DOI: 10.15849/IJASCA.240730.10

Int. J. Advance Soft Compu. Appl, Vol. 16, No. 2, July 2024
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Comparison of GARCH, LSTM, and Hybrid GARCH-LSTM Models for Analyzing Data Volatility

Laily Nissa Atul Mualifah, Agus Mohamad Soleh, Khairil Anwar Notodiputro

Department of Statistics, IPB University, 16680, Indonesia
e-mail: lailynissa@apps.ipb.ac.id
Department of Statistics, IPB University, 16680, Indonesia
e-mail: agusms@apps.ipb.ac.id
Department of Statistics, IPB University, 16680, Indonesia
e-mail: khairil@apps.ipb.ac.id

Abstract

Most statistical methods in time series analytics assume that the residuals are independently and identically distributed with zero mean and constant variance. In real cases, this assumption may be violated. Nowadays, data are dynamic and highly volatile, particularly in finance. The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is a statistical method for non-constant conditional variance that can capture the volatility data. Recently, artificial intelligence methods are gaining popularity and have promising performance, one of those is the Long Short-Term Memory (LSTM) method. However, due to the filtering process by forget gate in the LSTM cell some information is missing, which can decrease the prediction's accuracy. This study proposes a method, namely Hybrid GARCH-LSTM, to overcome those limitation. The performance of the proposed method is evaluated in the simulation and empirical data and compared with GARCH and LSTM model. The results show that the Hybrid GARCH–LSTM model is able to recognize the volatility pattern of data well and outperforms all the other models.

Keywords: *Autocorrelation, Heteroscedasticity, Machine Learning, Stock Price, Time Series.*

1 Introduction

Time series are datasets of arranged observations in time order [1]. One of the essential works with time series datasets is how to predict the future based on the past. Stationarity is the main characteristic of time series that must be fulfilled when using statistical models. The stationary diagnostic checks of those models are based on the assumptions of errors, which are independently and identically distributed with zero means and constant variance [2]. However, this assumption may be violated in many practical cases, particularly in finance or economics.

Todays, data are dynamic and highly volatile. The high volatility data generally has time-dependent variance (heteroskedastic). In this case, a linear model or even an ARIMA

Received 4 January 2024; Accepted 22 March 2024

model could not capture the volatility pattern of data and thereby would lead to erroneous conclusions. To overcome this limitation, Engle [3] introduced the *Autoregressive Conditional Heteroskedasticity* (ARCH) processes which allow the conditional variance to change over time. As the generalization of ARCH, *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) was initially introduced by Bollerslev [4]. GARCH has been widely used in the risk management of assets. Some studies [5, 6] show that GARCH can capture data volatility.

In addition to statistical methods, artificial intelligence methods are widely used in many prediction tasks. The best artificial intelligence method for sequential data is the *Recurrent Neural Network* (RNN). It has a memory that can remember important information about the input [7]. As the improvement of RNN, LSTM has a recurrent learning unit with gates to capture the longer states from the beginning unit and the shorter states from the last unit. These features allow LSTM to have a good memory for long data periods. Some studies [8, 9, 10] show a promising performance of LSTM in time series forecasting problems. Despite the advantages of LSTM above, LSTM has some problems. Under limited data, the accuracy of LSTM prediction will decrease with the increase of the prediction period. Besides, the existence of forget gate in LSTM would reduce the participation of previously hidden states and prioritizes the current state. The LSTM model can be improved by combining LSTM with other models [11]. Some studies have been combining the LSTM model with statistical [12], deep learning [13], and a hybrid of both methods [14].

This study will examine the performance of hybrid GARCH–LSTM models for volatility data and compares the performance to the GARCH, LSTM models. In the hybrid model, GARCH is the initial model used to model the main linear components of data, while the LSTM model is used to model the non-linear details in residuals of the GARCH model. The model's performances are evaluated using the *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), and *Mean Absolute Error* (MAE) metrics. This study contributes to the literature by comparing statistical, machine learning, and hybrid statistical and machine learning methods for analyzing the dynamical pattern of data volatility.

2 Literature Review

2.1. Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

Heteroscedasticity is defined as the condition where the residual variance is non-constant over time. Given a time series $\{Y_t\}$, the conditional variance of Y_t measures the uncertainty in the deviation of Y_t from its conditional mean. The phenomenon where the conditional variance of this series varies over time is called volatility [15]. *Autoregressive Conditional Heteroscedasticity* (ARCH) processes are serially uncorrelated processes with non-constant variances conditional on the past but constant unconditional variances [3]. The ARCH process allows the conditional variance to depend on the past realization of the series. The ARCH (q) denoted the ARCH process with order q. The conditional variance at time t of ARCH (q) process is defined as:

$$\sigma_{t|t-1}^2 = \omega + \sum_{i=1}^{q} \alpha_i r_{t-i}^2, \quad (1)$$

with $r_t = \sigma_{t|t-1}\varepsilon_t$ is a return at time t; ε_t is a random variable at time t that is independently and identically distributed with zero mean and unit variance; ω, α_i are the model parameters.

The natural generalization of the ARCH process, *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH), allows p lagged conditional variances to enter the model. The conditional variance at time t of $\text{GARCH}(q,p)$ process is defined as

$$\sigma_{t|t-1}^2 = \omega + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j|t-j-1}^2. \quad (2)$$

The condition of $\alpha_i \geq 0$ and $\beta_i \geq 0$ is necessary to make the conditional variance non-negative. The necessary and sufficient condition for weak stationary of GARCH process is $\sum_{i=1}^{max(p,q)}(\beta_i + \alpha_i) < 1$.

2.2. Long Sort-Term Memory (LSTM)

Long Short–Term Memory (LSTM) is a Recurrent Neural Networks (RNN) variation. RNN has been developed for tasks related to ordered sequence data. RNN accomplishes the task by recursive processes. The weight estimators, together with an additional latent variable called hidden state, are iteratively updated and keeps the memory of the previous steps [16]. Due to the repetitive nature of RNN, the gradient tends to shrink exponentially or become too large in the significant time step. It makes the model training process less effective, and the model has poor performance while dealing with long-term dependency. LSTM address to overcome the vanishing gradient problem of RNN.

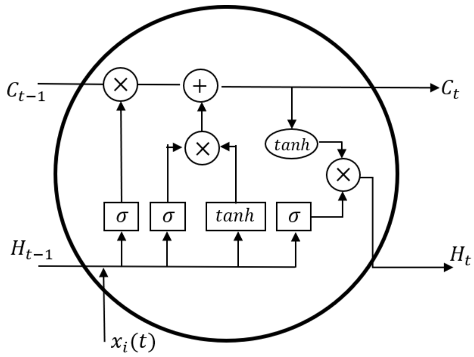


Fig 1. An LSTM cells

The LSTM block contains a memory cell and three gates with weights and bias vectors through which information passes serially [17]. Three gates of LSTM, called input gate, forget gate, and output gate, control the information flow, whether the information should be saved or forgotten. In general, the architecture of LSTM is presented in Fig 1. The workflow of LSTM cell consists of some stages. First of all, a combination of current block input and previous activation values are passed through a sigmoid function f_t (Equation 3) to filter the information that should be stored. In the next stage, other sigmoid functions i_t (Equation 4) in the input gate are used to filter the same combination of activation from the previous layer and the current block to be used as the updated input. Then, the primary set of information is passed through a tan hyperbolic function to produce new candidate values $(\tilde{c_t})$ (Equation 5) of the current block. The previous cell state (c_{t-1}) is then updated to a new cell state (c_t) (Equation 6) and served as the final values of the current block. Another sigmoid function filters the last stage, the final values of the current block o_t (Equation 7) in the output gate and passed through another *tanh* activation. All functions for each stage are defined as:

$$f_t = \sigma\left(w_{fx}x_t + w_{fh}h_{t-1} + b_f\right), \quad (3)$$

$$i_t = \sigma(w_{ix}x_t + w_{ih}h_{t-1} + b_i), \quad (4)$$

$$\widetilde{c_t} = tanh(w_c x_t + x_c h_{t-1} + b_c), \quad (5)$$

$$c_t = f_t \times c_{t-1} + i_t \times \widetilde{c_t}, \quad (6)$$

$$o_t = \sigma(w_{ox}x_t + w_{oh}h_{t-1} + b_o), \quad (7)$$

$$h_t = o_t \times tanh(c_t), \quad (8)$$

with f_t, i_t, o_t are filter in forget, input and output gate respectively, w is a weight matrix size $\left((t-1) \times (2t-2)\right), b$ is a bias vector size $\left((t-1) \times 1\right)$, x is an input, h is an output, c and $\tilde{c}$ are cell state values and candidate cell state values, respectively.

2.3. Evaluation Metrics

When various models are applied in the analysis, forecasting accuracy generally is used to decide the best model. Forecasting accuracy refers to the forecasting error that is the deviation between the actual and predicted values [18]. Here are several metrics used to access the forecasting accuracy in the time series modeling.

Mean Square Error (MSE)

Given data at time t is y_t and $\hat{y}_t$is the predicted value of y_t. MSE is calculated by:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2. \quad (9)$$

The best model will have an MSE value that is close to zero. MSE is also used to detect outliers. If there is a bad prediction result, the squaring part of the MSE function magnifies the error [19].

Root Mean Square Error (RMSE)

RMSE is related to the MSE (through the square root). RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}. \quad (10)$$

Mean Absolute Percentage Error (MAPE)

MAPE has an intuitive interpretation in terms of relative error. The formula of MAPE is presented in Equation 11. MAPE usually is used in tasks where being sensitive to relative variations are more essential than to absolute variations.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100. \quad (11)$$

Mean Absolute Error (MAE)

MAE or MAD (*Mean Absolute Deviation*) is the average absolute deviation between the actual and forecasted values. The formula for MAE is:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|. \quad (12)$$

3 The Proposed Method

The hybrid GARCH-LSTM combines two models to capture linear and non-linear features of data. The GARCH model is the initial model used to capture the linear elements of data, while the LSTM model is proposed to extract the non-linear features. The general procedure of the hybrid GARCH–LSTM model is illustrated in Fig 2. Firstly, data is modelled using the GARCH model. The order of the GARCH model is determined based on the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) identification. Secondly, the residuals of the GARCH model are then modelled using the LSTM model. Lastly, the final prediction of the hybrid model is calculated as the sum of predicted values of the GARCH model and LSTM model. The calculation formula of the hybrid model is defined as follows:

$$y_r(\mathrm{t}) = y_o(\mathrm{t}) - y_g(\mathrm{t}), \tag{13}$$

$$y_f(\mathrm{t}) = y_g(\mathrm{t}) + y_l(\mathrm{t}), \tag{14}$$

with $y_r(t)$ is the residual, $y_o(t)$ is the actual value of the return, $y_g(t)$ is the GARCH model prediction, $y_l(t)$ is the LSTM model prediction, $y_f(t)$ is the hybrid model prediction.

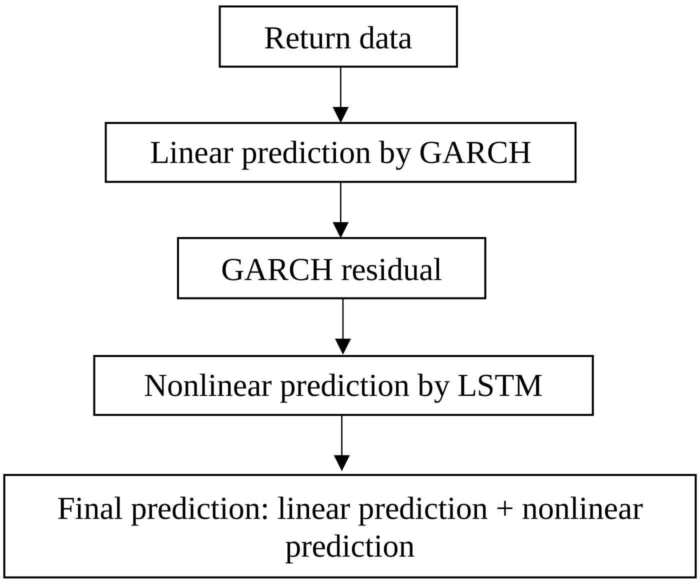


Fig 2. Flow chart for hybrid GARCH–LSTM model

4 Methodology

4.1. Data

This study evaluates the performance of GARCH, LSTM, and hybrid of both models in simulation and empirical data. The simulation data are obtained from the simulation process by generating data from the GARCH model. There are ten scenarios of the data generating process, presented in Table 1, with ten repetitions for each scenario. To generate returns r_t, for $t = 1, 2, \dots, 365$, the GARCH processes in (1) and (3) are used, with the initial value of the return $r_0 \sim Uniform\ (0, 1)$ and $\varepsilon_t \sim N(0, 1)$.

The empirical data is the stock closing price of PT. Bumi Resources Minerals Tbk. (BRMS) obtained from https://yahoo.finance.com. The closing price from 01st January to 31st

December 2021 is used for this study. The data is in the daily frequency with the stock price only available for the workdays, so there are 247 observations.

Table 1. The simulation scenario

Scenario	Variance	Dataset	Parameter		
			ω	α	β
1	High	GARCH (1,1)	0.5	0.1	0.7
2		GARCH (1,3)	0.7	0.2	0.4, 0.2, 0.1
3		GARCH (3,1)	0.2	0.3, 0.1, 0.1	0.4
4		GARCH (3,3)	0.2	0.3, 0.2, 0.05	0.2, 0.1, 0.05
5		Contaminated GARCH	0.5	0.1	0.7
6	Low	GARCH (1,1)	0.5	0.1	0.15
7		GARCH (1,3)	0.1	0.05	0.2, 0.1, 0.05
8		GARCH (3,1)	0.1	0.2, 0.1, 0.05	0.05
9		GARCH (3,3)	0.1	0.15, 0.1, 0.05	0.1, 0.05, 0.025
10		Contaminated GARCH	0.1	0.05	0.2, 0.1, 0.05

4.2. Analysis

For analysis, each data (i.e., simulation and empirical data) is divided into training and testing data with the proportion of 80% and 20%, respectively. Three models (i.e., GARCH, LSTM, and Hybrid) are applied to each training data. In the LSTM modeling, manual hyperparameter tuning is carried out. Due to the subjective inference of the manual process, the complete block factorial design is conducted to analyze the MAPE values obtained from all scenarios of simulation data. The design uses three factors: variance, model, and dataset, while the block is the repetition. The effect of the factors on the MAPE value is then analyzed using analysis of variance (ANOVA). The best model is obtained from Tukey's HSD analysis.

For empirical data, the stock closing price data is transformed into returns to be stationer. The returns can be obtained by log and differencing transformation based on the following formula:

$$r_t = log\, y_t - log\, y_{t-1}. \tag{15}$$

The ARCH effect on the returns is checked by the sample ACF and PACF and Ljung-Box test of the returns and the absolute returns. The absolute returns are examined if the returns are serially uncorrelated or provide the white noise model. If the absolute returns admit significant autocorrelations, these autocorrelations furnish evidence of the existence of the ARCH effect. The model's validation is based on the goodness of fit between the actual values and the model's prediction values of returns and stock closing price, which is evaluated by MSE, RMSE, MAPE, and MAE metrics. The seven days ahead predictions are calculated after the validation process.

5 Results and Discussion

5.1 Simulation Study

MAPE values from the modelling process using simulated data are analyzed using analysis of variance (ANOVA) based on the complete block factorial design with three factors.

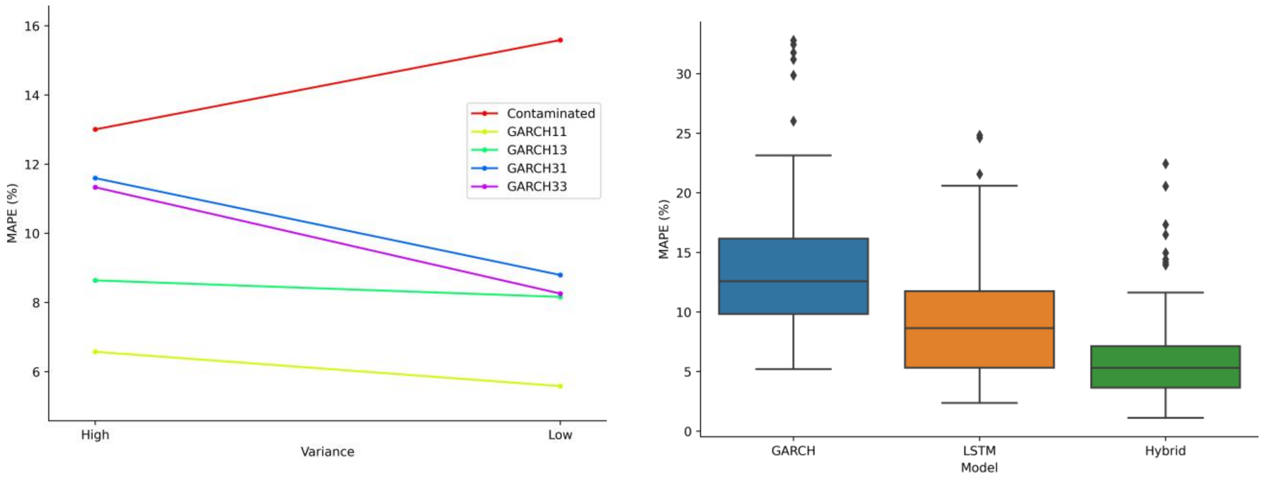
Table 2 presents the ANOVA results of MAPE. The results show that the only significant interaction is the interaction between variance and dataset, and all main factors significantly affect the value of MAPE. Meanwhile, the other factors (i.e., block and model interactions) are statistically insignificant. However, the significant interaction effect between variance and dataset implies that variance affects MAPE value differently. Still, these variance effects depend on the kind of dataset.

Table 2. ANOVA results of MAPE

Source	Degree of freedom	Sum Square	Mean Square	*F*	*p-value*[a]
Block	9	242.125	26.903	1.713	8.628e-02
Variance	1	68.055	68.055	4.332	3.838e-02[a]
Model	2	2948.542	1474.271	3.844	1.963e-31[a]
Dataset	4	2170.901	542.725	4.547	3.824e-23[a]
Variance: Model	2	56.145	28.073	1.787	1.695e-01
Variance: Dataset	4	309.972	77.493	4.933	7.535e-04[a]
Model: Dataset	8	88.880	11.110	0.707	6.851e-01
Variance: Model: Dataset	8	136.066	17.008	1.083	3.755e-01
Residual	261	4100.277	15.710		

[a] Factor is statistically significant at a 95% confidence level

Based on Fig 3A, the MAPE value of the Contaminated GARCH dataset has a different pattern from the other dataset. The MAPE values of Contaminated GARCH are low for high variance, and the MAPE values are high for low variance. Meanwhile, the MAPE values are high on the other dataset when the variance is high and low when the variance is low. It implies that the MAPE value from different variances depends on the kind of dataset. Compared to the other datasets, the MAPE values of the contaminated GARCH are the highest, while the MAPE values of the GARCH (1,1) are the lowest. According to Tukey's multiple comparisons, the Contaminated GARCH dataset from high variance has different effects on the MAPE value than other datasets from low variance. Moreover, GARCH (3,1) and GARCH (3,3) datasets from high variance have different effects on the MAPE value than the GARCH (1,1) dataset from low variance. Another result shows that the GARCH (1,3) and GARCH (3,3) datasets from high variance have different effects on the MAPE value than the Contaminated GARCH dataset from low variance. Furthermore, all comparisons show no significant effect on the MAPE value for the same datasets but with different variances. On the other hand, for the same variance (i.e., both are high variances), the Contaminated GARCH dataset has different effects on the MAPE value than the GARCH (1,1) and GARCH (1,3) datasets. Meanwhile, for both low variances, the Contaminated GARCH dataset has a different effect on the MAPE value than other datasets.



