Securities & Exchange Organization, Research, Development & Islamic Studies (RDIS)
Journal of Securities and Exchange, Spring 2025, V. 18, No.69, pp. 25-52

Stock Price Pumping and Dumping Manipulation Detection Using Deep Learning Methods in Tehran Stock Exchange[1]

Meysam Amiri[2], Moslem Peymani[3], Mohammad Ali Dehghan Dehnavi[4], Reza Jahanbin[5]

Received: 2024/09/23
Accepted: 2025/02/18

Research Paper

Abstract

One of the most important challenges in the stock market is the manipulation of securities. Manipulation of securities not only causes problems in the functionality of the capital market in the assets pricing, but also causes damages to the investors' confidence. Therefore, the design and use of smart surveillance tools, especially tools based on artificial intelligence, to detect the market manipulation is very important. The purpose of this research is to provide a model based on deep learning methods to detect stock pump and dump manipulation. For this purpose, we generated labeled manipulation data using designing surveillance triggers at Securities and Exchange Organization and checking suspicious transactions. In the following, by using the adversarial generative network, misbalancing of data has been solved, and by using multi-layer perceptron neural network models, recurrent neural network, long-short-term memory neural network, and radial basis function network, price manipulation has been detected. The results show that long-term memory network and recurrent neural network had a good performance at price manipulation detection by more than 92% F2 score and have the ability to be implemented in regulatory processes and future research in other areas of stock manipulations.

Key Words: Market Manipulation, Pump & Dump, Deep Learning
JEL Classification: G10, G14.

---

1. doi: 10.22034/JSE.2024.12351.2219
2. Assistant Professor, Department of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran. (amiry@atu.ac.ir).
3. Associate Professor, Department of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran. (m.peymany@atu.ac.ir).
4. Assistant Professor, Department of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran. (dehghandehnavi@atu.ac.ir).
5. Ph.D. Student, Department of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran. (Corresponding Author). (re.jahanbin@gmail.com).

Journal of Securities Exchange

سازمان بورس و اوراق بهادار، مرکز پژوهش، توسعه و مطالعات اسلامی

فصلنامه بورس اوراق بهادار، سال هجدهم، شماره ۶۹، بهار ۱۴۰۴، صص ۵۲-۲۵

شناسایی دستکاری قیمتی فروش پس از القای روند در بورس اوراق بهادار تهران با استفاده از روش‌های یادگیری عمیق[۱]

میثم امیری[۲]، مسلم پیمانی[۳]، محمدعلی دهقان‌دهنوی[۴]، رضا جهان‌بین[۵]

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۰۷/۰۲

تاریخ پذیرش: ۱۴۰۳/۱۲/۳۰

چکیده

یکی از مهمترین چالش‌ها و معضلات بازار سرمایه، دستکاری اوراق بهادار است. دستکاری اوراق بهادار نه تنها قیمت‌گذاری صحیح دارایی‌ها را با مشکل همراه می‌سازد بلکه سبب ایجاد خدشه در اعتماد سرمایه‌گذاران می‌شود. بنابراین طراحی و استفاده از ابزارهای هوشمند نظارتی برای شناسایی دستکاری قیمتی از اهمیت ویژه‌ای برخوردار است. هدف از انجام این پژوهش ارائه مدلی استوار بر روش‌های یادگیری عمیق جهت شناسایی تخلف فروش پس از القای روند است. برای این منظور در ابتدای امر بر اساس هشدار نظارتی طراحی شده در سامانه نظارتی سازمان بورس و اوراق بهادار و بررسی موارد مشکوک، داده‌های برچسب‌دار تخلف تولید شده است. در ادامه با استفاده از شبکه متخاصم مولد، مشکل عدم توازن داده‌ها رفع شده و با استفاده از مدل‌های شبکه عصبی پرسپترون چند لایه، شبکه عصبی بازگشتی، شبکه عصبی حافظه طولانی-کوتاه مدت و شبکه تابع پایه شعاعی به شناسایی دستکاری قیمتی پرداخته شده است. نتایج بررسی نشان می‌دهد مدل‌های شبکه حافظه طولانی-کوتاه مدت و شبکه عصبی بازگشتی با امتیاز F2 بالای ۹۲٪ از عملکرد مناسبی در شناسایی دستکاری قیمتی برخوردار بوده و دارای قابلیت پیاده‌سازی در فرآیندهای نظارتی و پژوهش‌های آتی سایر حوزه‌های شناسایی تخلف است.

واژه‌های کلیدی: دستکاری اوراق بهادار، تخلف فروش پس از القای روند، یادگیری عمیق.

طبقه‌بندی موضوعی: G10، G14.

---

۱. doi: 10.22034/JSE.2024.12351.2219

۲. استادیار، گروه مالی و بانکداری، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. (amiry@atu.ac.ir).

۳. دانشیار، گروه مالی و بانکداری، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. (m.peymany@atu.ac.ir).

۴. استادیار، گروه مالی و بانکداری، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. (dehghandehnavi@atu.ac.ir).

۵. دانشجوی دکتری، گروه مالی و بانکداری، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. (نویسنده مسئول). (re.jahanbin@gmail.com).

مقدمه

بازار سرمایه به عنوان یکی از دو بخش اصلی بازار مالی اقتصادی کشور نقش بسزایی در توسعه فعالیت‌های اقتصادی، تامین مالی و انتقال سرمایه به بخش‌های مولد و کارا دارد. از جمله مهمترین کارکردهای اصلی این بازار می‌توان به ایجاد زیرساخت برای تأمین مالی بنگاه‌ها و واحدهای اقتصادی، افزایش نقدشوندگی دارایی‌ها، امنیت معاملات، کاهش هزینه‌ها، ایجاد شفافیت اطلاعاتی، برطرف نمودن نیازهای بنگاه‌ها به ابزارهای مالی جدید و ... اشاره کرد. بنابراین توسعه و تعمیق این بازار لازمه رشد و توسعه اقتصادی کشور است.

لازمه و پیش‌نیاز توسعه و تعمیق این بازار، ایجاد بستری امن و مطمئن برای مشارکت سرمایه‌گذاران و عموم مردم است. بر اساس ماده ۲ قانون بازار اوراق بهادار جمهوری اسلامی ایران، حمایت از حقوق سرمایه‌گذاران، ساماندهی، حفظ و توسعه بازار شفاف، منصفانه و کارای اوراق بهادار و به منظور نظارت بر حسن اجرای قانون از جمله اهداف و ماموریت‌های تأسیس سازمان بورس اوراق بهادار است. بنابراین در صورتی که این مهم اجرایی نشود، اعتماد عمومی به شفافیت و کارایی بازار از بین رفته و زیان‌های جبران ناپذیری به سرمایه‌گذاران و مشارکت‌کنندگان در این بازار و در نهایت به اقتصاد کشور، تحمیل خواهد شد.

نتایج مطالعات نشان می‌دهد، اگر چه در اوایل قرن بیستم دستکاری قیمتی در بازار سرمایه کشورهای توسعه یافته یکی از بزرگترین مشکلات نهادهای نظارتی بوده است اما این کشورها از طریق وضع قوانین محدودکننده و توسعه سامانه‌های نظارتی و کشف دستکاری قیمتی سعی در کنترل این پدیده کرده‌اند. با این حال با گسترش و بلوغ بازار سرمایه روش‌های دستکاری قیمتی نیز پیچیده‌تر شده است به‌طوری‌که در حال حاضر نیز بخش بزرگی از تخلفات قابل شناسایی نبوده و تنها از طریق شکایت‌ها یا اطلاع‌رسانی مردمی (به اصطلاح سوت‌زنی) قابل شناسایی است. متاسفانه در کشورهای در حال توسعه، به دلایل مسائلی نظیر نبود سیستم‌های نظارتی قوی، افزایش حجم معاملات و تراکنش‌ها و افزایش ابزارهای مالی، شناسایی دستکاری قیمتی با مشکلات بیشتری همراه است (مارخام[1]، ۲۰۱۵).

---

1. Markham

اگرچه نهادهای نظارتی بازار سرمایه همواره سعی می‌کنند تا روش‌ها و مدل‌های دستکاری قیمتی را شناسایی و اقدام به محدود کردن آن کنند، با این حال شناسایی دستکاری‌های قیمتی با دشواری همراه است. همچنین به سبب گسترش و توسعه بازار سرمایه، افزایش تعداد شرکت‌های پذیرش شده و افزایش حجم معاملات و تراکنش‌ها، نظارت دستکاری قیمتی و اقدامات مجرمانه توسط نیروی انسانی محدود نهادهای نظارتی کمابیش ناممکن است. از این رو نهادهای نظارتی به برنامه‌های کامپیوتری سریعی برای تحت نظر قرار دادن معاملات و روند قیمتی سهام و طراحی هشدارهای نظارتی به منظور شناسایی دستکاری قیمتی نیازمند است (لینگارون، تنگامچیت و تاجچایاپونگ [1]، ۲۰۱۶).

اگرچه نهاد ناظر همواره در تلاش است تا از طریق شناسایی راه‌های دستکاری قیمتی اقدام به حذف آن‌ها کند، متأسفانه فرآیند شناسایی دستکاری با چالش‌های متعددی همراه است. نخستین چالش توسعه الگوریتم‌های معاملاتی جدید و پیچیده است که شناسایی و کشف دستکاری سهام را با دشواری همراه می‌سازد. همچنین با رشد و توسعه بازار سرمایه و افزایش تعداد شرکت‌های پذیرفته شده در آن و افزایش حجم تراکنش‌ها و معاملات، نظارت بر ناهنجاری‌ها و دستکاری سهام صرفاً توسط نیروی انسانی محدود و بدون بهره‌مندی از ابزارهای نرم‌افزاری جدید کمابیش غیر ممکن است. بنابراین نهاد ناظر به ابزارها و برنامه‌های کامپیوتری سریعی نیاز دارد تا بتواند ناهنجاری‌ها و تخلفات بازار سرمایه را در عین سرعت، با دقت بالایی شناسایی کند.

