EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

# Financial Fraud: Identifying Corporate Tax Report Fraud Under the Xgboost Algorithm

Xianjuan Li[1,*]

[1]Hunan City University, Yiyang, Hunan 413000, China

## Abstract

INTRODUCTION: With the development of economy, the phenomenon of financial fraud has become more and more frequent.
OBJECTIVES: This paper aims to study the identification of corporate tax report falsification.
METHODS: Firstly, financial fraud was briefly introduced; then, samples were selected from CSMAR database, 18 indicators related to fraud were selected from corporate tax reports, and 13 indicators were retained after information screening; finally, the XGBoost algorithm was used to recognize tax report falsification.
RESULTS: The XGBoost algorithm had the highest accuracy rate (94.55%) when identifying corporate tax statement falsification, and the accuracy of the other algorithms such as the Logistic regressive algorithm were below 90%; the F1 value of the XGBoost algorithm was also high, reaching 90.1%; it also had the shortest running time (55 s).
CONCLUSION: The results prove the reliability of the XGBoost algorithm in the identification of corporate tax report falsification. It can be applied in practice.

**Keywords:** financial fraud, corporate tax, falsification identification, XGBoost algorithm.

Received on 14 February 2023, accepted on 18 April 2023, published on 05 May 2023



doi: 10.4108/eetsis.v10i3.3033

## 1. Introduction

With the continuous development of the capital market [1], the development of enterprises has also been subject to great challenges. In order to further regulate the capital market, its supervision has become more and more strict [2]. However, there are still many enterprises that take the risk and choose financial fraud driven by profit [3]. Financial fraud refers to the act of intentionally fabricating false financial reports to make improper profits to mislead information users, which is a great threat to economic security [4]. Enterprises obtain illegal benefits by falsifying their income and expenses, which not only seriously affects the judgment of investors but also is detrimental to the stability of the capital market [5]. Therefore, if financial fraud can be identified in advance through some methods, it is of great importance to effectively avoid risks and maintain market order. From a machine learning perspective, the identification of financial fraud is a binary classification problem [6]. With the development of computer technology, artificial intelligence, machine learning and other methods have been widely used in the financial field [7], which can effectively detect and predict fraud [8]. There are also many studies on text mining [9], which focus on mining clues from textual information such as annual reports and announcements of enterprises. At the same time, the study of financial fraud from interdisciplinary perspectives, such as psychology and sociology, has become a new way of thinking. Houssou et al. [10] studied the situation of data imbalance in financial fraud prediction and analyzed fraud prediction using homogeneous and non-homogeneous Poisson processes. They found through experiments that the method exhibited a superior prediction ability than a baseline approach. Swa

*Corresponding author. Email: juanwei3312@yeah.net

et al. [11] designed a knowledge graph (KG) framework and then analyzed the performance of four machine learning algorithms on fraud detection and found that support vector machine (SVM) performed best on the test set. Akra et al. [12] analyzed the roles of Altman and Beneish models in detecting early profit manipulation and applied them to the Kuwaiti stock market. They found that the Beneish model had good power in predicting possible earnings manipulation or report falsification by firms. Zhou et al. [13] designed a convolutional neural network (CNN)-based method for fraud in supply chain finance and found through tests that the method had high accuracy and recall rate. Burke et al. [14] found that a brief online educational intervention could reduce fraud susceptibility. Davidson [15] analyzed 1805 executives and found that executives suspected of fraud had stronger equity incentives than executives in similar positions in non-fraudulent companies, and that equity incentives for all members of the top management team could be considered when identifying financial fraud. Novatiani et al. [17] analyzed the data of 90 state-owned enterprises by SEM-PLS and found that the effectiveness of the internal audit function could prevent financial statement falsification. The study of financial fraud focuses on how to detect, identify, and prevent companies or individuals who intentionally fabricate false information in their financial reports to deceive investors. For example, some companies may misrepresent their net profits to attract investments or fabricate false accounts to commit financial fraud. The identification of fraudulent corporate tax reports can effectively detect fraudulent behavior of enterprises, thus protecting the interests of investors and the public, which is of great importance to maintain the stability of the market and ensure compliance. This paper used a relatively novel machine learning algorithm, i.e., the XGBoost algorithm, to study financial fraud, screened the falsification identification indicators by the indicator information value (IV), and proved the reliability of the method by comparing it with other methods. The research in this paper provides a new method for identifying fraudulent behavior of enterprises in reports, which can be applied in practice to better detect fraudulent behavior of enterprises and thus promote the stability of capital market.

