DOI: 10.23977/jaip.2025.080114
ISSN 2371-8412 Vol. 8 Num. 1

Research on internal financial fraud identification model of enterprise based on ensemble learning

Yingjie Bu[1,a,*], Yi Wu[2,b], Guodong He[3,c], Qian Zhuge[4,d]

[1]Institute of Collaborative Innovation, University of Macau, Macau SAR, China
[2]Sussex Artificial Intelligence Institute, Zhejiang Gongshang University, Hangzhou City, Zhejiang Province, China
[3]College of Information Engineering, Wenzhou Business College, Wenzhou City, Zhejiang Province, China
[4]College of Finance and Trade, Wenzhou Business College, Wenzhou City, Zhejiang Province, China
[a]497213716@qq.com, [b]yw604@sussex.ac.uk, [c]1184136728@qq.com, [d]631127376@qq.com
**Corresponding author*

Keywords: Financial Fraud Detection, XGBoost, Machine Learning Model

Abstract: In recent years, financial fraud cases have been on the rise, prompting numerous scholars to explore relevant fields and contribute significantly to the practical oversight of the economy. Integrated learning models have also gained widespread application in the realm of financial fraud detection, proving their efficacy in identification. This paper provides a summary of existing research and methodologies employed by scholars. After reviewing pertinent literature, the Logistic Regression model, a Single Decision Tree, Gradient Boosting Decision Trees, Random Forest model, XGBoost model, and LightGBM model were selected as candidate models for studying financial fraud detection. A comparative analysis of their respective identification accuracies was conducted. The research findings indicate that across the overall detection models, the identification rates of all models exceed 70%. Among these, the XGBoost model exhibits the best performance, achieving an identification accuracy of 87.77%. From the comparative results, it is evident that the accuracy of ensemble learning models generally surpasses that of traditional classification models and basic machine learning models, effectively enhancing the efficiency of financial fraud detection. Furthermore, in terms of identification speed, ensemble learning models demonstrate advantages such as shorter processing times and the ability to accommodate larger datasets.

1. Introduction

With the increase in the size and complexity of businesses, financial fraud has become a focal point of concern for corporate management and regulatory authorities. Combining artificial intelligence methodologies to develop models for the identification of internal corporate financial fraud provides a potent tool for enterprises and regulatory bodies[1]. This can be utilized to more accurately detect potential instances of financial misconduct.

The logistic model, as one of the early models used for identifying financial fraud, has been widely applied in related fields. For instance, Pearson conducted research on financial fraud behaviors of listed companies using a stepwise logistic regression model[2]. They found that financial leverage, capital turnover rate, asset composition, and company size are important factors influencing financial fraud among listed companies[3]. The identification results were satisfactory, correctly identifying the majority of fraudulent cases. Y. Wen utilized a multinomial logistic regression model to select high-utility variables, resulting in a model that exhibited strong statistical significance and identification capabilities[4]. H. Jin established support vector machine and logistic regression models for financial fraud detection, evaluating their capabilities by analyzing identification efficiency[5]. Beneish, based on the logistic regression method, introduced innovative theory from the field of mathematics, specifically Taylor expansion, to create a nonlinear principal component logistic regression model. This addressed limitations imposed by linear assumptions and innovatively explored model setups that align better with patterns of financial misconduct. Spathis and Charalambos T constructed a model based on principal component analysis and logistic regression principles, demonstrating commendable scalability and accuracy. Their model holds instructive significance for real regulatory monitoring and preventive measures[6].

In recent years, the application of machine learning models has gradually garnered attention from scholars. Ys A and other researchers have combined traditional financial features with knowledge graph models, considering the interrelated information among various financial indicators. Research indicates that incorporating correlated information into financial feature representation significantly enhances the classification performance of SVM and K-NN models, yielding superior identification outcomes compared to decision trees and logistic regression[7]. W. Xu (2015) primarily employed classification algorithms to construct financial fraud detection models. Three single classifiers, namely C4.5, Bayesnet, libsvm, as well as two ensemble learning algorithms, Adaboost and random forest, were employed in model construction. Ultimately, the random forest model emerged as the optimal choice for achieving the best classification results. J. Zhang (2021) collected data encompassing 4 non-financial indicators and 26 financial indicators[8]. Employing methods like normality tests and factor analysis on the financial indicators, 6 common factors were extracted as identification signals. A data mining model was established, and the effects of neural networks, decision trees, and SVM models were compared. This led to the creation of a comprehensive identification model with higher credibility[9].

