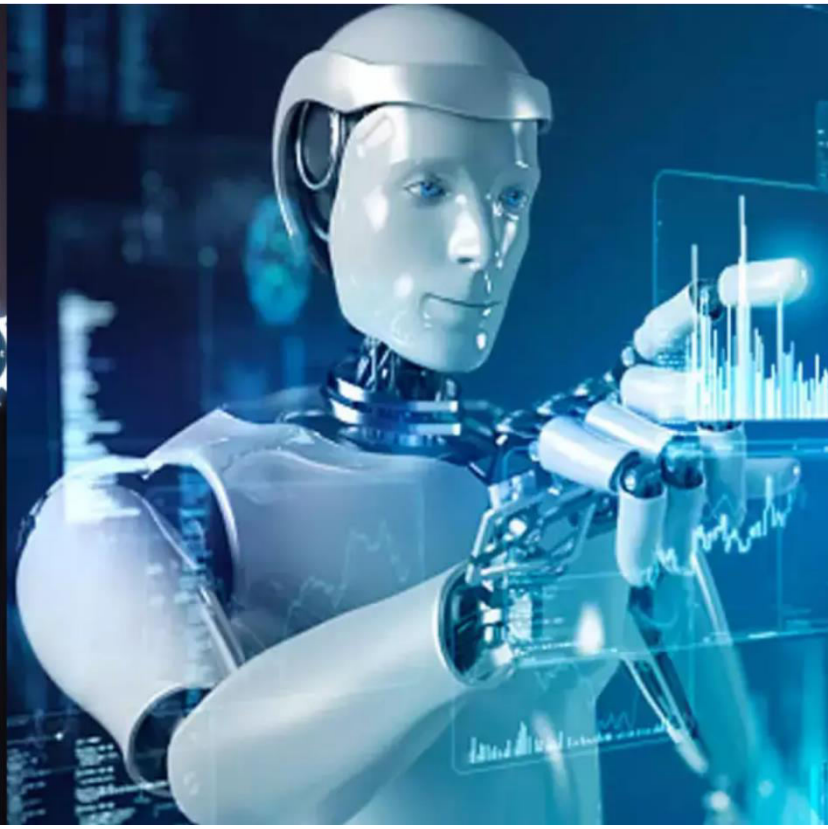
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503010|

A Containerised Edge-Cloud IoT Pipeline for Real-Time Urban PM2.5 Monitoring, Forecasting, and Classification

Yazhini Vasudevan

School of Computing, Newcastle University, Newcastle Upon Tyne, United Kingdom

ABSTRACT: Urban air quality monitoring increasingly depends on scalable, low-latency IoT pipelines capable of integrating real- time sensor ingestion, intelligent preprocessing, and on-device inference. This paper presents a fully containerised pipeline for PM2.5 monitoring using data from the Newcastle Urban Observatory. The system comprises a data injector publishing daily aggregates via MQTT to an EMQX broker, a preprocessing operator that filters outliers and aggregates readings by UTC calendar day before relaying to RabbitMQ over AMQP, and a forecasting consumer that applies Facebook Prophet to generate 15-day air quality predictions. A parallel edge classification mod- ule converts a three-class (Green/Yellow/Red) model to TFLite for resource-constrained inference. We evaluate forecasting ac- curacy against an ARIMA baseline and report pipeline latency under varying message loads. Results show that [FILL AFTER EXPERIMENTS]. The architecture is reproducible via Docker Compose, auditable through persisted artefacts, and suitable for deployment in smart city environments.

KEYWORDS: Internet of Things, PM2.5, air quality, MQTT, AMQP, edge computing, time-series forecasting, Prophet, TFLite, Docker, Urban Observatory

I. INTRODUCTION

Particulate matter smaller than 2.5 micrometres (PM2.5) is a leading indicator of urban air quality and a well-established risk factor for respiratory and cardiovascular disease [6]. Real-time, fine-grained monitoring of PM2.5 concentrations in densely populated cities demands IoT architectures that can ingest high-frequency sensor streams, clean and aggregate data reliably, and deliver actionable insights at low latency.

