Volume 1 Nomor 2 | Desember 2023 | halaman 73-88

Inside | Jurnal Sistem Informatika Cerdas

Optimizing Forecasting of Dow Jones Stock Index in New York amid Uncertain Global Conditions in 2023: A Combined Approach of ARIMA and Machine Learning Models

Ferry Wiranto[1*], Iqbal Sabilirrasyad[2], Mas'ud Hermansyah[3]
[1, 2,3] Institute Technology and Science Mandala, Jember, Indonesian

Corresponding Author:
Ferry Wiranto, Institute Technology and Science Mandala
, Sumatera 118-120 Street, Jember and 68121, Indonesian
Email: ferry@itsm.ac.id

Abstract

This research aims to optimize the forecasting of the Dow Jones stock index in New York amid the uncertain global conditions of 2023 using a combined approach of ARIMA and Machine Learning models. This research aims to analyze the movement of the Dow Jones stock index during the uncertainty of the global economic conditions in 2023. The study used a quantitative approach with secondary daily data sourced from yahoo.finance website from january 2020 to may 2023. The current global uncertainty poses challenges to accurate forecasting, and this combined approach offers an effective solution. In this approach, the ARIMA model is employed to capture trends and seasonal patterns in historical data that exhibit stationarity, while the Machine Learning model is used to address more complex patterns and interactions among variables that cannot be handled by ARIMA. The uncertain data of the Dow Jones stock index in New York undergoes preprocessing stages using Machine Learning techniques such as data cleaning, data transformation, feature extraction, and data labeling. The results of the research demonstrate that the combined approach of ARIMA and Machine Learning provides more accurate and reliable forecasts. The integration of the ARIMA and Machine Learning models enables the capture of complex patterns and relationships in the data, resulting in improved forecasting accuracy and valuable insights for investors and market participants in navigating uncertain global conditions. In conclusion, the combined approach of ARIMA and Machine Learning is an effective strategy for optimizing the forecasting of the Dow Jones stock index in New York during uncertain global conditions in 2023. This research contributes to the field of financial forecasting by expanding the understanding of the use of combined approaches to enhance forecasting accuracy and support better decision-making amidst global uncertainty.

Keywords : *Forecasting, Dow Jones Stock Index, Uncertain Global Conditions, ARIMA, Machine Learning*

1 INTRODUCTION

The stock index is a statistical measure used to represent the overall movement of stock prices in a particular market or sector. Stock indices generally combine the stock prices of several companies listed in the stock market, and changes in the index reflect changes in the overall market value. Stock indices can be used as performance indicators of the stock market and serve as a reference for investors to measure their investment performance. Stock indices are typically formed using specific calculation methods, where the stock prices of companies included in the index are given certain weights. These weights can be based on market capitalization (total value of outstanding shares), stock prices, or other methods. The most well-known stock indices are the Dow Jones Industrial Average (DJIA) in the United States, S&P 500, and NASDAQ Composite.

Forecasting stock indices is an important part of financial market analysis. In uncertain global conditions, investors and market participants often face challenges in predicting the movement of stock indices. In 2023, global conditions are expected to become even more uncertain with rapid economic fluctuations and changes in other factors affecting the stock market, such as the ongoing Russia-Ukraine conflict and global interest rate hikes. Therefore, it is important to develop an approach that can optimize stock index forecasting in uncertain global conditions. In this context, accurate and reliable forecasting becomes crucial for investors and market participants to anticipate and provide insights into stock price movements and make informed decisions. One of the most widely followed stock indices and a benchmark for global stock indices is the Dow Jones index in New York. Accurate forecasts for this index provide valuable information for investors and help them optimize investment decisions. However, amid the uncertain global conditions of 2023, achieving accurate and reliable forecasts poses complex and challenging tasks.

The Dow Jones Industrial Average in New York is the focus of this research due to its popularity as a key indicator in the financial market . The DJIA reflects the performance of the 30 largest companies' stocks in the United States. As long as there is no macroeconomic condition that causes a spike in interest rates, investors will remain more motivated to invest in stocks [1]. The Dow Jones Industrial Average is widely regarded as the benchmark stock price index for decision-making and investor information. This is due to its status as the longest-running market performance index in the United States that is still active to this day [2]. Accurate forecasts for this index provide valuable information for investors and assist them in making investment decisions in several other global indices. In uncertain global conditions, accurate forecasts for the Dow Jones index are crucial to anticipate market changes and optimize investment decisions as stock indices in other parts of the world are often influenced by this index. In this research, our focus is on 2023, predicted to be a year full of uncertainty in the global financial market. The dow jones index may impact some indexes in other countries such as, The Dow Jones Index has a significant influence on the Composite Stock Price Index (IHSG) in the Indonesia Stock Exchange (IDX) during the period 2015-2019 [3].

Our main objective is to optimize the forecast of the Dow Jones index in New York for this year using a combined approach of ARIMA and Machine Learning models. Various models are organised to deal with various types of problems. A type of regression analysis known as ARIMA assesses the strength of one dependent variable in relation to other fluctuating variables [4]. This approach is expected to overcome the limitations of ARIMA models in handling nonlinear patterns and high variability in data while leveraging the strengths of Machine Learning techniques in capturing complex patterns and interactions between variables. The use of the Dow Jones stock index as the focus of this research is chosen due to its popularity as an important indicator in the financial market. By studying and optimizing the forecast for the Dow Jones index in New York under uncertain global conditions, this research can make a significant contribution to understanding market dynamics and assist investors and market participants in making better decisions.

The combined approach of ARIMA and Machine Learning has shown great potential in improving forecast accuracy in various fields, including financial forecasting. In this research, we combine the ARIMA model, which has the ability to capture trends and seasonal components, with Machine Learning models, which can handle more complex patterns and nonlinear interactions in data. This approach allows us to obtain more accurate and reliable forecasts for the Dow Jones stock index in New York under uncertain global conditions. The data preprocessing process is an important stage in this combined approach, and one of the techniques used is Machine Learning. Through data preprocessing using Machine Learning techniques, we can clean the data, transform the data format, and extract relevant features for forecasting. Thus, Machine Learning plays a crucial role in optimizing the data before being applied to the ARIMA model, which in turn produces more accurate and valid forecasts.

