Environmental monitoring with the help of Internet of Things (loT) and Machine Learning.

by

Md. Sakirul Alam
16101233
Farshid Ahsan
16301088
Anik Sarker
16301007

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
April 2020

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Md. Sakirul Alam.
Date:- 13.07.2020 .

Md. Sakirul Alam
16101233

Farshid
13-07-2020

Farshid Ahsan
16301088

Anik Sarker
Date: 13. 07. 20

Anik Sarker
16301007

Approval

The thesis/project titled "Environmental monitoring with the help of Internet ofThings (loT) and Machine Learning" submitted by

1. Md. Sakirul Alam (16101233)
2. Farshid Ahsan (16301088)
3. Anik Sarker (16301007)

Of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on April 07, 2020.

Examining Committee:

Supervisor:
(Member)

Md. Motaharul Islam.

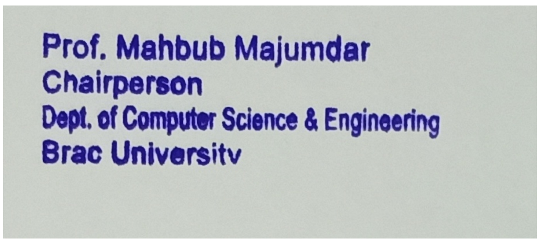
Md. Motaharul Islam, PhD
Professor
Department of Computer Science and Engineering
United International University
Ex. Associate Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Rabiul

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)



Mahbubul Alam Majumdar, PhD
Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

In recent day situation, the non-stop increase in air and noise pollution can be an enormous direful downside for our country. It has become an obligatory to manage and fittingly monitor true so, the specified steps to manage true may be undertaken. According to World Health Organization (WHO), 60dB is that the sound level human will tolerate while not gradual loss of hearing. Around 11.7% of the population in Asian country have lost their hearing thanks to pollution says the Department of Environment (DoE) study, that was conducted in 2017 and for pollution it affects the human body's metabolic process and cardiac systems, and additionally affects the eyes and different body organs.. Pollution additionally causes various diseases like cancer that have an effect on different body organs. Therefore, in our project, an IoT primarily based technique exploitation Raspberry Pi is employed to observe and check live the Air quality and also the noise pollution of a neighborhood, are projected. System uses air sensors to sense presence of harmful gases/compounds within the air and perpetually transmit this information to micro controller. Additionally system keeps measuring excessive sound level and reports it to the web server over IoT. We tend to additionally need to predict the pollution level of sound and harmful gases so it may be prevented more.This permits authorities to observe pollution in several areas and take action against it. For taking action, we've to research it and that we can analyze it by Machine learning algorithms

Keywords: IoT, Air Monitoring, Safety, Multi-Layer Perception, Principal Component Analysis, Long Short Term Memory.

Dedication

We want to dedicate our work to the people who are working day and night for the sake of environment. We tried to help them with this effort to win the battle against pollution and to save the environment. We also want to dedicate our work to our parents, without whom we could not achieve the anything in this world.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our advisor Md.Motaharul Islam sir for his kind support and advice in our work. He provided us with patient guidance, encouragement and advice and helped us whenever we needed help throughout our time as his students. We have been extremely lucky to have a supervisor who cared deeply about our work, and who responded to our questions and queries so promptly.

And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ϵ Epsilon

υ Upsilon

ANN Artificial Neural Network

AQI Air Quality Index

CNN Convolutional Neural Network

CSS Cascading Style Sheets

DoE Department of Environment

DTR Decision Tree Regression

$HTML$ Hypertext Markup Language

IoT Internet of Things

LR Linear Regression

$LSTM$ Long Short Term Memory

MLP Multi Layer Perception

PCA Principal Component Analysis

PM particulate Matter

RNN Recurrent Neural Network

SDG Sustainable Development Goals

WHO World Health Organization

XGB Extreme Gradient Boosting

Chapter 1

Introduction

In today's world it is impossible for us to live a single day without internet and now we are using this technology in all of our gadgets and those are somewhat linked to internet and we can control them from a remote distance. This technology is called Internet of Things (IoT). Internet of Things (IoT) has introduced a new paradigm of communication between humans and various devices. It constitutes a network between objects, devices, homes and so on equipped with electronics, sensors, software and network connectivity. These devices maintain a suitable protocol which enables them to interchange data with one another. We are using this technology to solve various problems in our life and our goal is to use this technology for the sake of environment and to improve the environmental condition and prevent it from being polluted. We also want to use machine learning algorithms in our devices to make it smarter and more robust. Today IoT is considered as one of the most significant emerging technology and gaining huge attention in almost all sectors.
We want to use these modern technologies to prevent and monitor one of the greatest Problem of modern life and a great threat to the world which is 'Air and Sound Pollution'. Developing country like Bangladesh is the worst victim of this problem. The capital of Bangladesh, Dhaka is second most polluted city in terms of air in the whole world. We want to use these modern technologies to prevent the problems and to make people more aware about the effects of the pollution.

1.1 Motivation

All living creatures have to breathe air in order to survive in the world. All creatures can survive few days or weeks without food or water but none live more than few minutes without breathing air into the lungs. That's how important air is to the world but sadly it is the most polluted element in the entire environment. Specially developing country like Bangladesh where air is polluting continuously because of rapid development. The SDG (Sustainable Development Goals) states that by the UN (United Nations) to build a clean world and to build a healthy environment there are seventeen goals we must achieve. According to SDG, Goal 3: Maintaining a safe life and promoting well-being for all ages, Goal 7: Maintaining access to affordable, secure, renewable and modern energy for everyone; Goal 12: Ensuring renewable consumption and development patterns; Goal 13: taking immediate steps to tackle climate change and its impact, Goal 14: Protect and sustainably utilize rivers, seas and aquatic resources, Goal 15: sustainably manage forests, counter desertification, stop and reverse land destruction, stop loss of biodiversity. Such targets cannot be met if, at a large stage, the key component of our ecosystem is contaminated. Mainly because of these reasons we chose this topic for our thesis.

1.2 Thesis Contribution

Previously many researches have done research and implemented many algorithms into the collected data from the environment but we developed an integrated system which is able to collect the data from the environment with the help of IoT and later on we have applied four machine learning algorithms to analyze the data and the predict the level of pollution for future. Main contributions of our thesis are given below:

• We have implemented Google cloud based architecture using Firebase to store and publish data into our web interface.

• We have implemented four machine learning algorithms and integrated them together.

• We have implemented a device which is very low cost and affordable comparing to the existing air quality monitoring devices out there in the market.

We have tried to build our whole system with affordable and available microcontrollers and sensors so that it can be used everywhere and people all over Dhaka city. Our main focus is to create awareness about the current condition of environment pollution and take proper step against it.

1.3 Thesis Outline

In this section we will briefly discuss the following chapters. In chapter 2 we have tried to show other people's work and give brief description about some conference and journal papers related to our work. Precious work regarding this subject can be understood by reading this chapter. Chapter 3 is all about the knowledge that we need for this thesis to conduct. We have briefly discussed the things which knowledge is needed to conduct the thesis further. In chapter 4 description of our proposed system is given. Along with this the description of the components which we have used in our thesis is also given in this section. All the description of our algorithm and how those will be used in our system is described in chapter 5. Chapter 6 includes the technology which we have used to build our website and brief description about the usage of those technology in our system. Chapter 7 is about the results we have got from our research. Finally, chapter 8 concludes our paper by future work and conclusion.

Chapter 2

Literature Reviews

Chen Xiaojun et al.[8] presented IoT based air pollution and forecasting system using artificial neural network. Their work is mainly based on two methods for different seasons according to the geography and climate. They worked on mainly mean temperature, average dew point temperature, average sea level pressure etc. They were using 24 input layer, 4 hidden layers and 1 output layer in artificial neural network algorithm. In their research the nodes of input increases to 29 and nodes of hidden layer increase to 6 and they had gained 90% confidence and it is happened because they were using five meteorological factors.

Yasser Alsouda et al.[19] presented a machine learning based noise classification using a low- power and inexpensive IoT unit. K-nearest neighbors and support vector machine algorithms had been used for noise classification. They used a dataset of about 3000 sound samples in eight sound classes. They had achieved a noise classification accuracy in the range 85-100%.

Aarti Rao Jaladi et al.[8] Presented Environmental Monitoring Using Wireless Sensor Networks (WSN) based on IoT. Universally they build a project for any scale to monitor the parameter of each environment. They were using raspberry pi and different sensors which helps to collect the data from the environment. After that they stored the data in a web server and whenever it's time for the need they fetch the data from the server and show it. In their project they were using Raspberry Pi which works as a base station because it connects the number of distributed sensor nodes via Zigbee protocol.

Ayaskanta Mishra [25] presented Forecasting and Predictive Modeling using Machine Learning. To send the data from the sensor he uses ThinkSpeak cloud and he also uses MQ series sensors which is connected to a NodeMCU, equipped with a ESP8266 WLAN. They were using machine learning algorithm for predicting air pollution level and forecasting model, which is basically a subset of predictive modeling. He built a IoT device for collecting the data and use a dataset for their research which is provided by the US govt. Three Machine learning algorithm were used for their project to find out the best predictive and forecasting model for calculating AQI of four different gases. They were used Linear Regression, Random Forrest, XGBoost for prediction and ARIMA model for time series forecasting. They were observed that the random forest had the best performance.

Arnab Kumar et al.[26] presented A Raspberry Pi powered cloud-based air and sound Temperature and Humidity monitoring system. Their project is mainly focused on to monitor air quality index and noise intensity of a region. There are four modules for example, the Air Quality Index Monitoring Module, the Sound Intensity Detection Module, the Cloud-based Monitoring Mod- ule and the Anomaly Notification Module, are mainly used on their project. At first, they were measuring the air quality index based on five criteria which pollutes the air. The sound intensity was measured on using respective sensor. After that, the cloud-based monitoring software ensures the process of data acquisition using the Raspberry Pi wi-fi support that fulfills the objective of periodically analyzing information. Finally, in the event of an undesired situation, the anomaly warning module warns the user.
Cuiping Wang et al.[29]monitored noise data with Zigbee and GPRS. They also used propagation neural network to classify data into different categories. They have developed their software system using VB language and flex network technology and they published their result on internet.

Somansh Kumar et al[15] presented a real-time air quality monitoring system which includes various parameters such as PM 2.5, carbon monoxide, carbon dioxide, temperature, humidity and air pressure. They have converged IoT with cloud computing for data retrieving from different sensor. Raspberry Pi was used for this model and the measured parameter was shown in IBM Bluemix Cloud.
Sandro Rodriguez Garzon et al.[23] Presented a system about awareness of air pollution and alert services that notifies citizens via a mobile about the air quality information when they enter an area with an air pollution that exceeds a user-defined level.

