International Journal of Finance, Accounting and Economics Studies, 00(00), 2022, 55-71.
Print ISSN: 2251-645X Online ISSN: 2383-2517

Islamic Azad University
Science and Research Branch
Faculty of Management and Economics

International Journal of Finance, Accounting and Economics Studies
Journal homepage: https://ijfaes.srbiau.ac.ir

A hybrid Model on the Basis of Data Envelopment Analysis and Data Mining Techniques to Analyze the Investment Behavior in Stock Exchange: A Real Case Study in Tehran Stock Exchange

Saiedeh Molla Hosseinagha [1*]
Kaveh Khalili-Damghani [2]

[1]*Department of Industrial Engineering, South-Tehran Branch, Islamic Azad University, Tehran, Iran*
Email: Smha67@gmail.com
[2]*Department of Industrial Engineering, South-Tehran Branch, Islamic Azad University, Tehran, Iran*
Email: kaveh.khalili@gmil.com
Email: k_khalili@azad.ac.ir

Article History
Received Date: 03 May 2020
Revised Date: 28 July 2020
Accepted Date: 20 September 2020
Available Online: 19 April 2022

JEL Classification:

Keyword:
Stock exchange
Data mining; Clustering;
Data envelopment analysis (DEA)
Efficiency

Abstract

Stock exchange, which is the core of investment market, transfers scattered resources to the productive sector of the service and industry. The analysis of stock exchange has attracted lots of researches in recent decades. The issue concern both of securities and investors. More formally both sides of investment (i.e., behavior of investors and fluctuations of returns on securities) form the markets. In this paper, a hybrid model on the basis of data envelopment analysis and data mining techniques is proposed to analyze the investment behavior in stock exchange. The proposed approach is applied in Tehran stock exchange. First, the financial data of investments on selective companies during 2013-2014 are analyzed using data mining approaches to recognize the behavioral patterns of investors and securities. Second, customers are clustered in 3 selling and 4 buying groups using data mining techniques. Then, the efficiency of active companies in stock exchange is evaluated using and input-oriented Data Envelopment Analysis (DEA) model considering variable returns-to-scale (VRS) assumption. Data mining techniques and DEA model are codified using MATLAB and LINGO software, respectively. The obtained results indicate that in 2013 and 2014, nine and six companies are efficient, respectively. Among these companies, six companies are efficient in both years. The most visited companies in reference sets are determined during periods of this study. The reference set of inefficient companies are determined in order to project them toward efficient frontier. The results of this study can provide insightful vision to investors in order to illustrate the behavior of previous investors and securities.

*.Corresponding Author: Smha67@gmail.com

1. Introduction

Creating a suitable environment for investors to invest in stock exchange markets is essential for economy progress. Stable economic condition usually leads to a predictable environment and all these will persuade investors for a more reliable investment. Investment market plays an important role in the allocation of financial source to companies in developing countries. Hence, Identification of key criteria for a safe investment is critically important. Processing big historical data of past investments in a specified filed can be very useful for making efficient investment decisions. Data mining is one of these processing methods [1]. Stock exchange, which can provide large data bases of historical data on previous investments, is a suitable resource for applying data mining approaches [2].

In today's world, stock exchange markets are rapidly growing. Every day, a large number of shares and huge amount of data are added to these markets. At the same time, global stock exchange markets are being created, which can be considered as an opportunity for investors. In such a situation, the main problem is the selection of proper investment criteria and the creation of optimal portfolios. Economists and experts believe that competition among actors in the market leads to mispricing of shares in the markets. This leads to failure in random selection of shares. It is very important for us to know which factors have an impact on the behavior of investors. Another issue which has created some problems in Iranian investment markets is the deficiencies in analyses. A lot of experts and investors believe that price divided by earn per share (P/E) is the main index of change in prices. Lack of long-term models has pushed investors to employ short-term models for analyses. When a large number of investors invest their money on the basis of a limited number of models, many people simultaneously invest in one industry and draw back their money at the same time. This intensifies problematic behaviors in the market. When there are a lot of fluctuations in economic variables, profitability of companies is affected by a lot of changes. In such a situation, investors do not trust the published data. Besides, operational conditions and efficiency might have an influence on share prices. Brigham and Houston [19] showed that rates in the financial statements have an impact on share prices. In other words, figures in the financial statements show performance and efficiency of the companies. Investment in efficient companies produces higher returns. Here, the main problem is that the evaluation of efficiency of companies with various criteria and indices is not an easy task. In addition, decision-making on the basis of limited number of indices is not true. Although the evaluation of companies by financial statements is difficult, DEA models can help incorporating several criteria and indices when assessing the companies [20]. In this study behavior of investors in 2013 and 2014 is analyzed. The data are gathered from Tehran stock exchange. The efficiency of companies whose shares have been bought are determined and compared.

This paper involves in examining the investment behaviors of investors in Tehran stock exchange. First, well-known data mining approaches such as clustering methods are used to investigate the transaction behavior of investors using 30 criteria. Afterward, data envelopment analysis (DEA) approach is employed to assess the efficiency of investments. The novelties of this study are divided into two parts. In the first part data mining methods are used to study the behavior of customers of stock markets in Iran on the basis of their transactions. The second part includes the evaluation of efficiency of active companies in Tehran Stock Exchange using DEA. The proposed approach is applied to all active industries in Tehran Stock Exchange.

The next parts of this paper are organized as follows. In Section 2 a brief literature of past researches works are reviewed. In Section 3 the proposed approaches are discussed. The case study and results are presented in Section 4. Finally, the paper will be concluded in Section 5.

2. Review of Past Research Works

In this section, a number of studies that have been conducted by clustering and DEA are reviewed.

2.1 Application of Data Mining Approaches in Stock Exchange

Bastlo et al. [3] clustered stock market companies via chaotic map synchronization. Doherty et al. [4] used hierarchical clustering algorithms to classify sectors of financial markets in a 10-year period. Nanda et al. [5] used the three clustering methods i.e., K-means, SOM, and fuzzy K-means, to classify shares. in order to investigate the Indian stock market data for portfolio management. Does et al. [6] used clustering methods and financial time series for enhanced index tracking portfolios. Song et al. [7] used shrinking-based clustering approach. It was a non-parametric method and used a simple clustering algorithm on the basis of error square of Euclidean distance. Jain et al. [8] presented a definition of clustering and reviewed various clustering methods.

2.2 Application of DEA in Stock Exchange

In 1957, Farrell [9] used a method on the basis of estimation of production function to measure the technical efficiency in a manufacturing company incorporating one input and one output. Farrell [9] also used the approach to evaluate the efficiency of farming industry in the USA compared to the other countries. In 1978, Charnes et al. [10] developed Farrell's idea and presented a model that had the ability to measure the efficiency in presence of several inputs and outputs. The model proposed by Charnes et al. [10] was the first official Data Envelopment Analysis (DEA) model. In 1984, Banker et al. [11] extended the proposed model of Charnes et al. [10] in presence of variable returns-to-scale assumptions, called BCC model.

