ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

"From Data to Impact: Leveraging Innovation in Data Science and Official Statistics for Data-Driven Policy Making and Global Competitiveness"

The Impact of Training-Testing Proportion on Forecasting Accuracy: A Case of Agricultural Export in Indonesia

T W Septiarini[1,*], M D P Martinasari[1] and E Pariyanti[3]

[1] Mathematics Study Program, Faculty of Science and Technology, Universitas Terbuka, South Tangerang, Banten, 15437 Indonesia
[2] Department of Doctor in Management, Graduate School, Universitas Terbuka, South Tangerang, Banten, 15437 Indonesia

*Corresponding author's email: tri.wijayanti@ecampus.ut.ac.id

Abstract. Accurate forecasting of agricultural exports is crucial for supporting trade policy and ensuring economic stability in Indonesia. This study investigates the impact of training–testing proportions on the forecasting accuracy of six models: linear regression, decision tree, optimized decision tree, neural network, Auto Regressive Integrated Moving Average (ARIMA), and exponential smoothing. Using Indonesia's agricultural export data, model performance was evaluated under two data-splitting schemes (80%:20% and 75%:25%) with error metrics including MAE, MSE, RMSE, and MAPE. The results consistently show that statistical time series models outperform regression-based and machine learning approaches. In particular, SES achieved the lowest forecasting errors across all evaluation criteria, with MAPE values as low as 0.93%, followed by ARIMA as the second-best performer. Machine learning models, on the other hand, produced relatively higher error values, suggesting their limited ability to capture temporal dependencies in the data. Importantly, the choice of training–testing proportion did not significantly alter the ranking of model performance, indicating that model selection plays a more critical role than data partitioning. Overall, this study highlights the robustness of exponential smoothing methods as reliable forecasting tools for Indonesia's agricultural exports and provides evidence-based insights for policymakers in designing effective trade strategies.

Keyword: ARIMA, agricultural exports, exponential smoothing, forecasting model, machine learning.

1. Introduction

In the early 1980s and early 1990s, Indonesia was one of the most productive agricultural nations. Since its population's demands are being exceeded by the amount of production, it is exporting this production [1]. Being an agricultural nation, Indonesia has the capacity to control the global trade market in terms of the volume of agriculture products [2]. Agriculture is a major source of income for many Indonesians, particularly those who reside in rural areas, and it plays a significant role in the daily life of several Indonesians [3]. Furthermore, Indonesia's agricultural exports are an important aspect of growth, rural income, and foreign exchange, but monthly export values are influenced to global rate movement [4], [5]. Agricultural products in Indonesia that are highly competitive and of high quality will be able to penetrate export markets in ASEAN countries, which will in turn boost domestic production, increase farmers' incomes, create job opportunities for young people, and generate foreign exchange for the country. In addition, an issue affecting Indonesia's agriculture industry is the aging of its farmers, which is linked to their inability to absorb new breakthroughs or technologies [6]. Forecast accuracy is a vital role for public authorities and agribusinesses, which apply predictions to allocate resources wisely and minimize climate-related risks. Hence, strengthening accuracy and communication of forecasts is key to support agricultural decision-making [7], [8].

The process of making the most accurate predictions about the future is called forecasting [9]. The main goal of a forecasting approach is to create a mathematical model that forecasts future production by utilizing a number of criteria [10]. According to trading value and policymakers, risk management, and price speculation in particular, in commodities trading should be more precisely quantified, computed, predicted, and analyzed [11]. Forecasting the agriculture export value is intriguing and difficult because it resembles a time-series data set with a complex and dynamic stability level. Having correct information on the agricultural situation is highly crucial. Making informed decisions in today's market is made possible by the ability to forecast agricultural product trends [12]. Since it helps to lower future uncertainty, forecasting future events is crucial in many fields to aid in decision-making [13]. In this study, forecasting is not only concerned with decreasing errors after the fact, but also with guaranteeing reliability before decisions are proposed. Fragile back-testing may cause agencies to issue allows at the wrong time or allocate buffers not appropriately, which could intensify volatility instead of stabilizing it [14], [15], [16], [17]. Consequently, this research extends beyond point accuracy to evaluating how resilient model performance remains under varying validation scenarios.

An often neglected but vital address of reliability is the dividing of historical data for validation. Different train–test ratio reshape the learning and testing horizons that bigger train proportion can decrease variance yet miss recent shifts, whereas smaller test proportion horizons test stability but risk under-fitting [18], [19], [20]. This impacts both accuracy measures and model rankings sensitive to the partition ratio [21]. For policymakers, such sensitivity indicates fragile reliability and heightens decision risks in monitoring and timing. To the best of our knowledge, this is the first study that analyze training-testing proportion in the forecasting agriculture export model. There is currently no agreement based on theoretical and numerical research regarding the ideal data splitting ratio [22]. A popular method in machine learning and predictive modeling is to divide a dataset into two separate halves, usually called the training and testing datasets [23]. There are several studies that concern in different splitting dataset 75%:25% training and testing [24], 50% testing data set [25], and based on asymptotic analysis, testing ratio ought to approach 100% when the volume of data increases significantly [26]. In addition, numerical research has suggested that a testing set of about 30% is a sensible option [27]. Therefore, specifically, this study evaluates how 75%:25% and 80%:20% chronological partition affect the forecasting model accuracy and whether model rankings remain stable across these validation choices.

This study addresses this gap by benchmarking six widely used approaches on Indonesia's agricultural export series under two schemes splitting data set, 80%:20% and 75%:25% with some techniques Simple Linear Regression, Decision Tree, Neural Network, Auto Regressive Integrated Moving Average (ARIMA), and Simple Exponential Smoothing (SES). The paper is organized as follows. Section 2 presents a brief discussion of research method which consist of data used, methods used, and performance evaluation criteria. Section 3 presents result and discussion which is the forecasting results of five. And accuracies of all models are also discussed. Section 4 contains conclusions and recommendations for the future research.

