American Journal of Financial Technology and Innovation (AJFTI)

ISSN: 2996-0975 (ONLINE)

VOLUME 3 ISSUE 1 (2025)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Financial Technology and Innovation (AJFTI)

Volume 3 Issue 1, Year 2025
ISSN: 2996-0975 (Online)
DOI: https://doi.org/10.54536/ajfti.v3i1.4570
https://journals.e-palli.com/home/index.php/ajfti

Financial Distress Multi-Classification Prediction: A Case Study in Vietnam

Tram Thi Hoai Vo[1], Pai-Chou Wang[2*]

Article Information

Received: February 15, 2025

Accepted: March 19, 2025

Published: July 26, 2025

Keywords

Adaptive Synthetic Sampling, Financial Distress, Multi-Classification

ABSTRACT

Stock investment remains one of the most attractive and profitable activities in financial markets. However, assessing a company's financial health is a complex task due to the vast amount of financial data involved. This study classifies companies into three financial status categories of safe, risky, and distressed by employing three key financial distress measures: Distance to Default (DD), Emerging Market Score (EMS), and Altman Z-score. A set of 68 financial ratios is utilized to predict the financial status of a company. We employ the Adaptive Synthetic Sampling (ADASYN) technique alongside advanced machine learning algorithms, including Random Forest, CatBoost, XGBoost, and Support Vector Machine to further improve model performance. Our results show that Random Forest yields highly accurate predictions in multi-class classification by integrating machine learning with the EMS method. The best-performing model achieves an exceptional ROC-AUC score of 99.26%. These findings provide a powerful decision-making tool for investors, traders, practitioners, and policymakers, enabling more precise assessments of corporate financial stability.

INTRODUCTION

Vietnam's stock market operates through two primary exchanges: the Hanoi Stock Exchange (HNX) and the Ho Chi Minh Stock Exchange (HOSE). While Vietnam remains a promising frontier market, it also presents significant risks. According to a report from Vietnam.vn (Triều, 2025), the number of listed enterprises on HNX has been declining, with an increasing number of firms being delisted due to stricter regulatory oversight and sanctions against violations. In the first ten months of 2024 alone, 15 companies were delisted, and 22 deregistered for trading. This trend raises concerns about transparency and financial stability, as companies often do not fully disclose the true state of their financial health in financial statements such as balance sheets, income statements, and cash flow reports. Traditional financial metrics like Price-to-Earnings (P/E) ratio, Price-to-Book (P/B) ratio, and Earnings Per Share (EPS) are commonly used to evaluate company performance. However, these indicators may not provide a reliable assessment of a company's financial health, especially in a market where institutional trading volume significantly influences stock prices like Vietnam. During financial crises or major economic events, institutional investors can drive sharp price swings, making it even more challenging for retail investors to make informed decisions.

Several methodologies have been developed to assess a company's financial distress, with Distance to Default (DD), Emerging Market Score (EMS), and Altman's Z-score being among the most popular. However, these methods differ in their analytical approach and effectiveness across varying market conditions and time periods. EMS and Altman's Z-score are accounting-based models, relying on financial statements to assess a company's solvency and financial stability. These methods evaluate distress risk based on historical financial data, providing insights into a firm's liquidity, profitability, and leverage. In contrast, the Distance to Default (DD) is a measurement metric based on the stock market, incorporating real-time market variables such as stock prices and volatility to estimate the probability of default. Unlike accounting-based models, which rely on past performance, Distance to Default (DD) reflects current market sentiment and forward-looking risk assessments, making it particularly valuable in rapidly changing financial environments. Each method focuses on different aspects of financial performance, and their effectiveness varies depending on market conditions and time periods. This study aims to address this gap by analyzing the predictive power of 68 financial ratios derived from company financial statements. Instead of relying on binary classification (distressed vs. non-distressed) based on interest coverage ratio (IC), as commonly seen in previous research, we introduce a more granular, multi-class prediction framework that categorizes companies into three financial health statuses: Green (Stable), Grey (At-Risk), and Red (Distressed), which leverage from three approaches: Distance to Default (DD), Emerging Market Score (EMS), and Altman's Z-score.

In addition, widely used machine learning models in classification tasks such as Random Forest, XGBoost, CatBoost and Support Vector Machine are applied in this study. Among the tested machine learning algorithms, Random Forest performed the best with an impressive ROC-AUC score of 99.26%. The EMS score emerged as a highly effective predictor of financial distress when integrated with machine learning models, surpassing traditional methods and demonstrating the potential

[1] College of Business, Southern Taiwan University of Science and Technology, Tainan 710301, Taiwan ROC
[2] Department of Information Management, Southern Taiwan University of Science and Technology, Tainan 710301, Taiwan ROC
[*] Corresponding author's e-mail: pwang@stust.edu.tw

of machine learning in financial health prediction. This research makes a significant contribution by identifying an effective method for predicting financial distress in Vietnam and verifying the effectiveness of the EMS score as an early warning indicator in the Vietnamese stock market. The results will provide useful tools for traders, investors, and policymakers, helping them make data-driven decisions. Moreover, our approach enhances existing distress prediction models by leveraging machine learning techniques, which have been underutilized in Vietnam's financial market analysis.

For this study, we collected financial statement data from 34 Vietnamese companies, spanning from their initial public offering (IPO) to 2023. The remaining of this paper is organized is organized as follows: The pertinent literature on predicting financial crisis in Vietnam is reviewed in Section 2. In Section 3, the method is described. The findings are covered in Section 4, and Section 5 provides the key findings, contributions, and further discussion of this study.

