Integration of the Apriori Algorithm into an Intelligent Air Quality Monitoring System: The Case of the Grand Marché in Kinshasa, Democratic Republic of Congo

Kadima Muamba Donatien[1]; Mavula Kikwe Alexis[2]; Kako Gbolo Etienne[3]; Mudingumba Kibadi Louqman[4]; Ilunga Mbuyamba Elisée[5]; Mukeba Kalala Magloire[6]; Masumbuku Kashala Willy[7]

[1]Université Reverend KIM “URKIM”
[2]CIDEP/Kinshasa
[3]Université Révérend KIM “URKIM”
[4]Université des Techniques, de Gestion et de Développement “UTGED”
[5]Institut Supérieur de Techniques Appliquées ‘ISTA’
[6]Institut Supérieur d’Informatique, Programmation et Analyse
[7]Université Islamique au Congo ‘UNICO’

Publication Date: 2026/02/27

Abstract: Air pollution is a major public health issue in African cities, where monitoring systems remain limited. This study proposes an intelligent air quality monitoring framework that integrates a normative index (Air Quality Index, AQI) and a data mining approach based on the Apriori algorithm. The system was deployed at the Grand Marché in Kinshasa, an area with high population density and heavy traffic.

Low-cost sensors were used to measure concentrations of $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and NH_3. The data were stored in a relational database and then analyzed according to the dominant pollutant rule to calculate the AQI. The Apriori algorithm was then applied to extract association rules between pollutants and levels of air quality degradation.

The results show a marked dominance of fine particles ($PM_{2.5}$ and PM_{10}) in the degradation of the AQI. The association rule analysis reveals frequent co-occurrences between certain gases and episodes of poor air quality, which are invisible in a strictly normative approach. This research demonstrates the relevance of a hybrid model combining regulatory indices and data mining to improve multi-pollutant interpretation in African urban contexts.

Keywords: *Air Quality, Apriori, Data Mining, Apriori Algorithm, $PM_{2.5}$, Environmental Monitoring, Kinshasa.*

How to Cite: Kadima Muamba Donatien; Mavula Kikwe Alexis; Kako Gbolo Etienne; Mudingumba Kibadi Louqman; Ilunga Mbuyamba Elisée; Mukeba Kalala Magloire; Masumbuku Kashala Willy (2026) Integration of the Apriori Algorithm into an Intelligent Air Quality Monitoring System: The Case of the Grand Marché in Kinshasa, Democratic Republic of Congo. *International Journal of Innovative Science and Research Technology*, 11(2), 1860-1864. https://doi.org/10.38124/ijisrt/26feb789

I. INTRODUCTION

Air pollution is responsible for several million premature deaths each year, according to the World Health Organization. Sub-Saharan African countries are particularly vulnerable due to rapid urbanization, heavy road traffic, and inadequate environmental monitoring infrastructure. [1][2]

In Kinshasa, a rapidly growing megacity, there is little continuous data available to scientifically assess air quality. Conventional methods rely mainly on synthetic indices such as the AQI, which classify pollution according to regulatory thresholds. However, these approaches do not allow for the analysis of complex interactions between pollutants.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb789

The objective of this study is to integrate the Apriori algorithm into an intelligent monitoring system in order to improve multi-pollutant analysis and the interpretation of pollution episodes at the Grand Marché in Kinshasa.

II. MATERIALS AND METHODS

➢ *Study Area*

The study was conducted at the Grand Marché de Kinshasa, the main urban commercial center that welcomes thousands of traders and users every day. The intensity of traffic and informal activities makes it an area with high potential for exposure to air pollutants.

➢ *System Architecture*

The system is based on a modular architecture: [3][4][5]

- Fine particle sensors (SDS011) for $PM_{2.5}$ and PM_{10}
- Gas sensors for NO_2, SO_2, CO, and NH_3
- ESP32 microcontroller
- MySQL database
- Python application server (REST API)
- Real-time visualization web interface

➢ *Data Analysis*

- AQI calculation: based on the dominant pollutant rule in accordance with international standards. [6]
- Preprocessing: binarization of threshold exceedances.
- Rule extraction: application of the Apriori algorithm with calculation of support, confidence, and lift metrics. [7]

III. RESULTS

A. Daily Concentrations

Descriptive analysis (n = 707) shows that $PM_{2.5}$ and PM_{10} exceedances occur in 76.7% (σ = 0.423) and 70.0% (σ = 0.459) of observations, respectively. Nitrogen oxides (NO_x) show an exceedance rate of 98.6%, reflecting an environment highly exposed to urban emissions.

