A

Mini Project Report

on

AQI Tracker Using Random Forest Regression

Submitted in partial fulfillment of the requirements for the

degree

Third Year Engineering – Computer Science Engineering (Data Science)

by

Sharayu Mahajan 22107051

Kalpana Mohanty 22107059

Rishi Mane 22107063

Avadhoot Virkar 22107064

Under the guidance of

Ms. Dipali Gat

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING (DATA SCIENCE)

A.P. SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W)-400615

UNIVERSITY OF MUMBAI

Academic year: 2024-25

CERTIFICATE

This to certify that the Mini Project report on AQI Tracker using Random Forest Regression has been submitted by Sharayu Mahajan(22107051), Kalpana Mohanty(22107059), Rishi Mane(22107063) and Avadhoot Virkar(22107064) who are bonafide students of A. P. Shah Institute of Technology, Thane as a partial fulfillment of the requirement for the degree in **Computer Science Engineering (Data Science)**, during the academic year **2024-2025** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Ms. Dipali Gat

Guide

Ms. Anagha Aher

HOD, CSE (Data Science)

Dr. Uttam D. Kolekar

Principal

External Examiner:

1.

Internal Examiner:

1.

Place: A. P. Shah Institute of Technology, Thane

Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Ms. Dipali Gat**. Expressing gratitude towards our HoD, **Ms. Anagha Aher**, and the Department of Computer Science Engineering (Data Science) for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our project coordinator **Ms. Sarala Mary** who gave us his/her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

TABLE OF CONTENTS

ABSTRACT

In response to the growing need for accurate and user-friendly air quality tracking, Air Quality Index (AQI) was developed as an AI-powered platform utilizing Random Forest Regression to predict air quality index (AQI) levels. Designed for both individuals and organizations, the platform provides real-time air pollution insights, helping users make informed decisions about outdoor activities and health precautions.

AQI leverages a robust database to collect, store, and analyze air quality data from multiple sources, including environmental sensors, weather reports, and historical pollution trends. By employing machine learning algorithms, particularly Random Forest Regression, the system predicts AQI levels based on key environmental factors such as temperature, humidity, wind speed, and pollutant concentration. This predictive model enhances the accuracy of air quality forecasting, improving user awareness and proactive planning.

The platform features an intuitive user interface, allowing users to check local AQI levels, receive personalized alerts, and access historical data trends. Additional functionalities include interactive maps, health recommendations based on air quality conditions, and customizable notifications for high pollution events. Secure cloud-based storage ensures seamless data management, while API integration allows businesses and researchers to utilize the platform for advanced environmental analysis.

Future enhancements for AQI include real-time traffic pollution monitoring, integration with wearable health devices, and AI-driven suggestions for reducing exposure to poor air quality. By combining advanced machine learning techniques with user-friendly design, AQI aims to enhance air quality awareness and promote healthier living environments worldwide.

Chapter 1

Introduction

This section introduces AQI, an AI-powered air quality tracking system designed to provide real-time and accurate Air Quality Index (AQI) predictions. Utilizing Random Forest Regression, AQI processes live environmental data from APIs to help individuals and communities make informed decisions about outdoor activities and health precautions. The platform was developed to meet the increasing demand for precise air quality monitoring systems that enhance public awareness through automation, smart forecasting, and alert-based monitoring.

The project addresses key challenges in air pollution tracking, including data accuracy, real-time forecasting, and personalized air quality alerts. To improve prediction accuracy, AQI integrates machine learning techniques, analyzing factors such as temperature, humidity, wind speed, and pollutant concentrations. By automating AQI predictions and providing visual insights through interactive dashboards, the platform enhances user awareness and enables proactive pollution management.

A standout feature of AQI is its prediction engine, which utilizes historical trends and real-time sensor data to forecast AQI levels effectively. The user-centric dashboard, built using libraries like Matplotlib, ensures an intuitive experience, allowing users to track air quality, receive health recommendations, and get real-time alerts when pollution levels exceed safety thresholds. As the platform evolves, future enhancements will include real-time traffic pollution monitoring, integration with wearable health devices, and AI-driven pollution mitigation strategies, positioning AQI as an advanced environmental monitoring tool for public health and safety.

A key highlight of AQI is its intelligent prediction engine, which utilizes a blend of live sensor data and historical AQI trends to forecast future pollution levels with high reliability. This forecasting capability transforms AQI into a preventive tool, allowing users to take proactive measures such as avoiding high-exposure areas or modifying travel routes.

In essence, AQI is not only a technological solution—it's a step toward environmental responsibility, sustainable urban planning, and improved public health awareness. By automating AQI predictions, delivering insightful visualizations, and enabling timely alerts, the system empowers individuals and communities to make informed, health-conscious decisions. It encourages proactive responses to pollution, supports long-term environmental strategies, and fosters a culture of awareness and accountability in the face of growing ecological challenges.

1.1 Purpose

The purpose of this AQI tracker is to serve as a comprehensive, intelligent platform for monitoring and analyzing air quality in real time. It is designed to provide users with a seamless and intuitive experience, allowing them to access accurate AQI data for their current location or any other region across the globe. Whether individuals are planning outdoor activities, commuting, or managing health concerns related to pollution exposure, the platform empowers them with the information needed to make safe and informed decisions.

By integrating environmental and meteorological data from Kaggle's datasets, the system can handle large volumes of information related to pollutant concentrations, temperature, humidity, wind patterns, and more. These datasets are vital in training the machine learning model and ensuring reliable forecasts. Additionally, the inclusion of user preferences helps tailor the experience, enabling features such as location-based alerts and personalized recommendations.

The AQI tracker not only emphasizes data accuracy and real-time accessibility but also focuses on performance, scalability, and user engagement. The platform is built to accommodate a growing number of users and data points without compromising speed or reliability. Through the combination of intelligent algorithms, scalable data handling, and a user-friendly interface, the AQI tracker enhances public awareness and helps users better understand the air quality in their surroundings—ultimately contributing to healthier and more informed communities.

1.2 Problem Statement

The primary purpose of the AQI tracker is to provide users with a holistic and intelligent platform for monitoring, tracking, and understanding air quality across different regions. Designed to offer a seamless and interactive user experience, the system allows individuals to access real-time air quality data based on their current location or explore conditions in other parts of the world. This empowers users to make better-informed decisions regarding outdoor activities, travel plans, and health precautions—especially in areas prone to high pollution levels.

