Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 10, No. 2, June 2024, pp. 217-234
ISSN: 2338-3070, DOI: 10.26555/jiteki.v10i2.28490

Performance of an AIOT-Particle Device for Air Quality and Environmental Data Prediction in Salatiga Area Using ARIMA Model

Johanes Dian Kurniawan, Suryasatriya Trihandaru, Hanna Arini Parhusip
Master of Data Science, Faculty of Science and Mathematics, Satya Wacana Christian University
Jl. Diponegoro No 52-60, Salatiga, Indonesia 50711

ARTICLE INFO

Article history:

Received February 29, 2024
Revised March 26, 2024
Published June 21, 2024

Keyword:

ARIMA;
Air Quality;
IoT;
Prediction;
Forecasting

ABSTRACT

This study introduces the AIOT-Particle, a compact device designed for comprehensive air quality and environmental monitoring in Tegalrejo, Salatiga, Indonesia. Addressing the need for real-time, multi-parameter environmental data, the device simultaneously tracks PM1.0, PM2.5, temperature, humidity, pressure, and altitude, utilizing a built-in data fusion algorithm to ensure accurate and coherent data collection. Air pollution standards classify air quality as "good" (0–50), "moderate" (51–100), "unhealthy" (101-200), "very unhealthy" (201-300), and "hazardous" (>300). The research contribution is the development and validation of the AIOT-Particle using the ARIMA model for precise environmental monitoring. The methods involved deploying the device in Salatiga and applying the ARIMA model to analyze the collected data for accuracy. The results demonstrated promising accuracy: for PM1.0, the RMSE was 8.13 with an MAE of 6.04; for PM2.5, the RMSE was 6.60 with an MAE of 4.49. Environmental data analysis showed an RMSE of 0.74 for temperature (MAE 0.43), 2.11 for humidity (MAE 1.36), 0.25 for pressure (MAE 0.19), and 2.18 for altitude (MAE 1.70). These findings highlight the device's potential to enhance environmental surveillance and public health assessments, advance the understanding of air quality dynamics, and support targeted interventions to mitigate environmental risks. The novelty of this study lies in the integration of multiple environmental parameters into a single monitoring device, validated for accuracy using the ARIMA model.

Corresponding Author:

Hanna Arini Parhusip, Master of Data Science, Faculty of Science and Mathematics, Satya Wacana Christian University, Jl. Diponegoro No 52 – 60, Salatiga 50711
Email: hanna.parhusip@uksw.edu

1. INTRODUCTION

As urbanization accelerates, the impact on air quality becomes a paramount concern for public health and environmental sustainability. One significant concern is particulate matter (PM), a mixture of small solid and liquid particles found in the air that contributes to air pollution [1], [2]. These particles originate from various sources, including fossil fuel combustion, dust, and industrial emissions [3], [4]. Due to their small size, particulate matter can penetrate the human respiratory system, causing adverse health effects. PM1.0, consisting of particles smaller than 10 micrometers, and PM2.5, consisting of particles smaller than 2.5 micrometers, are particularly harmful as they can reach the lungs and bloodstream, enter the body, and affect organs in the body [5], [6] and potentially lead to severe health issues such as respiratory diseases, heart disease, stroke, and lung cancer [7], [8], [9]. Consequently, monitoring the concentration of PM1.0 and PM2.5 in the environment is crucial to mitigating human exposure and health risks [10], [11].

The prevalence of PM pollution in urban areas globally underscores its widespread impact. Cities such as Delhi, Beijing, and Mexico City frequently report PM levels far exceeding the World Health Organization's recommended limits, posing severe health risks to their inhabitants [12]. In 2019, a global air quality report revealed that 92% of the world's population lives in places where air quality levels exceed WHO limits, with urban areas being the most affected [13].

The study on the AIOT-Particle device in Tegalrejo, Salatiga, Central Java, Indonesia, addresses air quality challenges stemming from vehicular emissions, industrial activities, and domestic burning. Salatiga's tropical climate, high population density, and mix of urban and rural demographics complicate air quality management, impacting public health, especially among vulnerable groups. The AIOT-Particle device, validated with the ARIMA model showing promising accuracy, provides precise environmental data essential for developing targeted interventions to mitigate pollution and improve health outcomes, highlighting its potential to enhance environmental surveillance and inform tailored policy decisions. The AIOT-Particle device is specifically designed to monitor PM1.0, PM2.5, temperature, humidity, pressure, and altitude simultaneously through its compact design, enabling efficient functioning in various environmental conditions based on IoT, as can be seen in Fig. 3.

The previous studies related to PM2.5 and PM1.0 showed that the study to monitor PM2.5 in China resulted in an average RMSE of the ARIMA (4,1,0) of 12.68 and an average MAPE of 10.5%, which shows that the models have good generalization ability [14]. Another study in Italy using pandemic data, PM10 concentrations in Nocera Inferiore, resulted in the best model, a simple AR(1) with only an order 1 autoregressive component. This model is derived using the slope of the autocorrelation and partial autocorrelation functions [15]. Both previous studies concluded that ARIMA showed good performance in the simulation of data and was suitable as a short-term prediction model.

Located in the Tegalrejo area of Salatiga City, Central Java Province, Indonesia, the AIOT-Particle leverages an onboard data fusion algorithm to synchronize data acquisition from all sensors, ensuring accurate and cohesive environmental parameter readings. This comprehensive monitoring capability distinguishes the AIOT-Particle from existing devices in the field. The location of AIOT-Particle devices is depicted in Fig. 1.

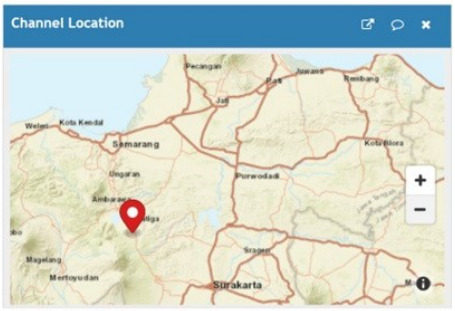


Fig. 1. Location of the AIOT-Particle device in Salatiga area

In Tegalrejo, Salatiga, Central Java, Indonesia, air quality challenges arise from vehicular emissions, industrial activities, and domestic burning, compounded by the region's tropical climate and high population density. Rapid urbanization and industrialization have historically increased pollution levels, particularly PM1.0 and PM2.5, from manufacturing and vehicular sources. Given the high PM emissions from vehicular emissions, industrial activities, and domestic burning, the study aims to mitigate PM pollution through real-time monitoring and accurate predictions, promote environmental sustainability by balancing industrial activities with preservation efforts, support industrial stakeholders in operating sustainably and complying with regulations, and ensure adherence to governmental environmental standards to prevent legal penalties and promote corporate responsibility.

The research contribution is twofold: First, the development and implementation of the AIOT-Particle device for simultaneous monitoring of multiple environmental parameters provides a comprehensive and accurate tool for real-time air quality assessment. Second, the validation of the device using the ARIMA model demonstrates its predictive capabilities, enhancing environmental surveillance and enabling targeted interventions to improve public health and sustainability. This innovative methodology, not previously achieved by other researchers, enhances understanding of air quality dynamics and supports the development of targeted interventions. The subsequent sections detail the methodology employed and analyze the results, advancing environmental monitoring and its implications for public health and sustainability.

2. METHODS

In the study, the process of creating an ARIMA model was carried out in several stages involving data collection, loading, and reading of the sensor. data preprocessing, data normalization, creating a supervised dataset, stationery test, differencing, splitting data train and test, building an ARIMA model, prediction, prediction performance evaluation, and forecasting as depicted in the flowchart in Fig. 2.

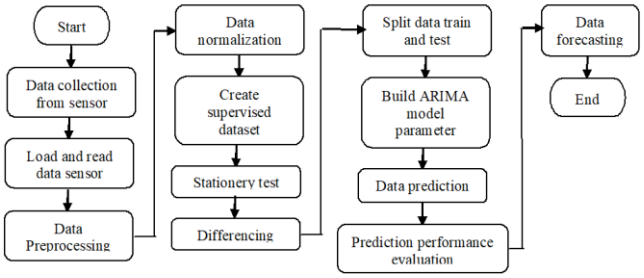


Fig. 2. Research Flowchart

2.1. Flowchart description

From Fig. 2, the brief description of each step in the research methodology is explained as listed:

2.1.1. Data collection: The first step to collecting all the research data derived from the sensors DSM501 and BME280, which are located in the Tegalrejo area of Salatiga. Calibration was made up using reference-grade equipment under controlled conditions to establish baseline measurements. Similarly, the BME280 sensors, responsible for measuring temperature, humidity, pressure, and altitude, were calibrated against standardized instruments to ensure their precision. The data was taken from November 2023 to February 2024, for a total of 310.080 pieces of data, with data intervals read every 15 seconds using the Thingspeak platform.

