

**EVALUATING ARIMA AND LSTM MODELS FOR ETHEREUM (ETH) PRICE IN
FORECASTING**

By

Alif Satria Adhi

12202026

BACHELOR'S DEGREE

in

Information Technology

Faculty of Engineering and Information Technology



SWISS GERMAN UNIVERSITY

The Prominence Tower

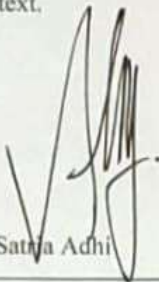
Jalan Jalur Sutera Barat No. 15, Alam Sutera

Tangerang, Banten 15143 - Indonesia

June 2026

STATEMENT FROM AUTHOR

I hereby declare that my thesis is entirely my own work. To the best of my knowledge and belief, it contains no material previously published or written by any other person, nor any material that has been accepted for the award of any other degree or diploma at any other educational institution, except where proper acknowledgement is made in the text.



Alif Satria Adhi

17 June 2026

Student

Date



Dr. Ir. Eka Budiarto ST, M.Sc.

17 June 2026

Advisor

Date

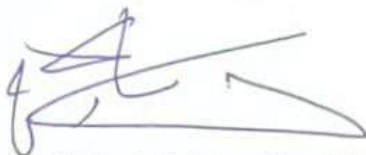


Dr. Maulahikmah Galinium S.kom M.Sc.

17 June 2026

Co-Advisor

Date



Dr. Maulahikmah Galinium S.kom M.Sc.

17 June 2026

Dean

Date

ABSTRACT

EVALUATING ARIMA AND LSTM MODELS FOR ETHEREUM (ETH) PRICE IN FORECASTING

By

Alif Satria Adhi

Dr. Ir. Eka Budiarto ST, M.Sc.

Dr. Maulahikmah Galinium S.kom M.Sc.

SWISS GERMAN UNIVERSITY

This research evaluates and compares forecasting models for daily Ethereum (ETH) closing prices from 2019 to 2025. The purpose is to determine which model produces more accurate predictions under volatile market conditions. The methodology includes ARIMA (with grid search and walk-forward validation), SARIMA, Holt's exponential smoothing, univariate and multivariate LSTM (with feature selection using correlation, permutation importance, and SHAP), a naive persistent baseline, and Random Forest. Performance is measured using RMSE, MAE, and MAPE. The findings reveal that Random Forest achieves the highest one-day-ahead accuracy (MAPE 1.03%), outperforming all other models. ARIMA with daily walk-forward retraining attains a MAPE of 2.74%, while the naive baseline (2.81%) is surprisingly competitive due to the strong autocorrelation of ETH prices (≈ 0.9966). Univariate LSTM (2.93%) offers only marginal improvement over persistence. Multivariate LSTM features add negligible benefit, and seasonal models or double differencing do not improve accuracy. In conclusion, Random Forest is the best model for one-day-ahead Ethereum forecasting, followed by ARIMA with daily retraining. LSTM provides stable performance without retraining but with limited advantage. The naive baseline remains a strong benchmark.

Keywords: *ARIMA, Ethereum, LSTM, Random Forest, walk-forward validation*



ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to Allah SWT, the Most Gracious and Most Merciful, for granting me strength, health, and perseverance throughout my academic journey and the completion of this thesis. I am profoundly grateful to my beloved mother, whose endless love, prayers, sacrifices, and unwavering support have been my greatest source of strength. Her encouragement kept me going through every challenge, and this achievement is dedicated to her.

