



<https://amf.ui.ac.ir>

Journal of Asset Management and Financing

E-ISSN: 2383-1189

Vol. 14, Issue 4, No. 55, 2027, p 109-132

Received: 06/08/2025

Accepted: 28/01/2026

Research Paper

Comparison of Two Machine Learning-Based Algorithms for the Two Stages of a Pairs Trading Strategy

Samira Asadian

PhD. Student, Department of Accounting, Faculty of Economics, Management and Accounting, Yazd University, Yazd, Iran

Hojjatollah Sadeqi

Associate Professor, Department of Industrial Management, Faculty of Economics, Management and Accounting, Yazd University, Yazd, Iran
sadeqi@yazd.ac.ir

Dariush Damoori * 

Associate Professor, Department of Accounting and Finance, Faculty of Economics, Management and Accounting, Yazd University, Yazd, Iran
d.damoori@yazd.ac.ir

Abstract

Market participants have shown increasing interest in quantitative trading models and the intelligent application of data science. These approaches can enhance the effectiveness of trading strategies such as pairs trading—a valuable investment method even during bear markets. This strategy comprises two key steps: selecting two securities (pair formation) and detecting anomalies in the price spread between them (trading execution). We evaluate two machine learning approaches to identify the most effective algorithm for each stage. To efficiently group related securities and uncover high-potential pairs—a task that even skilled investors may find challenging—we employ two density-based clustering methods: DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and OPTICS (Ordering Points to Identify the Clustering Structure). These algorithms define the search space by identifying meaningful clusters of securities. To reduce portfolio drawdowns while maximizing returns after detecting price divergences, we implement two time series forecasting models: Long Short-Term Memory (LSTM) and LSTM encoder-decoder. OPTICS outperforms DBSCAN, demonstrating greater efficiency with fewer variables, a higher Sharpe ratio, and a larger proportion of profitable pairs. The LSTM encoder-decoder outperforms the standard LSTM model, delivering higher returns, an improved Sharpe ratio, and fewer days of portfolio decline. This study introduces a novel approach by leveraging machine learning models for both phases of the pairs trading strategy while optimizing model selection. Other distinctive features include the use of high-frequency intraday data (5-minute intervals) across all market stocks (rather than a specific stock or industry), and a focus on net returns after accounting for transaction costs.

Keywords

Pairs Trading Strategy, Machine Learning, Density-Based Clustering, Forecasting-Based Model

Introduction

Pairs trading was first introduced to the financial literature by Gatev et al. (2006) and involves two steps: identifying two securities (a pair) and detecting an irregularity in the price gap between them (taking a trade). Given its proven robustness during market downturns and broad adoption (Do & Faff, 2010; Jacobs & Weber, 2015; Clegg & Krauss, 2018), the application of intelligent data science methods offers a promising path to enhance pairs trading performance. In the pair formation phase, we use two density-based clustering algorithms—DBSCAN and OPTICS—to determine the search space and identify clusters among a large number of securities, a task that is otherwise difficult to perform (Ester et al., 1996; Ankerst et al., 1999). In the trading phase, we employ two time-series prediction models—Long Short-Term Memory (LSTM) and an LSTM encoder-decoder network—to shorten portfolio drawdown periods and boost returns. The key contribution of this research lies in its systematic application of multiple machine learning algorithms across every phase of the strategy, aiming to identify the optimal models and enhance overall performance through predictive analytics. Accounting for transaction costs enables the reporting of realistic net profitability. Furthermore, the use of 5-minute intraday data

*Corresponding author

Asadian, S., Sadeqi, H. and Damoori, D. (2027). Comparison of two machine learning-based algorithms for two parts of pairs trading strategy. *Journal of Asset Management and Financing*, 14(4), 109-132.



2383-1189 © University of Isfahan

This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



10.22108/amf.2026.146240.2006

significantly improves the accuracy of our findings.

Materials & Methods

The data consist of 111 stocks traded on the Tehran Stock Exchange and the IranFaraBourse over the period 2014 to 2019. To ensure liquidity conditions consistent with the criteria adopted by Gatev et al. (2006) and Do & Faff (2012), data imputation was performed using the tsrobbrep open-source package in R software. To create meaningful clusters, we first performed dimensionality reduction using principal component analysis (PCA), followed by the application of DBSCAN and OPTICS clustering algorithms. Subsequently, four conditions were applied: (1) the components of each pair are co-integrated; (2) the Hurst component of the pair's price gap exhibits mean-reverting properties; (3) the price gap of each pair shows convergence and divergence over appropriate periods; and (4) the mean reversion of the pair's price gap occurs with sufficient frequency. The data were divided into a formation period and a trading period. The work includes three different test portfolios of securities: Portfolio 1 includes all pairs identified during the formation period; Portfolio 2 includes pairs with positive results; and Portfolio 3 is limited to investing in the 10 pairs with the highest returns in the validation set. To compare the robustness of the standard threshold-based model with the proposed prediction-based model, simulations were carried out using the LSTM and LSTM encoder-decoder models. All pairs in the portfolio were equally weighted. The code developed for this work was implemented in Python.

Findings

The first step in applying the proposed approach to stock clustering involves defining an appropriate number of principal components to reduce the dimensionality of the original data, which was set to five. OPTICS offers two distinct advantages over DBSCAN: (1) it effectively handles data clusters of varying densities, and (2) it eliminates the challenge of determining the optimal ε parameter. Analysis of the selected pairs shows that this method yields a promising portfolio with a satisfactory profit margin and a high average Sharpe ratio of 10.57. The results of the prediction-based trading models (LSTM and LSTM encoder-decoder) using the OPTICS algorithm for the period from April 2017 to March 2019 are presented in the table 1, alongside the results for the standard threshold-based model. In the standard trading model, the portfolio drawdown is significantly shorter; however, the drawdown in the LSTM encoder-decoder model is adequate and substantially shorter than that of the other model. The longer drawdown period for the standard LSTM is 190 days. The portfolio performs considerably better in terms of total return on investment (21,993.8%) and Sharpe ratio (35.50). The superior results from the LSTM encoder-decoder suggest that multi-stage prediction is effective in these settings.

Table 1. Comparing trading results using a 3-year formation period

Trading model	standard	LSTM-based model	LSTM encoder-decoder-based model
parameter	$\alpha_S = \mu_S + 2\sigma_S$ $\alpha_L = \mu_S - 2\sigma_S$	$\alpha_S = Q_{f-(0.10)}$ $\alpha_L = Q_{f+(0.90)}$	$\alpha_S = Q_{f-(0.20)}$ $\alpha_L = Q_{f+(0.80)}$
validation			
Drawdown numbers	2	12	12
SR	6.7	2.18	-0.51
ROI(%)	7911.1	29817.4	616.28
MDD(%)	8.43	33.97	170.24
Trading numbers(+, -)	46(12-34)	9044(8983-62)	9829(9749-80)
Profitable pairs (%)	100	22.7	9.1
test			
Drawdown numbers	10	190	29
SR	0.53	2.18	50.35
ROI(%)	1918.9	-90.12	21993.8
MDD(%)	23.13	99.80	79.17
Trading numbers(+, -)	25(12-13)	1196(1162-34)	150(143-7)
Profitable pairs (%)	86.4	0	100.0

Discussion & conclusion

In comparing the results of this study with those of others, the work of Sarmiento and Horta (2020) introduced advanced predictive modeling techniques—namely LSTM and its encoder-decoder variant—to the field of US ETF trading analysis. While these models successfully shortened portfolio drawdown periods, they also reduced overall returns. In their study, the traditional securities portfolio significantly outperformed the models in both total return and Sharpe ratio. Therefore, key features of the Iranian capital market's structure—such as its price fluctuation limits, base volume trading rules, and the surge in retail investor participation during 2019–2020—may be the primary reasons for the differing outcomes observed in the present study. Scientifically, however, this outcome is better explained by the limited adoption of machine learning models in the Iranian capital market. Even the use of classical algorithmic trading remains minimal, making the application of advanced models exceedingly rare. Flori and Regoli (2021) employed LSTM to analyze nonlinear interdependencies within and between financial time series for forecasting S&P 500 stock market performance. Their LSTM framework's predictive accuracy was benchmarked against several common machine learning models, including logistic regression, gradient boosted trees, and feedforward neural networks. The purpose of the LSTM application was to forecast the probability of a stock generating excess returns compared to similar stocks in the short term. Their analysis indicates that investment strategies incorporating such forecasts are effective for improving portfolio performance, as they generate predictive signals with exploitable information beyond basic price-yield gaps.

مقایسه دو الگوریتم مبتنی بر یادگیری ماشینی برای دو بخش راهبرد معاملات زوجی

سمیرا اسدیان

دانشجوی دکتری، گروه حسابداری، دانشکده اقتصاد مدیریت و حسابداری، دانشگاه یزد، یزد، ایران

asadian.samira@gmail.com

حجت‌الله صادقی

دانشیار، گروه مدیریت صنعتی، دانشکده اقتصاد مدیریت و حسابداری، دانشگاه یزد، یزد، ایران

sadegi@yazd.ac.ir

دارپوش دموری 

دانشیار، گروه حسابداری و مالی، دانشکده اقتصاد مدیریت و حسابداری، دانشگاه یزد، یزد، ایران

d.damoori@yazd.ac.ir

چکیده

باتوجه به علاقه مشارکت‌کنندگان بازار به مدل‌های معاملاتی کمی و استفاده از کاربردهای هوشمندانه علم داده می‌توان عملکرد راهبردهای معاملاتی از جمله معاملات زوجی را بهبود بخشید که ابزار سرمایه‌گذاری خوبی حتی در بازارهای نزولی است. با در نظر گرفتن دو مرحله این راهبرد شامل شناسایی دو ورقه بهادار (یک زوج) و کشف بی‌نظمی در شکاف قیمتی بین آنها (انجام معامله)، دو روش یادگیری ماشینی برای انتخاب الگوریتم بهتر در هر مرحله مقایسه شد. برای تعیین فضای جست‌وجو که سعی بر گروه‌بندی اوراق بهادار مرتبط در خوشه‌ها و شناسایی زوج‌های برتر دارد، از دو الگوریتم خوشه‌بندی مبتنی بر چگالی شامل خوشه‌بندی فاصله‌ای مبتنی بر چگالی کاربردها با نوفه (دی‌بی‌اسکن) و مرتب‌سازی نقاط برای شناسایی ساختار خوشه‌بندی (آپتیکز) استفاده شد که در غیر این صورت شناسایی آن حتی برای سرمایه‌گذار باتجربه هم دشوار است. از طرفی برای به حداقل رساندن تعداد روزهای کاهش ارزش سبد به همراه افزایش بازده پس از یافتن شکاف قیمتی، از دو مدل معاملاتی مبتنی بر پیش‌بینی سری‌زمانی، حافظه کوتاه‌مدت طولانی (الاس‌تی‌ام) و حافظه کوتاه‌مدت طولانی رمزگذار - رمزگشا استفاده شد. برتری آپتیکز به دی‌بی‌اسکن در تعداد متغیرهای کمتر و نسبت شارپ بالا و درصد زوج‌های سودآور بالاتر مشهود است. مدل الاس‌تی‌ام رمزگذار - رمزگشا در کل در کسب بازدهی و نسبت شارپ، عملکرد بهتر و درباره کاهش تعداد روزهای رکودی سبد بالاتر از مدل الاس‌تی‌ام بود. استفاده از داده‌های با تکرار زیاد (۵ دقیقه‌ای) (تمامی سهام بازار و نه سهم یا صنعت خاصی) به همراه استفاده از مدل‌های یادگیری ماشینی برای افزایش سرعت در تصمیم‌گیری و ارائه بازده‌های خالص به منظور بررسی مدل‌ها از موارد مهم این کار است.

واژه‌های کلیدی: راهبرد معامله زوجی، یادگیری ماشینی، خوشه‌بندی مبتنی بر چگالی، مدل مبتنی بر پیش‌بینی.

طبقه‌بندی JEL: C53، C61، G11، G12

* نویسنده مسئول

اسدیان، سمیرا، صادقی، حجت‌الله و دموری، دارپوش. (۱۴۰۵). مقایسه دو الگوریتم مبتنی بر یادگیری ماشینی برای دو بخش استراتژی معاملات زوجی، مدیریت دارایی و تأمین مالی، ۱۴ (۴)، ۱۰۹-۱۳۲.



مقدمه

راهبرد معامله زوجی در ادبیات مالی ابتدا توسط گتو و همکاران (Gatev et al., 2006) معرفی شد که شامل مراحل شناسایی دو ورقه بهادار (یک زوج) و سپس کشف بی‌نظمی در شکاف قیمتی بین آن دو جزء (انجام معامله) است. باتوجه به نتایج یافته‌های مطالعات پیشین (Clegg & Krauss, 2018; Jacobs & Weber, 2015; Do & Faff, 2010) درخصوص عملکرد خوب معامله زوجی در زمان تنگنای مالی و استفاده گسترده از آن، دیگر با روش‌های عادی اقتصادسنجی نمی‌توان به بازدهی مناسبی دست یافت؛ از این رو استفاده از کاربردهای هوشمندانه علم داده برای بهبود عملکرد این راهبرد معاملاتی در هر دو مرحله آن پیشنهاد می‌شود.

