Quarterly Financial Accounting
E-ISSN: 2783-4417
Vol. 16, Issue 3, No. 63, Autumn 2024, p 1-15
Received: 03.09.2024 Accepted: 27.04.2025

Research Paper

Iran Market Manipulation Detection Based On Machine Learning Algorithms Methods

Seyed Mohammad Reza Habibzadeh
Ph.D Student, Department of Financial Engineering, Qo.C, Islamic Azad University, Qom, Iran.
Em.habibzadeh@ iau.ac.ir

Reza Gholami-Jamkarani *
Associate Professor, Department of Accounting, Qo.C. Islamic Azad University, Qom, Iran.
reza.gholami@iau.ac.ir

Mohammad Ali Rastegar Sorkheh
Assistant Professor , Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.
ma_rastegar@modares.ac.ir

Seyed Kazem Chavoshi
Assistant Professor , Department of Insurance, Banking and Customs Management Faculty of Management, Kharazmi University, Tehran, Iran.
chavoshi@khu.ac.ir

Abstract
One of the most important prerequisites for the expansion and deepening of the capital market as the main economic market of the country is the trust of the actors and beneficiaries of this market in its efficiency and correctness as a basis for determining the price of financial assets fairly and away from fraud.therefore, the purpose of this research is to identify the manipulation of stock prices of companies accepted in the capital market of Iran, and with regard to the development of technology and complex trading algorithms based on artificial intelligence in stock transactions in the capital markets, taking advantage of Developed tools based on machine learning are considered essential by regulatory institutions in order to identify price manipulation in the capital market. In this research, using statistical methods such as sequence, skewness, kurtosis and textual anomaly detection test in order to prepare and label data, from a variety of machine learning methods including decision tree, random forest, support vector machine, network Multilayer neural and logistic regression have been used to identify stock manipulation during the second half of 1398 to the end of 1402 in the Iranian capital market. For this purpose, data on 73 stocks from 19 industries listed on TSE. The total number of trading days was about 71,300, of which manipulation occurred on 537 trading days and no manipulation occurred on the other days. The decision tree algorithm performed better than the other compared methods from the point of view of the balance between accuracy index, readability and F2 index. This result states that one of the most effective ways to identify price manipulation is to use predetermined rules that are extracted by decision tree models and can be updated at different time intervals. Based on the obtained results, the volume variable on the same day of manipulation (vol0) and the volume variable on a trading day before the manipulation (vol1) are the most important in identifying stock manipulation.

Keywords: Capital Market, Stock Price Manipulation, Artificial Intelligence, Data Labeling, Decision Tree.

Introduction: With the development of technology and complex trading algorithms, manipulation of stocks has occurred more easily, and this makes the use of computer such as AI and deep learning inevitable to detect manipulation by supervisor institutions. The purpose of this research is to identify stock manipulation in Iran's capital market. For this, the information of 73 stocks accepted in TSE between 2018 and 2022, the number of trading days is about 71,300.

* Corresponding author
Habibzadeh, S.M.R., Gholami-Jamkarani, R., Rastegar Sorkheh, M.A., Chavoshi,S. K. (2024). Iran Market Manipulation Detection Based On Machine Learning Algorithms Methods. *Quarterly Financial Accounting*, 16(63): 1-15.

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

Method and Data: Due to the non-announcement of manipulated stocks by the stock market supervisor, data identification has been done in two ways. 1) Using statistical tests such as abnormal returns, manipulated stocks and the exact date of manipulation have been determined. 2) Random data has been injected into the time series that are not manipulated with high confidence (expert's opnion) by simulating the stock manipulation pattern. In next step, using the algorithm is designed to compare with some ML models such as DT, SVM, etc. It calculates the probability of stock manipulation in each trading day.

Findings: After implementing the proposed research models and other deep learning models, based on the F2 performance evaluation index, decision tree, random forest, multilayer neural network, SVM, and logistic regression models showed better performance. The approximate value of the F2 functional evaluation index of the mentioned models is 72%, 69%, 50%, 41%, 40% and 26%, respectively.

Conclusion and discussion: The decision tree algorithm performed better than the other compared methods from the point of view of the balance between accuracy index, readability and F2 index. This result states that one of the most effective ways to identify price manipulation is to use predetermined rules that are extracted by decision tree models and can be updated at different time intervals. Based on the obtained results, the volume variable on the same day of manipulation (vol0) and the volume variable on a trading day before the manipulation (vol1) are the most important in identifying stock manipulation.

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

فصلنامه حسابداری مالی
سال شانزدهم، شماره ۶۳، پاییز ۱۴۰۳، ص ۱۵-۱
تاریخ دریافت: ۱۴۰۳/۰۶/۱۳ تاریخ پذیرش: ۱۴۰۴/۰۲/۰۷

مقاله پژوهشی

شناسایی دستکاری قیمت سهام در بازار سرمایه ایران با استفاده از الگوریتم‌های یادگیری ماشین

سید محمدرضا حبیب‌زاده
دانشجوی دکتری، گروه مهندسی مالی، واحد قم، دانشگاه آزاد اسلامی، قم، ایران.
Em.habibzadeh@ iau.ac.ir

رضا غلامی جمکرانی*
دانشیار، گروه حسابداری، واحد قم، دانشگاه آزاد اسلامی، قم، ایران.
reza.gholami@iau.ac.ir

محمدعلی رستگار سرخه
استادیار، گروه مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.
ma_rastegar@modares.ac.ir

سید کاظم چاوشی
استادیار، گروه آموزشی مدیریت بانک، بیمه و گمرک دانشکده مدیریت، دانشگاه خوارزمی، تهران، ایران.
chavoshi@khu.ac.ir

چکیده:
از با اهمیت‌ترین پیش‌نیازهای گسترش و تعمیق بازار سرمایه به‌عنوان اصلی‌ترین بازار اقتصادی کشور، اعتماد بازیگران این بازار به کارایی و درستی آن به‌عنوان مبنایی برای تعیین عادلانه و به دور از تقلب قیمت دارایی‌ها است. از این رو، هدف از انجام این پژوهش شناسایی دستکاری قیمت سهام شرکت‌های پذیرفته شده در بازار سرمایه ایران است و با توجه به توسعه تکنولوژی و الگوریتم‌های معاملاتی پیچیده مبتنی بر هوش مصنوعی در معاملات سهام در بازارهای سرمایه، بهره‌گیری از ابزارهای توسعه‌یافته مبتنی بر یادگیری ماشین به‌منظور شناسایی دستکاری قیمت‌ها در بازار سرمایه توسط نهادهای نظارتی ضروری تلقی می‌گردد. در این پژوهش با استفاده از روش‌های آماری مانند تسلسل، چولگی، کشیدگی و آزمون تشخیص ناهنجاری متنی به‌منظور آماده‌سازی و برچسب‌گذاری داده‌ها، از انواع روش‌های یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی، ماشین‌بردار پشتیبان، شبکه عصبی چندلایه و رگرسیون لجستیک برای شناسایی دستکاری سهام طی بازه زمانی نیمه دوم ۱۳۹۸ تا پایان ۱۴۰۲ پرداخته شده است. برای این کار از اطلاعات ۷۳ سهم از ۱۹ صنعت پذیرش شده در بورس اوراق بهادار 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. الگوریتم درخت تصمیم، از منظر توازن بین شاخص دقت، بازخوانی و شاخص F_2 نسبت به سایر روش‌ها عملکرد بهتری داشت. این نتیجه بیان می‌کند که یکی از مؤثرین راه‌ها برای شناسایی دستکاری قیمت استفاده از قواعد از قبل تعیین شده‌ای است که به‌وسیله مدل‌های درخت تصمیم استخراج شده و قابلیت به‌روزرسانی در فواصل زمانی مختلف را دارد. بر اساس نتایج به‌دست‌آمده متغیر حجم در همان روز دستکاری و متغیر حجم در یک روز معاملاتی قبل از دستکاری بیشترین اهمیت را در شناسایی دستکاری سهام دارند.

