The Caspian Sea Journal
ISSN: 1578-7899

Volume 10, Issue 1, Supplement 4 (2016) 111-117

Comparing Capabilities of K- Nearest Neighbor Predictive Models, Decision Tree and Bayes in Prediction of Bankruptcy of Tehran Stock Exchange Companies
(Case Study: Pharmaceutical Industry)

Mehdi Nazemi Ardakani,
Assistant Professor of Accounting, Faculty of Economics, Management and Accounting, University of Yazd, Iran
Email: nazemi@yazd.ac.ir
Vahid Zare Mehrjardi,
Master of Accountancy, Faculty of Economics, Management and Accounting, University of Yazd, Iran
Email: vahid69.zare@gmail.com
Amir Hossein Zare,
Master of Accountancy, Faculty of Economics, Management and Accounting, University of Yazd, Iran
Email: zare1026@gmail.com
Milad Kahdoy
Audit MSc Student, Faculty of Management and Accounting, Allameh Tabatabaei University, Iran.
Email: miladkahdooei@yahoo.com

ABSTRACT─ The main aim of the study is that by using the models of k- nearest neighbor, decision trees and Bayes and using the base year of prediction model of Suitable bankruptcy is designed for companies of pharmaceutical industry and then a comparative study of prediction accuracy to do of each designed models for this industry in the base year.
Methods: This study is based on objective, functional, statistical, modeling using models of k- nearest neighbor, decision tree and Bayes and in terms of method, descriptive study (semi-experimental) is considered as correlational type.
findings: The results of k- nearest neighbor, models, decision tree and designed Bayes for the industry, shows that these models are capable of predicting so carefully respectively equal to 95 / 95,96 and 95/95 percent, companies bankruptcy only by use of base year information .
Conclusion: The research results showed that decision tree model has prediction accuracy than the other two models and in result is a more appropriate tool for predicting bankruptcy of the industry companies, although regarding to obtained accuracy for the other two models also must be said that these two models are useful tools for predicting bankruptcy of companies in the industry.
KEYWORDS: bankruptcy, bankruptcy prediction, k- nearest neighbor model, decision tree model, Bayes model.

Introduction
Companies' bankruptcy is one of the most important discussions in financial area which always has focused most number of people and different groups include firms' owners and stakeholders, since the financial event can impose irreversible and unexpected losses to them, and this is unexpected aspect of bankruptcy that makes it more dangerous. So they were always trying to create a shield to protect themselves against such risks. These not only interested in understanding whether a company will be bankrupt or not, but also want to know when company will be bankrupted (1). One of very efficient way to accomplish the notion is to use predictive models to assess the company's financial performance. Because with the Relative awareness from future state of a company financially, can be done necessary actions in order to prevent any losses arising from the bankruptcy of companies. Hence in this research is trying to design and explain predicative suitable bankruptcy by use of these three models for Pharmaceutical industry companies and their ability to predict bankruptcy of the industry companies to be compared.

Research theoretical principals
KNN model
Method of kNN (2) is one of simple learning algorithms Supervised. In spite of simplicity the method has reported ideal and Competitive results which have required to predication compared to other predication methods (3).to more accurate investigate of the method, its two main phases are investigated meant learning phase and prediction phase in the following.
1. Learning phase: in this phase, part of data samples that are known as learning data, is given to the algorithm as input. This phase of algorithm tries to put each learning data in an m-dimensional space. For example, assume that R to be a sample data which has been shown in the form R = {r_1, r_2... r_m}. Here each of r_is is a value of one of the features of data sample.

It is obvious that with composition of these r1s can show that data sample as a single point in next m-dimensional space. It is obvious that each ris is that point image after i-th.

2. Predication phase: in this phase another part of data tilted as test data as input is given to algorithm and the algorithm tries to predict the target variable on it. KNN algorithm takes these steps to do that:

The first step: calculates the distance of test data sample with learning data samples in m-dimensional space. To calculate the distance can be used different criteria of distance, including the Euclidean distance or Manhattan distance. In done implementations of this study, the Euclidean distance has been used at this stage.

