Adv. Math. Fin. App., 2022, **7(**1), P. 149-167

Advances in Mathematical Finance & Applications

www.amfa.iau-arak.ac.ir
Print ISSN: 2538-5569
Online ISSN: 2645-4610
Doi: 10.22034/AMFA.2021.1933112.1608

Research Paper

Analysis of Stock Market Manipulation using Generative Adversarial Nets and Denoising Auto-Encode Models

Hamed Hamedinia[a], Reza Raei[a, *], Saeed Bajalan[a], Saeed Rouhani[b]

[a]*Department of Finance, Accounting and Management, Faculty of Tehran University, Tehran, Iran*
[b]*Department of IT Management, Faculty of Tehran University, Tehran, Iran*

ARTICLE INFO

Article history:
Received 2021-05-12
Accepted 2021-08-04

Keywords:
Anomaly Detection
Deep Learning
Generative Adversarial Net (GAN)
Stock Manipulation Detection

ABSTRACT

Market manipulation remains the biggest concern of investors in today's securities market. The development of technologies and complex trading algorithms seems to facilitate stock market manipulation and make it inevitable for regulators to use Deep Learning models to prevent manipulation. In this research, a Denoising GAN-based model has been designed. The proposed model (GAN-DAE4) consists of a three-layer encoder along with a 2-dimension encoder as the discriminator and a three-layer decoder as the generator. First, using statistical methods such as sequence, skewness, and kurtosis tests and some unsupervised learning methods such as Contextual Anomaly Detection (CAD) and some visual and graphical methods, the manipulated stocks have been detected in the Tehran Stock Exchange from 2015 to 2020; then GAN-DAE4 and some supervised deep learning models have been applied to the prepared data set. The results show that GAN-DAE4 outperformed other deep learning models (with F2-measure 73.71%) such as Decision Tree (C4.5), Random Forest, Neural Network, and Logistic Regression.

1 Introduction

The capital market as one of the subsystems of the economic-financial sector has a significant role in the development of economy in different countries [9]. The most vital prerequisite for the comprehensive development of the capital market is the trust of market participants and investors in its efficiency for determining the fair assets' price and allocating financial resources to various sectors optimally. Therefore, the main regulator task is to create a suitable environment for market participants to trade fairly [15]. From the beginning of 2019 to the middle of 2020 in the Iran capital market, the public's interest increased significantly. A 67 percent increase in active stock codes and a 350 percent increase in daily trading volume confirm this. Therefore, the role of the capital market in recent years has become much more prominent among different segments of the population in Iran. Therefore, the regulator must consider a mechanism that prevents people from losing confidence in the capital market [49].

* Corresponding author. Tel.: *+9821-61117621*
E-mail address: *Raei@ut.ac.ir*

Market manipulation is one of the most harmful activities for the capital market and its participants. Manipulation causes participants to lose confidence in the market. This problem harms the market's liquidity, increases the cost of transaction costs, and increases the financing rate for companies listed in the market [10]. According to Allen and Gale [2] and Jarrow [29], stock manipulation is classified as action-based, information-based, and trade-based manipulation. Some years later, Comerton-Forde and Putnins [9] augment those categories by adding order-based and submission-based techniques. Trade-based manipulation involves influencing the price of a financial instrument through trading. In information-based manipulation, a manipulator releases false information or rumors about a company to inflate or depress its share price. Action-based manipulation involves taking actions to affect the price of a firm without changing supply and demand.

For example, a company director may shut down one production line to decrease the share price. Order-based manipulation involves using submissions, amendments, and cancellations of orders that are not intended to execute. Submission-based manipulation consists of making false or misleading submissions to a financial benchmark calculation [38]. Although market manipulation is more difficult in the developed markets due to strict regulations and strict monitoring, manipulation problems in these markets are not entirely solved; Because, in most cases in developed markets, manipulation occurs in complex and secretive ways. In developing markets where there is no close monitoring, manipulation occurs more easily. So, manipulation in developing market is more important than developed market [12, 41]. Market manipulation is difficult to define precisely. In part, this is because it encompasses a comprehensive collection of highly varied trading strategies. But another reason, in particular for the vagueness in legal definitions, is to minimize the risk that manipulators circumvent the law by devising schemes that fall outside of a precise and narrow legal definition [38]. The law and financial economics literature have a considerable debate about how to define manipulation. In a departure from mainstream legal thought, Fischel and Ross [21] argue that market manipulation is too vague a concept to form the basis for criminal charges. They point out that there is no objective definition of manipulation and suggest that manipulation could only be defined as dishonest intent to move stock prices. Manipulation and Anomaly detection has been studied in various fields such as credit cards, telecommunications, and banking transactions, etc., still this issue has been less studied by researchers in the capital market. First of all, the volume of transactions in the capital market is extremely high; second, the time series of the stock price is one of the most complex time series available, making it challenging to identify manipulation and anomalies. Lack of access to Cleaned and Labeled data is another reason for the lack of research in this area [20, 24].

Although the regulator is always trying to identify and eliminate market manipulation methods, unfortunately, it is not easy to detect manipulations. Manipulation detection has become more complex than ever with the advent of new computer programs and sophisticated trading algorithms. Also, with the expansion of the financial market and increasing the volume of transactions, humans can't monitor these fraudulent actions. Therefore, Market regulators need an intelligent computer program to investigate these transactions [31]. In the Iran exchange market, despite the existence of numerous instructions to prevent manipulation, unfortunately, according to many participant's opinions, many stocks are manipulated. Previous studies related to the detection of price manipulation in the Tehran Exchange Securities have mainly examined small features such as return, volume, PE ratio, and floating share [13, 14,15,16,17]. The small number of features, lack of appropriate and reliable data and not using new artificial intelligence methods have caused the manipulation detection always to be a severe challenge in the Iranian capital market. Therefore, in this study, we try to detect market manipulation in the Tehran Exchange Securities using a GAN-based Denoising Auto-Encode-Decode Model (GAN-DAE4) and effective and extensive features that never be applied in previous researches. The Contributions of this paper mention as follows: In the present study, for the first time, the combined model of denoising auto-encode-decode and GAN has been used to detect manipulation. To our best knowledge, this combination is not used for stock manipulation detection. The Unsupervised Learning model has not yet been used to identify price manipulation in the stock market of Iran. As there are no clean and labeled data in stock manipulation in Iran exchange market, it is better to use unsupervised learning other than supervised one. Of the unsupervised, GAN-based models are promising for anomaly detection. For the first time in Iran, we divided stock into manipulated and not manipulated categories using Contextual Anomaly Detection (CAD) test. We have used many features in our model because the structure of the model allows it (a denoising 3-layer Auto-encode-decode). The number of features used in this article is broader and more comprehensive than previous researches.

The remainder of the paper is organized as follows. Section 2 reviews the literature and positions our work. Section 3 discusses the cases, the datasets, and the deep learning models used in this work in detail. Section 4 discusses our proposed GAN-based Model (GAN-DAE3) and shows the results. Finally, section 5 summarises the work and proposed future research directions.