Ke Hu et al.[6]presented a machine learning model that combines both sparse fixed station data with dense mobile sensor data to calculate air pollution for any given hour or any given day in Sydney. They have used SVR (Support Vector Regression) is similar to DTR (Decision Tree Regression) and RFR (Random Forest Regression), and higher than XGB (Extreme Gradient Boosting), MLP (MultiLayer Perceptrons), LR (Linear Regression) and ABR (Adaptive Boosting Regression). After calculating the air pollution from their model, the result shows us the boundary of the air polluted area and it is validated through a field trial.

Khaled Bashir Shaban et al.[9] given a system regarding watching and forecasting urban pollution. They used low price motes that area unit equipped with many gases and earth science sensors. These motes wireless communicate to an intelligent sensing platform that consists of many modules. These modules area unit responsible for receiving and storing the info, preprocessing and changing data into helpful information. They were victimization 3 machine learning algorithms for accurate the foretelling models for ballroom dancing and multi-step prior to concentrations of ground-level gas (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2). These cubic centimeter algorithms area unit support vector machines, M5P model trees, and artificial neural networks (ANN).

Yi-Ting Tsai et al.[29] bestowed a system regarding observance PM2.5 monitoring with machine learning algorithm like RNN (recurrent neural network) with LSTM (Long Short- Term Memory). They used keras that may be a high-level neural network Api written in python and it's conjointly capable of running on prime of Tensorflow. The network coaching information is collected by the Taiwan EPA (Environmental Protection Administration) from 2012 to 2016.They completed their experiment to work out the anticipated price of PM2.5 in sixty six stations in Taiwan for next four hours. The result shows that the planned technique will effectively predict the worth of PM2. 5 .

Chapter 3

Background Study

3.1 Importance of Monitoring Air Pollution

Air is the most important element for human being to survive in this world. In order to maintain a good and sound health, clean air is essential for every creature. Even though clean air is a primary need of people in every place but city area is being highly affected because of air pollution. The pollution level of almost every major cities in the whole world is at alarming rate[12] [33]. Various air contaminants understood or suspected adverse effects on human health and the environment. Such emissions spread primarily from space heating, power generation or from motor vehicles. Pollutants from these source are not only harming the city areas it is also wide spreading and travelling long distances which is causing trouble and harming people all over the country[18] [4]. Generally, when someone is young and in good health, moderate rates of air pollution are unlikely to have any significant consequences in the short term. Nevertheless, higher rates and long-term exposure to air pollution can lead to more severe health-causing symptoms and conditions. This affects not only the respiratory and inflammatory response processes, but can also lead to more severe conditions like heart diseases and cancer[36]. People who have critical heart and lungs condition are more vulnerable to air pollution[38] [2]. Doctors recognize air pollution as one of the greatest killers of world. Air pollution causes about 29000 premature deaths in the UK and 430,000 worldwide in a single year. Air pollution can create both short term and long term effects on health. An effective air pollution monitoring system can make people more concern about the condition of the current situation of pollution. The types of people who will be benefited with this system are given below:

• Vulnerable to air pollution and have critical heart and lung condition, whose health may be affected by air pollution.

• Health care professionals, parents who look after someone whose health may be affected due to air pollution.

• Researchers and environmentalists who want to do research and trying to find the continuous reason for constant pollution.

Environmental monitoring is also important for major public and private industries as their industries is one of main reasons for the pollution. If they want to make sure about the emission level they are creating through their industry they must focus on air pollution monitoring system. With air quality control changing the responsibility from publicly funded monitoring to industry-funded monitoring, the procurement of its own quality monitoring equipment has become increasingly important for companies. Air quality monitoring system is commonly used to provide the best environmental information on air pollution rates, toxicity and harmfulness for humans .

3.2 Machine Learning

Machine learning is a subset of Artificial Intelligence (AI) by which a machine or program learns from previous experience and make predictions based on that experience in the future. Machine learning mainly focuses on making prediction based on data not by explicitly written program to do certain task [22]. These algorithms get better over time and they learn from their mistake by improving the accuracy over time when tested against new data. Moreover, data are separated into test set and train set. Algorithms are trained by test set and those make predictions against the test data set [3] [28]. The algorithms are trained until a satisfactory result is obtained from the prediction. Machine learning algorithms can be categorized into three different types [28]

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

Description of these types of machine learning algorithms are given in the following subsections.

3.2.1 Supervised Learning

Supervised algorithms of machine learning are designed as an example learning. The term "supervised" learning comes from the very fact that this technique of coaching with algorithms is like having a lecturer manage the complete operation. Once coaching a supervised learning algorithmic program, the coaching information should accommodated inputs let alone the proper outputs. Throughout preparation, the algorithmic program scans patterns within the information that correspond to the required outputs. Upon coaching, a supervised learning algorithmic program ought to be able to soak up new unknown inputs and can confirm that mark ought to classify the new inputs supported previous coaching results. The goal of a supervised learning model is to predict the proper label for submitted new input file (see figure 3.1).

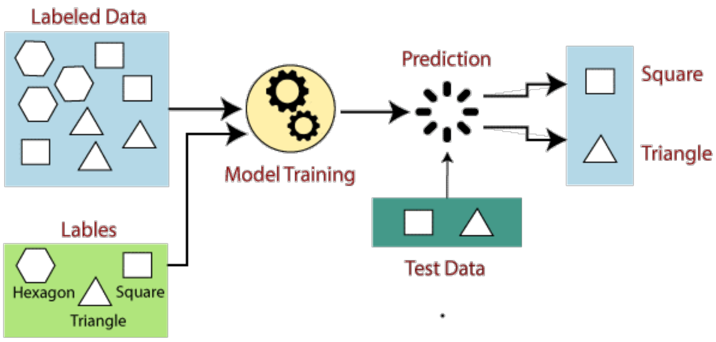


Figure 3.1: Supervised Learning

There are input variables (x) and an output variable (Y) in supervised learning, using which an algorithm is developed to learn the mapping function from input to output Y= f(x). The goal is to measure the mapping function in such a way that extracts the best result and the output variables (Y) can be estimated for that amount of data whenever new input data (x) is gathered. There are two types of supervised learning-Classification and Regression. A regression problem is when a true or constant value of the output variable is. Context- Pay, Weight, etc. Many different models can be employed, the simplest of which is linear regression. It tries to get t the best hyper-plane data going through the stages. A classification problem is when the output variable is in a subclass. A classification model tries to take an inference from observed values. A model of a classification may try to predict the value of one or more outcomes given one or more inputs.

3.2.2 Unsupervised Learning

Unsupervised learning is a type of algorithm in machine learning which are able to do more complex tasks compare to supervised learning. It uses information which are unlabeled and not classified allowing the algorithm to function without guidance on that information. However, unsupervised algorithm may be volatile comparing to other types of natural methods of learning. (See figure 3.3)

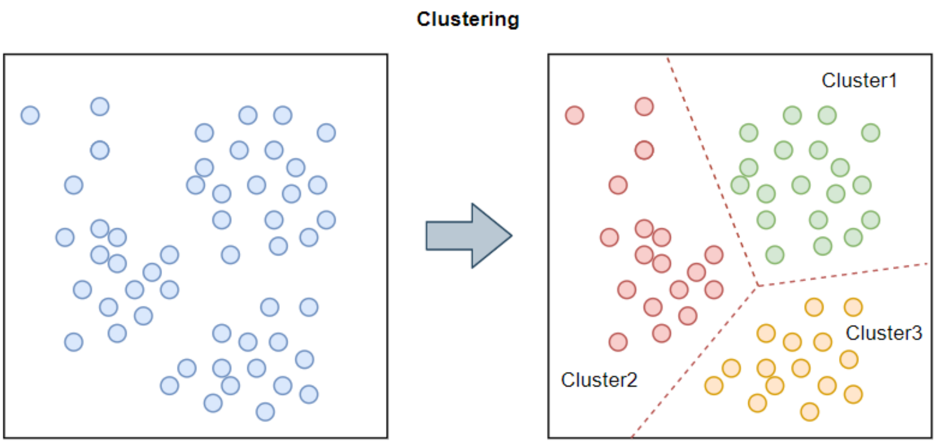


Figure 3.2: Unsupervised Learning

Unsupervised learning can be divided into two types which are: Clustering and Association problems. Clustering is mainly used to find structure or pattern in an unorganized data collection. Clustering algorithms will analyze the data and, if they exist in the data the algorithm will find natural clusters (groups) in the dataset. Number of clusters can also be adjusted by the algorithms. Association rule allows the user to set up association within large databases between data objects. This technique in unsupervised learning involves the discovery in large databases of important relationship between variables.

3.2.3 Reinforcement Learning

Reinforcement learning is teaching Machine learning models to make a set of choices. The agent learns to achieve a target in an unpredictable, potentially complex environment. It's about how the automated agents should take action in an environment. It is a part of deep learning method that lets the user to optimize any accumulated reward components. (See figure 3.3)

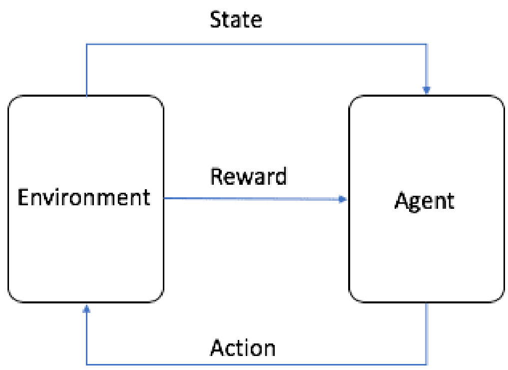


Figure 3.3: Reinforcement Learning

Chapter 4

Proposed Model

The proposed model of our system is shown below in figure 4.1. At first we need to set up the device in a suitable area where we want to monitor the quality of sound and air. We have to set a standard base for every area by which the device will be able to identify if the environment is polluted or not. In the next step data will be collected through our sensors and it will be monitored and machine learning algorithms will be run through the data to understand the pattern of the quality of air and in final stage the result will be shown in our website with multiple information and graphs to understand the quality air the current time.

