E-ISSN: 2717-4581 | P-ISSN: 2783-1345

Innovation Management and Operational Strategies

www.journal-imos.ir

Inov. Manage. Oper. Stra. Vol. 6, No. 4 (2025) 468–484.

Paper Type: Original Article

Pairs Trading with Machine Learning in Tehran Stock Exchange

Samira Asadian[1], Dariush Damouri[1,*], Hojatollah Sadeghi[1]

[1]Department of Accounting, Faculty of Economics, Management and Accounting, Yazd University, Yazd, Iran; asadian.samira@gmail.com; d.damoori@yazd.ac.ir; sadeqi@yazd.ac.ir.

Citation:

Asadian, S., Damouri, D., & Sadeghi, H. (2025). Pairs trading with machine learning in Tehran Stock Exchange. *Innovation management and operational strategies, 6*(2), 468–484.

Received: 23/05/2025 Reviewed: 12/07/2025 Revised: 17/08/2025 Accepted: 03/09/2025

Abstract

Purpose: A well-structured strategy can enhance trading efficiency. The interest of market participants in quantitative trading models has led to the increased use of statistical arbitrage strategies, including pairs trading, which involves two steps: identifying two securities (a pair) and discovering an anomaly in the price gap between two components of the pair (taking a trade).

Methodology: Machine learning can help define a search space by clustering correlated securities and then selecting the most promising pairs for trading opportunities. However, after finding a price gap in pairs, there may be divergence even after opening a position in the market, and the investor may witness a decrease in the value of their portfolio. To mitigate this risk, a time series forecasting-based trading model can be implemented to enhance returns while minimizing periods of portfolio drawdown.

Findings: In clustering with the OPTICS model, there are fewer trades (4) in exchange for identifying a higher percentage of profitable pairs (100) with a higher Sharpe ratio (10.25). Using deep learning and an LSTM model to provide trading signals, results indicate that this model has not performed very successfully in the Iranian capital market compared to the standard basic model.

Originality/Value: 5-minute data frequency, net returns including trading costs, and the use of machine learning for both components of strategy are some of the innovative features of this article.

Keywords: Pairs trading strategy, Clustering, Trading costs, Machine learning, LSTM.

Corresponding Author: d.damoori@yazd.ac.ir

10.22105/imos.2025.531130.1494

Licensee. **Innovation Managemrnt & Operational Strategeis.** This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

مدیریت نوآوری و راهبردهای عملیاتی
دوره ۶، شماره (۳)، (۱۴۰۴)، ۴۸۴-۴۶۸
www.journal-imos.ir

نوع مقاله: پژوهشی

استراتژی معاملات زوجی با یادگیری ماشینی در بورس اوراق بهادار تهران

سمیرا اسدیان[۱]، داریوش دموری[۱،*]، حجت الله صادقی[۱]
[۱] گروه حسابداری، دانشکده اقتصاد، مدیریت و حسابداری، دانشگاه یزد، یزد، ایران.

چکیده

هدف: جهت استفاده از کاربردهای هوشمندانه مالی در فرآیند طاقت‌فرسای شناسایی اوراق بهادار از نقطه‌نظر ارزش‌گذاری در تصمیم‌گیری معاملات با توجه به ایده قیمت‌گذاری نسبی، داشتن یک استراتژی قوی نسبت به شرایط مختلف بازار سرمایه پیشنهاد می‌شود. علاقه مشارکت‌کنندگان بازار به مدل‌های معاملاتی کمی، منجر به افزایش استفاده از استراتژی‌های آربیتراژ آماری، از جمله معاملات زوجی شد که طی دو مرحله، شناسایی دو ورقه بهادار (یک زوج) و کشف بی‌نظمی در شکاف قیمتی بین آن‌ها (انجام معامله) است.

روش‌شناسی پژوهش: از یادگیری ماشینی می‌توان برای تعیین فضای جست‌وجو جهت گروه‌بندی اوراق بهادار مرتبط در خوشه‌ها و شناسایی زوج‌های برتر درون آن‌ها استفاده نمود. از طرفی پس از یافتن شکاف قیمتی زوج‌ها، امکان دارد حتی پس از باز کردن یک موقعیت در بازار، واگرایی وجود داشته باشد و احتمالا سرمایه‌گذار شاهد کاهش ارزش سبدش باشد. در این مورد می‌توان از یک مدل معاملاتی مبتنی بر پیش‌بینی سری‌زمانی استفاده نمود تا ضمن افزایش بازده، تعداد روزهای کاهش ارزش سبد کاهش یابد.

یافته‌ها: در خوشه‌بندی با استفاده از مدل *OPTICS*، شاهد تعداد معاملات کم‌تری (۴) در ازای شناسایی درصد بالاتر زوج‌های سودآور (۱۰۰) همراه با نسبت شارپ بالاتر (۱۰/۲۵) بودیم. با استفاده از یادگیری عمیق و مدل *LSTM* در ارایه سیگنال‌های معاملاتی، نتایج خاطرنشان می‌سازند این مدل در مقایسه با مدل استاندارد پایه، در بازار سرمایه ایران چندان موفق عمل نکرده است.

اصالت/ارزش‌افزوده علمی: تواتر زمانی ۵دقیقه‌ای داده‌ها، بازده‌های خالص با احتساب هزینه‌های معاملاتی و استفاده از یادگیری ماشینی برای هر دو مرحله این استراتژی، از موارد نوآورانه این مقاله است.

کلیدواژه‌ها: استراتژی معامله زوجی، خوشه‌بندی، هزینه‌های معاملاتی، یادگیری ماشینی، *LSTM*.

۱- مقدمه

سری‌های قیمتی ایستا خصوصا برای معامله‌گران مالی جذاب هستند، زیرا زمانی‌که قیمت یک ورقه بهادار به‌صورت قابل‌توجهی از میانگین تاریخی خود انحراف دارد. با توجه به این مهم که قیمت (به میانگین خود) باز خواهد گشت، از ویژگی ایستایی هنگام ارسال سفارش، منفعت کسب می‌کنند. به‌هرحال این مورد به‌ندرت در سری‌های زمانی مالی یافت می‌شود. در حقیقت، سری‌های قیمتی اغلب فاقد ایستایی هستند و ازاین‌رو، در تئوری کاملا غیرقابل‌پیش‌بینی می‌باشند.

از سویی، علاقه مشارکت‌کنندگان بازار به مدل‌های معاملاتی کمی، به‌دلیل توسعه قابل‌توجه فناوری رایانه در دهه ۱۹۸۰، منجر به افزایش استفاده از استراتژی‌های آربیتراژ آماری، خصوصا میان صاحبان استراتژی صندوق‌های پوشش ریسک[۱] و معامله‌گران دارایی در شرکت‌های تامین سرمایه شد

[1] Hedge funds

[1]. معاملات زوجی یک استراتژی توسعه داده‌شده از سوی گروهی از ریاضی‌دانان، فیزیک‌دانان و مهندسین کامپیوتر بوده که توجه خاصی را در اوایل دهه مذکور دریافت کرده است [2]. ایده پشت این استراتژی، کسب مزیت از عدم کارایی بازار با یک قاعده معاملاتی آسان طی دو مرحله است:

۱. ابتدا، نیازمند شناسایی دو ورقه بهادار (برای نمونه دو سهم یا یک زوج) هستیم که سری قیمتی آنان، رفتار مشابهی را نشان می‌دهد. به‌محض شناسایی زوج‌ها، سرمایه‌گذار می‌تواند به گام دوم استراتژی رود.
۲. سپس، فرض اساسی این است که سری قیمتی دو ورقه بهادار در گذشته، حرکتی نزدیک به هم داشتند ازاین‌رو در آینده نیز این حرکت استمرار خواهد داشت. پس چنان‌چه بی‌نظمی‌ای رخ دهد یک موقعیت معاملاتی جذاب برای سود بردن از این اصلاح فراهم می‌آورد. برای یافتن چنین فرصت‌هایی، شکاف قیمتی بین دو جزو تشکیل‌دهنده زوج‌ها باید به‌صورت مداوم پایش شوند. زمانی‌که یک بی‌قاعدگی آماری شناسایی شد، ورود به بازار رخ می‌دهد که به‌محض اصلاح احتمالی شکاف قیمتی باید از بازار خارج شد [3].

با این‌وجود به‌دلیل تعداد زیاد داده در دسترس، امکان انتخاب زوج‌های برتر و بهترین روش معاملاتی اندکی پیچیده شده و دیگر با صرف استفاده از رویکردهای سنتی شناسایی ازجمله رویکرد فاصله و هم‌انباشتگی قادر اجرای موفقیت‌آمیز استراتژی نخواهیم بود؛ درنتیجه استفاده از روش‌های یادگیری ماشینی برای هر دو مرحله استراتژی معامله زوجی پیشنهاد می‌شود که شامل موارد زیر است:

۱. استفاده از الگوریتم یادگیری نظارت‌نشده مرتب‌سازی نقاط جهت شناسایی ساختار خوشه‌بندی[1] برای خوشه‌بندی اوراق بهادار و انتخاب زوج‌ها که در این صورت دیگر نیازی به تعیین تعداد خوشه‌ها از پیش نمی‌باشد و برای خوشه‌های با چگالی متغیر مناسب می‌باشد.
۲. استفاده از یادگیری عمیق جهت یادگیری خودکار ماشین برای شناسایی بهتر سیگنال‌های معاملاتی با استفاده از مدل حافظه کوتاه‌مدت طولانی[2] که مناسب یادگیری وابستگی‌های بلندمدت متوالی بوده و به‌منظور پیش‌بینی در مرحله معاملات مورداستفاده قرار می‌گیرد.

با توجه به ماهیت پیچیده و ارتباط متقابل اجزا در بازارهای مالی و حجم بالای تعاملات در مقایسه با مدل‌های سنتی پردازش داده که مایه زحمت هستند، ضروری است با کاربردهای هوشمندانه داده دارای یک استراتژی قوی نسبت به شرایط مختلف بازار سرمایه، صرف‌نظر از جهت آن (صعودی، نزولی یا بدون جهت) باشیم [3]. همچنین با توجه به عمق گرفتن بازار سرمایه ایران و روند افزایشی تعداد شرکت‌های پذیرفته‌شده در بازار و حجم سرمایه تحت مدیریت افراد (چه حقیقی چه حقوقی) دیگر نمی‌شود بدون داشتن سیستم معاملاتی و استراتژی سرمایه‌گذاری به بازده خوبی رسید. ازاین‌رو ایجاد یک سیستم معاملاتی که موقعیت‌های معاملاتی را به‌درستی تشخیص دهد، برای مدیریت فعال سبدهای اوراق بهادار ضروری است.

