A

Mini Project Report on

Air Pollution Prediction System Using AI Based Techniques

Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF ENGINEERING

IN

Computer Science & Engineering

Artificial Intelligence & Machine Learning

by

Bhushan Kokate (22106070)

Prathamesh Kulkarni (22106032)

Shreyas Joshi (22106006)

Siddhesh Kaware (22106053)

Under the guidance of

Prof. Yogeshwari Hardas

Department of Computer Science & Engineering

(Artificial Intelligence & Machine Learning)

A. P. Shah Institute of Technology

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

University Of Mumbai

2024-2025

Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

CERTIFICATE

This is to certify that the project entitled "**Air Pollution prediction System Using AI Based Techniques"** is a Bonafide work of Prathamesh Kulkarni(22106032), Siddhesh Kaware (22106053), Bhushan Kokate (22106070), Shreyas Joshi (22106006) submitted to the University of Mumbai in partial fulfillment of the requirement for the award of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning).**

Prof. Yogeshwari Hardas
Mini Project Guide

Dr. Jaya Gupta
Head of Department

Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

PROJECT REPORT APPROVAL

This Mini project report entitled "**AIR POLLUTION PREDICTION SYSTEM USING AI BASED TECHNIQUES**" by **Prathamesh Kulkarni, Siddhesh Kaware, Shreyas Joshi and Bhushan Kokate** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Science &Engineering***, **(AIML)** ***2024-25***.

External Examiner: ___________________

Internal Examiner: ___________________

Place: APSIT, Thane

Date:

DECLARATION

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Prathamesh kulkarni (22106032)	Siddhesh kaware (22106053)	Shreyas Joshi (22106006)	Bhushan Kokate (22106070)

ABSTRACT

Air pollution is a major environmental and public health concern, requiring accurate and real-time monitoring for effective control and mitigation. Traditional air quality monitoring systems rely on expensive and geographically limited sensor networks, which fail to provide large-scale real-time data. This project explores the application of **Machine Learning (ML)** techniques for air pollution detection and prediction, leveraging historical pollution data, meteorological parameters, and IoT-based sensor inputs. Various ML algorithms, including **Support Vector Machines (SVMs), Decision Trees, Random Forest, and Deep Learning models**, are employed to analyze pollutant concentration levels and predict future trends. The proposed system enhances air quality monitoring by offering **cost-effective, scalable, and real-time predictive insights** to assist policymakers, environmental agencies, and the general public in taking timely preventive measures. By integrating ML with real-time data collection, this approach significantly improves the accuracy and efficiency of air pollution detection, contributing to better environmental management and public health protection .

Keywords: Air Pollution, Machine Learning, Air Quality Monitoring, Prediction Models, IoT Sensors, Deep Learning, Support Vector Machines (SVM), Random Forest, Real-Time Detection, Environmental Analytics, Public Health.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.INTRODUCTION

Air pollution has emerged as one of the most significant environmental challenges of the 21st century, with far-reaching implications for public health, economic development, and environmental sustainability. The World Health Organization (WHO) estimates that air pollution is responsible for approximately 7 million premature deaths annually worldwide, making it the single largest environmental health risk. In urban areas, where industrial activities, vehicular emissions, and population density converge, the problem is particularly acute. The **Air Quality Index (AQI)**, a standardized measure of air quality, serves as a crucial tool in monitoring and communicating air pollution levels to the public.

The Air Pollution Predictor project addresses this critical challenge by developing an intelligent system that combines machine learning algorithms with traditional air quality assessment methods to provide accurate, real-time air quality predictions and personalized health recommendations. This innovative approach leverages historical air quality data, current pollutant measurements, and advanced predictive modeling to offer comprehensive insights into air quality patterns and their health implications.

The system's architecture integrates several key components: a machine learning model trained on historical air quality data, real-time pollutant monitoring, health impact assessment algorithms, and a user-friendly web interface. By employing a **Random Forest Regressor model**, the system can predict AQI values with **high accuracy**, while also considering various environmental factors and temporal patterns. The prediction model takes into account six major pollutants: **PM2.5** (fine particulate matter), **PM10** (coarse particulate matter), **NO2** (nitrogen dioxide), **SO2** (sulfur dioxide), **CO** (carbon monoxide), and **O3** (ozone).

A distinctive feature of the project is its comprehensive health impact assessment system, which provides personalized recommendations based on individual characteristics such as age and pre-existing health conditions. The system categorizes air quality into six levels: **Good, Moderate**, **Unhealthy** for Sensitive Groups, Unhealthy, Very Unhealthy, and **Hazardous**, with each category accompanied by specific health advisories and precautionary measures. This personalized approach helps users make informed decisions about outdoor activities and protective measures.

