J. Appl. Environ. Biol. Sci., 4(5S)55-64, 2015

ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Data Mining, a New Way to Encounter Challenges

Ehsan Zare Saadabadi[1], Maryam Zare Saadabadi[2]

[1]Academic Trainer of Technical and Engineering University of Bahonar of Shiraz.
[2]Academic Trainer of Zand University of Shiraz

Received: January 27, 2015
Accepted: March 31, 2015

ABSTRACT

In the current era famous for information era, without adequate and comprehensive and timely information, decision making for managers is not possible. By increasing development of science and technology and data collection methods, we are encountered with large amount of data. Optimal application of data bases is possible if we use efficient and standard tools beside an organized and futuristic plan to convert data to information. Thus, today, we are not encountered with the lack and shortage of information and the important case is using suitable and standard methods to keep, update, keep and discover new knowledge of existing information mass. Data mining is an efficient method to extract knowledge and information of the existing data. Data mining methods are used widely by experts and planners. This paper attempts to introduce data mining and its duties in various issues.
KEYWORDS: Data mining, Information Retrieval, knowledge Acquisition, challenges

1- INTRODUCTION

Data mining is extract of knowledge among extensive volume of data and is introduced as the most important stage in knowledge discovery process. This tool applies some scientific fields as database technology, statistics, artificial intelligence, machine learning, neural networks, model identification, Knowledge-Based System,

Knowledge-Acquisition, Information Retrieval, High-Performance Computing, Data Visualization and have received much attention as the efficient technique in discovery of useful organizational knowledge by researchers. Data mining techniques as descriptive and predictive tools are used in many fields and also many researches have been conducted in this regard. Some of the applied fields of data mining include commercial, management, medical, sport, econometric, financial management, web mining and text mining (Han & Kamber,2006).

This study attempted to evaluate the application of various data mining techniques in strategic management, banking industry, organizations agility, crimes modeling, CRM and human resources and an applied framework is also presented. The next section reviews the literature of data mining and the relevant concepts of these tools and various techniques can be defined briefly.

2-The review of literature of data mining

Data mining is one of the efficient methods in discovery of useful information among the great volume of information and by extraction of models and relations between the data, their latent values are revealed. By discovery of these values, we can predict the value of other variables and apply them in decision making (Hand, 1998, P.10).

Data mining is a new science. Regarding history of data mining, Lovell (1983) is the first one presenting “data mining simulation”. New researches on data mining started in the early 90s as the interest of researchers to this issue in the late 90s cannot be compared with the same early decades. Data mining is an attractive topic with various branches for researchers in various science fields as statistics, artificial intelligence, industries software and etc. (Saeedi, 2005).

3- Definition of data mining

According to Han & Kamber (2006), data mining is extraction of knowledge of great amount of data and is introduced as the most important stage in knoweldge discovery. The term data mining is equal to data mining, knoweldge extraction, data dredging and knoweldge discovery in database. The main osition of knoweldge extraction and data mining is shown in Figure 1.

* **Corresponding Author:** Ehsan Zare Saadabadi, Academic Trainer of Technical and Engineering University of Bahonar of Shiraz. Email: ehsanzare1359@gmail.com; Mobile: 09172616761