DEA models have impressively been extended in both theoretical and practical aspects since the work by Charnes et al. [10]. Edisinghe and Zhan [12] selected portfolios using DEA and financial strength indicators in USA portfolio market. In this study, the correlation among financial indices was investigated. Results of this study showed that DEA models with financial indices are efficient for evaluating the performance of companies in the technology sector of the USA [12]. Titko et al. [13] employed DEA and benchmarking to evaluate the performance of banks in Latvia and Lithonia. This study was conducted on the basis of the size and age of banks. Sungmook Lima et al. [14] investigated South Korean stock market using a new DEA cross-efficiency method to select the portfolio. Sungmook Lima et al. [14] used variances of cross-efficiencies under the mean-variance framework and showed efficiency of their method in comparison with other methods on benchmark problems. In 2015, Fenyves et al. [15] used DEA to evaluate the financial performance of farming industry. Khanjamali [16] conducted a study on relative efficiency of pricing of stock markets in a number of industries by DEA. Risk was taken as the input and expected returns were taken as the output in an input-oriented DEA model with variable return to scale assumption. Results indicated that shares of an industry had not been properly valued. Also, it was concluded that DEA can identify companies whose shares had been efficiently valued. Khajavi el al. [17] used an input-oriented DEA model with constant return to scale assumption for finding optimal portfolio of most efficient companies in stock exchange. The proportion of price to income, five-year beta coefficient, and five-year sigma coefficient were assumed as input variables. One-year return rate, three-year return rate, five year return rate, and income produced by every share were assumed as output variables. Ahmadzadeh [18] used DEA to choose shares and compared it with size-effect strategy in Tehran Stock Exchange. Ahmadzadeh [18] investigated several stocks such as car-manufacturing companies, drug producers, cement producers, and chemical companies. Results showed that portfolio selected by DEA models had a higher efficiency compared to industry return. Also, return rate of portfolio adjusting risk of DEA models did not have a significant difference with size-effect model and industry return model.

Mashyekhi el al. [24] have proposed a new multipurpose model which incorporated DEA cross-efficiency into mean-variance Markowitz model to select portfolio of investment. The model examined 52 companies operating in Iran's Stock Market. The results showed that the amounts of risk and return were considerably appropriate at the same time in comparison with Markowitz model and DEA. Huang el al. [25] proposed a new DEA model in which a large number of the inputs and outputs of financial indices were considered. The research was done on the companies listed in China's Stock Exchange. Edirisinghe el al. [26] developed a generalized DEA model to analyze the financial statements of a company over a period of time in order to determine the Relative Financial Strength Index (RFSI). A proposed model was applied on 230 companies operating in different businesses in America. Eero Pätäri el al. [27] applied DEA in portfolio selection problem. The results showed that DEA as a multi-criteria method was useful in cases when the number of the stocks of the sample was great.

3. Proposed Approach

The main steps of proposed approach of this study are presented here. First, some essential financial indices are determined and defined. Then, a clustering approach is conducted to cluster the considered companies of this study. The performance of each company and clusters are determined through DEA.

3.1 Main Financial Indices

In order to investigate the stock exchange some financial indices are required. Consulting a number of experts, who have enough experience in the related field, yielded to identifying the most essential financial indices. They are defined as follows.

Total assets=current assets + goods +pre-payments + fixed assets+ intangible assets + long-term investments + amortization (1)

Total debts=current debts – long term debts (2)

Payments to share-holders include capital, share reduction, savings, profits, and losses.

Operational profit (loss) = general profits – expenses of distribution and sale – general and administrative expenses ± net incomes and operational costs (3)

Return of equity shows the efficiency of management of company in employing the resources in order to obtain profits. It is an important index for showing income of the company. It is calculated using (4).

Return of equity =

$$\frac{profits\ and\ losses\ after\ reducing\ taxes}{total\ aeets} \times 100 \quad (5)$$

Return on assets shows amount of income produced by each share. It is obtained by dividing net income of company by payments to share-holders as shown in (5).

Return on assets = $\frac{profits\ and\ losses\ after\ reducing\ taxes}{payments\ to\ share-holders} \times 100$ (5)

3.2. Clustering Approach

The first phase of proposed approach is clustering of investment records in order to find subsets of data in a way that the variance within clusters is minimized and the variance between clusters is maximized [21]. One of the main problems in clustering is to determine the number of clusters. There are a several methods for determining the number of clusters. In this study, Wilk's lambda method as shown in (6) is used to determine the suitable number of clusters.

$$\text{Wilk's Lambda}=\frac{SS^{k}_{within}}{SS_{total}}, \quad k=2,\dots,k_{\max} \tag{6}$$

SS^{k}_{within} is within cluster variance and SS_{total} is total variance. If the diagram of Wilk's lambda coefficient is drawn on the basis of *k*, the first jump in the diagram is the optimal number of clusters.

Based on the data gathered from Tehran stock exchange, the following variables are used to cluster the customers.

- Online buying and selling of shares;
- time of buying and selling which is divided into four periods (i.e., the first six months of 2013, the second six months of 2013, the first six months of 2014, the second six months of 2014)
- Company and industry whose shares have been sold;
- The amount of sold or bought shares;
- Value of transactions, which is equal to the volume of transactions made by mean of share prices on that day.

3.3. Analyzing the Efficiency Scores

As the companies are assumed as DMUs should be analyzed using DEA approach in the next phases of this research, experts are asked to divide the financial indices into two main classes as inputs and outputs. Based on the views of experts of stock exchange, the indices are divided into inputs (total assets, total debts, payments to share-holders) and outputs (operational profits and losses, return of equity, return on assets, sales). A company is shown in Figure 1 as a DMU.

The input-oriented DEA considering variables return to scale assumption is used to evaluate the financial efficiency of companies. This model was first proposed by Banker et al. [11] and is as shown in (7).

