***Publisher: Shahid Beheshti University***
***Journal of Financial Management Perspective***
***2025, Vol. 14, No.48***
PP: 55-78
https://doi.org/10.48308/jfmp.2025.238860.1475
P-ISSN: 2645-4637
E-ISSN: 2645-4645


# Interpreting Forecast the Return of the Price Index of Manufacturing Industries in the Tehran Stock Exchange Using Explainable Ensemble Learning

**Reza Raei[*], Masoud Vahdati[**], Hossein Mohebbi[***], Amirhossein Heydari Delooei[****]**

**Abstract:**

**Purpose**: In recent years, machine learning has gained significant attention as an effective tool for forecasting financial time series. However, many of these models function as black boxes, and their lack of transparency has led to reduced trust in their predictions. To address this limitation, the use of explainable artificial intelligence (XAI) models-capable of providing detailed insights into the prediction mechanisms-has become essential. Accordingly, the aim of this study is to develop and evaluate an AI-based forecasting model that not only delivers high accuracy but also offers strong interpretability. In this context, the contribution and role of input variables in the model's predictions are explicitly identified, and the stability of the results in terms of both accuracy and explainability is assessed using cross-validation techniques, particularly time series splitting.

**Method:** This applied research adopts a descriptive-analytical method with a quantitative forecasting approach. For the first time in Iran, it investigates the explainability of optimized artificial intelligence models in forecasting the return of the price index for eight manufacturing industries listed on the Tehran Stock Exchange. The dataset, covering the period from 2018 to 2023, was collected from the BourseView database. The Random Forest algorithm, as an ensemble learning method, was trained using a combination of technical, fundamental, and macroeconomic variables as input features. A Genetic Algorithm was utilized to optimize the model's hyperparameters. To enhance transparency and model credibility, the SHAP technique was employed to analyze the influence and importance of each feature in the prediction process.

[*] Professor, Department of Finance, College of Management, University of Tehran, Tehran, Iran.
[**] MSc of Finance, College of Management, University of Tehran, Tehran, Iran.
[***] Assistant Professor, Industrial Management Department, Meybod University, Meybod, Iran. (Corresponding Author), Email: h.mohebbi@meybod.ac.ir
[****] MSc Student of Algorithms and Computations, College of Engineering, University of Tehran, Tehran, Iran.

**Findings:** The results demonstrate that combining the Random Forest algorithm with Genetic Algorithm-based hyperparameter optimization and incorporating explainability techniques such as SHAP values not only improves the prediction accuracy of the price index returns for Tehran's manufacturing industries but also enhances model transparency and reliability. The findings highlight those technical indicators-particularly the Exponential Moving Average (EMA), MACD index, trading volume, and free float shares-play the most significant role in enhancing predictive accuracy. In contrast, fundamental variables such as the price-to-earnings ratio and interest rates are influential but less impactful compared to technical indicators. Furthermore, time series cross-validation confirms the robustness and generalizability of the proposed model across different time periods.

**Conclusion:** In line with reputable international studies, the results suggest that explainable AI models not only outperform traditional models in predictive tasks but also assist financial analysts in making informed and effective decisions. These models can play a pivotal role in risk management and portfolio optimization. Therefore, the proposed model-featuring operational transparency and high reliability-is introduced as an effective tool for financial analysts, opening new horizons for the application of explainable AI in Iran's financial sector.

**Keywords: Explainable Artificial Intelligence, Random Forest, Genetic Algorithm, Cross-Validation, Tehran Stock Exchange.**

**ناشر: دانشگاه شهیدبهشتی**
**نشریه چشم انداز مدیریت مالی**
۱۴۰۳ دوره ۱۴، شماره ۴۸ صص ۷۸-۵۵

شاپای چاپی: ۴۶۳۷-۲۶۴۵
شاپای الکترونیکی: ۴۶۴۵-۲۶۴۵



# تفسیرِ پیش‌بینی بازده شاخص قیمت صنایع تولیدی بورس اوراق بهادار تهران با استفاده از یادگیری تجمیعی توضیح‌پذیر

**رضا راعی[*]، مسعود وحدتی[**]، حسین محبی[***]، امیرحسین حیدری دلوئی [****]**

## چکیده

**هدف:** امروزه، یادگیری ماشین به عنوان ابزاری کارآمد در پیش‌بینی سری‌های زمانی مالی مورد توجه قرار گرفته است. با این حال، اغلب این مدل‌ها به عنوان مدل‌های جعبه سیاه به دلیل عدم شفافیت، موجب کاهش اعتماد به نتایج پیش‌بینی شده‌اند. برای رفع این محدودیت، بهره‌گیری از مدل‌های هوش مصنوعی توضیح‌پذیر که امکان تحلیل دقیق ساز و کار پیش‌بینی را فراهم می‌آورند، ضروری است. بر این اساس، هدف این پژوهش، توسعه و ارزیابی یک مدل پیش‌بینی مبتنی بر هوش مصنوعی است که علاوه بر دقت بالا، از قابلیت توضیح‌پذیری نیز برخوردار باشد. در این راستا، نقش و سهم متغیرهای ورودی در پیش‌بینی‌های مدل به طور شفاف مشخص شده و پایداری نتایج آن از نظر دقت و قابلیت توضیح‌پذیری، با استفاده از روش‌های اعتبارسنجی متقاطع، به‌ویژه بخش‌بندی سری‌های زمانی، مورد ارزیابی قرار می‌گیرد.

**روش:** این پژوهش از نظر هدف، کاربردی و از نظر روش، توصیفی- تحلیلی با رویکرد پیش‌بینی کمّی است که برای نخستین بار در ایران به بررسی قابلیت توضیح‌پذیری هوش مصنوعی بهینه شده در پیش‌بینی بازده شاخص قیمت هشت صنعت تولیدی بورس اوراق بهادار تهران می‌پردازد. داده‌های پژوهش شامل شاخص‌های صنایع در بازه زمانی ۱۳۹۷ تا ۱۴۰۲ است که از پایگاه‌های اطلاعاتی (بورس ویو) جمع‌آوری شده‌اند. برای آموزش مدل جنگل تصادفی به عنوان یک مدل یادگیری تجمیعی، متغیرهای تکنیکال، بنیادی و کلان اقتصادی به عنوان ویژگی‌های مدل، مورد بررسی قرار گرفته‌اند. همچنین الگوریتم ژنتیک به منظور بهینه‌سازی هایپرپارامترهای این مدل به کار گرفته شده است. به منظور افزایش شفافیت و اعتمادپذیری مدل، از تکنیک تفسیرپذیری شاپ برای شناخت تأثیر و اهمیت ویژگی‌ها استفاده شده است.

**یافته‌ها:** نتایج این پژوهش نشان می‌دهد که ترکیب الگوریتم جنگل تصادفی با بهینه‌سازی هایپرپارامترها از طریق الگوریتم ژنتیک و استفاده از روش توضیح‌پذیری همچون مقادیر شاپ، علاوه بر افزایش دقت پیش‌بینی بازده شاخص قیمت صنایع تولیدی بورس تهران، شفافیت و اعتمادپذیری مدل را نیز ارتقا می‌دهد. یافته‌ها تأکید دارند که متغیرهای تکنیکال، به ویژه شاخص میانگین متحرک نمایی، شاخص همگرایی و واگرایی میانگین متحرک، حجم معاملات و میزان سهام شناور، بیشترین

[*] استاد، گروه مالی، دانشکدگان مدیریت، دانشگاه تهران، تهران، ایران
[**] کارشناسی ارشد مالی، دانشکدگان مدیریت، دانشگاه تهران، تهران، ایران.
[***] استادیار گروه مدیریت صنعتی، دانشکده علوم انسانی، دانشگاه میبد، میبد، ایران. (نویسندهٔ مسئول).
E-Mail: h.mohebbi@meybod.ac.ir
[****] دانشجوی کارشناسی ارشد الگوریتم‌ها و محاسبات، دانشکدگان فنی، دانشگاه تهران، تهران، ایران.

نقش را در بهبود دقت پیش‌بینی ایفا می‌کنند. در مقابل، متغیرهای بنیادی همچون نسبت قیمت به درآمد و نرخ بهره و نرخ تورم تأثیرگذارند، اما نقش آن‌ها نسبت به متغیرهای تکنیکال کمتر است. علاوه بر این، ارزیابی متقاطع سری زمانی، پایداری و تعمیم‌پذیری بالای مدل پیشنهادی را در دوره‌های مختلف تأیید می‌کند.

**نتیجه‌گیری:** با توجه به هم‌خوانی نتایج این پژوهش با مطالعات معتبر بین‌المللی می‌توان نتیجه گرفت که مدل‌های هوش‌مصنوعی توضیح‌پذیر نه‌تنها عملکرد خوبی نسبت به مدل‌های سنتی دارند، بلکه تحلیلگران مالی را در اتخاذ تصمیمات آگاهانه و مؤثر یاری کرده و می‌توانند نقشی کلیدی در مدیریت ریسک و بهینه‌سازی سبد دارایی‌ها ایفا کنند. بدین ترتیب، مدل پیشنهادی با شفافیت عملکرد و قابلیت اطمینان بالا، به عنوان ابزاری مؤثر برای تحلیلگران مالی معرفی شده و افق‌های تازه‌ای را در کاربرد هوش مصنوعی توضیح‌پذیر در صنعت مالی ایران می‌گشاید.

**کلیدواژه‌ها: هوش م صنوعی تو ضیح‌پذیر؛ جنگل ت صادفی؛ الگوریتم ژنتیک؛ ارزیابی متقاطع؛ بورس اوراق بهادار تهران.**

## ۱. مقدمه

پژوهش‌های مرتبط با پیش‌بینی بازارهای سهام، همواره جایگاهی کلیدی در حوزه اقتصاد مالی داشته‌اند [۱۰]. مطالعات تجربی متعدد نشان داده‌اند که اگرچه بازارهای مالی به طور کامل پیش‌بینی‌پذیر نیستند، اما می‌توانند الگوهای جزئی و قابل‌پیش‌بینی را از خود نشان دهند [۲۸]. در این زمینه، روش‌های اقتصادسنجی و آماری به‌عنوان ابزارهای اصلی در توسعه مدل‌های پیش‌بینی مورد استفاده قرار گرفته‌اند [۷]. با این حال، در سال‌های اخیر، ظهور و گسترش فناوری‌های هوش مصنوعی، تحولات چشمگیری در حل مسائل پیچیده حوزه مالی ایجاد کرده است. به‌ویژه، روش‌های یادگیری ماشین به دلیل برخورداری از دقت بالاتر و توانایی بیشتر در شناسایی الگوهای پیچیده، نسبت به مدل‌های کلاسیک، توجه گسترده‌ای را در میان پژوهشگران و متخصصان مالی به خود جلب کرده‌اند[ ۱۸]. کاربردهای یادگیری ماشین در حوزه مالی بسیار گسترده و متنوع است و شامل پیش‌بینی قیمت سهام، معاملات الگوریتمی [۸]، مدیریت ریسک [۲]، قیمت‌گذاری دارایی‌ها [۹] و کشف تقلب مالی [۳۲]. می‌شود. این فناوری‌ها با ارائه الگوریتم‌های بهینه و کارآمد، توانسته‌اند عملکرد سیستم‌های مالی را به‌طور قابل‌توجهی بهبود بخشند. یکی از حوزه‌های پرکاربرد یادگیری ماشین، پیش‌بینی سری‌های زمانی است [۶]. با وجود پیشرفت‌های چشمگیر و تلاش‌های فراوان برای ارائه مدل‌هایی با دقت بالاتر و خطای کمتر، هنوز چالش‌های بنیادینی در این زمینه وجود دارد که ذهن پژوهشگران و کاربران را به خود مشغول کرده است [۲۵]. یکی از این چالش‌های اساسی، نیاز به درک علل و دلایل پنهان پیش‌بینی‌های انجام‌شده توسط مدل‌ها است [۱۱]. در واقع، درک "چرایی" پیش‌بینی‌ها می‌تواند به اندازه "دقت" آن‌ها حائز اهمیت باشد. یکی از موانع اصلی در استفاده گسترده از سیستم‌های مبتنی بر هوش مصنوعی، فقدان شفافیت و قابلیت تفسیرپذیری این مدل‌ها است [۲۲]. به عبارت دیگر، با وجود دقت بالای پیش‌بینی‌های حاصل از مدل‌های یادگیری ماشین، این مدل‌ها اغلب به صورت جعبه‌های سیاه عمل می‌کنند و کاربران قادر به درک نحوه دستیابی مدل به یک پیش‌بینی خاص نیستند [۲۶]. این مسئله باعث کاهش اعتماد به نتایج مدل‌ها و محدودیت در کاربرد عملی آن‌ها می‌شود. در این راستا، مفهوم "هوش مصنوعی توضیح‌پذیر[1]" به عنوان مجموعه‌ای از روش‌ها و تکنیک‌ها مطرح شده است که هدف آن‌ها، افزایش شفافیت و تفسیرپذیری مدل‌های یادگیری ماشین و در نتیجه تقویت اعتماد به نتایج آن‌ها است [۱].

