pijmath
Pattimura International Journal of Mathematics (PIJMath)
Volume 1, issue 2, pages 89-100, November 2022
E-ISSN : 2830-6791
RESEARCH ARTICLE ◙ *OPEN ACCESS*
DOI: https://doi.org/10.30598/pijmathvol1iss1pp89-100

Forecasting The Composite Stock Price Index Using Autoregressive Integrated Moving Average Hybrid Model Artificial Neural Network

Muhidin Jaariyah[1], Lexy Janzen Sinay[2*], Norisca Lewaherilla[3], Yopi Andry Lesnussa[4]

[1,2,3]*Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Pattimura*
[4]*Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Pattimura*
[2]*Mathematics Laboratorium, Faculty of Mathematics and Natural Sciences, Universitas Pattimura*
Jl. Ir. M. Putuhena, Ambon, 97233, Maluku, Indonesia

*Corresponding author's e-mail: * lexyjz@gmail.com*

ABSTRACT

Article History

Received :23rd October 2022
Revised : 30th October 2022
Accepted :1st November 2022

Keywords

Artificial Neural Network; Autoregressive Integrated Moving Average; Backpropagation; Model Hybrid ARIMA-ANN; IHSG; Forecasting

A stock index is a statistical measure that reflects the overall price movement of a group of stocks selected based on certain criteria and methodologies and evaluated regularly. JCI is included in the composite index, which is the Headline index. The Headline Index is an index that is used as the main reference to describe the performance of the capital market. The JCI is very important in describing the current condition of the capital market because the JCI measures the price performance of all stocks listed on the Main Board and Development Board of the IDX. This study aims to predict JCI data using the time series method. The hybrid Autoregressive Integrated Moving Average–Artificial Neural Network (ARIMA-ANN) model combines the linear ARIMA model and the non-linear ANN model. The best models are the ARIMA model (2,1,1) and the ANN Backpropagation model with one input layer, one hidden layer with 20 neurons, and one output. The ARIMA-ANN hybrid model accurately predicts JCI data because it produces a MAPE value of less than 1%, with the level of forecasting accuracy from testing results being smaller than the level of accuracy during training. In addition, the forecast for the next five days is very accurate because it produces a very small RMSE and a MAPE below 1%, respectively, namely 56.99 and 0.72%.

[1]*How to cite this article:*
M. Jaariyah, L. J. Sinay, N. Lewaherilla, and Y. A. Lesnussa, "Peramalan Indeks Harga Saham Gabungan Menggunakan *Hybrid Autoregressive Integrated Moving Average* (ARIMA) – *Artificial Neural Network* (ANN)", *Pattimura Int. J. Math. (PIJMATH).*, vol. 1, iss. 2, pp. 89-100, November, 2022.

1. Introduction

Stocks are one of the most popular financial market instruments. Issuance of shares by a company is one option for funding the company. On the other hand, stocks can provide an attractive profit level so that they become an investment instrument many investors choose. Shares are defined as a sign of capital participation of a person or party (business entity) in a company or limited liability company. The party who invests (involves capital) has a claim on the company's income and claims on company assets and is entitled to attend the General Meeting of Shareholders (GMS). [1]

A stock index is a statistical measure that reflects the overall price movement of a group of stocks selected based on certain criteria and methodologies and evaluated regularly. The purpose and benefits of the stock index are 1) measuring market sentiment; 2) used as passive investment products such as Index Mutual Funds and Index ETFs as well as derivative products; 3) benchmarks for active portfolios; 4) proxy in measuring and modeling return on investment, systematic risk, and risk-adjusted performance; 5) Proxy for asset class on asset allocation. [2]

Today, the development of passive investment is increasing globally, and the stock index can be used as a reference for passive investment. Based on data from the Investment Company Institute in the 2010-2019 period, passive investment (index mutual funds and ETFs) in the United States increased by USD 1.8 trillion, while active investment (non-index mutual funds) decreased by USD 1.7 trillion. In Indonesia, the use of the IDX index for investment products increased significantly, from IDR 2.72 trillion in funds under management at the end of 2015 to IDR. 15.88 trillion by the end of 2020. In other words, the average growth is 42% per year. [2]

The Indonesia Stock Exchange (IDX), as the regulator and organizer of trading in the Indonesian capital market, provides a stock index that capital market players can use. In February 2021, there are 37 stock indices provided by the IDX, which are classified into four index categories, namely Headline (10 indexes), Sector (13 indexes), Thematic (7 indexes), and Factor (7 indexes). One of the indexes available on the IDX is the Composite Stock Price Index (IHSG) or JCI (Jakarta Composite Index). JCI is included in the composite index, which is the Headline index. The Headline Index is an index that is used as the main reference to describe the performance of the capital market. [2]

JCI was first introduced in April 1983. At that time, JCI was used as a reference for price movements of all shares listed on the IDX, both regular and preferred shares [3]. In addition, the JCI is usually used as a proxy for the market portfolio, which can then be used to calculate a portfolio's systematic risk and risk-adjusted-performance [2]. Thus, the JCI is very important in describing the current condition of the capital market because the JCI measures the price performance of all stocks listed on the Main Board and Development Board of the IDX [2]. Based on IDX data as of January 29, 2021, there are 713 constituents in the JCI. Return data for the last five years shows that the return (yoy) of the JCI has decreased significantly. In 2017, the JCI return reached 20%. In 2021, the return was -2%. The lowest JCI return condition occurred in 2020, which reached -5.1% [2]. Thus, it is important to observe and analyze JCI price movements using the right statistical model.

