FINANCIAL DISTRESS FORECASTING WITH A MACHINE LEARNING APPROACH

Hong Hanh Ha *, Ngoc Hung Dang **, Manh Dung Tran ***

* National Economics University, Hanoi, Vietnam
** Hanoi University of Industry, Hanoi, Vietnam
*** *Corresponding author,* National Economics University, Hanoi, Vietnam
Contact details: National Economics University, 207 Giai Phong Road, Hai Ba Trung district, Hanoi, Vietnam

OPEN ACCESS

How to cite this paper: Ha, H. H., Dang, N. H., & Tran, M. D. (2023). Financial distress forecasting with a machine learning approach. *Corporate Governance and Organizational Behavior Review, 7*(3), 90–104. https://doi.org/10.22495/cgobrv7i3p8

ISSN Online: 2521-1889
ISSN Print: 2521-1870

Received: 16.03.2023
Accepted: 09.06.2023

JEL Classification: F65, G21, O16
DOI: 10.22495/cgobrv7i3p8

Abstract

A highlighted issue relating to the financial distress of public companies raises more debate from both academic and current practice perspectives as financial markets are currently a key source of growth for the local and international economies. In the context of advanced technology and the digital revolution, forecasting and early detection of financial distress are important methods that contribute to increasing confidence between investors and the market and help to make sound decisions promptly to avoid reaching bankruptcy (Fuentes et al., 2023). This study employs machine learning algorithms to measure the probability of financial distress of listed firms on the Vietnam Stock Exchange by using a dataset with 4,936 observations from 2009 to 2020. The research has identified internal determinants such as debt-to-equity ratio, asset turnover ratio, and profit margin ratio as indicators that have the greatest impact on financial distress under different models. The results reveal that Model 1 — Altman and Model 3 — Zmijewski predict financial distress with an accuracy rate of 98%. In addition, we have determined the threshold when using the decision tree algorithm, which has an important impact on the financial distress of listed firms. This finding contributes to the existing literature review and is consistent with previous studies of Chen et al. (2021) and Martono and Ohwada (2023).

Keywords: Financial Distress, Machine Learning, Random Forest, Artificial Intelligence

Authors' individual contribution: Conceptualization — H.H.H.; Methodology — N.H.D.; Validation — H.H.H.; Writing — Review & Editing — N.H.D. and M.D.T.; Visualization — H.H.H.; Supervision — H.H.H., N.H.D., and M.D.T.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

1. INTRODUCTION

Financial distress forecasting has been a topic of interest in recent decades because of its importance for listed firms, investors, creditors, regulators, and the economy (Wanke et al., 2015). If financial distress prediction is reliable, firm managers can initiate remedial measures to avoid deterioration before the crisis hits, and investors can take advantage of the crisis to evaluate the financial position of listed firms and adjust their investment strategies for optimizing profit.

Financial distress forecasting and bankruptcy is getting more and more attention from investors, creditors, and management. Determination of a firm falling into financial distress is necessary because it helps firm managers give suitable management for maintaining operations. It also helps investors and creditors evaluate the risks that they encounter when a firm falls into financial distress. Almost all studies on financial distress are conducted in the circumstances of the United States of America and Europe. This theme is still new in emerging countries including Vietnam.

Vietnam is in the process of deep integration with the region and the world. This is a golden opportunity for Vietnamese firms to join but also a challenge. Being a developing economy, most Vietnamese firms are small and medium-sized, so investment opportunities are not high, competitiveness is low, and above all, the capital market has not been developed yet. Vietnamese firms are prone to face difficulties such as capital scarcity, unstable cash flow, low investment opportunities, risk of insolvency, and likely to be in financial distress. Therefore, it is worth mentioning that the proper implementation of financial distress forecasting provides a good opportunity for firms to compete in the market and offer qualitative products and services (Hallunovi, 2023).

According to Wruck (1990), financial distress is a term that describes a financial difficulty when a firm's cash flow is insufficient to pay its current financial liabilities. Research directions on financial distress to date have focused on both theoretical and empirical aspects. In terms of theory, the studies provide methods to measure the state of financial distress of firms. These works use analytical approaches to identify variables for measuring the probability of financial distress as well as to determine the cut-off for getting the threshold for financial distress and non-financial distress. Many scientists focus on univariate models to determine the separate effects of each variable to measure the likelihood of firm financial distress (Beaver, 1966). In addition, they have developed methods of multivariate analysis and conditional probability analysis to measure financial distress (Altman, 1968; Ohlson, 1980; Zmijewski, 1984). The extant literature on financial distress is rich and diverse such as the empirical studies of Campbell et al. (2008) and Tinoco and Wilson (2013) aim to analyze the determinants of the probability of financial distress.

Conducting this topic of financial distress forecasting with a machine learning approach is necessary for both theoretical and practical perspectives. It helps management know the impact levels of determinants influencing financial distress and then give some suggestions to overcome. The signals of financial distress are also recognized and quickly responded to for reducing expenses to solve this problem. These aspects of financial distress also affected by corruption risk can become the foundation for effective and proactive community fraud prevention measures (Marzuki et al., 2022; Julian et al., 2022; Malik & Yadav, 2020). Machine learning is a data analysis method providing an accuracy of 98% to forecast financial distress. In this study, machine learning algorithms are employed to predict the probability of financial distress in the case of listed firms on the Vietnam Stock Exchange. From there, we consider which financial indicators are most effective in forecasting and determining models as well as which algorithms are the most effective.

To achieve the above objectives, the rest of the paper is structured as follows. Section 2 presents a theoretical framework and literature review on research issues including variable definitions and prior findings in the extant literature review. Section 3 is the methodology which describes the variable measurement, the method of machine learning, the method of evaluation, and the research data. Section 4 includes the empirical results and some discussion of findings whereas Section 5 is the conclusion of the paper.

2. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1. Theoretical framework

2.1.1. Financial distress

Altman and Hotchkiss (2005) provide a complete description and definition of financial distress and show that bankruptcy is the closest legal definition of a financial crisis. "Bankruptcy" occurs when a firm submits a formal bankruptcy to a court and is approved by the court for bankruptcy. Zmijewski (1984) defines financial distress as the act of submitting for bankruptcy. However, many firms falling into financial distress have never filed for bankruptcy due to mergers or privatizations, whereas those in good standing often file for bankruptcy for avoiding taxes and costly lawsuits (Theodossiou et al., 1996). In practice, "bankruptcy", "financial failure", "default", and "financial distress" are used interchangeably. The terms "financial failure" or "financial distress" is used in many studies. "Financial distress" is a more flexible definition than "bankruptcy" and is used in the study with a bigger sample size. In contrast, "bankruptcy" is a special form of "financial distress"; and "bankruptcy" focuses on studies with smaller sample sizes. The use of "financial distress" provides more apt not only in practice but also in theory, because not all financially distressed firms go "bankrupt". "Bankruptcy" is only a last option for firms when they cannot solve their financial problems (Aktas & Mahaffy, 1996).

Others argue that financial distress refers to a firm's difficulty in repaying debt or meeting other financial obligations (Ghazali et al., 2015). In case of severe financial distress, the firm may go bankrupt. Binti, Zeni, and Ameer (2010) define financial distress as a term used when contractual arrangements with creditors cannot be executed due to a firm's financial difficulties. Meanwhile, Hu and Ansell (2006) state financially distressed firms as those with a debt ratio greater than 1, meaning that liabilities are greater than total assets, or an interest payment ratio (based on cash flow) is less than 1, meaning that the cash flow of a firm is not enough to pay interests.

2.1.2. Techniques for predicting financial distress

Discriminant analysis

Discriminant analysis (DA) is the method commonly employed before 1980 (Altman, 1968; Beaver, 1966).

Beaver's univariate discriminant analysis

Beaver (1966) employs financial indicators from the empirical study of 79 bankrupt firms and a number of firms that did not fail over 10 years (1954–1964).

The results show that the ratio of cash to total liabilities is the most important indicator in predicting signs of financial distress and bankruptcy. This indicator presents the balance between a firm's ability to generate cash and the amount of debt that the firm has to pay. In addition, return on assets and debt ratio are also important signals in detecting firm financial distress and bankruptcy because these signals reflect the firm performance and the level of financial risk.

A comparison of indicators which were drawn from Beaver's research illustrates that all financial indicators of firms in crisis are much lower than those in a normal situation. Thus, the findings of Beaver's (1966) study show the way to predict the financial distress of a firm in aspects of detecting signs of financial distress/bankruptcy of a business by comparing the firm financial ratios with the averages calculated by Beaver. These findings also have been applied in several fields of financial reporting and corporate governance (Biberaj et al., 2022; Nguyen & Ahmed, 2023).