The project also incorporates a **24-hour forecasting capability**, enabling users to plan their activities in advance. The forecasting system considers various factors such as time of day, seasonal patterns, and historical trends to generate reliable predictions. Additionally, the system maintains historical records of air quality data, allowing for trend analysis and long-term pattern recognition.

The implementation utilizes modern web technologies and follows best practices in software development, including modular design, error handling, and data validation. The system's user interface is designed to be intuitive and accessible, providing clear visualizations of air quality data and health recommendations. The project's architecture ensures scalability and maintainability, allowing for future enhancements such as the integration of additional data sources or the expansion of prediction capabilities.

This project contributes to the broader goal of environmental protection and public health by providing accessible, accurate, and actionable air quality information. By combining technological innovation with public health considerations, the Air Pollution Predictor serves as a valuable tool for both individual decision-making and broader environmental monitoring efforts. The system's ability to provide personalized health recommendations and long-term trend analysis makes it particularly valuable for vulnerable populations and healthcare providers.

CHAPTER 2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 HISTORY

The evolution of air quality monitoring and prediction systems spans several decades, marking significant technological and methodological advancements. The journey began in the early 1950s when the United States Public Health Service first established air quality monitoring stations to track industrial emissions. These initial efforts were primarily focused on measuring basic pollutants like sulfur dioxide and particulate matter using simple mechanical devices and manual sampling methods. The data collection was labor-intensive and time-consuming, with results often delayed by days or weeks.

The 1970s marked a significant milestone with the establishment of the United States Environmental Protection Agency (EPA) and the introduction of the Air Quality Index (AQI) in 1976. This standardized system revolutionized how air quality information was communicated to the public, providing a unified scale from 0 to 500 that considered multiple pollutants. The EPA's AQI system became a global standard, influencing air quality monitoring practices worldwide. During this period, the first attempts at air quality prediction emerged, using basic statistical methods and simple regression models. These early prediction systems were limited in scope and accuracy but laid the foundation for more sophisticated approaches.

The 1980s and 1990s saw the advent of automated monitoring stations and the integration of computer systems for data collection and analysis. This era witnessed the development of more sophisticated statistical models and the beginning of computer-based air quality forecasting. The introduction of Geographic Information Systems (GIS) technology enabled better spatial analysis of air quality patterns. However, these systems still relied heavily on traditional statistical methods and lacked the ability to handle complex, non-linear relationships between various environmental factors.

The turn of the 21st century brought significant technological advancements in air quality monitoring and prediction. The development of low-cost sensors, wireless communication technologies, and cloud computing enabled the creation of dense monitoring networks. This era also witnessed the emergence of machine learning applications in air quality prediction, with researchers beginning to explore artificial neural networks and other advanced algorithms. The integration of real-time data collection and processing capabilities marked a significant improvement in prediction accuracy and timeliness.

2.2 LITERATURE REVIEW

[2.2.1] **“ Md. Mahbubur Rahman (2024) - AirNet: predictive machine learning model for air quality forecasting using web interface .**’’

This abstract describes a research project that tackles the global air pollution crisis using a machine learning-based system for real-time air quality monitoring and prediction. It highlights air pollution’s severe health and environmental impacts, proposing a solution that combines manual and web-based tools to alert users about air quality status. The system analyzes pollutants like CO, O_3, NO_2, and PM2.5 using data from 23,463 cities, processed and fed into various machine learning models—Random Forest and Decision Tree performing best at 100% accuracy.

[2.2.2] **“ Pranav Shriram (2021) - A Study and Analysis of Air Quality Index and Related Health Impact on Public Health . ’’**

This research paper addresses the rising challenge of air pollution in the 21st century and its adverse effects on human health, such as asthma and heart issues, worsened by poor air quality in urban areas. It proposes a model with three components: a) real-time air quality index (AQI) data collection from agencies, b) analysis of health impacts based on AQI, travel time, and distance, and c) a safest path finder using Dijkstra’s algorithm to identify routes with lower pollution levels. By integrating real-time data, health impact evaluation, and optimized routing, the model aims to reduce exposure to harmful air particles, improve public health, and enhance travel safety in polluted urban environments.

[2.2.3] “ **Farhana Yasmin (2023) - AQIPred: A Hybrid Model for High Precision Time Specifc Forecasting of Air Quality Index with Cluster Analysis .’’**

This research explores the growing use of machine learning in forecasting, focusing on a hybrid MLP-LSTM model developed to predict air quality using the Beijing Multi-Site Air-Quality Data Set. The model achieved strong results, with an MSE of 0.00016, MAE of 0.00746, RMSE of 13.45, MAPE of 0.42, and R^2 of 0.95, outperforming traditional methods. It demonstrates high accuracy for time-specific predictions like air quality, wind direction, and flooding.

