MethodsX 14 (2025) 103080

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Navigating Samarinda's climate: A comparative analysis of rainfall forecasting models[☆,☆☆]

Mislan[a,*], Andrea Tri Rian Dani[b]

[a] Department of Physics, Faculty of Mathematics and Natural Science, Mulawarman University
[b] Department of Mathematics, Faculty of Mathematics and Natural Science, Mulawarman University

ARTICLE INFO

Method name:
Traditional and Machine Learning Models in Forecasting: Exponential Smoothing, ARIMA, NN

Keywords:
Exponential smoothing
ARIMA
Neural network
Time series modeling
Forecasting

ABSTRACT

Modeling rainfall data is critical as one of the steps to mitigate natural disasters due to weather changes. This research compares the goodness of traditional and machine learning models for predicting rainfall in Samarinda City. Monthly rainfall data was recapitulated by the Meteorology, Climatology, and Geophysics Agency from 2000 to 2020. The traditional models used are Exponential Smoothing and ARIMA, while the machine learning model is a Neural Network. Data is divided into training and testing with a proportion of 90:10. Evaluation of goodness-of-fit using Root Mean Squared Error Prediction (RMSEP). The research results show that the Neural Network has better accuracy in predicting rainfall in Samarinda. Forecasting results indicate that monthly rainfall trends suggest that the months with the highest rainfall occur around November to March. This research provides important implications for developing a warning system for hydrometeorological disasters in Samarinda. The superior points in this research are:

- Modeling rainfall data in Samarinda City using several forecasting methods: Exponential Smoothing, ARIMA, and Neural Network.
- The Neural-Network algorithm used is Backpropagation with data standardization.
- Information about predicted high rainfall can be used to issue early warnings of floods or landslides. Disaster mitigation through policies to regulate water discharge based on rainfall predictions to prevent floods and droughts.

Specifications table
This table provides general information on your method.

Subject area:	Environmental Science
More specific subject area:	Climatology; Hydrology; Statistics Modeling; Forecasting
Name of your method:	Traditional and Machine Learning Models in Forecasting: Exponential Smoothing, ARIMA, NN
Name and reference of original method:	R. S. Pontoh, T. Toharudin, B. N. Ruchjana, N. Sijabat, and M. D. Puspita, "Bandung Rainfall Forecast and Its Relationship with Niño 3.4 Using Nonlinear Autoregressive Exogenous Neural Network," Atmosphere (Basel), vol. 13, no. 2, Feb. 2022, doi: 10.3390/atmos13020302.

(continued on next page)

☆ **Related research article:** None.
☆☆ **For a published article:** None.
* Corresponding author.
E-mail address: mislan.jos@fmipa.unmul.ac.id (Mislan).

https://doi.org/10.1016/j.mex.2024.103080
Received 10 October 2024; Accepted 30 November 2024
Available online 4 December 2024

Resource availability	N. H. A. Rahman, M. H. Lee, Suhartono, and M. T. Latif, "Artificial neural networks and fuzzy time series forecasting: an application to air quality," Qual Quant, vol. 49, no. 6, pp. 2633–2647, Nov. 2015, doi:10.1007/s11135-014-0132-6. The data used is the rainfall data of Samarinda City from 2000 to 2020, monthly. The data collection technique used is secondary data collection, which is obtained directly from related agencies, in this case the Meteorology, Climatology and Geophysics Agency (BMKG) of Samarinda City.

Background

Rainfall is the height of rainwater collected in a flat place in a certain period, usually measured in millimeters (mm) per unit of time (BMKG) [1]. Rainfall is a natural phenomenon that plays a vital role in various aspects of life, including the agricultural sector and water resources, and can also be information for natural disaster mitigation. Rainfall is one of the most essential elements in climate patterns [2]. An accurate understanding and prediction of rainfall is needed in policy-making and early warning systems. Rainfall prediction can use time series models [3,4]. The time series model is a mathematical representation of data collected sequentially over time [5,6]. With the advancement of information technology today, the development of time series models is massive in obtaining the best accuracy, from traditional to machine learning models [7]. Researchers will use conventional and machine-learning models to model rainfall data in this study [8,9]. The forecasting models that will be used in this study are Exponential Smoothing (ES), Autoregressive Integrated Moving Average (ARIMA), and neural network (NN).

Exponential Smoothing (ES) is one of the simple smoothing methods, but it has a pretty good performance and can be used to forecast future time series [10]. The working principle of ES is to provide further weight to the latest observation time series data compared to older observation time series data. The advantage of the ES method is that it is simple and easy to implement in its application [11]. Several time series data studies that use ES include [illegible]. Autoregressive Integrated Moving Average (ARIMA) is a time series model with solid assumptions that require stationary data, so it is necessary to transform the data [13,14]. In addition, the residuals of the ARIMA model must be White Noise and Normally Distributed. Several studies of time series data using ARIMA include [15–19].

Neural Network (NN), a time series model inspired by Artificial Neural Networks, is known for its adaptability to data change patterns [8]. It adjusts the weight of connections between neurons based on the difference between the actual output and the output to be predicted, a process done iteratively [20]. This adaptability allows NN to identify complex data patterns that traditional models miss. Several time series data studies have successfully utilized NN are [16,21–25].

The primary goal of this study is to forecast rainfall data for the next 12 periods using the best time series model. This model, once identified, can serve as a valuable tool for obtaining future insights. Its potential benefits extend beyond the academic realm, as it can help the general public mitigate the negative impacts of extreme weather, making it a crucial step in disaster management.

Method details

Exponential smoothing

In the world of forecasting, the exponential smoothing method is divided into three parts, namely Single Exponential Smoothing, which is a development of the Single Moving Average; Double Exponential Smoothing, which is a development of the Double Moving Average method, and Triple Exponential Smoothing which is a method used to analyze data that has a trend or seasonal pattern. One of the Double Exponential Smoothing methods that is often used in forecasting is Double Exponential Smoothing Holt [10,11]. Double Exponential Smoothing (DES) Holt is an exponential smoothing method with two parameters, and its analysis uses trends and actual data patterns. DES Holt forecast uses the following formula in Eq. (1)–Eq. (3).