[1] Explainable Artificial Intelligence

هدف این پژوهش، توسعه و ارزیابی یک مدل پیش‌بینی مبتنی بر هوش مصنوعی است که علاوه بر دقت بالا، از قابلیت توضیح‌پذیری نیز برخوردار باشد. در این راستا، تلاش می‌شود تا اهمیت هر یک از متغیرهای ورودی در پیش‌بینی‌های مدل به طور شفاف مشخص شود. سپس، با استفاده از روش‌های اعتبارسنجی متقاطع به ویژه روش بخش‌بندی سری‌های زمانی، پایداری و پایایی نتایج مدل از نظر دقت پیش‌بینی و قابلیت توضیح‌پذیری در بازه‌های زمانی مختلف مورد ارزیابی قرار خواهد گرفت. این پژوهش با ترکیب دقت پیش‌بینی و شفافیت مدل، گامی مؤثر در جهت بهبود کاربردپذیری سیستم‌های هوش مصنوعی در حوزه مالی برداشته و می‌تواند به عنوان الگویی برای پژوهش‌های آینده در این زمینه، مورد استفاده قرار گیرد.

## ۲. مبانی نظری و پیشینه پژوهش

بازارهای مالی به طور ذاتی سیستمی پیچیده و پویا هستند که بازده آن‌ها تحت‌تأثیر عوامل متعددی قرار دارد. بازده شاخص‌ها به ویژه در بورس اوراق بهادار، از متغیرهای کلان اقتصادی مانند نرخ بهره، تورم و نرخ ارز گرفته تا متغیرهای خرد بازار مانند رفتار سرمایه‌گذاران، روندهای تکنیکال و اخبار مرتبط، تأثیر می‌پذیرند [۷,۱۸]. در این میان، شاخص قیمت صنایع تولیدی به عنوان یکی از شاخص‌های مهم بورس اوراق بهادار تهران، نمایانگر وضعیت کلی صنایع تولیدی و بازده آن‌ها است. نوسانات این شاخص به دلیل وابستگی شدید اقتصاد ایران به تحریم‌ها، سیاست‌های پولی و مالی داخلی و تغییرات جهانی، پیش‌بینی بازده آن را بسیار دشوار کرده است [۲۰]. هوش مصنوعی با استفاده از مدل‌هایی همچون شبکه‌های عصبی، یادگیری ماشینی و یادگیری عمیق، توانایی بالایی در تحلیل داده‌های پیچیده و کشف روابط غیرخطی دارد [۹]. با این حال، یکی از چالش‌های اساسی در کاربرد این مدل‌ها، شفافیت پایین و عدم امکان تفسیر خروجی‌ها است [۲۲]. به بیان دیگر، این مدل‌ها به صورت "جعبه سیاه" عمل می‌کنند، به طوری که تحلیل‌گران نمی‌توانند به درستی درک نمایند که چگونه یک مدل به نتیجه خاصی رسیده است [۲۶]. این عدم شفافیت، به ویژه در تصمیم‌گیری‌های مالی که نیازمند اعتماد و اطمینان بالا است، یک محدودیت جدی به شمار می‌رود [۱۱]. برای حل چالش شفافیت و جعبه سیاه بودن مدل‌های هوش مصنوعی، رویکرد هوش مصنوعی توضیح‌پذیر توسعه یافته است [۱]. این رویکرد تلاش می‌کند تا فرآیند تصمیم‌گیری مدل‌های پیچیده هوش مصنوعی را به زبان ساده و قابل فهم برای انسان‌ها توضیح دهد. هدف اصلی توضیح‌پذیری، افزایش شفافیت و اعتماد به مدل‌های هوش مصنوعی است تا تحلیل‌گران و تصمیم‌گیران بتوانند با اطمینان بیشتری از این ابزارها استفاده کنند [۱۵]. روش‌های مطرح در توضیح‌پذیری، شامل تکنیک‌هایی مانند SHAP[1] و LIME[2] هستند که به طور گسترده برای ارائه توضیحات معنادار درباره تصمیمات مدل‌های یادگیری ماشین در حوزه مالی مورد استفاده قرار گرفته‌اند [۲۴]. این روش‌ها امکان تحلیل تأثیر متغیرهای ورودی (مانند نرخ ارز، نرخ بهره و ...) بر خروجی مدل را فراهم می‌کنند. استفاده از هوش مصنوعی توضیح‌پذیر در پیش‌بینی مالی، مزایای متعددی از جمله شناسایی دقیق عوامل کلیدی مؤثر بر پیش‌بینی بازده شاخص، ارائه توضیحات شفاف برای تصمیم‌گیران و تحلیل‌گران مالی، کاهش سوگیری‌های احتمالی و افزایش اعتماد به نتایج مدل را به همراه دارد [۲۶, ۸].

با توجه به مرور ادبیات پژوهش، می‌توان دریافت که اگرچه استفاده از یادگیری ماشین در پیش‌بینی بازارهای مالی به طور گسترده مورد بررسی قرار گرفته است، اما تحقیقات اندکی به ترکیب دقت پیش‌بینی و قابلیت توضیح‌پذیری در مدل‌ها پرداخته‌اند. به‌ویژه، در بازارهای نوظهور مانند ایران که ویژگی‌های منحصربه‌فردی مانند نوسانات بالا و تأثیرپذیری از عوامل سیاسی و اقتصادی دارند، این موضوع کمتر مورد توجه قرار گرفته است. مقاله حاضر برای

[1] Shapley Additive Explanations
[2] Local Interpretable Model-agnostic Explanations

نخستین‌بار در ایران با توسعه و ارزیابی یک مدل پیش‌بینی مبتنی بر هوش مصنوعی توضیح‌پذیر، به بررسی قابلیت توضیح‌پذیری مدل جنگل تصادفی بهینه‌شده با الگوریتم ژنتیک در پیش‌بینی شاخص قیمت صنایع تولیدی بورس تهران پرداخته است که گامی مؤثر در جهت پُرکردن این خلأ پژوهشی است. این پژوهش نه تنها به بهبود دقت پیش‌بینی بازده شاخص قیمت صنایع تولیدی بورس اوراق بهادار تهران می‌پردازد، بلکه با به‌کارگیری مقادیر شپ، شفافیت و تفسیرپذیری مدل را نیز افزایش می‌دهد. در ادامه به پیشینه پژوهش مرتبط با این موضوع پرداخته می‌شود.

نالاکاروپان و همکاران[1] (۲۰۲۴) در مطالعه‌ای باعنوان "ارزیابی ریسک اعتباری و پشتیبانی تصمیم‌گیری مالی با استفاده از هوش مصنوعی توضیح‌پذیر"، یک مدل یادگیری ماشینی توضیح‌پذیر، برای مدیریت ریسک اعتباری پیشنهاد دادند که شفافیت و اعتماد به پیش‌بینی‌های هوش مصنوعی را افزایش می‌داد. این مدل از ابزار مقادیر شپ برای توضیح پیش‌بینی‌ها استفاده می‌کند و به سیاست‌گذاری داده‌محور در مدیریت ریسک اعتباری کمک می‌نماید. نتایج نشان داد که مدل‌های درخت تصمیم و جنگل تصادفی با دقت بالا (۰/۸۹ تا ۰/۹۳) به نهادهای مالی در ارتقای شفافیت، انصاف و اعتمادسازی کمک می‌کند [۲۷]. آرسنالت و همکاران[2] (۲۰۲۴) در پژوهشی باعنوان "مروری بر هوش مصنوعی توضیح‌پذیر در پیش‌بینی سری‌های زمانی مالی" به دسته‌بندی و تحلیل کاربردهای مختلف توضیح‌پذیری پرداخته و همچنین چارچوبی برای انتخاب رویکردهای مناسب آن در کاربردهای مالی ارائه می‌دهد [۳]. مقاله‌ای که توسط لیوو و همکاران[3] (۲۰۲۳) انجام شده، به بررسی قابلیت مدل‌های یادگیری ماشین در پیش‌بینی بازده ارزهای دیجیتال پرداخته است. این پژوهش با استفاده از داده‌های بازده روزانه چندین ارز دیجیتال طی سال‌های ۲۰۱۳ تا ۲۰۲۱، عملکرد مدل‌های مختلف از جمله رگرسیون خطی و مدل تقویت گرادیان شدید را مقایسه کرده است [۲۳]. نتایج نشان می‌دهد که مدل تقویت گرادیان شدید در شناسایی روابط غیرخطی میان ویژگی‌ها و بازده‌ها، عملکرد بهتری نسبت به روش‌های سنتی دارد. همچنین، بازده یک‌روزه قبلی به عنوان مهم‌ترین عامل در پیش‌بینی بازده، شناسایی شده است. مطالعه‌ای که توسط بابایی و همکاران (۲۰۲۲) منتشر شده، به کاربرد هوش مصنوعی توضیح‌پذیر در تخصیص دارایی‌های رمزنگاری شده پرداخته است. این پژوهش ترکیبی از مدل‌های بهینه‌سازی مارکویتز و یادگیری ماشین با استفاده از مقادیر شاپ برای توضیح نتایج را ارائه می‌دهد. یافته‌ها نشان می‌دهند که توضیح‌پذیری مدل‌ها، شفافیت، اعتماد کاربران و پذیرش بیشتر توسط تنظیم‌کنندگان را افزایش می‌دهد و به بهبود عملکرد سیستم‌های مدیریت دارایی‌های رمزنگاری کمک می‌کند [۴].

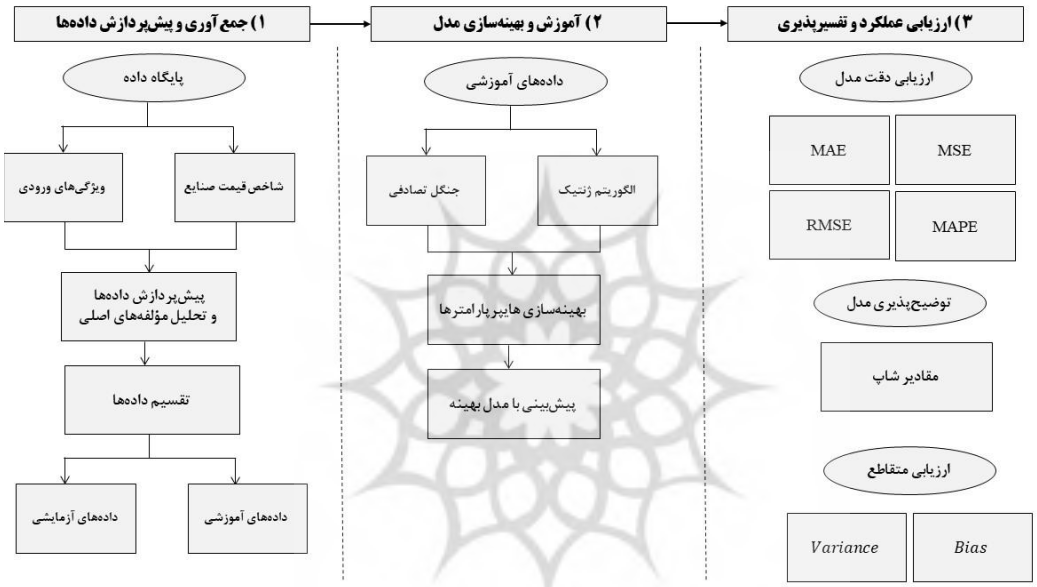
از سوی دیگر، تحقیقات در داخل کشور بیشتر مبتنی بر ارجحیت دقت مدل‌های هوش مصنوعی در پیش‌بینی سری زمانی، انجام شده است. در پژوهشی که توسط غلامی و شمس قارنه (۱۴۰۳) انجام شده است، از مدل‌های ترکیبی LSTM-CNN و CNN-LSTM برای پیش‌بینی قیمت سهام استفاده شد. این مطالعه، داده‌های ده ساله بورس تهران شامل متغیرهایی مانند قیمت سهام، شاخص‌های تکنیکال و نرخ دلار را تحلیل کرده است. الگوریتم PSO[4] برای بهینه‌سازی مدل‌ها به کار رفته و مدل LSTM-CNN بهترین عملکرد را از نظر دقت پیش‌بینی و افزایش بازده مالی داشته است [۱۷]. اقتصاد و محمدی (۱۴۰۲)، سه مدل یادگیری ماشین شامل مدل حافظه بلندمدت کوتاه‌مدت، جنگل تصادفی، مدل خودرگرسیو میانگین متحرک را برای پیش‌بینی بازده و بهینه‌سازی سبد سرمایه‌گذاری در پنج صنعت مختلف در بازه زمانی ۱۳۹۶ تا ۱۴۰۱ بررسی کردند. نتایج نشان داد که مدل جنگل تصادفی در کاهش خطا و افزایش

[1] Nallakaruppan et al.
[2] Arsenault et al.
[3] Liu et al.
[4] Particle Swarm Optimization

بازده عملکرد بهتری داشته و ترکیب آن با مدل میانگین- واریانس مارکویتز، اثربخشی فرآیند بهینه‌سازی سبد را بهبود می‌بخشد [۱۲].

