Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 9, No. 3, June, 2025

Machine Learning and Parameter Optimization for Banking Stability Prediction and Determinants Identification in ASEAN

Pham Thuy Tu [1], Dao Le Kieu Oanh [1], Do Doan Trang [2*]

[1] *Ho Chi Minh Universiy of Banking, 56 Hoang Dieu 2, Thu Duc, Ho Chi Minh City, Viet Nam.*

[2] *Binh Duong Universiy, 504 Binh Duong Boulevard, Thiep Thanh Ward, Thu Dau Mot City, Binh Duong Province, Viet Nam.*

Abstract

This study leverages machine learning and advanced variable selection techniques to enhance the prediction of the Bank Financial Stability Index (Z-score) in emerging ASEAN markets. Utilizing a comprehensive secondary dataset comprising macroeconomic and bank-specific indicators from 61 commercial banks across Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam (2010–2023), we systematically evaluate the predictive power of multiple machine learning models. A rigorous cross-validation framework is employed to optimize forecasting accuracy, integrating Linear Regression, Random Forest, K-Neighbors, Decision Tree, Gradient Boosting, AdaBoost, Support Vector Regression, and XGBoost with Lasso, Ridge, and Elastic Net regularization. Empirical results reveal that key drivers of financial stability include equity capital, financial leverage, return on equity, GDP growth, inflation, technological advancements, and systemic shocks like the COVID-19 pandemic. Notably, the Ridge-optimized XGBRegressor model achieves the highest predictive accuracy (~89%), demonstrating the efficacy of hybrid machine learning approaches in financial stability forecasting. These findings offer crucial insights for policymakers and regulators, facilitating data-driven strategies to strengthen banking resilience and mitigate systemic risks in volatile economic environments.

Jel Classifier: C45, C52, C55, G21, G32.

Keywords:

ASEAN; Bank;
Financial Stability;
Parameter Optimization;
Machine Learning;
Predicting.

Article History:

Received:	29	December	2024
Revised:	11	April	2025
Accepted:	17	April	2025
Published:	01	June	2025

1- Introduction

The financial health of enterprises has long been a cornerstone of economic research, with seminal contributions from Beaver (1966) [1] and Altman (1968) [2] establishing foundational frameworks for assessing financial distress and bankruptcy prediction. Subsequent refinements by Altman & Hotchkiss (2005) [3] underscored the critical interplay between financial stability, legal frameworks, and macroeconomic resilience. Within the banking sector, financial stability is not merely an institutional concern but a fundamental pillar of economic sustainability. As emphasized by Ben et al. (2022) [4], Thabet et al. (2024) [5], and Boubaker et al. (2024) [6], banks function as vital intermediaries in financial systems, facilitating capital allocation, mitigating systemic risks, and acting as buffers against economic volatility. Ensuring banking stability is thus imperative for fostering sustainable economic development, particularly in emerging markets where financial systems are more susceptible to external shocks and structural inefficiencies [7, 8].

Despite extensive academic efforts, significant research gaps persist in the study of banking stability. Prior studies have predominantly centered on three key areas: (1) constructing indices and methodologies to measure financial stability [2, 6, 9, 10]; (2) identifying and evaluating determinants influencing banking stability [7, 11-14]; and (3) developing forecasting models and early warning systems for financial distress [15-19]. However, these traditional

* **CONTACT**: doantrang.bolt@bdu.edu.vn

DOI: http://dx.doi.org/10.28991/ESJ-2025-09-03-04

approaches often encounter methodological limitations, including issues of data scarcity, multicollinearity, and endogeneity, which restrict the robustness and generalizability of findings [5, 13]. Additionally, the financial systems of emerging economies, particularly those in the ASEAN region, exhibit distinct structural vulnerabilities—ranging from high capital flow volatility to regulatory disparities—that necessitate a more nuanced analytical approach [8, 20]. These markets, characterized by rapid economic growth and financial liberalization, demand advanced predictive models that can accommodate their dynamic and complex nature.

Machine Learning (ML) and advanced optimization techniques have recently emerged as transformative tools in financial risk assessment, offering superior predictive accuracy over traditional econometric models [17]. By leveraging sophisticated algorithms, ML models can efficiently process high-dimensional financial data, detect complex patterns, and enhance forecasting precision. Techniques such as Lasso, Ridge, and Elastic Net regularization have proven particularly effective in addressing multicollinearity and improving variable selection, thereby enhancing model interpretability and stability [4, 17, 21, 22]. However, despite the demonstrated efficacy of these methodologies, their application to banking stability prediction in ASEAN economies remains largely unexplored. Given the unique financial dynamics and regulatory environments of emerging markets, there is a pressing need for tailored ML-driven models that can reliably assess banking stability while accommodating market-specific idiosyncrasies.

This study aims to bridge these research gaps by integrating ML algorithms with advanced parameter optimization techniques to construct a robust predictive model for banking stability in ASEAN's emerging economies. The key objectives of this research are threefold: (1) to systematically identify and analyze the most significant determinants of banking stability in ASEAN; (2) to enhance predictive accuracy by leveraging ML methodologies in conjunction with optimization techniques; and (3) to provide empirical insights and practical recommendations for policymakers and banking regulators to formulate more effective risk management and regulatory frameworks. By employing cross-validation techniques and rigorous model evaluation, this study seeks to ascertain the most suitable ML approaches and variable selection methods for financial stability assessment, ensuring both reliability and applicability across varying economic conditions.

This paper is structured as follows: *Section 2* presents a comprehensive literature review, highlighting theoretical perspectives and empirical studies relevant to banking stability and predictive modeling. *Section 3* details the data collection process, preprocessing techniques, and methodological framework employed in this study. *Section 4* discusses the empirical results and their implications, while Section 5 concludes with policy recommendations and future research directions. Through this research, we endeavor to contribute to the growing body of knowledge on financial stability by offering an innovative, data-driven approach tailored to the specific needs of ASEAN economies. By integrating ML with advanced optimization techniques, this study provides both a methodological advancement and actionable insights for enhancing financial resilience in emerging markets.

2- Theoretical Background and Literature Review

2-1- Bank Financial Stability and Its Relationship with Influencing Factors

The concept of bank financial stability has its theoretical foundation in earlier studies on corporate financial health, serving as a critical indicator of a bank's resilience to economic shocks and market volatilities [1, 2]. While there is no universally accepted definition, financial stability in the banking sector is broadly characterized by a bank's ability to absorb adverse shocks, manage risks effectively, and maintain operational soundness in the face of macroeconomic fluctuations and intensifying competition [13, 23, 24]. To quantify this stability, various financial ratios have been employed, including the Non-Performing Loan (NPL) ratio, Capital Adequacy Ratio (CAR), and Return on Assets (ROA). However, these metrics, while valuable, often fail to provide a comprehensive assessment of a bank's overall risk exposure and long-term financial viability. The fundamental rationale behind using the Z-score lies in its ability to capture the interplay between capital adequacy and earnings volatility, offering a forward-looking measure of a bank's default risk. According to Hafeez et al. (2022) [25], the Z-score essentially reflects the extent to which a bank's capital buffer can absorb potential fluctuations in profitability without leading to insolvency. A higher Z-score indicates greater financial stability, signifying that a bank possesses sufficient capital to withstand profit variations and adverse financial conditions. Conversely, a lower Z-score suggests increased financial fragility and a heightened risk of distress. Several compelling arguments justify the preference for the Z-score over alternative indicators. First, it provides a more holistic assessment of bank stability by integrating both capital strength and earnings volatility, as opposed to standalone measures like CAR or ROA that may not fully account for risk dynamics. Second, its applicability across diverse banking environments, particularly in economies with evolving financial structures, enhances its relevance in cross-country and longitudinal studies. Finally, its consistency in empirical research and strong theoretical underpinning makes it a widely recognized and credible metric for financial stability analysis. Given these advantages, the Z-score serves as a rigorous and comprehensive indicator for evaluating banking stability, aligning with the broader objective of understanding risk resilience in ASEAN economies. This study adopts the Z-score methodology to ensure a robust, comparative, and theoretically grounded assessment of financial stability across the banking sector.

Among the available methodologies, the Z-score and its variants have gained prominence as robust and theoretically sound measures of banking stability, particularly in the context of emerging markets and developing economies [2, 9, 25, 26]. The Z-score is calculated as follows:

$$Z_{score_{it}} = \frac{ETA_{it}+ROAA_{it}}{\partial_{ROAA_i}} \tag{1}$$

where: *ROAA:* Return on average assets, *ETA:* Equity to total assets ratio, *Standard Deviation (∂_{ROAA_i}):* calculated based on the standard deviation of each bank's ROAA by country during the study period.

Previous studies on bank financial stability are often considered in the multidimensional relationship between micro factors such as inherent bank characteristics and macroeconomic factors [11, 13, 14, 24-27]. A review of previous studies reveals the following micro and macro factors affecting bank financial stability (Figure 1):

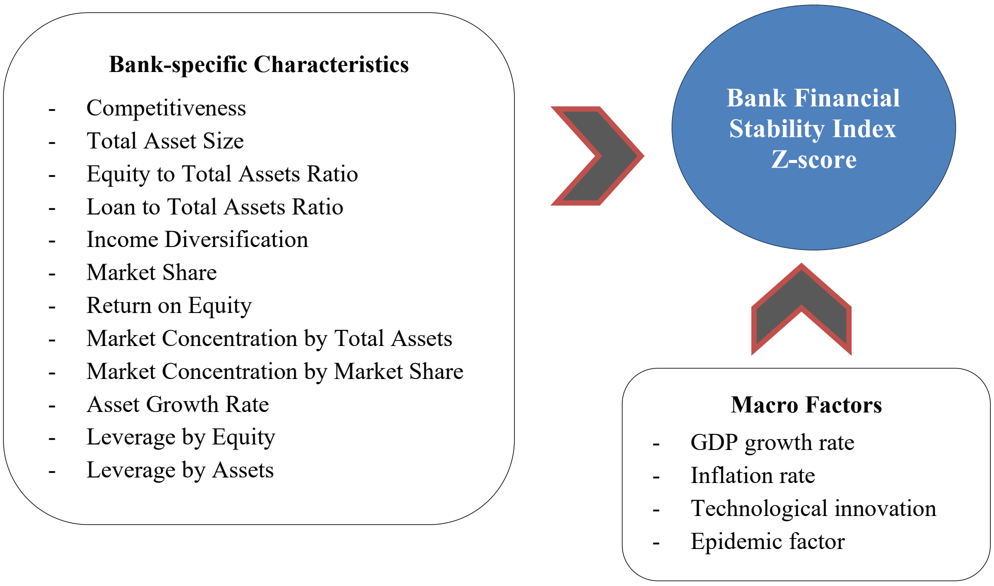

Figure 1. Factors influencing bank financial stability

2-2- Variable Selection Optimization

In research models, removing, adding, or creating new variables can lead to the loss of useful information or the addition of redundant information, affecting the accuracy of regression results and even leading to overfitting and spurious regression [4, 15, 22, 28]. To address these issues, variable selection optimization methods such as Lasso, Ridge, or Elastic Net are considered effective in regression problems [4, 5, 17, 21, 22].

Lasso regression (short for Least Absolute Shrinkage and Selection Operator) was first proposed in Tibshirani's (1996) [29]. The idea of Lasso is to add a penalty factor (called L1 Regularization) to the sum of squared errors of the linear regression model. L1 is calculated based on the sum of the absolute values of the regression coefficients. Adding L1 aims to minimize the sum of squared residuals, depending on the sum of the absolute values of the coefficients being less than a constant. Due to the nature of this constraint, it tends to create some regression coefficients exactly equal to 0 and remove the corresponding variables from the model. Reducing the number of features simplifies the model, minimizes the risk of overfitting, and improves the ability to predict accurately on new data. However, there is still a risk of potentially removing important information (Figure 2).

