



# A Multi-Stage Financial Distress Early Warning System: Analyzing Corporate Insolvency with Random Forest

Tanaka, Katsuyuki  
Higashide, Takuo  
Kinkyō, Takuji  
Hamori, Shigeyuki

---

## (Citation)

Journal of Risk and Financial Management, 18(4):195

## (Issue Date)

2025-04-04

## (Resource Type)

journal article

## (Version)

Version of Record

## (Rights)

© 2025 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license



## (URL)

<https://hdl.handle.net/20.500.14094/0100499689>



Article

# A Multi-Stage Financial Distress Early Warning System: Analyzing Corporate Insolvency with Random Forest

Katsuyuki Tanaka <sup>1,2</sup>, Takuo Higashide <sup>3</sup>, Takuji Kinkyō <sup>4</sup>  and Shigeyuki Hamori <sup>4,5,\*</sup> 

<sup>1</sup> Center for Computational Social Science, Kobe University, Kobe 657-8501, Japan; katsutanaka@puppy.kobe-u.ac.jp

<sup>2</sup> Data Science and AI Innovation Research Promotion Center, Shiga University, Hikone 522-8522, Japan

<sup>3</sup> au Asset Management Corporation, Tokyo 101-0065, Japan; takuo0823@gmail.com

<sup>4</sup> Graduate School of Economics, Kobe University, Kobe 657-8501, Japan; kinkyō@econ.kobe-u.ac.jp

<sup>5</sup> Faculty of Political Science and Economics, Yamato University, Suita 564-0082, Japan

\* Correspondence: hamori@econ.kobe-u.ac.jp

**Abstract:** As corporate sector stability is crucial for economic resilience and growth, machine learning has become a widely used tool for constructing early warning systems (EWS) to detect financial vulnerabilities more accurately. While most existing EWS research focuses on bankruptcy prediction models, bankruptcy signals often emerge too late and provide limited early-stage insights. This study employs a random forest approach to systematically examine whether a company's insolvency status can serve as an effective multi-stage financial distress EWS. Additionally, we analyze how the financial characteristics of insolvent companies differ from those of active and bankrupt firms. Our empirical findings indicate that highly accurate insolvency models can be developed to detect status transitions from active to insolvent and from insolvent to bankrupt. Furthermore, our analysis reveals that the financial determinants of these transitions differ significantly. The shift from active to insolvent is primarily driven by structural and operational ratios, whereas the transition from insolvent to bankrupt is largely influenced by further financial distress in operational and profitability ratios.



Academic Editors: Mihaela Brîndușa Tudose, Valentina Diana Rusu and Thanasis Stengos

Received: 17 January 2025

Revised: 23 March 2025

Accepted: 27 March 2025

Published: 4 April 2025

**Citation:** Tanaka, K., Higashide, T., Kinkyō, T., & Hamori, S. (2025). A Multi-Stage Financial Distress Early Warning System: Analyzing Corporate Insolvency with Random Forest. *Journal of Risk and Financial Management*, 18(4), 195. <https://doi.org/10.3390/jrfm18040195>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** random forest; data science; company insolvency and bankruptcy; financial distress; financial vulnerability; economic activity

## 1. Introduction

Financial stability, particularly regarding bankruptcy, is a crucial factor influencing the stability of economic systems, including the global economy. A company's economic activity forms the foundation of industrial and economic development, and its failure due to financial instability affects subsidiaries, employees, clients, and lenders, potentially causing significant economic disruptions (Chakraborty & Sharma, 2007; Jones et al., 2017). Consequently, extensive research has been conducted to develop reliable tools and models for identifying early signs of bankruptcy (Jayasekera, 2018; Tanaka et al., 2019; Altman, 1968; Beaver, 1966; Brédart, 2014). Assessing the financial stability of companies is a key component of measuring economic health and provides valuable insights for various stakeholders, including governments, fund managers, financial institutions, and regulators, enabling them to anticipate economic prospects and formulate measures to mitigate substantial and sometimes irreversible economic losses.

A common approach to assessing a company's financial vulnerability is to identify signs of bankruptcy—known as the bankruptcy modeling approach. Over the past few

decades, supervised machine learning-based (Alpaydin, 2014; Geron, 2022) early warning systems (EWS) have gained popularity as a method for detecting financial vulnerabilities (Chakraborty & Sharma, 2007; Jones et al., 2017; Jayasekera, 2018; Tanaka et al., 2016, 2018, 2019; Holopainen & Sarlin, 2017; Kristóf & Virág, 2022; Barboza et al., 2017; Liu et al., 2023; Climent et al., 2019). They serve as an alternative to the multivariate logistic and probit models, which have been widely used in economic studies (Altman, 1968; Beaver, 1966; Brédart, 2014; Drehmann & Juselius, 2014; Hillegeist et al., 2004; Lennox, 1999; Ohlson, 1980; Martin, 1977). These machine learning models, trained on large datasets of labeled bankruptcy and non-bankruptcy cases, have been shown to outperform conventional approaches in constructing more accurate EWS models.

Another strand of studies captures financial vulnerabilities as multi-stage financial distress based on financial adversity, as bankruptcy is neither a sudden nor a uniform event; rather, it results from progressive stages of financial distress (Tsai, 2013; Farooq & Qamar, 2019; Turetsky & McEwen, 2001; Pindado et al., 2008; Lin et al., 2012; Farooq et al., 2018; Sun et al., 2021). Statistical approaches have been employed to construct a multi-stage financial distress prediction model to capture the early signs of financial adversity.

The aforementioned studies have demonstrated the effectiveness and importance of EWS; the problems of current vulnerability approaches are twofold. First, while bankruptcy prediction provides some degree of early warning, it typically indicates that a company's financial situation has already deteriorated beyond recovery. Some practitioners argue that this approach is impractical, as it fails to offer precise insights into a company's financial health or effective measures to prevent economic damage, often providing warnings too late for recovery. Second, even though the multi-stage financial distress model provides the ability to detect early signs of financial adversity, most analyses have been conducted using statistical approaches with small national datasets (Tsai, 2013; Farooq & Qamar, 2019; Pindado et al., 2008; Lin et al., 2012; Farooq et al., 2018), and research on machine learning-based approaches in the multi-stage financial distress literature remains limited (Sun et al., 2021). Furthermore, the definition of financial distress and the selection of features vary and are often subjective. This is largely due to the nature of statistical modeling approaches, which require defining relationships between independent and dependent variables. Additionally, while statistical modeling approaches offer strong explanatory power compared to machine learning methods—often criticized as a “black box”—they struggle to capture the complex mechanisms underlying corporate financial distress. The motivation of this paper is to answer the research question: whether more effective and systematic methods or frameworks can be developed to detect early signs of financial adversity.

