BUKTI KORESPONDENSI
ARTIKEL JURNAL INTERNASIONAL BEREPUTASI

Judul Artikel	: Navigating Samarinda's climate: A comparative analysis of rainfall forecasting models
Jurnal	: MethodsX, Volume 14, June 2025, 103080
Penulis	: **Mislan**, Andrea Tri Rian Dani

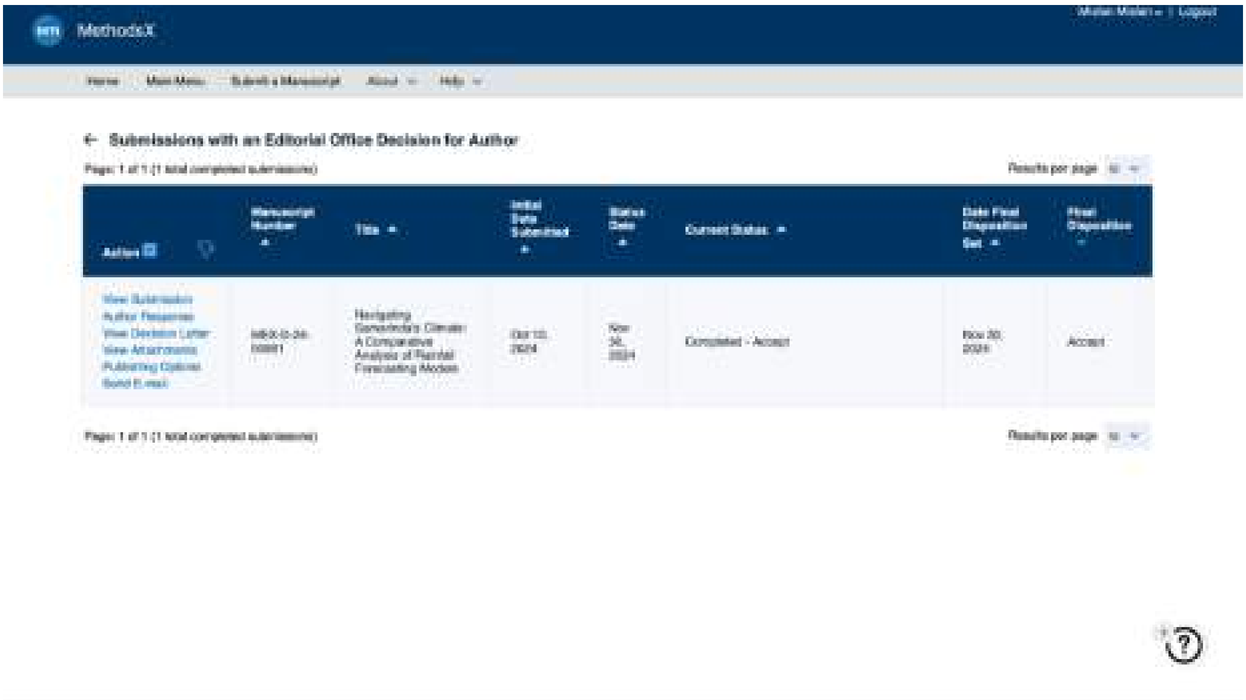
Ringkasan Proses Korespondensi :

No.	Perihal (Link)	Tanggal
1.	Bukti submission 1. Registrasi Akun 2. Bukti submit 3. Manuscript 4. Graphical Abstract	05 Oktober 2024 10 Oktober 2024
2.	Bukti email Submission Acknowledgement	10 Oktober 2024
3.	Bukti email Track your Elsevier submission	14 Oktober 2024
4.	Bukti email Editorial Decision Reviewer Comment 1: Reviewer #1: I reviewed the manuscript titled "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" which aims to compare traditional forecasting models (Exponential Smoothing and ARIMA) and a machine learning model (Neural Network) in predicting rainfall in Samarinda, Indonesia, using monthly data from 2000-2020. The research seeks to identify the best model for forecasting future rainfall trends to improve early warnings and disaster mitigation strategies. Some comments are provided below: -The study briefly mentions the choice of Exponential Smoothing, ARIMA, and Neural Networks, but it lacks a detailed justification for selecting these specific methods over others. A stronger rationale for why these models were chosen, compared to other possible alternatives like SARIMA or other machine learning models (e.g., Random Forest), would strengthen the methodology. -While the study discusses using 90% of the data for training and 10% for testing, it does not provide enough information on how missing data, outliers, or potential seasonal components in rainfall were handled during preprocessing. Addressing these issues would enhance the clarity and reliability of the results. -The description of the Neural Network architecture (2 hidden layers with 6-5 neurons) lacks a clear explanation of how this specific configuration was chosen. A discussion on how different architectures were tested and the impact of tuning hyperparameters would add depth to the model development process.	23 Oktober 2024

	-The forecasting results indicate that the highest rainfall is predicted for January, but the paper does not sufficiently explain the practical implications of these results. More emphasis on how these predictions can directly impact disaster mitigation or urban planning would better highlight the real-world application of the study. -Although the paper includes graphical representations, the visualizations could be improved for clarity. -The paper does not address any limitations of the study, such as potential model overfitting, reliance on historical data that may not account for future climatic changes, or the limited generalizability of the findings to other regions. A section outlining these constraints would provide a more balanced view of the research. Reviewer Comment 2: Reviewer #2: In general, this article is good, it needs a few additions or improvements to make it better. Please check in attached file Check again the english for grammatical and typos. In conclusion discuss more about the results and add future development of this research. * Abstract (very good) * Graphical Abstract (acceptable) * Methods (acceptable) * References (very good)	
5	Bukti Submission setelah Review Process 1 1. Bukti Submission 2. Manuskrip Revisi 1 3. Turnitin Check Revisi 1 4. Submission Confirmation dari MethodsX	26 Oktober 2024
6	Bukti email Track your Submission after Revision 1	29 Oktober 2024
7	Bukti email Editorial Decision Reviewer Comment 3: Reviewer #3: This is not a true method article becuase it does not present a new or improved method. It simply compares well-known methods used to predict rainfall (e.g., ARIMA and NN) to conclude that NN perform better in a region of Indonesia. Although the description of methods is Ok and the test to the selected region is also Ok, I cannot recommend approval of this submission becuase is not a method is a comparison of well-known methods that taken together do not form a new method. Reviewer Comment 4: Reviewer #4: I have read the revised manuscript with title "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" which has covered the initial comments of the reviewers. To my opinion is well explained method and deserve publication.	13 November 2024
8	Bukti Submission setelah Review Process 2 1. Bukti Submission	19 November 2024

	2. Manuskrip Revisi 2 3. Turnitin Check Revisi 2 4. Submission Confirmation dari MethodsX	
9	Bukti email Track your Submission after Revision 2	26 November 2024
10	Bukti email Notifikasi accepted	30 November 2024
11	Bukti email Notifikasi Invoice	03 Desember 2024
12	Bukti email Publishing Agreement	03 Desember 2024
13	Bukti email Galley Proof	07 Desember 2024
14	Bukti Published Naskah Final dan Link Akses: https://www.sciencedirect.com/science/article/pii/S2215016124005314	
15	Check Turnitin Final Manuscript: Turnitin Mislan	

Dokumen bukti terlampir pada halaman berikut.



Gmail Andrea T R Dani <andrikadoko@gmail.com>

Editorial Manager Registration for MethodsX

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 5 Oktober 2024 pukul 19.56
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

Dear Mr Mislan,

Thank you for registering for the Editorial Manager online submission and peer review tracking system for MethodsX.

Here is your username, which you need to access Editorial Manager at https://www.editorialmanager.com/mex/.
Username: mislanjoss

If you do not know your confidential password, you may reset it by clicking this link: click here to reset your password

Please save this information in a safe place.

You can change your password and other personal information by logging into the MethodsX website and clicking on the Update My Information link on the menu.

Best regards,
MethodsX

For more information on using EM, please refer to Aries Systems' user Tutorials and Video Library:
https://www.ariessys.com/for-current-customers/tutorials/
https://www.ariessys.com/views-and-press/resources/video-library/
You can also access these videos by clicking the "Play" icon in the top right corner of the EM site.

More information and support
FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

---

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

Gmail

Andrea T R Dani <andrikadoko@gmail.com>

Your PDF has been built and requires approval

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 10 Oktober 2024 pukul 20.38
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

MethodsX
Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
Authors: Mislan Mislan; Andrea Tri Rian Dani

Dear Mislan,

The PDF for your submission, "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" has now been built and is ready for your approval. Please view the submission before approving it, to be certain that it is free of any errors. If you have already approved the PDF of your submission, this e-mail can be ignored.

To approve the PDF please login to the Editorial Manager as an Author:

https://www.editorialmanager.com/mex/
Your username is: mislanjoss

Then click on the folder 'Submissions Waiting for Author's Approval' to view and approve the PDF of your submission. You may need to click on 'Action Links' to expand your Action Links menu.

You will also need to confirm that you have read and agree with the Elsevier Ethics in Publishing statement before the submission process can be completed. Once all of the above steps are done, you will receive an e-mail confirming receipt of your submission from the Editorial Office. For further information or if you have trouble completing these steps please go to: http://help.elsevier.com/app/answers/detail/a_id/88/p/7923.

Please note that you are required to ensure everything appears appropriately in PDF and no change can be made after approving a submission. If you have any trouble with the generated PDF or completing these steps please go to: http://help.elsevier.com/app/answers/detail/a_id/88/p/7923.

Your submission will be given a reference number once an Editor has been assigned to handle it.

Thank you for your time and patience.
Kind regards,
Editorial Office
MethodsX

****
For further assistance, please visit our customer support site at http://help.elsevier.com/app/answers/list/p/7923. Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EM via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

***MethodsX* methods article template Version 6 (April 2024)**

AUTHOR INSTRUCTIONS

Before you complete this template, a few important points to note:

- This template is for a methods article. If you want to describe a protocol or changes you've made to a protocol; for example, for use in a clinical or drug trial, please complete our protocol article template.
- The format of a methods article is very different from a traditional research article. To help you write yours, we have created this template. **We will only consider methods articles submitted using this template**.
- Please consult the *MethodsX* Guide for Authors; it highlights mandatory requirements and is packed with useful advice. We have also developed a step-by-step video guide to completing this template, designed to maximize your chances of acceptance.
- Brief **instructions are in blue italics**. Additional instructions appear in comment boxes to the right of the main text.
- Please do not change any section headings (e.g., *Article information*, *Specifications table*).

Still have questions?

- Email our editorial team at mex-me@elsevier.com

Now you are ready to fill in the template below. As you complete each section, it's important to read the associated instructions carefully. All sections are mandatory, unless marked OPTIONAL.

Please delete this line and everything above it before submitting your article. In addition, we ask you to delete the below instructions in blue, including those featured in the Article Information and Specifications table.

- -

MethodsX
Open access
Article Template

Article information

Article title
Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models.

Authors
Mislan, Andrea Tri Rian Dani*

Affiliations
Department of Physics, Faculty of Mathematics and Natural Science, Mulawarman University; Department of Mathematics, Faculty of Mathematics and Natural Science, Mulawarman University

Corresponding author's email address and Twitter handle
mislan.jos@fmipa.unmul.ac.id

Keywords
Exponential Smoothing; ARIMA; Neural Network; Time Series Modeling; Forecasting

Related research article
None

For a published article:

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.

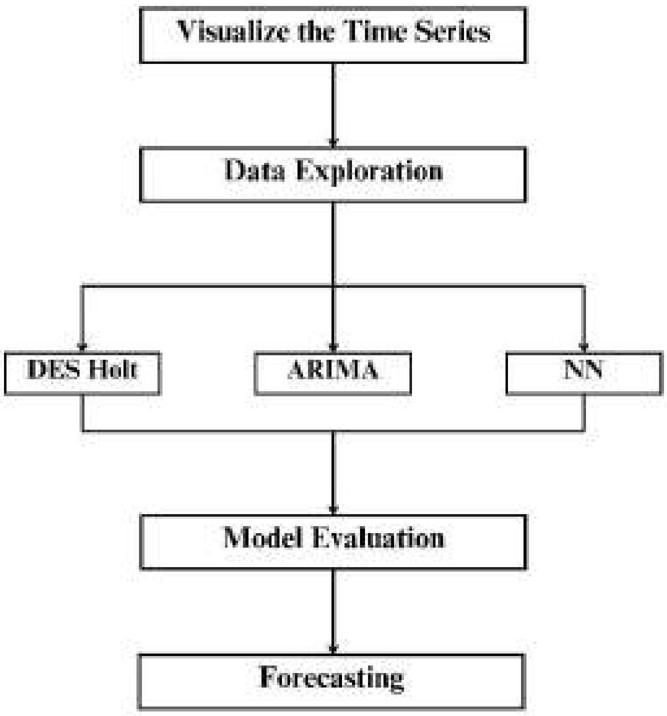
N. H. A. Rahman, M. H. Lee, Suhartono, and M. T. Latif, "Artificial neural networks and fuzzy time series forecasting: an application to air quality," Qual Quant, vol. 49, no. 6, pp. 2633–2647, Nov. 2015, doi: 10.1007/s11135-014-0132-6.

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 drought.*

Graphical abstract



The research design used was ex post facto, meaning data was collected after all the events. The stages of data analysis modeling rainfall data in Samarinda City are visualized in the Graphical Abstract. The researchers chose the three methods based on their advantages and flexibility in the modeling process. The modeling process uses R software.

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.* *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.*

Resource availability	*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 [10], [11], [12], 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], [16], [17], [18], [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 may miss. Several time series data studies have successfully utilized NN are [16], [21], [22], [23], [24], [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

A. *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 < α < 1

β : trend smoothing parameter, 0 < α < 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)

B. 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:

$$(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)Z_t=(1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)a_t \quad (5)$$

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:

1. *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).

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

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

$$\phi_p(B)Z_t=a_t \quad (8)$$

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

2. *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).

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

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

$$Z_t=\theta_q(B)a_t \quad (10)$$

   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:*

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*	*Nothings is out of bounds*

C. 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 generalize 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 laye*[32]*r. As an illustration, it can be seen in Figure 1.*

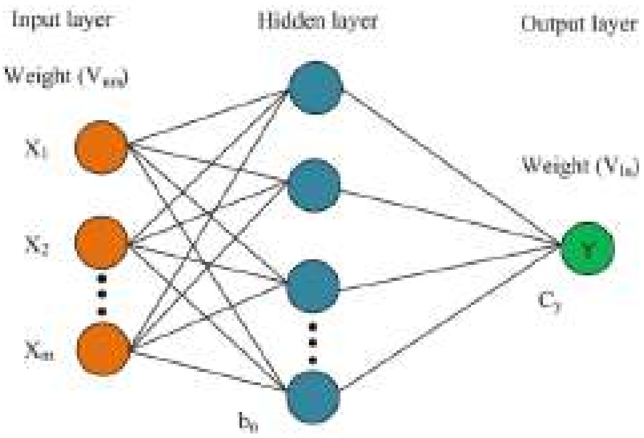


Fig. 1. Neural Network Structure

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]*. 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).*

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

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

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

D. 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).

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

E. 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 Figure 2.

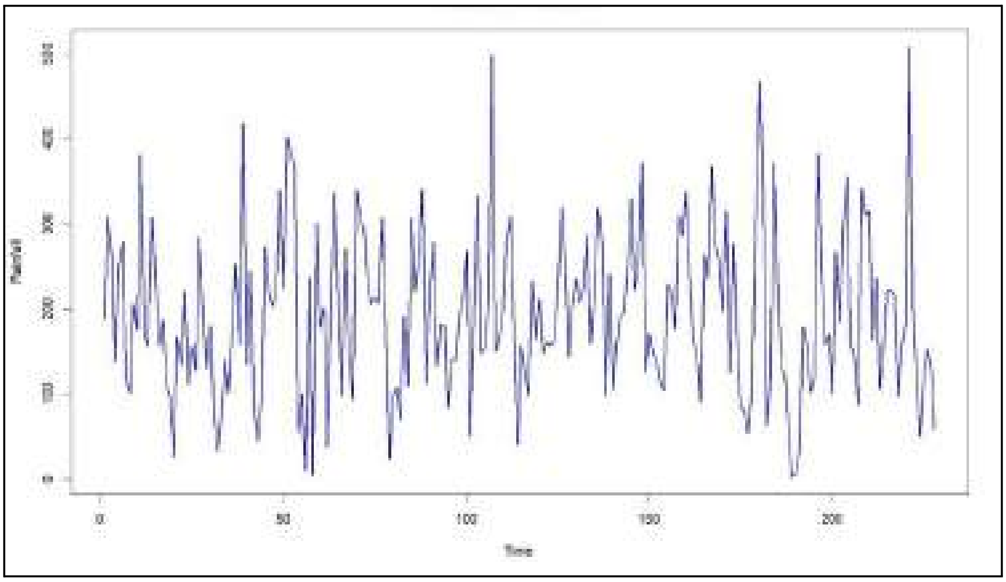


Fig. 2. Time series plot of rainfall data in Samarinda

Based on Figure 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

a. *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 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.

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	115.28	0.7	0.1	114.66
0.1	0.2	168.88	0.4	0.2	114.72	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	120.05	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	138.89	0.4	0.6	125.13	0.7	0.6	133.49
0.1	0.7	140.94	0.4	0.7	127.14	0.7	0.7	137.81
0.1	0.8	143.61	0.4	0.8	129.14	0.7	0.8	142.23
0.1	0.9	146.54	0.4	0.9	131.40	0.7	0.9	146.69
0.2	0.1	138.03	0.5	0.1	113.38	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	124.69	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	130.18	0.5	0.6	126.01	0.8	0.6	138.95
0.2	0.7	133.68	0.5	0.7	129.02	0.8	0.7	143.98
0.2	0.8	138.00	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	120.05
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	129.47
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	125.59	0.9	0.5	140.34
0.3	0.6	127.90	0.6	0.6	129.19	0.9	0.6	146.16
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	136.90	0.9	0.8	158.77
0.3	0.9	133.12	0.6	0.9	140.91	0.9	0.9	165.67

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. 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	113.40	121.13
α = 0.7 and β = 0.1	114.66	118.03

b. 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 Figures 3 and 4.

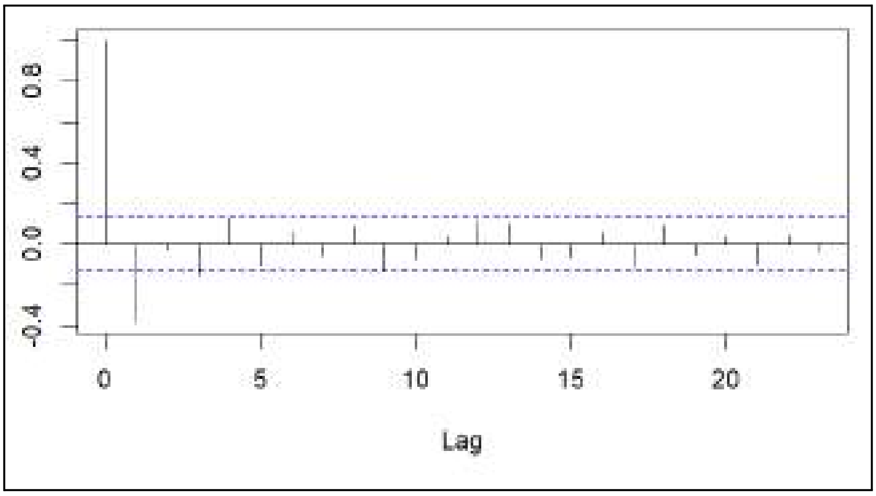


Fig. 3. ACF Plot of Rainfall Data Results of Differencing

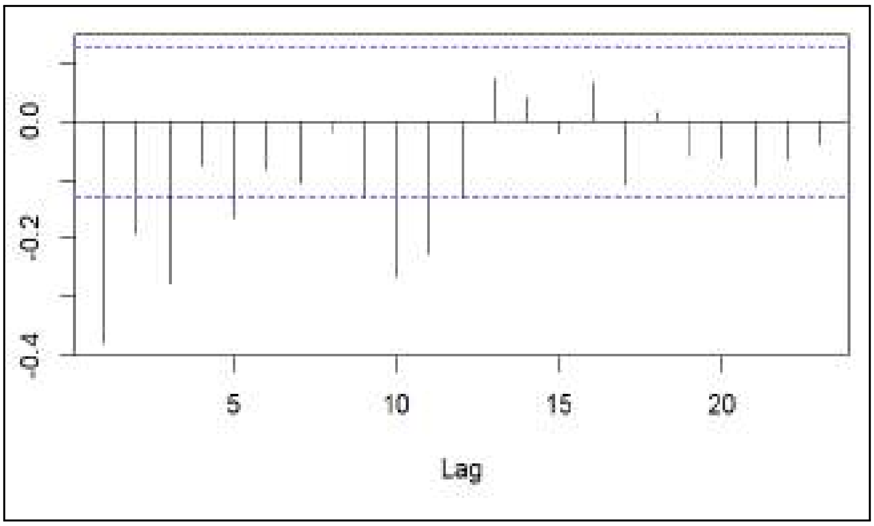


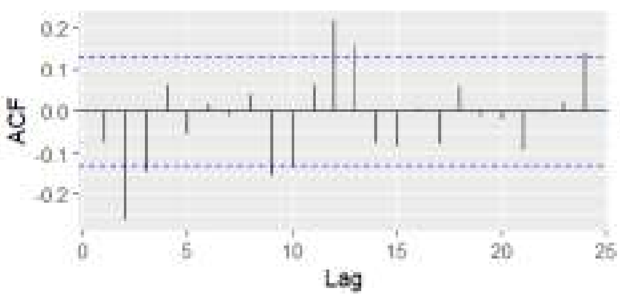
Fig. 4. PACF Plot of Rainfall Data Results of Differencing

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).

Table 5. Temporary ARIMA Model Parameter Estimation

Model	Parameter	Estimate	*p-value*	Conclusion
ARIMA(1,1,0)	$\hat{\phi}_1$	-0.380766	5.579e-10	Significant
ARIMA(2,1,0)	$\hat{\phi}_1$	-0.455156	2.982e-12	Significant
	$\hat{\phi}_2$	-0.193967	0.002867	
ARIMA(3,1,0)	$\hat{\phi}_1$	-0.507994	1,760e-15	Significant
	$\hat{\phi}_2$	-0.320263	3.421e-06	
	$\hat{\phi}_3$	-0.275761	1.507e-05	
ARIMA(0,1,1)	$\hat{\theta}_1$	-1,000000	< 2.2e-16	Significant
ARIMA(1,1,1)	$\hat{\phi}_1$	0.273909	2.071e-05	Significant
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(2,1,1,)	$\hat{\phi}_1$	0.266500	6.351e-05	Not Significant
	$\hat{\phi}_2$	0.028447	0.6701	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(3,1,1)	$\hat{\phi}_1$	0.268049	5.375e-05	Not Significant
	$\hat{\phi}_2$	0.048998	0.475	
	$\hat{\phi}_3$	-0.081457	0.220	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	

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.



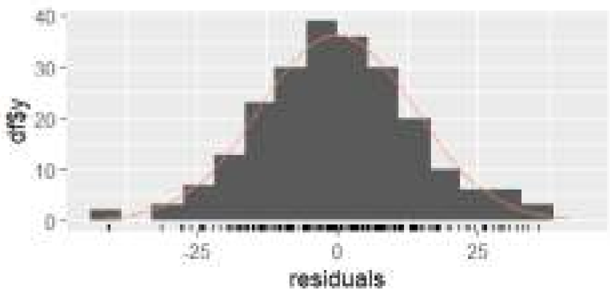


Fig. 5. Independence and Normality of ARIMA Residuals (1,1,0)

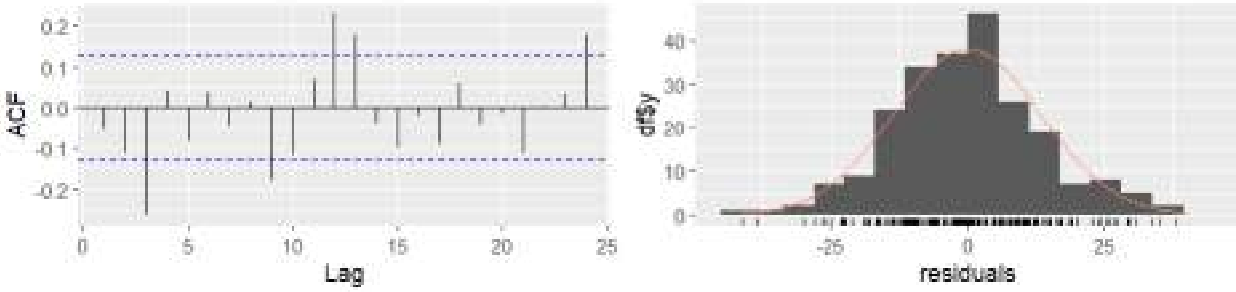


Fig. 6. Independence and Normality of ARIMA Residuals (2,1,0)

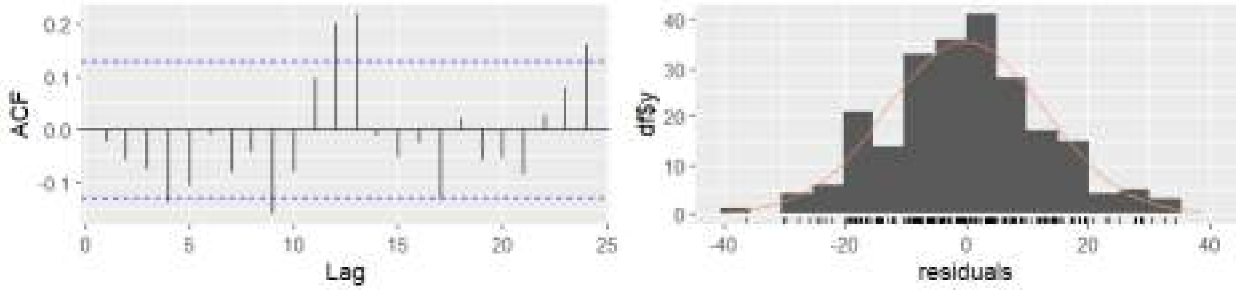


Fig. 7. Independence and Normality of ARIMA Residuals (3,1,0)

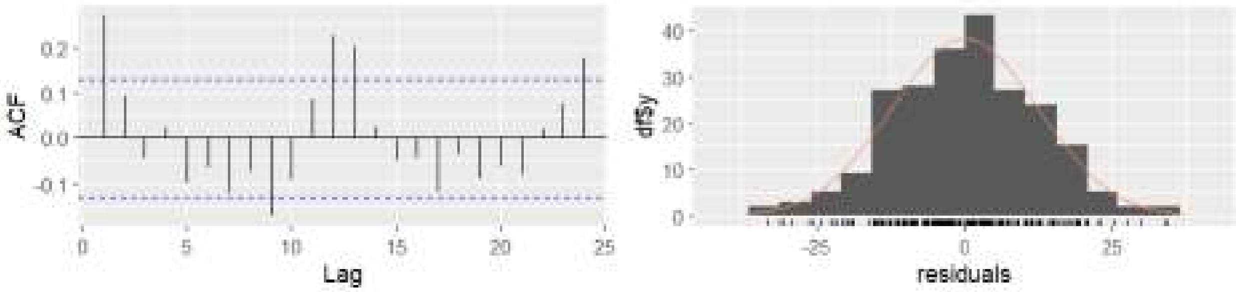


Fig. 8. Independence and Normality of ARIMA Residuals (0,1,1)

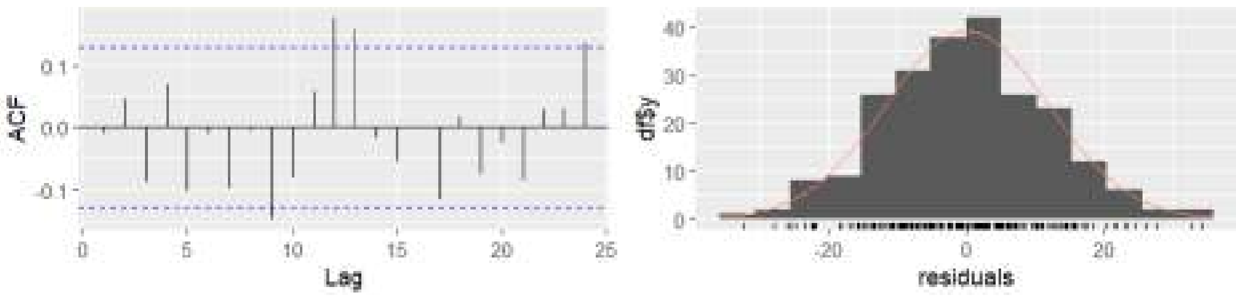


Fig. 9. Independence and Normality of ARIMA Residuals (1,1,1)

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.

