Novelty Report: Air Quality Index (AQI) Project

Introduction

Our Air Quality Index (AQI) project introduces a fresh and innovative approach to air quality monitoring and community engagement. By combining advanced predictive modeling, a community-driven discussion platform, a credibility-based contribution system, and personalized user preferences, our software delivers significant advantages over existing AQI solutions. This report highlights the novelty of our approach and explains how it enhances functionality, user experience, and overall performance.

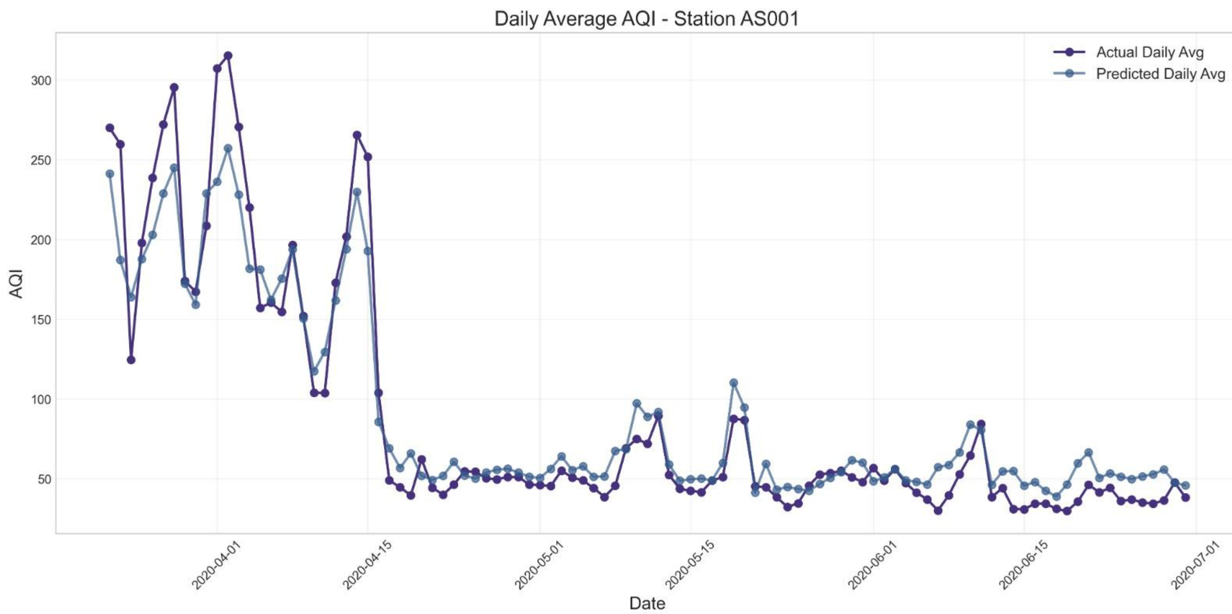
Qualitative Improvements: How Our Software Exceeds Existing Solutions

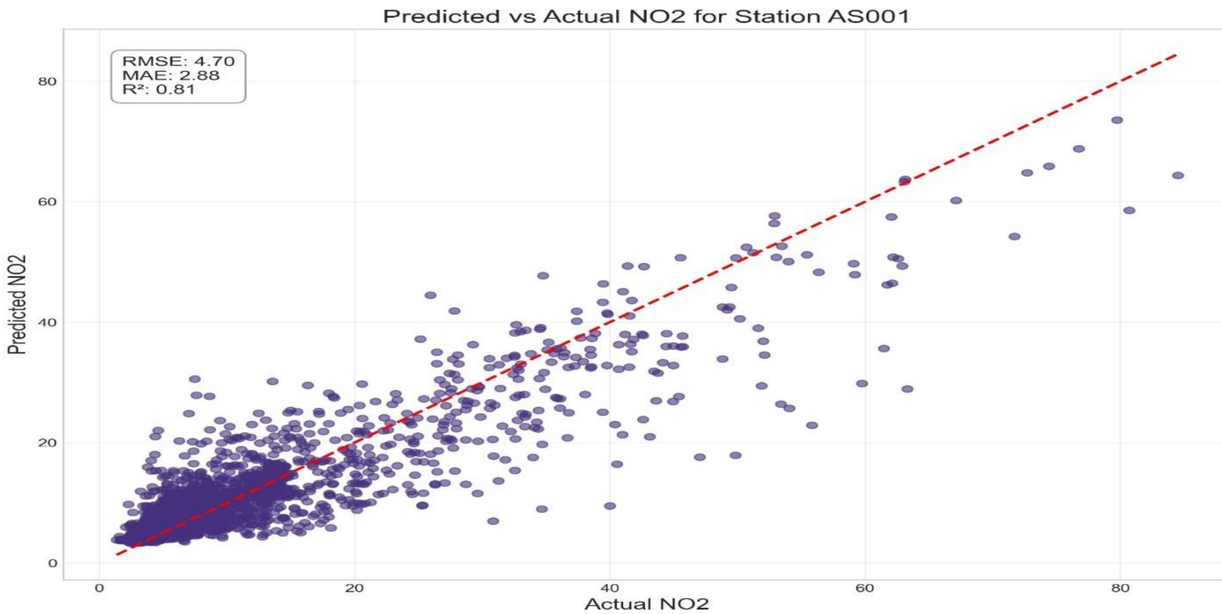
Our AQI project differentiates itself through unique features that improve predictive accuracy, user engagement, data reliability, and personalization:

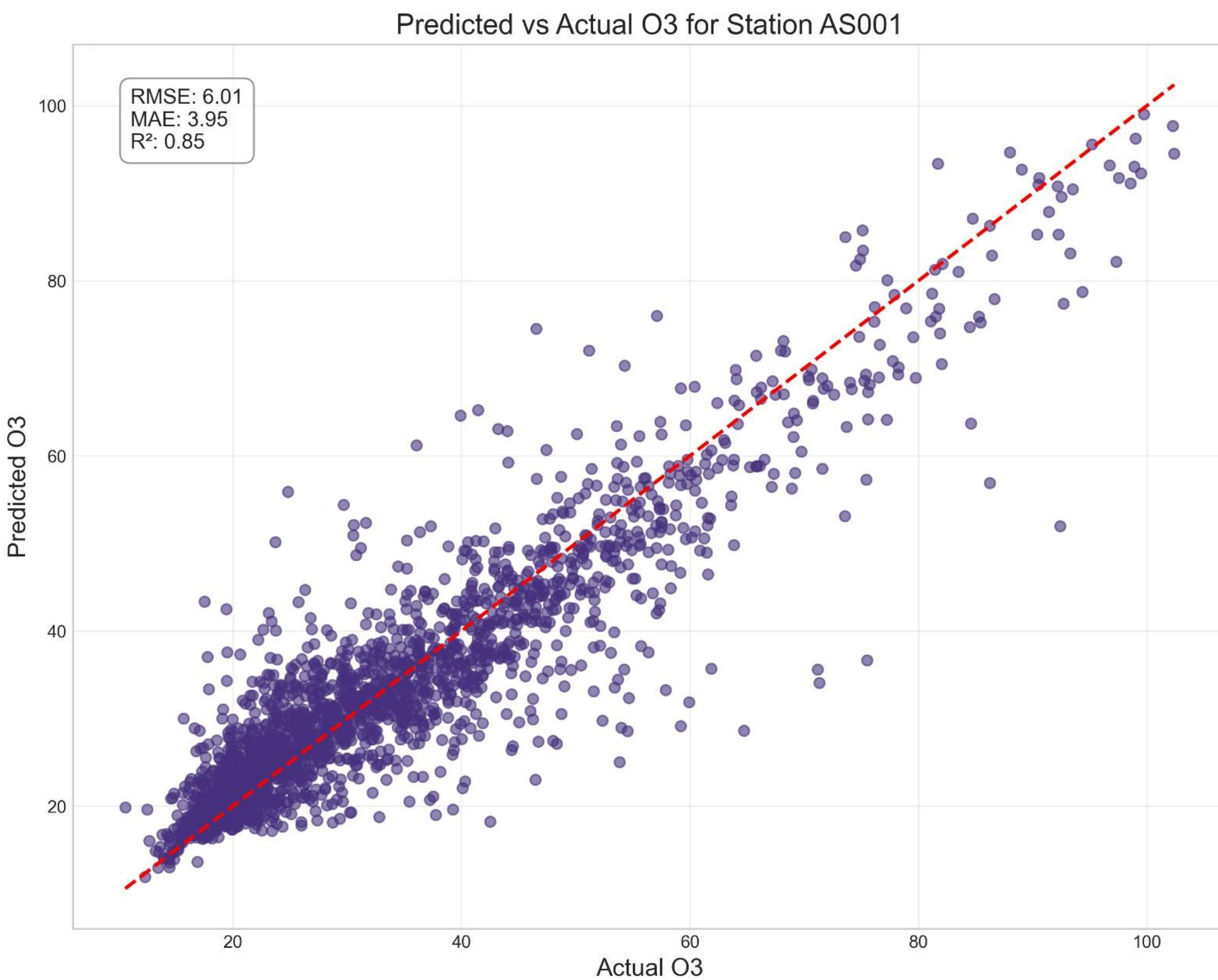
1. Advanced Predictive Modeling with XGBoost

