing code, a message area, a text console, a toolbar with buttons for common functions and a series of menus. It connects to the Arduino and Genuino hardware to upload programs and communicate with them. For this project we need two libraries: Software Serial and Liquid Crystal.

8.2 Dash

Dash is a productive Python framework for building web applications. Written on top of Flask, Plotly.js, and React.js, Dash is ideal for building data visualization apps with highly custom user interfaces in pure Python. It's particularly suited for building applications with real time and interactive graphs.

Through a couple of simple patterns, Dash abstracts away all of the technologies and protocols that are required to build an interactive web-based application. Dash is simple.

9. MACHINE LEARNING ALGORITHM:

Time Series is a set of observations on the values that a variable takes at different times. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus, it is a sequence of discrete-time data.

Line charts are very frequently used to plot time series. It is used in various fields like statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction and largely in any domain of applied science and engineering which involves temporal measurements. Time Series forecasting is the use of a model to predict future values based on previously observed values. Time series are widely used for non-stationary data.

Time Series analysis comprises methods for analysing time series data in order to extract meaningful statistics and other characteristics of the data. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called “Time Series analysis”, which focuses on comparing values of a single time series or multiple dependent time series at different points in time. Interrupted Time Series analysis is the analysis of interventions on a single time series.

10. TIME SERIES FORECASTING METHODS:

10.1 Auto-Regression (AR) Model

Linear Regression model predicts the dependent variable based on linear combination of independent variables.

For Example: $Y = MX + C$;

where Y is the dependent variable and X is independent variable.

Linear Regression can also be used in time series analysis where input variables at previous time steps can be used as one of the features called as lag feature to predict the output variable.

>For example:- We can predict the value for the next time step (t+1) given the observations at the last two-time steps (t & t-1). As a regression model, this would look as follows:-

$$X(t+1) = \beta 0 + \beta 1 X(t-1) + \beta 2 X(t-2)$$

Because the regression model uses data from the same input variable at previous time steps, it is referred to as an auto regression (regression of self).

The notation for the model involves specifying the order of the model p as a parameter to the AR function, e.g. AR(p). For example, AR (1) is a first-order Auto-Regression model.

The method is suitable for univariate time series without trend and seasonal components.

Example:- AR (1) : $X(t+1) = \beta 0 + \beta 1 X(t-1)$

AR (2) : $X(t+1) = \beta 0 + \beta 1 X(t-1) + \beta 2 X(t-2)$

10.2 Moving Average (MA)

A moving average term in a time series model is a past error (multiplied by a coefficient). The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.

The notation for the model involves specifying the order of the model q as a parameter to the MA function, e.g. MA(q). For example, MA (1) is a first-order moving average model.

The method is suitable for univariate time series without trend and seasonal components.

$$Xt = \beta + \varepsilon t + \theta 1\ \varepsilon t\text{-}1$$

Example:- MA(1) : $Xt = \beta + \varepsilon t + \theta 1\ \varepsilon t\text{-}1$

MA(2) : $Xt = \beta + \varepsilon t + \theta 1\ \varepsilon t\text{-}1 + \theta 2\ \varepsilon t\text{-}2$

AR or MA are not applicable on non-stationary series. The primary difference between an AR and MA model is based on the correlation between time series objects at different time points.

10.3 Autoregressive Moving Average (ARMA)

The Autoregressive Moving Average (ARMA) method models the next step in the sequence as a linear function of the observations and residual errors at prior time steps.

It combines both Auto-Regression (AR) and Moving Average (MA) models.

The notation for the model involves specifying the order for the AR(p) and MA(q) models as parameters to an ARMA function, e.g. ARMA(p, q). An ARIMA model can be used to develop AR or MA models.

The method is suitable for univariate time series without trend and seasonal components.

10.4 Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) method models the next step in the sequence as a linear function of the differenced observations and residual errors at prior time steps.

It combines both Auto-Regression (AR) and Moving Average (MA) models as well as a different pre-processing step of the sequence to make the sequence stationary, called integration (I).

The notation for the model involves specifying the order for the AR(p), I(d), and MA(q) models as parameters to an ARIMA function, e.g. ARIMA(p, d, q). An ARIMA model can also be used to develop AR, MA, and ARMA models. The method is suitable for univariate time series with trend and without seasonal components.

