Cloud Driven IoT based Big Data Solution: Air Quality Monitoring System and Predictive Analysis of Satellite Data using Machine Learning

By

Noor-E Sadman
Student ID : **1730008**

Spring, 2021

Supervisor :
Dr. Mahady Hasan
Associate Professor, Department Head
Department of Computer Science & Engineering
Independent University, Bangladesh

September 30, 2021

Dissertation submitted in partial fulfillment for the degree of Bachelor of Science in Computer Science

Department of Computer Science & Engineering
Independent University, Bangladesh

Attestation

I, Noor-E Sadman (ID :1730008), certify that the report is completed by me and submitted in partial fulfillment of the requirement for the Degree of Computer Science and Engineering from Independent University, Bangladesh (IUB). This project was completed under the supervision of Dr. Mahady Hasan sir within a time period of one year. All the sources of information used in this project are verified and the report has been duly acknowledged in it.

Noor-E Sadman

Student Name

1730008

Student ID

18/11/2021

Date

Declaration of Authorship

I , Noor-E Sadman, hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

Signed: ……

Date: ……

Acknowledgement

I acknowledge my profound indebtedness and express sincere gratitude to my senior project supervisor Dr. Mahady Hasan, Head of Department, Department of Computer Science & Engineering, School of Engineering, Technology & Sciences, Independent University, Bangladesh - who not only showered us with guidance, supervision, and valuable suggestions at all stages to carry out the project, but also provided us with a fully funded thesis. Without his inspiration and kind support the study could not be carried out. I am thankful to him for giving me the opportunity to work under his supervision and am proud to have him as my supervisor for the senior project .

I also would like to express my heartfelt gratitude to FabLab IUB for creating a space with qualitiful research opportunities and providing me with all the equipment, facility and support that was needed during the process of execution of the project.

My utmost appreciation to all the developers who built libraries for the sensors used in this project. Without their contribution my project would have been incomplete.

Finally, earnest gratitude to the Almighty, Allah (S.W.T), for giving me good health for which I could complete this report in due time. Also, my family members and friends, who supported me with ideas and suggestions.

Table of Contents

Abstract

The main objective of this project is to eliminate the climate change crisis around the world, by fulfilling the Sustainable Development Goal 13 which is Climate Action. The data we are extracting from this project about polluting compounds will be an extended research, which we believe will help us to find a solution to this worldwide crisis and spread awareness globally.

For the time being, we are proposing two solutions to this global problem. We propose to extract data from a primary source which we aspire to construct using IoT and a secondary source which is AQUA Satellite. For the primary source, we will demonstrate an IoT based device and compare it with the expensive industrial graded air monitoring device, AirVisual Pro by IQAir, to determine device compatibility and try to find an inexpensive solution to the problem. Since we have to make sure to measure the air quality of all the regions of Bangladesh, the inexpensive device that we will be making will no longer be cost efficient since we need to fill the infrastructure gap. To solve this problem, a successful approach has been noted from multiple research where the AOD PM2.5 product is predicted using Machine Learning Algorithms (MLAs). In the process of making an atmospheric forecast reporting system, we intend to incorporate station wise graphs and using MLAs our aim is to generate an 'AQI alert atmospheric map',with the help of the satellite data. Additionally, our project will extensively monitor the most polluting areas/routes - to give an alert when the pollution exceeds the favourable amount.

For the primary data source, our project aims to distinguish the particle concentration of PM1.0, PM2.5, PM10, CO, NO_2, CO_2 and VOCs like (CH_4,C_3H_8,C_4H_{10},C_2H_5OH). For detecting PM1.0, PM2.5 and PM10, PMS5003 PM2.5 Air Particle Dust Sensor is used and the $NO_2$and VOCs like (CH_4,C_3H_8, C_4H_{10},C_2H_5OH) compounds are measured by Grove Multichannel Gas Sensor. Since increment of concentration of CO is becoming a prime concern in Bangladesh, we have used a sensor called MQ 7 for the detection of CO - for more accuracy of the data. Similarly, for CO_2concentration observation, MH-Z19B Co2 Sensor Module is chosen to work with. Not to keep our system limited to only to the extraction of polluting compound data, we used the BME280 module to keep track of the temperature, humidity and barometric pressure. Since this is an IoT based air monitoring system, the user will be able to get all the information regarding the air quality.

In order to make the IoT based air monitoring system, Arduino Mega 2560 Wifi - which is a development board- is used as the microcontroller of the system and incorporated with the set of sensors to detect the concentration of pollutants and monitor the data patterns. It is turned to an IoT based device after the set of sensors are connected to the development board. Arduino Mega 2560 WiFi takes readings from the sensors and sends all the sensor data to the PC using Serial Communication. This data is then migrated to the cloud. All the concentration of polluting compound charts will be shown in a Google Sheet visualization.

The secondary AOD(Aerosol Optical Depth) 550nm data is aimed to be extracted from AQUA Satellite that will be later collected using NASA Giovanni data visualization platform. The available

AOD 550 nm product needs to be preprocessed from MODIS (Moderate Resolution Imaging Spectroradiometer) instrument aboard the Aqua Satellite, which passes from South to North poles of the earth. The selected shape of the data is Bangladesh, which means the obtained AOD 550 nm product will be collected and preprocessed by the AQUA satellite using the MODIS instrument while orbiting above Bangladesh. The obtained AOD 550 nm is the mean AOD 550 nm for each day processed by the AQUA MODIS.

Chapter 1 : Introduction

1.1 Overview

The atmosphere is a complicated system, and numerous elements, including air particles, solid and liquid molecules, influence the extinction effect. Air pollution is a worldwide crisis with limited solutions that is caused because of the presence of compounds in the atmosphere that are detrimental to the wellness of habitants, or pose a hazard to the ecosystem or objects. Starting from long term damage to the built environment to exposure of buildings to pollutants causing degradation, changes in soiling patterns due to rain impingement and corrosion metal and glass was discovered (Brimblecombe, n.d., #).

Chemical composition, reactive characteristics, emission, persistence in the environment, and capacity to be transmitted over long or short distances have all been found to differ across air pollutants. However, variants of polluting compounds with similar properties have been divided into four categories (Castanas & Kampa, 2007, #). The main anthropogenic sources are as follows :

1. Gaseous contaminants.
   Compounds like SO2, NOx, CO, ozone, Volatile Organic Compounds are the prime forms of air pollutants, which form due to combustion of fossil fuels, leakage of chemicals into the atmosphere in the making process of advanced technologies.
2. Persistent organic pollutants.
   Atmospheric toxins like dioxins are emitted in the atmosphere during incomplete combustion and whenever materials containing chlorine (e.g. plastics) are burned. Heavy metals like lead and mercury are natural components of the earth’s crust, thus they cannot be degraded or destroyed and can only be transported by air - infiltrating water and human food supply. Furthermore, toxins affect the atmosphere from a range of sources, such as combustion, waste water discharges and production plants.
3. Particulate Matter.
   This is the vital reason for cardiovascular, respiratory diseases leading to damaging of the nervous system since it is very tiny in size. Particulate pollution is mostly caused by industries, power plants, waste landfills, automobiles, building construction, fires, and natural dusts.

The rising fossil fuel combustion over the last century is to be accountable for the gradual change in climate patterns, which lead to air pollution. Intense toxic substance discharge to the environment by mostly the developing countries of Annex B of the Kyoto Protocol are responsible for being the prime sources of pollution - since their focus is to expand industrial areas on a larger scale, overuse of power generation and introducing advanced technologies in their countries which can elevate them from 3rd world countries. As a consequence of enhanced mechanization and automation, industrialization has occurred in every region of the world, raising the living standard for some people while also assisting the country's economic success. Aside from anthropogenic acts, different physical activities (volcanoes, fires, etc.)

produce various pollutants into the biosphere, which contributes equally to increased GHG emissions.

1.2 Objectives

The notion of our project is not against the development revolution of 3rd world countries - our goal is to spread awareness globally to make toxic discharge in a favourable amount. With Dhaka being entitled as one of the top most polluted cities among 106 countries according to IQAir from 2019 and onwards, the main objective of our research is to help fulfill the 13th SDG goal - Climate Action, with the help of IoT. For the introduction of automation, the idea of Industrial IoT (IoT) acquired significant attention. While this aspires to boost productivity, efficiency and real-time visibility, IoT has emerged among the most demanded research area. IoT is a system of interconnected internet - enabled devices that allows data to be transferred and received from one device to the next and enables device-to-device and human-to-device interactions in a reliable and resilient manner.

This project focuses is to improve air quality and to establish a mechanism to improve AQI by monitoring the air standard. Our goal is to diminish the problem efficiently within the limited resources economically.

1.3 Scope of the project

Our scope is to make an economically friendly device, predict PM2.5 concentration using satellite data with the help of Machine Learning and make a software as an atmospheric forecast reporting system using the device and satellite data.

This research is proposing two types of solutions to mitigate the worldwide crisis.It will consist of retrieving data from a primary source where IoT will be implemented, and a secondary source where data from AQUA satellite is extracted. The main objective of this paper is to make an inexpensive solution to the problem, which will be then compared with an industrial graded expensive device to determine the device compatibility. Additionally, we have to make sure to measure the air quality of all the regions of Bangladesh. Our aim is to make a software as an atmospheric forecast reporting system, where we intend to incorporate station wise-graphs using device and satellite data and generate an ‘AQI alert Atmospheric Map’. The diagram below shows a preliminary model of the proposed system.

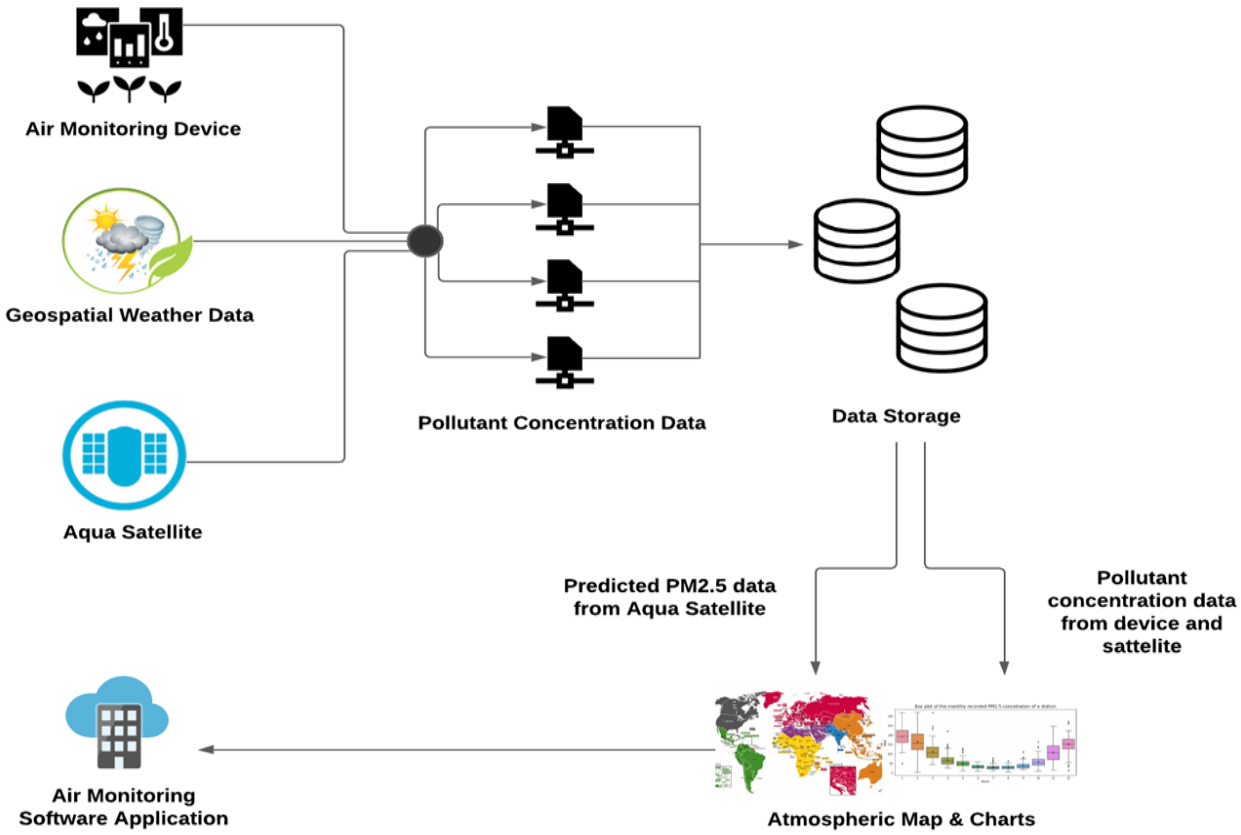

Figure 1.3 : Proposed System detailed representation diagram

In this paper, for the primary data, IoT has been proposed for the detection of polluting compounds in our atmosphere causing global warming and calculating the AQI with the help of data we get from the sensors that detect the toxic gas. The project intends to measure the temperature, humidity, barometric pressure, VOCs, multiple toxic gases and detect concentration of fine particles smaller than 1.0µm, 2.5µm and 10.0µm in diameter. With the set of sensors connected to the development board, Arduino Mega 2560 Wifi takes readings from the sensors and sends all the sensor data to the PC using Serial Communication. This data is then migrated to the cloud. All the concentration of polluting compound charts will be shown in a Google Sheet visualization for now.

Afterwards our intention is to compare the primary data of the pollutants with an industrial graded device by IQAir, to determine the device precision. Coming onto the industrial graded device we have purchased, AirVisual Pro - the vital goal of the IQAir technology is to operate the world's largest real time air quality platform. Not only has it taken the initiative to help people explore every country's real time air quality remotely through a world pollution map, but also has made it easier and more accessible to people to do instant indoor air quality monitoring by launching an application. Overall, the data of this device is known to have good accuracy and is used by people as a reference.

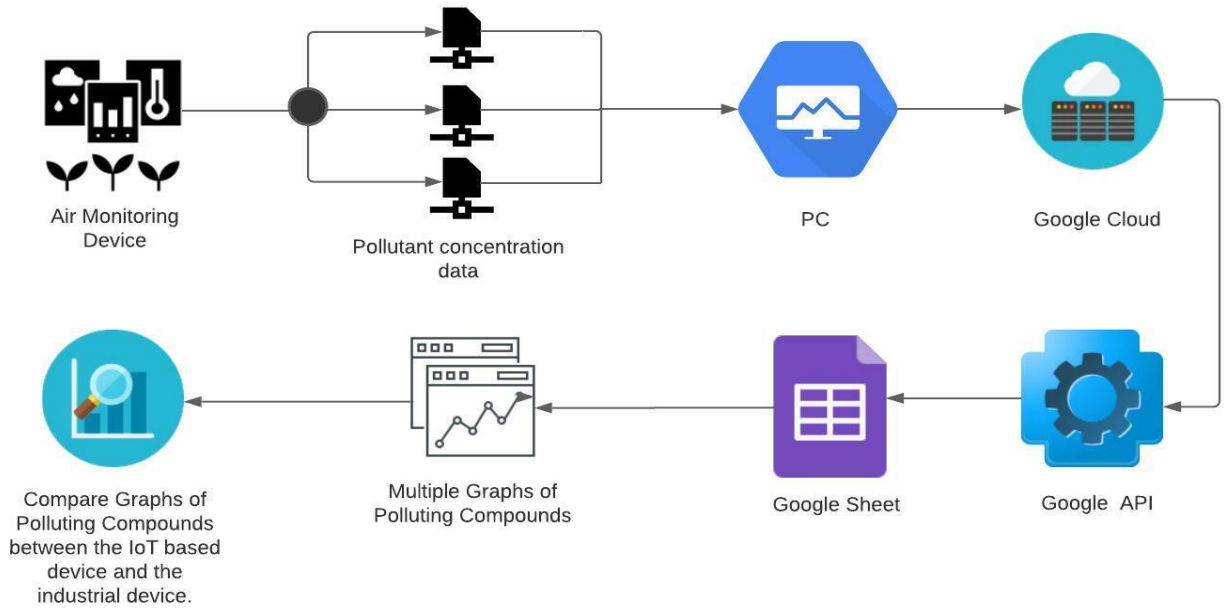

Figure 1.3.2 : IoT based device detailed representation diagram

In the process of covering all the areas of Bangladesh, the inexpensive device that we will be making will no longer be cost efficient since we have to expand the infrastructure. To solve this problem, a successful approach has been noted from multiple research where the secondary data - AOD PM2.5 product - is predicted using Machine Learning Algorithms (MLAs).

The secondary data is originally extracted from the AQUA Satellite. The AOD (Aerosol Optical Depth) 550 nm data is collected and preprocessed by the AQUA satellite using the MODIS(Moderate Resolution Imaging Spectroradiometer) instrument while orbiting above Bangladesh. AOD is the measure of 550 nm aerosols (e.g., urban haze, smoke particles, desert dust, sea salt) distributed within a column of air from the Earth's surface to the top of the atmosphere. The used AOD 550 nm is the dust aerosol optical thickness at 550 nanometers (nm). The satellite AOD (Aerosol Optical Depth) 550 nm data was collected using NASA Giovanni data visualization platform.

In the process of making an atmospheric forecast reporting system, we aspire to make a software incorporated with the device. We intend to integrate station wise graphs and using MLAs our aim is to generate an 'AQI alert Atmospheric Map',with the help of the satellite and device data. Additionally, our project will extensively monitor the most polluting areas/routes - to give an alert when the pollution exceeds the favourable amount.

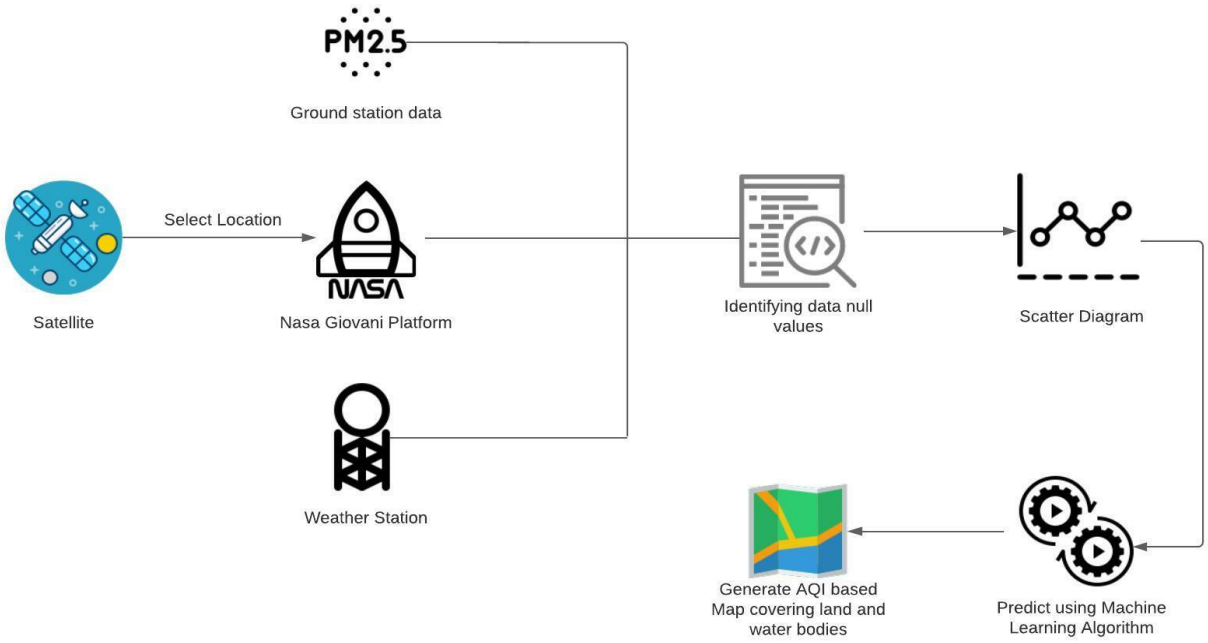

Figure 1.3.3 : Prediction of PM2.5 concentration using Satellite data detailed representation diagram

1.4 Organization of the report

In this section, we are going to determine the contents of the chapters in this paper.

The ‘Chapter 1’ denotes the introductory part of the research to give a clear idea and overview of the project. It also contains the scope of the project, where our targets for this research are discussed.

In the ‘Chapter 2’ section, the literature reviews on multiple papers will be done to find out the possibilities to achieve the goals of this project and this will help us build our knowledge in this field. The literature review will be divided into two sub-sections : one for the IoT based device and another for the satellite data.

Moving on to ‘Chapter 3’, it will contain the project breakdown, timings and schedules along with roles and responsibilities. It means to achieve goals and involves decision making, keeping in mind about all the potential risks and finding ways to diminish them by doing a critical path analysis.

The hypothesis of a vital part from our research will be found in ‘Chapter 4’. It will describe the problem statement to introduce the readers to the importance of the topic being studied. A step by step proposed system is added to this chapter to gain the consensus of the readers about the project.

The primary data source, IoT based Device which is also known as RefinAir, is discussed in ‘Chapter 5’. An introduction is given on the device and AirVisual Pro together to help the readers correlate the purpose of the project. Since the IoT based Device itself is a prototype, our aim is to make a pre-prototype for our desired system, to give a representation of the design of the finalized system. In the problem statement sub-section, we will present the expensive air quality devices available in the market as a problem and will express our desire to make an inexpensive device as a solution. Also, we will point out the fact that we will try to make our device more durable and sustainable and easier to maintain for the users. Another sub-section is made for the system description where the chosen device components are justified by comparing it with other industrially approved components. A hardware layout is shown using block and schematic diagrams - followed by the original picture of the hardware which will be assembled by us using the components. Afterwards, a performance analysis will be done to compare the data of the polluting compounds measured by both the IoT based device and AirVisual Pro. Both daytime and nighttime data will be taken to determine the data accuracy and IoT based device compatibility. Application of RefinAir will be stated and an overall comparison will be done between AirVisual Pro and RefinAir based on the difference between the functionalities and facilities given by them.

The prediction on PM2.5 data concentration using Machine Learning will be shown in the ‘Chapter 6’, where the collection of AOD and ground station data is discussed and the procedure to convert AOD 550 nm to PM2.5 data is stated. In this section, we will determine which Machine Learning Algorithms works best for prediction analysis of PM2.5 concentrations. All the Machine Learning Algorithms used and their outcomes will be discussed here.

In the last part, ‘Chapter 7’, the challenges faced are discussed for the hypothesis, primary data and the secondary data. The environmental, ethical and social issues are declared here. Authors’ interest in developing new proposals based on the same research are affirmed in the future work and recommendation sub-section.

Chapter 2 : Literature Review

This section of the report will consist of literature reviews on the two propositions we aim to execute to diminish and monitor the degradation of air quality. In order to fulfil the SDG, we proposed an IoT based inexpensive solution to solve the problem. It has been noticed that we have to create IoT based devices in industry in order to be able to monitor the air quality of all the regions of Bangladesh. Instead of adding up another cause of pollution, we came up with another promising proposition where we will extract data from AQUA satellite and use MLAs to estimate the AOD PM2.5 product. Our aim is to make an atmospheric forecast reporting system with station wise graphs.

The primary data here comes from the primary source,IoT based device, and the secondary data comes from the secondary source, AQUA Satellite.However, the literature review below is divided into two sections for the two different sources of data.

2.1 Literature Review on IoT Based Device

The following sensor system assures air quality monitoring via an array of sensors which transmit their reading through Bluetooth to the nearest smartphone. The readings get updated every time the application- installed in the smartphone- is clicked. This system also has the facility to track the location of the set of sensors by GPS in google maps. This function of the embedded system is controlled by NXP ARM mbed LPC1768 which is a microcontroller and RN-42 bluetooth module is used for communication. The advantages of this system are that this system has a wide range of toxic gases that can be detected within just a click and the GPS tracker helps more in mobile monitoring. The drawbacks of this system are that this system allows mobile monitoring but it does not have a database of data which can be later used for taking necessary measures and the bluetooth, RN-42 module, might not work efficiently if the user is out of its range and still wants to get the readings. (Yang & Li, 2015, #)

This smart vehicle monitoring system for air pollution detection using Wsn is a system for the moving vehicles to monitor the NO2, Humidity, Temperature, CO levels of air contamination by using NO2 sensor, Humidity sensor, Temperature sensor, CO sensor. MANET(Mobile Ad Hoc Network) routing algorithm is used, which has nearly 28 mobile nodes(Vehicles) that provide a coverage area of 300meters around the city. The sensor data of the vehicles will be sent to the smartphones of the appropriate drivers to monitor. The microcontroller of this system is PIC Microcontroller 16f877a and for communication Zigbee is used. This system has a high Coverage of areas with the help of 28 moving vehicles with a range of 300meters per vehicle, thus collecting a large range of data. By using a cloud network, large amounts of data of different vehicle records can be stored and retrieved for future purposes. Mobile monitoring is another plus point of this system. But this device can only detect some toxic gases and does not have a GPS facility, which can be counted as pitfalls of this system. (Suganya & Vijayashaarathi, 2016, #)

In this existing system, which is known to develop Arduino based embedded systems to detect toxic gases has an array of sensors are set within some parameters to detect the concentration of toxic gases and vapors in air. The parameters for the sensors will be set inside the code.If the parameters are breached, the system switches on the buzzer and red LED and outputs an alarm message to the LCD module and PC via the serial interface. Also sends SMS to the mobile device via a GSM module. Only few MQ sensors are used here and are incorporated with ATmega2560 which is the microcontroller of this system and for communication GSM module is used. But this system gives only real-time data, thus does not have a database which can be used for future purpose and the data can not be shown anytime on smartphones unless there's a breach, thus partial mobile monitoring. (Holovaty et al., 2018, #)

Another project about low cost IoT based air monitoring system using Raspberry Pi with MySQL Database is reviewed below. Here, a set of data is extracted from some sensors which had some threshold values set by the system. Raspberry-Pi is interfaced with various sensors (temperature, Humidity, MQ 5 Gas Sensor) and real-time data will be obtained and stored in MySQL Database. If the sensor value from the environment exceeds the threshold value, the communication module sends the message alert to the client by using ThinkSpeak open data IoT platform. In this system, NodeMCU and Raspberry Pi are used as microcontrollers and ESP8266 is used for data transmission. All the data is stored for future research purposes. However, this device has less variety of sensors and does not have any facility for sending SMS to the user in case the user does not have internet and the thresholds are breached. No location trackers are incorporated in this system which can be known as another drawback of this system. (Kiruthika & Umamakeswari, 2017, #)

The following literature review is about a project where a mobile GPRS-Sensors array is used for air pollution monitoring. They have developed an air pollution geo-sensor network Consisting of 24 sensors taking 24/7 readings of CO, NO2 and SO2 to obtain the AQI (Air Quality Index). They've used GPRS-Modem to transmit the data to a php and mysql based database server and GPS-Module to locate the location of the sensors and using google maps APIs they have obtained the real time readings of the sensors and the current location of sensors. They've developed a system which can be portable and can be attached to any public transport to obtain sensor readings remotely with live GPS location. The sensors are incorporated with HCS12 which is a microcontroller in this system. The collected data is stored for future research purposes and helps achieve statistical graphs to track the sensor readings. This method provides real-time location and sensor readings of the sensor array. Additionally, this device is portable and can be attached to any public transport to obtain sensor readings and GPS location remotely. However, they could’ve used a single microcontroller board with embedded GPS and GSM to make the system more compact and no sensor to detect PM2.5 concentration was used. (Al-Ali et al., 2010, #)

In this research, IoT is used for Mobile–Air pollution monitoring system where an IoT solution is developed to detect the AQI (Air Quality Index) and the concentration of Carbon Monoxide (CO), Methane (CH4) and Carbon dioxide (CO2) gases. They also used GPS to get the location and timestamp of the devices to get the sensor readings of the specific locations. They used an ESP8266 to send the obtained sensor readings and the GPS location to a cloud server. They have used Ubidots as a cloud server to receive the data from the devices remotely and then send the data to the android app they’ve built. In the android app the user can see the pollution level of any specific route the user is using. ATmega328P and ESP8266 are incorporated in this system as a microcontroller and communication purpose simultaneously. The collection of CO, CH4, CO2 concentration and AQI data is stored for future research purposes, but no sensor for the detection of PM2.5 compounds is used here. This device is portable and can be attached to any public transport to obtain sensor readings and GPS location remotely. (Dhingra et al., 2019, #)

2.2 Literature Review on AOD PM2.5 Product Prediction Analysis using Machine Learning Algorithms

A hybrid remote sensing and machine learning approach, named RSRF model is proposed to estimate daily ground-level PM2.5 concentrations, which integrates Random Forest (RF), one of machine learning (ML) models, and aerosol optical depth (AOD), one of remote sensing (RS) products. The model inputs consist of data sets for AOD, SO2, NO2, CO, O3, AT, RH,WS, P and WD. The RSRF model was compared with Multiple Linear Regression (MLR), Motivation, Abilities, Role Perception and Situational Factors (MARS) and Support Vector Machine (SVR) models. The RSRF, MARS and SVR models have higher prediction accuracies than the MLR model. The results indicated that the RSRF model had a relatively high prediction accuracy, outperforming other three models. The main conclusions derived for this study is that the prediction abilities of different predictors on PM2.5 concentrations vary seasonally. AOD is not a unique indicator for fine particle pollution. As a result, more directly relevant predictors such as important precursor pollutants should be considered. Although weather variables have less direct effects on PM2.5 pollution, they are essential in the RSRF model. (Li & Zhang, 2019, #)

In China, a geo-intelligent deep learning model, Geoi-DBN, was developed for better representation of the AOD-PM2.5 relationship using geographical distance and spatiotemporally correlating to PM2.5 in a deep belief network.. The geographical correlation was adopted to significantly improve the estimation accuracy. Geoi-DBN can capture the essential features associated with PM2.5 from latent factors. The results show that Geoi-DBN performs significantly better than the traditional neural network. However, in this research the Geoi-DBN achieved satisfactory performance and in the Geoi-DBN cross-validation slope of observed PM2.5 versus estimation indicated some evidence of bias. The reason behind this was groundlevel PM2.5 is greater than 60 μg/m3 in China. Additionally, this underestimation may be down to several reasons, including the possibility of mixed types and layers of aerosols in the atmosphere and the hygroscopicity of urban aerosols. While previous studies have used machine learning to simulate the AOD-PM2.5 relationship, this study further considers the geographical correlation to greatly improve model performance. (Li et al., 2017, #)

For mountainous regions, a recently developed algorithm, Multiangle Implementation of Atmospheric Correction (MAIAC) is used for the Moderate Resolution Imaging Spectroradiometer (MODIS), which provides Aerosol Optical Depth (AOD) at a high resolution of 1 km. A filtering scheme has been developed to reduce the two main sources of artifacts in MAIAC high resolution AOD from clouds and snow. MAIAC AOD has similar accuracy as MODIS Collection 5.1 AOD product but provides information at 2–3 km spatial scale and with better data coverage due to the higher resolution and less restrictive statistical filtering. The MAIAC AOD product can be used for climate related studies, i.e. the assessment of seasonal or annual averages. (Emili et al., 2011, #)

In a statistical model, that is trained to predict hourly concentrations of particles smaller than 10 m (PM10) by combining satellite-borne Aerosol Optical Depth (AOD) with meteorological and

land-use parameters, is shown that besides human emissions, concentrations of particles in the air are to a large extent driven by meteorological factors such as wind direction.With increasing data availability and computational power, machine learning methods, e.g. Artificial Neural Networks and Random Forests (RF) have been applied frequently in recent years. These machine learning models are beneficial as they efficiently reproduce non-linear relationships and interactions of input features.To this end, Gradient Boosted Regression Trees (GBRT) are used to understand and quantify the conditions driving air quality, as well as determinants of the relationship between AOD and PM10. GBRT has been successfully applied to study sensitivities of aerosol processes before.GBRT as implemented in python's scikit-learn module are used and merges several statistical approaches found in machine learning applications like decision trees, boosting and with gradient descent. The use of GBRT proved fruitful to understand interconnected processes and the approach presented here can be potentially expanded to other research questions focusing on the understanding of multivariate processes. Future e reports will further address the determination of mechanisms leading to high pollution events using machine learning not only for total PM10 concentrations, but for individual aerosol species. (Stirnberg et al., 2020, #)

In this paper, an improved high-spatial-resolution aerosol retrieval algorithm with land surface parameter support (I-HARLS) at 1-km resolution for MODIS images is developed. A precalculated global land surface reflectance (LSR) database is constructed using the MODIS 8-day synthetic surface reflectance (MOD09A1) products, and a prior seasonal global Land Aerosol Type (LAT) database is created using the MOD04 daily aerosol products. The main aerosol optical properties and types are determined based on the monthly average historical aerosol optical properties from local AERosol RObotic NETwork (AERONET) sites. For cloud screening, the Universal Dynamic Cloud Detection Algorithm (UDTCDA) is selected to mask cloud pixels in remote sensing images. Then, a 1-km-resolution AOD dataset is generated based on the I-HARLS algorithm. Successful AOD retrievals are available over dark and bright surfaces. To test and validate the performance of the I-HARLS algorithm, four typical regions (including Europe, North America, Beijing-Tianjin-Hebei and the Sahara) with different underlying surface and aerosol types are selected for aerosol retrieval experiments. Moreover, AERONET Version 2 Level 2.0 AOD measurements and MODIS daily AOD products at 3-km resolution (MOD04_3K) are selected for validations and comparisons.The results show that the I-HARLS algorithm performs well overall at both the site and regional scales, and AOD retrievals are highly correlated with AERONET AOD measurements.This study shows that although the new AOD retrieval algorithm performs well overall over land, certain problems remain. However, due to the long time series of MODIS data records, longer and wider-scale experiments and validations need to be undertaken. In addition, this paper only performs comparisons with current operational and free-open high-resolution MOD04 aerosol products; therefore, more comprehensive and effective comparison efforts with other high-resolution products (such as MAIAC products) need to be performed in future studies. (Wei et al., 2018, #)

This study aimed to develop machine learning-based models for predicting hourly street-level PM2.5 and NOx concentrations at three roadside stations in Hong Kong. This study highlights the capability of MLAs to produce high temporal resolution air pollution predictions, which can

supplement traditional methods (e.g., land use regression) in generating accurate and high-temporal-resolution estimations of air pollution concentration.The researchers comprehensively evaluated and compared the performance of six common machine learning algorithms (MLAs) including Random Forest (RF), Boosted Regression Trees (BRT), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Generalized Additive Model (GAM), and Cubist and hence applied the most suitable MLAs to apportion the contributions from emission and non-emission factors to hourly street-level PM2.5 and NOx concentrations. The results show that RF outperformed other MLAs and BRT, XGBoost and Cubist presented comparable predictive performances. SVM and GAM have worse predictions than other MLAs. (Li et al., 2020, #)

Chapter 3 : Project Planning

A project plan establishes project objectives, outlines activities, determines the way to accomplish goals and specifies what resources will be required - keeping in mind about the corresponding budgets and completion timeframes. A project plan specifies all of the work and steps initiated that will be done in a project and who will accomplish it.