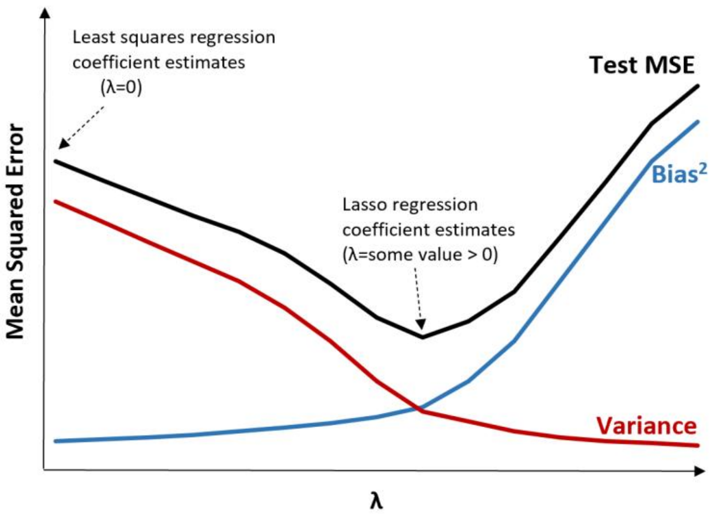

Figure 2. Comparison of bias and variance in Lasso regression

Ridge regression is quite similar to the Lasso method in its concept of adjusting the model's parameters towards a certain threshold. However, Ridge regression differs in that it minimizes the loss function by adjusting the estimated coefficients to simultaneously fit the training data well, based on the principle of retaining all input variables [28]. Accordingly, the Ridge regression technique adds a regularization component to the loss function (called L2 Regularization). This component is the sum of the squares of the model's coefficients, multiplied by a constant alpha (denoted λ), where λ is called the regularization coefficient. The goal of regularization is to minimize the values of the coefficients, thereby reducing overfitting, addressing multicollinearity, and increasing the model's stability. The performance of Ridge regression training depends on the optimal value of λ; therefore, during the training process, cross-validation is often used to help find the optimal λ for the model. However, this can lead to retaining variables that are not truly necessary in the model (Figure 3).

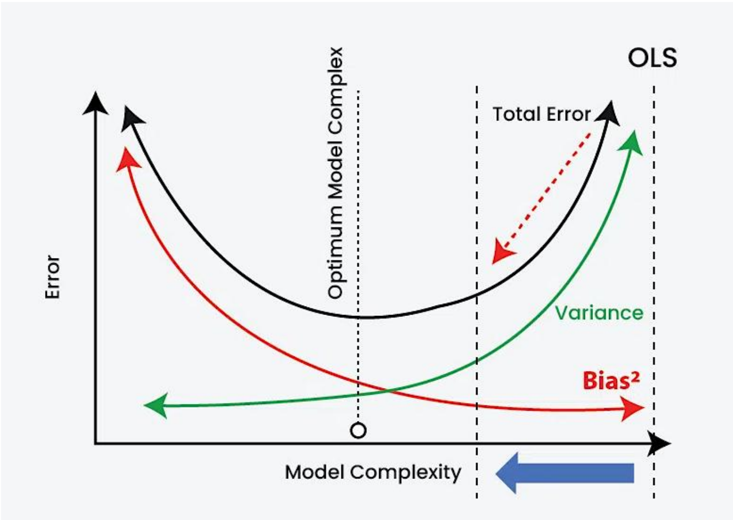

Figure 3. Comparison of bias and variance in Ridge Regression

Elastic Net regression, developed by Zou & Hastie (2005) [28], is considered a comprehensive method that leverages the advantages and overcomes the disadvantages of both Lasso and Ridge. The main idea of the Elastic Net method is to eliminate variables with estimated coefficients of 0 (after adjustment) and attempt to retain variables even if they are weakly correlated with the target variable [5, 21, 30]. Accordingly, the Elastic Net technique combines both L1 and L2 regularization components during the training process to adjust the set of estimated parameters (Figure 14).

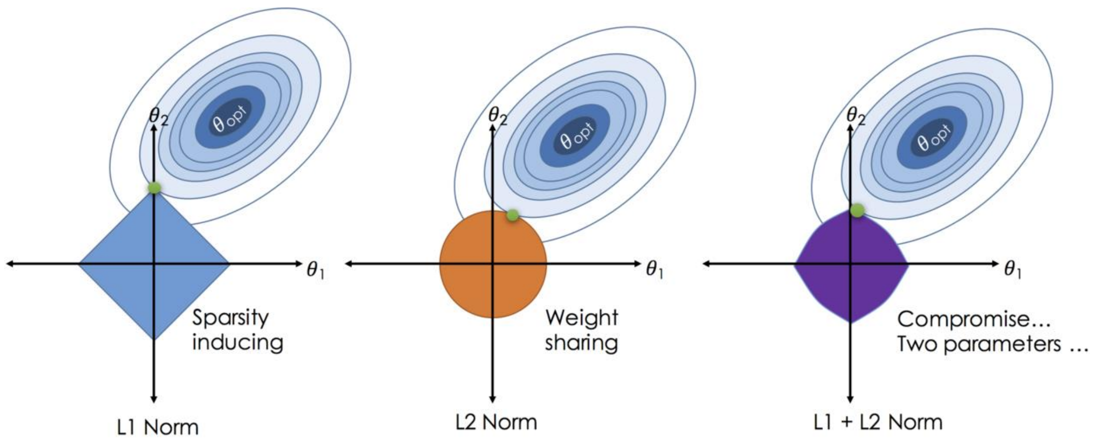

Figure 4. Concept of the Elastic Net technique1

2-3- Forecasting Models in Machine Learning

ML algorithms are increasingly used in forecasting models and have proven effective in risk forecasting models and early warning systems for bank financial instability [5, 31]. Barboza et al. (2017) [32] conducted experiments on a dataset of over 10,000 observations to forecast the bank stability of North American banks using Support Vector Machines (SVM) and Random Forest (RF), showing that ML methods provide better forecasting performance compared to traditional methods. The study by Santosh et al. (2020) [17], combining Lasso with ML algorithms such as AdaBoost, RF, and Logistic Regression to predict signs of financial decline in Indian banks, showed that Lasso helped eliminate noise and that RF and AdaBoost algorithms yielded better predictive performance than logistic regression in predicting the financial decline of banks. Thabet et al. (2024) [5] built a composite model based on the fundamental algorithms SVM, K-nearest neighbors (KNN), RF, and AdaBoost, providing empirical evidence that combining models achieves better predictive performance than training each model individually. Fernández (2020) [33] applied Tree Regression, RF, and parameter tuning techniques like Boosting and Bagging to examine the impact of factors on the financial stability of US banks, and the results showed that the RF method provided the best predictive performance for the financial stability index. Alessi & Detken (2018) [34], as well as Tanaka et al. (2016) [35], argued that applying the RF method can improve early warning predictions compared to logit models and signal-based approaches. Holopainen & Sarlin (2017) [36] emphasized this argument and extended the development of an ensemble learning model for four methods: Artificial Neural Networks (ANN), SVM, KNN, and Decision Tree. Another study by Francisco et al. (2019) [20] on forecasting the volatility of the financial stability index of European banks highlighted that the Extreme Gradient Boosting (XGBoost) method performed well in classification models aimed at identifying which variables should be monitored to predict a bank's financial distress, thereby predicting the likelihood of bank default.

Overall, prior research suggests that forecasting with ML algorithms yields more promising results compared to traditional methods. A key implication is the superior performance of ML solutions, including faster processing, efficiency with large datasets, predictions based entirely on collected real-world data, and the ability to detect data-specific issues compared to traditional statistical methods. In this study, the Z-score (target variable) is continuous, thus suitable ML algorithms include Linear Regression, Random Forest, K-Neighbors, Decision Tree, Gradient Boosting, AdaBoost, Support Vector Regression, and XGBoost. The characteristics and concepts of each algorithm are summarized below:

Linear Regression is a statistical model used to determine the linear relationship between a dependent variable (target variable) and one or more independent variables (explanatory variables). The goal of Linear Regression is to find the best-fit line through the data points that minimizes prediction error.

Decision Tree is a crucial technique in machine learning and data mining. This method creates a predictive model based on training data by generating a decision tree. A decision tree is a tree-like graph with nodes and edges, where each node represents a decision or a test on a data attribute, and each leaf represents a prediction [36].

Random Forest (RF) is a variant of Ensemble Learning that combines multiple Decision Trees to create a more robust and stable model. RF employs random sampling from the training dataset to build each individual tree. Each tree is constructed on a different subset of the data, with some samples potentially overlapping. After building multiple Decision Trees, RF combines the prediction results from all trees to make a final decision, typically by selecting the prediction value chosen by the majority of Decision Trees. This is a crucial statistical technique for estimating and constructing confidence intervals for sample-based statistics without requiring strong assumptions about the data distribution. This technique is particularly useful when the sample size is too small for traditional statistical methods. RF's use of multiple Decision Trees and random sampling helps mitigate overfitting, leading to better predictions on new data, especially tabular data without seasonality or dependence on past time periods. Consequently, RF is widely applied in financial stability forecasting models [5, 17, 32, 33].

K-Neighbors (KNN), also known as the K-Nearest Neighbors method, is an ML algorithm used for prediction based on measuring the distance between data points in a feature space. KNN operates on the assumption that data points close to each other in the feature space typically belong to the same class or group. The algorithm works by finding the K nearest data points (the "neighbors") to the data point needing prediction. It then uses the majority frequency or weights of these points to determine the class or predicted value of that point. KNN is commonly applied in building models to warn of anomalies in financial stability [5, 36].

AdaBoost, short for Adaptive Boosting, is a machine learning technique belonging to the boosting family, which improves the performance of weak learners into a stronger model. Typically, decision stumps are used as weak learners. AdaBoost's training process is sequential: after each iteration, the algorithm adjusts the weights of difficult-to-predict data samples to enhance model accuracy. The weak models are then combined to create a final model with higher accuracy. AdaBoost is often used in both regression and classification problems and is effective with datasets that are not overly complex, while being less prone to overfitting compared to more complex models [5, 17].

Support Vector Machine (SVM) is a widely used ML algorithm in regression problems. Designed to identify the regression relationship between data points, SVM's key feature is data distance optimization, enabling effective generalization on new data. SVM handles both linear and non-linear data using kernel functions to map data into higher-dimensional spaces for better hyperplane separation. It's applied in numerous studies for both regression and classification, notably in bank credit risk warning applications [5, 32, 36].

Proposed by Friedman (2000) [16], *Gradient Boosting* is an enhanced ML algorithm based on the concept of increasing the gradient of decision trees, creating objectively competitive and robust training processes. It uses multiple simple models (typically small decision trees) to form a more complex and accurate predictive model. Gradient Boosting gradually builds a sequence of models, each added to reduce the residual error of the previous one by optimizing along the error function's gradient, minimizing overall prediction error. Consequently, Gradient Boosting achieves high accuracy in classification and regression, suitable for complex and non-linear data.

Extreme Gradient Boosting (XGBoost), an independent machine learning algorithm developed by Chen & Guestrin (2016) [37], extends the Gradient Boosting algorithm. Its principle involves training improved new models by combining previous weaker models, primarily decision trees, to create a stronger ensemble. This is achieved by minimizing the error function through multiple training iterations, each new tree correcting errors of its predecessors. XGBoost addresses regression, classification, ranking, and user-defined problems. Its strengths include handling non-linear data, high performance, missing value handling, and sophisticated regularization methods (L1 and L2) to minimize overfitting. Consequently, XGBoost is widely used in early warning models for bank financial instability, demonstrating superior performance compared to other ML algorithms [20, 38, 39].

