Decision Making: Applications in Management and Engineering
Vol. 6, Issue 2, 2023, pp. 722-733.
ISSN: 2560-6018
eISSN: 2620-0104
DOI: https://doi.org/10.31181/dmame622023785

ENSEMBLE LEARNING ALGORITHM - RESEARCH ANALYSIS ON THE MANAGEMENT OF FINANCIAL FRAUD AND VIOLATION IN LISTED COMPANIES

Weihong Li[1] and Xiujuan Xu[1*]

[1] Jiangxi University of Engineering, Xinyu, Jiangxi 338000, China

Received: 5 May 2023;
Accepted: 25 July 2023;
Available online: 31 July 2023.

Original scientific paper

Abstract: *In recent years, despite the strict "zero tolerance" crackdown on financial fraud and violations by listed companies, there has been a continued exposure of cases involving financial fraud, revenue and profit overstatement, and suspected fraud. This study first established a financial fraud index system and used the XGBoost algorithm to construct a prediction model for financial fraud and violations in listed companies. The indicators were selected and inputted into the model. A dataset was obtained for the experiments. The XGBoost algorithm was compared to two other algorithms. The receiver operating characteristic (ROC) curves showed that the XGBoost algorithm had the best prediction performance among the three algorithms. It was found that the precision of the XGBoost algorithm was 93.17%, the recall rate was 92.23%, the F_1 value was 0.9270, and the area under the curve was 0.90. These results indicated better performance compared to the prediction models based on the Gradient Boosted Decision Tree (GBDT) algorithm and the Logistic algorithm. Considering the data from various evaluation indicators, it is found that the XGBoost algorithm produces the most accurate predictive effect for the financial fraud and violation prediction model.*

Key words: *Ensemble algorithm, listed companies, financial fraud and violation, XGBoost.*

1. Introduction

The act of financial fraud by listed companies is closely related to their own interests. Some listed companies falsify financial information or conceal the financial information that should be disclosed by manipulating their financial statements. These practices can lead to the violation of investors' rights and erode their trust in the financial system (Aslan, 2021). How to predict and manage the occurrence of financial fraud in listed companies in advance has become a crucial research topic.

* Corresponding author.
E-mail addresses: ehl15l@126.com (W. Li), juai2861@yeah.net (X. J. Xu)

The traditional method of detecting financial fraud involves manually reviewing financial statements. However, this approach not only requires reviewers to have extensive experience but also has low efficiency. With the advancement of computer performance and technology, machine learning algorithms have been increasingly utilized in various domains, including the detection of financial fraud. Machine learning algorithms can extract the relevant judgment rules from the training samples and utilize the rules to identify other financial statements. Some studies related to financial fraud are as follows. Adnovaldi and Wibowo (2019) analyzed fraudulent behavior in financial statements using the fraud diamond theory. They found that only external pressure variables, such as leverage, and industry variables, such as obsolete inventory accounts, significantly influenced the detection of potential fraudulent financial statements. Triyanto (2019) analyzed the false financial statements of listed companies in the food and beverage manufacturing industry in Indonesia. The study found that variables such as pressure, opportunity, rationalization, ability, and arrogance did not have a simultaneous effect on fraudulent financial statements. Irawan et al. (2019) proposed identifying the likelihood of financial statement fraud through earnings management and measured it using the F-score indicator. The research results showed that changes in financial objectives and the financial stability of the auditor had a significant positive impact on financial statement fraud. Ardhiansyah et al. (2019) conducted a study on all manufacturing companies listed on the Indonesia Stock Exchange (BEI) and discovered that financial distress, liquidity, leverage, and corporate governance have a significant impact on financial statement fraud. Li (2020) utilized a Back-Propagation Neural Network (BPNN) to develop a financial identification model for detecting instances of financial fraud in publicly traded companies. Wu et al. (2022) established a knowledge graph of audit information based on the relationships among enterprises, audit firms, and auditors. They also proposed a framework based on sub-feature extraction. They found that potential financial fraud companies could be well identified by analyzing the audit data and searching for known financial fraud companies. A variety of algorithms have been used to identify financial fraud in the aforementioned studies. However, all of these algorithms have utilized single recognition algorithms. The accuracy of a single recognition algorithm is limited, even after extensive training. Additionally, using an extensive number of filtered training samples not only increases the training workload but can also lead to overfitting problems. Therefore, to improve the accuracy of the algorithm, this paper combines several algorithms to create an ensemble algorithm. The XGBoost algorithm used in this paper is an ensemble algorithm. The purpose of this paper is to construct a financial fraud prediction model for listed companies using ensemble learning algorithms. This model will enable relevant departments to forecast the risk of financial fraud and implement appropriate preventive measures in a timely manner. Therefore, this article constructs a financial fraud indicator system for listed companies by reviewing the literature and uses XGBoost as an ensemble learning algorithm to construct a financial fraud prediction model. The final indicators were selected based on their features for training the model, and the validity of the model was verified. It is expected that the results of this paper can solve the problem of predicting financial fraud in listed companies and lay a theoretical foundation for managing financial fraud violations in listed companies using ensemble learning algorithms.