3.1 Work Breakdown Structure (WBS)

Work Breakdown Structure determines the steps that need to be initiated with a project which helps the user to accomplish strategic planning. It helps the researcher to monitor the phase of the project with an effective plan. It helps detect the risks and outbacks of the project, leading it to a more successful approach as a project planner.

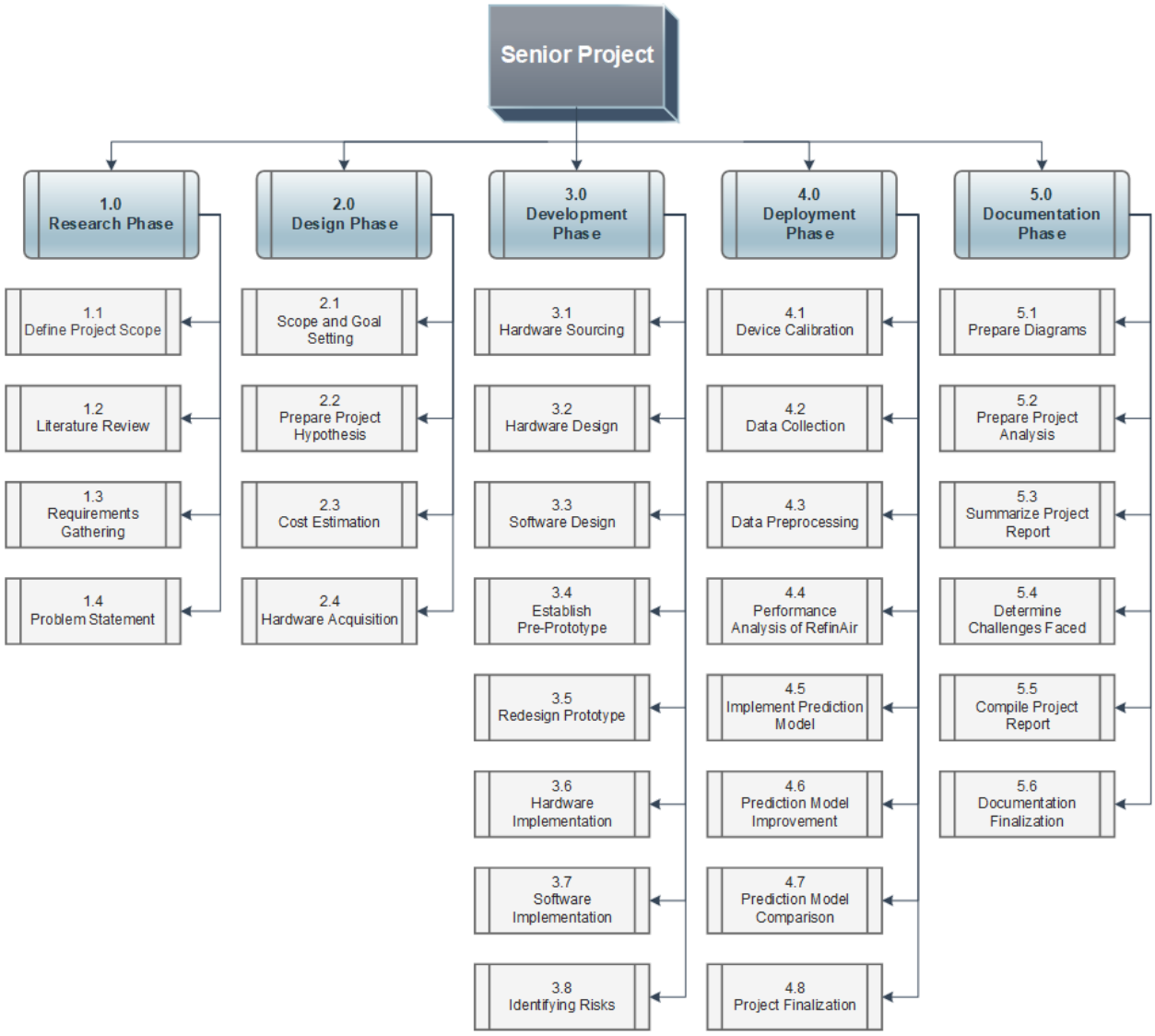

Figure 3.1 : Work Breakdown Structure

3.1.1 Air Quality Monitoring System

This demonstrates the proposed system we aspire to make in our project. All the hardware, software implementation we want to acquire in this project will be stated here. Please note that this section contains the idea we intend to accomplish.

In this section, we will be giving insights about the steps of the project we intend to implement in order to reach our goal. In order to choose the sensors for detecting the level of air toxins, we first need to determine which air pollutants should we target to monitor. The maximum parameters need to be researched that will be appropriate for maintaining a healthy surrounding for different factors, e.g. indoor and outdoor air pollutant levels. For choosing the appropriate sensor, it is important to notice if the sensor is suitable for the environmental condition in which it will inhabitat. The range of the sensor is necessary to identify to know if the sensor will be able to give it's best performance in terms of sensing according to the environment. The sensor shape is important because our aim is to make a compact device.
Please note that it is important to maintain the cost estimation for cost feasibility so that normal citizens can also get the opportunity to track the rise of the catastrophe and get aware of the situation. As most of the citizens in 3rd world countries are from middle class background, this engagement in air monitoring will encourage them to use less air polluting substances or at least reduce them.
The next step to help us lead to a successful project is to read multiple research papers that are published in different Journals and Conferences related to this topic. It is advised to do a literature review before starting the project to get prepared about the upcoming steps to reach a fruitful outcome.
After being able to choose the right sensors and other components for the project, the sensor data containing concentrations of air pollutants will be sent to the PC and simultaneously, the data will also be uploaded into the software we aspire to make. The set of data will be stored in a database and with the help of Machine Learning Algorithms our plan is to make an 'AQI Atmospheric Map'.

3.1.1.1 Identifying requirement

An essential step to establish the project's requirements in advance is to be inclusive about the project's requirements, which can lead to a coherent and realistic conclusion for the betterment of the society.

1. Identify the Stakeholders

Stakeholders are those who have a direct or indirect relationship with the organization. They are the people who make a contribution to the project in some capacity. The stakeholders of this project are the researchers whose sole aim is to run the project successfully; the machine learning engineers whose work is to input the prediction process into the system; the software engineer who makes sure to provide the customers a user friendly experience and lastly the customers who purchase the device to be aware of the alarming situation.

2. Gather information

Gathering information entails uncovering all conceivable scenarios that are important to the topic, followed by putting the best concept into action - in order to meet an effective end. This can be accomplished by reading a number of research articles and publications that are similar to the project outline. Prototypes can be created to assess which process is ideal for the project and to monitor the results. Obtaining assistance from faculty or other individuals with experience in the issue will further enlighten the researcher.

3. Verify and Validate

Time constraints can be set for each project milestone, and resources can be identified in order to complete the project as quickly as possible. Budget allocation is an important component of the project since we need to ensure that our technology is both sustainable and cost-effective because our goal is to raise awareness among people of all socioeconomic backgrounds.

4. Confirm the feasible sustainable requirements

Multiple research and web searches on other industrial-based devices can be conducted to choose the most sustainable components for the project in order to achieve the highest level of sustainability. It is critical to learn about the drawbacks of other industrial-based gadgets so that we can add those features into our own.

3.1.1.2 Implementation

- Indoor air monitoring system
- Outdoor air monitoring system
- Atmospheric Map
- Air monitoring application
- Area wise air monitoring reporting station

3.1.2 RefinAir

RefinAir is an IoT based device which is connected to multiple air pollutant detecting sensors for observing the level of air contamination level. This is considered as an inexpensive solution to the worldwide crisis for spreading awareness to people about the rising catastrophe. In order to

find out if the device is as fit as the industrially approved air monitoring device, we explored an existing air monitoring system, AirVisual Pro by IQAir, that can be compared with the IoT based device, RefinAir. However, the existing air monitoring system, AirVisual Pro, is discussed below.

The history of air cleaning and AQI improvement began in 1963 with the brothers Manfred and Klaus, who invented domestic air filters in Switzerland that are currently used in more than 70 nations. Customers and air quality experts all around the world have praised this item. It has been deployed in the Olympics because it refines 99.5 percent of all particulate matter and has been shown to deliver medical-grade air utilizing HyperHEPA technology. The 3D ultraseal is intended to prevent air leakage. IQAir Earth is the world's first 3D air pollution map. It provides interior and outdoor surveillance systems, as well as hourly weather and air pollution forecasts and alarms if the air becomes hazardous. IQAir launched the world's first air quality application, which provides real-time data as well as historical air quality information. to be compared to RefinAir, an IoT-based technology.

To make the Iot based devices as notable as the standard ones, we aimed to incorporate multiple sensors for detecting hazardous gases in the surroundings for both indoor and outdoor monitoring. Various literature reviews were done to help us find the appropriate sensors with more accuracy. For hardware sourcing, much advice was taken from the supervisors and the appropriate hardware was sourced from different e-commerce mediums. Before using the industrially graded sensors that we purchased, we decided to assemble local sensors with a microcontroller as a pre-prototype to get a glimpse of the final outcome of the project. After a successful attempt, we decided to assimilate more features to RefinAir. We took ideas from different research papers and came to a conclusion to upload our pollutant concentration data into Google Cloud. We browsed uncountable tutorials to find out what could be done with the data and with the help of our Project Manager we deduced a solution. According to the guidance of our supervisor, we decided to migrate both the data coming from AirVisual Pro and IQAir into Google Sheets and generate graphs of the same pollutants. Followed by the graph generation, accuracy errors can be calculated by looking into the graph trends and performance of RefinAir can be judged - making the AirVisual Pro’s data as the superior data.

3.1.2.1 Identifying requirement

It is important to identify the requirements of the project beforehand to be concized about the requirements of the project which can lead to an understandable and feasible outcome, keeping in mind about the benefits that can be provided to the society.

1. Identify the Stakeholders

Stakeholders are the category of people who have some participation with the organization directly or indirectly. They are the individuals who contribute to the project in some way. The stakeholders of this project are the researchers whose sole aim is to run the project successfully; hardware vendor who provides the hardware components of the project; maintenance team who makes sure about the sustainability of the device and responds to all customer queries and gives customer service; and lastly customers who purchases the device to be aware of the alarming situation.

2. Gather information

Gathering information is discovering all the scenarios possible that are relevant to the topic, followed by executing the best idea to it - in order to get the most fruitful outcome. This can be done by giving a read to multiple research papers and journals that are similar to the project abstract. Prototypes can be made to determine which process will be best for the project and observe the outcome. Getting help from the faculties or other individuals who have experience in the topic will enlighten the researcher more.

3. Verify and Validate

Time limitations can be fixed for each milestone of the project and resources can be recognised in order to run the project swiftly. Budget allocation plays an important part of the project where we have to make sure that our device is sustainable and budget friendly because our intention is to spread awareness to all kinds of social status.

4. Confirm the feasible sustainable requirements

Multiple research and web searches on other industrial based devices can be done to find out the most sustainable components for the project to get the utmost sustainability. It is important to find out the disadvantages of other industrial based devices so that we can incorporate those facilities as well in our device.

3.1.2.2 Implementation

- Indoor air monitoring system
- Outdoor air monitoring system

3.1.3 Prediction Model using Satellite Data

To cover all the regions of the country from land to water bodies for discovering the contamination level of the polluting compounds, our goal is to extract the 550nm AOD product from the satellite and convert the 550 nm product to PM2.5 and preprocess the data. We want to prepare a scatter diagram to determine the correlation. After the correlation, we aspire to find the correct MLA to predict the area-wise PM2.5 concentration in future.

After a literature review on predicting PM2.5 from the AOD product, our main goal was to gather AOD, station PM2.5, and weather data from available online resources. We intended to collect the AOD product data from NASA EARTH DATA's Giovanni data visualization platform. The available AOD 550 nm product was preprocessed from MODIS (Moderate Resolution Imaging Spectroradiometer) instrument aboard the Aqua Satellite, which passes from South to North poles of the earth. The obtained AOD 550 nm product is basically the level-3 atmosphere daily global product (MYD08_D3), which are derived from four level-2 MODIS AQUA atmosphere products MYD04_L2, MYD05_L2, MYD06_L2, and MYD07_L2. We selected U.S Embassy Dhaka as the region shape of the data in Giovanni. Aerosol Optical Depth 550 nm (Deep Blue, Land Only) was selected as the dataset for the AOD product. We extracted AOD data from 01-01-2017 to 31-05-2021 from the Giovanni platform. All the data was extracted as a csv file. Subsequently, PM2.5 raw concentration data were obtained from AirNow. AirNow is a partnership of the U.S EPA (Environmental Protection Agency), National Oceanic and Atmospheric Administration (NOAA), National Park Service, NASA, Centers for Disease Control, and tribal, state, and local air quality agencies. The extracted data from AirNow represents the raw concentration of PM2.5 as a unit of $\mu g/m^3$. We extracted PM2.5 data from 01-01-2017 to 31-05-2021 from the AirNow website. All the data was extracted as a csv file. Afterwards, we converted the per hour data to mean concentration of PM2.5 per day to fit with the AOD 550 nm product data obtained from the AQUA MODIS. Henceforth, we obtained the weather dataset from Visual Crossing Weather services web application. The dataset contains mean temperature, rain precipitation, wind speed, visibility, cloud coverage, and humidity data. All the data are mean data of the particular day. We extracted weather data from 01-01-2017 to 31-05-2021 from the Visual Crossing Weather services website. All the data was extracted as a csv file. After successful extraction of all the datasets we combined all the datasets day-to-day from 01-01-2017 to 31-05-2021. We manually preprocessed the dataset by eliminating all the outliers and null data for any particular date. After preprocessing the dataset we split the dataset on the basis of training and testing at 70:30 ratio respectively. Afterwards, we applied multiple regression linear models and various machine learning models to predict the PM2.5 raw concentration from the AOD data. The decided prediction models applied to predict the PM2.5 concentration from the AOD 550 nm and weather data are the following:

I. Multiple Regression Linear (MLR) Model
II. Artificial Neural Network (ANN)
III. Random Forest Model
IV. Gradient Boosting Regressor Model
V. Extreme Gradient Boosting (XGBoost) Model
VI. CatBoost Regression Model

3.1.3.1 Implementation

- Predict PM2.5
- Generate overall country AQI Atmospheric Map
- Cover monitoring the land and water transportation bodies.

3.2 Gantt Chart

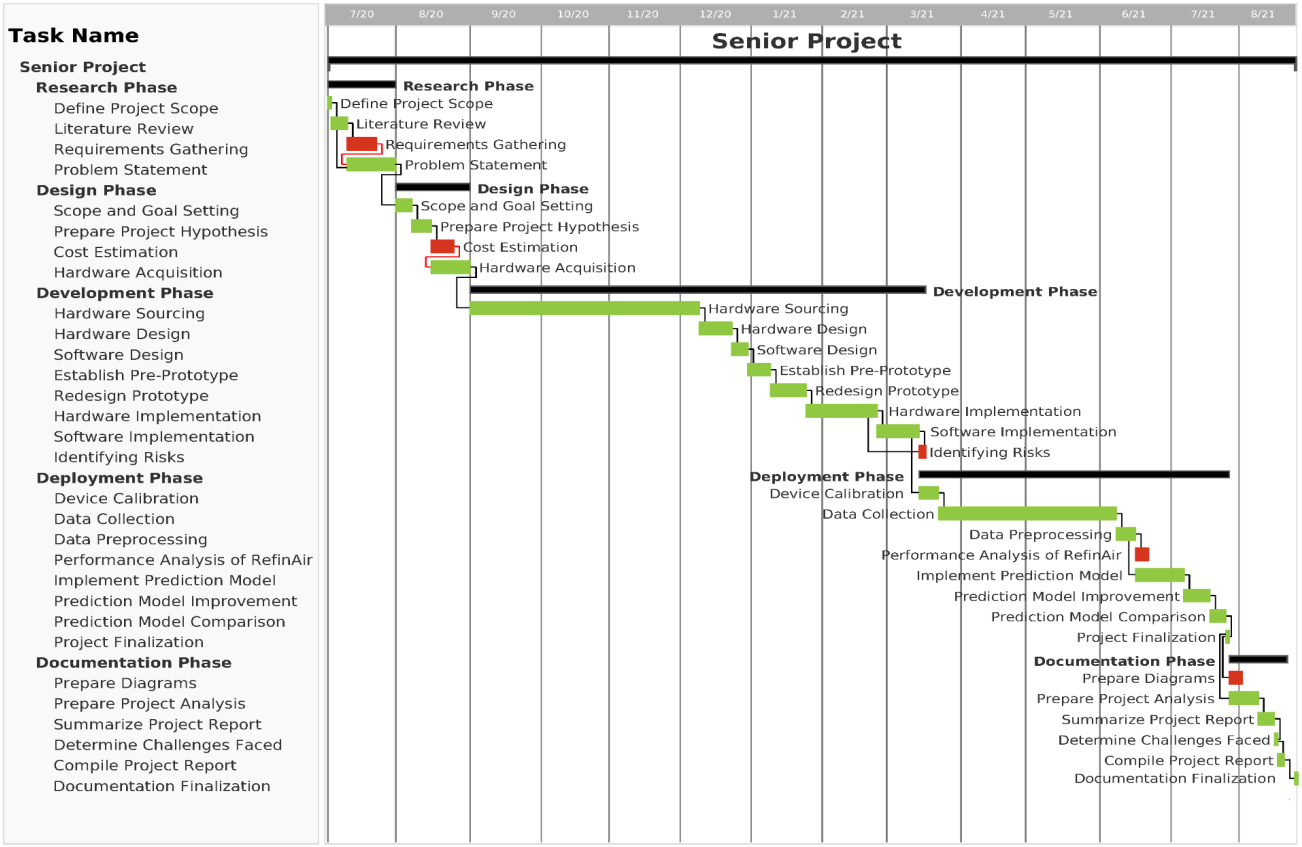

Figure 3.2 : Gantt Chart for RefinAir

3.3 Cost Estimation

We are offering three solutions to the problem. This section will demonstrate the component list and cost estimation of the device.

3.3.1 RefinAir Cost Estimation

SL No	Element / Component Name	Sensor / Item Name	Item Description	Unit Price
RefinAir Device Price				
1	Temperature	BME280	BME280 is a Breakout Board featuring a Bosch Sensortec ME280 Temperature, Humidity & Pressure Sensor.	৳ 700.00
2	Humidity			
3	Barometric Pressure			
4	PM1.0	PMS5003	PMS5003 is a kind of digital and universal particle concentration sensor, which can be used to obtain the number of suspended particles in the air, i.e. the concentration of particles, and output them in the form of digital interface.	৳ 2,700.00
5	PM2.5			
6	PM10			
7	Carbon Dioxide (CO_2)	MH-Z19B NDIR CO_2 Module	The MH-Z19B NDIR infrared gas module detects the presence of CO_2 in the air using the non-dispersive infrared (NDIR) principle.	৳ 2,800.00
8	Nitrogen Dioxide (NO_2)	Grove Multichannel Gas Sensor V2	Grove Multichannel Gas Sensor V2 uses MEMS technology to detect a variety of gases like Carbon monoxide (CO), Nitrogen dioxide (NO2), Ethyl alcohol(C2H5CH), Volatile Organic Compounds (VOC) and etc.	৳ 5,500.00
9	Total Volatile Organic Compounds (TVOC)			
10	Carbon Monoxide (CO)	Gravity: Analog Carbon Monoxide Sensor	This is a CO sensor. It detects carbon monoxide (CO) concentrations in the air between 20 and 2000ppm using the MQ7 probe.	৳ 1,550.00

11	Microcontroller Board	Arduino Mega 2560 WiFi Board	It is a development board using the Atmega2560 microcontroller. It has a built-in WiFi module incorporated with it.	৳ 1,200.00
12	GPS	U-Blox NEO-6M GPS Module	The NEO-6M GPS module is a well-performing complete GPS receiver with a built-in 25 x 25 x 4mm ceramic antenna, which provides a strong satellite search capability and accurate GPS coordinates.	৳ 900.00
13	Data Storage	Arduino SD Card Module	It stores data into an SD card from the arduino development board.	৳ 300.00
14	Wiring	Cables and board		৳ 500.00
Communication & Power				
16	Communication	4G WiFi Router Modem	Provide internet through WiFi to send data from the device to the cloud	৳ 1,400.00
17	Solar Power	Solar Charging Power Supply with solar panel	It charges the battery using the solar panel	৳ 2,500.00
18	Battery	Lipo Battery 3300mAh 11.1V 3S	It powers the device.	৳ 2,400.00
			Total:	৳ **22,450.00**

3.3.2 Air Visual Pro Cost Estimation

SL No	Element / Component Name	Sensor / Item Name	Item Description	Unit Price
AirVisual Pro Device				
1	PM1.0	IQAir Visual Pro	Air Quality monitoring device.	৳ 28,000.00
2	PM2.5			
3	PM10			
4	Temperature			
5	Humidity			
6	Carbon Dioxide (CO2)			
Additional Sensor & Components				
7	Nitrogen Dioxide (NO_2)	Grove Multichannel Gas Sensor V2	Grove Multichannel Gas Sensor V2 uses MEMS technology to detect a variety of gases like Carbon monoxide (CO), Nitrogen dioxide (NO2), Ethyl alcohol(C2H5CH), Volatile Organic Compounds (VOC) and etc.	৳ 5,500.00
8	Total Volatile Organic Compounds (TVOC)			
10	Carbon Monoxide (CO)	Gravity: Analog Carbon Monoxide Sensor	This is a CO sensor. It detects carbon monoxide (CO) concentrations in the air between 20 and 2000ppm using the MQ7 probe.	৳ 1,550.00
11	Microcontroller Board	Arduino Mega 2560 WiFi Board	It is a development board using the Atmega2560 microcontroller. It has a built-in WiFi module incorporated with it.	৳ 1,200.00
12	GPS	U-Blox NEO-6M GPS Module	The NEO-6M GPS module is a well-performing complete GPS receiver with a built-in 25 x 25 x 4mm ceramic antenna, which provides a strong satellite search capability and accurate GPS coordinates.	৳ 900.00
13	Data Storage	Arduino SD Card	It stores data into an SD card from the arduino development	৳ 300.00

		Module	board.	
14	Wiring	Cables and board		৳ 300.00
Communication & Power				
16	Communication	4G WiFi Router Modem	Provide internet through WiFi to send data from the device to the cloud	৳ 1,400.00
17	Solar Power	Solar Charging Power Supply with solar panel	It charges the battery using the solar panel	৳ 2,500.00
18	Battery	Lipo Battery 3300mAh 11.1V 3S	It powers the device.	৳ 2,400.00
			Total:	৳ **44,050.00**

3.3.3 Research and Development Staff Cost For Prediction Model

Research and Development Staff Cost For Prediction Model					
SL No	**Position**	**Base Cost**	**No. of Staff**	**Hours of Work**	**Total Cost**
1	Professor	৳ 2,800.00	1	50	৳ 140,000.00
2	Associate Professor	৳ 2,500.00	1	70	৳ 175,000.00
3	Lecturer	৳ 1,700.00	1	10	৳ 17,000.00
4	Junior Lecturer	৳ 1,350.00	1	70	৳ 94,500.00
5	R&D Officer	৳ 1,350.00	1	60	৳ 81,000.00
6	Research Assistant	৳ 100.00	2	720	৳ 144,000.00
				Total:	৳ **651,500.00**

3.3.4 Research and Development Staff Cost For Prediction Model

Research and Development Staff Cost For Device					
SL No	**Position**	**Base Cost**	**No. of Staff**	**Hours of Work**	**Total Cost**
1	Associate Professor	৳2,500.00	1	350	৳875,000.00
2	Lecturer	৳1,700.00	1	30	৳51,000.00
3	R&D Officer	৳1,350.00	1	350	৳472,500.00
4	Research Assistant	৳100.00	2	2040	৳204,000.00
				Total:	**৳1,602,500.00**

3.3.5 Cost Estimation Scenario

Alternate Cost Estimation Scenario									
	Alternate 1			**Alternate 2**			**Alternate 3**		
IoT Devices									
	No. of Devices	**Per Device Cost**	**Total Cost**	**No. of Devices**	**Per Device Cost**	**Total Cost**	**No. of Devices**	**Per Device Cost**	**Total Cost**
District-Wise Monitoring	64 x AirVisual Pro	৳ 44,050.00	৳ 2,819,200.00	64 x RefinAir	৳ 22,450.00	৳ 1,436,800.00	16 x RefinAir	৳ 22,450.00	359,200.00৳
Power Station & Surrounding Area Monitoring	755 x AirVisual Pro	৳ 44,050.00	৳ 33,257,750.00	755 x RefinAir	৳ 22,450.00	৳ 16,949,750.00	189 x RefinAir	৳ 22,450.00	4,243,050.00৳
Transportation Route									
i. National	197 x AirVisual Pro	৳ 44,050.00	৳ 8,677,850.00	197 x RefinAir	৳ 22,450.00	৳ 4,422,650.00	50 x RefinAir	৳ 22,450.00	1,122,500.00৳

	ii. Dhaka City	57 x AirVisual Pro	৳ 44,050.00	৳ 2,510,850.00	57 x RefinAir	৳ 22,450.00	৳ 1,279,650.00	15 x RefinAir	৳ 22,450.00	336,750.00৳
Sub Total :		**৳ 47,265,650.00**			**৳ 24,088,850.00**			**৳ 6,061,500.00**		
Web Development										
Web Development Fee		৳ 30,000.00			৳ 30,000.00			৳ 30,000.00		
Sub Total :		**৳ 30,000.00**			**৳ 30,000.00**			**৳ 30,000.00**		
Cloud Service										
		Monthly Cost	**Yearly Cost**		**Monthly Cost**	**Yearly Cost**		**Monthly Cost**	**Yearly Cost**	
Azure IoT Hub		৳ 2,133.00	৳ 25,596.00		৳ 2,133.00	৳ 25,596.00		৳ 0.00	৳ 0.00	
Azure IoT Edge		৳ 0.00	৳ 0.00		৳ 0.00	৳ 0.00		৳ 0.00	৳ 0.00	
Notification Hubs		৳ 853.20	৳ 10,238.40		৳ 853.20	৳ 10,238.40		৳ 853.20	৳ 853.20	
Azure Maps		৳ 0.00	৳ 0.00		৳ 0.00	৳ 0.00		৳ 0.00	৳ 0.00	
App Service		৳ 7,998.75	৳ 95,985.00		৳ 7,998.75	৳ 95,985.00		৳ 7,998.75	৳ 95,985.00	
Azure Machine Learning		৳ 0.00	৳ 0.00		৳ 0.00	৳ 0.00		৳ 2,695.25	৳ 32,103.00	
Yearly Sub Total :		**৳ 131,819.40**			**৳ 131,819.40**			**৳ 128,941.20**		
Sub Total for 5 Years :		**৳ 659,097.00**			**৳ 659,097.00**			**৳ 644,706.00**		
Research and Development										
Device R&D Cost		৳1,602,500.00			৳1,602,500.00			৳1,602,500.00		
Prediction Model R&D Cost		৳ 0.00			৳ 0.00			৳ 651,500.00		
Sub Total :		**৳ 1,602,500.00**			**৳ 1,602,500.00**			**৳ 2,254,000.00**		
System Maintenance										
		Yearly Cost / Salary			**Yearly Cost / Salary**			**Yearly Cost / Salary**		
Hardware Maintenance										
	IoT Maintenance Engineer	৳ 600,000.00			৳ 600,000.00			৳ 600,000.00		
	Additional Device Maintenance Cost (10%)	৳ 4,726,565.00			৳ 2,408,885.00			৳ 606,150.00		
Software Maintenance										
	Web Application Maintenance Engineer	৳ 600,000.00			৳ 600,000.00			৳ 600,000.00		

	Cloud Maintenance Engineer	৳ 600,000.00	৳ 600,000.00	৳ 600,000.00
	Additional Cloud Maintenance Cost (20%)	৳ 26,363.88	৳ 26,363.88	৳ 25,788.24
Yearly Sub Total :		**৳ 6,552,928.88**	**৳ 4,235,248.88**	**৳ 2,431,938.24**
Sub Total for 5 Years :		**৳ 32,764,644.40**	**৳ 21,176,244.40**	**৳ 12,159,691.20**
Total :		**৳ 82,321,891.40**	**৳ 47,556,691.40**	**৳ 21,149,897.20**

Chapter 4 : Cloud based Air Quality Monitoring System (CAQ)

4.1 Problem statement

This section of the research depicts the problem statement of the project, which has encouraged us to go on with the project. Ac-cording to research, the average annual PM 2.5 concentrations inBangladesh were 77.1 micrograms per cubic meter (µg/m3) of air,which is seven times higher than the WHO exposure guidelines,with Dhaka standing second among 106 countries. Investigatorsfrom IQAir, a worldwide air quality information and technology enterprise, evaluated pollution data from 106 nations, specifically detecting PM2.5, a microscopic pollutant that can pose serious health concerns. In this situation, we need a system that will col-lect data from different areas of Bangladesh and forecast the air quality. To eliminate the climate change crisis, we have chosen aset of sensors with a microcontroller to detect the concentration of pollutants and monitor the data patterns. Our collected data showed promising results with respect to the industrial-approved device called AirVisual Pro by IQAir. Therefore, we believe that a cloud-based big data analysis-driven air quality monitoring system(CAQ) can be developed which will forecast the air quality.