LITERATURE REVIEW

Literature includes a variety of studies focused on company bankruptcy or financial distress, aimed at assisting academics, investors, and policymakers in evaluating the stock market. However, most of these studies typically use binary target variables to provide simple "Yes" or "No" outcomes. A research on forecasting financial difficulties and insolvency among Vietnamese listed firms was carried out by Pham Vo Ninh *et al.* (2018). They used accounting, market, and macroeconomic aspects in their study, using a dataset of 800 Vietnamese companies from 2003 to 2016 with 6,736 observations. The effect of market data models, financial statement data models, and external economic variables on the probability of financial distress in Vietnamese enterprises was investigated using a logistic regression model. The results showed that while the leverage ratio shows a positive link with financial difficulty, larger enterprises had a reduced likelihood of default. Furthermore, it was discovered that financial hardship positively correlated with both inflation and short-term Treasury bill interest rates. Tran *et al.* (2022) applied various machine learning algorithms, including artificial neural networks, support vector machines (SVM), logistic regression, decision trees, and random forests to forecast risky credit in public companies in Vietnam over the period from 2010-2021. The dataset consisted of 3,277 observations in which 436 companies (13.3%) were identified as financially distressed, while 2,841 companies (86.7%) were categorized as non-distressed. The outcomes demonstrated that factors such as accounts payable to equity ratio, long-term debt to equity ratio, diluted earnings per share and enterprise value to revenue ratio significantly influenced the predictive outcomes and were largely aligned with established expert knowledge. Tran *et al.* (2023) conducted an examination of financial distress within a cohort of 500 publicly traded firms in Vietnam from 2012 to 2021. The results revealed that merely four financial indicators - Total Equity/Total Liabilities, Total Liabilities/Total Assets, Net Income/Total Assets, EBIT/Total Assets and - were proficient in forecasting financial distress in the Vietnamese context. According to the analysis, Altman Z'-score model is only applicable in Vietnam when financial difficulties is represented by the interest coverage ratio. Nguyen *et al.* (2024) utilized seven distinct analytical approaches - logistic regression, linear discriminant analysis, neural networks, support vector machines, decision trees, random forests, and the Merton model - to assess financial distress among publicly listed enterprises in Vietnam during the period from 2011 to 2021. The research incorporated five elements from Altman's model and nine from Ohlson's model. The variable deemed most critical in both Altman's model and the integrated Altman-Ohlson model was "reat", while "ltat" and "wcapat" emerged as the most significant variables in Ohlson's model. The outcomes further indicated that these models typically exhibited superior performance in forecasting financial distress for larger companies in comparison to smaller entities and demonstrated greater accuracy during periods of economic expansion than during recessions. Dinh *et al.* (2021) executed a study focused on predicting financial distress across the six largest nations within the ASEAN Economic Community (AEC). They inferred that the Distance to Default (DD) methodology is appropriate as an early warning signal for impending financial distress in the subsequent year.

Besides, the advanced machine learning models like Random Forest, XGBoost, CatBoost and Support Vector Machine are employed in much research of financial distress predictions. Hou *et al.* (2024) employed financial indicators such as working capital, operational funds, cash ratio, quick ratio, current ratio, and borrowing ratio for a Random Forest model to predict financial distress in 12 companies. The study revealed a strong predictive capability, with the model achieving an R^2 value of 0.77. This shows that 77% of the variation in the data could be explained by the model, demonstrating its effectiveness in identifying financial distress. This result highlights the potential of Random Forest as a robust tool for financial distress prediction, particularly when combined with key financial ratios. Lu and Hu (2024) utilized a set of 15 financial indices including net profit, total liabilities, monetary capital, fixed assets, main business income, total assets, short-term borrowings, total operating income, net fixed assets, basic earnings per share, bonds payable, net assets per share, investment income, and long-term borrowings as input features for a CatBoost model. The model was optimized using the Particle Swarm Optimization (PSO) method.

The results demonstrated promising performance, with a Mean Squared Error of 0.032 and a Mean Absolute Error of 0.124, highlighting the model's ability to effectively predict financial distress using these financial indicators. Carmona *et al.* (2022) analyzed 1,760 French firms to identify key indicators of business failure. Their findings highlighted the important indicators include equity per employee, solvency, current ratio, net profitability, and sustainable return on investment among 36 examined indicators. Adopting XGBoost as the primary model for failure prediction, the study achieved a high AUC score of 0.964, demonstrating the model's strong predictive capability. Doğan *et al.* (2022) applied Support Vector Machine and Logistic Regression to predict financial distress among 172 companies listed on Borsa İstanbul. A total of 24 financial indicators were used as input features for the predictive models. Their findings revealed that the hybrid model combining Logistic Regression and Grid Search-optimized SVM achieved the highest accuracy of 93.75% on the testing set, proving its usefulness in forecasting financial turmoil.

MATERIALS AND METHODS

Figure 1 illustrates the methodology employed in this study. The process begins with the collection of datasets which include financial ratios and the corresponding company status for each year. To address the issue of class imbalance, the Adaptive Synthetic Sampling (ADASYN) technique is applied to ensure that all three classes in the dataset are balanced. Next, the dataset is divided into training (70%) and testing (30%) subsets. The training set is then used to train machine learning models including XGBoost, CatBoost, Random Forest, and Support Vector Machine (SVM) to predict company financial status. The trained models generate predictions for the testing set, which are subsequently evaluated against actual values using performance metrics such as AUC-ROC, F1-score, precision, accuracy, specificity and recall.

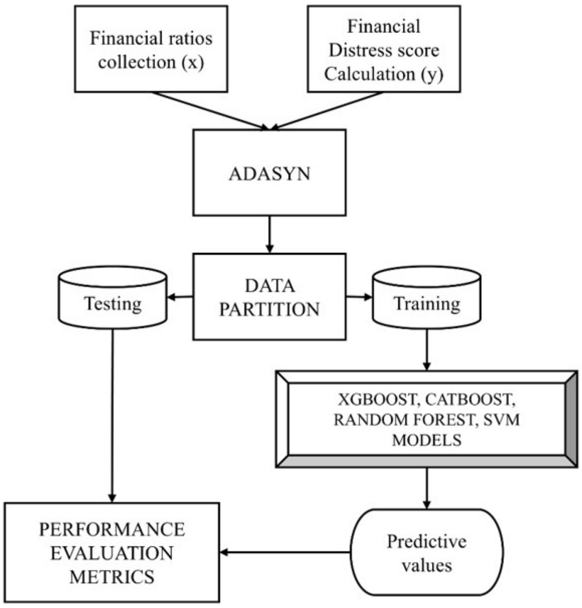

Figure 1: Process flow chart of this paper

Financial Distress Measurement

Financial distress prediction is a crucial area of research in corporate finance which helps investors, regulators, and policymakers in assessing the likelihood of firm bankruptcy. Among various financial distress prediction models, Altman's Z-score remains one of the most widely used methodologies. Developed by Altman (1968) in 1968, the Altman Z-score is a model based on multiple discriminant analysis (MDA) that uses financial ratios to predict corporate bankruptcy. The model was initially designed for publicly traded manufacturing firms and demonstrated high predictive accuracy in distinguishing between businesses who are bankrupt and those that are not. Two additional variants were introduced: Z' in 1983 (Altman, 1983) and Z" in 1995 (Altman *et al.*, 1995). These models are specifically designed for private companies and firms operating in emerging markets, with Z" particularly suited for non-manufacturing companies. In this study, the Z-score and Z"-score are utilized to assess the financial status of manufacturing and non-manufacturing companies. The sample consists of 18 manufacturing firms, representing approximately 53% of the total, and 16 non-manufacturing firms, making up the remaining 47% as shown in Table 1.

