



## Information disclosure by data mining approach

Naser Azad<sup>1</sup>, Vahid Ranjbar<sup>2</sup>, Davoud Khani<sup>3</sup> and Sara Taheri Moosavi<sup>4\*</sup>

<sup>1</sup>Faculty Member of Azad University, South Branch of Tehran, Iran; <sup>2</sup>University of Sistan & Baluchestan, Iran

<sup>3</sup>Ferdowsi University of Mashhad; <sup>4</sup>University of Payam noor, Iran

Ranjbar.v@gmail.com<sup>2</sup>, Davoud\_khani@yahoo.com<sup>3</sup>, sara.tahery\_moosavi@yahoo.com<sup>4\*</sup>

### Abstract

Undoubtedly, one of the most important factors in the profitable securities stock market investment is timely and regulated disclosure of companies' information. Several standards affect transparency and disclosure of information but understanding and effect percentage of these factors on the level of disclosure is very complicated and beyond the investor time limitation and costs. This study aims to investigate the effective factors on information disclosure level of present companies in Tehran stock exchange market and is one of the first studies to have undertaken empirical investigation of the effective factors on the level of information disclosure of presence companies in Tehran stock exchange market, Iran. Doing so, effective factors on information disclosure in the securities stock market, had been investigated nationally and internationally in previous works, reviewed and 19 factors selected regarding Iran capital market conditions. After using data mining approach and specifically decision trees algorithm and determining most important factors, resulted findings show that Audit Organization factor (AO) is inversely related to good information disclosure while factors such as "Return of Assets (RS)", "Share Percentage of Major Shareholders (SPMS)", "Dept Ratio (DR)", "Chairman Position (CP) (bounded/non-bonded)" and "Share Percentage of Governmental Companies (SPGC)" have direct relationship with it. Analysis of poor disclosure rules, on the other hand, did not provide an appropriate model.

**Keywords.** Information disclosure, Tehran Stock Exchange, Data Mining, C5.0

### Introduction

In recent years, plenty of financial scandals have occurred in famous global companies like Enron and World.com. With the collapse of these companies, corporate governance has been talked much more than before (Alberts & Doroffe 2003). Literature review shows no consensus definition of corporate governance. OECD (Organization of Economic Cooperation and Development) principles cover six key areas of corporate governance: Creation of conditions that ensures the effectiveness of the framework of corporate governance, shareholders rights, role of shareholders in corporate governance, disclosure, transparency and responsibility of the board (Bushee & Noe, 2000). One important aspect of corporate governance considered above is information transparency concept. Clear information is one of the managers' tools for acting their responsiveness role (Dicker, 2003). Presence of clear information reduces information asymmetry and has a significant role in reducing the cost of financing exchanges and can lead to facilitate risk management, identify the best investment opportunities, disciplined and controllable managers, and facilitate exchange of products and services, mobilization of society savings and the fair distribution of wealth (Donnelly & Mulkahy, 2008). Improving information disclosure to improve transparency is one of the best goals of global changes and main initiatives in corporate governance reform process in America, Britain and other countries (Ferris *et al.*, 2002). Understanding the different effective variables on information disclosure and companies transparency is very difficult and beyond the investors time limitation and costs. So according to

specified criteria the securities stock market in different countries releases report rankings of companies' information disclosure at specified time periods. Based on acquired scores from communication and disclosure quality window, to help more enlightened investing, Tehran stock exchange market announces companies' information disclosure ranking every quarter of year. Numerous studies have conducted at the international level about factors affecting the amount of information disclosure; factors such as: Corporate governance (board composition, leadership structure, board structure, and audit committee), ownership structure (the focus of family ownership, state ownership, foreign ownership, institutional ownership, and managers ownership) and the special features of the company (firm size, debt ratio, type of audit, current ratio, return of stockholders equity, etc.). Which of these factors has the greatest impact on the companies' information disclosure ranking and should be more attended by investors? This study has to do with factors that were extracted from these articles (as input or independent variable) and matching these with the Iranian capital market and use of data mining approach, provides a model for classifying listed companies in Tehran stock exchange market and compares the accuracy of this model to the stock ranking.

### Information disclosure

Corporate governance principles of OECD that was founded in 1999 by 30 member countries represented as a reference for all the world countries. Following extensive review process that led to the adoption of the revised principles of corporate governance in the spring of 2004, today, these principles are a reflection of a global