2. Research Method

2.1. Data used

The study looks at monthly data on agriculture exports from Indonesia between January 2012 and May 2025. Overall, as Figure 1 illustrates, the time series of Indonesia's agricultural exports shows an overall upward trend with noticeable seasonal and cyclical fluctuations, reflecting recurring peaks and troughs over time. In the later periods, exports display higher volatility and sharp growth, indicating both increasing variability and significant expansion in the sector. Moreover, in this research, it has been evaluated models under two chronological train–test partitions that are widely used in practice: **80%:20%** and **75%:25%**, respectively[22], [24], [28]. The proportion is determined to investigate the impact of different data grouping and to balance estimation sufficiency and evaluation robustness on our 161-month series. The data partitioning is as a method for internal validation, where the model is validated on the same dataset it was developed on [29]**.** In addition, Figure 2 presents two different approaches to splitting the agricultural export time series into training and testing sets. In panel (a), 80% of the data (January 2012–September 2022) is used to train the forecasting models, while the remaining 20% (October 2022–May 2025) is reserved for testing. In panel (b), the split is adjusted to 75% training (January 2012–January 2022) and 25% testing (February 2022–May 2025), allowing comparison of how training–testing proportions influence forecasting accuracy.

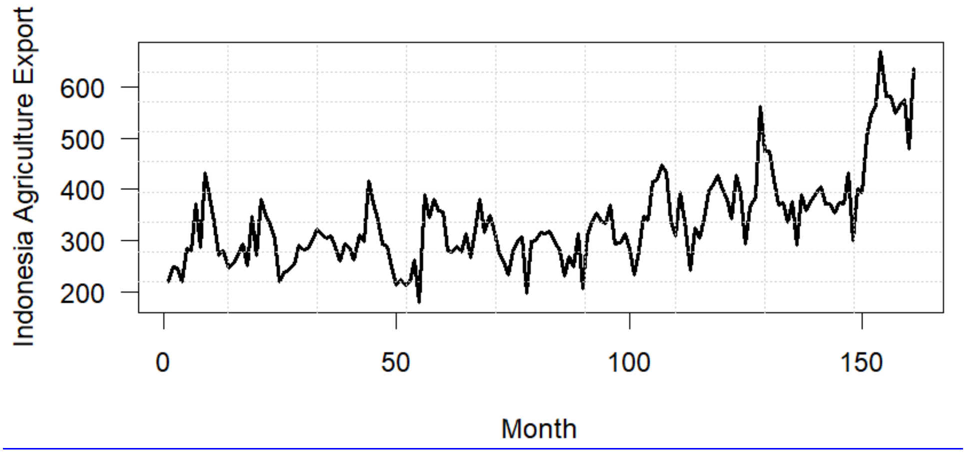


Figure 1. Time series plot of Indonesia's agriculture export.

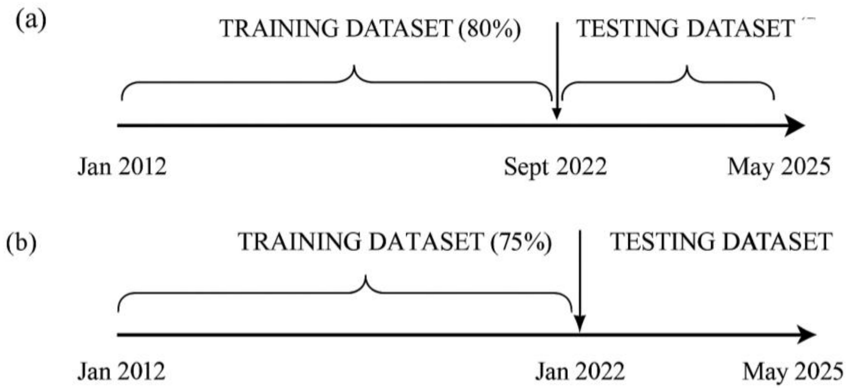


Figure 2. Training and testing dataset partition schemes for Indonesia's agricultural export forecasting (a) 80%:20%, (b) 75%:25%.

2.2. Method used

Simple Linear Regression (SLR)

By interpreting the explanatory factors in a linear function, linear regression aims to explain the fluctuation of a dependent variable [30]. Moreover, in SLR, output variable defined as Y_a is related linearly with input variable defined as X_a [31]. Equation (1) will be used to get the expected output Y for a given input X_a where m is the regression coefficient and c is the intercept of the regression line [32].

$$Y = m\, X_a + c \tag{1}$$

Decision tree

By using particular criteria during the decision-making process, the decision tree, a supervised simple classification tool can divide data records into predetermined groups [33]. It is a well-known tool that is among the most effective with comparatively low interpretability learning curves. It is frequently used in a variety of contexts, including image processing, machine learning, data mining, and pattern recognition. Decision trees are a type of machine learning algorithm that classifies or predicts a few small groups of people by charting many decision-making principles [34]. Figure 3 illustrates a decision tree's fundamental structure.

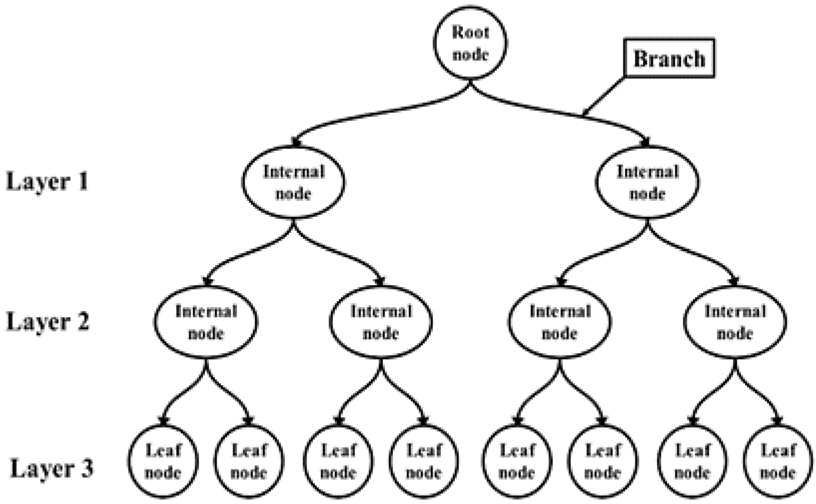


Figure 3. Decision tree basic structure [34].