Level smoothing

$$L_t = \alpha Z_t + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{1}$$

Trend smoothing

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{2}$$

With

$$F_{t+m} = L_t + T_t m \tag{3}$$

The Holt DES method estimates two smoothing values, which can be done using the following Eq. (4).

$$L_1 = Z_1 \text{ and } T_1 = Z_2 - Z_1 \tag{4}$$

Where:

α	level smoothing parameter, $0 < \alpha < 1$
β	trend smoothing parameter, $0 < \alpha < 1$
Z_t	actual data at time t
L_t	level smoothing at time t

T_t trend smoothing at time t
F_{t+m} forecasting at time (t+m)

ARIMA

The ARIMA model was introduced in 1970 by George EP Box and Gwilym M. Jenkins through their book entitled Time Series Analysis [5,26]. ARIMA is also often called the Box-Jenkins time series method. ARIMA is very accurate for both short-term and long-term forecasting. ARIMA can be interpreted as combining two models, namely the Autoregressive (AR) model integrated with the Moving Average (MA) model [27]. The ARIMA model is generally written with the notation ARIMA (p,d,q) where p is the degree of the AR process, d is the differencing order, and I is the degree of the MA process.

According to Box and Jenkins, the ARIMA (p,d,q) can be expressed in Eq. (5).

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \tag{5}$$

With:

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$	backshift operator(B) AR process
$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$	backshift operator(B) MA process
B	backshift operator
$(1-B)^d$	differentiating operator
d	order of differencing

Eq. (5) can be expressed in another form, namely:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)Z_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)a_t \tag{6}$$

The ARIMA (p,d,q) model is a combination of the AR (p) and MA (q) models with non-stationary data patterns, then differencing is performed with order d. Several time series models for stationary data are as follows:

Autoregressive (AR) Model

Autoregressive is a form of regression but not one that connects dependent variables, but rather connects them with previous values at a time lag, so that an autoregressive model will state a forecast as a function of previous values of the time series data. The autoregressive model with the order AR (p) or ARIMA model (p,0,0) is stated as follows in Eq. (7).

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t \tag{7}$$

Eq. (7) can be written using the backshift operator (B) as:

$$\phi_p(B)Z_t = a_t \tag{8}$$

With $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is called $AR(p)$ operator.

Moving Average (MA) Model

Another model of the ARIMA model is the moving average which is denoted as MA (q) or ARIMA (0,0,q) which is written in Eq. (9).

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \tag{9}$$

Eq. (9) can be written using the backshift operator (B), as:

$$Z_t = \theta_q(B)a_t \tag{10}$$

With $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ is called $MA(q)$ operator.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) that have been calculated are then used to identify the ARIMA model [15,28]. The identification stage is a stage used to find or determine other orders of p and q with the help of the autocorrelation function (ACF) and partial autocorrelation function (PACF) as follows:

Neural network

Neural Network (NN) is an information processing method that imitates how the human brain works [29]. NN has several simple processing units that are interconnected and work in parallel to complete complex tasks. The learning process in NN is carried out by adjusting the weight of the synapses that connect between units so that they can generate patterns in data and make predictions [30,31]. NN consists of neurons that have information flow. The NN structure consists of three layers of neural units, namely the input layer, the hidden layer, and the output layer [32]. As an illustration, it can be seen in Fig. 1. (Table 1).

Backpropagation is a core algorithm in NN learning that works by adjusting the connection weights between neurons to minimize prediction errors [33]. This process allows NN to learn complex patterns in data. The activation function, an essential component in neurons, plays a role in determining whether a neuron will be active. A good activation function must have continuous, differentiable, and non-monotonic properties for the gradient calculation during the backpropagation process. The derivative of this activation function is crucial in measuring how much each neuron contributes to the total error, allowing for more precise weight adjustments [34].

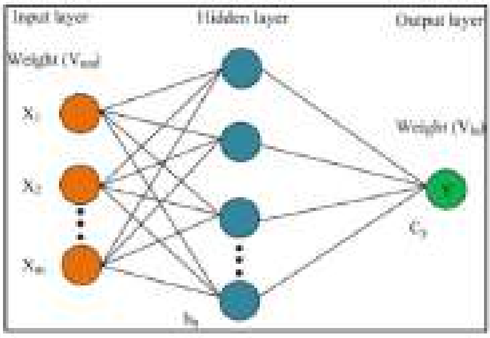


Fig. 1. Neural network structure.

Table 1
General ACF and PACF Patterns for AR and MA Models.

Process	ACF	PACF
AR (p)	Dies down (rapidly decreasing exponentially/sinusoidal)	Cuts off after lag p
MA (q)	Cuts off after lag q	Dies down (rapidly decreasing exponentially/sinusoidal)
ARMA (p,q)	Dies down (rapidly decreasing exponentially/sinusoidally)	Dies down (rapidly decreasing exponentially/sinusoidally)
AR (p) or MA (q)	Cuts off after lag q	Cuts off after lag p
White Noise (Random)	Nothing is out of bounds	Nothing is out of bounds

The activation function used in this study is the bipolar sigmoid function. The bipolar sigmoid activation function has a value range of -1 to 1 with the formula in Eq. (11).

$$f_1(z) = \frac{2}{1+e^{-z}} - 1 \tag{11}$$

With the derivative of Eq. (11) shown in Eq. (12).

$$f_1'(z) = \frac{1}{2}\left[1 + f_1(z)\right]\left[1 - f_1(z)\right] \tag{12}$$

Root mean square error prediction

In this study, to find the forecast accuracy value, the Root Mean Square Error Prediction (RMSEP) method is used. RMSEP can be interpreted as a measure of error based on the difference between two-value, actual and prediction. The RMSEP formula shown in Eq. (13).

$$RMSEP = \frac{1}{n}\sqrt{\sum_{t=1}^{n}\left(Z_t - \hat{Z}_t\right)} \tag{13}$$

Data and data sources

The data used is the rainfall data of Samarinda City from 2000 to 2020, monthly. The data collection technique used is secondary data collection, which is obtained directly from related agencies, in this case the Meteorology Climatology and Geophysics Agency (BMKG) of Samarinda City. Time series plot of the rainfall data in Samarinda for 2000 – 2020 can be seen in Fig. 2.

Based on Fig. 2, there is a significant fluctuation in rainfall in Samarinda in the period from January 2000 to December 2020. This indicates that rainfall in Samarinda has experienced quite significant changes over time during this period. This fluctuation can be caused by various factors, such as global climate change, human activities, and other natural phenomena.