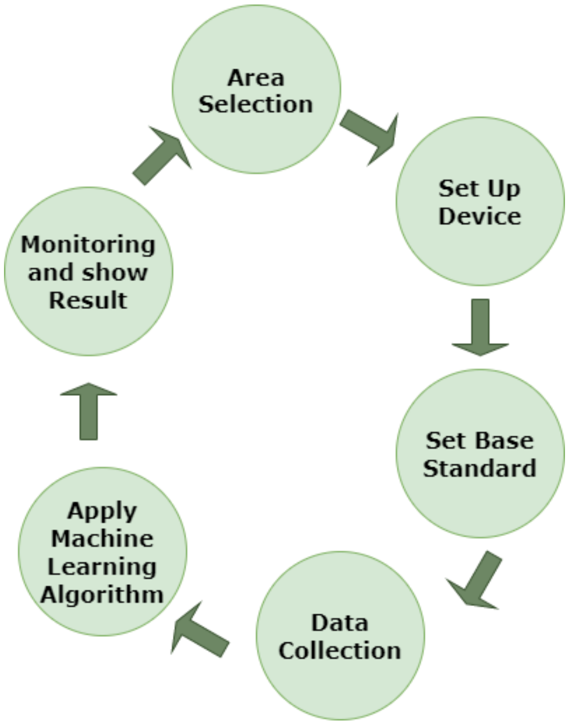


Figure 4.1: Proposed model of the system

In the above figure 4.1 we can get a general idea of our proposed model's functionality and how it will work. In the later part more detailed description of the entire system is given.

4.1 System Architecture Design

Our total system is divided into Hardware and software parts. After getting the data through the sensors using the device the test data will be stored in the allocated local host server. The server will ready with to compare the testing data with the training data. After that the system will predict the pollution level and it will show it into a web interface. Structural figure of our system architecture is given below.

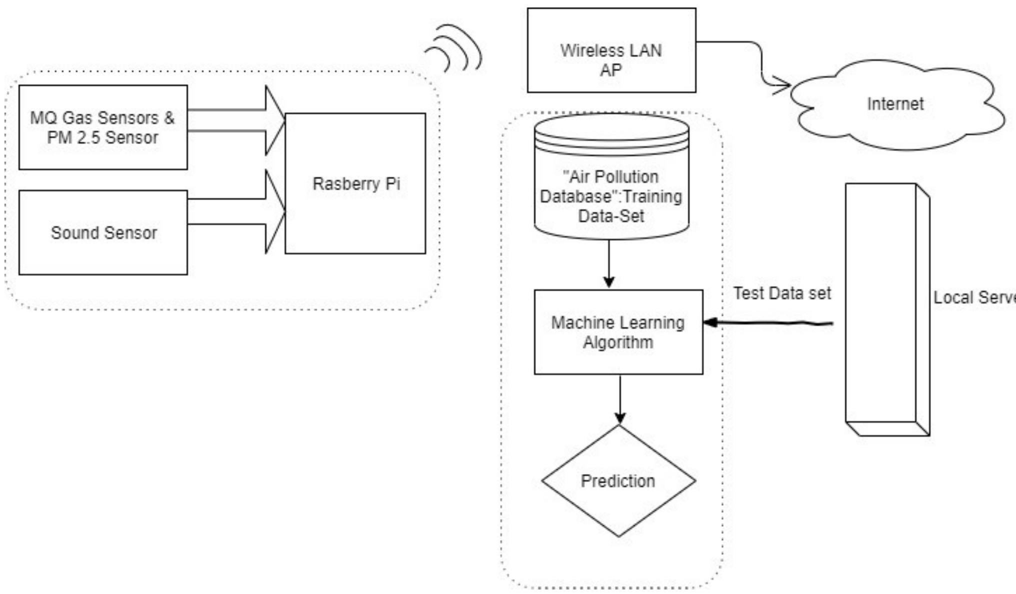


Figure 4.2: System architecture of the system

Multiple devices can be connected with our server for proper analyzing of the dataset we are continuously getting from the sensors.

4.2 Hardware Part of the System

In this section we are going to explain the hardware part of our proposed system. As mentioned earlier we have tried to use those components in our system which are both available in the market and affordable as well. Our main goal is to build a system which can be used everywhere and to maintain the cost at minimum level. In the upcoming sections details of the components that we have used in our system are given.

4.2.1 Components of the Device

We are using these four sensors in our research purpose. These sensors receive the data from the environment and send those to the server. Since we all know that Raspberry Pi does not have any analog pin so that is why we are using Arduino Mega. After reading the data from the sensors, the data are sent via Ethernet connection by the help of Raspberry Pi. The list of the components are given below which we have used:

1. Raspberry PI 3 (Model B+)

2. Arduino Mega

3. PM2.5/Dust Sensor Module

4. MQ135 Sensor Module

5. MQ7 Sensor Module

6. Sound Sensor Module

7. Breadboard

8. Wires

9. LED

10. Relay switches

These components are briefly discussed in the following sections.

4.2.2 Raspberry Pi

Raspberry Pi is a credit card sized computer that works with various devices like TV or a computer monitor. We can also use mouse and keyboard in raspberry pi. Raspberry pi is used for software and hardware based devices

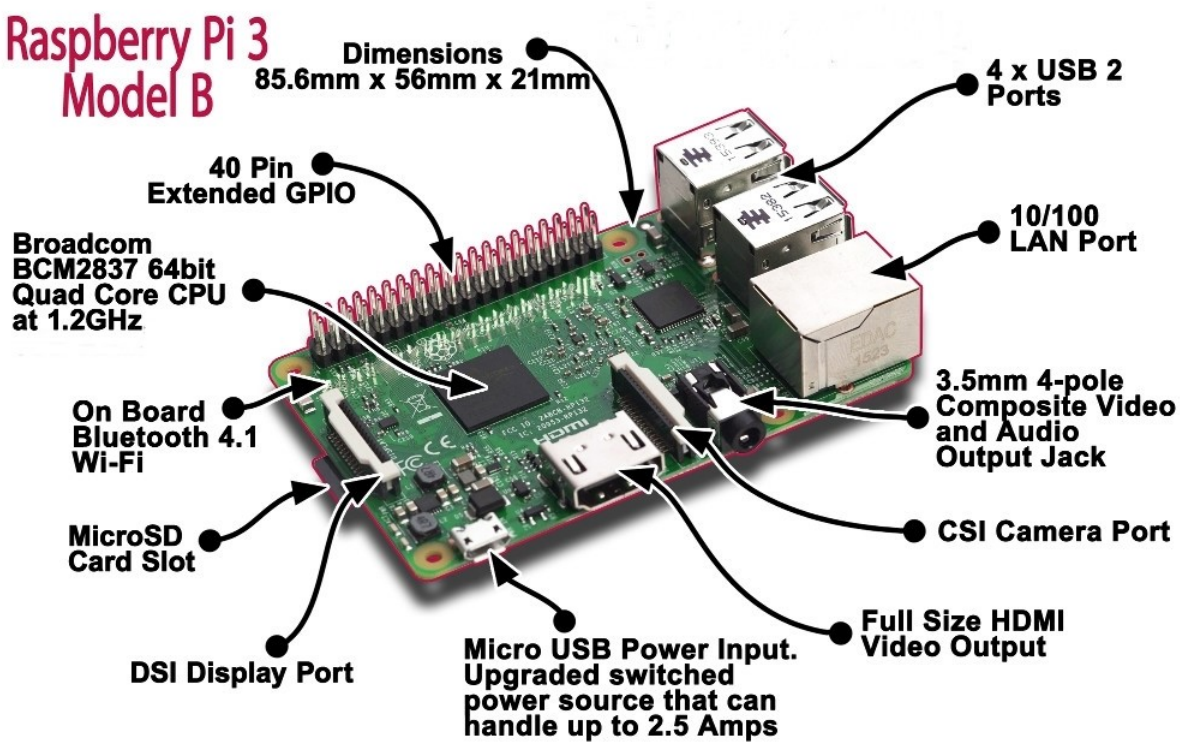


Figure 4.3: Raspberry PI 3 B+

Several versions are released and every one of them and every one models feature a Broad- com system on a chip (SoC) with associate integrated ARM-compatible central process (CPU) and on chip graphics process unit (GPU). In Gregorian calendar month 2016 Raspberry pi Model B was free and it's 1.2 GHz 64-bit quad-core processor, onboard, it's additionally 802.11n Wi-Fi and Bluetooth module. In PI day, raspberry pi Model B+ was and it's some additional feature than Model B. it's a quicker processor (1.4 GHz) and thrice quicker Ethernet [27][13]. Alternative options square measure Power over local area network (PoE) (with the add-on writer HAT), USB boot and network boot. This model additionally uses Broadcom BCM2837 SoC with a 1.2 GHz 64-bit quad-core ARM Cortex-A53 processor, with 512 K shared L2 cache. It has additionally four USB 2.0 ports,1 HDMI port, 3.5mm audio and video jack, USB power Input, DSI show port, CSI camera and forty –pin GPIO (Male Headers). Raspberry pi wants 5V DC voltage and 2Amps[17]. Raspberry pi Official software is Noobs and Raspbian, a Debian-based (32- bit) UNIX operating system distribution for transfer. It promotes on Python and Scratch as a programing language. A helpful feature of the raspberry pi has forty GPIO (general

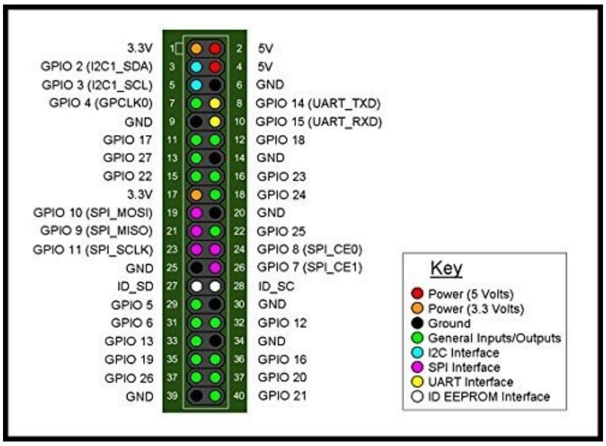


Figure 4.4: GPIO Pin-Out

Purpose input/output) pins. Every GPIO pin is of great significance. They're referred to as associate input or output pin (in software), and square measure used for various functions. There square measure 20 years V and 2 3V3 pins on the board furthermore as ground pins (6, 9, 14, 20, 25, 30, 34 and 39). A GPIO pin are often set to high (3V3) or low (0V). This can be created easier for the utilization of internal pull-up or pull-down electrical device. Pins GPIO2 and GPIO3 have mounted pull-up resistors.

4.2.3 Arduino Mega

The Arduino Mega is a microcontroller board supported the ATmega2560. It's designed for comes that require extra I/O lines, extra sketch memory and extra RAM. Arduino Mega has fifty four pins for optical input / output. Fourteen pins square measure used as outputs for PWM in those rows. It collectively has sixteen analog input pins and 4 UARTs, a sixteen rate per second generator, a USB interface, associate degree influence jack, associate in nursing ICSP header and a button. Arduino Mega simply connects it to a USB-cable device associate in nursing power it with associate degree AC-to-DC adapter or battery or Power Bank. Its computer storage of 256KB of that 8kb is employed by bootlader.