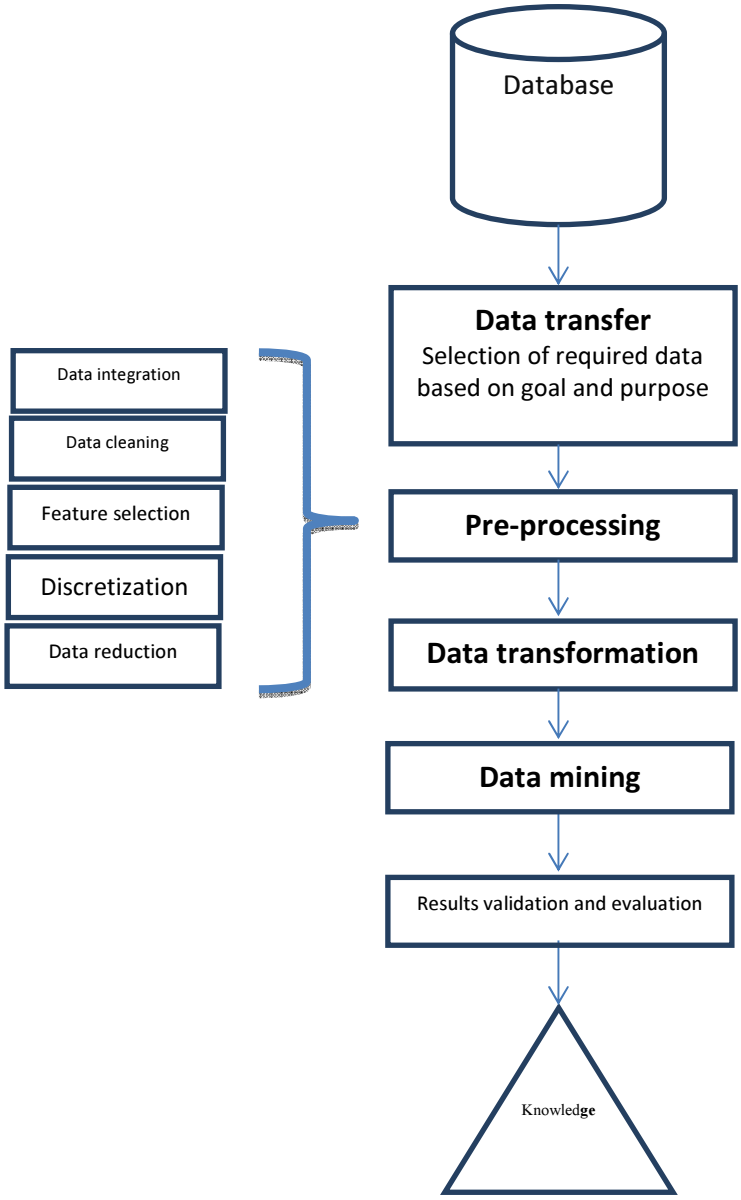

Figure 1- The stages of knowledge discovery process and data mining position (Han & Kamber,2006)

Knowledge discovery in database is accurate identification process and finally discovery of models and perceiving models in data. In other words, data mining refers to the extraction of useful unknown data among large amounts of data. In all cases, data amount is high and discovery of latent knowledge among large amount of data is the final goal of data mining. Simply, data mining is extraction or knowledge mining among large amount of raw data. A general view can lead to the discovery of considerable unexpected structures in this wide set of data and the identification of these relations of models and new trends of data mining can help the decision making process and design of a model for prediction and decision making.

Generally, data mining is searching and discovery process of various models, summarization and obtaining useful values among large amount of data.

4- Functions and duties of data mining

Data mining methods are divided into two general types: Descriptive and predictive. In descriptive methods, the goal is description of an event or a reality. In predictive methods, the goal is prediction of unknown variable of

future data. Figure 2 presents function and duties of data mining (Tan et al,2006). A brief explanation of each of techniques is presented.

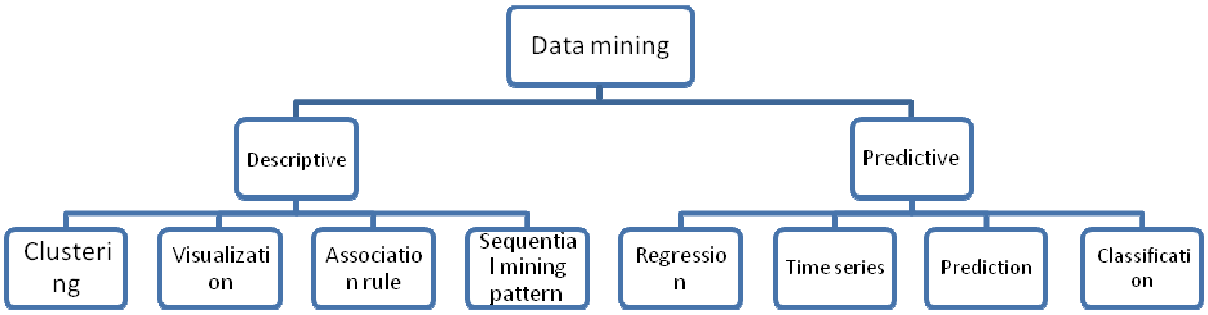

Figure 2- The functions of data mining (Tan et al, 2006)

- **Prediction**: Classification techniques are used to predict discrete features and predictive methods model the continuous functions (Tan et al., 2006),
- **Classification**: The process of finding a model in which by distinguishing the levels or data concepts, we can predict the unknown class of other objects. Some of common classification methods are Decision Tree, Bayesian Method (simple Bayesian and Bayesian networks), neural networks, Support Vector Machine (SVM) and etc.
- **Regression**: It is the best model linking output variables with input variables. The simplest case is linear fitness model, establishing the relation between input and output variables as linear. In Formula (1), y is output and dependent variable, x is input variable and a, b regression coefficients.
  Formula (1): $Y = a + (X_1 \times b_1) + (X_2 \times b_2) + \cdots + (X_n \times b_n)$

- **Time Series:** It is the sorted sequences of observations. Time series is sorted based on time (equal time intervals) and based on other dimensions as interval. Time series as electric pulses are continuous. Each time series describes the value of an object as a function of time in collected data set (Qazanfari et al., 2007).

