Implementation Of The ARIMA Method In Predicting LQ 45 Stock Prices (UNTR Issuer)

Tegas Hadiyanto[1], Sarjon Defit[2], Rini Sovia[3]
[1,2,3]Faculty of Computer Science, Universitas Putra Indonesia YPTK, Padang, 25221, Indonesia
Tegas.Hadiyanto82@gmail.com

ABSTRACT

The implementation of technology is used in running businesses or activities that generate profits, such as predicting investments on the stock exchange through transaction data in the transaction data base. Machine learning is an algorithm that produces an approximation function that connects input variables so that it has the potential to be implemented in stock predictions. Stock investment has the characteristics of high risk - high return. Losses are caused by investors' lack of knowledge. Stock value analysis is divided into two, namely fundamental analysis and technical analysis. Technical analysis uses data or records about the market to try to access the demand and supply of a particular stock or the market as a whole. Based on the problems found by investors or bankers, this research will use the autoregressive integrated moving average (ARIMA) method to predict stock price movements. The Arima method consists of four stages, namely identifying time series methods, estimating parameters for alternative methods, testing methods and estimating time series values. Based on these problems, the ARIMA method will be used to predict stock movements. The Arima model (1,0,2) with RMS: 2200.576849857124 successfully predicted for the next 180 days.

Keywords: Data Mining, Forecasting, ARIMA, Stock.

INTRODUCTION

In the era of society 5.0, technology is part of humans themselves [23]. The implementation of technology is used in running businesses or activities that generate profits, such as predicting investments on the stock exchange through transaction data in the transaction data base [29]. Knowledge Discovery in databases (KDD) is the process of obtaining knowledge from data by emphasizing the application of data mining methods applicatively [15].

Data mining is a method for analyzing future patterns and characteristics and collecting unexpected information that has never been seen before from a large database, in this case data mining explores knowledge and patterns in data through mathematical statistics and machine learning [28]. Machine learning is an algorithm that produces an approximation function that connects input variables [26,33] so that it has the potential to be implemented in stock predictions.

Stock prediction is an implementation of supervised machine learning features. This really helps investors and bankers to predict stock movements that often change [8, 30, 13]. Machine learning can be applied to identify trends, patterns and changes in stock price behavior through historical data to predict future stock prices on the capital market [15, 27]. Shares are certificates that provide proof of company ownership [according to Silaban & Siahaan in journal 22]. The capital market is a place for stock transactions, because the capital market has an important role for the economy of a country, including the first for companies as a means for business funding or as a means for companies as additional funds from investors or investors, second for the public to invest in instruments. financial matters such as shares, bonds, mutual funds, and so on [9, 24].

Investment is the act of assigning capital as cash or other important resources to an item, organization, or party with the expectation that the financial backer will benefit after a certain period of time [19]. Stock investment has the characteristics of high risk - high return [43]. Losses are caused by investors' lack of knowledge [25]. According to Snariyah (2006), stock value analysis is divided into two, namely fundamental analysis and technical analysis. Technical analysis uses data or records about the market to try to access the demand and supply of a particular stock or the market as a whole. This analysis is determined by the size of demand and supply in the short term [3]

Based on the problems found by investors or bankers, this research will use the autoregressive integrated moving average (ARIMA) method to predict stock price movements. The ARIMA method is part of machine learning. The ARIMA method is a periodic series analysis method known as Box-Jenkins. According to Box-Jenkins, the Arima method consists of four stages, namely identifying the time series method, estimating parameters for alternative methods, testing the method and estimating the time series value. The Box-Jenkins procedure has the advantage of short-term prediction by modeling the dependence of data points and forecast errors. [25]. The ARIMA method provides better output because this method is based on a stationary time series regression model on closing data for each trade [32, 38, Ainiyah & Bansuri 2021). Based on the above, the ARIMA method will be used to predict stock movements.

METHODS

Framework This research was carried out using the framework formation method where this framework needs to be designed so that the research can be carried out in a structured and directed manner in accordance with the expected objectives. The framework is the stages of the research process which are ordered based on interrelated steps. These steps begin with the process of problem identification, literature study, data collection, system requirements analysis, feature extraction, designing the Arima model, implementing the model, testing the ARIMA model, and measuring the Arima accuracy level. We can describe the framework of this research as follows:

The following are the specifications of some of the tools that will be used:

a. Hardware:
   1. MSI Modern 14 B5M Laptop
   2. AMD Ryzen 7 5700U CPU @ 1.80GHz
   3. 16 GB memory
   4. 500 GB SSD storage
b. Software:
   1. Windows 11 Home 64-bit Operating System
   2. Microsoft 2021
   3. Google Collaboratory
   4. Python3 programming language
   5. Python libraries: pandas, numpy, matplotlib, seaborn, streamlit, statmodels, scikit-learn, ploty, etc.

Continuing the discussion regarding the research framework, the following is a detailed explanation of the systematic stages included in this research framework, as can be seen in Figure 3.1:

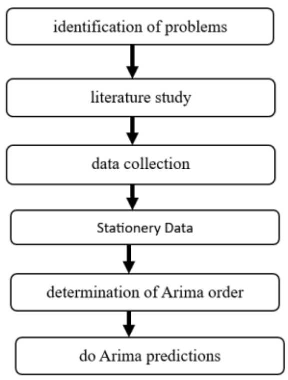

Figure 1. Research Methodology

Identifying the Problem

The initial step of this research starts from the problem identification stage. The problem identified is how an investor in the stock market gets input about future stock price predictions derived from analysis of stock transaction data in the previous period, as well as stock price movement patterns.