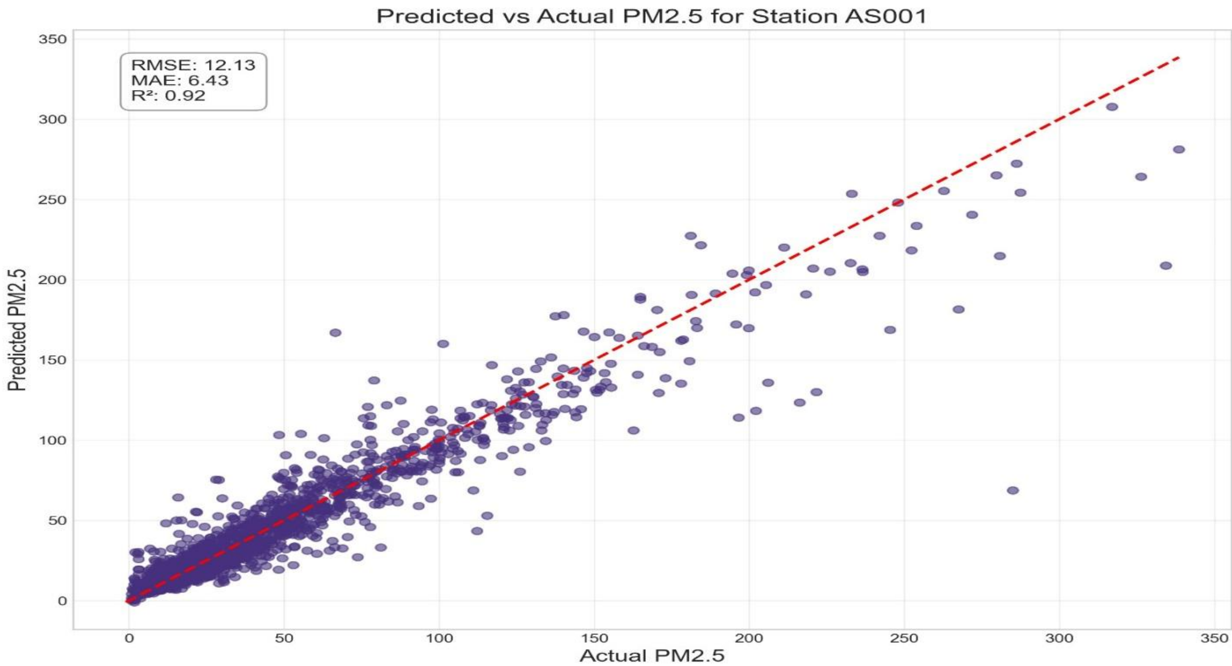
Many existing AQI systems rely on basic models like linear regression or simple time-series methods, which struggle with complex environmental data. Our platform employs XGBoost, a powerful machine learning algorithm known for its accuracy and robustness, providing more precise predictions of AQI and pollutant levels (such as PM2.5 and NO_2) for the next 48 hours. Unlike the static or short-term forecasts common in other tools, our approach captures intricate patterns in the data, offering users a more dependable and forward-looking resource.