To ensure reliable and insightful forecasting, the platform utilizes datasets sourced from Kaggle, which include extensive information on historical air quality, weather patterns, and atmospheric pollutants. These large-scale datasets are essential for training the underlying machine learning models and for analyzing key environmental factors such as temperature, humidity, wind speed, and particulate matter levels. The use of such high-quality data ensures a solid foundation for generating accurate AQI predictions.

Furthermore, the system is designed with scalability and performance in mind. It can handle increasing volumes of data and user interactions without compromising efficiency, thanks to optimized data processing and visualization techniques. The integration of advanced technologies and machine learning not only enhances the reliability of the platform but also adds a layer of personalization through features like customized alerts and health advisories.

The AQI tracker aims to raise public awareness, encourage health-conscious behavior, and support smarter environmental decisions. With its user-friendly design, it provides vital air quality information for both individuals and environmental stakeholders.

1.3 Purpose

The AQI Tracker is an AI-powered system that predicts real-time air quality in Indian cities using live API data and machine learning. It features weather forecasting, pollution level classification, and alert notifications to help users stay informed and safe.

- **Real-Time AQI Prediction:**
  Develop an AI-powered system to predict the Air Quality Index (AQI) for various Indian cities using real-time API data. The system aims to provide accurate, timely air quality updates, helping users make informed decisions. It also promotes public awareness about pollution and its health impacts.

- **Weather Forecasting:**
  Integrate weather prediction features to provide forecasts alongside AQI data, offering a more complete environmental overview. This helps users plan better and stay safe by receiving alerts about weather conditions that could affect air quality.

- **AQI Classification & Alerts:**
  Categorize AQI levels (Good, Moderate, Unhealthy, etc.) to help users understand air pollution risks. Implement an alert system to notify users when air quality surpasses safety thresholds.

- **Interactive Dashboard for Data Visualization:**
  Use data visualization tools like Matplotlib to present AQI trends and weather conditions in an interactive format. Allow users to analyze and compare historical and real-time data effectively.

- **Machine Learning for Improved Accuracy:**
  Train and deploy machine learning models like Random Forest Regression to improve the accuracy AQI predictions. Models analyze environmental factors to deliver reliable, real-time air quality.

1.4 Scope

The AQI Tracker system aims to provide real-time air quality monitoring and forecasting using AI-driven models. The project will focus on collecting and analyzing AQI data for major Indian cities, offering accurate predictions to enhance public awareness and safety.

The system will include a user-friendly interface displaying real-time AQI levels, weather forecasts, and historical trends. It will classify air quality into categories (Good, Moderate, Unhealthy, etc.) and send alerts when pollution levels exceed safety thresholds. Additionally, it will utilize machine learning models like Random Forest Regressor to improve prediction accuracy.

The AQI Tracker will cater to multiple user groups, including the general public, environmental researchers, and government agencies. Initially, the focus will be on providing accurate air quality predictions and visualization tools, with future potential for expanding to additional regions, integrating more advanced forecasting techniques, and enhancing mobile accessibility for wider user engagement.

Chapter 2

Literature Review

A literature survey involves reviewing existing research, academic papers, books, and scholarly articles to gain insights into the current state of knowledge on a topic. In the context of Air Quality Index (AQI) prediction, various studies have explored methodologies, challenges, and advancements in machine learning and AI-powered forecasting systems. This survey highlights key findings, limitations, and gaps in AQI-related research.

Existing Studies on AQI Prediction and Monitoring:

A notable study from 2023 titled "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis" investigates the effectiveness of various machine learning algorithms in forecasting AQI. The research compares models such as Random Forest, Support Vector Machines (SVM), and Neural Networks, concluding that ensemble models like Random Forest generally outperform others in terms of accuracy and robustness. Despite their performance, the study acknowledges several persistent challenges including data preprocessing difficulties, limited real-time data integration, and issues related to the generalizability of the models across different regions and conditions. To overcome these limitations, the study suggests focusing on better feature selection techniques and enhancing real-time forecasting capabilities to further improve predictive performance and adaptability of the models. [1]

In addition to prediction models, researchers have also delved into the public health implications of AQI. A 2020 study titled "A Study and Analysis of Air Quality Index and Related Health Impact on Public Health" analyzes the correlation between deteriorating air quality and the rising incidence of respiratory and cardiovascular ailments. The study underscores the urgency for implementing real-time AQI monitoring systems capable of issuing immediate alerts to the public, especially in urban areas with high pollution levels. However, one of the major limitations identified in the research is its reliance on historical AQI data, without incorporating meteorological variables such as wind speed, humidity, and temperature that significantly influence AQI values. As a way forward, the study recommends the integration of real-time weather data with AI-driven models to enable more accurate predictions and the development of personalized health risk assessments for sensitive populations such as children, the elderly, and individuals with pre-existing conditions. [2]

Earlier foundational work from 2015, titled "Air Quality Index – A Comparative Study for Assessing the Status of Air Quality", provides a comprehensive analysis of AQI trends across various cities. It identifies the primary sources of pollution and examines their seasonal variations in detail. The study critically assesses traditional AQI computation methods, highlighting their limitations in adapting to rapidly changing pollution dynamics.To address these limitations, the authors advocate for the implementation of advanced AI-based forecasting techniques, such as time-series analysis, which are more capable of capturing and predicting fluctuations in air quality. Additionally, the research emphasizes a significant gap in the current AQI monitoring systems—the lack of integration with IoT-enabled sensors. These sensors have the potential to support real-time data collection, thereby increasing the accuracy and responsiveness of AQI monitoring frameworks.Bridging this technological divide is deemed essential not only for improving predictive accuracy but also for enhancing public awareness and informing environmental policy-making efforts. [3]

Chapter 3

Proposed System

The proposed system aims to design and implement an AI-powered framework for real-time Air Quality Index (AQI) prediction and monitoring. Based on insights from research and analysis, this system enhances accessibility, accuracy, and public awareness of air quality data.

