Int. J. Nonlinear Anal. Appl. 16 (2025) 1, 341–358
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.28789.3989

Presenting a comprehensive model for predicting the type of audit opinion from machine learning algorithms: Evidence from Tehran Stock Exchange

Alireza Rahimzadeh[a], Mehran Matinfard[a,*], Zohreh Hajiha[b], Ehsan Rahmaninia[a]

[a]*Department of Accounting, North Tehran Branch, Islamic Azad University, Tehran, Iran*

[b]*Department of Accounting, East Tehran Branch, Islamic Azad University, Tehran, Iran*

(Communicated by Mohammad Bagher Ghaemi)

Abstract

The main aim of the represented research is to provide a comprehensive model for predicting the type of audit opinion based on a number of machine learning algorithms in some companies in the Tehran Stock Exchange. In order to achieve this goal, 1,606 company-years (146 companies for 11 years) observations collected from the annual financial reports of companies admitted to the Tehran Stock Exchange from 2010 to 2020 have been tested. In this study, six machine learning algorithms (decision tree and regression, random forest, neural network, nearest neighbor, logit regression, support vector machine) and also two methods of selecting the final variables of the research (two samples mean comparing test, forward step-by-step selection method) has been used for the model creation. The results show that the overall accuracy of decision tree and regression, random forest, neural network, nearest neighbor, logit regression, and support vector machine procedures respectively are 78.7%, 77.7%, 76.9%, 74.6%, 78.3%, and 76.7%. Regarding the obtained outcomes, the decision tree and regression algorithm outperform in forecasting the type of audit opinion compared to other studied methods. Meanwhile, in general, the result of variable selection techniques illustrates that the step-by-step method is far more effective. Hence, in the studied companies in the Tehran Stock Exchange, the step-by-step method and the decision tree and regression algorithm provide the most efficient model for the prediction of the audit opinion type.

Keywords: type of audit opinion, prediction, machine learning algorithm
2020 MSC: 68T07, 60G25

1 Introduction

The transition from an undeveloped economy requires investment facilitation and optimal allocation of excess funds and resources. In economic systems based on the private sector or systems that take steps to increase the share of the private sector in performing economic activities, This task is often performed appropriately by the capital markets. However, expanding the scope of economic activities, separating ownership from management, increasing competition

*Corresponding author
Email addresses: rahimzadeh.al@gmail.com (Alireza Rahimzadeh), mehran.matinfard2@gmail.com (Mehran Matinfard), drzhajiha@gmail.com (Zohreh Hajiha), ehsanrahmaninia@gmail.com (Ehsan Rahmaninia)

Received: September 2022 *Accepted:* November 2022

in the private sector capital market, pure profiteering, and the existence of grounds for abuse lead some companies (managers) to provide unhealthy and misleading information. Auditing as a controlling activity and as one of the mechanisms of corporate governance is necessary to gain (and maintain) investors' trust in the existence of information symmetry, the fairness of the market, and the fairness of available information. Conducting an audit, provided that it has a suitable quality level, solves the need of users of financial reports to evaluate the quality of information before using them in the decision-making process and provides a suitable basis for making favorable economic decisions. Financial statements audited by independent auditors are a very suitable means of transmitting reliable information. The independent auditor is the most qualified person to comment on the reliability of the financial reports of the economic unit. The merit of the auditor is due to the fact that he is an independent person and performs the audit according to the auditing standards to make sure that the items of the financial statements are prepared based on the accounting standards. Therefore, the auditor accredits the claims prepared by another person in the form of financial statements. Thereby, it leads to an increase in the reliability of the information used in economic decisions [39].

Auditors' comments are the final product of the audit process. The role of auditors has evolved to provide adequate and appropriate assurance to control the economic issues of companies. Undoubtedly, the management's responsibility in ensuring the adequacy of the company's internal controls is still firmly in place. On the other hand, auditing is one of the basic elements of the accountability process, because accountability requires the existence of reliable and valid information, and the information requires to be reviewed by someone independent from the information provider to guarantee its reliability. This important matter is done through the auditing process, which creates added value in the process of answering and auditing with relevant comments by determining the validity of the information. In this way, communication between the auditors' findings and people outside and inside the company is created in the form of an audit opinion, which plays an important role in warning the users of financial statements to recognize the problems facing the company. If the company has any flaws and distortions, the auditing firm will not easily ignore them and it will influence the audit opinion [13]. Auditors modify their opinions when they encounter problematic issues such as ambiguities, lack of agreement in the application of accounting standards, limitations in processing, and doubts about the continuity of activity. In this way, auditors' adjustment reports are a sign of low-quality financial statements from the users' perspective. Any decrease in the quality of the financial statements causes the adjustment of the audit report. Koskivaara [33] believes that advanced auditing tools can prevent the manipulation of companies' accounts, and help accountants respond to today's demands of the business environment. The increasing number of management frauds has doubled the need to use these tools and new decision-making methods. Bell and Tabor [6] and Chen and Church [10] believe that auditors can use the output of these decision-making methods to plan audit procedures in order to achieve an appropriate level of audit risk. Machine learning or data mining algorithms, whose purpose is to extract wanted information from a wide set of data, have been widely used as an active decision-making tool. This has increased the attention of researchers in enriching data in order to facilitate the discovery of wanted patterns [50]. Data mining is extracting or exploring knowledge from a large amount of information. Strong patterns or rules discovered through data mining techniques can be used for the non-obvious prediction of new data. Therefore, data mining is an interdisciplinary field that uses statistical model analysis tools, mathematical algorithms, and machine learning methods to discover previously unknown valid patterns in large data sets [15]. According to the stated content, the present research aims to provide a model for predicting the type of audit opinion using machine learning algorithms in companies admitted to the Tehran Stock Exchange. To achieve the mentioned goal, the information of 146 companies from 2010 to 2020 is used. The findings of the techniques display that the overall accuracy of decision tree and regression, random forest, neural network, nearest neighbor, logit regression, and support vector machine algorithms are 78.7%, 77.7%, 76.9%, and 74.6%, respectively. It can be concluded that the performance of the decision tree and regression algorithm is far better than other investigated algorithms. What is more, the outcomes of the variable selection process indicate that the step-by-step technique is effective in this research. Thus, in the Tehran Stock Exchange companies, the step-by-step method and the decision tree and regression algorithm provide the most efficient model for forecasting the type of audit report.

It is expected that the results of this research can have scientific achievements and added value as follows: First, the results of this research can lead to the expansion of theoretical foundations in the field of auditing, especially audit opinion. Second, the results of this research can provide auditors with appropriate information about the factors affecting the type of audit opinion through additional training. Thirdly, it provides guidelines to the authorities that compile audit standards and capital market rule-makers. Finally, the results of the research can suggest new ideas for conducting new research in the field of predicting the type of audit opinion.

2 Theoretical foundations and research background

Agency problems between shareholders and managers lead to the employment of auditors, who provide independent assurance to shareholders that the financial statements prepared by company managers are in accordance with generally accepted accounting principles [49]. In addition to this, auditing through the discovery of expropriation by people inside the organization [38], and benefiting the company's management through messaging about the reliability of financial information prepared by the management, plays an important role in law enforcement and rights protection played by investors [22]. Due to the provision of audit services, independent auditors play an important role in maintaining public interests in the capital markets. Audit reports are one of the things that are used as a guide for investors when making decisions, and the low quality of these reports can lead to inappropriate allocation of resources in society. Playing the role of the protector of society's interests makes the auditor remain independent. In fact, the quality and credibility of the audit report depend on the independence degree of the auditor from his employer; But, in practice, the management influence of the considered company, on the issues related to the selection and presentation of information to the auditor, along with the high motivations of the management to reach the set goals or to cross them, are important obstacles on the way of auditors to remain independent [48].