consensus about the critical importance of good corporate governance in preservation of countries life and economic stability (Lu, et al. 2009). With an overview of different definitions of corporate governance, most comprehensive definition of concept can be expressed as: rules, regulations, structures, processes, cultures and systems that facilitate achieving the goals of accountability, transparency, justice and the rights of beneficiaries (Goodwin & Prather, 2002). In this definition, transparency is one of the objectives of corporate governance (Ghose & Rajan, 2006). Despite the numerous definition of the concept of corporate governance, transparency has always been emphasized as a common aspect of these definitions. Financial transparency and full and fair information disclosure are considered as crucial factors for proper functioning of an efficient capital market. The disclosure term is referred to a diverse set of mandatory and voluntary information including: notes and financial statements, board reports, management analysis, management forecasting and etc (Gordon *et al.*, 2006). According to the guidelines of the Securities and Exchange Organization issued in 2007, all companies listed on the Tehran stock exchange market were required to disclose their financial information, ownership structure and other investors' requirements. Information asymmetry is one result of unfavorable companies' information disclosure. Information asymmetry refers to a situation that executives' awareness of companies' activities is much more than that of actual and potential investors and other stockholders (Makhija & Patton, 2004; Paul *et al.*, 2001). If required information distributed asymmetrically among individuals, different outcomes will be resulted from a single subject (Sweeney, 2003). Information asymmetry about companies' stocks causes that intrinsic stocks value and their value for stocks market investors be different. Then the real value of companies' stocks is inconsistent with the expected shares value in capital market. Companies that offer higher transparency will have better performance in terms of stocks price (Bushee & Noe, 2000). The result of a survey on a sample of 40 member companies listed in Tehran stock exchange market shows increasing in disclosure amount reduces the cost of common stocks. On the other hand, investors are willing to invest in companies that posses more information disclosure and less risk (Lu *et al.*, 2009). Review of published papers in the field of information disclosure explains that information disclosure is associated with corporate governance, ownership structure and firm characteristics.

*The Relationship between corporate governance, ownership structure and companies' characteristics with the level of information disclosure:* Companies prefer to disclose good rather than bad information. So, companies' information disclosure situation is very important for global markets. Information transparency evaluation of companies is much more complicated than

offering a series of financial statements and annual reports. The impact of corporate governance, ownership structure and firm characteristics on the level of disclosure has been investigated in several articles. Verrecchia in 1983 showed information disclosure was effective in improving corporate governance and reducing agency problems (Alsaeed, 2006). The relationship between institutional investors and the amount of information disclosure was investigated in a study and identified that institutional investors prefer to invest in firms with greater information transparency (Barako *et al.*, 2006). In examining the factors affecting the information transparency in developing countries, in particular Kenya, factors such as: Corporate governance, corporate ownership structure and characteristics were studied and determined that board composition, the united role of CEO and chairman of the board, percentage shares of the top 20 shareholders, company audited by four great auditors, return of stockholders equity and current ratio are the factors that had no effect on information disclosure extension in Kenya stock market, despite the factors such as: companies' audit committee, shares percentage of foreign shareholders, shares percentage of institutional investors, firm size and debt ratio which had significant effect on information disclosure (Core, 2001). In terms of such criteria: ownership structure (shares percentage of main shareholders (above 5%)), shares distribution of shareholders, firm size and debt ratio; 22 companies in Ghana Stock market, as a developing country, were studied and these results obtained: increasing the number of major shareholders reduces information disclosure, shares distribution is effective on disclosure of information, larger companies tend to have higher disclosure and finally debt ratio has no noteworthy impact on information disclosure of Ghana stock market companies (Damianides, 2005). In 2002, study of 51 companies from presented 61 companies in Ireland capital market showed that by increasing the number of non-duty members of the board, voluntary disclosure of information increases. Factors examined in this study included: percentage of non-duty to total members of board, chairman position (as being bounded or non-bounded), percentage of in-hands shares of institutional investors, percentage of in-hands shares of managers, firm size and finally board size (number of board members) (Eng & Mak, 2003). Considering these factors: firm size, ownership structure, government ownership, return of stockholders equity, intangible assets, bank dept and etc; the effect of ownership structure of presented 43 companies in Czech Republic stock exchange market was assessed, as well (Ge & McVay, 2005). In a study in France, in examining 207 companies in the period of 1998 to 2001, factors such as: ownership structure and firm characteristics (firm size, performance, industry type, dept ratio) were investigated in relation to information disclosure, and determined that: companies that are properties of major shareholders are less willing to



disclose information, increasing foreign shareholders leads to more information disclosure, the presence of more institutional investors reduces information disclosure, larger companies are more willing to disclose, the debt ratio is inversely related to disclosure and finally, companies that are in the high-tech field possess a higher disclosure. In a study of 40 companies in Saudi stock exchange market in 2003, the relationship between the level of companies' information disclosure and companies' characteristics such as: variables related to structure (company size, dept, distribution of property and companies lifetime), variable related to performance (the profit margin, return of stockholders equity and current ratio) and market-related factors (industry type and size of audit firm) was studied and these results obtained: significant positive relationship was between firm size and disclosure level, there was no specific relationship between dept ratio and the level of disclosure, there was not reported a significant relationship between property distribution and disclosure level, companies lifetime showed no difference in the level of their information disclosure, there was no significant relationship among companies performance, industry type and disclosure level, and finally size of auditor firm had no effect on the information disclosure level of companies (Lakhal, 2007). In 1995, based on a sample of 158 companies listed on Singapore stock exchange market, Ang and Mok studied the effect of ownership structure and board composition on voluntary disclosure of information and concluded that: loss of managers' ownership, also government ownership expansion enhances disclosure level, major shareholders ownership has no relationship with the level of disclosure, increasing the number of board managers that are outside the company decreases information disclosure, larger firms and companies with lower dept have a higher degree of disclosure (Tsamenyi *et al.*, 2007). Our variables used in this paper have been extracted from reviewing and refining these articles and other studies in the field of information disclosure.