The paper provides a comprehensive summary of existing research and methodologies. After conducting a thorough review of relevant literature, we selected the Logistic Regression model, Single Decision Tree model, Gradient Boosting Decision Tree model, Random Forest model, XGBoost model, and LightGBM model as candidate models for studying financial fraud detection[10]. We conducted a detailed comparison and analysis of these models in terms of their accuracy in fraud identification. The research findings indicate that all models achieve a recognition rate of over 70% in the overall detection framework. Particularly, the XGBoost model stands out with an impressive accuracy of 87.77%. Comparative results demonstrate that ensemble learning models generally outperform traditional classification models and basic machine learning models in terms of accuracy, significantly enhancing the efficiency of financial fraud detection. Furthermore, in terms of recognition speed, ensemble learning models exhibit advantages such as shorter processing times and adaptability to larger datasets.

2. XGBoost model

XGBoost (eXtreme Gradient Boosting), in the Kaggle Higgs Boson Signal Recognition competition, garnered significant attention from participants and scholars due to its accurate identification results and impressive recognition efficiency. The development of this algorithm is based on a gradient boosting framework, with the optimization focusing on eliminating the step of calculating coefficients for the weak learners during the iterative process[11].

XGBoost is an addition expression consisting of k base models:

$$\hat{y}_i = \sum_{t=1}^{k} f_t(x_i) \tag{1}$$

Where f_k represents the k-th base model and $\hat{y}_i$ represents the predicted value for the i-th sample.

The loss function can be expressed using the predicted value $\hat{y}_i$ and the true value y_i:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) \tag{2}$$

Where, n is the number of samples.

The XGBoost algorithm is built upon the principle of optimizing a structured loss function, with the regularization term serving as a crucial component within the loss function during the optimization process. This aids in optimizing the weak learners while directly calculating the first and second derivatives of the loss function. Additionally, mechanisms like pre-sorting and weighted quantile are introduced, greatly enhancing the algorithm's utility[12].

The following is the algorithm flow:

1) Initialize the predicted values for each sample.

2) Define the objective function.

The performance of the model can be assessed through bias and variance. Bias is derived from the magnitude of the computed loss function, and inevitably, low variance corresponds to a simpler model with simpler outcomes. Therefore, a reasonable objective function is composed of the loss function L and a regularization term Ω aimed at reducing the complexity of the model:

$$Obj = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{i=1}^{n} \Omega\,(f_i) \tag{3}$$

Where $\Omega(f_t)$ represents the regularization term.

$$\Omega(f_t) = \gamma T_t + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_J^2 \tag{4}$$

T_t represents the weights on the individual leaf nodes of the decision tree, while γ and λ are pre-defined hyperparameters. After introducing the regularization term, the algorithm will favor relatively simpler yet high-performing models by considering the regularization factor. The regularization term is employed within the model to curtail overfitting of the classifier $f_i(x)$ during each algorithm iteration, but it does not play a role in the final ensemble model.

3) Taylor expansion of the objective function

The Boosting model builds upon forward additive steps, where at the current step t, the final prediction for the i-th sample x_i when introduced into the model is given by:

$$\hat{y}_i^t = \hat{y}_i^{t-1} + f_t(x_i) \tag{5}$$

Where $\hat{y}_i^t$ represents the predicted values provided by the model at step t-1, which effectively act as constants during the t-th prediction.

During the t-th model prediction, $f_t(x_i)$ represents the new model to be incorporated in that prediction. Substituting $\hat{y}_i^t$ into the objective function leads to further simplification.

$$Obj^{(t)} = \sum_{i=1}^{n} l(\hat{y}_i^t, y_i) + \sum_{i=1}^{n} \Omega(f_i) = \sum_{i=1}^{n} l(\hat{y}_i^{t-1} + f_t(x_i), y_i) + \sum_{i=1}^{n} \Omega(f_i) \tag{6}$$

From the above equation, it is evident that optimizing this objective function is tantamount to solving for the current $f_t(x_i)$. In each iteration of the XGBoost system, a new decision tree is constructed, with its specific construction based on the differences between the previous predicted values and the true values, also known as residuals. As for the aforementioned objective function, according to the Taylor formula:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 \tag{7}$$

Expanding $l(\hat{y}_i^{t-1} + f_t(x_i), y_i)$ around $l(\hat{y}_i^{t-1}, y_i)$, the objective function can be further simplified to:

$$Obj^{(t)} \approx \sum_{i=1}^{n} \left[l(\hat{y}_i^{t-1}, y_i) + g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \sum_{i=1}^{t} \Omega(f_i) \tag{8}$$

Where g_i represents the first derivative of the loss function l with respect to the previous predicted values, akin to the term $f'(x_0)$ in the Taylor expansion, and hi stands for the second derivative of the loss function l.

