JOURNAL OF CULTURAL ANALYSIS AND SOCIAL CHANGE

Journal of Cultural Analysis and Social Change,
2025, 10(3), 209-222
ISSN: 2589-1316

Leveraging Machine Learning for Corporate Fraud Detection: A Random Forest Study

Szu-Hsien Lin[1], Tzu-Pu Chang[2], Huei-Hwa Lai[3]*, Yan Ting Chen[4]

[1]*Accounting and Information Systems, Asia University, Taiwan R.O.C.*
[2]*Department of Finance, National Yunlin University of Science and Technology, Taiwan R.O.C.*
[3]*Department of Business and Administration, ChaoYang University of Technology, Taiwan R.O.C*
[4]*Department of Finance, National Yunlin University of Science and Technology, Taiwan R.O.C.*

***Corresponding Author:** edithlai2005@gm.cyut.edu.tw

Citation: Lin, S.-H., Chang, T.-P., Lai, H.-H., & Chen, Y. T. (2025). Leveraging Machine Learning for Corporate Fraud Detection: A Random Forest Study. *Journal of Cultural Analysis and Social Change*, *10*(3), 209–222. https://doi.org/10.64753/jcasc.v10i3.2399

Published: November 26, 2025

ABSTRACT

The occurrence of corporate fraud often results in significant losses to stakeholders and society. Therefore, this study aims to construct a model to predict corporate fraud, with the goal of providing early warnings of potential fraudulent activities. The research focuses on fraudulent listed companies in Taiwan and selects matching non-fraudulent companies at a ratio of 1:2 as the research sample. To comprehensively capture the factors contributing to fraud, 53 indicators are selected from four dimensions: financial statements, corporate governance, market transactions, and the overall economy. This study further categorizes fraud methods into financial statement fraud and non-financial statement fraud (i.e., hollowing out/misappropriating assets/manipulating stock prices), and applies machine learning techniques, specifically decision tree and random forest algorithms, for prediction and analysis. The empirical results indicate that: (1) the random forest method, based on ensemble learning, achieves higher prediction accuracy than the decision tree model, and the prediction accuracy improves when fraud methods are distinguished; (2) the type I error of the random forest model is zero, meaning that if the model predicts a company as fraudulent, fraud will occur in the following year; and (3) the detailed techniques of fraud evolve structurally over time, leading to a relatively high type II error.

Keywords: Corporate fraud prediction, Machine learning, Random forest, Decision tree.

INTRODUCTION

The prediction of the occurrence of corporate financial crisis has always been a joint effort of the academic and practical circles. As early as the Z-score Model proposed by Altman (1968), the quantitative research on predicting the occurrence of corporate financial crisis has never stopped. In terms of foreign literature, the prediction models proposed by Ohlson (1980), Shumway (2001), Duffie *et al.* (2007) and Campbell *et al.* (2008) have attracted considerable attention; As for domestic literature, Chen et al. (2004), Xu et al. (2007) and Huang et al. (2012) also made a lot of contributions to local research. In fact, there are many reasons or categories behind the financial crisis of enterprises, among which "corporate fraud" has attracted the attention of the industry, government and academia.[1] Because the amount involved in corporate fraud is often quite large and seriously affects the rights and interests of minority shareholders, investors, employees and creditors, the causes and prediction of corporate fraud have aroused the research interest of scholars.

[1] For example, the Taiwan Economic Journal (TEJ) classifies corporate financial crises into 9 types of financial crises and 7 types of quasi financial crises. Among them, such as hollowing out and misappropriation, the chairman bounces or doubts about continuing operation may be caused by corporate fraud.

The information transparency of Taiwan's capital market was not high in the past, and due to the rising awareness of investors and minority shareholders at that time and the incompleteness of government laws and regulations, a series of corporate fraud incidents broke out in Taiwan due to false financial statements, false increase of revenue, hollowing out or manipulation of stocks by backdoor listed companies. For example, the famous cases of Tuntex Group, Rebar Group and New Magnitude Group in the early days, these corporate fraud events involved illegal facts of tens of billions of dollars and affected the operation and development of several listed OTC companies.[2] Although in recent years, the competent authorities have continuously improved the financial supervision system and information transparency, hoping to detect the possibility of corporate fraud in advance, the continuous innovation of corporate financial instruments in Taiwan has also made the fraud means more and more complicated, so fraud incidents still occur from time to time.

This paper uses Figure 1 to further illustrate the past and current situation of corporate fraud in Taiwan. It can be seen from the figure that the number of frauds of listed enterprises reached a peak in 1998 and 2004. Although Taiwan was not affected much by the Asian financial turmoil at that time, because the supervision system of Taiwan's enterprises was not yet perfect, the board of directors, major shareholders and senior management had the opportunity to escape loopholes, and took the opportunity to hype and empty out in the market, resulting in a series of fraud incidents. In view of this, the competent authorities began to reform the corporate governance and financial supervision system, and the SARS incident broke out in 2003, causing the domestic enterprises to face serious challenges. In mid-2004, Boda Company filed for reorganization without warning. And its 6 billion 300 million NTD cash disappeared. In the same year, fraudulent incidents of companies such as Zhongfu, Kolin, etc. broke out one after another. Since 2004, the number of frauds has gradually decreased. However, with the increasingly developed technology network and the continuous innovation of various industries in the market, there are still a few fraudulent incidents, and the fraudulent operation methods adopted are relatively varied. Therefore, the competent authority must still face the problem of how to regulate the internal and external supervision of the company to prevent or early alert the occurrence of fraud, which also prompted this study to try to establish a model for predicting corporate fraud.

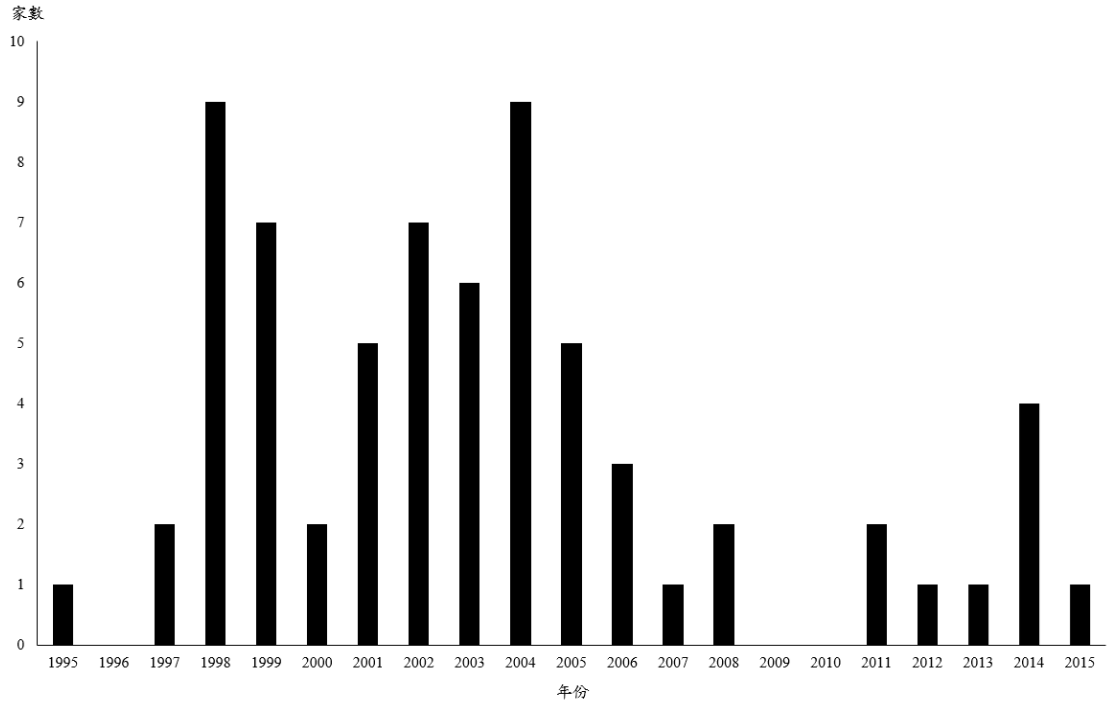

Figure 1. The number of fraudsters in listed companies from 1995 to 2015.

