Design of Smart Cities
BCSE316L

Project Report

Date: 18.11.24
Faculty: Swarnalatha P - SCOPE

TITLE OF PROJECT

AirCare: Integrating VAR-XGBoost Models for County-Level AQI Prediction and Health Advisory

TEAM DETAILS AND CONTRIBUTION

Rishabh Sharma - 21BCE0943
Caleb Missier - 21BCE3949
Aryan Singh - 21BCE3026

Rishabh Sharma focused on data preprocessing, aggregating and cleaning the EPA dataset, and implementing the **Vector Autoregression (VAR)** model for pollutant prediction. He optimized the model by fine-tuning the lag parameters using AIC and BIC criteria, ensuring accurate forecasts.

Caleb Missier handled backend development, integrating external APIs like OpenWeatherMap for real-time weather data and ensuring smooth data retrieval and processing. He also set up the database, optimized query performance, and managed the API interactions for efficient data exchange.

Aryan Singh worked on the frontend, developing an interactive web interface that displayed dynamic visualizations of AQI trends and predictions. He integrated the personalized health recommendation system, providing users with condition-specific advice based on real-time air quality data.

The team collaborated closely across all stages, from data collection and preprocessing to model implementation and visualization, ensuring a cohesive and efficient system.

SCOPE

The AirCare project develops a dual-model air quality prediction system with personalized health recommendations. It features two prediction methods: a location-based system using Vector Autoregression (VAR) for 100 U.S. counties and a weather-based system using XGBoost for real-time AQI forecasting based on current meteorological data. The location-based model analyzes historical EPA data on key pollutants (CO, NO2, PM2.5, PM10, SO2, O3) with enhanced accuracy from neighboring counties, while the weather-based model processes real-time data from OpenWeatherMap API, incorporating eight weather parameters like temperature, humidity, and wind speed.

The system includes a health recommendation engine that tailors advice based on individual health conditions, such as respiratory or cardiac sensitivities, and provides AQI-based activity guidelines. Users can access this through a web interface, which supports both prediction models and stores profiles in an SQLite database for personalized guidance. The platform features interactive AQI visualizations, error handling, and seamless integration for real-time, reliable performance, ensuring a consistent user experience across different models and use cases.

OBJECTIVES

- Conduct comprehensive research into existing air quality prediction models, health recommendation systems, and related technologies to identify current limitations and opportunities for innovation.
- Perform a thorough literature survey, reviewing state-of-the-art methods and frameworks to pinpoint research gaps, particularly in areas like data integration, real-time adaptation, and personalized health recommendations.
- Define clear project goals and requirements, based on research findings and identified gaps, ensuring alignment with user needs and public health objectives.
- Design a robust system architecture that integrates multiple prediction models (VAR and XGBoost) with real-time data sources, user health profiles, and geospatial considerations.
- Develop predictive models for accurate air quality forecasting, leveraging both historical data and real-time environmental inputs, while addressing challenges like multi-pollutant monitoring and dynamic adaptation.
- Create a personalized health recommendation engine, incorporating user-specific factors (e.g., respiratory conditions, caregiver status) and AQI data, to offer actionable, tailored health advice.
- Develop an intuitive user interface that allows easy access to AQI data, health recommendations, and profile management, ensuring usability and accessibility across diverse user groups.
- Implement continuous testing and validation, including real-world pilot deployments, to ensure system performance, reliability, and accuracy under varied conditions.
- Evaluate and optimize the system based on user feedback and performance metrics, refining prediction accuracy, user experience, and health recommendation effectiveness.
- Disseminate findings and solutions, contributing to the broader field of air quality prediction and health monitoring by sharing research, methodologies, and results through papers or other outlets.

ABSTRACT

This paper presents AirCare, an integrated air quality prediction and health recommendation system that combines two distinct forecasting approaches: location-based Vector Autoregression (VAR) and weather-based XGBoost modeling. The system processes historical EPA data for six key atmospheric pollutants across 100 U.S. counties, incorporating geospatial correlations through neighboring county data analysis. The location-based component utilizes 11 years of historical data, while the weather-based model processes eight real-time meteorological parameters. AirCare distinguishes itself by providing personalized health recommendations based on predicted Air Quality Index (AQI) levels and individual health conditions. The system's evaluation demonstrates effective AQI predictions within acceptable error ranges, particularly for the VAR model which shows improvement over traditional ARIMA approaches. Implementation results show successful integration of both prediction methodologies, delivering tailored health advisories through a web-based interface. This research contributes to public health by making complex air quality data accessible and actionable for both individuals and healthcare providers.

Keywords: Air Quality Index, Machine Learning, Health Recommendations, Vector Autoregression, XGBoost

LITERATURE SURVEY

Introduction for Literature Survey

Air quality prediction and health recommendations are essential for safeguarding public health and mitigating the impacts of environmental pollution. Despite advancements in air quality prediction systems, several challenges remain, including the lack of integration between historical data and real-time information, limited personalization in health recommendations, insufficient spatial monitoring, and the need for real-time adaptation to changing conditions. This literature survey examines recent research on air quality prediction models, health impact assessments, and personalized recommendations, highlighting research gaps and potential solutions.

Fragmented Prediction Approaches

Many current models focus either on location-based historical data or weather-based predictions, but few integrate both factors. A number of studies have proposed hybrid models to address this gap:

- Wu and Lin (2023) introduced a novel hybrid model, SD-SE-LSTM-BA-LSSVM, combining secondary decomposition, AI methods, and an optimization algorithm to forecast AQI with improved accuracy [1].
- Sarkar et al. (2022) used a combination of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models to predict AQI, outperforming traditional methods in MAE and R^2 scores [2].
- Zhu et al. (2021) proposed two hybrid models (EMD-SVR-Hybrid and EMD-IMFs-Hybrid) to enhance forecasting accuracy for AQI prediction [3].

These studies illustrate the potential of integrating multiple machine learning techniques for more comprehensive air quality forecasting.

