DECISION INFORMATION FOR AUDITORS TO ASSESS LITIGATION RISK: APPLICATION OF MACHINE LEARNING TECHNIQUES

Yu-Hsin Lu *, Yu-Cheng Lin **, Fang-Ci Gu ***

* *Corresponding author*, Department of Accounting, Feng Chia University, Taiwan
Contact details: Feng Chia University, No. 100, Wenhua Rd., Xitun Dist., Taichung City 40724, Taiwan
** Department of Accounting, National Yunlin University of Science & Technology, Taiwan
*** Department of Accounting, Feng Chia University, Taiwan

OPEN ACCESS

How to cite this paper: Lu, Y.-H., Lin, Y.-C., & Gu, F.-C. (2022). Decision information for auditors to assess litigation risk: Application of machine learning techniques. *Corporate Ownership & Control, 19*(3), 133–146. https://doi.org/10.22495/cocv19i3art10

ISSN Online: 1810-3057
ISSN Print: 1727-9232

Received: 15.03.2022
Accepted: 06.05.2022

JEL Classification: M4, C8, M1
DOI: 10.22495/cocv19i3art10

Abstract

Fraud cases have become more common in recent years, highlighting the role of auditors' legal liability. The competent authorities have called for stricter control and disciplinary measures for auditors, increasing auditors' legal liability and litigation risk. This study used machine learning (ML) techniques to construct a litigation warning model for auditors to assess audit risk when they evaluate whether accept or terminate an engagement, thus improving audit quality and preventing losses due to litigation. Otherwise, a sample matching method comprised of 64 litigated companies and 128 non-litigated companies was used in this study. First, feature selection technology was used to extract six important influencing factors among the many variables affecting auditors' litigation risk. Then a decision tree was used to establish a litigation warning model and a decision table for auditors' reference. The results indicated that the eight outcomes provided by the decision table could effectively distinguish the level of a litigation risk with an accuracy rate of 92.708%. These results can provide useful information to aid auditors in assessing engagement decisions.

Keywords: Litigation Risk, Machine Learning, Feature Selection, Classifier Technique

Authors' individual contribution: Conceptualization — Y.-H.L.; Methodology — Y.-H.L.; Formal Analysis — Y.-C.L.; Resources — F.-C.G.; Writing — Y.-H.L. and Y.-C.L.; Supervision — Y.-H.L.

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

1. INTRODUCTION

Recently, serious fraud cases around the world caused massive losses for investors and creditors and shook public confidence in the capital market. The integrity of management, as well as the profession and ethics of auditors, were also called into question. In 2004, Taiwan saw a quick succession of fraud cases involving Procomp Informatics, Infodisc Technology, and Summit Computer Technology which led the Financial Supervisory Commission (FSC) of Taiwan to issue warnings regarding or cancel the certification of auditors. Otherwise, according to deep pocket theory (Calabresi, 1970), when a company is charged with fraud, creditors and investors often pursue litigation against well-paid certified public accountants (CPAs), despite a lack of audit failure, to obtain more compensation for losses (Carcello & Palmrose, 1994). Fraud cases not only directly put auditors at risk from litigation or sanction but also come with large legal costs and can cause substantial harm to reputations (Bonner, Palmrose, & Young, 1998).

In Taiwan, auditor litigation was less and investors couldn't confront large companies or auditors alone since huge litigation expenses before the Enron. In 2002, the Securities Investor and Futures Trader Protection Act was announced by the government, and Securities and Futures Investors Protection Center was established at the same time. Recently, the Securities and Futures Investors Protection Center has helped investors to sue many illegal companies and their auditors. A stricter legal environment lets large audit firms have begun to emphasize client screening and risk management. Recently, Deloitte Taiwan established a Reputation and Risk Department to assess clients' industry status and level of risk; KPMG Taiwan established an independent assessment team and a client risk screening team to determine whether new clients should be accepted; Ernst & Young (EY) Taiwan also established a risk management committee to investigate clients (Liu, Wang, & Lai, 2009). However, this decision-making process is extremely complex as underestimating the risk may lead to future litigation and damage reputations. Therefore, investigating the factors influencing litigation against auditors and providing auditors with risk evaluation information is important for both practice and academics. Particularly, developing a user-friendly litigation warning model that can be used in everyday auditing is the most crucial for auditors. In the early phases of audit work, machine learning (ML) enables auditors to access unbiased and more accurate information by collecting data using rules developed with machine learning algorithms (Cho, Vasarhelyi, Sun, & Zhang, 2020).

Machine learning techniques, such as decision trees and artificial neural networks (NN), are superior to traditional statistical methods, such as logistic regression (LR) and discriminant analysis in constructing detection models (Varetto, 1998; Cristianini & Shawe-Taylor, 2000; Min & Lee, 2005). Mitchell (as cited in Cho et al., 2020, p. 1) provided a widely referenced definition of machine learning: "The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience". The main contribution of machine learning and the primary difference between machine learning and other algorithms is its predicting power, which arises from the processes of training and testing datasets (Cho et al., 2020). A common strategy used is to discover a pattern in a training dataset. This pattern is then used to classify and/or predict the behavior of new samples (Bose & Mahapatra, 2001; Burez & Van den Poel, 2007; Cho et al., 2020). This study used these techniques to extract important variables from factors influencing auditors' litigation risk; then, an accurate warning model was created for auditors to assess risk to aid in evaluating whether auditors should accept or terminate engagements.

This study contributes to the literature in the following aspects. First, different from prior litigation risk research (Casterella, Jensen, & Knechel, 2010; Boone, Khurana, & Raman, 2011; Schmidt, 2012; Kaplan & Williams, 2013) that examines the associations between auditors' litigation and audit firm characteristics or abnormal accruals, this study employed the feature selection approach to extract six critical factors to create our litigation warning model. As compared to audit firm characteristics and abnormal accruals, the credit risk index is the most important factor influencing the litigation risk for auditors in Taiwan and serves as a reference for other developing countries. Second, the prediction performance of our machine learning model is superior to that of the logistic regression and the discriminant analysis. This result underscores the value of applying ML to assist auditors in assessing their litigation risk. Finally, while extant litigation warning models focus on improving prediction accuracy, our study focused on constructing a classification model and a decision table that included eight classification rules from which the auditors can better assess litigation likelihood. The auditors can use this model to screen out potentially risky clients and decide which audit engagements can be accepted.

Section 2 of this study reviews the literature on the litigation risk against auditors and data mining techniques. Section 3 describes the steps taken to construct the warning model, variable measurements, sample selection, and sources of data. Section 4 summarizes the results and analysis, and Section 5 provides a conclusion and suggestions.

2. LITERATURE REVIEW

2.1. Affecting factors of auditors' litigation risk

Assessing auditors' litigation risk is a complex procedure with many affecting factors. Arens, Elder, and Beasley (2014) indicated that engagement risk analysis can provide a framework. Engagement risk is the risk that the auditor or audit firm will suffer harm after the audit is finished, even though the audit report was correct. For example, if a client declares bankruptcy after an audit is complete, the likelihood of a lawsuit against the CPA firm is reasonably high. When auditors modify audit evidence for engagement risk, it is done by control of acceptable audit risk. Acceptable audit risk is a measure of how willing the auditor is to accept that the financial statements may be materially misstated after the audit is completed and an unmodified opinion has been issued. According to Statement on Auditing Standards of Taiwan (hereafter, SAS of Taiwan) No. 51, audit risk is affected by inherent risk, control risk, and detection risk. Otherwise, not only audit risk but also the deep pocket theory and the stricter legal environment let the litigation risk against auditors increase. Therefore, this study explored the affecting factors of litigation against auditors from prior literature and classified them into the following types: *inherent risk*, *control risk*, *detection risk in audit risk*, and *legal environment*. All of the factors and related literature are shown in Table 1.