The main purpose of this study is to extend the multi-stage financial distress literature by focusing on insolvency and conducting a financial analysis of insolvency as an alternative and key financial distress status that lies between active and bankrupt states. We conduct this analysis using a random forest machine learning framework and a larger financial dataset from companies operating in Organization for Economic Co-operation and Development (OECD) countries. We construct an early warning system (EWS) with insolvency-based models and systematically investigate (1) the financial indicators that signal a company's transition into insolvency and (2) whether insolvent companies subsequently fall into bankruptcy. Additionally, we analyze (3) how the financial conditions of insolvent companies differ from those of active and bankrupt companies.

Insolvency is often treated as synonymous with bankruptcy in the early warning system (EWS) literature; however, it can also be considered part of a company's financial life cycle. Insolvency represents another unstable financial state, in which a company fails to meet its financial obligations when debt becomes due. Although further financial

deterioration may lead an insolvent company to file for bankruptcy, a company could also potentially recover and return to normal operations through appropriate measures, such as debt restructuring or financial and organizational restructuring. If a company is unable to pay off its debt, it enters default and faces two possible outcomes: (1) liquidation and bankruptcy or (2) receivership and an attempt to recover (known as Chapters 7 and 11, respectively, in the US). Hence, in this study, we treat insolvency as a distinct financial condition separate from bankruptcy and investigate whether it could serve as an EWS indicator, as well as how insolvency and bankruptcy differ.

We follow a similar approach to that of [Tanaka et al. \(2019\)](#) and build random forest models; however, instead of a traditional bankruptcy model, we develop and analyze insolvency models, including active versus insolvent and bankrupt versus insolvent classifications. We choose random forest modeling because it is conceptually simple and does not require complex hyperparameter tuning, which is necessary for neural network-based machine learning methods, particularly deep learning. Additionally, random forest has not been extensively explored in multi-stage financial distress research. Furthermore, unlike other machine learning methods, random forest provides insights into key variables, allowing for a better understanding and interpretation of model construction.

Although some research has investigated the effectiveness of various financial indicators, such as accounting metrics, market prices, macroeconomic variables, and corporate governance factors ([Farooq & Qamar, 2019](#); [Jones, 2017](#); [Manzaneque & Priego, 2016](#); [Miglani et al., 2015](#); [C. C. Chen et al., 2020](#); [Tian et al., 2015](#); [Tian & Yu, 2017](#)), collecting and integrating such information from multiple sources can be highly labor-intensive. Therefore, for simplicity, we use only financial ratios obtained from Orbis to construct the prediction models and conduct our financial analysis. This approach ensures that our primary focus remains on the application of random forest models to insolvency prediction, rather than on variable selection for optimizing predictive performance.

Furthermore, the primary objective of this empirical study is to investigate the feasibility of utilizing insolvency as an alternative to the financially fragile status of bankruptcy and to examine how the financial conditions of insolvent companies differ from those of active and bankrupt firms. This study does not aim to compare predictive performance (“horse race”) with other methodologies ([Almaskati et al., 2021](#)). Thus, our contributions to the early warning system (EWS) literature are as follows:

1. By focusing on insolvency status, we construct an EWS using an insolvency-based modeling approach and investigate whether insolvency status can serve as an alternative financial distress status between active and bankruptcy statuses. We examine whether precautionary financial signs of insolvency can be accurately identified using a random forest machine learning approach ([Breiman, 2001](#)).
2. We analyze how financial criteria and mechanisms differ among active, insolvent, and bankrupt companies by identifying and comparing key variables derived from random forest-based insolvency prediction models.
3. Overall, we introduce a systematic random forest machine learning framework to conduct multi-stage EWS construction and analysis based upon active, insolvency, and bankruptcy company statuses.

Our experimental results demonstrate that the three financial statuses—active, insolvent, and bankrupt—can be distinguished with high accuracy using a machine learning modeling approach. Key variables identified in each model help clarify the mechanisms through which active companies become insolvent and highlight how the financial conditions of insolvent firms differ from those of both active and bankrupt companies. These findings have important implications, suggesting that, although insolvency is often treated

as equivalent to bankruptcy (Purnanandam, 2008), financial data indicate that insolvency and bankruptcy represent distinct financial conditions.

The remainder of this paper is organized as follows. Section 2 reviews the existing literature on early warning systems (EWS). Section 3 describes the dataset and model construction process. Section 4 evaluates the predictive performance of the models and analyzes the key differences between the bankruptcy and insolvency EWS criteria. Finally, Section 5 presents the conclusions.

## 2. Literature Review

The literature on early warning systems (EWS) for financial distress can be broadly categorized into two strands: the bankruptcy prediction approach and the multi-stage financial distress modeling approach. The evolution of these models reflects a growing recognition that financial distress is not a binary event but a progressive process, requiring a more nuanced analytical framework. While early studies primarily focused on bankruptcy prediction, recent research has expanded to examine multiple stages of financial distress. However, despite this shift, the concept of insolvency remains underexplored in EWS literature. This section critically reviews existing studies, highlighting key contributions and identifying research gaps.

### 2.1. Bankruptcy Prediction Approach

The development of financial vulnerability models has a long history, dating back to the seminal works of Altman (1968) and Beaver (1966). Traditional bankruptcy prediction models predominantly employed statistical techniques such as discriminant analysis, multivariate linear logistic regression, and probit modeling (Altman, 1968; Beaver, 1966; Brédart, 2014; Drehmann & Juselius, 2014; Hillegeist et al., 2004; Lennox, 1999; Ohlson, 1980; Martin, 1977). Altman's Z-score model, for instance, became widely used for assessing the likelihood of corporate bankruptcy based on financial ratios such as working capital, retained earnings, and EBIT (Altman, 1968, 1984, 1993; Altman et al., 1994). While these models provided valuable predictive insights, their reliance on linear assumptions limited their flexibility and accuracy (Varian, 2014; Einav & Levin, 2014).

With advances in computing, machine learning techniques have been increasingly applied to bankruptcy prediction. Decision trees, random forests, and deep learning models have demonstrated improved classification performance. A decision tree (Breiman et al., 1984) is a widely used machine learning approach for predicting and analyzing companies' financial vulnerabilities and bankruptcy risk (Kim et al., 2008; Messier & Hansen, 1988; Shirata, 2003). For example, Shirata (2003) used decision trees to develop a linear bankruptcy model based on financial ratios, while Kim et al. (2008) compared decision trees and neural networks with conventional statistical models. Tanaka et al. (2019) extended the decision tree modeling approach by constructing random forest models, an ensemble of trees, to predict financial vulnerability. More recent studies have employed deep learning techniques, such as convolutional neural networks (CNN) (Hosaka, 2019; Jabeur & Serret, 2023) and recurrent neural networks (LSTM and GRU) (Vochozka et al., 2020; Thor & Postek, 2024), to enhance predictive accuracy. Despite their success, these models face limitations, including complexity, lack of interpretability, and high computational demands.