## 2. Overview of Financial Fraud and Fudging

As the economy grows, the number of frauds occurring in companies is increasing [18]. Table 1 shows a few classic examples of financial fraud.

Table 1. Recent financial fraud cases

| Enterprise | Counterfeiting Overview |
|---|---|
| Kangmei Pharmaceutical | From 2016 to 2018, inflated operating revenues of 29,128 million yuan and inflated monetary funds of 88,681 million yuan |
| Kangde Xin | Inflated profits of 11.531 billion yuan from 2015 to 2018 |
| Zhangzidao | Inflated profits of 130 million yuan in 2016 and 280 million yuan in 2017 |
| LETV | From 2007 to 2016, inflated revenues of 1.872 billion yuan and inflated profits of 1.737 billion yuan |
| Yihua Life | From 2016 to 2019, inflated revenues of 21.121 billion yuan and inflated profits of 2.816 billion yuan |
| Kingenta | From 2015 to 2018, inflated revenues of 23,073 million yuan and inflated profits of 300 million yuan |
| Lonkey | From 2018 to 2019, inflated operating revenues of 12.886 billion yuan and inflated profits of 412 million yuan |

It is seen from the current financial fraud cases that most of the frauds are related to the falsification and beautification of reports and show the following characteristics: 1 the more backward the economic development of the region, the higher the possibility of fraud in enterprises; 2 the more complex the way of fraud, and the more objects of manipulating profits; 3 the motives of fraud are more diversified, and some even involve criminal crimes; 4 the frauds are more difficult to detect and concealed.

The financial fraud and falsification of enterprises will lead to punishment, damage the corporate image, and make the stock fall, which is not conducive to the long-term development of the enterprises. For investors, financial fraud and falsification will lead them to make wrong judgments due to false information and suffer economic losses [19]; for the capital market, the endless fraud and falsification will disrupt the market order and affect investors' investment confidence, which is not conducive to the healthy development of the capital market [20]. Therefore, the identification of financial fraud is very important, not only to help improve the quality of the audit but also to reduce the risk of investment for investors, creditors and other information users, and for regulators, it is also conducive to the anti-fraud work.

## 3. Fraud identification based on the XGBoost algorithm

### 3.1. Data selection and processing

Reports can reflect the economic situation of enterprises [21]; therefore, this paper studied frauds from the tax reports of enterprises. Samples were selected from the

China Stock Market & Accounting Research (CSMAR) database, including 482 fraudulent tax reports of 227 Shanghai and Shenzhen A-share listed companies between 2011 and 2020 and another 482 normal tax reports of 227 non-fraudulent enterprises belonging to the same industry as the fraudulent enterprises in the same period.

For the experimental data, after eliminating invalid, duplicate, and abnormal data, the missing values were filled using the mean value. The data were normalized in order to avoid the errors caused by the index magnitude, and the corresponding formula is:

$$x' = \frac{x-x_{min}}{x_{max}-x_{min}},$$

where x is original data, $x_{max}$ and $x_{min}$ are the maximum and minimum values of original data, and $x'$ is the processed data, normalized to between 0 and 1.

In order to obtain a high identification accuracy, indicators related to fraud were selected from the tax reports. The selection of indicators was considered mainly from the following aspects.

(1) Debt service: In the case of a high proportion of enterprise liabilities, the management of an enterprise may be more inclined to disclose good news and avoid bad news. At this time, there is a possibility of falsifying the reports, including using accounts receivable and inventory to make adjustments, inflating profits, and beautifying the reports. Therefore, the selection of indicators needs to consider the debt service of an enterprise.

