PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 04 No. 1, April 2025
Page 81-94

APPLICATION OF THE ARIMA MODEL IN FORECASTING ETHEREUM PRICES

Romario Desouza Daniel Mangiwa[1*], Revina Siregar[2] , Seli Delima Sari[3], Neli Agustina[4]

[1,2,3,4] Politeknik Statistika STIS
Otto Iskandardinata No.64C 1 St, RT.1/RW.4, Bidara Cina, 13330,
Jakarta, Indonesia

E-mail Correspondence Author: *romariomangiwa@gmail.com*

Abstract
Ethereum is one of the leading cryptocurrencies utilizing blockchain technology for peer-to-peer financial transactions. This study aims to forecast Ethereum's price using the Autoregressive Integrated Moving Average (ARIMA) model. Historical price data from January 1, 2023, to January 15, 2025, covering 534 periods, was analyzed. The ARIMA (0,1,9) model was selected based on AIC, SC, and Adjusted R-squared criteria, with forecast evaluation showing a Mean Absolute Percentage Error (MAPE) of 15.01% and a Root Mean Squared Error (RMSE) of 649.702. Forecast results indicate an upward trend in Ethereum's price over the next 30 periods, with fluctuations being less pronounced compared to historical data. The study concludes that ARIMA provides reasonably accurate short-term predictions, although forecasting errors increase with longer prediction periods. These findings can serve as a reference for investors in developing short-term investment strategies for Ethereum.

Keywords: *Arima, Cryptocurrency, Ethereum, Forecast, Investment*

: https://doi.org/10.30598/parameterv4i1pp81-94

Research Articiel • ***Open Acces***
Email: jurnalparameter@gmail.com
Homepage: https://ojs3.unpatti.ac.id/index.php/parameter
Submitted: Januari 2025
Accepted: April 2025

1. INTRODUCTION

Cryptocurrency is one of the fastest-growing forms of investment in the last decade [1]. Ethereum is one of the largest cryptocurrencies besides Bitcoin [2]. Like Bitcoin, investing in Ethereum requires careful planning, including understanding price volatility, market trends, and forecasting methods [2]. Crypto investments like Ethereum offer high profit potential. However, the accompanying price volatility poses substantial risks that demand effective risk management strategies. As a result, accurate price prediction becomes essential for mitigating potential losses and making informed decisions.

Crypto assets have three key characteristics: they are decentralized, unregulated, and anonymous [3]. These unique characteristics have contributed to the growing interest of Indonesians in investing, as evidenced by the increasing number of crypto investors. According to data from the Commodity Futures Trading Regulatory Agency (Badan Pengawas Perdagangan Berjangka Komoditi—Bappebti), as of August 2024, there were 20.9 million registered crypto investors in Indonesia. The increase of 500,000 investors in just one month indicates a rising public interest in cryptocurrency investments. Younger generations, including Millennials and Gen Z, are also becoming more involved in crypto investments, highlighting the widespread appeal of these assets.

Ethereum is one of the largest cryptocurrencies, created by Vitalik Buterin in 2013 as a platform for executing smart contracts and decentralized applications [1]. Ethereum has grown increasingly popular in recent years, leading more people to invest in this crypto asset. Various algorithms for predicting and forecasting crypto asset prices have been developed to support crypto investments in the digital world [4].

Several forecasting methods are used to predict cryptocurrency prices, such as Triple Exponential Smoothing (TES) and ARIMA. ARIMA is effective for non-stationary data [5] and is used to predict stock and cryptocurrency prices, though its accuracy in the highly volatile crypto market requires further research. [6] found that ARIMA outperformed SARIMAX in predicting Bitcoin prices, with a Mean Squared Error (MSE) of 54,791. However, [7] noted that ARIMA is only accurate for short-term predictions as it struggles to capture sharp price fluctuations. Conversely, [8] found that LSTM was more accurate than ARIMA for crypto price prediction. [9] concluded that ARIMA outperformed deep learning models, achieving a Mean Absolute Percentage Error (MAPE) of 2.76%. [10]compared four univariate models and found that ARMA was suitable for certain crypto assets, emphasizing the importance of selecting forecasting models based on cryptocurrency characteristics. [11] stated that K-Nearest Neighbors (KNN) performed better than ARMA and GARCH in predicting cryptocurrency liquidity, whereas GARCH was more effective in emerging markets. [12] proposed a hybrid LSTM-VAR model, which was found to be superior to ARIMA due to its lower evaluation metrics, indicating the high potential of combining time series and deep learning methods for cryptocurrency price prediction.