استفاده از یادگیری ماشینی در هر دو مرحله این استراتژی به همراه مدل مبتنی بر پیش‌بینی و مدنظر کل بازار جهت شناسایی زوج‌ها و لحاظ نمودن هزینه‌های معاملاتی وجه تمایز این کار است. همچنین تواتر زمانی بین‌روزی شامل داده‌های ۵ دقیقه‌ای به عمق کار می‌افزاید. ادامه این مقاله بدین‌صورت است: بخش دوم به ادبیات موضوع معاملات زوجی و مدل یادگیری عمیق، بخش سوم به روش پژوهش، بخش چهارم به تحلیل داده‌ها، بخش پنجم به بحث و دید کلی، بخش ششم به پیشنهادها و آخرین بخش به ذکر منابع می‌پردازد.

۲- ادبیات موضوع

تجزیه‌وتحلیل معاملات زوجی، اولین بار از سوی [1] معرفی شد و به‌دنبال آن از سوی بسیاری [2]، [4–7] پیگیری شد. روش معامله زوجی، تکنیک مهم آربیتراژ آماری و یکی از مشهورترین استراتژی‌های سفته‌بازی کوتاه‌مدت است که به‌دلیل رویکرد آسان و کاربرد ساده آن، به‌صورت گسترده‌ای به‌عنوان استراتژی خنثی در بازار به‌کار گرفته می‌شود [8].

قاعده معاملاتی استراتژی معاملات زوجی آسان است: مدنظر قرار دادن دو ورقه بهادار که قیمت‌های آن‌ها باهم حرکت می‌کنند، دقت به زیاد شدن شکاف قیمتی (دوره تشکیل) و سپس خرید ورقه بهادار با قیمت نسبتا پایین‌تر و فروش ورقه بهادار با قیمت نسبتا بالاتر (دوره معاملاتی). چنان‌چه اوراق بهادار به الگوی شکاف قیمتی تاریخی خود همگرا شدند، معامله به سود منتهی می‌شود [8]. تلاش قابل‌ملاحظه‌ای از سوی دانشگاهیان و

[1] Ordering Points To Identify the Clustering Structure (OPTICS)
[2] Long Short-Term Memory (LSTM)

فعالین مالی در خصوص شناسایی و بهره‌برداری از ناهنجاری بازار صورت گرفته تا استراتژی‌های معاملاتی ارزش نسبی مربوط به زوج‌ها بر اساس بازده ایجاد شوند (برای نمونه [1]، [2]، [12–9] را ببینید).

در دوره تشکیل اگر سرمایه‌گذار هیچ محدودیتی بر فضای جست‌وجو تحمیل نکند ممکن است مجبور شود ترکیبات اضافی کشف کند و احتمالا روابط جعلی بیابد. با توجه به این‌که یادگیری ماشینی در حال تبدیل به ابزار بسیار بااهمیتی در کسب‌وکار و در ارتباط با استفاده از مجموعه داده‌های بزرگ می‌باشد تا ارتباط بین متغیرها را یاد بگیرد، پیش‌بینی کند و در یک محیط متغیر اخذ تصمیم نماید [13]، بااستفاده از الگوریتم یادگیری نظارت‌نشده می‌توان خوشه‌های معناداری از دارایی‌ها برای انتخاب زوج‌ها یافت. انگیزه این کار برمی‌گردد به این‌که اجازه دهیم داده‌ها به‌وضوح خود را نشان دهند به‌جای این‌که به‌صورت دستی گروه‌هایی برای هر ورقه بهادار تعریف کنیم که به آن تعلق داشته باشند [14]. برای استفاده از تکنیک خوشه‌بندی یادگیری نظارت‌نشده در معاملات زوجی موارد زیر را درنظر می‌گیریم:

۱. عدم نیاز به مشخص کردن تعداد خوشه‌ها از پیش
۲. عدم نیاز به گروه‌بندی تمام اوراق بهادار
۳. انتساب دقیقی که موارد پرت را مدنظر قرار دهد
۴. عدم وجود هیچ‌گونه مفروضی در خصوص شکل خوشه‌ها

پس به سمت انتخاب نوع مناسبی از خوشه‌بندی (خوشه‌بندی مبتنی بر چگالی) می‌رویم که دارای چنین مزایایی است: این خوشه‌ها می‌توانند شکل‌های دلخواهی داشته باشند و ازاین‌رو نیازی به هیچ فرض گاوسی در مورد شکل داده‌ها نیست. این مورد طبیعتا در خصوص داده‌های پرت بسیار قوی می‌باشد؛ چراکه هیچ الگوریتمی به گروه‌بندی هر نقطه در مجموعه داده‌ها نمی‌پردازد [3]. الگوریتم *OPTICS* از [15] پیشنهاد شد تا با مجاز نمودن رویه‌ای خودکار برای انتخاب مناسب‌ترین (فاصله قابل‌دسترسی) به‌ازای هر خوشه، مشکل مربوط به یافتن آن را برای هر خوشه مدنظر قرار می‌دهد. به‌عنوان یک محصول فرعی، ازآنجاکه سرمایه‌گذار تنها نیاز به تعیین تعداد خوشه‌ها (*min Pts*) دارد فرآیند خوشه‌بندی را به فرآیندی بدون پارامتر نزدیک‌تر می‌کند [3].

در بخش معاملات این استراتژی و جهت پیش‌بینی، استفاده از *LSTM* برای تخمین این احتمال است که سهمی در آینده نزدیک در مقایسه با همتایان خود، بازده افزایشی در بازار داشته باشد [16]. شبکه‌های *LSTM* جهت کاربردهای مالی برای نمونه در کارهای [20–17] درنظر گرفته شده‌اند. *LSTM* نوعی شبکه عصبی بازگشتی[1] است. شبکه‌های *LSTM* ابتدا توسط [21] با هدف حل برخی معایب خاص *RNN* یعنی مشکل ناپدید شدن/انفجار شیب که مانع از یادگیری وابستگی‌های بلندمدت با توالی توسط شبکه‌های بازگشت‌کننده می‌شد، معرفی شدند [22]. ایده پشت شبکه‌های *LSTM* استفاده از واحدهای دروازه‌ای جهت کنترل جریان اطلاعات برخاسته از وضعیت‌های گذشته است تا از مساله شناخته‌شده *RNN* استاندارد در یادگیری از دنباله‌های طولانی، اجتناب ورزیم [23]. مطابق نظر [24] شبکه‌های *RNN* نه‌تنها داده‌های انفرادی نقطه‌ای بلکه تمامی توالی داده‌ها را می‌توانند پردازش کنند. این یعنی فرض می‌شود وابستگی‌ای متوالی میان ورودی‌ها وجود دارد و ممکن است حالات قبلی بر تصمیم شبکه عصبی در یک نقطه زمانی متفاوت اثرگذار باشند. این مورد با حلقه‌های داخلی به‌دست می‌آیند که *RNNs* را مجاز می‌نمایند از محاسبات پیشین خود حافظه‌ای داشته باشند.

ساختار *RNN* که *LSTM* از آن منتج شده از یک‌لایه ورودی که دنباله‌ای از داده‌هاست $(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_\tau)$ طی یک مرحله زمانی به شبکه وارد می‌شود، یک یا تعداد بیش‌تری لایه مخفی برای هر مرحله زمانی $(h_1, \cdots, h_\tau)$ و یک‌لایه خروجی که به‌صورت مناسبی با توجه به مشکل انتخاب شده، تشکیل شده است[2]. یک خروجی طبقه‌بندی واحد y در انتهای دنباله درنظر گرفته می‌شود اما در کل *RNN* می‌تواند با پیش‌بینی‌های دنباله-به-دنباله سروکار داشته باشد. بنابراین برای مورد حاضر *RNN* سعی در مدل‌سازی ارتباط زیر دارد:

$$y_\tau = f(x_1, \cdots, x_\tau). \qquad (۱)$$

[1] Recurrent Neural Network (RNN)

[2] از حروف برجسته برای نمایش بردارها و از حروف بزرگ برجسته برای نمایش ماتریس‌ها نشان داده شده است.

ویژگی کلیدی *RNN*، برابر دیگر شبکه‌های عصبی مصنوعی این است که لایه‌های مخفی ماهیتی خودرگرسیو دارند، یعنی

$$h_t = \phi(h_{t-1}, x_t\,;\,\theta), \qquad (۲)$$

که بردار θ پارامترهای تعریف‌کننده خانواده توابع ϕ را بیان می‌کند. برای *RNN* "ساده" با یک‌لایه مخفی و خروجی طبقه‌بندی، مدل را می‌توان با سیستم معادلات زیر خلاصه نمود:

$$h_t = tanh(Wh_{t-1} + Ux_t + b^h) \quad t = 1, \cdots, \tau, \qquad (۳)$$

$$\hat{y}_\tau = softmax(Vh_\tau + b^y), \qquad (۴)$$

که (W، U، V، b^h، b^y) پارامترهایی هستند که باید از داده‌ها تنظیم شوند، "*tanh*" تابع فعال‌سازی لایه مخفی (که طبق معمول مطابق مولفه‌ها اعمال می‌شود) است و

$$\Sigma\Xi Z\Pi MA\Psi i\,(_\Theta < \frac{E^{\Theta}}{s\ {}_I E^{\Theta}} \qquad (۵)$$

برای احتمالات خروجی در مورد طبقه‌بندی ضروری است به‌نحو موثری در حالت باینری به *logit* کاهش یابد. مدل *LSTM* با معرفی سه واحد دروازه‌ای شامل یک دروازه فراموشی f، یک دروازه ورودی خارجی g و یک دروازه خروجی o، به‌همراه دیگر وضعیت داخلی s، کمی بیش‌تر از یک معادله است.

