Int. J. Nonlinear Anal. Appl. 13 (2022) 2, 1937–1951
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.6290

Application of data mining in assessing the level of corporate social responsibility disclosure compliant with financial performance and accounting criteria

Amin Alivandi Darani[a], Mehdi Arab Salehi[b,*], Hadi Amiri[c], Farsad Zamani Boroujeni[d]

[a] *Department of Accounting, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran*

[b] *Department of Accounting, University of Isfahan, Isfahan, Iran*

[c] *Department of Economics, University of Isfahan, Isfahan, Iran*

[d] *Department of Computer, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran*

(Communicated by Javad Vahidi)

Abstract

Utilizing new models instead of traditional statistical models can be quite useful in this field. Examples include data mining, which possesses both high speed and accuracy as well as nonlinear and non-parametric properties Therefore, in the present study, the effects of performance criteria on social responsibility disclosure level was investigated and analyzed via utilization of data mining, Henceforth, a model will be presented aiming to estimate/project the optimal level of social responsibility disclosure grounded on performance criteria. The present study is applied in terms of its nature and purpose/objective as well as based on the field research method of data collection. The statistical population are companies listed on the Tehran Stock Exchange and the required data were collected and analyzed using six data mining methods during the 2013-2019 period. The findings reveal it is possible to classify and predict the optimal level of corporate social responsibility disclosure within the Iranian economic circumstances/environment. Moreover, the utilization of the classification algorithm in the vicinity of the nearest neighbor can accurately predict the optimal level of corporate social responsibility disclosure based on performance criteria. The findings of this study indicate that performance metrics can be a positive predictor of the optimal level of corporate social responsibility disclosure. In addition, due to the potent strength of the proposed model, the used model in this research can be utilized to rank/rate the level of corporate social responsibility disclosure.

Keywords: social responsibility, nearest neighbourhood classification algorithm, supervised learning, data mining
2020 MSC: 26E60, 62H11, 91G15

1 Introduction

With the advent of the digital information age, the topic of data explosion, information and the need to constantly maintaining it is palpable more than ever before. Data analysis can provide additional knowledge and insight regarding

*Corresponding author
Email addresses: amin.48361@gmail.com (Amin Alivandi Darani), mehdi_arabsalehi@ase.ui.ac.ir (Mehdi Arab Salehi), h.amiri@ase.ui.ac.ir (Hadi Amiri), f.zamani@khuisf.ir (Farsad Zamani Boroujeni)

Received: November 2021 *Accepted:* February 2022

specific fields. This can be accomplished with increasing data transparency in order to extract knowledge. In fact, data in raw/unrefined form has little value. What is valuable is the knowledge and cognition that can be obtained and utilized from data. Larger data sets can induce stronger results. The business community is theoretically well aware of the overflow and the superfluous amount of information available today [12].

Among the most significant concerns of society today is obtaining credible information about the social performance of companies and their impact on the environment and society. The timeliness and reliability of the information about these companies is also important [38]. Reliable estimation and prediction of the social responsibility status of corporations provides an opportunity for society and investors to assess the extent of corporate social and environmental responsibilities in their decision making process. Furthermore, analyzing and evaluating data will enhance their information, knowledge and awareness in order to project and predict what could happen in the future, hence helping decision makers to make accurate/correct decisions [12]; Thus, the use of machine learning and data mining techniques can provide useful information to decision makers (investors, executives and other stakeholders) concerning the future of corporate social performance.

In previous studies [9], the effect of performance criteria on the level of social responsibility disclosure via using statistical analysis methods was the only topic assessed/investigated, however, no study was preformed on the possibility/potential of predicting and estimating the future status of corporate social responsibility disclosure; Hence, in light of the impact of various performance criteria on the level of corporate social responsibility disclosure in previous studies, the question arises as to which corporate performance criteria can affect the level of their social responsibility disclosure and whether the level of social responsibility disclosure can be determined/predicted by considering such criteria. Therefore, the objective/purpose of this study is to determine the level of corporate disclosure based on performance criteria and to assess whether the level of social responsibility disclosure can be predicted utilizing smart/intelligent algorithms. Accordingly, hereinafter in this study, pursuant to delineating the theoretical and experimental background, the research methodology shall b presented. Thereafter, the experimental findings of the research shall be described, and in the final part, the conclusions and proposals will be offered.

2 Theoretical foundations

The concept of social responsibility deals with the relationship between a corporation and society. On the one hand, it examines the impact of company activities on all members of society, and on the other hand, evaluates its long-term survival in a competitive world as a consequence the of its social responsibility [43].

In various studies, different dimensions and indexes have been identified as social indicators. For example, Qiu & Associates [43], analyzed in excess of 160 social responsibility indicators in two general environmental and social categories (based on GRI guidelines) to determine the relationship between disclosure and performance indicators, the contract or Everaert & Associates [16], identified principles and management values/vision as performance indicators of disclosure and described social dimension components as human rights, work practices, end-product and society.

In Iran, since there is not a comprehensive standard and criteria of social responsibility indicators, in studies conducted consistent with international institution guidelines, multiple dimensions of corporate social responsibility have been studied. For instance, Hassas, Yeganeh & Barzegar [43], examined 7 social dimension components including: 1-Work & employees, 2-Human rights, 3- Supply chain, customers & consumers of products & services, 4-Participation & societal development, 5-Business ethics, 6- Corruption, bribery & money laundering, and 7-Adherence of laws & regulations. They moreover analyzed 43 related indicators. Elsewhere, Alivandi & Associates [4] categorized 45 indicators into 5 social responsibility dimensions by validating the dimensions and indicators of social responsibility disclosure utilizing the Delphi method as well as taking into account the views of experts. In addition, Mahdavi & Associates [25] used 37 indicators in the form of 6 dimensions of environment, products & services, human resources, customers, society and energy to measure the level of social responsibility disclosure of companies listed in the Tehran Stock Exchange.

Researchers have expressed various theories about the effect of corporate performance on the level of social responsibility disclosure. Some theorists argue that the legitimacy derived from disclosing social responsibility and the public pressure to enhance social legitimacy have positive environmental and social effects on a company's operations and lead to further disclosures [28]. Contrary to the legitimacy viewpoint, other scholars, either implicitly or explicitly, based on resource-based theory ([17], [34]) and the voluntary disclosure economic theory [42], argue that superior performance or superior economic resources are likely to induce and bring about higher and better quality disclosures, including social responsibility disclosures [10].

Some other researchers argue that corporate performance can affect the timeliness and quality of financial reporting as well as disclosure of corporate information; Therefore, companies with success (good news) engage in financial reporting in a more timely manner and with better quality than companies with unsuccessful operations (bad news). Proponents of this position believe that performance measures the efficiency and productivity of companies and a company's performance has an effect on the company's fortunes in the stock market. It also displays the company's management skills. For example, a company with good news (optimal performance) experiences an increase in stock market value plus management credibility. The opposite is true for companies with bad news (negative performance) [27].