These various studies indicate that the best model for forecasting crypto asset prices may vary depending on the study period and dataset used. This aligns with the findings of [13], who emphasized that the choice of data period significantly impacts model performance in cryptocurrency forecasting, with some models being more effective for short-term predictions and others for long-term forecasts. ARIMA, which has been proven effective in analyzing non-stationary data, can be utilized to capture Ethereum's price patterns over specific periods. ARIMA-based research using recent data has the potential to provide deep insights into Ethereum market behavior while serving as a foundation for developing more targeted investment strategies.

Therefore, this study aims to use the ARIMA method with the best model to predict Ethereum's future value using the most recent data. This is done to provide research findings that can serve as a reference for investors before making Ethereum investment decisions. The study includes an evaluation of forecasting results using Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). The findings are expected to assist crypto investors in making informed decisions when investing in Ethereum.

2. METHODOLOGY

The data used in this study is secondary data. The analyzed data includes Ethereum (ETH) prices over 534 periods, spanning from January 1, 2023, to January 15, 2025, sourced from https://id.investing.com/crypto/ethereum/historical-data. To ensure model reliability, the dataset is divided into two categories: training data and testing data, with proportions of 80% and 20%, respectively. Thus, the training data consists of 428 periods, while the remaining 106 periods are reserved for performance testing. The training data is used to build the model, whereas the testing data is used to evaluate the model's performance. This study follows several key stages: data collection, model identification, residual diagnostics, and forecasting. Data analysis is conducted using the E-Views software.

2.1. Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a forecasting method that utilizes historical data for prediction. This method is widely used in financial forecasting because it often produces the best estimation models. ARIMA does not incorporate other independent variables; instead, it relies solely on current and past values to predict future values. The ARIMA method is highly accurate for short-term forecasting; however, its accuracy tends to decline in long-term forecasts as predictions often flatten or remain constant over extended periods. It is known to perform well for short-term forecasts, although its predictive power tends to diminish for longer horizons due to the tendency of forecasts to converge or flatten

ARIMA works by leveraging historical and current data from a single time series variable (univariate) to generate accurate forecasts. This method is particularly effective when observations in a time series are statistically dependent. The ARIMA model combines two key approaches: the Autoregressive (AR) model, which explains variable patterns using past data of the variable itself, and the Moving Average (MA) model, which analyzes variable movements based on residual values from previous periods [14].

2.2. Autoregressive (AR) Model

The Autoregressive (AR) model describes the forecasting process as a relationship between the current value and past values within a time series [15]. An autoregressive model of order p (AR(p)), also known as an ARIMA (p,0,0) model, has the general form as follows:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \quad (1)$$

Or expressed in backshift notation as follows:

$$\phi_p(B) Y_t = \varepsilon_t \quad (2)$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

Description:

Y_t : Value of the Variable at Period t

Y_{t-i} : Value of the Variable in the Previous Period at Time t-i, where $i = 1,2, \ldots, p$

$\phi_1, \dots, \phi_p$: Estimated AR Coefficients
ε_t : Error Value at period t
B^p : Backshift Operator, where $B^p Y_t = Y_{t-p}$

2.3. Moving Average (MA) Model

The Moving Average (MA) represents the value of a time series at a specific time, which is influenced by the current error as well as the weighted errors from previous periods **[15]**. A moving average model of order q (MA(q)) or ARIMA(0,0,q) can generally be expressed as follows:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (3)$$

Or expressed in backshift notation as follows:

$$Y_t = \theta_q(B)\varepsilon_t \quad (4)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

Keterangan:
Y_t : Value of the Variable at Period t
$\theta_1, \dots, \theta_q$: Estimated MA Coefficients
$\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-p}$: Error Values at Periods t to t-q
B^q : Backshift Operator, $B^q Y_t = Y_{t-p}$

2.4. Mixed Model

ARMA Model

The autoregressive moving average model, denoted as ARMA(p,q), is a combination of the autoregressive and moving average models without involving differentiation. In general, the ARMA model can be written as follows:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (5)$$

Or expressed in backshift notation as follows:

$$\phi_p(B) Y_t = \theta_q(B)\varepsilon_t \quad (6)$$

ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is an extension of the ARMA(p,q) model designed to analyze non-stationary data. In this model, non-stationary data is transformed into stationary data through differentiation d times. Once differentiation is completed, the ARMA(p,q) model is converted into an ARIMA(p,d,q) model. The mathematical form of the ARIMA model can be described as follows:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\varepsilon_t \quad (7)$$

2.5. Forecasting Steps with the ARIMA Method

1. Model Identification

   At this stage, the stationarity test of the data is conducted by analyzing the patterns in the time series data. The stationarity identification process can be carried out by observing data patterns using plots, ACF graphs **[16]**, or by applying the Augmented Dickey-Fuller (ADF) test. The hypotheses and test statistics used in the ADF test are as follows.