3- Data and Research Methods

3-1- Data

The research dataset comprises secondary data from 61 banks across six ASIAN countries: Indonesia, Malaysia, Philippines, Singapore, Thailand, and Vietnam, from 2010 to 2023. Bank characteristic data is sourced from financial reports; macroeconomic factors (GDP growth rate, inflation rate) and technological innovation indicators (Commercial bank branches/100,000 people, ATMs/100,000 people, Mobile subscribers/100 people, Internet banking accounts (%/population)) are from Worldbank and IMF. Technological innovation data from Worldbank and IMF is available until 2022; Linear Regression forecasts 2023 data. A Covid-19 dummy variable is included: 0 for pre-Covid years (2010-2019), 1 for Covid-affected years (2020-2023). The data is in balance sheet format, by fiscal year.

3-2- Research Methodology

The general research process is designed as follows (see Figure 5):

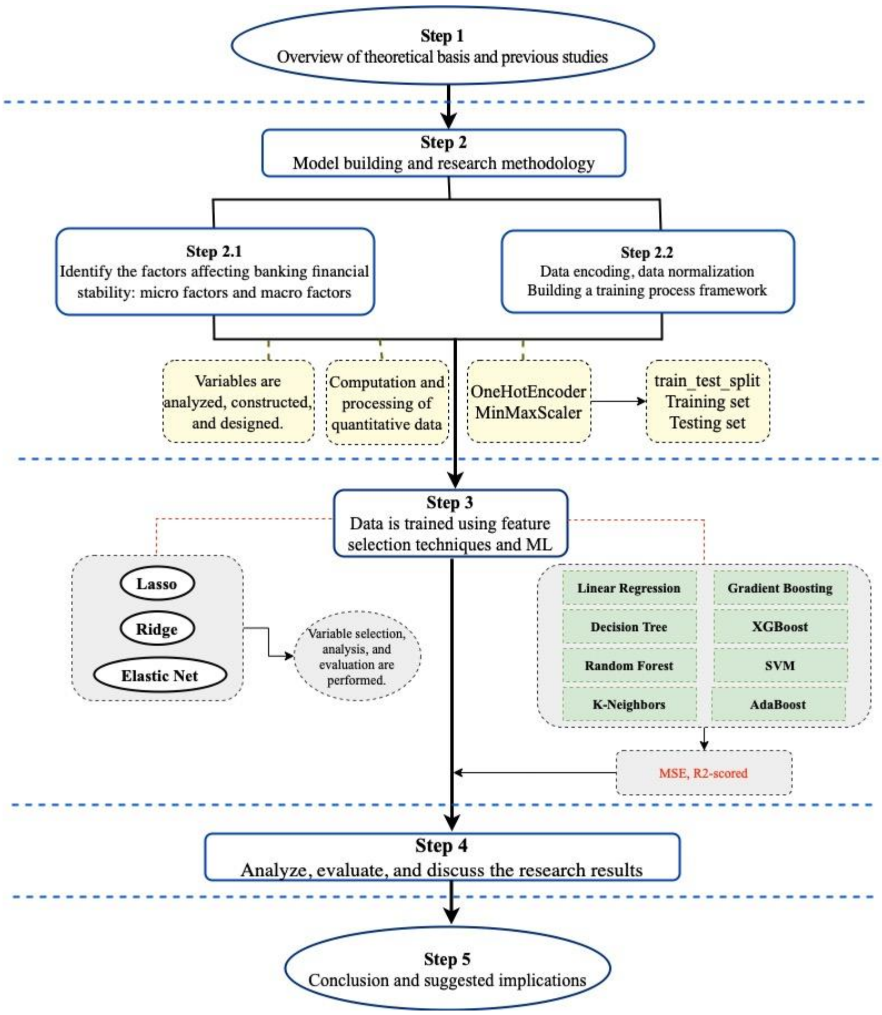

Figure 5. Proposed research process2

3-2-1- Research Model

Building upon previous research [13, 27, 40], the linear regression model has the following general form:

$$STABILITY\_BANK_{it} = \beta_0 + \beta_1\{BANK_{it}\} + \beta_2\{MACRO_t\} + e_{it}. \quad (2)$$

where; $STABILITY\_BANK_{it}$: the Z-$score_{it}$ index evaluates the stability of bank i in year t; $BANK_{it}$: A group of variables characterizing bank i in year t; MACRO: Macroeconomic variables: GDP growth rate, inflation rate, Covid-19 dummy variable, and Technological Innovation Index (TII) of each country over the observed years.

The meaning and calculation method of the factors in the research model (2) are summarized as follows:

Table 1. Calculation method of variables in the research model

Dependent Variable	**Z-score = (ROAA + ETA)/Standard Deviation (ROAA)**		
Independent Variable			
Meaning	**Symbol**	**Formula**	**Sign**
Past Stability Level	ZSCORE _1	= Lag (Z-score)	+
Competitive Capacity (formula from Berger (2009) [27])	LERNER	$= (P_{it} - MC_{it})/P_{it}$ Where: Pit is output price, MC_{it} is marginal cost.	+
Total Asset Size	SIZE	= Ln(Total Assets)	+
Equity Capital Size	ETA	= Equity/Total Assets	+
Loan to Asset Ratio	LTA	= Loan Portfolio/Total Assets	-
Market Share	MS	= Total Assets/Total Assets of the Credit System	+
Income Diversification	DDHI	$=1 - [(NET/(NII))^2 + (NON/(NII))^2]$ where: NON represents non-interest income; NET represents net interest income, and NII is the total income, defined as NII=NON+NET	-
Return on Equity and Average Assets	ROEA	Data collected from annual financial reports	+
Total Asset Growth Rate	GTA	= (Total Assets in current period - Total Assets in previous period) / Total Assets in previous period	-
Financial Leverage	Debt to Equity Ratio (DER)	= Total Debt/Equity	-
	Debt to Asset Ratio (DAR)	= Total Debt/Total Assets	-
Market Concentration	Herfindahl-Hirschman Index of Market Share (HHI_MS)	$=\sum_{i=1}^{n}(MS_i)^2$ • n: Number of banks in each country • MS: Market share of the bank in the national system	+
	Herfindahl-Hirschman Index of Bank Size (HHI_SIZE)	$=\sum_{i=1}^{n}(SIZE_i)^2$ • n: Number of banks in each country • SIZE: Total Asset Size of the bank in the national system	+
GDP Growth Rate	Gross Domestic Product	Data collected from World Bank and IMF sources	+
Inflation Rate	INF		-
Technological Innovation Index (formula from (Lumsden (2018) [30]; Gebregziabher & Makina (2019) [41], with the addition of two factors I2 and I4)	Technological Innovation Index	$TII = \frac{1}{2}\left[\frac{\sqrt{(I_1)^2+(I_2)^2+(I_3)^2+(I_4)^2}}{\sqrt{4}} + \left(1 - \frac{\sqrt{(1-I_1)^2+(1-I_2)^2+(1-I_3)^2+(1-I_4)^2}}{\sqrt{4}}\right)\right]$ *Where:* • I1: Number of commercial bank branches/100,000 people; • I2: Number of ATMs/100,000 people; • I3: Number of mobile subscribers/100 people; • I4: Number of Internet user accounts (%/population).	+
Covid pandemic	COVID-19	= 0: year without the Covid pandemic = 1: year with the Covid pandemic	+/-

3-2-2- Machine Learning Model Training Process Combined with Variable Selection Techniques

The study utilizes cross-validation to combine Lasso, Ridge, and Elastic Net with eight ML algorithms. Based on the MSE, R2-score, and MAE values, the best-performing parameter tuning technique and ML algorithm will be selected to predict the bank financial stability index. The general training process framework is simulated as follows (see Figure 6):

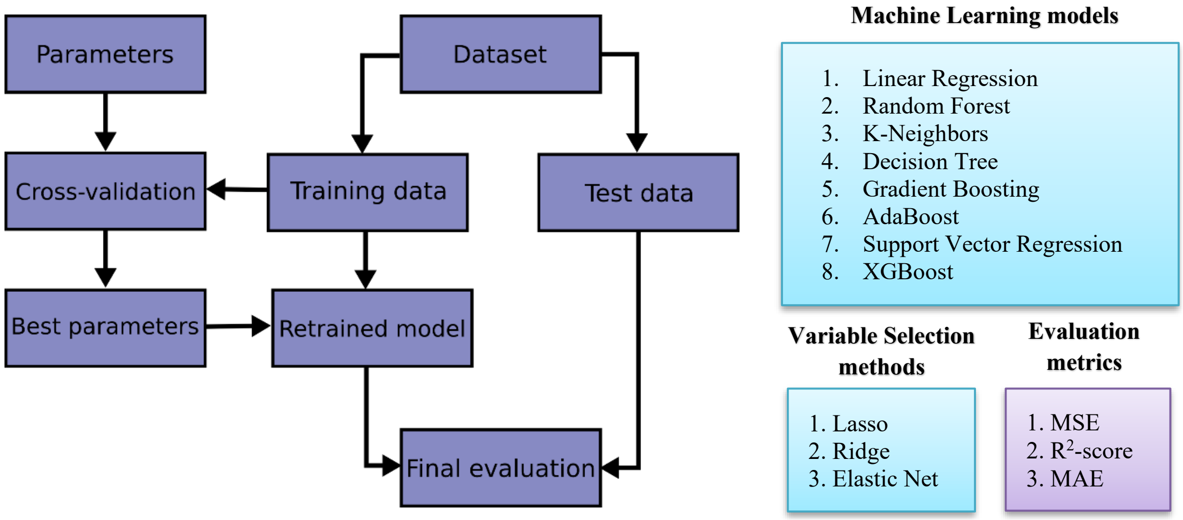

Figure 6. Machine Learning model training process

The study will conduct experimental training with Lasso, Ridge, and Elastic Net methods to determine the most suitable inputs for the research dataset. In the experimental process, the solution to the Lasso, Ridge, or Elastic Net estimation method is carried out according to the test loop mechanism on a set of coefficient values Lagrange - (noted lambda - λ) by cross-validation technique. The study uses a popular technique in statistics, the cross-validation technique combined with Lasso, Ridge, Elastic Net to find the lambda coefficient (λ) at which Lasso, Ridge, Elastic Net achieve the best performance. Cross-validation was employed to determine the optimal regularization parameter (λ) for each method, ensuring that the models achieve a balance between bias and variance. This step is critical for enhancing the generalizability of the results, particularly in the context of ASIAN's dynamic banking environment. The determination of the lambda (λ) coefficient is crucial in the training process of Lasso, Ridge, and Elastic Net models, as it regulates the trade-off between prediction accuracy and model complexity. λ directly influences the strength of regularization, with larger values in Lasso leading to the exclusion of variables by shrinking their coefficients to zero, while Ridge regularization maintains all variables but penalizes large coefficients. Elastic Net combines both approaches, and λ controls the relative contributions of Lasso and Ridge regularization. Achieving optimal λ, often via cross-validation, is essential for building a model that balances simplicity with high predictive performance [29].

This study will evaluate these methods empirically to identify the most relevant variables for the research dataset, with λ determined through iterative adjustments using cross-validation. Model performance will be assessed using metrics such as Mean Squared Error (MSE) and Adjusted R-squared (R^2), followed by a detailed analysis of the results.