Altman's multivariate discriminant analysis

Unlike Beaver (1966), Altman (1968) uses multivariate discriminant analysis (MDA) to find linear equations of financial ratios for determining whether firms are bankrupt or not. Altman (1968) employs MDA with the variables used as *X1* = Working capital/Total assets; *X2* = Retained earnings/Total assets; *X3* = Earnings before tax and interest/Total assets; *X4* = Market value equity/Book value of total liabilities; *X5* = Sales/Total assets; *Z* = Overal index based on data of 66 firms in the US, in which these firms are divided into two groups of 33 each. Group 1 includes 33 firms that went bankrupt from 1946 to 1965. Group 2 consists of 33 firms that did not go bankrupt and continued to operate normally (at least) until 1966. Non-bankrupt firms had similar sizes and sectors are paired with bankrupt firms. From balance sheets and income statements, 22 financial indicators are calculated and classified into five groups of liquidity, profit, leverage, solvency, and operating ratios.

Logistic analysis techniques

Unlike discriminant analysis, which only determines whether a firm is distressed or not, logit analysis can also determine the probability of a firm's financial distress. The coefficients of the logit model can be estimated using the "maximum likelihood" method. Logit analysis uses the logistic cumulative probability to predict financial distress. The result of a function is between 0 and 1, which is the probability of financial distress.

Using logistic models and data from financial statements of American firms for the period 1970–1976, Ohlson (1980) develops a model that estimates the probability of firm failure. Data were collected from 105 bankrupt firms and 2,058 non-bankrupt firms in the industrial sector, from 1970 to 1976 that have traded on the US Stock Exchange. The indicators calculated and selected for use in the model represent four groups of basic financial indicators in predicting bankruptcy, including size, financial structure, performance, and liquidity. From there, Ohlson (1980) selects nine independent variables in predicting bankruptcy/financial distress, including:

- *SIZE* = log (Total assets/(GNP price – level index);
- *TLTA* = Total liabilities/total assets;
- *WCTA* = Net working capital/total assets;
- *CLCA* = Current liabilities/current assets;
- *OENEG* = 1 if total liabilities > total assets, and vice versa;
- *NITA* = Profit after tax/total assets;
- *FUTL* = Active funds/total liabilities;
- *INTWO* = 1 if net income decreases in two consecutive years and vice versa.

$$CHIN = \frac{NI_t - NI_{t-1}}{|NI_t| + |NI_{t-1}|} \quad (1)$$

where, NI_t = net income.

Based on the theoretical framework and literature review of the possibility of bankruptcy/ financial distress, Ohlson (1980) proposed the variation of independent variables to a dependent variable as *TLTA*, *CLCA*, and *INTWO* have covariate properties; *SIZE*, *WCTA*, *NITA*, *FULT*, and *CHIN* are inverse; *OENEG* is unspecified. The three models include: 1) the first model predicts failure in 1 year; 2) the second model predicts failure in 2 years; and 3) the third model predicts failure in 1 or 2 years. Then Ohlson (1980) used logistic binary to predict the probability of a firm's bankruptcy for each model. The results show that the predictive accuracy is over 90%. The classification of firms is based on the calculated value of p (p is the probability that a firm is at risk of bankruptcy). If $p > 0.5$, the firm is assigned to bankrupt/financially distressed, and if $p < 0.5$, it is unlikely to go bankrupt/financially distressed.

Machine learning algorithms

Kumar and Ravi (2007) adopt various algorithms of intelligent techniques to solve the problems of financial distress. According to Serrano-Cinca (1996) and Fletcher and Goss (1993), neural network (NN) is the most commonly used technique. Other techniques including decision tree (DT), and support vector machines (SVM) are used to investigate financial distress prediction. A decision tree is a structured hierarchical tree employed to classify objects based on a series of rules. When given data about objects containing attributes along with their classes, the decision tree generates rules to predict the class of the unknown objects (unseen data). The support vector machines technique is a supervised machine learning model used to analyze and classify data. SVM takes incoming data and classifies them into two different classes. There are many studies using machine learning in predicting financial distress such as those of Anandarajan et al. (2001), Wang and Ma (2012), Kim and Upneja (2014), Geng et al. (2015), Gregova et al. (2020), and Tunio et al. (2021).

2.2. Literature review

2.2.1. Identification of financial distress

Most of the studies on forecasting financial distress have focused on predicting bankruptcy (Altman, 1968). However, recent studies have shown that financial distress is not the same as bankruptcy and suggest that not all firms undergoing financial distress will eventually submit for bankruptcy

(He et al., 2010). The existence of different views on financial distress in studies on forecasting financial distress is due to the difference in the selection of research samples as well as the variety and complexity of the financial distress (Wruck, 1990), including failure, insolvency, default, and bankruptcy (Altman & Hotchkiss, 2005). Therefore, there are some measures to identify the firm financial distress. Some studies identify the state of financial distress based on accounting and market data (Denis & Denis, 1995; Andrade & Kaplan, 1998). Some other studies rely on corporate actions such as cutting or stopping dividend payments, delisting, submitting for bankruptcy, or performing mergers and acquisitions with other firms (Turetsky & McEwen, 2001; Altman & Hotchkiss, 2005). Recently, many studies have confirmed that the Z-index (Altman, 1968) or the index of B (Zmijewski, 1984) can be employed as a metric to determine whether a firm is in financial distress or not. Among them, the index is the most commonly used because it is not sensitive to different states of financial distress and is also not sensitive to business lines (Munsif et al., 2011; Kim & Upneja, 2014).

2.2.2. Application of machine learning in financial distress prediction

The previous empirical evidence for forecasting financial distress illustrates that models have been improved in both predictability and accuracy over different periods, from univariate analysis (Beaver, 1966), multivariate discriminant analysis, MDA (Altman, 1968), and logistic conditional probability statistical analysis (Ohlson, 1980). Multivariate discriminant and logistic analyses are two popular methods because of their high accuracies. However, both models have weaknesses in assumptions that make use difficult. MDA assumes that independent variables have a normal distribution and a matrix of variance–covariance has to be the same between financial distress and non-financial distress, while logistic analysis assumes data variability homogeneity and sensitivity to multicollinearity.

Since the early 20th century, with the development of science and technology, machine learning models such as artificial neural networks (ANN), support vector machines (SVM), random forest (RF), decision tree (DT), and Bayesian models have been introduced. Studies of forecasting financial distress adopting machine learning methods have become non-parametric methods employed in forecasting financial distress. Updated studies confirm that ensemble algorithms from feature selection to predictor construction can achieve high accuracy according to the actual case, and the interpretation framework can meet the needs of external users by generating local explanations and global explanations (Zhang et al., 2022). An overview of financial distress using machine learning is presented in Table 1.

Table 1. Overview of some financial distress studies by machine learning method

Authors	*Dataset*	*Algorithms**	*Evaluation metrics*	*Findings*
Anandarajan et al. (2001)	American	ANN	Accuracy	73%
Wang and Ma (2012)	Chinese	LRA, DT, ANN, SVM,	Accuracy	67.52%–78.98%
Kim and Upneja (2014)	American	DT, AdaBoosted DT	Accuracy AUC	93.08%–98.1% 68.4%-98.8%
Heo and Yang (2014)	Korean	AdaBoost, ANN, SVM, DT	Accuracy	73.1%–78.5%
Geng et al. (2015)	Chinese	ANN, SVM, DT, LR, DA, RSA, CM, MC	Accuracy Precision Recall	73.96%–77.79% 70.39%–81.82% 67.45%–74.64%
Gregova et al. (2020)	Slovak	LR, NN, RF	AUC	87.1%–87.7%
Tunio et al. (2021)	Pakistani	LR, ANN, DT, SVM	Accuracy AUC	82.94%–89.07% 83.8%–94.3%

*Note: * Artificial neural networks (ANN), Lending risk analysis (LRA), Support vector machines (SVM), Decision tree (DT), Logistic regression (LR), Discriminant analysis (DA), Rough set analysis (RSA), Clustering methods (CM), Multiple classifiers (MC), Area under curve (AUC).*

Several highlighted studies such as Chen et al. (2021), Kuizinienė et al. (2022), and Martono and Ohwada (2023) illustrate that for the Z-score model, samples analyzed using the five classifiers in five groups (1:1 — 5:1) of different ratios of companies, the bagging classifiers scores a worse (40.82%) than when no feature selection model is used, while the logistic regression classifier and decision tree classifier (J48) result in better scores.

3. RESEARCH METHODOLOGY

3.1. Variable measurement

Currently, there are some ways to measure financial distress. However, each measure has its advantages and disadvantages. Ghazali et al. (2015) state that Altman Z-Score can be the most popular method to measure the financial condition and has been employed to determine financial distress. Therefore, in this study we determine financial distress based on three approaches 1) the Z-index of Altman (1968), 2) the dummy variable of Fich and Slezak (2008), and 3) the index B of Zmijewski (1984).

