# A DEEP LEARNING MODEL FOR AIR QUALITY PREDICTION

*A Project Report*

*Submitted to the APJ Abdul Kalam Technological University*

*in partial fulfillment of requirements for the award of degree*

***Bachelor of Technology***

*in*

***Information Technology***

*by*


**ABHIMANYU A S(LTRV20IT069)**

**GATHA REGHUNATH(TRV20IT033)**

**SAYANDH K S(TRV20IT058)**

**MERLYN DEENA PHILIPOSE(TRV20IT062)**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**GOVERNMENT ENGINEERING COLLEGE, BARTON HILL, THIRUVANANTHAPURAM-695035**

**KERALA**


**May 2024**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**GOVERNMENT ENGINEERING COLLEGE, BARTON HILL, THIRUVANANTHAPURAM-695035**

**2023-24**

## CERTIFICATE

This is to certify that the report entitled **A DEEP LEARNING MODEL FOR AIR QUALITY PREDICTION** submitted by **ABHIMANYU A S** (LTRV20IT069), **GATHA REGHUNATH** (TRV20IT033), **SAYANDH K S** (TRV20IT058) & **MERLYN DEENA PHILIPOSE** (TRV20IT062) to the APJ Abdul Kalam Technological University in partial fulfillment of the B.Tech. degree in Information Technology is a bonafide record of the project work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

**Prof.Jassim Rafeek**
Assistant Professor
Dept.of IT
Government Engineering College,
Barton Hill, Trivandrum

**Prof. Simi Krishna K R**
Assistant Professor
Dept.of IT
Government Engineering College,
Barton Hill, Trivandrum

**Prof.Josna V R**
Assistant Professor
Dept.of IT
Government Engineering College,
Barton Hill, Trivandrum

**Prof.Ajini A**
Assistant Professor
Dept.of IT
Government Engineering College,
Barton Hill, Trivandrum

**Dr. Haripriya A.P**
Associate Professor and Head
Dept.of IT
Government Engineering College,
Barton Hill, Trivandrum

## DECLARATION

We hereby declare that the project report **A DEEP LEARNING MODEL FOR AIR QUALITY PREDICTION** , submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under supervision of **Prof.Jassim Rafeek**.

This submission represents our ideas in our own words and where ideas or words of others have been included, we have adequately and accurately cited and referenced the original sources.

We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

**ABHIMANYU A S**
**GATHA REGHUNATH**
**SAYANDH K S**
**MERLYN DEENA PHILIPOSE**

Trivandrum
03 MAY 2024

# Abstract

Air quality prediction plays a crucial role in understanding and mitigating the impacts of pollution on public health and the environment.In prediction context, deep learning techniques have been used for several forecasting problems in big data. In this project, we propose a novel approach for air quality prediction using a Bi-Directional Long Short-Term Memory (Bi-LSTM) algorithm. The data utilized in this study is collected through a hardware setup comprising three sensors: MQ135, MQ5, and Dust sensors, which measure key air pollutants.

The primary objective of this project is to develop a user interface that provides predictive insights into PM 2.5 levels, a critical indicator of air quality, based on the input prediction date. We aim to forecast PM 2.5 concentrations accurately, enabling individuals and authorities to make informed decisions regarding air quality management. Forecasting air quality conditions through this interface is essential for proactive measures to mitigate the adverse effects of air pollution on human health and the environment. By providing timely and accurate predictions, individuals can take preventive actions, such as adjusting outdoor activities or using protective equipment, to minimize exposure to harmful pollutants.

Through this project, we demonstrate the effectiveness of employing advanced machine learning techniques, specifically Bi-LSTM, in predicting air quality parameters. The user interface serves as a valuable tool for raising awareness about air pollution and empowering individuals to make informed choices for healthier living environments.

Overall, this project underscores the significance of air quality forecasting as a strategy for protecting public health and promoting environmental sustainability by giving peak PM 2.5 alert through social media group which also consist of health advisory.

# Acknowledgement

We take this opportunity to express my deepest sense of gratitude and sincere thanks to everyone who helped us to complete this work successfully.We express our sincere gratitude to **Dr. Shiny G**, Principal,Government Engineering College, Barton Hill, , for all her support and guidance.

We express our sincere thanks to **Dr. Haripriya A.P**, Head of Department, Information Technology, Government Engineering College, Barton Hill, for providing us with all the necessary facilities and support.

We like to express my sincere gratitude to **Prof. Simi Krishna K R**,**Prof.Josna V R** and **Prof.Ajini A**, Department of Information Technology, Government Engineering College, Barton Hill,Trivandrum for their support and co-operation.

We would like to place on record my sincere gratitude to our project guide **Prof.Jassim Rafeek**, Assistant Professor, Information Technology, Government Engineering College, Barton Hill, for the guidance and mentorship throughout this work.

Finally we thank our family, and friends who contributed to the successful fulfilment of this project work.

**ABHIMANYU A S**
**GATHA REGHUNATH**
**SAYANDH K S**
**MERLYN DEENA PHILIPOSE**

# Contents

# List of Figures

# Chapter 1

# Introduction

It is impossible to overestimate the importance of PM2.5 level monitoring in the context of future urban development. Due to its profound penetration into the bloodstream and lungs, PM2.5 presents serious health hazards to urban populations. The ability to successfully handle air quality concerns is improved by incorporating real-time PM2.5 data into decision-making processes. Creating customized alert systems that quickly alert the public and appropriate authorities when PM2.5 concentrations approach hazardous levels is a crucial component of this strategy. Cities can prevent immediate hazards and protect public health by identifying peak PM2.5 readings and forecasting when they will occur. These alarm systems also go beyond simple notice, offering vital health alerts and advice on preventive steps people can take to reduce exposure. Giving locals access to useful information encourages awareness-raising and proactive participation in air quality control. Essentially, one of the most important steps toward developing healthier and more sustainable living environments is incorporating PM2.5 monitoring and forecast into urban development policies. Cities can improve public health, build resilient communities, and proactively combat air pollution by utilizing digital technologies and data-driven solutions.