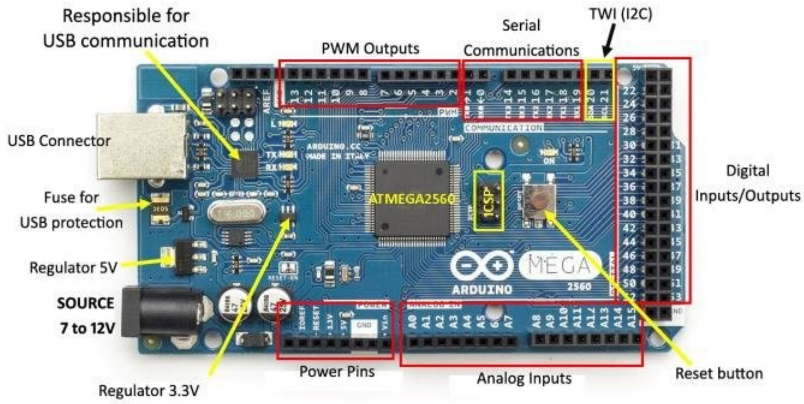


Figure 4.5: GPIO Arduino Mega

Each of the 54 digital pins on the Mega can be used for input or output. However, there are several functions we can use for input or output purpose and those are pinMode(), digitalWrite() and digitalRead(). The Arduino Mega allows serial communication on any digital pins. The ATmega2560 also supports TWI and SPI communication. All Arduino controllers programmed by Arduino IDE and it is a free software where we can use the different libraries for different sensors as it can integrate with different sensors.

4.2.4 PM2.5 Sensor/Dust Sensor

TThis module comes with a dirt sensing device supported the GP2Y1010AU0F sensing element. It's an inside infrared emitting diode (IERD) for detecting the mirrored light-weight of dirt within the air and manufacturing an output voltage proportional to the density of dirt to live dirt [31]. It will live the thing that is larger than 0.8m in diameter. It offers the analog voltage and also the operating voltage is two.5V 5.5V. Its dimensions area unit sixty three.2mm* 41.3mm. The Output Voltage of the module is proportional to dirt density within the range of 0 to 0.5mg/m3. This sensing element will find dirt within the air thus we are able to because the air setup, air monitor and PM2.5 Detector. It's six pins and these area unit V-LED, LED- GND, LED, S-GND, Vo (Analog pin), Vcc. These pins area unit connected with Arduino mega in our device.

Figure 4.6: Dust Sensor Module

4.2.5 MQ135 Air Quality Sensor Module

MQ135 is employed for measuring air quality and area unit appropriate for measuring NH3, NOx, Alcohol, Benzene, Smoke, and CO2. MQ135 includes a digital pin that creates this detector operate while not a microcontroller and it's simply detected a specific gas[20]. If we'd like the worth in PPM, then we'd like Associate in Nursing analog pin and that we connected MQ135 with Arduino to induce the analog worth. We are able to either use the digital pin or the analog pin to observe the gas and that we need to provide the facility light-emitting diode glow 5V to the present detector and once no gas is detected the output light-emitting diode will stay converted which suggests that the digital output is 0V.

Figure 4.7: MQ135 Gas Sensor Module

If we detect a particular gas by the sensor then we can see the output LED to go high along with the digital pin, if no then we can use the potentiometer until it gets high. Every each time the sensor is added to this particular gas the digital pin is going to go up. We can use the analog pin to do the same thing too. Using a microcontroller read out the meaning. This value will be directly proportional to the measured gas concentration by the sensor. We should experiment with these values to test how the sensor responds to a particular concentration of gas to improve the program accordingly.

4.2.6 MQ7 Carbon Monoxide Sensor Module

MQ7 is a carbon monoxide sensing module. It detects CO concentrations from anywhere in the air between 20 and 2000 ppm. This sensor very highly sensitive and it has a quick response time and also it uses analog resistance as an output and we can connect with the Arduino. We can operate it from 3.3V to 5V. Its dimension is 40.0mm*21.0mm

Figure 4.8: MQ7 Sensor Module

It has four pins and these are AOut, DOut, VCC and GND. AOut stands for analog output pin and DOut stands for Digital Output pin. We connect VCC to 3.3 5V and GND goes to the GND pin of any microcontroller. Its heating power is 350mW and its heating resistance RH is 313 (Room temperature). This sensor is also high sensitivity to CH_4, Natural gas and low sensitivity to smoke and alcohol. [10].

4.2.7 Sound Sensor Module

The sound sensor detects the surrounding sounds and then outputs the analog value. It has adjustable sensitivity with stable performance. Its working voltage is 3.3-5V and its dimension is 44 * 15 * 10 mm. Its response frequency is 50Hz-20 KHz and also its sensitivity is 48-66dB. This module has also a digital output and when output is 0 it goes low level and 1 indicates for high level and it has also a VCC and GND pin. We can have integrated this sensor to the Arduino or Raspberry Pi.

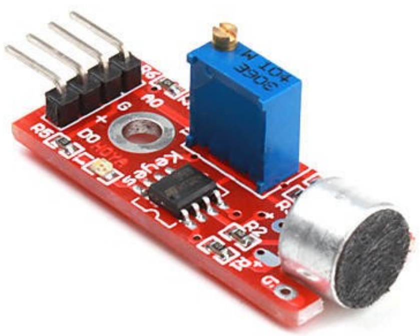


Figure 4.9: Sound Sensor Module

It has also an operating temperature and it is 40Cto +85C and its PCB size is 3.4cm*1. 6cm.It has also 3 indicator LEDs and one is for power indicator and another one is for comparator output indicator.

4.3 Schematic Diagram of the System

In the schematic diagram we showed here how we connected sensors to Arduino and Raspberry Pi. Firstly, we connected PM2.5 to the Arduino. PM2.5 has 6 pins as we mentioned earlier. However, V-Led connected to 1.5K resistance and then the other side of the resistor goes to the 3.3V of Arduino Mega. LED Pins goes to D30 of Arduino Mega. It is a digital output pin of Arduino Mega. V0 goes to the analog pin of Arduino and it is A0. LED-GND and S-GND both go to the GND pin through the positive side of the capacitor along with the resistor and the negative part of the capacitor goes to the GND of Arduino and then VCC goes to the 3.3V pin of Arduino. Secondly, we connect the MQ135 sensor to the Arduino. At first

A0 (analog pin) pin of the sensor goes to the Arduino A1 pin then GND goes to the GND of Arduino and VCC goes to the 5V of Arduino.

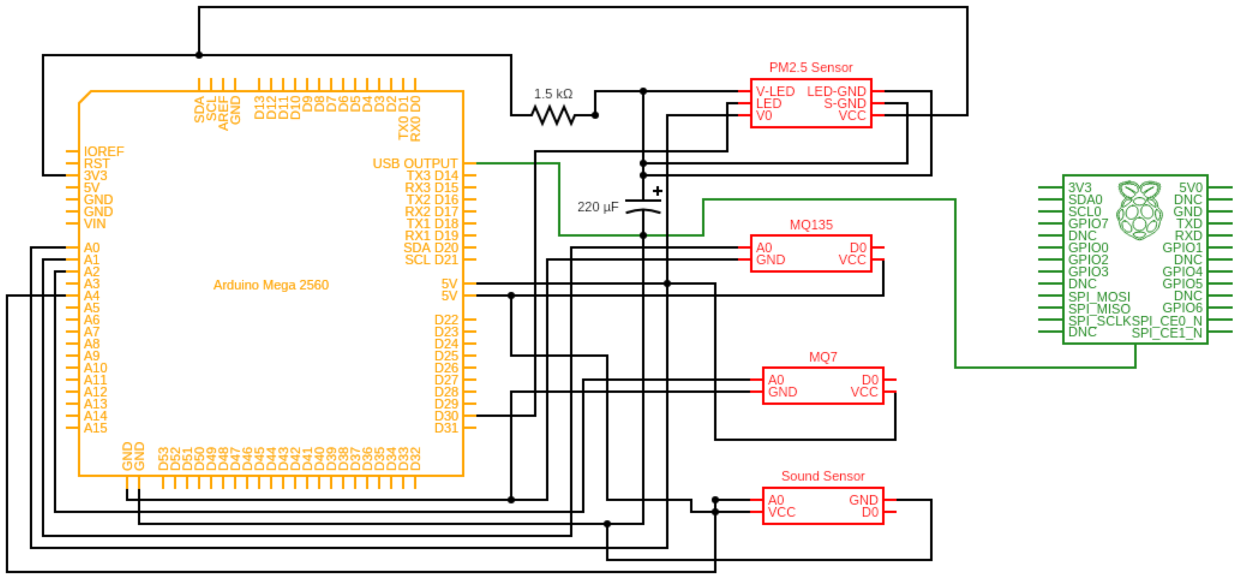


Figure 4.10: Schematic Diagram

After connecting PM2.5 and MQ135, we have connected the MQ7 sensor. It has 4 pins. The A0 of the sensor goes to the A2 pin of Arduino mega. Then VCC goes to the 5V pin of the Arduino after that we have connected the GND pin to the GND of the Arduino. Then, we integrated Sound Sensor to the Arduino. Firstly, the A0 pin of the sensor goes to the A4 pin of the Arduino Mega and VCC goes to the 5v of the Arduino and GND goes to the GND pin of Arduino. Then, we have connected Arduino to the Raspberry Pi as raspberry pi has no analog pin and we need the analog value for our research purpose that is why we have used Arduino but, we have to send the data to the server and that is the reason we have used Raspberry Pi. We have connected Arduino USB port to Raspberry Pi USB port and then we have done our server-side programming to send the data to server.

Chapter 5

Machine Learning Algorithms

In this section we will explain about the algorithms we have used in our system. To monitor, visualize the information, classification and prediction we have mainly used three algorithms. The algorithms we are using are PCA (Principal Component Analysis), LSTM (Long Short Term Memory) and Random Forest algorithm. Detailed explanation about the algorithms and how this algorithms will be used in our system are explained below.

5.1 Principal Component Analysis (PCA)

PCA or Principal Component Analysis is a technique of dimensionality reduction which uses orthogonal transformation to reduce a large set of variables to a smaller set of variables while retaining most of the present details. The method works by transforming the original dataset's strongly correlated variables into smaller uncorrelated linear variables, known as main components [5] [1]. Then, these key components account for much of the original database variance PCA is very useful when the data has 3 or higher dimensions because it is incredibly difficult to make predictions from such a large amount of knowledge. In addition, high-dimensional data visualization is often challenging. PCA also solves this data visualization problem[11].

5.1.1 Working Mechanism of Principal Component Analysis (PCA)

Several steps are followed in order to find Principal Component from a dataset. After completing these steps the Principal Components of the dataset can be figured out[7]. The steps which are followed to find the principal components are described below:

1. 1. We have to get data for finding PCA and we have to find the mean of the attribute of each dataset.