Table 1. Affecting factors of auditors' litigation risk

	Variable	*Reference*
Audit risk related: Inherent risk		
1	Asset scale	Kaplan and Williams (2013); Schmidt (2012); Boone et al. (2011); Palmrose and Scholz (2004); Pierre and Anderson (1984)
2	Period listed on stock exchange	Schmidt (2012); Bonner et al. (1998); Palmrose (1988); Pierre and Anderson (1984)
3	Credit risk index	Kaplan and Williams (2013); Boone et al. (2011); Stice (1991); Palmrose (1987); Pierre and Anderson (1984)
4	Debt ratio	Boone et al. (2011)
5	Accounts receivable ratio	Boone et al. (2011); Lys and Watts (1994)
6	Accounts receivable turnover rate	Francis and Krishnan (1999); Stice (1991)
7	Inventory ratio	Schmidt (2012); Boone et al. (2011)
8	Inventory turnover rate	Francis and Krishnan (1999); Stice (1991)
9	Inventory growth rate	Summers and Sweeny (1998)
10	Sales growth rate	Kaplan and Williams (2013); Schmidt (2012); Boone et al. (2011); Summers and Sweeny (1998)
11	Operating profit ratio	Stice (1991)
12	Operating profit growth rate	Stice (1991)
13	Return on operating assets	Kaplan and Williams (2013); Boone et al. (2011); Summers and Sweeny (1998)
14	Annual return on stock	Kaplan and Williams (2013); Boone et al. (2011); Bonner et al. (1998); Lys and Watts (1994)
15	Stock price fluctuation	Boone et al. (2011); Cahan and Zhang (2006)
16	Annual loss	Carcello and Palmrose (1994); McKeown, Mutchler, and Hopwood (1991)
17	Heavy losses	McKeown et al. (1991)
18	Working capital	McKeown et al. (1991)
19	Insufficient working cash flow	Kaplan and Williams (2013)
20	Type of industry	Schmidt (2012); Boone et al. (2011); Cahan and Zhang (2006); Bonner et al. (1998).
Audit risk related: Control risk		
21	Chairperson is an executive director	Kaplan and Williams (2013)
22	Change of chairperson	Casterella et al. (2010)
23	Ratio of stock owned by directors	Kaplan and Williams (2013)
24	Ratio of independent board members	Beasley (1996); Fama and Jensen (1983)
25	Ratio of stock owned by legal entities	Kaplan and Williams (2013)
26	Change of auditor	Schmidt (2012); Knechel, Naiker, and Pacheco (2007); Schwartz and Soo (1996); Titman and Trueman (1986)
27	Illegal behavior	Casterella et al. (2010); Bonner et al. (1998); Pierre and Anderson (1984)
28	Restatement of financial reports	Schmidt (2012); Palmrose and Scholz (2004); Lev, Ryan, and Wu (2008)
Audit risk related: Detection risk		
29	Large audit firm	Schmidt (2012); Kim, Chung, and Firth (2003); Krishnan and Krishnan (1997); Palmrose (1988)
30	Industry expert_CPA	Schmidt (2012); Beasley and Petroni (2001); Gramling and Stone (2001); Solomon, Shields, and Whittington (1999)
31	Client importance_CPA	Schmidt (2012)
32	Client importance_firm	Schmidt (2012)
33	Non-audit fees	Schmidt (2012); Cahan and Zhang (2006)
34	Audit report lag	Kaplan and Williams (2013); Blacconiere and DeFond (1997)
35	Audit firm tenure	Lys and Watts (1994)
36	CPA tenure	Kaplan and Williams (2013); Boone et al. (2011); Lys and Watts (1994); Stice (1991)
37	Auditor industry experience	Lys and Watts (1994)
38	Audit opinion type	Kaplan and Williams (2013); Blacconiere and DeFond (1997); Carcello and Palmrose (1994); Lys and Watts (1994); McKeown et al. (1991)
Legal environment		
39	China Rebar case	Lin and Lin (2010)

2.1.1. Affecting factors of litigation risk: Audit risk related

Inherent risk is the possibility of material misstatement in financial statements before considering the effectiveness of the client's related internal controls. Changes in the industry or complex transactions may influence the inherent risk of a company. Therefore, when an auditor assesses inherent risk, the company's organizational situation, financial situation, nature of services, type of transactions, fraud-related assets and liabilities, and socio-economic environment are often carefully considered (McKeown et al., 1991; Stice, 1991; Bonner et al., 1998; Palmrose & Scholz, 2004; Schmidt, 2012; Kaplan & Williams, 2013; Perols, Bowen, Zimmermann, & Samba, 2017).

Control risk is the possibility that financial statements have material misstatements that cannot be prevented or detected by the client's internal controls. SAS No. 99, SAS No. 110, and SAS of Taiwan No. 43 state that the management of clients must establish internal controls and maintain effective execution of policies and procedures to achieve internal control objectives. The governance level of the client should monitor the management to ensure the establishment and maintenance of internal control, reliable financial statements, efficient and effective operation, and compliance with relevant ordinances. The actualization of corporate governance monitors business performance prevents conflicts of interest and ensures compliance with various ordinances. Therefore, much prior literature indicated significant correlations between the quality of corporate governance and the prevention of fraud which, in turn, influences the litigation risk against auditors (Palmrose & Scholz, 2004; Schmidt, 2012; Kaplan &

Williams, 2013; Lev et al., 2008; Beasley, 1996; Fama & Jensen, 1983).

Detection risk is the risk that the audit evidence collected is unable to detect materiality misstatements. This risk is correlated to the auditor's or audit firmєs audit quality and characteristics. DeAngelo (1981) showed that the quality of audit services is defined to be the market-assessed joint probability that a given auditor will both 1) discover a breach in the client's accounting system, and 2) report the breach. The discovery of a breach is dependent on the professional competency of the auditor which is measured by audit firm size or industry expert usually (Kim et al., 2003; Schmidt, 2012; Beasley & Petroni, 2001; Gramling & Stone, 2001; Solomon et al., 1999; Krishnan & Krishnan, 1997; Palmrose, 1988). The reporting of the breach is contingent upon auditor independence and client importance, auditor tenure, non-auditor fee, and auditor opinion type are always proxy variables (Stice, 1991; Lys & Watts, 1994; Cahan & Zhang, 2006; Boone et al., 2011; Schmidt, 2012; Kaplan & Williams, 2013; Blacconiere & DeFond, 1997).

2.1.2. Affecting factors of litigation risk: Legal environment

The 2001 Enron case not only impacted the US capital market but also increased the attention paid to capital market regulatory and supervisory systems in other countries. Auditing standards require auditors to identify fraud risks during the planning stages of their audits and then design audit procedures to investigate the identified risks (American Institute of Certified Public Accountants [AICPA], 2002). In Taiwan, a similar requirement of fraud risks assessment is provided by SAS of Taiwan No. 43 and increases auditors' liability. Otherwise, fraud cases from 2004 led the FSC to issue warnings about or cancel the certification of several CPAs. The FSC made major revisions to the Certified Public Accountant Act, which increases auditors' civil and criminal liabilities. In 2008, the court issued criminal sentences to the two CPAs who were involved in the China Rebar. Lin and Lin (2010) warned the accounting and auditing profession that the criminal liability of auditors will increase in the future when cases similar to the China Rebar happen again.

2.2. Use of data mining techniques in the formation of the warning model

Studies on predicting or detecting models developed rapidly after the first use of univariate discriminant analysis by Beaver (1966), in which sample matching was used to predict financial crises in sample US companies. For example, Altman (1968) used multiple discriminant analysis (MDA) to construct a bankruptcy detection model; the results selected 22 financial ratios and developed the Z-score model often used in later studies (Merkevicius, Garšva, & Girdzijauskas, 2006; Kirkos, Spathis, & Manolopoulos, 2007). Ohlson (1980) used logistic regression to develop a bankruptcy detection model using 9 financial ratios. Subsequent studies (Keasey & Watson, 1987; Lussier, 1995; Sheppard, 1994; Slowinski & Zopounidis, 1995; Doumpos & Zopounidis, 1999) on crisis prediction models not only included non-financial variables in the model but also tried different statistical methods and tools to create warning models. Recently, related studies in many business domains have shown that machine learning techniques, such as neural networks, decision trees, and support vector machines (SVM), are superior to traditional statistical methods. They can be used to discover interesting patterns or relationships from a given dataset and predict or classify new unknown instances (Varetto, 1998; Cristianini & Shawe-Taylor, 2000; Min & Lee, 2005; Lu, Lin, & Lin, 2016).