برای تعیین فضای جست‌وجو و خوشه‌بندی در مرحله تشکیل زوج‌ها باتوجه به ویژگی‌های مناسب الگوریتم‌های مبتنی بر چگالی از جمله اینکه، این خوشه‌ها می‌تواند شکل‌های دلخواهی داشته و از این رو نیازی به هیچ فرض گاوسی درباره شکل داده‌ها نیست، عملکرد مناسب آنها در نپرداختن به داده‌های پرت و عدم نیاز به هیچ مشخصاتی درخصوص تعداد خوشه‌ها، از دو الگوریتم دی‌بی‌اسکن و آپتیکز استفاده شد که در غیر این صورت، شناسایی خوشه‌ها از میان تعداد زیادی اوراق بهادار حتی برای سرمایه‌گذار باتجربه هم دشوار است (Ester et al., 1996; Ankerst et al., 1999).

در مرحله معاملات برای به حداقل رساندن تعداد روزهای کاهش ارزش سبد اوراق بهادار پس از یافتن شکاف قیمتی، از دو مدل معاملاتی مبتنی بر پیش‌بینی سری‌زمانی (مدل‌های ناپارامتریک) استفاده شد. ابتدا شبکه‌های حافظه کوتاه‌مدت طولانی^۱ (ال‌اس‌تی‌ام) استفاده شد که نوعی از شبکه عصبی بازگشتی^۲ (آر‌ان‌ان) مطرح شده توسط هوخرایتر و اش‌میدهور (Hochreiter & Schmidhuber, 1997) با هدف حل برخی معایب خاص آر‌ان‌ان یعنی مشکل ناپدیدشدن/ انفجار شیب است که مانع از یادگیری وابستگی‌های بلندمدت با توالی توسط شبکه‌های بازگشتی می‌شد. شبکه‌های آر‌ان‌ان نه تنها داده‌های انفرادی نقطه‌ای، تمامی توالی داده‌ها را می‌تواند پردازش کند (Goodfellow et al., 2017). این یعنی ممکن است حالات قبلی بر تصمیم شبکه عصبی در یک نقطه زمانی متفاوت اثرگذار باشد که با حلقه‌های داخلی به دست می‌آید و آر‌ان‌ان‌ها را مجاز می‌کند از محاسبات پیشین خود حافظه‌ای داشته باشند.

سپس از شبکه‌های حافظه کوتاه‌مدت طولانی رمزگذار – رمزگشا^۳ استفاده شد؛ زیرا از دیدگاه معاملاتی، ممکن است استفاده از اطلاعات باتوجه به پیش‌بینی نه تنها لحظه زمانی بعدی، برای مراحل زمانی پس از آن نیز مفید باشد. برای انجام این مورد امکان دارد با استفاده از داده‌های جمع‌آوری شده تا زمان t جهت پیش‌بینی مستقیم N مرحله زمانی $(t + 1, \dots, t + N)$ به صورت یک‌سره بهره برد. یک پیشرفت در این کار بدین صورت است که نه تنها از داده‌های جمع‌آوری شده تا زمان t ، از داده‌های پیش‌بینی شده جدید نیز می‌توان بهره برد. سبکی از ال‌اس‌تی‌ام رمزگذار – رمزگشا به طور طبیعی متناسب با چنین سناریوهایی است. این سبک از دو گروه ال‌اس‌تی‌ام، یکی برای خواندن دنباله ورودی و رمزگذاری آن به یک بردار با طول ثابت، رمزگذار و دومی برای رمزگشایی بردار با طول ثابت و خروجی از دنباله پیش‌بینی شده، رمزگشا تشکیل شده است. دو شبکه به صورت مشترک آموزش می‌بینند تا زیان پیش‌بینی باتوجه به یک دنباله هدف ناشی از دنباله منبع حداقل شود.

برای کسب بازدهی بالاتر و کاهش تعداد روزهایی که بازدهی سبد سرمایه‌گذاری منفی است، نیاز به استفاده از راهبردهای معاملاتی با سرعت زیاد به همراه پیش‌بینی است. با عمق‌گیری هرچه بیشتر بازار سرمایه، دیگر با استفاده از مدل‌های کلاسیک نمی‌توان عملکرد خوبی ارائه کرد؛ در نتیجه نیاز امروزه هر سرمایه‌گذار حرفه‌ای و مدیر دارایی به استفاده از روش‌های ساده‌تر

¹ Long-Short Term Memory (LSTM)

² Recurrent Neural Network (RNN)

³ LSTM encoder-decoder

(ناپارامتریک) و کاراتر (مبتنی بر پیش‌بینی) است. این پژوهش به کاربرد و توسعه روش‌شناسی بخشی از کار سارمنتو و هورتا (Sarmiento & Horta, 2020) در بازار سرمایه ایران می‌پردازد. برای دو بخش راهبرد معاملات زوجی به آزمون دو الگوریتم برای یافتن بهترین آن پرداخته شده است؛ در بخش اول به بررسی ساخت سازه‌هایی اندازه‌پذیر با دو الگوریتم خوشه‌بندی فاصله‌ای مبتنی بر چگالی کاربردها با نوفه^۱ (دی‌بی‌اسکن) و مرتب‌سازی نقاط برای شناسایی ساختار خوشه‌بندی^۲ (آپتیکز) پرداخته شد. سپس در بخش معاملات از دو الگوریتم ال‌اس‌تی‌ام و ال‌اس‌تی‌ام - رمزگذار - رمزگشا برای یافتن بهترین مدل در پیش‌بینی برای بخش معاملات راهبرد استفاده شد.

وجه تمایز این پژوهش به استفاده از چند الگوریتم یادگیری ماشینی در تمامی مراحل این راهبرد برای یافتن بهترین آن‌ها و کمک به بهبود عملکرد راهبرد با کمک پیش‌بینی است؛ استفاده از هزینه‌های معاملاتی باعث می‌شود بازده‌های خالص حاصل شود. استفاده از داده‌های بین‌روزی (تواتر ۵ دقیقه‌ای) بر کیفیت خروجی افزوده است. مقاله حاضر بدین صورت ادامه می‌یابد: بخش دوم به مبانی نظری دو الگوریتم خوشه‌بندی مبتنی بر چگالی و مدل‌های پیش‌بینی برای معاملات می‌پردازد؛ بخش سوم روش پژوهش، بخش چهارم یافته‌ها، بخش پنجم نتایج و پیشنهادها را ارائه می‌دهد و در انتها منابع ارائه شده است.

مبانی نظری

سرمایه‌گذاری به صورت خرید/فروش هم‌زمان سهام مربوط می‌شود به اواخر دهه ۱۹۴۰ که سرمایه‌گذاری آلفرد وینزلو جونز^۳ سهام را می‌خرد و هم‌زمان فروش استقراضی انجام می‌داد. بعداً توسط نانزیو تارتاگلیا^۴ در مورگان استنلی و اواخر دهه ۱۹۸۰ اصلاح شد. این راهبرد به صورت گسترده‌ای در ادبیات جاری (Enders & Granger, 1998; Dunis & Ho, 2005; Vidyamurthy, 2004; Lin et al., 2006; Khandani & Lo, 2008) استناد شده است. درحقیقت این راهبردها میان مدیران سرمایه‌گذاری از زمان سیر نزولی ۲۰۰۰ خیلی محبوب شد؛ زیرا مشخصه اصلی آنان در صورتی که به درستی ایجاد و اجرا شود، این است که رفتار بازار سهام پایه نباید بر نتایج سبدها اثری بگذارد؛ یعنی بازده‌های ایجادشده با یک سبدی سهام بازار - خنثی باید مستقل از بازده‌های کلی بازار سهام باشد. روش معامله زوجی، تکنیک مهم آربیتراژ آماری و یکی از مشهورترین راهبردهای سفته‌بازی کوتاه‌مدت است که با توجه به عملکرد خوب آن در زمان تنگنای مالی، از سوی صندوق‌های پوشش ریسک بسیار استفاده می‌شود (Vidyamurthy, 2004; Gatev et al., 2006)؛ اما عمده مطالعات در حوزه استفاده از یادگیری ماشینی در معاملات زوجی به مطالعات هاک (Huck, 2009/2010) برمی‌گردد. تا جدیدترین آن که یک چارچوب یادگیری عمیق سلسله‌مراتبی است و به ترکیب یادگیری نظارت‌شده و یادگیری تقویتی عمیق^۵ (دی آر ال) می‌پردازد. این چارچوب از دو لایه تشکیل شده است: یک لایه بالایی که از یادگیری نظارت‌شده با شبکه‌های توجه گراف^۶ (گت) برای کشف روابط پیچیده دارایی‌ها و انتخاب جفت‌های بهینه از مجموعه دارایی‌ها استفاده می‌کند؛ یک لایه پایینی که از دی آر ال برای اجرای راهبردهای معاملاتی خودکار و پویا برای جفت‌های انتخاب‌شده استفاده می‌کند. در بورس نزدیک انجام شد که نتایج تجربی، بهبودهای چشمگیری را در مقایسه با راهبردهای سنتی نشان می‌دهد (Xia et al., 2025).

- الگوریتم‌های یادگیری ماشینی برای دوره تشکیل

¹ Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

² Ordering Points To Identify the Clustering Structure (OPTICS)

³ Alfred Winslow Jones

⁴ Nunzio Tartaglia

⁵ Deep Reinforcement Learning (DRL)

⁶ Graph Attention Networks (GAT)

درخصوص انتخاب یک نوع مناسب خوشه‌بندی از میان سه روش خوشه‌بندی تقسیمی، تاریخی و مبتنی بر چگالی، مورد آخر دارای چندین مزایا است: خوشه‌ها می‌توانند شکل‌های دلخواهی داشته باشند و از این‌رو نیازی به هیچ فرض گاوسی درباره شکل داده‌ها نیست؛ ۲) دارای عملکرد قوی درخصوص داده‌های پرت است؛ زیرا هیچ الگوریتمی به گروه‌بندی هر نقطه در مجموعه داده‌ها نمی‌پردازد و ۳) به هیچ مشخصاتی درخصوص تعداد خوشه‌ها نیاز نیست؛ در نتیجه دو نوع خوشه‌بندی مبتنی بر چگالی از یادگیری نظارت‌نشده برای این کار در نظر گرفته شد.

یک الگوریتم خوشه‌بندی مبتنی بر چگالی به تفسیر فضای خوشه‌بندی به عنوان مجموعه‌ای باز در فضای اقلیدسی می‌پردازد که می‌تواند به مجموعه‌ای از اجزای به هم مرتبط تقسیم شود. الگوریتم دی‌بی‌اسکن مربوط به استر و همکاران (Ester et al., 1996) تأثیرگذارترین آنها در این دسته است. مطابق نظر نویسنده، دلیل اینکه چرا به راحتی خوشه‌های نقاط شناسایی می‌شود این است که یک چگالی معمولی از نقاط در هر خوشه وجود دارد که به طور درخورتوجهی بیشتر از [سایر نقاط] خارج از خوشه است. ایده اصلی متکی بر مفهوم چگالی و اتصال است و دو پارامتر ϵ (فاصله) و $\min Pts$ (تعداد نقاط) در مدل وجود دارد. اجرای دی‌بی‌اسکن به روشی ساده را می‌توان این گونه بیان کرد: ۱) یافتن نقاطی در همسایگی ϵ - هر نقطه و تشخیص نقاط مرکزی با بیش از $\min Pts$ همسایه؛ ۲) یافتن اجزای متصل شده نقاط مرکزی در گراف همسایه، فارغ از تمامی نقاط غیر مرکزی و ۳) تخصیص هر نقطه غیر مرکزی به یک خوشه نزدیک، چنانچه خوشه یک همسایه ϵ - باشد؛ در غیر این صورت تخصیص آن به نوفه با وجود مزایای دی‌بی‌اسکن، برخی اختراها را نیز باید مدنظر قرار داد. ابتدا الگوریتم، بسیار به پارامترها از جمله تعریف ϵ حساس است؛ علاوه بر این، حتی پس از یافتن یک ϵ مناسب، الگوریتم فرض می‌کند خوشه‌ها به صورت یکنواخت متراکم است. این مورد منتج به ناتوانی در یافتن خوشه‌هایی با چگالی متفاوت می‌شود. انتظار وجود خوشه‌های با چگالی نامتوازن به عنوان کاتالستی برای حرکت به سمت آپتیکز را فراهم می‌آورد. آپتیکز که از جانب انکرست و همکاران (Ankerst et al., 1999) پیشنهاد شده، مشکل شناسایی خوشه‌های معنادار در داده‌های با چگالی متفاوت را مورد خطاب قرار می‌دهد. این الگوریتم که از دی‌بی‌اسکن الهام گرفته شده، اجرای آن اساساً با دی‌بی‌اسکن برابر است؛ با این استثنا که به جای حفظ یک مجموعه شناخته شده که قبلاً اعضای خوشه پردازش نشده است، از یک صف مقدم^۱ استفاده می‌کند. این الگوریتم نقاط و فاصله‌های قابل دسترسی مربوط را بازمی‌گرداند. این نقاط به روشی سفارش داده می‌شوند که از لحاظ فضایی، نزدیک‌ترین نقاط در همسایگی یکدیگر باشند. رویه استخراج خوشه‌ها در این الگوریتم عمدتاً به دو روش انجام می‌شود: ۱) بررسی طرح قابلیت دسترسی و تثبیت یک ϵ مناسب. بدین روش، هر مقدار انتخابی به هر خوشه در مجموعه داده‌ها تحمیل می‌شود که به راحتی منتج به همان نتایجی می‌شود که در کاربرد دی‌بی‌اسکن با همان مقدار برای ϵ و $\min Pts$ حاصل شد؛ ۲) مجاز کردن رویه‌ای خودکار برای انتخاب مناسب‌ترین ϵ به ازای هر خوشه که با تعریف حداقل شیب برای طرح قابل دسترسی اجرا می‌شود و یک مرز برای خوشه تشکیل می‌دهد. رویه مذکور، مشکل مربوط به یافتن مناسب‌ترین ϵ را برای هر خوشه مدنظر قرار می‌دهد. به عنوان یک محصول فرعی، از آنجاکه سرمایه‌گذار تنها نیاز به تعیین $\min Pts$ دارد، فرایند خوشه‌بندی را به [فرایندی] بدون پارامتر نزدیک‌تر می‌کند.