# Chapter 2

# Literature Review

The literature review explores various approaches to assess and predict air quality, focusing on various geographic locations. Researchers initially used biased networking and DAG formation in Kazakhstan, but noted its limitations in different regions. IoT-based vehicle data and LSTM help predict air pollution, with an emphasis on CO2 and NO. However, some studies overlook key pollutants such as NO, CO, sulfur dioxide, PM2.5, and PM10. Delhi's pollution is analyzed using K-means, identifying specific areas. Classification methods, including Naive Bayesian and random forest, demonstrate effectiveness in evaluating air quality. Hybrid machine learning methods improve the identification and accuracy of the model. Bengaluru's study employs ZeroR for carbon emissions assessment, relating impurities to air purity. New methods, such as multimodal categorization and neural network-based techniques, classify PM10 concentrations. Some studies present hybrid strategies, blending stochastic predictions for accuracy. A synthetic approach in a CH4 gas study considers various parameters, but has limitations in gas coverage. In Spain, SVM and logistic regression, incorporating inductive reasoning, forecast vehicle emissions and suggest preventive measures for air pollution.

## 2.1 A Deep Learning Model for Air Quality Prediction in Smart Cities

[1].In their work published in the IEEE International Journal on Big Data in 2022, İbrahim Kök, Mehmet Ulvi Şimşek, and Suat Özdemir[1] introduce a sophisticated deep learning model tailored for air quality prediction in smart cities. The study details the utilization of advanced neural network architectures to process intricate environmental data, enabling the accurate forecasting of air quality levels. Using deep learning techniques and integrating them with IoT sensor data, the proposed model showcases promising results in facilitating real-time monitoring and prediction of air quality. This integration not only enhances the precision of predictions but also holds potential for proactive measures in addressing air pollution challenges within urban environments.

## 2.2 Air Quality Prediction System Using ML and DL Techniques

[2].In their paper presented at the IEEE North Karnataka Subsection Flagship International Conference in 2020, S.V. Vasantha, K. Ramesh, and J. Vaishnavi[2] delve into the development of an innovative air quality prediction system that combines machine learning (ML) and deep learning (DL) techniques. The study underscores the prowess of DL models in capturing intricate patterns inherent in air quality data, exceeding the capabilities of conventional ML methods. Using DL algorithms, the proposed system demonstrates enhanced accuracy and reliability in forecasting air quality levels. This amalgamation of ML and DL methodologies represents a significant advancement in the field, promising more effective and precise air quality prediction capabilities that can contribute to better environmental management and public health outcomes.

## 2.3 Deepu B.P, Dr. Ravindra Rajput, Air Pollution Prediction Using Machine Learning, IRJET Volume 10, 2022

[3]. This study explores the application of machine learning techniques for air pollution prediction. Although the specific focus on deep learning is not mentioned, the research provides insight into the broader realm of machine learning algorithms utilized for air quality forecasting. The findings contribute to the understanding of machine learning's efficacy in predicting air pollutant levels, thereby laying the groundwork for further investigation into more advanced methodologies.

## 2.4 Marius, Andreea, Marin, Machine Learning Algorithms for Air Pollutants Forecasting, IEEE Journal, 2020

[4]. This research investigates various machine learning algorithms for forecasting air pollutants. Although the emphasis is not explicitly on deep learning, the study provides valuable insight into the application of machine learning in the prediction of air quality. By comparing different algorithms, the authors offer insights into their performance and effectiveness in forecasting air pollutant concentrations, contributing to the broader understanding of air quality prediction methodologies.

## 2.5 Prediction of Air Pollutants Using Supervised Machine Learning

[5]. In their research published in IEEE Paper Volume 9 in 2021, SriramKrishna Yarragunta, Mohammed Abdul Nabi, and Jeyanthi [5].investigate the use of supervised machine learning techniques to predict air pollutant concentrations. Although the study does not focus exclusively on deep learning, it provides a comprehensive exploration of various machine learning algorithms, including deep learning models, for air quality

forecasting. Through comparative analysis, the authors offer valuable insights into the performance of these algorithms, shedding light on their strengths and limitations in predicting air pollutant concentrations. This study contributes significantly to the advancement of the understanding of the role of deep learning in air quality forecasting applications, thus forming future research directions and practical implementations in environmental monitoring and management.

## 2.6 Shreyas Simu, Varsha Turkar, Rohit Martires, Air Pollution Prediction using Machine Learning, IEEE ACCESS, 2021

[6]. In this study published in IEEE ACCESS, the authors delve into air pollution prediction using machine learning techniques. Although deep learning is not the sole focus, the research sheds light on the application of machine learning in predicting air pollution levels. Using machine learning methodologies, the study demonstrates the potential for accurate prediction of air quality parameters, providing valuable insights into the field of air quality detection and prediction. These selected studies collectively contribute to the results

# Chapter 3

# Problem Definition

## 3.1 Problem Statement

To minimize or control air pollution, along with measurement, prediction also plays a major role. The first task is to perform quality measurement using the air quality index. Following this, the next task is to predict future air quality levels based on previous and present measurement data. The detection and prediction of air quality are crucial for ensuring public health and disease prevention. In addition, it is necessary for various aspects of a smart city, including urban planning, policy making, and traffic management.

## 3.2 Problem Description

The critical issue of air pollution in India is highlighted, highlighting its detrimental impact on human health and the environment. Primary and secondary pollutants that cause air pollution are identified, along with their adverse effects such as air quality degradation, acid rain, global warming, and smog production. The study advocates for the use of Deep Learning (DL) algorithms, particularly the Long-Short Term Memory (LSTM) algorithm, to predict air pollution.