A **B**

Fig 3. Simulation study results: **(**A) interaction plot between variance and dataset; (B) boxplot of models

Another factor that significantly affects the MAPE values is the model. Boxplot in Fig 3B shows the different values of MAPE based on the model. The lowest MAPE is obtained from the Hybrid GARCH–LSTM model with the mean value of 6.30%, while the highest MAPE is obtained from the GARCH model with the mean value of 13.89%, and for the LSTM model, the average value of MAPE is 9.07%. Tukey's multiple comparisons show that the mean difference for each pair of two models is significant. The mean difference of MAPE between GARCH and Hybrid models is -7.5879. The negative sign indicates that the mean MAPE of the Hybrid model is lower than that of the GARCH model. Meanwhile, the mean difference in MAPE between the Hybrid and LSTM models is 2.771. It shows that the mean MAPE of the LSTM model is higher than that of the Hybrid model. Therefore, it can be concluded that the Hybrid GARCH–LSTM model gives the lowest MAPE or is the best model among others.

5.2 Empirical Study

Fig 4A shows the BRMS daily stock closing prices for the study period. Generally, it shows an increasing trend and different variability for different levels of stock closing price; the higher stock closing prices the higher variability. On the financial asset, a return is the asset's price change, which is a stationary data. The closing price data is transformed into the return form and shown in Fig 4B. The return plot shows that the mean values are constant over zero. The Dickey-Fuller unit root (ADF) test result also indicates that returns are stationary with a $p\text{-}value < 0.05$. On the other hand, the plot shows that returns are highly volatile over some periods, especially at the beginning and the middle of the study period. The covid-19 cases might cause these results during these periods. This pattern gives a hint of the volatility clustering on the returns. The sample ACF and PACF of returns show serial uncorrelation over the data and suggest the white noise model. It can be said that the series is independently and identically distributed. Meanwhile, the ACF and PACF of the absolute returns preserve the serial dependence structure in the absolute returns. There are some significant autocorrelations in the absolute returns. Therefore, it can be concluded that the returns are not independently and identically distributed, or there is a volatility clustering on the returns.

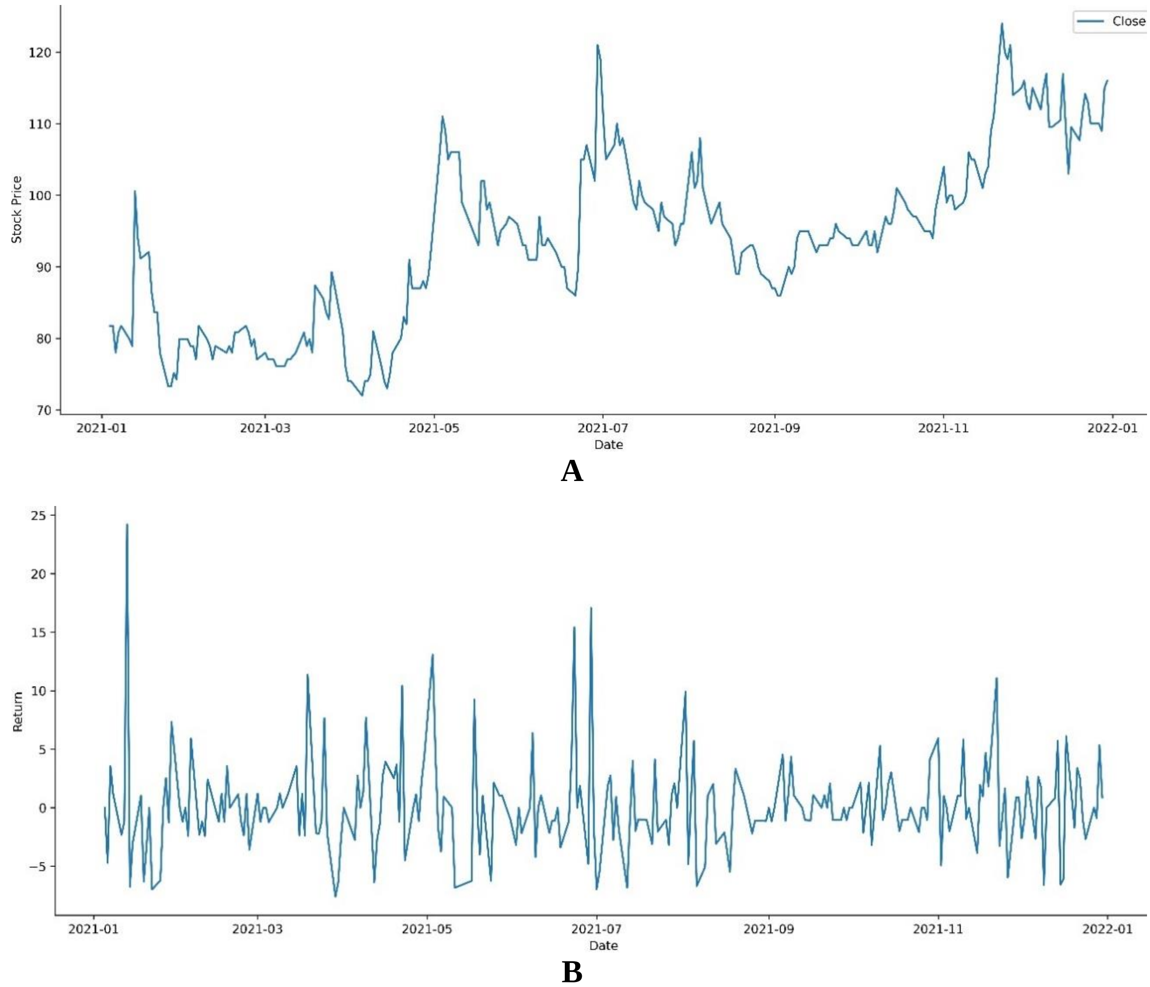
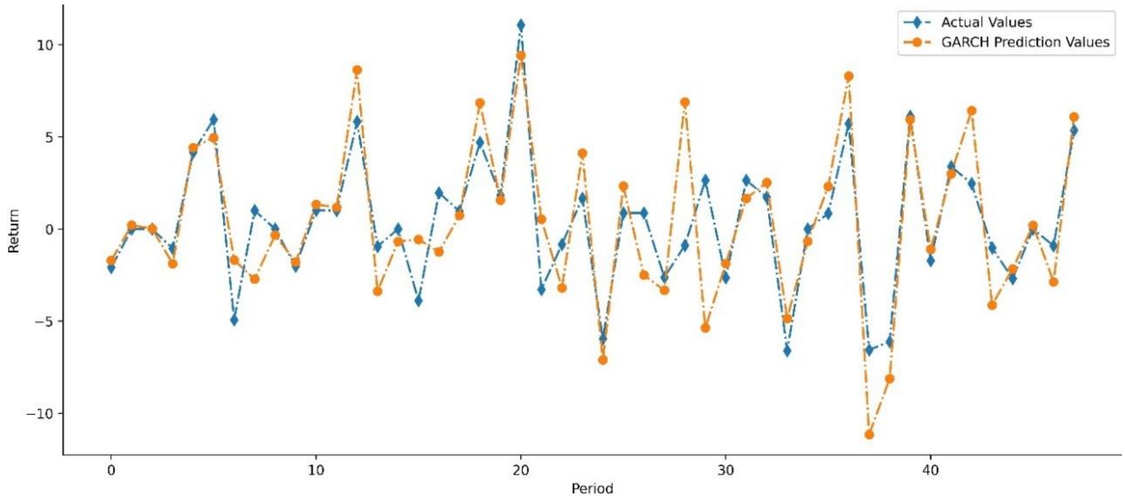


