Bachelor of Science in Computer Science & Engineering

Development of IoT Based System for Air Pollution Data Analysis and Visualization

by

Nayan Das

ID: 1504021

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

April, 2021

Development of IoT Based System for Air Pollution Data Analysis and Visualization

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Nayan Das
ID: 1504021

Supervised by

Dr. Asaduzzaman
Professor
Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)
Chattogram-4349, Bangladesh.

The thesis titled '**Development of IoT Based System for Air Pollution Data Analysis and Visualization'** submitted by ID: 1504021, Session 2019-2020 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

______________________________ Chairman

Dr. Asaduzzaman
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

______________________________ Member (Ex-Officio)

Dr. Asaduzzaman
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

______________________________ Member (External)

Dr. Md. Mokammel Haque
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

Signature of the candidate

Date:

Acknowledgements

I am grateful to Almighty God, who has given me the ability to complete my project and intend to perform the completion of B.Sc. Engineering degree. I am indebted to my supervisor Dr. Asaduzzaman, Professor, Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, for his motivation, proper guidance, constructive criticisms, and endless patience progress of the project. All these things contributed to my humble development as a network analysis practitioner. He supported me by providing books, conference and journal papers, and practical bits of advice. From the very beginning, sir always encouraged me with proper guidelines, so the project never seemed a burden. All my teachers helped me with their invaluable assistance in every phase of my learning process over the course of four years. Finally, I'd like to express my appreciation to my friends, seniors, and the department staff for their valuable suggestions and assistance in completing the project.

Abstract

Air quality has become the most concerning matter around the world. The World air quality report 2021 claimed that South Asia, including Bangladesh, is the worst air polluted country, where Dhaka is the second most polluted city. Air pollution causes everybody to see the dilemma as it causes severe diseases related to the lungs, which can be acute or chronic. Good air quality depends on the particular amount of different substances within a range in the air; otherwise, it becomes a matter of concern. These measurements rely on various ingredients such as the rate of CO, CO_2, NO_2, Alcohol, Benzine, Particular Matter, Dust, and so on. Our primary purpose is to build an IoT-based model to collect data from the sensors. We have used five parameters for monitoring pollution. These are the smoke, CO, particular matter, temperature, and humidity. A real-time monitoring system developed for front-end that allows knowing the current situation with exact values indicated by various parameters. This work also includes the data analysis, which can also predict the condition according to the previous data with a machine learning model. A comparison of various prediction algorithms such as linear regression, logistic regression, KNN, SVM, random forest and decision tree, including accuracy and test case analysis. We also analyze the correlation between air quality index and weather condition (temperature and humidity) for different location and time. All these things make people understand the gruesomeness of the pollution and support to be aware.

Keywords— Air Quality, IoT-based model, Real-time monitoring, Machine Learning, Prediction, Evaluation

Table of Contents

Acknowledgements **iii**

Abstract **iv**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**

1.1 Introduction 1

1.1.1 Objective 2

1.2 Framework/Design Overview 2

1.3 Difficulties 2

1.4 Applications 4

1.5 Motivation 4

1.6 Contribution of the thesis 5

1.7 Thesis Organization 5

1.8 Conclusion 6

2 Literature Review **7**

2.1 Introduction 7

2.2 Related Literature Review 7

2.2.1 Air Quality Index 7

2.2.2 AQI Measurement 7

2.2.3 Traditional Air Quality Monitoring 8

2.2.4 IoT Based Water Quality Monitoring System 9

2.2.5 Data collection module 10

2.2.6 Data transmission module 10

2.2.7 Model Prediction and Visualization 10

2.2.7.1 Visualization 12

2.2.8 Real Time monitoring 12

2.2.9 Background Study of the Problem 12

2.2.10 Present State of the Problem 14

2.3 Conclusion 14

2.3.1 Implementation Challenges 15

3 Methodology **16**
3.1 Introduction 16
3.2 Overview of Framework 16
3.2.1 Experimental Setup 16
3.2.2 System Design for Hardware 17
3.2.2.1 Arduino 17
3.2.2.2 NodeMcu 18
3.2.2.3 Sensors 19
3.2.2.4 MQ135 (Smoke) Sensor 19
3.2.2.5 MQ7 (CO) Sensor 20
3.2.2.6 GP2Y1010AU0F (Dust) Sensor 21
3.2.2.7 DHT11 (Temperature & Humidity) Sensor 21
3.2.3 Building the Setup 23
3.2.3.1 MQ135 sensor with Arduino 23
3.2.3.2 MQ7 sensor with Arduino 24
3.2.3.3 GP2Y1010AU0F sensor with Arduino 24
3.2.3.4 DHT11 sensor with Arduino 24
3.2.4 Arduino and NodeMcu Connection 26
3.3 Detailed Explanation 26
3.3.1 Data Flow of the Project 26
3.3.2 Device Design 28
3.3.3 Android.cc IDE Coding 28
3.3.4 System design For Software Implementation 30
3.3.5 Real-time Monitoring & Visualization 31
3.3.6 Implementation 33
3.4 Conclusion 33

4 Results and Discussions **34**
4.1 Introduction 34
4.2 Dataset Description 34
4.3 Impact Analysis 35
4.3.1 Social and Environmental Impact 36
4.3.2 Ethical Impact 36
4.4 Evaluation of Framework 37
4.5 Evaluation of Performance 39
4.6 Conclusion 40

5 Conclusion **44**
5.1 Conclusion 44

5.2 Future Work 45

List of Figures

1.1 System Overview . . . 3

2.1 AQI Index . . . 8
2.2 Serial Communication between Arduino Uno and Nodemcu . . . 11

3.1 Arduino Uno . . . 18
3.2 ESP8266 NodeMcu . . . 19
3.3 MQ135 . . . 20
3.4 MQ7 . . . 21
3.5 GP2Y1010AU0F(Dust) Sensor . . . 22
3.6 DHT11 Sensor . . . 22
3.7 MQ135 sensor with Arduino . . . 24
3.8 MQ7 sensor with Arduino . . . 25
3.9 GP2Y1010AU0F sensor with Arduino . . . 25
3.10 DHT11 sensor with Arduino . . . 26
3.11 Flow Chart . . . 27
3.12 Front View of the Device . . . 29
3.13 Side View of the Device . . . 29
3.14 Arduino.cc IDE for NodeMcu . . . 30
3.15 Home page . . . 31
3.16 Combo Chart according to Temperature . . . 32
3.17 Histogram . . . 32
3.18 Scatter Chart . . . 32
3.19 Combo Chart according to Humidity . . . 33

4.1 Dataset . . . 36
4.2 Scatter Chart (Value changes with respect to Time) . . . 38
4.3 Values changes respect to Temperature and Humidity . . . 39
4.4 Labelling changes respect to Temperature and Humidity . . . 40
4.5 Histogram for four Datasets . . . 42
4.6 Correlations for four Datasets . . . 43
4.7 Heat Map . . . 43

List of Tables

3.1 Connection between MQ135 and Arduino 23
3.2 Connection between MQ7 and Arduino 24
3.3 Connection between NodeMcu and Arduino 26

4.1 Air Condition Leveling 34
4.2 Accuracy Analysis based on five parameters 41
4.3 Accuracy Analysis based on three parameters 41
4.4 Test Case Comparison based on five parameters 41
4.5 Test Case Comparison based on three parameters 41

Chapter 1

Introduction

1.1 Introduction

Air pollution has become a significant concern among people around the world. The rate of air pollution is constantly increasing at an alarming rate. The reason behind the air pollution is an admixture of natural and man-made substances in the very air we breathe. Emission from vehicles, factories, Fuel combustion, Burning of wastes, fossil fuel, domestic coal, Forest fires, Dust storms, Smoke from power plants, Coal mining, Farming chemicals, Cigarette smoke, Pollen, etc. is the notable contribution of pollution.

Air pollution causes 7 million people to suffer each year from acute and chronic illnesses such as asthma, bronchial symptoms, lung infection, decreased lung capacity, cardiovascular diseases, and lung cancer. Most importantly, children are suffering badly due to nature's hazardous conditions, impacting neurodevelopment and cognitive ability.

As the population explosion rate increases daily, developing countries like Bangladesh are getting exposed to air pollution. 1.23 lakh people have been killed by air pollution, responsible for indoor and outdoor pollution in Bangladesh in 2017 [1]. The common air pollutants are CO (Carbon monoxide), CO_2 (Carbon dioxide), O_3 (ozone), NO_2 (Nitrogen dioxide), Particular Matter ($PM_{2.5}$, PM_{10}, $PM_{0.1}$), Pb (Lead), Benzene, Alcohol, Suspended Particular Matter (SPM), etc [2]. Air pollution can be measured by Air Quality Index (AQI), representing air quality in different ranges. In Bangladesh, five pollutants (PM_{10}, $PM_{2.5}$, NO_2, SO_2, CO, O_3) are mainly responsible for computing AQI.

Our intention is to make awareness of unhealthy air conditions and their terrible consequences by creating an IoT-based model to monitor the air quality for three pollutants, contaminate the air, and forecast the AQI level. IoT an interrelating system among computing devices, machines, and other objects without making human interaction. A real-time monitoring system is the primary concern of this project, and it also helps predict the AQI level. Prediction is a machine learning approach that will help analyze the datasets of AQI level at a different time. It may save innocent lives by taking necessary precautions.

1.1.1 Objective

1. To build an IoT device with specific sensors for particular pollutants.
2. Data from sensors would be collected and visualized in various location, time and weather condition.
3. Machine learning is being used to analyze the collected data and predicted the pollution level using AQI.
4. Build a real time monitoring and visualization model for diverse group of users.

1.2 Framework/Design Overview

The system overview is quite simple. With an IoT device, the flow starts. This device consists of several sensors. The device is connected to the server, where continuous data has been stored. From this data, a dataset can be made, which helps to future analysis of the system. This server also connected with a website to perform real-time monitoring and visualization.

1.3 Difficulties

Several difficulties were faced during project work. The main problem was the device configuration. This complex device has some limitations too. The connection was getting disconnected continuously. Maintaining the server was another difficulty. Server connection and interruption of collecting data with the device

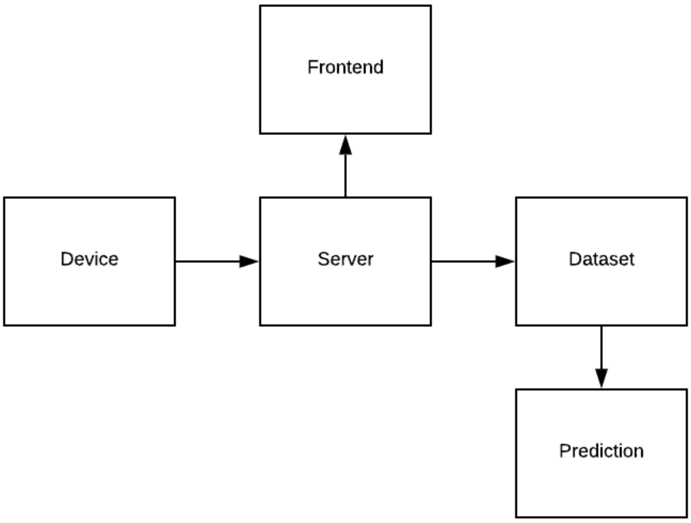

Figure 1.1: System Overview

was the toughest task. After that, a calculation AQI value of the individual sensor and their labeling at the dataset was also an arduous task. The below are some of the most challenging scenarios for IoT technology:

1. Signaling and Bandwidth: One of the most critical requirements for implementing the net of factors technologies is the availability of continuous, efficient, and bi-directional signaling. That is the first and most important condition for the IoT age to work flawlessly. It is the underlying principle that turns the concept of IoT facts streams into action. Both devices linked over a network and centralized servers and other third-party equipment necessitate continuous reliable signaling [3]. Whatever the case might be, the most important condition for applying IoT platforms is that data must transfer as quickly and reliably as possible from point a to point b. Consequently, bandwidth use is an important issue that all IoT devices will face in the future.

2. Boundaries to long-term deployment: Monitoring systems that may crash are sending in high-frequency data. It can break and lose power during connection build-up. For this reason, inaccurate results can be shown [4].

3. Security: One of the significant concerns with IoT technology is security. For continuous connection, its important to secure the passing information

[5]. One such vulnerability is well shown. A slew of 'matters' connected, exchanging data the whole generation might be sensitive to security issues.

1.4 Applications

Since air quality has become a significant concern around the world, It really necessary to monitor the rate of pollution to know the place is suitable for living or not. When accurate knowledge regarding indoor and outdoor air quality is accessible, the climate control device will provide the most efficient airflow, maintaining clean and healthy living conditions [6]. Some crucial area is more important to implement this kind of device. School, college, varsity and other institutions are essential to be under monitoring for good health. Most importantly, hospital areas are the most suitable location to check the condition of the natural air. People can also implement this monitoring process in their home for their very own health care. There are several application regarding this:

1. Indoor Air Quality Monitoring System
2. Outdoor Air Quality Monitoring System
3. Particulate Matter Monitoring
4. Gas Detection System

1.5 Motivation

The primary motivation working behind this thesis is summarized below:

1. To detect and control the air pollution before rising the pollution level at an alarming rate. As the pollution increases day-by-day and gets out of control, it needs to be prevented. For preventing the problem, the monitoring is the first thing to know the exact condition of the situation. By which it can be controlled by creating alerts.

2. To predict the pollution rate and determine whether the place is suitable for living or not based on collected data. Many places like highways, industrial areas, construction areas, etc. are heavily contaminated with pollution

which are not suitable for living without having safety measures. By this analysis, future planning can be done for the public health care by taking necessary steps.