The problem revolves around air pollution in urban areas and its adverse effects on public health. The focus is on monitoring and predicting air quality levels to safeguard the well-being of residents. Key issues addressed include the potential health hazards associated with poor air quality, especially for vulnerable groups such as children, the elderly, and individuals with respiratory conditions. The goal is to implement solutions using digital technologies, IoT sensors, data analytics, and communication networks to measure air quality, predict future levels, and take proactive measures to mitigate the impact of air pollution.

## 3.3 Scope and Objectives

The scope and objectives of the IoT-based prediction system are to address the increasing data produced by IoT-based applications and predict future values of air quality parameters, specifically PM 2.5. The system aims to leverage deep learning techniques, particularly Long Short Term Memory (LSTM) networks, to analyze IoT data and make accurate predictions. The primary focus is on addressing air pollution problems by using IoT data analysis to forecast air quality parameters. Additionally, the system aims to provide early warnings and precautionary measures for air pollution, particularly focusing on red alarm situations. Furthermore, it aims to demonstrate the effectiveness of the proposed LSTM-based prediction model and its potential for application in other prediction problems. Overall, the objectives include developing a robust deep learning model for IoT data analysis and prediction, with a specific focus on addressing air pollution challenges.

# Chapter 4

# Requirement Specification

## 4.1 Purpose

Air quality detection and prediction serve a multifaceted purpose in the context of healthcare. They actively monitor, analyze, and forecast air quality conditions with the primary objective of mitigating health impacts. By identifying and addressing potential risks associated with poor air quality, timely interventions can be implemented to protect residents. Furthermore, the integration of these technologies enables healthcare systems to allocate resources more efficiently, particularly by predicting the rises in respiratory issues related to air quality. The dissemination of real-time information to the public fosters awareness, encouraging people to take precautions during periods of compromised air quality. Moreover, the data-driven approach contributes to urban planning by unraveling patterns of air pollution, informing the development of policies and infrastructure aimed at enhancing overall air quality. In times of environmental emergencies, the predictive capabilities empower cities to respond rapidly, minimizing the impact of pollution spikes. In essence, the synergy of air quality management not only improves public health, but also optimizes resources and fosters sustainable urban development.

## 4.2 Software Requirement Specification

### 4.2.1 Functional Requirements

- Air quality Measurement:- The system must measure air quality data using IoT sensors connected to Arduino.
- Air quality Prediction:- The system must employ deep learning techniques (LSTM) for air quality prediction.
- Air Quality Alert:- The system automatically detects peak PM 2.5 value within the input range of date and provide health advisory along with an alert message through a WhatsApp group.

### 4.2.2 Non-Functional Requirements

**Performance Requirements**

- Precision, Recall, and F1-Score:The model's performance is evaluated using precision, recall, and F1-score metrics to measure its ability to predict air quality alarm colors accurately. The precision, recall, and F1-score values are calculated for each alarm color (red, yellow, green) to assess the model's performance.
- Error Criteria: The model's prediction accuracy is assessed using error criteria such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These metrics provide insights into the accuracy of the model's predictions and its ability to minimize errors in air quality forecasting.
- Prediction Accuracy: The prediction accuracy of the LSTM- based model is calculated . The accuracy of the model's predictions is a key performance requirement to evaluate its effectiveness in air quality prediction.

**Software System Attributes**

- Usability:- The system should have a user-friendly interface that allows city stakeholders to easily access and interpret the air quality predictions. The interface should provide visualizations and reports to facilitate decision-making.

- Scalability:- The proposed system should be scalable to handle the increasing volume of data generated by IoT devices . It should be able to accommodate the growing number of connected devices and data sources.

- Reliability:- The system should be reliable and available to provide continuous monitoring and prediction of air quality . It should have minimal downtime and be resilient to potential failure.

- Modularity : Emphasize a modular design to facilitate the ease of maintenance, upgrades, and the addition of new functionalities.

- Interoperability: Ensure that the system can seamlessly integrate with existing infrastructure and healthcare systems.

# Chapter 5

# Project Design

## 5.1 System Overview

The Air Quality Detection and Prediction System is a cutting-edge solution designed to address the health challenges arising from air pollution. Using advanced techniques, this system provides a comprehensive approach to monitoring, analyzing, predicting, and giving out air quality alerts and health warnings. The primary goal is to improve the health aspects by offering real-time insights and proactive measures to mitigate the impact of poor air quality on public health by considering PM 2.5 as the major air pollutant.

### 5.1.1 Key Components

- Air Quality Monitoring Infrastructure:-The system incorporates a network of strategically deployed sensors to monitor various air quality parameters, including pollutants such as particulate matter (MQ135 Sensor), nitrogen dioxide (MQ5 Sensor), Carbon Monoxide (Dust Sensor), and ozone (MQ5 Sensor).

- Data Collection Mechanisms:- Efficient mechanisms are implemented to collect data from sensors in real-time. Additionally, the system interfaces with external sources such as weather stations and traffic monitoring systems to enrich the dataset for more accurate predictions.

- Advanced deep learning models are integrated to analyze historical air quality data and identify patterns. These models continuously learn and adapt to changes, improving the accuracy of air quality predictions over time.Here we use LSTM Model for the time series prediction of PM 2.5 value.

- Prediction Accuracy Module:- The system defines stringent standards for prediction accuracy, ensuring that forecasts are reliable and actionable. Continuous evaluation and validation processes are in place to maintain the high quality of the prediction. The LSTM Model provides more accuracy compare to other models like SVR(Support Vector Regression).
- User Interface:-The User Interfacesconsist of a user friendly GUI,WhatsApp alert and voice broadcast system.

### 5.1.2 Benefits

- Proactive Healthcare Planning: The system empowers healthcare authorities with timely information to plan and implement proactive healthcare measures, reducing the impact of air pollution-related health issues by giving early warnings.
- Citizen Awareness and Engagement: Citizens receive real-time information on air quality through user-friendly interfaces and the WhatsApp group, fostering awareness and encouraging individuals to take preventive actions for their well-being.
- Policy Decision Support: Data generated by the system serves as valuable insights for policymakers, aiding in the formulation of effective environmental policies and regulations to improve overall air quality.
- Community health improvement:- By promptly addressing air quality concerns, the system contributes to the overall improvement of community health, reducing the burden on healthcare facilities, and improving the quality of life in smart cities.