(2) Operation: The operation of an enterprise reflects the use and management of its capital. In fraud, profits can be manipulated by reducing the inventory turnover rate and increasing the proportion of inventories. Therefore, the selection of indicators needs to consider the operation of an enterprise.

(3) Profitability: When a company has a poor level of profit, not only will managers' earnings be reduced, but the company's ability to raise capital will also be affected, and the possibility of fraud also exists at this time. Thus, the profit situation of an enterprise should be paid much attention to.

(4) Risk: In a certain period of time, the cost and structural changes of an enterprise directly affect the revenue; therefore, the risk profile of an enterprise also needs to be considered in the selection of indicators.

(5) Cash flow: The cash flow situation of an enterprise is related to its ability to pay, and abnormal changes in the relevant indicators are likely to indicate fraud; therefore, the cash flow situation of an enterprise can be used as one of the indicators for fraud identification.

Based on the above aspects, the falsification identification indicators shown in Table 2 were selected.

Table 2. Indicators for identifying fraudulent reports

| Type | Code | Name | Calculation formula |
|---|---|---|---|
| Debt service | X1 | Current ratio | Current assets/current liabilities |
| | X2 | Quick ratio | (Current assets - inventories/current liabilities) |
| | X3 | Asset-liability ratio | Total liabilities/total assets |
| Operation | X4 | Inventory turnover ratio | Cost of main operations/inventory closing balance |
| | X5 | Total assets turnover ratio | Revenue from main business/average total assets |
| | X6 | Accounts receivable turnover ratio | Revenue from main business/average occupancy of accounts receivable |
| | X7 | Current asset turnover ratio | Revenue from main business/average total current assets |
| | X8 | Ratio of accounts receivable to revenue | Accounts receivable/operating revenue |
| Profit ability | X9 | Return on assets | (Total profit + finance costs)/total average assets |
| | X10 | Ratio of expenses to sales | Selling expenses/operating income |
| | X11 | Net operating margin | Net operating profit/revenue from main business |
| | X12 | Return on net assets | Net income/average shareholders' income |
| | X13 | Earnings per share | Current value of net profit/current closing value of paid-in capital |
| Risk | X14 | Financial leverage | (Total profit + financial cost)/total profit |
| | X15 | Operating leverage | Profit from main business/(total profit + financial cost) |
| Cash flow | X16 | Cash ratio | (Monetary funds + financial assets held for trading)/current liabilities |
| | X17 | Net cash flow per share | Net increase in cash and cash equivalents/total number of shares |
| | X18 | Net cash content of net profit | Net cash flow from operating activities/net income |

The indicators in Table 2 were further screened. The predictive ability of the indicators was determined by calculating the indicator information value (IV). The IV was calculated based on the weight of evidence (WOE). The calculation formula of WOE is:

$$\mathrm{WOE_i} = \ln \frac{f(x_i|X_n)}{f(y_i|Y_n)}, \tag{1}$$

where $f(x_i|X_n)$ refers to the proportion of falsified samples in the current group to the total falsified samples after grouping and $f(y_i|Y_n)$ refers to the proportion of normal samples to the total normal samples after grouping. The larger the WOE value, the greater the number of falsified samples. On this basis, IV is calculated:

$$\mathrm{IV} = \sum\left(\mathrm{f}(\mathrm{x_i}|\mathrm{X_n}) - \mathrm{f}(\mathrm{y_i}|\mathrm{Y_n})\right) \ln\frac{\mathrm{f}(\mathrm{x_i}|\mathrm{X_n})}{\mathrm{f}(\mathrm{y_i}|\mathrm{Y_n})}. \quad (2)$$

The value of IV can reflect the contribution of an indicator to label differentiation. It is generally considered that indicators with IV values below 0.02 do not have valid information. The calculation results of the IV values of the indicators in Table 2 are shown in Table 3.

