Journal of Economics, Finance and Management Studies

ISSN (print): 2644-0490, ISSN (online): 2644-0504

Volume 07 Issue 12 December 2024

Article DOI: 10.47191/jefms/v7-i12-10, Impact Factor: 8.044

Page No: 7039-7048

Application of Machine Learning in Predicting Financial Distress Among Indonesian Banks

Bagas Saputra[1], Payamta Payamta[2]

[1,2]Faculty of Economic and Business, Sebelas Maret University, Surakarta, Indonesia

ABSTRACT: Banks play a crucial role in the economy. Consequently, systemic banking crises destabilize financial markets and hinder global economic growth. In this study, machine learning was used to predict bank distress in the Indonesian banking sector. Key variables relevant to the banking sector were identified. The data, spanning the period 2019 to 2023, was collected from Indonesian banks listed on the Indonesia Stock Exchange (IDX). Random Forest (RF) and XGBoost (XGB) models were employed to develop predictive frameworks, with their performance evaluated using overall prediction accuracy (OPA), Type I error, and Type II error. The results showed that the RF model outperformed the XGB model, achieving the highest accuracy and completely eliminating Type II errors. This study highlights the potential of machine learning to improve early warning systems for financial distress, contributing to the stability of the banking sector and the resilience of the economy.

KEYWORDS: Financial Distress, Bank, Machine Learning, XGBoost, Random Forest.

I. INTRODUCTION

The banking sector plays a critical role in economic stability, making financial distress prediction essential for both practitioners and policymakers. While traditional financial distress models provide valuable insights, they often lack the predictive power to address the complexities of the modern financial environment, especially in emerging markets like Indonesia (Purwanti et al., 2024). Current literature highlights significant advancements in machine learning for financial distress prediction (Kristanti et al., 2024). However, a clear gap remains regarding its application to Indonesian banking companies (Mahardini et al., 2022), where distinct economic and regulatory factors influence financial stability.

The banking system is essential to economic stability, as disruptions in its functions can impact the broader economy through money and credit distribution. Unlike crises in other industries, banking crises carry significant external costs, diminishing trust in the financial system and affecting other banks exposed to insolvent counterparts (Bhattacharya et al., 1998). This interdependence among banks means that individual bank failures can quickly escalate into systemic crises, with the potential for national and even global contagions (Benston & Kaufman, 1995). Reduced credit availability, interruptions in interbank lending, and declines in collateral value are factors that can drive this transmission, amplifying the impact across sectors (Laeven, 2011). The economic consequences of such crises include resource misallocation, reduced consumption, and decreased investment (Rajan & Zingales, 1998). Studies from the International Monetary Fund (1998) and the World Bank (2020) suggest that the fiscal cost of banking distress can reach up to 50% of a nation's annual GDP, reflected in lost output, rising unemployment, increased public debt, and fiscal expenditures aimed at supporting banks.

In this study, we try to predict the financial distress in the Indonesian banks that are listed on the Indonesian Stock Exchange (IDX) from 2019 to 2023 using XGBoost (XGB) and Random Forest (RF) models. This approach differs from previous research, which has primarily relied on traditional models (Purwanti et al., 2024) or focused on developed economies (Paule-Vianez et al., 2020). By examining Indonesia's banking sector, this study highlights how machine learning can enhance financial stability in emerging markets, offering valuable insights for predictive analytics in finance with both theoretical and practical applications. The main idea of this study is to evaluate the predictive power of advanced machine learning models, specifically XGB and RF, in identifying financial distress among Indonesian banks. This approach is founded on the theory that machine learning models can capture complex patterns in data that traditional statistical models may miss (Barboza et al., 2017; Florez-Lopez, 2007), especially in high-stakes areas like banking where early detection of distress is crucial. Empirically, this study follows the methodology outlined by Shrivastava et al. (2020), utilizing a set of carefully selected independent variables to analyze the predictive performance of XGB and RF models in identifying financial distress. Specifically, we incorporate financial ratios and indicators relevant to the banking sector, such as the measure of capital, asset quality, profitability metrics, and liquidity, which have been demonstrated as significant predictors of distress.