2.1.2. The load and read data sensor: Is the research step to load the data, which is stored in Google Drive, and then read the data to know the data content and characteristics. Data preprocessing: Aims to prepare and clean the raw data before it is fed into a machine learning algorithm or statistical model. In this research, data preprocessing includes dropping unnecessary data columns, indexing data datetime, resampling data into an hour, eliminating outliers using the interquartile range method, and dropping zero and NaN values. Data preprocessing is crucial for enhancing the quality and accuracy of machine learning models. These techniques streamline the dataset, ensure temporal accuracy, reduce noise, and clean the data from distortions and errors, leading to more reliable and robust models. The rationale behind these steps is to focus on relevant features, maintain consistent time intervals, and remove data points that could skew results. However, potential limitations include the risk of inadvertently removing valuable information, as these steps can also lead to the loss of significant data insights. A careful evaluation is necessary to balance data cleanliness with the preservation of important information.

2.1.3. Normalize data: The main purpose of data normalization is to set and ensure that the numerical values of different features are on a similar scale, which will contribute to stability.

2.1.4. Create supervised dataset: To predict the next data values based on the past value sequence, it will need to create a supervised dataset for a supervised learning scheme.

2.1.5. Stationery test: The stationery test is the step to check stationarity from each feature to satisfy ARIMA model assumptions to get robustness and accuracy in capturing the underlying patterns in time series, which will lead to more reliable predictions.

2.1.6. Differencing: Differencing is a process that aims to make a time series stationary. Stationarity is a key assumption in many timeseries models, including ARIMA. By differentiating a time series, it will remove trends and seasonality, common sources of non-stationarity.

2.1.7. Split data train and test: Split data train and data test is a common way to split the data into data train and data test before predicting the dataset. In this step, the total data after resampling is 2.818 and will be split into an 80% data test (2.254 data) and a 20% data test (564 data). The split data will be taken from the dataset, which is already resampled into hourly date times to reduce computation load.

2.1.8. Building ARIMA model parameter: Building the ARIMA model parameter is to search and find the p, d, and q parameter values to build the ARIMA model, which will be used for prediction. The ARIMA model (*p, d*, and *q*) will be determined by plotting the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation) graphs for each feature and then analyzing and observing the ACF and PACF graphs. This begins with checking the stationarity of the time series data using the Augmented Dickey-

Fuller (ADF) test, applying differencing as needed to achieve stationarity for unstationer features. Next, the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are analyzed to determine initial guesses for p and q values. Model validation involves comparing predictions against a holdout sample using RMSE and mean absolute error (MAE). The selected ARIMA models, such as ARIMA (3,0,1) for temperature and ARIMA (1,0,1) for PM2.5, were chosen for their satisfactory RMSE, ensuring reliable and accurate predictions.

2.1.9. Data prediction: Data prediction is the crucial step and the main step to predicting the next values of the dataset. Prediction will use the data train to predict the data test that is already set in split data steps (point 2.1.8).

2.1.10. Evaluation prediction performance: After the prediction, an evaluation of prediction performance will be made to evaluate ARIMA models. In this research, the performance evaluation is performed using RMSE and MAE.

2.1.11. Data forecasting: Data forecasting is predicting future trends or outcomes based on historical data. Forecasting models provide insights into what might happen in the future, aiding in planning and strategizing for optimal outcomes.

2.2. ARIMA

The Autoregressive Integrated Moving Average (ARIMA) method is a powerful time series forecasting technique that combines autoregression, differencing, and moving average components to capture and model the underlying patterns and trends within a time series [16], [17]. The ARIMA method is a powerful time series forecasting technique chosen for its flexibility and effectiveness in handling various time series data. It combines three key components: autoregression (AR), differencing (I), and moving average (MA). The AR part uses past values to predict future ones, the I part makes the data stationary by differencing, and the MA part smooths out noise by modeling the dependency between an observation and a residual error. Parameterization involves selecting appropriate values for these components by analyzing autocorrelation and partial autocorrelation plots and performing unit root tests. ARIMA's flexibility, strong theoretical foundation, proven empirical success, and robust diagnostic tools make it a versatile and reliable choice for forecasting, effectively capturing underlying patterns and trends within the data.

The autoregressive (AR) component, emphasizing the relationship between an observation and its lagged values, has been foundational in time series analysis [18]–[20]. The integrated (I) component, involving differencing to achieve stationarity, plays a critical role in removing trends and seasonality, ensuring the model's stability and effectiveness [21]. The autoregressive (AR) equation is mathematically described as follows [22];

$$y_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \epsilon_t \quad (1)$$

with y_t denotes dependent variable at time t, μ denotes intercept, $\varphi_1, \varphi_2, .., \varphi_p$ denotes autoregressive parameter, $X_{t-1}, X_{t-2}, ,, ,, X_{t-p}$ denotes lagged values of the time series, and ε_t denotes error value at time t.

The moving average (MA) component models the relationship between observations and residual errors from a moving average model applied to lagged observations, enhancing the model's capacity to capture short-term fluctuations [23], [24]. The moving average (MA) equation mathematically can be expressed as [25];

$$y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-k} \quad (2)$$

with y_t denotes dependent variable at time t, μ is intercept, $\theta_1, \theta_2, \ldots, \theta_p$ denotes moving average parameter, e_t $e_{t-1}, e_{t-2}, \ldots, e_{t-k}$ denotes error lagged error value at the time of *t-k*.

The ARIMA model is denoted as ARIMA (*p, d, q*), where '*p*' represents the order of the AR component, '*d*' is the order of differencing, and '*q*' is the order of the MA component [26], [27]. Proper parameter selection is crucial for accurate modeling. Validation and assessment of stationarity are essential prerequisites before applying ARIMA, with various tests available to ensure the suitability of the time series for modeling. such as the ADF (Augmented Dickey-Fuller Test) [28]–[30]. ARIMA can be expressed in mathematical equation form as follows [24]:

$$y_t = C + \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_p \epsilon_{t-q} + \epsilon_t \quad (3)$$

where y_t denotes the dependent variable at time *t*, *C* denotes the intercept of a model, $\varphi_1, \ldots, \varphi_p$ denotes autoregressive parameters, $\theta_1, \ldots, \theta_p$ denotes moving average parameters, y_{t-1}, $\ldots, y_{t-p}$, and $\epsilon_{t-1}, \ldots, \epsilon_{t-q}$ denotes lag error term at time *t* and current error term respectively.

2.3. IoT

IoT is a technological concept that connects various physical devices or objects to the internet or other communication networks [31], [32]. In this study, data was obtained from readings of BME280 and DSM501 sensors from the AIOT-Particle device, which were then transmitted using the IoT concept through MQTT brokers. IoT connects various physical devices or objects to the internet or other communication networks, allowing them to collect, transmit, and exchange data with other devices and data centers [33], [34]. The architecture and method of IoT in this study are depicted in Fig. 3. The sensor data collected by sensors is processed either locally on the device itself or sent to a data center (MQTT and Thingspeak) for further analysis. This may involve real-time data processing, big data analysis, or artificial intelligence [35], [36].

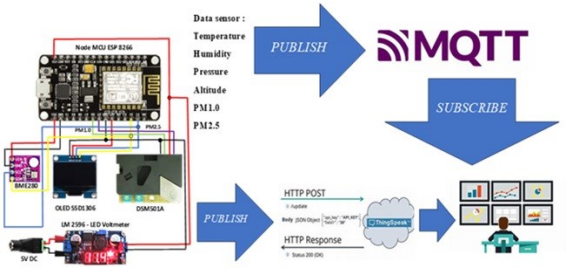


Fig. 3. IoT research architecture with specific connected devices produced by the authors in 2023

The AIOT-Particle infrastructure, as depicted in Fig. 3, consists of the following elements:

2.3.1. Sensors.

Two sensors were utilized:

a. DSM 501: The DSM501 sensor detects dust particles with a size concentration of more than 1 micron in the form of solid matter or particulate molecules (PM) to monitor air quality. DSM501 is also used to determine the concentration of particulate matter (PM). The DSM501 sensor detects and measures fine particle concentrations in the air using the laser scattering method.

b. BME280: The BME280 sensor module measures barometric pressure, altitude, temperature, and humidity. Because pressure varies with height, the BME280 sensor can estimate altitude. This sensor module comes in various different versions. Data is exchanged between the BME280 sensor module and microcontrollers using the I2C or SPI communication protocol.

2. Microcontroller ESP8266: Expressive Systems produces the ESP8266, a low-cost Wi-Fi microprocessor. The ESP8266 is a low-cost solution for adding Wi-Fi capability to devices such as IoT (Internet of Things) applications, home automation systems, and sensor networks. It has a full TCP/IP protocol stack, which allows it to communicate with both TCP and UDP. In this experiment, the ESP8266 functions as a standalone microcontroller, with programming executed directly on the chip.

3. OLED SSD1306: The OLED SSD1306 is a small, energy-efficient display module with high contrast and sharp visuals. It's commonly used in electronics projects due to its compact size, low power usage, and compatibility with various microcontrollers. The display controller, SSD1306, supports different communication protocols, making it easy to integrate into different devices.

4. LM2596: The LM2596 is a versatile voltage regulator integrated circuit (IC) commonly used in electronic circuits to efficiently step down voltage levels. It can take a higher input voltage, typically up to 40 volts, and regulate it down to a lower, adjustable output voltage, usually ranging from 1.2 volts to 37 volts.