#### *Data mining*

In researches, typically a hypothesis is formulated and then, based on observed reports proved or disproved. But real world problems are extraordinary non-linear and a specific model needed to be created that takes into consideration all the variables required for solving that specific problem. So, common statistical methods are almost ineffective to apply (Roiger & Geatz, 2003). Due to development of database systems and the high volume of data stored in them, new tools are required to process these data and making resulted information available for users. In this way many methods have come through to life idiomatically called "knowledge discovery" techniques. Data mining is one of the most important among them. For extracting hidden knowledge of huge amounts of data, data mining uses exploratory data analysis, one branch of statistics science. It also is closely related to artificial intelligence and machine

learning. Many definitions have been proposed in the literature of data mining. Perhaps the simplest can be expressed as: "Extracting knowledge from large volumes of data" (Kovalerchuk & Vityaev, 2002; TCC, 1999).

Different data mining strategies include: relationship, order, classification, clustering and forecasting. In classification, the most common data mining strategy, also used in this paper, historical data are applied for model creation. Produced model then, can be exploited to classify new data set (Wang, 2008; Islan & Brankovic, 2004). Data mining techniques generally used for classification contain neural networks and decision trees (Han & Kamber, 2006; Fayyad & Shapiro, 1996). But since classifications made by decision trees are more understandable than that of neural networks and because we want all investors to be beneficent of results of this study, it was preferred to use decision trees approach.

#### *Decision trees and rules extraction*

One of the common methods of classification and forecasting problems is the decision trees. Decision trees are a way to display a series of rules that lead to a class or value. Most decision tree algorithms have been developed and introduced in the last 30 years. These algorithms have successfully been used in the areas of banking and financial. An example is using these algorithms to check credit risk of bank loans applicants (Pedrycz & Sosnowski, 2000). ID3 is one of decision trees algorithms that introduced by Coelin at the end of 70's and early 80's. C4.5 algorithm was developed by the same person shortly after ID3 (Wang *et al.*, 2000). CART is another popular algorithm in this context developed by Berryman, Friedman, Aston, and Alshon (Abonyi *et al.*, 2005). The difference between C4.5 and CART is that CART trees produce only two branches at each node and each branch leads to another decision node or a leaf node while in C4.5 algorithm more than two branches can be established from each node. Rules extraction is a way of producing a set of rules that classifies cases. Although decision trees are capable of developing rule sets but rule extraction procedure produces a set of independent rules that is not necessarily a tree. Since rules maker has not to be divided in each level and can look at the next steps, it will be able to find different and sometimes better patterns (Chaturvedi *et al.*, 1993). Unlike trees, produced rules may not cover all possible cases. Also, rules may be conflicting in predictions, so, we have to select a rule to fallow in any case. One way to solve this conflict is appointing a confidence value to each rule and use of rules with higher confidence value (Chen *et al.*, 2000). In this paper the latest version of C4.5 algorithm, C5.0, has been used. The most common methods for increasing the accuracy of classification algorithms are Bagging and Boosting methods. In Boosting method, the output of several models are combined to create a higher and stronger model. In this approach, several predictions, each time with different training data, selected from previous runs, are run and finally the most repeated



Table 1. Effective factors on companies information disclosure

Data Source	Variable Definition	Variable Name
Audited financial statements and RAHAVARD NOWIN software	Ratio of non-bounded members to all members	Board Composition
Audited financial statements	.....	The number of board members
Audited financial statements and RAHAVARD NOWIN software	If chairman be non-bounded, number one else zero	Is chairman bounded or non-bounded?
Audited financial statements and RAHAVARD NOWIN software	If manager director be president , number one else zero	Is manager director president or not?
Audited financial statements	If audit organization balance the company, number one else zero	Does Audit organization balance the company or not?
Audited financial statements	Share percentage of major shareholders (above 5%) to total	Share Percentage of Major Shareholders (above 5%)
Audited financial statements	Share percentage of institutional shareholders to total shares	Share Percentage of Institutional Shareholders
Audited financial statements	Share percentage of governmental companies to total shares	Share Percentage of Governmental Companies
Audited financial statements	Natural logarithm of total assets	Companies Size
Audited financial statements	Ratio of total debt to total assets	Debt Ratio
Audited financial statements	Ratio of total net profit to sum of stockholders equity	Return of Stockholders Equity
Audited financial statements	Ratio of current assets to current	Current Ratio
Audited financial statements	Ratio of bank debt to total assets	Bank Debt Ratio
Audited financial statements	Ratio of total net profit to shares	EPS
Audited financial statements	Ratio of total net profit to total assets	Return of Assets
www.Irbourse.ir Website and Audited financial statements	Ratio of shares' end of year price to EPS	P/E
RAHAVARD NOVIN software	.....	Return of Shares
www.Irbourse.ir Website and Audited financial statements	Ratio of (each share's end of year price*the number of shares+ total	MBA
www.Irbourse.ir Website and Audited financial statements	Ratio of (each share's end of year price*the number of shares) to sum of	MBE
Ranking report presented by stock exchange market	.....	Company's Information Disclosure Ranking

answer is chosen. Although this process is time consuming, but the answers will be more reliable. This method was first used in 1996 and these days due to the extending computational power of computers has been increasingly accepted (Nemeth *et al.*, 2006). According to Roiger and Gates (2003), the results of applying C4.5 algorithm with Boosting method show higher accuracy than those of Bagging method.