The results of this research are expected to provide valuable insights for investors and market participants in facing uncertain global conditions and assist them in making better investment decisions. The dow jones index is also affected by other global conditions such as Changes in the Federal Funds Rate [5]. Furthermore, this research can contribute significantly to the understanding of using the combined approach of ARIMA and Machine Learning in stock index forecasting. The results of this research do not serve as personal investment advice or otherwise, the information provided is merely an objective academic study related to providing an overview of the possible direction of the global stock index movement at the present time. It is hoped that this research can serve as a foundation for the development of more effective forecasting methods in the future.The results of this research are expected to provide valuable insights for investors and market participants in navigating uncertain global conditions and assist them in making better investment decisions. Additionally, this research can contribute significantly to understanding the use of the combined approach of ARIMA and Machine Learning in stock index forecasting. The results of this research do not serve as personal investment advice or the like; instead, they provide an objective academic study that offers an overview of the potential movements of global stock market indices. It is hoped that this research can serve as a foundation for the development of more effective forecasting methods in the future.

Volume 1 Nomor 2 | Desember 2023 | halaman 73-88

1.1. Objective of the study This

Firstly, the study aims to enhance the accuracy and reliability of forecasting the Dow Jones stock index in New York, considering the challenging global economic landscape of 2023 characterized by rapid fluctuations and geopolitical tensions. By leveraging a combined approach of ARIMA and Machine Learning models, the researchers seek to overcome the limitations of traditional ARIMA models and capture complex patterns and nonlinear interactions in the data. Secondly, the study intends to provide valuable insights and guidance for investors and market participants facing the uncertain global conditions of 2023. By optimizing the forecast of the Dow Jones index, the research aims to assist in making informed investment decisions and anticipating market changes. The Dow Jones index holds significant influence over other global indices, making accurate forecasts for this benchmark particularly valuable in navigating the dynamic financial landscape. Lastly, the study aims to contribute to the understanding and advancement of forecasting methodologies by combining the strengths of ARIMA and Machine Learning models. By integrating the trend-capturing capabilities of ARIMA with the ability of Machine Learning models to handle complex patterns and nonlinear relationships, the researchers aim to develop a more effective forecasting approach. This research has the potential to lay the groundwork for future studies in optimizing stock index forecasting under uncertain global conditions, providing valuable insights into the dynamics of financial markets and supporting decision-making processes.

1.2. Contribution to the study This

Makes significant contributions to the field of stock market analysis and forecasting. Firstly, it offers valuable insights into the application of a combined approach of ARIMA and Machine Learning models in optimizing the forecasting of the Dow Jones index. By leveraging the strengths of both methodologies, the study provides a more accurate and reliable forecast, addressing the limitations of traditional forecasting models and capturing complex patterns and nonlinear relationships within the data. This contribution enhances the ability of investors and market participants to anticipate market trends and make well-informed investment decisions.

Furthermore, the research contributes to the understanding of how to navigate and optimize forecasting under uncertain global conditions. By focusing on the year 2023, which is predicted to be a year filled with economic fluctuations and geopolitical tensions, the study provides valuable guidance to investors and market participants in mitigating the challenges posed by an uncertain global landscape. The findings shed light on the importance of accurate and reliable forecasting in such conditions, emphasizing the role of the Dow Jones index as a crucial indicator in global financial markets. Overall, this research contributes to the advancement of forecasting methodologies and provides practical implications for investors and market participants in an increasingly volatile and unpredictable financial environment.

2 RESEARCH METHOD

The research aims to optimize the prediction of the Dow Jones stock index in New York amid uncertain global conditions in 2023. To achieve this goal, this study utilizes a combined approach of the ARIMA model and machine learning. Additionally, in terms of maintaining the quality of the data used, a robust data transformation method is employed using the Python language. The robust data transformation method can help improve the performance of the prediction model by eliminating outliers and reducing the impact of invalid data, allowing the model to focus more on the actual patterns and trends in the data. This can enhance the accuracy and precision of the predictions, thereby providing more optimal results.

The initial step in this methodology is data collection. The data used in this research is obtained from the yahoo.finance website. The collected data includes daily stock prices, trading volume, and other relevant key economic indicators pertaining to the Dow Jones stock index. In this case, the researcher chooses to use daily stock prices because in uncertain market conditions as described in the research title, stock prices can experience significant daily fluctuations. By using daily stock prices, the research can capture market changes more quickly and be more responsive to economic, political, or financial events that impact stock prices. This allows the researcher to identify real-time trends and market changes, enabling more accurate and timely decision-making.

After the data is collected, the next step is data transformation using the robust method with the Python language. The robust method is employed to eliminate outliers and obtain data that is more resistant to anomalies. In this stage, machine learning is applied in the data transformation process to ensure better quality for further analysis, which will be processed using Minitab. Next, modeling is conducted using the ARIMA model. The ARIMA model is used to capture trends and seasonal patterns in the Dow Jones stock index data. In this stage, machine learning is not directly utilized, but the results of the previous robust data transformation will influence the performance of the ARIMA model. After obtaining the ARIMA model, the training and validation process is performed, including the identification of ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function). ACF and PACF are used to determine whether the data is stationary with respect to the mean. After plotting ACF and PACF, it is necessary to reconsider whether differencing is required for the data. In this case, differencing is needed as the ACF plot indicates that the data is not stationary with respect to the mean. The differencing method used is "first difference," which involves subtracting each data point from the previous data point in the time series. Thus, each data point in the series represents the change in value from the previous data point.

By utilizing the combined approach of the ARIMA model and machine learning, as well as leveraging the robust data transformation method using the Python language, this research aims to provide optimal predictions for the Dow Jones stock index in New York amid uncertain global conditions in 2023. The following is a framework for describing the steps in the ARIMA procedure in **Figure 1**.