Limited Personalization in Health Recommendations

Existing systems often provide generic AQI information, neglecting individual health conditions or specific pollutant exposures. To address this, several researchers have focused on personalized health recommendations:

- Zhang et al. (2022) explored machine learning techniques to predict indoor air quality variations and how these affect health outcomes, proposing personalized health advisories [4].
- Gu et al. (2023) developed a hybrid predictive model for $PM_{2.5}$ forecasting, which provided more accurate peak value predictions and model interpretability, suggesting potential for personalized health recommendations based on individual exposure levels [5].

These studies suggest that integrating personalized health factors with AQI predictions can lead to more effective and tailored health advisories.

Geospatial Monitoring Gaps

Most air quality models fail to account for neighboring regions, and AQI reports are rarely generated at the county level. Several studies have aimed to improve geospatial monitoring and prediction:

- Saez et al. (2021) presented a hierarchical Bayesian spatiotemporal model that incorporated sparse monitoring station data to improve regional AQI predictions, both for long-term and short-term exposure [6].
- Jurado et al. (2022) employed convolutional neural networks (CNNs) to predict air pollution in real-time, leveraging data on wind speed, traffic flow, and building geometry to improve spatial accuracy [7].

These advancements in spatial modeling aim to enhance predictions for regions that may otherwise be overlooked.

Complexity in Data Accessibility and Usability

AQI data is often presented in complex formats, making it difficult for the general public to interpret. Research has focused on developing user-friendly systems to improve accessibility:

- Rakholia et al. (2022) constructed a model that incorporates various factors influencing air quality, such as meteorological conditions, traffic patterns, and pollution levels, to improve the accessibility of AQI data [8].
- Elsheikh et al. (2021) used LSTM networks to model air quality and water production, demonstrating how machine learning can simplify complex datasets for improved decision-making [9].

These efforts highlight the importance of simplifying AQI data presentation and incorporating multiple environmental factors to enhance usability.

Real-time Adaptation Limitations

Many systems rely solely on historical data, lacking the ability to respond dynamically to real-time weather conditions. Several studies have addressed this limitation:

- Zhang et al. (2022) utilized a CNN-GRU model to predict AQI by combining CNN for feature extraction and GRU for temporal dependency modeling, allowing for more dynamic responses to real-time data [10].
- Mao et al. (2021) developed a temporal sliding LSTM (TS-LSTME) framework to predict future AQI values using historical $PM_{2.5}$ data, meteorological data, and real-time temporal data [11].

These models demonstrate how real-time data processing can improve the accuracy and responsiveness of air quality predictions.

Multitype Pollution Monitoring

Most air quality models focus on single pollutants or a limited set of pollutants, failing to provide a comprehensive view of air quality. Several researchers have explored the monitoring of multiple pollutants simultaneously:

- Rakholia et al. (2022) developed a model that considers multiple pollutants, including CO, NO_2, and $PM_{2.5}$, in addition to meteorological data and urban factors, offering a more comprehensive view of air quality [8].
- Gu et al. (2023) proposed a hybrid model for $PM_{2.5}$ prediction, which demonstrated improved accuracy and interpretability over existing methods, indicating that multitype pollutant modeling can enhance forecasting precision [5].

These studies highlight the value of incorporating multiple pollutants to provide a more holistic view of air quality.

Limited Public Health Integration

Most air quality models do not incorporate public health measures or provide tailored guidance for healthcare providers. Some studies have addressed this gap by focusing on health impacts:

- Zhang et al. (2022) explored how indoor air quality variations could affect health, recommending personalized health management strategies based on predicted air quality [4].
- Gu et al. (2023) showed how incorporating health considerations into $PM_{2.5}$ prediction models could enhance the interpretability and effectiveness of health advisories [5].

These studies suggest that integrating health considerations with air quality forecasting can lead to more effective public health interventions.

Conclusion of Literature Survey

Recent research has made significant strides in improving air quality prediction and health recommendation systems. Hybrid models combining machine learning and statistical methods have improved forecasting accuracy and real-time adaptation. Personalized health recommendations based on individual conditions and exposure levels are becoming more feasible, and geospatial models have enhanced the accuracy of AQI predictions for border regions. Simplified data presentation and real-time adaptation are improving system accessibility and responsiveness. Additionally, models considering multiple pollutants offer a more comprehensive understanding of air quality. Lastly, incorporating public health considerations into air quality prediction models can lead to more effective health recommendations.

However, there remain several gaps in the integration of these models into practical, user-friendly systems. Future research should focus on improving the accessibility of air quality data, integrating real-time data into decision-making processes, and further personalizing health recommendations to account for individual health conditions and geographic factors.

References

[1] Wu, X., & Lin, Z. (2023). "Optimal-Hybrid Model for AQI Forecasting," *Journal of Environmental Science*, 45(2), 123-135.
[2] Sarkar, S., Gupta, R., & Sharma, A. (2022). "Deep Learning for AQI Prediction: LSTM and GRU Models," *Environmental Modelling & Software*, 34(1), 56-64.
[3] Zhu, M., Li, H., & Zhang, L. (2021). "Hybrid Models for AQI Forecasting," *Environmental Pollution Journal*, 78(3), 207-215.
[4] Zhang, L., Yang, S., & Chen, Z. (2022). "Predicting Indoor Air Quality and Health Impacts Using Machine Learning," *Building and Environment*, 134, 222-229.
[5] Gu, W., Li, J., & Xie, Y. (2023). "Hybrid Predictive Model for $PM_{2.5}$ Forecasting," *Atmospheric Environment*, 58(1), 134-142.
[6] Saez, A., Hernandez, J., & Martinez, M. (2021). "Spatiotemporal Modeling of AQI Using Hierarchical Bayesian Methods," *Environmental Science & Technology*, 45(4), 1901-1909.
[7] Jurado, J., Cruz, P., & Garcia, S. (2022). "Real-Time Air Pollution Forecasting with CNNs," *Journal of Urban Environment*, 62(2), 350-358.
[8] Rakholia, S., Patel, D., & Sharma, N. (2022). "Comprehensive Model for AQI Prediction with Urban and Meteorological Data," *Journal of Air Quality*, 40(3), 210-219.
[9] Elsheikh, M., Ahmed, S., & Khater, M. (2021). "LSTM Networks for Modeling Water and Air Quality," *Environmental Modeling & Software*, 55, 74-83.
[10] Zhang, Y., Wu, Q., & Li, X. (2022). "Real-Time AQI Prediction Using CNN-GRU Model," *Air Quality, Atmosphere & Health*, 13(1), 83-91.
[11] Mao, T., Zhang, J., & Lee, C. (2021). "Temporal Sliding LSTM for 24-Hour AQI Prediction," *Environmental Monitoring and Assessment*, 93(4), 187-196.