- **Association Rule:** Extraction of association rules is an undirected data mining searching the relation between the features in the dataset. In other words, dependencies analysis is the study of features with each other. Another name of dependency analysis is Market Basket Data extracting the rules among these features. In other words, these methods attempt to extract the rules to quantify the relation between two or some features.
- **Clustering**: It is one of descriptive methods as dividing the data to similar groups. Indeed, this method analyzes the data without label in the form of some groups. This division is as the data in the cluster have highest similarity to each other and highest difference with other clusters.
- **Visualization**: It includes all methods applied to achieve a compressed description of data and are used in report creation as obtaining mean and standard deviation for the required fields of multi-variate illustration techniques and discovery of function relations between the variables.
- **Sequential Pattern Mining:** Mining sequential patterns is the discovery of the incidents occurred frequently and sequentially. For example, It is highly possible a person purchases color printer andalso buys as digital camera at the same month. Also, this method in marketing, customer retention, weather forecast and many other industries can be applied mostly (Han &Kamber, 2006).

5- Data mining stages

In a brief and general view to data mining process, we can consider its process from the beginning to the end in Chart 3.

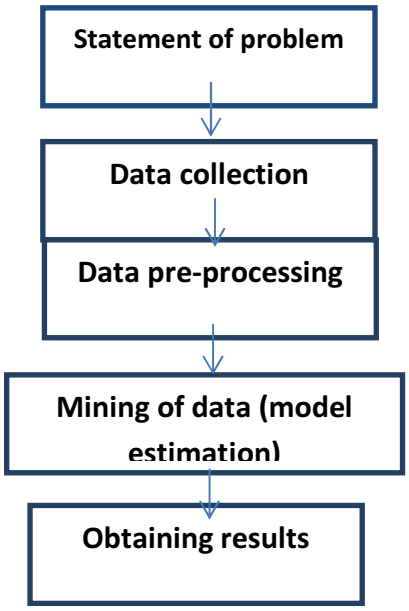

Figure 3- Data mining process (Nejat and Ali Akbari, 2008).

The above stages are defined in a detailed division as:

1- Goal identification: In this stage, the requirements of user are revealed and also the level of information obtained from database is defined.
2- Data selection: In this stage, the data should be selected based on definite criteria.
3- Data preparation: The data forma and identification of extra variables are the goals of this stage.
4- Data evaluation: Some criteria as data distribution, the feature and structure of data in general conditions of data determine the framework of this section.
5- Response framing: Presenting response format as image, chart, neural network and etc. is the output of this section.
6- Tools selection: In this stage, suitable tools for data mining can be selected and its consistency with computer can be evaluated.
7- Modelling: Data mining process starts from this stage. This section includes searching the models in a dataset, classification, evaluation of data and etc. Some cases as validity of model errors, model development and etc. are performed in this stage.
8- Validation of findings: This stage includes testing the models. In data mining analysis, we can discuss about the results of analysis with manager, executive manager or analyst.
9- Presenting the results: In this section, the final report is provided for user and this report should be based on total data mining process.
10- Using the results: The final goal of data mining is using discovered results for decision making the policy making and prediction to create a new and better situation.

6- Data mining structure

Indeed, the basis of data mining is divided into statistics and artificial intelligence and artificial methods are considered as machine learning methods. The basic difference between statistical methods and machine learning methods is based on the assumptions or the nature of the processed data. Normally, it is assumed that in statistical techniques, data distribution is definite and mostly the distribution is normal and finally the true or false final results depend upon the initial assumption. Machine learning methods don't use any assumption regarding data. Statistical methods are used mostly in data mining when the data are not very more and we can achieve much information about data. These methods use a tool to discover the relations between the data and despite neural network methods with ambiguous process, these methods are completely clear. In addition, the accuracy of conclusions and outputs in these methods are better. Generally, statistical methods are useful when the data interpretation is difficult by other methods. Table 4 indicates the difference of statistical methods with artificial intelligence methods.

Table 1- The comparison of statistical methods with other data mining methods (Nejat and Ali Akbari, 2008).

Other data mining methods	Statistical methods
Without initial hypothesis	Having initial hypothesis
Applied in wide range of data	Applied in small range of data
Application in various types of data	Much application for quantitative data
Using learning and artificial intelligence methods	Using math relations

7-Data mining applications

In this section, we deal with the application of data mining in strategic management, banking industry, agility of organizations, crimes modeling, CRM and human resources.