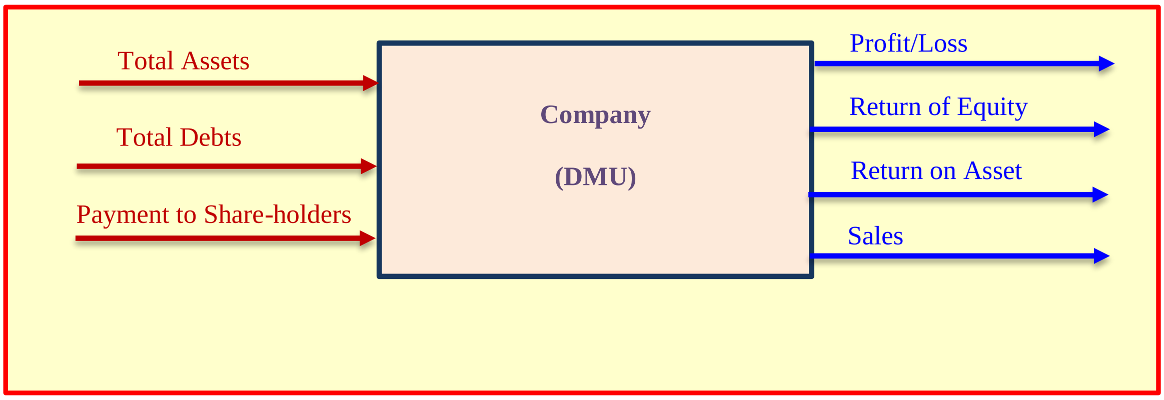

Figure 1. *Schematic View of a Company as a DMU*

$$Max \quad Z_p = \frac{\sum_{r=1}^{s} u_r y_{rp} + w}{\sum_{i=1}^{m} v_i x_{ip}}$$

s.t.

$$\frac{\sum_{r=1}^{s} u_r y_{rj} + w}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \quad j = 1,\dots,n$$

$$v_i \ge 0, \quad i = 1,\dots,m$$

$$u_r \ge 0, \quad r = 1,\dots,s$$

$$w \quad \text{free in sign} \tag{7}$$

The model (7) is a fractional mathematical programming problem and it global optimum solution is hard to find. So, Banker et al. [11] proposed the following linear form.

$$Max \quad Z_p = \sum_{r=1}^{s} u_r y_{rp} + w$$

s.t.

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + w \le 0, \quad j = 1,\dots,n$$

$$\sum_{i=1}^{m} v_i x_{ip} = 1$$

$$v_i \ge \varepsilon, \quad i = 1,\dots,m$$

$$u_r \ge \varepsilon, \quad r = 1,\dots,s$$

$$w \quad \text{free in sign} \tag{8}$$

In BCC model, the sign of variable W indicates return-to-scale for each DMU.
A: If W< 0, return-to-scale is descending
B: If W=0, return-to-scale is fixed
C: IF W> 0, return-to-scale is ascending
The dual of linear model (8) which is called the envelopment form is as model (9).

$$Min \quad w_p = \theta - \varepsilon \times \left(\sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^-\right)$$

s.t.

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{ip}, \quad i = 1,\dots,m$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{rp}, \quad r = 1,\dots,s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \ge 0, \quad j = 1,\dots,n$$

$$s_i^- \ge 0, \quad i = 1,\dots,m$$

$$s_r^+ \ge 0, \quad r = 1,\dots,s \tag{9}$$

These models can be used to evaluate the financial performance of clustered companies as DMUs which were schematically depicted in Figure 1.

4. Results and Discussion

The proposed approach of this study is applied in selected companies in Tehran stock exchange. Twenty-three representatives and active companies from various industries in Tehran Stock Exchange are considered. Those companies which have the highest number of shares in each industry are selected as representative of industry. Data of transactions of 56643 customers in 2013 and 2014 are gathered. After pre-screening, the data of those customers who have at least one buying and one selling record during this period are selected. Financial data of customers, companies, and industries are collected from Tehran stock exchange database. Excel and SPSS are used for manipulating and ordering of data. MATLAB is used for the clustering. GAMS software is utilized to codify the DEA models.

4.1. Proper Number of Clusters for Buying and Selling Data

As mentioned before Wilk's lambda coefficient is used to determine the suitable number of clusters. As shown in Figure 2, the first jumps for buying and selling data are occurred in $k=3$ and $k=4$ clusters, respectively. So, the suitable number of clusters on the basis of Wilk's lambda coefficient is 3, and 4 for buying and selling data, respectively.

4.2 Dimension Reduction by Principal Component Analysis

Principal Component Analysis (PCA), which is a method for reducing dimension of high-dimensional data on the basis of the direction of data dispersion [23], is used to reduce the dimension of buying and selling data in favor of plotting the results of clustering in two-dimension plot. PCA is used for both buying and selling data. The associated two-dimension clustering plots are shown in Figure 3.

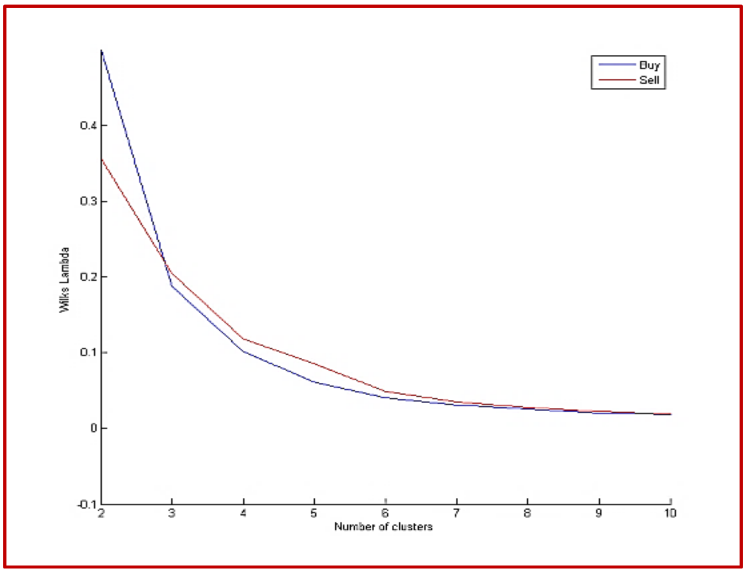

Figure 2. Number of Clusters VS Wilk's Lambda Coefficient

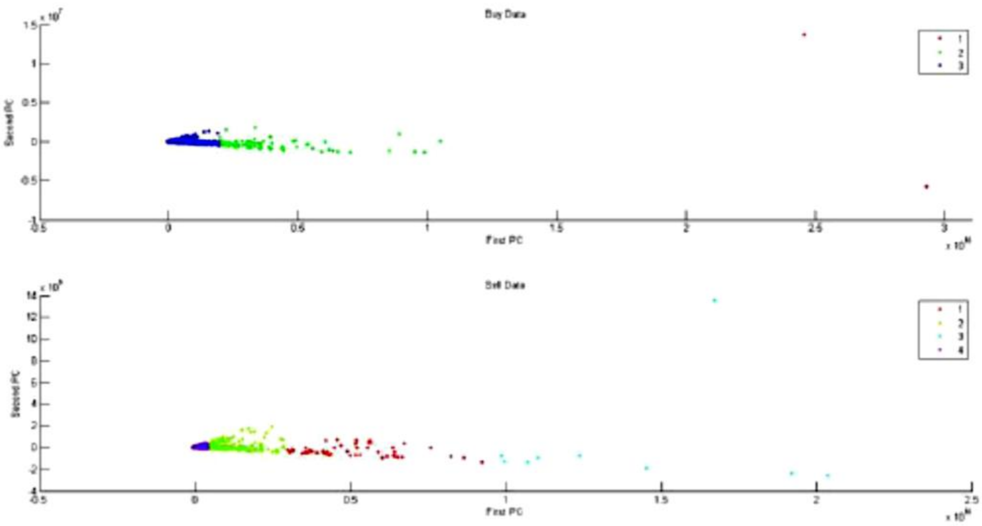

Figure 3. Data shown by Principal Component Analysis

4.3. Clustering Analysis using K-Means Method

Both buying and selling records are distinctively clustered using K-means method considering the suitable number of clusters suggested by Wilk's lambda Coefficient. It is notable that a record in the database is associated with a person who has accomplished a buy or a sell during 2013-2014. Results of clustering are shown in Table 1. It is notable that the clustering is accomplished in a 25-dimensional space.