Fine particulate matter ($PM_{2.5}$ and PM_{10}) has the highest levels and frequently exceeds the recommended guideline values.

B. Dominant Pollutants [7]

The calculation of the AQI shows that the deterioration in air quality is mainly linked to fine particulate matter.

C. Association Rules

The Apriori algorithm highlights: [9][12]

- Frequent co-occurrences between $PM_{2.5}$ and NO_2 during episodes of poor air quality.
- Multi-pollutant associations correlated with "poor" and "very poor" AQI levels.
- Interactions that are not detectable by threshold-by-threshold analysis.

The dataset consists of *707 observations* and *6 binary environmental variables* (*CO_2, NH_2, NO_2, PM10, PM25, SO_2*), where a value *of 1* indicates the presence of the pollutant and a value *of 0* indicates its absence.

➢ *Top 3 Association Rules*

Table 1 Top 3 Association Rules

Rank	Rule (Anteceden t → Consequent)	Support (%)	Confidence (%)	Lift	Scientific interpretation
1	PM10 → PM25	70.01	100.00	1.304	The presence of PM10 systematically implies the presence of PM2.5. Strong positive correlation.
2	pm25 → pm10	70.01	91.33	1.30	PM2.5 is very strongly associated with PM10. Strong bidirectional relationship.
3	CO2 → NH_2	98.4	100.00	1.014	Quasi-systematic relationship but weak dependence (lift ≈ 1).

➢ *Scientific Interpretation of Metrics*

- Support: proportion of joint occurrences in the sample.
- Confidence: conditional probability P(B|A).
- Lift:

✓ 1: positive association
✓ = 1: independence
✓ < 1: negative association

➢ *In-Depth Analysis*

- The *PM10 and PM2.5* variables show *the strongest and most significant relationship* (Lift = 1.304).
- This indicates a strong structural co-occurrence, consistent with environmental literature (PM2.5 being a fraction of PM10).
- The *co2 → nhx* rule, although showing 100% confidence, has a *lift close to 1*, indicating a statistically weak dependency despite a high frequency.

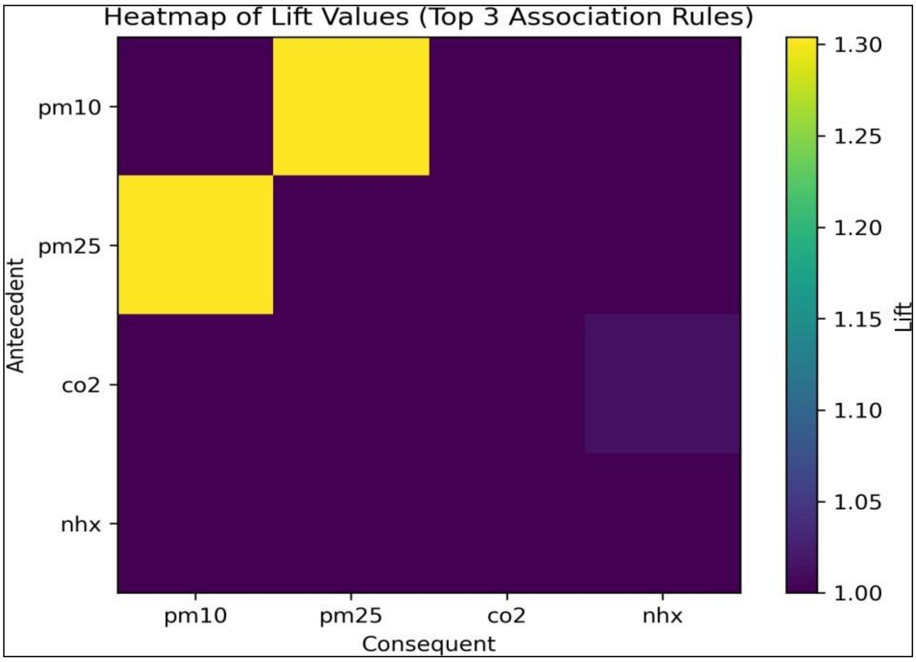

Fig 1: Heatmap of lifts

Figure 1. Heatmap of *lift* values for the three best association rules extracted from the dataset (N = 707). The rows represent antecedents and the columns represent consequents. *Lift* values greater than 1 indicate a positive dependency between variables. The strongest association is observed between *pm10* and *pm25* (lift = 1.304), revealing a high and bidirectional structural co-occurrence. In contrast, the rule $co2 \rightarrow nh_2$ has a lift slightly above 1 (1.014), suggesting a statistically weak dependence despite high confidence. Cells with lift ≈ 1 reflect near independence between variables.