11.CONCLUSION

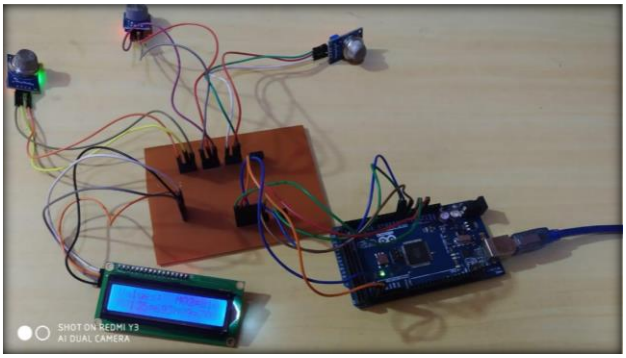
This research proposed a smart air pollution monitoring system that constantly keeps track of air quality in any particular area and displays the air quality measured on an LCD screen. It also sends data measured to the "Dash Board" platform. The system helps to create awareness of the quality of air that one breathes daily. This monitoring device can deliver Real-time as well as Forecasting measurements of air quality.

12. ACKNOWLEDGEMENT:

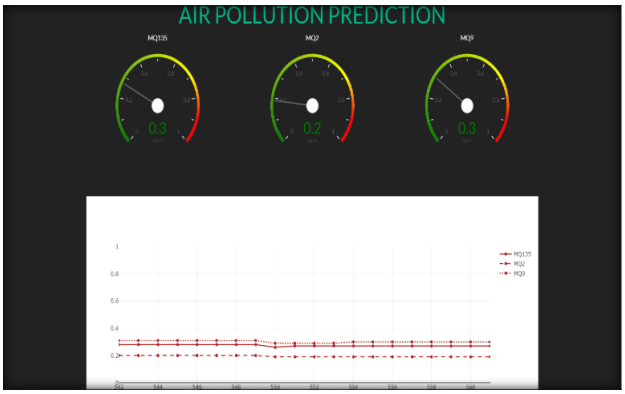
We hereby are very grateful to and wish to thank our guide and institute to give us this opportunity who supported us greatly and were willing to help us.

13. RESULTS

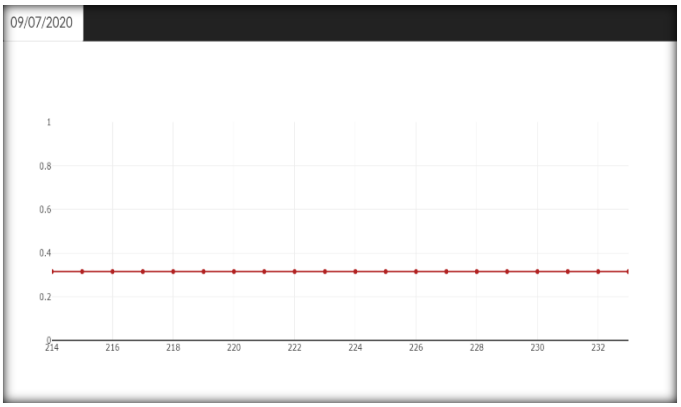
13.1 Hardware Connections



13.2 Real Time Processing Graph



13.3 Forecasting Graph



14. REFERENCES

[1] Xia Xi; Zhao Wei; Rui Xiaoguang; Wang Yijie; Bai Xinxin and Yin Wenjun, " A comprehensive evaluation of air pollution prediction improvement by a machine learning method", "IEEE International Conference on Service Operations and Logistics, And Informatics (SOLI)" 10.1109/SOLI.2015.7367615 , 2016.

[2] Yves Rybarczyk and Rasa Zalakeviciute, "Machine learning approach to forecasting urban pollution", "IEEE Ecuador Technical Chapters Meeting (ETCM)" 10.1109/ETCM.2016.7750810, 2016.

[3] Chen Xiaojun; Liu Xianpeng and Xu Peng, "IOT- Based Air Pollution Monitoring and Forecasting System", " International Conference on Computer and Computational Sciences (ICCCS)", 10.1109/ICCACS.2015.7361361, 2015.

[4] Himadri Nath Saha; SupratimAuddy; Avimita Chatterjee; Subrata Pal and Shivesh Pandey, "Pollution control using Internet Of Things (IOT)", "8th Annual Industrial Automation and Electromechanical Engineering

[5]https://www.youtube.com/watch?v=o-k7ZdjVGo&t=98s

[6] https://www.waveshare.com/mq-135-gas-sensor.html

