Fin. Manag. Persp. 2026, 16(2), 9-28; DOI: 10.48308/jfmp.2026.245590.1622

Financial Management Perspective
Journal homepage: https://jfmp.sbu.ac.ir/

Original Article

Testing Weak-Form Market Efficiency Based on Return Predictability: A Comparison of the Tehran Stock Exchange Composite Index and the S&P 500 Using Econometric and Machine Learning Approaches

Masoud Fazli[*]

Ahmad Shabani[**]

Abstract

Introduction: The main objective of this study is to test weak-form efficiency in capital markets, using the Tehran Stock Exchange composite index and the S&P 500. The main question is whether returns in these markets are consistent with weak-form efficiency when tested against past price and volume information. Out-of-sample return prediction provides the empirical testing framework, with a zero-return random walk as the statistical benchmark. Further analyses ask whether nonlinear models improve on linear regression and whether Tehran trading-condition indicators add information. These supplementary comparisons support the efficiency test; the indicators neither measure an order book nor isolate the effect of market rules. The design therefore treats statistical predictability as evidence about the information contained in market history, while keeping economic efficiency, risk adjustment, transaction costs, and implementability conceptually separate.

Methods: Daily index data are used to predict the next day's log return. The standard features include current returns, three return lags, ten-day volatility and five-day mean volume. For Tehran, a second set includes distance from a hypothetical band, its five-day mean, proximity to the band, positive or negative returns with low relative volume,

Received: 24, May, 2026 Accepted:21,September,2026

[*] M.Sc. Student, Department of Economics, Faculty of Islamic Studies and Economics, Imam Sadiq University, Tehran, Iran **(Corresponding Author). E-mail**: masud.fazli.shahri@gmail.com

[**]Associate Professor, Department of Financial Economics, Faculty of Islamic Studies and Economics, Imam Sadiq University, Tehran, Iran.

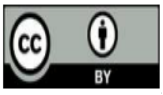
and a period indicator. The combined set contains both groups. Linear regression and LightGBM are compared with the zero-return benchmark in five expanding-window folds. Training and validation come before each test block, with one-row gaps at the selection boundaries. Preprocessing uses training data only. Performance is assessed with forecast errors, test-sample R-squared and direction accuracy. Paired squared-error losses are compared by a Diebold–Mariano test with a standard error robust to changing variance and serial dependence. Checks of result stability vary the volume threshold and initial training share and use a subsample before the period boundary. A separate analysis compares tuned and default LightGBM settings to assess sensitivity to hyperparameters. Models are re-estimated as the training window expands. Performance is summarized both by averaging the five test-fold metrics and by pooling test errors, so fold averages are distinguished from aggregate results.

Finding: For Tehran with standard features, mean fold RMSE is 0.011300 for the random walk, 0.010504 for linear regression and 0.010709 for LightGBM. Both fitted models improve on the benchmark, with HAC p-values of 0.0002 and 0.0014, respectively. Their difference is not significant in the full-sample baseline. Neither model shows a significant gain over the benchmark for the S&P 500. In Tehran, gains from tuned models remain under the tested volume thresholds and in the subsample before the boundary. However, default LightGBM settings lead to worse performance. This comparison demonstrates sensitivity to hyperparameters, not model robustness. Stability under the tested volume thresholds and subsample choices does not establish stability across model settings. Feature-attribution results are used to describe model behavior rather than to establish causal effects or incremental forecasting value.

Conclusions: The zero-return random walk is a statistical forecasting benchmark, not a complete test of weak-form market efficiency. Market efficiency does not require a zero expected return. The absence of significant gains for the S&P 500 therefore does not prove efficiency, while lower forecast errors in Tehran do not by themselves establish inefficiency or net trading profits. LightGBM performance is sensitive to hyperparameters; the default-setting comparison does not establish model robustness. No consistent advantage of nonlinear complexity is found, and a nonsignificant difference does not establish model equivalence. The cross-market comparison is descriptive and cannot identify causal effects of market rules or structure. Accordingly, the evidence supports a cautious, benchmark-dependent interpretation: historical data contain useful predictive content for Tehran under selected specifications, whereas the S&P 500 results do not provide comparable evidence within this sample and research design.

Keywords: Capital market; Efficient market hypothesis; Machine learning; Market microstructure; Return predictability.

How to Cite Fazli, M. and Shabani, A. (2026). Testing Weak-Form Market Efficiency Based on Return Predictability: A Comparison of the Tehran Stock Exchange Composite Index and the S&P 500 Using Econometric and Machine Learning Approaches. *Financial Management Perspective, 16(2). (In Persian).* 10.48308/jfmp.2026.245590.1622



چشم انداز مدیریت مالی ،۱۴۰۵ دوره ۱۶، شماره ۲ صص۹ -۲۸، DOI: 10.48308/jfmp.2026.245590.1622

چشم انداز مدیریت مالی
Journal homepage: https://jfmp.sbu.ac.ir/

نوع مقاله: پژوهشی

آزمون شکل ضعیف کارایی بازار بر اساس پیش‌بینی‌پذیری بازده: مقایسه شاخص کل بورس تهران و S&P ۵۰۰ با رویکرد اقتصادسنجی و یادگیری ماشین

مسعود فضلی*،
احمد شعبانی**

چکیده

هدف: پژوهش حاضر شکل ضعیف کارایی بازار سرمایه را در بورس تهران و بازار سهام ایالات متحده، با استفاده از شاخص کل تهران و شاخص اس‌اندپی ۵۰۰، آزمون می‌کند. پرسش اصلی این است که آیا پیش‌بینی‌پذیری بازده بر پایه اطلاعات تاریخی قیمت و حجم با شکل ضعیف کارایی سازگار است. برای پاسخ، پیش‌بینی بازده روز آینده در خارج از نمونه با گام تصادفی با بازده صفر مقایسه می‌شود. مقایسه مدل خطی و غیرخطی و بررسی نشانگرهای تقریبی شرایط معاملاتی تهران، تحلیل‌های تکمیلی این آزمون‌اند. این مقایسه‌ها نشان می‌دهند که بهبود پیش‌بینی تا چه اندازه به اطلاعات تاریخی و تا چه اندازه به نوع مدل وابسته است. بااین‌حال، از تفاوت نتایج دو بازار نمی‌توان درباره اثر علّی ساختار بازار نتیجه‌گیری کرد.

روش: از داده‌های روزانه دو شاخص برای پیش‌بینی بازده لگاریتمی روز بعد استفاده شد. داده‌های خام تهران شامل ۲۷۲۱ روز معاملاتی از ۷ مارس ۲۰۱۵ تا ۲۹ ژوئن ۲۰۲۶ و داده‌های اس‌اندپی ۵۰۰ شامل ۲۸۷۷ روز معاملاتی از ۱۶ ژانویه ۲۰۱۵ تا ۲۶ ژوئن ۲۰۲۶ بود؛ پس از آماده‌سازی، نمونه‌های قابل مدل‌سازی به‌ترتیب ۲۷۱۰ و ۲۸۶۶ مشاهده را دربر گرفتند. متغیرهای متعارف شامل بازده جاری، وقفه‌های یک تا سه‌روزه، نوسان ده‌روزه و میانگین متحرک پنج‌روزه حجم بودند. برای تهران، فاصله از باند مرجع فرضی، میانگین پنج‌روزه فاصله، نزدیکی به باند، بازده مثبت یا منفی همراه با کاهش نسبی حجم و متغیر مجازی دوره نیز ساخته شدند. مجموعه ترکیبی هر دو گروه را در بر گرفت. گام تصادفی، رگرسیون خطی و لایت‌جی‌بی‌ام در پنج مرحله با پنجره آموزش بسط‌یابنده مقایسه شدند. آموزش و اعتبارسنجی پیش از هر بلوک آزمون قرار گرفتند و برای جلوگیری از نشت اطلاعات بازده روز آینده، فاصله‌ای یک‌ردیفی در مرزهای انتخاب مدل اعمال شد. پیش‌پردازش تنها بر داده‌های آموزش تکیه

تاریخ دریافت: ۱۴۰۵/۰۳/۰۳ تاریخ پذیرش:۱۴۰۵/۰۶/۳۰

* دانشجوی کارشناسی ارشد، گروه اقتصاد، دانشکده معارف اسلامی و اقتصاد، دانشگاه امام صادق(ع)، تهران، ایران (نویسنده مسئول). E-mail: masud.fazli.shahri@gmail.com

** دانشیار، گروه اقتصاد مالی، دانشکده معارف اسلامی و اقتصاد، دانشگاه امام صادق(ع)، تهران، ایران.

داشت. ارزیابی با خطاهای پیش‌بینی، ضریب تعیین خارج از نمونه و دقت جهت انجام شد. معناداری اختلاف زیان مربعی نیز با آزمون دیبولد–ماریانو و خطای معیار مقاوم به ناهمسانی واریانس و خودهمبستگی سنجیده شد. برای تفسیر سهم ویژگی‌ها در پیش‌بینی مدل‌های برآوردی، مقادیر شاپ نیز برای مشاهدات آزمون محاسبه شد. برای سنجش پایداری نتایج، آستانه حجم و سهم اولیه آموزش تغییر داده شد و زیرنمونه پیش از مرز زمانی نیز ارزیابی شد. مقایسه تنظیمات منتخب و پیش‌فرض مدل درختی، جداگانه برای سنجش حساسیت عملکرد به ابرپارامترها انجام شد.

یافته‌ها: در بازاجرای داده‌های بازسازی‌شده، میانگین معیار جذر میانگین مربعات خطا در پنج مرحله آزمون برای گام تصادفی، مدل خطی و مدل درختی برای مجموعه متعارف تهران به‌ترتیب ۰/۰۱۱۳۰۰، ۰/۰۱۰۵۰۴ و ۰/۰۱۰۷۰۹ بود. مقایسه مستقیم مدل خطی با گام تصادفی با احتمال ۰/۰۰۰۲ و مدل درختی با گام تصادفی با احتمال ۰/۰۰۱۴ به نفع مدل‌های برآوردی بود. اختلاف مدل درختی و خطی در نمونه کامل معنادار نبود. در اس‌لندپی ۵۰۰، هیچ‌یک از دو مدل در مشخصات مبنا برتری معناداری نسبت به معیار مرجع نداشتند. در تهران، متغیرهای تقریبی به‌تنهایی بهبود معناداری نسبت به گام تصادفی ایجاد نکردند و افزودن آن‌ها به مجموعه متعارف نیز خطای مدل خطی را کاهش نداد و برای مدل درختی تنها بهبود محدودی ایجاد کرد. بهبود مدل‌های تنظیم‌شده تهران در آستانه‌های حجم ۶۰، ۷۰ و ۸۰ درصد و زیرنمونه پیش از مرز زمانی باقی ماند؛ اما عملکرد مدل درختی با تنظیمات پیش‌فرض ضعیف‌تر شد. این مقایسه نشان‌دهنده حساسیت عملکرد به ابرپارامترهاست، نه تأیید استواری مدل؛ حفظ بهبود در تغییرات بررسی‌شده حجم و نمونه نیز به معنای پایداری در برابر تغییر تنظیمات نیست. تحلیل شاپ برای توضیح سهم ویژگی‌ها در خروجی مدل به‌کار رفت؛ این سهم به‌تنهایی معیاری برای سنجش بهبود دقت پیش‌بینی نیست.