A critical limitation of most bankruptcy prediction models is their focus on terminal failure rather than the preceding stages of financial distress. Bankruptcy represents the final stage of financial decline, often leaving decision-makers with little room for intervention. This has led researchers to explore multi-stage financial distress models, recognizing that early intervention requires understanding financial fragility before bankruptcy occurs.

## 2.2. Multi-Stage Financial Distress Modeling Approach

To address the limitations of bankruptcy prediction models, researchers have examined financial distress as a multi-stage process (Tsai, 2013; Farooq & Qamar, 2019; Turetsky & McEwen, 2001; Pindado et al., 2008; Lin et al., 2012; Farooq et al., 2018). Turetsky and McEwen (2001) used survival analysis to track corporate health deterioration, while Lin et al. (2012) introduced intermediate distress categories based on financial indicators such as insolvency ratio and interest coverage. Tsai (2013) extended this approach by modeling different levels of distress, distinguishing between slight distress, severe distress, and bankruptcy using multinomial logistic regression. Farooq et al. (2018) and Farooq and Qamar (2019) further refined multi-stage models by integrating feature selection and stock market indicators to enhance predictive performance.

Despite these advancements, insolvency has not been adequately addressed as a distinct financial distress condition in the literature. Existing studies primarily focus on liquidity crises and deteriorating financial performance but fail to systematically incorporate insolvency into EWS models. While Lin et al. (2012) mentioned an insolvency ratio, its role in predicting corporate distress remains underexplored. Understanding insolvency as a transitional phase between financial stability and bankruptcy is crucial for improving risk assessment and timely intervention strategies.

## 2.3. Identifying Research Gaps and Contributions

Although multi-stage financial distress models provide a more detailed analysis than traditional bankruptcy models, significant gaps remain. First, most studies rely on traditional statistical techniques, which, while interpretable, may lack predictive accuracy. Second, machine learning applications in multi-stage financial distress modeling remain limited, with most research focusing on bankruptcy rather than earlier distress stages. Third, insolvency has not been adequately integrated into financial distress frameworks, despite its critical role in corporate financial health.

To address these gaps, this study proposes an insolvency-based EWS using a random forest machine learning approach. By systematically incorporating insolvency as a distinct distress status, we examine its role as an intermediate financial condition between active and bankrupt states. Unlike previous models that emphasize bankruptcy as the primary outcome, our approach seeks to improve predictive accuracy and interpretability while offering a more actionable framework for financial decision-making. This study contributes to the literature by extending the multi-stage financial distress framework, integrating machine learning, and emphasizing insolvency as a crucial yet overlooked financial condition.

# 3. Random Forest Insolvency Modelling Methodology

## 3.1. Data

This study used the financial statements of companies recorded in the Orbis database.<sup>1</sup> Orbis provides extensive coverage of billions of companies across countries, offering standardized data formats for various types of information such as financial accounts, status (active or inactive), and M and A activities. We collected the annual financial statements of each company based on the Global Ratio category, which was classified into three groups: profitability ratios, operational ratios, and structural ratios (a total of 26 indicators: 13, 7, and 6 variables, respectively).

Following the experimental data setup of Tanaka et al. (2019), we selected companies operating in OECD countries with consolidation codes C1 (consolidated accounts only), C2/U2 (both types of accounts), and U1 (unconsolidated accounts only). We collected data for companies with the following statuses: Active, Active (insolvency proc.)<sup>2</sup>, Dissolved

(bankruptcy), and Bankruptcy from Orbis and excluded companies missing more than 50% of the relevant values. Table 1 presents the total number of active, insolvent, and bankrupt companies.

**Table 1.** Number of active, insolvent, and bankrupt companies.

Status	Active	Insolvent	Bankrupt
Data size	5,398,234	43,274	273,275

### 3.2. Random Forest Modelling

Following the strategy of [Tanaka et al. \(2019\)](#), we built models using random forest modeling. Random forest is a nonlinear ensemble machine learning algorithm, composed of multiple classifiers, and is a variant of decision trees. Rather than using a single tree, random forest is built by combining a large number of trees, thus forming a “forest”. The key aspect of random forest is that each tree is trained on not only a randomly selected subset of the training data but also a randomly selected subset of features, known as feature bagging. This method contrasts with using all available features. Random sampling of both the training data and features accounts for potential variability in the data, thereby adding more diversity to the dataset. This prevents correlations among trees, reducing the risk of overfitting, bias, and overall variance, which ultimately leads to more accurate predictions. Each tree brings a unique and independent perspective on the decision. The random forest algorithm is summarized as follows:

1. Draw a subset of training data using random sampling with replacement (bootstrap).
2. Train a decision tree using this subset of training data. At each node of the tree, select the best split from a randomly chosen subset of variables (rather than using all available variables).
3. Repeat steps 1 and 2 to generate  $d$  decision trees.
4. Make predictions for new data by voting for (or taking the average of) the most frequent class among all the outputs of the  $d$  decision trees.

We chose random forest due to its simplicity, strong predictive accuracy, extensive use in the EWS literature, and its reported ability to outperform conventional EWS methods ([Tanaka et al., 2019, 2016, 2018](#)). Additionally, random forest modeling provides insights into variable importance by quantifying the relative contribution of each variable to the prediction, which aids in identifying and evaluating the criteria that distinguish between active, insolvent, and bankrupt companies. To demonstrate and confirm the robustness of the insolvency modeling approach, we also constructed models that are widely used in the EWS literature: one statistical method and two machine learning methods, namely logistic regression, neural networks, and decision trees.<sup>3</sup>

To avoid model bias caused by an imbalanced training set and to mitigate computational costs, we randomly down-sampled the number of active and bankrupt companies to match the size of the insolvent companies, ensuring that each class comprised 43,274 data points. We used the Global Ratio data from the most recent available year (year-1) to build the model and evaluate whether the status of the last available year can be accurately predicted. We constructed three binary classification EWS models: active versus insolvent, bankrupt versus insolvent, and active versus bankrupt. All models were trained using the default hyperparameters for simplicity, as the main goal of this paper is not to compare prediction performance. Additionally, we built a multi-class classification model: active versus bankrupt versus insolvent. We used the active versus bankrupt model as the baseline and evaluated the accuracy of different solvency-based models. The accuracy of our experiments was measured as the average of ten-fold cross-validation. Table 2 presents the

financial ratios used as indicators and their statistics for the training data. statistics for the training data.