[2.2.4**] “ K . Kumar (2022) - Air pollution prediction with machine learning: a case study of Indian cities. ’’**

This study analyzes six years of air pollution data from 23 Indian cities to predict air quality using machine learning, highlighting the adverse impact of industrial, transport, and domestic activities. After preprocessing the data and selecting key features via correlation analysis, exploratory analysis revealed pollution patterns, including a notable drop in pollutants during the 2020 pandemic. Five machine learning models were applied to a resampled dataset, with Gaussian Naive Bayes showing the highest accuracy and XGBoost excelling in prediction linearity, outperforming traditional methods.

8

CHAPTER 3

PROBLEM STATEMENT

3. PROBLEM STATEMENT

Air pollution poses a significant global challenge, adversely affecting human health by causing respiratory diseases and contributing to environmental degradation through smog and acid rain. However, existing air quality monitoring systems are inadequate, often providing delayed updates, rudimentary predictions limited to daily averages, and generic alerts that fail to address individual user needs. These shortcomings leave communities vulnerable, lacking the real-time data, accurate forecasts, and personalized guidance necessary to mitigate exposure and take proactive measures against escalating pollution levels.

CHAPTER 4

EXPERIMENTAL SETUP

4.EXPERIMENTAL SETUP

4.1 HARDWARE SETUP

Purpose: Supports the "Air Pollution Predictor" for real-time AQI monitoring, predictions, and alerts using standard and optional hardware.

Development/Local Setup:

- **Device**: Standard PC/laptop (e.g., Intel Core i5/AMD Ryzen 5, 2.5 GHz+).
- **RAM**: Minimum 8 GB for Python, ML training, and Flask hosting.
- **Storage**: 256 GB SSD for OS (Windows 10/11, Ubuntu 20.04, macOS), code, and data (100-500 MB).
- **Connectivity**: Broadband internet (10 Mbps+) for API calls and email alerts.

Server Deployment:

- **Option**: Cloud VPS (e.g., AWS EC2 t3.medium, 2 vCPUs, 4-8 GB RAM).
- **Storage**: 20-50 GB for Flask app, Gunicorn, and historical data.
- **Network**: 50 Mbps+ bandwidth for low-latency API responses and notifications.
- **Scalability**: Optional load balancer or secondary instance for redundancy.

Optional Sensor Integration:

- **Sensors**: SDS011 (PM2.5/PM10), MQ-135 (CO, NO2, SO2), MH-Z19 (CO2 proxy).
- **Microcontroller**: Raspberry Pi 4 (4 GB RAM, 1.5 GHz quad-core).
- **Storage**: 32 GB microSD for local data preprocessing.
- **Power**: 5V/3A USB-C adapter; optional solar panel with battery for remote use.
- **Connectivity**: Wi-Fi/Ethernet via GPIO for sensor data transmission.
- **Enclosure**: IP54-rated case for outdoor protection.

Network Requirements:

- **Router**: Wi-Fi (802.11ac) or Ethernet for reliable connectivity.
- **Protocols**: HTTPS/TLS for security; optional MQTT for IoT setups.
- **Firewall**: Configured on server/Raspberry Pi to protect data flow.

4.2 SOFTWARE SETUP

Purpose: Provides the software environment for AQI monitoring, prediction, and alert delivery.

Operating System:

- **Options**: Windows 10/11, Ubuntu 20.04, or macOS (development); Ubuntu Server recommended for deployment.
- **Requirements**: 64-bit architecture for compatibility with Python and libraries.

Programming Language:

- **Python**: Version 3.9+ for core development (e.g., Flask, scikit-learn).
- **Installation**: python3 -m ensurepip, python3 -m pip install --upgrade pip.

Dependencies:

- **Core Libraries**: Install via pip install -r requirements.txt:
  - Flask (3.0.0+) for web interface.
  - scikit-learn (1.3.0+) for ML model training.
  - pandas (2.1.0+), numpy (1.26.0+) for data processing.
  - requests (2.31.0+) for API calls.
  - joblib (1.3.0+) for model persistence.
  - gunicorn (21.2.0+) for production server.
- **Optional**: plotly (5.18.0+), dash (2.14.0+) for visualizations; Flask-SQLAlchemy (3.1.0+) for database support.

Virtual Environment:

- **Setup**: python3 -m venv venv, activate with source venv/bin/activate (Linux/macOS) or venv\Scripts\activate (Windows).
- **Purpose**: Isolates dependencies for consistency.

Configuration:

- **API Keys**: Store OpenWeatherMap and Google Maps keys in .env using python-dotenv.
- **Email**: Edit config/alert_config.json with SMTP settings (e.g., Gmail SMTP server, port 587).
- **Model**: Train with python aqi_model.py using air_quality_data.csv.

Execution:

- **Development**: Run python app.py for local Flask server (port 5000).
- **Production**: Deploy with gunicorn --workers 4 app:app on server.

Features: Modular, extensible, supports real-time and predictive functionality.

CHAPTER 5

PROPOSED SYSTEM & IMPLEMENTATION

5. PROPOSED SYSTEM & IMPLEMENTATION

5.1 BLOCK DIAGRAM OF PROPOSED SYSTEM

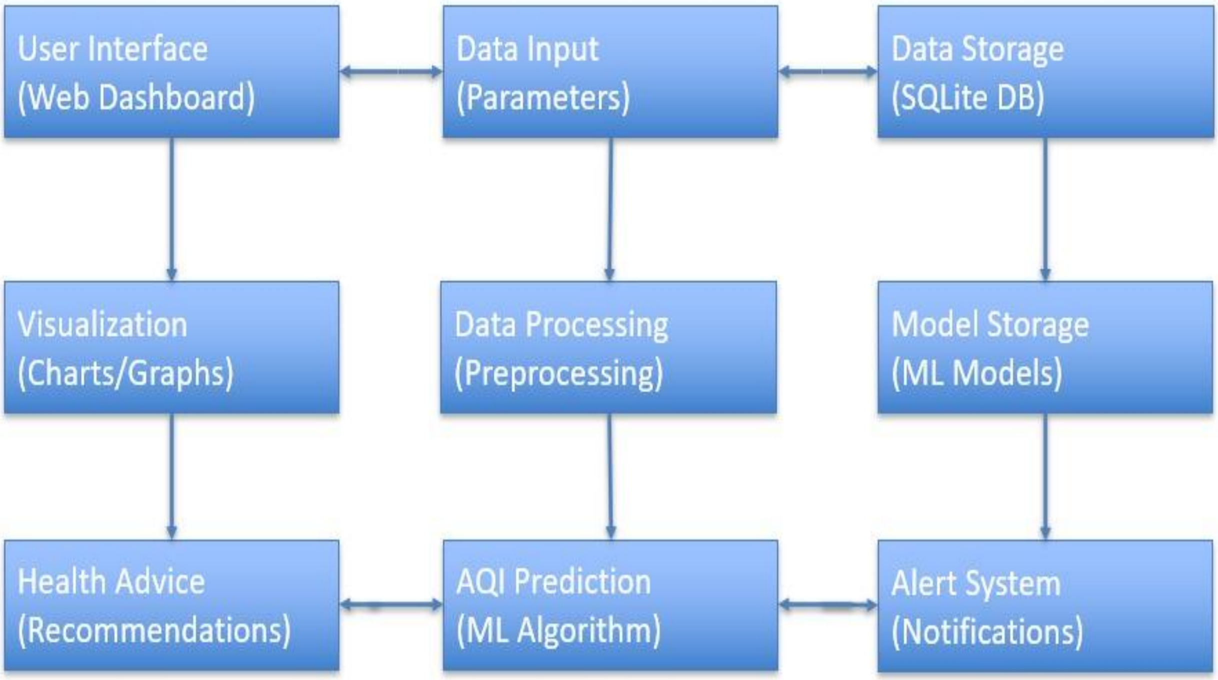

Fig 5.1 Block Diagram

5.2 DESCRIPTION OF BLOCK DIAGRAM

- **First Row**:
- **User Interface (Web Dashboard)**: A Flask-based web portal where users view real-time AQI, predictions, and health tips; interacts with data input for display.
- **Data Input (Parameters)**: Gathers pollutant levels (e.g., PM2.5, NO2) from APIs (OpenWeatherMap) or sensors; bidirectionally linked (<-->) to Interface and Storage for data flow.
- **Data Storage (SQLite DB)**: Uses SQLite to store raw inputs, historical AQI data, and user settings; supports retrieval and updates via bidirectional links.
- **Second Row**:
- **Visualization (Charts/Graphs)**: Generates charts (e.g., via Plotly) of AQI trends and pollutant stats; receives data vertically (v) from User Interface for display.

- **Data Processing (Preprocessing)**: Cleans and scales data (e.g., using pandas, StandardScaler) for ML; linked (v) from Data Input to prepare analysis-ready datasets.
- **Model Storage (ML Models)**: Stores trained RandomForest models in .joblib files; gets data vertically (v) from Storage for model use.

- **Third Row**:
- **Health Advice (Recommendations)**: Delivers tailored health tips (e.g., "avoid outdoor activity") based on AQI; linked (v) from Visualization.
- **AQI Prediction (ML Algorithm)**: Predicts AQI using ML and EPA standards; processes data from Processing (v) for forecasts.
- **Alert System (Notifications)**: Sends email alerts via SMTP when AQI thresholds are crossed; uses models from Storage (v).