3. To incorporate with the advancement of the smart city.

1.6 Contribution of the thesis

A specific set of goals is achieved by performing thesis or research work. The areas of contribution of the thesis can be summarized as follows:

1. Designing an IoT device with specific sensors.
2. Collecting data and maintaining the server with appropriate request.
3. Classify the data regarding the air pollution and predict them with several algorithm.
4. Designing a website for real time monitoring.

1.7 Thesis Organization

The following is how the remainder of this study paper is organized:

1. Chapter 2 discusses the previous studies in Air quality monitoring related works and different prediction algorithm based works.
2. Chapter 3 illustrates the detailed work of the proposed methodology including designing the device and how it works with a particular formula. AQI value calculation and the entire project works. It will also explain the classification and prediction of the dataset and how it goes with a different algorithm. It will finally describe the real-time monitoring website and visualization.
3. Chapter 4 describes the working dataset and analysis of the accuracy with random test cases.
4. Chapter 5 contains the overall summary of this thesis work and provides some future recommendations.

1.8 Conclusion

In this chapter, an overview is provided. Along with the difficulties, the epitome of this whole system has been described in this chapter. The motivation behind this work and contributions are also stated here. In the next chapter, the background and present state of the problem will be given.

Chapter 2

Literature Review

2.1 Introduction

The aim of this study is to look at the difficulties that are to be faced while researching the air quality monitoring system and its data analysis and visualization. By providing a summary of the previous study, this chapter discusses various applications previously done and their limitation and advantages. A detailed description of both these methods is provided in this chapter.

2.2 Related Literature Review

2.2.1 Air Quality Index

For measuring pollution level Air quality index has been used. It informs us about air conditions. It ranges between 0-500, leveling as good, moderate, cautious, unhealthy, very unhealthy, and extremely unhealthy. AQI value of more than 300 considers a hazardous situation that is totally unfavorable for living. An adequate amount of AQI value is below 100 for a healthy environment. It's getting unhealthy with the increasing rate of AQI. Air Quality Index (AQI) for Bangladesh:

2.2.2 AQI Measurement

The calculation of AQI can be done by following formula [7]:

$$I = \frac{I_{High} - I_{Low}}{C_{High} - C_{Low}} * (C - C_{Low}) + I_{Low} \tag{2.1}$$

AQI Range	Category		Color
	In English	*In Bengali*	
0-50	Good	ভালো	Green
51-100	Moderate	-	Yellow Green
101-150	Caution	-	Yellow
151-200	Unhealthy	অস্বাস্থ্যকর	Orange
201-300	Very Unhealthy	খুব অস্বাস্থ্যকর	Red
301-500	Extremely Unhealthy	অত্যন্ত অস্বাস্থ্যকর	Purple

Figure 2.1: AQI Index

Here,

I = Air Quality Index

C = the pollutant concentration

C_{low} = the concentration breakpoint which is $\leq C$

C_{high} = the concentration breakpoint which is $\geq C$

I_{low} = the index breakpoint corresponding to C_{low}

I_{high} = the index breakpoint corresponding to C_{high}

C_{low}, C_{high}, I_{low}, I_{high} are from the Pollutant Breakpoint

2.2.3 Traditional Air Quality Monitoring

Traditional Air Quality: Air quality can be measured traditionally. The traditional way is not that efficient and convenient. Sometimes it can be more costly. It may also provide low-resolution data. Monitoring stations are spread out more sparsely. This system can be built on wireless sensor network technology and is connected to the global mobile communications infrastructure [8].

Indicators for indoor air quality may be calculated physically, chemically, biochemically, or ecologically, depending on the observations. Indoor air quality is mostly determined by analyzing samples obtained from the atmosphere. Passive mechanisms collect the majority of air samples obtained by laboratories. In the absence of an industry norm, the evaluator's knowledge and skills determine the form of samples, number of samples, analytes, and methodological preferences. As a result, it's understandable to ask what metrics are used in air quality practice to determine sampling, monitoring, and other procedures. Day-to-day observations in and about the vicinity in enclosed areas can be used to determine indoor air quality. All these are described below [9]:

Physical: Functional markers are valuable information that can be obtained by sight and scent. Air quality is also defined as affecting withering, discoloration, and unpleasant odors in and around a closed system, among other things.

Chemical: Chemical sensors are also helpful in detecting low indoor air quality. Blackening of copper electric wire, irregular shading, splitting, etc. are some of the positive indications.

Biological: Asthma, allergy to aeroallergens, and other airborne illnesses can be the indicator of polluted air, especially for microbial pollutants.

Environmental: The presence of sour flora and fauna, as well as climatic and other weather conditions, habits, and appearance is a strong indication of air quality. Certain changes in weather levels in and near inhabitable regions are strong indicators of indoor air pollution owing to carbon gases from garbage in the vicinity, smoke from forest fires, waterlogging, floodwater runoff, hazardous waste, among other sources.

2.2.4 IoT Based Water Quality Monitoring System

As the air pollution is increasing day-by-day, an automatic monitoring system needs to be developed. For this whole process, we need to collect natural data and send it to server. An internet connectivity is required to implement this, and the devices should be integrated with wireless communication model. For this reason, IoT based system is the best solution of this problem.

2.2.5 Data collection module

Natural air has been used to collect raw data. Indoor and outdoor both sources could be the option here. It could be good, moderate, cautious, unhealthy, very unhealthy, incredibly unhealthy. A specific sensor is needed to collect the data. In this case, smoke sensor, carbon monoxide, particular matter sensor, temperature and humidity sensor. By using all these sensors, a collection module will be developed.

2.2.6 Data transmission module

Data will be collected from different types of sensors. A microcontroller, Arduino, is connected with the sensors. We have chosen Arduino as it has analog pins, unlike another microcontroller like Raspberry Pie. Arduino Uno has six analog input pins. But Arduino doesn't have any wifi module. So, it can't be connected to the internet alone. Another transmission module called NodeMcu has been used for this reason to make the connectivity with the internet. By this NodeMcu, data can be transferred to the webserver to store the data. Different types of communication protocols can be used for transmission. Three methods out there for analog transmission are:

1. SPI
2. I2C
3. UART

The devices should be linked as shown below to provide serial contact between them.

2.2.7 Model Prediction and Visualization

Predictive modeling and predictive analytics are the same things that are a mathematical technique that attempts to forecast future events or outcomes by evaluating patterns that are likely to predict future results. In our project, we have used several algorithms to check the accuracy and to know which are the best-fit algorithm in our works. A short description is illustrated below:

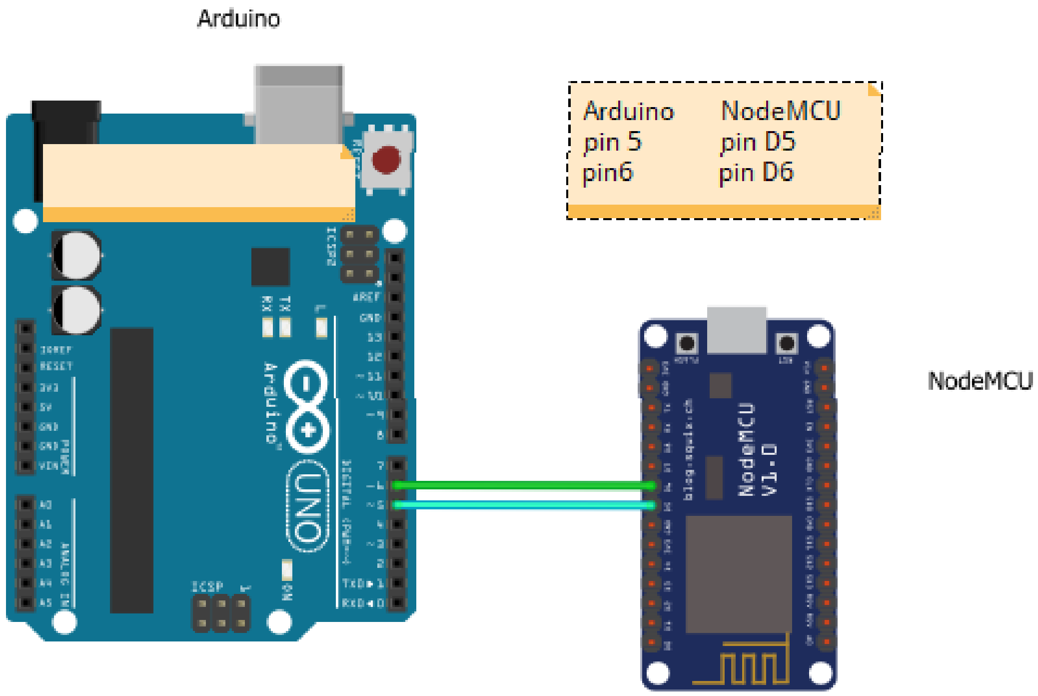

Figure 2.2: Serial Communication between Arduino Uno and Nodemcu

Decision Trees: Data (mined, open access, internal) is graphed out in branches by decision tree algorithms to show the potential outcomes of different decisions. Decision trees characterize response variables and forecast response variables based on previous choices, are readily explainable and open to beginner data scientists, and can be used for incomplete data sets.

Logistic Regression: This is a data processing approach that uses mathematical analysis. The algorithm's ability to sort and classify data increases as more data is used, allowing predictions to be made.

Linear Regression: Linear regression can obtain a relation linking dependant and independent variables. For making a prediction, linear regression can be used by fitting them. It predicts the dependent variable.

SVM: Classification and regression problems can be resolved by the Support Vector Machine. It is, however, primarily used to solve classification problems. Each data object is mapped in this algorithm. A drawing line has been plotted, and data has been split into both sides.

KNN: Another supervised learning algorithm named K-nearest neighbor is used for the same reason. It has also been used for both analyses. It is mostly used in industry to solve classification and prediction problems.

Random Forest: Random samples have been selected from the datasets. A decision has been built after the selection process, and finally, for each expected outcome, polling will be conducted.

2.2.7.1 Visualization

Data visualization provides a lot of information that data alone cannot provide. Python provides some of the most engaging data visualization tools available. The more basic plot styles are spread through libraries, but some are exclusive to specific libraries. We have used some visualization tools that are Matplotlib and Seaborn.

2.2.8 Real Time monitoring

An administrator may check, analyze, and change the inclusion, deletion, alteration, and use of data on software, a database, or a device using real-time data monitoring (RTDM). Via graphical maps and bars on a central interface/dashboard, data managers will review the entire procedures and tasks executed on the data in real-time, even when they happen.

2.2.9 Background Study of the Problem

As an alarming concern, some authors have worked on-air pollution-based research papers. Some IoT based system for air pollution are described below:

In [7] authors have developed an IoT model for monitoring air quality in Thailand. They have worked on several parameters like CO, SO2, O3, NO2, and PM10. They have used Raspberry Pi and Arduino Mega to complete the project. They have implemented in a city of Thailand and ends the project with a satisfactory result. Their works are quite costly and they didn't consider any analysis part for classification of the data. The authors [10] in this paper suggest the green

IoT, which can detect air pollution and be powered by renewable energy. They had worked on outdoor air pollution and a notification system to notify users when the pollution exceeded the acceptable level.They didn't use any real-time monitoring system to check the pollution rate continuously and also didn't do any kind of analysis. In this paper [11], the authors have done monitoring air pollutants' levels using a wide-ranged sensory network. Deployment of sensor nodes on a large scale is one of the key roles of this project. It describes a detailed survey work about air pollution monitoring. They haven't implemented the system instead finishing a survey work. In this paper [12], the authors have done monitoring using a portable sensory system's design and fabrication. Light alert signals and sound alert signals have been used to notify the user. They have worked on three parameters whereas we have worked on five parameter with no real-time monitoring system. They have missed two crucial parameter called smoke and carbon monoxide. The Authors [13] have built a simulation of air quality at a regional scale. They used a sophisticated forecasting method called the Community Multiscale Air Quality Model, which was developed by Weather Research and Forecasting (WRF) (CMAQ). CMAQ can underestimate the pollutant concentrations in Dhaka. Two-level nested domains were implemented using WRF and CMAQ. The authors [14] have predicted based on PM10 and PM2.5. Several machine learning model have been used here to predict the data. The autoregressive nonlinear neural network remains the best among them to predict air pollution with external input. The Authors [15] have used Global Positioning System (GPS) location for acquiring carbon dioxide and carbon monoxide levels as a parameter in the air. They have used Azure cloud service as a database. They have been used Microsoft's Azure Machine learning service. They have analyzed the previous data to make prediction. In this paper [16], a monitoring system have been built by the authors with only one pollutant detection sensor named MQ135 sensor which use Bluetooth communication to transfer data in a mobile application. In this paper [17], they have worked on three different sensors with real-time monitoring system. Their limitations are that they didn't consider particular matter which is the most crucial parameter for air quality detection. They didn't consider any kind of further analysis of the data.

2.2.10 Present State of the Problem

For monitoring the pollution rate, the real-time monitoring device will be built and for future analysis purposes, we have analyzed the whole datasets. From previous studies, we have found some lackings in their projects. Summary of them is described below:

1. The previous works had the bonus of using Raspberry Pi. But when it comes to cost efficiency, we are using Arduino Uno and NodeMCU ESP8266, which is cheaper than a Raspberry Pi and other microcontrollers.

2. Some of them have worked on particular pollutants monitoring systems. They worked with one or two parameters but we will work on different pollutants that affect our health directly. We have considered five pollutants in this case which are smoke, CO, particular matter 2.5, temperature and humidity. By this we can also understand the weather condition of that time.

3. Almost none of them have shown the visualization analysis whereas we have considered it for further analysis. We have shown the changes of values with respect to different parameters.

4. Machine learning techniques for data analysis will add a new dimension to this project. Multiple algorithms will be tested for prediction, and we will show the comparison analysis of their accuracy.