Hypothesis
H_0: $\delta = 0$ (there is a unit root or data is not stationary)
H_1: $\delta \neq 0$ (no unit root or data is stationary)

Test Statistic

$$\tau = \frac{\hat{\delta}}{se(\hat{\delta})} \sim \tau_n \tag{8}$$

Keterangan:
$\hat{\delta}$: Coefficient of Y_{t-1}
$se(\hat{\delta})$: Standard Error of the Coefficient Y_{t-1}

Based on the test statistic value, the research hypothesis leads to the rejection of H_0 when $|\tau| > |\tau_{\alpha,n}|$ or *p-value* $\leq \alpha$.

If the data is not yet stationary, differentiation must be applied to the original data (Yt) to make it stationary. Differentiation is performed by calculating the difference between the data value at a given period and the value from the previous period. The differentiation notation using the backshift operator is as follows:

$$(1 - B)^d Y_t \tag{9}$$

where d is the number of differentiations

After the data is declared stationary, a correlogram analysis is conducted to determine the order values of p, d, and q. These three orders will be used in the model-building process. The order p is obtained from the PACF data pattern, while the order q is taken based on the ACF data pattern. This step is essential because ARIMA requires stationarity for reliable modeling.

2. Parameter Estimation

After determining the orders for AR and MA, the next step is to estimate the AR and MA parameters to identify significant parameters. Parameter estimation is performed on all temporary ARIMA models to ensure whether the generated parameters are significant or not. A model is considered significant if the parameter significance value is less than alpha (α), with α set at 0.05. If more than one model meets the significance criteria, model selection is guided by three commonly used metrics that assess both model fit and complexity. Model selection is based on the values of Akaike's Information Criterion (AIC) and Schwarz Criterion (SC), as well as the adjusted R-Square value, which serve to identify the best model from the dataset. Akaike's Information Criterion (AIC) and Schwarz Criterion (SC) evaluate the trade-off between goodness of fit and the number of parameters, helping to avoid overfitting by penalizing overly complex models **[15], [17]**. Adjusted R-squared, meanwhile, measures how well the model explains variation in the dependent variable while accounting for the number of predictors, with higher values indicating better explanatory power **[17]**. These metrics serve as standard tools in model selection by quantifying both the fit of the model and its complexity, making them especially useful when comparing competing models.

Hypothesis
H_0: $\theta_i = 0$ (parameter is not significant)
H_1: $\theta_i \neq 0$ (parameter is significant)

Test Statistic

$$t_{statistic} = \frac{\hat{\theta}_l}{se(\hat{\theta}_l)} \quad \sim t_{(\alpha,df)} \tag{10}$$

Based on the test statistic value, the research hypothesis results in the decision to reject H_0 when $|t_{statistic}| > t_{\frac{\alpha}{2},n-p}$ or *p-value* $< \alpha$.

3. Residual Diagnostic Checking

Residual diagnostic checking is conducted to ensure that the model used is adequate and suitable for forecasting purposes. In this study, diagnostics include normality tests, white noise tests, and homoscedasticity tests.

a. Normality Test

The optimal ARIMA model is one whose residuals follow a normal distribution. Therefore, a normality test of the residuals is required to confirm that the residuals conform to a normal distribution. This test can be conducted using the Kolmogorov-Smirnov test, with the hypothesis and test statistics explained as follows.

Hypothesis

H_0: $Residuals\ follow\ a\ normal\ distribution$

H_1: $Residuals\ do\ not\ follow\ a\ normal\ distribution$

Test Statistic

$$D = Max\ |F(x) - S(x)| \tag{11}$$

Description:

$F(x)$: Cumulative distribution function of the sample

$S(x)$: Proportion of sample observations less than or equal to x

Based on the test statistic value, the research hypothesis results in the decision to reject H_0 when $D > D_{table}$ atau *p-value* $< \alpha$

b. White Noise Non-Autocorrelation Residual Test

A good ARIMA model has residuals that are white noise, meaning they are random, patternless, and free from autocorrelation. In other words, the residuals must be independent, non-autocorrelated, and have homogeneous variance. The white noise assumption can be tested using the Ljung-Box test with the following hypothesis and test statistics.

