International Journal of Knowledge Processing Studies
(KPS)
Homepage: http://kps.artahub.ir/

ORIGINAL RESEARCH ARTICLE

Application of Data Mining to Detect Accounting Fraud in Information Systems

Tina Malekolkalami[1], Khadijeh Khodabakhahi Parijan[2,*], Maliheh Alifari[3]

[1] *MSc. of Accounting, Department of Accounting, West Tehran Branch, Islamic Azad University, Tehran, Iran. tina_malek_1369@yahoo.com, 0000-0002-9737-2331*

[2] *Assistant Professor, Department of Accounting, West Tehran Branch, Islamic Azad University, Tehran, Iran. parijan.kh@gmail.com, 0000-0001-6653-8653*

[3] *Assistant Professor, Department of Accounting, West Tehran Branch, Islamic Azad University, Tehran, Iran. malihe.alifari@wtiau.ac.ir, 0000-0001-7738-782X*

ARTICLE INFO

Article History:
Received: 2022/04/22
Revised: 2022/05/09
Accepted: 2023/07/01
Published Online: 2023/07/04

Keywords:
data mining, fraud detection, accounting fraud, information system, stock exchange member companies.

Number of Reference: 45
Number of Figures: 4
Number of Tables: 14

DOI: 10.22034/kps.2023.394274.1127
DOR: 20.1001.1.27834611.2023.3.4.6.9

Publisher:
Ayande Amoozan -e- ATA (AAA)

ABSTRACT

The purpose of this research is to use data mining to detect accounting fraud in the database of stock exchange member companies. The combination of discrete and continuous data has increased the necessity of using data mining and machine learning methods in the field of fraud detection. This research is applied in terms of purpose and descriptive in terms of method. The document review method was used to collect information in the field of literature and research background. The prepared questionnaire includes 7 main indicators consisting of 48 questions for each of the variables. This questionnaire was made available by the researcher to 400 accountants of companies admitted to the Tehran Stock Exchange by sampling method. In order to fit the model, the structural equation method was used in SMARTPLS software. In the data mining section, all IB1, IBK, LWL, KSTAR, and KNN algorithms were used to simulate the proposed model in Rapidminer software. Effectiveness of internal control, compensation system, asymmetry of information, compliance with accounting rules, management ethics, and ethical principles are effective and meaningful on accounting fraud. In evaluating parameters and according to the graphs, K-STAR algorithm has better performance than other algorithms. The proposed data mining model for financial fraud detection showed that since the amount of data creation in financial companies is increasing day by day with the development of technology, it is possible to provide early detection of fraud by reviewing and analyzing the data. **©authors**

►**Citation:** Malekolkalami, T., Khodabakhahi Parijan, K., & Alifari, M. (2023). Application of data mining to detect accounting fraud in information systems. *International Journal of Knowledge Processing Studies (KPS)*, 3(4): 60-72. Doi: 10.22034/kps.2023.394274.1127

1. Introduction

Accounting knowledge has two aspects: recognition and analysis. Fraud detection requires knowledge of financial data analysis to uncover lack of transparency (Pargmann et al., 2023). Accounting fraud is an issue that has wide consequences in the field of economy, industry, as well as daily life (Fanai et al., 2023). Accounting fraud reduces confidence in the industry, destabilizes the economy, and affects the cost of people's daily lives (Akila et al., 2018). Traditional methods (manual techniques) such as auditing are inefficient and unreliable due to the high severity of the problem (Chen et al., 2020). Data mining-based methods are efficient and useful due to the ability to identify small anomalies in large data sets (Donyaee et al., 2019). Accounting knowledge based on the identification of fraudulent trends and unsubstantiated reports can be the basis for the necessary knowledge to detect fraud (Xie et al., 2023). There are many types of accounting fraud; In contrast, various data mining methods also give us the opportunity to use the best method in each field of financial fraud (Kewei et al., 2021). Accounting fraud is a broad term that potentially has different meanings; But in this research, accounting fraud generally means the use of illegal methods and methods in order to obtain financial profit (Ofori et al., 2014). Transparency is one of the important components of financial reporting. The detection of fraud caused by accounting knowledge comes from the detection of lack of transparency in financial reports (Liang et al., 2019). Accounting fraud has a very negative impact on stock exchange companies and society, so that credit card fraud alone causes billions of dollars in damages annually (Forough et al., 2021). Some graphs show that in the United States of America, credit card fraud alone costs billions of dollars annually (Wahid et al., 2023). Accountants need knowledge of fraud detection based on data mining and data-driven techniques. By extracting the necessary knowledge, they can prevent financial frauds to a large extent. Annually, accounting fraud has caused hundreds of millions of dollars in losses to business units around the world, and continuous rumors and promotions resulting from such acts and malicious deeds on a large scale can have long-term disastrous consequences. In recent years, advances in modern technologies such as the Internet and mobile computing have increased accounting fraud (Popat et al., 2018). Currently, data mining has shown that it plays a significant role in areas such as credit card approval, bankruptcy prediction, and stock market analysis. Fraud detection is also considered a problem of identification and classification, with the difference that there is a wide imbalance in identifying legitimate and fake transactions; So that its wrong classification will cause exorbitant costs. Data mining is a viable method for processing big data for fraud detection; So that it will be able to detect with high accuracy without the need of input variable information (Rai et al., 2020). Today, due to the critical issue of accounting in large companies and its impact on other company processes, the need to examine important and related matters at this level is quite tangible. In the meantime, the question that can be answered is why with the increase in the level of education of accountants and the creation of technologies that help the working person in this matter and facilitate the handling of the performance and financial status of the company, it is still witnessed There are reports of fraud and corruption, negligence, neglect of duty, failure to achieve effectiveness or efficiency, lack of employee compensation and satisfaction, and inability to manage the accounting group optimally; Issues that can affect the economy of a country.