As shown in Table 1, buying and selling include 3 and 4 clusters, respectively. In buying, the first cluster of buying section has lowest number of members among the others. Ii also has the highest average of trading value. Reconsidering the customers included in this cluster, it has been recognized that they were legal customers.

The third cluster in buying section has the highest number of members among the others. It is notable that the average of trading values is the lowest amount in comparison with other clusters. More formally, the customers included in this cluster are who invest in low volume and high diversity.

In selling section, third cluster has the lowest number of members and the highest average of trading value. The fourth cluster in this section has the highest number of members and the lowest average of trading value. Investigating among the customers of this cluster reveals that they are real customers with an average investment of 15617 USD.

Hit number and percentage of records in both buying and selling clusters are shown in Table 2 and Table 3, respectively.

It is notable that the sum of values in each column of Table 3 is not equal to 100. For instance, in fourth cluster of selling data, %80 of people used online shopping method. In other words, among 24503 people in this cluster, 19609 people used online method

Table 1. Results of Clustering

Data	**Time of clustering**	**Whole data**	**MSE ($\times 10^{16}$)**	**Clusters**	**Number of members of each cluster**	**Variance of each cluster ($\times 10^{16}$)**	**Mean of sales**	**Mean of transaction values***
Buying	1.52 seconds	31587	2.5806	First cluster	2	0.1122	13050000	77125415
				Second cluster	78	2.9596	938669	11374316
				Third cluster	31507	5.0798	17542	142916
Selling	0.78 second	25056	1.9059	First cluster	49	1.0443	1066943	14754968
				Second cluster	495	1.2555	292086	2610905
				Third cluster	9	1.3095	3892495	39812682
				Fourth cluster	24503	1.6608	15617	124063

***Mean of transaction values, One US dollar=35000 IRR**

4.4. Results of customer clustering

Buying data and selling data included 31587 and 25056 records, respectively. The third cluster of buying with 31507 records and fourth cluster of selling with 24503 records have the highest number of members among the clusters. This shows that behaviors of customers in these clusters are similar. Figures 4-7 clearly describe the resultant clusters.

Table 2. Number of people who used various industries in each cluster

	Data	Data of buying			Data of selling			
		First	Second	Third	First	Second	Third	Fourth
1	Online shopping	0	13	23594	7	215	1	19609
2	First sixth months of 2013	0	16	6688	7	102	2	4565
3	Second sixth months of 2013	2	39	14257	12	196	0	9569
4	First sixth months of 2014	0	14	6340	23	114	6	6131
5	Second sixth months of 2014	0	9	4222	7	83	1	4238
6	Telecommunication	0	2	1520	3	20	0	1271
7	Insurance	0	1	800	0	6	0	599
8	Machinery	0	5	1295	4	26	1	1119
9	Manufacturing of metallic	0	0	431	0	13	1	376
10	Basic metals	0	3	2651	0	56	1	2274
11	Candy and sugar	0	0	237	0	0	0	174
12	Drilling	0	0	363	0	12	0	337
13	Transportation	0	1	285	0	5	0	267
14	Other non-metallic products	0	0	415	0	2	0	386
15	Car	0	10	2524	4	41	0	2287
16	Tile	0	0	71	0	1	0	82
17	Technical and engineering	0	20	2141	16	80	3	1647
18	Computer	0	1	443	1	11	0	347
19	Mass, Real Estate	0	1	633	1	5	0	437
20	Cement, lime	0	1	830	2	7	0	731
21	Chemical products	0	19	5548	13	91	1	3657
22	Medical products	1	0	1540	1	12	0	1264
23	Bank	1	3	5184	0	53	1	3281
24	Food products except sugar	0	3	1030	0	6	0	912
25	Extraction of minerals	0	2	1141	0	19	0	847
26	Multi industrial companies	1	5	1491	4	20	2	1333
27	Mediators	0	1	774	0	8	0	726
28	Investment	0	0	160	0	1	0	149

Table 3. Percentage of people who used industry with a special method in each cluster

Variable	Data	Data of buying			Data of selling			
		First	Second	Third cluster	First	Second	Third	Fourth
1	Online shopping	0	16.7	74.9	14.3	43.4	11.1	80
2	First sixth months of 2013	0	20.5	21.2	14.3	20.6	22.2	18.6
3	Second sixth months of 2013	100	50	45.3	24.5	39.6	0	39.1
4	First sixth months of 2014	0	17.9	20.1	46.9	23	66.7	25
5	Second sixth months of 2014	0	11.5	13.4	14.3	16.8	11.1	17.3
6	Telecommunication	0	2.6	4.8	6.1	4	0	5.2
7	Insurance	0	1.3	2.5	0	1.2	0	2.4
8	Machinery	0	6.4	4.1	8.2	5.3	11.1	4.6
9	Manufacturing of metallic	0	0	1.4	0	2.6	0	1.5
10	Basic metals	0	3.8	8.4	0	11.3	11.1	9.3

Variable	Data	Data of buying			Data of selling			
		First	Second	Third cluster	First	Second	Third	Fourth
11	Candy and sugar	0	0	0.8	0	0	0	0.7
12	Drilling	0	0	1.2	0	2.4	0	1.4
13	Transportation	0	1.3	0.9	0	1	0	1.1
14	Other non-metallic products	0	0	1.3	0	0.4	0	1.6
15	Car	0	12.8	8	8.2	8.3	0	9.3
16	Tile	0	0	0.2	0	0.2	0	0.1
17	Technical and engineering	0	25.6	6.8	32.7	16.2	33.3	6.7
18	Computer	0	1.3	1.4	2	2.2	0	1.4
19	Mass, Real Estate	0	1.3	2	2	1	0	1.8
20	Cement, lime	0	1.3	2.6	4.1	1.4	0	3
21	Chemical products	0	24.4	17.6	26.5	18.4	11.1	14.9
22	Medical products	0	0	4.9	2	2.4	0	5.2
23	Bank	50	3.8	16.5	0	10.7	11.1	13.4
24	Food products except sugar	0	3.8	3.3	0	1.2	0	3.7
25	Extraction of minerals	0	2.6	3.6	0	3.8	0	3.5
26	Multi industrial companies	50	6.4	4.7	8.2	4	22.2	5.4
27	Mediators	0	1.3	2.5	0	1.6	0	3
28	Investment	0	0	0.5	0	0.2	0	0.6

