iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.81816

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Time-Series Forecasting of Air Quality Index (AQI) for Pune City Using Facebook Prophet Model

Dhruv Kanojia[1], Aditya Unde[2], Dnyaneshwar Sonawane[3], Narayan Raut[4], Dr. Shilpa Dange[5]
Department of Computer Science, Modern College of Arts, Science and Commerce, Pune, India

Abstract: Pune's air quality has drastically declined as a result of the city's rapid urbanization and growing car population, endangering public health. We need trustworthy models to forecast the future pollution levels before they occur in order to find solutions to this issue, not just the monitoring systems we currently have. This study uses the Facebook Prophet algorithm as a machine learning method to predict Pune's Air Quality Index. We chose this model because it is the model can handle the seasonal and complex nature of environmental data more effectively than conventional statistical techniques.

The study examines a historical dataset from 2017 to 2024 that included common data problems like sensor errors and missing dates. We used a meticulous cleaning procedure to prepare the data, standardizing the timestamps and filling in any missing values with linear interpolation. To prevent the model from being misled by gaps in the data, it was crucial to create a continuous and uninterrupted timeline. The Prophet model divides the data into daily, weekly, and annual patterns in order to identify hidden trends, in contrast to basic linear regression, which assumes a straight line trend.

Our findings revealed clear cycles in Pune's air quality, particularly how the monsoon rains naturally purify the air while wintertime traps pollutants and raises the Air Quality Index. Standard error metrics, such as Mean Absolute Error and Root Mean Squared Error are used to gauge how accurate our predictions were. The results show that our model effectively captures these regional patterns and provides a useful instrument for environmental monitoring. This system could effectively assist legislators and city planners in making data-driven decisions, such as controlling traffic flow before pollution levels become hazardous or promptly issuing health warnings.

Keywords: Air Quality Index (AQI), Facebook Prophet, Machine Learning, Time-Series Forecasting, Pune Pollution, Python.

I. INTRODUCTION

The health of millions has been effected negatively by India's severe air quality degradation brought on by urbanization and industrial growth. Particulate matter concentrations have significantly increased in Pune, the "Oxford of the East" and a growing IT center, as a result of ongoing metro construction projects and vehicle emissions. An essential tool for making the public aware about the seriousness regarding air pollution is the Air Quality Index (AQI). However, the intricate, non-linear patterns of pollutant dispersion influenced by meteorological factors are frequently not captured by conventional deterministic models [1].

Environmental modeling has recently seen an increase in the use of the Machine Learning (ML) techniques. Time-Series forecasting is still a crucial method among these. Despite their popularity, models such as LSTM (Long Short-Term Memory) networks and ARIMA (Auto-Regressive Integrated Moving Average) frequently require extensive parameter tuning and struggle with missing data or abrupt trend shifts that are frequent in environmental datasets [2]. This work explores the implication of the Facebook Prophet, an open-source time-series forecasting tool created by Meta, to estimate the AQI for the city of Pune. The model is ideal for air quality analysis as it separates time-series data into interpretable elements such as overall trends, recurring seasonal behavior, and holiday-related effects, while also dealing efficiently with missing entries and abnormal values. By utilizing daily AQI observations collected between 2017 and 2024, the study focuses to design robust forecasting framework that can assist local administrative bodies, including the Pune Municipal Corporation (PMC), in planning and taking proactive, data-driven decisions.

II. LITERATURE REVIEW

Air quality prediction research has evolved considerably, moving away from purely statistical models toward more advanced techniques powered by artificial intelligence and machine learning.

Dorage et al. [4] conducted an extensive analysis of Pune's Air Quality Index (AQI) using a multiple machine learning approaches. Their study demonstrated that ensemble-based models such as CatBoost and XGBoost are highly effective in capturing the non-linear characteristics commonly present in air pollution data. The authors also highlighted the critical role of meteorological factors, particularly temperature and humidity, in influencing pollutant concentrations.

Similarly, Cai [2] investigated the use of the Prophet algorithm for air quality forecasting and showed that it performs notably out performs than standard regression methods when modeling recurring seasonal effects. These include predictable variations such as reduced pollution during weekends and elevated levels during festival periods. The study further noted that Prophet's additive structure improves model transparency, which is an important consideration for environmental policy analysis.

In another related study, a case analysis from Northern Nigeria compared the forecasting performance of LSTM and Prophet models. The results indicated that while LSTM is better suited for datasets with abrupt structural changes, Prophet frequently delivers equal or superior accuracy in time series dominated by strong seasonal patterns. Such seasonal dominance closely reflects the climatic conditions observed across many regions in India [5].

Moreover, assessments released by the Centre for Science and Environment (CSE) have raised some concerns over the rising concentrations of ground-level ozone and PM2.5 within the Pune Metropolitan Region. These reports stress the necessity for more granular and data-driven air quality monitoring and forecasting systems [6]. Motivated by these findings, the present study applies the Prophet model specifically to Pune using an extensive dataset spanning eight years.

III. METHODOLOGY

A. Dataset Description

The primary dataset employed in this study is the *"Pune AQI Data (2017–2024)"*, which was obtained from the public Kaggle repository [9]. This dataset provides daily measurements of AQI along with corresponding categorical labels.

- Temporal Coverage: The data spans from January 1, 2017, to the year 2024.
- Attributes: The dataset includes Year, Month, Day, Date, AQI value, and AQI category.
- Volume: More than 2,500 daily AQI observations are available.

To assist in validating short-term pollution fluctuations, an additional dataset consisting of hourly pollutant concentrations for the year 2024 was also examined. This auxiliary dataset includes parameters such as PM2.5, PM10, and NO_2. Nevertheless, the primary forecasting and analysis in this work are based on the daily AQI time series.

B. Data Pre-processing

Environmental data collected from real-world monitoring systems often contains missing entries and irregularities. To improve data quality and ensure reliable modeling, several preprocessing part was performed in Python with the Pandas library.

1) Handling Missing Values: A portion of the dataset, particularly from the early years (2017–2018), contained AQI values marked as "Unknown." These missing observations were addressed using linear interpolation to preserve continuity in the time series while avoiding the introduction of artificial variance.
2) Date Conversion: The Date column was converted into a datetime format and set as the index to enable compatibility with time-series forecasting models.
3) Outlier Treatment: Extreme deviations caused by potential sensor anomalies were mitigated using the interquartile range (IQR) method, where unusually high or low val-
ues were capped to reduce their influence on long-term trends.

C. Facebook Prophet Model Architecture

The predictive framework adopted in this research is dependent on the Facebook Prophet algorithm, originally introduced by Taylor and Letham [3]. Prophet employs a decomposable time-series formulation that separates observed data into interpretable components, making it both robust to missing values and transparent in its predictions [2]. Unlike black-box neural network models, Prophet provides clear insights into underlying temporal patterns.

The model is defined by the equation:

$$y(t) = g(t)+s(t)+h(t)+\epsilon_t \qquad (1)$$

where:

- g(t) represents the trend component, capturing nonperiodic variations through piecewise linear or logistic growth functions [3].
- s(t) models recurring seasonal patterns, such as weekly and annual cycles, using Fourier series.
- h(t) accounts for the impact of holidays and special events, including Diwali and Ganesh Chaturthi.
- ϵt denotes the random error term.

For forecasting Pune's AQI, yearly seasonality was explicitly enabled to capture the influence of meteorological factors such as the monsoon, which has been identified an important determinant of the air quality variation in earlier studies [4].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

IV. EXPERIMENTAL RESULTS

This part shows the FB-Prophet model's predictions for the Pune AQI dataset. [1].

A. Trend Analysis

The time-series decomposition generated by the Prophet model reveals several notable patterns in Pune's air quality dynamics:

- Overall Trend: A gradual increase in baseline AQI levels is observed from 2017 to 2024, reflecting the effects of rapid urbanization and the rising number of vehicles in the city.
- Yearly Seasonality: A clear U-shaped annual pattern emerges, with AQI values reaching their lowest levels during the monsoon months of July and August due to enhanced pollutant dispersion, and peaking during December and January when atmospheric inversion restricts pollutant diffusion.
- Weekly Seasonality: Slight reductions in AQI are consistently observed on Sundays, which may be attributed to decreased traffic flow and lower industrial activity.

B. Forecasting

The model was trained using AQI data from 2017 to 2023, and forecasts were generated for the year 2024 to assess predictive performance against actual observations.

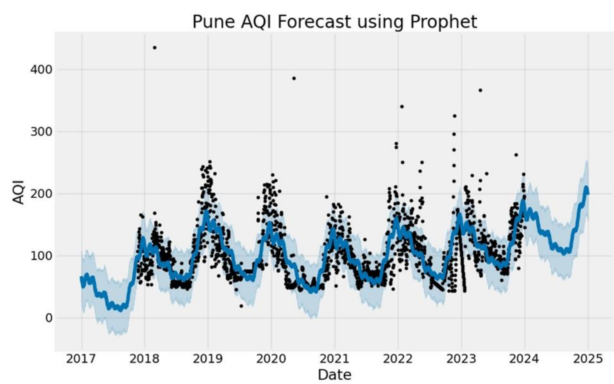


Figure 1: Forecast of Pune AQI using Facebook Prophet Model. Black dots shows the actual data, and the blue line represents the predicted trend.

Figure 2: Prophet Model Forecast for Pune AQI.

The forecast successfully captures the sharp decline in AQI beginning in June 2024, which coincides with the onset of the monsoon season. However, the model slightly underestimates the extreme pollution peaks observed in November 2023. These anomalies commonly due to the localized construction activities and episodic pollution events that are not represented in historical seasonal patterns.

C. Performance Metrics

The forecast accuracy was tested using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), which are widely used metrics for time-series prediction tasks.

Table 1: Model Evaluation Metrics

Metric	Value
RMSE	23.45
MAE	18.12

Given the large variability in AQI values, which often range from 50 to well above 300, an RMSE of 23.45 is considered acceptable.

[1] You can find the Jupyter Notebook that has the data preprocessing, model training, and evaluation metrics in the Project Repository.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

V. DISCUSSION

The results confirm that Pune's air quality is strongly influenced by seasonal meteorological conditions. The Prophet model identifies a pronounced U-shaped annual trend, with improved air quality during the monsoon months (July–August) and elevated pollution levels during winter (December–January). These findings are consistent with the observations of Dorage et al. [4], who attributed winter pollution spikes to temperature inversion and reduced wind speeds.

Furthermore, the 2024 forecasts suggest an rise in number of days classified under the "Poor" AQI category when compared to the 2017–2019 baseline period. This trend aligns with recent assessments by the Centre for Science and Environment (CSE), which reported rising particulate matter concentrations in the Pune Metropolitan Region [6].

A key limitation of the present univariate modeling approach is its inability to account for abrupt, non-cyclical events such as accidental fires or sudden construction activity without the inclusion of external regressors. As highlighted by Samad et al. [5], hybrid forecasting frameworks that combine statistical models with LSTM-based residual learning often yield improved performance under highly volatile conditions.

VI. CONCLUSION

This study effictively shows the efficieny of using the Facebook Prophet model in forecasting air quality trends for Pune. By utilizing AQI data spanning from 2017 to 2024, the analysis highlights a recurring improvement in air quality during the monsoon season, alongside a broader and concerning upward trend in pollution levels over time. The proposed approach offers a computationally efficient baseline forecasting solution that can be extended into a web-based platform for public awareness and policy support. Future enhancements will explore hybrid modeling techniques that combine Prophet's trend modeling capabilities with LSTM-based residual learning to further reduce prediction error.

REFERENCES

[1] F. Cai, "The Prediction of the Air Quality based on the Prophet Algorithm," Highlights in Science, Engineering and Technology, vol. 39, pp. 1056–1060, 2023. [2] S. J. Taylor and B. Letham, "Forecasting at Scale," The American Statistician, vol. 72, no. 1, pp. 37–45, 2018.