It is safe to say that correct and timely examination of financial statements and evaluation of the quality of accounting information of companies can prevent such issues from occurring, and accountants play a very important role in financial decisions by providing favorable results (Khalilpour, Sohedeh, Chashmi, 2018). For example, internal controls have been taken into consideration in order to increase investors'

confidence in the reliability of financial statements following the financial scandals of companies such as Enron and WorldCom. Abroad, the Sarbanes-Oxley Act in 2002 required companies to publish internal controls, and also in Iran in 2013, the Stock Exchange Organization, in line with Article 18, published internal control guidelines for issuers admitted to the Tehran Stock Exchange (Brandek and Mohammadi, 2018).

Therefore, to the authors' best knowledge, no research has been done to examine the relationship and impact of internal control, compensation compliance, accounting ethics, compliance with accounting laws, accounting fraud, information asymmetry, and management ethics together at the listed companies of the Tehran Stock Exchange, this research seeks to investigate the relationship between each of the mentioned components and the impact of each on accounting fraud. This research seeks answers to the following questions:

What are the factors affecting accounting fraud?

To what extent are the current data mining techniques used to detect accounting fraud reliable?

2. Literature Review

According to Figure 1, generally, fraud in its life cycle can be used as a model, in such a way that by analyzing it, a suitable answer can be obtained (Bose et al., 2007)

As with its analysis and analysis (data mining intelligence methods), a proper answer is given to this fraud, and again with the development of knowledge and providing new solutions and protocols, the way is opened for fraudsters, and new fraud methods are formed and the life cycle of fraud continues.

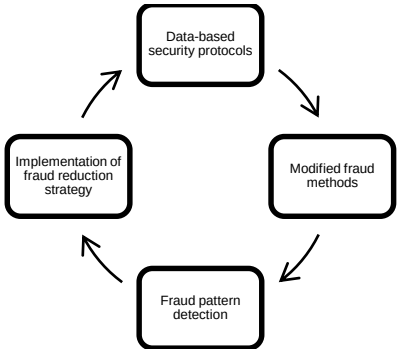

Figure 1. *Fraud lifecycle based on data mining (Lodi et al., 2014).*

The emerging trend of accounting fraud is generally detected through analysis and information extraction (data mining) from the database of transactions of financial institutions, which are flagged (Wieder et al., 2015), and this leads to Developing new security and authentication policies and protocols helps (Abdallah et al., 2016).

In contrast, fraudsters also change their methods based on the implementation of new security policies as well as the identification of alternative fraud opportunities on current platforms (Zhu et al., 2016) and then new fraud methods and patterns (Yue et al., 2007), which resulted from the implementation of new security policies, have been re-identified (Sundarkumar et al., 2015) and strategies to prevent their occurrence are predicted and applied (Drezewski et al., 2015). In Table 1, we will briefly show all the data mining methods and their features in discovering accounting fraud:

Table 1. *Examining the characteristics of data mining methods to detect financial fraud*

Algorithm	Characteristics	Requirements
Neural network	Has a good track record in detecting financial fraud and suitable for problems without algorithms and binary classifications Simple to implement Has a good track record in detecting financial fraud	Requires high computing power for training and implementation. Inappropriate for real functions. Can be fitted in models that have an unsupervised learning algorithm. Therefore, it needs constant retraining to adapt to new fraud methods.
Logistics model	Simple to implement Has a good track record in detecting financial fraud	Low classification performance compared to other methods Unusable in complex problems of fraud detection
Support vector machine	Ability to be used in solving non-linear classification problems in fraud detection Implementation and training of this method requires simple computer calculations and is	Difficulty for auditors to process results due to changes in input data sets

	suitable for real functions.	
Decision tree, forest and CART	Simple to implement and understand Implementation and training of this method requires simple computer calculations and is suitable for real functions.	Can be fitted in models that have an unsupervised learning algorithm. Therefore, it needs constant retraining to adapt to new fraud methods. Optimization during the initial setup requires high computing power.
Genetic algorithm	Simple to implement in classification problems with high accuracy and model fit Suitable for problems without algorithms and binary classifications	Requires high computing power for training and implementation. Inappropriate for real functions. Difficult to adapt to fraud models due to internal minimum and maximum problems
Text mining	Very useful for all kinds of cheating with large amounts of Textual data, such as fraudulent financial statements	Need another classification method in doing Real detection of fraud Textual data is usually more subjective so it is difficult to process.
Group data analysis technique	Easy implementation Guaranteed to find the best solution available	Difficult to classify noisy data, which includes a lot of fraud
Response level methodology	Ability to solve non-linear classification problems such as fraud detection	Low classification performance compared to other methods Difficult to detect complex fraud issues
Self-organized maps	Simple to implement and very easy for auditors to visually understand the results.	Use with auditor supervision and simply does not have the ability to run automatically.
Bayesian trust network	Suitable for problems without algorithms and binary classifications Can be used in real functions with powerful computer calculations	It requires a strong understanding of normal and abnormal behavior to investigate types of fraud.
mining process	Useful for internal fraud checks where information is available for each repeating step.	Ability to focus on an overall process chain rather than individual features
Artificial immune system	Very suitable for classification problems with unbalanced data, such as fraud detection	Requires high computing power for training and implementation. Inappropriate for real functions.
Combined technique	Adaptation in new cheating techniques with combination in Strengths of multiple traditional fraud detection methods	Fraud is a high cost problem and therefore new, under test Different methods are risky