## 5.2 Software Design Description

### 5.2.1 System Architecture

#### Hardware for data collection

The first step of this project is to detect the parameters necessary to predict the PM 2.5 value such as carbon monoxide (CO), nitrogen dioxide (N02), ozone (O3) and a sensor to measure the PM 2.5 value required to determine air quality. The sensors used for data collection are the MQ135 Sensor, the MQ5 Sensor and the Dust Sensor. These sensors are connected to the Arduino Module and thus the data are collected.

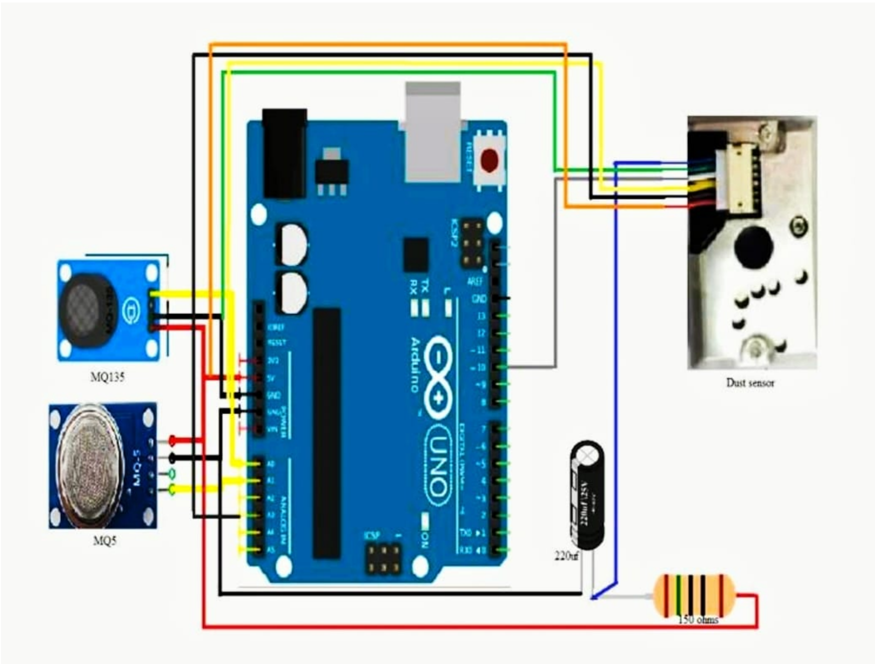


Figure 5.1: Hardware Connections

#### Basic Architecture of Proposed System

- **Data Collection**: Various air quality data are measured by sensors placed across the city of Trivandrum, such as PM2.5, NO2, CO, O3, and other pollutants. Each reading has a timestamp attached to it, making it possible to track how the quality of the air varies over time.

- **Data Storage**:The data collected is kept in a CSV file for ease of access and model training compatibility. To guarantee the quality and reliability of

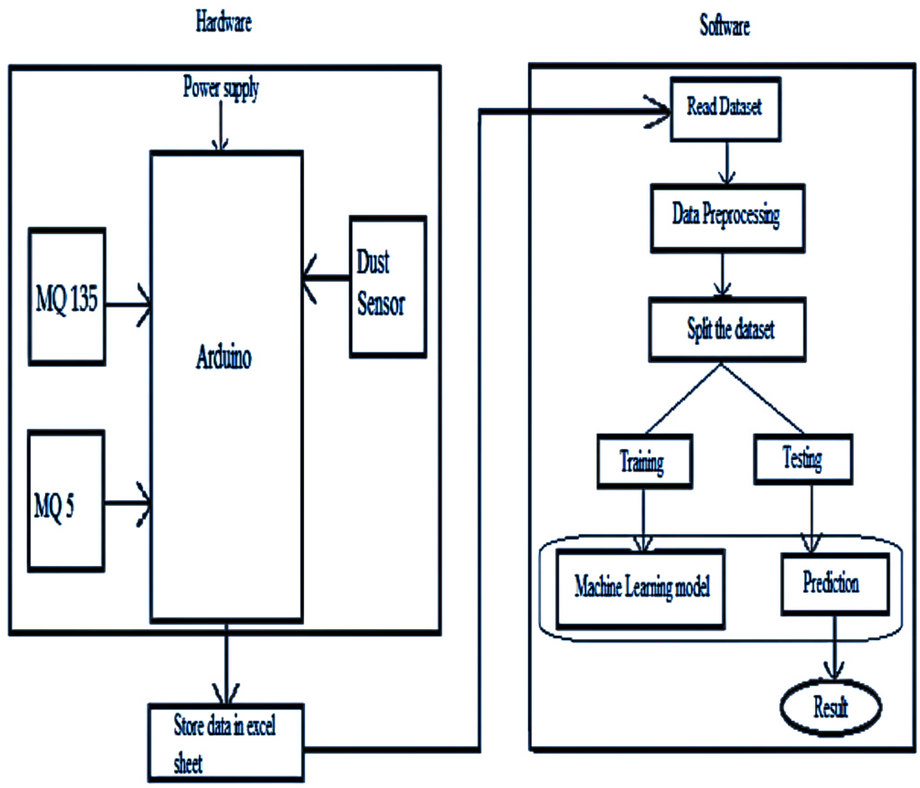


Figure 5.2: Basic Architecture of Proposed System

datasets, data preparation processes like cleaning, handling missing values, and identification of outliers are carried out.

- **Model Training:-**Using the gathered dataset, the selected model was trained. The model was trained with input characteristics and target values that corresponded to air quality parameters. This allowed the model to identify patterns and relationships in the data quicker. The model's capacity to predict air quality metrics was eventually enhanced by allowing it to repeatedly be exposed to historical data and related target values, which allowed it to identify patterns and correlations.