**Table 2.** Statistics on global ratio indicators.

		1st Qu.	Median	Mean	3rd Qu.
Profitability Ratios	ROE using P/L before tax	−0.97	8.35	2.63	26.87
	ROCE using P/L before tax	2.88	6.61	4.80	11.60
	ROA using P/L before tax	−5.67	0.87	−0.45	6.57
	ROE using Net income	−0.12	5.78	−0.61	19.53
	ROCE using Net income	2.94	6.02	3.74	10.04
	ROA using Net income	−4.68	0.53	−0.92	5.32
	Profit margin	−4.23	0.83	−0.84	4.58
	EBITDA margin	0.49	4.06	4.72	8.75
	EBIT margin	−2.46	1.99	1.29	6.22
	Cash flow/Operating revenue	−0.05	2.51	2.34	6.37
Operational Ratios	Net assets turnover	1.51	3.36	9.97	7.25
	Interest cover	0.17	1.18	9.75	2.17
	Stock turnover	5.84	9.72	30.95	17.06
	Collection period days	7.00	43.00	77.77	96.00
	Credit period days	9.00	32.00	63.53	73.00
Structure Ratios	Current ratio	0.79	1.13	2.35	1.76
	Liquidity ratio	0.44	0.84	1.81	1.35
	Shareholders liquidity ratio	0.19	0.59	7.09	1.42
	Solvency ratio (Asset based) (%)	3.35	16.63	20.37	38.66
	Solvency ratio (Liability based) (%)	14.21	19.86	24.81	27.03
	Gearing (%)	14.72	52.32	116.92	117.99

## 4. Experimental Results

### 4.1. Model Performances

#### 4.1.1. Active vs. Insolvent Model Results

The first row in Table 3 shows the accuracy results for the active versus insolvent model. We found that it is possible to build a comparatively accurate active versus insolvent model, as we achieved a model accuracy of 76.84% on random forest. This is as high as that of the conventional bankruptcy model, for which the performance of the active versus bankrupt model was 75.64%. The results of the neural network, tree, and logistic modelling approaches yielded classification accuracies of approximately 70%, 70.75%, 71.15%, and 69.41%, respectively. Although not as high as those of random forest, they are much higher than those of random prediction. These results indicate that there are significant financial differences between active and insolvent companies that machine learning models can systematically detect. Consequently, insolvent companies can be considered financially distressed and less healthy than active ones.

**Table 3.** Accuracies of the four models.

	Random Forest	Neural Network	Tree	Logistic
Active vs. Insolvent	76.84%	70.75%	71.15%	69.41%
Bankrupt vs. Insolvent	70.31%	60.97%	62.99%	56.70%
Active vs. Bankrupt	75.64%	69.53%	71.27%	68.42%
Active vs. Insolvent vs. Bankrupt	60.81%	50.55%	52.45%	48.98%

Note: For building the multiclass model, multinomial logistic regression is used.

#### 4.1.2. Bankrupt vs. Insolvent Model Results

As shown in the second row of Table 3, although the accuracy is not as high as that of the active versus bankrupt and active versus insolvent models (75.64% and 76.84%, respectively), the results of the bankrupt versus insolvent model for random forest also had relatively high accuracy, at 70.31%. The performance of the other modelling approaches was approximately 10% lower than that of random forest (neural network = 60.97%, tree = 62.99%, logistic = 56.70%); however, they are all still more accurate than random prediction. These findings are surprising and interesting. Although insolvency is often treated as equivalent to bankruptcy, these results have important implications, in that distinguishable differences can be observed in the financial condition of insolvent companies compared to that of bankrupt companies that machine learning can systematically detect.

#### 4.1.3. Active vs. Insolvent vs. Bankrupt Model Results

The results of the multiclass model of active versus bankrupt versus insolvent in the fourth row of Table 3 show accuracies of 60.81% for random forest, 50.55% for neural network, 52.45% for tree, and 48.98% for logistics. These values are much higher than the random prediction of 33%. The results again indicate the existence of notably different financial mechanisms among active, insolvent, and bankrupt states that machine learning can systematically detect.

#### 4.2. Criteria Differences in Financial Condition

The results presented in Section 4.1. indicate differences in financial condition among active, insolvent, and bankrupt companies. Machine learning has great potential for building more accurate models but has been criticized for its lack of interpretability, owing to its black-box modelling nature, for application in economic analysis. This section investigates the criteria differences distinguishing these statuses by analyzing the importance variables (Breiman, 2001) produced in the process of learning the random forest model.

Variable importance<sup>4</sup> measure the variables that have a greater influence on prediction outcomes. The importance of a variable is measured by comparing the performance of the original prediction error  $e_{org}$  with the prediction error  $e_{perm}$  after shuffling or permutating the values of the variable; hence, it is also known as permutation variable importance. Intuitively, if the model prediction error, quotient  $e_{perm}/e_{org}$  or difference  $e_{perm} - e_{org}$  is increased by the shuffling values, the feature is important because it indicates that the performance relies on the variable. However, if the two errors do not differ significantly, the variable does not have much influence on model performance and is thus unimportant.

Figures 1–3 present comparisons of the importance variables for active versus insolvent, bankrupt versus insolvent, and active versus bankrupt, respectively.

##### 4.2.1. Financial Criteria Differences Between Active and Insolvent Companies

The top five priority criteria for differentiating between active and insolvent companies are solvency ratio (asset-based), credit period days, gearing and cash flow/operating revenue, and interest cover. The results show that they were mainly structural ratios (solvency ratio, gearing, and liquidity ratio) and operational ratios (credit period days and interest cover). Notably, profitability ratios, such as ROA- and ROE-based indicators, are not key factors in distinguishing between active and insolvent statuses, since most are ranked very low. These findings contradict previous literature which has often relied on profitability ratios such as EBITDA or ROA as a main indicator to analyze a company's financial health and ability to meet financial obligations.