$$g_i \frac{\partial l\left((y_i, \hat{y}_i^{t-1})\right)}{\partial \hat{y}_i^{t-1}}, h_i = \frac{\partial^2 l\left((y_i, \hat{y}_i^{t-1})\right)}{\partial \hat{y}_i^{t-1}} \tag{9}$$

Since $\hat{y}_i^{t-1}$ is a known value during the t-th step prediction, it follows that $l((y_i, \hat{y}_i^{t-1}))$ is a constant, which will not affect function optimization. Thus, the objective function can be further expressed as:

$$Obj^{(t)} \approx \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2] + \sum_{i=1}^{t} \Omega(f_i) \tag{10}$$

4) Ultimate simplification of objective function based on decision tree

Based on the foundation of decision trees, XGBoost transforms the approach of traversing samples in a decision tree into the traversal of all leaf nodes to obtain optimal values. This leads to the derivation of:

$$\begin{aligned} \sum\nolimits_{i=1}^{t} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] &= \sum\nolimits_{i=1}^{n} \left[g_i \omega_{q(x_i)} + \frac{1}{2} h_i \omega_{q(x_i)}^2 \right] \\ &= \sum\nolimits_{j=1}^{T} \left[\left(\sum\nolimits_{i \in Ij} g_i \right) \omega_j + \frac{1}{2} \left(\sum\nolimits_{i \in Ij} h_i \right) \omega_j^2 \right] \end{aligned} \tag{11}$$

Where T is the total number of leaves in the decision tree.

Regarding the regularization term, it's necessary to constrain the complexity of the decision tree, ensuring that the weights of the leaf nodes are maintained at a reasonable level. Thus, the regularization term should be defined as:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} \omega_j^2 \tag{12}$$

Therefore, the total number of leaf nodes (controlled by the γ balance) to some extent represents the complexity and depth of the generated decision tree. Simultaneously, the norm of the vector composed of weights of each leaf node (controlled by the λ balance) also holds a certain influence. Upon substituting the results and simplification, the objective function becomes:

$$Obj^{(t)} = \sum_{j=1}^{T} \left[G_j \omega_j + \frac{1}{2} \left(H_j + \lambda \right) \omega_j^2 \right] + \gamma T \tag{13}$$

Here, G_j and H_j are known values calculated through t-1 steps, while w_j is an unknown value at this point. Therefore, when taking the first derivative of the objective function, it should yield:

$$\frac{\partial J(f_t)}{\partial \omega_j} = G_j + \left(H_j + \lambda \right) \omega_j = 0 \tag{14}$$

The weight corresponding to each leaf node can be calculated as follows:

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \tag{15}$$

Thus, the final objective function is given by:

$$Obj = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \tag{16}$$

5) Build the decision tree using the Best Split Point Partitioning algorithm.

6) Accumulate the new predicted values with the historical prediction results. As multiple decision trees are trained through additive training, the optimization of the objective function is performed step by step through iterative decision tree additions. Starting from the first tree and proceeding to the k-th tree, each one is optimized in sequence[13].

3. Data source and sample description

The research data in this paper are sourced from publicly available information provided by the China Securities Regulatory Commission (CSRC) and stock exchanges[13]. The financial misconduct data of listed companies used in this study, along with the financial data of all listed companies, are exclusively obtained from the CSMAR Guotaian database. The specific sources include the Chinese Listed Company Audit Research Database and the Chinese Listed Company Financial Indicator Analysis Database.[14]

The Chinese Listed Company Audit Research Database is a comprehensive and accurate professional database established based on in-depth research and understanding of audits conducted on listed companies[15]. It mainly comprises four sections: audit party information, listed company audit firms, violations, and restatement information. This database includes individual characteristics of Chinese certified public accountants and fundamental information about accounting firms, historical rankings of firms, and audit-related information of listed companies. Additionally, it encompasses audit violation information of both listed companies and accounting firms. The Chinese Listed Company Financial Indicator Analysis Database is derived from the CSMAR Chinese Listed Company Financial Statement Database. It involves derived calculations of authoritative and comprehensive financial indicators using scientific computation rules[16]. The indicators cover eleven different aspects of financial indicators, including solvency, disclosed financial indicators, and ratio structure. This database offers a more comprehensive, detailed, and intuitive understanding of the financial status and operational performance of listed companies. It provides vital data support for researchers to conduct empirical studies in a more convenient manner[17].

The selected samples of financial fraud in this study are sourced from the "Violation Information - Listed Company Financial Violations" table within the Chinese Listed Company Audit Research Database. This study considers companies involved in the following financial fraud behaviors as the subjects of analysis, as illustrated in Table 1.

