

Machine Learning Algorithm for Forecasting Banking Profitability in Southeast Asia

Achmad Kautsar^{1*}

¹ Digital Business, Universitas Negeri Surabaya, Indonesia, achmadkautsar@unesa.ac.id

Corresponding Author Email: achmadkautsar@unesa.ac.id

ABSTRACT

Received : 28 October 2025
Accepted : 28 November 2025
Published : 30 December 2025

Keywords:

Machine Learning, Forecasting, Banking Profitability, ROA, Support Vector Regression, Southeast Asia.

This study examines the use of machine learning algorithms to forecast banking profitability in Southeast Asia using Return on Assets (ROA) as the main performance indicator. The study is motivated by the central role of banks as financial intermediaries, liquidity providers, and key institutions supporting financial stability across the ASEAN region. A quantitative predictive approach is employed using secondary data from the annual financial statements of banks in Indonesia, Malaysia, Singapore, Thailand, the Philippines, and Vietnam for the 2017–2024 period. The analysis is carried out in Google Colab using the Support Vector Regression (SVR) algorithm. Historical ROA values are used as predictors, while ROA in 2024 serves as the prediction target. Model performance is evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results show an MAE of 1.6299, an RMSE of 3.6195, and a MAPE of 209.33%. These results suggest that SVR can capture part of the historical pattern in bank ROA, particularly for banks with values within a more typical range. Its performance, however, becomes less stable when actual ROA is negative, close to zero, or unusually high or low. The findings therefore indicate that SVR can serve as an initial forecasting model for banking profitability, although its predictive performance should be strengthened by incorporating additional financial variables.

INTRODUCTION

Banking is one of the most strategically important sectors in Southeast Asia's financial system. Banks perform essential functions as financial intermediaries, liquidity providers, credit distributors, and supporters of regional economic activity. Indonesia, Malaysia, Singapore, Thailand, the Philippines, and Vietnam have banking industries that continue to expand in terms of assets, digital services, and financing for the real sector. Within the ASEAN context, banking stability has become increasingly important because of deeper regional financial integration, the growth of the digital economy, cross-border capital mobility, and greater exposure to global risks such as interest-rate movements, inflation, exchange-rate volatility, and credit risk. The ASEAN+3 Macroeconomic Research Office, in its

ASEAN+3 Financial Stability Report 2024, emphasizes the need for regional financial systems to strengthen their resilience to external pressures, including asset-quality risks and property-market developments that may affect bank capital.

Bank profitability across Southeast Asia has shown mixed but notable developments. Several large regional banks continued to report strong earnings, although pressure on interest margins, funding costs, credit provisions, and digital competition has intensified. For example, Singapore's DBS Group recorded net profit of SGD 11.4 billion in 2024, an increase of 11%, with an ROE of 18.0%, illustrating the strength of a major regional bank in a high-interest-rate environment. Maybank Group in Malaysia reported a 7.9% increase in profit after tax and zakat attributable to shareholders to RM10.09 billion in

FY2024, supported by stronger net operating income and improved regional economic activity. In Indonesia, banking profitability indicators also remained relatively stable; Bank Indonesia reported an industry NIM of 4.48% and ROA of 2.58% in June 2025. These developments show that Southeast Asian banks retain substantial earnings potential while operating in a rapidly changing environment.

Forecasting bank profitability is important because profit reflects more than operating success. It is also associated with financial soundness, asset-management efficiency, resilience to credit risk, and long-term competitiveness. Bank earnings are influenced by many factors, including net interest income, fee-based income, credit quality, operating efficiency, provisioning expenses, policy rates, inflation, loan growth, and broader macroeconomic conditions. Because these factors are uncertain and often interact with one another, bank management requires reliable predictive tools to support revenue planning, risk control, capital allocation, liquidity management, and data-driven business strategy. Profit forecasting is also relevant to investors, regulators, and shareholders because it can help them assess bank prospects, financial-system stability, and potential systemic risks.