7-1 The role of data mining based on association rules in strategic management

To guide the organization to achieving more successes, the designers of organizations strategy determine organization strategies based on three common stages in strategic management (formulation, implementation and evaluation). The accuracy of data extraction to draw the existing condition and good condition by experts and final users is an important issue leading to strategic planning to achieve the required results. It is possible the beneficiaries give wrong answer to the questions (due to lack of information or different perception of questions) or due to much work and giving less importance to the questions , select the choices randomly or don't select any choice. To evaluate these items, we can apply new solutions and data mining algorithms in this regard. By mining the questionnaires of similar organizations and extraction of rules and latent knowledge, we can achieve a criterion regarding the relation between various sections in strategic planning. As a method of this work, we can extract doubtful responses based on current condition and by investigation of the responses of relevant people, we can achieve results. Taking decision and planning strategy in organizations required great care in extraction of required data and the data features due to its variation has challenged strategic management. To solve these challenges, the behavior of customers, experts, managers and leaders of organization is the best solution. To implement this solution, extraction of knowledge of data is raised. As data mining the association rules is one of the highly applied tools in this regard, it has great features as helping in decision making of managers. Rasulian et al., (2008) in a study "the role of data mining based on association rules in strategic management" besides description of data mining, its application method was stated in strategic planning. They showed that in an organization, how customers' behavior is modeled or by a real example, they showed that extracted strategies were useful for strategic designers of organization. Briefly, they found that extraction of weaknesses and strengths and threats with great amount of data needs another solution besides conventional solutions and data mining can be applied solution to eliminate this problem (Rasulian et al., 2008).

2-7 Data mining in crimes modeling

Data mining in the relevant issues with police can be considered as one of the efficient tools in information analysis and police data. One of the most important and applied functions of data mining is prediction that can be useful in prediction of crimes and prevention of crimes .Ozkan (2004) classified the application of data mining technique in criminology in Precrime and

Postcrime. The former is prediction and prevention of crime and latter is discovery of crime evidences after committing. The survey of published papers in application of data mining techniques shows that this field is taken into attention in recent years. Here, prediction techniques have the highest papers. Ahmadvand and Akhundzade (2009) in a study presented a framework for data mining techniques application in crimes modeling. In this study, the issues of criminology are classified in the form of identification, prediction and prediction of crime and the application of data mining techniques is investigated in three fields. Based on the studies, prediction techniques are used more than other data mining tools to predict crime committing and effective parameters. Among prediction algorithms, regression models dedicate large amount of studies. A combination of clustering and prediction techniques can be observed in some cases. Other techniques of data mining can be useful in this field. They proposed that using association rules to discover the cause of crime and effective factors can be useful. Also, the design of supporting Knowledge-Based System discovered by data mining techniques can be useful in this regard. The presented framework by them is shown in the followings.

Table 2- A review of data mining applications in various police fields (Ahmadvand and Akhundzade, 2010).

Reference	Applied technique	Applied fields	No
Karlis et al., (2007)	Clustering	Crimes identification	1
Adderley et al., (2007)	Clustering		
Murtagh et al., (2009)	Clustering		
Corapcıoglu et al., (2004)	Regression –prediction	Crimes prediction	2
Moon et al., (2010)	Regression –prediction		
Liu et al., (2003)	Prediction		
Corapcıoglu et al., (2004)	Regression –prediction		
Liu et al., (2003)	Criminal Incident Prediction		
D'Alessio et al., (2010)	Regression –prediction		
Deadman (2003)	Time series-prediction		
Freilich et al., (2007)	Regression –prediction		
Xue et al., (2006)	Prediction-clustering		
Hadjidj	Prediction-clustering		
Liu et al., (2010)	Fuzzy swarm	Crimes prediction	3
Oatley et al., (2003)	A combination of regression techniques, neural network and Bayesian network		

7-3 The design of the model of selection of human resources with data mining approach

The success or failure of organization has direct relation with attraction and holing its human resources. There are large amount of data in organizations regarding holding entrance tests and absorbing employees and they are not applied mostly. Data mining is a solution to such issues. Adel Azar et al., (2010) presented a study "designing the model of selection of human resources with data mining approach". The study was applied in terms of purpose, correlation-consistency study in terms of nature. It was attempted to apply data mining techniques to identify the rules between the scores of entrance tests and other personal and job variables (as defined before the entrance of any person to organization) and condition of employees with job performance and their promotion. By the evaluation of databases of test and human resources of a commercial bank for two consecutive years (2004, 2005), the human resources indices effective on performance or promotion were identified. The studied population was the "employed people in entrance test during 2004, 2005 in one of the commercial banks in Iran.

Data mining technique in this study is decision tree and rules extraction is performed by C5.0 ,CHAID , CART and QUEST. Finally, besides presenting a model for selection of effective variables, target variable and suitable algorithms, among the rules, non-axiomatic rules were defined and these rules were determined by experts. One of the results is elimination of performance evaluation variable as target variable in this study as the result of lack of precision of completing the performance evaluation forms in bank evaluation process. Also, this study showed that of total 26 investigated variables, five factors of "total score of test", "interview score", "academic level", "experience of talking" and province of service place" were effective on promoting the volunteers. These results led into the knowledge as they can be applied.