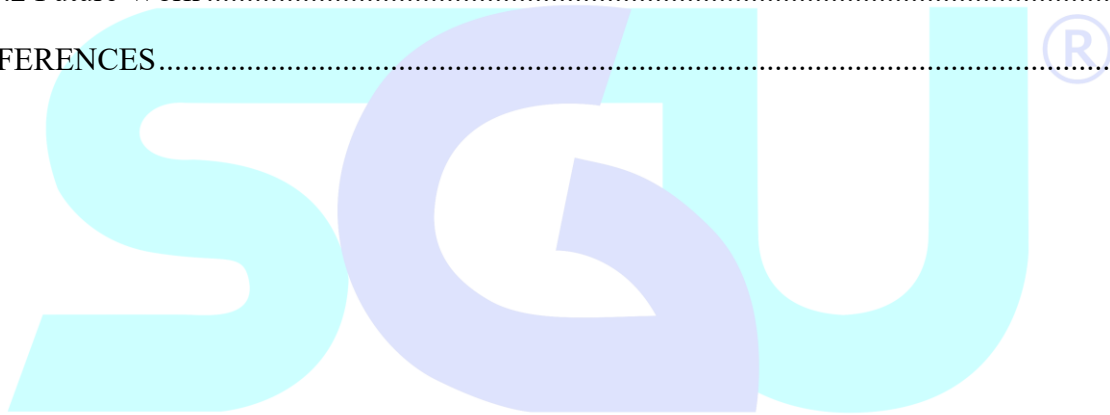
I would like to extend my sincere appreciation to my advisor, Dr. Ir. Eka Budiarto ST, M.Sc., and my co-advisor, Dr. Maulahikmah Galinium S.kom M.Sc., for their invaluable guidance, patience, and constructive feedback throughout this research. I also wish to thank the Indodax team for their insights and support, as well as my campus friends for their companionship, encouragement, and moral support during the ups and downs of this academic journey. May Allah SWT reward you all abundantly.

TABLE OF CONTENTS

	page
STATEMENT FROM AUTHOR	2
ABSTRACT	3
ACKNOWLEDGEMENT	5
TABLE OF CONTENTS	6
LIST OF TABLES	9
LIST OF FIGURES.....	10
CHAPTER 1 – INTRODUCTION	12
1.1 Background	12
1.2 Problem Statement	14
1.3 Research Objective.....	15
1.4 Research Question.....	15
1.5 Hypothesis.....	15
1.6 Scope	16
1.7 Limitation.....	16
1.8 Significance of the Study	16
CHAPTER 2 – LITERATURE REVIEW	17
2.1 Theoretical Perspective	17
2.2 Related Works	25
2.3 Research Gap & Novelty.....	29
CHAPTER 3 – METHODOLOGY.....	30
3.1 Research Overview	30
3.2 Data Collection.....	30
3.3 Data Preprocessing and Exploratory Data Analysis	31

3.3.1 Data Preprocessing	31
3.3.2 Selection of Forecasting Target Variable	32
3.3.3 Exploratory Data Analysis	33
3.4 Feature Preparation	33
3.5 Model Development (ARIMA & LSTM)	34
3.5.1 ARIMA	34
3.5.2 LSTM	35
3.6 Forecasting	36
3.7 Model Evaluation and Comparison	36
CHAPTER 4 - RESULTS	38
4.1 ARIMA	38
4.1.1 Data Pattern and Stationary Test	38
4.1.2 Model Identification (ACF and PACF)	39
4.1.3 ARIMA Model Selection	41
4.1.4 Residual Diagnostic Checking	42
4.1.5 Forecasting and Model Evaluation	43
4.1.6 ARIMA (3, 1, 2) Testing	45
4.1.7 ARIMA with Second-Order Differencing ($d=2$)	46
4.1.8 Walk-Forward Validation (Multi-Horizon Forecasting)	48
4.1.9 Walk-Forward Validation for ARIMA (4, 2, 5)	52
4.1.10 SARIMA	54
4.1.11 SARIMA with Second-Order Differencing ($d=2, D=2$)	57
4.1.12 Exponential Smoothing (Holt's Linear Trend)	59
4.2 LSTM	60
4.2.1 Data Preparation and Normalization	60
4.2.2 Sequence Preparation (Sliding Window)	61
4.2.3 LSTM Model Architecture	61