This study takes Taiwan listed companies with fraud events from 1997 to 2015 as the research object, removes the samples of the financial industry, and constructs the prediction model through the decision tree and random forest algorithm in the field of machine learning. Compared with the previous literature, this paper is expected to

[2] The case of Tuntex Group involved a hollowing out of 70 billion NTD, resulting in the withdrawal of Tun-Yun Textile and Chien Tai Cement from the market. The Rebar group case involved the hollowing out of 73.1 billion NTD, resulting in the delisting of China Rebar, Chia Hsin Food & Synthetic Fiber, Chung-Hwa Commercial Bank etc. in the group. The New Magnitude Group case involved Wu Tso Chin, the president of the group, who speculated on stocks in a backdoor listing mode, and transferred his personal shares to the head account and then maliciously defaulted and delivered, which involved hollowing out 5 billion NTD of assets, causing Taifang, Puda, Mingjia Leigh, Arthur, Xintaishen delisted and the loss of investors.

increase the depth and breadth of fraud research in three aspects: First, the literature believes that under the control of Generally Accepted Accounting Principles (GAAP) and International Financial Reporting Standards (IFRS), financial statements can best help investors and creditors understand the financial status and operating performance of enterprises; Therefore, financial statements and corporate governance were most often used for observation and analysis in the literature (Lin and Chang, 2009; Beasley, 1996). However, the falsity of financial statements itself is one of the fraud events, which means that the analysis of financial statements may not completely capture the possibility of fraud events. Therefore, in addition to the financial statements and corporate governance variables, this study increases the market transaction variables related to stock prices and transactions, as well as the overall economic variables such as interest rates and exchange rates that affect the company's financial operations. Second, corporate fraud is a general term, which contains a variety of different fraud means. Therefore, the Association of Certified Fraud Examiners (ACFE) mainly divided fraud into financial statement fraud, corruption and misappropriation of assets (ACFE, 2010). The existing foreign literature mostly studies financial statement fraud cases alone (such as Kotsiantis *et al.*, 2006; Kirkos *et al.*, 2007; Ravisankar *et al.*, 2011), and the domestic literature also lacks models that distinguish between different types of fraud. Thus, this study categorizes all fraud samples into two groups: financial statement fraud and non-financial statement fraud (including hollowing out, misappropriating assets, and manipulating stock prices), and develops corresponding prediction models for each. Third, the traditional prediction models mostly used logistic regression or probit regression to predict the probability of fraud (Summer and Sweeney, 1998; Chen *et al.*, 2006); in recent years, due to the maturity of data mining technology, there were also literatures using decision tree, neural network and other technologies as modeling tools (Kirkos *et al.*, 2007; Ravisankar *et al.*, 2011). As for this study, the "random forest" of machine learning is used to construct the prediction model. The random forest can be regarded as an enhanced decision tree algorithm, which has stronger learning ability and anti-noise ability. Since there is no application of random forest in fraud prediction in the literature, this study hopes to improve the accuracy and practicability of the prediction model.

The subsequent arrangements of this paper are as follows: Section 2 reviews the literature of enterprise fraud prediction in order to understand the possible candidate variables of the prediction model; section 3 describes the selection of research samples and variables, and introduces the prediction model and algorithm used; section 4 is the result of empirical analysis, and the conclusions and future research suggestions are made in section 5.

DISCUSSION ON THE LITERATURE OF ENTERPRISE FRAUD

Fraud refers to an act in which the fraudster intentionally or deliberately deceives others, resulting in a loss for the bona fide party or a gain for the fraudster. In accordance with Article 6 of Taiwan Auditing Standards Bulletin No. 43 "Investigation of Fraud in the Audit of Financial Statements", enterprise fraud refers to:

Fraud refers to the behavior of one or more of the management, governance units or employees who deliberately use deception and other methods to obtain improper or illegal benefits.

For example, when corporate fraud cases break out, the improper or illegal interests behind them are often hundreds of millions to tens of billions of dollars; compared with the other party, these losses should be borne by the whole society, especially the unwitting stakeholders, such as employees, shareholders, customers and creditors.

Compared with the research on corporate financial crisis, the research on corporate fraud started relatively late, about the end of the 1980s. The academic literature on corporate fraud can be divided into two main areas. The first focuses on exploring the factors that influence corporate fraud, while the second aims to establish models to predict fraud. The following literature review discusses these two areas separately.

Factors Affecting Corporate Fraud

There are many types of corporate fraud, and the causes behind it are naturally quite complex. Uzun *et al.* (2004) mentioned that according to the "Object of Fraud", it could be divided into four categories: fraud related to stakeholders, fraud related to the government, fraud in financial statements and violation of laws and regulations. According to "fraudulent practices", they can be divided into more categories. For example, Lin and Chang (2009) distinguished fraudulent methods into false financial reporting, misuse or misappropriation of company assets, obtaining assets at high prices for the purpose of benefiting others, and false and untrue transactions and false company, fake line numbers, insider transactions, breach of trust and bounced tickets and others. KPMG released the "Global Profiles of the Fraudster" report in 2016, which pointed out that 47% of the global fraud cases in 2013-2015 were misappropriation of assets and 22% were financial statement frauds. This was consistent with the aforementioned ACFE arguments (KPMG, 2016).

As for the various causes of fraud, Article 12 of the bulletin of Auditing Standards No. 43 summarizes it into three points: (1) Inducement or pressure refers to the individual's inability to meet the income in life or the pressure of the enterprise's internal and external on the management to set goals; (2) Opportunity refers to the fraudster's

ability to exceed the control or know the missing part of the internal control and (3) Rationalization of attitude or behavior, which refers to the deviation of the fraudster's personality and morality or psychological pressure to regard the fraudster as rationalization. Some scholars in the academic literature have put forward the same views. Early studies believed that the financial distress of enterprises would induce the motivation of management fraud; therefore, many empirical studies have found that the financial performance of enterprises is highly correlated with fraud related behaviors, such as Kinney and McDaniel (1989) and Loebbecke *et al.* (1989). Stice (1991) and Summers and Sweeney (1998) pointed out that Altman's Z-score, accounts receivable, and the ratio of inventory to total assets were significantly associated with fraud.

In addition to the fact that financial statements are a very key observation indicator, KPMG (2016) found that 61% of the reasons for fraud were due to the weak internal control of the organization; thus, scholars propose to examine the impact of the strength of internal control on fraud from the perspective of corporate governance. Beasley (1996) is a pioneer in this research direction. His study analyzed fraud cases in the United States and found that the higher the proportion of foreign directors in the composition of the board of directors, the lower the possibility of fraud. Uzun *et al.* (2004) also analyzed the fraud events in the United States and pointed out that the proportion of external directors and independent directors and whether to set up an audit committee and remuneration committee were helpful to explain the occurrence of fraud events. Chen *et al.* (2006) studied the impact of ownership structure in fraud cases in China from 1999 to 2003. The study found that external directors, the number of the board of directors and the seniority of the chairman had a significant impact on the possibility of fraud. As for the fraud research in Taiwan, Lin and Chang (2009) discussed the effects of abnormal changes of directors and supervisors and family enterprises. Although the results did not find the explanatory ability of the proportion of external directors, supporting abnormal changes of directors and supervisors was an important indicator to capture fraudulent companies. In addition, they also found that family enterprises were less prone to fraud.