What's more, [37] stipulate that profitability and performance are a measure of proper management. Conversely, when profitability (performance) is low, management may disclose less information in order to conceal the causes of losses (or reduce profits/revenues); Hence, it is expected from the management of a profitable company (with good performance) to report comprehensive details and quality information in order to describe its ability to maximize shareholder dividends as well as strengthen its position and earnings.

Another reason for performance's impact on the level of social responsibility disclosure is founded on the theory of messaging. According to this theory, executives are willing to report good news to the capital market in order to avoid undervaluation of their shares. Therefore, they report detailed and quality information in order to maintain their favorable positions and contracts. Inchausti (1997) believes that more profitable companies (with superior performance), report more and higher quality information. He argues that, consistent with agency/representation theory, for-profit corporate executives use disclosure of information for personal gain. Other reasons for the performance's effect on the level of social responsibility disclosures can be the political processes theory, which states that more profitable companies are interested in more quality information disclosure to justify the level of profit/earnings [28].

Data mining is deploying special algorithms to extract patterns and hidden knowledge in large chunks of data [19]. Data mining has three main objectives: To describe, predict, and provide a solution (version). Describing the data is achieved by focusing on finding interpretable patterns, prediction is accomplished via utilizing variables in the database to estimate the future values of other variables, with the objective being to solve the issue by identifying the best possible solution to the problem by examining all possibilities [15]. Contingent on the objectives, data mining tasks are varied. If the objective is to categorize, the data mining focuses on predefined qualitative characteristics; if the objective is predictive, then it is to find a numerical value in the future; And if the goal is detection, data mining focuses on finding data that deviates significantly from the normal state. Moreover, with the aim of optimization, data mining focuses on finding the best solution consistent with the available resources [19].

There is a wide range of data mining techniques such as artificial neural networks, decision-making tree, support vector devices, regression, classification algorithm in the nearest neighbor, etc. Each of these data mining techniques serves a specific objective, issue, and business requirement [5]. The decision-making tree is a retrospective structure for expressing an intermittent classification process described by a set of attributes and assigns a circumstance to a discrete set of categories. Each leaf of the decision-making tree represents a class/category. In the tree classifier, there is flexibility to select various subsets of traits in different internal groups of the tree, in such a manner wherein the selected subset optimally distinguishes between the categories of this group [19].

The classification algorithm in the immediate vicinity is an algorithm wherein test data are classified according to training patterns. The data that placed next to each other is called the neighbor, and each new data introduced to the algorithm, its distance from the other data is calculated and placed in the category that is closest to each other [21].

Linear separator analysis is closely related to variance and regression analysis, endeavoring to express an independent variable as a linear combination of other properties. This independent variable in linear separator analysis is in the form of a class label and additionally attempts to model the differences between various classes of data [31].

An artificial neural network is an information processing concept inspired by the biological nervous system and processes data like the brain. The key element of this concept is the new structure of the information processing system. This system is composed of a large number of extraordinarily interconnected processing elements called neurons working together to solve a problem [19].

A support vector machine is among the supervised learning methods utilized for linear classification. Linear classification methods endeavor to separate data by constructing a super surface (which is a linear equation). The backup vector machine classification method (among the linear classification methods), finds the best super surface that separates with maximum distance data related to two classes [19].

Among the most widely used data modeling methods, which also has a very simple mathematical basis, is linear regression. Once we can distinguish a linear relationship between two variables, we can deploy this type of regression to predict the values of these variables based on the value of the other variable. Linear relationship means observing when with the increase of one variable, the other variable increases (decreases) and with the decrease of the second variable, the variable decreases (increases), and this increase or decrease has a direct relationship (simple coefficient) with the value of the first variable, called an independent variable [19].

The utilization of data mining technology and artificial intelligence in accounting is an inevitable and irreversible trend that brings about extraordinary changes and developments as well as a new era. Undoubtedly, this intelligent accounting is a financial development that is definitely in line with the future trends of the sector. Examining and assessing data mining applications in past accounting studies reveals that according to the objectives of previous conducted researches, most of the previous studies were retrospective and much less about prediction and future estimations of values utilizing the findings of past and present studies, demonstrating a clear gap between the two empirical applied research categories (retrospective & predictive), revealing suitable opportunities to utilize and benefit from data mining in this category of research [5].

3 Research background

Up to this point, no domestic or international research has been conducted on modeling and predicting the optimal level of social responsibility disclosure based on performance criteria. Also, research preformed on the application of artificial intelligence and data mining in accounting has been very limited (please see hereinafter).

[40] utilized the decision-making tree technique to predict financial failure. Using the properties of financial ratios and the entropy-based discretization method, they designed a data mining model to predict corporate financial distress. By testing 35 financial ratios in 135 companies, they proved/verified the feasibility and validity of the data mining method to explore and predict the financial problems of the sampled companies. [41] assessed the prediction of profit/earning management via neural network and decision-making tree. The primary goal of their research was to analyze the utilization of neural networks to predict up or down profit/revenue management. They used stock exchange data and factors related to earnings management in previous Taiwanese research and after the validation stage, the research findings demonstrated the validity of the forecast of upward profit/earnings of 10.21%.

In their study, [22] used data mining methods to estimate financial failure predictions for a sample of companies listed on the China Stock Exchange. In their study, data collected from companies listed on the Shanghai Stock Exchange and the Shenzhen Stock Exchange were analyzed utilizing data mining algorithms such as decision-making tree, support vector device, nearest neighbor and logistic regression. Their findings indicated that by using data mining algorithms, it is possible to predict the business failure of Chinese companies. [45] used models based on meta data analysis versus prediction models based on predetermined models, which examined the performance of different models regarding the performance of different models vis a vis forecasting the financial distress of companies using data mining techniques and their research results showed that there is no significant difference between data mining models for classifying and forecasting financial distress.

However, a combination of knowledge and methods based on genetic algorithms can perform better in predicting corporate financial distress. They also discovered that the combination of predetermined and predictive models significantly improved the ability to predict the financial predominant of companies. [44] also utilized machine learning-based models to determine the optimal level of investment. using data mining algorithms, they compared the Shanghai Stock Exchange index with the Nasdaq index and found that the genetic algorithm performed better to determine the optimal level of investment than the decision-making tree and the Bayesian network methods. In their research, [6] assessed the impact of organizational systems, big data and data analysis on management accounting and provided a framework for analyzing management accounting of data based on balanced evaluation card theory and business intelligence.

[23] examined the application of artificial intelligence in accounting for Chinese companies and universities, and the findings revealed their increasing use (artificial intelligence) to effectively solve problems and help their future development.

[8] predicted earnings management via utilizing the decision-making tree. To verify the accuracy of their forecast, institutional shareholder ownership percentage, debt ratio, company size, income tax, sales variability, profit/earnings variability, operating cash, profit/earnings quality ratio, total asset turnover, sales returns, return on investment and returns equity were assessed as independent variables and optional accruals was considered as a dependent variable. Their research findings revealed that the decision-making tree's highest accuracy for profit/earnings management forecast is 74.7%.