Neural network (NN)

A type of artificial intelligence known as neural networks (NN) mimics the organic makeup of the human brain [35]. Applications for NN are not limited to a single field, rather, they are diverse which can be used for a number of purposes [36]. Moreover, NN is a non-linear function that maps from a measurable input set X to a measurable output set Y, f_θ, and is parameterized by model parameters θ, or the network weights [37], [38] as represented in Equation (2).

$$f_\theta : X \rightarrow Y \qquad f_\theta(x) = y \tag{2}$$

Simple exponential smoothing (SES)

A time-series forecasting technique for univariate data without trend or seasonality is called simple exponential smoothing, or SES. The smoothing factor or smoothing coefficient, alpha (α), is the only parameter needed [39], [40]. The exponential decay of the influence of the observations at previous time steps is controlled by this parameter. Typically, alpha is set to a number between 0 and 1. Furthermore, SES corresponds to Equation (4) written as

$$y'_t = \alpha\, \mathrm{y_t} + (1-\alpha) y'_{t-1} \tag{3}$$

when y_t is denoted as the observation value, y'_t is denoted as the fitted value of y_t, α is denoted as the weight place on the end observation ($0 \leq \alpha \leq 1$).

Auto Regressive Integrated Moving Average (ARIMA)

One kind of statistical model for evaluating and predicting time-series data is an ARIMA model [39]. When the data is univariate, stationary, and free of anomalies, ARIMA can be applied. The difference can be applied once or twice to eliminate non-stationarity cases [40]. Furthermore, ARIMA corresponds to Equation (4) written as

$$y'_t = c + \emptyset_1 y'_{t-1} + \cdots + \emptyset_p y'_{t-p} + \emptyset_p \in_{t-1} + \cdots + \emptyset_q \in_{t-q} + \in_t \tag{4}$$

when the differenced series denoted by y'_t. Both lagged errors and values of y'_t are included in the predictors. The following factors determine it:

i. p (lag order): the number of lag observations in the model
ii. d (degree of differencing): the number of times the raw observations are differentiable
iii. q (order of moving average): the size of the moving average window

2.3. *Performance evaluation criteria*

To assess the forecasting performance of each model, several error metrics were employed. These metrics are widely used in time series forecasting to capture different aspects of model accuracy. The following four criteria were adopted [41], [42], [43]:

1. Mean Absolute Error (MAE)
   Without taking into account the direction of the errors, MAE calculates the average size of the errors in a collection of forecasts. It is described in Equation (5) as follow:
   $$MAE = \frac{1}{m}\sum_{i=1}^{m} |X_i - Y_i| \tag{5}$$
   where Y_i is the actual ith value, and X_i is the forecasted ith value, and m is the number of observations.
2. Mean Squared Error (MSE)
   MSE squares the difference between actual and predicted values to penalize greater errors more severely. It is written in Equation (6) as follow:
   $$MSE = \frac{1}{m}\sum_{i=1}^{m} (X_i - Y_i)^2 \tag{6}$$
3. Root Mean Squared Error (RMSE)
   An interpretable metric in the same unit as the data is provided by RMSE, which is the square root of MSE. It is described in Equation (7) as follow:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \quad (7)$$

4. Mean Absolute Percentage Error (MAPE)
   Forecasting accuracy is expressed as a percentage using MAPE, making it simple to interpret across datasets. It is written in Equation (8) as follow:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{X_i - Y_i}{y_i}\right| \quad (8)$$

3. Result and Discussion

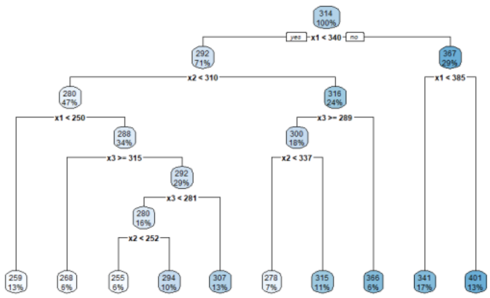
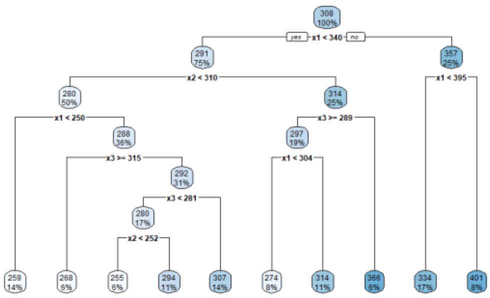
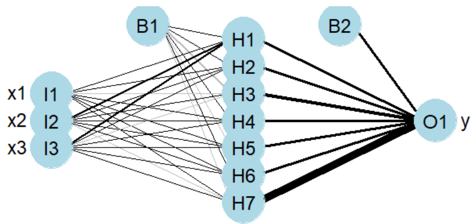
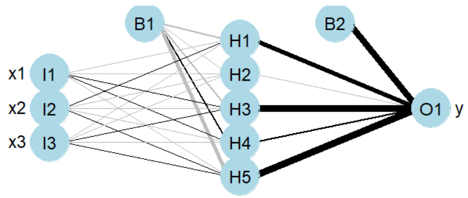
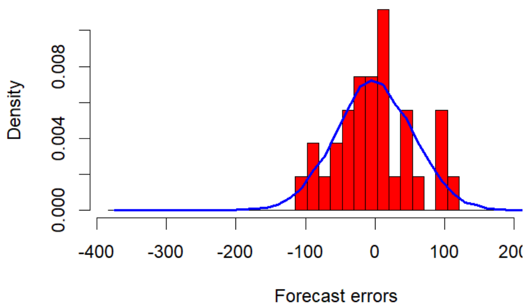
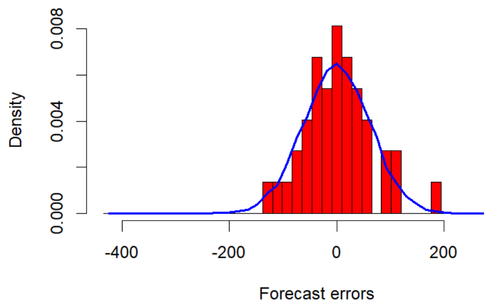
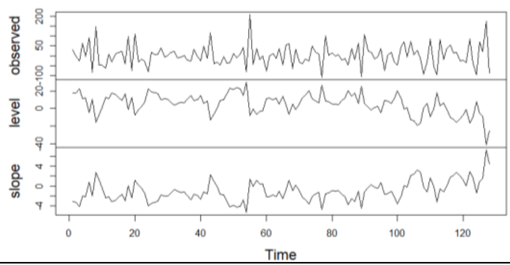
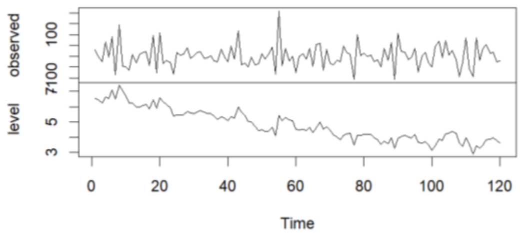
It has been conducted five models to forecast Indonesia's agricultural export which are linear regression, decision tree, neural network, ARIMA and simple exponential smoothing. The observation values have been divided into training and testing data set such that 80%:20% and 75%-25%, respectively. According to Table 1, The regression model for 80% training data $y = 109.545 + 0.526\,x_1 + 0.284\,x_2 - 0.154\,x_3$ shows that x_1 and x_2 significantly and positively influence the dependent variable, while x_3 has a negative but insignificant effect. The model explains about 39% of the variance ($R^2 = 0.3937$) and is highly significant overall ($p < 0.001$). In another words, the regression model for 75% train-data $y = 126.91038 + 0.48458\,\mathrm{x}_1 + 0.29048\,\mathrm{x}_2 - 0.18553\,\mathrm{x}_3$ suggests that x_1 and x_2 have significant positive effects on the dependent variable, whereas x_3 contributes negatively but without statistical significance. With an $R^2 = 0.368$ and an overall p-value of 2.402×10^{-11}, the model explains about 37% of the variance and is statistically significant as a whole.