واژه‌های کلیدی: بازار سرمایه، دستکاری قیمت سهام، برچسب‌گذاری داده، درخت تصمیم.

---

* نویسنده مسئول

حبیب‌زاده، سید محمدرضا، غلامی جمکرانی، رضا، رستگار سرخه، محمدعلی، چاوشی، سید کاظم. (۱۴۰۳). شناسایی دستکاری قیمت سهام در بازار سرمایه ایران با استفاده از الگوریتم¬های یادگیری ماشین. *فصلنامه حسابداری مالی*،۱۶(۶۳): ۱۵-۱.

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

مقدمه

بورس اوراق بهادار از اصلی‌ترین بازارها در فضای اقتصاد هر کشوری است. از با اهمیت‌ترین پیش‌نیازهای گسترش و تعمیق بازار سرمایه، اعتماد بازیگران و ذی‌نفعان این بازار به کارایی و درستی آن به‌عنوان مبنایی برای تعیین عادلانه و به دور از تقلب قیمت دارایی‌های مالی است و میزان اعتماد عمومی به بازار بورس متناسب با ریسک در این بازارها، به‌عنوان یکی از عوامل اصلی تأثیرگذار بر تصمیم‌گیری سرمایه‌گذاران محسوب می‌گردد. با توسعه بازارهای مالی و درک صحیح از مخاطرات که منجر به تصویب قوانین سخت‌گیرانه و نظارت دقیق شده است، دستکاری سهام سخت‌تر از گذشته اتفاق می‌افتد، اما نمی‌توان انتظار داشت که چالش دستکاری در بازارها به‌طورکلی حل شود؛ چرا که در اکثر موارد، دستکاری با روش‌های جدید و پیچیده انجام می‌گیرد. (ارگان، یالامان، ماناهوف و ژانگ[1]، ۲۰۲۰). کشف دستکاری سهام با توجه به توسعه برنامه‌های کامپیوتری جدید و الگوریتم‌های معاملاتی پیچیده، سخت‌تر از قبل شده است. همچنین با گسترش بازار مالی و افزایش تعداد شرکت‌های بورسی و افزایش حجم معاملات، نظارت بر اقدامات متقلبانه و دستکاری سهام صرفاً توسط نیروی انسانی و بدون کمک گرفتن از ابزارهای کامپیوتری تقریباً ناممکن است. از این رو نهاد ناظر به برنامه‌های کامپیوتری سریعی نیاز دارد تا معاملات و روند قیمتی سهام را تحت نظر قرار داده و برای شناسایی دستکاری هشدارهای لازم را کسب و اقدام متناسب نماید (لینگارون و تنگامچیت[2]، ۲۰۱۸).

در کشورها و بازارهای نوظهور که نظارت کمتری بر آن‌ها حاکم است و از فقر انتشار دقیق و به لحظه اطلاعات رنج می‌برند، لزوم کشف دستکاری قیمت سهام، روش‌های شناسایی و اقدام به حذف آن به مراتب از اهمیت ویژه‌ای برخوردار است.

در بازار بورس 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 در بازار سرمایه حداقل در مقاطعی از زمان در معرض دستکاری بوده یا به‌طور جدی دستکاری می‌شوند. 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 هوش مصنوعی برای شناسایی دستکاری قیمت‌ها در بازار سرمایه توسط نهاد نظارتی ضروری تلقی می‌گردد.

در این پژوهش تلاش شده است ضمن معرفی روش مناسب شناسایی به هنگام سهام دستکاری شده، برخی پارامترهای با اهمیت توسط روش‌های یادگیری ماشین به نهاد ناظر بورس معرفی گردد. همچنین برای برچسب‌گذاری سهام دستکاری شده علاوه بر متغیرهای بنیادی از متغیرهای تکنیکال و نموداری نیز استفاده نموده‌ایم. این کار منجر به افزایش اطمینان در پایگاه داده می‌شود. در نهایت نیز روش پیشنهادی، با توجه به عملکرد مقایسه‌ای بهتر نسبت به سایر مدل‌های مورد بررسی، علاوه بر ناظر بازار به سایر فعالین بازار بورس کمک خواهد کرد تا به صورت به هنگام از احتمال دستکاری سهام مطلع شده و به‌طور متناسب تصمیم لازم را اتخاذ نمایند.

مبانی نظری و پیشینه پژوهش

با توسعه بازارهای مالی و درک صحیح از مخاطرات که منجر به تصویب قوانین سخت‌گیرانه و نظارت دقیق شده است، دستکاری سهام سخت‌تر از گذشته اتفاق می‌افتد، اما نمی‌توان تصور کرد که مسئله دستکاری در بازارها به‌طورکلی حل شده است؛ چرا که در اکثر موارد در این بازارها، دستکاری با روش‌های جدید و پیچیده انجام می‌گیرد. (ارگان و همکاران، ۲۰۲۰). شاید این‌گونه به نظر برسد که دستکاری سهام، پدیده‌ای نوظهور است، اما قدمت دستکاری به شکل‌گیری اولین معاملات بازمی‌گردد. (دارستون و مک‌کئون[3]، ۲۰۲۰). تاریخ مملو است از تلاش‌های ناموفق که درصدد جلوگیری از سفته‌بازی، دستکاری قیمت و کنترل قیمت کالا بوده‌اند. از قوانین ضد احتکار تالمودیک[4] برای دریافت سود مازاد از سوی بازرگانان تا ممنوع نمودن دستکاری بازارهای در زمان امپراطور رم و نهی از فعالیت سفته‌بازانه غرر در اسلام که به‌عنوان یکی از اصول

[1] Ergün, Yalaman, Manahov & Zhang
[2] Leangarun & Tangamchit
[3] Durston & McKeon
[4] Talmudic

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

اصلی ممنوع نمود (پراکاش[1]، ۲۰۱۱). از اولین پژوهش‌های انجام گرفته در زمینه دستکاری قیمت، پژوهش هارت[2] در سال ۱۹۷۷ است که به‌طور رسمی دستکاری قیمت را بر اساس مدل‌های اقتصادسنجی در بازارهای دارایی مورد تحقیق قرار داده است. هارت، شرایطی را که سفته‌بازی سودآور در یک وضعیت غیر احتمالی ممکن است رخ دهد را مورد بررسی قرار داد. وی دریافت که اگر بازار در حالت تعادل پایدار نباشد و توابع غیرخطی باشند، سفته‌بازان قادر به انجام معاملات سودآور خواهند بود.

مطالعات دستکاری بازار با پژوهش آلن و گیل[3] (۱۹۹۲) به بررسی احتمال دستکاری بر پایه معامله را در چارچوب نظریه انتظارات منطقی منجر به معرفی دسته‌بندی دستکاری سهام: بر پایه اطلاعات، بر پایه اقدام و بر پایه معاملات شد. پس از آن نیز (جارو ۱۹۹۲) به معرفی استراتژی‌های دستکاری سودآور در بازار اشاره کرد (نورعلی دخت، تهرانی و فلاح‌پور، ۱۴۰۲).

