Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
DOI: 10.22034/jscit.2026.559514.2195
Volume 15, Number 2, Summer 2026, pp. 64-77
Received: 11/21/2025, Revised: 04/15/2026, Accepted: 05/26/2026

Predicting Artificial Inflation in Cryptocurrency Prices Using a Hybrid Approach

Faezeh Mousavi-Firouzkola[1], Hamid Moradi-Kamali[2] and Reza Entezari-Maleki[3*]
School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.
[1]Faezeh_mousavi@comp.iust.ac.ir, [2]ha_moradi@comp.iust.ac.ir, and [3]entezari@iust.ac.ir

[*] Corresponding author's address: Reza Entezari-Maleki, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.

Abstract- Pump-and-dump (P&D) schemes, orchestrated through social media, are a prevalent form of manipulation in cryptocurrency markets. While most research has focused on detecting these events, the critical challenge of predicting the specific target coin remains largely underexplored. Existing attempts at target prediction often analyze textual content or time-series anomalies for individual coins in isolation. This paper introduces a fundamentally different approach, framing target prediction as a sequential problem where the history of a P&D channel's activity informs the next target. In this paper, we firstly develop a robust labeling method that integrates textual signals from Telegram with financial data, identifying 1,200 P&D events from 2022-2024. In the second step, we design a sequence-based predictive model that uses self-attention to learn historical P&D patterns within a channel and cross-attention to correlate these patterns with the features of potential target coins. Evaluated on real-world data, our proposed model achieves an Area Under the Curve (AUC) of 89% and demonstrates a significant improvement of approximately 10% on the Hit Ratio at K (HR@K) metric compared to a state-of-the-art baseline. Our findings demonstrate that modeling the sequential context of manipulation campaigns is a superior strategy for proactively identifying target coins.
To access the code and data related to the approach presented in this paper, you may refer to the link below.
https://github.com/faezemousavi/Pump-and-Dump

Keywords- Cryptocurrency Market, Market Manipulation, Fraud Detection, Text Analysis.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
شناسه دیجیتال: 10.22034/jscit.2026.559514.2195
جلد ۱۵، شماره ۲، تابستان ۱۴۰۵، صفحه ۶۴-۷۷
دریافت: ۱۴۰۴/۰۸/۳۰، بازنگری: ۱۴۰۵/۰۱/۲۶، پذیرش: ۱۴۰۵/۰۳/۰۵

پیش‌بینی افزایش مصنوعی قیمت رمزارزها با استفاده از رویکرد ترکیبی

فائزه موسوی فیروزکلا[۱]، حمید مرادی کمالی[۲]، رضا انتظاری ملکی*[۳]
دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
[1]faezeh_mousavi@comp.iust.ac.ir, [2]ha_moradi@comp.iust.ac.ir, [3]entezari@iust.ac.ir

* نشانی نویسنده مسئول: رضا انتظاری ملکی، تهران، دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر.

چکیده- کلاهبرداری به سبک پامپ‌ودامپ (Pump-and-Dump) که عمدتاً از طریق شبکه‌های اجتماعی سازمان‌دهی می‌شود، یکی از رایج‌ترین انواع دستکاری در بازار رمزارزها به شمار می‌آید. در حالی که اغلب پژوهش‌ها بر تشخیص زمان وقوع این رویدادها تمرکز داشته‌اند، چالش اساسی پیش‌بینی رمزارز هدف همچنان کمتر مورد توجه قرار گرفته است. پژوهش‌های موجود معمولاً با تحلیل محتوای متنی یا ناهنجاری‌های سری‌زمانی به شناسایی این وقایع می‌پردازند، در حالی که رویکردهای ترکیبی کمتر بررسی شده‌اند. در این مقاله، رویکردی ترکیبی ارائه می‌شود که مسئله پیش‌بینی رمزارز هدف را در کنار پیش‌بینی زمان وقوع کلاهبرداری، به‌صورت یک مسئله دنباله‌ای بازتعریف می‌کند. این رویکرد با در نظر گرفتن تاریخچه فعالیت‌های پیشین کانال‌های کلاهبردار، قادر است اهداف بعدی را از نظر رمزارز هدف و زمان وقوع رویداد پیش‌بینی کند. بدین منظور، ابتدا یک روش برچسب‌گذاری باثبات توسعه داده شد که سیگنال‌های متنی استخراج‌شده از تلگرام را با داده‌های مالی ترکیب کرده و ۱۲۰۰ رویداد را در بازه زمانی ۲۰۲۲ تا ۲۰۲۴ شناسایی می‌کند. سپس، یک مدل پیش‌بینی مبتنی بر دنباله ارائه می‌شود که با بهره‌گیری از سازوکار خودتوجه، الگوهای تاریخی پامپ‌ودامپ را در سطح کانال می‌آموزد و با استفاده از سازوکار توجه متقابل، روابط غیرخطی میان دنباله و ویژگی‌های رمزارزهای کاندید را استخراج می‌کند. ارزیابی بر روی داده‌های واقعی نشان می‌دهد مدل پیشنهادی به AUC برابر با ۸۹٪ دست یافته و در معیار HR@K حدود ۱۰٪ بهبود مطلق نسبت به مدل معیار ایجاد کرده است. نتایج نشان‌دهنده برتری رویکرد ترکیبی پیشنهادی نسبت به روش‌های پیشین در شناسایی این نوع کلاهبرداری است.

https://github.com/faezemousavi/Pump-and-Dump

واژه‌های کلیدی: بازار رمزارزها، دستکاری بازار، تشخیص کلاهبرداری، تحلیل داده‌های متنی.

۱- مقدمه

بازار رمزارزها به‌عنوان یکی از پرنوسان‌ترین و در عین حال رو به رشدترین بازارهای مالی، همواره بستری مستعد برای فرصت‌طلبی و دستکاری مخربان و کلاهبرداران بوده است [۱و۲]. نبود نظارت‌های سنتی و ساختار متمرکز، در کنار طمع سرمایه‌گذاران برای کسب سودهای عظیم، این بازار را به طعمه‌ای مناسب برای انواع کلاهبرداری‌ها تبدیل کرده است. یکی از رایج‌ترین و خطرناک‌ترین این دستکاری‌ها، کلاهبرداری به روش پامپ‌ودامپ[۱] است [۳]. در این روش، گروهی از افراد، یک رمزارز با حجم معاملات و ارزش بازار پایین را انتخاب می‌کنند، قیمت آن را به‌صورت مصنوعی و از طریق خریدهای حجیم یا تبلیغات کاذب افزایش می‌دهند، و سپس در اوج قیمت، دارایی خود را می‌فروشند و به زیان اکثریت شرکت‌کنندگان، سودهای هنگفتی را دریافت می‌کنند [۴و۵]. شبکه‌های اجتماعی به‌ویژه تلگرام، در هماهنگی و تبلیغ این فرآیند نقش کلیدی دارند؛ کانال‌های کلاهبرداری پامپ‌ودامپ با انتشار پیام‌های فریبنده و

ادعاهای دروغین، سرمایه‌گذاران خرد و بی‌تجربه را به شرکت در این کلاهبرداری ترغیب می‌کنند. این پدیده، نه تنها اعتماد عمومی به بازار را تضعیف می‌کند، بلکه تهدیدی جدی برای امنیت سرمایه‌گذاران سالم در بازار رمزارزها محسوب می‌شود [۶-۸]. در راستای مقابله با این نوع از کلاهبرداری، توسعه ابزارهایی برای پیش‌بینی و پیشگیری از این رفتارهای سودجویانه، به ضرورتی اجتناب‌ناپذیر در مسیر شفاف‌سازی و سالم‌سازی بازار رمزارزها تبدیل شده است.

در سال‌های اخیر، پژوهشگران حوزه داده و بازارهای مالی تلاش کرده‌اند تا با استفاده از تحلیل‌های مبتنی بر سری‌های زمانی [۹]، نشانه‌های مبتنی بر حجم معاملات و رفتار یا پیام‌های کاربران در شبکه‌های اجتماعی، الگوهای رفتاری این نوع کلاهبرداران در بازارهای مالی را شناسایی کنند [۱۰و۱۱]. بیشتر این تحقیقات بر تشخیص پس‌نگرانه[۲] یا پیش‌بینی زمان تقریبی کلاهبرداری تمرکز داشته‌اند. اما این رویکردهای انفعالی، تنها زمانی مفید هستند که کلاهبرداری آغاز شده یا به پایان نزدیک شده باشد. در چنین شرایطی، اطلاعات برای پیشگیری از وقوع کلاهبرداری چندان مفید نخواهد بود، چرا که زمان مقابله با آن گذشته است. مهم‌تر آنکه، تقریباً هیچ‌یک از پژوهش‌های پیشین به یک سؤال کلیدی پاسخ نمی‌دهند: کدام رمزارز در آستانه پامپ قرار دارد؟ این در حالی است که اگر بتوان هدف بعدی را پیش‌بینی کرد، سرمایه‌گذار یا صرافی قادر خواهد بود تصمیمی پیش‌دستانه و محافظه‌کارانه اتخاذ کند. بنابراین، یک چالش مهم در این حوزه تحقیقاتی وجود دارد که هدف اصلی این پژوهش می‌باشد و آن رفع ناتوانی مدل در پیش‌بینی رمزارز هدف در یک رویداد پامپ‌ودامپ است. پر کردن این خلأ می‌تواند گامی مهم در ارتقاء امنیت مالی سرمایه‌گذاران و ساختارمندسازی مقابله با کلاهبرداری در بازار رمزارزها باشد [۱۲و۱۳].