Conducting Literature Studies

A literature study was carried out to obtain a theoretical basis for the research subject, namely UNTR issuer shares. Then, in this stage, knowledge is also gathered about capital markets, investors, stock data, stationering methods, regression methods, Neural Networks, Python programming software and other supporting materials, including:

Knowledge Discovery in Database (KDD)

KDD is a comprehensive approach to data processing, starting from data acquisition to obtaining the expected results [15]. This approach aims to uncover patterns, trends, or hidden knowledge contained in large data sets [44]. This allows the discovery of new relationships, rules, or models that were previously unknown [15]. The knowledge gained can be used to aid decision making, understand phenomena, or build new theories [22, 32, 44].

Data Mining

Data mining is the process of finding interesting patterns and knowledge from large amounts of data [41]. Data mining consists of searching for desired trends or patterns in large databases to help make decisions in the future, these patterns are recognized by certain tools that can provide useful and insightful data analysis which can then be studied more thoroughly which may be in support of in another Decision [4] .

Machine Learning

Data science or machine learning is developing very rapidly in combination with several fields that help to process data and obtain useful information, such as mathematics, statistics and computers. The goal of combining with other sciences is to create artificial intelligence (AI) that can predict data with sufficient accuracy (prediction), help to make decisions, and make it continue to learn, so that it can be used for (machine learning). Making predictions, namely making predictions in the form of values, probabilities and data, and recommending the results as decision support or directly by the system. Science literally means "to know" [20]. According to Cholissodin, I., & Soebroto, A. A. (2021) Machine Learning (ML) or Machine Learning is a branch of AI that focuses on learning from data, namely focusing on developing systems that are able to learn "independently" without having to repeatedly programmed by humans. Machine learning is a technique used in artificial intelligence to allow machines to learn from data without being explicitly programmed. One machine learning algorithm that can be used is a classification with a supervised learning approach [31].

Autoreggressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) is a supervised machine learning algorithm. The ARIMA algorithm has advantages in forecasting time series data based on historical data regression [6]. Autoregressive Integrated Moving Average (ARIMA), also known as the Box and Jenkins, 1976 method, is based on stochastic process theory and features fewer data requirements, simple structure and fast modeling. Additionally, they can handle non-stationary time series [23]. The advantages of the ARIMA method include that it is flexible (follows existing data patterns), has a high level of accuracy and tends to have small error values because the process is detailed [17]. In making predictions, you must go through the Time Series analysis stage (Timeseries Data). Time series analysis is a statistical analysis that uses time series data patterns, with the assumption that the time series are correlated or dependent, such as daily sales data [7]. Here are the steps:

1. Checking Data

Checking the data must go through the following stages:

a. Stationary Data Creation

there is testing in the form of testing using Rolmean and RolStd to determine stationary in the variant. According to (W. Wei, 2006) Data stationarity is a necessary condition in time series regression analysis because it can reduce model errors, so that if the data is not stationary, then stationarity transformation must be carried out through a differentiation process, if the trend is linear, whereas if it is not linear, then The transformation must be carried out first by transforming the linearity of the trend through a natural logarithm process if the trend is exponential, and a weighting process (simple exponential smoothing) if it is in another form, then a differentiation process on the data resulting from the linearity process. There are two forms of stationarity, namely strong stationary (strickly stationer), or first order stationary (primary stationer) and weak stationary (weakly stationer), or second order stationary (secondary stationer).

Meanwhile, data non-stationarity is classified into three forms, namely: 1) Non-stationary on average, if the trend is not flat (not parallel to the time axis) and the data is spread over a "band" that covers the trend equally. 2) Not stationary in variance, if the trend is flat or almost flat but the data is spread out to form a widening or narrowing pattern that covers the trend in balance (trumpet pattern). 3) Non-stationary in mean and variance, if the trend is not flat and the data builds a trumpet pattern. Stationarity in the average can be identified visually, the first stage can be carried out on a data map over time using a time series plot, because it is usually "easy", and if it is not yet clear.

One of the features of ARIMA time series analysis is stationary data. Stationary means that there is no increase or decrease in data, which means stationary in the mean and stationary in the variance. This is proven by the formula:

$$r_t = \Theta_1 \, r_{t-1} + \alpha_t \text{ ....(1)}$$

Where :

rt : return value at time t

rt-1 : return value at time t-1

ϴ1 : Autoregressive parameter at lag 1

αt : residual value at time t

If the data is not stationary based on visuals then what is done is a differencing process in the average. Differentiation is the difference between the t-th data and the t-1 data. The following is the differencing formula:

$$\Delta Zt = Zt - Zt\text{-}1 \text{ ....(2)}$$

Where :

Zt = data to t

Zt-1 = data to t -1

b. Box-Cox Testing

The stationarity of variance in a time series model can be done visually using a time series plot by looking at whether the data pattern is widening or narrowing, while the stationary variance test can be explained in the form of a Box-Cox plot [37]. According to (B. Bowerman, 1993) Stationary in variance, time series data Zt is said to be stationary in variance if the variance value is constant. Using the Box-Cox plot, data is said to be stationary if the value $\lambda = 1$ or the interval has passed one. If the data is not stationary in terms of variance, then treatment is carried out using the Box-Cox transformation, with the formula:

$$t(rt) = \begin{cases} \frac{r_t^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \ln rt & \lambda = 0 \end{cases} \quad \text{..................................(3)}$$

Where :

rt = return value at time t

λ = is the lambda parameter

hypothesis: if $\lambda = 1$, the p-value < α it can be said that the data is stationary.

Stationary on average, RT time series data is said to be stationary on average if the data does not fluctuate around the average. If the data is not stationary in the average, then differencing is handled. The differencing process at the 1st lag can be shown by the equation.

$$X'_t = X_t - X_{t-1} \quad \text{............................(4)}$$

Where :

X't : is the first difference

Xt : X value of order t

Xt-1 : X value of order t-1

c. Augmented Dickey Fuller (ADF) Testing

If the "visual" study of stationarity in the average is less convincing, then statistical hypothesis testing for data stationarity needs to be carried out using the Augmented Dickey-Fuller Test.

