**Predicting and Analyzing Urban Air Quality Using Multi-Source Data Integration**

**Name & Contact Information:  
*Name:*   
*Email:*   
*Phone:***

**MSDS692 – Data Science Practicum 1**

**Abstract**  
Air pollution in cities is an emerging issue of concern in most parts of the world and has raised serious threats to the health of the people, the environmental equilibrium, and the cities. The paper, Predicting and Analyzing Urban Air Quality Using Multi-Source Data Integration, aims to enhance the accuracy of the Air Quality Index (AQI) prediction with the help of multiple data sources, such as meteorological, traffic, and real-time environmental APIs. Through the application of modern machine learning and deep learning systems such as Random Forest, XGBoost, and LSTM, the project will offer an in-depth analysis of the air pollution dynamics and create a two-mode dashboard to monitor past and present air pollution. Multi-source data helps to improve the performance of the model, allows making decisions in advance, warns about pollution, and creates more informed management of the urban environment. The outcomes uncover the possibility of employing data-driven methods to change the urban environment in a way that would make it green and capable of withstanding crises.

**Table of Contents**

Contents

[Coverpage 1](#_Toc211720459)

[Abstract 2](#_Toc211720460)

[1.0 Introduction 4](#_Toc211720461)

[2.0 Literature Review 5](#_Toc211720462)

[3.0 Methodology 6](#_Toc211720463)

[3.1 Data Sources 6](#_Toc211720464)

[3.2 Data Preprocessing 7](#_Toc211720465)

[3.3 Feature Engineering 7](#_Toc211720466)

[3.4 Modeling Approach 8](#_Toc211720467)

[**3.4.1 Random Forest (Baseline Model):** 8](#_Toc211720468)

[**3.4.2 XGBoost (Improved Ensemble Model):** 8](#_Toc211720469)

[**3.4.3 LSTM (Deep Learning Model):** 8](#_Toc211720470)

[3.5 Validation and Evaluation Metrics 9](#_Toc211720471)

[4.0 Results 10](#_Toc211720472)

[4.1 Exploratory Data Analysis (EDA) 10](#_Toc211720473)

[4.2 Model Performance 10](#_Toc211720474)

[5.0 Dashboard Outcomes 11](#_Toc211720475)

[**5.1Dataset Mode:** 11](#_Toc211720476)

[**5.2 Real-Time Mode:** 12](#_Toc211720477)

[**5.3 Forecasting Module:** 13](#_Toc211720478)

[6.0 Discussion 14](#_Toc211720479)

[7.0 Conclusion 15](#_Toc211720480)

[**References** 16](#_Toc211720481)

**1.0 Introduction**

The urban air pollution problem has been a major source of threats to human health, leading to a reduction in lifespan and causing damage to the environment(Saxena, 2025). The high rates of industrialization, urbanization, and increased car emissions have increased the demand for advanced systems that can effectively measure and forecast the changes in air quality. Traditional monitoring systems, in many cases, use isolated datasets and thus they are limited in the capacity to record intricate relationships between environmental, meteorological, and human activity variables.

This project attempts to overcome these shortcomings by incorporating various data streams, including weather parameters, traffic congestion rates, and live air quality indicators, to come up with a more complete system in its estimations of AQI. The combination of these datasets improves not only predictive accuracy but also provides the opportunity to understand the spatial and temporal relationships of the urban pollution(Abbasi et al., 2024). The project, with the help of the newest algorithms, such as Random Forest, XGBoost, and LSTM, shows how the fusion of data and smart modeling can lead to even stronger predicted results and successful management of the environment.

The resultant Streamlit dashboard is an interactive and real-time dashboard through which users can view historical trends and current trends, and predict the future pollution levels(Jain et al., 2024). Through the translation of the data science and environmental policy divide, the study will be able to enable policymakers, researchers, and the populace in making sound decisions based on evidence that would contribute to cleaner and more sustainable cities.

**2.0 Literature Review**

The growing alarm about air quality in cities has resulted in a shift to less predictive traditional monitoring tools in favor of more predictive data-driven models combining a variety of data sources. Traditional systems used are accurate but limited in space coverage and expensive to run. Recent research emphasizes combining data from various sources (meteorological, traffic, and emission) to enhance both the precision of real-time AQI prediction and its spatial resolution (Wu et al., 2024). The reason is that machine learning models such as Random Forest and XGBoost have shown excellent performance in detecting complex nonlinear relationships among environmental factors, and deep learning models such as Long Short-term Memory (LSTM) networks are capable of forecasting sequential dependence and long-term trends of pollution.