Hypothesis

H_0: $\rho_1 = \rho_2 = \cdots = \rho_k = 0$ (residuals are not autocorrelated/white noise assumption is satisfied)

H_1: at least one $\rho_k \neq 0$ (residuals are autocorrelated/white noise assumption is not satisfied)

Test Statistic

$$Q = n(n+2)\sum_{k=1}^{K} \frac{\hat{p}_k^{\,2}}{(n-k)} \sim \chi^2_{(1-\alpha;K)} \tag{12}$$

Description:

n : Number of Observations

k : Time Lag

${\hat{p}_k}^2$: Sample Autocorrelation Function of the K-th Lag Residual

Based on the test statistic value, the research hypothesis results in the decision to reject H_0 when $Q > \chi^2_{(1-\alpha;K)}$ or *p-value* $< \alpha$.

c. Homoscedasticity Test

The homoscedasticity test aims to identify the variance pattern. Variance can be constant (homoscedasticity) or variable (heteroscedasticity). The homoscedasticity test is conducted using the Breusch-Pagan test with the following hypothesis and test statistics.

Hypothesis

H_0: $var(e_i) = \sigma^2$ (homoscedasticity is satisfied)

H_1: at least one $var(e_i) \neq \sigma^2$ (homoscedasticity is not satisfied/heteroscedasticity is present)

Test Statistic

$$BP = \frac{1}{2}\,(ESS) \tag{13}$$

Description:

ESS : Explained Sum of Squares with $ESS = \sum(\hat{y}_t - y_t)^2$

Based on the test statistic value, the research hypothesis results in the decision to reject H_0 when $BP > \chi^2_{(\alpha;m)}$ or *p-value* $< \alpha$.

4. Forecasting

The final stage is to perform data forecasting. After all tests and assumptions have been satisfied, the best model obtained will be used for forecasting. If the selected best ARIMA model exhibits violations of the homoscedasticity assumption in the residuals, the ARCH/GARCH approach can be applied to address this issue. However, if the homoscedasticity assumption is not violated, forecasting can be directly carried out using the ARIMA method based on the best-selected model.

2.6. Best Model Selection

In practical terms, AIC and SC are used to compare competing ARIMA models by penalizing models that are too complex. A lower AIC or SC value suggests a better-fitting model with fewer unnecessary parameters. Meanwhile, the Adjusted R-squared reflects how well the model explains variations in the data. A higher Adjusted R-squared indicates better explanatory power. When multiple models are statistically significant, the combination of low AIC/SC values and a high Adjusted R-squared is used to determine the most optimal model for forecasting **[15]**. According to **[17]**, a model with lower AIC and SC values typically demonstrates better predictive performance and greater suitability for the underlying data pattern. These criteria help ensure the chosen model not only fits the data well but also avoids overfitting, maintaining a balance between complexity and forecasting reliability.

Akaike's Information Criterion (AIC) Formula :

$$AIC = \ln\left(\frac{RSS}{n}\right) + \frac{2}{n}m \tag{14}$$

Description:

RSS : Residual Sum of Squares

m : Number of Parameters

n : Number of Observations

Schwarz Criterion (SC) Formula:

$$SC = \ln\left(\frac{RSS}{n}\right) + \frac{\ln(n)}{n}m \tag{15}$$

Description:

RSS : Residual Sum of Squares

m : Number of Parameters

n : Number of Observations

In addition, the best model can also be determined by a higher coefficient of determination (R-Square). [18] explains that the coefficient of determination represents the contribution of independent variables to the variation of the dependent variable or measures the percentage of total variation in the dependent variable explained by the regression model. The coefficient of determination is a non-negative value ranging from 0 to 1. If the value approaches 0, the model's ability to explain the dependent variable is very weak. Conversely, if the value approaches 1, the model becomes more effective in describing the dependent variable.

R-Square Formula

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{16}$$

$$R^2_{adjusted} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} \tag{17}$$

Description:

R^2 : Coefficient of Determination

SSR : Sum of Square Residual

SSE : Sum of Square Error

SST : Sum of Square Total

n : Number of Observations

k : Number of Independent Variables

2.7. Forecasting Accuracy

The optimal forecasting method is one with a relatively low error rate. According to [15], forecasting accuracy can be evaluated using several measures, such as Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). In this study, forecasting accuracy is measured using two metrics: MAPE and RMSE.

1. Mean Absolute Percentage Error (MAPE)

MAPE (Mean Absolute Percentage Error) is one of the measures used to evaluate the accuracy of forecasting models. MAPE indicates how large the forecasting error is compared to the actual value in percentage form.

Mean Absolute Percentage Error (MAPE):

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right|\right) \times 100 \quad (18)$$

Description:

Y_t : Actual Data at period t

$\hat{Y}_t$: Forecasted Data at period t

n : Number of Data Points

A smaller percentage of error indicates higher forecasting accuracy. Accuracy level can be calculated using the formula: accuracy level = 100% - MAPE.