## ۳. روش‌شناسی پژوهش

این پژوهش از نظر هدف، کاربردی و از نظر روش، توصیفی-تحلیلی است که هدف آن پیش‌بینی نوسانات بازار سرمایه و ارزیابی بازده شاخص قیمت هشت صنعت تولیدی بورس اوراق بهادار تهران است. داده‌های این پژوهش از منابع معتبر مالی همچون سازمان بورس و اوراق بهادار تهران و پایگاه‌های اطلاعات مالی (بورس‌ویو) براساس داده‌های آخرین معاملات شاخص‌ها در بازه زمانی پنج ساله ۱۳۹۷ تا ۱۴۰۲ جمع‌آوری و تحلیل مدل‌ها و داده‌ها با استفاده از زبان برنامه‌نویسی پایتون[1] و در محیط ژوپیتر پیاده‌سازی شده‌اند. شکل ۱، ساختار روش پیشنهادی پژوهش را نشان می‌دهد که در ادامه به‌صورت مبسوط به هر کدام از مراحل پراخته خواهد شد.




***شکل ۱.*** *ساختار و مراحل روش پیشنهادی پژوهش*

**جمع‌آوری و پیش‌پردازش داده‌ها:** در این پژوهش، برای بهبود کارایی مدل و حذف روندهای غیرضروری، از بازده لگاریتمی شاخص به‌عنوان ورودی اصلی مدل استفاده شده است. این روش، موجب ایستا شدن داده‌ها و هموار شدن نوسانات می‌شود که منجر به آموزش دقیق‌تر مدل خواهد شد. ویژگی‌های مدل که برای آموزش مدل وارد می‌شوند، به سه دسته اصلی متغیرهای تکنیکال، متغیرهای بنیادی و متغیرهای کلان اقتصادی تقسیم شدند. ویژگی‌ها به صورت روزانه یا ماهانه، گردآوری و به طور پله‌ای وارد مدل شدند. به عبارت دیگر، هر ویژگی ماهانه در طول همان ماه تکرار می‌گردد تا مدل بتواند از آن بهره‌برداری کند. همچنین، در روزهایی که بازار سهام و بازارهای جهانی هر دو باز نیستند، برای تعیین قیمت‌های بازار جهانی که به مدل وارد می‌شوند، از آخرین قیمت به عنوان قیمت مرجع استفاده شده است. در جدول ۱، تمامی ویژگی ورودی مدل و نحوه استخراج آن‌ها بیان شده است.

[1] Python

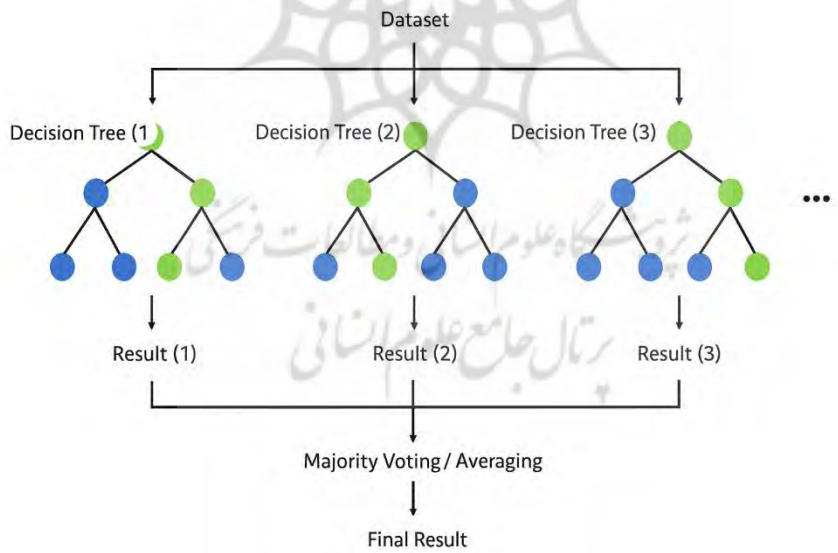
***جدول ۱.*** *نوع و نحوه استخراج ویژگی‌های ورودی*

| نوع ویژگی | نام ویژگی | نحوه استخراج یا فرمول محاسبه ویژگی | دوره محاسبه |
|---|---|---|---|
| **تکنیکال** | Volume | حجم معاملات به تعداد سهام یا قراردادهایی که در یک بازه زمانی مشخص (معمولاً یک روز) معامله شده‌اند، گفته می‌شود. | روزانه |
| **تکنیکال** | Trade Value | ارزش معاملات به مقدار پولی که در یک بازه زمانی مشخص در بازار برای خرید و فروش سهام یا قراردادها هزینه شده است، گفته می‌شود. | روزانه |
| **تکنیکال** | Free Float | درصدی از کل سهام شرکت است که در بازار آزاد، قابل معامله بوده و در دسترس عموم قرار دارد. | روزانه |
| **تکنیکال** | SMA | $SMA = \frac{P_1 + P_2 + \cdots + P_n}{n}$ | ۱۰، ۲۰، ۵۰ روزه |
| **تکنیکال** | EMA | میانگین متحرک نمایی (EMA) مانند SMA است، اما به قیمت‌های اخیر وزن بیشتری می‌دهد. این باعث می‌شود که EMA سریع‌تر به تغییرات قیمت واکنش نشان دهد. | ۱۰، ۲۰، ۵۰ روزه |
| **تکنیکال** | RSI | $RSI = 100 - \frac{100}{\frac{Average\ Gain}{Average\ Loss} + 1}$ <br> $\frac{Average\ Gain}{Average\ Gain}$: میانگین افزایش قیمت در دوره‌های بالارفتن قیمت <br> $\frac{Average\ Gain}{Average\ Gain}$: میانگین کاهش قیمت در دوره‌های پایین‌آمدن قیمت | ۱۴ روزه |
| **تکنیکال** | MACD | $MACD = EMA_{12} - EMA_{26}$ | ۱۴، ۳۰، ۵۰ روزه |
| **تکنیکال** | SD | $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(\mu - P_i)^2}{n}}$ | ۱۴، ۳۰، ۵۰ روزه |
| **تکنیکال** | OBV | اگر قیمت بسته شدن امروز بالاتر از قیمت بسته شدن دیروز باشد، حجم معاملات امروز به OBV قبلی اضافه می‌شود.<br>اگر قیمت بسته شدن امروز پایین‌تر از قیمت بسته شدن دیروز باشد، حجم معاملات امروز از OBV قبلی کم می‌شود.<br>اگر قیمت بسته شدن امروز برابر با قیمت بسته شدن دیروز باشد، OBV بدون تغییر باقی می‌ماند. | ۱۴، ۲۱، ۳۰ روزه |
| **تکنیکال** | VPT | حجم- قیمت روند (VPT) یک شاخص تکنیکال است که تغییرات حجم معاملات و قیمت را ترکیب می‌کند تا قدرت روند قیمت را اندازه‌گیری کند. این شاخص مشابه شاخص OBV است، اما تغییرات حجم معاملات را بر اساس درصد تغییرات قیمت، تنظیم می‌کند.<br>$VPT_{Current} = VPT_{Previous} + (Volume_{Current} * \%Price\ Change)$ | ۱۴، ۲۱، ۳۰ روزه |
| **بنیادی** | مبلغ فروش ماهانه | جمع مبلغ فروش ماهانه شرکت‌های موجود در یک صنعت | ماهانه |
| **بنیادی** | ارزش بازار صنعت | ارزش روز بازاری هر یک از صنایع تولیدی بورس تهران به تفکیک | روزانه |
| **بنیادی** | $\frac{P}{E}ttm$ | این نسبت، نشان‌دهنده قیمت یک سهم نسبت به سود هر سهم (EPS) است. برای محاسبه این نسبت، قیمت فعلی بازار یک سهم بر سود خالص هر سهم تقسیم می‌شود. | روزانه |
| **بنیادی** | $\frac{P}{B}$ | این نسبت، نشان‌دهنده نسبت قیمت بازار یک سهم به ارزش دفتری (کتابی) هر سهم است. برای محاسبه P/B، قیمت بازار یک سهم بر ارزش دفتری هر سهم (ارزش خالص دارایی‌های شرکت تقسیم بر تعداد سهام در جریان) تقسیم می‌شود. | روزانه |
| **بنیادی** | $\frac{P}{S}$ | این نسبت، نشان‌دهنده نسبت قیمت بازار یک سهم به درآمد (فروش) هر سهم است. برای محاسبه P/S، قیمت بازار یک سهم بر درآمد (فروش) هر سهم (درآمد کل شرکت تقسیم بر تعداد سهام در جریان) تقسیم می‌شود. | روزانه |
| **کلان اقتصادی** | قیمت جهانی طلا | معمولاً قیمت جهانی طلا بر اساس هر اونس (حدود ۲۸,۳۵ گرم) اعلام می‌شود و ممکن است در بازارهای مختلف به صورت لحظه‌ای تغییر کند. | روزانه |
| **کلان اقتصادی** | قیمت جهانی نفت | قیمت جهانی نفت به قیمتی اشاره دارد که برای نفت خام در بازارهای جهانی تعیین می‌شود. | روزانه |

| | | | |
|---|---|---|---|
| **کلان اقتصادی** | نرخ دلار آزاد | مفهوم "نرخ دلار آزاد" به قیمت دلار آمریکا در بازارهای آزاد ارز اشاره دارد. | روزانه |
| **کلان اقتصادی** | نرخ بهره بدون ریسک | نرخ بهره بدون ریسک به سودی اشاره دارد که در صورت سرمایه‌گذاری در اوراق بهادار با کمترین ریسک ممکن، انتظار می‌رود به دست آید. در ایران، به دلیل عدم وجود اوراق قرضه با پشتوانه دولتی با ریسک صفر، نرخ بهره بدون ریسک به صورت تئوریک و با استفاده از اوراق مشارکت یا سپرده‌های بانکی با سود قطعی و با در نظر گرفتن تورم، محاسبه می‌شود. | روزانه |
| **کلان اقتصادی** | نرخ تورم | نرخ تورم یا تورم، نشان‌دهنده افزایش عمومی قیمت‌ها در اقتصاد است و نشان‌دهنده تغییرات قیمت‌ها برای یک سبد از کالاها و خدمات است که مصرف کنندگان معمولاً آن‌ها را مصرف می‌کنند. | ماهانه |

همچنین، برای کاهش ابعاد ویژگی‌ها از تحلیل مولفه‌های اصلی استفاده شده است که با انتخاب مولفه‌هایی که ۸۰ درصد واریانس داده‌های اصلی را توضیح می‌دهند، نهایتاً چهار یا پنج ویژگی به عنوان ویژگی‌های نهایی انتخاب شده است. کل داده‌ها با استفاده از روش ارزیابی متقاطع سری زمانی[1] پنج بخش تقسیم شده‌اند. در هر پنج بخش از ارزیابی متقاطع، ۸۰ در صد داده‌ها برای آموزش مدل و ۲۰ در صد برای آزمایش مدل، ا ستفاده شده ا ست. این روش امکان ارزیابی عملکرد مدل در شرایط واقعی بازار را فراهم نموده و از داده نشت[2] جلوگیری می‌کند.

**آموزش و بهینه‌سازی مدل:** مدل توسعه یافته در این پژوهش، یک مدل جنگل تصادفی است که به‌عنوان یک الگوریتم یادگیری تجمیعی عمل می‌کند. هایپرپارامترهای این مدل با استفاده از الگوریتم ژنتیک بهینه‌سازی شده‌اند. این مدل به دلیل توانایی بالای خود در یادگیری الگوهای پیچیده به دلیل ترکیب چند یادگیر ضعیف و کاهش بیش‌برازش از طریق ترکیب چندین درخت تصمیم‌گیری، انتخاب شده است. مدل جنگل تصادفی (شکل ۲) با ساخت درختان مستقل، نتایج پیش‌بینی‌های هر درخت را با هم ترکیب می‌کند تا بر اساس میانگین نتایج پیش‌بینی درختان، پیش‌بینی نهایی را انجام دهد.




***شکل ۲.*** *مدل جنگل تصادفی*

[1] Time Series Split Cross-Validation
[2] Data Leakage

برای بهبود دقت پیش‌بینی‌ها، از الگوریتم ژنتیک برای تنظیم هایپرپارامترهای مدل جنگل تصادفی استفاده شده است. هایپرپارامترهای جنگل تصادفی شامل موارد زیر می‌باشد:

۱) تعداد برآوردگرها (درخت‌ها): تعداد درخت‌هایی که در مدل نهایی آموزش داده می‌شوند. این پارامتر، کنترل تعداد تکرارهای مدل و ترکیب نتایج را بر عهده دارد.

۲) حداکثر تعداد ویژگی‌ها برای تقسیم‌بندی: تعداد ویژگی‌هایی که در هر تقسیم‌بندی مجاز به انتخاب هستند. این پارامتر کمک می‌کند تا مدل، حساسیت کمتری به نویز داشته باشد و از بیش‌برازش جلوگیری کند.