نتیجه‌گیری: گام تصادفی با بازده پیش‌بینی‌شده صفر، معیار آماری مقایسه پیش‌بینی‌هاست و آزمون کامل شکل ضعیف کارایی محسوب نمی‌شود؛ زیرا کارایی بازار الزاماً بازده مورد انتظار صفر را ایجاب نمی‌کند. بنابراین، نبود بهبود معنادار در اس‌لندپی ۵۰۰ اثبات کارایی نیست و کاهش خطای پیش‌بینی در تهران نیز به‌تنهایی ناکارایی یا سودآوری خالص را اثبات نمی‌کند. عملکرد لایت‌جی‌بی‌ام به ابرپارامترها حساس است و مقایسه با تنظیمات پیش‌فرض، شاهد استواری مدل نیست. شواهد کافی از برتری پایدار مدل غیرخطی به‌دست نیامد و معنادار نبودن اختلاف، برابری مدل‌ها را اثبات نمی‌کند. مقایسه دو بازار توصیفی است و اثر علّی مقررات یا ساختار بازار را مشخص نمی‌کند.

کلیدواژه‌ها: پیش‌بینی‌پذیری بازده؛ بازار سرمایه؛ ریزساختار بازار؛ فرضیه بازار کارا؛ یادگیری ماشین

استناد دهی: فضلی، مسعود و شعبانی، احمد. (۱۴۰۵). آزمون شکل ضعیف کارایی بازار بر اساس پیش‌بینی‌پذیری بازده: مقایسه شاخص کل بورس تهران و S&P500 با رویکرد اقتصادسنجی و یادگیری ماشین. *چشم‌انداز مدیریت مالی*، ۱۶(۲)، ۹-۲۸.

۱. مقدمه

کارایی اطلاعاتی بازار به میزان انعکاس اطلاعات در قیمت دارایی‌ها و سرعت واکنش قیمت‌ها به اطلاعات جدید اشاره دارد. بر اساس فرضیه بازار کارا، قیمت دارایی‌های مالی باید اطلاعات موجود را به‌سرعت و به‌طور کامل منعکس کند؛ به‌گونه‌ای که استفاده از اطلاعات تاریخی قیمت‌ها و حجم معاملات نتواند به ایجاد بازده‌های غیرعادی و قابل‌پیش‌بینی منجر شود (Fama, 1970). در چارچوب شکل ضعیف فرضیه بازار کارا، بررسی اینکه آیا داده‌های تاریخی بازار حاوی اطلاعاتی درباره بازده‌های آینده هستند یا خیر، یکی از رویکردهای اصلی برای ارزیابی تجربی کارایی بازارها محسوب می‌شود. برای نمونه، بررسی داده‌های هفتگی بازار سهام ایالات متحده طی سال‌های ۱۹۶۲ تا ۱۹۸۵ به رد فرض گام تصادفی برای مجموعه‌ای از شاخص‌ها و پرتفوی‌های مرتب‌شده بر اساس اندازه منجر شده است (Lo & MacKinlay, 1988). این شاهد به نمونه و افق همان مطالعه مربوط است و نباید به همه بازارها تعمیم داده شود.

شواهد مقایسه‌ای نشان می‌دهد پیش‌بینی‌پذیری بازده در بازارهای مختلف یکسان نیست. بررسی بازارهای نوظهور نشان داده است که بازده این بازارها، در مقایسه با کشورهای توسعه‌یافته، بیشتر از اطلاعات محلی تأثیر می‌پذیرد (Harvey, 1995). این یافته زمینه‌ای برای بررسی تفاوت محتوای اطلاعاتی داده‌ها فراهم می‌کند؛ بااین‌حال، از آن نمی‌توان به‌تنهایی نتیجه گرفت که تفاوت نقدشوندگی، زیرساخت یا حضور سرمایه‌گذاران نهادی علت اختلاف دقت پیش‌بینی در نمونه حاضر است. در بازارهای نوظهور، سازوکارهای معاملاتی می‌توانند سرعت کشف قیمت و واکنش بازار به اطلاعات جدید را تغییر دهند. دامنه نوسان در بورس تهران با هدف مهار نوسان‌های شدید اعمال می‌شود، اما ممکن است تعدیل قیمت و رفتار معامله‌گران را نیز تحت تأثیر قرار دهد. بنابراین، ارزیابی پیش‌بینی‌پذیری بازده این بازار باید با توجه به ویژگی‌های نهادی و ریزساختاری آن انجام شود.

در سال‌های اخیر، توسعه روش‌های یادگیری ماشین و افزایش دسترسی به داده‌های مالی، چارچوب جدیدی برای مطالعه قابلیت پیش‌بینی بازارهای مالی فراهم کرده است. در مقایسه با رگرسیون خطی، الگوریتم‌های غیرخطی یادگیری ماشین امکان مدل‌سازی انعطاف‌پذیرتر روابط و تعامل میان متغیرهای توضیحی را فراهم می‌کنند. به همین دلیل، این روش‌ها در مطالعات مرتبط با پیش‌بینی بازده، قیمت‌گذاری دارایی‌ها و تحلیل رفتار بازار مورد توجه گسترده قرار گرفته‌اند (Gu et al., 2020). همچنین، الگوریتم‌های مبتنی بر درخت تقویت‌شده مانند لایت‌جی‌بی‌ام[3] به‌دلیل کارایی محاسباتی و توانایی در پردازش روابط غیرخطی، کاربرد گسترده‌ای در مسائل پیش‌بینی مالی پیدا کرده‌اند (Ke et al., 2017).

پیچیدگی مدل زمانی ارزش پیش‌بینی دارد که در داده‌های دیده‌نشده نیز خطا را نسبت به معیارهای ساده کاهش دهد. ازاین‌رو، مدل‌ها با ارزیابی خارج از نمونه و آزمون اختلاف دقت مقایسه می‌شوند. ادبیات موجود سه مسئله نزدیک اما متمایز را بررسی کرده است: آزمون گام تصادفی، پیش‌بینی بازده یا قیمت با روش‌های یادگیری ماشین و توضیح نقش ویژگی‌ها در خروجی مدل. تفاوت متغیر هدف، افق پیش‌بینی و سطح تجمیع داده‌ها، مقایسه مستقیم نتایج این مطالعات را دشوار می‌کند. پژوهش حاضر بازده یک‌روزه دو شاخص را با یک طراحی زمانی مشترک ارزیابی می‌کند و مدل‌های خطی و غیرخطی را در برابر معیار بازده صفر می‌سنجد. در تهران، نقش نشانگرهای تقریبی شرایط معاملاتی نیز جداگانه بررسی می‌شود.

داده‌های هر بازار جداگانه و با حفظ ترتیب زمانی تحلیل می‌شوند. این شیوه از جابه‌جایی مشاهدات میان تقویم‌های معاملاتی متفاوت جلوگیری می‌کند و ارزیابی خارج از نمونه را به اطلاعات در دسترس هر مقطع محدود نگه می‌دارد.

[3] Light Gradient Boosting Machine (LightGBM)

نتایج بر پایه تغییر خطای پیش‌بینی و آزمون اختلاف زیان گزارش می‌شوند. این طراحی سودآوری پس از هزینه معاملات یا اثر علّی مقررات بازار را نمی‌سنجد؛ بنابراین، تفسیر یافته‌ها به شواهد آماری همین نمونه محدود است.

پرسش اصلی این است که آیا شواهد پیش‌بینی خارج از نمونه با شکل ضعیف کارایی در دو بازار سازگار است. دو پرسش تکمیلی نیز بررسی می‌شود: آیا مدل غیرخطی نسبت به مدل خطی مزیت پایداری دارد و آیا نشانگرهای تقریبی شرایط معاملاتی تهران اطلاعات مکملی فراهم می‌کنند؟

۲. مبانی نظری و پیشینه پژوهش

نظریه مبنای پژوهش، فرضیه بازار کارای فاما است. در شکل ضعیف این فرضیه، قیمت‌ها اطلاعات تاریخی قیمت و حجم معاملات را منعکس می‌کنند؛ بنابراین، استفاده از این اطلاعات نباید امکان کسب بازده غیرعادی پایدار را پس از لحاظ ریسک و هزینه معاملات فراهم کند. بااین‌حال، بازده مورد انتظار می‌تولند در طول زمان تغییر کند و پیش‌بینی‌پذیری آماری بازده، به‌تنهایی، ناکارایی را اثبات نمی‌کند. آزمون تجربی کارایی به مدل بازده مورد انتظار نیز وابسته است. در پژوهش حاضر، این نظریه چارچوب تفسیر مقایسه‌های خارج از نمونه را فراهم می‌کند؛ برتری بر معیار بازده صفر فقط شاهد بهبود دقت نسبت به این معیار است و برای داوری اقتصادی به بررسی ریسک، هزینه معاملات و امکان اجرای راهبرد نیاز دارد (Fama, 1970 ;Campbell et al., 1997).

کاربرد آزمون نسبت واریانس برای بازده هفتگی شاخص‌ها و پرتفوی‌های بازار سهام ایالات متحده طی سال‌های ۱۹۶۲ تا ۱۹۸۵ به رد فرض گام تصادفی در نمونه مورد بررسی منجر شد و این رد عمدتاً با رفتار سهام کوچک مرتبط بود (Lo & MacKinlay, 1988). نتیجه این آزمون، وجود انحراف از یک مدل آماری مشخص را نشان می‌دهد و به‌تنهایی معادل اثبات سودآوری پس از هزینه معاملات نیست. در پژوهش حاضر نیز معیار گام تصادفی برای مقایسه دقت پیش‌بینی به‌کار می‌رود؛ تفاوت آن است که ارزیابی بر خطای بازده روز آینده در مشاهدات خارج از نمونه متمرکز است.

بررسی بازده بازارهای نوظهور نشان داده است که اطلاعات محلی در پیش‌بینی این بازارها نسبت به بازارهای توسعه‌یافته نقش بیشتری دارد (Harvey, 1995). این یافته ضرورت مقایسه محتوای اطلاعاتی داده‌های تاریخی در محیط‌های متفاوت را برای پژوهش حاضر روشن می‌کند؛ بااین‌حال، تفاوت در نمونه، متغیرها و مدل‌های دو مطالعه مانع از یکسان‌انگاری نتایج آن‌ها می‌شود. سازوکارهای معاملاتی در بازارهای نوظهور می‌توانند بر سرعت کشف قیمت اثر بگذارند. در بازار پایه فرابورس ایران، کاهش دامنه نوسان در همه تابلوها به کاهش نوسان‌پذیری درون‌روز منجر نشده است (Hasannezhad et al., 2023). همچنین، شواهدی از اثر مغناطیسی محدودیت قیمت گزارش شده است (Jafaripour et al., 2021). این یافته‌ها توجه به شرایط معاملاتی تهران را توجیه می‌کنند، اما نوسان درون‌روز یا اثر مغناطیسی مستقیماً پیش‌بینی‌پذیری بازده روز بعد شاخص کل را اندازه نمی‌گیرد. نسبت‌دادن اختلاف دو بازار به یک مقرره مشخص نیز به طراحی علّی جداگانه نیاز دارد.