Previous studies have extensively examined the application of machine learning techniques in predicting financial distress. This study contributes to the research literature by extending the application of machine learning models specifically XGB and RF in financial distress prediction, focusing on the Indonesian banking sector. By demonstrating the effectiveness of these models within an emerging market context, this research broadens the understanding of machine learning's adaptability and robustness for predictive tasks in finance. In particular, these studies collectively illustrate the evolution of machine learning techniques in predicting financial distress. The result from Kristanti et al. (2024) shows that the RF and XGB models are much more accurate than the other models, such as support vector classification and long short-term memory. Chou (2019) demonstrated the benefits of hybrid models, while Gregova et al. (2020) subsequently reinforced the superiority of advanced models such as neural networks. Bräuning et al. (2020) emphasized the practical applications and effectiveness of the decision tree model, while Malakauskas (2021) and Shen et al. (2020) demonstrated the superiority of RF. The remaining part of the paper is structured as follows. The following section gives an overview of the previous literature on financial distress prediction models. Section 3 highlights the data, sample, and methodology followed by empirical results in Section 4. The last section covers the conclusion and discussion of the result outcomes.

II. LITERATURE REVIEW

Bankruptcy prediction is a vital and widely studied topic, with various statistical and analytical methods developed to forecast the likelihood of failure in banks and firms. Altman (1968) pioneered the field by introducing the Z-score model, which uses multivariate analysis and five key financial and economic indicators to assess bankruptcy risk. Later, Martin (1977) and Ohlson (1980) applied logistic regression to predict failures in firms and banks. Martin, for instance, studied U.S. commercial bank failures from 1970 to 1976, analyzing 25 financial ratios and finding that logistic regression outperformed linear discriminant analysis in classification accuracy. Thomson (1991) investigated U.S. bank failures in the 1980s using statistical methods, while Greuning & Iqbal (2007) employed financial ratio analysis, peer group analysis, and econometric models for early warning systems. Similarly, Canbas et al. (2005) demonstrated that discriminant analysis achieved better results than Probit and Tobit models in predicting the distress of Turkish commercial banks using 49 financial ratios.

In recent years, machine learning has revolutionized bankruptcy prediction by addressing the limitations of traditional statistical methods. Techniques like XGBoost, Random Forest, and Artificial Neural Networks have gained popularity due to their ability to handle complex datasets, identify nonlinear relationships, and adapt to various data structures without relying on strict assumptions. Unlike statistical models, which require predefined structures and parameter estimation, machine learning algorithms learn the model structure directly from the data (Wang et al., 2014). Additionally, statistical analysis often depends on assumptions such as normal data distribution and no correlation between variables, which can reduce predictive accuracy. Some empirical studies compare traditional and machine learning prediction methods. A study by Irvan & Supriyanto (2024) shows that machine learning models, such as RF and Convolutional Neural Networks, outperform traditional methods in predicting financial distress, achieving accuracies above 90%. Similarly, Rahman & Zhu (2024) applied machine learning techniques to develop financial distress prediction models for Chinese A-listed construction companies and compared their classification performance with conventional Z-Score models. Their results confirm that machine learning classifiers are more effective and accurate than Z-Score models in predicting financial distress for Chinese A-listed construction companies.

Extreme Gradient Boosting or XGBoost is a variant of the gradient tree boosting algorithm that uses the tree ensemble principle, which combines a set of weak classifiers to create a strong classifier. Gradient boosting and XGB have the same principle, but XGB provides better performance by controlling the complexity of the tree using various regularization techniques (Patrous, 2018). XGB is effectively suitable for sparse data because it uses a sparsity-aware split-finding approach (Chemura et al., 2020). XGB has been increasingly applied in the field of financial distress prediction in recent years. Carmona et al. (2019) predicted bank failures in the U.S. banking sector and found that XGB demonstrated greater predictive power compared to both logistic regression and RF methods.

The random forest technique is based on decision tree models, also known as generalised classification and regression trees' (CART). The model created by Breiman (2001) has a level of precision similar to that of AdaBoost and, depending on the set, can provide better results than boosting can (Kruppa et al., 2013). RF has emerged as a robust tool for predicting financial distress across various contexts, demonstrating superior performance compared to traditional methods. A study by Barboza & Altman (2024) indicates that RF outperforms logistic regression in predicting corporate distress, demonstrating higher predictive power and lower error rates. This machine learning algorithm effectively handles complex datasets and identifies critical financial indicators, making it valuable for financial analysts and regulators. Additionally, RF not only classifies observations but also highlights the importance of variables driving group separation (Maione et al., 2016).