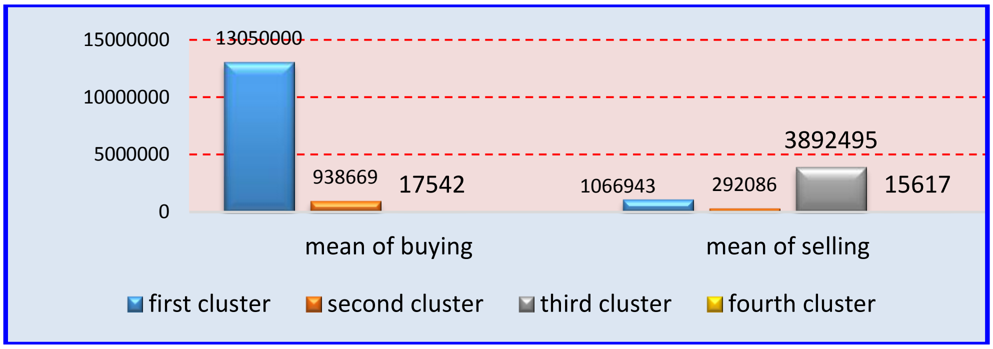

Figure 4. Mean of transacted shares in each cluster

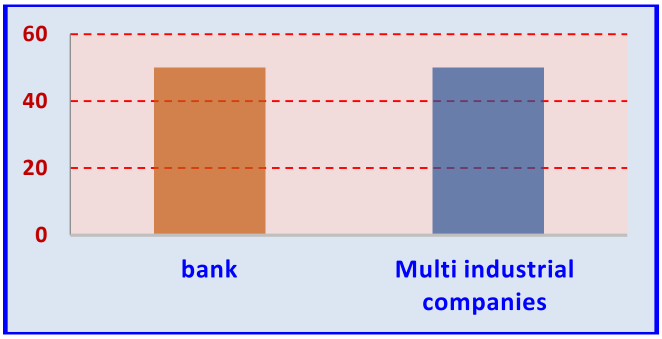

Figure 5. Transactions of buying shares in the first cluster

As can be seen in Figure 5, in the first cluster, investors have invested in two sectors (Bank and Multi Investment Company). This cluster has the highest value of transactions and the lowest dispersion of investment in various industries.

As can be seen in Figure 6, in the second cluster, investors have invested in 16 industries. Technical and engineering services and chemical products have the largest shares in buying transactions of this cluster.

As can be seen in Figure 7, in the third cluster, investors have invested in all 23 industries. Among these industries, bank and chemical products have the largest shares.