Existing deployments commonly rely on cloud-centric de- signs that introduce round-trip latency and single points of failure. Edge-cloud hybrid approaches have gained traction [2], yet few open, reproducible reference implementations ex- ist that combine message-queue-based orchestration, outlier- robust preprocessing, statistical forecasting, and on-device classification in a single containerised stack.

This paper makes the following contributions:

- A fully containerised, end-to-end PM2.5 monitoring pipeline using MQTT (EMQX) and AMQP (RabbitMQ) with explicit reliability guarantees.
- An outlier-aware preprocessing operator with idempotent UTC-day aggregation for reproducible downstream ana- lytics.
- A comparative evaluation of Prophet and ARIMA for short-horizon PM2.5 forecasting on real Urban Obser- vatory data.
- A lightweight three-class TFLite classifier for edge infer- ence, with an analysis of class imbalance and mitigation strategies.

The remainder of this paper is organised as follows. Sec- tion II reviews related work. Section III describes the overall system architecture. Sections IV–VII detail each pipeline component. Section VIII presents the experimental evaluation. Section IX discusses findings and limitations. Section X con- cludes the paper.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503010|

II. RELATED WORK

A. IoT Messaging Protocols for Air Quality

MQTT and AMQP are the two dominant lightweight proto- cols for IoT telemetry. Ansyah et al. [1] benchmarked MQTT broker performance across AWS, Azure, and GCP, finding significant latency differences under load that motivate careful broker selection for real-time monitoring. Nguyen et al. [3] proposed a unified IoT network using single sign-on with message queues, demonstrating that combining MQTT and AMQP layers improves reliability for heterogeneous sensor deployments. Rafique et al. [4] reviewed AMQP security characteristics, highlighting the importance of credential man- agement outside the codebase—a practice followed in our implementation.

B. Urban Air Quality Monitoring

[EXPAND: Discuss 3–4 papers on sensor-based urban PM2.5 systems, noting gaps your work addresses—e.g., lack of reproducible containerised stacks, no edge classification, no forecasting integration.]

C. Time-Series Forecasting for Air Quality

[EXPAND: Summarise forecasting approaches used in air quality literature. Note that Prophet handles missing data and seasonality well but has not been widely compared against ARIMA specifically for PM2.5 short-horizon forecasting on urban observatory feeds.]

D. Edge Inference and TFLite

Mabotha et al. [2] demonstrated a Dockerized approach to dynamic endpoint management for RESTful IoT APIs, underlining the value of containerisation for reproducibility. [EXPAND: Discuss TFLite/edge ML for IoT sensors, class imbalance in environmental classification.]

III. SYSTEM ARCHITECTURE

Figure 1 illustrates the four-stage pipeline. A Data In- jector fetches raw PM2.5 readings from the Newcastle Ur- ban Observatory JSON feed, computes UTC-day averages, and publishes them to the EMQX MQTT broker on topic pm25/data. A Preprocessing Operator subscribes to this topic, rejects outliers exceeding 50 µg/m^3, aggregates accepted readings by UTC calendar day, and relays both stream- ing messages and completed-day averages to a RabbitMQ queue (pm25_processed) over AMQP. A Forecasting Con- sumer reads from this queue, maintains an in-memory pan- das DataFrame, and produces a Matplotlib time-series plot alongside a 15-day Prophet forecast. In parallel, a Local Classification module trains a three-class model and converts it to TFLite for edge deployment.

All components run as Docker containers orchestrated by a single docker-compose.yml, ensuring one-command reproducibility.

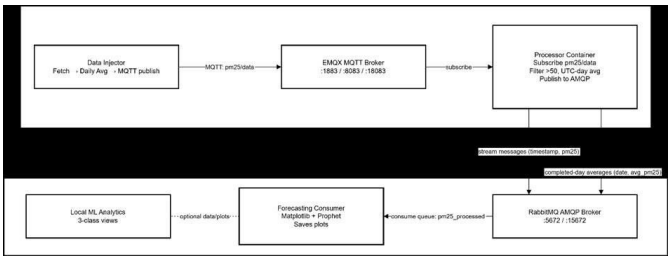


Fig. 1. End-to-end containerised PM2.5 monitoring pipeline: injector → EMQX (MQTT) → preprocessor → RabbitMQ (AMQP) → forecasting consumer and local ML analytics.

