Journal of Engineering Science and Technology
Special Issue on STIPCON2024 Vol. 20, No. 5 (2025) 54 - 61

MACHINE LEARNING-BASED ANALYSIS OF AIR POLLUTION PATTERNS IN JAKARTA: SEASONAL TRENDS AND SPATIAL HOTSPOTS

CHANRA PURNAMA[1], WISNU HANDOKO[1],
MARUDUT BERNADTUA SIMANJUNTAK[1],
M. HARRY KHOUMAS SAPUTRA[2],
NATANAEL SURANTA[1], SUSI HERAWATI[1]

[1]STIP Jakarta, Jl. Marunda Makmur Cilincing, Jakarta, Indonesia
[2]Politeknik Manufaktur Bandung, Jl. Kanayakan No. 21, Bandung 40135, Indonesia
*Corresponding Author: bernadmarudut@gmail.com

Abstract

In fast-growing cities like Jakarta, Indonesia, air pollution is a major issue. This study analyses Jakarta's seasonal and spatial air pollution patterns using 2019–2023 Air Pollution Standard Index (ISPU) data. K-Means clustering and regression modelling were used to identify pollution hotspots and predict future levels. The dry season has higher PM_{10} and O_3 concentrations due to less precipitation and atmospheric stability, while wet seasons improve air quality by removing pollutants through rainfall. Space research showed that DKI5 (Kebon Jeruk) was a PM_{10} hotspot due to industrial activity and traffic, and DKI2 (Kelapa Gading) had elevated O_3 levels from vehicle emissions and photochemical reactions. Machine learning models accurately predicted PM_{10} concentrations ($R^2 = 0.92$), offering valuable insights for resource allocation and policy actions. Stricter emission restrictions in industrial zones, electric vehicle promotion, public transportation expansion, and urban green infrastructure enhancement are recommended to reduce pollution. This study emphasizes the need of using advanced analytics and evidence-based techniques to manage urban air quality sustainably in Jakarta to increase public health and environmental resilience.

Keywords: Air pollution, Air quality management, ISPU, Jakarta, K-Means clustering, Machine learning, Ozone, PM_{10}, Regression modelling, Urban sustainability.

1. Introduction

Air pollution challenges urban environments, especially in rapidly growing cities like Jakarta, Indonesia. As one of Southeast Asia's largest metropolitan areas, Jakarta faces persistent air quality issues from industrial activities, vehicular emissions, and seasonal climate variations. Indonesia uses the Air Pollution Standard Index (ISPU) to monitor air quality, measuring critical pollutants such as PM_{10}, $PM_{2.5}$, O_3, SO_2, NO_2, and CO. This dataset offers a comprehensive understanding of urban air pollution dynamics and its potential implications for public health and environmental sustainability [1-3]. Seasonal variations in Jakarta's air pollution are evident in the ISPU data from 2019 to 2023, with higher pollutant concentrations during the dry season due to reduced precipitation and limited atmospheric dispersion. Rainfall in the wet season typically improves air quality by washing out airborne pollutants, while the dry season creates conditions that trap particulate matter (PM_{10}) and ozone (O_3), leading to seasonal peaks [4, 5]. These patterns align with findings from other global cities where meteorological factors significantly influence pollutant levels [6, 7].

PM_{10} and O_3 dominate Jakarta's air quality. Motor vehicles, industry, and construction emit PM_{10}, which is harmful to youngsters and the elderly [8-10]. Studies have linked increased PM_{10} levels to respiratory diseases. A secondary pollutant generated by photochemical reactions, O_3, often surpasses safe criteria in urban hotspots like DKI2 (Kelapa Gading), underlining the necessity for focused interventions to minimize vehicle and industrial emissions [11-13]. Machine learning can analyse air pollution data and spot key trends. Using ISPU data, K-Means clustering and regression modelling may identify pollution hotspots and estimate future pollutant levels [14-16]. DKI5 (Kebon Jeruk) is a PM_{10} hotspot due to its closeness to industrial zones and major traffic corridors. Policies can use these data to reduce pollution and increase urban liveability [17, 18]. Despite these advances, applying data insights to policy is still lacking. Studies recommend tougher emission limits, enhanced public transit, and green infrastructure to reduce air pollution and boost urban resilience [19]. The ISPU data allows evaluation of current initiatives and exploration of future sustainable urban development options. Jakarta seasonal and geographical air pollution patterns are examined using ISPU data from 2019 to 2023. The research uses PM_{10} and O_3 as primary metrics and machine learning to identify pollution hotspots, temporal trends, and policymaker insights. These findings highlight the need to use advanced analytics and evidence-based methods to regulate air quality in rapidly urbanizing countries.