In this regard, some recommendations are presented to the studied organization and the they are the results of analyses of outputs of model. Some of the recommendations include:

- Based on the elimination of test variables (scoring literature, religious book, math and general knowledge), it is necessary to change the type of selected textbooks or change their content and be designed based on the goal of measuring future potentials (not past data).
- Scoring methods of changed performance evaluation and to complete the information, we can use educational tools and cultural norms.
- We can evaluate the methods providing holding a part of test as explained.

Finally, the important point is optimized and wide application of data mining in other management fields and besides low cost application of existing data in organizations, we can clarify new horizon of latent knowledge in organization for managers.

Based on the stages in study and results, the final model of this study is presented in the following chart:

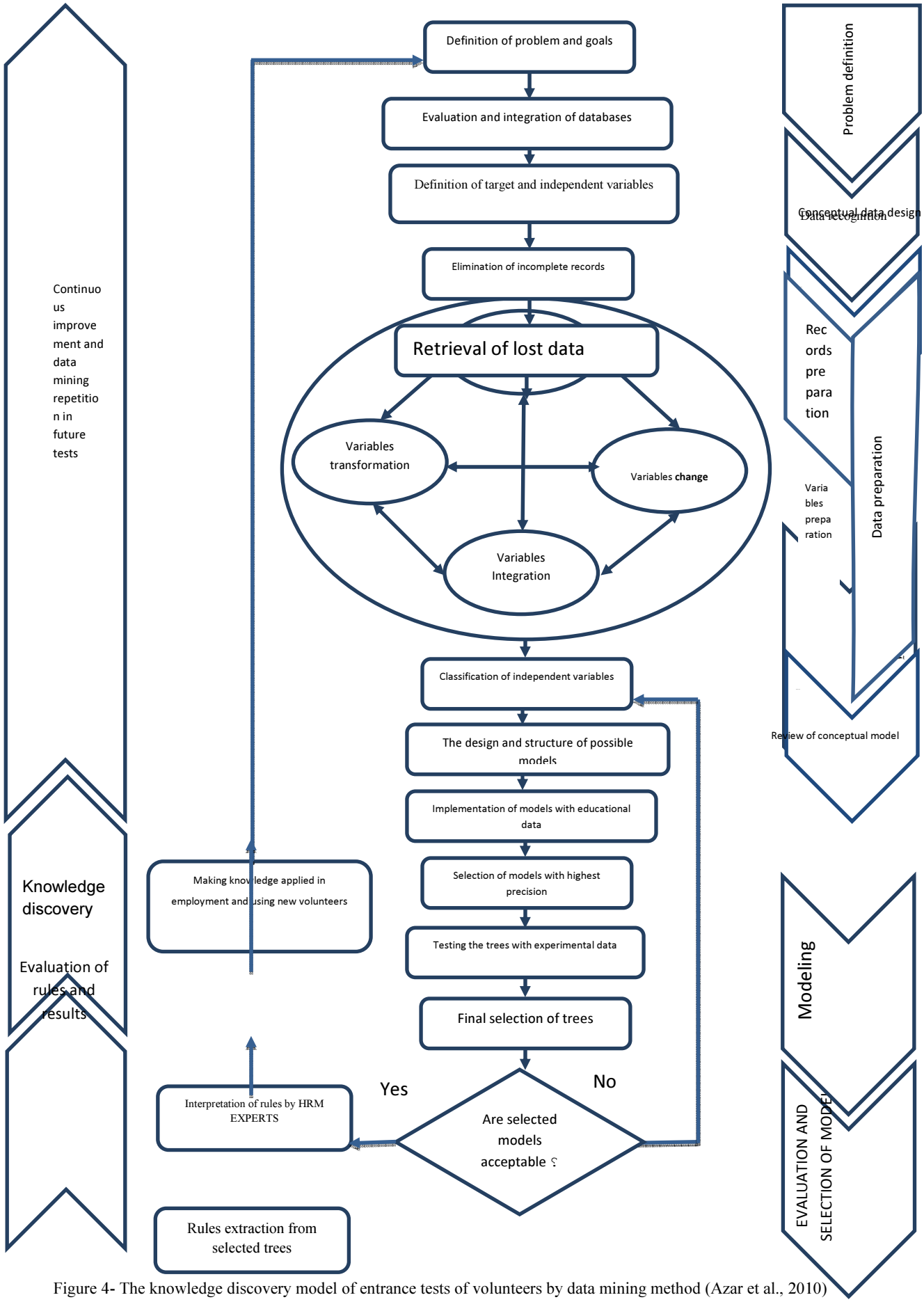

Figure 4- The knowledge discovery model of entrance tests of volunteers by data mining method (Azar et al., 2010)