Table 1: Types of violations involved in financial fraud.

Violation type	Violation code
Fictitious profit	P2501
Fictitious assets	P2502
False record	P2503
Deferred disclosure	P2504
Material omission	P2505
Mis-disclosure	P2506
Fraudulent listing	P2507
Illegal investment	P2508
Unauthorized change of use of funds	P2509

The selected sample period for this study spans from January 1, 2006, to September 30, 2021, encompassing a total of 761 fraud cases. Furthermore, to establish a financial fraud detection model, this paper selects matched samples from non-fraudulent companies for modeling. The criteria for selecting matched samples are as follows.

1) Choosing a ratio of 1:3 for the selection of non-fraudulent samples is mainly due to the relatively low occurrence of financial fraud cases among all listed companies. To establish a more objective and accurate fraud detection model that yields higher accuracy, this paper adopts a 1:3 ratio for the selection of fraudulent and non-fraudulent samples.

2) The matched non-fraudulent samples are selected from listed companies that are not categorized as ST or PT.

3) The financial statement data of the matched non-fraudulent samples share the same time period as the financial statement data of the non-fraudulent companies.

Considering the aforementioned selection criteria, this study has gathered a total of 761 fraudulent sample companies and 2117 non-fraudulent sample companies as research subjects, resulting in a cumulative dataset of 2879 financial data records. The results are shown in Figure 1. For model application, this paper randomly selects 80% of the data from the dataset for training purposes and reserves 20% as a test set for model training and identification.

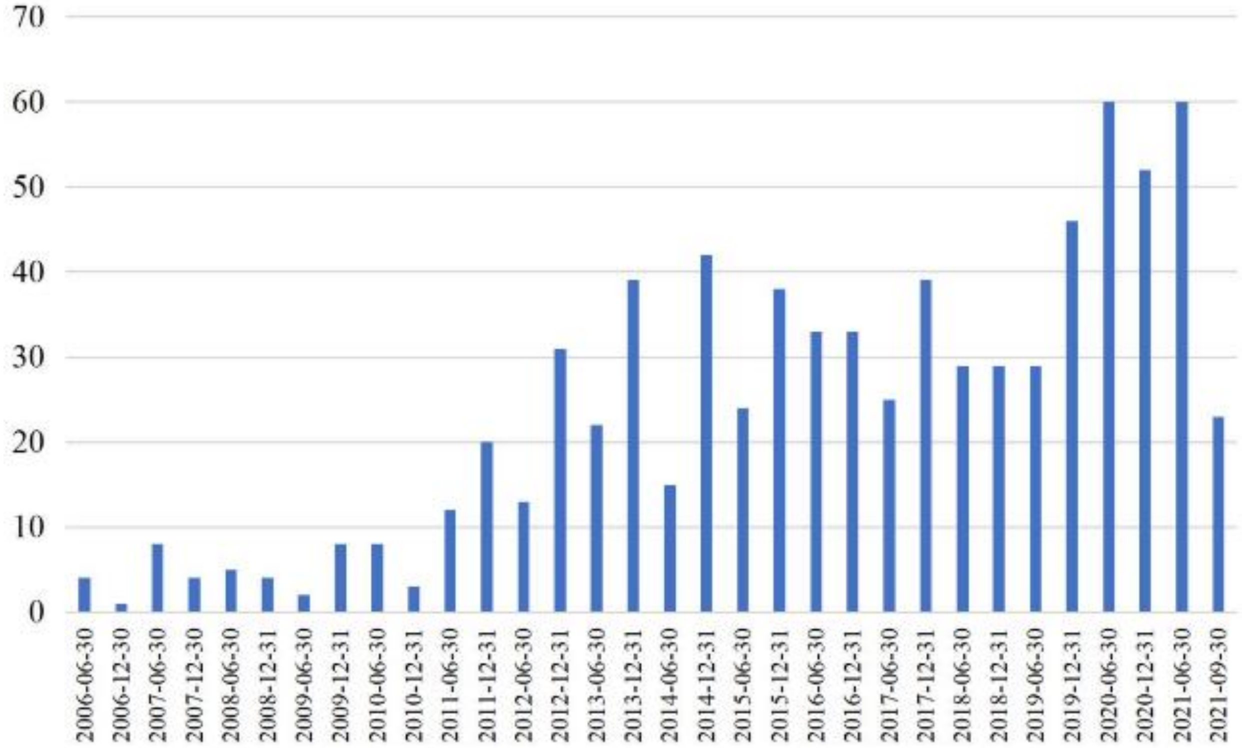

Figure 1: The time profile of the fake sample company.