As mentioned, the final product of the audit process is the auditors' comments. They modify their reports whenever they face challenging issues like uncertainties, lack of agreement in the accounting standards application, restrictions in processing, and suspicions about the continuousness of activity. Such a modification leads to the adjustment of the audit report which causes the auditors' reports to seem low-quality from the point of view of users of financial statements. In this way, the adjustments in the auditor's opinion indicate some accounting issues that have the potential to increase the amount of uncertainty in the company's current and future profits which subsequently causes a lower profit. By adjusting their comments, auditors can reflect problematic issues that ultimately change the content of companies' profits. Companies that receive a modified opinion have lower earnings than companies that receive an acceptable opinion. Furthermore, different types of audit adjustments are expected to be associated with different levels of information content.

Koskivaara [33] is of the opinion that progressive auditing implements are able to avoid the manipulation of the accounts by companies which assists accountants to meet the needs of today's business environment. The requirements for utilizing these implements and new decision-making approaches have doubled because management frauds have increased overwhelmingly. Bell and Tabor [6] and Chen and Church [10] claim that the results of decision-making methods can be employed by auditors for planning audit procedures to obtain a proper level of audit risk. In addition to this, the results of these methods can be used in the evaluation of potential businessmen, surveying of competitors, checking of quality control, forecasting of audit quality, and the statement of accountants in similar conditions and as a defense in law firms. Machine learning algorithms have been extensively utilized as an active decision-making instrument in different studies in the field of accounting.

Faizizadeh [16] investigated the effect of audit evidence on the auditor's report. The empirical findings of binary logistic regression in their work showed that the adequacy of audit evidence is negative but insignificant and the validity of audit evidence has a positive coefficient sign. The low value of McFadden's coefficient of determination indicated that the null hypothesis was rejected and the alternative hypothesis was accepted. This study showed that more experimental activities should be done in this field.

Faruzandeh et al. [18] in a study concluded that the low quality of financial statements based on the intensity of renewal criteria never affected the type of audit report, increasing the number of condition clauses, emphasizing specific items and other explanatory clauses. In this way, according to the theory of limited attention, the auditors have not issued an adjusted audit report with more conditional and explanatory clauses for financial statements containing many distortions and errors that are presented in the next year with high intensity, so that the user of the financial statements is of high quality. It is low and therefore the ability to rely upon and trust the audit reports based on the criterion of the intensity of renewal is low.

Amri et al. [4] in one research came to the conclusion that the short-sightedness and ability of the managers were not related to the auditor's opinion and the change of the auditor in the year after the modified opinion. They also concluded that the ability of the management was not able to establish the relation between the short-sightedness of the management and the issuance of the modified audit report and to moderate the relation between management's short-sightedness and the probability of auditor change after issuing a modified opinion. Among the reasons for the lack of communication, it can be said that it is difficult to identify real manipulations of profits by auditors due to their nature and the lack of influence of auditors on the ability of management, the different nature of Iran's audit market compared to other countries, etc.

In a study, Jamei and Lotfiju [25] discovered that political connections have a negative effect on the auditor's

opinion. The probability of issuing a favorable report by auditors in non-government companies is more than in government companies. The existence of institutional shareholders weakens the effect of political communication on the auditor's opinion, and managerial ownership also weakens the effect of political communication on the auditor's opinion, but if there are strict penalties for political communication, it does not have a significant effect on the auditor's opinion.

Fakhari and Amiri [17] in their research entitled "Causes and Consequences of Buying an Auditor's Opinion" came to the conclusion that the special characteristics of the audit work can be an important stimulus for intensifying the phenomenon of buying an auditor's opinion. Also, other results of the research indicate that buying the auditor's opinion leads to adverse consequences, including reducing the quality of auditing and the quality of financial reporting.

Mohammad Rezaei et al. [36] did a research titled "Type of audit report, number and type of conditional clause of qualified audit report: the role of economic crisis".They analyzed the information of 142 companies admitted to the Tehran Stock Exchange from 2007 to 2016, and they concluded that during the economic crisis compared to the pre-crisis period, the number of qualified audit reports and the paragraphs' number of the audit report do not increase. However, testing the types of conditional paragraphs in the audit opinion showed that there is a significant and direct relationship among the economic crisis and the paragraphs of insufficient reserves for income tax, incorrect calculation of the cost price, insufficient reserve for doubtful receivables, non-identification of bearing costs, and insufficient reserves to reduce the investment value.

Khajavi et al. [27] investigated and compared the usefulness of different variable selection methods in predicting the type of auditors' opinions of companies listed on the Tehran Stock Exchange. In this regard, the performance of variable selection methods (including correlation-based, step-by-step diagnostic analysis, t-test, Relief, and factor analysis) has been investigated and compared with each other. The used classifiers also include the support vector machine method and artificial neural networks. Empirical findings related to the review of 1214 observations (company-year) accepted in the Tehran Stock Exchange between 2007 and 1993 indicate the usefulness and positive effect of using variable selection methods on the performance of predicting the type of auditors' comments and also the existence of a significant difference between the usefulness of these methods. In other words, if the selected variables of these methods are used, compared to the use of 35 primary variables, the average accuracy increases, and the type 1(MSE "mean square error") and 2(RMSE) errors decrease. Finally, the findings of the research indicate a better and more suitable performance of support vector machine procedure than neural networks.

Abbaszadeh et al. [1] in a study with 980 observations through data analysis during the years 2009 to 2015 and using 33 financial and non-financial variables including liquidity indicators, asset management, profitability, debt management, market value, company growth and size, employee productivity, corporate governance, bankruptcy, company performance and other influential factors (company life and industry type) in line with the prediction of the auditor's opinion. They concluded that the neural network optimized with the colonial competition algorithm with 94.98 percent predictive accuracy compared to other methods has the best performance in predicting the type of independent auditor's opinion. Also, the results show that changing the type of independent auditor's opinion, the type of previous year's audit report, investment return, current ratio, debt ratio, loss of the company, price to income, and net profit ratio have the greatest effect in predicting the type of independent auditor's opinion.

Setayesh et al. [45] implemented a study through data analysis of 842 companies during the period from 2001 to 2010 and came to the conclusion that the support vector machine method with 76% prediction accuracy compared to other methods has the best performance in forecasting. It is the type of auditors' comments. Checking the error of the first and second types of each method also shows the better performance of the support vector machine.

Ahmadpour et al. [2] investigated the effect of financial and non-financial variables on the issuance of conditional audit opinions by using neural networks. The results showed that financial variables have the greatest impact on the issuance of conditional statements.

Pourheidari and Azami [41] identified the type of auditors' comments using neural networks. The results show that the neural network has a better performance in identifying the auditor's report type and logistic regression provides an unbalanced model in identifying the auditor's comment types.

Setayesh and Jamalianpour [46] investigated the Equation between financial and non-financial ratios and the type of auditor's opinion. The findings of the research indicated that the distribution of independent variables of companies with different comments is not the same in most cases. Also, the results show that among all the financial and non-financial variables, the type of performance has the most Equation with the type of auditor's opinion.

He et al. [24] conducted one research titled "Do goodwill impairments affect audit opinions? Evidence from China".Through the analysis of the information of companies listed on the Chinese stock exchange from 2007 to 2017,

they concluded that by reducing the amount of goodwill, the probability of providing a conditional audit opinion increases. Secondly, the positive association between goodwill impairment and conditional audit opinion is mainly driven by earnings management risks. Thirdly, this positive association is more pronounced when the auditors are industry experts and there is no mismatch between the auditor and the client. Fourthly, auditors are more sensitive to the amount of goodwill impairment. In general, empirical evidence showed that auditors perceive a decline in goodwill as a signal of information risks and communicate their concerns to investors to avoid lawsuits.

Golmohammadi Shuraki et al. [21] conducted a study entitled "Accounting comparability, financial reporting quality and audit opinions: evidence from Iran". By analysis of information on companies admitted to the Tehran Stock Exchange between 2015 and 2019, they concluded that there is an adverse correlation between accounting comparability and audit opinion criteria. The results also showed that there is an adverse relationship between the quality of financial reporting and audit opinion.

Averio [5] did one research entitled "The analysis of influencing factors on the going concern audit opinion–a study in manufacturing firms in Indonesia". Via analysis of information of companies listed on the Indonesian Stock Exchange from 2015 to 2019, they achieved that leverage has a positive effect on the going concern audit opinion. Whereas, audit quality, profitability, and liquidity have a negative effect on the audit opinion of the going concern. The size of the company and the delay of the audit do not affect the audit opinion of the going concern.

