

Constructing a Financial Distress Prediction Model for Listed and OTC Companies in Taiwan Using Machine Learning Techniques

Chen Lisi, Huang Zhao-Ying, and Lin Chien-Chih

Department of Banking and Finance, Faculty of Business and Management, Tamkang University, Tamsui Campus,
New Taipei City, Taiwan

*Corresponding author, E-mail: sd22348683@gmail.com

Abstract

Addressing the diminishing efficacy of linear models in financial distress prediction, this study leverages Taiwanese corporate data (2020-2024). To overcome the scarcity of "full-cash-delivery" stocks, a longitudinal panel dataset was constructed by tracking 45 unique firms (15 distressed and 30 healthy) over multiple quarters. This approach yielded 844 firm-quarter observations, providing sufficient data volume for robust model training and validation. The study compares a Logistic Regression baseline against a Random Forest (RF) and a Deep Neural Network (DNN). Using "full-cash-delivery stock" as the distress indicator, models are evaluated via metrics including AUC, precision, and recall. Empirical results demonstrate the superior predictive power of non-linear models (RF and DNN), which achieve performance metrics exceeding 96% and identify profitability indicators (EBT, ROA) as key risk determinants. This research concludes by proposing a novel two-stage hybrid framework that synergizes the high sensitivity of DNN for initial screening with the robustness of RF for confirmation, offering a more pragmatic and effective risk management paradigm.

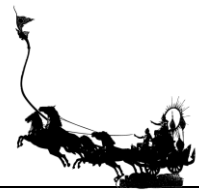
Keywords: *Financial Distress Prediction, Machine Learning, Random Forest, Logistic Regression, Deep Learning Network*

1. Introduction

Since the onset of the COVID-19 pandemic in 2019, the global economic system has grappled with multiple shocks, including supply chain disruptions, commodity price volatility, and persistent inflation, posing unprecedented challenges to corporate stability. This macroeconomic turmoil has amplified the financial fragility of corporations, a trend acutely visible in Taiwan's capital market. In recent years, a growing number of listed companies have been designated as "full-cash-delivery stocks" or have even faced delisting, underscoring the inadequacy of existing early-warning systems.

Historically, financial distress prediction was dominated by linear models like Multivariate Discriminant Analysis (Altman, 1968) and Logistic Regression (Ohlson, 1980). However, the efficacy of these static ratios has been challenged in the modern era. Barboza et al. (2017) highlighted that traditional models struggle to capture the complex, non-linear dynamics of contemporary markets. Consequently, a paradigm shift toward Artificial Intelligence (AI) has emerged. Early applications of neural networks demonstrated significant advantages over traditional statistical methods in forecasting bank failures (Tam & Kiang, 1992). As machine learning evolved, ensemble techniques such as neural network ensembles (Tsai & Wu, 2008) and Random Forests (Breiman, 2001) were proven to substantially increase predictive accuracy by mitigating overfitting (Kim & Kang, 2010). A comprehensive review by Alaka et al. (2018) further highlighted the necessity of selecting appropriate AI tools to handle specific financial data structures. More recently, the focus has shifted to deep learning, where advanced architectures are utilized to extract complex latent risk features from high-dimensional corporate data (Mai et al., 2019). Furthermore, deep learning models, such as residual networks (He et al., 2016), alongside robust algorithms like Support Vector Machines (Lin et al., 2011), have advanced the capability to capture critical signals, emphasizing that AI is essential for contemporary financial governance.

Against this backdrop, the ascent of Artificial Intelligence (AI) and big data analytics presents a new paradigm for resolving this long-standing challenge. Advanced machine learning algorithms, such as Random Forest (RF) and Deep Neural Networks (DNN), are adept at handling high-dimensional, non-linear data,



possessing the capability to automatically mine latent patterns and complex correlations. The core objective of this study is to systematically investigate and compare the efficacy of various machine learning models in predicting the financial distress of companies listed in Taiwan. To this end, the research establishes a progressive comparative framework, employing Logistic Regression as a performance baseline, introducing Random Forest as a high-performance benchmark, and exploring a DNN to assess the potential of end-to-end feature learning capabilities.