III. METHODOLOGY

This study uses secondary data, specifically financial reports, selected through purposive sampling. The data is collected for Indonesian banks that listed on the IDX over the period of 2019–2023 and is categorized into two categories distress or safe. In this study, a bank is classified as distressed if it has incurred negative net income over the past two years (Platt, 2008). Feature selection is used to assess the health of the bank, with the selected features detailed in Table 1. The number of banks in the data set was 215 out of which 195 banks were under the safe category and 20 were under the distress category. The dataset in this study is imbalanced, with significantly fewer distressed banks compared to safe banks in Indonesia. Imbalanced classification occurs when one class significantly outnumbers another, leading to uneven class distribution in the dependent variable (Chawla, 2005; Sun et al., 2009). To address this, the Synthetic Minority Oversampling Technique (SMOTE) was applied, a method that generates artificial samples based on feature similarities within the minority class. Unlike traditional oversampling, which duplicates data, SMOTE effectively balances the dataset, enhancing the model's ability to classify minority instances accurately.

Table 1. Data Description

Name	Definition
Bank Status	Binary indicator equivalent to 1 for the financial distress banks, if the bank has not experienced financial distress, the variable is assigned a value of 0.
Total Assets	Cash and assets due to banks, total earning assets, foreclosed real estate, fixed assets, and other assets.
Total Liabilities	Total assets minus total equity.
Total Equity	Common equity, non-controlling interest, securities revaluation reserves and foreign exchange revaluation reserves.
Total Provision	Net loans minus reserves for impaired loans.
Total Deposits	Customer deposits, bank deposits, other deposits, and short-term borrowings.
Total Capital	The sum of tier 1 capital and tier 2 capital.
Return on Assets	Net Income/Average Total Assets.
Net Income	Post-tax profit.
Other Operating income	Any other sustainable income which is related to the company’s core business.
Overheads	Personnel and other operating costs.
Loan Loss Reserves/Loans	It signifies how much funds have been put apart for potential losses.
Total Equity/Total Assets (TE/TA)	Evaluates the amount of security the bank enjoys by its equity.
Total Equity/Net Loans (TE/NL)	Measures the equity insulation available to take up losses on the loan manuscript.
Total Equity/Total Deposits (TE/TD)	Estimates the amount of everlasting funding relative to undersized funding.

Total Equity/Total Liabilities (TE/TL)	Also identified as the capitalization ratio and it is the inverse of the leverage ratio.
Cost/Income	Estimates the costs of managing the bank, the main element of salaries, as a proportion of income produced before provisions.
Net Loans/Total Assets (NL/TA)	Reveals what proportion of the resources of the bank are coupled up in loans.
Net Interest Income (NIM)	Net interest income calculated as a percentage of earning assets.

Source: Compiled by the authors

This study employed XGB and RF models and compared their performance by analyzing overall prediction accuracy (OPA), Type I error, and Type II error. A threshold of 0.3 was applied during classification to enhance the models' sensitivity, prioritizing the identification of distressed banks. This decision was made because reducing Type II errors (misclassifying distressed banks as safe) is critical for maintaining banking sector stability and preventing financial crises. These metrics were derived from the confusion matrix, also known as the error matrix which includes:

Table 2. Confusion Matrix

Actual	**Model Prediction**	
	Distressed	Safe
Distressed	TP	FN
Safe	FP	TN

Source: Compiled by the authors.

Where:

- TP (True Positive): Distressed banks correctly predicted as distressed.
- FP (False Positive): Safe banks incorrectly predicted as distressed.
- TN (True Negative): Safe banks correctly predicted as safe.
- FN (False Negative): Distressed banks incorrectly predicted as safe.

Overall prediction accuracy refers to the percentage of cases the model correctly predicts.

$$\text{OPA} = \frac{TP + TN}{TP + FP + TN + FN}$$

Type I error measures the percentage of safe banks wrongly classified as distressed, while Type II error measures the percentage of distressed banks wrongly classified as safe. Therefore, the two errors can be stated as follows:

$$\text{Type I Error} = \frac{FP}{FP+TN}$$

$$\text{Type II Error} = \frac{FN}{TP+FN}$$

Additionally, feature importance identification was conducted to determine the most influential variables contributing to the prediction of financial distress.

