sustainability

MDPI

Article

Financial Fraud Detection of Listed Companies in China: A Machine Learning Approach

Yasheng Chen and Zhuojun Wu *

Department of Accounting, School of Management, Xiamen University, Xiamen 361005, China
* Correspondence: anthony604329469@163.com

Abstract: As the focus of capital market supervision, financial report fraud has shown a development trend of enormous numbers, complex transactions, and hidden means in recent years. To improve audit efficiency and reduce the dependence on non-financial data, the study only uses the structured original data in the financial report to constructs a new fraud identification model, which can quickly detect fraud in China. This study takes the listed companies in China from 1998 to 2016 as research samples and selects 28 sets of raw data from financial reports. Then, this study compares the detection effectiveness of two single classification machine learning algorithms and five ensemble learning algorithms on fraud detection. Compared with single classification machine learning algorithms, the results show that ensemble learning algorithms are generally better at detecting fraud for Chinese listed companies, and the stacking algorithm performs the best. The study results provide direct evidence for rapid fraud detection using financial report raw data and ensemble learning algorithms. The study first proposes a stacking algorithm-based financial reporting fraud identification model for listed companies in China, which provides a simple and effective approach for investors, regulators, and management. It can also provide a reference for the detection of other fraud scenarios.

Keywords: artificial intelligence; machine learning; fraud detection; financial report; financial disclosures; China; securities fraud

Citation: Chen, Y.; Wu, Z. Financial Fraud Detection of Listed Companies in China: A Machine Learning Approach. *Sustainability* **2023**, *15*, 105. https://doi.org/10.3390/su15010105

Academic Editors: Albert Y.S. Lam and Yanhui Geng

Received: 10 November 2022
Revised: 9 December 2022
Accepted: 16 December 2022
Published: 21 December 2022

1. Introduction

Financial report fraud has occurred frequently in recent years. Due to the increasingly hidden fraud of listed companies, traditional fraud detection methods cannot meet the review requirements promptly. Limited by human resources and detection methods, it usually takes several months, or even years, for listed companies to obtain an administrative penalty notice from the China Securities Regulatory Commission (CSRC) for suspected financial report fraud. "The ACFE Global Fraud Investigation Report" is the most cited source of fraud and job-related crime data in the world. It has been published every two years since it was first published in 1996. The questionnaire was distributed to fraud investigators and anti-fraud practitioners in various countries worldwide, and ACFE sets up a special team to collect, analyze, summarize, and explain the survey. The survey results released in 2020 are based on 2504 survey reports from 125 countries. According to the 2020 ACFE Global Fraud Investigation Report [1], the average investigation time for each fraud case is 14 months. Financial report fraud has led to the highest economic loss for all types of fraud, with a median loss of USD 964,000. Regulatory authorities can effectively reduce the fraud of listed companies through timely intervention and construct a timely and effective fraud identification model for the financial report of listed companies, which is significant for investors, analysts, and regulators [2].

To improve audit efficiency, researchers introduce machine learning as an effective fraud detection tool. Machine learning can find hidden rules in massive data far more efficiently than manual work [3]. Additionally, the detection of financial fraud can be regarded as a classification task in machine learning [4,5]. Generally, machine learning

algorithms can be divided into traditional single classifiers and ensemble classifiers [6–8]. Specifically, the single classifier refers to the classification or prediction of samples by a single classification method. The ensemble classifier integrates the prediction results of multiple single classifiers based on the single classification to obtain the final prediction result. Different ensemble learning algorithms have been developed according to different integration processes [6–8].

However, in previous studies on listed companies in China, researchers only paid attention to how to select appropriate financial ratios or non-financial information to construct a fraud identification model, and hold the view that these financial ratios can cover the complete financial information, but they ignored the direct information in financial reports [9–11]. During the generation process, essential original information may be lost by financial ratios in the financial report [10], and non-financial information originates from different channels. Non-financial information may come from management news, board meeting minutes, contract details, etc. [12]. This information can be divided into three categories, including management characteristics and control environments, governance and organization structures, and relation and industry conditions [5]. For example, management characteristics and control environments include board mobility, highly complex transactions, excessive incentives to management, ineffective accounting, and information systems, etc. Governance and organization structures include the proportion of the largest shareholder, board of directors, board of supervisors, the proportion of independent directors, etc. [13]. Relation and industry conditions include significant related-party transactions not in the ordinary course of business, significant declines in increasing business failures in the industry, etc. The information is basically in an unstructured form, which is difficult to deal with in a unified and centralized way, thus increasing the difficulty of financial report audits [14]. This paper suggests that fraud detection from the financial reporting raw accounting numbers may be more practical and provide a simple and direct approach.

To fully utilize the vital information in raw accounting data, reduce the dependence on non-financial information, and achieve rapid fraud detection in financial reporting, it is necessary to develop a simple and effective fraud detection model using the raw financial data of listed companies in China. This paper selects the financial report raw data of listed companies in China from 1998 to 2016 and tests the applicability of the financial report fraud detection model developed by Bao et al. in China's capital market. Then, based on the research background of listed companies in China, two single classifiers and four ensemble classifiers, namely logistics, SVM, random forest, RUSBoost, XGBoost, AdaBoost, and stacking, are selected to construct financial reporting fraud identification models respectively; meanwhile, comparative research is conducted to determine financial reporting fraud identification models suitable for the capital market in China. The results indicate that the financial report fraud recognition model developed by Bao et al. does not achieve the expected effect in China's capital market. After further comparing the detection effects of various machine learning methods, it is found that the fraud detection effect of the ensemble learning method is generally better than that of the single classifier method. Especially, the financial reporting fraud identification model based on the stacking algorithm achieves the best detection effect, and its AUC reaches 0.742.