Method validation

Modeling with double exponential smoothing

Double Exponential Smoothing (DES) Holt is an exponential smoothing method that has two parameters, namely α and β. In this study, the data was divided into training data and testing data with a division of 90:10. The first step that must be taken is to find

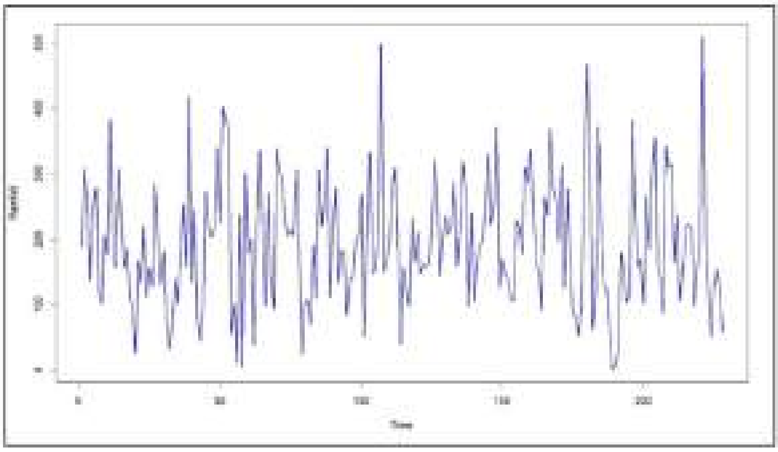


Fig. 2. Time series plot of rainfall data in Samarinda.

Table 2
Combination α and β Optimal.

Alpha (α)	Beta (β)	RMSEP	Alpha (α)	Beta (β)	RMSEP	Alpha (α)	Beta (β)	RMSEP
0.1	0.1	209.57	0.4	0.1	113.28	0.7	0.1	114.90
0.1	0.2	168.98	0.4	0.2	114.73	0.7	0.2	117.53
0.1	0.3	151.35	0.4	0.3	117.14	0.7	0.3	121.28
0.1	0.4	142.15	0.4	0.4	120.66	0.7	0.4	125.22
0.1	0.5	138.57	0.4	0.5	122.79	0.7	0.5	129.28
0.1	0.6	128.99	0.4	0.6	125.13	0.7	0.6	133.40
0.1	0.7	149.94	0.4	0.7	127.14	0.7	0.7	137.81
0.1	0.8	143.61	0.4	0.8	130.14	0.7	0.8	142.23
0.1	0.9	146.94	0.4	0.9	131.40	0.7	0.9	146.97
0.2	0.1	138.03	0.5	0.1	113.28	0.8	0.1	116.90
0.2	0.2	125.13	0.5	0.2	114.43	0.8	0.2	120.51
0.2	0.3	123.20	0.5	0.3	117.25	0.8	0.3	124.84
0.2	0.4	124.09	0.5	0.4	120.25	0.8	0.4	129.38
0.2	0.5	127.24	0.5	0.5	123.15	0.8	0.5	134.08
0.2	0.6	133.18	0.5	0.6	126.03	0.8	0.6	138.95
0.2	0.7	133.68	0.5	0.7	129.02	0.8	0.7	142.98
0.2	0.8	138.01	0.5	0.8	132.29	0.8	0.8	149.15
0.2	0.9	142.86	0.5	0.9	135.84	0.8	0.9	154.50
0.3	0.1	121.04	0.6	0.1	113.40	0.9	0.1	120.85
0.3	0.2	117.09	0.6	0.2	115.47	0.9	0.2	124.44
0.3	0.3	118.59	0.6	0.3	118.72	0.9	0.3	129.17
0.3	0.4	121.41	0.6	0.4	122.12	0.9	0.4	134.78
0.3	0.5	124.67	0.6	0.5	125.58	0.9	0.5	140.33
0.3	0.6	127.90	0.6	0.6	129.19	0.9	0.6	146.10
0.3	0.7	130.56	0.6	0.7	132.97	0.9	0.7	152.29
0.3	0.8	132.27	0.6	0.8	136.98	0.9	0.8	158.77
0.3	0.9	133.12	0.6	0.9	140.91	0.9	0.9	165.67

the combination value for α and β optimal by looking at the Root Mean Square Error Prediction (RMSEP) value on the training data, where the smaller the RMSEP value, the better the model's ability to predict accurately. The following is a table of combination results:

Based on Table 2, it can be seen that there are 4 combinations that have the smallest RMSE values, namely the following combinations: (Table 3).

Modeling with ARIMA

The rainfall data to be modeled with ARIMA is first transformed and differencing to make it stationary in mean and variance. The ACF and PACF plots of the transformed and differencing data of order 1 can be seen in Figs. 3 and 4.

Table 3
Optimal combination value of training and testing data.

Parameter Value	RMSEP Training	RMSEP Testing
α = 0.4 and β = 0.2	114.72	122.92
α = 0.5 and β = 0.1	113.38	127.50
α = 0.6 and β = 0.1	113.48	121.19
α = 0.7 and β = 0.1	114.64	118.03

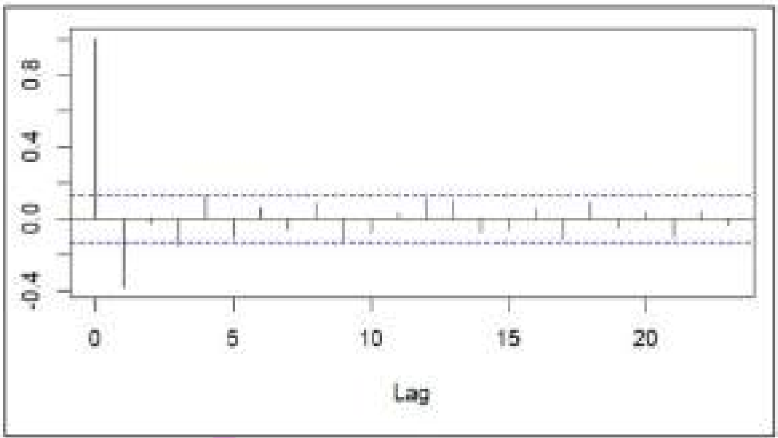


Fig. 3. ACF plot of rainfall data results of differencing.

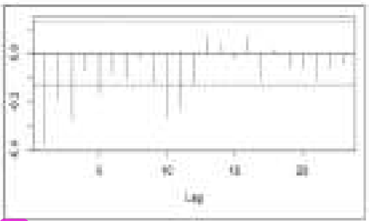


Fig. 4. PACF plot of rainfall data results of differencing.