The need for forecasting has become more pressing as banking undergoes structural change driven by digitalization, fintech, open banking, growing digital transactions, and shifts in customer behavior. HSBC, for example, launched a USD 1 billion ASEAN Growth Fund to support digital platform businesses in Southeast Asia, reflecting the increasing importance of the region's digital economy. These developments may create new revenue opportunities for banks, while at the same time increasing competition from fintech companies, digital banks, and non-bank financial platforms. For this reason, conventional forecasting approaches may be less effective when they are unable to capture nonlinear patterns, complex interactions, and rapid changes in banking data.

In this setting, machine learning is relevant because it can learn from historical financial and macroeconomic patterns and generate more adaptive predictions. Machine learning methods are particularly useful for handling large datasets, identifying hidden patterns, modeling nonlinear relationships, and potentially improving predictive performance when financial data are volatile or structurally complex.

Previous studies have increasingly applied machine learning to financial-performance analysis and banking prediction. Appiahene et al. (2020) combined Data Envelopment Analysis and machine learning to evaluate bank efficiency and performance, showing that machine learning can support a more comprehensive assessment of banking operations. Ryll and Seidens (2019), in a survey covering more than 150 studies, reported that machine learning algorithms often outperform a number of traditional stochastic methods in financial-market forecasting, partly because of their ability to capture temporal dependencies in financial data. Bolivar, Duran, and Lozano-Vivas (2024) also used machine learning to examine how business models contribute to bank profit performance in the European Union, further demonstrating the relevance of these techniques for explaining differences in bank profitability. Other studies have extended predictive modeling to banking risk. Rustam et al. (2025), for instance, developed bank-bankruptcy prediction models using Logistic Regression, Random Forest, and Support Vector Machine and showed how machine learning can help identify potential risk in financial institutions.

Even so, research on banking-profitability forecasting in Southeast Asia remains relatively limited compared with studies focusing on Europe, the United States, or individual countries. Southeast Asia presents a distinctive setting because its banking systems differ in market structure, level of digitalization, regulatory environment, asset quality, and exposure to macroeconomic conditions. Laili et al. (2023) show that ASEAN banking performance is influenced by bank-specific, industry-level, macroeconomic, and country-governance factors. This reinforces the need for forecasting models that move beyond historical profitability alone and incorporate broader financial and macroeconomic predictors.

Against this background, the present study examines the use of machine learning for forecasting banking profitability in Southeast Asia. The study seeks to evaluate the predictive performance of Support Vector Regression in estimating bank ROA across the region. The study contributes to the growing literature on machine-learning-based financial forecasting and offers practical insight for bank managers, investors, financial analysts, and regulators seeking to understand profitability patterns, anticipate performance risk, and support data-driven decision making. By applying a machine learning approach to

a cross-country Southeast Asian banking sample, the study also provides a basis for developing more accurate and context-sensitive forecasting models in future research.

The novelty of this study lies in applying a machine-learning-based profitability forecasting framework to a cross-country sample of banks from six Southeast Asian economies. While previous studies have predominantly examined developed markets or individual countries, this study considers heterogeneous banking systems within a single regional framework. By using historical ROA to predict subsequent profitability, the study provides an initial empirical foundation for developing more comprehensive machine learning models tailored to the Southeast Asian banking context.

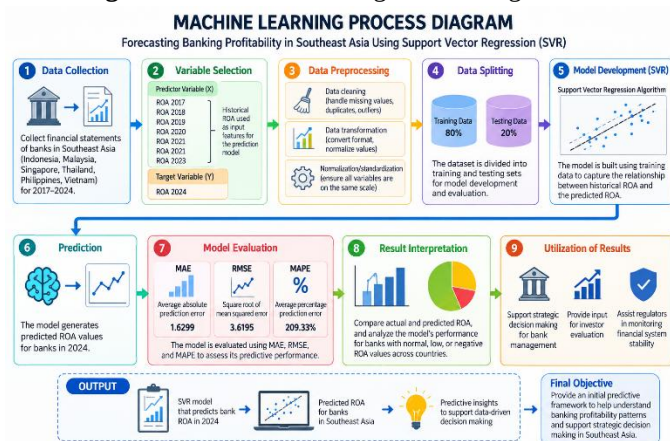
METHOD

This study adopts a quantitative predictive research design using machine learning. The quantitative approach is appropriate because the analysis is based on numerical banking data, including net income, total assets, interest income, operating expenses, loans, equity, and financial ratios. The predictive approach is used because the primary objective is to construct a forecasting model for bank profitability in Southeast Asia based on historical patterns.