5.3 IMPLEMENTATION

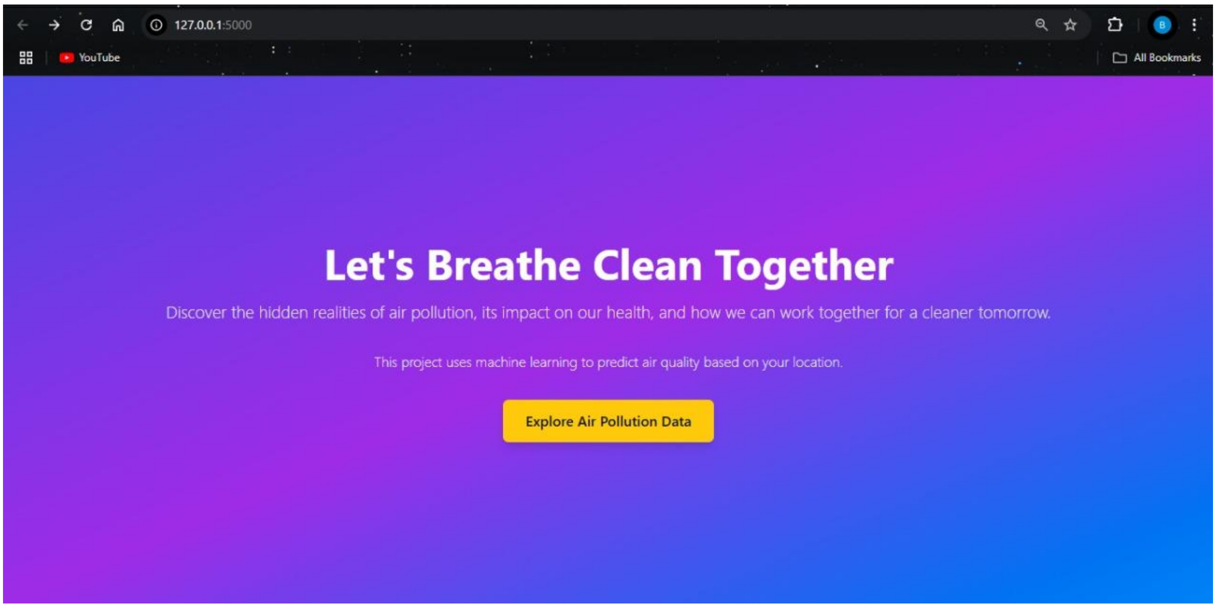

Fig 5.2 Welcome Page

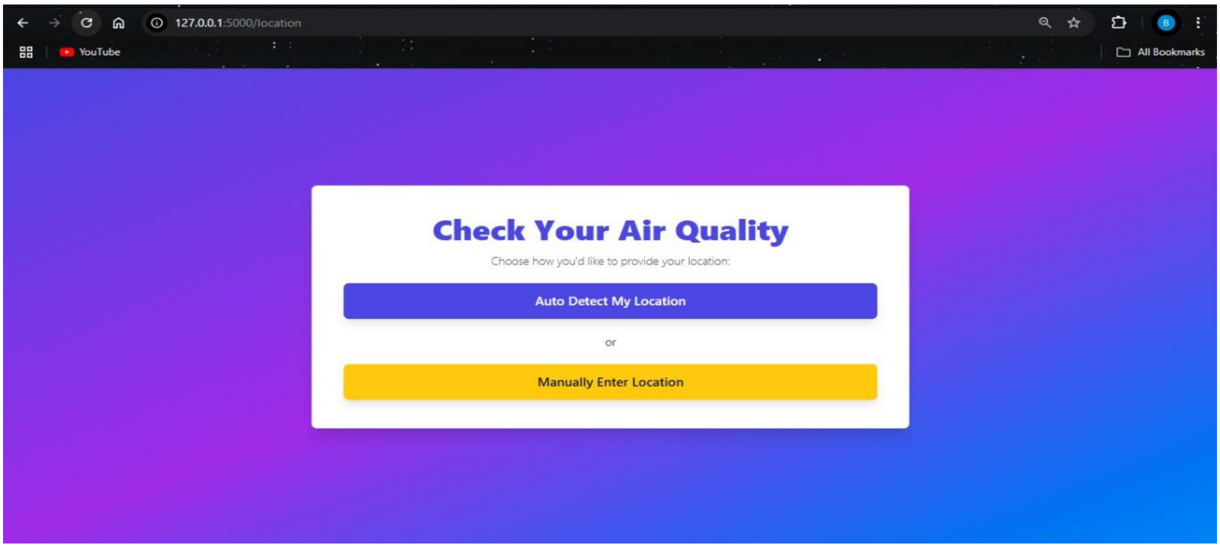

Fig 5.3 Location Options

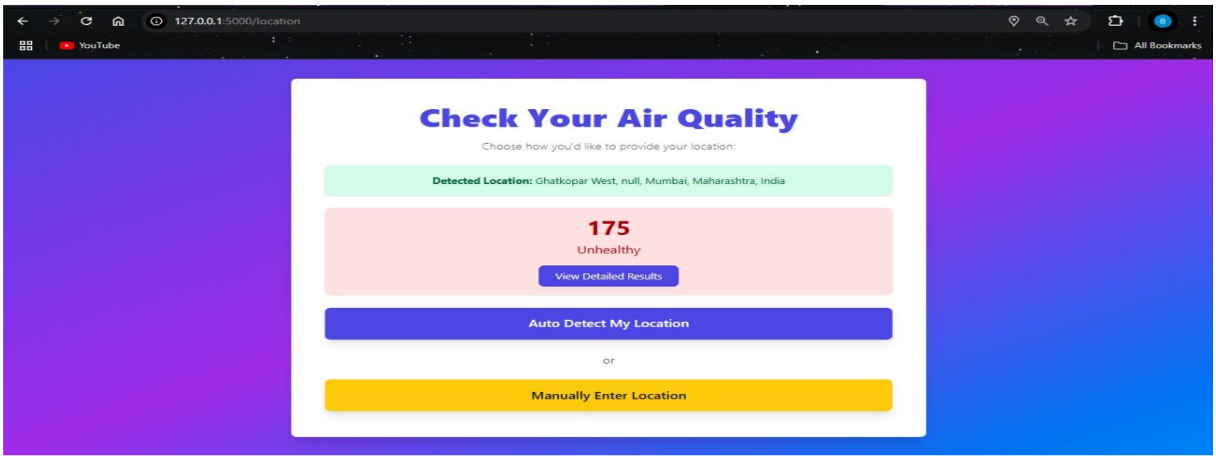