Based on the ACF plot, it can be seen that the cut-off after lag 1, so the q order used is 0 and 1. Meanwhile, based on the PACF plot, it can be seen that there is a cut-off after lag 3 so that the p order used is 0, 1, 2, and 3. So that the temporary ARIMA models that can be formed are ARIMA (1,1,0), ARIMA (2,1,0), ARIMA (3,1,0), ARIMA (0,1,1), ARIMA (1,1,1), ARIMA (2,1,1), and ARIMA (3,1,1).

The temporary ARIMA models that have significant parameters are ARIMA (1,1,0), ARIMA (2,1,0), ARIMA (3,1,0), ARIMA (0,1,1) and ARIMA (1,1,1). The following figures are visualizations of the residual independence and residual normality assumptions of the temporary ARIMA models that are formed (Table 4).

Based on the figures above, it can be seen that all temporary ARIMA models meet the residual normality assumption because they form a bell curve, meaning that the models have normally distributed residuals. However, all of these models can be indicated that there is autocorrelation or violation of the residual independence assumption because there are several lags that are outside the upper and lower limits of the ACF plot. The selection of the best temporary ARIMA model can be determined by looking at the smallest RMSEP value in the transformed data. Thus, the ARIMA model (1,1,1) is the best temporary ARIMA model (Fig. 5).

Table 4
Temporary ARIMA model parameter estimation.

Model	Parameter	Estimate	p-value	Conclusion
ARIMA(1,1,0)	ϕ_1	-0.390766	5.579e-10	Significant
ARIMA(2,1,0)	ϕ_1	-0.455156	2.980e-12	Significant
	ϕ_2	-0.190967	0.002967	
ARIMA(3,1,0)	ϕ_1	-0.507994	1.760e-15	Significant
	ϕ_2	-0.320263	3.421e-06	
	ϕ_3	-0.275763	1.505e-05	
ARIMA(0,1,1)	θ_1	-1,000000	< 2.2e-16	Significant
ARIMA(1,1,1)	ϕ_1	0.273969	2.071e-05	Significant
	θ_1	-1,000000	< 2.2e-16	
ARIMA(2,1,1)	ϕ_1	0.266593	6.151e-05	Not Significant
	ϕ_2	0.020447	0.6701	
	θ_1	-1,000000	< 2.2e-16	
ARIMA(3,1,1)	ϕ_1	0.260049	5.275e-05	Not Significant
	ϕ_2	0.040998	0.475	
	ϕ_3	-0.081457	0.220	
	θ_1	-1,000000	< 2.2e-16	

Table 5
Accuracy of Training and Testing Data for ARIMA Model (1,1,1).

Proportion	RMSEP
Data Training	92.79
Data Testing	90.82

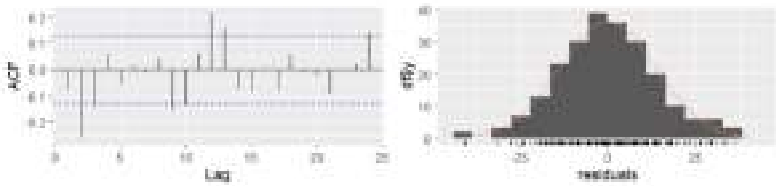


Fig. 5. Independence and Normality of ARIMA Residuals (1,1,0).

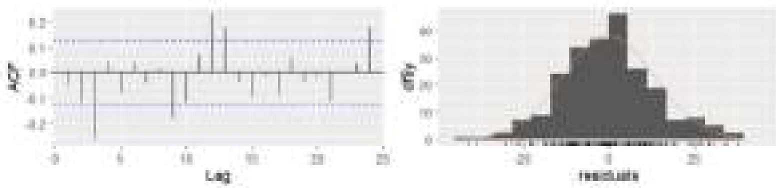


Fig. 6. Independence and Normality of ARIMA Residuals (2,1,0).

Modeling with neural network (NN)

- Determination of Input Variables

The determination of network input variables is done based on significant lags on the ACF graph or PACF graph. The ACF graph and PACF graph of rainfall data in Samarinda can be seen in Fig. 10 (Table 5).

Based on Fig. 10, it can be seen that there are several significant lags. In the ACF graph, there are significant lags at lag 1, lag 9, lag 12, and lag 13, while in the PACF graph, significant lags are at lag 1, lag 9, lag 11, and lag 12. This indicates a dependency between the value of an observation and the value of the previous observation up to 12 or 13 time periods. Therefore, this study uses 6 time lags as input variables, namely lag 1, lag 2, lag 3, lag 9, lag 11, and lag 12 (Fig. 6).

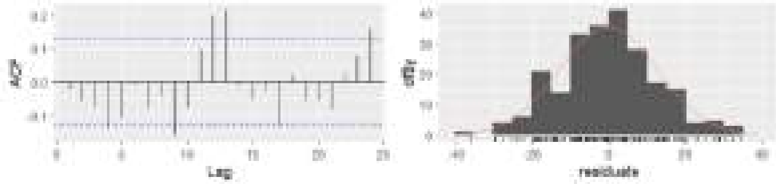


Fig. 7. Independence and Normality of ARIMA Residuals (3,1,0).

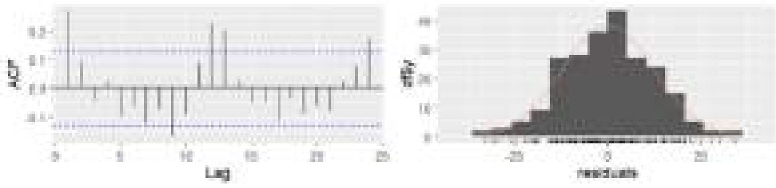


Fig. 8. Independence and Normality of ARIMA Residuals (0,1,1).

- Data Standarization

Standardization of research data is done to change the range of data values into a more uniform scale, thus facilitating comparison and analysis. In this study, the z-score standardization method is used to change the data into a standard score with an average of 0 and a standard deviation of 1 (Fig. 7).

- Best Model Selection in NN

The backpropagation training process is carried out by adjusting the NN architecture. In this study, two types of architectures are used, namely networks with one hidden layer and networks with two hidden layers where each architecture will try various combinations of the number of neurons in each layer. The criterion for stopping training is when it reaches a maximum iteration of 50,000,000 using a learning rate of 0.001. This training aims to minimize the error value and obtain a model with good generalization. After getting the results of the NN architecture, the next step is to perform a back transformation or destandardization. This process aims to change the predicted values that have been normalized back to their original scale, so that the predicted results can be interpreted in the context of the original data and can compare the predicted values with the actual values of the data (Fig. 8).

