

Research Article

Analyzing the Investment Behavior in the Iranian Stock Exchange during the COVID-19 Pandemic Using Hybrid DEA and Data Mining Techniques

Amir Homayoun Sarfaraz,¹ Amir Karbassi Yazdi ,² Thomas Hanne ,³ Özaydin Gizem,⁴ Kaveh Khalili-Damghani,¹ and Saiedeh Molla Husseinagha¹

¹School of Industrial Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

²Young Researchers and Elite Club, South Tehran Branch, Islamic Azad University, Tehran, Iran

³Institute for Information Systems, University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

⁴Social Sciences Institute, Izmir Katip Celebi University, Izmir, Turkey

Correspondence should be addressed to Thomas Hanne; thomas.hanne@fhnw.ch

Received 6 September 2021; Revised 1 December 2021; Accepted 10 December 2021; Published 28 January 2022

Academic Editor: Jianxu Liu

Copyright © 2022 Amir Homayoun Sarfaraz et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The main purpose of this paper is to investigate the effects of COVID-19 regarding the efficiency of industries based on data in the Tehran stock market. A hybrid model of Data Envelopment Analysis (DEA) and data mining techniques is used to analyze the investment behavior in Tehran stock market. Particularly during the COVID-19 pandemic, many companies face financial crises. That is why companies with inferior performance must be benchmarked with efficient companies. First, the financial data of investments on selective companies are analyzed using data mining approaches to recognize the behavioral patterns of investors and securities. Second, customers are clustered into 3 selling and 4 buying groups using data mining techniques. Then, the efficiency of active companies in stock exchange is evaluated using input-oriented DEA. The results indicate that, among 23 industries listed on the stock market in Iran, solely nine were efficient in 2019. Moreover, in 2020, the number of efficient industries further decreased to six industries. Comparing the obtained results with those of another study which was conducted in 2018 by other researchers revealed that COVID-19 strongly affects the performance of an industry and some industries which were efficient in the past such as the bank industry became inefficient in the following year.

1. Introduction

Creating a suitable environment for investors to invest in stock markets is essential for economic progress [1]. Stable economic conditions usually lead to a predictable environment, and all these will persuade investors for a more reliable investment [2]. The investment market plays an important role in the allocation of financial sources to companies in developing countries [3]. Hence, the identification of key criteria for a safe investment is critically important. Processing big data of past investments in a specified field can be very useful for making efficient investment decisions. Data mining is one of these processing methods [4]. Stock exchanges, which can provide large amounts of historical data related to previous investments,

are a suitable resource for applying data mining approaches [5]. Today stock exchange markets are rapidly growing. Every day, a large number of shares are traded, and huge amounts of data result. 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 improved pricing efficiency of shares [6]. Nevertheless, investors and other market participants usually assume that a random selection of shares is not a useful strategy. It is very important for us to know which factors have an impact on the behavior of investors. Another issue that has created some problems in Iranian investment markets is the

deficiencies in analyses. A lot of experts and investors believe that price divided by earnings 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 based on a limited number of models, many people simultaneously focus their investments on a small number of companies or industries and disinvest from other industries at the same time. This intensifies problematic behaviors in the market. When there are a lot of fluctuations in economic variables, the profitability of companies can also be significantly affected by these changes. In such a situation, investors do not trust the published data. Besides, operational conditions and efficiency might have an influence on share prices. Traina [7] showed that rates in the financial statements have an impact on share prices. In other words, figures in the financial statements show the performance and efficiency of the companies. Investment in efficient companies produces higher returns. Here, the main problem is that the evaluation of the efficiency of companies with various criteria and indices is not an easy task. In addition, decision-making based on a limited number of indicators is insufficient. Although the evaluation of companies by financial statements is difficult, DEA models can help incorporate several criteria and indicators when assessing the companies [8, 9]. The data are gathered from the Tehran stock exchange. The efficiency of companies whose shares have been bought is determined and compared.

Since the emergence of COVID-19 in 2019, many aspects of our lives have been affected. Many countries were forced to lock down, and subsequently, many industries did not work. The stock market values of many industries decreased dramatically, and many people lost their work. This research can help to find out which companies in the stock market of Iran are efficient and which of them are not. In addition, inefficient companies can benefit from benchmarking with efficient companies.

Data mining is a technique for extracting patterns and also finding correlations among big data sets for prediction purposes [10]. This technique can be applied for increasing revenue, customer satisfaction and relationship, reducing costs and risks, and so on. Data mining includes many methods. One of them is clustering, which is based on the idea of putting similar things into the same cluster.

There are many methods for evaluating the performance of companies or businesses such as the Balanced Scorecard or EFQM (European Foundation for Quality Management). One of the most popular methods for performance evaluation of companies is Data Envelopment Analysis (DEA), which is based on the linear programming method [11]. In this method, Decision-Making Units (DMUs) are evaluated by input and output factors. This method consists of the variants BCC by Banker et al. [12] and CCR by Charnes et al. [13].

COVID-19 has spread rapidly in many countries, including Iran, with a strongly increasing number of cases since the end of February 2020. From the beginning of this pandemic, all public places were closed in Iran, and social distancing and home quarantine were encouraged. Due to US sanctions against Iran and a resulting fragile economy,

the government could not shut down the country. Hence, the number of patients and death rates increased dramatically. The government reduced the weekly work hours in government organizations. Most of the private companies and factories could not maintain production levels. In some cases, the government ordered a shutdown in the capital of Iran and some other cities for one week. These issues led to a decreased performance of companies.

Several studies conducted in Iran investigate the effects of the COVID-19 pandemic. Ahmadi and Ramezani [14] studied the COVID-19 effects from an emotion-focused therapy perspective. Samadi et al. [15] used Wavelet Coherence Analysis to examine the comovements between markets in a time period from September 2014 to June 2020 as a period of intense uncertainty in Iran. According to our search, no study was conducted to assess the effect of COVID-19 on the Iran stock market.

The stock market of Iran is the main source of pricing the most important goods such as cement and other commodities and for allocating capital for companies. Besides, many people invest their money, and if stocks decrease abnormally, many people lose their properties. Therefore, the stock market of Iran plays a key role for both companies and private persons.

There are a lot of papers published not only about stock markets worldwide but also about the Iran stock market using diverse methods such as MCDM or DEA. However, previous studies on the Iranian stock market were carried out in normal situations, so that the following question remains: Does the pandemic have a significant effect on the Iran stock market?

The novelty of this paper is using a hybrid technique for clustering and efficiency assessment. We examine the performance of Iran's stock market during the COVID-19 pandemic and investigate the efficiency of Iranian companies in the stock market.

First, well-known data mining approaches such as clustering methods are used to investigate the transaction behavior of investors using 30 criteria. Afterward, the 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 in the stock markets in Iran based on their transactions. The second part includes the evaluation of the efficiency of active companies in the 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 review of past research works is provided. 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

2.1. Studies Related to Efficiency Assessment in Stock Markets. In this section, we review a number of studies employing neural networks, clustering and DEA, and other data mining approaches to stock market analysis.

Hiransha et al. [16] used novel deep learning architectures such as Multilayer Perceptrons (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) Networks, and Convolutional Neural Networks (CNN) to predict stock prices of companies on the National Stock Exchange of India and the New York Stock Exchange. As a result, the Convolutional Neural Network (CNN) method performed better than other methods due to its ability to catch sudden changes in the network during the forecasting phase.

Montenegro and Molina [4] started from the fact that the neural networks appeared superior to other methods in modeling nonlinear data and worked on the prediction of the day-to-day stock value of the S&P 500 Index. By creating a data set from the daily market activity values of the stocks between June 7, 2013, and June 6, 2018, for each company in the S&P 500 Index, the Deep Learning Neural Network method was used in the training of the network and the Feature Selection Analysis method in determining the behavioral tendencies of the companies. Thus, this paper helped decision-makers to improve investment behavior.