4- Results and Discussion

4-1-Model Parameter Tuning Using Lasso, Ridge, and Elastic Net

First, the data is trained sequentially with Lasso, Ridge, and Elastic Net techniques to find the best alpha coefficient (regularization coefficient λ). The results are shown in Table 2.

The empirical results derived from Lasso, Ridge, and Elastic Net regression techniques provide critical insights into the determinants of bank financial stability, measured by the Z-score. The findings highlight the differential influence of various bank-specific and macroeconomic factors, while also underscoring the importance of methodological selection in ensuring robust and reliable predictions. The results consistently indicate that past financial stability (Z-score_1), capital adequacy (ETA), operational efficiency (ROE), financial leverage (DAR, DER), macroeconomic conditions (GDP, INF), technological innovation (TII), and the COVID-19 pandemic exert a significant impact on bank stability. Across all three techniques, these factors retain substantial coefficients, confirming their role in shaping financial resilience. Lasso and Elastic Net produce comparable results, effectively identifying variables with minimal influence on bank stability. Factors such as LERNER, SIZE, LTA, MS, GTA, DDHI, HHI_MS, and HHI_SIZE exhibit near-zero coefficients, suggesting their limited explanatory power and potential redundancy in predictive modeling. This underscores the necessity of dimension reduction techniques in enhancing model efficiency. However, Ridge regression differs in its approach by retaining all variables, leading to slightly improved performance. The R-squared values

(76.23% for Lasso, 76.85% for Ridge, and 75.40% for Elastic Net) indicate that Ridge regression achieves the highest explanatory power, albeit marginally. Similarly, the Mean Squared Error (MSE) is lowest for Ridge (0.00288), confirming its superior predictive accuracy.

Table 2. Estimated coefficient results using Lasso, Ridge, and Elastic Net

Features	Lasso_Coefficients	Ridge_Coefficients	Elastic Net Coefficients
ZSCORE _1	0.2012	0.2077	0.2067
LERNER	**0.0000**	-0.0168	**0.0000**
SIZE	**0.0000**	0.0082	**0.0000**
ETA	0.2121	0.2438	0.2392
LTA	**0.0000**	0.0087	**0.0000**
ROEA	0.1201	0.1595	0.1054
MS	-0.0001	-0.0165	-0.0001
GTA	**0.0000**	-0.0154	**0.0000**
DDHI	**0.0000**	-0.0049	**0.0000**
DAR	0.2058	0.2294	0.1220
DER	-0.3954	-0.3690	-0.3517
GDP	-0.1354	-0.1565	-0.1279
INF	-0.1185	-0.1358	-0.1156
HHI_MS	**0.0000**	0.0139	**0.0000**
HHI_SIZE	0.0038	0.0082	0.0032
TII	-0.0457	-0.0559	-0.0446
COVID-19	-0.0885	-0.0896	-0.0867
Best Alpha	0.00037	**1.0000**	0.00079
Mean Squared Error	0.00297	**0.00288**	0.00307
R-squared	76,23%	**76,85%**	75,40%

The study conducted experiments to observe the change in the model's estimated parameter vector as the coefficient λ varies, obtaining the results (Figures 7 to 9). The trend of the estimated coefficients of the factors in the model using Lasso and Elastic Net methods does not show significant differences.

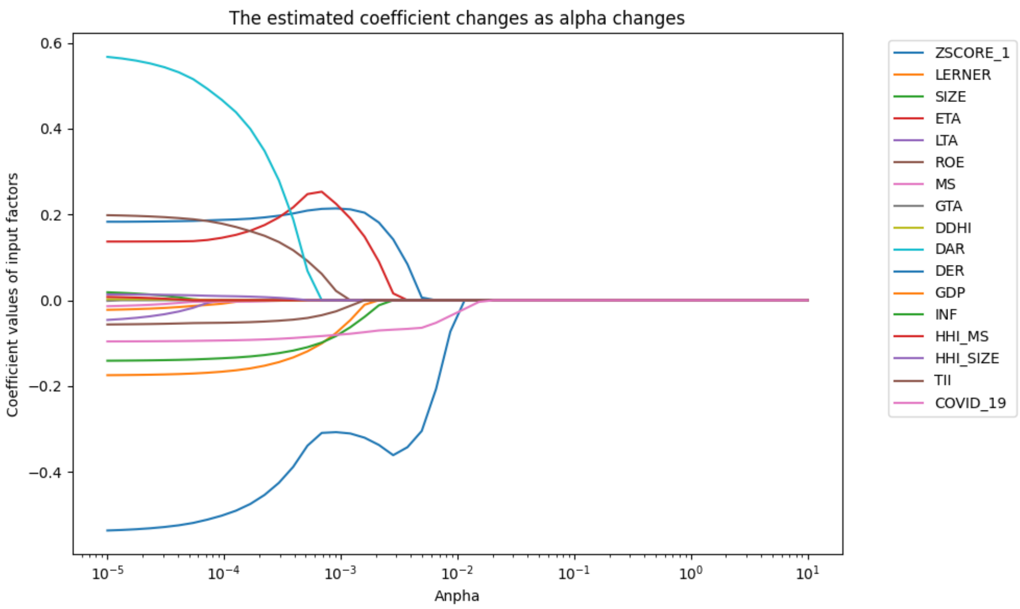

Figure 7. Training results using Lasso

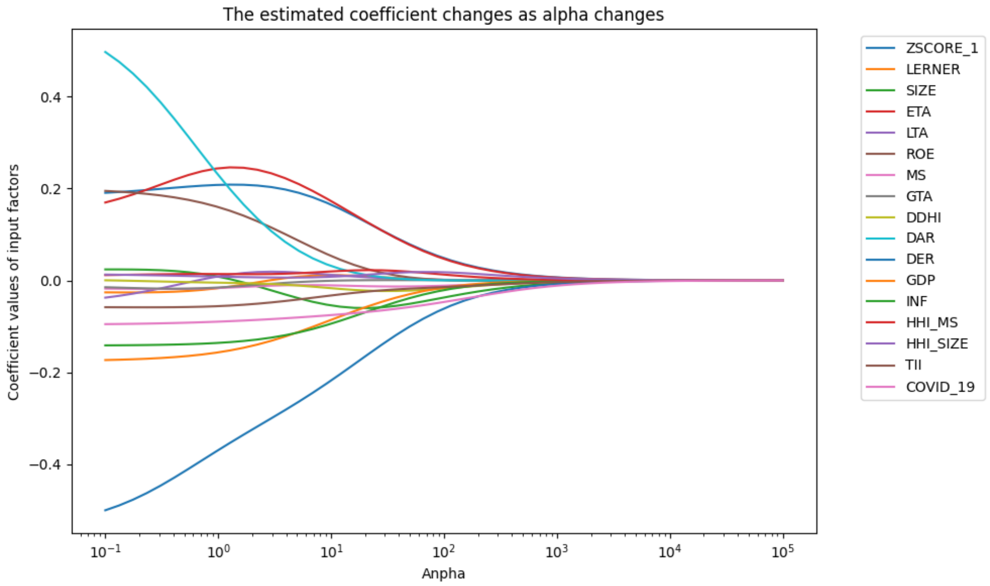

Figure 8. Training results using Ridge

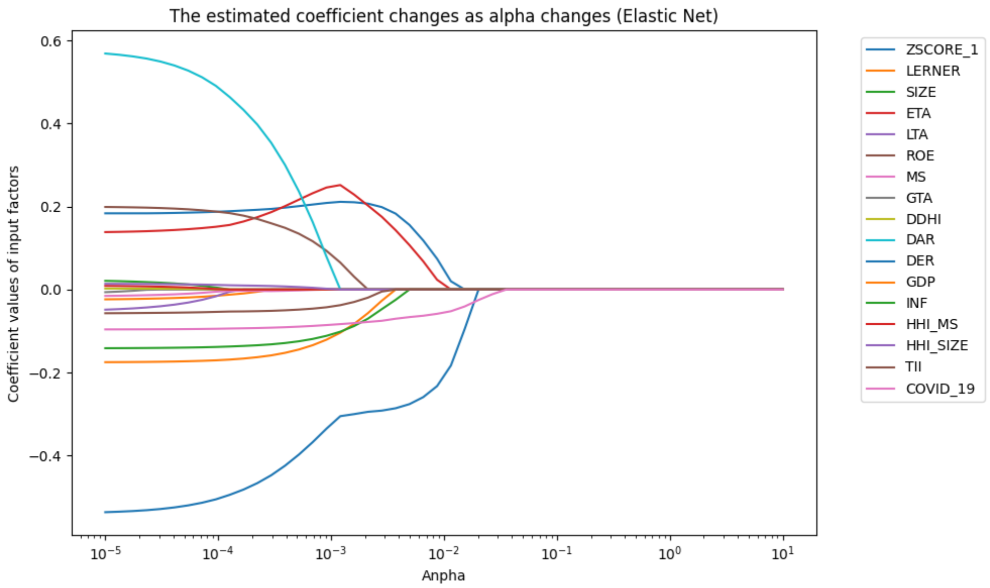

Figure 9. Training results using Elastic Net

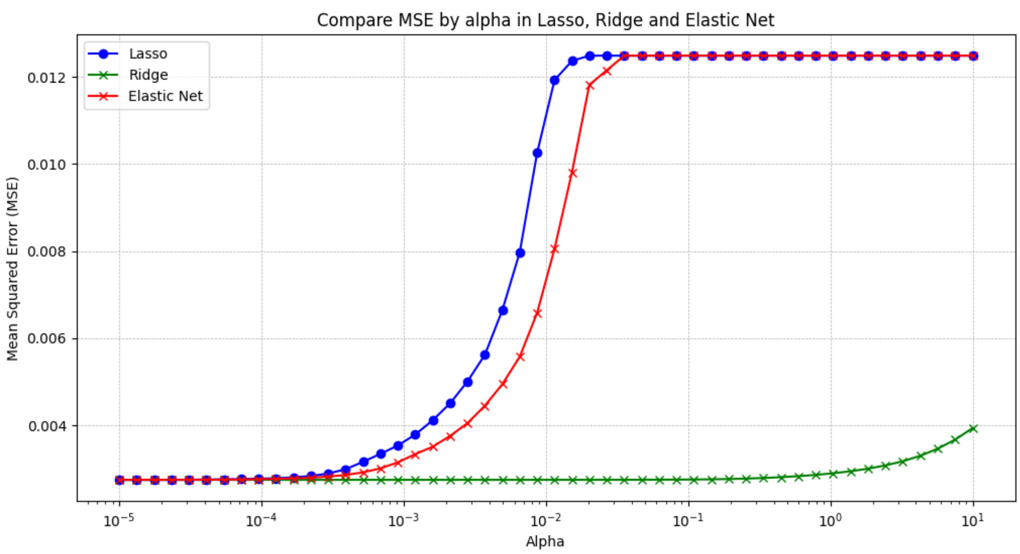

Figure 10. Variation of MSE with changing alpha coefficient

Figure 10 compellingly illustrates the superior stability and predictive performance of Ridge regression compared to Lasso and Elastic Net, as evaluated through MSE. The Ridge regression curve (green) maintains consistently low MSE values across all regularization strengths (α), underscoring its resilience and robustness against parameter variations. In contrast, both Lasso (blue) and Elastic Net (red) exhibit a pronounced escalation in MSE beyond a critical threshold ($\alpha \approx 10^{-2}$), signaling a substantial decline in predictive accuracy due to excessive coefficient shrinkage. This empirical evidence positions Ridge regression as the optimal choice for scenarios prioritizing model stability and predictive precision under varying levels of regularization.