Within this framework, the study conducts a rigorous comparison utilizing key performance indicators such as Precision, Recall, and AUC to identify the most suitable predictive model. Beyond predictive accuracy, this study further employs methods like Feature Importance analysis to investigate the key financial drivers affecting the stability of Taiwanese listed companies. The overall framework and process are illustrated in Figure 1. The research commences by establishing a theoretical foundation, followed by a systematic evaluation and comparison of the models' predictive results. Finally, based on empirical findings, the study draws conclusions and proposes managerial implications for regulators, investors, and corporate managers in navigating an increasingly complex financial landscape.

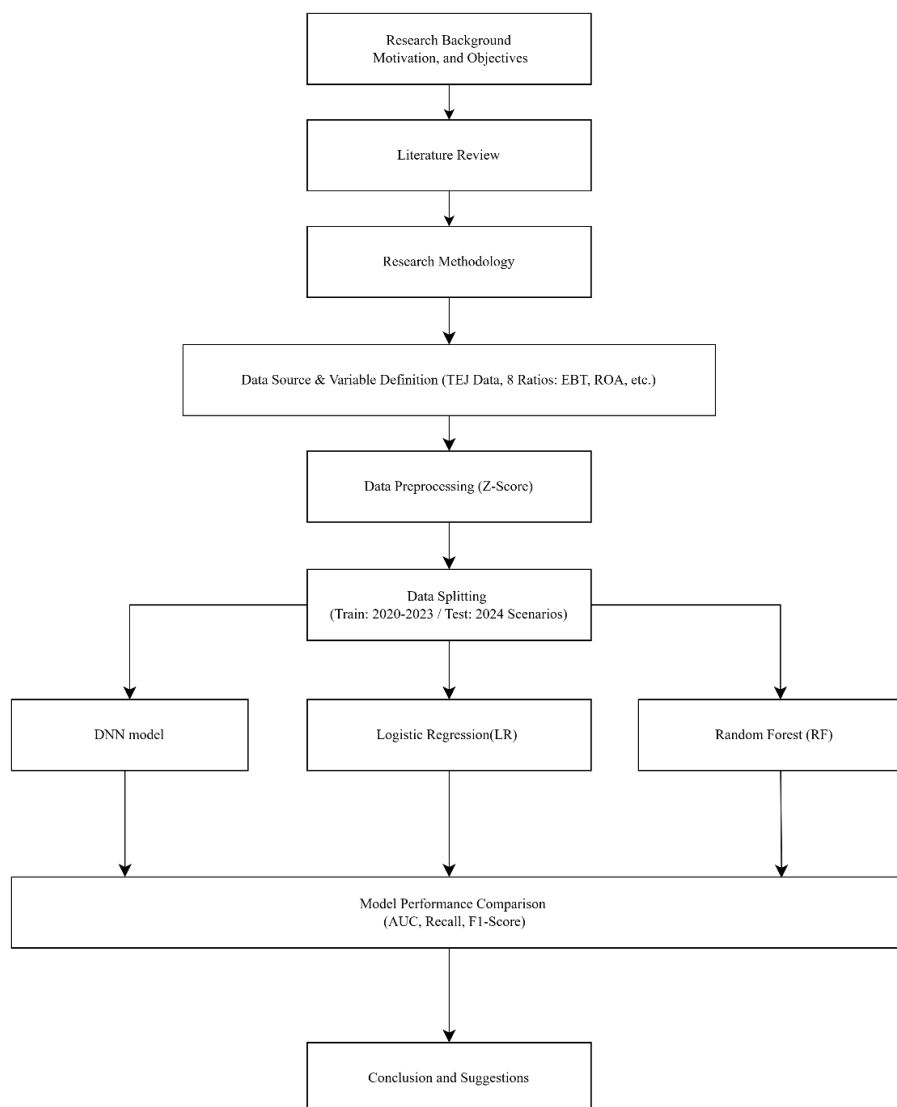


Figure 1 Research Framework and Flowchart



2. Objectives

- 1) To systematically compare the efficacy of Logistic Regression, Random Forest, and Deep Neural Networks in predicting the financial distress of companies listed in Taiwan.
- 2) To evaluate the predictive performance of these models using measurable indicators, specifically Precision, Recall, and the Area Under the Curve (AUC).
- 3) To identify and analyze the critical financial drivers affecting corporate stability through Feature Importance analysis.
- 4) To establish a robust AI-driven evaluation framework that provides reliable decision-making support for financial regulators and investors.