The contributions of this paper are summarized as follows. First, the study extracts raw data from financial reports to provide a new perspective to quickly detect fraud in Chinese listed companies from financial reports, improving audit efficiency and reducing the dependence on non-financial data. Then, this paper tests the applicability of the existing western financial reporting fraud identification model in the Chinese capital market. Subsequently, the study uniformly compares the effectiveness of two single classifiers and five ensemble classifiers in detecting fraud of listed companies in China; the results provide direct evidence for rapid fraud detection using financial report raw data and ensemble learning algorithms. Finally, under the background of listed companies in China, the study first proposes a stacking algorithm-based financial reporting fraud identification model

for listed companies in China, which provides a simple and effective detection method for investors, analysts, and regulators.

2. Literature Review

There is some important literature related to variable selection and method selection. Researchers have tried to construct an effective financial reporting fraud detection model to identify financial reporting fraud more efficiently and accurately, by using financial indicators and machine learning methods.

2.1. Raw Accounting Data and Financial Ratios

Researchers at home and abroad have paid much attention to the selection of financial ratios to establish fraud identification models. For example, Skousen (2008) [15] selected five pressure indicators and two opportunity indicators according to the fraud triangle theory and established a logistic regression model to evaluate the impact of each financial indicator on financial fraud. Additionally, Spathis (2002) [16], Kirkos (2006) [17], Ravisankar (2011) [18], and Kanapickienė (2015) [19], used financial ratios and non-financial data for fraud detection. Aiming at the raw financial report data, Bao et al. (2020) [10] used 28 sets of raw financial report data to construct a RUSBoost model that can effectively detect the fraud of listed companies in the United States, in which 28 groups of detailed raw data can be referred to the details of Section 4. However, few studies have proposed a fraud detection model for listed companies in China based on the original financial report data. Therefore, this paper puts forward Hypothesis 1: the RUSBoost model applicable to the listed companies in the US can achieve the same forecast effect in China.

2.2. Single Classifier and Ensemble Classifier

In the existing research, the single classifier algorithm such as logistics and support vector machine (SVM) is widely used to detect financial fraud of listed companies in China [4,9,19,20], and the adopted ensemble algorithm includes random forest, XGBoost, etc. [14,21,22]. Nagai (2010) [23] surveyed 49 related studies published from 1997 to 2008 and found that the single classifier method was still the mainstream in financial fraud detection research during this period. Logistic regression is a classic single-classifier method in fraud detection research [9]. Many researchers have also proposed to use ensemble algorithms for fraud detection. Kotsiantis (2006) [24] proposed to use several machine learning methods to integrate the stacking model to construct a fraud prediction model to predict financial fraud. Hajeck's (2017) [25] research uses the AdaBoost algorithm to classify financial fraud. Baesens (2021) [26] believes that the XGBoost algorithm is suitable for fraud detection. Bao et al. (2020) [10] applied the RUSBoost algorithm to construct a fraud detection model suitable for the US capital market.

However, there has been little research on using raw financial reporting data to construct machine learning models for listed companies in China, and there is no uniform comparison of the effectiveness of different machine learning methods. The algorithms taken for comparison in this study include two traditional machine learning methods (the logistics model and SVM) and four ensemble learning algorithms (random forest, AdaBoost, XGBoost, RUSBoost, and stacking). To construct a financial reporting fraud detection model suitable for the Chinese market, this paper compares the detection effects of single classifiers and ensemble classifiers and puts forward Hypothesis 2: based on the unified sample data set, the ensemble classifier can achieve better fraud detection results than the single classifier.

3. Overview of Machine Learning

Logical regression is often used in classification problems, and its essence is linear regression. The maximum likelihood estimation method is often employed to solve each parameter, and a layer of Sigmoid function is added to the feature-to-result mapping [27,28]. SVM is a supervised binary classification algorithm. Its general idea is to find a hyperplane with

the best generalization ability in the sample space to divide the samples, thus maximizing the distance between the sample points closest to the hyperplane in the two types of samples [29,30]. Compared with a single classifier using a single method, the ensemble classifier makes classifications by combining the results of multiple classifiers. According to the different integration approaches, the ensemble classifier can be divided into three categories: bagging, boosting and stacking [31]. Bagging sets the result of each classifier to the same weight, and the final classification result is obtained by voting, which is represented by the random forest algorithm [32,33]. Bagging gives different weights to different classifiers, and it allocates more weight to mis-divided samples in the classifier's training process, thus improving the performance of the classifier. This type of ensemble classifier includes AdaBoost [34,35], XGBoost [36,37], RUSBoost [38,39], and other variants. Stacking is an ensemble learning technology that integrates multiple classifications models through a meta-classifier. The classifiers selected by stacking are of different types. It first obtains the results of multiple classifiers, then takes the results as features, and uses a meta-classifier, such as logical regression, for training to obtain the final classification results [40,41].

In view of the complexity of the mathematical formulas of these algorithm models, Table 1 is attached to list some relevant literature on each algorithm in detail.

Table 1. Related literature list of algorithm model.