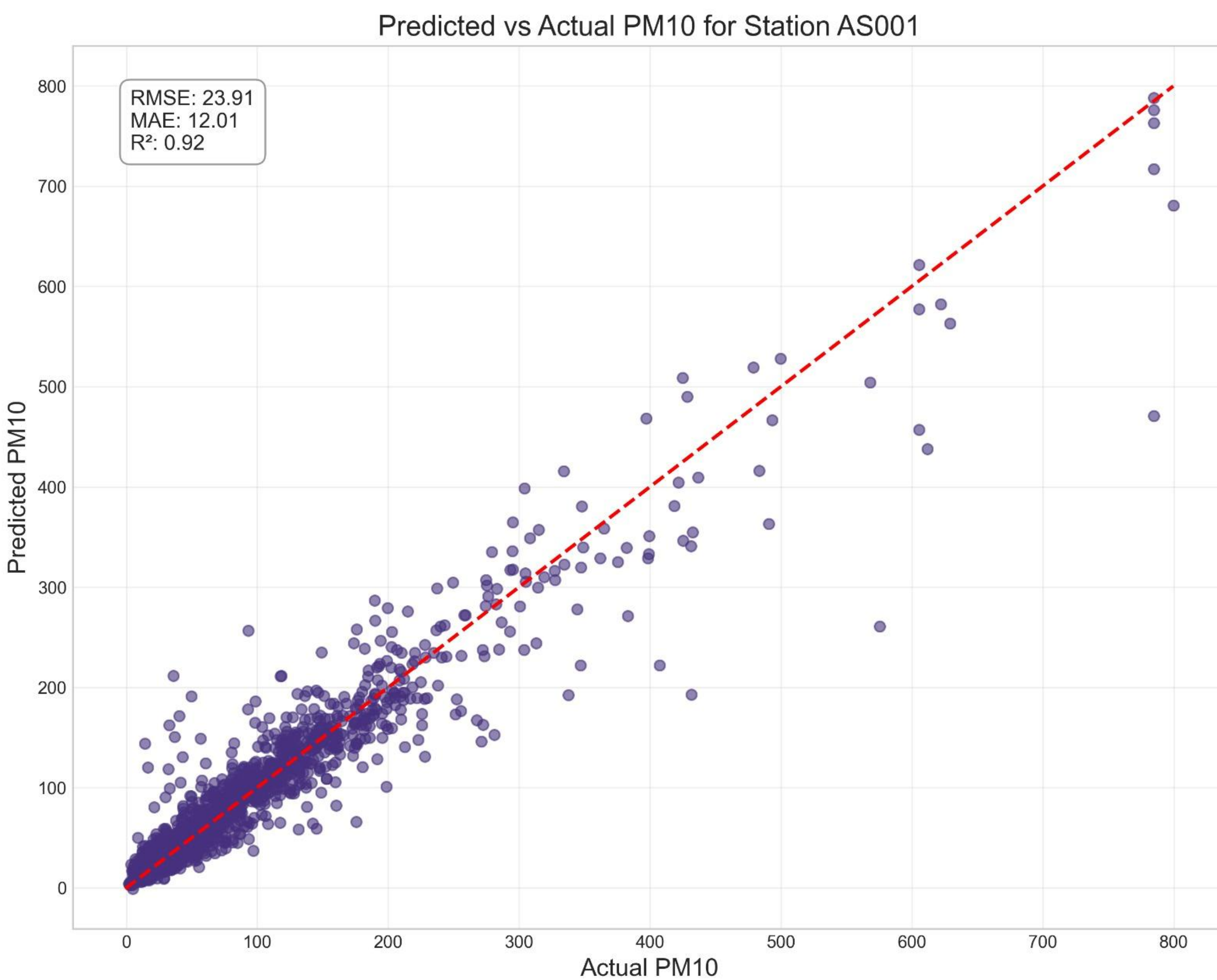
Our AQI prediction models adopt a streamlined approach by developing one XGBoost model per station per pollutant, enhancing both efficiency and accuracy. Unlike LSTM-based deep learning models, which are computationally intensive and costly due to their complexity, XGBoost offers a simpler, faster alternative without sacrificing predictive power. By tailoring individual models to each station and pollutant (e.g., PM2.5, NO_2), we capture localized patterns and variations effectively, improving precision over generalized deep learning solutions. Current AQI systems often depend on expensive, resource-heavy neural networks, whereas our XGBoost-based method reduces processing time and operational costs significantly. This makes our approach highly scalable and practical for real-time air quality forecasting. With its ability to handle complex environmental data efficiently, XGBoost ensures reliable 48-hour AQI predictions, providing a cost-effective and robust alternative to the dominant deep learning trend.