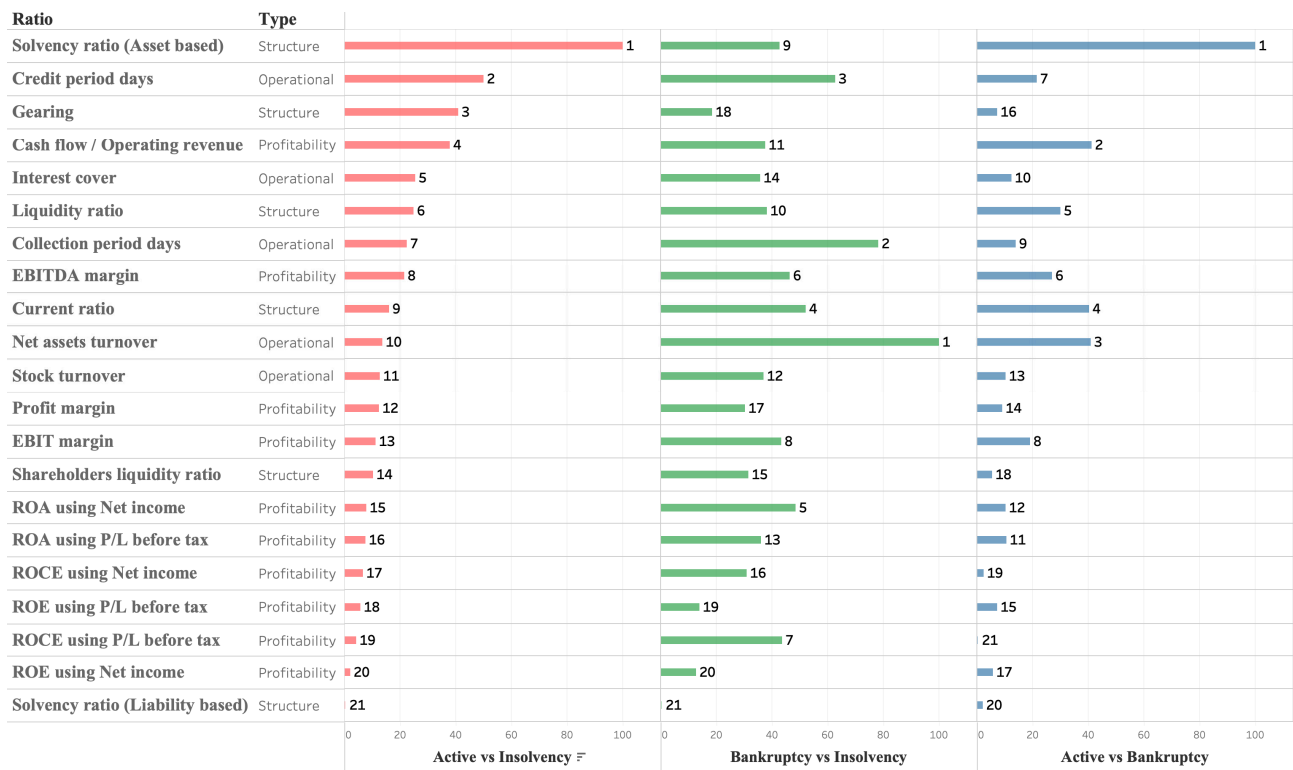


Figure 1. Variable importance in the active versus insolvent model.

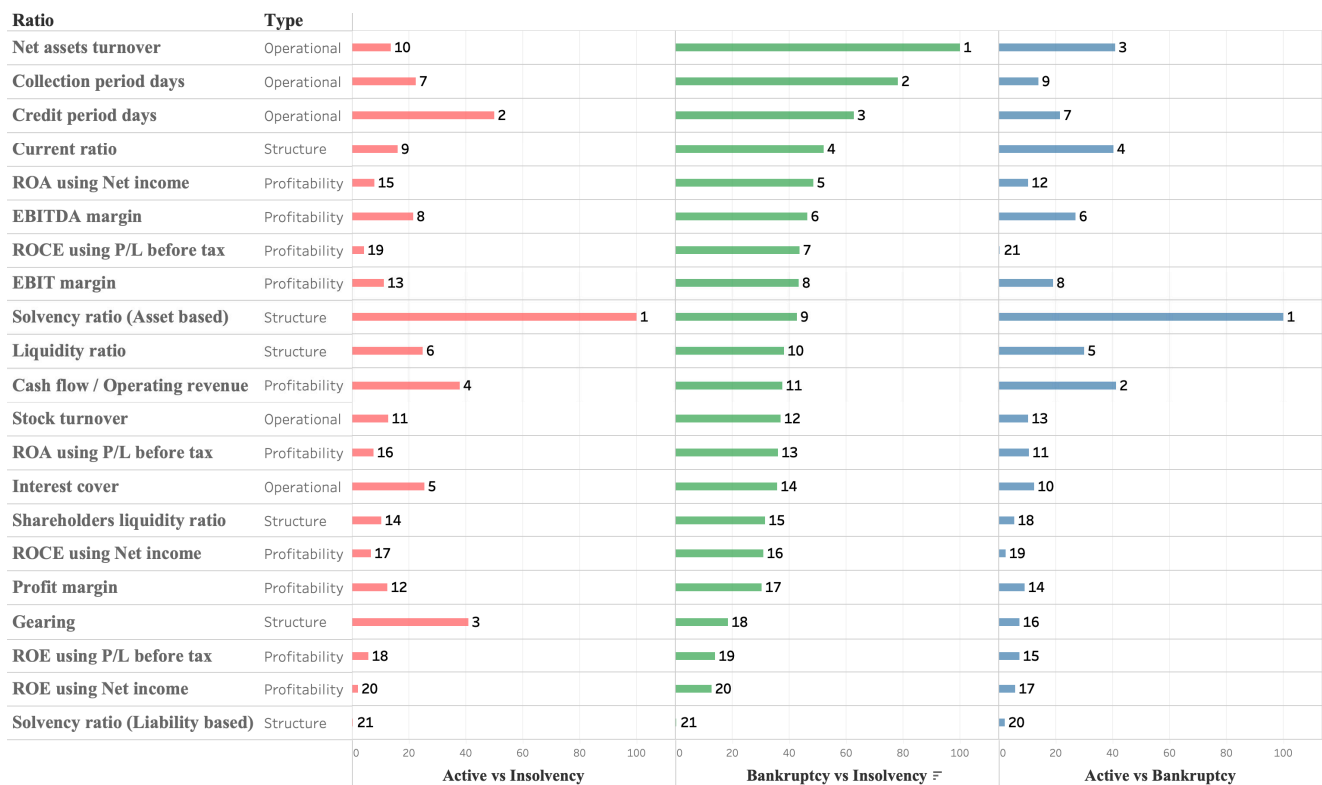


Figure 2. Variable importance in the bankrupt versus insolvent model.



Figure 3. Variable importance in the active versus bankrupt model.

As illustrated in Figure 1, these important variables are positioned very differently in bankrupt versus insolvent because credit period days is only within the top five (rank 3), while the positions of others reflect lower importance, ranked 9, 18, 11, and 14. The top five criteria of active versus insolvent are positioned relatively higher in the active versus bankrupt models, but only the solvency ratio (asset-based) and cash flow/operating revenue are ranked in the top five.

#### 4.2.2. Financial Criteria Differences Between Bankruptcy and Insolvency Companies

In contrast to the active versus insolvent model, as illustrated in Figure 2, the criteria distinguishing between the financial conditions of bankruptcy and insolvency are based mostly on operational ratios. Net assets turnover, collection period days, and credit period days are ranked as the top three (followed by current ratio and ROA using net income). Notably, these are different types of operational ratios found in the active versus insolvent model. Furthermore, the profitability ratios behave very differently in this model, as many of them are ranked in the top ten, whereas they were ranked as less important in the active versus insolvent model.