Although data mining approach has many applications in various fields such as: marketing, banking, insurance industry, manufacturing, CRM and etc, but many few researches have been done on its application in information disclosure field. One of the most important researches is a study by Lu and Chen in Taiwan stock exchange market (Lu *et al.*, 2009). In their study, regarding the type of companies ownership structure in East Asia, more emphasis has performed on ownership structure (family ownership, state ownership and etc.), while in this study most effective variables in the disclosure of information, regarding Iranian capital market characteristics, have been investigated. Furthermore, the study includes information of companies in Tehran stock

exchange market for the three fiscal years which is more comprehensive than that of aforementioned study (their information were related only to one fiscal year), and thus is more accurate and reliable.

### Methodology

According to companies' guidelines and standards for information disclosure, and to help more wisely financing, the Tehran stock exchange market announces companies' information disclosure ranking every three months. In this study, effective then using data mining approach and specifically decision trees algorithm most frequent factors in the produced set of rules extracted and considered as the most significant measures of companies' information transparency that investors must be aware of them. The following steps have gone to resolve the issue investigated in this paper:

#### *Determination of effective factors on companies' information disclosure*

As was discussed in section 2.1.1, in determining these factors, an extensive review of published articles in national and international level was performed and every factor which had positive or negative influence on the

disclosure of information considered. Then, by matching these factors to capital market of Iran (here, match means availability of these factors), a total of 19 independent variables and a dependent variable (qualitative ranking of information disclosure, presented by Tehran stock exchange market) were considered (Table 1).

*Determination of companies surveyed in the study*

To increase the study accuracy, presence companies in Tehran stock exchange market, from 2009 to 2011, were selected according to the following limitations (Table 2):

Table 2. Applied data records in this study

Year	Total Companies	Studied Companies (regarding mentioned limitations)
2009	812	370
2010	830	362
2011	632	313
Sum	2264	1045

- The fiscal year of company must be ended to 12.29
  - The company must not be part of an informal panel
  - The company must has research required information
- To collect companies' information, for the period of 2008-2011, Companies' financial statement CDs were used. In addition, required information related to 2008, was gathered via Tehran stock exchange market website. RAHAVARD NOVIN software was also used to collect our required information.

*Companies ranking based on information disclosure*

Quarterly reports of Tehran stock exchange market declare companies' ranking regarding publishers announcement scores (calculated based on the presenting time of related information to each share earnings forecasting, no audited financial statements for the periods of 4, 5, and 10 months, and etc.). This study has used the companies ranking report considering the quality of companies' disclosure and appropriate announcement at the end of 2009, 2010, and 2011 (Table 3). After ranking implementation, the records number of 362 (equivalent to 45.8%) were determined as good, and 370 records (equivalent to 54.2%) identified as bad disclosure.

Table 3. Companies ranking based on quantitative score (Presented by Stock exchange Organization)

Quantitative Score	Qualitative Ranking
SQ < 50	Poor
SQ >=50	Good

*The algorithm used in the ranking of companies' information disclosure*

In this study, the C5.0 algorithm of SPSS (Clementine) software package has been practiced. C5.0 is the last version of decision trees algorithm and rule extraction procedure that producing if-then rules simply can help investors to consider effective factors on the level of companies' information disclosure and consequently will lead to more informed investments.

**Results of C5.0**

In the form of an Excel file with 20 columns, including 19 independent variables (inputs) and a dependent variable (target), our data were entered into SPSS software (Table 4). After running the C5.0 algorithm, if/then rules were created in two levels of good and poor information disclosure (Table 5). According to these rules, 1045 records of 2264 data records were correctly classified (83.78) and the rest of them wrongly located in non-integer classes (Table 6). So, this set of information predicts companies' information disclosure with 83.87 percentage of accuracy.

Boosting and Bagging are the most common methods of increasing the accuracy of classification algorithms. In this paper, Boosting approach has been manipulated to escalate model accuracy. Different models with different accuracy were obtained using different K (here, K is the number of categories). The most proper model was related to K=3 with 91.73 accuracy (Table 7). This table indicates that increasing K does not necessarily bring higher accuracy. For example, the model with K=5, provides less accuracy. By increasing the accuracy of forecasting model, the number of records classified incorrectly reduced from 66 to 43 (Table 8). Investigating the rules set with highest accuracy shows that: the organization that audit the company, dept ratio, return of asset and chairman position, as being bound or non-bound, are the variables that have most frequency in the rules and investors should consider them in their assessment of most effective factors on companies' information disclosure (Table 9).

