miptah4@gmail.com 1

Kolaborasi.pdf

Class AA 9

Class AA

Universitas Pendidikan Muhammadiyah Sorong

Document Details

Submission ID
trn:oid:::1:3302043601

Submission Date
Jul 25, 2025, 7:32 AM GMT+7

Download Date
Jul 25, 2025, 7:35 AM GMT+7

File Name
Kolaborasi.pdf

File Size
1.6 MB

13 Pages

6,820 Words

32,433 Characters

6% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Small Matches (less than 10 words)

Match Groups

22 Not Cited or Quoted 6%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
4% Publications
4% Submitted works (Student Papers)

Match Groups

- **22** Not Cited or Quoted 6%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 5% Internet sources
- 4% Publications
- 4% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	
RMIT University		1%
2	Internet	
www.biorxiv.org		<1%
3	Internet	
journal.gpp.or.id		<1%
4	Publication	
M. Pięk, K. Pawłuszek-Filipiak. "Displacement time series forecasting and anomal...		<1%
5	Publication	
Mohammad Abu Baker Siddique, Balaram Mahalder, Mohammad Mahfujul Haqu...		<1%
6	Publication	
Utriweni Mukhaiyar, Nur'ainul Miftahul Huda, Indah Gumala Andirasdini, Udjiann...		<1%
7	Student papers	
Universitas Brawijaya		<1%
8	Internet	
ejournal.uin-malang.ac.id		<1%
9	Internet	
pubs.ascee.org		<1%
10	Student papers	
School of Business and Management ITB		<1%

11 Publication

Azadeh, A.. "Integration of genetic algorithm, computer simulation and design of... <1%

12 Student papers

UCL <1%

13 Internet

autodocbox.com <1%

14 Internet

ojs3.unpatti.ac.id <1%

15 Internet

pure.bond.edu.au <1%

16 Publication

Deby Fakhriyana, Irhamah, Kartika Fithriasari. "Modeling Jakarta composite inde... <1%

17 Publication

Olivier Darné. "The uncertain unit root in real GNP: A re-examination", Journal of ... <1%

18 Internet

arxiv.org <1%

19 Internet

www.mdpi.com <1%

turnitin Page 5 of 17 - Integrity Submission Submission ID trn:oid:::1:3302043601

CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 427-439
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Combining IoT and Time Series Model for Minute-Level Outlier Detection in Wind Speed Forecasting

Nur'ainul Miftahul Huda[1*], Nurfitri Imro'ah[2], Rahmi Hidayati[3], and Kartika Sari[4]

[1]*1Department of Mathematics, Faculty of Mathematics and Natural Sciences, Indonesia, 78124*
[2]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Indonesia, 78124*
[3,4]*Department of Computer System Engineering, Faculty of Mathematics and Natural Sciences, Indonesia, 78124*

Abstract

An integrated strategy that combines Internet of Things (IoT) technology with time series modelling is presented in this work. This approach aims to identify outliers and estimate wind speed over short distances. Wind speed data was gathered at one-minute intervals from December 24, 2024, to January 10, 2025, utilising IoT-based anemometer sensors outdoors. The raw data was subsequently pre-processed by aggregating it into 5-minute intervals, which functioned as input for the ARIMA (Auto-Regressive Integrated Moving Average) model. This model was employed to discern temporal patterns and produce short-term wind speed predictions. Outliers were identified by analysing residuals between observed and projected values, facilitating the detection of anomalous wind behaviour. Combining continuous sensing based on the Internet of Things (IoT) with time series modelling for real-time environmental monitoring makes this approach unique. Experimental findings indicated that the ideal temporal dependency for simulating wind speed dynamics was a 65-minute window, aligning with an autoregressive order of AR(13) when employing 5-minute intervals. In six iterations of the outlier detection method, 87 outliers were detected. Integrating these outliers as exogenous variables in the ARIMA model enhanced forecasting precision. This underscores the significance of rectifying anomalies in time series models to improve predictive accuracy. The augmented ARIMA model with outlier corrections delivers precise, minute-level forecasts and dependable anomaly identification, rendering it appropriate for applications like renewable energy optimisation and early warning systems. The findings of this study underline the importance of incorporating modern statistical approaches with the Internet of Things (IoT) to enhance real-time decision-making about environmental and energy-related domains.

Keywords: extreme, accuracy, autoregressive, sensor

Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

1 Introduction

The Autoregressive Integrated Moving Average (ARIMA) time series model is an established technique extensively employed in studying non-stationary data, particularly for modelling and forecasting meteorological phenomena such as temperature, precipitation, and wind velocity [1]–[4]. The model functions by integrating autoregressive (AR), differencing (I), and moving average (MA) elements to capture short-term patterns and long-term trends. Several prior studies have utilised ARIMA to model wind speed on

*Corresponding author. E-mail: nurainul@fmipa.untan.ac.id

Submitted: ... Reviewed: ... Accepted: ...
DOI: https://doi.org/10.18860/cauchy.v10i1.32446

turnitin Page 5 of 17 - Integrity Submission Submission ID trn:oid:::1:3302043601

both daily and monthly scales, including research by some researchers [5]–[7]. When applied to high-resolution data, however, conventional ARIMA models frequently face a loss in performance. This is because these models do not explicitly handle the presence of outliers, which are common in short-time-dimensional weather data.

In the West Kalimantan region, wind speed exhibits considerable variability, particularly at the year's end during the monsoon changeover [8]. By the conclusion of 2024, a rise in severe winds, together with meteorological abnormalities, was documented at multiple monitoring sites. This phenomenon affects various sectors, including shipping safety, agriculture, and forest fire management [9]. Consequently, the demand for wind prediction systems more responsive to swift fluctuations escalates. The action of minute-by-minute wind speed data offers intricate temporal insights, yet presents obstacles due to significant volatility and numerous outliers resulting from local factors and monitoring equipment interference.

Advancements in Internet of Things (IoT) technology facilitate the acquisition of real-time atmospheric data at minute intervals. This detailed data enhances the precision of forecasts, particularly within locally-focused early warning systems [10], [11]. The primary issue with high-resolution data is the significant noise and the occurrence of outliers, including Additive Outliers (AO) and Innovation Outliers (IO), which can affect the noise model parameters. Some studies demonstrated the significance of outlier detection and correction in minute-level data to enhance the precision of ARIMA models; however, the methodology remains confined to small-scale experiments and has not been evaluated under extreme climate conditions in West Kalimantan [12]–[14].