[25] provided a model for detecting fraud by auditors utilizing a feed-in artificial neural network with an error replication algorithm. The statistical population of the study consisted of supervisors, senior supervisors and managers/executives

of auditing firms (all members of the Iranian Society of Certified Public Accountants). Using the MATLAB software, the analysis of the collected questionnaires from the above-mentioned statistical population regarding fraudulent and non-fraudulent companies revealed that the artificial neural network pattern (designed with 9 hidden layers) is able to identify fraudulent and non-fraudulent companies with an 86.9% accuracy rate.

[3] proposed a model utilizing a decision-making tree and an algorithm with a tree structure to predict the current and future returns of companies. Accessing the information of 317 companies listed on the Tehran Stock Exchange (from 2002-2013), they used four decision-making tree algorithms to elucidate the current return and predict future returns. Their findings indicated that the power of the models to explicate current returns is greater than predicting future returns, but since in both instances the models are not statistically reliable, the hypothesis of an explanatory relationship between projected financial ratios and changes in current and future stock returns was not confirmed.

[39] used data mining algorithms to predict the systematic return and risk of stocks in companies listed on the Tehran Stock Exchange. In their research, deploying four linear separator analysis algorithms, nonlinear separator analysis algorithm, nearest neighbor K algorithm and decision-making tree, and with the assistance of 16 independent variables, the systematic return of stock yield and risk was projected. Their findings suggest that the utilization of selected independent variables (instead of all independent variables) enhances the ability of algorithms to predict systematic returns and risk.

[35] evaluated the accuracy of earnings management forecasting utilizing neural networks and decision-making tree and then compared them with linear models. Toward this goal, they used nine variables affecting earnings management as independent variables and optional accruals as a dependent variable. In their study, four sectors (agriculture, pharmaceutical, textile and petroleum byproducts) and 63 companies were dissected. Their findings showed that the neural network and decision-making tree method is more accurate in predicting profit/earnings management and with a lower error level than linear methods.

4 Research methodology

4.1 Data mining

For a given set $\mathcal{A}$ of m points in R^n represented by the matrix $A \in R^{m\times n}$ and a number k of desired clusters, we formulate the clustering problem as follows. Find cluster centers $C_\ell, \ell = 1, \cdots, k$ in R^n such that the sum of the minima over $\ell \in \{1, \cdots, k\}$ of the l-norm distance between each point $A_i, i = 1, \cdots, m$, and the cluster centers $C_\ell, \ell = 1, \cdots, k$, is minimized. More specifically we need to solve the following mathematical program:

$$\begin{array}{cc} \underset{C,D}{minimize} & \sum_{i=1}^{m} \min_{\ell=1,\cdots,k}\{e^T D_{i\ell}\} \\ subject\ to & -D_{i\ell} \le A_i^T - C_\ell \le D_{i\ell}, i = 1, \cdots, m, \ell = 1, \cdots, k \end{array} \tag{4.1}$$

Here $D_{i\ell} \in R^n$, is a dum my variable that bounds the components of the difference $A_i^T - C_\ell$ between point A_i^T and center C_ℓ and e is an $n \times 1$ vector of ones in R^n. Hence $s^T D_{i\ell}$ bounds the 1-norm distance between A_i and C_ℓ. We note that just as in the case of ro bust regression [?], [? , pp 82-87], the use of the 1-norm here to measure the error criterionleads to insensitivity to outliers such as those resulting from distributions with pronounced tails. We also note that since the objective function of (4.1) is the minimum of k linear (and hence concave) functions, it is a piecewise-linear concave function [? , Corollary 4.1.14]. This is not the case for the 2-norm or p-norm, $p \neq 1$. Although (4.1) is NP-hard, it can be reformulated as the following bilinear program which can be solved effectively by using a k-Median Algorithm that consists of solving a succession of simple linear programs in closed from. We state the bilinear programming formulation and k-Median Algorithm for solving the clustering problem.

Proposition 4.1 (Clustering as a Bilinear Program). The clustering problem (4.1) is equicalent to the following bilinear program:

$$\begin{array}{cc} \underset{C_\ell\in R^n, D_{i\ell}\in R^n, T_{i\ell}\in R^n}{minimize} & \sum_{i=1}^{m}\sum_{\ell=1}^{k} e^T D_{i\ell}T_{i\ell} \\ subjectto & -D_{i\ell} \le A_i^T - C_\ell \le D_{i\ell}, i = 1, \cdots, m, \ell = 1, \cdots, k \\ & \sum_{\ell=1}^{k} T_{i\ell} = 1 \quad T_{i\ell} \ge 0, i = 1, \cdots, m, \ell = 1, \cdots, k \end{array} \tag{4.2}$$

This esscntially obvious resull [? , Proposition 2.2] can be seen from the fact that, for a fixed i, setting all the components of $T_{i\ell}, \ell = 1, \cdots, k$ equal tozero except one corres ponding to a smallest $e^T D_{i\ell}$, with respect to ℓ, equal to 1, leads to the objective function of (4.1) from of (4.2). Note that the constaints of (4.1) are uncoupled in the variables (C, D) and the variable T. Hence the Uncoupled Bilinear Program Algorithm UBPA [? , Algorithm 2.1] is applicable. Simply stated, this algorithm algorithm alternates between solving alinear program in the variable T and alinear program in the variables (C, D).The algorithm terminates in afinite number of iterations at a stationarv point satisfying the minimum principle necessary optimality condition for problem (4.2) [**?** , Theorem 2.1].

We note however, because of the simple structure the bilinear program (4.2), the two linear programs can be solved explicitly in closed from. This leads to the following algorithmic implementation.

Algorithm$5 \cdot 1$ **k-Median Algorithm** Given the cluster centers $C_1^j, \cdots C_k^j$ at iteration compute $C_1^{j+1}, \cdots C_k^{j+1}$ by the following two steps:

(a) **Cluster Assignment**: For each $A_i^T, i = 1, \cdots, m$ determine $\ell(i)$ such that $C_{\ell(i)}^j$ is closest to A_i^T in the norm.

(b) **Cluster Center Update**: For $\ell = 1, \cdots, k$ choose C_ℓ^{j+1} as a median of all A_i^T assigned to C_ℓ^j.
Stop when $C_\ell^{j+1} = C_\ell^j$. Assign each point to a cluster whose center is the 1-norm to the point.