Beyond its statistical implications, these findings offer profound insights into the financial stability of emerging economies, particularly within the ASEAN region. The distinct divergence in MSE trends across the three regression methods suggests a strong parallel between model robustness and banking system resilience. Just as Ridge regression sustains reliable performance despite increasing regularization pressure, financial stability in emerging markets hinges on structural resilience to external shocks. Conversely, the sharp deterioration observed in Lasso and Elastic Net, where MSE surges beyond a certain threshold, mirrors the fragility of banking institutions in these economies when subjected to excessive regulatory or macroeconomic stress. This analogy underscores the critical need for a well-calibrated regulatory framework and adaptive financial policies to fortify long-term banking stability in ASEAN and other emerging markets.

4-2- Trends and Impact of Factors on the Bank Financial Stability Index

The results of training the model sequentially with Lasso, Ridge, and Elastic Net techniques are as follows (Figures 11 to 13):

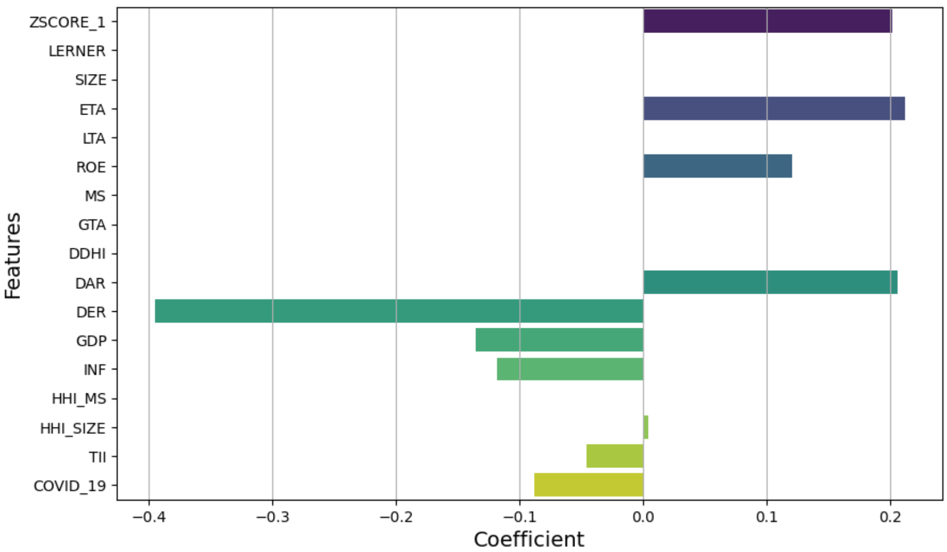

Figure 11. Trends and influence of factors on Z-score (Lasso)

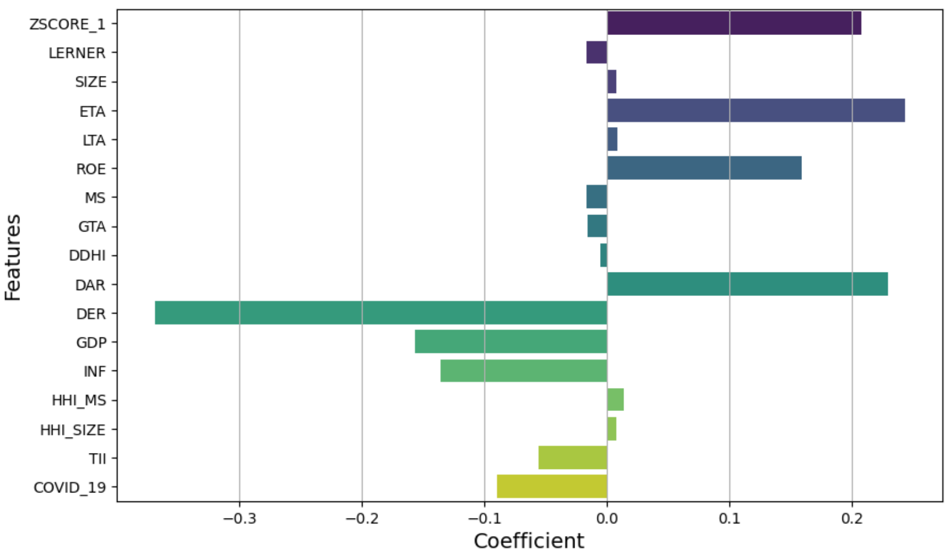

Figure 12. Trends and influence of factors on Z-score (Ridge)

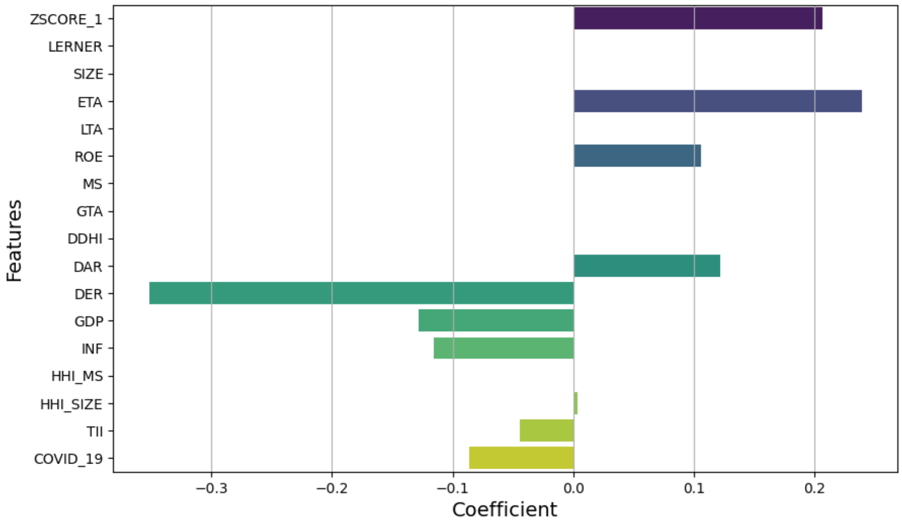

Figure 13. **Trends and influence of factors on Z-score (Elastic Net)**

The empirical findings derived from the three variable selection techniques—Lasso, Ridge, and Elastic Net—consistently highlight the critical determinants of bank financial stability. Specifically, past financial stability (ZSCORE_1), equity scale (ETA), operational efficiency (ROE), financial leverage (DAR, DER), macroeconomic conditions (GDP, INF), technological innovation (TII), and the COVID-19 pandemic exhibit significant influences on the Z-score. Notably, Lasso and Elastic Net yield comparable results, suggesting that variables such as LERNER, SIZE, LTA, MS, GTA, DDHI, HHI_MS, and HHI_SIZE contribute negligibly or are statistically insignificant in determining financial stability. Consequently, these variables may be excluded to enhance model interpretability and efficiency. However, the analysis also reveals that Ridge regression outperforms the other two methods in terms of predictive accuracy and model robustness. This raises an important consideration: while removing non-significant variables can simplify the model, retaining all variables—despite their minimal impact—may enhance predictive stability, particularly when using Ridge regression. To address this, cross-validation will be employed to rigorously re-evaluate the parameter tuning effectiveness of all three techniques (Lasso, Ridge, and Elastic Net) across multiple machine learning algorithms. Ultimately, the study aims to identify the optimal variable selection approach tailored to the research dataset, ensuring both predictive accuracy and practical applicability in assessing banking stability.

Additionally, the impact trends of both micro and macro factors on the Z-score need attention as empirical evidence for assessing the positive or destabilizing effects of these factors on bank financial stability. According to the detailed results in Table 2, Figures 11 to 13, the impact trends of these factors on the Z-score value are as follows:

- *Past stability* (ZSCORE_1) exhibits a strong positive correlation with current financial stability across all three models, reaffirming the path dependency of banking resilience. This finding aligns with previous studies, which highlight that a stable financial position serves as a foundation for future stability and buffers against external shocks [12, 13, 42]. It underscores the necessity for banks to build and maintain robust financial health over time to withstand market volatility.
- *Competitive capacity* (LERNER) representing market power, shows a negative but minor impact on the Z-score. This suggests that heightened competition may erode profit margins, intensifying risk-taking behavior, as indicated in prior research [11, 27]. While competitive strength can enhance efficiency and innovation, excessive competition may drive banks toward riskier strategies, compromising long-term stability [27, 42].
- *Total asset size* (SIZE) exhibits a minor positive effect on financial stability, supporting the "too-big-to-fail" hypothesis, where larger banks benefit from economies of scale and greater access to capital buffers [11, 13, 14, 42]. However, the findings echo concerns from Berger et al. (2009) that larger banks may pursue high-risk investments, potentially heightening instability [27].
- *Equity capital* (ETA) and return on equity (ROE) emerge as pivotal factors for financial stability, with consistently strong positive coefficients across all models. This aligns with the literature emphasizing the role of capital as a safeguard against financial distress [12, 19, 29, 40, 42]. Nevertheless, high capital reserves

might incentivize excessive investment, creating financial strain if rapid recovery proves challenging [40]. Likewise, high ROE reflects sound financial performance, reducing systemic pressure and enhancing resilience [40].

- *Loan portfolio size* (LTA) shows a minor positive correlation with the Z-score, indicating a limited impact. According to Beck et al. (2013) [11], excessive lending with ineffective recovery can lead to financial losses, pressuring the bank. This contrasts with Tu et al. (2021) [13] and Tin et al. (2023) [12] in their studies on Vietnam. However, Tu et al. (2021) only examined the relationship between lending effectiveness and financial stability without considering the COVID-19 context, which is considered to increase bad debts, pressuring the overall economy and the banking sector [13, 24].
- *Market share* (MS) and *total asset growth rate* (GTA) show negative estimated coefficients and have a minor impact on the Z-score. While bank size can be a prerequisite for maintaining financial stability against economic shocks, excessive asset growth or portfolio expansion can increase risks and financial pressure [13, 27, 40]. Therefore, asset growth needs controlled and strict management.
- *Financial leverage* presents a nuanced impact on stability: the debt-to-equity ratio (DER) shows a positive correlation with risk, indicating that higher leverage intensifies vulnerability. In contrast, the debt-to-assets ratio (DAR) exhibits a stabilizing effect, suggesting that asset growth can mitigate leverage risks, provided asset quality is maintained. This dual effect highlights the critical need for balanced leverage strategies to optimize risk-return trade-offs.
- *Market concentration* (HHI_MS and HHI_SIZE) is positively correlated with the Z-score. This finding is consistent with the research of Noman et al. (2017) [7] and Verma & Chakarwarty (2024) [43] on ASEAN countries, suggesting that concentrated markets may provide stability benefits by reducing competitive pressure and increasing pricing power.
- *Income diversification* (DDHI) results show that, in the context of recent research in ASIAN countries, income diversification activities are negatively correlated with the level of financial stability in banks. This is consistent with the research of Tin et al. (2023), Tu et al. (2021), and Wu (2020) [12-14], which states that for emerging market economies, especially those affected by Covid-19, diversification activities, if not well controlled or not aligned with actual financial capacity, can cause financial pressure, reduce operational efficiency, increase competitive pressure on banks, and decrease operating performance, leading to bank instability.
- *The macroeconomic landscape*—encompassing GDP growth, inflation, technological innovation, and the Covid-19 pandemic—exerts a profound influence on banking stability, with a strong negative correlation to the Z-score. This aligns with prior research and underscores a critical economic reality: while GDP growth and inflation drive economic expansion, they also introduce substantial risks to financial stability [11, 27]. Rooted in Minsky's Financial Instability Hypothesis (1992), the findings reveal how rapid economic expansion fuels excessive risk-taking, credit overextension, and financial fragility, culminating in instability [44]. Empirical evidence reinforces this: Beck et al. (2013) [11] associate high GDP growth in emerging markets with credit risk and asset bubbles, while Noman et al. (2017) [7] and Wu (2020) [14] highlight inflation's erosive impact on bank assets and its role in amplifying uncertainty and default rates. These trends echo Stiglitz & Weiss's (1981) [45] adverse selection and moral hazard theories, demonstrating how economic volatility distorts risk assessment and credit allocation. The vulnerabilities are even more acute in emerging markets like ASEAN, where underdeveloped financial systems and weaker regulatory frameworks magnify susceptibility to external shocks [8, 20]. The Covid-19 crisis further intensified financial pressures, particularly through costly yet slow-return technological investments [12, 24]. These findings underscore an urgent need for counter-cyclical policies, macroprudential regulations, and inflation-targeting frameworks to fortify banking resilience in an unpredictable economic environment.

