Analysis and Prediction of Air Quality Index (AQI) in Gorakhpur During Diwali Festival Using Python (2016–2027)

Er. Prateek Shahi[1], Vaibhav Kumar Mishra[2], Belal Ahmad Siddique[3], Aditya Kumar[4], Rajesh Singh[5] Mohammad Amanatullah[6]

[1,2,3,4,5,6]*KIPM College of Engineering and Technology, Gida, Gorakhpur*

***Abstract*—This research talks about Gorakhpur during festival time and how air quality changes every year. Diwali is a well-known problem across India, but most research and media attention goes to Delhi. The city is located near GIDA Industrial zone on flat plate area, where pollutants are trapped. When Diwali comes in November, cold air goes downward therefore pollution is not separated and remains frozen. Our group members collected AQI readings for the Diwali period (2016–2024) at three stations in Gorakhpur: MMMUT (Madan Mohan Malaviya University of Technology campus), JAL KAL (near Jal Nigam office with high traffic), and GIDA (near Gorakhpur Industrial Development Authority zone). For every station, we built a Linear Regression model using Python, then used those models to predict the Diwali AQI at each station in 2026 and 2027. MMMUT showed a downward trend. JAL KAL showed no clear trend due to weather conditions. GIDA gave the most serious result, with the highest AQI of all three stations and an upward trend. Predicted AQI for 2026: 95.57 (MMMUT), 174.46 (JAL KAL), and 260.53 (GIDA). For 2027: 93.87, 174.26 and 265.23. The difference between MMMUT and GIDA in 2026 is about 165 AQI units, meaning GIDA is predicted to be nearly three times more serious than MMMUT on the same Diwali night.**

***Index Terms*—QI, Diwali pollution, Gorakhpur, MMMUT, JAL KAL, GIDA, Python, Linear Regression, air quality prediction, Uttar Pradesh.**

I. INTRODUCTION

1.1 Why We Chose This Topic

This topic came from personal experience. Some group members lived in Gorakhpur, and every year around Diwali the air quality changes. There is a yellow colour that settles in the environment. The air smells like burnt material. When we come outside, we feel uncomfortable in our eyes and throat. It was experienced every year.

One member of our group has a younger sibling suffering with asthma. For this sibling, nebulizer medicine is required every year. That is a very poor condition in Gorakhpur. Some people we know have similar stories. When we discussed project topics, this came up naturally.

From our point of view, air quality is a genuine Civil and Environmental Engineering problem. It connects directly to how cities are planned and where industrial zones are located. We used Python, but this is not a computer science project. The actual question we are asking is why air quality is more polluted in some parts of a city than others.

When we searched for published research on Diwali AQI in Indian cities, Gorakhpur did not appear anywhere. Delhi and Lucknow appear everywhere. Gorakhpur does not, even though it also has an industrial zone, geography that is genuinely bad for pollutant dispersal, and a significant population that deals with Diwali pollution every year.

1.2 About Gorakhpur and the Three Monitoring Stations

Gorakhpur is located on flat alluvial land. It has a significant population, a functioning university, a medical college, and a growing industrial base through the Gorakhpur Industrial Development Authority. Despite all these factors, it does not feature in national-level environmental research.

Sitting on the flat Indo-Gangetic Plain means there are no hills. During October and November, temperature inversions are common in this part of UP. Cold dense air near the ground gets trapped under a warmer layer above, and without enough wind to break this inversion, pollutants stay concentrated at breathing level for extended periods. This is the atmospheric backdrop for every Diwali in Gorakhpur.

We chose three monitoring stations:

- MMMUT (Madan Mohan Malaviya University of Technology), established in 2013. The campus has banned vehicle access, more green cover than the surrounding city and no heavy industry nearby.
- JAL KAL is located near Jal Nigam office. This area has heavy traffic, closely packed residential areas, especially during Diwali.
- GIDA (Gorakhpur Industrial Development Authority zone), which covers a wide range of industries including textiles and construction material manufacturing. This zone has been expanding over the study period. GIDA has the highest AQI of all three stations every single year on Diwali.