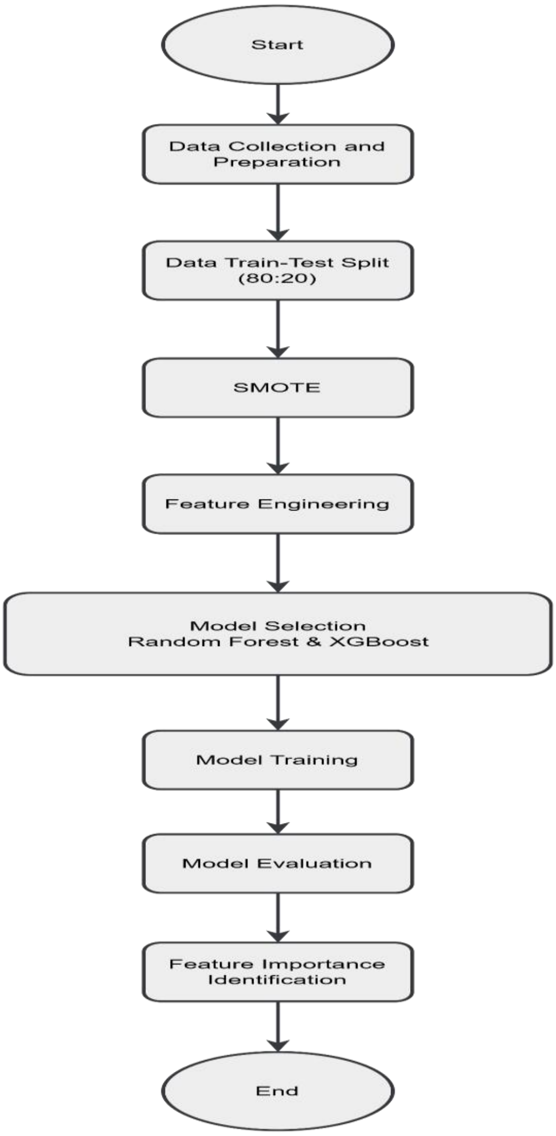

Figure 1. Flowchart outline of the study

IV. RESULTS AND DISCUSSION

Table 3 provides us with the descriptive statistics of the bank-specific variables considered in this study, as presented in Table 1. This study is based on an Indonesian bank that is listed on the IDX. Variables such as total assets, total liabilities, total equity, total provisions, deposits, total capital, net income, other operating income, and overheads are reported in millions of rupiah, while the remaining variables are expressed as financial ratios. Table 3 highlights considerable variation in size, profitability, and financial stability. Wide ranges in metrics such as total assets (mean: 198223625.33, maximum: 2174219449) and total liabilities (mean: 162391202.84, maximum: 1660442815) indicate the existence of both big and small banks in the dataset.

Table 3.

Bank Specific Variables	Mean	Median	Std. Deviation	Minimum	Maximum
Total Assets	198223625.33	26127820.00	414960073.696	953737	2174219449
Total Liabilities	162391202.84	21932578.00	336818355.451	248249	1660442815
Total Equity	30060648.00	4691449.00	62150502.627	299766	316472142
Total Provision	107648918.86	13579434.00	221426339.449	218547	1253634957

Deposits	144165109.65	20343719.00	298787304.234	233307	1441368542
Total Capital	27659591.11	4630113.00	54106855.513	206080	250568767
ROA	0.006081	0.008300	0.0282755	-0.1589	0.0478
Net Income	3485104.81	114941.00	10168330.469	-6055703	60425048
Other Operating Income	3217303.36	266400.00	8308486.028	1273	45625785
Overheads	6296288.94	921249.00	13473780.377	20440	76782291
Loan Loss Reserves/Loans	0.040286	0.031300	0.0322216	0.0027	0.2168
TE/TA	0.212799	0.165100	0.1458663	0.0553	0.9251
TE/NL	0.436781	0.310700	0.4157288	0.0809	3.7304
TE/TD	0.464963	0.223600	1.1835819	0.0731	13.1387
TE/TL	0.422351	0.203000	1.0658442	0.0622	12.3479
Cost/Income	0.846120	0.652200	1.0634899	0.2369	13.0042
NL/TA	0.543255	0.553700	0.1230882	0.1033	0.8009
NIM	0.046800	0.044200	0.0295793	-0.0352	0.2023