Model	Related Literature
Logistic Regression	Peng C Y J, Lee K L, Ingersoll G M. [27]. Bewick V, Cheek L, Ball J. [28]
SVM	Jakkula V. [29] Chen Y W, Lin C J. [30]
RUSBoost	Seiffert C, Khoshgoftaar T M, Van Hulse J, et al. [38] Seiffert C, Khoshgoftaar T M, Van Hulse J, et al. [39]
AdaBoost	Wu X, Kumar V, Ross Quinlan J, et al. [34] Schapire R E. [35]
Random Forest	Breiman L. [32] Biau G, Scornet E. [33]
XGBoost	Brownlee J. [36] Chen T, Guestrin C. [37]
Stacking	Ting K M, Witten I H. [40] Sill J, Takács G, Mackey L, et al. [41]

4. Verification of Applicability of the RUSBoost Model

4.1. Sample Selection and Data Sources

This paper uses the research sample of all listed companies from 1990 to 2019. The sample data are obtained from the CSMAR financial reporting database and violation database. Given the completeness and accessibility of the data, the financial data involved in this study is derived from the annual report. The CSMAR violations database covers 16 types of violations, as shown in Table 2. Most of the financial reporting frauds of listed companies in China focus on the whitewashing and manipulation of profits [42], so this study selects corporate annual reports with fictitious profits as fraud samples and manually collates 400 corporate annual reports involving fictitious profits, according to the announcement of irregularities by CSMAR.

Table 2. The type of violation for the CSMAR violation database.

Violation Category	
(1) Fictitious profit	(9) False disclosure (other)
(2) Fictitious assets	(10) insider dealing
(3) False records (misleading statements)	(11) Fraudulent listing
(4) Improper handling of general accounting	(12) Illegal trading of shares
(5) Deferred disclosure	(13) Changing the use of funds without authorization
(6) Major omissions	(14) Illegal guarantee
(7) Violation of capital contribution	(15) Manipulate stock prices
(8) Occupation of corporate assets	(16) Other

According to the CSMAR's violation database, this study statistically collates the proportion of the number of companies involved in fictitious profits in the annual financial report from 1990 to 2019 to the total number of listed companies in the database for the current year, as shown in Table 3. The analysis result indicates that according to the regulations of the Ministry of Finance of China, enterprises are required to stop preparing the statement of changes in financial position from 1998 and replace it with the statement of cash flows. Meanwhile, the announcement of financial report fraud is lagging. The proportion of fictitious profit companies announced from 2017 to 2019 to the total number of listed companies in the current year decreased by 0.69%, 0.35%, and 0.10%, respectively, indicating that there may be some companies whose fictitious profit behaviors have not been discovered during this period. Therefore, to ensure the sample's reliability, this study selects companies' annual financial reports from 1998 to 2016 as the sample data, with a total of 35,574 samples, including 337 samples of annual fraud of fictitious profit companies. The data show that between 1998 and 2016, the number of companies involved in fictional profits accounted for about 1% of all listed companies, and the specific proportion fluctuated around 1% each year.

Note that this paper labels listed companies as fraud according to the CSRC's fictional profit penalty announcement, and it is assumed that CSRC has successfully identified all false profit companies, and there is no misstatement in the announcement that has been issued. Correspondingly, in the fraud model built by Bao, there is the same assumption about the samples of American-listed companies. In practice, it is impossible to determine the correct extent to which CSRC can identify fraud, but this does not weaken the importance of the data or the conclusions of this study. For example, assuming that the model can quickly and effectively detect 1% of listed companies with fraud labels, even if fraudulent listed companies account for 2% of all listed companies, it can still save substantial time and help to reduce the risk of investors.

4.2. Selection of Raw Financial Data

Referring to the raw financial data selection by Bao et al. (2020), this paper constructs 28 characteristic variables, of which 25 variables are derived from the raw financial report data; the remaining three variables are "annual turnover of tradable shares", "closing price at the end of the year", and "market value of tradable shares at the end of the year", and they are derived from the CSMAR stock trading database. Due to the differences between some financial reporting items of China and the United States, the corresponding variables in China's capital market and their calculation formulas are listed in Table 4.

Table 3. The statistical distribution of annual reports of fictitious profit companies from 1990 to 2019.

Year	Number of Listed Companies with Fictitious Profits	Total Number of Listed Companies	Percentage of Listed Companies with Fictitious Profits
1990	0	11	0.00%
1991	0	17	0.00%
1992	0	78	0.00%
1993	0	229	0.00%
1994	0	361	0.00%
1995	0	389	0.00%
1996	7	605	1.16%
1997	17	819	2.08%
1998	14	929	1.51%
1999	12	1030	1.17%
2000	17	1175	1.45%
2001	22	1258	1.75%
2002	23	1319	1.74%
2003	24	1381	1.74%
2004	21	1469	1.43%
2005	11	1464	0.75%
2006	12	1547	0.78%
2007	7	1661	0.42%
2008	9	1715	0.52%
2009	7	1864	0.38%
2010	12	2218	0.54%
2011	19	2452	0.77%
2012	22	2580	0.85%
2013	26	2624	0.99%
2014	27	2739	0.99%
2015	41	2927	1.40%
2016	36	3221	1.12%
2017	25	3598	0.69%
2018	13	3690	0.35%
2019	4	3893	0.10%
Grand Total	428	49263	0.87%

Table 4. The selection of variables from raw financial data.