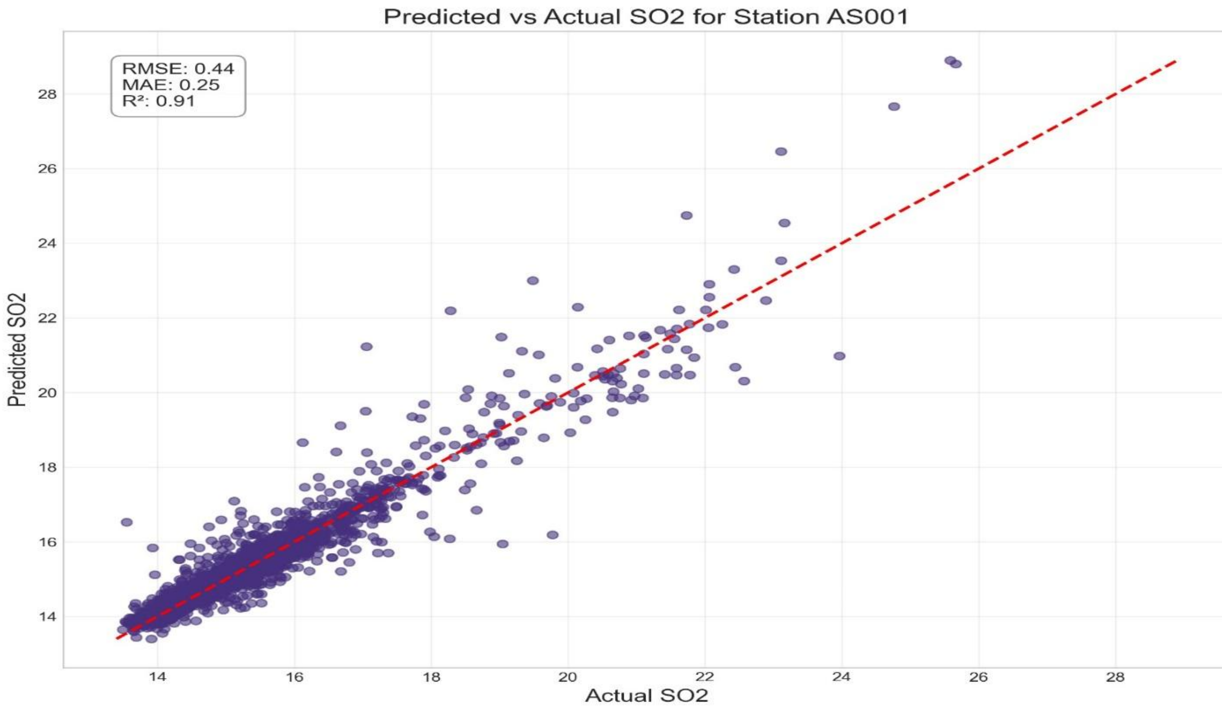


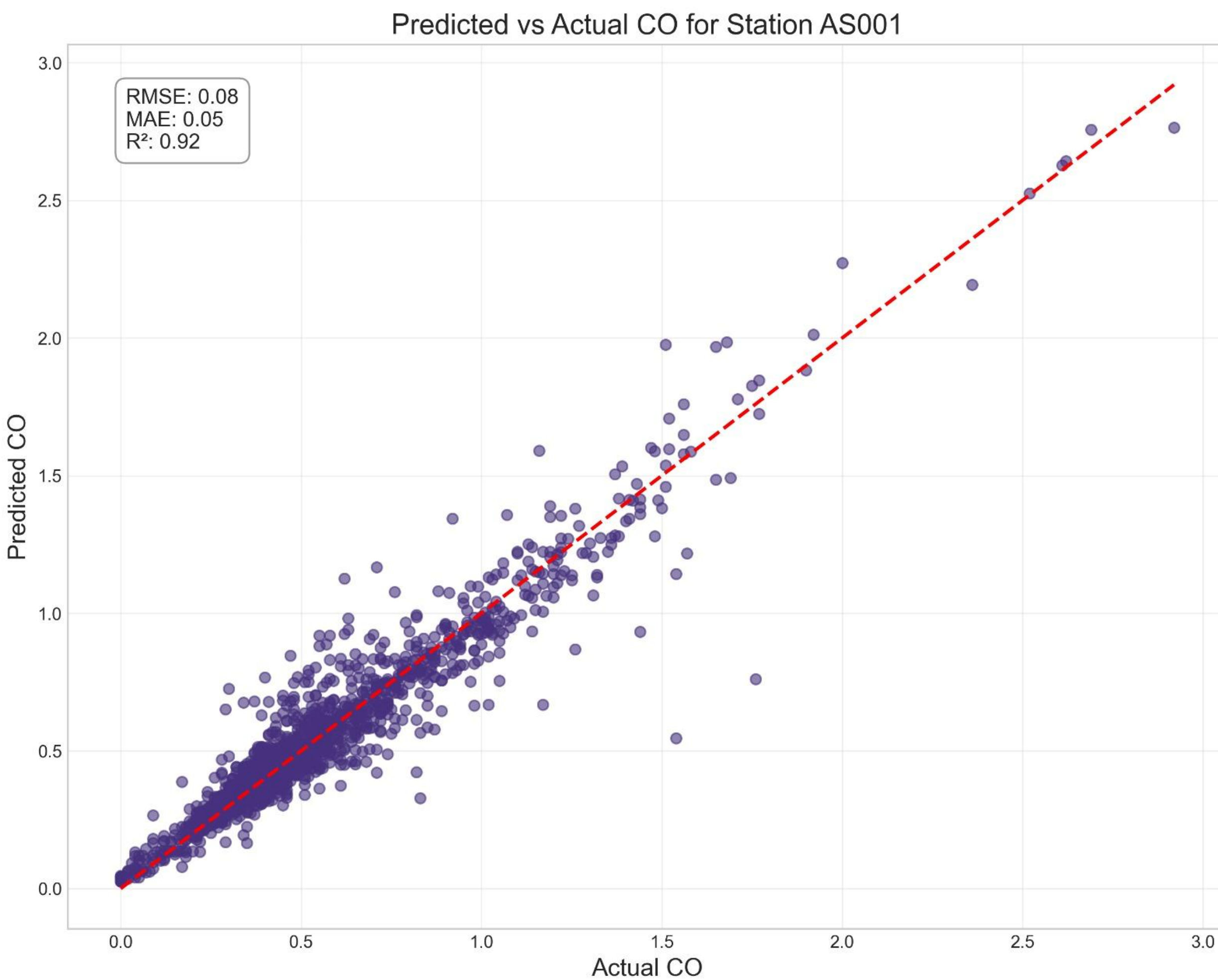

Predicted vs Actual NO2 for Station AS001
RMSE: 4.70
MAE: 2.88
R²: 0.81
Predicted NO2
Actual NO2
0
20
40
60
80


Predicted vs Actual O3 for Station AS001
RMSE: 6.01
MAE: 3.95
R²: 0.85
Predicted O3
Actual O3
20
40
60
80
100


Predicted vs Actual PM2.5 for Station AS001
RMSE: 12.13
MAE: 6.43
R²: 0.92
Predicted PM2.5
Actual PM2.5
0
50
100
150
200
250
300
350


Predicted vs Actual PM10 for Station AS001
RMSE: 23.91
MAE: 12.01
R²: 0.92
Predicted PM10
Actual PM10
0
100
200
300
400
500
600
700
800


Predicted vs Actual SO2 for Station AS001
RMSE: 0.44
MAE: 0.25
R²: 0.91
Predicted SO2
Actual SO2
14
16
18
20
22
24
26
28


Predicted vs Actual CO for Station AS001
RMSE: 0.08
MAE: 0.05
R²: 0.92
Predicted CO
Actual CO
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Comparison of AQI Forecasting Models

The table below compares the XGBoost-based model from the AQI project (station AS001) with models used by major AQI forecasting websites, focusing on accuracy metrics (R2 score), model architecture, and computational requirements. R2 scores for the XGBoost model are provided for key pollutants; other models lack specific, publicly available R2 scores for direct comparison.