Although the k-Median Algorithm is similar to the k-Mean Algorithm wherein the 2-norm distance is used [**?**]-[**?**], it differs from it computationally, and theoretically. In fact, the underlying problem (4.2) of the k-Median Algorithm is a concave minimization on a polyhedral set while the corresponding problem for a two-or p-norm, $p \neq 1$, is :

$$\begin{array}{ll} \underset{C,D}{minimize} & \sum_{i=1}^m \min_{i=1,\cdots,k} \|D_{i\ell}\|_p \\ subject\ to & -D_{i\ell} \leq A_i^T - C_\ell \leq D_{i\ell}, i = 1, \cdots, m, \ell = 1, \cdots, k. \end{array} \tag{4.3}$$

This is not a concave minimazion on a ployhedral set, because the minimum of a set of convex functions is not in general concave. We also note that thek-Mean Algorithm finds a stationary point not of problem (4.3) with $p = 2$, but of the same problem except that $\|D_{i\ell}\|_2$ is replaced by $\|D_{i\ell}\|_2^2$ andthus favoring outliers. Without this squared distance term, thesubproblem of the k-Mean Alggorithm becomes the considerably harder Weber problem [**?**]-[**?**] which locates a center in R^n closest in sum of Euclidean disstances (not their squares!) to a finite set of given points. The Weber problem has not closed from solution. However, using the mean as a cluster center of points assigned to the cluster, as done in the k-Mean Algorithm, minimizes the sum of the squares of the distances from the cluster center to the points.
Because there is no guaranteed way to ensure global optimality of the solution obtained by either the k-Mean Algorithms, different stating points can be used to initiate the algorithm. Random stating cluster centers or some other heuristic can be used such as placing k initial centers along the coordinate axes at densets, second densest ... k densest intervals on the axes. The latter heuristic was used in our computational results.

To test the effectiveness of the k-Median Algorithm it was used as a KDD tool [**?**] to mine the Wisconsin Prognostic Breast Cancer Database (WPBC) is order to discover medical knowledge. For such medical databases, extracting well-separated survival curves provides an essential prognostic tool. Survival curves [**?**]-[**?**] give expected percent of surviving patients as a function of time. The k-Median Algorithm was applied to WPBC to ext ract such curves. Survival curves were constructed for 194 patients using two clinically available features for each patient: tumor size and number of cancerous lymph nodes excised. Using $k = 3$, the k-Median Algorithm separated the points into 3 clusters. Thesurvival curve for each cluster is depicted in Figure 2(a). The key observation to make here is that curves are well separated, and hence the clusters can be used as prognostic indicators to assign a survival curve to a patient depending on the cluster into which the patient falls. By contrast, the k-Mean Algorithm obtained poorly separated survival curves as shown in Figure 2(b), and hence are not useful for prognosis.

a nother comperison of the k-Median and k-Mean Akgorithms was perpormed on databases with known classes. Correctness was measured by the ratio of the sum of the number of majority points in each cluster to the total number of points m in the data set. Table 1 shows results averaged over ten random starts for four data bases from the lrvine repository of databases[**?**]. We note that for two of the databeses the k-Median gave better correctness than the k-Mean and for the other two the k-Mean was better.

4.2 Methodology

The research methodology used in this research is correlational and predictive and its analysis is grounded on exploratory methods. Moreover, this is an applied type of study. To collect data related to regression review and data mining methods, the annual report of boards of directors to general assemblies plus the financial statements of companies listed on the Tehran Stock Exchange were used (2013-2019). Furthermore, in the statistical population of companies listed on the Tehran Stock Exchange, a sample of 134 companies was selected based on the non-probabilistic (targeted) sampling method. Considering that the objective of this study is to provide a model for predicting the level of disclosure of social responsibility based on performance criteria, the following three steps were undertaken to accomplish the aims of this research:

Step 1: By reading the reports and financial statements of boards of directors to general assemblies, utilizing the checklist prepared from the indicators of disclosure of social responsibility of Alivandi & Associates (2020), the level of disclosure of corporate social responsibility was determined. The dimensions and indicators of social responsibility disclosure used in this research are presented in Table 1.

In the second stage, based on the statistical sample, in order to determine which of the performance criteria is able to affect the level of social responsibility disclosure and can be included in the predictive model, regression and least squares method was utilized. The effect of each performance criterion on the level of social responsibility disclosure was assessed in separate models. Those criteria whose effect on the level of social responsibility disclosure was not proven were excluded from the model of optimal prediction of the social responsibility disclosure level. Toward analyzing the effect of performance criteria on corporate social responsibility disclosure level [consistent with the research by Qiu & Associates], the regression model with the combined data method was utilized, as delineated in Equation (4.4).

$$DisclpsureScore_{it} = \beta_0+\beta_1 PerformanceCriteria_{it}+\beta_2 PerformanceCriteria_{it-1}+\beta_3 Age_{it}+\beta_4 Size_{it}+\beta_5 Leverage_{it}+\varepsilon_{it} \tag{4.4}$$

In the third and final stage, using six data mining algorithms, the optimal level of social responsibility disclosure was predicted and the accuracy of each evaluation method was compared. RapidMiner software was deployed to classify and predict the optimal level of social responsibility disclosure based on performance criteria. The work process includes the following steps:

1- Categorizing the social disclosure level and labeling data as low (0-3), medium (3-5) and high (5 & above).
2- Data normalization.
3- Determining the percentage of training and test data for statistical models utilized in this study that require training (75% of the data as training data & 25% as test data).
4- Application of data mining methods including classification algorithm in the vicinity of the nearest neighbor, decision-making tree, linear separation analysis, linear regression, support vector machine and neural network.
5- Ultimately, compare the value predicted by the model with the actual value and determine the measurement models' accuracy (Figure 1).

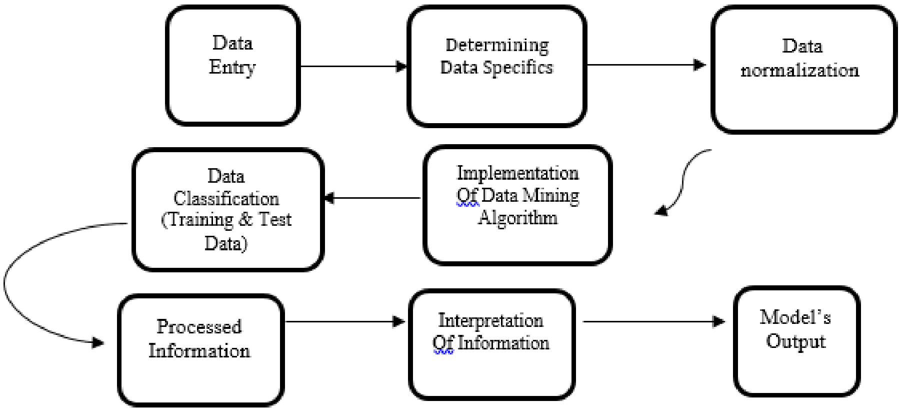
Data Entry
Determining Data Specifics
Data normalization
Implementation Of Data Mining Algorithm
Data Classification (Training & Test Data)
Processed Information
Interpretation Of Information
Model's Output

Figure 1: Implementation model of rapid miner software for predicting the optimal level of social responsibility disclosure compliant with performance criteria

In order to evaluate the validity of the algorithms in the classification, the accuracy index is used [obtained from the following equation [19]:

$$\text{Algorithm Prediction Accuracy} = \frac{\text{Number of companies that are properly categorized}}{\text{Total number of companies}} \tag{4.5}$$