This study presents an innovative approach to modelling minute-level wind speed forecasting with ARIMA, incorporating iterative outlier detection and adjustment. Few studies have thoroughly investigated the use of ARIMA for forecasting high-resolution data and an outlier-aware ARIMA methodology, which is suitable for integrating IoT-based real-time monitoring systems in extreme wind-prone areas like West Kalimantan.

2 Methods

ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is a prevalent statistical method in time series forecasting. Since its inception by Box and Jenkins in 1976, ARIMA has been extensively utilised across diverse domains, including economics, finance, meteorology, and epidemiology. Numerous prior studies have demonstrated the efficacy of ARIMA in modeling non-stationary data, including the research on stock price forecasting [15] and predicting daily COVID-19 case numbers [16]–[18]. The primary benefit of ARIMA is its proficiency in effectively managing historical data patterns via Autoregressive (AR), differencing (I), and Moving Average (MA) components. The ARIMA(p,d,q) model is characterized by [19],

$$\phi(B)(1-B)^d Y_t = \theta(B)e_t$$

where

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots)$$
$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots)$$
$$B^d Y_t = Y_{t-d}$$

The ARIMA approach begins by determining the stationarity of the data. If not, the differentiation process continues until the data attains stationarity. Subsequently, the identification and parameter estimation of the ARIMA(p,d,q) model are conducted, where p denotes the autoregressive order, d represents the degree of differencing, and q signifies the moving average order. Optimal model selection typically relies on the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Root Mean Square Error (RMSE), or Mean Absolute Percentage Error (MAPE) metrics. After establishing the model, a diagnostic test is conducted on the residuals to verify that the model satisfies the white noise

assumption [19]. The ultimate outcome of an effective ARIMA model can facilitate short-term forecasts with a comparatively narrow margin of error.

ARIMA Model with Outlier Factor

Outliers in time series data are points that drastically diverge from the overall trend and may result from multiple sources, including recorded inaccuracies, external disruptions, or infrequent occurrences (e.g., natural disasters or economic crises) [20], [21]. In time series analysis, outliers can be categorised into many categories, including Additive Outliers (AO), Innovation Outliers (IO), and level shifts. Identifying and categorising outliers compromises the model's accuracy [14].

The primary risk posed by outliers in time series data is their potential to undermine fundamental model assumptions, including stationarity and the normality of residuals. Outliers can lead to skewed parameter estimate findings, diminish forecasting accuracy, and elevate error metrics such as MAPE or RMSE [22]–[24]. In an ARIMA model, an extreme outlier might result in overfitting or erroneous model selection. Consequently, identifying and managing outliers is an essential phase before constructing a dependable time series model. A traditional visual technique for identifying outliers is the utilisation of box plots [25]. Upon detecting an outlier via the box plot, the subsequent action may involve eliminating the data if deemed erroneous or applying transformations such as winsorising, interpolation, or smoothing techniques to mitigate its impact. In time series analysis, it is crucial to ascertain whether outliers possess any particular significance (e.g., resulting from notable events) before making alterations or eliminations.

Two categories of outliers exist: Additive Outliers (AO) and Innovative Outliers (IO). Each sort of outlier serves a distinct function based on its influence. AO outliers directly influence when $t = T$, but IO outliers impact all subsequent observations following the detection of the outlier, specifically $Y_T, Y_{T+1}, \ldots$. The ARIMA model incorporating outlier factors is [26].

$$Y_t = \begin{cases} X_t + \omega I_t^{(T_A)} & \text{AO} \\ X_t + \frac{\theta(B)}{\phi(B)} \omega I_t^{(T_I)} & \text{IO} \end{cases}$$

with

$$I_t^{(T_u)} = \begin{cases} 1 & t = T \\ 0 & t \neq T \end{cases}$$

and $u = \{I, A\}$.

In general, the ARIMA model used when more than one outlier is detected is defined by

$$Y_t = \sum_{j=1}^{k} \omega_j v_j(B) I_t^{T_u} + X_t \tag{1}$$

where

$$X_t = \frac{\theta(B)}{\phi(B)} a_t$$

$$v_j(B) = \begin{cases} 1 & \text{AO} \\ \frac{\theta(B)}{\phi(B)} & \text{IO} \end{cases}$$

The methodology for identifying outliers in time series data is an iterative process [26], [27].

1. The algorithm's input consists of the residuals from the ARIMA model,

$$\hat{e}_t = \hat{\pi}(B) X_t$$
$$= \frac{\hat{\phi}(B)}{\hat{\theta}(B)} X_t$$

2. The detection phase commences with the computation of model residuals and their variance,

$$\hat{\sigma}_a^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t^2$$

be the initial estimate of σ_a^2

3. Determine the values of $\hat{\lambda}_{1,t}$ and $\hat{\lambda}_{2,t}$ for $t = 1, 2, \ldots, n$ where

$$\hat{\lambda}_{1,t} = \frac{\tau \hat{\omega}_{AT}}{\hat{\sigma}_a} \text{ and } \hat{\lambda}_{2,t} = \frac{\hat{\omega}_{IT}}{\hat{\sigma}_a}$$

Identifies the outlier type by establishing a criterion. If $\hat{\lambda}_T = |\hat{\lambda}_{1,t}| > C$, then AO is determined at time T, a constant valued at 3 or 4. If $\hat{\lambda}_T = |\hat{\lambda}_{2,t}| > C$, then IO is determined at the time T, where

$$\hat{\lambda}_T = \max\{\hat{\lambda}_{1,t}, \hat{\lambda}_{2,t}\}$$

4. Adjust the residuals using,

$$\tilde{e}_t = \begin{cases} \hat{e}_t - \hat{\omega}_A \hat{\pi}(B) I_t^{(T_A)} & AO \\ \hat{e}_t - \hat{\omega}_I I_t^{(T_I)} & IO \end{cases}$$

5. Reiterate stages three through four; the iterative process stops when no outliers are identified.

Once all outliers have been identified, the ARIMA modelling is re-executed by incorporating the outlier factor corresponding to the identified instances. After the model is conducted and all the parameters are estimated, repeat the first until the fifth process till no more outlier is detected. The ARIMA modelling technique using outlier factors is illustrated in Figure 1.