1.3 Problem Statement

Gorakhpur people can easily know weather forecasts every week. Temperature and rainfall are available on phones. However, many people cannot find AQI during Diwali time.

From an urban planning point of view, advance knowledge of Diwali AQI has many benefits. If any model predicts AQI exceeding 260 during Diwali, people can issue targeted health advisories for industrial zones. Industries can also be made aware to reduce emissions during Diwali time. It also increases enforcement of National Green Tribunal orders on firecracker sales.

The dataset was small, covering nine years of Diwali readings per station. The three stations behave very differently.

1.4 Objectives

The specific objectives are given below:

- To collect AQI data for the Diwali period at MMMUT, JAL KAL and GIDA stations from 2016 to 2024.
- To identify year-on-year trends at every station by fitting a separate Linear Regression model.
- To generate AQI predictions for Diwali 2026 and 2027 at every station using fitted models.

1.5 Structure of This Report

Chapter 2 discusses research on Diwali air pollution and machine learning AQI prediction. Chapter 3 describes where data came from and how models were built. Chapter 4 presents result for every station and comparisons. Chapter 5 summarizes key findings, limitations, and outlines directions for future work. The Python code used in this study is provided in Appendix A and the national AQI category table in Appendix B.

One important note: this study focuses only on the Diwali period. It does not model Gorakhpur air quality across the full winter season from October to February. The Diwali focus was chosen because it is a fixed annual event, and the small dataset was sufficient to identify patterns during that specific period.

II. LITERATURE REVIEW

2.1 Research on Air Pollution During Diwali

Scientists began research on air quality during Diwali in the early 2000s. Notable studies include Ravindra and Kaushik (2003), who studied PM10 levels in Patiala before, during, and after the festival. They found that concentrations during Diwali were about six times higher compared to normal days, proving that burning firecrackers during Diwali suddenly increases particulate matter.

Barman and colleagues also looked at Lucknow (2008). PM2.5 reached very poor quality. Singh et al. published a chemical analysis (2010).

Studies show that when Diwali comes in November, the situation is much worse compared to when it comes in October. The reason is atmospheric. In eastern UP, air is cold and stable in November. In October, firecracker pollution disperses within hours, but in November pollutants are trapped near the ground. We can observe this in our dataset. All three stations showed maximum AQI in years like 2018 and 2021 when Diwali was in November. In 2016 and 2022, when Diwali was in October, air quality was comparatively clean.

Despite all the above, we could not find a single published study that included Gorakhpur. This absence is notable given that the city has a major industrial zone, poor atmospheric conditions for pollutant dispersal during winter, and a large population experiencing pollution every year.

2.2 Machine Learning Approaches for AQI Prediction

Machine learning models use many years of data to make predictions, making them useful for future air quality forecasting. The basic idea is to train an algorithm on past data so it understands the relationship between input variables (weather, traffic) and the output variable (pollutant concentration). This differs from physics-based dispersion models.

Several approaches have been tested. Gu and colleagues used a neural network in 2020 for Shenzhen, reporting an R-squared value greater than 0.85. Random Forest has been applied in both Indian and Chinese city studies for daily AQI prediction. LSTM networks are used for AQI time-series because they capture temporal dependencies well.

In Delhi, Sharma et al. (2021) found that Gradient Boosting performed best in comparisons. Related to our work, Pandey et al. (2022) applied trend-based regression to Diwali air quality in Varanasi, concluding that historical data is sufficiently consistent to support simple linear trends for Diwali period forecasting.

The general lesson is that model complexity should match data size. Complex models like neural networks are appropriate when there are many observations. When the dataset is small, simpler models are safer because complex models overfit and give unreliable predictions. With only nine observations per station, Linear Regression was clearly the appropriate choice.

2.3 Why We Chose Linear Regression

With only nine data points per station, Linear Regression was a practical choice for four reasons.