7-4- Using data mining to implement CRM in banking industry

According to Porter model, banking sector (Chart 6) is encountered with five competitive force to activate in current competitive environment and one of them is customer bargaining power. The increase of competition between the banks (due to high amount of banks), changing the rules and introduction of new technology and namely internet infrastructure increasing awareness of customers increased bargaining power of customer as the customer can refer to another bank immediately. Thus, to absorb new customers and holding old customers, it is possible that the banks take the method based on new technologies as knowledge discovery in data bases, data mining, data warehouse and CRM for the analysis of the needs and customer behaviors. CRM is not used in Iranian banks for customer retention or it is not successful. It seems that knowledge discovery of data basis and data mining is a suitable approach to solve this problem.

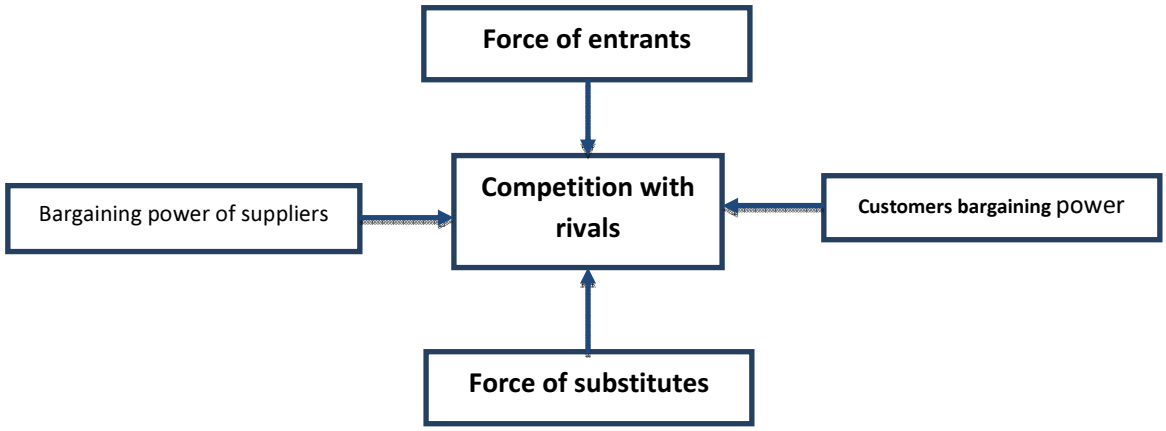

Figure 6: Competitive forces of Porter (Parsayian and Arabi, 2004)

In Iranian banks, customers are not satisfied with banks services and to solve this challenge, we should consider CRM and recognition of customers' features. The most important tool is data mining. In a study done by Amir Albadavi et al., "using data mining for implementation in banking industry of Iran" it was attempted to introduce a framework to use data mining to help banking sector for CRM implementation. By this framework, we can understand how CRM can take benefit of data mining and improve banking and CRM can also be measured. Banking sector can identify its customers better and design their suitable services by this tool and receive valuable price from them. Thus, it can implement dynamic pricing in the bank (Wangler, 1999). This framework considers bank customer as a high level set of features of solution to implement strategies and in this bank, bank features are determined by the customers and the bank itself and it is associated to interaction between bank and customer. The framework on objects and features of objects associated to the interaction between customer and bank than bank and by considering them can be successful in decision making and strategic planning process (Wangler et al., 1999; Locopoulos, P. and Kavakli, 1999)(Philipido et al., 1998; Wangler et al., 1999; Kardasis et al., 1999, Kavakli, 1999). The standard variables can be applied in independent and dependent variables. Independent variables are composed of two types of information needs: Required information of data system of bank regarding specific customers of bank and discovered knowledge of bank in an experience that is a rule in bank but dependent variables as the other half of this framework are composed of customer file and bank file. Customer file variables detect knowledge features regarding customer that the banks can identify them better and measure the profitability and hold them. Bank file needs the part of knowledge of bank to be responsive to the customers of various sectors of target market. This framework makes CRM clear in bank and then implementation can be performed (Abdollah et al., 2002) as summarized in the following six stages:

1- Development of commercial knowledge
2- Development of data warehouse
3- Using commercial modeling technology to define the data mining duties
4- Using technology to create Data Marts
5- The start and execution of data mining
6- Integration of the data mining results with developed knowledge models

The mentioned framework with its dependent and independent variables can facilitate steps 1 to 6. Fourth step is the detailed form of second step. In fifth step, the required data is extracted from data warehouse of bank and finally in sixth step, the analysts and experts of banking industry investigate the results of previous steps form

commercial view and by finding suitable and significant knowledge can use it in commercial activities of bank and creation of CRM in bank.