	Raw Financial Variables Selected in the American Financial Reports and Capital Markets (Bao, 2020)	Corresponding Raw Financial Variables in Chinese Financial Reports and Capital Markets
Balance Sheet	(1) Cash and short-term investments	(1) Cash + Short-term Investments
	(2) Receivables, total	(2) Accounts Receivable + Notes Receivable
	(3) Inventories, total	(3) Inventories
	(4) Short-term investments, total	(4) (Before 2007) Net Value of Short-term Investment; (5) (2007 and beyond) Fair Value Through Other Comprehensive Income
	(5) Current assets, total	(6) Current Assets

Table 4. *Cont.*

	Raw Financial Variables Selected in the American Financial Reports and Capital Markets (Bao, 2020)	Corresponding Raw Financial Variables in Chinese Financial Reports and Capital Markets
	(6) Property, plant and equipment, total	(7) Total Tangible Assets
	(7) Investment and advances, other	(8) (Before 2007) Net Value of Short-term Investment + Net Value of Long-term Investment; (9) (From 2007 to 2017) Investment Property + Trading Financial Assets + Available-for-sale Securities + Held-to-maturity Investments + Long-term Equity Investments (10) (2018) Investment Property + Trading Financial Assets + Other Equity Instrument Investment + Long-term Equity Investments
	(8) Assets, total	(11) Total Assets
	(9) Accounts payable, trade	(12) Accounts Payable
	(10) Debt in current liabilities, total	(13) Notes Payable + Non-current Liabilities maturing within one year
	(11) Income taxes payable	(14) Income Tax Expense − Deferred Income Tax Liabilities + Deferred Income Tax Assets
	(12) Current liabilities, total	(15) Total Current Liabilities
	(13) Long-term debt, total	(16) Total Long-term Liabilities
	(14) Liabilities, total	(17) Total Current Liabilities
	(15) Common/ordinary equity, total	(18) Total Owner's Equity − Preferred Stock
	(16) Preferred/preference stock (capital), total	(19) Preferred Stock
	(17) Retained earnings	(20) Surplus Reserve + Retained earnings
Income Statement	(18) Sales/turnover (net)	(21) Operating Income
	(19) Cost of goods sold	(22) Operating Cost

Table 4. *Cont.*

	Raw Financial Variables Selected in the American Financial Reports and Capital Markets (Bao, 2020)	Corresponding Raw Financial Variables in Chinese Financial Reports and Capital Markets
	(20) Depreciation and amortization	(23) Depreciation for Fixed Assets + Amortization of Intangible Assets + Amortization of Long-Term Expenses Prepayments
	(21) Interest and related expenses, total	(24) Selling Expenses + Administration Expenses + Financial Expenses
	(22) Income taxes, total	(25) Income Tax Expense
	(23) Income before extraordinary items	(26) Net Profit + Non-operating Revenue − Non-operating Expenses
	(24) Net income (loss)	(27) Net Profit
Cash Flow Statement	(25) Long-term debt issuance	(28) Bonds Payable
Market Value	(26) Sale of common and preferred stock	(29) Annual Turnover of Tradable Shares
	(27) Price close, annual, fiscal	(30) Annual Closing Price
	(28) Common shares outstanding	(31) Year-end Circulation Market Value

In February 2006, the Ministry of Finance issued 39 new accounting standards, and the listed companies in China were required to implement these accounting standards from 1 January 2007 [43]. Since then, the Ministry of Finance has successively issued three accounting standards for financial instruments. Since the sample covers the period from 1998 to 2016, this study considers the revision of the accounting standards, generates the financial variables involved according to the corresponding calculation formulas for different periods, and keeps the calculation method of the remaining financial variables unchanged.

4.3. Data Preprocessing

There are some missing values in the financial reporting data in the research samples obtained from the CSMAR database. To evaluate the performance of this model in the Chinese market, this study adopts the same approach as that of Bao et al. (2020) and retains the missing values to avoid noise caused by other filling forms. In this study, when the values of 28 characteristic variables in a sample are missing, the missing value is retained. For example, if "Preferred Stock" is null in a sample in 2012, the null value is retained instead of being filled with averages or other values.

Among all samples selected in this paper, the fraud samples involving fictitious profits announced by the CSRC are labeled as 1, and the remaining non-fraud samples are labeled as 0.

4.4. Evaluation Indicators

In the existing research, the detection of financial reporting fraud is often regarded as a typical two-class classification task, so the performance of the financial fraud model can be measured by using the evaluation indicators for classification tasks. The evaluation indicators mainly include the following two categories: The indicators based on the confusion matrix including accuracy, precision, true positive rate (sensitivity), true negative rate (specificity), false positive rate, false negative rate, and F1 score; the indicators based on the ROC curve, including AUC [44,45].

Due to the small proportion of fraud samples, the accuracy cannot properly reflect the model's prediction performance. Considering that the commit of fraud will bring significant harmful effects to the enterprise and the market. This paper selects precision, sensitivity, and AUC as evaluation indicators to explain more critical information. Specifically, precision measures the percentage of all samples predicted to be fraudulent are really fraudulent. Sensitivity measures the percentage of all fraud samples correctly predicted to be fraud samples. The ROC curve is a curve with FPR as the abscissa and TPR as the ordinate. The AUC is equal to the area under the ROC curve. The AUC considers both the accuracy of positive sample identification and the false positive rate of negative samples. Thus, it comprehensively investigates the true positive rate and true negative rate, and its value falls within [0, 1]. The closer the AUC value is to 1, the better the prediction effect of the model.