مطالعات در زمینه دستکاری معامله محور، به بررسی شرایطی پرداختند که در آن، دستکاری در یک بازار سودآور است یا خیر. فیچل و راس[4] (۱۹۹۲) ادعا کردند که دستکاری معامله محور در بازار کارا، امکان‌پذیر نیست. ژارو[5] (۱۹۹۲)، چریان و ژارو[6] (۱۹۹۵) بر پایه مدل هارت (۱۹۷۷)، شرایطی را یافتند که در آن، امکان دستکاری معامله محور وجود ندارد. در این مدل، دستکاری با حضور بازارگردان عقلایی، درصورتی‌که سفارشات وارد شده به سیستم متقارن باشد، امکان‌پذیر نیست. ژارو نتیجه گرفت تنها شرط کافی برای اینکه استراتژی دستکاری بازار ناموفق باشد، این است که قیمت‌ها تنها بر اساس جمع سهام در اختیار معامله‌گران و نه بر اساس روند گذشته معاملاتی سهم تعیین شود؛ به عبارت دیگر سهام، دارای تکانه آنی[7] نباشند.

آگاروال و وو[8] (۲۰۰۶)، چارچوب آلن و گیل را توسعه دادند و به این نتیجه رسیدند که اگرچه سفته‌بازان و آربیتراژ کنندگان سبب افزایش کارایی بازار می‌شوند، ولی زمانی که شرایط برای دستکاری مهیا است، رقابت بیشتری برای تصاحب سهم می‌نمایند و همین امر باعث می‌شود دستکاری‌کنندگان راحت‌تر از قبل اقدام به دستکاری بازار و در نتیجه کسب سود و ضربه زدن به کارایی بازار نمایند.

پژوهش اگراوال و وو (۲۰۰۶) جز اولین بررسی‌های انجام شده در زمینه بررسی سهام دستکاری شده بود. این مطالعه با جمع‌آوری دستی ادعاهای مطرح شده در دادگاه‌هایی که منجر به صدور رأی مبنی بر رخ دادن دستکاری شده، پایگاه داده‌ای از سهام دستکاری شده 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 بوده و محققین ۱۴۲ مورد دستکاری را طی این دوره شناسایی کردند. نتایج نشان می‌دهد، طی دوره دستکاری افزایش قیمت، پس از دوره دستکاری کاهش قیمت و طی دوره قبل و بعد از دستکاری افزایش نوسان قیمتی در سهم اتفاق افتاده است. نمونه‌های آن‌ها بیشتر شامل سهامی با نقدشوندگی کم بوده و اکثراً سهام به‌وسیله دارندگان اطلاعات نهانی مانند مدیران شرکت، سهامداران عمده، بازارگردانان و کارگزاران دستکاری شده‌اند. پس از آن کامرتون- فورد و پوتنینس[9] (۲۰۱۱)، نمونه‌ای شامل ۱۸۴ شرکت که طی سال‌های ۱۹۹۷ تا ۲۰۰۹ به‌وسیله کمیسیون بورس و اوراق بهادار آمریکا بابت دستکاری آخرین قیمت مورد رسیدگی قرار گرفته بود را بررسی کردند. اکثر دستکاری‌های اتفاق افتاده به‌وسیله مدیران سرمایه‌گذاری صندوق‌ها، 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. بر این اساس، شاخصی برای محاسبه احتمال دستکاری آخرین قیمت طراحی نمودند که از آن می‌توان برای شناسایی دستکاری آخرین قیمت بهره برد.

[1] Prakash
[2] Hart
[3] Allen & Gale
[4] Fischel & Ross
[5] Jarrow
[6] Cherian & Jarrow
[7] Momentum
[8] Aggarwal & Wu
[9] Comerton-Forde & Putniņš

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

کامرتون- فورد و پوتنینس (۲۰۱۴) سعی کردند مشکل فراگیر نبودن نمونه‌های دستکاری شده که در دادگاه تعیین می‌شود را حل کنند. آن‌ها از مدل شناسایی تخمین کنترل شده[1] که شامل احتمال دستکاری بازار و احتمال شناسایی دستکاری به‌وسیله نهاد ناظر و دادگاه‌های رسیدگی‌کننده بود، موفق شدند مشخصه‌های دستکاری داده‌ها را به کل بازار تعمیم داده و نگرانی از قلم افتادن برخی دستکاری‌ها را کم‌رنگ کنند. با استفاده از این روش، آن‌ها نشان دادند که رخداد دستکاری قیمت سهام با مشخصات زیر محتمل‌تر است: ۱) در سهامی با عدم تقارن اطلاعاتی بسیار بالا، ۲) در روزهای پایانی فصل و ماه ۳) سهامی با نقدشوندگی متوسط یا پایین. آن‌ها با استفاده از مدل طراحی شده نشان دادند که درصد کمی (یک از ۳۰۰ مورد) از دستکاری قیمت توسط نهاد ناظر شناسایی می‌شود. این مطالعه نشان می‌دهد با توجه به تورش داده‌ها، امکان اتکا به مدل شناسایی دستکاری به‌شدت دشوار است.

هانگ و چنگ[2] (۲۰۱۳) در مطالعه‌ای به بررسی دستکاری قیمت سهام در بازار سهام تایوان پرداختند. آن‌ها با مجموعه‌ای از داده‌های جمع‌آوری شده به‌صورت دستی، به بررسی الگوهای سهام دستکاری شده و اثرات آن بر بازار پرداختند. نتایج آن‌ها نشان داد که شرکت‌های دستکاری‌شده کوچک بوده و از حاکمیت شرکتی ضعیفی برخوردارند. اکثر موارد دستکاری در این پژوهش با استراتژی معاملاتی افزایش قیمت و خالی کردن در ارتباط بودند. این نوع منجر به افزایش نوسان، افزایش حجم معاملاتی، بازگشت قیمت بلندمدت طی دوره دستکاری می‌شود. این دستکاری‌ها تأثیر مهمی بر کارایی بازار خواهند داشت

دیاز، تئودولیدیس و سامپایو[3] (۲۰۱۱) در پژوهشی به بررسی چالش‌های بکارگیری روش‌های هوش مصنوعی برای شناسایی دستکاری بازار پرداخته و با استفاده از اطلاعات بازار سهام، دستکاری معامله محور را بررسی نمودند. آن‌ها با تشکیل سبدی از دارایی مشابه با سهام دستکاری شده، اختلاف بین بازده سهام دستکاری شده و بازده سبد مشابه را به‌عنوان یکی از متغیرهای اصلی در نظر گرفتند. متغیرهای استفاده شده شامل اختلاف بازدهی سهام دستکاری شده با سبد مرجع، بازدهی غیر نرمال سهام، نقدشوندگی و نوسان سهم می‌باشد. درنهایت پژوهش‌گران با استفاده از روش 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 این کار ابتدا به روش غربال‌گری نمونه‌ای به حجم ۳۹۷ از شرکت پذیرفته شده در بورس اوراق بهادار تهران انتخاب کرده 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 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 بروس اوراق بهادار طراحی کردند. نتایج نشان داد که قدرت پیش‌بینی مدل لاجیت و شبکه‌های عصبی مصنوعی به ترتیب ۹۲/۱٪ و ۹۴/۱٪ بوده است.

در پژوهشی دیگر گلمحمدی، زایان و دیاز[4] (۲۰۱۴) با استفاده از داده‌های مقاله دیاز و همکاران (۲۰۱۱) به بررسی انواع روش‌های نظارت یافته هوش مصنوعی برای شناسایی دستکاری قیمت پرداخته و مدل‌ها را با یکدیگر مقایسه کردند. برخی از مدل‌های مورد استفاده شامل درخت تصمیم، جنگل تصادفی، هوش مصنوعی، بیز ساده،KNN و ماشین‌بردار پشتیبان بودند. مدل بیز ساده از نظر معیار عملکرد F_2 با ۵۳٪ بهترین عملکرد را در شناسایی دستکاری قیمتی داشت.