Frawley, Piatetsky-Shapiro, and Matheus (1992) stated that machine learning techniques find potential data hidden in previously unknown valuable information. Simply put, machine learning techniques efficiently search databases for useful knowledge and principles by finding patterns and relationships. Bose and Mahapatra (2001) introduced machine learning techniques used to deal with four problem types in the business area. The first type consists of a prediction problem, which examines past observed values for an attribute to infer a future value for the attribute; for example, stock returns prediction model (Tsai, Lin, Yen, & Chen, 2011) or a credit rating prediction model (Tsai & Chen, 2010). The second type consists of classification problems, which define analyzed attributes and create classes; for example, Ravisankar, Ravi, Raghava Rao, and Bose (2011) used machine learning techniques such as multilayer feed forward neural network (MLFF), support vector machines, genetic programming (GP), group method of data handling (GMDH), logistic regression, and probabilistic neural network (PNN) to identify companies that resort to financial statement fraud. Tsai, Lu, and Yen (2012) used feature selection in data mining to screen important variables affecting intangible assets, creating an intangible asset assessment and classification model to aid investors in determining whether companies have intangible assets. Kuzey, Uyar, and Delen (2014) used a decision tree and neural network to create a corporate value classification model. The third type consists of an association problem, which determines which related items should be grouped; a commonly used technology of this type is association rules. For example, Lu, Tsai, and Yen (2010) used association rules to find six factors that influenced corporate values for Taiwanese businesses. The fourth type consists of a detection problem, which combines prediction and classification functions. In this problem, machine learning can infer future values of attributes according to past values' and then classify them. Common applications include financial statement fraud detection and crisis detection or warning models (Martens, Bruynseels, Baesens, Willekens, & Vanthienen, 2008; Kwak, Eldridge, Shi, & Kou, 2011; Eldridge, Kwak, Venkatesh, Shi, & Kou, 2012; Kwak, Shi, & Kou, 2012; Ngai, Hu, Wong, Chen, & Sun, 2011; Perols et al., 2017; Bao, Ke, Li, Yu, & Zhang, 2020). Zhou and Kapoor (2011) employed decision tree, neural networks and Bayesian networks to identify fraud. The effectiveness of these machine learning techniques (and their limitations) is examined, especially when new schemes of financial statement fraud adapt to the detection techniques.

Recently, studies have used machine learning techniques to construct prediction or detection models and have found that the prediction outcomes are superior to those of traditional statistical methods (Cristianini & Shawe-Taylor, 2000; Min & Lee, 2005; Lu et al., 2016). Coats and Fant (1993) used neural networks and MDA to create financial distress models base on the five ratios: working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, the market value of equity/book value of total debt, and sales/total assets. This study examined 282 firms that were in operation from the period 1970–1989. Half of the firms of the sample were used to develop NN and MDA models and the rest served as a test sample. The test results suggest that the NN approach is more effective than MDA for the early detection of financial distress. Chaveesuk, Srivaree-Ratana, and Smith (1999) explored three of the most well-known supervised neural network paradigms: backpropagation, radial basis function, and learning vector quantization, for the task of rating US corporate bonds. Using generally available historic data, bonds are assigned to ratings based on a classification scheme. Comparisons were made with logistic regression and multiple regression models on both the data set used to create the predictive models and on new data. The results indicated that back-propagation neural networks (BPNs) were the superior method. Min and Lee (2005) used 1,888 firms including bankruptcy and non-bankruptcy cases and applied SVM to the bankruptcy prediction problem in an attempt to suggest a new model with better explanatory power and stability. The study used a grid-search technique using 5-fold cross-validation to find out the optimal parameter values of kernel function of SVM and compared its performance with those of MDA, LR, and three-layer fully connected back-propagation neural networks. The experiment results show that SVM outperforms the other methods. In summary, prior research has found that the warning model prediction accuracy of machine learning techniques is superior to that of traditional statistical methods.

3. RESEARCH METHODOLOGY

3.1. Research sample

The sample for this study consisted of 64 companies listed on the Taiwan Stock Exchange Corporation which had litigation taken against their auditors. These companies were chosen from the 2002–2013 "*Summary of Indictments and Sentences for Major Securities Crimes*" issued by the Financial Supervisory Commission R.O.C. Securities and Futures Bureau (2009) and the litigation cases announced by the Securities and Futures Investors Protection Center. In 2002, Securities and Futures Investors Protection Center was established and has helped investors to sue many illegal companies and their auditors. However, the litigation cases are limited and there are not many samples of auditors being sued together in Taiwan. Major fraud cases in recent years have been discovered after many years (e.g., Wirecard, and Ya Hsin Industrial Co., Ltd.). This study tried to find out the characteristics of high litigation risk companies by using matching samples. Therefore, we confirm that the non-litigation companies are still legal after many years. Financial industries were not included in the sample since the risk assessment and industry characteristics of these industries are much different from others. Sample matching was conducted by the methods used in Beaver (1966); non-litigated companies within the same period, similar industries, and with similar asset scales acted as the control sample. A 1:2 sample matching method (Coats & Fant, 1993) comprised of 64 litigated companies and 128 non-litigated companies was used in this study.

Table 2 shows that the litigated companies covered 13 industries; 56.3% of this sample was the electronics industry sample (28 companies) and the construction industry sample (8 companies). Many litigation cases happened in 2007 and 2008, a period after the China Rebar case.

Table 2. Industry and year distributions of litigated companies

Industry	*2002*	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*Total*
Food	3											3
Rubber			1		1							2
Textile				1			5					6
Electrical equipment		1	1				1	1				4
Wire and cable	1					1						2
Biochemical						1					1	2
Steel	1	2			1							4
Glass and ceramics	1											1
Electronics	1	3	2		3	4	5	1	3	4	2	28
Construction	2	1	3			2						8
Aviation						1						1
Tourism	1											1
Trade and consumer goods		2										2
Total	10	9	7	1	5	9	11	2	3	4	3	64

Moreover, the cross-validation method is used to construct prediction models, to avoid sample variability and minimize any biasing effect (Tam & Kiang, 1992). Specifically, this study considers a 10-fold cross-validation method, since this is the most commonly used strategy to examine the performance of classifiers. Further, it is based on dividing the whole dataset into 10 equal parts, from which 90% of the dataset is selected and used for model training, and the other 10% is used for model testing. Therefore, every subset is trained 9 times and tested once, and from this, the average prediction performance is obtained.

3.2. Variable definition and measurement

This study examined the factors influencing audit or litigation. Influencing factors were divided into inherent risk, control risk, detection risk, and legal environment. Table 1 provides a large number of academic literature searched from business databases or top journals related to accounting and business. After referencing 35 years of relevant literature, 39 representative variables were selected (Table 3). The dependent variable was litigation against auditors, where 1 indicated litigation and 0 if not.

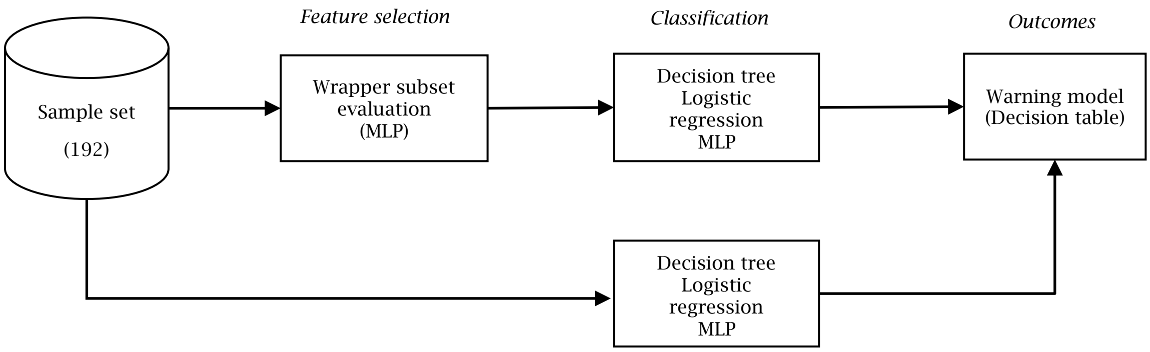
Table 3. The measurement of research variables