2. 2. We have to subtract the mean from the value of each data from the attribute and we have to multiply that and we also have to find the mean of that attribute.
3. 3. We have to calculate the covariance matrix from the result of previous step. The formula of finding the (x,y) element of the covariance matrix is:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

4. We have to find the eigenvectors and eigenvalue from the covariance matrix and we have to find the highest eigenvalue.
5.5. Finally we only have to consider the maximum eigenvalue and determine feature vector for that value.

5.1.2 Implementation of PCA in System

We will mainly use PCA in our project for finding the maximum level of sound which is harmful for human health. Among all the sounds in the environment PCA will detect the sound level which is harmful for human health and will show in the interface by a graphical plot of the presence of harmful sound level present in the environment. PCA will also be used to find the element which is causing maximum pollution currently at that time. It will also be showed by a graphical representation. So that, specific measures can be taken to prevent that element from doing harm in the environment.

5.2 Long Short Term Memory (LSTM)

Neural networks are cellular structures that take input vectors and process the inputs within the hidden layers using activation functions and different gating techniques in a series of propagation steps from one cell state to another until the vector output is produced[24]. Neural networks typically include RNN, but it has the drawback that due to its smaller memory, it cannot relay data from previous time steps which are considerably wide. LSTM is an enhanced and extended RNN variant that can cure this problem[14] [35]. LSTM is good for non-linear data processing because it employs several non-linear activation functions such as sigmoid, hyperbolic tangent. Generally speaking, neural networks perform better and have higher precision, but certain regions overlapping with other models indicate that they do not always outperform other models [30].

5.2.1 Working Mechanism of LSTM

A typical LSTM unit consists of a cell, an input gate, a gate of output and a gate of forgetting [37]. The cell recalls values over arbitrary periods of time, and the three gates control information flows into and out of the cell.
C_{t-1} stands for the input from a memory cell in time point t, x_t is an input in time point t an h_t is an output in time point t that goes to both the output layer and the hidden layer in the next time point. The following steps are required to run LSTM successfully:

1. We have new value x_t and value coming in from previous h_{t-1} node.

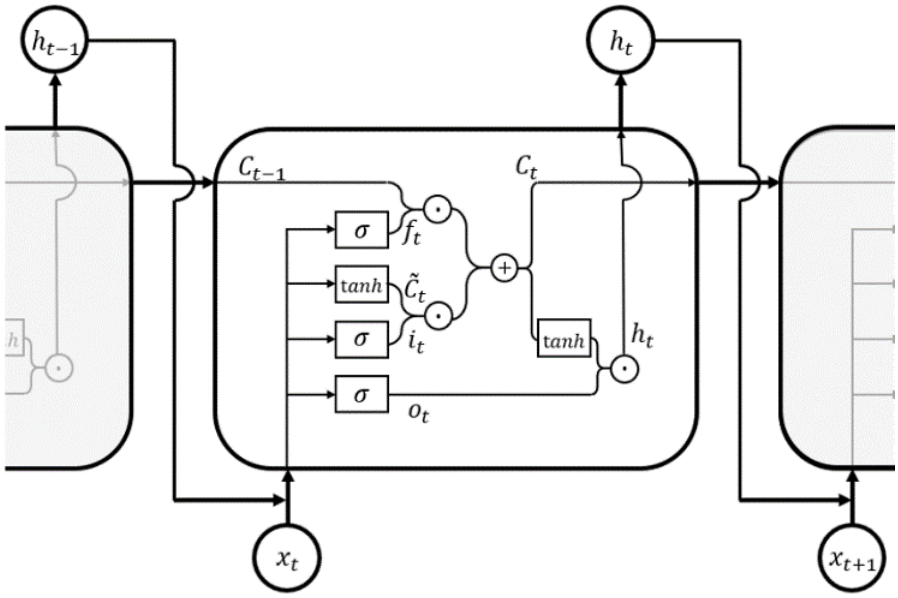


Figure 5.1: LSTM cell structure

2. These values are combined and go through the sigmoid activation function, where it is decided whether to some extent the forget valve should be open, closed, or open.

3. The same values, or in fact vectors of values, go in parallel through another "tanh" layer operation, where it is decided what value we are going to transfer to the memory pipeline, and also sigmoid layer operation, where it is decided if that value is going to be passed to the memory pipeline and to what degree.

4. Then, we got a memory that flows through the top pipeline. When we forget to open the valve and shut the memory valve, the memory won't change. Otherwise, the memory would be fully restored if we forget valve closed and memory valve open.

5. Finally, we have x_t and h_{t-1} combined to determine which part of the memory pipeline will become this module's output.

5.2.2 Implementation of LSTM in System

We will use LSTM for predicting the value of PM 2.5 in the air for the next hour by observing the value of previous 24 hour or 48 hours. It will make prediction by observing the previous values and factors for the PM 2.5 value and make the prediction. By this algorithm we will be able to determine the level of PM 2.5 present in the air for the next few hours so that the people who are sensitive to air pollution can avoid that area and measures can be taken to reduce the value of PM 2.5 in the area. We will also use this algorithm for predicting the level of sound pollution of next hour by analyzing the previous data so that in can be reduced.

5.3 Random Forest

Random forest is associate degree ensemble learning methodology for classification, regression and alternative tasks that works by making a large number of call trees at the time of coaching and generating the category that's the category mode (classification) or mean prediction (regression) of the individual trees[16] [21]. It corrects the habit of over fitting to their coaching set.

5.3.1 Working Mechanism of Random Forest

Random Forest algorithm is implemented by following some steps which are shown below:

1. 1. At first we have to bootstrap for random subset of the data. It means that we have to select few datasets from the entire data sets. The data can be repeated but it is recommended to not repeat too much data.

2. 2. We have to the find the GINI impurity from the bootstrapped dataset. We do not have to take the whole dataset. Instead, we have to take two or three attribute at a time of the data set and then find the Gini impurity. The process of finding the Gini impurity is: The GINI impurity can be determined by summing up the probability pi of an item with a mark I being picked times the likelihood of an error in categorizing the item [34]. The formula is:

$$\mathrm{I}_G(p) = \sum_{i=1}^{J} p_i \sum_{k \neq i} p_k = \sum_{i=1}^{J} p_i (1 - p_i) = \sum_{i=1}^{J} (p_i - {p_i}^2) = \sum_{i=1}^{J} p_i - \sum_{i=1}^{J} {p_i}^2 = 1 - \sum_{i=1}^{J} {p_i}^2$$

Which number gets the lowest value of the Gini impurity that will be the root node. This process will continue for other attributes as well and it will make multiple decision trees
3. N3. Now, for the bagging process we will compare our current data to the decision trees generated in the 2nd step and will see what the result is.

4. In the bagging process if the result of majority our data comparing the decision tree is positive then the answer is positive or it is negative.

5.3.2 Implementation of Random Forest in System

In our project for air pollution monitoring we will take multiple elements which cause air pollution. Such as SO_2, CO_2, CO, N_2O and P.M 2.5. Our result attribute will be the AQI (Air Quality Index) and it will indicate the level of pollution in the air by the presence of this element in the air.

By the Ideal dataset we will create the decision trees by finding the lowest gini impurity values comparing to the attributes[32].
After this we will do the bagging process for real time data which we will acquire from our sensor and we can see from the comparison if the data is polluted or not

by the value of AQI.
Finally, we can determine the quality of air by seeing the result of bagging process.

5.4 Flow Chart of Algorithms

As discussed in the earlier sections those were the algorithms we will use in our systems and each algorithms will do different work for the system which will lead to different results. In the following algorithms we have tried to show how the different algorithms will work and give us different output. The flow chart is in figure 5.2.

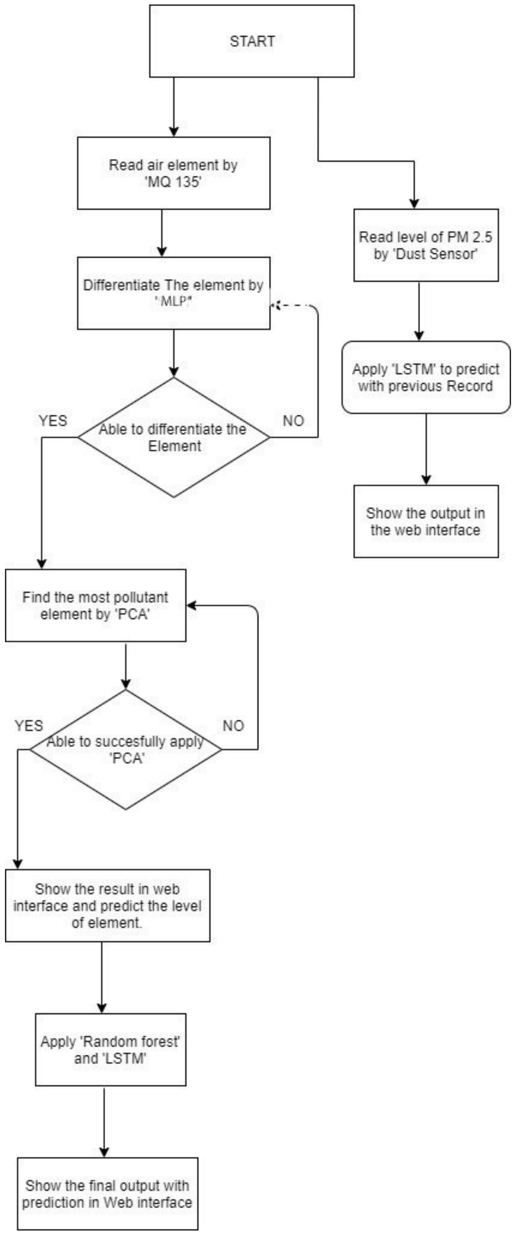


Figure 5.2: Implementation of Algorithms

In Figure 5.2 we have tried to show, how our algorithms is running sequentially and the analysis and prediction of the level of air elements are being done. At first by using 'MQ-135' sensor we can collect the reading of different elements of air. By using 'MLP' the different compounds will be separated. After the separation we will apply 'PCA' to the separated data to know the compound is causing the most pollution at that present time. We show the result of the analysis up to that point in the web interface. We apply 'LSTM' and 'Random Forest' to predict the pollution level for later time in future in our web interface too. With the help of dust sensor we will be able to read the level of PM 2.5 and we will apply only 'LSTM' to do time series analysis in the reading of PM 2.5 level and to predict the level for future and show it in the web interface to aware people.

Chapter 6

Implementation

6.1 Web Programming

Front-End and Back-End are two main parts of internet programming. Front end programming is about the design section of a web page where all of the functionality functions and the website layout is properly handled. Front end is all about the user interface and what is available to the user. In reality, the content design of the website appears in the front website here, which demonstrates how the web page will look properly. There is all the terminology on the back of the website that is meant to be used for programming to make a website effective and useful. Back-end programming is the web page layout, which is the essence of a web page. At first, the programming had to be counted on both sides. We had to find the front end portion at the very beginning, and then back end portion for the webpage.