In the NN architecture in this study, researchers tried to use a maximum of two hidden layers in the NN compartment, where a combination of the number of neurons from 1 to 10 will be carried out. This combination obtains the optimal number of neurons in the first and second hidden layers. Some architectures of combinations of neurons in each hidden layer are intended to perform hyperparameter tuning, limiting the learning method and the number of hidden layers and choosing the activation function used. Evaluation using RMSEP on training and testing data. Based on the calculation results, the RMSEP value of each model is obtained which can be seen in Table 6 and Fig. 9.

By considering various considerations such as choosing the smallest RMSEP value and the difference between the RMSEP values of the training data and testing data is not very significant, then based on Table 6, it is found that the NN model with a 2 hidden layer architecture model (6–5 neurons) is the best NN model to be used in predicting rainfall in Samarinda for the next 12 periods. In this model, the RMSEP of the training data is 52.473 and the RMSE of the testing data is 69.749. Some of the architectural results of the NN modeling can be seen in Fig. 11.

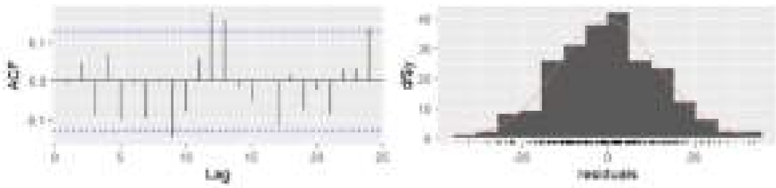


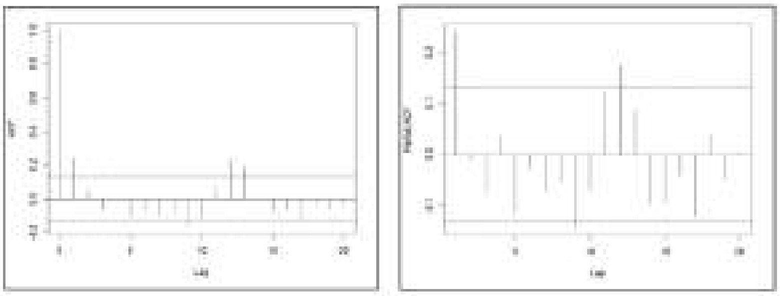
Fig. 9. Independence and Normality of ARIMA Residuals (1,1,1).

Table 6
RMSE calculation results.

Hidden Layer		RMSEP		Difference between Training and Testing RMSEP
Hidden Layer1	Hidden Layer2	Training	Testing	
1 neuron		86.292	88.595	2.303
2 neurons		85.018	90.397	5.380
3 neurons		78.101	110.632	32.531
4 neurons		75.907	100.158	24.251
5 neurons		66.349	97.531	31.182
6 neurons		68.059	96.447	28.388
7 neurons		69.542	103.764	34.222
8 neurons		58.299	102.908	44.609
9 neurons		57.471	114.915	57.444
10 neurons		46.560	110.919	64.359
2 neurons	1 neuron	83.746	73.969	9.777
3 neurons	2 neurons	77.546	100.936	23.390
4 neurons	3 neurons	70.852	97.384	26.532
5 neurons	4 neurons	61.167	131.724	70.557
6 neurons	**5 neurons**	**52.473**	**89.749**	**37.276**
7 neurons	6 neurons	49.585	121.518	71.933
8 neurons	7 neurons	29.696	135.890	106.194
9 neurons	8 neurons	27.541	116.387	88.846
10 neurons	9 neurons	17.895	181.869	163.974

Table 7
Model goodness of fit measures.

Method	RMSEP	
	Training	Testing
DES Holt ($\alpha = 0.7$ and $\beta = 0.1$)	104.66	118.03
ARIMA (1,1,1)	82.75	80.82
NN 2 HL (6–5 neurons)	**52.473**	**89.749**



(a) ACF graph (b) PACF graph

Fig. 10. ACF and PACF graphs of rainfall data in Samarinda.

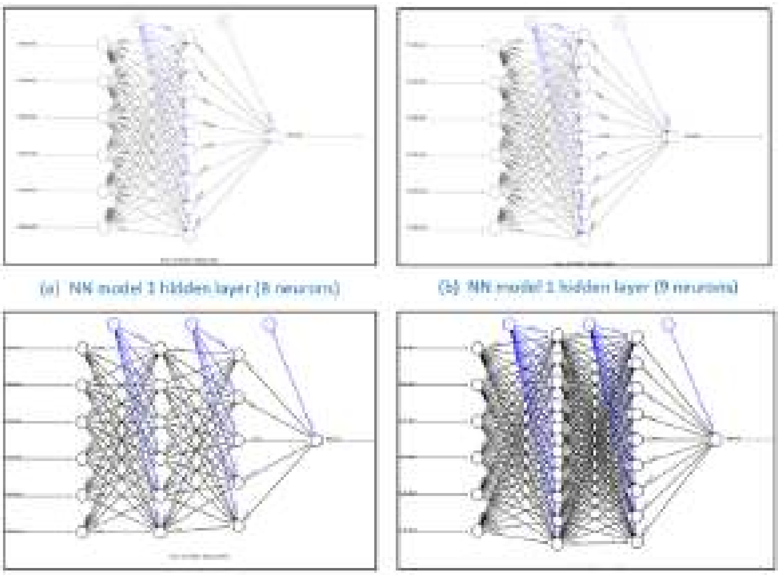
Best model selection

Based on the results of rainfall data modeling using DES Holt, ARIMA, and NN above, the next step is to determine the best model that can be used for forecasting. Table 7 displays the RMSEP values of the best models. The model with the smallest RMSE value will be selected as the best method.

In Table 7, it can be seen that the NN 2 HL (6–5 neurons) has a smaller RMSEP value for training data compared to other models, so the NN 2 HL (6–5 neurons) model will be used for forecasting the next 12 periods of rainfall data.

Forecasting and Discussion

Forecasting rainfall data for the next 12 periods using the NN 2 HL (6–5 neurons) model can be seen in Table 8.



(a) NN model 1 hidden layer (8 neurons) (b) NN model 1 hidden layer (9 neurons)

(c) NN model 2 hidden layers (6-5 neurons) (d) NN model 2 hidden layers (10-9 neurons)

Fig. 11. Some architectures of NN modeling.