In the literature on fraud factors, financial statement indicators and corporate governance are the main research variables. However, some scholars believe that observing the trading status of corporate stocks can explain fraud. The relationship between stock returns and their volatility and fraud had been found in previous studies. The reason behind this was that stocks were often a part of managers' compensation structure. Once there was abnormal excess returns or fluctuations, it would lead to managers' pressure or fraud-inducing behavior (Stice, 1991; Hackenbrack, 1993; Johnson *et al.*, 2009). Summers and Sweeney (1998) explored the relationship between insider trading behavior and fraud, and found that the number, quantity and amount of shares sold by insiders would affect the possibility of fraud. Ye et al. (2015) focused on the judgment cases of stock price manipulation in Taiwan, and pointed out that the common phenomenon of stock price manipulation was the slow rise of stock price in the initial stage, which meant that the trend of stock price was a factor to judge whether it was fraudulent or not.

Enterprise Fraud Detection and Prediction Model

Another main area of corporate fraud research is to establish a model to predict the probability of fraud, or use classifier technology to distinguish the samples into fraudulent / non-fraudulent companies, and finally evaluate the appropriateness and accuracy of the model. At the initial stage, relevant studies used econometric models, such as logistic regression and probit regression, to estimate the probability of fraud. Stice (1991) used variables such as financial reports, accounting firm attributes, stock price return fluctuations, and market capitalization to estimate the probability of companies being sued due to financial reports through probit regression, and calculated Type I and Type II errors. Summers and Sweeney (1998) used logistic regression to estimate the probability of fraud in corporate financial statements, with a probability of 0.5 as the threshold for identifying fraud; the results showed that the correct rate was 66.7%. Spathis (2002) also used logistic regression to analyze the financial statement fraud cases in Greece. The overall prediction accuracy of this study reached 84.21%.

In the past two decades, due to the gradual maturity of data mining and machine learning technology, a considerable number of scholars have introduced data mining technology into the research of corporate fraud prediction. Green and Choi (1997) first proposed to construct three types of neural networks with a single hidden layer based on eight financial indicators. The study found that the overall error rates of the three neural network models ranged from 37% to 69%. Kotsiantis et al. (2006) proposed up to seven data mining models and forecasted companies with fraudulent financial statements in Greece. The results of the study found that decision trees had the highest accuracy (91%) and support vector machines had the worst performance (73%). Similarly, using the data from Greece, kirkos *et al.* (2007) used decision tree, neural network and Bayesian belief network to construct prediction model, and the study showed that Bayesian belief network was the best and decision tree was the worst.

As for the research in China, Ravisankar *et al.* (2011) took 35 financial indicators as input variables and selected 10 or 18 significant variables with t-test, and then used six data mining models for prediction; on the whole, probabilistic neural networks performed best, and the accuracy could be as high as more than 90%. Song et al.

(2014) constructed four models — backpropagation neural network, logistic regression, decision tree, and support vector machine — using 23 financial indicators. Their results showed that the support vector machine achieved the highest accuracy, while linear logistic regression performed the worst in terms of predictive ability. Similarly, Yeh et al. (2008) developed a financial statement fraud detection model in Taiwan by collecting financial statement variables and corporate governance variables, and applied Bayesian belief networks, support vector machines, and decision tree models. These research findings are consistent with those of Kirkos et al. (2007), who found that the Bayesian belief network had the highest accuracy rate, followed by the support vector machine, while the decision tree performed the worst. Recent research has demonstrated the effectiveness of machine learning techniques in detecting corporate fraud. A study by Xu, Xiong, and An (2022) applied the GONE framework to predict corporate fraud in China, utilizing the Random Forest (RF) model among others. Their findings indicated that Exposure variables played a significant role in fraud prediction, highlighting the importance of incorporating such variables in predictive models.

Based on the above literature, it is evident that with the evolution of data mining and machine learning, new algorithmic technologies have emerged and are being applied to corporate fraud prediction. The random forest method used in this study falls within the broader field of machine learning techniques, which enhances the generalization ability of decision trees. It is also hoped that research on fraud models will lead to broader applications of these models.

DATA AND RESEARCH METHODOLOGY

Data Sources and Selection of Variables

The research object of this paper takes the fraudulent companies and matched normal companies (non-fraudulent) listed in Taiwan as samples, and the data research period is from 1997 to 2015. The samples of fraudulent companies are selected from the acts of false financial statements, manipulation of stock prices, hollowing out and breach of trust reported in newspapers and magazines, as well as the information from the indictments and judgments announced by courts and prosecutors' offices at all levels, and distinguish the fraudulent means used by the company. After excluding the financial industry with special financial reports, a total of 56 fraudulent companies were collected. The matching sample was 112 normal companies with a ratio of "1:2" between fraud and normal companies, and the total number of samples was 168 companies.[3] Referring to Lin, Chang (2009) and Chen *et al.* (2006), the matching sample selection principle was to select two companies with the same industry and similar asset scale as the matching sample of fraud companies one year before the fraud time point of each fraud company. The so-called fraud time point here refers to the exact year of fraud after reading the court indictment and judgment, not the time point reported by the media. Table 1 presents the fraudulent company samples used in this study, highlighting a significant gap between the initial year of fraud and the year of disclosure, with an average gap of 5.4 years. It is more than 5 years after the average company conducts fraud.

As for the fraud means in the last column of Schedule 1, they are also confirmed through media reports, court proceedings or judgments. This study divides the fraud means into four categories: financial fraud, hollowing out, misappropriation of assets and manipulation of stock prices; However, due to the small number of samples of hollowing out, misappropriating assets and manipulating stock prices, these three categories are combined into the same category in the subsequent construction of the model, that is, they belong to the category of non-financial statement fraud. In addition, the methods of some fraud cases used were not only a single method, so there were 39 companies with false financial statements and 78 matching samples, a total of 117 companies; the second category was hollowing out, misappropriating assets and manipulating stock prices. There were 34 companies engaged in this fraud, while the matching sample was 68, making a total of 102. The data sources for this research are the Taiwan Economic Journal Database, the General Accounting Office of the Executive Yuan, the Central Bank, the Knowledge Winner News Search and the Judicial Yuan Global Information Network.

Research Variables

The dependent variable in this paper is a dummy variable of whether the company is fraudulent, in which the fraudulent company is set to 1, and the normal company is set to 0. As for the variables used to predict whether an enterprise is fraudulent, a total of 53 variables are extracted from the four structural variables - financial indicators, corporate governance, market transactions and the overall economy by referring to the above-mentioned relevant literature on enterprise financial anomaly prediction and fraud prediction at home and abroad. The following is a brief description of the study variables.

[3] Lin and Chang (2009) pointed out that in the prior literature, normal non-fraudulent companies were often selected in the ratio of 1:1, but this practice will lead to the problem of over sampling or make the prediction results of the model overly optimistic. Thus, it is suggested to take samples of normal companies in the ratio of 1:2.

(1) Financial indicators: according to the previous literature review, the management level of enterprises will increase the motivation of fraud due to the poor financial performance of the company; therefore, this paper will select multiple financial statement variables to try to capture corporate performance completely, so as to echo the above literature. The financial statement variables of this study adopt the important subjects of the three major statements (balance sheet, income statement and cash flow statement) in the database of Taiwan Economic Journal, and select the variables in the categories of enterprise growth rate, profitability, solvency and operating ability. There are a total of 30 financial statement variables (see Table 1 for detailed variables).

(2) Corporate governance: The most important condition for the Tuntex Group and Rebar Group cases was to obtain absolute control in the past. As long as the dominant control was obtained, the more assets could be misappropriated or embezzled; and the corporate governance mechanism was for the purpose of reducing agency problems and avoiding the possibility of abuse of power as described above. There are many aspects of corporate governance. Among them, the literature most often talks about the impact of the composition of the board of directors on financial crisis or fraud. Therefore, we choose the size of the board of directors, the number of independent directors, the chairman concurrently serving as the general manager and the change status of the chairman. The shares held by directors and supervisors and the shareholding pledge ratio can reflect the management's views and confidence in the sustainable operation of the enterprise. Finally, the proportion of related party sales and the proportion of endorsement and guarantee represent the information transparency of corporate governance, which may also affect the possibility of fraud.