Hypothesis: H0 : $\gamma = 0$ (data is not stationary)

H1: $\gamma < 0$ (data is stationary)

Test statistics: Augmented Dickeye-Fuller Test

$\tau = \gamma$^^/(se(γ^^)) ……………..…………(5)

Test criteria:

Reject H0 if p– value < α, accept in other cases

With:

$\hat{\gamma}$ = least squares estimate of γ

se(γ) = standard error of γ

Searching for unit roots (root tests) in time series data can be identified using Augmented Dickey Fuller (ADF). Testing using ADF on data can be concluded to have a unit root if the t-statistic must be smaller when compared to the test critical value or what is called the critical value. If there is one variable that is not stationary, the thing that must be done is to make a first difference and continue making subsequent differences until the data can be said to be stationary [24]. According to (Gujarati, 2003) the unit root test is a concept for testing the stationarity of time series data, by using the Augmented Dickey Fuller Test (ADF) there are three possibilities where the ADF test is estimated from three different forms of equations, namely:

d. Level Data

$\Delta Y_t = \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-1} + u_t$: no intercept.........(6)

$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-1} + u_t$: Intercept........(7)

$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-1} + u_t$:Intercept+trend..(8)

The steps for testing variables with the ADF test, the test is as follows:

1. The hypothesis used is:

   $H_0 : \delta = 0$ (nonstationary), against $H_1 : \delta < 0$

   The test statistics used are:

   $\tau = \frac{\delta}{se(\delta^{\wedge})}$ ....(9)

2. Reject H0 if τ the calculation result is greater than τ in the table or if the probability of the calculation result is smaller than the degree of confidence we want.
3. If the variable we are testing turns out to be non-stationary, then the data is differentiated and then a test is carried out on the data as in Step 1.

2. Determining the ARIMA Model

ARIMA models can be classified as (p,d,q) models where p is the number of autoregressive (AR) orders, q is the number of moving average (MA) orders in the prediction equation, and d is the order of differences required for stationarity. The formula for calculating ARIMA is as follows [46].

$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$ …. (10) Description:

yt = forecast result value

C = constant value

$\phi 1, \phi 2, \ldots, \phi p$ = autoregressive model parameters

$\theta 1, \theta 2, \ldots, \theta q$ = moving average model parameters

p = gap between t and $t - p$ q = gap between t and $t - q$

The three parameters p, d, q in the ARIMA method act as: p (order of autoregression) to determine how many lag variables are needed. d (order of integration) to determine the amount of differencing. q (order of moving average) to determine the number of window sizes for the MA

These three parameters can be seen based on the decomposition graph and plotting the ACF and PACF graphs to determine the number of lags required (parameter p). parameter d can be seen from the trend graph and q from seasonality (Farosanti, 2022). Determination of p,q through the ACF and PACF creation process is carried out as follows:

A. Auto Correlation Function (ACF)

The auto correlation function plays a role in identifying a suitable time series model by checking whether the ACF value effectively goes to zero after a certain lag. According to Makridakis, et al. (1999) ACF on lag-k is defined as follows [19] :

$$\rho_k = \frac{\sum_{t=1}^{n-k}(Zt-\bar{Z})(Zt+k-\overline{Zt}}{\sum_{t=1}^{n-k}\ (Zt-\overline{Zt})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

Where :

Zt : correlation between observations at time t

Zt+k : Z correlation on sequence t+k

B. PACF Process

PACF Function The Partial Auto correlation function is a set of Partial Auto correlations for various lag-k. The data $Zt+k$ from the stationary process is then regressed on $Zt+k-1$, $Zt+k-2$,...,Z_t. PACF is defined as follows [19] :

$$\frac{[illegible]}{1-\sum_{j=1}^{k-1}\theta_{k-1,j}\rho_j} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(12)$$

Where :

K : lag from time (1,2,3,.... k-1)

Ɵ : partial autocorrelation

C. ARIMA model performance measurement with Root Mean Square Error (RMSE)

Root mean Square Error (RMSE) is a quadratic scoring rule that also measures the average size of the error. RMSE is the square root of the mean squared difference between predicted data and actual observations. The formula for RMSE is (Ashari & Sadikin, 2020). The formula used is:

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(At-Ft)^2}{n}} \quad ....(13)$$

Where :

At = value of data

Ft = predicted data value

N = Number of data

5 Capital Markets

Economic actors in a country in the world certainly really need the scope of the capital market as a source of financing for business actors who need additional capital to carry out their business [14].

The capital market is a means of bringing together parties who have excess funds with parties who need funds. The main activity of the capital market is to trade (buy and sell) securities in the form of securities such as shares, bonds, equity and other securities, both issued by the government and private companies [34].

The capital market is defined as a means of funding and as a means for investment activities through efforts to provide various facilities and infrastructure for buying and selling activities and other related activities. Investors certainly expect a profit (return) on the type of investment they choose. In investing activities there is also a risk of failure that investors will experience, but this risk can be anticipated by analyzing the causes of risk from any information that influences the capital market reaction [48].

Share

Shares are certificates that function as proof of investor ownership of a company [50]. Shares on the market have varying prices.

The definition of share price based on (Jogiyanto, 2016) is the price that occurs on the stock market at a certain time which is determined by market players and determined by the demand and supply of the shares concerned in the capital market.

According to (Tandelilin, 2010) defines that stock prices are a reflection of investors' expectations regarding earnings factors, cash flow and the level of return required by investors, where these three factors are also greatly influenced by the macroeconomic conditions of a country as well as global economic conditions.

Investment

Investment is a sacrifice made now with the aim of obtaining something of higher value in the future [12]. Investment is a solution that can be used to manage finances. Many investors in the capital market certainly have a variety of financial instruments that can be used to invest. One of the popular investment instruments in Indonesia is shares [25].