در این پژوهش با توجه به اهمیت شناسایی دستکاری قیمتی سهام و لزوم تشخیص صحیح و به موقع آن و اتخاذ سیاست‌های پیشگیرانه، به تشخیص دستکاری قیمتی سهام با استفاده از مدل‌های یادگیری عمیق پرداخته خواهد شد. همچنین در این پژوهش به پیروی از کائو، کولمن، بلاتریچ و مک‌جینیتی [2] (۲۰۱۴) و لینگارون، تنگامچیت و تاجچایاپونگ [3] (۲۰۱۸) تمرکز اصلی بر تخلف فروش پس از القای روند است. تخلف فروش پس از القای روند، به فعالیتی گفته می‌شود که مشتری در ابتدا اقدام به خریداری سهام کرده سپس از طریق فعالیت‌های گوناگون از جمله افزایش مظنه مکرر، سفارش‌چینی و ایجاد سفارش ترس و ... اقدام به القای روند افزایشی

1. Leangarun, Tangamchit, & Thajchayapong
2. Cao, Coleman, Belatreche, & McGinnity
3. Leangarun, Tangamchit & Thajchayapong

کرده و در ادامه پس از افزایش قیمت سهم، اقدام به فروش سهام خریداری شده خود می‌کند. بر این اساس در این مقاله در ادامه پیشینه پژوهش و خلاءهای پژوهشی ارائه می‌شود. سپس مدل‌های یادگیری عمیق[1] مورد بررسی قرار گرفته و نتایج عددی ارائه شده و در نهایت به بیان نتایج و پیشنهادها استوار بر نتایج عددی پرداخته می‌شود.

مبانی نظری و توسعه فرضیه‌ها

تا کنون مطالعات و پژوهش‌های متعدد داخلی و خارجی در حوزه شناسایی دستکاری قیمتی در بازار سرمایه صورت گرفته است اما با توجه به نبود داده‌های برچسب‌دار تخلف، بخش عمده‌ای از این پژوهش‌ها تنها به شناسایی رفتارهای غیرطبیعی در روند معاملات بسنده کرده و یا از طریق تزریق اختیاری داده نویز به مجموعه داده، به شناسایی آن اقدام کرده‌اند. همچنین در پژوهش‌های خارجی که از الگوریتم‌های نظارت شده بر روی داده‌های برچسب‌دار تخلف استفاده شده است، کمابیش تمامی پژوهش‌ها به شناسایی دستکاری قیمتی روزانه پرداخته و پژوهشی بر روی داده‌های درون روزی انجام نشده است چرا که داده‌های برچسب‌دار تخلف، در بهترین حالت تنها به روز وقوع تخلف اشاره کرده اما زمان و ساعت وقوع آن را مشخص نمی‌کند. در نتیجه در این مقالات با توجه به اطلاعات کل روز، دستکاری قیمتی پیش‌بینی می‌شود. حال آنکه ممکن است تخلف صورت گرفته به طور موقت بر متغیرهای روزانه و قیمت نماد تاثیر گذاشته و دوباره شرایط سهم به حالت عادی بازگشته باشد. درنتیجه در این حالت روش‌های هوش مصنوعی و یادگیری عمیق قادر به آموزش صحیح با استفاده از داده‌های روزانه نیستند. اما در این مقاله سعی شده است تا با بهره‌گیری از داده‌های برچسب‌دار تخلف درون‌روزی و استفاده از مدل‌های نظارت‌شده یادگیری عمیق، نقصان‌های بیان شده در بالا برطرف شده و به شناسایی دستکاری قیمتی در بازار سرمایه ایران بپردازیم. در ادامه برخی از پژوهش‌های صورت گرفته در حوزه دستکاری قیمتی مورد بررسی قرار خواهد گرفت.

اوگات، دوگانای و آکتاس[2] (۲۰۰۹) در پژوهش خود به شناسایی دستکاری قیمتی در بازار ترکیه با استفاده از شبکه عصبی مصنوعی و ماشین بردار پشتیبان[3] پرداختند. آن‌ها به دلیل عدم در اختیار داشتن داده‌های برچسب‌دار، در ابتدا مجموعه داده را دستکاری کرده سپس با اختلاف

---

1 . Deep Learning
2. Öğüt, Doğanay & Aktaş
3. Support Vector Machine

حاصل شده از میانگین بازدهی، دستکاری قیمتی را شناسایی کردند. نتایج بررسی آنها نشان می‌دهد شبکه عصبی مصنوعی و ماشین بردار عملکرد بهتری نسبت به رگرسیون لجستیک داشته است. گل‌محمدی و زایان[1] (۲۰۱۵) در مقاله خود به بررسی کاربرد هوش مصنوعی در شناسایی دستکاری قیمتی در پژوهش‌های علمی گذشته پرداخته‌اند. 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 بیان کردند. در پژوهشی دیگر گلمحمدی، زایان و دیاز[2] (۲۰۱۴) بر اساس داده‌های برچسب‌دار S&P 500 برای سال ۲۰۰۳، اقدام به شناسایی دستکاری قیمتی با استفاده از روش‌های درخت استنتاج شرطی[3]، جنگل تصادفی[4]، طبقه‌بندی بیژین[5]، شبکه عصبی مصنوعی، ماشین بردار پشتیبان و Kامین همسایه نزدیک[6] کردند. نتایج پژوهش آنها نشان می‌دهد روش طبقه‌بندی بیژین عملکرد بهتری در میان سایر روش‌ها داشته است. لینگارون و همکاران (۲۰۱۶) برای شناسایی دستکاری قیمتی داده‌های مورد بررسی را به ۲ بخش داده‌های عمومی و داده‌های غیرعمومی (در دسترس ناظر بازار) تقسیم کرده است و با استفاده از روش‌های شبکه عصبی و گاوسین دو بعدی دو تخلف فروش پس القای روند و سفارش‌چینی سنگین و حذف را مورد بررسی قرار داده است. نتایج بررسی نشان می‌دهد، تخلف فروش پس از القای روند با استفاده از داده‌های در دسترس عموم قابل شناسایی بوده اما تخلف سفارش‌چینی سنگین و حذف تنها با استفاده از داده‌های در اختیار ناظر قابل شناسایی است. لی، وو و لیو[7] (۲۰۱۷)، به شناسایی دستکاری قیمتی در ۲ بازه زمانی روزانه و درون روزی با استفاده از داده‌های برچسب دار پرداخته‌اند. نتایج بررسی آنها نشان می‌دهد روش‌های یادگیری تحت نظارت برای سری‌های زمانی روزانه عملکرد بهتری داشته اما برای داده‌های درون روزی عملکرد مناسبی ندارند. علت امر این است که داده‌های برچسب‌دار تولید شده توسط سازمان‌های نظارتی، تنها روز انجام تخلف را مشخص کرده اما ساعت و زمان انجام را تعیین نکرده است. همچنین نتایج بررسی

1. Golmohammadi & Zaiane
2. Golmohammadi, Zaiane, & Díaz
3. Conditional inference trees
4. Random Forest
5. Naïve Bayes
6. K-nearest neighborhood
7. Li, Wu & Liu

نشان می‌دهد روش‌های درخت تصمیم‌گیری و Kامین همسایه نزدیک نسبت به سایر روش‌ها عملکرد بهتری داشته‌اند. لینگارون و همکاران (۲۰۱۸) با استفاده از شبکه مولد تخاصمی[1] به منظور شناسایی رفتارهای غیر طبیعی به عنوان دستکاری قیمتی استفاده کرده است. در این پژوهش در ابتدا مدل با استفاده از داده‌های دستکاری نشده آموزش دیده و سپس در مجموعه داده آزمون از طریق تزریق داده پرت به عنوان دستکاری قیمتی، به شناسایی دستکاری پرداخته می‌شود. نتایج بررسی نشان می‌دهد مدل ارائه‌شده دقت ۶۸ درصدی در شناسایی دستکاری قیمتی فروش پس از القای روند را دارد. وانگ، ژو، هانگ و یانگ[2] (۲۰۱۹) با استفاده از شبکه عصبی بازگشتی همراه با آموزش دسته‌ای اقدام به شناسایی دستکاری قیمتی سهام در بورس چین کرده‌اند. نتایج پژوهش آنها نشان می‌دهد مدل شبکه عصبی بازگشتی در مقایسه با مدل‌های مشابه مانند شبکه عصبی مصنوعی به طور میانگین حدود ۲۹٪ عملکرد بهتری در شناسایی دستکاری دارد. کراجا، کیم و لسمن[3] (۲۰۲۰) با استفاده از تحلیل صحبت‌های مدیران شرکت در گزارش‌های سالانه مدیریتی و همچنین نسبت‌های مالی اقدام به شناسایی دستکاری قیمتی اطلاعاتی کرده‌اند. این پژوهش از مدل شبکه‌های توجه سلسله مراتبی[4] برای متن کاوی استفاده کرده است. اسریدهار و سیدارتها[5] (۲۰۲۲) به ارائه مدلی ترکیبی از شبکه عصبی حافظه طولانی- کوتاه مدت برای شناسایی دستکاری قیمت در بورس هند پرداختند. نتایج بررسی آنها نشان می‌دهد مدل طراحی شده قابلیت شناسایی ۹۶,۶٪ موارد تخلفاتی را داشته است. اوسلو و آکال[6] (۲۰۲۲) از ۶ روش با نظارت جهت شناسایی دستکاری قیمتی در بورس استفاده کرددند. روش‌های مورد استفاده آنها عبارت بودند از درخت تصمیم، رگرسیون لاجستیک، نزدیکترین همسایه، جنگل تصادفی، بیز ساده و ماشین بردار پشتیبان. بر اساس نتایج به دست آمده امتیاز F بیشتر روش‌های مورد استفاده بالاتر از ۹۰٪ بوده است. فیوره و همکاران[7] (۲۰۱۹) از شبکه متخاصم مولد[8] جهت رفع مشکل عدم توازن داده‌های تخلف در شرایط خوشه‌بندی دو-دو ای استفاده کردند. نتایج بررسی آنها نشان می‌دهد شبکه متخاصم مولد عملکرد مناسب‌تری نسبت

1. Generative Adversarial Networks
2. Wang, Xu, Huang & Yang
3. Craja, Kim & Lessmann
4. hierarchical attention networks
5. Sridhar & Mootha
6. Uslu & Akal
7. Fiore, De Santis, Perla, Zanetti & Palmieri
8. Generative Adversarial Networks

به سایر روش‌های رفع مشکل عدم توازن از جمله بیش‌نمونه‌گیری تصادفی داشته است. چاریتو، دراگیسیو و گارسز[1] (۲۰۲۱) نیز همانند فیوره و همکاران از شبکه متخاصم مولد برای رفع مشکل عدم توازن داده‌های تخلف پولشویی استفاده کردند. نتایج پژوهش آنها نیز نشان می‌دهد شبکه متخاصم مولد در مقایسه با سایر روش‌ها عملکرد بهتری داشته است.