(3) Market transactions: In the method of corporate fraud, companies would use the method of raising the company's stock price to hollow out or embezzle funds, so observing stock price changes would help predict whether the company may be fraudulent (Ye et al., 2005). In the prior literature, market transaction variables also included market value, excess return, standard deviation of return, and turnover rate as variables for predicting financial crisis (Huang et al., 2012). In addition, the more shares in circulation, the less likely the stock price is to be manipulated by interested people. Therefore, this paper also adds the number of shares in circulation as a prediction variable. As for the change of insider's shareholding, it is also a very important predictor variable, so please refer to Table 1 for the nine market transaction variables used in this paper.

(4) Overall economy: Although the importance of overall economic variables had not been mentioned in the prior fraud literature, it is a variable that is often mentioned in the study of corporate financial crisis. Because the overall economic situation may directly affect the company's operating policies, including investment opportunities and capital costs. Therefore, this paper selects six total economic variables, among which the annual growth rate of prosperity leading index is used as the variable to measure the future prosperity. The annual growth rate of money supply and interest rates are indicators for the central bank to observe and control inflation. For enterprises, they will be important variables that affect the evaluation of future investment opportunities and costs. Furthermore, Taiwan's economy depends on foreign trade, and the change of exchange rate will affect the operating performance of export-oriented enterprises. Hence, this paper selects the exchange rates of Taiwan dollar against US dollar, Japanese yen and Euro.

Table 1. Four dimensional study variables used in this paper.

Financial Statements			**Corporate Governance**	**Market Transactions**	**Overall Economy**
1. Short-term investment	11. Interest expense	21. Operating profit rate	1. The chairman concurrently serves as the general manager	1. Closing price	1. One-year fixed deposit rate
2. Current assets	12. Total non-operating income	22. Cash reinvested %	2. Related party sales	2. Annual turnover ratio	2. Change rate of leading indicators
3. Long-term investment	13. Total non-operating expenses	23. Current ratio	3.Endorsement and guarantee	3. Number of outstanding shares	3. M2 annual growth rate
4. Total assets	14. Net profit before interest and tax	24. Quick ratio	4. Number of chairman changes in three years	4. Market value	4. Exchange rate of Taiwan dollar against US dollar
5. Current liabilities	15. Earnings per share (dollar)	25. Interest expense rate	5. Shareholding ratio of directors and supervisors	5. Annual rate of return	5. Exchange rate of Taiwan dollar against Japanese dollar

6. Total liabilities	16. Cash flow from operations	26. Debt ratio	6. Equity pledge rate of directors and supervisors	6. Excess rate of return	6. Exchange rate of Taiwan dollar against Euro
7. Retained surplus	17. Cash flow from investing activities	27. Interest coverage ratio	7. Total number of directors	7. Annualized standard deviation	
8. Total shareholders' equity	18. Cash flow from financing activities	28. Total asset turnover	8. Number of independent directors	8. Increased number of shares held by insiders	
9. Net operating income	19. Return on Total Assets	29. Accounts receivable turnover		9. Decrease in number of shares held by insiders	
10. Business interests	20. Revenue growth rate	30. Inventory turnover rate			

Note: Short-term investment, current assets, long-term investment, current liabilities, total liabilities, retained earnings, total shareholders' equity, net operating income, operating benefits, interest expenses, total non-operating income, total non-operating expenses, net profit before tax , cash flow from operations, cash flow from investing activities, cash flow from financing activities, the above variables are divided by the total assets; the total assets variable is the natural logarithm of the total assets.

Different from other literatures that use univariate analysis of variance or means to test the significance of variables, this study does not screen for the significance of individual variables in advance. The main consideration is that the applied decision tree and random forest algorithm are nonlinear classifiers, and the nonlinear relationship between variables and fraud will be ignored by filtering variables by a linear pre-test method.

Research Methodology

Since it is a typical classification problem to predict whether enterprises will commit fraud, this study attempts to apply two common classification methods in the field of machine learning - decision tree and random forest algorithm to build the prediction model.[4] The decision tree and random forest methods are introduced as follows:

Decision Tree

Decision tree is a kind of supervised learning in machine learning. It is a technical tool for data mining. It classifies complex data and makes decisions in a tree-like form and converts it into a simple and easy-to-interpret representation, so that explanatory variables can predict the target variable. The generation process of decision tree is shown in Figure 2: (1) first, the best variable will be found as the root node among all classification variables (attributes) and start branching. (2) Then, the subsequent branches will be generated from top to bottom and from left to right, which are called child nodes. (3) If the specific growth stop conditions are met, the branch will be stopped and the leaf node will be obtained. (4) Finally, prune the decision tree to get the best decision tree.

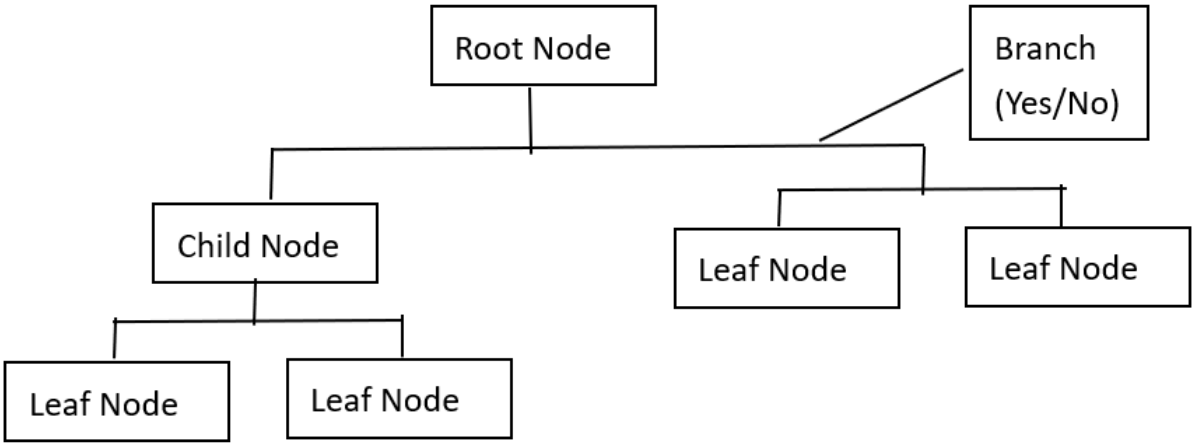

Figure 2. Structure of decision tree.

[4] There are some overlaps between the field of machine learning and the technologies used in data mining, such as decision trees and neural networks. This study does not examine the differences between the two, so decision tree is also regarded as a method of machine learning.

At present, there are three common decision tree algorithms, such as CART, CHAID, and C4.5. Since there is no absolute and optimal version of the three decision tree algorithms, this paper adopts the CART (classification and regression tree) algorithm. Cart is a technology to generate binary tree. Combined with the concepts of classification tree and regression tree, it can analyze continuous or category variables, take Gini index as the basis for selecting branch attributes, and finally use post pruning to complete the process of constructing decision tree. The Gini index is an indicator of impurity. Assuming that the data set S contains n categories, the Gini index is defined as the following formula:

$$\text{Gini(S)} = 1 - \sum_{n \in S} P_i^2 \quad (1)$$

When Gini (S) = 0, it means that the purity is the lowest, which means that all samples belong to the same classification; when Gini (S) is maximized, it means that all class nodes appear with the same probability. Therefore, when selecting node branch variables, CART will select the variables that can reduce the Gini coefficient (reduce impurity) the most among the variables to be selected.