4.2 System Analysis AS IS

4.2.1 Rich Picture AS IS

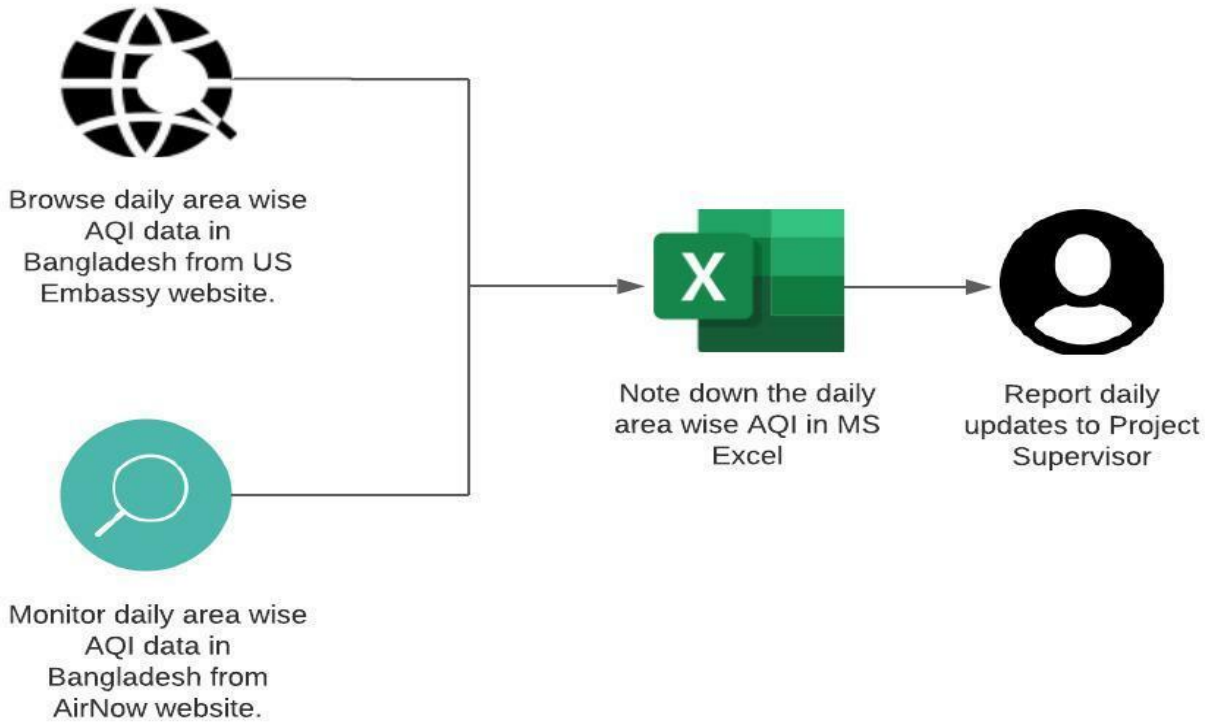

Figure 4.2.1 : Rich Picture of AS IS

4.2.2 Six Element Analysis AS IS

Stakeholder	Non-Computing Hardware	Computing Hardware	Software	Database	Communication & Network
I. Researcher ● Collects AQI data from multiple websites and notes down. ● Provides updates to the project supervisor. **II. Project Supervisor** ● Guides the researchers. ● Prepares and plans the project outline. ● Make sure the researcher is going in the right direction. ● Provides feedback to the project.	**I. Paper and Stationery** ● Paper & stationeries can be used by the researchers to note down the AQI. .	**I. PC** ● Helps the users to search for AQI. ● Helps the users to note down the raw collected data to MS Excel.	**I. Google Chrome** ● Helped search AQI data from multiple websites. **II. Microsoft Excel** ● Researchers store the AQI data found from multiple websites.	**I. Microsoft Excel** ● Researchers store the AQI data found from multiple websites.	**I. ISP** ● Provides internet connection to all the necessary activities that need internet support.

Table 4.2.2 : Six Element Analysis AS IS

4.3 System Analysis TO BE

We want to create an online web based big data driven solution which will help us monitor air quality and will help us make data driven decision making.

4.3.1 Rich Picture TO BE

The main purpose of our proposed cloud-based air quality monitor-ing software system (CAQ) is to improve air quality and to establish a mechanism to improve AQI by monitoring the air standard. In-spired by the success of IQAir, we propose to develop a Big Data Analysis driven cloud-based software system which will be used to ensure the air quality of Bangladesh. A high-level diagram of our proposed cloud-based air quality monitoring system (CAQ) is presented in Figure 4.3.1.

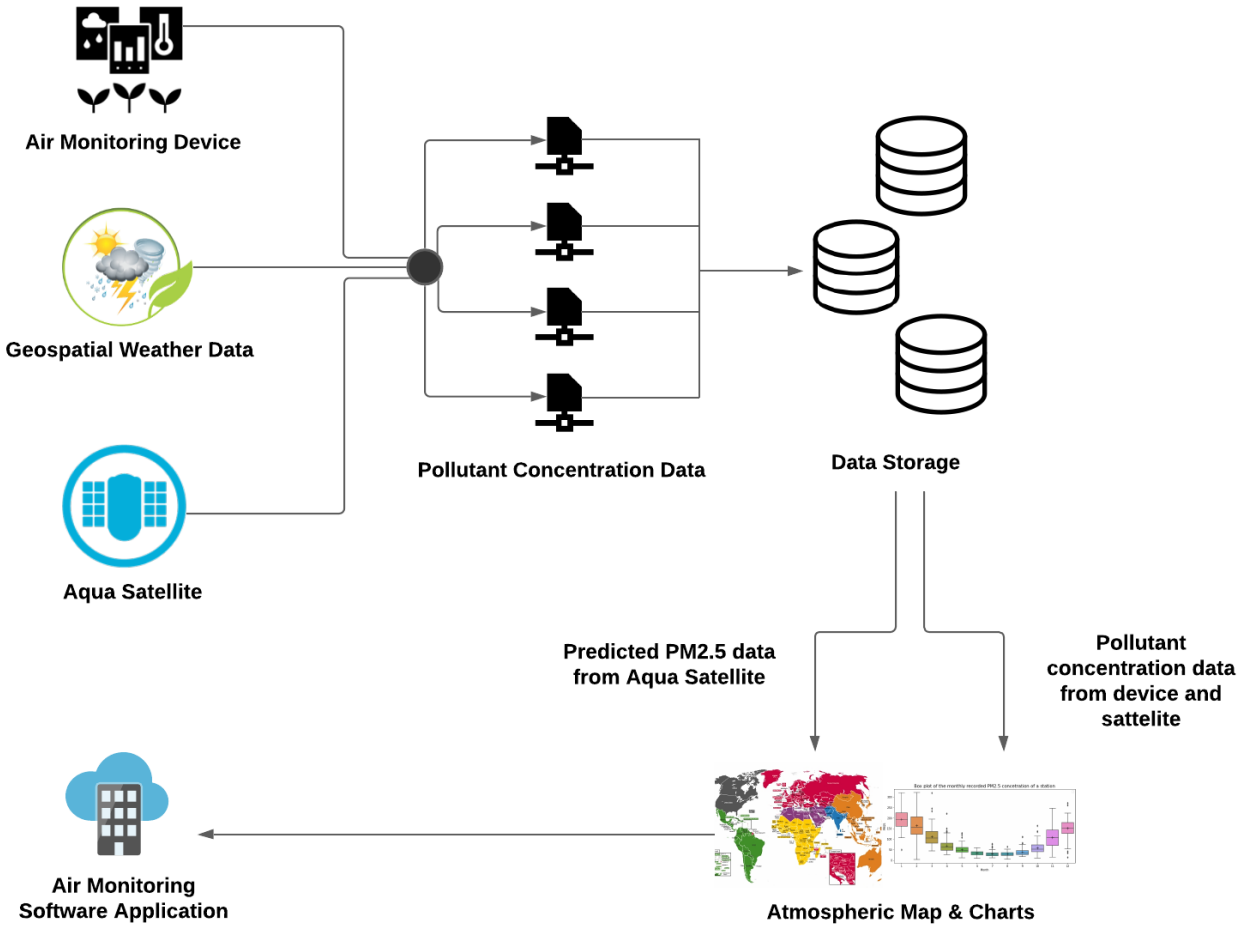

Figure 4.3.1 : Rich Picture TO BE for Proposed system

4.3.2 Six Element TO BE

Stakeholder	Non-Computing Hardware	Computing Hardware	Software	Database	Communication & Network
I. Researcher • Does research and chooses which polluting compounds to monitor. • Does literature review. • Reads multiple articles on finding sensors that are as good as industrial graded sensors. • Sources the appropriate hardwares. • Assembles the hardware and turns into a device. • Contacts the Software Engineer and discusses about the requirements needed to make the software. • Makes sure that the device migrates data to PC through serial monitor and to the 'Air monitoring Software'. • Makes sure the data is stored into a database. • Contacts the Machine Learning Engineer and plans ways to implement the data into an 'AQI Atmospheric Map'. **II. Software Engineer** • Plans the software	**I. Sensor** • All the air polluting compound detecting sensors were included to fetch their concentration data. • Communication based sensors were used to send the data to the PC and software. **II. Microcontroller** • This is the brain of the device. All the sensors are connected to it. • Sends and receives signals. • Follow the commands written in the code. • Stores a limited amount of data. **III. Connecting Wire** • Creates a pathway that connects the sensors and the microcontroller in order to work the device. **IV. Breadboard** • Helps to build and test circuits.	**I. PC** • Helps to do the research. • Helps to code for making the device work. • Helps monitor the fetched data from the serial monitor. • Helps in order to make the software. • Helps to preprocess data. • Migrate data to the database. • Helps to implement multiple MLA for prediction. **II. Server** • The servers help store data into the database.	**I. Air Monitoring Software** • Fetches data from the IoT based device. • Migrates the data into the database. • Lets users know about the concentration of the pollutants. • Shows an 'AQI Atmospheric Map'. **II. Arduino** • Receives command through code • Executes the commands into the hardwares. **III. Google Chrome** • Helped do research on the topic. **IV. Kaggle**	**I. Flash Memory in Arduino** • Retrieved data from sensors are stored in a text file. **II. Database for Air Monitoring System** • The stored data is transferred to a dedicated database to preserve the daily data.	**I. Communication Module** • Sends data from the device to PC and software. **II. ISP** • Provides internet connection to all the necessary activities that need internet support.

development process. ● Requirement analysis. ● Design ● Development and coding ● Integration ● Edits according to the requirement of the researcher. ● Testing ● Implementation ● Operations and maintenance. **III. Machine Learning Engineer** ● Preprocess data from the device. ● Separate data for testing and training. ● Choose multiple models to determine which model gives the best prediction. ● Determine the outliner and exclude them. ● Train the model. ● Evaluate Model. ● Parameter turning. ● Makes area-wise predictions. ● Hands over the prediction to the Software Engineer to implement it into an 'AQI Atmospheric Map'. **IV. Customers** ● Purchases the device and uses them.			● Many Machine Learning Algorithms were written to determine which algorithm works the best. ● Showed results for the inputted data. ● Helped create 'AQI Atmospheric Map'. **V. MatLab** ● Many Machine Learning Algorithms were written to determine which algorithm works the best. ● Showed results for the inputted data. ● Helped create 'AQI Atmospheric Map'. **VI. Microsoft Excel** ● Data was entered here. ● Data preprocessing was done. ● Scatter diagram was created. ● Correlations were done.		

Table 4.3.2 : Six Element Analysis TO BE for Proposed system

4.4 Proposed Solution

Our proposed system will collect data through different sensors which have been set in different regions of Bangladesh. Sequentially, the collected data will be stored in the cloud directly. Since it is not feasible to set the sensors in every single place, therefore, our proposed system will also collect Aerosol Optical Depth (AOD) 550 nm data from Aqua Satellite for those areas which are not covered under sensors and geospatial weather data from Visual Crossing weather data bank. All data collected from sensors, Aqua Satellite, and weather data bank will be stored into the cloud.Based on the stored data atmospheric maps and various charts representing the AQI of different areas will be generated. Our focus is to mainly surveil the air quality of distinctive states, industrial zones, and transportation routes for both land and marine bodies.

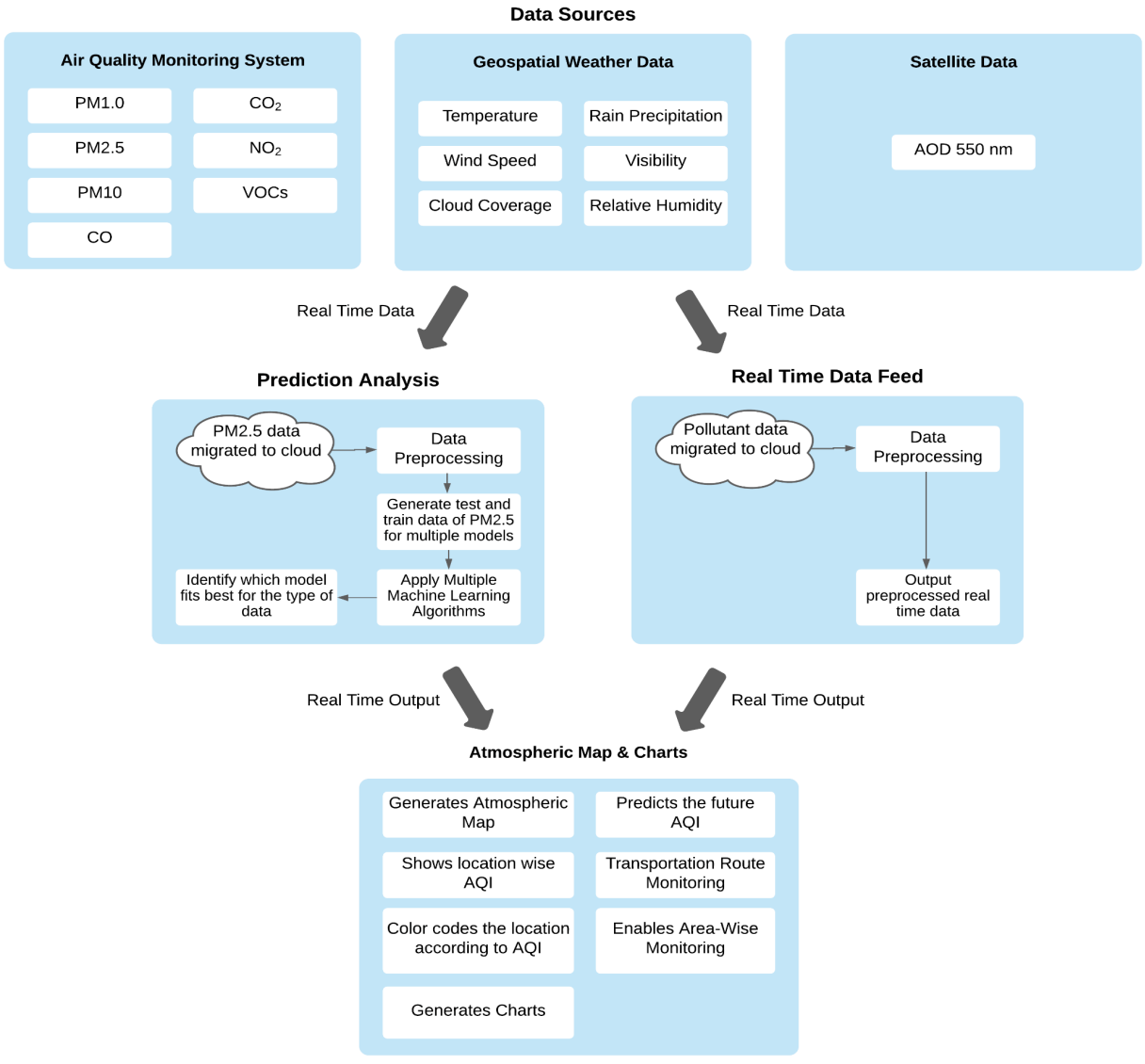

Figure 4.4 : Software System Architecture

Figure 4.4 depicts the software system architecture of our proposed system (CAQ) where the air quality data will be gathered from various sources i) Pollutant data from IoT based Air Quality Monitoring System ii) Aerosol Optical Depth (AOD) 550 nm data from Aqua satellite iii) Geospatial weather data from Visual Crossing weather data bank. Real time data discovered from different sources will be migrated to the cloud. Pollutant data will be preprocessed and generate atmospheric maps and charts for different areas such as distinctive states, industrial zones, transportation routes. Similarly, our proposed system will collect PM2.5 data and operate a prediction analysis based on the preprocessed data and fit the proper models according to data. Finally, generate AQI-based atmospheric maps and charts for the betterment of polluting areas and route monitoring.

4.5 Implementation Phases

4.5.1 Data Collection

For the detection of contaminating toxic gases, it is necessary to outsource the sensors with greater accuracy that is available and inexpensive, keeping in mind the budget-friendly factor. After sourcing the sensors, it is vital to narrow down the suitable methodologies that will lead the proposed system to its outcome. To determine the most-fitted procedure, it is better to do a literature review to execute the goal and fix the steps. The stated process above will be executed according to the proposition.

According to the proposition design, the sensors will be mounted to multiple cars as nodes. When the car is in motion, the device takes readings from sensors every minute and uploads the data to the website with the location and time stamp. Whenever the vehicle is in idle position, the set of data is only taken a few times an hour for data efficiency. When a car is within the coverage area of an available WiFi hotspot, all data is uploaded to the server, processed and published on the SensorMap portal. Given a sufficient number of nodes and diverse mobility patterns, a detailed picture of the air quality in a large area will be obtained at a low cost. These readings along with temperature and relative humidity data and GPS information are stored on the website. Aerosol Optical Depth (AOD) 550 nm data will be collected from Aqua Satellite for those locations not covered by sensors and Geospatial weather data from Visual Crossing weather data bank.

4.5.2 Data Preprocessing

In our experiment, We started with data cleaning during the data preprocessing phase. Initially, we identified all the missing data, noisy data, and outliers caused due to equipment malfunction and inconsistency with other recorded data. After identifying, we removed all missing data and outliers.

To accomplish the data integration, the acquired AOD 550 nm data from the aqua satellite, PM2.5 data from the ground station, and geospatial weather data were merged into a single coherent csv file. Data value conflicts such as different scales were removed during the extraction of the data as all the data were extracted in British Units. Later on, the data were split into a 70:30 ratio respectively for training and testing datasets.

Finally, data transformation was performed. Except for the AOD 550 nm data, all the integrated data was normalized by decimal scaling to two decimal points. Decimal scaling the AOD 550 nm impacted heavily in the prediction model thus it was discarded from decimal scaling.

4.5.3 Reporting

The demonstration will have two main parts. First, the prototype hardware platform will be shown and its operation demonstrated. Secondly, the visualization of sensormap. The following area wise reporting are shown in the below figures.

The most important part of our proposed cloud-based Big data analysis-driven air quality monitoring system (CAQ) is to generate atmospheric maps or graphical reports. Sensor mobility is handled using the SensorMap mobile proxy feature. The overall air quality will be displayed in the form of contour maps utilizing image overlays. The time-series data for a given sensor or a given geographic location will be also available.

4.5.3.1 Country wise overall monitoring

We have generated atmospheric maps using demo data. Our proposed system CAQ will generate atmospheric maps for states, industrial areas, and transportation routes for both land and marine which are presented below. Our generated atmospheric map for all the states of Bangladesh is presented in Figure 4.5.3.1.

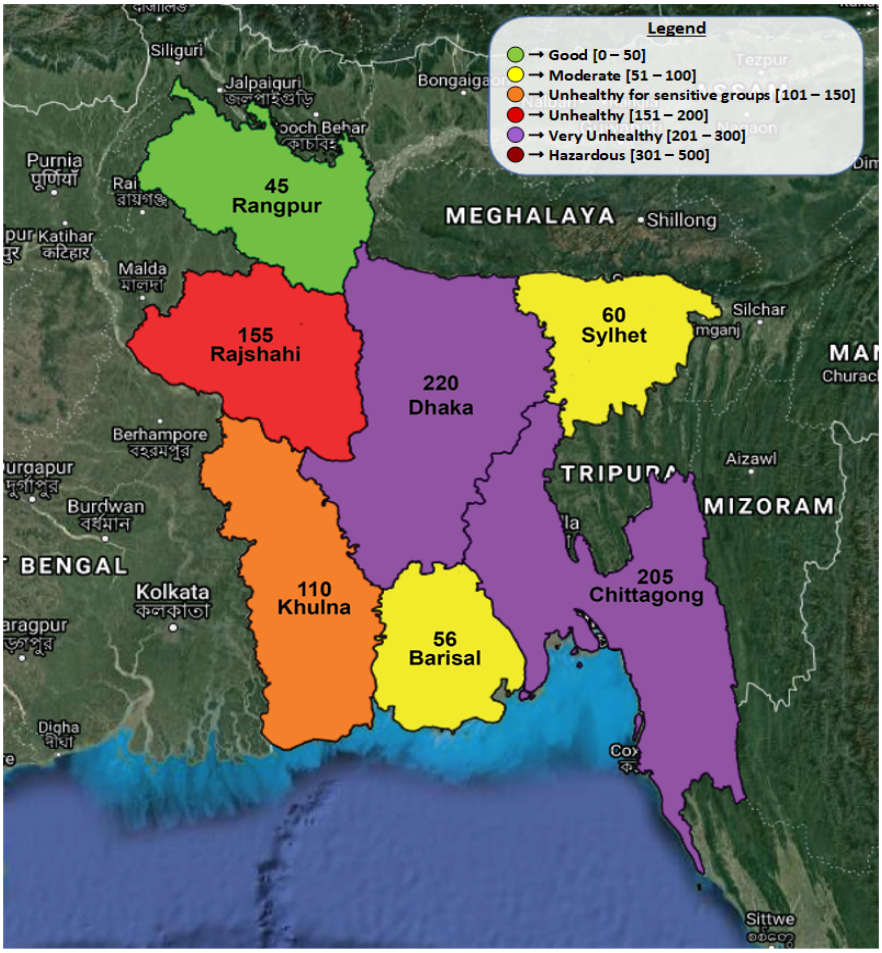

Figure 4.5.3.1 : Country wise monitoring

In Figure 4.5.3.1, air quality index (AQI) is shown for all the states of Bangladesh such as Dhaka, Sylhet, Chittagong, Barisal, Khulna, Rajshahi, and Rangpur. All the states were color coded according to their individual mean AQI. We have also generated a demo atmospheric map for Tongi industrial area of Dhaka City which is presented in Figure 4.5.3.2.

4.5.3.2 Specific Region Monitoring (for industrial areas / power plant)

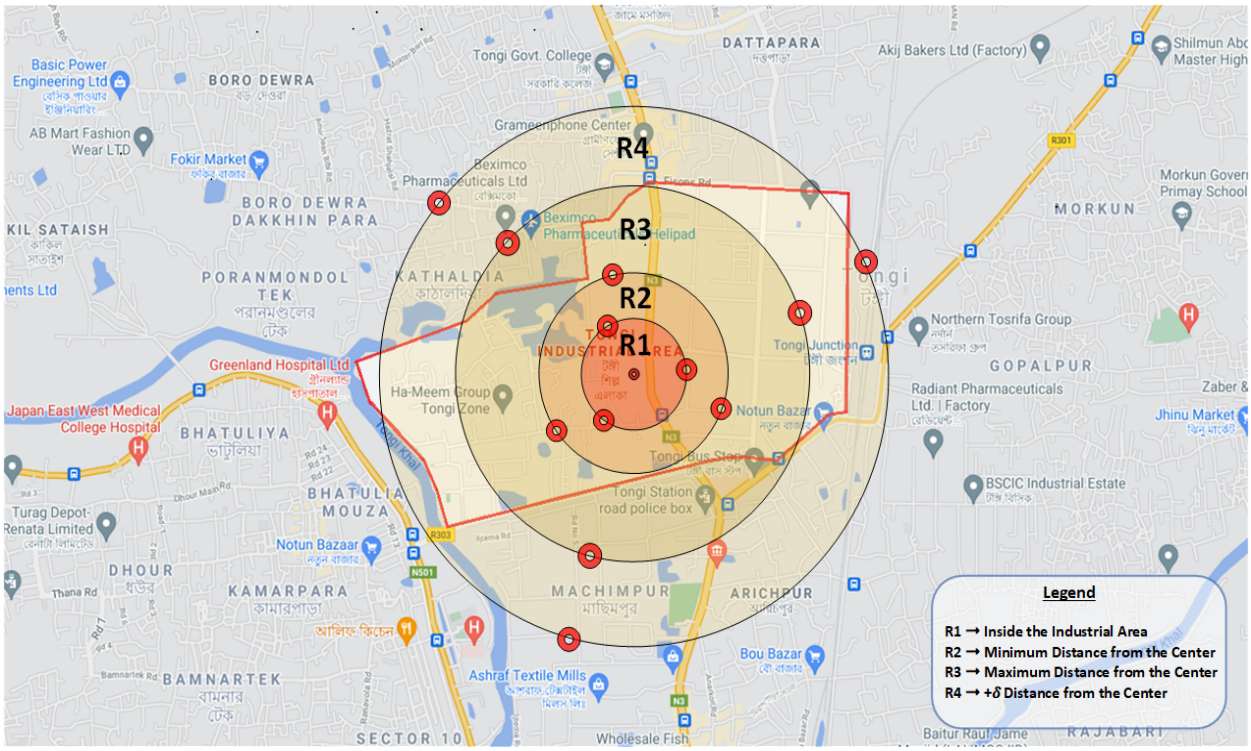

Figure 4.5.3.2 : Industrial Area monitoring

In Figure 4.5.3.2, R1 represents the center of the industrial area where the level of pollution is highest, R2 represents minimum distance from the center where all the sensor nodes are inside the industrial area, R3 represent the maximum distance from center where most of the sensor nodes are outside the industrial area, and R4 represent the $+\delta$ distance away from the center where all the sensor nodes will be outside the industrial area. We have divided the industrial area in circular areas and plan to place three sensor nodes in every single circle which will be used to read the AQI.

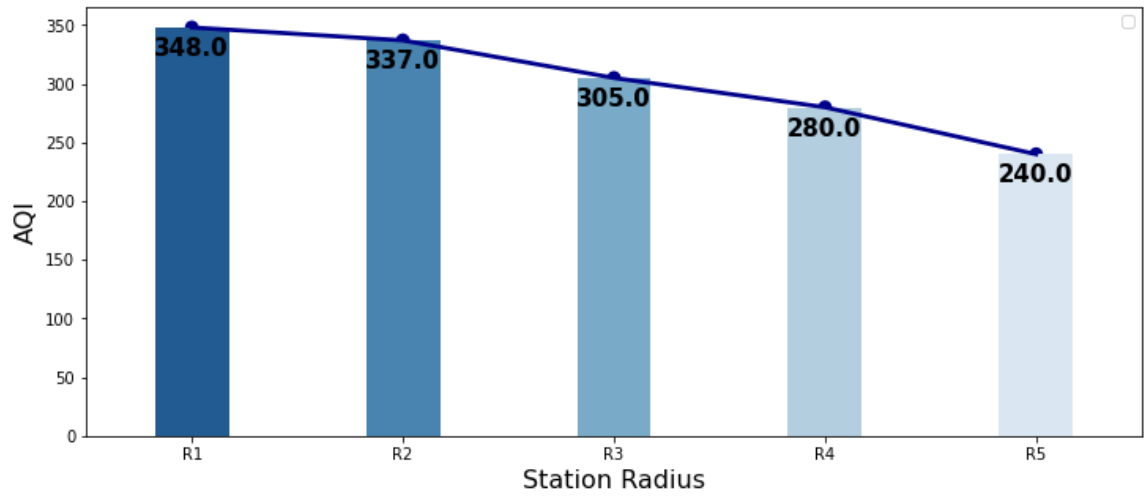

Figure 4.5.3.3 : Bar Chart of Mean AQI Per Station Radius

Figure 4.5.3.3 depicts the mean AQI per station radius for the industrial area which is shown in Figure 4.5.3.2. The AQI is highest in the center R1 and is falling as distance increases, indicating that air quality is improving.

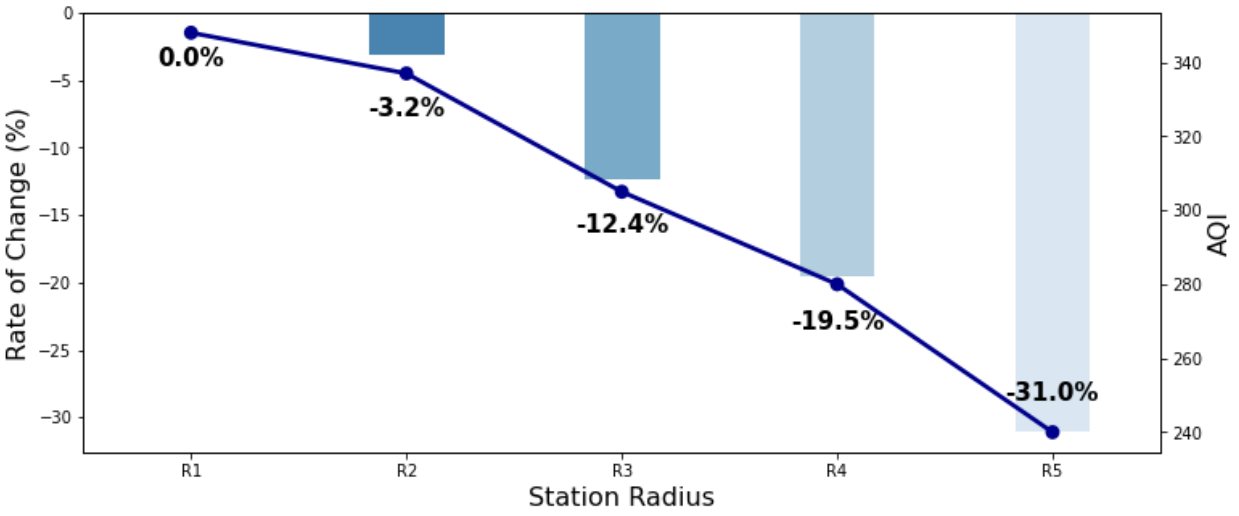

Figure 4.5.3.4 : Change of Rate of AQI Per Station Radius

Figure 4.5.3.4 shows the rate of change in the AQI per station radius of the industrial area depicted in Figure 4.5.3.2. The rate of change in the AQI increases negatively as distance increases.