2.3 Conclusion

All the background study has been discussed in this section. The limitations have also been discussed in this section. Several parameters have been used to develop the system. Besides this device, a real-time monitoring system and machine learning analysis has also done. In the next chapter, details of the work procedure will be illustrated. Finally, prediction models and it's result analysis adds a new dimension to the works.

2.3.1 Implementation Challenges

A few implementation challenges are designing a device, collecting data into a server, calculating the AQI values, modeling the dataset for future prediction, visualization of the collected data, and finally, the real-time monitoring system to show the current situation.

Chapter 3

Methodology

3.1 Introduction

An IoT device for air quality monitoring system has been developed. In this section, the details of the work procedure has been discussed here. Necessary snapshots will be attached to understand the total procedure easily. Pictures of the device from different angles have also been attached too. We express our system's layout, describe the building set up for hardware, and describe the software setup and analysis.We have also provided the location-wise visualization and its analysis.It also includes the detail of the real-time monitoring system.

3.2 Overview of Framework

3.2.1 Experimental Setup

The following is a list of methods that have been used to incorporate an IoT-based air pollution monitoring and visualization system:

1. Hardware

    - Desktop / Laptop computer
    - Arduino Uno
    - NodeMcu
    - MQ135(Smoke) Sensor
    - MQ7(CO) Sensor
    - GP2Y1010AU0F(Dust) Sensor

- DHT11(Temperature and Humidity) Sensor
- Bread Board and Cables
- Adapter

2. Software
    - Operating System: Windows 10
    - Arduino.cc IDE
    - Front End is developed by HTML, CSS, JavaScript, Bootstrap
    - MySQL is used as the backend for storing related data
    - Python for model prediction
    - Google colab for visualization
    - Google chart on website for visualization

3.2.2 System Design for Hardware

Before knowing the hardware level design, we need to know the hardware description. Simple discussion about all the hardwares have been described below which have been used for designing the device. In a nutshell, the description is given below:

3.2.2.1 Arduino

We use Arduino UNO to build our device, which a user-friendly platform for organizing projects. It has a physical programmable circuit board and an IDE platform to program the specific function. This IDE runs the CPU by writing code onto it. Several models are available in the market for Arduino. Arduino Uno is one of them, and we used it for our project. It is cost-efficient than the others. It has six(6) analog input pins. It also has 14 Digital I/O Pins(of which 3 provide PWM output). Its DC per I/O Pin is 40 mA.

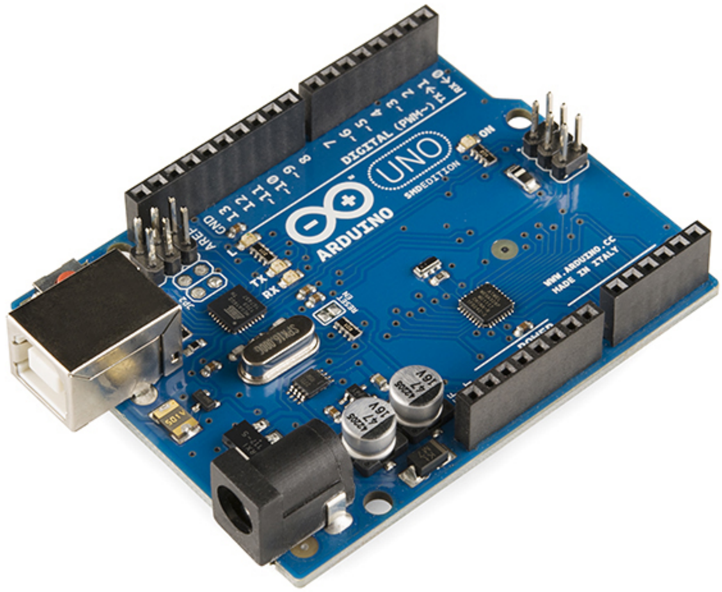

Figure 3.1: Arduino Uno

3.2.2.2 NodeMcu

The ESP8266 WiFi Module is a self-contained microchip (SOC) with an integrated TCP/IP protocol stack that can be used to bind any microcontroller to a WiFi network. The ESP8266 may be used to host an application or to offload all WiFi networking functions to another cpu. Every ESP8266 board has the AT command set firmware pre-programmed; you can simply connect it to the Arduino computer and get around as much WiFi functionality as a WiFi Shield. The ESP8266 module is a low-cost board with a burgeoning group. This module's on-board processing and storage capabilities are appropriate. The ESP8266 supports APSD for VoIP applications and Bluetooth co-existence protocols; it has a self-calibrated RF that allows it to run under all operating conditions and requires no additional RF components; and it has a self-calibrated RF that allows it to operate in all operating conditions and requires no additional RF parts. There is an almost infinite supply of content for the ESP8266, much of which has been rendered possible by tremendous community help.

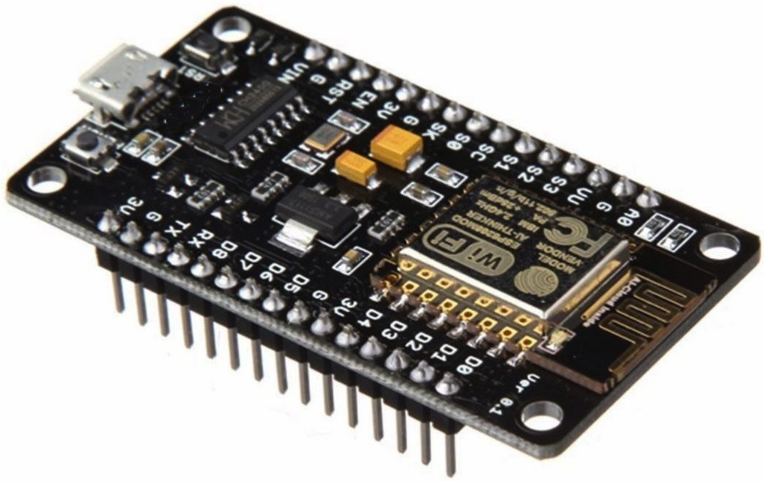

Figure 3.2: ESP8266 NodeMcu

3.2.2.3 Sensors

For this project, we have used four sensors to measure several parameters of air. These are Smoke sensor, CO sensor, Dust Sensor, Temperature and Humidity sensor. A brief description of sensors is described below:

3.2.2.4 MQ135 (Smoke) Sensor

The MQ-135 Gas Sensors have been used in air quality monitoring systems to identify or measure NH3, NOx, Alcohol, Benzene, Smoke, and CO2. Since you only need to sense one gas, the MQ-135 sensor module has a Digital Pin that enables it to operate without a microcontroller. Since the analog pin is TTL driven and runs at 5V, it can be used with a wide variety of microcontrollers.

1. MQ-135 Sensor FeaturesWide detecting scope

    - It has wide detecting scope
    - It also has the fast response and High Sensitivity
    - It includes stable and long life

- It has operating voltage is +5V.
- NH_3, NO_x, alcohol, Benzene, smoke, CO_2, etc., can be detected.
- Its Analog output voltage is 0V to 5V.
- Its Digital output voltage is 0V or 5V (TTL Logic)
- Its Preheat duration 20 seconds.
- It can be used as a digital or analog sensor.
- The Sensitivity of the Digital pin can be varied using the potentiometer.

Figure 3.3: MQ135

3.2.2.5 MQ7 (CO) Sensor

MQ7 is a Carbon Monoxide (CO) sensor that can detect PPM levels of CO in the air. CO concentrations of 20 to 2000 ppm can be measured by the MQ7 Gas sensor. The sensitivity of this sensor is high, and the response time is fast. The sensor's contribution is an analog resistance.

1. Sensitive for carbon monoxide.
2. Its output voltage rises as the concentration of the gases being tested rises.
3. It has a fast response and recovery.
4. It has adjustable sensitivity.
5. It has a signal output indicator.

6. It needs power around 2.5V $\sim$ 5.0V.

Figure 3.4: MQ7

3.2.2.6 GP2Y1010AU0F (Dust) Sensor

The Sharp Optical Dust Sensor (GP2Y1010AU0F) is a common component in air purifier systems, and it's especially good at detecting small particulate matter like cigarette smoke. To track the reflected light of dust in the air, this device uses an infrared emitting diode and a phototransistor mounted diagonally. This device's electrical scheme is quite simplistic. The air sensor consumes very little power (20mA at most, 11mA on average) and can be operated directly from the Arduino board. With a sensitivity of 0.5V/0.1mg/m3, the sensor's output is an analog voltage equivalent to the calculated dust density.

1. Shallow values make it more alert.
2. It consumes very little power and can be driven directly from the Arduino.
3. It is easily customizable.
4. It has an analog output, which allows it to be directly attached to a measurement instrument without any external circuitry.

3.2.2.7 DHT11 (Temperature & Humidity) Sensor

As a temperature and humidity monitor, the DHT11 has been used. A dedicated NTC for temperature calculation is included in the sensor, as well as an 8-bit microcontroller for serial data output of temperature and humidity values. Since the sensor is factory-tuned, it's easy to attach it to other microcontrollers. The

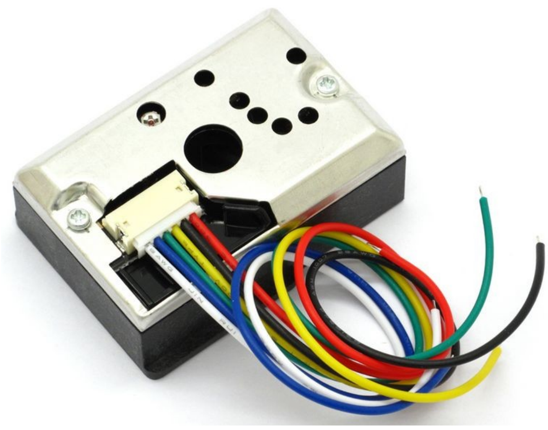

Figure 3.5: GP2Y1010AU0F(Dust) Sensor

sensor will calculate temperature from 0°C to 50°C and humidity from 20% to 90% with a precision of 1°C to 1 degree.

1. Its operating voltage is 3.5V to 5.5V.
2. Its operating current is 0.3mA (measuring) 60uA (standby)
3. It provides serial data as an output.
4. Its temperature range is 0°C to 50°C
5. Its humidity range is 20% to 90%.
6. Its accuracy is ±1°C and ±1%

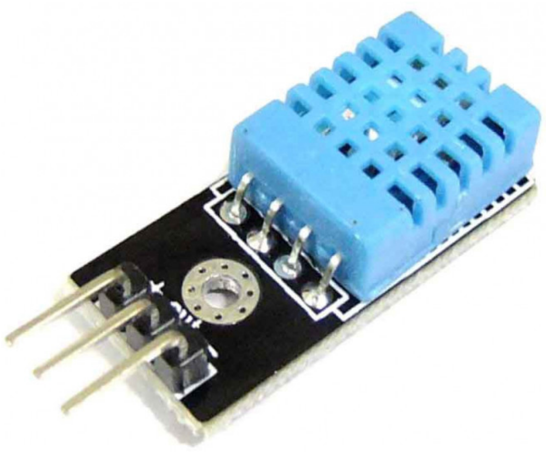

Figure 3.6: DHT11 Sensor

Table 3.1: Connection between MQ135 and Arduino

SL	MQ135	Arduino
1	+5 Vcc	+5 Vcc
2	GND	GND
3	DATA	A0

3.2.3 Building the Setup

This setup makes use of a NodeMcu for overall power and an Arduino Uno for sensing experiments. Despite the presence of NodeMcu, a microprocessor-based single-board computer (SBC), it can seem redundant and needless to integrate Arduino with the framework. Also in the early stages, only a single piece of NodeMcu was required to perform all of the system's sensing and regulating operations. The whole prototype can be broken down into two pieces. The NodeMcu is included in the central control and Internet access module, and the extensible sensor box module is used in the extensible sensor box package. The Arduino and all of the attached sensors are housed in the sensor case. There are actually four analog series sensors housed in it. The NodeMcu processes and uploads sensor readings from the sensor box, which are sent down to the central control module. The primary function of the main control module is to accept control instructions from an external controller and drive the device accordingly. Four analog sensors are wired to the Arduino Uno's analog input pins. A Wifi module is also included in the platform, which uses the 802.11n standard. Via serial communication of TX and RX ports, the Arduino is also linked to the NodeMcu. As a result, we've wired up all of the sensors to the Arduino and read them using the NodeMcu.

3.2.3.1 MQ135 sensor with Arduino

Mq135 is a conventional analog sensor. We'll need an analog pin in Arduino to attach the sensor. On the MQ135 sensor, we used the A0 pin on the Arduino Uno. To bind the +5V Vcc and GND, we used a breadboard.

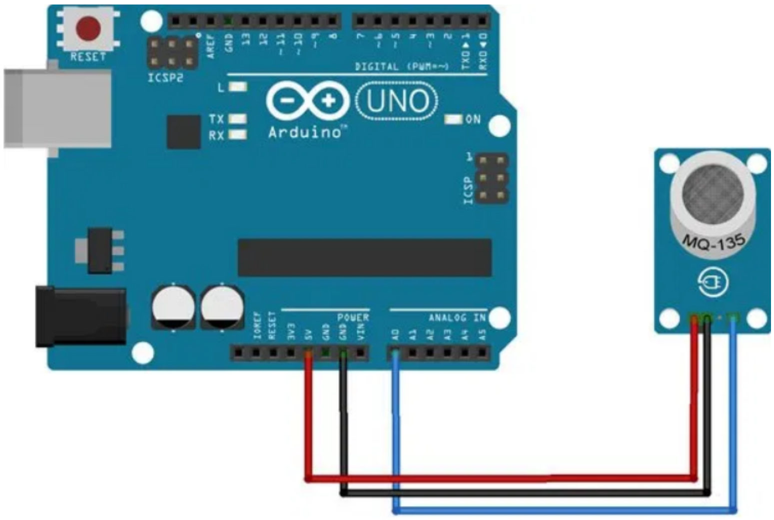

Figure 3.7: MQ135 sensor with Arduino

Table 3.2: Connection between MQ7 and Arduino

SL	MQ135	Arduino
1	+5 Vcc	+5 Vcc
2	GND	GND
3	DATA	A1

3.2.3.2 MQ7 sensor with Arduino

Mq7 is a traditional analog sensor. We'll need an analog pin in Arduino to attach the sensor. On the MQ7 sensor, we used the A1 pin on the Arduino Uno. To bind the +5V Vcc and GND, we used a breadboard.