- **Model Testing and Validation:-** To forecast air quality for the next few days, the model was evaluated using testing data. The predictions of the model were visualized by plotting graphs of PM2.5. Furthermore, the mean absolute error metric was employed to assess the model's performance and offer an understanding of the typical size of errors that exist between the model's predictions and the actual measurements of air quality. The usefulness of the model for air quality forecasting tasks was increased by this thorough testing and validation process, which guaranteed the accuracy and dependability of the model, and the GUI is created for the respective model.

### 5.2.2 Deep Learning Model

- LSTM:- LSTM, or Long Short-Term Memory, is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem that occurs in traditional RNNs. LSTMs are particularly effective for processing and predicting sequences of data due to their ability to capture long-term dependencies and remember information over extended time intervals. This makes them well-suited for tasks involving time series data, such as predicting air quality based on historical sensor readings.
  - Organize the data into sequences, where each sequence represents a temporal sequence of air quality measurements. This is crucial for LSTM to learn patterns over time.
  - Extract relevant features from the data set that could influence air quality. These could include meteorological data, time of day, and historical air quality readings.
  - Train the LSTM model using the prepared dataset. The model learns to capture patterns and dependencies in the time-series data.

**Algorithm for implementation of single step LSTM**

To forecast PM2.5 air quality using a single-step LSTM model, first collect historical air quality data, including PM2.5 measurements, and relevant characteristics such as temperature, humidity, and wind speed. Preprocess the data by cleaning it, removing duplicates, handling missing values, and normalizing to ensure consistency across features. Split the data into training and testing sets. Next, define the sequence length, determine the number of time steps to consider for prediction, and generate data sequences with past observations of PM2.5 and other characteristics.

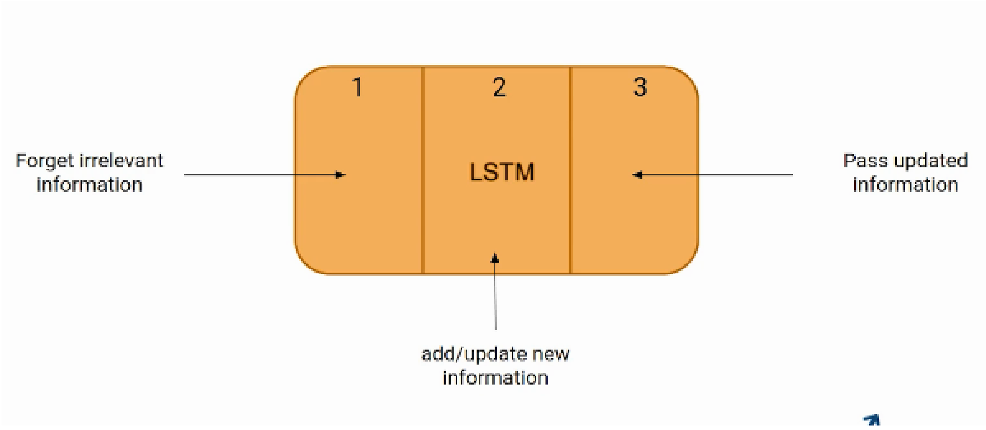


Figure 5.3: LSTM Block

The LSTM block has an input gate,a forget gate and an output gate.Build the LSTM architecture, including input layers for historical data sequences, LSTM layers to capture temporal dependencies, and a fully connected layer (dense) for prediction. Compile the model with appropriate loss functions (e.g., Mean Squared Error), optimizers (e.g., Adam), and evaluation metrics (e.g., Mean Absolute Error).

Then, train the model on the training data, adjusting the weights through backpropagation to minimize loss over multiple epochs. Evaluate the model's performance on a validation set, calculating metrics like the mean squared error and mean absolute error. Assess the model's performance on unseen data (testing set) using the same metrics. Visualize predicted PM2.5 values against actual values to assess accuracy and identify patterns. Optionally, fine-tune the model based on insights gained from validation and testing phases.

Once satisfied with performance, deploy the model for real-time or future PM2.5 forecasting, monitoring its accuracy over time, and updating as needed. Keep in mind that the specifics may vary depending on the characteristics of the dataset, the model architecture, and the domain expertise, requiring adjustments and optimizations through experimentation and refinement.

### 5.2.3 Tools and Methodologies used

- **Jupyter Notebook:-** Jupyter Notebook is utilized as a versatile platform for conducting data analysis and experimenting with code, facilitating an interactive environment for explo- ration and documentation.
- **Visual Studio Code:-**The main code editor used for project is Visual Studio Code (VS Code). Numerous features are available, including debugging tools, code completion, and syntax highlighting. Writing and managing Python code is a perfect fit for VS Code because of its adaptability and support for extensions.
- **Python:-** Python is the primary programming language for tasks ranging from data analysis and machine learning to web development, offering extensive libraries and frameworks for diverse applications.
- **Flask:-** Flask is chosen as a lightweight web framework to build web applications, offering simplicity and flexibility in creating scalable and customizable web solutions.
- **HTML:-** HTML is used as the standard markup language to structure web pages, providing a foundation for organizing and presenting content on the Web.
- **CSS:-** CSS is utilized as a stylesheet language to define the visual presentation and layout of web pages, enhancing the aesthetic appeal and user experience of web applications.
- **Pandas:-**The Pandas module is used mainly in this project to handle time series data related to the prediction of PM2.5. It has functions for manipulating data, handling dates and times, integrating NumPy arrays, analyzing and visualizing data, and loading and storing data. Pandas is useful for activities like date parsing, working with date/time columns, calculating statistics, combining data, making visualizations, and loading PM2.5 historical data for model testing and training.
- **Numpy:-**The Numpy module is utilized extensively for efficient data handling and mathematical operations.It makes it easier to store and interpret big datasets, such as PM2.5 measurements and model predictions, by offering fundamental data structures like arrays and matrices. Numpy is a library of mathematical functions that can be used for a wide range of activities, including array

manipulation, element-wise operations, linear algebra, statistical computations, and random number generation. These tasks include data normalisation, model training, and evaluation. Additionally, numpy's fluid integration with pandas DataFrames makes it easier for various project components to exchange data.