However, the problem of over fitting may occur in the process of decision tree learning, which leads to the high accuracy of decision tree in training sample classification (in sample prediction), but the accuracy of test sample classification (out of sample prediction) is greatly reduced. Therefore, the decision tree must be properly pruned afterwards to improve the generalization ability of the model. The CART algorithm uses the Cost Complexity Pruning method to form a cost complexity function with the leaf nodes of the decision tree and the error rate. When the cost complexity exceeds the threshold, the branch will be pruned.

Random Forest

Random forest is a branch of ensemble learning in machine learning. It adopts the skill of bagging algorithm, which is specifically aimed at the ensemble learning of decision tree algorithm.[5] As its name suggests, the random forest is a collection of many decision trees, and the classification results will be jointly determined by all decision trees, that is, similar to the voting method, based on the modal number. As for how to obtain many decision trees from samples, it is carried out by random sampling of samples and attributes (independent variables). The sampling process is as follows: (1) Assuming that the number of samples in the original training set is N, K samples with the number of samples of N are obtained by random extraction and return; (2) M variables are randomly selected from M input variables, where m is less than M. According to this, K decision trees can be obtained. The input variables of each decision tree are different, so K classification results can be obtained. Finally, the modal number is taken as the final classification result. Because the random sampling of random forest ensures the randomness, there will be no over adaptation problem, and it is less sensitive to multivariate collinearity. So it can be used to deal with the decision tree with insufficient identification or over adaptation of original data. Regarding the above two parameters of m and K, in this study, K is fixed to 500, that is, there are 500 trees in the forest, and m is the parameter with the lowest out-of-bag error (OOB error) selected from 5~30.

In addition, the importance of input variables can be further obtained through the classification accuracy of individual decision trees. Therefore, this study will make two models in the training of random forest. The first is the practice of traditional random forest, which considers the overall learning of all original input variables; the second is a two-stage random forest. Through the first model, we can know the importance of all variables. The importance part is sorted according to the reduction of Gini coefficient of each variable. Then in the second stage, only the first 15 variables with high importance are used for overall learning, and the m parameter of the second stage will select the one with the lowest out of bag error rate among 3 ~ 10 (K is still set to 500).

Evaluating Predictive Models

In order to evaluate the feasibility of the model, when constructing the model, this paper will divide the samples into a training group and a test group to evaluate the prediction accuracy, type I error and type II error of the two groups respectively. The accuracy rate was the proportion in which both the fraudulent company and the normal company were classified correctly, the type I error was the proportion in which the normal company was classified as the fraudulent company, and the type II error was the proportion in which the fraudulent company was classified as the normal company (Green and Choi, 1997; Song *et al*., 2014). Type I and II errors were of practical significance, in which type I error represented that financial supervisors or auditors would spend unnecessary time and cost; Type II error represented the failure to detect fraudulent companies, which might bring considerable costs to the whole society (Kirkos *et al*., 2007; Yeh *et al*., 2008).

Traditionally, data mining or machine learning will use (1) Random sampling to select the training group and the test group (i.e. holdout sampling), and substitute the test group into the model evaluation results; or (2) Use the full sample for *k*-fold cross validation). However, this study believes that the above methods are not applicable to the model verification of fraud prediction, because these methods must ensure that the fraud characteristics will

[5] Bagging is the abbreviation of bootstrap aggregating, which was proposed by Breiman (1996).

not change structurally over time, but in practice, the fraud means of enterprises are likely to derive new criminal methods over time. Therefore, this study adopts the ranking according to the year of fraud, takes the samples that occurred before 2007 as the training group and those that occurred later as the test group, and the number of samples in the two groups is about 4:1. Cutting samples by fraud time has the following advantages: First, if the fraud characteristic means does not produce structural changes over time, the prediction effect of dividing samples by fraud time will be little different from random sampling or *k*-fold cross validation. Second, if the fraud characteristics means will change structurally with time, random sampling or *k*-fold cross validation will overestimate the prediction effect of the model, and cutting samples according to the fraud time will be closer to the reality.

EMPIRICAL RESULTS

The empirical evidence of this paper is divided into three subsections. First, all samples do not distinguish fraudulent means, and then they are divided into samples with false financial reports and hollowing out / misappropriating assets / manipulating stock prices for model construction and evaluation.

The Whole Sample Does Not Discriminate Fraudulent Means

Figure 3 shows the decision tree constructed by the training group using the non-discrimination method. The root node of the first layer in the figure is the Return on Total Assets (ROA). The lower the ROA, the more likely it is to be classified as a fraudulent company; since ROA is a basic indicator for judging the operation of an enterprise, the decision tree regards it as the root node. That is to say, it shows that the business performance is too poor, and the management may start to commit fraud in the next year. This result verifies that Loebecke *et al.* (1989) proposed that the management's motivation for fraud increased due to poor financial performance. The second level sub nodes are cash flow and stock turnover ratio of financing activities. The more frequent financing activities and the higher stock turnover rate are, the more likely they are to be a sign of corporate fraud. The third and fourth sub-nodes are current liabilities and retained income, respectively. The higher the current liabilities and the lower the retained income, the more likely it is to be classified as a fraudulent company. The above branch variables are mainly from the financial statements, which is also consistent with the findings of the prior literature. Another market trading variable is included, which shows that the observation of stock trading has its discrimination ability.

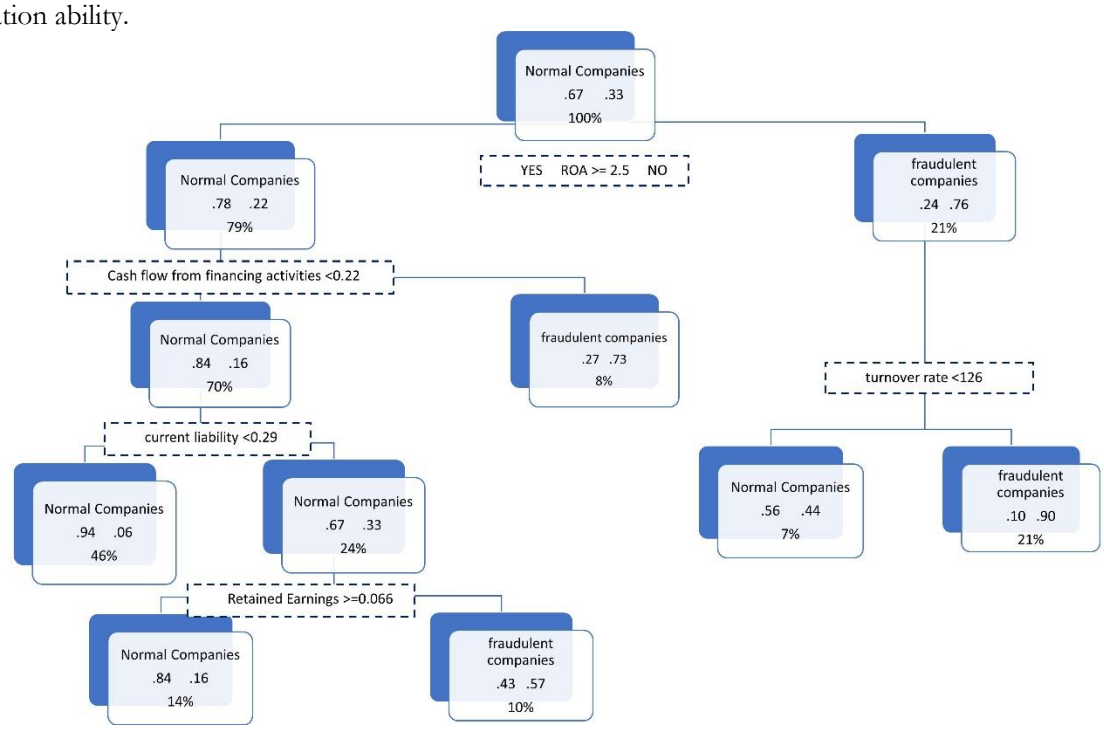

Figure 3. Decision trees of all samples.