Table 3. Results of the calculation of the IV value of the indicator

| | Indicator Code | IV value |
|---|---|---|
| Current ratio | X1 | 0.168 |
| Quick ratio | X2 | 0.049 |
| **Asset-liability ratio** | **X3** | **0.011** |
| Inventory turnover ratio | X4 | 0.246 |
| Total assets turnover ratio | X5 | 0.151 |
| Accounts receivable turnover ratio | X6 | 0.239 |
| Current asset turnover ratio | X7 | 0.172 |
| Ratio of accounts receivable to revenue | X8 | 0.082 |
| Return on assets | X9 | 0.179 |
| **Ratio of expenses to sales** | **X10** | **0.001** |
| Net operating margin | X11 | 0.212 |
| Return on net assets | X12 | 0.137 |
| Earnings per share | X13 | 0.156 |
| **Financial leverage** | **X14** | **0.012** |
| **Operating leverage** | **X15** | **0.015** |
| Cash ratio | X16 | 0.132 |
| Net cash flow per share | X17 | 0.146 |
| **Net cash content of net profit** | **X18** | **0.018** |

It is generally considered that indicators with IV values below 0.02 do not have predictive power; therefore, indicators X3, X10, X14, X15, and X18 with IV values less than 0.02 were excluded, and the final indicators used in this paper are shown in Table 4.

Table 4. Screened report falsification identification indicators

| Code | Indicator name |
|---|---|
| X1 | Current ratio |
| X2 | Quick ratio |
| X3 | Inventory turnover ratio |
| X4 | Total assets turnover ratio |
| X5 | Accounts receivable turnover ratio |
| X6 | Current asset turnover ratio |
| X7 | Ratio of accounts receivable to revenue |
| X8 | Return on assets |
| X9 | Net operating margin |
| X10 | Return on net assets |
| X11 | Earnings per share |
| X12 | Cash ratio |
| X13 | Net cash flow per share |

## 3.2. XGBoost recognition algorithm

The XGBoost algorithm is currently a relatively new machine learning algorithm, which is characterized by its ability to handle high latitude, unbalanced and complex data well, effectively avoiding the problem of overfitting, and having high accuracy and efficiency in solving classification and regression problems; therefore, this paper used the XGBoost algorithm for report falsification identification. The XGBoost algorithm is an optimization of the gradient boosted decision tree (GBDT) algorithm [22]. Compared with the GBDT algorithm, the XGBoost algorithm adds a regular term to the objective function, which reduces complexity and also improves efficiency [23]. The objective of the GBDT algorithm is to find every optimal single regression tree ($\varphi_j$), its objective function is written as:

$$\hat{\varphi}_\mathrm{j} = \underset{\varphi_\mathrm{j}}{\mathrm{argmin}}\left\{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{L}\left[\mathrm{y_i}, \hat{\mathrm{y}}_\mathrm{i}^{(\mathrm{j}-1)} + \mathrm{v}\mathrm{f_j}(\mathrm{x_i}; \varphi_\mathrm{i})\right] + \Omega(\varphi_\mathrm{j})\right\}, \quad (3)$$

where $N$ is the sample size, $L(y,\hat{y}) = (\hat{y} - y)^2$, and $\Omega(\varphi_j)$ is the regular term of the $j$-th regression tree:

$$\Omega(\varphi_\mathrm{j}) = \gamma \mathrm{M_j} + \frac{1}{2}\lambda \sum_{\mathrm{k}=1}^{\mathrm{M_j}} \left(\mathrm{w}_\mathrm{k}^{(\mathrm{j})}\right)^2, \quad (4)$$

where $M_j$ is the number of leaf nodes of the j-th regression tree, $\gamma$ represents the minimum loss per additional leaf node branch, $\lambda$ is the regular term, and $w_k^{(j)}$ is the leaf node value of the j-th regression tree.