The decision tree for 80% training data shows that x_1 is the most influential predictor, with further splits refined by x_2 and x_3. The terminal nodes represent forecasted agricultural export values for different data segments, capturing nonlinear relationships among the variables. While the decision tree model for 75% training data starts with x_1 as the root split, followed by x_2 and x_3 as secondary predictors. The structure shows how combinations of these variables define different paths, leading to terminal nodes that represent predicted agricultural export values for specific conditions. Moreover, NN model in this approach has structure with three input variables, one hidden layer containing seven neurons for 80% training data while five neurons for 75% training data, and a single output neuron. The model utilizes weighted connections and biases to capture nonlinear relationships between predictors and agricultural export values.

In this study, ARIMA model proposed with first differencing. It means that parameter d equals 1. The ARIMA residuals histogram for both training-testing data proportions indicate that the residuals are close to a normal distribution because the majority of forecast errors are centered around zero and essentially follow a bell-shaped curve. Given that errors appear random rather than systematic, this suggests that the ARIMA model has done a respectable job of capturing the underlying structure of the data. In addition, exponential smoothing model for both training-testing data proportions decompose the dataset into some components. This makes its model particularly effective for time series with non-linear trends and multiple seasonal cycles.

Table 1. Model construction under two training–testing proportions.

Model	Construction	
	80%:20%	75%:20%
Linear regression	$y = 109.545 + 0.526\,x_1 + 0.284\,x_2 - 0.154\,x_3$	$y = 126.91038 + 0.48458\,\mathrm{x}_1 + 0.29048\,\mathrm{x}_2 - 0.18553\,\mathrm{x}_3$

Decision tree		
Neural network		
ARIMA		
SES		

The forecasting plots in Figure 4 reveal notable differences in predictive performance among the models. Linear regression (a) and decision tree (b) tend to underestimate fluctuations, producing smoother predictions compared to the actual values. The neural network (c) shows improved adaptability to changing trends but still exhibits slight deviations in peak values. ARIMA (d) demonstrates strong alignment with the actual series, effectively capturing short-term variations, while simple exponential smoothing (e) provides stable predictions but fails to adequately follow rapid changes. Although actual values span the full time step in Figure 4(e), in the simple exponential smoothing plot they are obscured because the red forecast line overlaps the blue actual line beyond step 15, owing to plotting order and high similarity. Overall, ARIMA and neural network models exhibit the closest fit to the actual export data, indicating their suitability for capturing both trend and variability in Indonesia’s agricultural export series.