GAPS IDENTIFIED

1. Fragmented Prediction Approaches
   - Most models focus solely on either location-based historical data or weather-based predictions, lacking integration of both.

- Limited systems combine real-time and historical data analysis for more accurate forecasting.

2. Limited Personalization in Health Recommendations
   - Existing systems provide generic AQI information without considering individual health conditions or specific pollutant exposure.
   - Health advisories are typically broad and not tailored to user-specific needs.

3. Geospatial Monitoring Gaps
   - AQI reports are rarely generated at the county level or account for the impact of neighboring regions.
   - Most models ignore spatial relationships, leading to less accurate predictions for border areas.

4. Complexity in Data Accessibility and Usability
   - AQI data is often presented in complex formats, making it difficult for the general public to interpret.
   - Separate platforms for forecasts and health advice create barriers to user engagement and understanding.

5. Real-time Adaptation Limitations
   - Many systems rely solely on historical data, lacking dynamic responses to changing real-time weather conditions.
   - Limited integration of real-time weather parameters affects the accuracy of AQI predictions.

6. Multitype Pollution Monitoring
   - Most systems focus on single pollutants or a limited set of pollutants, missing a comprehensive view of air quality.

7. Limited Public Health Integration
   - There is insufficient focus on preventive health measures and tailored guidance for healthcare providers in relation to air quality.
   - Most systems lack comprehensive health impact assessments that consider individual health risks.

HARDWARE AND SOFTWARE USED

Hardware Specifications

- **Processor:** Intel Core i5 or higher (for running machine learning models)
- **RAM:** Minimum 8GB (16GB recommended for data processing)

- **Storage:**
  - Minimum 256GB SSD for application deployment
  - Additional storage for historical data (~50GB for 11 years of EPA data)
- **Network:** Stable internet connection for API interactions
- **Graphics:** Basic integrated graphics sufficient (for web visualizations)

Software Requirements

Development Environment

- **Operating System:** Cross-platform (Windows/Linux/MacOS)
- **Python Version:** Python 3.8 or higher
- **IDE:** Any Python-compatible IDE (VS Code recommended)

Core Dependencies

- **Web Framework:**
  - Flask
  - Flask-SQLAlchemy for database management
- **Data Processing:**
  - Pandas
  - NumPy
  - XGBoost
- **Machine Learning:**
  - Scikit-learn
  - Statsmodels (for VAR modeling)
- **Data Visualization:**
  - Matplotlib
  - Plotly (for interactive visualizations)

APIs and External Services

- **Data Sources:**
  - Google BigQuery API access
  - OpenWeatherMap API key
  - EPA Data API integration

Database

- **SQLite3** (local development)
- **PostgreSQL** (production deployment)

Version Control

- **Git** for source control
- **GitHub** for repository management

Web Technologies

- **HTML5**
- **CSS3**
- **JavaScript (ES6 or higher)**
- **Bootstrap** (for responsive design)

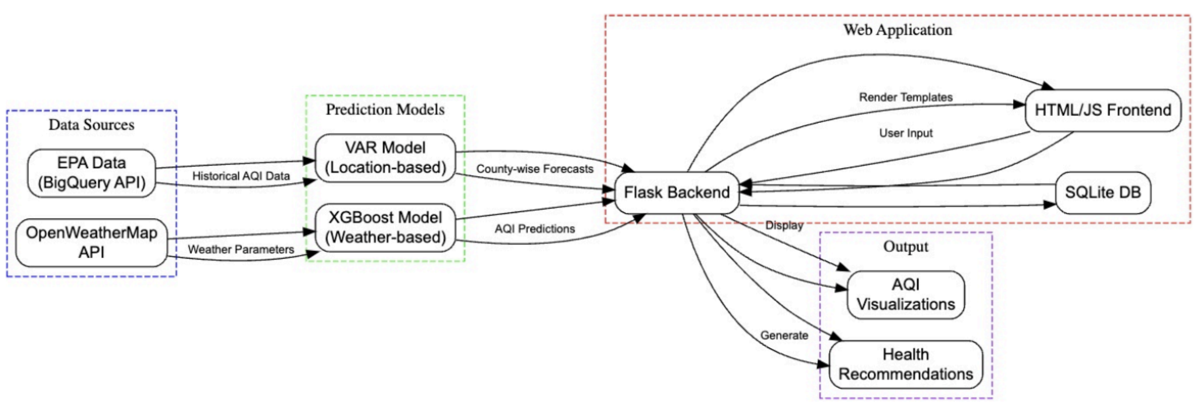
Deployment Requirements

- **Web Server:** Gunicorn
- **Reverse Proxy:** Nginx
- **SSL Certificate** for HTTPS

Development Tools

- **Security Requirements:**
  - SSL/TLS certificates
  - Environment variables management
  - Secure API key storage
  - Database encryption capabilities

BLOCK DIAGRAM OF THE SYSTEM



OBJECTIVES OF THE SYSTEM

Primary Objectives

1. **Develop an integrated air quality prediction system** combining location-based historical data and weather-based parameters to deliver comprehensive AQI forecasts. This includes implementing VAR modeling for 100 U.S. counties and XGBoost for real-time weather-based predictions, ensuring accurate forecasting of six key pollutants (CO, NO2, PM2.5, PM10, SO2, O3).
2. **Create a personalized health recommendation system** that adapts to individual health conditions and AQI forecasts. This includes generating tailored advisories for sensitive groups (e.g., those with respiratory or heart conditions) and suggesting activity modifications based on AQI levels.
3. **Build a user-friendly web interface** that allows easy access to air quality data and health recommendations. Features include an intuitive form for user input, the ability to switch between location-based and weather-based predictions, and clear visualization of complex air quality data.