2 Literature Review

Manipulation has drawn the attention of many researchers. Early theoretical trade-based manipulation literature establishes very general conditions under which pure trade-based manipulation in a single market (e.g., a series of buys followed by a series of sells) is and is not profitable. Fischel and Ross [21], among others, argue that trade-based manipulation is not possible in an efficient market. Jarrow [29], Cherian and Kuriyan [7], and Cherian and Jarrow [8] build on the model of Hart [27] and derive conditions under which trade-based manipulation is not possible. Early empirical asset pricing and market microstructure studies identify various abnormalities in closing prices but do not link the abnormalities to market manipulation. however, several studies attribute seasonal patterns and anomalies in day-end trading to closing price manipulation. Felixson and Pelli [19] examine whether closing prices are manipulated in the Finnish stock market using regression analysis. Although their results are consistent with the hypothesis that closing prices are manipulated, they concede that further research is required to be conclusive. Zhai et al. [46] proposed a paper based on analytics analyzing trading behavior data for manipulation detection in the NASDAQ stock market. The authors divided the methods into two groups: static models (k-nearest neighbor and one-class support vector machine) and dynamic models (adaptive dynamic model and Gaussian mixture model). Stock price manipulation strategies were classified and analyzed into two forms: spoofing trading and quote stuffing. Both models were effective to detect manipulative behaviors.

The literature also provides some indirect evidence about relatively high-frequency (intraday) spoofing manipulation. For example, Lee et al. [32] analyze account-level limit order book data from the Korean stock exchange. They find patterns of order submissions and cancelations consistent with spoofing, and more specifically, layering. Their results suggest that manipulators exploited a particular feature of the Korean stock exchange as part of their spoofing strategies. Namely, until 2002, the Korean stock exchange displayed the total quantity of orders on each side of the order book without displaying prices, which meant that displayed volume could be easily manipulated by placing orders very far from the best quotes with little chance of the orders executing. A recent paper by Griffin and Shams [26] provides a collection of circumstantial evidence that suggests the VIX Volatility index is manipulated around the time that VIX futures and options settle. VIX is calculated from the prices of SandP 500 index options, including highly illiquid deep-out-of-the-money options. The evidence suggests traders manipulate the

VIX index by aggressively trading the illiquid deep-out-of-the-money SandP 500 index options, presumably to profit from the settlement of VIX derivatives. The application of deep learning methods to detect stock manipulation is relatively new, but significant growth has occurred in recent years [11, 23]. The first researchers who used deep learning methods to detect manipulation in developed markets were Westphal and Blaxton [44]. more recently, new models of deep learning have been conducted by Ogut et al. [35] in developing markets. Diaz et al. [11] examined the challenges of using artificial intelligence methods to identify market manipulation. They used features like the difference in stock returns manipulated by the reference portfolio, abnormal stock returns, liquidity and stock volatility. Finally, the researchers used the decision tree to examine the manipulated stock and proposed an algorithm to detect market manipulation. Golmohammadi et al. [24] have studied anomaly detection using Contextual Anomaly Detection (CAD). the researchers showed that the CAD model outperformed KNN and Random Walk models in detecting anomalies. In a similar study, Al-Thani [1] examined the Qatar Stock Exchange and slightly changed the inputs and obtained far better results than the previous study. Leangarun et al. [31] designed a GAN-based model to identify stock manipulation in the capital market. In their model, first, the GAN is trained with normal data and then the trained network is used to detect the manipulated data. According to the results, the accuracy of the model is estimated at 68%. In their GAN, the LTSM model is used as a Discriminator and features such as the first, last, lowest, highest price, trading volume, etc. are used as features.

Zheng et al. [47] investigated fraudulent transactions in the banking network in China. Using a variety of GAN-based models (semi-supervised models), they designed a model that can be used to detect suspicious transfers. According to the authors, if the number of positive (manipulated) samples is too small, conventional deep learning models will not learn effectively. Therefore, the GAN-based model designed by researchers will eliminate this problem. Ergun et al. [12] examined stock manipulation in Turkey using trade-based manipulations announced by the Turkish Capital Markets Board for 2005 to 2013. two of the main questions of the researchers was how the manipulators choose the suitable stock for manipulation and what is the effects of manipulation on market efficiency. The results show that manipulators prefer low liquidity, weak performance and less volatility stocks to manipulate in emerging markets. They also observed that in most cases, liquidity, efficiency and volatility increased during the manipulation period and decreased after the manipulation period. The application of manipulation detection using deep learning techniques has also recently been extended to digital currency markets.

There is recent work by Xu et al. [45] analyzing pump-and-dump in the context of crypto-currency trading that these scams are conducted manually by humans coordinating over services like Telegram and Discord and without layering. Unfortunately, in Iran, stock manipulation data are not disclosed by the regulator. For this reason, there is little research in this area.

Fallah Shams and Teymouri Shendi [15], using stock market information in 2002-2004, investigated the factors affecting the manipulation and then the binary logistic regression model was fitted to predict the probability of manipulation. Falahshams and Kordloui [16] divided the 397 companies listed on the Tehran Stock Exchange from 2001 to 2009 into manipulated and unmanipulated categories, using Sequence, Kurtosis, Skewness and Duration tests. Then, using logit and artificial neural network and features like the company's size, information transparency, PE ratio and liquidity of stocks, they designed a model to predict the stock manipulation. The results showed that the performance measure of the logit model and artificial neural networks was 92.1% and 94.1%, respectively. Several studies have been conducted in the Iranian capital market to identify manipulation using deep learning methods. Poostforoush et al. [37] showed that the Quadratic Discriminant Analysis (QDA) model performs better than the artificial neural network. Fallahzadeh [13] used the Decision Tree Networks (C5.0, CART, …)

to detect manipulations. They show that C5.0 outperforms other DT networks. Rabiee et al. [39] investigated the effects of stock manipulation on market efficiency in the Tehran Security Exchange. The results showed that the supply and demand gap increase shortly before the manipulation begins and this gap decreases after the manipulation occurs. Stocks with high trading volumes are more disposed to manipulation. Nadiri et al. [34] examined companies capable of being manipulated in the Iranian capital market. Using two quarterly periods (bearing and bullish market), the researchers studied spoofing and order-based manipulation. The results show that small companies with high trading volume, low information transparency, high yield are more likely to be manipulated. They found as well that manipulation is inversely related to changes in the market index. Kazemi-Tabar and Shahbazzadeh [30], using the Chebyshev inequality, provided a way to identify people who have used inside information and have made a significant profit in a short time by fraudulent activities. YekeZare [33] investigated manipulation during the rights issues period. The results show that the major shareholders reduce the stock price by selling the stock during the rights issues period and then buy the rights issues at low prices. Some legal and jurisprudential studies have also been conducted about manipulation in the Iran capital market. The results show that manipulation is legally forbidden and Sharia prohibited it [3, 22, 43].

3 Research Methodology

The standard approach for manipulation detection with deep learning models is to use dataset containing manipulated stock prosecuted by regulator (court). In this research, we have studied several deep learning models containing Neural Networks, Logistic Regression, Decision Tree, Random Forest and Generative Adversarial Nets (GAN). The first four models are supervised and the last one is unsupervised. We have proposed an unsupervised denoising GAN-based model that outperforms other ones.

3.1 Models

We define the classification problem as predicting the class of $\mathrm{Y} \in \{0, 1\}$ based on a feature set of X_1, X_2, …, X_d where Y represents the class of a sample (1 implies a manipulated sample) and X_i represents features such as returns, number of shares in a transaction (i.e., volume), etc. The dataset is divided to training and testing dataset.