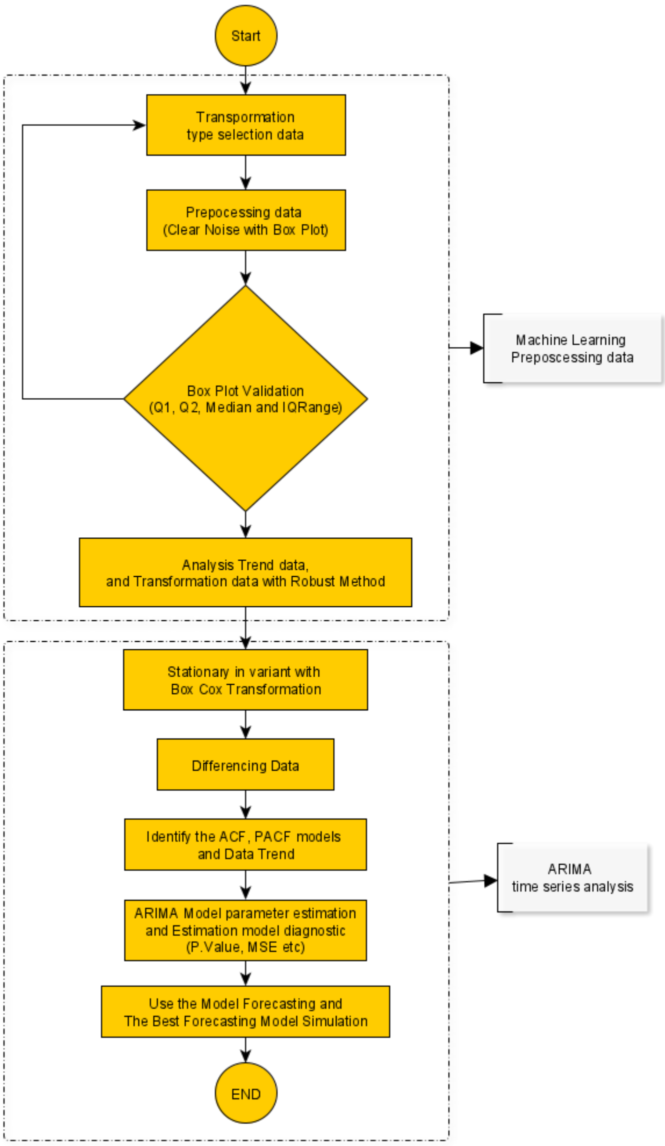

Figure 1 Framework procedure

3 RESULTS AND ANALYSIS

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily.

3.1. Machine Learning and Prepocessing Data

Machine learning is a branch of artificial intelligence that focuses on the development of algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed. Several activities within machine learning are extensive, ranging from data preparation, data processing, to final analysis and evaluation. The implementation of machine learning is currently very popular and widely developed for all fields and aspects of life. Some examples of implementation include: KNN method is used to find hotel recommendations [6]. Classification of political news with KNN [7], clustering of Umrah data [8], Neural Network Implementation Using the Perceptron Method for Recognition of Letters or Text [9] Etc. Machine Learning in Data Preprocessing is focused on the necessary steps to prepare data before using it in a Machine Learning model to predict the Dow Jones stock index in New York using the ARIMA model. Researchers utilize machine learning in the Detection and Handling of Outliers. Outliers are data points that significantly differ from the general pattern in the dataset. Outliers can impact the performance of the Machine Learning model and the prediction outcomes. Therefore, in data preprocessing, steps for outlier detection and handling are essential, such as utilizing statistical methods or other techniques, to minimize the influence of outliers on prediction results. **Figure 1** below illustrates the initial data plot, indicating that the data still contains outliers or a significant number of anomalies.

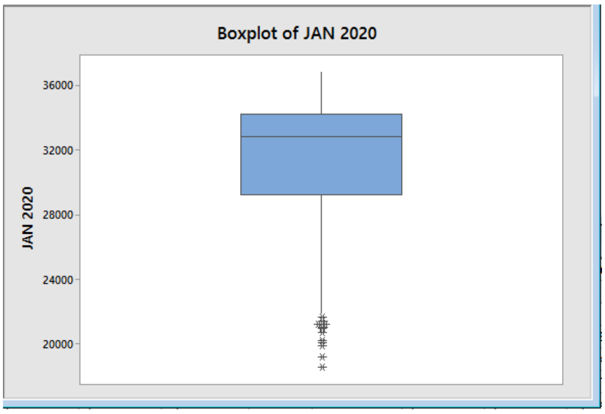

Figure 2 Box Plot Plot for Initial Data

This study also attempts to explore various data transformations to examine whether the data can be made outlier-free when transformed. However, after trying several transformation methods, no ideal Box Plot plots were found**. Figure 3** displays the different types of transformations used. As a result, the researcher decided to explore another option, which involves reducing the impact of outliers using a popular method typically suitable when the dataset contains significant outliers or when the presence of outliers can affect the analysis or the constructed model. In this case, the researcher heavily relies on the final model outcome and, therefore, opts for the Robust method.

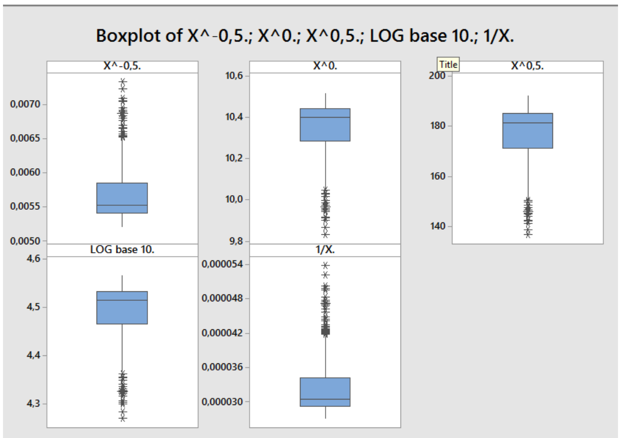

Figure 3 Various Types of Box Plot for Data Transformation

The robust method was implemented using the Python programming language. The method was analyzed to address outliers in situations where outliers can disrupt the analysis or the constructed model can be seen in **Figure 5**. This method provides better protection against the influence of outliers and can produce more consistent analysis and stable models, thus helping to shape the stationary data distribution in time series analysis. The implementation process of the robust method can be seen in **Figure 4**.