3.2.3.3 GP2Y1010AU0F sensor with Arduino

GP2Y1010AU0F is also an analog sensor. It also has a 220 µF capacitor and 150 ohm resistor. The analog pin of the sensor is connected with the A5 pin of arduino. The following diagram has the detail view of the connection.

3.2.3.4 DHT11 sensor with Arduino

Water vapor can be detected by the DHT11 by measuring the electrical resistance between two electrodes. The humidity sensing component is a moisture-holding substrate with electrodes on the top. The substrate emits ions when it consumes water vapor, increasing the conductivity between the electrodes. The relative humidity is equal to the change in resistance between the two electrodes. The resistance between the electrodes decreases as the relative humidity rises, while

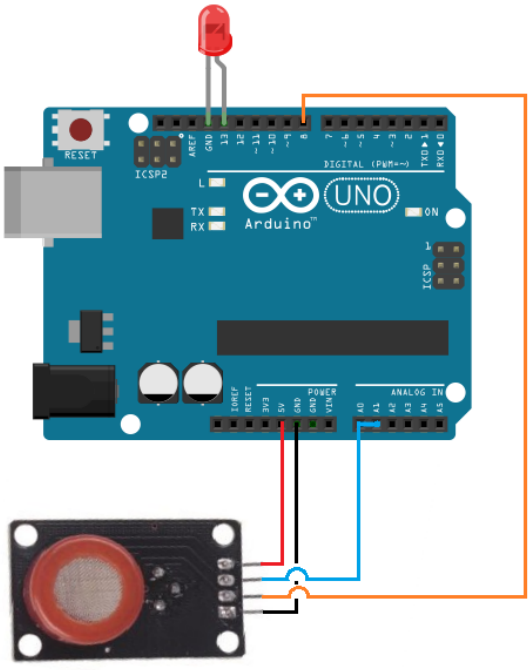

Figure 3.8: MQ7 sensor with Arduino

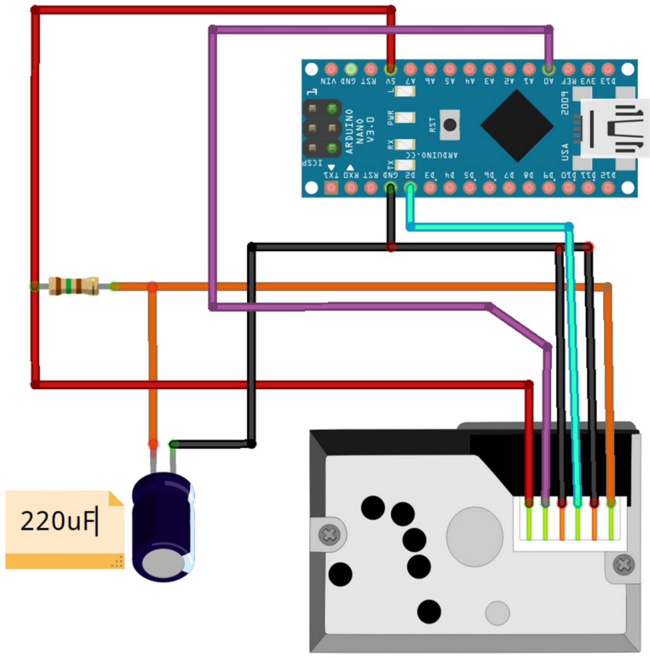

Figure 3.9: GP2Y1010AU0F sensor with Arduino

the resistance between the electrodes increases as the relative humidity falls. The connectivity of the sensor with arduino is described in the diagram.

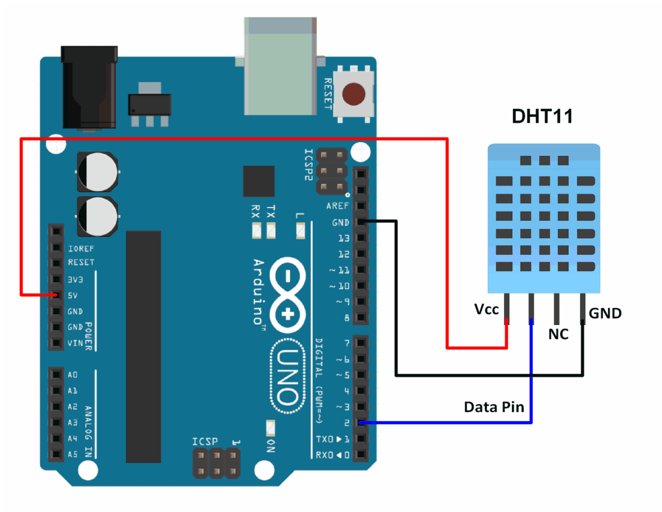

Figure 3.10: DHT11 sensor with Arduino

Table 3.3: Connection between NodeMcu and Arduino

SL	NodeMcu	Arduino
1	RX	TX
2	TX	RX
3	GND	GND

3.2.4 Arduino and NodeMcu Connection

To submit data from the Arduino to the NodeMcu, serial communication is needed. In the previous chapter, we discussed the three types of serial communication between Arduino and Arduino, as well as Arduino and NodeMcu. We may use UART contact since we only need a single byte stream for the outputs of four sensors. The Arduino's TX and RX ports are TX0 and RX1, respectively. In NodeMcu, RX and TX are included. For the GND to be attached on both sides of two microcontrollers, two things are needed. Another important consideration is that the RX port of the NodeMcu must be wired to the TX port of the Arduino and vice versa. The NodeMcu RX can process the data sent by the Arduino TX (T for transmission) (R for receive).

3.3 Detailed Explanation

3.3.1 Data Flow of the Project

The proposed flow chart or methodology is described here

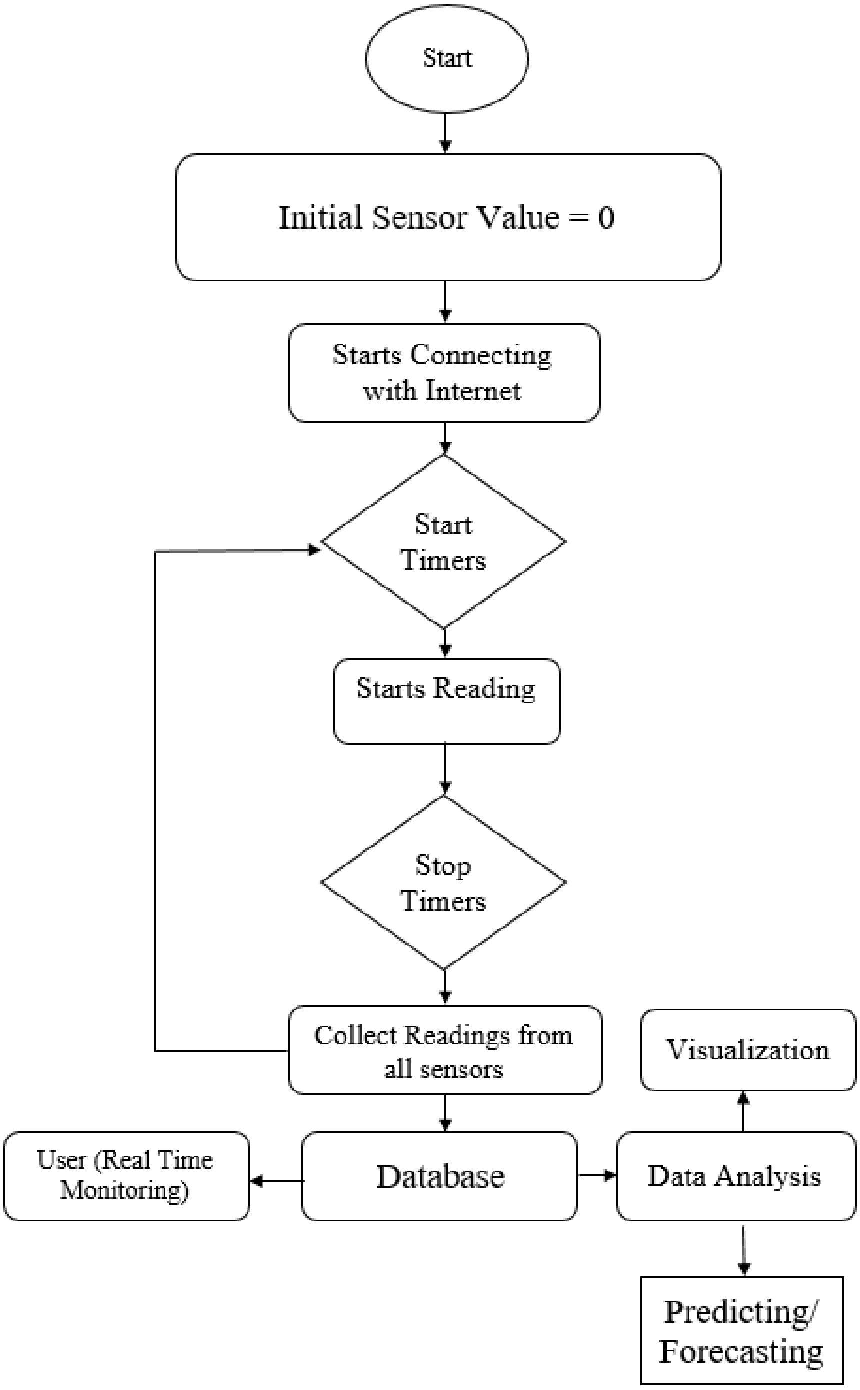

Figure 3.11: Flow Chart

1. The whole procedure starts after receiving data as an input by gas sensors respectively.

2. Initialization needs to be done. Initially all the sensor value is zero.

3. After initializing, device have to connect with internet to reach the server.

4. A timer will start after the connection building up. In this project, we setup one minute delay.
5. After this all the sensors starts to collect its data.
6. After one minute the timer will stop.
7. All data can be collected from the sensor and send it to server through internet. And it will go back to the step no 4.
8. Database store the data collected from the device.
9. This data can be shown in a real-time monitoring website with visualization.
10. From this data, data analysis can also be done.
11. After analyzing data, the system can predict the AQI values.
12. Visualization of different location data can also be done after data analysis.

3.3.2 Device Design

After giving all the connections, we combined them in a place. Arduino module has a connection setup with all sensors, and the WiFi module was also integrated after that. We placed both modules in a customized box. By giving a power supply, it was finally able to collect data. Then it sends the data to the server through NodeMcu. Data has been sent to the server continuously after one minute delay. Three sensors have been integrated into the front side of the device, and another sensor with a power supply have been integrated on another side of the device.

3.3.3 Android.cc IDE Coding

Arduino.cc is a forum for developing electronic projects that is both free and open-source.Arduino consists of two components: a basic programmable circuit board (also known as a microcontroller) and IDE (Integrated Development Environment) software that runs on a laptop and is used to write down and link laptop code to the physical board. For good cause, the Arduino platform has become quite common with people who are just getting started with electronics.