The random forest contains 500 decision trees, and the generation combinations of each decision tree are different, so it is impossible to draw a single structure as shown in Figure 3 above; but we can judge the importance of variables from the accuracy of each decision tree. In the process of two-stage random forest training, the most important 15 variables are: return on total assets, retained income, operating interests, cash flow from financing activities, cash flow from operation, total non-operating income, net profit before interest before tax, interest cover

multiple, cash flow from investment activities, current liabilities, shareholding of directors and supervisors, turnover rate, long-term investment, related party sales and annual remuneration are regarded as input variables in the second stage. Table 2 is sorting out the prediction ability of decision tree, random forest and two-stage random forest model. In the training group, the classification accuracy rate of decision tree is 83.7%, and the accuracy rate of random forest and two-stage random forest is 100%, showing the strong learning ability of random forest. However, in the test group, the accuracy rate of the three models is 69.7%, and the type II error is very high. This result indicates that the model established by using the samples before 2007 cannot successfully predict the fraud cases after 2007, which also means that the fraud characteristics and means have structural changes over time.

Table 2. The prediction accuracy of the full sample indiscriminate method.

Model	Training group			Test group		
	Accuracy rate	Type I error	Type II error	Accuracy rate	Type I error	Type II error
Decision tree	83.7%	12.2%	24.4%	69.7%	4.5%	81.8%
Random forest	100.0%	0%	0%	69.7%	0%	90.9%
Two-stage random forest	100.0%	0%	0%	69.7%	4.5%	81.8%

Note: Type I error means that normal companies are wrongly classified as fraudulent companies, and Type II error means that fraudulent companies are wrongly classified as normal companies. The number of samples in the training group is 135 (45 fraudulent companies), and the number of samples in the test group is 33 (11 fraudulent companies), and the year of fraud 2007 is used as the cutting condition.

Fraud Means is False Financial Statements

Figure 4 shows the decision tree constructed by the training group using the false financial statements method. The first root node in the figure is retained income. In the observation data, a large proportion of the retained income of fraudulent companies show a negative number, which means that these companies want to beautify their financial statements and business performance by means of financial statement fraud due to long-term losses. The second tier sub nodes are the cash flow of financing activities and the number of outstanding shares. It can be imagined that companies with high retained income generally have accumulated income, and the demand for financing should be low. Therefore, enterprises with high retained income and high financing demand may be fraudulent companies. In addition, companies with accumulated losses and a low number of outstanding shares are more likely to be manipulated by interested parties, resulting in fraudulent incidents.

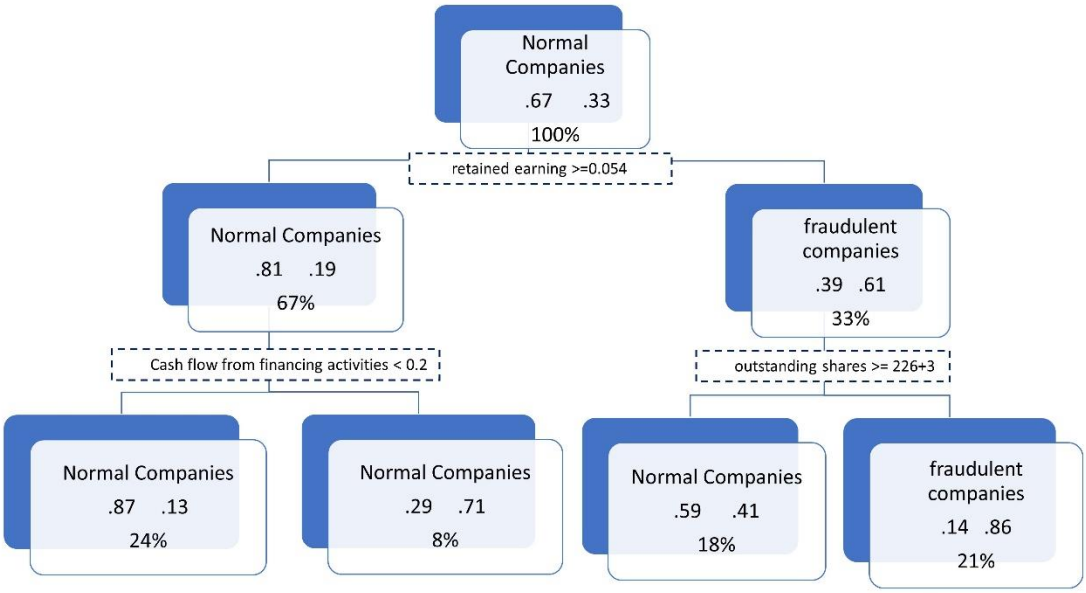

Figure 4. The decision tree of fraudulent means of financial fraud.

In the training samples with false financial reports, the random forest algorithm found that among the 53 variables, the 15 most important variables were cash flow from operations, retained income, shares held by directors and supervisors, total liabilities, and operating profits, current liabilities, cash flow from financing

activities, cash reinvestment, debt ratio, excess return, quick ratio, total shareholders' equity, interest coverage ratio, net profit before tax, and total non-operating income. The above-mentioned variables mainly come from financial statement technology, and there are also variables from corporate governance and market transactions, such as shareholding by directors and supervisors, excess remuneration, etc. Table 3 is sorting out the prediction ability of decision tree, random forest and two-stage random forest model. In the training group, the classification accuracy rate of decision tree is 80.6%, and the accuracy rate of random forest and two-stage random forest is 100%. This is consistent with the results in section 4.1. As for the test group, the accuracy of decision tree and two-stage random forest is 79.2%, and that of random forest is 66.7%. The random forest model representing unfiltered variables is affected by variables without discrimination ability. It is worth mentioning that the Type II error between the decision tree and the two-stage random forest is reduced to 62.5%, and the Type I error is 0. It shows that although more than half of the fraud cases after 2007 cannot be predicted one year in advance, once the model determines that the company is fraudulent, it will start to engage in fraud activities in the next year.

Table 3. The prediction accuracy of fraudulent financial reporting methods.

Model	Training group			Test group		
	Accuracy rate	Type I error	Type II error	Accuracy rate	Type I error	Type II error
Decision tree	80.6%	6.5%	45.2%	79.2%	0%	62.5%
Random forest	100.0%	0%	0%	66.7%	0%	100.0%
Two-stage random forest	100.0%	0%	0%	79.2%	0%	62.5%

Note: Type I error refers to a non-fraudulent company being incorrectly classified as a fraudulent one, while Type II error refers to a fraudulent company being incorrectly classified as a non-fraudulent one. The training set consists of 93 companies (31 fraudulent), and the testing set consists of 24 companies (8 fraudulent), with the year 2007 used as the cutoff for identifying the year of fraud.

Fraudulent Means Include Hollowing Out, Misappropriating Assets and Manipulating Stock Prices

Figure 5 shows the decision tree constructed by the training group using non-financial practices such as hollowing out, misappropriating assets and manipulating stock prices. In the figure, the root node of the first layer is ROA, which is the same as the non-discrimination method, but the threshold value of the branch is reduced to 1.5% (2.5% for the non-discrimination method). The sub-nodes of second and third layers are current liabilities and short-term investment. The decision tree will ROA > 1.5%, current liabilities > 25% of total assets and short-term investment < 0.42% of total assets. Finally, the sub-nodes of fourth layer is the endorsement guarantee. Interestingly, if the endorsement guarantee amount/net value ratio is less than 2%, it will be classified as a fraudulent company. Intuitively, it may be thought that the higher the endorsement guarantee, the more likely it is to hollow out and embezzle assets through subsidiaries. However, the classification criterion here is the result of a four-layer decision tree branch, and a nonlinear classification method appears.