4.5.3.3 Transportation Route Monitoring

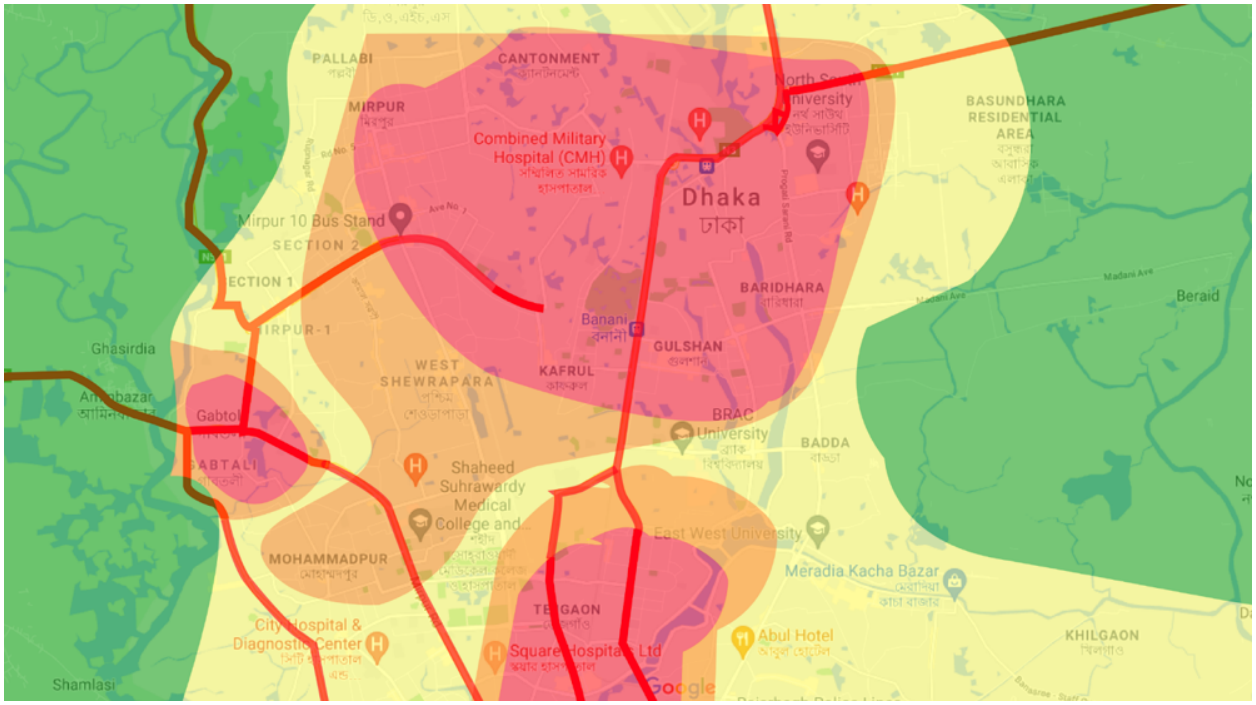

Figure 4.5.3.5 : Transportation Route Monitoring

We have also generated a demo atmospheric map for transporta-tion routes of Dhaka city which is presented in Figure 4.5.3.5 with different AQI-based color codes representing the AQI of the areas. Our proposed cloud-based air quality monitoring system (CAQ) will generate these types of atmospheric maps and charts for different states, industrial areas, and transportation routes across the country which will monitor and ensure the air quality of the country.

4.6 Result and Discussion

The purpose of making the IoT-based device is to present a cheaper option and easily accessible to the citizens of the country, so that they become alert of the alarming situation which is not being addressed yet. Our IoT-based device detects many of the pollutants like PM1.0, PM2.5, PM10, CO, CO2, NO2, Volatile Organic Compounds like Ethanol and helps us keep track of temperature and humidity.

However, in this experiment, the PM2.5 data were separately monitored to determine the pollutant's concentration level in our country for a one year time span. To observe the data, we incorpo-rated multiple PM2.5 detecting sensors by mounting them over the local transports which travel only through that specific route throughout the day. The recorded data is later uploaded to the cloud for each transportation stop, also known as stations. In the following parts of the discussion, the results are examined and user interfaces have been developed using

Python for analyzing station-wise, hourly, monthly, and season-wise data patterns. Like every other experimental result, some outliers have been distinguished into the data patterns - which we will be covering in the following research below.

4.6.1 Selected Transportation Route Map

With the help of the PMS50003 sensor, we recorded the PM2.5 concentration data in a busy transportation route shown in Figure 4.6.1. Busy transportation routes are chosen to predict the worst case scenario in our country. Multiple stations are placed on the pathway for extracting the PM2.5 data and determining the current condition of the alarming situation.

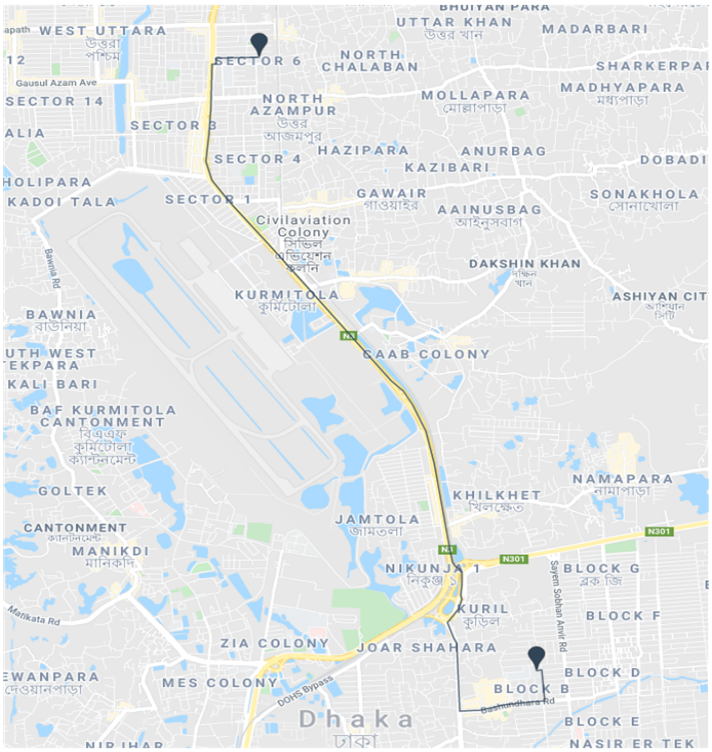

Figure 4.6.1 : Selected Transportation Route Map

4.6.2 Transportation Route Map Showing Mean PM2.5 of Each Stations

The PM2.5 data which have been kept under surveillance for a year was derived to a mean value found within a certain radius for each station, as shown in Figure 4.6.2. It is very evident in the data pattern that the most active local vehicle circulation has the highest amount of polluting compound.

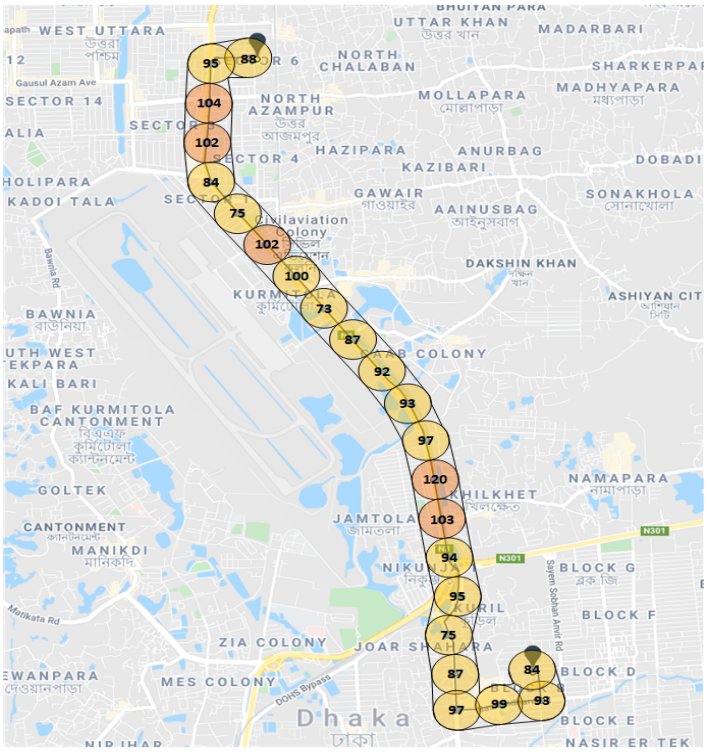

Figure 4.6.2 : Transportation Route Map Showing Mean PM2.5 of Each Stations

4.6.3 Data Distribution of PM2.5

Figure 4.6.3 shows the data patterns in PM2.5 concentration readings found in the busy transportation route. The data was taken for one year continually round the clock. We may deduce from the distribution of PM2.5 data that the data is positively skewed. The majority of the PM2.5 concentration data falls on the lower bound in this positively skewed distribution, although a rise in PM2.5 concentration during the daytime leads the distribution to be skewed positively.

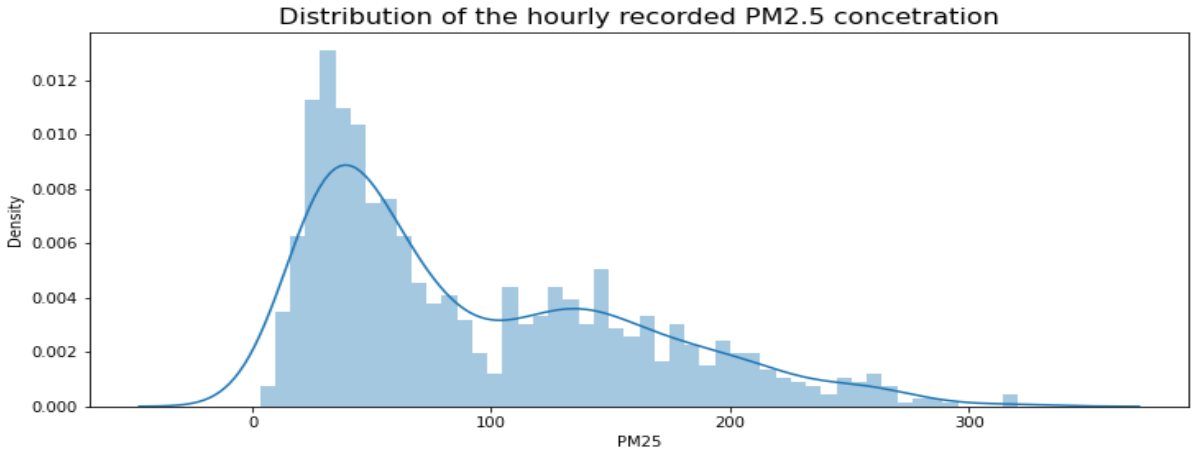

Figure 4.6.3 : Data Distribution of PM2.5

4.6.4 Box Plot User Interface

Figure 4.6.4 presents the user interface for the Box Plot Diagrams portraying the PM2.5 concentration behaviour in Bangladesh. The dropdowns depend on the time chosen by the user on day, night, hourly, or monthly basis and the users are independent to choose the season of their choice such as Winter, Spring, Summer, Autumn, All Season.

Figure 4.6.4 : Box Plot User Interface

4.6.5 Division-Wise Time Based User Interface

A division wise time based user interface for PM2.5 concentration behaviour in Bangladesh has been presented in Figure 4.6.5. This user interface displays PM2.5 for eight divisions such as Dhaka, Khulna, Barisal, Chittagong, Rangpur, Rajshahi, Sylhet, and Mymensingh over the years.

Figure 4.6.5 : Division-Wise Time Based User Interface

4.6.6 Box Plot of Station-Wise PM2.5 Data

Since the PM2.5 data are being updated to the cloud according to the station, a station-wise box plot diagram demonstrated in Figure 4.6.6 has the most upper extreme and upper quartile in the most busy routes.

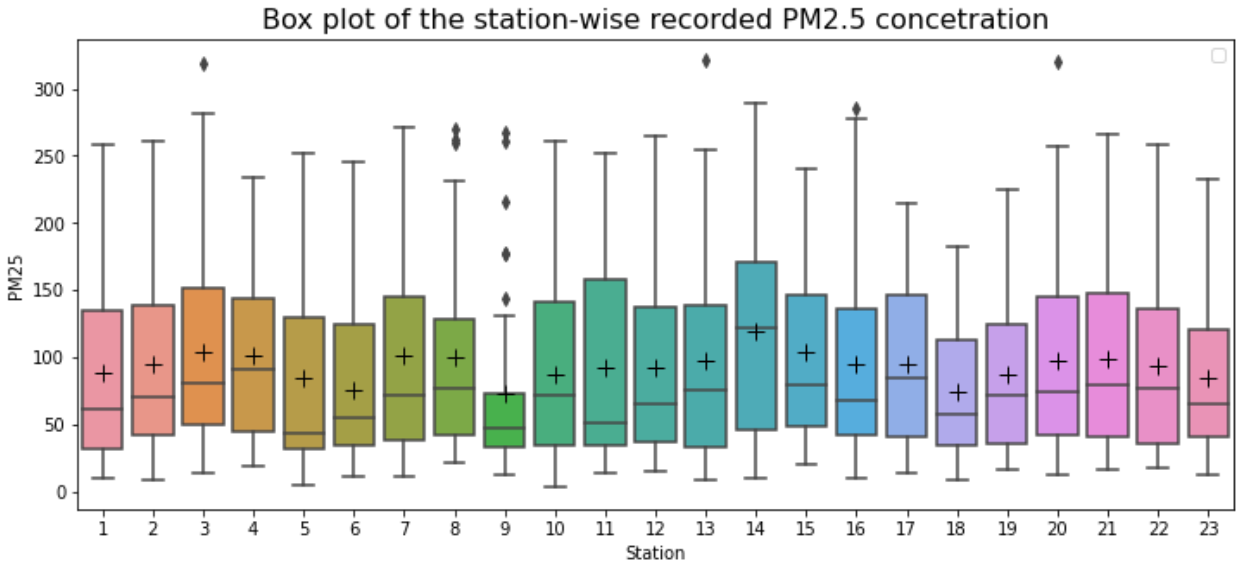

Figure 4.6.6 : Box Plot of Station-Wise PM2.5 Data

4.6.7 Box Plot of Hourly PM2.5 Data

Figure 4.6.7 indicates the hourly box plot diagram against PM2.5 and it has come to a surprise that the most generation of PM2.5 occurs at the earliest time of the day, evening and night-time. In the morning time, the routes are mostly used by citizens, during the evening working citizens take the road and at night-time, the paths are mostly occupied by trucks and pickup vans for carrying goods.

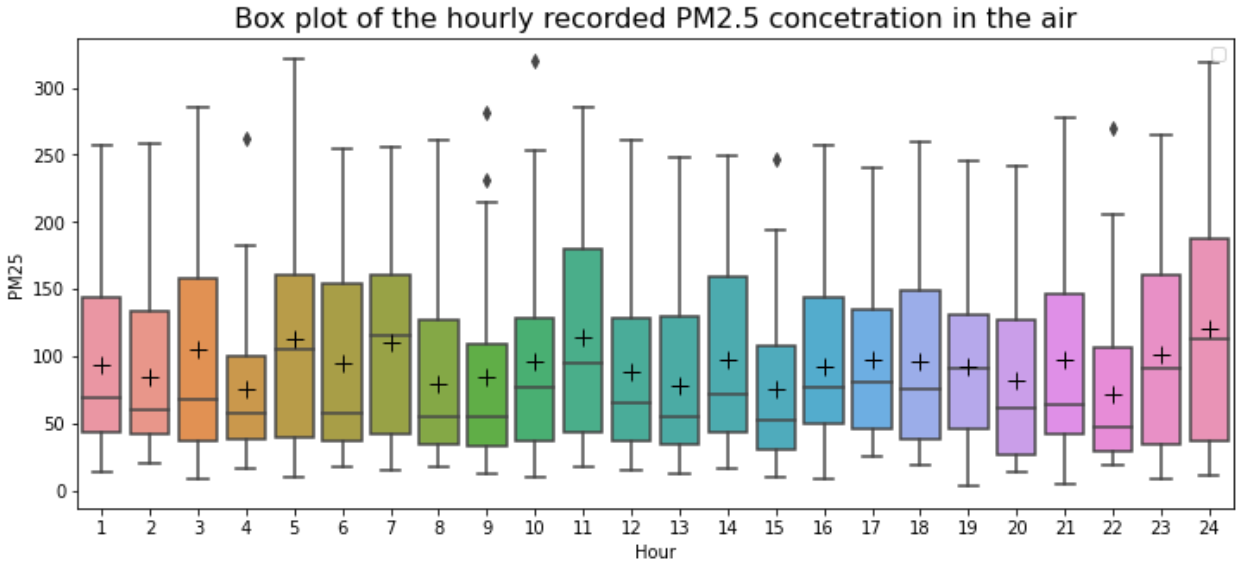

Figure 4.6.7 : Box Plot of Hourly PM2.5 Data

4.6.8 Box Plot of Monthly PM2.5 Data

The monthly box-plot diagram in Figure 4.6.8 explains that the maximum production of PM2.5 occurs during January, February, March, November and December since the winter and autumn seasons last within those months. The rest of the months have less concentration of this pollutant due to having the most amount of humidity and rain.

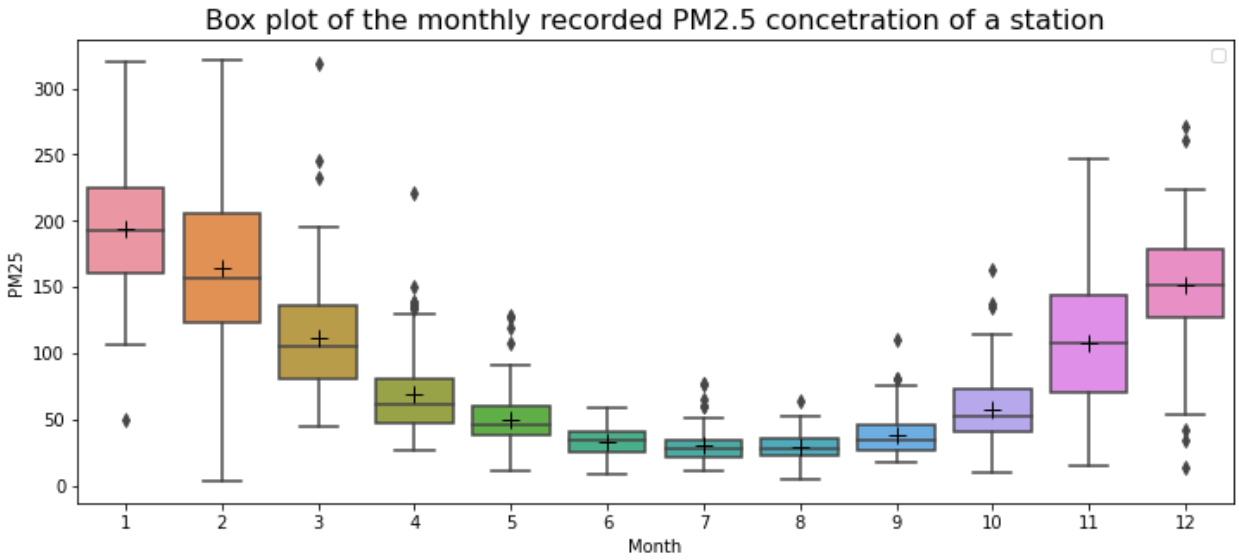

Figure 4.6.8 : Box Plot of Monthly PM2.5 Data

4.6.9 Box Plot of Season-Wise PM2.5 Data

Figure 4.6.9 validates our point placed in Figure 4.6.8 where we justified how humidity and higher precipitation are responsible for less generation of PM2.5. The lower wind speed and shallower boundary layer height causes less substantial amounts of PM2.5.

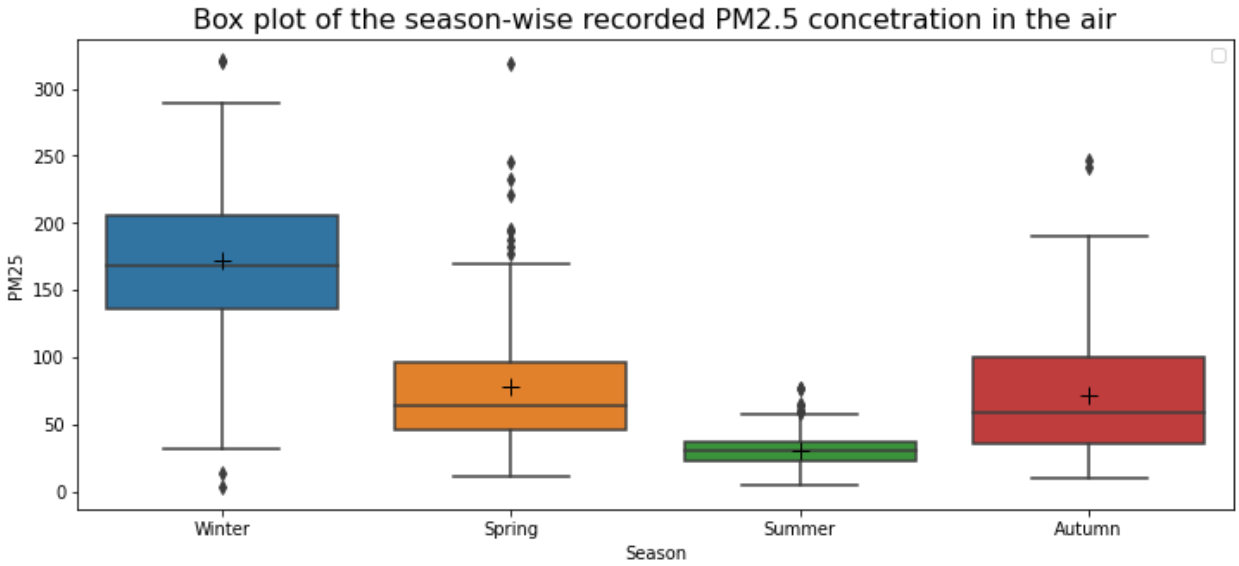

Figure 4.6.9 : Box Plot of Season-Wise PM2.5 Data

Chapter 5 : RefinAir

5.1 Overview

The main concern of this project is to focus on IoT based devices and to distinguish if the IoT based device is as equal as the industrial graded device, by looking into the data and evaluating accuracy of the device that we have made, and provide a low cost solution. Collation of data is determined to be done to examine the accuracy of the data retrieved from the IoT based device. Errors are measured and compared to justify that our IoT based device is as good as the industrial ones. Sidewise, heading to the goal of finding an inexpensive solution, we will be learning more about the advanced air monitoring technology used by AirVisual Pro and explore the reasons why this device is chosen.

In 1963, the history of cleaning air and improvisation of AQI started from the brothers, Manfred and Klaus, who created residential air filters in Switzerland that are now used by more than 70 countries. This device has received applause from customers and air quality experts around the globe. It is known to be used in Olympics because it refines 99.5% of all the ultrafine particles and is proven to deliver medical-grade air using HyperHEPA technology. The 3D ultraseal is designed to eliminate air leakage.

IQAir Earth is the first ever 3D air pollution map. It delivers indoor and outdoor monitoring systems - providing hourly weather and air pollution forecasts and alerts if air becomes unhealthy. The world’s first air quality application was introduced by IQAir, which gives real-time data and shows historic air quality information. to be compared with the IoT based device, RefinAir.

5.2 Problem Statement

This section of the research depicts the problem statement of the project, which has encouraged us to go on with the project. According to research, the average annual PM 2.5 concentrations in Bangladesh were 77.1 microgrammes per cubic metre (mcg/m3) of air, which is seven times higher than the WHO exposure guidelines, with Dhaka standing second among 106 countries. Investigators from IQAir, a worldwide air quality information and technology enterprise, evaluated pollution data from 106 nations, specifically detecting PM2.5, a microscopic pollutant that can pose serious health concerns.

To eliminate the climate change crisis, we have chosen a set of sensors with a microcontroller to detect the concentration of pollutants and monitor the data patterns. It will be turned to an IoT based device after the set of sensors are connected to the development board. The data from IoT based devices will then be compared with the industrial-graded device to determine the accuracy level of the experimental device.

Devices for air monitoring are available in the market but they are very expensive and not easily accessible in case of maintenance. The sustainability of these devices are also questionable, even if they attain industrial accommodation. Third world countries like Bangladesh need to attempt serious air monitoring strategies to avoid future catastrophe. Keeping in concern to this matter, our IoT based air monitoring device, RefinAir, is introduced with which we will be comparing the industrially-graded device, AirVisual Pro, to distinguish the device compatibility, durability and sustainability..

5.3 System Analysis

5.3.1 Rich Picture

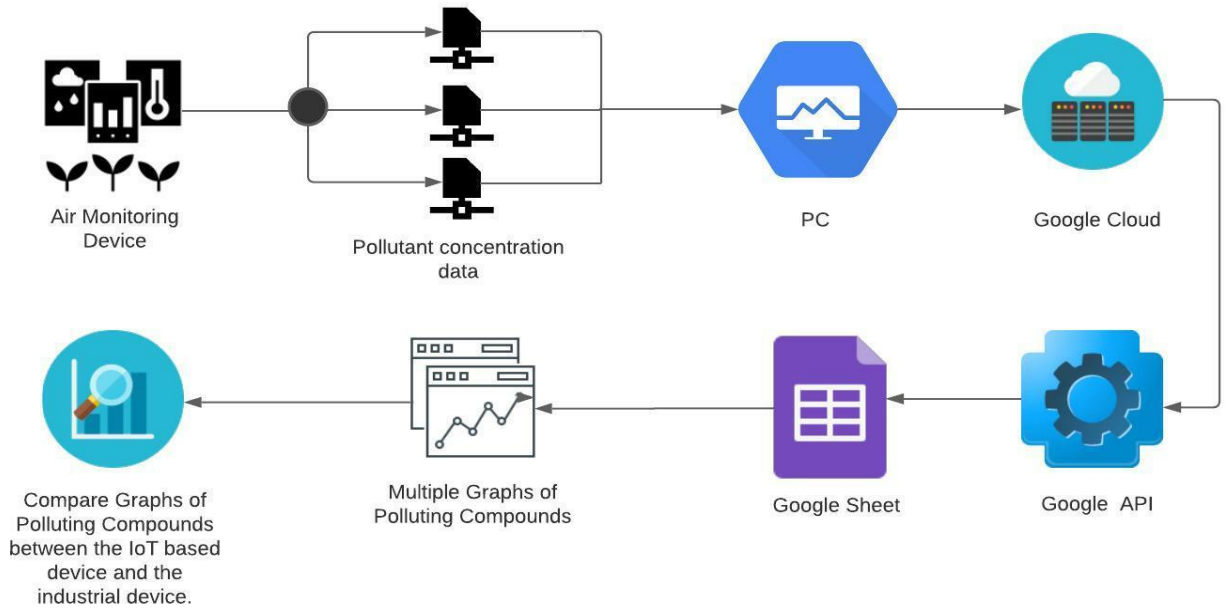

Figure 5.3.1 : Rich Picture for RefinAir

5.3.2 Six Elements Analysis

Stakeholder	Non-Computing Hardware	Computing Hardware	Software	Database	Communication & Network
I. Researcher • Does research and chooses which polluting compounds to mon itor. • Does literature review. • Reads multiple articles on finding sensors that are as good as industrial graded sensors. • Sources the appropriate hardwares. • Assembles the hardware and turns into a device. • Contacts faculties for work review and suggestions. • Makes sure that the device migrates data to PC through serial monitor and to the 'Air monitoring Software'. • Makes sure the data is stored into a database. • Contacts the Machine Learning Engineer and faculties to get help about machine learning approach. **II. Hardware Vendor** • Receives order for hardwares from customers. • Sends an order package of hardwares to their customer. • development purpose. **III. Maintenance Team**	**I. Sensor** • All the air polluting compound detecting sensors were included to fetch their concentration data. • Communication based sensors were used to send the data to the PC and software. **II. Microcontroller** • This is the brain of the device. All the sensors are connected to it. • Sends and receives signals. • Follow the commands written in the code. • Stores a limited amount of data. **III. Connecting Wire** • Creates a pathway that connects the sensors and the microcontroller in order to work the device. **IV. Breadboard**	**I. PC** • Helps to do the research. • Helps to code for making the device work. • Helps monitor the fetched data from the serial monitor. • Helps in order to make the software. • Helps to preprocess data. • Migrate data to the database. • Helps to implement multiple MLA for prediction. **II. Server** • The servers help store data into the database.	**I. Arduino** • Receives command through code • Executes the commands into the hardwares. **II. Google Chrome** • Helped do research on the topic. **III. Microsoft Excel** • Data was entered here. • Data preprocessin g was done. • Scatter diagram was created. • Correlations were done.	**I. Flash Memory in Arduino** • Retrieved data from sensors are stored in a text file. **II. Google Sheet** • Fetches pollutant data from the devices. • Pollutant data is stored here. • Comparison of data using graphs is done between the outcome of two platforms.	**I. Communi cation Module** • Sends data from the device to PC and software. **II. ISP** • Provides internet connection to all the necessary activities that need internet support.

• Fix and maintain electrical equipment. • Listens to customers' problems and fixes them. **IV. Customer** • Purchases the device and uses them.	• Helps to build and test circuits.				

Figure 5.3.2 : Six Element Analysis for RefinAir

5.3.3 Process Diagram

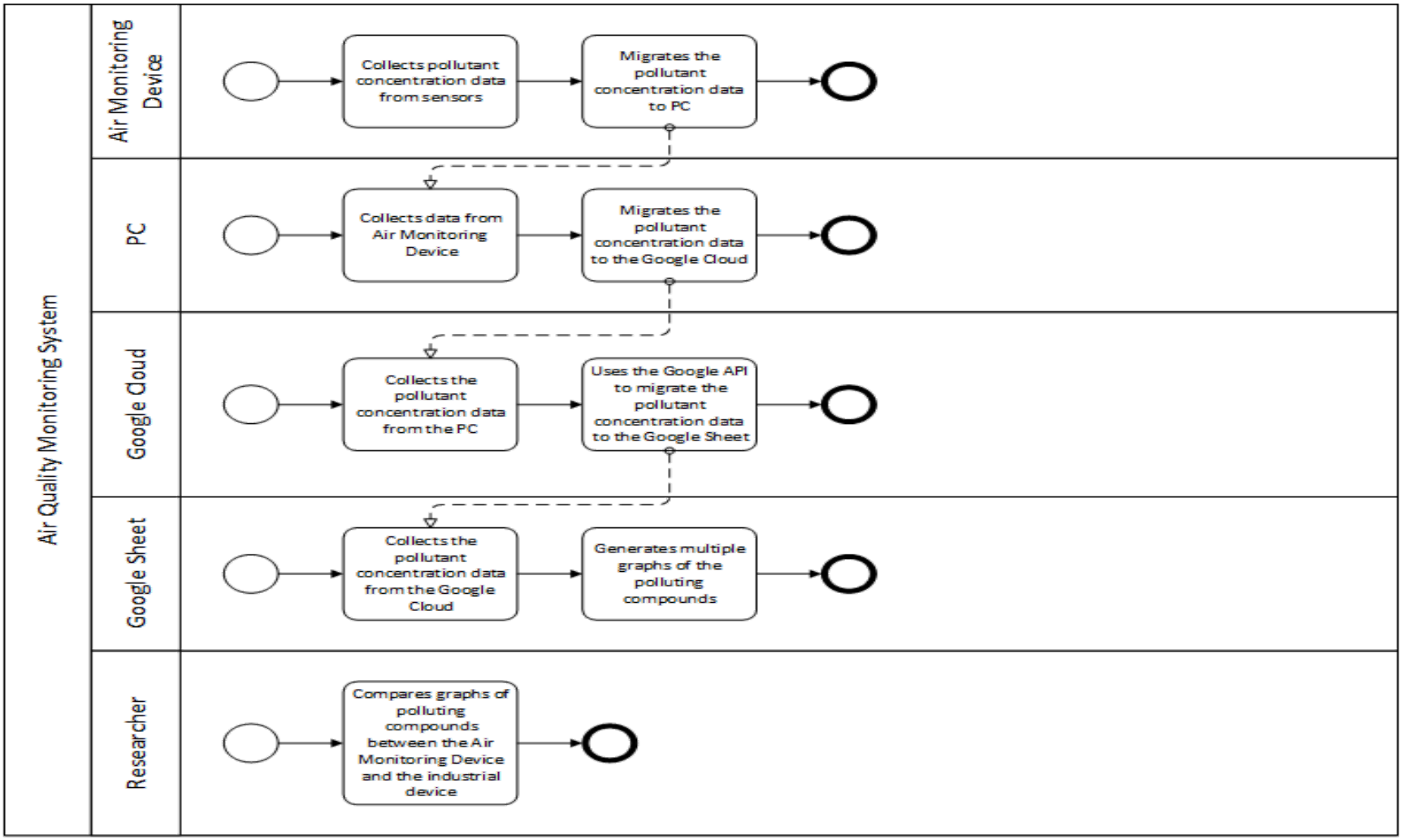

Figure 5.3.3 : Process Diagram TO BE for RefinAir

5.5 Proposed System Description

5.5.1 Component Compatibility Comparison with other Sensor

The purpose of making this IoT based device is to present a cheaper option and easily accessible to the citizens of this country, so that they become alert of the alarming situation which is not being addressed yet. However we will be discussing the components that are being used in the device along with the proposed solution.