2. Construction of Financial Fraud Indicator System

Currently, many scholars have conducted targeted research on effective indicators for managing financial fraud violations. This paper analyzed validated indicators for managing financial fraud violations and preliminarily determined the indicators for financial fraud as solvency, operational capacity, profitability, cash flow, development capacity, and non-financial variables. These five aspects were used to determine whether a listed company had committed financial fraud (Chen et al., 2020). At the same time, in order to avoid missing important indicators, as many indicators of financial fraud violations as possible were collected during the initial selection stage. Afterward, all indicators were screened to narrow down the scope. Therefore, this paper selected a total of 33 indicators. The specific names and definitions of these indicators are shown in Table 1.

Table 1 Initial financial falsification indicator system

Indicator type	Name of indicator	Definition and interpretation of indicators
Debt solvency	X1 Current ratio	Company current assets/company current liabilities
	X2 Quick ratio	(Company's current assets - inventories)/company's current liabilities
	X3 Asset-liability ratio	Total company liabilities/total company assets
	X4 Long-term debt to total assets ratio	Amount of long-term loans/total company assets
	X5 Equity ratio	Total company liabilities/total owner's equity
	X6 Tangible net worth to debt ratio	Total company liabilities/total tangible net assets
Operating capability	X7 Accounts receivable turnover to revenue ratio	Accounts receivable/operating revenue
	X8 Inventory turnover ratio	Operating costs/inventory ending balance
	X9 Accounts receivable turnover ratio	Operating income/average accounts receivable
	X10 Total asset turnover ratio	Operating income/average total assets
	X11 Current asset turnover ratio	Operating income/closing balance of current assets
	X12 Non-current asset turnover ratio	Total operating income/non-current assets ending balance
Profitability	X13 Return on equity	Net income/shareholders' equity balance
	X14 Return on invested capital	(Net profit + finance charges)/(assets + liabilities)
	X15 Gross profit margin	(Operating revenues - operating costs)/total operating revenues
	X16 Sales expense	(Operating + administrative + financial)

	ratio	expenses/total operating income
	X17 Return on investment	Current investment income/(long-term ending value + held-to-close value)
	X18 Return on assets	(Total profit + finance costs)/total assets
Cash flow	X19 Cash flow coverage ratio	Net cash flow from operating activities/average current liabilities
	X20 Cash coverage ratio	Net operating cash flow/net operating profit
	X21 Cash return on assets ratio	Net cash flow from operating activities/total assets
	X22 Cash flow per share	Net cash flow/number of common shares outstanding
	X23 Cash content of operating revenue	Cash received from sales of goods and services/operating income
	X24 Cash recovery ratio	Net cash flows from operating activities/total assets ending balance
Development capability	X25 Prime operating revenue growth rate	(Current year - last year's prime operating revenue)/last year's prime operating revenue
	X26 Total asset growth rate	Closing value of assets - opening value of assets/opening value of assets
	X27 Asset preservation and appreciation rate	Shareholders' equity at year-end/shareholders' equity at the beginning of the year
	X28 Growth rate of return on equity	(Current year - last year's return on net assets)/last year's return on net assets
	X29 Growth rate of earnings per share	(Current year - last year's earnings per share)/last year's earnings per share
	X30 Owner's equity growth rate	(Current year - last year's owner's equity)/last year's owner's equity
Non-financial variables	X31 Number of litigation cases	Total number of cases involving the company's lawsuits from 2020 to the study period
	X32 Board size	Total number of board members at year-end
	X33 Industry concentration ratio	Herfindahl-Hirschman Index (HHI) ∈ (0,1); the closer the HHI value is to 0, the more competitive the industry is.