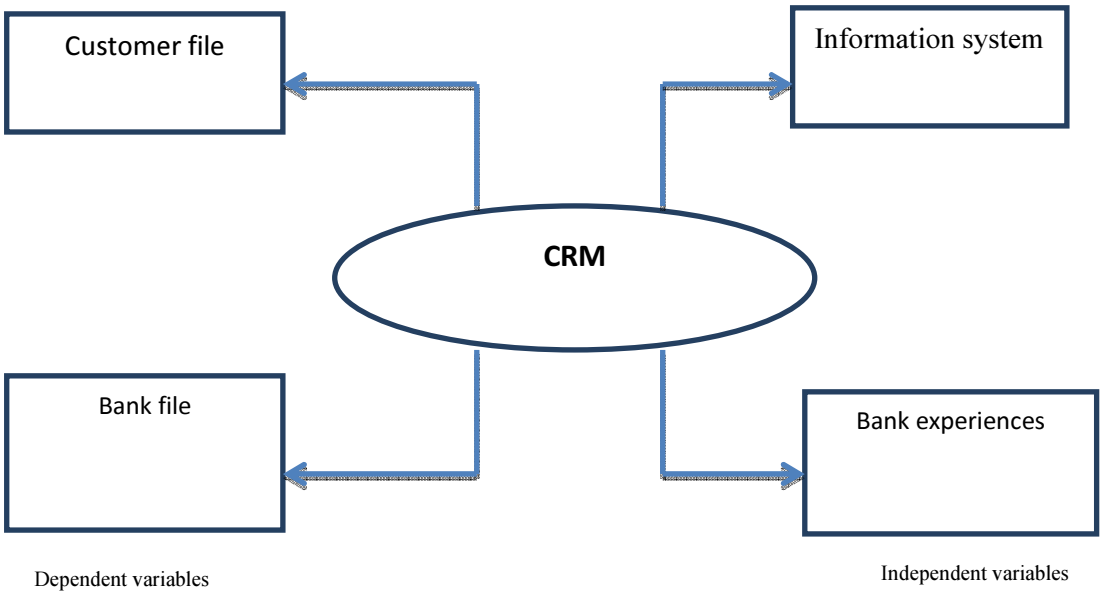

Figure 7: Effective variables on CRM implementation (Albadavi et al., 2005)

8-Conclusion

This study evaluated the applications of data mining in strategic management, crimes modeling, human resources and banking industry. Data mining reduced the relevant challenges. For example, to take decision and strategy planning in organizations, we need the extraction of required data but in terms of variability of data feature, strategic management encountered challenges and this challenge was removed by association rules data mining as one of the highly applied tools in this regard. Regarding the detection of crime and offence detection, many of inspection and law administration experts by data mining can identify the activities of fraudulent and detection of new methods of crime committing. Also, the experts can be helped in detection of behavioral crisis models regarding drugs, trading and money laundry, kidnapper activities, identification of smugglers in border areas. Regarding human resources, correct application of employees selected by entrance tests of institutions and companies, namely in service-oriented organizations has specific importance and using data mining and detection of tacit knowledge can be effective. Regarding CRM in banking industry, to solve this challenge that customers are not satisfied with banks services, using data mining tools can be useful in this regard. Thus, the importance of data mining in implementation of CRM in banking sector can be expressed. Also, required steps are presented to implement the required framework to eliminate the challenges of not implementation of CRM in their banks and take benefits of CRM in serious competitive conditions of banks in current conditions.

9- REFERENCES

9-1 Local references

1- Nejat, Seyed Amir Reza; Ali Akbari, Arash. Data mining, a method to unknown. Two journals of human development of police, Year 5. NO. 18, MOrdad and Shahrivar 2008.
2- Rasulian, Mohsen; Sharayei, Abolghasem, FathiGohardani, Mohammad Bagher. 2008. The role of data mining based on association rules in strategic management. Journal of Basirat. Year 15. NO. 39, Spring 2008.
3- Ahmadvand, Ali Mohammad; ElhamAkhundzade. 2010. Applied framework of data mining techniques in crimes. Two journals of police human resources development. Year 7. No. 30, MOrdad and Shahrivar. 2010.
4- Azar, Adel; ParvizAhmadi, Mohammad Vahid. Bast. 2010. The design of human resources selection model with data mining approach. IT management journal. Period 2, NO. 4, Spring and Summer 2010. 3-22.
5- Albadavi, Amir; Mehdi Fasghanri, Amir HosseinQapanchi. 2005. Using data mining technique to implement CRM in banking industry of Iran. Business investigation. NO. 14, Mehr and Aban 2005.

6- Fredara, David. 2004. Strategic management. Translated by Ali Parasain, Seyed Mohammad Arabi.
7- Saeedi, Ahmad. 2005. Data mining, concept and its application in high education. Journal of high education. NO. 18.