In the robust method, there is a value called MAD (Median Absolute Deviation) constraint of 1.4826. This value is used in the robust method because it is related to the assumption of the data distribution following a normal distribution. MAD is a statistical measure that quantifies how far individual data points deviate from the median value. In a normal distribution, MAD is related to the standard deviation with a scaling factor of 1.4826. Therefore, by multiplying MAD by this factor (1.4826), we obtain an estimation that approximates the standard deviation in a normal distribution. The use of the factor 1.4826 in the robust method allows for a direct comparison between MAD and the standard deviation in a normal distribution. In some implementations of the robust method, MAD values that exceed this limit may be considered as indications of outliers. By using the MAD constraint of 1.4826, the robust method can provide more consistent results, enabling further analysis of the Dow Jones index.

```
In [8]: data = np.array([28868,28634,28703,28583,28745,28956,28823,28907,28939,29030,29297,29348,29196,29186,29160,28989,28535,28722

In [9]: median = np.median(data)

In [10]: absolute_deviation = np.abs(data - median)

In [11]: MAD_constant = 1.4826  # Faktor konstanta untuk MAD
         threshold = 3 * MAD_constant * np.median(absolute_deviation)

In [12]: data_no_outliers = data[absolute_deviation <= threshold]
```

Figure 4 The process of removing outlier data with robust method

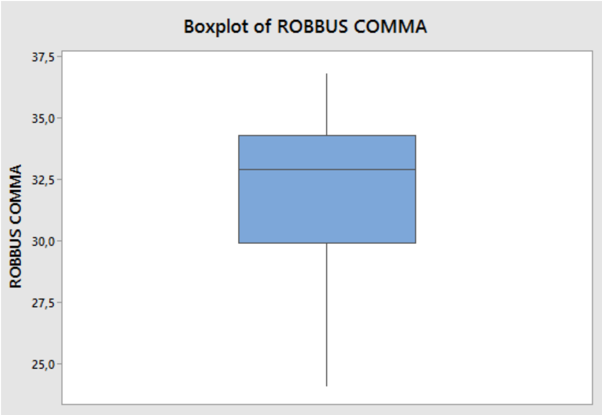

Figure 5 Plot Robust to eliminate outliers

3.2. Identification of Time Series Models

The application of the ARIMA method for forecasting Dow Jones is applied to the close data of the Dow Jones index, consisting of 816 data points from January 2020 to May 2023. The researchers used data starting from January 2020 because some data from 2019 showed severe anomalies due to the ongoing COVID-19 pandemic at that time. The impact of the Covid-19 pandemic has dropped many stock indexes in various countries to their lowest point in the last ten years [10]. The researcher utilizes daily closing transaction data, specifically the last point at the end of each session/closing for the Dow Jones index. This data provides more detailed information on point/price changes over time. With daily data, fluctuations in point/price can be observed within shorter periods, revealing trends, patterns, or changes that may be missed when only using weekly or monthly data. The pattern of the Dow Jones index's close data is shown in **Figure 6**.

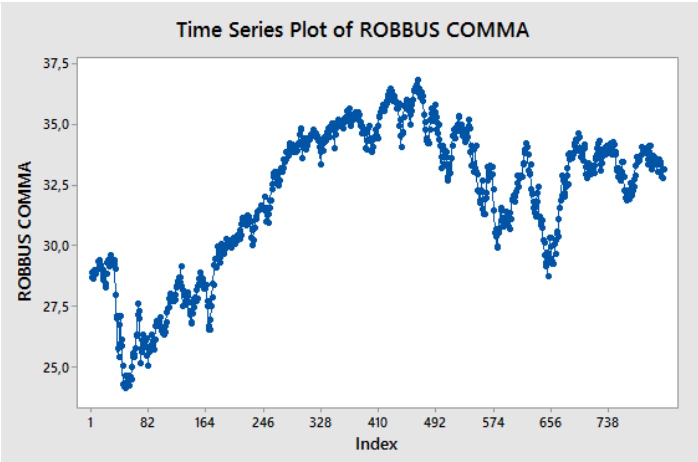

Figure 6 Plot of close data for the Dow Jones Index from January 2020 to May 2023

Based on the above **Figure 6**, it shows an increasing trend over time, indicating the presence of a fluctuating (non-stationary) trend pattern, often referred to as a stochastic trend. After understanding the data pattern, the next step is to examine whether the data has achieved stationarity in terms of mean and variance. If the data has not reached stationarity in terms of mean, a differencing process is required to reduce the data spread by taking the difference between the data at time *n* and the data at time *n+1*, commonly known as first differencing. On the other hand, if the data has not reached stationarity in terms of variance, a transformation process is needed. To evaluate stationarity in terms of variance, we can use the Box-Cox transformation in Minitab. If the rounded value or lambda (λ) is equal to 1, it can be said that the data has achieved stationarity in terms of variance. However, if lambda (λ) is not equal to 1, a transformation is performed until the rounded value in Box-Cox becomes 1. Thus, data transformation will be carried out to achieve stationarity in terms of variance before proceeding to the next step.

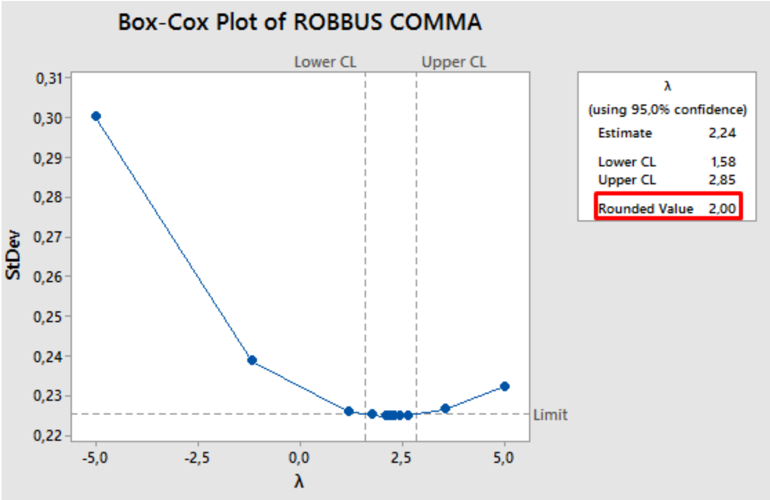

Figure 7 Box-Cox plot close index Dow Jones

Based on the results shown in **Figure 7**, it can be concluded that the data has not reached stationarity in terms of variance. This can be observed from the rounded value (lambda) obtained from the Box-Cox output, which is 2.00. This value indicates that the data is still not stationary in terms of variance because if the data had achieved stationarity in terms of variance, the rounded value should be 1. Therefore, data transformation is needed to achieve stationarity in terms of variance before proceeding with further analysis. The information provided in **Figure 7** provides a basis for applying the appropriate transformation to achieve the required stationarity in terms of variance for data analysis.