گلمحمدی و زایان[5] (۲۰۱۵) در پژوهشی دیگر و با استفاده از روش سری زمانی متنی[6](CAD) که یکی از روش‌های هوش مصنوعی بدون نظارت است به شناسایی داده‌های ناهنجار پرداختند. آن‌ها ادعا کردند که جمع‌آوری داده‌های واقعی در زمینه دستکاری سهام، به‌شدت سخت، زمان‌بر و غیرقابل اعتماد است (بسیاری از دستکاری‌ها توسط نهاد ناظر و دادگاه شناسایی نمی‌شود، تاریخ دقیق زمان دستکاری مشخص نیست و ...)؛ بنابراین از سهامی که احتمال دستکاری در آن‌ها وجود ندارد

[1] Detection Controlled Estimation
[2] Haung & Cheng
[3] Diaz, Theodoulidis & Sampaio
[4] Golmohammadi, Zaiane & Díaz
[5] Golmohammadi & Zaiane
[6] Contextual Anomaly Detection

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

استفاده کرده و با تزریق داده تصادفی به‌عنوان ناهنجاری، پایگاه داده جدیدی ساخته و مدل پیشنهادی را روی آن بررسی کردند. نتایج نشان می‌دهد که مدل CAD نسبت به مدل‌ها KNN[1] و RW[2] عملکرد بهتری دارد. در پژوهشی مشابه، الثانی[3] (۲۰۱۷) با بررسی بورس قطر و تغییر اندکی در متغیرهای ورودی مدل CAD، نتایج به مراتب بهتر نسبت به پژوهش قبلی به دست آورد.

شاه، اسماعیل و شاهرین[4] (۲۰۱۹) در مطالعه‌ای به بررسی دستکاری قیمت در بازارهای نوظهور و در حال توسعه شرق آسیا با استفاده از داده‌های سال ۲۰۰۱ الی ۲۰۱۷ و با جمع‌آوری دستی ۲۴۴ سهم دستکاری‌شده پرداختند. ابزار مورد استفاده پژوهش‌گران رگرسیون لجستیک بوده و نتایج برخلاف انتظار نشان می‌دهد که شرکت‌های بزرگ و نقدشونده با احتمال بیشتری مورد دستکاری واقع می‌شوند.

کلوس و کاشف[5] (۲۰۲۰) در پژوهشی با عنوان ترکیب مدل سیستم ایمنی مصنوعی و تحلیل خوشه‌بندی به کشف ناهنجاری در بازار سرمایه اقدام کردند. داده‌های استفاده شده 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، ماراگاتام و کرتیگا دیوی[6] (۲۰۲۲) با استفاده از رویکرد ترکیبی خودرمزگذار با LSTM در تشخیص ناهنجاری مجموعه داده‌های سرور فضای ابری، پرداختند. مقایسه مدل‌ها بر اساس شاخص دقت، یادآوری و F_1 صورت گرفت و بر اساس شاخص F_1 رتبه‌بندی شد. DA − LSTM با ۹۹.۱٪ و SVM با ۸۱٪ و درخت تصمیم ۷۹٪ اندازه‌گیری شد و در نتیجه مدل ترکیبی DA-LSTM بهترین عملکرد را از خود نشان داد.

کوین، تومان و کوران[7] (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.

فرضیه‌های پژوهش

فرضیه اصلی

۱. با استفاده از مدل درخت تصمیم امکان شناسایی به هنگام دستکاری سهام وجود دارد.
۲. مدل درخت تصمیم در شناسایی دستکاری قیمتی سهام نسبت به مدل شبکه عصبی چندلایه، جنگل تصادفی، ماشین بردار پشتیبان و رگرسیون لجستیک عملکرد بهتری دارد.

فرضیه فرعی

۱. متغیر حجم معاملات در شناسایی دستکاری قیمت سهام، اثر مستقیم دارد.
۲. متغیر بازده سهام در شناسایی دستکاری قیمت سهام، اثر مستقیم دارد.
۳. متغیر حجم معاملات نسبت به سایر متغیرها در شناسایی دستکاری قیمت سهام، با اهمیت‌تر است.

[1] K-Nearest Neighbors
[2] Random Walk
[3] Al-Thani
[4] Shah, Ismail & Shahrin
[5] Close & Kashef
[6] Pankajashan, Maragatham & Kirthiga Devi
[7] Quinn, Toman & Curran

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

روش پژوهش

برای اینکه بتوان از روش‌های داده‌کاوی و هوش مصنوعی به‌منظور شناسایی دستکاری سهام استفاده نمود، ابتدا باید اطمینان حاصل کرد که چه سهامی مورد دستکاری واقع شده و زمان دقیق دستکاری آن در چه دوره‌ای بوده است. در بازار سرمایه کشورهای توسعه‌یافته، معمولاً سهام شرکت‌های دستکاری شده، توسط نهاد ناظر و دادگاه‌های ذیصلاح اعلام می‌شود.

متأسفانه در کشور ما، سازمان بورس اوراق بهادار سهام دستکاری شده را مشخص و اعلام نمی‌نماید؛ بنابراین لازم است ابتدا نسبت به تشکیل پایگاه داده شامل سهام دستکاری شده و نشده و تاریخ دقیق دستکاری سهام دستکاری‌شده اقدام نماییم. برای این منظور با استفاده از انواع روش‌های آماری مانند آزمون تسلسل، کشیدگی و چولگی و همچنین روش تشخیص ناهنجاری متنی، سهامی که احتمال دستکاری در آن‌ها وجود دارد را شناسایی می‌کنیم. سپس با استفاده از روش‌های نموداری و جست‌وجوی دیداری بر روی متغیرهای مانند بازدهی غیر نرمال، افزایش ناگهانی حجم معاملات، افزایش نوسان‌پذیری سهم، مقایسه بازدهی سهم مربوطه با بازدهی صنعت آن و رهنمود مقالات قبلی در شناسایی دستکاری قیمتی و همچنین استفاده از اطلاعات سهام بررسی رخداد یا عدم رخداد اتفاق بنیادی خاص، تاریخ دقیق دستکاری شدن سهم را مشخص می‌نماییم. لذا این پژوهش، به لحاظ دسته‌بندی بر مبنای هدف، تحقیقی کاربردی و از حیث روش جمع‌آوری داده‌های مورد نیاز، جز پژوهش‌های توصیفی- تحلیلی است؛ بنابراین در نمونه‌گیری، نمونه تصادفی به اندازه کافی و به قصد تعمیم انتخاب می‌شود.

در این پژوهش به بررسی اطلاعات سهام پذیرفته شده و صندوق‌های سرمایه‌گذاری در سهام بین نیمه دوم سال‌های ۱۳۹۸ الی ۱۴۰۲ پرداخته شده است. طی دوره مذکور سه آزمون آماری تسلسل، کشیدگی، چولگی و چهار آزمون شناسایی ناهنجاری متنی روی حجم و بازده (۱۵ و ۳۰ روزه) برای تمامی سهام مورد بررسی انجام شده است. چنانچه در یک تاریخ، حداقل ۵ آزمون از ۷ آزمون انجام شده روی سهام مشخصی، دلالت بر دستکاری سهام داشته باشد، برچسب یک به آن می‌دهیم. سپس سهامی که برچسب یک گرفته‌اند را با روش‌هایی مانند مقایسه با سبد مشابه (هم صنعت) و بررسی بصری و عوامل شرکتی (صرفاً افزایش سرمایه از محل تجدید به‌عنوان رخداد در نظر گرفته شده است) ارزیابی می‌کنیم. اگر متغیرهای بررسی شده، تأییدکننده دستکاری بود، سهم مذکور در تاریخ موردنظر به‌عنوان سهام دستکاری شده منظور می‌شود.