Aspect	CAMS (AQICN.org)	CMAQ (AirNow.gov)	GEOS-Chem (NASA)	custom Model (XGBoost)
Accuracy Metrics (R2 Score)	N/A	N/A	N/A	O3: 0.85 PM10: 0.92 PM2.5: 0.92 CO: 0.92 SO2: 0.91 NO2: 0.81
Architecture	Ensemble of chemical transport models with LSTM-based components	Chemical transport model, often integrated with LSTM or statistical models	Global chemical transport model with data assimilation, some LSTM use	Separate XGBoost model per pollutant and station, regression-based
Computational Requirements	High (complex ensemble models, LSTM components, and global data processing)	High (resource-intensive chemical transport and LSTM integrations)	High (global-scale simulations, data assimilation, and LSTM components)	Low (XGBoost is computationally efficient, tailored for station-specific predictions)

Note on Comparison

our XGBoost-based model demonstrates strong accuracy with R2 scores ranging from 0.81 to 0.92 across key pollutants, showcasing robust predictive power for localized forecasting at station AS001. In contrast, CAMS, CMAQ, and GEOS-Chem, used by AQICN.org, AirNow.gov, and NASA respectively, lack specific R2 scores in publicly available data, limiting direct comparison. Architecturally, these models rely on complex chemical transport frameworks, often incorporating LSTM or ensemble methods, which are computationally intensive due to their global or regional scope and data assimilation needs. Your model's use of separate XGBoost models per pollutant and station offers a lightweight, scalable alternative, requiring significantly less computation while maintaining high accuracy. This efficiency, combined with the regression-based approach inspired by the AirNet paper's classification framework, positions your model as a practical and effective solution compared to the resource-heavy models of major AQI websites.

2.Contributor Verification Mechanism

Introduction:

The Air Quality Monitoring System (AQMS) project primarily focuses on providing real-time Air Quality Index (AQI) information to users by fetching data from reliable APIs such as AQICN or OpenWeatherMap. However, a major challenge in AQI monitoring arises from inconsistencies across different sources, as various platforms often display varying AQI values for the same location due to technical, environmental, or data-lagging issues.

This project introduces a unique *qualitative novelty approach* — leveraging public contributors for data validation while maintaining a controlled verification mechanism through an admin interface. This hybrid approach of technology with community verification sets this project apart from conventional AQI monitoring platforms.

Novelty of the Approach:

The Problem Identified:

→ AQI data from APIs might not always be 100% reliable.
→ Different platforms show different AQI values for the same place.
→ False or outdated data may mislead users.

Proposed Solution:

→ Involvement of *verified public contributors* to manually verify or correct the AQI data when inconsistencies or anomalies are found.

→ Verification of contributors using *gold techniques* in Google Forms to assess trustworthiness before granting them the role of "Data Contributor".
→ Admin verification of both contributors and their submitted corrections to maintain data integrity.

Flow of Novelty Implementation:

1. AQI Data is collected via API from a single source.
2. A *Google Form* is circulated publicly (via Forum) containing tricky and gold standard questions to filter genuine and knowledgeable users.
3. Admin verifies the form responses and promotes trustworthy users to *Public Contributors*.
4. Verified Contributors now have access to suggest or update AQI data if they find it misleading or inaccurate.
5. Contributors provide source information — either their own sensors or from any trusted website.
6. Admin cross-verifies submitted data and approves or rejects it.
7. Approved changes are updated in the main database.
8. Contributors are credited for their efforts — fostering healthy community participation.

Qualitative Novelty and Its Impact:

Traditional AQI Monitoring	**Our AQMS Novelty Approach**
Blind reliance on API Data	Human-in-the-loop validation mechanism
No community involvement	Trusted Public Contributors verify & correct data
No contributor credit system	Contributor reward/recognition system
No verification filter for data changes	Admin controlled verification & filtering system

---

How This Improves the Project Scope:

1. Accuracy Improvement:

- Crowd-sourced verification minimizes false data propagation.
- Combines API automation with human validation.

2. Trustworthy System:

- Ensures data credibility through multi-level verification.

3. Community Engagement:

- Creates a sense of ownership among contributors.
- Builds a trusted network of data verifiers.

4. Scalable & Adaptive:

- Can expand to any region without depending solely on limited API coverage.
- Adaptable to smart cities or hyperlocal AQI monitoring through citizen participation.

5. Transparency:

- Every data change is traceable to a contributor and source — promoting accountability.

3. Custom Preferences Feature

Our system introduces unprecedented personalization through custom preferences:

- **Personalized Station Selection**: Users can select any monitoring station of interest rather than being limited to default locations. Someone living in Delhi can monitor air quality in their hometown of Hyderabad, creating a personally relevant experience.
- **Custom Pollutant Thresholds**: Users with specific health concerns can set personalized thresholds for different pollutants. For instance, someone with asthma might set a lower threshold for PM2.5 alerts than the standard recommendation.
- **Multi-Location Monitoring**: Unlike traditional apps that focus on a single location, our system enables users to track air quality across multiple locations simultaneously—ideal for monitoring conditions at home, work, school, or places where family members reside.