First, Linear Regression is designed for exactly what we wanted to know: whether AQI at each station was going up or down, and by how much per year. The model coefficient directly answers this in a way that is easy for anyone to understand.

Second, Linear Regression does not overfit on small datasets, unlike complex models. When Random Forest is trained on nine points, it tends to memorize the data, and predictions for future values become unreliable.

Third, the results are easily explained. A slope of −1.70 for MMMUT means AQI dropped by about 1.7 units per year. A slope of +4.70 for GIDA means it is rising by about 4.7 units per year.

Fourth, the implementation using scikit-learn is transparent. Anyone who wants to check or replicate our work can do so easily without specialized expertise.

2.4 What Is Missing in Existing Research

After studying the above research carefully, we identified four gaps that this study addresses:

- Published research on air quality during Diwali is almost entirely focused on big cities and state capitals. Gorakhpur, which includes an important industrial zone, does not appear in the literature.
- Most city-level studies report a single figure per city. Nobody has studied intra-city variation, how AQI differs across different zones within a single city. Our study reveals a difference of over 165 AQI units within Gorakhpur on the same Diwali night.
- Many AQI prediction studies focus on short-term (next-day) forecasting. Multi-year Diwali-period prediction is rare. Our prediction horizon (2026 and 2027) is more useful for planning than next-day forecasts.
- Very limited research looks at AQI from a civil engineering perspective, combining results clearly with city planning, industrial zoning, and public health systems.

III. DATA COLLECTION AND METHODOLOGY

3.1 Data Sources

The data was taken from two publicly available government sources. The first is the Central Pollution Control Board (CPCB), accessible at cpcb.nic.in, which provides AQI data from monitoring stations across India and stores several years of historical data. The second is the Uttar Pradesh Pollution Control Board (UPPCB), which provides detailed data for monitoring stations within UP, including the three Gorakhpur stations used in this study.

For each year (2016 to 2024), we collected Diwali data covering approximately five days: two days before, the day itself, and two days after. From these five days, we selected the highest AQI as the representative value for that year. We used the peak rather than the average because during Diwali the most extreme air quality is what people actually experience and what matters most for health.

Data was not always available for every station and year combination. Where values were missing (primarily 2016 and 2017 at certain stations), we estimated using linear interpolation based on readings from the nearest available years.

3.2 Dataset Description

Table 3.1 below shows the complete dataset. Every

	MMMUT	JAL KAL	GIDA
2026	95.57 (Satisfactory)	174.46 (Moderate)	260.53 (Poor)
2027	93.87 (Satisfactory)	174.26 (Moderate)	265.23 (Poor)
	MMMUT	JAL KAL	GIDA
2026	95.57 (Satisfactory)	174.46 (Moderate)	260.53 (Poor)
2027	93.87 (Satisfactory)	174.26 (Moderate)	265.23 (Poor)

row represents one year and the peak Diwali AQI for the three stations.

Year	Diwali Date	MMMUT AQI	JAL KAL AQI	GIDA AQI
2016	Oct-30	67	110	123
2017	Oct-19	78	111	123
2018	Nov-07	177	270	356
2019	Oct-27	145	235	291
2020	Nov-14	104	163	248
2021	Nov-04	114	256	324
2022	Oct-24	78	131	226
2023	Nov-12	127	202	230
2024	Nov-01	62	103	170
2025	Oct-20	97.27	174.66	255.83

Table 3.1: Diwali-season peak AQI for Gorakhpur stations, 2016 to 2025.

Looking at this table, several things stand out. The year 2018 was the worst across all three stations: MMMUT (177), JAL KAL (270) and GIDA (356, Very Poor category). Diwali fell on November 7, 2018, meaning cold and still air created extreme conditions. The years 2016 and 2017 were the cleanest, with Diwali in October both years allowing pollution to disperse. The year 2024 had the lowest readings overall: MMMUT (62), JAL KAL (103) and GIDA (170). One pattern repeated every year without exception: GIDA always had the highest AQI.