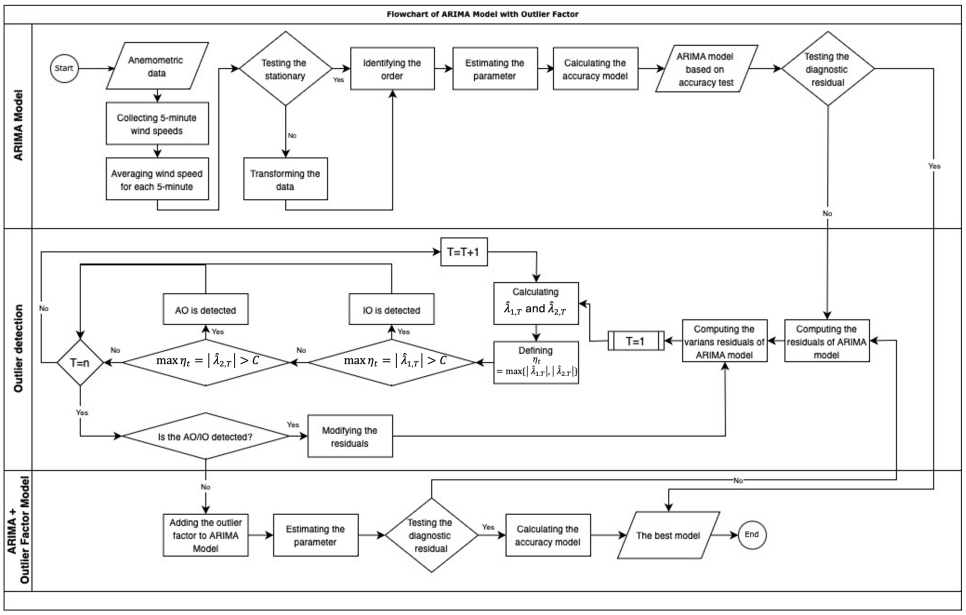


Figure 1: Flowchart ARIMA model, outlier detection, and ARIMA model + outlier factor model.

Wind Detection System

This rain and wind detection system is developed by merging multiple environmental sensors linked via a Wireless Sensor Network (WSN) [28]. Rain sensors employ optical or electrically conductive mechanisms to identify precipitation. In contrast, wind speed and direction are gauged using an anemometer and wind vane, which rely on magnetic or ultrasonic sensors [29]. Each sensor is linked to a microcontroller, such as the ESP32, which regularly gathers data. Sensor data is automatically calibrated for precision and transmitted via an Iot network utilising communication modules such as Wi-Fi, LoRa, or

GSM (e.g., SIM800L) to a server or cloud platform. Data is received and saved in a database for visualisation and analysis on the backend. The system facilitates real-time and historical monitoring, allowing for the prompt identification of severe weather and expeditious, data-informed decision-making [30].

3 Results and Discussion

Wind speed data is acquired by an automatic detection instrument, a cup-type anemometer that operates to catch wind gusts and turn them into rotational motion. Each cup rotation is measured by magnetic or optical sensors within the device, subsequently converted into speed units (meters per second). The apparatus is positioned at a conventional elevation of 10 meters above ground level, adhering to meteorological measurement regulations to prevent interference from structures or vegetation. Data recording was automated at one-minute intervals, yielding about 25,920 data points from 24 December 2024 to 10 January 2025. This data represents real-time wind speed variations, encompassing minor and significant swings.

The data was averaged at five-minute intervals to enhance analysis and mitigate the impact of brief, dramatic changes. This methodology was employed due to the propensity of 1-minute intervals to provide extremely volatile data, but not all minor fluctuations are pertinent for medium-term analytical requirements. A more stable representation of the wind speed trend is achieved by averaging every five minutes without compromising its fundamental properties. Consequently, the data volume was diminished to approximately 5,184 points, enhancing processing and visualisation ebiciency. The preprocessing phase involved addressing missing values via linear interpolation and identifying outliers using the interquartile range (IQR) approach, yielding a refined dataset prepared for subsequent analysis.

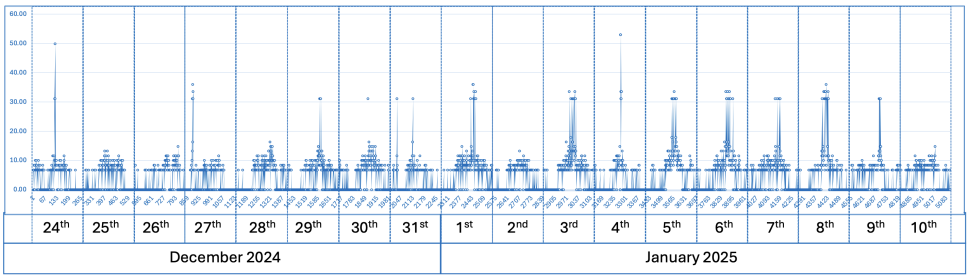
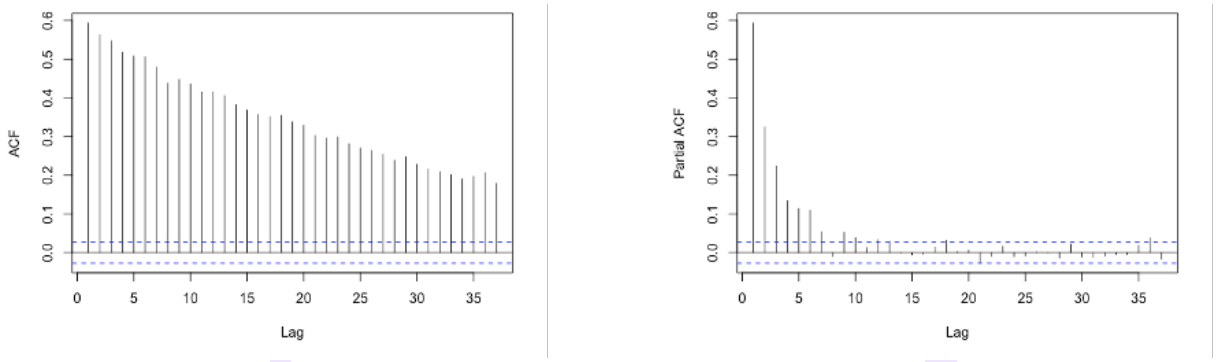


Figure 2: Time Series Data Plot.