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.

رابطه (۱) تعداد سلسله مورد انتظار:

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

رابطه (۲) انحراف معیار سلسله:

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

n_1 تعداد بازدهی‌های مثبت و n_2 تعداد بازدهی‌های منفی است. سپس معنادار بودن تفاوت تعداد سلسله‌های شمارش شده با تعداد سلسله‌های مورد انتظار برای متغیر تصادفی از طریق آزمون t مورد بررسی قرار می‌گیرد. اگر آماره آزمون (تفاوت بین تعداد سلسله‌های شمارش شده 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 نتیجه گرفته می‌شود که احتمالاً دستکاری اتفاق نیفتاده است؛ اما درصورتی‌که آماره آزمون در محدوده بحرانی قرار نداشته باشد، تعداد سلسله‌های شمارش شده با تعداد سلسله‌های مورد انتظار تفاوت معناداری داشته و بدین معنی است که طول سلسله‌های سری زمانی،

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

آنقدر طولانی است که با داده‌های تصادفی و مستقل هم‌خوانی ندارد و احتمال وقوع دستکاری وجود دارد. در جدول (۱) نتایج آزمون تسلسل برای یکی از شرکت‌های منتخب در نمونه ارائه شده است.

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
آماره Z	-۴/۸
P-value	تقریباً صفر

با توجه به p-value به‌دست‌آمده جدول (۱)، سهم موردنظر با سطح اطمینان ۹۹٪ در دوره زمان مذکور دستکاری شده است.

آزمون چولگی و کشیدگی

روش دیگر تشخیص احتمال بروز دستکاری قیمت سهام شرکت‌ها، بررسی ضرایب کشیدگی و چولگی بازدهی آن‌ها می‌باشد. اگر بازدهی روزانه شرکت‌ها به‌طور معنی‌داری 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) محاسبه می‌شود.

رابطه (۳) میانگین چولگی و کشیدگی:

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

رابطه (۴) واریانس چولگی و کشیدگی:

$$\text{کشیدگی} = \sqrt{\frac{6n}{(n-2)(n-1)(n+3)(n+5)}}$$

$$\text{چولگی} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$

جدول (۲): نمونه محاسبات چولگی (هرمز، ۳ ماهه سوم ۱۳۹۸)

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

همان‌طور که نتایج جدول (۲) نشان می‌دهد، سهم هرمز در سه ماهه سوم سال ۱۳۹۸ با سطح اطمینان ۹۵٪ دارای چولگی نمی‌باشد و سهم کاوه در سه ماهه چهارم سال ۱۴۰۰ با اطمینان ۹۹٪ دارای کشیدگی می‌باشد.

آزمون شناسایی ناهنجاری متنی

روش‌های عمده 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

[1] Contextual Anomaly Detection

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

سهم مشابه (به‌عنوان مثال سهام هم‌صنعت) به‌عنوان ورودی به مدل در نظر گرفته می‌شود. انتظار می‌رود که سهام مذکور به دلیل تشابه، دارای خواص تقریباً یکسانی باشند؛ بنابراین اگر یکی از خواص سهم انتخابی را بر اساس سایر سهام مشابه، تخمین بزنیم، نباید اختلاف معناداری بین خواص تخمین زده شده و خواص واقعی وجود داشته باشد. حال اگر اختلاف از سطح مشخصی بیشتر باشد، می‌توان ادعا نمود که در سهم مذکور ناهنجاری وجود دارد. به صورت ریاضی اگر چندین سری زمانی$\{X_i|i \in \{1,2,\dots,d\}\}$ از یک صنعت مشابه داشته باشیم، با انتخاب پنجره زمانی[۱]، می‌توان به تخمین سری زمان i–ام بر اساس سایر سری‌های زمانی مطابق با فرمول زیر اقدام نمود.

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

$\phi(X_t)$ تابعی از متغیر ورودی (مثلاً مقدار X_i در زمان t-1 و یا میانگین مقادیر گذشته سری زمانی)، c_t میانگین ویژگی مشخصی از سری‌های زمانی در هر لحظه می‌باشد. ψ، تابعی است که ارتباط بین سری زمانی موردنظر 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 فولاد را در اختیار داریم در هر روز میانگین بازدهی ۵ سهم را محاسبه می‌کنیم (C) حال برای هر یک از سری‌های زمانی، هم‌بستگی بین C و سری زمانی مذکور را محاسبه نموده و از رابطه ۶ که ساده شده رابطه ۵ می‌باشد، سری زمانی موردنظر را برآورد می‌کنیم.

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

حال فاصله اقلیدوسی بین مقدار برآوردی و واقعی ($|\widehat{X_{it}} - X_{it}|$) را محاسبه کرده و درصورتی‌که فاصله مذکور از انحراف معیار مقدار واقعی سهام در پنجره زمانی بیشتر باشد، داده مذکور را به‌عنوان ناهنجاری در نظر می‌گیریم. نتایج شناسایی تاریخ دستکاری دو سهم به‌وسیله 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 موردنظر از نقطه‌نظر حجم معاملات، ناهنجاری داشته و باید طبق رویه تشخیص دستکاری ذکر شده در قسمت قبل، مورد بررسی قرار گیرد.

جدول (۴): محاسبات آزمون CAD بر روی بازده سهم وسدید ۱۳۹۹/۰۴/۲۳

پنجره زمانی	۱۵ روز معاملاتی
هم‌بستگی بین بازده صنعت و سهم	-۰/۴۴۷
میانگین بازده سهم	٪۳/۲۷
تخمین بازده انتظاری سهم	-٪۱/۹۶
بازده واقعی سهم	٪۶/۴۲
فاصله اقلیدوسی	٪۸/۳۹
انحراف معیار بازدهی سهم	٪۵/۹۱

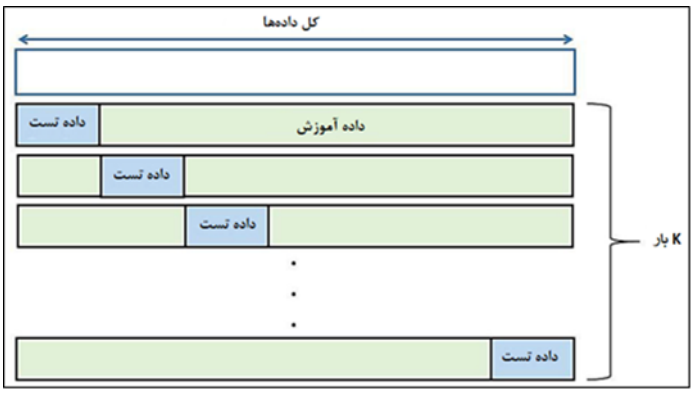
[۱] منظور از پنجره زمانی، تعداد داده‌ها و نحوه انتخاب آن‌ها از یک سری زمانی می‌باشد. به عنوان مثال فرض کنید سری زمانی بازده ۱۰ سهم را برای یک سال به صورت روزانه در اختیار دارید. اگر پنجره زمانی معادل ۲۰ روز باشد که از ابتدای روز اول شروع می‌شود، پنجره زمانی اول، داده‌های روز اول تا روز بیستم همان ماه را شامل می‌شود. پنجره زمانی دوم از روز بیستم همان ماه تا روز دهم ماه بعد می‌باشد و به همین ترتیب تا آخرین روز سال پنجره‌ها تشکیل می‌شود.