مطالعات داخلی کارایی بازار از نظر متغیر هدف و سطح تجمیع یکسان نیستند. برای نمونه، وابستگی بازده با خودرگرسیون چندکی سنجیده شده است (Alizadeh Chamazkoti et al., 2024). سنجش وابستگی در چندک‌های توزیع بازده با ارزیابی خطای پیش‌بینی مقدار بازده شاخص کل متفاوت است. ازاین‌رو، تفاوت نتایج این دو روش لزوماً به معنای تعارض در شواهد کارایی نیست. پژوهش حاضر بر ارزیابی زمانی خارج از نمونه و معیار خطای مشترک تمرکز دارد. در ادبیات داخلی قدیمی‌تر نیز ابعاد کارایی و عملکرد اقتصادی بورس تهران بررسی شده‌اند (Namazi, 2003 ;Namazi & Shooshtarian, 1996) و با رویکرد چندفراکتالی، ناهمگنی کارایی صنایع گزارش شده است (Oujimehr et al., 2020). این مطالعات از نظر دوره، سطح تجمیع و تعریف کارایی با آزمون پیش‌بینی

خارج از نمونه حاضر متفاوت‌اند؛ این تفاوت‌ها نشان می‌دهند که برای مقایسه شواهد داخلی باید روش سنجش و دامنه هر مطالعه را در نظر گرفت.

در حوزه قیمت‌گذاری تجربی دارایی‌ها، مقایسه روش‌های یادگیری ماشین برای پیش‌بینی صرف ریسک سهام بهبودهایی نسبت به راهبردهای رگرسیونی مرجع نشان داده است (Gu et al., 2020). موضوع آن مطالعه قیمت‌گذاری تجربی دارایی‌هاست؛ ازاین‌رو، یافته‌های آن نباید مستقیماً به برتری مدل غیرخطی در پیش‌بینی بازده روزانه یک شاخص تعبیر شود. این تمایز، مقایسه مستقل مدل خطی و غیرخطی در چارچوب پژوهش حاضر را ضروری می‌سازد.

الگوریتم لایت‌جی‌بی‌ام برای آموزش کارآمد درخت‌های تقویت‌شده معرفی شده است (Ke et al., 2017)؛ بااین‌حال، معرفی یک الگوریتم به‌خودی‌خود شاهد برتری آن در بازار مالی نیست. در مطالعات مربوط به ایران، از داده‌های چهار گروه بازار سهام ایران در فاصله نوامبر ۲۰۰۹ تا نوامبر ۲۰۱۹ برای پیش‌بینی با روش‌های یادگیری ماشین و یادگیری عمیق استفاده شده است (Nabipour et al., 2020). همچنین، بازده شاخص قیمت هشت صنعت تولیدی بورس تهران در دوره ۱۳۹۷ تا ۱۴۰۲ با جنگل تصادفی، تنظیم ابرپارامترها به کمک الگوریتم ژنتیک و مقادیر شاپ بررسی شده است (Raei et al., 2025). در آن مطالعه، متغیرهای تکنیکال، حجم معاملات و سهام شناور در تفسیر مدل برجسته بودند. این مطالعات به پیش‌بینی بازده و تفسیر خروجی مدل‌ها کمک می‌کنند؛ اما اهمیت متغیر در مدل با بهبود دقت نسبت به معیار مرجع یکسان نیست.

در یکی از مطالعات مرتبط با مدیریت مالی، پیش‌بینی بازده با شبکه حافظه بلند کوتاه‌مدت، جنگل تصادفی و آریما در بهینه‌سازی سبد سرمایه‌گذاری به‌کار گرفته شده است و بررسی شاخص پنج صنعت، عملکرد بهتر بهینه‌سازی میانگین–واریانس با پیش‌بینی جنگل تصادفی را نشان داده است (Eghtesad & Mohammadi, 2023). این نتیجه کاربرد پیش‌بینی در تصمیم تخصیص دارایی را نشان می‌دهد؛ اما عملکرد سبد، علاوه بر دقت پیش‌بینی، به قاعده تخصیص و سنجه ریسک وابسته است. ازاین‌رو، برتری یک سبد را نمی‌توان مستقیماً معادل رد شکل ضعیف کارایی یا برتری همان مدل در پیش‌بینی بازده شاخص کل دانست.

برای پیش‌بینی قیمت ده سهم بورس تهران، مدل‌های ترکیبی شبکه عصبی کانولوشن و حافظه بلند کوتاه‌مدت با تنظیم ابرپارامترها به کمک بهینه‌سازی ازدحام ذرات بررسی شده‌اند و بهبود معیارهای خطا و عملکرد برخی راهبردهای معاملاتی گزارش شده است (Gholami & Shams Gharneh, 2024). این مطالعه اهمیت تنظیم مدل را نشان می‌دهد، ولی هدف آن قیمت سهام منفرد است و از متغیرهای نرخ ارز و تورم نیز استفاده می‌کند. بنابراین، یافته‌های آن مستقیماً با پیش‌بینی بازده روزانه شاخص بر پایه اطلاعات تاریخی بازار قابل‌مقایسه نیست. همچنین، بهبود حاصل از تنظیم ابرپارامترها، به‌خودی‌خود شاهد پایداری عملکرد در برابر تغییر تنظیمات نیست؛ این تمایز در تحلیل حساسیت پژوهش حاضر رعایت می‌شود.

پیش‌بینی جهت شاخص کل بورس تهران با ترکیب تقویت گرادیانی و الگوریتم ژنتیک و با استفاده از متغیرهای تکنیکال و بنیادی بررسی شده و نتایج آن در چارچوب شکل ضعیف و نیمه‌قوی کارایی تفسیر شده است (Heydari Delooei et al., 2026). تفاوت پژوهش حاضر با آن مطالعه، تمرکز بر مقدار بازده روز آینده، مقایسه هم‌زمان دو شاخص و استفاده از مدل خطی و پیش‌بینی بازده صفر به‌عنوان معیارهای رقابتی است. دقت طبقه‌بندی جهت در آن پژوهش، با خطای پیش‌بینی مقدار بازده در این مطالعه مستقیماً قابل‌مقایسه نیست.

در مقایسه مدل‌های پیش‌بینی، کاهش عددی خطا لزوماً نشانه برتری آماری نیست. برای بررسی معناداری اختلاف دقت پیش‌بینی می‌توان از آزمون دیبولد–ماریانو استفاده کرد تا تفاوت عملکرد مدل‌ها با توجه به نوسانات نمونه‌ای تفسیر شود (Diebold & Mariano, 1995).

مرور مطالعات نشـان می‌دهد که تفسـیر نتایج به متغیر هدف، مجموعه اطلاعات و معیار ارزیابی وابسـته اسـت. پیش‌بینی قیمت، جهت شاخص، بازده صنایع و عملکرد سبد یک مسئله واحد نیستند و اهمیت یک ویژگی در توضیح خروجی مدل نیز ارزش افزوده آن برای دقت پیش‌بینی را اثبات نمی‌کند. پژوهش حاضر بازده روز آینده دو شاخص را با مجموعه‌های اطلاعاتی مشخص و یک چارچوب زمانی مشترک مقایسه می‌کند؛ تعمیم نتایج به دارایی‌ها، افق‌ها یا تنظیمات دیگر به آزمون جداگانه نیاز دارد.

بر این اساس، مسئله پژوهش حاضر سنجش دقت اضافی حاصل از افزایش پیچیدگی مدل و افزودن متغیرهای تقریبی شرایط معاملاتی در یک ارزیابی زمانی مشترک است. این مسئله با مقایسه مدل‌ها و مجموعه‌های اطلاعاتی در تهران و اس‌اندپی ۵۰۰ بررسی می‌شود. در این چارچوب، پاسخ به سؤال اصلی از مقایسه مستقیم مدل‌های مبتنی بر اطلاعات تاریخی با معیار گام تصادفی حاصل می‌شود. مقایسه مدل غیرخطی با مدل خطی و مقایسه مجموعه متعارف با مجموعه ترکیبی، به‌ترتیب پاسـخ‌گوی پرسـش‌های تکمیلی درباره پیچیدگی مدل و محتوای اطلاعاتی متغیرهای تقریبی‌اند. جهت و معناداری این تفاوت‌ها از پیش‌فرض نمی‌شود. پژوهش بر پرسش‌های فوق استوار است و فرضیه جهت‌دار مستقلی درباره برتری یک بازار، مدل یا مجموعه ویژگی‌ها مطرح نمی‌کند.

۳. روش‌شناسی پژوهش

طراحی پژوهش و داده‌ها: این پژوهش از نظر هدف کاربردی و از نظر روش کمی و تجربی اسـت. برای آزمون شکل ضعیف کارایی بازار سرمایه، بررسی می‌شود که اطلاعات تاریخی قیمت و حجم معاملات تا چه اندازه می‌تواند بازده آینده را در خارج از نمونه پیش‌بینی کند. هر مشاهده به یک روز معاملاتیِ شاخص مربوط است و متغیر هدف، بازده لگاریتمی روز معاملاتی بعد اسـت. نتایج پیش‌بینی با توجه به محدودیت‌های آزمون تجربی کارایی تفسـیر می‌شوند.

مقایسه با پیش‌بینی بازده صفر، دقت مدل‌ها را نسبت به معیار مرجع می‌سنجد. مقایسه رگرسیون خطی و مدل درختی نیز مشخص می‌کند که آیا انعطاف بیشتر مدل، دقت پیش‌بینی را افزایش داده است.

جامعه آماری پژوهش شامل داده‌های روزانه دو بازار مالی با ویژگی‌های نهادی متفاوت است. به‌عنوان نماینده یک بازار توسعه‌یافته، شاخص استاندارد اند پورز ۵۰۰ انتخاب شده است که یکی از شاخص‌های اصلی بازار سهام ایالات متحده محسـوب می‌شـود. در مقابل، شـاخص کل بورس اوراق بهادار تهران به‌عنوان نماینده یک بازار نوظهور مورد اسـتفاده قرار گرفته اسـت. انتخاب این دو بازار امکان مقایسه محیط‌هایی با تفاوت در عمق بازار، سـاختار معاملاتی، کیفیت جریان اطلاعات و محدودیت‌های نهادی را فراهم می‌کند.

برای مقایسه مدل‌ها در هر بازار، تاریخ‌های آزمون یکسان و مجموعه ورودی مشخصی در نظر گرفته می‌شود. بااین‌حال، مقایسه نتایج دو بازار ماهیت توصیفی دارد؛ زیرا تقویم معاملاتی، ترکیب شاخص و ویژگی‌های داده‌های آن‌ها متفاوت است و این تفاوت‌ها باید در تفسیر یافته‌ها لحاظ شوند.