**Sorting resulted C5.0 rules**

Although these four factors referred to in section 3.5, make investors able to assess companies' information disclosure in a better frame, but the remarkable thing is that they should consider these examined 19-fold factors (our variables in this study) to better understand a specific company's information disclosure and subsequently to take more correct decisions. With a glance at the rules presented in Table.5 and in particular, rules 5 and 13 related to the good disclosure, specified that these rules are quite general and can intrude investor decisions. To solve this problem and help resolve existing conflicts in the rules, this is tried to arrange the rules so that rules with higher priority would be considered.

Rules sorting approaches utilize scoring mechanisms. Among the most important of these approaches are CSA and ACS. In CSA approach, first, based on the confidence value of each rule all the extracted rules are ordered in descending, and then considering each rule support factor, rules with similar confidence value sorted in the same manner. Finally, if rules support factor also be alike, they are arranged in ascending regarding their number of presumptions. Unlike the CSA, ACS approach has different strategy.



Table 4. An example of data records used in research in the Excel file format

Board Composition (BC)	0.6	0.2	0.6	0.6	0.8	0.8	0.6	0.8	0.6	1	0.6
Number of Board Members (NBM)	5	5	5	5	5	5	5	5	5	5	5
Chairman Position (Bounded/non-bounded) (CP)	0	1	1	1	1	1	1	1	1	1	1
Managing Director (President/ Vice president) (MD)	0	0	0	0	1	0	0	0	0	0	0
Audit Organization (AO)	0	1	1	1	0	0	0	0	0	0	0
Share Percentage of Major Shareholders (above 5%) (SPMS)	73.07	83.97	79.58	62.8	84.19	73.1	61.133	82.45	69.9	83.46	83.46
Share Percentage of Institutional Shareholders (SPIS)	59.58	10.09	35.5	20.04	6.96	0	53.629	39.21	74.2	89.55	89.55
Share Percentage of Governmental Companies (SPGC)	0	0	41.91	43.29	4.63	0	0	2.49	2.5	0	0
Company Size (CS)	5.227	5.785	6.567	7.811	6.012	5.697	5.102	6.362	6.627	5.609	5.609
Debt Ratio (DR)	0.39	0.571	0.824	0.895	1.439	0.454	0.759	0.412	0.552	0.435	0.435
Return of Shareholders Equity (RSE)	0.338	0.453	0.386	0.334	0.573	0.124	0.568	0.102	0.267	0.594	0.594
Current Ratio (CR)	2.076	1.515	1.107	0.699	0.371	1.611	1.271	0.558	0.733	1.2	1.2
Ratio of Bank Dept (RBD)	0.016	0.395	0.212	0.412	1.314	0.03	0.024	0.299	0.301	0.081	0.081
EPS	697.0	1129	834.4	504.9	-3227.1	425.342	1729.3	500.891	422.84	1818.2	1818.2
Return of Assets (RA)	0.206	0.195	0.068	0.035	-0.251	0.068	0.137	0.06	0.12	0.336	0.336
P/E	8.096	1.909	9.582	5.907	-0.291	4.152	3.002	5.386	3.51	5.532	5.532
Return of Shares (RS)	62.66	-1.98	-1.18	22.68	-5.43	-74.36	5.62	-42.82	-33.8	38.81	38.81
MBA	2.062	0.942	1.474	1.102	1.512	0.736	1.17	0.736	0.972	2.291	2.291
MBE	2.74	0.866	3.702	1.973	-0.167	0.517	1.704	0.55	0.938	3.284	3.284
Score	39	43	14	4	51	70	66	58	63	70	70
Rank	0	0	0	0	1	1	1	1	1	1	1

In this approach first considering the number of each rule presumptions, they are sorted in descending and then rules with similar presumptions frequency ordered in descending based on their confidence value. If a number of rules had the same confidence value they would be ordered according to the amount of support (Wang *et al.*, 2000).

Kounen and Lang stated that: "the ACS approach ensure that more specific rules have higher priority than more general rules. To remove the general rules presented in Table.5, and prevent interference in the

investors' decision making process and since we want rules with lower presumption frequency to be of lower priority, the ACS approach has been used in this paper. Confidence and support values of C5.0 resulted rules are presented in Fig.1. Sorted rules by ACS approach are shown in Table 10, as well.

#### Analysis of related rules to good information disclosure

As Table.10 shows, in the ranking of ACS approach, rules No.5 and 13 have the lowest priority. These rules, according to Table.5, have only one presumption and from this point have the highest degree of universality.