Yazdi et al. [17] measured the performances of 25 insurance companies, four of which were public and the rest were private, in Iran during the 2014–2015 period. They developed and used a new approach based on Data Envelopment Analysis (DEA) together with fuzzy clustering. In the study, which considers employees, capital, and total assets as input variables and total costs, paid compensation, and profit as output variables, it is concluded that ten companies belong to an efficient cluster with the output-oriented CCR method. Similar to our study, the authors analyzed DMUs (banks and companies listed at the stock market) using DEA and clustering methods.

Khedr et al. [5] estimated the behavior in the stock market by using news sentiment analysis and data mining techniques, considering a data set containing the opening, high, low, and closing (OHLC) prices of the stocks of three companies traded in the NASDAQ market, and three news articles per day about the market and the company. According to the results obtained, a strong correlation was found between the news classified as positive or negative and the fluctuations in stock prices. In addition, it has been observed that using the opening, high, low, and closing prices of stocks increases the accuracy of the market behavior prediction up to 89.80%. The used method is similar to other studies, including our investigation, as a data mining approach is applied.

Karimi and Barati [18] evaluated the financial performances of 72 companies selected from the automotive, pharmaceutical, petrochemical, and cement sectors traded on the Tehran stock exchange using the negative data-limited adjusted measure in DEA [19]. In this study, financial ratios such as assets turnover ratio, quick ratio, and current ratio serve as input criteria, whereas earnings per share ratio, return on assets ratio, net profit margin ratio, and so on are output criteria. After all, fifty-eight companies were determined as efficient, and the rest were determined as inefficient companies; then, efficient companies were ranked using the Andersen and Petersen model. In contrast to our

study, this paper did not cluster industries, and the study is conducted under normal conditions, that is, without the effects of COVID-19 on the efficiency of companies.

Anouze and Bou-Hamad [20] considered MENA countries in the Middle East and North Africa and investigated the effects of 151 banks on environmental factors and performance in the years 2008–2010. They used data envelopment analysis and statistical data mining techniques such as classification and regression trees (CART), conditional inference trees (CIT), random forest based on CART and CIT, bagging, Artificial Neural Networks, and logistic regression. In the study, fixed assets, deposits, equity, and personnel expenses are input variables. Loans, net income, and liquid sources are determined as output variables. In summary, it was concluded that random forest and bagging methods are better statistical tools than other methods in measuring bank performance by using the importance rankings of the variables. The study used DEA and data mining methods, but there are differences in the used data mining techniques compared to our analysis.

Chang et al. [21] analyzed the portfolio performance valuation of 44 textile companies in Taiwan from 2011 to 2018 using the nested dynamic network (NDN) Data Envelopment Analysis (DEA) method. They applied a two-stage approach to their work. First, the additive dynamic DEA method was applied to find efficient financial assets. Then, an evaluation was made with the NDN DEA model to evaluate the portfolios consisting of selected effective financial assets in periods of three, four, five, or even six years. Also, here, the DEA technique was employed, but with a different data mining method compared to our study.

Zhong and Enke [22] evaluated 60 financial and economic factors in the S&P 500 Index using clustering and classification as data mining methods for daily stock market return prediction. Within the scope of this evaluation, 60 factors in the data set consisting of 2518 trading days between June 1, 2003, and May 31, 2013, were classified using the Principal Component Analysis (PSA) method to find the most important and most effective main components. In the next stage, by obtaining 12 new data sets from all adjusted data, the daily direction of returns was estimated by an Artificial Neural Network (ANN) and logistic regression methods. It has been concluded that the ANN achieves higher accuracy than logistic regression and classification and cluster mining are important in reducing the size of the data and increasing its efficiency.

Rezaee et al. [23] applied integrated methods to estimate the online financial performance of companies when online data were taken from the Tehran stock exchange in 2007–2012 at different time intervals. The used methods are Dynamic Fuzzy C-Means for updating cluster and membership numbers, Data Envelopment Analysis (DEA) for evaluating companies using financial ratios, and finally Artificial Neural Networks (ANN) for predicting the future performance of companies. The difference to our approach is that we used the K-means method, while Rezaee et al. [23] used fuzzy C-means.

Mehlawat et al. [24] evaluated 20 risky assets with their credibility by using variance and Conditional Value at Risk

(CVaR) as risk measurement tools together with liquidity and entropy criteria. Within the scope of this evaluation, randomly selected portfolios with different sample sizes, including risk and entropy as input values and return and liquidity as output values, were examined by the DEA method. Moreover, these risky assets were combined with the fuzzy multipurpose portfolio model, and performance evaluation was made. In addition, the result of this paper indicated how a portfolio could be selected considering the risk of companies by using DEA and VaR.

Mashayekhi and Omrani [25] have proposed a new multipurpose model, which incorporated DEA cross-efficiency into the mean-variance Markowitz model to select an investment portfolio. 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 the Markowitz model and DEA.

Table 1 provides an overview of methods used in this field.

As mentioned above, stock market evaluation is one of the most popular subjects in financial analysis. The reason is that the stock market is the engine of economics in most countries. In some related studies such as ours, the DEA method is used solely or in combination with data mining or Artificial Intelligence (AI) approaches. In DEA, the DMUs are evaluated and categorized by data analysis. In other methods, the regression shows the relationship between stock markets and factors that affect them. Some other approaches use AI to predict the performance of companies in the stock market. In our study, the effects of the COVID-19 pandemic as one of the substantial issues that affected economies worldwide and especially stock markets are evaluated. In particular, it is analyzed which industries are efficient. Inefficient industries should be benchmarked with efficient industries in order to improve.

2.2. Studies on Economic Effects of COVID-19. In the following, we provide a brief overview of studies related to the economic consequences of the COVID-19 pandemic in different countries.

Caraka et al. [27] studied the effects of COVID-19 on the environment and the economy of Indonesia. The conducted statistical analysis and results show that COVID-19 has a significant effect on the economy of Indonesia. Albu et al. [28] used a logistic model to predict the effects of the pandemic on the economy of Romania. Based on their simulation, pandemic evolution was classified into four distinct phases. Three scenarios were considered to estimate the economic impact of the epidemic at three levels. Results showed that, in the long term, an economic program based

on large investment could contribute to restoring growth levels both worldwide and in the case of the EU countries.

Grima et al. [29] studied the previous mechanisms which were applied to provide an understanding of the challenges related to GDP. A simple statistical analysis was used by adopting data that were collected from government websites, online statistics, published reports, trends, and internal data. It is mentioned in the study that the research will help risk managers and leaders to understand the devastating social and economic impact of such disruptions and act proactively to avoid repetition and the negative effects of being unprepared.

Thorbecke [30] studied the impact of COVID-19 on the United States economy. Stock returns for 125 sectors are considered during the COVID-19 crisis in this study. The paper investigates how both the macroeconomic environment and sector-specific factors affect returns. Several macroeconomic variables were used in this study. With the help of a regression technique, estimation equations were used to predict the stock price index. Finally, the study discusses that stock prices are useful because they provide a measure of how investors expect shocks to impact future cash flows across sectors. Results show that sectors impaired by idiosyncratic factors include airlines, aerospace, real estate, tourism, oil, brewers, retail apparel, and funerals.