3.3 Data Preparation

Since Linear Regression uses only one predictor, data preparation was straightforward with no complex preprocessing required. We created a pandas DataFrame in Python with five columns: Year, MMMUT, JAL KAL and GIDA. Each row represents one year (2016 to 2025). The Year column was used as the independent variable for all three models.

We built separate models for each station because the trend at each station differs in both direction and magnitude. Fitting a single combined model would hide these important differences.

We did not apply feature scaling, as it is unnecessary when there is only one numerical predictor and does not affect the results. For generating predictions, we applied the fitted models to future years 2026 and 2027.

We trained on the full available data rather than splitting into train and test sets. With only nine data points, a train-test split would leave too few samples for meaningful model training and would create an illusion of accuracy. We acknowledged this limitation explicitly.

3.4 The Linear Regression Model

Linear regression fits a straight line that minimizes the total squared error between actual and predicted values. The model equation is:

$$AQI = \beta_0 + \beta_1 \times Year + \varepsilon$$

The slope β_1 is the key number. If negative, AQI is trending downward. If positive, AQI is trending upward. The intercept β_0 is a mathematical constant. The term ε represents residual error, mostly caused by different weather conditions on each Diwali night.

We used scikit-learn's Linear Regression class in Python 3.10. Models were fitted using the. fit() method and predictions were generated using. predict () with input values of 2026 and 2027. The estimated slope values from the fitted models were approximately

Table 3.2: Predicted Diwali-period AQI for 2026 and2027 at the three Gorakhpur stations

MMMUT: slope ≈ −1.70 AQI units per year
- JAL KAL: slope ≈ −0.20 AQI units per year
- GIDA: slope ≈ +4.70 AQI units per year

3.5 Predictions for 2026 and 2027
Using the three fitted models, we produced AQI forecasts for both 2026 and 2027 at all three stations. The results are shown in Table 3.2.

IV. RESULTS AND DISCUSSION

4.1 MMMUT Station
MMMUT gave the most encouraging result of the three. The model found a negative slope of (−1.70) units per year, meaning Diwali air quality on campus has been slowly improving since 2016. This is not a dramatic change, but the direction is correct. Values were 67 in 2016, jumped to 177 in 2018 due to November timing and poor weather that year, and then came back down to 62 in 2024. Setting aside the 2018 exception, the general direction since 2016 has been downward.

Year	Actual AQI	Prediction	Category
2016	67	—	Satisfactory
2017	78	—	Satisfactory
2018	177	—	Moderate
2019	145	—	Moderate
2020	104	—	Moderate
2021	114	—	Moderate
2022	78	—	Satisfactory
2023	127	—	Moderate
2024	62	—	Satisfactory
2025 (est.)	97.27	—	Satisfactory
2026	—	95.57	Satisfactory
2027	—	93.87	Satisfactory

Table 4.1: MMMUT Station — Diwali AQI history and predictions

The predicted values of 95.57 for 2026 and 93.87 for 2027 both fall in the Satisfactory range (51–100 on the Indian AQI scale). This is not fully clean—people with asthma should still take precautions—but compared to the other two stations on the same night, MMMUT conditions are in a much better position.

The improving trend at MMMUT has a clear explanation. Vehicle access on campus is limited, so traffic-related pollution is very low. Campus residents also tend to have higher environmental awareness, resulting in lower firecracker use.

4.2 JAL KAL Station
JAL KAL was the hardest station to interpret. Raw values vary extremely: 103 in 2024, 270 in 2018, and a wide spread in between—a 167-unit range within a nine-year dataset. The regression slope came out at nearly zero, around (−0.20) per year, so the model sees no meaningful trend in any direction. Forecasts for 2026 (174.46) and 2027 (174.26) are almost identical, meaning the model is essentially predicting the historical average since there is no trend to project.