Table 5. C5.0 Extracted Rules

	Poor Information Disclosure Rules	Good Information Disclosure Rules
1	if (MBA > 1.869) and (RSE <= 0.363) then Disclosure is Poor	If EPS <= 1925.544 and (RSE) > 0.363 and (CP) > 0 and (AO) <= 0 and (SPMS) > 32.400 and (SPGC) <= 5.220 and (MD) <= 0 and (DR) <= 0.789 then disclosure is Good
2	if the (RA <= 0.061) and (CP) > 0 and (SPIS) > 3.100 and (DR) <= 0.789 and (RBD) <= 0.217 then Disclosure is Poor	If EPS <= 1925.544 and (RS) > 20.390 and (RA) > 0.068 and (CP) > 0 and (AO) <= 0 and (SPMS) > 32.400 and (SPGC) <= 3.870 and (MD) <= 0 and (DR) <= 0.789 then disclosure is Good
3	if (CP) > 0 and (AO) > 0 and (SPMS) <= 75.200 and (SPIS) <= 19 and (MD) <= 0 then disclosure is Poor	If MBA <= 1.018 and (RA) > 0.061 and (BC) > 0.167 and (CP) <= 0 and (AO) <= 0 and (SPMS) > 32.400 then disclosure is Good
4	if MBA > 1.090 and (DR) > 0.789 then disclosure is Poor	If MBA <= 1.803 and (CS) > 5.970 and (RA) > 0.061 and (CP) > 0 and (MD) > 0 then disclosure is Good
5	If (CS) > 6.733 and (AO) > 0 and (MD) <= 0 then disclosure is Poor	If (DR) <= 0.169 then disclosure is Good
6	If (RA) <= 0.061 and (AO) > 0 then disclosure is Poor	If (RA) > 0.061 and (BC) <= 0.667 and (AO) <= 0 and (SPMS) > 32.400 and (MD) > 0 and (DR) <= 0.789 then disclosure Good
7	If (EPS) < 482.408 and (DR) > 0.789 then disclosure is Poor	If (RSE) <= 0.165 and (AO) <= 0 and (SPIS) <= 49.080 and (DR) <= 0.789 and (RBD) > 0.471 then disclosure is Good
8	If (RS) <= 22.380 and (RA) > 0.061 and (CP) <= 0 and (AO) > 0 and (DR) > 0.416 then disclosure is Poor	If (RA) <= 0.052 and (BC) <= 0.800 and (CP) > 0 and (AO) <= 0 and (SPIS) <= 3.100 and (DR) <= 0.789 then disclosure is Good
9	If (RSE) > 0.330 and (CP) > 0 and (AO) > 0 and (SPIS) <= 19 and (MD) <= 0 and (DR) > 0.416 then disclosure is Poor	If EPS <= 591.867 and MBA <= 1.869 and (P/E) > 4.052 and (RS) <= 20.390 and (RA) > 0.061 and (BC) <= 0.857 and (AO) <= 0 and (SPMS) > 32.400 and (SPGC) <= 5.220 then disclosure is Good
10	If (RA) > 0.061 and (BC) <= 0.167 and (AO) <= 0 and (MD) <= 0 then disclosure is Poor	If (P/E) <= 11.192 and (RA) > 0.061 and (BC) > 0.167 and (CP) <= 0 and (AO) <= 0 and (SPMS) > 32.400 and (RBD) > 0.143 then disclosure is Good
11	If (RA) > 0.061 and (SPMS) <= 32.400 then disclosure is Poor	If (CS) <= 6.733 and (RA) > 0.061 and (BC) > 0.571 and (CP) > 0 and (AO) > 0 and (SPIS) > 19 and (DR) <= 0.789 then disclosure is Good
12	If (RS) <= 20.390 and (BC) > 0.857 and (DR) > 0.169 then disclosure is Poor	If (RS) > 0.061 and (AO) > 0 and (DR) <= 0.416 then disclosure is Good
13	If (RA) <= 0.252 and (SPGC) <= 9.170 then disclosure is Poor	If (DR) <= 0.789 then disclosure is Good

Table 6. Model forecasting with 87.3 % of accuracy

Forecasting Ranking	Actual Ranking	
	Poor	Good
	Poor	512
Good	78	402

Table 7. Accuracy comparison of C5.0 extracted rules regarding different K

Without Boosting Method	83.87
Model with 3 Classifier (K=3)	91.73
Model with 5 Classifier (K=5)	86.73

Table 8. Model Forecasting with 91.73 % of Accuracy

Forecasting Ranking	Actual Ranking	
	Poor	Good
	Poor	527
Good	22	448

So, such rules are not terribly useful to help investors make decision. Rules No.2 and 9 have the highest priority and each of them, with nine of nineteen variables, has the greatest presumption frequency. Meanwhile, compared to other rules, in terms of confidence value (power of rule) (with 96.7 and 88 percent), and the level of support (with 28 and 23), are in good condition. Examining rules No.2, 9, 1, 10, and 11, as five rules with top priorities shows:

- Audit Organization (AO) factor has presented in all of these rules. Four of these rules express that companies not audited by AO have a good level of disclosure and only one rule shows otherwise.
- Rule: "return of assets > 0.061" with four replicates in these rules is largely oblivious.
- Four of the five rules state that if "shares in-hands of stockholders (above > 5%) > 32.4" disclosure is good.

- Factors such as: "dept ratio  $\leq 0.789$ ", "chairman is bounded" and "government stock  $\leq 5.220$ " with three replicates, must be considered.