Social responsibility disclosure level is the dependent variable of this research. In this study, the scoring procedure for measuring corporate social responsibility is consistent with the [1]. If an item is disclosed according to the corporate social responsibility checklist, a score of 1 is given (0 if not disclosed). Hence, the number of items disclosed is compared to the whole corporate social responsibility checklist in line with the data contained in the annual reports of boards

Table 1: Social responsibility disclosure dimensions & indicators

Social Responsibility Dimensions	Disclosure Indicators
Surrounding/Workplace & Environment	Amount Of Recycled Raw Materials Utilized In Company's Ongoing/ Current Activities
Surrounding/Workplace & Environment	Saving Electricity In Company's Ongoing/Current Activities (MWh)
Surrounding/Workplace & Environment	Fuel Used For Company's Ongoing/Current Activities (Liters)
Surrounding/Workplace & Environment	Fuel Saving In Company's Ongoing/Current Activities (Liters)
Surrounding/Workplace & Environment	Solar energy Utilized In Company's Ongoing/Current Activities
Surrounding/Workplace & Environment	Utilization Of Other Renewable Energy In Company's Ongoing/Current Activities (Such As Wind, Geothermal, Waves, Etc.)
Surrounding/Workplace & Environment	Energy Efficiency Policy Utilized In The Company
Surrounding/Workplace & Environment	Water Consumption In Company's Ongoing/Current Activities (Cubic Meters)
Surrounding/Workplace & Environment	Use of Recycled Water In Company's Ongoing/Current Activities (Cubic Meters)
Surrounding/Workplace & Environment	Significant/Meaningful Impact Of Activities, Products & Services On Environmentally Protected & High-Biodiversity Valuable Areas
Surrounding/Workplace & Environment	Current Strategies, Actions & Plans/Programs For Managing Environmental Impacts & Biodiversity
Surrounding/Workplace & Environment	Emission Of Greenhouse & Hazardous Gases, Particulate Matter, Etc. By The Company
Surrounding/Workplace & Environment	Greenhouse & Hazardous Gases, Particulate Matter, Etc. Reduction Plans/Programs Implemented By The Company
Surrounding/Workplace & Environment	Hazardous Waste Generated By The Company
Surrounding/Workplace & Environment	Recycling Of Hazardous Waste Produced By The Company
Surrounding/Workplace & Environment	Company's Environmental Waste Disposal Policy
Surrounding/Workplace & Environment	Company's Waste Reduction Policy
Surrounding/Workplace & Environment	Initiatives On Reducing Detrimental Effects Of Products & Services On The Environment
Surrounding/Workplace & Environment	Total Investment & Costs Incurred For Environmental Protection
Surrounding/Workplace & Environment	Mechanism For Responding To Surrounding/Workplace & Environmental Complaints
Surrounding/Workplace & Environment	Approved Instances Of Compliance With Environmental Laws & Regulations
Human Resources	Occupational Health & Safety Policy Governing Company's Workplace Environment
Human Resources	Occupational Accidents Occurring To Company's Workforce/Employees
Human Resources	Death Of Personnel/Employees While Performing Duties/Tasks In The Workplace
Human Resources	Employee Training Expenditures Toward Increasing & Improving Their Performance
Human Resources	Fair Remuneration Policy Toward Increasing Personnel/Employee Efficiency
Human Resources	Certificates & Commendations For Compliance With Labor Laws & Regulations
Human Resources	Number Of Personnel/Employee Complaints About Workplace/Working Conditions
Human Resources	Formal Offenses/Crimes & Convictions Related To Company's Manpower/Human Resources Complaints
Human Rights	Measures Taken To Effectively Help Eliminate Child Labor
Society	Participation (Charitable Activities) In The Local Community & Assessment Of Local Impact (At The Geographical Level)
Society	Public Participation (Public Benefit Activities) & Public Impact Assessment (Domestically/Nationally & Internationally)
Society	Company's Anti-Corruption Policies & Procedures
Society	Actions Taken Against Uncovered/Exposed Corruption
Society	Crimes Related To Social Activities Rules & Regulations Non-Compliance
Product & Services	Measures For Improving Health & Safety Of Products & Services
Product & Services	Number Of Accidents Related Products & Services Health & Safety Regulations Non-Compliance
Product & Services	Supply Chain Of Company's Product(s)
Product & Services	After-Sales Service Provided By Company
Product & Services	Activities Undertaken To Increase Customer Satisfaction (Respect, Response, Etc.)
Product & Services	Procedures Pertinent To Measuring Customer Satisfaction, Including Surveys
Product & Services	System For Responding To Customer Complaints & Dissatisfaction
Product & Services	Protecting Privacy Of Consumers (Their Data/Info)
Product & Services	Certificates & Confirmations Of Compliance With Goods/Services & Rights Of Consumers Laws & Regulations
Product & Services	Crimes/Transgressions Pertinent To Goods/Services & Rights Of Consumers

Source: Social Responsibility Disclosure Dimensions & Indicators, [4].

of directors to general assemblies. The outcome indicates the level of corporate social responsibility disclosure (or corporate social responsibility rating).

Performance is the independent variable in this research. Toward measuring it, various indicators were utilized (as follows):

QTUBIN: The company's total market value (the number of shares in the market & its value at the end of the period & the "book value" of debts) divided by the "book value" of the company's assets

MTBR: The natural logarithm of the equity's market value ratio to the "book value" of equity

Return On Assets (ROA): Net profit divided by the company's total assets

Return On Equity (ROE): Net dividend divided by the sum of the company's equity

Earnings Per Share (EPS): Net income divided by the total issued common stock

P/E ratio : The ratio between the price and income/revenue of company's each share

P/B Ratio: Divide the market value of a stock by its "book value"

Current Ratio: Current assets divided by current liabilities

Instantaneous Ratio : The difference between current assets & inventories divided by current liabilities

Table 2: Performance measurement indicators

Evaluation Criteria Based On Market Value	QTUBIN Indicator	[20] & [33]
Evaluation Criteria Based On Market Value	MTBR	[14]
Return Ratios	ROA	[30], [7], [32], [33]
Return Ratios	ROE	[30], [33]
Return Ratios	EPS	[11], [32]
Criteria For Valuation Of Stock Value	P/B Ratio	[30], [13]
Criteria For Valuation Of Stock Value	P/E Ratio	[32]
Liquidity Ratios	Current/Present Ratio	[29], [30]
Liquidity Ratios	Future Ratio	[29], [30]

Source: Researcher's Findings

5 Control variables

Size : The natural logarithm of the "book value" of assets at the end of the financial period

Leverage : The ratio of the total "book value" of liabilities to the total "book value" of assets

Age : How many years since the company was listed on the Tehran Stock Exchange.