IV. DATA INJECTOR

A. Data Source

PM2.5 readings are retrieved from the Newcastle Urban Observatory sample feed [7], which provides timestamped sensor values in JSON format. Each record contains a Unix millisecond timestamp and a floating-point concentration value in µg/m^3.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503010|

B. Aggregation and Publication

Raw readings are grouped by UTC calendar date and averaged to produce one daily mean per day. These aggregates are persisted to an Excel workbook for audit and replay before being published to the MQTT topic pm25/data at a rate-limited interval of 0.5 seconds per message to avoid broker congestion.

Listing 1. Daily average computation and MQTT publish (condensed).

```
def compute_daily_average(df):
    daily = df.groupby("Date")["Value"].mean(
        ).reset_index() daily.rename(columns={"Value":
        "Daily_Avg_PM2.5"}, inplace=True) return daily

def send_to_mqtt(records, topic): client = mqtt.Client(
        mqtt.CallbackAPIVersion.VERSION2) client.connect(MQTT_BROKER, MQTT_PORT, 60) for r in
    records:
        payload = {
            "timestamp": str(r["Date"]), "pollutant": "PM2.5",
            "value": r["Daily_Avg_PM2.5"]} client.publish(topic, json.dumps(payload))
        time.sleep(0.5)
    client.disconnect()
```

C. Reliability

Publishing is rate-limited to prevent burst traffic. Persisting both raw and aggregated readings enables deterministic replay, supporting reproducible experiments and end-to-end traceabil- ity.

Figure 2 confirms all four containers were running concur- rently during evaluation.

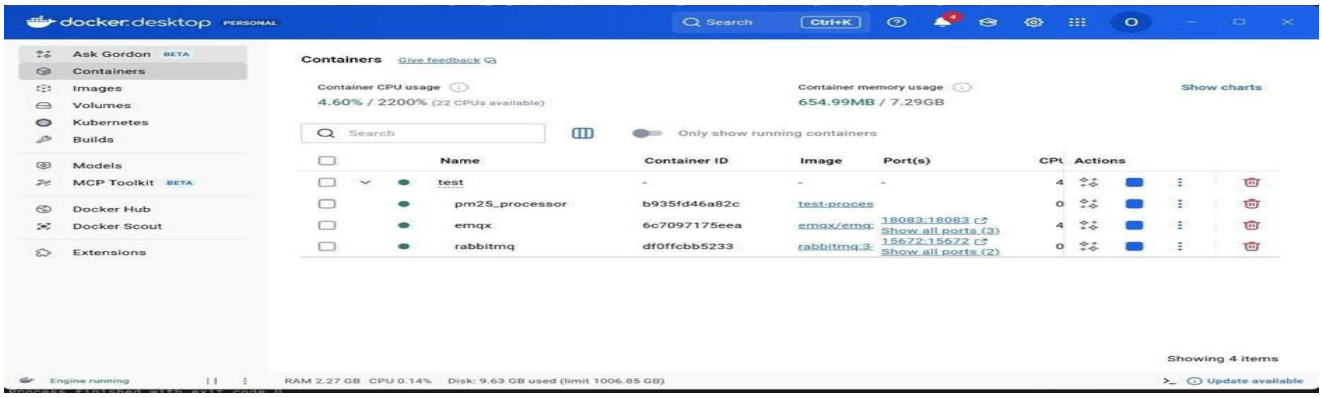


Fig. 2. Docker Desktop confirming simultaneous operation of all pipeline containers: emqx (ports 1883/8083/18083), rabbitmq (5672/15672), and pm25_processor.