Table 8
Forecasting results of 12 Periods of NN 2 HL (6–5 neurons).

Month	Prediction Results
January	301.959
February	245.819
March	269.822
April	82.904
May	173.226
June	209.721
July	207.147
August	204.006
September	201.438
October	133.406
November	217.897
December	306.919

Fig. 12 shows that in the time series graph for training and testing data, the predicted values almost follow the actual data pattern with a forecast accuracy level using RMSEP for training data of 52.473. Forecasting results for the following 12 periods show fluctuations in specific periods. Monthly rainfall trends indicate that the months with the highest rainfall occur around November to March. Based on the prediction results, it is known that the month with the highest rainfall is January. Rainfall patterns also tend to be seasonal, with peak rainfall at the beginning of the year and decreasing drastically in the middle of the year. The results of high rainfall predictions in certain months can undoubtedly be information and knowledge that brings several practical implications that need to be considered by various parties, including the government, society, and the private sector. For example, for the government, it can be an early warning system in facing the rainy season with high intensity, including in previous periods, by repairing drainage channels, building dams, and normalizing rivers. The government can also manage water resources through dams and irrigation (Table 8).

Samarinda, as one of the cities supporting the archipelago's capital, certainly faces challenges due to significant fluctuations in rainfall. It is hoped that the results of this prediction can become a mitigation strategy for the City of Samarinda in spatial management,

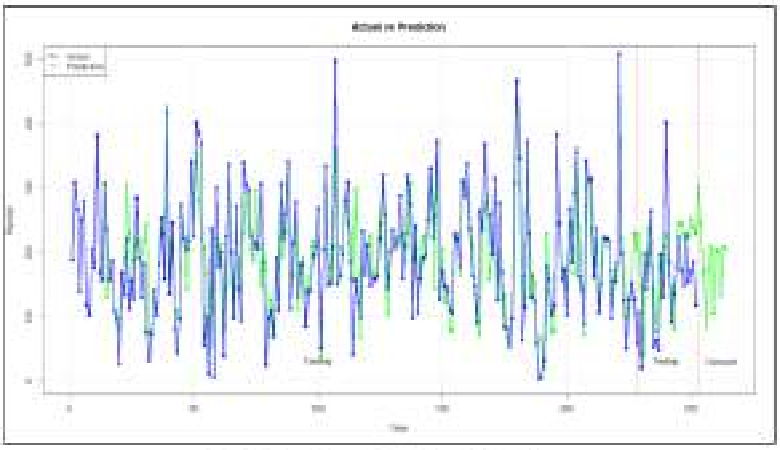


Fig. 12. Comparison plot of actual and predicted data.

an early warning system through weather monitoring, which monitors weather conditions in real time. Water resource management is critical to collect and absorb rainwater into the soil. Some steps that can be taken include:

- Adequate drainage system: The Samarinda City Government collaborates with related parties to evaluate and design an effective drainage system to drain rainwater smoothly and reduce puddles.
- Mapping flood-prone zones: The Samarinda City Government can map areas that are potentially flooded and establish appropriate regulations.
- Water resource management: The Samarinda City Government can build absorption wells and biopores to help absorb rainwater into the ground, reducing the risk of flooding. Rehabilitation of river basins through reforestation can increase water-holding capacity and reduce sedimentation.

Limitations

Limitations of this study include the potential for model overfitting, reliance on historical data that may not account for future climate change, and limited generalizability of the findings to other regions.

Ethics statements

The data used is the rainfall data of Samarinda City from 2000 to 2020, monthly. The data collection technique used is secondary data collection, which is obtained directly from related agencies, in this case the Meteorology, Climatology and Geophysics Agency (BMKG) of Samarinda City.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

None.

References

[1] M. Putra, M.S. Rosid, D. Handoko, A review of rainfall estimation in Indonesia: data sources, techniques, and methods, Signals (Basel) 5 (3) (2024) 542–561 Aug. doi:10.3390/signals5030030.

[2] R.S. Pontoh, T. Toharudin, B.N. Ruchjana, N. Sijabat, M.D. Puspita, Bandung rainfall forecast and its relationship with Niño 3.4 using nonlinear autoregressive exogenous neural network, Atmosphere (Basel) 13 (2) (2022) Feb. doi:10.3390/atmos13020302.

[3] R.F. Permata, R. Ni'mah, A.T.R. Dani, Daily rainfall forecasting with ARIMA exogenous variables and support vector regression, Jurnal Varian 7 (2) (2024) 177–188 Jun. doi:10.30812/varian.v7i2.3202.

[4] M. Mislan, A.T.R. Dani, Forecasting maximum water level data for post sungkuliman using an artificial neural network backpropagation algorithm, JTAM (Jurnal Teori dan Aplikasi Matematika) 8 (2) (2024) 465 May, doi:10.31764/jtam.v8i2.20132.

[5] G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, California, 1976.

[6] S. Makridakis, R.J. Hyndman, F. Petropoulos, Forecasting in social settings: the state of the art, Int. J. Forecast. 36 (1) (2020) 15–28 Jan. doi:10.1016/j.ijforecast.2019.05.011.

[7] S. Makridakis, E. Spiliotis, V. Assimakopoulos, V. Assimakopoulos, The Accuracy of Machine Learning (ML) Forecasting Methods Versus Statistical Ones: Extending the Results of the M3-Competition, 2017 [Online]. Available: https://www.researchgate.net/publication/320298859.

[8] A.F. Atiya, S.M. El-Shoura, S.I. Shaheen, M.S. El-Sherif, A Comparison Between Neural-Network Forecasting Techniques-Case Study: River Flow Forecasting, 1999.

[9] M. Narvekar, P. Fargose, Daily Weather Forecasting Using Artificial Neural Network, 2015.

[10] G.Puspita Mustarofilla, R.S. Zaidan, T. Wagyoko, The exponential smoothing model for squid production in Muara Angke Fishing Port, Jakarta, IOP Conference Series: Earth and Environmental Science, Institute of Physics, 2024, doi:10.1088/1755-1315/1359/1/012098.

[11] G. Airlangga, A. Rachmat, D. Lapihu, Comparison of exponential smoothing and neural network method to forecast rice production in Indonesia, Telkomnika (Telecommunication Computing Electronics and Control) 17 (3) (2019) 1367–1375, doi:10.12928/TELKOMNIKA.V17I3.11768.