– الگوریتم‌های یادگیری ماشینی برای دوره معاملات

در بررسی کاربردی بودن یک راهبرد معاملاتی مبتنی بر پیش‌بینی سری‌های زمانی و وجود مدل‌های متفاوتی در ادبیات موضوع، از مدل‌سازی ناپارامتریک شبکه‌های عصبی مصنوعی^۲ (ای‌ان‌ان) استفاده شد؛ زیرا هیچ فرض زمینه‌ای درخصوص فرایند ندارد. ای‌ان‌ان نمونه‌ای از مدل‌های ناپارامتریک است که نه تنها کاربرد آنها را در این زمینه به مطالعه موردی جذاب تبدیل می‌کند، نتایج بسیار امیدوارکننده‌ای در پیش‌بینی داده‌های مربوط به سری‌های زمانی مالی نشان داده است (Cavalcante et al., 2016).

¹ Priority queue

² Artificial Neural Network (ANN)

شبکه‌های ال‌اس‌تی‌ام برای کاربردهای مالی (Bao et al., 2017; Troiano et al., 2018; Fischer & Krauss, 2018) Borovkova & Tsiamas, 2019) در نظر گرفته شده است. آنها نوعی از شبکه عصبی بازگشتی^۱ (آر ان ان) هستند که ابتدا توسط هوخرایتر و اشمیدهور (Hochreiter & Schmidhuber, 1997) با هدف حل برخی معایب خاص آر ان ان یعنی مشکل ناپدید شدن/ انفجار شیب که مانع از یادگیری وابستگی‌های بلندمدت با توالی توسط شبکه‌های بازگشت کننده می‌شد، معرفی شد (Bengio et al., 1994). مطابق نظر گودفلو و همکاران (Goodfellow et al., 2017) شبکه‌های آر ان ان نه تنها داده‌های انفرادی نقطه‌ای، تمامی توالی داده‌ها را می‌توانند پردازش کنند. این یعنی ورودی‌ها مستقل از یکدیگر در نظر گرفته می‌شوند؛ درباره یک آر ان ان فرض می‌شود وابستگی متوالی میان ورودی‌ها وجود دارد و ممکن است حالات قبلی بر تصمیم شبکه عصبی در یک نقطه زمانی متفاوت اثرگذار باشد. این مورد با حلقه‌های داخلی به دست می‌آید که آنها را مجاز می‌کند تا از محاسبات پیشین خود حافظه‌ای داشته باشد. یک مشکل برخاسته از سبک آر ان ان مربوط به مشکل انتشار^۲ خطای پیش‌بینی است؛ در این خصوص، پس انتشار از خطای نهایی به سمت خروجی‌ها، وزن‌ها و ورودی‌های هر لایه مخفی با در نظر گرفتن آن اوزان برای بخشی از خطا، و محاسبه مشتقات جزئی آنان $\frac{\partial E}{\partial w_i}$ که E بیانگر تابع خطاست، حرکت به سمت عقب دارد. پس از آن مشتقات برای تعدیل اوزان در هر جهت که خطا را کاهش دهد، استفاده می‌شود. به هر حال، هنگام انتشار خطا در کل بعد زمانی شبکه، بسته به اینکه آیا مشتق تابع به روزرسانی، حالت خطای پس انتشار را بسیار شدید کاهش یا افزایش می‌دهد، امکان دارد شیب محو شود یا از بین برود. در مورد اول، بسیار طول می‌کشد تا فرایند یادگیری کامل شود؛ در حالی که مورد دوم دارای مشکلات عددی است. از نظر رسمی، معادلات شرح دهنده سلول ال‌اس‌تی‌ام را می‌توان این گونه خلاصه کرد:

$$\begin{cases} f_t = \sigma(W_x^f x_t + W_h^f h_{t-1} + b_f) \\ i_t = \sigma(W_x^i x_t + W_h^i h_{t-1} + b_i) \\ o_t = \sigma(W_x^o x_t + W_h^o h_{t-1} + b_o) \\ c_t = f_t \cdot c_{t-1} + i_t \cdot \sigma(W_x^c x_t + W_h^c h_{t-1} + b_c) \\ h_t = o_t \cdot \sigma(c_t) \end{cases} \quad (1)$$

که f_t , i_t و o_t به ترتیب بیانگر دروازه فراموشی، دروازه ورودی و دروازه خروجی است. با پذیرش انتخاب اطلاعات جدید و پاک کردن اطلاعات گذشته، این مکان‌شناسی قادر است جریان خطا را در طول شبکه حفظ کند. تعداد پارامترهای ال‌اس‌تی‌ام بیشتر از یک آر ان ان ساده است. به طور خاص برای یک ال‌اس‌تی‌ام تک‌لایه‌ای با m ویژگی ورودی و n گره مخفی، تعداد پارامترها $4(n(m+1) + n^2)$ است. لایه خروجی $K(n+1)$ پارامتر دیگر را در نظر می‌گیرد که K تعداد کلاس‌هاست. توجه داشته باشید که عمق دنباله زمانی بر پیچیدگی مدل اثرگذار نیست؛ زیرا پارامترها در تمام شبکه‌های آر ان ان طی زمان تغییرناپذیرند؛ وابستگی زمانی کاملاً توسط ماهیت خودرگرسیو وضعیت‌های پنهان توضیح داده می‌شود (Flori & Regoli, 2021).

در استفاده از ال‌اس‌تی‌ام رمزگذار - رمزگشا با پیش‌بینی مستقیم N مرحله زمانی $(t+1, \dots, t+N)$ و دو سبک، یکی برای خواندن دنباله ورودی و رمزگذاری آن به یک بردار با طول ثابت، رمزگذار و دومی برای رمزگشایی بردار با طول ثابت و خروجی از دنباله پیش‌بینی شده، رمزگشا، دو شبکه به صورت مشترک آموزش می‌بینند تا زیان پیش‌بینی باتوجه به یک دنباله هدف ناشی از دنباله منبع حداقل شود (Sarmiento & Horta, 2021). در این سناریوی پیش‌بینی چندمرحله‌ای، قوانین معاملاتی باید به روزرسانی شود. تغییر پیش‌بینی باید N بار از پیش محاسبه شود. به همین ترتیب آستانه‌های خرید و فروش باید باتوجه به توزیع تغییر درصدی بین $x(t)$ و $x(t-N)$ محاسبه شود.

¹ Recurrent Neural Network (RNN)

² propagating

در مرور ادبیات معامله زوجی در ایران، از رویکرد فاصله برای پیاده‌سازی راهبرد 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 شد (Tadi et al., 2018). تعیین میزان آستانه بهینه نیز یکی دیگر از اهداف این پژوهش بوده است. نتایج حاصل نشان داد پیاده‌سازی این راهبرد در سطوح خاصی از میزان آستانه (۰.۳ تا ۰.۹) رویکرد مدنظر پژوهش از کارایی لازم برخوردار است. در مطالعه‌ای دیگر به بررسی اثربخشی استفاده از راهبرد در معاملات آتی سکه پرداخته شد (Asgari & Abu, 2012). ابتدا با مدل تعمیم‌یافته ریشه واحد تصادفی^۱ (جی‌استور) تغییرات سری قیمت‌ها و شکاف قیمتی از لحاظ ایستایی و وجود ریشه واحد و سپس وجود رابطه هم‌انباشتگی بررسی شد. نتایج حاکی از آن است که سری شکاف قیمتی آتی سکه دارای خاصیت بازگشت به میانگین قوی بوده است و تلاطم به نسبت بالایی را نشان می‌دهد. پاکیزه و همکاران (Pakizeh et al., 2012) کاربرد راهبرد معاملات زوجی را در بازار قراردادهای آتی سکه طلای بهار آزادی بررسی کردند. نتیجه حاصل نیز کاربرد این راهبرد را در بازار مدنظر تأیید کرد.

رمضانی‌فر و همکاران (Ramezanifar et al., 2015) رویکرد هم‌انباشتگی فرکتالی را در راهبرد معاملات زوجی و عملکرد آن در قیاس با رویکرد هم‌انباشتگی روی داده‌های مربوط به قیمت‌های دقیقه‌ای بین‌روزی نفت خام (WTI) و نفت خام سبک (LCO) برای دو هفته از سال ۲۰۱۱ بررسی کردند. ابتدا دو سبد اوراق بهادار، FCI (برای رویکرد هم‌انباشتگی فرکتالی و با استفاده از مدل تصحیح خطای بردار فرکتالی^۲ (FVECM)) و CI (برای رویکرد هم‌انباشتگی و با استفاده از مدل تصحیح خطای برداری^۳ (VECM)) تشکیل شد. نتایج حاکی از آن است که رویکرد هم‌انباشتگی فرکتالی از نظر آماری به بازده‌های بالاتری در مقایسه با رویکرد هم‌انباشتگی منتهی می‌شود. تعداد معاملات برای رویکرد هم‌انباشتگی فرکتالی کمتر است که به دلیل عملکرد راهبرد است؛ در نهایت، در سطح تعدیل‌شده با ریسک، رویکرد هم‌انباشتگی فرکتالی دارای نسبت سورتینوی بالاتری در مقایسه با رویکرد هم‌انباشتگی بود.

فلاح‌پور و حکیمیان (Fallahpour & Hakimian, 2017) با محاسبه و بررسی بازده و نسبت سورتینو، عملکرد سیستم معاملات زوجی با استفاده از رویکرد هم‌انباشتگی در بورس اوراق بهادار تهران را بررسی کردند. نتایج آزمایش با داده‌های بین‌روزی (۵ دقیقه‌ای) روی زوج سهام منتخب نشان می‌دهد که استفاده از سیستم معاملات زوجی به عنوان 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 بوده است. از طرفی نتایج حاصل از این پژوهش نشان‌دهنده عملکرد چشمگیر راهبرد معاملات زوجی در قیاس با معاملات سهام و دیگر راهبردهای معاملات الگوریتمی به کارگرفته شده در پژوهش‌های پیشین است.

روش پژوهش

جامعه داده‌های پژوهش حاضر، سهام معامله‌شونده در بورس اوراق بهادار تهران و فرابورس ایران است. تعداد نمادهای مشترک معامله‌شونده از ابتدای سال ۱۳۹۳ تا انتهای سال ۱۳۹۹، ۲۳۱ نماد (غیر از عرضه‌های اولیه) بود که برخی از نمادها به دلیل در دسترس نبودن داده‌های کامل آنها و بسیاری از آنها به دلیل تعداد بالای روزهای بسته‌بودن نماد (بیش از ۵۰ درصد بازه زمانی مدنظر) حذف شد؛ در نهایت ۱۱۱ نماد مختلف، تأییدی بر شرایطی است که تشریح شد. باتوجه به شرط نقدشوندگی منطبق با معیار اتخاذی گتو و همکاران (Gatev et al., 2006) و دو و فاف (Do & Faff, 2012) که تمامی سهام با حداقل یک روز بدون معامله باید حذف شود و نیز امکان وجود داده‌های پرت در مجموعه داده، انتساب داده و پاک‌سازی

¹ Generalised Stochastic Unit Root (GSTUR)

² Fractional Vector Error Correction Model

³ Vector Error Correction Model

داده‌ها 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 است. برای انتساب داده‌ها یک مدل جایگزینی احتمالی پیشنهاد شد که ممکن است از اجزای خودرگرسیون و ورودی‌های خارجی تشکیل شده باشد و برای تشخیص داده پرت، یک الگوریتم خوشه‌بندی مبتنی بر مدل‌سازی ترکیبی محدود معرفی شده است که ویژگی‌های سری زمانی را برحسب شیب و فصلی بودن زیربنایی به‌عنوان ویژگی‌ها در نظر می‌گیرد (Narajewski et al., 2021).