داده‌ها، متغیرها و آماده‌سـازی داده‌ها: داده‌های خام تهران ۲۷۲۱ روز معاملاتی از ۷ مارس ۲۰۱۵ تا ۲۹ ژوئن ۲۰۲۶ و داده‌های اس‌اندپی ۵۰۰، ۲۸۷۷ روز معاملاتی از ۱۶ ژانویه ۲۰۱۵ تا ۲۶ ژوئن ۲۰۲۶ را پوشـش می‌دهد. پس از حذف ده مشاهده نخست برای تشکیل پنجره‌های گذشته و حذف آخرین مشاهده به‌دلیل نبود شاخص روز بعد، نمونه‌های قابل مدل‌سـازی تهران و اس‌لندپی به‌ترتیب ۲۷۱۰ و ۲۸۶۶ مشـاهده دارند. هر بازار بر اسـاس تقویم معاملاتی خود مدل‌سازی شد و داده‌های خام، داده‌های پردازش‌شده و تاریخ تحقق بازده هدف در فایل مکمل ارائه شده‌اند.

بازده روزانه شاخص‌ها به‌صورت بازده لگاریتمی محاسبه می‌شود:

$$R_t = ln\left(\frac{P_t}{P_{t-1}}\right) \quad (۱)$$

در رابطه (۱)، مقدار شاخص در روز جاری با P و بازده لگاریتمی با R نمایش داده می‌شود. متغیر هدف، بازده روز معاملاتی بعد است؛ بنابراین فقط اطلاعات در دسترس تا پایان روز جاری برای پیش‌بینی استفاده می‌شود.

متغیرهای پژوهش در سه مجموعه متعارف، تقریبی شرایط معاملاتی و ترکیبی دسته‌بندی می‌شوند. مجموعه متعارف از بازده جاری و وقفه‌های آن، نوسان و میانگین حجم تشکیل شده است. مجموعه دوم شامل نشانگرهایی است که از داده‌های تجمیعی شاخص تهران ساخته شده‌اند و یک متغیر مجازی دوره نیز دارد. این متغیرها نشانگرهای تقریبی شرایط معاملاتی‌اند و از الگوهای قیمت و حجم ساخته می‌شوند؛ بنابراین، مشاهده مستقیم ریزساختار بازار محسوب نمی‌شوند. مجموعه ترکیبی نیز با کنار هم قرار دادن این دو گروه ساخته می‌شود.

جدول ۱ *.تعریف متغیرهای مورد استفاده در پژوهش*

Table 1. Definition of Research Variables

گروه	متغیر	تعریف
هدف	بازده آتی	بازده لگاریتمی روز معاملاتی بعد
متعارف	بازده لگاریتمی	بازده لگاریتمی روز جاری
متعارف	وقفه‌های بازده	بازده‌های با وقفه یک تا سه روز
متعارف	نوسان تاریخی	انحراف معیار متحرک ده‌روزه بازده
متعارف	حجم معاملات	میانگین متحرک پنج‌روزه حجم معاملات
تقریبی	فاصله از باند مرجع فرضی	فاصله نرمال‌شده شاخص از نزدیک‌ترین حد باند فرضی
تقریبی	میانگین فاصله از باند فرضی	میانگین متحرک پنج‌روزه فاصله نرمال‌شده
تقریبی	نزدیکی به باند فرضی	نشانگر قرارگرفتن شاخص نزدیک حد باند فرضی
تقریبی	نشانگر بازده مثبت و کاهش حجم	بازده مثبت و نسبت حجم کمتر از آستانه
تقریبی	نشانگر بازده منفی و کاهش حجم	بازده منفی و نسبت حجم کمتر از آستانه
تقریبی	دوره زمانی	متغیر مجازی تقسیم نمونه در تاریخ تعیین‌شده

منبع: یافته‌های پژوهش.

تعریف عملیاتی متغیرها در جدول ۱ آمده است. برای سنجش اینکه متغیرهای تقریبی چه اطلاعاتی به مجموعه متعارف اضافه می‌کنند، عملکرد مجموعه متعارف و ترکیبی مقایسه می‌شود. این مقایسه، ارزش افزوده متغیرهای تقریبی را برای پیش‌بینی می‌سنجد، نه توانایی آن‌ها را در اندازه‌گیری مستقیم صف سفارش.

برای ساخت متغیرهای مرتبط با محدودیت دامنه نوسان، تاریخ ۲۱ مهر ۱۴۰۳ (۱۲ اکتبر ۲۰۲۴) به‌عنوان مرز تقسیم نمونه انتخاب شد. ضرایب در دو سوی این مرز، پارامترهای یک باند مرجع فرضی هستند. این ۰٫۰۳ و ۰٫۰۷ باند معیار یکسانی برای سنجش نزدیکی مقدار شاخص به یک محدوده مرجع فراهم می‌کند تا محتوای اطلاعاتی این نزدیکی در پیش‌بینی بازده بررسی شود. بنابراین، ضرایب یادشده نه حدود واقعی مجاز نوسان شاخص را نشان می‌دهند و نه تمام تغییرات تاریخی مقررات بازار را بازتاب می‌دهند.

در روابط (۲) تا (۶)، نمادهای H ،P و L به‌ترتیب نشان‌دهنده مقدار پایانی، بیشینه و کمینه روزانه شاخص در روز t هستند. همچنین، b ضریب باند مرجع، U و D به‌ترتیب حدود بالایی و پایینی این باند، d فاصله نرمال‌شده قیمت از باند و N نشانگر نزدیکی قیمت به باند را نشان می‌دهند. عملگر محدودسازی نیز برای قرار دادن مقادیر محاسبه‌شده در بازه صفر تا یک به‌کار رفته است.

$$b_t = \begin{cases} 0.07 & date < 2024-10-12 \\ 0.03 & date \geq 2024-10-12 \end{cases} \quad (۲)$$

$$U_t = P_{t-1}(1 + b_t); \quad D_t = P_{t-1}(1 - b_t) \quad (۳)$$

$$d_t = clip\left(\frac{min(U_t - H_t, L_t - D_t)}{P_{t-1} b_t}, 0, 1\right) \quad (۴)$$

$$m_t = \frac{1}{5}\sum_{j=0} 4d_{t-j} \quad (۵)$$

$$N_t = \mathbb{1}\left\{\frac{min(U_t - H_t, L_t - D_t)}{P_{t-1}} \leq 0.30 b_t\right\} \quad (۶)$$

در روابط (۷) تا (۱۰)، V حجم معاملات ثبت‌شـــده در روز t و ρ نســـبت این حجم به میانگین حجم معاملات گذشته است. نماد q آستانه تعیین‌شده را نشان می‌دهد و تابع 𝟙 مشخص می‌کند که شرط موردنظر برقرار است یا خیر. میانگین حجم در مخرج نسبت، تنها از پنج روز معاملاتی پیشین محاسبه می‌شود؛ بنابراین حجم روز جاری در این میانگین وارد نمی‌شـــود. این نکته، تفاوت آن را با میانگین متحرک پنج‌روزه حجم در مجموعه متعارف روشـــن می‌کند؛ میانگین اخیر، حجم روز جاری را نیز در بر می‌گیرد.

در مشخصات اصلی پژوهش، آستانه q برابر درصد میانگین حجم ۷۰ است. اگر حجم روز جاری به کمتر از ۰٫۷۰ پنج روز گذشته برسد، بسته به مثبت یا منفی بودن بازده، نشانگر متناظر فعال می‌شود. این نشانگرها فقط ترکیب جهت بازده و کاهش نسبی حجم را ثبت می‌کنند و وجود صف سفارش، حجم سفارش‌های اجرانشده یا فشار خالص خرید و فروش را اثبات نمی‌کنند. ازاین‌رو، تفسیر آن‌ها به اطلاعات قابل استخراج از قیمت و حجم معاملات محدود است. اگر میانگین حجم پنج روز گذشته صفر باشد، نسبت حجم تعریف نمی‌شود و مقدار نشانگر مربوطه صفر در نظر گرفته می‌شود.

$$\overline{V}_{t-1,5} = \frac{1}{5}\sum_{j=1} 5V_{t-j} \quad (۷)$$

$$\rho_t = \frac{V_t}{\overline{V}_{t-1,5}} \quad (۸)$$

$$I^+_t(q) = \mathbb{1}\{R_t > 0 \wedge \rho_t < q\} \quad (۹)$$

$$I^-_t(q) = \mathbb{1}\{R_t < 0 \wedge \rho_t < q\} \quad (۱۰)$$

در نسخه پیشین داده‌ها، بررسی‌های تشخیصی شامل آزمون دیکی—فولر تعمیم‌یافته[4] برای مانایی و استفاده از عامل تورم واریانس[5] و ماتریس همبستگی برای هم‌خطی بود. عامل تورم واریانس بیش از ۲۰ یا قدرمطلق همبستگی بیش از نشانه هم‌خطی شدید در نظر گرفته شد. این بررسی‌ها به همان نسخه پیشین مربوط‌اند و به معنای ۰/۹۵، اجرای مجدد آزمون‌ها روی نمونه بازسازی‌شده نیستند.

مدل‌های پیش‌بینی و طراحی ارزیابی: عملکرد سه رویکرد گام تصادفی بدون رانش، رگرسیون خطی چندمتغیره و درخت‌های تقویت‌شده گرادیانی مقایسه می‌شود. برای اینکه مقایسه دو مدل برآوردی بر مبنای یکسانی انجام شود، در هر مقایسه، مجموعه متغیرهای ورودی و مشاهدات آزمون آن‌ها یکسان در نظر گرفته می‌شود.

در گام تصادفی بدون رانش، مقدار شاخص روز آینده برابر مقدار امروز پیش‌بینی می‌شود؛ بنابراین، پیش‌بینی بازده لگاریتمی روز آینده صفر است. این تعریف، معیار آماری مشخصی برای مقایسه دقت مدل‌ها فراهم می‌کند. بااین‌حال، باید میان این معیار و فرضیه کلی کارایی تمایز گذاشت، زیرا کارایی بازار الزاماً به معنای صفر بودن بازده مورد انتظار نیست.