3. Materials and Methods

3.1 Data Sourcing, Preprocessing, and Variables

The empirical data for this study is primarily sourced from the Taiwan Economic Journal (TEJ) database, encompassing quarterly financial data for listed companies on the TWSE and TPE from 2020 to 2024. To overcome the inherent scarcity of distressed samples in the market, a longitudinal panel dataset was constructed. The final sample comprises 45 unique firms, consisting of 15 companies designated as "full-cash-delivery stocks" (distressed samples, $Y=1$) and 30 continuous trading firms (healthy samples, $Y=0$), maintaining a 1:2 ratio. By tracking these firms over 16 consecutive quarters, the study generated a total of 844 valid firm-quarter observations.

To mitigate biases from differing scales, Z-Score Standardization was applied to all numerical features. Regarding feature selection, this study adheres to the classic frameworks of Altman (1968) and Ohlson (1980), prioritizing solvency and profitability ratios. Furthermore, consistent with recent ML-based studies in the Taiwanese context (Lin et al., 2011; Tkáč & Verner, 2016), eight key financial indicators across four dimensions were selected as predictive variables, as detailed in Table 1.

Table 1 Definition of Research Variables

Symbol	Indicator Name	Description / Calculation
X1	Current Ratio	Measures short-term solvency
X2	Debt Ratio	Measures capital structure stability
X3	Return on Assets (ROA)	Measures asset utilization efficiency
X4	Return on Equity (ROE)	Measures return on shareholders' investment
X5	EBT (Earnings Before Tax)	A key indicator of core profitability
X6	Cash Equivalents	Reflects immediate liquidity status
X7	Revenue Growth Rate	Assesses operating growth momentum
X8	Asset Growth Rate	Reflects the trend of firm size expansion

3.2 Research Models and Hyperparameters

To construct a robust framework, three representative supervised learning models were employed. Logistic Regression (LR) serves as the baseline using L2 regularization and the 'liblinear' solver, a standard approach for establishing linear benchmarks (Hosmer & Lemeshow, 2000).

For the non-linear benchmark, Random Forest (RF) was selected due to its proven efficacy in handling noisy financial data and preventing overfitting through bagging mechanisms (Breiman, 2001; Barboza et al., 2017). The model aggregates multiple decision trees, with $n_{estimators}$ ranging from 100 to 1000.

Finally, a bespoke Deep Neural Network (DNN) with 30 epochs and the Adam optimizer was designed. This architecture incorporates Residual Connections (He et al., 2016) to capture complex, non-linear risk signals, a methodology supported by recent advancements in applying artificial neural networks to business applications (Tkáč & Verner, 2016).

**Table 2** Summary of Model Hyperparameter

Model Type	Model Name	Key Parameter	Value / Setting	Description
Baseline	Logistic Regression (LR)	Regularization	L2 (Ridge)	Penalty term (Ridge Regression) to prevent overfitting
		Solver	'liblinear'	Algorithm optimized for smaller datasets
Ensemble	Random Forest (RF)	n_estimators	100 - 1000	Total number of trees in the forest
		max_depth	10 - 40	Maximum depth per tree
Deep Learning	Deep Neural Network (DNN)	Optimizer	Adam	Adaptive learning rate optimization
		Epochs	30	Training duration (passes through data)
		Loss Function	BCEWithLogitsLoss	Binary Cross Entropy for classification

4. Results and Discussion

4.1 Performance Comparison of Predictive Models

This section elucidates the fundamental statistical characteristics of the dataset. Serving as the empirical cornerstone of this research, Table 3 delineates the descriptive statistics for the eight key financial indicators (X1-X8). These metrics, derived from the Training Set (N=703), underwent Z-score standardization to ensure numerical stability. As shown in the table, the significant variation in profitability metrics (EBT, ROA) suggests that these variables possess strong discriminatory potential for distinguishing between distressed and healthy firms.

Table 3 Descriptive Statistics of Research Variables

Variable	Definition	Mean	Std. Dev.	Min	Max
X1	Current Ratio	184.37	285.06	5.88	4,213.23
X2	Debt Ratio	54.90	19.92	2.76	94.54
X3	ROA	0.37	7.60	-43.52	27.09
X4	ROE	0.07	18.46	-98.06	80.14
X5	EBT	33,152,298	159,066,152	1,125	1,465,428,000
X6	Cash Equiv.	144.03	2,368.54	-98.32	56,552.24
X7	Rev. Growth	4.31	32.08	-52.13	509.86
X8	Asset Growth	2.83	5.51	-6.10	44.13

4.2 Performance Comparison of Predictive Models

The empirical analysis revealed a significant performance gap between linear and non-linear architectures, aligning with the comparative studies by Barboza et al. (2017) and Alaka et al. (2018). Random Forest (RF) and Deep Neural Networks (DNN) demonstrated superior predictive capabilities, while the Logistic Regression (LR) baseline failed to catch most distressed signals without manual intervention. The specific average risk probabilities for each test group across four forecasting horizons are summarized in Table 4.