در زمینه شناسایی دستکاری قیمتی در بازار سرمایه ایران، پژوهش‌هایی صورت گرفته است اما همانگونه که پیشتر توضیح داده شد به دلیل عدم دسترسی به داده‌های برچسب‌دار تخلف، کمابیش تمامی پژوهش‌ها تنها به شناسایی رفتار غیر نرمال روزانه بسنده کرده‌اند. در این راستا فلاح شمسی و کردلویی (۱۳۹۰)، از رگرسیون لجستیک باینری و شبکه عصبی مصنوعی جهت شناسایی دستکاری قیمتی استفاده کرده‌اند. در این مقاله به دلیل عدم دسترسی به داده‌های برچسب‌دار، شرکت‌ها با استفاده از آزمون‌های تسلسل، کشیدگی و وابستگی دیرش به دو دسته دستکاری شده و نشده تقسیم‌بندی شده و سپس با استفاده از لجستیک باینری و شبکه عصبی مصنوعی اقدام به شناسایی دستکاری قیمتی کرده‌اند. نتایج بررسی آنها نشان می‌دهد دستکاری قیمتی با شناوری، شفافیت اطلاعاتی و نقدشوندگی ارتباط مستقیم و با اندازه شرکت و p/e ارتباط معکوس دارد.در مقاله‌ای دیگر فلاح شمس و محمدی (۱۳۹۰)، با استفاده از محاسبه بازده غیر نرمال، شرکت‌ها را به دو دسته دستکاری شده و نشده تقسیم‌بندی کرده سپس با استفاده از رگرسیون باینری اقدام به شناسایی دستکاری قیمتی کرده‌اند. نتایج بررسی آنها نشان می‌دهد، شرکت‌های کوچک با شفافیت اطلاعاتی پایین و نقدشوندگی کم بیشتر در معرض دستکاری هستند و عدم شفافیت اطلاعاتی مهمترین فاکتور در شناسایی دستکاری قیمتی محسوب است. عطایی و همکاران (۱۳۹۵)، از الگوریتم ژنتیک بر مبنای شبکه عصبی مصنوعی و تابع درجه ۲ تعدیل شده برای شناسایی دستکاری قیمتی استفاده کرده‌اند. نتایج بررسی نشان می‌دهد، الگوریتم ژنتیک بر مبنای شبکه عصبی مصنوعی عملکرد بهتری در طبقه‌بندی شرکت‌ها به دو دسته دستکاری شده و نشده دارد. ربیعی و همکاران (۱۳۹۷) به بررسی تأثیر دستکاری قیمت با استفاده از ارسال سفارشات اغواکننده بر کارایی بازار پرداخته‌اند. در این پژوهش از رگرسیون داده‌های پنل برای بررسی فرضیه‌های پژوهش استفاده شده است. نتایج بررسی نشان می‌دهد، با وقوع دستکاری قیمتی، شکاف قیمتی کاهش می‌یابد. همچنین حجم بالای معاملات با دستکاری

1. Charitou, Dragicevic & Garcez

قیمتی رابطه مستقیم دارد. توحیدی و موسویان (۱۳۹۷) به بررسی و تحلیل فقهی بورس‌بازی مبتنی بر دستکاری قیمت در بازار اوراق بهادار پرداخته‌اند. آنها در پژوهش خود ضوابط مختلف فقهی شامل ممنوعیت تدلیس، غرر، غبن، ضرر، تبانی و معامله صوری را با روش‌های مختلف دستکاری تطبیق داده‌اند. نتایج بررسی آنها نشان می‌دهد اگر بورس‌بازی همراه با دستکاری قیمت‌ها به روش‌های فریبنده و گمراه‌کننده باشد، از نظر شرعی محل اشکال است. ندیری و همکاران (۱۳۹۷)، به شناسایی و بررسی شرکت‌های مستعد برای دستکاری قیمتی در بازار سرمایه ایران پرداخته‌اند. در این مقاله دستکاری قیمتی از طریق ورود سفارشات اغوا کننده مورد بررسی قرار گرفته و از روش رگرسیون پنل لاجیت برای آزمون داده‌ها استفاده شده است. نتایج به دست آمده نشان می‌دهد شرکت‌های با اندازه کوچک، شفافیت اطلاعاتی اندک، حجم معاملات بالا جهت دستکاری قیمتی مستعدتر هستند. همچنین دستکاری قیمتی با تغییرات شاخص رابطه معکوس داشته که در نتیجه آن امکان دستکاری در بازارهای رکودی بیشتر است. یکه‌زارع (۱۳۹۸) در رساله خود با عنوان دستکاری قیمتی در معاملات حق تقدم در بورس تهران، به بررسی دستکاری قیمتی زمان افزایش سرمایه و انتشار حق تقدم با استفاده از داده‌های مربوط به قیمت، دارایی و افزایش سرمایه پرداخته است. بر اساس نتایج به دست آمده، برخی از سهامداران عمده پس از افزایش سرمایه و انتشار حق تقدم اقدام به فروش سهام خود و کاهش قیمت نماد می‌کنند. پس از کاهش قیمت نماد، قیمت حق‌تقدم نماد نیز کاهش می‌یابد و در ادامه سهامدار عمده اقدام به خرید حق‌تقدم‌ها در قیمت‌های پایین‌تر می‌نماید. حامدی‌نیا و همکاران (۲۰۲۲) از شبکه متخاصم مولد برای شناسایی دستکاری قیمتی در بورس تهران استفاده کرده است. با توجه به آنکه نویسندگان به داده‌های دارای برچسب تخلف دسترسی نداشته‌اند از دو روش برای برچسب‌گذاری استفاده کرده‌اند. در روش اول با استفاده از آزمون‌های آماری تسلسل، چولگی و کشیدگی روز معاملاتی مشکوک به دستکاری برچسب‌گذاری شده‌اند. در روش دوم از طریق تزریق نویز به داده‌های اصلی اقدام به برچسب‌گذاری داده‌های تخلف نموده‌اند. نتایج بررسی آنها نشان می‌دهد شاخص امتیاز F2 برای مدل طراحی شده در حدود ۷۳٪ بوده است.

همانگونه که بررسی پژوهش‌های داخلی و خارجی نشان می‌دهد به دلیل نبود داده برچسب‌دار تخلف، اکثر مقالات بر شناسایی رفتار غیر طبیعی در روند معاملات سهام و یا شناسایی تخلف روزانه پرداخته‌اند که این امر فرآیند شناسایی تخلف و صحت نتایج

به‌دست‌آمده را به ویژه در مدل‌های یادگیری ماشینی و یادگیری عمیق با چالش همراه می‌سازد. بر این اساس در این مقاله برای نخستین بار، با استفاده از روش‌های نظارت شده یادگیری عمیق به شناسایی دستکاری قیمتی «فروش پس از القای روند» بر روی داده‌های برچسب‌دار درون روزی پرداخته شده است. لازم به بیان است بر اساس اطلاعات نویسندگان این مقاله، تا کنون از روش‌های یادگیری عمیق نظارت شده برای شناسایی دستکاری قیمتی در بورس اوراق بهادار تهران استفاده نشده است و استفاده از این الگوریتم‌ها به دلیل دقت و کارایی بالا، نقش بسزایی در شناسایی دستکاری قیمتی در بورس اوراق بهادار تهران خواهد داشت.

روش‌شناسی پژوهش

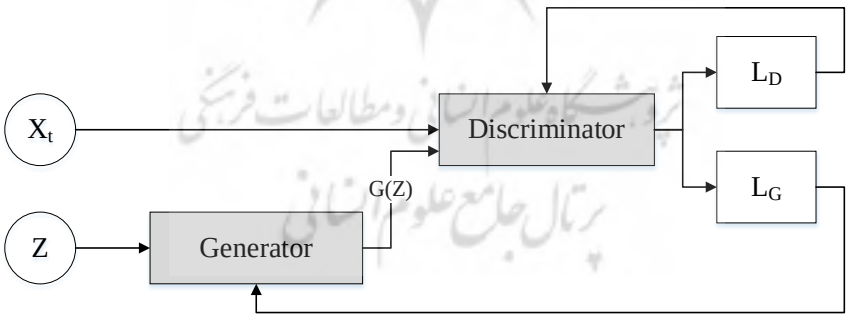
در این پژوهش از شبکه متخاصم مولد برای رفع مشکل عدم توازن داده‌ها و از ۴ روش یادگیری عمیق پرسپترون چندلایه ، شبکه عصبی بازگشتی، شبکه حافظه طولانی-کوتاه مدت و شبکه تابع پایه شعاعی به منظور شناسایی دستکاری قیمتی فروش پس از القای روند بر روی داده‌های دارای برچسب درون روزی استفاده شده است.

۱. رفع مشکل عدم توازن داده

یکی از بزرگ‌ترین چالش‌ها و عوامل لغزش در تشخیص ناهنجاری در پروژه‌های یادگیری ماشین و داده‌کاوی، حجم بالای تعداد داده‌ها و نامتوازن بودن توزیع آن‌هاست. تراکنش‌های تخلف به طور چشمگیری کم‌تر از تراکنش‌های معمولی و سالم هستند و به عبارتی، حدود ۱ تا دو درصد از تعداد کل مشاهدات هستند. روش‌های معمول ارزیابی مدل نمی‌توانند در روبرویی با داده‌های نامتوازن، عملکرد مدل را به طور دقیق اندازه‌گیری کنند. سمت‌گیری الگوریتم‌های دسته‌بندی بیشتر به سوی کلاس‌هایی است که تعداد آن‌ها در مجموعه داده بیشتر است. این الگوریتم‌ها فقط تمایل به پیش‌بینی کلاس اکثریت دارند و در روبرویی با ویژگی‌های کلاس اقلیت، با آن‌ها به‌عنوان نویز برخورد می‌کنند و بیشتر آن‌ها را نادیده می‌گیرند. به این ترتیب، احتمال دسته‌بندی اشتباه کلاس اقلیت در مقایسه با کلاس اکثریت بالاست. بنابراین در این پژوهش از یک روش ابتکاری با استفاده از شبکه متخاصم مولد برای رفع مشکل عدم توازن داده استفاده شده است (چاریتو و همکاران، ۲۰۲۱).

الف. شبکه متخاصم مولد

شبکه متخاصم مولد (GAN) یکی از مشهورترین شبکه‌های عصبی است که در سال ۲۰۱۴ توسط گودفلو، پاجت، میرزا، ژو، واردی و اوزایر[1] (۲۰۱۴) معرفی شد. شبکه متخاصم مولد از دو بخش اصلی یعنی متخاصم و مولد که خود به تنهایی شبکه عصبی هستند، تشکیل شده است. در شبکه عصبی GAN، شبکه مولد سعی می‌کند تا داده جعلی تولید کرده و شبکه متخاصم یا جداکننده سعی می‌کند داده‌های واقعی را از داده جعلی تشخیص دهد. فرآیند استفاده از شبکه عصبی GAN در رفع عدم توازن مجموعه داده به این صورت است که شبکه مولد با استفاده از داده‌های تصادفی به عنوان ورودی سعی در ایجاد داده‌های مصنوعی شبیه به داده‌های دسته اقلیت می‌کند. شبکه متخاصم نیز که ورودی آن داده‌های مربوط به دسته اقلیت است، سعی می‌کند تا داده‌های تصادفی تولید شده توسط شبکه مولد را از داده‌های اصلی دسته اقلیت تمیز دهد. در ادامه این فرآیند، هر دو شبکه متخاصم و مولد با توجه به توابع هزینه خود سعی در بهبود عملکرد خود می‌کنند به گونه‌ای که شبکه مولد سعی می‌کند داده‌هایی تولید کند که بیشتر شبیه داده‌های دسته اقلیت باشد و شبکه متخاصم نیز سعی می‌کند تا عملکرد خود را در تشخیص داده‌های جعلی از اصلی بهبود بخشد. پس از محاسبه توابع هزینه (معادله‌های ۱ و ۲) و پس‌انتشار مقادیر آن، از طریق کمینه کردن توابع هزینه، وزن‌های شبکه مولد و جداکننده بهبود می‌یابند و این امر تا زمانی ادامه می‌یابد که وزن‌های دو شبکه به حالت بهینه برسند.



شکل ۱. ساختار شبکه متخاصم مولد

1. Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley & Ozair

معادله (۱): $$L_D = l_{d1} + l_{d2} = -(log\sigma\big(D(x)\big) + \log(1 - \sigma(D\big(G(z)\big)))$$

معادله (۲): $$L_G = log\sigma(D\big(G(z)\big))$$

۲. مدل‌های شناسایی دستکاری قیمتی

الف. مدل پرسپترون چند لایه(MLP)

شبکه عصبی پرسپترون چند لایه (روزنبلات[1]، ۱۹۵۸)، دسته‌ای از شبکه عصبی‌های پیش‌خور بوده که از پشت سر هم قرار دادن چند لایه پرسپترون ایجاد می‌شود. یک شبکه عصبی پرسپترون چند لایه از حداقل ۳ لایه ورودی، خروجی و لایه پنهان تشکیل شده است. در این حالت در صورت افزایش تعداد لایه‌ها، به تعداد لایه‌های پنهان افزوده می‌شود. در این شبکه با توجه به چند لایه بودن نورون‌ها، مقدار خروجی در هر لایه از طریق مجموع موزون وزن‌ها در مقدار نورون‌های لایه قبلی محاسبه می‌شود (معادله ۳).

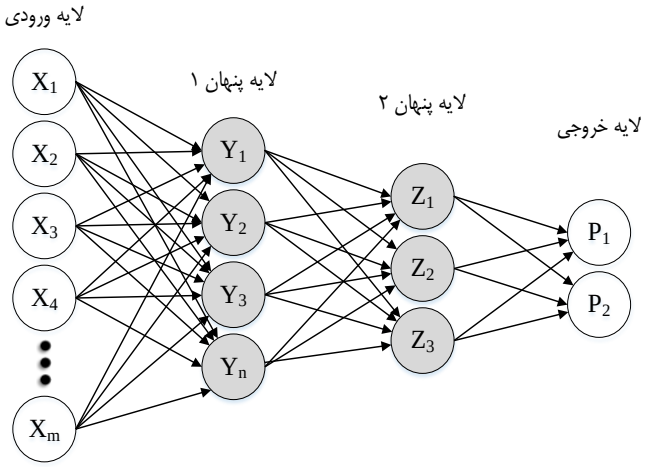
معادله (۳): $$Y_j = f(\sum_{i=1}^{m} x_i w_{ij} + w_0)$$

پس انجام محاسبات مربوط به تابع فعال‌ساز و تابع هزینه (معادله ۴)، بایستی وزن‌های شبکه از طریق الگوریتم‌های بهینه‌سازی بهبود یابند. اما در این شبکه نمی‌توان از فرآیند بهینه‌سازی استفاده شده در شبکه عصبی مصنوعی استفاده کرد چرا که برای لایه‌های میانی مقدار واقعی وجود ندارد که با استفاده از آن بتوان میزان خطا را بدست آورده و مشتق خطا نسبت به وزن آن لایه را محاسبه کرد. در این حالت از الگوریتم یادگیری پس‌انتشار خطا[2] استفاده می‌شود (معادله ۵).

معادله (۴):
$$\mathrm{e}_j(\mathrm{n}) = \mathrm{d}_j(\mathrm{n}) - \mathrm{Y}_j(\mathrm{n})$$
$$\varepsilon(n) = \frac{1}{2}\sum_j e_j^2(n)$$

معادله (۵):
$$\frac{\partial E}{\partial Y} = \frac{\partial E}{\partial P} \times \frac{\partial P}{\partial Z} \times \frac{\partial Z}{\partial Y}$$
$$\Delta w_{ji}(n) = -\eta \frac{\partial \varepsilon(\mathrm{n})}{\partial v_j(n)} y_i(n)$$

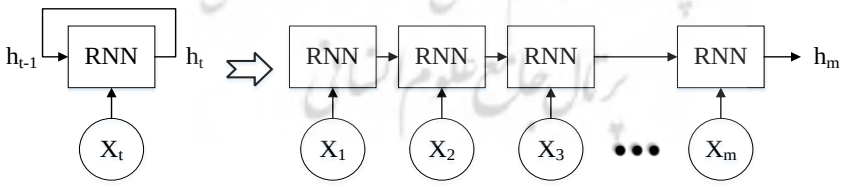
1. Rosenblatt
2. Backpropagation



شکل ۲. ساختار شبکه عصبی پرسپترون

ب. شبکه عصبی بازگشتی

مدل شبکه پرسپترون چند لایه که در بخش پیش توضیح داده شد، نمونه‌ای از مدل‌های پیش‌خور[1] هستند، یعنی جریان اطلاعات تنها در یک مسیر مستقیم از ورودی به لایه‌های پنهان و سپس به خروجی صورت می‌گیرد. اما در مدل شبکه عصبی بازگشتی (روملهارت و همکاران، ۱۹۸۶)، داده‌های خروجی مرحله قبلی به عنوان بخشی از ورودی‌های مدل در مرحله بعد در نظر گرفته می‌شود و در واقع شبکه دارای حافظه است. در این حالت ورودی تابع فعال‌ساز (معادله ۶) ترکیبی از نورون‌های ورودی و خروجی مرحله همراه با وزن‌های مرتبط با هر مجموعه است. از جمله مهمترین کاربردهای شبکه‌های عصبی بازگشتی می‌توان به پیش‌بینی سری‌های زمانی، پردازش زبان طبیعی و ترجمه ماشینی اشاره کرد.



شکل ۳. ساختار شبکه عصبی بازگشتی

1. Feed Forward

معادله (۶):
$$h_t = f(\sum_{i=1}^{m} x_i w_{ix} + h_{t-1} w_h + w_0)$$
$$Y_t = h_t w_h$$

ج. شبکه حافظه طولاتی-کوتاه مدت[1](LSTM)

شبکه عصبی بازگشتی دارای حافظه‌ای کوتاه مدت به میزان یک مرحله قبل است. همین امر سبب می‌شود تا مدل در مواردی که به خروجی چند مرحله قبل نیاز است، ضعیف عمل کند. این مشکل در شبکه حافظه طولاتی-کوتاه مدت (هوچریتر و اشمیت[2]، ۱۹۹۷) بر طرف شده است. در واقع این شبکه در طی مراحل آموزش خود، خروجی‌هایی که ارزشمند است را در حافظه خود نگاه داشته و موارد بی ارزش را از حذف می‌کند. LSTM بر خلاف شبکه عصبی بازگشتی دو ورودی و دو خروجی دارد که یکی خروجی مرحله قبل و دیگری حافظه شبکه است.

ساختار LSTM به ۴ بخش تقسیم می‌گردد. بخش اول، واحد فراموشی است. این بخش مربوط به حذف موارد بی‌ارزش از حافظه شبکه است (معادله ۷). بخش دوم، واحد ذخیره‌سازی است. در این بخش موارد جدید با ارزش در حافظه شبکه ذخیره می‌شوند (معادله ۸ و ۹). بخش سوم، بروزرسانی حافظه است. در این بخش با استفاده از بردارهای ۲ بخش قبل حافظه بروزرسانی می‌شود (معادله ۱۰). بخش چهارم، ایجاد خروجی است. در این مرحله آن بخش از حافظه که بایستی به عنوان خروجی ارسال شود، تعیین می‌شود (معادله ۱۱ و ۱۲).

معادله (۷):
$$f_t = \sigma(w_{hf} h_{t-1} + w_{if} x_t + b_{hf} + b_{if})$$

معادله (۸):
$$g_t = tanh(w_{hg} h_{t-1} + w_{ig} x_t + b_{hg} + b_{ig})$$

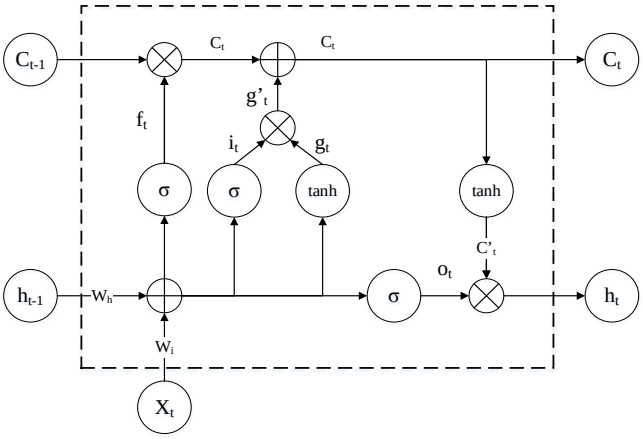
معادله (۹):
$$i_t = \sigma(w_{hi} h_{t-1} + w_{ii} x_t + b_{hi} + b_{ii})$$

معادله (۱۰):
$$C_t = \sigma(f_t \times C_{t-1} + i_t \times g_t)$$

معادله (۱۱):
$$o_t = \sigma(w_{ho} h_{t-1} + w_{io} x_t + b_{ho} + b_{io})$$

معادله (۱۲):
$$h_t = o_t \times \tanh(c_t)$$

1.Long-Short Term Memory
2. Hochreiter & Schmidhuber



شکل ۴. ساختار شبکه حافظه طولانی-کوتاه مدت

د. شبکه عصبی تابع پایه شعاعی[1](RBFN)

شبکه عصبی تابع پایه شعاعی (برومهد و لاو[2]، ۱۹۸۸) یکی از شبکه‌های عصبی قدرتمند در زمینه تخمین تابع و درون‌یابی است. منطق اصلی شبکه RBF بر پیاده‌سازی قضیه کاور استوار است. بر اساس این قضیه، یک فضای غیر خطی را می‌توان با استفاده از یک نگاشت غیر خطی، به یک فضای خطی تبدیل کرد. بدین منظور در لایه پنهان شبکه‌های RBF، از توابع فعال‌ساز غیر خطی استفاده می‌شود.

بر خلاف شبکه MLP که از چندین لایه پنهان بهره می‌برد، در این شبکه تنها یک لایه پنهان وجود دارد. همچنین در این شبکه وزن‌ها بر روی اتصالات بردارهای ورودی به لایه میانی نبوده بلکه بر بردار لایه پنهان به لایه خروجی قرار دارد. فرآیند اجرایی شبکه RBF به این صورت است که برای هر داده ورودی، فاصله برداری آن داده با بردار نمونه که برای هر ورودی مجزا است محاسبه شده و این فاصله به عنوان ورودی در اختیار تابع فعال‌ساز قرار می‌گیرد (معادله ۱۳). به عبارت دیگر تابع فعال‌ساز میزان شباهت یا نزدیکی بردار ورودی به بردار نمونه را محاسبه می‌کند. در ادامه توابع فعال‌ساز در وزن مربوطه ضرب شده به وسیله یک رابطه خطی خروجی مدل محاسبه می‌شود. به عبارت دیگر لایه خروجی یک تابع خطی از ترکیب وزنی نورون‌های لایه میانی یا همان نورون‌های RBF است. پس از محاسبه خروجی مدل، تابع هزینه

1. Radial Basis Function Network
2. Broomhead & Lowe

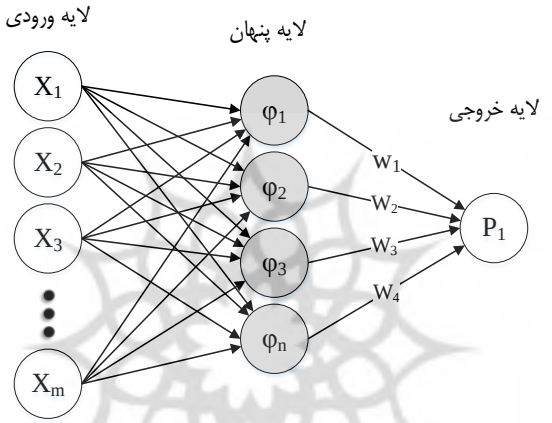
محاسبه شده و با استفاده از روابط گرادیان خطا و مشتق زنجیره‌ای وزن‌های نورون‌های لایه میانی و بردارهای نمونه بهبود می‌یابند (معادله ۱۴).

معادله (۱۳):

$$\varphi_j(x) = \exp\left(\frac{-\left|\left|x - c_j\right|\right|^2}{2\sigma_j^2}\right)$$

معادله (۱۴):

$$y_k(x) = \sum_{j=0}^{n} w_{kj}\varphi_j(x)$$



شکل ۵. ساختار شبکه عصبی تابع شعاعی

۳. داده‌ها

در این پژوهش برای نخستین بار از داده برچسب‌دار درون روزی برای شناسایی دستکاری قیمتی در بورس استفاده شده است. همین امر سبب شده است تا بتوان از روش‌های یادگیری عمیق که نیازمند داده برچسب‌دار برای آموزش هستند، در این پژوهش استفاده کرد. به منظور گردآوری داده‌های برچسب‌دار، هشدار تخلف «فروش پس از القای روند» برای نخستین بار در سامانه نظارتی سازمان بورس اوراق بهادار (سامانه بیدار)، برنامه‌نویسی شده است. در نتیجه این هشدار، موارد مشکوک به تخلف توسط سامانه نظارتی شناسایی می‌شود. پس از شناسایی و

اعلام موارد مشکوک، کارشناس مربوطه به بررسی ریز معاملات و تراکنش‌های ثبت شده پرداخته و در صورتی که تخلف اثبات شود، داده برچسب‌دار تخلف تهیه می‌شود.

در زمینه بررسی متغیرهای تاثیرگذار بر دستکاری قیمتی، تاکنون مطالعات داخلی و خارجی متعددی صورت گرفته و متغیرهای تاثیرگذار شناسایی شده‌اند. بنابراین در این پژوهش، از نتایج به دست آمده از این پژوهش‌ها برای انتخاب متغیرهای تاثیرگذار استفاده شده است. در جدول زیر متغیرهای مورد استفاده در این پژوهش نشان داده شده است.

جدول ۱. متغیرهای مدل

متغیرها	مقالات مرجع
بازده قیمت	گل‌محمدی و همکاران(۲۰۱۴)، وانگ و همکاران (۲۰۱۹) و ...
حجم معاملات	کائو و همکاران (۲۰۱۴)، لینگارون و همکاران (۲۰۱۶)، لی و همکاران (۲۰۱۷) و ...
نخستین قیمت	وانگ و همکاران (۲۰۱۹) ، لینگارون و همکاران (۲۰۱۶)، حامدی‌نیا و همکاران (۲۰۲۲) و ...
بیشترین قیمت	لینگارون و همکاران (۲۰۱۶)، گل‌محمدی و همکاران (۲۰۱۴)، شمس (۱۳۹۵) و ...
کمترین قیمت	حامدی‌نیا و همکاران (۲۰۲۲)، کائو و همکاران (۲۰۱۴)، کائو و همکاران (۲۰۱۴) و ...
آخرین قیمت	شمس و همکاران (۱۳۹۵)، لینگارون و همکاران (۲۰۱۶)، لی و همکاران (۲۰۱۷) و ...
اندازه شرکت	کلوز و همکاران (۲۰۲۰)، ندیری و همکاران (۱۳۹۷) و ...
بازده شاخص کل	ندیری و همکاران (۱۳۹۷)، شمس و همکاران (۱۳۹۵) و ...
سهام شناور آزاد	ندیری و همکاران (۱۳۹۷)، شمس و همکاران (۱۳۹۵) و ...
حجم خرید حقیقی به حقوقی	آگاراوال و همکاران (۲۰۰۶)، لئوز و همکاران (۲۰۱۷)
تعداد خریداران حقیقی	لئوز و همکاران (۲۰۱۷)، مک‌این‌تاش (۱۹۹۳) و ...
تعداد خریداران حقوقی	مک‌این‌تاش (۱۹۹۳) ، لئوز و همکاران (۲۰۱۷)و ...
تعداد فروشندگان حقیقی	لئوز و همکاران (۲۰۱۷)، مک‌این‌تاش (۱۹۹۳) و ...
تعداد فروشندگان حقوقی	مک‌این‌تاش (۱۹۹۳) ، لئوز و همکاران (۲۰۱۷)و ...

٤. سنجه‌های ارزیابی عملکرد

به منظور ارزیابی دقت مدل‌های ارائه شده، از چهار سنجه صحت[1]، حساسیت[2]، دقت[3] و امتیاز F [4] (ساساکی[5]، ۲۰۰۷) استفاده می‌شود. به‌این‌منظور در ابتدای امر بایستی ۴ پارامتر تشخیص صحیح دستکاری(TP)[6]، تشخیص کاذب دستکاری(FP)[7]، تشخیص صحیح عدم دستکاری(TN)[8]، تشخیص کاذب عدم دستکاری(FN)[9] محاسبه شد:

صحت: این نسبت توانایی مدل را در شناسایی سهام دستکاری شده نسبت به کل دستکاری‌های شناسایی شده را نشان می‌دهد.

$$\text{صحت } (Pr) = \frac{TP}{TP + FP}$$

حساسیت: این نسبت توانایی مدل را در شناسایی سهام دستکاری شده نسبت به کل دستکاری‌های اتفاق افتاده را نشان می‌دهد.

$$\text{حساسیت } (Se) = \frac{TP}{TP + FN}$$

دقت: این نسبت نشان می‌دهد که مدل تا چه میزان شناسایی صحیح داشته است.

$$\text{دقت } (\text{Ac}) = \frac{TP + TN}{TP + TN + FP + FN}$$

امتیاز F: سنجه امتیاز F، یک میانگین هارمونیک از سنجه‌های دقت و حساسیت است.

$$F_\beta = \frac{(1 + \beta^2) * \text{صحت} * \text{حساسیت}}{\left(\beta^2 * \text{صحت}\right) + \text{حساسیت}} = \frac{(1 + \beta^2) * \text{TP}}{(1 + \beta)^2 * \text{TP} + (\beta^2 * \text{FN}) + \text{FP}}$$

در این رابطه ضریب β اهمیت سنجه دقت در مقایسه با سنجه حساسیت در ارزیابی عملکرد مدل را نشان می‌دهد. اگر سنجه‌های صحت و حساسیت ارزش و اهمیت یکسانی داشته باشند،

---

1. Precision
2. Sensitivity
3. Accuracy
4. F-Score
5. Sasaki
6. True Positive
7. False Positive
8. True Negative
9. False Negative

ضریب β برابر با یک بوده و شاخص ارزیابی F1 است. اما در بازار سرمایه اشتباه در شناسایی سهام دستکاری شده و نشده، هزینه متقارن ندارند. بنابراین برای جریمه کردن عدم شناسایی دستکاری در مقایسه با شناسایی اشتباه دستکاری سهم دستکاری نشده، ضریب β، ۲ در نظر گرفته می‌شود.

یافته‌های پژوهش

۱. آموزش مدل‌های یادگیری عمیق و ارزیابی عملکرد آنها در شناسایی دستکاری قیمتی

فرآیند استفاده از مدل‌های یادگیری عمیق در شناسایی دستکاری قیمتی به اینصورت است که در ابتدا مدل با استفاده از داده‌های تاریخی و برچسب‌گذاری شده تحت آموزش قرار گرفته و سپس عملکرد آن مورد ارزیابی قرار می‌گیرد. در این راستا بایستی در ابتدای امر داده‌ها مورد پیش‌پردازش قرار گرفته و سپس جهت آموزش به مدل وارد شوند.

به منظور ورود داده به مدل‌های یادگیری عمیق بایستی ساختار داده مورد پیش‌پردازش قرار گیرند. بدین منظور هر متغیر در قالب یک بردار $n\times1$ در نظر گرفته شده که اندیس iام بردار مربوط به وضعیت آن متغیر در پنجره زمانی i می‌باشد. همچنین با توجه به اینکه مدل‌های مورد استفاده، مدل‌های یادگیری عمیق با ناظر هستند، یک بردار نیز به عنوان بردار خروجی در نظر گرفته می‌شود که در این پژوهش، متغیر خروجی، یک متغیر باینری بوده که ۱ نشان دهنده وقوع تخلف در پنجره زمانی مورد نظر و ۰ نشان‌دهنده عدم وقوع تخلف است. به منظور آنکه داده‌ها با توجه به توابع فعال‌سازی (که کران‌دار هستند) وارد ناحیه اشباع نشوند، بایستی نرمال‌سازی شوند که بدین منظور داده هر بردار را نسبت به همان بردار نرمال‌سازی شدند (معادله ۱۵).

معادله (۱۵):

$$x_i = \frac{X_i - \min(X)}{\max(X) - \min(X)}$$

پس از ساختاربندی داده‌های ورودی، بایستی فرآیند یادگیری مدل صورت گیرد. به منظور اجرای مدل، داده‌ها به ۲ دسته آموزش و تست تقسیم‌بندی شدند. از دسته اول به منظور آموزش الگوریتم‌های یادگیری عمیق استفاده شد و پس از اتمام فرآیند آموزش، عملکرد مدل با استفاده از داده‌های تست مورد بررسی قرار گرفت.

پس از اجرای مدل برای داده‌های تست، الگوریتم بایستی تنظیم می‌شد تا مشکل برازش بیش از حد و برازش کمتر از حد برطرف می‌شد. بدین منظور مولفه‌های مختلف مدل از جمله توابع فعال‌سازی و تعداد تکرار تنظیم شد تا حدی که بیشترین میزان دقت برای داده‌های تست بدست آمد و در عین حال الگوریتم دچار برازش بیش از حد نشد. منظور از برازش بیش از حد، حالتی است که مدل بیش از حد نسبت به داده‌های بخش اول، آموزش دیده و در نتیجه دقت پیش‌بینی مدل برای داده‌های بخش تست، نسبت به دقت آن برای داده‌های بخش آموزش کمتر می‌شود (جدول ۲).

جدول ۲. ساختار مدل‌های یادگیری عمیق

پرسپترون چند لایه

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 16)]	0
dense_2 (Dense)	(None, 50)	850
dropout (Dropout)	(None, 50)	0
dense_3 (Dense)	(None, 10)	510
dense_4 (Dense)	(None, 1)	11

Total params: 1,371
Trainable params: 1,371
Non-trainable params: 0

شبکه عصبی بازگشتی

Layer (type)	Output Shape	Param #
simple_rnn (SimpleRNN)	(None, 1, 16)	528
simple_rnn_1 (SimpleRNN)	(None, 30)	1410
dense (Dense)	(None, 1)	31

Total params: 1,969
Trainable params: 1,969
Non-trainable params: 0

شبکه حافظه طولانی-کوتاه مدت

Layer (type)	Output Shape	Param #
input_3 (InputLayer)	[(None, 4, 16)]	0
lstm_2 (LSTM)	(None, 100)	46800
dense_4 (Dense)	(None, 100)	10100
dense_5 (Dense)	(None, 1)	101

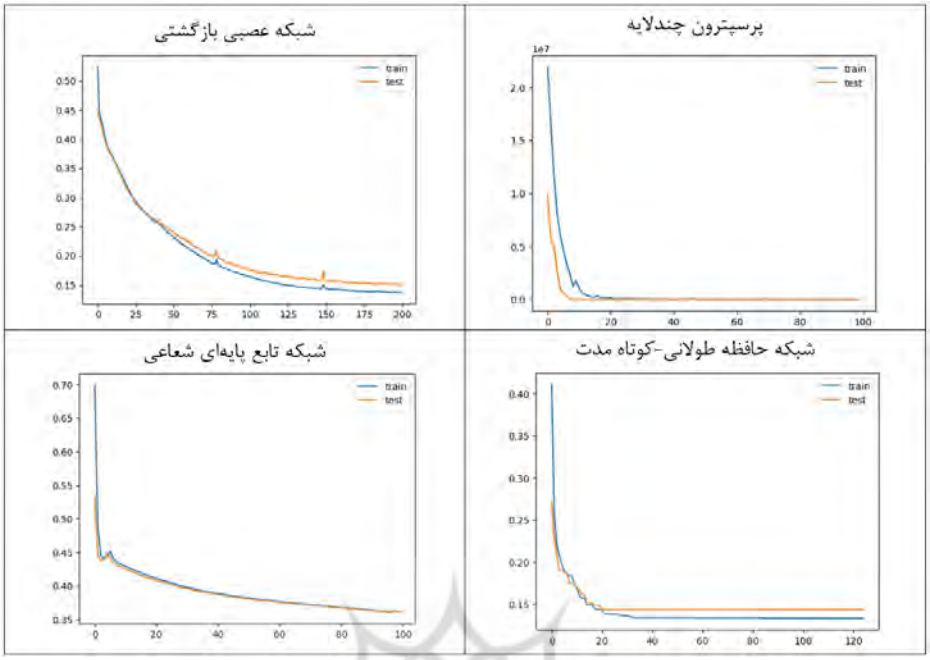
Total params: 57,001
Trainable params: 57,001
Non-trainable params: 0

شبکه تابع پایه‌ای شعاعی

Layer (type)	Output Shape	Param #
input_5 (InputLayer)	[(None, 16)]	0
dense_8 (Dense)	(None, 32)	544
dense_9 (Dense)	(None, 1)	33

Total params: 577
Trainable params: 577
Non-trainable params: 0

نمودار ۱ منحنی تابع زیان مدل‌های یادگیری عمیق متناسب با افزایش تعداد epoch را نشان می‌دهد. همانگونه که در نمودارها نیز قابل تشخیص است در هیچکدام از ۴ مدل مورد استفاده، مشکل برازش بیش از حد و برازش کمتر از حد مشاهده نمی‌شود و همگرایی تابع زیان مجموعه آموزش و مجموعه تست قابل قبول است.



نمودار ۱.منحنی‌های تابع زیان مدل‌های یادگیری عمیق (منبع: یافته‌های پژوهش)

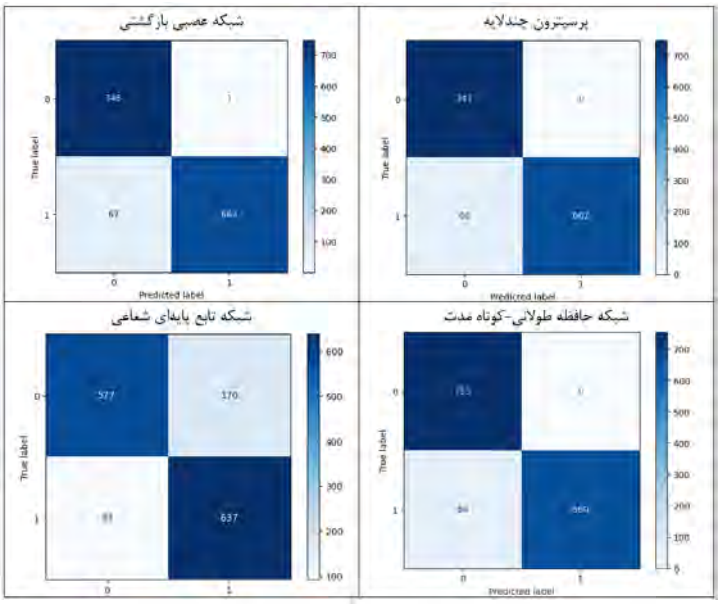
نمودار ۲ ماتریس درهم‌ریختگی مدل‌های یادگیری عمیق به کار رفته را نشان می‌دهد. ماتریس در هم‌ریختگی که از آن به عنوان ماتریس خطا نیز یاد می‌شود، اطلاعات خلاصه شده از عملکرد مدل را نشان می‌دهد. در این ماتریس تعداد پیش‌بینی‌های درست و غلط مدل بر اساس هر دسته نشان داده می‌شود. همانگونه که در نمودار زیر مشخص است، محور عمودی هر ماتریس داده‌های واقعی و محور افقی مقادیر پیش‌بینی شده را نشان می‌دهد. طبیعی است چنانچه پیش‌بینی‌های غیر متخلفانه تراکنش‌های سالم و تعداد پیش‌بینی‌های متخلفانه بودن داده‌های تخلف بیشتر باشد، عملکرد مدل در شناسایی دستکاری قیمتی بهتر بوده است. همچنین ماتریس در هم‌ریختگی این امکان را فراهم می‌آورد تا سنجه‌های ارزیابی عملکرد مدل‌ها به راحتی قابل محاسبه باشند. همانگونه که در نمودار زیر نشان داده شده است، شبکه عصبی بازگشتی ۶۶۳ مورد، پرسپترون چند لایه ۶۶۲ مورد، شبکه حافظه طولانی-کوتاه مدت ۶۶۰ مورد و شبکه تابع

پایه‌ای شعاعی ۶۳۷ مورد از تراکنش‌های متخلفانه را به درستی تشخیص داده‌اند. اما در مورد تراکنش‌های سالم شبکه حافظه کوتاه مدت ۷۵۳ مورد، پرسپترون چند لایه ۷۴۷ مورد و شبکه عصبی بازگشتی ۷۴۶ مورد و شبکه تابع پایه شعاعی ۵۷۷ مورد را به درستی، تراکنش سالم شناسایی کرده‌اند. با توجه به آنکه ترتیب مدل‌ها در تشخیص صحیح تراکنش‌های متخلفانه و سالم متفاوت است، بنابراین لازم است از سنجه‌های ارزیابی عملکرد برای شناسایی بهترین مدل استفاده کرد.

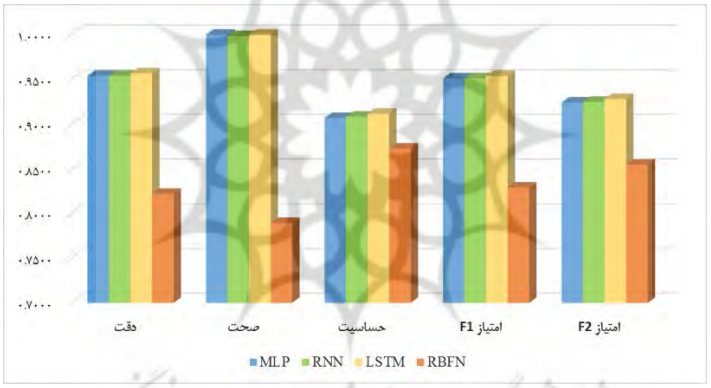
جدول۳ سنجه‌های ارزیابی عملکرد مدل‌های یادگیری عمیق را نشان می‌دهد. همانگونه که در بخش قبل توضیح داده شد، با توجه به اینکه هزینه عدم تشخیص تراکنش متخلفانه در مقایسه با تشخیص متخلفانه بودن تراکنش سالم یکسان نیست پس لازم است از سنجه‌ای استفاده کرد که در آن وزن عدم تشخیص تراکنش متخلفانه بیشتر باشد. بدین منظور در این پژوهش سنجه F2 به عنوان سنجه اصلی ارزیابی عملکرد مدل‌ها در نظر گرفته شده و در کنار آن از سنجه‌های دقت، صحت، حساسیت و امتیاز F1 نیز استفاده شده است. همانگونه که در جدول زیر نشان داده شده است، شبکه حافظه طولانی-کوتاه مدت از بیشترین امتیاز F2 برخوردار بوده و پس از آن شبکه عصبی بازگشتی، پرسپترون چند لایه و شبکه تابع پایه‌ای شعاعی قرار دارند که این مساله عملکرد بهتر شبکه حافظه طولانی-کوتاه مدت را نشان می‌دهد. نکته دیگری که از نتایج قابل برداشت است، نقش حافظه در شناسایی دستکاری است به گونه‌ای که دو مدل شبکه حافظه طولانی-کوتاه مدت و شبکه عصبی بازگشتی که از حافظه در فرآیند یادگیری خود استفاده می‌کنند، در مقایسه با دو مدل دیگر از عملکرد بالاتری برخوردار هستند.

جدول ۳. ارزیابی عملکرد مدل‌های یادگیری عمیق

مدل	دقت	صحت	حساسیت	امتیاز F1	امتیاز F2
MLP	۰/۹۵۴۰	۱	۰/۹۰۶۸	۰/۹۵۱۱	۰/۹۲۴۱
RNN	۰/۹۵۴۰	۰/۹۹۸۵	۰/۹۰۸۲	۰/۹۵۱۲	۰/۹۲۴۹
LSTM	۰/۹۵۶۷	۱	۰/۹۱۱۶	۰/۹۵۳۸	۰/۹۲۸۰
RBFN	۰/۸۲۱۹	۰/۷۸۹۳	۰/۸۷۲۶	۰/۸۲۸۹	۰/۸۵۴۶



نمودار ۲. ماتریس درهم‌ریختگی مدل‌های یادگیری عمیق (منبع: یافته‌های پژوهش)



نمودار ۳ .سنجه‌های ارزیابی عملکرد مدل‌های یادگیری عمیق (منبع: یافته‌های پژوهش)

۲. رفع مشکل عدم توازن داده

همانگونه که در بخش‌های قبل توضیح داده شد، یکی از بزرگ‌ترین چالشها در تشخیص ناهنجاری در پروژه‌های یادگیری ماشین و داده‌کاوی، حجم بالای تعداد داده‌ها و نامتوازن بودن توزیع آن‌هاست چراکه تراکنش‌های تخلف به طور چشمگیری کم‌تر از تراکنش‌های معمولی و سالم هستند و به عبارتی حدود ۱ تا دو درصد از تعداد کل مشاهده‌ها هستند. در این پژوهش

نیز تعداد کل تراکنش‌های مورد بررسی در حدود ۲۹۹۷ رکورد است که از این میزان تعداد ۳۰۵ رکورد مربوط به تراکنش‌های متخلفانه و تعداد ۲۶۹۲ رکورد مربوط به تراکنش‌های سالم و بدون تخلف است. همانگونه که از تعداد تراکنش‌ها مشخص است، تعداد تراکنش‌های متخلفانه در حدود ۱ درصد کل تراکنش‌ها است که این امر گویای غیر متوازن بودن مجموعه داده مورد بررسی است. چنانچه مدل‌های یادگیری عمیق با داده‌های فعلی آموزش ببینند، سمت‌گیری مدل‌ها به سمت تراکنش‌های سالم است. در این حالت مدل به احتمال زیاد تمامی داده‌ها را در دسته بدون تخلف طبقه‌بندی می‌کند و جالب‌تر اینکه دقت مدل نیز بسیار بالا خواهد بود چراکه تنها حدود ۱ درصد داده‌ها را اشتباه طبقه‌بندی می‌کند، حال آنکه هدف از طراحی مدل شناسایی همان ۱ درصد تراکنش‌های متخلفانه است.

به منظور رفع این مشکل، بایستی در ابتدا مشکل عدم توازن داده‌ها برطرف شده و سپس فرآیند یادگیری توسط مدل‌ها صورت گیرد. همانگونه که در بخش قبل توضیح داده شد، در این پژوهش از شبکه متخاصم مولد (GAN) برای رفع مشکل عدم توازن مجموعه داده استفاده شده است. در این پژوهش به منظور توزان بهتر داده‌ها از دو شبکه متخاصم مولد استفاده شده است به‌گونه‌ای که هدف شبکه اول تولید داده‌های مصنوعی سالم و غیر متخلفانه و هدف شبکه دوم تولید داده‌های مصنوعی متخلفانه است. بدین منظور با استفاده از شبکه اول حدود ۱۰۰۰ تراکنش سالم و غیر متخلفانه تولید می‌شود و با استفاده از شبکه دوم به میزان تفاوت تعداد تراکنش‌های سالم و متخلفانه یعنی ۳,۳۸۷ نمونه تراکنش متخلفانه مصنوعی تولید می‌شود.

به منظور بررسی عملکرد شبکه متخاصم مولد در رفع عدم توازن داده‌های ورودی، سنجه امتیاز F2 برای ۴ مدل اصلی پژوهش یعنی پرسپترون چندلایه، شبکه عصبی بازگشتی، شبکه حافظه طولانی-کوتاه مدت و شبکه تابع پایه شعاعی در سه حالت بدون توازن داده، توازن داده با بیش نمونه‌گیری تصادفی و توازن داده با شبکه متخاصم مولد، محاسبه شده است. همانگونه که در جدول ۴ نشان داده شده است، شبکه متخاصم مولد عملکرد بسیار مناسب‌تری درمقایسه با توازن داده با بیش‌نمونه‌گیری تصادفی و بدون توازن داده دارد. لازم به بیان است همانگونه که پیش‌تر توضیح داده شد در شرایطی که داده‌ها با مشکل عدم توازن همراه هستند استفاده از سنجه دقت مناسب نیست. در خصوص جزئیات مدل‌های یادگیری عمیق در ادامه به تفصیل پرداخته خواهد شد.