[4] Augmented Dickey–Fuller (ADF) test
[5] Variance Inflation Factor (VIF)

$$\hat{R}_{t+1|t} = 0 \qquad (۱۱)$$

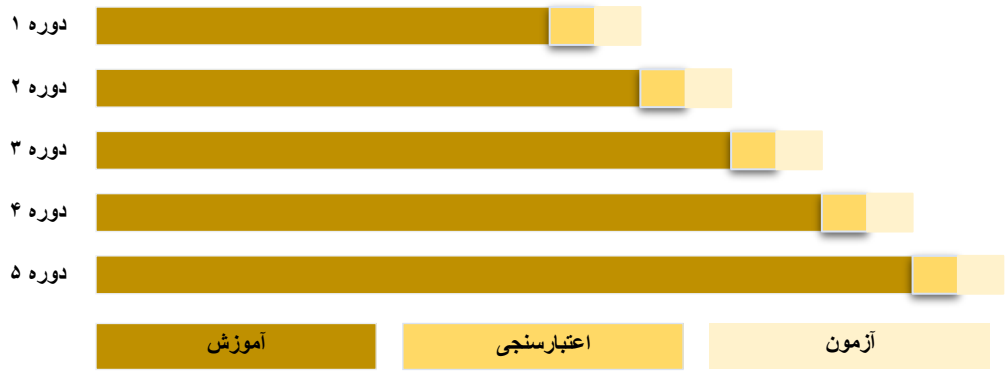
رگرسیون خطی، بازده آینده را به‌صورت ترکیب خطی ویژگی‌ها همراه با عرض از مبدأ برآورد می‌کند. مقیاس‌بندی ویژگی‌ها فقط با داده‌های آموزش هر مرحله انجام می‌شود.

$$R_{t+1} = \beta_0 + \sum_{j=1} k\beta_j X_{j,t} + \varepsilon_{t+1} \qquad (۱۲)$$

مدل لایت‌جی‌بی‌ام از درخت‌های تقویت‌شده برای یادگیری روابط غیرخطی و تعامل ویژگی‌ها استفاده می‌کند (Ke et al., 2017). توانایی آن در نمایش روابط انعطاف‌پذیرتر، تضمین‌کننده بهبود خارج از نمونه نیست.

ابرپارامترهای لایت‌جی‌بی‌ام در هر مرحله با استفاده از داده‌های آموزش و اعتبارسنجی زمانی انتخاب می‌شوند. در شبکه جست‌وجو، تعداد برگ‌ها برابر ۱۵ یا ۳۱، عمق درخت برابر ۴ یا ۶ و نرخ یادگیری برابر در نظر ۰/۰۵ یا ۰/۰۳ گرفته می‌شود. سقف تعداد درخت‌ها ۱۰۰۰ است. اگر عملکرد بهبود نیابد، توقف زودهنگام در مرحله جست‌وجو پس از ۳۰ دور و در برازش با تنظیمات منتخب پس از ۵۰ دور اعمال می‌شود. داده‌های آزمون در هیچ‌یک از این انتخاب‌ها دخالت ندارند.

ارزیابی در پنج مرحله و با پنجره آموزش بسط‌یابنده انجام شد؛ با پیشرفت زمان، داده‌های بیشتری وارد آموزش می‌شود. در مشخصات مبنا، ۵۰ درصد نخست داده‌ها مجموعه اولیه آموزش است. بخش باقی‌مانده به پنج بلوک زمانی تقسیم می‌شود و در هر بلوک، نیمه نخست برای اعتبارسنجی و نیمه دوم برای آزمون به‌کار می‌رود. آخرین ردیف هر بخش آموزش و اعتبارسنجی حذف می‌شود تا بازده هدفِ آن ردیف به روزی در بخش بعدی وابسته نباشد. در نتیجه، همه مقادیر هدف مورد استفاده در انتخاب مدل پیش از آغاز بخش بعدی مشاهده شده‌اند. این تقسیم‌بندی تصادفی نیست و تاریخ‌های دقیق آن در فایل مکمل آمده است. این طراحی با اصول ارزیابی زمان‌مند پیش‌بین‌های سری زمانی سازگار است (Tsay, 2010 ;Bergmeir & Benítez, 2012).



شکل ۱ *.طرح شماتیک پنجره‌های بسط‌یابنده تهران؛ فاصله یک‌ردیفی در مرزهای انتخاب مدل*

Figure 1. Schematic of Tehran expanding windows with one-row gaps at model-selection boundaries.

منبع: یافته‌های پژوهش.

یکسان بودن تاریخ‌های آزمون در هر بازار امکان مقایسه زوجی خطاها را فراهم می‌کند؛ یعنی خطای دو مدل برای بازده تحقق‌یافته در روزی واحد مقایسه می‌شود.

ارزیابی عملکرد و آزمون معناداری تفاوت مدل‌ها: عملکرد مدل‌ها در دو سطح ارزیابی می‌شود. معیارهای خطا، فاصله پیش‌بینی‌ها از بازده واقعی را نشان می‌دهند و آزمون مقایسه دقت، معناداری آماری اختلاف عملکرد را می‌سنجد. این دو سطح مکمل یکدیگرند، زیرا کاهش عددی خطا ممکن است حاصل نوسانات نمونه‌ای باشد. همین شیوه ارزیابی برای هر دو بازار به‌کار می‌رود.

مقایسه اصلی میان مدل‌های مبتنی بر اطلاعات تاریخی و معیار گام تصادفی انجام می‌شود. اندازه کاهش خطا و معناداری اختلاف زیان، در کنار هم مبنای تفسیر نتایج در چارچوب معیار و مجموعه اطلاعات منتخب قرار می‌گیرند.

معیارهای ارزیابی شامل جذر میانگین مربعات خطا،[6] میانگین قدرمطلق خطا،[7] ضریب تعیین خارج از نمونه[8] و دقت جهت هستند. برای محاسبه ضریب تعیین در هر بخش آزمون، مجموع مربعات خطا بر مجموع مربعات انحراف بازده واقعی از میانگین همان بخش تقسیم و حاصل از یک کم می‌شود. اگر مقدار این ضریب منفی باشد، مدل از پیش‌بینی بر پایه میانگین بازده آن بخش ضعیف‌تر عمل کرده است. میانگین مذکور فقط پس از مشاهده بازده‌های آزمون و برای ارزیابی به‌کار می‌رود و وارد آموزش مدل نمی‌شود. در فایل مکمل، ضریب بهبود نسبت به معیار بازده صفر نیز جداگانه گزارش شده است؛ مخرج این ضریب، مجموع مربعات بازده واقعی است.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1} n(y_t - \hat{y}_t)^2} \quad (۱۳)$$

در روابط ارزیابی، y بازده واقعی، $\hat{y}$ بازده پیش‌بینی‌شده و n تعداد مشاهدات آزمون است. میانگین بازده واقعی همان بخش آزمون با $\bar{y}$ نمایش داده می‌شود.

$$MAE = \frac{1}{n}\sum_{t=1} n|y_t - \hat{y}_t| \quad (۱۴)$$

$$R^2 = 1 - \frac{\sum_{t=1} n(y_t - \hat{y}_t)^2}{\sum_{t=1} n(y_t - \overline{y})^2} \quad (۱۵)$$

$$DA = (100/|S|)\sum_{t} \in S\, \mathbb{1}[sgn(\hat{y}_t) = sgn(y_t)];\ S = \{t : \hat{y}_t \neq 0\} \quad (۱۶)$$

برای سنجش معناداری اختلاف دقت، آزمون دیبولد–ماریانو بر خطاهای پیش‌بینی خارج از نمونه اجرا می‌شود.

در این آزمون، مربع خطای پیش‌بینی به‌عنوان زیان در نظر گرفته می‌شود. اختلاف زیان با کسر مربع خطای مدل دوم از مربع خطای مدل اول به‌دست می‌آید؛ بنابراین، آماره منفی به نفع مدل اول و آماره مثبت به نفع مدل دوم است. برای نتیجه‌گیری درباره معناداری این تفاوت، مقدار احتمال نیز بررسی می‌شود. معنادار نبودن اختلاف فقط به معنای کافی نبودن شواهد برای تفاوت دو مدل است و برابری آن‌ها را اثبات نمی‌کند (Diebold & Mariano, 1995).

$$\delta_t = {e_{1,t}}^2 - {e_{2,t}}^2 \quad (۱۷)$$

$$DM = \frac{\overline{\delta}}{SE_{HAC}(\overline{\delta})} \quad (۱۸)$$

خطای معیار میانگین اختلاف زیان با روشی مقاوم به ناهمسانی واریانس و خودهمبستگی[9] و با وزن‌های بارتلت برآورد شد. حداکثر وقفه از رابطه (۱۹) به‌دست می‌آید. در محاسبه خودکوواریانس، فاصله واقعی روزهای معاملاتی حفظ می‌شود؛ به‌عبارت دیگر، آخرین مشاهده یک بلوک آزمون و نخستین مشاهده بلوک بعد، به‌طور مصنوعی دو مشاهده متوالی در نظر گرفته نمی‌شوند. مقدار احتمال دوطرفه با تقریب نرمال محاسبه می‌شود و مقادیر گزارش‌شده اسمی‌اند؛ یعنی برای انجام چند آزمون به‌طور هم‌زمان تعدیل نشده‌اند. ازاین‌رو، به‌ویژه در مقایسه مدل‌های تودرتو، نتیجه آزمون همراه با اندازه خطا و شواهد آزمون‌های استواری تفسیر می‌شود.

$$L = \left\lfloor 4\left(\frac{n}{100}\right)^{2/9} \right\rfloor \quad (۱۹)$$

---

[6] Root Mean Squared Error (RMSE)
[7] Mean Absolute Error (MAE)
[8] Out-of-sample R-squared
[9] Heteroskedasticity and Autocorrelation Consistent (HAC)

برای توضیح سهم ویژگی‌ها در پیش‌بینی، مقادیر شاپ برای همه مشاهدات آزمون محاسبه شد (Štrumbelj & Kononenko, 2014 ;Lundberg et al., 2020). در هر مرحله، حداکثر ۱۰۰ مشاهده از داده‌های آموزش همان مرحله با بذر ثابت ۴۲ به‌عنوان پس‌زمینه انتخاب شد. برای مدل درختی، توضیح‌گر درخت با تعریف مداخله‌ای برای نحوه لحاظ‌کردن ویژگی‌های کنارگذاشته‌شده به‌کار رفت. در مدل خطی، سهم هر ویژگی از حاصل‌ضرب ضریب مدل در فاصله مقدار استانداردشده آن از میانگین پس‌زمینه محاسبه شد. این تعریف وابستگی میان ویژگی‌ها را در نظر نمی‌گیرد و تفسیر علّی ندارد. برای کنترل صحت محاسبات، جمع سهم ویژگی‌ها و مقدار پایه با پیش‌بینی مدل مقایسه شد و بیشینه اختلاف مطلق کمتر از چهار ضرب‌در ده به توان منفی ده بود.

شاپ نشان می‌دهد که هر ویژگی چه سهمی در خروجی مدل برازش‌شده دارد. این سهم، با اثر ویژگی بر دقت پیش‌بینی یکسان نیست و بزرگی آن نیز رابطه علّی یا امکان کسب سود را اثبات نمی‌کند. به همین دلیل، برای ارزیابی اطلاعات مکمل متغیرهای تقریبی، عملکرد خارج از نمونه مجموعه متعارف و ترکیبی با یکدیگر مقایسه می‌شود.