The time series plots illustrate wind speed variations from 24 December 2024 to 10 January 2025 (shown in Figure 2), with data averaged at five-minute intervals to mitigate noise and enhance analytical stability. Most wind speed measurements remain consistent within the range of 0-10 m/s, indicative of the climatic traits of West Kalimantan, which typically experiences low to moderate strength surface winds. Nevertheless, distinct spikes manifest at specific intervals, signifying the existence of outliers or extraordinary wind phenomena. This phenomenon may be attributed to local atmospheric dynamics during the rainy season, often in West Kalimantan from late December to early January. These spikes may indicate actual occurrences, such as abrupt strong gusts preceding substantial rainfall, or may result from sensor interference. It is crucial to acknowledge the presence of outliers, as they might impair the ebicacy of predictive models like ARIMA if left unaddressed. Consequently, outlier detection and remediation are executed during data preprocessing to enhance predictive models' precision and facilitate IoT-based wind monitoring systems in tropical locations. The subsequent steps involve the implementation of the ARIMA model considering outlier factors:

1. ARIMA modelling, which includes the process of identifying the order through ACF and PACF plots (Figure 3). From Figure 3, according to the criteria for determining the order of the ARIMA model (Wei, 2006), the possible model orders are $p = 1, 7, 10, 13$ and $q = 0$. After identifying the order, the parameter estimation uses the Maximum Likelihood Estimation (MLE) method.



(a) (b)

Figure 3: (a) ACF Plot and (b) PACF Plot.

The parameter estimation results for each possible model are given in Table 1.

Table 1: ARIMA model based on ACF and PACF plot (Figure 3).

Ordo	Model
AR(1)	$Y_t = 4.09 + 0.59\,Y_{t-1}$
AR(7)	$Y_t = 4.09 + 0.26\,Y_{t-1} + 0.17\,Y_{t-2} + 0.14\,Y_{t-3} + 0.07\,Y_{t-4} + 0.07\,Y_{t-5} + 0.09\,Y_{t-6} + 0.05\,Y_{t-7}$
AR(10)	$Y_t = 4.09 + 0.26\,Y_{t-1} + 0.17\,Y_{t-2} + 0.13\,Y_{t-3} + 0.06\,Y_{t-4} + 0.07\,Y_{t-5} + 0.09\,Y_{t-6} + 0.04\,Y_{t-7}$ $-0.03\,Y_{t-8} + 0.04\,Y_{t-9} + 0.04\,Y_{t-10}$
AR(13)	$Y_t = 4.09 + 0.26\,Y_{t-1} + 0.17\,Y_{t-2} + 0.13\,Y_{t-3} + 0.06\,Y_{t-4} + 0.07\,Y_{t-5} + 0.07\,Y_{t-6} + 0.08\,Y_{t-7}$ $+0.04\,Y_{t-8} - 0.04\,Y_{t-9} + 0.03\,Y_{t-10} + 0.03\,Y_{t-11} - 0.01\,Y_{t-12} + 0.03\,Y_{t-13}$

The final phase of the modelling process is the diagnostic assessment. According to the diagnostic test in Table 2, none of the models satisfied normality and residual independence criteria. Consequently, the model exhibits outliers, necessitating the implementation of an outlier detection algorithm based on model residuals. The selected model has the lowest RMSE value (bold text in Table 2), specifically the AR(13) model.

Table 2: Diagnostic Test.

Ordo	Dependency Test	Normality Test	RMSE
AR(1)	Not fulfilled	Not fulfilled	4.607
AR(7)	Not fulfilled	Not fulfilled	4.151
AR(10)	Not fulfilled	Not fulfilled	4.143
AR(13)	Not fulfilled	Not fulfilled	**4.138**

2. ARIMA Model with Outlier Factor

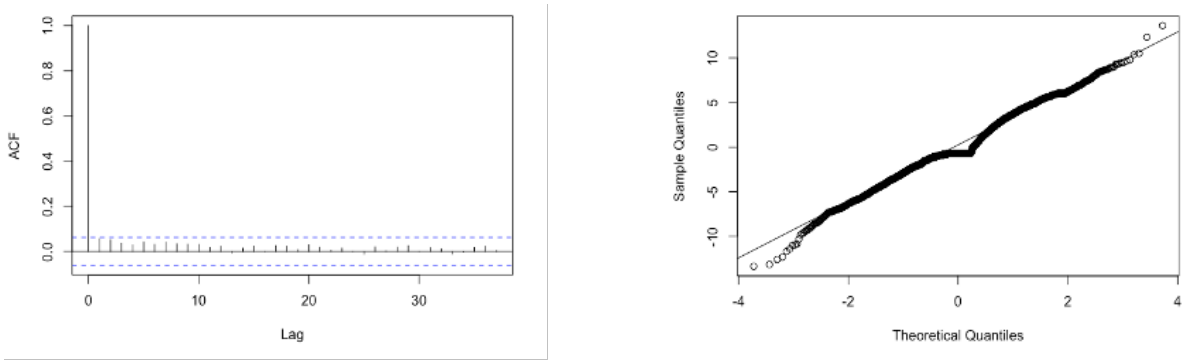
As seen in the flowchart in Figure 1, the subsequent step involves identifying outliers based on the residuals of the AR(13) model. Iteration 0 represents the AR(13) model. Iteration 1 results from outlier detection with the model derived from iteration 0, and so forth. Table 3 presents the outcomes of outlier detection throughout six iterations. No outliers were identified following the sixth round. Most outliers were identified in the initial round.

Table 3: Outlier's Time Detected. The blue text indicates the number of outliers detected in that iteration.

Itr.	Innovative Outlier	Additive Outlier
1	120; 121; 886; 888; 892; 1596; 2436; 2449; 2450; 2453; 2454; 2458; 2994; 3002; 3005; 3014; 3017; 3270; 3271; 3277; 3557; 3562; 3575; 3856; 3860; 3872; 3876; 3878; 4142; 4154; 4155; 4393; 4402; 4405; 4408; 4416; 4711; 4712 → (38)	122; 890; 891; 1599; 1601; 1864; 2026; 2114; 2452; 2463; 2465; 2985; 3016; 3021; 3268; 3269; 3276; 3567; 3570; 3578; 3582; 3863; 3867; 3881; 3883; 3896; 4131; 4156; 4160; 4407; 4417; 4422; 4428; 4717 → (34)
2	3006; 3272; 3558; 4406 → (4)	2460; 3015; 3564; 3576; 4413; 4426 → (6)
3	2459 → (1)	3009; 3563; 4713 → (3)
4	3275 → (1)	123 → (1)
5	3278 → (1)	→ (0)
6	→ (0)	3008 → (1)

Upon completing the iteration 1 process, proceed to estimate the parameters of the ARIMA model by incorporating the outlier component (refer to Equation 1). The resultant parameters are presented in Table 4. Upon the completion of parameter estimation in iteration 1, the outlier detection process for iteration 2 is conducted by examining the residuals of the ARIMA model, incorporating the outlier factor from iteration 1. The outlier detection time from the second iteration is subsequently incorporated into the ARIMA model along with the outlier factor derived from the first iteration. Additionally, parameter estimation for the ARIMA model is conducted by incorporating outlier components identified in iterations one and two. This procedure is reiterated until a model with residuals no longer identifiable as outliers is generated.