Stanisic et al. [47] in research titled "Predicting the type of auditor opinion: Statistics, machine learning, or a combination of the two?" commented on the relative forecasting performance of different forecasting methods with the aim of overcoming the methodological limitations of previous studies and predicting the type of auditor's opinion. The predictive performance of twelve models selected from the fields of statistics and machine learning is evaluated separately for two common real-world scenarios: a) when prior information about the employer (i.e., the type of statement received in the past) is available; and can be used in forecasting and b) if such information is not available (eg new companies). The results show that in the first scenario, the comparable predictive performance is around 0.89 which is measured by the area under the curve (AUC). However, in the second scenario, machine learning algorithms, especially tree-based examples, such as random forest, perform much better, achieving an AUC of up to 0.79. Finally, they developed the predictive performance of two hybrid models with the aim of combining the strengths of both statistics (i.e. interpreting results) and machine learning (i.e. handling a large number of predictors and improving accuracy). The full method is demonstrated to be reproducible using the largest empirical data ever used in this research stream and includes 13,561 pairs of annual financial statements and related audit reports.

Zhang [52] in a research titled "Predict audit quality using machine learning algorithms" analyzed the information of 14,028 company years during the period from 2008 to 2016 in the United States of America and used different machine learning algorithms. He concluded that the random forest algorithm can predict audit quality more accurately than logistic regression, and audit-related variables have a better function than financial variables in predicting audit quality.

Using a sample of 530 English and Irish companies, Zdolšek and Jagrič [51] presented a model for determining conditional audit opinion and evaluated it by receiver operating characteristic curve (ROC) and misclassification cost. They concluded that using auditors' subjective preferences is sufficient to evaluate the performance of the model, based on these performance criteria.

Using a sample of 450 Irish and English companies, Kirkos et al. [30] used three data mining methods including the multi-layer Perceptron method, decision tree and Bayesian network to classify auditors' comments. The results of this research showed the overall higher performance of the Bayesian network than other methods.

3 Research method

3.1 How to collect data

This research is practical in terms of purpose. The research design is quasi-experimental and post-event and it is done using historical information. Publications, books, and available databases have been used to collect information on the theoretical foundations of the research. Also, the data required for analysis have been collected from the software "Rehavard Navin" and the websites of "Research, Development and Islamic Studies Management of the Stock Exchange Organization", Kodal, Central Bank, and the Statistics Center of Iran.

3.2 Statistical population, statistical sample, and time frame of the research

The statistical population of the current research is the companies accepted in the Tehran Stock Exchange. The time period of the research is from 2010 to 2020. In this research, the systematic target method was used to determine the statistical sample, and the applied criteria are as follows:

1. The financial year of the company must end at the end of March of each year and the company has not changed the financial year during the review period.
2. Companies should not be part of the investment, holding, leasing, credit institutions, and banks.
3. Their information and data should be available during the research period.
4. According to the mentioned limitations, the statistical sample of this research includes 146 companies.

Table 1: Sampling method

505	The total number of companies admitted to the Tehran Stock Exchange
89	The number of companies whose financial year does not end at the end of March or have changed their financial year during the research period.
118	The number of companies that were in the group of investment companies, holding, leasing, credit institutions, and banks.
152	The number of companies whose complete information is not available during the research period.
146	Number of sample companies

4 The method of measuring research variables

Dependent variable (response):

The dependent variable is the type of audit opinion, which is a dummy variable; if the audit opinion is acceptable, it is given a value of one, and otherwise, a value of zero.

Independent variables (predictor):

One of the distinctive and innovative features of the current research is that it examines the factors affecting the type of audit opinion at the micro level and the macroeconomic level. Accordingly, the operational definition of these variables is mentioned below:

Size of the company (SIZE): It is equal to the natural logarithm of the company's total assets.

Liquidity of the company (LIQID): It is equal to the ratio of current assets to current liabilities.

Debt ratio of the company (DEBT): It is equal to the ratio of the total liabilities of the company to the total assets of the company.

Profitability of the company (PROF): It is equal to dividing the operating profit by the total assets of the company.

Age of the company (AGE): It is equal to the natural logarithm of the number of years the company has been established until the current year.

The company's growth opportunity (GROWTH): It is equal to the ratio of the market value of equity to the book value of equity.

Industry size (INDSIZE): It is equal to the natural logarithm of the total industrial assets that the company owns [29].

Number of firms in an industry (INDNUM): It is equal to the natural logarithm of the number of firms belonging to an industry [29].

Decentralization of the industry (product market competition) (PMC): In this research, the Herfindahl-Hirschman index is used to measure product market competition. The Herfindahl-Hirschman index is obtained from the square root of the market share of all companies active in the industry in the form of the following formula [11]:

$$HHI = \sum_{i=1}^{k} S_i^2 \tag{4.1}$$

where HHI is the Herfindahl-Hirschman index, k is the number of companies active in the market, and S_i is the market share of the i^{th} company, which is obtained from the following Equation:

$$S_i = \frac{X_i}{\sum_{l=1}^{n} X_j} \tag{4.2}$$

where X_j represents the sales of the company j and l represents the type of industry. The Herfindahl-Hirschman index measures the degree of industry concentration. The bigger this index is, the more concentration and the less competition exist in the industry, and vice versa. It should be noted that this index in the research of Chen et al. [11]; Chen et al. [12]; Namazi and Ebrahimi [37], and Khajavi et al. [28] have also been used.

The size of the board of directors (BS): It is equal to the natural logarithm of the number of members of the company's board of directors [3].

The independence of the board of directors (BI): It is equal to the ratio of the number of non-obligatory members in the board of directors of the company to the number of members of the board of directors of the company [3].

Financial expertise of the board of directors (FSPEC): It is a dummy variable, if at least one of the board of directors has a degree in accounting, auditing, financial management, and economics, the number is one, and otherwise, the number is zero [3].

Type of ownership (government or private) (TypeOWN): It is a dummy variable if more than 50% of the company's shares are owned by the private sector, it takes a value of one, and otherwise, it takes a value of zero [29].

Ownership of institutional shareholders (INSOWN): It is obtained by dividing the number of shares owned by institutional shareholders to the total number of issued shares of the company.

Major shareholder ownership (BOWN): It is obtained by dividing the number of shares owned by the largest shareholder of the company to the total number of issued shares of the company.

Managerial ownership (MOWN): It is obtained by dividing the number of shares owned by the company's managers to the total number of issued shares of the company.

Existence of audit committee (AC): It is a dummy variable, if the company has an audit committee, it is given a value of one and otherwise, a value of zero.

The size of the audit committee (ACSIZE): It indicates the number of members of the audit committee.

Audit Committee Independence (ACIND): It is obtained by dividing the independent members of the audit committee to the total number of members of the audit committee.

Financial expertise of the audit committee (ACSEP): It is obtained by dividing the members of the audit committee with financial expertise (having degrees in accounting, auditing, management, and economics) to the total number of members of the audit committee.

The size of the independent auditor (AUDSIZE): It is a dummy variable, if the independent auditor of the company, audit organization, or trusted auditing institutions of the stock exchange is ranked A, the number is one, and otherwise, the number is zero.

Independent Auditor Tenure (AUDTEN): It represents the natural logarithm of the number of years that an audit firm audits a company's financial statements consecutively.

Independent auditor's fee (AUDFEE): It is calculated through the natural logarithm of the fee paid to the auditor.

Independent auditor's report delay (ARL): It is Calculated through the natural logarithm of the difference between the number of days between the end of the financial year and the date of the audit report.