در استفاده از الگوریتم یادگیری نظارت‌نشده برای ایجاد خوشه‌های معنادار دارایی‌ها برای انتخاب زوج‌ها ابتدا با تحلیل مؤلفه‌های اصلی^۱ (پی سی ای) کاهش بُعد صورت پذیرفت. سپس از الگوریتم‌های خوشه‌بندی دی‌بی‌اسکن و آپتیکز استفاده شد؛ درنهایت زوج‌ها انتخاب شد؛ اما با ایجاد خوشه‌هایی برای دارایی‌ها، هنوز ضروری است مجموعه شرایطی برای انتخاب زوج‌های معاملاتی تعریف شود تا تعادل زوج‌ها ادامه یابد. بدین‌منظور زمانی که چهار شرط زیر برآورده شود، یک زوج انتخاب می‌شود: (۱) اجزای زوج هم‌انباشته باشد؛ (۲) جزء هرست مربوط به شکاف قیمتی زوج دارای ویژگی بازگشت به میانگین باشد؛ (۳) شکاف قیمتی مربوط به زوج در دوره‌های مناسب دارای هم‌گرایی و واگرایی باشد و (۴) بازگشت به میانگین مربوط به شکاف قیمتی زوج دارای فراوانی کافی باشد.

برای هر شبیه‌سازی معاملاتی باید داده‌ها به دو دوره تقسیم شود: دوره تشکیل و دوره معاملات. دوره تشکیل به شبیه‌سازی داده‌های در دسترس سرمایه‌گذار پیش از ورود به هر معامله‌ای می‌پردازد و برای یافتن جذاب‌ترین زوج‌های کاندید استفاده می‌شود و برای مدل‌های مبتنی بر پیش‌بینی، بخشی از داده‌های دوره تشکیل، مجموعه آموزش، برای آموزش الگوریتم پیش‌بینی استفاده می‌شود. دوره آموزش برای این مورد که آیا مدل معاملاتی اجرا شده با داده‌های مشاهده‌نشده آتی شبیه است یا خیر، به کار می‌رود. اساساً پیکربندی شبیه‌سازی شامل دوره تشکیل ۳ ساله است که از سال دوم تا آخر برای اعتبارسنجی عملکرد پیش از اجرای راهبرد در مجموعه آزمون استفاده می‌شود. به تعریف دوره معاملاتی ۱ ساله براساس یافته‌های دو و فاف (Do & Faff, 2010) پرداخته شده که مدعی هستند سودآوری را می‌توان با تمدید دوره معاملاتی اولین ۶ ماهه در گتو و همکاران (Gatev et al., 2006) به ۱ سال افزایش داد.

در مرحله تشکیل با دو الگوریتم خوشه‌بندی از یادگیری ماشینی نظارت‌نشده (دی‌بی‌اسکن و آپتیکز) به جست‌وجوی انتخاب زوج‌ها پرداخته شده است. در مرحله معاملات چون هدف اصلی مقایسه نتایج مربوط به تکنیک‌های انتخاب زوج‌ها است، بنابراین مدل استاندارد مبتنی بر آستانه با پارامترهای معرفی شده در جدول (۱) به کار گرفته شده است. پارامترهای استفاده‌شده در مدل همان‌هایی است که توسط گتو و همکاران (Gatev et al., 2006) به‌عنوان مرجعی برای چندین مطالعه در این زمینه به کار گرفته شده است. انحراف معیار مربوط به شکاف قیمتی σ_s و میانگین μ_s باتوجه به تمام دوره تشکیل محاسبه شده است. آستانه خروج به‌صورت مقدار میانگین تعریف شده که به معنای موقعیتی است که فقط با بازگشت شکاف قیمتی به میانگین خود بسته می‌شود؛ اما یک نقطه‌ضعف وجود دارد که نقاط ورود با دقت زیادی تعریف نشده است.

جدول (۱) پارامترهای مدل مبتنی بر آستانه

Table (1) Threshold-based model parameters

پارامترها	مقادیر
آستانه خرید	$\mu_s - 2\sigma_s$
آستانه فروش	$\mu_s + 2\sigma_s$
آستانه خروج	μ_s

¹ Principal Component Analysis (PCA)

برای ایجاد یک مدل معاملاتی مستحکم خوب است هنگام تعریف موقعیت‌های ورود به بازار، از داده‌های آتی پیش‌بینی‌شده استفاده کرد و بر پیش‌بینی سری‌های شکاف قیمتی متمرکز شد. S_t و S_t^* به ترتیب مقدار واقعی و پیش‌بینی برای شکاف قیمتی در زمان t تعریف می‌شود:

$$\Delta_{t+1} = \frac{S_{t+1} - S_t}{S_t} \times 100 \quad (2)$$

تغییر شده بین پیش

با محاسبه تغییر پیش‌بینی‌شده شکاف قیمتی، شرایط تسویه موقعیت معاملاتی در معادله (۳) شرح داده شده که α_s و α_l به ترتیب بیانگر موقعیت معاملاتی خرید و فروش فعال‌کننده آستانه‌ها است. به محض ورود به یک موقعیت معاملاتی، تا زمانی که جهت شکاف قیمتی پیش‌بینی‌شده ادامه دارد، موقعیت معاملاتی حفظ می‌شود؛ در غیر این صورت با تغییر جهت موقعیت معاملاتی بسته می‌شود.

$$\text{شرایط ورود به بازار} : \begin{cases} \text{if } \Delta_{t+1} \geq \alpha_l, & \text{باز کردن یک موقعیت خرید} \\ \text{if } \Delta_{t+1} \leq \alpha_s, & \text{باز کردن یک موقعیت فروش} \\ & \text{در غیر صورت این خارج از بازار باشید} \end{cases} \quad (3)$$

برای تعریف آستانه‌های α_s و α_l ، رویکردی ساده، غیرتکراری و مبتنی بر داده‌ها پیشنهاد می‌شود. برای هر سری مربوط به شکاف قیمتی با به دست آوردن $f(x)$ ، توزیع تغییر درصدی مربوط به شکاف قیمتی طی دوره تشکیل کار آغاز می‌شود. از توزیع $f(x)$ ، به ترتیب مجموعه مقادیر منفی (تغییرات درصدی منفی) و مقادیر مثبت (تغییرات درصدی مثبت) در نظر گرفته شده است. صدک‌ها و چندک‌های بالای $f^+(x)$ به عنوان کاندیدای تعریف α_l و موارد پایینی از $f^-(x)$ برای تعریف α_s استفاده می‌شود. سپس هر دو آستانه مبتنی بر چندک و صدک در مجموعه اعتبارسنجی آزمون می‌شود تا خوش‌بینانه‌ترین ترکیب انتخاب شود. فرایند تشریح شده را می‌توان به صورت رسمی‌تر این گونه ارائه داد:

$$\{\alpha_s, \alpha_l\} = \operatorname{argmax}_q R^{val}(q), q \in [\{Q_{f^-(x)}(0.20), Q_{f^+(x)}(0.80)\}, \{Q_{f^-(x)}(0.10), Q_{f^+(x)}(0.90)\}] \quad (4)$$

کار شامل سه سبد آزمایشی متفاوت اوراق بهادار شبیه به سناریوهای معاملاتی احتمالی است. سبد ۱ تمامی زوج‌های شناسایی‌شده در دوره تشکیل، سبد ۲ از بازخورد جمع‌آوری‌شده ناشی از اجرای راهبرد در مجموعه اعتبارسنجی با انتخاب زوج‌هایی که نتیجه مثبتی دارند و سبد ۳ محدود به سرمایه‌گذاری در تعداد ثابتی از k زوج با بازده کسب‌شده بالا در مجموعه اعتبارسنجی است. K برابر با ۱۰ در نظر گرفته می‌شود؛ زیرا بین انتخاب‌های گتو و همکاران (Gatev et al., 2006) است که از $k = 5$ و $k = 20$ استفاده کرده‌اند.

برای مقایسه استحکام فراهم‌آمده برای مدل استاندارد مبتنی بر آستانه با مدل پیشنهادی مبتنی بر پیش‌بینی، به شبیه‌سازی با استفاده از مدل ال‌اس‌تی‌ام و ال‌اس‌تی‌ام رمزگذار – رمزگشا پرداخته شده است. هنگامی که نتایج نشانی از برازش بیش از حد را نشان دهد، آموزش متوقف می‌شود. یک خط‌پایه ساده^۱ به تعریف هدف کرانه پایین‌تر می‌پردازد که برای پشتیبانی از فرایند 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:

$$\text{Naive} : Y_{t+1} = Y_t \quad (5)$$

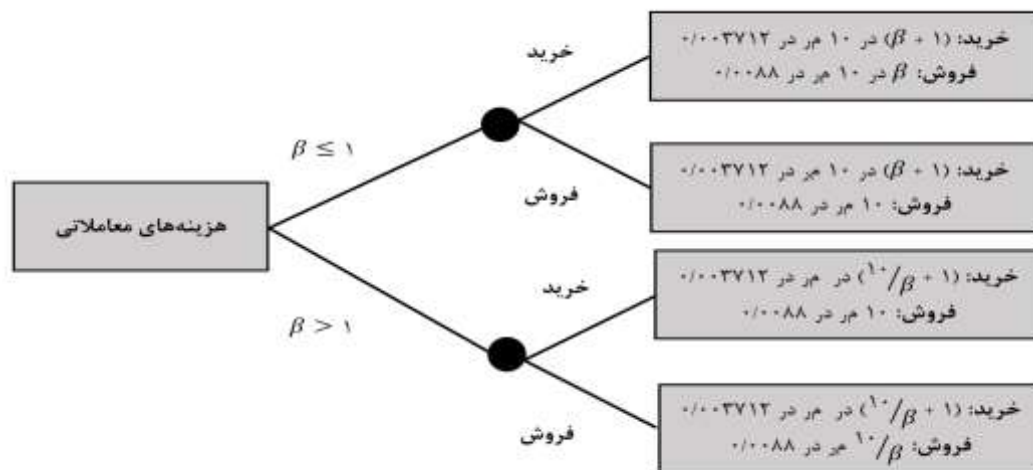
¹ Naive baseline

وزن تمامی زوج‌ها در سبد اوراق بهادار به صورت یکسان موزون شد. برای ساده کردن محاسبات در مطالعات بین‌المللی، مبلغ یک دلار سرمایه‌گذاری در هر زوج در نظر گرفته می‌شود. این رویکرد از سوی بیشتر نویسندگان در این زمینه (Gatev et al., 2016; Rad et al., 2013; Caldeira & Moura, 2010; Rudy et al., 2010; Avellaneda & Lee, 2006; al., 2006) اتخاذ شده است که برای اهداف مقایسه نیز آن را مناسب‌تر می‌کند؛ اما با توجه به حداقل مبلغ سرمایه‌گذاری در ایران (۵ میلیون ریال) برای اعمال نشدن حداقل کارمزد معامله (۲۰ هزار ریال) و سهولت محاسبات از مبلغ ۱۰ میلیون ریال استفاده شده که عدد ۱۰ در کدها و فرمول‌ها نماینده آن است؛ از این رو خلاف برخی پژوهش‌های علمی که تمایل به اتخاذ موقعیت خنثی – دلاری داشتند، همانند گتو و همکاران (Gatev et al., 2006) و رودی و همکاران (Rudy et al., 2010) که مبلغ مساوی در موقعیت معاملاتی خرید و موقعیت معاملاتی فروش سرمایه‌گذاری می‌کردند، روش به کار گرفته شده در این کار با توجه به نسبت هم‌انباشتگی β بین دو ورقه بهادار است؛ بنابراین، پیشنهاد می‌شود مبلغ سرمایه‌گذاری شده در X ، β برابر Y سرمایه‌گذاری شود. در این شرایط، به منظور تثبیت سرمایه‌گذاری اولیه ۱۰ میلیون ریالی، نیاز است اطمینان حاصل شود هزینه‌ها نه برای موقعیت معاملاتی خرید و نه برای موقعیت معاملاتی فروش بیش از ۱۰ میلیون ریال نیست. از نظر رسمی، شرط تحمیل شده این است که:

$$\max(\log 1, \log 2) = 10 \quad (6)$$

که $\log 1$ و $\log 2$ ارائه‌دهنده سرمایه سرمایه‌گذاری شده در هر بخش از زوج است. با پیشرفت معاملات، این طور در نظر گرفته می‌شود که تمامی سرمایه حاصل از یک زوج طی دوره معاملاتی در معامله بعدی سرمایه‌گذاری می‌شود؛ برای نمونه چنانچه برای اولین بار که زوجی مورد معامله قرار گرفت و بازده ۵ درصدی حاصل شد، بار دوم که معامله‌ای برای همان زوج انجام شد، سرمایه اولیه ۱۰/۵ میلیون ریال خواهد بود و نه مانند قبل ۱۰ میلیون ریال. این مکانیزم محاسبه بازده نهایی را تسهیل می‌کند.