Variable		Measurement
Affecting factors of audit risk: Inherent risk		
1	Asset scale	Natural logarithm of total assets
2	Period listed on stock exchange	Years listed on stock exchange
3	Credit risk index	Credit risk index listed in the Taiwan Economic Journal (TEJ) database
4	Debt ratio	Total debt/total assets
5	Accounts receivable ratio	Accounts receivable/total assets
6	Accounts receivable turnover rate	Annual sales/average accounts receivable and negotiable instruments
7	Inventory ratio	Inventory/total assets
8	Inventory turnover rate	Annual sales/average inventory
9	Inventory growth rate	(Average inventory for current period/average inventory for previous period) - 1
10	Sales growth rate	(Sales revenue for current period/sales revenue for previous period) - 1
11	Operating profit ratio	Operating profit ratio/sales revenue
12	Operating profit growth rate	(Operating profit for current period/operating profit for previous period) - 1
13	Return on operating assets	Operating profit for the past four quarters/average total assets
14	Annual return on stock	Market-adjusted annual return on stock
15	Stock price fluctuation	Fixed asset model beta coefficient
16	Annual loss	1 for annual net income after taxes that signified loss; 0 if not
17	Heavy losses	1 for shareholder equity less than authorized capital stock; 0 if not
18	Working capital	Working cash flow/total debt
19	Insufficient working cash flow	1 for negative working cash flow; 0 for positive
20	Type of industry	1 for companies in the electronics industry; 0 if not
Affecting factors of audit risk: Control risk		
21	Chairperson is an executive director	1 for if the chairperson is an executive director; 0 if not
22	Change of chairperson	Number of times a new chairperson was named within the past three years
23	Ratio of stock owned by directors	Number of stocks owned by directors/number of stocks sold to outsiders at year-end
24	Ratio of independent board members	Number of independent board members/total number of board members
25	Ratio of stock owned by legal entities	Number of stocks owned by legal entities/total number of stocks sold to outsiders at year-end
26	Change of auditor	Number of times a new auditor was appointed within the past three years
27	Illegal behavior	1 for any illegal activity committed by the company or auditor that was punished by the competent authority; 0 for none.
28	Restatement of financial reports	Number of times financial restatements were made
Affecting factors of audit risk: Detection risk		
29	Large audit firm	1 for companies audited by one of the four largest audit firms; 0 if not
30	Industry expert_CPA	Number of clients listed on the stock exchange in an industry/total number of companies listed on the stock exchange in that industry
31	Client importance_CPA	log(revenue of company under investigation)/log(revenues of all CPA clients listed on the stock market)
32	Client importance_firm	log(revenue of company under investigation)/log(revenues of all audit firm clients listed on the stock market)
33	Non-audit fees	1 for if the company's non-audit service fees account for over 1/4 of all fees paid to the audit firm; 0 if not
34	Audit report lag	Number of days between balance sheet date and audit report date
35	Audit firm tenure	Number of years the company has been audited by the same audit firm
36	CPA tenure	Number of years the company has been audited by the same certified accountant
37	Auditor industry experience	Number of years the auditor has audited within the industry to which the company belongs
38	Audit opinion type	1 for if the company received an unqualified opinion the previous year; 0 if not
Legal environment		
39	China Rebar case	1 for if the year of audit was after the China Rebar case (including 2007); 0 if not

3.3. Constructing the litigation warning model by using data mining techniques

After gathering and organizing the influencing factors of auditor litigation, feature selection extract critical factors which were then categorized to create the litigation warning model and decision table to help auditors determine if they should accept or terminate an engagement. The detailed research procedure is shown in Figure 1.

Figure 1. Model construction process



3.3.1. Feature selection

Improving prediction accuracy is an important part of creating an effective prediction or detection model. However, not all features and attributes within the database are associated with prediction accuracy. Methods to improve performance by effectively removing clutter and irrelevant attributes in data mining are called feature selection processes; commonly used machine learning techniques are decision tree and association rules (Questier, Put, Coomans, Walczak, & Heyden, 2005; Sugumaran, Muralidharan, & Ramachandran, 2007). In feature selection, a subset of representative features is selected from the training dataset for use in model construction. There are several advantages to be obtained by using feature selection. For example, the feature dimensionality is reduced in the feature space, which could enhance generalization because the overfitting problem is reduced. In addition, the computational cost of training a prediction model is also reduced (Questier et al., 2005; Sugumaran et al., 2007; Tsai et al., 2012; Lin, Lu, & Tsai, 2019).

Decision trees were first proposed by Quinlan (1986) in applied machine learning for dimension reduction and categorical data algorithms. Decision trees are comprised of roots, nodes, branches, and leaf nodes. While creating the decision tree, attribute selection measures screen for variables suitable for classifying data; the selected variables can be seen as key influencing factors for data sorting. The advantages of creating a decision tree are that parameters do not need to be set and it is applicable for exploring knowledge and finding key variables. Association rules are also known as shopping basket analysis; these rules extract intercorrelated knowledge hidden within the data to help find algorithms for important factors more closely associated with dependent variables. Two key measures are used to calculate the strength of the associations between variables. The first is support, which is the percentage of the number of times an item appears in the data. The second is confidence, which is the prediction strength of the rule. For example, the support of A for B is the percentage of $A \cup B$ and the confidence of A for B is the ratio of $A \cup B$ to A. The variables included in the rules that meet the minimum support and confidence are considered key influencing factors.

To assess the performance of decision tree and association rules in feature selection, the extracted features were then input into a multilayer perceptron (MLP) neural network which is most widely used in the many predictions and forecasting domains (Tsai & Wu, 2008). Three steps were taken in the evaluation of feature selection performance in this study. First, the dataset including all features was used to train and test the MLP model as the basis for the evaluation. Second, the extracted features from the decision tree and association rules were separately used to train and test the MLP models for comparison. Third, the performances of each model including prediction accuracy, type I and type II error rates, and the feature extraction rate

3.3.2. Classification

Classification is one of the most important techniques in machine learning and is used to categorize the data to be processed according to attributes. A classification technique commonly used in prior research, the decision tree, was chosen for this study. This method was compared with the traditional statistical methods of logistic regression and discriminant analysis. As a classification technique, the decision tree uses a known example to create a tree-shaped structure and induce rules for the example. Advantages of a decision tree not provided by logistic regression and discriminant analysis include the creation of a decision table and the easy interpretation of the extracted rules. The efficient data processing complies with the objective of this study to create an easily understood and convenient litigation early warning model and decision table.

A decision table was used to present and analyze decision situations. The columns in Table 4 can be seen as conditions and actions, whereas the rows are test items. The conditions are factors related to the decision, and the actions are possible outcomes for the decision (for example, whether or not litigation will be taken against the auditor). The value for the corresponding subset is presented under each condition; each action input is distributed to the corresponding action. Therefore, each row in the decision table is a classification rule (Martens et al., 2008). The decision rules produced by the decision tree in this study can help auditors assess the risk of litigation and decide whether to accept an engagement or expand audit procedures to reduce the risk of litigation.

Table 4. Decision table patterns

if			*then*	
Qualitative variable	***Quantitative variable 1***	***Quantitative variable 2***	***High risk of litigation***	***Low risk of litigation***
Yes	> 0.5	> 0	–	✓
		≤ 0	✓	–
	≤ 0.5	–	✓	–
No	–	> 1	–	✓
		≤ 1	✓	–

3.3.2. Model performance evaluation

For feature selection performance evaluations, the prediction accuracies, type I and type II error rates, and feature extraction rates of the two models created using feature selection tools were compared with the benchmark model made without feature selection. The method used to calculate prediction accuracy is shown in Table 5. This was the ratio of correct prediction data to total data (equation (1)). Type I errors are the incorrect rejection of a true null hypothesis. In the context of this study, this was the probability that the outcome where litigation would not be taken against the auditor was mistakenly classified as the outcome where litigation would be taken against the auditor (equation (2)). Type II errors are the incorrect acceptance of a false null hypothesis. This was the probability that the outcome where litigation would be taken against the auditor was mistakenly classified as the outcome where litigation would not be taken against the auditor (equation (3)). Analysis of variance (ANOVA) was also used to determine whether the differences in performance between the three models were significant.

For classification technique performance evaluations, the prediction accuracies of the early warning models created using the decision tree, logistic regression, and discriminant analysis were compared. The evaluation methods were the same as those for the feature selection processes. The area under the receiver operating characteristic (ROC) curve (AUC)[1] (equation (4)) was also used to determine the predictive accuracy of the models.