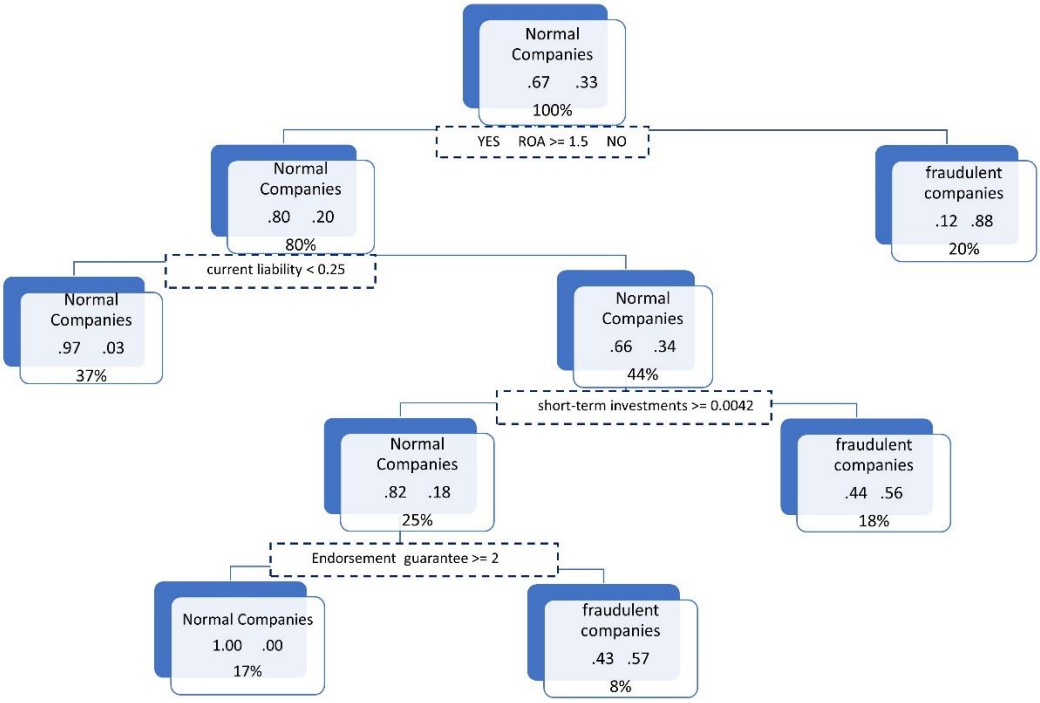

Figure 5. The decision trees of fraudulent means are hollowing out, misappropriating assets and manipulating stock prices.

The first-stage random forest calculus is conducted for the training samples that are not financially reported by fraudulent means, and the 15 most important variables are ROA, retained income, operating profits, pre-tax net profit, operating profit rate, interest guarantee multiple, current liabilities, shares held by directors and supervisors, interest expense ratio, total non-operating income, shares pledged by directors and supervisors, long-term investment, quick ratio, total shareholders' equity, short-term investment, and then conducts a second training with these 15 variables. Table 4 is sorting out the prediction ability of decision tree, random forest and two-stage random forest model. In the training group, the classification accuracy rate of decision tree is 85.1%, and the accuracy rate of random forest and two-stage random forest is 100%. This is consistent with the previous results. However, the prediction accuracy of the decision tree in the test group has been greatly reduced to 53.3%, which can be explained that the single decision tree established by the training group before 2007 is not suitable for subsequent cases, which is why it is necessary to use the random forest algorithm of overall learning. Among the three models, the result of two-stage random forest is the most reliable, and there is still 80% accuracy in the test group sample; although the error of Type II is still 60%, the error of Type I is 0%, that is, the companies determined to be fraudulent by the model will start to commit fraud in the next year.

Table 4. The prediction accuracy of fraudulent schemes such as asset misappropriation, embezzlement, and stock price manipulation.

Model	Training group			Test group		
	Accuracy rate	Type I error	Type II error	Accuracy rate	Type I error	Type II error
Decision tree	85.1%	20.7%	3.4%	53.3%	40.0%	60.0%
Random forest	100.0%	0%	0%	73.3%	0%	80.0%
Two-stage random forest	100.0%	0%	0%	80.0%	0%	60.0%

Note: Type I error refers to a non-fraudulent company being incorrectly classified as a fraudulent one, while Type II error refers to a fraudulent company being incorrectly classified as a non-fraudulent one. The training set consists of 87 companies (29 fraudulent), and the testing set consists of 15 companies (5 fraudulent), with the year 2007 used as the cutoff point based on the year of fraud.

Summary

At the same time, observing the prediction results in Tables 2, 3 and 4, this study finds the following conclusions: First, among the three models, the prediction ability of two-stage random forest is the best and stable, which shows that the overall learning technology is helpful to improve the classification effect and will not be limited by the results of a single decision tree. Second, the second-stage random forest is to select the variables

with high importance for the second training. However, all the overall economic variables have not been selected, which means that compared with other three-dimensional variables, the overall economy will not be the decisive factor for whether an enterprise commits fraud. This is quite different from Huang Ruiqing et al. (1912)'s prediction of financial crisis, because corporate fraud is more likely to be human harm, and financial crisis may be caused by the overall business environment. Third, after distinguishing the fraudulent means, the prediction performance of the two-stage random forest is better, including the improvement of the accuracy rate, the substantial reduction of the Type I error and the Type II error, which means that different fraudulent means have unique structures, and different fraudulent means should be considered to establish different model.

Fourth, no matter what kind of model or whether to distinguish fraud, the error of Type II is more than 50%. As mentioned above, the high Type II error reflects the great difference between the training group and the test group, that is, the fraud mentioned in this paper will appear new methods over time, resulting in structural changes and cannot be captured. Another explanation is that this study is to collect the data of the year before the beginning of fraud, so in fact, there was no fraud at all in the previous year, and there was only "mens rea" at most. Therefore, it is also possible to have a high Type II error. Fifthly, after distinguishing fraud means, the Type I errors of the two-stage random forest model are all 0, that is, no normal company is misjudged as a fraudulent company; on the other hand, it also means that those companies judged by the model as fraudulent companies must be fraudulent companies in the future, which is of great help to the competent authority of financial supervision in practice, because they can directly supervise the companies predicted to be fraudulent, which can save a considerable amount of resource misplacement cost in supervision.

CONCLUSIONS AND SUGGESTIONS

In the course of the development of Taiwan's capital market, corporate fraud broke out from time to time, involving hundreds of millions to tens of billions of improper interests and damaged the rights and interests of all stakeholders. Although the financial supervision system has been gradually completed in recent years, it is still unable to completely eliminate the occurrence of enterprise fraud; therefore, it is of practical and academic importance to establish an effective enterprise fraud early warning or prediction model. Thus, this study also attempts to establish a model to predict whether enterprises will commit fraud in the future.

In this study, two algorithms in the field of machine learning - decision tree and random forest are used as prediction models. A total of 53 prediction variables are selected from four aspects, including financial statements, corporate governance, market transactions and overall economy, so that the algorithm can learn how to correctly classify fraud and normal companies. Due to the diversity of fraud means, in order to enhance the prediction ability of the model, this paper also divides the fraud samples into two categories: false financial statements and hollowing out / misappropriating assets / manipulating stock price (non-financial reporting fraud). In addition, the criteria for evaluating the prediction ability of the model are different from the previous literature. This paper is divided into training group and testing group around the year of fraud in 2007. The purpose is to examine whether the detailed means of fraud events have structural changes with the evolution of time. The empirical results show that the prediction result of two-stage random forest is the best after the fraud is divided into financial fraud and non-financial fraud. However, its type II error is high in different methods, which shows that fraud does have structural changes, and it is difficult to capture new fraud samples with the past classification fraud criteria. However, it is worth mentioning that the two-stage random forest has 0% Type I error in different fraud methods, which implies that if the model determines that the company is a fraud company, the company will start to commit fraud in one year. This conclusion is of great help to the competent authority of financial supervision: in practice, the competent authority can focus on the company judged to be fraudulent, because no normal company will be misjudged.