3. Research Methodology

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

3.1. Main Financial Indicators. In order to investigate the stock exchange, some financial indicators are required. Consulting a number of experts, who have enough experience in the related field, yielded to identifying the most essential financial indicators. They are defined as follows (see, e.g., [31, 32]):

$$\begin{aligned} \text{total assets} &= \text{current assets} + \text{goods} + \text{prepayments} \\ &\quad + \text{fixed assets} + \text{intangible assets} + \text{long} \\ &\quad - \text{term investments} + \text{amortization}, \end{aligned} \quad (1)$$

$$\text{total debts} = \text{current debts} - \text{long term debts}.$$

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

$$\begin{aligned} \text{operational profit (loss)} &= \text{general profits} - \text{expenses of distribution and sale} \\ &\quad - \text{general and administrative expenses net incomes and operational costso.} \end{aligned} \quad (2)$$

TABLE 1: Overview of previous studies.

| Author(s) | Method(s) | Country(s) | Duration |
|---------------------------|---|----------------|---|
| Rezaee et al. [23] | Artificial Neural Networks and DEA | Iran | 2007–2012 |
| Zhong and Enke [22] | Principal Component Analysis, Artificial Neural Networks, and logistic regression | S&P 500 Index | June 1, 2003, and May 31, 2013 |
| Chang et al. [21] | Nested dynamic network and DEA | Taiwan | 2011–2018 |
| Anouze and Bou-Hamad [20] | Classification and regression trees and conditional inference trees | MENA countries | 2008–2010 |
| Karimi and Barati [18] | DEA | Iran | 2014 |
| Khedr et al. [5] | News sentiment analysis and data mining technique | NASDAQ | Did not mention the year |
| Asadifar and Kahani [26] | Intelligent data mining | Iran | Data gathering in 1 month; did not mention the year |
| Yazdi et al. [17] | DEA and fuzzy clustering | Iran | 2014–2015 |
| Hiransha et al. [16] | Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN) | India and USA | 1996–2015 |

Return of equity shows the efficiency of management of a company in employing the resources to obtain profits. It measures the rate of return on the owners' investments. Return on equity is calculated using

$$\text{return of equity} = \frac{\text{profits and losses after reducing taxes}}{\text{shareholder equity}} * 100. \quad (3)$$

Return on assets is calculated by dividing the net income of the company (income after taxes) by total assets as shown in

$$\text{return on assets} = \frac{\text{profits and losses after reducing taxes}}{\text{total assets}} * 100. \quad (4)$$

3.2. Methods

3.2.1. Clustering Approach. The first phase of the proposed approach is the 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 [33]. One of the main problems in clustering is to determine the number of clusters. There are several methods for determining the number of clusters. In this study, Wilk's lambda method [34, 35] as shown in (5) is used to determine the suitable number of clusters:

$$\text{Wilk's lambda} = \frac{SS_{\text{within}}^k}{SS_{\text{Total}}}, \quad k = 2, \dots, k_{\text{max}}. \quad (5)$$

SS_{within}^k is the variance within a cluster and SS_{Total} is the total variance. If the diagram of Wilk's lambda coefficient is drawn on the basis of k (cf. Figure 1), the first jump in the diagram is the optimal number of clusters. Based on the data gathered from the Tehran stock exchange, the following variables are used to cluster the customers:

- (i) Online buying and selling of shares
- (ii) Time of buying and selling, which is divided into four periods (i.e., the company and industry whose shares have been sold)

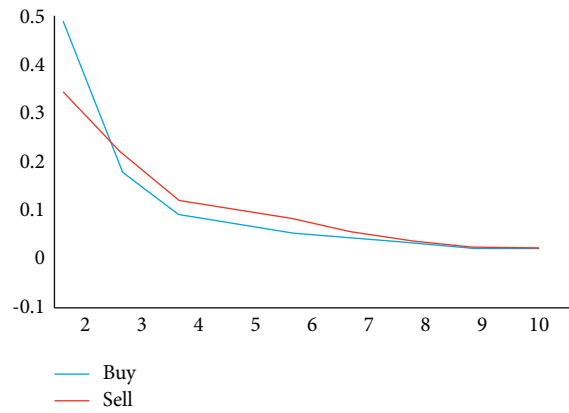


FIGURE 1: Number of clusters versus Wilk's lambda coefficient.

- (iii) The amount of sold or bought shares
- (iv) Value of transactions, which is equal to the volume of transactions made by means of share prices on that day

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

- (i) Online buying and selling of shares
- (ii) Time of buying and selling, which is divided into four periods (i.e., the first six months of 2019, the second six months of 2019, the first six months of 2020, and the second six months of 2020)
- (iii) Company and industry whose shares have been sold
- (iv) The amount of sold or bought shares
- (v) Value of transactions, which is equal to the volume of transactions made by mean of share prices on that day [36–38]

3.2.2. Analyzing the Efficiency Scores. Farrell [39] used a method on the basis of estimating production functions to measure the technical efficiency in a manufacturing company incorporating one input and one output. Farrell [39] also used

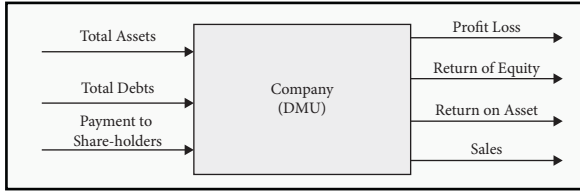


FIGURE 2: Schematic view of a company as a DMU.

the approach to evaluate the efficiency of the farming industry in the USA compared to the other countries. In 1978, Charnes et al. [13] developed Farrell's idea and presented a model that had the ability to measure efficiency in the presence of several inputs and outputs. The model proposed by Charnes et al. [13] was the first official Data Envelopment Analysis (DEA) model. In 1984, Banker et al. [12] extended the proposed model in the presence of variable return to scale assumptions, called the BCC model. The difference between the CCR and the BCC model is that the return to scale is constant in the CCR model, whereas in the BCC model, it is not.

As the companies are assumed as DMUs, they should be analyzed using the 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 the stock exchange, the indicators are divided into inputs (total assets, total debts, and payments to shareholders) and outputs (operational profits and losses, return of equity, return on assets, and sales). A company is shown in Figure 2 as a DMU.

The input-oriented DEA based on the variable return to scale assumption is used to evaluate the financial efficiency of companies. This model was first proposed by Banker et al. [12] and is formalized in (6). Let $x_j = (x_{1j}, \dots, x_{mj})$ and $y_j = (y_{1j}, \dots, y_{sj})$ be the input and output vectors of DMU j for $j = 1, 2, \dots, n$.

$$\begin{aligned} \text{Max } Z_p &= \frac{\sum_{r=1}^s u_r y_{rp} + w}{\sum_{i=1}^m v_i x_{ip}}, \\ \frac{\sum_{r=1}^s u_r y_{rj} + w}{\sum_{i=1}^m v_i x_{ij}} &\leq 1, \\ j &= 1, 2, \dots, n, \\ v_i &\geq 0, \\ i &= 1, 2, \dots, m, \\ u_r &\geq 0, \\ r &= 1, 2, \dots, s, \end{aligned} \quad (6)$$

w free on sign.

Here, p is the DMU being evaluated in the set of $j = 1, 2, \dots, n$. Z_p is the measure of the efficiency of DMU p . u_r and v_i are the weights assigned to output and input for solving the DEA model. w is the free variable. Model (6) is a

fractional mathematical programming problem, and its global optimum is hard to find. So, [12] proposed the following linear form:

$$\begin{aligned} \text{Max } Z_p &= \sum_{r=1}^s u_r y_{rp} + w, \\ \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} + w &\leq 0, \quad j = 1, 2, \dots, n, \\ \sum_{i=1}^m v_i x_{ip} &= 1, \\ u_r &\geq \varepsilon, \\ i &= 1, 2, \dots, m, \\ v_i &\geq \varepsilon, \\ r &= 1, 2, \dots, s, \\ w &\text{ free on sign.} \end{aligned} \quad (7)$$

In the BCC model, the sign of the variable W indicates the return to scale influence 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 the linear model (7), which is called the envelopment form, is specified as

$$\begin{aligned} \text{Min } w_p &= \theta - \varepsilon * \left(\sum_{r=1}^s s_r^+ + \sum_{i=1}^m s_i^- \right), \\ \sum_{j=1}^n \lambda_j x_{ij} + s_i^- &= \theta x_{ip}, \quad i = 1, 2, \dots, m, \\ \sum_{j=1}^n \lambda_j y_{rj} + s_r^+ &= y_{rp}, \quad r = 1, 2, \dots, s, \\ \sum_{j=1}^n \lambda_j &= 1, \lambda_j \geq 0, j = 1, 2, \dots, n, s_r^+ \geq 0, r = 1, 2, \dots, s. \end{aligned} \quad (8)$$

These models can be used to evaluate the financial performance of companies as DMUs, which were schematically depicted in Figure 2.