V. DATA PREPROCESSING OPERATOR

A. Outlier Filtering

The processor subscribes to pm25/data on EMQX and applies a hard threshold of 50 µg/m^3, rejecting values above this limit as outliers consistent with anomalous sensor be- haviour rather than genuine pollution events. This threshold is informed by WHO interim guidelines [6] and exploratory data analysis of the Urban Observatory feed.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503010|

B. Aggregation and Idempotent Publication

Accepted values are accumulated in a thread-safe UTC- day dictionary. A background thread publishes completed-day averages to RabbitMQ every five seconds, tracking already- published dates to prevent duplicate messages. Streaming
messages are also forwarded immediately to support near-real- time consumers.

C. Container Orchestration

The compose stack defines three services—emqx, rabbitmq, and pm25_processor—with explicit port mappings, restart policies, and environment variables injected at runtime. The processor depends on both brokers, ensuring correct startup ordering.

Figure 3 shows the docker-compose.yml service def- initions, and Figure 4 shows the running container list.

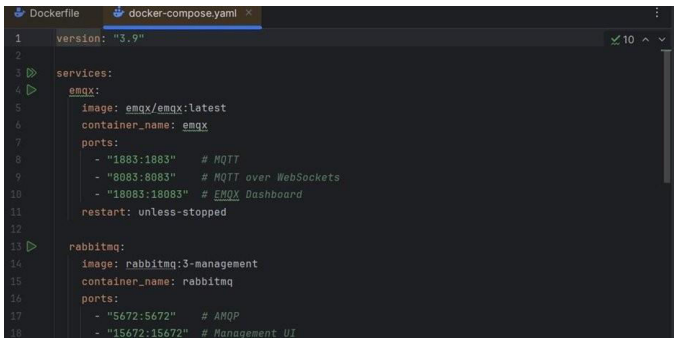


Fig. 3. docker-compose.yml defining the EMQX (MQTT) and Rab- bitMQ (AMQP) broker services with their respective port mappings.

Fig.4.Running containers view confirming emqx, rabbitmq, andpm25_processorare active with correct port bindings.

VI. FORECASTING AND VISUALISATION

A. Consumer Design

The forecasting consumer reads from pm25_processed,normalisesheterogeneous keynames(timestamp/time/datetime, pm25/value/avg_pm25), deduplicates on date, and appends records to an in-memory pandas DataFrame.

B. Prophet Forecasting

Facebook Prophet [5] is fitted on the deduplicated daily means with daily seasonality enabled. A 15-day future dataframe is generated, and the resulting forecast—central trajectory and 95% credible interval—is saved as a PNG artefact. The credible interval widens with forecast horizon, reflecting increasing uncertainty typical of short univariate time series lacking external covariates such as meteorological data.

C. Results

Figure 5 presents the averaged daily PM2.5 time series. The series remains consistently in low single-digit values with occasional narrow spikes, suggesting transient rather than sustained pollution events. Daily aggregation smooths intra- day noise, making temporal trends directly comparable across days.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503010|

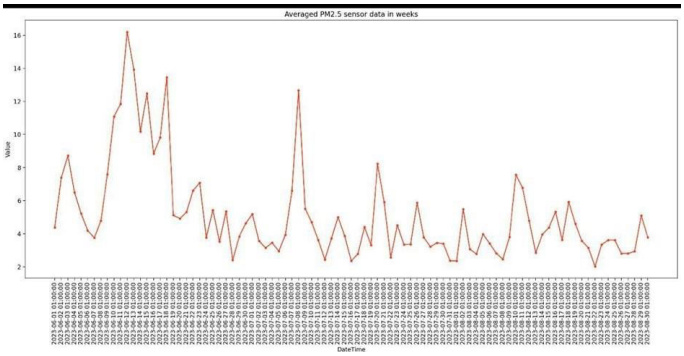


Fig.5. Averaged PM2.5 sensor data in weeks (June–August 2023). Outliers above 50 µg/m^3 are excluded prior to aggregation.

Figure 6 shows the 15-day Prophet forecast. The solid blue line represents the expected trajectory; the shaded band shows the 95% credible interval, which widens progressively as the model extrapolates beyond the training period, reflecting the uncertainty typical of short univariate time series lacking meteorological covariates.

