Publisher: Shahid Beheshti University
Journal of Financial Management Perspective
2024, Vol. 14, No.46
PP: 133-161
DOI: 10.48308/jfmp.2024.105131
P-ISSN: 2645-4637
E-ISSN: 2645-4645
Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by /4.0/)

Stock Price Manipulation in the Iran Stock Market Using VAE-LSTM Hybrid Model

Seyed Mohammadreza Habibzadeh*, Mohammad Ail Rastegar, Reza Golami Jamkarni***, Sayyed Kazem Chavoshi*****

Research Paper

Abstract

Purpose: The stock market, as one of the main economic sectors of countries, plays an important role in the development and expansion of economic activity. With the development of technology and complex trading algorithms, stock manipulation has become more easily, which makes the use of tools such as artificial intelligence and deep learning to identify manipulation by supervise institutions inevitable. The aim of this research is to identify stock manipulation in the Iran stock market. For this purpose, information on 73 stocks from 19 industries admitted to the stock exchange during 1398 to 1402, approximately 71,300 trading days, was used.

Method: Identifying manipulation in stock transactions poses a significant challenge due to the temporal correlation of stock price data and its dynamic. This challenge is also exacerbated by the unavailability of labeled data. Therefore, given the lack of announcement of manipulated stocks by the stock exchange supervise in the Iran stock market, data identification: 1) Statistical tests such as abnormal returns, manipulated stocks, and the exact date of manipulation have been determined. 2) Random data simulating the stock manipulation pattern has been injected into the time series of stocks that have not been manipulated with high confidence (expert questionnaire). In the next step, using a combination of variable autoencoding models and long short-

Received: 2024. September. 01 Accepted: 2024. December. 05.
* Ph.D. Candidate in Financial Engineering, Qom branch, Islamic Azad University, Qom, Iran. Email: em.habibzadeh@gmail.com
** Assistant Prof., Department of System and Productivity Management, Tarbiat Modares University, Tehran, Iran (Corresponding Author). Email: ma_rastegar@modares.ac.ir
*** Assistant Prof., Department of Accounting, Qom branch, Islamic Azad University, Qom, Iran. Email: reza.gholami@iau.ac.ir
**** Assistant Prof., Department of Financial Management, University of Kharazmi, Tehran, Iran. Email: chavoshi@khu.ac.ir

term memory, the VAE-LSTM algorithm has been designed to compare with some machine learning models such as decision tree, random forest, logistic regression, etc., which calculates the probability of stock manipulation.

Findings: After running the models, the accuracy and recall indices and F_1 and F_2 were calculated. Because in the stock market, the classification of manipulated and unmanipulated stocks is not of equal importance, the performance evaluation index $F_۲$ has been used to rank the models. In order, the VAE-LSTM, decision tree, random forest, multilayer neural network, support vector machine, and logistic regression models showed better performance. The approximate F_2 values of the mentioned models are: 72%, 69%, 50%, 41%, 4 % and 26%, respectively. Findings: After implementing the deep learning models, the accuracy and recall indices and F_1 and F_2 were calculated. Because in the capital market, the classification of manipulated and unmanipulated stocks is not of equal importance, the performance evaluation index F_2 was used to rank the models. The VAE-LSTM, decision tree, random forest, multilayer neural network, support vector machine, and logistic regression models performed better, respectively. The approximate F_2 values of the aforementioned models were: 72%, 69%, 50%, 41%, 40%, and 26%. After the VAE-LSTM hybrid model, the decision tree model is ranked next, which also has a good balance between the accuracy and recall indices. This indicates that perhaps one of the most effective ways to identify manipulation is to use predetermined rules that are extracted by decision tree models and can be updated at different time intervals.

Conclusion: Finally, the proposed model based on the F_2 performance evaluation index has shown a better ability to detect manipulation than other models. It is important to note that other machine learning models used in this study also performed well, especially in the accuracy evaluation index, but unfortunately, they performed poorly in terms of the more important recall performance index. After determining the proposed model as the selected model, based on the Tehran Stock Exchange's total index, we considered the capital market's bullish period in the period from 1398/12/01 to 1399/05/31, the capital market's bearish period in the period from 1399/05/21 to 1399/08/20, and the year 1400 as the capital market's equilibrium period. As expected, the probability of manipulation is higher in bullish, balanced, and bearish markets, respectively. These results are generally consistent with other previous studies. The results are conceptually consistent with reality. Since short selling is not possible in the Iranian capital market, manipulators can only make a profit by manipulating by “raising the price and emptying” and it is not possible to use the manipulation method of “lowering the price and buying back”. Therefore, in a bear market, creating a trend change in the capital market requires a lot of resources, which reduces the incentive to manipulate the share.

Keywords: Capital market, Stock price manipulation, Deep learning, VAE - LSTM

ناشر: دانشگاه شهیدبهشتی
نشریه چشم‌انداز مدیریت مالی
۱۴۰۳، دوره ۱۴، شماره ۴۶
ص‌ص. ۱۳۳- ۱۶۱

شاپای چاپی: ۴۶۳۷- ۲۶۴۵
شاپای الکترونیکی: ۴۶۴۵- ۲۶۴۵

Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

بررسی دستکاری قیمت‌ها در بازار بورس ایران با استفاده از مدل ترکیبی خودرمزگذار متغیر-حافظه کوتاه مدت طولانی

سید محمدرضا حبیب‌زاده*، محمدعلی رستگار، رضا غلامی جمکرانی***، سید کاظم چاوشی*****

مقاله پژوهشی

چکیده

هدف: بازار سرمایه به عنوان یکی از اصلی‌ترین بخش‌های اقتصادی کشورها، نقش با اهمیتی در توسعه و گسترش فعالیت اقتصادی دارد. با توسعه تکنولوژی و الگوریتم‌های معاملاتی پیچیده، دستکاری سهام با سهولت بیشتری رخ داده و این امر سبب می‌شود تا استفاده از ابزارهای مانند هوش مصنوعی و یادگیری عمیق برای شناسایی دستکاری توسط نهادهای ناظر ناگزیر باشد. هدف از این پژوهش، شناسایی دستکاری سهام در بازار بورس ایران است. برای این کار از اطلاعات ۷۳ سهم از ۱۹ صنعت پذیرش شده در بورس طی ۱۳۹۸ الی ۱۴۰۲، روزهای معاملاتی تقریبی ۷۱،۳۰۰ استفاده شده است.

روش: شناسایی دستکاری در معاملات سهام به دلیل همبستگی زمانی داده‌های قیمت سهام و پویا بودن آن، چالش مهمی را به وجود می‌آورد. این چالش با در دسترس نبودن داده‌های برچسب‌گذاری‌شده نیز تشدید می‌شود. از این رو باتوجه به عدم اعلام سهام دستکاری شده توسط ناظر بورس در بازار ایران، شناسایی داده‌ها: ۱) آزمون‌های آماری مانند بازدهی غیرنرمال، سهام

تاریخ دریافت مقاله: ۱۴۰۳/۰۶/۱۱، تاریخ پذیرش مقاله: ۱۴۰۳/۰۹/۱۵

* دانشجوی دکتری مهندسی مالی، واحد قم، دانشگاه آزاد اسلامی، قم، ایران.
E-Mail: em.habibzadeh@gmail.com
** استادیار، گروه مدیریت سیستم و بهره‌وری، دانشگاه تربیت مدرس، تهران، ایران (نویسنده مسئول).
E-Mail: rastegar@modares.ac.ir
*** دانشیار، گروه حسابداری، واحد قم، دانشگاه آزاد اسلامی، قم، ایران.
E-Mail: reza.gholami@iau.ac.ir
**** استادیار، گروه مدیریت بانک و بیمه، دانشگاه خوارزمی، ایران.
E-Mail:chavoshi@khu.ac.ir

دستکاری شده و تاریخ دقیق دستکاری مشخص شده است. ۲) داده‌های تصادفی با شبیه‌سازی الگوی دستکاری سهم، به سری زمانی سهامی که با اطمینان بالایی در آن‌ها دستکاری رخ نداده (پرسشنامه خبرگان)، تزریق شده است. در گام بعدی با استفاده از ترکیب مدل‌های خودرمزگذار متغیر و حافظه کوتاه مدت طولانی، الگوریتم VAE- LSTM برای مقایسه با برخی مدل‌ها یادگیری ماشین از قبیل درخت تصمیم، جنگل تصادفی، رگرسیون لجستیک و.. طراحی گردیده که احتمال دستکاری سهم را محاسبه می‌نماید.

یافته‌ها: پس از اجرای مدل‌های یادگیری عمیق شاخص‌های دقت و بازخوانی و F_1 و F_2 محاسبه شد. به دلیل اینکه در بازار سرمایه دسته‌بندی سهام دستکاری شده و نشده از اهمیت یکسانی برخوردار نیست، برای رتبه‌بندی مدل‌ها از شاخص ارزیابی عملکردی F_2 استفاده شده است. به ترتیب مدل‌های VAE-LSTM، درخت تصمیم، جنگل تصادفی، شبکه عصبی چندلایه، ماشین‌بردار پشتیبان، رگرسیون لجستیک عملکردی بهتری از خود نشان دادند. مقدار حدودی F_2 مدل‌های مذکور به ترتیب : ۷۲درصد، ۶۹درصد، ۵۰درصد، ۴۱درصد، ۴۰درصد و ۲۶درصد است.

نتیجه‌گیری: در نهایت مدل پیشنهاد شده براساس شاخص ارزیابی عملکرد F_2 نسبت به سایر مدل‌ها توانایی بهتری در شناسایی دستکاری از خود نشان داده است. ذکر این نکته ضروری است که سایر مدل‌های یادگیری ماشین استفاده شده در این پژوهش نیز عملکرد مناسبی به خصوص در شاخص ارزیابی دقت داشته اما متأسفانه از نظر شاخص عملکردی بازخوانی که مهم‌تر می‌باشد، عملکرد ضعیف تری داشته‌اند. پس از تعیین مدل پیشنهادی به عنوان مدل برگزیده، براساس شاخص کل بورس تهران، دوره صعودی بازار سرمایه را در بازه زمانی ۱۳۹۸/۱۲/۰۱ تا ۱۳۹۹/۰۵/۳۱، دوره نزولی بازار سرمایه را در بازه زمانی ۱۳۹۹/۰۵/۲۱ تا ۱۳۹۹/۰۸/۲۰ و سال ۱۴۰۰ را به عنوان دوره تعادل بازار سرمایه در نظر گرفته‌ایم که مطابق با انتظار، احتمال دستکاری به ترتیب در بازارهای صعودی، متعادل و نزولی بیشتر است.این نتایج به طور کلی با سایر مطالعات پیشین نیز هم‌سو می‌باشد.

کلیدواژه‌ها: بازار سرمایه، دستکاری قیمت، یادگیری عمیق، VAE - LSTM

۱. مقدمه

بازار سرمایه به عنوان یکی از اصلی‌ترین بخش‌های اقتصادی کشورها، نقش با اهمیتی در توسعه و گسترش فعالیت اقتصادی دارد. بازار سرمایه مکانی برای خرید و فروش اوراق بهادار بوده و توسعه و تعمیق بازار نیازمند اعتماد تمامی ذی‌نفعان به کارایی و درستی آن در تعیین عادلانه قیمت و به دور از هر عمل متقلبانه دیگری است. متأسفانه برخی بازیگرانی با انجام اقدامات فریبکارانه، سایر بازیگران را به گمراهی کشیده و از این محل کسب درآمد می‌نمایند. به طور کلی اقداماتی که منجر به گمراهی فعالان بازار گردد، دستکاری بازار نامیده می‌شود (لینگارن[1] و همکاران، ۲۰۱۸). دستکاری قیمت سهام تبدیل به موضوع بسیار مهمی در بازار سرمایه کشورهای توسعه‌یافته و در حال توسعه شده است. یکی از مهم‌ترین دلایل انحراف قیمت سهم از قیمت منصفانه آن، دستکاری می‌باشد (آگت[2] و همکاران، ۲۰۰۹).