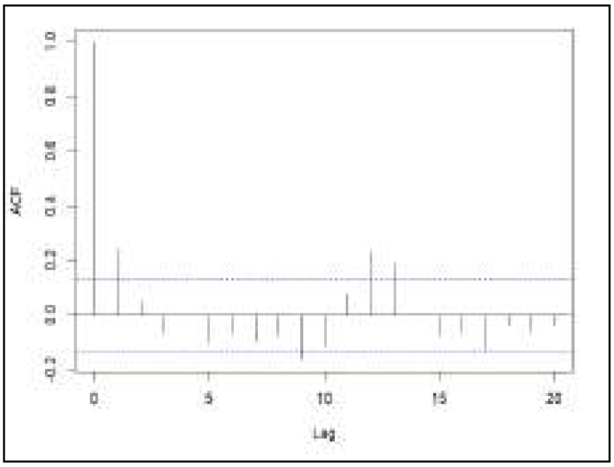
Table 6. Accuracy of Training and Testing Data for ARIMA Model (1,1,1)

Proportion	RMSEP
Data Training	92.75
Data Testing	80.82

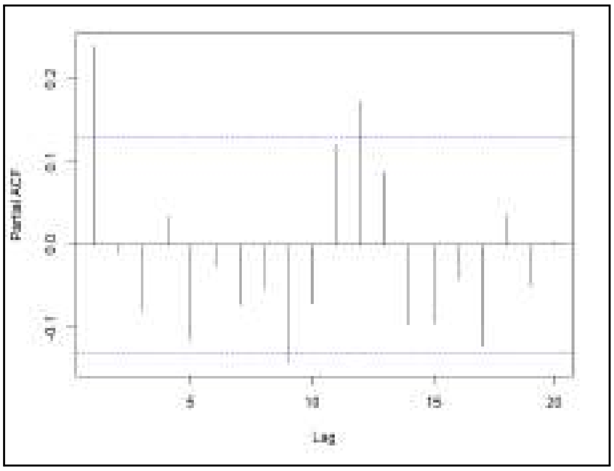
c. *Modeling with Neural Network (NN)*

- *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 Figure 10.



(a) ACF graph



(b) PACF graph

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

Based on Figure 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.

- *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.

- *Best Model Selection in NN*

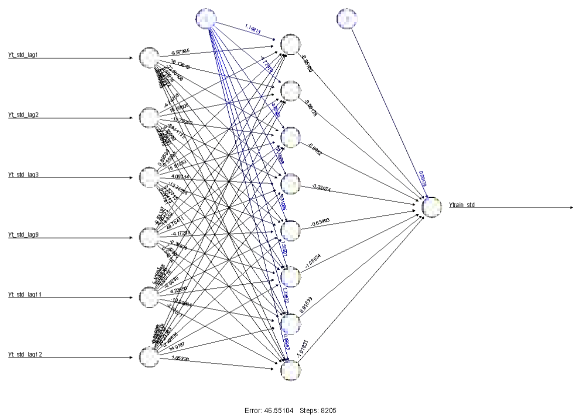
The backpropagation training process is carried out by adjusting the NN architecture. In this study, two types of architectures are used, namely networks with one hidden layer and networks with two hidden layers where each architecture will try various combinations of the number of neurons in each layer. 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.*

Determination of the best NN architecture model to be used in predicting rainfall in Samarinda is done based on the RMSE of training data. Based on the calculation results, the RMSEP value of each model is obtained which can be seen in Table 7.

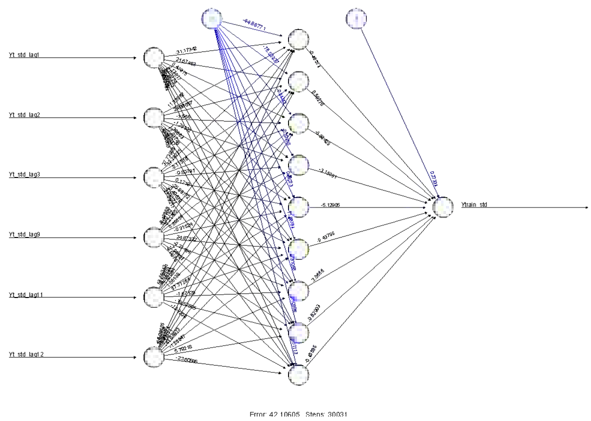
Table 7. RMSE Calculation Rsults

Hidden Layer		RMSEP		Difference between Training and Testing RMSEP
Hidden Layer1	**Hidden Layer2**	**Training**	**Testing**	
1 neuron		88.988	86.595	2.393
2 neurons		85.018	90.197	5.180
3 neurons		78.101	110.632	32.531
4 neurons		75.907	100.158	24.251
5 neurons		66.340	97.531	31.192
6 neurons		68.059	96.647	28.589
7 neurons		69.542	103.764	34.222
8 neurons		58.299	102.938	44.639
9 neurons		57.471	114.915	57.444
10 neurons		46.560	110.919	64.359
2 neurons	1 neuron	83.746	82.969	0.777
3 neurons	2 neurons	77.546	100.939	23.393
4 neurons	3 neurons	70.852	97.184	26.332
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	135.889	106.194
9 neurons	8 neurons	27.501	116.387	88.886
10 neurons	9 neurons	17.891	181.869	163.978

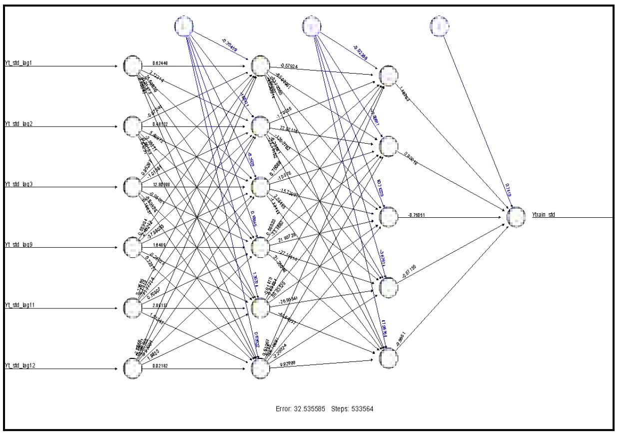
By considering various considerations such as choosing the smallest RMSEP value and the difference between the RMSEP values of the training data and testing data is not very significant, then based on Table 7, it is found that the NN model with a 2 hidden layer architecture model (6-5 neurons) is the 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 89.749. Some of the architectural results of the NN modeling can be seen in Figure 11.



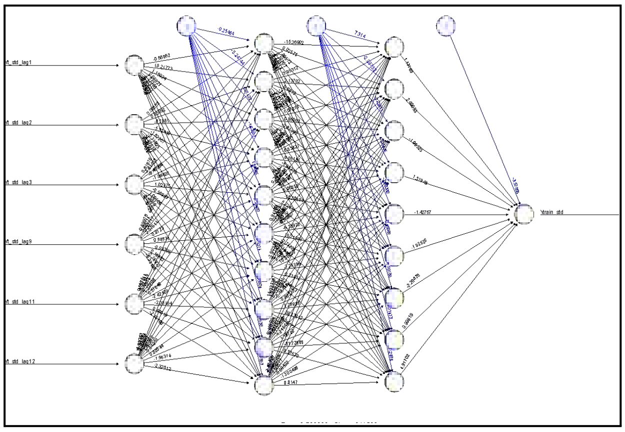
(a) NN model 1 hidden layer (8 neurons)



(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

d. 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 8 displays the RMSEP values of the best models. The model with the smallest RMSE value will be selected as the best method.

Table 8. Model Goodness of Fit Measure

Method	**RMSEP**	
	Training	Testing
DES Holt (α = 0.7 and β = 0.1)	114.66	118.03
ARIMA (1,1,1)	92.75	80.82
NN 2 HL (6-5 neurons)	**52.473**	**89.749**

In Table 8, 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.

e. 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.

Table 9. Forecasting Results of 12 Periods of NN 2 HL (6-5 neurons)

Month	Prediction Results
January	301.935
February	245.819
March	168.822
April	82.964
May	173.226
June	209.721
July	107.147
August	204.006
September	201.430
October	133.406
November	207.887
December	206,919

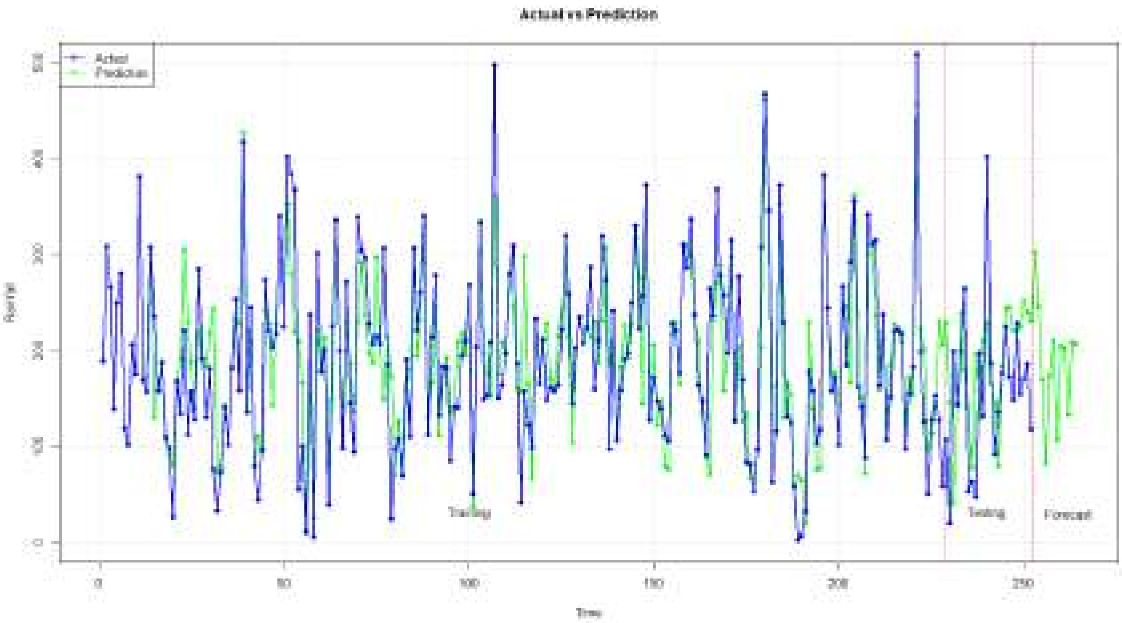


Fig. 12. Comparison plot of actual and predicted data

Figure 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.

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, 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.

Limitations

'Not applicable'.

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.

CRediT author statement

Mislan: Conceptualization, Methodology, Validity tests, Writing-Preparation of the first draft, and Supervision. **Andrea Tri Rian Dani**: Data curation, Analysis, Visualization, and Editing Draft, and writing original draft.

Acknowledgments

None

Declaration of interests

*Please **tick** the appropriate statement below (please do not delete either statement) and declare any financial interests/personal relationships which may affect your work in the box below.*

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Please declare any financial interests/personal relationships which may be considered as potential competing interests here.

Supplementary material *and/or* additional information [OPTIONAL]

None.

References

[1] M. Putra, M. S. Rosid, and D. Handoko, “A Review of Rainfall Estimation in Indonesia: Data Sources, Techniques, and Methods,” *Signals*, vol. 5, no. 3, pp. 542–561, Aug. 2024, doi: 10.3390/signals5030030.

[2] 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.

[3] R. P. Permata, R. Ni’mah, and A. T. R. Dani, “Daily Rainfall Forecasting with ARIMA Exogenous Variables and Support Vector Regression,” *Jurnal Varian*, vol. 7, no. 2, pp. 177–188, Jun. 2024, doi: 10.30812/varian.v7i2.3202.

[4] M. Mislan and A. T. R. Dani, “Forecasting Maximum Water Level Data for Post Sangkuliman using An Artificial Neural Network Backpropagation Algorithm,” *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 8, no. 2, p. 465, Mar. 2024, doi: 10.31764/jtam.v8i2.20112.

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

[6] S. Makridakis, R. J. Hyndman, and F. Petropoulos, “Forecasting in social settings: The state of the art,” *Int J Forecast*, vol. 36, no. 1, pp. 15–28, Jan. 2020, doi: 10.1016/j.ijforecast.2019.05.011.

[7] S. Makridakis, E. Spiliotis, V. Assimakopoulos, and 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, and M. S. El-Sherif, "A Comparison Between Neural-Network Forecasting Techniques-Case Study: River Flow Forecasting," 1999.

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

[10] Mustaruddin, G. Puspito, B. Z. Zanlee, and T. Nugroho, "The exponential smoothing model for squid production in Muara Angke Fishing Port, Jakarta," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2024. doi: 10.1088/1755-1315/1359/1/012098.

[11] G. Airlangga, A. Rachmat, and D. Lapihu, "Comparison of exponential smoothing and neural network method to forecast rice production in Indonesia," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 3, pp. 1367–1375, 2019, 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)*, vol. 14, no. 13, Jul. 2024, doi: 10.3390/app14135846.

[13] D. Zhao, R. Zhang, H. Zhang, and S. He, "Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models," *Sci Rep*, vol. 12, no. 1, pp. 1–13, 2022, 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*, vol. 23, no. 6, pp. 917–940, 2022, doi: 10.1007/s10198-021-01347-4.

[15] N. Mohamed, M. H. Ahmad, Z. Ismail, and 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. Atesongun and M. Gulsen, "A Hybrid Forecasting Structure Based on Arima and Artificial Neural Network Models," *Applied Sciences (Switzerland)*, vol. 14, no. 16, Aug. 2024, doi: 10.3390/app14167122.

[17] N. Suhermi, Suhartono, D. D. Prastyo, and 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, and 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, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/fi15080255.

[20] T. T. K. Tran, S. M. Bateni, S. J. Ki, and H. Vosoughifar, "A review of neural networks for air temperature forecasting," May 01, 2021, *MDPI AG*. doi: 10.3390/w13091294.

[21] W. Ge, P. Lalbakhsh, L. Isai, A. Lenskiy, and H. Suominen, "Neural Network-Based Financial Volatility Forecasting: A Systematic Review," Jan. 17, 2022, *Association for Computing Machinery*. doi: 10.1145/3483596.

[22] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.

[23] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K. W. Chau, "Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting," *Water (Switzerland)*, vol. 12, no. 5, May 2020, doi: 10.3390/w12051500.

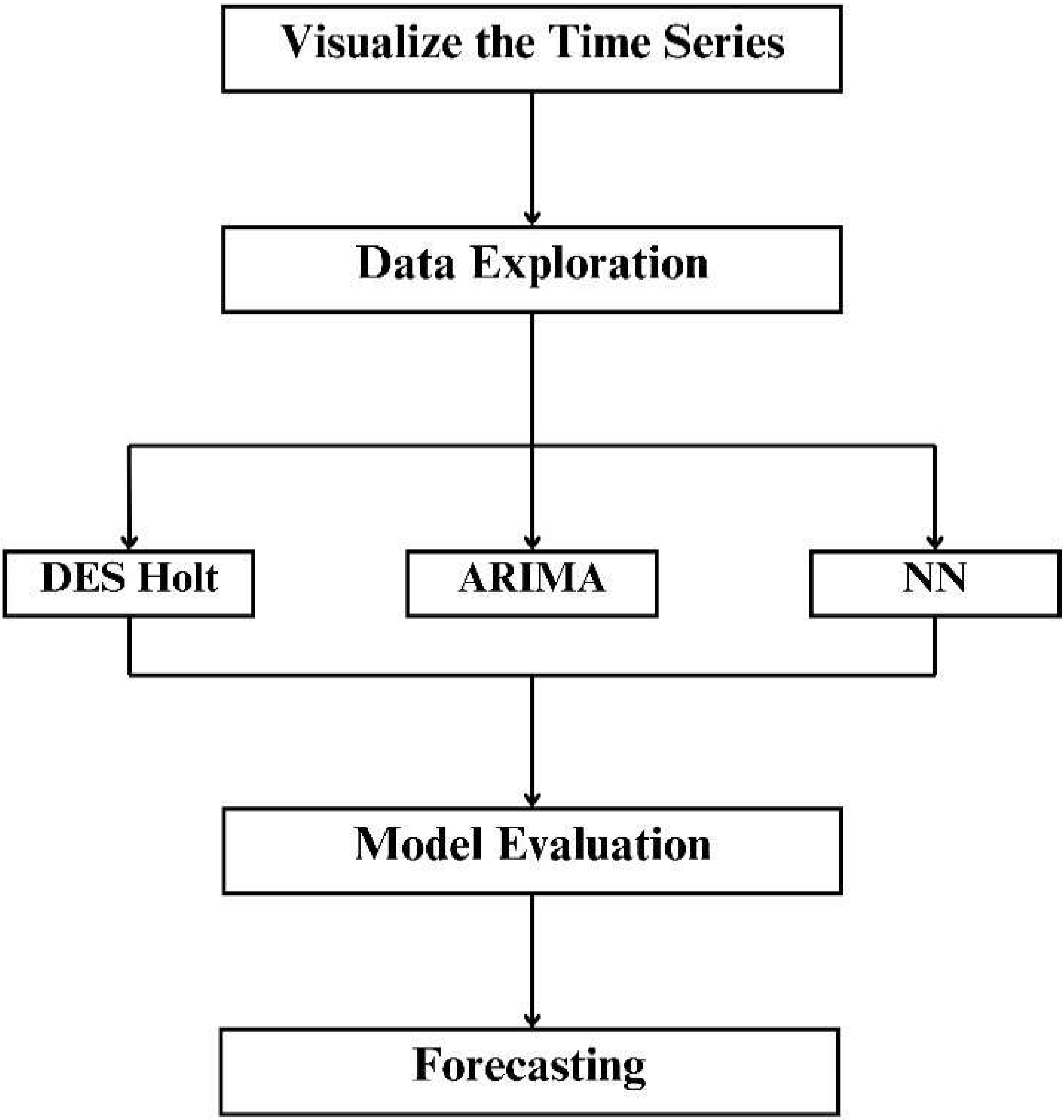
[24] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Syst Appl*, vol. 37, no. 1, pp. 479–489, Jan. 2010, doi: 10.1016/j.eswa.2009.05.044.

[25] S. Belhaiza and S. Al-Abdallah, "A Neural Network Forecasting Approach for the Smart Grid Demand Response Management Problem," *Energies* , vol. 17, no. 10, May 2024, doi: 10.3390/en17102329.

[26] A. T. R. Dani, M. Fauziyah, H. Sandariria, and Q. Q. A'yun, "Forecasting The Search Trends of The Keyword 'Sarung Wadimor' In Indonesia on Google Trends Data Using Time Series Regression with Calender Variation and Arima

Box-Jenkins," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 3, pp. 447–459, May 2023, doi: 10.20956/j.v19i3.24551.
[27] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan, and F. S. Al-Anzi, "On the accuracy of ARIMA based prediction of COVID-19 spread," *Results Phys*, vol. 27, p. 104509, 2021, doi: 10.1016/j.rinp.2021.104509.
[28] I. Mado, A. Soeprijanto, and Suhartono, "Applying of double seasonal arima model for electrical power demand forecasting at pt. pln gresik Indonesia," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 6, pp. 4892–4901, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4892-4901.
[29] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Appl Energy*, vol. 304, Dec. 2021, 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, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Appl Energy*, vol. 304, Dec. 2021, doi: 10.1016/j.apenergy.2021.117766.
[32] M. Adya and F. Collopy, "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation," *J Forecast*, vol. 17, pp. 481–495, 1998.
[33] 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.
[34] W. Sulandari, Subanar, M. H. Lee, and P. C. Rodrigues, "Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks," *Energy*, vol. 190, Jan. 2020, doi: 10.1016/j.energy.2019.116408.

Reminder: Before you submit, please delete all the instructional text in blue, wherever it appears – including here.
Thank you!


Visualize the Time Series
Data Exploration
DES Holt
ARIMA
NN
Model Evaluation
Forecasting

Gmail Andrea T R Dani <andrikadoko@gmail.com>

Your PDF has been built and requires approval

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 10 Oktober 2024 pukul 20.38
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

MethodsX
Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
Authors: Mislan Mislan; Andrea Tri Rian Dani

Dear Mislan,

The PDF for your submission, "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" has now been built and is ready for your approval. Please view the submission before approving it, to be certain that it is free of any errors. If you have already approved the PDF of your submission, this e-mail can be ignored.

To approve the PDF please login to the Editorial Manager as an Author:

https://www.editorialmanager.com/mex/
Your username is: mislanjoss

Then click on the folder 'Submissions Waiting for Author's Approval' to view and approve the PDF of your submission. You may need to click on 'Action Links' to expand your Action Links menu.

You will also need to confirm that you have read and agree with the Elsevier Ethics in Publishing statement before the submission process can be completed. Once all of the above steps are done, you will receive an e-mail confirming receipt of your submission from the Editorial Office. For further information or if you have trouble completing these steps please go to: http://help.elsevier.com/app/answers/detail/a_id/88/p/7923.

Please note that you are required to ensure everything appears appropriately in PDF and no change can be made after approving a submission. If you have any trouble with the generated PDF or completing these steps please go to: http://help.elsevier.com/app/answers/detail/a_id/88/p/7923.

Your submission will be given a reference number once an Editor has been assigned to handle it.

Thank you for your time and patience.
Kind regards,
Editorial Office
MethodsX

****
For further assistance, please visit our customer support site at http://help.elsevier.com/app/answers/list/p/7923. Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EM via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

Gmail Andrea T R Dani <andrikadoko@gmail.com>

Track the status of your submission to MethodsX

1 pesan

Track your Elsevier submission <no-reply@submissions.elsevier.com> 14 Oktober 2024 pukul 20.03
Kepada: andrikadoko@gmail.com

Manuscript Number: MEX-D-24-00981
Manuscript Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
Journal: MethodsX

Dear Mislan Mislan,

Your submitted manuscript is currently under review. You can track the status of your submission in Editorial Manager, or track the review status in more detail using Track your submission here:
https://track.authorhub.elsevier.com?uuid=5ffec470-a084-4828-92f7-424b7abccb82

This page will remain active until the peer review process for your submission is completed. You can visit the page whenever you like to check the progress of your submission. The page does not require a login, so you can also share the link with your co-authors.

If you are a WeChat user, then you can also receive status updates via WeChat. To do this please click the following link; you will be taken to Elsevier China's website where further instructions will guide you on how to give permission to have your submission's details made visible in WeChat. Note that by clicking the link no submission data is transferred to the WeChat platform. If you have any questions about using Track your submission with WeChat please visit 在线咨询 https://cn.service.elsevier.com/app/chat/chat_launch/supporthub/publishing/session/ - Journal Article Publishing 支持中心

https://webapps.elsevier.cn/st-wechat/subscribe?signature=1728907292-44c4966e69569b7fd46546babcf26c48&uuid=5ffec470-a084-4828-92f7-424b7abccb82

We hope you find this service useful.

Kind regards,
Journal Office of MethodsX
Elsevier B.V.

Gmail

Andrea T R Dani <andrikadoko@gmail.com>

Regarding your Manuscript MEX-D-24-00981

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 23 Oktober 2024 pukul 06.39
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

CC: "Andrea Tri Rian Dani" andreatririandani@fmipa.unmul.ac.id

Ref.: **MEX-D-24-00981**
Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
MethodsX

Dear Mislan,

We have received the reviewers` comments on your manuscript.

If you can suitably address their comments, below, I invite you to submit a revised version of your manuscript, for consideration.

Please carefully address all the issues raised in the comments. Please resubmit your revised manuscript by **Nov 12, 2024**.

To submit your revision, please go to https://www.editorialmanager.com/mex/ and log in to your Author account. You will find your paper in a folder entitled "Submission Needing Revision".

If you do not wish to submit a revision, please 'Decline to Revise' within this folder or let me know by response to this email.

If you are submitting a revised manuscript please also:
a) highlight any change in the text using the "Track Changes" function, and provide a point by point outline of the revisions, following the reviewers` comments
AND/OR
b) provide a suitable rebuttal to each reviewer comment not addressed in the text.

Include interactive data visualizations in your publication and let your readers interact and engage more closely with your research. Follow the instructions here: https://www.elsevier.com/authors/author-services/data-visualization to find out about available data visualization options and how to include them with your article.

I look forward to receiving your revised manuscript.