There are various standards to determine the performance and efficiency of detection techniques, which usually have three common characteristics: accuracy, sensitivity, and detection power (Wieder et al., 2015).

Accuracy: percentage of detection success; The ratio of successful cases to failed cases (Tarjo Herawati, 2015)

Sensitivity: percentage of true detection value; Comparing the number of cases correctly identified as fraud to those falsely identified as fraud is actually the ratio of true positives to false positives (Didimo et al., 2014).

Detection power: the percentage of the ability of the method used; In fact, the ratio of true negatives to false negatives (Ahmed et al., 2016).

Table 2 displays the classification of fraud detection methods based on performance measurement criteria; Based on the methods used in the types of fraud and comparing the performance of each one, it has been discussed:

Table 2. *Classification of fraud detection methods based on performance measurement criteria*

Researcher	Fraud investigated	Technique used	Precision
Bhattacharyya (2011)	Credit card fraud (real example)	logistic model (regression)	96.6-99.4%
Olszewski(2014)	Credit card fraud (case study: Bank of Warsaw)	Support vector machine Random forest Self-organized maps	99.5-99.6% 97.8-99.6% %100
Halvaiee & Akbari (2014)	Credit card fraud (case study: Brazilian bank)	Artificial immune system	94.6-96.4%
Bermúdez et al. (2008)	Car insurance fraud (case study: Spanish insurance	Logistic regression	60.68%

	company)		
Kirkos et al.(2017)	Fraud of financial statements (case study: Greek manufacturing companies)	Joule Bayesian regression decision tree	99.83% 73.6%
Ravisankar et al. (2011)	Fraud in financial statements (case study: Chinese companies)	neural network Bayesian trust network Support vector machine	%80 90.3% 41.73-41.70%
Glancy & Yadav (2011)	Fraud in financial statements (case study: American companies)	Genetic algorithm neural network Group data analysis technique logistic model (regression) Text mining	89.27-94.14% 77.78-75.32% 88.14-93% 66.70-86.86% 95.65%
Cecchini (2010)	Fraud in financial statements (case study: American companies)	Text mining Support vector machine and text mining	41.75-45.80% 95.65%
Humpherys et al. (2011)	Fraud in financial statements (case study: American companies)	The combined method of text mining and decision tree	67.3%
Huang (2013)	Fraud in financial statements (case study: legal cases of Taiwanese companies)	The combined model of text mining and Bayesian trust network Combined model of support vector machine and text mining logistic model (regression)	67.3% 65.8% 19-79%
Bose & Wang (2007)	Fraud in financial statements (case study: Chinese companies)	Support vector machine CART neural network	71-92% 72.38% 77.14%

By examining the researches in the field of fraud and fraud detection with data mining and text mining methods, it is clear that fraud detection is one of the most important activities in the accounting department of companies. Internal control in fraud detection ensures that errors and irregularities are prevented or detected at the right time to stop the fraud. Since the factors considered in this research have not been investigated before on the internal control of Tehran Stock Exchange companies, this research studies the effect of 6 influencing factors on the internal control of the mentioned companies based on data mining techniques.

3. Method

This mixed methodology uses two analysis techniques: correlation and data mining. In the first step, based on a standard questionnaire, the impact of specific components on accounting fraud has been identified. Library search and document review were used to collect information on literature and research background. Therefore, by reading books and articles and searching on the Internet, the required information was collected. The prepared questionnaire includes 7 main indicators, with 48 questions separately for each of the variables. The questionnaire was distributed by the researcher among the accountants of the companies admitted to the Tehran Stock Exchange. The statistical population of the research is all the accountants working in companies admitted to the Tehran Stock Exchange. In this step, 400 people were selected randomly from the available sample. Table 3 shows the descriptive statistics of the participants.

Table 3*. frequency of characteristics of sample members*

Characteristics of people	Subgroup	Percent	Number
Gender	Man	65	260
	Female	35	140
Marital status	Single	43.8	70
	Married	56.3	90
Age groups	21 to 29 years	27.8	112
	30 to 39 years	35.7	144
	40 to 49 years	31.8	128

	More than 50 years	4.7	19
Education	Sub-diploma and diploma	2.2	9
	Associate Degree	4.7	19
	Masters	45.7	184
	Masters	8.5	155
	PhD and above	8.9	36
Work Experience	1-2 years	9.9	40
	3-5 years	23.6	95
	6-10 years	22	90
	More than 10 years	44.2	178

Based on the library technique, the conceptual and operational definition of response and explanatory variables has been discussed. In the following, these variables and their operational definition have been examined.