Z-index of Altman (1968)

Altman's Z-index gives a calculation of the Z-index based on the following formula:

$$Z = 0.717 \times X1 + 0.847 \times X2 + 0.107 \times X3 + 0.420 \times X4 + 0.998 \times X5 \quad (2)$$

where:

- *X1* = Current assets minus current liabilities divided by total assets;
- *X2* = Retained earnings divided by total assets;
- *X3* = Profit before tax and interests divided by total assets;
- *X4* = Book value of equity divided by total debt;
- *X5* = Revenue divided by total assets.

If $Z < 1.81$, a firm is in the financial distress zone and the financial distress variable will have a value of 1, otherwise, it will have a value of 0.

Dummy variable of Fich and Slezak (2008)

Fich and Slezak (2008) measure financial distress through a dummy variable with a value of 1 if the ratio of return on interest expenses is less than 1 (that is, equal to evidence of financial distress), and has a value of 0 otherwise. This measure assumes that if a firm is unable to generate profit large enough to cover its interest expenses, it will soon face default on its debts. It can be said that this measure is based on book value, so it can overcome concerns about ambiguity in the index measurement method.

Index B of Zmijewski (1984)

Index B of Zmijewski (1984) defines index B as follows:

$$P(B = 1) = P(B * > 0); B * = -4.3 - 4.5ROA + 5.7FINL + 0.004LIQ \quad (3)$$

where:

- ROA = Net profit divided by total assets;
- $FINL$ = Total debt to total assets;
- LIQ = Current assets divided by current liabilities.

Firms are determined to be financially distressed when B > 0 (the cut-off point is 0) or the financially distressed variable will have a value of 1, otherwise, there is no financial distress.

In this study, 22 attributes of financial ratios are calculated, including the group of solvency, group of capital structure and debt serviceability, group of profitability, group of activity, group of growth indicators, and others, which are presented in detail in Appendix A, Table A.1.

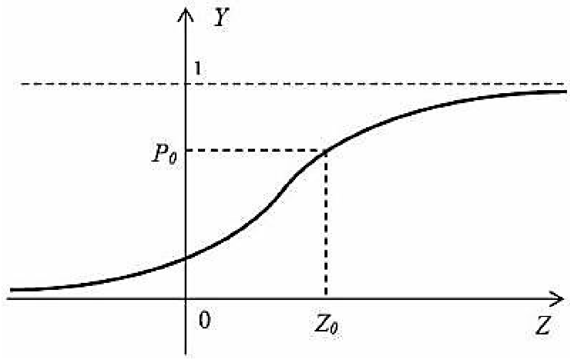
3.2. Machine learning methods

Machine learning is a means of artificial intelligence that employs algorithms to allow computers for learning from data for solving specific problems such as making computers have basic human cognitive (hearing, seeing, understanding, and solving math problems). Machine learning plays an important role in sciences and its applications are part of daily life. Machine learning is used to filter email spam, and predict the weather, in medical diagnostics, product recommendations, facial recognition, credit card fraud detection, financial distress prediction, or firm bankruptcy. In this research, we adopt some commonly used algorithms to predict financial distress such as logistic regression (LR), decision tree (DT), Bayesian network, support vector machine (SVM), K-nearest neighbor (KNN), and random forest (RF).

3.2.1. Logistic regression

Logit regression introduced by Berkson (1944) is a common tool in data analysis with binary variables. Some developments of Altman et al. (1994) and Flitman (1997) have been used in the analysis of multivariate regression models, and the analysis of differences. The binary logistic model employs a binary dependent variable to estimate the probability that an event will occur given the information of the independent variable. Data to be collected about the dependent variable is whether a certain event occurs or not (the dependent variable Y now has two values 0 and 1, with 0 being no event and 1 occurring) and of course data about the independent variables $X1$, $X2$, ..., Xk. From this binary dependent variable, a procedure will be employed to predict the probability of the event occurring according to the rule if the predicted probability is greater than 0.5 (default cut-off point), then the prediction result will be "yes", otherwise the predicted result will be given as "no". The binary logistic model is presented in Figure 1:

Figure 1. The binary logistic regression model



where, P is the probability that $Y = 1$ (which is the probability that the event will occur) when the independent variables take on a specific value. Accordingly, the probability that the event does not occur is:

$$1 - P = Prob(Y = 0) = 1 - \frac{e^z}{1 + e^z} = \frac{1}{1 + e^z} \quad (4)$$

The regression coefficients were estimated by the method of maximum likelihood (maximum likelihood-ML). The logit regression model can be adopted to estimate the log(odds) ratio for each independent variable of the model of Ohlson (1980). The parameters β_n were estimated by the method of maximum likelihood. The logit model is used with many types of data, few constraints, effective when applied in practice, easy to interpret results, and capable of monitoring, diagnosing, and adjusting to match results. consistent with reality.

3.2.2. Decision tree

Decision tree, a classification model introduced by Belson (1959), is widely adopted in different fields. After the introduction of the machine learning method system, the decision tree was further developed with the C4.5 algorithm by Quinlan (1996) and the ID3 algorithm by Quinlan (1986). A decision tree is a structured classification tree that classifies objects based on sequences of rules. Independent variables and attributes can be of different data types such as binary, nominal, ordinal, and quantitative data. To determine which variable to use classification first, the information weight (entropy) for each variable is calculated, the higher the information value, the more categorical information the variable carries.

3.2.3. Bayesian network

Bayesian network is applied for classification based on conditional probability. Like the logistic function, the Bayesian result is a probability that has a value from 0 to 1 (expressing the probability of an event occurring from 0% to 100%), the variables are linked together by a probability. The Bayesian method is developed from the Bayes theorem in statistical probability. According to Carlin and Louis (2000), the Bayesian method is more about statistics than regression. The Bayesian method is quite efficient and easy to use, does not require data conditions, and can work on both numeric and alphanumeric data. With small or unbalanced datasets, the method is more effective, when other methods cannot perform or have to process data with a lot of operations. For fraud detection, the Bayesian network will be built with the Bayesian rule along with the condition $P(Y = 1) + P(Y = 0) = 1$ written as below:

$$P(Y = 1 \mid X) = \frac{P(X \mid Y = 1)P(Y = 1)}{P(X)} \tag{5}$$

$$P(Y = 0 \mid X) = \frac{P(X \mid Y = 0)P(Y = 0)}{P(X)} \tag{6}$$

$$P(Y = 0 \mid X) = \frac{P(X \mid Y = 0)P(Y = 0)}{P(X)} \tag{7}$$

where, $(X) = P(Y = 1)P(X \mid Y = 1) + P(Y = 0); P(X \mid Y = 0)$

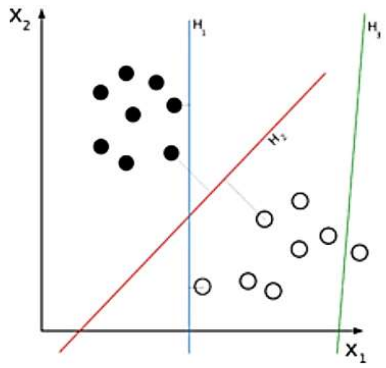
The component is calculated as follows: $P(Y = 1)$ is the error rate of the sample used to run the model, assuming the variables are independent.

3.2.4. Support vector machine

Support vector machine (SVM) is a binary classification algorithm. It takes input and classifies them into two different classes. Given a set of training examples belonging to two given categories, the SVM algorithm builds an SVM model to classify other examples into those two categories. The SVM builds/learns a hyperplane to classify the dataset into two separate classes. To do this SVM will construct a hyperplane or a set of hyperplanes in a multi-dimensional or infinite-dimensional space, which can be used for classification, regression, or other tasks. For the best classification, it is necessary to determine the optimal hyperplane located as far away from the data points of all classes as possible. In general, the larger the margin, the greater the generalization error of the algorithm the smaller the classification.

Figure 2 depicts the SVM algorithm. Given a training set represented in a vector space where each document is a point, this method finds a decision hyperplane that can best divide the points on this space into two separate layers, respectively, the layer with the data containing the feature simulated by the black dot and the layer with the data containing the feature simulated by the white dot. The quality of this hyperplane is determined by the distance (called the boundary) of the nearest data point of each layer to this plane. The larger the boundary distance, the better the decision plane and the more accurate the classification. The purpose of the SVM algorithm is to find the maximum boundary distance.