4.5. Model Construction

As the internal stakeholders of the enterprise will try their best to hide the fraud, external regulators need to spend much time to obtain sufficient evidence, and the verification process will consume a lot of staffing and material resources. Although the companies that are found to have committed fraud will be made public, in practice, due to a certain interval (i.e., the lag period) between the time when the listed companies in China commit fraud and the time when the fraud is found, it is not possible to obtain all the fraud samples in the lag period. To guarantee the generalization performance of the model, the lag period is considered when setting the training set and the testing set.

This paper uses the data of listed companies in China from 1998 to 2009 as the training set, and the data of listed companies in 2011, 2012, 2013, 2014, 2015, and 2016 as the testing set. Specifically, six groups of experiments are conducted. Each group of experiments selects the sample data of not less than 10 years as a training set and uses the sample data in the T+2 year as the test set. For example, if the sample data from 1998 to 2009 are used as the training set, the sample data in 2011 are used as the test set. If the sample data from 1998 to 2010 are used as the training set, the sample data in 2012 are used as the test set, etc.

According to the divided sample data set, this study uses Python language to call the RUSBoostClassifier module, and constructs a financial fraud detection model based on RUSBoost algorithm. Referring to the design parameters of the RUSBoost model, the number of decision trees is set at 3000, the minimum number of leaf samples is 5, the learning rate is 0.1, and the ratio of the number of fraud samples to the number of non-fraud samples is 1:1.

The RUSBoost algorithm is a common ensemble learning algorithm for unbalanced samples. It forms a new balanced sample data set by randomly under-sampling a certain proportion from a majority of data sets and then combining it with a minority of samples for unbalanced data sets.

4.6. Model Results

According to the evaluation indicators selected in this paper, the results are shown in Table 5.

Table 5. The results of the RUSBoost model.

Training Set Period	Testing Set Period	AUC	Sensitivity	Precision
1998~2009	2011	0.636	47.4%	61.4%
1998~2010	2012	0.656	51.8%	62.8%
1998~2011	2013	0.597	56.6%	67.5%
1998~2012	2014	0.586	52.9%	60.8%
1998~2013	2015	0.634	48.5%	61.5%
1998~2014	2016	0.493	54.1%	63.8%
	Average Value	0.600	51.9%	63.0%
	Bao's model value	0.725		

According to the model's results, the average AUC is 0.60, which is far lower than that of the model proposed by Bao et al. (with an AUC of 0.725). Thus, Hypothesis 1 is denied. The RUSBoost model suitable for the Western markets does not achieve the same prediction effect in China. Among them, the average sensitivity of the model is 51.9%, which indicates that among the listed companies in China, the fraud samples judged correctly by the model account for only half of the actual fraud samples, and the identification performance is low. The average precision of the model is 63%, indicating that 63% of the samples that are judged as fraud are fraudulent samples. Though the model can assist regulators in making pre-judgment to a certain extent, the identification ability is far lower than the effect applied to Western markets.

5. Improvement of the Financial Reporting Fraud Detection Model

5.1. Model Optimization

Through the analysis of the results, this study considers that the sample data set needs to be redesigned due to the extreme imbalance of positive and negative samples and the difference between the fraud paradigm in the Chinese market and that in the United States market, which has a significant impact on the detection performance. The following are two aspects of improvement considered in this paper:

1. The extreme imbalance of positive and negative samples.

Financial report fraud detection is a typical extreme imbalance sample problem. As stated in Section 4.1, among the Chinese listed companies, the proportion of companies with fictional profits fluctuates around 1% each year. Compared with non-fraudulent companies, fraudulent companies account for only a small proportion. Under an extreme imbalance of positive and negative samples, the random selection of non-fraudulent samples based on fraudulent samples brings great volatility. Therefore, this paper reduces the scope of non-fraud samples based on the industry, year, and asset size of the fraud samples to reduce the volatility caused by random sampling of non-fraud samples.

This paper selects matched non-fraud samples for each fraud sample under the following restrictions: first, the corresponding sample companies and the fraud sample companies are in the same industry; second, the corresponding sample and the fraud sample are in the same year; and third, the asset size of the corresponding sample ranges from 95% to 105% that of the fraud sample. Based on 337 fraud samples from 1998 to 2016, 4463 non-fraud samples were selected. In this study, six data sets are constructed, with the samples from 1998~2009 as the training set and the samples from 2011~2016 as the test set. In each constructed data set, non-fraud samples are randomly selected and matched with fraud samples. Aiming at the problem of continuous fraud, to avoid exaggerating the prediction effect, for the samples of listed companies with continuous fraud in the training set, this study changes the fraud sample mark "1" to the non-fraud sample mark "0" [10].

2. Adjustment of interval time between the training set and testing set.

When the previous model is built, the interval between the training set and the testing set is 2 years. However, after statistics and calculations, it is found that the average time spent in detecting fraud of listed companies in China in this study is about 4 years. The interval of some samples is presented in Table 6.

Table 6. The sample example of the time interval between fraud year and announcement time.