Figure 6.1: Two way programming deployment

6.1.1 Implementation on Languages

HTML and then PHP in addition to CSS and JavaScript are two main languages to take initiative in implementing the back-end software. HTML is the standard language used to build web pages which stands for "Hypertext Markup Language." The web page framework essentially stands with the help of HTML. HTML elements form blocks and are represented by tags. The CSS component is used for website design and outlook accessible to a customer. CSS (Cascading Style Sheets) explains exactly how the 30 html elements on the screen are represented. CSS describes web page look and style. The browser receives web server HTML documents and then sends them to web sites. The building blocks of Web pages are Web elements. HTML provides semicolon for testing such as chart, table, image and diagram. HTML can be written in JavaScript, a scripting language. JavaScript is a client side script that is executed before the webpage is visible after the server receives the returned information. One is known as PHP server-side script. Embedding PHP code into HTML code.

6.1.2 Server side scripting

Server-side programming is the term given to all types of programs operating on the webserver. We handle user feedback, communicate with databases, and monitor what information is being delivered out to the client as a response to the client's request. Mainly server-side programming is used in web development, android app application, etc. and server-side scripting is used in dynamic pages rather than a static page and that's why it is powerful. Server-side scripting can build APIs (Application Programming Interfaces) and users want to show something it can retrieve the data from the database and through JSON (Java Script Object Notation) format and show it to the browser.

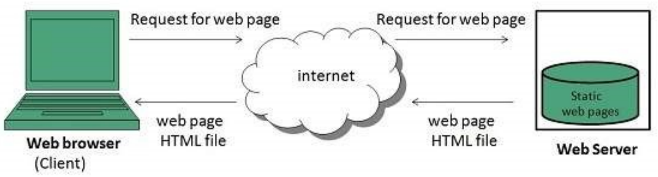


Figure 6.2: Server Side Programming

There are many popular server-side scripting languages for example, PHP, Java, Python, Ruby, C, C++, etc. and some framework which makes our work easier and more efficient. The popular framework of these is Laravel,, ASP.NET, Ruby on Rails, Django. We have NodeJs to do server-side scripting it is a run time environment of JavaScript. Now-a-days popular server-side scripting language is PHP and its framework Laravel. These are written on SQL language and keeps a connection with the database. We can store the data or retrieve the by different SQL queries with these server-side scripting languages..

6.1.3 Client-Side scripting

Client-side scripting is a part of front end development and it is run on the client computer (browser) and it deals its user interface, read and writing cookies and it also interacts with interactive web pages with temporary storage, submit data request to server, function as an interface between server and user. Application-side scripting makes a website user-friendlier and a website sensitive.

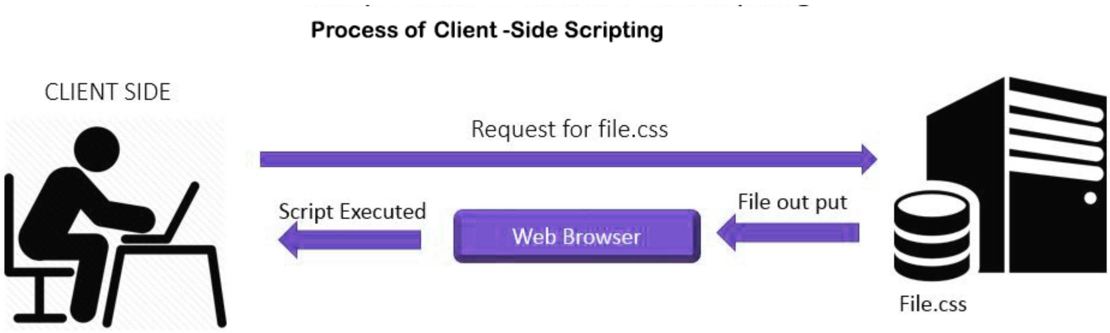


Figure 6.3: Client Side Programming

The popular client-side scripting language is HTML, CSS and JavaScript, Ajax. With HTML and CSS we can make the website more attractive and it makes a website responsive and also user friendly. Java script enables interactive web pages and it has application programming interfaces (API) for working with text, date and a regular expression and it has standard data structures and document object model (DOM). In our research project we make a front-end design with HTML, CSS and JavaScript it shows us the value of different sensors.

6.2 User Interface

User interface mainly focus on the data table and graph according to the table data. The main focus of the webpage is representing real time data of each sensor and the value of each gas in ppm unit so that user can easily understand about harmful pollutant. Besides, the pie chart shows some source from where we monitor air pollution like transport fuel, emission from industries, domestic cooking etc.

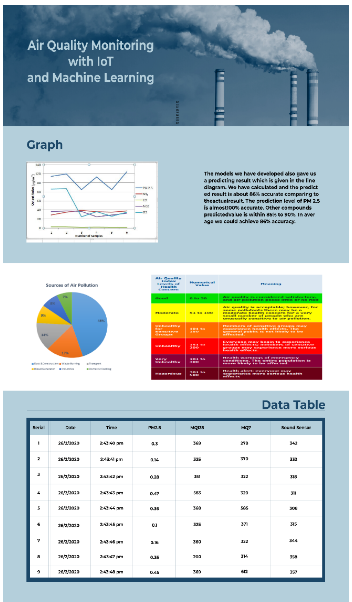


Figure 6.4: User Interface of Proposed System

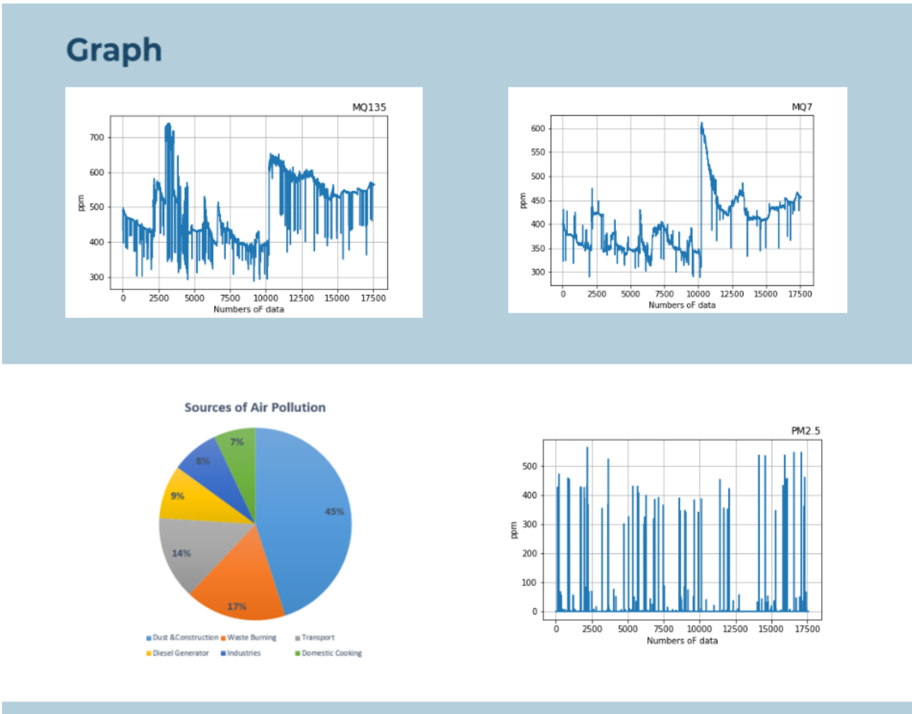
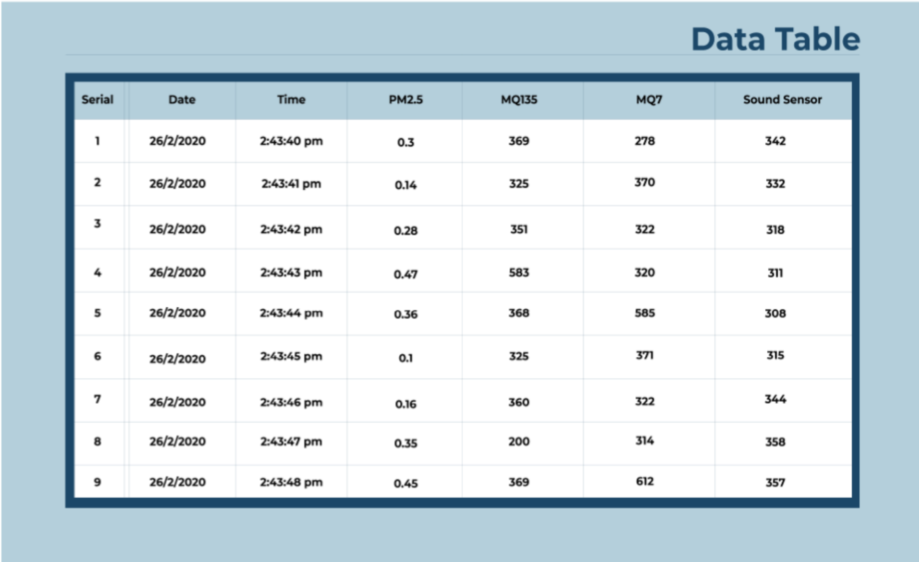


Figure 6.5: Pie chart and graph of sensor data

Above figure is representing graph of specific three sensors data which are MQ135, MQ7, and PM2.5 respectively. In every graph, numbers of data represent x-axis and ppm (Parts per Million) represent y-axis. From the close view of the graph, we can see the change of ppm value depending on the number of data changing. For MQ135 graph, when number of data is between 2500 to 5000 then the value is above 700ppm which is highest. Similarly, for MQ7 highest value exist in range 10000 to 12500 which is more than 600ppm and in this position ppm increased from lowest peak to highest peak. Lastly the pie chart represents the sources of air pollution percentage wise. Most of the pollutants are released by industries and factories, in percentage 45% which is almost half. Then transport and waste burning are almost same in number. And only 7% pollutant released from domestic cooking.



Data Table

Serial	Date	Time	PM2.5	MQ135	MQ7	Sound Sensor
1	26/2/2020	2:43:40 pm	0.3	369	278	342
2	26/2/2020	2:43:41 pm	0.14	325	370	332
3	26/2/2020	2:43:42 pm	0.28	351	322	318
4	26/2/2020	2:43:43 pm	0.47	583	320	311
5	26/2/2020	2:43:44 pm	0.36	368	585	308
6	26/2/2020	2:43:45 pm	0.1	325	371	315
7	26/2/2020	2:43:46 pm	0.16	360	322	344
8	26/2/2020	2:43:47 pm	0.35	200	314	358
9	26/2/2020	2:43:48 pm	0.45	369	612	357

Figure 6.6: Table containing sensor data in User Interface

Data table is representing the real time sensor data retrieving from the firebase database. Actually we use four sensors to analyze the environment polluted or not. PM2.5 sensor can measure particulate matter and represents that value in table according to time and date. MQ135 gas sensor can sense Sulfide gas, Carbon dioxide (CO_2), Benzes steam, Ammonia (NH_3) and also sensitive to smoke and other dangerous gases. On the other hand, MQ7 gas sensor has high sensitivity to detect Carbon monoxide (CO) gas. Actually the numerical value of each sensor is shown in data table.