3.1 Features and Functionality

Real-Time AQI Prediction & Monitoring:

- Utilizes APIs to fetch live AQI data from various Indian cities.
- Employs machine learning models like Random Forest Regressor for accurate forecasting.
- Provides users with real-time pollution levels, helping them take necessary precautions.

AQI Classification & Alert System:

- Classifies AQI levels into categories such as Good, Moderate, Unhealthy, and Hazardous.
- Sends automatic alerts when air quality surpasses safety thresholds.
- Helps users, policymakers, and environmental agencies take timely action.

Interactive Data Visualization:

- Features an interactive dashboard using Matplotlib and other visualization tools.
- Displays real-time and historical AQI trends, enabling users to analyze patterns.
- Offers graphical representation of pollution hotspots for better decision-making.

Weather Forecasting Integration:

- Incorporates meteorological data to provide accurate weather forecasts.
- Analyzes how weather conditions impact air quality over time.
- Enhances safety by alerting users to conditions that may worsen air pollution.

User-Friendly Dashboard & Accessibility:

- Provides an intuitive interface for users to track air quality easily.
- Enables sorting and filtering of AQI data based on location, severity, and trends.
- Ensures accessibility on multiple devices for seamless monitoring.

Continuous Model Improvement & Adaptability:

- Regularly updates machine learning models for improved accuracy.
- Integrates new data sources and adapts to changing environmental conditions.
- Ensures reliability through continuous evaluation and model optimization.

This system will significantly improve air quality awareness and support informed decision-making for both individuals and authorities.

Chapter 4

Requirements Analysis

Requirements analysis is the process of identifying, documenting, and analyzing the needs and expectations of stakeholders to define the system's scope, features, and constraints. It ensures the AQI Tracker is designed to meet user needs, function effectively, and operate reliably in real-world conditions.

- **Dataset:** The system will rely on real-time and historical datasets from trusted environmental agencies and weather APIs. These datasets will be essential for training AI models to predict AQI trends, personalize air quality recommendations, and optimize forecast accuracy based on regional pollution patterns.

  Your project uses real-time API data and datasets from Kaggle, which contain historical data on air quality, weather patterns, and pollutant levels. These datasets are crucial for training machine learning models like Random Forest Regressor to analyze environmental factors such as temperature, humidity, wind speed, and pollutant concentrations. This enables accurate AQI forecasting and personalization of health alerts based on regional pollution trends.

- **User Interface:** A user-friendly web and mobile interface will display real-time AQI levels, weather conditions, and forecasts. It will allow users to check air quality trends, receive alerts for unsafe conditions, and access recommendations for outdoor activities based on pollution levels.

  The system features a user-centric and interactive dashboard built using Matplotlib and other data visualization tools. This interface displays real-time AQI levels, weather conditions, historical trends, and forecast visualizations, allowing users to check conditions, receive safety alerts, and access health recommendations. The goal is to deliver a seamless and informative experience across web and mobile platforms.

- **Accuracy**: The system will provide reliable AQI and weather predictions based on sensor data, satellite imagery, and historical trends. The goal is to achieve an accuracy rate of at least 90% in air quality forecasts, improving over time with AI-driven models.

  Machine learning models such as Random Forest Regressor are used to enhance the prediction accuracy of AQI levels. By analyzing both historical trends and live sensor data, the system strives to maintain a high accuracy rate (targeting 90% or more), ensuring users get reliable and timely air quality forecasts for better decision-making.

- **Scalability:** The platform should be scalable to accommodate an increasing number of users and growing environmental datasets. As more cities and monitoring stations are integrated, the system should maintain seamless performance and efficient data processing.
  The system is designed to be robust, secure, and scalable, capable of processing increasing volumes of data and supporting more users as additional cities and monitoring stations are integrated. Efficient data handling, optimized model performance, and modular architecture contribute to scalable deployment across regions.

- **Search Interface:** The system should include an intuitive search interface where users can input their location, date, and specific AQI parameters (e.g., PM2.5, CO, NO2). Users should also be able to filter results by city, pollution levels, or health risk categories.
  While not explicitly detailed earlier, your project includes location-based search capabilities, allowing users to view AQI for their current city or explore other areas. This can be extended into a more advanced search interface, where users can filter by pollution types (PM2.5, CO, NO2) or health risk levels, and check AQI by date or forecast.

- **Forecasting Model:** The system should integrate AI-driven forecasting models to predict AQI levels and weather conditions for upcoming days. This will help users plan activities based on expected air quality and weather trends.
  A key feature of your platform is the AI-driven prediction engine, which combines real-time sensor input with historical trends to forecast AQI levels for the coming days. Integration with weather forecasting further improves contextual relevance, helping users plan activities based on predicted pollution and weather conditions.

Chapter 5

Project Design

Project design involves outlining and organizing the framework, elements, and features of a project to meet defined goals. This process translates the needs and objectives identified in the early stages, like requirement analysis, into a comprehensive plan or guide for execution.

5.1 Use Case Diagram

It is a visual representation that models the interactions between users (or other systems) and a system, describing its functionality and behavior from the user's perspective.