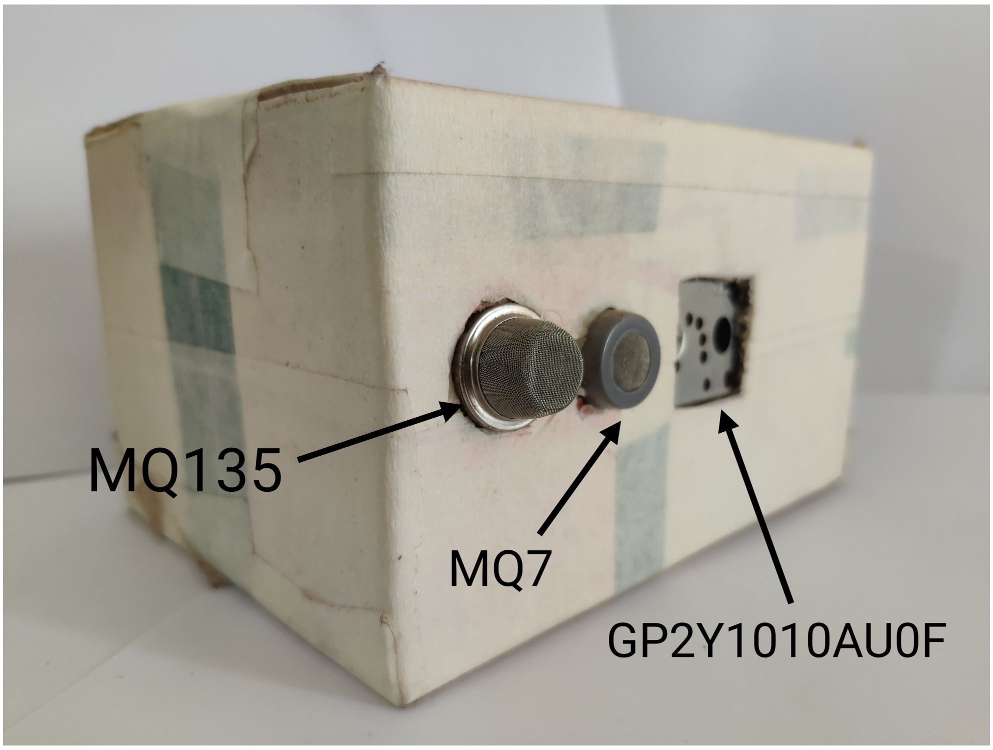

Figure 3.12: Front View of the Device

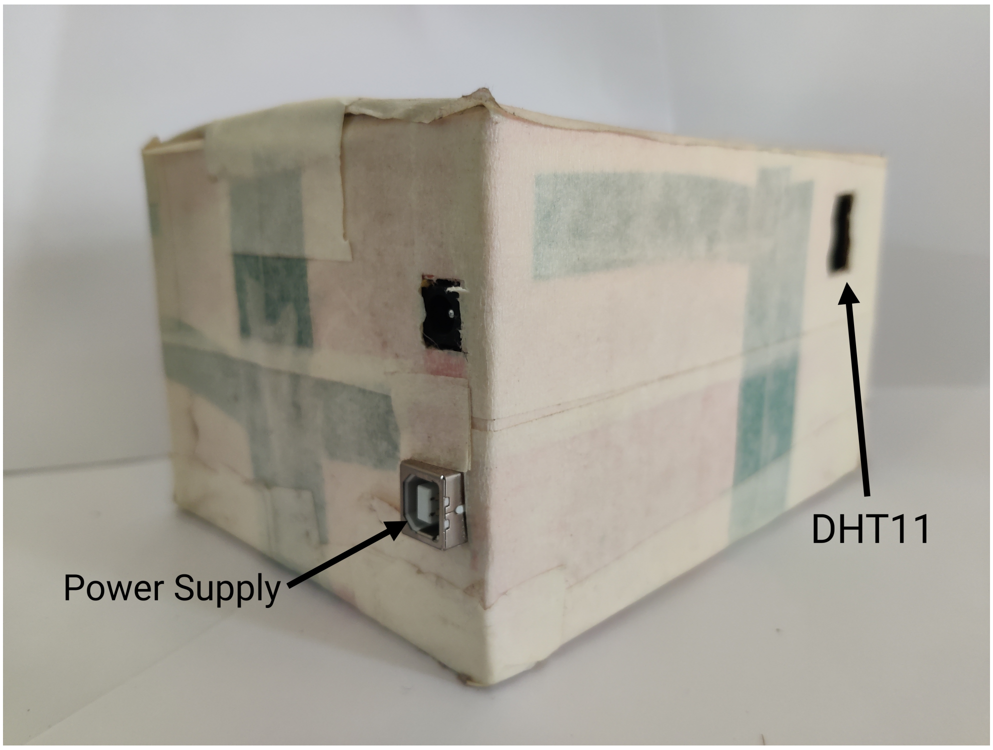

Figure 3.13: Side View of the Device

The Arduino, unlike previous programmable circuit boards, does not need a specific piece of hardware (known as a programmer) to load new programming onto the board; instead, all that is needed is a USB cable. Furthermore, the Arduino

IDE supports a simplified edition of C++, which makes it simpler to learn. As a consequence, Arduino has a fashionable form factor that compartmentalizes the microcontroller's functionality into a more manageable kit. We essentially used the analogRead() function to read an analog sensor's values in our project. We have used several default libraries for the WiFi module and, most notably, the link to the MySQL server.

3.3.4 System design For Software Implementation

For the use of the NodeMcu and Arduino dependent air quality monitoring system, various modular programs are written in C and C++ language in the Arduino.cc IDE. The critical device module sets up the air quality monitoring sensors, WiFi module, and serial terminal and verifies all sensor links at the current source stage. To conduct a real-time air quality monitoring system, a website has been created. The utilities are the smoke, CO, dust, temperature, humidity, and decision in the real-time update. Calibration mode presents the facility to calibrate the device with standard buffer samples. The web server is connected with the website to monitor the pollution rate and the visualization. The website offers a front view to show air quality parameters alongside smoke, CO, dust, temperature, and humidity. We store data for future prediction. Dataset provides a facility to collect location information and store them in the excel file format. Google chart has been used to visualize the data from the MySQL server.

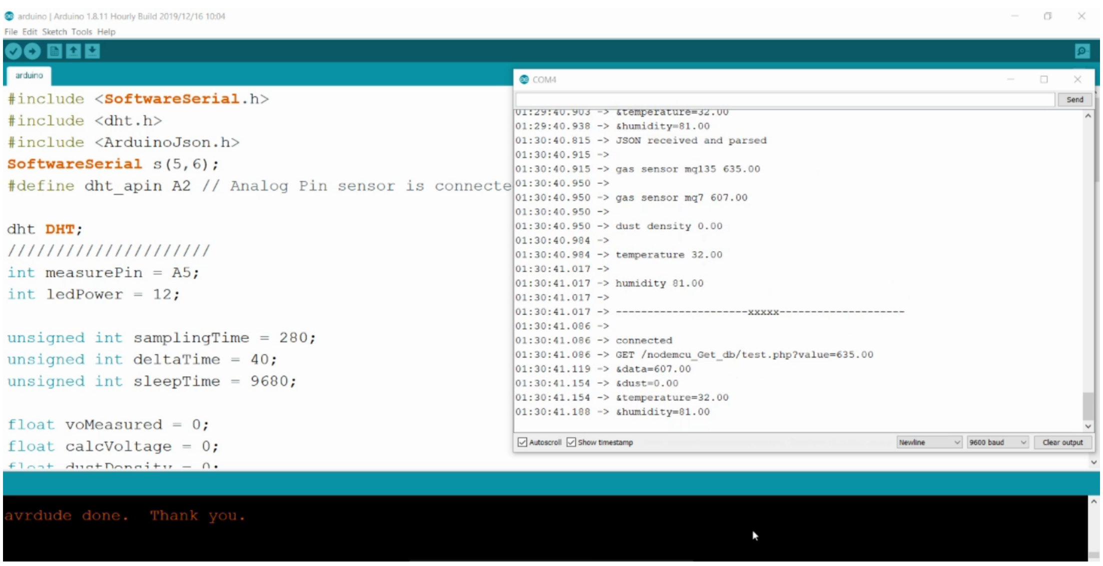

Figure 3.14: Arduino.cc IDE for NodeMcu

3.3.5 Real-time Monitoring & Visualization

Sensors collect all the raw data, and through NodeMcu, it sends data to the MySQL server. From this server, we fetch the value of all the sensors to show in a frontend. For this, we use HTML,CSS,Bootstrap and Javascript. For the backend, we use PHP to fetch the data using the get() method. Our simple website has five columns: the current value of smoke, CO, dust, temperature, and humidity of any specific location where the device has been set. The data will change continuously after one minute duration. Here are some color representations for three sensors: Good as Green color, Moderate as Blue Color, Caution as Yellow Color, Unhealthy as Red Color, Very Unhealthy as Ash Color, Extremely Unhealthy as Black Color. There are color indications for the temperature sensor too. Blue represents the cold temperature, yellow as the average one, and finally, red represents the hot temperature. For humidity sensor data, there is only one blue color indication. On the home page, there are some other options related to the graphs. There are four graphs on the page. These are the Histogram, Scatter Chart, and two combo charts for different parameters. For these charts, we have used google chart to integrate with our website. From this chart, anyone can see the real-time visualization of the total number of data on the dataset.