- **Matplotlib:-**An essential tool for data visualisation is the matplotlib package, which makes it possible to create informative plots and graphs that show model predictions and trends in PM2.5 data. Several plot types, such as line plots, are created using matplotlib's extensive features to display historical PM2.5 data, model forecasts, and peak value occurrences. Plot features like labels, titles, axes, and legends can be customised to make the visualisation process very user-friendly and instructive. Furthermore, plots created with matplotlib can be saved as picture files and then integrated into the Flask web application to be shown. Matplotlib's smooth integration into the project architecture improves data understanding, makes insights easier to communicate, and gives stakeholders an eye-catching visual depiction of PM2.5 patterns and predictions.

- **Seaborn:-**The seaborn module offers a high-level interface for making visually appealing and educational statistical visualisations, which is a supplement to matplotlib. By utilising seaborn's features, the project improves data visualisation by producing intelligent and aesthetically pleasing plots—like line plots—with less code complexity. The visualisation process is made more efficient and aesthetically pleasant by making use of Seaborn's pre-installed themes, colour schemes, and sophisticated statistical tools. Furthermore, Seaborn provides smooth interface with pandas DataFrames, enabling the direct display of model predictions and trends in PM2.5 data that are stored in tabular formats. Seaborn's user-friendly API and extensive visualisation toolkit improve the resulting plots' interpretability, giving stakeholders a better understanding of the PM2.5 data and model performance.

- **TensorFlow:-**The tensorflow module is essential because it makes machine learning models especially deep learning models for time series prediction implementable and deployable. The research forecasts PM2.5 pollution levels by using tensorflow to leverage pre-trained deep learning models, like LSTM

(Long Short-Term Memory) networks. These models are loaded and run in order to produce forecasts derived from past PM2.5 data. Tensorflow also offers tools for inference, model training, and data preprocessing, which simplify the development process and guarantee effective model deployment. Tensorflow's scalable and resilient architecture enables the project to use state-of-the-art deep learning methods for precise and dependable PM2.5 predictions, improving environmental monitoring and public health care.

- **Keras:-**The keras module is instrumental in this project, facilitating the development and deployment of deep learning models, particularly LSTM (Long Short-Term Memory) networks, for time series forecasting of PM2.5 pollution levels. Leveraging keras, the project benefits from a high-level neural networks API that allows seamless construction, training, and evaluation of complex neural network architectures. The keras API provides an intuitive and user-friendly interface, enabling rapid experimentation with different model architectures, optimization algorithms, and hyperparameters. Furthermore, keras integrates seamlessly with tensorflow, allowing the project to harness the power of tensorflow's computational graph optimization and execution capabilities while leveraging the simplicity and flexibility of keras's API. Through its ease of use and tight integration with tensorflow, keras empowers the project to efficiently develop and deploy accurate predictive models for PM2.5 forecasting, contributing to improved environmental monitoring and public health outcomes.
- **gtts:-**The gtts (Google Text-to-Speech) module plays a crucial role in this project by converting text data, specifically peak PM2.5 value and corresponding date information, into speech. This functionality enables the project to provide auditory feedback to users, enhancing the accessibility and usability of the application. Leveraging the gtts module, the project can dynamically generate spoken notifications or alerts, conveying important information in a clear and understandable manner. By converting text to speech, the application ensures that users, including those with visual impairments or those in situations where reading text may not be practical, can receive essential updates regarding peak PM2.5 levels. Overall, the integration of the gtts module enhances the project's

user experience and accessibility, making it more inclusive and effective in delivering timely information related to air quality.

- **Playsound:-**The playsound module is essential to this project because it allows audio files to be played back. Specifically, it plays back the speech produced by the Google Text-to-Speech (gtts) module along with the highest PM2.5 value and related date. The project improves the application's usability and accessibility by giving users audio input through the usage of the playsound module. With the help of this technology, the project may play the audio alerts that are generated dynamically, guaranteeing that users are informed of peak PM2.5 levels in a clear and audible manner. The playsound module's integration improves the user experience of the project and increases its efficacy in providing timely air quality updates in a way that is accessible to all users, irrespective of their visual impairments or other constraints.

- **Base64:-** The base64 module is used to make it easier to transform image data into a format that can be embedded in HTML content. It is specifically used to encode images produced by the matplotlib package into strings that are base64 encoded. By doing this, the process of producing visualizations within the web application is much simpler and eliminates the need to save images as separate files. These image representations may then be seamlessly inserted into HTML layouts. With this integration, the workflow is streamlined and the project can effectively deliver graphical data representations with textual content. This improves the user experience overall by offering detailed insights through an aesthetically pleasing interface.

- **Pyautogui:-**This project makes use of the pyautogui module to automate interactions with graphical user interfaces (GUIs) by mimicking keystrokes, clicks, and mouse movements. It is specifically used to browse the UI of the WhatsApp Desktop program and click on features like chat threads, attachment icons, and search bars. Pyautogui is also used to mimic typing actions for entering text, like a contact's name or the name of a group, and for indicating file paths when attaching documents. The project can increase efficiency and decrease manual involvement by automating the process of sending audio files to WhatsApp contacts or groups by utilizing pyautogui's features.

- **Pygetwindow:-**The project makes use of a Graphical User Interface (GUI) to give users an interactive and user-friendly platform for data access and visualization. Users can produce predictions and view matching visualizations by inputting parameters, such as date ranges, through the GUI, which improves their ability to effectively comprehend and evaluate the data. In addition, the incorporation of a speech broadcast system gives the application a special and useful function that lets users hear alerts regarding important discoveries and patterns in the data. The program ensures accessibility for users who prefer aural cues by converting pertinent information into speech through the use of text-to-speech (TTS) technology provided by the gtts module. Furthermore, the project's reach and usability are increased by the integration with WhatsApp broadcast capabilities, which enables users to easily share data.