3.3. Research Procedure. In our study, the following steps are conducted:

- Step 1. Extract inputs and outputs.
 - Step 2. Select companies from population.
 - Step 3. Collect data based on inputs and outputs.
 - Step 4. Run the DEA method.
 - Step 5. Use the C-mean method for clustering the result of DEA.
 - Step 6. Benchmark inefficient clusters from efficient cluster.
- Figure 3 shows the procedure of research.

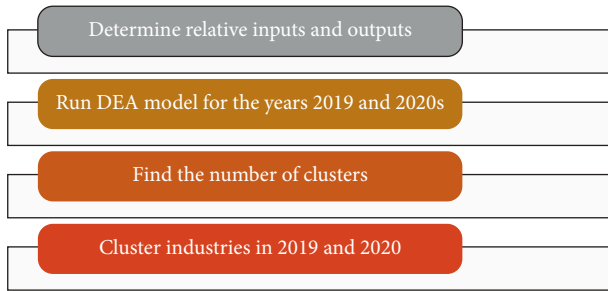


FIGURE 3: Research procedure.

3.4. Data Sample. The proposed approach of this study is applied to 214 companies in total, which are listed at the Tehran stock exchange and belong to 23 industries. The number of companies related to each industry is shown in Table 2. Data regarding the considered financial indicators are taken from published balance sheets of the companies. In addition, data of transactions of 56643 customers in 2019 and 2020 are gathered and used for our study.

4. Results and Discussion

After prescreening, 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 the Tehran stock exchange database. Excel and SPSS are used for preprocessing and analyzing the data. MATLAB is used for 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 1, the first jumps for buying and selling data occur for $k = 3$ and $k = 4$ clusters, respectively. So, the suitable number of clusters based on 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 the dimension of high-dimensional data on the basis of the direction of data dispersion [40, 41], is used to reduce the dimension of buying and selling data in favor of showing the results of clustering in a two-dimension plot. PCA is used for both buying and selling data. The associated two-dimensional clustering plots are shown in Figure 4.

4.3. Cluster Analysis Using the K-Means Method. Both buying and selling records are distinctively clustered using the 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 sell transaction from 2019 to 2020. Results of clustering are shown in Table 3. It is notable that the clustering is accomplished in a 25-dimensional

space as shown in Table 3, and buying and selling include 3 and 4 clusters, respectively. The variables which were considered for clustering the data are presented in Table 4.

In buying, the first buying-related cluster has the lowest number of members among the others. It also has the highest average of trading value. Reconsidering the customers included in this cluster, it has been recognized that they were corporate entities. The second buying-related cluster has a middle number of members compared to the other clusters. It is notable that the average of trading values is the lowest in comparison with the other clusters. The third buying-related cluster 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 the other clusters. More formally, the customers included in this cluster invest in low volume and high diversity.

Regarding selling, the first cluster has the lowest number of members and the highest average of trading value. The second cluster has the middle number of members and the lowest average of trading value. The third selling-related cluster has the highest number of members and the lowest average of trading value. An investigation of customers of this cluster reveals that they are normal persons with an average investment of 15617 USD.

The hit number and percentage of records in both buying and selling clusters are shown in Tables 2 and 5, respectively.

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 the fourth cluster of selling with 24503 records have the highest number of members among the clusters. This shows that the behaviors of customers in these clusters are similar. Figure 5 shows the mean of traded shares in each cluster. Figures 6–8 present transaction details of clusters 1–3. As can be seen in Figure 6, in the first cluster, investors have invested in two sectors (bank and multi industrial company). This cluster has the highest value of transactions and the lowest dispersion of investment in various industries. As can be seen in Figure 7, 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 8, in the third cluster, investors have invested in all 23 industries. Among these industries, bank and chemical products are most often traded in this cluster.

4.5. Results of Efficiency Measurement Using DEA in Each Cluster. In the prescreening phase, negative data are shifted using variable exchanges. Portela et al. [42] introduced a method for exchanging negative data in DEA with positive for solving the model. These variable exchanges are done by the following equations for both output and input variables.

First, consider (p) as the SP range, which measures the distance between a reference variable and the current variable. For output variables y_r , it is

TABLE 2: Number of people who buy or sell shares from various industries.

| Industry | No. of companies | Data of buying | | | Data of selling | | | Fourth cluster | |
|----------|--------------------------------------|----------------|----------------|---------------|-----------------|----------------|---------------|----------------|------|
| | | First cluster | Second cluster | Third cluster | First cluster | Second cluster | Third cluster | | |
| 1 | Telecommunication | 8 | 0 | 2 | 1520 | 3 | 20 | 0 | 1271 |
| 2 | Insurance | 6 | 0 | 1 | 800 | 0 | 6 | 0 | 599 |
| 3 | Machinery | 9 | 0 | 5 | 1295 | 4 | 26 | 1 | 1119 |
| 4 | Manufacturing of metallic products | 11 | 0 | 0 | 431 | 0 | 13 | 1 | 376 |
| 5 | Basic metals | 10 | 0 | 3 | 2651 | 0 | 56 | 1 | 2274 |
| 6 | Candy and sugar | 6 | 0 | 0 | 237 | 0 | 0 | 0 | 174 |
| 7 | Drilling | 7 | 0 | 0 | 363 | 0 | 12 | 0 | 337 |
| 8 | Transportation | 9 | 0 | 1 | 285 | 0 | 5 | 0 | 267 |
| 9 | Other nonmetallic products | 8 | 0 | 0 | 415 | 0 | 2 | 0 | 386 |
| 10 | Car | 12 | 0 | 10 | 2524 | 4 | 41 | 0 | 2287 |
| 11 | Tile | 9 | 0 | 0 | 71 | 0 | 1 | 0 | 82 |
| 12 | Technical and engineering services | 10 | 0 | 20 | 2141 | 16 | 80 | 3 | 1647 |
| 13 | Computer | 12 | 0 | 1 | 443 | 1 | 11 | 0 | 347 |
| 14 | Mass, real estate | 13 | 0 | 1 | 633 | 1 | 5 | 0 | 437 |
| 15 | Cement, lime | 8 | 0 | 1 | 830 | 2 | 7 | 0 | 731 |
| 16 | Chemical products | 9 | 0 | 19 | 5548 | 13 | 91 | 1 | 3657 |
| 17 | Medical products | 11 | 1 | 0 | 1540 | 1 | 12 | 0 | 1264 |
| 18 | Bank | 7 | 1 | 3 | 5184 | 0 | 53 | 1 | 3281 |
| 19 | Food products except sugar and candy | 12 | 0 | 3 | 1030 | 0 | 6 | 0 | 912 |
| 20 | Extraction of minerals | 6 | 0 | 2 | 1141 | 0 | 19 | 0 | 847 |
| 21 | Multi-industrial companies | 13 | 1 | 5 | 1491 | 4 | 20 | 2 | 1333 |
| 22 | Mediators | 8 | 0 | 1 | 774 | 0 | 8 | 0 | 726 |
| 23 | Investment | 10 | 0 | 0 | 160 | 0 | 1 | 0 | 149 |