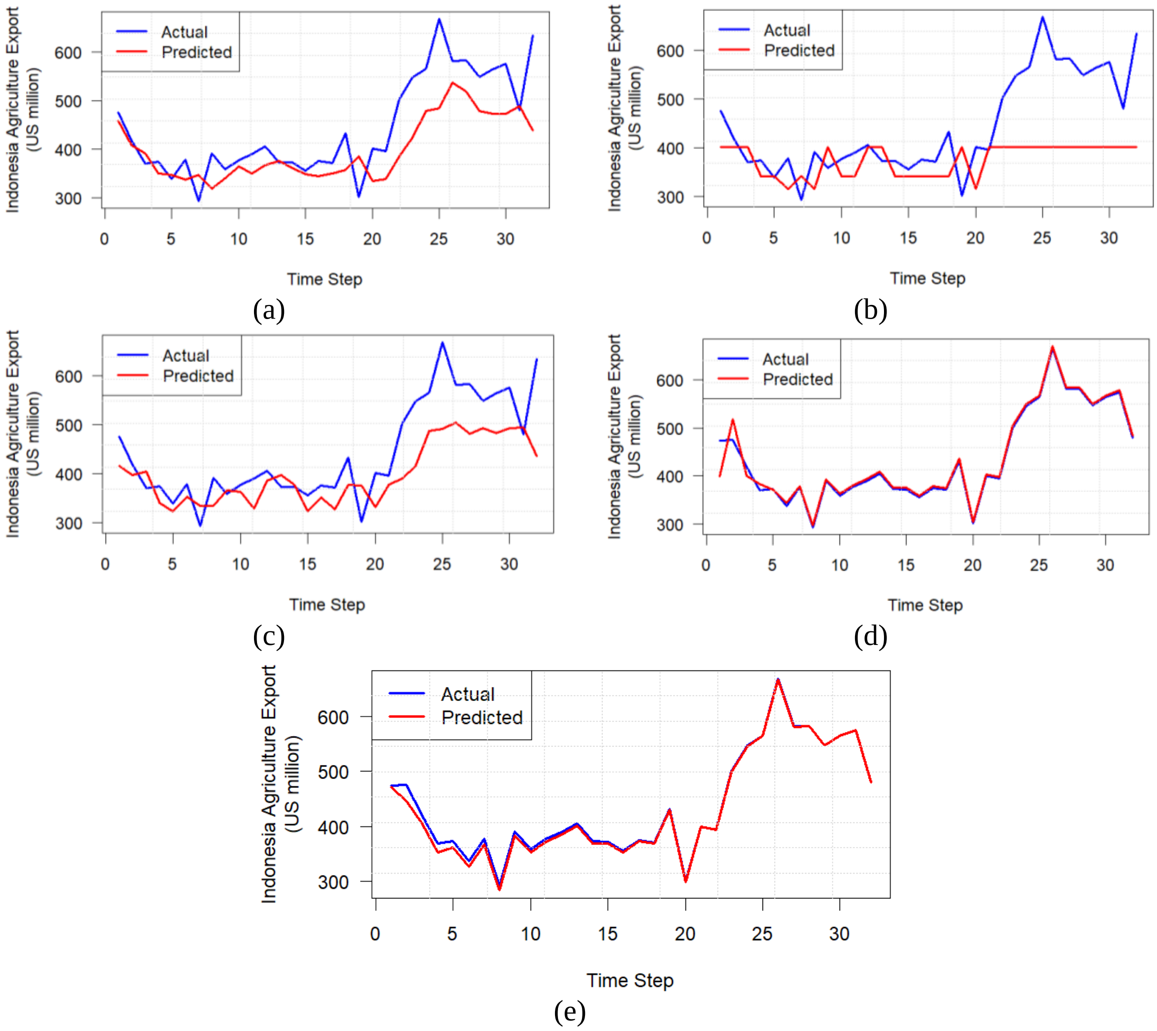


Figure 4. Forecasting performance on Indonesia's agricultural export data using the 80%:20% training–testing split: (a) simple linear regression, (b) decision tree, (c) neural network, (d) ARIMA, and (e) simple exponential smoothing.

Figure 5 compares the forecasting performance of five models against the actual agricultural export data. The simple linear regression (a), decision tree (b), and neural network (c) show noticeable deviations between actual and predicted values, particularly in capturing sharp fluctuations. ARIMA (d) demonstrates improved alignment, but still underestimates peak values. In contrast, simple exponential smoothing (e) exhibits the closest match to the actual series, effectively following both the overall trend and short-term variations, consistent with its superior error metrics reported in Table 1.

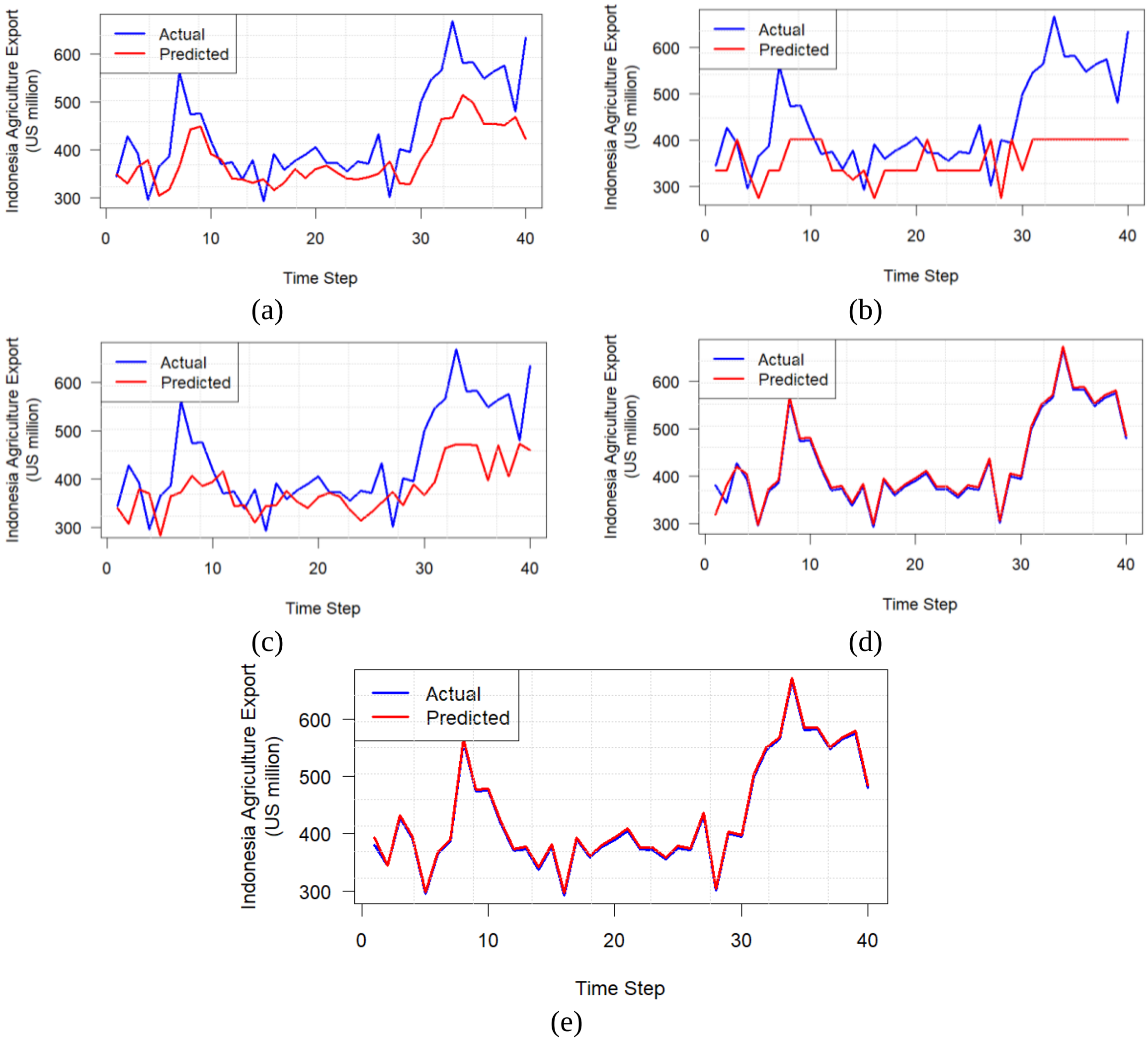


Figure 5. Forecasting performance on Indonesia's agricultural export data using the 75%:25% training–testing split: (a) simple linear regression, (b) decision tree, (c) neural network, (d) ARIMA, and (e) simple exponential smoothing.

Table 2 presents the forecasting performance of six different models—linear regression, decision tree, optimized decision tree, neural network, ARIMA, and simple exponential smoothing (SES)under two training–testing proportions (80%:20% and 75%:25%). Across both settings, the statistical time series models (ARIMA and SES) consistently outperform machine learning and regression-based approaches, highlighting their ability to effectively capture the temporal dependencies in agricultural export data. Specifically, SES achieves the lowest errors in all evaluation metrics, with MAE of 4.7699 and 3.8349, and MAPE of 1.2148% and 0.9322% for the 80%:20% and 75%:25% splits, respectively. These results demonstrate that SES is highly reliable and well-suited for forecasting data with relatively stable structures. ARIMA ranks as the second-best performer, also producing relatively low MAE and RMSE values, although slightly higher than those of SES.