Upon obtaining the ARIMA model with the outlier component, a diagnostic test is subsequently performed to assess the influence of the outlier factor on the model. According to Table 4, outliers influenced model residuals that failed to meet the diagnostic criteria. The diagnostic test outcomes for the ARIMA model incorporating the outlier component are presented in Figure 4, indicating that the ARIMA model with the outlier factor has satisfied both diagnostic tests.



(a) (b)

Figure 4: Diagnostic Test (a) Dependency Residual and (b) Normality Residual for ARIMA Model with Outlier Factors

In addition to the diagnostic test standpoint, the existence of outliers also influences the model's accuracy. The ARIMA model yields an RMSE of 4,138, but the ARIMA model incorporating outlier components demonstrates enhanced accuracy with an RMSE of 3,259. A comparison of the observed data to the fitted values for the ARIMA model and the ARIMA model incorporating outlier effects illustrates the differences between the two models (refer to Figure 5).

Table 4: ARIMA, Innovative Outlier (IO), and Additive Outlier (AO) per Iteration

Itr.	ARIMA $(\mu; \phi_1 \dots \phi_{13})$	IO $(\omega_{T_{I1}} I_t^{(T_{I1})}, \dots)$	AO $(\omega_{T_{A1}} I_t^{(T_{A1})}, \dots)$
0	4.09; 0.26; 0.16; 0.13; 0.06; 0.07; 0.08; 0.04; -0.04; 0.03; 0.03; -0.01; 0.03; 0.03		
1	3.87; 0.23; 0.17; 0.12; 0.07; 0.05; 0.06; 0.03; -0.01; 0.01; 0.04; -0.02; 0.03; 0.01	24.18; 20.18; 24.12; 25.46; -20.15; 24.72; 23.59; 27.26; 25.27; 9.93; 10.62; -23.94; 23.05; 20.95; 19.05; 18.44; -7.80; 37.79; 33.82; -10.18; 21.19; 14.96; 19.51; 25.46; 22.62; 18.94; 20.27; 19.93; 26.59; 24.05; 23.04; 23.13; 22.22; 19.57; 13.24; -17.86; 22.09; 18.65	35.01; 18.67; 12.18; 22.72; 19.92; 24.19; 26.04; 23.41; -9.97; 23.09; 20.17; 25.15; 19.28; 19.89; -8.68; -8.64; 14.35; 16.55; 19.13; 16.88; 15.75; 19.32; 23.09; 17.54; 16.95; 23.28; 22.89; 19.22; 19.09; -14.17; -18.64; -18.06; 14.35; 17.29
2	3.84; 0.21; 0.17; 0.13; 0.05; 0.05; 0.05; 0.05; -0.01; 0.01; 0.04; -0.01; 0.03; 0.01	24.32; 20.05; 25.50; 25.25; -12.63; 21.70; 23.58; 26.27; 24.48; 11.71; 11.21; -17.32; 21.91; 20.64; 19.53; 18.85; -12.75; 41.61; 39.21; -14.39; 21.14; 18.01; 19.13; 25.22; 20.55; 18.67; 21.99; 20.49; 27.14; 22.27; 21.16; 23.07; 24.35; 17.99; 11.01; -17.28; 24.52; 20.14; 13.21; 11.84; 16.86; 12.83	35.13; 20.52; 13.59; 21.84; 19/26; 24.17; 25.51; 24.71; -9.85; 20.85; 19.71; 25.11; 15.89; 20.19; -8.69; -7.57; 13.88; 16.67; 19.54; 17.14; 15.34; 19.22; 23.19; 17.61; 16.66; 22.49; 23.04; 20.48; 19.69; -16.23; -13.99; -15.79; 15.55; 18.65; 10.52; 15.61; 15.03; 15.71; -14.02; 16.49
3	3.84; 0.20; 0.18; 0.13; 0.06; 0.05; 0.05; 0.05; 0.04; -0.05; 0.03; 0.04; -0.01; 0.02; 0.01	23.62; 21.46; 25.69; 24.76; -13.03; 21.52; 22.85; 24.01; 21.93; 12.48; 11.97; -23.07; 21.30; 23.07; 15.58; 18.05; -13.90; 41.21; 33.58; -13.95; 21.09; 20.84; 20.49; 25.44; 20.81; 19.54; 20.63; 21.28; 27.05; 23.45; 19.75; 24.49; 24.80; 19.37; 10.33; -18.53; 24.29; 21.54; 12.15; 8.37; 18.02; 10.58; -8.23	37.59; 18.73; 12.85; 20.87; 19.72; 23.83; 26.79; 24.59; -9.32; 21.27; 20.18; 24.94; 15.76; 19.41; -7.76; -9.37; 11.93; 18.36; 19.56; 15.50; 16.72; 19.14; 23.19; 17.94; 17.75; 23.30; 22.27; 17.26; 19.04; -13.95; -17.11; -15.83; 14.22; 19.30; 17.58; 14.35; 17.54; 14.79; -15.17; 15.84; 11.98; 20.10; 15.81
4	3.84; 0.20; 0.18; 0.13; 0.6; 0.05; 0.05; 0.02; -0.01; 0.02; 0.04; -0.01; 0.03; 0.01	25.75; 22.03; 24.27; 24.22; -17.14; 22.34; 23.48; 25.13; 21.75; 10.56; 14.25; -21.11; 22.37; 21.14; 18.35; 17.84; -12.20; 42.53; 32.75; -12.83; 21.69; 19.62; 20.47; 27.76; 22.69; 19.84; 21.05; 23.68; 26.57; 22.29; 18.33; 24.70; 23.36; 14.98; 8.08; -20.39; 23.00; 21.91; 11.75; 4.88; 19.46; 17.43; -14.82; -9.76	36.44; 19.26; 13.13; 21.32; 19.05; 24.14; 26.92; 24.63; -10.66; 21.57; 19.63; 24.65; 16.84; 19.77; -6.99; -9.20; 7.61; 17.97; 20.01; 16.68; 17.44; 17.64; 21.93; 17.85; 18.28; 24.24; 22.38; 17.22; 18.57; -16.33; -19.30; -16.87; 13.86; 19.99; 9.57; 14.84; 17.65; 17.13; -16.35; 14.56; 10.56; 18.60; 18.19; 14.82
5	3.84; 0.20; 0.19; 0.13; 0.06; 0.06; 0.05; 0.03; -0.01; 0.01; 0.03; -0.01; 0.02; 0.01	25.87; 22.19; 25.07; 24.01; -16.98; 22.34; 23.24; 25.06; 21.81; 10.87; 14.06; -21.13; 21.87; 20.21; 17.76; 17.78; -11.98; 42.53; 34.53; -16.12; 21.67; 19.61; 19.57; 27.38; 23.07; 20.37; 23.57; 20.47; 26.77; 22.27; 18.44; 24.21; 22.99; 15.34; 7.11; -20.56; 22.65; 20.10; 14.96; 5.55; 19.00; 17.86; -12.99; -12.34; -16.88	36.49; 19.32; 13.15; 21.69; 19.14; 24.87; 27.55; 24.64; -10.93; 21.35; 19.49; 24.68; 16.57; 19.62; -7.21; -8.72; 7.26; 17.67; 20.05; 15.82; 17.76; 17.54; 22.17; 17.17; 17.04; 24.16; 22.32; 16.79; 18.16; -16.69; -19.25; -16.89; 14.04; 20.31; 10.55; 15.59; 18.08; 15.19; -17.78; 15.36; 11.92; 17.87; 18.47; 14.85
6	3.84; 0.20; 0.18; 0.12; 0.06; 0.05; 0.05; 0.03; -0.01; 0.01; 0.03; 0.01; 0.02; 0.01	24.33; 19.65; 24.67; 25.90; -12.86; 22.21; 23.31; 25.61; 22.41; 15.37; 11.49; -22.87; 21.03; 22.18; 19.02; 18.49; -13.55; 44.54; 32.70; -8.70; 21.07; 20.09; 18.96; 27.37; 19.39; 20.51; 20.28; 19.91; 26.35; 22.22; 17.89; 23.67; 23.99; 15.88; 7.66; -18.27; 25.01; 22.27; 15.61; 9.85; 16.56; 13.74; -4.67; -13.49; -7.19	34.40; 19.12; 13.33; 21.61; 19.97; 24.26; 25.55; 24.45; -8.67; 22.58; 20.06; 25.09; 15.65; 19.34; -8.54; -8.92; 13.91; 17.94; 19.79; 16.51; 15.95; 19.09; 22.26; 17.44; 16.56; 23.20; 22.53; 16.89; 18.45; -14.69; -13.67; -15.82; 14.97; 19.40; 12.86; 14.56; 16.97; 14.74; -17.33; 14.64; 13.15; 19.84; 18.79; 13.82; 15.41