Quality of accruals (ACCQ): In the present study, to calculate the quality of accruals, the adjusted model of Jones [26] which is based on discretionary accruals and presented by Dechow et al. [14] (error rate) is optional as an accrual item. The mentioned model is as follows:

$$\frac{TAC_{i,t}}{TA_{i,t-1}} = a_1 \left(\frac{1}{TA_{i,t-1}}\right) + a_2 \left(\frac{\Delta REV_{i,t} - \Delta REC_{i,t}}{TA_{i,t-1}}\right) + a_3 \left(\frac{PPE_{i,t}}{TA_{i,t-1}}\right) + \varepsilon_{i,t} \tag{4.3}$$

$TAC_{i,t}$ = total accruals of the company i in year t, which is calculated as follows:

$$TAC_{i,t} = (\Delta CA_{i,t} - \Delta CL_{i,t} - \Delta CASH_{i,t} + \Delta STDEBT_{i,t} - DEPN_{i,t}) \tag{4.4}$$

$\Delta CA_{i,t}$ = changes in current assets of the company i in year t compared to year t-1.

$\Delta CL_{i,t}$ = changes in current liabilities of the company i in year t compared to year t-1.

$\Delta CASH_{i,t}$ = changes in cash of the company i in year t compared to year t-1.

$\Delta STDEBT_{i,t}$= changes in the current share of receivables of the company i in year t.

$DEPN_{i,t}$= compared to year t-1.

$TA_{i.t-1}$= depreciation cost of the company i in year t.

$\Delta REV_{i,t}$ = total assets of the company i in year t-1.

$\Delta REC_{i,t}$ = changes in sales of the company i in year t compared to year t-1.

$PPE_{i,t}$ = changes in accounts receivable of the company i in year t compared to year t-1.

In the above model, the residuals of the regression will be optional accrual items, with negative numbers indicating optimal profit management and positive numbers indicating non-optimal profit management [8, 42].

Earnings Sustainability (EP): Earnings sustainability means the repeatability of current earnings. A profit that is not caused by unusual and unexpected activities is more stable [31]. Sustainable profits are desirable from the perspective of investors because they will continue in future periods [35, 40, 43, 44]. In this research, profit stability is measured based on the model of Kormendi and Lipe [32] as follows:

$$\frac{Earn_{j.t}}{Total\ Assets_{j.t-1}} = \alpha + \delta_1 \frac{Earn_{j.t}}{Total\ Assets_{j.t-1}} + V_{j.t} \tag{4.5}$$

$Earn_{j.t}$ = Operating profit of company j in year t

$Earn_{j.t-1}$ = Operating profit of company j in year t-1

$Total\ Assets_{j.t-1}$ = Total assets of company j in year t-1

If the explanatory variable coefficient of the profit stability model (δ_1) is close to one or greater than one, it indicates high-profit stability, and if it is close to zero or smaller than zero, it indicates the instability of profit.

Profit predictability (EPRED): Profit predictability is the ability to predict profit from oneself [31]. In other words, it is the ability of current profit to predict future profit. Francis et al. [19] used the square root of the variance (standard deviation) of the error of the above model's profit stability equation to calculate profit predictability as follows:

$$predictability_{j.t} = \sqrt{\sigma^2(v_{\widehat{j.t}})} \tag{4.6}$$

Its smaller value indicates high-profit predictability. As a result, the quality of profit is also at a high level (and vice versa).

Profit smoothing (ES): In this research, in accordance with the research of Leuz et al. [34] and Francis et al. [20], profit smoothing is measured by dividing the standard deviation of net profit by the standard deviation of operating cash flows.

ES = smoothing profit

SD = net income standard deviation of net income

$SDCFO$ = deviation of operating cash flow criterion

$$ES = \frac{\text{SD net income}}{SD\ CFO} \tag{4.7}$$

Thus, the higher value of the ES variable indicates the operating income of the company has higher fluctuations [23].

Trading volume (TURN): It is obtained through the ratio of annual trading volume to the number of issued shares of the company [9, 42].

Systematic risk (SYSR): It is obtained by using the market model estimate of beta for each firm during the year [7]. The market model is as follows:

$$R_m = \alpha + \beta R_i + \varepsilon \tag{4.8}$$

where in:

R_m: market return

R_i: return on specific share i

Unsystematic risk (NSYSR): Unsystematic risk is calculated by estimating the residual standard deviation from the daily market model [7].

Market return (MRET): It is obtained by dividing the difference between the price index and the cash return of the Tehran Stock Exchange at the end of the year and the beginning of the financial year by the price index and cash return of the Tehran Stock Exchange at the beginning of the financial year.

Stock return (RET): It is calculated based on the comprehensive formula of stock return, the information about which is extracted from Rahavard Navin software.

Stock Liquidity (LIQ): In order to measure this variable, Amihud's illiquidity criterion, which was introduced by Yako Amihud in 2002, is used.

$$ILLIQ_t = (|r_t|)/DVol_t \tag{4.9}$$

in the above Equation, r_t is the return on the t^{th} day and $DVOL_t$ is the volume of rials traded on the t^{th} day. Due to the anuality of the research variables, the simple average of the above variable is calculated every year and used in the calculations. Also, for the criteria to be comparable, the result of the above Equation is multiplied by one billion rials [7].

5 Research findings

5.1 Descriptive statistics of research variables

In Table 2, panel A, some concepts of descriptive statistics of variables, including mean, median, minimum observations, maximum observations, and standard deviation are presented. For example, the average of the company's liquidity variable is 1.547, which has relatively low volatility according to the standard deviation.

The average debt ratio of the company is 0.570, which indicates that 57% of the company's financial resources are financed through debt. The average profitability of the company is 0.149, which has relatively low volatility.

According to panel B, the results of the t-test show that at the 95% confidence level, the variables of company liquidity, company debt ratio, company profitability, company growth opportunity, number of companies in the industry, decentralization of the industry, ownership of institutional shareholders, ownership of major shareholders, managerial ownership, size of the audit committee, independence of the audit committee, financial expertise of the audit committee, independent auditor's fee, delay in submitting the independent auditor's report, profit predictability, unsystematic risk, market return and stock return in companies with acceptable and unacceptable audit opinion have a significant difference.

At the 95% confidence level, the other variables in panel B are not significantly different in fraudulent and non-fraudulent companies.

With regard to panel P, the results of the chi-square test show that at the 95% confidence level, the variables of type of ownership, the existence of an audit committee, and the size of the independent auditor in companies with acceptable and unacceptable audit opinions have a significant difference.

Table 2: Descriptive statistics of research variables

Panel A: Continuous variables						
Variable names	**Observation No.**	**Mean**	**Median**	**Max**	**Min**	**standard deviation**
Company liquidity	1606	14.3872	14.2064	20.7687	10.0312	1.5692
Debt ratio of the company	1606	1.5468	1.3236	3.8128	0.6408	0.7805
Company profitability	1606	0.5702	0.5872	0.9042	0.1857	0.2001
Life of the company	1606	0.1496	0.1242	0.8422	-0.3271	0.1413
Company growth opportunity	1606	3.8342	2.4044	14.9813	0.7190	3.6828
Industry size	1606	18.254	18.258	21.7156	12.9523	1.6808
Number of companies in the industry	1606	19.9346	22	31	8	9.2778
Decentralization of the industry (product market competition)	1606	0.1627	0.1537	0.3876	0.0121	0.1311
Board size	1606	5.0261	5	7	3	0.2379
Independence of the board of directors	1606	0.6612	0.6	1	0	0.1826
CEO tenure	1606	3.8076	3	12	1	3.2175
Ownership by institutional shareholders	1606	56.6578	67.65	93.76	0	32.5127

Major shareholder ownership	1606	50.2054	51	87.308	14.84	20.3338
Property management	1606	62.4581	68.0553	92.02	9.0022	23.9985
The size of the audit committee	1606	2.1382	3	5	0	1.4787
Independence of the audit committee	1606	0.4912	0.6667	1	0	0.3675
Financial expertise of the audit committee	1606	0.6354	1	1	0	0.4438
Auditor tenure	1606	4.2248	3	13	1	4.0942
Audit fees	1606	5.0312	6.4738	9.9331	3.5684	3.1991
Delay in submitting the report of the independent auditor	1606	4.3054	4.4006	5.4553	2.8332	0.3823
Quality of accrual items	1606	-0.0068	-0.0174	0.3951	-0.3621	0.1862
Sustainability of profits	1606	0.4550	0.3578	2.2196	-0.7641	0.7446
Ability to predict profit	1606	0.0774	0.0627	0.2002	0.0157	0.0508
Profit smoothing	1606	1.0721	0.8428	3.3466	0.1610	0.8419
Turnover	1606	0.5166	0.2835	2.1146	0.0112	0.5894
Systematic risk	1606	0.7086	0.6189	2.3259	-0.5490	0.7787
Unsystematic risk	1606	0.1333	0.1187	0.2936	0.0341	0.0716
Market return	1606	0.6472	0.4684	1.9294	-0.209	0.6464
Stock returns	1606	0.8457	0.3349	4.8057	-0.324	1.3324
Stock liquidity	1606	0.00030	0.00040	0.0031	0.000147	0.000747