هنگام ورود به یک موقعیت معاملاتی، سیستم به طور مداوم به پایش قیمت‌ها می‌پردازد تا به محض فعال شدن یک آستانه فرضی، یک موقعیت معاملاتی ترتیب داده شود؛ اما از آنجاکه یک سیستم معاملاتی به صورت آنی نمی‌تواند عمل کند، ممکن است تا اندازه‌ای انحراف در قیمت ورود وجود داشته باشد که تأخیری برای ورود به یک موقعیت معاملاتی است. در راستای گتو و همکاران (Gatev et al., 2006)، دو و فاف (Do & Faff, 2010) و رودی و همکاران (Rudy et al., 2010) فرض می‌شود تأخیر محافظه کارانه یک دوره‌ای برای ورود به یک موقعیت معاملاتی وجود دارد؛ بنابراین، چنانچه یک موقعیت معاملاتی در زمان $t = i$ فعال شود، ورود به موقعیت معاملاتی در زمان $t = i + 5$ دقیقه (داده با تکرار ۵ دقیقه‌ای) است. به این ترتیب اطمینان حاصل می‌شود در عمل، راهبرد کاربردی است.



شکل (۱) محاسبه هزینه‌های معاملاتی

Figure (1) Calculating of transaction costs

برای تخمین صحت مدل‌ها از بازده سرمایه‌گذاری $(ROI = \frac{\text{سود خالص}}{\text{سرمایه اولیه}} \times 100)$ ، نسبت شارپ $(SR_{\text{سال}})$ عامل سالانه $(\frac{R^{port} - R_f}{\sigma_{port}} \times 100)$ که بیانگر بازده سبدي، R_f بیانگر نرخ بدون ریسک و σ_{port} بیانگر انحراف معیار بازده سبدي است و کاهش حداکثری (حداکثر کاهش مشاهده‌شده از قله به سمت پایین از یک سری زمانی پیش از آنکه به قله جدید برسد که باتوجه به مانده حساب طی دوره معاملاتی محاسبه می‌شود) $MDD(T) = \max_{t \in (0, T)} \left[\max_{\tau \in (0, T)} \frac{X(t) - X(\tau)}{X(t)} \right]$ که $X(t)$ کل مانده حساب است) استفاده شده و برای تخمین صحت پیش‌بینی از سه مورد زیر استفاده شده است:

$$MSE = \frac{1}{n} \sum_{i=1}^n e_t^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_t|$$

$$RMSE = \sqrt{MSE}$$

(۷)

نرخ بدون ریسک، R_f بیانگر نرخ بازده موردانتظار یک سرمایه‌گذاری فرضی است که دارای هیچ ریسک زیان مالی نیست و باید از بازده کسب‌شده توسط یک راهبرد معین کسر شود تا بهره‌ای که همان مبلغ وجه نقد بدون ریسک ایجاد می‌کند به حساب آید. در این کار، از رویه معمول تعیین نرخ بدون ریسک در ایران استفاده می‌شود که برابر با متوسط بازده اوراق بدون ریسک (اخزا) مورد معامله در بورس اوراق بهادار تهران و فرابورس ایران است. جدول (۲) به توضیح نرخ‌های بدون ریسک سالانه‌شده مدنظر برای هر دوره آزمون می‌پردازد.

درخصوص هزینه‌ها در ایران این موارد قابل ذکر است: ۱) هزینه کارمزد: هزینه هر خرید ۰/۳۷۱۲ درصد و هزینه هر فروش ۰/۸۸ درصد است. سقف کارمزد کارگزاران در معاملات سهامی که ارزش زمان معامله آنها ۵ میلیون ریال یا کمتر باشد، «بیست هزار ریال» تعیین شده است؛ از این رو حداقل سرمایه برای شروع معاملات ۱۰ میلیون ریال در نظر گرفته شده است؛ ۲) هزینه محدودیت فروش استقرایی: در بورس اوراق بهادار تهران و فرابورس ایران ممنوعیت انجام فروش استقرایی وجود دارد؛ از این رو برای آن هزینه‌ای در نظر گرفته نشده است و فرض می‌شود برای بخش اتخاذ موقعیت فروش معامله زوجی، سبدي از کل سهام مدنظر به اندازه کافی در اختیار است.

جدول (۲) نرخ‌های بدون ریسک مدنظر برای دورهٔ آزمون

Table (2) Risk-free rates considered for test period

فروردین ۱۳۹۵ -	فروردین ۱۳۹۶ -	فروردین ۱۳۹۷ -	فروردین ۱۳۹۸ -
اسفند ۱۳۹۵	اسفند ۱۳۹۶	اسفند ۱۳۹۷	اسفند ۱۳۹۸
۰/۲۳۹۳	۰/۱۹۹۶	۰/۲۳۳۸	۰/۲۰۷۸

 R_f

کد توسعه داده‌شده در این کار در پایتون اجرا شده است. انگیزهٔ انجام آن مربوط به تعداد زیاد منابع در دسترس است که اجرای الگوریتم‌های انتخابی، یعنی رویه‌های داده‌کاوی و تجزیه و تحلیل داده را تسهیل می‌کند. برخی از کتابخانه‌های پایتون به‌طور خاص در اینجا مفید است. ابتدا، `sci-kit learn` به‌منظور اجرای پی‌سی‌ای و الگوریتم آپتیکز؛ دوم، `statsmodels` یک نسخهٔ تقریباً اجرایی از آزمون ADF و برای انجام آزمون هم‌انباشتگی؛ `tensorflow` برای پیاده‌سازی و آموزش مدل‌های یادگیری عمیق و در نهایت استفاده از `keras`، یک کتابخانهٔ یادگیری عمیق که بلوک‌های سازندهٔ یک شبکه عصبی را فراهم می‌آورد. برای اجرای ال‌اس‌تی‌ام از تابع فعال‌سازی رلو^۱ و برای معیار ارزیابی از میانگین قدرمطلق^۲ (MAE) خطا استفاده شده است.

یافته‌ها

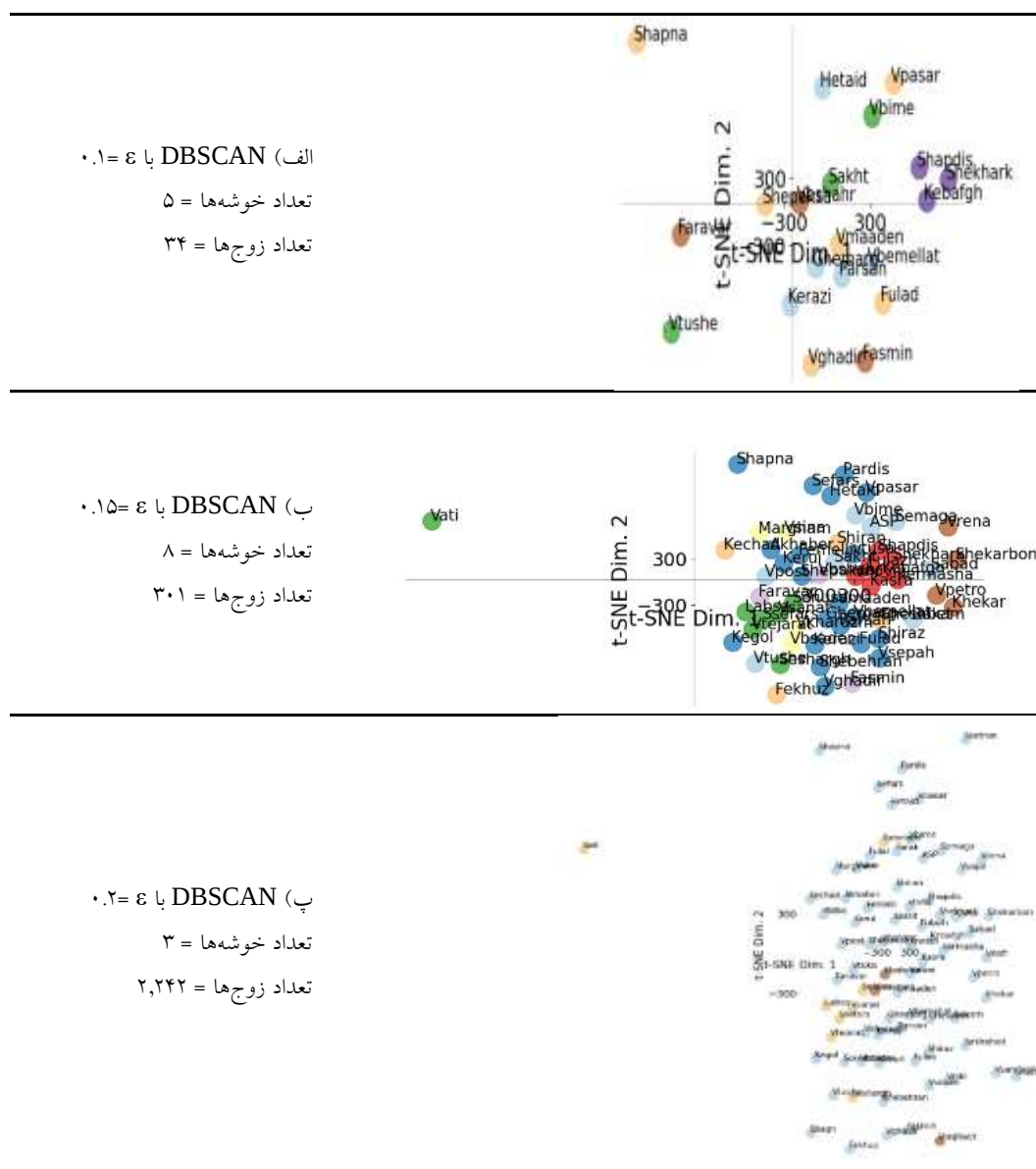
اولین گام در اعمال رویکرد پیشنهادی برای خوشه‌بندی سهام شامل تعریف تعداد مناسبی اجزای اصلی است تا بُعد داده‌های اصلی کاهش یابد. جدول (۳) به توضیح این مورد می‌پردازد. اعداد شرح داده‌شده مربوط به دورهٔ فروردین ۱۳۹۵ تا اسفند ۱۳۹۸ است. همان‌طور که مشهود است، از بُعد ۵ به بُعد ثبات در اعداد واریانس توضیحی حاصل شده است؛ از این‌رو، بعد ۵ انتخاب می‌شود.

جدول (۳) واریانس توضیحی هنگامی که تعداد اجزای اصلی متغیر است

Table (3) Explanatory variance when the number of principal components varies

تعداد اجزای اصلی	۳	۵	۸	۱۲
۰/۰۰۳۷۲	۰/۰۰۳۷۲	۰/۰۰۳۷۲	۰/۰۰۳۷۲	۰/۰۰۳۷۲
۰/۰۰۰۴۴۳	۰/۰۰۰۴۴۳	۰/۰۰۰۴۴۳	۰/۰۰۰۴۴۳	۰/۰۰۰۴۴۳
۰/۰۰۰۲۰	...	...	...	...
۰/۰۰۰۱۶	۰/۰۰۰۱۶	۰/۰۰۰۰۶	۰/۰۰۰۰۵	۰/۰۰۰۰۵
۰/۰۰۰۸۳	۰/۰۰۰۰۵	۰/۰۰۰۰۵	۰/۰۰۰۰۵	۰/۰۰۰۰۵

^۱ Relu^۲ Mean Absolute Error



شکل (۳) کاربرد دی‌بی‌اس‌کن با متغیر برای دوره تشکیل از فروردین ۱۳۹۵ تا اسفند ۱۳۹۸

Figure (3) Application of DBSCAN with variables for formation period from Farvardin 1395 to Esfand 1398

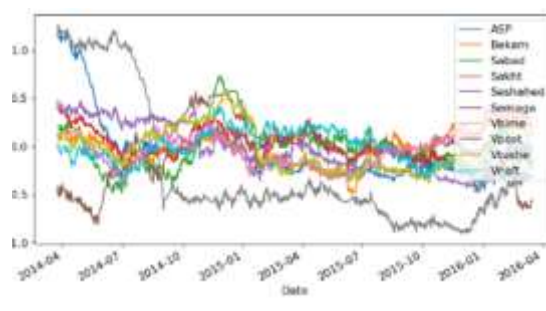
الگوریتم آپتیکز به دو دلیل مزیت خاصی در مقایسه با دی‌بی‌اسکن دارد: (۱) به‌خوبی خوشه‌های با چگالی متغیر را در مجموعه داده مشابه قرار می‌دهد؛ (۲) از فرایند دشواری اجتناب می‌ورزد که سرمایه‌گذار باید برای یافتن مقدار مناسب ϵ طی کند. برای توضیح اینکه چه‌طور این مورد در عمل بدیهی است، همانند شکل (۳) خوشه‌هایی ارائه شده که با استفاده از دی‌بی‌اسکن طی دوره تشکیل مشابه برای یک پارامترسازی متغیر شکل گرفته است

با مشاهده سه سناریو در شکل (۳)، اولین جنبه مشهود، حساسیت بالای الگوریتم به تغییرات در پارامتر ϵ است. این جنبه به تقویت این مورد می‌پردازد که سرمایه‌گذار باید هنگام انتخاب پارامتر بسیار محتاط باشد؛ علاوه‌براین، چه‌طور باید سرمایه‌گذار از سه ترکیب ارائه‌شده انتخاب کند. شاید زیاد تصمیم‌گیری ساده نباشد که کدام ترکیب به نتایج امیدوارکننده‌ای منتهی می‌شود و سپردن این تصمیم به سرمایه‌گذار بسیار طاقت‌فرسا است؛ با این وجود الگوریتم آپتیکز اجازه خودکارسازی این فرایند را می‌دهد.