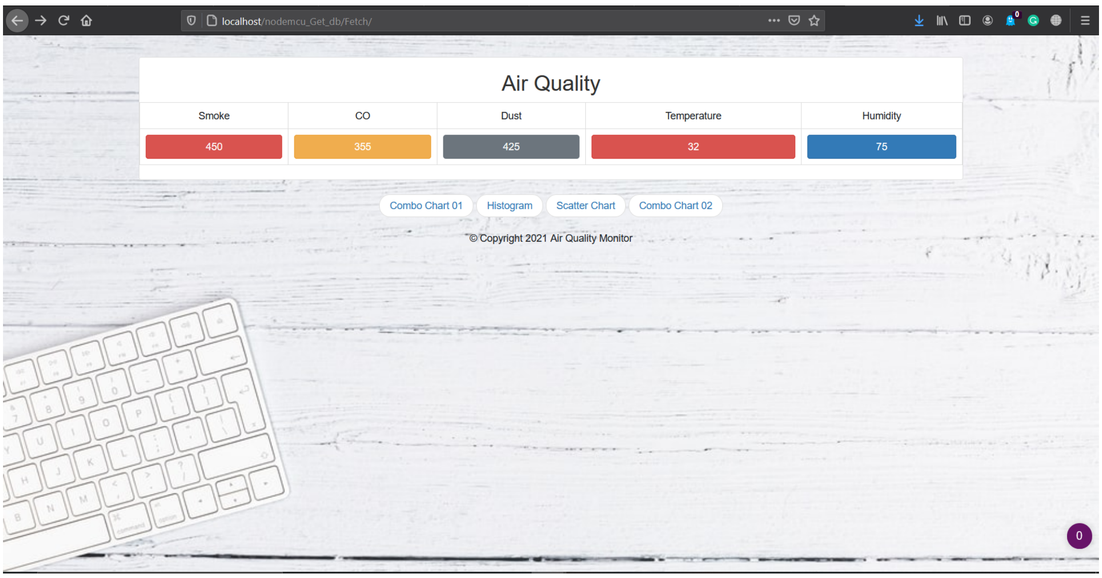

Figure 3.15: Home page

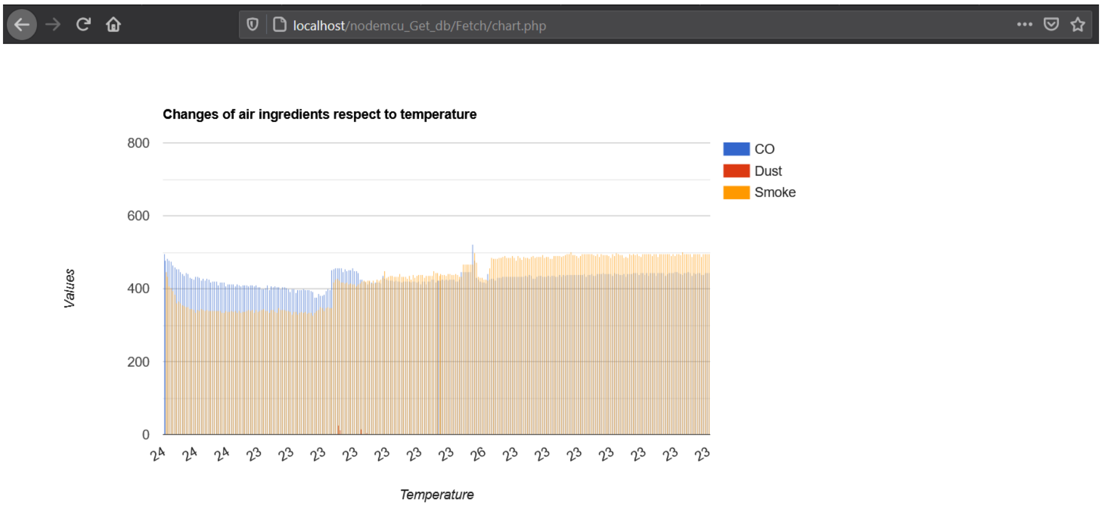

Figure 3.16: Combo Chart according to Temperature

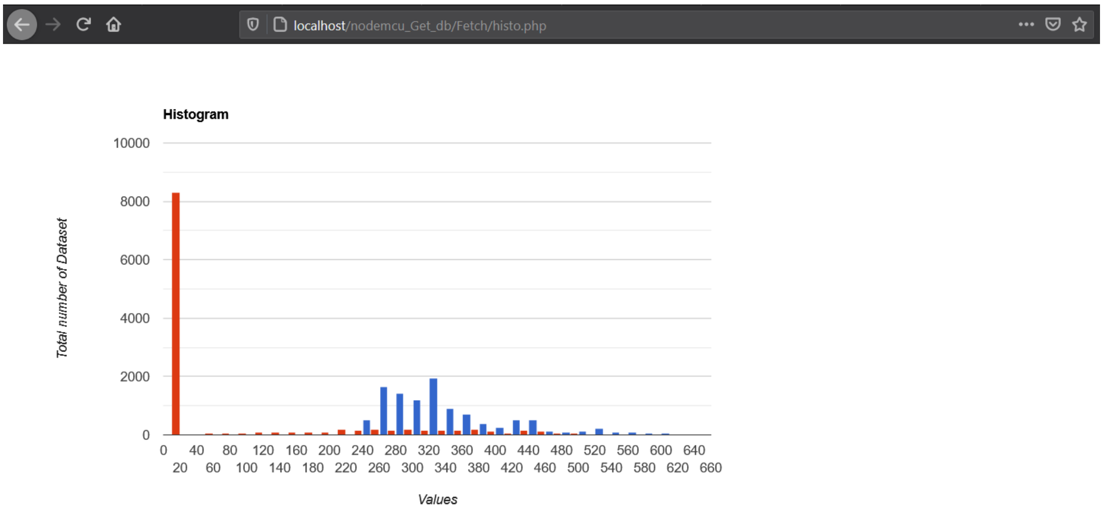

Figure 3.17: Histogram

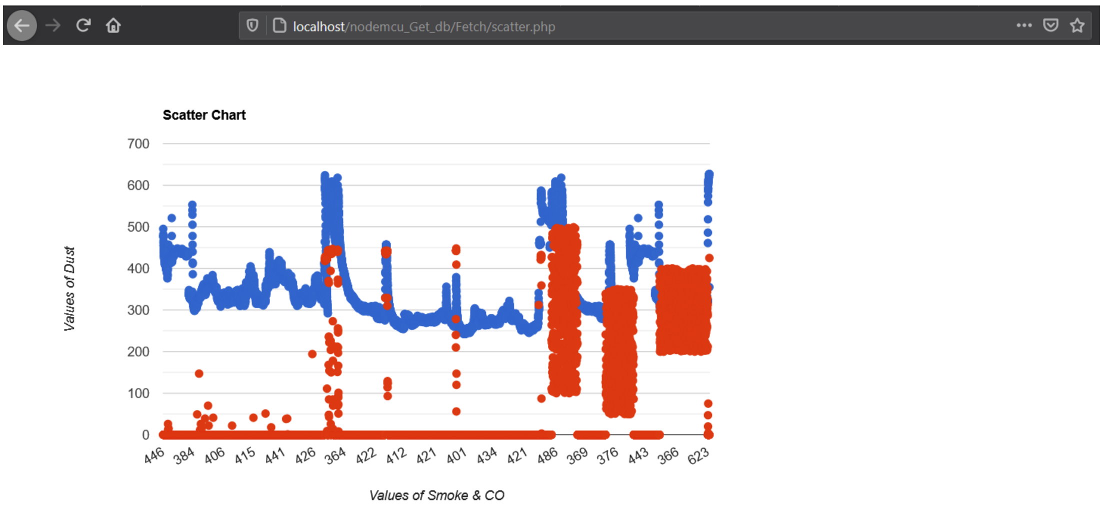

Figure 3.18: Scatter Chart

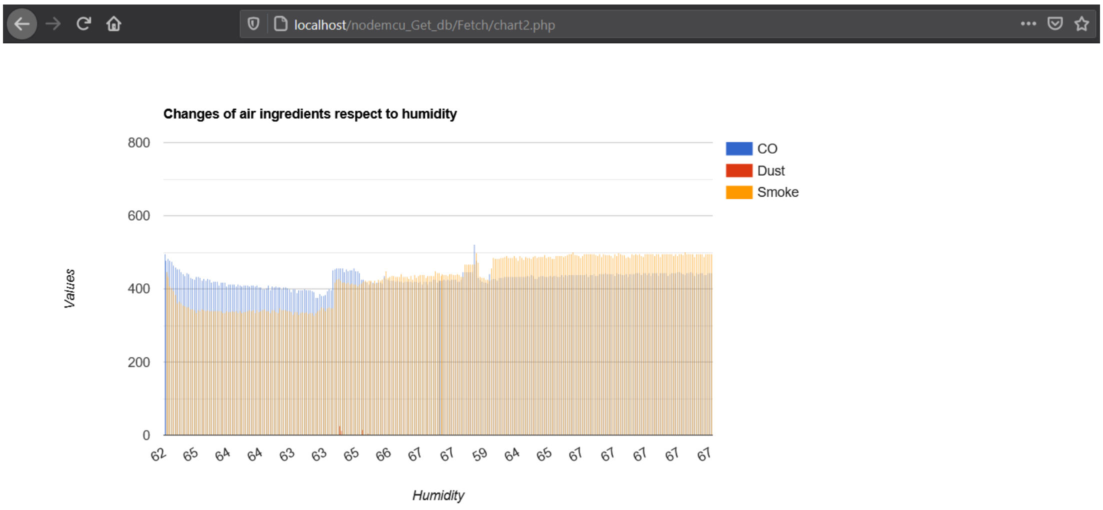

Figure 3.19: Combo Chart according to Humidity

3.3.6 Implementation

In this section, we have implemented the whole device setup. With the data flow diagram, we have completed the majority of the task in this section. We also describe the detailed description of the real-time monitoring section of the project, one of the leading implementations of this project.

3.4 Conclusion

We've provided a brief overview of the air quality monitoring system's hardware integration and visualization in this segment. The NodeMcu is used for universal controlling operations, and the Arduino Uno is used for sensing tasks. After sending and receiving the data, transmisson has ended. This data has been sent to server and then it will be ready for further processing. This whole analysis procedure will be discussed in the next chapter.

Chapter 4

Results and Discussions

4.1 Introduction

Two key goals will be discussed in this chapter as we stored data on a dataset. One is to analyze the data. For this, we train the dataset with several algorithms. We consider four different datasets for this analysis part. Three of them are the three additional location data, and the other one is the merged dataset. The second goal is the visualization based on a different dataset.

4.2 Dataset Description

In this project, we have analyzed different pollutants of the air to determine the air pollution. For this, we have selected five parameters for our analysis. And we have collected data from three different locations. One of the locations was in a rough condition, that's why we found a variation of the pollutants. We saw wild dust data in that location. That five pollutants are smoke, carbon monoxide, dust, temperature, and humidity. To make it easier, we have leveled the air condition with one to six numbers. We have collected ten thousands two hundreds and nineteen (10219) data. A table 4.1 is shown the details of the condition leveling.

Table 4.1: Air Condition Leveling

Level	Condition
1	Good
2	Moderate
3	Caution
4	Unhealthy
5	Very Unhealthy
6	Extremely Unhealthy

There are also some other attributes in the dataset, which is described below:

Value: This column contains the MQ135 sensor data, which is directly collected from the device. This data represents the smoke concentration in the air.

Data: MQ7 sensor data is stored here. This data represents the value ofcarbon monoxide.

Dust: Here dust density of air is stored from the device for any particular location.

Temperature: Current temperature value is stored in this column.

Humidity: This column contains the humidity value of the specific location.

Date: Current date is stored to know the timing of the location data.

Time: This database automatically updated after one minute delay. This whole time schedule is added in this column.

l135: Levelled data of the MQ135 sensor is stored here after calculating AQI values.

l7: Levelled data of the MQ7 sensor is stored here.

ldust: This column contains the leveled data of dust density.

Lebel: This column has the final leveling of the data, which indicates the air quality.

We have collected this type of three datasets for three different locations. After collecting all three dataset, a merged dataset has been built to analysis with total number of data. A snapshot of dataset is given below:

4.3 Impact Analysis

In this whole project, we have tried to build an air quality monitoring system. From this monitoring system, air pollution rates can be determined in a specific location. This project has a hardware part, a device, and a software part where analysis is done. It can be beneficial in both ways. By the monitoring part,

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id	value	data	dust	temperature	humidity	date	time	l135	l7	ldust	average	lebel	location
2	12704	489	515	0	28	42	2/21/2021	15:13:36	4	5	1	3.333333333	3	1
3	12705	494	516	0	28	42	2/21/2021	15:14:37	4	5	1	3.333333333	3	1
4	12707	555	541	442	28	42	2/21/2021	15:17:18	5	5	5	5	4	1
5	12708	575	556	443	28	43	2/21/2021	15:18:18	5	5	5	5	4	1
6	12709	589	570	439	28	44	2/21/2021	15:19:18	5	5	5	5	4	1
7	12710	580	578	440	28	44	2/21/2021	15:20:19	5	5	5	5	4	1
8	12711	583	584	441	28	44	2/21/2021	15:21:19	5	5	5	5	4	1
9	12712	574	583	441	28	44	2/21/2021	15:22:19	5	5	5	5	4	1
10	12713	558	580	444	28	44	2/21/2021	15:23:20	5	5	5	5	4	1
11	12714	557	576	442	28	44	2/21/2021	15:24:20	5	5	5	5	4	1
12	12715	546	568	443	28	44	2/21/2021	15:25:20	5	5	5	5	4	1
13	12716	536	563	442	27	43	2/21/2021	15:26:20	5	5	5	5	4	1
14	12717	524	550	370	27	43	2/21/2021	15:27:21	5	5	4	4.666666667	4	1
15	12718	515	552	168	28	44	2/21/2021	15:28:21	5	5	2	4	4	1
16	12719	504	545	6	27	43	2/21/2021	15:29:21	5	5	1	3.666666667	4	1
17	12720	498	542	0	27	43	2/21/2021	15:30:21	4	5	1	3.333333333	3	1
18	12721	497	532	0	27	43	2/21/2021	15:31:22	4	5	1	3.333333333	3	1
19	12722	500	535	48	27	43	2/21/2021	15:32:22	4	5	1	3.333333333	3	1
20	12723	508	537	236	27	45	2/21/2021	15:33:22	5	5	3	4.333333333	4	1
21	12724	513	538	365	28	45	2/21/2021	15:34:22	5	5	4	4.666666667	4	1
22	12725	509	537	235	27	44	2/21/2021	15:35:23	5	5	3	4.333333333	4	1
23	12726	504	532	85	27	44	2/21/2021	15:36:23	5	5	1	3.666666667	4	1
24	12727	501	530	43	27	44	2/21/2021	15:37:23	5	5	1	3.666666667	4	1
25	12728	488	527	0	28	45	2/21/2021	15:38:24	4	5	1	3.333333333	3	1
26	12729	492	524	0	27	44	2/21/2021	15:39:24	4	5	1	3.333333333	3	1
27	12730	486	522	3	27	44	2/21/2021	15:40:24	4	5	1	3.333333333	3	1
28	12731	501	523	222	27	44	2/21/2021	15:41:24	5	5	3	4.333333333	4	1
29	12732	498	516	154	27	44	2/21/2021	15:42:25	4	5	2	3.666666667	4	1
30	12733	500	517	442	27	44	2/21/2021	15:43:25	4	5	5	4.666666667	4	1

merge123

Figure 4.1: Dataset

real-time conditions can be known. And with the analysis part, the air quality condition can be analyzed by graph visualization. And the prediction can also help see the air quality condition in the future. This device can be installed anywhere with a power supply. The air quality of any location can make people aware of the place. This project has social, environmental, ethical impacts. This can be implemented in any critical site.

4.3.1 Social and Environmental Impact

Social and Environment impact: This project has a very crucial effect on society and the environment. Recently, the pollution rate is increasing a lot. Cigarette smoking, industrial pollution, smoke from vehicles, etc., are increasing day by day. In this situation, regular monitoring of air pollution can save people from all these sufferings. From this pollution rate, it can easily understand that the place is really suitable for living or not. This is also effective for kids and elderly people to make contact with fresh air, and by this, it can be easily known. It also saves the Environment by taking necessary steps to reduce the pollution rate after watching the graph of changes.

4.3.2 Ethical Impact

This project has some ethical consequences too. This monitoring system can be implemented in schools, colleges, varsity, and other institutions for a better environment. If the pollution rate increases, this will harm students. Most importantly, it can be installed in hospital areas where fresh air is badly needed. It can be installed in industrial and construction areas too. People can be able to know their environmental condition, which is directly related to their health.

4.4 Evaluation of Framework

The analysis part of the report is one of the important parts of this project. Machine learning implementation has been done here. The model has been trained on different algorithms. Four datasets have been used for this. Three of them are the individual datasets of locations, and the other one is the merged dataset. We have used some of the python libraries to use the model training. These are pandas, scikit-learn, etc. Generally, scikit-learn suitable for numeric data. Our dataset also contains all the numeric data. Scikit-learn also best for classification, regression analysis. Pandas is used for data analysis and manipulation. We have used SVM, KNN, Linear regression, Logistic Regression, Random Forest, and Decision Tree algorithm for model prediction. Using this library, we import the dataset. Dataset has been split into two parts, one part is train data and another one is the test data. Generally, 75% of data has been separated into the train data and another 25% of data has been separated into test data. Model training has been done on different algorithms, and at last, it has been tested on the test data to know the accuracy. Accuracy and random test case checking will be discussed in the next section. We implemented several graphs based on different parameters.

1. On figure 4.2, we have shown the scatter chart based on four different datasets. This charts represent the changes of values with respect to time.

2. On figure 4.3 shows the changes in the pollutant's value with respect to temperature and humidity. It is clearly seen that, dust and other values increase as the temperature increases and low humidity also has heavy dust density.

3. On figure 4.4 shows the changes of the pollutant's AQI value with respect

to temperature and humidity. It can easily help to understand the overall situations of the environment. This 6 levels distributed as the six overall conditions of the AQI lebelling.