3. Research Model

3.1. XGBoost prediction model construction

Ensemble learning algorithms can be understood as model frameworks composed of multiple machine learning algorithms. This makes the training process of the prediction model complex, but it also leads to higher accuracy in model predictions. Boosting algorithms is a common implementation of ensemble learning (An et al., 2021), with the AdaBoost algorithm (Zhang et al., 2019) and the XGBoost algorithm (Zúiga & Jesús, 2020) being the most frequently employed algorithms. In

this study, the XGBoost algorithm was used to construct a model for predicting financial fraud. The training of the model is to adjust the parameters of each financial indicator under supervised conditions so that it can accurately predict companies that engage in financial fraud. The XGBoost algorithm is an ensemble machine learning algorithm based on decision trees. It sequentially builds classification and regression trees to minimize the prediction error and approach the true value as closely as possible. The output of the final model result is:

$$\widehat{y_i} = \sum_{M=1}^{M} f_m(x_i). \quad (1)$$

To generate a good tree at each step, an objective function is needed. The definition of the objective function is:

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \ \widehat{y_i}) + \sum_{M=1}^{M} \Omega(f_m), \quad (2)$$

where $l(y_i, \ \widehat{y_i})$ represents the loss function, which calculates the prediction accuracy based on the difference between the actual and fitted values, $\Omega(f_m)$ represents the regularization function, and y_i represents the true result of the sample. To reduce the complexity of the model and avoid overfitting, a regularization term is added to the objective function:

$$\Omega(f_m) = \gamma T + \frac{1}{2}\lambda \sum_{T=1}^{T} w^2 \quad , \quad (3)$$

where γ stands for the regularization coefficient, T stands for the number of leaf nodes, and w stands for the number of outputs of leaf nodes.

During the training process of the model, the model parameters need to be adjusted according to the training results obtained in order to find the optimal prediction model. The final parameters of the XGBoost prediction model are as follows: the learning rate is 0.01, the number of iterations is 200, the maximum sub-tree depth is 3, and the sub-node weight threshold is 1.

3.2. Model evaluation indicators

3.2.1. Confusion matrix

The final experimental result studied in this paper is the occurrence of financial fraud and violations in listed companies within the dataset. These occurrences can be categorized as either into financial fraud or non-fraud. Therefore, the confusion matrix, which is the most commonly used metric for evaluating model performance in binary classification problems, was selected to assess the prediction performance of the model. Precision (P), recall rate (R), and F1 score were selected as the primary evaluation indicators from the confusion matrix (Sari et al., 2019). The specific calculation formulas are as follows:

$$P = \frac{TP}{TP+FP}, \quad (4)$$

$$R = \frac{TP}{TP+FN}, \quad (5)$$

$$F_1 = \frac{P*R}{P+R} * 2, \quad (6)$$

where TP denotes the number of samples of financial fraud correctly identified as fraud, FP indicates the number of samples of non-financial fraud wrongly identified

as fraud, and FN stands for the number of samples of financial fraud wrongly identified as non-fraud.

3.2.2. ROC and AUC

The receiver operating characteristic (ROC) curve (Westland, 2020) is a graphical representation of the true positive rate plotted against the false positive rate. The diagonal line Y=X (x is positive) is the boundary. Above the diagonal line, it means that the model has correctly identified more positive samples than it has incorrectly identified negative samples. The more the curve tilts towards the upper left corner, the higher the actual prediction accuracy of the model for companies with financial fraud, and the lower the prediction error rate for companies without financial fraud. Moreover, the steeper the ROC curve, the better the model's discrimination performance. Another evaluation indicator, the area under the curve (AUC), represents the area enclosed by the coordinate axis under the ROC curve. Its value is 0.5-1. If the value of the AUC approaches 1, it indicates that the model has better predictive performance.