Figure 2 shows that the important variables found in the bankrupt versus insolvent model are distributed differently than in the active versus insolvent and active versus bankrupt versus insolvent models. This indicates that the mechanism by which an active company falls into insolvency differs from that by which an insolvent company falls into bankruptcy. Unlike the active versus insolvent and bankrupt versus insolvent models, which have few dominant variables, the contribution of the top ten importance variables is relatively high. This indicates that a company that falls into bankruptcy is affected by multiple factors and may have a complex mechanism that it may not be possible to capture using only a few indicators.

#### 4.2.3. Financial Criteria Differences Between Active and Bankruptcy Status

The top five important variables of the active versus bankrupt model, solvency ratio (asset-based), cash flow/operating revenue, net assets turnover, current ratio, and liquidity ratio, as shown in Figure 3, indicate interesting behavior as important variables in other models. Although most of them differ from active to insolvent and insolvent to bankruptcy, some of the top five variables show similar behavior.

The solvency ratio (asset-based) and cash flow/operating revenue are ranked first and fourth in the active versus insolvent model (compared with ninth and eleventh in bankrupt versus insolvent), and the net asset turnover and current ratio are ranked first and fourth in the bankrupt versus insolvent model (tenth and ninth in the active versus insolvent model). Clearly, these are a mixture of active versus insolvent and bankrupt versus insolvent models. Furthermore, similar to the active versus insolvent model, most of the top five variables are structural ratios; however, the bottom half of the top ten variables are a mixture of profitability and operational ratios.

These results have important implications, as they suggest that active companies fall bankrupt in a two-step process of financial distress. Therefore, it would be beneficial to monitor the status of company insolvency as a different sign of financial distress.

#### 4.2.4. Financial Criteria Differences Among Active and Bankrupt and Insolvent Status

These results indicate that the criteria for determining the three statuses are considerably different. Notably, as shown in Figure 4, the multiclass model reflects these findings because the top five variables that differentiate active, insolvent, and bankrupt are solvency ratio (asset-based), liquidity ratio, credit period days, current ratio, and net assets turnover, which are a mixture of the active versus insolvent and bankrupt versus insolvent models and similar to the active versus bankrupt model.

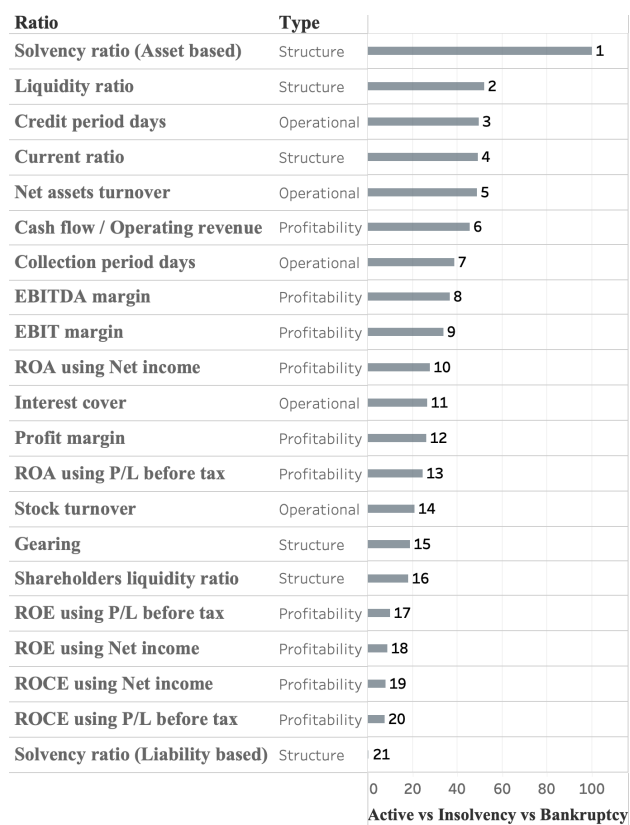


Figure 4. Variable importance in the multiclass model.

### 4.3. Discussion

The classification accuracy results in Section 4.1 demonstrate that the random forest machine learning approach generates more accurate models than conventional statistical methods (Tsai, 2013; Farooq & Qamar, 2019; Turetsky & McEwen, 2001; Lin et al., 2012; Farooq et al., 2018), as in line with previous machine learning-based EWS literature (Tanaka et al., 2019, 2016), even when applied to insolvency-based multi-stage financial distress EWS. This implies that machine learning is capable of identifying distinct patterns in the global ratios data obtained from Orbis, and therefore, the financial conditions of three statuses—active, insolvent, and bankrupt—are comparatively different.<sup>5</sup>

Furthermore, unlike other “black-box” machine learning methods (Sun et al., 2021; Vochozka et al., 2020; Hosaka, 2019; Jabeur & Serret, 2023; Thor & Postek, 2024), which are incapable of explaining the reasoning and causes behind the model’s decisions, the random forest modelling approach provides the overall importance of input variables, regardless of their dimensionality, which allows for a better understanding of the impact of all features. This study experimentally and evidentially found the three models show wide variation in their top five important variables. Active companies become insolvent due to the condition of the structural and operational ratios, whereas insolvent companies go bankrupt due to further condition changes in the different operational and profitability ratios. Furthermore, this is supported by the results of the status change from active to bankrupt, indicating a mixture of the two steps in the financial distress process described above. Feature importance may be more informative than the feature coefficients of linear models (Altman, 1968; Beaver, 1966; Brédart, 2014; Drehmann & Juselius, 2014; Hillegeist et al., 2004; Lennox, 1999; Ohlson, 1980; Martin, 1977; Vochozka et al., 2020; Hosaka, 2019; Jabeur & Serret, 2023; Thor & Postek, 2024), which only indicate the amount of output change relative to changes in specific features, as random forest offers more options for decision-making and policy development.

This framework is not only easy to implement but also flexibly captures changes in key factors. While statistical approaches require the identification of theoretical factors, feature engineering, and the definition of relationships between independent and dependent variables, the random forest method determines important features directly from the data to build effective models. This is important since factors that were important in the past may not be relevant in the future, or they may differ depending on the country or industry.

Additionally, random forest is less restrictive in terms of data assumptions, such as multicollinearity, which can cause interpretability problems due to feature dependence in statistical methods (Altman, 1968; Beaver, 1966; Brédart, 2014; Drehmann & Juselius, 2014; Hillegeist et al., 2004; Lennox, 1999; Ohlson, 1980; Martin, 1977; Vochozka et al., 2020; Hosaka, 2019; Jabeur & Serret, 2023; Thor & Postek, 2024). Since each tree in a random forest model typically uses univariate feature splits, the independence of each split helps avoid such issues.