Below are the MAPE value criteria [19]:

Tablel 1. MAPE Value Criteria

MAPE Value	Interpretation
<10%	Very Good Forecasting Model
10-20%	Good Forecasting Model
21-50%	Fair Forecasting Model
>50%	Poor Forecasting Model

These benchmarks provide an essential reference for interpreting forecast accuracy across studies and facilitate consistent evaluation, especially when comparing across different models and datasets.

2. Root Mean Square Error (RMSE)

RMSE (Root Mean Square Error) is a measurement method used to evaluate the level of error in forecasting models. RMSE calculates the square root of the average sum of squared differences between actual values and predicted values, making it more sensitive to large errors [15]. The goal is to determine how well the model predicts actual data, with a lower RMSE indicating a lower level of error [20].

Root Mean Square Error (RMSE) Formula :

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2} \quad (19)$$

Keterangan:

Y_t : Actual Data at period t

$\hat{Y}_t$: Forecasted Data at period t

n : Number of Data Points

3. RESULTS AND DISCUSSION

3.1 Data Patterns and Stationarity Test

The initial step in the forecasting process is identifying data patterns and testing data stationarity, both visually and through statistical tests. Data used in forecasting with the ARIMA method must be stationary to ensure homogeneous results and meet the assumption of no autocorrelation.

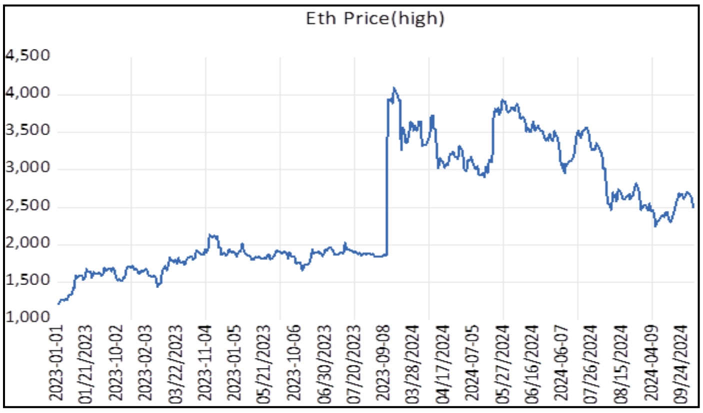


Figure 1. **Stationarity Test with graphic method**

The output in **Figure 1** shows that Ethereum's price exhibits both upward and downward trends and does not center around a mean value, indicating that the Ethereum data for the observed period is non-stationary. In March 2024, Ethereum's price experienced a significant surge, driven by a network upgrade known as **Dencun**, aimed at increasing throughput and reducing transaction costs. This upgrade was intended to enhance Ethereum's efficiency and competitiveness compared to other blockchains **[21]**. The stationarity test can be conducted using a statistical test, specifically the unit root test.

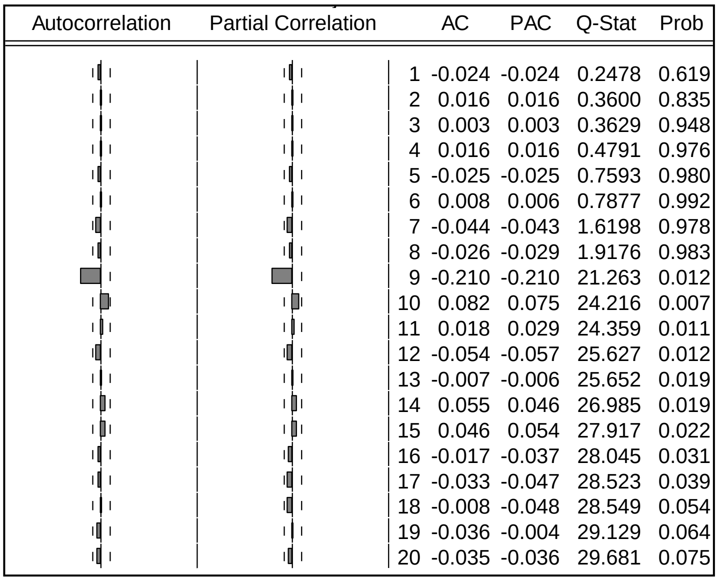
Table 2. **Summary of Stationarity Test at Level and 1st Differencing**

Variable	Level	1^{st} *Differencing*
t-Statistic	-2.0970	-21.0661
Prob. Value	0.2461	0.0000
Critical Value 5%	-2.8681	-2.8681
Conclusion	Non-Stationary	Stationary

The stationarity test conducted at the level, as shown in **Table 2**, indicates that the Augmented Dickey-Fuller (ADF) statistic for the Ethereum variable is -2.097 with a p-value of 0.2461. Since the p-value exceeds 0.05, the data is not stationary, which violates a key assumption for ARIMA modeling. To address this, first-order differencing is applied to remove trends and stabilize the mean over time. After differencing, the ADF statistic becomes -21.066 with a p-value of 0.000, indicating the data is now stationary and suitable for ARIMA analysis. Thus, all subsequent modeling uses the first-differenced data. Visual trends in **Figure 1** are supported by statistical confirmation in **Table 2**, reinforcing the appropriateness of differencing for ARIMA preprocessing. This dual-check strengthens the reliability of the transformation stage.