در این پژوهش، ما رویکردیی نوین برای پیش‌بینی رمزارز هدف در کلاهبرداری‌های پامپ‌ودامپ ارائه می‌دهیم. برخلاف رویکردهای گذشته که بر تحلیل‌های پس‌ازوقوع (پس‌نگرانه) یا مبتنی بر پیش‌بینی زمان وقوع تمرکز داشتند، مدل پیشنهادی در این مقاله از ترکیب داده‌های متنی استخراج‌شده از کانال‌های تلگرامی و شاخص‌های مبتنی بر بازار مالی، مانند حجم معاملات، نوسانات قیمتی و الگوهای قبلی پامپ استفاده می‌کند تا علاوه بر پیش‌بینی زمان وقوع، لیستی از رمزارزهای محتمل را نیز ارائه کند. در قلب این رویکرد، معماری مبتنی بر سازوکار توجه[۳] قرار دارد که قادر است وابستگی‌های زمانی و الگوهای غیرخطی را به‌خوبی یاد بگیرد. نوآوری‌های اصلی این مقاله شامل سه مورد است:

- ارائه یک رویکرد برچسب‌گذاری[۴] ترکیبی: رویکرد پیشنهادی با تحلیل هم‌زمان داده‌های متنی (برگرفته از شبکه‌های اجتماعی) و شاخص‌های مبتنی بر بازار مالی، رویکردی پایدار برای شناسایی رویدادهای پامپ‌ودامپ رمزارز هدف فراهم می‌آورد که بر محدودیت‌های رویکردهای انسانی غلبه می‌کند.
- توسعه یک مدل پیش‌بینی مبتنی بر دنباله: معماری پیشنهادی با استفاده از لایه‌های خودتوجه[۵] و توجه‌متقابل[۶]، الگوهای تاریخی در انتخاب رمزارزها توسط کلاهبرداران را به وسیله ویژگی‌های متنی و مالی مدل‌سازی کرده و رویدادهای آتی را پیش‌بینی می‌کند.
- ارزیابی مدل در مقایسه با مدل‌های معیار: مدل پیشنهادی در یک مجموعه‌داده متشکل از بیش از ۱۲۰۰ رویداد پامپ از سال‌های ۲۰۲۲ تا ۲۰۲۴، مورد ارزیابی قرار گرفته که نتایج حاصل نشان‌دهنده عملکرد قابل‌توجه این مدل نسبت به رقبا است. رویکرد پیشنهادی با دستیابی به مساحت زیر نمودار[۷] (AUC) برابر با ۸۹٪ و بهبود حدود ۱۰٪ در معیار نرخ برخورد[۸] (HR) نسبت به رویکردهای معیار، کارایی برتر خود را نشان می‌دهد.

۲- کارهای مرتبط

مطالعه کلاهبرداری در بازار رمزارزها، به‌ویژه رویدادهای پامپ‌ودامپ، در سال‌های اخیر توجه پژوهشگران زیادی را به خود جلب کرده است. بیشتر تحقیقات پیشین تمرکز خود را بر شناسایی و تشخیص این رویدادها گذاشته‌اند، در حالی که پیش‌بینی رویدادها، کمتر مورد توجه قرار گرفته است. در این بخش، آثار مرتبط را در سه محور اصلی بررسی می‌کنیم: تحقیقات در حوزه شناسایی رویدادهای پامپ‌ودامپ، تشخیص ناهنجاری (کلاهبرداری) در داده‌های بازار محور، و تحقیقات در حوزه پیش‌بینی رمزارز هدف.

۲-۱- شناسایی رویدادهای پامپ‌ودامپ

بیشتر پژوهش‌های انجام‌شده در حوزه دست‌کاری بازار رمزارزها، به تشخیص رویدادهای پامپ‌ودامپ در زمان وقوع یا پس از آن اختصاص دارد. این رویکردها نقش مهمی در آشکارسازی رفتارهای مخرب ایفا می‌کنند، اما ذاتاً پس‌نگرانه هستند و برای پیشگیری کاربرد محدودی دارند.

برخی از این مطالعات مبتنی بر تشخیص ناهنجاری در داده‌های بازار

هستند. برای مثال، کمپس و کلینبرگ [۱۴] یک رویکرد بدون‌ناظر[۹] را معرفی کردند که با تحلیل نوسانات ناگهانی در حجم معاملات و قیمت، رویدادهای مشکوک را شناسایی می‌کند. به‌طور مشابه، مورجیا و همکارانش [۱۲] با بهره‌گیری از ویژگی‌های کتابچه سفارش[۱۰] و توالی معاملات، مدل‌هایی را برای تشخیص رفتارهای مشکوک طراحی کرده‌اند. از سوی دیگر، دسته‌ای از پژوهش‌ها از داده‌های متنی و رسانه‌های اجتماعی بهره برده‌اند. به‌عنوان نمونه، میرطاهری و همکاران [۷] با تحلیل پیام‌ها و احساسات منتشرشده در X و تلگرام، به شناسایی رویدادهای هماهنگ دستکاری در بازارهای رمزارز پرداخته‌اند. افزون بر این، آن‌ها اثربخشی پیش‌بینی‌های خود را با بهره‌گیری از ویژگی‌های استخراج شده از رسانه‌های اجتماعی و شاخص‌های اقتصادی ارزیابی کرده‌اند؛ امری که بر اهمیت نظارت بر بحث‌های آنلاین برای شناسایی فعالیت‌های احتمالیِ تقلب در بازار ارزهای دیجیتال را برجسته می‌کند. همچنین ون و همکاران [۱۵] در مطالعه‌ای جامع، هفت گروه از داده‌های رسانه‌های اجتماعی را شناسایی کرده‌اند که در نوسانات قیمت رمزارزهای مورد بررسی نقش داشتند. این تحقیق نکاتی را در مورد عوامل مؤثر بر قیمت ارزهای دیجیتال ارائه کرده و رویکردی نوین برای ارزیابی تأثیرات رویدادهای کلاهبرداری در نوسانات قیمت ایجاد کرده که می‌تواند به عنوان معیاری برای مطالعات آینده استفاده شود. برخی پژوهش‌ها نیز به بررسی الگوهای فعالیت بات‌ها و نحوه انتشار اطلاعات در شبکه‌های اجتماعی پرداخته‌اند؛ برای مثال، فراتیک و همکاران [۱۶] مدلی ارائه کرده‌اند که با شبیه‌سازی رفتار عامل‌های مختلف، از جمله عامل متقلب، توانسته تغییرات قیمت و حجم معاملات را تبیین کند و رابطه میان نقدینگی و میزان کارایی دست‌کاری در بازار را نشان دهد. از مهم‌ترین مزایای این مدل می‌توان به کاربرد آن برای سیاست‌گذاران در ارزیابی راهکارهای مقابله با تقلب اشاره کرد. با این حال، مدل دارای ساده‌سازی‌هایی است که ممکن است به بروز خطا در پیش‌بینی یا کاهش دقت در برآورد تأثیر واقعی عامل متقلب منجر شود و نتواند برخی ناهنجاری‌ها را به‌طور کامل توضیح دهد. در ادامه Wu و همکارانش [۱۷] چارچوب PumpWatcher را پیشنهاد دادند که با بهره‌گیری از شبکه‌های عصبی گراف و یادگیری مقابله‌ای، رفتار تراکنش‌ها را در گراف‌های زمانی مدل‌سازی کرده و به شناسایی بلادرنگ پامپ‌ودامپ می‌پردازد. نتایج این مطالعه نشان می‌دهد که استفاده از مدل‌های گرافی و روش‌های پیشرفته یادگیری ماشین می‌تواند دقت تشخیص را به‌طور قابل توجهی افزایش دهد و نرخ خطاهای مثبت و منفی را کاهش دهد. با وجود اثربخشی این رویکردها در نظارت بر بازار، یک ضعف بنیادین در آن‌ها وجود دارد و آن ماهیت پس‌نگرانه در تشخیص کلاهبرداری است. در حالی‌که هدف مقاله حاضر ارائه‌ی رویکردی با قابلیت پیش‌بینانه و پیش‌گیرانه برای شناسایی رمزارزهای در معرض پامپ‌ودامپ است.

۲-۲- تشخیص ناهنجاری در داده‌های سری‌زمانی

مسئله پیش‌بینی رویداد پامپ‌ودامپ را می‌توان به عنوان یک مورد خاص از تشخیص ناهنجاری در داده‌های سری‌زمانی بازارهای مالی در نظر گرفت. در این حوزه، مدل‌های متنوعی برای شناسایی رفتارهای ناهنجار توسعه یافته‌اند. رویکردهای کلاسیک بر پایه مدل‌های آماری و یادگیری ماشین طراحی شده‌اند، در حالی‌که رویکردهای جدیدتر، از معماری‌های یادگیری عمیق برای کشف وابستگی‌های پیچیده غیرخطی استفاده می‌کنند. به‌عنوان مثال، ژو و همکارانش [۱۸] مدل Anomaly Transformer را معرفی کرده‌اند که با استفاده از سازوکار توجه، ناهنجاری‌های زمانی و زمینه‌ای را شناسایی می‌کند. همچنین، روی و همکاران [۱۹] رویکرد GAD-NR را توسعه داده‌اند که با مدل‌سازی روابط گرافی بین ویژگی‌های بازار محور، به کشف ناهنجاری‌های مشکوک در داده‌های بازار می‌پردازد. کلاف و همکاران [۲۰] ضمن بررسی ویژگی‌های پامپ‌ودامپ به تأثیرات بلندمدت رویدادهای پامپ‌ودامپ بر قیمت رمزارزها نیز پرداخته‌اند. نتایج پژوهش آن‌ها نشان می‌دهد که رویدادهای پامپ‌ودامپ تأثیر منفی بلند مدتی بر ارزش ارزهای پایه دارند. این یافته از این دیدگاه حمایت می‌کند که چنین طرح‌هایی در نهایت برای قیمت ارزهای دیجیتال مضر است؛ نتیجه‌ای که با برخی مطالعات پیشین که تأثیر مثبت و موقتی را گزارش کرده بودند، در تضاد است.

با این حال، این مدل‌ها ماهیتی کلی‌نگر (بدون در نظر گرفتن انواع خاص کلاهبرداری در بازارهای مالی) دارند و برای تشخیص دست‌کاری هدفمند و برنامه‌ریزی‌شده پامپ‌ودامپ در رمزارزها، فاقد دانش زمینه‌ای و تاریخی از اتفاقات مشابه هستند. علاوه بر این، این مدل‌ها عمدتاً به کشف رفتار غیرعادی بازار می‌پردازند، در حالی‌که پیش‌بینی پامپ‌ودامپ نیازمند درک الگوهای رفتاری متوالی و هدفمند کاربران است. این مسئله، انگیزه اصلی ما برای طراحی یک مدل ترکیبی مبتنی بر تحلیل داده‌های متنی و رفتارشناسی کاربران در کنار تحلیل داده‌های بازار بوده است.

۲-۳- پیش‌بینی رمزارز هدف در رویدادهای پامپ‌ودامپ

مهم‌ترین بخش از کارهای مرتبط، پژوهش‌هایی هستند که به‌صورت مستقیم به مسئله پیش‌بینی رمزارز هدف در رویدادهای پامپ‌ودامپ پرداخته‌اند. این دسته از تحقیقات نسبتاً جدید بوده و تلاش دارند

به مسئله‌ای دشوارتر و البته کاربردی‌تر نسبت به کارهای پیشین بپردازند.