Neural Network: A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so, the network generates the best possible result without needing to redesign the output criteria. The concept of neural networks, which has its roots in artificial intelligence, is swiftly gaining popularity in the development of trading systems. A neural network contains layers of interconnected nodes. Each node is a perceptron and is similar to a multiple linear regression. The perceptron feeds the signal produced by a multiple linear regression into an activation function that may be nonlinear. In a multi-layered perceptron (MLP), perceptrons are arranged in interconnected layers. The input layer collects input patterns. The output layer has classifications or output signals to which input patterns may map. Hidden layers fine-tune the input weightings until the neural network’s margin of error is minimal. It is hypothesized that hidden layers extrapolate salient features in the input data that have predictive power regarding the outputs. This describes feature extraction, which accomplishes a utility similar to statistical techniques such as principal component analysis. In the present study, we construct a structure with two hidden layers and a Relu activation function.

Decision Tree: Decision Tree Analysis is a general, predictive modelling tool that has applications spanning a number of different areas. In general, decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The decision rules are generally in form of if-then-else statements. The deeper the tree, the more complex the rules and fitter the model. There are several types of Decision Tree models like C4.5, C5.0, CART, etc. in this research, we use C4.5.

Logistic Regression: Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no). Mathematically, a logistic regression model predicts P(Y=1) as a function of X. It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc. Generally, logistic regression means binary logistic regression having binary target variables, but there can be two more categories of target variables that can be predicted by it. Based on those number of categories, Logistic regression can be divided into following types: Binary or Binomial, Multinomial and Ordinal.

Random Forest: Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result. Random forest works in four steps. first, start with the selection of random samples from a given dataset. second, this algorithm will construct a decision tree for every sample. Then, it will get the prediction result from every decision tree. Third, in this step, voting will be performed for every predicted result. forth, select the most voted prediction result as the final prediction result.

Generative Adversarial Net (GAN): GAN introduced by Goodfelow et al [25] in 2014 is a new framework which trains two models like a zero-sum game. In the adversarial process, the generator can be seen as a cheater to generate the similar data as the real data, while the discriminator plays the role of judge to distinguish the real data and generated data. They can reach an ideal point that the discriminator is unable to differentiate the two types of data. At this point, the generator can capture the data distributions from this game. GAN can be divided into supervised, unsupervised and semi-supervised models. In this article, we use unsupervised GAN to detect manipulation in stock market.

GANs consist of two primary components, a generator G and a discriminator D. They were introduced by Goodfellow et al. [25]. Two networks try to compete with each other. The generator G synthesizes a realistic sample that is similar to the training set by learning the distribution of input data from the latent space. The discriminator D performs a classification task. The discriminator D differentiates real data from the training set (class 1) and generated data from the generator G (class 0). Generator's weights are frozen while the discriminator D is training and vice versa. Both machines try to fine-tune their parameters and become better in their capabilities. Although each player depends on each other, each player cannot control the other's parameters. The objective function of GANs is as follows:

$$\min\max \mathrm{V(D,G)} = E_{x\to p_{data(x)}}[logD(x)] + \ E_{z\to p_{pz(z)}}\ \big[\log(1-D\big(G(z)\big))\big] \tag{1}$$

where z is vectors in the latent space. GANs solution involves minimization in the outer loop and maximization in the inner loop. The discriminator D aims to maximize those two terms. So, in the first term, the discriminator D should give an output '1' for real data. In the second term, the discriminator D should give an output '0' if the generator G cannot fool the discriminator D. On the other hand, the generator G aims to minimize only the second term. The discriminator D should give an output '1' for the generated data.

3.2 Data

Unfortunately, in Iran, the Exchange and Securities Organization does not announce manipulated stocks. Even in the developed markets, many stock manipulations are not found and prosecuted. So, there is not any reliable data for stock manipulation in the developed and developing markets. Therefore, it is necessary to create a database containing manipulated stocks. To generate the desired database, we use a variety of existing methods such as Sequence, Kurtosis and skewness test [16] and Contextual Anomaly Detection method [24], to identify stocks in which there is a possibility of manipulation. Then, considering some specific features such as abnormal returns, a sudden increase in trading volume, stock's volatility, difference between the returns of suspected stock and its industry (according to [11], we specify the exact date that stock is manipulated). We have analyzed the information of 69 stocks from 19 different industries from 2015 to 2020. During the mentioned period, three statistical tests containing of Sequence, Kurtosis, Skewness and four contextual anomaly detection tests (15-day and 30-day test on volume and return) have been conducted for all studied stocks. If on a trading day, at least 5 of the 7 tests indicate that the stock has been manipulated, we will consider that stock as manipulated.

Sequence Test: In the sequence test, if the negative and positive sequences are higher than the expected sequences, it indicates that there is a non-random pattern in the stock price trend. as a result, manipulation has probably occurred. we give a negative sign to the stock if daily returns are less than the one-month average period return and a positive sing vice versa. Each sequence contains one or more consecutive positive and negative signs. In other words, a new sequence is started when the sign changes. The total sequences are counted as the number of sequences. Then, the number of expected sequence and the standard deviation are calculated using the following formulas:

$$E(R) = \frac{2n_1n_2}{n_1+n_2} \tag{2}$$

$$Std = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} \tag{3}$$

Where n_1 is the number of positive returns and n_2 is the number of negative returns. Then the number of counted sequences and the number of expected sequences is examined by t-test. If the test statistic (difference between the number of counted sequences and the number of expected sequences divided by the standard deviation of the sequences) is in the critical range, then the number of sequences are not significantly different from the expected number of sequences and it is concluded that the number of the sequences is Random; Therefore, there is no possibility of manipulation. However, if the test statistic is not in the critical range, it means that the number of counted sequences is significantly different from the expected number of sequences, which means that it does not match random data. So, there is a possibility of manipulation. The sequence test results are presented in Table 1 for one sample.

Table 1: Sample Sequence Test Result (Foolad Khorasan, 2016/06/21 to 2016/09/21)

Period	3-month daily return
Average of daily return	0.73%
No. of return above average	14
No. of return under average	48
No. of sequences	13
No. of expected sequences	22.67
Standard deviation of expected sequences	2.71
Test statistic	-3.57
P-value	0.02%

According to the p-value mentioned in Table 1, the stock (Foolad Khorasan) has been manipulated even at the 99% confidence level.

Skewness and Kurtosis Test: Another way to detect market manipulation is to examine the kurtosis and skewness of stock returns. If the kurtosis and skewness of stocks are significantly different from the kurtosis and skewness of the normal distribution function, manipulation is likely to be occurred. The mean and variance of kurtosis and skewness are calculated from Equation 4 and Equation 5.

$$\text{Skewness}=\frac{\sum(x-\bar{x})^4}{(n-1).S^4}\ ,\ \text{Kurtosis}=\frac{\sum(x-\bar{x})^3}{(n-1).S^3}\quad \text{where } S=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \tag{4}$$

$$\text{Skewness}=\sqrt{\frac{6n}{(n-2)(n-1)(n+3)(n+5)}}\ ,\ \text{Kurtosis}=\sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \tag{5}$$

Table 2: Kurtosis calculation sample (Ayandeh Bank, 2020/09/22 to 2020/12/20) - Skewness calculation sample (Fanavard, 2016/03/20 to 2016/06/20)

Period	3-month daily return
Kurtosis	0.785
Std of kurtosis	0.343
Test statistic of kurtosis	2.288
P-value of kurtosis	1.11%
Skewness	2.349
Std of skewness	0.935
Test statistic of skewness	2.513
P-value of skewness	0.6%

As the results of Table 2 show, the kurtosis of Ayandeh Bank in the third quarter of 1399 with 95% confidence level and skewness of Fanavard stock in the first quarter of 1395 with 99% confidence is positive.