In practice, the ability to detect early signs of financial adversity would be more informative and beneficial for various stakeholders; overly complex EWS frameworks may be neither practical nor desirable for various stakeholders (Tsai, 2013; Farooq & Qamar, 2019; Turetsky & McEwen, 2001; Lin et al., 2012; Farooq et al., 2018). Furthermore, our approach may be beneficial to EWS researchers and practitioners aiming to capture relationships from large, high dimensional datasets, as our analysis is demonstrated on a larger dataset of companies operating in OECD countries.

In summary, these findings indicate that although an insolvent company is often considered bankrupt, an insolvent company sends different signals of financial distress compared to those that indicate bankruptcy. Hence, the financial mechanisms underlying the transition from an active company to insolvency differ considerably from those that lead

an insolvent company to bankruptcy. Therefore, more attention is required on insolvency conditions and monitoring the insolvency status of companies could provide a better EWS and aid in determining the mechanism of financial stability of companies and the economy.

Consequently, the random forest framework for assessing multi-stage financial distress—a combination of random forest models and financial ratios—offers a simple yet systematic framework for insolvency-based EWS modeling and financial criteria analysis. Even though conventional bankruptcy-based EWSs serve as good precautions to protect from bankruptcy but do not provide signals early enough to prevent bankruptcy, we managed to develop a more effective framework to detect early signs of financial vulnerability.

## 5. Conclusions

This study introduces a simple yet systematic random forest-based framework for insolvency financial analysis—an area largely unexplored in multi-stage financial distress early warning system (EWS) research. We investigate whether a company's insolvency status can serve as a key component of a multi-stage financial distress early warning system (EWS), which categorizes financial distress based on financial adversity. Using a random forest modeling framework, we systematically detect (1) when an active company transitions into insolvency and (2) when an insolvent company progresses to bankruptcy. Additionally, we analyze how the financial condition of an insolvent company differs from that of active and bankrupt companies based on the key variables identified in the random forest models.

Our empirical study demonstrates that accurate insolvency-based models can be developed to detect status changes from active to insolvent and from insolvent to bankrupt. In particular, the results reveal notable differences in data patterns between the financial conditions of insolvent and bankrupt companies, which can be effectively identified by a machine learning model.

Furthermore, our criteria analysis indicates that the financial factors driving the transition from active to insolvent differ significantly from those driving the transition from insolvent to bankrupt. The former is primarily influenced by structural and operational ratios, whereas the latter is driven by further deterioration in operational and profitability ratios.

Conclusively, the results provide important academic implications that bankruptcy is not the only indicator of a company's financial fragility, but insolvency can serve as a distinct financial distress condition or an alternative fragile financial status. Therefore, while bankruptcy-based EWSs serve as important tools for signaling financial vulnerability, incorporating insolvency into early warning systems could enhance the detection of financial distress at an earlier stage, underscoring the need to develop active-to-insolvent and insolvent-to-bankrupt prediction models.

In practice, financially healthy companies do not suddenly go bankrupt. Instead, a company's financial condition typically deteriorates through a series of stages, often exhibiting warning signs of distress. Therefore, early detection of these signs—before the company reaches an irrecoverable state—can provide an opportunity to implement appropriate corrective measures. Consequently, identifying financial distress at an early stage is crucial for preventing bankruptcy and mitigating the associated economic damage.

Our finding that the three financial statuses exhibit distinct financial conditions is promising, as it may offer valuable insights into the underlying causes of corporate financial instability. Furthermore, the simple and systematic framework presented in this study can be readily applied to forecasting financial distress, aiding various decision-makers

in practice and significantly reducing losses for stakeholders and mitigating economic damage.

The limitations of this study can be improved in the future studies as follows. First, the models of this study are based solely on financial factors. However, financial distress is a complex mechanism that also involves other factors; it can be extended by considering factors such as corporate governance, technological development, the stock market, and the accumulation of financial adversity. Second, this study only analyzes OECD companies; we investigate how the insolvency modelling approach is effective with non-OECD companies. Third, the characteristics of corporate failure often depend on underlying economic conditions and environments; future studies could be extended by incorporating countries, industries, and company size into insolvency models to conduct finer-grained analysis. The systematic and reproducible insolvency-based random forest framework, using standardized financial data and status labels from Orbis, can easily be adapted to these additional factors and serves as a baseline to develop a more comprehensive EWS for forecasting and/or managing the financial stability of companies and the broader economy.

**Author Contributions:** K.T., T.H., S.H. and T.K. conceived and designed the experiments; K.T. performed the experiments; T.H., S.H., T.K. and K.T. analyzed the data; K.T. contributed reagents/materials/analysis tools; K.T. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by JSPS KAKENHI (Grant Number 23K01335 and 22K01424).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflicts of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## Notes

- <sup>1</sup> See <https://www.bvdinfo.com/en-gb/our-products/data/international/orbis> (accessed on 26 March 2025).
- <sup>2</sup> Insolvency is categorized as one status of active companies and defined in Orbis as follows: ‘Active (insolvency proceedings): Here the company is declared insolvent. The company remains active, though it is in administration or receivership or under a scheme of arrangement (US—Chapter 11). During this period, the company is usually placed under the protection of a law and continues operating and repaying creditors and tries to reorganize and return to normal operation. At the end, the company will either return to normal operation (the default of payment was thus temporary); or will be reorganized (parts of its activity can be restructured or sold); or will be liquidated’.
- <sup>3</sup> We do not include deep learning because it is not very practical. It requires intensive hyper-parameter tuning to obtain good performance and does not provide off-the-shelf importance variable measurement; hence, it is not very interpolative.
- <sup>4</sup> A variable is also called an indicator or, more widely known as a feature in machine learning.
- <sup>5</sup> Random forest can be replaced by boosting approaches, such as gradient boosting (Friedman, 2001) and XGBoost (T. Chen & Guestrin, 2016), since it also provides easy hyper-parameter setting and interpolation of the model. In fact, as we conducted an experiment of XGBoost with the set experimental setup, the result is very similar to random forest (76.83%, 70.32%, and 75.83%). We used random forest because it is computationally lighter and more scalable owing to its better parallelization capability than the boosting approach, which is generally sequential learning.

## References

- Almaskati, N., Bird, R., Yeung, D., & Lu, Y. (2021). A horse race of models and estimation methods for predicting bankruptcy. *Advances in Accounting*, 52, 100513.
- Alpaydin, E. (2014). *Introduction to machine learning*. The MIT Press.