۳) حداقل تعداد نمونه برای تقسیم گره: حداقل تعداد نمونه‌هایی که یک گره باید داشته باشد تا بتوان آن را به دو گره تقسیم کرد. اگر تعداد نمونه‌ها در گره کمتر از این مقدار باشد، گره نهایی (برگ) محسوب می‌شود.

۴) حداقل تعداد نمونه در برگ نهایی: این پارامتر مشخص می‌کند که هر برگ نهایی در درخت تصمیم‌گیری باید حداقل چند نمونه داشته باشد. یعنی اگر بعد از تقسیم، در یکی از شاخه‌ها، تعداد نمونه‌ها کمتر از این مقدار باشد، آن تقسیم انجام نمی‌شود.

الگوریتم ژنتیک، الگوریتمی فرا ابتکاری است که با الهام از فرایند تکامل طبیعی به منظور حل مسائل بهینه‌سازی به کار می‌رود. در این مطالعه، از الگوریتم ژنتیک برای بهینه‌سازی هایپرپارامترهای مدل استفاده شده است. این الگوریتم با ایجاد جمعیتی از کروموزوم‌ها که در واقع یک رشته کد یا بردار است که مقادیر هایپرپارامترهای یک مدل را نمایش می‌دهد، با انتخاب، ترکیب و جهش تلاش می‌کند به مجموعه‌ای از هایپرپارامترها برسد که بهترین عملکرد را داشته باشند.

برای ارزیابی کیفیت هر کروموزوم، از تابع برازندگی استفاده می‌شود. در این پژوهش، معیار ریشه میانگین مربعات خطا[1] به‌عنوان تابع برازندگی به کار گرفته شده است. این معیار، میزان اختلاف بین مقادیر پیش‌بینی‌شده مدل و مقادیر واقعی را اندازه‌گیری می‌کند و هرچه مقدار آن کمتر باشد، کروموزوم برازنده‌تر محسوب می‌شود.

در تنظیمات الگوریتم، تعداد نسل‌ها[2] برابر با ۵ انتخاب شده است تا فرآیند تکامل در طی پنج مرحله به سمت بهبود عملکرد مدل هدایت شود. اندازه جمعیت اولیه[3] برابر ۵۰ در نظر گرفته شده است تا تنوع کافی در کروموزوم‌ها وجود داشته باشد و الگوریتم از گرفتار شدن در بهینه‌های محلی، جلوگیری کند. مقدار حالت تصادفی[4] برابر با ۴۲ تنظیم شده است تا امکان بازتولیدپذیری نتایج فراهم شود. میزان نمایش جزئیات خروجی[5] نیز بر روی مقدار ۲ تنظیم شده. این فرآیند در قالب کلاس TPOTRegressor در پایتون پیاده‌سازی شده است.

مراحل اجرای الگوریتم به شرح زیر است:

- ایجاد جمعیت اولیه: تولید مجموعه‌ای از کروموزوم‌ها با مقادیر تصادفی برای هایپرپارامترها.
- ارزیابی: اندازه‌گیری عملکرد هر کروموزوم با استفاده از تابع برازندگی.
- انتخاب: گزینش کروموزوم‌هایی با عملکرد بهتر برای انتقال به نسل بعد.
- ترکیب: ادغام اطلاعات دو کروموزوم والد برای تولید کروموزوم‌های جدید (فرزندان).
- جهش: تغییر تصادفی مقادیر بخشی از یک کروموزوم به منظور حفظ تنوع جمعیت و جلوگیری از همگرایی زودهنگام.

[1] RMSE
[2] generations
[3] population_size
[4] random state
[5] verbosity

- تکرار: انجام مراحل فوق تا زمانی که مجموعه‌ای بهینه از هایپرپارامترها به‌دست آید.

در نتیجه این فرآیند، مدل جنگل تصادفی بهینه‌سازی می‌شود و از گرفتار شدن در بهینه‌های محلی جلوگیری می‌کند، به‌گونه‌ای که مدل به سمت مقادیر بهینه‌تر در فضای جستجو هدایت می‌شود. برای تحلیل پیچیدگی محاسباتی فرآیند بهینه‌سازی جنگل تصادفی با استفاده از الگوریتم ژنتیک، ابتدا نمادهای اصلی مورد استفاده را تعریف می‌کنیم: $T$ نشان‌دهنده تعداد درخت‌ها، $n$ تعداد نمونه‌های آموزشی، $d$ ابعاد فضای ویژگی، $P$ اندازه جمعیت الگوریتم ژنتیک و $G$ تعداد نسل‌های فرآیند تکامل است.

با فرض اینکه هر درخت به طور میانگین از $\sqrt{d}$ ویژگی استفاده کرده و تا عمق $\log n$ رشد می‌کند، آموزش جنگل تصادفی دارای پیچیدگی زمانی در حدود $O(T * n * \sqrt{d} * \log n)$ است و پس از اتمام آموزش، زمان محاسباتی در حدود $O(T * \log n)$ نیاز دارد؛ زیرا مسیر از ریشه تا برگ تمامی درخت‌ها باید طی شود. زمانی که الگوریتم ژنتیک برای انتخاب ویژگی‌ها یا تنظیم ابرپارامترها به‌کار گرفته می‌شود، هر فرد جمعیت نمایانگر یک مدل جنگل تصادفی کامل است و این ارزیابی در تمام نسل‌ها تکرار می‌شود. در نتیجه، پیچیدگی زمانی کل فرآیند برابر است با $O(G * P * T * n * \sqrt{d} * \log n)$. با وجود این، از آنجا که معمولاً تنها تعداد محدودی از بهترین مدل‌ها ذخیره می‌شوند، میزان مصرف حافظه نسبتاً محدود مانده و در حدود $O(T * n)$ باقی می‌ماند. در نتیجه، انتخاب مقادیر میانه برای $P$ و $G$ (برای مثال جمعیتی حدود ۳۰ فرد در ۲۰ نسل) امکان پیمایش کارآمد فضای جستجو را فراهم می‌کند، بی‌آنکه زمان اجرای الگوریتم به طور نامطلوبی افزایش یابد.

## ارزیابی عملکرد و تفسیرپذیری

**ارزیابی دقت مدل:** برای ارزیابی دقت پیش‌بینی مدل‌ها در سری‌های زمانی، از معیارهای آماری رایج در حوزه پیش‌بینی مالی استفاده شده است (جدول ۲).

***جدول ۲.*** *معیارهای ارزیابی در سری‌های زمانی*

| معیار ارزیابی | نحوه محاسبه | تحلیل معیار ارزیابی |
|---|---|---|
| **ریشه میانگین مربعات خطا** | $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ | RMSE، مقدار خطای پیش‌بینی را به همان واحد متغیر اصلی می‌سنجد. هرچه مقدار RMSE کمتر باشد، دقت مدل بیشتر است. |
| **میانگین قدرمطلق درصد خطا** | $MAPE = \frac{100}{n}\sum_{i=1}^{n}\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|$ | MAPE، درصدی از خطای مطلق را نسبت به مقدار واقعی نشان می‌دهد. مقدار کمتر MAPE بیانگر پیش‌بینی دقیق‌تر است. |
| **میانگین قدر مطلق خطا** | $MAE = \frac{1}{n}\sum_{i=1}^{n}\|y_i - \hat{y}_i\|$ | MAE، میانگین قدر مطلق خطاها را ارائه می‌دهد. مقادیر کمترMAE ، نشان-دهنده دقت بالای مدل است. |
| **میانگین مربعات خطا** | $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ | MSE، میانگین مجذور خطاها را نشان می‌دهد و به مقادیر بزرگتر خطاها حساس‌تر است. مقدار کمتر MSE، بیانگر عملکرد بهتر مدل است. |

**ارزیابی متقاطع:** یکی از چالش‌های اساسی در پیش‌بینی سری‌های زمانی، ارزیابی میزان پایداری و تعمیم‌پذیری مدل است. در این پژوهش، برای سنجش دقت و استحکام مدل، از روش ارزیابی متقاطع مبتنی بر بخش‌بندی سری زمانی استفاده شد. این روش، که به طور گسترده در تحلیل سری‌های زمانی به کار می‌رود، ساختار زمانی داده‌ها را حفظ کرده و از نشت اطلاعات جلوگیری می‌کند. [۵]. در این رویکرد، داده‌ها به صورت ترتیبی بخش‌بندی شده و مدل به تدریج بر روی داده‌های بیشتری آموزش می‌بیند. این شیوه به‌ویژه در مسائل پیش‌بینی سری‌های زمانی، جایگزین کارآمدتری

نسبت به جداسازی داده‌های اعتبارسنجی[1] است زیرا از نشت اطلاعات آینده به مدل جلوگیری می‌کند و شرایط واقعی بازار را بهتر شبیه‌سازی می‌نماید .مزیت اصلی این روش آن است که هیچ‌گونه داده‌ای از آینده در مجموعه آموزشی قرار نمی‌گیرد؛ لذا منتج به شبیه‌سازی شرایط واقعی بازار می‌شود.

با اعمال فرآیند ارزیابی متقاطع در چندین بازه زمانی و بررسی عملکرد مدل در هر مرحله، می‌توان مطابق با معادلات ۱ و ۲ به تعادلی میان بایاس (میانگین دقت مدل در بخش‌ها) و واریانس (تغییرپذیری دقت مدل در بخش‌های مختلف) دست یافت. این ویژگی، اهمیت ویژه‌ای در مدل‌های مالی دارد، زیرا امکان افزایش پایداری تخمین‌ها و کاهش تأثیر نوسانات کوتاه‌مدت را فراهم می‌آورد.

$$bias(\hat{\theta}) = (n-1)(\bar{\hat{\theta}} - \hat{\theta}) \quad (۱)$$

$bias(\hat{\theta})$ : معیار بایاس برای ارزیابی متقاطع؛

n: تعداد بخش‌بندی ارزیابی متقاطع؛

$(\bar{\hat{\theta}} - \hat{\theta})$: میزان انحراف میانگین دقت‌ها از دقت در هر یک از بخش‌های مختلف؛

$$Var(\hat{\theta}) = \frac{n-1}{n}\sum_{i=1}^{n}(\bar{\hat{\theta}} - \hat{\theta})^2 \quad (۲)$$

$Var(\hat{\theta})$ : معیار بایاس برای ارزیابی متقاطع؛

n: تعداد بخش‌بندی ارزیابی متقاطع؛

$(\bar{\hat{\theta}} - \hat{\theta})$: میزان انحراف میانگین دقت‌ها از دقت در هر یک از بخش‌های مختلف؛

**توضیح‌پذیری یادگیری ماشینی:** یکی از جنبه‌های کلیدی در شفافیت عملکرد مدل‌های هوش مصنوعی، توضیح‌پذیری است که امکان درک بهتر روابط میان متغیرهای ورودی و خروجی مدل را فراهم می‌سازد. این امر نه‌تنها موجب افزایش اعتماد کاربران و ذینفعان به مدل می‌شود، بلکه به بهینه‌سازی و بهبود عملکرد مدل نیز کمک می‌کند. در این پژوهش، از مقادیر شاپ مبتنی بر تئوری بازی‌های مشارکتی برای محاسبه میزان تأثیرگذاری متغیرهای ورودی بر خروجی مدل استفاده شده است. تئوری بازی‌های مشارکتی، بر نحوه تخصیص منصفانه منابع و ارزش میان اعضای یک ائتلاف تمرکز دارد. روش ارزش شاپلی، به عنوان یکی از تکنیک‌های مهم در این حوزه، اهمیت هر متغیر را بر اساس تأثیر آن بر خروجی مدل و با در نظر گرفتن تمامی ترکیبات ممکن از متغیرها محاسبه می‌کند [۳۱]. مطابق با معادله ۳، مقدار شاپ یک متغیر خاص در یک مدل با N متغیر و تابع ارزش υ، از طریق میانگین ارزش افزوده‌ای که این متغیر به تمامی زیرمجموعه‌های ممکن از متغیرها اضافه می‌کند، محاسبه می‌شود:

$$\phi_i(\upsilon) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!}\left[\upsilon\big((S \cup \{i\})\big) - \upsilon(S)\right] \quad (۳)$$

$\phi_i(\upsilon)$ : مقدار شاپلی برای بازیکن i ؛

N: تعداد کل بازیکنان؛

S: زیرمجموعه‌ای از بازیکنان به غیر از بازیکن i؛

$|S|$: تعداد اعضای زیرمجموعه S ؛

[1] validation set

$\upsilon((S \cup \{i\})) - \upsilon(S)$: ارزش افزوده‌ای که بازیکن i به ائتلاف S اضافه می‌کند؛

$\frac{|S|!(|N|-|S|-1)!}{|N|!}$: ضریبی که نشان‌دهنده احتمال وقوع ترتیب‌های مختلف است؛