Contextual Anomaly Detection (CAD): The classic approach in anomaly detection is comparing the distance of given samples with a set of normal samples and assigning an anomaly score to the sample. Then, samples with significant anomaly scores are labeled as outliers/anomalies. The input to CAD is the set of time series $\{X_i | i \in \{1, 2, \dots, d\}\}$ from one sector such as steel industry stocks (time series that are expected to have similar behavior as they share similar characteristics including underlying factors which determine the time series values) and a window size. First, a subset of time series is selected based on the window size. Second, a centroid is calculated representing the expected behavior of the time series of the group within the window. The centroid is used along with statistical features of each time series X$_i$ (e.g., correlation of the time series with the centroid) to predict the value of the time series at time t (i.e., $\widehat{X_{it}}$). This is a lazy approach, which uses the centroid along with other features of the time series for predicting the values of X$_i$:

$$\widehat{X_{it}} = \psi(\phi\ (X_t), c_t) + \ \varepsilon \tag{6}$$

where X_{it} is the predicted values for the time series X$_i$ at time t, $\phi\ (X_t)$ is a function of time series features (e.g., the value of X$_i$ at time stamp t-1, drift, auto regressive factor etc.), ψ specifies the relationship of a given time series feature with the value of centroid at time t (i.e., C$_t$), and ε is the prediction error (i.e., $\left|\widehat{X_{it}} - X_{it}\right|$). The centroid time series C is the expected pattern which can be calculated by taking the mean or weighted mean of values of time series X$_i$ at each time stamp t. We define ϕ as the inner product of statistical features of each time series and its correlation with the centroid. For a better explanation of the model, suppose we have five stock return time series in the steel industry. We calculate the average return of 5 stocks per day for calculating centroid (C). Now, for each time series, we calculate the correlation between C and one of the mentioned time series, and from Equation 7, which is a special form of Equation 6, we estimate the time series of the stock in the time window.

$$\widehat{X_{it}} = X_{it-1} * Corr(X_i, C) + \ \varepsilon \tag{7}$$

Now we calculate the Euclidean distance between the estimated and actual value ($\left|\widehat{X_{it}} - X_{it}\right|$). if the distance is greater than the standard deviation of the stock in the time window, we consider the data as an anomaly. Golmohammadi et al. [24] showed that the recognition power of this type of CAD, although more seductive than other classification methods, is about 30%. However, Al-Thani [1] showed that if simple moving average (SMA) is used instead of X_{it-1}, the results will be much better and the success rate will increase to more than 95%. In the present study, to determine the manipulated points, we have used the second method CAD (Al-Thani [1]) for two variables: return and volume.

Table 3: CAD volume calculation of Gheshir on 2018/05/16

Time window	15 trading days
Correlation between stock and industry volume	83.13%
Average of stock volume	1,738,785
Stock volume estimation	1,445,454
Stock volume real	5,589,642
Euclidian distance	4,144,188
Standard deviation of stock volume	1,655,842

As Euclidian distance is far greater than the 15-day standard deviation of stock volumes on that particular day, the stock has anomalies and it should be examined according to the manipulation detection procedure mentioned in the previous section.

Table 4: CAD return calculation of Khousaz on 1395/07/24

Time window	30 trading days
Correlation between stock and industry return	24.17%
Average of stock return	0.91%
Stock return estimation	0.22%
Stock return real	4.62%
Euclidian distance	4.40%
Standard deviation of stock return	2.53%

As Euclidian distance is far greater than the 30-day standard deviation of stock returns on that particular day, the stock has anomalies and it should be examined according to the manipulation detection procedure mentioned in the previous section. In Table 5 we mention some features.

Table 5: Features used as inputs to deep learning

Variables	Calculation
Stock Volume	V_i
Stock Return	$R_i = \text{Ln}\ (p_i/p_{i-1})$
Industry Return	$R_{industry} = {}^{1}/_{n} \sum r_i$
Industry Volume	$V_{industry} = \sum V_i$
Market Cap	M_i
Average buys of real people	Vol. buys people / No. buys people
Average buys of firms	Vol. buys firms / No. buys firms
Average sells of real people	Vol. sells people / No. sells people
Average sells of firms	Vol. sells firms / No. sells firms
Capital Raising	Equal to 1 for 10 trading days before and 5 trading days after capital raising announcement

Because manipulation does not occur in just one day and there are several days or even several months for manipulation, in this study, for each feature mentioned in table 5 (except capital raising), we have considered the previous 10 trading days as input. Therefore, the number of inputs to the deep learning models is equal to 100 (= 9 * 11 + 1). According to the mentioned structure, about 71,200 data were analyzed in this research, of which about 500 data were manipulated.

3.3 Performance measure

Quantitative performance measures are needed to evaluate the efficiency of deep learning networks. Outputs of deep learning models are as follows:

True positive (TP): represent the number of manipulated cases classified correctly as positive

False Positive (FP): represent the number of non-manipulated samples that are incorrectly classified as positive

True Negative (TN): represent the number of non-manipulated samples that are correctly classified as positive

False Negative (FN): represent the number of manipulated samples that are incorrectly classified as negative

We have used four performance indicators including $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$, F1-measure and F2-measure. The F-measure formula is mentioned in the following equations:

$$F_\beta = \frac{(1+\beta^2) * \mathrm{P} * \mathrm{R}}{(\beta^2 * \mathrm{P}) + R} = \frac{(1+\beta^2) * \mathrm{TP}}{(1+\beta^2) * \mathrm{TP} + (\beta^2 * \mathrm{FN}) + \mathrm{FP}} \quad (8)$$

Where beta is the significance coefficient of the Recall versus Precision. If precision and recall are of equal importance, beta is equal 1 (F1-measure). In the capital market, incorrect classification of manipulated and normal stocks has no symmetrical cost. which means that if the stock is actually manipulated and incorrectly classified as non-manipulated (FN), the cost will be much higher than if the stock is not actually manipulated and incorrectly classified as manipulated (FP). Because, in the latter case, there is only checking cost for the regulator, but in the first case, the shareholders may be harmed by the manipulation activities [34]. Therefore, the most critical performance measure used in this study is F2-measure.

4 Result and Discussion

First, we describe the structure of the proposed model in this study, which is created by combining Generative Adversarial Nets with Denoising Auto-Encode-Decode (GAN-DAE4), and then compare the results of different models in detecting manipulation. Many features in the capital market may be effective in manipulation detection. The 3-layer encoders in the GAN-DAE4 model allows us to use a large number of features as inputs without being concern about performance reduction. With the rapid changes of the capital market, new methods of manipulation are emerging. In this situation, the old deep learning models have lost their capability since they have been trained with old features suited for those data, and it is essential to train the model with new features and new data again. On the other hand, it is necessary to pay attention to the fact that manipulators can deceive the deep learning models that have been trained by previous data with slight changes in some features. In the present study, we solved this issue by adding some noise to the inputs (Denoising).