Generally, JCI data is presented in the form of time series. Therefore, the time-series data modeling method is the right method. Based on the number of variables used, the time series model is classified into two, namely the univariate and multivariate time series models. Several previous studies have carried out the application of multivariate and univariate time series models in analyzing JCI data. The multivariate time series model used in previous research is a combination model between the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model and the Threshold Autoregressive Conditional Heteroscedasticity (TARCH) or ARIMAX-TARCH model [4]. Meanwhile, the time series model used by previous research is the Autoregressive Integrated Moving Average (ARIMA) [5] [6] [7]. Only now has research that compares multivariate and univariate time series models using JCI data. In a different case, some studies compare multivariate and univariate time series models, where the forecasting results using univariate time series models are better than multivariate models. [8] [9] [10].

The univariate time series model is a model that analyzes time series data consisting of one variable. ARIMA is a linear univariate time series model, so the ability to explain data that contains nonlinear components is not optimal. Nonlinear components must be analyzed using appropriate methods. It aims to minimize errors in predictions or avoid invalid prediction results. One method that can be used to analyze data containing nonlinear components is Artificial Neural Network (ANN). According to a review of several previous studies, there is the use of ANN algorithms for forecasting JCI data, such as the prediction of JCI using the backpropagation algorithm [11] and the use of the ANN algorithm without windowing to predict JCI more accurately because it produces smaller errors than using windowing [12]. Until now, there has been no research that discusses the prediction of the JCI by combining linear and nonlinear models.

Based on the description above, this study aims to model the JCI data by analyzing linear and nonlinear components in one model. The model is a hybrid model between ARIMA and ANN, where ARIMA is used to model linear components, while ANN is used to model nonlinear components [13] [14]. In its application, the ARIMA-ANN hybrid model produces accurate predictions [13]. Until now, the application of ARIMA-ANN is quite extensive in various fields, such as forecasting rainfall in Malang Regency, Indonesia [15], Banyuwangi [16], Robusta coffee price forecasting in India [17], inflation prediction in Indonesia[18], forecasting agricultural exchange rate data in Gorontalo

Indonesia [19], forecasting the IDR exchange rate against USD [20], and estimation of electromagnetic wave propagation in densely forested urban areas [21].

2. Research Methods

2.1 ARIMA

Autoregressive Moving Average (ARMA) is a combination of autoregressive (AR) and moving average (MA) models. The model is a model for stationary data which is denoted by $ARMA(p,q)$, which is defined as follows

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{1}$$

with $\phi_1, \ldots, \phi_p$ is the autoregressive coefficient, $\theta_1, \ldots, \theta_q$ is the coefficient of moving average, and ε_t is not correlated with $y_{t-1}, y_{t-2}, y_{t-3}, \ldots$. Using the backshift operator, equation (1) is expressed as follows.

$$\phi(B) y_t = \theta(B) \varepsilon_t \tag{2}$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$
$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$$

Process y_t is said to be $AR\ (p)$ if $\theta(B) = 1$, or it is said as $MA\ (q)$ if $\phi(B) = 1$.

In 1976, Box and Jenkins [22] introduced the Autoregressive Integrated Moving Average (ARIMA) model and forecasting using univariate time series data. The ARIMA model is an extension of the ARMA model. This model is a model for non-stationary data which is denoted by $ARIMA\,(p,d,q)$. The process y_t is called the ARIMA process (p,d,q) if $W_t = \nabla^d y_t = (1-B)^d y_t$ is a stationary $ARMA\,(p,q)$ process. In general, the model can be written as follows

$$\phi(B)\,(1 - B)^d y_t = \theta(B) \varepsilon_t \tag{3}$$

where d is the number of references that make the time series data stationary. The ARIMA modelling procedure is given as follows [23] :

(1) Identifying stationary using ACF/PACF plots and KPSS test statistics [24].
(2) Identifying the ARIMA model's order using the ACF/PACF plot.
(3) Estimating ARIMA model using Maximum Likelihood Estimator (MLE).
(4) Selecting the best ARIMA model based on significant parameter coefficients using *t*-test statistics, the smallest AIC value [25].
(5) Analyzing the residual of the selected ARIMA model and whether it meets the white noise assumption using the Ljung-Box test statistic.
(6) Forecasting

2.2 ANN

The artificial neural network (ANN) method was first devised by Warren McCulloch and Walter Pitts in 1943. They wrote papers on how neurons could work and modelled their ideas by creating simple neural networks using electrical circuits [26]. This finding is a breakthrough for sharing research on artificial neural networks, including the backpropagation algorithm, which was first introduced by Paul J. Werbos in 1974 [27], [28]. In 1975, Kunihiko Fukushima constructively developed a multi-layered artificial neural network [29].

The structure of the nervous network consists of synapses. Each synapse is assigned a weight indicating the corresponding neuronal effect, and all data passes through the neural network as a signal. Two things process signals: first, by an integration function that combines all incoming signals; second, by an activation function that alters the neuron's output. In simple terms, a multi-layer perceptron (MLP) consisting of *n* inputs and 1 outputs, the mathematical equation is

$$y_t = f\left(\alpha_0 + \sum_{i=1}^{n} \alpha_i y_{t-i}\right) = f(\alpha_0 + \boldsymbol{\alpha}^T \mathbf{y}) \tag{4}$$

where α_0 is an intercept, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ is a vector consisting of all synaptic weights without intercept, and $\mathbf{y} = (y_{t-1}, \ldots, y_{t-n})$ is a vector of all covariates [30]. To increase flexibility in modeling, a hidden layer is needed. MLP with hidden layer is shown as follows

$$y_t = f\left(\alpha_0 + \sum_{j=1}^{J} \alpha_i \cdot f\left(\beta_{0j} + \sum_{i=1}^{n} \beta_{ij} y_{t-i}\right)\right) = f\left(\alpha_0 + \sum_{j=1}^{J} \alpha_i \cdot f\left(\beta_{0j} + \boldsymbol{\beta}_j^T \mathbf{y}\right)\right) \quad (5)$$

Where α_0 is the intercept of the output neuron, β_{0j} is the intercept of the j-th hidden neuron, α_j is the synaptic weight corresponding to the synapse starting from the j-th hidden neuron to the output neuron, $\boldsymbol{\beta}_\mathbf{j} = (\beta_{1j}, \ldots, \beta_{nj})$ is a vector of all synaptic weights corresponding to the synapse pointing to the j-th hidden neuron [30].