3.2 Tentative Model Identification

The identification of a possible ARIMA(p,d,q) model is conducted by analyzing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. Based on the ACF and PACF plots, potential models can be determined based on the lags observed in the correlogram



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.024	-0.024	0.2478	0.619
		2	0.016	0.016	0.3600	0.835
		3	0.003	0.003	0.3629	0.948
		4	0.016	0.016	0.4791	0.976
		5	-0.025	-0.025	0.7593	0.980
		6	0.008	0.006	0.7877	0.992
		7	-0.044	-0.043	1.6198	0.978
		8	-0.026	-0.029	1.9176	0.983
		9	-0.210	-0.210	21.263	0.012
		10	0.082	0.075	24.216	0.007
		11	0.018	0.029	24.359	0.011
		12	-0.054	-0.057	25.627	0.012
		13	-0.007	-0.006	25.652	0.019
		14	0.055	0.046	26.985	0.019
		15	0.046	0.054	27.917	0.022
		16	-0.017	-0.037	28.045	0.031
		17	-0.033	-0.047	28.523	0.039
		18	-0.008	-0.048	28.549	0.054
		19	-0.036	-0.004	29.129	0.064
		20	-0.035	-0.036	29.681	0.075

Figure 2. ACF and PACF plot

The selection of p and q values is conducted by identifying lags that exceed the Bartlett line. In **Figure 2**, it is observed that Lag 9 surpasses the Bartlett line. Thus, the AR(p) value is 9 and the MA(q) value is 9. Since the ACF and PACF are stationary at the 1st differencing, the d value for the model is 1. Therefore, the potential models that can be formed are ARIMA (9,1,9), ARIMA (9,1,0), and ARIMA (0,1,9). The PACF spike at Lag 9 provides a strong signal for autoregressive order, while the ACF behavior justifies the moving average term. These diagnostics guide the structural configuration of the ARIMA model.

3.3 ARIMA Model Evaluation

To determine the best model, a comparison is made using Akaike's Information Criterion (AIC), Schwarz Criterion (SC), and Adjusted R-squared. A lower AIC and SC value indicate a more potential model, while a higher Adjusted R-squared value signifies a better fit.

Table 3. Comparison of AIC, SC, and Adjusted R-Squared Values

Lag	ARIMA(9,1,9)	ARIMA(0,1,9)	ARIMA(9,1,0)
AR(9)	0.0087	-	-0.2068*
MA(9)	0.2263	-0.2180*	-
AIC	12,570	12,565	12,567
SC	12,608	12,593	12,596
Adj R-squared	0,040	0,042	0,039

* Significance at $\alpha = 5\%$

Based on **Table 3**, the ARIMA (0,1,9) model has the lowest AIC value, while the lowest SC value is found in the ARIMA (9,1,9) model. The highest Adjusted R-squared value is also in the ARIMA (0,1,9) model. Given these results, ARIMA (0,1,9) is selected as the best model because it satisfies two out of the three evaluation criteria, offering an optimal trade-off between simplicity and predictive accuracy.

The performance evaluation reflects model parsimony and goodness-of-fit. By achieving the lowest AIC and highest Adjusted R-squared, ARIMA (0,1,9) offers the best trade-off between complexity and predictive reliability.

3.4 Residual Diagnostic Testing

Tablel 4. ARIMA Model Residual Tests

Test	Decision
Normality	Fulfilled
White Noise	Fulfilled
Homoscedasticity	Fulfilled

Table 4 shows that all residual assumptions are fulfilled, including normality, white noise, and homoscedasticity. These results indicate that the ARIMA (0,1,9) model is statistically adequate and suitable for forecasting. Meeting these diagnostic criteria confirms the model's validity and supports its use for short-term prediction, as residuals free from autocorrelation and heteroscedasticity reduce the risk of biased or inefficient forecasts. These diagnostic tests ensure that the residuals are random, normally distributed, and exhibit constant variance, fulfilling the classical assumptions that underpin reliable statistical inference in time series forecasting.