From the prediction results, the random forest is better than the decision tree, representing the overall learning, which is helpful to improve the prediction ability of a single decision tree. However, the random forest classification is the voting result of many trees, and lacks the clear and easy-to-understand structure of the original decision tree "If-Then". In addition, there is another technology called − boosting for holistic learning. In the calculation process, the weight of samples with classification errors will be increased, that is, learning from errors will be strengthened. According to the decision tree with false finance, the consistency of medium-sized Type II error is high no matter in training or test data. Therefore, it is suggested to use Gradient Boosting Decision Tree (GBDT) algorithm for fraud samples with false finance in the future to strengthen the samples of learning Type II error, that is, fraud companies that cannot be predicted in advance, it should make the overall accuracy or Type II error perform better. Finally, this study selects a total of 53 variables from four dimensions for model construction, but these variables belong to publicly available structured data. In the future, it is also suggested to add more variables "not from the database". For example, the literature on behavioral finance and corporate governance mentions the personality and attitude of the chairman or CEO, such as arrogance, overconfidence, etc. Because the root cause of the fraud is still from the internal and external pressure and incentives of the

fraudster, the company's operating performance is an external factor, while the personality characteristics of the chairman or CEO can be regarded as an internal factor; thus, perhaps increasing the variables related to people can also get more meaningful results.

Declarations
All authors declare that they have no conflicts of interest.

REFERENCES

ACFE, "Report to the Nation on Occupational Fraud and Abuse," Association of Certified Fraud Examiners, 2010.

Altman, E. I., "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy," *Journal of Finance*, Vol. 23, No. 4, 1968, pp. 589-609.

Beasley, M. S., "An Empirical Analysis of the Relation between the Board of Director Composition and Financial Statement Fraud," *Accounting Review*, Vol. 71, No. 4, 1996, pp. 443-465.

Breiman, L., "Bagging Predictors," *Machine learning*, Vol. 24, No. 2, 1996, pp. 123-140.

Campbell, J. Y., Hilscher, J., and Szilagyi, J., "In Search of Distress Risk," *Journal of Finance*, Vol. 63, No. 6, 2008, pp. 2899-2939.

Chen, G., Firth, M., Gao, D. N., and Rui, O. M., "Ownership Structure, Corporate Governance, and Fraud: Evidence from China," *Journal of Corporate Finance*, Vol. 12, No. 3, 2006, pp. 424-448.

Chen, Y.-N., Wang, Y.-Z., and Hsu, H.-Y., "Predicting the Financial Crisis of Taiwanese Enterprises: Which Is Better between Credit Scoring Method and Option Evaluation Method?" *Journal of Risk Management*, Vol. 6, No. 2, 2004, pp. 155-179.

Duffie, D., Saita, L., and Wang, K., "Multi-Period Corporate Default Prediction with Stochastic Covariates," *Journal of Financial Economics*, Vol. 83, No. 3, 2007, pp. 635-665.

Green, B. P. and Choi, J. H., "Assessing the Risk of Management Fraud through Neural Network Technology," *Auditing: A Journal of Practice & Theory*, Vol. 16, No. 1, 1997, pp. 14-28.

Hackenbrack, K., "The Effect of Experience with Different Sized Clients on Auditor Evaluations of Fraudulent Financial Reporting Indicators," *Auditing: A Journal of Practice & Theory*, 12(1), 1993, pp. 99-110.

Hsu, H.-N., Ouyang, H., and Chen, Q.-F., "Construction of Corporate Governance, Earnings Management and Financial Early Warning Model," *Accounting and Corporate Governanc*e, Vol. 4, No. 1, 2007, pp. 84-121.

Huang, R.-Q., Wu, C.-S., Lin, J.-L., and Hiao, C.-X., "An Empirical Study on Financial Crisis Factors of Taiwanese Enterprises," *Taiwan Financial Quarterly*, Vol. 13, No. 4, 2012, pp. 55-76.

Johnson, S. A., Ryan, H. E., and Tian, Y. S., "Managerial Incentives and Corporate Fraud: The Sources of Incentives Matter," *Review of Finance*, Vol. 13, No. 1, 2009, pp. 115-145.

Kinney, W. R. and McDaniel, L. S., "Characteristics of Firms Correcting Previously Reported Quarterly Earnings," *Journal of Accounting and Economics*, Vol, 11, No. 1, 1989, pp. 71-93.

Kirkos, E., Spathis, C., and Manolopoulos, Y., "Data Mining Techniques for the Detection of Fraudulent Financial Statements," *Expert Systems with Applications*, Vol. 32, No. 4, 2007, pp. 995-1003.

Kotsiantis, S., Koumanakos, E., Tzelepis, D., and Tampakas, V., "Forecasting Fraudulent Financial Statements Using Data Mining," *International Journal of Computational Intelligence*, Vol. 3, No. 2, 2006, pp. 104-110.

KPMG, "Global Profiles of the Fraudster," KPMG International, 2016.

Lin, C.-J., and Chang, C.-J., "The Abnormal Change of Directors and Supervisors, the Correlation between Family Businesses and Corporate Fraud," *Accounting Review*, No. 48, 2009, pp. 1-33.

Liu, X., Xiao, Y., & Li, Y. (2022). Using Machine Learning to Predict Corporate Fraud: Evidence Based on the GONE Framework. *Journal of Business Ethics*, 178(3), 705-726.

Loebbecke, J. K., Eining, M. M., and Willingham, J. J., "Auditors' Experience with Material Irregularities: Frequency, Nature, and Delectability," *Auditing: A Journal of Practice & Theory*, Vol. 9, No. 1, 1989, pp. 1-28.

Ohlson, J. M., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, Vol. 18, No. 1, 1980, pp. 109-131.

Ravisankar, P., Ravi, V., Rao, G. R., and Bose, I., "Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques," *Decision Support Systems*, Vol. 50, No. 2, 2011, pp. 491-500.

Shumway, T., "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *Journal of Business*, Vol. 74, No. 1, 2001, pp. 101-124.

Song, X. P., Hu, Z. H., Du, J. G., and Sheng, Z. H., "Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China," *Journal of Forecasting*, Vol. 33, No. 8, 2014, pp. 611-626.

Spathis, C. T., "Detecting False Financial Statements Using Published Data: Some Evidence from Greece," *Managerial Auditing Journal*, Vol. 17, No. 4, 2002, pp. 179-191.

Stice, J. D., "Using Financial and Market Information to Identify Pre-Engagement Factors Associated with Lawsuits against Auditors," *Accounting Review*, Vol. 66, No. 3, 1991, pp. 516-533.

Summers, S. L. and Sweeney, J. T., "Fraudulently Misstated Financial Statements and Insider Trading: An Empirical Analysis," *Accounting Review*, Vol. 73, No. 1, 1998, pp. 131-146.

Uzun, H., Szewczyk, S. H., and Varma, R., "Board Composition and Corporate Faud," *Financial Analysts Journal*, Vol. 60, No. 3, 2004, pp. 33-43.

Xu, X., Xiong, F., and An, Z., "Using Machine Learning to Predict Corporate Fraud: Evidence Based on the GONE Framework," Journal of Business Ethics, Vol. 186, No. 1, 2023, pp. 137-158.

Yeh, C.-C., Chi, D.-J., and Lin, S.-J., "A Study for Detecting Enterprise Financial Statement Fraud," Asian Journal of Management and Humanity Sciences, Vol. 3, No. 1-4, 2008, pp. 15-30.

Ye, J.-H., Lin, Y.-Z., and Chu, S.-Y., "Taiwan's Experience and Early Warning of Stock Price Manipulation of Raising and Swindling," *Economic Papers*, Vol. 43, No. 4, 2015, pp. 589-638.