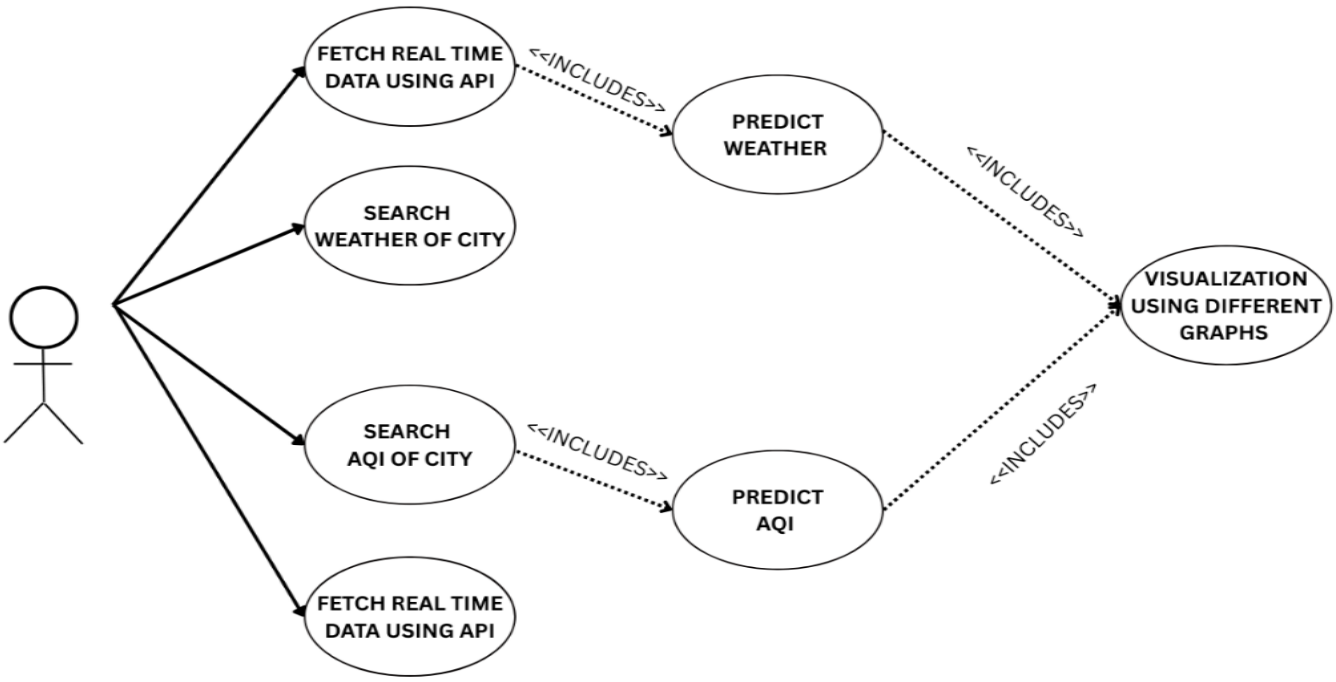


Figure 5.1: Use Case Diagram

In Figure 5.1, the central character is the **User**, who interacts with the AQI Tracker and Weather Forecast System to access real-time air quality and weather updates. The diagram highlights how users input their location and receive information such as AQI levels, weather forecasts, and safety alerts to support informed decision-making.

This flowchart breaks down the process into parallel paths based on the user's choice:

1. **User Selection**:

   Users start by selecting either Weather Prediction or AQI Prediction based on their need.

2. **Real-Time Data Fetching**:
   - The system fetches real-time data through APIs:
     - Weather data (e.g., temperature, rainfall).
     - Air Quality Index (AQI) data (e.g., pollutant levels).

3. **Prediction Pathways**:
   - **Weather Prediction**:
     - Historical data is used to forecast weather for the next 12 months using Random Forest Regression.
   - **AQI Prediction**:
     - Pollutant data is processed to predict AQI levels.
     - The AQI is categorized (e.g., Good, Moderate, Poor).

4. **Data Visualization**:
   - The results are visualized using various graph types (bar charts, line graphs, scatter plots) to improve data interpretability.

5. **Output Generation**:
   - Users receive clear, user-friendly forecast results to guide decision-making.

This flowchart illustrates a smart, user-driven system that integrates real-time API data with historical analysis, offering accurate and visually interpretable predictions for both weather and air quality.

5.2 DFD (Data Flow Diagram)

A Data Flow Diagram (DFD) visually represents how data moves within the AQI Tracker and Weather Forecast System, illustrating the interaction between external entities, processes, and data storage.

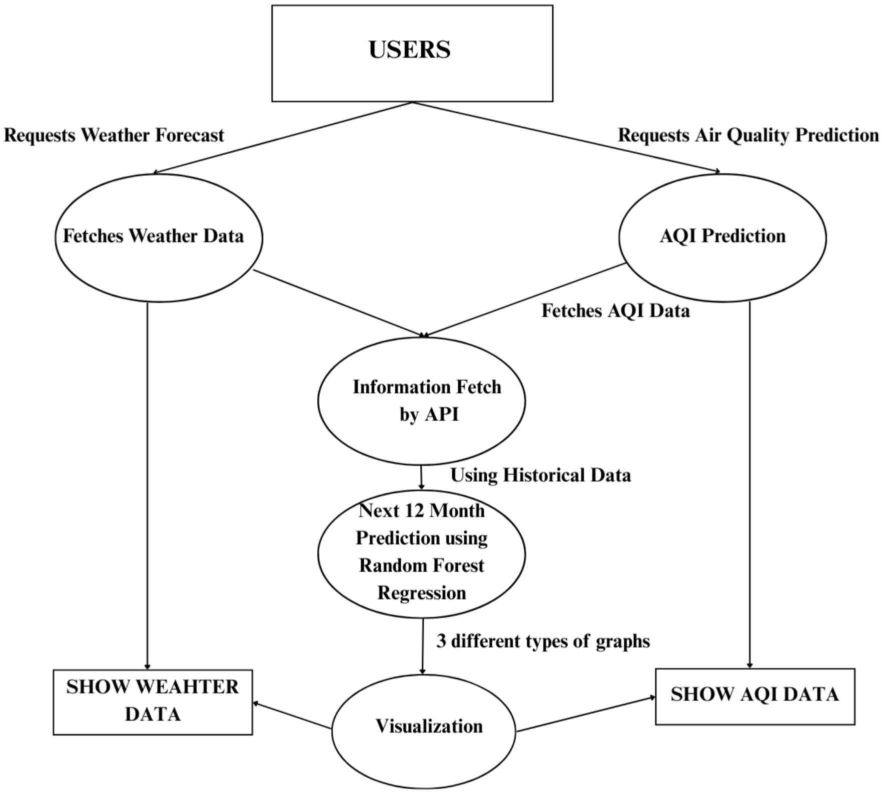


Figure 5.2: Data Flow Diagram

At the same time refering Figure 5.2, the system links to a centralized database, securely storing historical AQI data, weather forecasts, and user preferences. This database enables personalized recommendations, trend analysis, and automated alerts to help users make informed decisions about air quality and weather conditions.

This version simplifies the flow and emphasizes the shared backend for both prediction types:

Users: Initiate requests for weather or AQI predictions.

Data Fetching: APIs provide real-time weather and AQI data.

Prediction Engine: Combines historical and live data for 12-month forecasting.