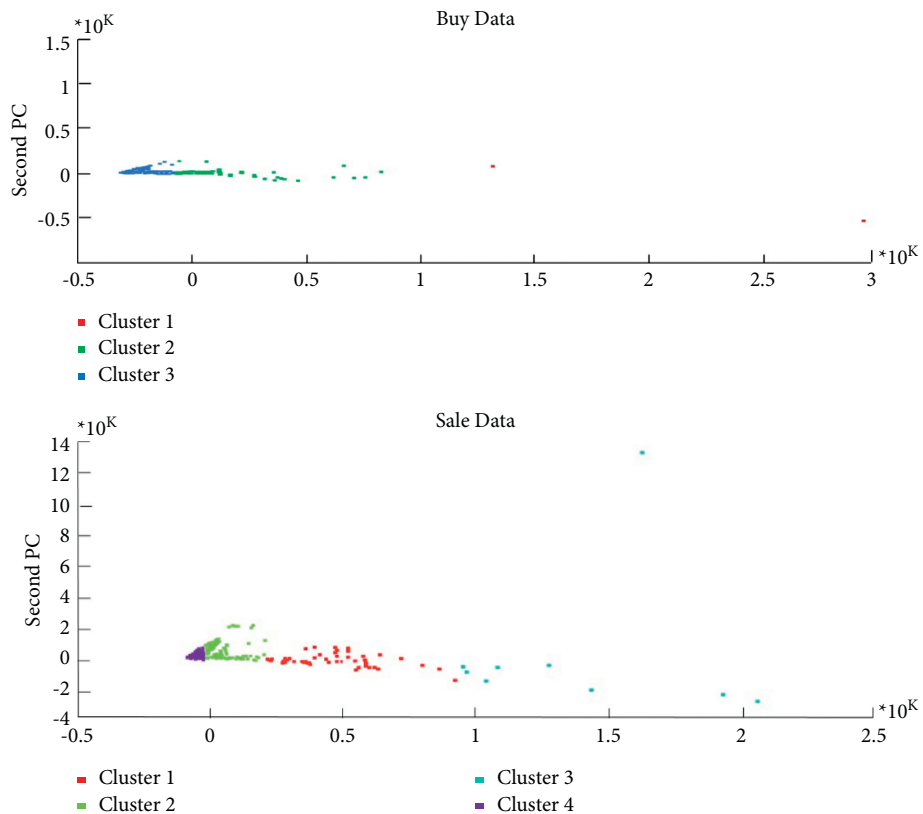


FIGURE 4: Data resulting from Principal Component Analysis.

TABLE 3: 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 | 13,050,000 | 77,125,415 |
| | | | | Second cluster | 78 | 2.9596 | 938,669 | 11,374,316 |
| | | | | Third cluster | 31,507 | 5.0798 | 17,542 | 142,916 |
| Selling | 0.78 seconds | 25056 | 1.9059 | First cluster | 49 | 1.0443 | 1,066,943 | 14,754,968 |
| | | | | Second cluster | 495 | 1.2555 | 292,086 | 2,610,905 |
| | | | | Third cluster | 9 | 1.3095 | 3,892,495 | 39,812,682 |
| | | | | Fourth cluster | 24503 | 1.6608 | 15,617 | 124,063 |

*Mean of transaction values, one US dollar = 42000 IRR

TABLE 4: Variables used for clustering.

| Variables used for clustering | |
|--------------------------------|--|
| EPS (earning per share) | P/E |
| DPS (dividend per share) | Total no. of transactions |
| Total sales | Percentage of capital increase |
| Total cost | Capital |
| Total profit | Equity share |
| No. of shares | Market share |
| No. of real person trades | Average no. of monthly trades |
| No. of legal person trades | Max. seasonal price increase |
| Market value | Max. monthly price increase |
| Average no. of seasonal trades | % of real person shareholders to legal person shareholders |
| Min. seasonal price decrease | % of monthly return |
| Min. monthly price decrease | % of yearly return |
| No. of staff | |

TABLE 5: Percentage of people who buy or sell shares from different industries with a special method in each cluster.

| | | Data of buying | | | Data of selling | | | |
|----|------------------------------------|----------------|----------------|---------------|-----------------|----------------|---------------|----------------|
| | | First cluster | Second cluster | Third cluster | First cluster | Second cluster | Third cluster | Fourth cluster |
| 1 | Telecommunication | 0 | 2.6 | 4.8 | 6.1 | 4 | 0 | 5.2 |
| 2 | Insurance | 0 | 1.3 | 2.5 | 0 | 1.2 | 0 | 2.4 |
| 3 | Machinery | 0 | 6.4 | 4.1 | 8.2 | 5.3 | 11.1 | 4.6 |
| 4 | Manufacturing of metallic products | 0 | 0 | 1.4 | 0 | 2.6 | 0 | 1.5 |
| 5 | Basic metals | 0 | 3.8 | 8.4 | 0 | 11.3 | 11.1 | 9.3 |
| 6 | Candy and sugar | 0 | 0 | 0.8 | 0 | 0 | 0 | 0.7 |
| 7 | Drilling | 0 | 0 | 1.2 | 0 | 2.4 | 0 | 1.4 |
| 8 | Transportation | 0 | 1.3 | 0.9 | 0 | 1 | 0 | 1.1 |
| 9 | Other nonmetallic products | 0 | 0 | 1.3 | 0 | 0.4 | 0 | 1.6 |
| 10 | Car | 0 | 12.8 | 8 | 8.2 | 8.3 | 0 | 9.3 |
| 11 | Tile | 0 | 0 | 0.2 | 0 | 0.2 | 0 | 0.1 |
| 12 | Technical and engineering services | 0 | 25.6 | 6.8 | 32.7 | 16.2 | 33.3 | 6.7 |
| 13 | Computer | 0 | 1.3 | 1.4 | 2 | 2.2 | 0 | 1.4 |
| 14 | Mass, real estate | 0 | 1.3 | 2 | 2 | 1 | 0 | 1.8 |
| 15 | Cement, lime | 0 | 1.3 | 2.6 | 4.1 | 1.4 | 0 | 3 |
| 16 | Chemical products | 0 | 24.4 | 17.6 | 26.5 | 18.4 | 11.1 | 14.9 |

TABLE 5: Continued.

| | | Data of buying | | | Data of selling | | | |
|----|----------------------------|----------------|----------------|---------------|-----------------|----------------|---------------|----------------|
| | | First cluster | Second cluster | Third cluster | First cluster | Second cluster | Third cluster | Fourth cluster |
| 17 | Medical products | 0 | 0 | 4.9 | 2 | 2.4 | 0 | 5.2 |
| 18 | Bank | 50 | 3.8 | 16.5 | 0 | 10.7 | 11.1 | 13.4 |
| 19 | Food products except sugar | 0 | 3.8 | 3.3 | 0 | 1.2 | 0 | 3.7 |
| 20 | Extraction of minerals | 0 | 2.6 | 3.6 | 0 | 3.8 | 0 | 3.5 |
| 21 | Multi-industrial companies | 50 | 6.4 | 4.7 | 8.2 | 4 | 22.2 | 5.4 |
| 22 | Mediators | 0 | 1.3 | 2.5 | 0 | 1.6 | 0 | 3 |
| 23 | Investment | 0 | 0 | 0.5 | 0 | 0.2 | 0 | 0.6 |

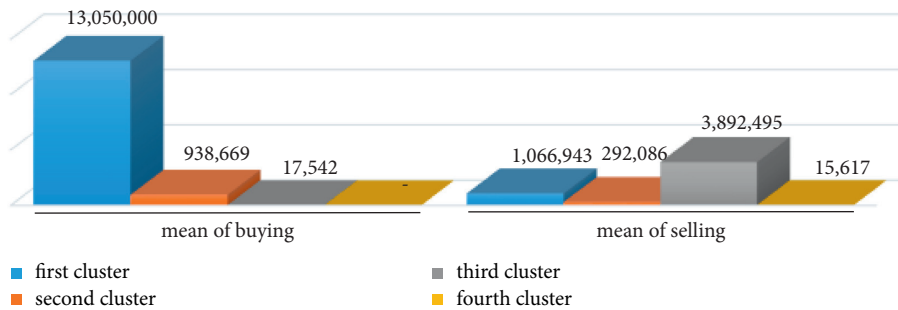


FIGURE 5: Mean of traded shares in each cluster.