## ٤. تحلیل داده‌ها و یافته‌ها

**عملکرد دقت مدل :** جدول ۳، نتایج حاصل از دقت مدل جنگل تصادفی بهینه‌شده را ارائه می‌دهد.

***جدول ۳.*** *نتایج حاصل از دقت مدل جنگل تصادفی بهینه‌شده*

| ردیف | صنایع | MSE | MAE | MAPE | RMSE | P_Value Significant level α = 0.05 |
|---|---|---|---|---|---|---|
| ۱ | دارویی | ۱/۶۲E-۰۶ | ۰/۰۰۰۹۸۵ | ۴/۱۰۰۲۳۶ | ۰/۰۰۱۲۳۳ | ۰/۰۰۵۸۸۵ |
| ۲ | فولاد | ۲/۶۱E-۰۶ | ۰/۰۰۱۲۳۵ | ۴/۷۲۹۴۳۱ | ۰/۰۰۱۵۴۳ | ۰/۰۱۹۲۴۸ |
| ۳ | کانی | ۲/۳۶E-۰۶ | ۰/۰۰۱۱۲۳ | ۳/۷۴۶۴۰۲ | ۰/۰۰۱۴۶۸ | ۰/۰۰۳۸۲۵ |
| ۴ | کاشی | ۲/۸۸E-۰۶ | ۰/۰۰۱۲۸۱ | ۱/۴۹۵۳۰۷ | ۰/۰۰۱۶۵۱ | ۰/۰۰۵۱۲ |
| ۵ | خودرو | ۶/۵۸E-۰۶ | ۰/۰۰۱۹۹۳ | ۳/۱۷۸۴۶۶ | ۰/۰۰۲۵۰۳ | ۰/۰۰۲۴۸۷ |
| ۶ | لاستیک | ۳/۰۲E-۰۶ | ۰/۰۰۱۳۳۲ | ۲/۲۵۴۸۳۷ | ۰/۰۰۱۶۸۹ | ۰/۰۰۳۳۲۸ |
| ۷ | شیمیایی | ۱/۸۲E-۰۶ | ۰/۰۰۱۰۰۲ | ۴/۵۲۷۵۳۸ | ۰/۰۰۱۲۷۹ | ۰/۰۱۳۶۵ |
| ۸ | سیمان | ۱/۹۱E-۰۶ | ۰/۰۰۱۰۵۹ | ۳/۱۳۴۴۷۳ | ۰/۰۰۱۳۶۱ | ۰/۰۰۵۶۸۴ |

نتایج به‌دست‌آمده از جدول ۳، بیانگر عملکرد مطلوب مدل است. بر اساس معیارهای ریشه میانگین مربعات خطا، میانگین قدر مطلق خطا و میانگین مربعات خطا، بالاترین دقت پیش‌بینی به شاخص صنعت دارویی اختصاص دارد، در حالی که شاخص صنعت خودرو، کمترین دقت را نشان داده است. همچنین، با توجه به معیار میانگین قدرمطلق درصد خطا ، مدل در پیش‌بینی شاخص صنعت کاشی بهترین عملکرد را داشته است، درحالی‌که صنعت فولاد، بالاترین میزان خطا را تجربه کرده است. این یافته‌ها بر کارایی بالای مدل پیشنهادی در پیش‌بینی بازده شاخص‌های مختلف صنایع تأکید دارند. همچنین برای معنی‌داری نتایج دقت مدل در سطح اطمینان ۹۵ درصد از آزمون آماری برای تمامی صنایع استفاده شد و مقادیر P-Value به ازای هر صنعت محاسبه گردید که با توجه به نتایج (مقادیر P-Value کمتر از ۰/۰۵) ، فرضیه صفر رد شده و نتایج دقت مدل به صورت آماری تأیید شد. همچنین، مقادیر هایپرپارامترهای بهینه شده مدل در جدول ۴ ارائه شده است.

***جدول ٤.*** *مقادیر هایپرپارامترهای بهینه‌شده توسط الگوریتم ژنتیک در هر صنعت*

| صنعت | *n_estimators* | *max_features* | *min_samples_split* | *min_samples_leaf* |
|---|---|---|---|---|
| دارویی | ۲۰۰ | sqrt | ۲ | ۱ |
| سیمان | ۱۰۰ | log۲ | ۲ | ۱ |
| فولاد | ۱۰۰ | sqrt | ۵ | ۲ |
| لاستیک | ۱۰۰ | log۲ | ۲ | ۱ |
| کانی | ۲۰۰ | sqrt | ۱۰ | ۲ |
| کاشی | ۱۰۰ | log۲ | ۲ | ۱ |
| خودرو | ۲۰۰ | log۲ | ۵ | ۲ |
| شیمیایی | ۲۰۰ | sqrt | ۲ | ۱ |

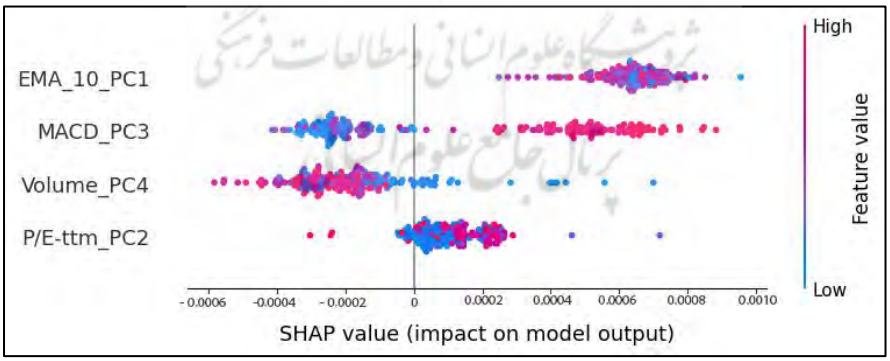
**توضیح‌پذیری مدل جنگل تصادفی:** برای درک بهتر نحوه تأثیرگذاری متغیرهای ورودی بر پیش‌بینی مدل، از نمودار خلاصه شاپ استفاده شده است. این نمودار به صورت بصری، میزان و جهت تأثیر ویژگی‌های مختلف را بر خروجی

مدل نمایش می‌دهد. در این نمودار، هر نقطه نمایانگر یک مشاهده در مجموعه داده‌ها است و رنگ نقاط بیانگر مقدار ویژگی مربوطه می‌باشد. تحلیل نمودار شاپ به شرح ذیل است [۲۴].

- مقادیر شاپ مثبت (نقاط قرمزتر)، نشان‌دهنده ویژگی‌هایی هستند که باعث افزایش مقدار پیش‌بینی‌شده توسط مدل می‌شوند. به عبارت دیگر، هر چه مقدار این ویژگی‌ها بالاتر باشد، خروجی مدل نیز بیشتر خواهد شد.
- مقادیر شاپ منفی (نقاط آبی‌تر)، نشان‌دهنده ویژگی‌هایی هستند که موجب کاهش مقدار پیش‌بینی‌شده می‌شوند. در واقع، این متغیرها تأثیر معکوسی بر خروجی مدل دارند.
- محور افقی، مقدار شاپ را نشان می‌دهد؛ مقدار مطلق بالاتر نشان‌دهنده اثرگذاری قوی‌تر آن ویژگی بر مدل است.
- محور عمودی، ترتیب اهمیت کلی ویژگی‌ها را از بالا به پایین نمایش می‌دهد، به طوری که متغیرهای قرار گرفته در بخش بالایی، بیشترین تأثیر را بر پیش‌بینی‌های مدل داشته‌اند.

این تحلیل به شفاف‌سازی نقش متغیرهای ورودی در مدل جنگل تصادفی کمک می‌کند و نشان می‌دهد که کدام ویژگی‌ها، بیشترین و کمترین اهمیت را در پیش‌بینی بازده لگاریتمی شاخص صنایع دارند. بنابراین، با توجه به نمودارهای خلاصه شاپ هر یک از صنایع (شکل‌های ۳ الی ۱۰)، تحلیل توضیح‌پذیری نتایج پیش‌بینی بازده شاخص قیمتی در هشت صنعت مورد مطالعه، به شرح ذیل ارائه می‌شود:

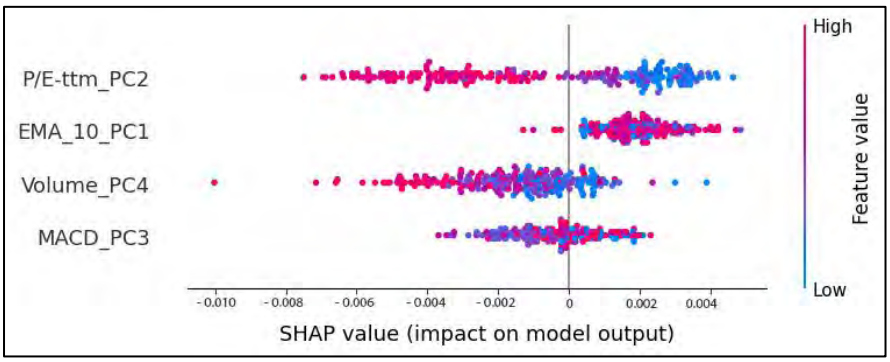
۱) **صنعت دارویی:** بر پایه شکل ۳، یافته‌های حاصل از تحلیل حاکی از آن است که ویژگی «میانگین متحرک نمایی ده‌روزه» بیشترین سهم را در تعیین خروجی مدل داشته و با افزایش مقدار آن، بازدهی پیش‌بینی‌شده نیز افزایش می‌یابد؛ ازاین‌رو، این متغیر نقشی مثبت و مؤثر در سازوکار مدل ایفا می‌کند. در مقابل، متغیر «همگرایی و واگرایی میانگین متحرک» از اهمیت کمتری برخوردار بوده و اثر آن بر خروجی مدل، بسته به مقدارش، می‌تواند مثبت یا منفی باشد. همچنین، ویژگی‌های مربوط به «حجم معاملات» و «نسبت قیمت به سود در دوازده ماهه گذشته» تأثیر چندانی بر پیش‌بینی بازده نداشته و نقش آن‌ها در مدل بسیار محدود ارزیابی می‌شود. به طور کلی، هرچه مقدار میانگین متحرک نمایی ده روزه افزایش یابد، اثر تقویتی آن بر پیش‌بینی بازده بیشتر خواهد بود، در حالی‌که سایر ویژگی‌های یادشده در تعیین نتیجه مدل، سهم اندکی دارند.




***شکل ۳.*** *یافته‌های حاصل از توضیح‌پذیری صنعت دارویی*

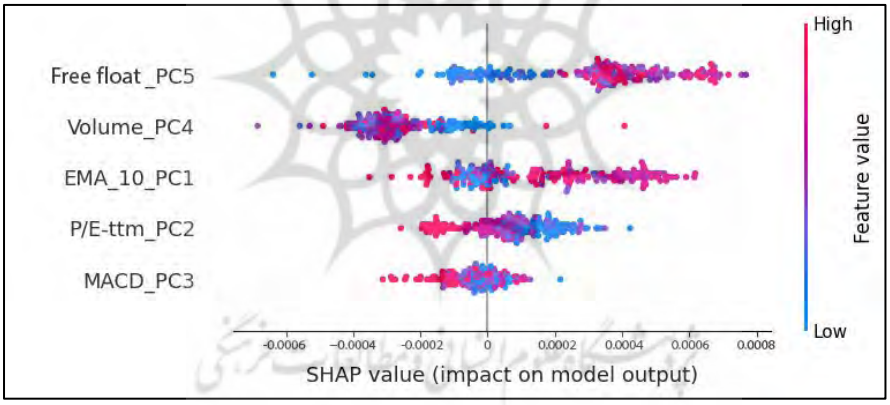
۲) **صنعت فولاد:** بر اساس شکل ۴، متغیر نسبت قیمت به سود دوره دوازده ماهه، آثار مثبت و منفی بر مدل دارد و نشان‌دهنده تأثیر وابسته به مقدار آن است؛ به‌طوری‌که سطوح بالاتر، اثر کاهنده و سطوح پایین‌تر، اثر تقویتی دارند. میانگین متحرک نمایی ده روزه، تأثیری مثبت و معنادار بر پیش‌بینی داشته و با افزایش مقدار آن، دقت

مدل بهبود می‌یابد. در مقابل، حجم معاملات، عمدتاً تأثیری منفی اما محدود دارد و ویژگی همگرایی و واگرایی میانگین متحرک نیز نقشی بسیار ناچیز در خروجی مدل ایفا می‌کند.