ANN is a data analysis method that produces a model that matches the target data (the smallest error) through a learning algorithm in a training process. Backpropagation is a popular learning algorithm. The algorithm uses gradient descent as a learning mechanism. The backpropagation algorithm calculates the network weight, makes small changes, and gradually makes adjustments determined by the error between the results obtained by the network and the desired results (minimizing errors). In other words, backpropagation modifies the weight of the neural network gradually to find the local minimum of the error function using the gradient descent technique [30, 31], where the learning algorithm is given through the following steps:

(1) Network initialization, is the process of determining the initial weight, generally using random weights.
(2) Feed Forward is the process of forwarding information in the network where information is passed from the input to the hidden layer and output through the nodes of the activation function and weights.
(3) Error assessment is the process of adjusting the output generated by the network, which must match the targeted output. The process depends on the error generated.
(4) Propagation is the process of modifying the weights based on errors in the output layer. It is the process of backward propagation of error through the network and calculating the gradient of the error change that corresponds to the change in the weight value.
(5) Adjustment is the process of adjusting the weights using a gradient of change to minimize errors. A factor adjusts each neuron's weight and bias based on the activation function's derivative, the difference between the network output and the actual target and neuron output. It is a network learning process.

2.3 Hybrid ARIMA-ANN

There is a time series model like the following:

$$y_t = L_t + N_t \quad (4)$$

where L_t is a linear component, N_t is a nonlinear component [13]. The linear component of L_t is estimated using ARIMA so that the residual is obtained. Suppose e_t is the model's residual at time *t*, then

$$e_t = y_t - \hat{L}_t \quad (5)$$

where $\hat{L}_t$ is the forecast value at time. Residual plays an important role in diagnosing the suitability of the linear model. The performance of the ARIMA model has limitations. If there is a significant nonlinear pattern in the residual e_t, one way to overcome this is to use an artificial neural network approach. It aims to find a nonlinear relationship in the model to improve the linear model's performance (ARIMA). ANN model for residuals with *n* input nodes

$$e_t = f(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \varepsilon_t \quad (6)$$

where f is a non-linear function obtained from the neural network approach and ε_t is a random error. If the forecasting result of equation (6) is denoted by $\hat{N}_t$, so

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (7)$$

Equation (7) shows a hybrid system consisting of ARIMA modeling to analyze linear components and neural network modeling using residuals from the ARIMA model. The ARIMA model residual is a nonlinear component that cannot be explained by the ARIMA model [13]. The procedure performed after obtaining the residual ARIMA model is to perform a linearity test using the Terasvirta test statistic and identification of the lag using a lag plot.

3. Results And Discussion

3.1 About IHSG

The data used was JCI data for the period January 2, 2018, to March 31, 2022, sourced from Yahoo Finance's publication [32]. The data is adjusted daily for the number of IDX working days, which is 1034 days. The JCI data can be visually displayed as a line plot, as shown in Figure 1.

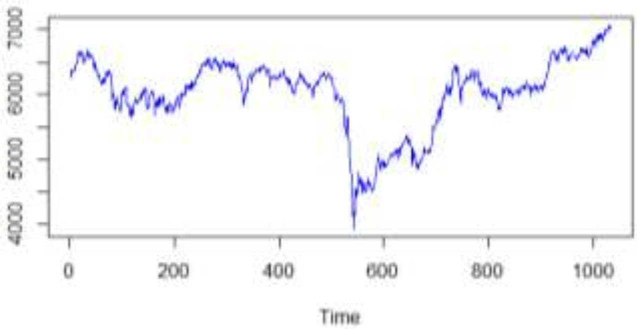


Figure 1. JCI on 2 January 2018 – 31 March 2022 (data in IDR)

Figure 1 shows that the JCI in that period experienced fluctuations, which reached a minimum price of 3,937.63 on March 24, 2020. The decline in the value of the JCI occurred during the COVID-19 pandemic in Indonesia. The decline in value formed a trend in the JCI data with a minimum value as a turning point. The average JCI in that period was 6,016.96. The increase in the JCI value exceeding the average value occurred on December 16, 2020, until it reached the highest value for the JCI in that period of 7,071.44, which occurred on March 31, 2022.

3.2 ARIMA Model

Stationary Identification

Figure 1 shows that the JCI data for the period 2 January 2018 – 31 March 2022 contains a trend so that the data is not stationary. Therefore, the ACF plot and the KPSS test in Table 1 below can be used to show clearly check the stationarity of the JCI data.

Table 1. Plot of ACF/PACF and KPSS test for stationarity

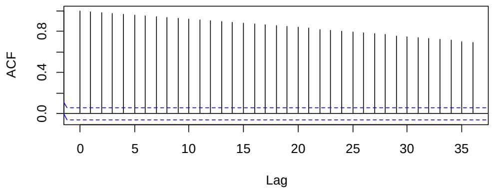
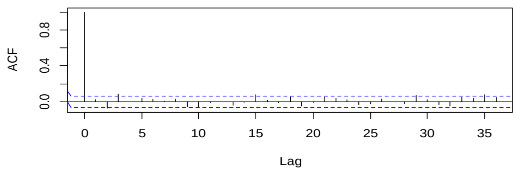
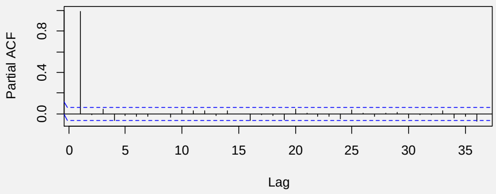
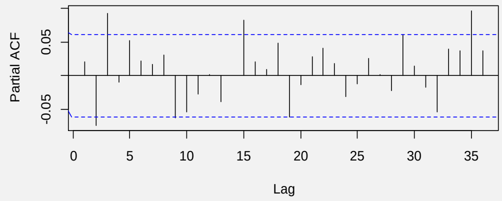
Tools	Data	
	Level	**1st Difference**
ACF		
PACF		
KPSS Test	Statistic Test: 1.5117 p value: 0.01	Statistic Test: 0.1954 p value: 0.1

Table 1 shows two ACF plots, namely for the actual JCI data (level conditions) and the results of the first difference from the JCI data. ACF plots of data in the level conditions decrease slowly. It means that there is a trend at the data level. On the other hand, the ACF plot does not decrease slowly. It indicates that the data from the first differentiation is stationary while the level data is not stationary.

The KPSS test was used to check the stationarity of the data to confirm the results shown by the ACF plot. Based on Table 1, the KPSS test statistic for level data is 1.5117 with a p-value of 0.01. The data level is not stationary because the p-value is less than the 5% significance level. Meanwhile, the statistical test of the JCI value after performing the first differentiation process was 0.1954 with a p-value of 0.1. It shows that the results of the first difference are stationary because the p-value is greater, with a significance level of 5%.

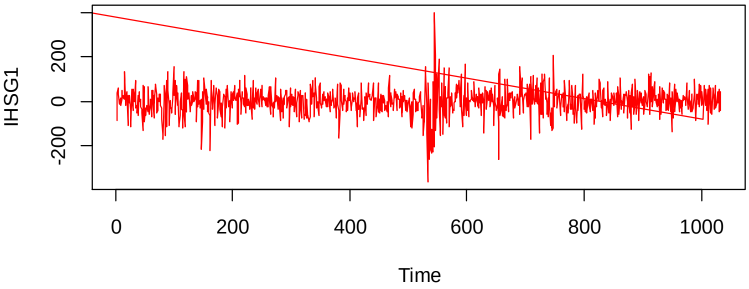


Figure 2. 1st Difference of JCI

b. ARIMA Model Estimation and Selection

The ARIMA model parameter estimation used the MLE method. The determination of the ARIMA order is based on the lag analysis on the ACF/PACF plot from the results of the first difference (Table 1), where eight possible models can be generated. The estimation results of the eight ARIMA models are summarized in Table 2.

Table 2. ARIMA Model Estimation Results

Model	Parameter	Estimation	Statistic t	*p value*	AIC
ARIMA (0,1,1)	θ_1	0.02474	0.7283	0.4664	11428.54
ARIMA (0,1,2)	θ_1	0.03866	1.2274	0.2196	11423.98
	θ_2	0.08096	−2.5839	0.0097	
ARIMA (1,1,0)	ϕ_1	0.02092	0.6721	0.5015	11428.62
ARIMA (1,1,1)	ϕ_1	−0.66891	−5.9673	0.0000	11422.43
	θ_1	0.73740	7.3028	0.0000	
ARIMA (1,1,2)	ϕ_1	−0.48804	−3.1937	0.0014	11420.25
	θ_1	0.22499	3.4333	0.0006	
	θ_2	−0.07373	−2.2084	0.0272	
ARIMA (2,1,0)	ϕ_1	0.02257	0.7269	0.46726	11424.93
	ϕ_2	−0.07426	−2.3883	0.01692	
ARIMA (2,1,1)	ϕ_1	−0.56105	−4.1325	0.0000	11419.62
	ϕ_2	−0.07834	−2.3131	0.0207	
	θ_1	0.59140	4.4279	0.0000	
ARIMA (2,1,2)	ϕ_1	−1.28347	−2.4670	0.0136	11419.26
	ϕ_2	−0.65551	−1.6976	0.0895	
	θ_1	1.32282	2.4764	0.0132	
	ϕ_1	0.63181	1.4167	0.15656	

Table 2 shows the estimation results of the eight ARIMA models to select the best ARIMA model. The selection of the best ARIMA model was determined based on the parameter significance test and the information criteria value. The parameter significance test uses the *t-test*, and the information criteria use the AIC value. Based on Table 2, the best model is ARIMA (2,1,1). This model has a significant coefficient at a significance level of 5% and has the smallest AIC value compared to other models.

ARIMA Model Residual Analysis

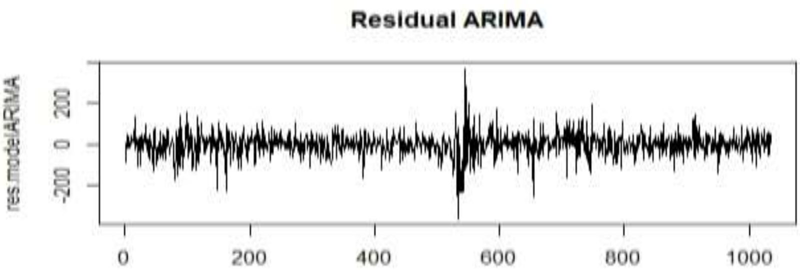


Figure 3. Residual ARIMA (2,1,1)

Visually, the residuals of the ARIMA model (2,1,1) are shown in Figure 3. One of the important assumptions that the best ARIMA model must meet is that the residuals must be white noise. The test used for this assumption was the Ljung-Box test. The test aims to see the serial correlation in the residual ARIMA model (2,1,1).

Table 3. ARIMA Model Residual Assumption Test (2,1,1)

Ljung-Box test		
Q Stat	**df**	**p-value**
20.494	33	0.1987

The results of the Ljung-Box test for the ARIMA model residuals (2,1,1) are shown in Table 3. The Ljung-Box test results show no serial correlation in the ARIMA residuals (2,1,1) because the p-value is greater than the 5% significance level. In other words, the ARIMA model residual (2,1,1) fulfills the assumption of white noise.

Table 4. Nonlinearity Test

Terasvirta Test		
χ^2	**df**	**p-value**
77.651	16	0.0000

The results of the nonlinearity test on the residual ARIMA model (2,1,1) shown in Table 4 show that the p-value of the Terasvirta test statistic is less than the 5% significance level. It means that the residual ARIMA model (2,1,1) is not linear. Thus, the residual ARIMA model (2,1,1) is a nonlinear component.

3.3 Artificial Neural Network (ANN) Model

ANN modeling uses the residual ARIMA model (2,1,1), a nonlinear component. The number of observations is 1034 data. It is divided into two processes: training and testing. The percentage of the distribution is 80% for the training process, which is 827 data, and 20% for the testing process, which is 207 data.

Training Results

The first step in this process was to evaluate the pattern of relationships between lags in the training data using a lag plot. The lag plot for lag 1 to lag 16 is shown in Figure 4.

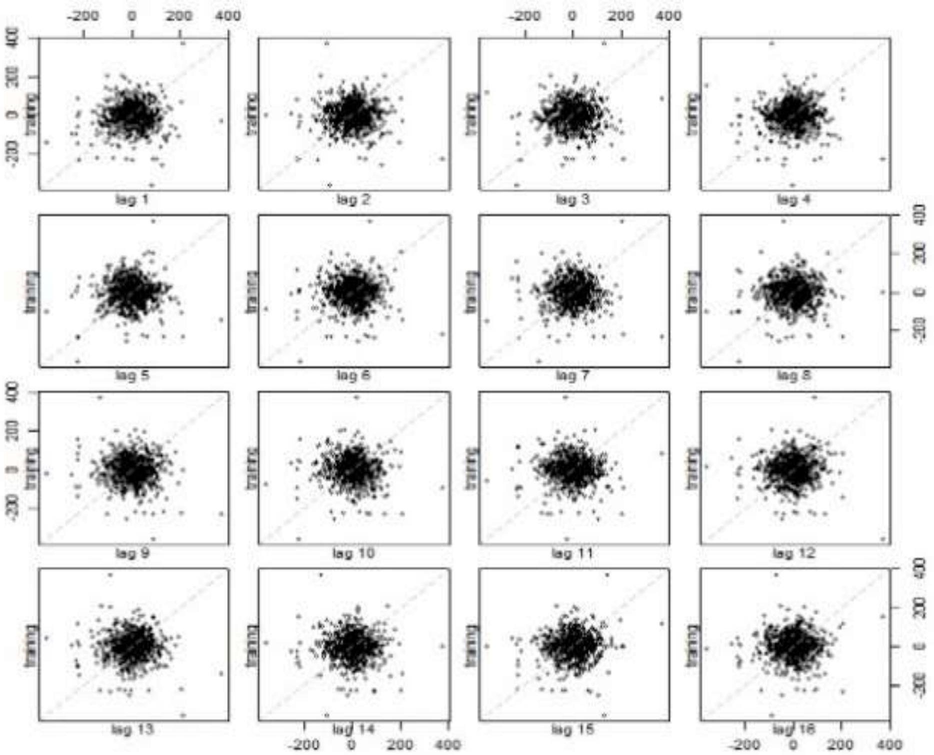


Figure 4. Lag plot of Training Data

Figure 4 shows that the lag plot for lag 1 to lag 16 forms a pattern, which tends to be in the middle and forms a cyclic (sinusoidal) shape. In addition, Figure 4 also identifies outliers in the lag plot for lag 1 to lag 16.

The next step was to standardize the training data to avoid large data variations. Then, the standardized data is trained using the backpropagation algorithm by applying one input layer and one output layer. Based on the characteristics of the data, the activation function used is tanh. The learning rate parameter used is 0.01. The network structure of the ANN model uses an input layer, namely lag 1 to lag five, based on the number of active working days. The ANN network architecture combines one input layer, one to 20 neurons, and one output layer. Based on this architecture, 100 ANN models with the smallest RMSE are obtained, as shown in Table 5.

Table 5. RMSE Network Architecture Structure

Number of neurons in hidden layer	*Input Layer*				
	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Neuron 1	0.9996616	1.0007020	0.9996616	0.9996616	0.9996616
Neuron 2	1.0007000	1.0007000	1.0007020	0.9996604	1.0007000
Neuron 3	0.9996609	0.9996612	1.0007000	0.9996602	0.9996600
Neuron 4	1.0007010	1.0006990	0.9996587	1.0007020	0.9996615
Neuron 5	0.9956640	0.9996598	0.9996599	0.9996609	1.0007010
Neuron 6	0.9951827	1.0007020	1.0007010	1.0007000	0.9979665
Neuron 7	0.9996616	0.9996622	1.0006990	0.9996611	0.9996616
Neuron 8	1.0007020	0.9996591	0.9996616	0.9996611	0.9996615
Neuron 9	1.0007010	0.9996616	1.0007010	1.0007020	0.9996615
Neuron 10	1.0007000	0.9996618	0.9996614	0.9996615	0.9996601
Neuron 11	0.9996588	0.9996612	0.9996616	0.9989785	1.0007020
Neuron 12	1.0007010	0.9996616	0.9996616	0.9996614	1.0006990
Neuron 13	1.0007040	0.9996611	1.0007000	1.0007020	0.9996616
Neuron 14	0.9996615	0.9996627	0.9996616	1.0007010	0.9996598
Neuron 15	0.9996590	0.9996614	0.9996609	1.0007010	1.0007020
Neuron 16	0.9956632	0.9996608	1.0007020	0.9996605	0.9996616
Neuron 17	1.0007020	0.9996615	0.9996618	1.0007020	0.9996616
Neuron 18	1.0007010	0.9996610	0.9971569	0.9996616	1.0007020
Neuron 19	1.0007010	0.9996616	1.0007020	0.9996616	1.0007000
Neuron 20	0.9951771	0.9996616	1.0006990	0.9996616	0.9996619

Based on Table 5, the smallest RMSE value is 0.9951771, which is found in the ANN model with the input layer at lag one and the number of neurons 20, where the network architecture is shown in Figure 5. Figure 5 is the best network architecture, where lag 1 is the input layer with one hidden layer, 20 neurons, and one output layer.

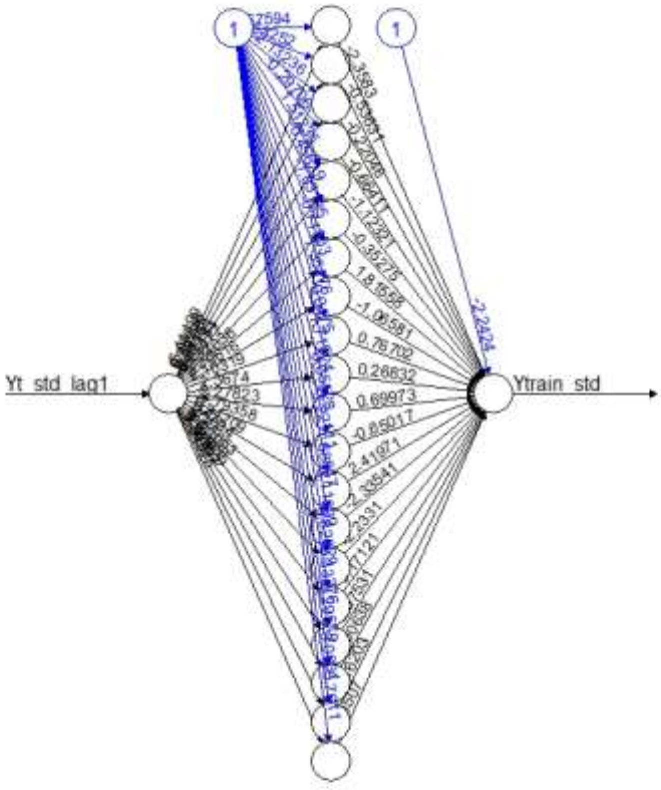


Figure 5. Best Network Architecture

Test result

The testing process aimed to recognize the data as well as to validate the built model and provide results with good accuracy with very small errors. The testing process is based on the best ANN model architecture obtained from the training process (Figure 5). The amount of data used in the testing process is 207. The accuracy of the results of the training and testing of the ANN model is shown in Table 5.

Table 6. Forecasting Accuracy

Data	RMSE
Training	63.7397
Testing	47.3166

Table 6 shows that the RMSE value of the testing results is smaller than the training results. It shows that the results of the training are effective because they can recognize the data that has been studied. Thus, the network architecture shown in Figure 5 is good to be combined with the ARIMA (2,1,1) model.

3.4 ARIMA-ANN Model

Based on the analysis above, the combination of ARIMA (2,1,1) and ANN models with one input layer, one hidden layer with 20 neurons, and one output layer is the best model, thus forming the ARIMA-ANN hybrid model. The model can be denoted in the form $y_t = \hat{L}_t + \hat{N}_t$ with $\hat{L}_t$ is the forecast result from ARIMA, which is a linear component and $\hat{N}_t$ is the forecast result from ANN, which is a nonlinear component. Visually, the comparison of the actual value and the forecast results of the ARIMA-ANN model is shown in Figure 6.

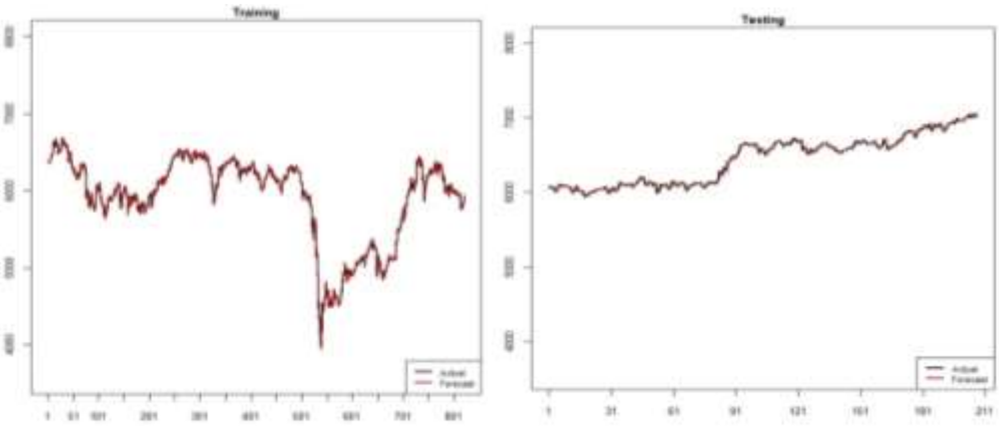


Figure 6. Comparison of Actual Data and Forecasting Data ARIMA-ANN Model

Figure 6 shows the value shown by the black line plot, while the forecast results from the ARIMA-ANN model are shown in the red line plot. Visually, the forecast results follow a pattern and merge with the actual training and testing data value. It indicates that the prediction results of the ARIMA-ANN model are accurate.

Table 7. ARIMA-ANN model MAPE value

Data	MAPE
Training	0.81%
Testing	0.58%

The comparison between the training and testing results based on the MAPE value is shown in Table 7. These results show that the MAPE value obtained from the ARIMA-ANN model is very small, less than 1%. The results in Table 7 also show that the ARIMA-ANN model accurately predicts JCI data, where the MAPE test results are smaller than the training results. Based on these results, the forecast results for the next five days are presented in Table 8.

Table 8. ARIMA-ANN Hybrid Forecast Results

Day / Date	**Forecast**	**Out-sample**
1-Apr-22	7,069.07	7,078.76
4-Apr-22	7,059.80	7,116.22
5-Apr-22	7,064.16	7,148.30
6-Apr-22	7,062.44	7,104.22
7-Apr-22	7,063.06	7,127.37

Table 8 shows the forecast results of the ARIMA-ANN model and the actual data outside the sample. These results show that the forecast value from April 1, 2022, to April 7, 2022, is close to the actual value with RMSE 56.99 and MAPE 0.72%.

4. Conclusions

This research applies the time series method in analyzing and predicting the JCI data. The model used was the ARIMA-ANN hybrid model. The combination of the ARIMA model and the ANN model aims to explain the linear and nonlinear forms contained in the JCI data. The resulting ARIMA-ANN hybrid model combines ARIMA (2,1,1) model and ANN Backpropagation model with one input layer, one hidden layer with 20 neurons, and one output. The ARIMA method is used to model the linear component of the actual JCI data, while the ANN model is used to model the nonlinear component of the residuals generated by the ARIMA model.

The ARIMA-ANN model is an effective model for forecasting JCI data because it has a very good level of accuracy. It is indicated by the resulting MAPE value of less than 1%, where the level of forecasting accuracy from the testing results is smaller than the level of accuracy during training. In addition, the forecast for the next five days is very accurate because it produces a very small RMSE and a MAPE below 1%, respectively. The two values are 56.99 and 0.72%.

References

[1] Indonesia Stock Exchange, “Bagaimana Menjadi Investor Saham,” [“How to Become a Stock Investor”] Jakarta, 2015. Accessed: Sep. 07, 2022. [Online]. Available: https://www.idx.co.id/Portals/0/StaticData/Information/ForInvestor/HowToBecomeAnInvestor/HTBAI-Ind-2015.pdf

[2] Indonesia Stock Exchange, “IDX Stock Index Handbook v1.2,” Jakarta, Jan. 2021. Accessed: Sep. 07, 2022. [Online]. Available: https://www.idx.co.id/media/9816/idx-stock-index-handbook-v12-_-januari-2021.pdf

[3] Indonesia Stock Exchange, “IDX Fact Book 2019,” Jakarta, 2019. [Online]. Available: www.idx.co.id

[4] E. Fauziyah, D. Ispriyanti, and T. Tarno, “PEMODELAN DAN PERAMALAN INDEKS HARGA SAHAM GABUNGAN (IHSG) MENGGUNAKAN ARIMAX-TARCH,” ["MODELING AND FORECASTING THE JOINT STOCK PRICE INDEX (JCI) USING ARIMAX-TARCH"] *Jurnal Gaussian*, vol. 10, no. 4, pp. 595–604, Dec. 2021, doi: 10.14710/j.gauss.v10i4.33102.

[5] P. Pardede, “PENERAPAN MODEL ARIMA DALAM MEMPREDIKSI INDEKS HARGA SAHAM GABUNGAN BURSA EFEK INDONESIA-JAKARTA,” ["APPLICATION OF THE ARIMA MODEL IN PREDICTING THE INDONESIA-JAKARTA JOINT STOCK PRICE INDEX"] *Jurnal Penelitian Manajemen*, vol. 1, no. 1, pp. 92–103, 2019, Accessed: Sep. 08, 2022. [Online]. Available: http://ojs.mputantular.ac.id/index.php/MPU/article/view/194/159

[6] R. Susanti and A. R. Adji, “ANALISIS PERAMALAN IHSG DENGAN TIME SERIES MODELING ARIMA,” ["JCI FORECASTING ANALYSIS WITH ARIMA TIME SERIES MODELING"] *Jurnal Manajemen Kewirausahaan*, vol. 17, no. 1, pp. 97–106, Jun. 2020, doi: 10.33370/jmk.v17i1.393.

[7] D. Gunawan and W. Astika, “The Autoregressive Integrated Moving Average (ARIMA) Model for Predicting Jakarta Composite Index,” *Jurnal Informatika Ekonomi Bisnis*, vol. 4, no. 1, pp. 1–6, Feb. 2022, doi: 10.37034/infeb.v4i1.114.

[8] J. du Preez and S. F. Witt, “Univariate versus multivariate time series forecasting: an application to international tourism demand,” *Int J Forecast*, vol. 19, no. 3, pp. 435–451, Jul. 2003, doi: 10.1016/S0169-2070(02)00057-2.

[9] I. A. Iwok and A. S. Okpe, “A Comparative Study between Univariate and Multivariate Linear Stationary Time Series Models,” *American Journal of Mathematics and Statistics*, vol. 6, no. 5, pp. 203–212, 2016, doi: 10.5923/j.ajms.20160605.02.

[10] D. Oktasari, “Perbandingan Model ARIMA Univariat dengan Multivariat Untuk Menduga Pola Absorbsi pada Stronsium Titanat Variasi Dop,” ["Comparison of Univariate and Multivariate ARIMA Models to Predict Absorption Patterns in Strontium Titanate Dop Variations"] Bogor Agricultural University (IPB), Bogor, 2018.

[11] A. Santoso and S. Hansun, “Prediksi IHSG dengan Backpropagation Neural Network,” ["JCI Prediction with Backpropagation Neural Network"] *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, pp. 313–318, Aug. 2019, doi: 10.29207/resti.v3i2.887.

[12] I. Halimi and W. A. Kusuma, “Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma Neural Network,” ["Prediction of the Composite Stock Price Index (JCI) Using the Neural Network Algorithm"] *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 4, no. 1, pp. 24–29, Jun. 2018, doi: 10.26418/jp.v4i1.25384.

[13] G. P. Zhang, “Time series forecasting using a hybrid ARIMA and neural network model,” *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003, doi: 10.1016/S0925-2312(01)00702-0.

[14] L. Wang, H. Zou, J. Su, L. Li, and S. Chaudhry, “An ARIMA-ANN Hybrid Model for Time Series Forecasting,” *Syst Res Behav Sci*, vol. 30, no. 3, pp. 244–259, May 2013, doi: 10.1002/sres.2179.

[15] B. A. Safitri, A. Iriany, and N. W. S. Wardhani, “Perbandingan Akurasi Peramalan Curah Hujan dengan menggunakan ARIMA, Hybrid ARIMA-NN, dan FFNN di Kabupaten Malang,” ["Comparison of Rainfall Forecasting Accuracy using ARIMA, Hybrid ARIMA-NN, and FFNN in Malang Regency"] in *Seminar Nasional Official Statistics*, Sep. 2021, pp. 245–253. doi: 10.34123/semnasoffstat.v2021i1.853.

[16] Y. Susanto and B. S. S. Ulama, “Pemodelan Curah Hujan dengan Pendekatan Model ARIMA, Feed Forward Neural Network dan Hybrid (ARIMA-NN) di Banyuwangi,” ["Rainfall Modeling with ARIMA Model Approach, Feed Forward Neural Network and Hybrid (ARIMA-NN) in Banyuwangi"] *JURNAL SAINS DAN SENI ITS*, vol. 5, no. 2, 2016, doi: 10.12962/j23373520.v5i2.16409.

[17] K. Naveena, S. Singh, S. Rathod, and A. Singh, “Hybrid ARIMA-ANN Modelling for Forecasting the Price of Robusta Coffee in India,” *Int J Curr Microbiol Appl Sci*, vol. 6, no. 7, pp. 1721–1726, Jun. 2017, doi: 10.20546/ijcmas.2017.607.207.

[18] W. S, “Prediksi Inflasi Indonesia Memakai Model ARIMA dan Artificial Neural Network,” ["Indonesia's Inflation Prediction Using the ARIMA Model and Artificial Neural Network"] *Jurnal Tata Kelola dan Kerangka Kerja Teknologi Informasi*, vol. 5, no. 1, Jan. 2019, doi: 10.34010/jtk3ti.v5i1.2297.

[19] I. K. Hasan and Ismail Djakaria, “Perbandingan Model Hybrid ARIMA-NN dan Hybrid ARIMA-GARCH untuk Peramalan Data Nilai Tukar Petani di Provinsi Gorontalo,” ["Comparison of Hybrid ARIMA-NN and ARIMA-GARCH Hybrid Models for Forecasting Farmers' Exchange Rate Data in Gorontalo Province"] *Jurnal Statistika dan Aplikasinya*, vol. 5, no. 2, pp. 155–165, Dec. 2021, doi: 10.21009/JSA.05204.

[20] C. Y. Rumaruson, L. J. Sinay, and M. I. Tilukay, “Indonesian rupiah exchange rate prediction using a hybrid ARIMA and neural network model,” in *AIP Conference Proceedings*, Nov. 2021, p. 020007. doi: 10.1063/5.0059512.

[21] R. L. Fraiha Lopes, S. G. C. Fraiha, H. S. Gomes, V. D. Lima, and G. P. S. Cavalcante, “Application of Hybrid ARIMA and Artificial Neural Network Modelling for Electromagnetic Propagation: An Alternative to the Least Squares Method and ITU Recommendation P.1546-5 for Amazon Urbanized Cities,” *Int J Antennas Propag*, vol. 2020, pp. 1–12, Mar. 2020, doi: 10.1155/2020/8494185.

[22] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Rev. San Francisco: Holden-Day, 1976.

[23] L. J. Sinay, F. K. Lembang, S. N. Aulele, and D. Mustamu, “Analisis Curah Hujan Bulanan di Kota Ambon menggunakan Model Heteroskedastisitas: SARIMA-GARCH,” ["Analysis of Monthly Rainfall in Ambon City using the Heteroscedasticity Model: SARIMA-GARCH"] *MEDIA STATISTIKA*, vol. 13, no. 1, pp. 68–79, Jun. 2020, doi: 10.14710/medstat.13.1.68-79.

[24] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, “Testing the null hypothesis of stationarity against the alternative of a unit root,” *J Econom*, vol. 54, no. 1–3, pp. 159–178, Oct. 1992, doi: 10.1016/0304-4076(92)90104-Y.

[25] H. Akaike, “Information Theory and an Extension of the Maximum Likelihood Principle,” in *Proceedings of the 2nd International Symposium on Information Theory*, 1973, pp. 267–281.

[26] W. S. Mcculloch and W. Pitts, “A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY,” 1990.

[27] P. J. Werbos, “Beyond regression : new tools for prediction and analysis in the behavioral sciences,” Harvard University, Massachusetts, 1974.

[28] P. J. Werbos, “Backpropagation through time: what it does and how to do it,” *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990, doi: 10.1109/5.58337.

[29] K. Fukushima, “Cognitron: A self-organizing multilayered neural network,” *Biol Cybern*, vol. 20, no. 3–4, pp. 121–136, 1975, doi: 10.1007/BF00342633.

[30] F. Günther and S. Fritsch, “neuralnet: Training of Neural Networks,” *R J*, vol. 2, no. 1, p. 30, 2010, doi: 10.32614/RJ-2010-006.

[31] N. D. Lewis, “BUILD YOUR OWN NEURAL NETWORK TODAY!,” 2015.

[32] Yahoo Finance, “Jakarta Composite Index (JKSE),” *https://finance.yahoo.com/quote/%5EJKSE/history?period1=1514851200&period2=1648684800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true*, Mar. 2022. https://finance.yahoo.com/quote/%5EJKSE/history?period1=1514851200&period2=1648684800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true (accessed Mar. 01, 2022).