Table 1: Investigated companies in Vietnam

No	Company Name	Industry	Catogories
1	Investment And Industrial Development Corporation (HOSE: BCM)	Real Estate Management & Development	Non-Manufacturing
2	Viettel Construction Joint Stock Corporation (HOSE: CTR)	Telecommunication Services	Non-Manufacturing

3	DHG Pharmaceutical Joint Stock Company (HOSE: DHG)	Pharmaceuticals, Biotechnology & Life Sciences	Manufacturing
4	FPT Corporation (HOSE: FPT)	Software & Services	Non-Manufacturing
5	PetroVietnam Gas Joint Stock Corporation	Utilities	Non-Manufacturing
6	Vietnam Rubber Group - Joint Stock Company (HOSE: GVR)	Materials	Manufacturing
7	Hoa Phat Group Joint Stock Company (HOSE: HPG)	Materials	Manufacturing
8	Vietnam Airlines JSC (HOSE: HVN)	Transportation	Non-Manufacturing
9	IDICO Corporation - JSC (HNX: IDC)	Real Estate Management & Development	Non-Manufacturing
10	Lam Dong Investment & Hydraulic Construction JSC (HNX: LHC)	Capital Goods	Non-Manufacturing
11	Masan Group Corporation (HOSE: MSN)	Food, Beverage & Tobacco	Manufacturing
12	Mobile World Investment Corporation (HOSE: MWG)	Consumer Discretionary Distribution & Retail	Non-Manufacturing
13	Tien Phong Plastic Joint Stock Company (HNX: NTP)	Materials	Manufacturing
14	Phuoc Hoa Rubber Joint Stock Company (HOSE: PHR)	Materials	Manufacturing
15	Viet Nam National Petroleum Group (HOSE: PLX)	Consumer Discretionary Distribution & Retail	Non-Manufacturing
16	Phu Nhuan Jewelry Joint Stock Company (HOSE: PNJ)	Consumer Discretionary Distribution & Retail	Manufacturing
17	PetroVietnam Power Corporation (HOSE: POW)	Utilities	Non-Manufacturing
18	Saigon Beer - Alcohol - Beverage Corporation (HOSE: SAB)	Food, Beverage & Tobacco	Manufacturing
19	Son La Sugar JSC (HNX: SLS)	Food, Beverage & Tobacco	Manufacturing
20	Thaiholdings Joint Stock Company (HNX: THD)	Capital Goods	Non-Manufacturing
21	TNG Investment and Trading JSC (HNX: TNG)	Consumer Durables & Apparel	Manufacturing
22	Vicostone JSC (HNX: VCS)	Materials	Manufacturing
23	Vinhomes JSC (HOSE: VHM)	Real Estate Management & Development	Non-Manufacturing
24	Vingroup Joint Stock Company (HOSE: VIC)	Real Estate Management & Development	Non-Manufacturing
25	Vietjet Aviation Joint Stock Company (HOSE: VJC)	Transportation	Non-Manufacturing
26	Viet Nam Dairy Products Joint Stock Company (HOSE: VNM)	Food, Beverage & Tobacco	Manufacturing
27	Vincom Retail Joint Stock Company (HOSE: VRE)	Real Estate Management & Development	Non-Manufacturing
28	Hoang Kim Tay Nguyen Group JSC (HNX: CTC)	Consumer Services	Manufacturing
29	Thien Nam Trading Import Export JSC (HOSE: TNA)	Consumer Discretionary Distribution & Retail	Non-Manufacturing
30	Hoang Anh Gia Lai Agricultural JSC (HOSE: HNG)	Food, Beverage & Tobacco	Manufacturing
31	Song Da 6 JSC (HNX: SD6)	Capital Goods	Non-Manufacturing
32	Dong A Plastic JSC (HOSE: DAG)	Capital Goods	Manufacturing
33	HTINVEST JSC (HNX: HTP)	Commercial & Professional Services	Non-Manufacturing
34	Vietnam Electric Cable Corporation (HOSE: CAV)	Capital Goods	Manufacturing

The original Altman Z-score consists of five key financial ratios that capture a company's profitability, liquidity, leverage, and efficiency:

• Liquidity A_1 is assessed by working capital as a proportion of total assets.

• Cumulative profitability A_2 is measured by the ratio of retained earnings to total assets

• Operating efficiency A_3 is computed by the ratio of EBIT to total assets.

• Leverage A_4 is calculated by the ratio of market value of equity to total liabilities

• Asset turnover A_5 is determined by the ratio of sales to total assets

The model assigns a Z-score based on a weighted sum of these ratios where $Z > 2.99$ indicates financial stability (Safe Zone). Scores between 1.81 and 2.99 suggest potential distress (Grey Zone). A $Z < 1.81$ signals a high risk of bankruptcy (Distress Zone). The formula of Z-score is expressed in Equation (1).

$$Z=1.2A_1+1.4A_2+3.3A_3+0.6A_4+1.0A_5 \quad (1)$$

In 1995, the third variation of the model, Z"-Score, was developed to account for the economic conditions in emerging markets, where companies often experience higher financial volatility. The model excludes the Sales/Total Assets (X_5) ratio, as emerging market firms often operate in diverse economic conditions with varying revenue structures. The formula is described in Equation (2).

$$Z''=6.56A_1+3.26A_2+6.72A_3+1.05A_4 \quad (2)$$

The classification thresholds are Z" score > 5.85 indicates financial stability (Safe Zone). Scores between 4.15 and 5.85 suggest uncertainty (Grey Zone). A Z" score < 4.15 signals high distress risk (Distress Zone).

Meanwhile, the Emerging Market Scoring (EMS) model (Altman, 2005) is founded on a fundamental financial analysis derived from a qualitative risk assessment framework. It functions as a refined rating system for evaluating specific credit risks. By building upon previous financial distress models, the EMS model integrates the strengths of established approaches, such as the Z-score, Z'-score, and Z"-score, while addressing their limitations. Notably, it incorporates an adjustment of +3.25 to better suit emerging markets, including developing economies like Vietnam. This enhancement enables a more precise and comprehensive evaluation of financial health in dynamic and volatile market conditions. The EMS model evaluates a company's financial stability using the following Equation (3).

$$EMS=6.56A_1+3.26A+6.72A_3+1.05A'_4+3.25 \quad (3)$$

where A'_4 is Book value of equity to total liabilities, indicating financial leverage and solvency.

The EMS score categorizes companies into three risk levels:

• Safe Zone (EMS>5.85) – Firms in this category are financially stable, with minimal bankruptcy risk.

• Warning Zone (4.15≤EMS≤5.85) – These firms face financial uncertainty, with potential distress risks.

• Distress Zone (EMS<4.15) – Companies in this range have a high probability of default or bankruptcy.

Unlike the accounting-based approach such as Altman Z-score or Emerging Market score (EMS), Distance to Default (DD) method is a structural credit risk model based on Merton's model (1974) (Merton, 1974). It measures how close a company is to default based on the value of its assets relative to its liabilities. The idea is that a company defaults when its asset value falls below its liabilities. However, the original model assumed constant debt which lacks empirical support. To address this limitation, the DD model was refined to incorporate default risk more effectively by including factors such as firm leverage ratio and equity volatility. By adopting a constant leverage ratio, the modified version of the DD model is considered more realistic and dynamic compared to the traditional Merton model. As a result, the updated DD model provides a more accurate estimation of default probability across various types of firms. Notably, this enhanced version is particularly effective in measuring financial distress in emerging and volatile markets (Byström, 2006; Pham Vo Ninh *et al.*, 2018).