Year	Actual AQI	Prediction	Category
2016	110	—	Moderate
2017	111	—	Moderate
2018	270	—	Poor
2019	235	—	Poor
2020	163	—	Moderate
2021	256	—	Poor
2022	131	—	Moderate
2023	202	—	Poor
2024	103	—	Moderate
2025 (est.)	174.66	—	Moderate
2026	—	174.46	Moderate
2027	—	174.26	Moderate

Table 4.2: JAL KAL Station — Diwali AQI history and predictions.

The large variation at JAL KAL is primarily driven by weather conditions on Diwali night. Pollution sources at this station do not change significantly year to year; what changes is the weather. A cold, still November night pushes AQI to 270. A breezy October night drops it to 103. The time-based model cannot predict this because weather data was not included.
For people who live near JAL KAL, the average estimate is 174, but on any given Diwali it can range from 100 to 270 depending mainly on weather. Adding wind speed, temperature and atmospheric mixing height data would substantially improve predictions

for this station and represents the most valuable improvement for future work.

4.3 GIDA Station

GIDA produced the most concerning result. It was not just that GIDA was the highest station every year. The main problem is that it is moving in the wrong direction. The slope is approximately +4.70 units per year. Over a decade, this produces a rise of 40–50 AQI units, which could push values that are currently in the Poor range into Very Poor territory.

Year	Actual AQI	Prediction	Category
2016	123	—	Moderate
2017	123	—	Moderate
2018	356	—	Very Poor
2019	291	—	Poor
2020	248	—	Poor
2021	324	—	Very Poor
2022	226	—	Poor
2023	230	—	Poor
2024	170	—	Moderate
2025 (est.)	255.83	—	Poor
2026	—	260.53	Poor
2027	—	265.23	Poor

Table 4.3: GIDA Station — Diwali AQI history and predictions.

The predicted values for 2026 (260.53) and 2027 (265.23) sit firmly in the Poor range (201–300 on the national scale). Prolonged outdoor exposure at this level affects both healthy individuals and those with respiratory conditions. For people with asthma, COPD, or heart disease, this level of air quality is dangerous and should be highlighted in health advisories.

The upward trend at GIDA is explained by industrial growth itself. More factories, more production activity, and more vehicle movement in and out of the zone means the baseline pollution in GIDA before Diwali is higher now than it was in 2016. When Diwali firecracker emissions are added to this elevated baseline, the combined peak is higher. The problem is not firecrackers alone; it is the interaction between firecrackers and rising industrial background pollution.

Another factor that may be relevant is enforcement during festival periods. Regular monitoring is typically reduced during holidays. If some factories in the GIDA zone take advantage of reduced oversight by operating at higher output levels, this would be reflected in higher AQI readings. Increasing monitoring intensity specifically during the Diwali period would be a practical and meaningful policy response.

4.4 Comparing the Three Stations

When we compare the predictions of all three stations, a clear pattern of intra-city inequality emerges.

Year	MMMUT	JAL KAL	GIDA
2026	95.57 — Satisfactory	174.46 — Moderate	260.53 — Poor
2027	93.87 — Satisfactory	174.26 — Moderate	265.23 — Poor

Table 4.4: Predicted 2026 and 2027 Diwali AQI for all three Gorakhpur stations compared

The difference in predictions between MMMUT and GIDA in 2026 is about 165 units. As a ratio, GIDA is 2.73 times higher. On the same Diwali night, in the same city, with the same weather, pollution near GIDA is nearly three times higher than inside the MMMUT campus. Both experience the same general weather. Both see the same Diwali celebrations. But their air quality is completely different.

JAL KAL sits in the middle at 174, which makes sense for a dense urban area. It is nearly double MMMUT but about two-thirds of GIDA.

This means that using a single city-wide AQI figure for Gorakhpur is misleading. Averaging MMMUT and GIDA gives approximately 178, which falls in the Moderate category and hides the fact that people near GIDA are experiencing Poor air quality while those at MMMUT are in Satisfactory conditions.

