

# Forecasting health sector stock prices using ARIMAX method

Muhammad Aprilianto<sup>1)\*</sup>, Mila Nirmala Sari Hasibuan<sup>2)</sup>, Syaiful Zuhri Harahap<sup>3)</sup>

<sup>1)2)3)</sup>Universitas Labuhanbatu, Indonesia

<sup>1)</sup>[crewcitycoss63@gmail.com](mailto:crewcitycoss63@gmail.com), <sup>2)</sup>[milanirmalasari7@gmail.com](mailto:milanirmalasari7@gmail.com), <sup>3)</sup>[syaifulzuhriharahap@gmail.com](mailto:syaifulzuhriharahap@gmail.com)

Submitted : Apr 28, 2022 | Accepted : May 1, 2022 | Published : May 2, 2022

**Abstract:** In daily stock trading activities, stock prices can experience ups and downs. The rise and fall of stock prices occurs due to changes in supply and demand for these shares. The COVID-19 pandemic did not have a negative effect, instead it had a positive impact on stock prices in health companies. Companies in the health sector experienced a fairly good profit of 10.46% in the fourth quarter of 2021. This fact made investors interested in buying shares in companies in the health sector in the hope of selling them when demand increased, resulting in doubled profits. Stock conditions continue to fluctuate every day, making investors need to pay attention and study the past data of the health sector company that will be selected before deciding to invest. Therefore, it is necessary to forecast stock prices in the health sector for the next several periods as a step in making investment decisions. The health sector companies that will be modeled are PT Kimia Farma (Persero) Tbk and PT Kalbe Farma Tbk. The method used in this study is the ARIMAX model. The test and analysis results show that based on the RMSE and MAPE values, the best model is ARIMAX(5,13) for PT Kalbe Farma Tbk shares with a MAPE value of 1% in in-sample data and 0.6% in out-sample data.

**Keywords:** ARIMAX; Company Stock; Economy; Health; Pharmacy.

## INTRODUCTION

The Indonesian capital market has a very large role for the country's economy. The existence of the capital market makes investors who have excess funds can invest their funds in various securities in the hope of getting a reward. This also benefits the company as a party that needs funds can use these funds to develop its business (Purnama & Juliana, 2019). In daily stock trading activities, stock prices can fluctuate. The rise and fall of stock prices occurs due to changes in supply and demand for these shares. Various factors can affect changes in demand and supply, for example the COVID-19 pandemic that has occurred since 2020 has resulted in a decline in company performance and the level of the economy, especially in Indonesia (Rusyida & Pratama, 2020).

In addition to having a negative impact on the level of the economy in Indonesia, the effects of the COVID-19 pandemic also have a positive impact on companies in several sectors, one of which is the health sector (Amanda, Ahmar, Sailendra, & Merawati, 2019). The high anticipation of the public in avoiding COVID-19 and the increasing awareness of health resulted in the health sector experiencing a significant advantage. According to the Central Statistics Agency, in the fourth quarter of 2021 the Indonesian economy grew by 3.69%, in terms of production, the largest growth occurred in the health services business and social activities by 10.46% (BPS, 2021). This makes investors interested in buying shares of health sector companies in the hope of selling them when demand increases. Stock conditions continue to fluctuate every day, making investors need to pay attention and study the past data of the health sector company that will be chosen to invest. This is very important for investors in knowing the prospect of future stock prices and minimizing various risks (Aziz, Kamaludin, & Pudjiastuti, 2020). Therefore, it is necessary to forecast stock prices in the health sector for the next several periods as a step in making decisions.

Research related to stock price forecasting has been widely discussed by a number of previous researchers. Forecasting stock prices using the Recurrent Neural Network method with seven variables using 1218 data has resulted in an accuracy of 94% (Suyudi, Djamal, & Masupah, 2019). By using the Long Short-Term Memory algorithm, stock price forecasting is carried out using the parameters of 215 neurons, 2 hidden layers, and 10 epochs to produce an MSE of 0.0015 (Arfan & ETP, 2019). Stock price forecasting analysis using the Support Vector Machine method has resulted in an accuracy of 96.41% with an RMSE value of 0.0932 (Fadilah,

\*name of corresponding author



Agfiannisa, & Azhar, 2020). An artificial neural network with the Backpropagaation method has also been able to predict stock prices with an average error of 3.38% (Untoro, 2020).

This study aims to forecast stock prices in the health sector in Indonesia. The health sector companies that will be modeled are PT Kimia Farma (Persero) Tbk and PT Kalbe Farma Tbk. The training data used in this study starts from January 1, 2021 to November 28, 2021 and the testing data used starts from November 29, 2021 to December 3, 2021. In this study the forecasting method chosen is the Autoregressive Integrated Moving Average with Exogenous Input (ARIMAX) model. because it is proven to have produced a better level of accuracy (Du, 2018). The formulation of the problem in this study is to determine the best model from ARIMAX in forecasting the stock price of the health sector based on the RMSE and MAPE values.

### LITERATURE REVIEW

The ARIMAX (Autoregressive Integrated Moving Average with Exogenous Input) model is an ARIMA model with the addition of predictor variables (Wei, 2006). Predictor variables can be factors that are thought to have a significant effect and dummy variables (Ling., Darmesah, Chong, & Ho, 2019). There are several steps in identifying the ARIMAX model. The first step is to identify the stationarity of the data in both the mean and variance. If the data is not stationary, then data transformation can be carried out, such as the Box-Cox transformation. Furthermore, the identification of ARIMAX orders based on ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) (Jing et al., 2018). A good ARIMAX model is one whose residual model meets the assumption of white noise and is normally distributed. The white noise condition was tested with the Ljung-Box test (Hossain, Ahmed, & Uddin, 2021).

Model evaluation is carried out to determine the goodness of the model in predicting stock prices for each issuer. One of the assessments that can be used for model evaluation is to calculate RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) (Dash & Dash, 2017). The RMSE provides information about the short-term performance of the model by allowing time-to-time comparisons of the actual difference between the estimated and measured values (Ma & Iqbal, 1984). MAPE is a calculation used to calculate the average absolute error, where the smaller the MAPE value indicates that the model's ability to forecast is getting better (Maricar, 2019).

### METHOD

The flow of research to forecast stock prices in the health sector in Indonesia using the ARIMAX method is illustrated in Figure 1. The data used in this study is secondary data in the form of stock price data of PT Kimia Farma (Persero) Tbk and PT Kalbe Farma Tbk obtained from <http://finance.yahoo.com>.

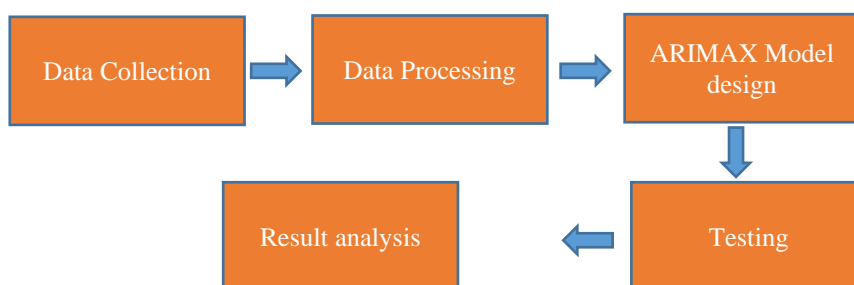


Fig 1. Research Flowchart

In-sample data starts from January 1, 2021 to November 28, 2021, while out-sample data starts from November 29, 2021 to December 3, 2021. The research variables used in this study are shown in Table 1.

Table 1. Research variable

Variable	Descriptions
Harga Saham PT Kimia Farma (Persero) Tbk	Harga saham periode 1 Januari 2021 s/d 3 Desember 2021
Harga Saham PT Kalbe Farma Tbk	Harga saham periode 1 Januari 2021 s/d 3 Desember 2021

The analytical model used in stock price modeling is the ARIMAX and NN models. The analytical steps used in this study are as follows:

1. Doing modeling and forecasting using the ARIMAX model with the following description:

\*name of corresponding author



- a. Divide in-sample data and out-sample data;
  - b. Checking the stationary and in-sample stock prices;
  - c. Modeling the stock price in-sample data by entering a dummy variable, then proceeding to the ARIMAX model, and checking the significance of parameters, white noise assumptions, and normal distribution assumptions;
  - d. If there is more than one ARIMAX model, it is necessary to choose based on the smallest RMSE value.
2. Choose the best model for forecasting stock prices for each issuer based on the RMSE and MAPE values, as well as forecasting from the best selected model.

### RESULT

This study models the stock prices of PT Kimia Farma (Persero) Tbk and PT Kalbe Farma Tbk using the ARIMAX model. In stock price modeling using the ARIMAX model, the exogenous variables used are the open price, high, low, and stock volume. Meanwhile, the metric variable is stock price. The first issuer to be analyzed is PT Kimia Farma Persero Tbk. Before doing the modeling, it is necessary to identify the data pattern using the time series plot shown in Figure 2.

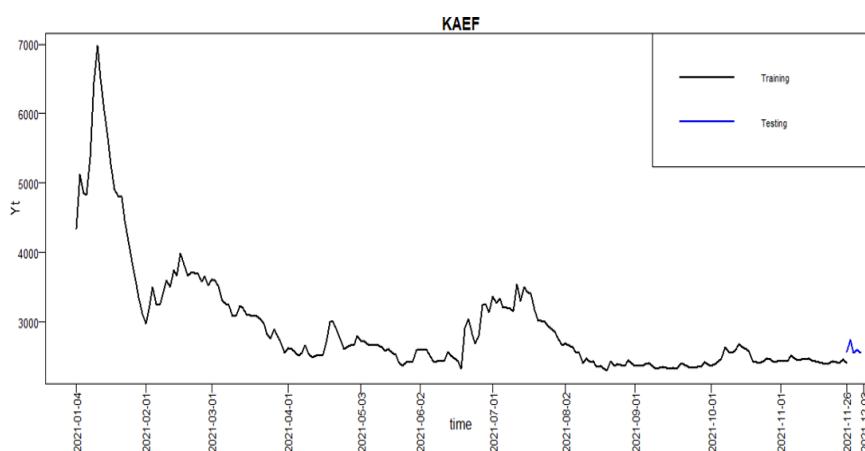


Fig 2. Share Price of PT Kimia Farma (persero) Tbk

Based on Figure 2, it can be seen that the data tended to trend down after January 12, 2021 until early February and after experiencing an increase, fell again until mid-June. Starting in mid-August, the share price of PT Kimia Farma (Persero) Tbk tends not to change significantly until December. This shows that the data is not stationary in the mean. Next, a stationary test of variance was carried out using the Box-Cox plot shown in Figure 3.

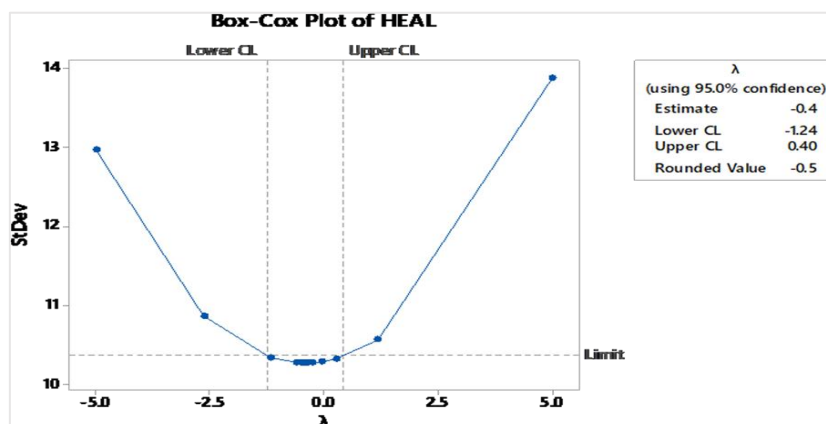


Fig 3. Plot Box-Cox Share Price of PT Kimia Farma (Persero) Tbk

Figure 3 shows that the rounded value is -2. This value is less than 1, so the data is not stationary in variance and needs to be transformed. The transformation used is  $= 1/Y$ . After the data is transformed, the rounded value is obtained with a value of 2 and the data is stationary in variance. Then the Augmented Dickey Fuller (ADF) test was carried out with the results shown in Table 2.

\*name of corresponding author



Table 2. ADF Test Results PT PT Kimia Farma (Persero) Tbk 's Share Price

Data	ADF	P-Value
Harga Saham PT Medikaloka Hermina Tbk	-3.413	0.053
Harga Saham PT Medikaloka Hermina Tbk setelah differencing	-5.512	0.010

Table 2 shows that the data needs to be differencing 1 to get a p-value less than of 0.05. The stock price that has been differencing 1 makes the data stationary in the mean. The results of estimation and testing of the significance of PT Kimia Farma (Persero) Tbk stock price data parameters using the ARIMAX model are shown in Table 3.

Table 3. Estimating and Testing the Significance of Parameters of the ARIMAX Model Share Price of PT Kimia Farma (Persero) Tbk

Model	Variable	Paramter	Estimate	Z	P-Value
ARIMAX(2,1,1)	MA (1)	$\theta_1$	-0,999	-68,945	< 2,2e-16
	Open	$X_{1,t}$	-0,554	-9,511	< 2,2e-16
	High	$X_{1,t}$	0,902	23,098	< 2,2e-16
	Low	$X_{2,t}$	0,653	11,018	< 2,2e-16
	Volume	$X_{4,t}$	-2,979	-2,347	0,01895

Table 3 shows the significant variables in the ARIMAX(2,1,1) model with a p-value of less than of 0.05. The next step is to test the white noise assumption for the modeling residuals with the results shown in Table 4.

Table 4. Residual White Noise Assumption Test ARIMAX Model Share Price of PT Kimia Farma (Persero) Tbk

Model	Hingga Lag	Chi-Square	P-Value
ARIMAX(2,1,1)	6	8.439	0.208
	12	14.399	0.276
	18	16.232	0.576
	24	17.622	0.821
	30	19.837	0.921
	36	24.275	0.932

Based on Table 4, it can be seen that the ARIMAX(2,1,1) model has reached the white noise condition, indicated by a p-value that is more than of 0.05. Then the normal distribution assumption is tested with the results shown in Table 5.

Table 5. Test the Normal Distribution Assumption of the ARIMAX Model Share Price of PT Kimia Farma (Persero) Tbk

Model	Kolmogorov-Smirnov	P-Value
ARIMAX(2,1,1)	0.096	0.034

Table 5 shows that the ARIMAX(2,1,1) model does not meet the normal distribution assumption because the p-value is less than of 0.05. The next issuer that will be modeled with ARIMAX is PT Kalbe Farma Tbk. Before doing the modeling, first identify the data pattern using the time series plot shown in Figure 4.

\*name of corresponding author



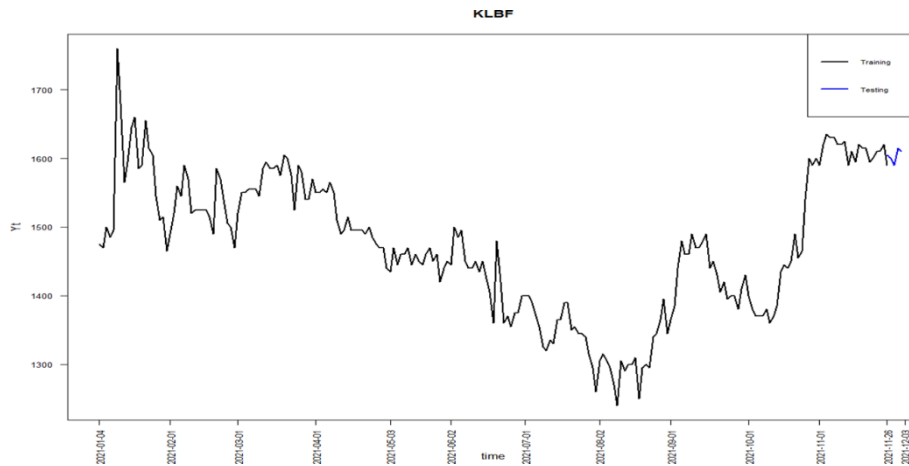


Fig 6. Share Price of PT Kalbe Farma Tbk

Based on Figure 6, it can be seen that the data tends to have a downward trend after January 11, 2021 until mid-August and after experiencing an increase, it declines again until mid-October. After that, the share price of PT Kalbe Farma Tbk tends to experience an upward trend until November and tends not to change significantly until December. This shows that the data is not stationary in the mean. Next, a stationary test of variance was carried out using the Box-Cox plot shown in Figure 7.

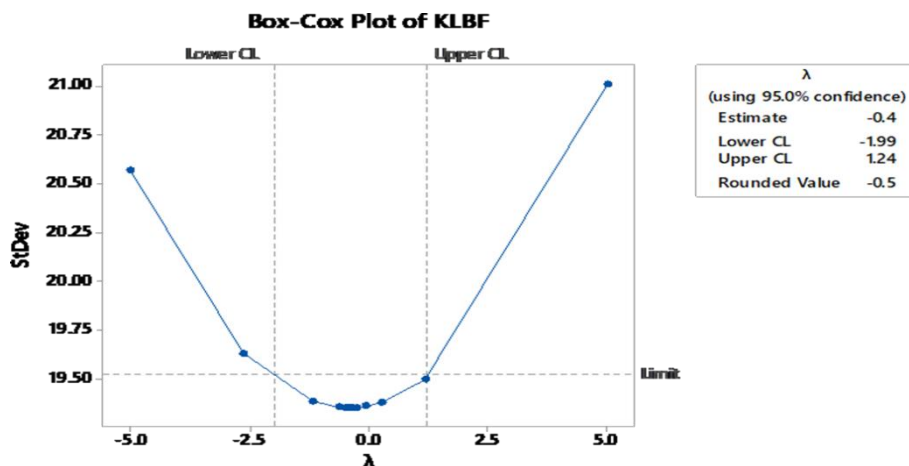


Fig 7. Plot Box-Cox Share Price of PT Kalbe Farma Tbk Before Transformation

Figure 7 shows that the rounded value is -0.5. This value is less than 1, so the data is not stationary in variance and needs to be transformed. The transformation used is  $= 1/\sqrt{Y}$ . After the data is transformed, it is obtained that the rounded value is 1 and the data is stationary in variance. Then the Augmented Dickey Fuller (ADF) test was carried out with the results shown in Table 14.

Table 6. ADF Test Results PT Kalbe Farma Tbk 's Share Price

Data	ADF	P-Value
Harga Saham PT Kalbe Farma Tbk	-1.123	0.917
Harga Saham PT Kalbe Farma Tbk setelah differencing	-6.090	0.010

Table 6 shows that the data needs to be differencing 1 to get a p-value less than of 0.05. The stock price that has been differencing 1 makes the data stationary in the mean. The results of estimation and significance testing of PT Kalbe Farma Tbk's stock price data parameters using the ARIMAX model are shown in Table 7.