Our basic idea is to use a deep neural network to extract latent representations that can support much more effective classification than raw input features. It employs adversarial learning to improve further the accuracy of discriminating between positive samples and negative samples in the data distribution. We take a three-layer autoencoder as the building block of our model. An autoencoder consists of an encoder that encodes an input vector X to a hidden (latent) representation $Z = f_\theta(x)$and a decoder that decodes z to a reconstructed vector $x' = g_{\theta'}(z)$, where f and g are affine mappings that can be sigmoid functions, and θ and θ' are vectors of weight and bias parameters of the encoder and the decoder, respectively. Autoencoder training consists of minimizing the reconstruction error:

$$\arg\min E_{x\to X}\left[L\left(x, g_{\theta'}(f_\theta(x))\right)\right] \tag{9}$$

Where X is the empirical distribution defined by the training set D, and L is the loss function. Typical choices for $L(x, x')$ include the squared error $\|x - x'\|^2$ for real-valued vectors and the fake one generated by G. A denoising autoencoder is a simple variant of the basic autoencoder where the encoder accepts a noised input $\tilde{x} = (x, \varepsilon)$ and transforms it to the latent $Z = f_\theta(\tilde{x})$. Denoising autoencoder training still consists of minimizing the average reconstruction error, still, the key difference is that the latent z is a function of $\tilde{x}$ rather than x and thus the result of a stochastic mapping of x:

$$\arg\min E_{x\to X}\left[L\left(x, g_{\theta'}(f_\theta(\tilde{x}))\right)\right] \tag{10}$$

D and G are simultaneously optimized through a two-player minimax game with the objective function mentioned in Equation 1. Fig. 1 shows the structure of the model designed in the present study.
In our GAN, the 3-layer encoder together with the last encoder acts as the discriminator D. That is, the 3-layer encoder accepts an input vector $\tilde{x}$ representing stock features along with some noise (the input features of which are summarized in Table 5) and transforms it into a latent vector Z, and encoder4 calculates from Z a possibility of the data being positive, i.e., fake data from the generator (x') is regarded as negative samples and normal data from real input ($\tilde{x}$) is regarded as positive sample. In this way, the 3-layer encoder is trained for discovering some important implicit features indicating a fraudulent transfer and encoding them into latent vectors to facilitate final detection. The decoder acts as the generator G, which accepts a latent vector Z, and outputs a set of features that constitute a (fake) feature. Finally, the 3-layer encoder together with encoder4 acts as the classifier that outputs a possibility of the input stock being a normal rather than a manipulated one. The discriminator D and the generator G are

simultaneously trained using iterative gradient descent that alternates between K steps of optimizing D and one step of optimizing G according to Equation (1) as shown in Algorithm (1).

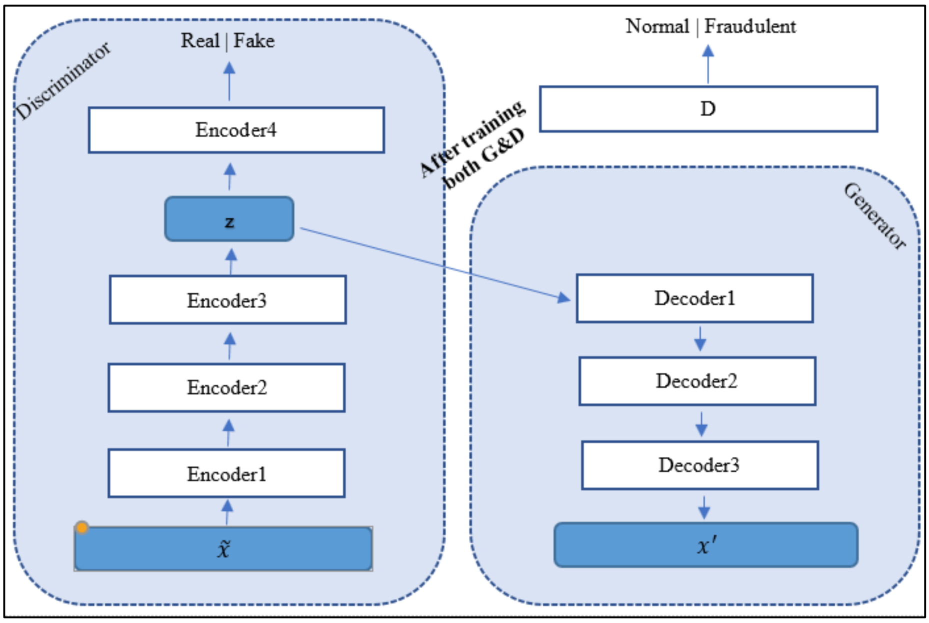

Fig. 1: Denoising Auto-Encode-Decode Generative Adversarial Nets (GAN-DAE4)

For our GAN-based model (denoted by GAN-DAE4), the input dimension is 100, and we empirically set the number of hidden neurons in the first, second and third layers to 96, 46 and 23, respectively. For the GAN training algorithm, we set k = 1 and the maximum number of iterations to 1000. For the last encoder (encoder4), the number of hidden neurons is equal to 2 (manipulated and normal). Once the model is fully trained, the 3-layer encoder with the encoder4 can be used to detect manipulations. Therefore, we first encode the test data into the latent Z with the 3-layer encoder and then encode Z with encoder4. The encoder4 returns a binary value (as mentioned before, the last encoder has just two dimensions) that indicates whether the stock is manipulated or not. It is worth noting that labeled data is only used to test the performance of GAN-DAE4.

Algorithm 1:

```
1: while the stop criterion of generative adversarial learning is not satisfied do
    2: for k=1 to k do
        3: Sample a minibatch Z of latent vector;
        4: Generate XA from generator for Z;
        5: Sample a minibatch of normal data XN;
        6: Update the discriminator's weights $\Theta_D$ by ascending along the gradient:
          $\nabla_{\Theta_D}\left[E_{x\in X^A}\log D(x) + E_{x\in X^N}\log\big(1 - D(x)\big)\right]$
        7: end for
    8: Sample a minibatch Z of latent vectors;
    9: Update the generator's weights $\Theta_G$ by ascending along the gradient:
      $\nabla_{\Theta_G}\left[E_{z\in Z}\log D\big(G(z)\big)\right]$
10: end while
```

Table 6 shows the performance of the deep learning models. As it is clear from the results, the highest value of F2-Measure belongs to the proposed model (GAN-DAE4). The performance of the decision tree is also ranked second and is considered very appropriate. It should be noted that GAN-DAE4 is unsupervised learning, which is compared to the supervised model. Supervised models usually perform better than unsupervised models, still, the GAN-based model has relatively high power in identifying the data distribution function, and this factor has helped to provide acceptable results.

Table 6: Performance Measure Comparison

Models	**Precision**	**Recall**	**F1-measure**	**F2-measure**
GAN-DAE4	87.01%	71.00%	78.19%	73.71%
Decision Tree	66.65%	69.06%	67.83%	68.56%
Random Forest	100.00%	44.00%	61.11%	49.55%
MLP	83.33%	36.53%	50.79%	41.15%
Logistic Regression	66.67%	22.22%	33.33%	25.64%