**Analysis of related rules to poor information disclosure**

Table 9. Determination of most frequent factors in extracted rules

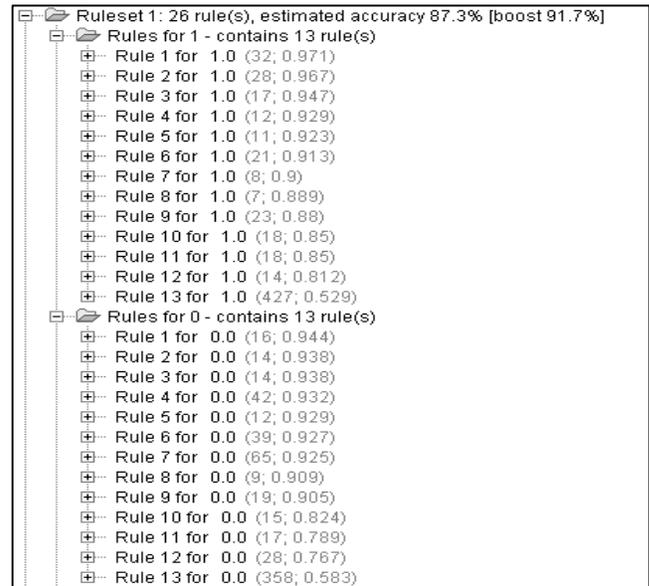
Effective factors on information disclosure	Frequency of poor information disclosure rules	Frequency of good information disclosure rules	Sum
BC	6	2	8
NBM	0	0	0
CP	7	4	11
MD	4	4	8
AO	10	6	16
SPMS	6	2	8
SPIS	3	3	6
SPGC	3	1	4
CS	2	1	3
DR	9	6	15
RSE	2	2	4
CR	0	0	0
RBD	2	1	3
EPS	3	1	4
RA	9	6	15
P/E	2	0	2
RS	2	2	4
MBA	3	2	5
MBE	0	0	0

Because high percentage of poor information disclosure rules (8 of 13 rules) included 2 or 3 of 19 effective variables, and this led to unnecessary general rules, analysis of poor disclosure rules did not provide an appropriate model. It is noteworthy that many conflicts have been seen in the resulted rules.

**Conclusion**

Transparency and disclosure concept is very determining in capital market. As was argued in this article, many factors play role in disclosure of companies' information. Identifying these factors is out of investors' expertise and capabilities. By reviewing the major published papers in national and international level, 19 effective factors on information disclosure were selected regarding capital market characteristics of Iran. Using C5.0 algorithm, the most important factors were extracted from 540 data records as 19 input variables. Next, to reduce forecasting error and improve model accuracy Boosting method was applied and 91.73 percentage of

Fig.1. Confidence and support values of C5.0 extracted rules



accuracy obtained. Finally, using ACS approach extracted rules were sorted based on their priorities and good solutions were provided to investors in the section of good information disclosure rules.

Our results show that Audit Organization (AO) factor has inverse relationship with good information disclosure. In other words, companies not audited by AO have a good level of disclosure. Furthermore, while (regarding "Good Disclosure" rules) factors such as "Return of Assets (RS)" and "Share Percentage of Major Shareholders (SPMS)" with four frequencies, and "Dept Ratio (DR)", "Chairman Position (CP)" and "Share Percentage of Governmental Companies (SPGC)" with three frequencies have direct relationship with good information disclosure level, analysis of poor disclosure rules did not provide an appropriate model and many conflicts were seen in the resulted rules.

**Recommendations**

In the present study, 2006, 2007 and 2008 fiscal year information were used. It is hoped that by increasing the number of fiscal years and present experts in Tehran stock exchange market, rules obtained from such studies can be developed in software packages and be available as an expert system approach to help shareholders make better and more accurate decisions. In Iran capital market very few studies have examined the factors are effective on investment by using data mining approach. So, it is recommended that since data mining is considered

Table10. Rules sorting using ACS approach

	Poor Disclosure										
ACS	9	2	3	8	10	5	12	1	4	6	7
Support of	19	14	14	9	15	12	28	16	42	39	65
Confidence	0.88	0.938	0.938	0.909	0.824	0.929	0.767	0.944	0.932	0.927	0.925
Number of	6	5	5	5	4	3	3	2	2	2	2

among the new functional approaches of statistic science and information of databases is rapidly enlarging, researchers in information technology try to use various data mining algorithms in the financial and stock market matters.

### Implications

The key implication of these findings for stock exchange market managers includes support for the use of clear information disclosure policies at a strategic level of their companies, so they will be able to finance their activities accurately and timely.

### Research limitations

The main limitation of this study was the lack of companies being cooperative. This makes the process of gathering information longer than usual. These companies argued that providing their information may risk the security of their businesses. On the other hand, some of these companies had little familiarity with the standards of transparency and information disclosure. As awareness of these standards was the key for gathering needed information, to familiarize companies with the standards in a way that would not undermine confidence in them, was another important challenge that we faced.