By contrast, machine learning models such as decision tree, optimized decision tree, and neural network exhibit considerably higher error values across MAE, MSE, RMSE, and MAPE. This suggests that these models may be less effective in modelling the time-dependent and auto correlated nature of agricultural export data, which is better suited to statistical approaches. These findings align with prior studies [44] that emphasize the robustness of exponential smoothing methods in economic and agricultural forecasting. Furthermore, although neural networks are capable of capturing nonlinearities, their performance is often dependent on larger datasets and extensive parameter tuning (Zhang et al., 2019), which may explain their lower accuracy in this case.

Table 2. Forecasting performance of different models under two training–testing proportions.

	MAE	MSE	RMSE	MAPE
80%:20%				
Linear regression	64.7610	6885.4812	82.9788	13.5192
Decision tree opt	92.8090	13404.4527	115.7776	19.1666
Neural network	57.8723	5465.6922	73.9303	12.1602
ARIMA	7.7257	261.8439	16.1816	1.7618
SES	4.7699	59.4099	7.7078	1.2148
75%:25%				
Linear regression	66.0883	7134.5643	84.4664	14.2112
Decision tree opt	89.0647	12671.9374	112.5697	18.6898
Neural network	70.5273	8015.4186	89.5289	15.0187
ARIMA	7.8641	159.8300	12.6424	1.9706
SES	3.8349	17.3444	4.1647	0.9322

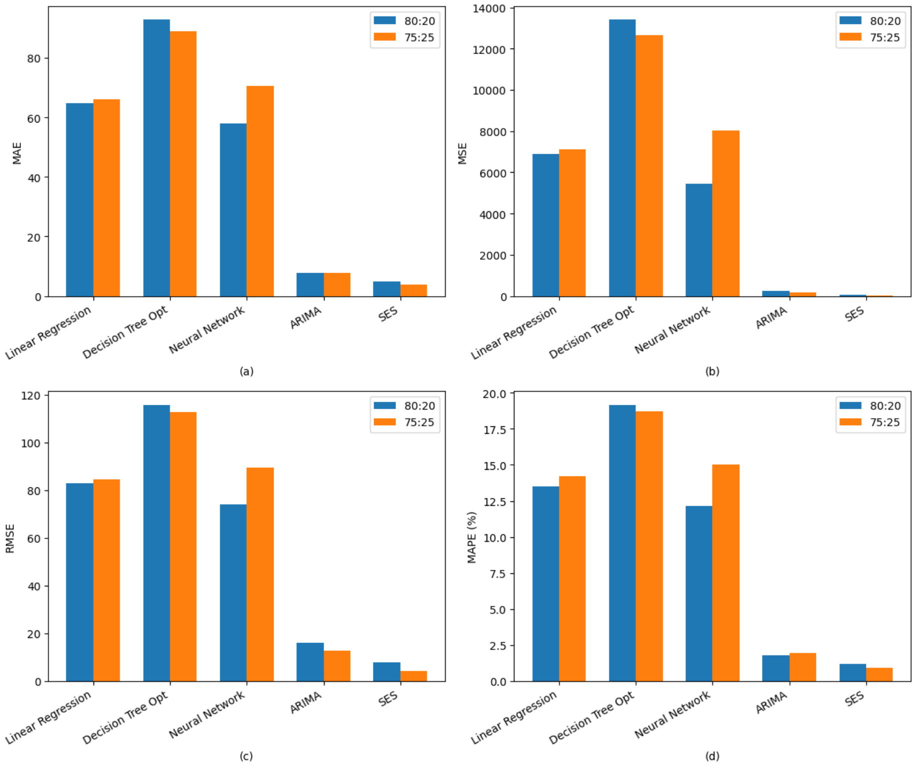


Figure 6. Comparison of forecasting performance across different models under two training–testing (a) MAE, (b) MSE, (c) RMSE, and (d) MAPE.

The results in Figure 6 further reinforce these observations by illustrating that SES consistently produces the lowest error metrics across both training–testing proportions, confirming its robustness. ARIMA also demonstrates strong performance, albeit slightly inferior to SES, while regression-based and machine learning models fall behind. Across both splitting schemes (75%:25% and 80%:20%), the relative ordering of models remains stable: SES obtains the lowest MAE/RMSE/MAPE, ARIMA is the second rank, and machine-learning/regression baselines trail behind. Absolute error levels change only marginally between the two proportions, indicating that the ratio selection within this range has a negligible effect on out-of-sample performance. Importantly, the choice of training–testing proportion (80%:20% vs. 75%:25%) appears to have only a minor influence on forecasting accuracy, as the ranking of models remains unchanged. This indicates that model selection has a far greater impact on forecasting performance than the specific partitioning of training and testing data. From a practical standpoint, this robustness underscores the reliability of SES and ARIMA as benchmark models for forecasting Indonesia's agricultural export trends. Taken together, these findings highlight that exponential smoothing methods not only deliver the most accurate forecasts but also maintain stable performance under different data-splitting schemes. This consistency is particularly valuable for policymakers and practitioners, as it ensures that forecasting outcomes remain dependable despite variations in data availability. While more complex machine learning approaches may be useful in other contexts, the results suggest that statistical models remain superior for forecasting agricultural export data with strong temporal patterns.

The results highlight that exponential smoothing methods provide the most accurate and reliable forecasts for Indonesia's agricultural export series. This finding is consistent with earlier studies that emphasize the robustness of exponential smoothing in modeling economic and trade-related time series [9]. Previous works have also shown that exponential smoothing can outperform more complex approaches such as ARIMA and neural networks, particularly in datasets with trend and seasonal structures [45], [46]. Another important finding is that varying the training–testing proportion (75%:25% vs. 80%:20%) does not significantly affect the ranking of forecasting performance. This result supports previous evidence that exponential smoothing yields stable and consistent accuracy across different data partitions, making it highly reliable for practical forecasting applications [47]. Taken together, these findings strengthen the position of exponential smoothing as a benchmark forecasting method due to its stability, interpretability, and computational simplicity. While advanced methods such as ARIMA, ANFIS, or neural networks may provide competitive results, exponential smoothing offers a dependable balance between accuracy and practicality, making it suitable for guiding trade-related decision-making in Indonesia.