Investment Knowledge is a basic understanding of investment such as the circumstances when investing, considerations in carrying out securities transactions, understanding the risks and profits after investing [4]. Shareholders also get 2 returns or profits in investing in shares, the first is capital gain, namely profits when the value of shares rises and the second is dividends, namely the distribution of profits from the company to its shareholders [8]. Capital gain itself is the difference in share prices when buying and selling shares that investors get.

Data Collection and Data Analysis

Data Collection

When viewed in terms of data collection methods or techniques, data collection techniques can be carried out using interviews, questionnaires, observation and documentation. The data collection technique in this research was carried out by means of documentation, namely by collecting secondary data obtained from the corner of the UPI "YPTK" stock exchange, with a time period of 01 January 2019 to 31 December 2023. The file is saved under the name UNTR.csv.

Analyzing Data

The next stage is to carry out an analysis of data requirements. This analysis is carried out by concentrating on the data preparation process. In this activity, the parameters and variables needed to carry out ARIMA training are determined. The data is analyzed to determine its suitability as a sample that will be used in the network training process. The data is prepared first by saving it in a file with the CSV extension format. By filling in the Date, Open, High, Low, Close, Adj Close, Volume columns. Data Analysis This stage is used to analyze the data in the file. At this stage the researcher consulted with the stock exchange corner at UPI YPTK Padang University, obtained a recommendation for the stock issuer whose data would be taken, namely the issuer UNTR.JK. In this consultation, the stock exchange said that stock data should contain the following fields: date of stock transactions, opening price of stock transactions, highest price of stock transactions, lowest price of stock transactions, highest price of stock transactions and closing price of stock transactions. Stock predictions based on closing stock price data.

1. Checking File Data

Data checking in the untr.csv file was carried out to ensure whether the data was suitable for processing in the ARIMA model. The steps taken are the following contents of the untr.csv file:

Date	Open	High	Low	Close	Adj Close	Volume
2019-01-01 00:00:00	27,350	27,350	27,350	27,350	16,418.375	0
2019-01-02 00:00:00	27,350	27,650	27,150	27,325	16,403.3691	2,171,300
2019-01-03 00:00:00	27,400	27,650	27,300	27,575	16,553.4453	3,808,500
2019-01-04 00:00:00	27,350	29,000	27,225	29,000	17,408.8789	10,285,000
2019-01-07 00:00:00	29,250	29,525	27,850	27,975	16,793.5684	8,798,500

Tabel 1. Untr csv

	Open	High	Low	Close	Adj Close	Volume
count	1,231	1,231	1,231	1,231	1,231	1,231
mean	24,267.9326	24,586.7384	23,913.1803	24,227.0715	17,060.4277	4,967,981.8034
std	4,149.5562	4,166.9449	4,123.3831	4,148.2492	4,015.9462	3,335,500.3808
min	12,800	13,250	12,000	12,600	7,956.9307	0
25%	21,675	21,875	21,375	21,625	14,270.9214	2,941,550
50%	23,700	24,050	23,300	23,700	15,968.1465	4,272,600
75%	26,987.5	27,237.5	26,562.5	26,875	20,535.6328	6,099,900
max	35,425	36,200	34,825	35,500	26,302.2832	50,516,600

Tabel 2. File Data Schema Untr csv

The contents of the untr.csv file are in accordance with recommendations from the UPI YPTK stock exchange corner, which contains: stock transaction date data, opening price data for stock transactions, highest price data for stock transactions, and closing stock price data.

The data schema obtained in the untr.csv file is as follows: that there are 1231 transaction data, the mean of the data on the opening price of stock transactions is 24,267.9326, the mean of the data on the highest price of stock transactions is 24,585.7384, the mean on the data of the lowest price of stock transactions is 23,913.1803, the mean on the closing price data for stock transactions was 24,227.0715, the mean for the Adj closing price data for stock transactions was 17,060.4277, the mean for the stock transaction volume data was 4,967,981.8034. The standard deviation (std) in the data on the opening price of stock transactions is 4,149.5562, the standard deviation (std) on the highest price data on stock transactions is 4,166.9449, the standard deviation (std) on the data on the lowest price of stock transactions is 4,123.3831, the standard deviation (std) on closing data stock transactions amounted to 4,148.2492, standard deviation (std) on closing Adj stock transaction data amounted to 4,015.9463, standard deviation (std) on stock transaction volume data amounted to 3,335,500.3808, minimum value (min) on opening stock transaction price data amounted to 12,800, minimum value (min) data on the highest price of share transactions is 13,250, the minimum value (min) on data on the lowest price of share transactions is 12,000, the minimum value (min) on closing stock transaction data is 12,600, the minimum value (min) on closing data Adj stock transactions is 7,956.9302 , the minimum value (min) in stock transaction volume data is 0. By processing 25% of the data, the value of the data on the opening price of stock transactions was 21,675, the data on the highest price of stock transactions was 21,875, on the data the lowest price on stock transactions was 21,875, on the data on the closing of stock transactions was 12,625, on the data on the closing Adj on stock transactions was 14,270.9224 , in stock transaction volume data of 2,941,550. By processing 50% of the data, the data value for the opening price of stock transactions was 23,700, the data for the highest price of stock transactions was 24,050, the data for the lowest price of stock transactions was 23,300, the data for closing stock transactions was 23,700, and the data for closing adj stock transactions was 15,968.1465 , in stock transaction volume data of 4,272,600. In processing 75% of the data, the data value for the opening price of stock transactions was 26,987, the data for the highest price of stock transactions was 247,237, the data for the lowest price of stock transactions was 26,562, the data for closing stock transactions was 26,875, and the data for closing adj stock transactions was 20,535.6328. , in stock transaction volume data of 6,099,900. By processing data as much as Max, the data value for the opening price of stock transactions is 35,425, the data for the highest price of stock transactions is 36,200, the data for the lowest price of stock transactions is 34,825, the data for closing stock transactions is 35,500, the data for closing adj stock transactions is 26,302, 2832, with stock transaction volume data of 50,516,600. The following is a visualization of the data above:

Figure 2. Visual Plot of Untr.CSV File Data

	0
Open	0
High	0
Low	0
Close	0
Adj Close	0
Volume	0

Tabel 3. File Data Checking Results

2. Indexing process

The next process is checking whether there are empty records in the data, the results of checking the data:

If it has been checked with the results that there is no empty data then the next process is the indexing process carried out between the close date and price, with the results:

Proses indexing

Date	Close
2019-01-01 00:00:00	27,350
2019-01-02 00:00:00	27,325
2019-01-03 00:00:00	27,575
2019-01-04 00:00:00	29,000
2019-01-07 00:00:00	27,975
2019-01-08 00:00:00	27,500
2019-01-09 00:00:00	28,100
2019-01-10 00:00:00	28,225
2019-01-11 00:00:00	27,800
2019-01-14 00:00:00	26,000

Tabel 4. Indexing Date with Close

3. Results and Discussion

If the data has been indexed then proceed to the next process, namely:

Stationary Process

In the stationering process, it goes through the following stages:

1. Checking Stationers

After ensuring that all data has been indexed, the next step is to check the stationarity of the data by meeting stationary criteria. Stationary means that data does not experience drastic changes within a certain time interval. Data is said to be stationary if data fluctuations are around a constant average value, independent of time and variations in these fluctuations. After ensuring that all data has been indexed, the next step is. Data plotting display results for the stationary test:

Figure 3. Data Stationary Test

Result of Dickey-Fuller Test:

	0
Test Statistic	-2.3466
p-value	0.1574
#Lags Used	1
Number of Observations Used	1,229
Critical Value (1%)	-3.4357
Critical Value (5%)	-2.8639
Critical Value (10%)	-2.568

Tabel 5. ADF Test Results

From the visual above it can be concluded that the data is not yet stationary because the data is still fluctuating and is not around a constant average value, it still depends on time and variations in these fluctuations. The next test by running the Adf test, obtained:This is proven by the ADF value, p-values > 0.05

2. Creating differentiated data

After knowing that the data is not stationary in the average, differencing is carried out. Differentiating means the original data value is replaced with the difference, the result is:

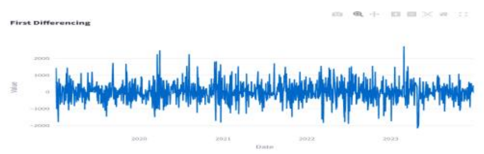

Figure 4. . Plot Visual Differencing Close

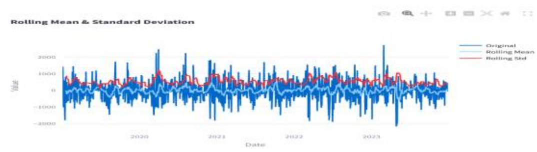

Figure 5. Visual Stationary Data Test

stationary data if data fluctuations are around a constant average value, independent of time and variations in these fluctuations, this is proven by the average value and variance value being parallel.

Box-Cox Transformation

The stationarity of variance in a time series model can be done visually using a time series plot by looking at whether the data pattern is widening or narrowing, while the stationary variance test can be explained in the form of a Box-Cox plot. Using a Box-Cox plot, data is said to be stationary if the value $\lambda = 1$ or the interval has passed one, box-cox test results

From the data above, the data is not stationary because the lambda value (λ) = 0.7438, $\lambda \neq 1$.

The next step is to differentiate the data:

The data is said to be visually stationary because the average value and variance value follow the original movement (close data). And the results of the box-cox process after being tested using ADF:The data is said to be stationary because p-value = 0.05. From the data above, it can be seen that the data movement based on the Cox box is not far from the closing stock price data.

Determination of ARIMA Order

In general, determining the order of the ARIMA time series model (p, d, q) can be seen based on the auto correlation function (ACF) and partial auto correlation function (PACF) plots.

1. Auto Correlation Function (ACF)

The auto correlation function plot is a graphical representation of the correlation of a time series with itself at different lags. The correlation coefficient is a measure of how closely two variables are related. A correlation coefficient of 1 indicates a perfect relationship, while a correlation coefficient of -1 indicates a perfect negative relationship. A correlation coefficient of 0 indicates there is no elationship between the two variables, the results are:

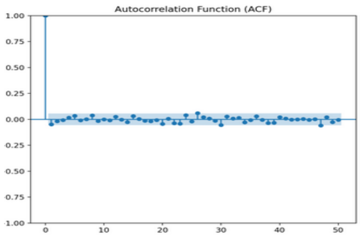

Figure 6. *Plot Visual ACF*

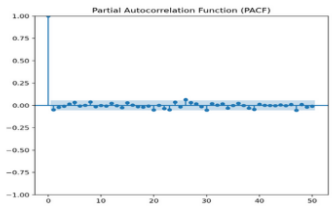

Figure 7. Plot Visual PACF

2. Partial Auto Correlation Function (PACF)

A PACF plot is a graphical representation of the correlation of a time series with itself at different lags, after removing the effects of previous lags. The PACF plot can be used to identify the MA model sequence. The order of the MA model is the number of lags included in the model. The PACF plot will show spikes at the lags included in the model

ARIMA Model Search (p,d,q) based on the smallest RMSE

One measure used to measure the accuracy of an ARIMA order is using the root mean square error (RMSE). The best model is the one with the smallest RMSE value. With results:

Prediksi Harga Saham untuk 180 hari ke depan:

	predicted_mean
1,401	23,861.1491
1,402	23,864.2943
1,403	23,867.4125
1,404	23,870.504
1,405	23,873.5688
1,406	23,876.6073
1,407	23,879.6197
1,408	23,882.6062
1,409	23,885.567
1,410	23,888.5024

Tabel 6. UNTR Share Price Prediction Results

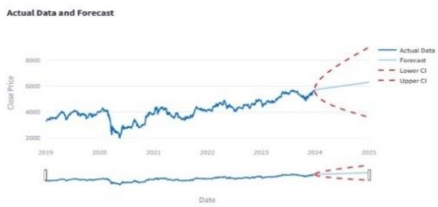

Figure 8. Visual Plot of UNTR Price Predictions

From the results above it can be concluded that the ARIMA model with order (1,0,2) is the best because it has the smallest RMS of: 2200, 576849857124.