Securities Code of Sample Company	Year of Fraud Practices	Year of Fraud Announcement	Interval Time
000020	2000	2004	4
000156	2001	2007	5
000405	1998	2002	4
000408	2017	2019	2
000430	1996	2001	5
000508	1996	1998	2
000509	2000	2005	5
000510	1997	2001	4
000511	2015	2017	2
000514	1998	2003	5
000519	2014	2018	4
000529	2003	2007	4

To avoid the influence of a lag period on model performance, this study set the T+4th year of the last year of the training set as the year of the testing set. In the newly constructed six data sets, the samples in 1998–2007 can be taken as the training set and those in 2011 as the testing set; the samples in 1998~2008 can be taken as the training set and those in 2012 as the testing set. Similarly, the average value of the model tests of the six data sets is taken as the test result. The improved data set is shown in Table 7.

Table 7. The division of the training set and test set.

Dataset Number	Training Set Period	Testing Set Period	Interval Time
1	1998~2007	2011	4 years
2	1998~2008	2012	4 years
3	1998~2009	2013	4 years
4	1998~2010	2014	4 years
5	1998~2011	2015	4 years
6	1998~2012	2016	4 years

5.2. The Results of the Improved Model

To construct a fraud detection model suitable for listed companies in China, this paper selects two single classifiers and five ensemble classifiers for comparison. The single classifiers include logistic regression, SVM, and ensemble classifiers involving random forest, AdaBoost, XGBoost, RUSBoost, and stacking. Among them, the stacking algorithm can form a robust classifier by combining multiple weak classifiers. Particularly, the weak classifier of the stacking algorithm can be a single classifier or an ensemble learning algorithm. The stacking algorithm used in this paper is ensembled by six classifier algorithms, including numbers 1–6 in Table 8. The evaluation indicators are AUC, sensitivity, and precision.

Table 8. The statistics of the improved model results.

Number	Model	AUC	Sensitivity	Precision
	RUSBoost (Before readjustment)	0.600	51.9%	63.0%
1	RUSBoost	0.721	64.7%	64.7%
2	Logistic Regression	0.697	58.8%	61.7%
3	SVM	0.679	70.6%	67.6%
4	AdaBoost	0.723	64.7%	70.6%
5	Stacking	0.742	76.5%	76.5%
6	Random Forest	0.739	70.6%	70.6%
7	XGBoost	0.740	70.6%	73.5%

Based on Python language, this study uses different modules for each algorithm functions. For single classifier, the study uses the LogisticRegression module and the svm_SVC module. For ensemble classifiers, the study uses the RUSBoostClassifier module, the RandomForestClassifier module, the AdaBoostClassifier module, the XGBClassifier module and the StackingClassifier module. This study evaluates the performance of the model through evaluation indicators such as AUC, sensitivity, and precision. The model results are shown in Table 8.

The results show that after the re-adjustment mentioned above of the sample data set of listed companies in China, the AUC of the fraud detection model based on the RUSBoost algorithm rises to 0.721, the sensitivity rises from 0.519 to 0.647, and the precision rate rises from 0.63 to 0.647. The change in sensitivity shows that the proportion of samples successfully detected by the model in the actual fraud samples has increased from 51.9% to 64.7%, indicating that the ability of the model to detect fraud samples has been significantly improved. Note that this study compares the detection performance of single classifiers and ensemble classifiers on the unified sample data set. The AUC of RUSBoost, random forest, AdaBoost, XGBoost, and stacking reaches 0.721, 0.739, 0.723, 0.740, and 0.742, while the AUC of logistic regression and SVM reaches 0.697 and 0.679, respectively. Therefore, the AUC of the ensemble classifier is generally larger than that of the single classifier. Considering that AUC comprehensively investigates the true positive rate and true negative rate, the comparison results of AUC indicate that the overall performance of the ensemble classifier is better than that of the single classifier, which supports Hypothesis 2. In addition, compared with the AUC value of the original RUSBoost model (i.e., 0.725), random forest, XGBoost, and stacking can achieve better detection performance for financial fraud report identification of listed companies in China.

Among them, the stacking model has achieved better performance. From the sensitivity index, the sensitivity of the stacking model reaches 76.5%, indicating that the fraud samples correctly judged by the model account for 76.5% of the actual fraud samples. Thus, the model can identify most fraud samples and has specific practical application value. From the precision index, the precision of the stacking model can reach 76.5%, indicating that the samples that are judged to be fraudulent account for 76.5% of actually fraudulent samples, and the judgment efficiency is relatively high.

Each test system will randomly select non-fraud samples equal to the number of fraud samples from the sample data set, which will bring certain fluctuations to each result. To evaluate the stability of each model, this study repeats the selection process 20 times to statistically analyze each result, and the standard deviation is calculated. The statistical results of stability are shown in Table 9.

Table 9. The standard deviation of the results of the improved model.

	Dataset Number						
Model	**1**	**2**	**3**	**4**	**5**	**6**	**Average Value**
RUSBoost (before readjustment)	0.057	0.053	0.062	0.032	0.037	0.051	0.049
RUSBoost	0.066	0.072	0.064	0.065	0.060	0.051	0.063
Logistic Regression	0.076	0.047	0.052	0.048	0.047	0.047	0.053
SVM	0.050	0.062	0.066	0.057	0.034	0.052	0.054
AdaBoost	0.071	0.071	0.065	0.035	0.047	0.050	0.057
Stacking	0.043	0.054	0.056	0.060	0.041	0.053	0.051
Random forest	0.066	0.074	0.048	0.047	0.049	0.041	0.054
XGBoost	0.078	0.063	0.055	0.051	0.043	0.036	0.054

The results indicate that, compared with other algorithms, the stacking model has better stability and can keep the detection performance consistent under multiple tests.