### Reference

1. Abonyi J, Feil B Nemeth S and Arva P (2005) Modified Gath-Geva clustering for fuzzy segmentation of multivariate time-series. *Fuzzy Sets and Systems, Data Mining Special Issue*. Available on-line from Science Direct.
2. Alberts C and Doroffe A (2003) *Managing information security risks: The OCTAVESm Approach*. Boston, MA, Addison-Wesley.
3. Alsaeed K (2006) The association between firm-specific characteristics and disclosure. The case of Saudi Arabia. *Managerial. Audit. J.* 21 (5), 476-496.
4. Barako, Dulacha G, Hancock, Phil, Izan, H Y (2006) Factors influencing voluntary corporate disclosure by Kenyan Companies. *Corporate Governance*.14 (2), 107-125.
5. Bushee BJ and Noe CF (2000) Corporate disclosure practices, institutional investors, and stock return volatility. *J. Accounting. Res.* 38, 171-202.
6. Chaturvedi AR, Hutchinson GK and Nazareth DL (1993) Supporting complex real-time decision making through machine learning. *Decision Support Systems*. 10 (2), 213-233.
7. Chen G, Wei Q and Kerre E (2000) Fuzzy data mining: Discovery of fuzzy generalized association rules. *Recent Issues on Fuzzy Databases*. 45-66.
8. Core (2001) A review of the empirical disclosure literature: Discussion. *J. Accounting & Economics*.3, 441-456.
9. Damianides M (2005) Sarbanes-Oxley and IT Governance: New guidance on IT control and compliance. *Information Systems Management*. 22, 77-85.
10. Dicker KM (2003) The evolution of data mining and related security correlation technology. SANS Institute. Domain site: [http://www.giac.org/practical/GSEC/Keith\\_Dickter\\_GS\\_EC.pdf](http://www.giac.org/practical/GSEC/Keith_Dickter_GS_EC.pdf).
11. Donnely R and Mulcahy M (2008) Board structure, ownership and voluntary disclosure in Ireland. *Corporate Governance*. 16 (5), 416-429.
12. Eng LL and Mak YT (2003) Corporate governance and voluntary disclosure. *J. Accounting .Public Policy*.22, 325-345.
13. Fayyad U, Shapiro GP and Smyth P (1996) From data mining to knowledge discovery in databases. *AI Magazine*. 17 (3), 37-54.
14. Ferris TA, Garrison GM and Lowe HJ (2002) A proposed Key Escrow System for secure patient information disclosure in biomedical research databases. In: Proc. AMIA Symp. pp:245-249.
15. Ge W and McVay S (2005) The disclosure of material weakness in internal control after the Sarbanes-Oxley Act. *Accounting Horizons*. 19 (3),137-158.
16. Ghose A and Rajan U (2006) The economic impact of regulatory information disclosure on information security investments, competition, and social welfare. In: Proc. of the Workshop on Economics of Information security, University of Cambridge.
17. Goodwin LK and Prather JC (2002) Protecting patient privacy in clinical data mining. *J. Health. Clin. Manage*. 16(4), 62-67.
18. Gordon LM, Loeb W, Lucyshyn and Sohail T (2006) The impact of the Sarbanes-Oxley Act on the corporate disclosures of information security activities. *J. Accounting & Public Policy*. 25 (5), 503-530.
19. Han J and Kamber M (2006) *Data mining: Concepts and techniques* (2nd ed.). Amsterdam: Morgan Kaufmann.
20. Islan MZ and Brankovic LA (2004) Framework for privacy preserving classification in data mining. In: Australasian Computer Science Week, School of Electrical Engineering and Computer Science.
21. Kovalerchuk B and Vityaev E (2002) Data mining in finance. In: *Advances in relational and hybrid methods*. Klawer Academic.
22. Lakhil F (2007) Ownership structure and voluntary disclosures: the case of French-listed firms. *Corporate Ownership & Control*. 5 (1), 131-138.
23. Lu, Chi Lin, Chen, Ta Cheng (2009) A study of applying data mining approach to the information disclosure for Taiwan's stock market investors. *Expert Systems with Appl*. 36 (2), 3536-3542.
24. Makhija AK, Patton J M (2004) The impact of firm ownership structure on voluntary disclosure: Empirical evidence from Czech annual reports. *J. Bus.* 77 (3), 457-491.
25. Nemeth P, Arva J. Abonyi FP and Pach F Szeifert (2006) Fuzzy association rule mining for data-driven analysis of dynamical systems. *Hungarian J. Industrial*



*Chem. Special Issue on Recent Advances in Computer Aided Process Engineering.*

26. Paul M, Healy Krishna and Palepu G (2001) Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *J. Accounting & Economics*. 31, 405-440.
27. Pedrycz W and Sosnowski A (2000) Designing decision trees with the use of fuzzy granulation. *IEEE Trans. Syst., Man, Cybern. A*. 30, 151-159.
28. Roiger RJ and Geatz MW (2003) Data mining: A tutorial-based primer. Addison Wesley, Boston.
29. Sweeney L (2003) Navigating computer science research through waves of privacy concerns: Discussions among computer scientists at Carnegie Mellon University. *ACM Computers & Soc.* 34 (1), 1-18.
30. Tsamenyi M, Enninful-Adu E and Onumah J (2007) Disclosure and corporate governance in developing countries: evidence from Ghana. *Managerial Auditing J.* 22 (3), 319-334.
31. Verrecchia A (1983) Discretionary disclosure. *J. Accounting & Economics*. 5, 179-194.
32. Wang YJ (2008) Hybrid rule ordering in classification association rule mining. *Transact. on Machine Learning & Data Mining*. 1 (1), 1-15.
33. Wang X, Chen B, Qian G and Ye F (2000) On the optimization of fuzzy decision trees. *Fuzzy Sets Syst.* 112, 117-125.