The second step: At this step, algorithm chooses of all learning data samples k to those which have the closest distance to the test data sample.

The third step: in this step algorithm of the selected k data sample chooses a value of the target variable (being healthy or being bankrupt) that most observed as target variable for the test data.

The decision tree model

Decision tree is one of the supervised learning algorithms. The model has many applications in prediction and classification regarding to features such as display possibility of prediction as if-then-else simple rules or use it when all the characteristics of the samples are not available. This model allows predicting with regard to features such as for if-then-else or using it when all the characteristics of the samples are not available, there are many applications in classification and prediction. Like any other learning algorithm this algorithm is composed of two phases, learning phase and predication phase.

Learning phase: In this phase, the algorithm tries to create a tree from learning data. An example of the trees has been displayed in figure "1" (4).

Predication Phase (test): In this phase the target variable is predicted based on the characteristics of a new sample data and taking the path of the tree. For example, in the case of a middle-aged individual target variable is projected "yes."

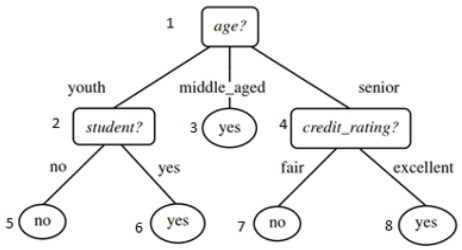

Figure 1: Example of a decision trees

Bayes model

simple classification of Bayesian (NBC8) is one of the classification algorithms that needs to only once the scan of learning data according to simple features such as computing simplicity data and the possibility of using it when all properties the sample are not available (8), it has been become one of the most widely used method of classification and prediction. The method is based on calculating the conditional probabilities based on Bayesian rule despite the assumption of independence between the characteristics of the samples. However the assumption is a Simplifier assumption and practically may not be provided but experimental results show that this classification can act better in many issues than other classifier algorithms (9). This algorithm also like other classifier algorithms consists of two phases, learning phase and predication phase. In the Following first explain predication phase and then learning phase. Predication Phase (test): assume that a data sample X (in our case a company) to has n features include on $x_1, x_2, \dots, x_n$. (In our case n is equal to 13, for example, x_1 is the ratio of current debt to equity). In addition, assuming that the values of these n feature is for the sample data are values $A_1, A_2, \dots, A_n$. in general state C target variable also includes value m $C_1, C_2, \dots, C_m$ (in our case m is the number 2 and c also get the value of zero to itself for healthy companies and the value one to bankrupted companies). NBC puts X sample in C_i groups if and only to be provided following conditions:

$P(C_i|X) > P(C_j|X)$ for $1 \le j \le m, j \ne i$

But Bayesian rule is used according to the following formula to calculate $P(C_i|X)$

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \qquad \text{Equation (2)}$$

It is obvious that here denominator is fixed for all C_i of ($1 \le i \le m$) and for maximization of $P(C_i|X)$ numerator of the $P(X|C_i)\,P(C_i)$ also should be maximized .calculation of this expression is the responsibility of learning phase which we investigate it in the following.

Learning phase: Learning phase: In this phase, information of number of samples that the target variable is known in advance about them and call them learning instances, are selected as input of this phase. Learning phase Outputs must to calculate two sections $P(C_i)$ and $P(X|C_i)$ for predicted phase. Calculation of $P(C_i)$ that is clear, it is enough to divide total number of learning data which the target variable about them is C_i on whole learning data. But it is certain that calculation of $P(X|C_i)$ with respect to the X consisting of n feature and x not to been among learning data, is a time taking and sometimes impossible work. Therefore,

by applying the simplifier assumption of independence between all x_i can be calculated $P(X|C_i)$ according to the following equation:

$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times ... \times P(x_n|C_i)$ Equation (2)

But $P(x_1|C_i)$ and $P(x_2|C_i)$ and ... $P(x_n|C_i)$ calculation from learning data will be easy work. It is obvious that here $P(x_k|C_i)$ is calculated for the value $x_k = A_k$ which A_k is the observed value of x_k feature about test data of X. So for calculation of $P(x_k|C_i)$ will is taken one of the two following procedure:

1. If x_k feature is Discrete, the total number of learning samples which were $x_k = A_k$ and target variable of C_i must be divided to the total number of learning samples which target variable was C_i about them.