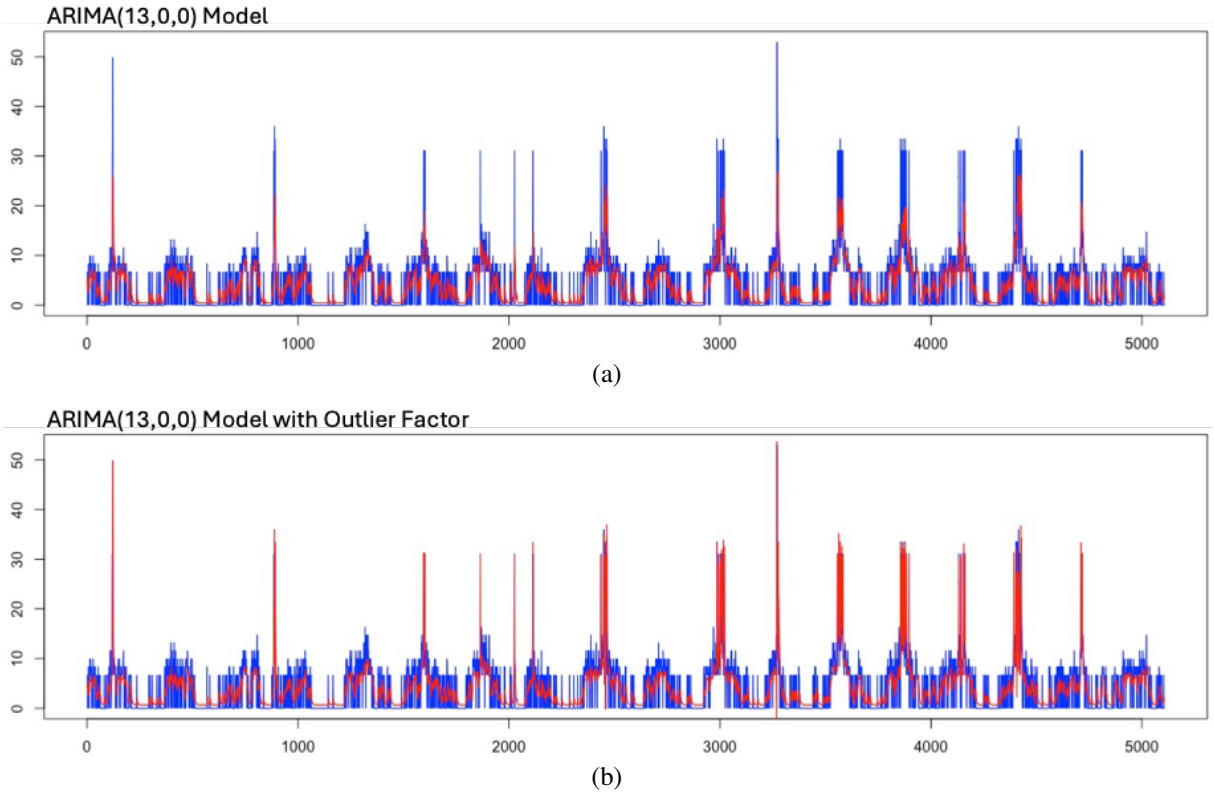


Figure 5: Wind Speed Data (Blue Line) and Fitted Values (Red Line) (a) ARIMA Model without Outlier Factor and (b) ARIMA Model with Outlier Factor

Figure 5 illustrates that the ARIMA model incorporating outlier effects significantly influences the ability to get extreme data points. Considering several factors, including diagnostic tests, model correctness, and the capacity to identify extreme values, the model incorporating the outlier component is highly effective for analysing wind speed data patterns. The model can be employed to forecast wind speed.

Discussion

The ARIMA model utilised in this study demonstrates significant efficacy in modelling high-resolution wind speed data patterns, particularly following an iterative outlier detection and adjustment process. This procedure involves detecting outliers from the original model residuals, examining the ACF and PACF patterns, and assessing the RMSE value for each iteration. The results indicate that the AR(13) model exhibits the optimal performance, evidenced by the lowest RMSE, suggesting that the wind speed at any point is significantly affected by the speed data from the preceding 13 periods, around 65 minutes prior (with a 5-minute interval every observation). This indicates that wind speed dynamics exhibit significant autoregressive behaviour beyond one hour, which is crucial for short-term forecasting. The use of this comparatively elevated AR order signifies that local atmospheric processes exhibit significant temporal dependency at high time resolution.