4. Experimental Analysis

4.1. Data Acquisition and Preprocessing

The financial fraud and violation data in the dataset were obtained from the China Listed Companies Financial Annual Report Database in the China Stock Market & Accounting Research Database, as well as the violation information table in the violation event database. Data from companies that violated regulations by "fabricating profits," "falsely recording assets," and "making false statements" during the violation year were selected as financial fraud and violation data. After obtaining the data, integration processing was required due to the large amount of messy data, such as missing value treatment. In the process of data integration, it was found that some data were missing. To address this, data from other years of the same company were used to fill in the missing values. Regarding the issue of unbalanced data, there is a significant disparity between the number of companies with and without financial fraud in the dataset, with a ratio of 1:50, so the undersampling method was used to randomly select and reduce the impact of companies without financial fraud on the analysis. The processed data was separated into training and test sets in a ratio of 7:3, as displayed in Table 2. The training and test sets in Table 2 contained the financial annual statements of the same enterprise for different periods.

Table 2. Distribution of the experimental data set

Dataset	Total number	Financial fraud	No financial fraud	Percentage of companies with financial fraud
Training set	6055	1211	4844	20.00%
Test set	2595	505	2090	19.46%
Total	8650	1716	6934	19.84%

4.2. Mathematical Statistics

The collected data was statistically analyzed using SPSS software, and then the indicators were screened. The correlation coefficient and population stability index (PSI) value were used to measure measuring indicators. The calculation formulas for these measures are:

$$\begin{cases} r_s = 1 - \dfrac{6\sum d_i^2}{n(n^2-1)} \\ PSI = \sum_i (p_{target}^i - p_{base}^i) \cdot \ln\left(\dfrac{p_{target}^i}{p_{base}^i}\right) \end{cases}, \tag{7}$$

where r_s stands for the correlation coefficient, PSI stands for the stability of features, d_i stands for the difference between the sequences of two feature data, n is the number of features, p_{target}^i is the percentage of the i-th sample in the target data set, p_{base}^i is the percentage of the i-th sample in the basic data set.

4.3. Results Analysis

Before the experiment, the indicators were screened. In the previous section, the financial indicator system for managing financial fraud and violations of listed companies was preliminarily organized and constructed. However, including too many indicators in the initial selection can easily lead to inconvenience in the subsequent collection and integration of data information. This can result in a decrease in the training accuracy of the prediction model and an increase in computation time. Therefore, the financial indicators were screened first.

Table 3. Financial fraud indicator screening

Indicator type	Name of indicator	Correlation coefficient	Population stability index
Debt solvency	X1 Current ratio	0.7512	0.1482
	X2 Quick ratio	0.7431	0.6197
	X3 Asset-liability ratio	**0.5943**	**0.0754**
	X4 Long-term debt to total assets ratio	0.8034	0.9413
	X5 Equity ratio	0.7921	0.5132
	X6 Tangible net worth to debt ratio	0.7785	0.3349
Operating capability	X7 Accounts receivable turnover to revenue ratio	**0.6983**	0.4713
	X8 Inventory turnover ratio	**0.6284**	**0.0610**
	X9 Accounts receivable turnover ratio	**0.6529**	**0.0993**
	X10 Total asset turnover ratio	0.7021	0.3135
	X11 Current asset turnover ratio	0.7526	0.1644

	X12 Non-current asset turnover ratio	0.7784	0.1853
Profitability	X13 Return on equity	0.7956	0.5712
	X14 Return on invested capital	0.7413	0.1768
	X15 Gross profit margin	**0.5842**	**0.0456**
	X16 Sales expense ratio	0.7423	**0.0428**
	X17 Return on investment	0.7619	0.2678
	X18 Return on assets	0.8426	0.1947
Cash flow	X19 Cash flow coverage ratio	**0.6159**	**0.0651**
	X20 Cash coverage ratio	0.7099	0.1685
	X21 Cash return on assets ratio	0.7421	0.5512
	X22 Cash flow per share	0.7563	0.1972
	X23 Cash content of operating revenue	0.7951	0.1639
	X24 Cash recovery ratio	0.8127	0.2271
Development capability	X25 Prime operating revenue growth rate	0.7139	0.1734
	X26 Total asset growth rate	**0.5753**	**0.0792**
	X27 Asset preservation and appreciation rate	0.7864	0.6251
	X28 Growth rate of return on equity	0.7954	0.1577
	X29 Growth rate of earnings per share	0.7763	0.4911
	X30 Owner's equity growth rate	0.7226	0.1560
Non-financial variables	X31 Number of litigation cases	**0.3428**	**0.0196**
	X32 Board size	**0.4211**	0.1673
	X33 Industry concentration ratio	**0.4673**	**0.0914**