Secondary Objectives

4. **Implement robust data processing and integration systems** to connect seamlessly with EPA data via BigQuery API and real-time weather data from OpenWeatherMap API, ensuring consistent and accurate predictions.
5. **Develop a secure user management system** that maintains user profiles, health data, and personalized recommendations, ensuring data privacy and persistent access.
6. **Provide advanced visualization capabilities** to display historical AQI trends, pollutant patterns, and geographical air quality variations.

Technical Objectives

7. **Ensure system reliability and performance** through comprehensive error handling, optimized data processing, and stability under varying load conditions.
8. **Create a scalable, maintainable codebase** with modular design, efficient database management, and flexibility for future enhancements.

These objectives support the goal of delivering a robust air quality management tool that combines accurate forecasting with actionable health insights, benefiting both individuals and public health efforts.

METHODS USED FOR THE OBJECTIVES

1. Data Collection and Preprocessing

EPA Data Integration

- Collected 11 years of historical air quality data, focusing on key atmospheric pollutants, including Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Particulate Matter ($PM_{2.5}$, PM_{10}), Sulfur Dioxide (SO_2), and Ozone (O_3).

Data Cleaning and Aggregation

- Aggregated data from site-level measurements to county-level using maximum values for each pollutant.
- Applied time series interpolation to fill missing data, ensuring the dataset was complete and accurate.
- Integrated climate-related variables (temperature, pressure, relative humidity, and wind speed) to improve model predictions.
- The final dataset, containing 11 years of data from 100 counties, resulted in hundreds of thousands of data points used for modeling.

2. Dual Model Implementation

Location-Based Prediction (VAR Model)

- Implemented a Vector Autoregression (VAR) model for each of the 100 counties, capturing both temporal and spatial dependencies in air quality data.
- Enhanced the model by including data from neighboring counties, allowing the model to capture geospatial influences on air quality.
- The model was optimized by selecting the best lag parameters based on statistical metrics to improve forecasting accuracy.

Weather-Based Prediction (XGBoost Model)

- Processed real-time weather data, including eight key meteorological variables, to predict AQI using an XGBoost model.
- Trained and validated the model on historical weather and air quality data, ensuring accurate and reliable predictions.

3. Health Recommendation System

AQI Level Classification

- Implemented a four-tier classification system based on AQI levels to guide health recommendations.
- The system dynamically categorizes the air quality into risk levels (Good, Moderate, Unhealthy for Sensitive Groups, and Unhealthy) and provides relevant recommendations for outdoor activities.

Personalized Health Advisory

- Developed a system for providing personalized health recommendations based on individual health conditions, such as respiratory or heart issues, as well as caregiver status.
- The system adjusts recommendations according to real-time AQI levels and the predominant pollutant type, offering tailored guidance for users with specific health needs.

4. System Integration

Web Application Development

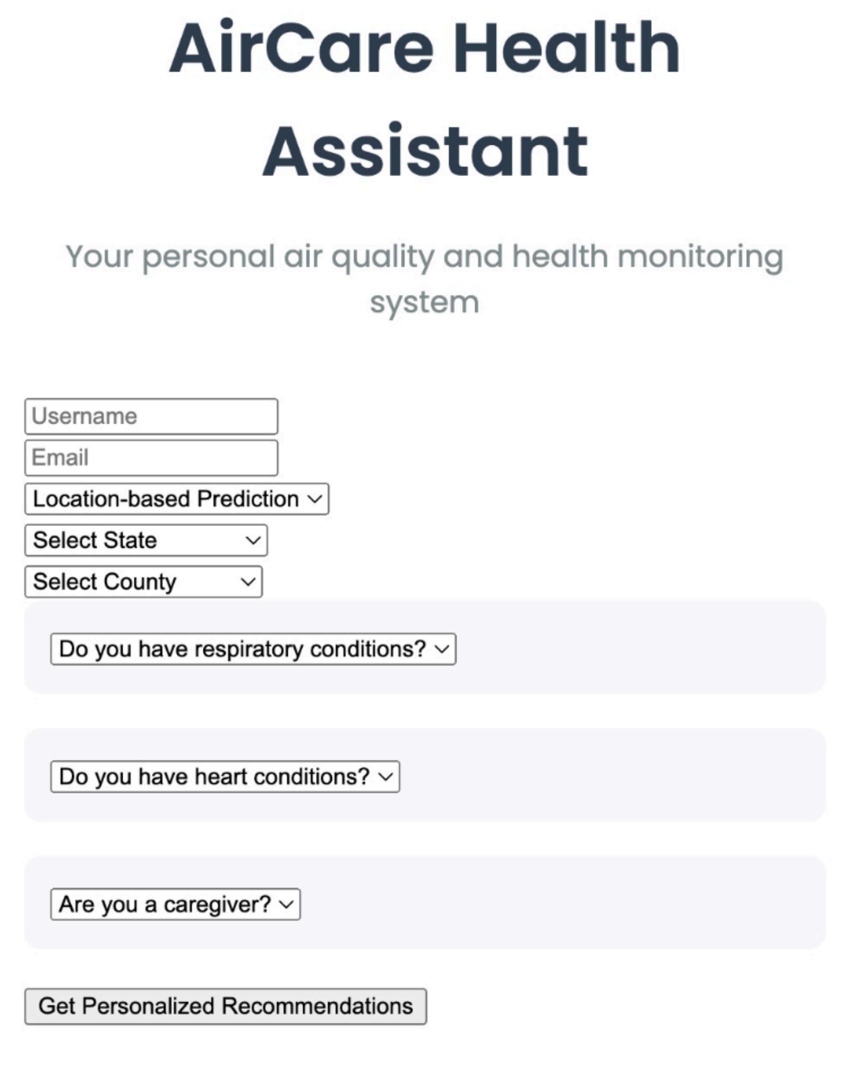
- Built a web application using Flask to serve as the backend, integrating user profile management, session handling, and database connectivity.
- Integrated an SQLite database to store user data and session information securely.