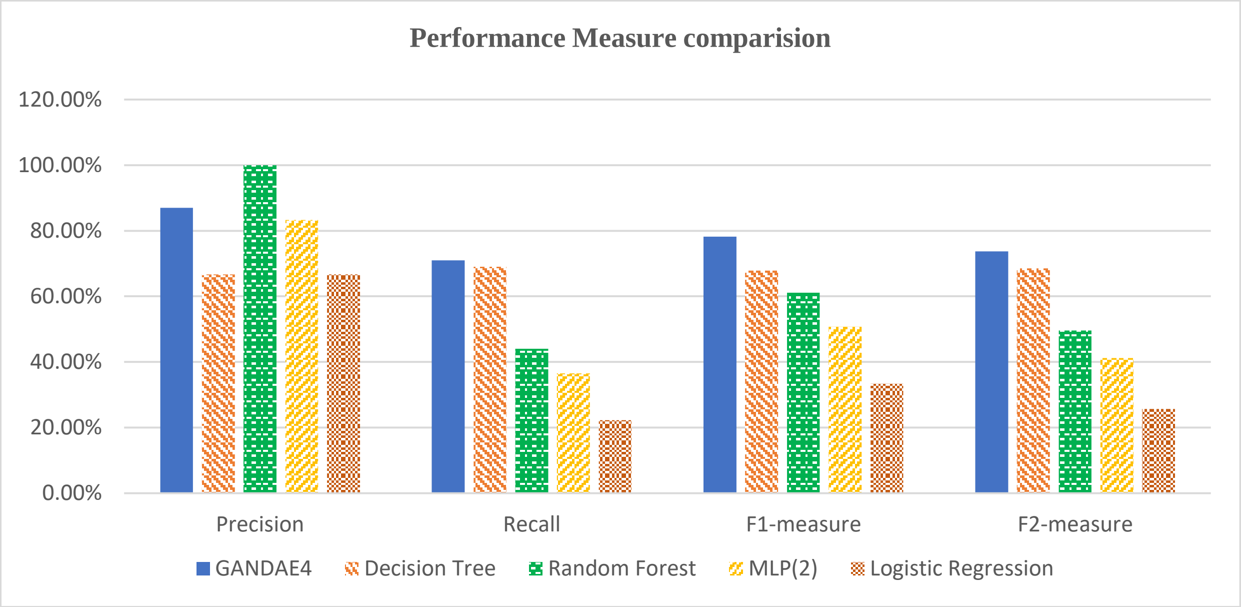

Fig. 2: Performance Measure Comparison

In this article, we use 10-fold cross-validation which means, the data are divided into 10 parts and the model runs 10 times. every time, nine parts are selected for training and one part for testing. Finally, we consider the average of 10 tests as output. therefore, about 7120 days have been used to test models, of which, 500 days are manipulated and 6620 days are unmanipulated.

Table 7: The Performance Measure of GAN-DAE4

Data	**No. of Days**	**GAN-DAE4**	**performance measure**
Manipulated	500	355	True Positive (TP)
		145	False Negative (FN)
Unmanipulated	6620	53	False Positive (FP)
		6567	True Negative (TN)
Total Data	**7120**	**7120**	-

As Table 7 shows, out of the 500 manipulated days, GAN-DAE4 has detected 355 days as manipulated correctly and 145 trading days as unmanipulated days incorrectly. Also, out of 6620, GAN-DAE4 has considered 6567 days as unmanipulated correctly and 53 days as manipulated incorrectly. So, the precision and Recall for GAN-DAE4 are about 87% (=355/ (355+53)), 70% (=355/ (355+145)) respectively. Each Precision and Recall represent the performance of the GAN-DAE4 in detecting manipulation, but to classify deep learning models, we have to use a combination of both Precision and Recall simultaneously. To do this, F1-measure which is the harmonic of Precision and Recall is used. In the GAN-DAE4, F_1-measure is about 78.19% (=2*0.87*0.71/ (0.71+0.87)). As described in Section 3.3, In the capital market, incorrect classification of manipulated and unmanipulated stocks has no symmetrical cost. therefore, in the harmonic, the Recall weight is considered twice the Precision and the F_2-measure is calculated 73.71% (=5*0.87*0.71/ (4*0.71+0.87)). For further investigation, it is necessary to explain the discriminative and Generative models. Discriminative models learn to find the decision boundary that separates the classes in an optimal way, while the generative models learn about the characteristics of each class. In the other word, discriminative models predict the labels conditioned on the input $p(y|x)$, whereas generative models learn the joint probability distribution $p(x, y)$. Examples of discriminative models include logistic regression, Support Vector Machine (SVM), Decision Tree and so on, where we can directly estimate $p(y|x)$, from the training set. Examples of generative models include Markov random fields and naive Bayes, where first we estimate $p(x, y)$ to determine $p(y|x)$. (Figure 3)

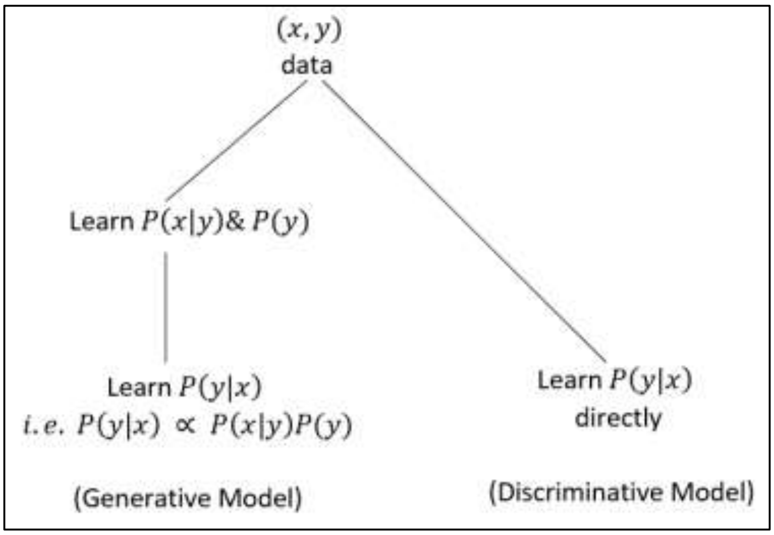

Fig. 3: Learning Discriminative and Generative Deep Learning Models. Reference [32]

GAN-based models are generative and the model proposed in the present article (GAN-DAE4) is no exception. Since the features involved in stock manipulation detection are very broad, it seems that learning the characteristics of each class (generative) is more efficient than learning to find the decision boundary (discriminative). Therefore, the GAN-DAE4 has performed better in market manipulation detection than other discriminative deep learning models like Decision Tree, Random Forest, Logistic Regression, etc. In Golmohammadi et al. [23], The best model to detect manipulation is Naïve Bayes that outperformed other supervised models. Therefore, the results of the present study are in line with the results of previous research. In the Iranian capital market, the manipulation detection has not been done in a practical way presented in this article and it is not possible to compare the results with previous research. In the “data” section, we explained that the number of manipulated data is 500, and the number of normal data is about 71,000. This asymmetry and having few positive samples make most deep learning models unable to learn manipulated stock characteristics. However, in the GAN-based Models, a lot of artificial data (which have features similar to manipulated stocks) are generated that help deep learning model train well. It is recommended to use the GAN-based model in situations where the number of positive data is significantly less than the total data. Finally, it should be noted that the model presented in this article is fully applicable and if we give the current stock information listed in the capital market (mentioned in Table 5) as input to the GAN-DAE4, the probability of stock manipulation on that day is calculated. Using historical data, the manipulation probability of Foolad and Zekesht stocks on 1398/07/20 is equal to 2% and 67%, respectively.

5 Conclusion

Stock manipulation detection is challenging research for machine learning when there are not many manipulation cases available for training. To deal with this challenge, unsupervised learning techniques had to be used in order to learn the trading data with no labels. This paper proposed a combination of Denoising Auto-encode-decode and GANs for stock manipulation detection. The proposed model denoted by GAN-DAE4 can take many features as input because of the 4-layer encoder without being concern about performance reduction of the model. Also, a little noise is added to the input features, making it possible to use the current model for new data. In other words, since the features may change slightly in the new data, our proposed model, which has already been trained with noise, takes these changes into account. The F2-measure of GAN-DAE4 was 73.71% which had a better performance than other models of deep learning. The F2-measure for Decision Tree (C4.5), Random Forest, Neural Network and Logistic Regression was equal to 68.56%, 49.55%, 41.15% and 25.64%, respectively.