Source: Calculated by the authors

In Table 4, we compare the mean difference in accounting information between safe and distressed banks. At the 1% significance level, distressed banks have much lower total assets (23217682 vs. 216172953), total liabilities (19318331 vs. 177065344), and total equity (3768868 vs. 32757241). Lower profitability and inefficiencies are evident in distressed banks, which report a negative return on assets (ROA: -4.56% vs. 1.14%), net income (-965846 vs. 3965613), lower NIM (3.45% vs. 4.81%), and higher cost-to-income ratios (221.7% vs. 70.6%). Overall, distressed banks reveal significantly weaker financial performance, lower profitability, and greater inefficiencies compared to safe banks, as reflected by key metrics such as total assets, ROA, net income, and cost-to-income ratios.

Table 4.

Bank Specific Variables	**Safe**	**Distress**
	216172953	23217682***
Total Liabilities	177065344	19318331***
Total Equity	32757241	3768868***
Total Provision	117469548	11897784***
Deposits	157217007	16909107***
Total Capital	30131231	3561100***
ROA	0.0114	-0.0456***
Net Income	3941613	-965846**
Other Operating Income	3526555	202103***
Overheads	6824275	1148427***
Loan Loss Reserves/Loans	0.0373	0.0696**
TE/TA	0.2116	0.2242
TE/NL	0.4244	0.5577**
TE/TD	0.4735	0.3819
TE/TL	0.4302	0.3458
Cost/Income	0.7055	2.2167***
NL/TA	0.5473	0.5036
NIM	0.0481	0.0345**

(***,**, implies significance at the 1% and 5% respectively)

Source: Calculated by the authors

The dataset has 215 records in which the proportion of minority and majority classes is 9.3% and 90.7% respectively. The imbalanced data set was divided into a two-part train and test in the proportion of 80% and 20%. To address class imbalance, SMOTE were used. We compared the overall accuracy of XGB and RF models, along with their Type I and Type II errors. Type I error refers to mistakenly classifying safe banks as distressed, while Type II error is when distressed banks are incorrectly classified as safe. As shown in Table 5, the RF model achieved the highest overall accuracy at 98%, with a Type I error of 3% and a significantly lower Type II error of 0%. In comparison, the XGB model demonstrated slightly lower accuracy at 95%, with a Type I error of 3% and a Type II error of 25%. Since Type II error is more serious for banks, because misjudging which banks are moving toward bankruptcy can create bigger problems over time, therefore the best prediction model is the one with the highest overall accuracy and the lowest Type II error rate (Shrivastava et al., 2020). Based on these criteria, the RF model is the better choice. It not only achieved the highest accuracy but also completely eliminated Type II errors compared to the XGB model. This makes RF more effective at identifying distressed banks, reducing the risk of missing those on the verge of financial trouble, which is essential for maintaining the stability of the banking sector and preventing broader financial crises.

Table 5. Model Comparison

Parameter	XGBoost	Random Forest
Accuracy	95%	98%
Type I Error	3%	3%
Type II Error	25%	0%

Source: Calculated by the authors

Understanding feature importance is crucial for interpreting predictions in machine learning models, especially after fine-tuning their hyperparameters. Analyzing feature importance in the RF and XGB models reveals key differences in how they prioritize variables for predicting financial distress in Indonesian banks. As shown in Figure 2, the RF model identifies ROA, Cost/Income, Net Income as the most important features, indicating that profitability and efficiency metrics are central to its predictions. In contrast, as shown in Figure 3, the XGBoost model places overwhelming importance on ROA, with Cost/Income and TE/TL as secondary contributors. These findings show that both models heavily rely on profitability and efficiency indicators but differ in how they weigh other financial ratios. XGBoost focuses primarily on ROA as the dominant feature, while RF spreads importance more evenly across multiple variables. Ultimately, both models confirm that ROA is the most influential feature for predicting financial distress, highlighting its strong impact on model performance.