By comparing various machine learning methods, this study finds that the AUC of the stacking model can reach 0.742, indicating that the model is effective for detecting financial reporting fraud of listed companies in China and can keep the detection performance relatively stable. The financial reporting fraud detection model of listed companies in China based on the stacking algorithm proposed in this paper can predict whether financial reporting fraud has occurred in listed companies in practice, thus helping supervisors, investors, and others to focus on listed companies that are predicted to be fraudulent. With the help of the model, the labor and time costs in the preliminary investigation can be significantly reduced, and the judgment efficiency of investors and regulators can be greatly improved.

6. Conclusions and Implication

This study makes an empirical study on the detection of financial fraud based on machine learning model. Based on the background of China's capital market, the study selects Chinese listed companies from 1998 to 2016, and selects 28 groups of raw data from financial reports. For financial reporting of fraud of Chinese listed companies, the study constructs machine learning models based on two single-classification machine learning algorithms and five ensemble learning algorithms, and compares the performance of each model. The research results indicate that the ensemble classifier generally has a better detection effect than the single classifier. AdaBooat, XGBoost, RUSBoost, random forest and stacking can effectively detect fraud of listed companies in China. Among them, a financial reporting fraud detection model suitable for listed companies in China can be better constructed based on the stacking algorithm, with an overall AUC of 0.742 and sensitivity and precision of 76.5%, which can effectively detect financial reporting fraud.

There are certain theoretical and practical implications of the research findings presented in this paper. From the perspective of theoretical implications, this paper extracts raw data from financial reports to provide a new perspective to quickly detect fraud in Chinese listed companies from financial reports. This is also the first time to propose a stacking algorithm-based financial reporting fraud identification model for listed companies in China, which enriches the application of the machine learning method in financial reporting fraud identification. Meanwhile, for financial report fraud detection, this study finds that the ensemble classifier has better detection performance than the single classifier, which can provide a specific reference for choosing the machine learning method in other similar scenarios. Besides, there are also certain practical implications of this study. Firstly, the study proposes that the raw financial report data can be directly used for fraud detection in China, which reduces the work of data processing and the practical difficulty of fraud detection. Secondly, this study uses the stacking algorithm to construct a new, effective financial reporting fraud identification model for listed companies in China, which provides a simple and effective approach for investors, regulators, and management.

Author Contributions: Conceptualization, Z.W.; Methodology, Z.W.; Validation, Y.C.; Writing—review & editing, Y.C.; Supervision, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 72172132.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 2020 Global Study on Occupational Fraud and Abuse. Available online: https://legacy.acfe.com/report-to-the-nations/2020/ (accessed on 2 November 2022).
2. Jia, C.; Ding, S.; Li, Y.; Wu, Z. Fraud, enforcement action, and the role of corporate governance: Evidence from China. *J. Bus. Ethics* **2009**, *90*, 561–576. [CrossRef]
3. Mitchell, T.M. *The Discipline of Machine Learning*; Carnegie Mellon University, School of Computer Science, Machine Learning Department: Pittsburgh, PN, USA, 2006.
4. Perols, J. Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Audit. A J. Pract. Theory* **2011**, *30*, 19–50. [CrossRef]
5. Song, X.-P.; Hu, Z.-H.; Du, J.-G.; Sheng, Z.-H. Application of machine learning methods to risk assessment of financial statement fraud: Evidence from China. *J. Forecast.* **2014**, *33*, 611–626.
6. Dietterich, T.G. Ensemble Methods in Machine Learning. In *International Workshop on Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15.
7. Ballings, M.; Van den Poel, D.; Hespeels, N.; Gryp, R. Evaluating multiple classifiers for stock price direction prediction. *Expert Syst. Appl.* **2015**, *42*, 7046–7056. [CrossRef]
8. Duin, R.P.W. The combining classifier: To train or not to train? In *Object Recognition Supported by User Interaction for Service Robots*; IEEE: Piscataway, NJ, USA, 2002; Volume 2, pp. 765–770.
9. Dechow, P.M.; Ge, W.; Larson, C.R.; Sloan, R.G. Predicting material accounting misstatements. *Contemp. Account. Res.* **2011**, *28*, 17–82. [CrossRef]
10. Bao, Y.; Ke, B.; Li, B.; Yu, Y.J.; Zhang, J. Detecting accounting fraud in publicly traded US firms using a machine learning approach. *J. Account. Res.* **2020**, *58*, 199–235. [CrossRef]
11. Sun, Y.; Ma, Z.; Zeng, X.; Guo, Y. A Predicting Model for Accounting Fraud Based on Ensemble Learning. In Proceedings of the 2021 IEEE 19th International Conference on Industrial Informatics, Palma de Mallorca, Spain, 21–23 July 2021.
12. Tang, J.; Karim, K.E. Financial fraud detection and big data analytics—Implications on auditors' use of fraud brainstorming session. *Manag. Audit. J.* **2018**, *34*, 324–337. [CrossRef]
13. Yao, J.; Zhang, J.; Wang, L. A financial statement fraud detection model based on hybrid data mining methods. In Proceedings of the 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 26–28 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 57–61.
14. Bao, Y.; Hilary, G.; Ke, B. Artificial Intelligence and Fraud Detection. In *Innovative Technology at the Interface of Finance and Operations*; Babich, V., Birge, J.R., Hilary, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2022; Volume 1, pp. 223–247.
15. Skousen, C.J.; Smith, K.R.; Wright, C.J. Detecting and Predicting Financial Statement Fraud: The Effectiveness of the Fraud Triangle and SAS No. 99. In *Corporate Governance and Firm Performance*; Hirschey, M., John, K., Makhija, A.K., Eds.; Emerald Group Publishing Limited: Bingley, UK, 2009; Volume 13, pp. 53–81.
16. Spathis, C.T. Detecting false financial statements using published data: Some evidence from Greece. *Manag. Audit. J.* **2002**, *17*, 179–191. [CrossRef]
17. Kirkos, E.; Spathis, C.; Manolopoulos, Y. Data mining techniques for the detection of fraudulent financial statements. *Expert Syst. Appl.* **2007**, *32*, 995–1003. [CrossRef]
18. Ravisankar, P.; Ravi, V.; Rao, G.R.; Bose, I. Detection of financial statement fraud and feature selection using data mining techniques. *Decis. Support Syst.* **2011**, *50*, 491–500. [CrossRef]
19. Kanapickienė, R.; Grundienė, Ž. The model of fraud detection in financial statements by means of financial ratios. *Procedia Soc. Behav. Sci.* **2015**, *213*, 321–327. [CrossRef]
20. Cecchini, M.; Aytug, H.; Koehler, G.J.; Pathak, P. Detecting management fraud in public companies. *Manag. Sci.* **2010**, *56*, 1146–1160. [CrossRef]
21. Purda, L.; Skillicorn, D. Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemp. Account. Res.* **2015**, *32*, 1193–1223. [CrossRef]