Table 4. Items related to measuring accounting fraud

	object
Accounting fraud detection	An individual's probability of committing fraud depends on their moral concern for the fraudulent behavior.
	A person's ambition to beat his competitor can be a possible motive for committing fraud.
	A variety of personal reasons can lead to the tendency of accounting fraud.
	A person's feeling of jealousy can motivate them to commit cheating behavior.
	Jealousy is also one of the most common drivers of accounting fraud.
	Strong dislike and hatred towards people or organizations can lead someone to commit fraudulent acts.
Effectiveness of internal control	The structure of the organization, the responsibility of the work task and the accessibility of the organization are among the most important factors of the internal control system of the organization.
	The competence of a management in an organization leads to a healthy and controlled management.
	The level of management accountability functions shows the effectiveness of the organization's internal control.
	Effective communication in management helps to avoid misunderstandings and allows management in organizations to access information.
	Frequent communication between employees and management gives both participants a better understanding of existing problems and allows management to track progress and updates.
	The authenticity of the information indicates that the information is reliable and therefore proven to be correct and real.
	Activity evaluation allows management to track and control all activities in the organization.
	Internal audit helps the organization to detect errors and fraud reports.
	The effectiveness of the internal control system in the organization depends on the internal control structure and how it is implemented in the organization.
Compensation system	The compensation system is essential in the company.
	The compensation system helps to recognize the organization in the successful implementation of work tasks.
	The compensation system has an effect on promotion and career advancement.
	The compensation system is essential in completing work tasks and achieving company goals.
	There should be equality and justice in the distribution of compensation system among employees.
Asymmetry of information	Who has better information about the activities under the responsibility of the employees than the outsiders of the company in this field?
	Who is familiar with input and output relations of part of the responsibilities towards groups outside the company?
	Who has better knowledge of the potential performance of responsibilities than groups outside the company?
	Who is more familiar with the technical work of responsibility than the groups outside the company?
	Who has more knowledge about the impact of external factors in the workplace than groups outside the company?
	Who has a better understanding of what can be done in the workplace than groups outside the company?
Compliance with accounting rules	Does compliance with global standards create executive responsibility?
	Does compliance with accounting standards have an effect on public interest?
	Does compliance with accounting standards create integrity, trust and integrity?
	Does compliance with accounting standards affect objectivity and impartiality and reality?
	Does compliance with accounting standards have an effect on prudence and consideration?
	Does compliance with accounting standards affect the confidentiality of accounting issues?
	Does compliance with accounting standards create stability in the company?
	Is it effective to have a standard accounting method in the company?
Management ethics	The manager is oriented in following the rules and punishments and does not escape punishment.
	The manager does not have a personal orientation in terms of benefit and profit for himself.
	In order to avoid not being accepted, the manager does not follow the customs and rules of the group and does not follow the morals of the group.
	The manager with regard to the social contract in line with the welfare of all and the benefit of all; Respecting

	the dignity and individuality of people respects ethics.
	The manager adheres to moral principles regardless of his position due to his conscience.
Ethical principles of accounting	I am an honest person in performing accounting services.
	I am careful, competent and persistent in performing my duties and always increase my knowledge and professional skills.
	I will not use or disclose information without the express permission of the employer or employer, and I will pay attention to the confidentiality of the information.
	It reports the services performed without any interest and completely neutral and based on reality.
	I avoid conflicts of interest, from simple to complex issues such as fraud and illegal practices.

In the next step, based on the data mining method, the impact of 7 explanatory components on accounting diagnosis were identified. The analysis was done in Rapidminer software. The steps of conducting research to implement the model are as follows (West et al., 2016):

First step: data determination

In this step, the data set is determined.

Second step: initial data review

By using expert knowledge and by calculating information such as data weight, average, data center, data analysis is performed.

Third step: creating and training the model

After creating the model, it can be trained.

Fourth step: Creating knowledge

The created model has the knowledge that it has learned from the training data set.

This knowledge contains the data structure and knows the patterns in it.

Step 5: Test the model

The obtained knowledge is tested for the data set from which no information is available.

The research variables were categorized and entered into the software

4. Findings

Inferential findings

In PLS software, one of the indicators for confirming relationships and investigating the effects of variables (in a structural model) is the significance of path coefficients. The significance of the path coefficient is the complement of the magnitude and direction of the beta coefficient of the model. If the obtained value above the minimum statistic is considered at the 95% confidence level (that is, the t-value is not in the range: -1.96 to 1.96), that relationship or hypothesis is confirmed.

In the present study, in order to investigate the simultaneous effects of the independent variables of the research on the dependent variable, the t-value output of the model is taken and presented in the following figure.