Yours sincerely,

Damià Barceló
Editor-in-Chief
MethodsX

****

Comments from the editors and reviewers:

Reviewer #1: I reviewed the manuscript titled "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" which aims to compare traditional forecasting models (Exponential Smoothing and ARIMA) and a machine learning model (Neural Network) in predicting rainfall in Samarinda, Indonesia, using monthly data from 2000-2020. The research seeks to identify the best model for forecasting future rainfall trends to improve early warnings and disaster mitigation strategies. Some comments are provided below:
-The study briefly mentions the choice of Exponential Smoothing, ARIMA, and Neural Networks, but it lacks a detailed justification for selecting these specific methods over others. A stronger rationale for why these models were chosen, compared to other possible alternatives like SARIMA or other machine learning models (e.g., Random Forest), would strengthen the methodology.
-While the study discusses using 90% of the data for training and 10% for testing, it does not provide enough information on how missing data, outliers, or potential seasonal components in rainfall were handled during preprocessing. Addressing these issues would enhance the clarity and reliability of the results.
-The description of the Neural Network architecture (2 hidden layers with 6-5 neurons) lacks a clear explanation of how this specific configuration was chosen. A discussion on how different architectures were tested and the impact of tuning hyperparameters would add depth to the model development process.
-The forecasting results indicate that the highest rainfall is predicted for January, but the paper does not sufficiently explain the practical implications of these results. More emphasis on how these predictions can directly impact disaster mitigation or urban planning would better highlight the real-world application of the study.
-Although the paper includes graphical representations, the visualizations could be improved for clarity.
-The paper does not address any limitations of the study, such as potential model overfitting, reliance on historical data that may not account for future climatic changes, or the limited generalizability of the findings to other regions. A section outlining these constraints would provide a more balanced view of the research.

Reviewer #2: In general, this article is good, it needs a few additions or improvements to make it better. Please check in attached file
Check again the english for grammatical and typos.
In conclusion discuss more about the results and add future development of this research.

* Abstract (very good)
* Graphical Abstract (acceptable)
* Methods (acceptable)
* References (very good)

At Elsevier, we want to help all our authors to stay safe when publishing. Please be aware of fraudulent messages requesting money in return for the publication of your paper. If you are publishing open access with Elsevier, bear in mind that we will never request payment before the paper has been accepted. We have prepared some guidelines (https://www.elsevier.com/connect/authors-update/seven-top-tips-on-stopping-apc-scams) that you may find helpful, including a short video on Identifying fake acceptance letters (https://www.youtube.com/watch?v=o5l8thD9XtE). Please remember that you can contact Elsevier s Researcher Support team (https://service.elsevier.com/app/home/supporthub/publishing/) at any time if you have questions about your manuscript, and you can log into Editorial Manager to check the status of your manuscript (https://service.elsevier.com/app/answers/detail/a_id/29155/c/10530/supporthub/publishing/kw/status/).

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

Gmail Andrea T R Dani <andrikadoko@gmail.com>

Your PDF has been built and requires approval

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 26 Oktober 2024 pukul 21.33
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

MethodsX
Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
Authors: Mislan Mislan; Andrea Tri Rian Dani

Dear Mislan,

The PDF for your submission, "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" has now been built and is ready for your approval. Please view the submission before approving it, to be certain that it is free of any errors. If you have already approved the PDF of your submission, this e-mail can be ignored.

To approve the PDF please login to the Editorial Manager as an Author:

https://www.editorialmanager.com/mex/
Your username is: mislanjoss

Then click on the folder 'Submissions Waiting for Author's Approval' to view and approve the PDF of your submission. You may need to click on 'Action Links' to expand your Action Links menu.

You will also need to confirm that you have read and agree with the Elsevier Ethics in Publishing statement before the submission process can be completed. Once all of the above steps are done, you will receive an e-mail confirming receipt of your submission from the Editorial Office. For further information or if you have trouble completing these steps please go to: http://help.elsevier.com/app/answers/detail/a_id/88/p/7923.

Please note that you are required to ensure everything appears appropriately in PDF and no change can be made after approving a submission. If you have any trouble with the generated PDF or completing these steps please go to: http://help.elsevier.com/app/answers/detail/a_id/88/p/7923.

Your submission will be given a reference number once an Editor has been assigned to handle it.

Thank you for your time and patience.
Kind regards,
Editorial Office
MethodsX

****
For further assistance, please visit our customer support site at http://help.elsevier.com/app/answers/list/p/7923. Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EM via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

MethodsX **methods article template Version 6 (April 2024)**

- -

Article information

Article title
Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models.

Authors
Mislan, Andrea Tri Rian Dani*

Affiliations
Department of Physics, Faculty of Mathematics and Natural Science, Mulawarman University; Department of Mathematics, Faculty of Mathematics and Natural Science, Mulawarman University

Corresponding author's email address and Twitter handle
mislan.jos@fmipa.unmul.ac.id

Keywords
Exponential Smoothing; ARIMA; Neural Network; Time Series Modeling; Forecasting

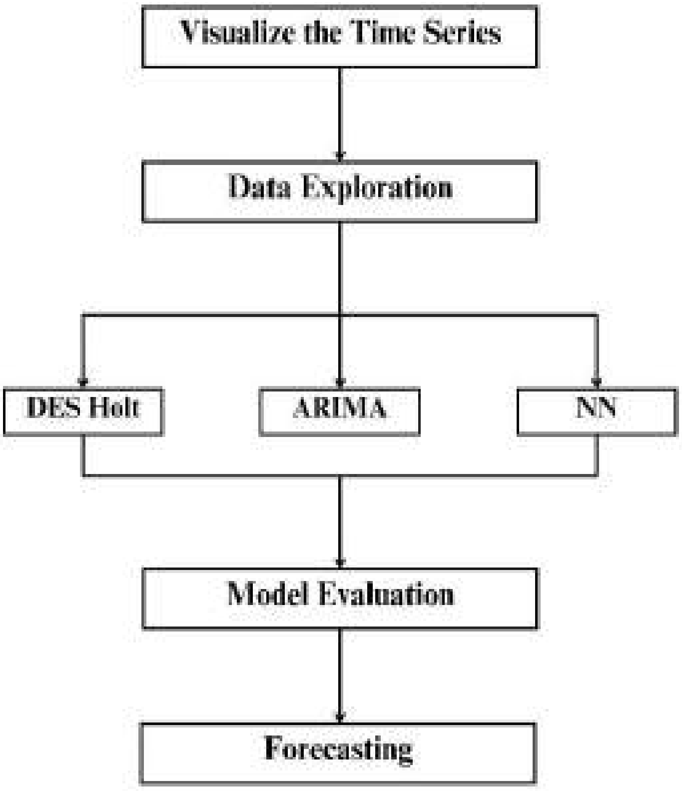
Related research article
None

For a published article:
None

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 drought.*

Graphical abstract



The research design used was ex post facto, meaning data was collected after all the events. The stages of data analysis modeling rainfall data in Samarinda City are visualized in the Graphical Abstract. The researchers chose the three methods based on their advantages and flexibility in the modeling process. The modeling process uses R software.

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.* *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.*
Resource availability	*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 [10], [11], [12], 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], [16], [17], [18], [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 may miss. Several time series data studies have successfully utilized NN are [16], [21], [22], [23], [24], [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

A. *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

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

Trend smoothing

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

With

$$F_{t+m} = L_t + T_t m \quad (3)$$

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

$$L_1 = Z_1 \text{ and } T_1 = Z_2 - Z_1 \quad (4)$$

Where:

α : *level smoothing parameter, 0 < α < 1*

β : *trend smoothing parameter, 0 < α < 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)*

B. 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).

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \quad (5)$$

With:

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$: *backshift operator(B) AR process*

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$: *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:

1. *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.*

2. *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:*

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)*

Process	ACF	PACF
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	Nothings is out of bounds

C. 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 generalize 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 Figure 1.

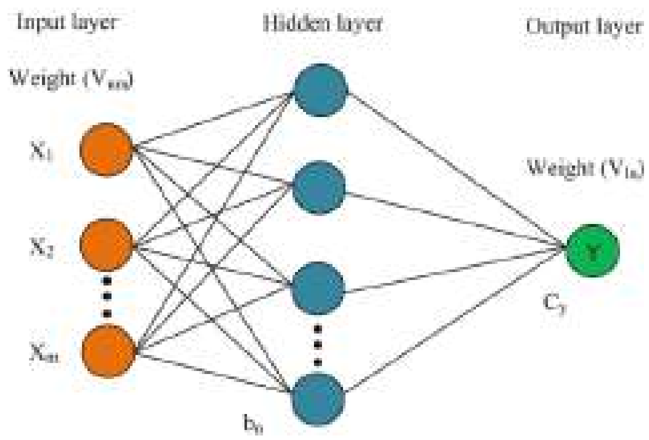


Fig. 1. Neural Network Structure

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]. 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).

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

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

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

D. 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).

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

E. 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 Figure 2.

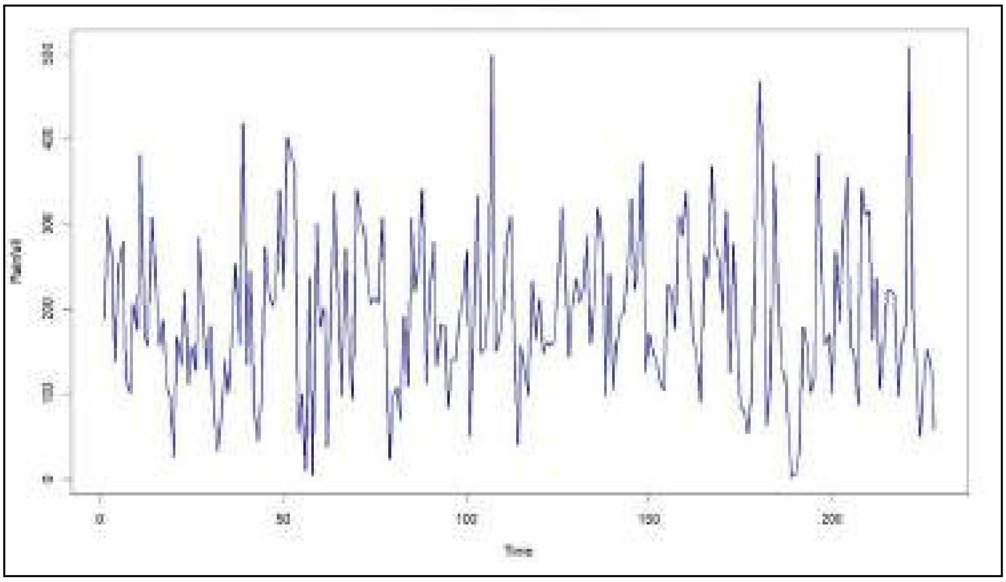


Fig. 2. Time series plot of rainfall data in Samarinda

Based on Figure 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

a. 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 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.*

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	115.28	0.7	0.1	114.66
0.1	0.2	168.88	0.4	0.2	114.72	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	120.05	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	138.89	0.4	0.6	125.13	0.7	0.6	133.49
0.1	0.7	140.94	0.4	0.7	127.14	0.7	0.7	137.81
0.1	0.8	143.61	0.4	0.8	129.14	0.7	0.8	142.23
0.1	0.9	146.54	0.4	0.9	131.40	0.7	0.9	146.69
0.2	0.1	138.03	0.5	0.1	113.38	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	124.69	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	130.18	0.5	0.6	126.01	0.8	0.6	138.95
0.2	0.7	133.68	0.5	0.7	129.02	0.8	0.7	143.98
0.2	0.8	138.00	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	120.05
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	129.47
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	125.59	0.9	0.5	140.34
0.3	0.6	127.90	0.6	0.6	129.19	0.9	0.6	146.16
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	136.90	0.9	0.8	158.77
0.3	0.9	133.12	0.6	0.9	140.91	0.9	0.9	165.67

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. 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	113.40	121.13

α = 0.7 and β = 0.1	114.66	118.03

b. 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 Figures 3 and 4.

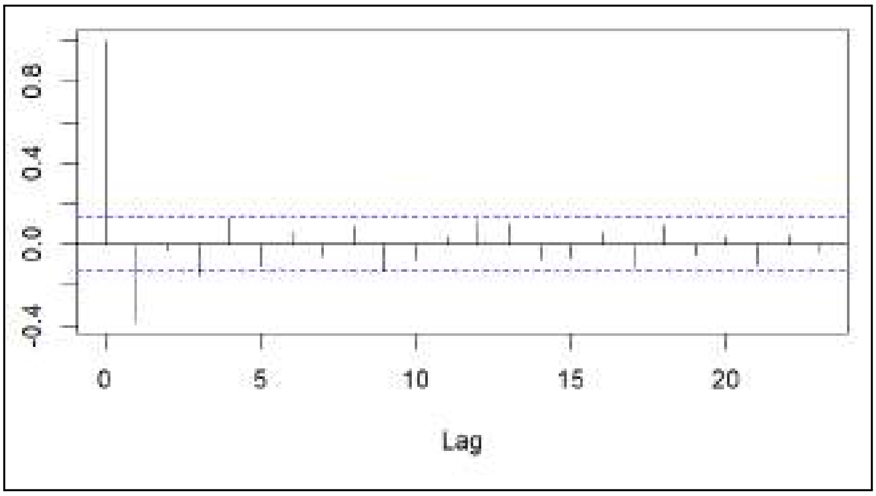


Fig. 3. ACF Plot of Rainfall Data Results of Differencing

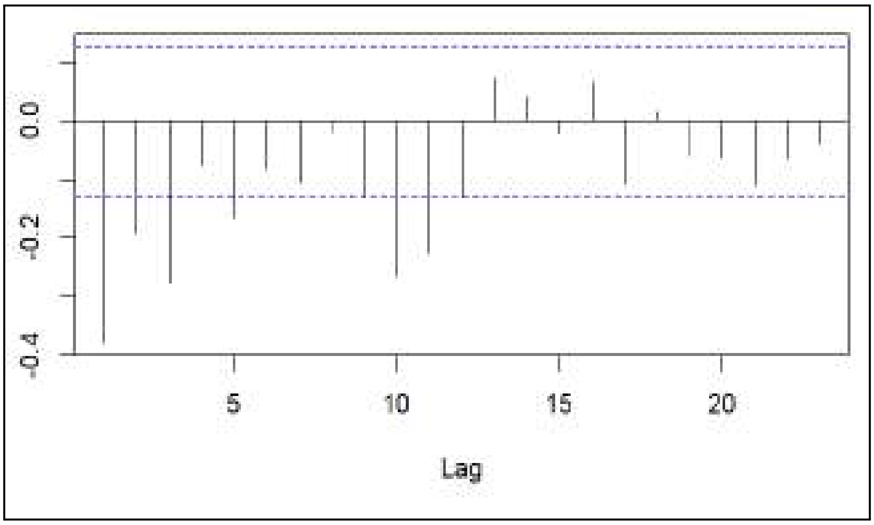


Fig. 4. PACF Plot of Rainfall Data Results of Differencing

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).

Table 5. Temporary ARIMA Model Parameter Estimation

Model	Parameter	Estimate	*p-value*	Conclusion
ARIMA(1,1,0)	$\hat{\phi}_1$	-0.380766	5.579e-10	Significant

Model	Parameter	Estimate	*p-value*	Conclusion
ARIMA(2,1,0)	$\hat{\phi}_1$	-0.455156	2.982e-12	Significant
	$\hat{\phi}_2$	-0.193967	0.002867	
ARIMA(3,1,0)	$\hat{\phi}_1$	-0.507994	1,760e-15	Significant
	$\hat{\phi}_2$	-0.320263	3.421e-06	
	$\hat{\phi}_3$	-0.275761	1.507e-05	
ARIMA(0,1,1)	$\hat{\theta}_1$	-1,000000	< 2.2e-16	Significant
ARIMA(1,1,1)	$\hat{\phi}_1$	0.273909	2.071e-05	Significant
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(2,1,1,)	$\hat{\phi}_1$	0.266500	6.351e-05	Not Significant
	$\hat{\phi}_2$	0.028447	0.6701	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(3,1,1)	$\hat{\phi}_1$	0.268049	5.375e-05	Not Significant
	$\hat{\phi}_2$	0.048998	0.475	
	$\hat{\phi}_3$	-0.081457	0.220	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	

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.

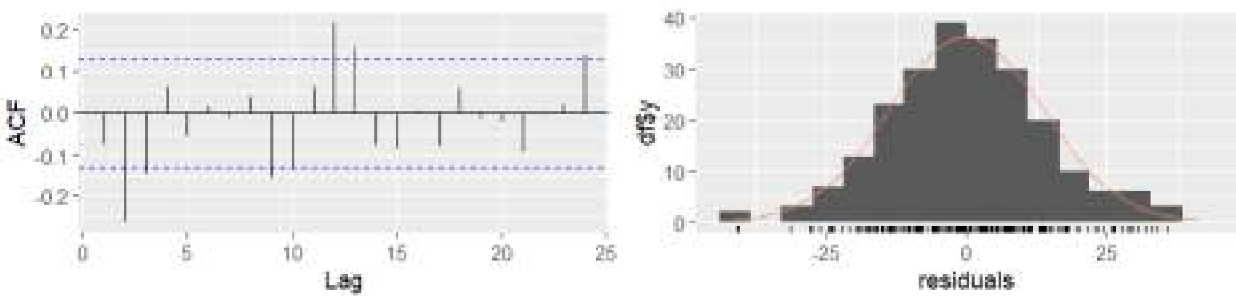


Fig. 5. Independence and Normality of ARIMA Residuals (1,1,0)

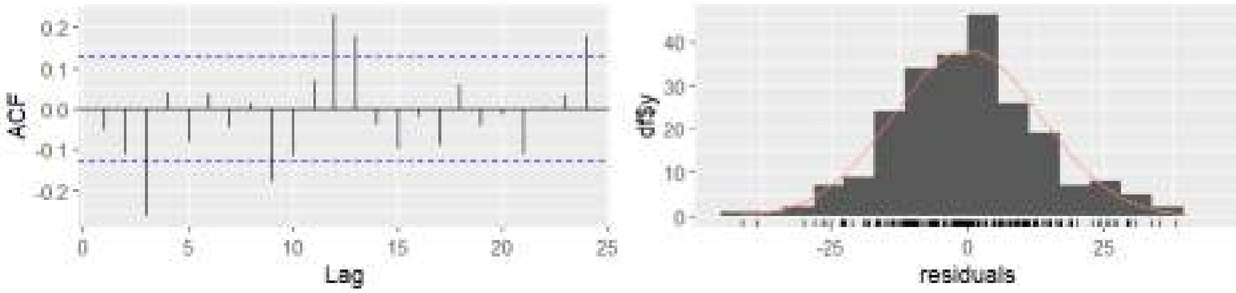


Fig. 6. Independence and Normality of ARIMA Residuals (2,1,0)

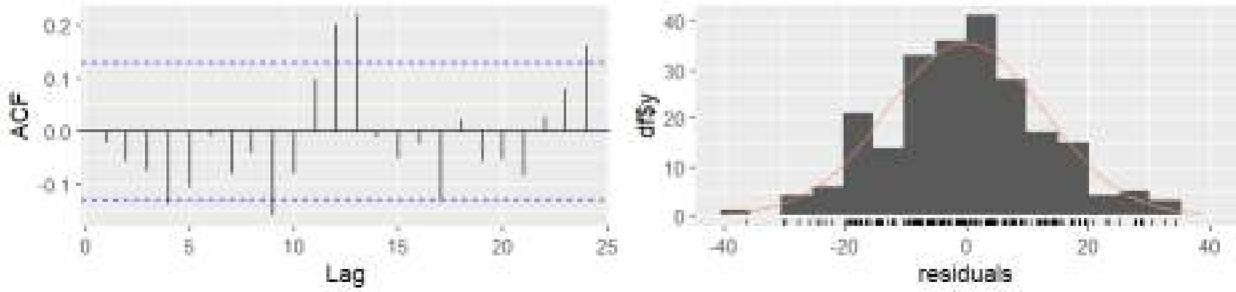


Fig. 7. Independence and Normality of ARIMA Residuals (3,1,0)

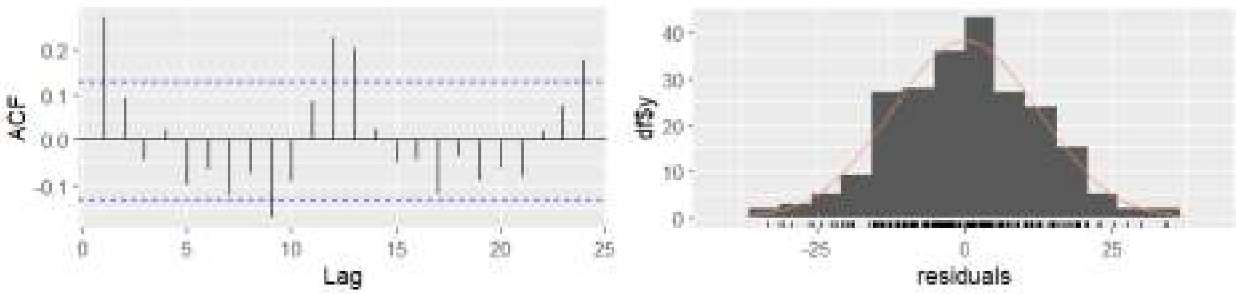


Fig. 8. Independence and Normality of ARIMA Residuals (0,1,1)

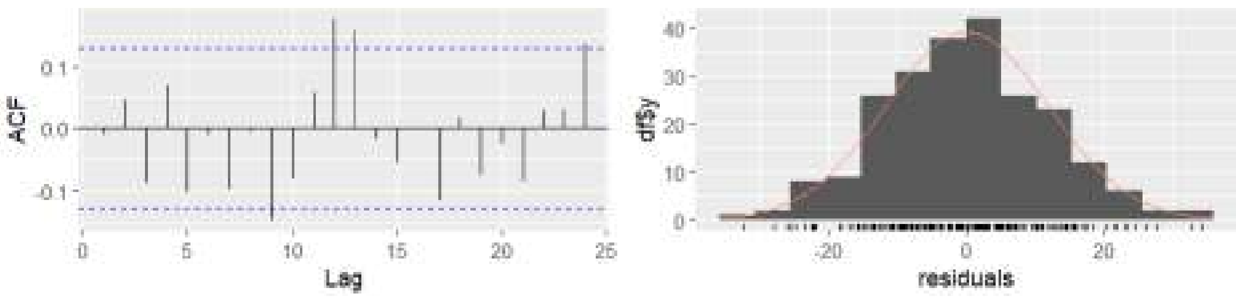


Fig. 9. Independence and Normality of ARIMA Residuals (1,1,1)

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.

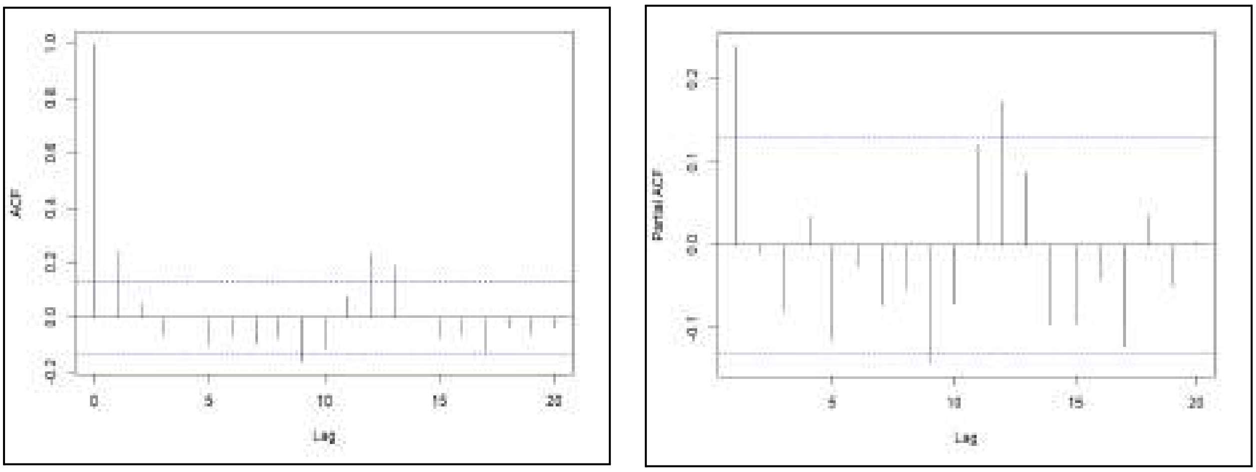
Table 6. Accuracy of Training and Testing Data for ARIMA Model (1,1,1)

Proportion	RMSEP
Data Training	92.75
Data Testing	80.82

c. *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 Figure 10.



(a) ACF graph

(b) PACF graph

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

Based on Figure 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.

- *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.

- *Best Model Selection in NN*

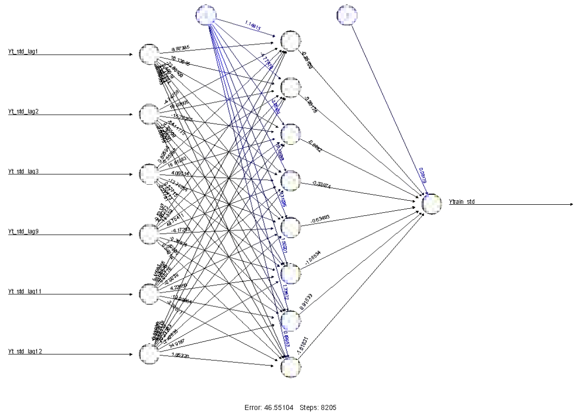
The backpropagation training process is carried out by adjusting the NN architecture. In this study, two types of architectures are used, namely networks with one hidden layer and networks with two hidden layers where each architecture will try various combinations of the number of neurons in each layer. 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.

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 rate 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 7.

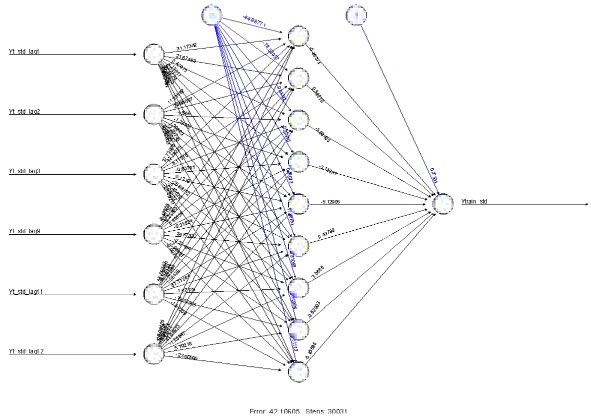
Table 7. RMSE Calculation Rsults

Hidden Layer		RMSEP		Difference between Training and Testing RMSEP
Hidden Layer1	**Hidden Layer2**	**Training**	**Testing**	
1 neuron		88.988	86.595	2.393
2 neurons		85.018	90.197	5.180
3 neurons		78.101	110.632	32.531
4 neurons		75.907	100.158	24.251
5 neurons		66.340	97.531	31.192
6 neurons		68.059	96.647	28.589
7 neurons		69.542	103.764	34.222
8 neurons		58.299	102.938	44.639
9 neurons		57.471	114.915	57.444
10 neurons		46.560	110.919	64.359
2 neurons	1 neuron	83.746	82.969	0.777
3 neurons	2 neurons	77.546	100.939	23.393
4 neurons	3 neurons	70.852	97.184	26.332
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	135.889	106.194
9 neurons	8 neurons	27.501	116.387	88.886
10 neurons	9 neurons	17.891	181.869	163.978

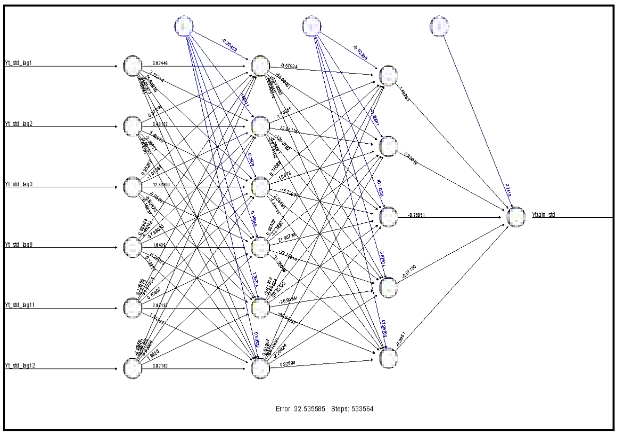
By considering various considerations such as choosing the smallest RMSEP value and the difference between the RMSEP values of the training data and testing data is not very significant, then based on Table 7, it is found that the NN model with a 2 hidden layer architecture model (6-5 neurons) is the 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 89.749. Some of the architectural results of the NN modeling can be seen in Figure 11.