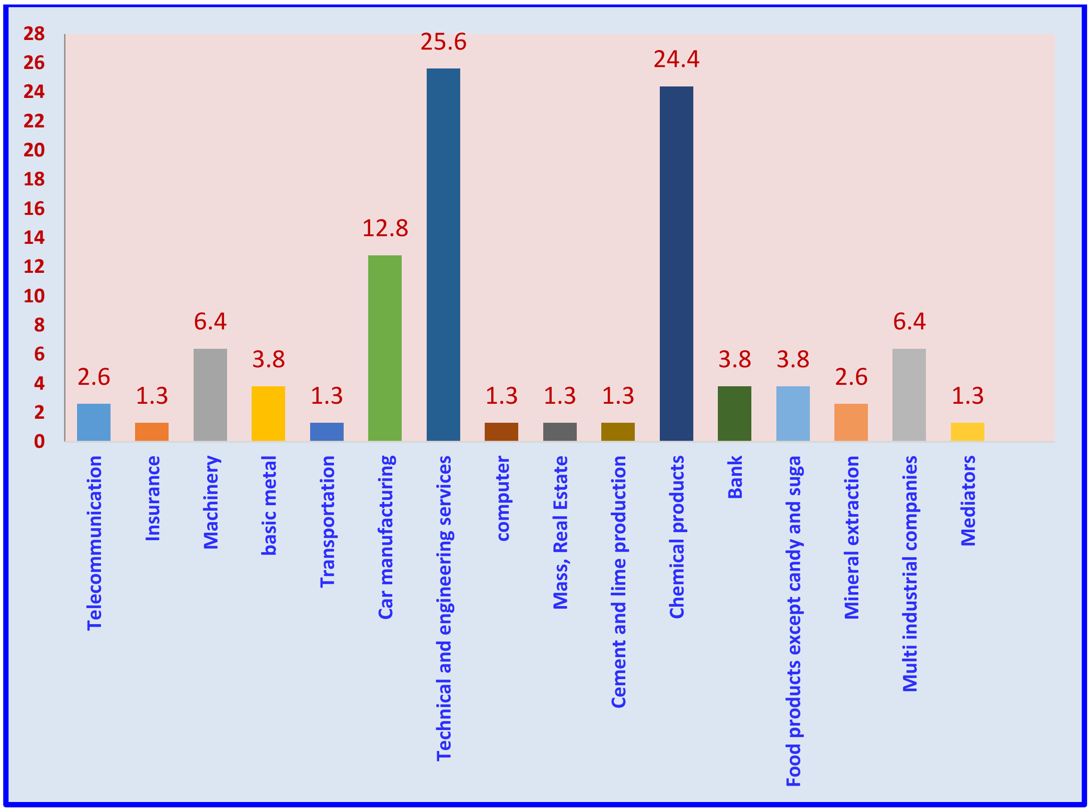

Figure 6. Buying Transactions in the second cluster

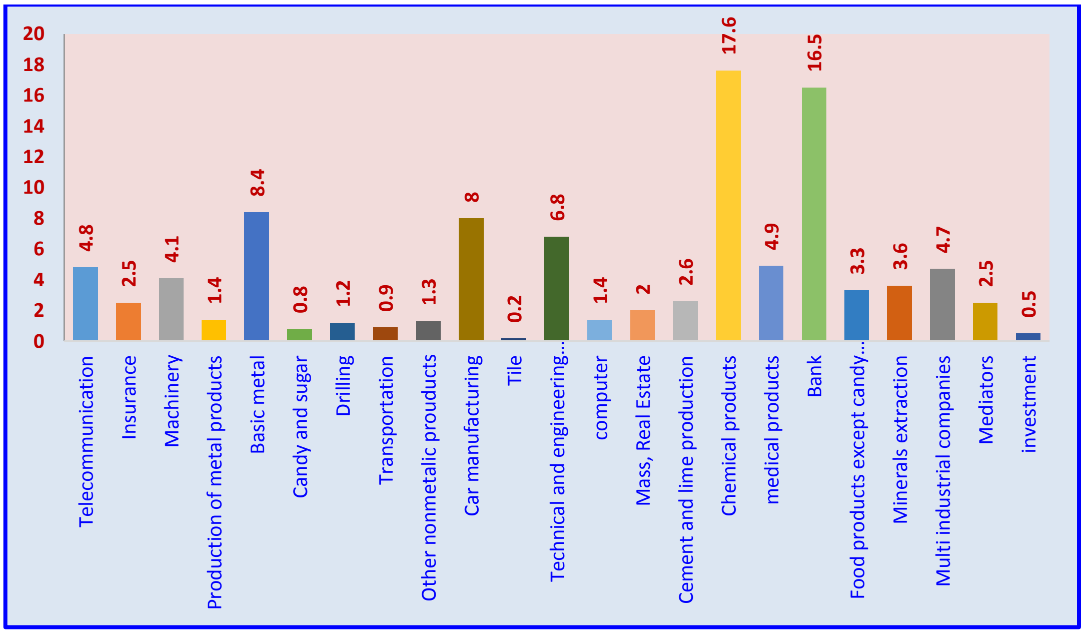

Figure 7. Buying Transactions in the thirds cluster

4.5. Results of Efficiency Measurement using DEA in each Cluster

In the pre-screening phase, negative data are shifted using a suitable map. Data are also normalized. The DEA models are codified in GAMS software. The input-oriented DEA model considering variable return-to-scale is used to evaluate the efficiency of the companies in Tehran Stock Exchange. Table4 presented the efficiency scores for 2013. According to data presented in Table4, among 23 companies, 9 companies are efficient. Seventy-nine percent of inefficient companies have an efficiency lower than 0.6. Transportation Company with an efficiency of 0.03 is the most inefficient company.

In order to suggest practical benchmark for increasing the efficiency of inefficient companies, reference set (linear combination of efficient companies) is used. Each reference set includes efficient DMUs which can construct efficient projection of the associated inefficient DMU. According to Table 3, the Tile Company and the multi Investment Company have the highest number of presence in the reference sets of all inefficient companies. For instance, the inefficient company of Computer company can follow methods of Tile Company, multi Investment Company, and car company in the selection of inputs and outputs in order to be projected toward efficient frontier.

On the basis of efficiency scores reported in Table 4, and using reference sets, one can project the inefficient DMUs toward efficient frontier. This can be assumed as practical benchmark for inefficient DMUs.

Tables 5-7 represents the efficiency scores and fraction of total investment in each company. For instance, as shown in Table 5, the portfolio of investment includes 50 percent of DMU10 and 50 percent of DMU23. The efficiency scores of both investments are equal to unit.

Table 4. Efficiency Score for 2013

DMU	Company	Efficiency	Reference units		
DMU01	Sugar and Candy	1	DMU01		
DMU02	Other non-metallic products	1	DMU02		
DMU03	Tile	1	DMU03		
DMU07	Medicinal Investment	1	DMU07		
DMU10	Multi industrial Investment	1	DMU10		
DMU19	Chemical Investment	1	DMU19		
DMU20	Car company	1	DMU20		
DMU21	Basic metals	1	DMU21		
DMU23	Bank	1	DMU23		
DMU09	computer	0.7	DMU03	DMU10	DMU20
DMU11	Extraction of minerals	0.62	DMU01	DMU02	DMU10
DMU14	Drilling Company	0.6	DMU03	DMU10	DMU20
DMU08	Manufacturing Company	0.56	DMU03	DMU10	DMU20
DMU06	Food products except sugar	0.51	DMU01	DMU02	DMU10
DMU05	Investment	0.47	DMU01	DMU02	DMU10
DMU04	cement and lime company	0.4	DMU01	DMU02	DMU10
DMU15	mediators	0.3	DMU01	DMU03	DMU20
DMU12	Machinery	0.29	DMU03	DMU10	DMU20
DMU17	Telecommunication	0.29	DMU03	DMU10	DMU20
DMU22	Technical and engineering	0.25	DMU03	DMU10	DMU20
DMU13	Mass, Real Estate	0.22	DMU01	DMU03	
DMU16	Insurance Company	0.15	DMU01	DMU03	
DMU18	Transportation	0.03	DMU01	DMU03	

Table 8 presented the efficiency scores for 2014. As can be seen in Table 8, the efficiency of 6 companies is equal to 1. These companies were also efficient in 2013 on the basis of data presented in Table 4.

Table 5. First Cluster: Efficiency Score and Investment Percentage in 2013

DMU	Company	Efficiency	Percentage
DMU10	Multi industrial Investment	1	50.0%
DMU23	Bank	1	50.0%

Table 6. Second Cluster: Efficiency Score and Investment Percentage in 2013

DMU	Company	Efficiency	Percentage
DMU21	Basic metals	1	3.8%
DMU20	Car company	1	12.8%
DMU19	Chemical Investment	1	24.4%
DMU23	Bank	1	3.8%
DMU10	Multi industrial Investment	1	6.4%
DMU09	computer	0.7	1.3%
DMU11	Extraction of minerals	0.62	2.6%
DMU06	Food products except sugar	0.51	3.8%
DMU04	cement and lime company	0.4	1.3%
DMU15	mediators	0.3	1.3%
DMU17	Telecommunication	0.29	2.6%
DMU12	Machinery	0.29	6.4%
DMU22	Technical and engineering	0.25	25.6%
DMU13	Mass, Real Estate	0.22	1.3%
DMU16	Insurance Company	0.15	1.3%
DMU18	Transportation	0.03	1.3%

Table 7. Third Cluster: Efficiency Score and Investment Percentage in 2013

DMU	Company	Efficiency	Percentage
DMU01	Sugar and Candy	1	0.8%
DMU02	Other non-metallic products	1	1.3%
DMU03	Tile	1	0.2%
DMU07	Medicinal Investment	1	4.9%
DMU10	Multi industrial Investment	1	4.7%
DMU19	Chemical Investment	1	17.6%
DMU20	Car company	1	8.0%
DMU21	Basic metals	1	8.4%
DMU23	Bank	1	16.5%
DMU09	computer	0.7	1.4%
DMU11	Extraction of minerals	0.62	3.6%
DMU14	Drilling Company	0.6	1.2%
DMU08	Manufacturing Company	0.56	1.4%
DMU06	Food products except sugar	0.51	3.3%
DMU05	Investment	0.47	0.5%
DMU04	cement and lime company	0.4	2.6%
DMU15	mediators	0.3	2.5%
DMU12	Machinery	0.29	4.1%
DMU17	Telecommunication	0.29	4.8%
DMU22	Technical and engineering	0.25	6.8%
DMU13	Mass, Real Estate	0.22	2.0%
DMU16	Insurance Company	0.15	2.5%
DMU18	Transportation	0.03	0.9%