In Iran, manipulated stocks are not disclosed by the regulator. This has led authors to use statistical tests including sequencing, Skewness, Kurtosis, and unsupervised methods such as contextual anomaly detection (CAD) and visual and graphical analysis to create a database of manipulated and unmanipulated stocks. For this purpose, the information of 69 stocks listed on the Tehran Stock Exchange in 2015-2020 has been analyzed. The total trading days in that period was about 71,200, of which about 500 days are manipulated days. If the number of positive data in traditional deep learning models is low, the model is usually not well trained, still, in GAN-based models, because the artificial data is generated similar to the original ones, this problem is less severe. The low Recall of the supervised models used in this study confirms this claim. Therefore, it is suggested that the proposed model in this research (GAN-DAE4) be used for situations where the positive samples are deficient.

The proposed model in this article is an unsupervised model that has a better performance in detecting manipulation even than most famous supervised models. Because the proposed model uses Auto-encode-decode models, it is possible to increase the number of features and that makes users sure that all vital features important to manipulation detection are used. Given that noise is added to the input data, this model can detect manipulation even with a slight change in the parameters, which ensures that manipulators cannot mislead the model into detecting manipulation by slightly changing the variables.

It seems that the models presented in previous Iranian researches cannot be used momentarily in the capital market for manipulation detection, but the proposed model determines the probability of stock manipulation by getting daily stock information. The Exchange and Securities Organization, as a supervisor, can use the current model to calculate the probability of stocks manipulation daily, and if it exceeds a threshold, can close the symbol and investigates further. Investors and investment managers can also use the current model to be aware of the possibility of stock manipulation and avoid buying stocks that are likely to be manipulated. Due to the rapid and significant boom in algorithmic trading, the designers of these algorithms should consider manipulation detection models like GAN-AED4 to be sure that manipulators cannot mislead those algorithmic models by making deceptive actions and make unusual profits. There are several ways to validate deep learning models. K-fold cross-validation is undoubtedly one of the most effective validation tests. In this article, we use 10-fold cross-validation which means, the data are divided into 10 parts and the model runs 10 times. every time, nine parts are selected for training and one part for testing. Cross-validation helps us be sure that the results are constant and reliable. Also, to select the appropriate models for comparison with the proposed model, we have used several different deep learning models.

For example, Support Vector Machine has been conducted, which has been omitted due to poor performance compared to other models. Therefore, we can be sure that most of the famous deep learning models have been conducted in the present study. Like many studies, this one has its limitations. The most important limitation is the classification of stocks into manipulated and unmanipulated by hand. Although different methods have been used to ensure that labeling data is correct, this type of labeling may be associated with errors. In this research, stocks listed on the stock exchange are examined and can’t be generalized to other types of companies. Future researches can be stated as follows: It is suggested that the present model (GAN-DAE4) be used in the other financial sectors where abnormal data are significantly less than normal data and conventional deep learning models cannot be trained sufficiently.

Due to the expansion of algorithmic trading in the Tehran Exchange Securities, it is suggested that the deep learning models (specially GAN-AED4) be examined with intraday data to detect stock manipulation in the minute timeframes. It is suggested that the Semi-supervised GAN-based models be used to detect stock manipulation, as this would probably make the model better able to detect manipulation. See Zheng et al. [47] for more information. One of the limitations of using deep learning models to manipulation detection is labeling data by hand. To overcome this problem, free-manipulated stocks (like Foolad, Mellat Bank, etc.) can be selected and random data can be injected to stock time series as anomalies. Then using deep learning models to anomalies detection. For more information, refer to Golmohammadi et al. [24]. It is suggested to study the use of evolutionary algorithms to improve the training efficiency for such complex deep networks with a large number of parameters. For more information, refer to Zheng et al. [47].

References

[1] Al-Thani, H.A., *Detecting market manipulation in stock market data, Detecting market manipulation in stock market data,* 2017. Doi: https://www.proquest.com/openview/8d73380cd0dd690f46a37b2db715def0/1?pq-origsite =gscholarandcbl=2026366

[2] Allen, F.,Gale, D., *Stock-price manipulation.* The Review of Financial Studies, 1992, **5**(3), P. 503-529. Doi: 10.1093/rfs/5.3.503

[3] Amini, M., Isaei Tafreshi, M., Shahbazinia, M., *Securities Market Manipulation: Concept, Principles, and Methods (A Comparative Study of Iranian and American Law),* Comparative Law Research, 2014, **19**, P. 1-26. (in persian)

[4] Banimahd, B., *The Ranking of Fraudulent Financial Reporting By Using Data Envelopment Analysis: Case of Pharmaceutical Listed Companies,* Advances in Mathematical Finance and Applications, 2020, **5**(1), P.69-80. Doi: 10.22034/amfa.2019.1863571.1193

[5] Ben-David, I., *Do hedge funds manipulate stock prices?,* The Journal of Finance, 2013, **68**(6), P. 2383-2434. Doi: 10.1111/jofi.12062

[6] Carhart, M. M., Kaniel, R., Musto, D. K., Reed, A. V., *Leaning for the tape: Evidence of gaming behavior in equity mutual funds*, The Journal of Finance, 2002, **57**(2), P. 661-693. Doi: 10.1111/1540-6261.00438

[7] Cherian, J.,Kuriyan, V., *Informationless manipulation in a market maker-type economy*, 1995, Available at SSRN 6501. Doi: https://ssrn.com/abstract=6501

[8] Cherian, J.A.,Jarrow, R.A., *Market manipulation*, Handbooks in Operations Research and Management Science, 1995, **9**, P. 611-630. Doi: 10.1016/S0927-0507(05)80064-7

[9] Comerton-Forde, C., Putniņš, T.J., *Measuring closing price manipulation*, 2011, Journal of Financial Intermediation, **20**(2), P. 135-158. Doi: 10.1016/j.jfi.2010.03.003

[10] Comerton-Forde, C., Putniņš, T.J., *Stock price manipulation: Prevalence and determinants*, Review of Finance, 2014, **18**(1), P.23-66. Doi: 10.1093/rof/rfs040

[11] Diaz, D., Theodoulidis, B., Sampaio, P., *Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices*, Expert Systems with Applications, 2011, **38**(10), P.12757-12771. Doi: 10.1016/j.eswa.2011.04.066

[12] Ergün, H. O., Yalaman, A., Manahov, V., and Zhang, H., *Stock market manipulation in an emerging market of Turkey: how do market participants select stocks for manipulation?* 2020, Applied Economics Letters, P. 1-5. Doi: 10.1080/13504851.2020.1753874

[13] Falahzadeh, H., *Detection of stock manipulation using data mining techniques*, University of Science and Art affiliated with University Jahad. (in persian)

[14] Fallah Shams Layalestani, M., Badpa, A., Badpa, I., *Designing a model to identify the possibility of final price manipulation in the Tehran Stock Exchange*, Financial Engineering and Securities Management, 2015, **6**, P. 135-152. (in persian)

[15] Fallah Shams, M.,Teymouri Shandi, A., *Designing a price manipulation forecasting model in Tehran Stock Exchange*, Research Quarterly of Imam Sadegh University, 2005, 27. (in persian)

[16] Fallah Shams, M., Kordloui, H.R., *Test of logit models and artificial neural network to predict price manipulation in Tehran Stock Exchange*, Financial Engineering and Securities Management, 2011, **2**(7), P. 37-69. (in persian)

[17] Fallah Shams, M., Kordloui, H.R., Rashno, M., *Investigating price manipulation in Tehran Stock Exchange using the support vector machine model*, Financial Research, 2012, **14**(1), P. 69-84. (in persian)