During the process of selecting indicator features, in addition to calculating the correlation value (Sihombing & Cahyadi, 2021), the population stability index (PSI) value (Udhayakumar et al., 2021) was also calculated. The stability of the model is primarily determined by the stability of its features. Unstable features can lead to overfitting problems in the model. The financial fraud indicators that had a correlation greater than 0.7 and a PSI value greater than 0.1 were deleted. After the calculation, eight indicators met the stability criterion, and ten indicators met the correlation criterion. These values were presented in bold in Table 3. After integrating these bold values, it was found that only eight financial fraud indicators met all the criteria. Therefore, the following eight indicators were selected based on their significance, correlation, and stability: asset-liability ratio (X3), inventory turnover ratio (X8), accounts receivable turnover ratio (X9), gross profit margin (X15), cash flow coverage ratio (X19), total asset growth rate (X26), number of litigation cases (X31), and industry concentration ratio (X33). These indicators were then inputted into the model to analyze the dataset, and the results are presented in Figure 1.

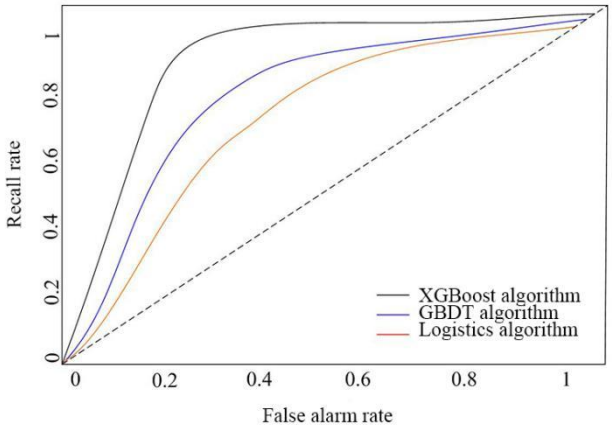

Figure 1. The ROC curve of different algorithms.

The ROC curves for different algorithms were collected and are shown in Figure 1. According to the evaluation criteria, the higher the tile of ROC curve toward the upper left corner, the greater the probability of accurately predicting financial fraud in listed companies. Additionally, a steeper ROC curve indicates better discrimination performance of the model. It is evident from Figure 1 that the ROC curve of the XGBoost algorithm is highest in the upper left corner. This indicates that the XGBoost algorithm has the highest probability of correctly predicting cases of financial fraud in listed companies. Additionally, the steepness of the curve is also the highest, which demonstrates that the model discrimination performance constructed by this algorithm is optimal.

Table 4. Experimental results comparison of different algorithms

Model category	Precision	Recall rate	F_1 score	AUC
XGBoost algorithm	93.17%	92.23%	0.9270	0.90
Gradient boosted decision tree algorithm (Bai et al., 2022)	87.27%	89.63%	0.8794	0.86
Logistic algorithm (Liu, 2021)	80.33%	81.39%	0.8086	0.81

The experimental results of the XGBoost algorithm were compared with those of the Gradient Boosted Decision Tree (GBDT) and Logistic algorithms, and the results are presented in Table 4. The precision, recall rate, F_1 score, and AUC value of the GBDT algorithm were 87.27%, 89.63%, 0.8794, and 0.86, respectively. The precision, recall rate, F_1 score, and AUC value of the Logistic algorithm were 80.33%, 81.39%, 0.8086, and 0.81, respectively. The XGBoost algorithm demonstrated a precision of 93.17%, a recall rate of 92.23%, a F_1 score of 0.9270, and an AUC value of 0.90. These results indicate that it outperformed both the GBDT and Logistic algorithms. The closer the F_1 value and AUC value are to 1, the better the model's prediction effect. Therefore, it can be concluded that the financial fraud and violation prediction

model constructed using the XGBoost algorithm has excellent performance. The comparison of numerical values for various evaluation indicators among the three algorithms demonstrated that the XGBoost algorithm performed the best in constructing a financial fraud and violation prediction model for listed companies.