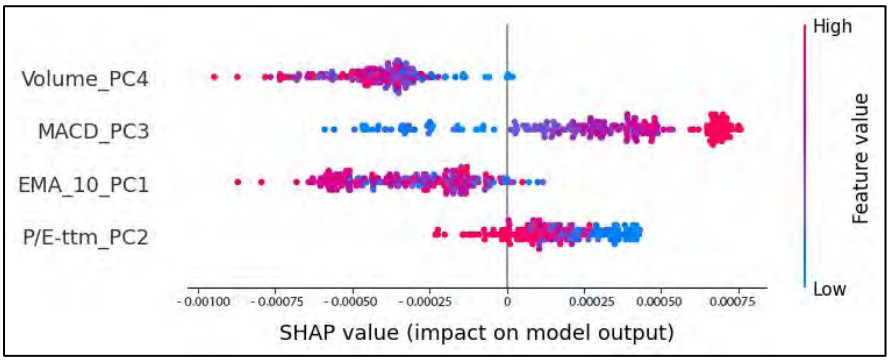

*شکل ٤. یافته‌های حاصل از توضیح‌پذیری صنعت فولاد*

۳) **صنعت کانی:** بر پایه شکل ۵، ویژگی سهام شناور، تأثیر معنادار و دگرگون‌پذیری بر مدل دارد؛ به‌گونه‌ای که مقادیر بالاتر آن، اثر تقویتی و مقادیر پایین‌تر، اثر تضعیفی دارند. نسبت قیمت به سود دوازده ماهه نیز عمدتاً نقشی مثبت و مؤثر ایفا می‌کند. در مقابل، میانگین متحرک نمایی ده روزه، بیشتر تأثیر منفی داشته و به‌ویژه در سطوح پایین‌تر، موجب کاهش دقت مدل می‌شود. حجم معاملات دارای اثری میانه با گرایش نسبی به سمت کاهش بازده در مقادیر پایین است. همچنین، همگرایی و واگرایی میانگین متحرک، نقشی ضعیف و عمدتاً منفی بر جای می‌گذارد.



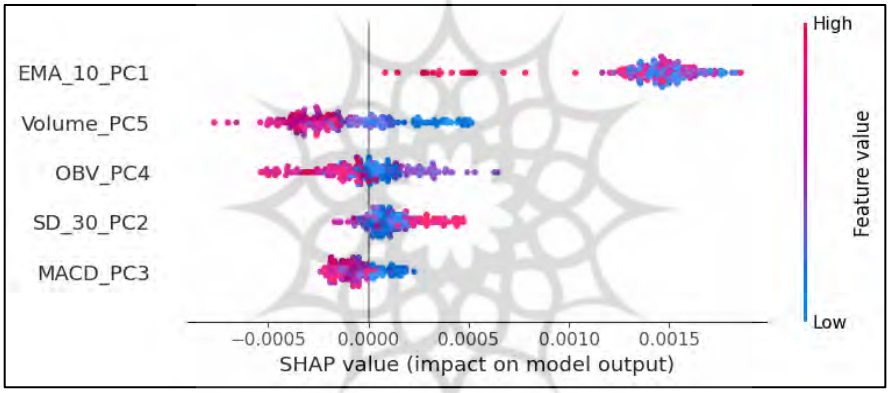

*شکل ۵. یافته‌های حاصل از توضیح‌پذیری صنعت کانی*

۴) **صنعت کاشی:** بر اساس شکل ۶، ویژگی حجم معاملات دارای اثری میانه و دوسویه است و بیشتر مقادیر آن، پیرامون منفی قرار دارند. متغیر همگرایی و واگرایی میانگین متحرک نیز اثری بسیار اندک دارد و نقش آن در مدل ناچیز است. در مقابل، میانگین متحرک نمایی ده روزه، اثری منفی و نسبتاً قوی دارد؛ به ویژه در مقادیر بالاتر که کاهش بازده را در پی دارد. نسبت قیمت به سود دوازده ماهه، عمدتاً دارای اثری مثبت و مؤثر است، به‌ویژه در سطوح بالاتر آن. در مجموع، میانگین متحرک نمایی و نسبت قیمت به سود از اثرگذارترین متغیرهای مدل هستند؛ عامل اول با نقش تضعیفی و عامل دوم با اثر تقویتی. سایر متغیرها نقش کمتری در پیش‌بینی مدل دارند.



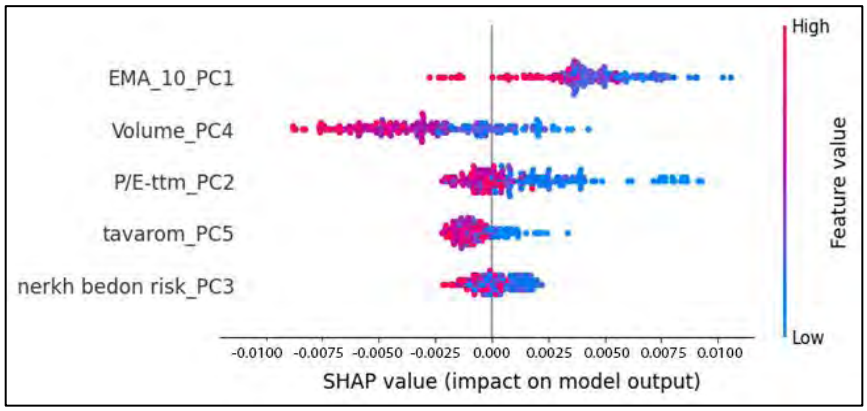

***شکل ٦.*** *یافته‌های حاصل از توضیح‌پذیری صنعت کاشی*

۵) **صنعت خودرو:** بر پایه شکل ۷، ویژگی میانگین متحرک نمایی ده روزه، بیشترین اثر مثبت را بر مدل دارد و افزایش مقدار آن با بهبود بازده پیش‌بینی‌شده همراه است. متغیر همگرایی و واگرایی میانگین متحرک، در مقابل، اثری منفی دارد و در سطوح بالاتر، موجب کاهش بازده می‌شود. سایر ویژگی‌ها از جمله حجم معاملات، حجم تعادلی و انحراف معیار سی روزه، عمدتاً اثری محدود و نزدیک به صفر داشته و نقش کم‌رنگی در عملکرد مدل ایفا می‌کنند.



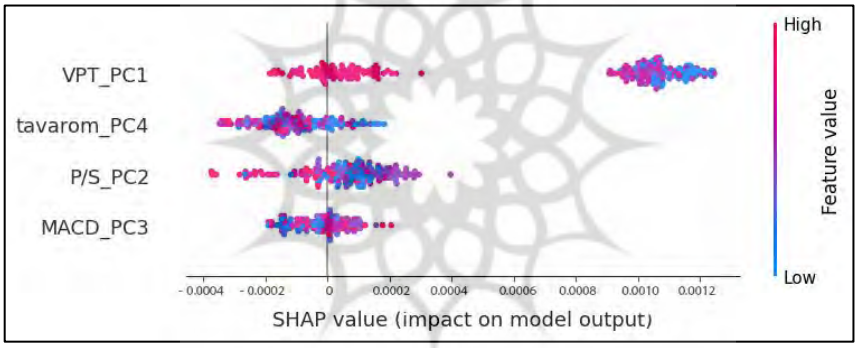

***شکل ۷.*** *یافته‌های حاصل از توضیح‌پذیری صنعت خودرو*

۶) **صنعت لاستیک:** بر اساس شکل ۸، ویژگی میانگین متحرک نمایی ده‌روزه، بیشترین اثر مثبت را بر مدل داشته و افزایش مقدار آن به بهبود خروجی منجر می‌شود. نسبت قیمت به سود دوازده ماهه نیز نقشی معنادار و عمدتاً مثبت دارد؛ هرچند آثار آن بسته به مقدار متغیر، تغییر می‌کند. حجم معاملات، اثری میانه و محدود دارد و نقش آن در پیش‌بینی، کم‌رنگ است. ویژگی تورم، تأثیر منفی بر مدل دارد و در سطوح بالاتر، موجب کاهش بازده پیش‌بینی‌شده می‌شود. در نهایت، نرخ بدون ریسک، نقشی بسیار اندک داشته و در مجموع، اثرگذاری آن، ناچیز ارزیابی می‌شود.



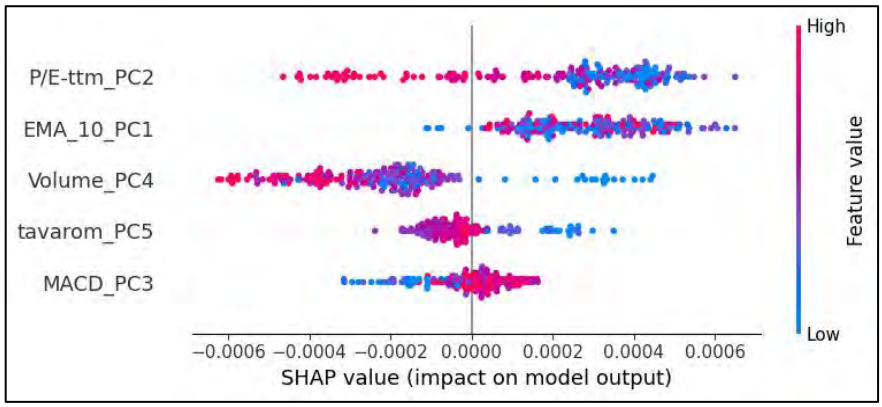

***شکل ۸.*** *یافته‌های حاصل از توضیح‌پذیری صنعت لاستیک*

۷) **صنعت شیمیایی:** بر پایه شکل ۹، متغیر حجم معاملات هم‌زمان با قیمت، بیشترین اثر مثبت را بر مدل داشته و در بیشتر موارد، موجب افزایش بازده پیش‌بینی‌شده می‌شود. متغیرهای تورم و نسبت قیمت به فروش، دارای آثار دگرگون‌پذیر هستند و بسته به مقدار خود، می‌توانند آثار تقویتی یا تضعیفی بر مدل داشته باشند. ویژگی همگرایی و واگرایی میانگین متحرک، کمترین نقش را در مدل ایفا کرده و آثار آن عمدتاً در نزدیکی صفر متمرکز است. در مجموع، حجم معاملات هم‌زمان با قیمت، اثرگذارترین متغیر در این مدل است؛ در حالی‌که سایر متغیرها، نقش محدود یا قابل تغییری دارند.




***شکل ۹.*** *یافته‌های حاصل از توضیح‌پذیری صنعت شیمیایی*

۸) **صنعت سیمان:** بر اساس شکل ۱۰، ویژگی‌های نسبت قیمت به سود دوازده ماهه و میانگین متحرک نمایی ده روزه، بیشترین اثر مثبت را بر مدل دارند و با افزایش مقدار آن‌ها، بازده پیش‌بینی‌شده، تقویت می‌شود. حجم معاملات دارای آثار متغیر، اما عمدتاً کاهنده است. متغیرهای تورم و همگرایی و واگرایی میانگین متحرک نقش محدودی در عملکرد مدل داشته و آثار آن‌ها بیشتر پیرامون صفر قرار دارد. در مجموع، دو ویژگی نخست از مهم‌ترین عوامل تقویت‌کننده مدل به شمار می‌روند، در حالی که سایر متغیرها تأثیر اندکی دارند یا آثارشان منفی است.




***شکل ۱۰.*** *یافته‌های حاصل از توضیح‌پذیری صنعت سیمان*

همچنین میانگین مقادیر شپ به ازای هر ویژگی در مدل‌سازی شاخص صنایع در جدول ۵ ارائه شده است:

***جدول ۵.*** *میانگین مقادیر شپ به ازای هر ویژگی در صنایع مختلف*

| | دارویی | سیمان | فولاد | لاستیک | کانی | کاشی | خودرو | شیمیایی |
|---|---|---|---|---|---|---|---|---|
| ***EMA*** | ۰/۰۰۰۹۱ | ۰/۰۰۰۴۵ | ۰/۰۰۰۲۳ | ۰/۰۰۵۱ | ۰/۰۰۰۲۷ | -۰/۰۰۰۲ | ۰/۰۰۱۵ | – |
| ***MACD*** | ۰/۰۰۰۴۵ | ۰/۰۰۰۱۱ | ۰/۰۰۰۱۱ | – | ۰/۰۰۰۲۵ | ۰/۰۰۰۳۱ | -۰/۰۰۰۱ | -۰/۰۰۰۳ |
| ***Volume*** | -۰/۰۰۰۳۱ | ۰/۰۰۰۳۴ | -۰/۰۰۰۱۸ | -۰/۰۰۲۵ | -۰/۰۰۰۳ | -۰/۰۰۰۵ | ۰/۰۰۰۱۰ | ۰/۰۰۰۴۵ |
| ***OBV*** | – | – | – | – | – | – | ۰/۰۰۰۵ | – |
| ***STd*** | – | – | – | – | – | – | ۰/۰۰۰۲ | – |
| ***Free Float*** | – | – | – | – | ۰/۰۰۰۶۱ | – | – | – |
| قیمت به سود | ۰/۰۰۰۲۱ | ۰/۰۰۰۵۵ | ۰/۰۰۰۳۱ | ۰/۰۰۲۱ | ۰/۰۰۰۲۱ | ۰/۰۰۰۲۱ | – | – |
| قیمت به فروش | – | – | – | – | – | – | – | ۰/۰۰۰۳۵ |
| نرخ بهره | – | – | – | ۰/۰۰۰۵ | – | – | – | – |
| نرخ تورم | – | -۰/۰۰۰۱ | – | ۰/۰۰۰۵۷ | – | – | – | ۰/۰۰۰۴۱ |

**ارزیابی متقاطع سری زمانی:** در ارزیابی مدل‌های پیش‌بینی سری زمانی، روش تقسیم متوالی یکی از رایج‌ترین روش‌ها است. در این روش، داده‌های قدیمی‌تر به عنوان مجموعه آموزشی و داده‌های جدیدتر به عنوان مجموعه آزمایشی در نظر گرفته می‌شوند. با هر تکرار، بخشی از داده‌های جدید به مجموعه آموزشی اضافه شده، مدل تحت آموزش قرار می‌گیرد و عملکرد آن بر مجموعه آزمایشی ارزیابی می‌گردد [۲۱]. این شیوه به‌ویژه در مسائل پیش‌بینی سری‌های زمانی، جایگزین کارآمدتری نسبت به جداسازی صلب داده‌های اعتبارسنجی است، زیرا از نشت اطلاعات آینده به مدل جلوگیری می‌کند و شرایط واقعی بازار را بهتر شبیه‌سازی می‌نماید.

برای سنجش پایداری مدل، میانگین دقت مدل در هر بخش و واریانس دقت پیش‌بینی در مراحل مختلف، محاسبه می‌شود. برابری میانگین دقت و واریانس پایین دقت بخش‌های مختلف، نشان‌دهنده تعمیم‌پذیری مطلوب مدل است. اگر واریانس نزدیک به صفر باشد، عملکرد مدل در تمامی بخش‌های داده، پایدار تلقی شده و دقت آن در مواجهه با داده‌های جدید قابل اطمینان خواهد بود. در مقابل، واریانس بالا نشان‌دهنده بی‌ثباتی مدل و نیاز به بازنگری در روش ارزیابی یا معماری مدل است. در نهایت، واریانس کمتر از ۱ درصد، در بسیاری از مسائل پیش‌بینی سری زمانی، قابل قبول است، به ویژه اگر میانگین دقت نیز بالا باشد [۵]. نتایج حاصل از ارزیابی متقاطع در بخش‌های مختلف در جدول ۶ قابل مشاهده است.

**جدول ٦**. *نتایج حاصل از ارزیابی متقاطع در بخش‌های مختلف*

| ردیف | صنعت | Fold | MSE | RMSE | MAE | MAPE |
|---|---|---|---|---|---|---|
| ۱ | دارویی | ۱ | ۰/۰۰۰۰۰۰۷۴۱ | ۰/۰۰۰۸۶۱ | ۰/۰۰۰۶۳۵ | ۷/۲۹۶۸۷۷ |
| | | ۲ | ۰/۰۰۰۰۰۰۸۵۴ | ۰/۰۰۰۹۲۴ | ۰/۰۰۰۷۰۶ | ۲/۸۸۷۲۱۶ |
| | | ۳ | ۰/۰۰۰۰۰۱۴۵ | ۰/۰۰۱۲۰۳ | ۰/۰۰۰۹۷۲ | ۱/۷۳۸۶۶۴ |
| | | ۴ | ۰/۰۰۰۰۰۲۶ | ۰/۰۰۱۶۱۳ | ۰/۰۰۱۳۰۸ | ۲/۳۰۰۱۰۴ |
| | | ۵ | ۰/۰۰۰۰۰۲۴۵ | ۰/۰۰۱۵۶۶ | ۰/۰۰۱۳۰۲ | ۶/۲۷۸۳۲۱ |
| | | Average | ۰/۰۰۰۰۰۱۶۲ | ۰/۰۰۱۲۳۳ | ۰/۰۰۰۹۸۵ | ۴/۱۰۰۲۳۶ |
| | | Std | ۰/۰۰۰۰۰۰۷۸ | ۰/۰۰۰۳۱۳ | ۰/۰۰۰۲۸۵ | ۲/۲۴۷۲۸۶ |
| ۲ | فولاد | ۱ | ۰/۰۰۰۰۰۰۶۷۷ | ۰/۰۰۰۸۲۳ | ۰/۰۰۰۵۳۶ | ۲/۸۲۳۲۳۱ |
| | | ۲ | ۰/۰۰۰۰۰۱۵۶ | ۰/۰۰۱۲۴۹ | ۰/۰۰۱۰۰۴ | ۳/۶۲۷۵۹۶ |
| | | ۳ | ۰/۰۰۰۰۰۴۸۷ | ۰/۰۰۲۲۰۷ | ۰/۰۰۱۸۶ | ۴/۷۹۲۰۵۶ |
| | | ۴ | ۰/۰۰۰۰۰۳۳۵ | ۰/۰۰۱۸۳ | ۰/۰۰۱۵۵۳ | ۱۰/۱۷۳۳۶ |
| | | ۵ | ۰/۰۰۰۰۰۲۵۸ | ۰/۰۰۱۶۰۶ | ۰/۰۰۱۲۲۱ | ۲/۲۳۰۹۱۲ |
| | | Average | ۰/۰۰۰۰۰۲۶۱ | ۰/۰۰۱۵۴۳ | ۰/۰۰۱۲۳۵ | ۴/۷۲۹۴۳۱ |
| | | Std | ۰/۰۰۰۰۰۱۴۵ | ۰/۰۰۰۴۷۶ | ۰/۰۰۰۴۵۵ | ۲/۸۵۴۱۳۸ |
| ۳ | کانی | ۱ | ۰/۰۰۰۰۰۱۳۲ | ۰/۰۰۱۱۴۹ | ۰/۰۰۰۹۰۳ | ۱۱/۰۴۲۱۵ |
| | | ۲ | ۰/۰۰۰۰۰۰۰۸ | ۰/۰۰۰۸۹۵ | ۰/۰۰۰۶۵۹ | ۲/۴۶۷۲۷۱ |
| | | ۳ | ۰/۰۰۰۰۰۴ | ۰/۰۰۲ | ۰/۰۰۱۶۳۷ | ۱/۰۲۶۴۹۸ |
| | | ۴ | ۰/۰۰۰۰۰۱۷۱ | ۰/۰۰۱۳۰۷ | ۰/۰۰۰۹۷۴ | ۲/۷۷۴۷۵ |
| | | ۵ | ۰/۰۰۰۰۰۳۹۶ | ۰/۰۰۱۹۹۱ | ۰/۰۰۱۴۴۲ | ۱/۴۲۱۳۳۶ |
| | | Average | ۰/۰۰۰۰۰۲۳۶ | ۰/۰۰۱۴۶۸ | ۰/۰۰۱۱۲۳ | ۳/۷۴۶۴۰۲ |
| | | Std | ۰/۰۰۰۰۰۱۳۶ | ۰/۰۰۰۴۵ | ۰/۰۰۰۳۶۱ | ۳/۷۰۴۳۷۸ |
| ۴ | کاشی | ۱ | ۰/۰۰۰۰۰۱۲۷ | ۰/۰۰۱۱۲۷ | ۰/۰۰۰۸۷۷ | ۱/۵۵۰۲۵۹ |
| | | ۲ | ۰/۰۰۰۰۰۱۹۸ | ۰/۰۰۱۴۰۷ | ۰/۰۰۰۹۵۵ | ۷/۴۷۲۰۴۳ |
| | | ۳ | ۰/۰۰۰۰۰۲۳۹ | ۰/۰۰۱۵۴۶ | ۰/۰۰۱۲۴۱ | ۲/۰۰۶۰۱۸ |
| | | ۴ | ۰/۰۰۰۰۰۴۷۳ | ۰/۰۰۲۱۷۴ | ۰/۰۰۱۷۶۴ | ۱/۸۹۰۰۲۳ |
| | | ۵ | ۰/۰۰۰۰۰۴۰۱ | ۰/۰۰۲۰۰۳ | ۰/۰۰۱۵۶۷ | ۱/۴۲۴۷۶۳ |
| | | Average | ۰/۰۰۰۰۰۲۸۸ | ۰/۰۰۱۶۵۱ | ۰/۰۰۱۲۸۱ | ۱/۴۹۵۳۰۷ |
| | | Std | ۰/۰۰۰۰۰۱۲۹ | ۰/۰۰۰۳۸۵ | ۰/۰۰۰۳۴۲ | ۲/۹۹۲۰۳۴ |
| ۵ | خودرو | ۱ | ۰/۰۰۰۰۰۳۴۵ | ۰/۰۰۱۸۵۹ | ۰/۰۰۱۴۷۶ | ۲/۶۱۴۲۴ |
| | | ۲ | ۰/۰۰۰۰۰۳۹۱ | ۰/۰۰۱۹۷۷ | ۰/۰۰۱۴۴۳ | ۲/۸۰۸۵۹۳ |
| | | ۳ | ۰/۰۰۰۰۱۱۷ | ۰/۰۰۳۴۱۹ | ۰/۰۰۲۶۶ | ۲/۲۳۵۲۵۳ |
| | | ۴ | ۰/۰۰۰۰۰۶۹۲ | ۰/۰۰۲۶۳ | ۰/۰۰۲۲۱۱ | ۵/۶۶۷۱۰۱ |
| | | ۵ | ۰/۰۰۰۰۰۶۹۳ | ۰/۰۰۲۶۳۲ | ۰/۰۰۲۱۷۴ | ۲/۵۶۷۱۴۵ |
| | | Average | ۰/۰۰۰۰۰۶۵۸ | ۰/۰۰۲۵۰۳ | ۰/۰۰۱۹۹۳ | ۳/۱۷۸۴۶۶ |
| | | Std | ۰/۰۰۰۰۰۲۹۴ | ۰/۰۰۰۵۵۹ | ۰/۰۰۰۴۶۸ | ۱/۲۵۷۹۲۱ |
| ۶ | لاستیک | ۱ | ۰/۰۰۰۰۰۱۹۳ | ۰/۰۰۱۳۹ | ۰/۰۰۱۰۶۵ | ۴/۰۴۹۵۳۷ |
| | | ۲ | ۰/۰۰۰۰۰۱۳۳ | ۰/۰۰۱۱۵۴ | ۰/۰۰۰۸۲۶ | ۲/۰۵۶۵۹۶ |
| | | ۳ | ۰/۰۰۰۰۰۲۵۶ | ۰/۰۰۱۵۹۹ | ۰/۰۰۱۲۷۶ | ۱/۵۱۶۸۲۹ |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | ۴ | ۰/۰۰۰۰۰۵۰۶ | ۰/۰۰۲۲۴۹ | ۰/۰۰۱۸۵۶ | ۱/۵۶۹۷۵۸ |
| | | ۵ | ۰/۰۰۰۰۰۴۲۱ | ۰/۰۰۲۰۵۲ | ۰/۰۰۱۶۳۸ | ۲/۰۸۱۴۶۶ |
| | | Average | ۰/۰۰۰۰۰۳۰۲ | ۰/۰۰۱۶۸۹ | ۰/۰۰۱۳۳۲ | ۲/۲۵۴۸۳۷ |
| | | Std | ۰/۰۰۰۰۰۰۱۴ | ۰/۰۰۰۴۰۷ | ۰/۰۰۰۳۷۴ | ۰/۹۲۷۸۲۵ |
| | | ۱ | ۰/۰۰۰۰۰۰۶۴۲ | ۰/۰۰۰۸۰۱ | ۰/۰۰۰۶ | ۱۲/۵۹۳۵۶ |
| | | ۲ | ۰/۰۰۰۰۰۰۸۸۷ | ۰/۰۰۰۹۴۲ | ۰/۰۰۰۷۳۸ | ۲/۷۳۱۳۱۲ |
| | | ۳ | ۰/۰۰۰۰۰۳۴۲ | ۰/۰۰۱۸۵ | ۰/۰۰۱۵۰۳ | ۳/۰۳۹۷۸۶ |
| ۷ | شیمیایی | ۴ | ۰/۰۰۰۰۰۱۱۴ | ۰/۰۰۱۰۶۸ | ۰/۰۰۰۸۱۵ | ۱/۴۱۸۷۹۳ |
| | | ۵ | ۰/۰۰۰۰۰۳۰۱ | ۰/۰۰۱۷۳۵ | ۰/۰۰۱۳۵۱ | ۲/۸۵۴۲۳۸ |
| | | Average | ۰/۰۰۰۰۰۱۸۲ | ۰/۰۰۱۲۷۹ | ۰/۰۰۱۰۰۲ | ۴/۵۲۷۵۳۸ |
| | | Std | ۰/۰۰۰۰۰۱۱۶ | ۰/۰۰۰۴۲۹ | ۰/۰۰۰۳۵۸ | ۴/۰۷۳۴۴۵ |
| | | ۱ | ۰/۰۰۰۰۰۹۹۸ | ۰/۰۰۰۹۹۹ | ۰/۰۰۰۰۷۶ | ۵/۸۳۸۵۲۲ |
| | | ۲ | ۰/۰۰۰۰۰۱۳۶ | ۰/۰۰۱۱۶۷ | ۰/۰۰۰۸۴۴ | ۳/۳۰۵۵۱۸ |
| | | ۳ | ۰/۰۰۰۰۰۲۵۷ | ۰/۰۰۱۶۰۳ | ۰/۰۰۱۲۴۱ | ۳/۰۷۹۷۶۶ |
| ۸ | سیمان | ۴ | ۰/۰۰۰۰۰۲۳۲ | ۰/۰۰۱۵۱۷ | ۰/۰۰۱۲۴۵ | ۱/۴۹۶۸۰۹ |
| | | ۵ | ۰/۰۰۰۰۰۲۳۱ | ۰/۰۰۱۵۲۶ | ۰/۰۰۱۱۹۷ | ۱/۹۵۱۷۵۱ |
| | | Average | ۰/۰۰۰۰۰۱۹۱ | ۰/۰۰۱۳۶۱ | ۰/۰۰۱۰۵۹ | ۳/۱۳۴۴۷۳ |
| | | Std | ۰/۰۰۰۰۰۰۶۱۴ | ۰/۰۰۰۲۳۵ | ۰/۰۰۰۲۰۹ | ۱/۵۱۱۶۱۶ |

طبق جدول ۶، نتایج حاصل از پیاده‌سازی ارزیابی متقاطع، حاکی از آن است که میانگین دقت در تمامی بازه‌های زمانی با دقت مدل مطابقت دارد. بنابراین (Bias = 0) است و انحراف معیار دقت مدل در این بازه‌ها نزدیک به صفر بنابراین (Variance = 0) است. این امر نشان‌دهنده ثبات نتایج و دقت مدل در طول زمان است که از طریق پیاده‌سازی ارزیابی متقاطع، تأیید می‌شود. علاوه بر این، نتایج حاصل نشان می‌دهد که مشارکت ویژگی‌های ورودی مدل در این بازه‌ها، ثابت باقی می‌ماند و تأثیری در پیش‌بینی‌های مدل در بازه‌های زمانی مختلف ندارد.