From the time distribution graph of fraudulent sample companies in Figure 1, it can be observed that the number of fraudulent companies has been increasing annually from 2006 to 2021. Starting in 2013, there was a growing gap between the number of fraudulent companies and the previous years. In 2020, the number of fraudulent companies reached its peak at a total of 112. According to recent literature on the phenomenon of financial fraud, this trend is not baseless. In recent years, incidents of fraud have been emerging frequently. There are several reasons for this. Looking at the overall economic situation of listed companies, the downward economic cycle in recent years has placed various pressures on these companies. Simultaneously, issues such as governance failures, imbalances between costs and benefits, and rigid delegation systems exist, all of which enhance the incentives for financial fraud. On the other hand, considering the audit situation of listed companies in our country, limitations in audit scope and problems stemming from rapid expansion leading to the dominance of audit firms, as well as unfavorable conditions for fraud detection and prevention, result in diminished audit effectiveness. The combination of these internal and external factors contributes to the growing frequency and severity of fraudulent activities.

4. Analysis of experimental results

4.1. XGBoost algorithm model analysis

Before constructing the XGBoost algorithm model, the present study also carried out data preprocessing on a dataset of 4087 samples, resulting in a finalized dataset for the XGBoost identification model. In order to accurately assess the model's performance, the data was similarly divided into training and testing sets in a 3:7 ratio, thereby facilitating the introduction of the XGBoost model for training.

Prior to model training, this study employed XGBoost for feature selection, assessing the feature importance of various financial indicators within the XGBoost model. The following Figure 2 shows the results:

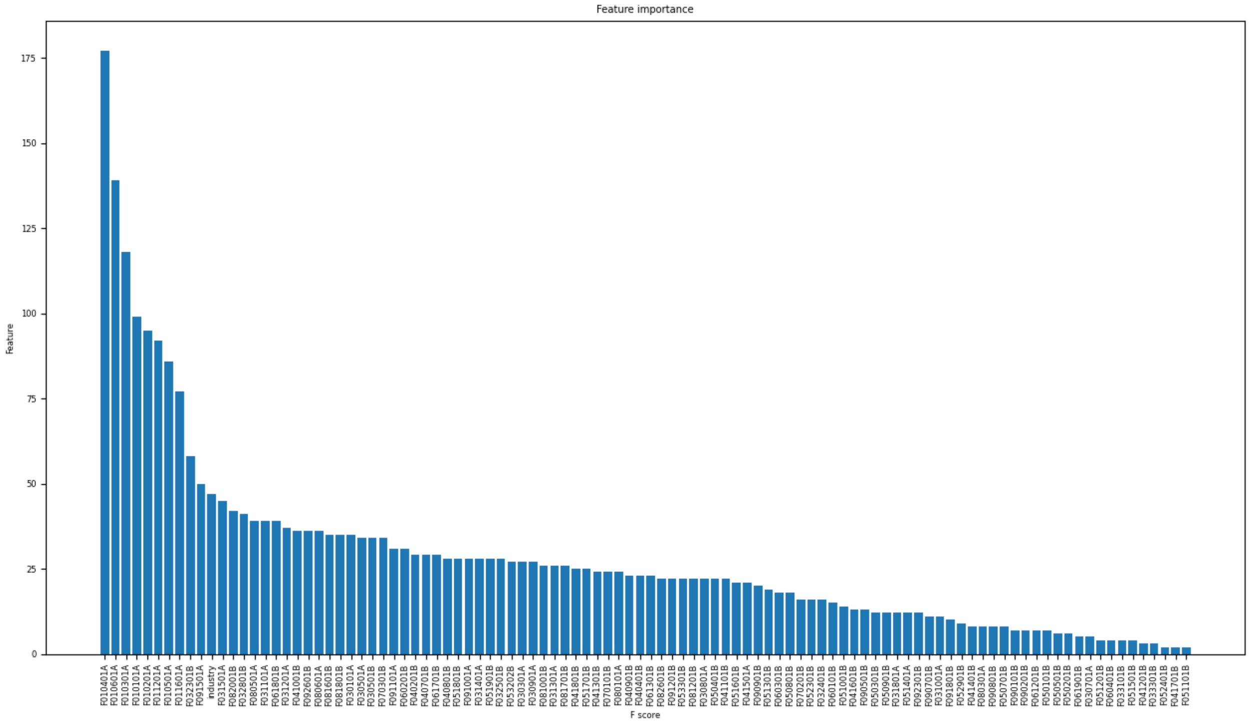

Figure 2: Xbgoost Model - Feature importance.