5. Discussion

Listed companies are required to release their financial statement. Investors will then decide whether to continue investing or divest based on the information provided in these statements. Under pressure from investors and internal business demands, listed companies may choose to falsify their financial statements. Once there is falsified data in the financial statement, whether it is manipulated to show positive or negative results, it will mislead investors and ultimately impact the growth of the market economy. Therefore, it is necessary to accurately determine the authenticity of financial statements. The traditional manual audit is demanding on auditors and has low efficiency. With the development of computer power and technology, machine learning algorithms are being used to assist in the identification of the accuracy of the financial statements. Machine learning algorithms will initially analyze the correlation between each index in the financial statement and the authenticity of the financial statement using training samples. Subsequently, these algorithms will use the discovered patterns to determine the authenticity of financial statements whose status is unknown. The XGBoost algorithm used in this paper is an ensemble machine learning algorithm that combines multiple classifiers with weak classification performance. In the subsequent experimental analysis, the index system for financial fraud analysis was first constructed. Then, the indexes were screened using correlation coefficient and PSI. Eight indices were selected for financial fraud analysis. Then, the data from the eight indicators was used to train the XGBoost algorithms. Additionally, a comparison was also made with the GBDT and Logistic algorithms. The final results are shown above. The results revealed that the XGBoost algorithm had better performance in detecting financial fraud compared to the other two algorithms. The reasons are analyzed. The Logistic algorithm is a traditional regression analysis method that is commonly used in the detection of financial fraud due to its simple principles. Regression analysis can be used to explain the relationship between different indicators and financial fraud. However, this algorithm is more suitable for analyzing low-dimensional data. In the face of financial indicators, such as high-dimensional data, the computational difficulty will be significantly increased. The GBDT algorithm is a gradient boosting decision tree algorithm, which is also an ensemble algorithm. The advantage of the algorithm lies in its fast recognition, and although its accuracy is not low, this algorithm is prone to overfitting due to the association and integration of weak classifiers. The XGBoost algorithm adopted in this paper will use the regular term as a constraint to prevent overfitting during the training process.

In this paper, the XGBoost algorithm was used to identify financial fraud, providing an effective reference for maintaining the stability of the market economy. In the face of financial fraud, managers can introduce machine learning algorithms to preliminarily screen many financial statements. This can improve supervisory efficiency, enhance the status of auditing institutions, perfect the supervision and punishment policies, and attempt to solve the problem of enterprise financing difficulties. Securities supervision agencies should also pay attention to the involvement of listed companies in litigation.

5. Conclusion

This article mainly introduces the financial fraud index system and the XGBoost algorithm. By conducting a thorough review of the relevant literature, this paper constructed a financial fraud index system. The XGBoost algorithm was then employed to construct a prediction model for financial fraud and violation in listed companies. The most significant indicators were selected to be input into the model after the indicator screening. The results of the XGBoost algorithm were then compared with the results of the models constructed by the other two algorithms. The ROC curves showed that the XGBoost algorithm had the steepest curve among the three algorithms, indicating that it had the best detection performance. Moreover, the precision of the XGBoost algorithm was 93.17%, the recall rate was 92.23%, the F_1 value was 0.9270, and the AUC value was 0.90. The XGBoost algorithm showed signs of outperforming the GBDT and Logistic algorithms. This paper demonstrates the effectiveness and feasibility of using the XGBoost algorithm in ensemble learning to develop a financial fraud and violation prediction model for listed companies.

Author Contributions: Conceptualization, W.L. and X.X.; methodology, W.L.; software, X.X.; validation, W.L.; formal analysis, W.L.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, W.L.; writing—review and editing, X.X.; visualization, W.L.; supervision, W.L.; project administration, W.L. Both authors have read and agreed to the published version of the manuscript.

Funding*:* This research received no external funding.

Data Availability Statement*:* The data generated in this paper are available from the corresponding author.

Acknowledgments: None.