5. MQTT: MQTT (messaging Queuing Telemetry Transport) is a standardized messaging protocol, or set of rules, used for machine-to-machine communication. Wearables, smart sensors, and Internet of Things (IoT) devices commonly require data transfer via networks with limited bandwidth and resources. These IoT devices use MQTT for data transmission because it is simple to set up and can efficiently communicate IoT data. MQTT enables message transmission between devices and the cloud, as well as from the cloud to devices. In this investigation, the data is sent through a mosquito broker. This mosquito is a free platform that is simple to install, making it the ideal solution for this research. With the presence of IoT in this research, data can be sent through the MQTT broker using the test.mosquitto.org broker and visualized on Thingspeak, providing precise data analysis. It can be utilized for automated decision-making or as a decision-support tool [37], [38].

2.4. Thingspeak

ThingSpeak is an Internet of Things (IoT) platform developed by MathWorks. It allows users to collect, analyze, and visualize data from sensors and other IoT devices in real time [39], [40]. The platform provides a cloud-based infrastructure that enables the creation of IoT applications without requiring extensive programming or hardware expertise.

2.5. Mean Absolute Error (MAE)

It is a method used to calculate the mean value of the absolute difference between the predicted and actual values. In general, if the MAE value is smaller, the quality of a model will be better [41], [42]. MAE is chosen because it is straightforward to interpret and does not consider the direction of the errors, focusing solely on their magnitude. MAE represents the average error magnitude in prediction models without considering the trend of errors. The general equation for MAE can be expressed as [43]:

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|Y_{pred} - Y_{obs}| \quad (4)$$

where N denotes the number of data points, Y_{pred} denotes the prediction value, and Y_{obs} denotes the actual value.

2.6. Root Mean Square Error (RMSE)

It is a method used to measure the performance of a model. RMSE calculates the square root of the average of the squared differences between actual values and values predicted by the model. RMSE is selected as it provides a comprehensive measure of model performance by penalizing larger errors more heavily. If the RMSE value is smaller, the quality of the model is better [44], [45]. In general, RMSE can be expressed by the equation [46];

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{n}(Y_{obs} - Y_{pred})^2} \quad (5)$$

where N denotes the number of data points, Y_{obs} denotes the actual value, and Y_{pred} denotes the prediction value.

2.7. Standard Deviation

The standard deviation (σ) is a measure of the amount of variation or dispersion in a set of values. It indicates how spread out the data points are around the mean. In the context of model evaluation, standard deviation is used to normalize the RMSE, allowing for comparison of error magnitudes relative to the inherent variability in the data. The standard deviation is calculated as [47];

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{obs,i} - \dot{Y_{obs}})^2} \quad (6)$$

where N denotes the number of data points, $Y_{obs,i}$ denotes the actual value, and Y_{obs} denotes the mean of the observed values.

2.8. Air Pollution Standards

Air pollution standards are regulatory measures established by governments or international organizations to control the concentration of pollutants in the air. These standards are designed to protect public health, the environment, and air quality. The air pollution standards in this research are shown in Table 1.

Table 1. Research air quality standards (references from: https://ditppu.menlhk.go.id/portal/read/indeks-standar-pencemar-udara-ispu-sebagai-informasi-mutu-udara-ambien-di-indonesia)

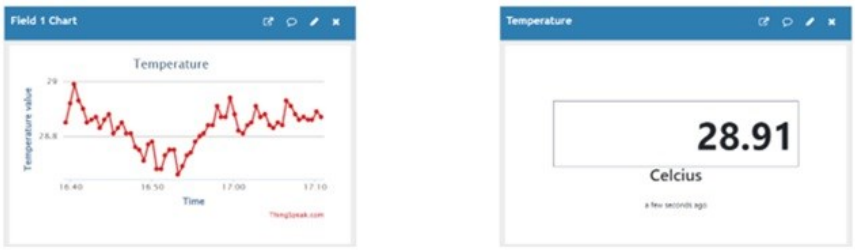
Category	PM 2.5 (ug/m3)	PM 1.0 (ug/m3)
Good (0-50)	0-15.5	0-50
Moderate (51-100)	15,6-55,4	51-150
Unhealthy (101-200)	55,5-150,4	151-350

Very Unhealthy (201-300)	150,5-250,4	351-420
Dangerous (>300)	>250	>420

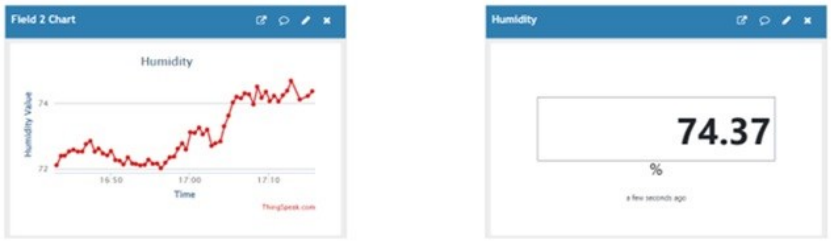
3. RESULTS AND DISCUSSION

3.1. IoT Implementation Result

The implementation results of the IoT architecture as shown in Fig. 3 for data visualization using MQTT and Thingspeak can be seen in Fig. 4 and Fig. 5. As Fig. 4 and Fig. 5 show, all features are displayed in real-time and visualized. The pattern that can be observed from Fig. 4a and Fig. 4b is that during the day, temperatures rise and relative humidity typically decreases. Conversely, at night, as temperatures drop, relative humidity tends to increase. and it may be concluded that temperature, humidity pressure, and altitude characteristics have an inverse correlation. In Figs. 4c and 4d, it is shown that when the pressure decreases with increasing altitude, it means that atmospheric pressure and altitude are inversely related, with pressure decreasing exponentially as altitude increases. This is because as altitude rises, the weight of the air above decreases, resulting in lower pressure. While in Figs. 4e and 4f, PM1.0 and PM2.5 are relatively positively correlated. This means that when the concentration of PM1.0 particles increases, the concentration of PM2.5 particles tends to increase as well, and vice versa. This is because PM1.0 particles are a subset of PM2.5 particles.



(a) Thingspeak visualization of temperature at February,29 2024 in the period time 16.40-17.10 WIB (left). The value of temperature in the last point (right)



(b) Thingspeak visualization of humidity at February,29 2024 in the period time 16.40-17.15 WIB (left). The value of humidity in the last point (right)

(c) Thingspeak visualization of pressure at February,29 2024 in the period time 17.10-17.50 WIB (left). The value of pressure in the last point (right)

(d) Thingspeak visualization of altitude at February,29 2024 in the period time 17.10-17.50 WIB (left). The value of altitude in the last point (right)

(e

(e) Thingspeak visualization of PM1.0 at February,29 2024 in the period time 17.10-17.50 WIB (left). The value of PM1.0 in the last point (right)

(f) Thingspeak visualization of PM2.5 at February,29 2024 in the period time 17.15-18.00 WIB (left). The value of PM2.5 in the last point (right)

Fig. 4. Thingspeak visualization of PM2.5 at February,29 2024

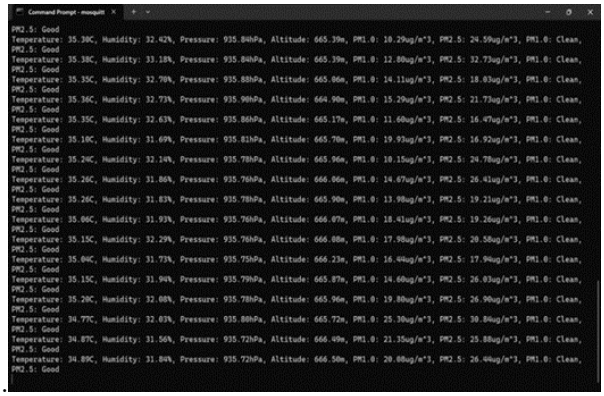


Fig. 5. MQTT display result

3.2. Data Preprocessing

Timeseries data plots for each feature are shown in Fig. 6–Fig. 11. From Fig. 6–Fig. 11, the plotted data for temperature, humidity, pressure, altitude, PM1.0, and PM2.5 reveal various trends, seasonal patterns, and irregularities. Temperature and humidity show diurnal and seasonal variations, with temperature fluctuations and humidity often higher in the morning. Pressure is relatively stable with minor weather-related fluctuations, while altitude should remain constant but may vary due to sensor drift. PM1.0 and PM2.5 indicate air quality, with higher concentrations potentially linked to pollution events such as traffic or industrial activities. Meanwhile, data preprocessing results for dropping unnecessary data columns, indexing data datetime, and resampling data into an hour are shown in Fig. 12–Fig. 13. From Fig. 12, it is clearly seen that the original data has NaN values with four unused columns, i.e., latitude, longitude, elevation, and status, then it drops, resulting in a in a new dataframe without NaN values.