From the above chart (as shown in Figure 2), it can be observed that in terms of the contribution to the AUC metric, the XGBoost model considers the following financial indicators as having high feature importance: F010401A (Cash Ratio), F010601A (Operating Capital), F010301A (Conservative Quick Ratio), F010101A (Current Ratio), F010201A (Quick Ratio), F011201A (Debt-to-Asset Ratio), F011601A (Equity Multiplier). Among these, the Cash Ratio holds the greatest influence within the XGBoost model, with a feature score of 177. This metric is calculated by combining cash and cash equivalents with current liabilities. Thus, it is evident that factors related to cash flow capability and debt scale, which impact the solvency of listed companies, exert a significant influence on their financial fraud behaviors.

Based on the ranking of the importance of the above functions, the parameter settings of the XGBoost model are shown in Table 2 below:

Table 2: XGBoost model parameter values.

Argument	Numerical value
base_score	0.5
colsample_bylevel	1
colsample_bynode	1
colsample_bytree	1
learning_rate	0.3
max_depth	6
min_child_weight	1
scale_pos_weight	1

The recognition results of the XGBoost model are as follows: there are 514 correctly identified instances of non-financial fraud with an accuracy of 0.90, and 204 correctly identified instances of financial fraud with an accuracy of 0.82. The overall accuracy of the model in identifying financial fraud is 87.74%, which shows improvement compared to the aforementioned models. In order to further enhance the precision of the model's recognition, this study continues to fine-tune the parameters through grid parameter tuning, aiming to achieve a higher accuracy XGBoost model.

The optimal parameter values obtained are shown in Table 3.

Table 3: Optimal parameter configuration.

Argument	Numerical value
colsample_bytree	0.6
learning_rate	0.6
max_depth	8
subsample	0.9

In the XGBoost model after parameter tuning, there are 513 correctly identified instances of non-financial fraud with an accuracy of 90.63%, and 208 correctly identified instances of financial fraud with an accuracy of 82.53%. The overall accuracy of the model in identifying financial fraud is 88.14%. It is evident that the model's accuracy has improved after parameter tuning, resulting in a better fitting performance.

4.2. Comparative analysis of financial fraud identification models

Through the application of various ensemble learning models for financial fraud detection, this study has established financial fraud recognition models using the Logistic model, a standalone decision tree model, gradient boosting decision tree model, XGBoost model, and LightGBM model. The summarized accuracy results for these models are presented in Figure 3 and Figure 4.

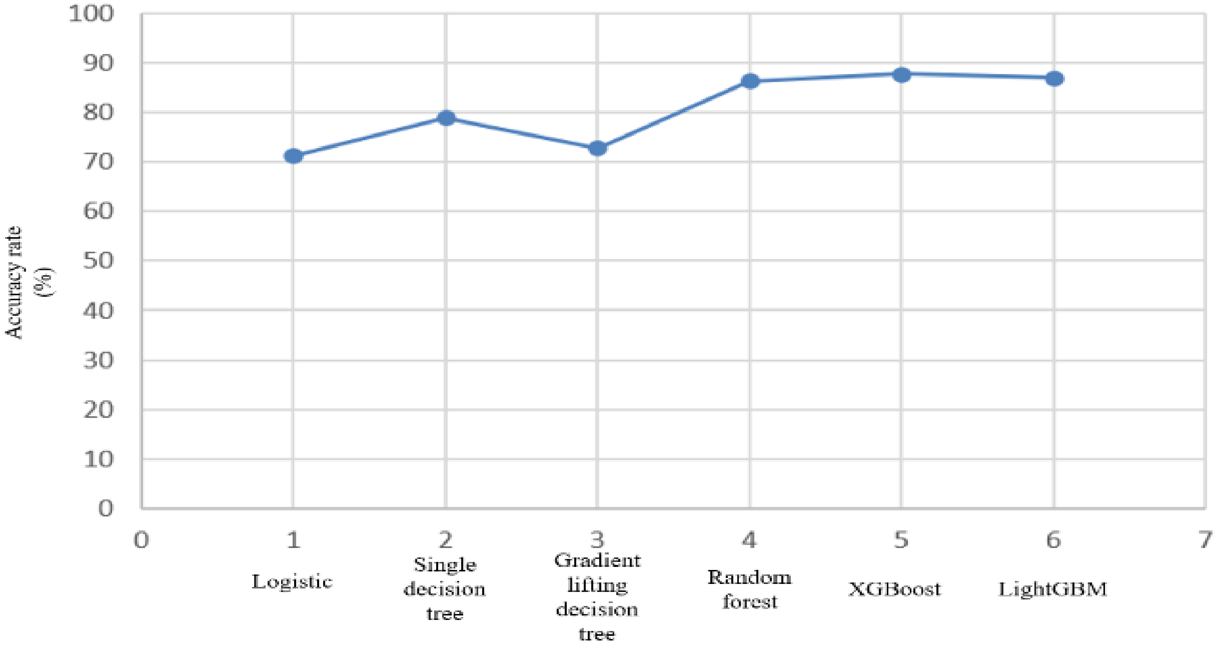

Figure 3: Overall recognition accuracy.