### 5.2.4 User Interface Design

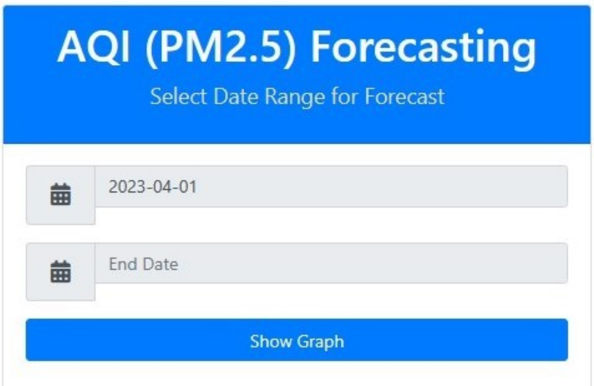


Figure 5.4: User Interface

This project makes use of a Graphical User Interface (GUI) to give users an interactive and user-friendly platform for data access and visualization. Users can produce predictions and observe matching visualizations by inputting parameters, such as date ranges, through the GUI, which improves their ability to properly comprehend and analyze data. Furthermore, the addition of a voice broadcast system to the application gives it a special and useful function that lets users hear alerts regarding important discoveries and patterns in the data. Through the use of the gtts module's text-to-speech (TTS) technology, the program speaks pertinent information, making it accessible to users who are more comfortable with auditory cues. Further, the project's reach and usefulness are increased by the integration with WhatsApp broadcast capabilities, which enables users to easily share data insights with selected groups or contacts. This feature makes it easier for people to collaborate and communicate with one another, which helps them effectively spread critical information. Overall, the project's accessibility, usefulness, and collaboration skills are improved by the voice broadcast system, user-friendly GUI, and WhatsApp connection, making it a flexible tool for communication and data analysis.

# Chapter 6

# Results

The project's evaluation focuses on assessing the impact of its key features: the Graphical User Interface (GUI), voice broadcast system, and WhatsApp integration. User feedback, usability testing, and metrics such as user engagement are used to evaluate the effectiveness, usability, and utility of these features. The evaluation aims to identify strengths, areas for improvement, and opportunities for enhancement to inform iterative refinements and future iterations of the project.

## 6.1 Outputs

**Graphical User Interface(GUI):-** A user-friendly interface consisting of two options to provide a starting date and an ending date and a predict button. The result will be shown as a graph when the predict button is pressed. The graph depicts the relationship between the date (x-axis) and PM2.5 concentration (y-axis), where PM2.5 represents particulate matter with a diameter of 2.5 micrometers or smaller, a significant contributor to air pollution. The start date and end date must be provided to get the prediction. Currently, the Air Quality Index (AQI) serves as a metric for gauging the levels of pollutants in the atmosphere, categorizing air quality from good to hazardous. Generally, PM2.5 concentration and AQI exhibit an inverse correlation. As PM2.5 concentration increases, AQI values tend to escalate, signaling deteriorating air quality. In contrast, decreasing PM2.5 concentrations correspond to decreased AQI values, indicative of improved air quality conditions. This inverse relationship is rooted in the substantial role that PM2.5 plays in air pollution.

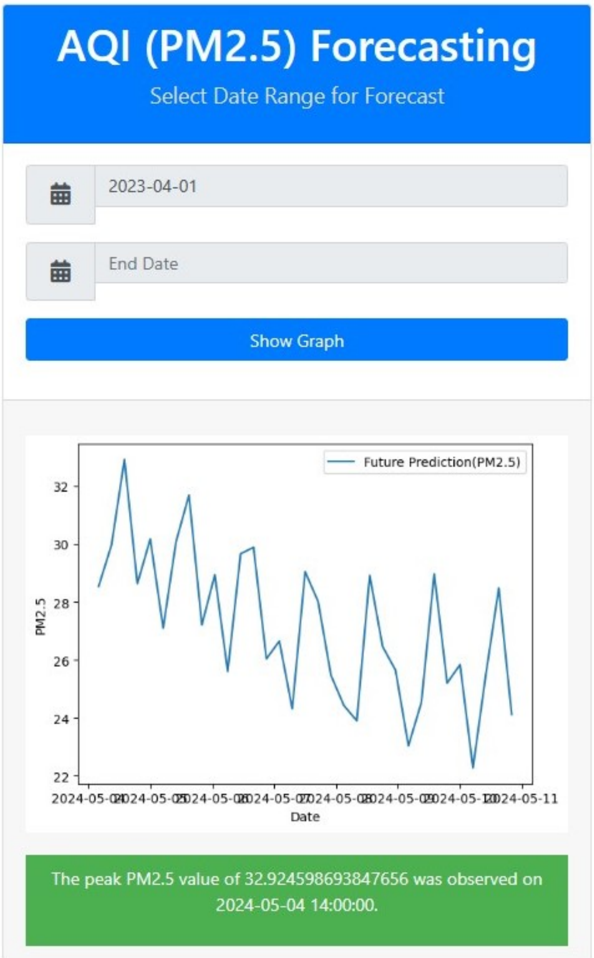


Figure 6.1: Graphical User Interface

Elevated PM2.5 concentrations amplify pollutant levels, with adverse implications for human health and the environment. Consequently, elevated levels of PM2.5 correlate with elevated AQI values, reflecting compromised air quality. In contrast, lower PM2.5 concentrations align with reduced AQI values, signaling favorable air quality conditions. Understanding this interplay facilitates the evaluation and management of air quality, vital to safeguarding public health and environmental well-being.

**Voice Broadcast System:-**The voice broadcast system integrated into the project provides users with auditory notifications about key insights and trends in the data. Utilizing text-to-speech (TTS) technology from the gtts module, relevant information is converted into speech, ensuring accessibility for users who prefer auditory cues. This feature adds a unique and convenient dimension to the application, enhancing user experience and providing an alternative mode of interaction for accessing important data insights.