The updated Distance-to-Default (DD) model is rebuilt using the Merton model as a basis (Pham Vo Ninh *et al.*, 2018), where leverage ratio $L=D/((E+D))$ (E is the market value of ewuity and D is the book value of debt) and σ_E is represented for the volatility of the firm's equity. The formula of DD score is illustrated in Equation (4).

$$DD = \frac{\ln\left(\frac{1}{L}\right)}{\sigma_E(1-L)} = \frac{\ln(L)}{(L-1)}\frac{1}{\sigma_E} \quad (4)$$

A robust correlation is evident between the Distance-to-Default (DD) metric and the likelihood of default, commonly referred to as the Expected Default Frequency (EDF). The EDF is derived from the DD value through the application of the cumulative normal distribution. Subsequently, this EDF outcome is aligned with the credit ratings provided by Standard & Poor's (S&P). S&P rating categories divide companies into twenty-two different tiers, ranging from 0% to 20%, within the KMV-EDF paradigm. Through bivariate analysis, these classifications are segmented into three categories: Safe Zone (0.00 – 0.52%], Grey Zone (0.52 – 6.94%], and Distress Zone (6.94 – 20.00%]. This categorization adheres to the guidelines established by Lopez (2004) in 2004. Consequently, the revised DD model is conceived to be more straightforward and structurally analogous to its predecessor. Nonetheless, it provides enhanced precision compared to the earlier iteration and is more appropriately tailored for emerging economies such as Vietnam.

Dataset Pre-processing

Table 2: Investigated Financial Ratios

No	Financial Ratios	No	Financial Ratios
1	Trailing EPS	35	Number of days of payables

2	Book value per share (BVPS)	36	Fixed asset turnover
3	P/E	37	Total asset turnover
4	P/B	38	Equity turnover
5	P/S	39	Short-term liabilities to total liabilities
6	Dividend yield	40	Debt to assets
7	Beta	41	Liabilities to assets
8	EV/EBIT	42	Equity to assets
9	Gross profit margin	43	Short-term liabilities to equity
10	EBIT margin	44	Debt to equity
11	EBITDA/Net revenue	45	Liabilities to equity
12	Net profit margin	46	Accrual ratio CF
13	ROE	47	Cash to income
14	Return on capital employed (ROCE)	48	Net cash flows/Short -term liabilities
15	ROA	49	Cash return to assets
16	ROE Trailling	50	Cash return on equity
17	ROA Trailling	51	Cash to income
18	Net revenue	52	Debt coverage
19	Gross profit	53	Cash flow per share (CPS)
20	Profit before tax	54	Cost of goods sold/Net revenue
21	Profit after tax for shareholders of the parent company	55	Selling expenses/Net revenue
22	Total assets	56	General and Administrative expenses/Net revenue
23	Long-term liabilities	57	Interest expenses/Net revenue
24	Liabilities	58	Short-term assets/Total assets
25	Owner's equity	59	Cash/Short-term assets
26	Cash ratio	60	Short-term investments/Short-term assets
27	Quick ratio	61	Short-term receivables/Short-term assets
28	Short-term ratio	62	Inventory/Short-term assets
29	Interest coverage	63	Other Short-term assets/Short-term assets
30	Receivables turnover	64	Long-term assets/Total assets
31	Days of sales outstanding	65	Fixed assets/Total assets
32	Inventory turnover	66	Tangible fixed assets/Fixed assets
33	Days of inventory on hand	67	Intangible fixed assets/Fixed assets
34	Payables turnover	68	Construction in progress/Fixed assets

In this study , the set of data was collected from Vietstock.vn and comprises 318 observations, with 68 independent variables as shown in Table 2 and one target variable, which was classified into three categories -safety, risky, and distress- based on three financial distress measurement methods: Distance to Default (DD), EMS, and Altman Z-score. However, as illustrated in Figure 2, the dataset exhibits a significant class imbalance:

- Distance to Default (DD) method: 13 distressed cases, 73 risky cases, 232 safety cases
- EMS method: 42 distressed cases, 86 risky cases, 190 safety cases
- Z-score method: 74 distressed cases, 84 risky cases, 160 safety cases

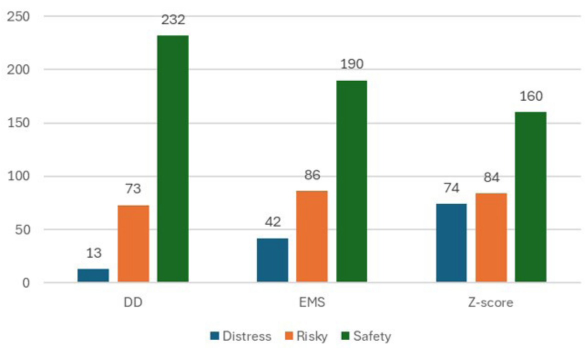

Figure 2: Dataset Description

As illustrated in Figure 2, the number of companies classified as distressed using the Z-score method is nearly twice that of the EMS method and six times higher than the DD method. Meanwhile, the number of companies categorized as risky remains relatively consistent across the three methods. However, the classification of safety status varies significantly - the DD method identifies the highest number of safe companies, while the count decreases when using the EMS method and further declines under the Z-score method. This discrepancy highlights the differing sensitivity of each method in assessing financial distress. In this study, approximately 53% of the firms investigated are non-manufacturing companies, which significantly influences the results of different financial distress measurement methods. The DD method relies on stock prices to estimate distress, making it more sensitive to market fluctuations and investor sentiment. Non-manufacturing firms often have more stable stock prices because they are less exposed to factors like raw material costs, production delays, and supply chain disruptions. Since DD uses stock price volatility in its calculations, lower volatility leads to fewer distress classifications, explaining why the number of distressed firms is significantly lower under the DD method compared to Z-score and EMS. The imbalances indicate that the dataset is skewed toward the majority class, which can lead to biased model predictions. To address this issue, the Adaptive Synthetic Sampling (ADASYN) is applied to balance the dataset, which guarantees a better representative distribution of all classes for machine learning models. This step is crucial to improve model performance and ensuring more reliable financial distress predictions.