عدم توانایی در کشف دستکاری قیمت، به کارایی بازار صدمات جبران ناپذیری وارد می‌نماید. از این رو نهادهای نظارتی همواره در تلاشند تا با وضع قوانین و دستورالعمل، ضمن پیشگیری از وقوع دستکاری، مجازات‌های متناسب را تعیین نمایند.

با افزایش ابعاد و پیچیدگی‌های معاملاتی، امکان نظارت و شناسایی به صورت سنتی امکان‌پذیر نمی‌باشد. از این رو می‌بایست با بهره‌گیری از رویکردهای شناسایی مبتنی بر هوش مصنوعی ضمن افزایش سرعت در شناسایی نمونه‌های با حجم بالا، سازو کاری فراهم نماید تا اپراتور انسانی نیز متناسب با تشخیص انجام شده پاسخ دهد (دنگ و هوی[3]، ۲۰۲۱).

در بازار سرمایه ایران علی‌رغم وجود دستورالعمل‌ها و قوانین متعدد برای جلوگیری از بروز دستکاری سهام، متأسفانه به نظر تعداد زیادی از فعالان بازار، بسیاری از سهام مورد معامله حداقل در مقاطعی از زمان در معرض دستکاری بوده یا به طور جدی دستکاری شده‌اند. این در حالی است که با توسعه تکنولوژی و الگوریتم‌های معاملاتی پیچیده در معاملات سهام این اتفاق محتمل‌تر خواهد بود، بنابراین لزوم بهره‌گیری از ابزارهای توسعه‌یافته مبتنی بر یادگیری ماشین برای شناسایی دستکاری قیمت‌ها در بازار سرمایه توسط نهاد نظارتی ضروری تلقی می‌گردد.

در این پژوهش برای شناسایی دستکاری قیمت سهام در بازار سرمایه ایران، ضمن ترکیب مدل خودرمزگذار با مدل اصلی، متغیرهای بهینه به صورت خودکار توسط شبکه انتخاب شده و شبکه بر پایه آن‌ها آموزش می‌بینند که باعث عملکرد بهینه درصورت بروز نویز نیز خواهد شد. همچنین برای برچسب‌گذاری سهام دستکاری‌شده علاوه بر متغیرهای بنیادی از متغیرهای تکنیکال و

---

[1] Leangarun
[2] Öğüt
[3] Deng & Hooi

نموداری نیز استفاده نموده‌ایم. این اتفاق اطمینان خاطر بیشتری در پایگاه داده ایجاد می‌کند. در نهایت نیز روش پیشنهادی، باتوجه به عملکرد مقایسه‌ای بهتر نسبت به سایر مدل‌های مورد بررسی، علاوه بر ناظر بازار به سایر فعالین کمک خواهد کرد تا زودتر از دستکاری سهم مطلع شده و تصمیم لازم را اتخاذ نمایند.

۲. مبانی نظری و پیشینه پژوهش

دستکاری قیمت در بازار اوراق بهادار از دیرباز یکی از چالش‌های اصلی برای داشتن معاملات سالم و بدون تقلب بوده است و به زمان شکل‌گیری اولین معاملات در شکل نوین آن یعنی بازار اوراق بهادار آمستردام باز می‌گردد. بورس اوراق بهادار از اصلی‌ترین بازارها در فضای اقتصاد کشورهاست و یکی از با اهمیت‌ترین پیش نیازهای گسترش و تعمیق بازار سرمایه، اعتماد بازیگران و ذی‌نفعان این بازار به کارایی و درستی آن به عنوان مبنایی برای تعیین عادلانه و به دور از تقلب قیمت دارایی‌های مالی است و میزان اعتماد عمومی به بازار بورس, به عنوان یکی از عوامل اصلی تاثیرگذار بر تصمیم‌گیری سرمایه‌گذاران محسوب می‌گردد.

امکان تعریف دقیقی از دستکاری سهام دشوار است چرا که از یک طرف دستکاری، مجموعه وسیعی از استراتژی‌های معاملاتی که به سرعت تغییر می‌یابد را شامل شده و از طرف دیگر، درصورتی که تعریفی دقیق و قانونی ارائه گردد، برای دستکاری‌کنندگان این امکان فراهم می‌گردد تا با استفاده از طرح‌ها و استراتژی‌های معاملاتی که خارج از تعریف مذکور قرار می‌گیرد، اقدام به دستکاری سهام نمایند (پوتنینس[1]، ۲۰۲۰). با این وجود به منظور جلوگیری از تبعات منفی این عمل تعاریف مختلفی از دستکاری سهام در ادبیات موضوع و قوانین مربوطه ارائه شده است.

از اوایل قرن ۱۷ میلادی تاکنون شاهد پژوهش‌های فراوانی به منظور شناسایی عوامل رخداد، روش‌های شناسایی و راهکارهای جلوگیری از دستکاری بازار بوده‌ایم. پژوهش (هارت[2]، ۱۹۷۷) یکی از اولین پژوهش‌های مستند در خصوص دستکاری از دید تئوری است که در آن امکان و شرایط دستکاری بحث شده است. او شرایطی را که سفته بازی در یک وضعیت غیراحتمالی ممکن است رخ دهد را مورد بررسی قرار داد. وی دریافت که اگر بازار در حالت تعادل پایدار نباشد و توابع غیرخطی باشند، سفته بازان قادر به انجام معاملات سودآور خواهند بود.

مطالعات دستکاری بازار با پژوهش (آلن و گیل، ۱۹۹۲) و بررسی احتمال دستکاری بر پایه معامله را در چارچوب (نظریه) انتظارات منطقی منجر به معرفی دسته‌بندی دستکاری سهام به سه دسته:

[1] Putnins
[2] Hart

بر پایه اطلاعات، بر پایه سهام و بر پایه معاملات شد. پس از آن نیز (ژارو[1]، ۱۹۹۲) به معرفی استراتژی‌های دستکاری سودآور در بازار اشاره کرد.

مطالعات در زمینه دستکاری معامله‌محور، به بررسی شرایطی پرداختند که در آن، دستکاری در یک بازار سودآور است یا خیر. فیچل و راس[2] (۱۹۹۲) ادعا کردند که دستکاری معامله‌محور در بازار کارا، امکان‌پذیر نیست. ژارو (۱۹۹۲) و چریان و کوریان[3] (۱۹۹۵) بر پایه مدل هارت، شرایطی را یافتند که در آن، امکان دستکاری معامله‌محور وجود ندارد. ژارو نتیجه گرفت شرط کافی برای اینکه استراتژی دستکاری بازار ناموفق باشد، این است که قیمت‌ها تنها براساس سهام در اختیار معامله‌گران و نه بر اساس روند گذشته معاملاتی سهم تعیین شود؛ به عبارت دیگر سهام، دارای تکانه آنی[4] نباشند. پس از آن بسیاری از محققین در تلاش بودند تا شرایطی که دستکاری معامله‌محور منجر به کسب منفعت می‌شود را بررسی نمایند. آلن و گورتون[5] (۱۹۹۲) بحث کردند که عدم تقارن اطلاعاتی سفارش‌ها، احتمال کسب سود از طریق دستکاری معامله‌محور را افزایش می‌دهد. بعضی از مطالعات نشان دادند که چگونه اوراق بهادر مشخصی (مانند مشتقات)، اقدامات شرکتی (مانند افزایش سرمایه) و مواردی از این دست، می‌تواند سبب افزایش سود دستکاری معامله‌محور شود. ژارو، (۱۹۹۴) به ذکر استراتژی‌هایی پرداخت که به سبب ابزارهای مشتقه ایجاد می‌شوند. کیل[6] (۱۹۸۴) ، ویلا[7] (۱۹۸۷) و آلن و همکاران (۲۰۰۶)، دستکاری به روش انحصار و تحت فشار قرار دادن را که در آن دستکاری‌کنندگان درصد بالایی از عرضه را در اختیار می‌گیرند و بازار را با کمبود عرضه مواجه می‌کنند، مورد بررسی قرار دادند. بگنولی و لیپمن[8] (۱۹۹۶) مدلی را توسعه دادند که در آن دستکاری‌کننده با ایجاد شایعات غلط، سبب افزایش قیمت شده و کسب سود می‌نماید. به منظور تکرار دستکاری به این شکل، اگر بازار تشخیص دهد که مرجع اطلاعات نادرست، دستکاری‌کننده بوده، آن فرد بی‌اعتبار شده و استراتژی دستکاری وی در آینده با مشکل مواجه می‌شود.

مطالعات بعدی، الگوهای فصلی و ناهنجاری در معاملات انتهای روز و پایانی را به دستکاری قیمت پایانی ارتباط دادند. کومار و سپی[9] (۱۹۹۲) ساختاری را تشریح کردند که در آن دستکاری‌کننده، در قرارداد آتی تعدادی زیادی موقعیت خرید گرفته و سپس با خرید مداوم دارایی

---

[1] Jarrow
[2] Fischel & Ross
[3] Cherian & Kuriyan
[4] Momentum
[5] Allen & Gorton
[6] Kyle
[7] Vila
[8] Bagnoli & Lipman
[9] Kumar & Seppi

پایه قیمت آن را افزایش داده و هم‌زمان با افزایش قیمت قرارداد آتی به واسطه افزایش دارایی پایه، با گرفتن موقعیت معکوس در قرارداد آتی کسب سود می‌نماید.

گاهی اوقات طراحی پارامترهای بازار به‌گونه‌ای است که برای دستکاری انگیزه ایجاد می‌نماید. به عنوان مثال در بازارهای مالی، کاملاً پذیرفته شده که قیمت آخرین معامله و دیگر پارامترهای تأثیرگذار بر شاخص، بسیار مستعد دستکاری اند. فلیکسون و پلی[۱] (۱۹۹۹) با استفاده از تحلیل رگرسیونی، دستکاری آخرین قیمت را بررسی کردند. نتایج نشان داد که دستکاری در قیمت پایانی اتفاق می‌افتد و این نتایج همسو با مفروضات بود. نتایج مطالعات هیلیون و سومینن[۲] (۲۰۰۴) نشان داد که در لحظات پایانی معاملات در بورس پاریس، نوسانات، حجم معاملات و اختلاف مظنه خرید و فروش به میزان معناداری افزایش می‌یابد که پژوهش‌گران این تغییرات را به دستکاری قیمت مربوط دانستند. یافته‌های کارهارت و همکاران[۳] (۲۰۰۲) نشان می‌دهد که در بازار سرمایه آمریکا، افزایش قیمت در طول روز، اکثراً در نیم ساعت پایانی بازار اتفاق افتاده و شدت افزایش قیمت در یک چهارم پایانی بیشتر است. برنهارد و دیویس[۴] (۲۰۰۹) نشان دادند که مدیران صندوق‌های سرمایه‌گذاری انگیزه لازم برای دستکاری آخرین قیمت سهام موجود در پرتفو را با نزدیک شدن به فصل گزارشگری دارند به نحوی که با افزایش قیمت سهام به روش دستکاری، ارزش دارایی‌های موجود در صندوق در کوتاه‌مدت بالا نشان داده شود.