Table 5. Early warning model performance classification matrix

Actual/prediction	***No litigation against the auditor***	***Litigation against the auditor***
No litigation against the auditor	(a)	(b)
Litigation against the auditor	(c)	(d)

Equations for prediction accuracy and type I/type II error rates:

$$Prediction\ accuracy\ = \frac{a+d}{a+b+c+d} \quad (1)$$

$$Type\ I\ error\ rate\ = \frac{b}{a+b} \quad (2)$$

$$Type\ II\ error\ rate\ = \frac{c}{c+d} \quad (3)$$

$$AUC\ = \frac{1}{2}\left[\left(\frac{a}{a+b}\right)+\left(\frac{d}{c+d}\right)\right] \quad (4)$$

4. EMPIRICAL RESULTS AND ANALYSIS

4.1. Warning model analysis

4.1.1. Feature selection results

The results in Table 6 show that as the benchmark model without a feature selection tool, the average training and testing times were the longest; moreover, the large number of dependent variables caused interference which lead to the poorest prediction accuracy (74.167%), type I error rate (19.375%), and type II error rate (38.750%) among the three models. The 10 variables chosen using the association rules effectively reduced the average training and testing times, but this model had poor prediction accuracy (71.292%), type I error rate (18.094%), and type II error rate (49.938%). The 6 variables chosen using the decision tree effectively reduced the average training and testing times and had the best prediction accuracy (85.625%), type I error rate (7.250%), and type II error rate (28.625%) among the three models. In addition, ANOVA was used for comparative analysis of the performances. Table 7 shows that the performances of the decision tree, association rules, and benchmark models were significantly different. The decision tree model had the best performance, followed by the association rules and benchmark models. According to the above, feature selection reduced both the training and testing times as well as interference from an excessive number of variables, improving the model's accuracy, efficiency, and effectiveness (Questier et al., 2005; Sugumaran et al., 2007; Lin, Lu, & Tsai, 2019). Moreover, the accuracy rate of 85.625% confirms that the six factors including *credit risk index*, *stock price fluctuation*, *client importance_firm*, *accounts receivable ratio*, *audit report lag*, and *CPA tenure* extracted from the decision tree can be seen as key factors for determining auditors' litigation risk.

The first factor was the *credit risk index*[2] with the greatest information gain in the decision tree. Credit risk mainly measures the corporate risk of bankruptcy. Prior studies have found that the main reason for litigation against auditors was related to client bankruptcy or financial distress (Pierre & Anderson, 1984; Palmrose, 1987; Lys & Watts, 1994). The second key factor was *stock price fluctuation*;

[1] Sokolova and Lapalme (2009) pointed out that the ROC curve indicates the trade-off between the true positive rate and false positive rate for the performance evaluation of a binary variable. The ROC curves for each classification model can be drawn for comparison, where the AUC serves as the indicator for the models' performance. The AUC values range from 0 to 1. An AUC of 1 indicates a perfect model, an AUC between 0.5 and 1 indicates that the model is better than random guessing and has predictive value, and an AUC less than or equal to 0.5 indicates that the model is equivalent to random guessing and has no predictive value.

[2] The TEJ Taiwan Corporate Credit Risk Index (TCRI) is a credit risk index first developed in 1991. TCRI rating uses semi-professional judgment to assess the credit risk of all public companies in Taiwan from public data. This risk index differs from traditional external credit assessment institutions because in Taiwan, credit emphasizes loans and ignores the bond market. The party who most requires credit information is not the debt holder, but the bank lender and stock market investor; as such, this indicator is of useful reference to banks and investors.

when a company's stock prices fluctuate greatly, the possibility for litigation against the auditor increases. Because stock prices are determined by the company's financial situation and negative information, failed shareholder investments often involve litigation against auditors (Carcello & Palmrose, 1994). The third factor was *client importance_firm*. This study found that the importance of each client was a significant factor affecting litigation against auditors. The distribution of profits in Taiwanese audit firms is correlated to the contribution of each auditor's fees; therefore, auditors accept auditing cases with high risks of litigation to contribute more to the firm (Lee & Chen 2004). The fourth key factor was the *accounts receivable ratio*. The uncertainty of accruals may result in potential errors in assets evaluation or operational doubts; for example, underestimation of allowance for doubtful accounts, or manipulation of accruals to cover up financial difficulties (Francis & Krishnan, 1999). The fifth factor was *audit report lag*. This lag is defined as the time between the last day of the fiscal year and the day of an audit report. The possibility of fraud and manipulation increases as a company's financial situation worsens. To avoid the risk of litigation, auditors embellish auditing procedures, which lengthens the audit period (Bamber, Bamber, & Schoderbek, 1993). The final key factor was *CPA tenure*. Chen, Lin, and Lin (2008) and Lee and Lin (2005) found that the lengths of tenure of the firm and the individual CPA both influence audit quality positively. *CPA tenure* helps maintain the quality of financial statements; in consideration of the risk of litigation, auditors retain better clients to reduce risk.

Table 6. Feature selection tool performance assessment

Model	*Number of features*	*Rate of extraction (%)*	*Accuracy (%)*	*Type I error rate (%)*	*Type II error rate (%)*
Comparison model: Decision tree	6	17.949	85.625	7.250	28.625
Comparison model: Association rules	10	25.641	71.292	18.094	49.938
Benchmark model: All variables	39	100.000	74.167	19.375	38.750

Table 7. Comparison of feature selection tool performances

	Comparison model: Decision tree		*Comparison model: Association rules*		*Benchmark model: All variables*		*Main effect*	*Comparison*
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*		
Accuracy (%)	85.625	1.145	71.292	2.913	74.167	1.591	349.837***	Decision tree > Benchmark model > Association rules
Type I error rate (%)	7.250	1.998	18.094	3.724	19.375	1.444	166.834***	Decision tree < Association rules < Benchmark model
Type II error rate (%)	28.625	1.844	49.938	2.571	38.750	4.134	314.533***	Decision tree < Benchmark model < Association rules

*Note: Main effects in this table are F-values; * p < 0.05; ** p < 0.01; *** p < 0.001.*

4.1.2. Classification results

After analysis of the feature selection results, the six extracted features were used to create an early warning model for litigation against auditors and a decision table using a decision tree. These were then compared to logistic regression and discriminant analysis methods which use the same six extracted features to provide auditors with a reference when selecting clients. The performance assessment results are shown in Table 8. Inspection of the AUC values for the three models revealed that all AUC was greater than 0.5, indicating that all three models have predictive value. The decision tree model developed in this study had an accuracy of 92.708% and the type I and type II error rates were 1.563% and 18.750%, respectively. The performance of this model was superior to those created using logistic regression and discriminant analysis.

Table 8. Classification performance assessment

Model	*Accuracy (%)*	*Type I error rate (%)*	*Type II error rate (%)*	*AUC value*
Machine learning model: Decision tree	92.708	1.563	18.750	0.915
Statistical model: Logistic regression	86.979	8.594	21.875	0.895
Statistical model: Discriminant analysis	83.333	14.844	20.313	0.824

4.2. Warning model decision information

Table 9 shows the decision table with eight key rules for decision-makers to assess the litigation risk. Rule 1 in the decision table indicates that when the *credit risk index* is 9[3], auditors have a high litigation risk. Thus, when a company faces worsening operations and is labeled a high credit risk, a financial crisis may lead to a lawsuit from

[3] TCRI rates on a scale from 1 to 9 rather than the international practice of using the English alphabet. Grading is relative; i.e., 1 is better than 2, 2 is better than 3, etc., and 9 is the worst. Scores of 7–9 mark the high risk group. These companies usually have had long-term losses, have broken even but have poor quality accounting information, or have broken even but have weak financial structure and poor fluidity. Therefore, scores of 7–9 indicate high risk and high financial stress. Scores of 5–6 mark the moderate risk group. These companies usually have a stable financial structure but poor or unstable profits, or have good profits but a weak financial structure; these companies are less able to withstand financial downturns than companies with the top four scores. Scores of 1–4 mark the low risk group. These companies usually have stable profits and financial structures, maintain moderate to high fluidity, and are able to withstand financial downturns. Therefore, scores of 1–4 indicate low risk.

the investors against the management and auditors for compensation (Pierre & Anderson, 1984; Palmrose, 1987; Lys & Watts, 1994). However, Rule 8 indicates that when the *credit risk index* is less than or equal to 6 (i.e., a company labeled a low or moderate credit risk), the company's financials are more stable and credit risk and the risk of the financial crisis are low; thus, auditors have a low litigation risk.