Figure 2. Support vector machine



3.2.5. K-nearest neighbors

The k-nearest neighbors (KNN) algorithm is very commonly employed in the field of data mining. KNN is a method to classify objects based on the closest distance between the object to be classified (query point) and all the objects in the training data. An object is classified based on its K-neighbors. K is a positive integer that is determined before the execution of the algorithm. Euclidean distance is often used to calculate the distance between objects.

3.2.6. Random forest

Random forest is an attribute classification method developed by Leo Breiman at the University of California, Berkeley (Breiman, 2001). Breiman is also the co-author of the classification and regression trees method which is rated as one of ten data mining methods. In a random forest, a significant improvement in classification accuracy results from the growth of a set of trees, each of which "votes" for the most popular class. To develop these sets of trees, normally random vectors are generated, which will govern the growth of each tree term in the sets. For the *kth* tree in the set of trees, a random vector *Vk* is generated, which is independent of the previously generated vectors *V1*, *V2*, ..., *Vk-1* but the distribution of the vectors is similar. A tree is grown based on the training set and the resulting vector *Vk* is a subclass *h(x, Vk)* where *x* is the input vector. After a large number of trees are created these trees "vote" for the most popular class.

3.3. Evaluation methods

In this study, in addition to measuring accuracy, in the case of severely imbalanced data, the use of accuracy as a measure of model evaluation is often ineffective because most of them are all very accurate. A stochastic model that predicts that the label belongs to the majority group will also yield results close to 100%. Then we consider a number of alternative metrics such as Precision, Recall, and F1-score. These indicators will not be too large to lead to a misconception of accuracy, and at the same time, they will focus more on getting accurate results to evaluate the accuracy of the minority group, which we want to forecast more accurately than the majority group.

Figure 3. The cross-statistical results between the forecast labels

		Actual		
		Positive	**Negative**	
Predicted	**Positive**	True positive (TP)	False positive (FP) Type I error	Precision = TP/(TP + FP)
	Negative	False negative (FN) Type II error	True negative (TN)	
		Recall = TP/(TP + FN)	False positive rate (FPR) = FP/(FP + TN)	

Positive corresponds to label 1 (financial distress) and Negative corresponds to label 0 (normal). From Figure 3, the meanings indicators as below:

- Precision: The level of prediction accuracy in the forecasted cases is Positive: *Precision = TP/(TP + FP).*
- Recall: The level of accurate prediction of cases is positive in actual cases is Positive: *Recall = TP/(TP + FN).*
- F1-score: Harmonic mean between Precision and Recall. This is an ideal surrogate metric for accuracy when the model has a high sample imbalance rate: *F1-score = 2/(1/Precision + 1/F Recall).*
- AUC (area under curve): Represents the relationship between sensitivity (sensitivity) and specificity (specificity). Assess the ability to classify financial distress and normality predicted from the model. Values of AUC less than 0.6 indicate poor predictive ability of the model, AUC between 0.8 and 0.9 is quite good, and above 0.9 is good.

A model with all the above indicators in the high range has better predictive quality. In this study, we use Accuracy, Precision, Recall, F1-score, and AUC as a measure of model evaluation.

3.4. Research data

This research uses data collected from the Vietnam Stock Exchange in the period from 2009 to 2020. Data are collected from audited financial statements of listed firms after excluding firms in the fields of banking, securities, and insurance sectors since these fields' characteristics are much different from other fields. After determining the indicators, the data used to perform the analysis and forecast is 4,936 observations, presented in Table 2.

Table 2. Statistics of research samples

Panel A: Data by year			*Panel B: Data by sectors*		
Year	***No. observations***	***Percentage***	***Sectors***	***No. of observations***	***Percentage***
2009	218	4.4%	Real estate and construction	1,788	36.2%
2010	316	6.4%	Technology	143	2.9%
2011	415	8.4%	Industry	568	11.5%
2012	424	8.6%	Service	565	11.4%
2013	443	9.0%	Consumer goods	416	8.4%
2014	440	8.9%	Energy	388	7.9%
2015	470	9.5%	Agriculture	421	8.5%
2016	490	9.9%	Materials	488	9.9%
2017	506	10.3%	Medical	159	3.2%
2018	508	10.3%			
2019	443	9.0%			
2020	263	5.3%			
Total	**4,936**	**100.0%**	**Total**	**4,936**	**100.0%**

Source: Authors' assessment.

Based on using financial management measures according to each model, financial distress data are presented in Table 3, whereby when measured according to the models of Altman, Fich and Slezak, and Zmijewski, it is 50.61%, 25.65%, and 9.83%, respectively. Appendix A, Table A.2 and Table A.3 provide information on mean, standard deviation, and minimum and maximum values between the financial accounting firms and the normal for the three models, respectively.

Table 3. Financial distress by models

Types	*Model 1 — Altman*		*Model 2 — Fich and Slezak*		*Model 3 — Zmijewski*	
	Freq.	***Percent***	***Freq.***	***Percent***	***Freq.***	***Percent***
Normal	2,438	49.39	3,670	74.35	4,451	90.17
Financial distress	2,498	50.61	1,266	25.65	485	9.83

Source: Authors' assessment.

The next step is to select and test the results of the model. We randomly divide the dataset into a pilot set and a test set.

• Pilot set: Based on the input and target variables of the train set, we train the financial distress classification model. The obtained model will be evaluated on other independent data sets such as a test set.

• Test set: This is also a dataset with fields similar to the train set that are considered completely new observations. The test set should have the most similar distribution to the actual data that the user will generate to evaluate the applicability of the model in practice.

4. RESULTS AND DISCUSSION

Based on the random forest algorithm, which belongs to the class of ensemble models. The results of the algorithm are based on majority election from many decision trees, so the model has higher reliability and better accuracy than simple linear classification models such as logistic or linear regression. The results in Appendix B, Figure B.1, Figure B.2., and Figure B.3 have determined the importance of the attributes of 22 variables in the model.

In the next step, we select the important variables for regression instead of selecting as many variables as possible because of the limitations of having too many features: increased cost and computation time; too many explanatory variables can lead to over-fitting (i.e., the phenomenon that the model works very well on the train set but poorly on the test set); among the variables will be those that cause interference and reduce the quality of the model. In consequence, eight attributes are selected, which are performed data transformations through sklearn.preprocessing. The coefficients of variables in each model are illustrated in Table 4.

Table 4. Effect coefficients of each attribute on financial distress by model

Model 1 — Altman		*Model 2 — Fich and Slezak*		*Model 3 — Zmijewski*	
Variables	*Coeff.*	*Variables*	*Coeff.*	*Variables*	*Coeff.*
X13: Asset turnover	-3.5960	*X8: Margin*	-0.4141	*X5: Debt-to-equity ratio*	2.1389
X5: Debt-to-equity ratio	2.6649	*X9: Marginal operating income*	-4.6605	*X1: Current ratio*	-3.1924
X3: Receivable turnover	0.0261	*X11: Net return on equity*	-13.0787	*X8: Margin*	29.6188
X1: Current ratio	-1.0672	*X5: Debt-to-equity ratio*	1.0379	*X9: Marginal operating income*	-34.4099
X14: Inventory turnover period	-0.0077	*X10: Net return to equity book value*	-7.5967	*X2: Quick ratio*	-2.4527
X2: Quick ratio	0.9964	*X12: Ratio of operating income to book value of equity*	-7.4179	*X3: Receivable turnover*	-0.2399
X7: Operating cash flow to total debt	-0.1734	*X22: Earnings per share*	0.0001	*X14: Inventory turnover period*	0.0002
X15: Fixed asset turnover	-0.0010	*X1: Current ratio*	-0.1625	*X13: Asset turnover*	-0.9261

Source: Authors' assessment

Table 5 reveals the results of the accuracy of models (Accuracy), logistic regression, support vector machine, decision tree, random forest, K-nearest neighbors, Bayesian network algorithms in Model 1 — Altman is the lowest 0.81, the highest is 0.98, in Model 2 — Fich and Slezak the lowest is 0.81, the highest is 0.90 and in Model 3 — Zmijewski the lowest is 0.81 and the highest is 0.98. Of the six algorithms used to forecast financial information, the random forest algorithm achieved the highest accuracy with a rate of 98%.

Table 5. Accurate prediction results of each algorithm and model

No.	*Method*	*Model 1 — Altman*	*Model 2 — Fich and Slezak*	*Model 3 — Zmijewski*
1	Logistic regression	0.93	0.89	0.97
2	Support vector machine	0.94	0.89	0.97
3	Decision tree	0.97	0.86	0.98
4	Random forest	**0.98**	0.90	**0.98**
5	K-nearest neighbors	0.81	0.84	0.93
6	Bayesian network	0.87	0.81	0.81

Source: Authors' assessment.