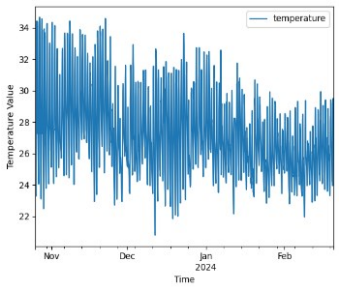


Fig. 6. Temperature time series data time series data observed in Nov2023-February 2024 and recorded daily

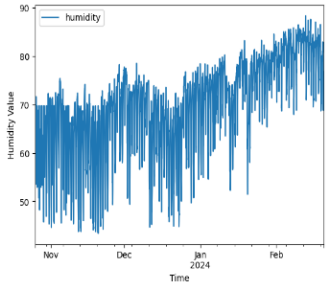


Fig. 7. Humidity time series data time series data observed in Nov2023-February 2024 and recorded daily

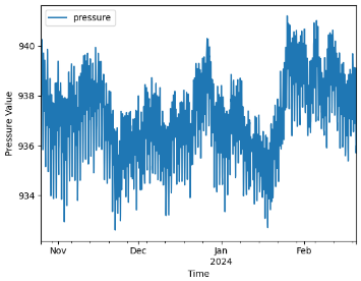


Fig. 8. Pressure time series data time series data observed in Nov2023-February 2024 and recorded daily

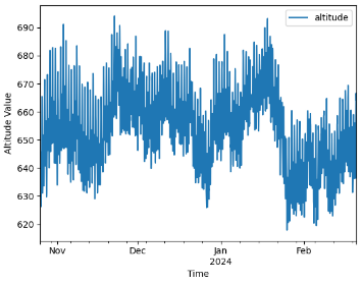


Fig. 9. Altitude time series data observed in Nov 2023-February 2024 and recorded daily

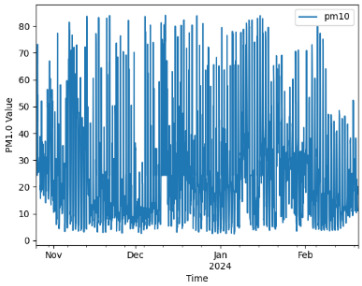


Fig. 10. PM1.0 time series data observed in Nov 2023-February 2024 and recorded daily

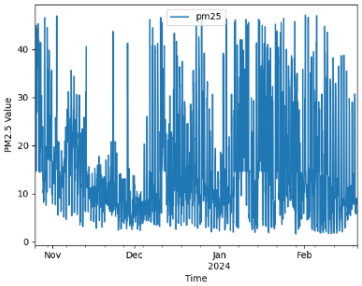


Fig. 11. PM2.5 time series data observed in Nov 2023-February 2024 and recorded daily

From Fig. 14 and Fig. 15, it is shown that the data was preprocessed to resample the original data after removing NaN values into hourly intervals to reduce the number of data points in order to shorten computation time. The original data consisted of 309153 rows × 6 columns, which were then reduced to 2818 rows × 6 columns.

created_at	entry_id	field1	field2	field3	field4	field5	field6	latitude	longitude	elevation	status
2023-10-25 14:22:21	1	30.25	55.29883	940.12378	627.43036	3.11108	10.54670	NaN	NaN	NaN	NaN
2023-10-25 14:22:53	2	30.29	55.22168	940.17499	626.97729	4.65205	23.41554	NaN	NaN	NaN	NaN
2023-10-25 14:23:25	3	30.31	55.25781	940.21686	626.60712	5.33077	26.00724	NaN	NaN	NaN	NaN
2023-10-25 14:23:59	4	30.34	55.35645	940.24994	626.31433	3.89070	20.32056	NaN	NaN	NaN	NaN
2023-10-25 14:24:33	5	30.36	55.26465	940.25983	626.22729	4.81816	20.28843	NaN	NaN	NaN	NaN

Fig. 12. Original data with NaN values

datetime	temperature	humidity	pressure	altitude	pm10	pm25
2023-10-25 14:22:21	30.25	55.29883	940.12378	627.43036	3.11108	10.54670
2023-10-25 14:22:53	30.29	55.22168	940.17499	626.97729	4.65205	23.41554
2023-10-25 14:23:25	30.31	55.25781	940.21686	626.60712	5.33077	26.00724
2023-10-25 14:23:59	30.34	55.35645	940.24994	626.31433	3.89070	20.32056
2023-10-25 14:24:33	30.36	55.26465	940.25983	626.22729	4.81816	20.28843

Fig. 13. Data without Nan after preprocessing

datetime	temperature	humidity	pressure	altitude	pm10	pm25
2023-10-25 14:22:21	30.25	55.29883	940.12378	627.43036	3.11108	10.54670
2023-10-25 14:22:53	30.29	55.22168	940.17499	626.97729	4.65205	23.41554
2023-10-25 14:23:25	30.31	55.25781	940.21686	626.60712	5.33077	26.00724
2023-10-25 14:23:59	30.34	55.35645	940.24994	626.31433	3.89070	20.32056
2023-10-25 14:24:33	30.36	55.26465	940.25983	626.22729	4.81816	20.28843
...	...	...	...	...	...	...
2024-02-19 23:57:38	24.26	84.75488	938.31421	643.45105	23.65566	5.90028
2024-02-19 23:58:10	24.29	84.84082	938.28888	643.67529	23.50842	6.02359
2024-02-19 23:58:44	24.27	84.83691	938.28564	643.70416	24.08286	5.50573
2024-02-19 23:59:16	24.24	85.22363	938.28314	643.72614	22.43792	5.36404
2024-02-19 23:59:47	24.28	85.34570	938.30420	643.53937	24.98106	5.23587

309153 rows × 6 columns

Fig. 14. Original data with 309153 rows × 6 columns

datetime	temperature	humidity	pressure	altitude	pm10	pm25
2023-10-25 14:00:00	30.235522	56.438302	940.209413	626.677679	4.619386e+01	87.042663
2023-10-25 15:00:00	28.820000	61.992210	939.981103	628.697674	3.167633e+01	61.124179
2023-10-25 16:00:00	26.889722	67.488065	939.227603	635.362746	1.494000e+10	32.610734
2023-10-25 17:00:00	26.150091	68.413450	938.725489	639.813467	4.806097e+01	37.875024
2023-10-25 18:00:00	25.372232	68.702245	938.268517	643.855538	4.577520e+01	35.981206
...	...	...	...	...	...	...
2024-02-19 19:00:00	24.215804	80.431275	937.092879	654.278114	1.535785e+01	9.042224
2024-02-19 20:00:00	24.235877	81.019489	936.889732	656.079702	2.121988e+01	7.423256
2024-02-19 21:00:00	24.194513	80.774449	937.098456	654.229108	2.277330e+01	6.799083
2024-02-19 22:00:00	23.966579	80.730529	937.587148	649.898048	1.705176e+01	6.175659
2024-02-19 23:00:00	23.954867	82.506827	938.209286	644.383358	1.947449e+01	5.752773

2818 rows × 6 columns

Fig. 15. Data after resample with 2818 rows × 6 columns

3.3. Stationery Test

The result of the stationery test on this research for each feature is depicted in Table 2. Based on Table 2, the null hypothesis of non-stationarity has been rejected for temperature, PM1.0, and PM2.5 since the p-value for each feature is less than the significant value of 0.05, which means that those features will not need any differencing process to reach stationarity, while for humidity, pressure, and altitude, differencing processes will be applied to reach stationarity with the results shown in Table 3.

Table 2. ADF feature statistics result

Features	p-value	Stationery Result
Temperature	0.002391	Stationery
Humidity	0.128732	Non- Stationery
Pressure	0.133664	Non- Stationery
Altitude	0.236962	Non- Stationery
PM1.0	0.000018	Stationery
PM2.5	6.043795e-07	Stationery

Table 3. ADF results after 1st differencing

Features	p-value	Stationery Result
Humidity	3.338514263078631e-24	Stationery
Pressure	2.4108133522831695e-19	Stationery
Altitude	2.002588237480572e-17	Stationery

From Table 3, it can be seen that the p-values for humidity, pressure, and altitude are already less than the significant value of 0.05, which means that those features have already reached stationarity. Stationarity is essential for time series modeling, especially for ARIMA models, because it ensures that the series' statistical properties (mean, variance, and autocorrelation) remain constant over time, enabling stable and reliable predictions. Non-stationary data can result in biased and unreliable forecasts due to changing data patterns.

The differencing process transforms non-stationary data into stationary data by subtracting the previous observation from the current one, removing trends and non-stationary components. If necessary, further differencing can be applied until stationarity is achieved. The results in Table 3 demonstrate that after applying first differencing to the initially non-stationary features (humidity, pressure, and altitude), they became stationary, as evidenced by their p-values being below the 0.05 significance level.

3.4. ARIMA Models

ACF (Autocorrelation Function) is a function that will give information on the value of the autocorrelation for time series data with lagged values [48], [49]. The ACF will describe the relationship between present and past values using time-series components including seasonality, trend, cyclic, and residual. At the same time, PACF finds the correlations using residuals with the next lag [50]. Based on the ACF and PACF plots for each feature, it can be derived with ARIMA models (p, q) for each feature. ACF and PACF plots for each feature are shown in Fig. 16-Fig. 21.

From Fig. 12–Fig. 17, the p values for the ARIMA model are derived by observing significant spikes in the PACF plots that exceed the boundary level. For temperature (Fig. 11), p is 3 based on spikes at the 1st, 2nd, and 3rd lags. Humidity (Fig. 12) has a p value of 2 from the first lag spike. Pressure (Fig. 13) and altitude (Fig. 14) both have p values of 5, with pressure showing spikes at the 1st, 2nd, and 3rd lags and altitude at the 1st, 2nd, and 4th lags. PM1.0 (Fig. 15) has a p value of 2, and PM2.5 (Fig. 16) has a p value of 1, both based on spikes at the 1st lag. Meanwhile, to derive the q order from the moving average (MA) model from ACF, this plot has a sharp cut-off after lag q. The q values for the ARIMA model, based on the ACF plots, are as follows: For temperature (Fig. 11), q is 1 due to a significant spike at the first lag. Humidity (Fig. 12) has a q value of 2 from the first lag spike. Pressure (Fig. 13) and altitude (Fig. 14) both have q values of 2, with spikes at the 1st and 2nd lags. PM1.0 (Fig. 15) has a q value of 2, and PM2.5 (Fig. 16) has a q value of 1, both based on spikes at the 1st lag.

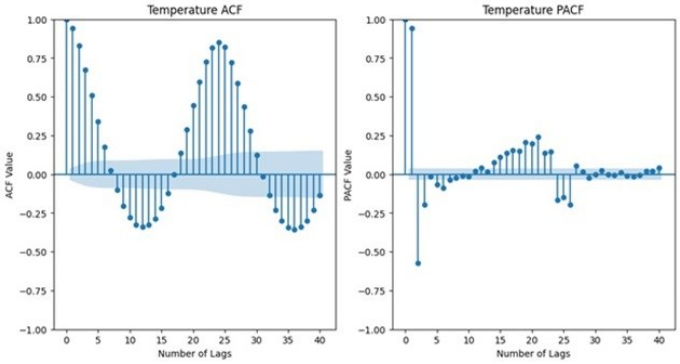


Fig. 16. Temperature ACF and PACF plot

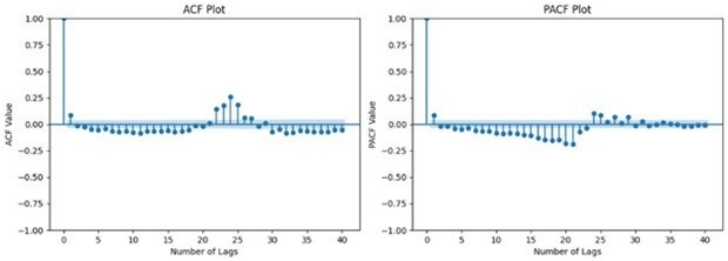


Fig. 17. Humidity ACF and PACF plot

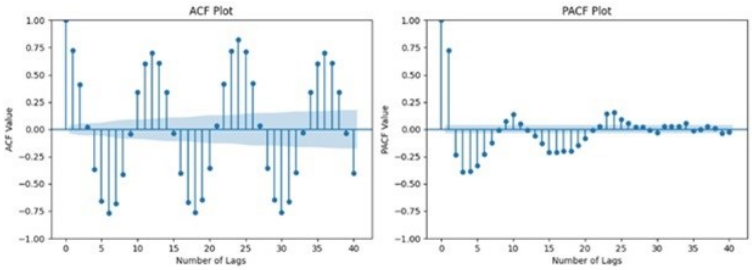


Fig. 18. Pressure ACF and PACF plot

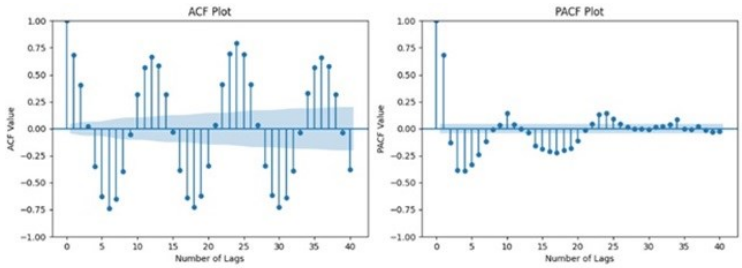


Fig. 19. Altitude ACF and PACF plot

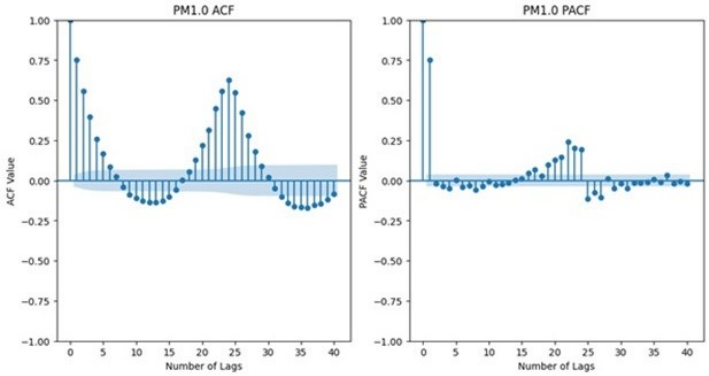


Fig. 20. PM1.0 ACF and PACF plot

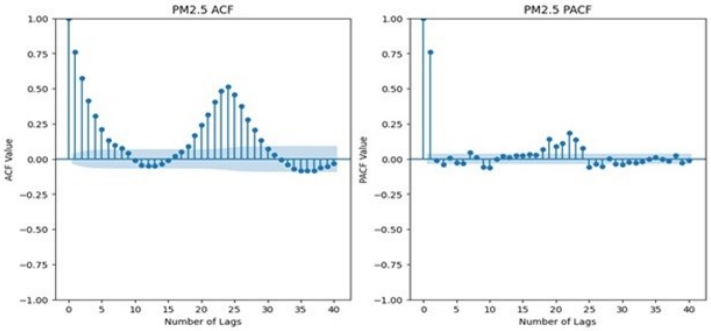


Fig. 21. PM2.5 ACF and PACF plot

In selecting the specific ARIMA orders for each feature, the rationale is based on analyzing the significant spikes in the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, as well as theoretical considerations of the underlying data patterns. The order p for the autoregressive (AR) model is determined from the PACF plots, where significant spikes above the boundary level indicate the number of lags that have a meaningful correlation with the current value of the time series. Similarly, the order q for the moving average (MA) model is derived from the ACF plots, where significant spikes indicate the lags that are relevant for modeling the error terms. These orders were selected to capture the relevant temporal dependencies and autocorrelations, ensuring accurate and reliable predictions. ARIMA models based on ACF, PACF plot observation, and ARIMA equations based on equation (3) for each feature are shown in Table 4.

The best ARIMA models for each environmental feature exhibit distinct characteristics and forecasting capabilities. The ARIMA (3,0,1) model for temperature effectively captures short-term fluctuations, leveraging three lagged temperature values and one lagged residual term. Humidity's ARIMA (1,1,1) model, with one lagged humidity value and one lagged residual term, adeptly considers persistence and trend in humidity changes. Pressure and altitude, both modeled with ARIMA (3, 1 ,2), demonstrate significant autoregressive and moving average effects, reflecting their temporal dynamics and trends. PM1.0 and PM2.5, modeled with ARIMA (1,0,1), showcase moderate autoregressive and moving average effects, crucial for short-term forecasting of particulate matter concentrations.

Table 4. ARIMA model and equations

Features	ARIMA Model (p,d,q)	ARIMA model (p,d,q) equation
Temperature	(3,0,1)	$y_t = 1.5444 + 1.9152y_{t-1} - 1.1041y_{t-2} + 0.1323y_{t-3} - 0.5480\varepsilon_{t-1} + \varepsilon_t$
Humidity	(1,1,1)	$y_t = 0.078y_{t-1} - 1.0000\varepsilon_{t-1} + \varepsilon_t$
Pressure	(3,1,2)	$y_t = 1.9772y_{t-1} - 1.4248y_{t-2} - 0.2456y_{t-3} + 1.7055\varepsilon_{t-1} + 1.9784\varepsilon_{t-2} + \varepsilon_t$
Altitude	(3,1,2)	$y_t = 1.8996y_{t-1} - 1.2906y_{t-2} - 0,1680y_{t-3} - 1.7014\varepsilon_{t-1} - 0.9790\varepsilon_{t-2} + \varepsilon_t$
PM1.0	(1,0,1)	$y_t = 24.3149 + 0.7414y_{t-1} + 0.0229\varepsilon_{t-1} + \varepsilon_t$
PM2.5	(1,0,1)	$y_t = 15.0288 + 0.7660y_{t-1} - 0.0234\varepsilon_{t-1} + \varepsilon_t$

3.5. Prediction and forecasting

From the research simulation, the prediction results for each feature are shown in Fig. 22- Fig. 27, and prediction evaluation performance using RMSE, MAE, and error ratio will be shown in Table 5.