$$\begin{aligned}
Z_T &< \ {}_\triangle l' O^Z H_{\Pi\,\S} * \ 'H^Z \Psi_I * \ B^Z o \quad T < 0 + \text{ⓓ} \Pi \ +\text{ⓐ} + \\
H_I &< \ {}_\triangle{}' O^H H_{\Pi\,\S} * \ 'H^H \Psi_I * \ B^H(\quad T < 0 + \text{ⓓ} \Pi \ +\text{ⓐ} + \\
\Xi_T &< \ {}_\triangle{}' O^\Xi H_{\Pi\,\S} * \ 'H^\Xi \Psi_I * \ B^\Xi(\quad T < 0 + \text{ⓓ} \Pi \ +\text{ⓐ} + \\
\Sigma_I &< Z_T \mathcal{K} \Sigma_{\Pi\,\S} * \ H_I \mathcal{K} TAN\Pi' O^\Sigma H_{\Pi\,\S} * \ 'H^\Xi \Psi_I * \ B^\Xi(\quad T < 0 + \text{ⓓ} \Pi \ +\text{ⓐ} + \\
H_T &< TAN\Pi' \Sigma_I(\ \mathcal{K} \ \Xi_T \quad T < 0 + \text{ⓓ} \Pi \ +\text{ⓐ} + \\
\text{ⓚ}_\S &< \Sigma\Xi Z\Pi MA\Psi' I\ H_{\lrcorner} * \ B^{\text{ⓚ}}(,
\end{aligned} \qquad (۶)$$

که ⊗ بیانگر عامل ضرب مطابق مولفه‌ها و σ تابع فعال‌سازی سیگموئید[1] می‌باشد.

$$\sigma(\Psi\text{ⓙ}) < \frac{0}{0 * \ E^{\ \Psi}} \qquad (۷)$$

نکته مهم در این خصوص، ایجاد خودحلقگی (شرطی) در وضعیت داخلی (اولین معادله بالا) در جایی است که تنها مشروط به فعال‌سازی دروازه فراموشی f، اطلاعات از گام $t-1$ به‌صورت مستقیم به گام t رانده می‌شوند. به‌طور خاص دروازه فراموشی f تعیین می‌کند چه مقدار وضعیت پیشین s_{t-1} مستقیما به s_t تبدیل شده، دروازه ورودی خروجی g به‌تعیین روشی می‌پردازد که لایه مخفی گذشته با s مشارکت می‌کند؛ درحالی‌که دروازه خروجی o، وضعیت روشن/خاموش لایه مخفی فعلی h_t را محاسبه می‌کند. پس، اطلاعات به‌راحتی طی زمان و در امتداد دنباله از وضعیت s_t به s_{t-1} از طریق دروازه f به عقب پس‌انتشار می‌یابد که این مورد با ضرب متوالی شیب‌های کوچک بدون موضوع "از دست دادن حافظه" است (مدل‌های با حافظه مستقیم خود حلقگی و واحدهای دروازه‌ای به کنترل جریان اطلاعات از طریق گام‌های زمانی با نام *RNN* دروازه‌ای می‌پردازند). *RNN* دروازه‌ای پیشرفته، شبکه واحد بازگشتی دروازه‌ای است [24].

با بررسی کارهای انجام‌شده در خصوص آزمون استراتژی معاملات زوجی در بازار سرمایه ایران به این موارد می‌رسیم: [25] از رویکرد فاصله (حداقل‌رسانی معیار فاصله مربع اقلیدسی[2] برای پیاده‌سازی استراتژی معاملات زوجی در صنعت استخراج کانه‌های فلزی (سهام با نمادهای معاملاتی کبافق، کروی، کگل، کچاد و کاما) با داده‌های روزانه قیمتی در سال ۱۳۹۵ استفاده کردند. همچنین تعیین میزان آستانه بهینه نیز یکی دیگر از اهداف این پژوهش بوده است. نتایج حاصله نشان داد پیاده‌سازی این استراتژی در سطوح خاصی از میزان آستانه (۰٫۳ تا ۰٫۹) رویکرد مدنظر

[1] Sigmoid activation function
[2] Sum of Euclidean Squared Distances (SSD)

پژوهش از کارایی لازم برخوردار است. با اعمال استراتژی زوجی و پس‌آزمایی آن، عملکرد با استراتژی خرید و نگهداری مقایسه شد که نتایج حاکی از آن است چنان‌چه سیستم فروش استقراضی وجود داشته باشد و در محدوده آستانه مطلوب، بازدهی معاملات زوجی از استراتژی خرید و نگهداری بیش‌تر خواهد بود. عسگری و ابو [26] در مطالعه‌ای با عنوان بررسی استراتژی معاملات زوجی روی قراردادهای آتی سکه با ترکیب رویکردهای هم‌انباشتگی و تصادفی با استفاده از قیمت تسویه روزانه قراردادهای آتی از ابتدای سال ۱۳۹۱ تا آذر ماه ۱۳۹۱ به بررسی اثربخشی و قابلیت استفاده از استراتژی در معاملات آتی سکه پرداختند. ابتدا با مدل تعمیم‌یافته ریشه واحد تصادفی[1] تغییرات سری قیمت‌ها و شکاف قیمتی از لحاظ ایستایی و وجود ریشه واحد و سپس وجود رابطه هم‌انباشتگی مورد بررسی قرار گرفتند. پس از تحلیل نتایج اجرای مدل خودرگرسیو مرتبه اول[2] و تعیین ضرایب مدل، وجود خاصیت بازگشت به میانگین با میانگین بلندمدت برای داده‌ها بررسی شدند. پس از آن احتمال سود موردانتظار از استراتژی معاملات زوجی با استفاده از مدل بازگشت به میانگین و تابع چگالی زمان هم‌گرایی محاسبه و با روش مونتی‌کارلو دوره و طول میان دوره معاملات برآورد شد. نتایج حاکی از آن است که سری شکاف قیمتی آتی سکه دارای خاصیت بازگشت به میانگین قوی بوده و تلاطم به نسبت بالایی را نشان می‌دهد. پاکیزه و همکاران [27] نیز در پژوهشی به بررسی کاربرد استراتژی معاملات زوجی در بازار قراردادهای آتی سکه طلای بهار آزادی پرداختند. نتیجه حاصله نیز کاربرد این استراتژی را در بازار مدنظر تایید نمود. فلاح پور [28] با محاسبه و بررسی بازده و نسبت سورتینو، عملکرد سیستم معاملات زوجی را با استفاده از رویکرد هم‌انباشتگی در بورس اوراق بهادار تهران مورد بررسی قرار داد. نتایج آزمایش با داده‌های بین روزی (۵ دقیقه‌ای) روی زوج سهام منتخب نشان می‌دهد استفاده از سیستم معاملات زوجی به‌عنوان یک سیستم معاملاتی خنثی نسبت به تغییرات و روندهای بازار، بازدهی چشمگیری نسبت به بازدهی معمولی سهام در مدت مشابه دارد. این در حالی است که بازدهی سهام اغلب شرکت‌های منتخب در این پژوهش در بازه زمانی مورد بررسی منفی بوده‌اند. از طرفی نتایج حاصل از این پژوهش نشان‌دهنده عملکرد چشمگیر استراتژی معاملات زوجی در قیاس با معاملات سهام و دیگر استراتژی‌های معاملات الگوریتمی به‌کار گرفته شده در پژوهش‌های قبلی می‌باشد. همه این موارد را می‌توان در دسته رویکرد سنتی استراتژی معاملات زوجی قرار داد و جای خالی رویکرد نوین مربوط به استفاده از تکنیک‌های علم داده بیش‌ازپیش احساس می‌شود.

۳- روش پژوهش

جامعه داده‌های این پژوهش، سهام معامله‌شونده در بورس اوراق بهادار تهران و فرابورس ایران از فروردین ماه ۱۳۹۳ تا اسفند ماه ۱۳۹۹ برای فواصل زمانی ۵ دقیقه‌ای هستند. تعداد نمادهای مشترک معامله‌شونده از ابتدای سال ۱۳۹۳ تا انتهای سال ۱۳۹۹، ۲۳۱ نماد (غیر از عرضه‌های اولیه) می‌باشند که برخی از نمادها به‌دلیل در دسترس نبودن داده‌های کامل آن‌ها و بسیاری از آن‌ها به‌دلیل تعداد بالای روزهای بسته بودن نماد (بیش از ۵۰% بازه زمانی مدنظر) حذف شدند. در نهایت ۱۱۱ نماد مختلف تاییدی بر شرایطی هستند که تشریح شد (**جدول ۱**).

با توجه به شرط نقدشوندگی منطبق با معیار اتخاذی [1]، [12] که تمامی سهام با حداقل یک روز بدون معامله باید حذف شوند و ۱۱۱ داده موجود در مجموعه این پژوهش که هر نماد کم‌تر از ۵۰% داده از دست‌رفته دارد و نیز امکان وجود داده‌های پرت در مجموعه داده، انتساب داده و پاک‌سازی آن با استفاده از بسته با منبع باز *tsrobprep* در نرم‌افزار *R* صورت پذیرفت که روش‌های کارآمدی برای مدیریت مقادیر ازدست‌رفته و نقاط پرت با استفاده از رویکردهای مبتنی بر مدل معرفی می‌کند. برای انتساب داده‌ها یک مدل جایگزینی احتمالی پیشنهاد شد که ممکن است از اجزای خودرگرسیونی و ورودی‌های خارجی تشکیل شده باشد. برای تشخیص داده پرت، یک الگوریتم خوشه‌بندی مبتنی بر مدل‌سازی ترکیبی محدود معرفی شده که ویژگی‌های سری زمانی را برحسب شیب و فصلی بودن زیربنایی به‌عنوان ویژگی‌ها در نظر می‌گیرد [29].

در ادامه برای هر شبیه‌سازی معاملاتی، باید داده‌ها به دو دوره تقسیم شوند. دوره تشکیل و دوره معاملات. دوره تشکیل به شبیه‌سازی داده‌های در دسترس سرمایه‌گذار پیش از ورود به هر معامله‌ای می‌پردازد و نقش آن سه‌گانه است:

۱. این دوره برای یافتن جذاب‌ترین زوج‌های کاندید مورداستفاده قرار می‌گیرد.
۲. بخش کوچک‌تری از داده‌ها (مجموعه اعتبارسنجی) برای شبیه‌سازی چگونگی عملکرد استراتژی در زمان‌های اخیر مورداستفاده قرار می‌گیرد.