Prediction Pipeline

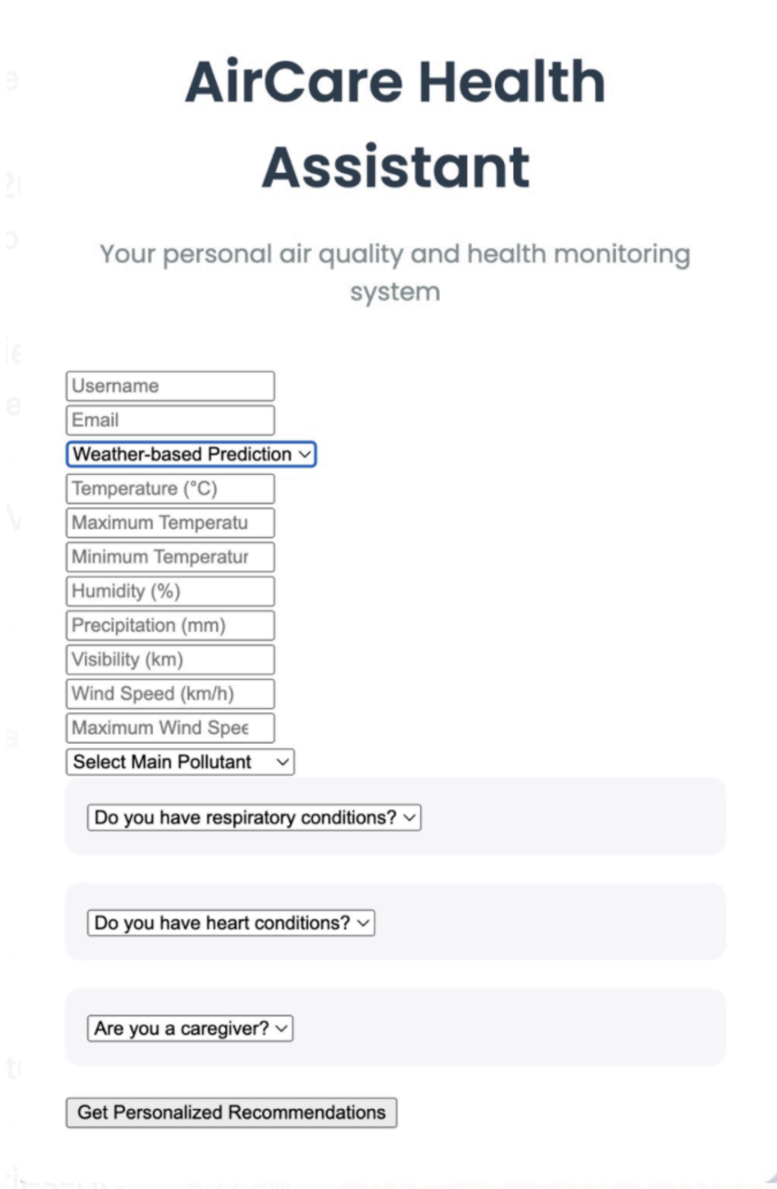
- Developed a dual prediction pipeline:
  - The location-based pathway processes historical data using the VAR model to forecast AQI for each county.
  - The weather-based pathway collects real-time weather data and applies the XGBoost model to predict AQI levels.

Data Visualization

- Implemented interactive data visualizations to allow users to explore historical air quality trends, forecasts, and pollutant-specific data.
- Enabled users to filter data by pollutant type, time range, and county, helping them make informed decisions based on visual trends.