Table 8. Efficiency Score for 2014

DMU	Company	Efficiency	Reference units		
DMU01	Sugar and Candy	1	DMU01		
DMU02	Other non-metallic products	1	DMU02		
DMU03	Tile	1	DMU03		
DMU07	Medicinal Investment	1	DMU07		
DMU10	Multi industrial Investment	1	DMU10		
DMU19	Chemical Investment	1	DMU19		
DMU20	Car company	0.92	DMU10	DMU19	
DMU21	Basic metals	0.86	DMU02	DMU07	DMU10
DMU23	Bank	0.72	DMU03	DMU10	DMU19
DMU09	Computer	0.66	DMU03	DMU10	DMU19
DMU11	Extraction of minerals	0.59	DMU02	DMU03	DMU10
DMU14	Drilling Company	0.54	DMU03	DMU10	DMU19
DMU08	Manufacturing Company	0.53	DMU03	DMU10	DMU19
DMU06	Food products except sugar	0.52	DMU02	DMU03	DMU10
DMU05	Investment	0.48	DMU03	DMU10	DMU19
DMU04	cement and lime company	0.44	DMU02	DMU03	DMU19
DMU15	mediators	0.42	DMU03	DMU10	DMU19
DMU12	Machinery	0.25	DMU03	DMU10	DMU19
DMU17	Telecommunication	0.23	DMU01	DMU03	DMU19
DMU22	Technical and engineering	0.17	DMU03	DMU19	
DMU13	Mass, Real Estate	0.15	DMU03		
DMU16	Insurance Company	0.08	DMU03	DMU19	
DMU18	Transportation	0.03	DMU01	DMU03	

Table 9. First Cluster: Efficiency Score and Investment Percentage in 2014

Cluster	DMU	Company	Efficiency	Percentage
cluster 1	DMU10	Multi industrial Investment	1	50.0%
	DMU23	Bank	0.72	50.0%

Table 10. Second Cluster: Efficiency Score and Investment Percentage in 2014

Cluster	DMU	Company	Efficiency	Percentage
cluster 2	DMU19	Chemical Investment	1	24.4%
	DMU10	Multi industrial Investment	1	6.4%
	DMU20	Car company	0.92	12.8%
	DMU21	Basic metals	0.86	3.8%
	DMU23	Bank	0.72	3.8%
	DMU09	Computer	0.66	1.3%
	DMU11	Extraction of minerals	0.59	2.6%
	DMU06	Food products except sugar	0.52	3.8%
	DMU04	cement and lime company	0.44	1.3%
	DMU15	mediators	0.42	1.3%
	DMU12	Machinery	0.25	6.4%
	DMU17	Telecommunication	0.23	2.6%
	DMU22	Technical and engineering	0.17	25.6%
	DMU13	Mass, Real Estate	0.15	1.3%
	DMU16	Insurance Company	0.08	1.3%
	DMU18	Transportation	0.03	1.3%

Table 11. Third Cluster: Efficiency Score and Investment Percentage in 2014

Cluster	DMU	Company	Efficiency	Percentage
cluster 3	DMU01	Sugar and Candy	1	0.8%
	DMU02	Other non-metallic products	1	1.3%
	DMU03	Tile	1	0.2%
	DMU07	Medicinal Investment	1	4.9%

Cluster	DMU	Company	Efficiency	Percentage
	DMU10	Multi industrial Investment	1	4.7%
	DMU19	Chemical Investment	1	17.6%
	DMU20	Car company	0.92	8.0%
	DMU21	Basic metals	0.86	8.4%
	DMU23	Bank	0.72	16.5%
	DMU09	Computer	0.66	1.4%
	DMU11	Extraction of minerals	0.59	3.6%
	DMU14	Drilling Company	0.54	1.2%
	DMU08	Manufacturing Company	0.53	1.4%
	DMU06	Food products except sugar	0.52	3.3%
	DMU05	Investment	0.48	0.5%
	DMU04	cement and lime company	0.44	2.6%
	DMU15	mediators	0.42	2.5%
	DMU12	Machinery	0.25	4.1%
	DMU17	Telecommunication	0.23	4.8%
	DMU22	Technical and engineering	0.17	6.8%
	DMU13	Mass, Real Estate	0.15	2.0%
	DMU16	Insurance Company	0.08	2.5%
	DMU18	Transportation	0.03	0.9%

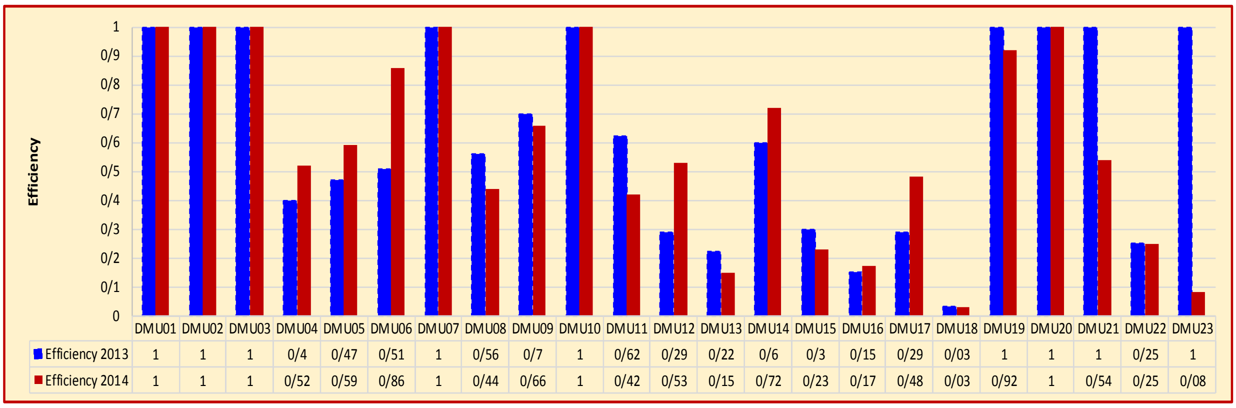

Figure 8. Efficiency Scores of 2013-2014

Figure 8 presents the efficiency scores of DMUs in both 2013 and 2014. Figure 8 prepare a suitable structure in order to compare the situation of a DMU during time period 2013-2014.

Figure 8 shows that DMU23 and DMU21 have a major reduction of efficiency (%92 and %46 compared to their efficiency in 2013). DMU06 have a %35 growth of efficiency compared to its efficiency in 2013.

5. Conclusion

In this study, a hybrid procedure on the basis of clustering analysis and DEA was proposed to investigate the selling and buying behavior of investors. The whole procedure was applied in some financial records in Tehran stock exchange. The main steps of the proposed procedure are as follows. In the first stage, a pre-screening method was accomplished on data. Then, a clustering approach was conducted on the basis of 25 variables to investigate the main clusters of selling and buying records. Finally, DEA was used to measure the efficiency score of each cluster. The results were analyzed on the basis of the financial data of the customers in Tehran stock exchange in 2013 and 2014. The efficient and inefficient DMUs (companies) were determined on the basis of efficiency scores in each buying and selling clusters. The reference set of each inefficient DMU was proposed in order to achieve the projection toward efficient frontier. The reference set can help the managers of inefficient companies to move toward best benchmarks in the market. In future studies, more clustering approaches can be considered. Uncertainty in data and clusters as well as fuzziness in inputs and outputs can be modeled through fuzzy clustering approaches and fuzzy DEA modeling.

References

Witten I. Frank E., and Hall M. (2011), “Data Mining: Practical Machine Learning Tools and Techniques”, 3rd Edition, Morgan Kaufmann Publishers.

Papanastasopoulos, G., Thomakos, D., and Wang, T. (2010). The implications of retained and distributed earnings for future profitability and stock returns, Review of Accounting and Finance, 9(4): 395-423.

Baslto, N-Bellotti,R-De cardo,F-FFacchi,P-Pascasio,S.(2005).Clustering Stock market companies via chaotic map synchronization.

Doherty,K-adams,R-Dawey,N-Pensuwon,W.(2005). Hierarchical topological clustering learns stock market sectos,ICSC congess on computational intelligence method & applications.Istanbul.

Nanda,D-Mahanty, B-Tiwan,M. (2010), Clustering Indian stock market data for portfolio management.

Does,C.& Cincotti,S.(2005).Clustering of financial time series with application to index and enhanced index tracking portfolio, Physical A: statistical Mechanics and irs Application, vol 355. [7] Yuqing Song, Yong Shi ,Aidong Zhang.(2005), A Shrinking-Based Clustering Approach for Multidimensional Data, IEEE Transactions on Knowledge and Data Engineering archive, Volume 17 Issue 10, Page 1389-1403

Jain,A,K,Musry,M.N,Flynn,P.J, (1999),Data clustering: A review. ACM computing sur vegs,31.

Farrell, M.J. (1957). The Measurement of Productive Efficiency. Journal of the Royal Statistical Society (A, general), 120: 253–281.

Charnes, A., Cooper, W.W., Rohdes, E. (1978) "Measuring the Efficiency of Decision Making Units", European Journal of Operational Research, 2(6): 429-444.

Banker, R. D., Charnes, A., Cooper, W.W., (1984): "Some Models for Estimation Technical and Scale in Data Envelopment Analysis", Management Science, 30 (9): 1078-1092.

Edirisnghe and X. Zhan, (2008), ‘Portfolio selection under Dea-based relative financial strength indicators: case of US industries’.Journal of Operational Research Society 59.pp:842-856.

Jelena Titko a, Daiva Jureviciene. (2014), DEA Application at Cross-country Benchmarking: Latvian vs Lithuanian Banking Sector, The 2-dn International Scientific conference „Contemporary Issues in Business, Management and Education 2013“,Volume 110, 24 January 2014, Pages 1124-1135.

Sungmook Lima, Kwang Wuk Oha, Joe Zhub. (2014), Use of DEA cross-efficiency evaluation in portfolio selection: An application to Korean stock market, European Journal of Operational Research, Volume 236, Issue 1, 1 July 2014, Pages 361–368.

[15] Veronika Fenyves, Tibor Tarnóczi, Kinga Zsidó. (2015), Financial Performance Evaluation of Agricultural Enterprises with DEA Method, Procedia Economics and Finance, Volume 32, 2015, Pages 423-431.

[16] Khanjamali, Hassan. (2006). A study of relative efficiency of pricing in stock exchange in industries by using DEA, MA thesis, Shahid Beheshti University, Faculty of Manageemnt.

Khajavi, Shokrollah; Salimifard, Alireza; & Rabi'eh, Mas'od. (2005). Using DEA to find portfolio of most efficient companies accepted in Tehran Stock Exchange, Journal of Social Sciences and Humanities, Shiraz University, issue 22, p 75-88.

Ahmadzadeh, Mas'od. (2010). Using DEA to select shares and comparing it with size-effect strategy, MA thesis, Imam Khomeini International University, Department of Industrial Management.

Brigham, Eugene; Houston, Joel. (2007), Fundamentals of Financial Management, Eleventh Edition. Thomson South-Western.

[20] Peter Wankea, C.P. Barrosb, Ali Emrouznejad, (2016) , Assessing productive efficiency of banks using integrated Fuzzy-DEA and bootstrapping: A case of Mozambican banks, European Journal of Operational Research, Volume 249, Issue 1, 16 February 2016, Pages 378–38

Xiaohui Huang, Yunming Yeb, Liyan Xionga, Raymond Y.K. Lauc, Nan Jianga, Shaokai Wang, (2016) , Time series k-means: A new k-means type smooth subspace clustering for time series data, Information Sciences, Volumes 367–368, 1 November 2016, Pages 1–13.

Rajiv D. Banker, William W. Cooper, Lawrence M. Seiford, Robert M. Thrall, Joe Zhu, (2004), Returns to scale in different DEA models, European Journal of Operational Research, Volume 154, Issue 2, 16 April 2004, Pages 345-362.

Jolliffe I.T., (2002), Principal Component Analysis, Springer-verlag New York, Second edition.

Zahra Mashayekhi, Hashem Omrani,(2016), An integrated multi-objective Markowitz–DEA cross-efficiency model with fuzzy returns for portfolio selection problem, Applied Soft Computing, Volume 38, January 2016, Pages 1–9

Chao Huang, Chong Dai, Miao Guo, (2015),A hybrid approach using two-level DEA for financial failure prediction and integrated SE-DEA and GCA for indicators selection, Applied Mathematics and Computation, Volume 251, 15 January 2015, Pages 431-441

N.C.P. Edirisinghe, X. Zhan, (2007), Generalized DEA model of fundamental analysis and its application to portfolio optimization, Journal of Banking & Finance, Volume 31, Issue 11, November 2007, Pages 3311-3335

Eero Pätäri, Timo Leivo, Samuli Honkapuro ,(2012), Enhancement of equity portfolio performance using data envelopment analysis, European Journal of Operational Research, Volume 220, Issue 3, 1 August 2012, Pages 786-797.

HOW TO CITE THIS ARTICLE
Molla Hosseinagha, S., Khalili-Damghani, K. (2021). A hybrid Model on the Basis of Data Envelopment Analysis and Data Mining Techniques to Analyze the Investment Behavior in Stock Exchange: A Real Case Study in Tehran Stock Exchange. International Journal of Finance, Accounting and Economics Studies, 3(1): 52-67.
DOI:
Url: **https://ijfaes.srbiau.ac.ir/article_16618.html**
Journal homepage: **https://ijfaes.srbiau.ac.ir**