4-3- Training with Machine Learning Algorithms

The study uses Cross-Validation techniques combined with parameter tuning methods (Lasso, Ridge, Elastic Net) with 8 Machine Learning algorithms to train the dataset. Concurrently, the study also trains the dataset without using Lasso, Ridge, or Elastic Net (Original) to compare the training effectiveness of ML algorithms with and without parameter adjustments. The detailed results are as follows (Table 3):

Table 3. Training results combining variable selection techniques and Machine Learning

1Machine Learning Algorithms		Model Performance	Lasso	Ridge	Elastic Net	Original
Linear Regression	Training set	*MSE*	0.0027	0.0027	0.0028	0.0044
		R-squared	0.7910	0.7925	0.7899	0.6620
	Testing set	*MSE*	0.0029	0.0029	0.0029	0.0045
		R-squared	0.7075	0.7054	0.7071	0.5438
Random Forest	Training set	*MSE*	0.0004	0.0005	0.0004	0.0030
		R-squared	0.9693	0.9652	0.9705	0.9478
	Testing set	*MSE*	0.0020	0.0023	0.0023	0.0050
		R-squared	0.8005	0.7714	0.7705	0.4936
KNN	Training set	*MSE*	0.0019	0.0025	0.0018	0.0030
		R-squared	0.8585	0.8089	0.8654	0.7703
	Testing set	*MSE*	0.0028	0.0040	0.0026	0.0057
		R-squared	0.7138	0.5980	0.7346	0.4283
Decision Tree	Training set	*MSE*	0.0000	0.0000	0.0000	0.0000
		R-squared	1.0000	1.0000	1.0000	1.0000
	Testing Set	*MSE*	0.0028	0.0035	0.0063	0.0087
		R-squared	0.7167	0.6481	0.3639	0.1249
Gradient Boosting	Training set	*MSE*	0.0006	0.0005	0.0006	0.0023
		R-squared	0.9528	0.9607	0.9514	0.8252
	Testing set	*MSE*	0.0015	0.0012	0.0020	0.0048
		R-squared	0.8523	0.8777	0.8019	0.5225
AdaBoost	Training set	*MSE*	0.0033	0.0029	0.0029	0.0042
		R-squared	0.7448	0.7747	0.7821	0.6781
	Testing set	*MSE*	0.0041	0.0032	0.0035	0.0049
		R-squared	0.5856	0.6801	0.6455	0.5088
Support Vector Machine	Training set	*MSE*	0.0034	0.0035	0.0035	0.0043
		R-squared	0.7389	0.7362	0.7301	0.6744
	Testing set	*MSE*	0.0039	0.0037	0.0039	0.0045
		R-squared	0.6126	0.6251	0.6130	0.5498
XGBoost	Training set	*MSE*	0.0000	0.0000	0.0000	0.0000
		R-squared	0.9998	0.9999	0.9999	0.9972
	Testing set	*MSE*	0.0012	0.0011	0.0019	0.0053
		R-squared	0.8840	0.8889	0.8085	0.4718

The results obtained (detailed in Table 3) after performing Cross-validation combined with Lasso, Ridge, Elastic Net, and 8 ML algorithms show some important findings:

Firstly, the scatter plot (Figure 14) provides compelling evidence of the superior predictive capability of the XGBoost regression model, as demonstrated by the strong alignment of the predicted values with the actual values along the ideal diagonal reference line. The dense clustering of data points near this line indicates a high degree of predictive accuracy, with minimal deviations, underscoring the model's effectiveness in capturing complex relationships within the dataset. The presence of only a few outliers further highlights the robustness and reliability of the model in real-world applications. Notably, the XGBoost algorithm, when optimized through Ridge parameter tuning, achieves an outstanding predictive performance of nearly 89%, reaffirming its status as one of the most powerful machine learning techniques for structured data regression. This result is consistent with previous studies [20, 38, 39], which have widely recognized XGBoost as a leading method in predictive analytics, particularly in financial modeling and risk assessment. The strong performance exhibited here not only validates the selection of XGBoost for this study but also emphasizes its applicability in high-stakes decision-making environments where precision and stability are paramount.

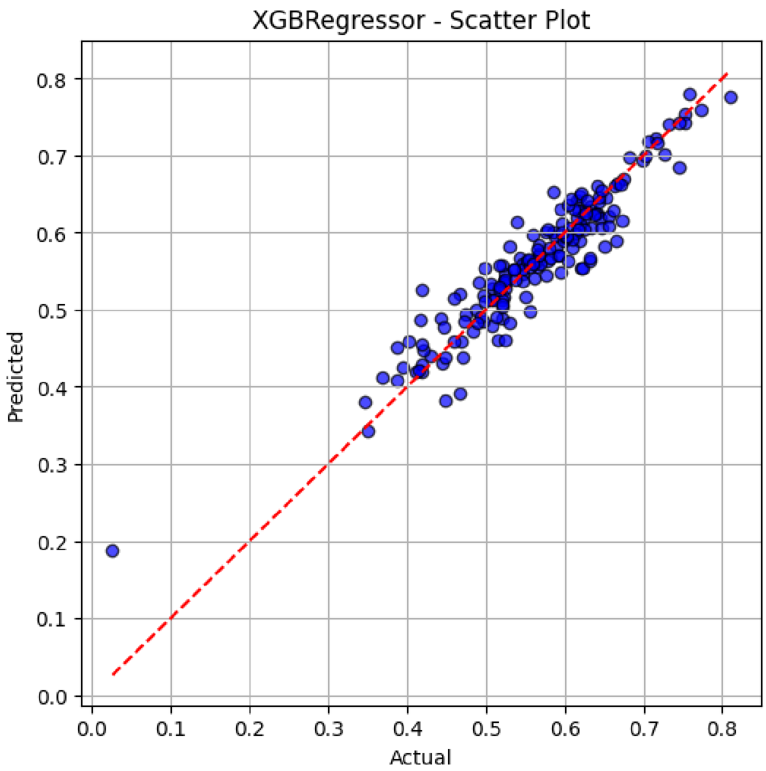

Figure 14. Forecasting effectiveness using XGBoost combined with Ridge

Secondly, the experimental results emphasize that using variable selection parameter tuning techniques like Lasso, Ridge, or Elastic Net significantly improves model performance compared to training solely with ML algorithms. The aggregated statistical results in Figures 15 and 16 show that the predictive performance of each ML model (evaluated based on R^2-score and MSE) combined with parameter tuning techniques is markedly improved compared to conventional training (denoted as Original). Furthermore, empirical evidence suggests that the predictive performance of each ML algorithm varies when combined with Lasso, Ridge, or Elastic Net. For instance, Random Forest combined with Ridge yields the highest MSE, while Gradient Boosting shows the opposite.