5.5.1.1 Arduino Mega 2560 WiFi

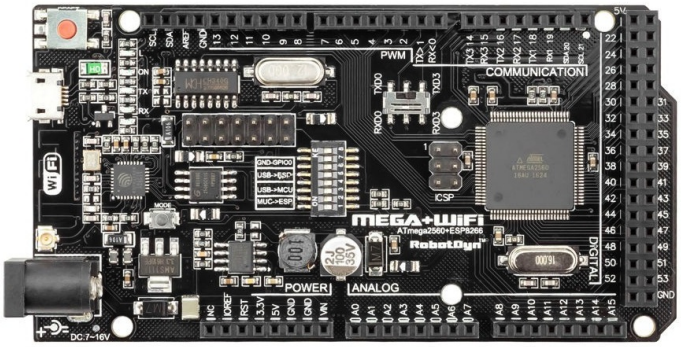

Figure 5.5.1.1 : Arduino Mega 2560 WiFi

Arduino Mega 2560 WiFi is the microcontroller of the system and a development board with a facility of inbuilt wifi system, which helps users to transmit data easily and store it to a database for future research purposes. This board also has a selector switch that allows the ESP to interleave the connection between TX0 and TX3, remembering that the ATmega has four serials. A second selector switch is the DIP Switch, and we also have a key recording mode of the ESP8266. All the pinning is completely compatible with the ATmega pinout. It has a memory flash of 32mb and the Esp8266 incorporated into the board has a memory of 8mb. The input voltage for ATmega 2560 are 5V/7-12V. The price for this microcontroller was BDT1,372. NodeMCU was not chosen to work with in this project because it has less voltage, which is of 4.5V-10V, than Arduino Mega 2560 WiFi - making it less efficient. Also it has a flash memory of 4MB/64kB only.

5.5.1.2 MQ 7 sensor

Figure 5.5.1.2 : MQ 7 sensor

This Carbon Monoxide (CO) gas sensor detects the concentrations of CO in the air and outputs its reading as an analog voltage. The sensor can measure concentrations of 10 to 10,000 ppm.The sensor can operate at temperatures from -10 to 50°C and consumes less than 150 mA at 5 V. This sensor is highly sensitive to CO and has a stable and long life. Compared to the MQ 7 sensor, the CCS811 also detects CO in the air. But MQ 7 sensor more efficient than CCS811 because of it's higher voltage than CCS811 which has a voltage range of 1.8 to 3.3V. In terms of pricing, MQ 7 is still feasible because it costs less than CCS811 which is BDT 180 and BDT1,600 simultaneously.

5.5.1.3 PMS5003 sensor

Figure 5.5.1.3 : PMS5003 sensor

The PMS5003 air quality sensor, which is known to detect the concentration of PM2.5 particles, is connected with the Arduino WiFi development board through UART communication. It's TX and RX are connected with the Arduino mega's RX and TX pins. The PMS5003 transmits data of PM2.5 to the Arduino mega.

According to research, the PM-900M sensor of the Temtop M2000C,known to be the leading air quality monitoring system in the world, is used for measuring the concentration of PM2.5. This industrial graded sensor costs $29.99 and the PMS5003 air quality sensor we have used in our project is of $30.83. Both PM-900M and PMS5003 sensors have an accuracy of $\pm$10%

(100-500 ug/m³), but PM-900M has a maximum concentration range of 0-999 ug/m³ and PMS5003 has more than 1000 μ g/m³.

5.5.1.4 BME280 sensor

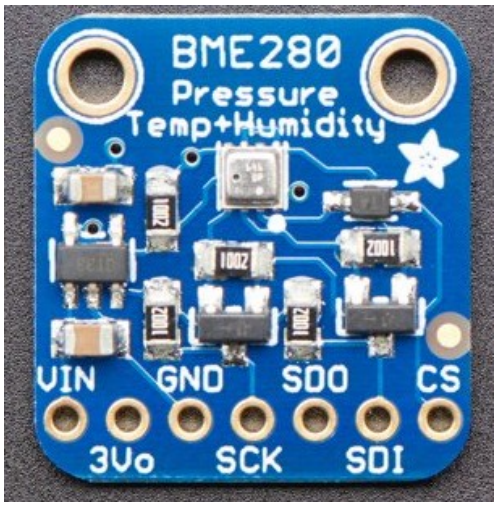

Figure 5.5.1.4 : BME280 sensor

BME280 is a product of BOSCH which is used to provide temperature, humidity and air pressure data to the Arduino mega using I2C communication, it's SDA and SCL pins are connected to Arduino mega's SDA and SCL pins.

Similar to BME280, which has an accuracy tolerance of ± 3% relative humidity , Groove TH02 temperature and humidity has an accuracy of ±0.5C. Sidewise, the Groove MPX5700AP sensor has a maximum error of 2.5% and the BME280 has sensitivity error of ±0.25%. The BME280 has cost $10.63 and the Groove TH02 has a price of $12.70.

5.5.1.5 MH-Z19B sensor

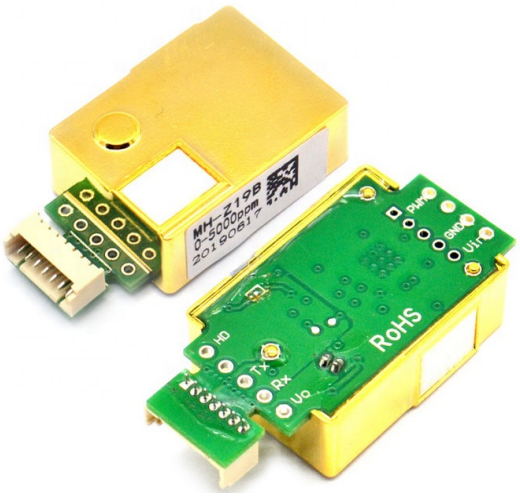

Figure 5.5.1.5 : MH-Z19B sensor

The MH-Z19B sensor is a CO2 detecting sensor, which operates in UART communication, and it's RX and TX pins are connected with the Arduino's TX and RX pins. It has three detection ranges, from 0-2000/5000/10000 ppm and the accuracy for all the three detection ranges is ± (50ppm+5% reading value), whereas an industrial certified SprintIR6S 100% CO2 Sensor has an accuracy of ±(300ppm+5% reading value). With higher accuracy comes less voltage and more current of 3.3 to 5.5V and 33mA simultaneously, leaving the SprintIR6S 100% CO2 Sensor less efficient in terms of energy loss. However, MH-Z19B sensor has more voltage of 4.5 to 5.5 V which means a decrease in current, which is 20mA, means less energy loss from resistance. The price for MH-Z19B is BDT. 3,500 and for SprintIR6S 100% CO2 Sensor is BDT 21,083.86. The weight also comes as a plus point for MH-Z19B, which is 5g, making another reason to not choose SprintIR6S 100% CO2 Sensor which weighs 16g.

5.5.1.6 Grove Multichannel Gas Sensor V2

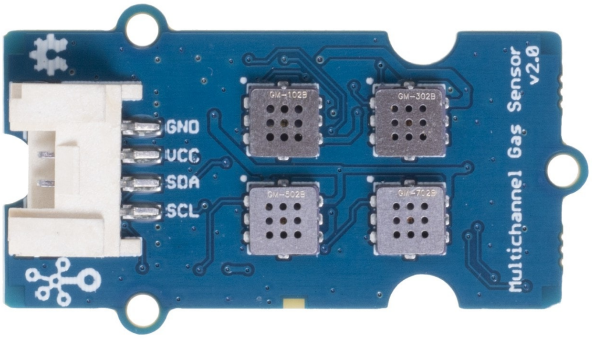

Figure 5.5.1.6 : Grove Multichannel Gas Sensor V2

Grove Multichannel Gas Sensor V2 is a sensor with four measuring units, which are sensitive to Carbon monoxide (CO), Nitrogen dioxide (NO2), Ethyl alcohol (C2H5CH), and Volatile Organic Compounds (VOC). This sensor can provide four sets of data at the same time. It operates in I2C communication, it's SDA and SCL pins are connected with the Arduino Mega's SDA and SCL pins. It has a voltage up to 3.3-5v leaving one of it's market competitor GC310 portable gas detector by Chicheng, which has a voltage of 3.7V and accuracy of <±3%, behind in case of efficiency. Grove Gas Sensor also has an advantage of being lighter in weight, which is 7g, compared to GC310 which weighs 220g. GC310 costs BDT 30,472 and Grove Gas sensor's price is BDT 3,600, which concludes that GC310 has also lost in case of cost feasibility with Grove Gas Sensor.

For the IoT based air monitoring system, Arduino Mega 2560 Wifi - which is a development board- is used as the microcontroller of the system and incorporated with the set of sensors to detect the concentration of pollutants and monitor the data patterns. It is turned to an IoT based device after the set of sensors are connected to the development board.

5.5.2 Hardware Layout

The Arduino Mega 2560 WiFi is the microcontroller of the system and a development board with a facility of inbuilt wifi system, which helps users to transmit data easily and store it to a database for future research purposes. All the sensors are connected to the development board for transmission of data. The PMS5003 air quality sensor is connected with the Arduino mega development board through UART communication, it's TX and RX are connected with the Arduino mega's RX and TX pins. The PMS5003 transmits data of PM2.5 to the Arduino mega. BME280 provides temperature, humidity and air pressure data to the Arduino mega using I2C communication, it's SDA and SCL pins are connected to Arduino mega's SDA and SCL pins. MH-Z19B is a CO2 detecting sensor, which operates in UART communication, and it's RX and TX pins are connected with the Arduino's TX and RX pins. CCS811 detects Carbon Monoxide CO and VOCs level of the air. It operates in I2C communication, it's SDA and SCL pins are connected to Arduino mega's SDA and SCL pins. Grove Multichannel Gas Sensor V2 is a sensor with four measuring units, which are sensitive to Carbon monoxide (CO), Nitrogen dioxide (NO2), Ethyl alcohol (C2H5CH), and Volatile Organic Compounds (VOC). This sensor can provide four sets of data at the same time. It operates in I2C communication, it's SDA and SCL pins are connected with the Arduino Mega's SDA and SCL pins.

Arduino Mega 2560 WiFi takes readings from the sensors. Arduino Mega connected with a PC sends all the sensor data to the PC using Serial Communication. A python application runs on the PC, which opens a serial communication between the Arduino Mega 2560 and the PC. The Python script then adds a timestamp with every single data and posts every single data simultaneously to the google cloud using RestAPIs and GoogleAPIs with very low latency. Then the inbuilt Google Sheet APIs fetches the data from the RestAPIs and adds the data into a google sheet. All the data is dynamically added to the google sheet automatically. Using the google sheet charts all the sensor data are visualized in different linear time series data charts according to the different obtained gas compounds and weather elements.

5.5.2.1 Block Diagram

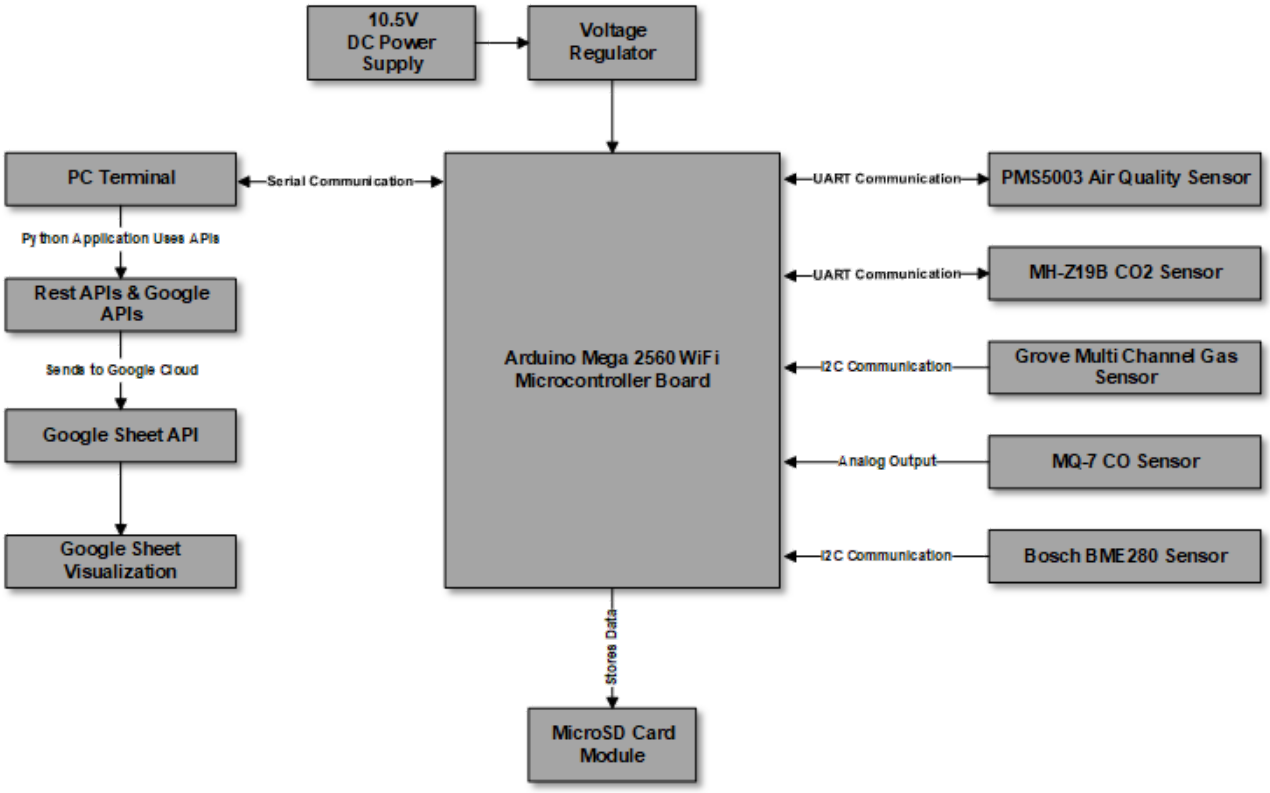

Figure 5.5.2.1 : Block Diagram of the proposed system

5.5.2.2 Schematic Diagram

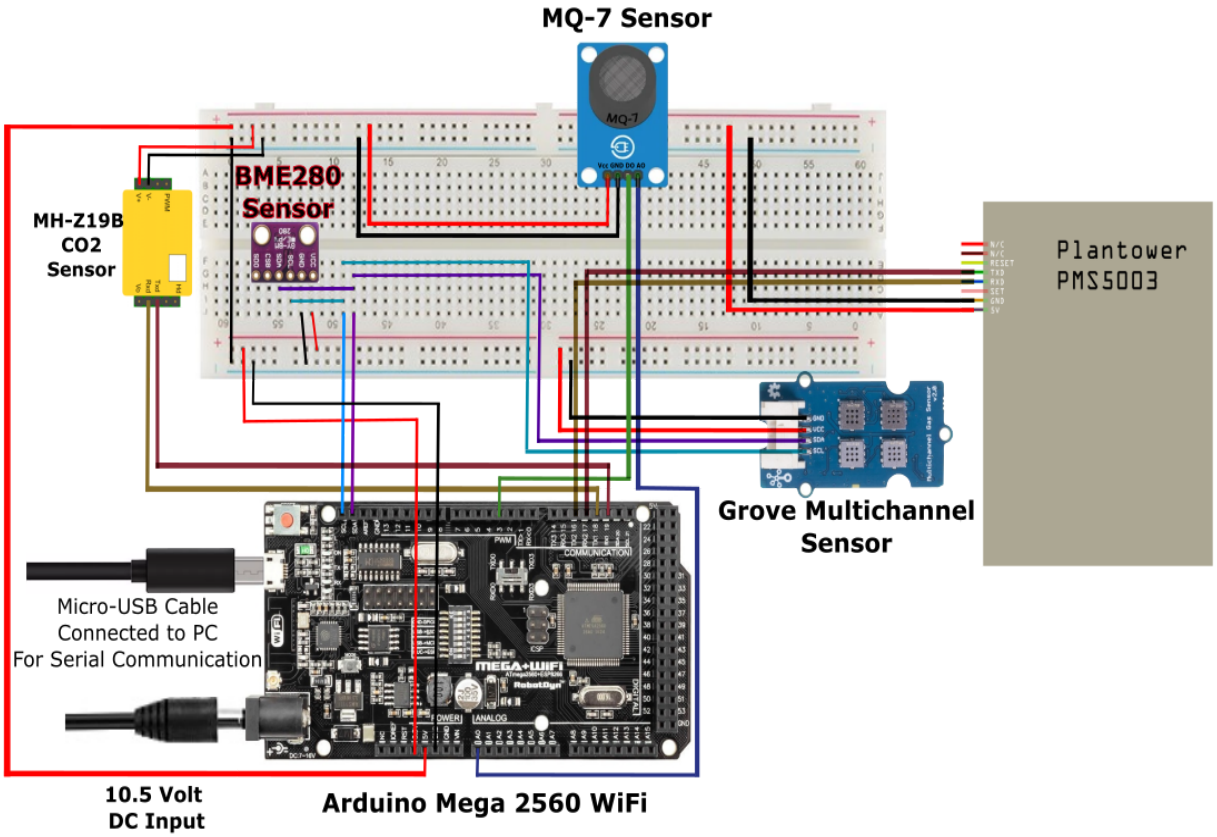

Figure 5.5.2.2 : Schematic Diagram of the proposed system

5.5.2.3 Hardware implementation of IoT air quality monitoring system

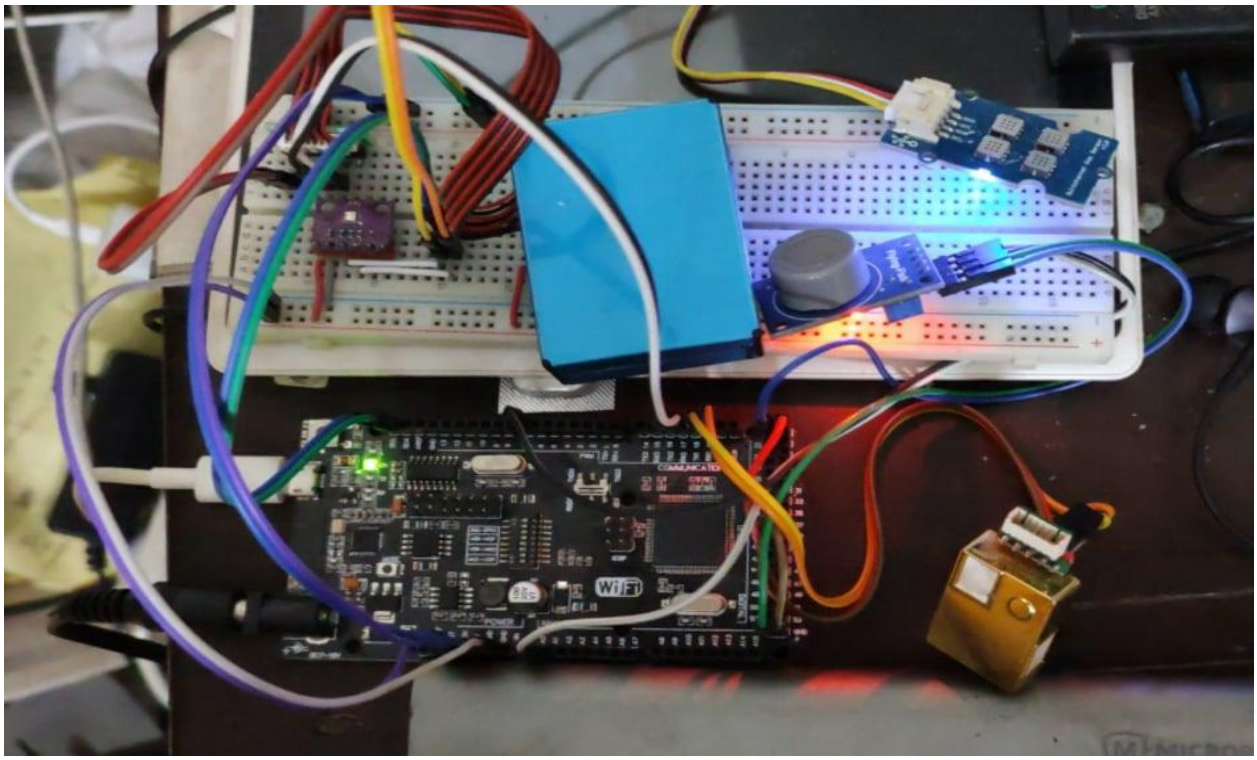

Figure 5.5.2.3 : Hardware implementation of the proposed system

5.6 Performance Analysis

This part of the report will determine the device compatibility with the help of the data that has been found from the IoT based device. The data taken from Iot based device, is generated by the sensors connected to a microcontroller. The microcontroller sends the data to the PC and through the PC the data is updated into Google Cloud. The inbuilt Google Sheet APIs fetches the data and adds data to the Google Sheet with timestamps.

On the other hand, the purchased device - AirVisual Pro- stores their data to a text file. The text file is then stored into Microsoft Excel. The data is then uploaded to the Google Cloud and the inbuilt Google Sheet APIs fetches the data and adds data to the Google Sheet.

5.6.1 Graphical Data Patterns

In this performance analysis, the collection of data of different polluting compounds are compared between the IoT based device and AirVisual Pro. Please note that all the data were collected for 8 hours after proper calibration of both AQI Air Visual Pro and our Air monitoring system device. Within the 8 hour timeline, after every 5 minute data was fetched from both the

devices automatically. Between the two devices, data difference is shown under the same category of sensors. Percentage error is done between temperature, humidity and CO_2. Particle concentration difference is between PM1.0, PM2.5 and PM10.0.

We have taken data from daytime, nighttime and have done a collation between the series of compounds to determine the sensitivity of the sensors.

5.6.1.1 Daytime Reading

The data were collected within an 8 hour period for the day time reading, and the device delivered data every 5 minutes to the Google Sheet using Google APIs.

5.6.1.1.1 Temperature Difference in Percentage

Average Difference: 0.180%

Minimum Difference: 0.129%

Maximum Difference: 0.295%

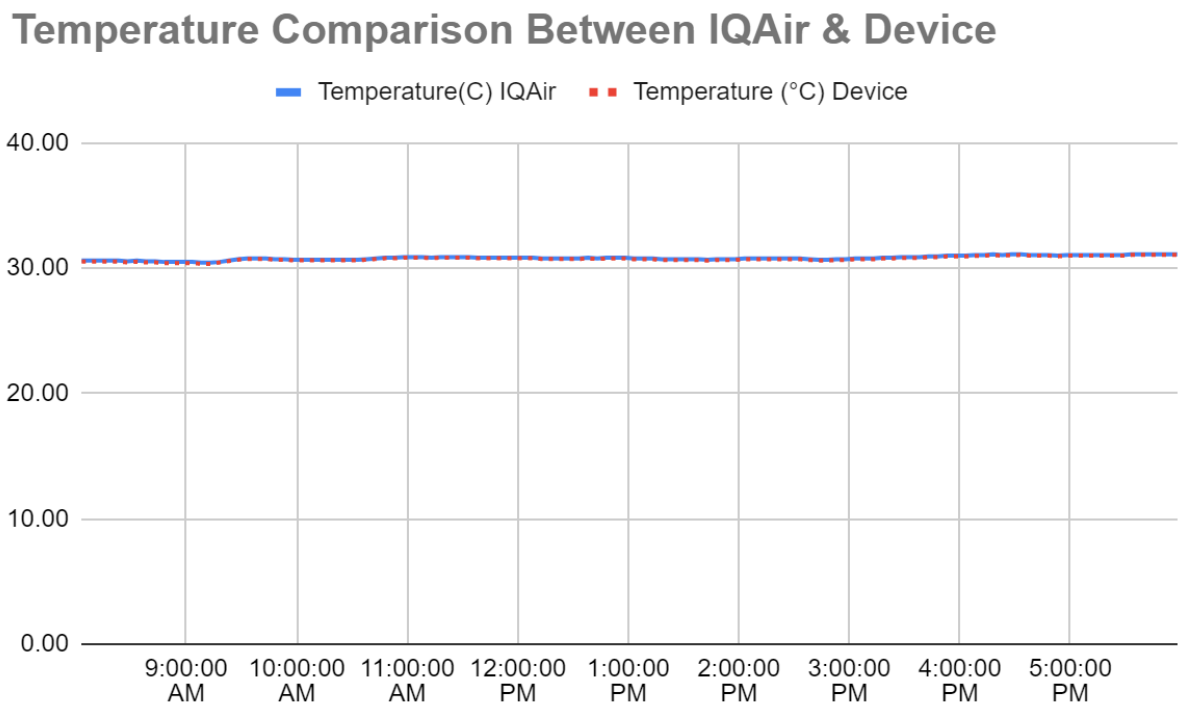

Figure 5.6.1.1.1 : Day Time Temperature Comparison.

5.6.1.1.2 Humidity Difference in Percentage

Average Difference: 0.267%

Minimum Difference: 0.126%

Maximum Difference: 0.701%

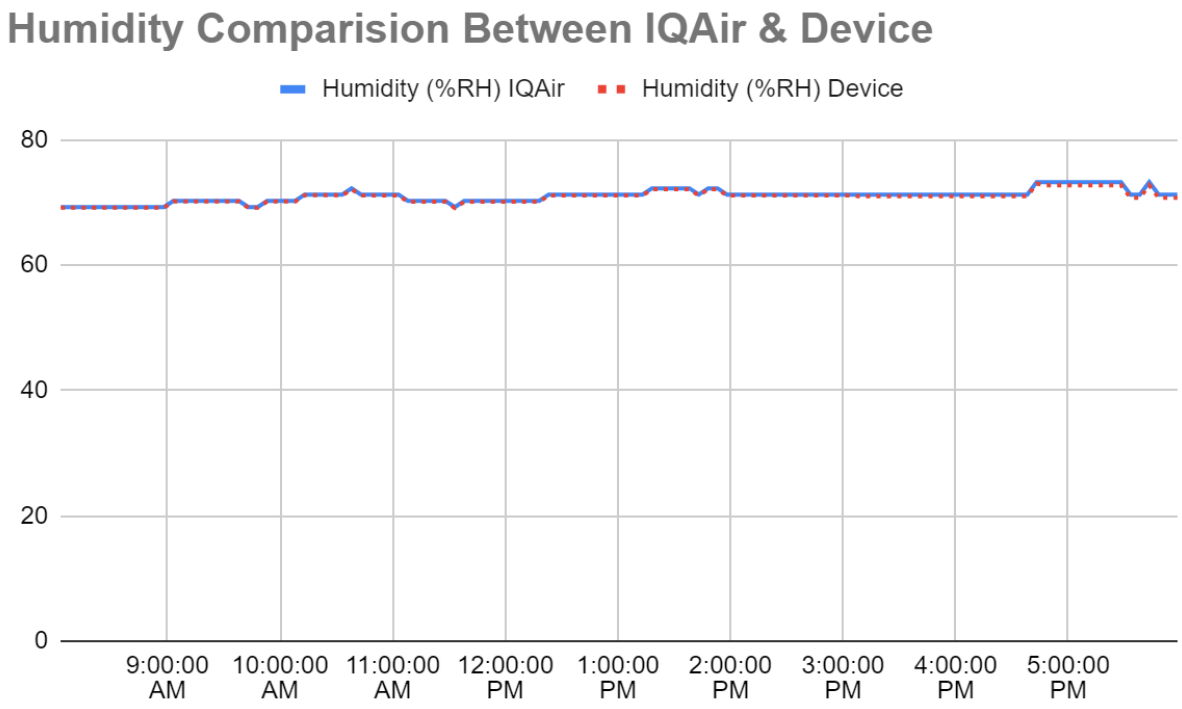

Figure 5.6.1.1.2 : Day Time Humidity Comparison.

5.6.1.1.3 PM1.0 (ug/m3) Difference in raw concentration

Average Difference: 0.712 ug/m3

Minimum Difference: 0 ug/m3

Maximum Difference: 2 ug/m3

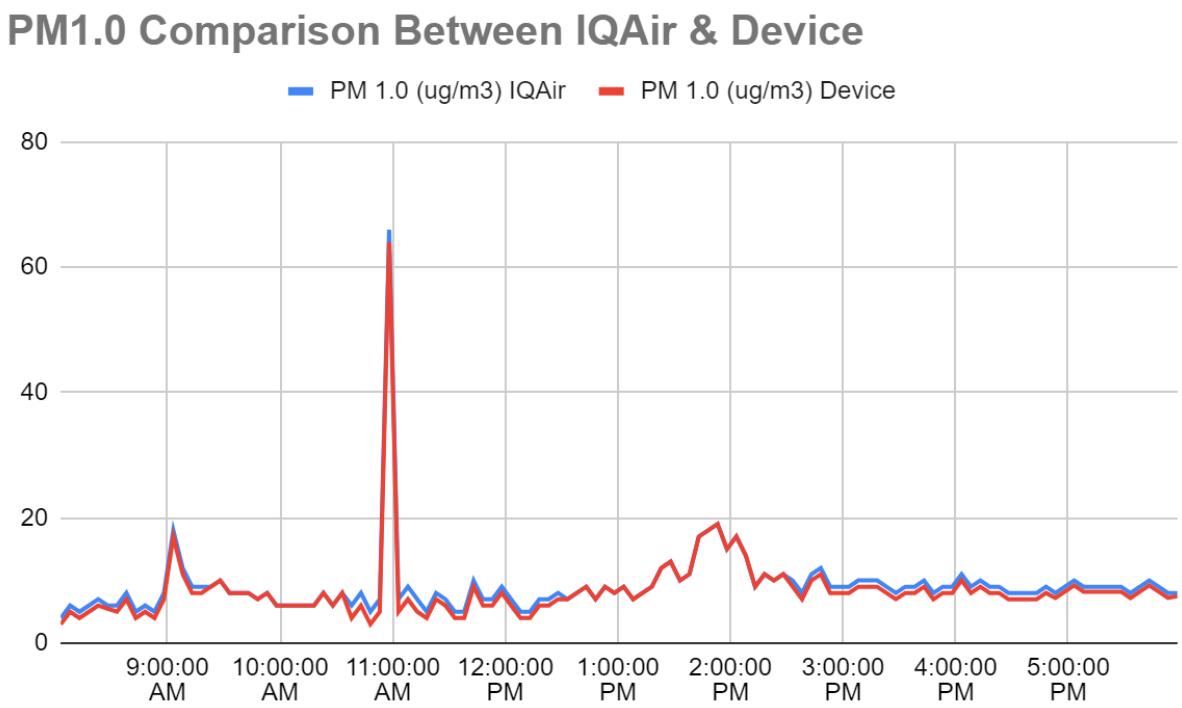

Figure 5.6.1.1.3 : Day Time PM1.0 Comparison.