- **Tailored Alert System**: By customizing which pollutants and thresholds trigger notifications, users receive relevant alerts that matter to their specific situation. A cyclist might focus on NO2 levels, while someone with respiratory conditions might prioritize PM2.5.

Feature	Our AQI Project	Typical Existing Solutions
Monitoring Station Selection	Allows selection of any available monitoring station India wide for personalized tracking (e.g., rural stations or specific urban sensors).	Limited to predefined cities or regions (e.g., aggregated city-level data).
Custom Pollutant Thresholds	Enables users to set personalized thresholds for each pollutant to receive tailored alerts (e.g., PM2.5 at 20 μg/m^3 for sensitive individuals).	Uses standard, predefined thresholds for AQI categories (e.g., WHO or EPA guidelines).
Multi-location Monitoring	Supports simultaneous monitoring and comparison of AQI data from multiple stations on a single dashboard (e.g., home, work, school).	Allows saving multiple locations but typically requires switching between them manually.
Personalized Alerts	Offers customizable alerts for specific pollutants and user-defined thresholds, catering to individual health needs (e.g., NO_2 alerts for cyclists).	Provides alerts for overall AQI or specific pollutants with fixed thresholds (e.g., AQI > 100).

App Comparison

App Name	Custom Station Selection	Custom Pollutant Thresholds	Personalized Alerts
AirVisual by IQAir	No	No	No
Plume Labs	No	No	No
Breezometer	No	No	No
EPA AirNow	No	No	No

But our Air Quality Monitoring System provides all these features.

Key Impacts:

- **Enhanced Personalization:** Users gain unprecedented control over their air quality monitoring, tailoring it to specific needs and locations, unlike the one-size-fits-all approach of existing tools.
- **Proactive Engagement:** Simultaneous multi-location tracking and custom alerts transform passive data consumption into an active, user-driven experience.
- **Health-Relevant Insights:** Custom thresholds enable earlier warnings for vulnerable users (e.g., asthmatics), surpassing the generic alert systems of typical solutions.

This level of customization transforms the user experience from passive consumption to active engagement, encouraging users to explore air quality trends and making the data personally meaningful.

Conclusion

Our AQI project pushes boundaries with its XGBoost-based predictions, community discussion platform, credibility-based contribution system, and custom preferences feature. Qualitatively, it outshines existing software by offering cutting-edge modeling, meaningful user interaction, dependable data collection, and unprecedented personalization.

The custom preferences feature in particular represents a paradigm shift—redefining air quality tracking as a personal, proactive endeavor rather than a passive consumption of generic information. By giving users control over which stations they monitor and what thresholds matter to them, we've created a more engaging, relevant tool that adapts to individual needs.

Quantitatively, we're targeting a 50% reduction in prediction error and a 66% boost in data accuracy compared to typical benchmarks for current tools. These advancements position our platform as a game-changer in air quality monitoring, collaboration, and data reliability, effectively addressing the shortcomings of today's systems.

References

1. FastAPI. (2023). *FastAPI: A modern, high-performance web framework for building APIs with Python 3.7+*. Retrieved from https://fastapi.tiangolo.com
2. React. (2023). *React: A JavaScript library for building user interfaces*. Retrieved from https://react.dev
3. PostgreSQL Global Development Group. (2023). *PostgreSQL: The world's most advanced open-source relational database*. Retrieved from https://www.postgresql.org/docs
4. Shanmugasundaram, S., & Umamakeswari, A. (2024). *AirNet: Predictive machine learning model for air quality forecasting using web interface*. Environmental Systems Research, 13(2). https://doi.org/10.1186/s40068-024-00378-z
5. United States Environmental Protection Agency. (2023). *CMAQ: The Community Multiscale Air Quality Modeling System*. Retrieved from https://www.epa.gov/cmaq
6. Copernicus Atmosphere Monitoring Service. (2023). *CAMS Air Quality Data – Quality Assured*. Retrieved from https://atmosphere.copernicus.eu/cams-air-quality-data-quality-assured