UNTR issuer share predictions for the next 180 days

Make predictions with the ARIMA model with order (1,0,2) with rmse: 282.89935310267634. produce :

CONCLUSION

Based on the descriptions in the previous chapters, the following conclusions can be drawn:

1. Based on consultations with the UPI YPTK stock exchange, it was found that the issuers recommended for processing were UNTR shares. After going through the analysis process, the stock transaction data can be processed because there are no empty columns. The next step is to index the close column. The next process will be to carry out a visual stationary data process which indicates that the average stationary value and the variant stationary value move parallel to the time axis and box-cox transformation, so that the data is categorized as stationary, the p-value is <0.05. before determining the data model test, go through the auto correlation function and partial autocorrelation function process to ensure there is a correlation between each data. If each data has been correlated then the next step is to choose a model based on the smallest root mean sequare error (rmse). It was found that the best Arima model was (1,0,2) with an RMSE of 2200.576849857124.
2. Based on the Arima model (1,0,2) with RMSE: 2200.576849857124, successfully predicted UNTR shares for the next 365 years from January 1 2024 to December 31 2024, with the final price prediction of: 22991.100523919995.

REFERENCES

W. *Wei, 2006, Time Series Analysis Univariate and Multivariate. London:*

Gujarati, Damodar N. 2003. Basic Econometrics. Forth Edition, Mc Graw Hill.

Agustin, I. N., & Fariono, F. (2023). Perbandingan Analisis Teknikal dengan Pendekatan Moving Average dan Parabolic SAR dalam Memprediksi Pengembalian Saham pada Indeks Saham LQ45. Ekonomis: Journal of Economics and Business, 7(1), 606-613. Doi: http://dx.doi.org/10.33087 /ekonomis.v7i1.769

Adiningtyas, S., & Hakim, L. (2022). Pengaruh pengetahuan investasi, motivasi, dan uang saku terhadap minat mahasiswa berinvestasi di pasar modal syariah dengan risiko investasi sebagai variabel intervening. Jurnal Ilmiah Ekonomi Islam, 8(1), 474–482. DOI: http://dx.doi.org/10.29040/jiei.v8i1.4609

Andini, Y., Hardinata, J. T., & Purba, Y. P. (2022). Penerapan Data Mining Terhadap Tata Letak Buku Di Perpustakaan Sintong Bingei Pematangsiantar Menggunakan Metode Apriori. Jurnal Times, 11(1), 9–15. https://doi.org/10.51351/jtm.11.1.2022661

Ariyanti, V. P., & Yusnitasari, T. (2023). Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(2), 405–413. DOI: https://doi.org/10.29207/resti.v7i2.4895

Asrirawan, A., Permata, S. U., & Fauzan, M. I. (2022). Pendekatan Univariate Time Series Modelling untuk Prediksi Kuartalan Pertumbuhan Ekonomi Indonesia Pasca Vaksinasi COVID-19. *Jambura Journal of Mathematics*, *4*(1), 86–103. DOI: https://doi.org/10.34312/jjom.v4i1.11717

Chusanudin, A., & Munandar, A. (2022). Pengaruh Pengetahuan Keuangan Terhadap Niat Investasi Saham di Indonesia dengan Persepsi Risiko sebagai Variabel Moderasi. *Syntax Literate; Jurnal Ilmiah Indonesia*, *7*(4), 5051–5071. DOI: https://doi.org/10.36418/syntax-literate.v7i4.6880

Desy, D. I. P., Qur'ana, T. W., & Dharmawati, A. (2023). Pemodelan Spasial untuk Analisa Produksi Padi Integrasi Machine Learning. *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, *14*(2), 128-137. DOI: https://doi.org/10.31849/digitalzone.v14i2.16256

Damaliana, A. T., Muhaimin, A., & Riyantoko, P. A. (2023, November). Peramalan Lonjakan Kasus Harian Covid-19 Di Indonesia Dengan Model Arima. In *Prosiding Seminar Nasional Sains Data* (Vol. 3, No. 1, pp. 184-189). Doi : https://doi.org/10.33005/senada.v3i1.112

Elga, R., Murni, S., & Tulung, J. E. (2022). Reaksi Pasar Modal Terhadap Peristiwa Sebelum Dan Sesudah Pengumuman Covid-19 Di Indonesia (Event Study Pada Indeks LQ45). *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, *10*(1), 1052-1060. https://doi.org/10.35794/emba.v10i1.38703

Ernayani, R., Asnawi, M. I., Lumentah, N. R., Moridu, I., & Lestari, W. (2023). Literature Review: Prospek Peningkatan Suku Bunga terhadap Nilai Perusahaan dan *Return* Saham. *Journal of Economic, Bussines and Accounting (COSTING)*, *6*(2), 1168–1180. DOI: https://doi.org/10.31539/costing.v6i2.4906

Farosanti, L. (2022). Analisa Peramalan Penjualan Alat Kesehatan dan Laboratorium di PT. Tristania Global Indonesia Menggunakan Metode ARIMA. *JIMP (Jurnal Informatika Merdeka Pasuruan)*, *7*(1). DOI: http://dx.doi.org/10.37438/jimp.v7i1.428

Fillahi, A. F. B., & Maika, M. R. (2023). ANALISIS BIBLIOMETRIK PASAR MODAL SYARIAH: PUBLIKASI ILMIAH DI NEGARA-NEGARA SELURUH DUNIA.