فریدر و زیترین[۵] (۲۰۰۶) و هانک و هاوسر[۶] (۲۰۰۶) اثربخشی روش دستکاری تبلیغ زیاد و فروش سهام را بررسی کردند. آن‌ها به این نتیجه رسیدند که فرستنده ایمیل تبلیغاتی، سهام را در روز قبل از آغاز کمپین خریداری نموده و سپس آن‌ها را پس از کسب بازده به فروش می‌رساند. براساس نتایج آن‌ها، تبلیغات تأثیر قابل ملاحظه‌ای به‌روی بازدهی، نوسان، شکاف عرضه و تقاضای بین روز و حجم معامله سهم دارند. به‌طور خاص، آن‌ها نشان دادند که حجم معامله در سهام تبلیغ‌شده به میزان قابل ملاحظه‌ای در روز و حوالی روزهای تبلیغات بیشتر است.

پژوهش آگراوال و وو (۲۰۰۶) جز اولین پژوهش‌ها در زمینه بررسی سهام دستکاری شده بود. این مطالعه، پایگاه داده‌ای از سهام دستکاری شده ناشی از صدور آرا مبنی بر رخ دادن دستکاری از سال ۱۹۹۰ تا ۲۰۰۱ را تشکیل دادند. دستکاری‌های اتفاق افتاده در این پایگاه داده از نوع دستکاری به روش افزایش قیمت و خالی کردن بود. بر این اساس، حداقل، میانگین و حداکثر طول دوره دستکاری سهم به ترتیب ۲، ۲۰۲ و ۱۳۷۳ روز بوده که بیانگر گستردگی دستکاری به روش مذکور است. نتایج نشان می‌دهد که در مورد نمونه‌های دستکاری شده، طی دوره دستکاری شاهد

---

[1] Felixson & Pelli
[2] Hillion & Suominen
[3] Carhart
[4] Bernhardt & Davies
[5] Frieder & Zittrain
[6] Hanke & Hauser

افزایش قیمت، پس از دوره دستکاری شاهد کاهش قیمت و طی دوره قبل و بعد از دستکاری افزایش نوسان قیمتی در سهم اتفاق افتاده است. پس از آن کامرتون و پوتنینس (۲۰۱۱)، نمونه‌ای شامل ۱۸۴ شرکت که طی سال‌های ۱۹۹۷ تا ۲۰۰۹ که SEC بابت دستکاری آخرین قیمت مورد رسیدگی قرار گرفته بود را بررسی کردند. اکثر دستکاری‌های اتفاق‌افتاده به وسیله مدیران سرمایه‌گذاری صندوق‌های سرمایه‌گذاری، مدیران رده بالای شرکت، کارگزاران و سهامداران عمده انجام شده است. آن‌ها بدین نتیجه رسیدند که دستکاری آخرین معامله معمولاً با افزایش قیمت در آخر روز معاملاتی، سپس کاهش قیمت پس از دستکاری، افزایش حجم معاملات، افزایش دامنه بین بهترین مظنه خرید و فروش همراه می‌باشد.

با گسترش بازارهای مالی و افزایش حجم معاملات در بورس، همچنین توسعه روزافزون روش‌های کامپیوتری و مبتنی بر یادگیری ماشین، برخی پژوهشگران به دنبال بهره‌گیری از تکنولوژی‌های جدید در شناسایی دستکاری شدند. گل‌محمدی و زایان[1] (۲۰۱۲) در پژوهشی تحت عنوان کاربرد هوش مصنوعی در شناسایی دستکاری در بازار سرمایه به بررسی مروری پژوهش‌های پیشین در این زمینه پرداختند. آن‌ها مطالعات شناسایی دستکاری به وسیله هوش مصنوعی را به پنج دسته کلی شامل شناسایی الگوی دستکاری، شناسایی داده‌های پرت، استخراج قوانین ضد دستکاری، تحلیل شبکه‌های اجتماعی در زمینه دستکاری و شناسایی دستکاری از طریق نمودار و تصویرسازی تقسیم کرده و داده‌های مورد نیاز برای هر نوع پژوهش را ذکر کردند. دیاز و همکاران[2] (۲۰۱۱) در پژوهشی به بررسی چالش‌های به‌کارگیری روش‌های هوش مصنوعی برای شناسایی دستکاری بازار پرداخته و با استفاده از اطلاعات بازار سهام، دستکاری معامله‌محور را بررسی نمودند. متغیرهای استفاده شده شامل اختلاف بازدهی سهام دستکاری شده با سبد مرجع، بازدهی غیرنرمال سهام، نقدشوندگی و نوسان سهم می‌باشد. درنهایت پژوهشگران با استفاده از درخت تصمیم به بررسی سهام دستکاری شده پرداخته و الگوریتمی برای شناسایی دستکاری به نهاد ناظر پیشنهاد دادند.

فلاح‌شمس و کردلوئی (۱۳۹۰) مدلی برای تخمین دستکاری قیمت در بورس اوراق بهادار تهران ارائه نمودند. برای این‌کار ابتدا به روش غربال‌گری ۳۹۷ شرکت پذیرفته‌شده در بورس تهران انتخاب و اطلاعات مربوط به قیمت و حجم معاملات آن‌ها را طی سال‌های ۱۳۸۰ تا ۱۳۸۸ گردآوری نمودند. سپس با استفاده از آزمون‌های تسلسل، کشیدگی و آزمون وابستگی دیرش شرکت‌های منتخب را به دو دسته دستکاری شده و دستکاری نشده تقسیم کردند. سپس با استفاده از آزمون‌های لاجیت و شبکه عصبی مصنوعی و با در نظر گرفتن اندازه شرکت، شفافیت اطلاعات، نسبت P/E و نقدشوندگی سهام، مدلی برای پیش‌بینی دستکاری قیمت سهام شرکت‌ها در بورس

[1] Golmohammadi & Zaiane
[2] Diaz

اوراق بهادار طراحی کردند. نتایج نشان داد که قدرت پیش‌بینی مدل لاجیت و شبکه‌های عصبی مصنوعی به ترتیب ۹۲,۱درصد و ۹۴,۱درصد بوده است. پوست‌فروش و همکاران (۱۳۹۳)، دستکاری قیمت‌ها با استفاده از مدل‌های ترکیبی تحلیل ممیزی درجه دوم و الگوریتم ژنتیک بر مبنای شبکه عصبی مصنوعی را بررسی کردند. آن‌ها با روش غربال‌گری، ۳۴۵ شرکت پذیرفته‌شده در بورس تهران را در بازده زمانی ۱۳۸۷ الی ۱۳۹۱ انتخاب نموده و اطلاعات مربوط به شاخص‌های قیمت و بازده نقدی، قیمت پایانی، نوسان قیمت پایانی و حجم معاملات را گردآوری کردند. سپس با بررسی نمودار روند تغییرات شاخص قیمت و بازده نقدی و حجم معاملات، تاریخ شروع دستکاری را تعیین و با اطلاعات مربوط به یک‌سال گذشته از زمان دستکاری سهم، مدل مذکور را توسعه دادند. نتایج نشان داد مدل تحلیل ممیزی درجه دوم نسبت به مدل الگوریتم ژنتیک بر مبنای شبکه عصبی مصنوعی، توانایی بهتری در دسته‌بندی سهام دستکاری شده و نشده دارد. شمس و عطایی (۱۳۹۳)، شناسایی دستکاری قیمت سهام از طریق مدل ترکیبی الگوریتم ژنتیک–شبکه عصبی مصنوعی (ANN-GA) و مدل تابع تفکیکی درجه دو تعدیل شده SQDF را با استفاده از داده‌های مربوط به ۳۱۶ شرکت، مورد بررسی قرار دادند. نتایج نشان داد که مدل ترکیبی ANN-GA عملکرد بهتری نسبت به مدل SQDF دارد.

گل‌محمدی و دیاز (۲۰۱۵) با استفاده از داده‌های جمع‌آوری‌شده در مقاله دیاز و همکاران (۲۰۱۱) به بررسی انواع روش‌های نظارت شده هوش مصنوعی برای شناسایی دستکاری قیمت پرداخت و مدل‌ها را با یکدیگر مقایسه کردند. برخی از ابزارهای مورد استفاده شامل درخت تصمیم، جنگل تصادفی، هوش مصنوعی، بیز ساده، KNN و ماشین‌بردار پشتیبان بوده است. مدل بیز ساده بهترین عملکرد را از نظر معیار عملکردی F_2 برابر با ۵۳ درصد بهترین عملکرد را در شناسایی دستکاری قیمتی داشت. گل‌محمدی و زایان (۲۰۱۵) در پژوهشی با استفاده از روش سری زمانی متنی[1] (CAD) به شناسایی داده‌های ناهنجار پرداخته است. آنان ادعا کرده‌اند که جمع‌آوری داده‌های واقعی در زمینه دستکاری سهام، سخت، زمان‌بر و غیرقابل اعتماد است (بسیاری توسط نهاد ناظر و دادگاه شناسایی نشده، تاریخ دقیق زمان دستکاری مشخص نیست و ...). بنابراین از سهامی که احتمال دستکاری در آن‌ها وجود ندارد استفاده کرده و با تزریق داده تصادفی به عنوان ناهنجاری، پایگاه داده جدیدی ساخته و مدل پیشنهادی را روی آن بررسی کردند. نتایج نشان می‌دهد که مدل CAD نسبت به مدل‌های KNN[2] و RW[3] عملکرد بهتری دارد. در پژوهشی مشابه، الثانی (۲۰۱۷) با بررسی بورس قطر و تغییر اندکی در متغیرهای ورودی مدل CAD، نتایج به مراتب بهتر نسبت به پژوهش قبلی به دست آورد.

[1] Contextual Anomaly Detection
[2] K-Nearest Neighbors
[3] Random Walk

کلوس و کاشف[1] (۲۰۲۰)، در پژوهشی با عنوان ترکیب مدل سیستم ایمنی مصنوعی و تحلیل خوشه‌بندی به کشف ناهنجاری در بازار سرمایه اقدام کردند. داده‌های استفاده‌شده مربوط به شرکت‌های بزرگی مانند گوگل، آمازون و ... بوده که با توجه به اندازه شرکت و حجم معاملات احتمال دستکاری در آن‌ها بسیار کم است. به همین دلیل در اقدامی مشابه با پژوهش گل‌محمدی به ایجاد داده تصادفی پرداختند. پانکاجاشان و همکارن[2] (۲۰۲۲)، با استفاده از رویکرد ترکیبی خودرمزگذار با LSTM در تشخیص ناهنجاری مجموعه داده‌های سرورOpenStack فضای ابری، پرداختند. مقایسه مدل‌ها براساس شاخص دقت، یادآوری و F_1 صورت گرفت و براساس شاخص F_1 رتبه بندی شد. DA-LSTM با ۹۹,۱درصد و SVM با ۸۱درصد و درخت تصمیم ۷۹درصد اندازه‌گیری شد و در نتیجه مدل ترکیبی DA-LSTM بهترین عملکرد را از خود نشان داد.