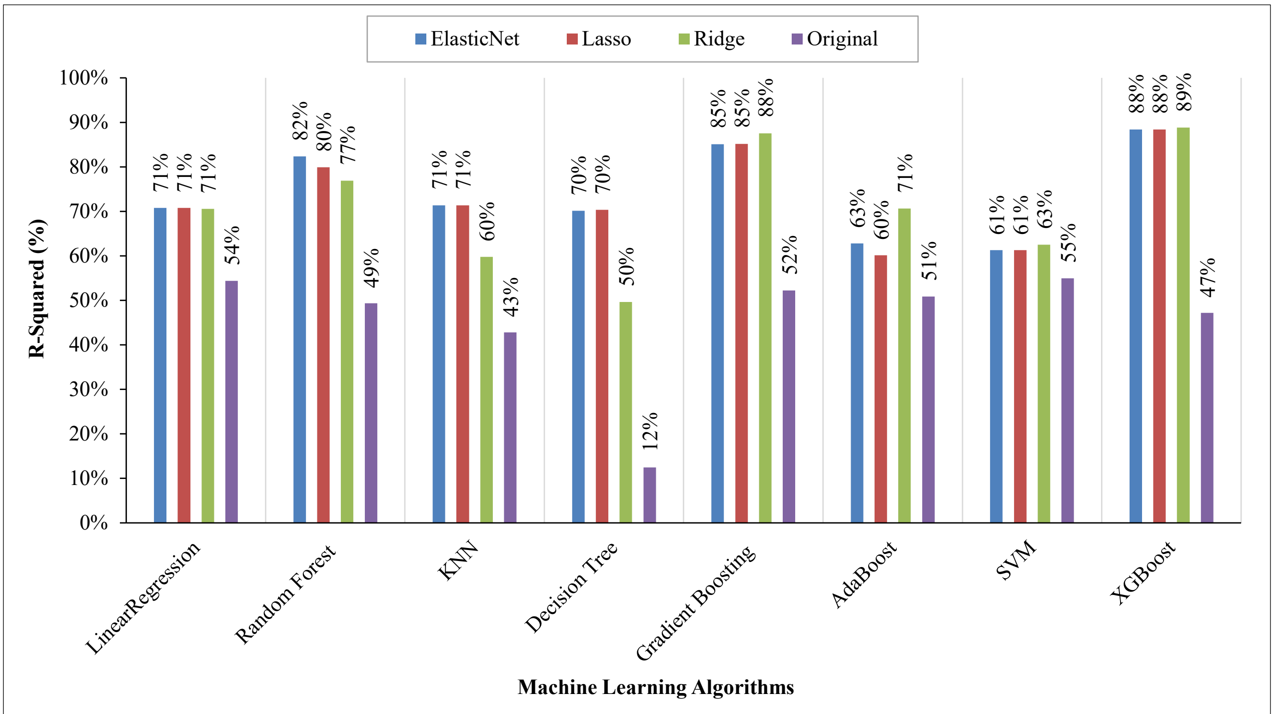

Figure 15. Model forecasting performance evaluated by R-squared

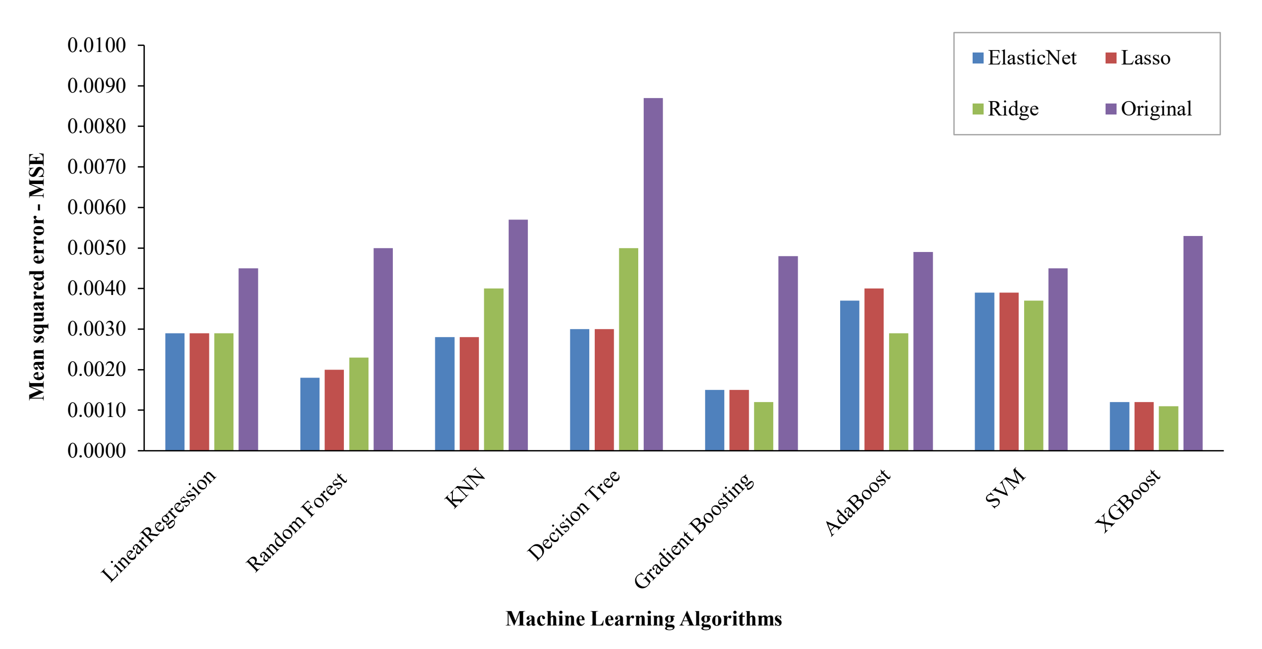

Figure 16. **Model forecasting performance evaluated by MSE**

Thirdly, without applying Lasso, Ridge, and Elastic Net, the ML model performance is inconsistent across both training and testing sets, with significant discrepancies observed.

The empirical findings further underscore the indispensable role of regularization techniques such as Lasso, Ridge, and Elastic Net in enhancing model stability and predictive accuracy. In the absence of these regularization methods, the performance of machine learning models exhibits substantial inconsistencies across both training and testing datasets, leading to significant discrepancies and suboptimal generalization capabilities (Figures 17 and 18). The lack of constraint on coefficient magnitudes results in overfitting to the training data, thereby reducing the model's ability to effectively capture underlying patterns when applied to unseen data. This not only compromises predictive robustness but also amplifies sensitivity to noise and irrelevant features, ultimately degrading model reliability. These findings strongly reaffirm the necessity of employing regularization techniques to mitigate variance, improve generalization, and ensure stable performance across diverse data distributions.

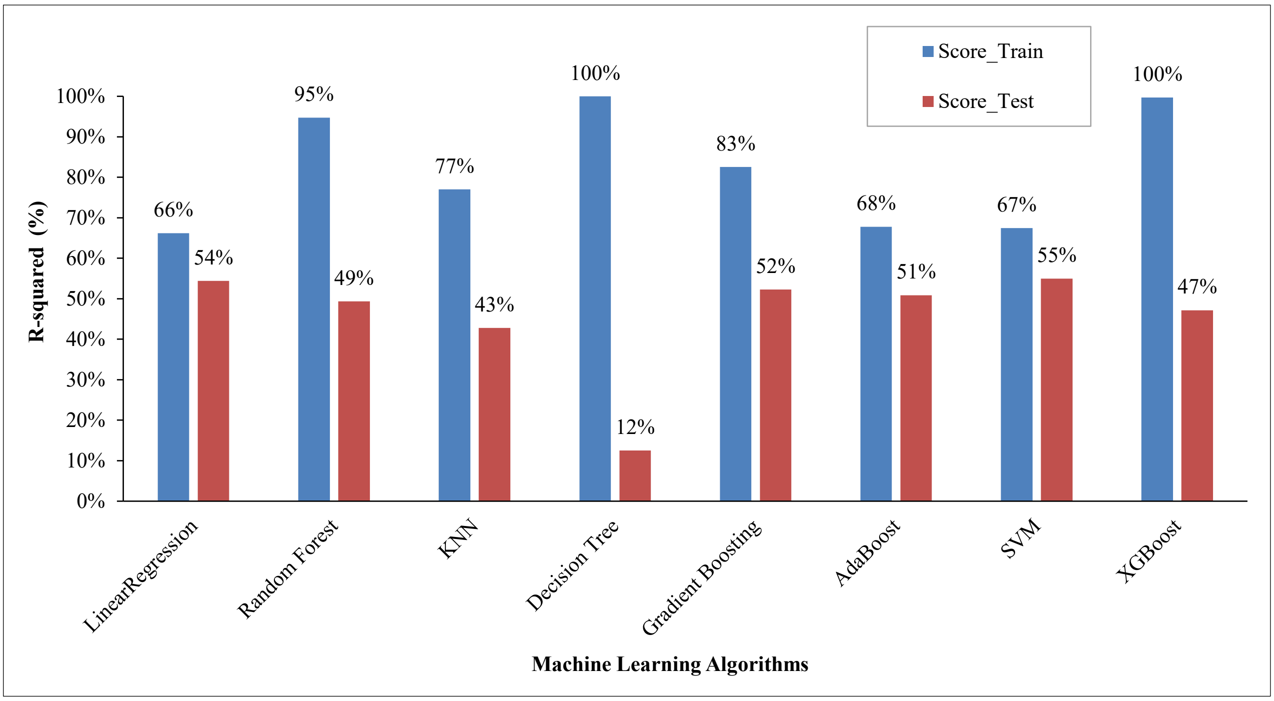

Figure 17. **Original model forecasting performance evaluated by R-squared**

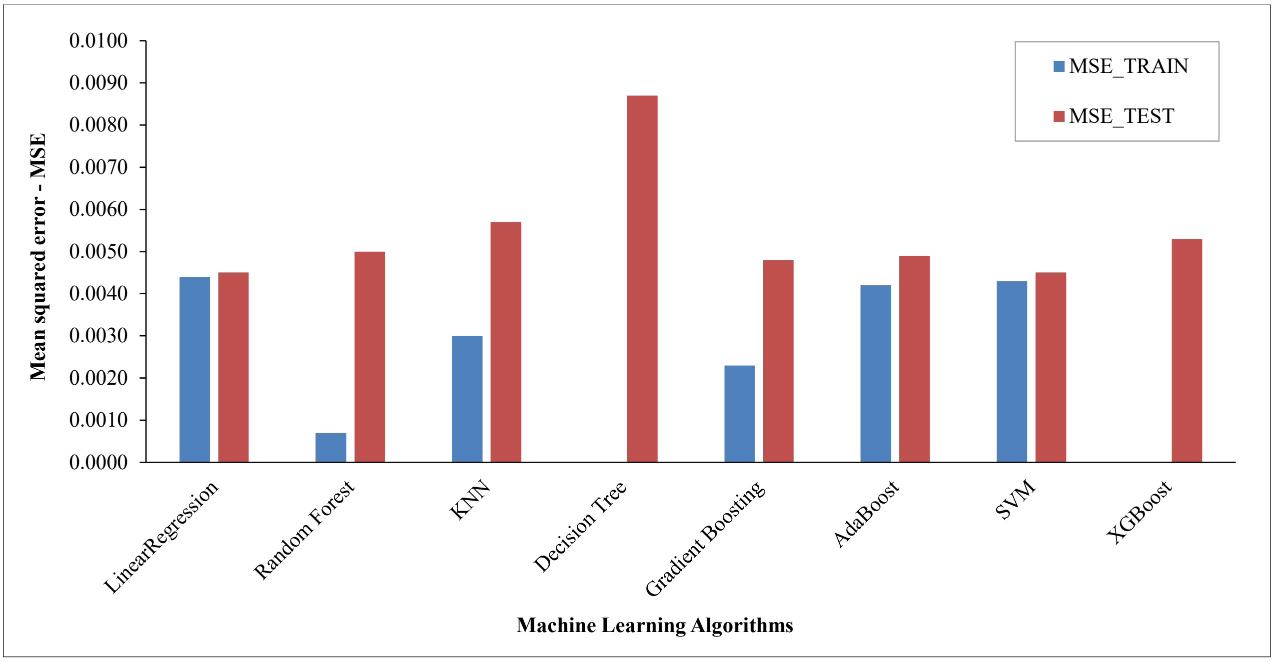

Figure 18. Original model forecasting performance evaluated by MSE

5- Conclusions and Recommendations

This study successfully fulfilled its objectives by systematically identifying key trends and assessing the impact of various factors on bank financial stability index forecasting in ASEAN economies. The empirical findings underscore the effectiveness of integrating Lasso, Ridge, and Elastic Net techniques to enhance the predictive performance of machine learning (ML) algorithms. Notably, among these methodologies, the Ridge regression technique demonstrates superior performance in optimizing parameter estimation, thereby yielding the most accurate forecasts. Moreover, combining the Ridge method with the XGBoost algorithm results in the best-performing predictive model, achieving an R^2-score of approximately 89%. These insights provide robust empirical evidence that can serve as a valuable reference for future research while offering actionable recommendations for bank executives and policymakers striving to strengthen financial stability.

Based on these findings, the study presents the following key recommendations:

Firstly, leveraging empirical insights for strategic banking stability management. The study provides comprehensive evidence on how microeconomic and macroeconomic factors influence bank financial stability. Specifically, variables such as Z-score_1, ETA, ROE, and financial leverage (DAR, defined as total debt over total assets) exhibit a positive correlation with the Z-score and should be reinforced to enhance stability. Conversely, financial leverage (DER, measured as total debt over equity) and macroeconomic indicators such as GDP growth rate, inflation (INF), trade integration index (TII), and exogenous shocks like the COVID-19 pandemic have a detrimental impact on banking stability. These risk factors necessitate stringent control and regulatory oversight. Additionally, although variables such as market competition (LERNER), bank size (SIZE), loan portfolio composition (LTA), market structure (MS), income diversification (DDHI), total asset growth (GTA), and market concentration exhibit a marginal impact on the Z-score, they require consistent monitoring to prevent systemic vulnerabilities. While banks can actively adjust microeconomic factors within their operational framework, macroeconomic determinants necessitate coordinated efforts with regulatory authorities. Encouraging financial innovation, efficiency enhancement, and strategic adaptability is paramount in fostering a resilient banking sector within developing economies.

Secondly, optimizing predictive modeling approaches in financial stability assessment. The study highlights the importance of methodological rigor in selecting and refining predictive models. While the Ridge method delivers the highest forecasting accuracy, the inclusion of all input variables poses the risk of overfitting and information redundancy. Therefore, variable selection should be meticulously evaluated in alignment with regulatory frameworks, national policies, and the specific banking landscape of each country. Future research should consider hybrid approaches that integrate statistical inference with domain expertise, ensuring that model adjustments are contextually appropriate and enhance predictive reliability. Additionally, reliance solely on statistical probability is insufficient; expert consultation remains indispensable in refining model robustness and ensuring practical applicability in real-world banking environments.

Thirdly, integrating advanced machine learning techniques with expert-driven oversight. The findings reaffirm the growing relevance of sophisticated ML methodologies such as Lasso, Ridge, and Elastic Net in developing predictive and early warning models for financial stability. However, while technological advancements offer significant improvements in predictive accuracy, overdependence on algorithmic solutions without human oversight presents risks. Effective implementation necessitates a balanced approach that incorporates expert judgment from both banking professionals and data scientists. Moreover, the expansion of predictive modeling frameworks through Deep Learning techniques and Big Data analytics is recommended to enhance model adaptability, reliability, and accuracy in dynamic financial environments.