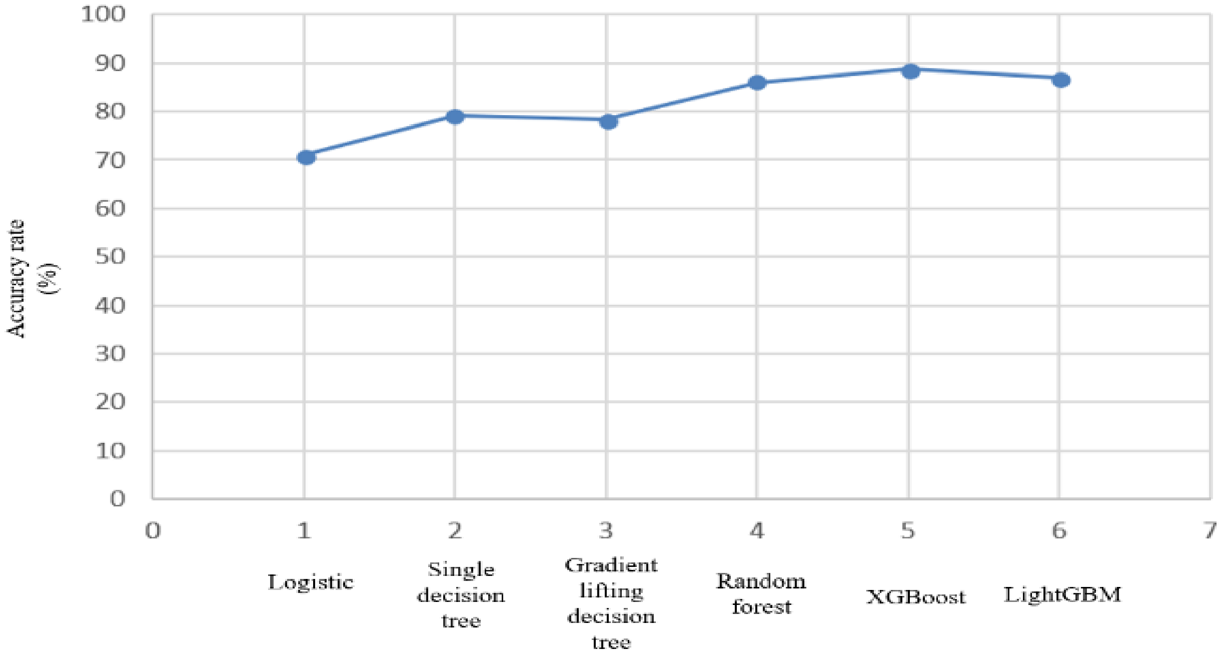

Figure 4: Distinguish manufacturing identification accuracy.

When not considering industry-specific factors, this study initially established a dataset based on overall financial indicators and trained various models for recognition. The recognition rates of all models exceeded 70%, with the XGBoost model achieving the highest accuracy at 87.77%. It demonstrated the best performance in identifying financial fraud, followed closely by the LightGBM model with an accuracy of 87.04%, showing a marginal difference from the XGBoost model. From the results of the six aforementioned models, ensemble learning-based models such as Random Forest, XGBoost, and LightGBM consistently outperformed other non-ensemble learning models in terms of recognition rate. This indicates that the application of ensemble learning can effectively enhance the efficiency of financial fraud detection. Moreover, when considering the overall time taken for model recognition, ensemble learning-based models exhibited shorter processing times, showcasing superior performance in terms of data handling capacity and other aspects.

When industry-specific factors were considered, this study introduced indicators to differentiate the manufacturing sector based on the original financial indicators, and then established various models. After accounting for industry distinctions, the accuracy of the single decision tree model, gradient boosting decision tree model, XGBoost model, and LightGBM model in recognizing instances improved to varying degrees. Notably, the industry factor had the most pronounced impact on enhancing the gradient boosting decision tree model's performance. This suggests that industry factors contribute significantly to the discriminative power of the gradient boosting decision tree model. From the results of the models incorporating indicators to differentiate the manufacturing sector, the XGBoost model still achieved the highest recognition accuracy. Based on this, the study concludes that the XGBoost model is an effective learning model for financial fraud detection, and its recognition accuracy is further enhanced when considering industry-specific factors.