From an information-systems perspective, this study falls within the area of data-driven decision support, in which data and computational algorithms are used to support managerial and analytical decision making. The machine learning model developed in this study is intended to assist bank managers, investors, and regulators in estimating future profitability more objectively and systematically.

The research focuses on banks in Southeast Asia, particularly in Indonesia, Malaysia, Singapore, Thailand, the Philippines, and Vietnam. The sample consists of banks with complete and publicly accessible annual financial statements for the observation period, covering 2017–2024.

Figure 1. Machine Learning Model Diagram



As illustrated in Figure 1, the processed dataset is divided into training and testing subsets, with 80% of the observations used for model development and 20% used for model evaluation. SVR is considered suitable for this study because banking profitability may exhibit nonlinear patterns that are difficult to capture using purely linear forecasting approaches. The separation between training and testing data also allows the model to be evaluated using observations that are not used directly during model estimation.

The study uses secondary data collected from banks' annual financial statements, investor reports, stock-exchange publications, financial-services authorities, central banks, and other relevant financial-data sources. The dataset has a panel structure because it combines cross-sectional observations across banks with time-series observations across years. This structure makes it possible to capture both differences in profitability between banks and changes in performance over time.

The study includes a target variable and predictor variables. The target variable is banking profitability, measured using net income or the Return on Assets (ROA) ratio.

$$ROA = \frac{\text{Laba Bersih}}{\text{Total Aset}} \times 100\%$$

A higher ROA indicates that a bank is more efficient in using its assets to generate earnings. A lower ROA, by contrast, suggests that the bank's asset base is producing a relatively weaker level of profit.

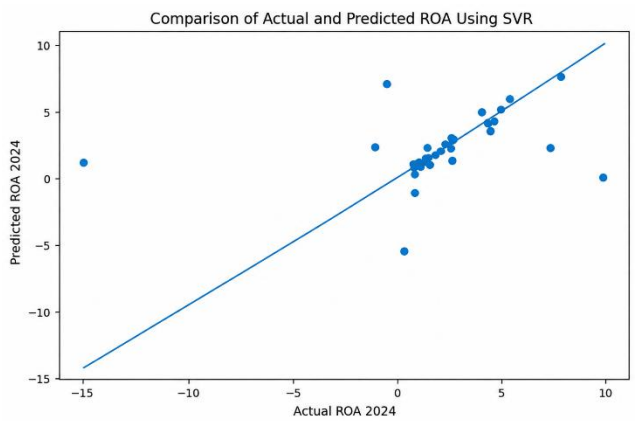
Data processing is conducted in several stages. First, the financial-statement data are collected and organized in tabular form. Second, data cleaning is carried out to remove missing, duplicated, or inconsistent observations. Third, the data are

transformed where necessary, including standardizing currency formats, calculating financial ratios, and constructing the target variable. Fourth, normalization or standardization is applied so that the variables are expressed on comparable scales.

The analytical procedure is implemented in Google Colab, a cloud-based computing platform that allows Python code to be executed online without additional local software installation. The forecasting model used in this study is Support Vector Regression (SVR). SVR is selected because it is capable of modeling nonlinear relationships and can be applied effectively in high-dimensional data settings. Model performance is evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

RESULT

The Support Vector Regression (SVR) model is used to predict banks' Return on Assets (ROA) for 2024 from historical ROA values in previous years. SVR is employed because of its ability to handle high-dimensional data and model nonlinear relationships. In this study, model performance is assessed by comparing actual ROA in 2024 with the corresponding SVR predictions.