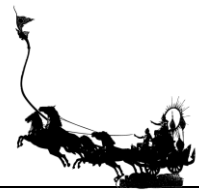


Table 4 Multi-Period Forecasted Average Risk Probability Comparison

Model	Group	3 Months	6 Months	9 Months	12 Months
DNN	A1 (Healthy)	21.81%	22.81%	23.79%	25.61%
	A2 (Mixed)	45.34%	45.58%	44.44%	43.88%
	A3 (Distressed)	71.88%	71.19%	69.38%	71.36%
LR	A1 (Healthy)	30.43%	31.56%	32.68%	34.64%
	A2 (Mixed)	59.55%	58.72%	58.75%	59.05%
	A3 (Distressed)	80.27%	79.06%	79.20%	81.23%
RF	A1 (Healthy)	14.89%	15.37%	15.76%	17.13%
	A2 (Mixed)	38.63%	37.53%	37.01%	36.58%
	A3 (Distressed)	75.61%	73.97%	72.75%	73.11%

As detailed in Table 4, the Distressed Group (A3) consistently exhibited the highest risk probabilities across all horizons, which validated the models' capacity to identify critical risk factors associated with financial insolvency. Notably, the Random Forest model maintained the lowest risk percentages for the Healthy Group (A1), ranging from 14.89% to 17.13%. This showcases its exceptional robustness in filtering out market noise, a characteristic advantage of ensemble techniques as noted by Kim & Kang (2010).

The discriminatory power of these models was further quantified through ROC curve analysis. As illustrated in Figure 2 and Figure 3, the RF model achieved a near-perfect AUC value of 0.997, significantly outperforming the unweighted LR baseline. While the LR model showed improved results after class-weight intervention, it remained less autonomous than the ensemble algorithm. Furthermore, the training stability of the deep learning architecture was monitored via the learning curves presented in Figure 4, where the convergence of loss functions confirmed the model's ability to assimilate complex, non-linear patterns within the financial dataset. These empirical findings solidly corroborate the assertions made by Barboza et al. (2017) and Tkáč & Verner (2016), demonstrating that machine learning algorithms significantly outperform traditional statistical baselines in capturing corporate bankruptcy dynamics.