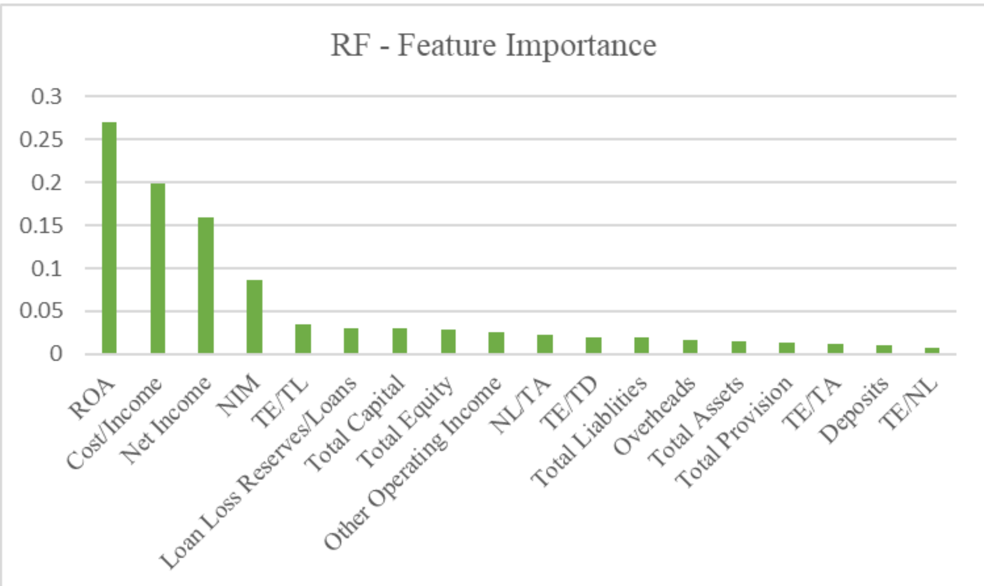

Figure 2. Feature importance of Random Forest

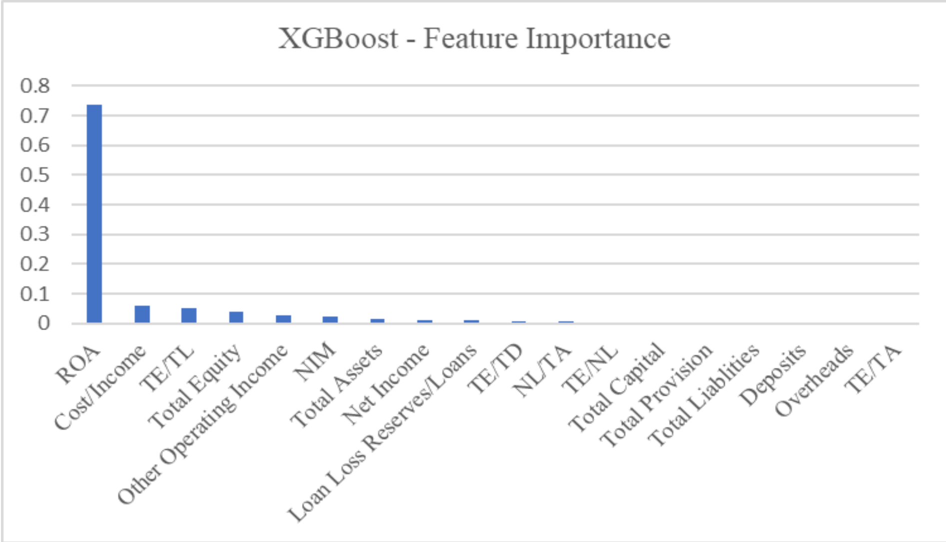

Figure 3. Feature importance of XGBoost

CONCLUSIONS

This study provides valuable insights for financial institutions, regulators, and policymakers by improving tools for detecting financial distress in Indonesia's banking sector. Accurate prediction models like XGB and RF can help banks identify and address financial risks early, supporting overall stability. For regulators, these models offer a data-driven approach to monitor financial health, enabling proactive supervision and regulation. The findings support better decision-making in risk management and policy creation, ensuring the resilience of Indonesia's banking system amid economic fluctuations. Among the models tested, RF demonstrated superior performance, achieving the highest accuracy and completely eliminating Type II errors, making it especially effective in identifying distressed banks.