Fig 4. BRMS daily stock (A) closing price; (B) return

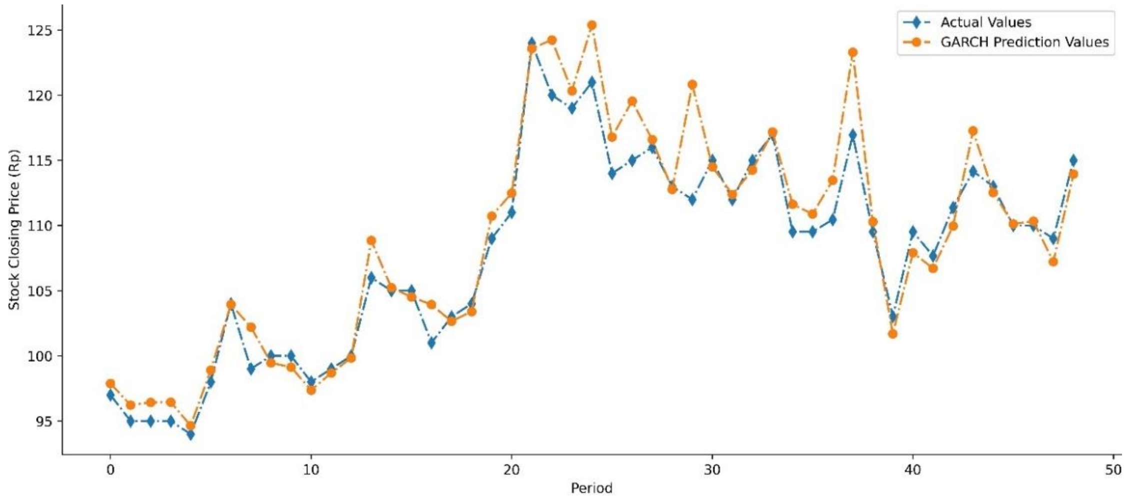
On the model specification stage, the best model for the returns is the GARCH (4, 3) model, which has the lowest AIC (*Akaike Information Criterion*) and BIC (*Bayesian Information Criterion*). The sample ACF and PACF of the standardized residuals of the model and the absolute values of those show that the white noise model is appropriate. This implies that the standardized residuals are independently and identically distributed, and the model is correctly specified. The GARCH model is then constructed using the order of (4, 3). The GARCH model for conditional variance of BRMS daily stock return is:

$$\sigma_{t|t-1}^2 = 1.6235 + 0.1599r_{t-1}^2 + 6.5111e^{-3}r_{t-2}^2 + 0.372r_{t-4}^2 + 0.0113\sigma_{t-1}^2 + 0.4503\sigma_{t-3}^2. \quad (16)$$

According to the Equation (16), the estimated parameters of α_3 and β_2 are zero. It means that the conditional variance value at t period does not depend on the third period ago, (r_{t-3}^2), and the conditional variance on the second period ago, (σ_{t-2}^2). All positive values of those estimated parameters indicate that the squared value of the returns and the conditional variance in the previous period positively affect the return value at t period. For example, the estimated parameter of β_3 is 0.4503. It implies that the conditional variance at t period is expected to increase by 0.4503, assuming that the squared value of the returns and the conditional variance in the previous period are constant. Note that the coefficient of r_{t-2}^2 is very small close to zero $(6.5111e^{-3})$. It might indicate that the conditional variance value at time t period also does not rely on the second previous period.



A



B

Fig 5. Fitted values of GARCH model for daily (A) return; (B) stock closing price

The GARCH model validation is presented in fitted values of prediction visualization in Fig 5A. Based on the Fig, the return predictions do not fit well with the actual returns, especially when the volatility is low. For some periods, the prediction pattern does not follow the pattern of the actual values well. In addition, the overestimated and underestimated predictions occur in the period when the volatility is high. Fig 5B presents the visualization of the GARCH predictions after being transformed into the stock closing prices. The stock closing price predictions of the GARCH model do not fit well. However, the differences between the actual stock closing price and the prediction values are small enough, and the prediction pattern still follows the actual pattern.