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

در جدول فوق، با توجه به اینکه فاصله اقلیدوسی از انحراف معیار بازدهی سهم بیشتر است، در تاریخ محاسبات، سهم موردنظر از نقطه‌نظر بازدهی، ناهنجاری داشته و باید بر اساس رویه تشخیص دستکاری گفته شده در قسمت قبل، مورد بررسی قرار گیرد. در نهایت مطابق با رویه شناسایی دستکاری، سهام دستکاری شده و تاریخ دستکاری سهم را شناسایی کرده و به آن‌ها برچسب یک می‌دهیم. به سایر داده‌ها برچسب صفر اختصاص می‌دهیم.

صحت‌سنجی و پایایی مدل‌های یادگیری ماشین

یکی از مهم‌ترین اقدامات لازم در زمینه ساخت مدل یادگیری ماشین، صحت‌سنجی جامع می‌باشد. صحت‌سنجی و پایایی باعث می‌شود ارزیابی غیرمتعصبانه از مدل طراحی شده داشته باشیم. در مدل‌های یادگیری ماشین روش‌های مختلفی برای صحت‌سنجی وجود دارد که ساده‌ترین آن جداسازی داده‌های آموزش و تست می‌باشد. در پژوهش حاضر علاوه بر جداسازی داده‌های آموزش و تست، از روش اعتبارسنجی متقابل ۵ برابر استفاده کرده‌ایم. در این روش، داده‌ها را به ۵ قسمت مساوی تقسیم کرده و در هر مرحله از ۴ قسمت برای آموزش و از یک قسمت برای تست استفاده می‌کنیم. اجرای مدل را ۵ مرتبه تکرار کرده و میانگین محاسبات عملکردی مدل را به‌عنوان خروجی گزارش نموده‌ایم. شکل (۱) جزییات نحوه انجام اعتبارسنجی k برابر را نمایش می‌دهد.



شکل (۲): چگونگی انجام اعتبارسنجی k برابر

ابزار اندازه‌گیری عملکرد

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

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

[1] True Positive
[2] False Positive
[3] True Negative
[4] False Negative

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

شکل (۲): ماتریس درهم‌ریختگی در مدل‌های یادگیری ماشین

برای اینکه بتوان بین مدل‌ها و داده‌های مختلف مقایسه انجام داد، معیارهای ارزیابی عملکرد بر پایه چهار متغیر ذکر شده طراحی شده است. معیار ارزیابی دقت[1] که از طریق فرمول $P = \frac{TP}{TP+FP}$ و بازخوانی[2] از طریق فرمول $R = \frac{TP}{TP+FN}$ به دست می‌آید. این دو متغیر به صورت جداگانه و برحسب 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 ۷ استفاده می‌شود.

رابطه (۷):
$$F_\beta = \frac{(1+\beta)^2 * P * R}{(\beta^2 * P) + R} = \frac{(1+\beta)^2 * TP}{(1+\beta)^2 * TP + (\beta^2 * FN) + FP}$$

در رابطه ۷، 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. 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.

متغیرهای پژوهش

در جدول (۵) متغیرهای مورد استفاده در پژوهش برای شناسایی دستکاری بازار ذکر شده‌اند.

جدول (۵): متغیرهای مورد استفاده در پژوهش برای شناسایی دستکاری بازار

متغیرها	نحوه محاسبه	دریافت اطلاعات	مقالات مرجع
حجم معاملات	V_i	Tseclient	کارول[3] (۲۰۰۹)
بازده	$R_i = Ln\ (p_i/p_{i-1})$	Tseclient	دیاز و همکاران (۲۰۱۱)
بازده صنعت	$R_{industry} = 1/n \sum r_i$	Tseclient	دیاز و همکاران (۲۰۱۱)
حجم صنعت	$V_{industry} = \sum V_i$	Tseclient	دیاز و همکاران (۲۰۱۱)
ارزش بازار	M_i	Tseclient	ژانگ، ژُنگ، دُنگ، وانگ و وانگ[4] (۲۰۱۹)
متوسط حجم خرید حقیقی	حجم خرید حقیقی به تعداد خریدار حقیقی	Tsetmc.com	نظر خبرگان
متوسط حجم خرید حقوقی	حجم خرید حقوقی به تعداد خریدار حقوقی	Tsetmc.com	نظر خبرگان
متوسط حجم فروش حقیقی	حجم فروش حقیقی به تعداد فروشنده حقیقی	Tsetmc.com	نظر خبرگان
متوسط حجم فروش حقوقی	حجم فروش حقوقی به تعداد فروشنده حقوقی	Tsetmc.com	نظر خبرگان
افزایش سرمایه از محل تجدید	برابر با یک، ۱۰ روز معاملاتی قبل و ۵ روز معاملاتی بعد از اعلام افزایش سرمایه	کدال، بورس ۲۴، بتاسهم	نظر خبرگان

با توجه به اینکه دستکاری تنها در یک روز اتفاق نمی‌افتد و دوره چند روزه و حتی چند ماهه برای دستکاری وجود دارد، در این پژوهش برای هر یک از متغیرهای ذکر شده در جدول ۵، ۱۰ روز معاملاتی قبل به‌عنوان ورودی در نظر گرفته شده است. به‌عنوان مثال در یکی از سطرهای ورودی در تاریخ مشخص، حجم معاملات همان روز و حجم معاملات ۱۰ روز قبل سهم نیز به مدل داده می‌شود. برای سایر متغیرها (به غیر از افزایش سرمایه) هم این اتفاق رخ می‌دهد.

[1] Precision
[2] Recall
[3] Carroll
[4] Zhang, Zhong, Dong, Wang & Wang

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

یافته‌های پژوهش

با توجه به مبانی نظری بررسی شده 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 ۸ مشخص است، متغیرهای مورد بررسی مانند بازدهی سهام، میانگین حجم معاملات و غیره، طی دوره دستکاری به‌وسیله متغیر مجازی آزمون می‌شود:

رابطه (۸): $\text{Indicator} = \alpha_0 + \alpha_1 * \text{Manipulated} + \varepsilon$

در رابطه فوق، متغیر مجازی در زمان وقوع دستکاری، مقدار یک و در زمان‌های دیگر مقدار صفر می‌گیرد. درصورتی‌که ضریب α_1 معنی‌دار باشد بدین معنی است که در دوره دستکاری مقدار متغیر مورد بررسی به میزان قابل‌توجهی نسبت به سایر دوره‌ها متفاوت است.

نتایج رگرسیون به‌دست‌آمده از روابط فوق در جدول شماره (۶) به شرح زیر ارائه شده است.

جدول (۶): رگرسیون خطی جهت تطبیق داده‌ها

رگرسیون خطی (رابطه ۵) $\text{Indicator} = \alpha_0 + \alpha_1 * \text{Dummy} + \varepsilon$	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	۰/۰۲۴۷	۰/۳۴۹	۰/۱۶۸**	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 ۹۹٪ معنی‌دار است به‌جز ** که در سطح ۹۵٪ معنی‌دار بوده و * که معنی‌دار نمی‌باشد.

با توجه به اطلاعات جدول (۶) درمی‌یابیم که بازده و حجم معاملات که به‌نوعی نشان‌دهنده نقدشوندگی سهام می‌باشد، در صورت وقوع دستکاری افزایش‌یافته 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 قابل‌توجه در نتایج، منفی بودن ضریب α_1 در متغیر میانگین فروش حقیقی به میزان ۰/۱۲۴۴- می‌باشد که باید دلیل آن را در اتفاقات دوره دستکاری پی گرفت. معمولاً در ابتدای فرآیند دستکاری، حقیقی‌ها دستکاری کننده اقدام به خرید سهام با حجم بالا نموده و در نیمه دوم دوره دستکاری، به‌سرعت سهام دستکاری شده 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 به بررسی تفکیک نمونه به دو دسته سهام دستکاری شده و دستکاری نشده می‌پردازیم. همان‌طور که از نتایج مشخص است تمامی متغیرها در سطح ۹۵٪ 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 داده‌های آماده‌سازی شده می‌پردازیم. رتبه‌بندی مدل‌ها بر اساس شاخص اندازه F_2 در نظر 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.

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

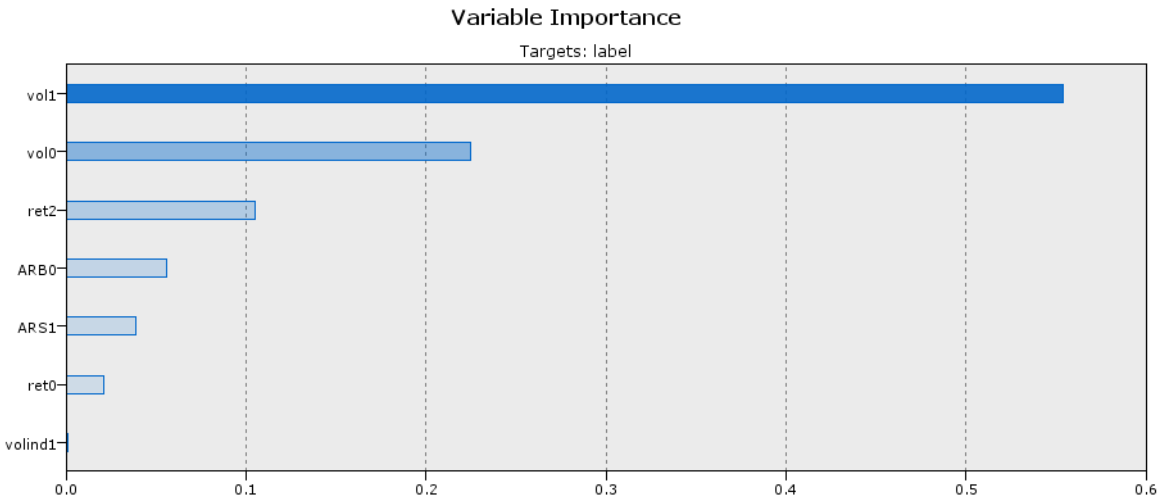
نتایج آزمون فرضیه‌های پژوهش

جدول (۷): 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	٪۴۴/۰۰	٪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

بر اساس نتایج جدول فوق درخت تصمیم در رتبه اول قرار گرفته است که از نقطه‌نظر توازن بین شاخص دقت و بازخوانی نیز وضعیت مناسبی دارد. این مورد بیان می‌کند که یکی از مؤثرین راه‌ها برای شناسایی دستکاری قیمت استفاده از قواعد از قبل تعیین شده‌ای باشد که به‌وسیله مدل‌های درخت تصمیم استخراج شده و قابلیت به‌روزرسانی در فواصل زمانی مختلف دارد. به نظر می‌رسد هم‌اکنون نیز در سطح مقدماتی و با قواعد ساده شده، شناسایی دستکاری قیمتی در بازار سرمایه ایران انجام می‌گیرد.

با استفاده از خروجی درخت تصمیم حتی می‌توان سیاست‌ها و قواعد مناسبی برای شناسایی دستکاری پیشنهاد داد. این مدل‌ها همچنین می‌توانند متغیرهای تأثیرگذار را بر اساس اهمیت طبقه‌بندی نمایند.



شکل (۳): متغیرهای تأثیرگذار بر شناسایی دستکاری سهام در بازار سرمایه بر اساس مدل درخت تصمیم

همان‌طور که ملاحظه می‌شود متغیر حجم در همان روز دستکاری (vol0) و متغیر حجم در یک روز معاملاتی قبل از دستکاری (vol1) بیشترین اهمیت را در شناسایی دستکاری سهام دارند. حتی حجم یک روز معاملاتی قبل اهمیت بالاتری دارد چرا که معمولاً اگر دو روز متوالی حجم معاملات بالا باشد، احتمال دستکاری افزایش می‌یابد.

بازدهی دو روز گذشته سهم (ret2) در رتبه بعدی اهمیت قرار می‌گیرد. ممکن است سهم دو روز مثبت باشد و متعادل شود، ولی نتایج حاضر نشان می‌دهد که اگر سهم سه روز متوالی بازدهی بالایی ثبت کرد لازم است از نقطه‌نظر دستکاری بررسی شود. میانگین خرید حقیقی در همان روز (ARB0) و میانگین فروش حقیقی در یک روز قبل (ARS1) و بازدهی در همان روز (ret0) در درجات بعدی اهمیت قرار دارند. توضیح اهمیت فروش حقیقی در یک روز قبل از دستکاری در این باشد که دستکاری‌کنندگان در روز قبل از دستکاری به‌آرامی شروع به جمع کردن سهم از بازار کرده و برخی از سهامداران بزرگ حقیقی که در جریان بازی دستکاری قرار ندارند فرصت مناسبی برای فروش سهم خود پیدا کنند. به هر ترتیب با کاوش بیشتر در درخت تصمیم، سازمان بورس و اوراق بهادار می‌تواند استراتژی‌های خاصی تبیین کرده که با احتمال مناسبی زنگ خطر قبل از دستکاری سهم به صدا درآید.

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

بحث و نتیجه‌گیری

همان‌گونه که پیشتر گفته شد موضوع دستکاری قیمت در بازارها همواره یکی از اصلی‌ترین چالش‌ها برای داشتن بازاری سالم و بدون تقلب است. بازار سرمایه نیز به‌عنوان یکی از اصلی‌ترین مصادیق بازار نیز از ابتدای ایجاد با چالش دستکاری قیمت توسط بازیگران روبه‌رو بوده است. با توجه به توسعه تکنولوژی و الگوریتم‌های معاملاتی پیچیده مبتنی بر هوش مصنوعی در معاملات سهام، لزوم بهره‌گیری از ابزارهای توسعه‌یافته مبتنی بر یادگیری ماشین برای شناسایی دستکاری قیمت‌ها در بازار سرمایه توسط نهادهای نظارتی ضروری تلقی می‌گردد. بدین منظور ابتدا به دلیل عدم اعلام سهام دستکاری شده در بازار توسط نهاد ناظر بورس نیازمند تشکیل پایگاه داده‌ای مناسب و قابل‌اتکا برای سهام دستکاری شده و دستکاری نشده بودیم که با روش‌های آماری آزمون تسلسل، چولگی، کشیدگی و آزمون تشخیص ناهنجاری متنی به‌منظور آماده‌سازی و برچسب‌گذاری داده‌ها استفاده شد. سپس بر اساس متغیرهای احصا شده برخی روش‌های یادگیری ماشین برای شناسایی دستکاری سهام، شامل درخت تصمیم، جنگل تصادفی، ماشین‌بردار پشتیبان، شبکه عصبی چندلایه و رگرسیون لجستیک را طی بازه زمانی نیمه دوم ۱۳۹۸ تا پایان ۱۴۰۲ در بازار سرمایه ایران اجرا نمودیم. برای این کار از اطلاعات ۷۳ سهم از ۱۹ صنعت پذیرش شده در بورس اوراق بهادار استفاده شده است. تعداد کل روزهای معاملاتی حدود ۷۱،۳۰۰ بوده که از این بین در ۵۳۷ روز معاملاتی دستکاری رخ داده و سایر روزها برای سهام منتخب دستکاری اتفاق نیفتاده است.