The objective function of the XGBoost algorithm is written as:

$$\hat{\varphi}_j \approx \underset{\varphi_j}{\mathrm{argmin}}\left\{\sum_{i=1}^{N} L\left[y_i, \hat{y}_i^{(j-1)} + v g_i^{(j)} f_j(x_i; \varphi_i)\right] + \frac{1}{2} v^2 h_i^{(j)} f_j^2(x_i; \varphi_i) + \gamma M_j + \frac{1}{2}\lambda \sum_{k=1}^{M_j} \left(w_k^{(j)}\right)^2\right\}, \quad (5)$$

$$g_i^{(j)} = 2\left(\hat{y}_i^{(j-1)} - y_i\right), \tag{6}$$

$$h_i^{(j)} = 2. \tag{7}$$

After finding every optimal single regression tree, the training of the XGBoost algorithm is completed, and it can be used for the identification of report falsification.

## 4. Results and analysis

There were 964 research samples from 454 enterprises, and there were 13 indicators used for report falsification identification, as shown in Table 4. After data normalization, the XGBoost algorithm was trained by using the ten-fold cross-test method, and then the parameters were optimized by the Harmonica algorithm. The final parameter settings are shown in Table 5.

Table 5. Parameter settings of the XGBoost algorithm

| Parameters | Value |
|---|---|
| learning_rate | 0.05 |
| max_depth | 6 |
| n_esimators | 300 |
| min_chid_weight | 1 |
| sub_sample | 0.9 |
| scale_pos_weight | 1 |
| reg_lambda | 300 |

The effectiveness of the algorithm on tax report falsification identification was evaluated on the basis of the confusion matrix (Table 6), and the relevant indicators are shown in Table 6.

Table 6. Confusion matrix

| | | Real category | |
|---|---|---|---|
| | | True | False |
| Identification result | Positive | True Positive (TP) | False Positive (FP) |
| | Negative | False Negative (FN) | True Negative (TN) |

(1) Accuracy: $\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}}$.

(2) Precision: $\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$.

(3) Recall rate: $\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$.

(4) F1-score: $\text{F1} - \text{score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}}$.

To better understand the performance of the XGBoost algorithm, it was compared with Logistic regressive [24], Support Vector Machine (SVM) [25], Random Forest (RF) [26], and GBDT algorithms. These algorithms were all implemented in SAS/EM software. A binary logistic regression process was used. The dependent variables were 0 and 1, 0 for normal reports and 1 for falsified reports. The independent variables were the 13 indicators filtered in Table 4; the RBF kernel function was used in the SVM algorithm, and the default values were used for the rest of the parameters. The default values were also used for the parameters of the RF algorithm as well as the GBDT algorithm. The parameters in Table 5 were used for the XGBoost algorithm. In addition, the method proposed in this paper was compared with the method proposed by Ma et al. [27]. The performance of these algorithms for report falsification identification is shown in Table 7.

Table 7. Performance comparison of different algorithms on report falsification identification

| | Accuracy/% | Precision/% | Recall rate/% | F1-score/% |
|---|---|---|---|---|
| Logistic regressive | 70.12 | 68.44 | 65.36 | 66.86 |
| SVM | 70.33 | 64.64 | 63.22 | 63.92 |
| RF | 80.07 | 80.56 | 76.54 | 78.50 |
| GBDT | 89.77 | 85.64 | 75.12 | 80.04 |
| The method proposed by Ma et al. [27] | 91.32 | 88.36 | 80.33 | 84.15 |
| XGBoost | 94.55 | 91.26 | 88.97 | 90.10 |

According to Table 7, among the five algorithms, Logistic regressive and SVM algorithms had low accuracy, both around 70%, and the RF algorithm had an accuracy of 80.07%, which was about 10% higher than Logistic regressive and SVM algorithms. The GBDT algorithm had an accuracy of 89.77%. The accuracy of the method proposed by Ma et al. [27] was 91.32%. The XGBoost algorithm was 94.55%, which was 4.78% higher than the GBDT algorithm and 3.23% higher than the method proposed by Ma et al. [27].