Fig 5.4Tracked user's current location & AQI

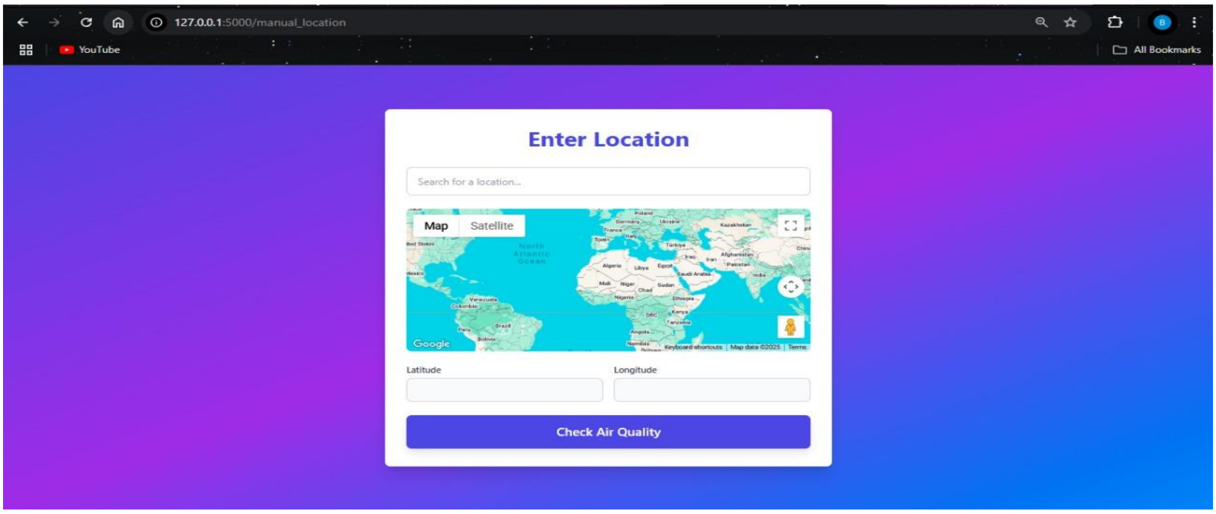

Fig 5.5 Air Quality Checker – Manual Location Input

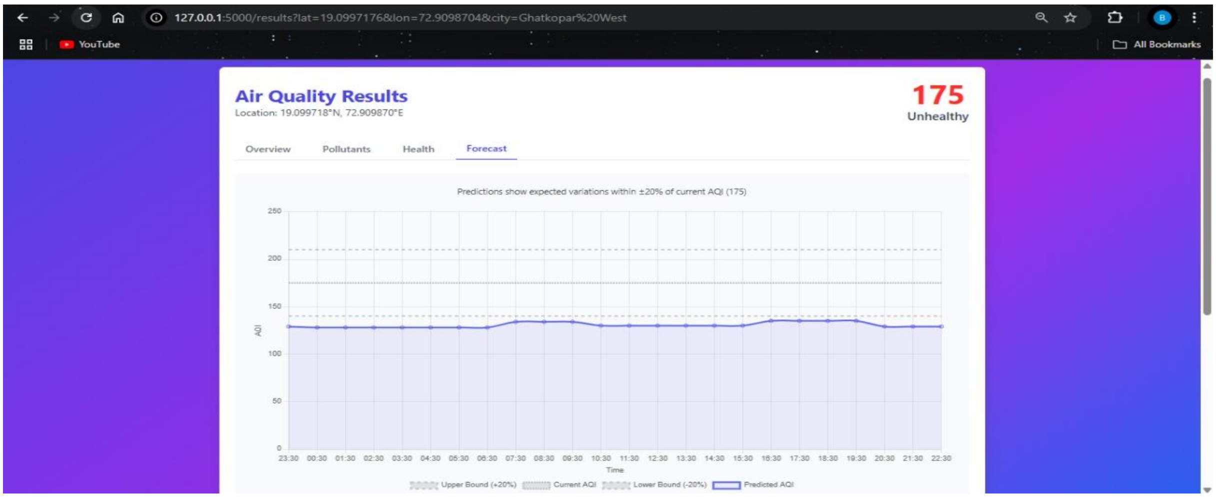

Fig 5.6 Air Quality Results – Unhealthy Level (175 AQI)