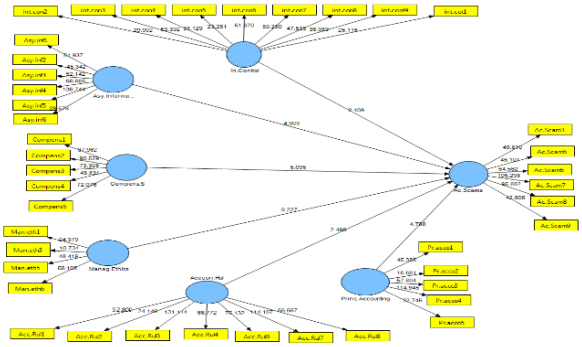

Figure 2. The output of t-values in the significance mode of the main effects model (the effect of independent variables on dependent variables)

Accounting fraud (Ac.Scams), effectiveness of internal control (In.control), asymmetry of information (Asy.Information), compensation system (Compens.S), compliance with accounting rules (Accoun.Rul), management ethics (Manag.Ethics), and ethical principles of accounting (Princ.Accounting)

As it can be seen in Figure (2), the values of path coefficients (t) of all independent variables affecting the dependent variable (accounting fraud) are not in the range of -1.96 to 1.96 and are significant. Therefore, it can be said that in the current study, the independent research variables included: effectiveness of internal control, compensation system, information asymmetry, compliance with accounting rules, management ethics, and ethical principles of accounting affect the dependent variable of the research (accounting fraud).

Further, in order to find out the extent of these effects, the output of the path coefficients (beta) was taken from the PLS software, which is shown in the figure (3).

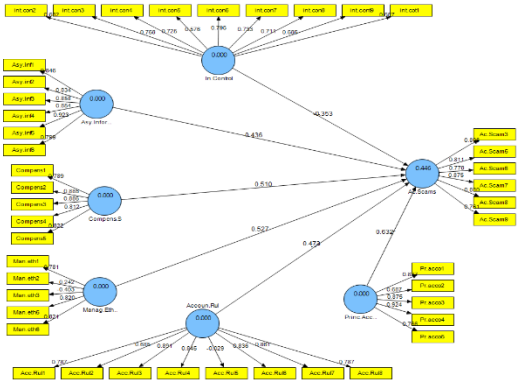

Figure 3. The output of the path coefficients of the main effects model (the effect of independent variables on dependent

The test results show that the six components identified based on the PLS outputs are verified. In the following, according to the confirmation of the six primary relationships of the research, we will examine the qualitative indicators of the structural and general model's desirability. In this regard, we use the coefficient of determination R^2 and predictor correlation criterion to check the desirability of the structural model, as well as the general model goodness of fit test (GOF) for the desirability of the overall model.

The degree of desirability of the structural model

The basic criterion for evaluating the endogenous variables of the model is the coefficient of determination R^2. In this research, to determine this value, we refer to the R Square table from the output of the PLS algorithm.

Table 5. The value of determining the dependent variable of accounting fraud

Main effects model	coefficient of determination R^2	Model usefulness
The effects of independent research → variables Accounting fraud	0.446	Appropriate and acceptable

Also, the desirability of the structural model can be calculated using the predictive correlation criterion Q^2or Aston-Geisler. Based on this criterion, the model should be able to predict the indicators of the reflective endogenous current variables.

Table 6. The output of the value of the predictor correlation criterion (Construct Cross Validated Redundancy)

Total	SSO	SSE	1-SSE/SSO
Ac.Scams Accounting fraud	396.96	275.39	0.31

Table 5 shows the output of the value of the correlation criterion for predicting the dependent variable of accounting fraud; Since the value of the accounting fraud variable is +0.31, it shows the strong predictive power of the model regarding this variable and confirms the appropriate fit of the structural model of the research once again.

The degree of desirability of the overall model

GOF= $\sqrt{\mu Communalities * \mu R^2} = 0.519$

According to the calculation of the goodness of fit, which is equal to 0.519, it can be concluded that the overall model has a suitable and strong fit.

Data mining

Data preprocessing is an important part of data mining analysis. To perform data mining, the databases of the mentioned companies were used in the period from 2019 to 2021. Data is entered into the database on a daily basis, and outliers are removed from the database monthly. Finally, the data used before being entered into the software to detect accounting fraud has been confirmed by the software in terms of accuracy. The pre-processing steps for each algorithm are described in the evaluation section. The algorithm for each model is available in the Rapid miner software. Modeling with software is shown in Figure 1. If necessary, functions and features can be added to the software.

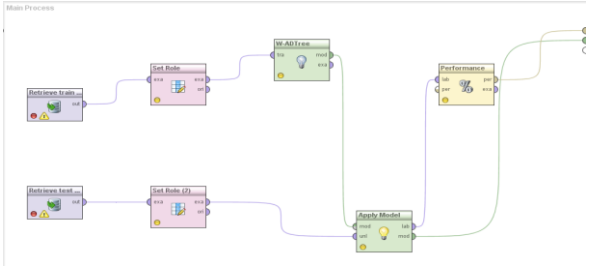

Figure 3. Tree modeling with Rapidminer software

In this section, all IB1, IBK, LWL, KSTAR and KNN algorithms are simulated and evaluated in the data mining model using software, and the evaluation results are shown in Tables 7 to 14. In this model, according to finding based on the nearest neighborhood, the only data set that gave the best answer for the models was that the data was entered into the software as a database, and the default data is in nomial format. They turn around. So, for all the algorithms of this model, the fourth type of data set has been used.

Algorithm IB1

This algorithm uses the normalized Euclidean distance to find the nearest neighbor.

Table 7. Confusion matrix of IB1 algorithm

	true anomaly	true normal	class precision
pred.anomaly	8681	334	96.30%
pred. normal	4152	9377	69.31%
class recall	67.65%	96.56%	

Table 8. Evaluation criteria and results of IB1 algorithm

F-measure	squared_error	precision	recall	accuracy
0.8244	0.199 +/- 0.399	82.80%	82.10%	80.10%

This algorithm uses the normalized Euclidean distance to find the nearest neighbor and also the value of k is obtained based on cross-validation. Distances in this algorithm can be weighted.

Table 9. Evaluation criteria and results of IBK algorithm

F-measure	squared_error	precision	recall	accuracy
0.8277	0.186 +/- 0.380	83.10%	82.45%	%۴۷80.

Table 10. Confusion matrix of IBK algorithm

	true anomaly	true normal	class precision
pred.anomaly	8745	315	96.52%
pred. normal	4088	9396	69.68%
class recall	68.14%	96.76%	

LWL algorithm

This algorithm uses the normalized Euclidean distance to find the nearest neighbor. In such a way that weight has been attributed to the features.

Table 11. Confusion matrix of LWL algorithm

	true anomaly	true normal	class precision
pred.anomaly	8623	746	92.04%
pred. normal	4210	8965	68.05%
class recall	67.19%	92.32%	

Table 12. Evaluation criteria and results of LWL algorithm

F-measure	squared_error	precision	recall	accuracy
0.8244	0.369 +/- 0.212	80.04%	79.76%	78.02%

KSTAR algorithm

Algorithm based on distance in which the entropy function is used to calculate the distance. The criterion of this algorithm is the similarity of the data with the training data.

Table 13. Confusion matrix of KST algorithm

	true anomaly	true normal	class precision
pred.anomaly	8814	316	96.54%
pred. normal	4019	9395	70.04%
class recall	68.68%	96.75%	

Table 14. Evaluation criteria and results of KSTAR algorithm

F-measure	squared_error	precision	recall	accuracy
0.8299	0.361 +/- 0.177	83.29%	%۷۱.82	80.77%

Figure 4 shows all the criteria of accuracy, precision, recall and F for the data mining model.

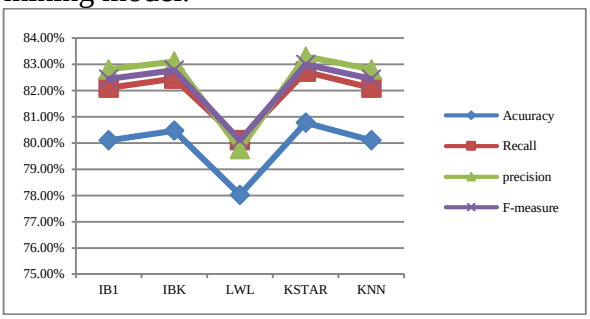

Figure 4. Evaluation chart of Kahl model algorithms according to different parameters

In evaluating parameters and according to the graphs, K-STAR algorithm has better performance than other algorithms.

5. Discussion

The purpose of the current research was to use data mining to detect accounting fraud in the information system of stock exchange member companies. First, the effectiveness of the identified components was tested using correlation and regression techniques.

Later on, the effectiveness of the proposed model based on the identified components was investigated using the data mining techniques and the related algorithms.

The use of automated fraud detection systems is essential for stock exchanges that manage online transactions. This reduces losses and increases customer confidence. As shown in the proposed research model, with the use of big data and artificial intelligence, new opportunities have been created in the use of advanced machine learning models to detect fraud. KNN, KSTAR, LWL, IBK and IB1 algorithms are popular methods in identifying and detecting fraud in financial statements. Zhang et al., (2021), showed that the combination of data mining methods reduces financial errors and accounting performance. Yang et al. (2021), in a data mining weighting method, determined that data mining and artificial intelligence methods have a much higher efficiency than traditional and time-consuming methods. Fanai et al. (2023) showed that the used deep learning classifiers trained on the transformed dataset obtained by the deep autoencoder significantly outperformed their baseline classifiers trained on the original data in terms of all performance measures. Rai et al. (2020), proved the effectiveness of the KSTAR method among fraud detection techniques.

6. Conclusion

Nowadays, accounting frauds are always changing and evolving; Therefore, fraud detection machine mechanisms should also continuously increase their effectiveness and efficiency by using available specialized knowledge. Detecting accounting fraud using current detection methods is becoming more and more difficult. An all-knowing CEO who intends to commit a fraud has enough resources to easily circumvent the system and is able to find any neutralize the type of discovery mechanism. The proposed data mining model for financial fraud detection showed that since the amount of data creation in financial companies is increasing day by day with the development of technology, it is possible to recognize the data and analyze it early detection of fraud can be achieved. Collections related to fraud in the stock market generally have a combination of several discrete and continuous variables. These data contain information about users in computer networks. This combination of discrete and continuous variables has increased the necessity of using data mining and machine learning methods in this field. The value of fraud detection is a function of time. In this way, the sooner the fraud is detected, the easier it will be to prevent it. Therefore, there is a need to find techniques that have the ability to detect fraud as quickly as possible in various fields, including accounting fraud.