ژو و لیوشیتس [۲۱] از اولین کسانی بودند که ساختار رویدادهای پامپ‌ودامپ را تحلیل کرده و با استفاده از رویکردهایی مانند LASSO و درخت تصادفی[۱۱] تلاش کردند تا رمزارز هدف را از روی ویژگی‌های بازار پیش‌بینی کنند. اگرچه مطالعه آن‌ها نشان داد که این پیش‌بینی تا حدی امکان‌پذیر است، اما تمرکز آن بر ویژگی‌های ایستا بوده و از جنبه‌های زمانی و رفتاری غافل مانده است، زیرا الگوهای بررسی‌شده حاصل از بررسی پویای بازار نبوده و از قوانین ثابت استفاده می‌کند. در کار ارائه شده توسط هو و همکاران [۲۲] پیشرفت فراتری نسبت به [۲۱] حاصل شده است. در [۲۲] مدل شبکه عصبی متوالی[۱۲] (SNN) معرفی شده است که اولین معماری برای مدل‌سازی رمزارزهای پامپ‌شده در گذشته به‌منظور پیش‌بینی رمزارز آینده می‌باشد. این رویکرد نقش الگوهای زمانی را در رفتار سازمان‌دهندگان پامپ‌ودامپ برجسته کرده و بنیانی نوین برای پژوهش‌های آتی بنا نهاده است. ما نیز در این مقاله، از رویکرد ارائه شده در [۲۲] به عنوان یکی از رویکردهای مورد مقایسه استفاده کرده‌ایم.

با این حال، مدل SNN نیز محدودیت‌هایی دارد. از جمله می‌توان به ناتوانی در یادگیری وابستگی‌های بلندمدت، ضعف در ادغام مؤثر داده‌های متنوع، مانند داده‌های مالی به همراه داده‌های متنی، و اتکای آن به یک رویکرد برچسب‌گذاری ساده اشاره کرد. این موارد، نیاز به مدلی پیشرفته‌تر و مجموعه‌داده‌ای دقیق‌تر را آشکار می‌سازد.

این محدودیت‌ها به‌ویژه زمانی پررنگ‌تر می‌شوند که بدانیم روش‌های توالی‌محور مانند SNN، اگرچه برای پردازش داده‌های ترتیبی طراحی شده‌اند، عمدتاً بر یادگیری الگوهای محلی و وابستگی‌های کوتاه‌مدت تمرکز دارند و در مواجهه با نویز بالا، وقفه‌های زمانی نامنظم یا توالی‌های بلندمدت در داده‌های مالی، کارایی خود را از دست می‌دهند. در چنین شرایطی، ناتوانی این مدل‌ها در تشخیص همبستگی‌های معنایی عمیق میان داده‌های ساختاریافته (مانند سری‌های زمانی قیمت) و داده‌های بدون‌ساختار (مانند متن) باعث می‌شود که نتوانند از تمام ظرفیت اطلاعاتی موجود استفاده کنند [۲۳و۲۴].

مدل پیشنهادی (SNN-TA)، نسخه توسعه‌یافته مدل SNN است که با معرفی یک سازوکار توجه متقابل بین ویژگی‌های متنی خاص هر هدف (رمزارز کاندید) و بازنمایی دنباله رویدادهای گذشته عمل می‌کند. این معماری به مدل اجازه می‌دهد تا به‌صورت انتخابی بر عناصری از دنباله تاریخی تأکید کند که بیشترین ارتباط را با هدف پیش‌بینی دارند و در نتیجه قدرت تفکیک‌پذیری مدل را بهبود می‌بخشد.

پژوهش حاضر، دقیقاً در همین حوزه ورود کرده و با توسعه یک معماری مبتنی بر سازوکار توجه و امکان یادگیری تعاملات پیچیده میان توالی‌های زمانی و مفاهیم متنی، به یک مدل‌سازی بهتری از روابط زمانی و مفهومی می‌رسد. همچنین با طراحی یک رویکرد برچسب‌گذاری ترکیبی (متنی-مالی)، مجموعه‌داده‌ای قابل اعتمادتر و دقیق‌تر برای یادگیری مدل فراهم می‌نماید. این نوآوری‌ها باعث می‌شود تا مقاله حاضر گامی مهم و ملموس در مسیر مقابله پیش‌گیرانه با دستکاری‌های بازار رمزارزها ارائه کند.

۳- رویکرد پیشنهادی

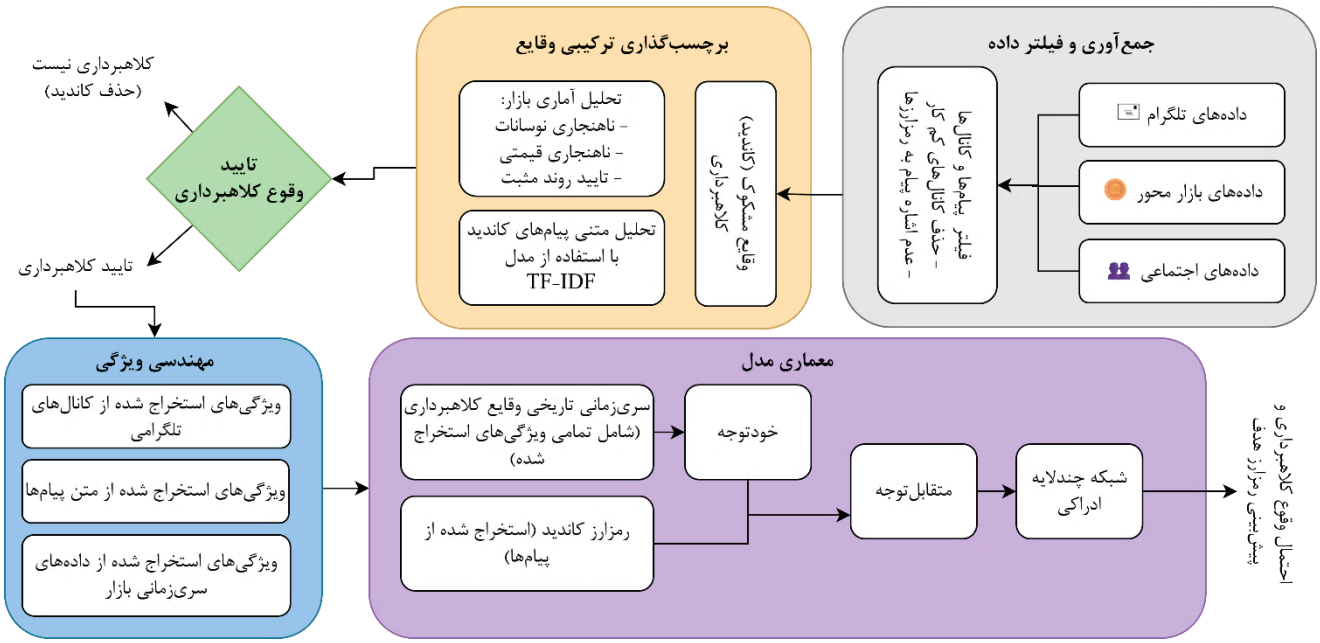
در این بخش، رویکرد پیشنهادی با جزئیات کامل در شکل ۱ شرح داده می‌شود. همان‌گونه که در شکل ۱ نشان داده شده، این رویکرد شامل چهار قسمت اصلی است.

۱. جمع‌آوری و پیش‌پردازش داده: شامل گردآوری داده‌های متنی از کانال‌های تلگرامی منتخب، داده‌های بازار برحسب دقیقه از صرافی‌ها و اطلاعات مکمل مربوط به کانال‌ها و اعضای آن. این داده‌ها سپس با هدف استفاده در مراحل بعد پیش‌پردازش شده و مورد استفاده قرار می‌گیرند. این لایه در شکل ۱ با عنوان جمع‌آوری و فیلتر داده نمایش داده شده است.

۲. برچسب‌گذاری ترکیبی رویدادهای کلاهبرداری: ارائه‌ی رویکردی برای شناسایی رویدادهای پامپ‌ودامپ معتبر و برچسب‌های معیار[۱۳] با ترکیب تحلیل محتوای متنی و معیارهای آماری ناهنجاری بازار. این مرحله با عنوان برچسب‌گذاری ترکیبی وقایع در شکل ۱ نمایش داده شده است.

۳. مهندسی ویژگی: استخراج و بازنمایی ویژگی‌های چندوجهی[۱۴] شامل اطلاعات کانال، ویژگی‌های رمزارز، و شاخص‌های سری‌زمانی. این مرحله در شکل ۱ با عنوان مهندسی ویژگی مشخص شده است.

۴. مدل پیش‌بینی: ارائه‌ی یک مدل دنباله‌محور[۱۵] با سازوکارهای خودتوجه و متقابل‌توجه برای یافتن احتمال انتخاب رمزارز کاندید به عنوان هدف بعدی با بررسی الگوی رویدادهای گذشته. قسمت معماری مدل در شکل ۱ به این موضوع اشاره دارد.



شکل ۱: ساختار رویکرد پیشنهادی

در ادامه، هر یک از این قسمت‌ها در راستای تسهیل بازتولید نتایج به تفصیل توضیح داده می‌شوند.

۳-۱- گردآوری و آماده‌سازی داده‌ها

بر اساس مطالعات پیشین، در بررسی پدیده‌ی پامپ‌ودامپ، در ابتدا از وب‌سایت «PumpOlymp» استفاده می‌شد که امروزه دیگر در دسترس نیست. در حال حاضر، تلگرام به بستر اصلی هماهنگی این‌گونه طرح‌ها تبدیل شده و به‌طور گسترده مورد سوءاستفاده کلاهبرداران قرار می‌گیرد. بخش قابل‌توجهی از طرح‌های پامپ‌ودامپ به‌صورت سازمان‌یافته در کانال‌ها و گروه‌های تلگرام اجرا می‌شود. این کانال‌ها اطلاعاتی نظیر زمان دقیق پامپ، نام کوین هدف، اهداف قیمتی و پیام‌های هیجانی را پیش از وقوع یا هم‌زمان با اجرای پامپ در بازار منتشر می‌کنند. ازاین‌رو، تلگرام به‌عنوان منبع اولیه اطلاعات در این حوزه شناخته می‌شود [۲۵]. با این حال، طبق بررسی‌های انجام‌شده، داده‌های مورد نیاز برای تحلیل را می‌توان در سه دسته‌ی کلی طبقه‌بندی کرد.

داده متنی (تلگرام): پیام‌ها از کانال‌های عمومی تلگرامی که به انتشار سیگنال‌های پامپ‌ودامپ شناخته شده‌اند جمع‌آوری شده است. در مجموع، مجموعه خام پیام‌ها شامل بیش از ۱ میلیون پیام از ۱۵۳ کانال عمومی در بازه اوت ۲۰۲۲ تا اوت ۲۰۲۴ است [۱۱]. پیام‌ها از طریق رابط برنامه کاربردی[۱۶] (API) تلگرام دریافت شده و اطلاعات تکمیلی اعم از زمان ارسال، شناسه کانال، متن پیام و هر نماد رمزارز یا لینک موجود در کانال مربوطه ذخیره شده‌اند.

داده مالی (بازار): برای سیگنال‌های بازار، سری‌های زمانی نمودار شمعی[۱۷] در بازه یک دقیقه‌ای به همراه حجم معاملات برای حدود ۴۰۰ رمزارز در بازه اوت ۲۰۲۲ تا اوت ۲۰۲۴ از API صرافی بایننس[۱۸] استخراج شده است. انتخاب داده یک دقیقه‌ای به علت سرعت بالای وقوع رویدادهای پامپ‌ودامپ و واکنش سریع بازار ضروری است.

ویژگی‌های مکمل/اجتماعی برای رمزارزها: ویژگی‌های مکمل مانند تعداد دنبال‌کنندگان رمزارز در شبکه‌های اجتماعی، تعداد اعضای فعال در شبکه اجتماعی ردیت[۱۹] و ارزش کل بازار[۲۰] از APIهایی مانند CryptoCompare استخراج شده‌اند. این ویژگی‌ها به‌عنوان اطلاعات مفید در ارزیابی اعتبار رمزارز (محبوبیت، اندازه جامعه) در مدل پیش‌بینی به کار می‌روند.

در ادامه برای پیش‌پردازش[۲۱] داده‌های متنی خام تلگرام که شامل محتوای بی‌ربط (بحث‌های عمومی، اسپم، بازنشر) است دو مرحله گزینش به‌صورت زنجیره‌ای اعمال شده است:

۱. گزینش کانال: حذف کانال‌های با فعالیت بسیار پایین (کمتر از ۱۰۰ پیام در کل بازه زمانی). این مرحله کانال‌های غیرفعال یا نامرتبط را کنار گذاشته و تمرکز محاسباتی را بر کانال‌های فعال در رویدادهای پامپ‌ودامپ حفظ می‌کند.

۲. گزینش پیام: بسیاری از پیام‌های این گونه کانال‌ها تبلیغات و پیام‌های نامرتبط است. ازاین‌رو در کانال‌های باقی‌مانده، تنها پیام‌هایی که به‌طور صریح یکی از ۴۰۰ رمزارز هدف را (با نماد یا نام) ذکر کرده یا حاوی کلیدواژه‌های مرتبط با عملیات

پامپ‌ودامپ باشند، نگه داشته می‌شوند. این کار حجم پیام‌ها را به‌شدت کاهش می‌دهد ولی پوشش رویدادهای بالقوه را با دقت بالا حفظ می‌کند.

پس از استخراج و پیش‌پردازش داده‌های خام در ادامه به برچسب‌گذاری رویدادها با رویکرد ترکیبی پیشنهادی می‌پردازیم.

۳-۲- برچسب‌گذاری رویدادها با رویکرد ترکیبی

برچسب‌گذاری خودکار رویدادهای پامپ‌ودامپ یکی از نوآوری‌های کلیدی این مقاله می‌باشد. برچسب‌گذاری انسانی، پرهزینه و وابسته به نظر اشخاص است؛ همینطور رویکردهای مبتنی بر داده‌های بازار نیز شامل نرخ اشتباهات بالا و ناتوانی در پیدا کردن روابط غیرخطی و پیچیده است. از این رو در این مقاله، یک رویکرد نیمه‌خودکار ترکیبی طراحی شده است که سیگنال‌های متنی (اعلان رویداد پامپ‌ودامپ) را با معیارهای مرسوم بازار ترکیب می‌کند تا وقوع رویداد پامپ‌ودامپ را به طور هم‌زمان و از دو جنبه متفاوت آماری و متنی تأیید شده و رمزارز هدف و زمان رویداد برچسب‌گذاری کند.

همانطور که اشاره شد، رویکردهای پیشین یا متکی بر برچسب‌گذاری انسانی بوده‌اند یا بر اساس قواعد ساده بازار (مثلاً جهش قیمت با حد آستانه از پیش تعیین‌شده) عمل می‌کنند. این گونه رویکردها مقیاس‌پذیر نبوده و مستعد مثبت‌های کاذب (برچسب‌گذاری رشد معمولی بازار به عنوان یک رویداد کلاهبرداری) هستند. رویکرد ترکیبی ما در میان سیل پیام‌های ارسال شده در کانال‌های انتخابی تلگرام با استفاده از رویکردهای تحلیل متن، پیام‌های مشکوک را انتخاب می‌کند، سپس در بازه زمانی اطراف آن پیام، رفتار بازار را با استفاده از رویکردهای مبتنی بر بازار بررسی می‌کند. تنها کاندیدهایی که تمام شروط مکمل مالی و متنی را برآورده کنند، به‌عنوان رویداد پامپ‌ودامپ تأیید می‌شوند.

رویکرد تشخیص پیام‌های مشکوک به ترتیب زیر انجام می‌شود:

۱. در میان کانال‌های منتخب در مرحله پیش‌پردازش پیام‌های خام، نشانه‌بندی[۲۲]، ریشه‌یابی[۲۳] و تقلیل[۲۴] می‌شوند.

۲. در میان پیام‌های پیش‌پردازش شده فقط پیام‌های حاوی اسامی یکی از ۴۰۰ رمزارز منتخب یا پیام‌های حاوی کلمات مشکوک از جمله:

pump, moon, target, signal, x100, buy now

گزینش می‌شود.

۳. پیام‌های کاندید به همراه اطلاعات مرتبط از جمله زمان ارسال، اطلاعات کانال و فرستنده برای بررسی در مرحله بعد استفاده می‌شوند.

حال نیاز داریم از میان پیام‌های کاندید با استفاده از ویژگی‌های مبتنی بر بازار و رویکردهای آماری، در بازه انتشار پیام، وقوع رویدادها کلاهبرداری را تأیید یا رد کنیم. اما پیش از آن به تعریف برخی از معیارهای مرسوم آماری در تشخیص ناهنجاری در بازارهای مالی نیاز است. تشخیص پامپ‌ودامپ در بازار رمزارزها معمولاً در قالب مسئله تشخیص ناهنجاری مدل‌سازی می‌شود، زیرا این رویدادها به‌صورت جهش‌های ناگهانی و غیرعادی در قیمت و حجم معاملات بروز می‌کنند. بر این اساس، فرمول ها و شاخص های آماری به کاررفته در این پژوهش با الهام از ادبیات موجود و مناسب با ماهیت پرنوسان بازار رمزارزها انتخاب شده اند [۱۴]. لازم به ذکر است که در فرمول‌های زیر P(t) قیمت دارایی در زمان t و V(t) حجم معاملات را در زمان t نشان می‌دهد.

- بازده تجمعی[۲۵]: برای یک بازه زمانی از t_a تا t_b بازده تجمعی $R_{\{t_a \to t_b\}}$ با استفاده از رابطه ۱ محاسبه می‌شود.

$$R_{\{t_a \to t_b\}} = \frac{P(t_b) - P(t_a)}{P(t_a)} \qquad (۱)$$

این معیار، برای بازه‌های داخل پنجره‌ای به طول ۵ روز محاسبه می‌شود تا اندازه جهش‌های قیمتی[۲۶] را مشخص کرده و در فرآیند تشخیص ناهنجاری در بازار به ما کمک کند.

- فاصله نوسان‌پذیری[۲۷]: انحراف معیار متحرک[۲۸] در زمان t با نماد $\sigma_\tau(t)$ برای بازه کوتاه مدت τ محاسبه شده و میانگین متحرک $\overline{\sigma}_\kappa(t)$ بر اساس بازه بلندتر κ محاسبه می‌شود. سپس فاصله نرمال‌شده نوسان‌پذیری $\Delta\sigma$ به صورت رابطه ۲ تعریف می‌شود.

$$\Delta\sigma(t) = \frac{|\sigma_\tau(t) - \overline{\sigma}_\kappa(t)|}{\overline{\sigma}_\kappa(t) + \epsilon} \qquad (۲)$$

که در آن ϵ یک مقدار کوچک برای جلوگیری از تقسیم بر صفر است. در این رابطه مقدار بزرگ $\Delta\sigma$ نشان‌دهنده یک دوره غیرعادی از نوسان‌پذیری در بازه زمانی اخیر است. این رابطه به ما کمک می‌کند علاوه بر جهش قیمتی نسبت به نوسان‌پذیری بازار نیز اطلاعاتی کسب کنیم. نمونه ای از عملکرد این معیار در شکل ۲ قابل مشاهده است که تغییرات قیمت ارز و نحوه عملکرد شاخص‌هایی مانند میانگین متحرک کوتاه‌مدت و بلندمدت، انحراف

استاندارد (نوسان) و بازده تجمعی را در شناسایی حرکات غیرعادی قیمت به تصویر می‌کشد.

- شاخص روند: برای میانگین‌های متحرک کوتاه‌مدت و بلندمدت در زمان t، به ترتیب $MA_s(t)$ و $MA_l(t)$ (برای مثال ۱۰ دقیقه‌ای و ۵۰ دقیقه‌ای) تعریف می‌شوند. یک شرط ساده برای تشخیص روند صعودی می‌تواند به صورت تعریف شده در رابطه ۳ باشد.

$$trend_up(t) = MA_s(t) > MA_l(t) \qquad (۳)$$

این ویژگی‌ها از جمله مرسوم ترین معیارهای آماری مبتنی بر بازار هستند که در رویکرد ترکیبی برچسب‌گذاری پیشنهادی از آن‌ها استفاده شده است.

حال به توصیف مراحل دقیق رویکرد برچسب‌گذاری می‌پردازیم. به عنوان پیش‌نیازی برای برچسب‌گذاری، مراحل زیر برای هر پیام انجام می‌شود:

۱. تعریف پنجره زمانی: با در نظر گرفتن زمان ارسال پیام کاندید به عنوان t_0 یک بازه ۵ روزه از ۴۸ ساعت قبل تا ۷۲ ساعت بعد تعریف می‌شود که به ترتیب به عنوان t_{start} و t_{end} نمایش داده می‌شوند.

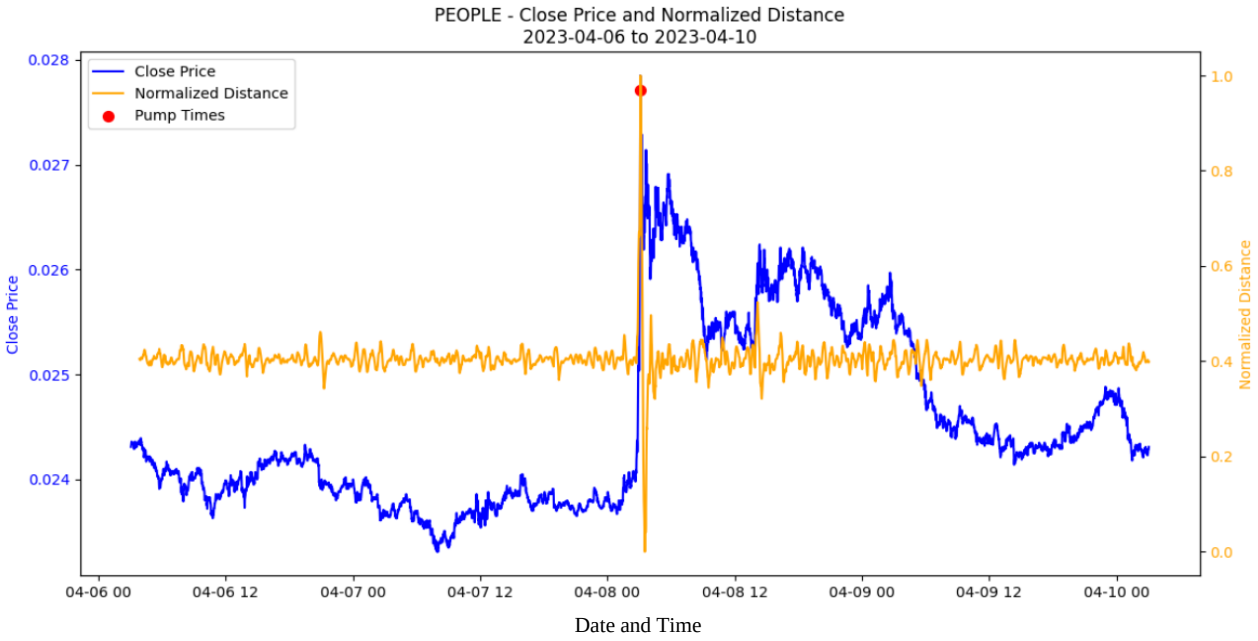
۲. محاسبه مقادیر آماری: پس از استخراج داده‌های مورد نیاز از صرافی بایننس، برای رمزارز c در بازه تعیین شده، مقادیر شاخص‌های آماری برای تشخیص ناهنجاری در بازار را محاسبه می‌کنیم. این مقادیر شامل موارد زیر است:

- قیمت بیشینه P_{max} و زمان وقوع آن t_{max} را در بازه t_0 تا t_{end} پیدا می‌کنیم.
- بازده تجمیعی $R_{\{t_{start}\to t_{max}\}}$ را محاسبه می‌کنیم.
- پیوستگی روند از t_{start} تا t_{max} را محاسبه می‌کنیم.
- بیشینه فاصله نوسان‌پذیری $\Delta\sigma(t)$ را در بازه پنج روزه محاسبه می‌کنیم.

به منظور ارزیابی استحکام رویکرد برچسب‌گذاری پیشنهادی، تحلیل حساسیتی بر روی مقادیر آستانه‌های جایگزین انجام شد. لازم به ذکر است که برای تعیین آستانه‌ها در رویکرد ترکیبی برچسب‌گذاری، ابتدا مقادیر اولیه از مطالعات پیشین [۳و۱۴] استخراج گردید. سپس با انجام جستجوی شبکه‌ای[۲۹] و بهینه‌سازی بر اساس معیار F1، مناسب‌ترین مقادیر متناسب با مجموعه داده این پژوهش مشخص شده و مورد استفاده قرار گرفته‌اند.

۳. اعتبارسنجی رویداد پامپ‌ودامپ: یک پیام فقط در صورتی به عنوان رویداد پامپ پذیرفته می‌شود که شرایط زیر همزمان برقرار باشند:

- بیشینه $\Delta\sigma(t)$ > مقدار آستانه نوسانات معیار
- $R_{\{t_{start}\to t_{max}\}}$ > مقدار بازده معیار
- حفظ شدن پیوستگی روند از t_{start} تا t_{max}
- وقوع قله قیمتی بعد از t_0



شکل ۲: نمونه واقعی از رویداد پامپ‌ودامپ و عملکرد معیار فاصله نوسان‌پذیری در تشخیص رویداد

۴. رفع رویدادهای تکراری: برای جلوگیری از شمارش چندباره، رویدادهایی که کمتر از ۱ ساعت با هم فاصله دارند ادغام شده و به‌عنوان یک رویداد واحد ثبت می‌شوند.

با اجرای این رویکرد، مجموعه‌ای شامل حدود ۱۲۰۰ رویداد پامپ‌ودامپ تأیید شده حاصل شده است که شامل رمزارز هدف، زمان اعلان، زمان اوج، شناسه کانال، پیام اعلان و ویژگی‌های بازار است. شایان ذکر است که در فرآیند برچسب‌گذاری این رویدادها، برای جلوگیری از نشت داده[۳۰]، تمایز دقیقی بین اطلاعات قابل دسترس در لحظه تصمیم‌گیری و اطلاعات مورد استفاده برای ساخت برچسب‌ها اعمال شده است. بدین صورت که مدل در لحظه انتشار پیام (t_0) صرفاً بر اساس ویژگی‌های استخراج‌شده از محتوای پیام و داده‌های بازار تا پیش از آن زمان، اقدام به پیش‌بینی وقوع پامپ در بازه آتی $[t_0, t_0+\Delta t]$ می‌کند. این بدان معناست که پیش‌بینی نهایی به شکل رابطه ۴ تعریف می‌شود:

$$\hat{y} = f(X(t_0)) \qquad (۴)$$

همانطور که در این رابطه مشخص شده اطلاعات قیمتی این بازه پسینی صرفاً برای تعیین برچسب‌های آموزشی به کار رفته و هرگز به عنوان ورودی در اختیار مدل قرار نمی‌گیرد؛ بنابراین مرز میان پیش‌بینی و تشخیص پسینی به طور کامل حفظ شده است.

۳-۳- معماری مدل پیش‌بینی

در این بخش، با دسترسی به رویدادهای برچسب‌گذاری شده پامپ‌ودامپ، به تشریح معماری پیشنهادی برای پیش‌بینی رمزارز هدف در عملیات پامپ‌ودامپ می‌پردازیم. مدل پیشنهادی با نام SNN-TA[۳۱]، یک مدل یادگیری عمیق مبتنی بر سازوکارهای خودتوجه و متقابل‌توجه است که برای استخراج الگوهای تاریخی از دنباله رویدادهای گذشته و ارزیابی الگوهای ارتباطاتی آن‌ها با ویژگی‌های استخراج شده طراحی شده است. در معماری پیشنهادی، توجه موقعیتی برای حفظ اطلاعات ترتیب زمانی نشانه‌ها به‌کار گرفته شده است، اما به‌تنهایی کافی در نظر گرفته نشده است. به همین دلیل، از سازوکار خودتوجه استفاده شده تا هر نشانه بتواند به‌صورت مستقیم با سایر نشانه‌های دنباله تعامل داشته باشد و وابستگی‌های غیرمحلی و غیرخطی میان پیام‌ها را به‌درستی مدل‌سازی شود. افزون بر این، با به‌کارگیری سازوکار متقابل توجه، ارتباط میان دنباله پیام‌ها و ویژگی‌های خارجی مرتبط با ارزها به‌صورت صریح مدل‌سازی شده است. این انتخاب معماری به مدل اجازه می‌دهد تعاملات معنادار میان داده‌های متغیر با زمان و ویژگی‌های ثابت را به‌طور هم‌زمان در نظر بگیرد؛ امری که برای تحلیل رفتارهای دستکاری بازار ضروری است.

برای تشریح عملکرد مدل فرض کنیم دنباله‌ای از رویدادهای قبلی به صورت $E = e_1, e_2, \dots, e_n$ در اختیار داریم. هر رویداد e_i با یک بردار ویژگی x_{e_i} شامل ویژگی‌های مبتنی بر بازار و بازنمایی[۳۲] پیام در قالب متن تعبیه‌شده[۳۳] نشان داده می‌شود. هدف مدل این است که با توجه به دنباله E و بردار ویژگی x_{e_i} برای رمزارز کاندید c، احتمال پامپ بودن آن را با استفاده از رابطه ۵ پیش‌بینی می‌کند:

$$\hat{y} = Probability(c \text{ is target coin} \mid P, x_{e_i}) \qquad (۵)$$

در این راستا معماری پیشنهادی از چهار بخش اصلی تشکیل شده است:

۱. لایه تعبیه[۳۴]: ویژگی‌های عددی و متنی ابتدا نرمال‌سازی شده و سپس به فضای برداری با بعد ثابت d نگاشت می‌شوند. برای پیام‌های متنی از یک مدل زبانی از پیش آموزش‌دیده[۳۵] استفاده می‌شود و بردار خروجی به همان فضای تعبیه منتقل می‌شود.

۲. سازوکار خودتوجه: دنباله رویدادهای پامپ گذشته از طریق لایه‌های چندسر[۳۶] پردازش می‌شود. این لایه روابط بلندمدت بین رویدادها (مثل تمایل به انتخاب متناوب رمزارزهای بزرگ و کوچک در یک کانال) را مدل‌سازی می‌کند و یک دنباله غنی از الگوهای پامپ‌های گذشته گذشته را تولید می‌کند.

۳. سازوکار متقابل‌توجه: با در دست داشتن بازنمایی دنباله به‌دست‌آمده از لایه خودتوجه، رمزارز کاندید c از یک لایه متقابل‌توجه نیز می‌گذرد. در این مرحله مدل یاد می‌گیرد که تا چه اندازه رمزارز کاندید با الگوهای تاریخی رویدادها (E) هماهنگ است.

۴. شبکه چندلایه ادراکی: بردار حاصل از مرحله قبلی به یک شبکه چندلایه ادراکی[۳۷] (MLP) داده می‌شود که پس از اعمال حذف تصادفی[۳۸] و نرمال‌سازی، احتمال نهایی $\hat{y}$ را با استفاده از تابع سیگموید[۳۹] برمی‌گرداند.

این ساختار چندلایه با هدف تشخیص روابط پیچیده در یک مجموعه‌داده چندوجهی طراحی شده و می‌تواند عملکرد خوبی در تشخیص الگوهای تاریخی داشته باشد. استفاده از لایه خودتوجه مدل پیشنهادی را قادر می‌سازد تا بر خلاف مدل‌های بازگشتی، وابستگی‌های بلندمدت را بهتر شناسایی کند. از طرفی لایه متقابل

توجه یک رویکرد اصولی برای سنجش ارتباط میان دنباله و کاندید ارائه می‌دهد، در حالی که مدل‌های پایه مثل SNN [۲۲] تنها از الحاق ساده ویژگی‌ها استفاده می‌کنند. از مزایای مدل پیشنهادی این است که به صورت طبیعی، ترکیبی از اطلاعات مالی و متنی و نیز اطلاعات تکمیلی استخراج شده از کانال‌های ارتباطی را در پیش-بینی لحاظ می‌کند.

با توجه به نوع وظیفه مدل که از نوع طبقه‌بندی[۴۰] است، مدل با تابع هزینه کراس‌انتروپی دودویی[۴۱] آموزش داده می‌شود. برای هر رویداد مثبت، ۳۹۹ رمزارز دیگر به‌عنوان نمونه‌های منفی[۴۲] انتخاب می‌شوند. بهینه‌سازی Adam به عنوان رویکرد بهینه‌سازی استفاده شده و از رویکردهایی مانند حذف تصادفی و پایان سریع[۴۳] برای جلوگیری از بیش‌برازش[۴۴] استفاده می‌شود.

۴- نتایج ارزیابی

این بخش به ارزیابی رویکرد پیشنهادی برای پیش‌بینی رمزارز هدف در رویدادهای پامپ‌ودامپ می‌پردازد. در این راستا، ابتدا پیکربندی آزمایش‌ها شامل آماده‌سازی داده، مدل‌های مبنا و معیارهای ارزیابی را شرح می‌دهیم و سپس نتایج اصلی عملکرد را ارائه کرده و در پایان، تحلیل رفتار مدل و انتخاب‌های طراحی آزمایش را بیان می‌کنیم. پیکربندی آزمایش

مدل پیشنهادی با استفاده از مجموعه‌داده برچسب‌گذاری‌شده حاصل از رویکرد ترکیبی پیشنهادی (بخش ۲-۳) ارزیابی شده است. این مجموعه شامل حدود ۱۲۰۰ رویداد پامپ‌ودامپ تأییدشده در میان حدود ۴۰۰ رمزارز است که در بازه اوت ۲۰۲۲ تا اوت ۲۰۲۴ گردآوری شده‌اند. برای یادگیری و اعتبارسنجی، داده‌ها با نسبت ٪۷۰ یادگیری (اوت ۲۰۲۲ تا نوامبر ۲۰۲۳)، ۱۵٪ اعتبارسنجی (نوامبر ۲۰۲۳ تا آوریل ۲۰۲۴) و ۱۵٪ آزمون (آوریل ۲۰۲۴ تا اوت ۲۰۲۴) و در بازه‌های زمانی پیوسته تقسیم شدند تا از نشت اطلاعات جلوگیری شود. همینطور برای ارزیابی مدل پیشنهادی دو مدل مبنا برای مقایسه استفاده شد:

- مدل معرفی‌شده توسط هو و همکاران [۲۲] با نام SNN: شبکه عصبی دنباله‌ای پیشرفته که دنباله تاریخی رویدادهای کلاهبرداری را مدل‌سازی می‌کند، اما از سازوکار خودتوجه یا متقابل‌توجه استفاده نمی‌کند.
- CLSTM [۲۶]: معماری ترکیبی قدرتمند که شامل لایه‌های پیچشی[۴۵] برای استخراج ویژگی‌های محلی و LSTM برای مدل‌سازی دنباله است که به عنوان نماینده یک مدل رقابتی بدون سازوکار توجه استفاده شده است.

همینطور دو معیار ارزیابی به عنوان معیارهای مرسوم در این حوزه برای ارزیابی رویکردها در نظر گرفته شده است.

- مساحت زیر منحنی (AUC): معیاری مستقل از آستانه برای سنجش عملکرد طبقه‌بندی دودویی که توان تفکیک مدل را نشان می‌دهد.
- نرخ برخورد (HR@K): نسبت رویدادهای تأیید شده در میان K پیش‌بینی برتر مدل را نشان می‌دهد.

$$HR@K = \frac{\#\ of\ events\ in\ top-K}{\#\ of\ total\ events} \qquad (۶)$$

این معیار برای $K \in \{1,3,5,10,20,30\}$ مطابق با کارهای پیشین محاسبه می‌شود تا دقت مدل در شرایط متفاوت آزموده شده و مقایسه شود. همینطور در راستای تسهیل تکرار نتایج در آزمایش‌های دیگر، مقادیر مربوط به مدل و فرآیند یادگیری عمیق که در ادامه در مدل‌های پیشنهادی استفاده شده‌اند به تفصیل در جدول ۱ ارائه شده‌است.

۴-۱- مقایسه عملکرد مدل پیشنهادی

در آزمایش اول، عملکرد مدل پیشنهادی را با مدل‌های معیار در پیش‌بینی رمزارز هدف برای ۱۲۰۰ رویداد منتخب ارزیابی می‌کنیم.

جدول ۱: مقادیر اولیه برای پارامترهای مدل

نماد	مفهوم	مقدار
SD-Window	پنجره لغزان برای محاسبه انحراف معیار	۳۰
MA-Window	پنجره لغزان برای محاسبه میانگین متحرک	۲۰
MA-Short-Window	پنجره لغزان برای محاسبه میانگین متحرک کوتاه مدت	۱۰
MA-Long-Window	پنجره لغزان برای محاسبه میانگین متحرک بلند مدت	۳۰
Volatility-Threshold	آستانه نوسان	۰/۹
Price-Change-Threshold	آستانه تغییر قیمت	۰/۰۳
Trend-Window	پنجره لغزان برای شناسایی روند قیمت	۵
Batch-Size	تعداد نمونه‌های پردازش شده با هم در یک تکرار در طول آموزش	۲۵۶
Epoch	تعداد دفعات پردازش مدل بر کل مجموعه داده آموزشی	۳۰
Embedding-Size	ابعاد نمایش برداری برای داده‌ها	۷۶۸
Learning Rate	اندازه گام برای به روز رسانی وزن مدل در طول بهینه‌سازی	$۸e^{-۵}$
Patience	تعداد ارزیابی‌های متوالی بدون بهبود معیار اعتبارسنجی	۵
Hidden-Size1	تعداد نورون‌ها در اولین لایه پنهان مدل	۱۲۸
Hidden-Size2	تعداد نورون‌ها در دومین لایه پنهان مدل	۶۴

نتایج این آزمایش در راستای مقایسه عملکرد رویکرد پیشنهادی و دو مدل مبنا شامل SNN [۲۲] و CLSTM [۲۶] طراحی شده است. مدل‌سازی غیرتوالی برای این مسئله با استفاده از یک مدل DNN[۴۶] به نتیجه ناامیدکننده ۵۷٪ مساحت زیر نمودار در فرایند یادگیری را به همراه دارد که نشان می‌دهد به‌کارگیری ویژگی‌های ایستا بدون در نظر گرفتن توالی، قدرت پیش‌بینی محدودی برای این مسئله دارد. از این‌رو، همانند پژوهش‌های پیشین در این حوزه، از رویکردهای مبتنی بر مدل‌سازی توالی استفاده شده است. مدل CLSTM به‌عنوان نماینده مدل‌های توالی، عملکردی به‌مراتب بهتر از مدل DNN ارائه می‌دهد که این امر اهمیت در نظر گرفتن پویایی‌های زمانی را تأیید می‌کند. بااین‌حال، عملکرد این مدل همچنان پایین‌تر از مدل SNN است که بر اساس پژوهش [۲۲] به‌عنوان مرسوم‌ترین معماری در این حوزه شناخته می‌شود. برتری مدل SNN نشان می‌دهد که به‌کارگیری سازوکار خودتوجهی در یک معماری توالی‌محور، با تأکید بر گام‌های زمانی حاوی اطلاعات مفید، قدرت تفکیک مدل را در مقایسه با معماری‌های فاقد این سازوکار افزایش می‌دهد. در نهایت، مدل کامل SNNTA با ترکیب هم‌زمان سازوکارهای خودتوجهی، توجه متقابل و ویژگی‌های متنی، به بالاترین عملکرد در پیش‌بینی رمزارز هدف دست می‌یابد. این برتری، ارزش افزوده مؤلفه‌های پیشنهادی را در کنار یکدیگر تأیید کرده و ضرورت به‌کارگیری هم‌زمان آن‌ها برای دستیابی به پیش‌بینی دقیق را نشان می‌دهد. همان‌طور که جدول ۲ نشان می‌دهد، مدل پیشنهادی تقریبا در تمام مقادیر HR@K به‌طور پایدار بهتر از هر دو مدل مبنا عمل کرده است. بیشترین بهبود در $k = 5$مشاهده می‌شود، جایی که مدل ما به مقدار ۰/۶۳۷ دست یافته که معادل ۱۲,۴٪ بهبود مطلق نسبت به مدلSNN است. ارجحیت این معیار نسبت به معیارهای مرسوم‌تر طبقه‌بندی به این دلیل است که انتخاب دقیق رمزارز هدف در میان ۴۰۰ رمزارز بسیار سخت بوده و منجر به نتایج بسیار ضعیف در معیارهایی مانند دقت و صحت می‌شود، اما معیار نرخ برخورد نتایج بهتری برای مقایسه

جدول ۲: مقایسه نتایج نرخ برخورد مدل پیشنهادی با مدل‌های معیار

مدل / معیار	CLSTM	SNN	SNN-TA (پیشنهادی)
HR@1	0.277	0.217	0.281
HR@3	0.414	0.308	0.527
HR@5	0.513	0.484	0.637
HR@10	0.623	0.651	0.726
HR@20	0.739	0.762	0.815
HR@30	0.823	0.893	0.863

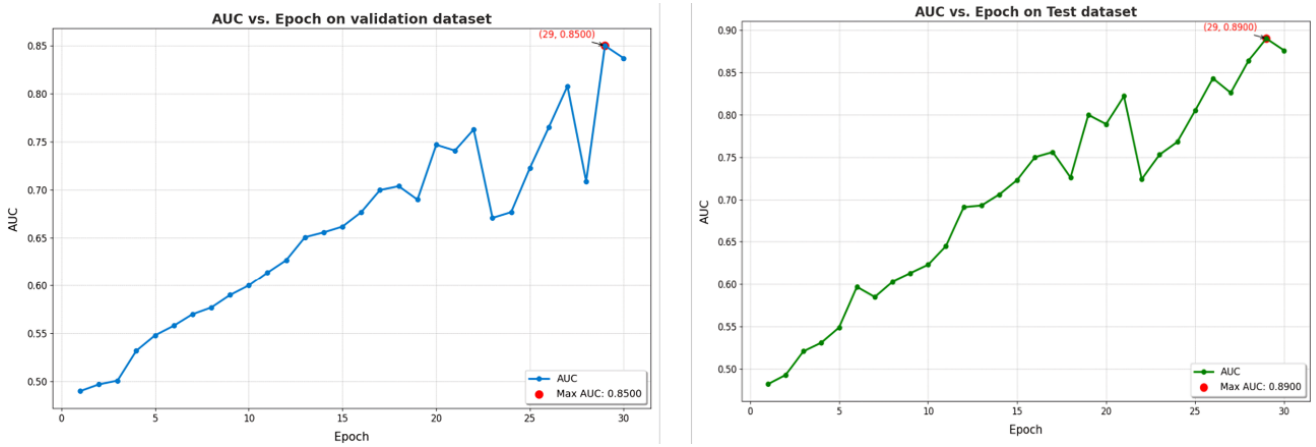
روش‌های مختلف ارائه می‌کند. معمولاً کاربران سیستم‌های تشخیص یا پیش‌بینی کلاهبرداری با استفاده از نتایج برتر پیش‌بینی شده نیاز به دخالت انسانی در تشخیص نهایی رمزارز دارند. در نتیجه عملکرد چشم‌گیر مدل پیشنهادی در نرخ برخورد ۵تایی اهمیت کاربردی ویژه‌ای برای کاربران نهایی دارد، زیرا تحلیل‌گر یا متخصص مربوطه با بررسی چند گزینه برتر پیش‌بینی شده نیز می‌تواند نتایج مطلوبی را دریافت کند. همچنین در مقادیر $K = 10$ و $K = 20$ نیز برتری مدل ما تداوم دارد که بیانگر پایداری عملکرد در مجموعه‌های کاندید گسترده‌تر است.

۴-۲- توان تفکیک‌پذیری (تحلیل مساحت زیر نمودار)

یکی از اساسی‌ترین مشکلات پیش‌بینی رمزارز در این حوزه، محدودیت گنجایش مدل‌های یادگیری است که خود را در قالب بیش‌برازش و کم‌برازش[۴۷] در زمان یادگیری نمایان می‌کند. عموماً برای تشخیص این نوع از محدودیت‌ها در مدل‌های یادگیری ماشین، تحلیل نمودارهای مساحت زیر نمودار در فرآیند یادگیری و اعتبارسنجی انجام شده و اصلاح می‌شود. در همین راستا، توان تفکیک کلی مدل با استفاده از مقدار مساحت زیر نمودار روی مجموعه آزمون و اعتبارسنجی در شکل ۳ قابل مشاهده است. با استفاده از شاخص مساحت زیر منحنی در آزمایشات بررسی شده مدل پیشنهادی روی مجموعه‌داده یادگیری به مقدار بیشینه مساحت زیر نمودار برابر ۰/۸۹دست یافته است که بیانگر توان بالای آن در تفکیک رمزارزهای هدف واقعی از موارد غیرهدف است.

همینطور مقدار بیشینه مساحت زیر نمودار برای مجموعه‌داده اعتبارسنجی معادل با ۰/۸۵ بوده که برحسب انتظار مقدار کمتری نسبت به مجموعه‌داده یادگیری ارائه می‌کند. همینطور شباهت منحنی در مراحل اعتبارسنجی[۴۸] و یادگیری نشان از عدم بیش‌برازش مدل دارد؛ این موضوع بیانگر توان تعمیم‌پذیری مناسب مدل بر داده‌های خارج از مجموعه‌داده یادگیری است. این مشاهدات در نمودار مربوط به اعتبارسنجی نیز تکرار شده و تا حد قابل قبولی نتایج مشابه با نمودار یادگیری را ارائه می‌کند. در ادامه با بررسی تأثیر استفاده از سازوکار منظم‌ساز در بهبود عملکرد مدل، به معیارها و فرآیند انتخاب مدل نهایی می‌پردازیم. منظم‌سازی و انتخاب مدل

با توجه به وجود موارد ناهنجار در رویدادهای کلاهبرداری، نیازمند استفاده از سازوکار منظم‌سازی[۴۹] برای کاهش تأثیر ناهنجاری‌ها در فرآیند یادگیری مدل هستیم. برای تحلیل نقش این انتخاب در عملکرد کلی مدل، یک آزمون فرسایشی[۵۰] با هدف اندازه‌گیری تأثیر

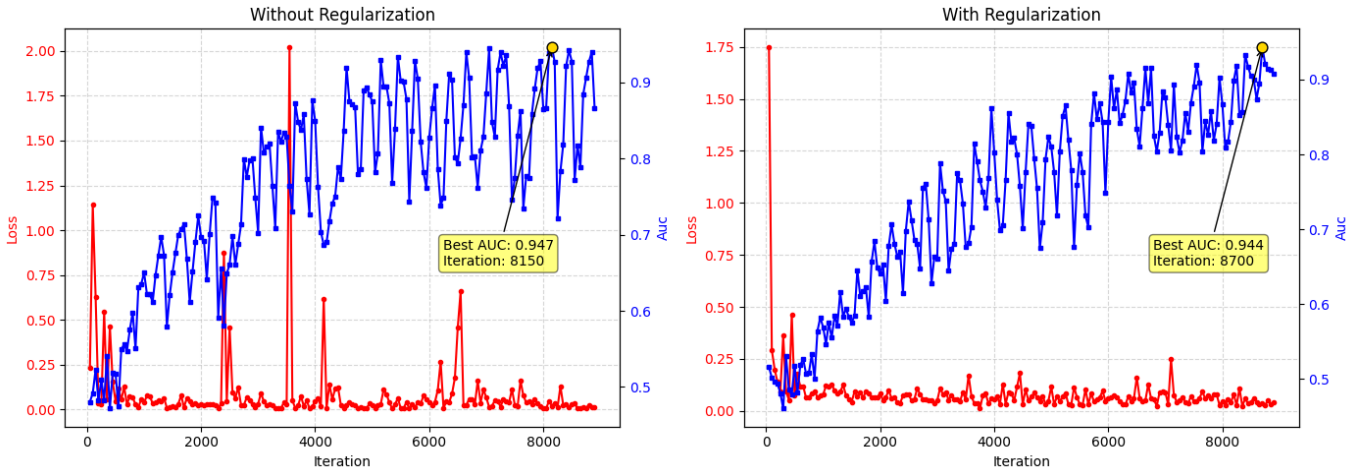


شکل ۳: عملکرد مدل در مراحل اعتبارسنجی و یادگیری از دیدگاه مساحت زیر نمودار

سازوکار منظم‌سازی انجام شده است. در این آزمایش، نتایج یادگیری و آزمون را با و بدون استفاده از منظم‌ساز L2 مقایسه می‌کنیم. بدون منظم‌ساز، خطای یادگیری نوسانات شدیدی داشته و مقدار AUC آزمون به‌طور محسوسی کاهش می‌یابد که نشانه بیش‌برازش است. در مقابل، افزودن منظم‌ساز L2 باعث پایداری فرآیند یادگیری، همگرایی نرم‌تر و بهبود AUC نهایی شده است. این موضوع اهمیت استفاده از منظم‌سازی برای استخراج الگوهای قابل تعمیم از داده‌های متنی و مالی را نشان می‌دهد. نتایج آزمون فرسایشی منظم‌سازی در شکل ۴ نشان داده شده است.

برای اطمینان از ارزیابی دقیق و قابل اعتماد، رویکرد ذخیره و انتخاب نقطه‌ذخیره‌سازی[۵۱] در طول آموزش به‌کار گرفته شده است. در این رویکرد، مدل در دوره‌های مختلف ذخیره و بر روی مجموعه اعتبارسنجی ارزیابی شده است. مدلی که بالاترین مقدار AUC در اعتبارسنجی را داشته برای ارزیابی نهایی روی مجموعه آزمون انتخاب گردیده است. در مجموع، نتایج ارائه‌شده شواهد تجربی محکمی برای کارایی رویکرد پیشنهادی فراهم می‌کند، بطوریکه مدل در تمام معیارها، چه معیار کلی مساحت زیر نمودار و چه معیارهای عملیاتی مانند نرخ برخورد، عملکردی بهتر یا حداقل معادل با مدل‌های معیار دارد. استفاده از مکانیزم‌های توجه موجب مدل‌سازی بهتر وابستگی‌های بلندمدت در دنباله‌های پامپ و یکپارچه‌سازی مؤثرتر ویژگی‌های کاندید شده است.

منظم‌سازی و نقطه‌ذخیره‌سازی به‌عنوان انتخاب‌های طراحی، نقش کلیدی در پایداری و قابلیت تعمیم‌پذیری مدل داشته‌اند. این یافته‌ها فرضیه اصلی پژوهش را تأیید می‌کنند که تأکید داشت: مدل‌های دنباله‌محور مبتنی بر توجه، همراه با فرآیند برچسب‌گذاری ترکیبی (متنی-آماری)، ابزار قدرتمند و عملی برایتشخیص و پیش‌بینی کلاهبرداری پامپ‌ودامپ در بازار رمزارزها را فراهم می‌آورند.



(الف) (ب)

شکل ۴: مقایسه عملکرد مدل در آزمون فرسایشی با استفاده از سازوکار منظم‌سازی:
(الف) با استفاده از منظم‌سازی – (ب) بدون استفاده از منظم‌سازی

۵- نتیجه‌گیری و کارهای آتی

در این مقاله، یک رویکرد نوآورانه برای پیش‌بینی رمزارز هدف در رویدادهای آتی پامپ‌ودامپ ارائه گردید. این رویکرد با ترکیب یک رویکرد ترکیبی برچسب‌گذاری که از سیگنال‌های متنی تلگرام و معیارهای آماری ناهنجاری بازار استفاده می‌کند و یک معماری پیش‌بینی دنباله محور مبتنی بر سازوکارهای خودتوجه و متقابل‌توجه ارائه می‌کند، توانست محدودیت‌های رویکردهای موجود از جمله استفاده از داده‌های چندوجهی (متنی-عددی) را برطرف سازد. مدل پیشنهادی این پژوهش (SNN-TA) با بهره‌گیری هم‌زمان از دو سازوکار کلیدی، گامی فراتر از روش‌های توالی‌محور صرف برمی‌دارد: نخست، استفاده از ترکیب خودتوجهی و توجه متقابل که امکان پردازش کارآمد توالی‌های بلند داده‌ها را بدون محدودیت‌های محاسباتی رایج فراهم می‌کند؛ دوم، به‌کارگیری توجه موقعیتی نسبی که انعطاف‌پذیری مدل در مواجهه با دنباله‌های با طول متغیر را افزایش می‌دهد. این ترکیب منحصربه‌فرد، مدل پیشنهادی را قادر می‌سازد تا هم از نظر محاسباتی مقیاس‌پذیرتر عمل کند و هم درک عمیق‌تری از وابستگی‌های بلندمدت در داده‌های مالی و متنی داشته باشد. این در حالی است که روش‌های پیشین مانند SNN عمدتاً به یکی از این جنبه‌ها توجه داشته و در مواجهه هم‌زمان با چالش‌های مقیاس‌پذیری و پیچیدگی داده‌های چندوجهی با محدودیت مواجه هستند. ارزیابی‌ها بر روی حدود ۱۲۰۰ رویداد پامپ تأییدشده و نزدیک به ۴۰۰ رمزارز نشان داد که مدل پیشنهادی در مقایسه با مدل‌های معیار، از جمله SNN [۲۲] و CLSTM [۲۶]، بهبود قابل‌توجهی در معیارهای مساحت زیر نمودار و نرخ برخورد ایجاد می‌کند. به‌ویژه، این مدل توانست عملکرد و پایداری خود را در طیف وسیعی از بازه‌های زمانی حفظ کند که بیانگر قابلیت کاربردی آن در شرایط واقعی بازار است. اگرچه نتایج به‌دست‌آمده نشان‌دهنده موفقیت رویکرد پیشنهادی است، فرصت‌های متعددی برای توسعه این حوزه وجود دارد. در این راستا می‌توان به موارد زیر اشاره کرد:

- افزودن منابع داده جدید: ترکیب داده‌های لحظه‌ای از پلتفرم‌هایی مانند X، دیسکورد و فروم‌های تخصصی می‌تواند پوشش بهتری از فعالیت‌های هماهنگ ایجاد کند.
- یادگیری چندوظیفه‌ای[۵۲]: آموزش همزمان مدل برای پیش‌بینی زمان، مدتِ زمان و رمزارز هدف می‌تواند دقت پیش‌بینی‌ها را بهبود بخشد.
- ارزیابی در بازارهای غیرمتمرکز: آزمایش مدل در صرافی‌های غیرمتمرکز که ساختار سفارش و نقدینگی متفاوتی دارند، می‌تواند دامنه کاربرد آن را گسترش دهد.
- کاربرد در سیستم‌های هشدار بی‌درنگ: ادغام مدل در پلتفرم‌های مانیتورینگ بازار برای ارائه اعلان‌های فوری به سرمایه‌گذاران و نهادهای نظارتی

این مسیرها می‌توانند به تقویت توان مدل در پیش‌بینی دقیق‌تر و زودهنگام، افزایش مقاومت در برابر انواع جدید دستکاری بازار، و گسترش دامنه عملیاتی آن در محیط‌های متنوع منجر شوند. اگرچه چالش‌های پیش‌روی این حوزه همچنان پابرجاست، اما رویکرد پیشنهادی با نشان‌دادن پتانسیل بالای یادگیری عمیق در کشف الگوهای پیچیده کلاهبرداری، مسیر جدیدی را برای پژوهش‌های آینده و توسعه سیستم‌های نظارتی هوشمند گشوده است. ادامه کار در مسیرهای پژوهشی پیشنهادی می‌تواند به ایجاد زیست بومی امن‌تر و شفاف‌تر برای سرمایه‌گذاران در فضای پرنوسان رمزارزها بینجامد.

مراجع

[1] K. Mishev, A. Gjorgjevikj, I. Vodenska and L. T. Chitkushev, "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," *IEEE Access*, vol. 8, no. 1, pp. 131662–131682, 2020. https://doi.org/10.1109/ACCESS.2020.3009626

[2] M. M. Patel, S. Tanwar, R. Gupta and N. Kumar, "A Deep Learning-based Cryptocurrency Price Prediction Scheme for Financial Institutions," *Journal of Information Security and Applications*, vol. 55, no. 1, p. 102583, 2020. https://doi.org/10.1016/j.jisa.2020.102583

[3] A. M. Khedr, I. Arif, P. R. P V and M. El-Bannany, "Cryptocurrency Price Prediction Using Traditional Statistical and Machine-Learning Techniques: A survey," *Intelligent Systems in Accounting, Finance and Management*, vol. 28, no. 1, pp. 3-34, 2021. https://doi.org/10.1002/isaf.1488

[4] L. Abednego, "Forex Trading Robot with Technical and Fundamental Analysis," *Journal of Computers*, vol. 13, no. 9, pp. 1089-1097, 2018. https://doi.org/10.17706/jcp.13.9.1089-1097

[5] Nam, D., and D. B. Skillicorn. "Detecting Pump &Dump Stock Market Manipulation from Online Forums." *Digital Finance*, vol. 7, no. 1, pp. 1–20, 2025. https://doi.org/10.1007/s42521-024-00121-4

[6] L. Charfeddine and A. Mahrous, "What Drives Cryptocurrency Pump and Dump Schemes: Coin Versus Market Factors*?," Finance Research Letters*, vol. 67, no. 1, p. 105861, 2024. https://doi.org/10.1016/j.frl.2024.105861

[7] M. Mirtaheri, S. Abu-El-Haija, F. Morstatter, G. V. Steeg and A. Galstyan, "Identifying and Analyzing Cryptocurrency Manipulations in Social Media," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 3, pp. 607–617, 2021. https://doi.org/10.1109/TCSS.2021.3059286

[8] A. Kalacheva, P. Kuznetsov and Y. Yanovich, "Detecting Rug Pulls in Decentralized Exchanges: The Rise of Meme Coins," *Blockchain: Research and Applications*, no. 1, pp. 50-62, 2024. https://doi.org/10.1016/j.bcra.2025.100336

[۹] ف. حسینی، ن. نوروزی «استفاده از روشهای یادگیری ماشین جهت پیش بینی بازده سهام با در نظر گرفتن تاثیر تغییر دامنه نوسان قیمت»، مجله علمی رایانش نرم و فناوری اطلاعات، جلد ۱۴، شماره ۲، تابستان ۱۴۰۴، صفحه ۵۹-۴۵. https://doi.org/10.22034/jscit.2025.506154.2107

[10] V. Chadalapaka, K. Chang, G. Mahajan and A. Vasil, "Crypto Pump and Dump Detection via Deep Learning Techniques," *arXiv*, no. 1, 2022. https://doi.org/10.48550/arXiv.2205.04646

[11] H. Nghiem, G. Muric, F. Morstatter and E. Ferrara, "Detecting Cryptocurrency Pump and-Dump Frauds Using Market and Social Signals," *Expert Systems with Applications*, vol. 182, no. 1, p. 115284, 2021. https://doi.org/10.1016/j.eswa.2021.115284

[12] M. L. Morgia, A. Mei, F. Sassi and J. Stefa, "The Doge of Wall Street: Analysis and Detection of Pump and Dump Cryptocurrency Manipulations," *ACM Transactions on Internet Technology*, vol. 23, no. 11, pp. 1 - 28, 2022. https://doi.org/10.1145/3561300

[13] D. Ardia and K. Bluteau, "Twitter and Cryptocurrency Pump-and-Dumps," *International Review of Financial Analysis*, vol. 95, no. 1057-5219, pp. 103479-103479, 2024. https://doi.org/10.1016/j.irfa.2024.103479

[14] J. Kamps and B. Kleinberg, "To The Moon: Defining and Detecting Cryptocurrency Pump and-Dumps," *Crime Science*, vol. 7, no. 1, pp. 1-18, 2018. https://doi.org/10.1186/s40163-018-0093-5

[15] L. Wen, "Market Manipulation of Cryptocurrencies: Evidence from Social Media and Transaction Data," *ACM Transactions on Internet Technology*, vol. 24, no. 2, pp. 1-26, 2024. https://doi.org/10.1145/3643812

[16] Fratrič, P., Sileno, G., Klous, S. et al." Manipulation of the Bitcoin Market: An Agent-Based Study," *Financial Innovation*, vol. 8, no. 60, 2022. https://doi.org/10.1186/s40854-022-00364-3

[17] C. Wu *et al*., "Profit or Deceit? Mitigating Pump and Dump in DeFi via Graph and Contrastive Learning," in *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 8994-9008, 2025. https://doi.org/10.1109/TIFS.2025.3594873

[18] J. Xu, H. Wu, J. Wang and M. Long, "Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy," *arXiv*, 2021. https://doi.org/10.48550/arxiv.2110.02642

[19] A. Roy, et al. "GAD-NR: Graph Anomaly Detection via Neighborhood Reconstruction." *ACM International Conference on Web Search and Data Mining*, pp. 576–585, Merida, Mexico, 19 Mar. 2024. https://doi.org/10.1145/3616855.3635767

[20] J. Clough and M. Edwards, "Pump, Dump, and then What? The Long-Term Impact of Cryptocurrency Pump-and-Dump Schemes," in 2023 *APWG Symposium on Electronic Crime Research*, pp. 1-17, Barcelona, Spain, 1 Apr. 2024. https://doi.org/10.1109/ecrime61234.2023.10485505

[21] J. Xu and B. Livshits, "The Anatomy of a Cryptocurrency Pump-and-Dump Scheme," in 28th USENIX Security Symposium, pp. 1609—1625, Santa Clare, CA, USA, 14 Aug. 2019. https://www.usenix.org/conference/usenixsecurity19/presentation/xu-jiahua

[22] S. Hu, Z. Zhang, S. Lu, B. He and Z. Li, "Sequence-Based Target Coin Prediction for Cryptocurrency Pump-and-Dump," *Proceedings of the ACM on Management of Data*, vol. 1, no. 1, pp. 1-19, 2023. https://doi.org/10.1145/3588686

[23] Bolz, Manuel, et al. "Real-Time Machine Learning Detection of Telegram-Based Pump-And-Dump Schemes." *Proceedings of the 2025 Workshop on Decentralized Finance and Security*, 13 Oct. 2025, pp. 1–15. https://doi.org/10.1145/3733815.3764042

[24] Lukasik, Michal, et al. "Text Segmentation by Cross Segment Attention." *ArXiv*, pp. 4707-4716, 1 Jan. 2020, https://doi.org/10.18653/v1/2020.emnlp-main.380

[25] Wang, Yu-An, and Yun-Nung Chen. "What Do Position Embeddings Learn? An Empirical Study of Pre-Trained Language Model Positional Encoding." *ArXiv*, pp. 6840-6849, 11 June 2024. https://doi.org/10.18653/v1/2020.emnlp-main.555

[26] X. Shi, et al, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, vol. 28, pp. 1, 2015. https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html

پاورقی‌ها:

[1] Pump and Dump
[2] Retrospective
[3] Attention
[4] Labeling Method
[5] Self-Attention
[6] Cross-Attention
[7] Area Under Curve
[8] Hit Ratio
[9] Unsupervised
[10] Order Book
[11] Random Forest
[12] Sequence Neural Networks
[13] Ground Truth
[14] Multimodal
[15] Sequence-Aware
[16] Application Program Interface
[17] Candle Stick
[18] Binance
[19] Reddit
[20] Total Market Cap
[21] Preprocessing
[22] Tokenized
[23] Stemmed
[24] Lemmatized
[25] Cumulative Return
[26] Price Spike
[27] Volatility Distance
[28] Rolling Standard Deviation
[29] Grid Search
[30] Data Leakage
[31] Sequence Neural Networks with Target Attention
[32] Representation
[33] Text Embedding
[34] Embedding Layer
[35] Pretrained Sentence Encoder
[36] Multi-Head
[37] Multi-Layer Perceptron
[38] Dropout
[39] Sigmoid
[40] Classification
[41] Binary Cross Entropy
[42] Negative Sampling
[43] Early Stopping

[44] Overfitting
[45] Convolution
[46] Deep Neural Network
[47] Underfitting
[48] Validation
[49] Kernel Regularization
[50] Ablation Study
[51] Checkpoint
[52] Multi-Task Learning