Table 7. Estimating and Testing the Significance of Parameters of the ARIMAX

\*name of corresponding author



Model	Variable	Paramter	Estimate	Z	P-Value
ARIMAX(5,1,3)	AR(1)	$\phi_1$	-1,058	-9,842	< 2,2e-16
	AR(2)	$\phi_2$	-0,545	-3,520	0,0004308
	AR(3)	$\phi_3$	0,453	2,795	0,0051862
	AR(4)	$\phi_4$	0,321	2,588	0,0096662
	AR(5)	$\phi_5$	0,276	3,578	0,0003457
	MA(1)	$\theta_1$	0,328	3,939	8,178e-05
	MA(2)	$\theta_1$	-0,383	-5,618	1,934e-08
	MA(2)	$\theta_1$	-0,921	-12,218	< 2,2e-16
	Open	$X_{1,t}$	-0,616	-11,599	< 2,2e-16
	High	$X_{2,t}$	0,764	20,365	< 2,2e-16
	Low	$X_{3,t}$	0,813	14,446	< 2,2e-16

Table 7 shows the significant variables in the ARIMAX(5,1,3) model with a p-value of less than of 0.05. The next step is to test the white noise assumption for the modeling residuals with the results shown in Table 8.

Table 8. Residual White Noise Assumption Test ARIMAX  
Model Share Price of PT Kalbe Farma (Persero)

Model	Hingga Lag	Chi-Square	P-Value
ARIMAX(5,1,3)	6	0,559	0,997
	12	3,502	0,991
	18	8,824	0,964
	24	17,258	0,837
	30	17,258	0,837
	36	19,570	0,927

Based on Table 8, it can be seen that the ARIMAX(5,1,3) model has reached the white noise condition, indicated by a p-value that is more than of 0.05. Then the normal distribution assumption is tested with the results shown in Table 9.

Table 9. Test the Normal Distribution Assumption of the ARIMAX  
Model Share Price of PT Kalbe Farma (Persero) Tbk

Model	Kolmogorov-Smirnov	P-Value
ARIMAX(5,1,3)	0.074	0.171

Table 9 shows that the ARIMAX(5,1,3) model has met the assumption of a normal distribution because the p-value is greater than of 0.05. Furthermore, the selection of the best model to predict PT Kimia Farma (Persero) Tbk and PT Kalbe Farma Tbk is done by using the RMSE and MAPE values. Comparison of models for each issuer is shown in Table 10.

Table 3. Share Price Model Comparison of each Emiten

Emiten	Model	Data in-sample		Data out-sample	
		RMSE	MAPE	RMSE	MAPE
PT Kimia Farma (Persero) Tbk	ARIMAX(2,1,1)	58,797	0,011	56,046	0,020
PT Kalbe Farma Tbk	ARIMAX(5,1,3)	19,887	0,010	11,452	0,006

## DISCUSSIONS

The input used in the test model is the lag of the close, low, high, and open stock prices. Determination of the input used is based on the plot on each ARIMAX model of the share price of each issuer. In this study, it was found that there was a non-linear relationship between the data and the lags for the stock price of PT Kimia Farma (Persero) Tbk. However, the ARIMAX(2,1,1) model does not meet the normal distribution assumption because the p-value is less than of 0.05. While the stock price and PT Kalbe Farma Tbk provide a large p-value of of 0.05 or it can be said that there is no non-linear relationship in the stock price data.

\*name of corresponding author



## CONCLUSION

Based on the results of the tests and analyzes that have been carried out in this study, it is concluded that the best model in predicting the stock prices of PT Kimia Farma (Persero) Tbk and PT Kalbe Farma Tbk is the ARIMAX (5,1,3) model with a Mean Absolute Percentage Error of 1% on in-sample data and 0.6% on out-sample data.