5.6.1.1.4 PM2.5 (ug/m3) Difference in raw concentration

Average Difference: 0.51 ug/m3

Minimum Difference: 0 ug/m3

Maximum Difference: 1.5 ug/m3

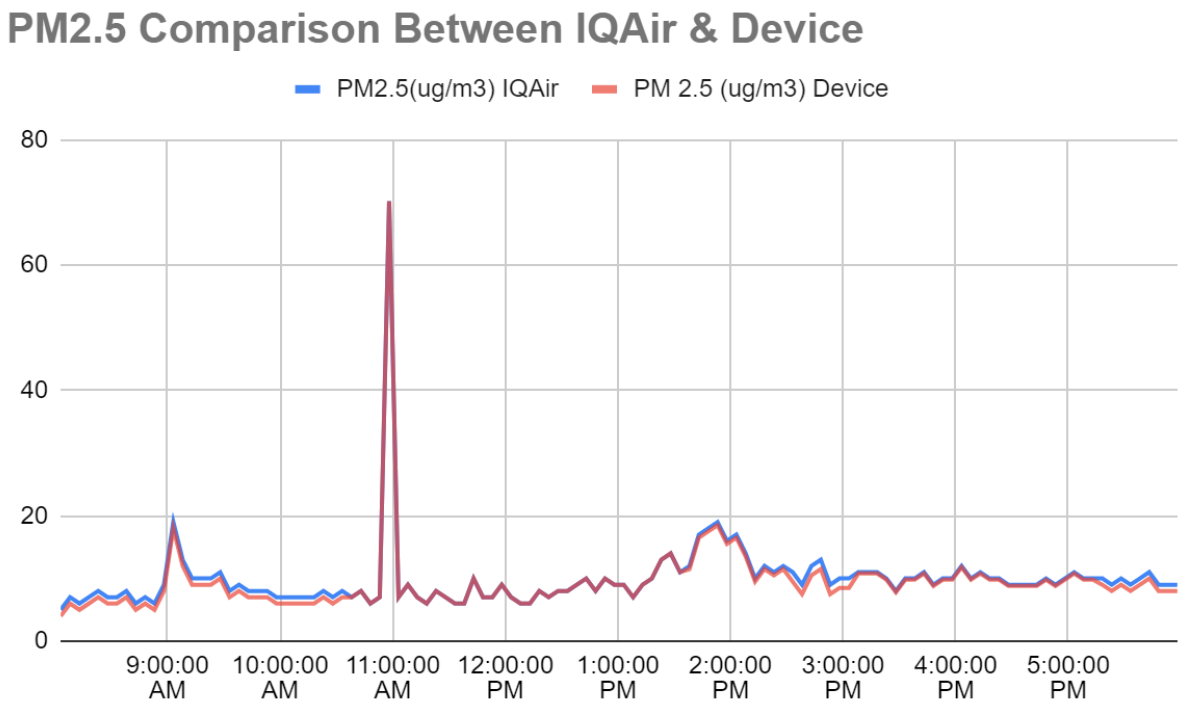

Figure 5.6.1.1.4 : Day Time PM2.5 Comparison.

5.6.1.1.5 PM10.0 (ug/m3) Difference in raw concentration

Average Difference: 0.475 ug/m3

Minimum Difference: 0 ug/m3

Maximum Difference: 2 ug/m3

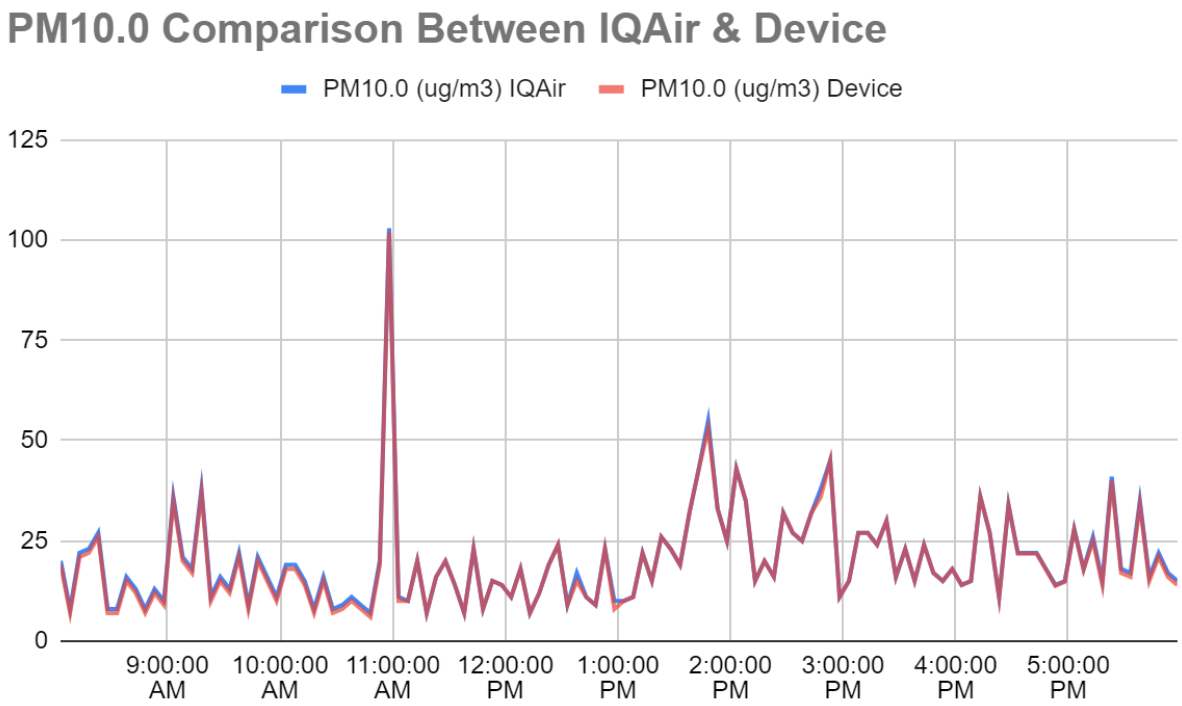

Figure 5.6.1.1.5 : Day Time PM10.0 Comparison.

5.6.1.1.6 CO_2 Difference in Percentage

Average Difference: 0.37%

Minimum Difference: 0.00%

Maximum Difference: 2.33%

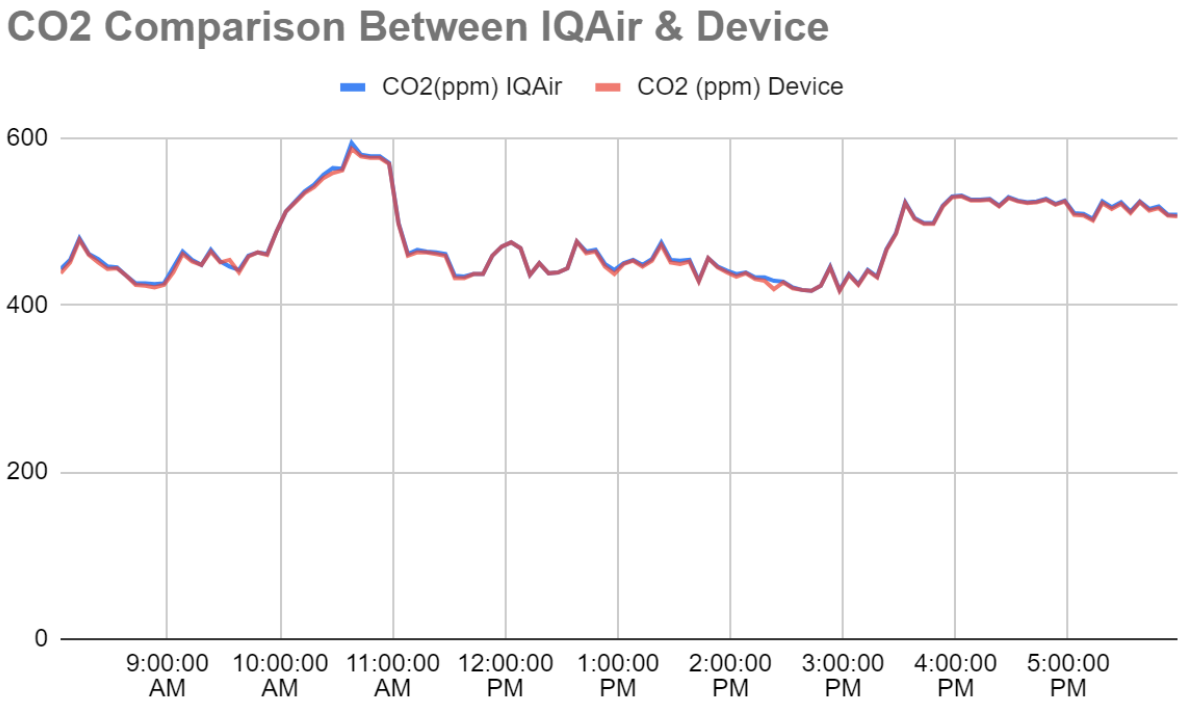

Figure 5.6.1.1.5 : Day Time CO_2 Comparison.

5.6.1.2 Night Time Reading

The data were collected within an 8 hour period for the night time reading, and the device delivered data every 5 minutes to the Google Sheet using Google APIs.

5.6.1.2.1 Temperature Difference in Percentage

Average Difference: 1.474%

Minimum Difference: 0.748%

Maximum Difference: 1.902%

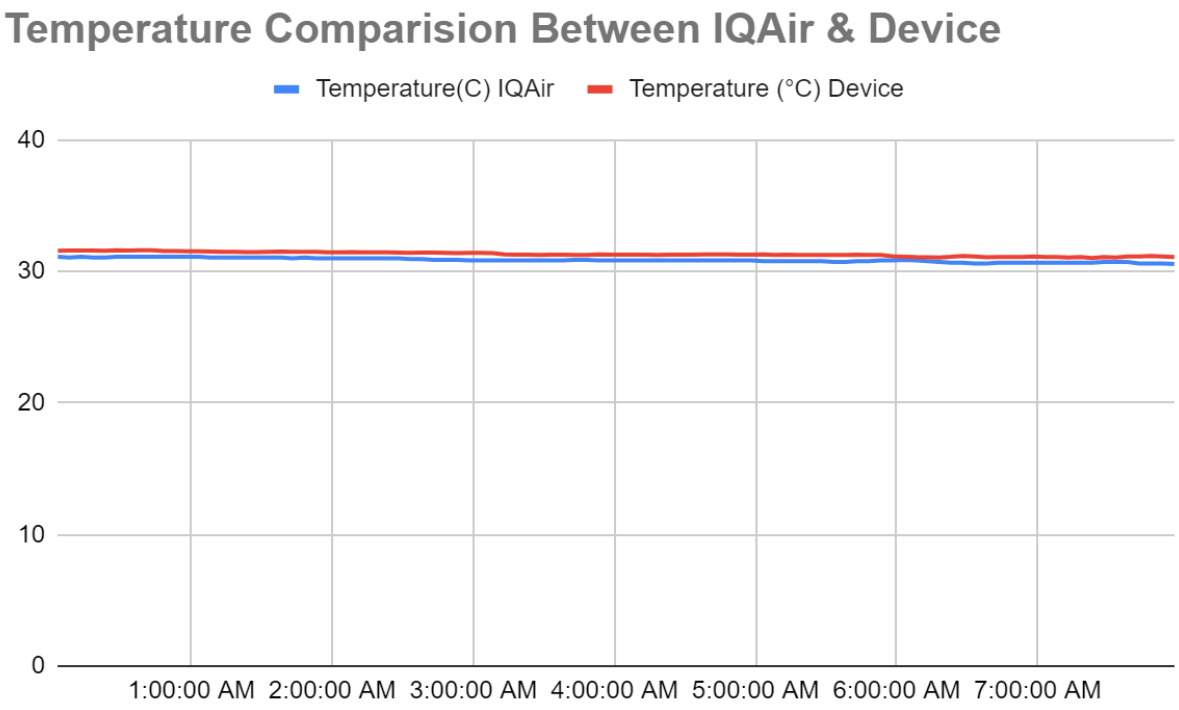

Figure 5.6.1.2.1 : Night Time Temperature Comparison.

5.6.1.2.2 Humidity Difference in Percentage

Average Difference: 0.979%

Minimum Difference: 0.973%

Maximum Difference: 0.989%

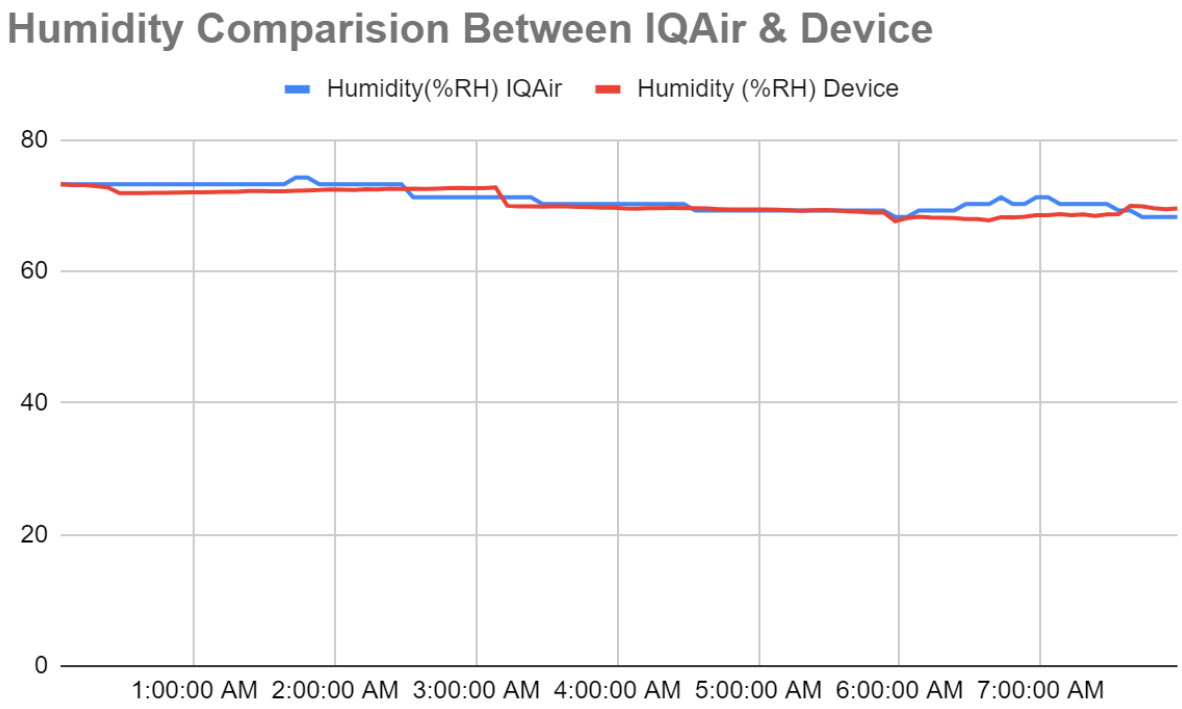

Figure 5.6.1.2..2 : Night Time Humidity Comparison.

5.6.1.2.3 PM1.0 (ug/m3) Difference in raw concentration

Average Difference: 0.791 ug/m3

Minimum Difference: 0 ug/m3

Maximum Difference: 3 ug/m3

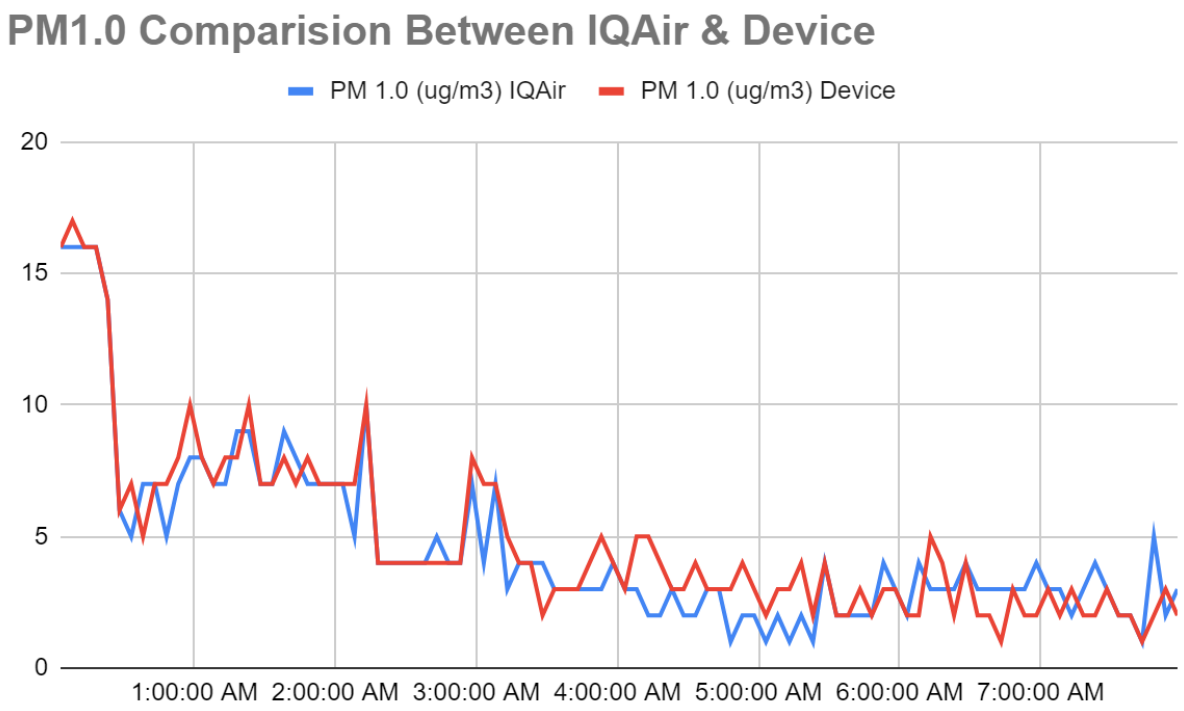

Figure 5.6.1.2.3 : Night Time PM1.0 Comparison.

5.6.1.2.4 PM2.5 (ug/m3) Difference in raw concentration

Average Difference: 0.604 ug/m3

Minimum Difference: 0 ug/m3

Maximum Difference: 2 ug/m3

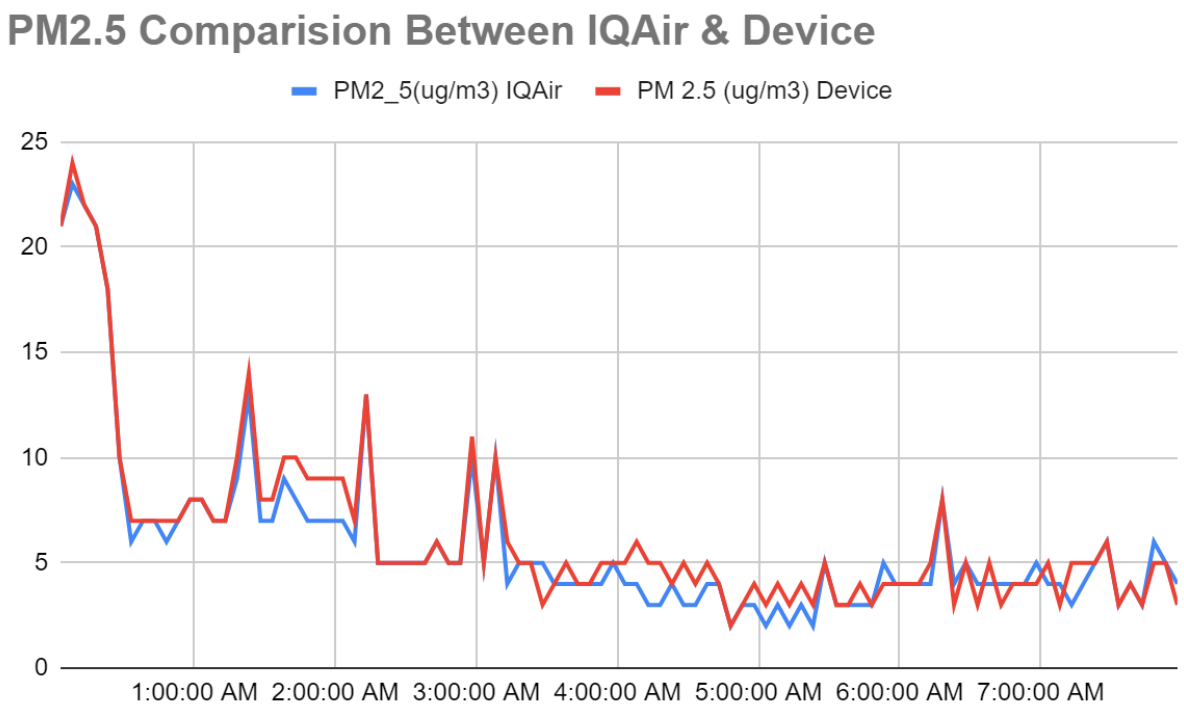

Figure 5.6.1.2.4 : Night Time PM2.5 Comparison.

5.6.1.2.5 PM10.0 (ug/m3) Difference in raw concentration

Average Difference: 0.646 ug/m3

Minimum Difference: 0 ug/m3

Maximum Difference: 3 ug/m3

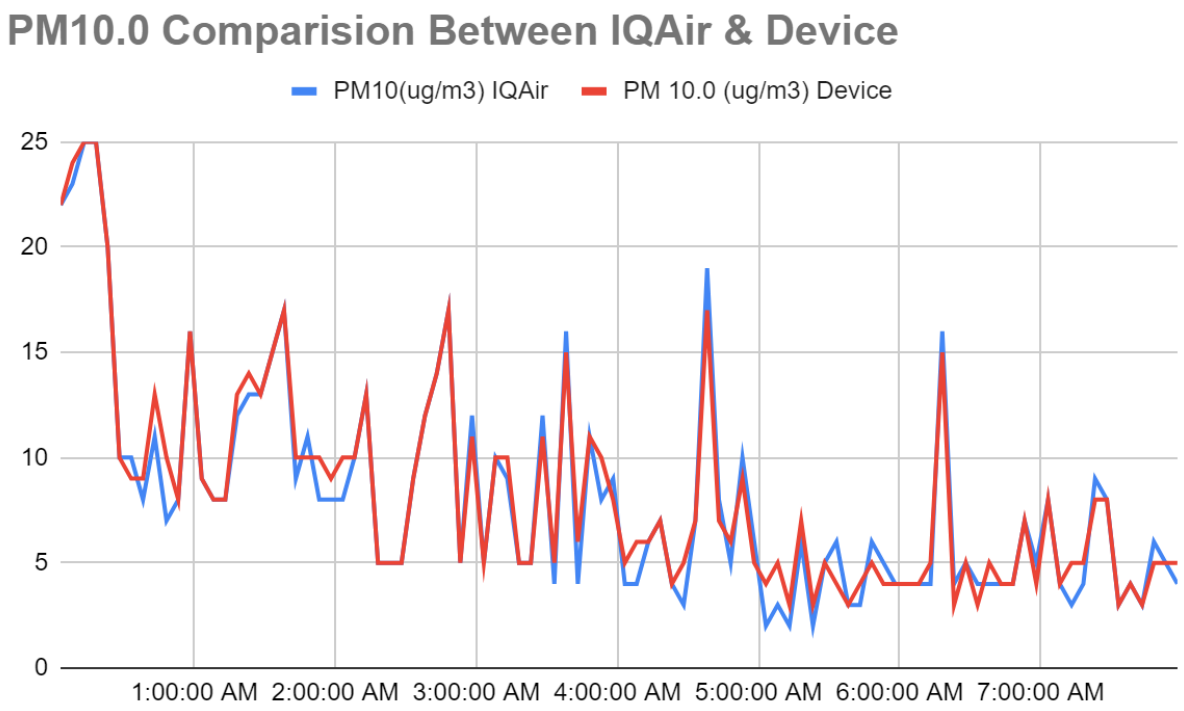

Figure 5.6.1.2.5 : Night Time PM10.0 Comparison.

5.6.1.2.6 CO_2 Difference in Percentage

Average Difference: 0.76%

Minimum Difference: 0.00%

Maximum Difference: 2.70%

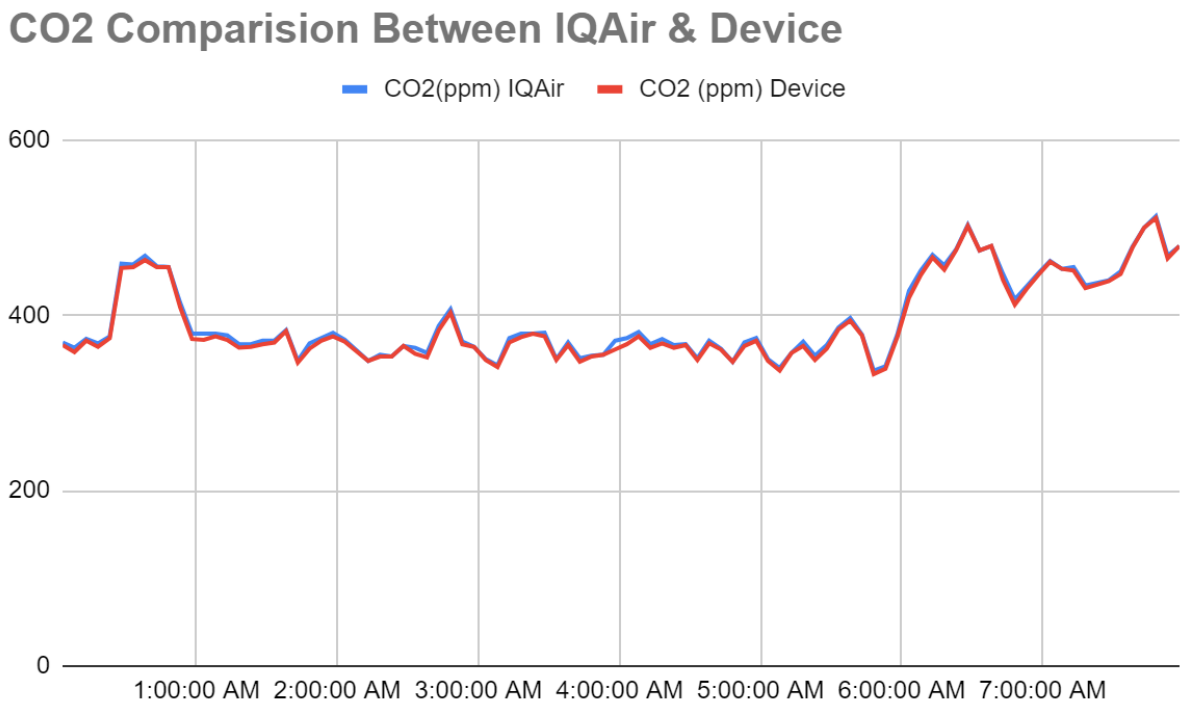

Figure 6.4.1.2.6 : Night Time CO_2 Comparison.

5.6.2 Discussion on Graphical Patterns

There are two lines of the same polluting components shown for determining the collection between RefinAir and AirVisual Pro. Please note that all the data were collected for 24 hours after proper calibration of both the air monitoring system devices. Within the 24 hour timeframe, data was fetched within every 5 minutes from both the devices automatically. Between the two devices, data difference is shown under the same category of sensors. Percentage error is done between temperature, humidity and CO_2. Particle concentration difference is between PM1.0, PM2.5 and PM10.0.

An observation is shown on the graphs above and we can conclude by saying that, both the devices have a very slim difference in data among all the sensors, proving the accuracy of our

device and the goal to diminish the worldwide crisis by introducing a low-cost and cheap solution.

5.7 Application of RefinAir

Around 3.8 million people die due to indoor air pollution because of the presence of Particulate Matter and increasing levels of toxic gases. Severe respiratory diseases have been reported which could lead to cancer. It is vital to keep the emission of toxins around the atmosphere in control, which is why the following places should be using our device, RefinAir.

- Indoor Air Quality Monitoring System or Gas Detection System

Monitoring indoor air can ensure a good working environment. It can be used in hospitals, homes, shops or any other indoor areas. The gas detection system can alert the people about the rising levels of pollutants and become health conscious.

- Outdoor Monitoring System

We will be able to fetch data from the most polluting outdoor sources like industries, untreated sewage, lead-acid battery recycling, metal smelting and processing, etc.. Data from transportation routes and surface areas of water bodies can also be collected by mounting our devices on the transportants.

- Atmospheric Map

An Atmospheric Map can be generated with the help of satellite data using MLAs, and this can help the citizens to determine the most polluting areas and take precautions accordingly. The industrialists will get alarmed about the situation and control the pollution in favourable conditions.

5.8 Comparison between RefinAir and AirVisual Pro

Properties	RefinAir	AirVisual Pro
Industrial Grade Sensors	✓	✓
Sensor Durability	✓	✗
Air Quality Forecasting	✓	✓
Device Maintenance Facility	✓	✗
Cost Feasibility	✓	✗
Real Time Monitoring	✓	✗
Indoor Air Monitoring System	✓	✓
Outdoor Air Monitoring System	✓	✗
Toxic Data Collection for extensive research	✓	✗

Table 5.8 : Comparison table for RefinAir and AirVisual Pro

Chapter 6 : Prediction of PM2.5 concentrations using Machine Learning

6.1 Overview

With rapid economic development and urbanization, the main concern of air pollution, PM2.5, is known to be fine particulate matter consisting of an aerodynamic equivalent diameter of less than 2.5 nm, has been proven to be the prime reason of many cardiovascular and respiratory diseases. The widespread anthropogenic and industrial emissions are a primary source of poor air quality in urbanized areas. An accurate assessment of PM2.5 pollution level is essential for analyzing public health and the environment hazards.

It is noticed that if the pollutant emission is not taken under control, this can cause untimely deaths of individuals with heart or lung conditions followed by nonlethal heart attack, nonuniform heartbeat, aggravated asthma, deteriorations in lung function, escalated respiratory symptoms and etc. - since fine particles less than 10 micrometers in diameter are the root of most problems in the human body as they have the ability to penetrate deep inside the lungs and enter the human body's bloodstream. Besides human bodily effects, the rise of PM2.5 has impaired visibility due to haze as one of it’s many other consequences. Repercussion like environmental damage causes acid rain turning water bodies acidic, changing and depleting the nutrients in ecological or environmental areas, impacting natural habitats and ecosystem diversification negatively. Due to acidic rains, materialistic destructions like staining, stone and other material mutilation are observed.

In order to diminish the global catastrophe, the Aerosol Optical Depth (AOD) 550nm data was collected from NASA Giovanni and the Ground Station data which was sourced from AQUA Satellite via a picture, was taken from AirNow website with US Embassy selected as the Ground Station. A total of 4.5 years of data was collected by the researchers for the project for factors like average temperature, windspeed, visibility, cloud coverage, rain precipitation, relative humidity through the Visual Crossing platform. Please note that all the gathered data are only of one region, Bangladesh.

For finding a solution to the worldwide rising predicamental situation, our aim of the project is to use the massive collection of data and to prepare a model to predict the upcoming concentration of PM2.5 data. Subsequently, different variables responsible for the rise of AQI have been known from several literature reviews. Multiple propositions have been put forward for the prediction of future behaviour of PM2.5 according to the criteria of data collection. By operating multiple models to predict the PM2.5, we have gained a fruitful outcome from employing the Multiple Regression Linear (MLR) model which utilizes several independent variables to predict the outcome of a dependent variable and its objective is to model the linear relationship between dependent variables and independent variables. Additionally, we have also placed a proposition of using Artificial Neural Network (ANN) as another suitable solution because of its favourable outcome. ANN is chosen for engineering nonlinear problems and estimating output values for provided input parameters based on their training values.