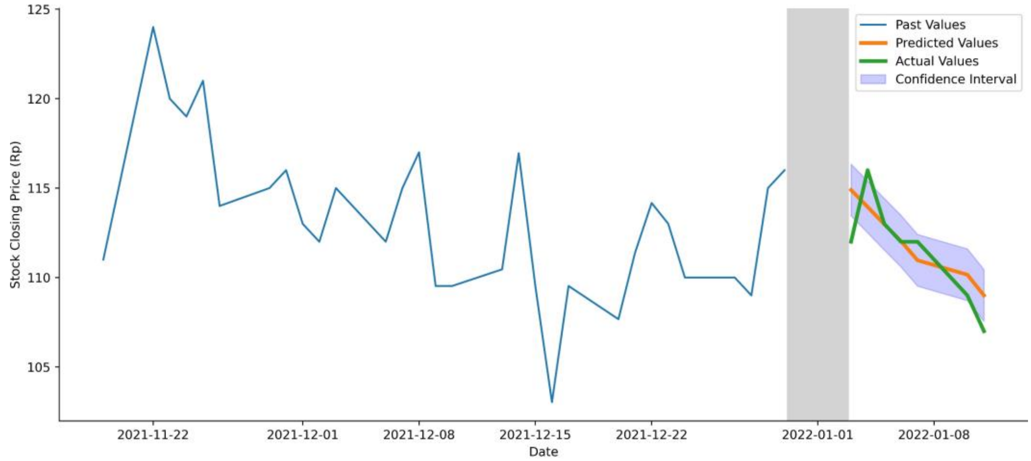
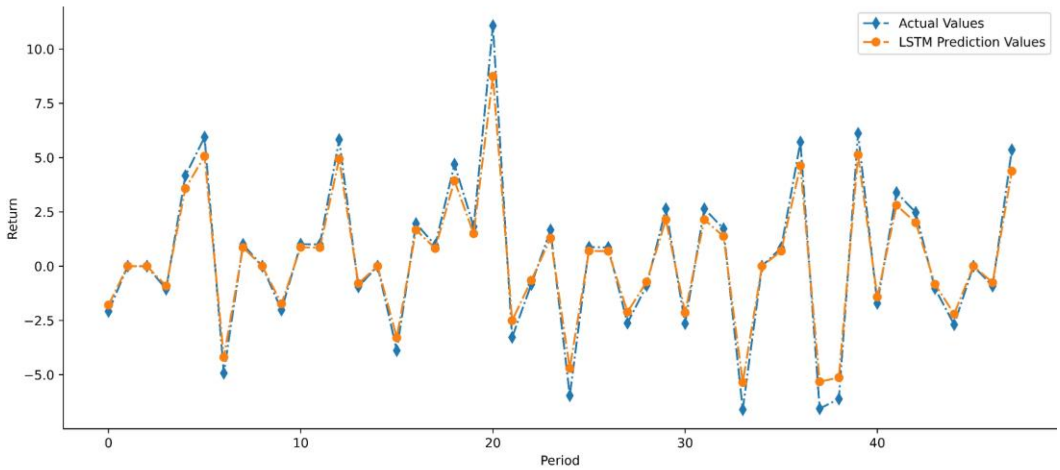
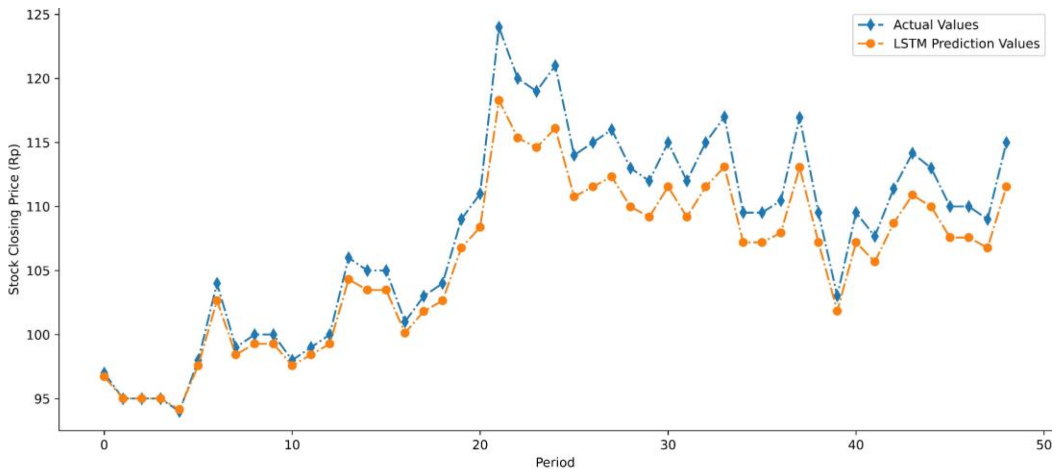


Fig 6. Stock closing price predictions of the GARCH model

The predictions of stock closing price seven days ahead using the GARCH model are shown in Fig 6, the first day of stock closing price prediction on 3rd January 2022 decrease from that of the last study period with the prediction value of Rp 114.775 and the confidence interval of *[113.360, 116.191]*. This prediction is overestimated than the actual value. On the second day, the actual value of the stock closing price increased. However, the stock closing price predictions decrease from the first day until the seventh day, 11st January 2022, with the prediction value of Rp 108.953 and the confidence interval of *[107.537, 110.368]*.



A



B

Fig 7. Fitted values of LSTM model for (A) return; (B) stock closing price

For the LSTM model, the network structure is predefined as the LSTM layer–Dropout Layer–Dense Layer. Some hyperparameters are tuned to get the best LSTM model, which gives the lowest MAPE value. The best hyperparameter combination is the number of neurons of 200, the learning rate of 200, and epochs of 100. The goodness of fit between the actual values and the LSTM model prediction values of returns and stock closing price is presented in Fig 7A. The plots show that the LSTM model can capture the pattern of the volatility data well. For returns, the prediction values of the LSTM model fit well with the actual values. However, when the volatility is high, the predictions are underestimated and overestimated. The return predictions are then transformed into the stock closing price prediction. Based on Fig 7B, the stock closing price predictions tend to be underestimated from the middle until the end of the study period. However, the predictions fit well at the beginning of the study period. Overall, the prediction pattern follows the actual data pattern.