## REFERENCES

- Amanda, T. R., Ahmar, N., Sailendra, & Merawati, E. E. (2019). Dampak COVID-19 Terhadap Tren Sektoral Harga Saham Syariah di Indonesia. *Fair Value : Jurnal Ilmiah Akuntansi Dan Keuangan*, 1(2), 307–313. Retrieved from <http://journal.uin-alauddin.ac.id/index.php/Iqtisaduna/article/view/20214>
- Arfan, A., & ETP, L. (2019). Prediksi Harga Saham di Indonesia Menggunakan Algoritma Long Short-Term Memory. *SeNTIK*, 3(1), 225–230.
- Aziz, M. A. A., Kamaludin, A., & Pudjiastuti, S. (2020). PENGARUH INDEKS HARGA KONSUMEN, INFLASI, DAN BI RATE PADA INDEKS HARGA SAHAM SYARIAH SEKTOR KESEHATAN Malik. *Iqtisadiya: Jurnal Ilmu Ekonomi Islam*, 7(14), 61–74. Retrieved from <https://journal.uinsgd.ac.id/index.php/iqtisadiya/article/view/10169>
- BPS. (2021). Pertumbuhan Ekonomi Indonesia Triwulan IV-2020. In [www.bps.go.id](http://www.bps.go.id). Retrieved from <https://www.bps.go.id/pressrelease/2021/02/05/1811/ekonomi-indonesia-2020-turun-sebesar-2-07-persen--c-to-c-.html>
- Dash, R., & Dash, P. K. (2017). *Chapter 25 - MDHS-LPNN: A Hybrid FOREX Predictor Model Using a Legendre Polynomial Neural Network with a Modified Differential Harmony Search Technique* (P. Samui, S. Sekhar, & V. E. B. T.-H. of N. C. Balas, eds.). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-811318-9.00025-9>
- Du, Y. (2018). Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network. *2018 Chinese Control And Decision Conference (CCDC)*, 2854–2857. <https://doi.org/10.1109/CCDC.2018.8407611>
- Fadilah, W. R. U., Agfiannisa, D., & Azhar, Y. (2020). Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine. *Fountain of Informatics Journal*, 5(2), 45. <https://doi.org/10.21111/fij.v5i2.4449>
- Hossain, M. S., Ahmed, S., & Uddin, M. J. (2021). Impact of weather on COVID-19 transmission in south Asian countries: An application of the ARIMAX model. *Science of The Total Environment*, 761, 143315. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.143315>
- Jing, Q. L., Cheng, Q., Marshall, J. M., Hu, W. B., Yang, Z. C., & Lu, J. H. (2018). Imported cases and minimum temperature drive dengue transmission in Guangzhou, China: evidence from ARIMAX model. *Epidemiology and Infection*, 146(10), 1226–1235. <https://doi.org/10.1017/S0950268818001176>
- Ling., A. S. C., Darmesah, G., Chong, K. P., & Ho, C. M. (2019). Application of ARIMAX Model to Forecast Weekly Cocoa Black Pod Disease Incidence. *Mathematics and Statistics*, 7(4), 29–40. <https://doi.org/10.13189/ms.2019.070705>
- Ma, C. C. Y., & Iqbal, M. (1984). Statistical comparison of solar radiation correlations Monthly average global and diffuse radiation on horizontal surfaces. *Solar Energy*, 33(2), 143–148. [https://doi.org/https://doi.org/10.1016/0038-092X\(84\)90231-7](https://doi.org/https://doi.org/10.1016/0038-092X(84)90231-7)
- Maricar, M. A. (2019). Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ. *Jurnal Sistem Dan Informatika*, 13(2), 36–45.
- Purnama, J., & Juliana, A. (2019). Analisa Prediksi Indeks Harga Saham Gabungan Menggunakan Metode Arima. *Cakrawala Management Business Journal*, 2(2), 454. <https://doi.org/10.30862/cm-bj.v2i2.51>
- Rusyida, W. Y., & Pratama, V. Y. (2020). Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode ARIMA. *Square : Journal of Mathematics and Mathematics Education*, 2(1), 73. <https://doi.org/10.21580/square.2020.2.1.5626>
- Suyudi, M. A. D., Djamal, E. C., & Maspupah, A. (2019). Prediksi Harga Saham menggunakan Metode Recurrent Neural Network. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, 33–38.
- Untoro, A. B. (2020). Prediksi Harga Saham Dengan Menggunakan Jaringan Syaraf Tiruan. *Jurnal Teknologi Informatika Dan Komputer*, 6(2), 103–111. <https://doi.org/10.37012/jtik.v6i2.212>
- Wei, W. (2006). *Time Series Analysis: Univariate and Multivariate Methods*, 2nd edition, 2006.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.