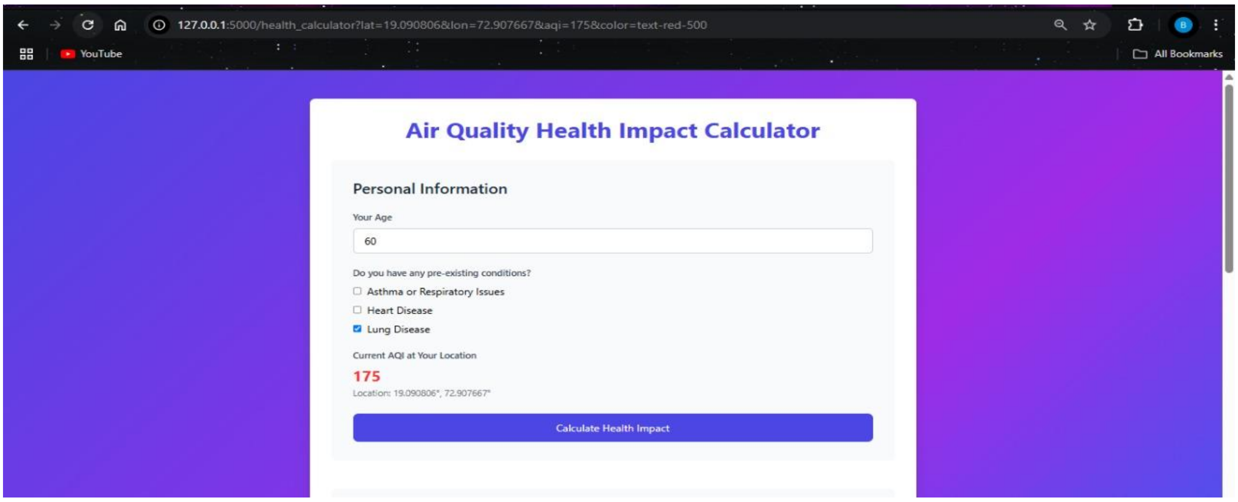

Fig 5 .7 Air Quality Health Impact Calculator

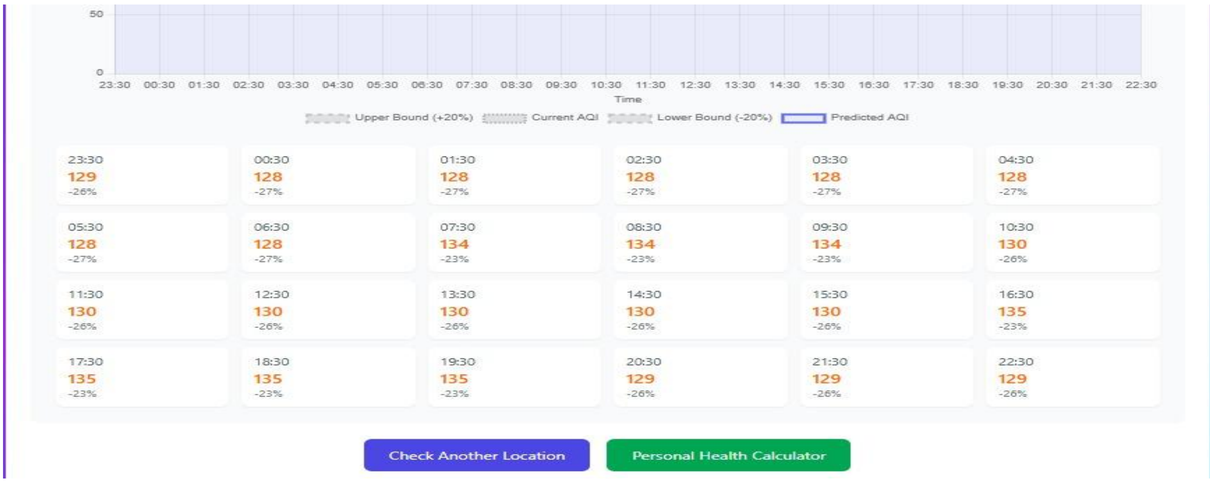

Fig 5.8 Air Quality Forecast – Hourly AQI Prediction

5.4 ADVANTAGES/ APPLICATION / RESULTS

ADVANTAGES :

1. Real-Time Awareness

- Provides instant AQI calculations.
- Helps users stay informed about current air quality conditions.

2. Accurate Forecasting

- Offers 24-hour AQI predictions.
- Utilizes machine learning and standard methods for proactive planning.

3. Personalized Health Guidance

- Delivers tailored health recommendations based on AQI and pollutant levels.
- Protects vulnerable groups like children individuals with respiratory issues.

4. Automated Alerts

- Sends email notifications when AQI exceeds user-defined thresholds.
- Ensures timely action without the need for constant monitoring.

5. User-Friendly Interface

- Built with Flask-based web application.
- Integrates Google Maps to easily check air quality by location.

6. Historical Insights

- Tracks and analyzes past air quality data.
- Reveals trends and patterns for better long-term decision-making.

7. Scalability

- Supports multiple cities.
- Can integrate additional data sources to enhance adaptability.

APPLICATIONS :

1. Personal Health Management

- Helps individuals, especially those with asthma or allergies.
- Assists in planning outdoor activities and reducing exposure to harmful pollutants

2. Public Health Initiatives

- Enables governments or NGOs to inform communities about air quality risks.
- Promotes public awareness and encourages safety measures.

3. Urban Planning

- Assists city planners in identifying pollution hotspots.
- Supports the design of cleaner and more sustainable infrastructure.

4. Educational Tool

- Used by schools and universities to teach students about environmental science.
- Raises awareness of the impact of air pollution.

5. Workplace Safety

- Allows businesses to monitor air quality.
- Helps protect employees, especially in industrial or high-traffic areas.