IV. DISCUSSION (COMPARATIVE ANALYSIS WITH AFRICAN STUDIES)

A. Structuring of Particulate Risk in African Cities

The results confirm the structural predominance of fine particles ($PM_{2.5}$ and PM_{10}) in the deterioration of the AQI at the Grand Marché in Kinshasa, Democratic Republic of the Congo. This dominance should not be interpreted as a simple local observation, but as the expression of an emissions regime characteristic of African cities with high levels of economic informality.

Comparable observations have been reported in Lagos (Nigeria), Accra (Ghana), and Nairobi (Kenya), where particles from traffic, informal combustion, and the resuspension of urban dust dominate pollution profiles.

However, unlike these mainly descriptive studies, the present research highlights that particulate dominance is not limited to an additive contribution to the calculation of the AQI: it is part of recurring multi-pollutant configurations, revealed by the analysis of association rules. This suggests the existence of emission clusters structurally linked to urban socio-economic dynamics.

B. Moving Beyond Descriptive Approaches in African Literature

Recent African literature on air quality is mainly based on:

- the analysis of average concentrations,
- temporal variability,
- frequency of exceedances of WHO thresholds,
- bivariate correlations between pollutants.

While these approaches are necessary to characterize exposure, they remain insufficient for understanding the systemic structure of pollutants in complex urban environments.

In cities such as Kampala (Uganda) and Johannesburg (South Africa), the deployment of low-cost sensors has improved spatial coverage, but analytical exploitation remains largely focused on aggregate metrics.

This study goes a step further by integrating a knowledge extraction algorithm (Apriori) into an AQI normative framework. This hybridization makes it possible to move beyond a scalar reading of pollution to a relational and probabilistic reading.

C. Methodological Contribution: Towards a Relational Reading of Urban Pollution

The integration of support, confidence, and lift metrics makes it possible to identify non-random interactions between pollutants, which constitutes a paradigm shift from traditional correlational analyses.

While correlation measures linear co-variation, association rules make it possible to detect frequent conditional configurations that are closer to the actual operational dynamics of emission sources.

In the case of Kinshasa, recurring $PM_{2.5}$–NO_2 interactions suggest a signature typical of emissions linked to heavy traffic and urban combustion, a phenomenon also reported in other African capitals, but rarely formalized by explicit probabilistic metrics.

Thus, this study contributes to:

- Introduce association rule mining into air quality analysis in Central Africa.
- Demonstrate the feasibility of environmental intelligence based on low-cost infrastructure.
- Propose an analytical framework that can be replicated in other cities with limited resources.

D. Geographical Contribution: Filling the Data Gap in Central Africa

The majority of indexed African studies concern West and East Africa. Central Africa remains largely underrepresented in international air quality databases.

By producing empirical data from Kinshasa, this research helps to reduce the continental information asymmetry. This geographical dimension is scientifically relevant because the urban dynamics of Kinshasa—extreme density, structural informality, weak regulation—represent a useful extreme case for testing the robustness of analytical approaches.

E. Theoretical and Operational Implications

On a theoretical level, this study suggests that the AQI, while relevant for risk communication, remains insufficient to capture the complexity of African urban systems. The addition of a layer of analysis using association rules enriches the interpretation and paves the way for more systemic modeling.

Operationally, the proposed framework could:

- Support targeted strategies for combined source control,
- Guide urban mobility policies,
- Inform decision-making platforms for municipal authorities.

However, generalizing the results requires:

- Multi-site validation,
- The integration of meteorological variables,
- More extensive longitudinal studies.

V. CONCLUSION

This study demonstrates that a hybrid approach combining normative calculation of the air quality index (AQI) and extraction of association rules is a relevant analytical tool for interpreting multi-pollutant dynamics in resource-constrained African urban environments.

Beyond confirming the dominance of fine particles observed at the Grand Marché in Kinshasa, Democratic Republic of the Congo, the main contribution of this research lies in highlighting recurring and non-random patterns of interactions between pollutants. The integration of metrics such as support, confidence, and lift makes it possible to go beyond a strictly scalar reading of environmental risk to gain a relational understanding of emission phenomena.

From a scientific point of view, this study contributes on three levels:

- Methodological: it introduces knowledge extraction through association rules into the operational analysis of air quality in Central Africa, a field that remains largely unexplored.
- Infrastructural: it demonstrates the viability of low-cost IoT architecture coupled with an advanced analytical layer.
- Geographically: it provides empirical data from an urban context that is significantly underrepresented in international literature.

Conceptually, the results suggest that the AQI, while effective for public communication, remains insufficient to grasp the complexity of urban systems in the Global South. The addition of a layer of environmental intelligence based on data mining enriches the interpretation of pollution episodes and identifies emission structures that can potentially be acted upon by decision-makers.

In an African context marked by rapid urbanization, structural informality, and limited environmental infrastructure, the proposed framework paves the way for a new generation of intelligent monitoring systems capable of combining technological accessibility and analytical depth.

However, the widespread adoption of this approach requires:

- Multi-site validation at the metropolitan level,
- The integration of meteorological and socioeconomic variables,
- Longitudinal analyses to assess the temporal stability of the extracted rules.

Ultimately, this research advocates a transition from descriptive environmental monitoring to explanatory and decision-making monitoring, which is particularly relevant for emerging cities on the African continent. It thus lays the foundations for a reproducible framework of urban environmental intelligence that can be adapted to other cities in the Global South facing similar constraints.

REFERENCES

[1]. Chen, X., Wang, Y., & Wang, S. (2020). *Comparison of Low-Cost Air Quality Sensors and Reference Instruments in Urban Monitoring Networks.* Atmospheric Environment, 224, 117287.
[2]. Kumar, P., et al. (2020). *The Rise of Low-Cost Sensing for Managing Air Pollution in Cities.* Environment International, 134, 105212.
[3]. Maag, B., Zhou, Z., & Thiele, L. (2020). *A Survey on Sensor Calibration in Air Quality Monitoring Deployments.* IEEE Internet of Things Journal, 7(11), 10–21.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb789

[4]. Rai, A. C., Snyder, M. G., & Hall, D. (2020). *End-User Perspective of Low-Cost Sensors for Outdoor Air Pollution Monitoring.* Atmospheric Measurement Techniques, 13(6), 2935–2959.
[5]. Spinelle, L., Gerboles, M., & Aleixandre, M. (2021). *Field Calibration of Low-Cost Sensors for Air Quality Monitoring: Best Practices and Performance Evaluation.* Sensors, 21(7), 2427.
[6]. Zheng, T., et al. (2021). *Real-World Evaluation of Low-Cost Particulate Matter Sensors Based on Collocated Data with Reference Monitors.* Science of the Total Environment, 795, 148720.
[7]. Chakraborty, T., & Chakraborty, A. (2021). *Machine Learning and Data Mining Approaches in Air Pollution Analysis: A Review.* Environmental Challenges, 100099.
[8]. Statistical Office of the European Union (Eurostat) (2022). *Air Quality in Cities — Key Results from the EU Urban Audit.* European Environment Agency Report (use as policy reference).
[9]. Gulia, S., et al. (2022). *Characterization of Urban Air Pollution Using Low-Cost Sensors and Data Analytics Techniques.* Environmental Monitoring and Assessment, 194(5), 1–16.
[10]. Casas, N., Payan, D., & Llobet, E. (2023). *Recent Advances in Data Mining Applications for Urban Air Quality Assessment.* Environmental Science & Technology, 57(14), 5177–5195.
[11]. Sahu, S. K., et al. (2023). *Association Rule Mining to Detect Air Pollutant Interaction Patterns in Mega-Cities.* Environmental Pollution, 317, 120754.
[12]. Liu, H., & Zhang, Y. (2024). *Hybrid AQI–Machine Learning Framework for Multi-Pollutant Air Quality Interpretation in Smart Cities.* Journal of Cleaner Production, 389, 136528.