[12] A.L.M. Serrano et al., Statistical Comparison of Time Series Models for Forecasting Brazilian Monthly Energy Demand Using Economic, Industrial, and Climatic Exogenous Variables, Applied Sciences (Switzerland) 14 (13) (2024) Jul. doi:10.3390/app14135846.

[13] D. Zhao, R. Zhang, H. Zhang, S. He, Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models, Sci. Rep. 12 (1) (2022) 1–13, doi:10.1038/s41598-022-23154-4.

[14] G. Perone, Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy, European Journal of Health Economics 23 (6) (2022) 917–940, doi:10.1007/s10198-021-01347-4.

[15] N. Mohamed, M.H. Ahmad, Z. Ismail, S. Suhartono, Short term load forecasting using double seasonal ARIMA Model, in: Proceedings of the Regional Conference on Statistical Sciences, 2010, 2010, pp. 57–73.

[16] A. Alsongan, M. Gokcen, A hybrid forecasting structure based on arima and artificial neural network models, Applied Sciences (Switzerland) 14 (16) (2024) Aug. doi:10.3390/app14167122.

[17] N. Suhermi, D.D.Prastyo Suhartono, B. Ali, Roll motion prediction using a hybrid deep learning and ARIMA model, in: Procedia Computer Science, Elsevier B.V., 2018, pp. 251–258, doi:10.1016/j.procs.2018.10.526.

[18] C.B. Aditya Satrio, W. Darmawan, B.U. Nadia, N. Hanafiah, Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET, in: Procedia Computer Science, Elsevier B.V., 2021, pp. 524–532, doi:10.1016/j.procs.2021.01.036.

[19] V.I. Kontopoulou, A.D. Panagopoulos, I. Kakkos, G.K. Matsopoulos, A Review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks, Multidiscip. Digital Publish. Institute (MDPI) (2023), doi:10.3390/fi15080255.

[20] T.T.K. Tran, S.M. Bateni, S.J. Ki, H. Vosoughifar, A review of neural networks for air temperature forecasting, MDPI AG (2021), doi:10.3390/w13091294.

[21] W. Ge, P. Lalbakhsh, L. Isai, A. Lenskiy, H. Suominen, Neural network-based financial volatility forecasting: a systematic review, Assoc. Comput. Mach. (2022), doi:10.1145/3483596.

[22] I.D. Mienye, T.G. Swart, G. Obaido, Recurrent neural networks: a comprehensive review of architectures, variants, and applications, Information 15 (9) (2024) 517 Aug. doi:10.3390/info15090517.

[23] H. Apaydin, H. Feizi, M.T. Sattari, M.S. Colak, S. Shamshirband, K.W. Chau, Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting, Water (Switzerland) 12 (5) (2020) May. doi:10.3390/w12051500.

[24] M. Khashei, M. Bijari, An artificial neural network (p, d, q) model for timeseries forecasting, Expert. Syst. Appl. 37 (1) (2010) 479–489 Jan. doi:10.1016/j.eswa.2009.05.044.

[25] S. Belhaiza, S. Al-Abdallah, A Neural Network Forecasting Approach for the Smart Grid Demand Response Management Problem, Energies (Basel) 17 (10) (2024) May. doi:10.3390/en17102329.

[26] A.T.R. Dani, M. Fauziyah, H. Sandariria, Q.Q. A'yun, Forecasting the search trends of the keyword 'Sarung Wadimor' in Indonesia on google trends data using time series regression with calendar variation and Arima Box-Jenkins, Jurnal Matematika, Statistika dan Komputasi 19 (3) (2023) 447–459 May. doi:10.20956/j.v19i3.24551.

[27] H. Alabdulrazzaq, M.N. Alenezi, Y. Rawajfih, B.A. Alghannam, A.A. Al-Hassan, F.S. Al-Anzi, On the accuracy of ARIMA based prediction of COVID-19 spread, Results Phys. 27 (2021) 104509, doi:10.1016/j.rinp.2021.104509.

[28] I. Mado, A. Soeprijanto, Suhartono, Applying of double seasonal arima model for electrical power demand forecasting at pt. pln gresik Indonesia, International Journal of Electrical and Computer Engineering 8 (6) (2018) 4892–4901 Dec. doi:10.11591/ijece.v8i6.pp4892-4901.

[29] Y. Wang, R. Zou, F. Liu, L. Zhang, Q. Liu, A review of wind speed and wind power forecasting with deep neural networks, Appl. Energy 304 (2021) Dec. doi:10.1016/j.apenergy.2021.117766.

[30] G. Tkacz, "Neural network forecasting of Canadian GDP growth," 2001. [Online]. Available: www.elsevier.com/locate/ijforecast.

[31] Y. Wang, R. Zou, F. Liu, L. Zhang, Q. Liu, A review of wind speed and wind power forecasting with deep neural networks, Appl. Energy 304 (2021) Dec. doi:10.1016/j.apenergy.2021.117766.

[32] M. Adya, F. Collopy, How effective are neural networks at forecasting and prediction? A review and evaluation, J. Forecast. 17 (1998) 481–495.

[33] N.H.A. Rahman, M.H. Lee, Suhartono, M.T. Latif, Artificial neural networks and fuzzy time series forecasting: an application to air quality, Qual. Quant. 49 (6) (2015) 2633–2647 Nov. doi:10.1007/s11135-014-0132-6.

[34] W. Sulandari, Subanar, M.H. Lee, P.C. Rodrigues, Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks, Energy 190 (2020) Jan. doi:10.1016/j.energy.2019.116408.

Naskah Mislan

ORIGINALITY REPORT

19%	13%	14%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ime.unicamp.br Internet Source	1%
2	isroset.org Internet Source	1%
3	ghhin.org Internet Source	1%
4	Submitted to University of Southampton Student Paper	1%
5	mafiadoc.com Internet Source	1%
6	Submitted to University of Western Sydney Student Paper	1%
7	Yuli Rahmawati, Peter Charles Taylor. "Empowering Science and Mathematics for Global Competitiveness", CRC Press, 2019 Publication	1%
8	Vita Ratnasari, Purhadi, Marisa Rifada, Andrea Tri Rian Dani. "Explore Poverty with Statistical Modeling: The Bivariate Polynomial Binary Logit Regression (BPBLR) .", MethodsX, 2024 Publication	1%
9	Ansari Saleh Ahmar, Eva Boj del Val, M. A. El Safty, Samirah AlZahrani, Hamed El-Khawaga. "SutteARIMA: A Novel Method for Forecasting the Infant Mortality Rate in Indonesia", Computers, Materials & Continua, 2022 Publication	1%