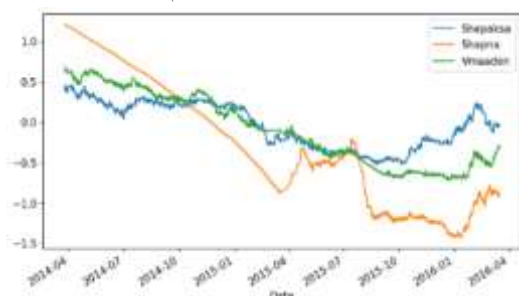
برای حصول دیدی بهتر از اوراق بهادار تشکیل دهنده هر خوشه، تجزیه و تحلیل سری های قیمتی مربوط پیشنهاد می شود؛ به همین منظور، نه خوشه برای تجسم انتخاب می شود. شکل (۴) به توضیح درباره الگوریتم سری های قیمتی هر سهم می پردازد. سری های قیمتی شرح داده شده ناشی از کسر شدن میانگین از سری های قیمتی اصلی است تا تجسم آنها تسهیل شود. با شروع از شکل ۴ الف، جالب است دیده شود چگونه سری های نزدیک هم، یکدیگر را دنبال می کنند که همین موضوع تمایز آنها را از هم دشوار می سازد. ده ورقه بهادار شناسایی شده به دسته انبوه سازی، املاک و مستغلات و سرمایه گذاری تعلق دارد، اما با هم در ارتباط هستند. این اولین مدرکی است که رویکرد مبتنی بر آپتیکز را می تواند یک سطح عمیق تر در گروه بندی سهام پیش برد؛ زیرا قادر است دسته بندی و نیز بخش های خاص را شناسایی کند. موقعیت مشابهی در شکل ۴ خ دیده می شود. شکل های ۴ ب، پ، ت، ث، ج، چ و ح ارائه دهنده قابلیت های خوشه بندی آپتیکز است که فراتر از انتخاب سهام درون همان بخش است. در خوشه ۳، سهامی یافت می شود که مربوط به بخش شیمیایی (پارسان)، حمل و نقل (حتاید)، غذایی (غمارگ)، سایر کانه های غیرفلزی (کرازی) و بانکی (وبملت) باشد. ارتباطی مشهود میان سری های قیمتی شناسایی شده وجود دارد. حتی اگر آنها تماماً به دسته مشابهی تعلق نداشته باشند. همین مورد درخصوص خوشه ۸ اعمال می شود که از سه دسته متفاوت املاک و مستغلات (ثباغ و ثتران)، خودرو (خکار و ورنه) و شیمیایی (شکربن و وپترو) تشکیل شده است.



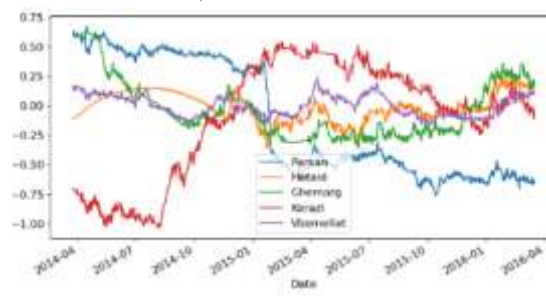
شکل ۲ (ب) قیمت های نرمال شده سهام خوشه ۲.



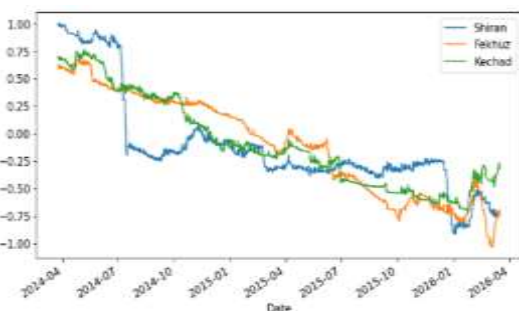
شکل ۴ الف) قیمت های نرمال شده سهام خوشه ۱.



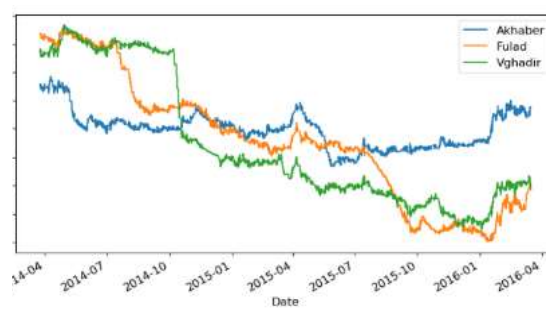
شکل ۲ (ت) قیمت های نرمال شده سهام خوشه ۴.



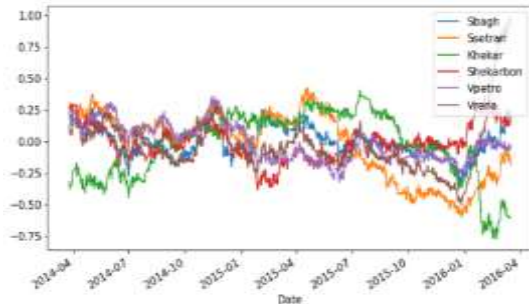
شکل ۳ (پ) قیمت های نرمال شده سهام خوشه ۳.



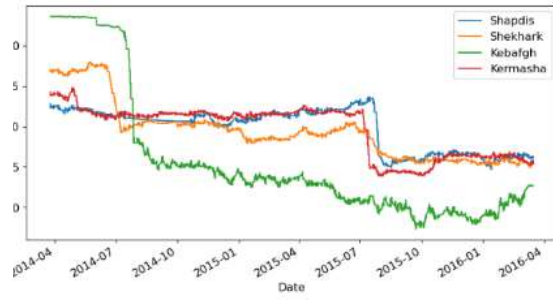
شکل ۴ (ج) قیمت های نرمال شده سهام خوشه ۶.



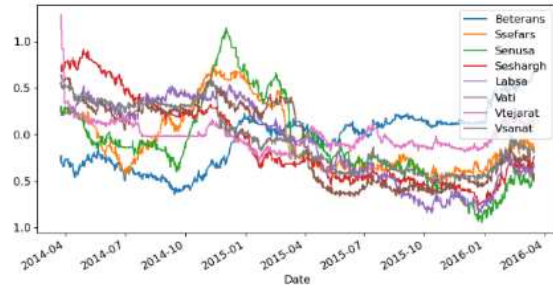
شکل ۴ (ث) قیمت های نرمال شده سهام خوشه ۵.



شکل ۴ ح) قیمت‌های نرمال‌شده سهام خوشه ۸



شکل ۴ چ) قیمت‌های نرمال‌شده سهام خوشه ۷



شکل ۴ خ) قیمت‌های نرمال‌شده سهام خوشه ۹

شکل (۴) ترکیب سری‌های قیمتی چند خوشه شکل گرفته در دوره‌ها از فروردین ۱۳۹۵ تا اسفند ۱۳۹۶

Figure (4) Combination of multi-cluster price series formed from April 2016 to March 2017

با تجزیه و تحلیل زوج‌های منتخب مشخص می‌شود که این روش، سبد اوراق بهادار امیدوارکننده‌ای را با حاشیه سودآوری رضایت‌بخشی ایجاد می‌کند (جدول ۴).

جدول (۵) نتایج مربوط به داده‌های مشاهده‌نشده را در دوره‌های آزمون آشکار می‌سازد. در این مورد برای هر دوره آزمون هنگام اتخاذ هر سه سبد اوراق بهادار آزمون‌شده به شرح نتایج پرداخته می‌شود. برای جمع‌آوری اطلاعات به‌روشی خلاصه‌تر، میانگین کل سال‌ها و سبدهای اوراق بهادار در ستون سمت چپ تشریح شده است. می‌توان تأیید کرد که سودآوری در هر محیط آزمون‌شده پایدار است. این نتیجه تأییدی بر این ایده است که رویه انتخاب زوج‌ها، مستحکم است. رویکرد مبتنی بر آپتیکز قادر است تا میانگین نسبت شارپ بالایی ۱۰/۵۷ مربوط به سبد اوراق بهادار ایجاد کند. نشان‌دهنده سازگاری بیشتری باتوجه به نسبت زوج‌های سودآور در سبد اوراق بهادار، با میانگین ۱۰۰٪ زوج‌های سودآور است.

جدول (۴) نتایج اعتبارسنجی برای تکنیک خوشه‌بندی آپتیکز

Figure (4) Validation results for the OPTICS clustering technique

دوره اعتبارسنجی	۱۳۹۵	۱۳۹۶	۱۳۹۷
آپتیکز			
SR	۹/۸۹	۲۲/۸۰	۱۰/۲۵
ROI (%)	۱۱۰۳۱/۸۲	۴۶۰۸/۴۳	۲۹۷۸/۵۳
MDD (%)	۱۷/۶۵	۹/۲۳	۱۷/۲۳
زوج‌های سودآور (%)	۱۰۰/۰۰	۹۲/۸۶	۱۰۰/۰۰
تعداد معاملات (مثبت-منفی)	۵۵ (۴۵-۱۰)	۲۸ (۲۰-۸)	۴ (۳-۱)

جدول (۵) نتایج آزمون برای تکنیک گروه‌بندی زوجها

Table (5) Test results for pair grouping technique

دوره آزمون	۱۳۹۵	۱۳۹۶	۱۳۹۷	میانگین
سبد آزمون	۱	۲	۳	-
آپتیکز	۱	۲	۳	-
SR	۱۱/۴۶	۱۱/۱	۱۱/۴۶	۱۰/۵۷
ROI (%)	۴۲۴۰	۴۶۱۸	۴۲۴۰	۵۵۷۴
MDD(%)	۱۶/۲۳	۱۷/۰۱	۱۶/۲۳	۲۱/۹۱
زوج‌های کل	۸	۷	۸	۱۲/۴۴
زوج‌های سودآور (%)	۱۰۰	۱۰۰	۱۰۰	۱۰۰
کل معاملات	۱۳	۱۱	۱۳	۱۹/۸۹
معاملات سودآور	۱۰	۹	۱۰	۱۶/۷۸
معاملات بدون سود	۳	۲	۳	۳/۱۱

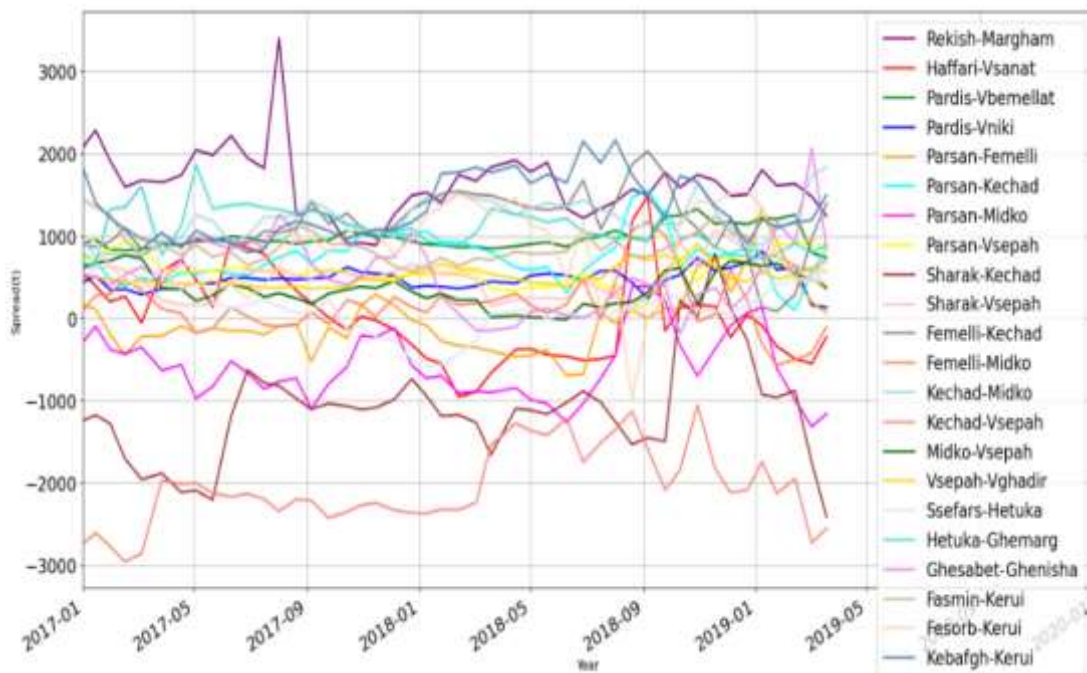
در پاسخ به پرسش «یک مدل معاملاتی مبتنی بر پیش‌بینی می‌تواند عملکرد مستحکم‌تری داشته باشد؟» این موضوع ارزیابی می‌شود که چگونه مدل مبتنی بر پیش‌بینی، هنگام مقایسه با مدل مبتنی بر آستانه که شبیه‌سازی شد عمل می‌کند. با استفاده از خوشه‌بندی آپتیکز زوج‌های منتخب در جدول (۶) ارائه شده است.