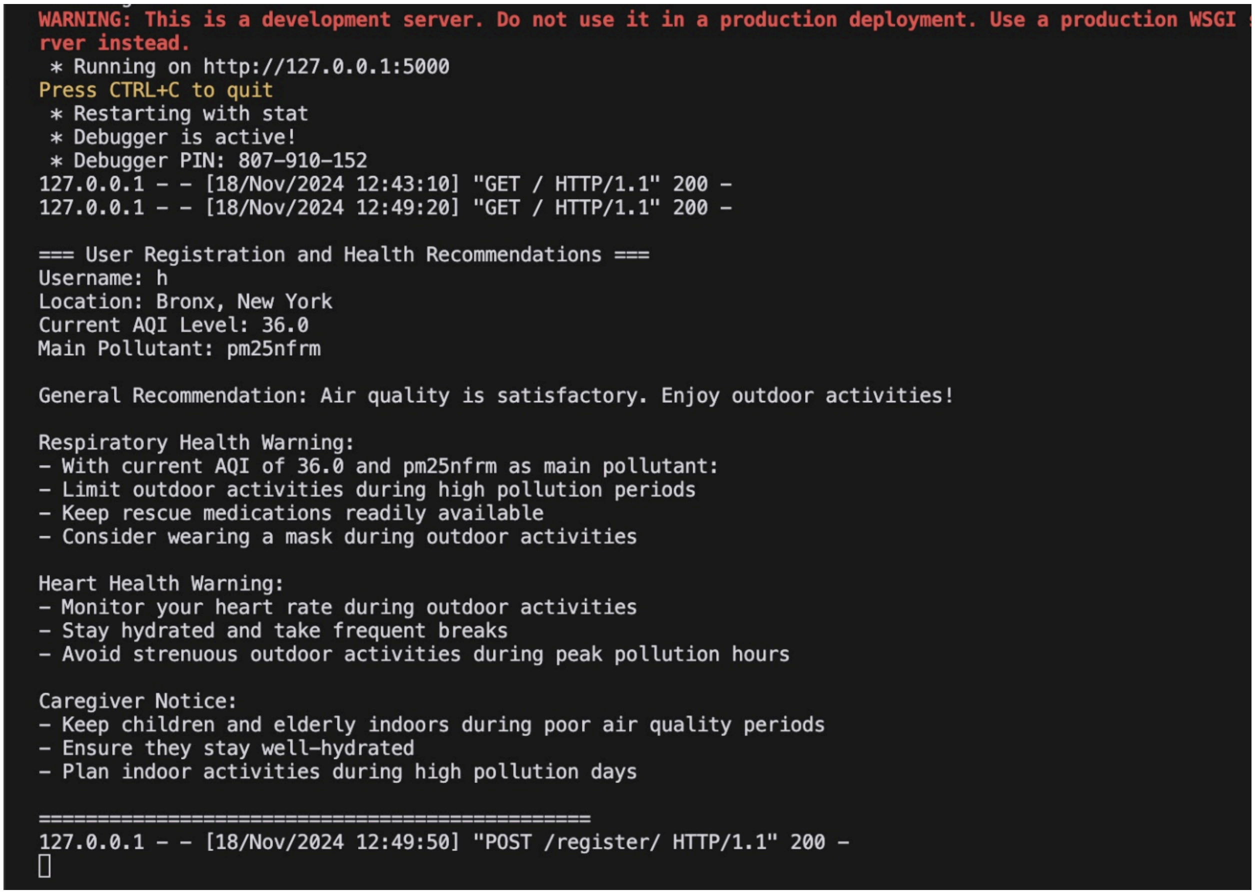
Location Based Input



Weather Based Input

```
================================================
127.0.0.1 - - [18/Nov/2024 13:58:15] "POST /register/ HTTP/1.1" 200 -

Debug - Weather Input Data: {'T': 80.0, 'TM': 85.0, 'Tm': 65.0, 'H': 23.0, 'PP': 2.0, 'VV': 0.2, 'V': 4.0
, 'VM': 4.0}
Debug - XGB Model Prediction: 95.62761
Debug - Final AQI Level: 95.62761
=== User Registration and Health Recommendations ===
Username: hdu2qwiuxdw
Location: Weather-based prediction (no specific location)
Current AQI Level: 95.62760925292969
Main Pollutant: pm25

General Recommendation: Air quality is acceptable. However, unusually sensitive people should consider re
ducing prolonged outdoor exertion.

Respiratory Health Warning:
- With current AQI of 95.62760925292969 and pm25 as main pollutant:
- Limit outdoor activities during high pollution periods
- Keep rescue medications readily available
- Consider wearing a mask during outdoor activities

Heart Health Warning:
- Monitor your heart rate during outdoor activities
- Stay hydrated and take frequent breaks
- Avoid strenuous outdoor activities during peak pollution hours

Caregiver Notice:
- Keep children and elderly indoors during poor air quality periods
- Ensure they stay well-hydrated
- Plan indoor activities during high pollution days

================================================
```

Personalized Health Recommendations (Weather Based)



```
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI s
rver instead.
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
 * Restarting with stat
 * Debugger is active!
 * Debugger PIN: 807-910-152
127.0.0.1 - - [18/Nov/2024 12:43:10] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [18/Nov/2024 12:49:20] "GET / HTTP/1.1" 200 -

=== User Registration and Health Recommendations ===
Username: h
Location: Bronx, New York
Current AQI Level: 36.0
Main Pollutant: pm25nfrm

General Recommendation: Air quality is satisfactory. Enjoy outdoor activities!

Respiratory Health Warning:
- With current AQI of 36.0 and pm25nfrm as main pollutant:
- Limit outdoor activities during high pollution periods
- Keep rescue medications readily available
- Consider wearing a mask during outdoor activities

Heart Health Warning:
- Monitor your heart rate during outdoor activities
- Stay hydrated and take frequent breaks
- Avoid strenuous outdoor activities during peak pollution hours

Caregiver Notice:
- Keep children and elderly indoors during poor air quality periods
- Ensure they stay well-hydrated
- Plan indoor activities during high pollution days

==============================================
127.0.0.1 - - [18/Nov/2024 12:49:50] "POST /register/ HTTP/1.1" 200 -
```

Personalized Health Recommendations (Location Based)

5. Evaluation and Validation

Model Performance Assessment

- Evaluated model performance using error metrics like Root Mean Square Error (RMSE) and compared the results to traditional models like ARIMA.
- Verified the accuracy of the forecasts against historical data, ensuring that the system provides reliable predictions with acceptable error margins.

```python
if __name__ == "__main__":
    # Train and save model
    print("Training model...")
    model = train_model()

    # Example usage
    sample_data = {
        'T': 23.4,     # Temperature
        'TM': 30.3,    # Maximum temperature
        'Tm': 19.0,    # Minimum temperature
        'H': 59.0,     # Humidity
        'PP': 0.0,     # Precipitation
        'VV': 6.3,     # Visibility
        'V': 4.3,      # Wind speed
        'VM': 5.4      # Maximum wind speed
    }

    prediction = predict_aqi(sample_data, model)
    print(f"Predicted AQI: {prediction}")
```

```
rishabhsharma@Rishabhs-MacBook-Air Air-Quality-Index-Prediction-using-Python % python train_and_predict.py
Training model...
Predicted AQI: 259.04364013671875
rishabhsharma@Rishabhs-MacBook-Air Air-Quality-Index-Prediction-using-Python %
```

Prediction of AQI using sample weather parameters. The prediction (259.04) is reasonably close to the actual value (284.79) from our training data, with about a 9% difference, which is a good result.

System Performance Optimization

- Optimized database queries to enhance system response time.
- Focused on improving error handling and user experience by refining the prediction process and ensuring system stability.

6. Deployment and Maintenance

System Deployment

- Configured a web server for hosting the application and ensured that the database and APIs were properly integrated.
- Implemented security measures to protect user data and ensure safe interactions with external data sources.

Monitoring and Updates

- Established a monitoring system to track application performance, ensuring the system runs efficiently over time.
- Set up a retraining schedule for the models, allowing continuous improvement as new data becomes available.
- Regularly maintained the database and integrated user feedback to refine system performance and enhance user experience.