Panel B: t-test in companies with acceptable and unacceptable audit opinion ($*$ Significance at 90% confidence level, $**$ significance at 95% confidence level, and $***$ significance at 99% confidence level)

Variable names	**Acceptable comment (number of observations = 825)**	**Unacceptable comment (number of observations = 781)**	**t-statistics**
Size of the company	14.3735	14.4017	-0.358
Company liquidity	1.6232	1.466	4.067***
Debt ratio of the company	0.5538	0.5876	-3.392***
Company profitability	0.1859	0.1113	10.978***
Life of the company	3.617	3.6391	-1.218
Company growth opportunity	4.3157	3.3255	5.457***
Industry size	18.2514	18.257	-0.066
Number of companies in the industry	20.6594	19.169	3.222***
Decentralization of the industry (product market competition)	0.1368	0.1902	-8.323***
Board size	5.0339	5.0179	1.355
Independence of the board of directors	0.6611	0.6613	-0.031
CEO tenure	3.8824	3.7286	0.958
Ownership by institutional shareholders	62.7412	50.2317	7.823***
Major shareholder ownership	54.1665	46.0212	8.156***
Property management	66.8775	57.7897	7.684***
The size of the audit committee	2.3115	1.9552	4.845***
Independence of the audit committee	0.5386	0.4411	5.349***
Financial expertise of the audit committee	0.6929	0.5746	5.373***
Auditor tenure	4.257	4.1908	0.324
Audit fees	5.3663	4.6772	4.332***
Delay in submitting the report of the independent auditor	4.2021	4.4145	-11.613***
Quality of accrual items	0.0011	-0.0153	1.767*
Sustainability of profits	0.4901	0.418	1.951*
Ability to predict profit	0.0802	0.0744	2.276**
Profit smoothing	1.0728	1.0714	0.033
Turnover	0.4921	0.5425	-1.715*
Systematic risk	0.6774	0.7415	-1.647*
Unsystematic risk	0.1368	0.1297	-2.011**
Market return	0.7142	0.5764	4.307***
Stock returns	0.9467	0.739	-3.141***
Stock liquidity	0.0003	0.0003	-1.368

Panel P: Chi-square test in companies with acceptable and unacceptable audit opinions

Variable names	**Acceptable comment (number of observations = 825)**	**Unacceptable comment (number of observations = 781)**	**Chi-square statistic**
Type of ownership (private vs. government)	279	404	52.655***
Existence of an audit committee	621	492	28.423***
The size of the independent auditor	236	176	7.753***

5.2 Examining the validity of research variables

According to Table 3, based on the "Generalized Dickey-Fuller" and "Phillips-Prone" tests, as the probability value of all variables was less than 5%, all independent, dependent, and control variables were at a stable level during the research period. Reliability means that the average and variance of the research variables over time and the variance of the variables have been constant between different years. As can be seen in Table 3, all variables are stable and there is no need for a collinearity test. Therefore, the problem of false regression in estimated coefficients will not exist. In meaningful false regression, the coefficients are pseud.

Table 3: Results of Manay test of research variables

Variables	Generalized Dickey Fuller		Results	Pilips-Peron		Results
	statics	Possibility		statics	Possibility	
Type of audit opinion	-18.7318	0.0000	Stable	-20.6000	0.0000	Stable
Size of the company	-8.1904	0.0000	Stable	-12.8992	0.0000	Stable
Company liquidity	-15.3458	0.0000	Stable	-15.3686	0.0000	Stable
Debt ratio of the company	-14.0237	0.0000	Stable	-14.2461	0.0000	Stable
Company profitability	-7.1454	0.0000	Stable	-20.2018	0.0000	Stable
Life of the company	-7.7101	0.0000	Stable	-10.8929	0.0000	Stable
Company growth opportunity	-7.6919	0.0000	Stable	-24.8883	0.0000	Stable
Industry size	-8.1230	0.0000	Stable	-8.3598	0.0000	Stable
Number of companies in the industry	-2.9315	0.00420	Stable	-3.1064	0.0263	Stable
Decentralization of the industry (product market competition)	-10.7722	0.0000	Stable	-5.7241	0.0000	Stable
Board size	-18.1918	0.0000	Stable	-14.2000	0.0000	Stable
Independence of the board of directors	-27.2163	0.0000	Stable	-18.8808	0.0000	Stable
CEO tenure	-18.9844	0.0000	Stable	-15.5850	0.0000	Stable
Type of ownership (private or public)	-11.7293	0.0000	Stable	-11.9679	0.0000	Stable
Ownership by institutional shareholders	-9.3695	0.0000	Stable	-9.6146	0.0000	Stable
Major shareholder ownership	-10.6323	0.0000	Stable	-11.1101	0.0000	Stable
Property management	-12.9715	0.0000	Stable	-13.0327	0.0000	Stable
Existence of an audit committee	-7.8739	0.0000	Stable	-18.3123	0.0000	Stable
The size of the audit committee	-7.8732	0.0000	Stable	-17.8094	0.0000	Stable
Independence of the audit committee	-6.8755	0.0000	Stable	-19.7699	0.0000	Stable
Financial expertise of the audit committee	-7.4031	0.0000	Stable	-18.8893	0.0000	Stable
The size of the independent auditor	-11.2745	0.0000	Stable	-11.6676	0.0000	Stable
Auditor tenure	-15.1103	0.0000	Stable	-15.2839	0.0000	Stable
Audit fees	-15.3238	0.0000	Stable	-15.4479	0.0000	Stable
Delay in submitting the report of the independent auditor	-12.3532	0.0000	Stable	-17.5561	0.0000	Stable
Quality of accrual items	-37.0543	0.0000	Stable	-38.8338	0.0000	Stable
Sustainability of profits	-28.0488	0.0000	Stable	-28.0215	0.0000	Stable
Ability to predict profit	-16.6856	0.0000	Stable	-16.9509	0.0000	Stable
Profit smoothing	-20.2019	0.0000	Stable	-20.7886	0.0000	Stable
Turnover	-6.0407	0.0000	Stable	-20.2115	0.0000	Stable
Systematic risk	-8.8193	0.0000	Stable	-35.2746	0.0000	Stable
Unsystematic risk	-5.5988	0.0000	Stable	-33.5259	0.0000	Stable
Market return	-17.8342	0.0000	Stable	-39.9380	0.0000	Stable
Stock returns	-8.0851	0.0000	Stable	-48.2162	0.0000	Stable
Stock liquidity	-31.6963	0.0000	Stable	-31.7857	0.0000	Stable

5.3 The process of creating research models

In order to achieve the goals of the research, first of all, the desired models in the current research include models created with decision tree and regression (CART), random forest (RF), neural network (KNN), nearest neighbor (ANN), logit regression (LR), and support vector machine (SVM) algorithms, and using the selected variables based

on the mean comparison test of two samples, we will create the forward step-by-step selection method and compare the results.

5.3.1 The final variables of the research

In order to determine the final variables of the research among the primary variables of the research, the average comparison test of two samples and the forward step-by-step selection method has been used. The difference between these two methods is that in the method of comparing the mean of two samples, the significance of the difference between the means for each independent variable is checked without considering the relation between other independent and dependent variables. While, in the forward stepwise selection method, the process starts with an empty set of features, and in each repetition step, the best main features are selected and added to the previous set, and its purpose is to select variables that maximize the coefficient of determination (R^2) of the model. SPSS software was used to perform the above statistical tests.