Rules 2–7 in the decision table consider multiple variables in determining whether a client is a high or low litigation risk. Rule 2 indicates that when the *credit risk index* is 7 or 8 and the *stock price fluctuation* is less than 0.37, the company has stable performance in the securities market as the stock prices have not caused large fluctuations; therefore, auditors have a low litigation risk. Rule 3 indicates that when the *credit risk index* is 7 or 8 and the *stock price fluctuation* is greater than 0.37, if the *client importance_firm* is greater than 0.52 (the client's fees account for no less than 52% of the firm's entire income), then auditors have a high litigation risk. The main reason for this is that auditors may rely on economic factors and lose their independence (Reynolds & Francis, 2000); auditors should avoid economic dependence on a single client which may affect their audit quality. Rules 4–7 indicate that when the *credit risk index* is 7 or 8, the *stock price fluctuation* is greater than 0.37, and the *client importance_firm* is less than 0.52, then the risk of facing litigation depends on the *accounts receivable ratio*, *audit report lag*, and *CPA tenure*. Rule 5 indicates that under the conditions above, if the accounts receivable rate is greater than 16.18 (i.e., the *accounts receivable ratio* account for over 16.18% of the total assets), and the *audit report lag* is over 84.5 days, then auditors have a high litigation risk. Rule 7 indicates that in the conditions above, if the *audit report lag* is less than 84.5 days and the auditor has been appointed for less than 5 years, then auditors also have a high litigation risk. In addition, Rules 4 and 6 indicate that auditors have a low litigation risk. The main discrepancy between these rules is in the influences of *CPA tenure* and *audit report lag*; when the client has a higher ratio of *accounts receivable ratio*, if the CPA has been newly appointed and requires more time to complete the audit, this indicates that the auditor has little knowledge of the company, which increases the risk of litigation (Pierre & Anderson, 1984). Thus, auditors should broaden the scope of audits to reduce this risk.

Table 9. Auditors' litigation risk decision table

<table>
<tr><th></th><th colspan="6">if</th><th colspan="2">then</th></tr>
<tr><th>Rule</th><th>Credit risk index</th><th>Stock price fluctuation</th><th>Client importance_firm</th><th>Accounts receivable ratio</th><th>Audit report lag (days)</th><th>CPA tenure</th><th>High risk of litigation</th><th>Low risk of litigation</th></tr>
<tr><td>1</td><td>9</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>✓</td><td></td></tr>
<tr><td>2</td><td rowspan="6">7 ≤ 8</td><td>≤ 0.37</td><td>–</td><td>–</td><td>–</td><td>–</td><td></td><td>✓</td></tr>
<tr><td>3</td><td rowspan="5">> 0.37</td><td>> 0.52</td><td>–</td><td>–</td><td>–</td><td>✓</td><td></td></tr>
<tr><td>4</td><td rowspan="4">≤ 0.52</td><td>≤ 16.18</td><td>–</td><td>–</td><td></td><td>✓</td></tr>
<tr><td>5</td><td rowspan="3">> 16.18</td><td>> 84.5</td><td>–</td><td>✓</td><td></td></tr>
<tr><td>6</td><td rowspan="2">≤ 84.5</td><td>> 5</td><td></td><td>✓</td></tr>
<tr><td>7</td><td>≤ 5</td><td>✓</td><td></td></tr>
<tr><td>8</td><td>≤6</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td></td><td>✓</td></tr>
</table>

4.3. Results discussion

According to the above results, in addition to having better accuracy than logistic regression and discriminant analysis, the decision tree also provides a decision table that illustrates the associations and rules between important factors and the litigation risk.

Recently years have seen a growing trend of artificial intelligence (AI), especially machine learning, application in auditing (Perols et al., 2017; Bao et al., 2020). Global Big 4 public accounting firms are actively exploring adopting AI and machine learning techniques in their audit services. For example, KPMG is constructing an intelligent audit platform, Clara, which embodies cognitive and predictive technologies[4]. EY is embedding AI technologies in their audit process, especially applying AI to document reading and interpretation, adopting automation to improve audit efficiency, and using drones to assist inventory examination[5]. Deloitte uses Deloitte AI Robot in relevant audit processes to improve audit efficiency, including document review and analysis and accounting knowledge inquiry[6]. Furthermore, PricewaterhouseCoopers (PwC) is building AI platforms to detect abnormal transactions in the general ledger, especially for cash-related accounts[7].

The reference created from the research also can be used as a supplementary tool when auditors evaluate litigation risk to reduce risk probability and prevent damage to the reputations of both the auditor and the audit firm.

5. CONCLUSION

After the recent series of fraud cases, investor protection mechanisms have urged the competent authority to revise the legal system to increase the legal liability of auditors, thus preventing fraud and litigation due to audit failure and improving audit quality and the degree of confidence in financial statements. Changes in the legal environment and laws have made it more difficult for auditors to assess whether to accept clients and to avoid indemnification and a damaged reputation due to litigation. The warning model and decision table for auditor litigation developed in this study can serve as a reference for auditors. Prior literature regarding litigation risk against auditors was reviewed to collect affecting factors for both audit

[4] htpps://home.kpmg/xx/en/home/insights/2017/05/kpmg-clara-automated-agile-intelligent-and-scalable.html
[5] https://www.ey.com/en_us/audit/innovation
[6] https://www2.deloitte.com/cn/en/pages/audit/articles/explore-audit-innovation-with-deloitte-ai-robot-vol-6.html
[7] https://www.pwc.com/gx/en/about/stories-from-across-the-world/harnessing-ai-to-pioneer-new-approaches-to-the-audit.html

risk-related and the legal environment. Feature selection was then used to extract critical variables, after which, categorization techniques constructed a representative and convenient warning model and decision table for auditors' reference.

Two procedures were used to construct the warning model. First, 39 influencing variables were collected from prior literature and a decision tree was used to select six key factors: *credit risk index*, *stock price fluctuation*, *client importance_firm*, *accounts receivable ratio*, *audit report lag*, and *CPA tenure*. Compared to the extracted variables using other selection tools, the variables chosen using the decision tree had higher accuracy and lower type I and type II error rates; therefore, these variables can be seen as key influencing factors for auditors' litigation risk. Second, the six extracted factors were used in a decision tree to construct a litigation warning model.. The results showed that the accuracy rate was 92.708% and the type I and type II error rates were below 10%; eight categorization rules were extracted which were compiled in a decision table.

This study contributed a new concept regarding the use of machine learning to determine the influencing factors of auditors' litigation risk. This study reviewed relevant literature and collected all influencing factors. Then, key affecting factors were extracted using feature selection methods, effectively increasing the prediction accuracy of the model. The warning model established using machine learning techniques in this study was higher than that for models created using logistic regression and discriminant analysis, demonstrating the value of applying machine learning techniques in other relevant areas.

Because the sample used in this study consisted of companies listed on the Taiwan Stock Exchange Corporation, the legal liability of Taiwanese auditors was weaker than that of auditors in developed countries. However, the complexity of auditor litigation is increasing. The factors extracted during feature selection can be seen as the key factors influencing the litigation risk for auditors in Taiwan and serve as a reference for other developing countries. Additionally, the objective of this study was to create a warning model for auditor litigation. The difference between this and prior studies predicting auditor litigation risk was that this study did not aim to improve model accuracy but to create an understandable classification model founded on rules for auditors to use when assessing audit-related risk. According to the valuable information from the warming model, following audit planning strategies may improve audit quality and lower the litigation risk.

Finally, while this study collected all influencing factors for auditors' litigation risk from previous literature, some other factors may have been left out or measured using different methods. Therefore, future studies can include other variables to construct a more complete and accurate warning model. Affecting factors of litigation risk may exist in huge variations in different industries. Therefore, the issue of auditor litigation risk in various industries is interesting in future studies. Otherwise, the litigation cases are limited and there are not many samples of auditors being sued together in Taiwan. Future studies could extend the research period and increase the research sample to confirm the research results.