5. Conclusion

After processing the data samples, we selected several candidate models for financial fraud detection, including Logistic Regression, Single Decision Tree, Gradient Boosting Decision Tree, Random Forest, XGBoost, and LightGBM models. We conducted a detailed comparison and analysis of these models' performance in identifying fraud. The research results indicate that within the entire detection framework, all models achieved a recognition rate exceeding 70%. Particularly noteworthy is the outstanding performance of the XGBoost model, achieving an impressive accuracy of 87.77%. In comparison with the control results, ensemble learning models generally outperformed traditional classification models and basic machine learning models in terms of accuracy, significantly enhancing the efficiency of financial fraud detection. Additionally, in terms of identification speed, ensemble learning models demonstrated advantages such as shorter processing times and adaptability to large-scale datasets.

References

[1] Oh J S, Shong I, "A case study on business model innovations using Blockchain: focusing on financial institutions," Asia Pacific Journal of Innovation & Entrepreneurship, vol. 11, no. 3, pp. 335-344, 2017

[2] W. Luo, "Innovation and application of blockchain technology in the financial field," Technoeconomics & Management Research, no. 8, pp. 90-95, 2018.

[3] Firdaus A, Razak M F A, and Feizollah A, et al, "The rise of "blockchain": bibliometric analysis of blockchain study," Scientometrics, vol. 120, no. 3, pp. 1289-1331, 2019.

[4] H. Cheng, and Y. Yang, "The development trend of block chain and commercial banks should study the policy," Financial Regulation Research, no. 6, pp. 73-91, 2016.

[5] Schutz A, Fertig T, and Weber K, et al, "Vertrauen ist gut, Blockchain ist besser – Einsatzmöglichkeiten von Blockchain für Vertrauensprobleme im Crowdsourcing," HMD Praxis der Wirtschaftsinformatik, vol. 55, no. 6, pp. 1155-1166, 2018.

[6] J. Zhang, "Construction and application of enterprise financial sharing service under cloud computing environment," Friends of Accounting, vol. 597, no. 21, pp. 136-140, 2018.

[7] N. Li, and J. C. Mitchell, "RT: a Role-based Trust-management framework," Proceedings DARPA Information Survivability Conference and Exposition, vol. 1, pp. 201-212, 2013.

[8] Ouaddah A, Abou Elkalam A, and Ait Ouahman A, "FairAccess: a new Blockchain-based access control framework for the Internet of Things," Security and Communication Networks, vol. 9, no. 18, pp. 5943-5964, 2016.

[9] Aitzhan N Z, and Svetinovic D, "Security and Privacy in Decentralized Energy Tradingthrough Multi-signatures, Blockchain and Anonymous Messaging Streams," IEEE Transactions on Dependable & Secure Computing, pp. 1-3, 2016.

[10] Wadud M, and Ali Ahmed H J, "Factors affecting delinquency of household credit in the U.D. Does consumer sentiment paly a role," The North American Journal of Economis and Finance, no. 52, pp.101132-101134, 2021.

[11] Patil S, Nemade V, and Soni P K, "Predictive modelling for Credit card fraud detection using data analytics," Procedia Computer Science, no. 132, pp. 385-395, 2021.

[12] Lukas. M, "The goal gradient effect and repayments in consumer credit," Economics Letters, no. 171, pp.208-210, 2020.

[13] Aitzhan N Z, and Svetinovic D, "Security and Privacy in Decentralized Energy Tradingthrough Multi-signatures, Blockchain and Anonymous Messaging Streams," IEEE Transactions on Dependable & Secure Computing, pp. 1-3, 2016.

[14] Y. Zheng, and R. Du, "Application of block chain technique in the management of digital knowledge assets," Science Citation Database, vol. 26, no. 3, pp. 97-104, 2018.

[15] Y. Zhou, and H. Li, "Data management solution based on blockchain," Information Security Research, vol. 6, no. 1, pp. 37-45, 2020.

[16] Y. Huang, Z. Liang, and Y. Sun, "Supply chain trusted data management based on blockchain," Journal of Computer Science and Technology, vol. 27, no. 12, pp. 9-17, 2018.

[17] Heilman E, Baldimtsi F, and Goldberg S, "Blindly Signed Contracts: Anonymous On-Blockchain and Off-Blockchain Bitcoin Transactions," International Conference on Financial Cryptography and Data Security, pp. 43-50, 2016.