Finally, reinforcing macroeconomic policy frameworks to sustain banking stability. The long-term stability of the banking sector is intrinsically linked to the broader economic environment and the efficacy of macroeconomic governance. A stable and resilient banking system can only flourish within a well-structured economic and regulatory framework. Thus, governments and regulatory bodies must continuously refine legal frameworks, enforce prudent risk management policies, and align national economic development strategies with financial sector stability objectives. Strengthening institutional governance, fostering financial transparency, and enhancing risk mitigation mechanisms are imperative in safeguarding banking stability within the evolving economic landscape of ASEAN.

In summary, this study underscores the vital interplay between empirical data-driven insights, advanced predictive methodologies, and strategic regulatory policies in ensuring financial stability. By integrating innovative ML techniques with macroeconomic prudence and institutional oversight, policymakers and banking practitioners can collectively drive sustainable financial resilience in developing economies.

5-1- Limitations and Future Research Directions

This study acknowledges several limitations that present opportunities for future research. Banks have not been grouped by country, limiting insights into country-specific stability dynamics. Deep learning algorithms remain unexplored, and threshold values for key determinants have not been established, which could enhance risk assessment frameworks. Additionally, the study does not explicitly examine the impact of macroeconomic factors such as oil prices and monetary policy. Expert consultations in machine learning were not incorporated, which could refine model selection and optimization. The analysis also lacks a country-level breakdown within ASEAN and comparative studies with similar emerging regions like Latin America or Eastern Europe, which could offer broader insights.

Future research will address these gaps by integrating advanced machine learning techniques, expert input, country-specific evaluations, and cross-regional comparisons to refine predictive models and enhance the understanding of banking stability determinants.

6- Declarations

6-1- Author Contributions

Conceptualization, P.T.T., D.L.K.O., and D.D.T.; methodology, P.T.T., D.L.K.O., and D.D.T.; software, P.T.T.; validation, D.L.K.O. and D.D.T.; formal analysis, P.T.T.; investigation, P.T.T. and D.L.K.O.; writing—original draft preparation, P.T.T., D.L.K.O., and D.D.T.; writing—review and editing, P.T.T., D.L.K.O., and D.D.T. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available in the article.

6-3- Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Not applicable.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. Journal of Accounting Research, 4, 71. doi:10.2307/2490171.

[2] Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23(4), 589. doi:10.2307/2978933.

[3] Altman, E. I., & Hotchkiss, E. (2011). Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt, Third Edition. In Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt. John Wiley & Sons, New Jersey, United States. doi:10.1002/9781118267806.

[4] Ben Lahouel, B., Taleb, L., Ben Zaied, Y., & Managi, S. (2024). Financial stability, liquidity risk and income diversification: evidence from European banks using the CAMELS–DEA approach. Annals of Operations Research, 334(1–3), 391–422. doi:10.1007/s10479-022-04805-1.

[5] Thabet, H. H., Darwish, S. M., & Ali, G. M. (2024). Measuring the efficiency of banks using high-performance ensemble technique. Neural Computing and Applications, 36(27), 16797–16815. doi:10.1007/s00521-024-09929-y.

[6] Boubaker, S., Ngo, T., Samitas, A., & Tripe, D. (2024). An MCDA composite index of bank stability using CAMELS ratios and shannon entropy. Annals of Operations Research, 1–24. doi:10.1007/s10479-024-06023-3.

[7] Noman, A. H. M., Gee, C. S., & Isa, C. R. (2017). Does competition improve financial stability of the banking sector in ASEAN countries? An empirical analysis. PLoS ONE, 12(5), 176546. doi:10.1371/journal.pone.0176546.

[8] Sun, X., Yuan, P., Yao, F., Qin, Z., Yang, S., & Wang, X. (2024). Financial Fragility in Emerging Markets: Examining the Innovative Applications of Machine Learning Design Methods. Journal of the Knowledge Economy, 1-22. doi:10.1007/s13132-023-01731-w.

[9] Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. Journal of Banking and Finance, 1(1), 29–54. doi:10.1016/0378-4266(77)90017-6.

[10] Ben Jabeur, S., Stef, N., & Carmona, P. (2023). Bankruptcy Prediction using the XGBoost Algorithm and Variable Importance Feature Engineering. Computational Economics, 61(2), 715–741. doi:10.1007/s10614-021-10227-1.

[11] Beck, T., De Jonghe, O., & Schepens, G. (2013). Bank competition and stability: Cross-country heterogeneity. Journal of Financial Intermediation, 22(2), 218–244. doi:10.1016/j.jfi.2012.07.001.

[12] Ho, T. H., Nguyen, D. T., Luu, T. B., Le, T. D. Q., & Ngo, T. D. (2023). Bank performance during the COVID-19 pandemic: does income diversification help? Journal of Applied Economics, 26(1), 26 1. doi:10.1080/15140326.2023.2222964.

[13] Pham, T. T., Dao, L. K. O., & Nguyen, V. C. (2021). The determinants of bank's stability: a system GMM panel analysis. Cogent Business and Management, 8(1), 1–18. doi:10.1080/23311975.2021.1963390.

[14] Wu, J., Chen, L., Chen, M., & Jeon, B. N. (2020). Diversification, efficiency and risk of banks: Evidence from emerging economies. Emerging Markets Review, 45(3), 1007–1020. doi:10.1016/j.ememar.2020.100720.

[15] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., Ishwaran, H., Knight, K., Loubes, J. M., Massart, P., Madigan, D., Ridgeway, G., Rosset, S., Zhu, J. I., Stine, R. A., Turlach, B. A., Weisberg, S., & Johnstone, I. (2004). Least angle regression. Annals of Statistics, 32(2), 407–499. doi:10.1214/009053604000000067.

[16] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. doi:10.1214/aos/1013203451.

[17] Shrivastava, S., Jeyanthi, P. M., & Singh, S. (2020). Failure prediction of Indian Banks using SMOTE, Lasso regression, bagging and boosting. Cogent Economics and Finance, 8(1), 17259–17269. doi:10.1080/23322039.2020.1729569.

[18] Zaghdoudi, T. (2013). Bank failure prediction with logistic regression. International Journal of Economics and Financial Issues, 3(2), 537–543.

[19] Shi, Y., Charles, V., & Zhu, J. (2024). Bank financial sustainability evaluation: Data envelopment analysis with random forest and Shapley additive explanations. European Journal of Operational Research, 321(2), 614–630. doi:10.1016/j.ejor.2024.09.030.

[20] Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. Journal of Business Research, 101, 885–896. doi:10.1016/j.jbusres.2018.11.015.

[21] Herawati, N., Wijayanti, A., Sutrisno, A., Nusyirwan, & Misgiyati. (2024). The Performance of Ridge Regression, LASSO, and Elastic-Net in Controlling Multicollinearity: A Simulation and Application. Journal of Modern Applied Statistical Methods, 23(2), 1–13. doi:10.56801/Jmasm.V23.i2.4.

[22] Safi, S. K., Alsheryani, M., Alrashdi, M., Suleiman, R., Awwad, D., & Abdalla, Z. N. (2023). Optimizing Linear Regression Models with Lasso and Ridge Regression: A Study on UAE Financial Behavior during COVID-19. Migration Letters, 20(6), 139–153. doi:10.59670/ml.v20i6.3468.

[23] Radchenko, N., Rubtsova, N., Chkan, I., & Yakysheva, I. (2019). Methodical approaches to implementation of financial bank stability. In Modern Development Paths of Agricultural Production: Trends and Innovations, 547–559. doi:10.1007/978-3-030-14918-5_54.

[24] Yin, F., Jiao, X., Zhou, J., Yin, X., Ibeke, E., Iwendi, M. G. P., & Biamba, C. (2022). Fintech application on banking stability using Big Data of an emerging economy. Journal of Cloud Computing, 11(1). doi:10.1186/s13677-022-00320-7.

[25] Hafeez, B., Li, X., Kabir, M. H., & Tripe, D. (2022). Measuring bank risk: Forward-looking z-score. International Review of Financial Analysis, 80, 1020–1039. doi:10.1016/j.irfa.2022.102039.

[26] Arzova, S. B., & Sahin, B. S. (2024). The effect of financial soundness variables on bank performance: a macro-level analysis in MSCI Emerging Market Index countries. Kybernetes, 53(8), 2605–2623. doi:10.1108/K-02-2023-0237.

[27] Berger, A. N., Klapper, L. F., & Turk-Ariss, R. (2009). Bank competition and financial stability. Journal of Financial Services Research, 35(2), 99–118. doi:10.1007/s10693-008-0050-7.

[28] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 67(2), 301–320. doi:10.1111/j.1467-9868.2005.00503.x.

[29] Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society. Series B: Methodological, 58(1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

[30] Lumsden, E. (2018). The Future Is Mobil: Financial Inclusion and Technological Innovation in the Emerging World. Stanford Journal of Law, Business & Finance, 23(1), 1-44.

[31] Puli, S., Thota, N., & Subrahmanyam, A. C. V. (2024). Assessing Machine Learning Techniques for Predicting Banking Crises in India. Journal of Risk and Financial Management, 17(4), 141. doi:10.3390/jrfm17040141.

[32] Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, 405–417. doi:10.1016/j.eswa.2017.04.006.

[33] Fernández Fernández, J. A. (2020). United States banking stability: An explanation through machine learning. Banks and Bank Systems, 15(4), 137–149. doi:10.21511/bbs.15(4).2020.12.

[34] Alessi, L., & Detken, C. (2018). Identifying excessive credit growth and leverage. Journal of Financial Stability, 35, 215–225. doi:10.1016/j.jfs.2017.06.005.

[35] Tanaka, K., Kinkyo, T., & Hamori, S. (2016). Random forests-based early warning system for bank failures. Economics Letters, 148, 118–121. doi:10.1016/j.econlet.2016.09.024.

[36] Holopainen, M., & Sarlin, P. (2017). Toward robust early-warning models: a horse race, ensembles and model uncertainty. Quantitative Finance, 17(12), 1933–1963. doi:10.1080/14697688.2017.1357972.

[37] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 785–794. doi:10.1145/2939672.2939785.

[38] Tan, B., Gan, Z., & Wu, Y. (2023). The measurement and early warning of daily financial stability index based on XGBoost and SHAP: Evidence from China. Expert Systems with Applications, 227, 1–17. doi:10.1016/j.eswa.2023.120375.

[39] Pham, X. T. T., & Ho, T. H. (2021). Using boosting algorithms to predict bank failure: An untold story. International Review of Economics and Finance, 76, 40–54. doi:10.1016/j.iref.2021.05.005.

[40] Nguyen, D. T., Le, T. D. Q., & Tran, S. (2024). Competition and bank financial stability: evidence from an emerging economy. Cogent Business and Management, 11(1), 1–16. doi:10.1080/23311975.2024.2365422.

[41] Gebregziabher Gebrehiwot, K., & Makina, D. (2019). Macroeconomic Determinants of Financial Inclusion: Evidence Using Dynamic Panel Data Analysis. Extending Financial Inclusion in Africa, 167–191. doi:10.1016/B978-0-12-814164-9.00008-6.

[42] Nguyen, T. D., & Du, Q. L. T. (2022). The effect of financial inclusion on bank stability: Evidence from ASEAN. Cogent Economics and Finance, 10(1), 1–15. doi:10.1080/23322039.2022.2040126.

[43] Verma, D., & Chakarwarty, Y. (2024). Impact of bank competition on financial stability-a study on Indian banks. Competitiveness Review, 34(2), 277–304. doi:10.1108/CR-07-2022-0102.

[44] Toporowski, J. (2020). The financial instability hypothesis. Credit and Crisis from Marx to Minsky, *74*. doi:10.4337/9781788972154.00022.

[45] Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. The American economic review, 71(3), 393-410.