10	www.grafiati.com Internet Source	1 %
11	Submitted to Universitas Pendidikan Indonesia Student Paper	<1 %
12	www.intechopen.com Internet Source	<1 %
13	journal.arrus.id Internet Source	<1 %
14	Submitted to Cardiff University Student Paper	<1 %
15	123dok.com Internet Source	<1 %
16	Submitted to Universitas Brawijaya Student Paper	<1 %
17	Submitted to University of Mauritius Student Paper	<1 %
18	www.ssca.org.in Internet Source	<1 %
19	Submitted to Assumption University Student Paper	<1 %
20	Gary Feng, Stacy Cobb, Zaid Abdo, Daniel K. Fisher, Ying Ouyang, Ardeshir Adeli, Johnie N. Jenkins. "Trend Analysis and Forecast of Precipitation, Reference Evapotranspiration, and Rainfall Deficit in the Blackland Prairie of Eastern Mississippi", Journal of Applied Meteorology and Climatology, 2016 Publication	<1 %
21	M. Imron, Wika Dianita Utami, Hani Khaulasari, Firman Armunanto. "ARIMA MODEL OF OUTLIER DETECTION FOR FORECASTING CONSUMER PRICE INDEX	<1 %

(CPI)", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2022
Publication

22	Lathifah Lathifah, Dina Agustina. "ADDITIVE HOLT-WINTERS METHOD FOR FORECASTING GROSS REGIONAL DOMESTIC PRODUCT AT CONSTANT PRICES OF EXPENDITURE OF WEST SUMATRA", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2024 Publication	<1 %
23	eurusd.site Internet Source	<1 %
24	ideas.repec.org Internet Source	<1 %
25	iris.imtlucca.it Internet Source	<1 %
26	"Swarm, Evolutionary, and Memetic Computing", Springer Science and Business Media LLC, 2013 Publication	<1 %
27	Submitted to Universitas Diponegoro Student Paper	<1 %
28	journals.plos.org Internet Source	<1 %
29	docplayer.net Internet Source	<1 %
30	Chaâbane, Najeh. "A novel auto-regressive fractionally integrated moving average–least-squares support vector machine model for electricity spot prices prediction", Journal of Applied Statistics, 2014. Publication	<1 %
31	Submitted to Liverpool John Moores University	<1 %

Student Paper

32	knm21.unram.ac.id Internet Source	<1 %
33	library.iugaza.edu.ps Internet Source	<1 %
34	www.walshmedicalmedia.com Internet Source	<1 %
35	Ade Gafar Abdullah, Isma Widiaty, Cep Ubad Abdullah. "Medical Technology and Environmental Health", CRC Press, 2020 Publication	<1 %
36	Gao, D.. "Neural networks for event extraction from time series: a back propagation algorithm approach", Future Generation Computer Systems, 200507 Publication	<1 %
37	Yongtao Wei, Jinkuan Wang, Cuirong Wang, Junwei Wang. "Network Traffic Prediction by Traffic Decomposition", 2012 Fifth International Conference on Intelligent Networks and Intelligent Systems, 2012 Publication	<1 %
38	heca-analitika.com Internet Source	<1 %
39	macsphere.mcmaster.ca Internet Source	<1 %
40	pdfs.semanticscholar.org Internet Source	<1 %
41	www.accentsjournals.org Internet Source	<1 %
42	www.ihp-microelectronics.com Internet Source	<1 %
43	www.tandfonline.com Internet Source	<1 %

44	Görkem Ekleme, Funda Yercan. "Assessment of the Black Sea Grain Initiative: Crisis Management via Maritime Transportation", Transport Policy, 2025 Publication	<1 %
45	Ismit Mado, Adi Soeprijanto, Suhartono Suhartono. "Applying of Double Seasonal ARIMA Model for Electrical Power Demand Forecasting at PT. PLN Gresik Indonesia", International Journal of Electrical and Computer Engineering (IJECE), 2018 Publication	<1 %
46	Kavasseri, R.G.. "Day-ahead wind speed forecasting using f-ARIMA models", Renewable Energy, 200905 Publication	<1 %
47	Shigeru Thomas Otsubo, Christian Samen Otchia. "Designing Integrated Industrial Policies Volume I - For Inclusive Development in Asia", Routledge, 2020 Publication	<1 %
48	Veerasak Punyapornwithaya, Orapun Arjkumpa, Noppawan Buamithup, Noppasorn Kuatako et al. "Forecasting of daily new lumpy skin disease cases in Thailand at different stages of the epidemic using fuzzy logic time series, NNAR, and ARIMA methods", Preventive Veterinary Medicine, 2023 Publication	<1 %
49	Zhang, W., W. Hou, P. Hu, and Z. Ma. "The Nonlinear Compressive Response and Deformation of an Auxetic Cellular Structure under In-Plane Loading", Advances in Mechanical Engineering, 2015. Publication	<1 %

50	alphanumericjournal.com Internet Source	<1 %
51	bura.brunel.ac.uk Internet Source	<1 %
52	discovery.researcher.life Internet Source	<1 %
53	erepository.uoeld.ac.ke Internet Source	<1 %
54	export.arxiv.org Internet Source	<1 %
55	isda2001.softcomputing.net Internet Source	<1 %
56	liu.diva-portal.org Internet Source	<1 %
57	mospace.umsystem.edu Internet Source	<1 %
58	www.frontiersin.org Internet Source	<1 %
59	Bagui, S.. "Classification with multiple independent measurements under a separate sampling scheme", Statistical Methodology, 200607 Publication	<1 %
60	Erick Meira, Fernando Luiz Cyrino Oliveira, Lilian M. de Menezes. "Point and interval forecasting of electricity supply via pruned ensembles", Energy, 2021 Publication	<1 %
61	Man-Chun Tan. "An Aggregation Approach to Short-Term Traffic Flow Prediction", IEEE Transactions on Intelligent Transportation Systems, 03/2009 Publication	<1 %

62	Navid Toufan, Aliakbar Niknafs. "Robot path planning based on laser range finder and novel objective functions in grey wolf optimizer", SN Applied Sciences, 2020 Publication	<1%
63	Usha Sakthivel, Nagarajan Susila. "Applied Learning Algorithms for Intelligent IoT", CRC Press, 2021 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

Naskah Mislan

GRADEMARK REPORT

FINAL GRADE

/100

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12