We use other metrics for more comprehensive testing. As shown in Table 5, the random forest algorithm gives the highest prediction accuracy, so we base this algorithm to measure the accuracy according to the measures of Precision, Recall, and F1-score. Table 6 shows that Model 1 and Model 3 give the best findings, especially Model 3.

Table 6. Forecasting results of each model according to the random forest algorithm

Models		*Precision*	*Recall*	*F1-score*
Model 1 — Altman	Normal	0.98	0.97	0.97
	Financial distress	0.97	0.98	0.97
Model 2 — Fich and Slezak	Normal	0.92	0.95	0.93
	Financial distress	0.84	0.75	0.79
Model 3 — Zmijewski	Normal	0.99	0.99	0.99
	Financial distress	0.93	0.90	0.92

Source: Authors' assessment.

However, in Model 3, it reveals that the results of measuring the accuracy of the financial distress group and the normal group have a small difference; in normal conditions, the accuracy reaches 99%, while the accuracy for the case of financial distress is only 93%, 90%, and 92%, respectively for Models 1, 2, and 3. The reason for the small differences may be due to the imbalance of data for Model 3. Based on the data in Table 3, the proportion of observations is subject to financial distress for only 9.83%. To deal with unbalanced data, we use these techniques: under-sampling, over-sampling, and synthetic minority over-sampling (SMOTE). The results of Table 7, after processing the unbalanced data, give very good results and there is no big difference in the measurement of the predictive level of Model 3 for the normal group and the financial distress group.

Table 7. Prediction results of Model 3 — Zmijewski with unbalanced samples

Methods		***Precision***	***Recall***	***F1-score***
Original data	Normal	0.99	0.99	0.99
	Financial distress	0.93	0.9	0.92
Method 1: Under-sampling	Normal	1.00	0.95	0.97
	Financial distress	0.95	1.00	0.97
Method 2: Over-sampling	Normal	1.00	0.98	0.99
	Financial distress	0.98	1.00	0.99
Method 3: SMOTE	Normal	1.00	0.98	0.99
	Financial distress	0.98	1.00	0.99

Source: Authors' assessment.

The AUC index measures the area under the receiver operating characteristic (ROC) curve, indicating whether the classification ability of the normal/financial distress group of the algorithms presented above is strong or weak. AUC ∈ [0, 1], the larger its value, the better the model. The GridSearch random forest algorithm (using GridSearch is a technique to help find the right parameters for the model) and decision tree achieve a high prediction accuracy rate, AUC = 0.97 for Model 1, and AUC = 0.95 for Model 3 (see Appendix C, Figure C.1, Figure C.2, and Figure C.3). This means that the model's predictive ability is good and can be applied in practice.

For the decision tree algorithm, we consider the three most important indicators of each model. According to Appendix D, Figure D.1, Figure D.2, and Figure D.3, for Model 1, the expenditure *X13: Total asset turnover* is the most important attribute to forecast financial results for firms, which shows that when total asset turnover is less than 1.461, the firm is forecasted as financially distressed. Similar to Model 2, when *X8: Profit margin of the firm* is less than 2.1% and Model 3, when *X5: Debt-to-equity ratio* is greater than 3.208, the firm falls into financial distress. This result is consistent with the results of Kim and Upneja (2014) when the debt-to-equity ratio decreases, the level of productivity is high, the asset turnover and profit margin are low, and the firm will be led to a state of financial distress.

5. CONCLUSION

This study aims to determine the direction of the impact of determinants on the possibility of financial distress and predict the probability of financial distress for listed firms on the Vietnam Stock Exchange in the period from 2009 to 2020. The results show that the debt-to-equity ratio has a positive impact on financial distress; but asset turnover, and profit margin negatively influences financial distress. This forecasting model has an overall correct prediction rate of 98%. Model 1 — Altman and Model 3 — Zmijewski are models capable of predicting financial distress at a high level. The research reveals that internal indicators in each firm will directly affect the probability of financial distress, corresponding to each model. The study has added to practice about issues management in firms which are the most important determinants determining the "health" status of firms.

Based on the findings, the regression coefficients of the independent variables in the models illustrate that as the debt ratio increases, the financial distress increases. Therefore, the more debt a firm has, the higher the risk of default increases, and increase the risk of firm financial distress. The greater the efficiency of asset usage, the lower the financial distress. This is completely consistent with the fact that the more revenue a firm produces, the less likely it is to become financially distressed as the firm sells more products. The higher the profit margin, the lower the probability of financial distress. When a firm has internal funding available, the firm will be proactive in investing, limiting external debt, thereby minimizing the possibility of financial distress.

This study has some limitations. First, no comprehensive data on the financial distress and bankruptcy of firms is collected. Second, no data of non-financial information such as corporate governance or some macro issues are put in the models. In the future, to achieve better and more comprehensive results, we continue to add macro-environmental and market determinants that influence firm distress as well as compare between different business lines. Therefore, in-depth study by sector helps managers realize the importance of investment, financing, and business operations management.

REFERENCES

1. Aktas, B., & Mahaffy, H. J. (1996). A two-phase level tracking method. *Nuclear Engineering and Design, 162*(2–3), 271–280. https://doi.org/10.1016/0029-5493(95)01132-3
2. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x
3. Altman, E. I., & Hotchkiss, E. (2005). *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt* (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781118267806

4. 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. https://doi.org/10.1016/0378-4266(94)90007-8
5. Anandarajan, M., Lee, P., & Mnandarajan, A. (2001). Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of artificial neural networks. *Intelligent Systems in Accounting, Finance & Management, 10*(2), 69-81. https://doi.org/10.1002/isaf.199
6. Andrade, G., & Kaplan, S. N. (1998). How costly is financial (not economic) distress? Evidence from highly leveraged transactions that became distressed. *The Journal of Finance, 53*(5), 1443–1493. https://doi.org/10.1111/0022-1082.00062
7. Beaver, W. H. (1996). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71–111. https://doi.org/10.2307/2490171
8. Belson, W. A. (1959). Matching and prediction on the principle of biological classification. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 8*(2), 65–75. https://doi.org/10.2307/2985543
9. Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association, 39*(227), 357–365. https://doi.org/10.1080/01621459.1944.10500699
10. Biberaj, A., Sheme, E., Rakipi, A., Xhaferllari, S., Kushe, R., & Alinci, M. (2022). Cyber attack against e-Albania and its social, economic and strategic effects. *Journal of Corporate Governance, Insurance, and Risk Management, 9*(2), 341–347. https://doi.org/10.56578/jcgirm090204
11. Binti, S., Zeni, M., & Ameer, R. (2010). Turnaround prediction of distressed companies: Evidence from Malaysia. *Journal of Financial Reporting and Accounting, 8*(2), 143–159. https://doi.org/10.1108/19852511011088398
12. Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324
13. Campbell, J. Y., Hilscher, J., & Szilahyi, J. (2008). In search of distress risk. *The Journal of Finance, 63*(6), 2899–2939. https://doi.org/10.1111/j.1540-6261.2008.01416.x
14. Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis* (2nd ed.). Chapman & Hall/CRC.
15. Chen, Y.-S., Lin, C.-K., Lo, C.-M., Chen, S.-F., & Liao, Q.-J. (2021). Comparable studies of financial bankruptcy prediction using advanced hybrid intelligent classification models to provide early warning in the electronics industry. *Mathematics, 9*(20), Article 2622. https://doi.org/10.3390/math9202622
16. Denis, D. J., & Denis, D. K. (1995). Causes of financial distress following leveraged recapitalizations. *Journal of Financial Economics, 37*(2), 129–157. https://doi.org/10.1016/0304-405X(94)00792-Y
17. Fich, E. M., & Slezak, S. L. (2008). Can corporate governance save distressed firms from bankruptcy? An empirical analysis. *Review of Quantitative Finance and Accounting, 30*(2), 225–251. https://doi.org/10.1007/s11156-007-0048-5
18. Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: An application using bankruptcy data. *Information & Management, 24*(3), 159–167. https://doi.org/10.1016/0378-7206(93)90064-Z
19. Flitman, A. M. (1997). Towards analysing student failures: Neural networks compared with regression analysis and multiple discriminant analysis. *Computers & Operations Research, 24*(4), 367–377. https://doi.org/10.1016/S0305-0548(96)00060-3
20. Fuentes, F., Herrera, R., & Clements, A. (2023). Forecasting extreme financial risk: A score-driven approach. *International Journal of Forecasting, 39*(2), 720–735. https://doi.org/10.1016/j.ijforecast.2022.02.002
21. Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research, 241*(1), 236–247. https://doi.org/10.1016/j.ejor.2014.08.016
22. Ghazali, A. W., Shafie, N. A., & Sanusi, Z. M. (2015). Earnings management: An analysis of opportunistic behaviour, monitoring mechanism and financial distress. *Procedia Economics and Finance, 28*, 190–201. https://doi.org/10.1016/S2212-5671(15)01100-4
23. 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), Article 3954. https://doi.org/10.3390/su12103954
24. Hallunovi, A. (2023). Financial results vs. implementation of accounting practices. *Journal of Governance & Regulation, 12*(2), 19–26. https://doi.org/10.22495/jgrv12i2art2
25. He, Q., Chong, T. T.-L., Li, L., & Zhang, J. A. (2010). A competing risks analysis of corporate survival. *Financial Management, 39*(4), 1697–1718. https://doi.org/10.1111/j.1755-053X.2010.01127.x
26. Heo, J., & Yang, J. Y. (2014). AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing, 24*, 494–499. https://doi.org/10.1016/j.asoc.2014.08.009
27. Hu, Y.-C., & Ansell, J. (2006). Developing financial distress prediction models: A study of US, Europe and Japan retail performance (Working paper). University of Edinburgh. http://www.centerforpbbefr.rutgers.edu/2006/Paper%202006/07AS-127-Developing%20Financial.pdf
28. Julian, L., Johari, R. J., Said, J., & Wondabio, L. S. (2022). Fraud risk judgement measurement scale development [Special issue]. *Journal of Governance & Regulation, 11*(1), 303–311. https://doi.org/10.22495/jgrv11i1siart10
29. Kim, S. Y., & Upneja, A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling, 36*, 354–362. https://doi.org/10.1016/j.econmod.2013.10.005
30. Kuizinienė, D., Krilavičius, T., Damaševičius, R., & Maskeliūnas, R. (2022). Systematic review of financial distress identification using artificial intelligence methods. *Applied Artificial Intelligence, 36*(1), Article 2138124. https://doi.org/10.1080/08839514.2022.2138124
31. Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques — A review. *European Journal of Operational Research, 180*(1), 1–28. https://doi.org/10.1016/j.ejor.2006.08.043
32. Malik, C., & Yadav, S. (2020). Forecasting and asymmetric volatility modeling of sustainability indexes in India. *Corporate Governance and Sustainability Review, 4*(1), 38–53. https://doi.org/10.22495/cgsrv4i1p5
33. Martono, N. P., & Ohwada, H. (2023). Financial distress model prediction using machine learning: A case study on Indonesia's consumers cyclical companies. In I. Koprinska, P. Mignone, R. Guidotti, S. Jaroszewicz, H. Fröning, F. Gullo, P. M. Ferreira, D. Roqueiro, G. Ceddia, S. Nowaczyk, J. Gama, R. Ribeiro, R. Gavaldà, E. Masciari, Z. Ras, E. Ritacco, F. Naretto, A. Theissler, P. Biecek, … S. Pashami (Eds.), *Machine learning and principles and practice of knowledge discovery in databases* (pp. 53–61). https://doi.org/10.1007/978-3-031-23633-4_5