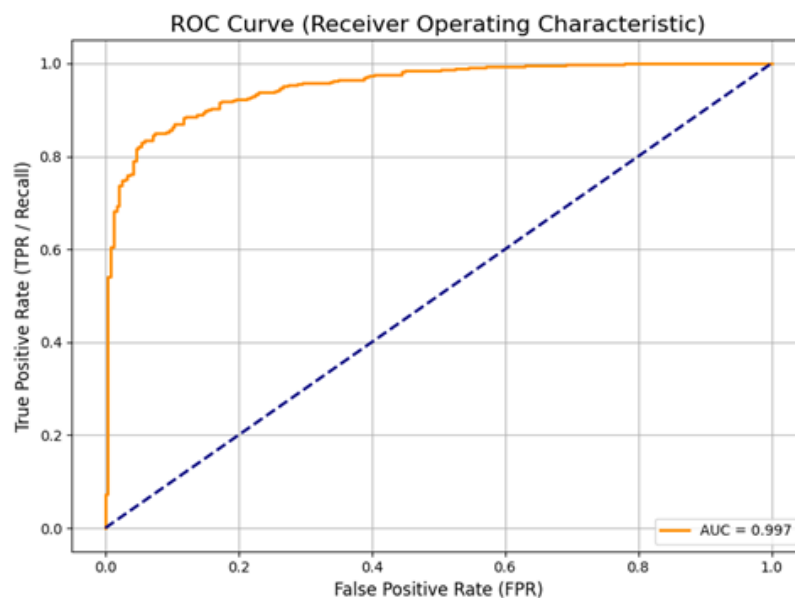


Figure 2 ROC Curve and AUC Value for the RF Model

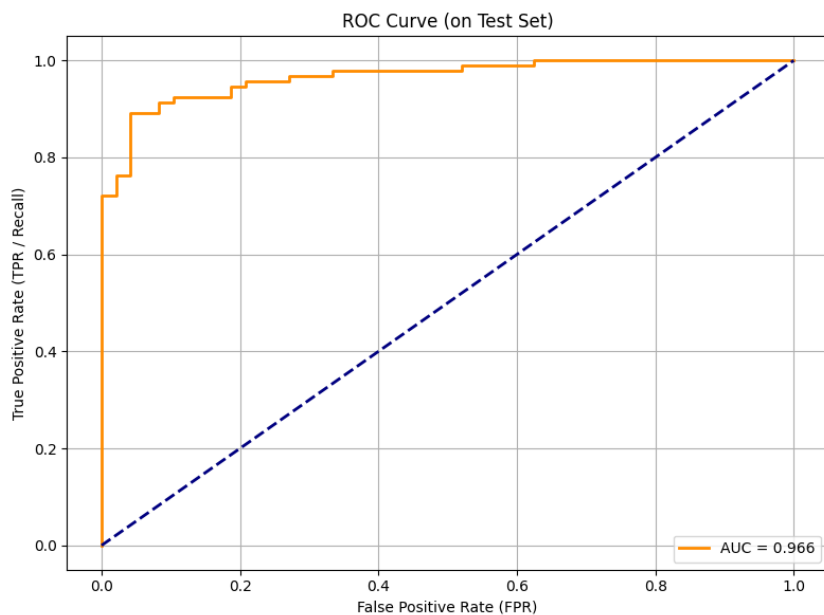
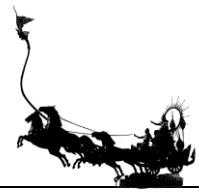


Figure 3 ROC Curve and AUC Value of the Single-Layer Neural Network (LR)

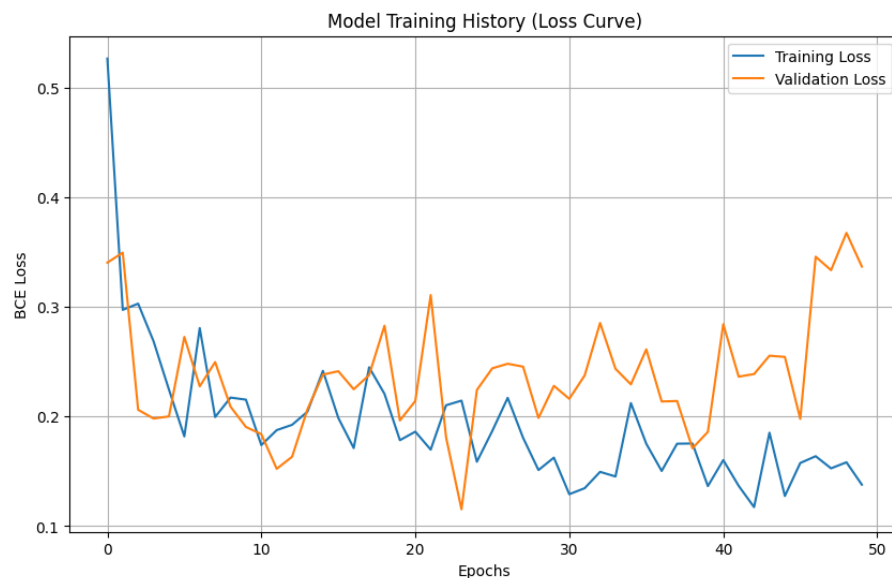


Figure 4 Training Process Learning Curve for the DNN Model

4.3 Feature Importance Analysis

Beyond predictive performance, the specific contribution of individual financial indicators was examined through feature importance analysis within the Random Forest model. As illustrated in Figure 5, "EBT" and "ROA" emerged as the most influential predictors, capturing the highest importance scores among the top eight features. This finding strongly aligns with fundamental financial theory and echoes the empirical results of Lin et al. (2011) in the Taiwanese market, reinforcing that profitability remains the primary determinant of corporate financial health. Furthermore, the significance of "Cash Equivalents" validates the foundational arguments of classical models (Ohlson, 1980) regarding the critical role of liquidity buffers in preventing distress.