The scatter plot shows that a substantial share of observations lies close to the diagonal line. This line represents the ideal case in which the predicted value is equal or very close to the actual value. Observations located near the diagonal therefore indicate relatively accurate predictions. Nevertheless, several points remain far from the line, particularly for banks with very low, negative, or unusually high actual ROA. This pattern suggests that the SVR model has

difficulty forecasting extreme observations or banks whose profitability patterns differ substantially from those of the majority of the sample.

Overall, the visualization suggests that SVR performs more reasonably for banks whose actual ROA falls within a relatively typical range of approximately 0% to 5%. Prediction accuracy becomes weaker for banks with large negative ROA or exceptionally high positive ROA. One possible explanation is the considerable variation in profitability across banks and countries in Southeast Asia. Banks may also differ in asset size, operating efficiency, credit quality, income structure, and other financial characteristics that are not fully represented when historical ROA is used as the main predictor.

Analysis of Mean Absolute Error, Root Mean Squared Error, and Mean Absolute Percentage Error

Evaluasi Model Support Vector Regression	

Mean Absolute Error (MAE)	: 1.6299
Root Mean Squared Error (RMSE)	: 3.6195
Mean Absolute Percentage Error (MAPE)	: 209.33 %

The model produces a Mean Absolute Error (MAE) of 1.6299. This means that the average absolute difference between actual and predicted ROA is approximately 1.63 percentage points. For example, if a bank has an actual ROA of 3%, the model's prediction may, on average, differ from that value by about 1.63 percentage points. The result suggests that SVR can provide a useful initial estimate in a dataset with considerable variation, although the size of the error also shows that the model is not yet highly precise. For banks with a low ROA, particularly below 2%, an error of 1.63 percentage points can be economically meaningful because even small changes in ROA may reflect substantial differences in asset-use efficiency. The model therefore captures broad patterns but still requires improvement before it can provide more precise forecasts.

The Root Mean Squared Error (RMSE) is 3.6195. Because RMSE assigns a larger penalty to large errors, its value being substantially higher than the MAE indicates that several observations have relatively large prediction errors. The difference between the MAE of 1.6299 and the RMSE of 3.6195 suggests that forecast errors are not evenly distributed. Most predictions may be reasonably close to actual values, while a smaller number of banks display much larger deviations. In this study, the comparatively high

RMSE indicates that SVR is sensitive to extreme ROA observations. Such extremes may arise from large losses, higher loan-loss provisions, weak operating efficiency, or bank-specific events.

The Mean Absolute Percentage Error (MAPE) is 209.33%, indicating a very high average percentage error. This metric should, however, be interpreted cautiously in the present setting. ROA values can be very small, close to zero, or negative, and percentage-based errors become unstable when the actual denominator is close to zero. The high MAPE therefore does not necessarily mean that every prediction is poor; rather, it highlights the limitations of MAPE for datasets containing near-zero and negative profitability observations.

Therefore, the MAPE value should not be interpreted in isolation as evidence that all predictions are highly inaccurate. In datasets containing negative or near-zero profitability values, a relatively small absolute prediction error can generate a disproportionately large percentage error. For this reason, greater emphasis should be placed on MAE, RMSE, and the actual-versus-predicted visualization when evaluating the predictive performance of the SVR model. Future studies may also consider alternative scale-independent measures such as symmetric MAPE or Mean Absolute Scaled Error.

Taken together, the MAE, RMSE, MAPE, and scatter-plot results suggest that SVR can capture part of the historical ROA pattern among Southeast Asian banks, especially for observations within a normal profitability range. The model nevertheless remains less reliable for extreme values, whether very low or very high. This result is consistent with the visual pattern in which several predictions are located close to the diagonal while a smaller number of observations show large deviations.

The findings indicate that SVR has potential for forecasting bank ROA in Southeast Asia, but its performance depends heavily on data quality, cross-bank heterogeneity, outliers, and the completeness of predictor variables. ROA is influenced by more than its own historical values. Net income, total assets, operating efficiency, credit quality, income structure, funding costs, macroeconomic conditions, and monetary policy can all affect profitability. A forecasting model based primarily on historical ROA is therefore unlikely to explain the full variation in bank performance.