4. Conclusion

This study shows that, for Indonesia's monthly agricultural exports, changing the train–test split between 75%:25% and 80%:20% has little effect on forecasting accuracy and does not alter the ranking of models. Simple Exponential Smoothing (SES) consistently provides the best results, followed by ARIMA, while regression, decision tree, and neural network models perform less effectively. In practice, this suggests that policymakers and agencies should focus more on selecting the robust model which may apply SES or ARIMA as reliable baselines across multiple seasonal cycles treating the exact proportion of training data within the 75–80% range as less critical. More complex techniques may only be crucial when monitoring indicates unusual patterns or structural shifts in the data. However, this study also has several limitations: it relies on a univariate series from a single country at a monthly frequency, it evaluates only two train–test splits, and it excludes external drivers as well as rolling-origin or block cross-validation methods. Future research could expand by testing a wider range of split ratios, adopting rolling or blocked validation techniques, incorporating exogenous factors and regime-shift diagnostics, and exploring decision-oriented loss functions and prediction intervals. Extending the analysis to multivariate, deep learning, or hybrid models (such as ARIMAX, VAR, LSTM, or ANFIS), as well as applying forecast combinations and testing other commodities or frequencies, would help assess the generalizability and robustness of these findings.

Acknowledgement

The authors would like to express their sincere gratitude to the anonymous reviewers for their valuable insights and constructive feedback, which have significantly improved the quality of this manuscript. The authors are also grateful to all colleagues and institutions that supported this research directly or indirectly.

References

[1] D. S. Dewanti and F. P. Purna, "Indonesia's Agriculture Tax: An Approach to the GTAP Model," *J. Ekon. dan Stud. Pembang.*, vol. 14, no. 1, 2022, doi: 10.17977/um002v14i12022p096.

[2] G. Widhiyoga, H. Wijayati, and R. Alma'unah, "Export Performance Of Indonesia's Leading Tropical Fruit Commodities To Main Destination Countries," *IQTISHADUNA J. Ilm. Ekon. Kita*, vol. 12, no. 1, pp. 128–148, Jun. 2023, doi: 10.46367/iqtishaduna.v12i1.1126.

[3] K. F. Arifah and J. Kim, "The Importance of Agricultural Export Performance on the Economic Growth of Indonesia: The Impact of the COVID-19 Pandemic," *Sustain.*, vol. 14, no. 24, Dec. 2022, doi: 10.3390/su142416534.

[4] E. D. Satriana, Harianto, and D. S. Priyarsono, "Pengaruh volatilitas nilai tukar terhadap kinerja ekspor utama pertanian indonesia," vol. 13, no. 2, pp. 163–186, 2019, doi: 10.30908/BILP.V13I2.424.

[5] U. Kalsum, N. Djamil, T. Nirmala, S. Wifasari, and I. G. P. R. Andaningsih, "The Effect of Global Commodity Price Volatility on Indonesia's Trade Balance," *Nomico.*, vol. 2, no. 7, pp. 1–7, 2025, doi: 10.62872/y3eepe50.

[6] R. R. Rachmawati and E. Gunawan, "Peranan Petani Milenial mendukung Ekspor Hasil Pertanian di Indonesia,"

Forum Penelit. Agro Ekon., vol. 38, no. 1, 2020, doi: 10.21082/fae.v38n1.2020.67-87.
[7] P. Calanca, "Weather Forecasting Applications in Agriculture," 2014, pp. 437–449. doi: 10.1016/B978-0-444-52512-3.00234-5.
[8] D. Lemoine and S. Kapnick, "Financial markets value skillful forecasts of seasonal climate," *Nat. Commun.*, vol. 15, 2024, doi: 10.1038/s41467-024-48420-z.
[9] R. J. Hyndman and Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.
[10] S. A. Haider *et al.*, "LSTM neural network based forecasting model for wheat production in Pakistan," *Agronomy*, vol. 9, no. 2, Feb. 2019, doi: 10.3390/agronomy9020072.
[11] R. R. Erlina and R. Azhar, "FORECASTING MODEL OF AGRICULTURE COMMODITY OF VALUE EXPORT OF COFFEE : APPLICATION OF ARIMA MODEL," vol. 9, no. 3, pp. 257–263, 2020.
[12] P. Phakdi, "Forecasting Model for the Value of Areca Nut's Export of Thailand," *Webology*, vol. 18, no. Special Issue, pp. 1241–1253, 2021, doi: 10.14704/WEB/V18SI04/WEB18195.
[13] I. Ilic, B. Görgülü, M. Cevik, and M. G. Baydoğan, "Explainable boosted linear regression for time series forecasting," *Pattern Recognit.*, vol. 120, Dec. 2021, doi: 10.1016/j.patcog.2021.108144.
[14] T. Aditya, S. Jaipuria, and P. K. Dadabada, "Model selection for long-term load forecasting under uncertainty," *J. Model. Manag.*, 2024, doi: 10.1108/jm2-09-2023-0211.
[15] L. Wang *et al.*, "Robust Nonparametric Distribution Forecast with Backtest-Based Bootstrap and Adaptive Residual Selection," *ICASSP 2022 - 2022 IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 3903–3907, 2022, doi: 10.1109/icassp43922.2022.9747701.
[16] A. Stratigakos, P. Andrianesis, A. Michiorri, and G. Kariniotakis, "Towards Resilient Energy Forecasting: A Robust Optimization Approach," *IEEE Trans. Smart Grid*, p. 1, 2023, doi: 10.1109/tsg.2023.3272379.
[17] M. Zanotti, "The effects of retraining on the stability of global models in retail demand forecasting," 2025, doi: 10.48550/arxiv.2506.05776.
[18] B. Jafrasteh, E. Adeli, K. M. Pohl, A. Kuceyeski, M. R. Sabuncu, and Q. Zhao, "Statistical variability in comparing accuracy of neuroimaging based classification models via cross validation," *Dent. Sci. reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-12026-2.
[19] G. Gallitto *et al.*, "External validation of machine learning models—registered models and adaptive sample splitting," *Gigascience*, vol. 14, 2025, doi: 10.1093/gigascience/giaf036.
[20] H. R. Maier *et al.*, "On how data are partitioned in model development and evaluation: Confronting the elephant in the room to enhance model generalization," *Environ. Model. Softw.*, vol. 167, p. 105779, 2023, doi: 10.1016/j.envsoft.2023.105779.
[21] C. J. Brown, C. A. Buelow, R. D. Stuart-Smith, N. S. Barrett, G. J. Edgar, and E. Oh, "Assessing predictive accuracy of species abundance models in dynamic systems," *Methods Ecol. Evol.*, 2025, doi: 10.1111/2041-210x.70105.
[22] V. R. Joseph, "Optimal ratio for data splitting," *Stat. Anal. Data Min.*, vol. 15, no. 4, pp. 531–538, Aug. 2022, doi: 10.1002/sam.11583.
[23] C. Vilette, T. Bonnell, P. Henzi, and L. Barrett, "Comparing dominance hierarchy methods using a data-splitting approach with real-world data," *Behav. Ecol.*, vol. 31, no. 6, pp. 1379–1390, Nov. 2021, doi: 10.1093/beheco/araa095.
[24] M. M. Oghaz, M. A. Maarof, M. F. Rohani, A. Zainal, and S. Z. Mohd Shaid, "A hybrid color space for skin recognition for real-time applications," *J. Comput. Theor. Nanosci.*, vol. 14, no. 4, 2017, doi: 10.1166/jctn.2017.6516.
[25] G. Afendras and M. Markatou, "Optimality of training/test size and resampling effectiveness in cross-validation," *J. Stat. Plan. Inference*, vol. 199, pp. 286–301, Mar. 2019, doi: 10.1016/j.jspi.2018.07.005.
[26] A. Dubbs, "Test Set Sizing via Random Matrix Theory," *Oper. Res. Forum*, vol. 5, no. 1, Mar. 2024, doi: 10.1007/s43069-024-00292-1.
[27] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/4832864.
[28] T. W. Septiarini, M. R. Taufik, and T. A. E. Prasetya, "A comparative forecasting model of COVID-19 case in Indonesia," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jun. 2021. doi: 10.1088/1742-6596/1918/4/042020.
[29] B. Van Calster, E. W. Steyerberg, L. Wynants, and M. van Smeden, "There is no such thing as a validated prediction model," *BMC Med.*, vol. 21, no. 1, 2023, doi: 10.1186/s12916-023-02779-w.
[30] M. Flores-Sosa, E. Leon-Castro, E. Aviles-Ochoa, and J. M. Merigo, "FORECASTING VOLATILITY WITH SIMPLE LINEAR REGRESSION AND ORDERED WEIGHTED AVERAGE OPERATORS," *Econ. Comput. Econ. Cybern. Stud. Res.*, vol. 56, no. 3, 2022, doi: 10.24818/18423264/56.3.22.13.
[31] V. Veeramsetty, A. Mohnot, G. Singal, and S. R. Salkuti, "Short term active power load prediction on A 33/11 kV substation using regression models," *Energies*, vol. 14, no. 11, Jun. 2021, doi: 10.3390/en14112981.
[32] H. S. Al-Zboon and M. I. Y. Alharayzeh, "The Impact of Guessing on the Accuracy of Estimating Simple Linear