Logistic regressive and SVM algorithms also performed poorly in terms of precision and recall rate, resulting in low F1-scores. The F1-score is a combination of precision and recall rate, and the comparison of F1-scores showed that the F1-score of Logistic regressive, SVM, and RF algorithms were below 80%, indicating that these three algorithms performed poorly in report falsification identification. The F1-score of the GBDT algorithm was 80.04%, and the F1-score of the method proposed by Ma et al. [27] was 84.15%. The F1-score of the XGBoost algorithm was 90.1%, which was about 10% higher than the GBDT algorithm and 5.95% higher than the method proposed by Ma et al. [27].

The comparison of the comprehensive performance found that the XGBoost algorithm had better performance in terms of accuracy and precision, indicating that it could identify and classify falsified and normal reports more accurately and thus help determine whether there is fraud in the enterprise.

Finally, the running time of these algorithms was compared, and the results are shown in Figure 1.

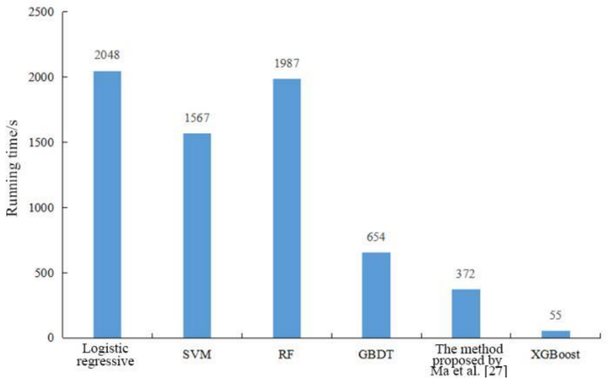


**Figure 1.** Comparison results of the running time of different algorithms

It was found from Figure 1 that the running time of Logistic regressive, SVM, and RF algorithms were long, above 1500 s, while the running time of the GBDT algorithm was 654 s, which was obviously shorter than the first three algorithms. The running time of the method proposed by Ma et al. [27] was 372 s, which was 43.12% shorter than the GBDT algorithm. The running time of the XGBoost algorithm was 55 s in report falsification identification, which was about one-fortieth of the Logistic regressive algorithm. Compared with the GBDT algorithm, the running time of the XGBoost algorithm was improved by 91.6%; compared with the method proposed by Ma et al. [27], it was improved by 85.22%. It was found from Table 7 and Figure 1 that the XGBoost algorithm had not only good recognition performance but also higher recognition efficiency in the identification of corporate tax report falsification, so it can be applied in practice to achieve better and faster identification of financial fraud.

## 5. Discussion

Financial fraud is a common problem in countries around the world and is becoming more prevalent as the economy grows [28]. It often takes a long period of time from the implementation of fraud to its exposure, and during this process, for investors and other stakeholders, the wrong decisions can no longer be changed, and economic losses have long been caused [29]. Moreover, with the development of technology, the means of fraud are becoming more and more diverse and hidden [30], and the amount involved is getting larger and larger, which seriously threatens the stability of the capital market [31]. Therefore, it is of great practical importance for both investors and regulators to identify fraudulent behavior of enterprises in advance [32].

The main means of fraud is the manipulation of report entries [33], such as fictitious profit, fictitious reduction of liabilities, etc. At present, commonly used report falsifications include the following aspects: ①fictitious profit: increase revenue by forging contracts and other means, confirm revenue in advance by taking advantage of time lags, record less costs, or not record some costs; ② fictitious assets: fictitious monetary funds, cash flow, etc., or recognize the potential loss as impairment through asset restructuring or appraisal; ③fictitious liabilities: conceal the liabilities of the enterprise, not record bank debits and repayments, etc.; ④related transactions: change the profit situation through the transfer of assets between the parent company and subsidiaries or related purchases and sales.

Faced with the complex fraudulent means and the increasingly massive data, it is increasingly difficult to recognize fraudulent enterprises. This paper identified whether the reports are fraudulent through the XGBoost algorithm and compared it with the Logistic regressive algorithm and other algorithms. The results of experimental analysis suggested that the XGBoost algorithm was more advantageous in terms of falsification recognition accuracy and running time, proving the usability of this algorithm in the actual enterprise report falsification identification.