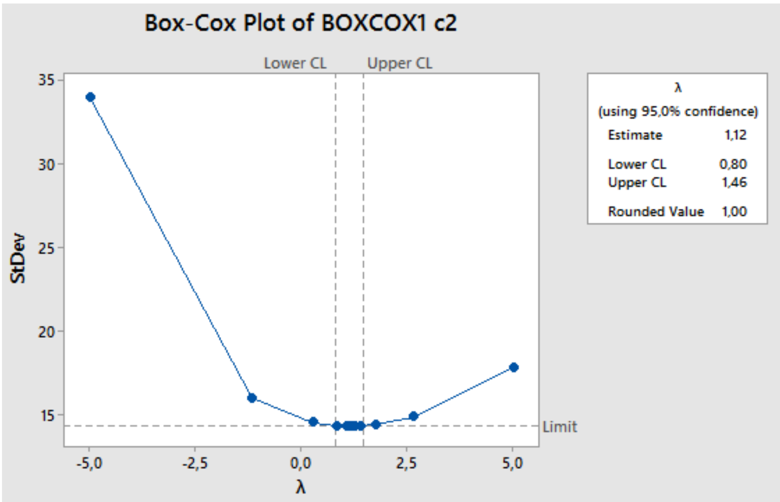

Figure 8 First Transformation Box-Cox data close Index Dow Jones

The researcher performed two Box-Cox transformations because the first transformation did not result in a rounded value of 1. The second transformation, as shown in **Figure 8**, yielded a rounded value of 1. Consequently, it can be concluded that the close data has achieved stationarity in variance. To ensure stationarity in mean, it is important to examine the time series plot and the ACF plot. Autocorrelation Function (ACF) or autocorrelation function is a function that shows the closeness of the relationship between observations at the t-time (Z_t) and observations at the previous time $(Z_{t-1}, Z_{t-2}, \ldots, Z_{t-k})$. Suppose $\bar{Z}$ is the sample mean where $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_t$, then the sample autocorrelation at lag 1 of Z_t is as follows.

$$\hat{\rho}_1 = \frac{\sum_{t=1}^{n-1}(Z_t - \bar{Z})(Z_{t+1} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2} \qquad (1)$$

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2} \quad (2)$$

where k is the time lag, $\hat{\rho k}$ is the estimated value of the sample autocorrelation function at lag-k, Z_t is the actual value at the t-time, $\bar{Z}$ is the average of observations of all periods in the data and Z_{t+k} is the observation at time to $(t + k)$ or the time after [11]. The Partial Autocorrelation Function (PACF) or the partial autocorrelation function between observations at time t (Zt) and observations at later times (Z_{t+k}) is defined as the correlation between Z_t and Z_{t+k} after the influence of variables Z_{t+1}, Z_{t+2}, …,Z_{t+k-1} has been removed. PACF is denoted by ϕ_{kk} The calculation of the lag k PACF sample value begins by calculating $\hat{\phi}_{1,1}$=$\hat{\rho}_k$, while the formula for calculating $\phi_{\text{k,k}}$ is as follows

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\phi}_{k,j} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\phi}_{k,j} \hat{\rho}_j} \quad (3)$$

And $$\hat{\phi}_{k+1,j} = \hat{\phi}_{k,j} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+1-j} \quad (4)$$

If the time series plot does not exhibit any trends or the ACF plot rapidly approaches zero, particularly after the second or third lag, it can be said that the data is stationary in mean. Additionally, another approach is to check if the number of lags in the ACF plots that exceed the confidence coefficient is not more than 3, indicating that the data is stationary in mean.

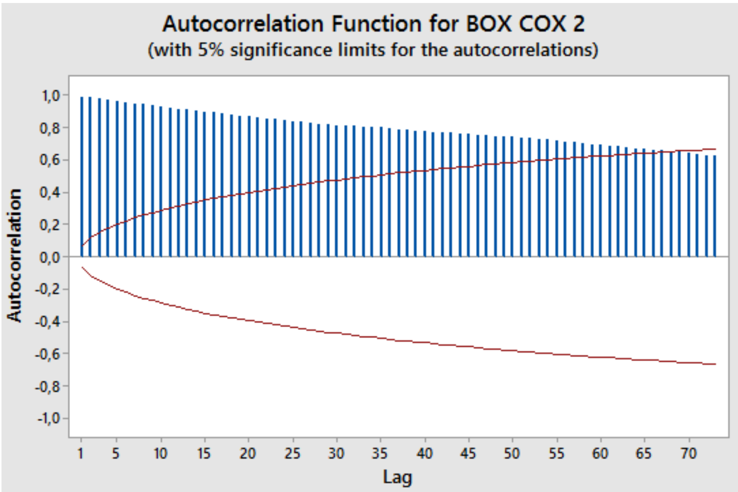

Figure 9 The results of the first transformation autocorrelation function data close Index Dow Jones

Figure 9 shows multiple consecutive time lags that exceed the significance threshold. This indicates that the close data of Dow Jones is still non-stationary in mean and requires data differencing. Differencing the data can help reduce existing trends or patterns in the data to achieve stationarity in mean. By taking the differences between observations at different time points, it produces data that is more stationary in mean.

.

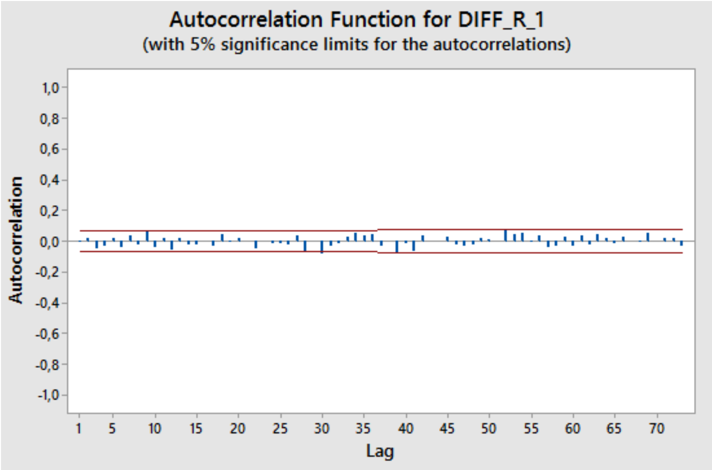

Figure 10 The results of the first differencing Autocorrelation function data close Index Dow Jones