3.5 Model Verification and Forecasting Results

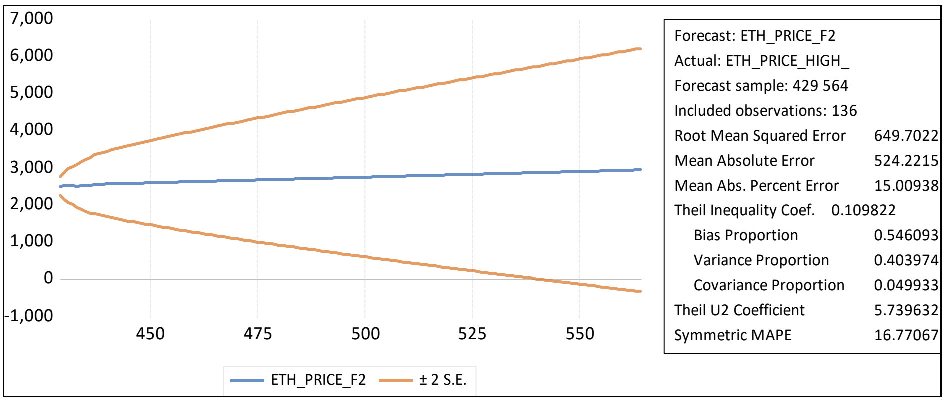


Figure 3. Ethereum Forecasting Results Using the ARIMA (0,1,9) Model

Figure 3 displays the forecasted Ethereum prices using the ARIMA (0,1,9) model, along with ±2 standard error bands, which represent the 95% confidence interval around the forecast. The blue line shows the predicted values, while the orange lines depict the upper and lower confidence bounds. The boxed panel on the right summarizes key forecast accuracy metrics such as RMSE, MAE, MAPE, and Theil coefficients. After selecting the best model and confirming that all model assumptions are met, the next step is to perform forecasting and assess its accuracy using MAPE and RMSE. As shown in Figure 3, the Ethereum price forecast using the ARIMA (0,1,9) model yields a MAPE of 15.01% and an RMSE of 649.702. Based on the MAPE classification in Table 1, as proposed by [19], the model falls within the "Good Forecasting Model" category. Given the high volatility of cryptocurrency markets, a MAPE in the 10%–20% range is considered reasonable for short-term forecasts. For context, [9] reported a MAPE of 2.76% using ARIMA for Bitcoin, while [14] obtained a MAPE of 44.8% for Ethereum using ARIMA (1,1,0), highlighting the variability in model performance across datasets and timeframes. In addition, [7] noted that ARIMA models typically yield lower forecasting errors in short-term horizons, but their accuracy declines as the forecast window extends. This is also reflected in the present study, where the standard error bands grow wider over the 30-period forecast, indicating increased uncertainty. Consequently, the ARIMA (0,1,9) model is considered suitable for short-term Ethereum forecasting, while more robust or hybrid approaches may be required for long-term prediction.

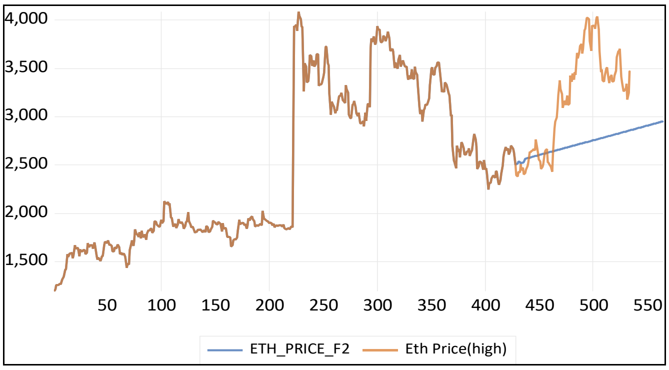


Figure 4. Comparison of Actual and Forecasted Ethereum Values.

Figure 4 displays a visual comparison between the actual Ethereum high prices and the forecasted values produced by the ARIMA (0,1,9) model. The solid brown line represents actual market data, while the dashed blue line shows the model's forecast for the next 30 periods. As shown in **Figure 4**, the forecast for Ethereum's value over the next 30 periods indicates an upward trend. Some of the forecasted values overlap with the actual data, suggesting that the model's predictions align well with observed values. The ARIMA (0,1,9) model estimates that Ethereum's value will rise over the next 30 periods, with a minimum forecasted value of $2,861.46 and a maximum of $2,949.68. However, the fluctuations in the forecast are not as pronounced as in historical data. If the forecasting period is extended further, the predicted Ethereum values will continue to rise and deviate more from actual data, making long-term forecasts increasingly inaccurate.