الگوریتم‌های پیش‌بینی آموزش می‌بینند تا شکاف‌های قیمتی مربوط به زوج‌های منتخب طی دوره تشکیل را (فروردین ۱۳۹۶ تا اسفند ۱۳۹۸) همان‌گونه که در شکل (۵) ارائه شده برآزش کنند. همان‌طور که انتظار می‌رفت شکاف‌های قیمتی، ایستا به نظر می‌رسد؛ علاوه‌براین، تفاوت محسوسی در نوسان میان سری‌های زمانی وجود دارد که از این ایده بیشتر پشتیبانی می‌کند که پارامترهای مدل داده‌محور، همانند آستانه‌های پیشنهادی برای مدل معاملاتی مبتنی بر پیش‌بینی، برای این شرایط مناسب به نظر می‌رسد.

جدول (۶) زوج‌های کاندید شناسایی شده از فروردین ۱۳۹۶ تا اسفند ۱۳۹۸

Table (6) Candidate pairs identified from Farvardin 1396 to Esfand 1398

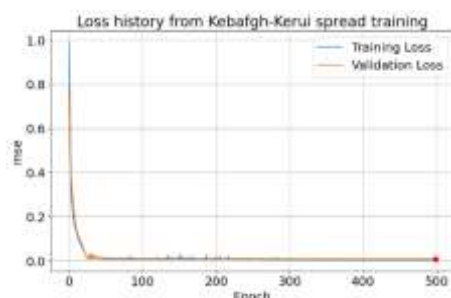
زوج‌ها	گروه	زوج‌ها	گروه
فملی	فلزات اساسی	فملی	فلزات اساسی
کچاد	استخراج کانه‌های فلزی	میدکو	فلزات اساسی
پارسان	محصولات شیمیایی	حتوکا	حمل و نقل، انبارداری و ارتباطات
فملی	فلزات اساسی	غمارگ	محصولات غذایی و آشامیدنی جز قند و شکر
پارسان	محصولات شیمیایی	رکیش	رایانه و فعالیت‌های وابسته به آن
کچاد	استخراج کانه‌های فلزی	مرقام	رایانه و فعالیت‌های وابسته به آن
پردیس	سرمایه‌گذاری	وساپا	سرمایه‌گذاری
ویملت	بانک‌ها و مؤسسات اعتباری	وغدیر	شرکت‌های چندرشته‌ای صنعتی
حفاری	استخراج نفت، گاز و خدمات جنبی	فاسمین	فلزات اساسی
وصنت	سرمایه‌گذاری‌ها	کروی	استخراج کانه‌های فلزی
پارسان	محصولات شیمیایی		
میدکو	فلزات اساسی		



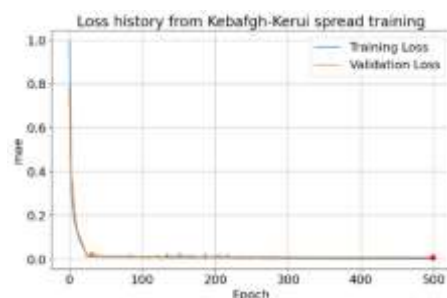
شکل (۵) شکاف قیمتی مربوط به ۲۲ زوج شناسایی شده بین فروردین ۱۳۹۶ و اسفند ۱۳۹۸

Figure (5) Price gap for 22 pairs identified from Farvardin 1396 to Esfand 1398

الاس تی ام و الاس تی ام رمز گذار - رمز گشا از دوره یادگیری برای آموزش استفاده می کنند که مجموعه اعتبارسنجی به صورت جداگانه به کنترل برازش بیش از حد می پردازد. از طریق توقف زودرس، سیستم به صورت مستمر به نظارت نمرة اعتبارسنجی می پردازد و چنانچه نشانه هایی از برازش بیش از حد شناسایی شود، فرایند آموزش متوقف می شود. شکل (۶) یک سناریوی معمولی آموزش را با نمونه ای از برازش شکاف قیمتی الاس تی ام رمز گذار - رمز گشا با استفاده از زوج کبافق - کروی نشان می دهد.



شکل ۶الف) خطای مربعات میانگین طی دوره آموزش



شکل ۶ب) خطای قدرمطلق میانگین طی دوره آموزش

شکل (۶) تاریخچه زیان آموزش الاس تی ام رمز گذار - رمز گشا برای شکاف قیمتی کبافق - کروی

Figure (6) Encoder-decoder LSTM training loss history for Kebafgh-Kegol price gap

در شکل ۶الف، چگونگی آموزش و اعتبارسنجی MSE ارائه شده است. همان طور که انتظار می رفت، هر دو با پیشرفت آموزش کاهش یافتند؛ اما به نظر می رسد الگوریتم به ترکیب بهینه خود در زمان ۵۰۰ رسیده است. از آن زمان به بعد، با وجود اینکه هنوز MSE آموزش در حال کاهش است، اما هیچ بهبودی در اعتبارسنجی MSE وجود ندارد. این یعنی الگوریتم در حال برازش بیش از حد است. شکل ارائه دهنده نقاط زمانی است (گرچه قابل مشاهده نیست)؛ زیرا توقف زودرس با مکتی در ۵۰۰

نقطه زمانی رخ داده است؛ بنابراین، آموزش برای ۵۰۰ نقطه زمانی ادامه می‌یابد؛ زیرا هیچ امتیاز بهتری در این بین کسب نشده است. لازم است تصویر سمت راست (شکل ۶ب) را مشاهده کنید که تکامل MAE را ترسیم کرده است. هر چند مدل به منظور حداقل‌رسانی MSE ترکیب‌بندی شده، مشهود است که پیشرفت MAE نیز الگوی مشابهی را نمایش می‌دهد.

برای انتخاب بهترین سبک‌های مدنظر مدل‌های مبتنی بر پیش‌بینی از کمترین میانگین مربع خطای اعتبارسنجی استفاده شده است. دلیل استفاده از نمره اعتبارسنجی به عنوان مرجع این است که مانند اطلاعاتی است که در دسترس سرمایه‌گذار پیش از شروع معامله قرار دارد، نقطه‌ای که باید تصمیم گرفته شود. میانگین مربع خطا به عنوان سنج‌ای برای اهداف مقایسه‌ای مدنظر قرار می‌گیرد؛ زیرا تابع زیانی است که با مدل‌ها به حداقل می‌رسد. مطابق این معیار، بهترین سبک‌های عملکرد در جدول (۷) ارائه شده است.

برای ارزیابی اثربخشی روش پیشنهادی در ایجاد سبد اوراق بهادار مستحکم‌تری با کاهش تعداد روزهایی که سبد مذکور در معرض کاهش ارزش خود قرار گیرد، به تشریح نتایج حاصل توسط مدل معاملاتی مبتنی بر پیش‌بینی با استفاده از الگوریتم‌های متفاوت پیش‌بینی طی دوره زمانی فروردین ۱۳۹۶ تا اسفند ۱۳۹۸ پرداخته شده است. نتایج در جدول (۸) کنار نتایج مربوط به مدل استاندارد مبتنی بر آستانه قرار گرفته است.

جدول (۷) مقایسه نتایج پیش‌بینی

Table (7) Comparison of prediction results

مدل	الاس‌تی‌ام	الاس‌تی‌ام رمزگذار-رمزگشا
	طول ورودی: ۱۲	طول ورودی: ۲۴
تشکیل	لایه‌های مخفی: ۱	گره‌های رمزگذار: ۱۰
	لایه‌های مخفی: ۶۰	گره‌های رمزگشا: ۵
اعتبارسنجی		
MSE(E-03)	۸/۱۰	۴/۲۲
RMSE(E-02)	۹/۰۰	۶/۵۰
MAE(E-02)	۵/۵۴	۴/۰۶
آزمون		
MSE(E-03)	۰/۷۵	۰/۹۴
RMSE(E-02)	۲/۷۴	۳/۰۶
MAE(E-02)	۱/۸۰	۲/۰۴

با تمرکز بر نتایج آزمون که شبیه یک محیط معاملاتی عملی است، می‌توان تأیید کرد در مدل معاملاتی استاندارد به‌طور چشمگیری دوره کاهش سبد اوراق بهادار کوتاه‌تر است؛ اما با این وجود و در نظر گرفتن دوره آزمون، دوره کاهش سبد اوراق بهادار در مدل الاس‌تی‌ام رمزگذار – رمزگشا مناسب است و در مقایسه با مدل دیگر به‌طور درخور توجهی کوتاه‌تر است. دوره طولانی‌تر مربوط به الاس‌تی‌ام، ۱۹۰ روز است. به هر حال، تمام آنچه که باید باشد نیست؛ زیرا از نظر بازده کل سرمایه‌گذاری (۲۱۹۹۳/۸ درصد) و نسبت شارپ (۵۰/۳۵) سبد اوراق بهادار عملکرد بسیار بهتری دارد. این حقیقت که الاس‌تی‌ام رمزگذار – رمزگشا به‌وضوح از مدل ساده‌تر الاس‌تی‌ام فراتر رفته است، دلالتی بر این مورد است که ظرفیت‌هایی برای استفاده از پیش‌بینی چندمرحله‌ای در این ترتیبات وجود دارد (جدول ۸).

جدول (۸) مقایسه نتایج معاملاتی با استفاده از یک دوره تشکیل ۳ ساله
 Table (8) Comparing trading results using a 3-year formation period

مدل معاملاتی	استاندارد	مدل مبتنی بر الاس تی ام	مدل مبتنی بر الاس تی ام رمز گذار - رمز گشا
پارامترها	$\alpha_S = \mu_S + 2\sigma_S$ $\alpha_L = \mu_S - 2\sigma_S$	$\alpha_S = Q_{f-(0.10)}$ $\alpha_L = Q_{f+(0.90)}$	$\alpha_S = Q_{f-(0.20)}$ $\alpha_L = Q_{f+(0.80)}$
اعتبارسنجی			
تعداد روزهای کاهش	۲	۱۲	۱۲
SR	۶/۷	۲/۱۸	-۰/۵۱
ROI(%)	۷۹۱۱/۱	۲۹۸۱۷/۴	۶۱۶/۲۸
MDD(%)	۸/۴۳	۳۳/۹۷	۱۷۰/۲۴
تعداد معاملات (مثبت-منفی)	۴۶ (۳۴-۱۲)	۹۰۴۴ (۶۱-۸۹۸۳)	۹۸۲۹ (۸۰-۹۷۴۹)
زوج های سودآور (%)	۱۰۰	۲۲/۷	۹/۱
آزمون			
تعداد روزهای کاهش	۱۰	۱۹۰	۲۹
SR	۰/۵۳	۲/۱۸	۵۰/۳۵
ROI(%)	۱۹۱۸/۹	-۹۰/۱۲	۲۱۹۹۳/۸
MDD(%)	۲۳/۱۳	۹۹/۸۰	۷۹/۱۷
تعداد معاملات (مثبت-منفی)	۲۵ (۱۳-۱۲)	۱۱۹۶ (۳۴-۱۱۶۲)	۱۵۰ (۷-۱۴۳)
زوج های سودآور (%)	۸۶/۴	۰	۱۰۰/۰

با بررسی نتایج جدول (۸)، باتوجه به سنجش های انتخابی، مدل های مبتنی بر الاس تی ام رمز گذار - رمز گشا از مدل دیگر پیشی می گیرد. طبیعتاً نتایج به طرح این پرسش منجر می شود که الاس تی ام ها برای شرایط این کار مناسب بود که خود به پژوهش عمیق تری در ادبیات موضوع منجر می شود. هرچند برازش الگوریتم های متفاوت پیش بینی قویاً به مجموعه داده وابسته است، ادبیات موضوع حاکی از آن است که هیچ اجماعی درخصوص مناسب بودن الاس تی ام ها برای پیش بینی سری های زمانی وجود ندارد. از یک طرف، برخی پژوهش های علمی از برتری آن ها بر تکنیک های ساده تر مانند مدل های ادراک کننده های چندلایه^۱ (MLP) یا خودرگرسیون حملیت می کنند؛ برای نمونه البویر و همکاران (Olivier et al., 1993) و تنتی (Tenti, 1996) ادعا می کنند این شبکه ها به طور کلی بهتر از شبکه های MLP است؛ هرچند از زمان طولانی محاسباتی رنج می برند. از طرف دیگر، برخی نویسندگان مانند گرز و همکاران (Gers et al., 2001) ادعا می کنند مشکلات عادی سری های زمانی که در ادبیات موضوع یافت شده است، اغلب از نظر مفهومی ساده تر از بسیاری از مواردی است که توسط الاس تی ام ها حل شده است و بیشتر اوقات به هیچ وجه به آن ها نیازی نیست؛ زیرا چند رویداد اخیر تمامی اطلاعات مرتبط درخصوص رویداد بعدی را منتقل می کند. به طور خاص، نویسندگان شواهدی یافتند که تنها با مشاهده چند ورودی اخیر، MLP مبتنی بر پنجره زمانی از رویکرد الاس تی ام در معیارهای خاص پیش بینی سری های زمانی عملکرد بهتری را داشت؛ بنابراین، نتیجه گرفتند قابلیت های خاص الاس تی ام برای چنین شرایطی ضروری نیست. نتایج با این دیدگاه تطابق ندارد. هرچند یک مدل MLP اجرا نشده است، نتایج به وضوح خاطرنشان می سازد که براساس موقعیت های مربوط به این مشکل، مدل الاس تی ام رمز گذار - رمز گشا استحکام بیشتری باتوجه به سنجش تعریف شده در جدول (۸) دارد.