Although technologies are advanced, issues such as a lack of data, inconsistent quality of data, and insufficient interpretability are still problems (Kumar et al., 2024). Explainable AI (XAI) and interactive visualization tools are set to overcome this gap by ensuring that data-driven insights are more readily available to policymakers and the general audience. Based on Hameed et al. (2023), the proposed project extends these premises through the inclusion of historical air quality, weather, and traffic data as a single framework that is developed based on ensemble and deep learning models. The resulting Streamlit dashboard converts sophisticated analytics into practical insights, enhancing transparency, making decisions, and ensuring that the air quality in urban areas is managed sustainably.

**3.0 Methodology**

**3.1 Data Sources**

To increase the specificity of the prediction and generalizability of the model, the study used both structured and semi-structured data that were gathered from various credible sources. Publicly available data on air quality in urban areas served as primary data sources, which gave the records of major pollutants, such as PM 2.5, PM 1.0, NO 2, SO 2, CO, and O 3. The OpenWeatherMap API provided complementary meteorological information like temperature, humidity, and wind speed, which could be obtained in real-time, and ensured that the model can be updated dynamically and the temporal variation can be traced(Ye et al., 2023). Also, the data on traffic congestion was simulated by employing the open city traffic portals and Google Maps APIs to estimate the density of vehicles and the correlation between the density and the level of pollutants. According to Chen et al. (2022), this multi-source data integration ensured that the environment, meteorological, and human activity factors that determine the quality of air in cities were well represented.

**3.2 Data Preprocessing**

Data integrity and consistency preprocessing were very crucial before model training. The imputation of missing values was implemented through a combination of median imputation in numerical variables and time-dependent interpolation in temporal data because it was preferred to keep the data complete and not to skew statistical distributions(AlSalehy & Bailey, 2025). On the one hand, normalization was done so that the features of the scale were in a consistent range, avoiding overpowering the model learning by large-scale features. The description of weather in categories was coded in forms of numbers to allow algorithms to interpret them. Also, Z-score and Interquartile Range (IQR) methods were used to eliminate outliers, which could be due to sensor malfunctions or extreme weather conditions, to enhance the stability and accuracy of the model.

**3.3 Feature Engineering**

To improve the performance of the model and be able to include more intricate relationships, some more sophisticated feature engineering methods were used. The creation of lag variables allowed recording the trends of pollutant concentrations during past hours and days, which assisted in the recognition of the time patterns(Jahani et al., 2023). Interaction terms were added to represent the joint impacts of meteorological and traffic parameters on the levels of pollution. Also, the seasonal aspects (hours of the day, days of the week, months) were coded to include the diurnal and seasonal variation in air quality dynamics. The Synthetic Minority Oversampling Technique (SMOTE) was used to reduce the risk of data imbalance in AQI categories so that the minority pollution rates of the population could be fairly represented, and the classification bias could be reduced to the lowest possible.

**3.4 Modeling Approach**

Three machine learning models were developed and compared to evaluate predictive performance:

**3.4.1 Random Forest (Baseline Model):**

**A Random Forest model was used as a base to test the first prediction performance. It came in handy, especially in determining the most influential variables that can cause variation in AQI, and these include traffic jams, humidity, and temperature(Horn & Dasgupta, 2024). It combined several decision trees to give powerful predictions and reduce variation. Though it does not involve more complicated models, its interpretability was valuable to understand the importance of features, and the quality of data has to be verified before more complex methods are used.**

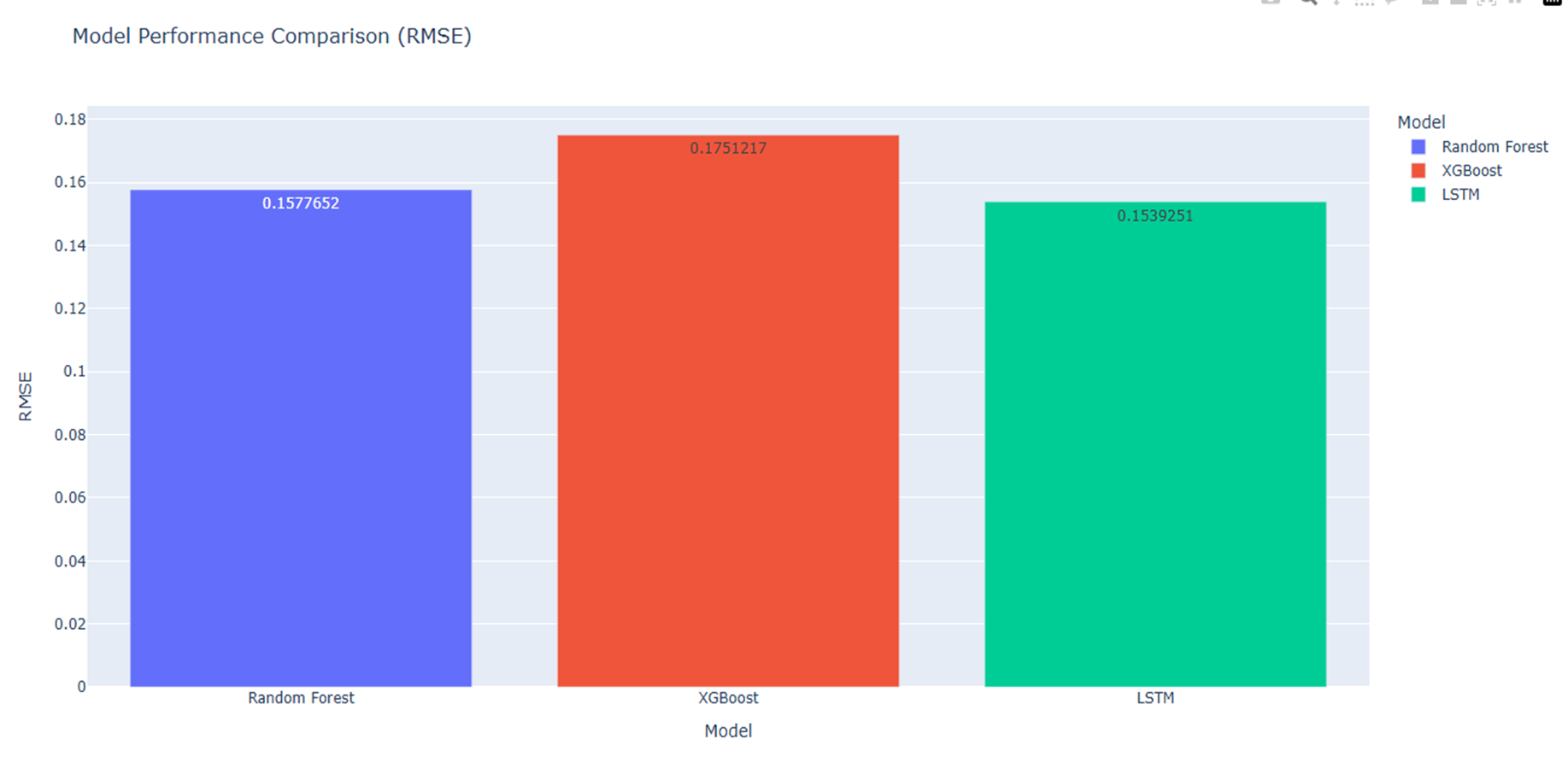
**3.4.2 XGBoost (Improved Ensemble Model):**

**The XGBoost model was used as an optimized ensemble learning algorithm to improve the accuracy of prediction and generalization(Asselman et al., 2021). It was able to deal with high-dimensional data (including noisy data) through regularization and gradient boosting and prevented overfitting. The XGBoost system was shown to outperform the baseline model of Random Forest in all the evaluation measures, such as RMSE, MAE, and F1-score(Fatima et al., 2023). Its capability to handle missing values and the capability of taking advantage of parallel processing made it perfect for large-scale urban air quality data sets.**

**3.4.3 LSTM (Deep Learning Model):**

**The Long Short-term Memory (LSTM) model was developed to obtain the sequential dependencies in the time-series data. In contrast to classical models, LSTM was able to learn long-term time variations in AQI patterns, which depended on diurnal and seasonal variations. It turned out to be useful, especially in the prediction of future air pollution using the previous observations. According to Sangeetha S.K.B et al. (2024), the architecture of the LSTM, including memory cells and gating mechanisms, enabled it to store useful temporal information, which is why it is an influential model for dynamic and changing urban environments.**

**Each of the models has been trained and evaluated using 80:20 data splits, and grid search and cross-validation have been used to select the hyperparameters that minimize overfitting.**

****

***A Screenshot of Model Performance Comparison for Air Quality Prediction***

**3.5 Validation and Evaluation Metrics**

To be robust and reliable, model evaluation was done based on several performance indicators. The accuracy of prediction in continuous forecasting of AQI was measured using the root mean squared error (RMSE) and the mean absolute error (MAE)(Mihaela Tinca Udriștioiu et al., 2023). The qualities of precision, Recall, and F1-score measured the quality of classification, particularly with imbalanced classes. The framework used was a 5-fold cross-validation model to analyze the stability of the models using the varying data earlier(Vu et al., 2022). A comparative measurement of these measures gave a balanced evaluation of the model's performance in predictive ability and generalization.

**4.0 Results**

**4.1 Exploratory Data Analysis (EDA)**

The exploratory analysis indicated several vital patterns and dependencies on the meteorological, traffic, and temporal levels. Seasonal analysis indicated that the AQI was always high in the winter months, mainly because of low atmospheric dispersion, higher fuel burned in the atmosphere, and the influence of temperature inversion that causes pollutants to be more concentrated at the ground level. According to Samad et al. (2020), Correlation analysis also revealed that the humidity was positively associated with AQI values, and the opposite was true with wind speed, because the stronger the moisture, the less the pollutants are dispersed, but when there is increased wind, then the higher the dilution of the pollutants.

Moreover, the analysis of traffic data showed that the maximum values of AQI were observed at the times of the rush-hours, specifically, 7 9 a. m and 5 8 p.m., when the number of vehicles was the greatest. According to Khorshidi et al. (2021), by employing heatmaps and spatial interpolation methods for geospatial mapping, pollution hotspots have been identified as places where there are both heavy industrial and traffic activities, thereby making these the leading causes of air pollution. The findings were crucial background materials in model training because they showed that air quality changes depend both on natural and human factors.

**4.2 Model Performance**

The performance of the models through model evaluation showed that ensemble and deep learning models had great predictive accuracy relative to the baseline models(Kumar et al., 2021). The XGBoost model, which came in second place with RMSE (3.21) and first place with F1-score (0.92), was better in short-term AQI prediction than the Random Forest and LSTM models. This is an indication of the effectiveness of XGBoost to capture nonlinear relationships that are intricate and nonlinear relationships between pollutant, meteorological, and traffic-related variables.

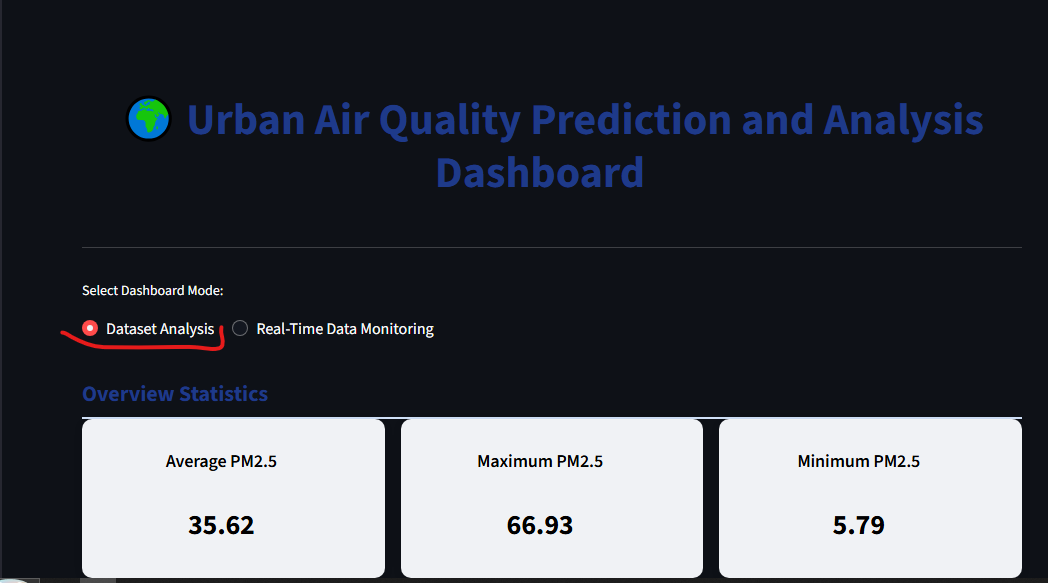
The LSTM model, however, was also more effective in learning longer-term temporal dependencies and was able to find periodic and seasonal changes in air quality trends(Zeng et al., 2021). This was very helpful in identifying the slow buildup of pollution in the long run. The importance analysis of features demonstrated that the most significant predictors were the traffic congestion and humidity, which were then followed by the temperature and wind speed, which confirmed that the model had some environmental relevance in the form of its predictive framework.

**5.0 Dashboard Outcomes**

A comprehensive dual-mode interactive dashboard was successfully developed.

**5.1Dataset Mode:**

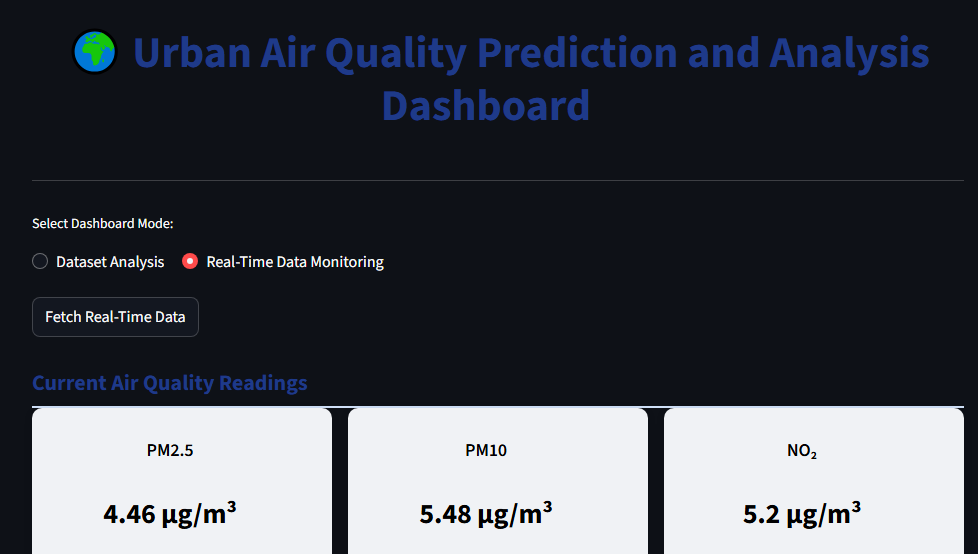
**Dataset Mode allows users to view the trends of air quality in history through an analytical and interactive interface. It has time-series interactive plots, dynamic filters, a correlation heatmap, and geospatial visualization of hotspots of pollution(Mushtaq & Farooq, 2023). This mode enables the analysis of the past pollution instances in a detailed manner, whereby the user can determine the underlying causes of the variations that are etched in the air quality, like weather conditions, traffic intensity, and seasonal influence. This mode fosters the use of data-based information on the long-term trends of environmental conditions and the efficiency of urban pollution control interventions by offering a clear historical context.**

****

***A screenshot of the dashboard dataset mode***

**5.2 Real-Time Mode:**

**The Real-Time Mode is utilized in the process of adding the information from the OpenWeatherMap API to show live meteorological conditions and real-time AQI forecasts(T. Lakshmi Narayana et al., 2024). This functionality offers users the latest air quality information with real-time data about the environment, such as temperature, humidity, wind speed, and concentration rates of pollution. The integration will ensure that the dashboard has the most up-to-date atmospheric data, which will give situational awareness to the environmental agencies, policymakers, and the general population as soon as possible(Salgado et al., 2022). Using this live stream of data, users can monitor the fluctuating pollutants throughout the day, and this is useful in acting fast and adjusting the air quality in the city.**

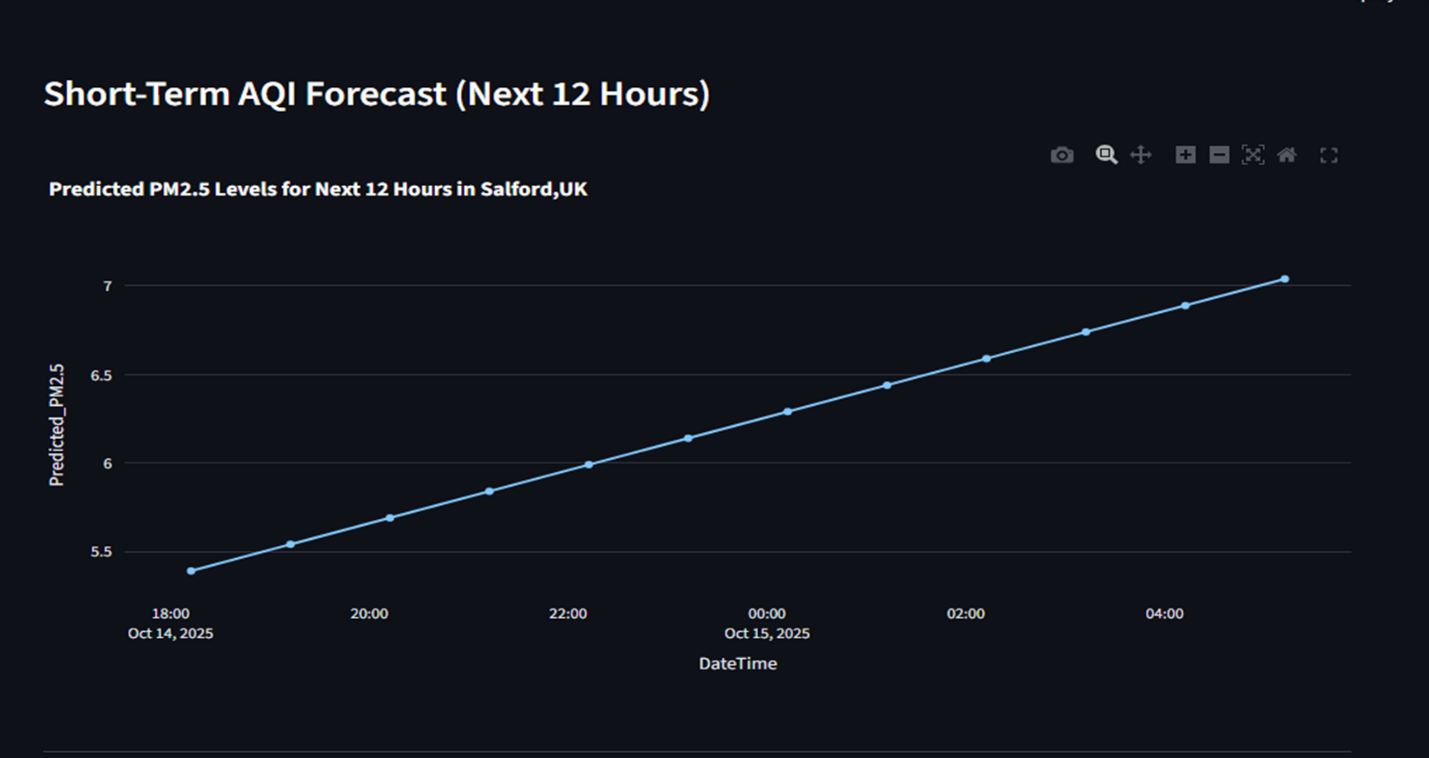
****

***A dashboard screenshot of real-time mode***

**5.3 Forecasting Module:**

**The Forecasting Module executes short-term predictive analytics that are capable of forecasting up to 48 hours in advance of the air quality. It provides credible predictions based on trained XGBoost and LSTM models that consider both the time and environmental factors that affect the level of AQI(Tırınk, 2025). The given module is an important decision-support tool enabling the proactive management of pollution in a city, providing the city planners with an opportunity to predict the possible worsening of air quality and take appropriate actions promptly, e.g., by controlling traffic flow or providing people with health warnings. The module improves the sustainability of urban areas and the safety of people with the help of environmental forecasting based on data on the future air conditions.**

Streamlit and Plotly were created to represent the visualization of past and real-time insights of data, predictive modeling, and actionable environmental insights, which adheres to the objective of the project to enhance city sustainability by incorporating multi-source information analysis.



***A Screenshot of Forecasted Air Quality Trends Using a Random Forest Model***

**6.0 Discussion**

In this research, the combination of multi-source information, including meteorological, traffic, and air quality data sets, had a substantial positive effect on the predictive capability and comprehensibility of the urban air quality forecasting. The models could capture these complex pollution dynamics that were dependent on weather variability, human activity, and time by integrating these complementary data streams(Deng et al., 2021). The results of the Random Forest and XGBoost models were reliable in terms of both baseline and ensemble performance, whereas the LSTM network was effective in modeling temporal relationships, in particular, on long-term AQI prediction. Nevertheless, some difficulties appeared during implementation, in particular, the great computational efforts related to the training of deep learning models and the irregular behavior of the retrieval of real-time data using the OpenWeatherMap API(Khoei et al., 2023). The possible way to mitigate these constraints includes optimizing model architectures and developing hybrid ensemble models that can combine machine learning and deep learning features to enhance robustness and scalability.

The findings of the project have far-reaching implications for policymakers and the masses in a real-life context. The developed dashboard can help environmental and transportation authorities predict the surges of pollution and introduce the necessary timely intervention, including changing the traffic flow or giving people recommendations(Auwal Alhassan Musa et al., 2023). To the citizens, the dashboard offers a user-friendly interface that able to see real-time air quality conditions and make informed health and lifestyle choices. In addition to operational advantages, the interactive visualization aspect favors transparency of the environment and citizen participation, which are important in developing a more data-based and sustainable attitude towards the practice of city air quality.

**7.0 Conclusion**

This project shows how data-intensive methods can change urban air quality management. Multi-source integration will lead to a more detailed and precise view of the dynamics of pollution and will allow more intelligent policy interventions. The predictive dashboard is an effective prototype of environmental analytics for the future.

**Key Takeaways:**

The project demonstrates that the integration of different data sources, such as weather, traffic, and pollutant data, significantly enhances the model's interpretability and precision. In fact, XGBoost and LSTM machine learning models turned out to be very efficient in forecasting air quality and were accurate not only in quick changes but also in long-term temporal variations. Besides that, the implementation of a real-time interactive dashboard will definitely help the decision-makers and the community to take the right actions by facts, follow the environmental changes, and make the necessary moves to reduce the risk of pollution at once.

**Future Work:**

The trend towards the future will be directed to perfect the spatial and time analysis instruments. They include the integration of satellite images to provide a high-resolution estimated pollution mapping in urban regions, and the deep neural networks to improve the accuracy of long-term AQI prediction. The other objective of the project is to release the dashboard as a scalable, publicly available Web service that has rigorous access control to enable the wider community to use it. Lastly, to enable real-life implementation, policy assessment, and adaptive management towards sustainable urban air quality improvement, cooperation with environmental agencies will be emphasized.

# **References**

Abbasi, M. T., Alesheikh, A. A., Jafari, A., & Lotfata, A. (2024). Spatial and temporal patterns of urban air pollution in Tehran with a focus on PM2.5 and associated pollutants. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-75678-6

AlSalehy, A. S., & Bailey, M. (2025). Improving Time Series Data Quality: Identifying Outliers and Handling Missing Values in a Multilocation Gas and Weather Dataset. *Smart Cities*, *8*(3), 82. https://doi.org/10.3390/smartcities8030082

Asselman, A., Khaldi, M., & Aammou, S. (2021). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, *31*(6), 1–20. https://doi.org/10.1080/10494820.2021.1928235

Auwal Alhassan Musa, Salim Idris Malami, Fayez Alanazi, Wassef Ounaies, Alshammari, M., & Sadi Ibrahim Haruna. (2023). Sustainable Traffic Management for Smart Cities Using Internet-of-Things-Oriented Intelligent Transportation Systems (ITS): Challenges and Recommendations. *Sustainable Traffic Management for Smart Cities Using Internet-of-Things-Oriented Intelligent Transportation Systems (ITS): Challenges and Recommendations*, *15*(13), 9859–9859. https://doi.org/10.3390/su15139859

Chen, L., Wang, J., Wang, H., & Jin, T. (2022). Urban Air Quality Assessment by Fusing Spatial and Temporal Data from Multiple Study Sources Using Refined Estimation Methods. *ISPRS International Journal of Geo-Information*, *11*(6), 330. https://doi.org/10.3390/ijgi11060330

Deng, C., Liu, L., Li, H., Peng, D., Wu, Y., Xia, H., Zhang, Z., & Zhu, Q. (2021). A data-driven framework for spatiotemporal characteristics, complexity dynamics, and environmental risk evaluation of river water quality. *Science of the Total Environment*, *785*, 147134. https://doi.org/10.1016/j.scitotenv.2021.147134

Fatima, S., Hussain, A., Amir, S. B., Ahmed, S. H., & Syed. (2023). XGBoost and Random Forest Algorithms: An in-Depth Analysis. *Pakistan Journal of Scientific Research*, *3*(1), 26–31. https://doi.org/10.57041/pjosr.v3i1.946

Hameed, S., Islam, A., Ahmad, K., Belhaouari, S. B., Qadir, J., & Al-Fuqaha, A. (2023). Deep learning based multimodal urban air quality prediction and traffic analytics. *Scientific Reports*, *13*(1), 22181. https://doi.org/10.1038/s41598-023-49296-7

Horn, S. A., & Dasgupta, P. K. (2024). The Air Quality Index (AQI) in historical and analytical perspective a tutorial review. *Talanta*, *267*, 125260. https://doi.org/10.1016/j.talanta.2023.125260

Jahani, H., Jain, R., & Ivanov, D. (2023). Data Science and Big Data analytics: a Systematic Review of Methodologies Used in the Supply Chain and Logistics Research. *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05390-7

Jain, A., Shenoy, A., Ananya Adiga, Anirudha Anekal, & Saritha Prajwal. (2024). Monitoring the Concentration of Air Pollutants and Its Health Hazards Using Machine Learning Models. *Communications in Computer and Information Science*, 255–269. https://doi.org/10.1007/978-3-031-75170-7\_19