However, this study has limitations. The relatively small sample size and inherent data imbalance, even with SMOTE applied, may impact the models' reliability and generalizability. Additionally, complex models like RF and XGB are prone to overfitting, especially with limited data. The data collection period (2019–2023) spans various economic cycles, including the effects of the COVID-19 pandemic, which may influence observed financial distress patterns. Future research should aim to use larger, more balanced datasets, incorporate macroeconomic variables, and consider the effects of different economic cycles to enhance model robustness and applicability.

REFERENCES

1) 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/10.2307/2978933
2) World Bank. (2020). Global economic prospects. *World Bank Group*.
3) Barboza, F., & Altman, E. (2024). Predicting financial distress in Latin American companies: A comparative analysis of logistic regression and random forest models. *The North American Journal of Economics and Finance*, *72*, 102158. https://doi.org/https://doi.org/10.1016/j.najef.2024.102158
4) Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. https://doi.org/https://doi.org/10.1016/j.eswa.2017.04.006
5) Benston, G. J., & Kaufman, G. G. (1995). Is the banking and payments system fragile? *Journal of Financial Services Research*, *9*(3–4). https://doi.org/10.1007/BF01051747
6) Bhattacharya, S., Boot, A. W. A., & Thakor, A. V. (1998). The Economics of Bank Regulation. *Journal of Money, Credit and Banking*, *30*(4). https://doi.org/10.2307/2601127
7) Bräuning, M., Malikkidou, D., Scalone, S., & Scricco, G. (2020). A New Approach to Early Warning Systems for Small European Banks. In G. Nicosia, V. Ojha, E. La Malfa, G. Jansen, V. Sciacca, P. Pardalos, G. Giuffrida, & R. Umeton (Eds.), *Machine Learning, Optimization, and Data Science* (pp. 551–562). Springer International Publishing.
8) Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.
9) Canbas, S., Cabuk, A., & Kilic, S. B. (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, *166*(2), 528–546. https://doi.org/https://doi.org/10.1016/j.ejor.2004.03.023