٤. تحلیل داده‌ها و یافته‌های پژوهش

پیش از اجرای مجدد مدل‌ها، ترتیب و یکتایی تاریخ‌ها، نحوه ساخت بازده و متغیر هدف و امکان بازسازی ویژگی‌ها بررسی شد. در داده‌های تهران، مقدار پایانی شاخص در ۴۱۶ ردیف خارج از بازه کمینه تا بیشینه ثبت‌شده بود و اندازه این اختلاف بین ،واحد شاخص قرار داشت. چون مبنای مستندی برای اصلاح این مقادیر در دست نبود ۲۵/۵ تا ۰/۱ .داده‌های خام بدون تصحیح حدسی حفظ شدند و این ناسازگاری در تفسیر محدودیت‌های پژوهش لحاظ شد

عملکرد خارج از نمونه مدل‌های پیش‌بینی: جدول ۲ عملکرد مدل‌ها را بر اساس میانگین معیارهای پنج بخش آزمون نشان می‌دهد. برای تهیه جدول، هر معیار ابتدا در هر بخش محاسبه شده و سپس میانگین آن با وزن یکسان برای همه بخش‌ها به‌دست آمده است. بنابراین، این اعداد لزوماً با معیار محاسبه‌شده از مجموع خطاهای تمام دوره‌های آزمون برابر نیستند. نتایج تهران برای سه مجموعه متغیر و نتایج اس‌اندپی ۵۰۰ برای مجموعه متعارف گزارش شده‌اند. در هر بازار و برای هر مشخصات، همه مدل‌ها در تاریخ‌های آزمون یکسان ارزیابی شده‌اند.

***جدول ۲** .میانگین عملکرد خارج از نمونه مدل‌های پیش‌بینی*

Table 2. Average Out-of-Sample Performance of Forecasting Models

بازار / متغیرها	مدل	RMSE	MAE	R^2	جهت (٪)
اس‌اندپی ۵۰۰ / متعارف	گام تصادفی	۰/۰۱۰۳۱۶	۰/۰۰۷۴۸۸	−۰/۰۰۶۷	—
اس‌اندپی ۵۰۰ / متعارف	رگرسیون خطی	۰/۰۱۰۴۲۸	۰/۰۰۷۶۲۴	−۰/۰۳۶۰	۵۱/۱۳
اس‌اندپی ۵۰۰ / متعارف	مدل درختی	۰/۰۱۰۲۷۱	۰/۰۰۷۴۶۳	۰/۰۰۲۶	۵۲/۷۸
تهران / متعارف	گام تصادفی	۰/۰۱۱۳۰۰	۰/۰۰۸۲۸۷	−۰/۰۲۸۷	—
تهران / متعارف	رگرسیون خطی	۰/۰۱۰۵۰۴	۰/۰۰۷۶۶۵	۰/۱۰۲۰	۵۹/۷۱
تهران / متعارف	مدل درختی	۰/۰۱۰۷۰۹	۰/۰۰۷۸۲۷	۰/۰۷۱۷	۵۹/۴۱
تهران / تقریبی	گام تصادفی	۰/۰۱۱۳۰۰	۰/۰۰۸۲۸۷	−۰/۰۲۸۷	—
تهران / تقریبی	رگرسیون خطی	۰/۰۱۱۵۱۲	۰/۰۰۸۶۹۴	−۰/۰۶۰۶	۵۱/۳۲
تهران / تقریبی	مدل درختی	۰/۰۱۱۱۲۳	۰/۰۰۸۲۴۰	۰/۰۰۴۵	۵۱/۹۱
تهران / ترکیبی	گام تصادفی	۰/۰۱۱۳۰۰	۰/۰۰۸۲۸۷	−۰/۰۲۸۷	—

بازار / متغیرها	مدل	RMSE	MAE	R^2	جهت (٪)
تهران / ترکیبی	رگرسیون خطی	۰/۰۱۰۶۴۹	۰/۰۰۷۹۵۳	۰/۰۸۰۱	۶۰/۵۹
تهران / ترکیبی	مدل درختی	۰/۰۱۰۶۵۰	۰/۰۰۷۸۴۵	۰/۰۸۴۴	۵۷/۷۹

منبع: یافته‌های پژوهش؛ یادداشت: اعداد ستون‌های عملکرد، میانگین پنج مرحله ارزیابی هستند. ضریب تعیین هر مرحله با استفاده از میانگین بازده همان بخش آزمون محاسبه شده است. دقت جهت فقط برای پیش‌بینی‌های غیرصفر محاسبه می‌شود؛ به همین دلیل، برای گام تصادفی که بازده صفر پیش‌بینی می‌کند، این معیار گزارش نشده است.

در اس‌اندپی ۵۰۰، میانگین جذر میانگین مربعات خطای گام تصادفی، مدل خطی و مدل درختی به‌ترتیب ۰/۰۱۰۳۱۶، ۰/۰۱۰۴۲۸ و ۰/۰۱۰۲۷۱ است. بهبود عددی مدل درختی نسبت به معیار مرجع محدود است و باید در کنار آزمون اختلاف زیان تفسیر شود.

در تهران و با استفاده از مجموعه متعارف، میانگین جذر میانگین مربعات خطا برای مدل خطی و برای ۰/۰۱۰۵۰۴ کمترند. با استفاده از ۰/۰۱۱۳۰۰ است. هر دو مقدار از خطای گام تصادفی، برابر با ۰/۰۱۰۷۰۹ مدل درختی می‌رسد. این مقایسه نشان ۰/۰۱۰۶۵۰ و ۰/۰۱۰۶۴۹ مجموعه ترکیبی، خطای مدل خطی و درختی به‌ترتیب به می‌دهد که افزودن متغیرهای تقریبی، خطای مدل خطی را کاهش نداده و برای مدل درختی تنها کاهش محدودی ایجاد کرده است. در نتیجه، صرف این تغییر عددی برای تأیید مزیت پایدار مجموعه ترکیبی کافی نیست.

نقش متغیرهای تقریبی شرایط معاملاتی بورس تهران: مقایسه مجموعه‌های متغیر نشان می‌دهد که ارزش افزوده نشانگرهای تقریبی به مدل و معیار ارزیابی وابسته است. این نشانگرها به‌تنهایی برتری معناداری نسبت به گام تصادفی ندارند. در مجموعه ترکیبی، دقت جهت مدل خطی و مدل درختی به‌ترتیب به حدود ۶۰/۵۹ و ۵۷/۷۹ درصد می‌رسد؛ اما همان‌طور که در جدول ۲ دیده می‌شود، افزودن آن‌ها خطای مدل خطی را کاهش نمی‌دهد و کاهش خطای مدل درختی نیز محدود است. بنابراین، رتبه بالای این متغیرها در تحلیل شاپ برای تأیید ارزش پیش‌بینی آن‌ها کافی نیست.

معناداری آماری تفاوت عملکرد مدل‌ها: برای بررسی اینکه آیا تفاوت‌های مشاهده‌شده در جدول ۲ از نظر آماری قابل‌اتکاست یا صرفاً ناشی از نوسانات نمونه‌ای است، آزمون دیبولد–ماریانو[10] بر خطاهای خارج از نمونه تجمیع‌شده اجرا شد؛ نتایج در جدول ۳ گزارش شده است.

[10] Diebold–Mariano test

جدول ۳ *.نتایج تجمیعی آزمون دیبولد–ماریانو*

Table 3. Aggregate Results of the Diebold–Mariano Test

بازار / متغیرها	مقایسه مدل اول / دوم	آماره مقاوم	احتمال	نتیجه ۵٪
اس‌اندپی ۵۰۰ / متعارف	رگرسیون خطی / گام تصادفی	۰/۶۶۰	۰/۵۰۹۰	معنادار نیست
اس‌اندپی ۵۰۰ / متعارف	مدل درختی / گام تصادفی	−۰/۵۶۴	۰/۵۷۲۸	معنادار نیست
اس‌اندپی ۵۰۰ / متعارف	مدل درختی / رگرسیون خطی	−۱/۲۵۸	۰/۲۰۸۵	معنادار نیست
تهران / متعارف	رگرسیون خطی / گام تصادفی	−۳/۷۲۶	۰/۰۰۰۲	به نفع رگرسیون خطی
تهران / متعارف	مدل درختی / گام تصادفی	−۳/۱۹۶	۰/۰۰۱۴	به نفع مدل درختی
تهران / متعارف	مدل درختی / رگرسیون خطی	۱/۸۴۷	۰/۰۶۴۷	معنادار نیست
تهران / تقریبی	رگرسیون خطی / گام تصادفی	۰/۸۲۹	۰/۴۰۶۸	معنادار نیست
تهران / تقریبی	مدل درختی / گام تصادفی	−۰/۸۹۴	۰/۳۷۱۳	معنادار نیست
تهران / تقریبی	مدل درختی / رگرسیون خطی	−۲/۱۲۲	۰/۰۳۳۸	به نفع مدل درختی
تهران / ترکیبی	رگرسیون خطی / گام تصادفی	−۲/۸۸۸	۰/۰۰۳۹	به نفع رگرسیون خطی
تهران / ترکیبی	مدل درختی / گام تصادفی	−۳/۱۵۰	۰/۰۰۱۶	به نفع مدل درختی
تهران / ترکیبی	مدل درختی / رگرسیون خطی	۰/۱۳۲	۰/۸۹۴۶	معنادار نیست

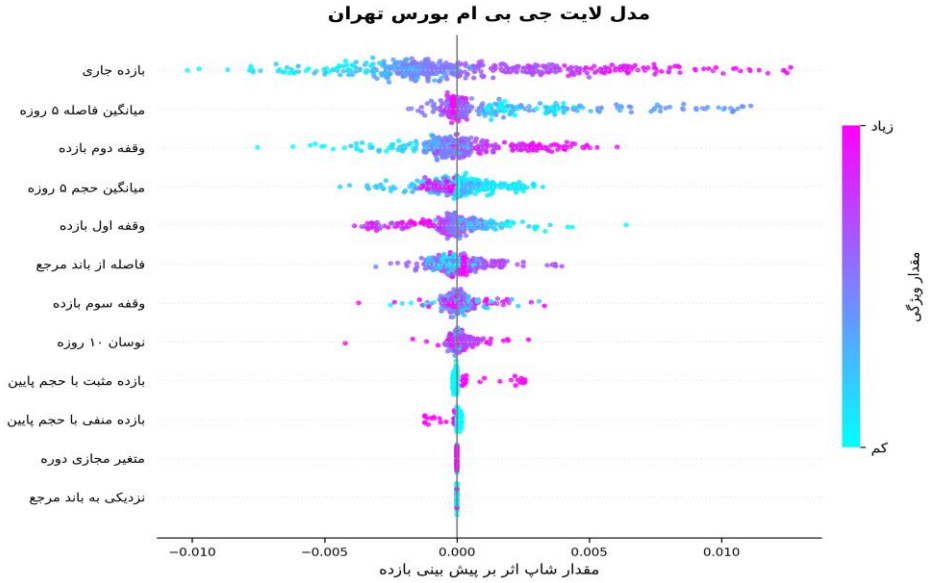
منبع: یافته‌های پژوهش.

در اس‌اندپی ۵۰۰، مقایسه مستقیم مدل خطی با گام تصادفی معنادار نبود و مقدار احتمال ۰/۵۰۹۰ به‌دست آمد. تفاوت مدل درختی با گام تصادفی و مدل خطی نیز معنادار نبود. مقدار احتمال این دو مقایسه به‌ترتیب ۰/۵۷۲۸ و ۰/۲۰۸۵ بود.