The country-level forecasting results also show that predicted profitability does not move in the same direction across Southeast Asia. The Philippines and Vietnam display a positive tendency because average predicted ROA for 2024 is higher than ROA in 2023. Indonesia, Malaysia, and Thailand show a negative tendency because average predicted ROA is lower than in the previous year, while Singapore appears relatively stable with only a small change in average ROA.

The more dominant negative pattern suggests that the SVR model identifies possible profitability pressure or slower performance among a number of banks. Potential explanations include lower interest margins, higher funding costs, rising operating expenses, credit risk, or historically unstable ROA. These interpretations should nevertheless be treated cautiously because the model relies mainly on historical ROA and country information. Future models would benefit from the inclusion of variables such as NIM, NPL, BOPO or cost-to-income measures, total assets, equity, loan growth, and net interest income.

CONCLUSION

The empirical results show that the Support Vector Regression model produces an MAE of 1.6299, an RMSE of 3.6195, and a MAPE of 209.33%. The MAE indicates an average ROA prediction error of approximately 1.63 percentage points, while the higher RMSE confirms the presence of several observations with large prediction errors. The very high MAPE further indicates instability when actual ROA is close to zero, negative, or highly dispersed.

SVR can therefore be regarded as an initial model for forecasting bank ROA in Southeast Asia, but further development is required. Potential improvements include adding more predictor variables, handling outliers more carefully, segmenting observations by country or bank group, and comparing SVR with alternative machine learning algorithms. These steps are important if future profitability models are to achieve higher accuracy and greater practical relevance for data-driven decision making.

From a theoretical perspective, the findings show that machine learning, particularly SVR, can be incorporated into information-systems research to

support predictive analysis of banking financial performance. Such a model can form part of a data-driven decision-support system that helps users identify profitability patterns and estimate future performance.

From a practical perspective, ROA forecasts may help bank management evaluate performance, develop profitability strategies, and identify signs of weakening performance at an earlier stage. For investors, the forecasts can serve as an additional input when assessing a bank's outlook. For regulators, predictive results can support the monitoring of banking-sector stability, particularly for banks showing declining ROA or unstable profitability patterns.

References

- Appiahene, P., Missah, Y. M., & Najim, U. (2020). Predicting bank operational efficiency using machine learning algorithm: Comparative study of decision tree, random forest, and neural networks. *Advances in Fuzzy Systems, 2020*, 1–12. <https://doi.org/10.1155/2020/8581202>
- ASEAN+3 Macroeconomic Research Office. (2024). *ASEAN+3 financial stability report 2024: Strengthening resilience to challenges ahead*. AMRO.
- Bank Indonesia. (2025). *Republic of Indonesia presentation book: August 2025*. Bank Indonesia.
- Bolivar, F., Duran, M. A., & Lozano-Vivas, A. (2024). *Business model contributions to bank profit performance: A machine learning approach*. arXiv. <https://doi.org/10.48550/arXiv.2401.12334>
- DBS Group Holdings. (2025). *DBS annual report 2024*. DBS Group Holdings Ltd.
- Laili, N. N., Prasetyo, D. A., & Winarno. (2023). ASEAN banking industry performance analysis. *Indonesian Journal of Business and Entrepreneurship, 9*(1), 1–12.
- Maybank. (2025). *Maybank annual report 2024: Financial statements*. Malayan Banking Berhad.
- Noviandy, T. R. (2024). Forecasting bank stock trends using artificial intelligence. *Indonesian Journal of Financial Studies, 3*(1), 1–12.
- Rustam, Z., Hartini, S., Islam, S. M. N., Aszhari, F. R., & Rifqi, M. (2025). *Enhancing bankruptcy prediction of banks through advanced machine learning techniques: An innovative approach and analysis*. arXiv. <https://doi.org/10.48550/arXiv.2510.06852>
- Ryll, L., & Seidens, S. (2019). *Evaluating the performance of machine learning algorithms in financial market forecasting: A comprehensive survey*. arXiv. <https://doi.org/10.48550/arXiv.1906.07786>