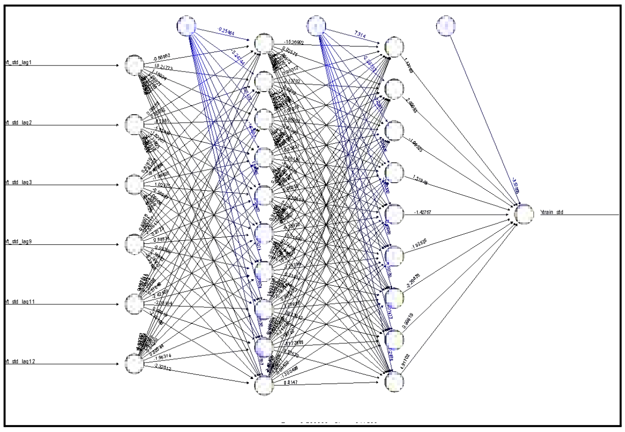
(a) NN model 1 hidden layer (8 neurons)



(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

d. 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 8 displays the RMSEP values of the best models. The model with the smallest RMSE value will be selected as the best method.

Table 8. Model Goodness of Fit Measure

Method	**RMSEP**	
	Training	Testing
DES Holt (α = 0.7 and β = 0.1)	114.66	118.03
ARIMA (1,1,1)	92.75	80.82
NN 2 HL (6-5 neurons)	**52.473**	**89.749**

In Table 8, 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.

e. 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.

Table 9. Forecasting Results of 12 Periods of NN 2 HL (6-5 neurons)

Month	Prediction Results
January	301.935
February	245.819
March	168.822
April	82.964
May	173.226
June	209.721
July	107.147
August	204.006
September	201.430
October	133.406
November	207.887
December	206,919

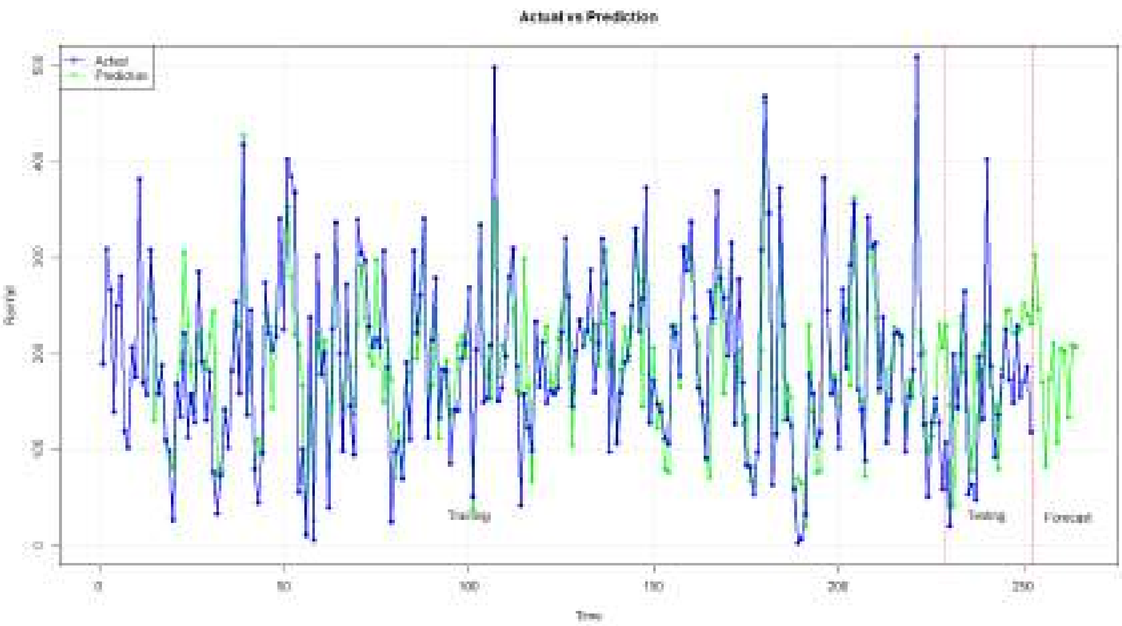


Fig. 12. Comparison plot of actual and predicted data

Figure 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.

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, 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.

CRediT author statement

Mislan: Conceptualization, Methodology, Validity tests, Writing-Preparation of the first draft, and Supervision. **Andrea Tri Rian Dani**: Data curation, Analysis, Visualization, and Editing Draft, and writing original draft.

Acknowledgments

None

Declaration of interests

*Please **tick** the appropriate statement below (please do not delete either statement) and declare any financial interests/personal relationships which may affect your work in the box below.*

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Please declare any financial interests/personal relationships which may be considered as potential competing interests here.

Supplementary material *and/or* additional information [OPTIONAL]

None.

References

[1] M. Putra, M. S. Rosid, and D. Handoko, “A Review of Rainfall Estimation in Indonesia: Data Sources, Techniques, and Methods,” *Signals*, vol. 5, no. 3, pp. 542–561, Aug. 2024, doi: 10.3390/signals5030030.

[2] 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.

[3] R. P. Permata, R. Ni'mah, and A. T. R. Dani, "Daily Rainfall Forecasting with ARIMA Exogenous Variables and Support Vector Regression," *Jurnal Varian*, vol. 7, no. 2, pp. 177–188, Jun. 2024, doi: 10.30812/varian.v7i2.3202.

[4] M. Mislan and A. T. R. Dani, "Forecasting Maximum Water Level Data for Post Sangkuliman using An Artificial Neural Network Backpropagation Algorithm," *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 8, no. 2, p. 465, Mar. 2024, doi: 10.31764/jtam.v8i2.20112.

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

[6] S. Makridakis, R. J. Hyndman, and F. Petropoulos, "Forecasting in social settings: The state of the art," *Int J Forecast*, vol. 36, no. 1, pp. 15–28, Jan. 2020, doi: 10.1016/j.ijforecast.2019.05.011.

[7] S. Makridakis, E. Spiliotis, V. Assimakopoulos, and 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, and M. S. El-Sherif, "A Comparison Between Neural-Network Forecasting Techniques-Case Study: River Flow Forecasting," 1999.

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

[10] Mustaruddin, G. Puspito, B. Z. Zanlee, and T. Nugroho, "The exponential smoothing model for squid production in Muara Angke Fishing Port, Jakarta," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2024. doi: 10.1088/1755-1315/1359/1/012098.

[11] G. Airlangga, A. Rachmat, and D. Lapihu, "Comparison of exponential smoothing and neural network method to forecast rice production in Indonesia," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 3, pp. 1367–1375, 2019, 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)*, vol. 14, no. 13, Jul. 2024, doi: 10.3390/app14135846.

[13] D. Zhao, R. Zhang, H. Zhang, and S. He, "Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models," *Sci Rep*, vol. 12, no. 1, pp. 1–13, 2022, 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*, vol. 23, no. 6, pp. 917–940, 2022, doi: 10.1007/s10198-021-01347-4.

[15] N. Mohamed, M. H. Ahmad, Z. Ismail, and 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. Atesongun and M. Gulsen, "A Hybrid Forecasting Structure Based on Arima and Artificial Neural Network Models," *Applied Sciences (Switzerland)*, vol. 14, no. 16, Aug. 2024, doi: 10.3390/app14167122.

[17] N. Suhermi, Suhartono, D. D. Prastyo, and 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, and 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, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/fi15080255.

[20] T. T. K. Tran, S. M. Bateni, S. J. Ki, and H. Vosoughifar, "A review of neural networks for air temperature forecasting," May 01, 2021, *MDPI AG*. doi: 10.3390/w13091294.

[21] W. Ge, P. Lalbakhsh, L. Isai, A. Lenskiy, and H. Suominen, "Neural Network-Based Financial Volatility Forecasting: A Systematic Review," Jan. 17, 2022, *Association for Computing Machinery*. doi: 10.1145/3483596.

[22] I. D. Mienye, T. G. Swart, and G. Obaido, “Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications,” *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.

[23] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K. W. Chau, “Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting,” *Water (Switzerland)*, vol. 12, no. 5, May 2020, doi: 10.3390/w12051500.

[24] M. Khashei and M. Bijari, “An artificial neural network (p, d, q) model for timeseries forecasting,” *Expert Syst Appl*, vol. 37, no. 1, pp. 479–489, Jan. 2010, doi: 10.1016/j.eswa.2009.05.044.

[25] S. Belhaiza and S. Al-Abdallah, “A Neural Network Forecasting Approach for the Smart Grid Demand Response Management Problem,” *Energies* , vol. 17, no. 10, May 2024, doi: 10.3390/en17102329.

[26] A. T. R. Dani, M. Fauziyah, H. Sandariria, and Q. Q. A’yun, “Forecasting The Search Trends of The Keyword ‘Sarung Wadimor’ In Indonesia on Google Trends Data Using Time Series Regression with Calender Variation and Arima Box-Jenkins,” *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 3, pp. 447–459, May 2023, doi: 10.20956/j.v19i3.24551.

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

[28] I. Mado, A. Soeprijanto, and Suhartono, “Applying of double seasonal arima model for electrical power demand forecasting at pt. pln gresik Indonesia,” *International Journal of Electrical and Computer Engineering*, vol. 8, no. 6, pp. 4892–4901, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4892-4901.

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

[32] M. Adya and F. Collopy, “How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation,” *J Forecast*, vol. 17, pp. 481–495, 1998.

[33] 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.

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

MethodsX Submission 1

by dani Andrea

Submission date: 25-Oct-2024 09:38AM (UTC+0700)
Submission ID: 2686128923
File name: Revisi_1-MethodsX-Manuscript_siap_Submitted_2024.docx.pdf (1.61M)
Word count: 6140
Character count: 29682

***MethodsX* methods article template Version 6 (April 2024)**

MethodsX
Open access

Article Template

Article information

Article title
Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models.

Authors
Mislan, Andrea Tri Rian Dani*

Affiliations
Department of Physics, Faculty of Mathematics and Natural Science, Mulawarman University; Department of Mathematics, Faculty of Mathematics and Natural Science, Mulawarman University

Corresponding author's email address and Twitter handle
mislan.jos@fmipa.unmul.ac.id

Keywords
Exponential Smoothing; ARIMA; Neural Network; Time Series Modeling; Forecasting

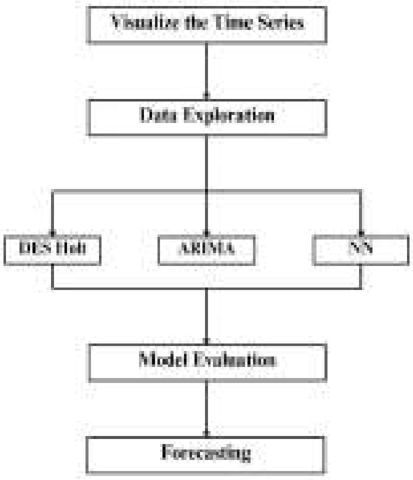
Related research article
None

For a published article:
None

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 drought.*

Graphical abstract



The research design used was ex post facto, meaning data was collected after all the events. The stages of data analysis modeling rainfall data in Samarinda City are visualized in the Graphical Abstract. The researchers chose the three methods based on their advantages and flexibility in the modeling process. The modeling process uses R software.

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.* *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.*
Resource availability	*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*

MethodsX
Open access
Article Template

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 [10], [11], [12]. 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], [16], [17], [18], [19].

Neural Network (NN), a time series model inspired by Artificial Neural Network, 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 may miss. Several time series data studies have successfully utilized NN are [16], [21], [22], [23], [24], [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

A. *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)

B. 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 q 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:

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

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:

1. 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).

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

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

$$\phi_p(B)Z_t=a_t \quad (8)$$

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

2. 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).

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

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

$$Z_t=\theta_q(B)a_t \quad (10)$$

   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:

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)

Process	ACF	PACF
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	Nothings is out of bounds

C. 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 generalize 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 Figure 1.

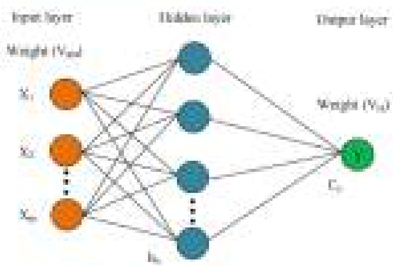


Fig. 1. Neural Network Structure

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]. 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).

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

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

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

D. 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).

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

E. 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 Figure 2.

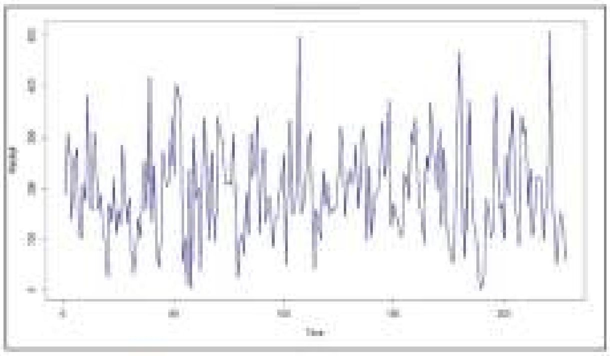


Fig. 2. Time series plot of rainfall data in Samarinda

Based on Figure 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

a. 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 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.

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	115.28	0.7	0.1	114.66
0.1	0.2	168.38	0.4	0.2	114.72	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	120.05	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	138.89	0.4	0.6	125.11	0.7	0.6	133.49
0.1	0.7	140.94	0.4	0.7	127.14	0.7	0.7	137.81
0.1	0.8	143.61	0.4	0.8	129.14	0.7	0.8	142.23
0.1	0.9	146.54	0.4	0.9	131.40	0.7	0.9	146.69
0.2	0.1	138.03	0.5	0.1	113.38	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	124.69	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	130.18	0.5	0.6	126.01	0.8	0.6	138.95
0.2	0.7	133.68	0.5	0.7	129.02	0.8	0.7	143.98
0.2	0.8	138.00	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	120.05
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	129.47
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	125.59	0.9	0.5	140.34
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	136.90	0.9	0.8	158.77
0.3	0.9	133.12	0.6	0.9	140.91	0.9	0.9	165.67

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. Optimal Combination Value of Training and Testing Data

Parameter Value	RMSEP Training	RMSEP Testing
α = 0.4 and β = 0.2	114.72	122.32
α = 0.5 and β = 0.1	113.38	127.50
α = 0.6 and β = 0.1	113.40	121.13

α = 0.7 and β = 0.1		114.66	118.03

B. *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 Figures 3 and 4.

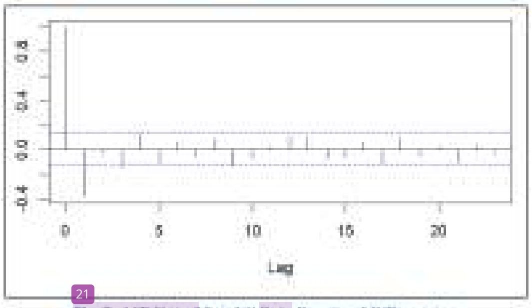


Fig. 3. ACF Plot of Rainfall Data Results of Differencing

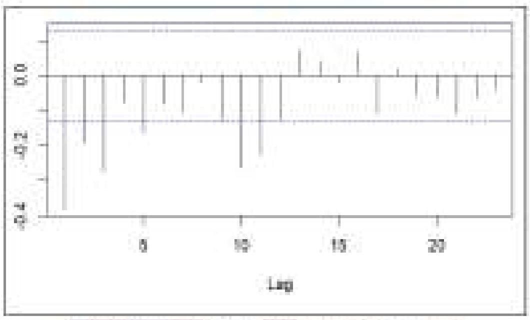


Fig. 4. PACF Plot of Rainfall Data Results of Differencing

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).

Table 5. Temporary ARIMA Model Parameter Estimation

Model	Parameter	Estimate	p-value	Conclusion
ARIMA(1,1,0)	$\hat{\phi}_1$	-0.380766	5.579e-10	Significant

Model	Parameter	Estimate	p-value	Conclusion
ARIMA(2,1,0)	$\hat{\phi}_1$	-0.455156	2.382e-12	Significant
	$\hat{\phi}_2$	-0.193967	0.002867	
ARIMA(3,1,0)	$\hat{\phi}_1$	-0.507994	1.760e-15	Significant
	$\hat{\phi}_2$	-0.320263	3.421e-06	
	$\hat{\phi}_3$	-0.275761	1.507e-05	
ARIMA(0,1,1)	$\hat{\theta}_1$	-1,000000	< 2.2e-16	Significant
ARIMA(1,1,1)	$\hat{\phi}_1$	0.273909	2.071e-05	Significant
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(2,1,1,)	$\hat{\phi}_1$	0.266500	6.351e-05	Not Significant
	$\hat{\phi}_2$	0.028447	0.6701	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(3,1,1)	$\hat{\phi}_1$	0.268049	5.375e-05	Not Significant
	$\hat{\phi}_2$	0.048998	0.475	
	$\hat{\phi}_3$	-0.081457	0.220	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	

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.

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

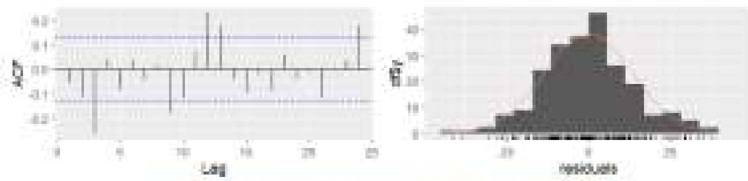


Fig. 6. Independence and Normality of ARIMA Residuals (2,1,0)

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

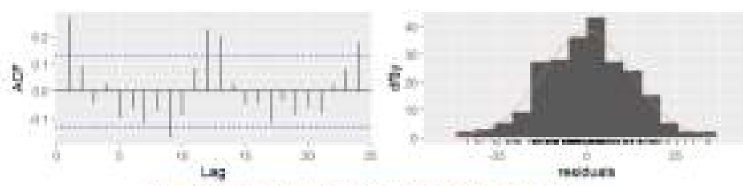


Fig. 8. Independence and Normality of ARIMA Residuals (0,1,1)

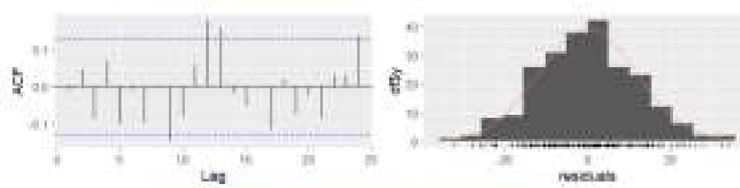


Fig. 9. Independence and Normality of ARIMA Residuals (1,1,1)

56

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.

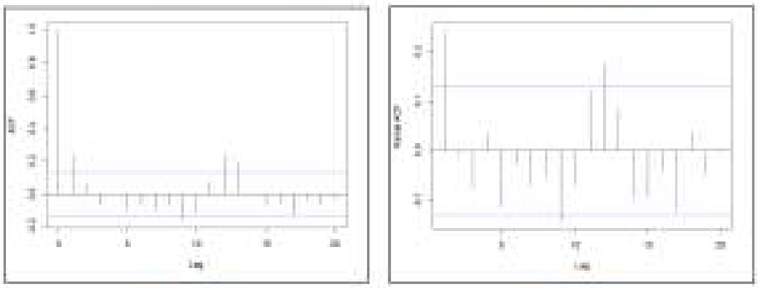
Table 6. Accuracy of Training and Testing Data for ARIMA Model (1,1,1)

Proportion	RMSEP
Data Training	92.75
Data Testing	80.82

c. *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 Figure 10.



(a) ACF graph (b) PACF graph

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

Based on Figure 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.

- *Data Standardization*

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.

- *Best Model Selection in NN*

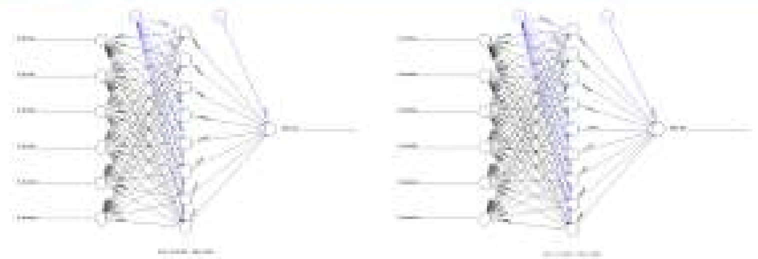
The backpropagation training process is carried out by adjusting the NN architecture. In this study, two types of architectures are used, namely networks with one hidden layer and networks with two hidden layers where each architecture will try various combinations of the number of neurons in each layer. 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.

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 contains 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 rate 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 7.

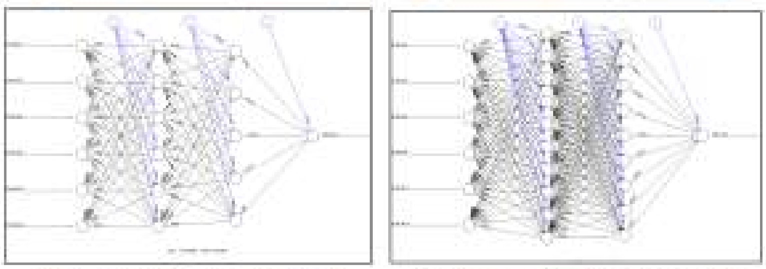
Table 7. RMSE Calculation Results

Hidden Layer		RMSEP		Difference between Training and Testing RMSEP
Hidden Layer1	Hidden Layer2	Training	Testing	
1 neuron		88.988	86.595	2.393
2 neurons		85.018	90.197	5.180
3 neurons		78.101	110.632	32.531
4 neurons		75.907	100.158	24.251
5 neurons		66.340	97.531	31.192
6 neurons		68.058	96.647	28.589
7 neurons		69.542	103.764	34.222
8 neurons		58.299	102.938	44.639
9 neurons		57.471	114.915	57.444
10 neurons		46.560	110.919	64.359
2 neurons	1 neuron	83.746	82.969	0.777
3 neurons	2 neurons	77.546	100.939	23.393
4 neurons	3 neurons	70.852	97.184	26.332
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	135.889	106.194
9 neurons	8 neurons	27.501	116.387	88.886
10 neurons	9 neurons	17.891	181.869	163.978

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 7, 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 89.749. Some of the architectural results of the NN modeling can be seen in Figure 11.



(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

d. *...r 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 8 displays the RMSEP values of the best models. The model with the smallest RMSE value will be selected as the best method.

Table 8. Model Goodness of Fit Measure

Method	RMSEP	
	Training	Testing
DES Holt (α = 0.7 and β = 0.1)	114.66	118.03
ARIMA (1,1,1)	92.75	80.82
NN 2 HL (6-5 neurons)	**52.471**	**89.749**

In Table 8, 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.

e. *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.

Table 9. Forecasting Results of 12 Periods of NN 2 HL (6-5 neurons)

Month	Prediction Results
January	301.935
February	245.819
March	168.822
April	82.964
May	173.226
June	209.721
July	107.147
August	204.006
September	201.430
October	131.406
November	207.887
December	206.919

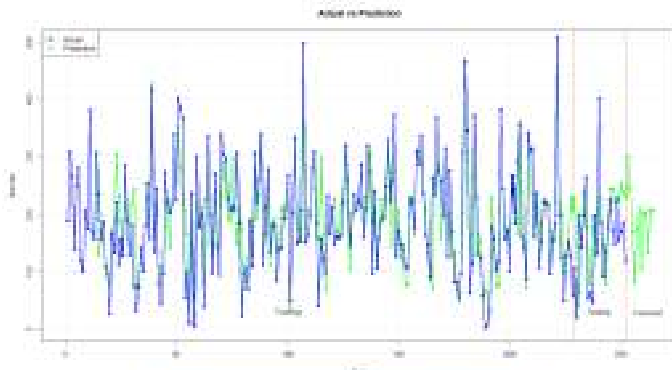


Fig. 12. Comparison plot of actual and predicted data

Figure 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.

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, 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.

CRediT author statement

Misdan: Conceptualization, Methodology, Validity tests, Writing-Preparation of the first draft, and Supervision. **Andrea Tri Rian Dani**: Data curation, Analysis, Visualization, and Editing Draft, and writing original draft.

Acknowledgments

None

Declaration of interests

Please **tick** *the appropriate statement below (please do not delete either statement) and declare any financial interests/personal relationships which may affect your work in the box below.*

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Please declare any financial interests/personal relationships which may be considered as potential competing interests here.

Supplementary material *and/or* additional information [OPTIONAL]

None.

References

[1] M. Putra, M. S. Rosid, and D. Handoko, "A Review of Rainfall Estimation in Indonesia: Data Sources, Techniques, and Methods," *Signals*, vol. 5, no. 3, pp. 542–561, Aug. 2024, doi: 10.3390/signals5030030.

[2] 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.

[3] R. P. Permata, R. Ni'mah, and A. T. R. Dani, "Daily Rainfall Forecasting with ARIMA Exogenous Variables and Support Vector Regression," *Jurnal Varian*, vol. 7, no. 2, pp. 177–188, Jun. 2024, doi: 10.30812/varian.v7i2.3202.

[4] M. Mislan and A. T. R. Dani, "Forecasting Maximum Water Level Data for Post Sangkuliman using An Artificial Neural Network Backpropagation Algorithm," *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 8, no. 2, p. 465, Mar. 2024, doi: 10.31764/jtam.v8i2.20112.

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

[6] S. Makridakis, R. J. Hyndman, and F. Petropoulos, "Forecasting in social settings: The state of the art," *Int J Forecast*, vol. 36, no. 1, pp. 15–28, Jan. 2020, doi: 10.1016/j.ijforecast.2019.05.011.