9-2 Foreign references

7- Han, J. &Kamber, M. (2006), Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufman Publisher.
8- Hand. D. J (1998)" Review of Data mining", The American statistician, 52-112-118.-
9 - Tan, P.N., Steinbach, M. & Kumar, V. (2006), Introduction to Data Mining, Boston
10 - Adderley, R., Townsley, M. & Bond, J. (2007), Use of Data Mining Techniques to- Model Crime Scene Investigator Performance, Knowledge-Based Systems, No. 20, PP. 170–176..
11 - Chung, W., Chen, H., Ch., Chang, W. & Chou, SH. (2006), Fighting Cybercrime- A Review and the Taiwan Experience, Decision Support Systems, No. 41, PP. 669.
12 - Corapcıoglu, A. &Erdogan, S. (2004), A Cross-Sectional Study on Expression of Anger and Factors Associated With Criminal Recidivism in Prisoners With Prior Offences, Forensic Science International, No. 140, PP.167–174..
13 - D'Alessio, S.J. &Stolzenberg, L. (2010), Do Cities Influence Co-Offending?- Journal of Criminal Justice, No. 38, PP. 711–719..
14 - Deadman, D. (2003), Forecasting Residential Burglary, International Journal of-
Forecasting, No. 19, PP. 567–578.
15 - Freilich, J.D. &Pridemore, W.A. (2007), Politics, Culture, and Political Crime Covariates of Abortion Clinic Attacks in the United States, Journal of Criminal Justice, No. 35, PP. 323–336..
16 - Hadjidj, R. , Debbabi, M., Lounis, H., Iqbal, F., Szporer, A. &Benredjem, D. (2009) , Towards an Integrated E-Mail Forensic Analysis Framework, digital investigation, No. 5, PP. 1 2 4 – 1 3 7..
17 - Karlis, D. &Meligkotsidou, L. (2007), Finite Mixtures of Multivariate Poisson- Distributions With Application, Journal of Statistical Planning and Inference, No. 137,PP. 1942 – 1960.
18 - Kavakli, V., and Loucopoulos, P., Focus Issue on - Legacy Information System and Business Proces Chan Modelling of Organizational Change Using the EKD Framework, CAIS, Vol. 2(6), 1999, pp.1-30
19 - Kardasis, P., Prekas, N., and Loucopoulos, P., A - Framework of Patterns for Banking Sector, Manchester M601 QD, Department of Computation, UMIST, U.K., 1999, pp.1-5.
20 - Locopoulos, P. and Kavakli, V., Enterprise - Knowledge Management and Conceptual Modeling, Manchester, UMIST, U.K., 1998, pp.1-9.
21 - Wangler, B., Holmin, S., Loucopoulos, P., Kardasis, P., Xini, G., and Filippidou, D., A Customer Profiling Framework for the Banking Sector, 9th European- Japanese Conf. on Information Modelling and Knowledge Base, IOS Press, 1999
22 - Khan, J.I. & Shaikh, S.S. (2008), Computing in Social Networks with Relationship- Algebra, Journal of Network and Computer Applications, No. 31, PP. 862–878.
23 - Li, SH.T., Kuo, SH.CH. & Tsai, F.CH. (2010), An Intelligent Decision-Support -Model Using FSOM and Rule Extraction for Crime Prevention, Expert Systems with Applications, No. 37, PP. 7108–7119
24 - Liu, H. & Brown, Donald E. (2003), Criminal Incident Prediction Using a Point-Pattern-Based Density Model, International Journal of Forecasting, No. 19, PP. 603 – 622.
25 - Moon, B., McCluskey, J.B. &McCluskey, C.P. (2010), General Theory of Crime and Computer Crime: An Empirical Test, Journal of Criminal Justice, No. 38, pp.767-772.
26 – Murtagh, F., Ganz, A. &McKie, S. (2009), The Structure of Narrative: The Case of Film Scripts, Pattern Recognition, No. 42, PP. 302 – 312
27 - Oatley, G.C. &Ewart, B.W. (2003), Crimes Analysis Software: 'Pins in Maps,' Clustering and Bayes Net Prediction, Expert Systems with Applications, No.25, PP. 569–588
28 - Ozkan, K. (2004), Managing Data Mining at Digital Crime Investigation, Forensic Science International, No. 146, PP. S37–S38.
29 –Tan, P.N., Steinbach, M. & Kumar, V. (2006), Introduction to Data Mining, Boston, pearson.
30 - Xue, Y. & Brown , D.E. (2006), Spatial Analysis with Preference Specification of Latent Decision Makers for Criminal Event Prediction, Decision Support Systems, No. 41, PP. 560 – 573.