34. Marzuki, H., Hasnan, S., & Ali, M. M. (2022). Comtemporary review of corruption risk studies [Special issue]. *Corporate Governance and Organizational Behavior Review, 6*(2), 255–267. https://doi.org/10.22495/cgobrv6i2sip10
35. Munsif, V., Raghunandan, K., Rama, D. V., & Singhvi, M. (2011). Audit fees after remediation of internal control weaknesses. *Accounting Horizons, 25*(1), 87–105. https://doi.org/10.2308/acch.2011.25.1.87
36. Nguyen, L. Q. T., & Ahmed, R. (2023). The impact of economic sanctions on foreign direct investment: Empirical evidence from global data. *Journal of Economics and Development, 25*(1), 79–99. https://doi.org/10.1108/JED-10-2022-0206
37. 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
38. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106. https://doi.org/10.1007/BF00116251
39. Quinlan, J. R. (1996). Bagging, boosting, and C4.5. In *Proceedings of the AAAI-96* (pp. 725–730). The AAAI Press. https://cdn.aaai.org/AAAI/1996/AAAI96-108.pdf
40. Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems, 17*(3), 227–238. https://doi.org/10.1016/0167-9236(95)00033-X
41. Theodossiou, P., Kahya, E., Saidi, R., Philippatos, G. (1996). Financial distress and corporate acquisitions: Further empirical evidence. *Journal of Business Finance & Accounting, 23*(5–6), 699–719. https://doi.org/10.1111/j.1468-5957.1996.tb01149.x
42. Tinico, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis, 30*, 394–419. https://doi.org/10.1016/j.irfa.2013.02.013
43. Tunio, F. H., Ding, Y., Agha, A. N., Agha, K., & Panhwar, H. U. R. Z (2021). Financial distress prediction using AdaBoost and bagging in Pakistan Stock Exchange. *The Journal of Asian Finance, Economics, and Business, 8*(1), 665–673. https://doi.org/10.13106/jafeb.2021.vol8.no1.665
44. 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. https://doi.org/10.1023/A:1011291425075
45. Wang, G., & Ma, J. A. (2012). A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine. *Expert Systems with Applications, 39*(5), 5325–5331. https://doi.org/10.1016/j.eswa.2011.11.003
46. Wanke, P., Barros, C. P., & Faria, J. R. (2015). Financial distress drivers in Brazilian banks: A dynamic slacks approach. *European Journal of Operational Research, 240*(1), 258–268. https://doi.org/10.1016/j.ejor.2014.06.044
47. Wruck, K. H. (1990). Financial distress, reorganization, and organizational efficiency. *Journal of Financial Economics, 27*(2), 419–444. https://doi.org/10.1016/0304-405X(90)90063-6
48. Zhang, Z., Wu, C., Qu, S., & Chen, X. (2022). An explainable artificial intelligence approach for financial distress prediction. *Information Processing and Management, 59*(4), 34–51. https://doi.org/10.1016/j.ipm.2022.102988
49. Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research, 22*, 59–82. https://doi.org/10.2307/2490859

APPENDIX A

Table A.1. Attributes in the research models

Index group	Attribute name	Code	Measurement
Solvency	Current ratio	X1	Current assets/Short-term liabilities
	Quick ratio	X2	(Current assets – inventories)/Current liabilities
	Receivables turnover	X3	Net revenue/Receivables
	Ratio of operating cash flow to short-term debt	X4	Operating cash flow/Short-term debt
Capital structure and debt serviceability	Debt-to-equity ratio	X5	Total liabilities/Equity
	Ratio of fixed assets to long-term capital	X6	Fixed assets/(Total capital – Current liabilities)
	Operating cash flow to total debt ratio	X7	Operating cash flow/Total liabilities
Profitability	Margin profit	X8	Net profit/Net revenue
	Marginal operating income	X9	Operating income/Net revenue
	The ratio of net return to book value of equity	X10	Net profit/Book value of shares
	Net return on equity	X11	Net profit/Equity
	Ratio of operating income to book value of equity	X12	Operating income/Book value of shares
Activity	Asset turnover	X13	Net Revenue/Total assets
	Inventory turnover period	X14	Inventory x 365/Cost of goods sold
	Fixed assets turnover	X15	Net revenue/Fixed assets
Growth indicators	Increase the revenue	X16	Net sales in year *t*/Net sales in year *t–1*
	Asset growth	X17	Total assets in year *t* /Assets in year *t–1*
	Operating income growth	X18	Operating income year t / Operating income year *t–1*
	Net profit growth	X19	Net profit year *t*/Net profit year *t–1*
	Equity growth	X20	Equity year *t*/Equity year *t–1*
Others	Stock price trend	X21	Stock price in year *t*/Share price in year *t–1*
	Earnings per share	X22	Net profit/Average number of shares outstanding

Table A.2. Descriptive statistics of attributes

Variables	Obs.	Mean	Std. Dev.	Min	Max
X1: Current ratio	4,936	2.037	1.740	0.382	17.298
X2: Quick ratio	4,936	1.437	1.560	0.151	15.740
X3: Receivable turnover	4,936	7.827	10.497	0.095	104.052
X4: Operating cash flow to current debt ratio	4,936	0.268	0.595	-2.248	4.706
X5: Debt-to-equity ratio	4,936	1.525	1.473	0.039	10.683
X6: Ratio of fixed assets to long-term capital	4,936	0.347	0.299	0.000	2.074
X7: Operating cash flow to total debt	4,936	0.197	0.468	-1.667	4.122
X8: Margin	4,936	0.075	0.103	-0.569	0.658
X9: Marginal operating income	4,936	0.092	0.120	-0.604	0.777
X10: Net return to equity book value	4,936	0.123	0.105	-0.368	0.548
X11: Net return on equity	4,936	0.119	0.102	-0.368	0.513
X12: Ratio of operating income to book value of equity	4,936	0.153	0.125	-0.368	0.684
X13: Asset turnover	4,936	1.171	0.907	0.028	6.855
X14: Inventory turnover period	4,936	-171.654	391.325	-4725.147	0.000
X15: Fixed asset turnover	4,936	21.198	49.483	0.168	593.817
X16: Revenue growth	4,936	1.154	0.568	0.130	10.091
X17: Asset growth	4,936	1.125	0.311	0.401	5.336
X18: Operating income growth	4,936	1.178	1.961	-13.866	27.579
X19: Net profit growth	4,936	1.209	2.570	-19.508	37.683
X20: Equity growth	4,936	1.126	0.330	0.433	6.493
X21: Stock price trend	4,936	1.175	0.562	0.289	4.250
X22: Earnings per share	4,936	2,293.371	2,330.574	-3,106.000	14,163.000