Based on the results, the following suggestions are made:

- Since internal control affects the amount of accounting fraud, it is suggested that managers identify the individual values of the organization's accountants by using scientific methods and create and strengthen the values causing a sense of belonging and organizational commitment in them. Moreover, managers can try to prioritize employees whose individual values are more consistent with the organization's values.

- Improving customer service by managing knowledge related to the customers. Investment in designing secure accounting software which categorizes and detects hidden knowledge in this field. This software design allows for various reports required by stakeholders, or creating the necessary facilities and equipment for modern archiving and considering the necessary solutions to reduce direct contact between employees but increase knowledge share among them with the least contact.

Disseminating financial knowledge clearly can be effective in improving the relationship between internal environment control and company efficiency. Therefore, there should be a clear policy for managers and shareholders to disseminate financial information and knowledge with higher quality and better transparency.

-By applying a proper knowledge system, managers can manage a counter-measure system that affects the amount of accounting fraud adequately. Furthermore, this software

should be designed such that decreases the need of conducting an annual audit and show the knowledge and information situation up to date. In other words, it should be able to decrease the reliance on annual audits.

- Making accessible knowledge and information by providing appropriate policies and tools can lead the companies to better profit. So that information asymmetry can be reduced and investors have more incentive to invest due to the possibility of informed decisions.

- Since conservatism is one of the components of the quality characteristic of reliability; it can hide the required knowledge needed by managers and decision-makers. Therefore, it can be useful for accounting standard setters who are trying to reduce conservatism.

- Holding training courses, and regular classes and providing the latest legal changes to learn and comply with legal requirements by accountants can help the progress. Accountants also need to be taught the importance of financial information which can make critical knowledge for an organization. Knowledge management courses in line with software classes can help. No matter how strong a system is, the role of human resources cannot be ignored, specifically in this field. Strengthening social responsibility, professional ethics, material and moral incentives and motivations, and disciplinary proceedings are significant in training accountants.

As mentioned earlier, it is suggested that managers increase confidentiality and impartiality by improving the atmosphere of trust and establishing interaction and mutual trust. This is one of the essential factors in the knowledge management field which can make profit and advantage for organizations. Organizations are required to have a knowledge management team to manage all connectable strings by knowledge

The accounting department needs to be updated and apply tools and methods to obtain knowledge from its sources. Also, using a deep auto-encoder as a pre-processing step to obtain a better representation of the input data, exploiting new deep learning classification models to detect credit card fraud, using the Bayesian optimization algorithm to optimize the meta-parameters of the used deep learning models, and conducting extensive experiments to evaluate the effectiveness of the proposed approach are suggested

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

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

References

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, *68*, 90-113. http://dx.doi.org/10.1016/j.jnca.2016.04.007, 2016.

Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems*, *55*, 278-288. https://doi.org/10.1016/j.future.2015.01.001

Akila, S., & Reddy, U. S. (2018). Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection. *Journal of computational science*, *27*, 247-254.https://doi.org/10.1016/j.jocs.2018.06.009.

Bermúdez, L., Pérez, J. M., Ayuso, M., Gómez, E., & Vazquez, F. J. (2008). A Bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics*, *42*(2), 779-786. https://doi.org/10.1016/j.insmatheco.2007.08.002

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision support systems*, *50*(3), 602-613. https://doi.org/10.1016/j.dss.2010.08.008

Bose, I., & Wang, J. (2007). Data mining for detection of financial statement fraud in Chinese Companies. In *International joint Conference on e-Commerce, e-*

Administration, e-Society, and e-Education. International Business Academics Consortium (IBAC) and Knowledge Association of Taiwan (KAT).

Brandek, S., Mohammadi, F. (2018). The effect of weak internal control on the asymmetric behavior of general, administrative and sales costs. *Accounting and Management Perspective Quarterly*. (2) 17, 38-53. [in Persian]

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Making words work: Using financial text as a predictor of financial events. *Decision support systems*, *50*(1), 164-175. https://doi.org/10.1016/j.dss.2010.07.012

Chen, H., Ai, H., Yang, Z., Yang, W., Ye, Z., & Dong, D. (2020, September). An Improved XGBoost Model Based on Spark for Credit Card Fraud Prediction. In *2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)* (pp. 1-6). IEEE. https://doi.org/10.1109/IDAACS-SWS50031.2020.9297058

Didimo, W., Liotta, G., & Montecchiani, F. (2014). Network visualization for financial crime detection. *Journal of Visual Languages & Computing*, *25*(4), 433-451. https://doi.org/10.1016/j.jvlc.2014.01.002

Donyaee, H., Hossini, S. E., & Mohamadi, N. (2019). Identifying of Influential Factors on Money Laundering in Sports. *Sport Management Studies*, *11*(55), 17-34.