¹ Multi-layer perceptrons

نتایج و پیشنهادها

برای مقایسه دو الگوریتم یادگیری نظارت نشده درخصوص خوشه بندی سهام در بازار بدون هیچ پیش فرضی درباره آنها، آپتیکز در مقایسه با دی بی اسکن از متغیرهای کمتری استفاده می کند و نتیجه از منظر ریسک تحمیلی (نسبت شارپ بالا) و میانگین بازده کسب شده و سازگاری بیشتر با توجه به نسبت زوج های سودآور در سبد اوراق بهادار، عملکرد خوبی داشته است. در استفاده از آپتیکز می توان به راحتی در بستر بازار به دنبال زوج های مناسب برای تشکیل و بهروزرسانی سبد بود، بدون آنکه نگرانی درخصوص بخش، صنعت و سهام باشد.

درخصوص استفاده از مدل های مبتنی بر پیش بینی با استفاده از یادگیری ماشین برای کاهش تعداد روزهایی که ارزش سبد اوراق بهادار کاهش می یابد، می توان تأیید کرد که در مدل معاملاتی استاندارد به طور چشمگیری دوره کاهش سبد اوراق بهادار کوتاه تر است؛ اما با این وجود و در نظر گرفتن دوره آزمون، دوره کاهش سبد اوراق بهادار در مدل ال اس تی ام رمزگذار - رمزگشا در مقایسه با مدل دیگر ال اس تی ام مناسب است. سارمنتو و هورتا (Sarmiento & Horta, 2020) مدل معاملاتی مبتنی بر پیش بینی (ال اس تی ام و ال اس تی ام رمزگذار - رمزگشا) را معرفی کردند که در جامعه آنان (بازار صندوق های قابل معامله ایالات متحده) این مدل ها قادر به کاهش دوره های رکود سبد بود؛ با این حال این مورد به قیمت کاهش سودآوری کلی انجامید. از نظر بازده کل سرمایه گذاری و نسبت شارپ، سبد اوراق بهادار استاندارد عملکرد بسیار بهتری دارد. این مورد که ال اس تی ام رمزگذار - رمزگشا به وضوح از مدل ساده تر ال اس تی ام فراتر رفته، دلالتی بر این مورد است که ظرفیت هایی برای استفاده از پیش بینی چند مرحله ای در این ترتیبات وجود دارد. در این خصوص می توان ساختار بازار سرمایه ایران، از جمله محدودیت در دامنه نوسان، وجود حجم مبنا برای معاملات و درگیر بازار شدن بیشتر مردم طی سال های ۱۳۹۸ و ۱۳۹۹ را از دلایل اصلی این نتیجه دانست؛ اما از منظر علمی به نظر استفاده اندک دست اندرکاران بازار سرمایه از مدل های یادگیری ماشینی چنین نتیجه ای به بار آورده است؛ زیرا هنوز حتی سهم معاملات الگوریتمی کلاسیک اندک است، چه برسد به مدل های پیشرفته آن؛ بنابراین، ممکن است ادعا شود هرچند مدل مبتنی بر پیش بینی در تمامی موارد مناسب تر نیست، در این شرایط افزایش در سودهای مشاهده شده جبرانی برای استحکام افزایش یافته است. در کار فلوری و رگولی (Flori & Regoli, 2021) ارتباطات غیرخطی میان و بین سری های زمانی مالی با استفاده از ال اس تی ام برای پیش بینی عملکرد بازار سهام موجود در شاخص S&P 500 مطالعه شد. پیش بینی ال اس تی ام در چارچوب راهبردهای سرمایه گذاری معامله زوجی آزمون شد که به مقایسه پیش بینی های چارچوب ال اس تی ام پیشنهادی برابر دیگر رویکردهای عادی ML از جمله رگرسیون منطقی، درخت های مدرج تقویت شده^۱ و شبکه های عصبی فیدفوروارد^۲ پرداخته شد. در این خصوص از ال اس تی ام برای تخمین احتمال بازده افزایشی سهمی در بازار در آینده نزدیک و در مقایسه با هم تایان خود استفاده شده است. تجزیه و تحلیل نشان می دهد راهبردهایی شامل چنین پیش بینی هایی می تواند در بهبود عملکرد سبد با فراهم آوردن سیگنال های پیش بینی کننده حاوی اطلاعاتی برای حرکت به سمت بالا و فراتر از شکاف های قیمتی و بازده، مؤثر باشد.

چندین مسیر احتمالی برای مطالعه می تواند پیرو این پژوهش باشد: (۱) افزایش تکرار داده ها (برای نمونه تکرار ادقیقه ای) برای کاهش در دوره تشکیل مورد نیاز و در نتیجه یافتن زوج های بیشتر. البته این مورد از منظر محاسباتی بسیار طاقت فرسا است؛ (۲) استفاده از مدل هایی با پیچیدگی کمتر (مانند ARMA) تا شاید بتوان ضمن انجام محاسبات کمتر به نتایج بهتری در زمان اندکی رسید؛ (۳) بررسی دو دوره زمانی از ۱۳۹۵ تا ۱۳۹۹ و از ۱۳۹۹ تا ۱۴۰۳ تا بتوان اثر پیش بینی کنندگی مدل ها را در دو دوره حرکت صعودی و حرکت نزولی بازار مقایسه کرد و (۴) ترکیب ETF های مربوط به کالاهای اساسی و انواع دیگر

¹Gradient-Boosted-Trees (GBT)

²Feedforward Neural Networks (FNN)

اوراق بهادار (مانند قراردادهای آتی) در زوج مشابه و بررسی این مورد که آیا سرمایه‌گذار از نوسان اضافی موردانتظار می‌تواند سود ببرد.

منابع

پاکیزه، کامران، اخوان چایجان، کوثر، و صالحی، پیام (۱۳۹۱). کاربرد راهبرد معاملات زوجی در بازار قراردادهای آتی سکه طلای بهار آزادی. *نهمین کنفرانس بین‌المللی مهندسی صنایع*، تهران.

طادی، مسعود، آبکار، مجید، و مطهری‌نیا، وحید (۱۳۹۷). ارزیابی استراتژی معاملات زوجی با رویکرد فاصله‌ای در بورس اوراق بهادار تهران. *فصلنامه علمی و پژوهشی دانش سرمایه‌گذاری*، ۷(۲۶)، ۹۹-۱۱۲.

عسگری، محسن، و ابو، زهرا (۱۳۹۱). بررسی اثربخشی استراتژی معاملات جفتی بر روی قراردادهای آتی سکه با ترکیب رویکردهای تصادفی و هم‌تابستگی. *سومین کنفرانس ریاضیات مالی و کاربردها*، سمنان.

فلاح‌پور، سعید، و حکیمیان، حسن (۱۳۹۶). بررسی عملکرد سیستم معاملات زوجی در بورس اوراق بهادار تهران: رویکرد هم‌تابستگی و بررسی نسبت سورتینو. *مجله مهندسی مالی و مدیریت اوراق بهادار*، ۸، ۱-۱۷.

References

- Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (1999). OPTICS. *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, 49–60. <https://doi.org/10.1145/304182.304187>
- Asgari, M., & Abu, Z. (2012). *Investigating the effectiveness of Pairs Trading strategy on coin futures contracts by combining stochastic and cointegration approaches*. Third Conference on Financial Mathematics and Applications, Semnan. [In Persian]
- Avellaneda, M., & Lee, J. H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7), 761–782. <https://doi.org/10.1080/14697680903124632>
- 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>
- 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>
- 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>
- Caldeira, J., & Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2196391>
- Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. F., Nobrega, J. P., & Oliveira, A. L. I. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, 55, 194–211. <https://doi.org/10.1016/j.eswa.2016.02.006>
- Clegg, M., & Krauss, C. (2018). Pairs Trading with partial cointegration. *Quantitative Finance*, 18(1), 121–138. <https://doi.org/10.1080/14697688.2017.1370122>
- Do, B., & Faff, R. (2010). Does simple Pairs Trading till Wwork?. *Financial Analysts Journal*, 66(4), 83–95. <https://doi.org/10.2469/faj.v66.n4.1>
- 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>
- Dunis, C. L., & Ho, R. (2005). Cointegration portfolios of European equities for index tracking and market neutral strategies. *Journal of Asset Management*, 6(1), 33–52. <https://doi.org/10.1057/palgrave.jam.2240164>
- Enders, W., & Granger, C. W. J. (1998). Unit-Root tests and asymmetric adjustment with an example sing the term structure of interest rates. *Journal of Business & Economic Statistics*, 16(3), 304–311. <https://doi.org/10.1080/07350015.1998.10524769>
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of International Conference on Knowledge Discovery and Data Mining*, 226–231.
- Fallahpour, S., & Hakimian, H. (2017). Evaluating the performance of a Pairs Trading system in Tehran Stock Exchange: The cointegration approach and Sortino ratio analysis. *Financial Engineering and Securities Management (Portfolio Management)*, 8, 1–17. [In Persian].

- 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>
- 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>
- Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs Trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies*, 19(3), 797–827. <https://doi.org/10.1093/rfs/hhj020>
- Gers, F. A., Eck, D., & Schmidhuber, J. (2001, August). Applying LSTM to time series predictable through time-window approaches. In *International conference on artificial neural networks* (pp. 669-676). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Goodfellow, I., Bengio, Y., & Courville, A. (2017). *Deep Learning*. The MIT Press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huck, N. (2009). Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research*, 196(2), 819–825. <https://doi.org/10.1016/j.ejor.2008.03.025>
- Huck, N. (2010). Pairs Trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research*, 207(3), 1702–1716. <https://doi.org/10.1016/j.ejor.2010.06.043>
- 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>
- Khandani, A. E., & Lo, A. W. (2008). What happened to the quants in August 2007?; Evidence from factors and transactions data. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1288988>
- Lin, Y. X., McCrae, M., & Gulati, C. (2006). Loss protection in pairs Trading through minimum profit bounds: A cointegration approach. *Journal of Applied Mathematics and Decision Sciences*, 1–14. <https://doi.org/10.1155/JAMDS/2006/73803>
- Maaten, L. V., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9.
- Narajewski, M., Kley-Holsteg, J., & Ziel, F. (2021). Tsrobpreg — an R package for robust preprocessing of time series data. *SoftwareX*, 16, 100809. <https://doi.org/10.1016/j.softx.2021.100809>
- Olivier, A., Jean-Luc, Z., & Maurice, M. (1993, June). Identification and prediction of non-linear models with recurrent neural network. In *International workshop on artificial neural networks* (pp. 531-535). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-56798-4_198
- Pakizeh, K., Akhavanhajian, K., & Salehi, P. (2012). *Application of Pairs Trading strategy in the Bahar Azadi gold coin futures market*. 9th International Conference on Industrial Engineering, Tehran. [In Persian].
- 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>
- Ramezanifar, E., Mohammadi, S., Rad, H., & Beyty, S. (2015). Pairs Trading using fractional cointegration approach and its comparison with cointegration approach. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2614240>
- Rudy, J., Dunis, C., Giorgioni, G., & Laws, J. (2010). Statistical arbitrage and igh-frequency data with an application to Eurostoxx 50 equities. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2272605>
- Sarmiento, 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>
- Sarmiento, S., & Horta, N. C. G. (2021). *A Machine Learning Based Pairs Trading Investment Strategy*. Springer.
- Tadi, M., Abkari, M., & Motaharinia, V. (2018). Evaluation of Pairs Trading strategy with distance approach in Tehran Stock Exchange. *Investment Knowledge*, 7(26), 88-112. [In Persian]
- Tenti, P. (1996). Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence*, 10(6), 567–582. <https://doi.org/10.1080/088395196118434>
- Troiano, L., Villa, E. M., & Loia, V. (2018). Replicating a Trading strategy by means of LSTM for inancial industry applications. *IEEE Transactions on Industrial Informatics*, 14(7), 3226–3234. <https://doi.org/10.1109/TII.2018.2811377>
- Vidnyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis*. J. Wiley.
- Xia, Y., Wang, T., Liew, S. C., & Zhang, S. (2025). A hierarchical deep learning framework for pair trading with attention and graph networks. *Expert Systems with Applications*, 291, 128496. <https://doi.org/10.1016/j.eswa.2025.128496>