جدول ۴. بررسی عملکرد شبکه متخاصم مولد در رفع مشکل عدم توازن داده

مدل یادگیری عمیق	بدون توازن داده	توازن داده با بیش‌نمونه‌گیری تصادفی	توازن داده با GAN
MLP	۰/۸۱۰۳	۰/۹۱۱۳	۰/۹۵۱۱
RNN	۰/۸۰۹۷	۰/۹۱۰۳	۰/۹۵۱۲
LSTM	۰/۸۲۶۳	۰/۹۲۱۸	۰/۹۵۳۸
RBFN	۰/۷۱۹۸	۰/۷۸۱۴	۰/۸۲۸۹

بحث و نتیجه‌گیری

هدف از نگارش این مقاله شناسایی دستکاری قیمتی در بازار بورس اوراق بهادار تهران با استفاده از مدل‌های یادگیری عمیق بود. در این راستا از ۴ مدل پرسپترون چند لایه، شبکه عصبی بازگشتی، شبکه حافظه طولانی-کوتاه مدت و شبکه تابع پایه‌ای شعاعی استفاده شد. به منظور آموزش و تست عملکرد مدل‌های بیان شده در بالا از داده‌های درون روزی دارای برچسب تهیه شده در سازمان بورس استفاده شد.

به منظور ارزیابی عملکرد مدل‌های یادگیری عمیق از سنجه امتیاز F2 به عنوان سنجه اصلی و سنجه‌های دیگر مانند دقت، صحت و حساسیت استفاده شد. نتایج بررسی نشان داد مدل شبکه حافظه طولانی-کوتاه مدت در مقایسه با سایر مدلها از امتیاز F2 بالاتری برخوردار بوده و پس از آن مدل‌های شبکه عصبی بازگشتی، پرسپترون چند لایه و شبکه تابع پایه‌ای شعاعی قرار داشتند. همچنین نتایج بررسی نشان داد مدل پرسپترون چند لایه در مقایسه با مدل شبکه عصبی بازگشتی از صحت بالاتری برخوردار بوده اما در مقابل میزان حساسیت مدل شبکه عصبی بازگشتی بالاتر بوده است. این نتایج اهمیت سنجه سنجه امتیاز F را در ارزیابی عملکرد مدل‌ها نشان می‌دهد.

نکته دیگری که از نتایج به دست آمده قابل برداشت بود، نقش الگوهای بازگشتی و دارای حافظه در شناسایی تخلف فروش پس از القای روند است. همانگونه که از نتایج مشخص است مدل شبکه حافظه طولانی-کوتاه مدت و شبکه عصبی بازگشتی که هر دو از الگوی بازگشتی و حافظه در الگوریتم خود استفاده می کنند، نسبت به مدل‌های پرسپترون چند لایه و شبکه تابع پایه‌ای شعاعی عملکرد بهتری داشتند.

یافته دیگر پژوهش مربوط به رفع مشکل عدم توازن داده‌ها است. همانگونه که پیشتر توضیح داده شد، داده‌های تخلف به صورت معمول نامتوازن هستند چرا که نسبت داده‌های دارای

برچسب تخلف در مقایسه با داده‌های دارای برچسب بدون تخلف به مراتب کمتر و در حدود ۱ تا ۲ درصد است که این امر فرآیند آموزش مدل‌های یادگیری عمیق را با چالش همراه می‌سازد و مدل تمایل دارد تا تمامی داده‌ها را در دسته بدون تخلف قرار دهد که این مسئله با هدف پژوهش یعنی شناسایی موارد تخلف در تضاد است. لذا در این پژوهش از یک روش ابتکاری یعنی به کارگیری شبکه متخاصم مولد برای رفع عدم توازن داده‌ها استفاده شد. نتایج به دست آمده نشان داد عملکرد مدل‌های یادگیری عمیق در حالت رفع عدم توازن داده‌ها با شبکه متخاصم مولد نسبت به عدم رفع توازن داده‌ها و رفع توازن داده‌ها با استفاده از روش بیش‌نمونه‌گیری تصادفی به مراتب بهتر است.

مهمترین محدودیت در مسیر انجام پژوهش حاضر، تعداد داده‌های دارای برچسب تخلف بود که همانطور که پیش‌تر توضیح داده شد این محدودیت با تولید داده‌های مصنوعی تخلف توسط شبکه عصبی متخاصم مولد برطرف شد. محدودیت دیگر مربوط به بازه زمانی پژوهش است. روشن است هرچه بازه زمانی پژوهش و پیرو آن حجم داده‌های بخش آموزش بیشتر باشد، یادگیری مدل بهتر خواهد بود.

یافته‌های این پژوهش می‌تواند به عنوان ابزاری نظارتی برای سازمان بورس قابل استفاده باشد. همچنین سرمایه‌گذارانی که به دنبال سرمایه‌گذاری در نمادهای بدون دستکاری هستند می‌توانند از یافته‌های این پژوهش استفاده کنند. با توجه به عملکرد مناسب مدل‌های یادگیری عمیق به ویژه مدل‌های دارای حافظه در شناسایی تخلف فروش پس از القای روند، پیشنهاد می‌شود در پژوهش‌های آتی از این مدل‌ها برای شناسایی سایر تخلفات و دستکاری‌های بازار نیز استفاده شد. همچنین پیشنهاد می‌شود از سایر مدل‌های یادگیری عمیق در شناسایی دستکاری قیمتی استفاده شود.

References

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

Broomhead, D. S; Lowe, D. J. R. s; & Malvern, r. e. (1988). Radial basis functions, multi-variable functional interpolation and adaptive networks. C*omplex Systems*, 25(3), 1-8.

Cao, Y; Li, Y; Coleman, S; Belatreche, A; McGinnity, T. M. J. I. t. o. n. n; & systems, l. (2014). Adaptive hidden Markov model with anomaly states for price manipulation detection. *Transactions on Neural Networks and Learning Systems*, 26(2), 318-330.

Charitou, C; Dragicevic, S; & Garcez, A. d. A. J. a. p. a. (2021). Synthetic data generation for fraud detection using gans. arXiv.org.

Craja, P; Kim, A; & Lessmann, S. J. D. S. S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, 139, 113421.

Fallah, M. & Kordlouyi, H. (2011). Test of logit models and artificial neural network to predict price manipulation in Tehran Stock Exchange. *Financial Engineering and Portfolio Management*, 2(7):37-72. (In Persian).

Fallah, M. & Mohammadi, M. (2011). A Model for Price Manipulation Prediction Case Study: Tehran Stock Exchange. *Journal of The Economic Research*, 11(2):30-41. (In Persian).

Fiore, U; De Santis, A; Perla, F; Zanetti, P; & Palmieri, F. J. I. S. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, 479, 448-455.

Golmohammadi, K; & Zaiane, O. R. (2015). Time series contextual anomaly detection for detecting market manipulation in stock market. Paper presented at the 2015 IEEE international conference on data science and advanced analytics (DSAA).

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

Goodfellow, I; Pouget-Abadie, J; Mirza, M; Xu, B; Warde-Farley, D; Ozair, S; Bengio, Y. J. A. i. n. i. p. s. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Hamedinia, H; Raei, R; Bajalan, S; & Rouhani, S. (2022). Analysis of Stock Market Manipulation using Generative Adversarial Nets and Denoising Auto-Encode Models. *Advances in Mathematical Finance and Applications*. 7(1), 133-151.

Hochreiter, S; & Schmidhuber, J. J. N. c. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

Leangarun, T; Tangamchit, P; & Thajchayapong, S. (2018). Stock price manipulation detection using generative adversarial networks. Paper presented at the 2018 IEEE Symposium Series on Computational Intelligence (SSCI).

Leangarun, T; Tangamchit, P; Thajchayapong, S. J. I. j. o. t; economics, & finance. (2016). Stock price manipulation detection based on mathematical models. *International journal of trade, economics and finance*, 7(3), 81-88.

Leuz, C; Meyer, S; Muhn, M; Soltes, E; & Hackethal, A. (2017). Who falls prey to the wolf of Wall Street? Investor participation in market manipulation. Retrieved from National Bureau of Economic Research.

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

MacIntosh, J. G. J. O. H. L. (1993). The Role of Institutional and Retail Investors in Canadian Capital Markets. Osgoode Hall Law Journal, 31, 371.

Markham, J. (2015). Law enforcement and the history of financial market manipulation: Routledge.

Nadiri, M, Alavi Nasab, M, Peymani, M. (2018). Investigating Some of Effective Factors on Spoofing Manipulation in Iranian Stock Market. *Financial Research Journal*, 20(3):327-342. (In Persian).

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

Rabiee, R, Nadri, M, Peymani, M, Jaberzadh, A. (2018). Determining manipulation effect on market efficiency in Tehran Stock Exchange. *Journal of Securities and Exchange*, 11(43): 113-131. (In Persian).

Rosenblatt, F. J. P. r. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.

Rumelhart, D. E; Hinton, G. E; & Williams, R. J. J. n. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536.

Sridhar, S; Mootha, S. J. I. J. o. N; & Organisations, V. (2022). Stock market manipulation detection using feature modelling with hybrid recurrent neural networks. *International Journal of Networking and Virtual Organisations*, 26(1-2), 47-79.

Sasaki, Y. J. T. t. m. (2007). The truth of the F-measure. Teach tutor mater, 1(5), 1-5.

Shams, Sh. & Ataei, B. (2016). The detection of the stock price manipulation by hybrid genetic algorithm: Artificial Neural Network model (ANN-GA) and SQDF model. *Journal of Financial Management Strategy*, 4(3):149-171. (In Persian).

Uslu, N. C; & Akal, F. J. C. E. (2022). A machine learning approach to detection of trade-based manipulations in Borsa Istanbul. *Computational Economics*, 60(1), 25-45.

Tohidi, M. & Mosavian. A. (2018). Study and analysis of speculation based on manipulation in view point of shariah principles. *Journal of Securities and Exchange*, 11(43): 79-113. (In Persian).

Wang, Q; Xu, W; Huang, X; & Yang, K. J. N. (2019). Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning. *Neurocomputing*, 347, 46-58.

Yekehzare, M. (2019). Market Manipulation During Rights Offering: Evidence from the Tehran Stock Exchange (Master's thesis), under the guidance of Ebrahim Nezhad Ali, Sharif University of technology. (In Persian).