Most machine learning algorithms perform optimally when class distributions are relatively balanced; however, imbalanced data often lead to the majority class dominating the learning process, resulting in biased models and unreliable predictions. Adaptive Synthetic Sampling (ADASYN) is an oversampling technique used to handle imbalanced datasets especially in classification problems where certain classes are underrepresented. It improves upon the Synthetic Minority Over-sampling Technique (SMOTE) by focusing more on generating synthetic samples for harder-to-classify minority class instances. SMOTE is another oversampling method that generates synthetic samples for the minority class to address class imbalance in classification problems. It achieves this by selecting samples from the minority class, identifying their nearest neighbors, and generating new samples that interpolate between these points in the feature space (Hanafy & Ming, 2021). SMOTE interpolates instances of the minority class to create synthetic samples rather than duplicating them; therefore, it does not change the expected value of the minority class but decreases its variability. Similarity to strategy of SMOTE, ADASYN (Adaptive Synthetic Sampling) addresses class imbalance by generating synthetic samples, focusing more on difficult-to-classify minority instances. It first identifies the imbalance, determines the number of new samples needed, and assigns higher weights to minority instances that are harder to classify. Using nearest neighbors, ADASYN generates new data points in these challenging regions, ensuring a more balanced and representative dataset that improves model performance while reducing bias. To adopt ADASYN over SMOTE, we consider the outstanding attributes from ADASYN in handling imbalanced data. ADASYN is predicated on the principle of adaptively producing synthetic data instances for minority classes in accordance with their respective distributions: a greater volume of synthetic data is generated for minority class instances that present greater challenges for learning in comparison to those minority instances that are less complex to learn. The ADASYN technique possesses the capacity to not only mitigate the learning bias engendered by the initial imbalanced data distribution, but it also has the ability to dynamically adjust the decision boundary to concentrate on those instances that are particularly challenging to learn (Haibo *et al.*, 2008).

Classification Models

To compare the suitability of Altman Z-score, EMS and DD method in evaluating the listed companies in Vietnam, we employed four widely used machine learning models: Random Forest, XGBoost, CatBoost and Support Vector Machine models.

The Support Vector Machine (SVM) represents a formidable supervised learning algorithm introduced in 1995 by Cortes and Vapnik (1995), which is extensively employed for both classification and regression tasks. Its efficacy is particularly pronounced in high-dimensional spaces, and it is distinguished by its capability to manage intricate decision boundaries. Support Vector Machines (SVM) function by pinpointing an optimal hyperplane that maximizes the margin between divergent classes. This hyperplane acts as a decision boundary that delineates data points associated with separate categories. The data points that are in closest proximity to the hyperplane, termed support vectors, are pivotal in establishing the optimal decision boundary. SVM can utilize a variety of kernel functions to transform data into elevated-dimensional spaces, facilitating the separation of data that is not linearly separable. Furthermore, regularization and slack variables are integral in addressing noisy data. The C parameter regulates the balance between maximizing the margin and minimizing misclassification, while slack variables permit SVM to accommodate some degree of misclassification, thereby enhancing its resilience to noise. A higher C value prioritizes correct classification over a large margin, while a lower C allows a larger margin but tolerates some misclassification. This combination enhances SVM's ability to generalize, particularly in real-world datasets where achieving perfect separation is challenging.

Random Forest model is introduced by Breiman (2001) in 2001. It is a widely used ensemble learning algorithm for both classification and regression tasks, as it can handle both continuous and categorical datasets. This study employs a Random Forest Classifier (RFC), which consists of multiple decision trees that operate collectively. A bootstrap sample of the dataset is used to train each tree in the forest, and features are randomly selected at each split to ensure diversity among trees. A classification tree serves the purpose of forecasting a categorical response as opposed to a numerical one. A classification tree asserts that each instance is allocated to the predominant category of training instances within its respective domain. A majority voting mechanism is implemented for the classification process. Compared to a single decision tree, RFC offers a significant advantage: rather of depending on a single model, it works as a group of professionals, where each tree contributes to the final decision. This collective approach enhances the model's accuracy, robustness, and generalization ability.

Extreme Gradient Boosting (XGBoost) constitutes a potent machine learning algorithm that builds upon the foundational concepts of gradient boosting, which was advanced in 2016 by Chen and Guestrin (Chen & Guestrin, 2016). Analogous to gradient boosting, XGBoost amalgamates the predictive capabilities of multiple learners into a singular model through an iterative process. XGBoost adheres to the principle of boosting, wherein weak learners (specifically decision trees) are sequentially trained to diminish the residual errors generated by preceding trees. It has gained widespread popularity due to its efficiency, scalability, and superior performance in classification and regression tasks. XGBoost is designed to handle missing data, feature sparsity, and large datasets while maintaining high predictive accuracy. The primary reasons for its effectiveness include advanced regularization techniques, parallelization, and efficient handling of missing values. XGBoost incorporates L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting, making it more robust than traditional gradient boosting models. Additionally, while gradient boosting minimizes overall model error by optimizing the loss function of its base models, XGBoost enhances this process by incorporating both first- and second-order partial derivative approximations, known as the gradient and Hessian. This approach provides more precise information about the gradient direction, leading to faster convergence and more efficient loss minimization.

CatBoost (Categorical Boosting) is a gradient boosting algorithm developed in 2018 by Yandex, a Russian multinational technology company, that is optimized for both categorical and numerical datasets (Dorogush *et al.*, 2018; Dorogush *et al.*, 2017). CatBoost enhances gradient boosting through ordered boosting, which reduces prediction shift by using properly permuted datasets to prevent target leakage, leading to more stable predictions. It also employs symmetric trees, ensuring balanced splits at each depth, which improves training efficiency and reduces overfitting compared to traditional asymmetric trees. Additionally, CatBoost efficiently handles pure numerical features without requiring extensive normalization or scaling, allowing it to learn directly from raw data. Its optimized handling of categorical features eliminates the need for one-hot encoding or complex preprocessing, making it a powerful choice for datasets with mixed data types. According to Kaggle's "State of Data Science and Machine Learning" surveys, CatBoost has been recognized among the most frequently used machine learning frameworks globally. In the 2020 survey, it ranked within the top 8, and in the 2021 survey, it moved up to the top 7 (Mooney, 2020). While CatBoost is popular, other frameworks like Scikit-learn, TensorFlow, and Keras have higher usage rates among data scientists. For instance, in the 2021 survey, over 80% of respondents reported using Scikit-learn, making it the most widely adopted framework (Đạt, 2021). This highlights CatBoost's significant role in the machine learning community.

To tune the hyperparameter for these models, Randomized Search Cross-Validation technique is applied to find the appropriate hyperparameter. Randomized Search Cross-Validation represents a sophisticated optimization methodology employed for the fine-tuning of hyperparameters within machine learning frameworks. This technique involves the stochastic selection of hyperparameter configurations from a specified distribution and subsequently assesses the efficacy of the model through the application of cross-validation. Unlike Grid Search, which exhaustively searches all possible parameter combinations, Random Search selects a limited number of random combinations, making it more computationally efficient while still providing good results (Alibrahim & Ludwig, 2021). The search spaces for predictive models are presented in Table 3.