Apart from the proposed solutions, we have also tried implementing Catboost Regressor , Gradient Boosting, Random forest and Xgboost but the mean squared error and root mean squared error value derived from the testing and training data was not found convincing.

In the upcoming sections of the paper, we will be discussing the collection process of data along with a description of the dataset, methodology, result analysis with findings in detail.

6.2 Problem Statement

In recent decades, our planet underwent rapid urbanization, industrialization, and global economic integration. Poor air quality has garnered considerable attention ever since. Nevertheless, the foremost reason for this problem is the particulate matter in air pollution, also known as PM2.5. Airborne particles or droplets as small as two and a half microns in width can travel deep into the respiratory tract and reach the lungs. Fine particle exposure can cause short-term health consequences such as irritation of the eyes, nose, throat, and lungs, coughing, sneezing, runny nose, and breathlessness. Fine particle exposure can also impair lung function and exacerbate medical conditions such as asthma and heart disease. Increases in daily PM2.5 exposure have been linked in scientific studies to an increase in respiratory and cardiovascular hospital admissions, emergency department visits, and deaths.Long-term exposure to fine particulate matter may also be associated with increased rates of chronic bronchitis, decreased lung function, and higher mortality from heart and lung diseases, according to studies. Individuals with breathing and heart difficulties, children, and the aged people may be highly susceptible to PM2.5.

As an outcome, finding solutions to accurately and effectively predict PM2.5 has become a prime concern for experts and scholars.In the following section, methodologies for predicting atmospheric PM2.5 concentrations are proposed and evaluated based on time series and interactive multiple models as a proposed solution to the challenge. Satellite data are extracted from AQUA Satellites of multiple regions, and based on the type of data the following models are applied.

6.3 Description of the Dataset

To establish a prediction model to predict the PM2.5 of a specific area we have extracted datasets from NASA Giovanni data visualization platform, AirNow, and Visual Crossing. The collected datasets are respectively, AOD (Aerosol Optical Depth) 550 nm, PM2.5, and geospatial weather data. All of the datasets are described below along with the data distribution of PM2.5, and data preprocessing methods used to establish a collective dataset to predict PM2.5. The datasets are summarized in Table 6.3.1.

Data	Source	Variables
Satellite Data	Nasa Earth Data (Giovanni)	Aerosol Optical Depth 550 nm
Air Quality Data	AirNow	PM 2.5
Weather Data	Visual Crossing	Temperature
		Rain Precipitation
		Wind Speed
		Visibility

		Cloud Coverage
		Relative Humidity

Table 6.3.1 : Data description used to develop prediction model.

Since we are going to predict PM2.5 concentration, we divided the data set into 70% and 30% for training and testing data sets respectively.

6.3.1 Collection of AOD (Aerosol Optical Depth) 550 nm Data from Aqua Satellite

Aerosol Optical Depth (AOD) is the measure of aerosols (e.g., urban haze, smoke particles, desert dust, sea salt) distributed within a column of air from the Earth's surface to the top of the atmosphere. The used AOD 550 nm is the dust aerosol optical thickness at 550 nanometers (nm). The satellite AOD (Aerosol Optical Depth) 550 nm data was collected using NASA Giovanni data visualization platform. The available AOD 550 nm product was preprocessed from MODIS (Moderate Resolution Imaging Spectroradiometer) instrument aboard the Aqua Satellite, which passes from South to North poles of the earth. The obtained AOD 550 nm product is basically the level-3 atmosphere daily global product (MYD08_D3), which are derived from four level-2 MODIS AQUA atmosphere products MYD04_L2, MYD05_L2, MYD06_L2, and MYD07_L2. It contains daily 1 x 1 degree grid average values of atmospheric parameters related to atmospheric aerosol particle properties, total ozone burden, atmospheric water vapor, cloud optical and physical properties, and atmospheric stability indices. The selected shape of the data was Dhaka, which means the obtained AOD 550 nm product was collected and preprocessed by the AQUA satellite using the MODIS instrument while orbiting above Dhaka, Bangladesh. The collected data timeframe was from 2021-03-01 to 2021-03-31. The obtained AOD 550 nm is the mean AOD 550 nm for each day processed by the AQUA MODIS.

6.3.2 Collection of Ground Station PM2.5 Data

PM2.5 refers to a category of particulate matter or pollutants that is 2.5 micrometers or smaller in size. The data of PM2.5 raw concentration was obtained from AirNow. AirNow is a partnership of the U.S EPA (Environmental Protection Agency), National Oceanic and Atmospheric Administration (NOAA), National Park Service, NASA, Centers for Disease Control, and tribal, state, and local air quality agencies. The extracted data from AirNow represents the raw concentration of PM2.5 as a unit of μg/m3 for Dhaka and the timeframe was from 2021-03-01 to 2021-03-31. The initial data was of the mean raw concentration of PM2.5 per hour for 24 hours of each day, later it was converted to mean concentration of PM2.5 per day to fit with the AOD 550 nm product data obtained from the AQUA MODIS.

6.3.3 Collection of Geospatial Weather Data

We've collected data of a number of geospatial weather variables to establish the PM2.5 prediction model. The collected weather data variables are the following: average temperature, rain precipitation, wind speed, visibility, cloud coverage, and relative humidity. The weather data was extracted from Visual Crossing. Visual Crossing provides preprocessed authentic weather data from government agencies. The extracted weather variables data from visual crossing represents the following units: average temperature as degree celsius (°C), rain precipitation as millimeters (mm), wind speed as miles per hour (mph), visibility as kilometer (km), cloud coverage as percentage (%), and relative humidity as percentage (%). The geospatial weather data was collected from Dhaka and the timeframe was from from 2021-03-01 to 2021-03-31. All the data are mean data of 24 hours of each day. Later, the data was preprocessed to fit the prediction model.

6.3.4 Summarized Statistics of the Data Set.

Summarized Statistics of the Data Set

Variables	Count	Maximum	Minimum	Median	Mean	Standard Deviation	Skewness
AOD 550 nm	1476	2.959325	0.048922	0.478594	0.573474	0.399341	2.014310
$PM_{2.5}$	1476	381.3	3.1	69.25	93.647358	68.134591	0.913241
Temperature	1476	33.4	12.8	27.4	26.2	3.998857	-0.685136
Rain Precipitation	1476	207.06	0.0	0.0	9.998720	22.590237	3.758339
Wind Speed	1476	152.10	2.50	11.10	11.675339	7.992685	9.237594
Visibility	1476	2.8	0.4	2.4	2.304336	0.401330	-1.862870
Cloud Coverage	1476	96.4	0.0	50.95	45.723916	29.397158	-0.212261
Relative Humidity	1476	94.07	43.72	71.415	71.426524	8.783704	-0.145035

Table 6.3.4 : Summarized Statistics of the Data Set.

From the summarized statistics table in **Table 6.3.4** we can conclude the total count, maximum value, minimum value, median, mean, standard deviation, and skewness of all the variables present in the data set,

6.3.5 Data Distribution of the PM2.5 in the Air

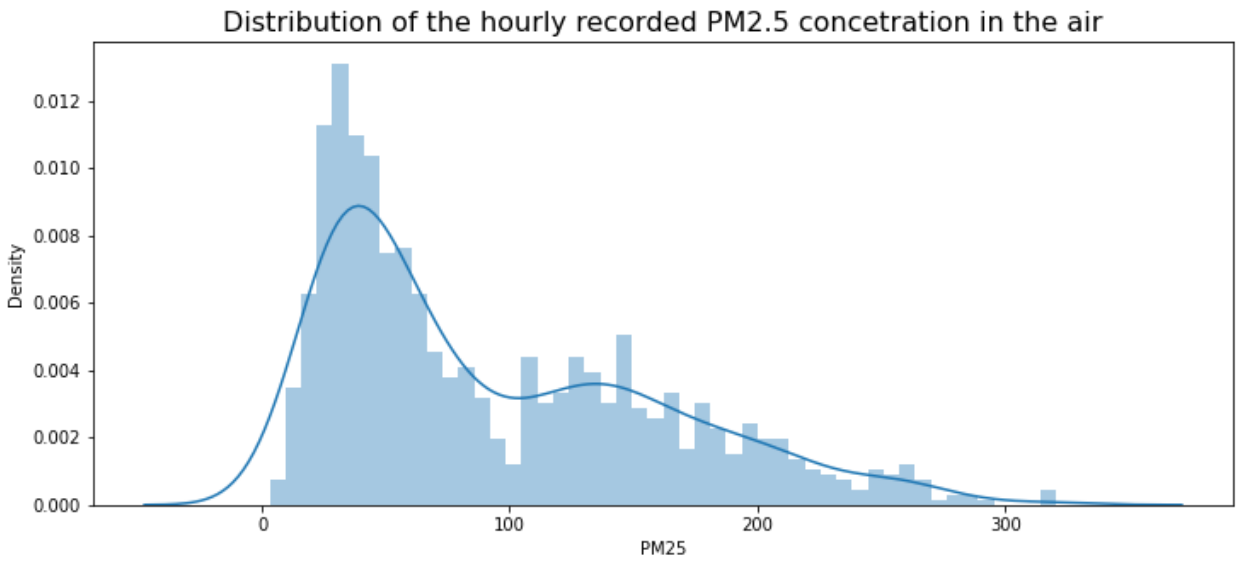

Figure 6.3.4.1 : Data Distribution of the PM2.5 Concentration in the Air.

From the data distribution diagram in Figure 6.3.4.1 we can conclude that the data of PM2.5 concentration in the air is positively skewed. And, most of the data falls in the range of 0 - 100 ppm.

6.3.6 Data Preprocessing

Respectively, AOD 550 nm data from aqua satellite, PM2.5 from ground station, and geospatial weather data was preprocessed to maintain the accuracy, completeness, and consistency of the data to establish the prediction model.

We started with data cleaning during the data preprocessing phase. Initially we identified all the missing data, noisy data, and outliers caused due to equipment malfunction and have inconsistency with other recorded data. After identifying, using we removed all missing data and outliers. The noisy data were smooth by means of it's nearest obtained data.

To accomplish the data integration, the acquired AOD 550 nm data from the aqua satellite, PM2.5 data from the ground station, and geospatial weather data were merged into a single coherent csv file. Data value conflicts such as different scales were removed during the extraction of the data as all the data were extracted in British Units. Later on, the data were split into a 70:30 ratio respectively for training and testing datasets.

Finally, data transformation was performed. Except the AOD 550 nm data all the integrated data was normalized by decimal scaling to two decimal points. Decimal scaling the AOD 550 nm impacted heavily in the prediction model thus it was discarded from decimal scaling.

Correlation analysis was performed to identify the correlation and a linear relationship between the variables. Following the correlation analysis all the positively correlated, negatively correlated, and independent variables were identified. The correlation matrix in Figure 6.3.6.1 shows that the ground station PM2.5 data has high negative correlation with temperature, visibility, and cloud coverage. The ground station PM2.5 also has partial negative correlation with AOD 550 nm, rain precipitation, wind speed, and relative humidity. We can also conclude from the correlation matrix that the AOD 550 nm, Average Temperature, Rain Precipitation, Wind Speed, Visibility, Cloud Cover, Relative Humidity are independent variable in the collective dataset and have a linear relation with the dependent variable PM2.5.

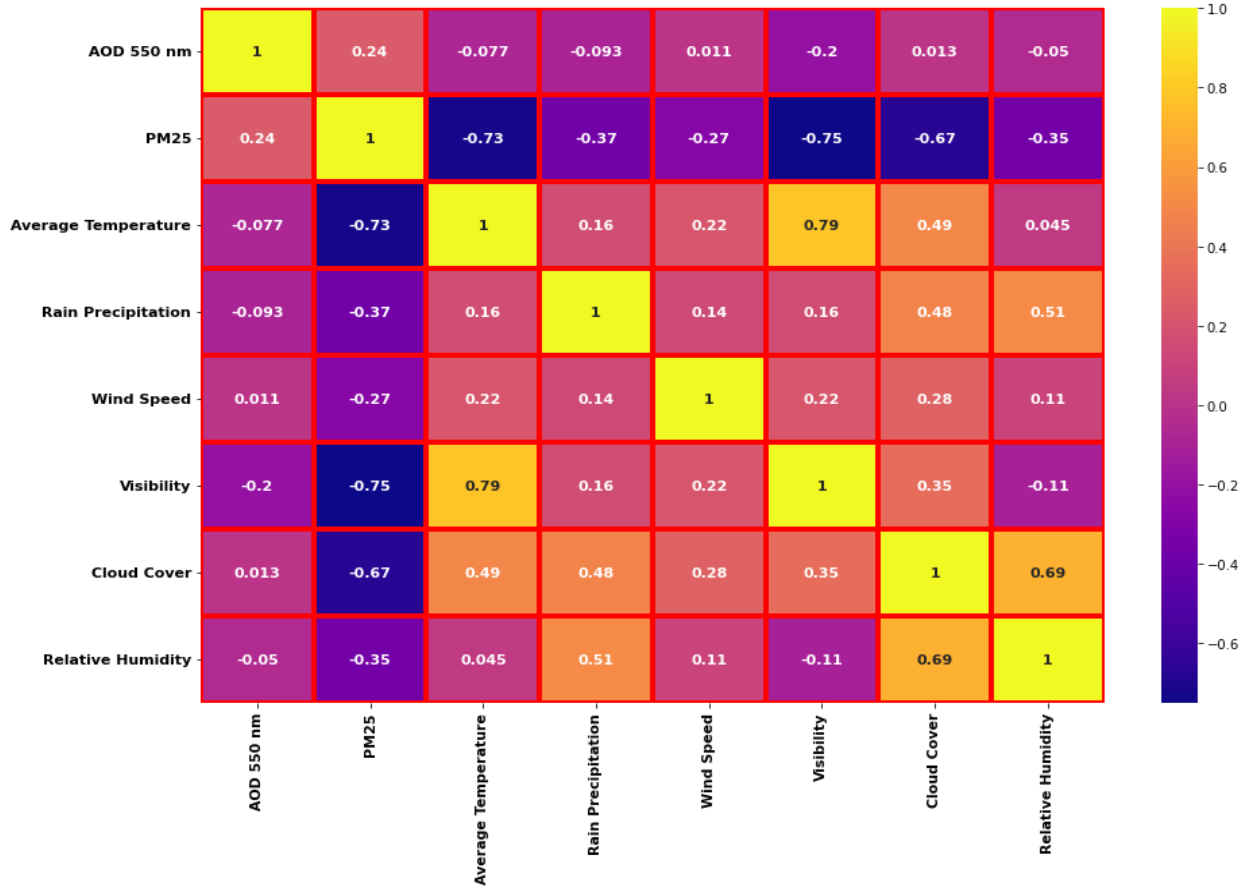

Figure 6.3.6.1 : Correlation Matrix Heatmap of the variables.

6.3.7 Data Visualization of Training and Testing Data Set

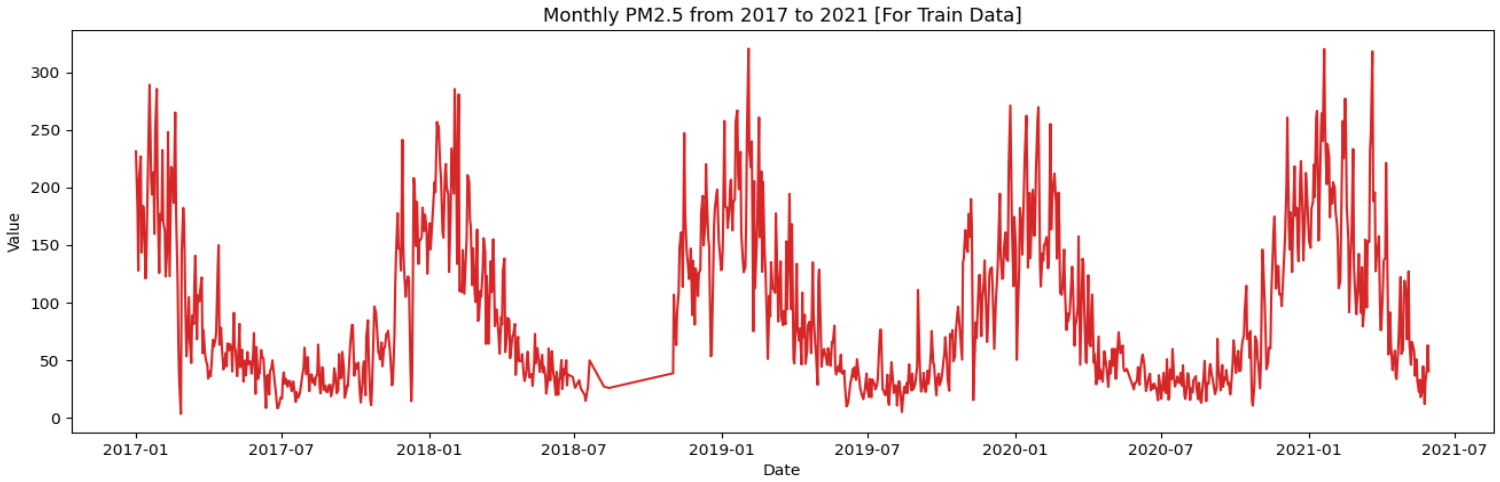

Figure 6.3.7.1 : Data Visualization for PM2.5 Train Dataset.

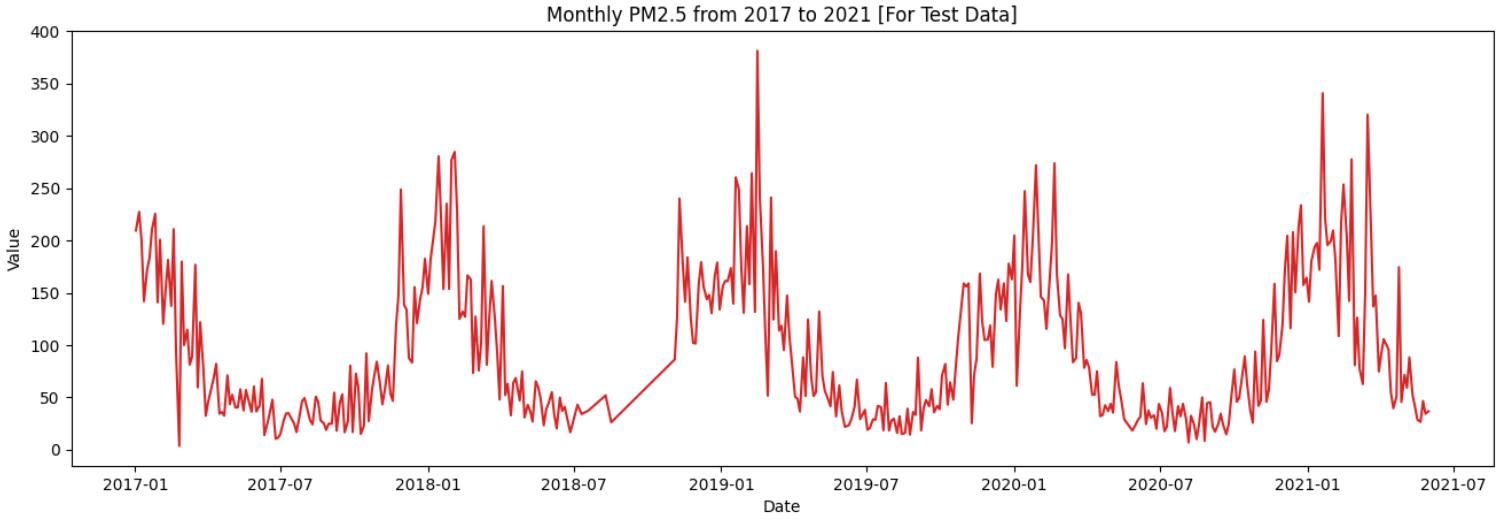

Figure 6.3.7.2 : Data Visualization for PM2.5 Test Dataset

The graphical pattern is observed to have chaotic behaviour and sine waves are found. From data visualization of PM2.5 concentrations for both the Train and Test data set in Figure 6.3.7.1 and Figure 6.3.7.2, we observed that a seasonal effect may be present in our data set.

6.4 Methodology

This section presents an expanded discussion on the Machine Learning Algorithms used to predict the concentration of PM2.5.

6.4.1 Multiple Regression Linear (MLR) Model

This model is an extension of the two-variable linear regression model that uses several independent variables to predict the outcome of a dependent variable. The objective is to model the linear relationship between dependent variables and independent variables. The MLR model is used comprehensively in econometrics, weather forecast, financial inference and many others. An MLR model can be written as:

$$PM2.5_i = \beta_0 + ct + \beta_1 DBA_{i1} + \beta_2 Temp_{i2} + \beta_3 RainP_{i3} + \beta_4 WS_{i4} + \beta_5 Visi_{i5} + \beta_6 CC_{i6} + \beta_7 Humid_{i7} + e_i i = 1, 2,..., n$$

Where PM2.5 is the dependent variable, Xp are the independent variables (e.g. DBA, ATem, RainP, WS, WD, Visi and Humd), β0 is the constant, c is the coefficient of time trend, β_i (i = 1, 2, …, 8) are the regression coefficients for each explanatory variable and et is the error term (known also as residuals). The MLR model is based on the following assumptions:

(i) There is a linear relation between dependent variable and independent variables
(ii) The independent variables are not too highly correlated with each other and
(iii) Residuals should be normally distributed with a mean of 0 and variance 1.

The most common method for fitting a regression line is the method of least squares, which calculates the best-fitting line by minimizing the sum of the squares of errors. The estimated multiple regression line becomes based on random samples:

$$\widehat{PM2.5}_i = \hat{\beta}_0 + \hat{c}t + \hat{\beta}X_i,$$

where $\hat{\beta}_0 = \overline{PM2.5} - \hat{c}t - \hat{\beta}\overline{X}$, and $\hat{\beta} = X'X^{-1}(PM2.5)$.

Divide methodology to all the methods used

6.4.2 Artificial Neural Network (ANN)

The ANN algorithm is a popular machine learning method for analyzing complex correlations between predictors and predictands. The number of layers and neurons in each layer may be readily changed because of the flexible architecture. Furthermore, throughout the model-building process, ANN does not require any previous assumptions, such as data stationarity. As a result, the data characteristics heavily influence the network model. Three layers make up the architecture of the most frequently used ANN model in time series forecasting, often known as multilayer perceptrons. The neurons in the processing units are

connected in a cyclic manner. To use such a network to represent time series data, a nonlinear function f of y_t sequence from y_{t-1} to y_{t-N} is built as shown in the following equation :

$$y_t = \omega_0 + \sum_{j=1}^{H} \omega_{jf}\left(\omega_{0j} + \sum_{i=1}^{N} \omega_{ij} y_{t-i}\right) + e_t,$$

where ω_{ij} and ω_{0j} are model weights at any given time t, and H and N are the number of hidden and input nodes, respectively. e_t represents a noise or error term in this equation. In ANN construction, the transfer function of the hidden layers f is usually a sigmoid function. The ability of ANN to approximate any continuous function by altering the number of layers N and hidden nodes H is its strength. The number of layers and nodes in each layer have a significant impact on the predicting performance of ANNs. Large numbers of N and H can provide extremely high training accuracies, but it suffers from overfitting since it memorizes the training data. An overly simple ANN network, on the other hand, results in poor generalization. In this study, we have used the Rectified Linear Unit (ReLU) activation function along with an input layer dimension of 11, two hidden layers with batch normalization, and a single output layer.

```
Model: "sequential_4"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_16 (Dense)             (None, 64)                768
_________________________________________________________________
batch_normalization_12       (None, 64)                256
_________________________________________________________________
dense_17 (Dense)             (None, 128)               8320
_________________________________________________________________
batch_normalization_13       (None, 128)               512
_________________________________________________________________
dense_18 (Dense)             (None, 256)               33024
_________________________________________________________________
batch_normalization_14       (None, 256)               1024
_________________________________________________________________
dense_19 (Dense)             (None, 1)                 257
=================================================================
Total params: 44,161
Trainable params: 43,265
Non-trainable params: 896
```

Figure 6.4.2.1 : Proposed ANN Architecture.

6.4.3 Random Forest

Random forest is a collection of decision trees that may be used to classify and predict data. It is extremely flexible and efficient. Random forest has just two input parameters to tune as a user-friendly algorithm: the number of trees to grow (n_{tree}) and the number of predictors randomly selected as candidates at each split (m_{try}). The hyperparameters of a model should be adjusted in order to undertake random forest analysis. The random forest hyperparameter ranges and optimal values used to predict PM2.5 from the dataset are shown in Table 6.4.3.1.

Parameter	Range	Optimal Value
n_estimators	800 - 900	800
max_depth	5 - 100	5
max_features	[auto, sqrt, log2]	sqrt
min_samples_split	[2, 4, 8]	2
min_samples_leaf	[1, 2, 4]	2
bootstrap	[True, False]	True

Table 6.4.3.1 : Random Forest hyperparameters.

6.4.4 Gradient Boosting Regressor

Another frequently used ensemble regression model is Gradient Boosting Regression (GBR). The basic concept behind this approach is to build new base-learners that are maximally correlated with the loss function's negative gradient, which is connected with the entire ensemble. Here, we have used five important parameters to tune the Gradient Boosting Regressor model. The Gradient Boosting Regressor parameters used to predict the PM2.5 are shown in Table 6.4.4.1.

Parameters	Optimal Value
n_estimators	100
max_depth	5

max_features	auto
min_samples_split	85
min_samples_leaf	15
learning_rate	0.05

Table 6.4.4.1 : Gradient Boosting Regressor parameters.

6.4.5 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) works on the same idea as the Gradient Boosting Regressor model. XGBoost, on the other hand, has a better handle on over-fitting because of a more regularized model technique (Chen et al., 2015). In comparison to previous algorithms, it offers greater control over overfitting due to the use of more regularized model formalization. In Kaggle contests, it has a high success rate, especially for structured features. The XGBoost parameter ranges and optimal values used to predict PM2.5 from the dataset are shown in Table 6.4.5.1.

Parameter	**Range**	**Optimal Value**
n_estimators	70 - 1600	1500
max_depth	1 - 10	3
gamma	0.0 - 1.0	0.0
min_child_weight	1 - 10	1
learning_rate	0.1 - 0.5	0.1
verbosity	[True, False]	False

Table 6.4.3.1 : Extreme Gradient Boosting (XGBoost) parameters.

6.4.6 CatBoost Regression

In comparison to previous implementations of gradient boosted decision trees, Catboost is the newest of the prominent gradient boosting libraries released by Yandex researchers in 2017, delivering state-of-the-art performance on a broad range of common machine learning tasks (Prokhorenkova et. al.). The introduction of ordered boosting, a modification of gradient standard boosting methods to minimize target leaking, and a novel algorithm approach to

handle categorical data are the primary aspects of CatBoost. The CatBoost Regression parameters used to predict the PM2.5 are shown in Table 6.4.6.1.

Parameters	Optimal Value
n_estimators	1000
depth	7
random_seed	42
bagging_temperature	0.2
metric_period	50
learning_rate	0.03

Table 6.4.3.1 : CatBoost Regression model parameters.

6.5 Result Analysis and Discussion

In this study, we have used Aerosol Optical Depth (AOD) 550 nm derived satellite data, ground station observed PM2.5 concentration data, and geospatial weather data to establish a prediction model to predict the PM2.5 concentration. The results from the aforementioned prediction models are to be discussed in this section.

6.5.1 Multiple Regression Linear (MLR) Model Results

Using the following MLR model equation we have predicted the PM2.5.

$$PM2.5_i = \beta_0 + ct + \beta_1 DBA_{i1} + \beta_2 Temp_{i2} + \beta_3 RainP_{i3} + \beta_4 WS_{i4} + \beta_5 Visi_{i5} + \beta_6 CC_{i6} + \beta_7 Humid_{i7}$$

The estimated coefficients are shown in Table 6.5.1.

	Estimate	SE	tStat	pValue
[Intercept]	512.58	18.941	27.062	$2.2571e^{-122}$
Timetrend	0.0060922	0.0034194	1.7817	0.075099
DBA	19.124	2.5976	7.3624	$3.6806e^{-13}$

Temp	-2.6173	0.46676	-5.6074	$2.6352e^{-08}$
RainP	-0.021017	0.053282	-0.39444	0.69334
WS	-0.33217	0.13483	-2.4635	0.013919
Visi	-89.147	4.6278	-19.263	$6.7337e^{-71}$
CC	-0.5807	0.067509	-8.6019	$2.8726e^{-17}$
Humid	-1.7884	0.20226	-8.842	$3.983e^{-18}$

Table 6.5.1 : Estimated Coefficients.

The total number of observations were 1043 and error degrees of freedom were 1034. We achieved a root mean square error of 0.322. R-squared value was 0.784 and adjusted R-Squared value was 0.782.

Measurement of errors of the applied MLR model is shown in Table 6.5.2.

Measures	Training Data	Testing Data
$RMSE$	32.2	28.8547
R^2	0.784	0.9344

Table 6.5.2 : Error measures of the MLR model.

Data visualization of the training and testing datasets along with their respective prediction of PM2.5 is shown in Figure 6.5.1.1 and Figure 6.5.1.2.

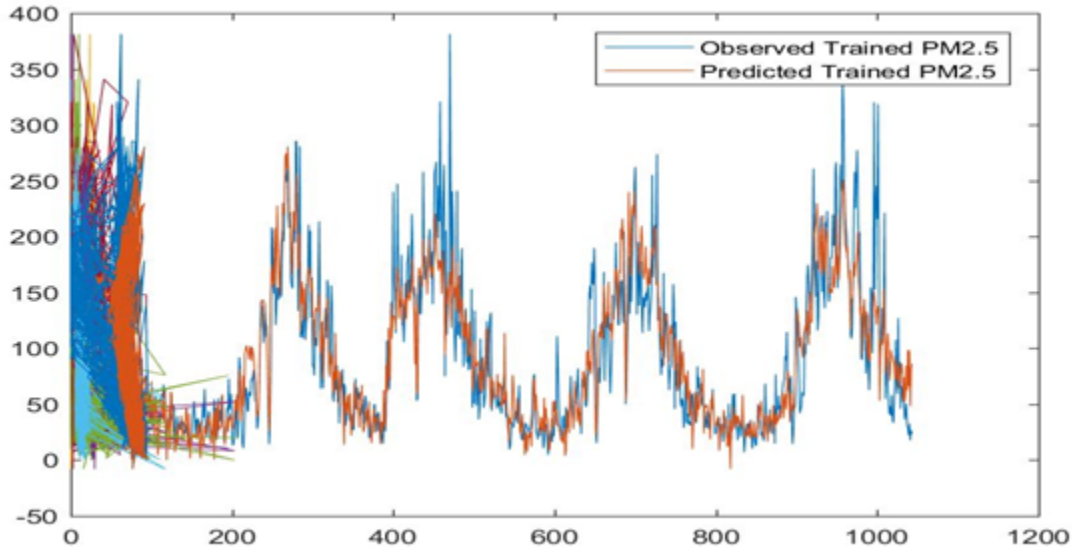

Figure 6.5.1.1 : Data visualization of observed and predicted training data of PM2.5.