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. [[CrossRef](#)]
- Altman, E. I. (1984). The success of business failure prediction models: An international survey. *Journal of Banking & Finance*, 8(2), 171–198.
- Altman, E. I. (1993). *Corporate financial distress and bankruptcy: A complete guide to predicting and avoiding distress and profiting from bankruptcy* (Vol. 18. 3). Wiley Finance Edition. Wiley.
- Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*, 18(3), 505–529.
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417.
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. [[CrossRef](#)]
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. [[CrossRef](#)]
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. In *Monterey: Wadsworth*. Chapman & Hall.
- Brédart, X. (2014). Bankruptcy prediction model: The case of the United States. *International Journal of Economics and Finance*, 6(3), 1–7. [[CrossRef](#)]
- Chakraborty, S., & Sharma, S. K. (2007). Prediction of corporate financial health by artificial neural network. *International Journal of Electronic Finance*, 1(4), 442–459. [[CrossRef](#)]
- Chen, C. C., Chen, C. D., & Lien, D. (2020). Financial distress prediction model: The effects of corporate governance indicators. *Journal of Forecasting*, 39(8), 1238–1252. [[CrossRef](#)]
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery.
- Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An extreme gradient boosting approach. *Journal of Business Research*, 101, 885–896.
- Drehmann, M., & Juselius, M. (2014). Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting*, 30(3), 759–780.
- Einav, L., & Levin, J. (2014). The data revolution and economic analysis. *Innovation Policy and the Economy*, 14, 1–24. [[CrossRef](#)]
- Farooq, U., Jibrán Qamar, M. A., & Haque, A. (2018). A three-stage dynamic model of financial distress. *Managerial Finance*, 44(9), 1101–1116. [[CrossRef](#)]
- Farooq, U., & Qamar, M. A. J. (2019). Predicting multistage financial distress: Reflections on sampling, feature and model selection criteria. *Journal of Forecasting*, 38(7), 632–648. [[CrossRef](#)]
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.
- Geron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems* (3rd ed.). O'Reilly Media, Inc.
- Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1), 5–34. [[CrossRef](#)]
- Holopainen, M., & Sarlin, P. (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance*, 17(12), 1933–1963.
- Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, 117, 287–299. [[CrossRef](#)]
- Jabeur, S. B., & Serret, V. (2023). Bankruptcy prediction using fuzzy convolutional neural networks. *Research in International Business and Finance*, 64, 101844.
- Jayasekera, R. (2018). Prediction of company failure: Past, present and promising directions for the future. *International Review of Financial Analysis*, 55, 196–208. [[CrossRef](#)]
- Jones, S. (2017). Corporate bankruptcy prediction: A high dimensional analysis. *Review of Accounting Studies*, 22(3), 1366–1422. [[CrossRef](#)]
- Jones, S., Johnstone, D., & Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting*, 44(1–2), 3–34.
- Kim, C. N., Yang, K. H., & Kim, J. (2008). Human decision-making behavior and modeling effects. *Decision Support Systems*, 45(3), 517–527. [[CrossRef](#)]
- Kristóf, T., & Virág, M. (2022). EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks. *Research in International Business and Finance*, 61, 101644. [[CrossRef](#)]
- Lennox, C. (1999). Identifying failing companies: A re-evaluation of the logit, probit and DA approaches. *Journal of Economics and Business*, 51(4), 347–364. [[CrossRef](#)]

- Lin, S. M., Ansell, J., & Andreeva, G. (2012). Predicting default of a small business using different definitions of financial distress. *Journal of the Operational Research Society*, 63(4), 539–548. [\[CrossRef\]](#)
- Liu, J., Li, C., Ouyang, P., Liu, J., & Wu, C. (2023). Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach. *Journal of Forecasting*, 42(5), 1112–1137. [\[CrossRef\]](#)
- Manzaneque, M., & Priego, A. M. (2016). Corporate governance effect on financial distress likelihood: Evidence from Spain. *Revista de Contabilidad*, 19(1), 111–121. [\[CrossRef\]](#)
- Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, 1(3), 249–276.
- Messier, W. F., Jr., & Hansen, J. V. (1988). Inducing rules for expert system development: An example using default and bankruptcy data. *Management Science*, 34(12), 1403–1415.
- Migliani, S., Ahmed, K., & Henry, D. (2015). Voluntary corporate governance structure and financial distress: Evidence from Australia. *Journal of Contemporary Accounting & Economics*, 11(1), 18–30.
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109.
- Pindado, J., Rodrigues, L., & De la Torre, C. (2008). Estimating financial distress likelihood. *Journal of Business Research*, 61(9), 995–1003.
- Purnanandam, A. (2008). Financial distress and corporate risk management: Theory and evidence. *Journal of Financial Economics*, 87(3), 706–739. [\[CrossRef\]](#)
- Shirata, C. (2003, January). *Predictors of bankruptcy after bubble economy in Japan: What can you learn from Japan case?* The Proceedings of the 15th Asian-Pacific Conference on International Accounting Issues, Bangkok, Thailand.
- Sun, J., Fujita, H., Zheng, Y., & Ai, W. (2021). Multiclass financial distress prediction based on support vector machines integrated with the decomposition and fusion methods. *Information Sciences*, 559, 153–170. [\[CrossRef\]](#)
- Tanaka, K., Higashide, T., Kinkyo, T., & Hamori, S. (2019). Analyzing industry-level vulnerability by predicting financial bankruptcy. *Economic Inquiry*, 57(4), 2017–2034.
- Tanaka, K., Kinkyo, T., & Hamori, S. (2016). Random forests-based early warning system for bank failures. *Economics Letters*, 148, 118–121.
- Tanaka, K., Kinkyo, T., & Hamori, S. (2018). Financial hazard map: Financial vulnerability predicted by a random forests classification model. *Sustainability*, 10(5), 1530. [\[CrossRef\]](#)
- Thor, M., & Postek, L. (2024). Gated recurrent unit network: A promising approach to corporate default prediction. *Journal of Forecasting*, 43(5), 1131–1152. [\[CrossRef\]](#)
- Tian, S., & Yu, Y. (2017). Financial ratios and bankruptcy predictions: An international evidence. *International Review of Economics & Finance*, 51, 510–526.
- Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89–100.
- Tsai, B.-H. (2013). An early warning system of financial distress using multinomial logit models and a bootstrapping approach. *Emerging Markets Finance and Trade*, 49(Suppl. S2), 43–69.
- Turetsky, H. F., & McEwen, R. A. (2001). An empirical investigation of firm longevity: A model of the ex ante predictors of financial distress. *Review of Quantitative Finance and Accounting*, 16(4), 323–343.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28.
- Vochozka, M., Vrbka, J., & Suler, P. (2020). Bankruptcy or success? The effective prediction of a company's financial development using LSTM. *Sustainability*, 12(18), 7529. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.