4. On figure 4.5 shows that the histogram analysis of the four datasets.

5. On figure 4.6 shows the correlation between the parameters. It's shown for four datasets too.

6. Finally on figure 4.7, the heat map which shows us that there are no null values on the merged dataset.

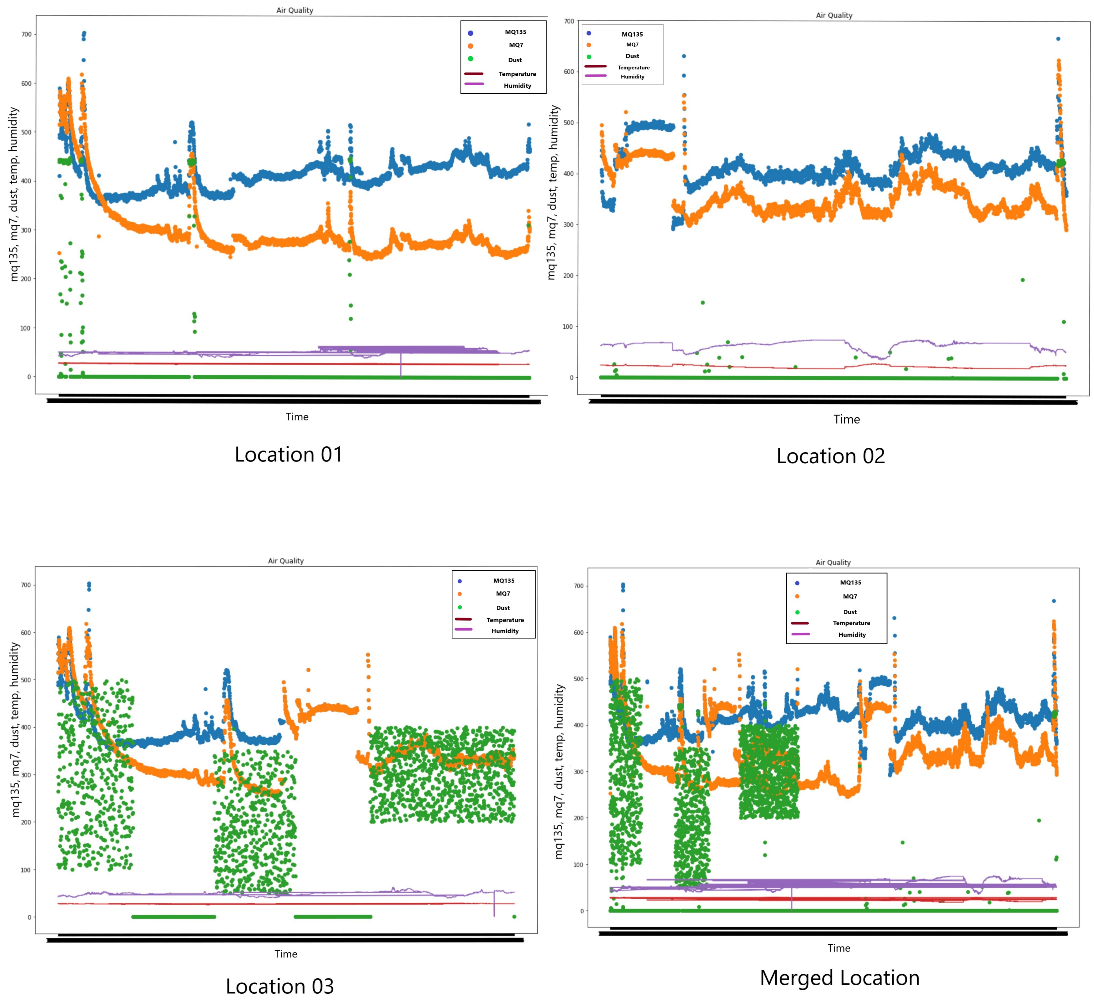

Figure 4.2: Scatter Chart (Value changes with respect to Time)

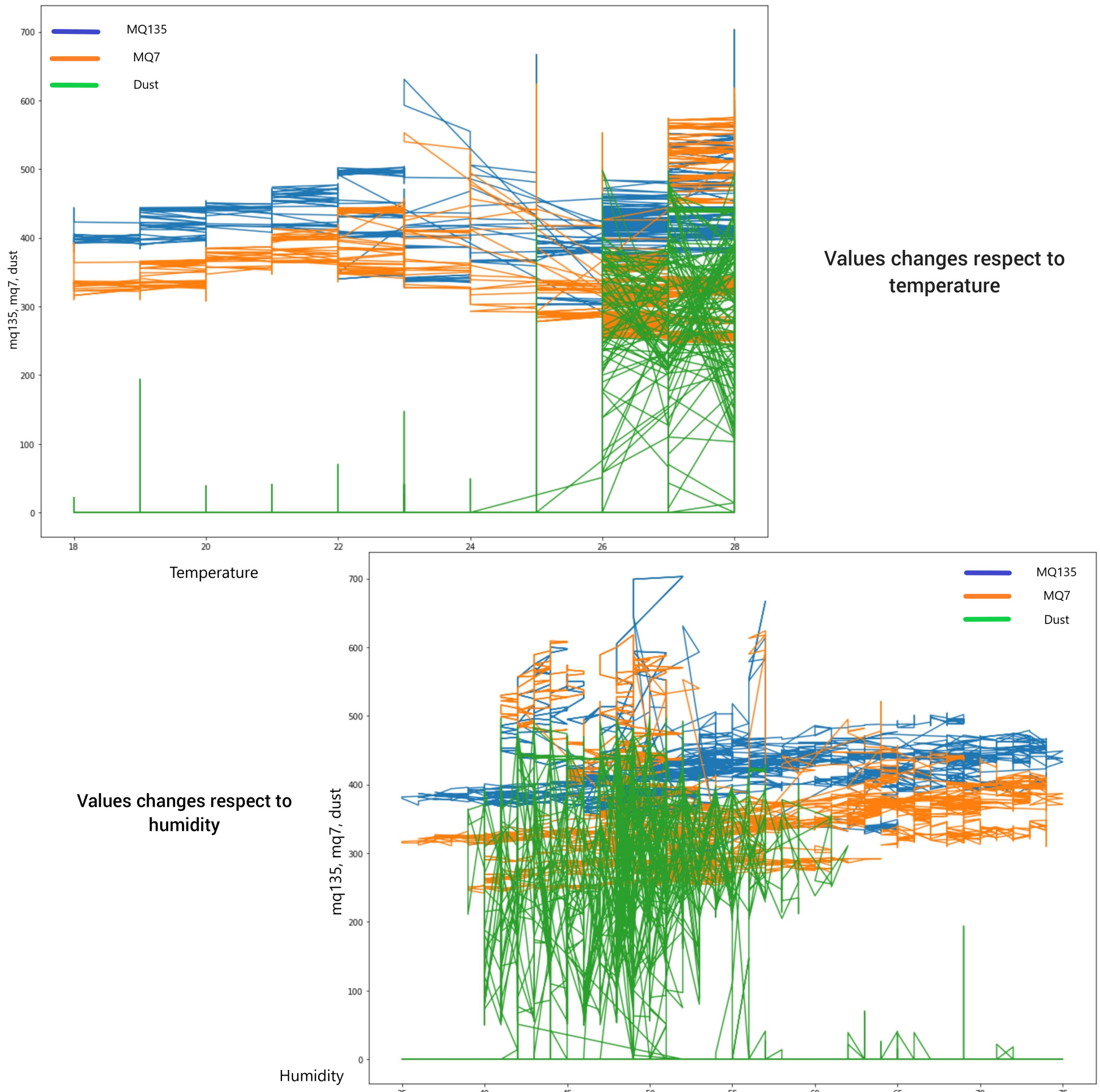

Figure 4.3: Values changes respect to Temperature and Humidity

4.5 Evaluation of Performance

Previously, we discussed about the training part where data has been split for model training. A comparison will be discussed between several algorithms. In table 4.2 shows all the algorithm's accuracy with different datasets based on five parameters. Three of them are the individual location data, and the other one is the merged dataset. In table 4.3 shows the algorithm's accuracy with merged dataset with three parameters. In table 4.4 shows the output result of two random test cases for four different datasets based on five parameters. In table 4.5 shows the output result of two random test cases for merged datasets based on three parameters. Here we have used some color indication to understand the best,

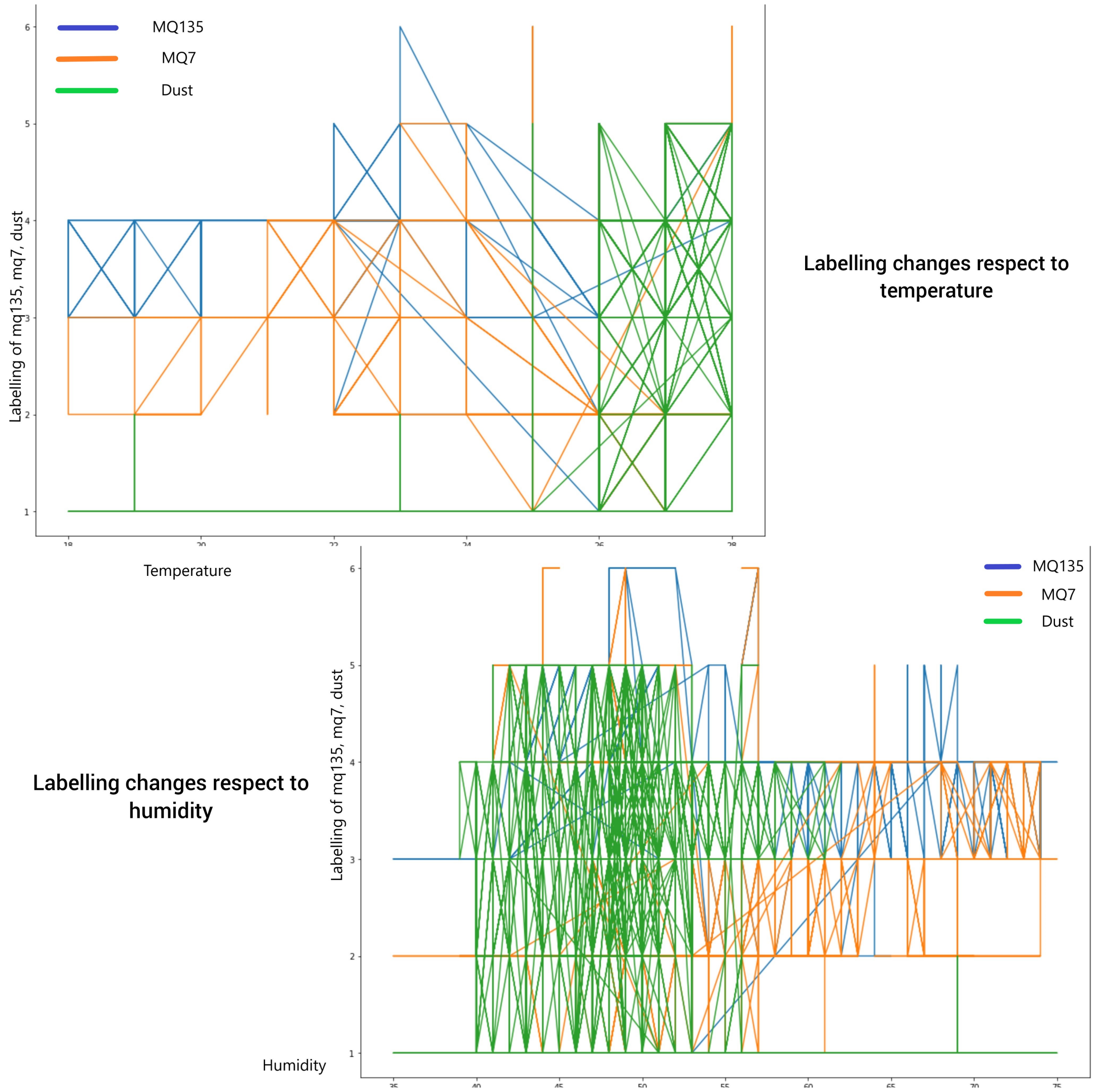

Figure 4.4: Labelling changes respect to Temperature and Humidity

average and worst match compared to the actual result. Here Green shows the accurate match with the actual result, Red Shows the worst match with the actual result, and Blue shows the partial match with the precise result. It is visible that a minority of the results are the worst match compared to the actual result. This is a clear comparison of the algorithms with different test cases.

4.6 Conclusion

Finally, it was discovered that an IoT-based air quality monitoring device would forecast air quality parameters with incredible precision. The established device produces brilliant results with the precision required by air quality standards.

Table 4.2: Accuracy Analysis based on five parameters

Location	SVM	Random Forest	Logistic Regression	Linear Regression	KNN	Decision Tree
1	0.90	0.99	0.96	0.89	0.99	0.98
2	0.94	0.99	0.86	0.62	0.98	0.99
3	0.69	0.98	0.85	0.85	0.96	0.98
Merged	0.88	0.99	0.877	0.83	0.98	0.99

Table 4.3: Accuracy Analysis based on three parameters

Location	SVM	Random Forest	Logistic Regression	Linear Regression	KNN	Decision Tree
Merged	0.848	0.997	0.86	0.849	0.986	0.99

Table 4.4: Test Case Comparison based on five parameters

Location	Actual Result	SVM	Random Forest	Logistic Regression	Linear Regression	KNN	Decision Tree
1	Very Unhealthy	Good	Very Unhealthy	Unhealthy	Unhealthy	Very Unhealthy	Very Unhealthy
1	Caution	Good	Unhealthy	Unhealthy	Moderate	Unhealthy	Good
2	Very Unhealthy	Moderate	Very Unhealthy	Very Unhealthy	Unhealthy	Very Unhealthy	Very Unhealthy
2	Caution	Caution	Caution	Unhealthy	Caution	Unhealthy	Caution
3	Very Unhealthy	Moderate	Unhealthy	Unhealthy	Very Unhealthy	Unhealthy	Unhealthy
3	Caution	Moderate	Caution	Moderate	Moderate	Caution	Caution
Merged	Very Unhealthy	Moderate	Very Unhealthy	Unhealthy	Very Unhealthy	Very Unhealthy	Very Unhealthy
Merged	Caution	Caution	Unhealthy	Moderate	Caution	Caution	Caution

Table 4.5: Test Case Comparison based on three parameters

Location	Actual Result	SVM	Random Forest	Logistic Regression	Linear Regression	KNN	Decision Tree
Merged	Very Unhealthy	Very Unhealthy	Very Unhealthy	Unhealthy	Very Unhealthy	Very Unhealthy	Very Unhealthy
Merged	Caution	Caution	Caution	Moderate	Moderate	Caution	Caution