در تهران، مدل خطی در مجموعه متعارف بهتر از گام تصادفی عمل کرد. آماره آزمون منفی ۳/۷۲۶ و مقدار احتمال ۰/۰۰۰۲ بود. مدل درختی نیز نسبت به گام تصادفی بهبود معناداری نشان داد و مقدار احتمال این مقایسه ۰/۰۰۱۴ بود. بااین‌حال، اختلاف مدل درختی و خطی در سطح پنج درصد معنادار نبود و مقدار احتمال ۰/۰۶۴۷ به‌دست آمد. در مجموعه ترکیبی، هر دو مدل از معیار مرجع بهتر بودند، اما اختلاف عملکرد آن‌ها معنادار نبود و مقدار احتمال ۰/۸۹۴۶ بود. در مجموعه تقریبی، مدل درختی از مدل خطی بهتر عمل کرد، ولی هیچ‌یک بر گام تصادفی برتری معناداری نداشتند. بنابراین، بهتر بودن یک مدل نسبت به مدلی ضعیف‌تر، به‌تنهایی شواهد کافی برای پیش‌بینی‌پذیری نسبت به معیار مرجع فراهم نمی‌کند.

در چارچوب فرضیه بازار کارا، بهبود دقت پیش‌بینی تهران در مشخصات متعارف و ترکیبی، محتوای پیش‌بینانه اطلاعات تاریخی را نسبت به معیار بازده صفر نشان می‌دهد (Fama, 1970). این نتیجه به‌تنهایی شکل ضعیف کارایی را رد نمی‌کند؛ زیرا تغییرات بازده مورد انتظار و ریسک می‌توانند با پیش‌بینی‌پذیری سازگار باشند و سودآوری پس از هزینه معاملات در این پژوهش ارزیابی نشده است. برای اس‌اندپی ۵۰۰ نیز نبود بهبود معنادار نسبت به معیار منتخب، صرفاً نبود شواهد کافی در این طراحی را نشان می‌دهد و اثبات کارایی محسوب نمی‌شود.

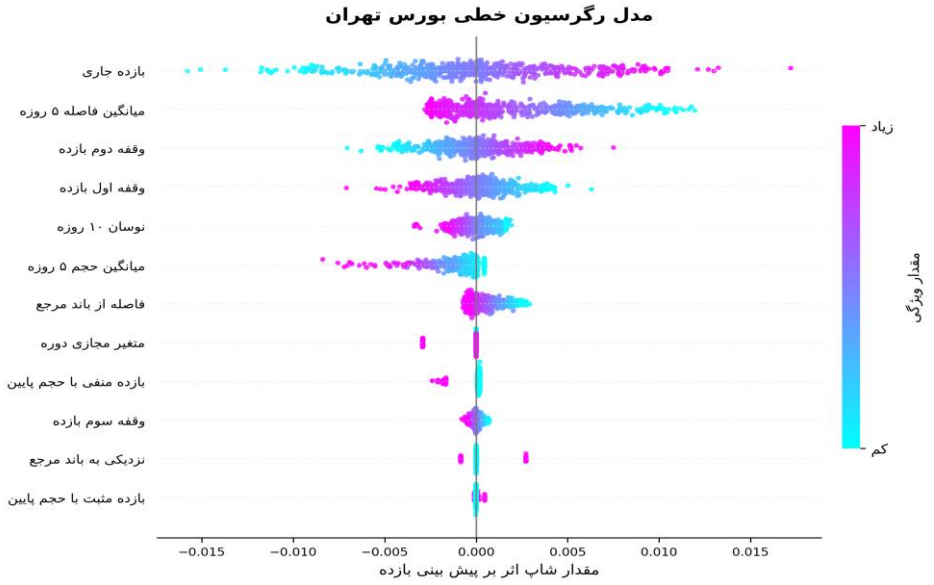
تفسیر مدل با روش شاپ[11]: شکل‌های ۲ و ۳ توزیع سهم‌های شاپ را به‌ترتیب برای مدل درختی و مدل خطی تهران در پنج بخش آزمون نشان می‌دهند. هر نقطه یک مشاهده آزمون است؛ رنگ، مقدار نسبی ویژگی و محور افقی، جهت و اندازه سهم آن در پیش‌بینی بازده را نشان می‌دهد.



شکل ۲ *.نمودار خلاصه شاپ مدل درختی تهران؛ مجموعه ترکیبی*

Figure 2. Out-of-sample SHAP summary: Tehran LightGBM, combined features.

منبع: یافته‌های پژوهش.



شکل ۳ *.نمودار خلاصه شاپ مدل خطی تهران؛ مجموعه ترکیبی*

Figure 3. Out-of-sample SHAP summary: Tehran linear regression, combined features.

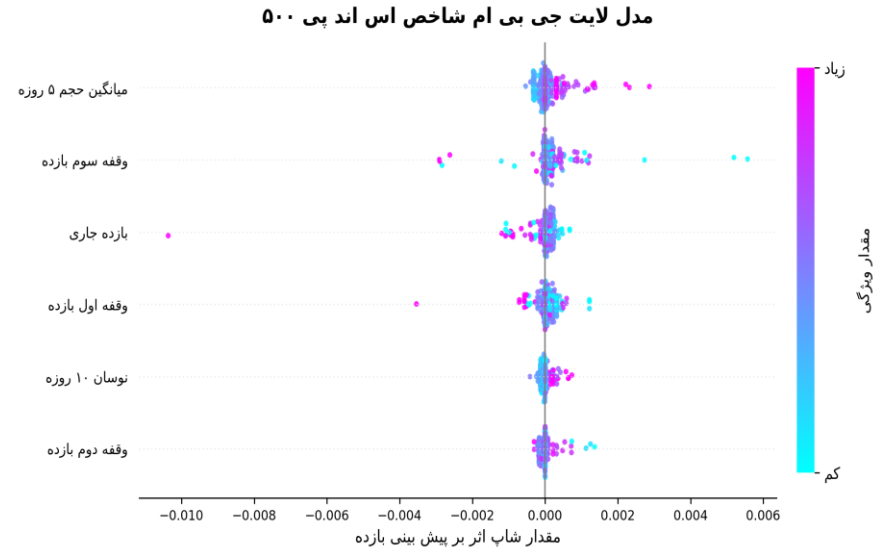
منبع: یافته‌های پژوهش.

نقاط نمودار از خروجی مدل برازش‌شده در هر یک از مراحل آزمون گردآوری شده‌اند. در هر مرحله، مدل دوباره برآورد می‌شود و مقدار پایه آن نیز ممکن است تغییر کند؛ ازاین‌رو، نمودار سهم ویژگی‌ها را در چند مدل متوالی

[11]SHapley Additive exPlanations (SHAP)

خلاصه می‌کند. اهمیت هر ویژگی نیز به تعریف پس‌زمینه و مشخصات مدل بستگی دارد و به‌تنهایی رابطه علّی یا مزیت آن ویژگی در پیش‌بینی را نشان نمی‌دهد.

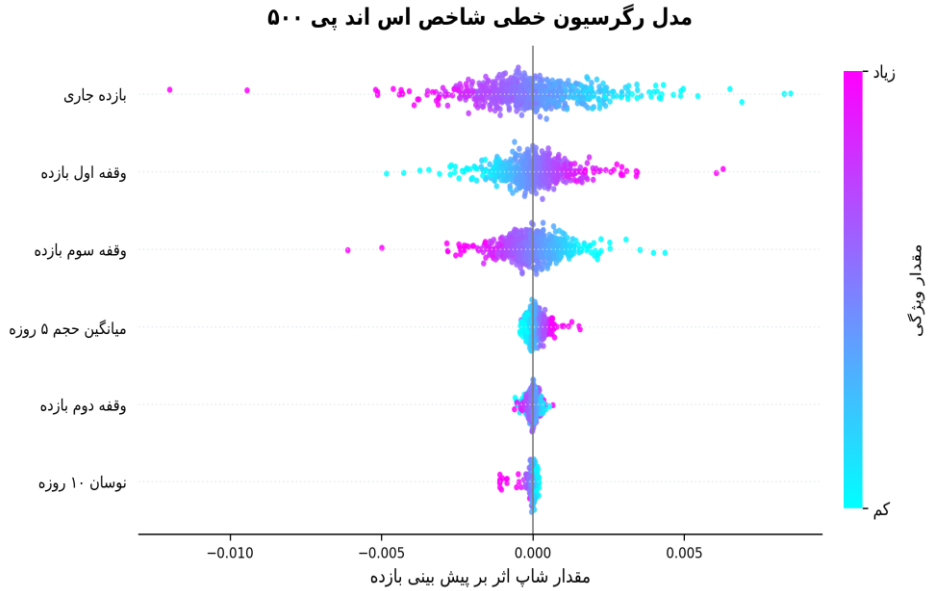
شکل‌های ۴ و ۵ نتایج متناظر اس‌اندپی ۵۰۰ را ارائه می‌کنند. برای مقایسه این نمودارها باید به تفاوت مقیاس سهم‌ها میان بازارها و مدل‌ها توجه کرد؛ مقایسه مستقیم بزرگی سهم‌ها، آزمونی برای تفاوت کارایی بازارها نیست. همچنین، چون گام تصادفی در این پژوهش همواره بازده صفر پیش‌بینی می‌کند، سهم همه ویژگی‌ها در آن صفر است و رسم نمودار جداگانه، اطلاعاتی به تحلیل اضافه نمی‌کند.



شکل ۴. *نمودار خلاصه شاپ مدل درختی اس‌اندپی ۵۰۰؛ مجموعه متعارف*

Figure 4. Out-of-sample SHAP summary: S&P 500 LightGBM, classical features.

منبع: یافته‌های پژوهش.



شکل ۵. *نمودار خلاصه شاپ مدل خطی اس‌اندپی ۵۰۰؛ مجموعه متعارف*

Figure 5. Out-of-sample SHAP summary: S&P 500 linear regression, classical features.

منبع: یافته‌های پژوهش.

در مجموع، نمودارهای شاپ الگوی استفاده مدل‌ها از ویژگی‌ها را توصیف می‌کنند. استنباط درباره دقت پیش‌بینی همچنان بر نتایج خارج از نمونه و آزمون اختلاف زیان استوار است.

تحلیل حساسیت و آزمون‌های استواری: برای سنجش حساسیت نتایج به تعریف کاهش نسبی حجم، آستانه‌های ۰/۶۰، ۰/۷۰ و ۰/۸۰ با تاریخ‌های آزمون یکسان بررسی شدند. تغییر مقدار مبنا به دو آستانه جایگزین، نشانگرها را به‌ترتیب در ۱۶۵ و ۲۹۷ ردیف تغییر داد؛ ازاین‌رو، ویژگی‌ها برای هر آستانه دوباره محاسبه شدند. در مجموعه ترکیبی تهران، جذر میانگین مربعات خطای تجمیعی مدل خطی در این سه حالت به‌ترتیب ۰/۰۱۰۸۲۲، ۰/۰۱۰۸۳۱ و ۰/۰۱۰۸۲۶ بود. مقادیر متناظر مدل درختی نیز ۰/۰۱۰۸۹۷، ۰/۰۱۰۸۵۳ و ۰/۰۱۰۸۳۰ به‌دست آمد. در هر سه آستانه، هر دو مدل از گام تصادفی بهتر بودند و اختلاف آن‌ها معنادار نبود. این نتیجه فقط برای آستانه‌های بررسی‌شده معتبر است و نباید به همه تعریف‌های ممکن تعمیم داده شود.