[7] S. Makridakis, E. Spiliotis, V. Assimakopoulos, and 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, and M. S. El-Sherif, "A Comparison Between Neural-Network Forecasting Techniques-Case Study: River Flow Forecasting," 1999.

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

[10] Mustaruddin, G. Puspito, B. Z. Zaniee, and T. Nugroho, "The exponential smoothing model for squid production in Muara Angke Fishing Port, Jakarta," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2024. doi: 10.1088/1755-1315/1359/1/012098.

[11] G. Airlangga, A. Rachmat, and D. Lapihu, "Comparison of exponential smoothing and neural network method to forecast rice production in Indonesia," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 3, pp. 1367–1375, 2019, 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)*, vol. 14, no. 13, Jul. 2024, doi: 10.3390/app14135846.

[13] D. Zhao, R. Zhang, H. Zhang, and S. He, "Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models," *Sci Rep*, vol. 12, no. 1, pp. 1–13, 2022, 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*, vol. 23, no. 6, pp. 917–940, 2022, doi: 10.1007/s10198-021-01347-4.

[15] N. Mohamed, M. H. Ahmad, Z. Ismail, and 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. Abasongun and M. Gulsen, "A Hybrid Forecasting Structure Based on Arima and Artificial Neural Network Models," *Applied Sciences (Switzerland)*, vol. 14, no. 16, Aug. 2024, doi: 10.3390/app14167122.

[17] N. Suhermi, Suhartono, D. D. Prastyo, and 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, and 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, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/fi15080255.

[20] T. T. K. Tran, S. M. Bateni, S. J. Ki, and H. Vosoughifar, "A review of neural networks for air temperature forecasting," May 01, 2021, *MDPI AG*. doi: 10.3390/w13091294.

[21] W. Ge, P. Lalbakhsh, L. Isai, A. Lenskiy, and H. Suominen, "Neural Network-Based Financial Volatility Forecasting: A Systematic Review," Jan. 17, 2022, *Association for Computing Machinery*. doi: 10.1145/3483596.

[22] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.

[23] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K. W. Chau, "Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting," *Water (Switzerland)*, vol. 12, no. 5, May 2020, doi: 10.3390/w12051500.

[24] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Syst Appl*, vol. 37, no. 1, pp. 479–489, Jan. 2010, doi: 10.1016/j.eswa.2009.05.044.

[25] S. Belhaiza and S. Al-Abdallah, "A Neural Network Forecasting Approach for the Smart Grid Demand Response Management Problem," *Energies* , vol. 17, no. 10, May 2024, doi: 10.3390/en17102329.

[26] A. T. R. Dani, M. Fauziyah, H. Sandariria, and Q. Q. A'yun, "Forecasting The Search Trends of The Keyword 'Sarung Wadimor' in Indonesia on Google Trends Data Using Time Series Regression with Calender Variation and Arima Box-Jenkins," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 3, pp. 447–459, May 2023, doi: 10.20956/j.v19i3.24551.

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

[28] I. Mado, A. Soeprijanto, and Suhartono, "Applying of double seasonal arima model for electrical power demand forecasting at pt. pln gresik Indonesia," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 6, pp. 4892–4901, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4892-4901.

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

[32] M. Adya and F. Collopy, "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation," *J Forecast*, vol. 17, pp. 481–495, 1998.

[33] 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.

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

MethodsX Submission 1

ORIGINALITY REPORT

17%	13%	14%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	sipeg.unj.ac.id Internet Source	1%
2	isroset.org Internet Source	1%
3	M. Imron, Wika Dianita Utami, Hani Khaulasari, Firman Armunanto. "ARIMA MODEL OF OUTLIER DETECTION FOR FORECASTING CONSUMER PRICE INDEX (CPI)", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2022 Publication	1%
4	erepository.uonbi.ac.ke Internet Source	1%
5	cdn.repository.uisi.ac.id Internet Source	1%
6	ijimai.org Internet Source	1%
7	investigacion.unirioja.es Internet Source	<1%
8	jurnal.goretanpena.com Internet Source	<1%
9	journal.uad.ac.id Internet Source	<1%
10	jurnal.lapan.go.id Internet Source	<1%
11	Komi Mensah Agboka, Emmanuel Peter, Erion Bwambale, Bonoukpoè Mawuko Sokame. "Biological reinforcement learning simulation	<1%

for natural enemy -host behavior: Exploring deep learning algorithms for population dynamics", MethodsX, 2024
Publication

12	iopscience.iop.org Internet Source	<1 %
13	www.grafiati.com Internet Source	<1 %
14	www.ssca.org.in Internet Source	<1 %
15	Rakesh N Pillai, Aleena Alex, Narassima M S, Vivek Verma, Ajil Shaji, Keechilat Pavithran, D. K. Vijaykumar, Denny John. "Economic burden of breast cancer prevalence by 2030 in India: Projections using Autoregressive Integrated Moving Average (ARIMA) model", Springer Science and Business Media LLC, 2024 Publication	<1 %
16	jestec.taylors.edu.my Internet Source	<1 %
17	Hoang T. Tran, Dong TL. Tran, Minh T. Nguyen, Thang C. Vu. "Intelligent Mobile Robot for Contagious Disease Treatments in Hospitals", MethodsX, 2024 Publication	<1 %
18	M I Mawardi, W S Winanti, T W Sudinda, K Amru, A A Saraswati, S I Sachoemar, Z Arifin, A Alimin. "Analysis of net-zero emission index for several areas in Indonesia using individual carbon footprint and land use covered", IOP Conference Series: Earth and Environmental Science, 2023 Publication	<1 %
19	ideas.repec.org Internet Source	<1 %
20	"CS2-PC-22_HR", ActEd	

Publication

<1%

21 Sanjita Jaipuria, S.S. Mahapatra. "An improved demand forecasting method to reduce bullwhip effect in supply chains", Expert Systems with Applications, 2014
Publication

<1%

22 journals.plos.org
Internet Source

<1%

23 geekflare.com
Internet Source

<1%

24 M. Kumazawa, Y. Imanishi, Y. Fukao, M. Furumoto, A. Yamamoto. "A theory of spectral analysis based on the characteristic property of a linear dynamic system", Geophysical Journal International, 1990
Publication

<1%

25 Siti Azirah Asmai, Burairah Hussin, Mokhtar Mohd Yusof, Abdul Samad Shibghatullah. "Time Series Prediction Techniques for Estimating Remaining Useful Lifetime of Cutting Tool Failure", International Review on Computers and Software (IRECOS), 2014
Publication

<1%

26 Yao, J.. "A case study on using neural networks to perform technical forecasting of forex", Neurocomputing, 200009
Publication

<1%

27 addi.ehu.es
Internet Source

<1%

28 knm21.unram.ac.id
Internet Source

<1%

29 library.iugaza.edu.ps
Internet Source

<1%

30 unitesi.unive.it
Internet Source

<1%

31	Gao, D.. "Neural networks for event extraction from time series: a back propagation algorithm approach", Future Generation Computer Systems, 200507 Publication	<1 %
32	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 %
33	Kishore Bingi, Ramadevi Bhukya, Venkata Ramana Kasi. "Fractional-Order Activation Functions for Neural Networks", Springer Science and Business Media LLC, 2025 Publication	<1 %
34	Richard B. Miller, Lee N. Johnson. "Advanced Methods in Family Therapy Research - A Focus on Validity and Change", Routledge, 2025 Publication	<1 %
35	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 %
36	bth.diva-portal.org Internet Source	<1 %
37	edu.stats.gov.cn Internet Source	<1 %
38	ijrrjournal.com Internet Source	<1 %
	pdfs.semanticscholar.org	

39	Internet Source	<1%
40	ph02.tci-thaijo.org Internet Source	<1%
41	sigma.yildiz.edu.tr Internet Source	<1%
42	www.accentsjournals.org Internet Source	<1%
43	www.tandfonline.com Internet Source	<1%
44	Elmanani Simamora, Wahyunita Yusardi, Abil Mansyur. "Planting Pattern Modeling Based on Rainfall Prediction Using Backpropagation Artificial Neural Network (Case Study: BMKG Rainfall Data, Deli Serdang Regency)", Journal of Physics: Conference Series, 2021 Publication	<1%
45	Endah Retnowati, Anik Ghufron, Marzuki, Kasiyan, Adi Cilik Pierawan, Ashadi. "Character Education for 21st Century Global Citizens", Routledge, 2018 Publication	<1%
46	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%
47	Yu Li, Dong Wei, Xiannan Peng, Hao Shen, Li Ni, Guoyou Wei. "High-efficiency ammonium recovery and high biomass yield from ultrahigh-ammonium industrial effluent via online-controlled photo-fermentation with thermoacidophilic Galdieria sulphuraria", Algal Research, 2025 Publication	<1%

48	digilib.iain-palangkaraya.ac.id Internet Source	<1 %
49	discovery.researcher.life Internet Source	<1 %
50	erepository.uoeld.ac.ke Internet Source	<1 %
51	hal.science Internet Source	<1 %
52	journal2.um.ac.id Internet Source	<1 %
53	jurnal.unipasby.ac.id Internet Source	<1 %
54	mospace.umsystem.edu Internet Source	<1 %
55	p.pdfkul.com Internet Source	<1 %
56	rumahstudio.com Internet Source	<1 %
57	www.ijmlc.org Internet Source	<1 %
58	www.intechopen.com Internet Source	<1 %
59	"Innovations in Information and Communication Technologies (IICT-2020)", Springer Science and Business Media LLC, 2021 Publication	<1 %
60	Khusnia Nurul Khikmah, Kusman Sadik, Indahwati Indahwati. "TRANSFER FUNCTION AND ARIMA MODEL FOR FORECASTING BI RATE IN INDONESIA", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2023 Publication	<1 %

61	"Precision agriculture 25 (SET TWO VOLUMES)", Walter de Gruyter GmbH, 2025 Publication	<1 %
62	"Swarm, Evolutionary, and Memetic Computing", Springer Science and Business Media LLC, 2013 Publication	<1 %
63	Veronica Livescu, Cameron M. Knapp, George T. Gray, Ramon M. Martinez, Benjamin M. Morrow, Bineh G. Ndefru. "Additively Manufactured Tantalum Microstructures", Materialia, 2018 Publication	<1 %
64	Wenhe Chen, Hanting Zhou, Longsheng Cheng, Min Xia. "Condition Monitoring and Anomaly Detection of Wind Turbines using Temporal Convolutional Informer and Robust Dynamic Mahalanobis Mahalanobis Distance", IEEE Transactions on Instrumentation and Measurement, 2023 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

Gmail

Andrea T R Dani <andrikadoko@gmail.com>

Submission Confirmation for MEX-D-24-00981R1

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 26 Oktober 2024 pukul 21.37
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

Ms. Ref. No.: **MEX-D-24-00981R1**
Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
MethodsX

Dear Mislan,

This message is to acknowledge that we have received your revised manuscript for reconsideration for publication in MethodsX.

You may check the status of your manuscript by logging into the Editorial Manager as an author at https://www.editorialmanager.com/mex/.

Thank you for submitting your work to MethodsX.

Kind regards,

Editorial Manager
MethodsX

****
For further assistance, please visit our customer support site at http://help.elsevier.com/app/answers/list/p/7923. Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EM via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

Gmail Andrea T R Dani <andrikadoko@gmail.com>

Track the status of your submission to MethodsX

1 pesan

Track your Elsevier submission <no-reply@submissions.elsevier.com> 29 Oktober 2024 pukul 21.52
Kepada: andrikadoko@gmail.com

Manuscript Number: MEX-D-24-00981
Manuscript Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
Journal: MethodsX

Dear Mislan Mislan,

Your submitted manuscript is currently under review. You can track the status of your submission in Editorial Manager, or track the review status in more detail using Track your submission here:
https://track.authorhub.elsevier.com?uuid=5ffec470-a084-4828-92f7-424b7abccb82

This page will remain active until the peer review process for your submission is completed. You can visit the page whenever you like to check the progress of your submission. The page does not require a login, so you can also share the link with your co-authors.

If you are a WeChat user, then you can also receive status updates via WeChat. To do this please click the following link; you will be taken to Elsevier China's website where further instructions will guide you on how to give permission to have your submission's details made visible in WeChat. Note that by clicking the link no submission data is transferred to the WeChat platform. If you have any questions about using Track your submission with WeChat please visit 在线咨询 https://cn.service.elsevier.com/app/chat/chat_launch/supporthub/publishing/session/ - Journal Article Publishing 支持中心

https://webapps.elsevier.cn/st-wechat/subscribe?signature=1730209710-04ab10ae9bee6096a65fdce555ed8f0d&uuid=5ffec470-a084-4828-92f7-424b7abccb82

We hope you find this service useful.

Kind regards,
Journal Office of MethodsX
Elsevier B.V.

Gmail

Andrea T R Dani <andrikadoko@gmail.com>

Regarding your Manuscript MEX-D-24-00981R1

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 13 November 2024 pukul 19.24
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

CC: "Andrea Tri Rian Dani" andreatririandani@fmipa.unmul.ac.id

Ref.: **MEX-D-24-00981R1**
Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
MethodsX

Dear Mislan,

We have received the reviewers` comments on your manuscript.

If you can suitably address their comments, below, I invite you to submit a revised version of your manuscript, for consideration.

Please carefully address all the issues raised in the comments. Please resubmit your revised manuscript by **Nov 27, 2024**.

To submit your revision, please go to https://www.editorialmanager.com/mex/ and log in to your Author account. You will find your paper in a folder entitled "Submission Needing Revision".

If you do not wish to submit a revision, please 'Decline to Revise' within this folder or let me know by response to this email.

If you are submitting a revised manuscript please also:
a) highlight any change in the text using the "Track Changes" function, and provide a point by point outline of the revisions, following the reviewers` comments
AND/OR
b) provide a suitable rebuttal to each reviewer comment not addressed in the text.

Include interactive data visualizations in your publication and let your readers interact and engage more closely with your research. Follow the instructions here: https://www.elsevier.com/authors/author-services/data-visualization to find out about available data visualization options and how to include them with your article.

I look forward to receiving your revised manuscript.

Yours sincerely,

Damià Barceló
Editor-in-Chief
MethodsX

****

Comments from the editors and reviewers:

Reviewer #3: This is not a true method article becuase it does not present a new or improved method. It simply compares well-known methods used to predict rainfall (e.g., ARIMA and NN) to conclude that NN perform better in a region of Indonesia. Although the description of methods is Ok and the test to the selected region is also Ok, I cannot recommend approval of this submission becuase is not a method is a comparison of well-known methods that taken together do not form a new method.

Reviewer #4: I have read the revised manuscript with title "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" which has covered the initial comments of the reviewers. To my opinion is well explained method and deserve publication.

At Elsevier, we want to help all our authors to stay safe when publishing. Please be aware of fraudulent messages requesting money in return for the publication of your paper. If you are publishing open access with Elsevier, bear in mind that we will never request payment before the paper has been accepted. We have prepared some guidelines (https://www.elsevier.com/connect/authors-update/seven-top-tips-on-stopping-apc-scams) that you may find helpful, including a short video on Identifying fake acceptance letters (https://www.youtube.com/watch?v=o5l8thD9XtE). Please remember that you can contact Elsevier s Researcher Support team (https://service.elsevier.com/app/home/supporthub/publishing/) at any time if you have questions about your manuscript, and you can log into Editorial Manager to check the status of your manuscript (https://service.elsevier.com/app/answers/detail/a_id/29155/c/10530/supporthub/publishing/kw/status/).

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

MethodsX
Open access

Article Template

***MethodsX* methods article template Version 6 (April 2024)**

- -

MethodsX
Open access
Article Template

Article information

Article title
Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models.

Authors
Mislan, Andrea Tri Rian Dani*

Affiliations
Department of Physics, Faculty of Mathematics and Natural Science, Mulawarman University; Department of Mathematics, Faculty of Mathematics and Natural Science, Mulawarman University

Corresponding author's email address and Twitter handle
mislan.jos@fmipa.unmul.ac.id

Keywords
Exponential Smoothing; ARIMA; Neural Network; Time Series Modeling; Forecasting

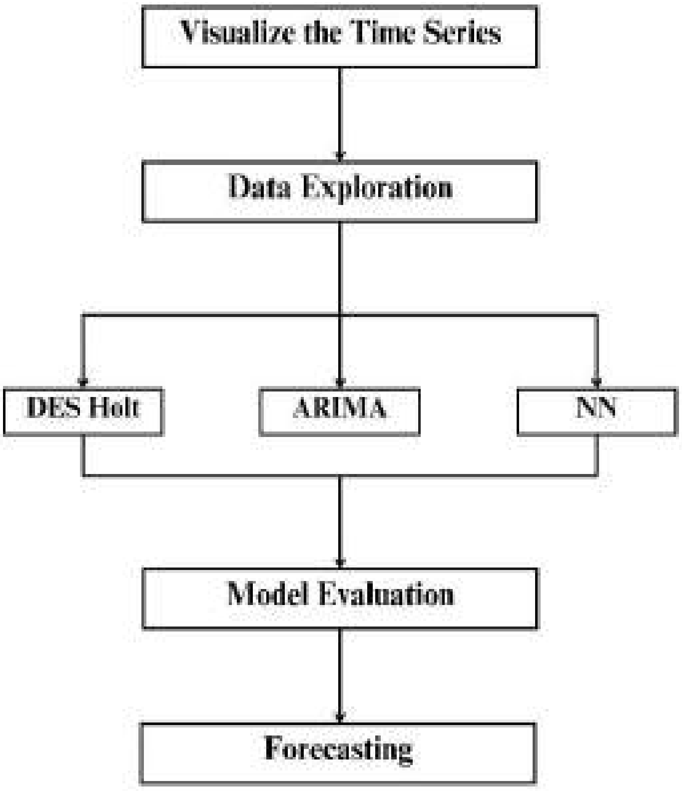
Related research article
None

For a published article:
None

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 drought.*

Graphical abstract



The research design used was ex post facto, meaning data was collected after all the events. The stages of data analysis modeling rainfall data in Samarinda City are visualized in the Graphical Abstract. The researchers chose the three methods based on their advantages and flexibility in the modeling process. The modeling process uses R software.

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.* *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.*
Resource availability	*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 [10], [11], [12], 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], [16], [17], [18], [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 may miss. Several time series data studies have successfully utilized NN are [16], [21], [22], [23], [24], [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

A. *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 < α < 1

β : trend smoothing parameter, 0 < α < 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)

B. 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:

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

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:

1. *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).

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

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

$$\phi_p(B)Z_t=a_t \quad (8)$$

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

2. *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).

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

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

$$Z_t=\theta_q(B)a_t \quad (10)$$

   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:*

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)*

Process	ACF	PACF
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	Nothings is out of bounds

C. 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 generalize 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 Figure 1.

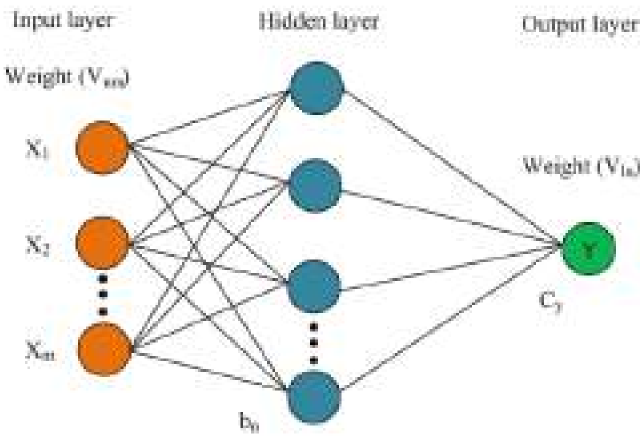


Fig. 1. Neural Network Structure

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]. 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).

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

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

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

D. 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).

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

E. 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 Figure 2.

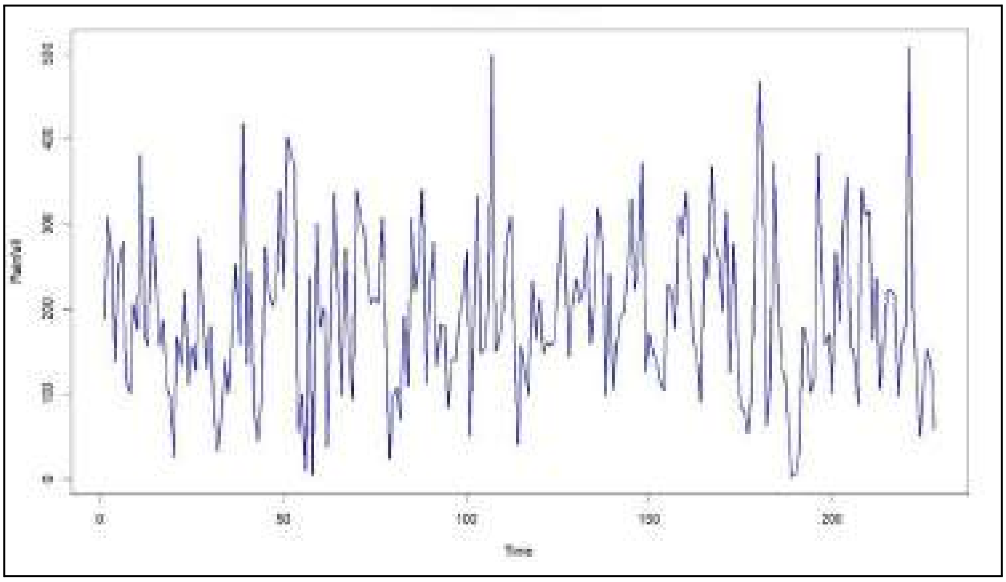


Fig. 2. Time series plot of rainfall data in Samarinda

Based on Figure 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

a. 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 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.*

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	115.28	0.7	0.1	114.66
0.1	0.2	168.88	0.4	0.2	114.72	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	120.05	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	138.89	0.4	0.6	125.13	0.7	0.6	133.49
0.1	0.7	140.94	0.4	0.7	127.14	0.7	0.7	137.81
0.1	0.8	143.61	0.4	0.8	129.14	0.7	0.8	142.23
0.1	0.9	146.54	0.4	0.9	131.40	0.7	0.9	146.69
0.2	0.1	138.03	0.5	0.1	113.38	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	124.69	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	130.18	0.5	0.6	126.01	0.8	0.6	138.95
0.2	0.7	133.68	0.5	0.7	129.02	0.8	0.7	143.98
0.2	0.8	138.00	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	120.05
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	129.47
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	125.59	0.9	0.5	140.34
0.3	0.6	127.90	0.6	0.6	129.19	0.9	0.6	146.16
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	136.90	0.9	0.8	158.77
0.3	0.9	133.12	0.6	0.9	140.91	0.9	0.9	165.67

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. 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	113.40	121.13

α = 0.7 and β = 0.1	114.66	118.03

b. 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 Figures 3 and 4.

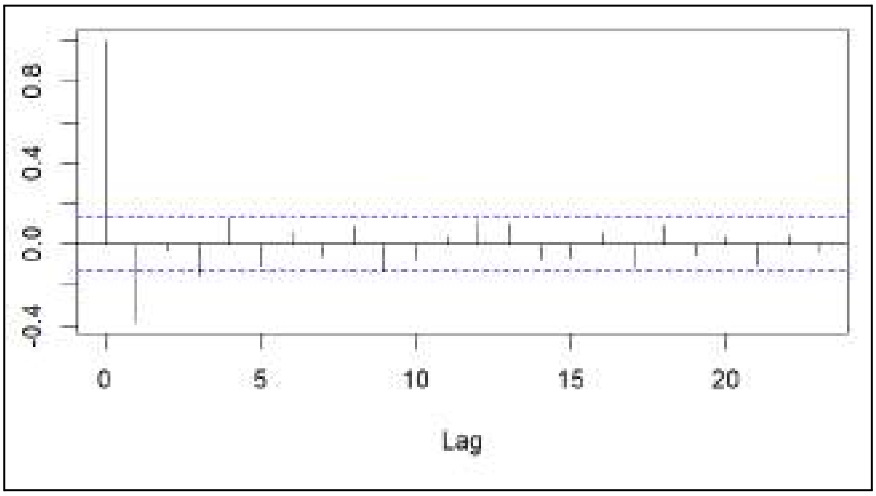


Fig. 3. ACF Plot of Rainfall Data Results of Differencing

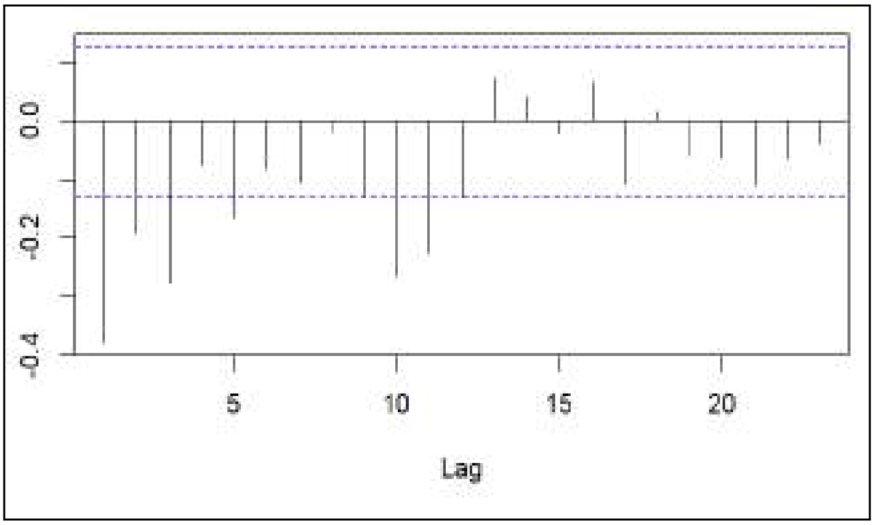


Fig. 4. PACF Plot of Rainfall Data Results of Differencing

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).

Table 5. Temporary ARIMA Model Parameter Estimation

Model	Parameter	Estimate	*p-value*	Conclusion
ARIMA(1,1,0)	$\hat{\phi}_1$	-0.380766	5.579e-10	Significant

Model	Parameter	Estimate	*p-value*	Conclusion
ARIMA(2,1,0)	$\hat{\phi}_1$	-0.455156	2.982e-12	Significant
	$\hat{\phi}_2$	-0.193967	0.002867	
ARIMA(3,1,0)	$\hat{\phi}_1$	-0.507994	1,760e-15	Significant
	$\hat{\phi}_2$	-0.320263	3.421e-06	
	$\hat{\phi}_3$	-0.275761	1.507e-05	
ARIMA(0,1,1)	$\hat{\theta}_1$	-1,000000	< 2.2e-16	Significant
ARIMA(1,1,1)	$\hat{\phi}_1$	0.273909	2.071e-05	Significant
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(2,1,1,)	$\hat{\phi}_1$	0.266500	6.351e-05	Not Significant
	$\hat{\phi}_2$	0.028447	0.6701	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(3,1,1)	$\hat{\phi}_1$	0.268049	5.375e-05	Not Significant
	$\hat{\phi}_2$	0.048998	0.475	
	$\hat{\phi}_3$	-0.081457	0.220	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	

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.

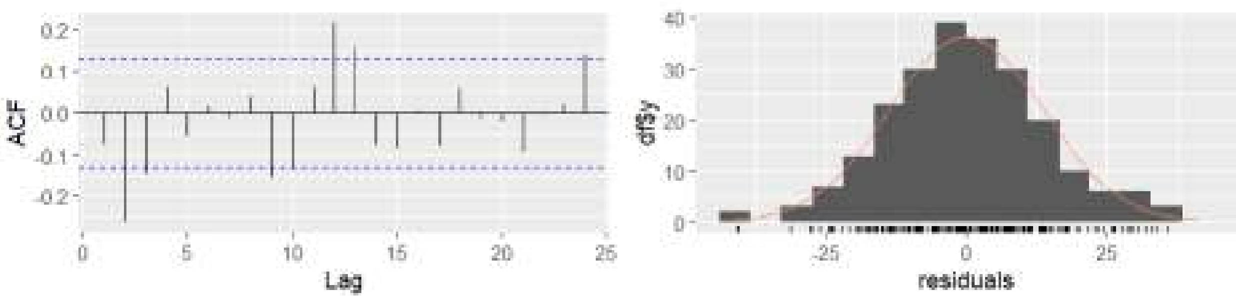


Fig. 5. Independence and Normality of ARIMA Residuals (1,1,0)

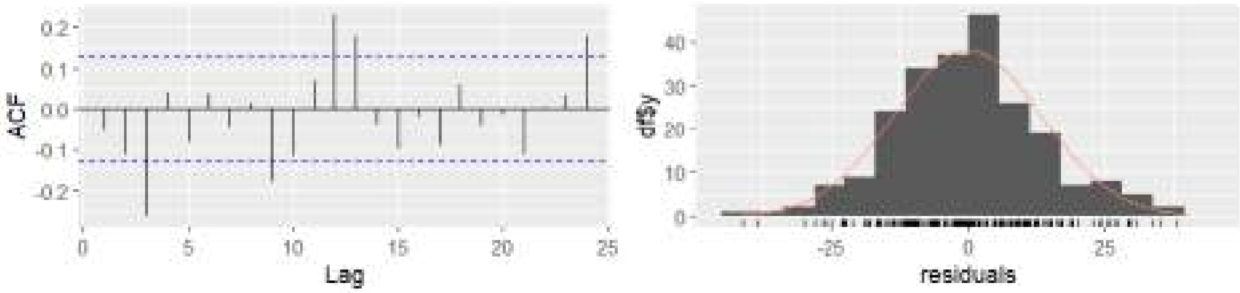


Fig. 6. Independence and Normality of ARIMA Residuals (2,1,0)

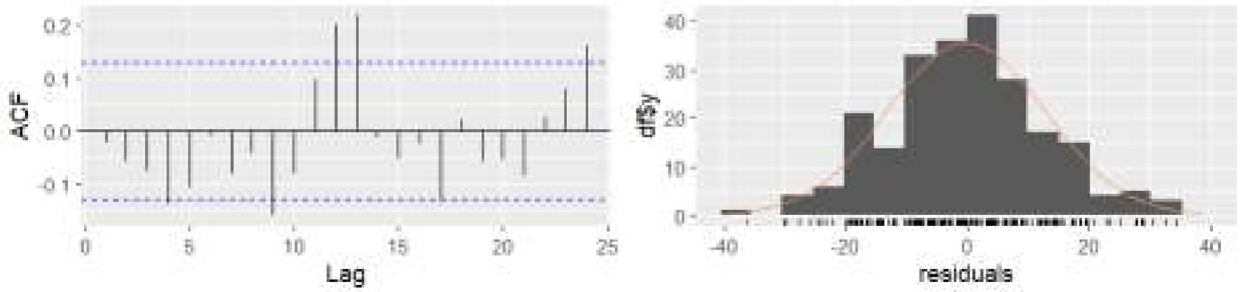


Fig. 7. Independence and Normality of ARIMA Residuals (3,1,0)

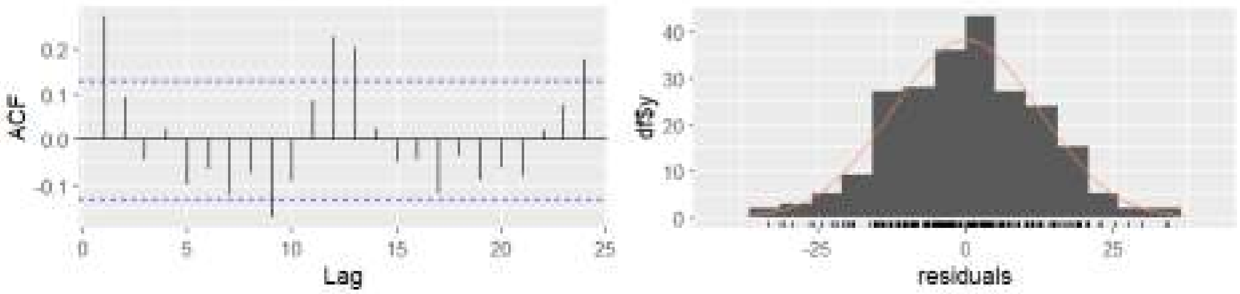


Fig. 8. Independence and Normality of ARIMA Residuals (0,1,1)

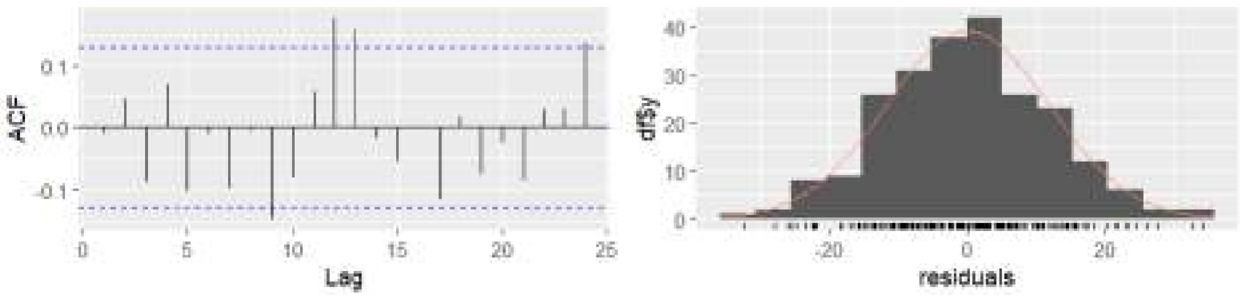


Fig. 9. Independence and Normality of ARIMA Residuals (1,1,1)

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.

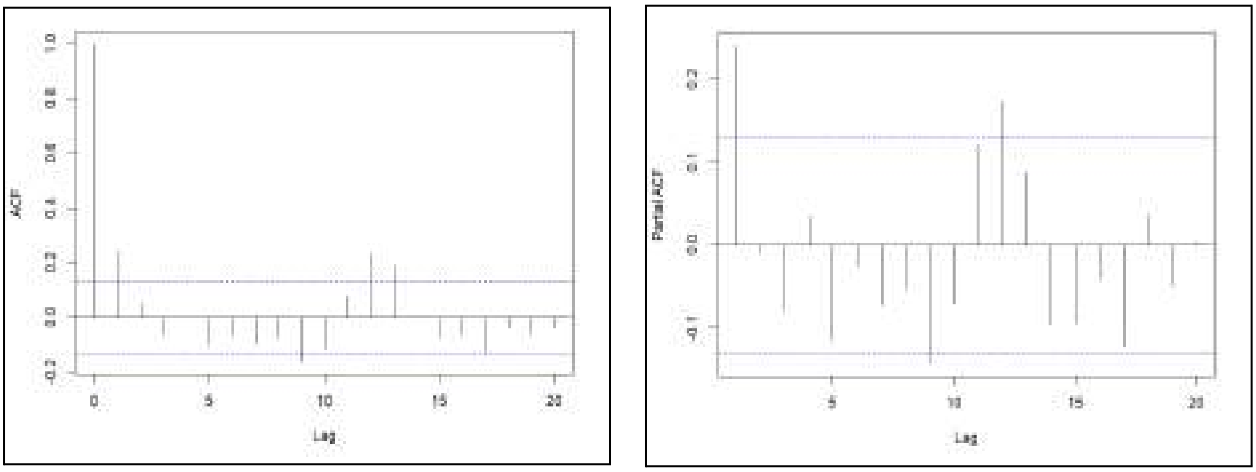
Table 6. Accuracy of Training and Testing Data for ARIMA Model (1,1,1)

Proportion	RMSEP
Data Training	92.75
Data Testing	80.82

c. *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 Figure 10.



(a) ACF graph (b) PACF graph

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

Based on Figure 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.

- *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.

- *Best Model Selection in NN*

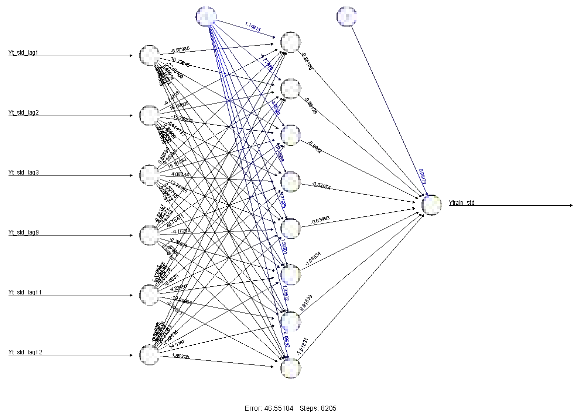
The backpropagation training process is carried out by adjusting the NN architecture. In this study, two types of architectures are used, namely networks with one hidden layer and networks with two hidden layers where each architecture will try various combinations of the number of neurons in each layer. 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.

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 rate 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 7.

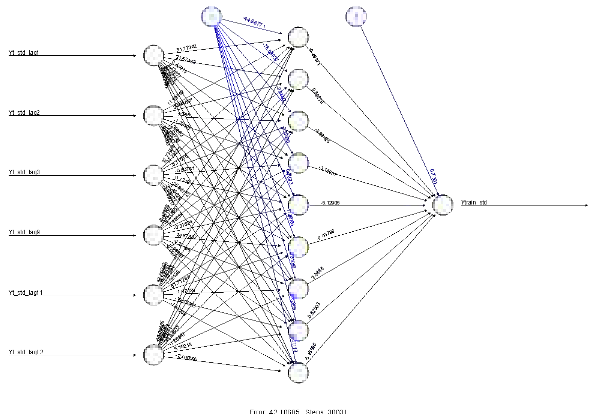
Table 7. RMSE Calculation Rsults

Hidden Layer		RMSEP		Difference between Training and Testing RMSEP
Hidden Layer1	**Hidden Layer2**	**Training**	**Testing**	
1 neuron		88.988	86.595	2.393
2 neurons		85.018	90.197	5.180
3 neurons		78.101	110.632	32.531
4 neurons		75.907	100.158	24.251
5 neurons		66.340	97.531	31.192
6 neurons		68.059	96.647	28.589
7 neurons		69.542	103.764	34.222
8 neurons		58.299	102.938	44.639
9 neurons		57.471	114.915	57.444
10 neurons		46.560	110.919	64.359
2 neurons	1 neuron	83.746	82.969	0.777
3 neurons	2 neurons	77.546	100.939	23.393
4 neurons	3 neurons	70.852	97.184	26.332
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	135.889	106.194
9 neurons	8 neurons	27.501	116.387	88.886
10 neurons	9 neurons	17.891	181.869	163.978

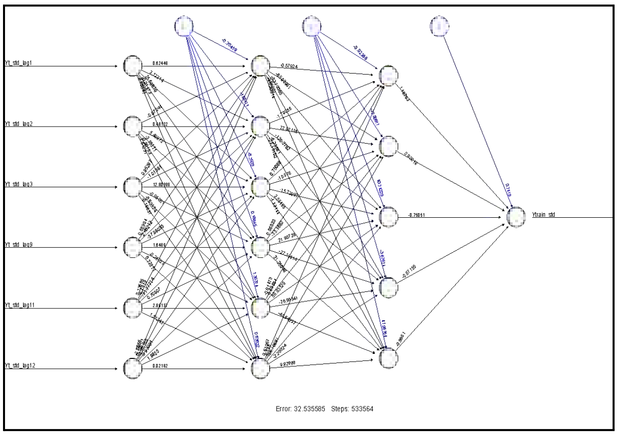
By considering various considerations such as choosing the smallest RMSEP value and the difference between the RMSEP values of the training data and testing data is not very significant, then based on Table 7, it is found that the NN model with a 2 hidden layer architecture model (6-5 neurons) is the 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 89.749. Some of the architectural results of the NN modeling can be seen in Figure 11.



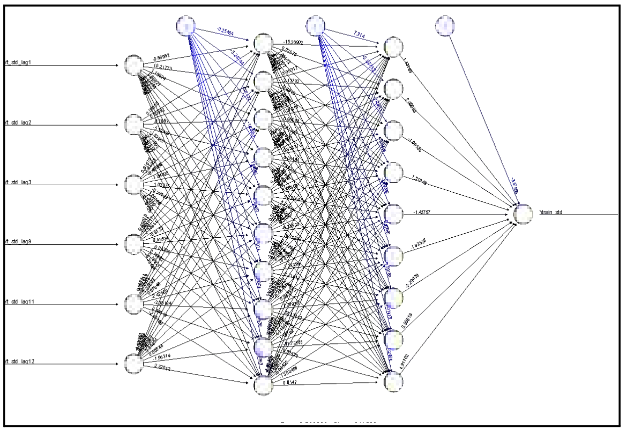
(a) NN model 1 hidden layer (8 neurons)



(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

d. 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 8 displays the RMSEP values of the best models. The model with the smallest RMSE value will be selected as the best method.

Table 8. Model Goodness of Fit Measure

Method	**RMSEP**	
	Training	Testing
DES Holt (α = 0.7 and β = 0.1)	114.66	118.03
ARIMA (1,1,1)	92.75	80.82
NN 2 HL (6-5 neurons)	**52.473**	**89.749**

In Table 8, 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.

e. 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.

Table 9. Forecasting Results of 12 Periods of NN 2 HL (6-5 neurons)

Month	Prediction Results
January	301.935
February	245.819
March	168.822
April	82.964
May	173.226
June	209.721
July	107.147
August	204.006
September	201.430
October	133.406
November	207.887
December	206,919

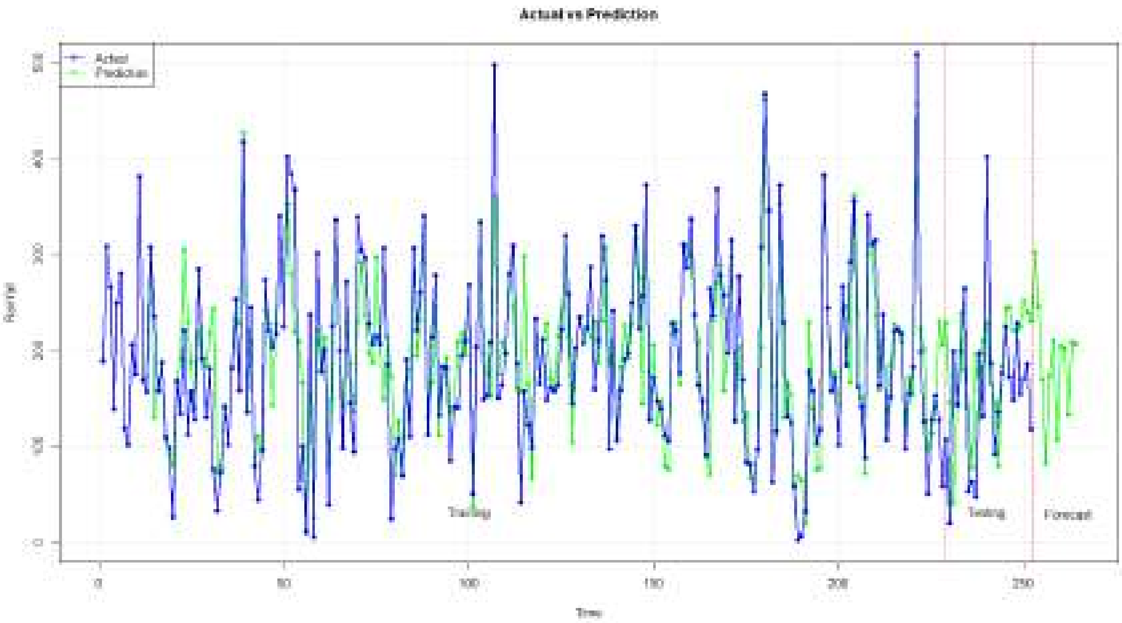


Fig. 12. Comparison plot of actual and predicted data

Figure 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.

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, 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.

CRediT author statement

Mislan: Conceptualization, Methodology, Validity tests, Writing-Preparation of the first draft, and Supervision. **Andrea Tri Rian Dani**: Data curation, Analysis, Visualization, and Editing Draft, and writing original draft.

Acknowledgments

None

Declaration of interests

*Please **tick** the appropriate statement below (please do not delete either statement) and declare any financial interests/personal relationships which may affect your work in the box below.*

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Please declare any financial interests/personal relationships which may be considered as potential competing interests here.

Supplementary material *and/or* additional information [OPTIONAL]

None.

References

[1] M. Putra, M. S. Rosid, and D. Handoko, “A Review of Rainfall Estimation in Indonesia: Data Sources, Techniques, and Methods,” *Signals*, vol. 5, no. 3, pp. 542–561, Aug. 2024, doi: 10.3390/signals5030030.

[2] 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.

[3] R. P. Permata, R. Ni'mah, and A. T. R. Dani, "Daily Rainfall Forecasting with ARIMA Exogenous Variables and Support Vector Regression," *Jurnal Varian*, vol. 7, no. 2, pp. 177–188, Jun. 2024, doi: 10.30812/varian.v7i2.3202.

[4] M. Mislan and A. T. R. Dani, "Forecasting Maximum Water Level Data for Post Sangkuliman using An Artificial Neural Network Backpropagation Algorithm," *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 8, no. 2, p. 465, Mar. 2024, doi: 10.31764/jtam.v8i2.20112.

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

[6] S. Makridakis, R. J. Hyndman, and F. Petropoulos, "Forecasting in social settings: The state of the art," *Int J Forecast*, vol. 36, no. 1, pp. 15–28, Jan. 2020, doi: 10.1016/j.ijforecast.2019.05.011.

[7] S. Makridakis, E. Spiliotis, V. Assimakopoulos, and 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, and M. S. El-Sherif, "A Comparison Between Neural-Network Forecasting Techniques-Case Study: River Flow Forecasting," 1999.

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

[10] Mustaruddin, G. Puspito, B. Z. Zanlee, and T. Nugroho, "The exponential smoothing model for squid production in Muara Angke Fishing Port, Jakarta," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2024. doi: 10.1088/1755-1315/1359/1/012098.

[11] G. Airlangga, A. Rachmat, and D. Lapihu, "Comparison of exponential smoothing and neural network method to forecast rice production in Indonesia," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 3, pp. 1367–1375, 2019, 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)*, vol. 14, no. 13, Jul. 2024, doi: 10.3390/app14135846.

[13] D. Zhao, R. Zhang, H. Zhang, and S. He, "Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models," *Sci Rep*, vol. 12, no. 1, pp. 1–13, 2022, 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*, vol. 23, no. 6, pp. 917–940, 2022, doi: 10.1007/s10198-021-01347-4.

[15] N. Mohamed, M. H. Ahmad, Z. Ismail, and 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. Atesongun and M. Gulsen, "A Hybrid Forecasting Structure Based on Arima and Artificial Neural Network Models," *Applied Sciences (Switzerland)*, vol. 14, no. 16, Aug. 2024, doi: 10.3390/app14167122.

[17] N. Suhermi, Suhartono, D. D. Prastyo, and 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, and 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, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/fi15080255.

[20] T. T. K. Tran, S. M. Bateni, S. J. Ki, and H. Vosoughifar, "A review of neural networks for air temperature forecasting," May 01, 2021, *MDPI AG*. doi: 10.3390/w13091294.

[21] W. Ge, P. Lalbakhsh, L. Isai, A. Lenskiy, and H. Suominen, "Neural Network-Based Financial Volatility Forecasting: A Systematic Review," Jan. 17, 2022, *Association for Computing Machinery*. doi: 10.1145/3483596.

[22] I. D. Mienye, T. G. Swart, and G. Obaido, “Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications,” *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.

[23] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K. W. Chau, “Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting,” *Water (Switzerland)*, vol. 12, no. 5, May 2020, doi: 10.3390/w12051500.

[24] M. Khashei and M. Bijari, “An artificial neural network (p, d, q) model for timeseries forecasting,” *Expert Syst Appl*, vol. 37, no. 1, pp. 479–489, Jan. 2010, doi: 10.1016/j.eswa.2009.05.044.

[25] S. Belhaiza and S. Al-Abdallah, “A Neural Network Forecasting Approach for the Smart Grid Demand Response Management Problem,” *Energies* , vol. 17, no. 10, May 2024, doi: 10.3390/en17102329.

[26] A. T. R. Dani, M. Fauziyah, H. Sandariria, and Q. Q. A’yun, “Forecasting The Search Trends of The Keyword ‘Sarung Wadimor’ In Indonesia on Google Trends Data Using Time Series Regression with Calender Variation and Arima Box-Jenkins,” *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 3, pp. 447–459, May 2023, doi: 10.20956/j.v19i3.24551.

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

[28] I. Mado, A. Soeprijanto, and Suhartono, “Applying of double seasonal arima model for electrical power demand forecasting at pt. pln gresik Indonesia,” *International Journal of Electrical and Computer Engineering*, vol. 8, no. 6, pp. 4892–4901, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4892-4901.

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

[32] M. Adya and F. Collopy, “How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation,” *J Forecast*, vol. 17, pp. 481–495, 1998.

[33] 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.

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

Mislan MethodsX Revisi 2

by Mislan Mislan

Submission date: 18-Nov-2024 09:50AM (UTC+0700)
Submission ID: 2735369825
File name: Revisi_2-MethodsX-Manuscript_siap_Submitted_2024.docx.pdf (1.61M)
Word count: 6140
Character count: 29682

MethodsX
Open access

Article Template

MethodsX methods article template Version 6 (April 2024)

MethodsX
Open access

Article Template

Article information

Article title

Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models.

Authors

Mislan, Andrea Tri Rian Dani*

Affiliations

Department of Physics, Faculty of Mathematics and Natural Science, Mulawarman University; Department of Mathematics, Faculty of Mathematics and Natural Science, Mulawarman University

Corresponding author's email address and Twitter handle

mislan.jos@fmipa.unmul.ac.id

Keywords

Exponential Smoothing; ARIMA; Neural Network; Time Series Modeling; Forecasting

Related research article

None

For a published article:

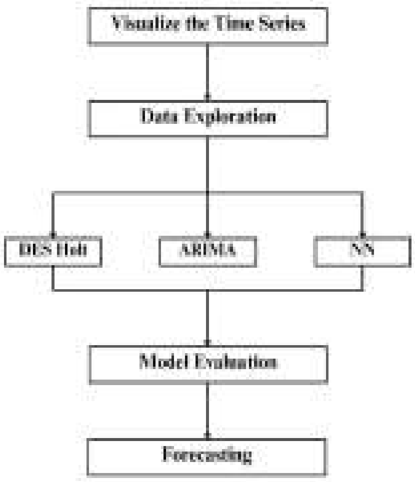
None

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 drought.*

Graphical abstract



The research design used was ex post facto, meaning data was collected after all the events. The stages of data analysis modeling rainfall data in Samarinda City are visualized in the Graphical Abstract. The researchers chose the three methods based on their advantages and flexibility in the modeling process. The modeling process uses R software.

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.* *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.*
Resource availability	*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*

MethodsX
Open access

Article Template

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 [10], [11], [12]. 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], [16], [17], [18], [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 may miss. Several time series data studies have successfully utilized NN are [16], [21], [22], [23], [24], [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

A. *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

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

Trend smoothing

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

With

$$F_{t+m} = L_t + T_t m \quad (3)$$

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

$$L_1 = Z_1 \text{ and } T_1 = Z_2 - Z_1 \quad (4)$$

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)*

B. 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 l is the degree of the MA process.*

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

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \quad (5)$$

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*

$(1-B)^d$: differenciating operator

d : order of differencing

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

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

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:

1. *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).

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

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

   $$\phi_p(B)Z_t=a_t \quad (8)$$

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

2. *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).

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

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

   $$Z_t=\theta_q(B)a_t \quad (10)$$

   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:

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)

Process	ACF	PACF
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	Nothings is out of bounds

C. 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 generalize 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 Figure 1.*

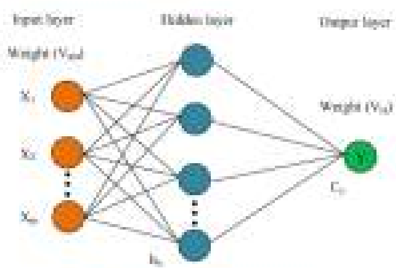


Fig. 1. Neural Network Structure

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]. *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).*

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

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

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

D. 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).

$$RMSEP = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Z_i - \hat{Z}_i\right)} \qquad (13)$$

E. 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 Figure 2.

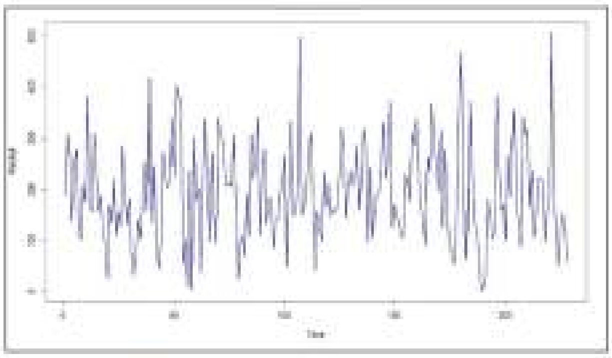


Fig. 2. Time series plot of rainfall data in Samarinda

Based on Figure 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

a. 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 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.

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	115.28	0.7	0.1	114.66
0.1	0.2	168.88	0.4	0.2	114.72	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	120.05	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	138.89	0.4	0.6	125.13	0.7	0.6	133.49
0.1	0.7	140.94	0.4	0.7	127.14	0.7	0.7	137.81
0.1	0.8	143.61	0.4	0.8	129.14	0.7	0.8	142.23
0.1	0.9	146.54	0.4	0.9	131.40	0.7	0.9	146.69
0.2	0.1	138.03	0.5	0.1	113.38	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	121.20	0.5	0.3	117.25	0.8	0.3	124.84
0.2	0.4	124.69	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	130.18	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	143.98
0.2	0.8	138.00	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	120.05
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	129.47
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	125.59	0.9	0.5	140.34
0.3	0.6	127.90	0.6	0.6	129.19	0.9	0.6	146.16
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	136.90	0.9	0.8	158.77
0.3	0.9	133.12	0.6	0.9	140.91	0.9	0.9	165.67

16

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. Optimal Combination Value of Training and Testing Data

Parameter Value	RMSEP Training	RMSEP Testing
α = 0.4 and β = 0.2	114.72	122.32
α = 0.5 and β = 0.1	113.38	127.50
α = 0.6 and β = 0.1	113.40	121.13

α = 0.7 and β = 0.1	114.66	118.03

6. 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 Figures 3 and 4.

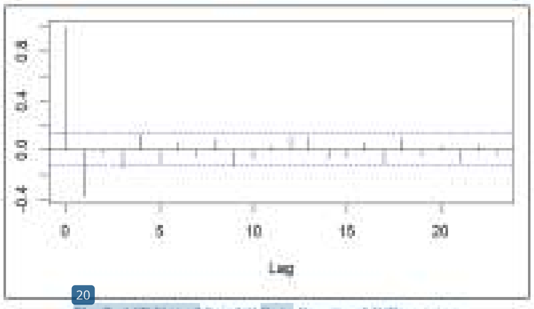


Fig. 3. ACF Plot of Rainfall Data Results of Differencing

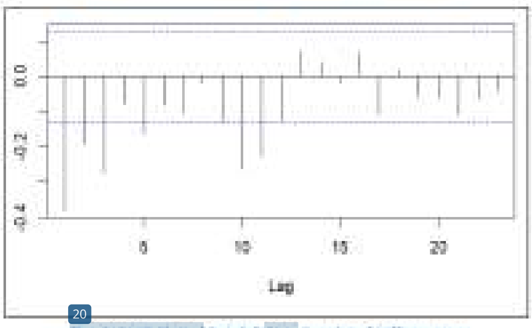


Fig. 4. PACF Plot of Rainfall Data Results of Differencing

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).

Table 5. Temporary ARIMA Model Parameter Estimation

Model	Parameter	Estimate	*p-value*	Conclusion
ARIMA(1,1,0)	$\hat{\phi}_1$	-0.380766	5.579e-10	Significant

Model	Parameter	Estimate	p-value	Conclusion
ARIMA(2,1,0)	$\hat{\phi}_1$	-0.455156	2.982e-12	Significant
	$\hat{\phi}_2$	-0.193967	0.002867	
ARIMA(3,1,0)	$\hat{\phi}_1$	-0.507994	1,760e-15	Significant
	$\hat{\phi}_2$	-0.320263	3.421e-06	
	$\hat{\phi}_3$	-0.275761	1.507e-05	
ARIMA(0,1,1)	$\hat{\theta}_1$	-1,000000	< 2.2e-16	Significant
ARIMA(1,1,1)	$\hat{\phi}_1$	0.273909	2.071e-05	Significant
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(2,1,1,)	$\hat{\phi}_1$	0.266500	6.351e-05	Not Significant
	$\hat{\phi}_2$	0.028447	0.6701	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(3,1,1)	$\hat{\phi}_1$	0.268049	5.375e-05	Not Significant
	$\hat{\phi}_2$	0.048998	0.475	
	$\hat{\phi}_3$	-0.081457	0.220	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	

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.

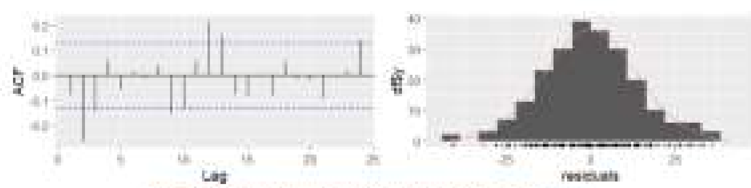


Fig. 5. Independence and Normality of ARIMA Residuals (1,1,0)

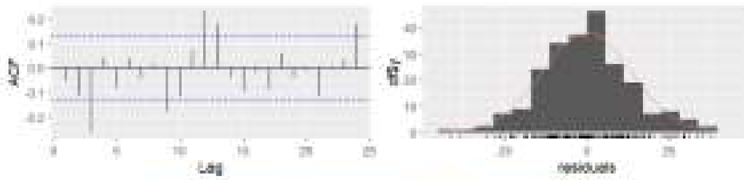


Fig. 6. Independence and Normality of ARIMA Residuals (2,1,0)

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

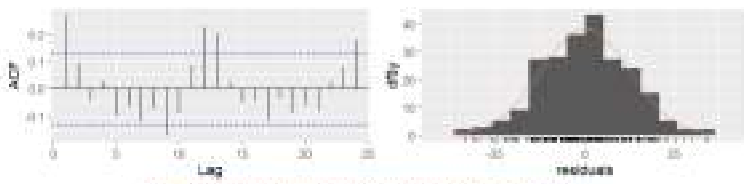


Fig. 8. Independence and Normality of ARIMA Residuals (0,1,1)

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

56

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.

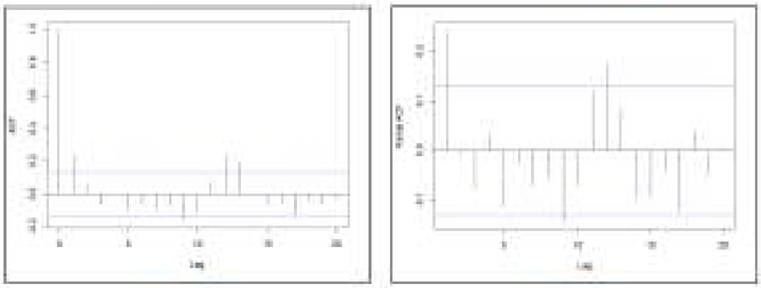
Table 6. Accuracy of Training and Testing Data for ARIMA Model (1,1,1)

Proportion	RMSEP
Data Training	92.75
Data Testing	80.82

c. *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 Figure 10.



(a) ACF graph

(b) PACF graph

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

Based on Figure 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.

- *Data Standardization*

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.

- *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.

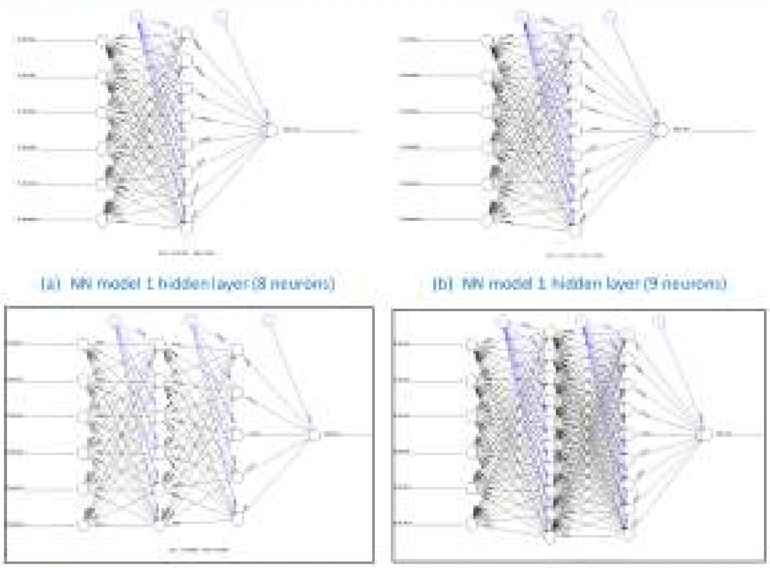
In the NN architecture in this study, researchers tried to use a maximum of two hidden layers in the NN compartment, where a combination of the number of neurons from 1 to 10 will be carried out. This combination obtains the optimal number of neurons in the first and second hidden layers. Some architectures of combinations of neurons in each hidden layer are intended to perform hyperparameter tuning, limiting the learning rate 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 7.

Table 7. RMSE Calculation Results

Hidden Layer		RMSEP		Difference between Training and Testing RMSEP
Hidden Layer1	Hidden Layer2	Training	Testing	
1 neuron		88.988	86.595	2.393
2 neurons		85.018	90.197	5.180
3 neurons		78.101	110.632	32.531
4 neurons		75.907	100.158	24.251
5 neurons		66.340	97.531	31.192
6 neurons		68.058	96.647	28.589
7 neurons		69.542	103.764	34.222
8 neurons		58.299	102.938	44.639
9 neurons		57.471	114.915	57.444
10 neurons		46.560	110.919	64.359
2 neurons	1 neuron	83.746	82.969	0.777
3 neurons	2 neurons	77.546	100.939	23.393
4 neurons	3 neurons	70.852	97.184	26.332
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	135.889	106.194
9 neurons	8 neurons	27.501	116.387	88.886
10 neurons	9 neurons	17.891	181.869	163.978

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 7, 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 89.749. Some of the architectural results of the NN modeling can be seen in Figure 11.

MethodsX
Open access
Article Template



(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

d. r 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 8 displays the RMSEP values of the best models. The model with the smallest RMSE value will be selected as the best method.

Table 8. Model Goodness of Fit Measure

Method	RMSEP	
	Training	Testing
DES Holt (α = 0.7 and β = 0.1)	114.66	118.03
ARIMA (1,1,1)	92.75	80.82
NN 2 HL (6-5 neurons)	**52.473**	**89.749**

In Table 8, 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.

e. 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.

Table 9. Forecasting Results of 12 Periods of NN 2 HL (6-5 neurons)

Month	Prediction Results
January	301.935
February	245.819
March	168.822
April	82.964
May	173.226
June	209.721
July	107.347
August	204.006
September	201.430
October	131.406
November	207.887
December	206.919

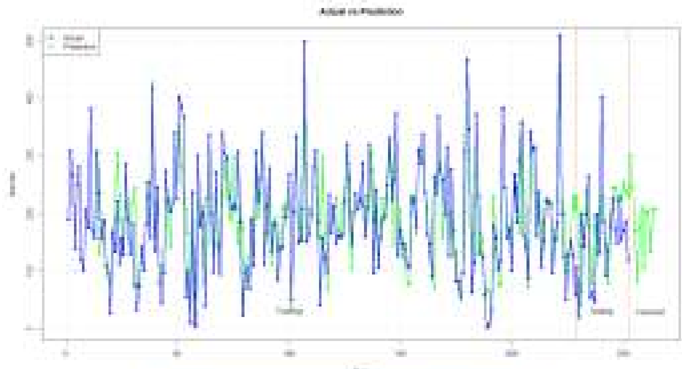


Fig. 12. Comparison plot of actual and predicted data

Figure 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.

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, 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.

CRediT author statement

Misdan: Conceptualization, Methodology, Validity tests, Writing-Preparation of the first draft, and Supervision. **Andrea Tri Rian Dani**: Data curation, Analysis, Visualization, and Editing Draft, and writing original draft.

Acknowledgments

None

Declaration of interests

Please tick the appropriate statement below (please do not delete either statement) and declare any financial interests/personal relationships which may affect your work in the box below.

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Please declare any financial interests/personal relationships which may be considered as potential competing interests here.

Supplementary material *and/or* **additional information [OPTIONAL]**

None.

References

[1] M. Putra, M. S. Rosid, and D. Handoko, "A Review of Rainfall Estimation in Indonesia: Data Sources, Techniques, and Methods," *Signals*, vol. 5, no. 3, pp. 542–561, Aug. 2024, doi: 10.3390/signals5030030.

[2] 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.

[3] R. P. Permata, R. Ni'mah, and A. T. R. Dani, "Daily Rainfall Forecasting with ARIMA Exogenous Variables and Support Vector Regression," *Jurnal Varian*, vol. 7, no. 2, pp. 177–188, Jun. 2024, doi: 10.30812/varian.v7i2.3202.

[4] M. Mislan and A. T. R. Dani, "Forecasting Maximum Water Level Data for Post Sangkuliman using An Artificial Neural Network Backpropagation Algorithm," *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 8, no. 2, p. 465, Mar. 2024, doi: 10.31764/jtam.v8i2.20112.

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

[6] S. Makridakis, R. J. Hyndman, and F. Petropoulos, "Forecasting in social settings: The state of the art," *Int J Forecast*, vol. 36, no. 1, pp. 15–28, Jan. 2020, doi: 10.1016/j.ijforecast.2019.05.011.

[7] S. Makridakis, E. Spiliotis, V. Assimakopoulos, and 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, and M. S. El-Sherif, "A Comparison Between Neural-Network Forecasting Techniques-Case Study: River Flow Forecasting," 1999.

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

[10] Mustaruddin, G. Puspito, B. Z. Zanlee, and T. Nugroho, "The exponential smoothing model for squid production in Muara Angke Fishing Port, Jakarta," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2024. doi: 10.1088/1755-1315/1359/1/012098.

[11] G. Airlangga, A. Rachmat, and D. Lapihu, "Comparison of exponential smoothing and neural network method to forecast rice production in Indonesia," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 3, pp. 1367–1375, 2019, 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)*, vol. 14, no. 13, Jul. 2024, doi: 10.3390/app14135846.

[13] D. Zhao, R. Zhang, H. Zhang, and S. He, "Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models," *Sci Rep*, vol. 12, no. 1, pp. 1–13, 2022, 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*, vol. 23, no. 6, pp. 917–940, 2022, doi: 10.1007/s10198-021-01347-4.

[15] N. Mohamed, M. H. Ahmad, Z. Ismail, and 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. Atesongun and M. Gulsen, "A Hybrid Forecasting Structure Based on Arima and Artificial Neural Network Models," *Applied Sciences (Switzerland)*, vol. 14, no. 16, Aug. 2024, doi: 10.3390/app14167122.

[17] N. Suhermi, Suhartono, D. D. Prastyo, and 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, and 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, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/fi15080255.

[20] T. T. K. Tran, S. M. Bateni, S. J. Ki, and H. Vosoughifar, "A review of neural networks for air temperature forecasting," May 01, 2021, *MDPI AG*. doi: 10.3390/w13091294.

[21] W. Ge, P. Lalbakhsh, L. Isai, A. Lenskiy, and H. Suominen, "Neural Network-Based Financial Volatility Forecasting: A Systematic Review," Jan. 17, 2022, *Association for Computing Machinery*. doi: 10.1145/3483596.

[22] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.

[23] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K. W. Chau, "Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting," *Water (Switzerland)*, vol. 12, no. 5, May 2020, doi: 10.3390/w12051500.

[24] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Syst Appl*, vol. 37, no. 1, pp. 479–489, Jan. 2010, doi: 10.1016/j.eswa.2009.05.044.

[25] S. Belhaiza and S. Al-Abdallah, "A Neural Network Forecasting Approach for the Smart Grid Demand Response Management Problem," *Energies* , vol. 17, no. 10, May 2024, doi: 10.3390/en17102329.

[26] A. T. R. Dani, M. Fauziyah, H. Sandariria, and Q. Q. A'yun, "Forecasting The Search Trends of The Keyword 'Sarung Wadimor' in Indonesia on Google Trends Data Using Time Series Regression with Calender Variation and Arima Box-Jenkins," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 3, pp. 447–459, May 2023, doi: 10.20956/j.v19i3.24551.

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

[28] I. Mado, A. Soeprijanto, and Suhartono, "Applying of double seasonal arima model for electrical power demand forecasting at pt. pln gresik Indonesia," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 6, pp. 4892–4901, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4892-4901.

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

[32] M. Adya and F. Collopy, "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation," *J Forecast*, vol. 17, pp. 481–495, 1998.

[33] 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.

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

Mislan MethodsX Revisi 2

ORIGINALITY REPORT

16%	12%	11%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.eng.auburn.edu Internet Source	1%
2	jurnal.lapan.go.id Internet Source	1%
3	ijimai.org Internet Source	1%
4	Nidhi Sharma, Krishanu Biswas, Shikhar Krishn Jha. "Dry sliding wear behaviour of metal toughened nanoceramics: A case study of aluminium-alumina nanocermets", Wear, 2022 Publication	1%
5	cdn.repository.uisi.ac.id Internet Source	1%
6	ia801403.us.archive.org Internet Source	1%
7	Komi Mensah Agboka, Emmanuel Peter, Erion Bwambale, Bonoukpoè Mawuko Sokame. "Biological reinforcement learning simulation for natural enemy -host behavior: Exploring deep learning algorithms for population dynamics", MethodsX, 2024 Publication	<1%
8	journal.uad.ac.id Internet Source	<1%
9	repository.unib.ac.id Internet Source	<1%
10	iopscience.iop.org Internet Source	<1%

11	www.grafiati.com Internet Source	<1 %
12	www.ssca.org.in Internet Source	<1 %
13	Xu-Ran Zhang, Hua-Yong Zhang. "Holocene vegetation dynamics and climate reconstruction of Chahannur Lake in the Inner Mongolian Plateau: Pollen-based reconstructions", Ecological Indicators, 2025 Publication	<1 %
14	jestec.taylors.edu.my Internet Source	<1 %
15	Eunju Hwang. "A partially periodic oscillation model combined with heterogeneous autoregression and its application to COVID-19", Applied Mathematical Modelling, 2024 Publication	<1 %
16	M I Mawardi, W S Winanti, T W Sudinda, K Amru, A A Saraswati, S I Sachoemar, Z Arifin, A Alimin. "Analysis of net-zero emission index for several areas in Indonesia using individual carbon footprint and land use covered", IOP Conference Series: Earth and Environmental Science, 2023 Publication	<1 %
17	c.coek.info Internet Source	<1 %
18	ideas.repec.org Internet Source	<1 %
19	"CS2-PC-22_HR", ActEd Publication	<1 %
20	Sanjita Jaipuria, S. S. Mahapatra. "Chapter 7 Reduction of Bullwhip Effect in Supply Chain through Improved Forecasting Method: An	<1 %

Integrated DWT and SVM Approach", Springer Nature, 2013
Publication

21	Wenhe Chen, Hanting Zhou, Longsheng Cheng, Min Xia. "Condition Monitoring and Anomaly Detection of Wind Turbines using Temporal Convolutional Informer and Robust Dynamic Mahalanobis Mahalanobis Distance", IEEE Transactions on Instrumentation and Measurement, 2023 Publication	<1 %
22	journals.plos.org Internet Source	<1 %
23	geekflare.com Internet Source	<1 %
24	M. Kumazawa, Y. Imanishi, Y. Fukao, M. Furumoto, A. Yamamoto. "A theory of spectral analysis based on the characteristic property of a linear dynamic system", Geophysical Journal International, 1990 Publication	<1 %
25	Siti Azirah Asmai, Burairah Hussin, Mokhtar Mohd Yusof, Abdul Samad Shibghatullah. "Time Series Prediction Techniques for Estimating Remaining Useful Lifetime of Cutting Tool Failure", International Review on Computers and Software (IRECOS), 2014 Publication	<1 %
26	Yao, J.. "A case study on using neural networks to perform technical forecasting of forex", Neurocomputing, 200009 Publication	<1 %
27	addi.ehu.es Internet Source	<1 %
28	dev.journal.ugm.ac.id Internet Source	<1 %

29	knm21.unram.ac.id Internet Source	<1 %
30	library.iugaza.edu.ps Internet Source	<1 %
31	unitesi.unive.it Internet Source	<1 %
32	Gao, D.. "Neural networks for event extraction from time series: a back propagation algorithm approach", Future Generation Computer Systems, 200507 Publication	<1 %
33	Richard B. Miller, Lee N. Johnson. "Advanced Methods in Family Therapy Research - A Focus on Validity and Change", Routledge, 2025 Publication	<1 %
34	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 %
35	bth.diva-portal.org Internet Source	<1 %
36	edu.stats.gov.cn Internet Source	<1 %
37	heca-analitika.com Internet Source	<1 %
38	ijrrjournal.com Internet Source	<1 %
39	pdfs.semanticscholar.org Internet Source	<1 %
40	ph02.tci-thaijo.org Internet Source	<1 %

41	sigma.yildiz.edu.tr Internet Source	<1 %
42	www.accentsjournals.org Internet Source	<1 %
43	www.tandfonline.com Internet Source	<1 %
44	Elmanani Simamora, Wahyunita Yusardi, Abil Mansyur. "Planting Pattern Modeling Based on Rainfall Prediction Using Backpropagation Artificial Neural Network (Case Study: BMKG Rainfall Data, Deli Serdang Regency)", Journal of Physics: Conference Series, 2021 Publication	<1 %
45	Endah Retnowati, Anik Ghufron, Marzuki, Kasiyan, Adi Cilik Pierawan, Ashadi. "Character Education for 21st Century Global Citizens", Routledge, 2018 Publication	<1 %
46	Görkem Ekleme, Funda Yercan. "Assessment of the Black Sea Grain Initiative: Crisis Management via Maritime Transportation", Transport Policy, 2025 Publication	<1 %
47	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 %
48	Yu Li, Dong Wei, Xiannan Peng, Hao Shen, Li Ni, Guoyou Wei. "High-efficiency ammonium recovery and high biomass yield from ultrahigh-ammonium industrial effluent via online-controlled photo-fermentation with	<1 %

thermoacidophilic Galdieria sulphuraria", Algal Research, 2025
Publication

49	archive.org Internet Source	<1 %
50	discovery.researcher.life Internet Source	<1 %
51	erepository.uoeld.ac.ke Internet Source	<1 %
52	hal.science Internet Source	<1 %
53	journal2.um.ac.id Internet Source	<1 %
54	manglar.uninorte.edu.co Internet Source	<1 %
55	p.pdfkul.com Internet Source	<1 %
56	rumahstudio.com Internet Source	<1 %
57	www.epj-conferences.org Internet Source	<1 %
58	www.ijmlc.org Internet Source	<1 %
59	www.intechopen.com Internet Source	<1 %
60	"Innovations in Information and Communication Technologies (IICT-2020)", Springer Science and Business Media LLC, 2021 Publication	<1 %
61	Khusnia Nurul Khikmah, Kusman Sadik, Indahwati Indahwati. "TRANSFER FUNCTION AND ARIMA MODEL FOR FORECASTING BI	<1 %

RATE IN INDONESIA", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2023
Publication

62 Mihaela Oprea, Marian Popescu, Elia Georgiana Dragomir, Sanda Florentina Mihalache. "Models of particulate matter concentration forecasting based on artificial neural networks", 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2017 <1%
Publication

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

Gmail Andrea T R Dani <andrikadoko@gmail.com>

Your PDF has been built and requires approval

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 19 November 2024 pukul 10.49
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

MethodsX
Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
Authors: Mislan Mislan; Andrea Tri Rian Dani

Dear Mislan,

The PDF for your submission, "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" has now been built and is ready for your approval. Please view the submission before approving it, to be certain that it is free of any errors. If you have already approved the PDF of your submission, this e-mail can be ignored.

To approve the PDF please login to the Editorial Manager as an Author:

https://www.editorialmanager.com/mex/
Your username is: mislanjoss

Then click on the folder 'Submissions Waiting for Author's Approval' to view and approve the PDF of your submission. You may need to click on 'Action Links' to expand your Action Links menu.

You will also need to confirm that you have read and agree with the Elsevier Ethics in Publishing statement before the submission process can be completed. Once all of the above steps are done, you will receive an e-mail confirming receipt of your submission from the Editorial Office. For further information or if you have trouble completing these steps please go to: http://help.elsevier.com/app/answers/detail/a_id/88/p/7923.

Please note that you are required to ensure everything appears appropriately in PDF and no change can be made after approving a submission. If you have any trouble with the generated PDF or completing these steps please go to: http://help.elsevier.com/app/answers/detail/a_id/88/p/7923.

Your submission will be given a reference number once an Editor has been assigned to handle it.

Thank you for your time and patience.
Kind regards,
Editorial Office
MethodsX

****
For further assistance, please visit our customer support site at http://help.elsevier.com/app/answers/list/p/7923. Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EM via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

Gmail

Andrea T R Dani <andrikadoko@gmail.com>

Track the status of your submission to MethodsX

1 pesan

Track your Elsevier submission <no-reply@submissions.elsevier.com> 26 November 2024 pukul 18.40
Kepada: andrikadoko@gmail.com

Manuscript Number: MEX-D-24-00981
Manuscript Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
Journal: MethodsX

Dear Mislan Mislan,

Your submitted manuscript is currently under review. You can track the status of your submission in Editorial Manager, or track the review status in more detail using Track your submission here:
https://track.authorhub.elsevier.com?uuid=5ffec470-a084-4828-92f7-424b7abccb82

This page will remain active until the peer review process for your submission is completed. You can visit the page whenever you like to check the progress of your submission. The page does not require a login, so you can also share the link with your co-authors.

If you are a WeChat user, then you can also receive status updates via WeChat. To do this please click the following link; you will be taken to Elsevier China's website where further instructions will guide you on how to give permission to have your submission's details made visible in WeChat. Note that by clicking the link no submission data is transferred to the WeChat platform. If you have any questions about using Track your submission with WeChat please visit 在线咨询 https://cn.service.elsevier.com/app/chat/chat_launch/supporthub/publishing/session/ - Journal Article Publishing 支持中心

https://webapps.elsevier.cn/st-wechat/subscribe?signature=1732588366-57a9129a06ccfce7a876372888564d52&uuid=5ffec470-a084-4828-92f7-424b7abccb82

We hope you find this service useful.

Kind regards,
Journal Office of MethodsX
Elsevier B.V.

Gmail Andrea T R Dani <andrikadoko@gmail.com>

Regarding your Manuscript MEX-D-24-00981R2

1 pesan

em@editorialmanager.com <em@editorialmanager.com> 30 November 2024 pukul 20.16
Balas Ke: mexjm@elsevier.com
Kepada: Mislan Mislan <andrikadoko@gmail.com>

CC: "Andrea Tri Rian Dani" andreatririandani@fmipa.unmul.ac.id

Ref.: **MEX-D-24-00981R2**
Title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
MethodsX

Dear Mislan,

I am pleased to inform you that your paper "Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models" has been accepted for publication in MethodsX.
We aim to send you the proofs of your paper within 10 working days.

MethodsX is a community effort, by researchers for researchers. We hope you enjoyed being part of the community as an author, and hope to welcome you back as a reviewer in the future. If you wish to be added to our reviewers database, please e-mail MethodsX@elsevier.com to express your interest.

Also, have you collected research data or developed software as part of your research project? Why not publish that work in addition to your 'MethodsX' article? Go to the Data in Brief and SoftwareX websites to learn how you can publish individual data sets and software, making your work searchable, peer reviewed, citable and reproducible.

Congratulations again and thank you for submitting your work to MethodsX.

We appreciate and value your contribution to MethodsX. We regularly invite authors of recently published manuscript to participate in the peer review process. If you were not already part of the journal's reviewer pool, you have now been added to it. We look forward to your continued participation in our journal, and we hope you will consider us again for future submissions.