کوین و همکاران[3] (۲۰۲۳)، در پژوهشی تحت عنوان شناسایی دستکاری بازار سهام با مدل‌های ترکیبی در بازار آمریکا و انگلستان، در کل نه مدل ترکیبی بر پایه روش LSTM ARIMA معرفی کردند و با معیار ارزیابی F_1 به مقایسه مدل‌ها پرداختند. نتایج نشان داد روش‌های مبتنی بر LSTM در مجموع F_1 بالاتری داشتند.

۳. روش‌شناسی پژوهش

برای اینکه بتوان از روش‌های داده‌کاوی و یادگیری عمیق استفاده کرد، ابتدا لازم است پایگاه داده‌ای از سهام دستکاری‌شده و دستکاری‌نشده ایجاد نمود. در بورس سایر کشورها سهام شرکت‌های دستکاری شده و تاریخ تقریبی دستکاری آن، توسط نهاد ناظر اعلام می‌شود. متأسفانه در کشور ما، سازمان بورس و اوراق بهادار این کار را انجام نمی‌دهد. بنابراین در این پژوهش لازم است ابتدا پایگاه داده شامل سهام دستکاری شده و نشده و تاریخ دقیق سهام دستکاری‌شده تشکیل دهیم. به این منظور از روش زیر استفاده می‌نماییم.

۱) با استفاده از روش‌های آماری، آزمون تسلسل، کشیدگی، چولگی و همچنین چهار آزمون شناسایی ناهنجاری متنی روی حجم و بازده (۱۵ و ۳۰ روزه) سهامی که احتمال دستکاری در آن‌ها وجود دارد را شناسایی می‌کنیم. اگر در یک تاریخ، حداقل ۵ آزمون از ۷ آزمون انجام‌شده روی سهام مشخصی، دلالت بر دستکاری سهم داشته باشد، برچسب یک و در غیر این‌صورت برچسب صفر به آن سهم در تاریخ مذکور می‌دهیم. سپس با استفاده از روش‌های نموداری و بر روی متغیرهایی مانند بازدهی غیرنرمال، افزایش ناگهانی حجم معاملات، افزایش نوسان‌پذیری سهم،

[1] Close & Kashef
[2] Pankajashan et al
[3] Quinn et al

مقایسه بازدهی سهم مربوطه با بازدهی صنعت، همچنین استفاده از اطلاعات سهام برای بررسی رخداد یا عدم اتفاق بنیادی خاص، تاریخ دقیق دستکاری سهم را مشخص می‌نماییم.

۲) با استفاده از پرسش‌نامه و نظر خبرگان، سهامی که از نظر نقدشوندگی و اندازه بازار به اندازه‌ای کارا هستند که امکان دستکاری قیمتی در این سهام وجود ندارد، انتخاب شده و به صورت تعمدی در برخی سهام و در روزهای معاملاتی، با استفاده از تابع توزیع داده‌ها در دوره زمانی مشخص (مثلاً دوره سه ماهه، شش ماهه و یک ساله) و همچنین میانگین و انحراف معیار داده‌ها در دوره مذکور، داده‌های پرت به صورت تصادفی به مدل تزریق می‌شود تا شبیه‌ساز دستکاری سهام باشند. روزهای معاملاتی تزریق‌شده، به عنوان روزهای معاملاتی دستکاری‌شده در نظر گرفته می‌شود.

در پژوهش حاضر با استفاده از ترکیب مدل حافظه کوتاه‌مدت- طولانی و روش خودرمزگذار متغیر، برای ۷۳ سهم و صندوق سرمایه‌گذاری (فولاد، ثنوسا، کاوه و...) از ۱۹ صنعت (فلزات اساسی، شیمیایی و...) پذیرش‌شده در بورس اوراق بهادار ایران طی بازه زمانی ۱۳۹۸ تا ۱۴۰۲ مدلی برای شناسایی دستکاری طراحی شده‌است. بدین منظور ابتدا سهامی که با توجه به نقدشوندگی و کارایی و نظر خبرگان، اطمینان نسبی از عدم دستکاری آن‌ها وجود دارد انتخاب و آموزش می‌دهیم. سپس داده‌های تست را به شبکه تزریق و مقدار عددی احتمال دستکاری شدن سهم در روزهای معاملاتی مختلف را محاسبه می‌کنیم.

آزمون تسلسل

در این آزمون چنانچه سلسله‌های بازدهی منفی و مثبت بیشتر از سلسله‌های مورد انتظار باشد، بیانگر وجود الگوی غیر تصادفی در روند قیمت سهم و در نتیجه احتمال بروز دستکاری قیمت خواهد بود. در تشکیل سلسله‌ها به بازده‌های روزانه کمتر از میانگین علامت منفی و بازده‌های بالاتر از میانگین علامت مثبت می‌دهیم. هر سلسله شامل توالی یک یا چند علامت مثبت یا منفی خواهد بود. به عبارت دیگر، زمانی که علامت تغییر می‌کند یک سلسله جدید شروع شده است. تعداد کل مثبت و منفی‌ها نیز در سری زمانی مورد بررسی، شمارش می‌شود. بعد از این مرحله، تعداد سلسله مورد انتظار و انحراف معیار آن‌ها از طریق فرمول‌های ذیل (فلاح‌شمس و کردلوئی، ۱۳۹۰) محاسبه می‌شود.

$$E(R) = \frac{2n_1n_2}{n_1 + n_2}$$ رابطه (۱)

$$STD = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$ رابطه (۲)

n_1 تعداد بازدهی‌های مثبت و n_2 تعداد بازدهی‌های منفی است. سپس معنادار بودن تفاوت تعداد سلسله‌های شمارش‌شده با تعداد سلسله‌های مورد انتظار برای متغیر تصادفی از طریق آزمون t مورد بررسی قرار می‌گیرد. اگر آماره آزمون (تفاوت بین تعداد سلسله‌های شمارش‌شده و تعداد سلسله‌های مورد انتظار تقسیم بر انحراف معیار سلسله‌ها) در محدوده بحرانی قرار داشته باشد، در این صورت تعداد سلسله‌ها با تعداد سلسله‌های مورد انتظار تفاوت معناداری نداشته و در نتیجه احتمالاً دستکاری اتفاق نیفتاده است. اما در صورتی که آماره آزمون در محدوده بحرانی قرار نداشته باشد، تعداد سلسله‌های شمارش شده با تعداد سلسله‌های مورد انتظار تفاوت معنی‌داری داشته و بدین معنی است که طول سلسله‌های سری زمانی، آنقدر طولانی است که با داده‌های تصادفی و مستقل هم‌خوانی ندارد و احتمال وقوع دستکاری وجود دارد.

جدول ۱. نمونه نتایج آزمون تسلسل سهام ثنوسا، ۳ ماهه اول ۱۳۹۹

دوره مورد بررسی	سه ماهه با محاسبات بازدهی روزانه
میانگین بازدهی روزانه	۰٪/۷۹
موارد بالای میانگین	۳۲
موارد زیر میانگین	۲۴
تعداد سلسله‌ها	۱۱
تعداد سلسله مورد انتظار	۲۸/۴
انحراف معیار سلسله مورد انتظار	۳/۶
Zآماره	-۴/۸
P-value	تقریباً صفر

با توجه به p-value ذکر شده در جدول (۱)، سهم مورد نظر با سطح اطمینان ۹۹درصد در دوره زمان مذکور دستکاری شده است.

آزمون چولگی و کشیدگی

روش دیگر تشخیص احتمال بروز دستکاری قیمت سهام، بررسی ضرایب کشیدگی و چولگی بازدهی می‌باشد. اگر بازدهی روزانه به طور معناداری از ضریب کشیدگی و چولگی تابع توزیع نرمال انحراف داشته باشد، امکان بروز دستکاری قیمت وجود خواهد داشت. میانگین و واریانس چولگی و کشیدگی از رابطه (۳) و رابطه (۴) (فلاح‌زاده و همکاران، ۱۳۹۴) محاسبه می‌شود.

$$S=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}\ ,\ \textbf{کشیدگی}=\frac{\sum(x-\bar{x})^4}{(n-1).S^4}\ ,\ \textbf{چولگی}=\frac{\sum(x-\bar{x})^3}{(n-1).S^3}$$ رابطه (۳)

$$\textbf{کشیدگی}=\sqrt{\frac{6n}{(n-2)(n-1)(n+3)(n+5)}}\ ,\ \textbf{چولگی}=\sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$ رابطه (۴)

جدول ۲. چولگی (هرمز، ۳ ماهه سوم ۱۳۹۸)

دوره مورد بررسی	سه ماهه با محاسبات بازدهی روزانه
چولگی	۰/۱۴۹
انحراف معیار چولگی	۰/۳۲۱
چولگیZآماره	۰/۴۶۳
چولگی P-value	۳۲٪/۱۴
کشیدگی (کاوه، ۳ ماهه چهارم ۱۴۰۰)	
کشیدگی	۸/۳۳۴
انحراف معیار کشیدگی	۰/۶۳۳
کشیدگیZآماره	۱۳/۱۵
کشیدگی P-value	نزدیک صفر

همانطورکه نتایج جدول نشان می‌دهد، سهم هرمز در سه ماهه سوم سال ۱۳۹۸ با سطح اطمینان ۹۵درصد دارای چولگی نمی‌باشد و سهم کاوه در سه ماهه چهارم سال ۱۴۰۰ با اطمینان ۹۹درصد دارای کشیدگی می‌باشد.

آزمون شناسایی ناهنجاری متنی

عمده روش‌های شناسایی ناهنجاری در سری‌های زمانی، مقایسه فاصله هر یک از داده‌ها با داده‌های نرمال و اطلاق ناهنجاری به داده‌هایی که فاصله بیشتر از حد معمول دارند، می‌باشد. در آزمون شناسایی ناهنجاری متنی[1] (CAD)، سری زمانی چندین سهم مشابه (به عنوان مثال سهام هم‌صنعت) به عنوان ورودی به مدل در نظر گرفته می‌شود. بنابراین اگر یکی از خواص سهم انتخابی را براساس سایر سهام مشابه، تخمین بزنیم، نباید اختلاف معنی‌داری بین خواص تخمین و خواص واقعی وجود داشته باشد. چنانچه اختلاف از سطح مشخصی بیشتر بود، می‌توان ادعا کرد که در سهم مذکور ناهنجاری وجود دارد. به صورت ریاضی اگر چندین سری زمانی از یک صنعت مشابه داشته باشیم، با انتخاب پنجره زمانی، می‌توان به تخمین سری زمان iام بر اساس سایر سری‌های زمانی مطابق با فرمول زیر (گل‌محمدی و زایان، ۲۰۱۵) اقدام نمود.

$$\widehat{X_{it}} = \psi(\phi\ (X_t), c_t) + \ \varepsilon$$ رابطه (۵)

$\phi\ (X_t)$ تابعی از متغیر ورودی (مقدار Xi در زمان t-1، و یا میانگین مقادیر گذشته سری زمانی و غیره)، c_t میانگین ویژگی مشخصی از سری‌های زمانی در هر لحظه می‌باشد. ψ ، تابعی است که ارتباط بین سری زمانی مورد نظر برای تخمین و c_t را مشخص می‌کند. برای توضیح بهتر و ساده‌تر مدل، فرض کنید سری زمانی بازده روزانه ۵ سهم متعلق به صنعت شیمایی را در اختیار داریم. در هر روز میانگین بازدهی ۵ سهم را محاسبه می‌کنیم (C). حال برای هر یک از سری‌های

[1] Contextual Anomaly Detection

زمانی، هم‌بستگی بین C و سری زمانی مذکور را محاسبه نموده و از رابطه (۶) که ساده شده رابطه (۵) می‌باشد سری زمانی مورد نظر را برآورد می‌کنیم.

$$\widehat{X_{it}} = X_{it-1} * Corr(X_i, C) + \varepsilon$$ رابطه (۶)

سپس فاصله اقلیدوسی بین مقدار برآوردی و واقعی ($\left|\widehat{X_{it}} - X_{it}\right|$) را محاسبه کرده و در صورتی که فاصله مذکور از انحراف معیار مقدار واقعی سهام در پنجره زمانی بیشتر باشد، داده مذکور را به عنوان ناهنجاری در نظر می‌گیریم.