Regression Equation Parameters and the Ability to Predict," *Int. J. Instr.*, vol. 16, no. 2, pp. 927–944, Apr. 2023, doi: 10.29333/iji.2023.16249a.

[33] C. S. Lee, P. Y. S. Cheang, and M. Moslehpour, "Predictive Analytics in Business Analytics: Decision Tree," *Adv. Decis. Sci.*, vol. 26, no. 1, pp. 1–29, Mar. 2022, doi: 10.47654/V26Y2022I1P1-30.

[34] M. Do, W. Byun, D. K. Shin, and H. Jin, "Factors influencing matching of ride-hailing service using machine learning method," *Sustain.*, vol. 11, no. 20, Oct. 2019, doi: 10.3390/su11205615.

[35] T. T. K. Tran, S. M. Bateni, S. J. Ki, and H. Vosoughifar, "A review of neural networks for air temperature forecasting," May 01, 2021, *MDPI AG*. doi: 10.3390/w13091294.

[36] M. Madhiarasan and M. Louzazni, "Analysis of Artificial Neural Network: Architecture, Types, and Forecasting Applications," *J. Electr. Comput. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/5416722.

[37] J. Gawlikowski *et al.*, "A survey of uncertainty in deep neural networks," *Artif. Intell. Rev.*, vol. 56, 2023, doi: 10.1007/s10462-023-10562-9.

[38] A. Almosova and N. Andresen, "Nonlinear inflation forecasting with recurrent neural networks," *J. Forecast.*, vol. 42, no. 2, 2023, doi: 10.1002/for.2901.

[39] A. Bhattacharjee, G. K. Vishwakarma, N. Gajare, and N. Singh, "Time Series Analysis Using Different Forecast Methods and Case Fatality Rate for Covid-19 Pandemic," *Reg. Sci. Policy Pract.*, vol. 15, no. 3, pp. 506–519, Apr. 2023, doi: 10.1111/rsp3.12555.

[40] T. W. Septiarini, M. R. Taufik, M. Afif, and A. Rukminastiti Masyrifah, "A comparative study for Bitcoin cryptocurrency forecasting in period 2017-2019," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1511/1/012056.

[41] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, 2021, doi: 10.7717/PEERJ-CS.623.

[42] D. Munandar, B. N. Ruchjana, and A. S. Abdullah, "Principal component analysis-vector autoregressive integrated (PCA-VARI) model using data mining approach to climate data in the West Java Region," *BAREKENG J. Ilmu Mat. Dan Terap.*, vol. 16, no. 1, pp. 99–112, 2022.

[43] R. Apriliyanti, N. Satyahadewi, and W. Andani, "Application of Extreme Learning Machine Method on Stock Closing Price Forecasting Pt Aneka Tambang (Persero) Tbk," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 1057–1068, 2023.

[44] R. Hyndman, A. Koehler, K. Ord, and R. Snyder, *Forecasting with exponential smoothing: the state space approach*. Springer, 2008.

[45] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *Int. J. Forecast.*, vol. 14, no. 1, 1998, doi: 10.1016/S0169-2070(97)00044-7.

[46] M. R. Abonazel and A. I. Abd-Elftah, "Forecasting Egyptian GDP using ARIMA models," *Reports Econ. Financ.*, vol. 5, no. 1, pp. 35–47, 2019.

[47] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and artificial neural networks models for stock price prediction," *J. Appl. Math.*, vol. 2014, no. 1, p. 614342, 2014.