## ۵. بحث و نتیجه‌گیری

این پژوهش، نخستین تلاش علمی در ایران است که به بررسی قابلیت توضیح‌پذیری هوش مصنوعی در پیش‌بینی شاخص‌های مالی پرداخته و بدین ترتیب، نقطه عطفی در مطالعات مرتبط با کاربرد مدل‌های یادگیری ماشین در بازار سرمایه محسوب می‌شود. مطالعات پیشین عمدتاً بر بهبود دقت پیش‌بینی در مدل‌های یادگیری ماشین تمرکز داشته‌اند، اما غالباً، عدم شفافیت و تفسیرپذیری این مدل‌ها را نادیده گرفته‌اند. از این‌رو، این پژوهش با ترکیب الگوریتم جنگل تصادفی، بهینه سازی هایپرپارامترها از طریق الگوریتم ژنتیک و توضیح‌پذیری مدل با مقادیر شپ و ارزیابی نتایج در ادوار زمانی به‌طور هم‌زمان دو هدف کلیدی را دنبال کرده است: ارتقای شفافیت مدل‌های یادگیری ماشینی با دقت بالا و تعمیم‌پذیری زمانی آنها.

نتایج این پژوهش با مطالعات معتبر بین‌المللی تطابق دارد. به عنوان مثال، مطالعه‌ای توسط چن و همکاران (۲۰۲۲) نشان داد که مدل‌های مبتنی بر یادگیری ماشین مانند جنگل تصادفی و تقویت گردایان شدید، به دلیل توانایی بالا در مدل‌سازی روابط پیچیده، عملکرد بهتری نسبت به مدل‌های سنتی اقتصادسنجی دارند [۹]. با این‌حال، لاندبرگ و همکاران (۲۰۲۰) بر اهمیت روش‌های توضیح‌پذیری مانند شاپ تأکید کرده‌اند و نشان داده‌اند که مدل‌های توضیح‌پذیر نه‌تنها دقت را افزایش می‌دهند، بلکه اعتماد سرمایه‌گذاران و سیاست‌گذاران را نیز جلب می‌کنند [۲۴]. این نتایج با

یافته‌های پژوهش حاضر مطابقت دارد؛ زیرا مدل پیشنهادی این پژوهش نیز نشان داد که با افزایش شفافیت مدل، تحلیل‌گران مالی قادر خواهند بود تصمیمات آگاهانه‌تری اتخاذ کنند. بررسی متغیرهای ورودی نشان داد که متغیرهای تکنیکال، تأثیر بیشتری بر دقت پیش‌بینی دارند. اندیکاتورهایی مانند میانگین متحرک نمایی، شاخص همگرایی و واگرایی میانگین متحرک و حجم معاملات و میزان سهام شناور به عنوان مهم‌ترین متغیرهای مؤثر در پیش‌بینی، شناخته شدند. این نتایج هم‌راستا با پژوهش‌های بین‌المللی مانند هارل و هارپاز[1] (۲۰۲۱) است که اهمیت متغیرهای تکنیکال را در پیش‌بینی بازده بازارهای مالی تأیید کرده‌اند [۱۹]. درحالی‌که برخی مطالعات مانند پاتل و همکاران[2] (۲۰۱۵) بر نقش متغیرهای بنیادی تأکید دارند، نتایج این پژوهش نشان داد که اگرچه متغیرهایی مانند نسبت قیمت به درآمد، نرخ تورم و نرخ بهره بدون ریسک تأثیرگذار هستند، اما نقش آن‌ها نسبت به متغیرهای تکنیکال کمتر است. علاوه بر این، ارزیابی متقاطع سری زمانی نشان داد که مدل پیشنهادی دارای پایداری و تعمیم‌پذیری بالایی در بازه‌های زمانی مختلف است [۲۹]. این مسئله از نظر عملی حائز اهمیت است، زیرا بسیاری از مدل‌های پیش‌بینی، پس از مدتی دچار افت عملکرد می‌شوند. اما در این تحقیق مشخص شد که با انتخاب ویژگی‌های مناسب و بهینه‌سازی هایپرپارامترها، می‌توان مدلی پایدار برای تصمیم‌گیری‌های مالی طراحی کرد. بر اساس یافته‌های این پژوهش، مدل پیشنهادی نه‌تنها یک ابزار پیشرفته برای پیش‌بینی شاخص‌های مالی است، بلکه با ویژگی توضیح‌پذیری خود، می‌تواند به تحلیلگران و مدیران مالی کمک کند تا تصمیمات آگاهانه‌تری بگیرند.

بر اساس نتایج حاصل از این پژوهش، افق‌های متعددی برای تحقیقات آتی گشوده می‌شود که می‌توانند به ارتقای دقت مدل‌های پیش‌بینی بازارهای مالی و افزایش شفافیت در فرآیندهای تصمیم‌گیری منجر شوند. از جمله حوزه‌های کلیدی پیشنهادی، توسعه مدل‌هایی است که با بهره‌گیری از تحلیل اخبار، به بررسی و پیش‌بینی روندهای روزانه بازار سهام می‌پردازند. در این زمینه، ترکیب تکنیک‌های پردازش زبان طبیعی[3] و یادگیری عمیق، امکان تحلیل دقیق‌تر تأثیر اخبار اقتصادی، سیاسی و مالی را بر رفتار بازار فراهم می‌سازد و زمینه‌ساز بهبود عملکرد مدل‌های پیش‌بینی مبتنی بر داده‌های متنی خواهد بود. علاوه بر این، مطالعه نقش نهادهای ناظر و چارچوب‌های مقرراتی بازار سرمایه در شناسایی پدیده‌هایی نظیر دستکاری قیمت سهام، به‌ویژه از طریق به‌کارگیری رویکردهای توضیح‌پذیر در حوزه هوش مصنوعی، از اهمیت بالایی برخوردار است. بهره‌گیری از قابلیت توضیح‌پذیری مدل‌ها همچنین می‌تواند در سایر زمینه‌های مرتبط مانند مدیریت ریسک سرمایه‌گذاری، بهینه‌سازی پرتفوی و طراحی استراتژی‌های معاملاتی، به‌عنوان مسیرهای ارزشمند برای تحقیقات آینده مدنظر قرار گیرد.

یکی از محدودیت‌های اصلی این پژوهش، عدم دسترسی به مجموعه کاملی از داده‌های مرتبط با قیمت سهام است. در تحقیق حاضر، تنها قیمت پایانی شاخص‌ها در دسترس بوده است که منجر به محدود شدن تحلیل‌ها به این بُعد از داده‌ها شده است. به‌طور خاص، برخی از اندیکاتورهای تکنیکال کلیدی که نیازمند داده‌هایی نظیر قیمت باز، کمترین و بیشترین قیمت روزانه هستند، نمی‌توانستند در تحلیل‌ها مورد استفاده قرار گیرند. یکی دیگر از محدودیت‌ها، عدم دسترسی به اطلاعات دقیق رفتار خریداران و فروشندگان در صنایع است که می‌تواند تحلیل روانشناسی بازار و نقدشوندگی سهام را بهبود بخشد. همچنین، فقدان داده‌های مربوط به بحران‌های مالی یا تغییرات عمده اقتصادی

[1] Harel & Harpaz
[2] Patel et al.
[3] Natural Language Processing (NLP)

همچون تحریم می‌تواند بر دقت پیش‌بینی‌ها در شرایط نوسانی تأثیر بگذارد و استفاده از آن‌ها در تحقیقات آینده، می‌تواند به بهبود عملکرد مدل‌ها و شناخت بهتر آن‌ها کمک کند.

**تشکر و قدردانی**

از استاد برجسته، جناب آقای دکتر شاپور محمدی که در مسیر اجرای این پژوهش، نویسندگان را یاری نمودند، تشکر و قدردانی می‌شود.

## Reference

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI*). IEEE Access*, 6, 52138–52160.
2. Ammann, M., Coqueret, G., & Schade, F. (2022). Machine learning in asset pricing. *Journal of Financial Economics*, 144(2), 424-453.
3. Arsenault, P., Wang, S., & Patenande, J. (2024). A Survey of Explainable Artificial Intelligence (XAI) in Financial Time Series Forecasting. *ArXiv*, abs/2407.15909.
4. Babaei, G., Giudici, P., & Raffinetti, E. (2022). Explainable artificial intelligence for crypto asset allocation. *Finance Research Letters*, 47, 102941.
5. Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83.
6. Box, G. E. P. (2013). Box and Jenkins: Time series analysis, forecasting and control. In A Very British affair: Six Britons and the development of time series analysis during the 20th century (pp. 161–215). Palgrave Macmillan.
7. Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average, *The Review of Financial Studies*, 21(4), 1509-1531.
8. Cartea, Á., Jaimungal, S., & Penalva, J. (2015). *Algorithmic and high-frequency trading*. Cambridge University Press.
9. Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, 365, 112395.
10. Deng, S., Huang, X., Zhu, Y., Su, Z., Fu, Z., & Shimada, T. (2023). Stock index direction forecasting using an explainable eXtreme Gradient Boosting and investor sentiments. *The North American Journal of Economics and Finance*, 64, 101848.
11. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv*:1702.08608.
12. Eghtesad, A. and Mohammadi,. (2023). Portfolio optimization with return prediction using LSTM, Random forest, and ARIMA. *Financial Management Perspective*, 13(43), 9-28. (in Persian)
13. Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 427–465.

14. 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.
15. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *In 2018 IEEE 5th International Conference on Data Science and Advanced Analytics* (DSAA) (pp. 80–89). IEEE.
16. Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*, 32, 100577.
17. Gholami, N. and Shams Gharne, N. (2024). Presenting an Optimized CNN-LSTM Model for Stock Price Forecasting in the Tehran Stock Exchange. *Financial Management Perspective*, 14(45), 123-147. (in Persian)
18. Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273.
19. Harel, A., & Harpaz, G. (2021). Forecasting stock prices. *International Review of Economics & Finance*, 73, 249–256.
20. Heidari, M. and Amiri, H. (2022). Inspecting the Predictive Power of Artificial Intelligence Models in Predicting the Stock Price Trend in Tehran Stock Exchange. *Financial Research Journal*, 24(4), 602-623. (in Persian)
21. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
22. Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36-43.
23. Liu, Y., Li, Z., Nekhili, R., & Sultan, J. (2023). Forecasting cryptocurrency returns with machine learning. *Research in International Business and Finance*, 64, 101905.
24. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56-67.
25. Moayedfar, S., Mohebbi, H., Mozaffaree Pour, N., & Sharifi, A. (2025). Developing a localized resilience assessment framework for historical districts: A case study of Yazd, Iran. *PLoS One*, 20(2), e0317088.
26. Molnar, C. (2022). *Interpretable machine learning*. Independently published.
27. Nallakaruppan, M., Chaturvedi, H., Grover, V., Balusamy, B., Jaraut, P., Bahadur, J., Meena, V., & Hameed, I. (2024). Credit Risk Assessment and Financial Decision Support Using Explainable Artificial Intelligence. *Risks*.
28. Orte, F., Mira, J., Sánchez, M. J., & Solana, P. (2023). A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction. *Research in International Business and Finance*, 64, 101829.
29. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index fusing fusion of machine learning techniques. *Expert Systems with Applications*, 42(4), 2162–2172.

30. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005-2019. *Applied Soft Computing*, 90, 106181.
31. Strumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3), 647-665.
32. West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, 57, 47-66.
33. Zhang, C. A., Cho, S., & Vasarhelyi, M. (2022). Explainable artificial intelligence (XAI) in auditing. *International Journal of Accounting Information Systems*, 46, 100572.