در پژوهش حاضر برای تعیین نقاط ناهنجاری و احتمالاً دستکاری شده، از روش الثانی (۲۰۱۷) بر روی دو متغیر بازده و حجم با پنجره زمانی ۱۵ و ۳۰ روزه، استفاده کرده‌ایم. نتایج شناسایی تاریخ دستکاری به وسیله روش CAD در جدول (۳) نمایش داده شده است.

جدول ۳. محاسبات آزمون CAD بر روی حجم سهم زدشت در تاریخ ۱۳۹۸/۰۹/۰۲

پنجره زمانی	۱۵ روز معاملاتی
هم‌بستگی بین حجم صنعت و سهم	۴۸/۴%
میانگین حجم معاملات سهم	۲،۹۹۲،۹۲۵
تخمین حجم معاملات انتظاری سهم	۱،۴۴۹،۵۳۸
حجم واقعی معاملات سهم (تعدیل شده بر اساس قیمت)	۶،۲۲۳،۳۰۹
فاصله اقلیدوسی	۴،۷۷۳،۷۷۲
انحراف معیار حجم سهم	۱،۶۷۴،۱۸۰

در جدول فوق، با توجه به اینکه فاصله اقلیدوسی از انحراف معیار حجم سهم بیشتر است، در این تاریخ، سهم مورد نظر از نقطه نظر حجم معاملات، ناهنجاری داشته و باید طبق رویه تشخیص دستکاری، بررسی شود. در نهایت مطابق با رویه شناسایی دستکاری، سهام و تاریخ دستکاری سهم را شناسایی و به آن‌ها برچسب یک و به سایر داده‌ها برچسب صفر می‌دهیم.

حافظه کوتاه مدت طولانی

شبکه عصبی بازگشتی LSTM نوع توسعه یافته شبکه عصبی ساده است که در تجزیه و تحلیل داده‌های سری زمانی عملکرد بهتری را نشان داده است. در این شبکه به جای یک لایه شبکه عصبی، چهار لایه وجود دارد و این لایه‌ها با روشی کاملاً خاص با یکدیگر در تعامل هستند. شبکه عصبی بازگشتی LSTM دارای ۴ گیت: فراموشی، یادآوری، یادگیری و خروجی. همچنین دارای سه ورودی حافظه بلندمدت، حافظه کوتاه مدت و نمونه آموزشی یا داده جدید (E) است. ساختار LSTM به صورتی است که لایه‌ی گیت فراموشی، تصمیم‌گیری را راجع بـه کنترل حافظه و اطلاعات انجام می‌دهد. رابطه ۷ ، h_{t-1} و x_t برای هر عدد در وضعیت سلول C_{t-1} عددی بین

صفر و یک را نشان می‌دهد. عدد یک نشان دهنده «به طور کامل نگه دار» و صفر به معنای «نگه ندار» است. (هوخرایتر و اشمیدهوبر، ۱۹۹۷)

$$f_t = \sigma\left(W_{f0}[h_t, X_t]\right) + b_f$$ رابطه (۷)

در گام بعدی اطلاعات جدید در وضعیت سلول ذخیره می‌شود. در بخش اول یک لایه سیگموئید به نام لایه‌ی گیت «ورودی» برای تعیین به روزرسانی کدام مقادیر رابطه (۸) و در بخش دوم یک لایه، tanh بُرداری از مقادیر کاندیدشده‌ی جدید ساخته می‌شود، رابطه (۹) در نهایت با ترکیب این دو بخش می‌توان در سلول، اطلاعاتی افزود یا به‌روزرسانی کرد.

$$i_t = \sigma(W_i[h_{t-1}, X_t]) + b_i$$ رابطه (۸)

$$\tilde{C}_t = tanh\,(W_{c0}[h_{t-1}, X_t]) + b_c$$ رابطه (۹)

براساس رابطه (۱۰)، می‌بایست وضعیت سلول به وضعیت جدید C_t به روزرسانی شود. در این بخش اطلاعاتی که شبکه تصمیم گرفته بود، حذف می‌شود. بنابراین C_{t-1} (حالت قدیم) را در f_t ضرب می‌کنیم. سپس $i_t * \tilde{c_t}$ را به فرمول اضافه می‌کنیم. با این کار مقادیر انتخابی برای به روزرسانی وضعیت، مشخص می‌شوند.

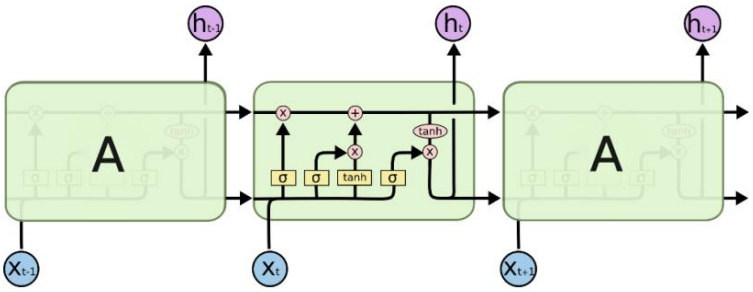
$$C_t = f_t * C_{t-1} + i_t$$ رابطه (۱۰)

در پایان تعیین می‌کنیم که از چه نتیجه‌ای خروجی خواهیم گرفت. این خروجی براساس وضعیت سلول ساخته می‌شود. این قسمت نیز شامل دو بخش است. در بخش اول لایه سیگموئید اجرا می‌گردد که تعیین می‌کند چه قسمتی از وضعیت سلول به خروجی فرستاده شود (رابطه (۱۱)). سپس لایه tanh (خروجی بین ۱ و −۱) اجرا می‌شود (رابطه (۱۲)). نتیجه در خروجی لایه سیگوئید ضرب شده و در نهایت خروجی مورد نظر به‌دست می‌آید.

$$O_t = \sigma(W_0[h_{t-1}, X_t]) + b_0$$ رابطه (۱۱)

$$h_t = O_t * tanh(C_t)$$ رابطه (۱۲)

در شکل زیر نمونه عملکرد مدل حافظه کوتاه مدت طولانی نشان داده شده است.



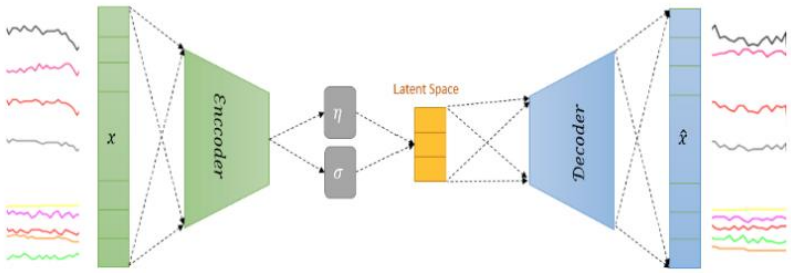
شکل ۱. ماژول تکرار شونده در مدل LSTM

در این پژوهش زبان پایتون و کتابخانه‌های آن از جمله Pandas, Numpy, Matplotlib, Scikit Learn و کتابخانه Keras که بر روی کتابخانه TensorFlow استفاده شده است.

خودرمزگذار متغیر

خود رمزگذارها از دو بخش رمزگذار و رمزگشا تشکیل شده‌اند و با نگاشت ورودی $x \in R^{d_x}$ به فضای نهان یعنی $z \in R^{d_z}$ سپس رمزگشایی فضای نهان به ابعاد ورودی، ورودی را بازسازی می‌کنند و ورودی بازسازی شده $\hat{x}$ را به‌دست می‌آورند. در مرحله آموزش با کمینه کردن تابع ضرر به دنبال بهبود نمونه بازسازی شده و یافتن پارامترهای حداقل‌کننده فاصله $\hat{x}$ با x می باشیم. به طور معمول فضای نهان z ابعاد کمتری نسبت به فضای ورودی x دارد و از این رو، خودرمزگذار مجبور است نمایش فشرده داده‌های ورودی را آموزش ببیند.

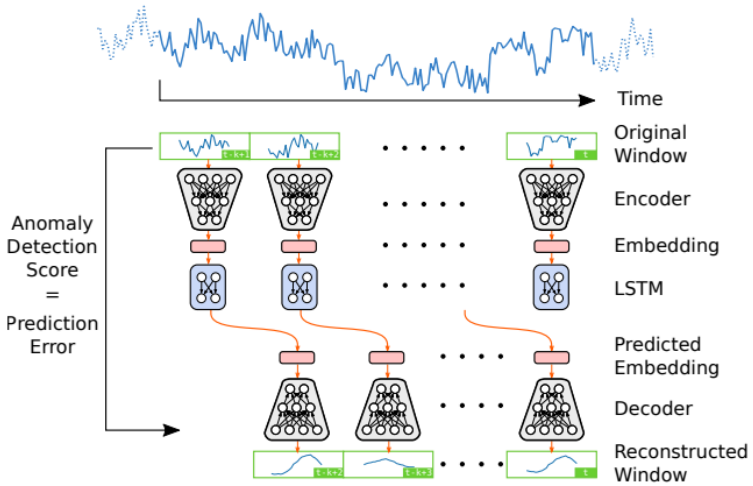
خود رمزگذار متغیر شکل (۲) یک مدل احتمالی است و در واقع VAE یک مدل مولد عمیق است که فضای نهان خود رمزگذار را به یک متغیر تصادفی توزیع شده بر اساس توزیع پیشین $p_\Theta(Z)$، محدود می‌کند که معمولاً یک توزیع نرمال استاندارد $Normal(0; I)$ است. ارائه یک معیار احتمالی به جای خطای بازسازی، مزیت اصلی یک مدل تشخیص ناهنجاری مبتنی بر VAE است؛ زیرا احتمالات بازسازی اصولی و عینی‌تر هستند.



شکل ۲. مدل خودرمزگذار متغییر در بازسازی سری زمانی

مدل ترکیبی VAE_LSTM

شکل (۳) خلاصه‌ای از الگوریتم تشخیص مدل هیبرید ما را نشان می‌دهد، که شامل یک ماژول VAE برای استخراج ویژگی‌های محلی یک پنجره کوتاه و یک ماژول LSTM برای تخمین روندهای بلند مدت است.



شکل ۳. مدل VAE-LSTM در توالی K پنجره زمانی

ابتدا از مجموعه کل داده‌ها، داده‌های مربوط به تست و آموزش را جدا سپس بخشی از سری زمانی قیمت را در نظر می‌گیریم که شامل هیچ ناهنجاری به عنوان داده‌های آموزشی نباشد و بقیه سری‌های زمانی با ناهنجاری به عنوان داده‌های آزمون برای ارزیابی نگه‌داشته می‌شوند.