[1] Generalised Stochastic Unit Root (GSTUR)
[2] Auto Regressive (AR)

۳. کاربردی برای مدل‌های مبتنی بر پیش‌بینی، بخشی از داده‌های دوره تشکیل (مجموعه آموزش) برای آموزش الگوریتم پیش‌بینی مورداستفاده قرار می‌گیرند. دوره آموزش برای این مورد که آیا مدل معاملاتی اجرا شده با داده‌های مشاهده‌نشده آتی شبیه هستند یا خیر، کاربرد دارد.

جدول ۱- شرح داده‌های مرتبط با تقسیم‌بندی داده‌ها.
Table 1- Data description related to data segmentation.

سال	تعداد روزهای معاملاتی	تعداد دوره‌های ۵ دقیقه‌ای
1393	240	10080
1394	240	10080
1395	236	9912
1396	241	10122
1397	239	10038
1398	238	9996
1399	242	10164

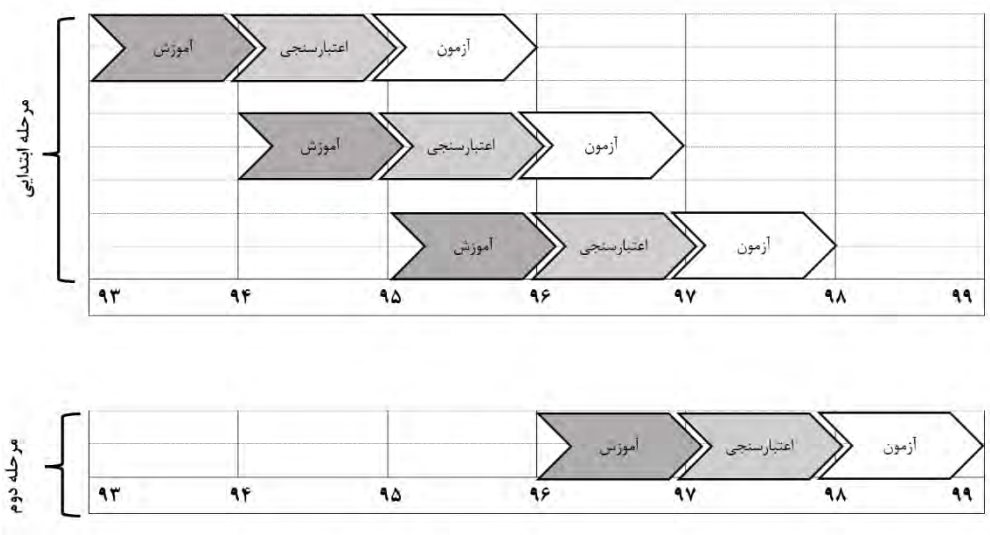
در این کار با استفاده از هزینه‌های معاملاتی، بازده‌ها خالص از هزینه هستند، هزینه‌ها شامل هزینه کارمزد برای هر خرید ۰٫۳۷۱۲% و برای هر فروش ۰/۸۸% است که سقف کارمزد کارگزاران در معاملات سهامی باارزش زمان معامله ۵ میلیون ریال یا کم‌تر، "بیست هزار ریال" تعیین‌شده است. ازاین‌رو حداقل سرمایه برای شروع معاملات ۱۰ میلیون ریال درنظر گرفته شده است. برای کارمزد هزینه اجاره اوراق بهادار در فروش استقراضی با توجه به ممنوع بودن فروش استقراضی در بازار سرمایه ایران، فرض می‌شود سبد متنوعی داریم که می‌توان طرف فروش معامله زوجی را از آن انتخاب نمود، ازاین‌رو همان هزینه فروش برای این بخش درنظر گرفته می‌شود.

پس از آماده‌سازی اولیه داده‌ها، با استفاده از تجزیه‌وتحلیل مولفه‌های اصلی PCA[1] به کاهش ابعاد داده‌ها (۵ بعد) می‌پردازیم، سپس با اعمال ۴ قانون زیر بر خروجی حاصل، به زوج‌های مناسب برای معامله می‌رسیم:

۱. خاصیت هم‌انباشتگی، با استفاده از آزمون انگل و گرنجر [30]، و انتخاب ترکیبی که کم‌ترین آماره t را ایجاد کند.
۲. جز هرست (H) کوچک‌تر از ۰٫۵ که گام اعتبارسنجی مازادی برای ویژگی بازگشت به میانگین شکاف قیمتی می‌باشد. H به کمی‌نمودن گرایش نسبی یک سری زمانی به رگرس‌شدن قوی به میانگین یا دنبال نمودن یک روند می‌پردازد [31].
۳. نیمه‌عمر میانگین یک روز تا یک‌سال، به‌منظور رهایی از زوج‌های ایستا با زمان‌بندی نامناسب می‌باشد. یک شکاف با بازگشت به میانگین به‌خودی‌خود، لزوما سود تولید نمی‌کند. باید بین زمان بازگشت به میانگین و دوره معاملاتی چسبندگی وجود داشته باشد. نیمه‌عمر بازگشت به میانگین شاخصی از زمانی است که یک سری زمانی به میانگین بازمی‌گردد [32]. بنابراین پیشنهاد بر حذف نیمه‌عمر کم‌تر از یک روز و بیش‌تر از یک سال می‌باشد.
۴. عبور از میانگین بیش از ۱۲ بار در سال تا به‌طور متوسط یک معامله در هرماه داشته باشیم.

برای اولین مرحله پژوهش (مرحله تشکیل) پیشنهاد بر استفاده از ۳ دوره متفاوت است تا شواهد آماری بیش‌تری در خصوص نتایج حاصله داشته باشیم. در دومین مرحله پژوهش (مرحله معاملات) به‌دلیل بار محاسباتی آموزش الگوریتم‌های پیش‌بینی، این امر امکان‌پذیر نیست. دوره ۳-ساله برای مدل معاملاتی استفاده می‌شود، هرچند این دوره مقداری نسبت به ادبیات کار [1]، [7]، [33] از یک دوره تشکیل یک ساله و از یک دوره تشکیل سه ماهه استفاده کردند) کمی بزرگ‌تر است (**شکل ۱**) [34].

[1] Principal Component Analysis (PCA)



شکل ۱- دوره‌های تشکیل و معاملاتی.
Figure 1- Formation and trading periods.

چون یک سیستم معاملاتی نمی‌تواند به‌سرعت عمل کند، ممکن است انحراف کوچکی در نقطه ورود باشد که ذات مربوط به تاخیر در ورود به یک موقعیت است. جهت مدنظر قرار دادن این عامل و اطمینان از امکان‌پذیر بودن استراتژی در عمل، فرض می‌کنیم یک دوره (۵ دقیقه) تاخیر برای ورود به موقعیت است. که این کار به‌معنای اجرای یک سیستم توقف-زیان نیست. بلکه بدین معناست که چنان‌چه زوج همگرا شود یا دوره معاملاتی خاتمه یابد از موقعیت خارج شده‌ایم. تمامی کدها در محیط برنامه‌نویسی پایتون نوشته و اجرا شدند. برخی از کتابخانه‌های پایتون به‌طور خاص در اینجا مفید هستند. ابتدا، *sci-kit learn* به‌منظور اجرای *PCA* و الگوریتم *OPTICS*، دوم، *statsmodels* یک نسخه تقریبا اجرایی از آزمون *ADF* و برای انجام آزمون هم‌انباشتگی، *tensorflow* برای پیاده‌سازی و آموزش مدل‌های یادگیری عمیق و درنهایت استفاده از *keras*، یک کتابخانه یادگیری عمیق که بلوک‌های سازنده یک شبکه عصبی را فراهم می‌آورد. برای اجرای *LSTM* از تابع فعال‌سازی رلو[1] و برای معیار ارزیابی از میانگین قدرمطلق[2] خطا استفاده شده است.

کار را با سه سبد آزمایشی متفاوت اوراق بهادار شروع می‌کنیم. سبد اوراق بهادار یک تمامی زوج‌های شناسایی‌شده در دوره تشکیل را شامل می‌شود. سبد اوراق بهادار دو از بازخورد جمع‌آوری‌شده ناشی از اجرای استراتژی در مجموعه اعتبارسنجی با انتخاب زوج‌هایی که نتیجه مثبتی دارند، استفاده می‌کند. سبد اوراق بهادار سه به موقعیتی مربوط می‌شود که سرمایه‌گذار محدود به سرمایه‌گذاری در تعداد ثابتی از k زوج است. در چنین موقعیتی، پیشنهاد به انتخاب k زوج با بازده کسب‌شده بالا است که مربوط به مجموعه اعتبارسنجی می‌باشد. K را ۱۰ درنظر می‌گیریم چراکه بین انتخاب‌هایی که از $k = 5$ و $k = 20$ استفاده کرده‌اند قرار می‌گیری [1]، [14]. با آزمون ساخت سبدهای متفاوت اوراق بهادار، نه‌تنها رویه خوشه‌بندی بهینه را می‌یابیم بلکه همچنین از بهترین شرایط کاربرد آن‌ها نیز ارزیابی صورت می‌پذیرد.

[1] Relu
[2] Mean Absolute Error (MAE)

جهت مقایسه استحکام فراهم‌آمده برای مدل *LSTM* از یک مدل استاندارد مبتنی بر آستانه استفاده می‌کنیم. هنگامی‌که نتایج نشانی از برازش بیش‌ازحد را نشان دهند، آموزش متوقف می‌شود. یک خط‌پایه ساده[1] به تعریف هدف کرانه پایین‌تر می‌پردازد که برای پشتیبانی از فرآیند آموزش ایجاد شده است. برای این منظور یک الگوریتم پیش‌بینی مداوم انتخاب شد، که از مقدار مربوط به مرحله زمانی گذشته برای پیش‌بینی نتیجه موردانتظار در مرحله زمانی آتی مطابق زیر استفاده می‌کند:

$$Naive: Y_{t+1} = Y_t \ . \qquad (۸)$$

این مدل باید به مدیریت دوره‌های کاهش طولانی ارزش سبد اوراق بهادار به‌دلیل وجود زوج‌هایی که به‌مدت طولانی واگرا بوده‌اند، بپردازد. رویه کلی مدل مبنا (مدل مبتنی بر آستانه) این است که ابتدا شکاف قیمتی بین دو زوج و میانگین و انحراف معیار شکاف قیمتی محاسبه شده، سپس آستانه‌هایی برای خرید، α_l، فروش، α_s و خروج از موقعیت معاملاتی، α_{exit} به این صورت تعریف شود:

$$\begin{cases} \mu_s - 2\sigma_s, & \text{آستانه خرید} \\ \mu_s + 2\sigma_s, & \text{آستانه فروش} \\ \mu_s. & \text{آستانه خروج} \end{cases} \qquad (۹)$$

در این مرحله شکاف قیمتی بررسی می‌شود تا اینکه آستانه‌ها قطع شوند تا معامله انجام شود. مدل دیگر (*LSTM*) مبتنی بر پیش‌بینی شکاف قیمتی می‌باشد که شکاف قیمتی و آستانه‌ها در آن به‌ترتیب زیر تعریف می‌شوند:

$$\text{تغییر پیش‌بینی‌شده}: \Delta_{t+1} = \frac{S^*_{t+1} - S_t}{S_t} \times 100, \qquad (۱۰)$$

که S^*_{t+1} قیمت پیش‌بینی‌شده و S_t قیمت واقعی است.

$$\text{شرایط ورود به بازار}: \begin{cases} if\ \Delta_{t+1} \geq \alpha_l, & \text{باز کردن یک موقعیت خرید،} \\ if\ \Delta_{t+1} \leq \alpha_s, & \text{باز کردن یک موقعیت فروش،} \\ & \text{در غیر این‌صورت،} \\ & \text{خارج از بازار باشید} \end{cases} \qquad (۱۱)$$

حالا باید آستانه‌ها را در مدل تعریف کنیم: با توجه به مدنظر بودن تغییرات ناگهانی با فراوانی کافی، به جست‌وجوی چندک‌های حداکثری می‌پردازیم که به‌نظر کافی هستند. به‌موازات اینکه چندک‌ها با نوسان مربوط به شکاف قیمتی تطابق می‌یابند، درنهایت آستانه‌ها با ویژگی‌های مربوط به شکاف قیمتی مرتبط می‌شوند. صدک‌ها و چندک‌های بالا را از $f^+(x)$ (که $f(x)$ توزیع تغییر درصد شکاف قیمتی است) به‌عنوان کاندیدای تعریف α_l و موارد پایینی را از $f^-(x)$ برای تعریف α_s به‌کار می‌بریم. فرآیند تشریح‌شده در اینجا را می‌توان به‌صورت رسمی‌تر این‌گونه ارایه داد:

$$\text{[illegible]} \qquad (۱۲)$$

که R^{val} بازده کسب‌شده در مجموعه اعتبارسنجی است.

جهت تخمین صحت مدل‌ها از بازده سرمایه‌گذاری ($ROI = \frac{\text{سود خالص}}{\text{سرمایه اولیه}} \times 100$)، نسبت شارپ ($SR_{\text{سال}} = \frac{R^{port} - R_f}{\sigma_{port}} \times \sqrt{\text{عامل سالانه}}$) و کاهش حداکثری (حداکثر کاهش مشاهده‌شده از قله به‌سمت پایین از یک سری زمانی پیش از آنکه به قله جدید برسد که با توجه به مانده‌حساب طی دوره معاملاتی محاسبه می‌شود) ($MDD_{(T)} = Max_{\tau \in (0,T)} \left[Max_{\tau \in (0,T)} \frac{X_{(t)} - X_{(\tau)}}{X_{(t)}} \right]$ که $X_{(t)}$ کل مانده‌حساب است) استفاده می‌کنیم و جهت تخمین صحت پیش‌بینی از سه مورد زیر استفاده می‌کنیم:

$$\text{[illegible]} \qquad (۱۳)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |e_t|. \qquad (۱۴)$$

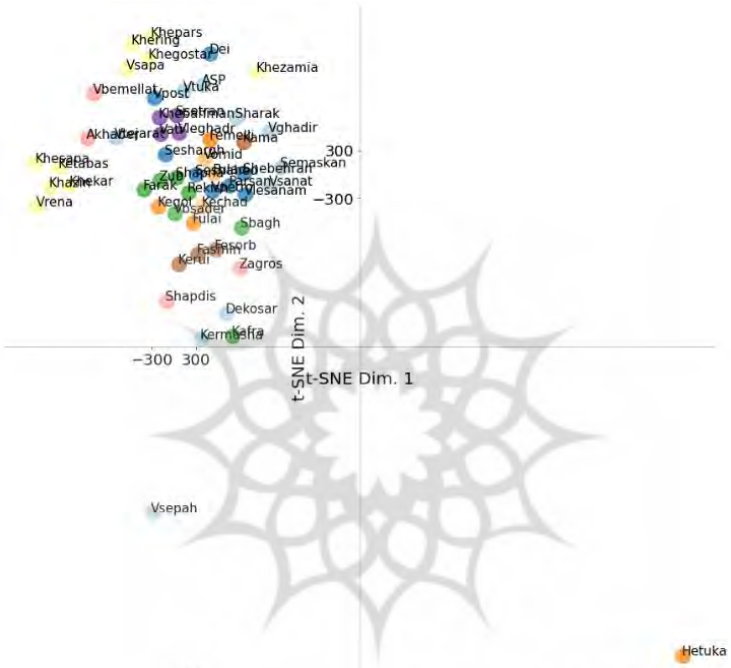
$$RMSE = \sqrt{MSE}. \qquad (۱۵)$$

[1] Naive baseline

۴- تحلیل داده‌ها

تجزیه‌وتحلیل نشان می‌دهد استراتژی‌های پیش‌بینی می‌توانند در بهبود عملکرد سبد اوراق بهادار با فراهم آوردن سیگنال‌های پیش‌بینی‌کننده حاوی اطلاعاتی برای حرکت به سمت بالا و فراتر از شکاف‌های قیمتی و بازده، موثر باشند. ازاین‌رو نتایج *LSTM* در تقویت مربوط به سیگنال‌های خرید یا فروش با شناسایی احتمال افزایش یا کاهش شکاف بازار هر سهم نسبت به همتایانش مشارکت دارد.

برای شروع و در خصوص اعتبارسنجی خوشه‌های *OPTICS* با تجزیه‌وتحلیل دوره‌ها از فروردین ۱۳۹۵ تا اسفند ۱۳۹۸ شروع می‌کنیم. داده‌ها با پنج بعد (تعداد اجزای اصلی) توصیف شده‌اند. برای مواجه شدن با مشکل نمایش آن‌ها از الگوریتم همسایه تصادفی محاط با توزیع-T[1] مربوط به پژوهش [35] استفاده می‌شود. **شکل ۲** به تشریح خوشه‌های شکل‌گرفته طی دوره تشکیل، در فضایی تبدیل‌شده می‌پردازد. هر خوشه با یک رنگ نمایش داده شده و سهام گروه‌بندی‌نشده نمایش داده نشده‌اند.



شکل ۲- کاربرد t-SNE برای خوشه‌های تولیدشده توسط OPTICS در دوره فروردین ۱۳۹۵ تا اسفند ۱۳۹۸.
Figure 2- t-SNE usage for clusters of OPTICS (Farvardin 1395-Esfand 1398).

چنان‌چه رویکرد خوشه‌بندی با *OPTICS* را با رویکرد بدون خوشه (تمامی سهام)مقایسه کنیم، می‌توان تایید نمود نه‌تنها سودآوری در هر دوره پایدار است (**جدول ۲**) بلکه نتایج خوشه‌بندی به مراتب بهتر از درنظر گرفتن کل سهام موجود می‌باشد. این نتیجه تاییدی بر این ایده است که رویه انتخاب زوج‌ها، مستحکم می‌باشد. رویکرد مبتنی بر *OPTICS* قادر است میانگین نسبت شارپ ۱۰٫۵۷ مربوط به سبد اوراق بهادار را کسب نماید که عملکرد بسیار خوبی است. همچنین با انتخاب تعداد مناسبی زوج، ۱۰۰% آن‌ها سودآور بودند و با توجه به دوره مورد آزمون تعداد روزهای کاهش ارزش سبد اوراق بهادار، کاهش مناسبی داشته است. در کل، استراتژی مبتنی بر *OPTICS* ثابت می‌کند قابلیت حفظ یک میانگین *MDD* مطلوب را در کل دارد. برای حصول دیدی بهتر از اوراق بهادار تشکیل‌دهنده هر خوشه، تجزیه‌وتحلیل سری‌های قیمتی مربوطه (نه خوشه) پیشنهاد می‌شود.

[1] T-distributed Stochastic Neighbor Embedding (T-SNE) algorithm

جدول ۳- نتایج اعتبارسنجی برای هر تکنیک گروه‌بندی.

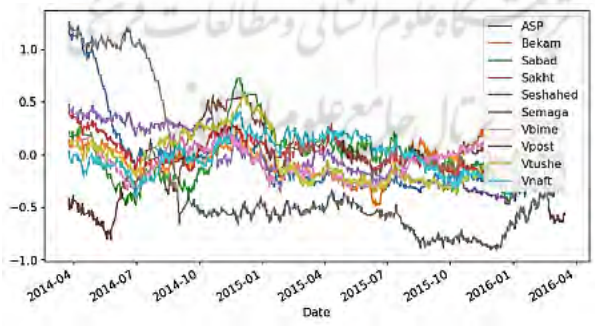
Table 3- Validation results for each grouping technique.

دوره اعتبارسنجی	1395	1396	1397
بدون خوشه‌بندی			
SR	6.17	5.54	7.12
ROI (%)	20804.22	12817.77	9717.12
MDD (%)	41.36	5.06	7.08
زوج‌های سودآور (%)	96.32	98.09	98.84
تعداد معاملات (مثبت-منفی)	364 (249-115)	295 (208-87)	254 (143-120)
OPTICS			
SR	9.89	22.80	10.25
ROI (%)	11031.82	4608.43	2978.53
MDD (%)	17.65	9.23	17.23
زوج‌های سودآور (%)	100.00	92.86	100.00
تعداد معاملات (مثبت-منفی)	55 (45-10)	28 (20-8)	4 (3-1)

شکل ۳ به توضیح پیرامون الگوریتم سری‌های قیمتی هر سهم می‌پردازد. سری‌های قیمتی شرح‌داده‌شده ناشی از کسر شدن میانگین از سری‌های قیمتی اصلی هستند تا تجسم آن ها تسهیل شود.

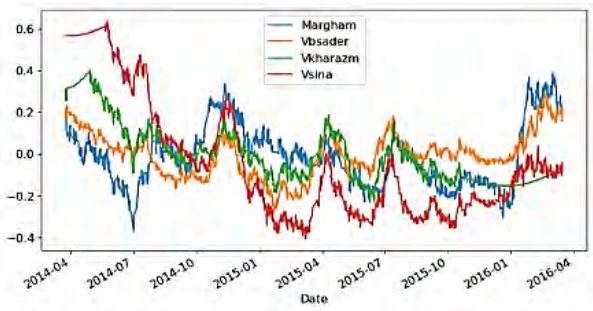
با شروع از **شکل ۳-الف**، جالب است ببینیم چگونه سری‌های نزدیک هم، یک‌دیگر را دنبال می‌کنند که همین موضوع تمایز آن ها را از هم دشوار می‌سازد. ده ورقه بهادار به دسته انبوه‌سازی، املاک و مستغلات و سرمایه‌گذاری تعلق دارند اما باهم در ارتباط هستند که با موقعیت مشابهی در شکل ۳ روبه‌رو هستیم.

شکل ۳-ب تا **شکل ۳-ح** ارایه‌دهنده قابلیت‌های خوشه‌بندی *OPTICS* می‌باشند که فراتر از انتخاب سهام درون همان بخش است. در خوشه ۳، سهامی را می‌یابیم که مربوط به بخش شیمیایی (پارسان)، حمل‌ونقل (حتاید)، غذایی (غمارگ)، سایر کانه‌های غیرفلزی (کرازی) و بانکی (وبملت) باشند. ارتباطی مشهود میان سری‌های قیمتی شناسایی‌شده وجود دارد حتی اگر آن‌ها تماما به دسته مشابهی تعلق نداشته باشند. همین مورد در خصوص خوشه ۸ اعمال می‌شود که از سه دسته متفاوت املاک و مستغلات (ثباغ و ثتران)، خودرو (خکار و ورنا) و شیمیایی (شکربن و وپترو) تشکیل شده است.



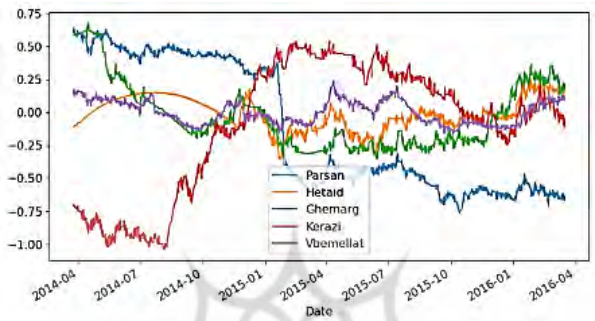
الف- قیمت‌های نرمال‌شده سهام خوشه ۱.

a- Normalized stock prices of cluster 1.



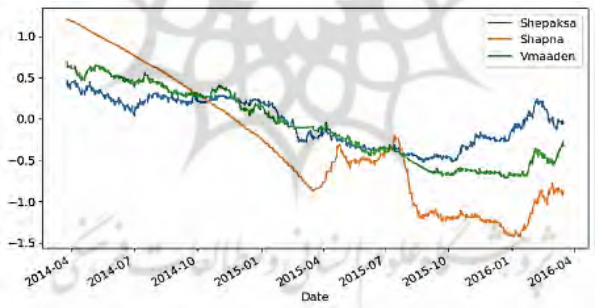
ب- قیمت‌های نرمال‌شده سهام خوشه ۲.

b- Normalized stock prices of cluster 2.



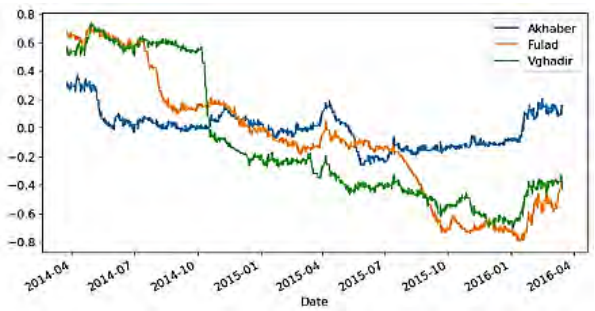
پ- قیمت‌های نرمال‌شده سهام خوشه ۳.

c- Normalized stock prices of cluster 3.



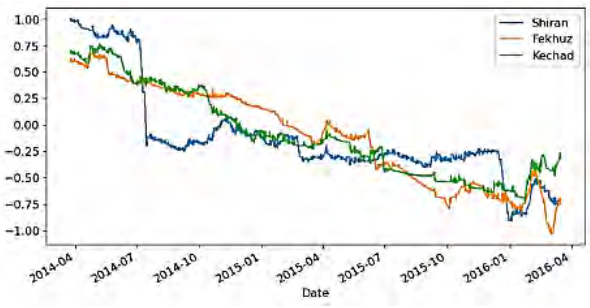
ت- قیمت‌های نرمال‌شده سهام خوشه ۴.

d- Normalized stock prices of cluster 4.



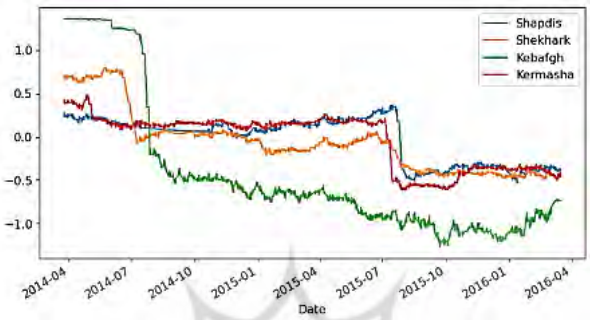
ث- قیمت‌های نرمال‌شده سهام خوشه ۵.

e- Normalized stock prices of cluster 5.



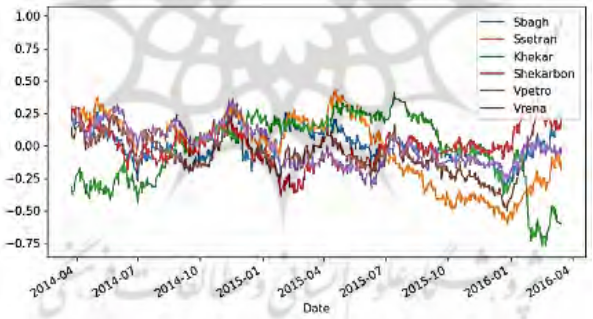
ج- قیمت‌های نرمال‌شده سهام خوشه ۶.

C- Normalized stock prices of cluster 6.



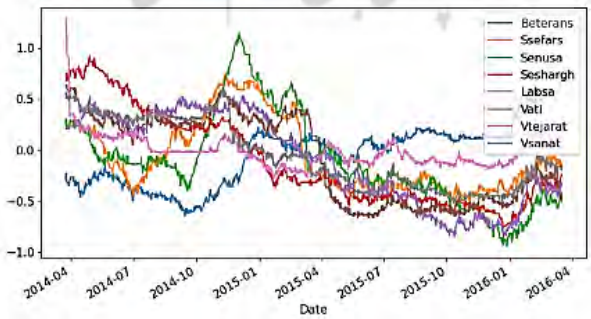
چ- قیمت‌های نرمال‌شده سهام خوشه ۷.

G- Normalized stock prices of cluster 7.



ح- قیمت‌های نرمال‌شده سهام خوشه ۸.

H- Normalized stock prices of cluster 8.



خ- قیمت‌های نرمال‌شده سهام خوشه ۹.

X- Normalized stock prices of cluster 9.

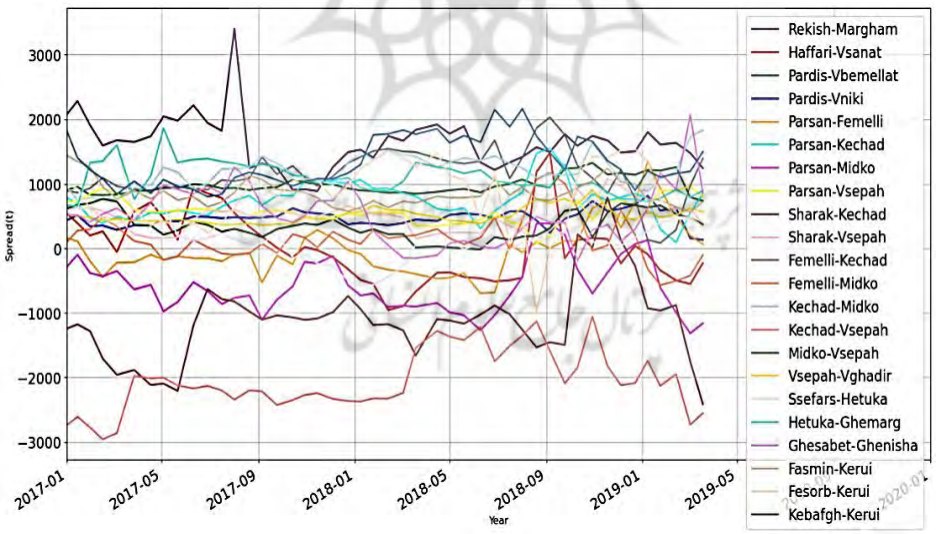
شکل ۳- ترکیب سری‌های قیمتی چند خوشه شکل‌گرفته در دوره‌ها از (فروردین ۱۳۹۵ تا اسفند ۱۳۹۶).

Figure 3- Combination of price series formed by multiple clusters (farvardin 1395-esfand 1398).

حال برای رسیدن به هدف عملکرد مناسب مدل مبتنی بر پیش‌بینی، به ارزیابی این‌که چگونه این مدل هنگام مقایسه با مدل مبتنی بر آستانه عمل می‌کند، می‌پردازیم. با این کار به‌تدریج به توصیف چگونکی پیشرفت شبیه‌سازی، از شناسایی زوج‌ها تا معامله آن‌ها خواهیم پرداخت. بر دوره‌ای از فروردین ۱۳۹۶ تا اسفند ۱۳۹۸ که ۲۲ زوج واجد شرایط برای معامله در آن یافت شد متمرکز می‌شویم. این عدد با مقادیر یافت‌شده در دوره‌های پیشین با همین دیرش تطابق دارد. این زوج‌ها را به‌عنوان مقیاسی از عملکرد مدل معاملاتی الگو که در محدوده بازه زمانی استانداردش اعمال شد (۳ سال) مورد استفاده قرار دادیم. الگوریتم‌های پیش‌بینی، آموزش می‌بینند تا شکاف‌های قیمتی مربوط به زوج‌های منتخب **جدول ۳** طی دوره تشکیل را برازش کنند. همان‌طور که انتظار می‌رفت شکاف‌های قیمتی، ایستا به‌نظر می‌رسند. علاوه بر این، تفاوت محسوسی در نوسان میان سری‌های زمانی وجود دارد که از این ایده که پارامترهای مدل داده-محور، همانند آستانه‌های پیشنهادی برای مدل معاملاتی مبتنی بر پیش‌بینی، برای این شرایط مناسب به‌نظر می‌رسند، بیش‌تر پشتیبانی می‌کند **شکل ۴**.

جدول ۳- زوج‌های کاندید شناسایی‌شده از فروردین ۱۳۹۶ تا اسفند ۱۳۹۸.
Table 3 - Identified candidate pairs (farvardin 1395-esfand 1398).

زوج‌ها	گروه	زوج‌ها	گروه
فملی	فلزات اساسی	فملی	فلزات اساسی
کچاد	استخراج کانه‌های فلزی	میدکو	فلزات اساسی
پارسان	محصولات شیمیایی	حتوکا	حمل‌و.نقل، انبارداری و ارتباطات
فملی	فلزات اساسی	غمارگ	محصولات غذایی و آشامیدنی جز قند و شکر
پارسان	محصولات شیمیایی	رکیش	رایانه و فعالیت‌های وابسته به آن
کچاد	استخراج کانه‌های فلزی	مرقام	رایانه و فعالیت‌های وابسته به آن
پردیس	سرمایه‌گذاری	وساپا	سرمایه‌گذاری
وبملت	بانک‌ها و مؤسسات اعتباری	وغدیر	شرکت‌های چندرشته‌ای صنعتی
پارسان	محصولات شیمیایی	فاسمین	فلزات اساسی
میدکو	فلزات اساسی	کروی	استخراج کانه‌های فلزی



شکل ۴- شکاف قیمتی مربوط به زوج‌های شناسایی‌شده بین فروردین ۱۳۹۶ و اسفند ۱۳۹۸.
Figure 4 - Price gaps for the 22 identified pairs (Farvardin 1395-Esfand 1398).

با تشریح داده‌های آموزش، به سمت سبک‌های پیش‌بینی آزمایش‌شده پیش می‌رویم. این سبک‌ها پیچیده‌تر می‌شوند تا اینکه هیچ شواهدی از بهبود یافت نشود. در مجموع ۲۱ سبک مدل پیش‌بینی اجرا شدند به این معنی که ۴۶۲ مدل (۲۱ سبک با ۲۲ زوج) آموزش دیدند. مهم است تصریح کنیم بهترین سبک‌های مدنظر آن‌هایی هستند که میانگین مربع خطای اعتبارسنجی را به‌حداقل می‌رسانند. میانگین مربع خطا به‌عنوان سنجه‌ای برای اهداف مقایسه‌ای مدنظر قرار می‌گیرد؛ زیرا تابع زیانی است که با مدل‌ها به حداقل می‌رسد. مطابق این معیار، بهترین سبک عملکرد در **جدول ۴** ارایه شده است.

جدول ۴- مقایسه نتایج پیش‌بینی
Table 4- Comparison of prediction results.

مدل معاملاتی	استاندارد	مدل مبتنی بر LSTM
پارامترها	$\alpha_S = \mu_s + 2\sigma_s$ $\alpha_L = \mu_s - 2\sigma_s$	$\alpha_S = Q_{f-(0.10)}$ $\alpha_L = Q_{f+(0.90)}$
اعتبارسنجی		
تعداد روزهای کاهش	2	12
SR	6.7	2.18
ROI (%)	7911.1	29817.4
MDD (%)	8.43	33/97
تعداد معاملات (مثبت-منفی)	46 (12-34)	9044 (8983-61)
زوج‌های سودآور (%)	100	22.7
آزمون		
تعداد روزهای کاهش	10	190
SR	0.53	2.18
ROI (%)	1918.9	-90.12
MDD (%)	23.13	99.80
تعداد معاملات (مثبت-منفی)	25 (12-13)	1196 (1162-34)
زوج‌های سودآور (%)	86.4	0

با این یادآوری شروع می‌کنیم که انگیزه اصلی برای دنبال کردن رویکرد معاملاتی زوج‌های مبتنی بر پیش‌بینی، ایجاد سبد اوراق بهادار مستحکم‌تری با کاهش تعداد روزهایی است که سبد مذکور در معرض کاهش ارزش خود قرار دارد. جهت ارزیابی اثربخشی روش پیشنهادی، به تشریح نتایج حاصله می‌پردازیم. نتایج در **جدول ۵** کنار نتایج مربوط به مدل استاندارد مبتنی بر آستانه قرار گرفته‌اند.

جدول ۵- مقایسه نتایج معاملاتی با استفاده از یک دوره تشکیل ۳-ساله.
Table 5- Comparison of trading results using a 3-year formation period.

مدل	LSTM
تشکیل	طول ورودی: ۱۲ لایه‌های مخفی: ۱ لایه‌های مخفی: ۶۰
MSE(E-03)	8.10
RMSE(E-02)	9.00
MAE(E-02)	5.54
MSE(E-03)	0/75
RMSE(E-02)	2.74
MAE(E-02)	1.80

با تمرکز بر نتایج آزمون که شبیه یک محیط معاملاتی عملی می‌باشد می‌توان تایید نمود در مدل معاملاتی استاندارد به‌طور قابل‌ملاحظه‌ای دوره کاهش سبد اوراق بهادار کوتاه‌تر است اما بااین‌وجود و درنظر گرفتن دوره آزمون، دوره کاهش سبد اوراق بهادار در مدل *LSTM* طولانی است. به‌هر حال، تمام آنچه که باید باشد نیست و از نظر بازده کل سرمایه‌گذاری و نسبت شارپ سبد اوراق بهادار نیز عملکرد مناسبی ندارد. در نتیجه مدل‌های مبتنی بر *LSTM* از مدل معاملاتی استاندارد پیشی نمی‌گیرد. طبیعتا نتایج منجر به طرح این پرسش برای ما می‌شوند که *LSTMs* برای شرایط این کار مناسب نبودند. هر چند برازش الگوریتم‌های متفاوت پیش‌بینی قویا به مجموعه داده وابسته است، اما در ادبیات موضوع هم هیچ اجماعی در خصوص مناسب بودن *LSTMs* برای پیش‌بینی سری‌های زمانی وجود ندارد. از یک طرف، برخی تحقیقات علمی از برتری *RNNs* بر تکنیک‌های ساده‌تر مانند مدل‌های ادراک‌کننده‌های چند لایه[1] یا خود-رگرسیو حمایت می‌کنند. برای نمونه [36] و[37] ادعا می‌کنند این شبکه‌ها به‌طور کلی بهتر از شبکه‌های *MLP* هستند هر چند از زمان طولانی محاسباتی رنج می‌برند.

[1] Multi-Layer Perceptrons (MLP)

۵- نتیجه‌گیری

در این کار به بررسی استفاده از الگوریتم‌های یادگیری ماشینی برای دو مرحله استراتژی معاملات زوجی (انتخاب زوج‌ها و انجام معاملات) پرداختیم. برای مرحله اول نتایج بسیار امیدوارکننده بود. چنان‌چه یک سرمایه‌گذار فارغ از ریسک متحمل‌شده، بر کسب *ROI* با بالاترین احتمال متمرکز شود تقریبا می‌توان روش مبتنی بر *OPTICS* را برگزید. بااین‌حال، چنان‌چه سرمایه‌گذار نگران ریسک تحمیلی بر سبد اوراق بهادار خود باشد، رویکرد مبتنی بر *OPTICS* قادر است تا بالاترین میانگین نسبت شارپ مربوط به سبد اوراق بهادار را در مقایسه با هنگامی‌که هیچ خوشه‌بندی صورت نگرفته، کسب کند. همچنین نشان‌دهنده سازگاری بیش‌تری با توجه به نسبت زوج‌های سودآور در سبد اوراق بهادار، در استفاده از *OPTICS* در مقابل هنگامی‌که اصلا هیچ خوشه‌بندی صورت نگرفته، است. سرانجام، کاهش‌های پایدارتری در سبد اوراق بهادار حاصل می‌شود و مقادیر *MDD* در محدوده قابل‌قبولی است. در کل، استراتژی مبتنی بر *OPTICS* ثابت می‌کند قابلیت حفظ یک میانگین *MDD* مطلوب را دارد. در این مورد [3] نیز با استفاده از این الگوریتم توانست در بازار صندوق‌های قابل‌معامله بازار سهام امریکا به نتایج مطلوبی دست یابد.

برای انجام معاملات با مدل پیش‌بینی *LSTM* به نتایج دلخواهی از منظر بازدهی و کاهش تعداد روزهای افول سبد نرسیدیم. بااین‌وجود آزمون سایر مدل‌های یادگیری ماشینی برای انجام پیش‌بینی در معاملات در بازار سرمایه ایران پیشنهاد می‌شود. در کار [14] به معرفی مدل معاملاتی مبتنی بر پیش‌بینی (*LSTM*) پرداختند که قادر به کاهش دوره‌های رکود پرتفو بود. با این حال این مورد به قیمت کاهش سودآوری کلی انجامید. در کار [16] ارتباطات غیرخطی میان و بین سری‌های زمانی مالی با استفاده از *LSTM* برای پیش‌بینی عملکرد بازار یک نمونه بزرگ از سهام مطالعه شد. پیش‌بینی *LSTM* در چارچوب استراتژی‌های سرمایه‌گذاری معامله زوجی آزمون شدند. که به مقایسه پیش‌بینی‌های چارچوب *LSTM* پیشنهادی برابر دیگر رویکردهای عادی *ML* یعنی، رگرسیون منطقی، درخت‌های-مدرج-تقویت‌شده[1] و شبکه‌های عصبی فیدفوروارد[2] پرداخته شد. در این خصوص از *LSTM* برای تخمین احتمال بازده افزایشی سهمی در بازار در آینده نزدیک و در مقایسه با همتایان خود استفاده شده است. تجزیه‌وتحلیل نشان می‌دهد استراتژی‌هایی شامل چنین پیش‌بینی‌هایی می‌توانند در بهبود عملکرد پرتفو با فراهم آوردن سیگنال‌های پیش‌بینی‌کننده حاوی اطلاعاتی برای حرکت به سمت بالا و فراتر از شکاف‌های قیمتی و بازده، موثر باشند.

۵-۱- پیشنهادها

در این خصوص و تنها با اتکا به یک مدل مبتنی بر پیش‌بینی نمی‌توان نتیجه گرفت چنین مدل‌هایی مناسب استفاده در بازار سرمایه ایران نیستند، در نتیجه استفاده از پتانسیل‌های مدل‌های پیش‌بینی چند مرحله‌ای نیز باید مورد بررسی قرار گیرند. چندین مسیر احتمالی برای مطالعه می‌تواند پیرو این کار باشد:

۱. بهبود مدل پیشنهادی مبتنی بر پیش‌بینی با اضافه کردن ویژگی‌های بیش‌تر برای پیش‌بینی تغییرات قیمتی به‌جای محدود کردن ویژگی‌ها به قیمت‌های با تاخیر از جمله حجم معاملات
۲. افزایش تکرار داده‌ها (برای نمونه تکرار ۱ دقیقه‌ای) برای اینکه قادر باشیم کاهشی در دوره تشکیل موردنیاز ایجاد کنیم و در نتیجه زوج‌های بیش‌تری را بیابیم.
۳. در کنار استفاده از این الگوریتم‌های پیچیده از الگوریتم‌های ساده‌تر پیش‌بینی از جمله رگرسیون‌ها و مدل‌های خودرگرسیون نیز استفاده شود به این امید که شاید نیازی به پیچیدگی بالا نباشد.

تشکر و قدردانی

از تمام افرادی که در فرآیند نگارش مقاله کمک کرده‌اند، تشکر و قدردانی می‌نمایم.

[1] Gradient-Boosted-Tree (GBT)
[2] Feedforward Neural Networks (FNN)

حمایت مالی

نویسندگان از هیچ‌گونه حمایت مالی استفاده نکرده‌اند.

تعارض با منافع

نویسندگان اعلام می‌دارند هیچ تضادی در منافع در مورد انتشار این نسخه وجود ندارد، همه نویسندگان، نسخه نهایی ارسال‌شده را مشاهده و تایید کرده‌اند. نویسندگان تضمین می‌کنند که مقاله، اثر اصلی آن‌ها بوده، قبلا چاپ نشده، و در حال حاضر تحت انتشار نمی‌باشد.

منابع

[1] Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The review of financial studies*, *19*(3), 797–827. https://doi.org/10.1093/rfs/hhj020

[2] Vidyamurthy, G. (2004). *Pairs Trading: quantitative methods and analysis*. John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=MBYsEAAAQBAJ&oi=fnd&pg=PR9&dq=%5B2%5D++Vidyamurthy,+G.+(2004).+Pairs+Trading:+quantitative+methods+and+analysis.+John+Wiley+%5C%26+Sons.&ots=If9ahwNCEk&sig=BBNAPkhtaR3qaKQesXX_OIV-Pac

[3] Sarmento, S. M., & Horta, N. (2021). *A machine learning based pairs trading investment strategy*. Springer. https://link.springer.com/content/pdf/10.1007/978-3-030-47251-1.pdf

[4] Clegg, M., & Krauss, C. (2018). Pairs trading with partial cointegration. *Quantitative finance*, *18*(1), 121–138. https://doi.org/10.1080/14697688.2017.1370122

[5] Liew, R. Q., & Wu, Y. (2013). Pairs trading: A copula approach. *Journal of derivatives & hedge funds*, *19*(1), 12–30. https://doi.org/10.1057/jdhf.2013.1%0A%0A

[6] Puspaningrum, H., Lin, Y.-X., & Gulati, C. M. (2010). Finding the optimal pre-set boundaries for pairs trading strategy based on cointegration technique. *Journal of statistical theory and practice*, *4*(3), 391–419. https://doi.org/10.1080/15598608.2010.10411994%0A%0A

[7] Rad, H., Low, R. K. Y., & Faff, R. (2016). The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative finance*, *16*(10), 1541–1558. https://doi.org/10.1080/14697688.2016.1164337

[8] Tokat, E., & Hayrullahoglu, A. C. (2022). Length Pairs trading: is it applicable to exchange-traded funds? https://aperta.ulakbim.gov.tr/record/262381/files/bib-de913e70-f289-43f9-9c74-17781f7e96a6.txt?download=1

[9] Blitz, D., Huij, J., Lansdorp, S., & Verbeek, M. (2013). Short-term residual reversal. *Journal of financial markets*, *16*(3), 477–504. https://doi.org/10.1016/j.finmar.2012.10.005

[10] Chen, H., Chen, S., Chen, Z., & Li, F. (2019). Empirical investigation of an equity pairs trading strategy. *Management science*, *65*(1), 370–389. https://doi.org/10.1287/mnsc.2017.2825

[11] Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of financial research*, *35*(2), 261–287. https://doi.org/10.1111/j.1475-6803.2012.01317.x

[12] Jacobs, H., & Weber, M. (2015). On the determinants of pairs trading profitability. *Journal of financial markets*, *23*, 75–97. https://doi.org/10.1016/j.finmar.2014.12.001

[13] Hull, J. C. (2021). *Machine learning in business: An introduction to the world of data science*. Independently Published. https://www.amazon.com/Machine-Learning-Business-Introduction-Science/dp/B095L19RG7?utm_source=chatgpt.com

[14] Sarmento, S. M., & Horta, N. (2020). Enhancing a pairs trading strategy with the application of machine learning. *Expert systems with applications*, *158*, 113490. https://doi.org/10.1016/j.eswa.2020.113490

[15] Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. *ACM sigmod record*, *28*(2), 49–60. https://doi.org/10.1145/304181.304187

[16] Flori, A., & Regoli, D. (2021). Revealing pairs-trading opportunities with long short-term memory networks. *European journal of operational research*, *295*(2), 772–791. https://doi.org/10.1016/j.ejor.2021.03.009

[17] Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one*, *12*(7), e0180944. https://doi.org/10.1371/journal.pone.0180944

[18] Troiano, L., Villa, E. M., & Loia, V. (2018). Replicating a trading strategy by means of LSTM for financial industry applications. *IEEE transactions on industrial informatics*, *14*(7), 3226–3234. https://doi.org/10.1109/TII.2018.2811377

[19] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, *270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

[20] Borovkova, S., & Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. *Journal of forecasting*, *38*(6), 600–619. https://doi.org/10.1002/for.2585

[21] Ullah, S., Ullah, N., Siddique, M. F., Ahmad, Z., & Kim, J.-M. (2024). Spatio-temporal feature extraction for pipeline leak detection in smart cities using acoustic emission signals: a one-dimensional hybrid convolutional neural network--long short-term Memory approach. *Applied sciences*, *14*(22), 10339. https://doi.org/10.3390/app142210339

[22] Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, *5*(2), 157–166. https://doi.org/10.1109/72.279181

[23] Heaton, J. (2018). Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618. *Genetic programming and evolvable machines*, *19*(1), 305–307. https://doi.org/10.1007/s10710-017-9314-z%0A%0A

[24] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *ArXiv preprint arxiv:1406.1078*. https://doi.org/10.48550/arXiv.1406.1078

[25] Tadi, M., Abkar, M., & Motaharinia, V. (2018). Evaluation of pairs trading strategy using distance approach at Tehran stock exchange. *Journal of investment knowledge*, *7*(26), 99–112. (**In Persian**). http://www.jik-ifea.ir/article_12603_3e7095814ff7ba9fb9fcbba2683174d6.pdf?lang=en

[26] Asgari, M., & Abu, Z. (2012). Investigating the effectiveness of pair trading strategy on coin futures contracts by combining stochastic and cointegration approaches. *Third conference on financial mathematics and applications*. Tehran, Iran. Civilica. (**In Persian**). https://civilica.com/doc/352540/

[27] Pakizeh, K., Payjan, K. A., & Salehi, P. (2012). Applying the pair trading strategy in the bahar azadi gold coin futures market. *9th international conference on industrial engineering*. Tehran, Iran. Civilica. (**In Persian**). https://civilica.com/doc/189168/

[28] Fallahpour, S., & Hakimian, H. (2017). Evaluating the Performance of a Pairs Trading System in Tehran Stock Exchange: the Cointegration Approach and Sortino Ratio Analysis. *Journal of financial engineering and securities management*, *8*(30), 1-17. (**In Persian**). https://www.sid.ir/paper/197646/en

[29] Narajewski Michałand Kley-Holsteg, J., & Ziel, F. (2021). tsrobprep—An R package for robust preprocessing of time series data. *SoftwareX*, *16*, 100809. https://doi.org/10.1016/j.softx.2021.100809

[30] Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the econometric society*, 251–276. https://doi.org/10.2307/1913236

[31] Kleinow, T. (2002). *Testing continuous time models in financial markets.* Humboldt university. berlin. https://edoc.hu-berlin.de/server/api/core/bitstreams/2c27a3d4-9df7-483d-8925-3aa77c451edd/content

[32] Chan, E. (2013). *Algorithmic trading: winning strategies and their rationale*. John Wiley & Sons.

[33] Do, B., & Faff, R. (2010). Does simple pairs trading still work? *Financial analysts journal*, *66*(4), 83–95. https://doi.org/10.2469/faj.v66.n4.1

[34] Rudy, J., Dunis, C., Giorgioni, G., & Laws, J. (2010). Statistical arbitrage and high-frequency data with an application to eurostoxx 50 equities. *Available at SSRN 2272605*, 1-31. https://dx.doi.org/10.2139/ssrn.2272605

[35] Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(Nov), 2579–2605. http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

[36] Olivier, A., Jean-Luc, Z., & Maurice, M. (1993). Identification and prediction of non-linear models with recurrent neural network. *International workshop on artificial neural networks* (pp. 531-535). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-56798-4_198

[37] Tenti, P. (1996). Forecasting foreign exchange rates using recurrent neural networks. *Applied artificial intelligence*, *10*(6), 567–582. https://doi.org/10.1080/088395196118434

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی