تشخیص ناهنجاری در بازار سهام با استفاده از تحلیل رفتاری

زهرا شعیری[۱]، سید جواد کاظمی تبار[۲*]، سروش حق وردی[۳]

دانش آموختهٔ دکترای مهندسی برق، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران[۱]
دانشیار گروه مخابرات، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران[۲*]
دانش آموختهٔ دکتری رشتهٔ مدیریت دانشگاه تهران، تهران، ایران[۳]

چکیده

یکی از روش‌های تقلب در بازار سهام، فرانت رانینگ است که در آن یک معامله‌گر، با علم (سفارشی) به یک سفارش بزرگ اقتصادی، به خریدوفروش سهام مبادرت می‌کند. در این مقاله، رویکردی برخط و بدون مربی، مبتنی بر تحلیل رفتاری، برای تشخیص ناهنجاری در دادهٔ بازار سهام پیشنهاد می‌شود که در تشخیص فرانت رانینگ موفق است. ابتدا برای هر کاربر پروفایلی حاوی ویژگی‌های رفتاری در خرید/فروش سهام ساخته می‌شود؛ سپس یک روش آماری پیشنهاد می‌شود که از آن برای محاسبهٔ ریسک هر تراکنش جدید استفاده می‌شود. این عدد ریسک به میزان تغییرات رفتار کاربر از رفتار مورد انتظار او بستگی دارد. برای تشکیل تابع ریسک از مفهوم نسبت درست‌نمایی در تئوری تشخیص استفاده می‌کنیم. احتمال شرطی برای طبیعی یا ناهنجاربودن هر تراکنش جدید محاسبه می‌شود؛ سپس ریسک را به‌صورت نسبت این دو احتمال در مقیاس لگاریتمی تعریف می‌کنیم. در محاسبهٔ ریسک از مفهوم بیز در تئوری احتمالات و درواقع قانون بیز استفاده می‌کنیم؛ همچنین فرض می‌کنیم ویژگی‌ها مستقل از یکدیگرند. در بخش شبیه‌سازی از دادهٔ خریدوفروش سهام شانزده ماه استفاده شده‌است. ویژگی‌های مربوط به مبالغ، ساعات انجام معامله، عجول‌بودن و معامله با یک معامله‌گر در خرید/فروش سهام در محاسبهٔ ریسک مورد استفاده قرار گرفته‌اند. نتایج نشان می‌دهد که روش پیشنهادی در تشخیص فرانت رانینگ موفق عمل می‌کند.

واژگان کلیدی: کشف تقلب بورس، پروفایل رفتاری، تحلیل داده، فرانت رانینگ، درست‌نمایی لگاریتمی، قانون بیز، آشکارسازی ناهنجاری.

Stock Market Anomaly Detection Using Behavioral Analysis

Zahra Shaeiri[1], Javad Kazemitabar[2]*, Soroush Haghverdi[3]
PHD, Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran[1]
Associate Professor, Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran[2]*
PHD, Faculty of Management, University of Tehran, Tehran, Iran[3]

Abstract

Stock market fraud, particularly front-running, is a deceptive practice in which traders exploit prior knowledge of significant orders placed by others to profit from stock price movements. Front-running is considered illegal because it involves using confidential or non-public information to manipulate the market for personal gain. This paper tries ti propose a novel, unsupervised, and real-time anomaly detection method based on behavioral analysis, specifically designed to identify front-running fraud within stock market transactions. The method focuses on building individual behavioral profiles for

* نویسندهٔ عهده‌دار مکاتبات
* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۹/۱۰/۲۲ • تاریخ پذیرش: ۱۴۰۳/۱۲/۱۴ • تاریخ انتشار: ۱۴۰۴/۶/۲۲ • نوع مطالعه: پژوهشی

[DOI: 10.61882/jsdp.22.2.31]

each trader, capturing their specific traits and patterns in stock buying and selling. These profiles serve as baselines for what is considered as 'normal' trading behavior for each trader.
To detect anomalies, we introduce a statistical framework where the risk of each transaction will be calculated by evaluating the deviation from the expected behavior based on the trader's historical actions. This deviation is a measure of how unusual the current transaction is in comparison to the trader's typical actions. The risk calculation involves the use of the log-likelihood ratio, a concept derived from detection theory, which compares the likelihood of a transaction being normal or fraudulent. The conditional probability of a transaction being either fraudulent or non-fraudulent is computed, and the ratio of these probabilities has been taken on a logarithmic scale to define the transaction risk. This risk metric is then utilized to flag potentially suspicious behavior for further investigations.
Bayesian probability theory underpins the model, specifically employing Bayes' rule to update the likelihood of fraud as more data will be accumulated over time. The model assumes the independence of risk components, which simplifies the complexity of the system and improves computational efficiency. Despite the potential limitation of assuming independence, empirical studies have shown that this assumption often yields reliable results for detecting anomalous behavior, making the approach both practical and effective.
Behavioral profiling plays a key role in this method. By observing the individual's trading history—such as the frequency, timing, and amounts of trades—the system learns a trader's typical behavior. This behavioral information is critical because it accounts for the natural variance in a trader's actions over time, allowing the model to distinguish between normal fluctuations and abnormal activities that might indicate fraud. Key behavioral indicators include the timing of trades, the volume of trades, the frequency of transactions with specific counterparties, and the trader's overall market engagement. Traders whose actions significantly deviate from their established patterns—such as purchasing large quantities of stocks at unusual times or interacting with the same trader excessively—are flagged as high-risk.
The simulation section of the paper uses 16 months of stock market transaction data, where features such as transaction amounts, time of trade, urgency, and consistency in trading with particular traders are analyzed to calculate the risk profile. The system ranks traders based on the risk scores of their transactions, enabling the detection of front-running activities in near real-time.
The results from the simulation indicate that the proposed method is highly effective in identifying front-running fraud. The use of behavioral profiling ensures that the system is adaptive to individual trading patterns, making it resistant to the evolving nature of fraud in financial markets. The methodology also provides a significant advantage over traditional rule-based systems, which often struggle to adapt to new fraud techniques as they emerge. Furthermore, this approach can be applied in live trading environments, making it a practical tool for regulatory bodies and market surveillance.
This paper contributes to the growing field of financial fraud detection by introducing an innovative approach that combines behavioral analysis with advanced statistical techniques. The findings underline the importance of real-time monitoring and adaptive fraud detection systems in maintaining market integrity. In the simulation section, stock market data of 16 months is used. Features related to amounts, hours, urgency, and trading with one trader in buying/selling have been used to obtain the ranking. Results show that the proposed method is effective in detecting front running cases.

Keywords: Stock market fraud detection, behavioral profiling, data analytics, Front running, log-likelihood, Bayes Rule, anomaly detection.

۱-مقدمه

در سال ۲۰۲۲ سرمایهٔ کل بازار[۱] در ایالات متحدهٔ امریکا ۴۰.۳ تریلیون دلار، در چین ۱۱.۴۷ تریلیون دلار، در کانادا ۲.۷۴ تریلیون دلار و در روسیه ۵۳۰ بیلیون دلار بوده‌است[۱]. در سال ۲۰۱۰ تنها در کانادا بیش از دویست نفر از صد شرکت، مورد پیگرد قانونی قرار گرفتند و سبب هزینه‌ای بیش از ۱۲۰ میلیون دلار شدند[۲]؛ بااین‌حال، مقدار واقعی هزینه‌های ناشی از کلاه‌برداری در اقتصاد و در بازار سهام بسیار بیشتر از این رقم است. تقلب در بازار سهام به شیوه‌های فریب‌دهنده در ارتباط با پیشنهاد خریدوفروش اوراق بهادار مربوط می‌شود. سوء استفاده از بازارهای مالی به شکل‌های متفاوتی انجام می‌شود و نتایج مخربی در یک‌پارچگی و عملکرد مناسب بازار ایجاد می‌کند. یکی از این روش‌ها دست‌کاری مبتنی بر انتشار اطلاعات است که در آن شایعه‌هایی ایجاد می‌شود که بر قیمت‌های عادلانه تأثیر می‌گذارد و تغییرات مخربی ایجاد می‌کند. مورد دیگر دست‌کاری مبتنی بر اقدام است که در آن عرضه یا تقاضای سهام را به‌نوعی منقبض می‌کنند و سبب تغییر قیمت سهام می‌شوند. یکی دیگر از روش‌های کلاه‌برداری در این زمینه روش فرانت رانینگ است که در آن یک معامله‌گر با علم (سفارشی) به یک سفارش بزرگ اقتصادی، به خریدوفروش سهام مبادرت می‌کند. فرانت رانینگ اقدامی غیرقانونی محسوب می‌شود. افراد سودجو با این‌گونه اقدامات قیمت سهام

[۱] کل سرمایهٔ بازار شرکت‌های سهامی عام

[DOI: 10.61882/jsdp.22.2.31]

را بر اساس انتظارات خود تطبیق و تغییر می‌دهند. دیده‌بانی تراکنش‌ها و تشخیص موارد پرریسک دارای اهمیت ویژه‌ای است و در عین حال با چالش‌های زیادی نیز همراه است؛ از جملهٔ این چالش‌ها تغییر مستمر روش‌های کلاه‌برداری است.

در گذشته تشخیص فرانت رانینگ با استفاده از روش‌های کل به جزء مانند مجموعه‌ای از قوانین و توسط دانش فرد خبره انجام می‌شد[۳، ۴]. یکی از ضعف‌های روش‌های مبتنی بر قانون، محدودبودن آن‌هاست. با تغییر و تحولات وسیع مانند واردشدن سرمایه‌گذاران جدید و تغییر بازار، قوانین می‌توانند کارایی خود را از دست بدهند. برای رفع این مشکل، روش‌های مبتنی بر داده‌کاوی و یادگیری ماشین پیشنهاد می‌شوند. این روش‌ها برخلاف روش‌های مبتنی بر قانون راه حل‌هایی جزء به کل را ارائه می‌دهند. این روش‌ها مدلی را از روی تاریخچهٔ داده تشکیل می‌دهند که برای تشخیص ناهنجاری مورد استفاده قرار می‌گیرد. در بیشتر موارد، دو رویکرد کلی برای تشکیل مدل داده وجود دارد؛ رویکرد با مربی و رویکرد بدون مربی. در رویکرد با مربی نمونه‌های داده دارای برچسب‌اند که نرمال یا تقلب‌بودن آن‌ها را مشخص می‌کند. در رویکرد بدون مربی برچسبی برای نمونه‌های فضای داده‌ها موجود نیست و تنها دانش موجود، ویژگی‌های نمونه‌هاست. دو چالش مهم در تشخیص تقلب، نبود برچسب داده‌ها و نامتعادل‌بودن فضای داده‌هاست؛ به‌طورمعمول داده‌های خامی که وجود دارند بدون برچسب‌اند و به‌دست‌آوردن برچسب برای نمونه‌ها نیز کاری پرهزینه و زمان‌بر است؛ از سوی دیگر نسبت نمونه‌های تقلب به نرمال بر حسب معمول عددی بسیار کوچک است؛ بنابراین حتی اگر داده دارای برچسب باشد، روش‌های با مربی راه حل‌های مناسبی نیستند. در مقالات زیادی از روش‌های با مربی برای کشف تقلب بورس استفاده شده‌است، اما به دلایلی که بیان شد، نیاز به مراحلی برای بیش یا کم‌نمونه‌برداری یا حتی استفاده از الگوریتم‌های پیچیده وجود دارد که درکل باعث پیچیده‌ترشدن روش‌ها و ارزیابی می‌شوند؛ همچنین پیشرفت فناوری اطلاعات و طبیعت در حال تغییر روش‌های تقلب، نیاز به تحول مستمر مدل را برای تشخیص نمونه‌های تقلب جدید بیشتر می‌کند. با توجه به موارد یادشده، روش‌های بدون مربی گزینهٔ مناسبی برای حل این‌گونه مسائل هستند. در پژوهش‌های انجام‌شده در این حوزه روش‌های گوناگونی در مدیریت ریسک در بازار سهام استفاده شده‌اند. روش‌های مؤثر در تجزیه‌وتحلیل و کشف تقلب همچنان در حال توسعه‌اند. در [۵-۹] مجموعه‌ای از روش‌ها در زمینهٔ مدل‌سازی و پیش‌بینی بازارهای مالی با استفاده از شیوه‌های هوش محاسباتی ارائه شده‌اند. این روش‌ها پس از بحران‌های مالی ۲۰۰۷-۲۰۰۹ مورد استقبال زیادی قرار گرفتند.

در مقالات بیشتر کارهایی که در حوزهٔ کشف تقلب مالی صورت‌گرفته در حوزهٔ کشف تقلب کارت‌های اعتباری است. تعداد کمتری از مقالات به کشف تقلب داده‌های بورس پرداخته‌اند؛ از سوی دیگر تنوع و تعدد تقلب‌های بورس بسیار زیاد است و همچنان روش‌های جدیدتری از تقلب نیز با رشد فناوری در حال شکل‌گیری هستند؛ به همین دلیل حوزهٔ آشکارسازی تقلب فرانت رانینگ نیز در تعداد خیلی کمتری از مقالات مورد توجه قرار گرفته است. روش‌های پیشنهادی در بین این مقالات نیز برخی دارای پیچیدگی بالایی بوده و برای اهداف برخط قابل استفاده نیستند.

یکی از رویکردهای نوین (بدون مربی) در تشخیص بدرفتاری، روش‌های کشف تقلب مبتنی بر تحلیل رفتاری و فناوری پروفایلینگ است. در این رویکرد برای هر کاربر پروفایلی ساخته می‌شود که حاوی رفتارها و نحوهٔ استفادهٔ او در موضوع مورد بحث است. با توجه به سامانه‌ای که کاربر در حال ارتباط با آن است، انواع مختلفی از داده‌های پروفایل در نظر گرفته می‌شوند. رُخ‌نمانگاری[1] در سامانه‌های تجارت الکترونیک، شبکه‌های اجتماعی، سامانه‌های توصیه‌گر، گوشی‌های تلفن همراه و فعالان در سهام از جمله مثال‌هایی در این زمینه‌اند. تعاملاتی که کاربر با سامانه دارد، به‌طور احتمالاتی مدل و پروفایل کاربر ساخته می‌شود. در [۱۰] برای دیوارهٔ آتش برنامه‌های تحت وب[2] از فناوری رُخ‌نمانگاری برای یادگیری رفتار مورد انتظار کاربران وب استفاده شده‌است. در سامانه حفاظت پیشرفتهٔ تهدیدِ لایت سایبر [۱۱] رفتار کاربران و دستگاه‌ها یاد گرفته‌شده و حمله‌های غیرعادی با این روش استخراج شده‌اند. شرکت HP برای اهداف کشف ناهنجاری از روش‌های رُخ‌نمانگاری استفاده می‌کند[۱۲]. سامانهٔ امنیت شبکهٔ Failsafe از پروفایلینگ برای یادگیری رفتارها استفاده می‌کند[۱۳].

در بین مقالاتی که در آن‌ها به تشخیص فرانت رانینگ مبادرت شده‌است، می‌توان [۱۴و ۱۵] را نام برد. این دو مقاله نیز رویکردی مشابه با رویکرد مقالهٔ حاضر دارند. تفاوت آن‌ها در نحوهٔ درنظرگرفتن ویژگی‌های کاربران است. در آنجا برای کشف ناهنجاری، رفتار فرد با کل جامعه مقایسه می‌شود؛ درحالی‌که در مقالهٔ حاضر رفتار فرد تنها با رفتارهای پیشین خودش مقایسه می‌شود.

در این مقاله، ناهنجاری[3] در قالب بدرفتاری فرانت رانینگ با استفاده از تحلیل رفتاری به‌طور برخط تشخیص داده می‌شود. در این رویکرد مدلی آماری از رفتار هر فرد تشکیل می‌شود. این مدل برای هر کاربر منحصربه‌فرد و مستقل از کاربران دیگر به‌دست می‌آید؛ سپس مدلی برای ریسک در نظر گرفته می‌شود و رفتارهای متعاقب کاربران مورد تجزیه‌وتحلیل

[1] Profiling
[2] Web Application Firewall
[3] Anomaly

[DOI: 10.61882/jsdp.22.2.31]

قرار می‌گیرد. برای تمام کاربران در تراکنش‌های متعاقب عدد ریسک محاسبه می‌شود. کاربرانی که رفتارهای آن‌ها از وضعیت مورد انتظار فاصلهٔ زیادی پیدا کند، ریسک بزرگ‌تری دارند. کاربران بر اساس میزان ریسک مرتب‌سازی‌شده و کاربرانی که تراکنش انجام‌گرفته توسط آن‌ها بیشترین ریسک را نشان داده‌است برای رسیدگی‌های بیشتر علامت‌گذاری می‌شوند؛ درحالی‌که نمایهٔ[1] کاربرانی که عدد ریسک پایینی دارند، با درنظرگرفتن تراکنش جدید به‌روزرسانی می‌شود. نتایج شبیه‌سازی‌ها نشان می‌دهد که روش پیشنهادی به‌خوبی تغییرات رفتاری را کمی کرده و شناسایی می‌کند.

در ادامه در بخش دوم مسئلهٔ کشف تقلب داده‌های بازار سهام تشریح می‌شود؛ سپس در بخش سوم کارهای انجام‌شده در این حوزه مورد بررسی قرار می‌گیرد؛ پس از آن در بخش چهارم روش پیشنهادی این مقاله ارائه و در انتها نیز بحث و نتیجه‌گیری انجام می‌شود.

۲- تعریف مسئله

تعریف و بیان مفهوم دست‌کاری بازار، برخلاف تبیین روش‌های دست‌کاری، اهداف و طرف‌های درگیر آن دشوار است[۱۶]؛ به همین دلیل درعمل و در بیشتر موارد، دست‌کاری را بر اساس روش‌ها و اهداف آن تعریف می‌کنند. دست‌کاری بازار «عمل آگاهانه به‌منظور تشویق دیگران به خرید سهام یا تغییر قیمت به‌صورت ساختگی یا کنترل قیمت اوراق بهادار با استفاده از معاملات ساختگی» تعریف شده‌است. درکل دست‌کاری بازار به فعالیت‌هایی گفته می‌شود که به هر طریق ممکن، کارکرد آزادانهٔ عرضه و تقاضای بازار را دچار اخلال کند و به خلق قیمت‌های ساختگی و نمایش کاذبی از فعالیت بازار سهام و درنهایت گمراه‌ساختن فعالان بازار منتهی شود[۱۷].

بی‌تردید هدف نهایی متخلفان از دست‌کاری بازار به‌دست‌آوردن سود است؛ اما برای رسیدن به این مقصود، اهداف میانی خاصی را پی‌گیری می‌کنند که تعدادی از آن‌ها به‌شرح زیر است:

- تأثیرگذاری بر قیمت اوراق بهادار یا ابزارهای مشتقه از آن، به‌منظور فراهم‌آوردن امکان خرید در قیمت‌های پایین‌تر و فروش در قیمت‌های بالاتر؛
- تأثیرگذاری بر قیمت سهامی که در قرارداد اختیار معامله وجود دارد، به‌منظور سودآور نبودن معامله برای طرف دیگر قرارداد و انصراف وی از اجرای قرارداد؛
- تأثیرگذاری بر قیمت پذیره‌نویسی عمومی یا غیرعمومی سهام؛
- تحت تأثیر قراردادن نرخ تبدیل سهام به‌منظور به‌دست‌آوردن سهم بیشتر؛
- تأثیرگذاری بر قیمت سهام به‌منظور تحت تأثیر قراردادن قیمت آن در معاملات عمده؛
- تأثیرگذاری بر قیمت یا نرخ تبدیل در ارتباط با ادغام شرکت‌ها؛
- ترغیب یا تحذیر افراد به پذیره‌نویسی، خرید یا فروش سهام و یا حق تقدم خرید آن؛
- تأثیرگذاری بر توصیه‌های مالی و سرمایه‌گذاری به‌منظور هدایت افکار عمومی [۱۸].

از دیرباز تأثیرگذاری مصنوعی بر قیمت سهام موضوع مهمی بوده‌است؛ پس از تأسیس بورسِ سهامِ آمستردام، در اوایل قرن هفدهم، کارگزاران دریافتند که می‌توانند قیمت سهام را به‌نحوی دست‌کاری کنند و سودی به‌دست آورند؛ آن‌ها در یک جریان متمرکز فروش مشارکت می‌کردند. سرمایه‌گذارانی که ترسیده بودند، شروع به فروش سهام خود می‌کردند و درنتیجه قیمت کاهش پیدا می‌کرد. پس از آن کارگزاران می‌توانستند سهام را در قیمت پایین دوباره بخرند و موقعیت قبلی را با قیمت کمتری به‌دست آورند؛ به این کار به اصطلاح حملهٔ خرس[2] گویند. کارگزاران با ساختن شایعات دروغین، روش حملهٔ خرس را محقق می‌کردند. این نوع از دست‌کاری، در بسیاری از بازارهای سهامی که در آن زمان تأسیس شده بودند، رخ داده‌است؛ برای نمونه در طول جنگ‌های ناپلئون، قیمت اوراق قرضه و سهام در بورسِ سهامِ لندن، به جریان جنگ بسیار حساس شده بود. متقلبان با استفاده از روزنامه‌ها اقدام به پخش اخبار و اطلاعات منفی در خصوص نتایج جنگ می‌کردند که قیمت سهام نیز به تبع آن کاهش پیدا می‌کرد. در بسیاری از کشورها این گونه فعالیت‌ها غیرقانونی در نظر گرفته شده‌است؛ صحبت‌های زیادی در خصوص دست‌کاری قیمت سهام شده‌است. هوبنر دست‌کاری قیمت سهام را موضوعی مهم در بازار سهام دانست[۱۹]. پس از سقوط بزرگ بازار آمریکا در ۱۹۲۹، نگرانی زیادی در مورد فراگیرشدن روش حملهٔ خرس که منجر به کاهش قیمت‌های سهام شده بود، به‌وجود آمده بود. بدین‌منظور کمیتهٔ بانک‌داری و ارز مجلس سنا پژوهش‌های گسترده‌ای در خصوص عملیات بازارهای اوراق بهادار انجام داد. این بررسی‌ها سبب شد تا علاوه‌بر آشکارسازی روش حملهٔ خرس، شواهد گسترده‌ای از سایر انواع دست‌کاری نیز بر آن‌ها نمایان شود. از این شواهد برای تنظیم مقررات، جهت شناسایی دست‌کاری‌ها استفاده شد. دست‌کاری‌های مزبور به سه طبقهٔ کلی تقسیم می‌شوند: طبقهٔ نخست، دست‌کاری مبتنی بر اقدام[3] است. این روشِ دست‌کاری با انجام اقداماتی همراه است که منجر به انحراف موقت ارزش سهام می‌شود. طبقهٔ

[1] profile

[2] Bear raid

[3] Action based manipulation

[DOI: 10.61882/jsdp.22.2.31]

دوم، دست‌کاری مبتنی بر اطلاعات[1] است. این روش دست‌کاری با استفاده از پخش اخبار، شایعات و یا اطلاعات غلط رخ می‌دهد. طبقهٔ سوم، دست‌کاری مبتنی بر معاملات است. در این روش دست‌کاری، معامله‌گر سعی می‌کند با انجام معامله و بدون انجام اقدام عمومی مشهود، قیمت را دست‌کاری کند.

قانون اوراق بهادار با انجام برخی اقدامات از قبیل ممنوع‌کردن فروش استقراضی برای اعضای هیئت مدیره و مدیران اجرایی شرکت‌ها سعی در ریشه‌کن‌کردن دست‌کاری مبتنی بر اقدام کرد. قوانینی نیز برای حذف روش مبتنی بر اطلاعات طراحی شده بود؛ شرکت‌ها می‌بایست در مقاطع زمانی مشخص، اقدام به انتشار اطلاعات به عموم می‌کردند که این اقدام پخش شایعه را سخت‌تر می‌کرد؛ همچنین قانون، اقدامات اشخاصی را که تلاش می‌کردند قیمت سهام را با استفاده از اطلاعات نادرست بالا یا پایین بیاورند، غیرقانونی اعلام کرد[۲۰].

امروزه با وجود تصویب قوانین و مقررات، دست‌کاری بازار همچنان وجود دارد. توسعهٔ فناوری‌های نوین توانمندی‌های فوق‌العاده‌ای در حوزه‌های گوناگون ایجاد کرده‌است، اما درعین‌حال روش‌های تقلب نیز با رشد فناوری دچار رشدونمو زیادی شده‌اند. در بازارهای مالی نوظهور که از سامانه‌های نظارتی کارا و اثربخش برای جلوگیری از دست‌کاری بازار اوراق بهادار بهره نمی‌برند، دست‌کاری بازار و تحت تأثیر قراردادن قیمت اوراق بهادار از سوی کسانی که از قدرت لازم برای این منظور برخوردارند، به‌طور گسترده و در همهٔ اشکال آن وجود دارد[۲۱]

در ادامه مسئله‌ای که در مقالهٔ حاضر به آن پرداخته شده‌است، توضیح داده می‌شود؛ داده‌ای از خریدوفروش سهام در مدت شانزده ماه در اختیار است. مشخصات هویتی خریدار و فروشنده رمز شده‌است و در اختیار تحلیل‌گر قرار ندارد. هدف تعیین‌شده، ارائهٔ روشی برای تشخیص به‌هنگام ناهنجاری در تراکنش‌های انجام‌گرفته توسط این کاربران است. معاملات اشخاص حقیقی و حقوقی پیش و پس از معاملهٔ بزرگ شخص حقوقی و بهره‌مندشدن از نوسان ایجادشده توسط شخص حقوقی می‌تواند فرانت رانینگ باشد و قابل رسیدگی است. خریدوفروش اشخاص حقوقی به‌طور معمول در حجم بالا اتفاق می‌افتد و بنابراین خریدوفروش با حجم کمِ این اشخاص مشکوک به دست‌کاری و تخلف است. در معاملات شخص حقوقی توجه به این نکته حائز اهمیت است که آیا طرف معامله، اشخاص متفاوتی هستند و یا یک فرد خاص است؛ برای مثال اگر شخص حقوقی نوددرصد معاملات خود را با یک شخص به‌خصوص انجام داده باشد، می‌تواند مورد بررسی قرار بگیرد. برای اطمینان بیشتر لازم است زمان ورود سفارش دو سمت معامله (خریدار و فروشنده) بررسی شود؛ اگر این دو زمان خیلی به یکدیگر نزدیک بودند (با درنظرگرفتن استثنائات) احتمال انجام معامله‌ای هماهنگ‌شده تشدید می‌شود و می‌تواند مشکوک تلقی شود. حالت مشکوک دیگر زمانی رخ می‌دهد که شخص حقوقی تمام معاملات خود را در دقایق پایانی بازار انجام دهد. شخص حقوقی می‌تواند معاملات خود را در حین بازار انجام دهد و حضور پر حجم او در دقایق پایانی بازار می‌تواند فرصت واکنش را از بازیگران آن روز نماد گرفته و دید مثبت و یا منفی نسبت به معاملات روز بعد ایجاد کند. اگر شخص حقوقی در بیشتر معاملات نماد حضور داشته باشد و یا تنها در یک روز خاص با حجم و خرید بالا وارد معاملات شده و پس از آن دیگر حضوری نداشته باشد، مشکوک بوده و می‌تواند مورد بررسی قرار گیرد. در مورد اشخاص حقیقی نیز مواردی وجود دارد که قابل پی‌گیری بیشتر است. مواردی که یادشد، ممکن است، به‌حتم نشانهٔ تقلب نباشند، اما تغییر رفتار محسوبمی‌شود.

با توجه به داده‌ها و امکاناتی که برای استخراج ویژگی‌های مؤثر از داده‌ها در اختیار داریم، در مورد اشخاص حقیقی نیز ویژگی‌هایی را استخراج می‌کنیم که بتوانند تغییر رفتاری را در خریدوفروش نشان دهند؛ برای مثال اینکه معاملات شخص به عاملیت یک کارگزاری انجام می‌شود و یا به‌صورت برخط انجام می‌شود، حائز اهمیت است. این‌که فردی که همیشه از طریق برخط مبادرت به انجام معامله می‌کرد، برای انجام معامله‌ای خاص به معامله‌گر کارگزاری رجوع کرده‌است، می‌تواند تغییر رفتار در نظر گرفته شود. زمان انجام معامله در ماه نیز مورد توجه است. حضور در سمت خرید معاملات در جلسات معاملاتی در اواسط ماه برای اشخاص حقیقی در موارد مشکوک قابل بررسی است؛ همچنین حضور شخصی که همواره در معاملات با ارزش پایین فعالیت داشته در معامله‌ای بزرگ تغییر رفتاری محسوب می‌شود؛ درصورتی‌که این تغییرات رفتاری از حدی بیشتر باشد، می‌تواند ناهنجاری در نظر گرفته شود[۲۲ و۲۳].

در شبیه‌سازی‌ها از زمان یادگیری رفتار کاربران (مرحلهٔ آموزش) احتمال خطر (ریسک) محاسبه می‌شود، اما این مقدار در زمان‌های ابتدایی قابل اتکا نیست. لازم است زمان مشخصی سپری شود تا حجم داده‌ای که در آموزش سامانه استفاده می‌شود، به حد کافی رسیده باشد. با استفاده از دادهٔ آموزش برای هر کاربر، رخ‌نمایی ساخته می‌شود که حاوی احتمالات رفتاری کاربر در خریدوفروش سهام است؛ سپس مدلی آماری برای رتبه‌بندی کاربران در تراکنش‌های جدید بر اساس ریسک آن‌ها ارائه می‌شود؛ بر اساس نتایج به‌دست‌آمده از این رتبه‌بندی در انتها با درنظرگرفتن یک مقدار آستانه که تجربی به‌دست آمده‌است، نمونه‌ها برچسب‌گذاری می‌شوند.

[1] Information based manipulation

[DOI: 10.61882/jsdp.22.2.31]

۳- کارهای انجام‌شدهٔ پیشین

در بسیاری از کشورها، نظارت و مشاهدهٔ برخط وضعیت خریدوفروش سهام مورد توجه قرار دارد، اما با توجه به مسائل امنیتی، اطلاعات کمی آزادانه در اختیار قرار دارد. روش‌های مبتنی بر کشف ناهنجاری در امور متعددی همچون تشخیص نفوذ[1]، تشخیص تقلب در سامانه‌های مخابراتی و کشف تقلب در بازار سهام مورد استفاده قرار گرفته است. سامانه‌های متعدد کشف تخلف یا سامانه‌های مشاهده-کنترل سهام به‌منظور رصد تغییرات غیرعادی قیمت سهام در راستای کشف تقلب توسعه داده شده‌اند. انجمن ملی معامله‌گران سهام در ایالات متحدهٔ آمریکا، سامانهٔ پیشرفتهٔ آشکارسازی[2] (ADS) را توسعه داده بود که برای رصد معاملات و مظنه‌ها در بازار سهام در بازه زمانی ۱۹۵۰ تا ۱۹۹۷ مورد استفاده قرار می‌گرفت[۲۴]. این سامانه حاوی دو بخش تطبیق‌دهندهٔ قواعد و تطبیق‌دهندهٔ توالی زمانی است.

هنگامی که تطبیق‌دهندهٔ توالی، ارتباطات موقتی بین رویدادهایی را که در آن‌ها پتانسیل نقض مقررات وجود دارد، تشخیص دهد، تطبیق‌دهندهٔ قواعد در آن، رفتارهای ازپیش تعریف‌شدهٔ مشکوک را (با مقایسه با مجموعه قوانین) درصورت وجود، کشف می‌کند؛ علاوه‌بر ADS، انجمن ملی معامله‌گران سهام امریکا، سامانهٔ مشاهدهٔ اوراق بهادار تحلیل اخبار و مقررات یا SONAR[3] را توسعه داده‌است. این سامانه قابلیت رصد و شناسایی معاملات دارای پتانسیل تقلب در بازار سهام را دارد[۲۵]. سامانهٔ SONAR از روش‌های متعدد هوش مصنوعی و آماری متعددی از قبیل پردازش زبان طبیعی[4]، رگرسیون آماری و منطق فازی استفاده می‌کند. به‌صورت مشابه، بازارهای مختلف سهام در کشورهای دیگر نیز سامانه‌های خود را دارند که به‌طورمعمول بنا به دلایل امنیتی به‌صورت عمومی توضیح داده نمی‌شوند.

پیش‌تر در مورد روش‌های با مربی و بدون مربی صحبت شد. از یک دیدگاه خاص تفاوت الگوریتم‌های با/بدون مربی از حیث وضعیت رخدادن رویداد متقلبانه در گذشته نیز می‌تواند بیان شود. الگوریتم با مربی وضعیت نمونه‌های برچسب‌گذاری‌شده در گذشته را به حافظه می‌سپارد؛ سپس با ورود یک نمونهٔ جدید، عدد ریسک را بر اساس دانش محفوظ خود محاسبه می‌کند. الگوریتم‌های درخت تصمیم، جنگل تصادفی و شبکه‌های عصبی مثال‌های معروفی در این زمینه‌اند؛ گرچه امروزه روش‌های با مربی توانمندی وجود دارند، اما با توجه به کمبود داده‌های دارای برچسب در این مسئله، مبادرت به آن‌ها کم است. از سوی دیگر، از آن‌جایی که این روش‌ها تنها بر اساس الگوهای شناخته‌شده در داده‌های گذشته عمل می‌کنند، ممکن است، برای کشف الگوهای جدید (تقلب‌های جدید) ناکارا باشند.

در مقابل برخی روش‌های یادگیری بدون مربی، تغییر در رفتار و کشف رویدادهای غیرطبیعی را با مقایسه با وضعیت نرمال به‌دست می‌آورند؛ لذا می‌توانند برای یافتن الگوهای جدید از تقلب مناسب باشند.

روش‌های رگرسیون ترابری، شبکهٔ عصبی، ماشین‌های بردار پشتیبان، درخت تصمیم رگرسیون، درخت تصمیم با الگوریتم C5.0، جنگل تصادفی و نزدیک‌ترین همسایگی، و تعدادی روش جدیدتر یادگیری ماشینی از جمله روش‌های مبتنی بر متن‌کاوی جهت تشخیص تقلب در بازار سهام استفاده شده‌است. در مقالهٔ دیگری نیز برای معاملات پرتواتر از شبکه‌های عصبی عمیق جهت آشکارسازی ناهنجاری استفاده شده‌است[۲۶، ۲۷، ۲۸ و ۲۹].

یکی از رویکردهای پیشرفته در تشخیص تقلب، روش‌های مبتنی بر تحلیل رفتاری هستند که به مشخصه‌های رفتاری منحصربه‌فرد کاربران وابسته‌اند[۲۲، ۲۳]. کاربران ناآگاهانه عادات و رفتارهای خود را در خریدوفروش نشان می‌دهند. راه حل‌های تحلیل رفتاری برای کشف رفتارهای طبیعی هر کاربر، محاسبهٔ ریسک هر فعالیت جدید و تولید برچسب نرمال و ناهنجاری برای نمونه‌های جدید طراحی شده‌اند. مشخصهٔ کلیدی که روش‌های تحلیل رفتاری را متمایز می‌کند، نظارت بر تمام تراکنش‌ها و نه‌تنها نظارت بر تراکنش‌های مجزا برای هر کاربر است. روش‌های تحلیل رفتاری با دسترسی به داده‌های بدون برچسب، به خودی خود تاریخچهٔ دقیقی از رفتارهای ناهنجار را دربردارند. هر چقدر یک رفتارِ ناهنجار زودتر شناسایی شود، جلوگیری از خسارت‌های ناشی از آن نیز سریع‌تر و ساده‌تر صورت خواهد گرفت. روش‌های تحلیل رفتاری قادرند پیش از این‌که خسارتی وارد شود، نخستین نشانه‌های رفتار ناهنجار را تشخیص دهند. از آنجا که این روش‌ها مبتنی بر رفتار فرد هستند، جدای از نوع تقلب می‌توانند حتی انواع جدیدتر ناهنجاری را نیز تشخیص دهند.

۴- روش پیشنهادی

اساس روش پیشنهادی این است که هر کاربر دارای رفتاری منحصربه‌فرد و یکتا در خریدوفروش است؛ درواقع رفتار یک فرد می‌تواند مانند اثر انگشت یا قرنیهٔ او به‌عنوان مشخصه‌ای یکتا، او را توصیف کند؛ البته رفتار کابران می‌تواند در طول زمان تغییراتی نیز داشته باشد، اما بر

[1] Intrusion detection
[2] Advanced Detection System (ADS)
[3] Securities Observation, News Analysis and Regulation
[4] Natural Language Processing

[DOI: 10.61882/jsdp.22.2.31]

حسب معمول این تغییرات ملایم و تفاوتی غیرعادی با رفتار طبیعی فرد ندارند. یکی از مسائل مهم، انتخاب ویژگی‌های رفتاری است، به‌طوری که بتواند به‌خوبی همهٔ جوانب را دربرگرفته و هر کاربر را به‌طور یکتا توصیف کند. با مدل‌کردن این ویژگی‌ها و مشخصه‌ها، برای هر کاربر، پروفایلی منحصربه‌فرد تشکیل می‌شود.

جهت تشکیل پروفایل رفتاری برای هر کاربر لازم است ویژگی‌ها و مشخصه‌های مؤثر و وابسته به امنیت را از رفتار کاربران که از طریق داده‌ها قابل دسترسی است، استخراج کنیم؛ برای مثال تعداد و حجم خرید در بخش‌های زمانی مختلف شبانه‌روز (هر ساعت)، حاصل ضرب قیمت در حجم سهام خریداری‌شده در بخش‌های زمانی مختلف شبانه‌روز و احتمال خرید روزانه و یا احتمال خرید در بخش‌های زمانی مختلف شبانه‌روز از جمله مثال‌هایی از این مشخصه‌هاست[۲۲، ۲۳].

در این کار از مفهوم لگاریتم نسبت درست‌نمایی[1] استفاده می‌شود. استفاده از این مفهوم در آزمون فرضیهٔ بیز معمول است. توجیه استفاده از نسبت، از یک تئوری در حوزهٔ تشخیص به نام لم نیمن-پیرسون[2] می‌آید[۳۰]. در حوزهٔ کدینگ و به‌طور خاص در تبادل پیام[3] نیز از لگاریتم نسبت درست‌نمایی برای تشخیص صفر یا یک‌بودن سیگنال دریافتی استفاده می‌شود[۳۱]؛ همچنین فرض استقلال متغیرها نیز در این روش در نظر گرفته می‌شود. با این‌که ممکن است، در ظاهر این فرض همیشه برقرار نباشد، اما نتایج فراوانی از پژوهش‌های گذشته مبنی بر این‌که این فرض ساده‌ساز نتایج قابل اتکایی تولید می‌کند، وجود دارد که می‌توان دقیق‌نبودن فرض استقلال را با آسودگی نادیده گرفت[۳۲، ۳۳و ۳۴]. در این مقاله به‌طور مشابه، موضوعی که بررسی می‌شود، ناهنجاربودن و یا طبیعی‌بودن سفارشات درخواست شده‌است. برای هر سفارش، نسبت درست‌نمایی زیر را در نظر می‌گیریم:

$$\frac{\Pr(Fraudster|Event)}{\Pr(User|Event)} = \frac{\Pr(F|E)}{\Pr(U|E)} \qquad (۱)$$

عبارت بالا با استفاده از قانون بیز به صورت زیر در می‌آید:

$$\frac{\Pr(E|F)\Pr(F)}{\Pr(E|U)\Pr(U)} \qquad (۲)$$

از آنجایی‌که $\Pr(F)/\Pr(U)$ مستقل از رخداد سفارش مزبور است آن را حذف کرده و تنها $\Pr(E|F)/\Pr(E|U)$ را مورد توجه قرار می‌دهیم؛ همچنین برای سادگی محاسباتی، لگاریتم نسبت درست‌نمایی $\log\{\Pr(E|F)/\Pr(E|U)\}$ را مورد استفاده قرار می‌دهیم. برای محاسبهٔ $\Pr(E|F)$ به سراغ تاریخچهٔ رفتار فرد در گذشته می‌رویم؛ برای مثال اگر رخداد E خریدی با مبلغ AMOUNT باشد که توسط فردی با مشخصات NATIONALCODE در ساعت ده صبح انجام گرفته لازم است بررسی کنیم که چنین رفتاری از طرف آن فرد چقدر محتمل است. جهت انجام این کار لازم است بدانیم این فرد در گذشته با چه احتمالی در ساعت دَه صبح و با چه احتمالی به میزان AMOUNT سفارش خرید ثبت کرده‌است؛ سپس از قانون بیز استفاده کرده و داریم:

$$\Pr(E|U) = \Pr(10{:}00\ AM\ |U).\Pr(AMOUNT\ |U)$$

به این ترتیب عدد ریسک برابر است با مجموع ریسک اجزای مختلف (لگاریتم حاصل‌ضرب برابر است با مجموع لگاریتم‌ها). به‌عبارت دقیق‌تر عدد ریسک زمان (ساعت خرید) با عدد ریسک مبلغ جمع می‌شود.

براساس تاریخچهٔ رفتار خریدارانِ سهام در خرید سهام و تاریخچهٔ رفتار فروشندگانِ سهام در فروش سهام برای هر کاربر رخ‌نمایی تشکیل می‌شود. در زیر ویژگی‌هایی که در مورد هر کاربر محاسبه می‌شود، آورده شده‌است:

- احتمال خرید/فروش در یکی از بازه‌های مبلغ
- احتمال خرید/فروش در ساعات مختلف
- میزان ریسک عجول‌بودن فرد
- میزان ریسک معامله با یک معامله‌گر (بیش از ۹۰ درصد معاملات با یک معامله‌گر بخصوص باشد)

مشخصه‌های نامبرده برای هر فرد به‌دست آمده و در رخ‌نمای او جمع‌آوری می‌شود. در این مقاله وزن تمام ویژگی‌ها یکسان است و همه به یک اندازه در محاسبهٔ ریسک دخالت می‌کنند. در تشکیل پروفایل‌ها، داده به‌صورت جریان داده در اختیار قرار می‌گیرد و از روش‌های پردازش جریان داده[4] برای تشکیل پروفایل‌ها استفاده می‌شود. با آمدن هر رکورد جدید از داده احتمالات محاسبه‌شده برای هر فرد دقیق‌تر شده و نتایج، قابل اتکا می‌شود؛ به‌عبارتی در ابتدا لازم است، اجازه دهیم تا زمان کافی بر سامانه بگذرد تا دانش خود را در مورد افراد غنی‌تر سازد. با گذشت زمان کافی می‌توانیم مطمئن باشیم که حجم کافی از داده در تشکیل رخ‌نماها دخالت کرده‌است. پس از گذشت زمان کافی از تشکیل پروفایل‌ها می‌توان بر صحت و دقت عدد ریسک اطمینان کرد. با رسیدن هر تراکنش جدید از داده، مقایسه‌ای بین آن و پروفایل فرد صورت می‌گیرد و بر اساس این مقایسه یک عدد ریسک به تراکنش اختصاص می‌یابد. عدد ریسک یادشده به‌صورت زیر محاسبه می‌شود:

$$Risk = \sum_i \log\left(\frac{\Pr(F)}{\Pr(U_i)}\right) \qquad (۳)$$

[1] Log Likelihood Ratio (LLR)
[2] Neyman-Perason Lemma
[3] Message Passing
[4] Stream Data Processing

[DOI: 10.61882/jsdp.22.2.31]

$\Pr(U_i)$ برابر است با احتمال ویژگی شمارهٔ i ام در خرید/فروش (برای مثال احتمال اینکه فرد با مبلغ AMOUNT خرید کند) و $\Pr(F)$ برابر با احتمال همان ویژگی زمانی که فرد متقلب اقدام به خرید/فروش کند است.

برای توضیح بیشتر، فردی را درنظر می‌گیریم که بر اساس رخ‌نمای خود خرید با مبالغ کم و در ساعات مشخصی داشته است؛ اگر این فرد اقدام به خرید با مبالغ بالا و ساعت دیگری بکند، رفتار او تغییر کرده‌است. این موضوع با رصد عدد ریسک مشخص می‌شود؛ به‌عبارتی مشاهده خواهیم کرد که چنین فردی ریسک مشخصی با نوسانات ملایمی داشته است. با تغییر رفتار یادشده، ریسک او افزایش می‌یابد. این افزایش مقدار ریسک، تغییر رفتار فرد را نسبت به گذشتهٔ او نشان می‌دهد. عدد ریسک آستانه‌گذاری و نمونه برچسب‌گذاری می‌شود (برچسب طبیعی/ناهنجاری). تعیین مقدار آستانه علاوه‌بر دانش تحلیل داده نیاز به استفاده از مشورت فرد خبرهٔ آشنا به حوزهٔ کاری مربوطه (خریدوفروش سهام) دارد. در اینجا مقدار آستانه به‌صورت زیر محاسبه می‌شود:

$$\text{Threshold} = \text{Range}(\boldsymbol{R}) \times \theta \qquad (۴)$$

در این رابطه $\boldsymbol{R}$ برداری است که در بردارندهٔ ریسک تمام نمونه‌هاست و θ عددی است که بر اساس تجربهٔ فرد خبره به‌دست آمده‌است و وابسته به داده است. هر چقدر عدد آستانه کوچکتر باشد، شرایط سخت‌گیرانه‌تری برای ناهنجاری و به‌دنبال آن تقلب در نظر گرفته شده‌است؛ بنابراین انتخاب مقدار آستانه دارای اهمیت به‌سزایی است. برای هر نمونهٔ جدید عدد ریسک محاسبه می‌شود. هر چقدر رفتار فرد مغایرت بیشتری با رفتار تاریخچه‌ای او داشته باشد ریسک رفتار فرد بیشتر می‌شود. پس از اعمال مقدار آستانه و برچسب‌گذاری نمونه‌ها، رخ‌نمای نمونه‌های طبیعی بر اساس تراکنش جدید به‌روزرسانی و در مورد نمونه‌های ناهنجار گزارشی از مشخصات آن‌ها تهیه می‌شود تا برای بررسی بیشتر مورد استفاده قرار گیرد[۲۲ و۲۳].

۵- محدودیت‌های پژوهش

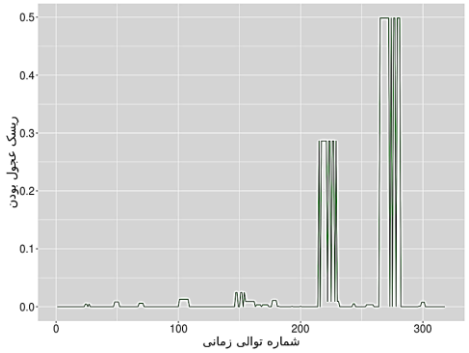
در این مقاله از قانون بیز در محاسبهٔ رتبه‌بندی ناهنجاری استفاده می‌شود. یکی از محدودیت‌های این روش اتکا بر فرض استقلال ویژگی‌ها از یکدیگر است؛ بنابراین درصورتی‌که ویژگی‌ها به حد کافی غیر وابسته به یکدیگر نباشند، نمی‌توان از این روش بهره جست؛ علاوه‌بر این، این روش ممکن است نسبت به محاسبهٔ احتمال پیشین نیز حساس باشد و لازم است، این احتمال به‌درستی و بدون خطا محاسبه شود.

۶- آزمایش‌ها و نتایج

در این مقاله داده‌های مربوط به تراکنش‌های بورس در مدت شانزده ماه مورد استفاده قرار گرفته است. این داده دارای ۵۵۲.۹۹۳ تراکنش و بیش از سی ویژگی (ستون) از جمله زمان، مبلغ، حجم و نوع معامله‌گر معامله (خرید/فروش) است. از R، پایتون و اسپارک[1] برای انجام شبیه‌سازی‌ها استفاده شده‌است. برای شروع، سهم‌های خریداری‌شدهٔ هر فرد در هر روز تجمیع می‌شود؛ به این معنی که سطرهایی که زمان تراکنش و نام خریدار یکسانی دارند با هم ادغام شده و سهام خریداری‌شدهٔ آن‌ها با هم جمع می‌شوند. ریسک نهایی که برای هر تراکنش محاسبه می‌شود، حاوی دو نوع معیار است. ریسک عجول‌بودن فرد و ریسک معامله‌کردن با یک تاجر خاص (در بیش از نوددرصد معاملات) به‌صورت دو عدد محاسبه می‌شوند. بزرگ‌بودن این عدد نشان‌دهندهٔ بالابودن میزان ریسک عجول‌بودن یا ریسک معامله با یک تاجر خاص است. زمانی که فردی در معاملات پیاپی عامل انجام معامله بوده باشد (سفارش خرید پس از سفارش فروش درج شده باشد) فردی عجول دانسته می‌شود.

معیار دیگر بر اساس احتمال رفتار فرد به‌دست می‌آید؛ برای هر تراکنش احتمال انجام تراکنش در بازه قیمت در حجم مورد توجه به‌دست می‌آید و لگاریتم عکس آن به‌عنوان ریسک لحاظ می‌شود؛ همچنین به همین شکل ریسک نوع معامله‌گر (از ایستگاه یا برخط) و ریسک زمان معامله به‌دست می‌آید. تمام معیارهای ریسک به بین صفر و یک تغییر مقیاس داده شده و جمع می‌شوند و به این ترتیب ریسک نهایی به‌دست می‌آید. در ادامه تعدادی از خریداران که در طول زمانِ معاملاتِ خود ریسک بالا داشته‌اند، مورد بررسی قرار گرفته‌اند.

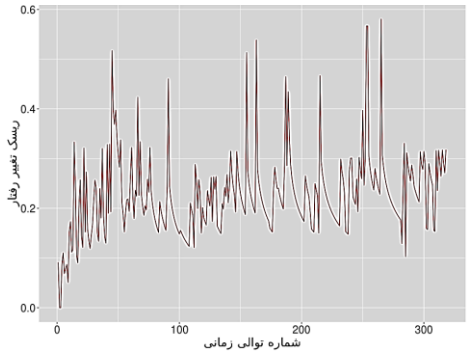
در شکل (۱) ریسک عجول‌بودن، ریسک تغییر رفتار (تغییر مبلغ در حجم و نوع معامله‌گر خرید) و ریسک نهایی برای یکی از ده مورد با ریسک بالا نشان داده شده‌است؛ همان‌طور که شکل (۱-الف) نشان می‌دهد میزان عجول‌بودن فرد در دو مقطع زمانی افزایش ناگهانی داشته‌است. ریسک نهایی نیز افزایش در گذر زمان را نشان می‌دهد.



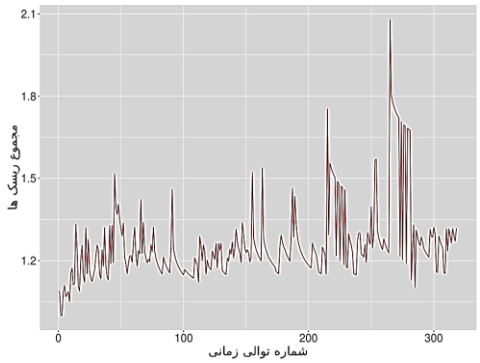
الف) ریسک عجول‌بودن
A: Risk of rushing

[1] Apache Spark

[DOI: 10.61882/jsdp.22.2.31]



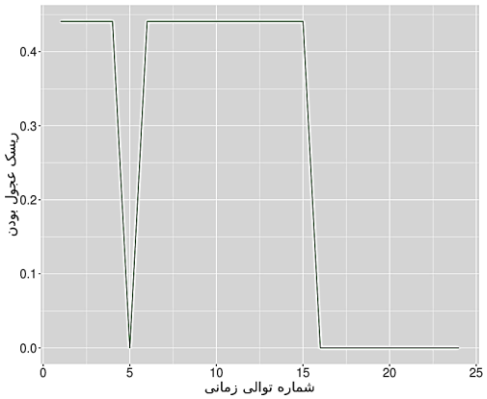
ب) ریسک تغییر رفتار
B: Risk of Behavior Change



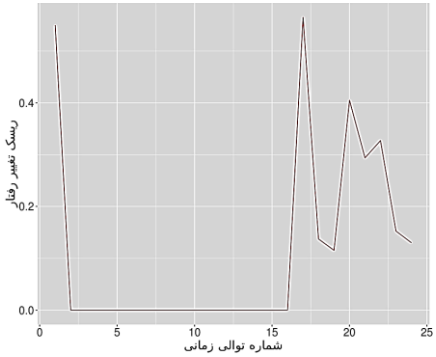
ج) مجموع ریسک‌ها
C: Total Risk

(شکل-۱): خریدار با کد مشتری رمزشده ۱۳۲۷۹ که به‌وسیلهٔ روش پیشنهادی یکی از دَه مورد با بالاترین ریسک تشخیص داده شده‌است.

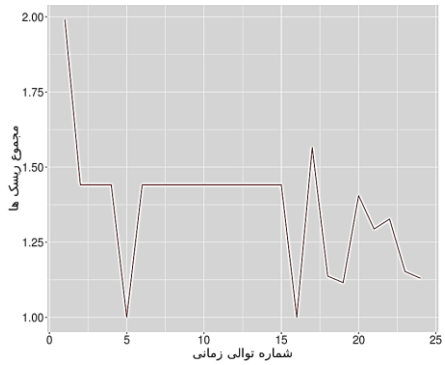
(Figure-1): Buyer with 13279 encrypted customer code, which is identified by one of the 10 items with the highest risk by the proposed method.



الف) ریسک عجول‌بودن
A: Risk of rushing



ب) ریسک تغییر رفتار
B: Risk of Behavior Change



ج) مجموع ریسک‌ها
C: Total Risk

(شکل-۲): خریدار با کد مشتری رمزشده ۱۴۳۰۳ که به‌وسیلهٔ روش پیشنهادی یکی از ده مورد با بالاترین ریسک تشخیص داده شده‌است.

(Figure-2): Buyer with 14303 encrypted customer code, which is identified by one of the 10 items with the highest risk by the proposed method.

شکل (۲) نیز به‌طور مشابه برای نمونهٔ دیگری از بین دَه نمونه با بالاترین ریسک رسم شده‌است. ریسک تغییر رفتار این فرد نشان می‌دهد که در ابتدا ریسک بالا بوده‌است. در روش‌های پروفایلینگ درعمل به‌طورمعمول بر صحت میزان ریسک در زمان‌های ابتدایی نمی‌توان اطمینان داشت. با گذر زمان مقدار عدد ریسک دقیق‌تر و قابل اتکاتر شده و نشان‌دهندهٔ مشکوک به تقلب‌بودن تراکنش است.

۷- نتیجه‌گیری

در این مطالعه، روشی مبتنی بر رفتار برای پیش‌بینی بلادرنگ دست‌کاری بازار ارائه شده‌است. برای همهٔ خریداران/فروشندگان سهم، رخنمای رفتاری تشکیل می‌شود؛ ویژگی‌هایی همچون بازه‌های مبلغ خریدوفروش، زمان خریدوفروش، معامله با یک معامله‌گر خاص و عجول‌بودن به‌عنوان مشخصه برای توصیف هر فرد در نظر گرفته می‌شود. این اطلاعات برای یک بازه زمانی به‌قدرکافی، طولانی پردازش شده و در رخنمای افراد ذخیره می‌شود. در این قسمت، کافی‌بودن حجم داده حائز اهمیت بالایی است؛ چراکه در غیر این صورت مدلِ سازندهٔ رخنماهای کاربران قادر نخواهد بود تمام نوسانات رفتاری طبیعی فرد را ذخیره کند تا در آینده بتواند رفتارهایی را که به‌درستی نابه‌هنجار هستند، از نوسانات رفتاری عادی شناسایی کند؛ سپس از روشی آماری برای تشخیص تغییرات ناگهانی رفتار افراد استفاده می‌شود. برای هر رفتار خرید/فروش جدید از هر کاربر عدد ریسک محاسبه می‌شود که تعیین‌کنندهٔ میزان کژرفتاری از رفتار معمول فرد است. درصورتی‌که این تغییر رفتار قابل ملاحظه باشد، فرد برای بررسی‌های بیشتر مورد توجه قرار می‌گیرد و در غیر این صورت رخنمای او

[DOI: 10.61882/jsdp.22.2.31]

به‌روزرسانی می‌شود. نتایج نشان می‌دهند که روش پیشنهادی برای تشخیص فرانت‌رانینگ مفید است؛

پس از پایدارشدن این روش کشف تقلب، به‌عنوان پیشنهاد ادامهٔ کار به منظور افزایش وضوح تشخیص در صورت وجود داده‌های کافی می‌توان ویژگی‌های رفتاری بیشتری را وارد آزمایش کرد و نتایج برچسب‌گذاری پس از آستانه‌گذاری را بهبود داد.

یک پساپیشنهاد برای کار ارائه‌شده در این مقاله می‌تواند مدل‌های ترکیبی (هیبریدی) باشد؛ برای مثال می‌توان از ترکیب روش حاضر (عدم قطعیت) با GBT[1] (مدل‌سازی غیرخطی) استفاده کرد. پیشنهاد دیگر بهینه‌سازی مقدار آستانهٔ استفاده‌شده در این مقاله است.

8-References ۸-مراجع

[1] *World Bank Open Data*, 2023, Available: [Online] https://data.worldbank.org.

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

[3] Aksenov, A., Grebenchshikova, E., Fayzrakhmanov, R. "Front-running Model in the Stock Market", *ieeexplore*, 2020.

[4] V. Azevedo, Ch. Hoegner, "Enhancing stock market anomalies with machine learning", *Quantitative finance and accounting*, vol. 60, pp. 195-230, 2023

[5] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances", *Expert systems with applications*, Elsevier, vol.193, pp. 1-34, 2022.

[6] C. Poutre, D. Chetelat, M. Morales, "Deep unsupervised anomaly detection in high-frequency markets", *The journal of finance and data science*, vol. 10, pp. 1-18, 2024.

[7] D. Y. Chiu, J. Y. Zhou, and Zh. Ch. Wang, "Appling artificial immune algorithm to explore the seasonal effect in the stock market", *International Conference on Software Intelligence and Applications*, 2014.

[8] Y. Cao, Y. Li, S. Coleman, A. Belareche, and T. M. McGinniti, "Hidden Markov model with abnormal states for detecting stock price manipulation", *IEEE International Conference on Systems, Man, and Cybernetics*, 2013.

[9] Y. Cao, Y. Li, S. Coleman, A. Belareche, T. M. McGinniti, "Adaptive hidden Markov model with anomaly states for price manipulation detection", *IEEE Transactions on Neural Networks and learning systems*, vol. 26, pp. 318-330, 2015.

[10] *Imperva's Web Application Firewall data sheet*. Available: https://www.imperva.com/docs/TB_Dynamic_Profilin g.pdf.

[11] *LightCyber and Check Point Advanced Threat Protection solution brief*. Available: https://www.checkpoint.com/download/downloads/products/solution-brief/SB_LightCyber.pdf.

[12] *HP Unifies Network Security Detection to Identify, Contain and Neutralize Patient Zero Infections*. Available: http://www.hp.com/hpinfo/newsroom/press_kits/2014/HPProtect2014/HPTippingPoint_Advisory.pdf.

[13] *Finding Advanced Threats Before They Strike: A Review of Damballa Failsafe Advanced Threat Protection and Containment*. Available :http://www.sans.org/reading-room/whitepapers/analyst/finding-advanced-threats-strike-review-damballa-failsafe-advanced-threat-protecti-34705.

[14] کاظمی تبار، سیدجواد، شهباززاده، مجید، «کشف تقلب در بازار بورس اوراق بهادار با استفاده از کاربرد نامساوی چبیشف»، *فصلنامهٔ پردازش علائم و داده‌ها*، دورهٔ ۱۷، شمارهٔ ۱، صص ۳-۱۴، ۱۳۹۹.

[14] S. J. Kazemitabar, M. Shahbazzadeh, "Stock market fraud detection, a probabilistic approach", *Signal and data processing*, vol. 17, no. 6, 2020.

[15] Y. Kim, S. Y. Sohn, "Stock fraud detection using peer group analysis", *Expert Systems with Applications*, pp. 8986-8992, 2012.

[16] B. Baesense, et al, "Fraud analytics using descriptive, predictive, and social network techniques", *Wiley*, 2015.

[17] V. Goldwasser, "Stock market manipulation and short selling", *Centre of corporate law and securities regulation, Faculty of law*, The University of Melborn, 1999.

[18] *International Organization of Securities Commissions (IOSC),* 2023, Available: https://www.iosco.org.

[19] S. S. Huebner, "The Stock Marke", *Kessinger Publishing*, 2006.

[20] A. Franklin, and D. Gale, "Stock-Price Manipulation", *Review of FinancialStudies*, vol. 5, pp. 503-529, 1992.

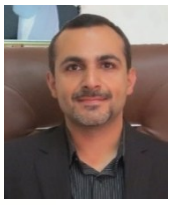
[21] H. Hamedinia, R.Raei, S. Bajalan, S. Rouhani, "Analysis of Stock Market Manipulation using Generative Adversarial Nets and Denoising Auto-Encode Model", *Advances in Mathematical Finance and Applications*, 7 (1), pp. 133-151, 2021.

[22] Z. Shaeiri, S. J. Kazemitabar, "Fast unsupervised autimobile insurance fraud detection based on spectral ranking of anomalies", *International Journal of Engineering*, vol. 33, no. 7, pp. 1240-1248, 2020.

[1] Gradient Boosted Trees

[23] Z. Shaeiri, S. J. Kazemitabar, Sh. Bijani, M. Talebi, "Behavior-Based online anomaly detection for a nationwide short message service", *Journal of AI and data mining*, vol. 7, no. 2, pp. 239-247, 2019.
[24] J. D. Kirkland et.al., "The NASD regulation advanced detection system (ASD)", *AI Magazine*, vol. 20, 1999.
[25] H. Goldberg et.al., "The ANSD securities observation, news analysis and regulation systems (SONAR)", *American Association for Artificial Intelligence (AAAI)*, 2003.
[26] K. Golmohammadi, O. R. Zaiane, "Data mining applications for fraud detection in securities market", *Intelligence and Security Informatics Conference (EISIC),* 2012.
[27] A. Kr, S. Yadav, and Marpe Sora, "Fraud Detection in Financial Statements using Text Mining Methods: A Review", *IOP conference series*, 2021.
[28] Z. Yi, et.al, "Fraud detection in capital markets: A novel machine learning approach", *Epert Systems with Applications*, Elsevier, vol. 231, 2023.
[29] S. Kim, J. Hong, Y. Lee, "A GANs-Based Approach for Stock Price Anomaly Detection and Investment Risk Management", *Fourth ACM international conference on AI in finance*, pp. 1-9, 2023.
[30] J. Neyman, and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses", *Phil. Trans*., pp. 694–706, 1993.
[31] S. Lin, and D. J. Costello, "Error Control Coding (2nd Edition)", *Pearson*, 2014.
[32] S. Viaene, G. Dedene, and R. Derig, "Auto claim fraud detection using Bayesian learning neural networks", *Expert systems with applications*, vol. 29, no. 3, pp. 653-666, 2005.
[33] S. Viaene, R. Derrig, and G. Dedene, "A case study of applying boosting naive bayes to claim fraud diagnosis", *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 5, pp. 612-620, 2004.
[34] S. Viaene, R. Derrig, B. Baesens, and G. Dedene, "A comparison of state-of-the- art classification techniques for expert automobile insurance claim fraud detection" *The Journal of Risk and Insurance*, vol. 69, no. 3, pp. 373-421, 2002.

زهرا شعیری دانش‌آموختهٔ دکترا در رشتهٔ مهندسی برق از دانشگاه صنعتی نوشیروانی بابل است. از جمله علائق پژوهشی وی پردازش سیگنال، تئوری اطلاعات، شناسایی آماری الگو و یادگیری ماشین است.
نشانی رایانامهٔ ایشان عبارت است از:
shaeiri.zahra@gmail.com



سید جواد کاظمی تبار دورهٔ کارشناسی خود را در سال ۱۳۸۲ در دانشگاه صنعتی شریف به پایان رسانده‌است. ایشان در سال ۱۳۸۷ مدرک دکترای خود را در رشتهٔ مخابرات در دانشگاه کالیفرنیا در شهر ارواین کسب کرد و تا سال ۱۳۹۳ در شرکت‌های مختلف مهندسی آمریکا به فعالیت پرداخت. ایشان از سال ۱۳۹۱ تا سال ۱۳۹۳ در «سیلیکون ولی» به‌عنوان متخصص داده‌کاوی در شرکت «گاردین آنالیتیکس» به کشف تقلب‌های بانکی کمک می‌کرد. ایشان از سال ۱۳۹۴ عضو هیئت علمی دانشگاه صنعتی نوشیروانی بابل است.
نشانی رایانامهٔ ایشان عبارت است از:
j.kazemitabar@nit.ac.ir

سروش حق وردی مدرک دکترای رشتهٔ مدیریت را از دانشگاه تهران کسب کرده‌است. ایشان تجربه چندین سال بازرسی پرونده‌های تخلف را در فرابورس ایران در کارنامهٔ خود دارد. تمرکز ایشان در بازرسی، رصد فضای مجازی و استفاده از روش‌های آماری برای کشف تقلب است.
نشانی رایانامهٔ ایشان عبارت است از:
haghverdi@ifb.ir

[DOI: 10.61882/jsdp.22.2.31]