4.5 Key Observations

Several key observations emerge from this analysis:

- Within Gorakhpur during the Diwali period, pollution variation across locations is larger than most people realize. A gap of 165 AQI units within one city on one night is significant.
- Pollution levels at MMMUT have been improving slowly, and if this pattern continues, they will remain within an acceptable range. The campus

environment—limited traffic, green cover, lower firecracker use—is the likely reason. This is the positive finding in the data.

- GIDA shows no sign of improvement without intervention. The combination of industrial growth and Diwali-related emissions creates a situation that is becoming more serious year by year.

V. CONCLUSION AND FUTURE SCOPE

5.1 Summary of Findings

This study began with a simple question: can nine years of Diwali AQI data from Gorakhpur monitoring stations reveal patterns useful for predicting future years? Based on what we observed, the answer is yes. The trend direction at individual stations is identifiable, and predictions for 2026 and 2027 are consistent with prior patterns.

The three stations tell three different stories. MMMUT is slowly improving, and it is expected to remain in a good range in 2026 and 2027. JAL KAL has no meaningful trend; the prediction of approximately 174 for both years represents a long-term average, while actual values will go up or down depending on weather conditions. GIDA is the most problematic: highest every year, trending upward, and predicted to remain in the Poor range in 2026 and 2027. If this trend continues, GIDA could reach Very Poor territory during Diwali within a few years.

The 165-unit gap between MMMUT and GIDA on the same Diwali night is this study's most important finding. It demonstrates that using a single city-wide AQI average is misleading and that location within Gorakhpur matters enormously for Diwali air quality exposure.

From a practical standpoint, the data suggests several interventions. Before Diwali, factories within the GIDA zone should be audited to identify which ones contribute most to baseline pollution. During the festival period, the most polluting factories should be required to reduce output, so that when Diwali firecracker emissions are added, the combined peak is lower. The public, especially those near GIDA, should be clearly informed that air quality on Diwali night will be poor, and should be advised to close windows and doors and wear N95 masks when going outside.

5.2 Limitations of This Study

Several limitations should be noted:

- The model uses only Year as a predictor, not weather variables such as temperature, wind speed, and humidity. These factors heavily influence the variation in pollution level from year to year, particularly at JAL KAL, where weather-driven variation is most visible.
- We used only the peak AQI for the five-day Diwali period. This captures the most serious pollution level but does not provide information about pollution buildup and dispersal patterns across the five days.
- The model is trained on past data, making predictions for 2026 and 2027 less reliable the further they are from the training period. These numbers describe where the trend is going, not guaranteed future values.
- Only three stations were covered. Other parts of Gorakhpur—including the old city near Golghar, the area near the railway station, and residential areas adjacent to the GIDA zone—are not represented in this study.

5.3 Scope for Future Work

Several directions for future work would significantly improve on this study:

- Adding weather data (temperature, humidity, wind speed from IMD Gorakhpur) would be the most valuable improvement, particularly for JAL KAL where weather-driven variation currently dominates.
- Using daily data for October to November rather than only the Diwali peak would increase the dataset size and allow the use of more advanced models such as ARIMA or LSTM networks.
- Expanding coverage to include more monitoring stations in Gorakhpur, and including similar cities in eastern UP such as Varanasi, Ayodhya and Basti, would allow zone-wise comparison across the region.
- Developing a publicly accessible website where people can see predicted Diwali AQI by station for the coming year would bring this research out of academic papers and into practical use for residents.
- Disentangling the effects of firecracker restrictions from weather conditions would help clarify whether current regulations are actually reducing pollution

or whether observed improvements in some years are mainly due to favourable weather.

ACKNOWLEDGEMENT

Every group member who helped with this project deserves recognition, because a piece of work like this does not come together without collective support.

Firstly, thanks go to Er. Prateek Shahi, Head of Department, Civil Engineering, KIPM College of Engineering and Technology. He guided the entire project, especially in the early stages when we were not sure how to proceed. His guidance on methodology and interpretation was invaluable, and he gave considerable time to ensure this project was satisfactory.

Group members also thank the faculty of the Civil Engineering Department at KIPM. The subjects studied, including Environmental Engineering and Data Analysis, are directly connected to this project. This project made those connections very clear in practice.

