Automated Air Pollution Forecasting Using Time Series Models

B.L. Velammal
Dept. of Computer Science and Engineering
College of Engineering Guindy, Anna University
Chennai, India
velammalbl@annauniv.edu

Akshaya Srikrishna
Dept. of Computer Science and Engineering
College of Engineering Guindy, Anna University
Chennai, India
2022103065@student.annauniv.edu

Anagha Srikrishna
Dept. of Computer Science and Engineering
College of Engineering Guindy, Anna University
Chennai, India
2022103066@student.annauniv.edu

Krishnendu M R
Dept. of Computer Science and Engineering
College of Engineering Guindy, Anna University
Chennai, India
2022103081@student.annauniv.edu

***Abstract*—An automated air quality forecasting system for Chennai has been developed by integrating time series modeling with real-time data acquisition. Historical pollutant concentrations—particulate matter less than 2.5 micrometers ($PM_{2.5}$), particulate matter less than 10 micrometers (PM_{10}), carbon monoxide (CO), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2)—sourced from the Central Pollution Control Board (CPCB) form the training foundation. The system employs a Seasonal Autoregressive Integrated Moving Average (SARIMA) model, selected after hyperparameter tuning and comparison with baseline approaches including Autoregressive Integrated Moving Average (ARIMA) and Facebook Prophet. The SARIMA model, trained on Air Quality Index (AQI) values using CPCB's sub-index methodology, undergoes daily retraining to adapt to evolving pollution patterns. After optimization, it achieves a Mean Absolute Percentage Error (MAPE) of 13.57%, demonstrating strong predictive performance. The framework incorporates real-time Application Programming Interface (API) connectivity, automated retraining, and forecast generation, ensuring continuous operation with minimal intervention. By integrating SARIMA modeling with automated retraining and real-time data acquisition while outperforming baseline methods, the system demonstrates scalability and adaptability for reliable short-term forecasting. This enables proactive environmental management and informed public health decision-making.**

***Index Terms*—air quality index, forecasting, time series analysis, SARIMA, automation, environmental monitoring**

I. Introduction

Air pollution remains a pressing environmental and public health concern in many urban areas, particularly in rapidly expanding cities like Chennai, India. Prolonged exposure to elevated levels of pollutants such as $PM_{2.5}$, PM_{10}, CO, NO_2, and SO_2 has been well-documented to cause adverse health effects [1]. Timely and accurate forecasting of air quality is essential for issuing health advisories, implementing pollution control strategies.

The AQI provides a standardized measure to quantify air pollution levels and communicate potential health risks to the public. Forecasting AQI involves modeling the temporal patterns of pollutant concentrations, which are influenced by complex factors including traffic emissions, meteorological conditions, industrial activity, and seasonal variations. Machine learning techniques, particularly time series models, offer promising approaches to capturing these dynamics and generating reliable short-term predictions.

However, traditional forecasting approaches often struggle to adapt to rapidly changing environmental conditions and typically require manual updates, making them inefficient for continuous operation. Therefore, there is a need for an automated and adaptive system capable of handling real-time data streams, dynamically retraining itself, and maintaining high forecasting accuracy over time.

In this work, an automated AQI forecasting system based on the SARIMA model is developed. After comparing multiple time series models, SARIMA is selected based on its initial superior performance. Following model selection, extensive hyperparameter tuning is conducted to further optimize forecasting accuracy. The final SARIMA model achieves a MAPE of 13.57%. The system integrates real-time pollutant data ingestion via APIs, daily model retraining, and automated forecasting pipelines, ensuring robust and adaptive AQI predictions. Through this framework, accurate, scalable, and real-time AQI forecasts are delivered to support proactive environmental management and public health decision-making.

The key contributions of this work are as follows:

- Design and development of a fully automated AQI forecasting system for Chennai using SARIMA for urban air quality prediction.
- Integration of real-time pollutant data ingestion using APIs for up-to-date inputs.
- Daily automated model retraining to maintain adaptability and forecasting accuracy.
- Extensive hyperparameter tuning and model evaluation, achieving a MAPE of 13.57%.

- Deployment-ready framework that aids proactive environmental and health-related interventions.
- Comparative analysis of multiple time series forecasting models to identify the most suitable approach for AQI prediction in the context of Chennai.

II. Related Work

Accurate forecasting of air quality is critical for safeguarding public health and guiding environmental policies. As air pollution patterns grow increasingly complex, research has transitioned from classical statistical models to deep learning, hybrid architectures, and sensor-integrated systems. These innovations aim to better capture the complex spatio-temporal dynamics of pollutants like $PM_{2.5}$ and PM_{10}, which are critical for health risk mitigation and environmental planning.

A. Classical and Statistical Approaches

Bhatti et al. [2] used SARIMA and factor analysis to study $PM_{2.5}$ in Lahore, revealing a worsening air quality trend. Chen et al. [3] employed a hybrid LSTM–SARIMA model enriched with contextual features such as weather, though it was focused on energy demand forecasting rather than AQI. Vanshay et al. [4] also evaluated COVID-era air quality shifts using SARIMAX, Prophet, and DLNM, showing a positive influence of lockdowns on AQI but with limited post-lockdown generalizability.

B. Deep Learning Models

Sani et al. [5] implemented a stacked LSTM model in Dhaka, achieving over 90% forecasting accuracy. Sofiah et al. [6] compared LSTM and SARIMA for AQI modeling in Gurugram, finding LSTM superior in capturing nonlinear behavior. Zeyu Yang et al. [7] introduced a hybrid model in which an LSTM component learned from SARIMA residuals, improving overall forecasting performance—though the focus was energy data. Mudhanai et al. [8] developed a Bi-LSTM architecture integrating sensor fusion, attaining an RMSE of 21.7 in Chennai and setting a new benchmark for spatial-temporal modeling. Terroso-Saenz et al. [9] explored spatio-temporal learning using Heterogeneous Graph Neural Networks (GNNs), enhancing AQI predictions by incorporating sensor-based topologies. Qin et al. [10] proposed a novel Efficient Group-Aware Graph Neural Network (EGAGNN), targeting small-scale, data-limited environments. Their model uses a differentiable grouping mechanism and message-passing strategy to improve AQI prediction in localized areas. Thomas et al. [11] conducted a health-focused analysis linking pollution to schizophrenia-related hospital admissions using DLNM and VAR in Bangalore, uncovering lagged effects from PM_{10}, NO_2, and SO_2.

C. Machine Learning and Hybrid Approaches

Walia et al. [12] proposed a dual-model strategy using Random Forest for AQI classification and Prophet for forecasting, achieving 99.57% accuracy. Volkova et al. [13] enhanced 7-day forecast accuracy by 24.4% using a composite model combining SARIMA, Prophet, and LightGBM. Yang et al. [14] introduced a Seasonal GRU with Kalman filtering, significantly improving robustness against seasonal variation and noise in Taiwan. Ankeshit et al. [15] designed a real-time AQI monitoring system in Vijayawada using Arduino-based IoT nodes and cloud analytics; ARIMA emerged as the most accurate among baseline models for short-term predictions. Ganguli et al. [16] presented a pan-India comparison of RNN, LSTM, and ARIMA, concluding RNN models offered the best accuracy for $PM_{2.5}$ time series due to robustness to spikes.

A summary illustrating the key points have been given in Table I.

III. Proposed Method

This study investigates the effectiveness of time-series forecasting models, Prophet, ARIMA, and SARIMA for accurate short-term AQI prediction in Chennai using CPCB data. Specifically, it seeks to determine whether SARIMA can outperform ARIMA and Prophet in capturing both seasonal and non-seasonal patterns in AQI data. The complete workflow is illustrated in the architecture diagram shown in Fig. 1.

The dataset used in this study was sourced from the Central Pollution Control Board (CPCB) [17], spanning from January 2021 to January 2025, and includes hourly concentrations for pollutants such as $PM_{2.5}$, PM_{10}, CO, NO_2, SO_2, and O_3. Preprocessing steps were performed to prepare the data for time-series forecasting. Invalid entries were removed, and pollutant concentrations were standardized to a common unit (e.g., converting CO from mg/m^3 to μg/m^3). Additionally, missing values were handled through forward-fill imputation to maintain continuity in the time series. After preprocessing, the data was retained in its hourly time-series format for further analysis.

To calculate the Air Quality Index (AQI), sub-index statistics were computed for each pollutant in accordance with the CPCB guidelines. The sub-index for a given pollutant p is determined using the following equation 1:

$$I_p = \left(\frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}} \right) \times (C_p - BP_{Lo}) + I_{Lo} \tag{1}$$

where I_p is the sub-index for pollutant p, C_p is the concentration of pollutant p, BP_{Hi} and BP_{Lo} are the upper and lower breakpoints for the pollutant, and I_{Hi} and I_{Lo} are the AQI breakpoints corresponding to these concentrations. The overall AQI for each time point is then derived by selecting the maximum sub-index from all pollutants. This approach ensures that the AQI reflects the pollutant with the highest concentration, which is most critical for air quality at that specific time.

For time-series forecasting, three models were considered: Prophet, ARIMA, and SARIMA.

A. PROPHET

The Prophet model is designed to handle time-series data with strong seasonal patterns. The model decomposes the

TABLE I
COMPARATIVE SUMMARY OF PRIOR WORK IN AIR QUALITY FORECASTING

Study	Model(s) Used	Key Contributions and Results	Limitations or Gaps
Bhatti et al. [2]	SARIMA, Factor Analysis	Analyzed $PM_{2.5}$ in Lahore and observed worsening trends.	Single-city scope; lacks real-time forecasting.
Chen et al. [3]	LSTM–SARIMA Hybrid	Combined contextual features with time series models for improved accuracy.	Applied to energy forecasting, not AQI.
Vanshay et al. [4]	SARIMAX, Prophet, DLNM	Modeled AQI shifts during COVID lockdowns across major Indian cities.	Limited generalization to post-lockdown scenarios.
Sani et al. [5]	Stacked LSTM	Achieved over 90% AQI prediction accuracy in Dhaka.	Narrow feature set; localized study.
Sofiah et al. [6]	LSTM, SARIMA	Demonstrated LSTM's superiority for nonlinear AQI modeling in Gurugram.	No spatial modeling; short-term focus.
Yang et al. [7]	SARIMA–LSTM Hybrid	Modeled SARIMA residuals with LSTM for improved performance.	Focused on energy data, not pollution metrics.
Mudhanai et al. [8]	Bi-LSTM with Sensor Fusion	Achieved RMSE of 21.7 using sensor-augmented Bi-LSTM in Chennai.	Single-city; lacks long-term evaluation.
Terroso-Saenz et al. [9]	Heterogeneous GNN	Incorporated sensor-based topology for spatio-temporal AQI prediction.	High complexity; scalability concerns.
Qin et al. [10]	EGAGNN (Hierarchical GNN)	Designed for small-scale, data-limited areas; captures direct and indirect influences using grouping and message passing.	Model complexity; needs real-world deployment validation.
Thomas et al. [11]	DLNM, VAR	Linked pollution exposure to schizophrenia hospitalizations in Bangalore.	Epidemiological focus; not designed for AQI forecasting.
Walia et al. [12]	Random Forest, Prophet	Achieved 99.57% accuracy in AQI classification and forecasting.	Potential overfitting; limited validation.
Volkova et al. [13]	SARIMA, Prophet, LightGBM	Improved 7-day forecasting accuracy by 24.4% using hybrid ML pipeline.	Not tested for long-term or rare events.
Yang et al. [14]	Seasonal GRU, Kalman Filter	Enhanced seasonal robustness and noise reduction for AQI prediction.	Focused only on NO_2 and SO_2; no AQI integration.
Ankeshit et al. [15]	ARIMA, IoT Monitoring System	Developed real-time AQI pipeline using Arduino and cloud analytics.	Short-term forecasting only; limited geographic scope.
Ganguli et al. [16]	RNN, LSTM, ARIMA	Pan-India comparison showing RNN as most robust for $PM_{2.5}$.	Focused only on $PM_{2.5}$; lacks multi-pollutant modeling.

time series into trend, seasonality, and holiday components, as represented by the following equation 2:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{2}$$

where $y(t)$ is the observed value at time t, $g(t)$ is the trend component, $s(t)$ is the seasonal component, $h(t)$ represents holiday effects, and ϵ_t is the error term. Prophet is particularly effective for handling data with strong seasonal and holiday effects, but may not capture all complex relationships in pollutant concentrations. Prophet uses piecewise linear or logistic growth models for the trend component, and a Fourier series to model seasonal effects, enabling it to flexibly capture multiple seasonalities in the data.

B. ARIMA

The ARIMA model predicts future values based on past observations. The model is given by equation 3:

$$\begin{aligned} Y_t = {} & \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} \\ & + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \end{aligned} \tag{3}$$

where Y_t is the observed value at time t, $\phi_1, \ldots, \phi_p$ are the autoregressive parameters, $\theta_1, \ldots, \theta_q$ are the moving average parameters, and ϵ_t is the error term. The ARIMA model was initially configured with an order of (2, 1, 2) for the non-seasonal parameters. The choice of these parameters was informed by exploratory data analysis, which revealed significant autocorrelation in the time series data. Specifically, an autoregressive order of 2 ($p = 2$) and a moving average order of 2 ($q = 2$) were selected to effectively capture the underlying temporal dependencies, balancing model complexity and performance. The first-order differencing ($d = 1$) was applied to achieve stationarity in the series, which is crucial for the ARIMA model's predictive accuracy.

C. SARIMA

Recognizing the seasonal patterns in the AQI data, the SARIMA model was configured with an order of (1, 1, 1) for the non-seasonal components and (1, 1, 1, 7) for the seasonal components. The seasonal period was set to 7 ($s = 7$) to capture weekly patterns, which were identified through visual analysis and autocorrelation plots, as air quality tends to exhibit weekly cycles due to factors like traffic and industrial activity. The SARIMA model's inclusion of both seasonal and non-seasonal components allows it to better handle the multi-faceted dependencies in the AQI data. The selected parameters (1, 1, 1) for both the seasonal and non-seasonal parts are relatively simple but effective, providing a good balance between capturing essential patterns and avoiding overfitting. The model is given in equation 4.

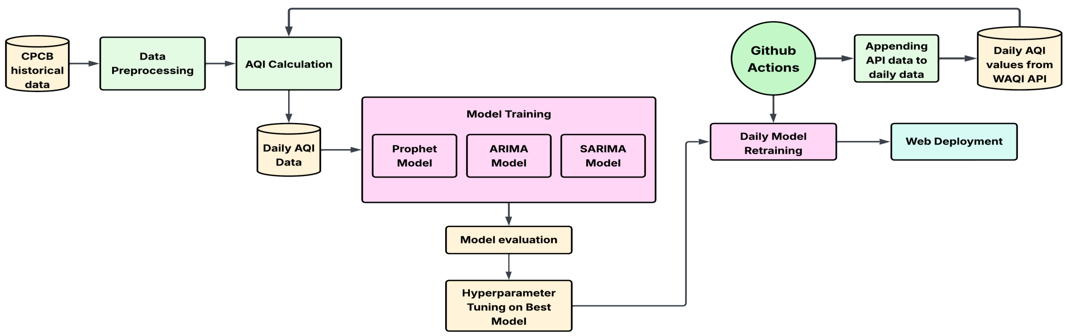

Fig. 1. System architecture for the proposed automated AQI forecasting system

$$Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \sum_{i=1}^{P} \Phi_i Y_{t-s} + \cdots + \sum_{j=1}^{Q} \Theta_j \epsilon_{t-s} + \epsilon_t \quad (4)$$

where Y_t is the observed value at time t, μ is the intercept, $\phi_1, \ldots, \phi_p$ are the sive parameters, $\theta_1, \ldots, \theta_q$ are the moving average parameters, $\Phi_1, \ldots, \Phi_P$ are the seasonal autoregressive parameters, $\Theta_1, \ldots, \Theta_Q$ are the seasonal moving average parameters, and ϵ_t is the error term. The inclusion of both seasonal and non-seasonal components allows SARIMA to effectively model both short-term fluctuations and long-term dependencies in the AQI data.

For model evaluation, the performance of each model was assessed using standard metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics were computed for both training and test datasets to evaluate model accuracy and generalizability. SARIMA, with its ability to model both seasonality and trends, was selected as the best-performing model for AQI forecasting.

D. Hyperparameter Tuning

Hyperparameter optimization for the SARIMA model was performed using grid search. The grid search considered the following range of hyperparameters for the non-seasonal and seasonal components:

$$(p, d, q) \in \{0, 1, 2, ...15\}, \quad (P, D, Q) \in \{0, 1, 2, 3\}, \quad s = 7$$

The hyperparameter tuning ensures that the model can effectively handle the non-stationary and seasonal nature of the AQI data. To ensure the model remains up to date, the SARIMA model is retrained daily after receiving new AQI data via an API. The latest AQI observations are appended to the existing dataset, and the model is retrained at three specific times each day: 5 AM, 1 PM, and 10 PM. This approach allows the model to continuously incorporate the most recent air quality data and improve forecasting accuracy over time. GitHub Actions are used to automate the retraining process and scheduling.

E. System Implementation and User Application

The developed AQI forecasting system is designed to serve Chennai residents through multiple accessible interfaces. The primary application is a real-time AQI dashboard that provides hourly updates and 24-hour forecasts for different zones across Chennai. To illustrate practical usability, the forecasting system is integrated with a web-based dashboard designed for daily use by Chennai residents. This dashboard displays predicted AQI levels for the next 5 days, providing early warnings through mail for poor air quality and actionable recommendations.

IV. Results and Discussion

The study was carried out on a Windows 11 system with an Intel Core i7-12700H CPU, 16 GB RAM and Python 3.10. The implementation utilized key libraries, including Pandas, NumPy, Matplotlib, StatsModels, Prophet, and scikit-learn. All models were trained and evaluated using Jupyter notebooks without GPU acceleration.

The dataset comprised 1460 daily observations from January 2, 2021, to December 31, 2024, with the final 90 days (January 1, 2025 to March 31, 2025) used exclusively for testing. This 90-day split ensured robust evaluation on unseen data. The primary goal was to predict AQI values for the testing period based on the historical data. To evaluate model performance, three key metrics were used: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The model parameters used are summarized in Table II.

The ARIMA(2,1,2) model, when applied to the AQI data (Fig. 2), provided reasonable forecasting performance. The MAE value of 19.1693 indicates that, on average, the model's predictions deviated from the actual AQI values by approximately 19 units. The RMSE of 22.0456 suggests that, while the model performed adequately, larger prediction errors occurred more frequently. The MAPE of 29.8500% further confirms that the ARIMA model struggles with maintaining accuracy across different levels of AQI values, particularly when dealing with high levels of variability in the data.

The SARIMA(1,1,1)x(1,1,1,7) model (Fig. 3) achieved improved performance over the ARIMA model. The MAE of

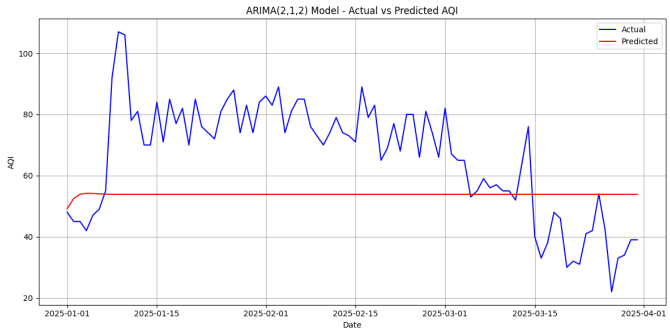

Fig. 2. ARIMA (2,1,2) Model: Actual vs Predicted AQI Values

18.0194 indicates that the SARIMA model's predictions were, on average, closer to the actual AQI values. The RMSE of 20.7107 further illustrates that the model successfully reduced the magnitude of larger prediction errors compared to ARIMA. The MAPE of 28.9064% highlights the model's better accuracy, especially for medium-to-low AQI values. The seasonal component of SARIMA, combined with the non-seasonal ARIMA structure, allowed the model to capture both short-term fluctuations and long-term trends in AQI data.

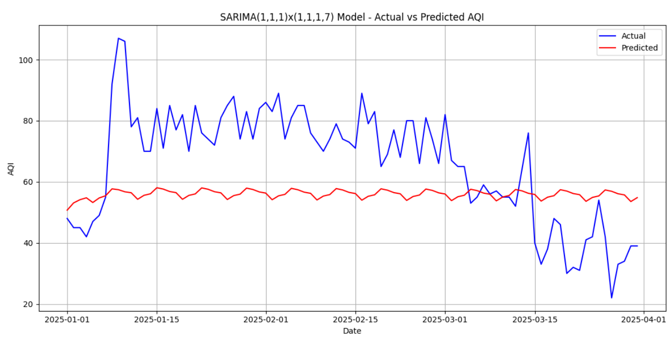

Fig. 3. SARIMA (1,1,1)x(1,1,1,7) Model: Actual vs Predicted AQI Values

The Prophet model (Fig. 4), designed to handle seasonality and trends, showed a higher degree of error. With an MAE of 26.5433, the Prophet model's predictions were less accurate compared to the ARIMA and SARIMA models. The RMSE of 29.6840 suggests that the model struggled with larger prediction errors, while the MAPE of 39.0652% demonstrates that the Prophet model was less effective in predicting AQI values accurately across the testing period.

TABLE II
MODEL PARAMETERS USED IN THE STUDY

Model	Configuration
ARIMA	(2, 1, 2)
SARIMA	(1, 1, 1)(1, 1, 1, 7), tuned: (0, 3, 15)(0, 0, 0, 7)
Prophet	Default (trend + yearly seasonality)

Despite the performance differences, the results of the SARIMA model, after tuning, demonstrate its superiority

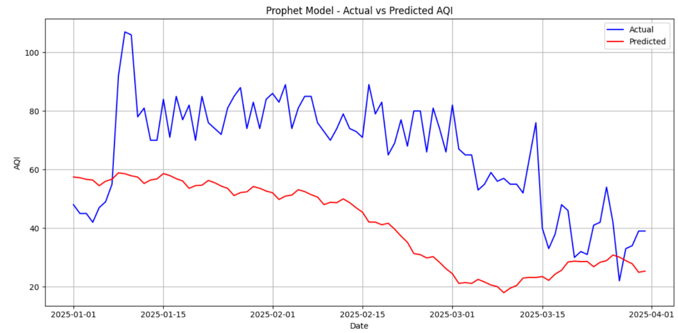

Fig. 4. Prophet Model: Actual vs Predicted AQI Values

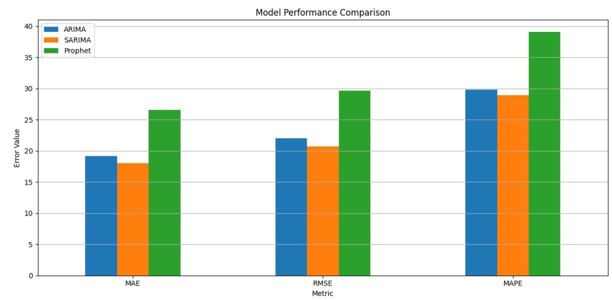

Fig. 5. Model Performance Comparison (MAE,RMSE,MAPE)

for this specific time-series forecasting task. The ability of SARIMA to capture both seasonal and non-seasonal components in the data proved to be beneficial in generating accurate predictions. Notably, SARIMA achieved these results without the need for exogenous inputs such as meteorological data, highlighting its robustness for univariate AQI forecasting.

TABLE III
MODEL PERFORMANCE COMPARISON

Metric	ARIMA	SARIMA	Prophet
MAE	19.1693	18.0194	26.5433
RMSE	22.0456	20.7107	29.6840
MAPE	29.8500%	28.9064%	39.0652%

The results (Fig. 5 and Table III) show that the SARIMA model performed better than ARIMA and Prophet across all metrics. It recorded the lowest MAE (18.0194), RMSE (20.7107), and MAPE (28.9064%), indicating smaller and more consistent prediction errors. This marks a relative improvement of 6.0% in MAE and 6.1% in RMSE over ARIMA, and nearly 32.6% improvement in MAPE over Prophet.

Hyperparameter tuning was performed on the SARIMA model to improve forecast accuracy. The optimal configuration was identified as $(p, d, q) = (0, 3, 15)$ and $(P, D, Q, s) = (0, 0, 0, 7)$, selected based on cross-validation results. The evolution of error metrics during tuning is illustrated in Fig. 6. Table IV presents a side-by-side comparison of the SARIMA model before and after tuning. Compared to the untuned configuration, the tuned SARIMA model demonstrates a significant reduction in all error metrics. Specifically, MAE dropped from 26.5433 to 11.7992, RMSE from 29.6840 to 19.4125, and MAPE from 39.0652% to 13.5732%. These improvements underscore the effectiveness of hyperparameter optimization in enhancing predictive accuracy and overall model reliability.

TABLE IV
COMPARISON OF SARIMA MODEL PERFORMANCE BEFORE AND AFTER TUNING

Metric	Untuned SARIMA	Tuned SARIMA
MAE	26.5433	**11.7992**
RMSE	29.6840	**19.4125**
MAPE	39.0652%	**13.5732%**

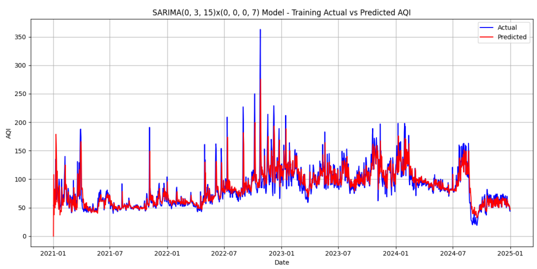

Fig. 6. SARIMA (0,3,15)x(0,0,0,7) after Hyperparameter Tuning

To maintain high prediction quality in real-time applications, the SARIMA model is retrained daily through an automated GitHub Actions workflow. This enables continuous adaptation to evolving AQI patterns using real-time air quality data without requiring manual intervention.

V. CONCLUSION

This study presents a robust framework for short-term AQI forecasting using real-time pollutant data, CPCB-based AQI computation, and time series models. Among the models evaluated (Prophet, ARIMA, and SARIMA), the SARIMA model outperformed others by effectively capturing both seasonal trends and short-term dynamics without requiring any exogenous variables. It achieved a MAPE of 13.57% after hyperparameter tuning, representing a 54% improvement over Prophet (MAPE: 39.07%) and a 54.5% improvement over the baseline SARIMA (MAPE: 28.91%). Daily retraining ensured adaptability to changing conditions, making the system suitable for urban deployment. Future work includes integrating meteorological variables to enhance accuracy and exploring deep learning approaches to model complex temporal dependencies, further advancing proactive air quality management.

REFERENCES

[1] World Health Organization, “Air quality guidelines: Global update 2021,” https://www.who.int/publications/i/item/9789240034228, Geneva, 2021, accessed: 2025-06-18.

[2] U. A. Bhatti, Y. Yan, M. Zhou, S. Ali, A. Hussain, Q. Huo, Z. Yu, and L. Yuan, “Time series analysis and forecasting of air pollution particulate matter (PM2.5): An SARIMA and factor analysis approach,” *IEEE Access*, vol. 9, pp. 41 019–41 031, 2021.

[3] F. Chen, C. Wu, J. Zhang, and G. Liu, “LSTM-based short-term data center electrical consumption forecasting,” in *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*. New York, NY, USA: Association for Computing Machinery, 2023, pp. 730–735. [Online]. Available: https://doi.org/10.1145/3594739.3612915

[4] V. Gupta, S. Kapadia, and C. Bhadane, “Time series analysis and forecasting of air quality in india,” in *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2023, pp. 1–5.

[5] S. Hossain Sani, M. Shopon, M. Hossain Khan, M. Hasan, and M. Mridha, “Short-term and long-term air quality forecasting technique using stacked LSTM,” in *Proceedings of the 6th International Conference on Communication and Information Processing*, 2021, pp. 165–171.

[6] N. A. Sofiah, K. D. Tania, A. Meiriza, and A. Wedhasmara, “A comparative assessment SARIMA and LSTM models for the gurugram air quality index’s knowledge discovery,” in *2024 International Conference on Electrical Engineering and Computer Science (ICECOS)*, 2024, pp. 26–31.

[7] Z. Yang, C. Wu, G. Liu, and J. Zhang, “Data center peak electrical demand forecasting: A multi-feature SARIMA-LSTM model,” in *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, 2023, pp. 700–705. [Online]. Available: https://doi.org/10.1145/3594739.3612910

[8] B. M. Mudhanai Sanjeevirayar, G. Mudhana, M. K. Rajagopal, P. Arumugam, V. K, and V. Varadarajan, “Integrative deep learning forecasting of air quality index in india: A fusion of bidirectional LSTM and sensor data,” in *Proceedings of the 2023 11th International Conference on Information Technology: IoT and Smart City*, 2024, pp. 194–199. [Online]. Available: https://doi.org/10.1145/3638985.3639017

[9] F. Terroso-Saenz, J. Morales-García, and A. Muñoz, “Nationwide air pollution forecasting with heterogeneous graph neural networks,” *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 1, 2024. [Online]. Available: https://doi.org/10.1145/3637492

[10] J. Qin, Z. Liu, W. Sui, T. Peng, and S. Zhao, “Efficient group-aware graph neural network for air quality forecasting in small-scale spaces,” in *2024 International Applied Computational Electromagnetics Society Symposium (ACES-China)*, 2024, pp. 1–3.

[11] J. S. Thomas and M. Parameswaran, “Time series lag models to analyze impact of air pollution on daily incidence of schizophrenia,” in *2023 International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS)*, vol. 1, 2023, pp. 1–7.

[12] A. Walia, A. Paliwal, S. Patidar, and R. Mahto, “Prediction of air quality index using random forest and Prophet tool,” in *2024 19th Annual System of Systems Engineering Conference (SoSE)*, 2024, pp. 275–280.

[13] E. Volkova, A. Ryabyshenkov, A. Gorelik, S. Muratchaev, and I. Zhuravlev, “Development of an air quality forecasting model based on neural networks and machine learning methods,” in *2024 Conference of Young Researchers in Electrical and Electronic Engineering (ElCon)*, 2024, pp. 1003–1006.

[14] C.-H. Yang, P.-H. Chen, C.-S. Yang, and L.-Y. Chuang, “Analysis and forecasting of air pollution on nitrogen dioxide and sulfur dioxide using deep learning,” *IEEE Access*, vol. 12, pp. 165 236–165 252, 2024.

[15] A. Srivastava, A. Ahmad, S. Kumar, and M. A. Ahmad, “Air pollution data and forecasting data monitored through google cloud services by using artificial intelligence and machine learning,” in *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, 2022, pp. 804–808.

[16] I. Ganguli, M. Nakum, B. Das, and N. Kshetrimayum, “Comprehensive analysis of air quality trends in india using machine learning and deep learning models,” in *Proceedings of the 26th International Conference on Distributed Computing and Networking*, 2025, pp. 313–318. [Online]. Available: https://doi.org/10.1145/3700838.3703681

[17] K. M. R., “Air quality dataset used for time series forecasting (alandur region),” https://github.com/KrishnenduMR/ChennaiAir/blob/main/static/data/alandur_daily.csv, 2025, gitHub repository, accessed April 29, 2025.