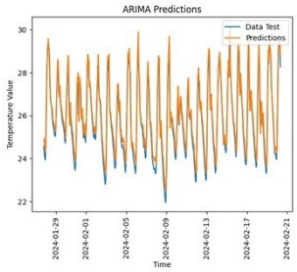


Fig. 22. Temperature prediction result

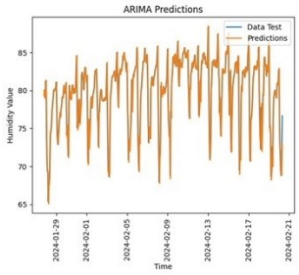


Fig. 23. Humidity prediction result

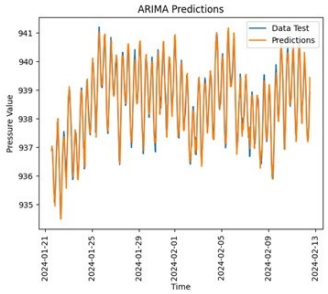


Fig. 24. Pressure prediction result

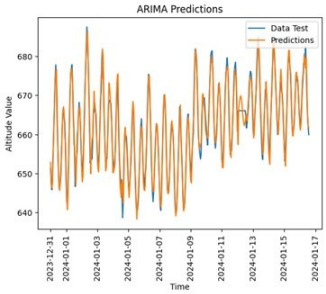


Fig. 25. Altitude prediction result

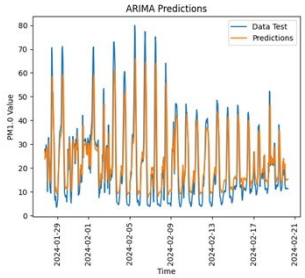


Fig. 26. PM1.0 prediction result

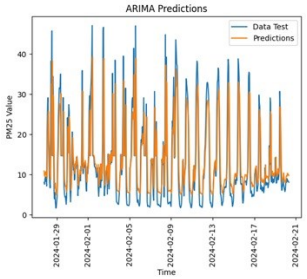


Fig. 27. PM2.5 prediction result

Table 5. Prediction Performance Evaluation

Features	RMSE	MAE
Temperature	0.74	0.43
Humidity	2.11	1.36
Pressure	0.25	0.19
Altitude	2.18	1.70
PM1.0	8.29	6.16
PM2.5	6.60	4.49

Based on Table 5, the prediction evaluation metrics indicate that the ARIMA model performs well for temperature, humidity, pressure, and altitude, with low RMSE and MAE values, signifying accurate and reliable predictions relative to the data's variability. However, the model's performance for PM1.0 and PM2.5 is moderate, with higher RMSE and MAE values reflecting the challenges in predicting particulate matter concentrations influenced by emissions, weather changes, and human activities, suggesting that these predictions are less reliable and could benefit from further refinement. Overall, while the model is effective for most features, improvements are needed for predicting particulate matter concentrations.

These findings highlight the strengths and limitations of the ARIMA models used in the AIOT-Particle device, emphasizing the need for model refinement, additional variables, or advanced machine learning techniques to improve prediction accuracy in the next research. Accurate predictions of parameters like temperature, humidity, and particulate matter can inform policy formulation, resource allocation, and health advisories, thereby improving public health outcomes and environmental sustainability.

The ARIMA model that has been tested to predict each feature is then used to forecast each feature for the next 24 hours. ARIMA's accuracy rate when forecasting is highly dependent on the data fluctuation based on external factors such as weather conditions, particulate matter concentration in the air, and sudden weather change. The forecasting result for each feature is shown in Table 6, while the forecast visualization for each feature is shown in Fig. 28–Fig. 33.

Table 6. ARIMA Forecasting Values

Features	1st hours	2nd hours	3rd hours	4th hours	5th hours	…	24th hours
Temperature	26.36	26.62	26.92	27.25	27.59	…	27.47
Humidity	82.49	82.11	81.74	81.42	81.17	…	80.26
Pressure	936.34	936.87	937.52	938.1	938.46	…	938.12
Altitude	639.56	636.53	636.04	638.2	642.43	…	644.34
PM1.0	14.56	17.08	18.95	20.34	21.36	…	24.29
PM2.5	9.65	10.87	11.8	12.5	13.03	…	14.65

The forecast data indicates a general warming trend, with temperatures rising from 26.36°C to a peak of 27.59°C by the fifth hour, then slightly cooling to 27.47°C by the 24th hour. Humidity is expected to decrease steadily from 82.49% to 80.26%, likely due to the rising temperatures. Pressure shows a minimal initial increase from 936.34 hPa to 938.46 hPa before gradually declining to 938.12 hPa, possibly indicating changing weather patterns. Altitude readings fluctuate slightly, suggesting potential sensor recalibrations or minor atmospheric pressure changes. Both PM1.0 and PM2.5 levels are forecast to rise significantly, indicating worsening air quality, which will become a concern for industrial stakeholders.

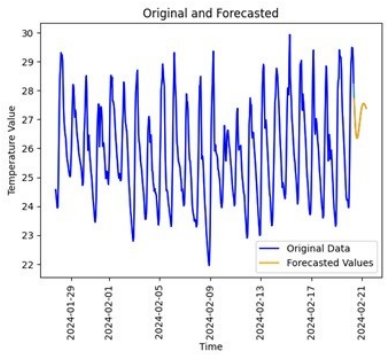


Fig. 28. Temperature forecasting result

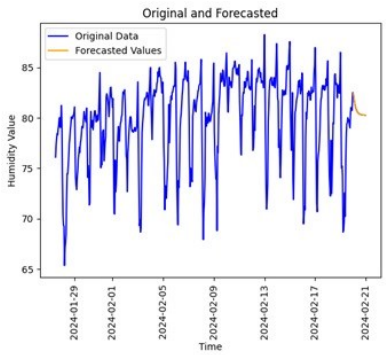


Fig. 29. Humidity forecasting result

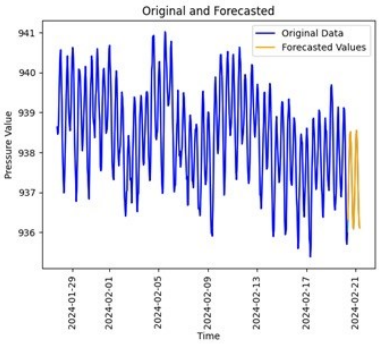


Fig. 30. Pressure forecasting result

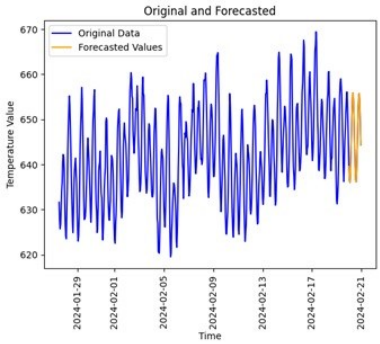


Fig. 31. Altitude forecasting result

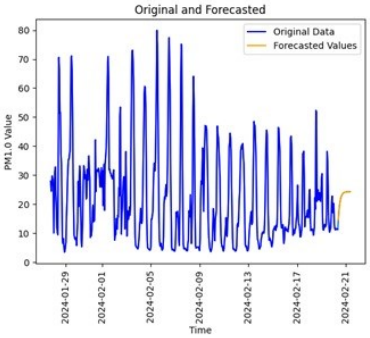


Fig. 32. PM1.0 forecasting result

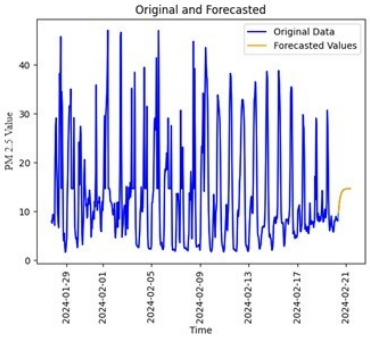


Fig. 33. PM2.5 forecasting result

However, this study has some limitations, including potential sources of error in sensor data, assumptions made during the modeling process, and notably, the lack of display for forecast data to users. This limitation could impede practical decision-making and the usability of the forecasting results for real-time applications. To address this, future work should focus on developing an integrated display system for forecast data, ensuring that the information is accessible and actionable for users. Additionally, refining the ARIMA models for better accuracy, incorporating more features or external variables such as wind speed or industrial activity levels, and exploring alternative forecasting techniques like machine learning models could enhance the robustness and reliability of predictions for future research.