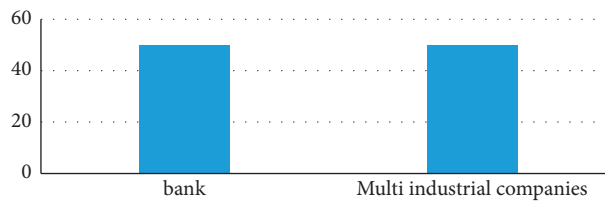


FIGURE 6: Transactions of buying shares in the first cluster.

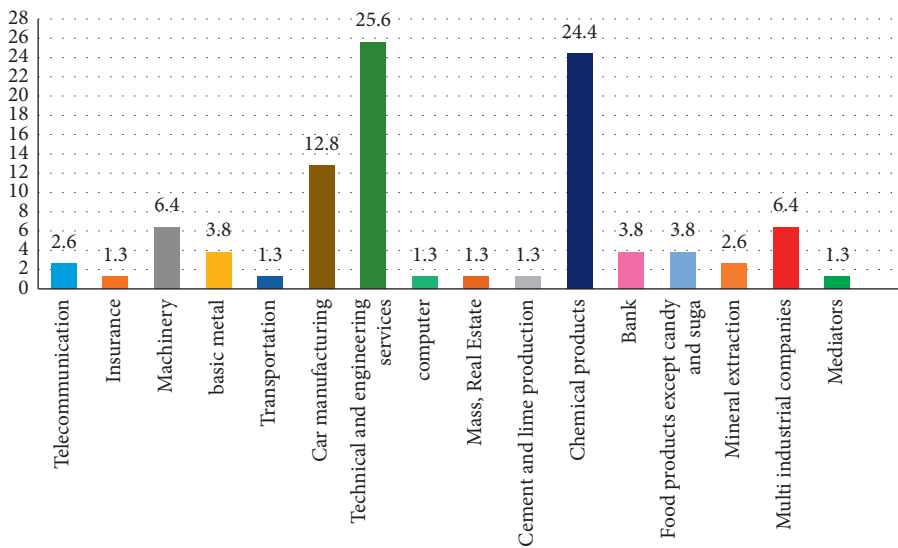


FIGURE 7: Buying transactions in the second cluster.

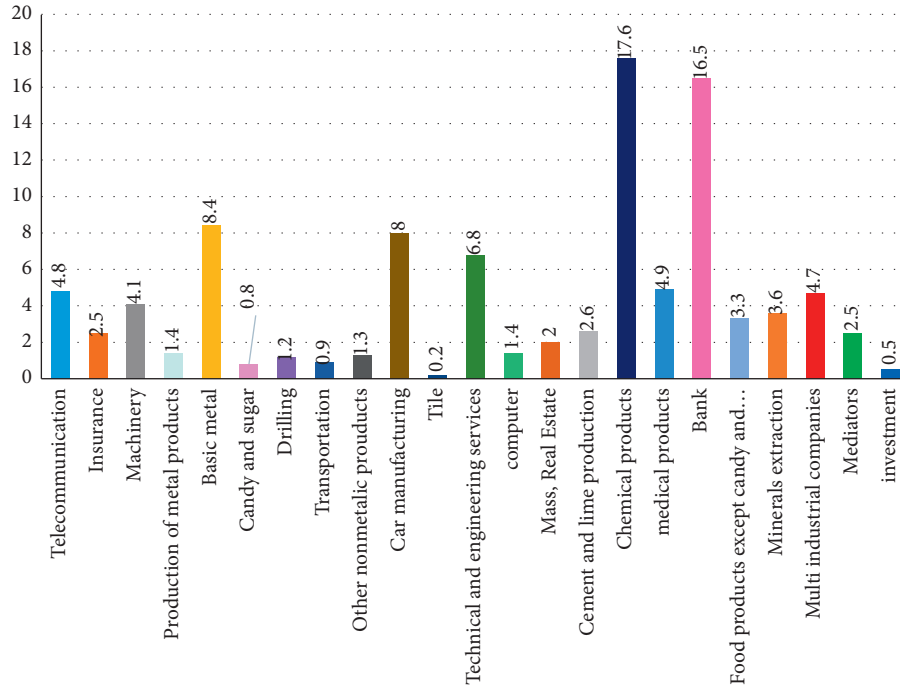


FIGURE 8: Buying transactions in the third cluster.

$$P_{r0}^+ = \max_j (y_{rj}) - y_{r0}, \quad r = 1, 2, \dots, s. \quad (9)$$

For input variables x_i , we have

$$P_{i0}^- = x_{i0} - \min_j (x_{ij}), \quad i = 1, 2, \dots, m. \quad (10)$$

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 the Tehran stock exchange. Table 6 presents the efficiency scores for 2019. According to the data presented in Table 6, among 23 industries, 9 are efficient. Seventy-nine percent of inefficient industries have an efficiency lower than 0.6. The transportation industry with an efficiency of 0.03 is the most inefficient company.

In order to suggest a practical benchmark for increasing the efficiency of inefficient industries, a reference set (linear combination of efficient companies) is used.

Each reference set includes efficient DMUs, which can construct efficient projections of the associated inefficient DMU. According to Table 6, the tile industry and the multi-investment industry have the highest presence in the reference sets of all inefficient industries; that is, they occur most often in nonefficient clusters.

For instance, the inefficient computer industry can follow methods of the tile industry, the multi-investment industry, and the car industry in the selection of inputs and outputs to be projected toward the efficient frontier. Based on efficiency scores reported in Table 6 and using reference sets, one can project the inefficient DMUs toward the efficient frontier. This can be assumed as a practical benchmark for inefficient DMUs. Tables 7–9 represent the efficiency scores and fraction of total investment in each industry.

TABLE 6: Efficiency score for 2019.

| DMU | Company | Efficiency |
|-------|-----------------------------|------------|
| DMU01 | Sugar and candy | 1 |
| DMU02 | Other nonmetallic products | 1 |
| DMU03 | Tile | 1 |
| DMU07 | Medicinal investment | 1 |
| DMU10 | Multi-industrial investment | 1 |
| DMU19 | Chemical investment | 1 |
| DMU20 | Car company | 1 |
| DMU21 | Basic metals | 1 |
| DMU23 | Bank | 1 |
| DMU09 | Computer | 0.7 |
| DMU11 | Extraction of minerals | 0.62 |
| DMU14 | Drilling company | 0.6 |
| DMU08 | Manufacturing company | 0.56 |
| DMU06 | Food products except sugar | 0.51 |
| DMU05 | Investment | 0.47 |
| DMU04 | Cement and lime company | 0.4 |
| DMU15 | Mediators | 0.3 |
| DMU12 | Machinery | 0.29 |
| DMU17 | Telecommunication | 0.29 |
| DMU22 | Technical and engineering | 0.25 |
| DMU13 | Mass, real estate | 0.22 |
| DMU16 | Insurance company | 0.15 |
| DMU18 | Transportation | 0.03 |

Table 10 presents the efficiency scores of industries in 2020. An efficiency of 1 indicates that the DMU is efficient, whereas efficiency of less than 1 indicates inefficiency.

In Table 11, the industries are categorized according to the efficiency scores.