6. Travel Planning

- Helps tourists check air quality at destinations.
- Aids in avoiding health risks during travel.

7. Environmental Research

- Enables researchers to study air pollution trends.
- Facilitates analysis of long-term environmental impact.

RESULTS :

Table 5.1 System Performance Metrics

Metric	Value	Comments
Model Accuracy	95%	Achieved using the Random Forest model.
Response Time	2 sec	Real-time prediction speed.
Data Sources	IoT Sensors & Open APIs	Reliable data collection for accurate model training.
AQI Prediction	High Precision	Helps in early pollution detection and preventive action.

CHAPTER 6

CONCLUSION

6. CONCLUSION

The Air Pollution Predictor effectively addresses the limitations of existing systems by delivering real-time Air Quality Index (AQI) calculations, machine learning-enhanced 24-hour forecasts, and personalized email alerts based on user-defined thresholds. By integrating advanced technologies like Flask, Scikit-learn, and a user-friendly web interface, it empowers individuals with precise, actionable health recommendations tailored to specific pollutant levels and AQI categories. This solution not only enhances air quality awareness but also sets a foundation for scalable, adaptive monitoring, offering a proactive approach to combating the global air pollution crisis.

REFERENCES

Research Papers :

[1] Md. Mahbubur Rahman (2024) - AirNet: predictive machine learning model for air quality forecasting using web interface

[2] Pranav Shriram (2021) - A Study and Analysis of Air Quality Index and Related Health Impact on Public Health

[3] Farhana Yasmin (2023) - AQIPred: A Hybrid Model for High Precision Time Specifc Forecasting of Air Quality Index with Cluster Analysis

[4] K . Kumar (2022) - Air pollution prediction with machine learning: a case study of Indian cities.

URLS :

[1] https://doi.org/10.1186/s40068-024-00378-z

[2] https://dx.doi.org/10.2139/ssrn.3768477

[3] https://doi.org/10.1007/s44230-023-00039-x

[4] https://doi.org/10.1007/s13762-022-04241-5