Besides their statistical efficacy, ARIMA models have demonstrated considerable reliability when applied to 5-minute data, capturing short-term dynamics more precisely than daily or hourly data. The model's responsiveness to minute-scale variations renders it appropriate for real-time monitoring or control systems, including wind turbine regulation and innovative grid applications. This study's minute-level forecasting methodology is a distinctive feature of the research focused on hourly or daily predictions. Minute-level forecasting offers substantial benefits in enhancing the accuracy of swift decision-making, particularly during abrupt weather fluctuations. The model employed for minute-level wind speed forecasting for the subsequent hour is the ARIMA(13,0,0) model, incorporating outlier compo-

nents (refer to Table 4, at the sixth iteration of the model). However, because

$$I_t^{(T_u)} = \begin{cases} 1 & t = T \\ 0 & t \neq T \end{cases}$$

Then, for forecasting, $I_t^{(T_u)}$ it will always be 0. Consequently, the model at the sixth iteration becomes

$$\begin{aligned} \hat{Y}_{5107+h} = 3.831 &+ 0.202Y_{5107+h-1} + 0.176Y_{5107+h-2} + 0.120Y_{5107+h-3} \\ &+ 0.057Y_{5107+h-4} + 0.052Y_{5107+h-5} + 0.049Y_{5107+h-6} + 0.033Y_{5107+h-7} \\ &- 0.005Y_{5107+h-8} + 0.011Y_{5107+h-9} + 0.025Y_{5107+h-10} + 0.006Y_{5107+h-11} \\ &+ 0.020Y_{5107+h-12} + 0.014Y_{5107+h-13} \end{aligned} \tag{2}$$

The prediction results for the next hour are given in Figure 6 below.

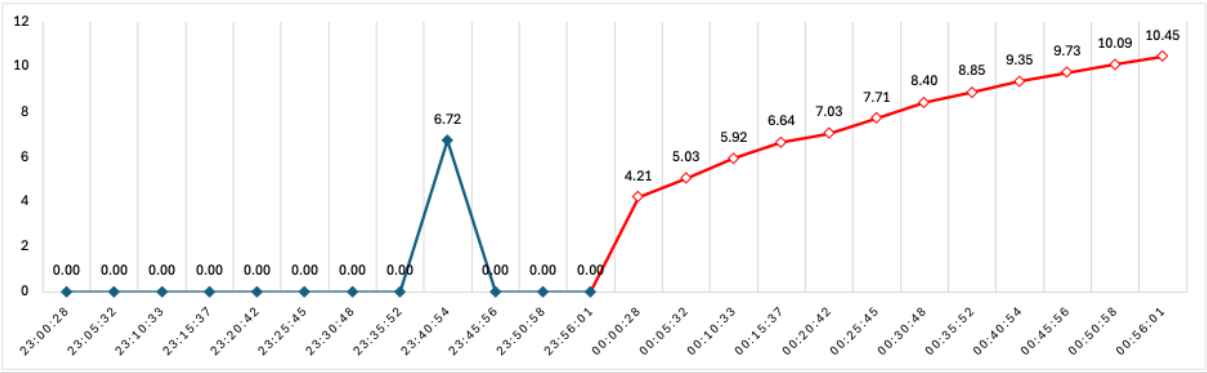


Figure 6: Minute-Level Prediction Result for 1 Hour (Red Line) based on The ARIMA Model With Outlier Factors

The efficacy of this concept is predominantly contingent upon the backing of an Internet of Things (IoT)-driven data gathering system, facilitating continuous, automated, and real-time documentation of wind speed. The Internet of Things (IoT) technology offers benefits related to temporal precision, seamless system integration, and enhanced operational efficiency. Incorporating minute-level data, iterative outlier detection and adjustment, ARIMA modelling, and IoT data collecting establishes a refined and practical framework for accurate short-term wind speed modelling.

4 Conclusions

The investigation indicates that ARIMA models incorporating outlier identification and adjustment, particularly Additive Outlier (AO) and Innovative Outlier (IO), enhance the model's accuracy in forecasting wind speed at a 5-minute interval. The AR(13) model was identified as the optimal choice using a six-stage iterative procedure, achieving the lowest RMSE value of 0.222. This signifies that short-term variations in minute-level data may be adequately accounted for by autoregressive dependence extending to 13 lags. The support of granular Iot-based data is crucial, as it facilitates the identification of micro-patterns and transient disturbances that may be missed at broader time resolutions. The amalgamation of ARIMA models and Iot technologies utilising minute-level data demonstrates benefits in accurate and prompt short-term forecasting, particularly pertinent for applications like real-time wind condition monitoring or early warning systems.

CRediT Authorship Contribution Statement

Author One: Conceptualization, Methodology, Formal Analysis, Writing–Original Draft, Visualization. **Author Two:** Data Curation, Writing–Review & Editing. **Author Three:** Software, Validation. **Author Four:** Project Administration, Validation

Declaration of Generative AI and AI-assisted technologies

No generative AI or AI-assisted technologies were used during the preparation of this manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request and subject to confidentiality agreements.

References

[1] J. K. B. Rogers, T. C. R. Mercado, and F. A. J. Galleto, "Comparison of arima boost, prophet boost, and tslm models in forecasting davao city weather data," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, pp. 1092–1101, 2024. DOI: 10.11591/ijeecs.v34.i2.pp1092-1101.

[2] H. Liu, J. Meng, W. Li, and S. Liu, "Extreme weather risk assessment and insurance decision making based on arima model and visual analysis," *Highlights in Science, Engineering and Technology*, vol. 101, pp. 305–313, 2024. DOI: 10.54097/93qwm898.

[3] I. Tyass, A. Bellat, A. Raihani, K. Mansouri, and T. Khalili, "Wind speed prediction based on seasonal arima model," in *E3S Web of Conferences*, vol. 336, 2022, p. 00 034. DOI: 10.1051/e3sconf/202233600034.

[4] Z. Zhou, Y. Dai, J. Xiao, M. Liu, J. Zhang, and M. Zhang, "Research on short-time wind speed prediction in mountainous areas based on improved arima model," *Sustainability*, vol. 14, no. 22, p. 15 301, 2022. DOI: 10.3390/su142215301.