Figure 9 presents the efficiency scores of DMUs in both 2019 and 2020. Thus, the figure provides suitable information to compare the situation of a DMU during the period

TABLE 7: First cluster: efficiency score and investment percentage in 2019.

| DMU | Company | Efficiency | Percentage |
|-------|-----------------------------|------------|------------|
| DMU10 | Multi-industrial investment | 1 | 50.0 |
| DMU23 | Bank | 1 | 50.0 |

TABLE 8: Second cluster: efficiency score and investment percentage in 2019.

| 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 9: Third cluster: efficiency score and investment percentage in 2019.

| DMU | Company | Efficiency | Percentage |
|-------|-----------------------------|------------|------------|
| DMU01 | Sugar and candy | 1 | 0.8 |
| DMU02 | Other nonmetallic 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 10: Efficiency score 2020.

| DMU | Company | Efficiency |
|-------|-----------------------------|------------|
| DMU 1 | Sugar and candy | 1 |
| DMU 2 | Other nonmetallic products | 1 |
| DMU 3 | Tile | 1 |
| DMU 7 | Medicinal investment | 1 |
| DMU10 | Multi-industrial investment | 1 |
| DMU19 | Chemical investment | 1 |
| DMU20 | Car company | 0.92 |
| DMU21 | Basic metals | 0.86 |
| DMU23 | Bank | 0.72 |
| DMU09 | Computer | 0.66 |
| DMU11 | Extraction of minerals | 0.59 |
| DMU14 | Drilling company | 0.54 |
| DMU 8 | Manufacturing company | 0.53 |
| DMU 6 | Food products except sugar | 0.52 |
| DMU 5 | Investment | 0.48 |
| DMU 4 | Cement and lime company | 0.44 |
| DMU15 | Mediators | 0.42 |
| DMU12 | Machinery | 0.25 |
| DMU17 | Telecommunication | 0.23 |
| DMU22 | Technical and engineering | 0.17 |
| DMU13 | Mass, real estate | 0.15 |
| DMU16 | Insurance company | 0.08 |
| DMU18 | Transportation | 0.03 |

TABLE 11: First cluster: efficiency score and investment percentage 2020.

| Cluster | DMU | Company | Efficiency | Percentage |
|-----------|-------|-----------------------------|------------|------------|
| Cluster 1 | DMU10 | Multi-industrial investment | 1 | 50.0 |
| | DMU23 | Bank | 0.72 | 50.0 |

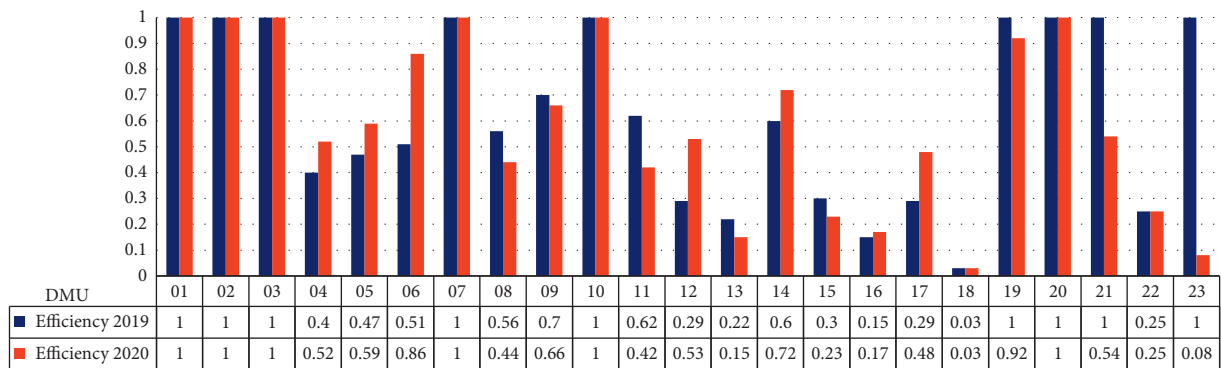


FIGURE 9: Efficiency scores of 2019–2020.

TABLE 12: Second cluster: efficiency score and investment percentage in 2020.

| 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 13: Third cluster: efficiency score and investment percentage in 2020.

| Cluster | DMU | Company | Efficiency | Percentage |
|-----------|-------------------|-----------------------------|------------|------------|
| Cluster 3 | DMU01 | Sugar and candy | 1 | 0.8 |
| | DMU02 | Other nonmetallic 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 | 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 | |

2019–2020. Figure 9 shows that DMU23 and DMU21 have a major reduction of efficiency (92% and 46% compared to their efficiency). DMU06 has a 35% growth of efficiency compared to its efficiency in 2019.

Tables 12 and 13 present the efficiency scores for the second and the third cluster in 2020.

5. Conclusions and Future Research Directions

In this study, a hybrid procedure based on clustering analysis and DEA was proposed to investigate the selling and buying behavior of investors. The whole procedure was applied to some financial records of the Tehran stock exchange. The main steps of the proposed procedure are as follows. In the first stage, a prescreening method was accomplished on data.

Then, a clustering approach was conducted to investigate the main clusters of selling and buying records. Finally, DEA was used to measure the efficiency scores of each cluster. The results were analyzed based on the financial data of the customers in the Tehran stock exchange. The efficient and inefficient DMUs (companies) were determined based on efficiency scores in each buying and selling cluster. The reference set of each inefficient DMU was proposed to achieve the projection toward the efficient frontier. The reference set can help the managers of inefficient industries to move toward the 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.

The phenomenon of COVID-19 has strongly affected our lives. Production and service companies are no exemptions from this observation. In Iran, the pandemic has strong effects on the fragile economy. Rates of death people, lockdowns, and decreased productivity are some of the various factors with economic impact. Hence, undoubtedly, the stock market is affected by this pandemic. In this paper, the performance of companies in the stock market during two years of the pandemic is evaluated.

The result indicated that, among 23 industries listed on the stock market in Iran, solely nine were efficient in 2019, and it shows that the COVID-19 pandemic had a strong effect on the stock market. However, in 2020, the number of efficient industries is less than in 2019 and reached six industries only. Among efficient categories based on the clustering method, the banking and investment industries were the most efficient among them. The reason is that some of them are supported by the government, and others had a suitable financial backup to tackle this problem. However, in 2020, the bank industry was not efficient because the effect of spreading COVID-19 was stronger than effects by government support or other issues. The worst industry in both 2019 and 2020 is transportation. The reason is that, according to rules resulting from the COVID-19 pandemic, many countries forbid or limit entrance to their territories; hence, the performance of transportation decreased dramatically. Another industry that is at the bottom of inefficiency in both 2019 and 2020 is the insurance industry. The reason is that many people in Iran became infected with COVID-19 and, hence, high costs result for these insurance companies to cover medical treatment of insurers. The result of the clustering pointed out that industries with a strong direct effect from COVID-19 had the least performance and were inefficient. For increasing their performance and efficiency, they must pay more attention to financial issues and indicators. In other words, the outbreak of the COVID-19 pandemic diminished the efficiency of companies in various industries in Iran. Comparing our results with those of another study conducted during a normal economic situation [18], we find that usually the three industries, automobile, pharmacy, and cement, had the highest efficiency among other industries. However, in 2019, solely automobile and pharmacy industries remained efficient, whereas the cement industry became inefficient. The reason is that the demand for building houses and thus for cement was decreased dramatically. In 2020 during the outbreak of the pandemic, the automobile section became inefficient as well. The reason is that people do not invest in buying cars in situations of crisis. The pharmacy industry remained efficient in both years because obviously many people were looking forward to treatment; hence, they needed pharmacy and other facilities for treatment.

For future research, we suggest further development of the employed methodology. In particular, researchers can investigate an uncertain environment, for example, by using approaches based on fuzzy numbers, D numbers, or Z numbers.