REFERENCES

1. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x
2. American Institute of Certified Public Accountants (AICPA). (2002). *Consideration of fraud in a financial statement audit* (SAS No. 99 and SAS No. 113, AU Section 316, pp. 1719–1770). New York, NY: AICPA. Retrieved from https://us.aicpa.org/content/dam/aicpa/research/standards/auditattest/downloadabledocuments/au-00316.pdf
3. Arens, A. A., Elder, R. J., & Beasley, M. S. (2014). *Auditing and assurance services* (15th ed). London, the UK: Pearson.
4. Bamber, E. M., Bamber, L. S., & Schoderbek, M. P. (1993). Audit structure and other determinants of audit report lag: An empirical analysis. *AUDITING: A Journal of Practice & Theory, 12*(1), 1–23. https://www.proquest.com/scholarly-journals/audit-structure-other-determinants-report-lag/docview/216733635/se-2?accountid=10820
5. Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research, 58*(1), 199–235. https://doi.org/10.1111/1475-679X.12292
6. Beasley, M. S. (1996). An empirical analysis of the relation between the board of director composition and financial statement fraud. *The Accounting Review, 71*(4), 443–465. Retrieved from https://www.jstor.org/stable/248566
7. Beasley, M. S., & Petroni, K. R. (2001). Board independence and audit-firm type. *AUDITING: A Journal of Practice & Theory, 20*(1), 97–114. https://doi.org/10.2308/aud.2001.20.1.97
8. Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71–111. https://doi.org/10.2307/2490171
9. Blacconiere, W. G., & DeFond, M. L. (1997). An investigation of independent audit opinions and subsequent independent auditor litigation of publicly-traded failed savings and loans. *Journal of Accounting and Public Policy, 16*(4), 415–454. https://doi.org/10.1016/S0278-4254(96)00042-7
10. Bonner, S. E., Palmrose, Z.-V., & Young, S. M. (1998). Fraud type and auditor litigation: An analysis of SEC accounting and auditing enforcement releases. *The Accounting Review, 73*(4), 503–532. Retrieved from https://www.marshall.usc.edu/sites/default/files/sbonner/intellcont/BonnerPalmroseYoung1998-1.pdf
11. Boone, J. P., Khurana, I. K., & Raman, K. K. (2011). Litigation risk and abnormal accruals. *AUDITING: A Journal of Practice & Theory, 30*(2), 231–256. https://doi.org/10.2308/ajpt-50003
12. Bose, I., & Mahapatra, R. K. (2001). Business data mining — A machine learning perspective. *Information & Management, 39*(3), 211–225. https://doi.org/10.1016/S0378-7206(01)00091-X

13. Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications, 32*(2), 277–288. https://doi.org/10.1016/j.eswa.2005.11.037
14. Cahan, S. F., & Zhang, W. (2006). After Enron: Auditor conservatism and ex-Andersen clients. *The Accounting Review, 81*(1), 49–82. https://doi.org/10.2308/accr.2006.81.1.49
15. Calabresi, G. (1970). *The cost of accidents: A legal and economic analysis.* London, the UK: Yale University Press.
16. Carcello, J. V., & Palmrose, Z.-V. (1994). Auditor litigation and modified reporting on bankrupt clients. *Journal of Accounting Research, 32*, 1–30. https://doi.org/10.2307/2491436
17. Casterella, J. R., Jensen, K. L., & Knechel, W. R. (2010). Litigation risk and audit firm characteristics. *AUDITING: A Journal of Practice & Theory, 29*(2), 71–82. https://doi.org/10.2308/aud.2010.29.2.71
18. Chaveesuk, R., Srivaree-Ratana, C., & Smith, A. E. (1999). Alternative neural network approaches to corporate bond rating. *Journal of Engineering Valuation and Cost Analysis, 2*(2), 117–131. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.42.4331&rep=rep1&type=pdf
19. Chen, C.-Y., Lin, C.-J., & Lin, Y.-C. (2008). Audit partner tenure, audit firm tenure, and discretionary accruals: Does long auditor tenure impair earnings quality? *Contemporary Accounting Research, 25*(2), 415–445. https://doi.org/10.1506/car.25.2.5
20. Cho, S., Vasarhelyi, M. A., Sun, T., & Zhang, C. (2020). Learning from machine learning in accounting and assurance. *Journal of Emerging Technologies in Accounting, 17*(1), 1–10. https://doi.org/10.2308/jeta-10718
21. Coats, P. K., & Fant, L. F. (1993). Recognizing financial distress patterns using a neural network tool. *Financial Management, 22*(3), 142–155. https://doi.org/10.2307/3665934
22. Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods.* https://doi.org/10.1017/CBO9780511801389
23. DeAngelo, L. E. (1981). Auditor size and audit quality. *Journal of Accounting and Economics, 3*(3), 183–199. https://doi.org/10.1016/0165-4101(81)90002-1
24. Doumpos, M., & Zopounidis, C. (1999). A multicriteria discrimination method for the prediction of financial distress: The case of Greece. *Multinational Finance Journal, 3*(2), 77–145. https://doi.org/10.17578/3-2-1
25. Eldridge, S., Kwak, W., Venkatesh, R., Shi, Y., & Kou, G. (2012). Predicting auditor changes with financial distress variables: Discriminant analysis and problems with data mining approaches. *Journal of Applied Business Research, 28*(6), 1357–1372. https://doi.org/10.19030/jabr.v28i6.7349
26. Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *The Journal of Law and Economics, 26*(2), 301–325. https://doi.org/10.1086/467037
27. Financial Supervisory Commission R.O.C. Securities and Futures Bureau. (2009). *Summary of indictments and sentences for major securities crimes.* Retrieved from https://www.tpex.org.tw/storage/governance/Summary_of_Indictments_and_Sentences_for_Major_Securities_Crimes.doc
28. Francis, J. R., & Krishnan, J. (1999). Accounting accruals and auditor reporting conservatism. *Contemporary Accounting Research, 16*(1), 135–165. https://doi.org/10.1111/j.1911-3846.1999.tb00577.x
29. Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine, 13*(3), 57–70. https://doi.org/10.1609/aimag.v13i3.1011
30. Gramling, A. A., & Stone, D. N. (2001). Audit firm industry expertise: A review and synthesis of the archival literature. *Journal of Accounting Literature, 20*, 1–29.
31. Kaplan, S. E., & Williams, D. D. (2013). Do going concern audit reports protect auditors from litigation? A simultaneous equations approach. *The Accounting Review, 88*(1), 199–232. https://doi.org/10.2308/accr-50279
32. Keasey, K., & Watson, R. (1987). Non-financial symptoms and the prediction of small company failure: A test of Argenti's hypotheses. *Journal of Business Finance & Accounting, 14*(3), 335–354. https://doi.org/10.1111/j.1468-5957.1987.tb00099.x
33. Kim, J.-B., Chung, R., & Firth, M. (2003). Auditor conservatism, asymmetric monitoring, and earnings management. *Contemporary Accounting Research, 20*(2), 323–359. https://doi.org/10.1506/J29K-MRUA-0APP-YJ6V
34. Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications, 32*(4), 995–1003. https://doi.org/10.1016/j.eswa.2006.02.016
35. Knechel, W. R., Naiker, V., & Pacheco, G. (2007). Does auditor industry specialization matter? Evidence form market reaction to auditor switches. *AUDITING: A Journal of Practice & Theory*, *26*(1), 19–45. https://doi.org/10.2308/aud.2007.26.1.19
36. Krishnan, J., & Krishnan, J. (1997). Litigation risk and auditor resignations. *The Accounting Review, 72*(4), 539–560. Retrieved from https://www.jstor.org/stable/248174
37. Kuzey, C., Uyar, A., & Delen, D. (2014). The impact of multinationality on firm value: A comparative analysis of machine learning techniques. *Decision Support Systems, 59*, 127–142. https://doi.org/10.1016/j.dss.2013.11.001
38. Kwak, W., Eldridge, S., Shi, Y., & Kou, G. (2011). Predicting auditor changes using financial distress variables and the multiple criteria linear programming (MCLP) and other data mining approaches. *Journal of Applied Business Research, 27*(5), 73–84. https://doi.org/10.19030/jabr.v27i5.5597
39. Kwak, W., Shi, Y., & Kou, G. (2012). Bankruptcy prediction for Korean firms after the 1997 financial crisis: Using a multiple criteria linear programming data mining approach. *Review of Quantitative Finance and Accounting, 38*(4), 441–453. https://doi.org/10.1007/s11156-011-0238-z
40. Lee, J. Z., & Chen, J. F. (2004). Importance on magnitude of earnings management: From the perspective of audit groups within the Big Five. *The International Journal of Accounting Studies, 38*, 59–80.
41. Lee, J. Z., & Lin, H. F. (2005). The relations between auditor tenure and abnormal accruals. *Management Review, 24*(4), 103–126.
42. Lev, B., Ryan, S. G., & Wu, M. (2008). Rewriting earnings history. *Review of Accounting Studies, 13*(4), 419–451. https://doi.org/10.1007/s11142-007-9041-4
43. Lin, C. C., & Lin, H. L. (2010). Auditor's liability for financial statement fraud in Taiwan from Li-Bar verdict: An empirical analysis. *National Taiwan University Law Journal, 39*(3), 223–288.
44. Lin, W.-C., Lu, Y.-H., & Tsai, C.-F. (2019). Feature selection in single and ensemble learning-based bankruptcy prediction models. *Expert Systems, 36*(1), e12335. https://doi.org/10.1111/exsy.12335