Jurnal Tabarru': Islamic Banking and Finance, *6*(1), 164–178. https://doi.org/10.25299/jtb.2023.vol6(1).12110

Głowania, S., Kozak, J., & Juszczuk, P. (2023). Knowledge Discovery in Databases for a Football Match Result. *Electronics (Switzerland)*, *12*(12), 2712. https://doi.org/10.3390/electronics12122712

Hasibuan, E., & Karim, A. (2022). Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web. *Jurnal Ilmiah Komputasi*, *21*(4), 595-602. Doi : https://doi.org/10.32409/jikstik.21.4.3327

HIDAYAT, M. Z. (n.d.). *Peramalan Jumlah Wisatawan Menggunakan Metode Time Series Arima dan Feed-Forward Neural Network (Studi Kasus: Tempat Wisata Pasir Putih, Situbondo)*.

Hijrah, M., Rachman, H., & Rahayu, P. I. (2023). Peramalan Harga Saham Perusahaan Perbankan dengan Market Capitalization Terbesar di Indonesia Pasca-Covid19. *Journal of Mathematics: Theory and Applications*, *5*(2), 95-99. Doi : https://doi.org/10.31605 /jomta.v5i2.3238

Huda, M., Azizah, R. N. N., & Setyana, A. N. (2023). IMPLEMENTASI METODE ARMA DALAM PERAMALAN INFLASI PROVINSI BANTEN. *Jurnal Bayesian: Jurnal Ilmiah Statistika Dan Ekonometrika*, *3*(2), 210–221. https://doi.org/10.46306/bay.v3i2.66

Idris, M., Adam, R. I., Brianorman, Y., Munir, R., & Mahayana, D. (2022). Kebenaran dalam Perspektif Filsafat Ilmu Pengetahuan dan Implementasi dalam Data Science dan Machine Leaning. *Jurnal Filsafat Indonesia*, *5*(2), 173–181. https://doi.org/10.23887/jfi.v5i2.42207

Kristhy, M. E., Febrizha, M., Satya, O. R. S., Kumala, L., Ande, P. P., & Hidayah, N. (2022). ERLINDUNGAN HUKUM BAGI INVESTOR ASING DI INDONESIA. *Jurnal Komunikasi Hukum (JKH)*, *8*(2), 79-106. Doi. https://doi.org/10.23887/jkh.v8i2.47127

Jansevskis, M., & Osis, K. (2023). Knowledge Discovery Frameworks and Characteristics. *Baltic Journal of Modern Computing*, *11*(4), 686–702. https://doi.org/10.22364/bjmc.2023.11.4.08

Lestari, W., Juliza, M., Angreni, P., Rahmawati, S., Fitria, I. M., & Hendrianingsih, N. (2023). Peramalan Harga Obligasi Pemerintah Mengunakan Model ARIMA Box-Jenkins.

JIIP-Jurnal Ilmiah Ilmu Pendidikan, *6*(11), 9502–9506. http://dx.doi.org/10.54371/jiip.v6i11.3280

Marufi, A. (2023). Arif Ma'rufi UJI PERFORMA MODEL AUTO-ARIMA UNTUK PRAKIRAAN CURAH HUJAN BULANAN TAHUN 2021 DI PROVINSI JAMBI: ASSESSMENT OF AUTOMATED ARIMA MODEL FOR MONTHLY RAINFALL PREDICTION OVER JAMBI PROVINCE. *Buletin Meterologi, Klimatologi Dan Geofisika*, *3*(1), 1–9.

Mauliza, A. Y. I., & Canggih, C. (2023). MINAT BERINVESTASI PADA SAHAM SYARIAH: PENGARUH PENGETAHUAN INVESTASI, SISTEM ONLINE TRADING SYARIAH, DAN MOTIVASI. *ECONBANK: Journal of Economics and Banking*, *5*(1), 36–50. DOI: https://doi.org/10.35829/econbank.v5i1.284

Marlina, W. A., Armijal, A., & Nisa, M. K. (2023). ANALISIS PERAMALAN BOX JENKINS TERHADAP PENJUALAN DI UMKM IM LELE, PAYAKUMBUH. *Industri Inovatif: Jurnal Teknik Industri*, *13*(2), 105-115. Doi: https://doi.org/ 10.36040/industri.v13i2.6526

Nabella, S. D., Munandar, A., & Tanjung, R. (2022). Likuiditas, solvabilitas, aktivitas dan profitabilitas terhadap harga saham pada perusahaan sektor tambangan batu bara yang terdaftar di Bursa Efek Indonesia. *Measurement Jurnal Akuntansi*, *16*(1), 97-102. DOI:https://doi.org/10.33373/mja.v16i1.4264

Nada, D. Q., & Syaiful, S. (2022). Faktor-Faktor Yang Mempengaruhi Minat Mahasiswa Berinvestasi Di Pasar Modal Pada Masa Pandemi Covid-19. *Maksimum: Media Akuntansi Universitas Muhammadiyah Semarang*, *12*(1), 42-52. Doi : https://doi.org/10.26714/mki.12.1.2022.42-52

Nasution, S. A., Lasmi, A., Silalahi, P. R., & Nasution, A. (2023). Efektivitas Galeri Investasi Syariah Bursa Efek Indonesia (GIS BEI) UINSU Medan Dalam Meningkatkan Literasi Pasar Modal. *El-Mal: Jurnal Kajian Ekonomi & Bisnis Islam*, *4*(3),548-559. Doi:https://doi.org/10.47467/elmal.v4i3.1931

Nurani, A. T., Setiawan, A., & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains Dan Edukasi Sains*, *6*(1), 34–43. https://doi.org/10.24246/juses.v6i1p34-43

Nurcahyo, J. A., & Sasongko, T. B. (2023). Hyperparameter Tuning Algoritma Supervised Learning untuk Klasifikasi Keluarga Penerima Bantuan Pangan Beras. *Indonesian Journal of Computer Science*, *12*(3).

Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile. *Entropy*, *23*(4), 485. https://doi.org/10.3390/e23040485

Pangaribuan, J. J., Fanny, F., Barus, O. P., & Romindo, R. (2023). Prediksi Penjualan Bisnis Rumah Properti Dengan Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA). *Jurnal Sistem Informasi Bisnis*, *13*(2), 154-161. Doi: https://doi.org/10.21456 /vol13iss2pp154-161

Prasetio, R., Iswanaji, C., & Khotijah, S. A. (2023). Pengaruh Persepsi *Return*, Risiko, Pengetahuan dan Religiusitas Terhadap Minat Berinvestasi di Pasar Modal Syariah. *Al-Intaj: Jurnal Ekonomi Dan Perbankan Syariah*, *9*(1), 88–108. DOI: http://dx.doi.org/10.29300/aij.v9i1.2706

Pratama, B. B., & Wijayanti, D. (2023). PERANCANGAN MODEL EKONOMI KREATIF 5.0 BERBASIS DIGITAL SOCIAL INNOVATION. *Jurnal Fokus Manajemen Bisnis*, *13*(1), 76-90.

Putri, D. I., & Putra, M. Y. (2023). Komparasi Algoritma dalam Memprediksi Perubahan Harga Saham GOTO Menggunakan RapidMiner. *Jurnal Khatulistiwa Informatika*, *11*(1), 14-20. DOI: https://doi.org/10.31294/jki.v11i1.16153.g6018

Rahayu, P. I. R., Hidayatullah, M., & Hijrah, M. (2023). Implementation Vector Autoregressive (Var) On Rice Production and Rice Productivity Data in Indonesia. *Jurnal Matematika, Statistika Dan Komputasi*, *19*(3), 580–592. https://doi.org/10.20956/j.v19i3.24881

Rapi, A. (2022). Penerapan Model ARIMA Dalam Memprediksi Penjualan Produk Minuman Teh Botol Sosro Ukuran 350 mL. *Inventory-Industrial Vocational E-Journal On Agroindustry*, *3*(2), 69-82. Doi : http://dx.doi.org/10.52759/inventory.v3i2.99

Riyandi, A., Aripin, A., Ardiansyah, I. N., Dany, R., & Yusrizal, Y. (2023). Analisis Data Mining untuk Prediksi Harga Saham: Perbandingan Metode Regresi Linier dan Pola Historis. *Jurnal Teknologi Sistem Informasi*, *4*(2), 278–288. DOI: https://doi.org/10.35957/jtsi.v4i2.5158

Rizal, M., & Yustanti, W. (2023). Peramalan Close Price Mata Uang Crypto Solana Menggunakan Jaringan Syaraf Tiruan Model Backpropagation. *Journal of Emerging Information System and Business Intelligence (JEISBI)*, *4*(4), 37-46.

Samosir, V. B., Widodo, A. M., Anwar, N., Sekti, B. A., & Erzed, N. (2024). Identifikasi Outlier Menggunakan Teknik Data Mining Clustering Untuk Analisis Data Tracer Study Pada Fakultas Ilmu Komputer Universitas Esa Unggul. *IKRA-ITH Informatika: Jurnal Komputer Dan Informatika*, *8*(1), 162–174. https://doi.org/10.37817/ikraith-informatika.v8i1.3211

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Setyaningtyas, S., Nugroho, B. I., & Arif, Z. (2022). TINJAUAN PUSTAKA SISTEMATIS PADA DATA MINING: STUDI KASUS ALGORITMA K-MEANS CLUSTERING. *J Teknoif Teknik Informatika*, *10*, 52-61. DOI 10.21063/jtif.2022.V10.2.52-61

Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, *110*, 102817. https://doi.org/10.1016/j.ssresearch.2022.102817

Soepriyono, G., & Triayudi, A. (2023). Perbandingan Kinerja Algoritma Clustering Data Mining Untuk Prediksi Harga Saham Pada Reksadana dengan Davies Bouldin Index. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *7*(4), 2061-2073. DOI : http://dx.doi.org/10.30865 /mib.v7i4.6623

Sun, J. (2021). Forecasting COVID-19 pandemic in Alberta, Canada using modified ARIMA models. *Computer Methods and Programs in Biomedicine Update*, *1*, 100029. https://doi.org/10.1016/j.cmpbup.2021.100029

Suparmadi, S., & Ramadhani, A. (2022). SISTEM ESTIMASI PENCAPAIAN TARGET PROFIT MENGGUNAKAN MODEL REGRESI BERBASIS MACHINE LEARNING. *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, *5*(3), 703-708.DOI: https://doi.org/10.54314/ jssr.v5i3.1042

Suryani, A. M., & Noviari, N. (2023). Reaksi Pasar Modal terhadap Pengumuman Kebijakan Insentif Pajak Penjualan atas Barang Mewah. *E-Jurnal Akuntansi*, *33*(2), 302. http://dx.doi.org/10.24843/EJA.2023.v33.i02.p02

Yutanesy, J., & Suhendah, R. (2022). Perubahan Harga, Volume Saham, dan Kapitalisasi Pasar Selama COVID-19 pada Sektor Keuangan. *Jurnal Ekonomi*, *27*(03), 159–181. https://doi.org/10.24912/je.v27i03.871

Utami, D., Mutmainah, K., & Jannati, N. B. (2023). Analisis Penilaian Harga Wajar Saham Dan Keputusan Investasi Saham Secara Fundamental Dengan Menggunakan

Metode Price Earning Ratio (Per) Dan Price To Book Value. *Journal Of Economic, Business And Engineering (Jebe)*, *5*(1), 167-184. Doi : https://doi.org/10.32500/jebe.v5i1.5646

Yutanesy, J., & Suhendah, R. (2022). Perubahan Harga, Volume Saham, dan Kapitalisasi Pasar Selama COVID-19 pada Sektor Keuangan. *Jurnal Ekonomi*, *27*(03), 159-181. Doi : https://doi.org/10.24912 /je.v27i03.871