4. CONCLUSION

The AIOT-Particle device was developed and tested to monitor real-time air quality and environmental data. The ARIMA model was used to analyze this data, with optimal parameters identified for each feature, demonstrating effective prediction performance. The ARIMA models showed varying levels of effectiveness for different environmental parameters: temperature (ARIMA(3,0,1), RMSE 0.74, MAE 0.43), humidity (ARIMA(1,1,1), RMSE 2.11, MAE 1.36), pressure and altitude (both ARIMA(3,1,2), RMSE 2.18, MAE 0.70), PM1.0 (ARIMA(1,0,1), RMSE 8.13, MAE 6.04), and PM2.5 (ARIMA(1,0,1), RMSE 6.60, MAE 4.49). While effective for most parameters, the models need refinement for particulate matter predictions.
Meanwhile, the ARIMA model equations for each feature, respectively, are;

$$y_t = 1.5444 + 1.9152y_{t-1} - 1.1041y_{t-2} + 0.1323y_{t-3} - 0.5480\varepsilon_{t-1} \quad \text{for temperature,}$$
$$y_t = 0.078y_{t-1} - 1.0000\varepsilon_{t-1} + \varepsilon_t \quad \text{for humidity,}$$
$$y_t = 1.9772y_{t-1} - 1.4248y_{t-2} - 0.2456y_{t-3} + 1.7055\varepsilon_{t-1} + 1.9784\varepsilon_{t-2} + \varepsilon_t \quad \text{for pressure,}$$
$$y_t = 1.8996y_{t-1} - 1.2906y_{t-2} - 0.1680y_{t-3} + 1.7014\varepsilon_{t-1} + 0.9790\varepsilon_{t-2} + \varepsilon_t \quad \text{for altitude,}$$
$$y_t = 24.3149 + 0.7414y_{t-1} + 0.0229\varepsilon_{t-1} + \varepsilon_t \quad \text{for PM1.0 and}$$
$$y_t = 15.0288 + 0.7660y_{t-1} - 0.0234\varepsilon_{t-1} + \varepsilon_t \quad \text{for PM2.5}$$

The AIOT-Particle device and ARIMA models offer promising results for real-time environmental monitoring and air quality forecasting due to their ease of implementation. However, more advanced methods provide higher accuracy but are more complex and resource-intensive. The AIOT-Particle device, with its potential for scalability and integration with machine learning, could balance simplicity and accuracy, making it valuable for environmental monitoring. This article introduces a novel, compact IoT-based device that monitors multiple environmental parameters, including PM1.0 and PM2.5, in real time.

Despite limitations such as sensor inaccuracies and the need for improved data display, the study highlights the potential of ARIMA models in environmental monitoring. Future research should refine these models, incorporate additional variables, and explore advanced machine learning techniques to enhance accuracy. This will help confirm the reliability and effectiveness of the AIOT-Particle device and ARIMA models across various geographical locations and environmental conditions, facilitating their widespread adoption for air quality monitoring and environmental data analysis.

Acknowledgments

The research here is supported by internal research funding from Satya Wacana Christian University for the year 2023.

REFERENCES

[1] M. Surdacki and M. Sobieszczańska, "The impact of particulate matter on prenatal and infant child development," *Alergol. Pol. - Polish J. Allergol.*, vol. 9, no. 3, pp. 179–185, 2022, https://doi.org/10.5114/pja.2022.119230.

[2] L. Zhang *et al.*, "Indoor particulate matter in urban households: Sources, pathways, characteristics, health effects, and exposure mitigation," *Int. J. Environ. Res. Public Health*, vol. 18, no. 21, 2021, https://doi.org/10.3390/ijerph182111055.

[3] B. Bessagnet *et al.*, "Emissions of Carbonaceous Particulate Matter and Ultrafine Particles from Vehicles—A Scientific Review in a Cross-Cutting Context of Air Pollution and Climate Change," *Appl. Sci.*, vol. 12, no. 7, 2022, https://doi.org/10.3390/app12073623.

[4] P. Maciejczyk, L. C. Chen, and G. Thurston, "The role of fossil fuel combustion metals in PM2.5 air pollution health associations," *Atmosphere (Basel).*, vol. 12, no. 9, pp. 1–34, 2021, https://doi.org/10.3390/atmos12091086.

[5] A. L. Moreno-Ríos, L. P. Tejeda-Benítez, and C. F. Bustillo-Lecompte, "Sources, characteristics, toxicity, and control of ultrafine particles: An overview," *Geosci. Front.*, vol. 13, no. 1, 2022, https://doi.org/10.1016/j.gsf.2021.101147.

[6] F. de A. Lima, G. B. Medeiros, P. A. M. Chagas, M. L. Aguiar, and V. G. Guerra, "Aerosol Nanoparticle Control by Electrostatic Precipitation and Filtration Processes—A Review," *Powders*, vol. 2, no. 2, pp. 259–298, 2023, https://doi.org/10.3390/powders2020017.

[7] T. F. Mebrahtu *et al.*, "The effects of exposure to NO 2 , PM 2 . 5 and PM 10 on health service attendances with respiratory illnesses : A time-series analysis ☆," *Environ. Pollut.*, vol. 333, no. 2, p. 122123, 2023, https://doi.org/10.1016/j.envpol.2023.122123.

[8] I. Chandra, K. Nisa, and E. Rosdiana, "Preliminary study : health risk analysis of PM2 . 5 and PM10 mass concentrations in Bandung Metropolitan," *IOP Conference Series: Earth and Environmental Science*, vol. 824, no. 1, p. 012049, 2021, https://doi.org/10.1088/1755-1315/824/1/012049.

[9] S. Dimitroulopoulou *et al.*, "Indoor air quality guidelines from across the world: An appraisal considering energy saving, health, productivity, and comfort," *Environ. Int.*, vol. 178, no. June, 2023, https://doi.org/10.1016/j.envint.2023.108127.

[10] S. Lim *et al.*, "Comparing human exposure to fine particulate matter in low and high-income countries: A systematic review of studies measuring personal PM2.5 exposure," *Sci. Total Environ.*, vol. 833, no. January, p. 155207, 2022, https://doi.org/10.1016/j.scitotenv.2022.155207.

[11] J. A. Saju, Q. H. Bari, K. A. B. M. Mohiuddin, and V. Strezov, "Measurement of ambient particulate matter (PM1.0, PM2.5 and PM10) in Khulna City of Bangladesh and their implications for human health," *Environ. Syst. Res.*, vol. 12, no. 1, 2023, https://doi.org/10.1186/s40068-023-00327-2.

[12] L. T. Molina, "Introductory lecture: Air quality in megacities," *Faraday Discuss.*, vol. 226, pp. 9–52, 2021, https://doi.org/10.1039/d0fd00123f.

[13] Health Effects Institute, "State of Global Air 2019," *Heal. Eff. Institute.*, p. 24, 2019, https://www.stateofglobalair.org/sites/default/files/soga_2019_report.pdf.

[14] M. Wang *et al.*, "Design of PM2.5 monitoring and forecasting system for opencast coal mine road based on internet of things and ARIMA Mode," *PLoS One*, vol. 17, no. 5 May, pp. 1–28, 2022, https://doi.org/10.1371/journal.pone.0267440.

[15] S. Mancini, A. Francavilla, G. Graziuso, and C. Guarnaccia, "An Application of ARIMA modelling to air pollution concentrations during covid pandemic in Italy," *J. Phys. Conf. Ser.*, vol. 2162, no. 1, 2022, https://doi.org/10.1088/1742-6596/2162/1/012009.

[16] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," *Futur. Internet*, vol. 15, no. 8, pp. 1–31, 2023, https://doi.org/10.3390/fi15080255.

[17] M. Wati, A. Haviluddin, A. Masyudi, H. Septiarini, and H. Rahmania, "Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Indonesian Crude Oil Price," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 3, pp. 720–730, 2023, https://doi.org/10.26555/jiteki.v9i3.22286.

[18] A. L. Schaffer, T. A. Dobbins, and S. A. Pearson, "Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions," *BMC Med. Res. Methodol.*, vol. 21, no. 1, pp. 1–13, 2021, https://doi.org/10.1186/s12874-021-01235-8.

[19] J. Wang, Z. K. Pei, Y. Wang, and Z. Qin, "An investigation of income inequality through autoregressive integrated