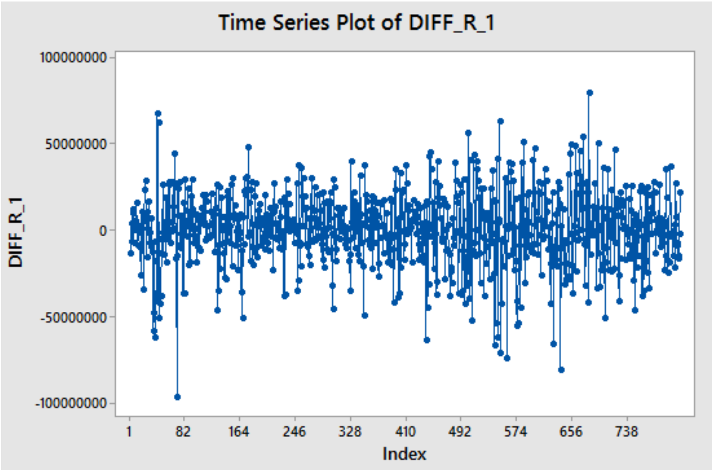

Figure 11 The output of the time series data for the close price Dow Jones after first differencing

After differencing, the data distribution becomes more uniform and there are no more than 3 lag values that exceed the confidence interval. Additionally, **Figure 10** illustrates that the data fluctuates around zero (constant). Based on these figures, it can be concluded that the data has achieved stationarity in both mean and variance. To identify the model for the data, the close Dow Jones data is plotted after differencing on the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) graphs. The following are the ACF in Figure 10 and PACF in **Figure 12** graphs used to estimate candidate models.

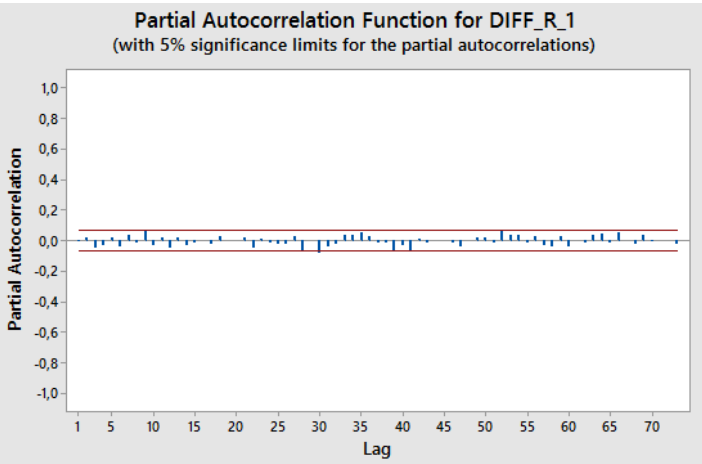

Figure 12 The results of the first differencing Partial Autocorrelation function data close Index Dow Jones

Based on **Figure 10**, it can be observed that in the ACF plot, the autocorrelation values at lag 30 and around 38 exceed the significant threshold (confidence interval) or fall outside the dashed lines. Similarly, in **Figure 12**, the PACF plot indicates that the partial autocorrelation value at lag 30 also exceeds the significant threshold. Therefore, it can be concluded that there are two lags in the ACF plot that surpass the significant threshold, and one lag in the PACF plot that surpasses the significant threshold. This suggests the presence of an Autoregressive (AR) process of order 2 and a Moving Average (MA) process of order 1. Consequently, several possible ARIMA models can be generated and included in the list of candidate models, which can be executed without Data Error in Minitab, as follows.

1. Model 1 : ARIMA (1,1,0)
2. Model 2 : ARIMA (2,1,0)

3.3. Estimation and Test of Significance of Parameters

The estimation phase is used to obtain coefficient estimates and the resulting model. Significance tests of parameters are conducted, where the model with the smallest MSE is chosen as the best model to be used in forecasting the Dow Jones index. This significance test of parameters is performed to evaluate the significance of each parameter in the candidate ARIMA model. Generally, the model with the lowest Mean Squared Error (MSE) is chosen as the best model because it provides a lower level of error in predicting the data. In the context of ARIMA, MSE is used to measure how accurately the model can predict the data. By selecting a model with the smallest MSE, we can generate more accurate forecasts that closely approximate the actual data.

Table 1 ARIMA model estimation close data Index Dow Jones

No	Model Estimation	Parameters	P. Value	Significance Test Results
1	ARIMA (1,1,0)	AR(1)	0,000	Significant
2	ARIMA (2,1,0)	AR(1) AR(2)	0,000 0,000	Significant

The significance threshold for p-values in time series analysis is not fixed. As a general rule, the commonly used significance thresholds are p-values less than 0.05 or 0.01, indicating a statistically significant level of analysis [12].In this study, a significance threshold below 0.01 was used to provide a higher level of confidence in decision-making. When the p-value falls below this threshold, researchers or decision-makers tend to be more confident that their findings can be replicated or applied more broadly. This provides additional assurance in the validity and generalizability of the findings.

.

3.4. Model Verification

After performing estimation and significance tests of parameters, all models exhibit significant results, leading to their selection as the best individual models. Next, the Mean Squared Error (MSE) values of each of these models will be evaluated. The following are the MSE values of the significant models.

Table 2 MSE value of the ARIMA model Dow Jones

No	Model Estimation	Mse	Selected
1	ARIMA (1,1,0)	5,641	N
2	ARIMA (2,1,0)	5,541	Y

Based on Table 2, the best model with the lowest MSE value is identified. The ARIMA (2,1,0) model exhibits the smallest MSE of 5.541. Subsequently, a check is conducted to determine whether the residuals of this model exhibit white noise properties or random behavior by examining the *p-value* and *Ljung-Box* statistics. The *Ljung-Box* test indicates that the ARIMA (2,1,0) model satisfies the assumption of white noise, meaning that the residuals are independent of each other or randomly distributed. The goodness of fit of the model is also assessed through the normality test of residuals.

The assumption of normal distribution in residuals allows for the use of statistical methods that rely on this assumption, such as hypothesis testing and confidence intervals. A normal distribution in residuals indicates that the ARIMA model has captured important patterns in the data and no significant information has been missed. Having normally distributed residuals facilitates the interpretation of results and makes the ARIMA model simpler for analysis and decision-making purposes.