However, the research in this paper also has some shortcomings, for example, the research samples were not comprehensive due to the insufficient public data, the report content was lagging, and the performance of the algorithm needs further improvement. Therefore, in the future work, more in-depth research on these content is needed to better improve the report falsification identification method.

## 6. Conclusion

This paper focused on the identification of corporate tax statement falsification. In order to better judge the financial fraud enterprises, this paper selected indicators from corporate tax reports and used the XGBoost algorithm for report falsification identification. Through experimental analysis, it was found that, compared with algorithms such as the Logistic regressive algorithm, the XGBoost algorithm had the highest accuracy in report falsification identification, reaching 94.55%, its F1-score was 90.1%, and its running time was also short, only 55 s, which shows good performance and can be further applied in practice.

## References

[1] Wang D, Lin J, Cui P, Jia Q, Wang Z, Fang Y, Yu Q, Zhou J, Yang S, Qi Y. A Semi-Supervised Graph Attentive Network for Financial Fraud Detection. 2019 IEEE International Conference on Data Mining (ICDM); 2019. p. 598-607.

[2] Cheng C H, Kao Y F, Lin H P. A financial statement fraud model based on synthesized attribute selection and a dataset with missing values and imbalanced classes. Appl. Soft Comput., 2021; 108(3):1-19.
[3] Heneke E, Valentine R, Jourdan Z. Predictive Factors in Financial Fraud and Malfeasance from 1950-2018. J. Bus. Econ. Perspect., 2021; 48(1):1-21.
[4] Voznyak H V. Financial Fraud in the Budget Sphere: Economic Essence and Varieties. Bus. Inform, 2020; 4(507):334-339.
[5] Wu H, Chang Y, Li J, Zhu X. Financial fraud risk analysis based on audit information knowledge graph. Proc. Comput. Sci., 2022; 199:780-787.
[6] Hppner S, Baesens B, Verbeke W, Verdonck T. Instance-dependent cost-sensitive learning for detecting transfer fraud. Eur. J. Oper. Res., 2022; 297(1):291-300.
[7] Hilal W, Gadsden S A, Yawney J. A Review of Anomaly Detection Techniques and Applications in Financial Fraud. Expert Syst. Appl., 2021; 193(8):1-34.
[8] Jain A, Shinde S. A Comprehensive Study of Data Mining-based Financial Fraud Detection Research. 2019 IEEE 5th International Conference for Convergence in Technology (I2CT); 29-31 March 2019; Bombay, India. New York: IEEE; 2019. p. 1-4.
[9] Humpherys S L, Moffitt K C, Burns M B, Burgoon JK, Felix WF. Identification of fraudulent financial statements using linguistic credibility analysis. Decis. Support Syst., 2011; 50(3):585-594.
[10] Houssou R, Bovay J, Robert S. Adaptive Financial Fraud Detection in Imbalanced Data with Time-Varying Poisson Processes. J. Financ. Risk Manag., 2019; 08(4):286-304.
[11] Wen S, Li J, Zhu X, Liu M. Analysis of financial fraud based on manager knowledge graph. Proc. Comput. Sci., 2022; 199:773-779.
[12] Akra R M, Chaya J K. Testing the Effectiveness of Altman and Beneish Models in Detecting Financial Fraud and Financial Manipulation: Case Study Kuwaiti Stock. Int. J. Bus. Manag., 2020; 15(10):1-70.
[13] Zhou H, Sun G, Fu S, Fan X, Jiang W, Hu S, Li L. A Distributed Approach of Big Data Mining for Financial Fraud Detection in a Supply Chain. Comput. Mater. Con., 2020; 64(2):1091-1105.
[14] Burke J, Kieffer C, Mottola G, Perez-Arce F. Can educational interventions reduce susceptibility to financial fraud?. J. Econ. Behav. Organ., 2022; 198(Jun):250-266.