4.2.4 Model Training.....	63
4.2.5 Model Evaluation	65
4.2.6 LSTM Model Testing.....	67
4.2.7 Naïve Baseline.....	68
4.2.8 Multivariate LSTM	70
4.2.9 Random Forest	79
4.3 Comparison Analysis	82
4.4 Practical Insights	88
CHAPTER 5 – CONCLUSION.....	90
5.1 Conclusion.....	90
5.2 Future Work	92
REFERENCES.....	93



LIST OF TABLES

Table 4. 1 Differencing results.....	39
Table 4. 2 ARIMA (3,1,2) Model Evaluation	44
Table 4. 3 Top ten ARIMA models with d=2 by grid search.....	47
Table 4. 4 Walk-Forward Validation results for ARIMA (3, 1, 2)	48
Table 4. 5 Walk-forward validation results for ARIMA (4, 2, 5)	53
Table 4. 6 SEASONAL ARIMA (SARIMA) with d = 1 Model Selection Table	54
Table 4. 7 Top ten SARIMA models with d=2, D=2 (white-noise residual prioritized)	58
Table 4. 8 LSTM Model Architecture Description	62
Table 4. 9 Performance Comparison for Different Epoch Values	64
Table 4. 10 LSTM Model Evaluation	65
Table 4. 11 Feature Selection.....	74
Table 4. 12 Performance comparison of LSTM models	77
Table 4. 13 Feature Importance for Random Forest	80
Table 4. 14 Model comparison between all models.....	84

LIST OF FIGURES

Figure 4. 1 Ethereum (ETH) Prices.....	38
Figure 4. 2 ACF and PACF plots	39
Figure 4. 3 ACF and PACF Values with Ljung-Box Test	40
Figure 4. 4 AIC comparison.....	41
Figure 4. 5 Residual Diagnostic Plots of ARIMA(3,1,2) Model	42
Figure 4. 6 Forecasting Results of the ARIMA (3,1,2) Model for Ethereum Prices	43
Figure 4. 7 ARIMA 30-Day Future Forecast with unseen data	45
Figure 4. 8 ARIMA Forecast with actual data (30-Day Period)	45
Figure 4. 9 ARIMA (4, 2, 5) result.....	47
Figure 4. 10 Forecasting Results of the ARIMA Walk-Forward (3,1,2) Model for Ethereum Prices with 1 horizon.....	49
Figure 4. 11 ARIMA Walk-Forward Forecast with actual data (30-Day Period).....	52
Figure 4. 12 ARIMA Walk-Forward (4, 2, 5) result.....	53
Figure 4. 13 Forecasting Results of the SARIMA (2,1,2)x(1,1,2,7) Model for Ethereum Prices	55
Figure 4. 14 Testing SARIMA (2, 1, 2) x (1, 1, 2, 7) for 30-Day.....	57
Figure 4. 15 SARIMA (1, 2, 1)(0, 2, 2, 7) results	59
Figure 4. 16 Exponential Holt-Winter result.....	60
Figure 4. 17 Sliding Window Representation for LSTM Input Sequences	61
Figure 4. 18 LSTM Model Architecture	62
Figure 4. 19 Training and Validation Loss of the LSTM Model	64
Figure 4. 20 Forecasting Result for the LSTM model	66
Figure 4. 21 LSTM Forecasting 30-Days with unseen data.....	67
Figure 4. 22 LSTM Forecasting 30-Days with actual data	67
Figure 4. 23 Naive Baseline Graph	69
Figure 4. 24 Multivariate Heat-Map for Correlation.....	71
Figure 4. 25 LSTM Multivariate Permutation	72
Figure 4. 26 LSTM Multivariate SHAP	73
Figure 4. 27 Forecasting Result for Model A.....	75
Figure 4. 28 Forecasting Result for Model B.....	76
Figure 4. 29 Forecasting Result for Model C.....	76

Figure 4. 30 Forecasting Result for Model D.....	76
Figure 4. 31 Forecasting Result for Model E	77
Figure 4. 32 Random Forest forecast result	82