Table 3: Hyperparameter Search Spaces for Predictive Models

Hyperparameters	**Search Spaces**	**Hyperparameters**	**Search Spaces**
XGBoost		**Random Forest**	
Learning rate	(0.01, 0.3)	n_estimators	50, 500
n_estimators	(100, 500)	Max depth	(3, 20)
Max depth	(3, 15)	Min samples split	(2, 10)
Min child weight	(1, 6)	Min samples leaf	(1, 5)
gamma	(0, 1)	Max features	sqrt, log2
subsample	(0.5, 0.5)	bootstrap	True, False
Colsample bytree	(0.5, 0.5)		
Reg alpha	(0.0001, 10) (logarithmic scale)		
Reg lambda	(0.0001, 10) (logarithmic scale)		

CatBoost		SVM	
iterations	(100, 1000)	C	(0.1, 10)
depth	(3, 12)	kernel	linear, poly, rbf, sigmoid
Learning rate	(0.01, 0.3)	gamma	scale, auto
l2 leaf reg	(1, 10)		
Border count	(32, 255)		
Random strength	(1, 10)		
Bagging temperature	(0, 1)		

Performance Evaluation

This segment delineates the evaluative metrics employed to gauge the efficacy of each model, specifically Precision, Recall, Balanced Accuracy, F1 Score, and Specificity. Their corresponding formulations are articulated in Equations (5)-(9). An elevation in these metrics signifies an enhancement in model performance. True Positive (TP) denotes the scenario in which the model accurately discerns a positive class, while a False Positive (FP) transpires when the model mistakenly classifies a negative instance as positive. In a similar vein, True Negative (TN) signifies that the model correctly identifies a negative class, whereas False Negative (FN) emerges when the model neglects to acknowledge a positive instance. The term β^2 denotes a variable incorporated within the F-beta score, which represents a weighted harmonic mean that integrates both Precision and Recall.

$$Balanced\ Accuracy = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \tag{5}$$

$$Precision = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FP_i)} \tag{6}$$

$$Recall = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FN_i)} \tag{7}$$

$$\text{Specificity} = \frac{\sum_{i=1}^{N} TN_i}{\sum_{i=1}^{N}(TN_i + FP_i)} \tag{8}$$

$$\text{F1 Score} = \frac{(\beta^2 + 1)Precision * Recall}{\beta^2 Precision + Recall} \tag{9}$$

Besides, we utilized confusion matrix to plot the detail results for the best methods to see clearly the values in each part of confusion matrix. Confusion Matrix with (n*n) table (n is the number of classes) is created to measure the performance of our classification models. In multiclass classification, the positive class refers to the specific label being evaluated, while the negative class encompasses all remaining labels.

The model's capacity to differentiate between classes is gauged by the Receiver Operating Characteristic-Area Under the Curve (ROC-AUC). The ROC-AUC score ranges from 0 to 1, where a score closer to 1 indicates a highly effective model in distinguishing between classes. For multiclass classification problems, the one-vs-all methodology is applied to calculate individual ROC-AUC scores for each class, treating each class as the positive class while combining all other classes as negative. The final weighted ROC-AUC score is then computed, which provides a more comprehensive evaluation of the model's overall performance across all classes. A model with a ROC-AUC scores close to 1 is considered optimal, demonstrating strong classification power. We employed this one-vs-all approach to obtain the weighted ROC-AUC score for our models, ensuring a robust and accurate assessment of performance. The detailed performance analysis of each method will be presented in the following section.

RESULTS AND DISCUSSION

The performance of four machine learning models—XGBoost, CatBoost, SVM, and Random Forest - using different target variables from DD score, Altman Z-score, and EMS score is presented in Table 4, Table 5 and Table 6, respectively. The best-performing machine learning model in this study was Random Forest, followed by XGBoost, then CatBoost, with SVM ranking last. The underperformance of SVM can be attributed to several factors. Unlike decision tree-based models, which inherently perform feature selection by identifying the most relevant variables during training, SVM lacks built-in feature selection. Furthermore, financial ratios often exhibit complex, non-linear relationships with financial distress, making tree-based models more effective at capturing these interactions compared to SVM. The best performance in predicting financial distress for Vietnamese companies is achieved when using the EMS score as the target variable, with a ROC-AUC value of 99.26%. The Z-score follows in ranking, while the DD score performs the worst in predicting financial distress. In Vietnam, stock markets may be less developed, with lower liquidity, information asymmetry, and possible government influence on firms, leading to less reliable market-based distress signals, reducing the effectiveness of the DD score. Although the results section does not provide a detailed comparison between the SMOTE and ADASYN techniques, our experiments during the research showed that ADASYN outperforms SMOTE in improving the accuracy of predictive models. The specicific results of the findings are described in the following parts.

Table 4: Prediction accuracy of each model when utilizing DD score

DD SCORE	XGBOOST	RANDOM FOREST	CATBOOST	SVM
Accuracy	0.7604	0.7604	0.7396	0.5729
Precision (WEIGHTED)	0.7257	0.7189	0.6837	0.6016
F1 Score (WEIGHTED)	0.7279	0.7238	0.6901	0.5849
Recall (WEIGHTED)	0.7604	0.7604	0.7396	0.5729
Specificity (CLASS 0, 1, 2)	[0.9783, 0.9296, 0.4483]	[0.9891, 0.9296, 0.4138]	[0.9783, 0.9296, 0.4483]	[0.8913, 0.7887, 0.4483]
ROC-AUC Score (WEIGHTED)	0.8437	0.8594	0.8333	0.7188

Table 4 presents XGBoost and Random Forest are the best-performing models, with the highest accuracy (76.04%) and balanced precision, recall, and F1-scores. CatBoost is slightly worse but still performs well, especially for class 2. SVM is the weakest model, performing poorly for class 1 and class 2, with the lowest accuracy (57.29%).

Table 5: Prediction accuracy of each model when utilizing Z- score

Z-SCORE	XGBOOST	RANDOM FOREST	CATBOOST	SVM
Accuracy	0.7812	0.7812	0.7812	0.6562
Precision (WEIGHTED)	0.7964	0.7971	0.7961	0.7282
F1 Score (WEIGHTED)	0.7821	0.7633	0.7739	0.6704
Recall (WEIGHTED)	0.7812	0.7812	0.7812	0.6562
Specificity (CLASS 0, 1, 2)	[0.8590, 0.9014, 0.9302]	[0.8462, 0.95772, 0.8605]	[0.8590, 0.9014, 0.9302]	[0.7564, 0.8451, 0.9302]
ROC-AUC Score (WEIGHTED)	0.9028	0.8989	0.8912	0.6737

Table 5 describes XGBoost performing the best, achieving the highest F1-score (0.7821) and ROC-AUC (0.9028). Random Forest & CatBoost are close contenders, with Random Forest showing the highest precision (0.7971) and specificity for Class 1 (0.9577). SVM performs the worst, with significantly lower accuracy (65.62%), recall, and ROC-AUC (0.6737), indicating poor predictive power.