Visualization: Generates visual representations of forecasts.

Centralized Database: Stores historical weather and AQI data.

The system efficiently combines live and historical data to deliver accurate 12-month forecasts with clear visual outputs.

5.3 System Architecture

The system architecture for the AQI Tracer and Weather Forecast platform defines the conceptual design and structural blueprint that governs the interaction between its various components and subcomponents.

Figure 5.3: System Architecture

The application follows a modular architecture to ensure scalability, maintainability, and flexibility, with a strong backend framework, this flowchart illustrates the AQI prediction process using machine learning, particularly the Random Forest algorithm. Here's a brief description:

1. **Load Historical Data (Trained Model)**: The system starts by loading previously collected and trained AQI and weather datasets.

2. **Preprocess Data**: This step involves cleaning and formatting the historical data to ensure consistency and quality.

3. **Fetch Live Data (API)**: Real-time weather and pollution data is collected from an external API to enhance predictions.

4. **Train Historical Data of Weather & AQI**: Both live and historical data are used together to build a robust training set for the model.

5. **Predict Future Trends (Using Random Forest)**: The Random Forest algorithm is used to forecast upcoming AQI trends based on the combined dataset.

6. **Analyze and Visualize**: Finally, the predicted results are presented through bar and line charts for better interpretation and decision-making.

This process enables accurate, real-time AQI forecasting and supports public awareness through clear visual insights.

5.4 Implementation

This section outlines the process of turning the project design into a working system. It includes the tools, technologies, and steps used to build and integrate the components, ensuring the final product meets the intended objectives.

This section outlines the detailed workflow of the AQI Tracer and Weather Forecast:

1. Home Page

Explore real-time environmental data through our sleek and intuitive homepage. Choose between checking the weather forecast or monitoring the Air Quality Index (AQI) — both just a click away. With a high-tech design featuring a stunning Earth-from-space background and easy-to-navigate, color-coded buttons (yellow for weather, blue for AQI), your experience is both informative and immersive.

Figure 5.4.1: Home Page

The homepage of the "Air Quality & Weather Tracker" platform, as seen in Figure 5.4.1, features a modern and visually engaging design. It presents two main options—checking the weather forecast and viewing AQI levels—each displayed with a distinct image and color-coded icon (yellow for weather, blue for AQI) for easy navigation. Below these options, the page highlights three major Indian cities—Pune, Kolkata, and Visakhapatnam—using artistic illustrations that reflect local landmarks. The entire interface is set against a striking image of Earth from space, creating a futuristic and immersive experience. Overall, the layout is clean, intuitive, and user-focused.

2. Weather Prediction

This dashboard offers a dynamic and user-friendly interface for weather forecasting, allowing users to access essential climate data for their selected location. It displays real-time weather conditions, including temperature, humidity, wind speed, and atmospheric status such as clear sky or smoke. Alongside current updates, the platform provides future weather forecasts in a structured format, enabling users to anticipate changes and plan accordingly. A visually engaging graph helps users track temperature and humidity trends over the coming months, making weather analysis both intuitive and informative.

Figure 5.4.2: Weather Prediction

On refering Figure 5.4.2, this shows that the weather forecasting feature has been implemented using a combination of real-time API integration and predictive modeling. Current weather data is fetched from a reliable weather API, ensuring accurate and timely updates. Future predictions are generated and presented in a clear tabular layout, showcasing monthly trends in temperature and humidity. A responsive graph is integrated to visualize these changes over time, enhancing user comprehension. The entire interface is designed with a space-themed background of Earth, providing a futuristic look while maintaining clarity and ease of navigation for users.

3. Air Quality Index Page

The platform introduces users to an advanced air quality monitoring interface, designed with both functionality and visual appeal in mind. Set against a stunning background image of Earth from space, the display creates an immersive and futuristic ambiance. The centerpiece is a sleek, semi-transparent card that provides real-time data for Thane, IN, prominently featuring the Air Quality Index (AQI) reading. This combination of informative content and modern design ensures that users receive critical environmental insights in a visually engaging manner.

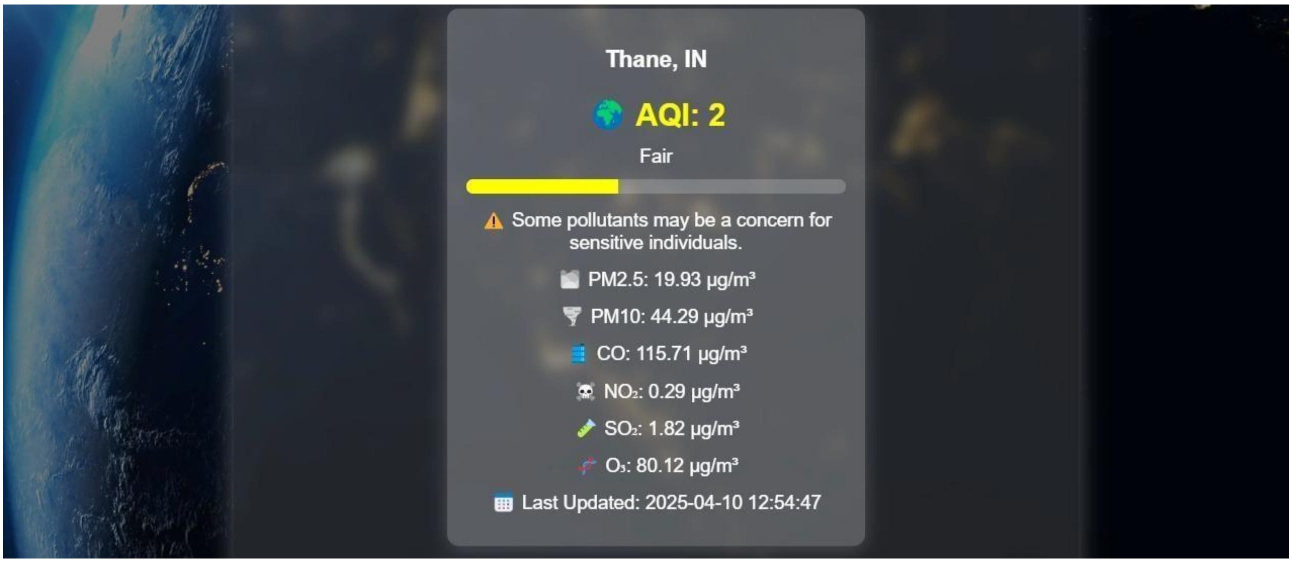


Figure 5.4.3: Air Quality Index Page

This Figure 5.4.3, the interface offers detailed information on the air quality status of a selected location. The AQI score is clearly displayed, along with a descriptive label such as "Fair," making it easy for users to understand current conditions. A color-coded progress bar gives a quick visual cue on the quality level, and a warning icon highlights potential health concerns for sensitive individuals.