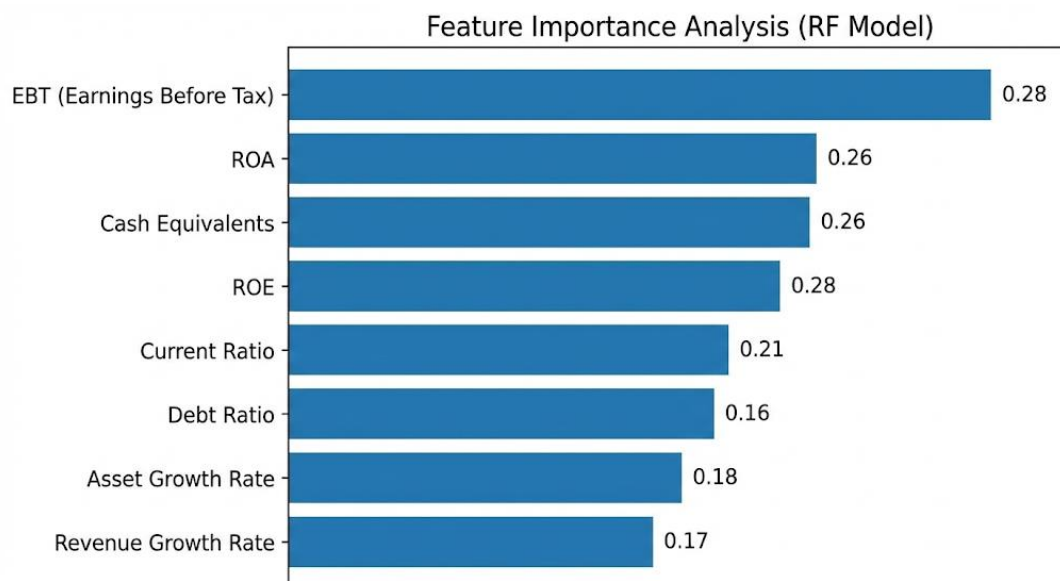


Figure 5 Feature Importance Analysis for the RF Model

5. Conclusion

As the global financial landscape in 2026 faces unprecedented volatility driven by rapid AI transformation, this study provides a timely evaluation of distress prediction models for the Taiwanese market. Through a "baseline-benchmark-frontier" framework, the research validates the complementary strengths of Random Forest (RF) as a robust "gold standard" and Deep Neural Networks (DNN) as high-sensitivity sentinels. By identifying EBT and ROA as critical predictors, these findings offer a sophisticated strategic shield for investors and regulators navigating an era characterized by high-frequency economic asymmetry.

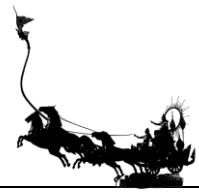
Looking toward "Financial AI 2.0," the research identifies a vital paradigm shift: the transition from complex algorithmic discovery to human-centric operational accessibility. Future research will prioritize the development of intuitive User Interface (UI) systems and simplified dashboards to transform sophisticated "black-box" outputs into actionable, user-friendly insights for decision-makers. By integrating explainable AI (XAI) within these streamlined platforms, technology can better serve as a transparent, manageable guardian of market stability, ensuring that advanced predictive power becomes a bridge to efficient financial governance.

6. Acknowledgements

The author would like to express sincere gratitude to the Department of Banking and Finance at Tamkang University for providing a rigorous academic environment and the necessary resources to complete this research. Special thanks are extended to the Taiwan Economic Journal (TEJ) for the provision of the comprehensive financial database used in the empirical analysis. Most importantly, I would like to express my deepest appreciation to my advisor for their dedicated teaching, unwavering support, and professional mentorship throughout this journey. I am also profoundly grateful for the help and guidance received from all faculty members, fellow researchers, and individuals who contributed to the refinement of this manuscript. Their collective wisdom and encouragement have been instrumental to the success of this research.

7. References

- Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, 94, 164-184.
- Altman, E. I. (1968). Financial ratios, discrimination analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.



- Barboza, F., Kimura, H., & Altman, E. I. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of 2016 the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, Nevada, U.S.A. Retrieved from <https://www.computer.org/csdl/proceedings/cvpr/2016/12OmNqH9hnp>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. (2nd ed). U.S.A.: John Wiley & Sons.
- Kim, M. J., & Kang, D. K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications*, 37(4), 3373-3379.
- Lin, F., Liang, D., & Chen, E. (2011). Financial ratio selection for business crisis prediction. *Expert Systems with Applications*, 38(12), 15094-15102.
- Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743-758.
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131.
- Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, 38(7), 926-947.
- Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, 38, 788–804.
- Tsai, C. F., & Wu, J. W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639-2649.