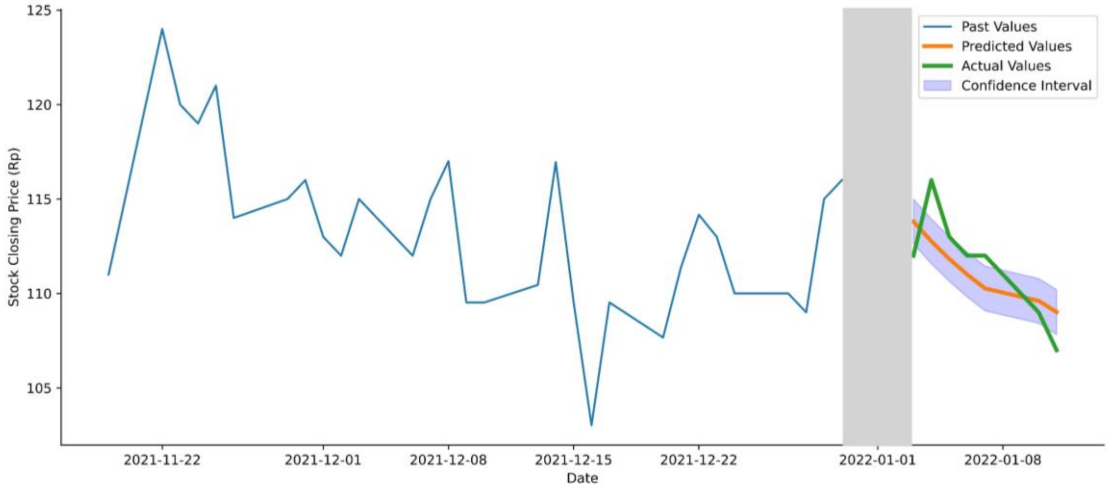
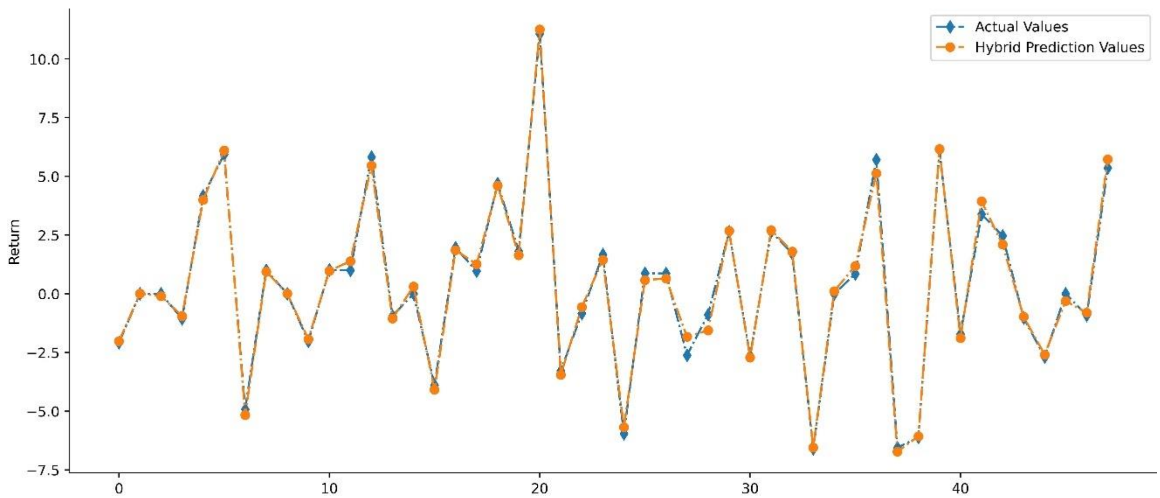
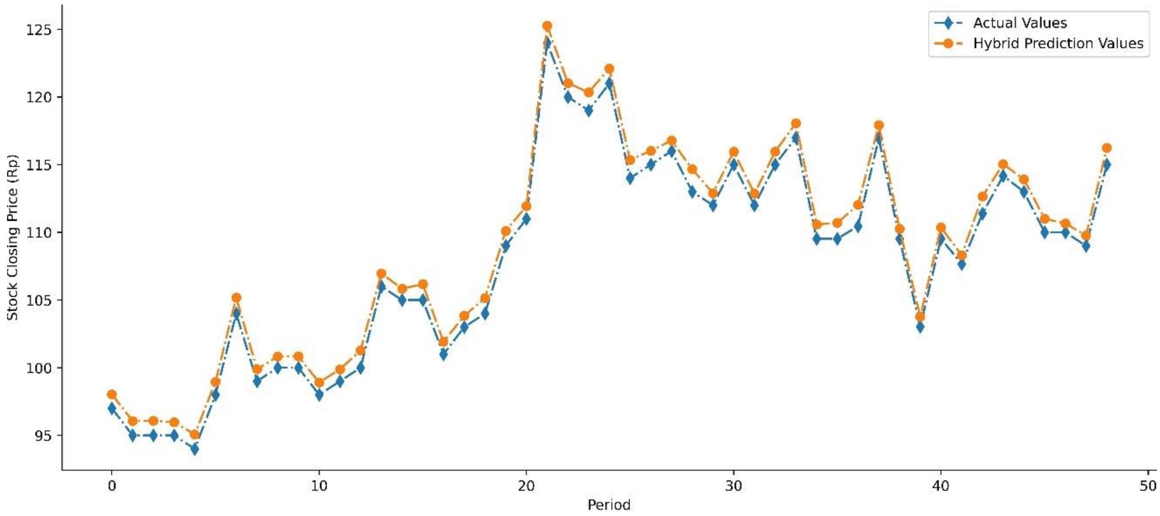


Fig 8. Stock closing price predictions of the LSTM model

The visualization of stock closing price predictions for seven days ahead using the LSTM model is presented in Fig 8. Based on the figure, the stock closing price predictions decrease from the first day until the seventh day. The stock closing price prediction on 3rd January 2022 is Rp 113.784 with a confidence interval of *[112.585, 114.983]*. This prediction is overestimated than the actual value. Meanwhile, the stock closing price prediction on 11th January 2022 is Rp 108.924 with a confidence interval of *[107.725, 110.123]*. Based on the figure, the predictions of the LSTM model tend to be underestimated than the actual values.



A



B

Fig 9. Fitted values of Hybrid GARCH-LSTM model: (A) return; (B) stock closing price

In the hybrid model, the residuals of the GARCH model are modelled using the LSTM model. The network structure for the LSTM model in the hybrid model is similar to that of the LSTM model, LSTM layer–Dropout Layer–Dense Layer. The best hyperparameter combination for the LSTM network in the hybrid model is the best number of neurons, learning rate, and epochs of the hybrid model of 40, 0.001, and 100, respectively. The hybrid GARCH–LSTM model validation is presented in Fig 9. The return predictions (Fig 9A) fit very well with the actual return values. However, few underestimated and overestimated predictions on the low volatility period exist. Fig 9B shows the predictions after being transformed into the stock closing prices. The stock closing price predictions of the hybrid model seem to be a little overestimated in all study periods. Still, the differences between the actual and the prediction values are very small. Generally, the prediction pattern follows the actual value pattern and fits very well.

The hybrid GARCH–LSTM model is then used to make predictions of the stock closing price seven days ahead. Based on Fig 10, the first stock closing price prediction on 3rd January 2022 decrease from that of the last study period with the prediction value of Rp 112.735 and the confidence interval of *[111.517, 113.953]*. This prediction is close to the actual value of Rp 112. The stock closing price prediction increases on the second day, 04th January 2022. The actual value of the stock closing price also increases on the second day. The predictions decrease from the third day until the seventh day, 11st January 2022, with the prediction value of Rp 108.692 and the confidence interval of *[107.474, 109.910]*.

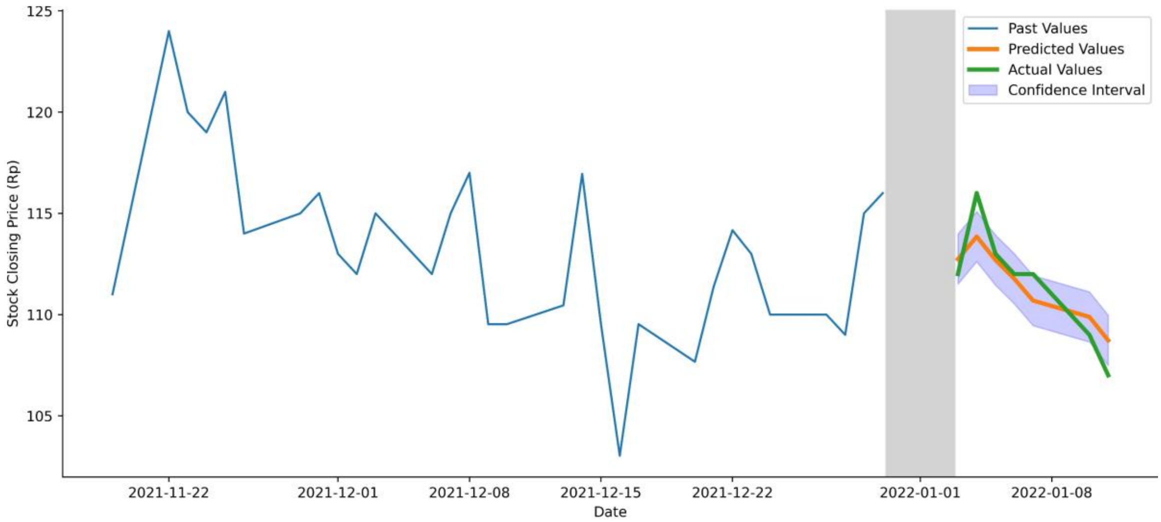


Fig 10. Stock closing price predictions of the Hybrid GARCH-LSTM model

Generally, all three models; GARCH, LSTM, and hybrid GARCH-LSTM; are able to capture the volatility pattern. Table 3 presents statistical values of model's performance evaluation in the form of MSE, RMSE, MAPE, and MAE. The GARCH and LSTM models have quite similar performance indicated by error values that are not significantly different. Meanwhile, the proposed model, the hybrid GARCH – LSTM, provides the lowest error values and thus is in line with the simulation study. In other words, the hybrid GARCH – LSTM model can recognize the volatility patterns better than the GARCH or LSTM models individually.