در مدل VAE، پنجره‌ای از خوانش‌های متوالی p به عنوان ورودی، یک نگاشت کم بعدی از ابعاد q را از طریق رمزگذار برآورد می‌کند و پنجره اصلی را از طریق رمزگشا بازسازی می‌نماید. برای آموزش، از پنجره غلتان برای تولید داده‌ها استفاده می‌کنیم. در مدل ترکیبی پیشنهادی از توالی k پنجره غیر هم پوشان از فرمول $W_t = [w_{t-(k-1)\times p, wt-(k-2)\times p}, \cdots, w_t]$ استفاده می‌شود، برای نشان‌دادن توالی پنجره‌های منتهی به زمان t و $E_t = [e_t^1, \ldots, e_t^k]$ برای جایگذاری متناظر در W_t با e_t^i از نشان دادن جایگذاری پنجره i ام در W_t استفاده می‌کنیم. برای داده‌های آموزشی خوانش‌های N_{train}، از پنجره‌های غلتان $N_{\text{train}} - p$ را برای آموزش مدل VAE و دنباله غلتان $N_{\text{train}} - pk$ برای آموزش مدل LSTM استفاده می‌کنیم. ۱۰ درصد از توالی‌های تولید شده را به صورت تصادفی از داده‌های آموزش به عنوان مجموعه اعتبارسنجی ذخیره، سپس پنجره‌ها و توالی‌ها در مجموعه اعتبارسنجی از آموزش حذف می‌شوند. با پنجره‌های باقی مانده در مجموعه آموزش، پارامترهای مدل VAE را بهینه‌سازی می‌کنیم. پس از بهینه‌سازی

مدل VAE، از کدگذار مدل VAE آموزش‌دیده برای برآورد تمام توالی جای‌گذاری E_t در مجموعه آموزش استفاده می‌کنیم. پس از آموزش، مدل قابلیت استفاده برای داده‌ها را پیدا می‌کند. در ابتدا از رمزگذار VAE برای برآورد E_tدر W_t و سپس اولین جایگذاری $k-1$ را به مدل LSTM برای پیش‌بینی جایگذاری $k-1$ بعدی $[\hat{e}_t^2, \dots, \hat{e}_t^k]$ استفاده می‌کند.

در تست مدل اگر مقدار احتمال بالاتر از ۵۰درصد بود آن را دستکاری شده و پایین‌تر را، دستکاری‌نشده تلقی می‌کنیم. شایان ذکر است داده‌های برچسب‌دار فقط برای تست مدل و سنجش کارایی استفاده می‌شود و برای آموزش و استفاده از مدل در واقعیت نیازی به آن‌ها نیست.

ابزار اندازه‌گیری عملکرد

به طور کلی، در بررسی خروجی هر مدل مبتنی بر یادگیری ماشین یکی از چهار حالت: ۱) به درستی مثبت[1] (TP) واقعاً دستکاری شده و مدل درست تشخیص دهد. ۲) به اشتباه مثبت[2] (FP) واقعاً دستکاری نشده و مدل دستکاری شده تشخیص دهد. ۳) به درستی منفی[3] (TN) واقعاً دستکاری نشده و مدل دستکاری شده تشخیص ندهد. ۴) به اشتباه منفی[4] (FN) واقعاً دستکای‌شده و مدل آن را دستکاری شده تشخیص ندهد، رخ خواهد داد (شکل (۴)). به طور منطقی شبکه‌ای که TP و TN را افزایش داده و FP و FN را کاهش دهد، عملکرد بهتری خواهدداشت (استمن، ۱۹۹۷).

برای مقایسه بین مدل‌ها و داده‌های مختلف، معیارهای ارزیابی عملکرد بر پایه چهار متغیر ذکرشده طراحی شده است. معیار ارزیابی دقت[5] که از طریق فرمول $P = \frac{TP}{TP+FP}$ و بازخوانی[6] از طریق فرمول $R = \frac{TP}{TP+FN}$ به دست می‌آید. این دو متغیر به صورت جداگانه و بر حسب درصد محاسبه شده و هر چه مقدار این دو متغیر بیشتر باشد، نشان‌دهنده عملکرد بهتر مدل می‌باشد. برای ترکیب این دو شاخص عملکردی و تبدیل آن به یک عدد واحد، از میانگین هارمونیک شاخص دقت و بازخوانی مطابق با رابطه (۱۳) استفاده می‌شود (استمن، ۱۹۹۷).

$$F_\beta = \frac{(1+\beta)^2 * \mathrm{P} * \mathrm{R}}{(\beta^2 * \mathrm{P}) + R} = \frac{(1+\beta)^2 * \mathrm{TP}}{(1+\beta)^2 * \mathrm{TP} + (\beta^2 * \mathrm{FN}) + \mathrm{FP}} \qquad \text{رابطه (۱۳)}$$

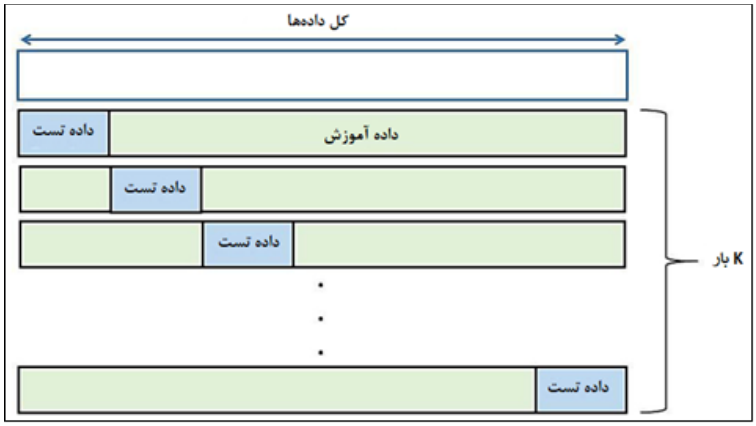
در رابطه (۱۳)، β ضریب اهمیت معیار ارزیابی عملکرد بازخوانی در مقابل دقت می‌باشد. اگر دقت و بازخوانی اهمیت یکسانی داشته باشند، β برابر با یک و شاخص ارزیابی F_1 ایجاد می‌گردد.

[1] True Positive
[2] False Positive
[3] True Negative
[4] False Negative
[5] Precision
[6] Recall

در بازار سرمایه اشتباه در دسته‌بندی سهام دستکاری شده و دستکاری نشده، هزینه متقارن ندارند. بنابراین در این پژوهش از شاخص‌های ارزیابی دقت، بازخوانی، F_1 و F_2 (در حالتی که بتا معادل با ۲ باشد) استفاده شده و برای رتبه‌بندی مدل‌ها از شاخص ارزیابی عملکردی F_2 استفاده شده است.

صحت‌سنجی و پایایی مدل‌های یادگیری ماشین

در مدل‌های یادگیری ماشین روش‌های مختلفی برای صحت‌سنجی وجود دارد که ساده‌ترین آن جداسازی داده‌های آموزش و تست می‌باشد. در این پژوهش علاوه بر جداسازی داده‌های آموزش و تست، از روش اعتبارسنجی متقابل ۱۰ برابر استفاده کرده‌ایم. در این روش، داده‌ها را به ۱۰ قسمت مساوی تقسیم و در هر مرحله از ۹ قسمت برای آموزش و از یک قسمت برای تست استفاده می‌کنیم. اجرای مدل را ۱۰ مرتبه تکرار کرده و میانگین محاسبات عملکردی مدل را به عنوان خروجی گزارش نموده‌ایم.



شکل ۴. چگونگی انجام اعتبارسنجی k برابر

۴. تحلیل داده‌ها و یافته‌ها

باتوجه به مبانی نظری بررسی‌شده طی دوره دستکاری اغلب، بازدهی، حجم معاملات و نوسان‌پذیری افزایش می‌یابد. برای آزمون دقیق‌تر این موضوع، معمولاً از مدل‌های رگرسیونی و رگرسیون لجستیک استفاده می‌شود. همانطور که از رابطه (۱۴) مشخص است، متغیرهای مورد بررسی (Indicator) مانند بازدهی سهام، میانگین حجم معاملات و غیره، طی دوره دستکاری به وسیله متغیر مجازی (Manipulated) آزمون می‌شود:

$$Indicator = \alpha_0 + \alpha_1 * Manipulated + \varepsilon$$ رابطه (۱۴)

در رابطه فوق، متغیر مجازی در زمان وقوع دستکاری، مقدار یک و در زمان‌های دیگر مقدار صفر می‌گیرد. در صورتی که ضریب α_1 معنی‌دار باشد بدین معنی است که در دوره دستکاری

مقدار متغیر مورد بررسی به میزان قابل توجهی نسبت به سایر دوره‌ها متفاوت است. به طور معکوس این مورد تست می‌شود که آیا با استفاده از متغیرهای مرتبط با سهام (بازده، حجم معاملات، نوسان‌پذیری و غیره)، می‌توان بین دستکاری و عدم دستکاری در سهام تمایز قایل شد. برای این منظور مطابق با روابط (۱۵) و (۱۶) از رگرسیون لجستیک استفاده می‌شود.

رابطه (۱۵) $$Manipulated = \beta_0 + \beta_1 * Indicator_1 + \cdots + \beta_n * Indicator_n + \varepsilon$$

رابطه (۱۶) $$f(z) = \frac{1}{1 + e^{Manipulated}}$$

متغیر مجازی در صورت دستکاری یک و در غیر اینصورت مقدار صفر می‌گیرد. مطالعات نشان داده‌اند که در صورت دستکاری، متغیرهای تأثیرگذار بر سهام ثابت نبوده و افزایش پیدا می‌کنند.

نتایج رگرسیون به دست آمده از روابط فوق در جدول (۴) به شرح زیر ارایه شده است.

جدول ۴. رگرسیون خطی جهت تطبیق داده‌ها

رگرسیون خطی (رابطه ۵) $Indicator = \alpha_0 + \alpha_1 * Dummy + \varepsilon$	بازده	حجم معاملات	میانگین خرید حقوقی	میانگین فروش حقوقی	میانگین خرید حقیقی	میانگین فروش حقیقی
α_0	۰,۰۰۲۴	۰,۴۴۲	۰,۴۱۶	۰,۴۱۲	۰,۴۵۱	۰,۴۵
انحراف معیار	(۰,۰۰۰۱)	(۰,۰۰۰۸)	(۰,۰۰۰۸)	(۰,۰۰۰۸)	(۰,۰۰۰۸)	(۰,۰۰۰۸)
P-value	۰,۰۰۰۰	۰,۰۰۰۰	۰,۰۰۰۰	۰,۰۰۰۰	۰,۰۰۰۰	۰,۰۰۰۰
α_1	۰,۰۲۷۴	۰,۳۴۹	۰,۰۱۶۸**	۰,۰۱۲۳*	۰,۲۲۱	−۰,۱۲۴۴
انحراف معیار	(۰,۰۰۰۹)	(۰,۰۰۸۳)	(۰,۰۰۸۲)	(۰,۰۰۸۲)	(۰,۰۰۸۴)	(۰,۰۰۷۹)
P-value	۰,۰۰۰۰	۰,۰۰۰۰	۰,۰۳۹۴	۰,۱۳۴۱	۰,۰۰۰۰	۰,۰۰۰۰

تمامی ضرایب در سطح ۹۹درصد معنی‌دار است به جز ** که در سطح ۹۵درصد معنی‌دار بوده و * که معنی‌دار نمی‌باشد.