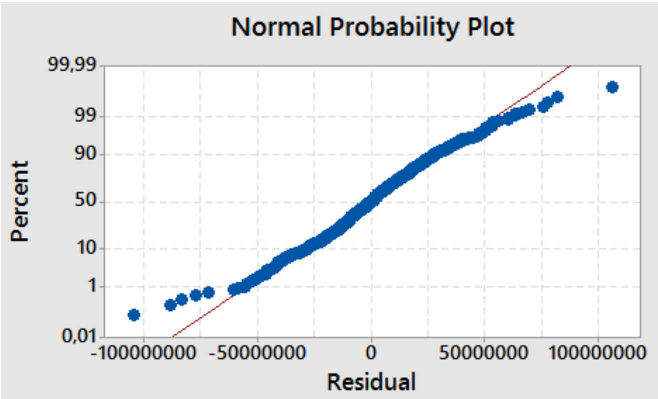

Figure 13 Output probability plot residual Minitab ARIMA (2,1,0)

In **Figure 13**, it can be observed that the distribution of residuals has approached a normal distribution as the data tends to cluster around the reference line. Therefore, the ARIMA (2,1,0) model can be considered suitable for forecasting the Dow Jones index. Furthermore, these results indicate that the model has a good ability to explain the variability and patterns in the available data.

3.5. Forecasting Data Close Dow Jones

Based on the selected model, which is ARIMA (2,1,0), a forecast has been made for the closing values of the Dow Jones data from January 2020 to May 2023. The equation used for this forecast is as follows:

$$Yt = Yt - 1 + (-0{,}6628) - 0{,}2933\varepsilon t - 1 \rightarrow \boldsymbol{Yt = Yt - 1 - 0,6628 - 0,2933\varepsilon t - 1} \quad (5)$$

Inside | Jurnal Sistem Informatika Cerdas
Volume 1 Nomor 2 | Desember 2023 | halaman 73-88

Table 3 Forecasting of the Dow Jones Index in New York for the period Juni 2023 to Oktb 2023 (150 days ahead)

Day	Forecasting		Day	Forecasting		Day	Forecasting
1	33,0516		16	33,1607		31	33,2348
2	33,0974		17	33,1657		32	33,2397
3	33,092		18	33,1706		33	33,2447
4	33,1027		19	33,1756		34	33,2496
5	33,1058		20	33,1805		35	33,2545
6	33,1116		21	33,1854		36	33,2595
7	33,1162		22	33,1904		37	33,2644
8	33,1213		23	33,1953		38	33,2693
9	33,1262		24	33,2002		39	33,2743
10	33,1311		25	33,2052		40	33,2792
11	33,1361		26	33,2101		41	33,2842
12	33,141		27	33,215		42	33,2891
13	33,1459		28	33,220		43	33,294
14	33,1509		29	33,2249		44	33,299
15	33,1558		30	33,2299		45	33,3039
Day	**Forecasting**		**Day**	**Forecasting**		**Day**	**Forecasting**
46	33,3088		61	33,3829		76	33,4569
47	33,3138		62	33,3878		77	33,4619
48	33,3187		63	33,3928		78	33,4668
49	33,3236		64	33,3977		79	33,4717
50	33,3286		65	33,4026		80	33,4767
51	33,3335		66	33,4076		81	33,4816
52	33,3385		67	33,4125		82	33,4866
53	33,3434		68	33,4174		83	33,4915
54	33,3483		69	33,4224		84	33,4964
55	33,3533		70	33,4273		85	33,5014
56	33,3582		71	33,4323		86	33,5063
57	33,3631		72	33,4372		87	33,5112
58	33,3681		73	33,4421		88	33,5162
59	33,373		74	33,4471		89	33,5211
60	33,378		75	33,452		90	33,526
Day	**Forecasting**		**Day**	**Forecasting**		**Day**	**Forecasting**
91	33,531		106	33,605		121	33,6791
92	33,5359		107	33,610		122	33,684
93	33,5409		108	33,6149		123	33,689
94	33,5458		109	33,6198		124	33,6939
95	33,5507		110	33,6248		125	33,6988
96	33,5557		111	33,6297		126	33,7038

.

97	33,5606		112	33,6347		127	33,7087
98	33,5655		113	33,6396		128	33,7136
99	33,5705		114	33,6445		129	33,7186
100	33,5754		115	33,6495		130	33,7235
101	33,5803		116	33,6544		131	33,7284
102	33,5853		117	33,6593		132	33,7334
103	33,5902		118	33,6643		133	33,7383
104	33,5952		119	33,6692		134	33,7433
105	33,6001		120	33,6741		135	33,7482
Day	**Forecasting**		**Day**	**Forecasting**		**Day**	**Forecasting**
136	33,7531		141	33,7778		146	33,8025
137	33,7581		142	33,7827		147	33,8074
138	33,763		143	33,7877		148	33,8124
139	33,7679		144	33,7926		149	33,8173
140	33,7729		145	33,7976		150	33,8222

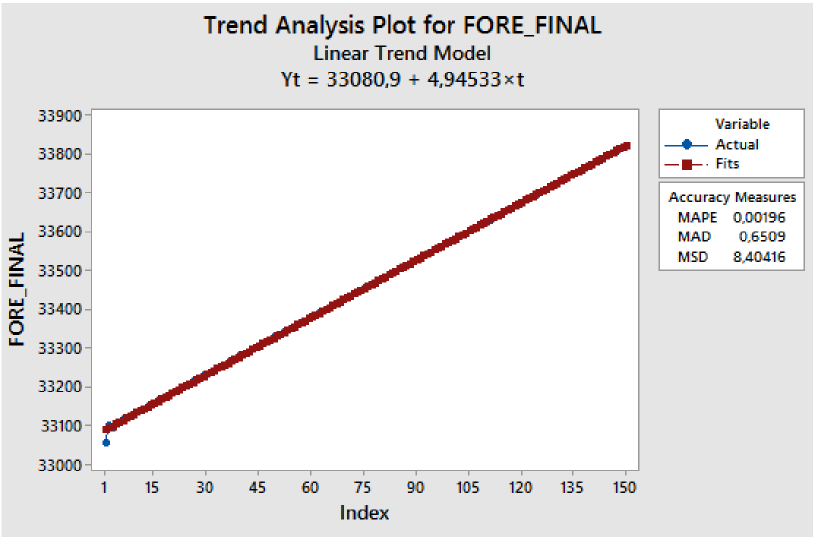

Figure 14 Plot prediction of the Dow Jones index score

Based on **Table 3** and **Figure 14**, it is found that the closing price of the Dow Jones Index has been increasing every day. In this case, there is a possibility that the stock prices in this index will continue to rise. This increase is not solely an invitation to invest in the Dow Jones index, but rather additional information considering that the fluctuations, whether up or down, are influenced by various external factors that need to be considered. Therefore, this information can serve as a guideline for investors to be more selective and cautious in their investments, especially in stock investments in the current global market benchmark, the Dow Jones Index. Although the index is predicted to gradually continue to rise, investors should still be selective in choosing which stocks are still good in the Dow Jones Index, as it is also certain which sectors will support the increase in the Dow Jones Index. This information consists solely of objective data and does not include any investment recommendations. In addition to being aware of the Dow Jones index trend, it is important to consider other factors such as the impact of the Fed rate and the domestic interest rate of the respective country [13].