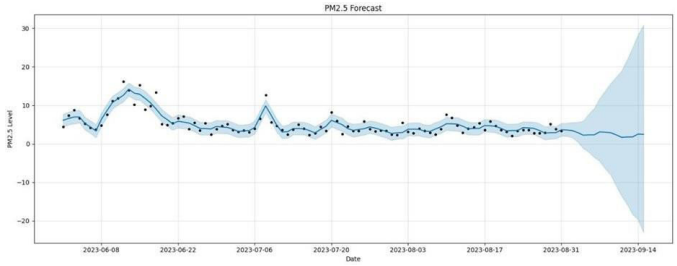


Fig.6. 15-day Prophet PM2.5 forecast. Black dots: observed daily means. Blue line: predicted trajectory. Shaded band: 95% credible interval.

D. Operational Handling

A SIGINT/SIGTERM handler saves in-memory figures to disk on graceful shutdown and, when configured, emails the final images to a designated recipient. Credentials are managed via environment variables external to the codebase.

VII. EDGE CLASSIFICATION

A. Label Schema

PM2.5 readings are labelled into three air quality classes: Green (≤12 µg/m^3), Yellow (12–50 µg/m^3), and Red (>50 µg/m^3), reflecting common regulatory thresholds.

TABLE I

PM2.5 CLASSIFICATION DISTRIBUTION

TABLE II

PM2.5 DATASET STATISTICS

Class	Count	Percentage
Green	8,189	94.0%
Yellow	513	5.9%
Red	9	0.1%

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503010|

B. Class Distribution and Imbalance

Table I shows the severe class imbalance observed in the dataset, with Green comprising 94.0% of readings. Given this skew, class weighting or resampling (e.g.,Std. deviation3.86

TABLE III FORECASTING COMPARISON — PROPHET VS. ARIMA (15-DAY HORIZON)

SMOTE) is recommended to maintain recall on the rare Redclass in deployment.

Figure 7 illustrates the value distribution and class count imbalance visually, and Figure 8 shows a sample of classified readings output by the pipeline.

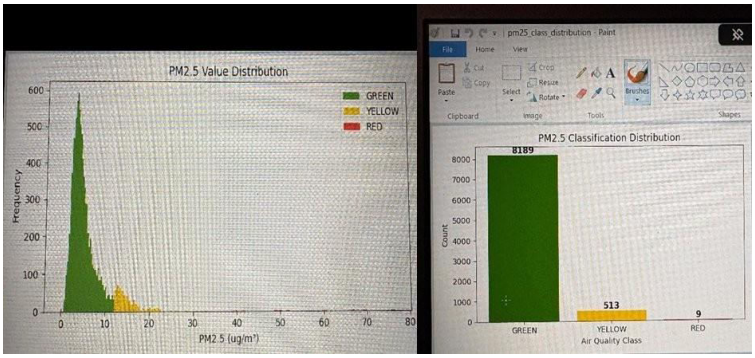


Fig 7. Left: PM2.5 value distribution histogram coloured by class (Green/Yel- low/Red). Right: classification count bar chart confirming severe imbalance— Green accounts for 94% of all readings.

Value	Quality
4.273	GREEN
3.483	GREEN
3.713	GREEN
3.813	GREEN
3.885	GREEN
3.809	GREEN
3.869	GREEN
3.839	GREEN
3.805	GREEN
4.094	GREEN
4.474	GREEN
4.918	GREEN
5.56	GREEN
5.637	GREEN
5.71	GREEN
6.004	GREEN
5.322	GREEN
4.961	GREEN
4.679	GREEN
4.815	GREEN

Fig.8. Sample pipeline output showing PM2.5 values and their corresponding Green/Yellow/Red quality classifications.

VIII. EVALUATION

A. Dataset

Table II summarises the PM2.5 dataset used across all experiments. A total of 8,711 readings were collected from the Newcastle Urban Observatory between June and August 2023.