در ادامه پس از اجرای مدل‌های یادگیری ماشین پیشنهادی در پژوهش، با توجه به اینکه هزینه اشتباه در شناسایی و عدم شناسایی سهام دستکاری شده، هزینه یکسانی ندارد، بر اساس شاخص ارزیابی عملکردی F_2 به مقایسه روش‌های انتخابی پرداختیم که به ترتیب مدل‌های درخت تصمیم، جنگل تصادفی، شبکه عصبی چندلایه، ماشین‌بردار پشتیبان و رگرسیون لجستیک عملکردی بهتری داشتند. مقدار حدودی شاخص ارزیابی عملکردی F_2 مدل‌های مذکور به ترتیب عبارت‌اند از: ۶۹٪، ۵۰٪، ۴۱٪، ۴۰٪ و ۲۶٪. این مورد بیان می‌کند که یکی از مؤثرین راه‌ها برای شناسایی دستکاری قیمت استفاده از قواعد از قبل تعیین شده‌ای است که به‌وسیله مدل‌های درخت تصمیم استخراج شده و قابلیت به‌روزرسانی در فواصل زمانی مختلف دارد. همچنین از بررسی خروجی مدل درخت تصمیم، ترتیب اهمیت پارامترهای مبنا برای تدوین استراتژی مناسب توسط ذی‌نفعان برای مقابله با دستکاری اعلام شد. به‌طورکلی می‌توان گفت طی دوره دستکاری افزایش قیمت، پس از دوره دستکاری کاهش قیمت و طی دوره قبل و بعد از دستکاری افزایش نوسان قیمتی در سهم اتفاق افتاده است. ذکر این نکته ضروری است که سایر مدل‌های یادگیری ماشین استفاده شده در این پژوهش نیز عملکرد مناسبی به‌خصوص در شاخص ارزیابی دقت داشته اما متأسفانه از نظر شاخص عملکردی بازخوانی که مهم‌تر می‌باشد، عملکرد ضعیفی داشته‌اند.

محدودیت‌های پژوهش

مشابه بسیاری از پژوهش‌های انجام شده، این پژوهش نیز دارای محدودیت‌هایی می‌باشد که در ادامه ذکر می‌گردد.

- مهم‌ترین محدودیت، دسته‌بندی سهام به دو نوع دستکاری شده و نشده به صورت دستی می‌باشد. معمولاً راه دقیق‌تر این است که سهام دستکاری شده توسط مراجع قانونی و نهاد ناظر افشا شود که در ایران این مهم صورت نمی‌پذیرد؛ هرچند همان‌طور که قبلاً اشاره شد، حتی در کشورهای پیشرفته نیز، بسیاری از دستکاری‌ها توسط نهاد ناظر شناسایی نمی‌شود و اطمینان کامل نسبت به داده‌های دستکاری شده 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 به‌شدت دشوار کرده است چرا که وجود محدودیت در داده‌ها، بعضاً عملکرد مدل‌های یادگیری ماشین را تحت تأثیر قرار می‌دهد.

پیشنهادات

- 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. شرکت‌های پردازش اطلاعات

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

مالی و مشاوره سرمایه‌گذاری تحت نظارت سازمان بورس و اوراق بهادار می‌توانند با بهره‌گیری از نتایج مدل ارائه شده و در راستای قوانین موجود، با استفاده از نرم‌افزار و اپلیکیشن نسبت به اطلاع‌رسانی لازم اقدام نمایند.

- نهاد ناظر بازار می‌توانند با بهره‌برداری از خروجی درخت تصمیم ارائه شده در نتایج این پژوهش، به تدوین استراتژی شناسایی دستکاری سهام با استفاده از چندین قاعده ساده اقدام نمایند؛ که باعث افزایش سرعت واکنش و کاهش ضریب خطا خواهد شد.
- در مدل فعلی برچسب‌گذاری داده‌های دستکاری شده، از روش‌های آماری و پارامتراهای بنیادی و تکنیکال استفاده شده است. در پژوهش‌های می‌توان با رعایت پیچیدگی مدل از سایر اندیکاتورهای تکنیکال نیز بهره گرفت.

منابع

فلاح شمس، میرفیض؛ و حمیدرضا کردلوئی. (۱۳۹۰). آزمون مدل‌های لاجیت و شبکه عصبی مصنوعی جهت پیش‌بینی دستکاری قیمت در بورس اوراق بهادار تهران. **مهندسی مالی و مدیریت اوراق بهادار** ۲(۷): ۶۹-۳۷.

نورعلی دخت، حمید؛ رضا تهرانی و سعید فلاح‌پور. (۱۴۰۲). الگوریتم تشخیص معاملات مشکوک در بورس اوراق بهادار تهران بر اساس مدل معاملات جعلی. **مجله تحقیقات مالی** ۱(۲۵): ۶۲-۲۷.

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

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

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

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

Close, L., & R. Kashef. (2020). Combining Artificial Immune System and Clustering Analysis: A Stock Market Anomaly Detection Model. **Journal of Intelligent Learning Systems and Applications** 12(04): 83-108. DOI:10.4236/jilsa.2020.124005.

Comerton-Forde, C.,Putniņš, T.J., (2011). Measuring closing price manipulation. **Journal of Financial Intermediation**. 20(2): p. 135-158. Measuring closing price manipulation

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

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

Durston, G.J., & A. McKeon. (2020). **The Little Book of Market Manipulation: An Essential Guide to the Law**. Publisher: Waterside Press.

Ergün, H.O., A. Yalaman, V. Manahov & H. Zhang. (2020). Stock market manipulation in an emerging market of Turkey: how do market participants select stocks for manipulation? **Applied Economics Letters** 28(5): 354-358.

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

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

Golmohammadi, K., O.R. Zaiane. (2015). 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.

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]

Hart, O.D. (1977). On the profitability of speculation. **The Quarterly Journal of Economics** 91: 579–597.

Jarrow, R.A. (1992). Market manipulation, bubbles, corners, and short squeezes. **Journal of Financial and Quantitative Analysis** 27(03): 311-336.

Leangarun, T., & P. Tangamchit. (2018). Stock Price Manipulation Detection using Generative Adversarial Networks. **Conference: 2018 IEEE Symposium Series on Computational Intelligence (SSCI)** DOI:10.1109/SSCI.2018.8628777.

Pankajashan, S., G. Maragatham & T. Kirthiga Devi. (2021). Hybrid approach with Deep Auto-Encoder and optimized LSTM based Deep Learning approach to detect anomaly in cloud logs. **Journal of Intelligent & Fuzzy Systems** 42(2): 1-15. DOI:10.3233/JIFS-201707.

Prakash, A.,l'agriculture, O.d.N.U.p.l.a.e., (2011). Safeguarding food security in volatile global markets. 2011: **Food and Agriculture Organization of the United Nations Rome.** Safeguarding food security in volatile global markets

Quinn, P., M. Toman, & K. Curran. (2023). Identification of stock market manipulation using a hybrid ensemble approach. **Applied Research and Smart Technology (ARSTech)** 4(2): 53-63. DOI:10.23917/arstech.v4i2.2576.

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

Zhang, K., G. Zhong, J. Dong, Sh. Wang & Y. Wang. (2019). Stock market prediction based on generative adversarial network. **Procedia Computer Science** 147: 400-406.

[Downloaded from qfaj.mobarakeh.iau.ir on 2025-08-15]