6 Research hypotheses & questions

Whenever a researcher seeks to solve a problem of a relationship between two or more variables, it is possible to formulate a hypothesis and predict/project the relationship between them; However, in cases where the goal is not the relationship between variables, but only to find out the status of a variable, the hypothesis is not necessary [26]. Since the purpose of this study is to compare the power of six data mining techniques in classifying and predicting/estimating social responsibility disclosure levels, this study's questions are presented as follows:

Question 6.1. Considering the economic environment/circumstances of Iran, is it possible to predict the optimal level of corporate social responsibility disclosure based on performance criteria utilizing data mining techniques?

Question 6.2. Which of the six algorithms of linear separation analysis, nearest neighbor classification algorithm, decision-making tree algorithm, neural network, linear regression and support vector machine is able to predict more accurately the social responsibility disclosure level of companies listed on the Tehran Stock Exchange?

7 Research findings

Table 3 displays the descriptive statistics of the model variables, which include data about mean, standard deviation, maximum & minimum.

Table 3: Descriptive statistics of model's variables

Name	Mean	Standard Deviation	Minimum	Maximum
Disclosure Level	0.220	0.108	0.022	0.555
QTUBIN	1.591	0.701	1.007	5.161
MTBR	2.708	1.673	1.011	9.846
ROA	0.095	0.133	-0.597	0.628
ROE	0.248	0.539	-2.947	4.761
EPS	775	1390	-2049	8449
P/B	6.588	8.079	0.41	64.516
P/E	12.674	120.148	-1425.4	595
Current/Present Ratio	1.372	0.699	0.131	5.210
Future Ratio	0.850	0.530	0.015	4.997
Financial Leverage	0.638	0.185	0.068	0.896
Company Size	14.634	2.881	7.352	26.960
Firm's Longevity	3.070	0.385	2.303	4.127

Source: Researcher's Findings

The primary central indicator is the average, signaling the equilibrium point and center of gravity of the distribution, and is a good indicator of showing the centrality of the data. For example, the average value for the return on assets (ROA) variable is 0.095, indicating that most of the data is centered around this point. One of the most important scattering parameters is the standard deviation. Among the variables, the level of disclosure is the lowest, with Earnings Per Share (EPS) having the highest dispersion (demonstrating that these two variables have the least and the most changes, respectively).

Regression analysis is founded on several fundamental and uncomplicated assumptions. If one or more of these assumptions are not met, the interpretation of regression analysis will be incorrect/deficient and the predictions made based on it will be weak/off the mark. These assumptions include: normality of errors, lack of alignment, lack of variance heterogeneity and lack of auto-correlation. Because of the fact that in this research, combined data has been utilized, only the hypothesis of non-alignment, non-heterogeneity of variance and lack of auto-correlation is sufficient to validate the least squares regression, as presented below:

Table 4: Residual auto-correlation test (voldridge test)

Model	Independent Variable	Voldridge Stat	Probability	Outcome/Result
1	QTUBIN	10.534	0.001	Homogeneity (Utilization of AR1)
2	MTBR	10.363	0.001	Homogeneity (Utilization of AR1)
3	ROA	10.430	0.001	Homogeneity (Utilization of AR1)
4	ROE	10.573	0.001	Homogeneity (Utilization of AR1)
5	EPS	10.270	0.001	Homogeneity (Utilization of AR1)
6	P/B	10.016	0.001	Homogeneity (Utilization of AR1)
7	P/E	10.231	0.001	Homogeneity (Utilization of AR1)
8	Current/Present Ratio	9.669	0.002	Homogeneity (Utilization of AR1)
9	Future Ratio	10.378	0.001	Homogeneity (Utilization of AR1)

Source: Researcher's Findings

Table 5: Variance heterogeneity (non-homogeneity) wald test

Model	Independent Variable	Wald Stat	Probability	Outcome/Result
1	QTUBIN	7626.79	0.000	Variance Heterogeneity (Utilization of GLS)
2	MTBR	8232.61	0.000	Variance Heterogeneity (Utilization of GLS)
3	ROA	9118.32	0.000	Variance Heterogeneity (Utilization of GLS)
4	ROE	7973.05	0.000	Variance Heterogeneity (Utilization of GLS)
5	EPS	8953.06	0.000	Variance Heterogeneity (Utilization of GLS)
6	P/B	8308.78	0.000	Variance Heterogeneity (Utilization of GLS)
7	P/E	18/8914	0.000	Variance Heterogeneity (Utilization of GLS)
8	Current/Present Ratio	9664.22	0.000	Variance Heterogeneity (Utilization of GLS)
9	Future Ratio	9127.45	0.000	Variance Heterogeneity (Utilization of GLS)

Source: Researcher's Findings

Table 6: Non-alignment test among independent variables

Model	Independent Variable	VIF Quantity	Outcome/Result
1	QTUBIN	1.06	Non-Alignment
2	MTBR	1.01	Non-Alignment
3	ROA	1.28	Non-Alignment
4	ROE	1.01	Non-Alignment
5	EPS	1.00	Non-Alignment
6	P/B	1.15	Non-Alignment
7	P/E	1.00	Non-Alignment
8	Current/Present Ratio	1.02	Non-Alignment
9	Future Ratio	1.00	Non-Alignment

Source: Researcher's Findings

Furthermore, the effect of performance criteria on social responsibility disclosure level findings utilizing generalized least squares regression at the 95% confidence level are summarized in Table 7 below.

Table 7: Summary of study's findings regrading performance criteria's effect on social responsibility disclosure level

Independent Variable	T Stat	Probability	Regression Findings
QTUBIN	1.769	0.044	Effect Confirmed
MTBR	1.521	0.105	Effect Rejected
ROA	2.745	0.006	Effect Confirmed
ROE	−2.134	0.033	Effect Confirmed
EPS	−0.381	0.703	Effect Rejected
P/B	2.270	0.023	Effect Confirmed
P/E	−3.618	0.000	Effect Confirmed
Current/Present Ratio	1.765	0.077	Effect Rejected
Future Ratio	1.330	0.183	Effect Rejected

Source: Researcher's Findings

According to the regression test findings, 5 performance criteria (QTUBIN, ROA, ROE, P/B, P/E) had an effect on social responsibility disclosure level as independent variables along with 3 criteria of company's size, financial leverage and firm's longevity. Control variables were included in data mining algorithms for the classification model and prediction of the optimal level of social responsibility disclosure.

First off, the classification algorithm in the vicinity of the nearest neighbor with the number of different neighbors (K) was investigated. The number of neighbors and distance measurement methods are the most important determining factors within this algorithm. As is evidenced, the number of neighbors is 5, 6, 7, 8 & 9 and the corresponding distance measurement methods are Manhattan, Euclidean respectively [**?**]. Table 8 shows the results of applying the classification model in the vicinity of the nearest neighbor to the training sample. The optimal outcome is obtained when the number of neighbors is 6. The best result is equal to 81.59% and the method of measuring used is the Euclidean distance.