2. If x_k feature is continuous process will has different. Usually continuous data are modeled with Gaussian distribution with mean μ and standard deviation σ with this equation:

$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ Equation (3)

So we have:

$P(x_k|C_i) = g(x_k, \mu_{Ci}, \sigma_{Ci})$ Equation (4)

To calculate the above expression is sufficient calculate that the values μ_{Ci} and σ_{Ci} for which training data that target variable about them is C_i and also x_k feature about those that have A_k value.

Research background

Esmailzadeh Maghari and Shaker (10) tried to predict financial distress of accepted companies which were active In Tehran Stock Exchange since 2010 to 2012 using two different predication models, simple Bayesian network model of artificial intelligence and expert systems and DEA pattern of Operational Research techniques. The results showed that it can be used from simple Bayesian network model (which reaches to 91 percent accuracy at best state) with more confidence than the data envelopment analysis method. Husseini and Rashidi (12) used decision tree and logistic regression in the study to predict the bankruptcy of accepted companies on the Tehran Stock Exchange. The research results showed that both of models have higher accuracy in predicting companies' bankruptcy but totally prediction accuracy of logistic regression model is more than CART decision tree model and finally cleared that the logistic regression model is the perfect tool compared to CART decision tree model to predict companies' bankruptcy and acts more conservative . Saeedi and Aghaie (13) tried to predict financial distress of accepted companies on Tehran Stock Exchange using Bayesian networks. Research results showed that simple Bayesian network model which is based upon conditional correlation, is able to predict companies' bankruptcy with an accuracy of up to 90% percent. As simple Bayesian network model based on the conditional probability is able to predict companies' bankruptcy with 93% accuracy. The results also showed that the logistic regression model also has 90% accuracy in predicting the company's bankruptcy. Gerdo and Habibi Kand Bon (14) evaluated financial management capabilities of pharmaceutical companies based on the Springett model. The obtained results of its analysis showed that the overall accuracy of regression models to assess the companies' financial management capabilities with Springett and non-Springett variables in four year period respectively is 85.6 and 79.8 percent and in two year period is 98.1 and 96.2 percent respectively . Also they tested Altman's Z Function in order to investigate accuracy of research pattern using the information of the year 2006 that predication accuracy obtained 96.2 and 92.3 percent. Nagaraj & Sridhar (15) comparatively studied the capabilities of predictive logistic regression models, Claire Forrest 10, Bayes and neural networks in predicting the bankruptcy of companies. The research results showed that all four of the models have high prediction accuracy over 90% and between them; the neural network has the highest prediction accuracy. Finally determined that all four above prediction models are very good patterns to predict bankruptcy. Y. Kim & Upenja (16) in a study sought to predict restaurants financial distress by use of decision tree models and Adaboosted 12 decision tree. The research results showed that full-service decision tree model has prediction accuracy of 96.99% and limited service decision tree model has prediction accuracy of 96.73%. Similarly, results showed full-service decision tree model, limited service Adaboosted decision tree model and general model of Adaboosted decision tree , respectively have predict accuracy of 98.1, 93.08and 97.69 percent. And finally determined that Adaboosted decision tree model has prediction accuracy more than the decision tree model.

Chen (17) tries to predict companies' financial distress, using decision tree and logistic regression. The research results showed that all three Decision Tree algorithms with logistic regression model are an appropriate tool to predict companies' financial distress. Chen et al (18) tried to use a new method to predict companies' bankruptcy by applying fuzzy KNN model in a study. The research results showed that all presented models in this study have a high potential in prediction of companies' bankruptcy and fuzzy KNN model has prediction accuracy than other models. Sarkar & Srirman (19) sought to predict banks bankruptcy in America by use of two models of Bayes simple networks and Bayes complex networks. The results showed that Bayes simple model is capable of predict companies' bankruptcy with accuracy equivalent to 80 percent and being healthy with 93 percent accuracy. Thus, it was revealed that Bayes complex model is capable of predict bankruptcy with accuracy of 88 percent and being healthy with accuracy of 93 percent. So it was revealed that Bayes complex model totally

has better performance than Bayes simple model in predicting banks performance. Finally it was determined that each model is useful tools for predicting bank bankruptcy.

Research Methodology

Research Hypotheses

1. KNN Model is a proper method to predict companies' bankruptcy of pharmaceutical industry.
2. The decision tree model is a proper method to predict companies' bankruptcy of pharmaceutical industry.
3. Bayes model is a proper method to predict companies' bankruptcy of pharmaceutical industry.
4. The decision tree model has higher ability than KNN to predict companies' bankruptcy of pharmaceutical industry.
5. Bayes model has higher ability than the KNN, to predict companies' bankruptcy of pharmaceutical industry.
6. Decision tree model has higher ability to predict companies' bankruptcy of pharmaceutical industry.

Research Variables

In this study, variables selected based on best relations in previous research results which has been shown in table (1).

Table 1: Research variables

X_1	**Ratio of Current debt to equity**	X_5	**Return on capital ROE**	X_9	**Inventory turnover**	X_{13}	**Ratio of debt to equity**
X_2	**Ratio of net income to sales**	X_6	**Current ratio**	X_{10}	**Receivables collection period**		
X_3	**Ratio of gross profit to sales**	X_7	**Quick ratio**	X_{11}	**Fixed asset turnover**		
X_4	**Return on assets ROA**	X_8	**Ratio of current assets**	X_{12}	**Total circulating of assets**		

Statistical population and research sample

Statistical Society of the research consists of all companies of pharmaceutical industry. Systematic elimination method was used to select sample. So, for the design of prediction model of bankruptcy of industry Khasayn using three models were used from information of all companies of this industry as an example (regarding to the following limitations).

1. Since 2001 have been accepted in Tehran Stock Exchange
2. The end of their fiscal year is end of March of each year.
3. During reminded financial years do not have activity change or fiscal year change.
4. All required information is available to calculate research variables in studied year.

Data collection method

All required data and information of the research extracted through Novin rahavard software. Applied bankruptcy Criterion of the research is debt ratio. So that the companies for which this ratio is greater than one are considered as bankrupt and companies that their ratio is less than one healthy.

Research design and statistical methods used

This research method is a descriptive research (semi experimental) of correlation type in term of method based on objective, functional and statistically is considered as decision tree and Bayes using KNN models. In this research has been tried to design proper bankruptcy prediction models for companies of this industry by use of above models and a comparative investigation to be done toward their performance in terms of prediction accuracy. In general, different criteria can be used in prediction phase to investigate the efficiency of the KNN, the decision trees and Bayes models. In this research the criterion "Wrong Estimation Percent16" was used which we call wep. To calculate this percent, we divided companies' available data for the industry into two training data and test data (The data are based on company-years, means each company is considered in each year). Training data consist 75% of data means 191 data which contains 16 bankrupt companies and 175 healthy companies. Test data also constitute 25 percent of data means 64 data from these 6 companies are bankrupt and the rest are healthy. After doing learning phase on experimental data and doing prediction step on test data, we measured incorrect prediction percent to the whole test data. This number is that web which this criterion is calculated of the two criterions wep0 and wep1 with the following definitions:

Wep0: is percent of test data which is occurred prediction wrongly and this prediction was that in fact the company was healthy but was predicted bankrupt. Wep1: is percent of test data which is occurred prediction wrongly and this prediction was that in fact the company was bankrupt but was predicted healthy. In this research, in order to investigate prediction wrong percent of KNN models and decision tree an experiment was designed as follows. In this experiment prediction wrong percent of model, with value change of n-neighbor and msl respectively was measured for KNN and decision tree models. In addition, in this

experiment the best amount of n-neighbor and msl for these two models were obtained respectively. Mean of the best value was amount of n-neighbor and msl respectively these two models had the least wep for this industry. About Bayes model must be said that the model has fixed wrong percent unlike two KNN and decision tree models and it is not as the two models which their prediction percent change respectively with change of n-neighbor and msl. To implement the research models soft wares are available such as MATLAB, Rapid miner and libraries in programming language, such as Java or Python or C. To implements of the research model was used library scikit-learn (20) in Python language. The library since has been decompiled, utilizes of its beneath layers, provides proper speed in performance of algorithms. In this research since the number of algorithms repetition was high, scikit-learn library was chosen as a suitable option.

Research findings

The findings of the KNN model

Obtained results of the experiment related to prediction accuracy with n_neighbor model change has been come for the industry in figure "2". As can be seen in this experiment n_neighbor has been changed from 1 to 30 with step 1. For each of the n_neighbor values, the experiment was done 1000 times and was measured average of wep.

As it is predicted about this industry from amount of n_neighbor to the next wep is fixed and practically n_neighbor effect disappears in model. In addition, it is seen that with change of n_neighbor, prediction accuracy of model changes about 9.91 percent to 95.95 percent that in the best case means it reaches from n_neighbor = 6 to n_neighbor = 15 the highest prediction accuracy means 95.95 turns. The average of prediction model is 94.05 percent for the industry.

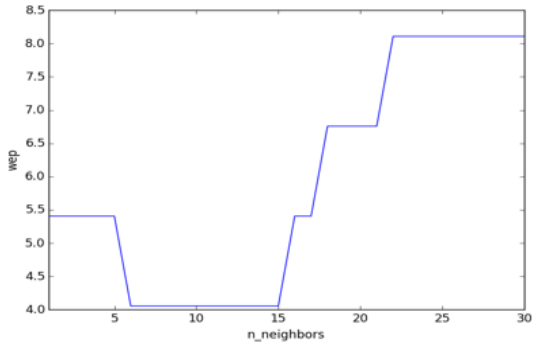

Figure 2: Prediction accuracy of model according to amount of n- neighbor

Here's mentioning a case which seems necessary is that for values of n_neighbor = 6 to next, wep0 value reaches zero. It means that the model has no error in bankruptcy prediction of healthy companies and model error value is due to wep1. So it can be said that by selecting an appropriate range of values of n_neighbor can be reached prediction accuracy of healthy companies to 100 percent on the proposed model.

Obtained results of the decision tree model

The results of experiment for prediction accuracy of decision tree model with amount change of msl for this industry have come in Table 2.As can be considered in this experiment msl has been changed from 1 to 40 with step 1. For each of the msl values, the experiment was done 1000 times and was measured the average of wep. As it is also predicted in the case of this model, wep is fixed from amount of msl to next and practically msl effect disappears in the model. In addition, it is seen that with change of msl, prediction accuracy of the model changes from 90.54% to 96% which in the best means will change that in the best case, means msl =2 reaches to 96 percent. The correct prediction percentage average of in all cases equals to 94.4 percent.

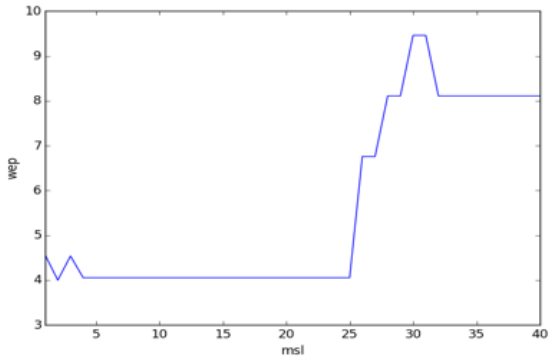

Figure 3: prediction accuracy of model based on the amount of msl

Here's mentioning a case which seems necessary is that for values of msl=4 to msl=13 and also values of msl =4 to up the wep0 value reaches zero. It means that the model has no error in bankruptcy prediction of healthy companies and model error value is due to wep1. So it can be said that by selecting an appropriate range of values of msl values can be reached prediction accuracy of healthy companies to 100 percent on the proposed model.

Obtained results of Bayes model

designed Bayes Model for the industry in terms of prediction accuracy, unlike two decision tree and KNN models which respectively each of them were influenced by amount of msl and n_neighbor, it is not influenced by any parameter, and by its performance obtained an overall accuracy of that. Accordingly, the obtained prediction accuracy of performance or the model for this industry was obtained 95.95. the point which is important here is that wep0 is zero for this model ,means the model has no error in prediction of healthy companies and all errors of model prediction is wep1.

Compare obtained results of these models performance

Obtained results of three designed models performance to predict the companies' bankruptcy of the industry has been come in Table 2.

Table 2: Obtained results of models performance

	Bayes model	**Decision tree model**	**KNN model**
Wep0	0	0.33	0
wep1	4/05	3.67	4.05
Wep	4/05	4	4.05

As Table 2 shows all three KNN, Bayes and decision tree models, have prediction accuracy over than 95%. Therefore, based on these results can be said that all three models have a high ability in predicting companies' bankruptcy in this industry. Also it is observed that the decision tree model has prediction accuracy compared to KNN and Bayes models. So can be said that it has more ability to predict pharmaceutical companies' bankruptcy compared to the other two models. Two Bayes and KNN models however have equal prediction accuracy but because the average of prediction error of KNN model is less compared to, so we can be said that Bayes model has more ability to predict pharmaceutical companies' bankruptcy than the KNN model.

Conclusion and discussion

In this study was tried to design prediction model of proper bankruptcy pharmaceutical Industry specific by use of predictive KNN, decision tree and Bayes models and then performance of each one of them to be compared in terms of prediction accuracy. For this purpose, according to previous results, 13 financial ratios were selected as predictor variables. Then the research sample was selected during the period of 2001to 2013 and was determined by use of the debt ratio, being bankrupt or healthy of selected companies as research sample. As the results show the prediction accuracy of designed KNN, Decision Tree and Bayes models for this industry are respectively 95.95, 96 and 95.95 percent which is due to high accuracy of the models. Also high accuracy of all three models in predicting of this industry companies showed that all three models are proper method to predict companies of the industry, especially the decision tree model that had the highest prediction accuracy.

Suggestions

It is recommended that other researchers who seek to design prediction models of companies' bankruptcy for other industries that there is possibility of designing and derivation of bankruptcy prediction model for them, to design and compare proper bankruptcy prediction models by use of these techniques as well as other techniques and patterns. Also researchers can investigate the effect of corporate governance variables, conservatism and firm size on the prediction accuracy of bankruptcy prediction models in order to do future researches in this field.

References

1-R.sandin, A., and M. Porporato (2007). "Corporate Bankruptcy Prediction Models applied to emerging economies". International Journal of Commerce and Management, Vol 17, Issue 4, Pp. 295-311.
2-Simard, P., LeCun, Y., and J. S. Denker (1993). "Efficient pattern recognition using a new transformation distance" .In Advances in neural information processing systems, Pp. 50-58.
3-Cover, T. M., and P. E. Hart (1967). "Nearest neighbor pattern classification" .Information Theory, IEEE Transactions on, Vol 13, Issue1, Pp. 21-27.
4-Jiawei, H., Kamber, M., and J. Pei (2011) .Data mining: concepts and techniques, Elsevier.
5-Quinlan, J. R (1986). "Induction of decision trees" .Machine learning, Vol 1, Issue 1, Pp. 81-106.
6-Quinlan, J. R (1993). C4.5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann.
7-Crawford, S. L (1989). "Extensions to the CART algorithm" .International journal of man-machine studies, Vol 31, Issue 2, Pp. 197-217.
8-Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., andR. Strachan (2014). "Hybrid decision tree and naive Bayes classifiers for multi-class classification tasks" .Expert Systems with Applications, Vol 41, Issue 4, Pp. 1937-1946.
9-Rish, I. (2001). "An empirical study of the naive Bayes classifier.In IJCAI 2001". workshop on empirical methods in artificial intelligence, Vol. 3, No. 22, Pp. 41-46.
10-Esmailzade moghri, A., and H. Shakery (2015). "Predicting Financial Distress of the listed Companies in Tehran Stock Exchange using Simple Bayesian Network and Compairing it with Data Covering Analysis", Journal of Financial Engineering and Stock Exchange Management, Vol 5, Issue 22. [In Persian]
11-Mousavi Shiri, M., Bafandeh Imandoust, S., and M. Bolandraftar Pasikhani (2013). "Application of K-Nearest Neighbor (kNN) for Predicting Corporate Financial Distress in Tehran Stock Exchange" .Journal ofMonetary EconomicsFinance, Vol 20, No. 6. [In Persian]
12-Hosseini, S. M., and Z. Rashidi (2013). "Bankruptcy Prediction of Companies listed Corporations in Tehran Stock Exchange by Using Decision Tree and Logistic Regression", Journal of Financial Accounting Researches, Vol 5, Issue3, Pp. 105-128. [In Persian]
13-Saeedi, A., and A. Aghaei (2009). "Predicting Financial Distress of firms Listed in Tehran Stock Exchange Using Bayesian networks". Journal of the Accounting and Auditing Review, Vol 16, Issue 3, Pp. 59-78. [In Persian]
14-Gerd, A., and V. Habibi Kandaben (2009). "Investigating the financial ability of Pharmaceutical companies using Springett Model", Journal of Economic Research and Policies, Vol 5, Issue 52, Pp. 115-134. [In Persian]
15-Nagaraj, K., A. Sridhar (2015). "A Predictive System for detection of Bankruptcy using Machine Learning techniques". International Journal of Data Mining & Knowledge Management Process (IJDKP) .Vol 5, No 1, Pp 29-40.
16-Kim, S. Y., and A. Upneja (2014). "Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models" .Economic Modelling, Vol 36, Pp. 354-362.
17-Chen, M (2011). "Predicting corporate financial distress based on integration of decision tree classification and logistic regression" .Journal of Expert Systems withApplications, Vol 9, Issue 38, Pp. 11261-11272.
18-Chen, H. L., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S., andD. Lio (2011). "A novel bankruptcy prediction model based on an adaptive fuzzy K-nearest neighbor method", Knowledge-Based Systems, Vol 24, Issue 8, Pp. 1348-1359.
19-Sarkar, S., andR. Sriram (2001). "Bayesian Models for Early Warning of Bank Failures" .Management Science, Pp. 1457-1475.
20-Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and E. Duchesnay (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, Vol 12, Pp. 2825-2830.
Persian sources:
21-Esmailzadeh Moghri, Ali & H.Shakeri, (2015). "Predicting financial distress Drbvrs Exchange listed companies using simple Bayesian network and compare it Bathlyl DEA." Financial engineering and management portfolios, Volume 5, Issue 22.
22-Mousavi, Shiri, Mahmoud,Bafandeh, Imandust, Sadegh and Mohammad boland Raftar Psikhany (2013). "K- nearest neighbor method in predicting financial distress companies listed on Tehran Stock Exchange" .aqtsad monetary, fiscal, the twentieth year, No. 6.
23-Hosseini, Mohsen and Rashidi Z. (2013). "Bankruptcy prediction listed in the Tehran Stock Exchange using logistic regression and decision trees" .pzhvhsh financial accounting, Volume 5, Number 3, Pages 105-130 .
24-Saeedi, Ali and ArezuAghei. (2009). "Financial distress prediction of listed companies in Tehran Stock Exchange using Bayesian networks" .brrsy accounting and auditing, Volume 16, Issue 56, 2009, pp. 59-78.
25-Gerd, Aziz and Vahid Habibikandaban (2009). "Pharmaceutical companies based on financial capability evaluation Springett." Research and economic policy, in Issue 52, pp. 115-134.