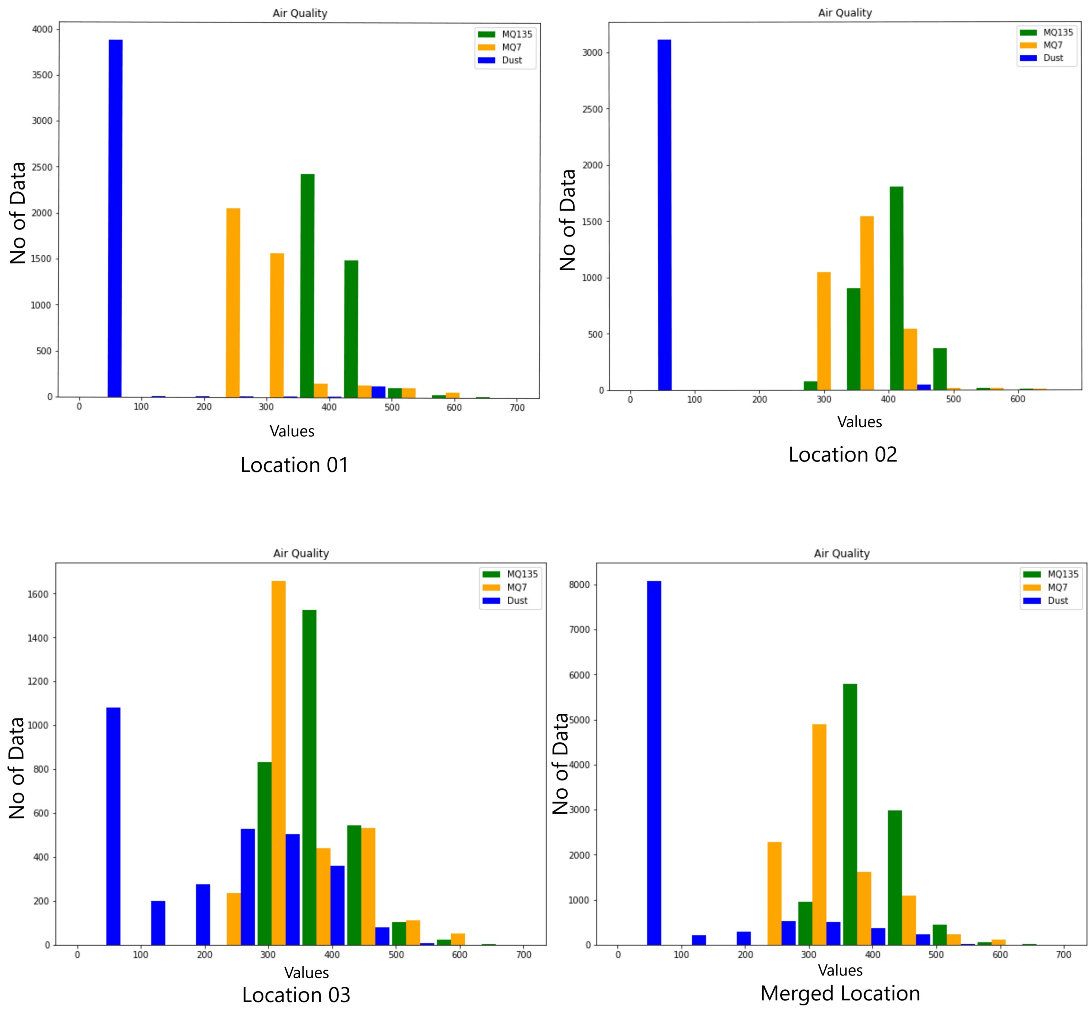

Figure 4.5: Histogram for four Datasets

The presented study was related to previous literature. This air quality monitoring device might be a low-cost monitoring system that could also be utilized in potential experiments. A visualization part gives a clear view of the changes of the different parameters. Accuracy comparison and the comparison of the test cases make the analysis more precise. The user can be benefited by knowing the current condition from the real-time monitoring system and from the analysis graphs as well. This work can also make a huge impact on the development of the smart city. Easily anyone can know the current situation with proper indication. People can easily know where to find the fresh air for good health and can decide the place they are currently living is suitable or not.

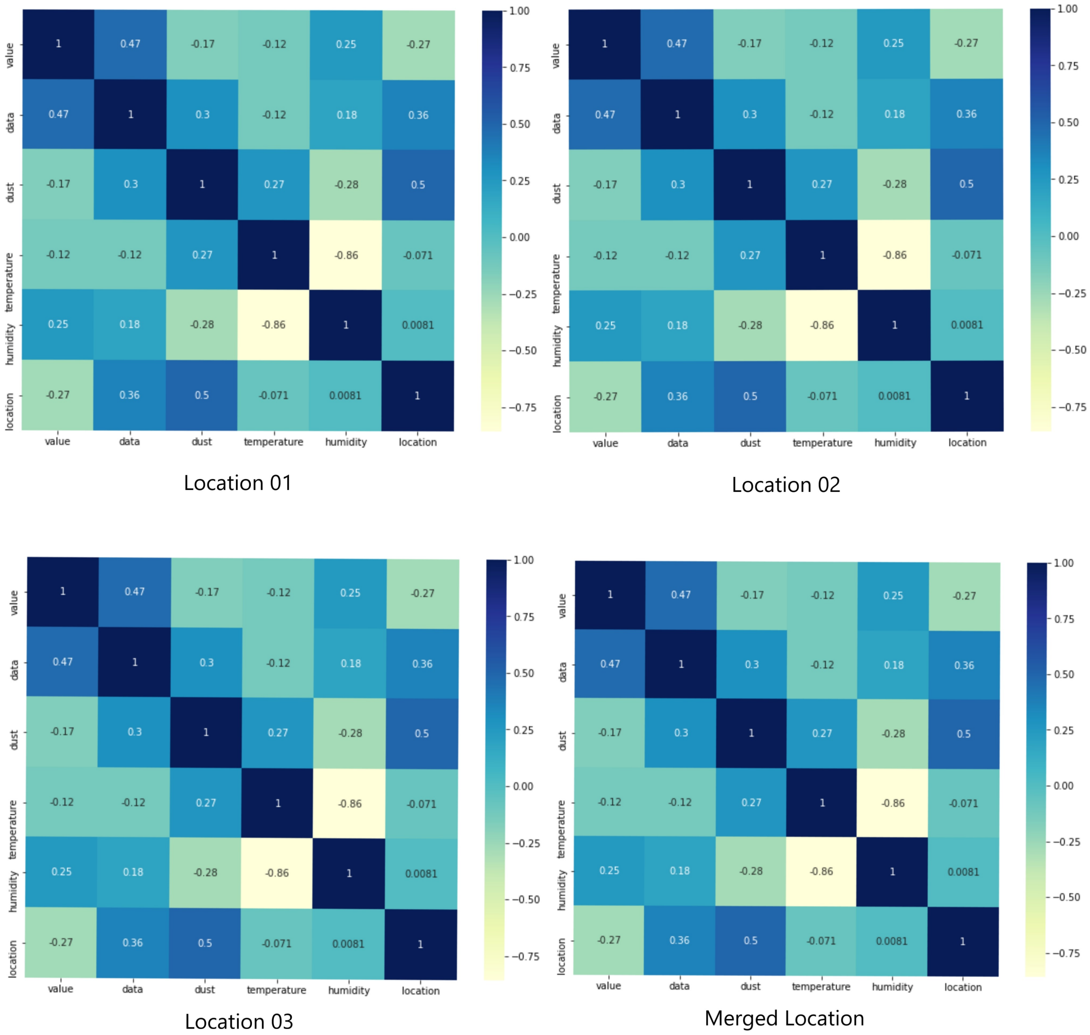

Figure 4.6: Correlations for four Datasets

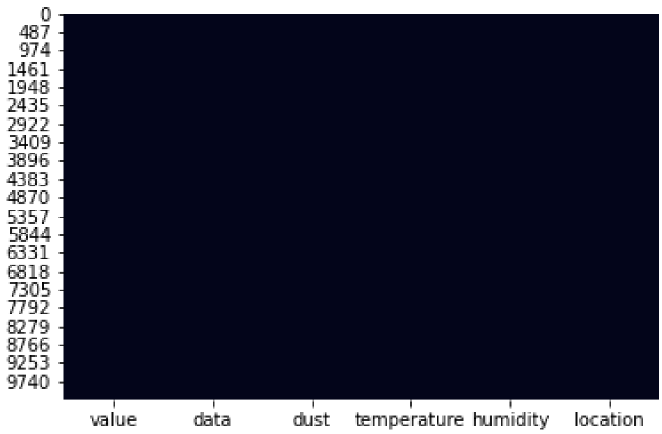

Figure 4.7: Heat Map

Chapter 5

Conclusion

5.1 Conclusion

Traditional air quality monitoring is not as efficient as this developed system. Air is an essential thing, fresh air to be precise. In the era of modernization, industrialization, the air has been contaminated a lot by people. For this, we need a regular monitoring system to know the exact condition of the environment as it is directly related to our health. This indication may make us a little conscious of the surroundings. We have built an IoT device connected to the internet to reach out to the server for this purpose. This real-time data with color indication make it more user-friendly, which is easily understandable. Visualization of real-time data is also a better part of this. Those who can't know the values can quickly understand the pollutants' real-life condition changes with different parameters. Another most exciting thing is the analysis part and its model visualization. Model training and prediction is the final touch of this project. For prediction, we have used several algorithms to make a =comparison of the accuracy among them. We have also tested for different test cases and check the result with the actual result where a minority number of the worst match was there. The majority of them have a better result, and by this, it terminates our works.

In summary, an IoT device has been built to collect data stored in a database. From this database, a real-time monitoring system has been built. Real-time visualization also helps to understand the condition precisely. From the dataset, a machine learning analysis has been done. Prediction analysis and its visualization has been done with test case analysis.

5.2 Future Work

The device can be updated in the future. We have worked on some parameters of the air. The parameter numbers can be added to the list, such as particular matter 0.1, particular matter 10 are which are very effective but costly. A message alert system can also be integrated with the project, informing the people about the pollution alert. If the pollution rate crosses the threshold level, it can suggest that people avoid the location or take some necessary steps. A mobile app based on real-time monitoring system can also be developed.

References

[1] *Air pollution in bangladesh killed 1.23 lakh in 2017 | daily star*, `https://www.thedailystar.net/frontpage/air-pollution-in-bangladesh-killed-1%2C23-lakh-2017-1724689`, (Accessed on 04/15/2021) (cit. on p. 1).

[2] J. A. Bernstein, N. Alexis, C. Barnes, I. L. Bernstein, A. Nel, D. Peden, D. Diaz-Sanchez, S. M. Tarlo and P. B. Williams, 'Health effects of air pollution,' *Journal of allergy and clinical immunology*, vol. 114, no. 5, pp. 1116–1123, 2004 (cit. on p. 1).

[3] *What is the iot? everything you need to know about the internet of things right now | zdnet*, `https://www.zdnet.com/article/what-is-the-internet-of-things-everything-you-need-to-know-about-the-iot-right-now/`, (Accessed on 04/15/2021) (cit. on p. 3).

[4] *What is iot (internet of things) and how does it work?* `https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT`, (Accessed on 04/15/2021) (cit. on p. 3).

[5] *What is internet of things- iot for beginners*, `https://internetofthingswiki.com/internet-of-things-definition/`, (Accessed on 04/15/2021) (cit. on p. 4).

[6] S. Zampolli, I. Elmi, F. Ahmed, M. Passini, G. Cardinali, S. Nicoletti and L. Dori, 'An electronic nose based on solid state sensor arrays for low-cost indoor air quality monitoring applications,' *Sensors and Actuators B: Chemical*, vol. 101, no. 1-2, pp. 39–46, 2004 (cit. on p. 4).

[7] S. Duangsuwan, A. Takarn and P. Jamjareegulgarn, 'A development on air pollution detection sensors based on nb-iot network for smart cities,' in *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*, IEEE, 2018, pp. 313–317 (cit. on pp. 7, 12).

[8] J.-H. Liu, Y.-F. Chen, T.-S. Lin, D.-W. Lai, T.-H. Wen, C.-H. Sun, J.-Y. Juang and J.-A. Jiang, 'Developed urban air quality monitoring system based on wireless sensor networks,' in *2011 Fifth International Conference on Sensing Technology*, IEEE, 2011, pp. 549–554 (cit. on p. 8).

[9] *What are indicators for good or bad indoor air quality (iaq)? - pure air control services inc.* `https://pureaircontrols.com/indicators-good-bad-indoor-air-quality-iaq/`, (Accessed on 04/15/2021) (cit. on p. 9).

[10] M. Ghoneim and S. M. Hamed, 'Towards a smart sustainable city: Air pollution detection and control using internet of things,' in *2019 5th International Conference on Optimization and Applications (ICOA)*, IEEE, 2019, pp. 1–6 (cit. on p. 12).

[11] S. Nagaraj and R. V. Biradar, 'Applications of wireless sensor networks in the real-time ambient air pollution monitoring and air quality in metropolitan cities—a survey,' in *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, IEEE, 2017, pp. 1393–1398 (cit. on p. 13).

[12] X. Zhao, S. Zuo, R. Ghannam, Q. H. Abbasi and H. Heidari, 'Design and implementation of portable sensory system for air pollution monitoring monitoring,' in *2018 IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia)*, IEEE, 2018, pp. 47–50 (cit. on p. 13).

[13] M. Muntaseer Billah Ibn Azkar, S. Chatani and K. Sudo, 'Simulation of urban and regional air pollution in bangladesh,' *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D7, 2012 (cit. on p. 13).

[14] M. R. Delavar, A. Gholami, G. R. Shiran, Y. Rashidi, G. R. Nakhaeizadeh, K. Fedra and S. Hatefi Afshar, 'A novel method for improving air pollution prediction based on machine learning approaches: A case study applied to the capital city of tehran,' *ISPRS International Journal of Geo-Information*, vol. 8, no. 2, p. 99, 2019 (cit. on p. 13).

[15] N. S. Desai and J. S. R. Alex, 'Iot based air pollution monitoring and predictor system on beagle bone black,' in *2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2)*, IEEE, 2017, pp. 367–370 (cit. on p. 13).

[16] S. R. Enigella and H. Shahnasser, 'Real time air quality monitoring,' in *2018 10th International Conference on Knowledge and Smart Technology (KST)*, IEEE, 2018, pp. 182–185 (cit. on p. 13).

[17] M. M. Ahmed, S. Banu and B. Paul, 'Real-time air quality monitoring system for bangladesh's perspective based on internet of things,' in *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, IEEE, 2017, pp. 1–5 (cit. on p. 13).