[15] Davidson R H. Who did it matters: Executive equity compensation and financial reporting fraud. J. Account. Econ., 2022(2/3):73.
[16] Achmad T, Ghozali I, Pamungkas I D. Hexagon Fraud: Detection of Fraudulent Financial Reporting in State-Owned Enterprises Indonesia. Economies, 2022; 10(1):1-16.
[17] Novatiani R A, Afiah N N, Sumantri R. Risk Management and other Factors Preventing Fraudulent Financial Reporting by State-Owned Enterprises in Indonesia. Asian Econ. Financ. Rev., 2022; 12(8):686-711.
[18] Kumar A, Mishra G S, Nand P, Chahar MS, Mahto SK. Financial Fraud Detection in Plastic Payment Cards using Isolation Forest Algorithm. Int. J. Innov. Technol. Explor. Eng., 2021; 10(8):132-136.
[19] Zhang J, Yao J, Wang L, Chen Y, Pan Y. A Financial Fraud Detection Model Based on Organizational Impression Management Strategy. J. Phys. Conf. Ser., 2020; 1616: 1-11.
[20] Amina Z. Financial Fraud Detection and the Importance of Internal Control. Int. J. Account. Financ. Rep., 2021; 11(4):28-36.
[21] Yadav A, Sora M. Fraud Detection in Financial Statements using Text Mining Methods: A Review. IOP Conf. Ser. Mater. Sci. Eng., 2021; 1020(1):1-9.
[22] Furui K, Ohue M. Compound Virtual Screening by Learning-to-Rank with Gradient Boosting Decision Tree and Enrichment-based Cumulative Gain. 2022 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB); 15-17 August 2022; Ottawa, ON, Canada. New York: IEEE; 2022. p. 1-7.
[23] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining;t 13-17 Augus, 2016; San Francisco California USA. New York, NY, United States: Association for Computing Machinery. 2016. p. 785-794.
[24] Millán O, Domenech LQ, Colom H, Fortuna V, Budde K, Sommerer C, López-Púa Y, Brunet M. Early prognostic performance of miR155-5p monitoring for the risk of rejection: Logistic regression with a population pharmacokinetic approach in adult kidney transplant patients. PLoS ONE, 2021; 16(1):1-20.
[25] Bernardo L S, Damasevicius R, de Albuquerque V, Maskeliunas R. A hybrid two-stage squeezenet and support vector machine system for parkinson's disease detection based on handwritten spiral patterns. Int. J. Ap. Mat. Com.-Pol, 2021; 31(4):549-561.
[26] Shanmugarajeshwari V, Ilayaraja M. Chronic Kidney Disease for Collaborative Healthcare Data Analytics using Random Forest Classification Algorithms. 2021 International Conference on Computer Communication and Informatics (ICCCI); 27-29 January 2021; Coimbatore, India. New York: IEEE; 2021. p. 1-14.
[27] Ma J, Sun L, Wang H, Zhang Y, Aickelin U. Supervised Anomaly Detection in Uncertain Pseudoperiodic Data Streams. ACM T. Internet Techn., 2016; 16(1):1-20.
[28] Hilal W, Gadsden S A, Yawney J. Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. Expert Syst. Appl., 2022; 193:1-34.
[29] Zhu X, Ao X, Qin Z, Chang Y, Liu Y, He Q, Li J. Intelligent financial fraud detection practices in post-pandemic era. Innovation, 2021; 2(4):1-11.
[30] Bahaweres R B, Trawally J, Hermadi I, Suroso AI. Forensic Audit Using Process Mining to Detect Fraud. J. Phys. Conf. Ser., 2021; 1779(1):1-10.
[31] Qiu S, Luo Y, Guo H. Multisource Evidence Theory-based Fraud Risk Assessment of China's Listed Companies. J. Forecast., 2021; 40(8):1524-1539.
[32] Xia H, Ma H. A Novel Structure-based Feature Extraction Approach for Financial Fraud Detection. J. Phys. Conf. Ser., 2021; 1865(4):1-7.
[33] Jan C L. Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry. Sustainability, 2021; 13(17):1-20.