Two samples mean comparison test

In order to compare the average of two samples, the variance of the two samples must first be compared, in other words, the test of equality of variances precedes the test of equality of means. In order to equalize variances, we use Levin's test, which is based on Fisher's statistics. In this test, the data distribution does not need to be normal.

The t-statistic is calculated to test the equality of the mean of two samples, in the case of equal and unequal variance of the two samples. In the case of equality of variances, we use equation (5.1) to calculate the t statistic, in which the degree of freedom is equal to $df = n_1 + n_2 - 2$.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - u_2)}{S_p \left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)} \tag{5.1}$$

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \tag{5.2}$$

In the case of inequality of variances, the t statistic is calculated from equation (5.3) and the degree of freedom is calculated from equation (5.4):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - u_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{5.3}$$

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_1}\right)^2}{n_2 - 1}} \tag{5.4}$$

where n_1 and n_2 are the number of members of the first and second samples, S_1 and S_2 are the standard deviations of the first and second samples. Then, based on the t number calculated according to the above relations and the t number according to the table, we will interpret the results according to the 5% error level, if the calculated t number is smaller than the t number according to the table ($p - value < 0.05$), is accepted that the difference between the averages of the two groups is meaningful, and otherwise, it will be rejected. The results of comparing the averages of the two samples are shown in table 3. According to the above table, it can be stated, the difference between the averages of two samples of companies with acceptable and unacceptable audit opinions for the variables of company liquidity, company debt ratio, company profitability, company growth opportunity, number of companies in the industry, decentralization of the industry, type Ownership, ownership of institutional shareholders, ownership of major shareholders, managerial ownership, the existence of audit committee, size of the audit committee, independence of audit committee, financial expertise of audit committee, size of the independent auditor, independent auditor's fee, delay in presenting independent auditor's report, ability to predict profit, unsystematic risk, market returns, stock returns are meaningful at the 95% confidence level. In this regard, according to the results of the above test, it can be stated that 21 variables mentioned in table 3, which p-value is less than 5%, have been introduced as final variables according to the two samples' average comparison test.

Forward step-by-step selection

In the forward step-by-step selection method, we are looking for the estimation of a logit regression based on the primary variables of the research in such a way that the value of the coefficient of determination of the regression is

maximum, for this purpose, the process starts with a null set of features and in each repetition step, the best feature is selected and added to the previous set, and its purpose is to select variables that do not maximize the coefficient of determination of the model, this process continues until the entry of each remaining variable into the model is not changed in order to improve the coefficient of determination of the regression.

Using the above method and after 13 steps, the following variables were selected as final variables:

- Size of the company
- Company profitability
- Company growth opportunity
- Decentralization of the industry
- Board size
- CEO tenure
- Ownership by institutional shareholders
- Ownership of major shareholders
- Financial expertise of the audit committee
- The size of the independent auditor
- Independent auditor's fee
- Delay in submitting the report of the independent auditor
- Market returns

5.3.2 Implementation of models

In order to implement and evaluate the techniques, we must divide the research data into two categories: training data and test data. We will use the training data set for construction and test data to evaluate the model created in the training phase. Classification techniques can be compared based on criteria such as accuracy, speed, and stability. The accuracy of a classification method depends on the number of correct predictions that the model makes. The speed of a classification method is the time required to build and use the model in classification, and it shows the stability of the model's ability to deal with unusual data or missing values.

In order to evaluate the accuracy and stability of the models, we divided the research data 50 times randomly into two categories of training and testing data and created the model, and evaluated the results. In this way, 75% of the data is used for training the model and 25% of it is used for evaluating and testing the model, and the average results obtained from 50 times running of each model are considered as its final result.

Table 4: Simulation parameters

The amount	**Simulation parameter**
1606 year-company	The number of samples is the total data
21 features	The number of characteristics of each sample (T-test) - the first set
13 features	The number of features of each sample (Stepwise) - the second set
1205	The number of model training examples
401	The number of model test samples

5.3.3 Model evaluation parameters

The detection of a classification method on data that has only two classes is expressed by one of the following states: true positive (TP), false negative (FN), true negative (TN), and false positive (FP). Based on this, the accuracy (or precision) parameter can be defined. Accuracy is the standard metric for summarizing the recognition performance in all classes, which is calculated as the following formula.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.5}$$

As shown in table 5, to calculate the classification accuracy, the disturbance matrix must be calculated first. Data on the matrix main diameter are used as correct detections of the classification method.

Table 5: disturbance matrix

True Positives (TP)	False Positives (FP)
False Negatives (FN)	True Negatives (TN)

In the following, the confusion matrix of each classification method in predicting financial fraud (with two different sets of features) is shown.

The results of the decision tree and regression algorithm (CART), which is presented in the form of a disturbance matrix in Table 6, show the amount of model error in identifying companies with acceptable audit opinions according to the T-test and Stepwise test. It is approximately 25.1% and 10.7%, respectively, so according to the method of selecting variables to implement the decision tree and regression algorithm (CART), the Stepwise method is more efficient. The results of the overall accuracy index of the model show that CART and Stepwise compared to CART and T-test has higher accuracy. In general, the results of the decision tree and regression algorithm (CART) show that the accuracy of this model in detecting and predicting the type of audit opinion of companies admitted to the Tehran Stock Exchange is approximately 78.7%. This shows that according to the variables selected by the Stepwise method and the decision tree and regression algorithm (CART), we can identify companies with acceptable audit opinions with a probability of 78.7%.

Table 6: Confusion matrix in detecting the type of audit opinion (decision tree and regression algorithm)

CART & T-test			**CART & Stepwise**		
	Positive	Negative		Positive	Negative
Positive	0.749	0.2	Positive	0.893	0.5
Negative	0.251	0.8	Negative	0.107	0.5
Accuracy	0.763		Accuracy	0.787	

The results of the random forest (RF) algorithm, presented in the form of a disturbance matrix in Table 7, show the error rate of the model in identifying companies with acceptable audit opinions according to the T-test and Stepwise test; the error rate is approximately 9.3% for both methods. So according to the method of selecting variables, in implementation of the Random Forest (RF) algorithm the T-test and Stepwise methods are not different. The results of the overall accuracy index of the model show that FR and T-test is more accurate than FR and Stepwise. In general, the results of the Random Forest (RF) algorithm show that the accuracy of this model in detecting and predicting the type of audit opinion of the companies admitted to the Tehran Stock Exchange is approximately 77.7%. This illustrates that with Paying attention to the variables selected by T-test method and random forest algorithm (RF), we can identify companies with acceptable audit opinion with a probability of 77.7%.

Table 7: Confusion matrix in detecting the type of audit opinion (random forest algorithm)

RF & T-test			**RF & Stepwise**		
	Positive	Negative		Positive	Negative
Positive	0.907	0.577	Positive	0.906	0.576
Negative	0.093	0.423	Negative	0.093	0.423
Accuracy	0.777		Accuracy	0.775	

The results of the neural network (KNN) algorithm, which is presented in the form of a disturbance matrix in Table 8, show the error rate of the model in identifying companies with acceptable audit opinions according to the T-test and Stepwise test is approximately 12.1%. Thus, according to the method of selecting variables, T-test and Stepwise methods are not different to implement the neural network (KNN) algorithm. The results of the overall accuracy index of the model show that KNN and T-test has higher accuracy than KNN and Stepwise. In general, the results obtained from the neural network algorithm (KNN) show that the accuracy of this model in detecting and predicting the type of audit opinion of the companies admitted to the Tehran Stock Exchange is approximately 76.9%, this shows Considering the variables selected by T-test method and neural network algorithm (KNN), the probability of identifying companies with acceptable audit opinion is 76.9%.