4. CONCLUSION

The ARIMA (0,1,9) model provides reasonable short-term predictions for Ethereum's price, with a MAPE value indicating good accuracy. However, its performance diminishes over longer prediction periods, highlighting the challenge of forecasting in highly volatile markets like cryptocurrency. This underscores the model's suitability for short-term investment strategies. Future research could explore hybrid models or integrate external variables to improve long-term forecasting accuracy. This study contributes to the application of ARIMA in cryptocurrency forecasting and offers valuable insights for investors making short-term decisions.

REFERENCES

[1] M. G. Solomon, *Ethereum For Dummies*. United States: Wiley, 2019.

[2] I. Santo, *Melangkah ke Dunia Cryptocurrency: Memulai Perjalanan Anda ke Investasi Digital di Indonesia*. Indonesia: Bintang Semesta Media, 2023.

[3] Meliza and D. Hastalona, *Investasi di Pasar Cryptocurrency*. Indonesia: Penerbit P4I, 2022.

[4] A. Singh, P. Kumar, and S. Taneja, *Algorithmic Approaches to Financial Technology: Forecasting, Trading, and Optimization*. IGI Global, 2023.

[5] F. Rojas, H. Pomares, I. Rojas, and O. Valenzuela, *Theory and Applications of Time Series Analysis: Selected Contributions from ITISE 2018*. Springer International Publishing, 2019.

[6] A. Aryanusa and S. Zahara, "Analisis Prediksi Harga Bitcoin Menggunakan Metode ARIMA," *SUBMIT: Jurnal Ilmiah Teknologi Infomasi dan Sains*, vol. 4, no. 1, pp. 15–18, Jun. 2024, doi: 10.36815/submit.v4i1.3339.

[7] A. Azari, "Bitcoin Price Prediction: An ARIMA Approach," Apr. 2019.

[8] S. M. Raju and A. M. Tarif, "Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis," Jun. 2020.

[9] A. Listiani and D. Al Mahkya, "Peramalan Cryptocurrency dengan Autoregressive Integrated Moving Average (ARIMA) dan Risiko Kerugian dengan Value at Risk (VaR)," *Journal of Science and Applicative Technology*, vol. 6, no. 2, pp. 85–91, 2022.

[10] M. Jayasankar, "Comparative Forecasting of Major Cryptocurrencies: An Empirical Study Using Four Timeseries Forecasting Models," in *Proc. Int. Conf. Financial Markets & Corporate Finance*, Singapore: Springer Nature Singapura, Jul. 2023, pp. 41–57.

[11] K. Cortez, M. D. P. Rodríguez-García, and S. Mongrut, "Exchange market liquidity prediction with the K-nearest neighbor approach: Crypto vs. fiat currencies," *Mathematics*, vol. 9, no. 1, p. 56, 2020.

[12] P. Sharma and R. M. Pramila, "Price prediction of Ethereum using time series and deep learning techniques," in *Proc. Emerging Trends and Technologies on Intelligent Systems: ETTIS 2022*, Singapore: Springer Nature SIngapore, 2022, pp. 401–413.

[13] D. L. John, S. Binnewies, and B. Stantic, "Cryptocurrency Price Prediction Algorithms: A Survey and Future Directions," *Forecasting*, vol. 6, no. 3, pp. 637–671, Aug. 2024, doi: 10.3390/forecast6030034.

[14] N. Liunardo, D. Gunawan, and N. Murniati, "Analisis Peramalan Nilai Ethereum Menggunakan Model Autoregressive Integrated Moving Average," *Accounting Progress*, vol. 2, no. 2, pp. 49–61, Jan. 2024, doi: 10.70021/ap.v2i2.99.

[15] S. G. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: Methods and Applications*, 2nd ed. New York, NY: Wiley, 1997.

[16] J. E. Hanke and D. W. Winchern, *Business Forecasting*, 8th ed. Englewood, NJ: Cliffs Prentice Hall, 2005.

[17] J. D. Cryer and K.-S. Chan, *Time Series Analysis with Applications in R*, 2nd ed. Berlin, Germany: Springer, 2008.

[18] J. Supranto, *Statistik Teori dan Aplikasi*, 7th ed. Jakarta, Indonesia: Penerbit Airlangga, 2008.

[19] P.-C. Chang, Y.-W. Wang, and C.-H. Liu, "The development of a weighted evolving fuzzy neural network for PCB sales forecasting," *Expert Syst Appl*, vol. 32, no. 1, pp. 86–96, Jan. 2007, doi: 10.1016/j.eswa.2005.11.021.

[20] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. Hoboken, NJ: Wiley, 2008.

[21] I. Apriansyah, "Penyebab Harga Ethereum (ETH) Naik: Didorong Upgrade Dencun dan ETF," Bittime.