10) Carmona, P., Climent, F., & Momparler, A. (2019). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, *61*, 304–323. https://doi.org/https://doi.org/10.1016/j.iref.2018.03.008
11) Chawla, N. V. (2005). Data Mining for Imbalanced Datasets: An Overview. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 853–867). Springer US. https://doi.org/10.1007/0-387-25465-X_40
12) Chemura, A., Rwasoka, D., Mutanga, O., Dube, T., & Mushore, T. (2020). The impact of land-use/land cover changes on water balance of the heterogeneous Buzi sub-catchment, Zimbabwe. *Remote Sensing Applications: Society and Environment*, *18*, 100292. https://doi.org/https://doi.org/10.1016/j.rsase.2020.100292
13) Chou, T.-N. (2019). An Explainable Hybrid Model for Bankruptcy Prediction Based on the Decision Tree and Deep Neural Network. *2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII)*, 122–125. https://doi.org/10.1109/ICKII46306.2019.9042639
14) Florez-Lopez, R. (2007). Modelling of insurers' rating determinants. An application of machine learning techniques and statistical models. *European Journal of Operational Research*, *183*(3), 1488–1512. https://doi.org/https://doi.org/10.1016/j.ejor.2006.09.103
15) IMF. (1998). *World Economic Outlook. Financial Crises: Causes and Indicators*. International Monetary Fund Washington, DC.
16) Gregova, E., Valaskova, K., Adamko, P., Tumpach, M., & Jaros, J. (2020). Predicting Financial Distress of Slovak Enterprises: Comparison of Selected Traditional and Learning Algorithms Methods. *Sustainability*, *12*(10). https://doi.org/10.3390/su12103954
17) Greuning, H. Van, & Iqbal, Z. (2007). Banking and the risk environment. *Islamic Finance: The Regulatory Challenge*, 9–39.
18) Irvan, F., & Supriyanto, R. (2024). Comparative Analysis Of Machine Learning And Deep Learning Models Integrated With Altman Z-Score For Financial Distress Prediction In Companies Listed On The Indonesia Stock Exchange (IDX). *EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi Dan Bisnis*, *12*(2). https://doi.org/10.37676/ekombis.v12i2.5478
19) Kristanti, F. T., Febrianta, M. Y., Salim, D. F., Riyadh, H. A., & Beshr, B. A. H. (2024). Predicting Financial Distress in Indonesian Companies using Machine Learning. *Technology & Applied Science Research*, *14*(6), 17644–17649. https://doi.org/10.48084/etasr.8520
20) Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, *40*(13), 5125–5131. https://doi.org/https://doi.org/10.1016/j.eswa.2013.03.019
21) Laeven, L. (2011). Banking crises: A review. *Annual Review of Financial Economics*, *3*. https://doi.org/10.1146/annurev-financial-102710-144816
22) Mahardini, N. Y., Bandi, B., Payamta, & Honggowati, S. (2022). Financial Distress and Bankruptcy Prediction Research in Indonesia: Last 20 Years. *Proceedings of the International Colloquium on Business and Economics (ICBE 2022)*, 231–240. https://doi.org/10.2991/978-94-6463-066-4_20
23) Maione, C., de Paula, E. S., Gallimberti, M., Batista, B. L., Campiglia, A. D., Jr, F. B., & Barbosa, R. M. (2016). Comparative study of data mining techniques for the authentication of organic grape juice based on ICP-MS analysis. *Expert Systems with Applications*, *49*, 60–73. https://doi.org/https://doi.org/10.1016/j.eswa.2015.11.024
24) Malakauskas, A., & Lakstutiene, A. (2021). Financial distress prediction for small and medium enterprises using machine learning techniques. *Engineering Economics*, *32*(1), 4–14. https://doi.org/10.5755/j01.ee.32.1.27382
25) Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, *1*(3), 249–276. https://doi.org/https://doi.org/10.1016/0378-4266(77)90022-X
26) Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131. https://doi.org/10.2307/2490395
27) Patrous, Z. S. (2018). *Evaluating XGBoost for User Classification by using Behavioral Features Extracted from Smartphone Sensors*. KTH Royal Institute of Technology.
28) Paule-Vianez, J., Gutiérrez-Fernández, M., & Coca-Pérez, J. L. (2020). Prediction of financial distress in the Spanish banking system: An application using artificial neural networks. *Applied Economic Analysis*, *28*(82), 69–87. https://doi.org/10.1108/AEA-10-2019-0039
29) Platt, H. D., & Platt, M. B. (2008). Financial Distress Comparison Across Three Global Regions. *Journal of Risk and Financial Management*, *1*(1), 129–162. https://doi.org/10.3390/jrfm1010129

30) Purwanti, L., Arief Effendi, S., Ibrahim, M., Tam Cahyadi, R., & Prakoso, A. (2024). Prediction and Comparison of Bankruptcy Models in Banking Sector Companies in Indonesia. *Educational Administration: Theory and Practice*, *30*(4), 4428–4440. https://doi.org/10.53555/kuey.v30i4.2224
31) Rahman, M. J., & Zhu, H. (2024). Predicting financial distress using machine learning approaches: Evidence China. *Journal of Contemporary Accounting & Economics*, *20*(1), 100403. https://doi.org/https://doi.org/10.1016/j.jcae.2024.100403
32) Rajan, R. G., & Zingales, L. (1998). Power in a theory of the firm. *Quarterly Journal of Economics*, *113*(2). https://doi.org/10.1162/003355398555630
33) Shen, F., Liu, Y., Wang, R., & Zhou, W. (2020). A Dynamic Financial Distress Forecast Model with Multiple Forecast Results under Unbalanced Data Environment. *Knowledge-Based Systems*, *192*, 105365. https://doi.org/https://doi.org/10.1016/j.knosys.2019.105365
34) Shrivastava, S., Jeyanthi, P. M., & Singh, S. (2020). Failure prediction of Indian Banks using SMOTE, Lasso regression, bagging and boosting. *Cogent Economics & Finance*, *8*(1). https://doi.org/10.1080/23322039.2020.1729569
35) Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of Imbalanced Data: A Review. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*(04), 687–719. https://doi.org/10.1142/S0218001409007326
36) Thomson, J. (1991). Predicting bank failures in the 1980s. *Economic Review*, *27*(Q I), 9–20. https://EconPapers.repec.org/RePEc:fip:fedcer:y:1991:i:qi:p:9-20:n:v.27no.1
37) Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*, *41*(5), 2353–2361. https://doi.org/https://doi.org/10.1016/j.eswa.2013.09.033