برای سنجش وابستگی نتایج به مرز زمانی، زیرنمونه‌ای شامل بازده‌های هدف پیش از ۱۲ اکتبر ۲۰۲۴ انتخاب و با همان ساختار زمانی دوباره ارزیابی شد. در مجموعه ترکیبی، خطای تجمیعی گام تصادفی، مدل خطی و مدل درختی به‌ترتیب ۰/۰۱۳۹۲۴، ۰/۰۱۲۶۷۷ و ۰/۰۱۳۱۶۱ بود. هر دو مدل از معیار مرجع دقیق‌تر بودند. مدل خطی نیز از مدل درختی بهتر عمل کرد و مقدار احتمال این مقایسه ۰/۰۰۲۵ بود. این نتیجه نشان می‌دهد شواهد پیش‌بینی‌پذیری تهران به دوره پس از مرز زمانی محدود نیست.

با افزایش سهم اولیه آموزش به ۶۰ و ۷۰ درصد، هر دو مدل تنظیم‌شده در مجموعه‌های متعارف و ترکیبی تهران همچنان از گام تصادفی بهتر بودند و اختلاف مدل خطی و درختی در نمونه کامل معنادار نشد. این تغییر، تاریخ‌های آزمون را نیز جابه‌جا می‌کند؛ پس نتیجه هم به نحوه تقسیم داده‌ها و هم به دوره ارزیابی وابسته است. در اس‌اندپی ۵۰۰، مدل درختی برتری نداشت. در تقسیم ۷۰ درصدی، مدل خطی نیز از گام تصادفی ضعیف‌تر بود و مقدار احتمال اسمی ۰/۰۴۸۳ به‌دست آمد.

برای سنجش حساسیت عملکرد به ابرپارامترها، مدل درختی با تنظیمات آماری پیش‌فرض و ۱۰۰ درخت اجرا شد؛ در این اجرا جست‌وجوی ابرپارامتر و توقف زودهنگام به‌کار نرفت. بذر تصادفی و تعداد پردازشگرها نیز ثابت نگه داشته شدند. در مجموعه ترکیبی تهران، خطای تجمیعی مدل ۰/۰۱۲۵۲۰ بود و از خطای ۰/۰۱۱۵۲۱ گام تصادفی فراتر رفت. این تفاوت به نفع گام تصادفی معنادار بود و مقدار احتمال ۰/۰۰۳۶ به‌دست آمد. در اس‌اندپی ۵۰۰ نیز مدل پیش‌فرض از گام تصادفی ضعیف‌تر عمل کرد. این یافته حساسیت عملکرد لایت‌جی‌بی‌ام به شیوه تنظیم را نشان می‌دهد و شاهد استواری مدل نیست. چون چند تنظیم هم‌زمان تغییر کرده‌اند، افت عملکرد را نمی‌توان به یک ابرپارامتر خاص نسبت داد. جزئیات معیارها، آزمون‌های اختلاف زیان، تقسیم‌های زمانی و تنظیمات منتخب در فایل اکسل مکمل آمده است.

۵. بحث و نتیجه‌گیری

یافته‌ها باید با تمایز میان پیش‌بینی‌پذیری آماری و کارایی اقتصادی تفسیر شوند. معیار بازده صفر، مبنای مشخصی برای مقایسه دقت فراهم می‌کند، اما کارایی بازار الزاماً بازده مورد انتظار صفر را ایجاب نمی‌کند. همچنین، این پژوهش سودآوری پس از تعدیل ریسک و کسر هزینه معاملات را ارزیابی نکرده است.

در اس‌اندپی ۵۰۰، هیچ‌یک از مدل‌ها در مشخصات مبنا برتری معناداری نسبت به گام تصادفی نداشتند. در تهران، رگرسیون خطی با مجموعه متعارف و ترکیبی خطای کمتری داشت و این نتیجه در تغییرات بررسی‌شده آستانه حجم، تقسیم زمانی و زیرنمونه پیش از مرز زمانی نیز مشاهده شد. بنابراین، شواهد از محتوای پیش‌بینانه اطلاعات تاریخی تهران در مشخصات منتخب حمایت می‌کنند؛ نبود شواهد مشابه برای اس‌اندپی ۵۰۰ نیز به همین نمونه و طراحی محدود است.

در نمونه کامل تهران، اختلاف مدل خطی و درختی برای مجموعه‌های متعارف و ترکیبی در سطح پنج درصد معنادار نبود. مدل خطی در زیرنمونه پیش از مرز زمانی بهتر عمل کرد، درحالی‌که لایت‌جی‌بی‌ام با تنظیمات

پیش‌فرض افت کرد. این نتایج شواهدی از مزیت پایدار پیچیدگی غیرخطی فراهم نمی‌کنند و نشان می‌دهند عملکرد مدل درختی به شیوه تنظیم آن وابسته است.

متغیرهای تقریبی به‌تنهایی بر معیار مرجع برتری معناداری نداشتند. افزودن آن‌ها به مجموعه متعارف نیز خطای مدل خطی را کاهش نداد و در مدل درختی فقط بهبود محدودی ایجاد کرد. چون باند مرجع فرضی است، این نتایج برای توضیح سازوکارهای ریزساختاری بازار کافی نیستند.

مقایسه دو بازار توصیفی است و آزمون مستقیمی برای سنجش تفاوت درجه کارایی آن‌ها انجام نشده است. بنابراین، اختلاف نتایج را نمی‌توان به ساختار بازار یا مقررات نسبت داد. برتری مدل غیرخطی در مطالعاتی با هدف، افق یا داده‌های متفاوت نیز برای انتظار نتیجه مشابه در این پژوهش کافی نیست.

از نظر کاربردی، انتخاب مدل باید بر مقایسه با معیارهای رقابتی در داده‌های دیده‌نشده و بررسی حساسیت به تنظیمات تکیه داشته باشد. کاهش خطای آماری تا زمانی که هزینه معاملات، نقدشوندگی و امکان اجرا بررسی نشده باشد، مبنای توصیه معاملاتی نیست.

محدودیت‌های اصلی شامل باند مرجع فرضی، افق یک‌روزه، وابستگی نتایج به ویژگی‌ها و تقسیم زمانی، نبود ارزیابی سودآوری خالص و ماهیت توصیفی مقایسه دو بازار است. داده‌های تجمیعی ممکن است بخشی از اطلاعات ریزساختاری را ثبت نکنند، اما جهت و اندازه این خطای اندازه‌گیری از داده‌های حاضر مشخص نیست. همچنین، ناسازگاری محدود قیمت پایانی با بیشینه و کمینه تهران و ابهام در تعریف و واحد حجم باید با خروجی اولیه تأمین‌کننده تطبیق داده شود؛ در توضیحات مکمل، حجم اس‌اندپی ۵۰۰ به صندوق قابل معامله SPY نسبت داده شده است.

پژوهش‌های بعدی می‌توانند افق‌های زمانی دیگر، اطلاعات سهام تشکیل‌دهنده شاخص، معیارهای جایگزین بازده مورد انتظار و سودآوری پس از کسر هزینه معاملات را بررسی کنند.

تعارض منافع: برای انجام و نگارش این پژوهش، هیچ‌گونه کمک مالی از فرد، نهاد یا سازمانی دریافت نشده است. نویسندگان اعلام می‌کنند هیچ‌گونه تعارض منافعی ندارند.

References

Alizadeh Chamazkoti, M., Fathabadi, M., Ghavidel Doostkouei, S., & Mahmoodzadeh, M. (2024). Tehran stock market efficiency: A quantile autoregression approach. *Financial Statistical Journal, 7*(1), Article 7543. https://doi.org/10.24294/fsj.v7i1.7543

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Eghtesad, A., & Mohammadi, E. (2023). Portfolio optimization with return prediction using LSTM, random forest, and ARIMA. *Financial Management Perspective, 13*(43), 9–28. https://doi.org/10.48308/jfmp.2024.104191 (In Persian).

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383–417. https://doi.org/10.2307/2325486

Gholami, N., & Shams Gharneh, N. (2024). Presenting an optimized CNN-LSTM model for stock price forecasting in the Tehran Stock Exchange. *Financial Management Perspective, 14*(45), 123–147. https://doi.org/10.48308/jfmp.2024.104892 (In Persian).

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Harvey, C. R. (1995). Predictable risk and returns in emerging markets. *The Review of Financial Studies, 8*(3), 773–816. https://doi.org/10.1093/rfs/8.3.773

Hasannezhad, M., Davallou, M., & Shabani, F. (2023). Investigation of the effects of price limit changes on the intraday volatility of Iran's stock market using realized variance and discrete Fourier transform. *Asset Management and Financing, 11*(2), 19–34. https://doi.org/10.22108/amf.2023.134778.1755 (In Persian).

Heydari Delooei, A., Vahdati, M., Mohebbi, H., & Bagherpour, N. (2026). Predictability of the Tehran Stock Exchange total index using a hybrid machine learning approach: Analysis of market efficiency and influential variables. *Financial Research Journal, 28*(2), 464–493. https://doi.org/10.22059/frj.2025.394747.1007738 (In Persian; translated title).

Jafaripour, M., Ramezan Ahmadi, M., Mazaheri, E., & Arman, S. A. (2021). Adjusted capital asset pricing models (CAPMs) with respect to the magnet effect factor (MEF) caused by the range of stock price fluctuations. *Journal of Asset Management and Financing, 9*(3), 65–88. https://doi.org/10.22108/amf.2022.130725.1699 (In Persian).

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies, 1*(1), 41–66. https://doi.org/10.1093/rfs/1.1.41

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., & Shahab, S. (2020). Deep learning for stock market prediction. *Entropy, 22*(8), Article 840. https://doi.org/10.3390/e22080840

Namazi, M. (2003). *An investigation of the economic performance of the stock exchange market in Iran*. Deputy for Economic Affairs, Ministry of Economic Affairs and Finance. (In Persian).

Namazi, M., & Shooshtarian, Z. (1996). An investigation and analysis of the efficiency of the Tehran Stock Exchange market. *Financial Research Journal, 2*(7–8), 82–104. (In Persian).

Oujimehr, S., Montakhab, A., & Samadi, A. H. (2020). Ranking the efficiency of selected active industries in the Tehran Stock Exchange market: An application of the multifractal detrended fluctuation analysis method. *Industrial Economics Research, 4*(14), 11–26. (In Persian).

Raei, R., Vahdati, M., Mohebbi, H., & Heydari Delooei, A. (2025). Interpreting forecast the return of the price index of manufacturing industries in the Tehran Stock Exchange using explainable ensemble learning. *Financial Management Perspective, 14*(48), 55–78. https://doi.org/10.48308/jfmp.2025.238860.1475 (In Persian).

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems, 41*(3), 647–665. https://doi.org/10.1007/s10115-013-0679-x

Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). John Wiley & Sons.