[5] A. Demirtop and O. Sevli, "Wind speed prediction using lstm and arima time series analysis models: A case study of gelibolu," *Turkish Journal of Engineering*, vol. 8, no. 3, pp. 524–536, 2024. DOI: 10.31127/tuje.1431629.

[6] L. A.-j. Dawod, "Forecasting short-term wind speed using arima, and convolution neural network models at international airports in saudi arabia," *Communications on Applied Nonlinear Analysis*, vol. 31, no. 6s, pp. 81–96, 2024. DOI: 10.52783/cana.v31.1162.

[7] S. Sheoran and S. Pasari, "Efficacy and application of the window-sliding arima for daily and weekly wind speed forecasting," *Journal of Renewable and Sustainable Energy*, vol. 14, no. 5, 2022. DOI: 10.1063/5.0108847.

[8] M. Roem, M. Musa, Rudianto, and Y. Risjani, "The surface wind regimes on the northeast coastal of kalimantan during 2016-2018," in *IOP Conference Series: Earth and Environmental Science*, vol. 763, 2021, p. 012 013. DOI: 10.1088/1755-1315/763/1/012013.

[9] K. Furtak and A. Wolińska, "The impact of extreme weather events as a consequence of climate change on the soil moisture and on the quality of the soil environment and agriculture – a review," *CATENA*, vol. 231, p. 107 378, 2023. DOI: 10.1016/j.catena.2023.107378.

[10] R. Chikte, A. Mathew, B. Ramachandran, K. Almaeeni, N. Aljasmi, and J. Vandavasi, "Short-term time series energy forecasting for smart buildings using machine learning models," in *Volume 6: Energy*, Nov. 2024. DOI: 10.1115/IMECE2024-145349.

[11] H. Darmawan, M. Yuliana, and M. Z. Samsono Hadi, "Realtime weather prediction system using gru with daily surface observation data from iot sensors," in *2022 International Electronics Symposium (IES)*, 2022, pp. 221–226. DOI: 10.1109/IES55876.2022.9888468.

[12] U. Mukhaiyar, N. M. Huda, R. K. Novita Sari, and U. S. Pasaribu, "Modeling dengue fever cases by using gstar(1;1) model with outlier factor," in *Journal of Physics: Conference Series*, vol. 1366, 2019, p. 012 122. DOI: 10.1088/1742-6596/1366/1/012122.

[13] U. Mukhaiyar, N. M. Huda, K. N. Sari, and U. S. Pasaribu, "Analysis of generalized space time autoregressive with exogenous variable (gstarx) model with outlier factor," in *Journal of Physics: Conference Series*, vol. 1496, 2020, p. 012 004. DOI: 10.1088/1742-6596/1496/1/012004.

[14] T. Umairah, N. Imro'ah, and N. M. Huda, "Arima model verification with outlier factors using control chart," *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, vol. 18, no. 1, pp. 0579–0588, 2024. DOI: 10.30598/barekengvol18iss1pp0579-058.

[15] S. A. Panchal, L. Ferdouse, and A. Sultana, "Comparative analysis of arima and lstm models for stock price prediction," in *2024 IEEE/ACIS 27th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2024, pp. 240–244. DOI: 10.1109/SNPD61259.2024.10673919.

[16] S. O. Areh and N. H. Miswan, "ARIMA Models For Kijang Emas Price Forecasting: Pre- and Post-COVID Analysis," *Malaysian Journal of Mathematical Sciences*, vol. 18, no. 1, pp. 127–140, 2024. DOI: 10.47836/mjms.18.1.08.

[17] K. Holakouie-Naieni, M. Sepandi, B. Eshrati, S. Nematollahi, and Y. Alimohamadi, "Comparative performance of hybrid model based on discrete wavelet transform and arima models in prediction incidence of covid-19," *Heliyon*, vol. 10, no. 13, e33848, 2024. DOI: 10.1016/j.heliyon.2024.e33848.

[18] 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, p. 3069, 2023. DOI: 10.3390/math11143069.

[19] J. D. Cryer and K.-S. Chan, *Time Series Analysis*. Springer New York, 2008. DOI: 10.1007/978-0-387-75959-3.

[20] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, "Outlier detection for temporal data: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250–2267, 2014. DOI: 10.1109/TKDE.2013.184.

[21] N. M. Huda, U. Mukhaiyar, and N. Imro'ah, "An iterative procedure for outlier detection in gstar(1;1) model," *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, vol. 16, no. 3, pp. 975–984, 2022. DOI: 10.30598/barekengvol16iss3pp975-984.

[22] X. Du, E. Zuo, Z. Chu, Z. He, and J. Yu, "Fluctuation-based outlier detection," *Scientific Reports*, vol. 13, no. 1, p. 2408, 2023. DOI: 10.1038/s41598-023-29549-1.

[23] L. F. Maldaner, J. P. Molin, and M. Spekken, "Methodology to filter out outliers in high spatial density data to improve maps reliability," *Scientia Agricola*, vol. 79, no. 1, 2022. DOI: 10.1590/1678-992x-2020-0178.

[24] C. Wang, Z. Liu, H. Gao, and Y. Fu, "Vos: A new outlier detection model using virtual graph," *Knowledge-Based Systems*, vol. 185, p. 104 907, 2019. DOI: 10.1016/j.knosys.2019.104907.

[25] H. Wang, M. J. Bah, and M. Hammad, "Progress in outlier detection techniques: A survey," *IEEE Access*, vol. 7, pp. 107 964–108 000, 2019. DOI: 10.1109/ACCESS.2019.2932769.

[26] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd. Addison Wesley, 2006.

[27] N. M. Huda, U. Mukhaiyar, and U. S. Pasaribu, "Forecasting dengue fever cases using autoregressive distributed lag model with outlier factor," in *AIP Conference Proceedings*, 2020. DOI: 10.1063/5.0018450.

[28] S. Armiento, F. Meder, and B. Mazzolai, "Device for simultaneous wind and raindrop energy harvesting operating on the surface of plant leaves," *IEEE Robotics and Automation Letters*, vol. 8, no. 4, pp. 2269–2276, 2023. DOI: 10.1109/LRA.2023.3250006.

[29] J. S. Sarjerao, G. Sudhagar, and R. N. Jadhav, "Sustainable smart irrigation in precision farming using iot," in *2025 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2025, pp. 1–7. DOI: 10.1109/ESCI63694.2025.10987956.

[30] D. Kim and T.-Y. Heo, "Anomaly detection with feature extraction based on machine learning using hydraulic system iot sensor data," *Sensors*, vol. 22, no. 7, p. 2479, 2022. DOI: 10.3390/s22072479.