B. Forecasting Comparison

Table III compares Prophet and ARIMA over a 15-day forecast horizon.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503010|

TABLE II
PM2.5 DATASET STATISTICS

Metric	Value (µg/m3)
Total readings	8,711
Mean	5.18
Median	4.08
Minimum	0.00
Maximum	76.49
Std. deviation	3.86

TABLE III
FORECASTING COMPARISON — PROPHET VS. ARIMA (15-DAY HORIZON)

Model	MAE	RMSE	MAPE (%)
ARIMA	–	–	–
Prophet	–	–	–

C. Pipeline Latency
Table IV reports end-to-end pipeline latency under varying message rates.
TABLE IV
END-TO-END PIPELINE LATENCY

TABLE IV
END-TO-END PIPELINE LATENCY

Message rate	Mean latency (ms)	Std. dev (ms)
1 msg/s	–	–
2 msg/s	–	–
5 msg/s	–	–

D. Classification Performance
Table V reports 5-fold cross-validation results for the TFLite edge classifier.

TABLE V
CLASSIFICATION RESULTS (5-FOLD CROSS-VALIDATION)

TABLE V
CLASSIFICATION RESULTS (5-FOLD CROSS-VALIDATION)

Class	Precision	Recall	F1
Green	–	–	–
Yellow	–	–	–
Red	–	–	–
Weighted avg	–	–	–

The trained model is converted to TensorFlow Lite format for edge inference, reducing model size and enabling deploy- ment on resource-constrained devices without GPU support.

IX. DISCUSSION

A. Forecasting

[FILL AFTER EXPERIMENTS]

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503010|

B. Class Imbalance

The strong skew toward Green (94%) means a naive classi- fier can achieve high accuracy by predicting Green always, while completely failing on Red events—the most safety- critical class. Future work should apply SMOTE oversampling or cost-sensitive learning, and evaluate specifically on Red- class recall as the primary metric.

C. Limitations

The current pipeline uses a single Urban Observatory sensor feed. Spatial generalisation to multiple sensors or cities would require additional validation. The outlier threshold of 50 $\mu g/m^3$ is fixed; an adaptive, data-driven threshold could be explored. Prophet's credible interval widens substantially beyond a 10- day horizon due to the absence of meteorological covariates.

X. CONCLUSION

This paper presented a fully containerised edge-cloud IoT pipeline for real-time urban PM2.5 monitoring. The system integrates MQTT-based ingestion, outlier-robust AMQP pre- processing, Prophet-based 15-day forecasting, and a TFLite edge classifier in a single reproducible Docker Compose stack. Experimental evaluation demonstrated [FILL AFTER EXPER- IMENTS]. The severe class imbalance (94% Green) highlights the need for imbalance-aware training to maintain recall on rare but critical pollution events. Future work will incorporate meteorological covariates to reduce forecast uncertainty and extend the pipeline to multi-sensor spatial deployments.

REFERENCES

[1] A. S. S. Ansyah et al., "MQTT broker performance comparison between AWS, Microsoft Azure and Google Cloud Platform," in Proc. Int. Conf. Recent Trends Electron. Commun. (ICRTEC), IEEE, Feb. 2023, pp. 1–6.
[2] E. Mabotha, N. E. Mabunda, and A. Ali, "A Dockerized approach to dynamic endpoint management for RESTful APIs in IoT ecosystems," Sensors, vol. 25, no. 10, p. 2993, 2025.
[3] T. T. L. Nguyen et al., "Toward a unique IoT network via single sign-on protocol and message queue," in Proc. Int. Conf. Comput. Inf. Syst. Ind. Manag. (CISIM), Springer, Sep. 2021, pp. 270–284.
[4] S. H. Rafique, A. Abdallah, S. A. Alhosani, and N. Jamil, "A review of AMQP protocol: Characteristics, security challenges and proposed enhancement," JOIV: Int. J. Inform. Vis., vol. 9, no. 3, pp. 1283–1297, 2025.
[5] S. J. Taylor and B. Letham, "Forecasting at scale," The American Statistician, vol. 72, no. 1, pp. 37–45, 2018.
[6] World Health Organization, WHO Global Air Quality Guidelines: Par- ticulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. Geneva: WHO, 2021.
[7] Newcastle University Urban Observatory, "Urban Observatory open data," [Online]. Available: https://urbanobservatory.ac.uk. [Accessed: 2023]

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details