moving average and regression analysis," *Healthc. Anal.*, vol. 5, no. October 2023, p. 100287, 2024, https://doi.org/10.1016/j.health.2023.100287.
[20] J. Kaur, K. S. Parmar, and S. Singh, "Autoregressive models in environmental forecasting time series: a theoretical and application review," *Environ. Sci. Pollut. Res.*, vol. 30, no. 8, pp. 19617–19641, 2023, https://doi.org/10.1007/s11356-023-25148-9.
[21] M. Amjad, A. Khan, K. Fatima, O. Ajaz, S. Ali, and K. Main, "Analysis of Temperature Variability, Trends and Prediction in the Karachi Region of Pakistan Using ARIMA Models," *Atmosphere (Basel).*, vol. 14, no. 1, p. 88, 2022, https://doi.org/10.3390/atmos14010088.
[22] R. N. Putri, M. Usman, Warsono, Widiarti, and E. Virginia, "Modeling Autoregressive Integrated Moving Average (ARIMA) and Forecasting of PT Unilever Indonesia Tbk Share Prices during the COVID-19 Pandemic Period," *J. Phys. Conf. Ser.*, vol. 1751, no. 1, 2021, https://doi.org/10.1088/1742-6596/1751/1/012027.
[23] F. R. Alharbi and D. Csala, "A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach," *Inventions*, vol. 7, no. 4, 2022, https://doi.org/10.3390/inventions7040094.
[24] R. Ospina, J. A. M. Gondim, V. Leiva, and C. Castro, "An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil," *Mathematics*, vol. 11, no. 14, pp. 1–18, 2023, https://doi.org/10.3390/math11143069.
[25] M. H. Jabardi, H. T. Kurmasha, and R. R. Al-Khalidy, "Forecasting Weekly COVID-19 Infection and Death Cases in Iraq Using an ARIMA Model," *Tech. Rom. J. Appl. Sci. Technol.*, vol. 4, no. 2, pp. 64–75, 2022, https://doi.org/10.47577/technium.v4i2.6191.
[26] A. A. Ibrahim, B. N. Saeed, and M. A. Fadil, "Forecasting Stock Prices with an Integrated Approach Combining ARIMA and Machine Learning Techniques ARIMAML," *J. Comput. Commun.*, vol. 11, no. 08, pp. 58–70, 2023, https://doi.org/10.4236/jcc.2023.118005.
[27] O. M. Salah, G. J. M. Mahdi, and I. A. A. Al-Latif, "A modified ARIMA model for forecasting chemical sales in the USA," *J. Phys. Conf. Ser.*, vol. 1879, no. 3, 2021, https://doi.org/10.1088/1742-6596/1879/3/032008.
[28] A. S. Ahmar, M. Botto-Tobar, A. Rahman, and R. Hidayat, "Forecasting the Value of Oil and Gas Exports in Indonesia using ARIMA Box-Jenkins," *JINAV J. Inf. Vis.*, vol. 3, no. 1, pp. 35–42, 2022, https://doi.org/10.35877/454ri.jinav260.
[29] J. Wang, T. Ji, and M. Li, "A combined short-term forecast model of wind power based on empirical mode decomposition and augmented dickey-fuller test," *J. Phys. Conf. Ser.*, vol. 2022, no. 1, 2021, https://doi.org/10.1088/1742-6596/2022/1/012017.
[30] R. R. Barry and I. Bernarto, "Spurious Regression Analysis on Time Series Data From Factors Affecting Indonesian Human Development Indexs in 1990 – 2017," *JMBI UNSRAT (Jurnal Ilm. Manaj. Bisnis dan Inov. Univ. Sam Ratulangi)*., vol. 7, no. 3, 2021, https://doi.org/10.35794/jmbi.v7i3.30608.
[31] Y. Al Moaiad *et al.*, "The Internet of Things as A Revolution to Enhance the Technology of the Future," *Int. J. Spec. Educ.*, vol. 37, no. 3, pp. 2022–6521, 2022, https://doi.org/10.13140/RG.2.2.19815.11686.
[32] L. Stošić, M. Dimitrovska, L. Pushova Stamenkova, and M. Smelcerović, "From Concept To Reality: Understanding the Internet of Things," *Sci. Int. J.*, vol. 2, no. 4, pp. 181–184, 2023, https://doi.org/10.35120/sciencej0204181s.
[33] Z. Abbas "Internet of Things (IoT) in the Cloud : Connecting and Managing Smart Devices," 2023, https://osf.io/preprints/osf/8k69a.
[34] S. A. Bkheet and J. I. Agbinya, "A Review of Identity Methods of Internet of Things (IOT)," *Adv. Internet Things*, vol. 11, no. 4, pp. 153–174, 2021, https://doi.org/10.4236/ait.2021.114011.
[35] G. Mohindru, K. Mondal, H. Banka, "Internet of Things and data analytics: A current review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1341, 2019, https://doi.org/10.1002/widm.1341.
[36] T. Gusman, M. Naeemullah, and A. M. Qasim, "Big Data Processing: A review," *Mesopotamian J. Big Data*, pp. 23–30, 2022, https://doi.org/10.58496/mjbd/2022/003.
[37] M. E. E. Alahi *et al.*, "Integration of IoT-Enabled Technologies and Artificial Intelligence (AI) for Smart City Scenario: Recent Advancements and Future Trends," *Sensors*, vol. 23, no. 11, 2023, https://doi.org/10.3390/s23115206.
[38] H. Allioui and Y. Mourdi, "Exploring the Full Potentials of IoT for Better Financial Growth and Stability: A Comprehensive Survey," *Sensors*, vol. 23, no. 19, 2023, https://doi.org/10.3390/s23198015.
[39] T. Zachariah, N. Klugman, and P. Dutta, *ThingSpeak in the Wild: Exploring 38K Visualizations of IoT Data*, vol. 1, no. 1. Association for Computing Machinery, 2022. https://doi.org/10.1145/3560905.3567766.
[40] N. Sindhwani, R. Anand, R. Vashisth, S. Chauhan, V. Talukdar, and D. Dhabliya, "Thingspeak-Based Environmental Monitoring System Using IoT," *PDGC 2022 - 2022 7th Int. Conf. Parallel, Distrib. Grid Comput.*, no. April, pp. 675–680, 2022, https://doi.org/10.1109/PDGC56933.2022.10053167.
[41] T. Christen, M. Hess, D. Grichnik, and J. Wincent, "Value-based pricing in digital platforms: A machine learning approach to signaling beyond core product attributes in cross-platform settings," *J. Bus. Res.*, vol. 152, no. August 2021, pp. 82–92, 2022, https://doi.org/10.1016/j.jbusres.2022.07.042.
[42] A. Jierula, S. Wang, T. M. Oh, and P. Wang, "Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data," *Appl. Sci.*, vol. 11, no. 5, pp. 1–21, 2021, https://doi.org/10.3390/app11052314.
[43] S. M. Robeson and C. J. Willmott, "Decomposition of the mean absolute error (MAE) into systematic and unsystematic components," *PLoS One*, vol. 18, no. 2 February, pp. 1–8, 2023,

https://doi.org/10.1371/journal.pone.0279774.
[44] P. K. Jacobson, L. Lind, and H. L. Persson, "Unleashing the Power of Very Small Data to Predict Acute Exacerbations of Chronic Obstructive Pulmonary Disease," *Int. J. COPD*, vol. 18, no. June, pp. 1457–1473, 2023, https://doi.org/10.2147/COPD.S412692.
[45] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, 2022, https://doi.org/10.5194/gmd-15-5481-2022.
[46] A. Jadon, A. Patil, and S. Jadon, "A Comprehensive Survey of Regression Based Loss Functions for Time Series Forecasting," 2022, [Online]. Available: http://arxiv.org/abs/2211.02989.
[47] K. Przystupa *et al.*, "Standard Deviation in the Simulation of Statistical Measurements," *Metrol. Meas. Syst.*, vol. 30, no. 1, pp. 17–30, 2023, https://doi.org/10.24425/mms.2023.144403.
[48] U. A. Yakubu, M. Panji, and A. Saputra, "Time Series Model Analysis Using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for E-wallet Transactions during a Pandemic," vol. 3, no. 3, pp. 80–85, 2022, https://doi.org/10.47194/ijgor.v3i3.168.
[49] G. Vishnu, D. Kaliyaperumal, P. B. Pati, A. Karthick, N. Subbanna, and A. Ghosh, "Short-Term Forecasting of Electric Vehicle Load Using Time Series, Machine Learning, and Deep Learning Techniques," *World Electr. Veh. J.*, vol. 14, no. 9, pp. 1–17, 2023, https://doi.org/10.3390/wevj14090266.
[50] C. H. Weiß, B. Aleksandrov, M. Faymonville, and C. Jentsch, "Partial Autocorrelation Diagnostics for Count Time Series," *Entropy*, vol. 25, no. 1, pp. 1–21, 2023, https://doi.org/10.3390/e25010105.

BIOGRAPHY OF AUTHORS

Johanes Dian Kurniawan, Bachelor's degree in faculty of Electrical Engineering and Master's degree in Data Science at Satya Wacana Christian University. He is interested in Machine Learning, Deep Learning, Neural Network and IoT. He can be contacted by email: 632022001@student.uksw.edu.

Hanna Arini Parhusip holds a Doctor of Mathematics degree from ITB, 2005. She received her B.Sc. of Mathematics from Gadjah Mada University in Indonesia and her M.Sc. (industrial mathematics) from Kaiserslautern University, Germany in 1992 and 1997, respectively. She is currently an associate professor in the Master of Data Science Department at Satya Wacana Christian University in Indonesia. She has been one of the founders in the Master program in Data Science at this university since 2020 and has been developing the application of artificial intelligence for mining and the Internet of Things for several purposes in some surrounding companies as a collaboration between the university and the industries. Her research includes machine learning, artificial intelligence, internet of things, and the application of ordinary differential equations. She has published over 10 papers in international journals and conferences in the last 5 years. She has been the head of the Master program in Data Science. She can be contacted at email: hanna.parhusip@uksw.edu.

Suryasatriya Trihandaru holds a Doctor of Mathematics degree from ITB, 2005. He received her B.Sc. in Physics from Gadjah Mada University in Indonesia and M.Sc. (industrial Mathematics) from Kaiserslautern University, in 1992 and 1998, respectively. He is currently an associate professor in the Master of Data Science Department at Satya Wacana Christian University in Indonesia. He is one of the founders of Master program in Data Science in this university since 2020 and has been developing the application of Artificial Intelligence for mining and the Internet of Things for several purposes in some surrounding companies as a collaboration between the university and the industries. His research includes machine learning, artificial intelligence, internet of things, and medical physics. He has published over 10 papers in international journals and conferences in the last 5 years. He can be contacted by email: suryasatriya@uksw.edu.