45. Liu, C., Wang, T., & Lai, S.-T. (2009). Litigation risk and large audit firms' acceptable level of clients' financial risk. *NTU Management Review, 20*(1), 1–40. Retrieved from https://www.proquest.com /openview/292901ca62629a7389ce2bf5b9fe2d41/1?pq-origsite=gscholar&cbl=2049105
46. Lu, Y. H., Lin, Y. C., & Lin, Y. L. (2016). Going-concern opinion: The application of data mining technologies. *Journal of Accounting Review, 63*, 77–108. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2815247
47. Lu, Y.-H., Tsai, C.-F., & Yen, D. C. (2010). Discovering important factors of intangible firm value by association rules. *The International Journal of Digital Accounting Research, 10*(3), 55–85. https://doi.org/10.4192/1577-8517-v10_3
48. Lussier, R. N. (1995). A nonfinancial business success versus failure prediction model for young firms. *Journal of Small Business Management, 33*(1), 8–20.
49. Lys, T., & Watts, R. L. (1994). Lawsuits against auditors. *Journal of Accounting Research, 32*, 65–93. https://doi.org/10.2307/2491440
50. Martens, D., Bruynseels, L., Baesens, B., Willekens, M., & Vanthienen, J. (2008). Predicting going concern opinion with data mining. *Decision Support Systems, 45*(4), 765–777. https://doi.org/10.1016/j.dss.2008.01.003
51. McKeown, J. C., Mutchler, J. F., & Hopwood, W. (1991). Towards an explanation of auditor failure to modify the audit opinions of bankrupt companies. *AUDITING: A Journal of Practice & Theory, 10*, 1–13.
52. Merkevicius, E., Garšva, G., & Girdzijauskas, S. (2006). A hybrid SOM-Altman model for bankruptcy prediction. In V. N. Alexandrov, G. D. van Albada, P. M. A. Sloot, & J. Dongarra (Eds.), *Computational science — ICCS 2006: 6th International Conference* (pp. 364–371). https://doi.org/10.1007/11758549_53
53. Min, J. H., & Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications, 28*(4), 603–614. https://doi.org/10.1016/j.eswa.2004.12.008
54. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems, 50*(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006
55. 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
56. Palmrose, Z.-V. (1987). Litigation and independent auditors: The role of business failures and management fraud. *AUDITING: A Journal of Practice & Theory, 6*(2), 90–103.
57. Palmrose, Z.-V. (1988). 1987 competitive manuscript co-winner: An analysis of auditor litigation and audit service quality. *The Accounting Review, 63*(1), 55–73. Retrieved from https://www.jstor.org/stable/247679
58. Palmrose, Z.-V., & Scholz, S. (2004). The circumstances and legal consequences of non-GAAP reporting: Evidence from restatements. *Contemporary Accounting Research, 21*(1), 139–180. https://doi.org/10.1506/WBF9-Y69X-L4DX-JMV1
59. Perols, J. L., Bowen, R. M., Zimmermann, C., & Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review, 92*(2), 221–245. https://doi.org/10.2308/accr-51562
60. Pierre, K. S., & Anderson, J. A. (1984). An analysis of the factors associated with lawsuits against public accountants. *The Accounting Review, 59*(2), 242–263. Retrieved from https://www.jstor.org/stable/247297
61. Questier, F., Put, R., Coomans, D., Walczak, B., & Heyden, Y. V. (2005). The use of CART and multivariate regression trees for supervised and unsupervised feature selection. *Chemometrics and Intelligent Laboratory Systems, 76*(1), 45–54. https://doi.org/10.1016/j.chemolab.2004.09.003
62. Quinlan, J. R. (1986). Introduction of decision trees. *Machine Learning, 1*, 81–106. https://doi.org/10.1007/BF00116251
63. Ravisankar, P., Ravi, V., Raghava Rao, G., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems, 50*(2), 491–500. https://doi.org/10.1016/j.dss.2010.11.006
64. Reynolds, J. K., & Francis, J. R. (2000). Does size matter? The influence of large clients on office-level auditor reporting decisions. *Journal of Accounting and Economics, 30*(3), 375–400. https://doi.org/10.1016/S0165-4101(01)00010-6
65. Schmidt, J. J. (2012). Perceived auditor independence and audit litigation: The role of nonaudit services fees. *The Accounting Review, 87*(3), 1033–1065. https://doi.org/10.2308/accr-10217
66. Schwartz, K. B., & Soo, B. S. (1996). Evidence of regulatory noncompliance with SEC disclosure rules on auditor changes. *The Accounting Review, 71*(4), 555–572. Retrieved from https://www.jstor.org/stable/248571
67. Sheppard, J. P. (1994). Strategy and bankruptcy: An exploration in to organizational death. *Journal of Management, 20*(4), 795–833. https://doi.org/10.1177/014920639402000406
68. Slowinski, R., & Zopounidis, C. (1995). Application of the rough set approach to evaluation of bankruptcy risk. *Intelligent Systems in Accounting, Finance and Management, 4*(1), 27–41. https://doi.org/10.1002/j.1099-1174.1995.tb00078.x
69. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002
70. Solomon, I., Shields, M. D., & Whittington, O. R. (1999). What do industry-specialist auditors know? *Journal of Accounting Research, 37*(1), 191–208. https://doi.org/10.2307/2491403
71. Stice, J. D. (1991). Using financial and market information to identify pre-engagement factors associated with lawsuits against auditors. *The Accounting Review, 66*(3), 516–533. Retrieved from https://www.jstor.org/stable/247807
72. Sugumaran, V., Muralidharan, V., & Ramachandran, K. I. (2007). Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing, 21*(2), 930–942. https://doi.org/10.1016/j.ymssp.2006.05.004
73. Summers, S. L., & Sweeney, J. T. (1998). Fraudulently misstated financial statements and insider trading: An empirical analysis. *The Accounting Review 73*(1), 131–146. Retrieved from https://www.jstor.org/stable/248345
74. Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure prediction. *Management Science, 38*(7), 926–947. https://doi.org/10.1287/mnsc.38.7.926

75. Titman, S., & Trueman, B. (1986). Information quality and the valuation of new issues. *Journal of Accounting and Economics, 8*(2), 159–172. https://doi.org/10.1016/0165-4101(86)90016-9
76. Tsai, C.-F., & Chen, M.-L. (2010). Credit rating by hybrid machine learning techniques. *Applied Soft Computing, 10*(2), 374–380. https://doi.org/10.1016/j.asoc.2009.08.003
77. Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications, 34*(4), 2639–2649. https://doi.org/10.1016/j.eswa.2007.05.019
78. Tsai, C.-F., Lin, Y.-C., Yen, D. C., & Chen, Y.-M. (2011). Predicting stock returns by classifier ensembles. *Applied Soft Computing, 11*(2), 2452–2459. https://doi.org/10.1016/j.asoc.2010.10.001
79. Tsai, C.-F., Lu, Y.-H., & Yen, D. C. (2012). Determinants of intangible assets value: The data mining approach. *Knowledge-Based Systems, 31*, 67–77. https://doi.org/10.1016/j.knosys.2012.02.007
80. Varetto, F. (1998). Genetic algorithms applications in the analysis of insolvency risk. *Journal of Banking & Finance, 22*(10–11), 1421–1439. https://doi.org/10.1016/S0378-4266(98)00059-4
81. Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. *Decision Support Systems, 50*(3), 570–575. https://doi.org/10.1016/j.dss.2010.08.007