Table 8: Classification method accuracy comparative table in the vicinity of the nearest neighbor (with distance type & number of different neighbors)

Distance Type	K=6	K=7	K=8	K=9
Manhattan Distance	79.92%	81.17%	81.17%	79.50%
Euclidean Distance	81.59%*	80.75%	80.33%	79.92%
Chebyshov Distance	81.59%	80.33%	79.50%	79.08%

Source: Researcher's Findings (* Highest Level Of Accuracy, 95% Confidence Level)

The findings of estimating/projecting social responsibility disclosure level based on performance criteria using six classification algorithms including nearest neighbor (KNN), decision-making tree, linear separation analysis, linear regression, support vector machine algorithm (SVM) and neural network is presented below in Table 9:

Table 9: Accuracy comparison table of supervised learning methods for predicting the social responsibility disclosure level based on performance criteria

Learning Method	KNN (Euclid & K=6)	Decision-Making Tree	Linear Separation Analysis	Linear Regression	SVM	Neural Network
ClassificationAccuracy	81.59%*	79.08%	78.66%	76.92%	80.75%	81.17%

Source: Researcher's Findings (* Highest Level Of Accuracy, 95% Confidence Level)

Question 7.1. In accordance with the obtained findings, it can be stated that data mining methods are able to predict the level of social responsibility disclosure based on performance criteria with good/favorable accuracy (at least 76%) and can be put in use to predict the future level of disclosure of corporate social responsibility by taking into account the criteria of effective performance and considering it in future analysis and decisions.

Question 7.2. Among the six linear separation analysis algorithms, the nearest neighbor classification algorithm, the decision-making tree algorithm, the linear regression algorithm, the support vector machine algorithm and the neural network algorithm, the one algorithm that is able to predict/project social responsibility disclosure level of companies listed on the Tehran Stock Exchange with greater accuracy (more than 81%) than other algorithms and data mining techniques is the nearest neighbor classification algorithm. Moo rover, in addition to the possibility of comparing various algorithms, the accuracy of the algorithms can also be utilized as a criterion for evaluating the validity of the (social responsibility disclosure level of companies).

8 Conclusion

With the development and expansion of the corporate sector, there is increasingly more focus and attention to the effects of corporate performance on society and the environment. Hence, predicting/estimating corporate social responsibility disclosure levels can help inform investors, stakeholders as well as society in general regarding corporate performance. In previous studies, traditional models have been used to evaluate interactions between social responsibility disclosure level and performance, however, due to the high accuracy and nonlinear and non-parametric properties of new models, including data mining (compared to traditional statistical models), data mining models have become ever more popular. Therefore, in the current study, six supervised learning algorithms (among data mining methods), have been utilized to classify and predict/project the optimal level of social responsibility disclosure of companies listed on the Tehran Stock Exchange.

In this study, after extracting data from corporate financial statements, assessing and reviewing similar researches as well as the regression test findings, ultimately, five performance criteria (QTUBIN, ROA, ROE, P/B, P/E) were selected as effective variables and the following three (company size, financial leverage & firm's longevity) were determined as control variables. In the next stage, the data were divided into two training and experimental categories for the models, and pursuant to utilizing them and obtaining the best outcome in the training sample, the number of neighbors and the distance measurement function were attained. Thereafter, these characteristics were used in the control sample to measure the power/strength of the model.
In regard to the first question of this research, according to the acquired findings, it can be stated that the data mining method with good accuracy (more than 76%) is able to categorize corporate social responsibility disclosure level and predict/estimate the optimal level of corporate social responsibility disclosure within the Iranian economic realities/environment. Concerning the second question, compliant with the obtained findings, the classification algorithm in the vicinity of the nearest neighbor is able to project/predict the social responsibility disclosure level of companies listed on the Tehran Stock Exchange with greater accuracy (more than 81%) than other methods. Moreover, this study's findings indicate that performance metrics can be a good predictor of the optimal level of corporate social responsibility disclosure. In addition, due to the high strength/power of the proposed model, the model utilized in this research can be used to rank the level of corporate social responsibility disclosure.

Given the growing importance of corporate social responsibility and new methods in evaluating such information, in addition to past and present information, stakeholders should also discern and be able to have a good idea concerning what will happen in the future in the field of corporate social responsibility and the business environment in their decision making process. This can also improve the business environment, enhance and develop social performance, etc. What's more, by transmitting positive news about their performance and disclosing their future social responsibility to stakeholders, corporations can also enjoy benefits (present & future) based on theories of signage and legitimacy.

Considering that the corporate social responsibility field is a multifaceted and interdisciplinary subject, for future research it is suggested/recommended that other criteria, including economic and financial criteria, be included in the model toward creating a comprehensive and complete model; Also, according to the characteristics of the genetic algorithm, it is suggested/recommended that the optimal level of corporate social responsibility disclosure be estimated/projected based on performance criteria and via utilizing the genetic algorithm.

One of the present study's constraints was the limitation of the analyzed performance criteria, thereby making the inferences made (based on the research findings) cautious/conservative; In addition, to determine corporate social responsibility disclosure level, annual reports and from boards of directors to general assemblies were used, even though companies may choose to disclose their social responsibility through other means such as newspapers, the internet, audio and video files as well.

References

[1] W. Abbott and R.J. Monsen, *On the measurement of corporate social responsibility: Self reported disclosures as a method of measuring corporate social involvement*, Acad. Manag. J. **1979** (1979), no. 22, 501–515.

[2] R. Agarwal and E. Karahanna, *Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage*, MIS Quart. **24** (2000), 665–694.

[3] A.M. Alimohamadi, A.M. Abbasimehr and A. Javaheri, *Prediction of stock return using financial ratios: A decision tree approach*, Financ. Manag. Strategy **3** (2016), no. 11, 125–146.

[4] A. Alivandi Darani, M. Arabsalehi , H. Amiri and F. Zamani Boroojeni, *Validation tool for assessing the level of disclosure of corporate social responsibility for the companies listed on the Tehran Stock Exchange*, J. Health Account. **9** (2020), no. 1, 78–100.

[5] F.A. Amani and A.M. Fadlalla, *Data mining applications in accounting: A review of the literature and organizing framework*, Int. J. Account. Inf. Syst. **24** (2017), 32–58.

[6] D. Appelbaum, A. Kogan, M. Vasarhelyi and Z. Yan, *Impact of business analytics and enterprise systems on managerial accounting*, Int. J. Account. Inf. Syst. **25** (2017), 29–44.

[7] M. Arabmazar Yazdi, A. Nasseri, M. Nekoee Zadeh and A. Moradi, *The impact of accounting information system flexibility on firm performance with dynamic capabilities approach*, Account. Audit. Rev. **24** (2017), no. 2, 221–242.

[8] P. Chalaki and M. Yosefi, *Predicting earnings management using the decision tree*, Account. Audit. Stud. **1** (2012), no. 1, 110–123.

[9] P.M. Clarkson, Y. Li, G.D. Richardson and F.P. Vasvari, *Does it really pay to be green? Determinants and consequences of proactive environmental strategies*, J. Account. Public Policy **30** (2011), no. 2, 122–144.