The data came from CPCB and UPPCB monitoring portals, both available at no cost. Without freely accessible government data, a student-level project of this kind would not be possible. Group members did not require institutional funding.

We also acknowledge the value of working as a group. During the few difficult points in this project—mainly when data was missing or results were initially confusing—group discussion made progress much easier. The combined effort and mutual encouragement throughout the semester made this work possible.

REFERENCES

[1] Ravindra, K., Mor, S., & Kaushik, C.P. (2003). Impact of firecrackers on air quality during Diwali night: A case study. Journal of Environmental Monitoring, 5(2), 260–264.

[2] Barman, S.C., Singh, R., Negi, M.P.S., & Bhargava, S.K. (2008). Ambient air quality of Lucknow during Diwali festival. Environmental Monitoring and Assessment, 137(1–3), 495–504.

[3] Singh, D.P., Gadi, R., Mandal, T.K., Dixit, C.K., Singh, K., Saud, T., & Saxena, M. (2010). Study of air quality changes during Diwali period. Environmental Monitoring and Assessment, 169(1–4), 1–13.

[4] Gu, K., Zhou, Y., Sun, H., Zhao, L., & Liu, S. (2020). Air quality prediction using neural network algorithm. Neural Computing and Applications, 32(7), 1975–1982.

[5] Sharma, N., Agarwal, A.K., Eastham, S., & Shukla, P.R. (2021). Air pollution and its control. In Air Pollution and Control (pp. 1–23). Springer, Singapore.

[6] Pandey, A.K., Singh, S.K., & Tiwari, A. (2022). Prediction of air pollution during Diwali using AI techniques. Environmental Science and Pollution Research, 29(15), 21002–21013.

[7] CPCB (2014). National Air Quality Index: Launch document. Central Pollution Control Board, Ministry of Environment, Forest and Climate Change, Government of India, New Delhi.

[8] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[9] WHO (2021). WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulphur Dioxide and Carbon Monoxide. World Health Organization, Geneva.

[10] UPPCB (2023). Air Quality Monitoring Reports – Gorakhpur District. Uttar Pradesh Pollution Control Board, Lucknow. Available at: uppcb.up.nic.in

APPENDIX A: PYTHON CODE

The code below was used to build the Linear Regression models and generate the AQI predictions for 2026 and 2027. It was written and run in Python 3.10 using Jupiter Notebook on a Windows system.

```
import pandas as pd import NumPy as np import matplotlib.pyplot as plt from sklearn.linear_model import Linear Regression   # Load the dataset data = pd.read_excel('AQI_prediction.xlsx')   data['year'] = [2016,2017,2018,2019,2020,2021,2022,2023,2024,2025]  # Set predictor variable X = data[['YEAR']] stations = ['MMMUT', 'JAL KAL', 'GIDA'] model = Linear Regression()  # Train and predict for each station for s in stations: y = data[s]    model. Fit(X, y)    preds = model. Predict ([[2026], [2027]])    print(f'{s}: 2026={round(preds [0],2)}, 2027={round(preds[1],2)}')  # Output: # MMMUT:
```

2026 = 95.57, 2027 = 93.87 # JAL KAL: 2026 = 174.46, 2027 = 174.26 # GIDA: 2026 = 260.53, 2027 = 265.23

APPENDIX B: INDIAN AQI CATEGORY CHART

For reference, the Indian National AQI scale as defined by CPCB uses the following categories

AQI Range	Category	Health Implications
0 – 50	Good	Minimal impact
51 – 100	Satisfactory	Minor discomfort for sensitive people
101 – 200	Moderate	Discomfort for people with respiratory or heart conditions
201 – 300	Poor	Breathing discomfort for most people on prolonged exposure
301 – 400	Very Poor	Serious respiratory illness for most people
401 – 500	Severe	Health emergency, affects everyone seriously

Table B.1: Indian National AQI Categories and Health Implications. Source: CPCB (2014).