Table 3. Goodness of fit metrics

	MSE	RMSE	MAPE	MAE
GARCH	21.4082	4.6269	3.2093	3.5208
LSTM	20.6886	4.5485	3.1641	3.5195
Hybrid GARCH-LSTM	1.0652	1.0321	0.9378	1.0102

5 Conclusion

This study compares statistical, machine learning, and a hybrid of statistical and machine learning methods for analyzing the dynamic pattern of volatility data. The simulation study is carried out to learn how the proposed method work compared to the baseline models, GARCH and LSTM. The results show that the LSTM model is better than the GARCH model, and furthermore the Hybrid GARCH–LSTM model is the best model (i.e., it produces lower MAPE values). As the application of those models, the empirical study is applied to the daily stock closing price of BRMS. Generally, the GARCH model can capture the volatility pattern of the BRMS stock closing prices, but it has the overestimated and underestimated predictions in all study periods. Moreover, the LSTM model is able to recognize the volatility of the BRMS stock closing price. Furthermore, the hybrid GARCH–LSTM model performs very well in grasping the volatility pattern of the BRMS daily stock closing price. Both simulation and empirical studies give similar conclusion that the proposed model, Hybrid GARCH – LSTM, outperforms other two models.

For the future studies, this study can be extended in several ways. First, this study examines the dynamic pattern of volatility data with the normally distributed residuals or assumed to be normal. In practice, the data may have non-normal residuals. A study of non-normal residuals can be developed for the following research. Second, this study simulates the volatility data from the GARCH model; innovation can be done for the kind of data by simulating volatility data using models other than the symmetric GARCH model. Third, due to the manual hyperparameter tuning on the LSTM modelling process, the results of this study may not be optimum. The automatic hyperparameter tuning is needed. Last, the performance of the proposed model needs to be evaluated by comparing this model with other models, such as the hybrid of ARIMA-GARCH model.

ACKNOWLEDGEMENTS

We would like to thank all those who have helped provide valuable suggestions for improving the quality of this paper.

References

[1] Auffarth, B. (2021). Machine Learning for Time-Series with Python. *Introduction to Time-Series with Python* (pp. 1–33). Packt Publishing.

[2] Montgomery, D. C., Jennings, C. L., Kulahsi, M. (2008). Introduction to Time Series Analysis and Forecasting. *Autoregressive Integrated Moving Average*

(ARIMA) Models (pp. 231–298). John Wiley & Sons New Jersey.

[3] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica, 50 (4)*, 987–1007.

[4] Bollerslev. (1986). Generalized autoregressive conditional heteroscedasticity. *Econometric, 3*, 307–327.

[5] Borkowski, B., Krawiec, M., Karwański, M., Szczesny, W., Shachmurove, Y. (2021). Modeling garch processes in base metals returns using panel data. *Resources Policy, 74*, 1–13.

[6] Guo, Z. Y. (2022). Risk management of Bitcoin futures with GARCH models. *Finance Research Letter, 45*, 1–6.

[7] Busari, G. A., Lim, D. H. (2021). Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. *Computers and Chemical Engineering, 155*, 1–9.

[8] Shahid, F., Zameer, A., Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons and Fractals, 140*, 1–9.

[9] Luo, J., Zhang, Z., Fu, Y., Rao, F. (2021). Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. *Results in Physics, 27*, 1–9.

[10] Pacella, M., Papadia, G. (2021). Evaluation of deep learning with long short-term memory networks for time series forecasting in supply chain management, *Procedia CIRP, 99*, 604–609.

[11] ArunKumar, K. E., Kalaga, D. V., Kumar, Ch. M. S., Kawaji, M., Brenza, T. M. (2022). Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria Engineering Journal, 61 (10)*, 7585–7603.

[12] Ma, R., Zheng, X., Wang, P., Liu, H., Zhang, C. (2021). The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method. *Scientific Reports, 11 (1)*, 1–14.

[13] Elmaz, F., Eyckerman, R., Casteels, W., Latré, S., Hellinckx, P. (2021). CNN-LSTM architecture for predictive indoor temperature modeling. *Building and Environment, 206*, 1–9.

[14] He, K., Ji, L., Wu, C. W. D., Tso, K. F. G. (2021). Using SARIMA–CNN–LSTM approach to forecast daily tourism demand. *Journal of Hospitality and Tourism Management, 49*, 25–33.

[15] Cryer, J. D., Chan, K. S., (2008). Time Series Analysis: With Applications to R Second Edition. *Time Series Models of Heteroscedasticity* (pp. 277–318). Springer Science+Business Media.

[16] Gandin, I., Scagnetto, A., Romani, S., Barbati, G. (2021). Interpretability of time-series deep learning models: A study in cardiovascular patients admitted to Intensive care unit. *Journal of Biomedical Informatics, 121*, 1–8.

[17] Nabi, K. N., Tahmid, M. T., Rafi, A., Kader, M. E., Haider, M. A. (2021). Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks. *Results in Physics, 24*, 1–13.

[18] Karmaker, C. L., Halder, P. K., Sarker, E. (2017). A Study of Time Series Model for Predicting Jute Yarn Demand: Case Study. *Journal of Industrial Engineering*, 1–8.

[19] Chicco, D., Warrens, M. J., Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science, 7*, 1–24.

Notes on contributors

Laily Nissa Atul Mualifah is a lecturer at the Department of Statistics, IPB University, Bogor, Indonesia. Her main teaching and research interests include Time Series and Forecasting, and Statistical Machine Learning.

Agus Mohamad Soleh is an Associate Professor at the Department of Statistics, IPB University, Bogor, Indonesia. His main teaching and research interests include Classification Modeling, Statistical Downscaling, Neural Network, and Deep Learning. He has published several research articles in national and international journals of statistics.

Khairil Anwar Notodiputro is a Professor at the Department of Statistics, IPB University, Bogor, Indonesia. His main teaching and research interests include Small Area Estimation, Generalized Linear Mixed Models, Time Series and Forecasting, and Statistical Machine Learning. He has published several research articles in national and international journals of statistics.