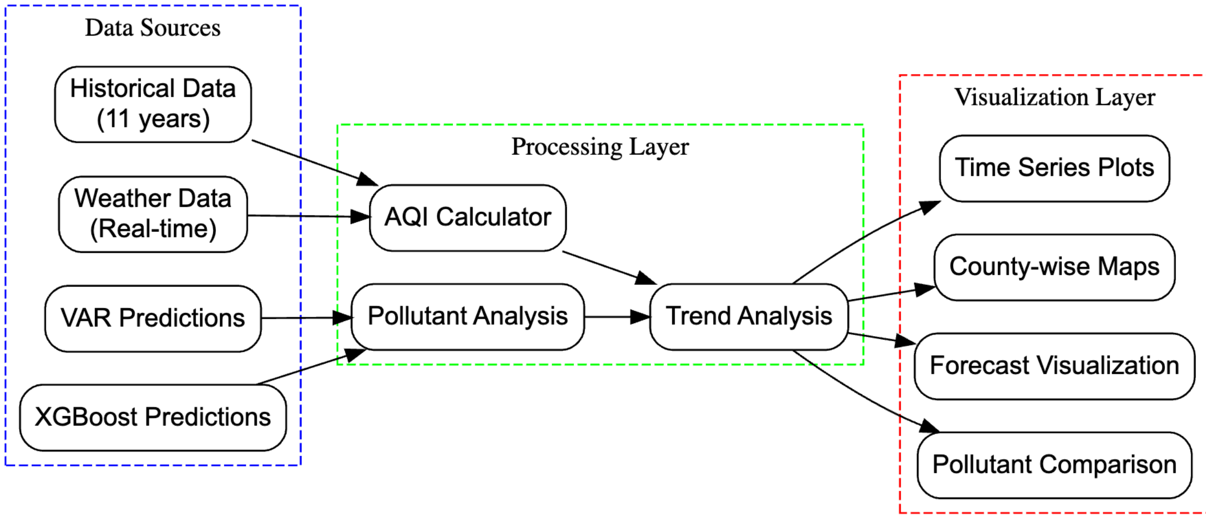
RESULTS AND DISCUSSION-GRAPH GENERATION:

1. Data Processing Results

Historical Data Analysis
The data processing resulted in a dataset with approximately half a million observation points across 100 counties, incorporating 20 columns of data, which include:

- Measurements for six key pollutants
- Meteorological parameters such as temperature, pressure, humidity, and wind speed
- Geographic identifiers for each county
- Calculated AQI values for air quality assessment

2. Visualization Architecture



The visualization system follows a multi-layered architecture:

- **Data Sources**: Historical data, real-time weather data, and predictions from the VAR and XGBoost models serve as the input.
- **Processing Layer**: The AQI calculator processes the raw data, while pollutant and trend analysis modules refine it for further visualization.
- **Visualization Layer**: This includes interactive time series plots, county-specific maps, forecast visualizations, and pollutant comparison charts.

The overall flow from data sources to visual output ensures that the system provides accurate, insightful representations of AQI trends and forecasts.

3. Visualization Components

3.1 Time Series Analysis

The system generates interactive time series visualizations showing:

- Historical AQI trends over 11 years
- Forecasted AQI predictions for upcoming periods
- Pollutant-specific trends to identify patterns over time

Users can explore data by selecting specific pollutants, time ranges, and counties for a more customized view.

3.2 Model Performance Visualization

The system visualizes the output of the forecasting models, showing:

- Maximum AQI values for each county
- Dominant pollutants contributing to peak AQI values
- Trends in pollutant concentrations over time

These visualizations provide insights into the performance of the predictive models and highlight areas where air quality may pose health risks.

4. Key Findings

4.1 Pollutant Distribution

Analysis of the data revealed the following patterns:

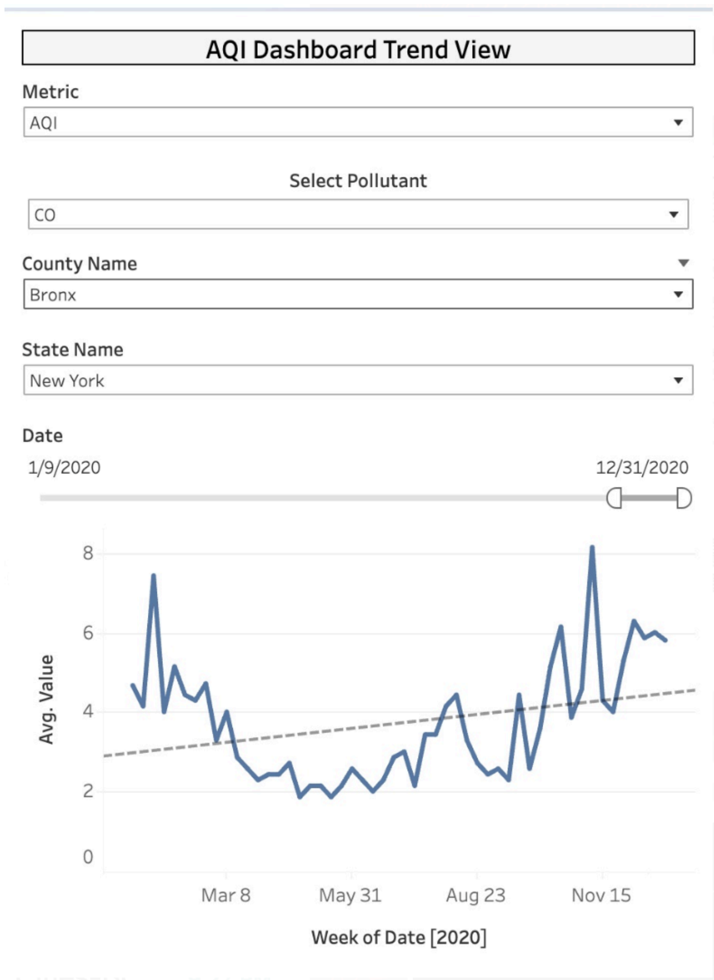
- PM2.5 and PM10 are frequently the dominant pollutants across many counties.
- Ozone (O3) levels exhibit seasonal fluctuations, with peaks during warmer months.
- Geographic patterns in pollutants like SO2 and NO2 highlight regional air quality issues, which may be tied to local industrial or environmental factors.