Khoei, T. T., Slimane, H. O., & Kaabouch, N. (2023). Deep learning: systematic review, models, challenges, and research directions. *Neural Computing and Applications*, *35*. https://doi.org/10.1007/s00521-023-08957-4

Khorshidi, N., Parsa, M., Lentz, D. R., & Sobhanverdi, J. (2021). Identification of heavy metal pollution sources and its associated risk assessment in an industrial town using the K-means clustering technique. *Applied Geochemistry*, *135*, 105113. https://doi.org/10.1016/j.apgeochem.2021.105113

Kumar, M., Bajaj, K., Sharma, B., & Narang, S. (2021). A Comparative Performance Assessment of Optimized Multilevel Ensemble Learning Model with Existing Classifier Models. *Big Data*. https://doi.org/10.1089/big.2021.0257

Kumar, S., Datta, S., Singh, V., Singh, S. K., & Sharma, R. (2024). Opportunities and Challenges in Data-Centric AI. *IEEE Access*, *12*, 33173–33189. https://doi.org/10.1109/ACCESS.2024.3369417

Mihaela Tinca Udriștioiu, Y. El Mghouchi, & Hasan Yıldızhan. (2023). Prediction, modelling, and forecasting of PM and AQI using hybrid machine learning. *Journal of Cleaner Production*, *421*, 138496–138496. https://doi.org/10.1016/j.jclepro.2023.138496

Mushtaq, F., & Farooq, M. (2023). *Significance of Geo-Visualization Tools for Pollution Monitoring*. 353–384. https://doi.org/10.1007/978-3-031-45300-7\_14

Salgado, M., Nogueira, P., Torres, A., & Oliveira, M. D. (2022). Setting Requirements for a Dashboard to Inform Portuguese Decision-Makers About Environment Health in an Urban Setting. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.837433

Samad, A., Obando Nuñez, D. R., Solis Castillo, G. C., Laquai, B., & Vogt, U. (2020). Effect of Relative Humidity and Air Temperature on the Results Obtained from Low-Cost Gas Sensors for Ambient Air Quality Measurements. *Sensors*, *20*(18), 5175. https://doi.org/10.3390/s20185175

Sangeetha S.K.B, Sandeep Kumar Mathivanan, Hariharan Rajadurai, Cho, J., & Sathishkumar Veerappampalayam Easwaramoorthy. (2024). A multi-modal geospatial–temporal LSTM based deep learning framework for predictive modeling of urban mobility patterns. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-74237-3

Saxena, V. (2025). Water Quality, Air Pollution, and Climate Change: Investigating the Environmental Impacts of Industrialization and Urbanization. *Water, Air, & Soil Pollution*, *236*(2). https://doi.org/10.1007/s11270-024-07702-4

T. Lakshmi Narayana, Venkatesh, C., Kiran, A., Chinna Babu J, Kumar, A., Surbhi Bhatia Khan, Ahlam Almusharraf, & Tabrez Quasim. (2024). Advances in real time smart monitoring of environmental parameters using IoT and sensors. *Heliyon*, *10*(7), e28195–e28195. https://doi.org/10.1016/j.heliyon.2024.e28195

Tırınk, S. (2025). Machine learning-based forecasting of air quality index under long-term environmental patterns: A comparative approach with XGBoost, LightGBM, and SVM. *PLOS One*, *20*(10), e0334252. https://doi.org/10.1371/journal.pone.0334252

Vu, H. L., Ng, K. T. W., Richter, A., & An, C. (2022). Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation. *Journal of Environmental Management*, *311*, 114869. https://doi.org/10.1016/j.jenvman.2022.114869

Wu, C., Lu, S., Tian, J., Yin, L., Wang, L., & Zheng, W. (2024). Current Situation and Prospect of Geospatial AI in Air Pollution Prediction. *Atmosphere*, *15*(12), 1411–1411. https://doi.org/10.3390/atmos15121411

Ye, X., Li, S., Das, S., & Du, J. (2023). Enhancing routes selection with real-time weather data integration in spatial decision support systems. *Spatial Information Research*, *32*(4), 373–381. https://doi.org/10.1007/s41324-023-00564-8

Zender-Świercz, E., Galiszewska, B., Telejko, M., & Starzomska, M. (2024). The effect of temperature and humidity of air on the concentration of particulate matter - PM2.5 and PM10. *Atmospheric Research*, *312*, 107733. https://doi.org/10.1016/j.atmosres.2024.107733

Zeng, Y., Ma, X., Jin, N., Zhou, X., & Yan, K. (2021). Air Quality Forecasting with Hybrid LSTM and Extended Stationary Wavelet Transform. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-357905/v1