Table 6: Prediction accuracy of each model when utilizing EMS score

EMS SCORE	XGBOOST	RANDOM FOREST	CATBOOST	SVM
Accuracy	0.9271	0.9375	0.9167	0.8333
Precision (WEIGHTED)	0.9276	0.9407	0.9162	0.8438
F1 Score (WEIGHTED)	0.9270	0.9376	0.9158	0.8354
Recall (WEIGHTED)	0.9271	0.9375	0.9167	0.8333
Specificity (CLASS 0, 1, 2)	[0.9884, 0.9444, 0.9412]	[1.0, 0.9444, 0.9412]	[0.9884, 0.94444, 0.9412]	[0.9186, 0.9306, 0.8824]
ROC-AUC Score (WEIGHTED)	0.9881	0.9926	0.9908	0.9055

Table 6 points out that the highest accuracy is achieved by Random Forest (0.9375), followed closely by XGBoost (0.9271) and CatBoost (0.9167) when using Altman Z-score as target classification. The SVM model performs the worst (0.8333). SVM performs the worst across all metrics, making more false positives and false negatives, and having the lowest ROC-AUC score. Since Random Forest (0.9926), CatBoost (0.9908), and XGBoost (0.9881) are close to 1 in ROC-AUC score, all three models are highly effective in distinguishing different classes.

To illustrate the detailed performance of the best method, we plot confusion matrix as Figure 3 . It shows that all models performed well in Class 2, with 60 correctly classified instances. CatBoost performed slightly better in Class 0 but worse in Class 1 compared to XGBoost and RF. Random Forest had the best balance for Class 1, with the highest correct classifications (22). XGBoost and RF had similar performance, but RF had fewer misclassifications in Class 1. SVM has the worst performance among investigated models.

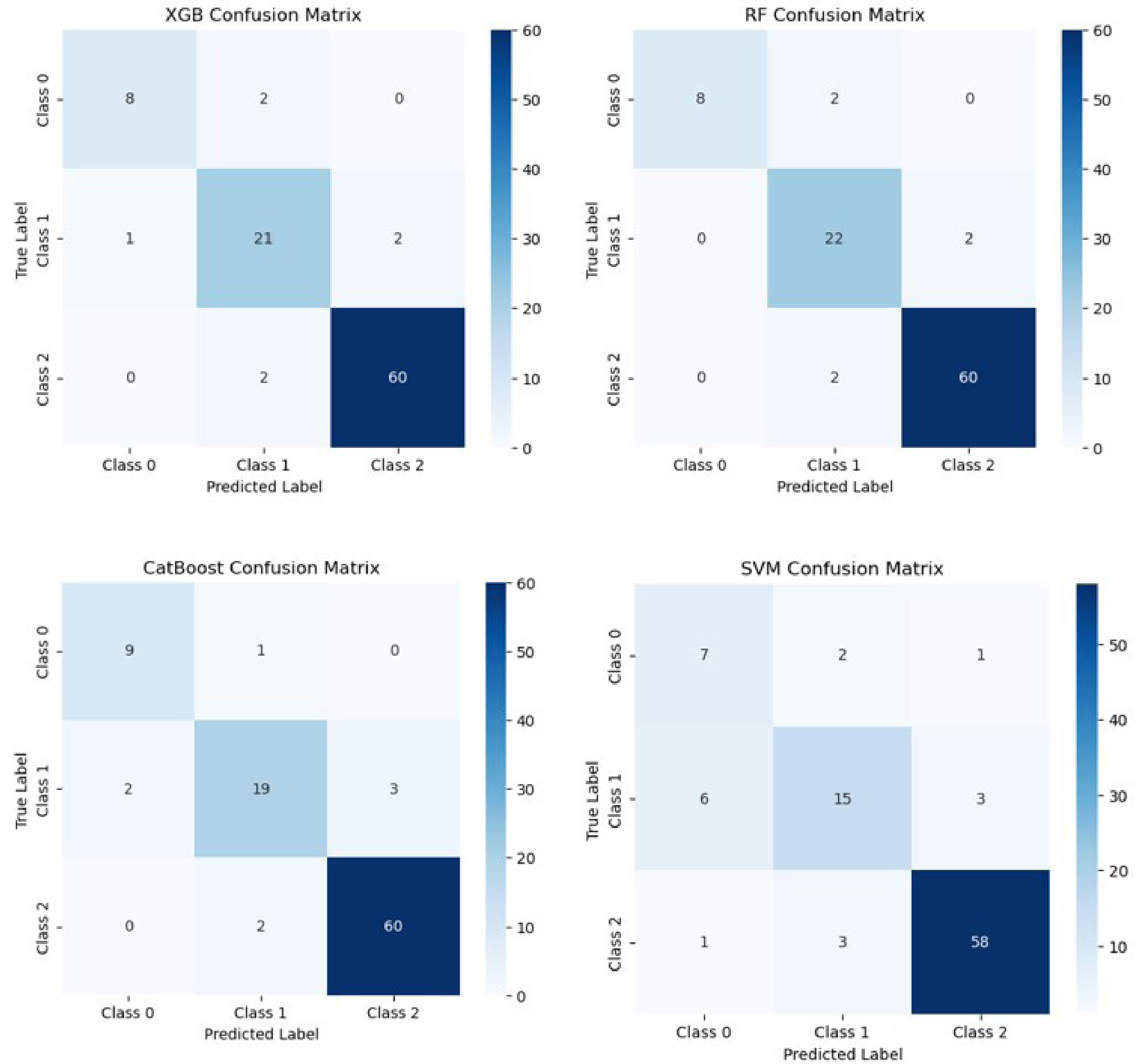

Figure 3: Confusion Matrix of predictive model performance when using EMS scores

Random Forest excels in handling multiple classes effectively by constructing multiple decision trees and aggregating their predictions. This approach makes the multiclass decision-making process straightforward and efficient. One of the key strengths of Random Forest is its resilience to noisy data. Compared to other models like XGBoost, CatBoost, and SVM, Random Forest is better equipped to manage noisy datasets. By combining the results of several trees, it minimizes the impact of individual errors or outliers, leading to a more stable and reliable model, particularly in real-world scenarios such as financial data, where noise and outliers are common.

The best performance with 99.26% at ROC-AUC value achieves from models using EMS score as the target variable to predict. This implies that the Vietnamese stock market can be explained by EMS score. The Emerging Market Score is a composite measure that accounts for various economic and financial factors specific to emerging markets. In the case of Vietnam, which is considered an emerging market, the EMS score captures key macroeconomic indicators, financial health, and market performance indicators that are critical in understanding the stability and risk of companies within that market. Given the unique economic conditions, volatility, and risks associated with emerging markets like Vietnam, the EMS score serves as an effective predictor of financial distress and stability for companies listed in such markets. By using this score as the target variable in the model, the machine learning algorithms can leverage the specific characteristics and risk factors of the Vietnamese market, leading to higher prediction accuracy. The ROC-AUC score of 99.26% indicates that the EMS score is highly informative and predictive in this context.