**WhatsApp Alert:-**

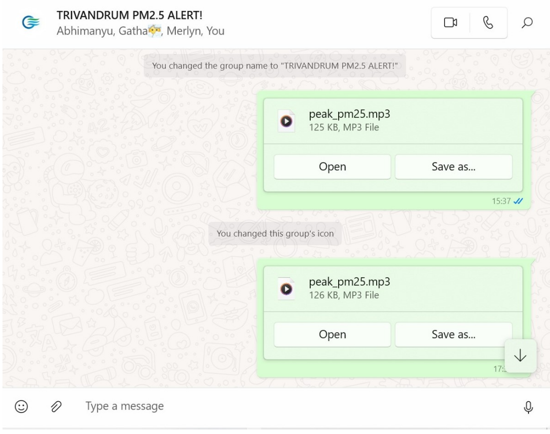


Figure 6.2: WhatsApp Alert

Sends a voice note to users by sending alerts on maximum PM 2.5 values consisting of a health advisory.These alerts, triggered by significant changes in air quality levels, ensure that users stay informed about environmental risks. Using WhatsApp as a communication channel, the system delivers timely notifications, empowering users to take proactive measures to protect their health.

# Chapter 7

# Analysis

## 7.1 Evaluation Matrices

To assess the efficiency and performance of the machine learning models that are being used quantitatively, we use various metrics called Evaluation Metrics. These metrics act as a benchmark for the model's performance on a specific task. Few among the popular metrics used to evaluate machine learning models are accuracy, precision, recall, and F1 score.

### 7.1.1 Accuracy

Measures the amount of correct predictions the model makes and focuses on the overall performance of the model. It is the ratio between the amount of correct predictions and the total number of predictions.

### 7.1.2 Precision

Measures the ratio of the number of positive instances correctly predicted out of the total positive predictions.

### 7.1.3 Recall

Measures the ratio of the number of correctly predicted positive instances out of the total positive instances present.

### 7.1.4 F1-Score

It evaluates by combining the scores of precision and recall. Rather than focusing on overall performance, it focuses on class-wise performance of the model. It is the harmonic mean of Precision and Recall.

### 7.1.5 Evaluation Results

| Accuracy | | Precision | | Recall | | F1-Score | |
|---|---|---|---|---|---|---|---|
| LSTM | SVR | LSTM | SVR | LSTM | SVR | LSTM | SVR |
| 0.80 | 0.75 | 0.84 | 0.80 | 0.82 | 0.80 | 0.81 | 0.78 |

Table 7.1: Comparison of results of the evaluation metrices of LSTM nand SVR

The prediction accuracy of the LSTM and SVR based model were calculated as 80% and 75%respectively.The best Precision84%, Recall82% and F1-Score%81 results were achieved for the LSTM model.Therefore, this scores have shown that LSTM based model outputs has a better performance than the SVR model.The overall results inidcate that LSTM based prediction model takes its advantages of memorizing of long historical data and achieve higher prediction accuracy.

# Chapter 8

# Conclusion

In conclusion, the implementation of an air quality prediction system emerges as a crucial initiative with far-reaching implications for public health and urban well-being. By serving as a vigilant guardian of air quality levels, the system empowers residents, especially those in vulnerable groups, with timely warnings and the ability to adopt precautionary measures. Beyond its immediate health benefits, the system plays an essential role in shaping urban planning, policy making, traffic management, environmental monitoring, and emergency response.

The holistic impact of this project extends well beyond preventive health measures, contributing significantly to the enhancement of overall quality of life in the city. As we enter an era where environmental concerns are of utmost importance, the air quality prediction system stands as a beacon of progress, safeguarding not only the present but also the future well being of our communities.

The paramount importance of the project lies in its ability to provide accurate and timely information, particularly to vulnerable populations. The early warnings generated by the Air Quality Prediction System empower individuals to make proactive choices, fostering a culture of health-conscious decision-making within the community. The ripple effects of this extend beyond individual well-being to encompass economic productivity, healthcare efficiency, and societal resilience.

# References


[1] ˙Ibrahim K¨OK,Mehmet Ulvi S¸˙IMS¸EK,Suat ¨OZDEM˙IR, *A deep learning model for air quality prediction in smart cities*, IEEE International Journal on BigData, 2022

[2] S.V.Vasantha, K.Ramesh, J.Vaishnavi,*Air Quality Prediction System using ML and DL Techniques*, IEEE North Karnataka Subsection Flagship International Conference,2020

[3] Deepu B.P, Dr. Ravindra Rajput,*Air Pollution Prediction Using Machine Learning*, IRJET volume 10 2022

[4] Marius, Andreea, Marin,*Machine Learning algorithms for air pollutants forecasting*, IEEE Journal 2020

[5] SriramKrishna Yarragunta, Mohammed Abdul Nabi, Jeyanthi.P,*Prediction of Air Pollutants Using Supervised Machine Learning*, IEEE paper volume 9 2021

[6] Shreyas Simu,Varsha Turkar, Rohit Martires,*Air Pollution Prediction using Machine Learning*, IEEE ACCESS 2021

turnitin

**Similarity Report ID:** oid:8268:58714175

PAPER NAME

**Deep_Learning_AQ_Report.pdf**

| WORD COUNT | CHARACTER COUNT |
| --- | --- |
| **6428 Words** | **37236 Characters** |
| PAGE COUNT | FILE SIZE |
| **34 Pages** | **829.0KB** |
| SUBMISSION DATE | REPORT DATE |
| **May 5, 2024 4:02 PM GMT+5:30** | **May 5, 2024 4:02 PM GMT+5:30** |

**19% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.

- 8% Internet database
- 8% Publications database
- Crossref database
- Crossref Posted Content database
- 16% Submitted Works database

Summary