Conflicts of Interest*:* The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adnovaldi, Y., & Wibowo, W. (2019). Analisis determinan fraud diamond terhadap deteksi fraudulent financial statement. Jurnal Informasi Perpajakan, Akuntansi dan Keuangan Publik, 14(2), 125-146. https://doi.org/10.25105/jipak.v14i2.5195

An, X., Hu, C., Liu, G., & Lin, H. (2021). Distributed online gradient boosting on data stream over multi-agent networks. Signal Processing: The Official Publication of the European Association for Signal Processing (EURASIP), 189(4), 108253. https://doi.org/10.1016/j.sigpro.2021.108253

Ardhiansyah, A. S., Kusuma, H., & Sa'Dani, O. S. (2019). Analisa pengaruh kinerja keuangan dan corporate governance terhadap kemungkinan terjadinya financial statement fraud. Jurnal REKSA: Rekayasa Keuangan, Syariah, dan Audit, 6(2), 149-165. https://doi.org/10.12928/j.reksa.v6i1.1375

Aslan, L. (2021). Financial statement fraud in the turkish financial services sector. Istanbul Business Research, 50(2), 385-409. https://doi.org/10.26650/ibr.2021.50.844527

Bai, M., Zheng, Y., & Shen, Y. (2022). Gradient boosting survival tree with applications in credit scoring. Journal of the Operational Research Society, 73(1), 39-55. https://doi.org/10.1080/01605682.2021.1919035

Chen, L., Xiu, B., & Ding, Z. (2020). Finding misstatement accounts in financial statements through ontology reasoning. IEEE Access, 1-14. https://doi.org/10.1109/ACCESS.2020.3014620

Irawan, P. A., Susilowati, D., & Puspasari, N. (2019). Detection analysis on fraudulent financial reporting using fraud score model. SAR (Soedirman Accounting Review): Journal of Accounting and Business, 4(2), 161-180. https://doi.org/10.20884/1.sar.2019.4.2.2467

Li, S. L. (2020). Data mining of corporate financial fraud based on neural network model. Computer Optics, 44(4), 665-670. DOI: 10.18287/2412-6179-CO-656

Liu, X. (2021). Empirical analysis of financial statement fraud of listed companies based on logistic regression and random forest algorithm. Journal of Mathematics, 2021(2), 1-9. https://doi.org/10.1155/2021/9241338

Sari, N. S., Sofyan, A., & Fastaqlaili, N. (2019). Analysis of fraud diamond dimension in detecting financial statement fraud. Jurnal Akuntansi Trisakti, 5(2), 171-182. https://doi.org/10.25105/jat.v5i2.4861

Sihombing, T., & Cahyadi, C. C. (2021). The effect of fraud diamond on fraudulent financial statement in asia pacific companies. Jurnal ULTIMA Accounting, 13(1), 143-155. https://doi.org/10.31937/akuntansi.v13i1.2031

Triyanto, D. N. (2019). Fraudulence financial statements analysis using pentagon fraud approach. Journal of Accounting Auditing and Business, 2(2), 26. https://doi.org/10.24198/jaab.v2i2.22641

Udhayakumar, K., Rakkiyappan, R., Li, X., & Cao, J. (2021). Mutiple psi-type stability of fractional-order quaternion-valued neural networks. Applied Mathematics and Computation, 401, 126092. https://doi.org/10.1016/j.amc.2021.126092

Westland, J. C. (2020). Predicting credit card fraud with Sarbanes-Oxley assessments and Fama-French risk factors. Intelligent Systems in Accounting, Finance & Management, 27(2), 95-107. https://doi.org/10.1002/isaf.1472

Wu, H., Chang, Y., Li, J., & Zhu, X. (2022). Financial fraud risk analysis based on audit information knowledge graph. Procedia Computer Science, 199, 780-787. https://doi.org/10.1016/j.procs.2022.01.097

Zhang, Z., Qiu, J. X., & Dai, W. (2019). A new improved boosting for imbalanced data classification. IOP Conference Series: Materials Science and Engineering, 533, 012047. https://doi.org/10.1088/1757-899X/533/1/012047

Zúiga, E., & Jesús, J. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. Ingeniería Investigación y Tecnología, 21(3), 1-16. https://doi.org/10.22201/FI.25940732E.2020.21.3.022