Chapter 7

Experimental Evaluations

We have conducted experimental analysis in various areas of 'BRAC University' campus. The summary of the reading and analysis of the data are given below in Figure 7.1, 7.2 and 7.3. We have placed our device in various areas of our university area to collect sample data to test our system. It is a bit concern that almost in every area the concentration level of PM 2.5 is slightly higher than other compounds. In Table 7.1 sample reading of the sensors are given below. We have run our device for about 6 hours to get enough dataset for getting accurate result by machine learning algorithm.

Table 7.1: Sample Gathered Data for Air Pollution and Particulate Matter Detector Testing

Date	Time	PM2.5 (µg/m3)	MQ135 (PPM)	MQ7 (PPM)
2/26/2020	2:43:40 PM	0.3	369	278
2/26/2020	2:43:41 PM	0.14	325	370
2/26/2020	2:43:42 PM	0.28	351	322
2/26/2020	2:43:43 PM	0.47	583	320
2/26/2020	2:43:44 PM	0.36	368	585
2/26/2020	2:43:45 PM	0.1	325	371
2/26/2020	2:43:46 PM	0.16	360	322
2/26/2020	2:43:47 PM	0.35	200	314
2/26/2020	2:43:48 PM	0.45	369	612

In our proposed system the users will be able to see plot about the data gathered by our sensors for a fixed amount of time and it will give the users idea about the current environment pollution. The plot of the data taken by our all sensors are shown below.

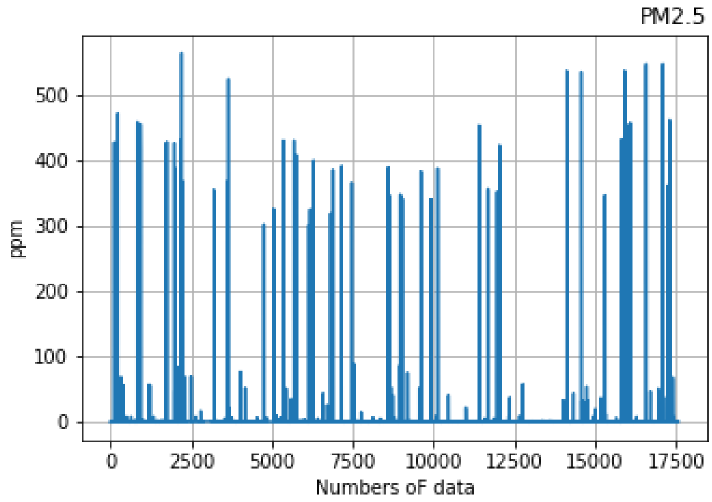


Figure 7.1: Plot of dust sensor reading

In the X axis number of data is being shown and in the Y axis the amount of pollution is being shown to give an idea to the user about the behavior of the pollution in recent times. In the same way other two sensor's data are also plotted.

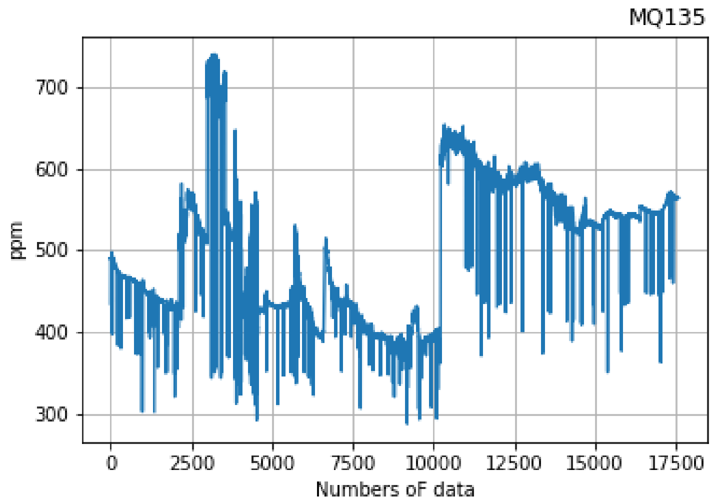


Figure 7.2: Plot of 'MQ135' sensor data

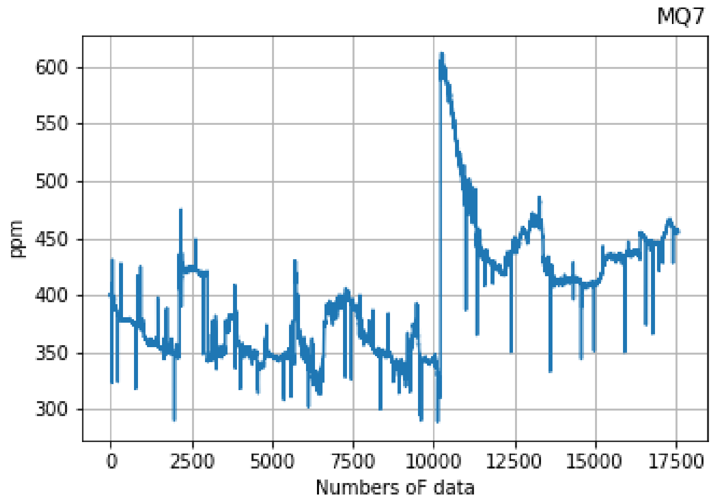


Figure 7.3: Plot of 'MQ7' sensor data

Besides these, the output of individual algorithms are also available in the system for the users to see. As mentioned earlier we are identifying the most pollutant element of the air by applying 'PCA'. A visual graph of only the most two pollutant components of the entire dataset is also available in the system and it will be updated time to time. An example of the output of PCA algorithm is given below. The value of the first principal component will be in the X axis and the value the second principal component will be in the Y axis of the graph. By using PCA algorithm we have reduced the dimension of our data set into 2D. Because of this in figure 7.4 we can easily differentiate between the qualities of air easily. The quality is divided into three conditions which are 'Fresh', 'Moderate' and 'Polluted'.

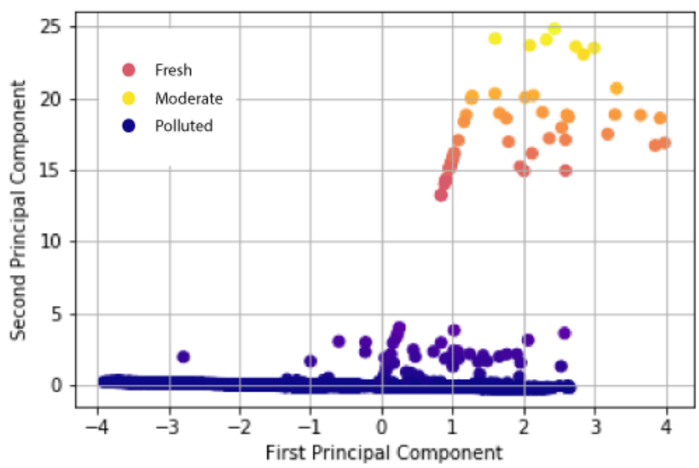


Figure 7.4: Output graph of 'PCA'

For prediction and time series analysis 'LSTM' is being used in our system. A graph is available to show train and test Loss of our dataset from the multivariate LSTM training. In the graph plot we can see that the 'Loss' is very minimal between our test and train dataset. Which means accuracy will be higher. A plot of the information is attached below in figure 7.5.

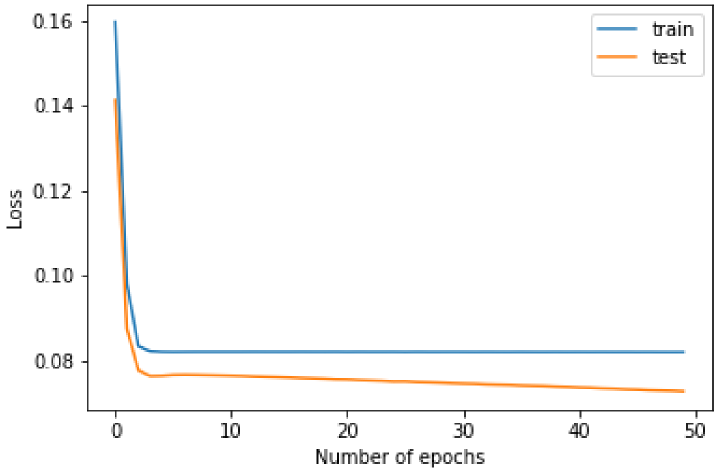


Figure 7.5: Line Plot of Train and Test Loss from the Multivariate LSTM during Training

The X axis is indicating the number of 'Epochs' of our train and test dataset and the Y axis is indicating the value of loss function. We can see that the loss is decreasing as the epochs are increasing. By applying 'Random Forest' algorithm in our dataset we achieved 88% precision and 86% f-1 score in average in the dataset we have used recently. It can be varied by the character of the dataset as the data is very dynamic. A table is attached below of the result we have got in Table 2.

Table 7.2 Accuracy rate of the proposed model

	Precision	Recall	F1- score	Support
Polluted	0.95	0.86	0.90	22
Not Polluted	0.40	0.67	0.50	03
Avg/Total	0.88	0.84	0.86	25

Finally, a summary of the total concentration of the element of our sensors are able to get will be showed in a graphical plot in Fig 7.6. Which will be useful for the users to understand the current situation of the quality of the air and everyone will be able to understand it just by looking into it. The prediction value of the elements is also being showed in another graph for the next day for users' convenience.

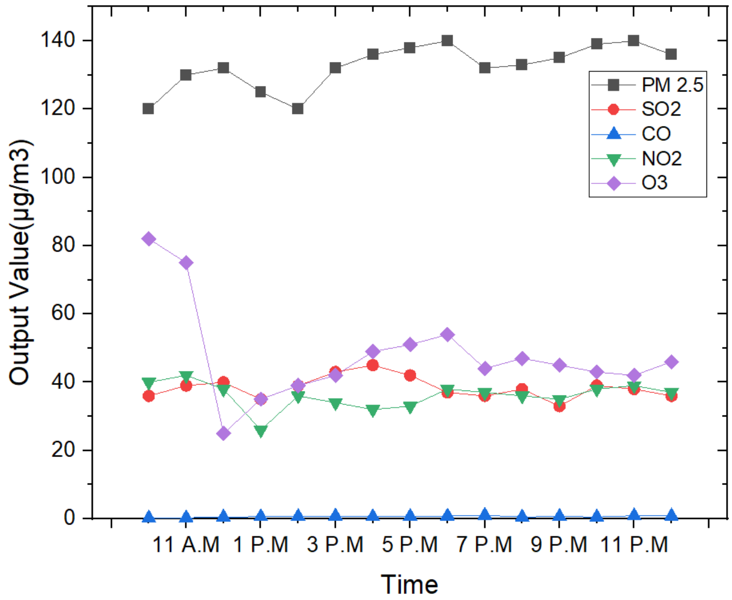


Figure 7.6: Reading from the sensors

The reading of the sensors is in microgram/cubic meter. In both graphs the X axis is indicating the time and the Y axis is indicating the output value of our sensors. The models we have developed also gave us a predicting result which is given in Fig 7.7. We have calculated and the predicted result is about 88% accurate comparing to the actual result.

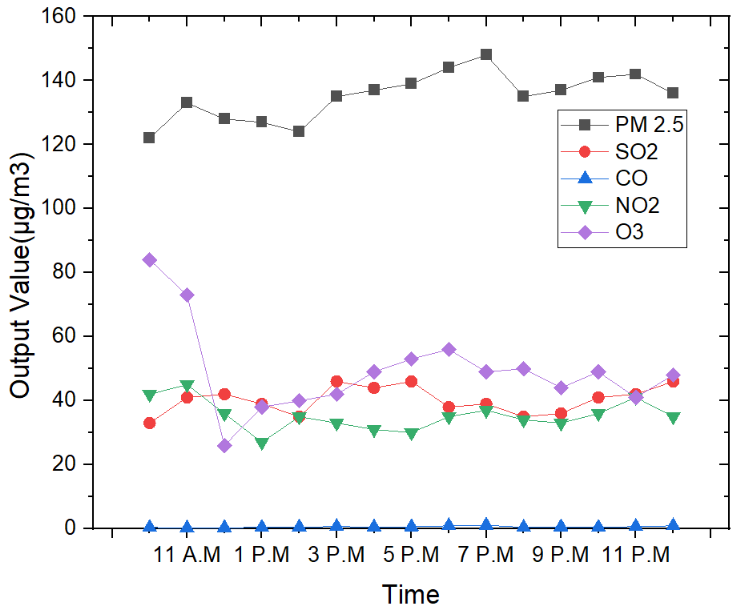


Figure 7.7: Predicted result of the model

Chapter 8

Conclusion and Future Works

8.1 Conclusion

IoT and machine learning algorithms have begun to play a very crucial role in our life. Our main goal and objective is to build a cheap and simple IoT device which can be used in everywhere to monitor the environment for our own safety and betterment. There is no other way for us to solve our problem without machine learning algorithms now a day. Recently Dhaka was announced to be the most polluted city in terms of air by the global organizations like WHO and UNDP has also showed their concern about Dhaka.
We want to grow more awareness among people that how polluted our environment is getting only because of us. We hope that through this device we will be able to contribute to our society for a better future.

8.2 Future Work

There is a huge scope for development in our work in future. We can optimize the device with more sensors and if we can use more expensive and good sensors we will be able to get more accurate result in the reading and the algorithms that we have used will give us more accuracy. There is a huge scope for development in the website too. We can certainly add more options in the user interface and we can build the website more dynamic at the same time. We will build a mobile app for the users for more convenience in the future. Alongside this, we have thought of another device which will be connected with these device and it will play a big role for decrease the pollution dynamically. We will optimize the algorithm further for more efficient and more accuracy and there is a scope for using deep learning in our device to get almost full accurate result in the future.

References

[1] L. Wei-min and C. Chein-I, "Variants of principal components analysis", in *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2007, pp. 1083–1086.

[2] L. Chuang, "Enlightenment of environmental regulation policy and practice of developed countries", in *2009 International Conference on Environmental Science and Information Application Technology*, vol. 3, 2009, pp. 79–82.

[3] T. Ayodele, "Introduction to machine learning", in. Feb. 2010, ISBN: 978-953-307-034-6. DOI: 10.5772/9394.

[4] H. Zhou, Y. Li, and Z. Cui, "Distribution characteristics of air ions and their relationship with environmental factors in xi'an", in *2011 International Symposium on Water Resource and Environmental Protection*, vol. 3, 2011, pp. 2185–2188.

[5] K. Rujirakul, C. So-In, B. Arnonkijpanich, K. Sunat, and S. Poolsanguan, "Pfp-pca: Parallel fixed point pca face recognition", in *2013 4th International Conference on Intelligent Systems, Modelling and Simulation*, 2013, pp. 409–414.

[6] K. Hu, T. Davison, A. Rahman, and V. Sivaraman, "Air pollution exposure estimation and finding association with human activity using wearable sensor network", Dec. 2014. DOI: 10.1145/2689746.2689749.

[7] E. C. Ozan, S. Kiranyaz, and M. Gabbouj, "M-pca binary embedding for approximate nearest neighbor search", in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 2, 2015, pp. 1–5.

[8] C. Xiaojun, L. Xianpeng, and X. Peng, "Iot-based air pollution monitoring and forecasting system", in *2015 International Conference on Computer and Computational Sciences (ICCCS)*, 2015, pp. 257–260.

[9] K. Bashir Shaban, A. Kadri, and E. Rezk, "Urban air pollution monitoring system with forecasting models", *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2598–2606, 2016.

[10] S. Karamchandani, A. Gonsalves, and D. Gupta, "Pervasive monitoring of carbon monoxide and methane using air quality prediction", in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 2498–2502.

[11] B. F. Almubarak, Y. I. Aleisa, B. Liu, and A. H. Muqaibel, "Global pca for in-field compression of seismic data acquisition", in *2017 9th IEEE-GCC Conference and Exhibition (GCCCE)*, 2017, pp. 1–4.

[12] R. R. Appannagari, "Environmental pollution causes and consequences: A study", vol. 3, Aug. 2017.

[13] B. Harish Kumar, "Wsn based automatic irrigation and security system using raspberry pi board", in *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, 2017, pp. 1097–1103.

[14] Y. Heryadi and H. L. H. S. Warnars, "Learning temporal representation of transaction amount for fraudulent transaction recognition using cnn, stacked lstm, and cnn-lstm", in *2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, 2017, pp. 84–89.

[15] S. Kumar and A. Jasuja, "Air quality monitoring system based on iot using raspberry pi", May 2017, pp. 1341–1346. DOI: 10.1109/CCAA.2017.8230005.

[16] P. N. Samarakoon, E. Promayon, and C. Fouard, "Light random regression forests for automatic multi-organ localization in ct images", in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 371–374.

[17] N. S. Yamanoor and S. Yamanoor, "High quality, low cost education with the raspberry pi", in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, 2017, pp. 1–5.

[18] E. Ahmed, "Environmental pollution", *Agricultural Research Technology: Open Access Journal*, vol. 18, Oct. 2018. DOI: 10.19080/ARTOAJ.2018.18.556049.

[19] Y. Alsouda, S. Pllana, and A. Kurti, "A machine learning driven iot solution for noise classification in smart cities", Aug. 2018.

[20] T. M. Amado and J. C. Dela Cruz, "Development of machine learning-based predictive models for air quality monitoring and characterization", in *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2018, pp. 0668–0672.

[21] X. Fan, Z. Feng, X. Yang, T. Xu, J. Tian, and N. Lv, "Haze weather recognition based on multiple features and random forest", in *2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, 2018, pp. 485–488.

[22] V. Jatana, *Machine Learning For Beginners*. Oct. 2018.

[23] J. Kaminska, "The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: A case study in wrocław", *Journal of environmental management*, vol. 217, pp. 164–174, Mar. 2018. DOI: 10.1016/j.jenvman.2018.03.094.

[24] A. H. Mirza and S. Cosan, "Computer network intrusion detection using sequential lstm neural networks autoencoders", in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 2018, pp. 1–4.

[25] A. Mishra, "Air pollution monitoring system based on iot: Forecasting and predictive modeling using machine learning", Oct. 2018.

[26] A. Saha, S. Sircar, P. Chatterjee, S. Dutta, A. Mitra, A. Chatterjee, S. Chattopadhyay, and H. Saha, "A raspberry pi controlled cloud based air and sound pollution monitoring system with temperature and humidity sensing", Jan. 2018, pp. 607–611. DOI: 10.1109/CCWC.2018.8301660.

[27] F. Salih and M. S. A. Omer, "Raspberry pi as a video server", in *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, 2018, pp. 1–4.

[28] X. Teng and Y. Gong, "Research on application of machine learning in data mining", *IOP Conference Series: Materials Science and Engineering*, vol. 392, p. 062 202, Aug. 2018. DOI: 10.1088/1757-899X/392/6/062202.

[29] Y. Tsai, Y. Zeng, and Y. Chang, "Air pollution forecasting using rnn with lstm", in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2018, pp. 1074–1079.

[30] W. Yao, P. Huang, and Z. Jia, "Multidimensional lstm networks to predict wind speed", in *2018 37th Chinese Control Conference (CCC)*, 2018, pp. 7493–7497.

[31] Y. Zhao, L. Fu, and L. Wang, "Design of pm2.5 monitoring system under the human micro-environment", in *2018 IEEE 27th International Symposium on Industrial Electronics (ISIE)*, 2018, pp. 218–223.

[32] A. Behnamian, S. Banks, L. White, K. Millard, D. Pouliot, J. Pasher, and J. Duffe, "Dimensionality reduction in the presence of highly correlated variables for random forests: Wetland case study", in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 9839–9842.

[33] J. Gebicki and B. Szulczynski, "Remote monitoring of environmental pollutants", in. Aug. 2019, pp. 325–352, ISBN: 978-981-13-9104-0. DOI: 10.1007/978-981-13-9105-7_11.

[34] V. Jain and A. Phophalia, "Exponential weighted random forest for hyperspectral image classification", in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 3297–3300.

[35] S. Poh, Y. Tan, X. Guo, S. Cheong, C. Ooi, and W. Tan, "Lstm and hmm comparison for home activity anomaly detection", in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 2019, pp. 1564–1568.

[36] P. Sebastian, C. Romeo, and G.-A. Ursan, "Environmental pollution control", Oct. 2019, pp. 1–5. DOI: 10.1109/SIELMEN.2019.8905843.

[37] X. Song, J. Huang, and D. Song, "Air quality prediction based on lstm-kalman model", in *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2019, pp. 695–699.

[38] L. Bláha and J. Hofman, "Ecotoxicology of environmental pollutants", in. Jan. 2020, pp. 549–572, ISBN: 978-3-030-29839-5. DOI: 10.1007/978-3-030-29840-1_27.