[10] D. Cormier and M. Magnan, *Environmental reporting management: A Europeanperspective*, J. Account. Public Policy **22** (2003), no. 1, 43–62.

[11] F. Daniel, F. Lohrke, C. Fornaciari and A. Turner, *Slack resources and firm performance: A meta-analysis*, J. Bus. Res. **57** (2004), no. 6. 565–574.

[12] M. Dastgir and M. Shafiei Sardashti *Data mining, a new Approach in financial field*, Auditor **11** (2011), no. 5, 6–27.

[13] N.C.P. Edirisinghe and X. Zhang, *Portfolio selection under DEA-based relative financial strength indicators: Case of US industries*, J. Oper. Res. Soc. **59** (2008), no. 6, 842–856.

[14] H. Etemadi, R. Hesarzadeh, M. Mohammadabadi and A. Bazrafshan, *Disclosure and firm value: Evidence from Iran's Emergence Stock Market*, Manag. Account. **5** (2015), no. 13, 67–77.

[15] J.R. Evans, *Business Analytics: Methods, models, and decisions*, Prentice-Hall, Boston, MA, 2013.

[16] P. Everaert, *Discovering patterns in corporate social responsibility (CSR) reporting: A transparent framework based on the Global Reporting Initiative's (GRI) SustainabilityR eporting Guidelines*, Faculty of Economics and Business Administration, Ghent University, Belgium, 2009.

[17] S.L. Hart, *A natural resource based view of the firm*, Acad. Manag. Rev. **20** (1995), no. 4, 986–1014.

[18] B.J. Inchausti, *The influence of company characteristics and accounting regulation on information disclosed by Spanish firms*, Eur. Account. Rev. **6** (1997), no. 1, 45–68.

[19] H. Jiawei, M. Kamber and J. Pei, *Data mining: Concepts and techniques*, 3rd edition, Morgan Kaufmann, 2011.

[20] A.F. Jurkus, J.C. Park and L.S. Woodard, *Women in top management and agency costs*, J. Bus. Res. **64** (2011), no. 2, 180–186.

[21] L. Kozma, *K Nearest neighbor algorithm (Knn)*, Helsinki University of Technology, Special Course in Computer and Information Science, Available online at: ww.lkozma.net/knn2.pdf, 2008.

[22] H. Li, J. Sun and J. Wu, *Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods*, Expert Syst. Appl. **37** (2010), no. 8, 5895–5904.

[23] J. Luo, M. Meng and C. Yan, *Analysis of the impact of artificial intelligence application on the development of accounting industry*, Open J. Bus. Manag. **2018** (2018), no. 6, 850–856.

[24] G. Mahdavi, A. Daryaei, R. Alikhani and M. Maranjory, *The relation of firm size, industry type and profitability to social and environmental information disclosure*, Empir. Res. Account. **4** (2015), no. 3, 87–103.

[25] G.H. Mahdavi and A. Ghahramani, *Providing a pattern of fraud detection by auditors using artificial neural network*, Audit Sci. **7** (2015), no. 67, 45–70.

[26] S.M.S. Mousavi Nasab, *Hypothesis; Where is it necessary?*, Res. **3** (2011), no. 2, 5–14.

[27] A. Owusu-Ansah, *Timeliness of corporate financial reporting in emerging capital markets: Empirical evidence from the Zimbawe Stock Exchange*, Account. Bus. Res. **30** (2000), no. 3, . 241–254.

[28] D.M. Patten, *Media exposure, public policy pressure, and environmental disclosure: An examination of the impact of TRI data availability*, Account. Forum **26** (2002), no. 2, 152–171.

[29] M. Namazi and R. Gholami, *Comprehensive ranking model of companies via accounting information, balanced scorecard and PAPRIKA technique*, Account. Knowledge **7** (2017), no. 27, 7–33.

[30] M. Namazi and N.R. Namazi, *Ranking firms based on the performance evaluation criteria via multiple attribute TOPSIS technique and comparing evaluation criteria (Evidence from companies listed in Tehran Stock Exchange)*, Financ. Account. Res. **8** (2016), no. 2, 39–64.

[31] G. Perriere and J. Thioulouse, *Use of correspondence discriminant analysis to predict the subcellular location of bacterial proteins*, Comput. Meth. Prog. Biomed. **2003** (2003), no. 70, 99–105.

[32] Y. Qiu, A. Shaukat and R. Tharyan, *Environmental and social disclosures: Link with corporate financial performance*, Br. Account. Rev. **48** (2016), no. 1, 102–116.

[33] M. Rodriguez-Fernandez, *Social responsibility and financial performance: The role of good corporate governance*, BRQ Bus. Res. Quart. **19** (2016), no. 2, 137–151.

[34] M.V. Russo and P.A. Fouts, *A resource-based perspective on corporate environmental performance and profitability*, Acad. Manag. J. **40** (1997), no. 3, 534–559.

[35] M. Salehi and L. Farrokhi Pilehrood, *Predicting earnings management using neural network and decision tree*, Financ. Account. Audit. Res. **10** (2017), no. 37, 1–24.

[36] H. Servaes, and A. Tamayo, *The impact of corporate social responsibility on firm value: The role of customer avareness*, Manag. Sci. **59** (2013), no. 5, 1045–1061.

[37] S. S. Singhavi and H. B. Desai, *An empirical analysis of the quality of corporate financial disclosure*, Account. Rev. **46** (1971), no. 1, 129–138.

[38] A. Skouloudis, K. Evangelinos and F. Kourmousis, *Assessing non-financial reports according to the global reporting initiative guidelines: Evidence from Greece*, J. Clean. Prod. **18** (2010), no. 5, 426–438.

[39] A. Soroushyar and M. Akhlaghi, *The comparative assessment of data mining methods effectiveness to forecasting return and risk of stock in companies Listed in Tehran Stock Exchange*, Financ. Account. Res. **9** (2017), no. 1, 57–76.

[40] J. Sun and H. Li, *Data mining method for listed companies' financial distress prediction*, Knowledge-Based Syst. **21** (2008), no. 1, 1–5.

[41] C. F. Tsai and Y. J. Chiou, *Earnings management prediction: A pilot study of combining neural networks and decision trees*, Expert Syst. Appl. **36** (2009), no. 3, 7183–7191.

[42] R. Verrecchia, *Essays on disclosure*, J. Account. Econ. **32** (2001), no. 1-3, 97–180.

[43] Y. Hassas Yeganeh and G. Barzegar, *Identifying the components and indicators of corporate social responsibility in Iran*, Soc. Cult. Dev. Stud. **2** (2013), no. 1, 209–234.

[44] X.Z. Zhang, Y. Hu, K. Xie, W.G. Zhang, L.J. Su and M. Liu, *An evolutionary trend reversion model for stock trading rule discovery*, Knowledge-Based Syst. (2015), no. 79, 27–35.

[45] L.G. Zhou, D. Lu and H. Fujita, *The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches*, Knowledge-Based Syst. **85** (2015), 52–61.