Table A.3. Comparison of attributes among the research models

Variables	Model 1 — Altman		Model 2 — Fich and Slezak		Model 3 — Zmijewski	
	Normal	Financial distress	Normal	Financial distress	Normal	Financial distress
X1: Current ratio	2.605	1.482	2.267	1.369	2.138	1.109
X2: Quick ratio	1.910	0.975	1.637	0.856	1.525	0.630
X3: Receivable turnover	11.221	4.515	8.770	5.095	8.236	4.080
X4: Operating cash flow to current debt ratio	0.406	0.133	0.333	0.078	0.292	0.052
X5: Debt-to-equity ratio	0.877	2.157	1.185	2.512	1.152	4.953
X6: Ratio of fixed assets to long-term capital	0.324	0.369	0.313	0.444	0.333	0.474
X7: Operating cash flow to total debt	0.344	0.054	0.249	0.046	0.216	0.019
X8: Margin	0.079	0.071	0.099	0.006	0.082	0.017
X9: Marginal operating income	0.096	0.088	0.120	0.013	0.100	0.024
X10: Net return to equity book value	0.152	0.094	0.155	0.030	0.127	0.080
X11: Net return on equity	0.149	0.090	0.151	0.029	0.124	0.077
X12: Ratio of operating income to book value of equity	0.188	0.118	0.190	0.044	0.157	0.108
X13: Total asset turnover	1.678	0.677	1.246	0.955	1.199	0.914
X14: Inventory turnover period	-72.681	-268.249	-152.311	-227.727	-162.362	-256.928
X15: Fixed asset turnover	26.351	16.169	23.296	15.116	21.646	17.088
X16: Revenue growth	1.134	1.174	1.167	1.116	1.157	1.127
X17: Asset growth	1.093	1.156	1.141	1.076	1.121	1.158
X18: Operating income growth	1.214	1.142	1.346	0.690	1.208	0.905
X19: Net profit growth	1.230	1.188	1.396	0.665	1.243	0.896
X20: Equity growth	1.120	1.131	1.146	1.069	1.130	1.083
X21: Stock price trend	1.201	1.150	1.220	1.045	1.180	1.132
X22: Earnings per share	2935.285	1666.875	2898.908	537.982	2399.224	1321.918

APPENDIX B

Figure B.1. Importance of attributes in Model 1 — Altman

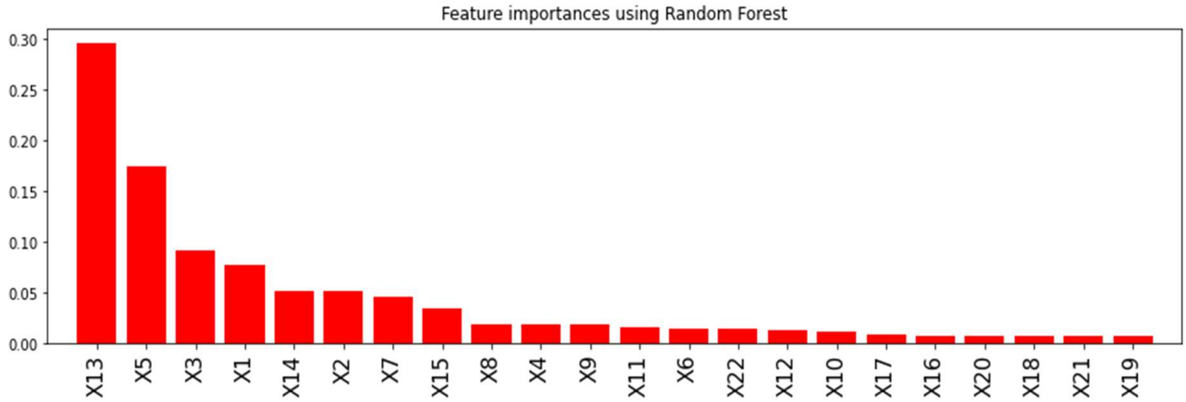


Figure B.2. Importance of attributes in Model 2 — Fich and Slezak

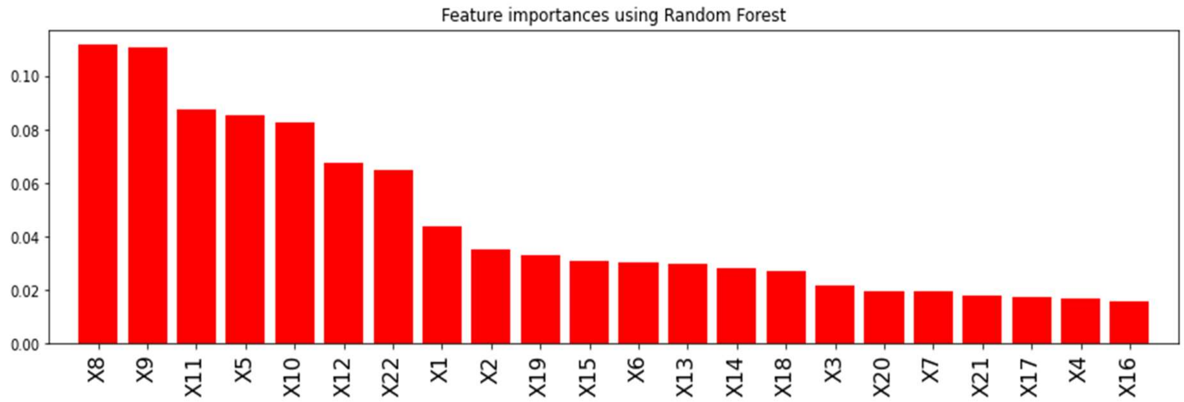
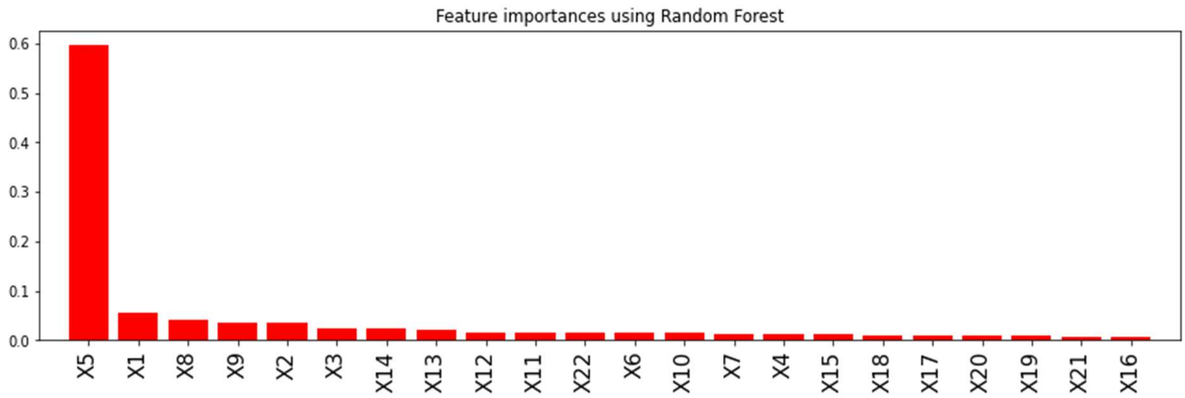


Figure B.3. Importance of attributes in Model 3 — Zmijewski



APPENDIX C

Figure C.1. AUC prediction results of algorithms for Model 1 — Altman

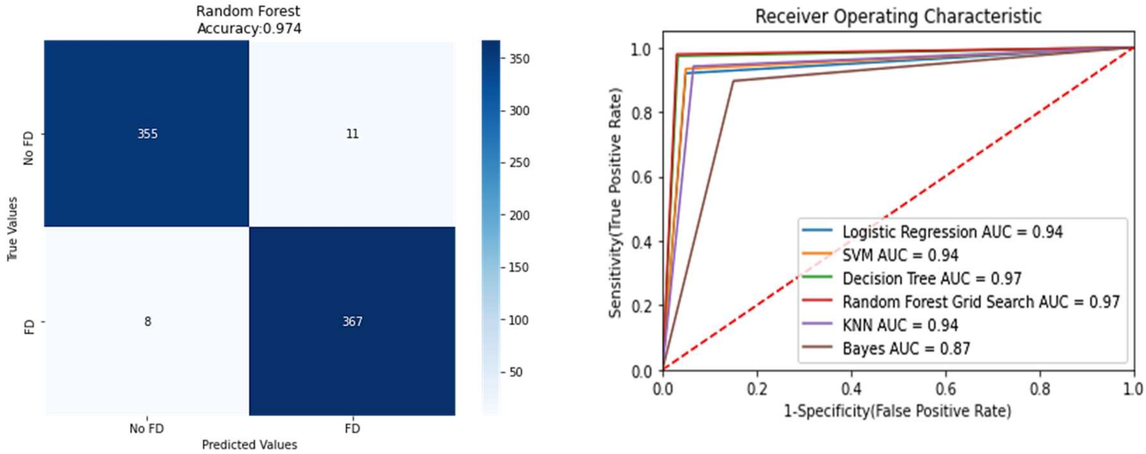


Figure C.2. AUC prediction results of algorithms for Model 2 — Fich and Slezak

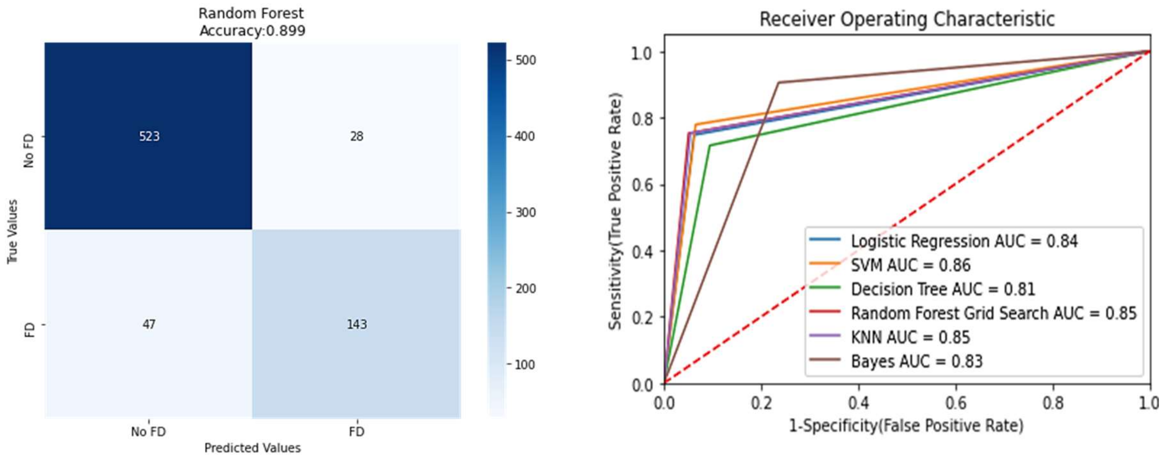
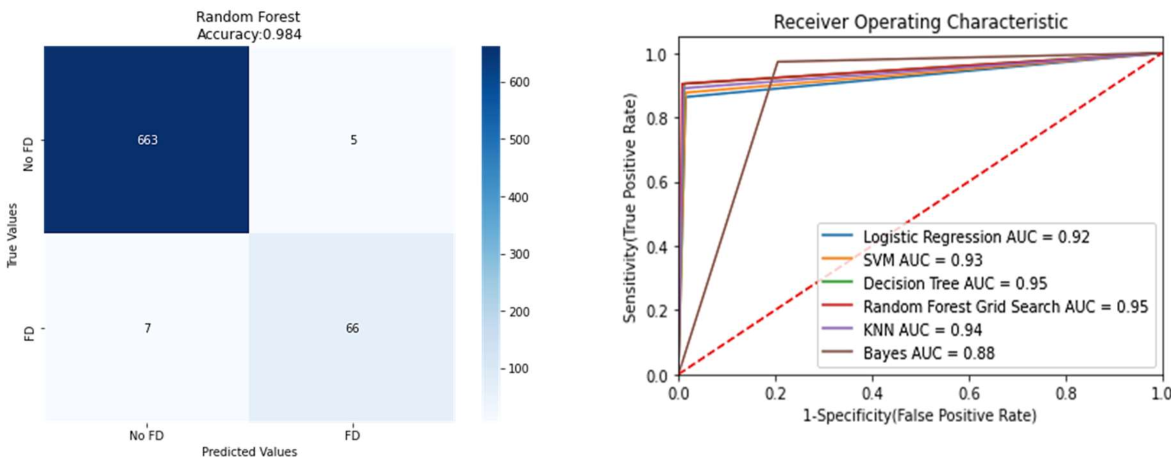


Figure C.3. AUC prediction results of algorithms for Model 3 — Zmijewski



APPENDIX D

Figure D.1. Decision tree algorithm results for Model 1 — Altman

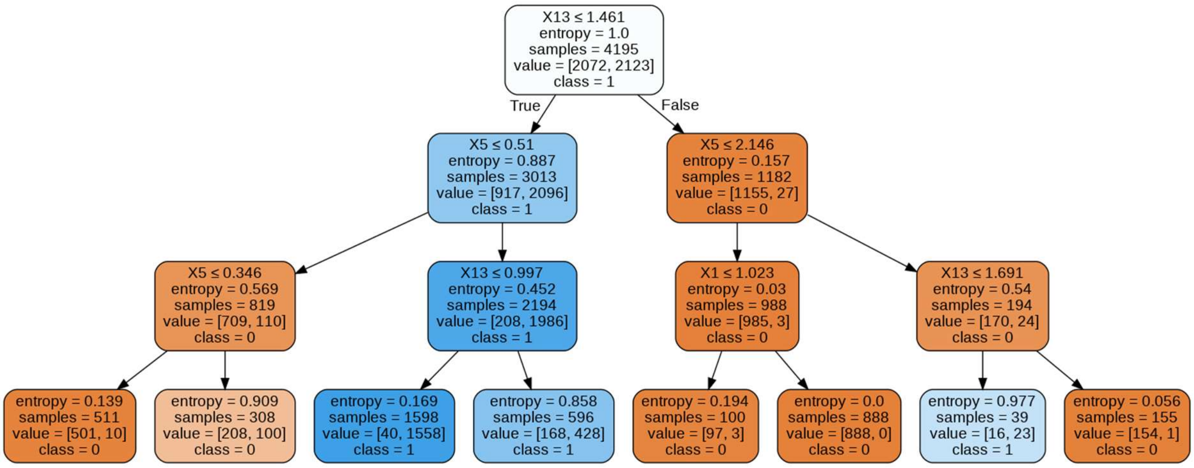


Figure D.2. Decision tree algorithm results for Model 2 — Fich and Slezak

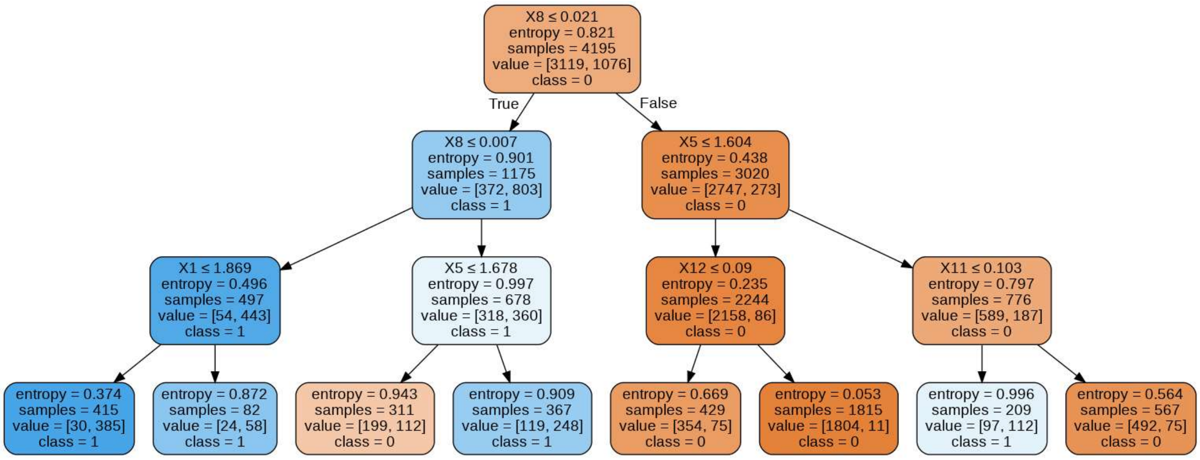


Figure D.3. Decision tree algorithm results for Model 3 — Zmijewski