CONCLUSION

In this study, we aimed to predict the financial health of listed companies in Vietnam using a combination of financial ratios and advanced machine learning techniques. By leveraging methods such as Distance to Default (DD), Emerging Market Score (EMS), and Altman Z-score, we developed models capable of assessing the financial status of companies. Among the machine learning algorithms tested, Random Forest demonstrated superior performance, achieving an impressive ROC-AUC score of 99.26%. The EMS score, when combined with machine learning models, proved to be an effective predictor of financial distress, outperforming traditional methods and highlighting the potential for applying machine learning in the context of financial health prediction.

Our study also emphasized the importance of data balancing techniques, specifically the Adaptive Synthetic Sampling (ADASYN) method, which significantly enhanced the performance of the models in predicting financial distress. This underscores the critical role of data preprocessing in improving model accuracy, especially in cases involving imbalanced datasets common in financial markets.

While Random Forest excelled in this study due to its resistance to noise and ability to handle multi-class classification, XGBoost and CatBoost remain powerful alternatives. The integration of the EMS score demonstrated its value in financial distress prediction, especially in emerging markets like Vietnam. Future research could explore expanding these models to other markets or incorporating additional economic indicators to improve predictions. This study provides valuable insights for financial analysts and policymakers in assessing corporate health with higher accuracy.

REFERENCES

Alibrahim, H., & Ludwig, S. A. (2021, 28 June-1 July 2021). Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization. *2021 IEEE Congress on Evolutionary Computation (CEC),* pp. 1551-1559. IEEE

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance, 23*(4), 589-609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Altman, E. I. (1983). *Corporate financial distress: A complete guide to predicting, avoiding, and dealing with bankruptcy.* Wiley.

Altman, E. I. (2005). An emerging market credit scoring system for corporate bonds. *Emerging Markets Review, 6*(4), 311-323. https://doi.org/10.1016/j.ememar.2005.09.007

Altman, E. I., Eom, Y. H., & Kim, D. W. (1995). Failure Prediction: Evidence from Korea. *Journal of International Financial Management & Accounting, 6*(3), 230-249. https://doi.org/10.1111/j.1467-646X.1995.tb00058.x

Breiman, L. (2001). *Random Forests. Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Byström, H. (2006). Merton Unraveled: A Flexible Way of Modeling Default Risk. *The Journal of Alternative Investments, 8,* 39-47. https://doi.org/10.3905/jai.2006.627849

Carmona, P., Dwekat, A., & Mardawi, Z. (2022). No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. *Research in International Business and Finance, 61,* 101649. https://doi.org/10.1016/j.ribaf.2022.101649

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297. https://doi.org/10.1007/BF00994018

Đạt, T. M. T. (2021). *Kaggle's state of machine learning and data science 2021.* Scribd. https://www.scribd.com/document/554216647/Kaggle-s-State-of-Machine-Learning-and-Data-Science-2021

Dinh, D. V., Powell, R. J., & Vo, D. H. (2021). Forecasting corporate financial distress in the Southeast Asian countries: A market-based approach. *Journal of Asian Economics, 74,* 101293. https://doi.org/10.1016/j.asieco.2021.101293

Doğan, S., Koçak, D., & Atan, M. (2022). Financial Distress Prediction Using Support Vector Machines and Logistic Regression. In M. K. Terzioğlu (Ed.), *Advances in Econometrics, Operational Research, Data Science and Actuarial Studies: Techniques and Theories* (pp. 429-452). Springer International Publishing. https://doi.org/10.1007/978-3-030-85254-2_26

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). *CatBoost: Gradient boosting with categorical features support.* arXiv. https://doi.org/10.48550/arXiv.1810.11363

Dorogush, A. V., Gulin, A., Gusev, G., Kazeev, N., Ostroumova, L., & Vorobev, A. (2017). *Fighting biases with dynamic boosting.* arXiv. https://doi.org/10.48550/arXiv.1706.09516

Haibo, H., Yang, B., Garcia, E. A., & Shutao, L. (2008, June 1–8). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (pp. 1322–1328). IEEE.

Hanafy, M., & Ming, R. (2021). Comparing SMOTE Family Techniques in Predicting Insurance Premium Defaulting using Machine Learning Models. *International Journal of Advanced Computer Science and Applications, 12.* https://doi.org/10.14569/IJACSA.2021.0120970

Hou, G. D., Tong, D. L., Liew, S. Y., & Choo, P. Y. (2024, September 3–4). Exploring random forest regression for financial distress detection. In *2024 5th International Conference on Artificial Intelligence and Data Sciences (AiDAS)* (pp. 7–12). Aconf.

Lopez, J. A. (2004). The empirical relationship between average asset correlation, firm probability of default, and asset size. *Journal of Financial Intermediation, 13*(2), 265-283. https://doi.org/10.1016/S1042-9573(03)00045-7

Lu, H., & Hu, X. (2024). Enhancing Financial Risk Prediction for Listed Companies: A Catboost-Based Ensemble Learning Approach. *Journal of the Knowledge*

Economy, 15(2), 9824-9840. https://doi.org/10.1007/s13132-023-01601-5

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The journal of finance, 29*(2), 449-470. https://doi.org/10.1111/j.1540-6261.1974.tb03058.x

Mooney, P. (2020). *2020 Kaggle machine learning & data science survey. Kaggle.* https://www.kaggle.com/competitions/kaggle-survey-2020

Nguyen, M., Nguyen, B., & Liêu, M.-L. (2024). Corporate financial distress prediction in a transition economy. *Journal of Forecasting, 43*(8), 3128-3160. https://doi.org/10.1002/for.3177

Pham Vo Ninh, B., Do Thanh, T., & Vo Hong, D. (2018). Financial distress and bankruptcy prediction: An appropriate model for listed firms in Vietnam. *Economic Systems, 42*(4), 616-624. https://doi.org/10.1016/j.ecosys.2018.05.002

Tran, K. L., Le, H. A., Nguyen, T. H., & Nguyen, D. T. (2022). Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam. *Data, 7*(11), 160. https://doi.org/10.3390/data7110160

Tran, T., Nguyen, N. H., Le, B. T., Thanh Vu, N., & Vo, D. H. (2023). Examining financial distress of the Vietnamese listed firms using accounting-based models. *PloS one, 18*(5), e0284451. https://doi.org/10.1371/journal.pone.0284451

Triều, T. (2025). *There are still many reasons to be optimistic.* Vietnam.vn. https://www.vietnam.vn/en/van-co-nhieu-ly-do-de-lac-quan