The results of the nearest neighbor (ANN) algorithm, which is presented in the form of a disturbance matrix in Table 9, show the amount of model error in identifying companies with acceptable audit opinions according to the T-test and Stepwise is approximately 16.7% and 16.4%. Hence, according to the method of selecting variables to

Table 8: Confusion matrix in detecting the type of audit statement (neural network algorithm)

KNN & T-test			**KNN & Stepwise**		
	Positive	Negative		Positive	Negative
Positive	0.879	0.531	Positive	0.878	0.554
Negative	0.121	0.469	Negative	0.121	0.446
Accuracy	0.769		Accuracy	0.762	

implement the nearest neighbor (ANN) algorithm, the Stepwise method is more efficient. The results of the overall accuracy index of the model depict that ANN and T-test is more accurate than ANN and Stepwise. In general, the results of the nearest neighbor algorithm (ANN) show that the accuracy of this model in detecting and predicting the type of audit opinion of companies listed on the Tehran Stock Exchange is approximately 74.6%, this shows that paying attention to the variables selected by T-test method and the nearest neighbor algorithm (ANN), the possibility of identifying companies with acceptable audit opinion is 74.6%.

Table 9: Confusion matrix in detecting the type of audit opinion (nearest neighbor algorithm)

ANN & T-test			**ANN & Stepwise**		
	Positive	Negative		Positive	Negative
Positive	0.833	0.492	Positive	0.836	0.554
Negative	0.167	0.508	Negative	0.164	0.446
Accuracy	0.746		Accuracy	0.731	

The results of the Logit Regression (LR) algorithm, which is presented in the form of a disturbance matrix in Table 10, show the amount of model error in identifying companies with acceptable audit opinions according to the T-test and Stepwise test is approximately 10.2% and 9.9%, respectively. So, according to the method of selecting variables to implement the Logit Regression (LR) algorithm, the Stepwise method is more efficient. The results of the overall accuracy index of the model show that LR and T-test is more accurate than LR and Stepwise. In general, the results of the Logit Regression (LR) algorithm show that the accuracy of this model in detecting and predicting the type of audit opinion of companies listed on the Tehran Stock Exchange is approximately 78.3%, this indicates that Considering the variables selected by T-test method and Logit Regression (LR) algorithm, identifying companies with acceptable audit opinion with a probability of 78.3% is probable.

Table 10: Confusion matrix in detecting the type of audit opinion (logit regression algorithm)

LR & T-test			**LR & Stepwise**		
	Positive	Negative		Positive	Negative
Positive	0.898	0.531	Positive	0.901	0.569
Negative	0.102	0.469	Negative	0.099	0.431
Accuracy	0.783		Accuracy	0.775	

The results of the support vector machine (SVM) algorithm, which is presented in the form of a disturbance matrix in Table 11, show that the error rate of the model in identifying companies with acceptable audit opinion according to the T-test and Stepwise test is approximately 8.2% and 15.3%, so according to the method of selecting variables to implement the Logit Regression (LR) algorithm, the T-test method is more effective. The results of the overall accuracy index of the model show that SVM and Stepwise has higher accuracy than SVM and T-test. In general, the results of the support vector machine (SVM) algorithm show that the accuracy of this model in detecting and predicting the type of audit opinion of the companies listed in the Tehran Stock Exchange is approximately 76.7%, indicating that according to the variables selected by the Stepwise method and the Support Vector Machine (SVM) algorithm, it is possible to identify companies with acceptable audit opinions with a probability of 76.7%.

Table 11: Confusion matrix in detecting the type of audit opinion (support vector machine algorithm)

SVM & T-test			**SVM & Stepwise**		
	Positive	Negative		Positive	Negative
Positive	0.918	0.908	Positive	0.847	0.454
Negative	0.082	0.092	Negative	0.153	0.546
Accuracy	0.696		Accuracy	0.767	

In short, the results of machine learning techniques show that the overall accuracy of CART, RF, KNN, ANN, LR, and SVM techniques is 78.7%, 77.7%, 76.9%, 74.6%, 78.3%, and 76.7% respectively. which shows the efficiency of the decision tree and regression algorithm (CART) compared to other machine learning algorithms. Meanwhile, in

general, the results obtained from the variable selection method show the efficiency of the stepwise method. Therefore, in the companies admitted to the Tehran Stock Exchange, the stepwise method and the decision tree and regression algorithm (CART) provide the most efficient model for predicting the type of audit opinion.

6 Discussion and conclusion

The current research is looking for an answer to the question of whether it is possible to provide an effective and efficient approach for predicting the type of audit opinion using machine learning algorithms and selected variables based on theoretical foundations and research background. Through the analysis of information related to 146 companies admitted to the Tehran Stock Exchange from 2010 to 2020, the findings of machine learning techniques show that the overall accuracy of decision tree and regression, random forest, neural network, the closest Neighborhood, logit regression, support vector machine techniques are 78.7%, 77.7%, 76.9%, 74.6%, 78.3%, and 76.7%, respectively. The results depict that the decision tree and regression algorithm is efficient compared to other studied machine learning algorithms. Meanwhile, in general, the results of the variable selection method show the effectiveness of the step-by-step method. Therefore, in the companies admitted to the Tehran Stock Exchange, the step-by-step method and the decision tree and regression algorithm provide the most effective model for predicting financial fraud. These results are somewhat similar to the findings obtained in the research of Stanisic et al., [47]; Zhang, [52]; Khajavi et al., [27]; Setayesh et al., [45]; Setayesh and Jamalianpour, [46] are consistent.

The outcome of the presented work can lead to some achievements including:

(a) Causing the improvement of theoretical foundations in auditing, -particularly the audit viewpoint.

(b) Providing proper information for auditors about the elements that impact their opinion type with extra training.

(c) Making a set of effective guidelines available for the experts who arrange the audit standards and for the lawmakers of the capital market.

(d) Providing novel visualization for doing new studies in forecasting the audit opinion type.

References

[1] M. Abbaszadeh, M. Muftoonian, M. Babaei Kalarijani, and M. Fadaei, *Investigating the accuracy of heuristic algorithms and logit linear regression in predicting the type of independent auditor's opinion*, Quart. J. Modern Res. Account. Audit. **1** (2016), no. 4, 39–73.

[2] A. Ahmadpour, A.A. Taher Abadi, and S. Abbasi, *The effect of financial and non-financial variables on the issuance of conditional audit opinion (companies admitted to the Tehran Stock Exchange)*, Stock Exchange J. **3** (2009), no. 9.

[3] H.K. Alkdai and M.M. Hanefah, *Board of directors' characteristics and value relevance of accounting information in Malaysian shariah-compliant companies: A panel data analysis*, Econ. Finance Rev. **2** (2012), no. 6, 31–44.

[4] O. Amri, M. Shurvarzi, A. Masihabadi and A.R. Mehr Azin, *The role of management characteristics on the type of audit opinion and auditor-employer equations*, Audit Knowledge **21** (2021), no. 83, 246–266.

[5] T. Averio, *The analysis of influencing factors on the going concern audit opinion–a study in manufacturing firms in Indonesia*, Asian J. Account. Res. **6** (2020), no. 2, 152–164.

[6] T.B. Bell and R.H. Tabor, *Empirical analysis of audit uncertainty qualifications*, J. Account. Res. **29** (1991), no. 2, 350–370.

[7] B.M. Blau, T.J. Brough, and T.G. Griffith, *Bank opacity and the efficiency of stock prices*, Journal of Banking & Finance, **76** (2017), 32–47.

[8] J.L. Callen, M. Khan, and H. Lu, *Accounting quality, stock price delay, and future stock returns*, Contemp. Account. Res. **30** (2013), no. 1, 269–295.

[9] C. Cao, B. Liang, A.W. Lo, and L. Petrasek, *Hedge fund holdings and stock market efficiency*, Rev. Asset Pric. Stud. **8** (2018), no. 1, 77–116.

[10] K.C.W. Chen and B.K. Church, *Default on debt obligations and the issuance of going-concern opinions*, Audit.: J. Practice Theory **11** (1992), no. 2.

[11] J. Chen, W. Dong, S. Li, and Y. Zhang, *Perceived audit quality, state ownership, and stock price delay: Evidence from China*, Asia-Pac. J. Account. Econ. **25** (2018), no. 1–2, 253–275.

[12] L.R. Chen and H.L. Tao, *Empirical research on the impact of product market competition on internal control quality*, China Acad. J. **2013** (2013), 52–59.

[13] L.E. DeAngelo, *Auditor size and audit quality*, J. Account. Econ. **3** (1981), no. 3, 183–199.

[14] P.M. Dechow, R.G. Sloan, and A.P. Sweeney, *Detecting earnings management*, Account. Rev. **70** (1995), no. 2, 193–225.

[15] S. Dua and X. Du, *Data Mining and Machine Learning in Cybersecurity*, CRC Press, 2016.

[16] A. Faizizadeh, *The effect of audit evidence on the auditor's report*, Account. Knowledge Manag. Audit Quart. **38** (2021), 159–166.

[17] H. Fakhari and I. Amiri, *Causes and consequences of purchasing an auditor's opinion*, Account. Audit. Stud. **9** (2019), no. 36, 5–26.

[18] J. Faruzandeh, N. Izadinia, and S. Daie Karimzadeh, *The effect of the severity of financial statement renewal on the type of auditor's opinion and the paragraphs of the audit report (evidence of the quality of financial statements)*, Audit. Knowledge **82** (2021), no. 4, 298–321.

[19] J. Francis, R. LaFond, P.M. Olsson, and K. Schipper, *Costs of equity and earnings attributes*, Account. Rev. **79** (2004), no. 4, 967–1010.

[20] J. Francis, P. Olsson, and K. Schipper, *Earnings quality*, Foundations and Trends® in Accounting, **1** (2008), no. 4, 259–340.

[21] M. Golmohammadi Shuraki, O. Pourheidari, and M. Azizkhani, *Accounting comparability, financial reporting quality and audit opinions: evidence from Iran*, Asian Rev. Account. **29** (2021), no. 1, 42–60.

[22] A. Habib, A.H. Muhammadi, and H. Jiang, *Political connections, related party transactions, and auditor choice: Evidence from Indonesia*, J. Contemp. Account. Econ. **13** (2017), no. 1, 1–19.

[23] S.A. Hashemi, S. Farahmand, and N. Jashouqani, *The effect of conditional conservatism on cost of equity capital*, Iran. J. Account. Knowledge **7** (2011), 20.

[24] Z. He, D. Chen and J. Tang, *Do goodwill impairments affect audit opinions? Evidence from China*, China J. Account. Res. **14** (2021), no. 2, 151–182.

[25] R. Jamei and N. Lotfijo, *The effect of political communication on the auditor's opinion with an emphasis on the market characteristics of Tehran Stock Exchange companies*, Karafan Sci. Quart. **18** (2022), no. 2, 55–72.

[26] J. Jones, *Earnings management during import relief investigations*, J. Account. Res. **29** (1991), 193–228.

[27] S. Khajavi, M. Kazem Nejad, A.A. Dehghani Saadi, and A. Mumtazian, *Investigating the usefulness of different methods of selecting predictor variables in predicting the type of auditors' comments*, Exper. Account. Rese. **7** (2017), no. 27, 81–102.

[28] S. Khajavi, G. Mohsenifard, G. Rezaei, and S.D. Hosseini Rad, I*nvestigating the effects of competition in the product market on the profit management of companies listed in the Tehran Stock Exchange*, Sci.-Res. Quart. Asset Manage. Financ. **1** (2012), no. 3, 119–134.

[29] J.B. Kim, Z. Yu and H. Zhang, *Can media exposure improve stock price efficiency in China and why?*, China J. Account. Res. **9** (2016), no. 2, 83–114.

[30] E. Kirkos, C. Spathis, A. Nanopoulos, and Y. Manolopoulos, *Identifying qualified auditors' opinions: a data mining approach*, J. Emerg. Technol. Account. **4** (2007), no. 1, 183–197.

[31] G. Kordestani and Z. Majdi, *Investigation of the relationship between quantitative return attributes and the expected return rate of common stock investors*, Account. Audit. Investigations **14** (2007), no. 48, 85–104.

[32] R. Kormendi and R. Lipe, *Earnings innovations, earnings persistence, and stock returns*, J. Bus. **60** (1987), no.

3, 323–345.

[33] E. Koskivaara, *Artificial neural networks in analytical review procedures*, Manage. Audit. J. **19** (2004), no. 2, 191–223.

[34] C. Leuz, D. Nanda, and P.D. Wysocki, *Earnings management and investor protection: an international comparison*, J. Financ. Econ. **69** (2003), no. 3, 505–527.

[35] F. Li, I. Abeysekera, and S. Ma, *Earnings management and the effect of earnings quality in relation to stress level and bankruptcy level of Chinese listed firms*, Corp. Ownership Control **9** (2011), no. 1, 366–391.

[36] F. Mohammad Rezaei, Z. Dayanti Deilmi, and R. Darvand, *The type of audit report, the number and type of condition clauses of the conditional audit report: the role of the economic crisis*, Knowledge Account. Manage. Audit, **9** (2019), no. 33, 25–39.

[37] M. Namazi and S. Ebrahimi, *Investigating the relationship between product market's competitive structure and stock return of the listed companies on the Tehran Stock Exchange*, Financ. Account. Knowledge **2** (2012), no. 3, 9–27.

[38] D.P. Newman, E.R. Patterson, and J.R. Smith, *The role of auditing in investor protection*, Account. Rev. **80** (2005), no. 1, 289–313.

[39] A. Nikkhah Azad, *Statement of the basic concepts of auditing*, Issued by the committee responsible for developing statements of fundamental auditing concepts of the American Accounting Association (AAA), Tehran: Audit Organization, Audit Guidelines Development Committee Publication, 2000.

[40] S.H. Penman and X.-J. Zhang, *Accounting conservatism, the quality of earnings, and stock returns*, Account. Rev. **77** (2002), no. 2, 237–264.

[41] O. Pourheidari and Z. Azami, *Identification of auditors' comments using neural networks*, J. Account. Knowledge **3** (2009), 77–97.

[42] M. Qian, P.-W. Sun, and B. Yu, *High turnover with high price delay? Dissecting the puzzling phenomenon for China's A-shares*, Finance Res. Lett. **22** (2017), 105–113.

[43] S. Richardson, *Earnings quality and short sellers*, Account. Horizons **17** (2003), 49–61.

[44] M. Scott, *Financial Accounting Theory*, (3 rd ed.), New Jersey: Prentice Hall, 2000.

[45] M.H. Setayesh, F. Ebrahimi, S.M. Saif, and M. Sarikhani, *Predicting the type of auditors' comments with an approach based on data mining methods*, Manag. Account. Res. J. (2011), no. 15, 69–82.

[46] M.H. Setayesh and M. Jamalianpour, *Investigating the equation between financial ratios and non-financial variables with the auditor's opinion*, Account. Audit. Res. **1** (2009), no. 2, 130–157.

[47] N. Stanisic, T. Radojevic and N. Stanic, *Predicting the type of auditor opinion: Statistics, machine learning, or a combination of the two?*, Eur. J. Appl. Econ. **16** (2019), no. 2, 1–58.

[48] B. Van Tendeloo and A. Vanstraelen, *Earnings management and audit quality in Europe: Evidence from the private client segment market*, Eur. Account. Rev. **17** (2008), no. 3, 447–469.

[49] R.L. Watts and J.L. Zimmerman, *Agency problems, auditing, and the theory of the firm: Some evidence*, J. Law Econ. **26** (1983), no. 3, 613–633.

[50] D.G. Whiting, J.V. Hansen, J.B. McDonald, C. Albrecht, and W.S. Albrecht, *Machine learning methods for detecting patterns of management fraud*, Comput. Intell. **28** (2012), no. 4, 505–527.

[51] D. Zdolšek and T. Jagrič, *Audit opinion identification using accounting ratios: experience of United Kingdom and Ireland*, Aktual. Prob. Ekon. **115** (2011), no. 1, 285–310.

[52] C.A. Zhang, *Predict audit quality using machine learning algorithms*, Proc. Seventeenth Int. Florida Artific. Intelli. Res. Soc. Conf. FLAIRS, 2004, no. 2, pp. 1–6.