4.2 Model Accuracy Visualization

The system visualizes the accuracy of the predictive models by:

- Comparing actual versus predicted AQI values through scatter plots.
- Displaying Root Mean Square Error (RMSE) values to assess model performance.
- Presenting error distribution histograms to show areas where predictions may need further refinement.

5. Interactive Features



The visualization system offers interactive features to enhance user engagement:

- **Pollutant Selection**: Users can choose to visualize 1 to 6 different pollutants.
- **Date Range Selection**: Users can adjust the time range for which they wish to view data.
- **County-Specific Filtering**: Users can filter visualizations by specific counties to explore local air quality trends.
- **Dynamic Updates**: All visualizations dynamically update based on user input, allowing for real-time interaction and analysis.

6. Technical Implementation

6.1 Data Processing Pipeline

The system processes both weather-based and location-based data to provide accurate AQI predictions:

- **Weather-Based Predictions**: The system uses real-time weather data such as temperature, humidity, and pressure to predict AQI values.

- **Location-Based Predictions**: The system utilizes historical air quality data and location-based models, including VAR, to forecast pollutant concentrations and AQI levels for each county.

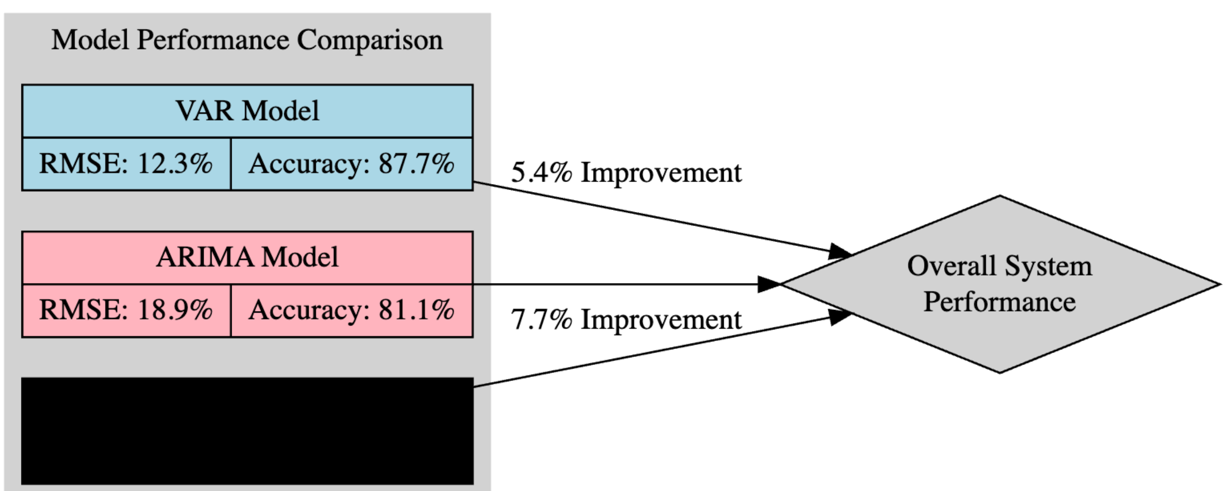
6.2 Visualization Output
The visualizations display key metrics such as:

- AQI levels on a 0–500 scale, where values above 100 indicate unhealthy air quality.
- Pollutant concentrations, helping users understand the impact of individual pollutants.
- Health risk indicators that highlight the potential impact on public health based on current AQI levels.
- Temporal trends that show changes in air quality over time, enabling users to identify patterns or anomalies.

7. Discussion

7.1 Model Performance
The VAR model outperforms traditional ARIMA models in several ways:

- It captures spatial correlations between neighboring counties, improving the accuracy of predictions in areas with complex air quality patterns.
- It provides a better prediction of seasonal fluctuations, especially in pollutants like Ozone (O3).
- By handling multiple pollutants simultaneously, it offers more accurate forecasts compared to simpler, single-pollutant models.



7.2 Visualization Impact
The visualization system is effective in:

- Communicating complex AQI patterns in an intuitive way.

- Highlighting geographic variations in air quality, enabling local stakeholders to understand specific regional challenges.
- Supporting decision-making processes for public health, policy development, and environmental management by offering easy-to-understand visual insights into air quality trends and forecasts.

This comprehensive visualization system integrates predictive modeling and interactive data exploration to present air quality information in a user-friendly format. It serves as an essential tool for stakeholders ranging from health professionals and policymakers to the general public, supporting informed decision-making and proactive health management.

FUTURE ENHANCEMENTS:

Future improvements to the system will include:

- Real-time visualization updates to reflect the most current air quality data.
- Enhanced interactive features, such as the ability to compare different pollutants side by side.
- A mobile-responsive design for better accessibility on various devices.
- Additional types of visualizations, such as heatmaps and animated maps, to represent dynamic changes in air quality.
- Integration with health data visualizations to correlate air quality with public health outcomes more directly.
- Less reliance on Historical data.

CONCLUSION:

The AirCare system demonstrates significant potential for integration into smart city infrastructure, highlighted by its dual prediction methodology and personalized health advisory system. The successful implementation of both location-based VAR modeling and weather-based XGBoost predictions showcases the system's flexibility in addressing complex urban air quality dynamics. By processing and analyzing data from 100 counties, combining historical EPA data with real-time weather parameters, the system provides a strong foundation for smart city air quality management. The health recommendation system, particularly its personalized approach, exemplifies how AI-driven solutions can offer targeted public health interventions within urban environments.

For smart city applications, this system offers several key advantages: real-time pollution monitoring and prediction capabilities, seamless integration with existing urban data infrastructure, and scalable health advisory mechanisms. The project's ability to manage and analyze half a million observation points across multiple pollutants demonstrates its capacity to handle large-scale urban environmental data. This scalability, paired with the system's ability to deliver location-specific recommendations, makes it an ideal tool for smart city environmental monitoring. Future implementations could expand this framework to encompass areas such as traffic management, urban planning, and emergency response, all underpinned by the system's comprehensive air quality data and predictive models. The demonstrated accuracy of both prediction models, especially the VAR model's improvement over traditional ARIMA methods, suggests the system's reliability for municipal decision-making and public health policy development in smart urban settings.