The display further breaks down the presence of key atmospheric pollutants, including:

- **PM2.5** (fine particulate matter)
- **PM10** (coarse particulate matter)
- **CO** (carbon monoxide)
- **NO_2** (nitrogen dioxide)
- **SO_2** (sulfur dioxide)
- **O_3** (ozone)

Each pollutant is listed with its concentration in micrograms per cubic meter (µg/m³), providing users with comprehensive environmental data. The module also includes the timestamp of the latest update, ensuring users are accessing the most recent information available.

4. Air Quality Index Page

The platform also features a robust weather forecasting section, aimed at delivering detailed meteorological insights in a user-friendly visual format. Positioned against the same high-tech backdrop, this section helps maintain visual continuity while shifting focus to weather data. Users are greeted with the heading “Weather Prediction Graph,” followed by a trio of well-organized charts. These visual tools not only enhance the user experience but also empower individuals to make weather-informed decisions quickly and confidently.

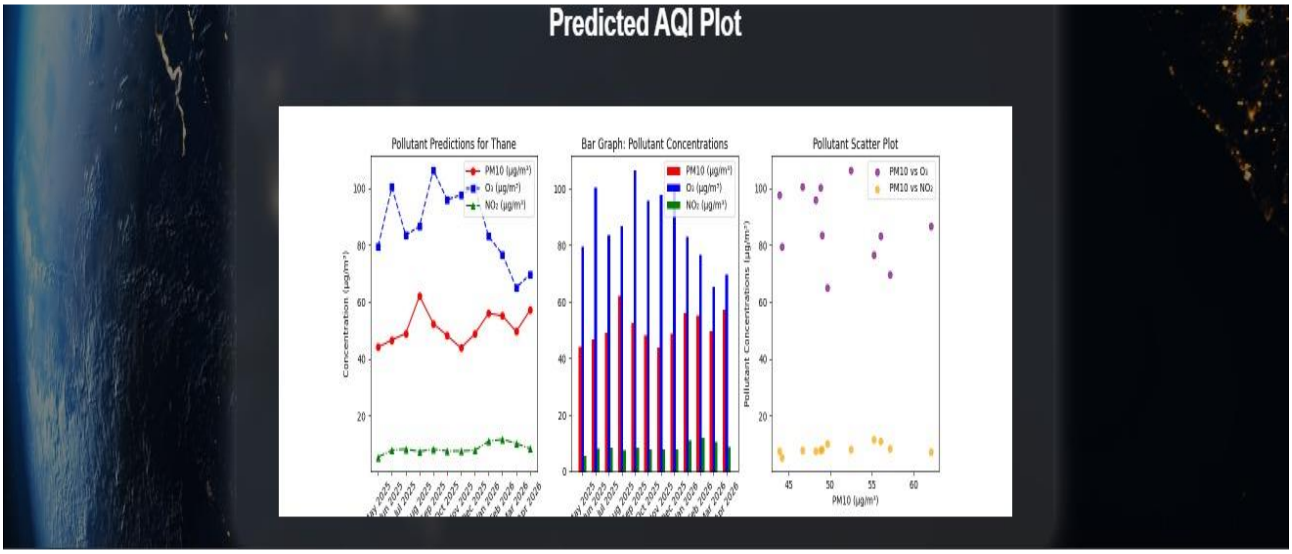


Figure 5.4.4: Prediction of Air Quality on different Plot

This section of Figure 5.4.4 provides three distinct types of graphical representations that work together to present a clear picture of upcoming weather trends:

1. **Line Graph** – This graph tracks daily predictions for temperature (in °C) and humidity (in %), offering a temporal overview that reveals patterns such as warming trends or moisture fluctuations.
2. **Bar Graph** – A comparative bar chart illustrates the relationship between temperature and humidity, making it easy to visually interpret changes in both variables across several days.
3. **Scatter Plot** – This plot maps temperature against humidity to examine how the two variables interact. It serves as an analytical tool to spot any correlation, such as higher humidity at lower temperatures.

Together, these graphs provide a multi-angle view of the weather, enabling users to understand not just what the weather will be, but how different factors relate to each other — an ideal feature for both everyday users and data enthusiasts.

Chapter 6

Technical Specification

Our project specifications encompass the selection of programming languages and frameworks to ensure that the project is equipped with the appropriate resources for compatibility, scalability, and efficiency throughout its development and deployment phases.

Front-end:

- **Development Framework:**
  HTML5, CSS 4.15
- **Functionality:**
  User Interface (UI) for displaying AQI and weather forecast information.
  Receiving user input for location-based searches.
  Presenting real-time AQI levels, historical trends, and weather conditions.

Back-end:

- **Development Framework:**
  Python : Flask, Matplotlib, Numpy , Pandas, Skitlearn
- **Functionalities:**
  - Fetching real-time AQI and weather data from external APIs.
  - Implementing a Random Forest Regression Model for AQI prediction.
  - Processing user requests for specific city-based forecasts.
  - Generating alerts when AQI surpasses safety thresholds.
  - Storing user preferences for personalized forecasts.

Database Management:

Datasets:

- **Datasets:**
  Source: Kaggle dataset (Weather & AQI historical data)
- **Collections:**
  - **Users:** Stores user preferences, login credentials, and search history.
  - **AQI Data:** Stores historical air quality index values, pollutant concentrations, and trends.
  - **Weather Data:** Maintains weather parameters (temperature, humidity, wind speed, etc.).
  - **Predictions:** Records AQI predictions generated by the machine learning model.

