

**STOCK PRICE FORECASTING OF PT. BANK CENTRAL ASIA USING HYBRID
AUTOREGRESSIVE INTEGRATED MOVING AVERAGE-NEURAL NETWORK
(ARIMA-NN) METHOD**

Apipah Nur Azizah*, Fatkhurokhman Fauzi, Prizka Rismawati Arum
Department of Statistics, Universitas Muhammadiyah Semarang, Semarang, Indonesia

*e-mail: afifahnurazizah78@gmail.com

Article Info:

Received: May 31, 2024

Accepted: July 31, 2024

Available Online: July 31, 2024

Keywords:

*BBCA Shares, Forecasting,
Hybrid ARIMA-NN, MAPE*

Abstract: PT. Bank Central Asia is a private company that has superior shares in the Lq45 category but has share prices that fluctuate every period. So forecasting is needed to predict stock prices in the next period. These fluctuations can cause linear and nonlinear relationships in historical stock price data. This research uses the Hybrid ARIMA-NN approach, where the ARIMA model is able to overcome data non-stationarity while the Neural Network is used to capture nonlinear patterns that cannot be explained by the ARIMA model by using the residuals as NN input, the hybrid model can increase forecasting accuracy. The data used is weekly data on closing stock prices for the period January 2019 to June 2024. Prediction measurements use Mean Absolute Percentage Error. The research results show that forecasting with Hybrid ARIMA(2,1,2)-NN(1-5-1) obtained a MAPE value of 3.99% smaller than the ARIMA(2,1,2) a MAPE value of 4.13%, that the accuracy of the forecasting model is very good.

1. INTRODUCTION

The increasing public interest in investing makes capital market conditions a major concern for business people, especially investors. Currently, investors in the capital market are an alternative form of investment that is easy for people to do since the opening of the BEI (Indonesian Stock Exchange). Investment is a form of shifting current consumption to the future which has uncertain risks. The capital market has several types of investment instruments including shares. Shares are securities that represent a person's capital ownership of a corporation[1]. Currently, shares are one of the investment options that investors often like because of their ability to generate attractive profits. But stock investment is also a form of investment that can produce quite high risks. This is due to the nature of shares which are sensitive to changes in economic conditions which can influence the rise and fall of share prices. One type of stock that is of great interest to investors is banking shares [2].

PT. Bank Central Asia is the largest privately owned bank in Indonesia and was founded in Jakarta on February 21, 1957 by Sudono Salim. PT shares. Bank Central Asia is one of the leading stocks on the Indonesian stock market because the shares are owned by PT. Bank

Central Asia is included in the Lq45 stock index [1]. Therefore, many investors are interested in investing in PT. Central Asia Bank. Development of PT share prices. Bank Central Asia continues to experience increases and decreases all the time [2].

The difficulty in estimating the selling price of a share is caused by the share price continuing to change from period to period. This causes investors to suffer losses if the share selling price is executed at the wrong time. Stock price predictions function to see how investment in a company's shares will progress in the future in order to minimize investors' risks when investing [3]. Continuous fluctuations can result in linear and nonlinear relationships in historical stock price data [4].

There are many time series methods for forecasting, one of which is the ARIMA model developed by Box-Jenkins. The ARIMA method produces predictions based on previous data patterns [5]. As a univariate model, ARIMA is suitable for use if time series observations are not correlated with each other. The ARIMA model consists of non-seasonal and seasonal models. The non-seasonal model which is a stationary model consists of AR(p), MA(q) and ARMA(p,q), while the ARIMA(p,d,q) model is a form of non-stationary model which is assumed to be linear. This causes ARIMA models to have difficulty handling nonlinear forms that generally exist in time series data [6]. As a result, ARIMA models require models that can handle non-linear patterns. One method that can handle linear and non-linear data patterns is Artificial Neural Networks (ANN).

Neural Network is a computational model inspired by the way the human brain works. A neural network model is an adaptive model that is able to capture nonlinear patterns and complex problems that are difficult to solve with a mathematical approach. Combining linear and nonlinear models is an alternative in forecasting. There are several reasons to use a hybrid model. First, there are difficulties in applying linear and nonlinear models to time series data. Second, time series in fact often contain both, not just linear or nonlinear. Third, there is some forecasting literature which says that there is no one model that is suitable for every condition [7].

In previous research, Setiawan and friends predicted share price data PT. Unilever with the hybrid ARIMA-Neural Network method. The purpose to predict linear and nonlinear time series data and obtain a MAPE 1.03% lower than the ARIMA model [4]. Similar research was also carried out by Anggaraeni and friends regarding forecasting Indonesian crude oil prices using the ARIMA-Neural Network hybrid approach. The research results show that the hybrid ARIMA-NN method produces a MAPE value of 6.8692%, while the ARIMA method itself produces a MAPE value of 7.7816% [8]. Amrullah and friends conducted similar research regarding the ARIMA-ANN hybrid to predict the closing share price of PT. Bank Negara Indonesia Tbk. Producing the best model, namely ARIMA(2,1,2)-ANN(5-10-1) with a learning rate of 0.3 to predict the closing stock price for the next 25 days, obtained a forecasting accuracy value using MAPE of 6.349% [9].

Hybrid methods are important especially when dealing with non-stationary time series data and have non-linear patterns which are often found in economic and financial time series data. The combination of linear and non-linear models produces more accurate predictions compared to using linear or non-linear models individually to forecast time series data [10]. Hybrid methods can also increase forecasting accuracy values [4]. Because stock prices have volatile movements, making predictions is expected to help investors make decisions when investing.

2. LITERATURE REVIEW

2.1. Forecasting

Forecasting in business is the process of estimating future value using past data. Forecasting is generally used to make decisions and policies that arise from the uncertainty of the future. The hope of the forecasting process is to obtain the minimum possible difference between the forecast results and what will happen in the future [11].

2.2. Stasionerity

Stationarity of time series data is defined as the condition where the mean and variance of the data are constant. Stationary data has no trend. Stationary means that the data does not experience significant changes over a certain period of time [12]. There are two stationarities in the data, namely stationary in the mean and stationary in the variance.

If the data is not stationary with respect to the mean differentiation is needed, the differencing process is carried out repeatedly if the first differencing data is still not stationary. The differencing time series data is formulated as follows [13]:

$$W_t = Y_t - Y_{t-1} \quad (1)$$

Where:

W_t : first differencing

Y_t : data at time t

$Y_{(t-1)}$: data at time t-1

Data are said to be stationary in the variance when the confidence interval $\lambda = 1$ [11]. If the data is not stationary with respect to variance, a Boxcox Lambda transformation is needed. The following is the transformation formula [14]:

$$T(Y_t) = \frac{Y_t^\lambda - 1}{\lambda}, \lambda \neq 0 \quad (2)$$

Where $T(Y_t)$ is the data transformation function Y at time t and λ is the transformation parameter value.

2.3. Autoregressive Integrated Moving Average (ARIMA)

The ARIMA method is a forecasting method based on historical data patterns that was first developed by Box-Jenkins in 1970 [15]. The ARIMA model is a univariate model, meaning the model only uses the dependent variable. The ARIMA (p,d,q) model is a combination of Autoregressive (AR) and Moving Average (MA) models which are differentiated d times. The following is the form of the ARIMA equation (p,d,q) [14]:

$$\phi_p(B)(1 - B)^d Y_t = \delta + \theta_q(B)\varepsilon_t \quad (3)$$

2.4. Artificial Neural Network

Artificial Neural Network is a flexible computational model and does not require assumptions for nonlinear modeling in various applications [16]. The neural network function was created to design a computer that can carry out a learning process based on examples of events. Feedforward Neural Network (FFNN) is a form of artificial neural network model that is commonly used because FFNN is very flexible for nonlinear functions [17]. FFNN consists

of one input layer, one or more hidden layers, and one output layer which is generally written with the following equation [18]:

$$\hat{y} = f^0\left[\sum_{j=1}^q \{w_j^o f_j^h [\sum_{i=1}^p w_{ij}^h y_{t-i} + b_j^h] + b^0\}\right] \quad (4)$$

2.5. Hybrid ARIMA-NN

Hybrid modeling is an analysis that is capable of solving complex problems, the aim of which is to increase forecast accuracy. The ARIMA model in the hybrid model is used as a solution to the linear model, where the residual results of the ARIMA model contain non-linear components. The hybrid model equation is written as follows [19]:

$$\hat{Y}_t = \hat{L}_t + \hat{N}_t \quad (5)$$

Where:

\hat{Y}_t : Forecast value of hybrid ARIMA-NN model

\hat{L}_t : ARIMA model prediction (linear model)

\hat{N}_t : NN model prediction (nonlinear model)

2.6. Mean Absolute Percentage Error (MAPE)

To produce accurate forecasts, an evaluation is carried out to determine the effectiveness of the forecasting model used. This evaluation is used to measure the error of the forecasting model used, one of which is the Mean Absolute Percentage Error (MAPE). MAPE is a calculation used to calculate the average residual absolute percentage error rate which serves to indicate forecasting accuracy. The following is the MAPE calculation formula [9]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\% \quad (6)$$

Where:

n : The amount of data

Y_t : Actual value for the t period

\hat{Y}_t : Forecast value for the t period

3. METHODOLOGY

This research uses secondary data from the finance.yahoo.com site in the form of closing price shares data for PT. Bank Central Asia Tbk with weekly time intervals from January 2019 to June 2024, as many as 287 data. This research data is divided into two, namely 80% train data from January 2019 to May 2023 and 20% test data from May 2023 to June 2024.

The steps in this research were carried out as follows:

a. ARIMA Models:

1. Enter data on the closing price of BCA shares.
2. Conduct descriptive analysis on closing stock price data PT. Bank Central Asia.
3. View the characteristics of the data with a time series plot.
4. Divide the train data into 80% and test data into 20%.
5. Carry out a stationarity test based on the ADF and BoxCox lambda tests.
6. Determine the temporary ARIMA model based on the ACF and PACF plots.
7. Estimating parameters using MLE and testing parameter significance.

8. Carry out a diagnostic check to test that the model residuals meet the assumptions of white noise and normality.
9. Determining the best ARIMA model based on the lowest AIC value.
10. Perform the best ARIMA residual linearity test

b. Hybrid Models:

1. Inputting the ARIMA model residuals sees a significant residual lag.
2. Normalize the data.
3. Determine the NN architecture.

At this stage, the architectural design that will be used in the training stage is determined.

4. Training and testing the NN model.
5. Denormalize the data.
6. Predict the hybrid ARIMA-NN model with equation (5).
7. Evaluate the hybrid ARIMA-NN model using MAPE values.
8. Make forecasts for the next 10 periods.

4. RESULTS AND DISCUSSION

Characteristics of closing price shares data for PT. Bank Central Asia used from January 2019 to June 2024 can be seen from descriptive analysis. The results of descriptive analysis show that the average closing price shares of PT. Bank Central Asia for 287 weeks amounted to 7340 rupiah with a standard deviation of 1428.614. This shows that the closing price shares variance of PT. Bank Central Asia is quite high.

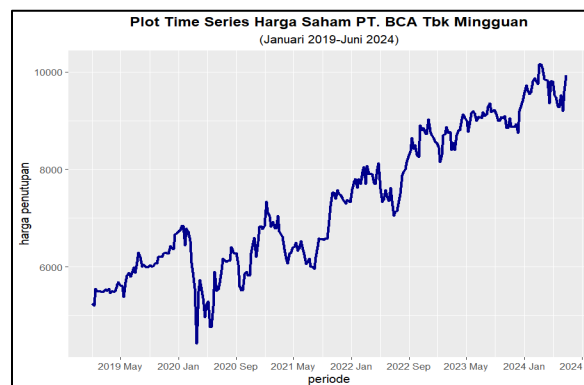


Figure 1. Plot of BCA Share Closing Prices from January 2019-June 2024

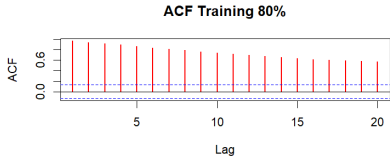
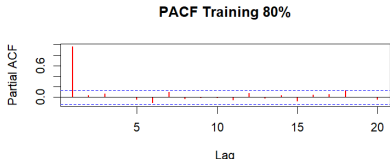
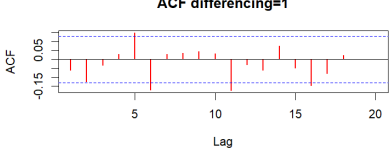
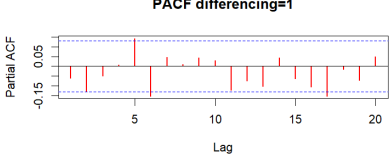
Based on Figure 1, it can be seen that the closing price shares movement of PT. Bank Central Asia experiences increases every week or has a fluctuating upward trend pattern.

4.1. ARIMA Model

Stationarity

Stationarity indicates that the data does not contain trend patterns and the data moves around constant mean and variance values. To see the stationarity of the data, it can be seen using the ACF PACF plot, ADF test, and Boxcox Lambda.

Table 1. Stationarity Test Results

	Not Differences	Differences = 1
	 	 
Boxcox λ	$\lambda = 1$	$\lambda = 1$
ADF Test	$t_{statistic} = -2,699$ $p - value = 0,282$	$t_{statistic} = -5,707$ $p - value = 0,01$

Based on table 1, it is known that the previous data was stationary on the variance with $\lambda = 1$ but not yet stationary on the mean, so differencing was carried out. After differencing once, the data was stationary in both variance and mean with the Augmented Dickey Fuller (ADF) Test results showing a pvalue of $0,01 > \alpha = 0,05$.

Estimation and Significance of Model Parameters

Parameter estimation is carried out for each model and it can be seen whether the model parameters are significant or not. A good model is a model that has significance or p-value $< \alpha = 0,05$. The following are the results of temporary model parameter estimation with MLE:

Table 2. Parameter Estimations

Models	Parameter	Estimations	AIC	<i>p-value</i>
ARIMA (1,1,1)	ϕ_1	0,4724	3108,39	0,1447
	θ_1	-0,5627		0,0617
ARIMA (1,1,2)	ϕ_1	-0,6189	3107,76	0,0023
	θ_1	0,5580		0,0080
	θ_2	-0,1542		0,0365
ARIMA (2,1,2)	ϕ_1	0,3808	3104,38	0,0000
	ϕ_2	-0,9060		0,0000
	θ_1	-0,4207		0,0000
	θ_2	0,8355		0,0000

Based on table 2, there are two models whose parameters are significant, the ARIMA (1,1,2) and ARIMA (2,1,2) models.

Diagnostic Checking

At the diagnostic checking stage, checks are carried out regarding testing the residual assumptions of the ARIMA model whose parameters are significant so that the selected model can represent the data pattern. There are two assumption tests on the ARIMA model residuals, namely white noise and normality.

Table 3. Ljung-Box Test

Models	χ^2	df	<i>p-value</i>	Information
ARIMA (1,1,2)	0,0005	1	0,9815	White Noise
ARIMA (2,1,2)	0,3136	1	0,5755	White Noise

Table 4. Kolmogorov Smirnov Test

Models	D_{stat}	df	<i>p-value</i>	Information
ARIMA (1,1,2)	0,0908	1	0,0458	Abnormal
ARIMA (2,1,2)	0,0844	1	0,0764	Normal

Based on tables 3&4, the ARIMA (2,1,2) model is a model that meets all assumptions, both white noise and normality, and has an AIC value that is smaller than the ARIMA (1,1,2) model, so the ARIMA (2,1,2) is selected as the best model.

Linearity Test

The linearity test is used to determine whether models has a linear or nonlinear pattern. Test the linearity of the closing price shares of PT. Bank Central Asia uses the Terasvirta test. Data can be said to be linear if the Terasvirta test results have a $p\text{-value} > \alpha = 0,05$.

Table 5. Terasvirta Test

Data	χ^2	df	<i>p-value</i>	Information
Residuals ARIMA (2,1,2)	21,442	2	0,00002	Nonlinier

4.2. Neural Network Model

After modeling with ARIMA, the next step is modeling with a Neural Network using the residual values of the selected ARIMA model as input for the NN model. In determining the number of inputs for the NN, it can be seen from the PACF plot that the residuals of the ARIMA model are significant.

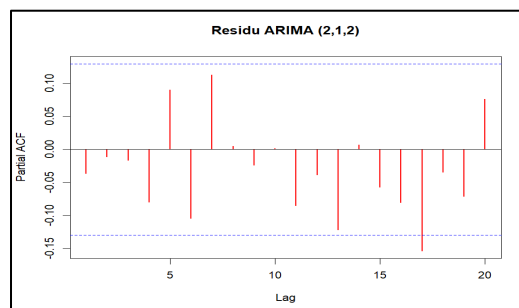


Fig 2. Plot PACF Residuals

Based on the residual PACF plot, there is only 1 significant lag, meaning that the NN input only uses 1 input neuron. In the NN model, this research uses an FFNN artist with 1

input neuron, 1-10 hidden layer neurons, 1 output neuron using a sigmoid activation function with a learning rate of 0.01 and a stepmax of 1000. The training results obtained are as below:

Table 6. Training Architecture Neural Network

<i>Neuron Hidden layer</i>	Error	Step
1	0,992	487
2	0,993	51
3	0,992	109
4	0,993	86
5	0,992	63
6	0,992	93
7	0,993	87
8	0,993	122
9	0,996	87
10	0,993	82

From table 6 it is known that the best architecture, namely NN(1-5-1), was chosen based on small error values and a small number of iterations.

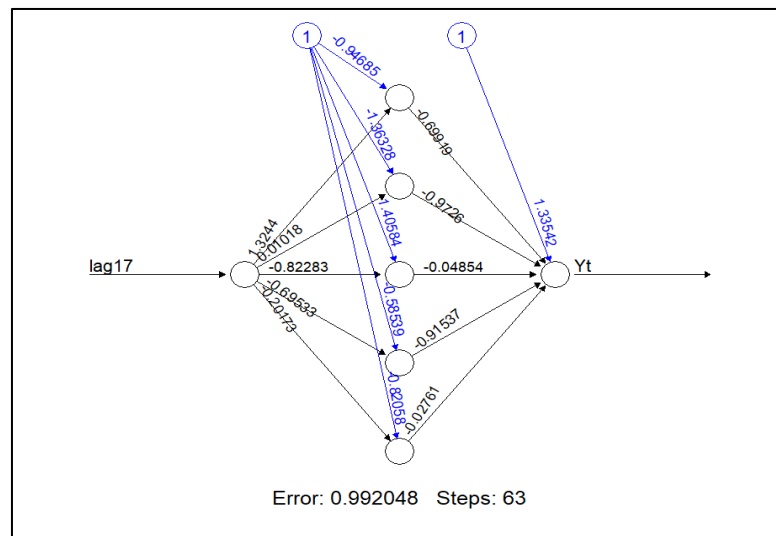


Fig 3. Architecture NN (1-5-1)

4.3. Hybrid ARIMA-NN

Prediction with the hybrid ARIMA-NN approach is the prediction of the best ARIMA model plus the best NN as in equation (5).

ARIMA (2,1,2) Equation:

$$Y_t = 1,3808Y_{t-1} - 1,2868Y_{t-2} + 0,9060Y_{t-3} + \varepsilon_t - 0,4207\varepsilon_{t-1} + 0,8355\varepsilon_{t-2}$$

NN (1-5-1) Equation:

$$\hat{N}_t = -0,699f_1^h(.) - 0,973f_2^h(.) - 0,049f_3^h(.) - 0,915f_4^h(.) - 0,028f_5^h(.) + 1,335$$

With :

$$f_1^h(.) = (1 + \exp(-(-0,947 + 1,324\varepsilon_{t-17})))^{-1}$$

$$f_2^h(.) = (1 + \exp(-(-1,363 + 0,010\varepsilon_{t-17})))^{-1}$$

$$f_3^h(.) = (1 + \exp(-(1,406 - 0,823\varepsilon_{t-17})))^{-1}$$

$$f_4^h(.) = (1 + \exp(-(-0,585 - 0,695\varepsilon_{t-17})))^{-1}$$

$$f_5^h(.) = (1 + \exp(-(0,821 - 0,202\varepsilon_{t-17})))^{-1}$$

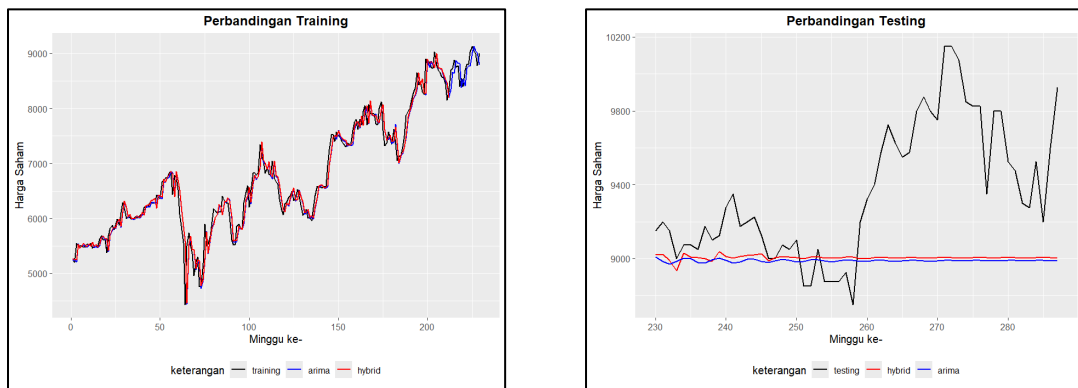


Fig 4. Comparison Training & Testing Models Prediction

After getting the prediction results, to find out how accurate the model is in forecasting, it is evaluated using MAPE.

Table 7. Evaluation Models

Model	Nilai MAPE	
	Training	Testing
ARIMA (2,1,2)	2,3%	4,13%
Hybrid ARIMA (2,1,2)-NN(1-5-1)	2,3%	3,99%

Based on table 7, the MAPE testing value of the Hybrid ARIMA(2,1,2)-NN(1-5-1) model is 3.99% lower than the ARIMA(2,1,2) model of 4.13%. This proves that the hybrid model approach can increase model accuracy compared to using a single model. So the Hybrid ARIMA(2,1,2)-NN(1-5-1) model will be used for forecasting the next period.

Following are the forecasting results for the next 10 weeks using hybrid ARIMA (2,1,2)-NN(1-5-1):

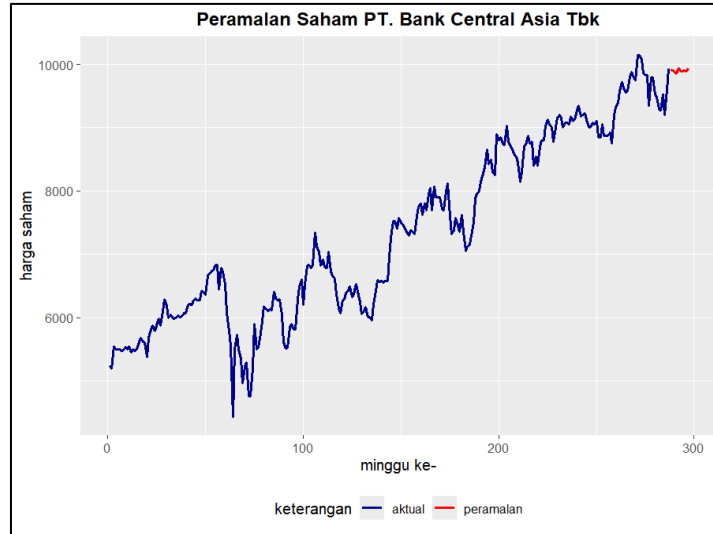


Fig 4. Time Series Plot Forecasting BBCA Shares

5. CONCLUSION

The characteristics of the closing price movement of BCA shares over the last 5 years show that the shares have an upward trend even though they fluctuate every week with the lowest price being 4430 on March 17 2020 or the 64th week and the highest price being 10150 on March 5 2024 or the 271st week. The best Hybrid ARIMA-NN model is a combination of the ARIMA(2,1,2) model using the residual NN(1-5-1) producing a MAPE value for model training of 2.3% and model testing of 3.99%, where the hybrid MAPE value is smaller than ARIMA (2,1,2). A MAPE value of less than 10% indicates that the forecasting model used is very good for predicting closing prices. Forecasting results for the next 10 weeks closing price shares of PT. Bank Central Asia Tbk showed fluctuations with the highest price occurring in the 297th week at 9943.82 rupiah and the lowest price at 9856.59 rupiah per share.

REFERENCES

- [1] M. P. Abdillah, E. Zukhronah, and Respatiwan, "Peramalan Harga Saham PT. Bank Central Asia Tbk Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA)," in *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, Yogyakarta, 2021.
- [2] C. N. R. Putri, "Peramalan Harga Saham Subsektor Perbankan Menggunakan Metode Geometric Brownian Motion (GBM) dan Estimasi Value at Risk (VaR)," Universitas Muhammadiyah Semarang, 2022.
- [3] W. R. U. Fadilah, D. Agfiannisa, and Y. Azhar, "Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine," *Fountain of Informatics Journal*, vol. 5, no. 2, pp. 45–51, Sep. 2020, doi: 10.21111/fij.v5i2.4449.
- [4] C. D. Setiawan, W. Sulandari, and Y. Susanti, "Peramalan Harga Saham PT Unilever Indonesia Menggunakan Metode Hibrida ARIMA-NEURAL NETWORK," 2023, *Seminar Nasional Riset dan Inovasi Teknologi (SEMNAS RISTEK)*.
- [5] D. Ayu Rezaldi and Sugiman, "Peramalan Metode ARIMA Data Saham PT. Telekomunikasi Indonesia," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4,

- pp. 611–620, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [6] E. Dave, A. Leonardo, M. Jeanice, and N. Hanafiah, “Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM,” in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 480–487. doi: 10.1016/j.procs.2021.01.031.
- [7] I. K. Hasan and I. Djakaria, “Perbandingan Model Hybrid ARIMA-NN dan Hybrid ARIMA-GARCH untuk Peramalan Data Nilai Tukar Petani di Provinsi Gorontalo,” *Jurnal Statistika dan Aplikasinya*, vol. 5, no. 2, 2021.
- [8] D. A. Anggraini, S. Wahyuningsih, and M. Siringoringo, “Peramalan Harga Minyak Mentah Indonesia Jenis Sepinggian Yakin Mix Menggunakan Model Hybrid Autoregressive Integrated Moving Average-Neural Network,” 2023. [Online]. Available: <https://migas.esdm.go.id/>.
- [9] A. Amrullah, O. Soesanto, and Maisarah, “Penerapan Metode Hybrid ARIMA-ANN Untuk Memprediksi Harga Saham PT. BNI (Persero) Tbk,” *RAGAM: Journal of Statistics and Its Application*, vol. 01, no. 01, 2022.
- [10] N. Sadeghi Gargari, H. Akbari, and R. Panahi, “Forecasting Short-term Container Vessel Traffic Volume Using Hybrid ARIMA-NN Model,” *INTERNATIONAL JOURNAL OF COASTAL & OFFSHORE ENGINEERING (IJCOE)*, vol. 4, no. 3, pp. 47–52, 2019.
- [11] M. Al Haris and P. R. Arum, “Peramalan Harga Emas Dengan Model Generalized Autoregressive Conditional Heteroskedasticity (GARCH),” *Jurnal Sains dan Matematika Unpam*, vol. 3, no. 1, pp. 19–30, 2020.
- [12] M. K. Possumah and A. A. Rohmawati, “Prediksi Harga Saham Menggunakan Vector Autoregressive (VAR) Non-Stasioner (Studi Kasus Saham Perusahaan PT United Tractors Tbk),” in *e-Proceeding of Engineering*, 2020, p. 8361. [Online]. Available: <https://www.bi.go.id/>.
- [13] M. Mardhatillah, “Peramalan Curah Hujan Di Provinsi Sumatera Barat Menggunakan Vector Autoregressive-Generalized Space Time Autoregressive (VAR-GSTAR),” 2021. [Online]. Available: <http://repository.unimus.ac.id>
- [14] S. Deviana, Nusyirwan, D. Azis, and P. Ferdias, “Analisis Model Autoregressive Integrated Moving Average Data Deret Waktu Dengan Metode Momen Sebagai Estimasi Parameter,” *Jurnal Siger Matematika*, vol. 02, no. 02, 2021.
- [15] Y. W. A. Nanlohy and S. B. Loklomin, “Model Autoregressive Integrated Moving Average (ARIMA) Untuk Merelekan Inflasi Indonesia,” *VARIANCE: Journal of Statistics and Its Applications*, vol. 5, no. 2, pp. 201–208, Oct. 2023, doi: 10.30598/variancevol5iss2page201-208.
- [16] U. C. Buyuksahin and S. Ertekin, “Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition,” *Neurocomputing*, vol. 361, pp. 151–163, Oct. 2019, doi: 10.1016/j.neucom.2019.05.099.
- [17] M. E. Sonata, “Peramalan Produksi Padi Di Provinsi Jawa Timur Menggunakan Metode Feedforward Neural Network,” *Universitas Muhammadiyah Semarang*, 2021. [Online]. Available: <http://repository.unimus.ac.id>

- [18] Y. Rismawanti, “Peramlan Nilai Tukar Rupiah Terhadap Dolar Amerika Menggunakan Feedforward Neural Network Dengan Algoritma Backpropagation,” Universitas Muhammadiyah Semarang, 2019. [Online]. Available: <http://repository.unimus.ac.id>
- [19] V. Sari and D. A. Maulidany, “Prediksi Kecepatan Angin Dalam Mendeteksi Gelombang Air Laut Terhadap Skala Beaufort Dengan Metode Hybrid ARIMA-ANN,” *Statistika*, vol. 8, no. 1, 2020, [Online]. Available: <http://jurnal.unimus.ac.id>