Yours sincerely,

Damià Barceló
Editor-in-Chief
MethodsX

Comments to Authors:

Reviewer #3: I accepted the authors' justification and have no further questions

At Elsevier, we want to help all our authors to stay safe when publishing. Please be aware of fraudulent messages requesting money in return for the publication of your paper. If you are publishing open access with Elsevier, bear in mind that we will never request payment before the paper has been accepted. We have prepared some guidelines (https://www.elsevier.com/connect/authors-update/seven-top-tips-on-stopping-apc-scams) that you may find helpful, including a short video on Identifying fake acceptance letters (https://www.youtube.com/watch?v=o5l8thD9XtE). Please remember that you can contact Elsevier s Researcher Support team (https://service.elsevier.com/app/home/supporthub/publishing/) at any time if you have questions about your manuscript, and you can log into Editorial Manager to check the status of your manuscript (https://service.elsevier.com/app/answers/detail/a_id/29155/c/10530/supporthub/publishing/kw/status/).

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Remove my information/details). Please contact the publication office if you have any questions.

Gmail

Andrea T R Dani <andrikadoko@gmail.com>

Your Elsevier invoice OAD0000509909

1 pesan

OAElsevierInvoicesNoReply@elsevier.com <OAElsevierInvoicesNoReply@elsevier.com>
Balas Ke: OAElsevierInvoicesNoReply@elsevier.com
Kepada: andrikadoko@gmail.com

3 Desember 2024 pukul 18.02

Dear Sir/Madam,

Please find your invoice **OAD0000509909** attached to this email.

Transaction Number	Due Date	Currency	Transaction Amount
OAD0000509909	02-JAN-2025	USD	459.00

Can we assist you with anything?

For payment and bank account information please refer to the attached invoice. If you have a question regarding the invoice please follow the link to our customer support center.

Kind regards,

Researcher Support
ELSEVIER | Research Content Operations
www.elsevier.com

*** This is an automatically generated email, please do not reply ***

This email is for use by the intended recipient and contains information that may be confidential. If you are not the intended recipient, please notify the sender by return email and delete this email from your inbox. Any unauthorized use or distribution of this email, in whole or in part, is strictly prohibited and may be unlawful. Any price quotes contained in this email are merely indicative and will not result in any legally binding or enforceable obligation. Unless explicitly designated as an intended e-contract, this email does not constitute a contract offer, a contract amendment, or an acceptance of a contract offer.

Invoice_OAD0000509909.pdf
48K

Gmail Andrea T R Dani <andrikadoko@gmail.com>

Publishing Agreement completed for your article [MEX_103080]

1 pesan

Elsevier - Author Forms <Article_Status@elsevier.com> 3 Desember 2024 pukul 16.00
Kepada: andrikadoko@gmail.com

ELSEVIER

Dear Mr. Mislan,

Thank you for completing the Publishing Agreement Form for your article *Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models*. Please find attached a copy of the "Journal Publishing (License) Agreement" which you completed online on December 03, 2024.

If you have any questions, please do not hesitate to contact us. To help us assist you, please quote our article reference MEX_103080 in all correspondence.

Now that your article has been accepted, you will want to maximize the impact of your work. Elsevier facilitates and encourages authors to share their article responsibly. To learn about the many ways in which you can share your article whilst respecting copyright, visit: www.elsevier.com/sharing-articles.

We are committed to publishing your article as quickly as possible.

Kind regards,
Elsevier Researcher Support

Have questions or need assistance?
Please do not reply to this automated message.
For further assistance, please visit our Elsevier Support Center where you can search for solutions on a range of topics and find answers to frequently asked questions.
From here you can also contact our Researcher Support team via 24/7 live chat, email or phone support.

© 2024 Elsevier Ltd | Privacy Policy http://www.elsevier.com/privacypolicy
Elsevier Limited, 125 London Wall, London, EC2Y 5AS, United Kingdom, Registration No. 1982084. This e-mail has been sent to you from Elsevier Ltd. To ensure delivery to your inbox (not bulk or junk folders), please add Article_Status@elsevier.com to your address book or safe senders list.

MEX103080.html
25K

Gmail Andrea T R Dani <andrikadoko@gmail.com>

Review the proofs of your article [MEX_103080] in MethodsX

1 pesan

s.singh.7@elsevier.com <s.singh.7@elsevier.com> 7 Desember 2024 pukul 06.48
Kepada: andrikadoko@gmail.com

PLEASE DO NOT ALTER THE SUBJECT LINE OF THIS EMAIL

Our reference: MEX 103080
Article reference: MEX_MEX-D-24-00981
Article title: Navigating Samarinda's Climate: A Comparative Analysis of Rainfall Forecasting Models
To be published in: MethodsX

Dear Mr. Mislan,

We recently sent you proofs of the above-mentioned article. Our records indicate that we have not yet received your proof corrections or your approval to publish the article without change.

Please review the proof and submit any corrections within two business days to help us publish your article as quickly and accurately as possible.

Please click on the link below to access your proofs and the instructions for returning your corrections.

https://elsevier.proofcentral.com/en-us/landing-page.html?token=569cb798d0fa4fd037cd2584e3e199

If this link does not work, please copy the entire URL (noting that it may run on to a second line in this message) into your browser.

If you have already returned your corrections within the past 24 hours then please ignore this reminder.

Please note that once we receive your corrections, your article is considered finalized and further amendments are no longer possible.

Thank you for your cooperation. Please contact us if you have any questions, and quote the reference for your article, MEX 103080, in all of your messages to us.

Kind regards,

Singh Satendra
Data Administrator
Elsevier
E-Mail: s.singh.7@elsevier.com

HAVE QUESTIONS OR NEED ASSISTANCE?

For further assistance, please visit our Customer Support site, where you can search for solutions on a range of topics, such as Open Access or payment queries, and find answers to frequently asked questions. You can also talk to our customer support team by phone 24 hours a day from Monday-Friday and 24/7 by live chat and email.

Get started here: http://service.elsevier.com/app/home/supporthub/publishing
----
Copyright © 2015 Elsevier B.V. | Privacy Policy http://www.elsevier.com/privacypolicy

Elsevier Limited, The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, United Kingdom, Registration No. 1982084

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 drought.

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
2215-0161/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

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 [10–12], 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 may 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

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

Trend smoothing

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

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).

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

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 generalize 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].

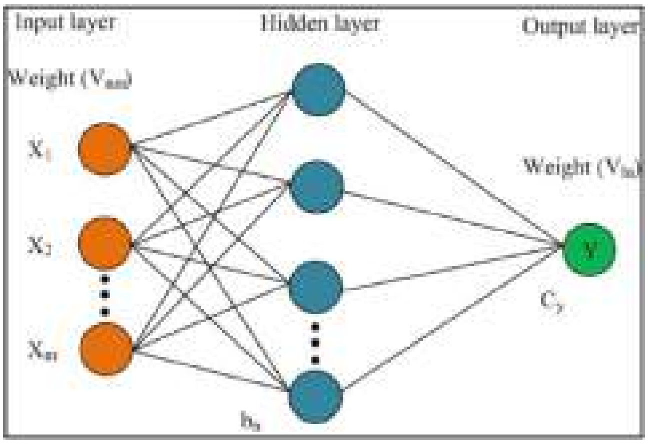


Fig. 1. Neural network structure.

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	Nothings 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).

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

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

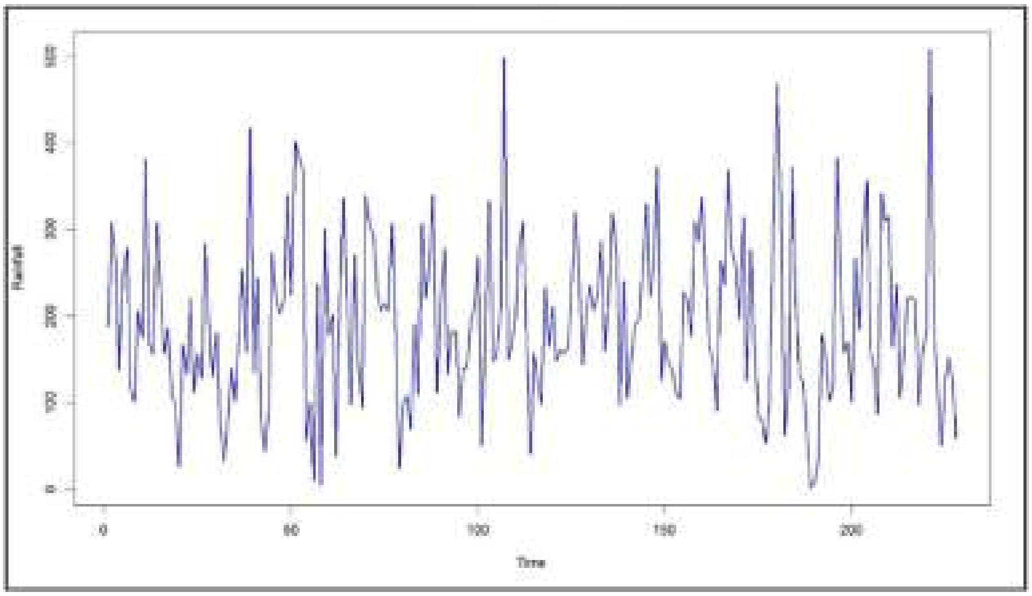


Fig. 2. Time series plot of rainfall data in Samarinda.

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	115.28	0.7	0.1	114.66
0.1	0.2	168.88	0.4	0.2	114.72	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	120.05	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	138.89	0.4	0.6	125.13	0.7	0.6	133.49
0.1	0.7	140.94	0.4	0.7	127.14	0.7	0.7	137.81
0.1	0.8	143.61	0.4	0.8	129.14	0.7	0.8	142.23
0.1	0.9	146.54	0.4	0.9	131.40	0.7	0.9	146.69
0.2	0.1	138.03	0.5	0.1	113.38	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	124.69	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	130.18	0.5	0.6	126.01	0.8	0.6	138.95
0.2	0.7	133.68	0.5	0.7	129.02	0.8	0.7	143.98
0.2	0.8	138.00	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	120.05
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	129.47
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	125.59	0.9	0.5	140.34
0.3	0.6	127.90	0.6	0.6	129.19	0.9	0.6	146.16
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	136.90	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
$\alpha = 0.4$ and $\beta = 0.2$	114.72	122.92
$\alpha = 0.5$ and $\beta = 0.1$	113.38	127.50
$\alpha = 0.6$ and $\beta = 0.1$	113.40	121.13
$\alpha = 0.7$ and $\beta = 0.1$	114.66	118.03

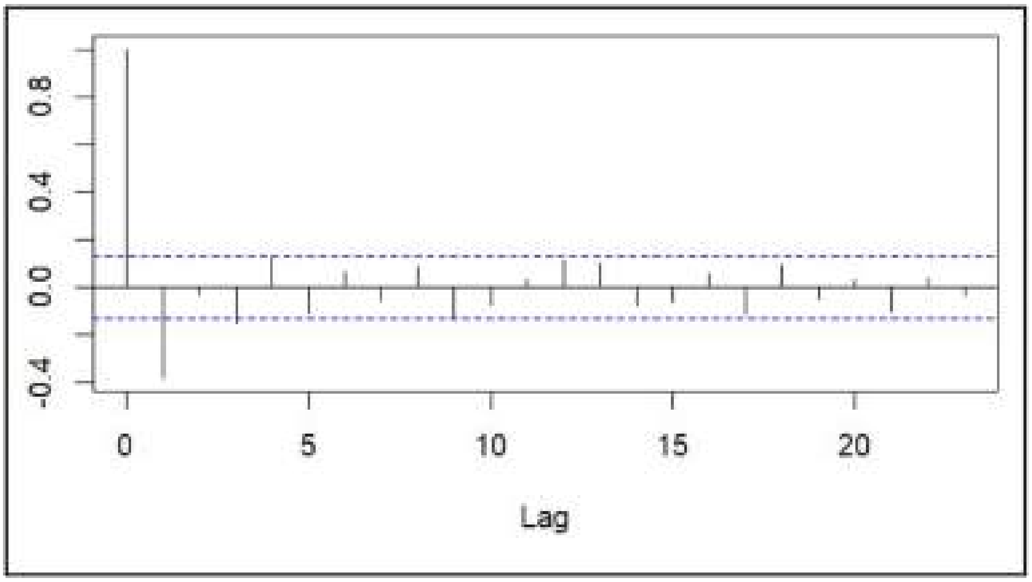


Fig. 3. ACF plot of rainfall data results of differencing.

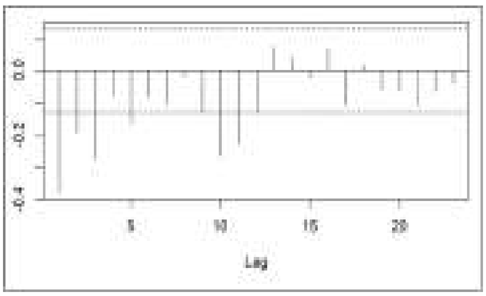


Fig. 4. PACF plot of rainfall data results of differencing.

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)	$\hat{\phi}_1$	-0.380766	5.579e-10	Significant
ARIMA(2,1,0)	$\hat{\phi}_1$	-0.455156	2.982e-12	Significant
	$\hat{\phi}_2$	-0.193967	0.002867	
ARIMA(3,1,0)	$\hat{\phi}_1$	-0.507994	1,760e-15	Significant
	$\hat{\phi}_2$	-0.320263	3.421e-06	
	$\hat{\phi}_3$	-0.275761	1.507e-05	
ARIMA(0,1,1)	$\hat{\theta}_1$	-1,000000	< 2.2e-16	Significant
ARIMA(1,1,1)	$\hat{\phi}_1$	0.273909	2.071e-05	Significant
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(2,1,1,)	$\hat{\phi}_1$	0.266500	6.351e-05	Not Significant
	$\hat{\phi}_2$	0.028447	0.6701	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	
ARIMA(3,1,1)	$\hat{\phi}_1$	0.268049	5.375e-05	Not Significant
	$\hat{\phi}_2$	0.048998	0.475	
	$\hat{\phi}_3$	-0.081457	0.220	
	$\hat{\theta}_1$	-1,000000	< 2.2e-16	

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

Proportion	RMSEP
Data Training	92.75
Data Testing	80.82

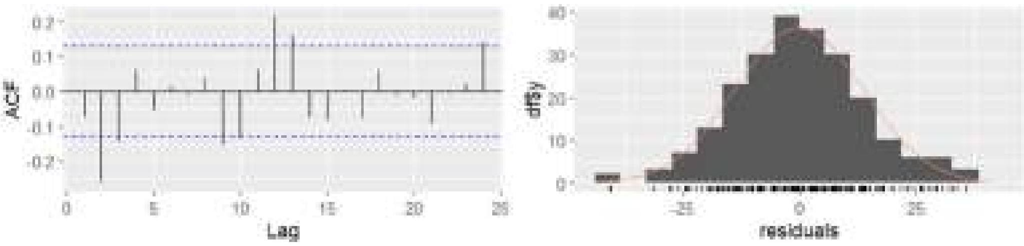


Fig. 5. Independence and Normality of ARIMA Residuals (1,1,0).

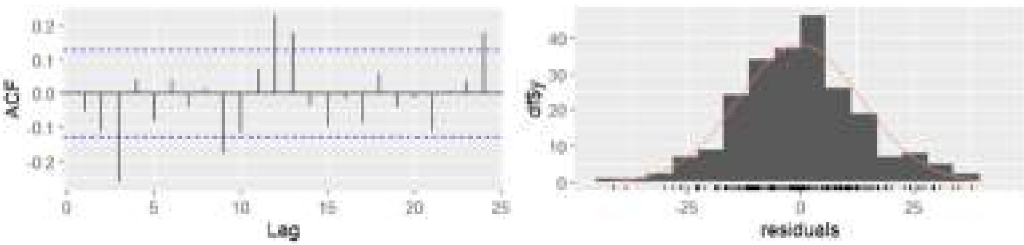


Fig. 6. Independence and Normality of ARIMA Residuals (2,1,0).

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).

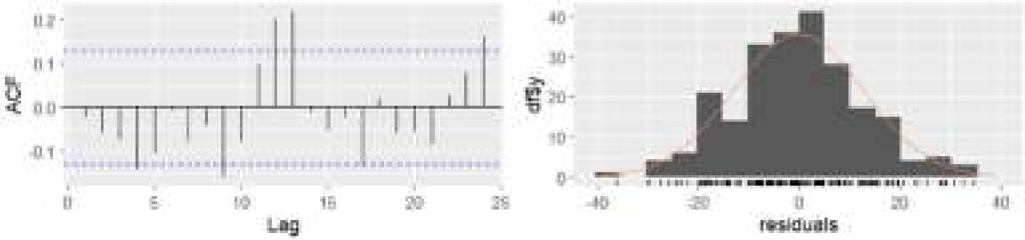


Fig. 7. Independence and Normality of ARIMA Residuals (3,1,0).

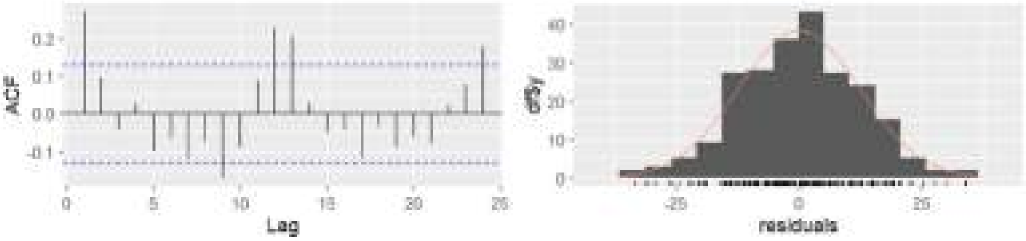


Fig. 8. Independence and Normality of ARIMA Residuals (0,1,1).

- 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 rate 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 89.749. Some of the architectural results of the NN modeling can be seen in Fig. 11.

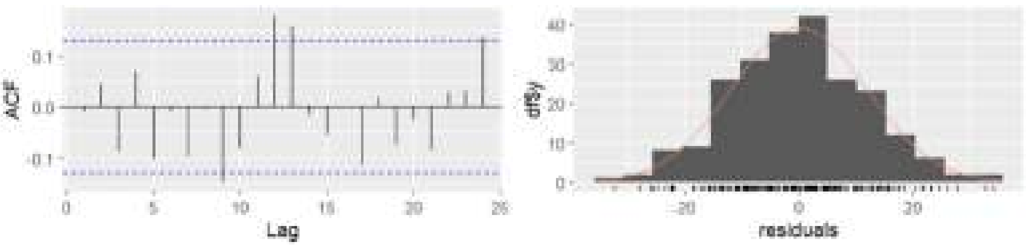


Fig. 9. Independence and Normality of ARIMA Residuals (1,1,1).

Table 6
RMSE calculation rsults.

Hidden Layer		RMSEP		Difference between Training and Testing RMSEP
Hidden Layer1	Hidden Layer2	Training	Testing	
1 neuron		88.988	86.595	2.393
2 neurons		85.018	90.197	5.180
3 neurons		78.101	110.632	32.531
4 neurons		75.907	100.158	24.251
5 neurons		66.340	97.531	31.192
6 neurons		68.059	96.647	28.589
7 neurons		69.542	103.764	34.222
8 neurons		58.299	102.938	44.639
9 neurons		57.471	114.915	57.444
10 neurons		46.560	110.919	64.359
2 neurons	1 neuron	83.746	82.969	0.777
3 neurons	2 neurons	77.546	100.939	23.393
4 neurons	3 neurons	70.852	97.184	26.332
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	135.889	106.194
9 neurons	8 neurons	27.501	116.387	88.886
10 neurons	9 neurons	17.891	181.869	163.978

Table 7
Model goodness of fit measure.

Method	RMSEP	
	Training	Testing
DES Holt ($\alpha = 0.7$ and $\beta = 0.1$)	114.66	118.03
ARIMA (1,1,1)	92.75	80.82
NN 2 HL (6–5 neurons)	**52.473**	**89.749**

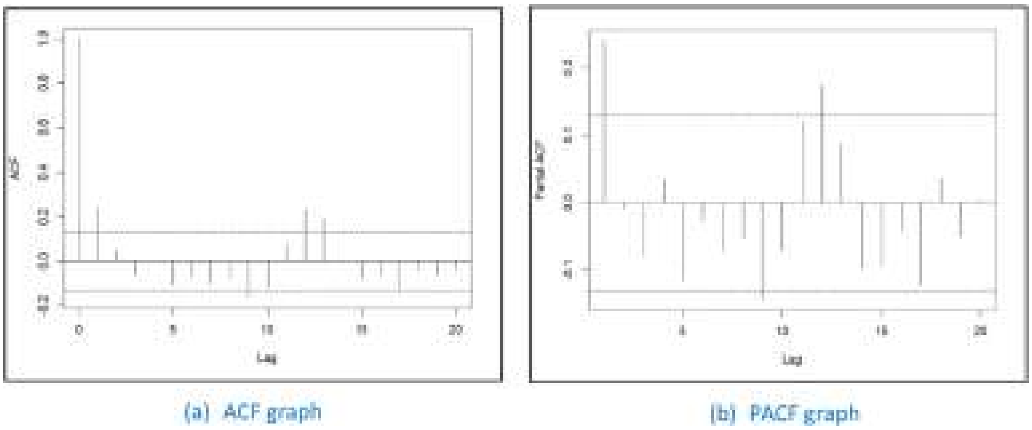


Fig. 10. ACF and PACF graphs of rainfall data in Samarinda.

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.

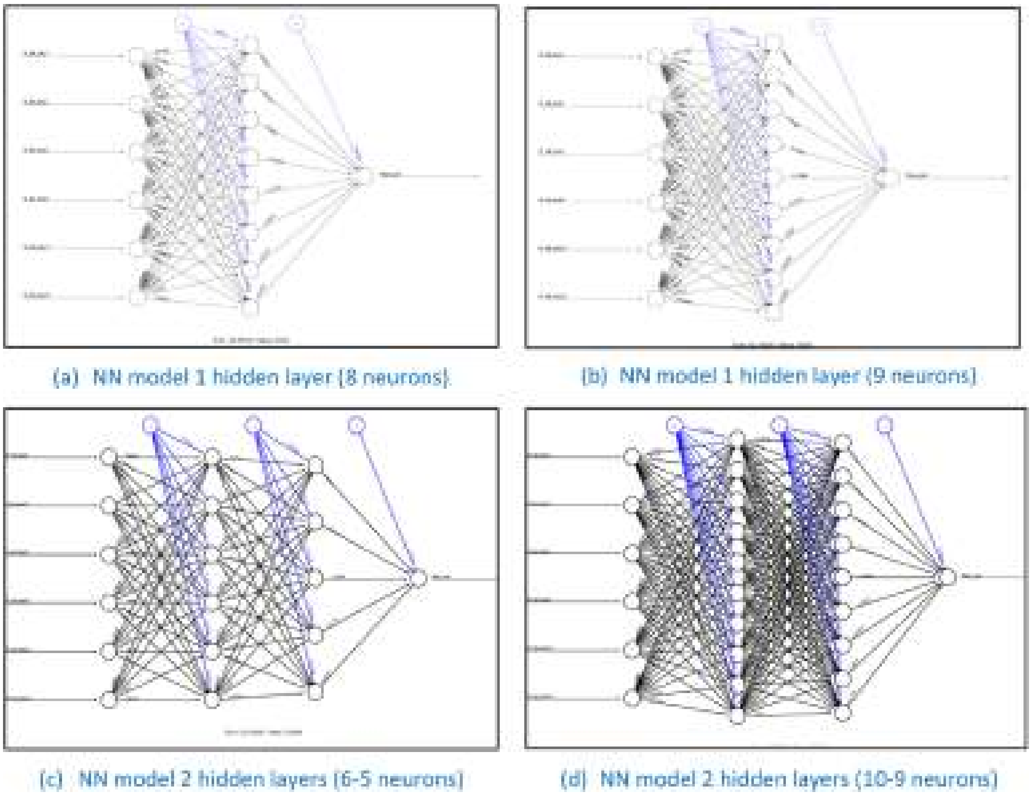


Fig. 11. Some architectures of NN modeling.

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

Month	Prediction Results
January	301.935
February	245.819
March	168.822
April	82.964
May	173.226
June	209.721
July	107.147
August	204.006
September	201.430
October	133.406
November	207.887
December	206,919

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,

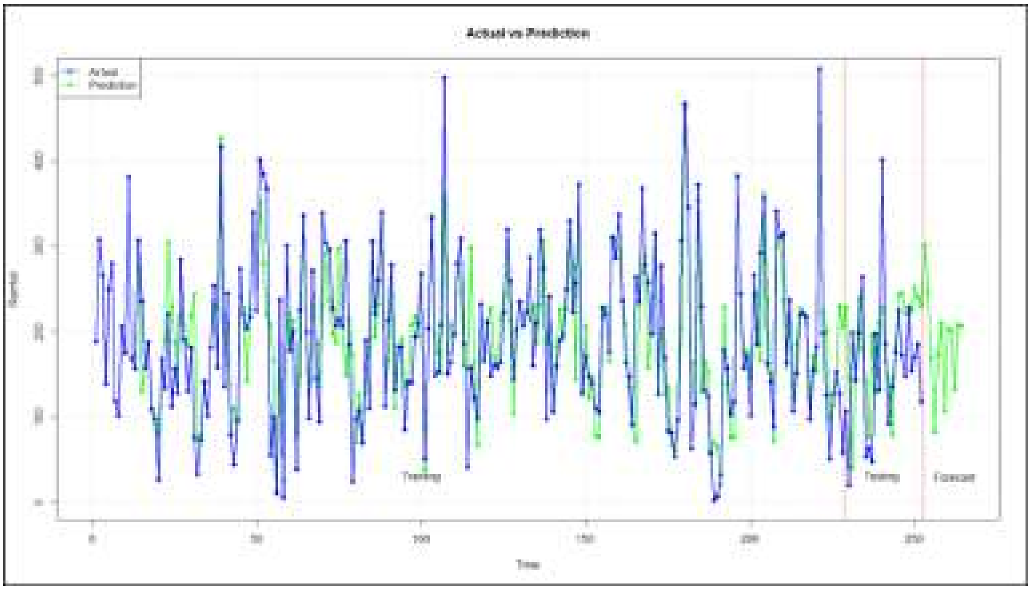


Fig. 12. Comparison plot of actual and predicted data.

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.P. 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 sangkuliman using an artificial neural network backpropagation algorithm, JTAM (Jurnal Teori dan Aplikasi Matematika) 8 (2) (2024) 465 Mar, doi:10.31764/jtam.v8i2.20112.

[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.Puspito Mustaruddin, B.Z. Zanlee, T. Nugroho, 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. Atesongun, M. Gulsen, 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 calender 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.