Chapter 7

Project Scheduling

In our project, the Gantt chart outlines key activities where each task would be represented by a bar on the chart, indicating its start and end dates, duration, and dependencies, allowing project stakeholders to track progress, identify potential delays, and timely completion of project objectives.

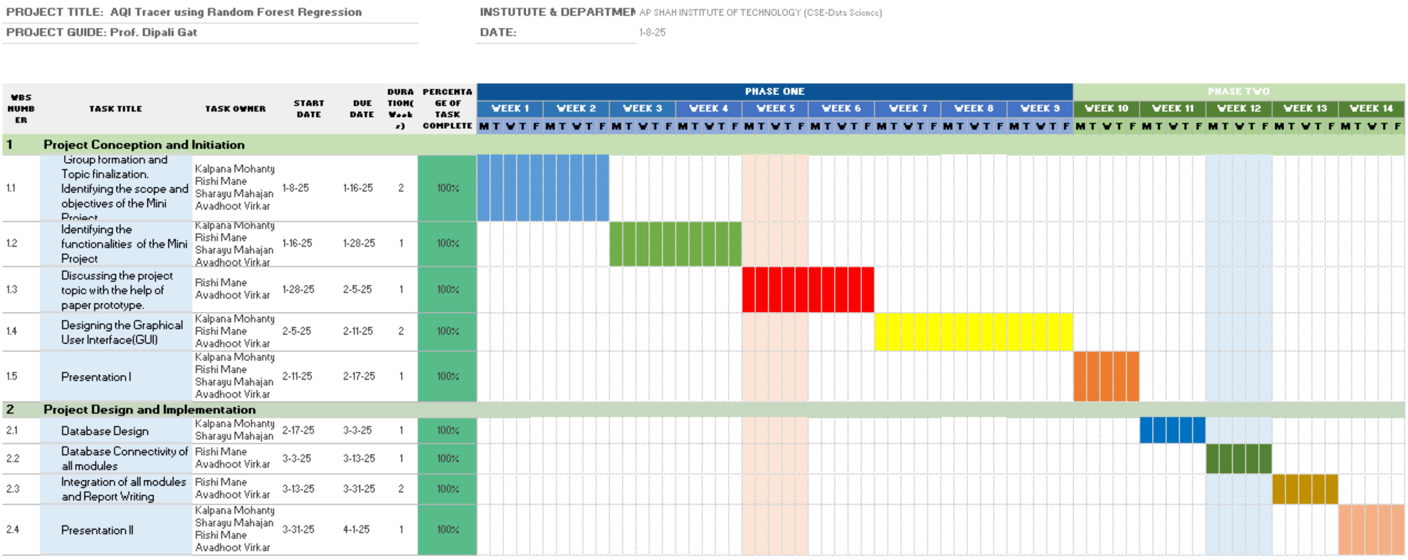


Figure 7.2: Gantt Chart

The details of the finalized project are outlined in section 7.2, in the second week of January, Kalpana Mohanty, Rishi Mane, Avadhoot Virkar, and Sharayu Mahajan formed a group for their mini project. They discussed and finalized the project's topic, scope, and objectives during this meeting. In the following weeks, Rishi Mane and Avadhoot Virkar used a paper prototype to explore and refine project ideas, completing this phase by the 2nd week of February.

In early February, Rishi Mane, Avadhoot Virkar, and Kalpana Mohanty executed the design and integration of the graphical user interface (GUI). Afterward, on the 11th of February, the first project review took place, and the faculty suggested some changes to the GUI, which were subsequently approved. Following this, Kalpana Mohanty and Sharayu Mahajan collaborated to create a structured database system, facilitating the systematic storage of information.

This, in turn, made it easier for Rishi Mane and Avadhoot Virkar to connect the database to the project. This database work was completed by the 3rd week of March. Finally, the team integrated all modules and completed the report writing, resulting in their final presentation on 01st April, which was approved by the faculty.

Chapter 8

Project Results

The Weather and AQI Prediction System was designed to provide accurate forecasts of environmental conditions, including temperature, humidity, and Air Quality Index (AQI), helping users plan their activities and protect their health. After evaluating multiple models, Random Forest Regressor was selected as the final model due to its high accuracy and reliability.

Model Evaluation and Selection

Initially, two models were considered:

- **Random Forest Regressor:**
  Random Forest Regression is a machine learning algorithm that uses an ensemble of decision trees to predict continuous values. It works by averaging the predictions of multiple decision trees to improve accuracy and reduce overfitting.
- **FB Prophet**
  FB Prophet is an open-source forecasting tool developed by Facebook, designed for time series data. It automatically detects trends, seasonality, and holidays to make accurate future predictions, and is especially useful for data with strong seasonal effects and historical patterns.

Final Model Performance:

Accuracy Achieved:

- **Random Forest Regressor**: 85%
- **FB Prophet**: 60%

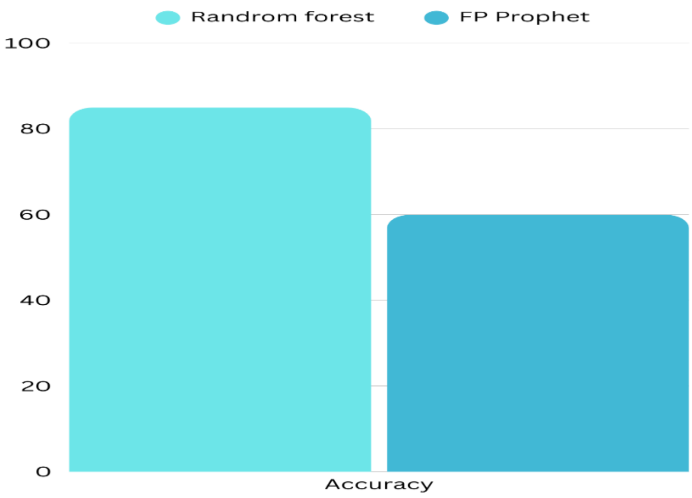
These results indicate that the Random Forest model significantly outperforms FB Prophet in terms of prediction accuracy.

Statistical Performance Metrics: Random Forest Regression

- **Mean Squared Error (MSE):** Low
- **R^2 Score:** Close to 0.98, indicating strong model fit

These metrics affirm the robustness of the Random Forest model in capturing data variability and minimizing prediction errors.