4 CONCLUSION

The prediction results indicate that the Dow Jones Index is expected to continue its upward trend until the end of 2023. This provides an opportunity for investors to strategize their stock selection within the Dow Jones Index. The impact of the Dow Jones movement will also be felt by stock exchanges in several countries, such as Indonesia, Singapore, London, Hong Kong, etc., as the Dow Jones Index currently attracts the attention of global investors. Furthermore, the combined approach of ARIMA and Machine Learning models has shown promising results in optimizing predictions. Despite not utilizing an excessive number of significant models, this approach can generate an optimal model for predicting the movement of the Dow Jones Index. By harnessing the strengths of both methods, accurate predictions can be made, aiding investors in making better investment decisions.

Thus, this research makes a significant contribution to improving the quality of forecasting for the Dow Jones Stock Index amidst uncertain global conditions. Investors can utilize these prediction results as a guide in structuring their investment strategies, enhancing the potential for success and reducing risks in the stock market. In general, this study reveals that the movement of the Dow Jones Index is closely related to the global economic conditions. The findings of this research provide important quantitative insights into the movement of the Index, which can provide valuable information for investors in facing the situations and conditions of the global economy that can affect stock markets worldwide. In this era of globalization, understanding the connection between the local market and global economic conditions is crucial in making smart and effective investment decisions. By considering global economic factors in their investment strategies, investors can prepare themselves with data and knowledge to face potential changes and adjust their portfolios to the ongoing global stock market conditions. These findings can serve as a guide for investors to understand and anticipate the movement of the Dow Jones Index, which serves as a benchmark for indices in various countries, and then adapt to the conditions in their respective stock markets.

REFERENCES

[1] F. Fuad and I. Yuliadi, "Determinants of the Composite Stock Price Index (IHSG) on the Indonesia Stock Exchange," *J. Econ. Res. Soc. Sci.*, vol. 5, no. 1, pp. 27–41, 2021, doi: 10.18196/jerss.v5i1.11002.

[2] S. F. Tamara, "Pengaruh Dow Jones Industrial Average , Deutscher Aktienindex , Shanghai Stock Exchange Composite Index , dan Straits Times Index Terhadap Indeks Harga Saham Gabungan di Bursa Efek Indonesia (Periode 2010 – 2012)," 2012.

[3] D. Herlianto, "Pengaruh Indeks Dow Jones , Nikkei 225 , Shanghai Stock Exchange , Dan Straits Times Index Singapore Terhadap Indeks Harga Saham Gabungan (IHSG) Di Bursa Efek Indonesia (BEI)," vol. 03, no. Jones 2006, pp. 211–229, 2020.

[4] A. S. Salihin, "Pengaruh Indeks Bursa Global Dan Variabel Makro Ekonomi Terhadap Indeks Harga Saham Gabungan (Ihsg) Priode 2018-2020," *J-EBIS (Jurnal Ekon. dan Bisnis Islam.*, vol. 6, no. April, pp. 1–17, 2021, doi: 10.32505/j-ebis.v6i1.2390.

[5] K. Garg, "The Effect of Changes in the Federal Funds Rate on Stock Markets : A Sector-wise Analysis The Effect of Changes in the Federal Funds Rate on Stock Markets : A Sector-wise," vol. 4, no. 1, 2008.

[6] A. Muliawan, T. Badriyah, and I. Syarif, "Membangun Sistem Rekomendasi Hotel dengan Content Based Filtering Menggunakan K-Nearest Neighbor dan Haversine Formula," *Technomedia J.*, vol. 7, no. 2, pp. 231–247, 2022, doi: 10.33050/tmj.v7i2.1893.

[7] D. Afreyna Fauziah, A. Maududie, and I. Nuritha, "Klasifikasi Berita Politik Menggunakan Algoritma K-nearst Neighbor (Classification of Political News Content using K-Nearest Neighbor)," *Berk. Sainstek*, vol. 6, pp. 106–114, 2018, [Online]. Available: https://jurnal.unej.ac.id/index.php/BST/article/view/9256.

[8] M. Hermansyah, R. A. Hamdan, F. Sidik, and A. Wibowo, "Klasterisasi Data Travel Umroh di Marketplace Umroh.com Menggunakan Metode K-Means," *J. Ilmu Komput.*, vol. 13, no. 2, p. 8, 2020, doi: 10.24843/jik.2020.v13.i02.p06.

[9] J. O. F. Infocum, "IMPLEMENTATION OF NEURAL NETWORK WITH PERCEPTRON," vol. 10, no. 1, pp. 313–321, 2021.

[10] L. Winny and Y. Yulfiswandi, "Macroeconomics and the LQ45 Index: Is the COVID-19 pandemic making a difference?," *J. Manaj. Strateg. dan Apl. Bisnis*, vol. 5, no. 2, pp. 217–230, 2022, doi:

.

10.36407/jmsab.v5i2.612.

[11] S. C. Hillmer and W. W. S. Wei, “Time Series Analysis: Univariate and Multivariate Methods.,” *J. Am. Stat. Assoc.*, vol. 86, no. 413, p. 245, 1991, doi: 10.2307/2289741.

[12] G. S. Gowda, S. Komal, T. N. Sanjay, S. Mishra, C. N. Kumar, and S. B. Math, “Sociodemographic, legal, and clinical profiles of female forensic inpatients in Karnataka: A retrospective study,” *Indian J. Psychol. Med.*, vol. 41, no. 2, pp. 138–143, 2019, doi: 10.4103/IJPSYM.IJPSYM.

[13] I. S. Wicaksono and G. W. Yasa, “Pengaruh Fed Rate, Indeks Dow Jones, Nikkei 225, Hang Seng Terhadap Indeks Harga Saham Gabungan,” *E-Jurnal Akunt.*, vol. 18, no. 1, pp. 358–385, 2017.