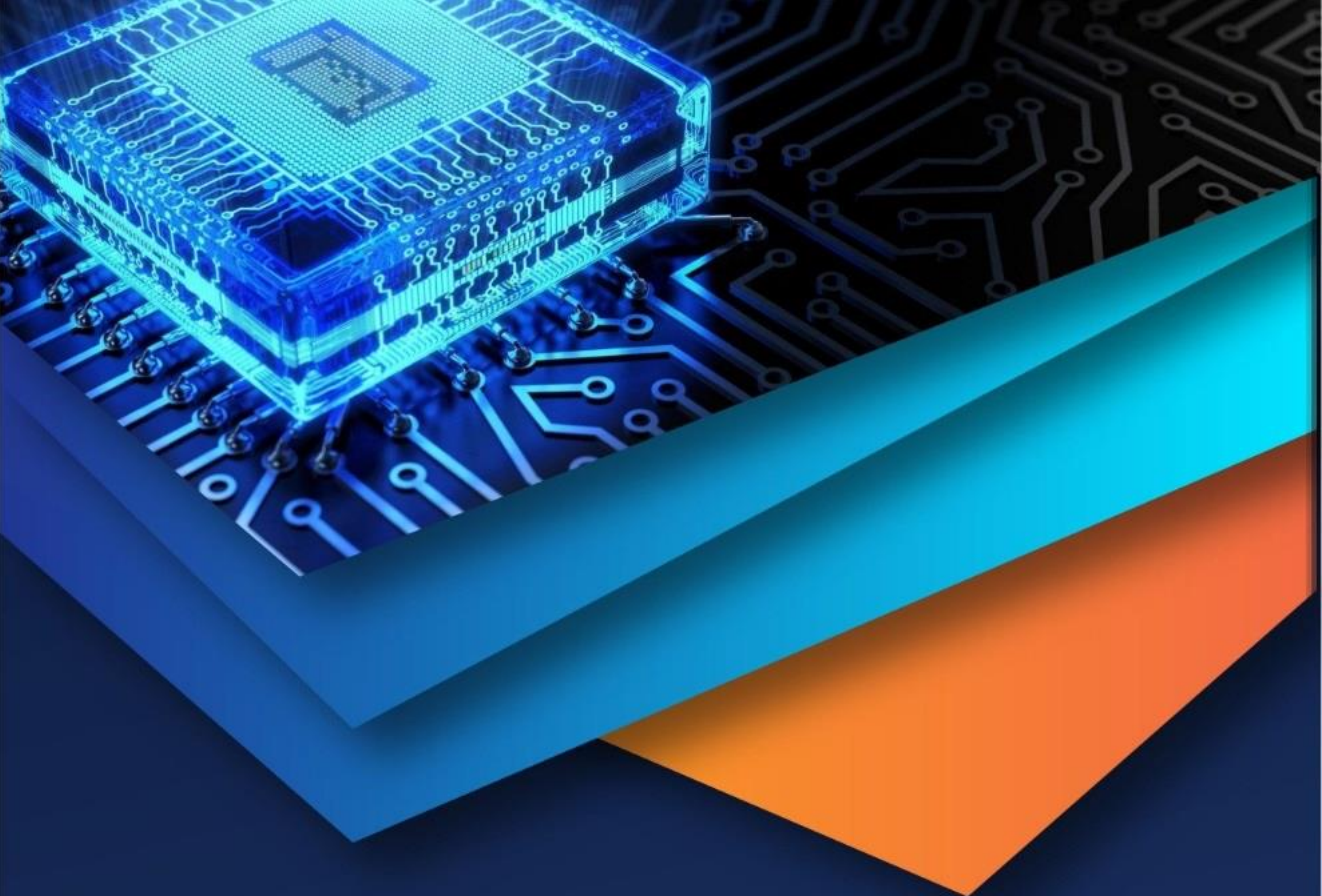
[2] A. Dorage, A. Chopkar, T. Kulkarni, and A. M. Deshpande, "Predictive Analysis of Pune City's Air Quality Index: A ML and Time Series Approach," in 2024 15th Conference on Computing Communication and Networking Technologies (ICCCNT), 2024.

[3] A. Samad et al., "Facebook's Prophet vs LSTM for Air Pollution Forecasting in Data-Constrained Northern Nigeria," arXiv preprint arXiv:2508.16244, 2025.

[4] Centre for Science and Environment, "Air Quality Tracker: Pune Metropolitan Region," Aug. 2024. [Online]. Available: https://www.cseindia.org/air-quality-trackerpune-metropolitan-region-12320.

[5] Pune Municipal Corporation, "Environmental Status Report 2024-25," PMC, Pune, India, Tech. Rep., 2025.

[6] B. Alam, A. Hussain, and M. Fayaz, "An Effective Approach for Air Quality Prediction in Bishkek Based on Machine Learning techniques," Applied Sciences, vol. 12, no. 15, 2022. [8] P. Raizada, "Pune Air Quality Index Dataset (2017-2024)," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/pranavraizada/pune-airquality-index-dataset.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY