International SAMANM Journal of Finance and Accounting
ISSN 2308-2356 January 2014, Vol. 2, No. 1

Use of Clustering Approach For Portfolio Management

Saeid Fallahpour [1], Mohammad Hendijani Zadeh & Eisa Norouzian Lakvan

Abstract:

In this article, clustering approach was implemented to classify 79 selected stocks of Iran stock market into a number of clusters. Collected data were referred to the initiation of currency crisis in Iran's economy that influenced Iran Stock Exchange dramatically (22/09/2012- 22/03/2013). Applied clustering techniques were K-medoids ,K-means and X-means and valuation ratios and returns of stocks in determined time periods had been considered investment variables. Mentioned clustering techniques were evaluated by the application of Intraclass inertia, which illustrates the compactness of each clustering method. With comparison of Intraclass inertia, it was deducted that K-Means algorithm has a better quality (most compact clusters) in comparison with K-medoids and X-means techniques. The efficient number of clusters as the initial output in mentioned algorithm was achieved by the use some defined indexes namely called Silhouette andDavis-Bouldin. According to brought results, five stocks from the most desirable cluster resulted by best applied algorithm(K-means), were chosen to create an efficient portfolio regarding to Markowitz model, which meets the dimension of portfolio risk minimization by portfolio diversification, and it's weekly returns for chosen 6-month period were compared with a certain benchmark (TSE50). The main contribution of this approach is the reduction of possibilities,when we (investors) are supposed to make efficient portfolios owing to similar stocks are classified in the identical clusters.

Keyword: *Clustering techniques, K-Medoids, K-means, X-means, portfolio management, Markowitz model*

1. Introduction

In most of developed countries, stock market is regarded as one of the significant parameters for enlargement of capital market and a great number of companies are funded through it.

[1] Saeid Fallahpour is the assistant professor of Finance Department,Faculty of Management, University of Tehran (FALAHPOR@ut.ac.ir)

According to belief of major proportion of economists, recession in this market is interpreted as economic recession and when it is booming, it is considered that economy is in a good circumstance as a whole system. In the realm of finance, portfolio management has been put forward as a great concern for financial institutions in terms of helping investors to create an optimized portfolio regarding to returns and risks of portfolio's assets. Markowitz (1952) presented a model for creation of an efficient portfolio,based on returns of assets, as well as thestandard deviation of asset returns as the symbol of portfolio risk. Therefore, returns of assets and standard deviation of these returns in a determined timeperiod, constitute the essence of Markowitz model. Furthermore, quantification of the defined risk enables the investors to take some measures for risk reduction by diversification of their investment in various stocks.

In this paper, Iran Stock market has been perused to assist investors for having an efficient portfolio. Even though, the scale of Iranian market has augmented dramatically in 2000s, Iran's economy is afflicted with numerous financial issues. On one hand international sanctions imposed by western countries, resulted in dropping down of RIAL(currency of Iran) compared with foreign currencies severely. On the other hand, high rate of inflation is an influential factor in Iran's economy. Because of stock market is a chief constituent of economy, stocks of Iranian companies were affected remarkably. Selected and determined time period for this research is simultaneous to commencement of currency crisis in Iran. The main intention of Data mining procedure is to elicit valuable information from great quantities of data sets and it tries to result in the data modeling creation. Generally, there are two kinds of data modeling called predictive and descriptive modeling (Hand,2001). Predictive modeling is usually applied to anticipate the target value based on the experiences of known target values, while descriptive modeling is usually associated with the unsupervised learning functions. Instead of predicting the target value, descriptive modeling discovers patterns in existing data, concentrates on the structure and relation of the data to create meaningful classes and clusters. Some representatives of descriptive modeling are: density estimation, data segmentation, and clustering (Hand, 2001). Density estimation aims to find out the density of population around each individual. It is a more complex case of clustering. Whereas, segmentation intends to set apart data into homogenous clusters and quantity of the clusters are defined beforehand.

Clustering is considereda remarkableway to extract some valuable information from a data set. According to the Handbook of computational statistic (Gentle, 2004), clustering is the most frequently used technique. Without having predefined classes, a clustering algorithm groups

similar objects into the same clusters or subgroups. Clustering techniques aim to maximize the intra-class similarity and minimize the inter-class similarity. In addition, they are good for a quick overview of data or when there are many groups in the data (Romesburg, 2004). Han and Kamber (2006) categorized techniques of clustering developed for dealing with miscellaneousstatic data into five segments: model based methods (Cheeseman,1996) ,Segmentation (Partitioning) techniques, hierarchical algorithms, density-based(Esterm,1996) andgrid-Based(Wang,1997) approaches.Applied data mining approach (clustering) is implemented to tackle the issue of well diversified stocks selection.A data set is organized to clusters and points of data possessed bycertain cluster seem to beanalogous. Contradictorily, the other points pertain to various clusters are unlike (Chui-Yu Chiu, 2009). Three well known clustering techniques namely called K-means, K-Medoids, X-means are exerted to cluster 79 stocks from different segments of Iran's industries. The stock data include valuation ratios and some attributes, which represent the timely returns of them. These chosen stocks came from different parts of Iran's economy and they could portray the status of Iran's economy transparently. By the use of some defined indexes, optimal number of clusters is elicited. Performance comparison of three applied techniques is carried out through the Intraclass Inertia (Michaud,1997) explained thoroughly in this article. In order to establish an efficient portfolio concerned to Markowitz model, clustering methods are applied to opt stocks with some characteristics in this article. This paper embodies fournotable sections; first of all, there are some reviews about former applications of clustering methods and portfolio management. Next segment of this article is devoted to some clarifications about utilized methods namely called K-means,K-Medoids,X-means. Third section of this paper contains problem description and implementation of three mentioned clustering techniques. Conclusion is made in the 4th section of this paper.

2. Literature Review:

2.1. Portfolio Management

Markowitz (1952) developed a model of creating an optimized portfolio. In his model, the stock's return is represented by the average return and stock's risk is quantified by standard deviation of returns of stocks in a specific time period. We compute the portfolio's return by sum of weighted returns of it's stocks. In this model, there are some efficient frontiers, which suggest highest return for each level risk (for a given level of return, lowest risk is shown).Diversification of investment results in risk reduction and it is one of the most popular

economic measures among investors. A lot of works have been conducted on portfolio management henceforth. Fernandez (2005) developed a stochastic control model that contains ecological and economic uncertainty for jointly management in all types of natural resources. In dynamic portfolio management area, Ostermark (1996) stated fuzzy models. Evolutionary algorithms such as Genetic Algorithm (GA) have been implemented abundantly for instance: Oh, Kim and Min (2005) which was done for index fund management.

2.2. Clustering Techniques

Data clustering can provide a suitable way toward an ideal solution or even many lead directly to it (Paranjape, Voditel; 2013). Over the last several decades,owing to the development of artificial intelligence and soft computing,clustering techniques based on other theories or methods have occurred (Backer,1995; Fayyad & Piatesky & Smith,1996) and increasing importance of data clustering techniques in its widespread applications, has led to the development of a variety of algorithms with different quality/complexity tradeoffs(Jain,M.Murty,1999)(A. Qin,2005).A good clustering method will produce high-quality clusters with high intraclass similarity and low interclass similarity. The quality of a clustering algorithms depends on both: similarity measure used by the method, and it's implementation (H.W. Shin,2004).Most of clustering algorithms are based on two popular techniques known as hierarchical and partition clustering (J.Han,2006).These methods can be further extended to supervised and unsupervised clustering algorithms. The main difference between supervised and unsupervised approach is the necessity of specifying the number of clusters in supervised clustering algorithms.In this article, three well known methods are stated for clustering selected stocks and making an efficient portfolio according to Markowitz model and diversification concept. A great number of works have been carried out in the domain of clustering techniques like finance, mathematic and biology. Lots of works illustrate comparisons of different clustering techniques (Mingoti&Lima,2006; Mastopoulos,Nikita & Marsh,1999).

3. Methodology

In this section, some explanations about applied techniques as well as defined procedure of mentioned approach are made.

3.1. Markowitz Model

Markowitz model enabled investors to take some measures aiming to risk reduction by the use of investment diversification. Following model represents his approach:

$$\text{Min } \sigma_p^2 = W^T \text{S W} \quad (1)$$

$$\text{Subject to } \quad W^T \text{I} = 1 \quad (2)$$

$$W^T R = R_E \quad (3)$$

Where σ_p^2 is theportfolio risk and equals to:

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \quad (4)$$

σ_{ij}demonstrates covariance returns of stocks*i andj*. *n* denotes the number of stocks and *w* depicts weight vector and it has a quantity between o and 1 inclusive. *S* represents the covariance matrix of stock and R_E and *R*are defined as expected return and mean return of each stock respectively.R_t is calculated according to following formula;

$$R_t = \log \left({}^{S_t}/_{S_{t-1}} \right) \quad (5)$$

S_t illustrates the price of stock at defined time period(*t*) .

3.2. K-Means Algorithm

K-means technique is broadly used in terms of fast processing ability of largeamount of data and considered a non hierarchical method. It intends to segment*n* observation into *k* partitions regarding to each observation which pertains to the part with the nearest mean. Following order is executed to proceed K-means clustering (Mingoti, S. A., & Lima, J. O.,2006):

1- Among all of the observations (data), *k* data are indiscriminately chosen regarding to the quantity of clusters.

2- For all of the remaining data (*N-K* observations), the nearest cluster is found according to the Euclidian distance with respect to $x_i=(x_{i1}\ldots x_{ip})$

3- Each observation is attributed to the closest cluster. After allocation of all observations, Euclidian distance related to each data and cluster's centroidis calculated and corroborate the possibility that it has been assigned to the closest cluster or not.

4-Finally the insignificant clusters are put away and rest of segments are iteratively trained to reach the ultimate clusters (Mingoti, S. A., & Lima, J. O., 2006).

3.3. K-Medoids Technique

The *k*-medoids method is considered a clustering technique associated with the *k*-means approach and the medoids shift algorithm. Both of the *k*-means and *k*-medoids techniques break the dataset up into groups and intend to minimize squared errorwhich is defined as the distance among points labeled to be in a cluster and a point designated as the center of that cluster (Chen et al, 2002). *k*-medoids technique chooses data points as medoids (centers).*K*-medoids is considered a partitioning approach of clustering as well that clusters the data set of *n* observations into *k* segments with *k* known *a priori*. The most prevalent realization of *k*-medoid clustering techinque is the Partitioning Around Medoids (PAM) procedure and is as follows(Chen et al, 2002):

a. Initializing step: select randomly *k* of the *n* observations as the centers (medoids)

b. Attribute each observation to the closest center.

c. For each center, *m*andeach data point *o* associated to *m* swap *m* and *o*, and compute the total cost of the arrangement that is the average dissimilarity of *o* to all the data points associated to *m*.(Chen et al, 2002). Choose the center *o* with the lowest cost of the arrangement.(Chen et al, 2002).

Steps *b* and *c are* repeated until no alternation in the assignments is achieved.

3.4. X-Means Algorithm

The X-means clustering technique is considered an improvement of the k-means clustering method(Pelleg and Moore 2000;Tucker,2010). This algorithm intends to enhance on three key features of the previous form (k-means) (Tucker,2010). First of all, it tries to omit the necessity of being informant about the number of clusters formerly (Tucker,2010). Secondly, expediting the computational scalability and ultimately, intensifying the criteria of search to renewcluster

centroid. The procedurethat X-means bring these enhancements regarding to it's ability ofselection criterion to determine what time to supplement or substitute a specific cluster center with child centers(Kass and Wasserman ,1995; Pelleg and Moore ,2000;Tucker,2010). Posterior probabilities (Tucker, 2010) are applied to specify the rank of all the built models. Pr[*Mj* |*D*], where *D* symbolizes the our data set and *Mj* represents each with a specific size of cluster (Pelleg and Moore, 2000;Tucker,2010).The Bayesian information criterion (BIC) has been applied by X-means technique to rank the created models and determine that which model is a more precise depiction of initial set of data. BIC is shown as follows (Kass and Wasserman, 1995; Pelleg and Moore 2000):

$$\text{BIC}(M_j) = l_j(D) + P_j/2 \ \log \text{R} \qquad (6)$$

l_j denotesthe log likelihood of the data brought by the maximum likelihood point(Tucker,2010). $P_j$mentions to the quantity of parameters in M_j. D represents the given set of data and*R* denotes the number of data points of candidate centroids.

3.5. Algorithms of Cluster Validity

Cluster validity algorithms are applied to measure the quality of clustering results built by various clustering methods or by applying dissimilar setting of values of parameters in similar methods. Some explanations about applied validity algorithms have been made in next part.

3.5.1.Validity Indexes

Multiple indexes and functions have been provided to measure validity of every segment. Additionally, mentioned indexes illustrate a comprehensible view on the efficient quantity of clusters.

3.5.1.1. Davies Bouldin Index

Davies Bouldin (DB) index is expressed mathematically as follow:

$$DB=1/(n\sum_{i=1,i\neq j}^{n} max\ (\sigma_i + \sigma_j/d(c_i + c_j)))(7)$$

n denotes the quantity of clusters,σ_i *and* σ_j represents the average distance of all objects in cluster *i*-*j*to center of their clusterrespectively, and d(c_i ,c_j) shows the distance of cluster centers c_i-c_j . Low values of Davies-Bouldin index associate withcompact clusters containing very similar objects, and whose centers are not close to each other. Consequently, the number of clusters minimizing Davies-Bouldin index is recognized as the optimum quantity of clustersThe cluster configuration refers to different factors that need to be set in different clustering algorithms such as specifying number of clusters for *k*-means method. For negative quantities of DB, the absolute amount is considered and lower amounts have better qualities.

3.5.1.2 Silhouette Index

One of the cluster validity algorithms is Silhouette validation method, which its' index is:

$$S(i)=(b(i)-a(i))/(max(a(i),b(i))) \quad (8)$$

a(i) implies the average unlikeliness of *i*-observation to all other observations in the same cluster and *b(i)* represents the minimum of average unlikeliness*of i*-observation to all observations in other segments (in the closest segment). Unlikeliness generally is regarded as the complement of similarity, and it's result consists of the number of attributes that two objects uniquely have compared with the total number of attributes among them. If value of Silhouette index becomes near to 1, it is interpreted that all the objects in the sample are clustered well. On contrary, If value of Silhouette index becomes near to 0, it is deduced that objects could be arranged to the other cluster that has the same distance with the current cluster. If value of Silhouette index becomes near to -1,it implies that sample point has been misclassified‖ it is merely somewhere among the clusters.

Table1 represents the two determined indexes for evaluation in this article:

Table 1: Indexes description

Name of Index	Evaluation criterion	Reference
Silhouette index	Larger Silhouette value is interpreted as better quality of clustering	(Chen et al,2002)
Davis-Bouldin index	Smaller value demonstrates higher clustering quality	(Kasturi,Acharya,Ramanathan,2003)

4. Problem Definition

4.1 Description of data

In this article, stock data, attributed to some firms in Iran Stock Exchange have been extracted from a financial information database. These stock data of 79 companies were referred to the start of currency crisis (22/09/2012- 22/03/2013) in Iran's economy.It was attempted to select these stocks from miscellaneoussegments of Iran's industries. Therefore, they quite depicted the condition of the market during that specific 6- month period.

Validation ratios and returns of stocks during determined intervals have been presumed as variables and they were given in terms of the dimensions of selected stocks. These applied dimensions have been mentioned in Table 2.

Table 2: Stock classification parameters

Parameter Type	Parameters	Significance
Periodic Returns	Daily	Short period
	weekly	
	monthly	
	3 months	Middle period
	6 months	
Validation ratios	Price to earning(P/E)	This ratiodemonstrates the eagerness intensity of investors to pay per dollar earning
	Price to Sale(P/S)	It shows the importance of the stock sales.
	Price to Book Value (P/BV)	It illustrates the price compared with its book value

4.2. Data preparation

Raw data always has problems with invalid or missing values. Hence, data pre-processing is regarded as an indispensable state in any Data Mining procedure to detect and correct the deviant information. In the data-cleaning phase, some records have been detected that cannot

be inserted into the database due to invalid or missing values. Thus, a modification has been made to set these bad stocks and ultimately we achieved 79 clean stocks,which had complete variables. Furthermore, this process contains the data normalization which transforms our data to the range of (0-1) inclusive that is one of the applications of applied software (Rapid Miner).

4.3. Applied Techniques

In this paper, comparisons on performances of three clustering algorithms including: K-means, K-medoids and X-means approaches were made on the same dataset. They are the most common approaches to clustering, found in literature. K-means and K-medoids were selected for their simplicity and fast execution, whereas X-means clustering algorithm was chosen for its scalability and ability to handle large datasets. Three selected clustering techniques were performed and regarding to chosen evaluating indexes,the optimum quantity of clusters for each ofmentioned technique was inferred. It is important to mention that, by considering one of thementioned appraisement indexes lonely, it will be imprecise to make a deduction about the efficient number of clusters. Hence, the optimum is detected through comparisons of all brought outcomes of validity indexes.

4.4. K-Means

By the use of Rapid Miner software, K-means clustering approach was implemented and brought results have been demonstrated in table 3.The number of clusters varies between 2 to 10 inclusive.This table implies that the optimal number of clustersin implementation of K-means method is eight for our given stock data. As it is shown, according to DB index,(-2.04) has been calculated for eight –cluster separation and it is the minimum (For negative amounts, the absolute amount is considered and lower amount has a better quality) amount among the other figures. On the other hand, it is compatible with results brought from Silhoutte index which considers 8 –cluster output (regarding to .434) as the optimal consequence. Figure 1 illustrates the diversification percentage of stocks in each cluster deduced by K-means algorithm.

Table 3: Validity indexes of K-means

K-means Method	Number of clusters								
Applied indexes	2	3	4	5	6	7	8	9	10

Davies-Bouldin	-3.56	-2.90	-3.66	-2.67	-2.94	-2.29	-2.04	-2.48	-2.46
Silhoutte	.305	.338	.317	.398	.366	.417	.434	.402	.400

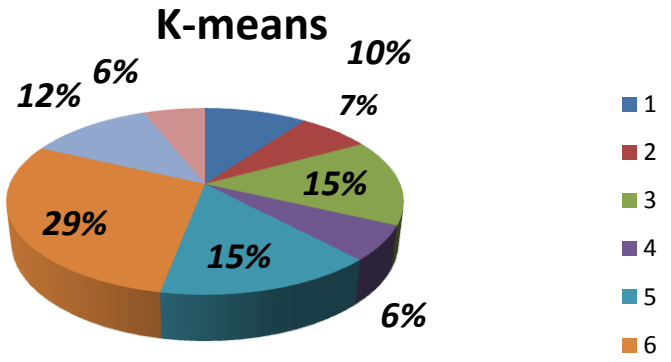

Figure 1: Diversification percentage of stocks in each cluster deduced by K-means method

4.5. K-Medoids

By the use of Rapid Miner software, K-medoids clustering approach was implemented and brought results were illustrated in table 4. The number of clusters varies between 2 to 10 inclusive .Regarding to achieved results; the optimal number of clusters in carrying out of K-medoids method is 4 for our given stock data.As it was stated,the optimum quantity of clusters is inferred withcomparison of all applied indexes. Four- cluster output has less DB amount in comparison with ten-cluster separation, but it has a so much better situation in silhouette index and is considered the optimum. Figure 2 depicts the diversification percentage of stocks in each cluster deduced by K-medoids algorithm.

Table 4: Validity indexes of K-Medoids

K-Medoids mehod	Number of clusters								
Applied Indexes	2	3	4	5	6	7	8	9	10
Davies-Bouldin	-2.26	-2.15	-1.98	-2.02	-2.21	-2.16	-2.43	-2.15	-1.97

Silhoutte	.310	.305	.342	.335	.312	.315	.284	.318	.334

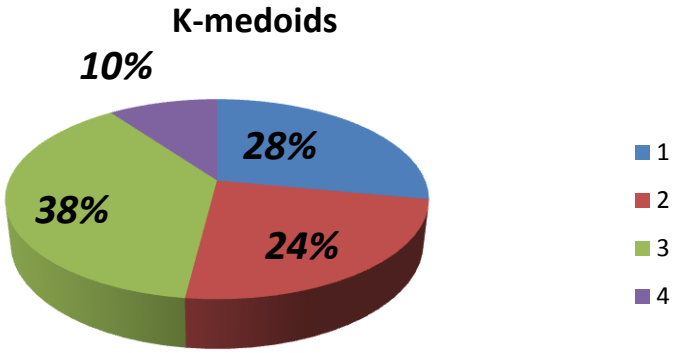

Figure 2: Diversification percentage of stocks in each cluster deduced by K-medoids method

4.6. X-Means

By the use of Rapid Miner software, X-means clustering approach was implemented and brought results have been shown in table 5. The number of clusters varies between 2 to 10, inclusive .This table implies that the optimal quantity of clusters in execution of X-means method is 9 for our given stock data. As it is shown, according to DB index,(-3.22) has been calculated for nine – cluster separation and it is the maximum quantity among the other figures. On the other hand, it is compatible with results brought from Silhoutte index which consider 9 –cluster output (regarding to .418) as the optimal consequence. Figure 3 illustrates the diversification percentage of stocks in each cluster deduced by K-means algorithm.

Table 5: Validity index of X-means

X-means Method	Number of clusters								
Applied Indexes	2	3	4	5	6	7	8	9	10
Davies-Bouldin	-6.55	-5.24	-4.81	-4.40	-4.44	-3.87	-4.12	-3.22	-3.26
Silhoutte	.267	.302	.335	.342	.342	.395	.364	.418	.401

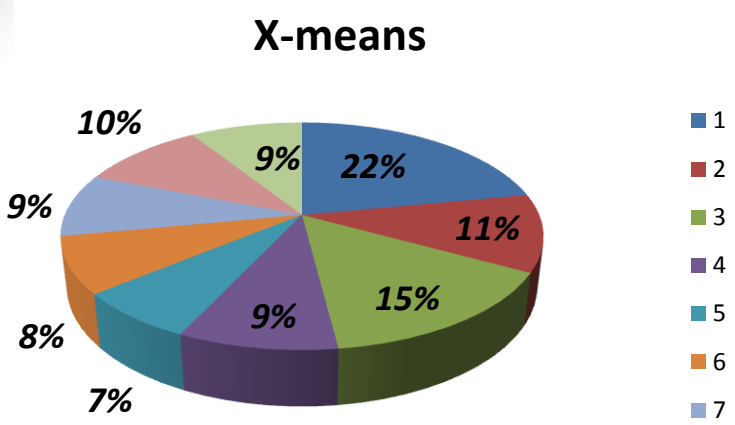

Figure 3: Diversification percentage of stocks in each cluster deduced by X-means method

4.7. Appraisement of Performance By The Use Of Intraclass Inertia

Intraclass inertia(*F(k)*) (Michaud,1997) is used to compare the performance of clustering algorithms(Michaud,1997). It is interpreted as a measure, showing the compactness of each cluster subject to fixed number of clusters. More ever, it is interpreted as the average squared Euclidean distance between every object and the mean of it's cluster. The consequences of using Intraclass inertia have been demonstrated in figure 4.It can be understood that K-means has a better compactness quality, because of having higher amount of *F(K)*rather than two other techniques.The horizontal vector represents the number of clusters and vertical vector shows the Intraclass inertia.

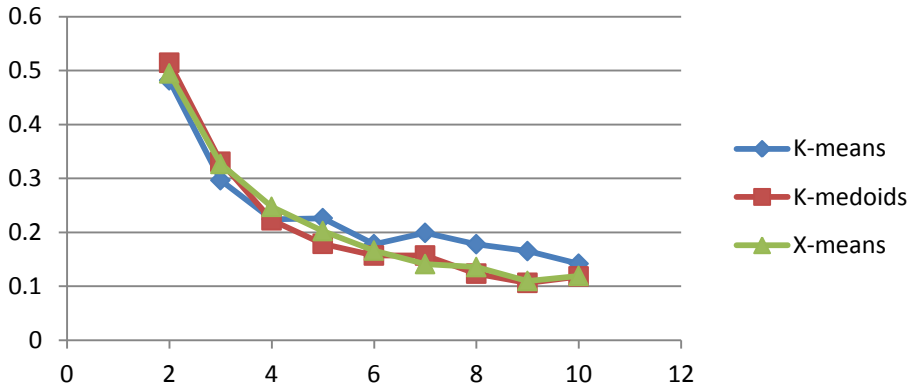

Figure 4: The Intraclass inertia for the three applied clustering algorithms regarding to different number of clusters

4.8. Building An Efficient Portfolio

According to brought clustering results, stocks from the most desirable cluster resulted by K-means algorithm (best method for our given data)are chosenand their pertinent weights are calculated regarding to Markowitz model.As it was adduced,the optimal number of cluster in K-means method was 8.After evaluation of these brought clusters, a specific cluster, which comprised stocks with high rates of profitability and stability during chosen 6-month period, was selected.This desirable cluster included 12 stocks(15% of total number of stocks) and five stocks were chosen to create a portfolio. These stocks had a sustainable trend in their weekly returns. Additionally, the domains of their weekly return fluctuations were subtle in comparison with other stocks. These properties of stocks were our criteria for building portfolio. Figures (5,6,7,8, and 9) portray the weekly returns of five selected stocks, which constitute the components of our portfolio. The horizontal vector demonstrates the number of weeks and vertical vector shows the weekly returns of each stocks. these stocks were:Vemadan,Fark,Folad,Shabehrn,Shfar.

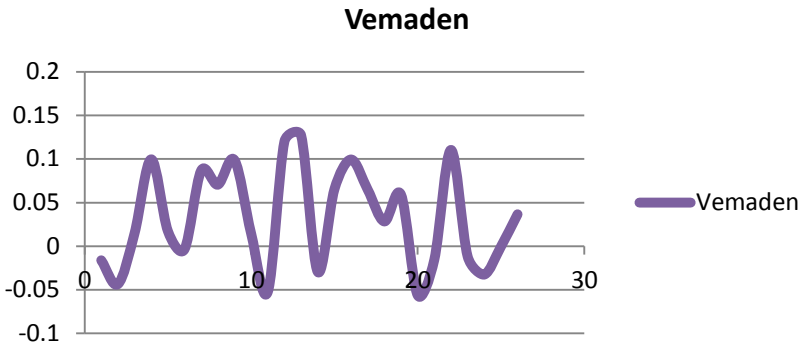

Figure 5: Weekly returns of one of the five selected stocks (Vemaden)

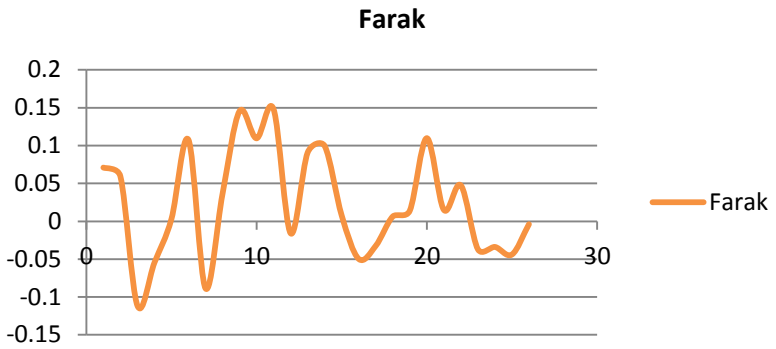

Figure 6: Weekly returns of one of the five selected stocks (Farak)

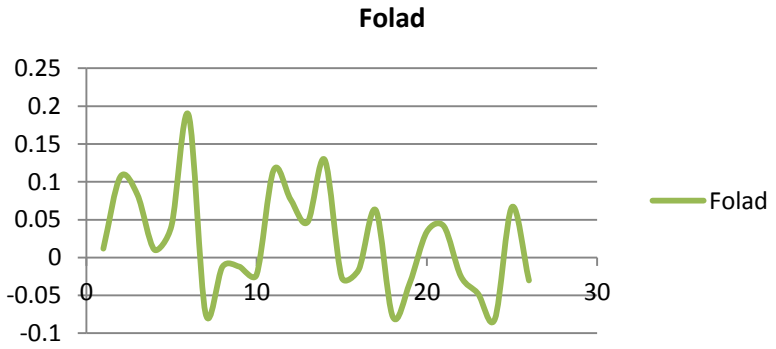

Figure 7: Weekly returns of one of the five selected stocks (Folad)

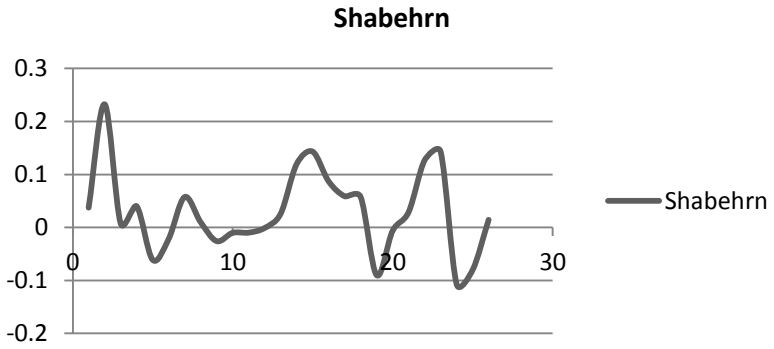

Figure 7: Weekly returns of one of the five selected stocks (Shabehrn)

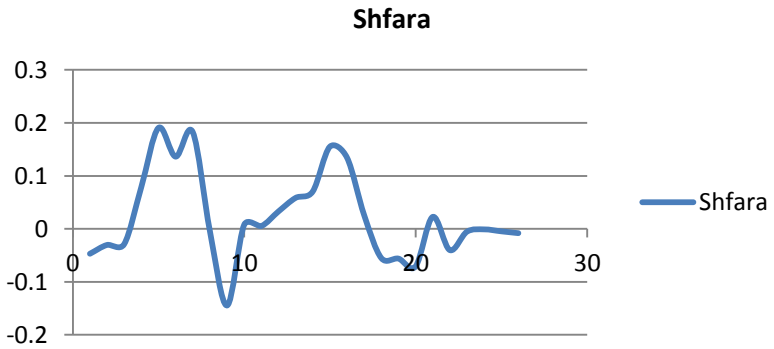

Figure 7: Weekly returns of one of the five selected stocks (Shfara)

By the use of this procedure, there is a great advantage in terms of reducing the number of possibilities in building an efficient portfolio. This approach cab practically be used by investors, specifically at the time of economic crisis, when all sections of economy such as stock market are influenced tragically. Built portfolio is portrayed in table 5.Their weights were calculated by implementation of Markowitz model and concept of risk minimization. Figurer 8 demonstrates the scaled plot which containsweekly returns of created portfolio and chosen benchmark (TSE50, which includes fifty active companies in Iran Stock Exchange). The chosen time period was six months (2/09/2012 - 22/03/2013) .This scaled plot implies that built portfolio had a better condition rather thanTSE50 as a whole comparison.Ultimately, because of diversification of portfolio (regarding to Markowitz model)the level of risk was

optimized (minimzing direction) and it also reacted fairly desirable in terms of brought returns in comparison with the applied index(TSE50) without considerable downward drift in our data.

Table 5: Created portfolios and weights of it's components

Portfolio(formed by results of K-means method)	
Names of stocks	Weights
Shfara	.13
Folad	.21
Vemaden	.25
Farak	.09
Shabehrn	.32

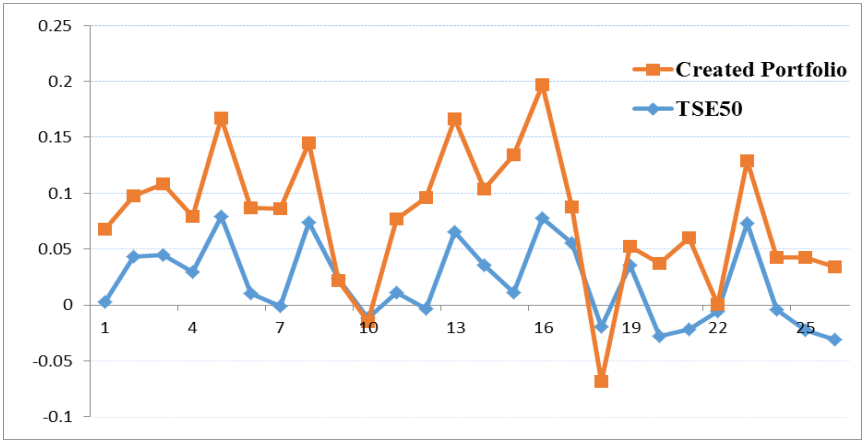

Figure 8: Scaled plots depicting weekly returns of the built portfolio respect to TSE50

Figure 9 represents a schematic depiction of proposed approach for solving a portfolio management problem and it includes all the mentioned procedures.

Fig.9.Schematic depiction of proposed approach

5. Conclusion

In this paper,Clustering approach was applied to classify 79 stocks in Iran Stock Exchange. Selected time period was simultaneous with the beginning of currency crisis in Iran's economy. First method is namely called K-means and second and third onesare called K-medoids and X-means respectively.According to defined indexes(Davis-Bouldin and Silhoutte) and intraclass inertia, optimal number of clusters was attained and it was inferred that K-means algorithm had a better quality in terms of generating more compact clusters(High quantity of intraclass inertia) in comparison with other methods in our data set. Reduction of possibilities for making efficient portfolios, which results in great contribution for investors, is considered a remarkable advantage of applied approach; specifically, implementation of Markowitz model satisfies the necessity of portfolio diversification dimension.Putting similar stocks in one cluster leads the investors to have a better chance to diversify their portfolios as well as having high rates of return in their investments. As an instance, five stocks associated with the most desirable cluster (Based two guidelines: high rate of profitability, existence of stability in their return trends) engendered by the most efficient algorithm (K-means) were selected.Weights of these stocks were calculated according to Markowitz modeland a sample portfolio was formed. The weekly returns of this sample portfolio were compared with TSE50 (benchmark index) returns.The determined time period was six months including (22.09.2012- 22/03/2013).This scaled plot implied that created portfolio had a better condition rather than TSE50 as a whole comparison. Proposed approach can result in a lot of employment in risk management domain and devising trading strategies in financial markets as it tries to find an efficient solution for portfolio management problems.

References

A. Qin, P. Suganthan, Enhanced neural gas network for prototype-based clustering,*Pattern Recognition* 38 (8) (2005) 1275–1288

Backer, U. E. (1995).Computer-assisted reasoning in cluster analysis.*PrenticeHall*

Bezdek, J. C., & Pal, N. (2005). Cluster validation with generalized Dunn's indices. In Proceedings of *the 2nd New Zealand two-stream international conference onartificial neural networks and expert systems* (p. 190).

Cheeseman P., Stutz J., (1996), Bayesian Classification (AutoClass): theory andresults. In: Fayyad U., Piatelsky-Shapiro G.,

Smyth P., Uthurusamy R., editors, *Advances in Knowledge Discovery and Data Mining*, Cambridge, MA: AAAI Press/MIT

Chen, G., Jaradat, S. A., Banerjee, N., Tanaka, T. S., Ko, M. S. H., & Zhang, Q.(2002).Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data. *Statistical Sinica*, 12, 241–262.)

Chui-Yu Chiu *, Yi-Feng Chen, I-Ting Kuo, He Chun Ku,An intelligent market segmentation system using k-means and particle swarm optimization.*Expert Systemswith Applications*, 36 (2009) 4558–4565.

Delibasis, K. K., Mouravliansky, N., Matsopoulos, G. K., Nikita, K. S., & Marsh,

A.(1999). MR functional cardiac imaging: Segmentation, measurement and WWW based visualisation of 4D data. *Future Generation Computer Systems*,15(2),185–193

Ester, M., Kriegel, H.-P., Sander, J., Xu X., (1996), A density based algorithm for discovering clusters in large spatial databases, *Proceedings of the 1996 International Conference on Knowledge Discovery and Data Mining (KDD'96)*,Portland, pp. 226–231

Fayyad, U., Piatetsky-Shapiro, G., & Smith, P. (1996). From data minin toknowledgediscovery in database.*American Association for Artificial Intelligence*, August, 37–54,Intelligence, August, 37–54.

Fernandez, L. (2005). A diversified portfolio: Joint management of non-renewableand renewable resources offshore. *Resource and Energy Economics*, 27, 65–82

Gentle, J.E., Härdle, W., Mori, Y., (2004), *Handbook of Computational Statistics*, SpringerHand, D., Mannila, H., Smyth, P., (2001), *Principles of Data Mining*, A BradfordBookthe MIT Press Cambridge, Massachusetts London, England.

H.W. Shin, S.Y. Sohn,Segmentation of stock trading customers according to potential value,*Expert Systems with Applications* 27 (2004) 27–33

Jain, M. Murty, P. Flynn, Data clustering: a review, ACM Computing Surveys 31 (1999) 264–323.

J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd ed.Morgan Kaufmann, 2006.

Kass, R.E. andWasserman, L., 1995. A reference Bayesian test for nested hypothesesand its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90 (431), 928–934.

Kasturi, J., Acharya, R., &Ramanathan, M. (2003).An information theoretic approach foranalyzing temporal patterns of gene expression.*Bioinformatics*, 19, 449–458.

Kuo, R.J (2011). "Application of a hybrid of geneticalgorithm and particle swarm optimizationalgorithm for order clustering", *Decision Support Systems*, 201011

Michaud, P. (1997). Clustering techniques.*Future Generation Computer System*, 13(2), 135–147.

Mingoti, S. A., & Lima, J. O. (2006).Comparing SOM neural network with fuzzy C means,K-means and traditional hierarchical clustering algorithms.*EuropeanJournal of Operational Research*, 174, 1742–1759.

Markowitz, H. (1952). Portfolio selection.*Journal of Finance*, 7, 77–91.

Östermark, R. (1996). A fuzzy control model (FCM) for dynamic portfoliomanagement.*Fuzzy Sets and Systems*, 78, 243–254.

Oh, K. J., Kim, T. Y., & Min, S. (2005). Using genetic algorithm to suppoportfolio optimization for index fund management.*Expert Systems with Applications*, 28,371–379

Pelleg, D. and Moore, A., 2000. X-means: Extending k-means with efficient estimationof the number of clusters. *Proceedings of the 17th international conference on machine*

learning, 29 June–2 July, Stanford University,California. San Francisco,CA: Morgan Kaufmann, 727–734.

PreetiParanjape-Voditel, UmeshDeshpande, A stock market portfolirecommender system based on association rule mining,*Applied Soft Computing*, 13 (2013) 1055– 1063

Romesburg, H.C., (2004), *Cluster analysis for researchers*, Lifetime LearningPublications,Belmont,CA,UnitedState

Tucker, Conrad S. , Kim, Harrison M. , Barker, Douglas E. and Zhang, Yuanhui(2010) 'AReliefF attribute weighting and X-means clustering methodology for top-down productfamily optimization', *Engineering Optimization*, 42: 7, 593 — 616, First published on: 08 April 2010

Wang, W., Yang, J., Muntz, R., (1997), STING: a statistical information grid approach tospatial data mining, *Proceedings of the 1997 International Conference on Very Large Data Base (VLDB'97)*, Athens, Greek, pp. 186–195.