22. Liu, C.; Chan, Y.; Alam Kazmi, S.H.; Fu, H. Financial fraud detection model: Based on random forest. *Int. J. Econ. Financ.* **2015**, *7*, 178–188. [CrossRef]
23. Ngai, E.W.T.; Hu, Y.; Wong, Y.H.; Chen, Y.; Sun, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decis. Support Syst.* **2011**, *50*, 559–569. [CrossRef]
24. Kotsiantis, S.; Koumanakos, E.; Tzelepis, D.; Tampakas, V. Forecasting fraudulent financial statements using data mining. *Int. J. Comput. Intell.* **2006**, *3*, 104–110.
25. Hajek, P.; Henriques, R. Mining corporate annual reports for intelligent detection of financial statement fraud—A comparative study of machine learning methods. *Knowl. Based Syst.* **2017**, *128*, 139–152. [CrossRef]
26. Baesens, B.; Höppner, S.; Verdonck, T. Data engineering for fraud detection. *Decis. Support Syst.* **2021**, *150*, 113492. [CrossRef]
27. Peng, C.Y.J.; Lee, K.L.; Ingersoll, G.M. An introduction to logistic regression analysis and reporting. *J. Educ. Res.* **2002**, *96*, 13–14. [CrossRef]
28. Bewick, V.; Cheek, L.; Ball, J. Statistics review 14: Logistic regression. *Crit. Care* **2005**, *9*, 1–7. [CrossRef]
29. Jakkula, V. Tutorial on support vector machine (svm). *Sch. EECS Wash. State Univ.* **2006**, *37*, 3.
30. Chen, Y.W.; Lin, C.J. Combining SVMs with Various Feature Selection Strategies. In *Feature Extraction*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 315–324.
31. Pavlyshenko, B. Using stacking approaches for machine learning models. In Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 21–25 August 2018; pp. 255–258.
32. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
33. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [CrossRef]
34. Wu, X.; Kumar, V.; Ross Quinlan, J.; Yang, Q. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]
35. Schapire, R.E. Explaining Adaboost. In *Empirical Inference*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 37–52.
36. Brownlee, J. *XGBoost with Python: Gradient Boosted Trees with XGBoost and Scikit-Learn*; Machine Learning Mastery: San Francisco, CA, USA, 2016.
37. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
38. Seiffert, C.; Khoshgoftaar, T.M.; Van Hulse, J.; Napolitano, A. RUSBoost: Improving classification performance when training data is skewed. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–4.
39. Seiffert, C.; Khoshgoftaar, T.M.; Van Hulse, J.; Napolitano, A. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2009**, *40*, 185–197. [CrossRef]
40. Ting, K.M.; Witten, I.H. Stacking Bagged and Dagged Models. In Proceedings of the International Conference on Machine Learning, Nashville, TN, USA, 8–12 July 1997.
41. Sill, J.; Takács, G.; Mackey, L.; Lin, D. Feature-weighted linear stacking. *arXiv* **2009**, arXiv:0911.0460.
42. Huang, S.; Ye, Q.; Xu, S.; Ye, F. Analysis of Financial Fraud of Listed Companies in China from 2010 to 2019. *Financ. Account. Mon.* **2020**, *14*, 153–160. (In Chinese)
43. Wang, M. The Study of Accounting Conservatism after the Promulgation of New Accounting Standards for Enterprises-Based on the Evidence in China's Capital Market. *Int. Bus. Res.* **2013**, *6*, 183. [CrossRef]
44. Elmrabit, N.; Zhou, F.; Li, F.; Zhou, H. Evaluation of machine learning algorithms for anomaly detection. In Proceedings of the 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Dublin, Ireland, 15–19 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–8.
45. Grandini, M.; Bagli, E.; Visani, G. Metrics for multi-class classification: An overview. *arXiv* **2020**, arXiv:2008.05756.