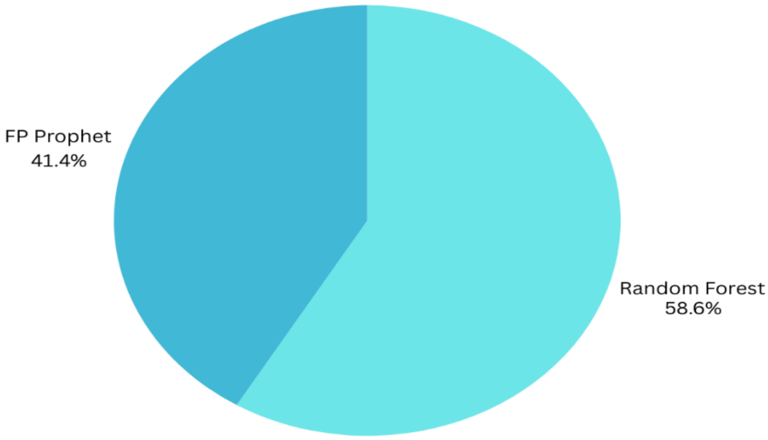
The bar graph shows that Random Forest Regression achieved higher accuracy than FB Prophet in predicting AQI values. Due to its better performance, Random Forest was selected for our project.



- **Figure 8.1: Accuracy Comparison (Random Forest vs FB Prophet)**

(Random Forest shown as superior in a bar chart)

In Figure 8.1, the bar graph compares the accuracy scores of Random Forest Regression and FB Prophet. The height of each bar reflects the model's performance in predicting future AQI (Air Quality Index) values. As shown, Random Forest achieved a higher accuracy than FB Prophet, indicating its better ability to learn complex patterns in the data.



- **Figure 8.2: Accuracy Distribution**

(Pie chart reflecting Random Forest's dominance)

In Figure 8.2, the pie chart compares the accuracy of two forecasting models: Random Forest and FB Prophet. Random Forest outperformed FB Prophet with an accuracy of 58.6% versus 41.4%. While FB Prophet showed limited accuracy during evaluation (around 60%), the Random Forest Regressor achieved approximately 85% accuracy and was selected for the final implementation due to its superior performance. This choice significantly improved the reliability of weather predictions in the system.

Key Insights

1. **Random Forest Regressor** demonstrates higher accuracy and more stable predictions across different parameters such as temperature, humidity, and AQI.
2. **FB Prophet**, while useful for trend visualization and seasonality detection, showed relatively lower predictive accuracy.
3. Visual comparison of model accuracy highlights the dominance of the Random Forest model in this use case.

Significance of the Results

The results emphasize that:

1. The Random Forest model is highly reliable for real-world AQI and weather prediction use cases.
2. The comparative study confirms that FB Prophet, while effective for capturing trends, is less accurate in complex multi-feature environments.
3. Accurate AQI predictions enable better health planning, especially in urban regions with fluctuating air quality.

Challenges & Solutions:

1. **Data Irregularities:**
   Incomplete AQI readings and missing weather values were addressed using interpolation and imputation techniques.
2. **Data Cleaning:**
   Missing and noisy data were cleaned using imputation and smoothing techniques.

Future Enhancements

- **Real-time Prediction**:
  Deploying models as APIs to provide live weather and AQI updates.

- **Geo-spatial Insights:**
  Integrating location-based modeling for city-wise AQI heatmaps.

- **Hybrid Models:**
  Combining deep learning (LSTM) with ensemble methods for even better prediction accuracy.

Chapter 9

Conclusion

The AQI Tracer project highlights the potential of integrating real-time air quality monitoring with advanced data analytics to provide users with accurate and meaningful insights. By leveraging machine learning models such as Random Forest Regression, the system effectively predicts Air Quality Index (AQI) levels and helps individuals make informed decisions regarding outdoor activities and health precautions.

The project successfully integrates real-time API data, ensuring that users receive up-to-date air quality information. The use of visual representations like color-coded AQI categories and graphical trends enhances the accessibility of the data for both experts and the general public.

Furthermore, the modular architecture of AQI Tracer ensures flexibility and scalability, allowing future improvements such as dynamic pollutant tracking, predictive alerts, and advanced health recommendations. The system's commitment to accuracy, efficiency, and user engagement positions it as a valuable tool in the field of environmental monitoring and public health awareness.

Chapter 10

Future Scope

The future development of AQI Tracer envisions expanding its capabilities to provide more precise and personalized air quality insights. One major enhancement includes the implementation of deep learning techniques to refine AQI predictions, enabling more accurate long-term forecasts. Additionally, the integration of multiple government and private air quality APIs will improve data reliability and consistency, ensuring users receive the most accurate real-time information.

To further enhance its impact, health risk assessments based on AQI trends and individual health profiles will be incorporated, providing personalized recommendations for users, especially those with respiratory conditions. Geospatial analysis will also be improved, allowing for interactive maps with real-time AQI visualization and pollutant dispersion patterns, helping users understand local air quality variations more effectively.

Expanding accessibility, a mobile application will be developed to offer real-time AQI monitoring, push notifications, and location-based air quality alerts, making crucial air quality data readily available to users. Furthermore, IoT integration will enable connectivity with low-cost air sensors and wearable devices, allowing hyper-local AQI readings and improved data granularity.

By implementing these advancements, AQI Tracer aims to become a comprehensive air quality monitoring solution, empowering individuals, communities, and policymakers to take proactive measures toward a healthier and more sustainable environment.

References

[1] Chen, Y., Zhang, L., ‘Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis’, March 15, 2023.

[2] Kumar, R., Sharma, P., ‘A Study and Analysis of Air Quality Index and Related Health Impact on Public Health’, July 10, 2020.

[3] Gupta, S., Verma, A., ‘Air Quality Index – A Comparative Study for Assessing the Status of Air Quality’, April 5, 2015.

[4] Wang, J., Lee, C., ‘Machine Learning Approaches for Weather Prediction and AQI Forecasting’, September 14, 2022.

[5] Patel, R., Kumar, N., ‘Enhancing Accuracy in Weather Forecasting Using AI and Big Data’, February 3, 2024.