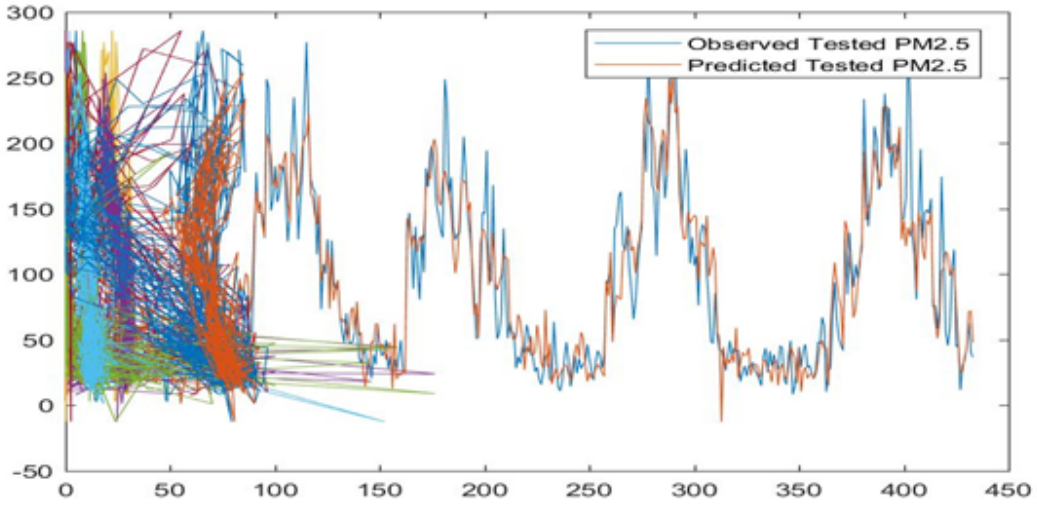

Figure 6.5.1.2 : Data visualization of observed and predicted testing data of PM2.5.

From, the Table 6.5.2 and Figure 6.5.1.2, we can conclude that the MLR model is perform well for the testing data with a R^2 value of 0.9344, which is close to 1. The training data shows a R^2 value of 0.784.

6.5.2 Artificial Neural Network (ANN) Results

Using the aforementioned ANN architecture with Rectified Linear Unit (ReLU) activation function for 30 epoches we were able to achieve a Root Mean Squared Error value of 31.112. With an overall model accuracy of 80.10%. The $RMSE$ per epoch of the ANN is shown in Figure 6.5.2.1 and a data visualization of the real data and predicted data is shown in Figure 6.5.2.2.

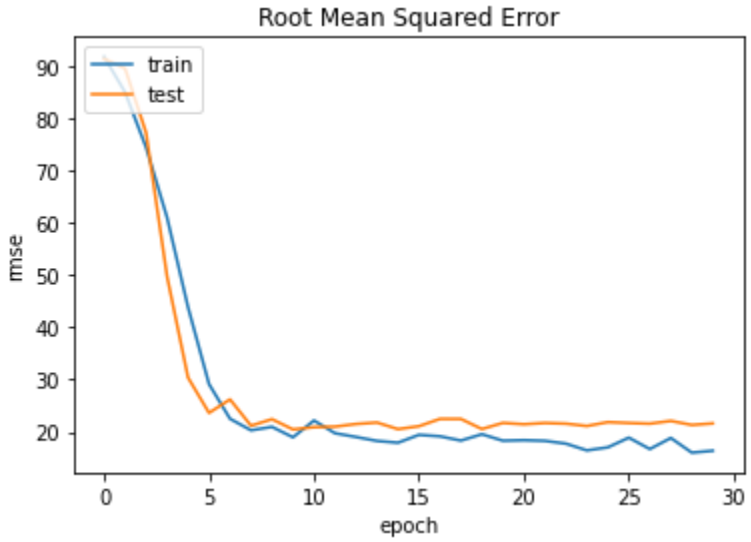

Figure 6.5.2.1 : RMSE per epoch.

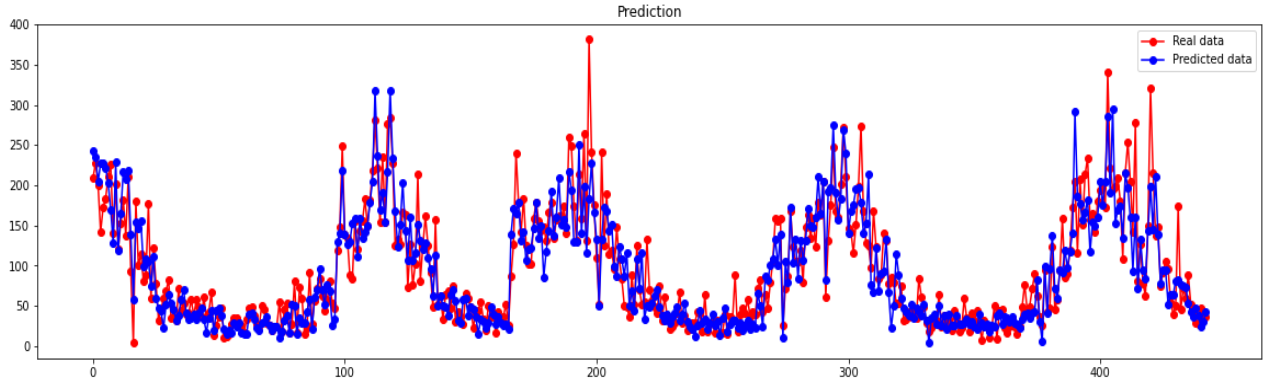

Figure 6.5.2.2 : Data visualization of the real data v/s predicted data for ANN.

6.5.3 Random Forest Results

Using the aforementioned hyperparameters we were able to achieve the result shown in Table 6.5.3.1.

Measures	Training Data	Testing Data
$RMSE$	24.64	31.71
R^2	0.87	0.79
MSE	607.26	1005.65
MAE	17.63	21.36
$MAPE$	30.58	39.70

Table 6.5.3.1 : Error measures of the Random Forest model.

The data visualization for the Random Forest model of the targeted PM2.5 data along with the PM2.5 predicted data is shown in Figure 6.5.3.1.

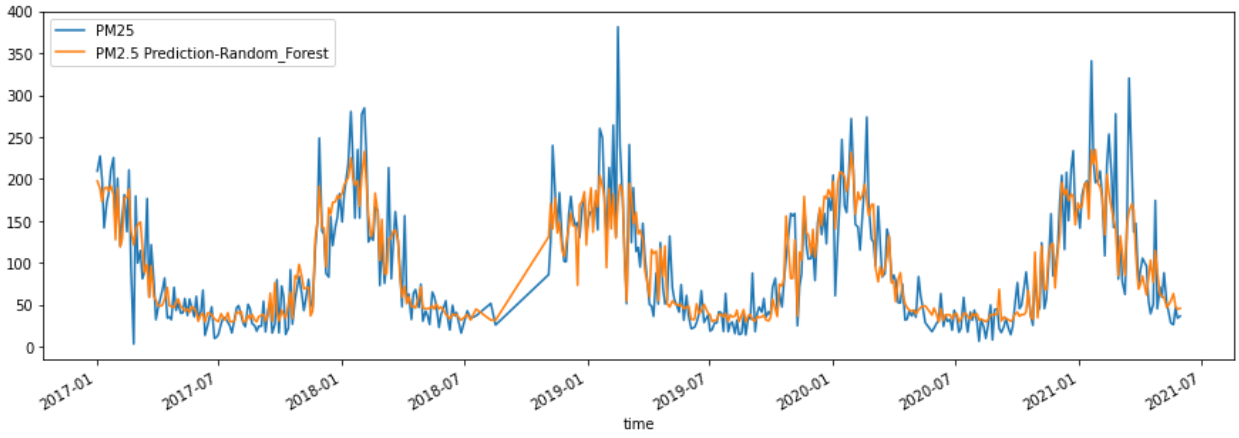

Figure 6.5.3.1 : Data visualization of targeted PM2.5 v/s predicted PM2.5 for Random Forest.

Using the Random Forest model along with the hyperparameters a model accuracy of 79.40% was obtained.

6.5.4 Gradient Boosting Regressor Results

Using the aforementioned parameters we were able to achieve the result shown in Table 6.5.4.1.

Measures	Training Data	Testing Data
$RMSE$	20.82	30.94
R^2	0.90	0.80
MSE	433.33	957.51
MAE	14.85	20.72
$MAPE$	25.99	35.32

Table 6.5.4.1 : Error measures of the Gradient Boosting Regressor model.

The data visualization for the Gradient Boosting Regressor model of the targeted PM2.5 data along with the PM2.5 predicted data is shown in Figure 6.5.4.1.

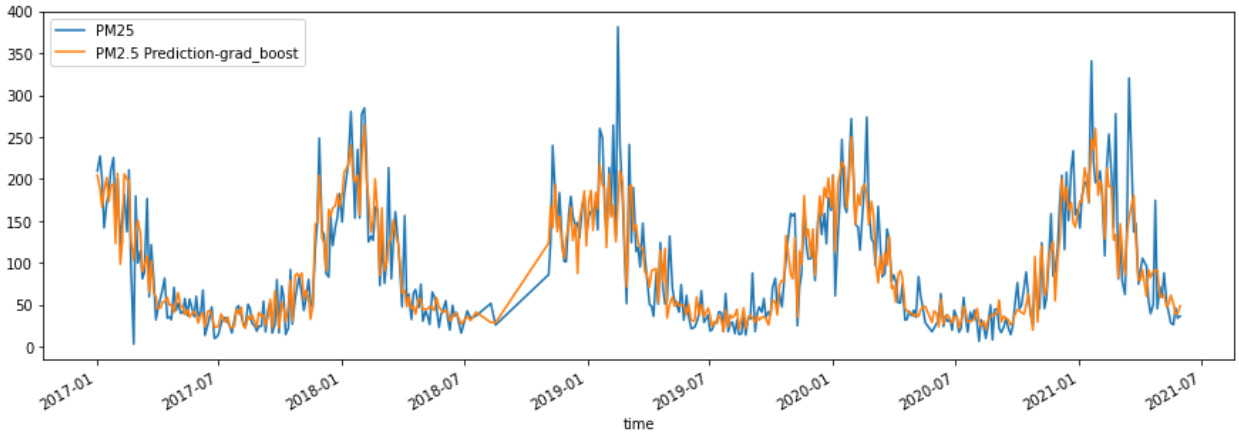

Figure 6.5.4.1 : Data visualization of targeted PM2.5 v/s predicted PM2.5 for Gradient Boosting Regressor.

Using the Gradient Boosting Regressor model along with the parameters a model accuracy of 80.32% was obtained.

6.5.5 Extreme Gradient Boosting (XGBoost) Results

From the Figure 6.5.5.1 we can conclude that the AOD 550 nm data has the most feature importance for the XGBoost model, followed by the Relative Humidity and Temperature to predict the PM2.5 concentration.

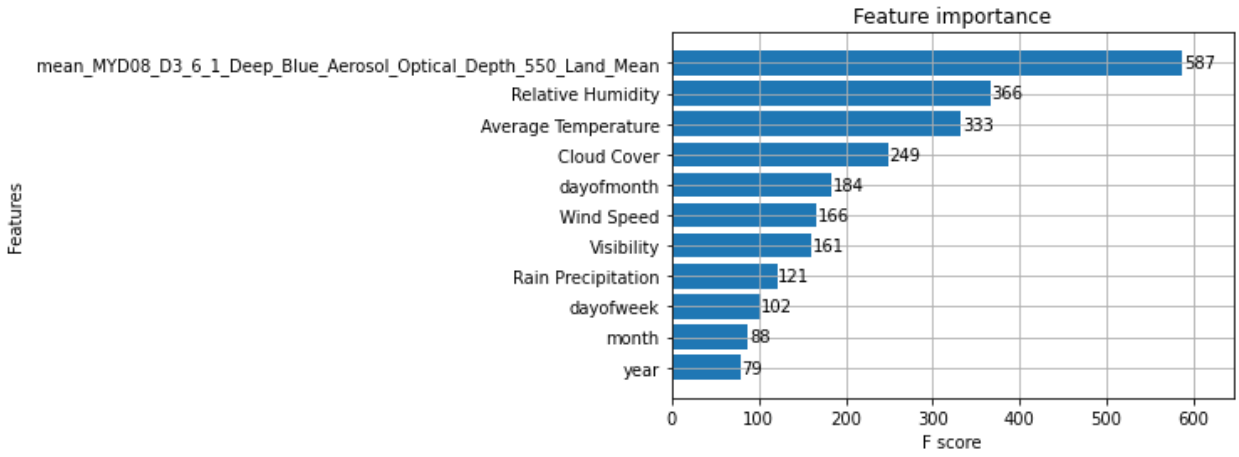

Figure 6.5.5.1 : Feature importance for XGBoost to predict PM2.5 concentration.

Using the aforementioned parameters we were able to achieve the result shown in Table 6.5.5.1.

Measures	Training Data	Testing Data
$RMSE$	13.34	31.22
R^2	0.96	0.80
MSE	177.99	974.65
MAE	9.82	21.21
$MAPE$	17.68	34.49

Table 6.5.5.1 : Error measures of the XGBoost model.

The data visualization for the XGBoost model of the targeted PM2.5 data along with the PM2.5 predicted data is shown in Figure 6.5.5.2.

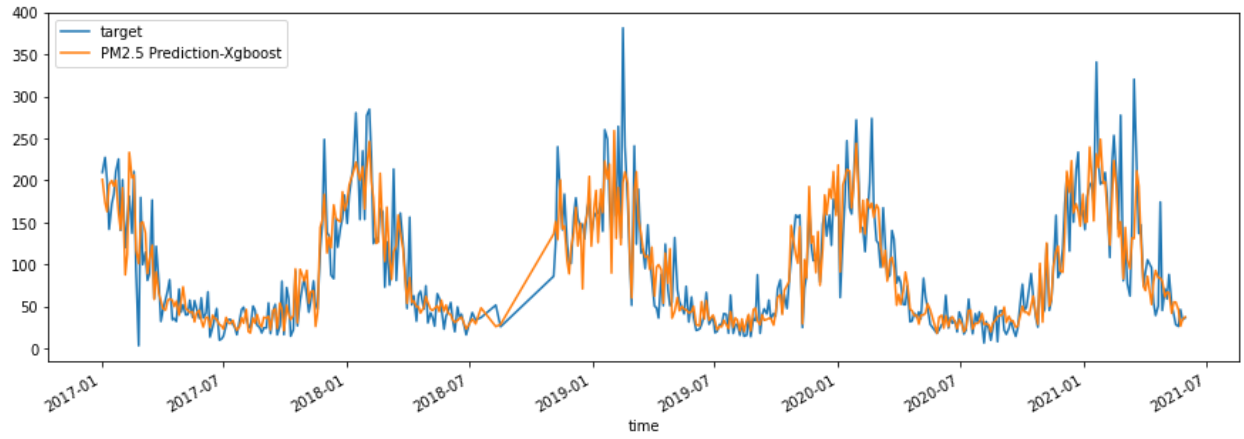

Figure 6.5.5.2 : Data visualization of targeted PM2.5 v/s predicted PM2.5 for XGBoost.

Using the XGBoost model along with the parameters a model accuracy of 80.65% was obtained.

6.5.6 CatBoost Regression Results

Using the aforementioned parameters we were able to achieve the result shown in Table 6.5.6.1.

Measures	Training Data	Testing Data
$RMSE$	8.90	29.46
R^2	0.98	0.82
MSE	79.14	867.69
MAE	6.82	20.05
$MAPE$	12.82	31.57

Table 6.5.6.1 : Error measures of the CatBoost Regression model.

The data visualization for the CatBoost Regression model of the targeted PM2.5 data along with the PM2.5 predicted data is shown in Figure 6.5.6.1.

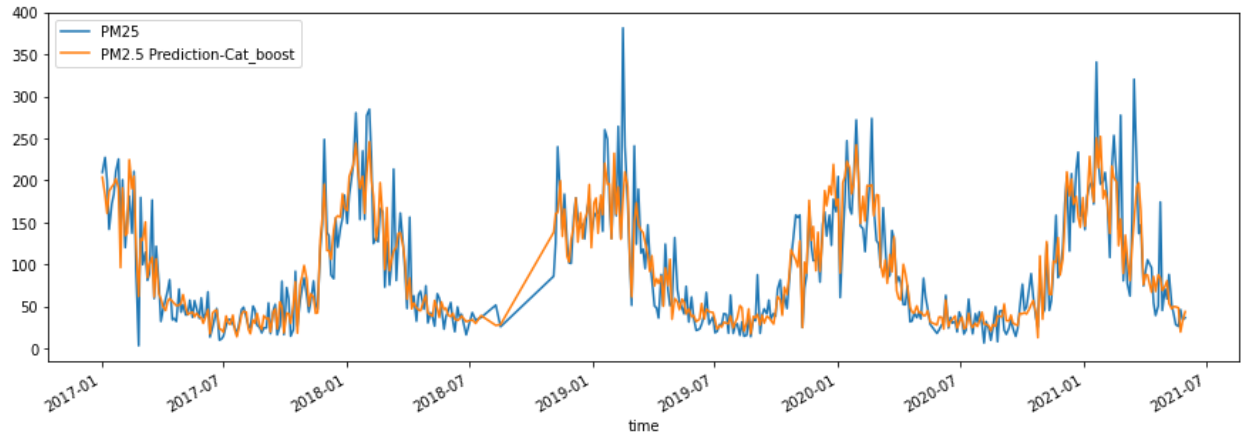

Figure 6.5.6.1 : Data visualization of targeted PM2.5 v/s predicted PM2.5 for CatBoost Regression.

Using the CatBoost Regression model along with the parameters a model accuracy of 82.17% was obtained.

6.5.6 Discussion

In Table 6.5.6, a comparison of the implemented algorithms to predict the PM2.5 are shown.

	RMSE		R^2		MSE		MAE		$MAPE$	
Algorithm	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
MLR	32.2	28.86	0.78	0.93	-	-	-	-	-	-
ANN	17.74	30.07	0.93	0.81	314.63	904.42	12.71	20.36	128.19	132.32
Random Forest	24.64	31.71	0.87	0.79	607.26	1005.65	17.63	21.36	30.58	39.70
GBR	20.82	30.94	0.90	0.80	433.33	957.51	14.85	20.72	25.99	35.32
XGBoost	13.34	31.22	0.96	0.80	177.99	974.65	9.82	21.21	17.68	34.49
CatBoost	8.90	29.46	0.98	0.82	79.14	867.69	6.82	20.05	12.82	31.57

Table 6.5.6: Result Comparison Table of the Implemented Algorithms.

From, Table 6.5.6 we can conclude that the Multiple Regression Linear (MLR) model was the most effective model among all to predict the PM2.5 using the AOD 550 nm satellite derived data and geospatial weather data with a R^2 value of 0.9344 for the testing dataset. Followed by the CatBoost Regression model achieved a R^2 value of 0.82 for the testing dataset and a model accuracy of 82.17%. The XGBoost model achieved a similar results to CatBoost Regression model, XGBoost model achieved a R^2 value of 0.80 and a model accuracy of 80.65%. The other prediction models have shown similar results to the XGBoost model.

So, we can conclude that using the Multiple Regression Linear (MLR) model to predict the PM2.5 will be the most accurate model, followed by the CatBoost Regression model.

Chapter 7 : Conclusion

This final section is dedicated to the discussion on the objective and proven hypothesis of the research. The reason why the project was finalized at first place was as a concern to the worldwide climate crisis and with an urge to find a solution to this problem. The 13th SDG goal is one of the causes why our planet is and will be more unliveable in future, and this will create more catastrophe as years will pass by.

The hypothesis of the project was to introduce a sustainable and inexpensive solution to the world, which could spread awareness on global warming and climate change. Often normal citizens can not afford fancy devices and the weather information that appears on their smartphones does not show the polluting compound statistics, which does not send any message to the people about our dying planet. If we compare the pricing between our IoT based device and AirVisual Pro, the total cost for making our device was approximately BDT 16,150 which in contrast to the industrial-graded device is more than half the amount less in terms of costing. With our proven hypothesis, if we can make our IoT based device on an industrial level amount, our costing will decrease.

At the beginning of our project, we made a pre-prototype, where some low-graded sensors were used and there was a major fluctuation in data. With not so satisfying results, we sourced some sensors with accuracies that could match any other industrial sensors, but made sure the price did not go off-limit. After a successful procurement of sensors, we implemented all the sensors to a microcontroller and started receiving data - transferring it to PC. Using a Python script, the data was migrated from PC to Google Cloud, which was then uploaded to Google Sheet. On the Google Sheet the graphs were shown and data comparison was done between the AirVisual Pro and IoT based devices. During our solution analysis, it was seen that the data difference was negligible between the two devices, proving our proposition and making our device eligible as the industrial-graded ones.

Apart from working on our hypothesis, we have also fetched data from AQUA satellite and ground station which can be known as another proposed solution to the problem. This is another solution to the crisis which is cost effective and has more coverage of area in case of gathering the pollutant concentration data. This approach will not only cover land but also water bodies and help us determine the pollution rate in transportation routes. Our alternative proposal is to use machine learning to predict PM2.5 using satellite and weather data.

Establishing the prediction models we found the following findings. With an R^2 value of 0.9344 for the testing dataset, we can infer that the Multiple Regression Linear (MLR) model was the most successful model among all for predicting PM2.5 utilizing the AOD 550 nm satellite derived data and geospatial meteorological data. The CatBoost Regression model was then used, which had an R^2 value of 0.82 for the testing dataset and an accuracy of 82.17 %. The

XGBoost model produced comparable findings to the CatBoost Regression model, with an R^2 value of 0.80 and a model accuracy of 80.65%. The findings of the other prediction models were identical to those of the XGBoost model. As a result, we may deduce that the MLR model, followed by the CatBoost Regression model, will be the most accurate in predicting PM2.5.

Most significantly, to our knowledge, this is the first study to look into the relevance of PM2.5 concentration prediction characteristics and create a system to monitor air quality in a developing nation like Bangladesh. Previous work has not included new features such as the idea of an air quality monitoring system to monitor the air quality of Bangladesh on a wide scale and be able to forecast and display PM2.5 concentrations on a single platform, as well as analyzing the impacts of meteorological data on PM2.5. However, there are certain drawbacks to be aware of. Although we were able to predict PM2.5 with fair accuracy, satellite-derived AOD 550nm data was difficult to obtain. Future work will focus on developing a more accurate prediction model, creating an Atmospheric Map, establishing weather stations around the country, and developing a mobile application to raise public awareness about air pollution.

7.1 Challenges faced

The following section is to give an insight of the boundaries that have been faced during the one year timespan of the ongoing research period. I believe, with every good thing, comes difficulty. However, the problems are mentioned below.

7.1.1 Air Quality Monitor

- Difficulty to distinguish the work breakdown structure.
- Listing out all the initiatives that can be done in the project.
- Determining the start date, due date of the initiatives done in thep project in Gantt Chart.

7.1.2 RefinAir

- Finding a proper library for all the sensors.
- Calibrating all the sensors.
- Establishing smooth UART communication between PMS5003 and MH-Z19B sensors.
- Troubleshooting the python application to decrease errors in serial communication.
- Unavailability of components due to pandemic situation.
- Sudden increase of cost and shipping charge due to lockdown and pandemic.
- Unable to go to the lab and do the project together.
- Virtual screening of the project.

7.1.3 Prediction of PM2.5 concentrations using Machine Learning

- Finding out the correct procedure to extract data.
- Retrieving data from satellites.
- Retrieving satellite derived AOD 550nm data for short radius distance. Due to high missing values for short radius distances.
- Finding out the formula to convert AOD to PM2.5 concentration.
- Finding out the correct prediction model to predict the PM2.5 concentration.

7.2 Environmental, Social, Ethical issues

We declare no conflict of interest. The environmental impact given by the product we made. Pestle Web Analysis is presented below that shows the environmental, social and ethical issues.

PestleWeb Analysis

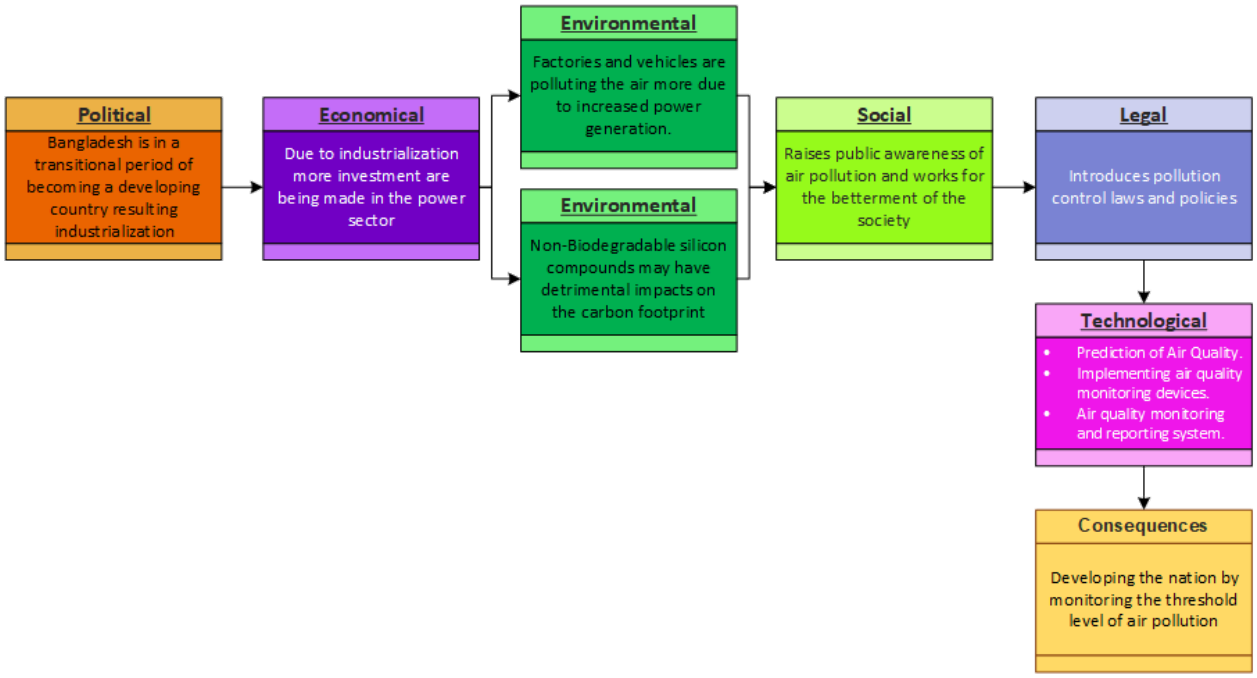

Figure 7.2 : Environmental, Social, Ethical Issues

7.3 Recommendation / Future Scope of Work

In the future, we aspire to make an ‘Atmospheric Map’ that can be generated with the help of satellite data using MLAs, and this can help the citizens to determine the most polluting areas and take precautions accordingly. The industrialists will get alarmed about the situation and control the pollution in favourable conditions.

Additionally, our aim is to take an initiative on making a mobile application that will indicate the state of pollution of the regions around the country using the standard color code of the Air Quality Index. We would like to incorporate satellite data here to cover all the areas around the country, in case our device can not reach those places, mostly water bodies. We wish to inaugurate weather stations all around the country for fetching data and spreading awareness to the citizens of this country.

References

I. Al-Ali, A. R., Zualkernan, I., & Aloul, F. (2010, October). A Mobile GPRS-Sensors Array for Air Pollution Monitoring. *IEEE Sensors Journal*, *10*, 1666-1671. 10.1109/JSEN.2010.2045890

II. Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5-32.

III. Brimblecombe, P. (Ed.). (n.d.). *The Effects Of Air Pollution On The Built Environment* (Vol. 2). Imperial College Press. www.books.google.com.bd/books?id=m--3CgAAQBAJ&lpg=PP1&pg=PR5#v=onepage&q&f=false

IV. Castanas, E., & Kampa, M. (2007). Human health effects of air pollution. *Environmental Pollution*, *151*, 362-367. 10.1016/j.envpol.2007.06.012

V. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

VI. Dhingra, S., Madda, R. B., Gandomi, A. H., Patan, R., & Daneshmand, M. (2019, June). Internet of Things Mobile–Air Pollution Monitoring System (IoT-Mobair). *IEEE Internet of Things Journal*, *6*(3), 5577-5584. 10.1109/JIOT.2019.2903821

VII. Emili, E., A.Lyapustin, Y.Wang, C. Popp, S.Korkin, M.Zebisch, S. Wunderle, & M. Petitta. (2011, December 16). High spatial resolution aerosol retrieval with MAIAC: Application to mountain regions. *JOURNAL OF GEOPHYSICAL RESEARCH*, *116*(D23). 10.1029/2011JD016297

VIII. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine,. *Ann. Statist*, *29*(5), 1189-1232.

IX. Holovaty, A., Teslyuk, V., Lobur, M., Pobereyko, S., & Sokolovsky, Y. (2018). Development of Arduino-Based Embedded System for Detection of Toxic Gases in Air.

2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 139-142. 10.1109/STC-CSIT.2018.8526672.

X. Khashei, M., & Bijari, M. (2010). An artificial neural network (p,d,q) model for timeseries forecasting. *Expert Systems with Applications*, *37*(1), 479-489. https://doi.org/10.1016/j.eswa.2009.05.044

XI. Kiruthika, R., & Umamakeswari, A. (2017). Low cost pollution control and air quality monitoring system using Raspberry Pi for Internet of Things. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 2319-2326. 10.1109/ICECDS.2017.8389867

XII. Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating Ground-Level PM2.5 by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach. *Geophysical Research Letters*, *44*(23), 11985-11993. 10.1002/2017GL075710

XIII. Li, X., & Zhang, X. (2019, June). Predicting ground-level PM2.5 concentrations in the Beijing-Tianjin-Hebei region: A hybrid remote sensing and machine learning approach. *Environmental Pollution*, *249*, 735-749. doi.org/10.1016/j.envpol.2019.03.068

XIV. Li, Z., -LamYim, S. H., & Ho, K. -F. (2020, September 20). High temporal resolution prediction of street-level PM2.5 and NOx concentrations using machine learning approach. *Journal of Cleaner Production*, *268*(121975). https://doi.org/10.1016/j.jclepro.2020.121975

XV. Mayer, H. (1999, October). Air pollution in cities. *Atmospheric Environment*, *33*(24-25), 4029-4037. https://doi.org/10.1016/S1352-2310(99)00144-2

XVI. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*.

XVII. Stern, A. C. (1977). *Air Pollution: The Effects of Air Pollution* (3rd ed., Vol. 2). A Subsidiary of Harcourt Brace Jovanovich. www.books.google.com.bd/books?id=07FA4WIBkf8C&lpg=PP1&ots=REoqwig1oj&dq=air%20pollution%20effects&lr&pg=PR3#v=onepage&q&f=false

XVIII. Stirnberg, R., Cermak, J., Fuchs, J., & Andersen, H. (2020, February 6). Mapping and Understanding Patterns of Air Quality Using Satellite Data and Machine Learning. *Journal of Geophysical Research: Atmospheres*, *125*(4). https://doi.org/10.1029/2019JD031380

XIX. Suganya, E., & Vijayashaarathi, S. (2016). Smart vehicle monitoring system for air pollution detection using Wsn. *International Conference on Communication and Signal Processing (ICCSP) 2016*. 10.1109/ICCSP.2016.7754238

XX. Wei, J., Sun, L., Peng, Y., Wang, L., Zhang, Z., Bilal, M., & Ma, Y. (2018, October 22). An Improved High-Spatial-Resolution Aerosol Retrieval Algorithm for MODIS Images Over Land. *Journal of Geophysical Research: Atmospheres*, *123*(21), 12291-12307. https://doi.org/10.1029/2017JD027795

XXI. Yang, Y., & Li, L. (2015). A Smart Sensor System for Air Quality Monitoring and Massive Data Collection. *ICTC 2015*, 147-152. 10.1109/ICTC.2015.7354515