Fanai, H., & Abbasimehr, H. (2023). A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. *Expert Systems with Applications*, 119562. https://doi.org/10.1016/j.eswa.2023.119562

Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing*, *99*, 106883.https://doi.org/10.1016/j.asoc.2020.106883

Glancy, F. H., & Yadav, S. B. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems*, *50*(3), 595-601. https://doi.org/10.1016/j.dss.2010.08.010

Halvaiee, N. S., & Akbari, M. K. (2014). A novel model for credit card fraud detection using Artificial Immune Systems. *Applied soft computing*, *24*, 40-49. https://doi.org/10.1016/j.asoc.2014.06.042

Herawati, N. (2015). Application of Beneish M-Score models and data mining to detect financial fraud. *Procedia-Social and Behavioral Sciences*, *211*, 924-930. https://doi.org/10.1016/j.sbspro.2015.11.122

Huang, S. Y. (2013). Fraud detection model by using support vector machine techniques. *International Journal of Digital Content Technology and its Applications*, *7*(2), 32. https://doi.org/10.4156/jdcta.vol7.issue2.5

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, *50*(3), 585-594. doi.org/10.1016/j.dss.2010.08.009

Kewei, X., Peng, B., Jiang, Y., & Lu, T. (2021, January). A hybrid deep learning model for online fraud detection. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 431-434). IEEE. https://doi.org/10.1109/ICCECE51280.2021.9342110

Khalilpour, M., Kamyabi, Y., Nabavi Chashmi, A. (2018). The effect of ethical approaches of accountants on the quality of accounting information with the moderating role of company size and financial leverage. *Financial accounting and audit research*. 11(43), 118-91. [in Persian].

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert systems with applications*, *32*(4), 995-1003. doi.org/10.1016/j.eswa.2006.02.016

Liang, Z., Pan, D. & Xu, R. (2019). Knowledge representation framework of accounting event in corpus-based financial report text. *Cluster Comput* **22** (4), 9335–9346. https://doi.org/10.1007/s10586-018-2153-8

Lodi, G., Aniello, L., Di Luna, G. A., & Baldoni, R. (2014). An event-based platform for collaborative threats detection and monitoring. *Information Systems*, *39*, 175-195. https://doi.org/10.1016/j.is.2013.07.005

Ofori-Dwumfuo, G. O., & Gyimah, E. (2013). Business intelligence as an ICT tool for financial crime detection in Ghana. *Research Journal of Information Technology*, *5*(3), 62-71. https://doi.org/10.19026/rjit.5.5789

Olszewski, D. (2014). Fraud detection using self-organizing map visualizing the user profiles. *Knowledge-Based Systems*, *70*, 324-334. https://doi.org/10.1016/j.knosys.2014.07.008

Pargmann, J., Riebenbauer, E., Flick-Holtsch, D. *et al*. (2023). Digitalisation in accounting: a systematic literature review of activities and implications for competences. *Empirical Res Voc Ed Train* **15**, 1. https://doi.org/10.1186/s40461-023-00141-1

Rai, A. K., & Dwivedi, R. K. (2020, July). Fraud detection in credit card data using unsupervised machine learning based scheme. In *2020 international conference on electronics and sustainable communication systems (ICESC)* (pp. 421-426). IEEE. https://doi.org/10.1007/978-981-15-6318-8_31.

Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision support systems*, *50*(2), 491-500. doi.org/10.1016/j.dss.2010.11.006

Sundarkumar, G. G., & Ravi, V. (2015). A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Engineering Applications of Artificial Intelligence*, *37*, 368-377. https://doi.org/10.1016/j.engappai.2014.09.019

Wahid, A., Msahli, M., Bifet, A., & Memmi, G. (2023). NFA: A neural factorization autoencoder based online telephony fraud detection. *Digital Communications and Networks*.doi.org/10.1016/j.dcan.2023.03.002

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, *57*, 47-66. https://doi.org/10.1016/j.cose.2015.09.005

Wieder, B., & Ossimitz, M. L. (2015). The impact of Business Intelligence on the quality of decision making–a mediation model. *Procedia Computer Science*, *64*, 1163-1171. https://doi.org/10.1016/j.procs.2015.08.599

Xie, Z., Chiu, D.K.W. & Ho, K.K.W. (2023) The Role of Social Media as Aids for Accounting Education and Knowledge Sharing: Learning Effectiveness and Knowledge Management Perspectives in Mainland China. *J Knowl Econ*. doi.org/10.1007/s13132-023-01262-4

Yang, Z., Yin, C., Jin, C., Ning, J., & Zhou, J. (2021, May). Lightweight delegated authentication with identity fraud detection for cyber-physical systems. In *Proceedings of the 7th ACM on Cyber-Physical System Security Workshop* (pp. 17-28). https://doi.org/10.1145/3457339.3457984

Yue, D., Wu, X., Wang, Y., Li, Y., & Chu, C. H. (2007, September). A review of data mining-based financial fraud detection research. In *2007 International Conference on Wireless Communications, Networking and Mobile Computing* (pp. 5519-5522). Ieee. https://doi.org/10.1109/WICOM.2007.1352

Zhang, X., Han, Y., Xu, W., & Wang, Q. (2021). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*, *557*, 302-316. https://doi.org/10.1016/j.ins.2019.05.023

Zhu, X., Vanden Broucke, S., Zhu, G., Vanthienen, J., & Baesens, B. (2016). Enabling flexible location-aware business process modeling and execution. *Decision Support Systems*, *83*, 1-9. https://doi.org/10.1016/j.dss.2015.12.003