Data Availability

Used data can be made available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] K. Wendt, "Social stock exchanges-democratization of capital investing for impact," in *Proceedings of the 30th Australasian Finance and Banking Conference*, Sydney, Australia, December 2017.
- [2] J. Górecki, P. Núñez-Cacho, F. A. Corpas-Iglesias, and V. Molina, "How to convince players in construction market? Strategies for effective implementation of circular economy in construction sector," *Cogent Engineering*, vol. 6, no. 1, Article ID 1690760, 2019.
- [3] R. Kumari and A. K. Sharma, "Determinants of foreign direct investment in developing countries: a panel data study," *International Journal of Emerging Markets*, vol. 12, 2017.
- [4] C. Montenegro and M. Molina, "Improving the criteria of the investment on stock market using data mining techniques: the case of S & P 500 index," *International Journal of Machine Learning and Computing*, vol. 10, no. 2, pp. 309–315, 2020.
- [5] A. E. Khedr, N. Yaseen, and S. E. Salama, "Predicting stock market behavior using data mining technique and news sentiment analysis," *International Journal of Intelligent Systems and Applications*, vol. 9, no. 7, p. 22, 2017.
- [6] N. Koptuyug, L. Persson, and J. Tåg, "Should we worry about the decline of the public corporation? a brief survey of the economics and external effects of the stock market," *The North American Journal of Economics and Finance*, vol. 51, Article ID 101061, 2020.
- [7] J. Traina, "Is aggregate market power increasing? production trends using financial statements," *Production Trends Using Financial Statements*, 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3120849.
- [8] P. Wanke, C. P. Barros, and A. Emrouznejad, "Assessing productive efficiency of banks using integrated fuzzy-DEA and bootstrapping: a case of mozambican banks," *European Journal of Operational Research*, vol. 249, no. 1, pp. 378–389, 2016.
- [9] A. K. Yazdi and F. Abdi, "Designing robust model for banks benchmarking based on rembrandt method and DEA," *Benchmarking: An International Journal*, vol. 24, 2017.
- [10] K. Jayamalini and M. Ponnavaikko, "Research on web data mining concepts, techniques and applications," in *Proceedings of the 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)*, pp. 1–5, Chennai, India, February 2017.
- [11] A. Boussofiane, R. G. Dyson, and E. Thanassoulis, "Applied data envelopment analysis," *European Journal of Operational Research*, vol. 52, no. 1, pp. 1–15, 1991.
- [12] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078–1092, 1984.
- [13] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

- [14] K. Ahmadi and M. A. Ramezani, "Iranian emotional experience and expression during the COVID-19 crisis," *Asia Pacific Journal of Public Health*, vol. 32, no. 5, pp. 285-286, 2020.
- [15] A. H. Samadi, S. Owjimehr, and Z. N. Halafi, "The cross-impact between financial markets, Covid-19 pandemic, and economic sanctions: the case of Iran," *Journal of Policy Modeling*, vol. 43, no. 1, pp. 34-55, 2021.
- [16] M. Hiransha, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "NSE stock market prediction using deep-learning models," *Procedia Computer Science*, vol. 132, pp. 1351-1362, 2018.
- [17] A. K. Yazdi, Y. J. Wang, and A. Alirezaei, "Analytical insights into firm performance: a fuzzy clustering approach for data envelopment analysis classification," *International Journal of Operational Research*, vol. 33, no. 3, pp. 413-429, 2018.
- [18] A. Karimi and M. Barati, "Financial performance evaluation of companies listed on Tehran stock exchange: a negative data envelopment analysis approach," *International Journal of Law and Management*, vol. 60, 2018.
- [19] K. Tone, T. S. Chang, and C. H. Wu, "Handling negative data in slacks-based measure data envelopment analysis models," *European Journal of Operational Research*, vol. 282, no. 3, pp. 926-935, 2020.
- [20] A. L. M. Anouze and I. Bou-Hamad, "Data envelopment analysis and data mining to efficiency estimation and evaluation," *International Journal of Islamic and Middle Eastern Finance and Management*, 2019.
- [21] T. S. Chang, K. Tone, and C. H. Wu, "Nested dynamic network data envelopment analysis models with infinitely many decision making units for portfolio evaluation," *European Journal of Operational Research*, vol. 291, no. 2, pp. 766-781, 2021.
- [22] X. Zhong and D. Enke, "A comprehensive cluster and classification mining procedure for daily stock market return forecasting," *Neurocomputing*, vol. 267, pp. 152-168, 2017.
- [23] M. J. Rezaee, M. Jozmaleki, and M. Valipour, "Integrating dynamic fuzzy C-means, data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange," *Physica A: Statistical Mechanics and its Applications*, vol. 489, pp. 78-93, 2018.
- [24] M. K. Mehlatat, P. Gupta, A. Kumar, S. Yadav, and A. Aggarwal, "Multiobjective fuzzy portfolio performance evaluation using data envelopment analysis under credibilistic framework," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 11, pp. 2726-2737, 2020.
- [25] Z. Mashayekhi and H. Omrani, "An integrated multi-objective Markowitz-DEA cross-efficiency model with fuzzy returns for portfolio selection problem," *Applied Soft Computing*, vol. 38, pp. 1-9, 2016.
- [26] S. Asadifar and M. Kahani, "Semantic association rule mining: a new approach for stock market prediction," in *Proceedings of the 2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*, pp. 106-111, Kerman, Iran, March 2017.
- [27] R. E. Caraka, Y. Lee, R. Kurniawan et al., "Impact of COVID-19 large scale restriction on environment and economy in Indonesia," *Global Journal of Environmental Science and Management*, vol. 6, pp. 65-84, 2020.
- [28] L. L. Albu, C. I. Preda, R. Lupu, C. E. Dobrotă, G. M. Călin, and C. M. Boghicevici, "Estimates of dynamics of the covid-19 pandemic and of its impact on the economy," *Romanian Journal of Economic Forecasting*, vol. 23, no. 2, pp. 5-17, 2020.
- [29] S. Grima, R. Dalli Gonzi, and E. Thalassinos, "The impact of COVID-19 on Malta and it's economy and sustainable strategies," *Romanian Journal of Economic Forecasting*, vol. 23, 2020.
- [30] W. Thorbecke, "The impact of the COVID-19 pandemic on the US economy: evidence from the stock market," *Journal of Risk and Financial Management*, vol. 13, no. 10, p. 233, 2020.
- [31] A. Yahya and S. Hidayat, "The influence of current ratio, total debt to total assets, total assets turn over, and return on assets on earnings persistence in automotive companies," *Journal of Accounting Auditing and Business*, vol. 3, no. 1, 2020.
- [32] M. Monea, "Financial ratios-reveal how a business is doing?" *Annals of the University of Petroșani Economics*, vol. 9, no. 2, pp. 137-145, 2009.
- [33] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191-203, 1984.
- [34] A. El Ouardighi, A. El Akadi, and D. Aboutajdine, "Feature selection on supervised classification using wilks lambda statistic," in *Proceedings of the 2007 International Symposium on Computational Intelligence and Intelligent Informatics*, pp. 51-55, Agadir, Morocco, March 2007.
- [35] I. G. Hatvani, J. Kovács, I. S. Kovács, P. Jakusch, and J. Korponai, "Analysis of long-term water quality changes in the kis-balaton water protection system with time series-, cluster analysis and wilks' lambda distribution," *Ecological Engineering*, vol. 37, no. 4, pp. 629-635, 2011.
- [36] N. R. Pal, K. Pal, and J. C. Bezdek, "A mixed c-means clustering model," in *Proceedings of the 6th International Fuzzy Systems Conference*, vol. 1, pp. 11-21, Barcelona, Spain, July 1997.
- [37] K. L. Wu and M. S. Yang, "Alternative c-means clustering algorithms," *Pattern Recognition*, vol. 35, no. 10, pp. 2267-2278, 2002.
- [38] J. Yu, "General C-means clustering model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1197-1211, 2005.
- [39] M. J. Farrell, "The measurement of productive efficiency," *Journal of the Royal Statistical Society: Series A*, vol. 120, no. 3, pp. 253-281, 1957.
- [40] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433-459, 2010.
- [41] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9, pp. 2812-2831, 2014.
- [42] M. C. A. S. Portela, E. Thanassoulis, and G. Simpson, "Negative data in DEA: a directional distance approach applied to bank branches," *Journal of the Operational Research Society*, vol. 55, no. 10, pp. 1111-1121, 2004.