[18] Farshadfar, Z., Prokopczuk, M., *Improving Stock Return Forecasting by Deep Learning Algorithm*, Advances in Mathematical Finance and Applications, 2019, **4**(3), P.1-13. Doi: 10.22034/amfa.2019.584494.1173

[19] Felixson, K., Pelli, A., *Day end returns—stock price manipulation*, Journal of Multinational Financial Management, 1995, **9**(2), P. 95-127. Doi: 10.1016/s1042-444x(98)00052-8

[20] Ferdousi, Z., Maeda, A., *Anomaly Detection Using Unsupervised Profiling Method in Time Series Data*, in ADBIS Research Communications, 2006. Doi: 10.1.1.89.4348andrep=rep1andtype=pdf

[21] Fischel, D.R., Ross, D.J., *Should the Law Prohibit" Manipulation" in Financial Markets?* Harvard Law Review, 1991, **105**(2), P.503-553. Doi: 10.2307/1341697

[22] Ghorbani, M., Bagheri, A., *Securities market manipulation*, Public Law Research, 2008, **12**, P. 301-326. (in persian)

[23] Golmohammadi, K., Zaiane, O.R., Díaz, D., *Detecting stock market manipulation using supervised learning algorithms*, in 2014 International Conference on Data Science and Advanced Analytics (DSAA), 2014, IEEE. Doi: 10.1109/DSAA.2014.7058109

[24] Golmohammadi, K., Zaiane, O.R., *Time series contextual anomaly detection for detecting market manipulation in stock market*, in 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015, IEEE. Doi: 10.1109/DSAA.2015.7344856

[25] Goodfellow, I.J., *Generative adversarial networks.* arXiv preprint arXiv:1406.2661, 2014. https://papers.nips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

[26] Griffin, J.M., Shams, A., *Manipulation in the VIX?* The Review of Financial Studies, 2018, **31**(4), P. 1377-1417. Doi: 10.2139/ssrn.2972979

[27] Hart, O.D., *On the profitability of speculation.* The Quarterly Journal of Economics,1977, P. 579-597. Doi: 10.2307/1885883

[28] Hillion, P., Suominen, M., *The manipulation of closing prices,* Journal of Financial Markets, 2004, **7**(4), P. 351-375. Doi: 10.1016/j.finmar.2004.04.002

[29] Jarrow, R.A., *Market manipulation, bubbles, corners, and short squeezes*, Journal of financial and Quantitative Analysis, 1992, P.311-336. Doi: 10.2307/2331322

[30] Kazemi Tabar, S.J., Shahbazzadeh, M., *Detection of fraud in the stock market using the Chebyshev inequity,* Quarterly Journal of Signal and Data Processing, 2019, **17**, P. 3-14. (in persian)

[31] Leangarun, T., Tangamchit, P., Thajchayapong, S., *Stock Price Manipulation Detection using Generative Adversarial Networks,* in 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018, IEEE. Doi: Doi:10.1109/SSCI.2018.8628777

[32] Lee, E.J., Eom, K.S., Park, K.S., *Microstructure-based manipulation: Strategic behavior and performance of spoofing traders,* Journal of Financial Markets, 2013, **16**(2), P. 227-252. Doi: 10.1016/j.finmar.2012.05.004

[33] MehdiYekeZare, M., *manipulation in rights offering trading on the Tehran Exchange Securities*, Sharif University of Technology, 2019. (in persian)

[34] Nadiri, M., Alavi Nasab, S., Peymani, M., Rabiee, R., *Investigating the effect of some factors affecting price manipulation in the Iranian stock market,* Financial Research, 2018, **20**(3), P. 327-342. (in persian)

[35] Öğüt, H., Doğanay, M.M., Aktaş, R., *Detecting stock-price manipulation in an emerging market: The case of Turkey.* Expert Systems with Applications, 2009, **36**(9), P. 11944-11949. Doi: 10.1016/j.eswa.2009.03.065

[36] Phua, C., Lee, V., Smith, K., Gayler, R., *A comprehensive survey of data mining-based fraud detection research.* arXiv preprint arXiv, 2010. Doi: 10.1016/j.chb.2012.01.002

[37] Poostforooshan, M.H., Sadrabadi, N., Moinuddin, M., *Evaluation of price manipulation using quadratic discriminant analysis models and genetic algorithm based on artificial neural network*, Financial Knowledge of Securities Analysis, 2014, P. 119-143. (in persian)

[38] Putnins, T., *An Overview of Market Manipulation,* Corruption and Fraud in Financial Markets Malpractice, Misconduct and Manipulation, 2020. Doi: https://ssrn.com/abstract=3398258

[39] Rabiee, R., Nadiri, M., Peymani, M., Jaberizade, A., *Investigating the effect of price manipulation on market efficiency in Tehran Stock Exchange,* Stock Exchange Quarterly, 2018, **11**(43), P. 113-131. (in persian)

[40] Rahmani, A., Zanjirdar, M., Ghiabi, H., *Effect of Peer Performance, Future Competitive Performance, and Factors of Correlation with Peer Companies on Manipulation of Abnormal Real Operations*, Advances in Mathematical Finance and Applications, 2021, **6**(1), P.57-70. Doi: 10.22034/amfa.2020.1875478.1272

[41] Roodposhti, F.R., Shams, M.F., Kordlouie, H., *Forecasting stock price manipulation in capital market.* International Journal of Economics and Management Engineering, 2011, **5**(8), P. 957-967. Doi: 10.5281/zenodo.1329717

[42] Takahashi, S., Chen, Y., Tanaka-Ishii, K., *Modeling financial time-series with generative adversarial networks*, Physica A: Statistical Mechanics and its Applications, 2019, **527**, P. 121261.
Doi: 10.1016/j.physa.2019.121261

[43] Tohidi, M., Mousavian, S.A., *Jurisprudential study and analysis of stock exchange based on price manipulation in the securities market*, Stock Exchange Quarterly, 2018, **11**, P. 79-113. (in persian)

[44] Westphal, C., Blaxton, T., *Data mining solutions: methods and tools for solving real-world problems*, John Wiley and Sons, Inc. ISBN: 978-0-471-25384-6

[45] Xu, I., Livshits, B., *The anatomy of a cryptocurrency pump-and-dump scheme*, In28th USENIX Security Symposium (USENIX Security 19), 2019, P.1609–1625. https://www.usenix.org/system/files/sec19-xu-jiahua_0.pdf

[46] Zhai, J., Cao, Y., and Ding, X., *Data analytic approach for manipulation detection in stock market*. Review of Quantitative Finance and Accounting, 2018, **50**(3), P.897-932. Doi: 10.1007/s11156-017-0650-0

[47] Zheng , Y.J., Zhou, X.H., Sheng, W.G., Xue, Y., Chen, S.Y., *Generative adversarial network based telecom fraud detection at the receiving bank.* Neural Networks, 2018, **102**, P. 78-86. Doi: 10.1016/j.neunet.2018.02.015

[48] Zheng, Y. J., Ling, H. F., Xue, J. Y., and Chen, S. Y., *Population classification in fire evacuation: A multi-objective particle swarm optimization approach*, IEEE Transactions on Evolutionary Computation, 2013, **18**(1), P.70-81. Doi: 10.1109/TEVC.2013.2281396

[49] Interview with Capital Market Expert (Part 1), *How is the loss of microfinance on the stock exchange minimized*?, https://plus.irna.ir/news/84140636/, 2021.