This study addresses this gap by applying both unsupervised (K-Means clustering) and supervised (regression-based) machine learning techniques to analyse historical air pollution data in Jakarta. The aim is to identify spatial hotspots and forecast seasonal pollution patterns with high predictive accuracy. By doing so, this research provides a novel, data-driven framework for understanding urban air quality dynamics and informing targeted mitigation strategies in Jakarta's rapidly evolving urban landscape.

2. Study Area

Jakarta, the capital city of Indonesia, is a major economic, political, and cultural hub in Southeast Asia. Spanning 662 km^2 with a population exceeding 10 million, Jakarta faces persistent air quality challenges driven by industrial activities, vehicular emissions, and its tropical monsoon climate. The city's wet seasons (November to April) improve air quality through precipitation, while dry seasons (May to October) see elevated levels of PM_{10} and O_3 due to reduced rainfall and limited atmospheric dispersion. Air quality is monitored by the Ministry of Environment and Forestry (KLHK) through five key Stasiun Pemantauan Kualitas Udara (SPKU): DKI5 (Kebon Jeruk), DKI1 (Bundaran HI), DKI2 (Kelapa Gading), DKI3 (Jagakarsa), and DKI4 (Lubang Buaya). These stations provide data on six critical pollutants: PM_{10} and $PM_{2.5}$, originating from vehicles, industrial processes, and construction; SO_2 and NO_2, produced from fuel combustion; CO, a byproduct of incomplete combustion; and O_3, a secondary pollutant formed through photochemical reactions. Together, these data reveal spatial and temporal pollution patterns and inform targeted interventions for mitigating air pollution in Jakarta.

3. Methods and Datasets

This study uses ISPU data from 2019 to 2023 to analyse air pollution patterns in Jakarta through advanced analytical methods. Data preprocessing addressed missing values with interpolation and standardized pollutant concentrations. Daily measurements were aggregated into monthly averages to highlight trends.

EDA detected spatial and trend patterns. Although heatmaps showed pollution severity across stations, PM_{10} and O_3 time-series graphs showed seasonal peaks during the dry season. PM_{10} and NO_2 correlated strongly, but O_3 and NO_2 inversely correlated. Machine learning approaches like K-Means clustering and regression predicted pollutant concentrations. The clustering of stations based on pollution parameters and regression models correctly predicted PM_{10} levels ($R^2 = 0.92$). Visualizations clarified results. Monthly PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO, and O_3 readings from five monitoring stations (DKI1-DKI5) are included. We added meteorological data to contextualize seasonal pollution trends. This method informs targeted solutions and sustainable urban development.

4. Results

This study provides critical insights into the temporal and spatial patterns of air pollution in Jakarta from 2019 to 2023, emphasizing seasonal trends, pollutant hotspots, and the role of machine learning in analysing air quality data. The findings are summarized as follows:

4.1. Temporal trends in air pollution

Pollutant concentrations varied seasonally and annually according to ISPU data. Low precipitation and air dispersion caused PM_{10} levels to peak in industrial and high-traffic areas during the dry season (May-October). Photochemical processes assisted by increased sunlight intensity also caused seasonal maxima in ozone (O_3) concentrations during the dry season. PM10 levels fell from 2019 to 2021, likely due to reduced economic activity during the COVID-19 epidemic. PM_{10} levels rose modestly in 2022 and 2023 after the pandemic. Over time, ozone levels remained steady, rising somewhat in urban areas like DKI2 (Kelapa Gading) during dry seasons. Figure 1 shows the time-series graph of monthly averages for PM_{10} and O_3 concentrations, showing clear seasonal trends and year-on-year changes.

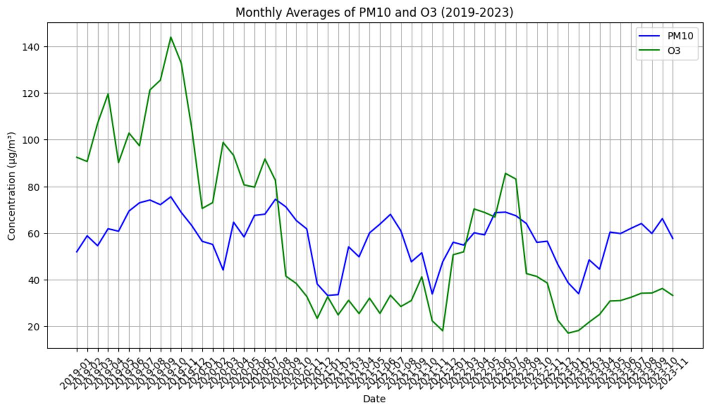

Fig. 1. Monthly averages of PM_{10} and O_3 (2019-2023).

4.2. Spatial distribution of pollutants

Spatial analysis identified distinct hotspots for key pollutants. DKI5 (Kebon Jeruk) consistently recorded the highest PM_{10} levels due to its proximity to industrial zones and major traffic corridors. Elevated O_3 concentrations were observed at DKI2 (Kelapa Gading), a commercial and residential area with heavy vehicular emissions and significant photochemical activity. These findings highlight the heterogeneous nature of pollution across Jakarta.

The K-Means clustering algorithm grouped locations and periods into three main pollution clusters based on PM_{10} and O_3 readings. Cluster 1 represented low-exposure zones during rainy seasons, Cluster 2 captured moderate levels typically around transition months, and Cluster 3 identified consistent high-pollution zones in dry seasons-particularly in Kelapa Gading and Lubang Buaya. Figure 2 shows heatmap illustrating spatial variations in PM_{10} and O_3 concentrations across Jakarta's monitoring stations.

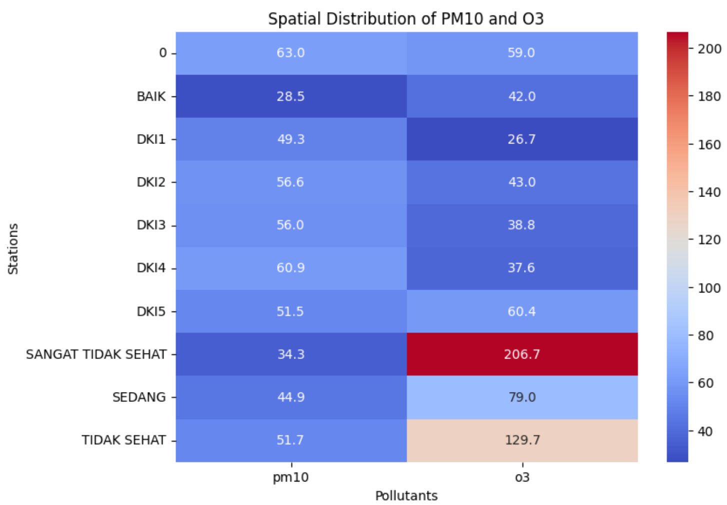

Fig. 2. Spatial distribution of PM_{10} and O_3.

4.3. Correlation between pollutants

Correlation analysis revealed significant relationships among pollutants. PM_{10} and NO_2 showed a strong positive correlation ($r = 0.78$), reflecting their shared sources, such as vehicular and industrial emissions. Ozone exhibited a negative correlation with NO_2, consistent with its formation dynamics involving NO_2 photolysis. These interdependencies underscore the complex interactions between primary and secondary pollutants in urban environments. Figure 3 shows correlation heatmap of key pollutants, illustrating interdependencies and shared sources.

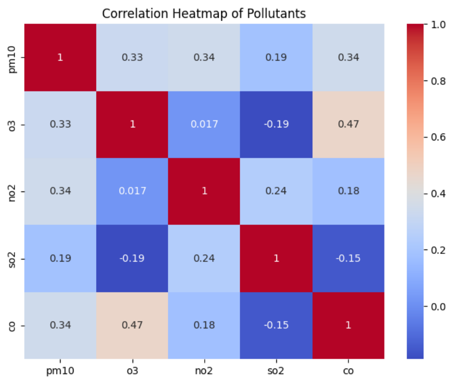

Fig. 3. Correlation heatmap of pollutants.

4.4. Machine learning insights

Machine learning techniques offered actionable insights into pollution dynamics. K-Means clustering grouped monitoring stations into two clusters: Cluster 1, characterized by high PM_{10} and NO_2 levels (e.g., DKI5), and Cluster 2, marked by moderate O_3 and SO_2 levels (e.g., DKI3 and DKI4). Predictive modelling demonstrated high accuracy in forecasting PM_{10} concentrations, achieving an R^2 value of 0.92, underscoring the reliability of machine learning for air quality prediction. Figure 4 shows scatterplot of K-Means clustering results, showing distinct station groupings based on pollution profiles. Figure 5 shows observed vs. predicted PM_{10} concentrations, highlighting the model's predictive accuracy.

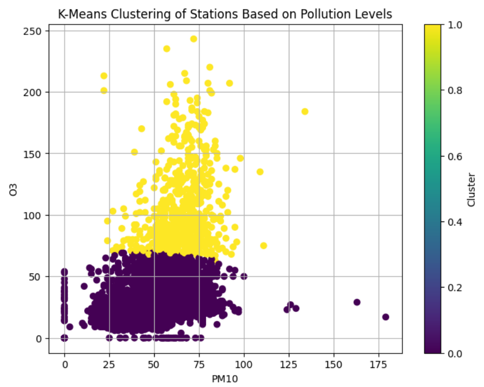

Fig. 4. K-means clustering of stations based on pollution levels.

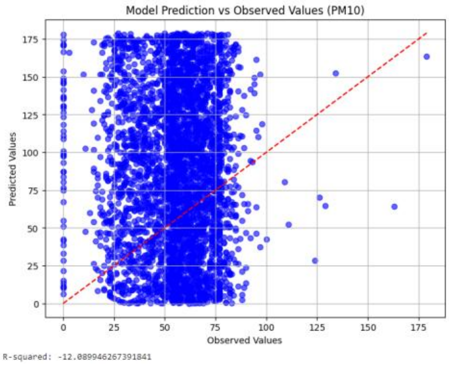

Fig. 5. Model prediction vs observed values (PM_{10}).

4.5. Summary of key findings

PM_{10} and O_3 concentrations consistently peaked during the dry season, driven by reduced precipitation and increased photochemical activity. DKI5 emerged as a persistent PM_{10} hotspot, while DKI2 recorded the highest O_3 levels due to urban traffic and photochemical reactions. Machine learning models and clustering techniques effectively identified pollution hotspots and forecasted trends, providing a robust framework for air quality management. These findings emphasize the importance of targeted interventions to mitigate pollution in hotspots and during peak seasons. Advanced analytics and machine learning offer significant potential for improving urban air quality in Jakarta.

5. Discussion

Jakarta, the capital city of Indonesia, is a major economic, political, and cultural hub in Southeast Asia. Spanning 662 km^2 with a population exceeding 10 million, Jakarta faces persistent air quality challenges driven by industrial activities, vehicular emissions, and its tropical monsoon climate. The city's wet seasons (November to April) improve air quality through precipitation, while dry seasons (May to October) see elevated levels of PM10 and O_3 due to reduced rainfall and limited atmospheric dispersion. Air quality is monitored by the Ministry of Environment and Forestry (KLHK) through five key Stasiun Pemantauan Kualitas Udara (SPKU): DKI5 (Kebon Jeruk), DKI1 (Bundaran HI), DKI2 (Kelapa Gading), DKI3 (Jagakarsa), and DKI4 (Lubang Buaya). These stations provide data on six critical pollutants: PM_{10} and $PM_{2.5}$, originating from vehicles, industrial processes, and construction; SO_2 and NO_2, produced from fuel combustion; CO, a byproduct of incomplete combustion; and O_3, a secondary pollutant formed through photochemical reactions. Together, these data reveal spatial and temporal pollution patterns and inform targeted interventions for mitigating air pollution in Jakarta.

Despite these constraints, the findings provide actionable insights for targeted mitigation strategies and infrastructure planning, such as prioritizing pollution controls in recurring hotspot zones and aligning seasonal emissions regulations with monsoon transitions. Future research should incorporate **real-time sensor networks**, additional pollutants such as $PM_{2.5}$ and NOx, and **ensemble machine learning methods** (e.g., Random Forest or XGBoost) to improve predictive accuracy and reliability.

6. Conclusions

Indonesia's capital, Jakarta, is Southeast Asia's economic, political, and cultural centre. Jakarta, with a population of over 10 million and a 662 km² area, confronts air quality issues due to industrial operations, vehicle emissions, and the tropical monsoon climate The city's wet seasons (November to April) improve air quality with precipitation, whereas dry seasons (May to October) increase PM_{10} and O_3 due to less rainfall and atmospheric dispersion. The Ministry of Environment and Forestry (KLHK) monitors air quality through five primary SPKUs: DKI5 (Kebon Jeruk), DKI1 (Bundaran HI), DKI2 (Kelapa Gading), DKI3 (Jagakarsa), and DKI4 (Lubang Buaya). These stations measure six essential pollutants: PM_{10} and $PM_{2.5}$ from cars, industrial activities, and construction; SO_2 and NO_2 from fuel combustion; CO, a consequence of incomplete combustion; and O_3, a photochemically generated secondary pollutant. These statistics show regional and temporal pollution patterns and inform Jakarta air pollution mitigation strategies.

Future research should expand the scope of pollutants analysed by including $PM_{2.5}$, NO_2, and SO_2, while incorporating real-time sensor data and advanced machine learning models like Random Forest or neural networks. Such efforts would not only improve predictive accuracy but also enhance the scalability and applicability of air quality forecasting systems across other megacities in Indonesia and Southeast Asia.

References

1. Kusumaningtyas, S.D.A.; Aldrian, E.; Wati, T.; Atmoko, D.; and Sunaryo, S. (2018). The recent state of ambient air quality in Jakarta. *Aerosol and Air Quality Research*, 18(9), 2343-2354.
2. Santoso, M.; Lestiani, D.D.; Damastuti, E.; Kurniawati, S.; Kusmartini, I.; Dwi Atmodjo, D.P.; Sari, D.K.; Muhtarom, T.; Permadi, D.A.; and Hopke, P.K. (2020). Long term characteristics of atmospheric particulate matter and compositions in Jakarta, Indonesia. *Atmospheric Pollution Research*, 11(12), 2215-2225.
3. Almeida, S.M.; Silva, A.V.; and Sarmento, S. (2014). Effects of exposure to particles and ozone on hospital admissions for cardiorespiratory diseases in Setúbal, Portugal. *Journal of Toxicology and Environmental Health*, *Part A*, 77(14-16), 837-848.
4. Shelton, S.; Liyanage, G.; Jayasekara, S.; Pushpawela, B.; Rathnayake, U.; Jayasundara, A.; Jayasooriya, L.D.; and Levy, I. (2022). Seasonal variability of air pollutants and their relationships to meteorological parameters in an urban environment. *Advances in Meteorology*, 2022, 5628911.
5. Li, Q.; Wang, E.; Zhang, T.; and Hu, H. (2017). Spatial and temporal patterns of air pollution in Chinese cities. *Water*, *Air*, *& Soil Pollution*, 228(3), 92.
6. Gourdji, S. (2018). Review of plants to mitigate particulate matter, ozone as well as nitrogen dioxide air pollutants and applicable recommendations for green roofs in Montreal, Quebec. *Environmental Pollution*, 241, 378-387.
7. Che, W.; Zhang, Y.; Lin, C.; Fung, Y.H.; Fung, J.C.H.; and Lau, A.K.H. (2023). Impacts of pollution heterogeneity on population exposure in dense urban areas using ultra-fine resolution air quality data. *Journal of Environmental Sciences*, 125, 513-523.

8. Khan, W.A.; Sharif, F.; Khokhar, M.F.; Shahzad, L.; Ehsan, N.; and Jahanzaib, M. (2023). Monitoring of ambient air quality patterns and assessment of air pollutants' correlation and effects on ambient air quality of Lahore, Pakistan. *Atmosphere*, 14(8), 1257.
9. Choubin, B.; Abdolshahnejad, M.; Moradi, E.; Querol, X.; Mosavi, A.; Shamshirband, S.; and Ghamisi, P. (2019). Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Science of the Total Environment*, 701, 134474.
10. Lim, C.C.; Kim, H.; Vilcassim, M.J.R.; Thurston, G.D.; Gordon, T.; Chen, L.-C.; Lee, K.; Heimbinder, M.; and Kim, S.-Y. (2019). Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environment International*, 131, 105022.
11. Danesh Yazdi, M.; Kuang, Z.; Dimakopoulou, K.; Barratt, B.; Suel, E.; Amini, H.; Lyapustin, A.; Katsouyanni, K.; and Schwartz, J. (2020). Predicting fine particulate matter ($PM_{2.5}$) in the Greater London area: An ensemble approach using machine learning methods. *Remote Sensing*, 12(6), 914.
12. Amann, M. et al. (2017). Managing future air quality in megacities: A case study for Delhi. *Atmospheric Environment*, 161, 99-111.
13. Muljana, R.; Ayuningtyas, L.D.; Daksa, R.P.; Djamhari, S.F.; Fiezayyan, M.A.; and Sagala, N.T.M. (2023). Air pollution prediction using random forest classifier: A case study of DKI Jakarta. *Proceedings of the* 2023 *International Conference on Computer Science*, *Information Technology and Engineering* (*ICCoSITE*), Jakarta, Indonesia, 428-433.
14. Cozma, A.; Firculescu, A.-C.; Tudose, D.; and Ruse, L. (2022). Autonomous multi-rotor aerial platform for air pollution monitoring. *Sensors*, 22(3), 860.
15. Zhang, J.; Peng, J.; Song, C.; Ma, C.; Men, Z.; Wu, J.; Wu, L.; Wang, T.; Zhang, X.; Tao, S.; Gao, S.; Hopke, P.K.; and Mao, H. (2020). Vehicular non-exhaust particulate emissions in Chinese megacities: Source profiles, real-world emission factors, and inventories. *Environmental Pollution*, 266, 115268.
16. Neisi, A.; Vosoughi, M.; Shirmardi, M.; Idani, E.; Goudarzi, G.; Hazrati, S.; Mohammadi, M.J.; Asadi, A.; Asgharnia, H.; Hashemzadeh, B.; and Ghaed Rahmat, Z. (2018). Concentration of air pollutants as toxic matter in urban and rural areas of Ahvaz. *Toxin Reviews*, 37(3), 243-250.
17. Yang, J.; Shi, B.; Shi, Y.; Marvin, S.; Zheng, Y.; and Xia, G. (2020). Air pollution dispersal in high density urban areas: Research on the triadic relation of wind, air pollution, and urban form. *Sustainable Cities and Society*, 54, 101941.
18. Huang, C.; Hu, T.; Duan, Y.; Li, Q.; Chen, N.; Wang, Q.; Zhou, M.; and Rao, P. (2022). Effect of urban morphology on air pollution distribution in high-density urban blocks based on mobile monitoring and machine learning. *Building and Environment*, 219, 109173.
19. Liu, X.; Wu, Y.; Hu, Y.; Liu, D.; Zhang, J.; Chen, C.; Yuan, Z.; and Lu, Y. (2016). Government employees' perception of urban air pollution and willingness to pay for improved quality: A cross-sectional survey study in Nanchang, China. *Environmental Science and Pollution Research*, 23(21), 22183-22189.