با توجه به اطلاعات جدول (۴) درمی‌یابیم که بازده و حجم معاملات که به نوعی نشان‌دهنده نقدشوندگی سهام می‌باشد، در صورت وقوع دستکاری افزایش یافته است. همانطور که انتظار می‌رفت در صورت وقوع دستکاری، افزایش در میانگین خرید و فروش شرکت‌های حقوقی نسبت به افزایش در میانگین خرید و فروش افراد حقیقی کمتر بوده است. به عبارتی دستکاری سهام بیشتر توسط افراد حقیقی بازار سرمایه اتفاق افتاده است.

نکته قابل توجه در نتایج، منفی بودن ضریب α_1 در متغیر میانگین فروش حقیقی به میزان۰,۱۲۴۴− می‌باشد که باید دلیل آن را در اتفاقات دوره دستکاری پی گرفت. معمولا در ابتدای فرآیند دستکاری، حقیقی‌ها دستکاری کننده اقدام به خرید سهام با حجم بالا نموده و در نیمه دوم

دوره دستکاری، به سرعت سهام دستکاری‌شده را می‌فروشند. این مورد سبب می‌شود که برآیند میانگین فروش افراد حقیقی در دوره دستکاری کمتر از دوره‌های عدم دستکاری باشد.

با استفاده از نتایج روابط (۱۵) و (۱۶) به بررسی تفکیک نمونه به دو دسته سهام دستکاری شده و دستکاری نشده می‌پردازیم. نتایج در جدول (۵) ارائه گردیده است. همانطور که از نتایج مشخص است تمامی متغیرها در سطح ۹۵درصد معنی‌دار بوده و می‌توان بین داده‌های دستکاری‌شده و دستکاری نشده تمایز قائل شد.

پس از اطمینان از تطبیق داده‌ها با مطالعات پیشین، به آزمون مدل‌های مورد بررسی بر روی داده های آماده‌سازی‌شده می‌پردازیم. رتبه‌بندی مدل‌ها بر اساس شاخص اندازه F_2 درنظر گرفته شده است، زیرا اشتباه در دسته‌بندی، هزینه متقارن ندارد. به عبارتی اگر سهام واقعاً دستکاری شده باشد و به اشتباه در دسته دستکاری نشده قرار گیرد به مراتب هزینه بیشتری نسبت به حالتی که واقعاً دستکاری نشده باشد و به اشتباه در دسته دستکاری شده قرار گیرد، به سیستم تحمیل می‌شود.

جدول ۵. مقایسه عملکرد مدل‌های یادگیری ماشین

مدل	دقت (P)	بازخوانی (R)	اندازه F1	اندازه F2
Decision Tree	٪۶۶/۶۵	٪۶۹/۰۶	٪۶۷/۸۳	٪۶۸/۵۶
Forest Random	٪۱۰۰	٪۴۴	٪۶۱/۱۱	٪۴۹/۵۵
MLP	٪۸۳/۳۳	٪۳۶/۵۳	٪۵۰/۷۹	٪۴۱/۱۵
SVM	٪۱۰۰	٪۳۴/۶۱	٪۵۱/۴۲	٪۳۹/۸۲
Regression Logistic	٪۶۶/۶۷	٪۲۲/۲۲	٪۳۳/۳۳	٪۲۵/۶۴
VAE-LSTM	۷۸%	۴۰%	۲۳,۱%	۷۱/۴۹%

همانطور که از نتایج ارائه‌شده استنتاج می‌شود، پس از مدل ترکیبی VAE-LSTM، مدل درخت تصمیم در رتبه بعدی قرار گرفته است که از نقطه‌نظر توازن بین شاخص دقت و بازخوانی نیز وضعیت مناسبی دارد. این مورد بیان می‌کند که شاید یکی از موثرین راه‌ها برای شناسایی دستکاری قیمت استفاده از قواعد از قبل تعیین شده‌ای باشد که به وسیله مدل‌های درخت تصمیم استخراج شده و قابلیت به‌روزرسانی در فواصل زمانی مختلف دارد.

جدول (۶) جزییات توانایی مدل VAE-LSTM (بهترین مدل) در شناسایی دستکاری سهام را توضیح می‌دهد. برای هر بار تست مدل، از ۷،۱۳۰ داده استفاده شده که ۵۳۷ داده آن دستکاری شده و مابقی دستکاری نشده، بوده‌اند. با توجه به نتایج، مدل توانسته است از ۵۳۷ روز معاملاتی دستکاری شده، حدود ۳۶۷ روز را به درستی شناسایی کند و تنها در شناسایی ۱۷۰ مورد ناتوان بوده که آمار قابل قبولی می‌باشد. همچنین مدل مذکور تنها ۵۳ روز معاملاتی را به اشتباه به عنوان دستکاری شده تشخیص داده که در واقعیت در آن روزها دستکاری اتفاق نیفتاده است.

جدول ۶. جزییات عملکرد مدل پیشنهادی در شناسایی دستکاری سهام

داده‌ها	تعداد روز معاملاتی	VAE-LSTM	شاخص ارزیابی
دستکاری شده	۵۳۷	367	True Positive (TP)
		۱۷۰	False Negative (FN)
دستکاری نشده	6593	۵۳	False Positive (FP)
		6540	True Negative (TN)
کل داده‌ها	۷۱۳۰	۷۱۳۰	–

پس از مشخص شدن اینکه تابع احتمال محاسبه شده توسط مدل پیشنهادی VAE-LSTM احتمال دستکاری سهام را به درستی و با درصد خطای کمی نمایش می‌دهد، مستقل از تعداد سهام مورد آزمون، با بررسی احتمال دستکاری سهام در بازار صعودی، نزولی و متعادل نتایج ارزشمندی به‌دست آمد. برای این منظور براساس روند شاخص کل بازار، دوره صعودی بازار سرمایه را در بازه زمانی ۱۳۹۸/۱۲/۰۱ تا ۱۳۹۹/۰۵/۳۱، دوره نزولی بازار سرمایه را در بازه زمانی ۱۳۹۹/۰۵/۲۱ تا ۱۳۹۹/۰۸/۲۰ و سال ۱۴۰۰ را به عنوان دوره تعادل بازار سرمایه در نظر گرفته‌ایم.

احتمال دستکاری برای حدود ۲۵۰ سهم مورد معامله در بازار سرمایه در آخرین روز هر دوره توسط مدل محاسبه شده که اطلاعات توصیفی آن در جدول ذیل قابل مشاهده است.

جدول ۷. آمار توصیفی شرکت‌های مورد بررسی در دوره صعود، نزول و تعادل بازار سرمایه

وضعیت بازار	تعداد شرکت مورد بررسی	احتمال دستکاری	
		میانگین	انحراف معیار
متعادل	۲۰۷	٪۰,۱۶۶۳۸	٪۰,۰۰۵۹۵
صعودی	۲۴۰	٪۲,۱۸۹۵۱	٪۱,۲۱۷۳۷
نزولی	۲۵۰	٪۰,۰۱۷۹۳	٪۰,۰۰۰۲۵

براساس آزمون های انجام‌شده احتمال دستکاری سهام در بازار صعودی بیشتر از بازار متعادل و در بازار متعادل بیشتر از بازار نزولی است. نتایج از نظر مفهومی با واقعیت منطبق است. از آن‌جایی که در بازار سرمایه ایران امکان فروش استقراضی وجود ندارد، دستکاری‌کنندگان صرفا می‌توانند با روش دستکاری «افزایش قیمت و خالی کردن» کسب سود نموده و استفاده از روش دستکاری «کاهش قیمت و خرید مجدد» امکان‌پذیر نیست. بنابراین در بازار نزولی ایجاد تغییر روند در بازار سرمایه منابع زیادی لازم دارد که همین عامل انگیزه دستکاری کردن سهم را کاهش می‌دهد.

۵. بحث و نتیجه‌گیری

دستکاری سهام از دیرباز وجود داشته و قدمت آن تقریبا به زمان تأسیس بازارهای مالی باز می‌گردد. یکی از اصلی‌ترین مواردی که می‌تواند منجر به عدم کارایی بازار در ایفای وظایفش شود، دستکاری سهام با هدف کسب منفعت از طریق اقدامات فریبکارانه است. با توجه به توسعه تکنولوژی و

الگوریتم‌های معاملاتی پیچیده مبتنی بر هوش مصنوعی در معاملات سهام، لزوم بهره‌گیری از ابزارهای توسعه یافته مبتنی بر یادگیری ماشین برای شناسایی دستکاری قیمت‌ها در بازار سرمایه توسط نهادهای نظارتی ضروری تلقی می‌گردد. باتوجه به اینکه در بازار بورس اوراق بهادار ایران، سهام دستکاری شده توسط ناظر اعلام نمیگردد، نیازمند تشکیل پایگاه داده‌ای مناسب و قابل اتکا بودیم. بر این اساس از دو طریق ۱) با روش‌های آماری آزمون تسلسل، چولگی، کشیدگی و همچنین چهار آزمون شناسایی ناهنجاری متنی بر روی حجم و بازده (۱۵ و ۳۰ روزه) چنانچه در یک تاریخ، حداقل ۵ آزمون از ۷ آزمون انجام شده روی سهام مشخصی، دلالت بر دستکاری سهم داشته باشد، برچسب یک و در غیر این‌صورت برچسب صفر به آن سهم در تاریخ مذکور می‌دهیم. ۲) سهامی که نظر به مشخصات خاص آن‌ها مانند نقدشوندگی و شفافیت بالا، از نظر خبرگان (پرسشنامه) با احتمال خیلی کمی مورد دستکاری قرار می‌گیرند، انتخاب شده و متغیرهای تصادفی مطابق با الگوی دستکاری قیمت (مانند الگوی افزایش قیمت و خالی کردن)، به صورت تعمدی به داده‌های نرمال تزریق شده‌است. در نهایت داده‌های تزریق‌شده در روز معاملاتی مشخص، به عنوان سهام دستکاری شده و سایر سهام به عنوان سهام دستکاری نشده در نظر گرفته شده است.

برای این کار از اطلاعات ۷۳ سهم از ۱۹ صنعت پذیرش شده در بورس اوراق بهادار تهران در فاصله زمانی ۱۳۹۸ الی ۱۴۰۲ استفاده شده است. تعداد کل روزهای معاملاتی حدود ۷۱،۳۰۰ بوده که از این بین در ۵۳۷ روز معاملاتی دستکاری رخ داده و در سایر روزها دستکاری اتفاق نیفتاده است.

پس از اجرای مدل‌های پیشنهادی پژوهش و سایر مدل‌های یادگیری عمیق، بر اساس شاخص ارزیابی عملکردی F2، به ترتیب مدل‌های VAE-LSTM، درخت تصمیم، جنگل تصادفی، شبکه عصبی چندلایه، ماشین‌بردار پشتیبان، رگرسیون لجستیک عملکردی بهتری از خود نشان دادند. مقدار حدودی شاخص ارزیابی عملکردی F2 مدل‌های مذکور به ترتیب : ۷۲درصد، ۶۹درصد، ۵۰درصد، ۴۱درصد، ۴۰درصد و ۲۶درصد. است. نتایج به مراتب بهتر نسبت به پژوهش قبلی به‌دست آمد.

ذکر این نکته ضروری است که سایر مدل‌های یادگیری ماشین استفاده شده در این پژوهش نیز عملکرد مناسبی به خصوص در شاخص ارزیابی دقت داشته اما متأسفانه از نظر شاخص عملکردی بازخوانی که مهم‌تر می‌باشد، عملکرد ضعیفی داشته‌اند.

پس از تعیین مدل پیشنهادی به عنوان مدل برگزیده، براساس شاخص کل بورس تهران، دوره صعودی بازار سرمایه را در بازه زمانی ۱۳۹۸/۱۲/۰۱ تا ۱۳۹۹/۰۵/۳۱، دوره نزولی بازار سرمایه را در بازه زمانی ۱۳۹۹/۰۵/۲۱ تا ۱۳۹۹/۰۸/۲۰ و سال ۱۴۰۰ را به عنوان دوره تعادل بازار سرمایه درنظرگرفته‌ایم. احتمال دستکاری در بازار صعودی نسبت به متعادل و در بازار متعادل نسبت به

نزولی بیشتر بوده و p-value آزمون‌های مذکور به ترتیب ۰/۰۰۲۴ و ۰/۰۰۱۹ می‌باشد که در سطح اطیمنان ۹۹درصد معنی‌دار می‌باشد. این نتایج به طور کلی با سایر مطالعات پیشین نیز هم‌سو می‌باشد.

۶. پیشنهادها و محدودیت‌ها

- افراد حقیقی بازار سرمایه با سرمایه خرد، معمولاً هدف اصلی دستکاری‌کنندگان و سودجویان بازار سرمایه می‌باشد و از بی‌اطلاعی این افراد در جهت رسیدن به اهداف دستکاری سهام بهره می‌برند. پیشنهاد می‌شود شرکت‌های پردازش اطلاعات مالی در کنار نهاد ناظر بازار با استفاده از نرم افزار و اپلیکیشن نسبت به اطلاع‌رسانی لازم اقدام نمایند.
- در مدل فعلی برچسب‌گذاری داده‌های دستکاری‌شده، از روش‌های آماری و پارامتراهای بنیادی و تکنیکال استفاده شده‌است. در پژوهش‌های می‌توان با رعایت پیچیدگی مدل از سایر اندیکاتورها نیز بهره گرفت.
- مشابه بسیاری از پژوهش‌های انجام شده، این پژوهش نیز دارای محدودیت‌هایی می‌باشد که در ادامه ذکر می‌گردد.
- مهم‌ترین محدودیت، دسته‌بندی سهام به دو نوع دستکاری‌شده و نشده به‌صورت دستی می‌باشد. معمولاً راه دقیق‌تر این است که سهام دستکاری‌شده توسط مراجع قانونی و نهاد ناظر افشا شود که در ایران این مهم صورت نمی‌پذیرد؛ هرچند همان‌طورکه قبلاً اشاره شد، حتی در کشورهای پیشرفته نیز، بسیاری از دستکاری‌ها توسط نهاد ناظر شناسایی نمی‌شود و اطمینان کامل نسبت به داده‌های دستکاری‌شده در آن کشورها نیز وجود ندارد.
- وجود دامنه نوسان در بازار سرمایه ایران و تشکیل صف‌های خرید و فروش متوالی، روند شناسایی دستکاری سهام را به شدت دشوار کرده است چرا که وجود محدودیت در داده‌ها، بعضاً عملکرد مدل‌های یادگیری ماشین را تحت تأثیر قرار می‌دهد.

به نظر می‌رسد با بررسی‌های آماری و احتمال شناسایی هر یک از مدل‌ها، می‌توان استراتژی‌های مختلفی با استفاده از ترکیب نتایج مدل‌های یادگیری عمیق طراحی نمود که در شناسایی دستکاری سهام، کاراتر باشد. بررسی این موضوع به پژوهش‌های آتی واگذار می‌گردد.

سپاسگزاری

از کلیه افرادی که ما را در انجام این پژوهش یاری نمودند تشکر می‌نماییم. در این پژوهش از سازمان، نهاد یا شخصی کمک مالی دریافت نشده است.

References

Aggarwal, R.K.,Wu, G., (2006). Stock market manipulations. *Journal of Business.* **79**(4), 1915-1953. Stock market manipulations.

Ahmed, M., Choudhury, N., Uddin, S. Anomaly detection on big data in financial markets (2017) IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2017. IEEE.

Al-Thani, H.A., Detecting market manipulation in stock market data (2017). Detecting market manipulation in stock market data.

Allen, F.,Gale, D., (1992). Stock-price manipulation. *Review of Financial Studies.* 5(3), 503-529. Stock-price manipulation.

Brownlee, J., (2016). Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras. 2016: Machine Learning Mastery. Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras.

Cao, Y., et al (2014). Detecting price manipulation in the financial market. in 2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr). IEEE.

Cherian, J.,Kuriyan, V., (1995). Informationless manipulation in a market maker-type economy. Available at SSRN 6501. Informationless manipulation in a market maker-type economy.

Cherian, J.A.,Jarrow, R.A., (1995). Market manipulation. Handbooks in Operations Research and Management Science. 9, 611-630. Market manipulation.

Chollet, F., (2017). Deep learning with Python. Simon and Schuster. Deep learning with Python.

Comerton-Forde, C.,Putniņš, T.J (2014). Stock price manipulation: Prevalence and determinants. *Review of Finance*, 18(1), 23-66. Stock price manipulation: Prevalence and determinants.

Cumming, D., et al. (2020). Market manipulation and innovation. *Journal of Banking & Finance.* 120, 105957. Market manipulation and innovation.

Diaz, D ,.Theodoulidis, B., Sampaio, P. (2011). Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices. Expert Systems with Applications. 38(10), 12757-12771. Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices.

Ergün, H.O., et al. (2020). Stock market manipulation in an emerging market of Turkey: how do market participants select stocks for manipulation? Applied Economics Letters: 1-5. Stock market manipulation in an emerging

market of Turkey: how do market participants select stocks for manipulation.

Fallah Shams, Mirfaiz; Teymouri Shandi, Ali (2005). Designing a model for predicting price manipulation in the stock exchange. Imam Sadeq University (AS) Research Quarterly. 27. (in persian)

Fallah Shams, Mirfiz; Kordloui, Hamid Reza; Rashno, Mehdi (2012). Investigating price manipulation in Tehran Stock Exchange using support vector machine model. Financial Research, 14(1), 69-84. (in persian)

Fallahzadeh, Hadith (2018). Detecting Stock Manipulation Using Data Mining Techniques. University of Science and Art affiliated with Jihad Daneshgahi. Tehran. (in persian)

Fischel, D.R.,Ross, D.J., (1991). Should the Law Prohibit "Manipulation" in Financial Markets? *Harvard Law Review.* 105(2), 503-553. Should the Law Prohibit" Manipulation" in Financial Markets.

Géron, A., (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. 2019: O'Reilly Media. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.

Golmohammadi, K., Zaiane, O.R., Díaz, D. (2014). Detecting stock market manipulation using supervised learning algorithms. 2014 International Conference on Data Science and Advanced Analytics (DSAA). IEEE.

Golmohammadi, K.,Zaiane, O.R. (2015). Time series contextual anomaly detection for detecting market manipulation in stock market. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE.

Griffin, J.M.,Shams, A. (2018). Manipulation in the VIX: *Review of Financial Studies.* 31(4), 1377-1417. Manipulation in the VIX.

Hart, O.D. (1977). On the profitability of speculation. *Quarterly Journal of Economics*, 579-597. On the profitability of speculation.

Hillion, P.,Suominen, M. (2004). The manipulation of closing prices. *Journal of Financial Markets.* 7(4), 351-375. The manipulation of closing prices.

Jarrow, R.A. (1992). Market manipulation, bubbles, corners, and short squeezes. *Journal of financial and Quantitative Analysis*: 311-336. Market manipulation, bubbles, corners, and short squeezes.

Lee, E.J., Eom, K.S., Park, K.S., (2013). Microstructure-based manipulation: Strategic behavior and performance of spoofing traders. Journal of Financial Markets. 16(2), 227-252. Microstructure-based manipulation: Strategic behavior and performance of spoofing traders.

Li, A., Wu, J., Liu, Z., (2017). Market manipulation detection based on classification methods. *Procedia Computer Science*. 122, 788-795. Market manipulation detection based on classification methods.

Öğüt, H., Doğanay, M.M., Aktaş, R. (2009). Detecting stock-price manipulation in an emerging market: The case of Turkey. *Expert Systems with Applications.* 36(9), 11944-11949. Detecting stock-price manipulation in an emerging market: The case of Turkey.

Postforosh, Mohamadhossin; Naser sadrabadi, Alireza; Moendin, Mahmoud (2013). Measuring price manipulation using quadratic audit analysis models and genetic algorithms based on artificial neural networks. *Journal of Financial knowledge of securities analysis*, 143-119.(in persian)

Putnins, T. (2020). An Overview of Market Manipulation. Corruption and Fraud in Financial Markets Malpractice, Misconduct and Manipulation. An Overview of Market Manipulation.

Rabiei,Reyhaneh; Nadiri, Mohammad; Peymani, Muslim; Jaberizadeh, Ali (2018). Investigating the effect of price manipulation on market efficiency in Tehran Stock Exchange. *Quarterly Journal of Tehran Stock Exchange*, 11(43), 113-131. (in persian)

Roodposhti, F.R., Shams, M.F., Kordlouie ,H., (2011). Forecasting stock price manipulation in capital market. *International Journal of Economics and Management Engineering.* 5(8), 957-967. Forecasting stock price manipulation in capital market.

Shah, S., Ismail, I., Shahrin, A., (2019). Stock market manipulation: A comparative analysis of East Asian emerging and developed financial markets. *Management Science Letters*. 9(1), 183-192. Stock market manipulation: A comparative analysis of East Asian emerging and developed financial markets.

Shams, Shahabuddin; Ataei, Behrouz (2016). Identifying stock price manipulation through a combined genetic algorithm-artificial neural network model and the SQDF model. *Financial Management Strategy*, 4(3), 149-171.(in persian).

Smith, M.V. (1980). Preventing the Manipulation of Commodity Futures Markets: To Deliver or Not to Deliver. Hastings LJ. 23, 1569. Preventing the Manipulation of Commodity Futures Markets: To Deliver or Not to Deliver.

Westphal, C.,Blaxton, T., (1998). Data mining solutions: methods and tools for solving real-world problems. 1998: John Wiley & Sons, Inc. Data mining solutions: methods and tools for solving real-world problems.

Zhang, K. (2019). Stock market prediction based on generative adversarial network. Procedia computer science. 147, 400-406. Stock market prediction based on generative adversarial network.

Zheng, Y.-J. (2018). Generative adversarial network based telecom fraud detection at the receiving bank. *Neural Networks*. 102, 78-86. Generative adversarial network based telecom fraud detection at the receiving bank.

استناد

حبیب‌زاده، سید محمدرضا؛ رستگار، محمدعلی؛ غلامی جمکرانی، رضا و چاوشی، سید کاظم (۱۴۰۳). بررسی دستکاری قیمت‌ها در بازار بورس ایران با استفاده از مدل ترکیبی خودرمزگذار متغیر–حافظه کوتاه مدت طولانی. *چشم‌انداز مدیریت مالی*، ۱۴(۴۶)، ۱۳۳–۱۶۱.

Citation

Habibzadeh, Seyed Mohammadreza; Rastegar, Mohammad Ail; Golami Jamkarni, Reza & Chavoshi, Sayyed Kazem (2024). Stock Price Manipulation in the Iran Stock Market Using VAE-LSTM Hybrid Model. *Journal of Financial Management Perspective*, 14(46), 133 - 161. (*in Persian*)