


Providing a Portfolio Optimization Model Based on Prediction Considering Investment Sentiment

1. Ali Ghasemi Kian*: Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran. Email: alikian7777777@gmail.com (Corresponding Author)

2. Amirabbas Najafi: Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran

Article history



Received: 22 December 2025

Revised: 04 May 2026

Accepted: 11 May 2026

Initial Publish: 19 June 2026

Final Publish: 22 June 2027

Abstract:

The purpose of this study was to develop an integrated model for stock return prediction and portfolio optimization by combining historical market data, technical indicators, and investor sentiment extracted from social media in order to improve investment decision-making accuracy in financial markets. This study was conducted using daily data from 16 selected U.S. stock market symbols during the period from September 30, 2021 to September 29, 2022. Price data, trading volume, and social media textual information were collected and preprocessed into 105 features, including 92 technical features and 13 sentiment-related features. Binary Grey Wolf Optimizer (BGWO) was employed for feature selection, while the XGBoost algorithm was used for stock return prediction. Subsequently, the predicted 31-day returns were incorporated into the Markowitz mean–variance optimization framework. Monte Carlo simulation with 100,000 admissible weight vectors was used to determine the optimal portfolio composition. Model performance was evaluated using MSE, RMSE, MAE, MAPE, and R^2 indices. The findings demonstrated that integrating technical indicators with investor sentiment features significantly improved prediction performance. The highest R^2 values were observed for KO and COST, with values of 0.8945 and 0.8915, respectively, whereas TSLA and BX showed comparatively weaker predictive performance. The BGWO–XGBoost framework successfully selected informative features and generated a stable low-redundancy predictive structure. In the portfolio optimization stage, only symbols with R^2 values above 0.70 were included in the Markowitz framework. Simulation results revealed that the optimal portfolio achieved an annual Sharpe ratio of 3.1585, an expected return of 0.7956, and volatility of 0.2424. The highest portfolio weights were assigned to CRM, PG, KO, and AAPL, indicating their superior risk–return balance. The results indicated that integrating machine learning algorithms with investor sentiment analysis and classical financial optimization frameworks can substantially enhance stock return prediction accuracy and investment decision quality. The proposed model successfully reduced the gap between theoretical financial analysis and real market behavior by generating a balanced, executable, and high-performing portfolio in terms of risk and return. These findings highlight the strong potential of behavioral data and text analytics in the development of intelligent investment management systems.

Keywords: Portfolio Optimization, Machine Learning, Investor Sentiment Analysis, Grey Wolf Optimizer, XGBoost, Markowitz Model

Citation: Ghasemi Kian, A. & Najafi, A. (2027). Providing a Portfolio Optimization Model Based on Prediction Considering Investment Sentiment. *Accounting, Finance and Computational Intelligence*, 5(2), 1-23.



Copyright: © 2027 by the authors. Published under the terms and conditions of Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Extended Abstract**Introduction**

Financial markets have become increasingly complex due to rapid information diffusion, behavioral fluctuations among investors, and the growing influence of digital communication platforms. Traditional financial theories, which assume rational decision-making and efficient markets, have faced substantial criticism because they often fail to explain abnormal volatility, speculative bubbles, and sudden market reactions. Behavioral finance emerged as an alternative perspective emphasizing that investors are influenced by emotions, cognitive biases, and psychological reactions when making financial decisions (Pompian, 2009; Raei & Fallahpour, 2004). This perspective argues that investor sentiment can significantly affect asset prices, trading behavior, and market dynamics, thereby creating deviations from intrinsic values (Eslami Bidgoli & Kordlouie, 2010). Empirical evidence has demonstrated that emotions such as fear, anger, optimism, and anxiety alter risk-taking behavior and influence investment choices (Fessler et al., 2004; Gambetti & Giusberti, 2012). Consequently, investor sentiment has become an essential factor in explaining stock market movements and portfolio performance.

The rapid development of artificial intelligence and machine learning technologies has transformed financial forecasting methodologies. Machine learning algorithms are capable of identifying nonlinear relationships and hidden structures within large-scale financial datasets, outperforming many conventional econometric approaches in predictive accuracy (Lin & Marques, 2024). Models such as artificial neural networks, gradient boosting algorithms, and deep learning frameworks have been extensively used in stock market prediction (Qiu & Song, 2016). In recent years, hybrid models integrating optimization algorithms with machine learning techniques have demonstrated particularly promising results. For instance, the GA-XGBoost framework achieved high predictive performance in forecasting stock price direction by optimizing model parameters and reducing prediction error (Yun et al., 2021). Similarly, comprehensive forecasting frameworks incorporating technical, economic, and structural indicators have improved prediction reliability in financial markets (Asghari et al., 2024; Rezaeian et al., 2024).

At the same time, the expansion of social media and online investment communities has created unprecedented opportunities for extracting behavioral information from textual data. Sentiment analysis techniques based on natural language processing enable researchers to quantify investor emotions from news articles, tweets, and online discussions. These approaches have become increasingly important because social media reflects collective expectations, optimism, fear, and market reactions in real time (Salehi Chegini et al., 2025). Studies have shown that combining textual sentiment indicators with historical financial data significantly improves forecasting performance (Gu et al., 2024). Models integrating FinBERT and LSTM architectures have successfully captured market sentiment from financial news and enhanced stock price prediction accuracy (Gu et al., 2024). Furthermore, social media sentiment analysis has been proven effective in forecasting stock market volatility and identifying behavioral fluctuations among investors (Saravanos & Kanavos, 2025).

Research in emerging markets has also highlighted the importance of investor sentiment. In the Tehran Stock Exchange, investor sentiment has been found to affect market liquidity, volatility, and trading behavior (Aghababaei, 2022). Behavioral reactions among investors influence market dynamics and contribute to fluctuations in trading activity (Hosseini & Morshedi, 2019). Moreover, investor sentiment combined with governmental behavior has been identified as a major source of market instability and uncertainty (Dadgar et al., 2023). Studies have also emphasized the relationship between corporate social responsibility, environmental awareness, and investor perceptions in shaping financial decisions (Rezaei-Kalidbari et al., 2013;

Zare Bahman-Miri et al., 2022). These findings suggest that sentiment-related information can provide valuable insights beyond traditional market indicators.

Another major challenge in financial prediction involves high-dimensional feature spaces. Financial datasets often contain redundant or irrelevant variables that reduce model efficiency and increase the risk of overfitting. Therefore, feature selection has become a critical step in machine learning-based forecasting models (Li et al., 2017; Miao & Niu, 2016). Optimization-based feature selection methods, including evolutionary and metaheuristic algorithms, have demonstrated strong capabilities in reducing dimensionality while preserving predictive information (Ramirez-Gallego et al., 2017; Yu & Liu, 2003). Among these approaches, the Grey Wolf Optimizer has shown remarkable effectiveness in feature selection and optimization tasks (Emary et al., 2016). Enhanced versions of this algorithm have further improved exploration and exploitation balance in complex search spaces (Liu et al., 2023). Integrating such optimization algorithms with machine learning frameworks can substantially improve forecasting performance and portfolio construction efficiency.

Although previous studies have separately examined stock prediction, sentiment analysis, and portfolio optimization, limited research has integrated all these dimensions into a unified framework. Many forecasting studies focus solely on prediction accuracy without connecting outputs to portfolio decision-making. Likewise, several sentiment analysis studies consider behavioral indicators independently from technical and optimization frameworks. Therefore, the present study attempts to bridge this gap by proposing an integrated portfolio optimization model that combines investor sentiment analysis, machine learning prediction, and classical mean–variance optimization.

Methods and Materials

This study employed daily financial and textual data from 16 selected U.S. stock market symbols during the period from September 30, 2021 to September 29, 2022. The selected companies included Apple, Microsoft, Amazon, Tesla, Salesforce, Coca-Cola, Procter & Gamble, Boeing, and several other major firms representing technology, industrial, and consumer sectors. Daily OHLCV data consisting of open, high, low, close, and volume values were collected and synchronized with the official trading calendar.

A comprehensive preprocessing procedure was applied to ensure data consistency and stability. Financial records were cleaned, normalized, and aligned temporally. Based on these data, a large set of technical indicators was generated, including moving averages, momentum oscillators, volatility indicators, volume-based features, logarithmic returns, and lagged price ratios. Simultaneously, textual data extracted from social media platforms were processed through sentiment analysis techniques. Tweets related to each stock symbol were cleaned, filtered, and aggregated daily. Sentiment features included average sentiment score, median sentiment, sentiment variance, positive-to-negative ratio, tweet counts, and sentiment class averages.

In total, 105 features were constructed for each observation, including 92 technical indicators and 13 sentiment-related variables. Feature selection was conducted using the Binary Grey Wolf Optimizer (BGWO). The optimization process employed a population size of 20, feature dimension of 105, and 50 iterations. After selecting optimal features, the XGBoost algorithm was trained to predict 31-day future stock returns. Hyperparameter tuning included adjustments to tree depth, learning rate, subsampling ratio, column sampling ratio, regularization parameters, and number of estimators.

The forecasting outputs were then integrated into the Markowitz mean–variance optimization framework. Only symbols with test-set R^2 values greater than 0.70 were selected for portfolio construction. Monte Carlo simulation with 100,000

admissible weight vectors was employed to estimate efficient portfolios under standard constraints. Portfolio performance was evaluated using expected return, annualized volatility, and Sharpe ratio with a 3% annual risk-free rate.

Findings

The results demonstrated that integrating technical indicators with investor sentiment significantly improved forecasting performance. Among the examined symbols, Coca-Cola (KO) and Costco (COST) achieved the highest prediction accuracies with R^2 values of approximately 0.89, indicating strong explanatory power of the BGWO-XGBoost framework. In contrast, more volatile stocks such as Tesla (TSLA) and Blackstone (BX) produced comparatively lower predictive performance. Error metrics including MSE, RMSE, MAE, and MAPE confirmed that the proposed model maintained relatively stable prediction accuracy across multiple symbols.

The feature selection process substantially reduced redundancy within the dataset. Depending on stock characteristics, the optimization framework selected between 29 and 41 effective features from the original 105 variables. The selected features typically included combinations of trend indicators, momentum variables, volatility measures, volume-based signals, and sentiment indicators. Sentiment-related variables such as average sentiment score, positive-to-negative tweet ratio, and social interaction intensity emerged as influential predictors for several stocks.

The hyperparameter optimization procedure produced distinct configurations for different stock symbols. Some stocks required deeper tree structures and stronger regularization, while others achieved optimal performance using simpler parameter settings. These findings reflected the heterogeneous and nonlinear nature of stock market behavior across industries and market conditions.

In the portfolio optimization stage, 11 symbols with R^2 values above 0.70 were selected as candidates for portfolio construction. Among them, nine stocks with positive predicted average returns entered the Markowitz optimization framework. Monte Carlo simulation generated the efficient frontier and identified the optimal portfolio with the highest Sharpe ratio. The final portfolio achieved an annualized Sharpe ratio of approximately 3.16, expected annual return of nearly 79.56%, and annualized volatility of around 24.24%.

The optimal portfolio weights concentrated primarily on Salesforce (CRM), Procter & Gamble (PG), Coca-Cola (KO), and Apple (AAPL), while smaller allocations were assigned to Microsoft, Amazon, Ford, AMD, and Boeing. The results indicated that combining machine learning prediction with sentiment analysis and classical optimization could produce a balanced and executable portfolio with strong risk-return characteristics.

Discussion and Conclusion

The findings demonstrated that integrating investor sentiment with technical and historical market data substantially improves stock return prediction and portfolio optimization performance. The proposed BGWO-XGBoost framework successfully captured nonlinear relationships among financial and behavioral variables while reducing dimensionality through optimized feature selection. The incorporation of sentiment indicators provided additional explanatory power beyond traditional market variables and enhanced the model's ability to identify short-term behavioral fluctuations.

The study also showed that social media sentiment can serve as a valuable information source for financial decision-making. Investor emotions reflected in online discussions contributed meaningfully to forecasting future stock movements, confirming the growing importance of behavioral signals in modern financial markets. These results suggest that financial

markets are increasingly influenced not only by economic fundamentals but also by collective psychological reactions and information diffusion across digital platforms.

From a methodological perspective, the study highlighted the effectiveness of combining metaheuristic optimization algorithms with machine learning frameworks in financial forecasting. The Grey Wolf Optimizer improved feature selection efficiency and reduced redundancy, while XGBoost provided strong predictive capability in complex nonlinear environments. The integration of these methods with mean–variance portfolio optimization created a unified decision-making framework capable of translating predictive outputs into practical investment strategies.

The portfolio optimization results further demonstrated that prediction-driven investment strategies can significantly improve portfolio efficiency. The optimized portfolio achieved a high Sharpe ratio while maintaining acceptable volatility levels, indicating a favorable balance between return and risk. This finding suggests that combining artificial intelligence techniques with classical portfolio theory can bridge the gap between theoretical financial models and real market behavior.

Overall, the proposed framework provides evidence that machine learning, sentiment analysis, and behavioral finance can collectively contribute to more intelligent and adaptive investment systems. As financial markets continue to evolve under the influence of digital communication and behavioral dynamics, integrated predictive models are likely to become increasingly important for portfolio management, risk assessment, and strategic investment decision-making.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

ارائه مدل بهینه‌سازی پرتفوی مبتنی بر پیش‌بینی با در نظرگیری احساسات سرمایه‌گذاری

تاریخچه مقاله
تاریخ دریافت: ۱ دی ۱۴۰۴
تاریخ بازنگری: ۱۴ اردیبهشت ۱۴۰۵
تاریخ پذیرش: ۲۱ اردیبهشت ۱۴۰۵
تاریخ چاپ اولیه: ۲۹ خرداد ۱۴۰۵
تاریخ چاپ نهایی: ۱ تیر ۱۴۰۶

۱. علی قاسمی کیان* (ORCID)، گروه مهندسی صنایع، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران. ایمیل: alikian777777@gmail.com (نویسنده مسئول)

۲. امیرعباس نجفی (ORCID)، گروه مهندسی صنایع، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران

چکیده

هدف این پژوهش، ارائه یک مدل یکپارچه برای پیش‌بینی بازده سهام و بهینه‌سازی پرتفوی با استفاده از ترکیب داده‌های تاریخی بازار، شاخص‌های تکنیکال و احساسات سرمایه‌گذاران استخراج‌شده از شبکه‌های اجتماعی بود تا دقت تصمیم‌گیری سرمایه‌گذاری در بازارهای مالی بهبود یابد. این پژوهش با استفاده از داده‌های روزانه ۱۶ نماد منتخب بورس آمریکا در بازه زمانی ۳۰ سپتامبر ۲۰۲۱ تا ۲۹ سپتامبر ۲۰۲۲ انجام شد. داده‌های قیمت، حجم معاملات و اطلاعات متنی مرتبط با شبکه‌های اجتماعی گردآوری و پس از پیش‌پردازش، در قالب ۱۰۵ ویژگی شامل ۹۲ ویژگی فنی و ۱۳ ویژگی احساسی سازمان‌دهی شدند. برای انتخاب ویژگی‌های بهینه از الگوریتم گرگ خاکستری دودویی (BGWO) و برای مدل‌سازی و پیش‌بینی بازده سهام از الگوریتم XGBoost استفاده شد. در ادامه، بازده پیش‌بینی شده ۳۱ روزه به‌عنوان ورودی مدل میانگین-واریانس مارکوویتز به کار گرفته شد و با استفاده از شبیه‌سازی مونت کارلو و تولید ۱۰۰ هزار بردار وزنی، ترکیب بهینه پرتفوی تعیین گردید. شاخص‌های ارزیابی شامل MSE، RMSE، MAE، MAPE و R^2 برای سنجش عملکرد مدل استفاده شدند. نتایج نشان داد که ترکیب هم‌زمان ویژگی‌های فنی و شاخص‌های احساسات سرمایه‌گذاران موجب بهبود معنادار عملکرد پیش‌بینی مدل شد. بیشترین مقدار ضریب تعیین مربوط به نمادهای KO و COST به ترتیب برابر با ۰.۸۹۴۵ و ۰.۸۹۱۵ بود، در حالی که نمادهای TSLA و BX عملکرد ضعیف‌تری نشان دادند. همچنین، مدل BGWO-XGBoost توانست با انتخاب ویژگی‌های مؤثر، ساختاری کم‌افزونگی و پایدار برای پیش‌بینی ایجاد کند. در مرحله بهینه‌سازی پرتفوی، تنها نمادهایی با مقدار R^2 بیشتر از ۰.۷۰ وارد مدل مارکوویتز شدند. نتایج شبیه‌سازی نشان داد که پرتفوی نهایی دارای نسبت شارپ سالانه ۳.۱۵۸۵، بازده موردانتظار ۰.۷۹۵۶ و نوسان ۰.۲۴۲۴ بود. بیشترین وزن‌های پرتفوی به نمادهای CRM، PG، KO و AAPL اختصاص یافت که بیانگر برتری این نمادها در توازن میان ریسک و بازده بود. نتایج پژوهش نشان داد که ادغام الگوریتم‌های یادگیری ماشین با تحلیل احساسات سرمایه‌گذاران و چارچوب‌های کلاسیک بهینه‌سازی مالی، می‌تواند دقت پیش‌بینی بازده سهام و کیفیت تصمیم‌گیری سرمایه‌گذاری را به‌طور قابل‌توجهی افزایش دهد. مدل پیشنهادی توانست ضمن کاهش فاصله میان تحلیل‌های نظری و رفتار واقعی بازار، پرتفوی متوازن، اجرایی و دارای عملکرد مطلوب از نظر ریسک-بازده ارائه کند. این یافته‌ها بیانگر ظرفیت بالای استفاده از داده‌های رفتاری و تحلیل متن در توسعه سیستم‌های هوشمند مدیریت سرمایه‌گذاری است.

کلیدواژه‌گان: بهینه‌سازی پرتفوی، یادگیری ماشین، تحلیل احساسات سرمایه‌گذاران، الگوریتم گرگ خاکستری، XGBoost، مدل مارکوویتز

شبهه استناددهی: قاسمی کیان، علی، و نجفی، امیرعباس. (۱۴۰۶). ارائه مدل بهینه‌سازی پرتفوی مبتنی بر پیش‌بینی با در نظرگیری احساسات سرمایه‌گذاری. *حسابداری، امور مالی و هوش محاسباتی*، ۵(۲)، ۱-۲۳.



بازارهای مالی در دهه‌های اخیر به دلیل افزایش حجم داده‌ها، پیچیدگی رفتار سرمایه‌گذاران و سرعت بالای گردش اطلاعات، وارد مرحله‌ای شده‌اند که تحلیل‌های سنتی مبتنی بر روابط خطی و الگوهای کلاسیک دیگر پاسخگوی نیازهای سرمایه‌گذاران و مدیران دارایی نیستند. نوسانات شدید بازار، گسترش شبکه‌های اجتماعی، تغییرات سریع در انتظارات سرمایه‌گذاران و تأثیر عوامل روان‌شناختی بر رفتار معامله‌گران، موجب شده است که پیش‌بینی قیمت سهام و بهینه‌سازی پرتفوی به یکی از پیچیده‌ترین مسائل در حوزه مالی تبدیل شود. در چنین شرایطی، رویکردهای مبتنی بر هوش مصنوعی، یادگیری ماشین و تحلیل احساسات به‌عنوان ابزارهایی نوین برای تحلیل رفتار بازار مورد توجه قرار گرفته‌اند (Asghari et al., 2024; Lin & Marques, 2024). مطالعات جدید نشان می‌دهند که بازارهای مالی نه تنها تحت تأثیر متغیرهای اقتصادی و بنیادی قرار دارند، بلکه واکنش‌های احساسی سرمایه‌گذاران نیز نقش تعیین‌کننده‌ای در شکل‌گیری روندهای قیمتی و نوسانات بازار ایفا می‌کند (Pompian, 2009; Raei & Fallahpour, 2004).

از این رو، ترکیب تحلیل داده‌های تاریخی با شاخص‌های رفتاری و احساسی می‌تواند افق جدیدی برای بهبود دقت پیش‌بینی و تصمیم‌گیری سرمایه‌گذاری فراهم سازد. مبانی نظری مالی رفتاری بر این فرض استوار است که سرمایه‌گذاران همواره منطقی عمل نمی‌کنند و تصمیمات آنان اغلب تحت تأثیر سوگیری‌های شناختی، هیجانات و واکنش‌های روان‌شناختی قرار دارد. برخلاف نظریه‌های کلاسیک مالی که فرض می‌کنند سرمایه‌گذاران دارای عقلانیت کامل هستند، نظریه مالی رفتاری نشان می‌دهد که احساساتی مانند ترس، طمع، اضطراب و خوش‌بینی می‌توانند موجب انحراف قیمت‌ها از ارزش‌های بنیادی شوند (Eslami Bidgoli & Kordlouie, 2010; Raei & Fallahpour, 2004). در همین راستا، پژوهش‌های متعددی اثر احساسات بر ریسک‌پذیری و تصمیمات سرمایه‌گذاری را بررسی کرده‌اند. برای مثال، فسler و همکاران نشان دادند که هیجانات مختلف می‌توانند رفتارهای متفاوتی در پذیرش ریسک ایجاد کنند (Fessler et al., 2004). همچنین گامبتی و جیوسبرتی نشان دادند که ویژگی‌های هیجانی مانند اضطراب و خشم بر انتخاب دارایی‌ها و رفتار سرمایه‌گذاری افراد اثر معنادار دارد (Gambetti & Giusberti, 2012). این یافته‌ها بیانگر آن است که تحلیل رفتار سرمایه‌گذاران و استخراج شاخص‌های احساسی می‌تواند به‌عنوان یک منبع اطلاعاتی مکمل در مدل‌های مالی مورد استفاده قرار گیرد.

با ظهور رسانه‌های اجتماعی و توسعه فناوری‌های پردازش زبان طبیعی، امکان استخراج احساسات سرمایه‌گذاران از داده‌های متنی فراهم شده است. شبکه‌هایی مانند توییتر، ردیت و پلتفرم‌های خبری مالی به منابعی غنی برای سنجش نگرش بازار تبدیل شده‌اند و پژوهشگران تلاش کرده‌اند تا از این اطلاعات برای پیش‌بینی روندهای قیمتی استفاده کنند (Salehi Chegini et al., 2025; Saravanos & Kanavos, 2025). در این میان، تحلیل احساسات به‌عنوان یکی از شاخه‌های مهم داده‌کاوی متنی، قادر است نگرش مثبت، منفی یا خنثی کاربران را از متن استخراج کرده و آن را به شاخص‌های کمی تبدیل کند. مطالعات اخیر نشان داده‌اند که ترکیب داده‌های احساسی با اطلاعات تکنیکال و بنیادی، عملکرد مدل‌های پیش‌بینی را به‌طور چشمگیری بهبود می‌دهد (Gu et al., 2024). به‌ویژه مدل‌های مبتنی بر FinBERT و شبکه‌های LSTM توانسته‌اند با استفاده از اخبار و متون مالی، الگوهای پنهان بازار را شناسایی کرده و بازده سهام را با دقت بیشتری پیش‌بینی کنند (Gu et al., 2024). همچنین، پژوهش ساراوانوس و کاناووس نشان داد که تحلیل احساسات شبکه‌های اجتماعی می‌تواند در پیش‌بینی نوسانات بازار سهام عملکرد مؤثری داشته باشد (Saravanos & Kanavos, 2025).

در بازار سرمایه ایران نیز نقش احساسات سرمایه‌گذاران به‌طور گسترده مورد بررسی قرار گرفته است. آقابابائی نشان داد که احساسات سرمایه‌گذاران بر نقدشوندگی بازار و نوسانات آن در بورس تهران تأثیر معناداری دارد (Aghababaei, 2022). حسینی و مرشدی نیز بیان کردند که رفتارهای احساسی سرمایه‌گذاران می‌تواند پویایی معاملات و روندهای معاملاتی بازار را تغییر دهد (Hosseini & Morshedi, 2019). همچنین دادگر و همکاران با رویکرد اقتصاد رفتاری نشان دادند که احساسات سرمایه‌گذاران و رفتار دولت نقش مهمی در ایجاد نوسانات بازار سهام دارند (Dadgar et al., 2023). پژوهش‌های داخلی دیگر نیز بر ارتباط مسئولیت اجتماعی شرکت‌ها، ادراک سرمایه‌گذاران و واکنش‌های رفتاری بازار تأکید کرده‌اند (Rezaei-Kalidbari et al., 2013; Zare Bahman-Miri et al., 2022). افزون بر این، وکیلی‌فرد و همکاران نشان دادند که واکنش‌های رفتاری سرمایه‌گذاران در بورس تهران می‌تواند ساختار تصمیم‌گیری مالی را تحت تأثیر قرار دهد (Vakilifard et al., 2013). مجموعه این مطالعات نشان می‌دهد که تحلیل احساسات در بازار سرمایه ایران نیز دارای ظرفیت بالایی برای بهبود مدل‌های پیش‌بینی و تصمیم‌گیری سرمایه‌گذاری است.

در کنار تحلیل احساسات، پیشرفت‌های حوزه یادگیری ماشین و داده‌کاوی نقش مهمی در تحول مدل‌های پیش‌بینی مالی ایفا کرده‌اند. مدل‌های سنتی آماری معمولاً در مواجهه با روابط غیرخطی و حجم عظیم داده‌ها با محدودیت مواجه هستند، در حالی که الگوریتم‌های یادگیری ماشین توانایی استخراج الگوهای پیچیده و پنهان را دارند (Lin & Marques, 2024). در سال‌های اخیر، مدل‌هایی مانند شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان، درخت تصمیم و الگوریتم‌های تقویتی برای پیش‌بینی قیمت سهام مورد استفاده قرار گرفته‌اند (Qiu & Song, 2016). پژوهش یون و همکاران نشان داد که ترکیب الگوریتم ژنتیک با XGBoost می‌تواند جهت حرکت قیمت سهام را با دقت بالایی پیش‌بینی کند (Yun et al., 2021). همچنین رضاییان و همکاران با استفاده از مدل‌سازی ساختاری تفسیری، مدلی جامع برای پیش‌بینی قیمت سهام ارائه کردند که نشان‌دهنده اهمیت ترکیب متغیرهای مختلف در تحلیل بازار است (Rezaeian et al., 2024). اصغری و همکاران نیز تأکید کردند که استفاده هم‌زمان از عوامل بنیادی، تکنیکال و اقتصادی می‌تواند دقت پیش‌بینی قیمت سهام را افزایش دهد (Asghari et al., 2024).

یکی از چالش‌های اساسی در مدل‌های یادگیری ماشین، مسئله انتخاب ویژگی است. داده‌های مالی معمولاً دارای ابعاد بالا و متغیرهای متعدد هستند و وجود ویژگی‌های غیرضروری می‌تواند موجب کاهش دقت مدل و افزایش بیش‌برازش شود. از این رو، انتخاب ویژگی به‌عنوان مرحله‌ای کلیدی در توسعه مدل‌های پیش‌بینی شناخته می‌شود (Li et al., 2016; Miao & Niu, 2016). روش‌های متعددی برای انتخاب ویژگی توسعه یافته‌اند که از جمله آن‌ها می‌توان به الگوریتم‌های مبتنی بر اطلاعات متقابل، الگوریتم‌های تکاملی و روش‌های فراابتکاری اشاره کرد (Ramirez-Gallego et al., 2017; Yu & Liu, 2003). در این میان، الگوریتم گرگ خاکستری به‌عنوان یکی از الگوریتم‌های فراابتکاری نوین، عملکرد مطلوبی در مسائل بهینه‌سازی و انتخاب ویژگی نشان داده است (Emary et al., 2016). نسخه‌های توسعه‌یافته این الگوریتم نیز توانسته‌اند کارایی بالاتری در جست‌وجوی فضای ویژگی‌ها و جلوگیری از گیر افتادن در بهینه‌های محلی ارائه دهند (Liu et al., 2023). استفاده از این الگوریتم‌ها در کنار مدل‌های یادگیری ماشین می‌تواند به ساخت مدل‌هایی با دقت بالاتر و پیچیدگی کمتر منجر شود.

علاوه بر مسئله پیش‌بینی، موضوع بهینه‌سازی پرتفوی نیز یکی از ارکان اساسی مدیریت سرمایه‌گذاری محسوب می‌شود. نظریه میانگین-واریانس مارکوویتز، اگرچه همچنان یکی از مهم‌ترین چارچوب‌های تصمیم‌گیری مالی است، اما در محیط‌های پرنوسان امروزی نیازمند ترکیب با ابزارهای هوشمند و داده‌محور است. مدل‌های کلاسیک عمدتاً بر داده‌های تاریخی متکی هستند و معمولاً اطلاعات رفتاری و احساسی بازار را نادیده می‌گیرند. در حالی که پژوهش‌های جدید نشان داده‌اند که استفاده از احساسات بازار می‌تواند در تخصیص دارایی و ساخت پرتفوی‌های کارا تر نقش مؤثری ایفا کند (Xing et al., 2018). همچنین شیانگ نشان داد که احساسات سرمایه‌گذاران می‌تواند بر عملکرد شرکت‌ها و رفتار سرمایه‌گذاری تأثیرگذار باشد (Xiang, 2022). بنابراین، ادغام تحلیل احساسات با چارچوب‌های بهینه‌سازی مالی می‌تواند به ایجاد پرتفوی‌هایی منجر شود که علاوه بر بازده بالاتر، از پایداری و مدیریت ریسک بهتری نیز برخوردار باشند.

مرور ادبیات پژوهش نشان می‌دهد که اگرچه مطالعات متعددی در زمینه پیش‌بینی قیمت سهام، تحلیل احساسات و بهینه‌سازی پرتفوی انجام شده است، اما هنوز شکاف قابل توجهی در ادغام هم‌زمان این سه حوزه وجود دارد. بسیاری از پژوهش‌ها تنها بر پیش‌بینی قیمت تمرکز کرده‌اند و خروجی مدل‌های پیش‌بینی را وارد فرآیند تصمیم‌گیری پرتفوی نکرده‌اند. از سوی دیگر، برخی مطالعات تحلیل احساسات را صرفاً به‌عنوان یک شاخص جانبی در نظر گرفته‌اند و از ظرفیت کامل داده‌های متنی بهره نبرده‌اند. همچنین در بخش قابل توجهی از مطالعات، انتخاب ویژگی و کاهش ابعاد داده‌ها به‌صورت بهینه انجام نشده است. بنابراین، نیاز به مدلی یکپارچه که بتواند با استفاده از یادگیری ماشین، تحلیل احساسات و بهینه‌سازی هوشمند پرتفوی، فرآیند تصمیم‌گیری سرمایه‌گذاری را بهبود بخشد، بیش از پیش احساس می‌شود.

بر این اساس، پژوهش حاضر با هدف ارائه مدل بهینه‌سازی پرتفوی مبتنی بر پیش‌بینی در نظرگیری احساسات سرمایه‌گذاران انجام شده است تا از طریق ترکیب داده‌های تاریخی بازار، شاخص‌های تکنیکال، تحلیل احساسات شبکه‌های اجتماعی و الگوریتم‌های یادگیری ماشین، مدلی جامع و هوشمند برای پیش‌بینی بازده سهام و تشکیل پرتفوی بهینه ارائه دهد.

روش پژوهش و مواد

پیش بینی قیمت سهام نوعی مدل پیش بینی محسوب می شود که در آن مقدار Y'_{t+1} به عنوان قیمت پایانی پیش بینی شده در روز $t + 1$ بر اساس مجموعه ای از ویژگی های ورودی X_t در روز t محاسبه شده و سپس با مقدار واقعی Y_{t+1} که یک عدد اعشاری است مقایسه می گردد. به بیان دیگر، مقدار Y'_{t+1} تابعی از ویژگی های ورودی X_t در زمان t است. به این معنا که:

$$Y'_{t+1} = f(X_t) \quad (1)$$

که در آن f یک تابع غیرخطی است که مجموعه ای از ویژگی های ورودی X_t را در روز t به مقادیر نتیجه عدد اعشاری در روز $t + 1$ ترسیم می کند. سری زمانی قیمت سهام X_t یک فرآیند مارکوف از مرتبه n است که تحت تأثیر n حالت قبلی قیمت ها قرار می گیرد که می تواند به صورت زیر بیان شود:

$$p(x_t | x_{t-1}, x_{t-2}, \dots, x_1) = p(x_t | x_{t-1}, x_{t-2}, \dots, x_{t-n}) \text{ for } t > n \quad (2)$$

که در آن $p(A|B)$ یک احتمال شرطی از رویداد A را نشان می دهد. این سری زمانی قیمت سهام X_t از فرآیندهای کلاسیک مارکوف متفاوت است.

$$p(x_t | x_{t-1}, x_{t-2}, \dots, x_1) = p(x_t | x_{t-1}) \quad (3)$$

از آنجا که در زنجیره مارکوف کلاسیک، حالت های گذشته، حال و آینده مستقل از یکدیگر در نظر گرفته می شوند، وضعیت بعدی تنها به حالت فعلی وابسته است و تحت تأثیر توالی حالت های قبلی قرار نمی گیرد. اما در سری های زمانی مربوط به قیمت سهام، در چارچوب فرآیند مارکوف با حافظه n ، قیمت روز آینده به مقادیر ثبت شده در n روز پیشین بستگی دارد. در نگاه مارکوفی به داده های قیمتی، سری زمانی سهام معمولاً بر پنج متغیر اصلی روزانه یعنی «بالا ترین»، «پایین ترین»، «قیمت باز»، «قیمت بسته» و «حجم معاملات» در روز t برای پیش بینی قیمت روز بعد متکی است. بنابراین، برای بهره گیری از اطلاعات موجود در n حالت گذشته، شاخص های فنی و تحلیل احساسات متنوعی بر پایه این پنج داده تاریخی اولیه و متون شبکه های اجتماعی ایجاد می شوند.

مقادیر تاریخی اصلی: $H_t^H, H_t^L, H_t^O, H_t^C, H_t^V$ مقادیر قیمت روز جاری را نشان می دهند: به ترتیب «بالا»، «کم»، «باز»، «بسته» و «حجم» و k شاخص فنی $T_t^1, T_t^2, \dots, T_t^k$ مقدار توابع غیرخطی مقادیر تاریخی n روز گذشته $H_{t-1}, H_t, \dots, H_{t-n}, H_{t-n+1}$ از روز t را نشان می دهد. در نتیجه، نتیجه پیش بینی شده روز $t + 1$ را می توان به عنوان یک تابع غیرخطی از ویژگی های ورودی روز t که شامل مقادیر تاریخی $H_t^H, H_t^L, H_t^O, H_t^C, H_t^V$ و k شاخص های فنی $T_t^1, T_t^2, \dots, T_t^k$ که اطلاعات تاریخی روزهای $t, t-1, \dots, t-n, t-n+1$ را منعکس می کند و بنابراین معادله ۳-۱ را می توان با معادله شماره ۴ جایگزین کرد.

$$Y'_{t+1} = f(H_t^H, H_t^L, H_t^O, H_t^C, H_t^V, H_t^1, H_t^2, \dots, T_t^K) \quad (4)$$

که در آن $H_t^H, H_t^L, H_t^O, H_t^C, H_t^V$ عبارتند از «بالا»، «کم»، «باز»، «بسته»، «حجم» و $T_t^1, T_t^2, \dots, T_t^k$ عبارتند از k شاخص فنی و تحلیل احساسات روز t هستند. در اینجا، شاخص فنی و تحلیل احساسات j ام روز t برابر با:

$$T_t^j = g^j(H_{t-n}, H_{t-n+1}, \dots, H_{t-1}, H_t) \quad (5)$$

که در آن g^j یک تابع غیرخطی از مقادیر تاریخی n روز گذشته: $H_{t-n}, H_{t-n+1}, \dots, H_{t-1}, H_t$ از روز t است.

الگوریتم گرگ خاکستری (Grey Wolf Optimizer – GWO) یکی از روش های فراابتکاری مبتنی بر جمعیت است که با الهام از ساختار اجتماعی و رفتار شکار گرگ های خاکستری طراحی شده و به طور گسترده در حل مسائل پیچیده ی بهینه سازی مورد استفاده قرار می گیرد.

در GWO، فرآیند شکار طعمه (یافتن بهترین جواب) از طریق مدل‌سازی رفتار محاصره‌ای گرگ‌ها انجام می‌شود. این رفتار به صورت ریاضی با استفاده از بردارهای موقعیت و فاصله تعریف می‌شود.

$$X(t+1) = X_p(t) - \quad (6)$$

که در آن X_p موقعیت طعمه، A بردار ضریب، و D به صورت زیر تعریف شده است.

$$D = |CX_p(t) - X(t)| \quad (7)$$

که در آن C بردار ضریب، X موقعیت گرگ خاکستری و t تعداد تکرارها است. بردارهای ضریب A و C ، به صورت زیر تعیین می‌شوند.

$$A = 2a \cdot r_1 - a \quad (8)$$

$$C = 2r_2 \quad (9)$$

که در آن r_1 و r_2 دو عدد تصادفی مستقل هستند که به طور یکنواخت بین $[0, 1]$ توزیع شده‌اند، و a ضریب حلقه‌ای است که برای ایجاد تعادل بین اکتشاف و بهره‌برداری استفاده می‌شود. طبق رابطه زیر در GWO، پارامتر a به صورت خطی از ۲ به ۰ کاهش می‌یابد.

$$a = 2 - 2\left(\frac{t}{T}\right) \quad (10)$$

که در آن t تعداد تکرارها و T حداکثر تعداد تکرارها است. در GWO، گرگ‌های آلفا، بتا و دلتا پیشرو به داشتن اطلاعات بهتر در مورد موقعیت بالقوه طعمه شناخته شده‌اند. بنابراین، رهبران گرگ‌های امگا را راهنمایی می‌کنند تا به سمت موقعیت مطلوب حرکت کنند. از نظر ریاضی، موقعیت جدید گرگ مانند رابطه زیر به روز می‌شود.

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (11)$$

که در آن X_1 ، X_2 و X_3 به صورت زیر محاسبه می‌شوند:

$$X_1 = |X_\alpha - A_1 \cdot D_\alpha| \quad (12)$$

$$X_2 = |X_\beta - A_2 \cdot D_\beta| \quad (13)$$

$$X_3 = |X_\delta - A_3 \cdot D_\delta| \quad (14)$$

که در آن X_α ، X_β و X_δ موقعیت آلفا، بتا و دلتا در تکرار t هستند. A_1 ، A_2 و A_3 مطابق با معادله (۹) محاسبه می‌شوند. و D_α ، D_β و D_δ به ترتیب در معادلات (۱۵) تا (۱۷) تعریف شده‌اند.

$$D_\alpha = |C_1 \cdot X_\alpha - X| \quad (15)$$

$$D_\beta = |C_2 \cdot X_\beta - X| \quad (16)$$

$$D_\delta = |C_3 \cdot X_\delta - X| \quad (17)$$

که در آن C_1 ، C_2 و C_3 مانند رابطه (۱۰) محاسبه می‌شوند (Liu et al., ۲۰۲۳; Mirjalili et al., ۲۰۱۴). به طور کلی، الگوریتم گرگ خاکستری (Grey Wolf Optimizer – GWO) به عنوان یک روش فراابتکاری مبتنی بر جمعیت، برای حل مسائل بهینه‌سازی در فضای پیوسته طراحی شده است و عملکرد مؤثری در یافتن پاسخ‌های بهینه در مسائل غیرخطی و پیچیده از خود نشان داده است. این الگوریتم با الهام از رفتار اجتماعی گرگ‌های خاکستری در طبیعت، تلاش می‌کند فرآیند جست‌وجوی پاسخ بهینه را از طریق شبیه‌سازی سلسله‌مراتب گله و نحوه‌ی شکار طعمه مدل‌سازی کند. در ساختار این الگوریتم، سه راه‌حل برتر در هر مرحله به ترتیب با عنوان آلفا (α)، بتا (β) و دلتا (δ) شناخته می‌شوند. سایر راه‌حل‌ها که از نظر کیفیت پایین‌تر هستند، در دسته‌ی امگا (w) قرار می‌گیرند.

در این مدل، نقش هدایت‌گر فرآیند بهینه‌سازی بر عهده‌ی سه عضو اصلی گله است—آلفا، بتا و دلتا—که موقعیت‌های آن‌ها به عنوان مرجع برای به‌روزرسانی موقعیت سایر اعضا در نظر گرفته می‌شود. امگاها، که نماینده‌ی راه‌حل‌های ضعیف‌ترند، با پیروی از موقعیت‌های سه رهبر، مسیر حرکت خود را تنظیم می‌کنند. این ساختار سلسله‌مراتبی، الگوریتم

را قادر می‌سازد تا همزمان از سه نقطه‌ی مرجع برای هدایت جست‌وجو استفاده کند؛ رویکردی که به‌طور طبیعی تعادل میان اکتشاف (exploration) و بهره‌برداری (exploitation) را حفظ می‌کند.

رفتار محاصره‌ای گرگ‌ها در هنگام شکار، که یکی از ویژگی‌های کلیدی این الگوریتم است، به‌صورت ریاضی با استفاده از بردارهای فاصله و ضرایب تصادفی مدل‌سازی می‌شود. در هر مرحله، موقعیت گرگ‌ها با توجه به فاصله‌ی آن‌ها از طعمه (که در اینجا همان بهترین راه‌حل شناخته‌شده است) و با در نظر گرفتن عوامل کنترلی، به‌روزرسانی می‌شود. این فرآیند به الگوریتم اجازه می‌دهد تا به‌تدریج به سمت نواحی بهینه در فضای جست‌وجو حرکت کند، بدون آن‌که در مراحل اولیه بیش‌ازحد به یک نقطه خاص متمرکز شود.

الگوریتم XGBoost پیاده‌سازی گسترش‌یافته‌ای از چارچوب گرادین بوس‌تینگ تعمیم‌یافته است که در آن یک عبارت منظم‌سازی نیز به فرآیند یادگیری اضافه می‌شود. وجود این جزء کنترلی، به الگوریتم کمک می‌کند تا از بروز مشکل برازش بیش از حد جلوگیری کرده و در عین حال امکان کار با انواع توابع زیان قابل تفکیک‌پذیر را فراهم آورد. برخلاف بهینه‌سازی صرف خطای مربعات ساده، در XGBoost یک تابع هدف تعریف می‌شود که از دو بخش اصلی تشکیل شده است: تابع زیان محاسبه‌شده روی داده‌های آموزشی و عبارت منظم‌سازی که با محدود کردن پیچیدگی مدل، از افزایش غیرضروری آن جلوگیری می‌کند.

$$Obj = \sum_i L(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (18)$$

$L(y, \hat{y})$ می‌تواند هر تابعی در ارتباط با ضرر متمایز محذب باشد که تفاوت بین پیش‌بینی و برچسب واقعی را برای یک نمونه آموزشی معین اندازه‌گیری می‌کند. $\Omega(f_k)$ پیچیدگی درخت f_k را توصیف می‌کند و در الگوریتم (XGBoost و Chen و Guestrin، ۲۰۱۶) به صورت زیر تعریف شده است:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda w^2 \quad (19)$$

که در آن، T تعداد برگ‌های درخت f_k بوده و w وزن برگ است (یعنی مقادیر پیش‌بینی‌شده ذخیره‌شده در گره‌های برگ). هنگامی که $\Omega(f_k)$ در تابع هدف گنجانده می‌شود، مجبور می‌شویم تا یک درخت با پیچیدگی کمتر را بهینه‌سازی کنیم. در این بین، مقدار $L(y, \hat{y})$ باید به حداقل میزان ممکن برسد. این امر به کاهش بیش‌برازش کمک می‌کند. γT یک جریمه ثابت برای هر برگ درخت اضافی ارائه می‌دهد و λw^2 وزن‌های شدید را جریمه می‌کند. γ و λ پارامترهای قابل تنظیم توسط کاربر هستند. با توجه به اینکه تقویت به صورت تکراری انجام می‌شود، می‌توانیم تابع هدف را برای تکرار فعلی m برحسب پیش‌بینی تکرار قبلی $\hat{y}_i^{(m-1)}$ بیان کنیم. این روند توسط جدیدترین درخت f_k تنظیم می‌شود:

$$Obj^m = \sum_i L(y_i, \hat{y}_i^{(m-1)} + f_k(x_i)) + \sum_k \Omega(f_k) \quad (20)$$

سپس می‌توانیم برای یافتن f_k که هدف ما را به حداقل می‌رساند، روند بهینه‌سازی را انجام دهیم.

استفاده از توسعه تیلور برای ارتقای این تابع به مرتبه دوم به ما این امکان را می‌دهد تا به راحتی توابع مختلف ضرر را تطبیق دهیم:

$$Obj^m \approx \sum_i [L(y_i, \hat{y}_i^{(m-1)}) + g_i f_k(x) + \frac{1}{2} h_i f_k(x)^2] + \sum_k \Omega(f_k) + constant \quad (21)$$

در اینجا، g_i و h_i به ترتیب مشتقات مرتبه اول و دوم تابع ضرر برای مثال i هستند:

$$g_i = \frac{dL(y_i, \hat{y}_i^{(m-1)})}{d\hat{y}_i^{(m-1)}} \quad h_i = \frac{d^2 L(y_i, \hat{y}_i^{(m-1)})}{d(\hat{y}_i^{(m-1)})^2} \quad (22)$$

توجه داشته باشید که مدل $\hat{y}_i^{(m-1)}$ در طول این فرآیند بهینه‌سازی بدون تغییر باقی می‌ماند.

تابع هدف ساده شده با حذف ثابت‌ها به شکل زیر خواهد بود:

$$Obj^m = \sum_i [g_i f_k(x) + \frac{1}{2} h_i f_k(x)^2] + \sum_k \Omega(f_k) \quad (23)$$

می‌توان به راحتی مشاهده کرد که درخت تصمیم مقادیر ثابت را در یک برگ پیش‌بینی می‌کند. سپس $f_k(x)$ را می‌توان به صورت $w_{q(x)}$ نشان داد که در آن، w بردار حاوی امتیازات برای هر برگ است و $q(x)$ نمونه x را برای یک برگ به تصویر می‌کشد.

سپس تابع هدف را می‌توان به گونه‌ای تغییر داد که بر روی برگ درخت و عبارت منظم‌سازی معادله ۳-۳۳ جمع‌بندی شود

$$Obj^m = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_{q(x)} + \frac{1}{2} \left(\sum_{i \in I_j} h_i \right) w_{q(x)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w^2 \quad (24)$$

در اینجا، I_j به مجموعه‌ای از نمونه‌های آموزشی در برگ j اشاره دارد. مجموع مشتقات در هر برگ را می‌توان به صورت زیر تعریف کرد:

$$G_j = \sum_{i \in I_j} g_i \quad H_j = \sum_{i \in I_j} h_i \quad (25)$$

همچنین توجه داشته باشید که $w_{q(x)}$ در هر برگ ثابت است و می‌تواند به صورت w_j نمایش داده شود. با ساده‌سازی این فرآیند، معادله زیر به دست می‌آید:

$$Obj^m = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \quad (26)$$

وزن w_j برای هر برگ تابع هدف را به حداقل می‌رساند:

$$\frac{\partial Obj^m}{\partial w_j} = G_j + (H_j + \lambda) w_j = 0 \quad (27)$$

بهترین وزن برگ w_j با توجه به ساختار درخت فعلی آن است.

$$w_j = -\frac{G_j}{H_j + \lambda} \quad (28)$$

با استفاده از بهترین w_j در معادله (۲۸)، تابع هدف برای یافتن بهترین ساختار درختی به شکل زیر خواهد بود:

$$Obj^m = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (29)$$

معادله (۲۹) در XGBoost به عنوان معیاری برای کیفیت یک درخت معین استفاده می‌شود.

مدل میانگین‌وارینانس برای بهینه‌سازی پورتفولیو

مدل میانگین-وارینانس که نخستین بار توسط مارکوویتز ارائه شد، چارچوبی اساسی برای انتخاب پرتفوی محسوب می‌شود. در این مدل، بازده سرمایه‌گذاری با بازده مورد انتظار و ریسک با واریانس یا انحراف معیار بازده‌ها سنجیده می‌شود. بر اساس این رویکرد، سرمایه‌گذاران عقلایی همواره به دنبال آن هستند که یا برای سطح معینی از بازده، ریسک را به حداقل برسانند، یا در سطح مشخصی از ریسک، بیشترین بازده ممکن را به دست آورند. از این رو، تصمیم‌گیری در انتخاب پرتفوی به گونه‌ای انجام می‌شود که مطلوبیت مورد انتظار سرمایه‌گذار حداکثر شود. در واقع، هدف اصلی این مدل ایجاد توازن میان حداکثرسازی بازده و حداقل‌سازی ریسک است که با یک تابع بهینه‌سازی چندهدفه متعارف نمایش داده می‌شود (Chen et al., ۲۰۲۱):

$$\min \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij} \quad (30)$$

$$\max \sum_{i=1}^n x_i \mu_j$$

$$s. t. \begin{cases} \sum_{i=1}^n x_i = 1, \\ 0 \leq x_i \leq 1, \forall i = 1, \dots, n, \end{cases}$$

در جایی که σ_{ij} کوواریانس بین دارایی i و دارایی j است، x_i و x_j نشان دهنده نسبت ارزش اولیه است و μ_j بازده مورد انتظار دارایی i است. چانگ و همکاران ضریب ریسک‌گریزی را برای تغییر فرمول چند هدفه به فرمول تک‌هدف معرفی کرد (Chang et al., ۲۰۰۹):

$$\min \lambda \left[\sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij} \right] - (1 - \lambda) \left[\sum_{i=1}^n x_i \mu_j \right]$$

$$s. t. \begin{cases} \sum_{i=1}^n x_i = 1, \\ 0 \leq x_i \leq 1, \forall i = 1, \dots, n, \end{cases} \quad (31)$$

در این چارچوب، ضریب ریسک‌گریزی (λ) در بازه‌ای بین صفر تا یک تعریف می‌شود. زمانی که $\lambda = 0$ ، سرمایه‌گذار تنها بر حداکثرسازی بازده تمرکز دارد و سطح ریسک برای او اهمیتی ندارد. در نقطه مقابل، هنگامی که $\lambda = 1$ ، سرمایه‌گذار تنها به حداقل‌سازی ریسک توجه کرده و بازده را در نظر نمی‌گیرد. مقادیر میانی این ضریب، بیانگر حالتی هستند که در آن سرمایه‌گذار بین افزایش بازده مورد انتظار و کاهش ریسک، نوعی تعادل برقرار می‌کند. بدین ترتیب، هر سرمایه‌گذار می‌تواند متناسب با سطح ترجیحات ریسک خود، یکی از این راه‌حل‌های ممکن را انتخاب کرده و پرتفوی مناسب تشکیل دهد.

در این پژوهش از داده‌های روزانه بازه ۳۰ سپتامبر ۲۰۲۱ تا ۲۹ سپتامبر ۲۰۲۲ استفاده شده است. مجموعه نمادهای تحت بررسی عبارت‌اند از: AMD, AAPL (Apple), (Advanced Micro Devices), AMZN (Amazon), BA (Boeing), BX (Blackstone), COST (Costco Wholesale), CRM (Salesforce), DIS, TSM (Walt Disney), TSLA (Tesla), PG (Procter & Gamble), NIO (NIO Inc.), MSFT (Microsoft), KO (Coca-Cola), F (Ford Motor), (Taiwan Semiconductor Manufacturing Company), ZS (Zscaler). برای هر نماد، تاریخچه قیمت و حجم معاملات در بسامد روزانه گردآوری و با تقویم معاملاتی هماهنگ شده است. اطلاعات مربوط به تعداد داده‌های هر نماد در جدول ۴-۱ ارائه شده است.

جدول ۱- مشخصات نمادهای بورسی مورد استفاده در پژوهش، به همراه تعداد داده‌های مربوط به هر نماد

نام شرکت	مخفف	تعداد داده‌ها	تعداد توییت‌ها
Apple	AAPL	۲۵۲	۵۰۵۶
Advanced Micro Devices	AMD	۲۵۲	۲۲۲۷
Amazon	AMZN	۲۵۲	۴۰۸۹
Boeing	BA	۲۵۲	۳۹۹
Blackstone	BX	۲۵۲	۵۰
Costco Wholesale	COST	۲۵۲	۳۹۳
Salesforce	CRM	۲۵۲	۲۳۳
Walt Disney	DIS	۲۵۲	۶۳۵
Ford Motor	F	۲۵۲	۳۱
Coca-Cola	KO	۲۵۲	۳۱۰
Microsoft	MSFT	۲۵۲	۴۰۸۹
NIO Inc.	NIO	۲۵۲	۳۰۲۱
Procter & Gamble	PG	۲۵۲	۴۰۸۹
Tesla	TSLA	۲۵۲	۳۷۴۲۲
Taiwan Semiconductor Mfg.	TSM	۲۵۲	۱۱۰۳۴
Zscaler	ZS	۲۵۲	۱۹۳

حسابداری، امور مالی و هوش محاسباتی

در پیش پردازش ابتدا داده‌های قیمت و حجم به صورت روزانه برای ۱۶ نماد منتخب خوانده و یکدست می‌شود. تاریخ‌ها به تقویم معاملاتی هماهنگ می‌گردد، قالب زمانی یکسان می‌شود و نمادها نرمال سازی اسمی می‌شوند تا هر ردیف یکتا با ترکیب تاریخ و نماد تعریف شود. سپس بازه زمانی مورد نظر اعمال و ردیف‌های ناقص یا تکراری حذف می‌گردد تا یک پنل پایدار و مرتب برحسب زمان به دست آید.

بر پایه سری قیمت، مجموعه گسترده‌ای از ویژگی‌های تکنیکال ساخته می‌شود؛ از میانگین‌های متحرک و فاصله از آن‌ها تا نوسانگرها و سنجه‌های مومنتوم، شاخص‌های نوسان و جریان/انباشت حجم، بازده لگاریتمی محاسبه و با تاخیرهای روزانه گسترش می‌یابد و نسبت‌های قیمتی مانند نسبت‌های سقف به باز و کف به باز با لگ‌های زمانی استخراج می‌شود. این ویژگی‌ها تصویری غنی از روند، نوسان و مومنتوم هر نماد در مقیاس روزانه فراهم می‌کنند.

در گام ادغام، جدول ویژگی‌های قیمتی با جدول ویژگی‌های متنی براساس تاریخ و نماد ترکیب می‌شود. پس از ادغام، ستون‌ها بازبینی و یکدست، ترتیب زمانی تثبیت و هر نوع ناسازگاری برطرف می‌گردد تا یک ماتریس ویژگی نهایی برای هر نماد در هر روز شکل بگیرد. در پایان، خروجی پیش پردازش به صورت یک فایل کش محلی ذخیره می‌شود تا در اجراهای بعدی بدون محاسبات تکراری و پرهزینه، مستقیماً بارگذاری گردد. این جریان باعث می‌شود داده نهایی برای آموزش مدل و ارزیابی پرتفوی، پایدار، همگن و آماده استفاده باشد.

در این پژوهش برای هر مشاهده ۱۰۵ ویژگی ساخته شده است که ۹۲ مورد آن فنی و بر پایه شاخص‌های تکنیکال (میانگین‌های متحرک، مومنتوم و نوسان‌سازها مانند MACD و RSI و استوکاستیک، سنجه‌های نوسان مانند ATR/NATR، ویژگی‌های مبتنی بر حجم و جریان نقدینگی، تبدیلات هیلبرت، بازده‌های لگاریتمی و نسبت‌های لگ شده قیمت) و ۱۳ مورد آن مرتبط با احساس بازار و تعاملات شبکه‌های اجتماعی است. ویژگی‌های احساسی از تجمیع روزانه توئیتهای به دست آمده و شامل میانگین، میانه و واریانس مرکب احساس، شمارش توئیتهای مثبت و منفی، لگ شمارش توئیتهای مثبت و منفی و میانگین‌های کلاس‌های مثبت، منفی و خنثی می‌شود. این ترکیب، تصویری هم‌زمان از رفتار قیمت و سیگنال‌های متنی بازار فراهم می‌کند.

یافته‌ها

در این بخش نتایج مربوط به الگوریتم XGBoost با مهندسی ویژگی توسط BGWO به تفکیک نمادهای بورسی ارائه می‌شود. در جدول ۲ پارامترهای پایه الگوریتم BGWO نشان داده شده است. اندازه جمعیت برابر با ۲۰ در نظر گرفته شده تا تنوع جستجو حفظ شود و در عین حال هزینه محاسباتی کنترل شود. بعد هر عضو جمعیت ۱۰۵ است که با تعداد ویژگی‌های ورودی همخوانی دارد و امکان جستجوی هم‌زمان در فضای ویژگی گسترده را فراهم می‌کند. تعداد تکرار ۵۰ در نظر گرفته شده تا فرایند همگرایی با توازن میان دقت و زمان اجرا انجام شود و از گرفتار شدن در بهینه‌های محلی کاسته شود. این تنظیمات مبنای اجرای روال بهینه‌سازی و انتخاب ویژگی در ادامه پژوهش است. بدین منظور برای ۱۶ نماد مورد بررسی به تفکیک شاخص‌های بهینه انتخابی و نتایج ارزیابی ارائه می‌شود.

جدول ۲- مقادیر پارامترهای قابل تنظیم در الگوریتم BGWO

پارامتر قابل تنظیم	مقادیر مربوطه
تعداد جمعیت اولیه	۲۰
ابعاد اعضای جمعیت	۱۰۵
تعداد تکرار	۵۰

در جدول ۲، نتایج شاخص‌های ارزیابی برای نمادها گزارش شده است. مقدار R2 برای بخش قابل توجهی از نمادها بالا است؛ برای نمونه KO با ۰/۸۹۴۵ و COST با ۰/۸۹۱۵ کارایی پیش‌بینی قوی را نشان می‌دهند، در حالی که TSLA با ۰/۵۴۶۸ و BX با ۰/۵۴۸۵ در سطح پایین تری قرار دارند. از نظر خطا، کمترین RMSE مربوط به F با ۰/۴۷۸۸ و پس از آن KO با ۰/۷۲۲۶ و NIO با ۰/۸۸۲۷ است، در حالی که CRM با ۰/۹۶۴۳ و سپس COST و MSFT دارای RMSE بزرگ تری هستند. الگوهای MAPE

قاسمی کیان و نجفی

نیز از دقت بالاتر برای KO و COST و دقت پایین تر برای برخی نمادهای پرنوسان حکایت دارد. این تصویر کلی نشان می‌دهد ترکیب ویژگی‌های فنی و احساسات بازار برای بسیاری از نمادها بهبود قابل ملاحظه‌ای در برآزش ایجاد کرده است.

جدول ۳- نتایج شاخص‌های ارزیابی در الگوریتم BGWO-XGBoost برای نمادهای بورسی

نام نماد	MSE	RMSE	MAE	MAPE	R ²
AAPL	۹۳۵۷۶	۳۰۵۹۰	۲۴۹۶۳	۱۵۶۹۱	۰.۸۴۷۸
AMD	۲۲۵۷۵۹	۴۷۵۱۳	۳۳۵۵۳	۴۲۸۵۵	۰.۸۳۴۹
AMZN	۱۳۲۶۴۵	۳۶۴۰۸	۲۸۶۶۷	۲۱۸۰۲	۰.۸۳۲۳
BA	۲۳۱۰۳۹	۴۸۰۶۲	۳۹۱۰۷	۲۵۵۶۶	۰.۸۵۹۴
BX	۲۳۸۴۲۴	۴۸۸۲۴	۳۵۳۵۴	۳۸۷۹۳	۰.۵۴۸۵
COST	۶۲۶۸۷۵	۷۹۱۹۷	۶۴۶۸۲	۱۲۳۶۳	۰.۸۹۱۵
CRM	۸۰۴۶۲۱	۸۹۶۴۳	۷۰۴۵۳	۴۴۸۴۸	۰.۷۰۰۳
DIS	۷۱۳۴۰	۲۶۷۱۱	۲۲۴۷۸	۲۰۱۸۸	۰.۸۵۶۸
F	۰.۲۲۹۲	۰.۴۷۸۸	۰.۳۶۳۷	۲.۵۵۰۴	۰.۸۵۵۱
KO	۰.۵۲۲۱	۰.۷۲۲۶	۰.۵۶۶۰	۰.۹۳۱۴	۰.۸۹۴۵
MSFT	۶۲۲۱۵۶	۷۸۸۹۵	۶۴۲۶۰	۲۵۰۳۱	۰.۸۱۳۲
NIO	۰.۷۷۹۲	۰.۸۸۲۷	۰.۶۴۷۸	۳۳۸۲۴	۰.۶۳۷۸
PG	۷.۹۳۱۰	۲.۸۱۶۲	۲.۱۱۹۶	۱.۵۳۹۴	۰.۷۰۹۸
TSLA	۵۹.۲۲۰۵	۷.۶۹۴۸	۶.۲۲۸۳	۲.۱۵۸۳	۰.۵۴۶۸
TSM	۹۳۱۶۴	۳۰۵۲۳	۲۲۰۹۴	۲.۷۹۶۶	۰.۶۸۹۱
ZS	۴۲.۵۳۶۱	۶.۵۰۲۸	۴.۸۲۲۷	۲.۸۷۵۱	۰.۶۷۰۳

در جدول ۴، دامنه پارامترهای قابل تنظیم برای جستجوی XGBoost ارائه شده است. عمق درخت از ۴ تا ۱۲، نرخ یادگیری از ۰/۱ تا ۰/۱، و نسبت‌های نمونه برداری سطر و ستونی در بازه ۰/۶ تا ۱/۰ تعریف شده تا طیف مناسبی از پیچیدگی مدل و کنترل واریانس پوشش داده شود. پارامترهای منظم‌سازی نیز با مقادیر ۰/۰ تا ۵/۰ برای reg_lambda و ۰/۰ تا ۲/۰ برای reg_alpha در نظر گرفته شده تا امکان مهار بیش برآزش فراهم باشد. دامنه تعداد درخت‌ها بین ۵۰۰ تا ۲۰۰۰ تنظیم شده تا مدل بتواند با نرخ یادگیری‌های پایین تر به تدریج همگرا شود.

جدول ۴- مقادیر پارامترهای قابل تنظیم در الگوریتم XGBoost

پارامتر قابل تنظیم	دامنه مقادیر
max_depth	۴، ۶، ۸، ۱۰، ۱۲
learning_rate	۰.۰۱، ۰.۰۳، ۰.۰۵، ۰.۱
Subsample	۰.۶، ۰.۸، ۱.۰
colsample_bytree	۰.۶، ۰.۸، ۱.۰
reg_lambda	۰.۰، ۰.۵، ۱.۰، ۲.۰، ۵.۰
reg_alpha	۰.۰، ۰.۵، ۱.۰، ۲.۰
n_estimators	۵۰۰، ۱۰۰۰، ۱۵۰۰، ۲۰۰۰

در جدول ۵، فرآیندهای بهینه شده برای هر نماد گزارش شده است. تکرار توقف زود هنگام برای همه موارد برابر با ۵۰ ثابت نگه داشته شده و سایر پارامترها متناسب با ویژگی‌های هر سری زمانی انتخاب شده‌اند؛ برای نمونه عمق‌های بالاتر مانند ۱۲ برای AAPL و AMZN و عمق‌های محدودتر مانند ۴ برای AMD و BX برگزیده شده که نشان از تفاوت در پیچیدگی الگوهای درون داده دارد. همچنین در برخی نمادها از منظم‌سازی قوی تر استفاده شده است، مانند DIS و TSLA با reg_lambda برابر ۵/۰، در حالی که

برای نمادهایی با ساختار ساده تر، مقادیر کمتر انتخاب شده است. ترکیب‌های نهایی نرخ یادگیری و تعداد درخت‌ها نیز از ۰/۱ تا ۰/۱ و ۵۰۰ تا ۲۰۰۰ متغیر بوده تا توازن میان سرعت یادگیری و کنترل خطا حاصل شود.

جدول ۵- نتایج فرآیندهای بهینه شده در الگوریتم XGBoost برای نمادهای بورسی

نام نماد	early_stopping_rounds	max_depth	learning_rate	subsample	colsample_bytree	reg_lambda	reg_alpha	n_estimators
AAPL	۵۰	۱۲	۰.۰۳	۰.۶۰	۰.۸۰	۲.۰۰	۱.۰۰	۱۰۰۰
AMD	۵۰	۴	۰.۰۱	۱.۰۰	۰.۸۰	۰.۵۰	۰.۵۰	۱۰۰۰
AMZN	۵۰	۱۲	۰.۰۳	۰.۶۰	۰.۸۰	۲.۰۰	۱.۰۰	۱۰۰۰
BA	۵۰	۴	۰.۰۳	۰.۶۰	۰.۸۰	۱.۰۰	۲.۰۰	۱۵۰۰
BX	۵۰	۴	۰.۰۳	۱.۰۰	۱.۰۰	۱.۰۰	۰.۵۰	۲۰۰۰
COST	۵۰	۸	۰.۰۱	۰.۶۰	۰.۶۰	۱.۰۰	۲.۰۰	۱۵۰۰
CRM	۵۰	۱۲	۰.۱۰	۱.۰۰	۰.۸۰	۱.۰۰	۰.۵۰	۱۰۰۰
DIS	۵۰	۱۰	۰.۰۱	۰.۸۰	۰.۸۰	۵.۰۰	۱.۰۰	۵۰۰
F	۵۰	۱۰	۰.۱۰	۰.۶۰	۰.۸۰	۰.۵۰	۰.۵۰	۱۵۰۰
KO	۵۰	۴	۰.۰۳	۱.۰۰	۱.۰۰	۱.۰۰	۰.۵۰	۲۰۰۰
MSFT	۵۰	۱۲	۰.۰۱	۱.۰۰	۱.۰۰	۰.۰۰	۱.۰۰	۱۵۰۰
NIO	۵۰	۴	۰.۰۱	۱.۰۰	۰.۸۰	۰.۰۰	۰.۰۰	۵۰۰
PG	۵۰	۴	۰.۰۳	۰.۶۰	۰.۸۰	۱.۰۰	۲.۰۰	۱۵۰۰
TSLA	۵۰	۱۰	۰.۰۱	۰.۸۰	۰.۸۰	۵.۰۰	۲.۰۰	۱۵۰۰
TSM	۵۰	۶	۰.۰۳	۱.۰۰	۱.۰۰	۰.۰۰	۰.۵۰	۲۰۰۰
ZS	۵۰	۴	۰.۰۳	۱.۰۰	۱.۰۰	۱.۰۰	۰.۵۰	۲۰۰۰

الگوی کلی بیشتر نمادها نوسان پیرامون صفر با جهش‌های مثبت مقطعی است. در میان سهام فناوری، AMD از میانه‌ی ماه جهش پله‌ای و چشمگیری دارد و تا حوالی ۰/۱۵- رشد می‌کند؛ CRM نیز روندی صعودی را تا حدود ۰/۱۲ تجربه می‌کند و MSFT از ناحیه‌ی منفی ابتدای ماه به شیب صعودی باثباتی می‌رسد (پایان ماه نزدیک ۰/۰۶-). در مقابل، AAPL عمدتاً در محدوده‌ای محدود نوسان می‌کند (تقریباً ۰/۰۲±) و تغییراتش کم‌دامنه‌تر از سایر نمادهای رشد است. به‌طور کلی، براساس شکل ۲-۴ گروه فناوری سیگنال‌های مثبتی در نیمه دوم ماه نشان می‌دهد.

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (۳۲)$$

در بین نمادهای چرخه‌ای و تدافعی، الگوها متنوع‌تر است. AMZN افت تند میانه ماه (نزدیک -۰/۰۳) را با بازگشت تدریجی و سپس کاهش انتهای ماه جبران می‌کند؛ BA نوسان شدید با ریزش مقطعی (حدود -۰/۰۴) و جهش‌های مثبت کوتاه‌مدت دارد؛ COST پس از کف میانه ماه (حدود -۰/۰۲۵) جهشی تا نزدیک ۰/۰۳ را تجربه می‌کند؛ F قله‌ای تیز حوالی ۲۰ سپتامبر (تقریباً ۰/۰۴) و بعد افت انتهای ماه دارد؛ KO با رفت‌وبرگشت سریع از حدود -۰/۰۲ به حدود ۰/۰۱۷ مواجه است و DIS عمدتاً نزدیک صفر نوسان کرده و اواخر ماه یک جهش مثبت کوتاه‌مدت نشان می‌دهد.

در این پژوهش از چارچوب BGWO-XGBoost برای پیش‌بینی بازده قیمت پایانی در افق ۳۱ روز آینده استفاده شد و خروجی پیش‌بینی‌ها به‌عنوان ورودی مدل میانگین-واریانس مارکوفیتز برای ساخت یک پرتفوی بهینه به کار رفت. از میان ۱۶ نماد بررسی شده، ۱۱ نماد که مقدار R2 آن‌ها در بخش آزمون بالاتر از ۰/۷ بود برای مرحله بهینه‌سازی

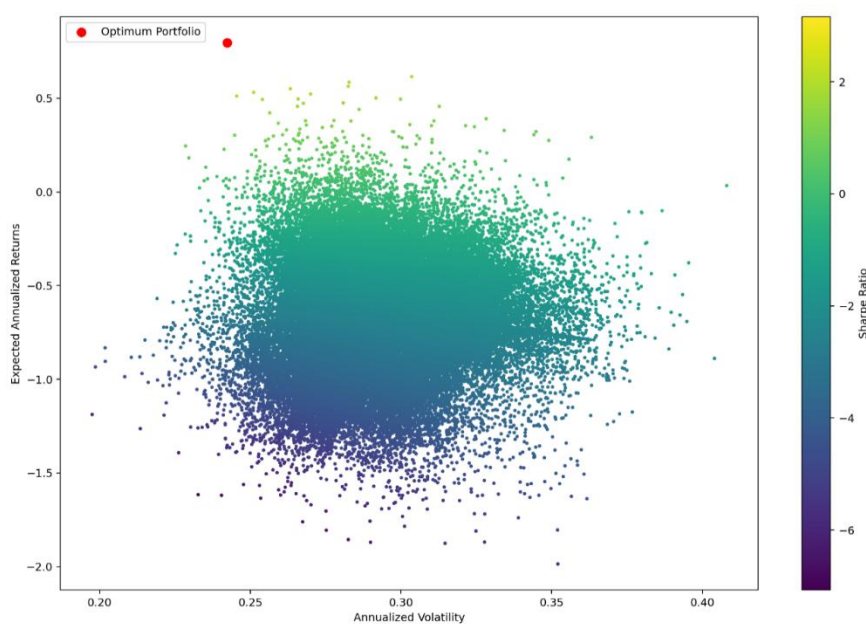
قاسمی کیان و نجفی

انتخاب شدند. این آستانه با هدف ورود نمادهای دارای قدرت پیش‌بینی بالاتر و کاهش خطای ورودی در فرآیند بهینه‌سازی تعیین شد؛ در نتیجه وزن‌دهی نهایی پرتفوی بر مبنای پیش‌بینی‌های قابل اعتمادتر انجام گرفته است.

جدول ۶- نتایج میانگین بازده پیش‌بینی شده نمادهای منتخب برای دوره ۳۱ روزه

نام نماد	میانگین بازدهی ماهانه
CRM	۰.۰۶۴۴
AMD	۰.۰۲۹۶
PG	۰.۰۱۳۲
MSFT	۰.۰۰۶۳
F	۰.۰۰۶۲
AAPL	۰.۰۰۳۵
AMZN	۰.۰۰۲۸
BA	۰.۰۰۱۳
KO	۰.۰۰۱۱
COST	-۰.۰۰۱۲
DIS	-۰.۰۰۷۶

گام بعدی، تعیین ترکیب بهینه‌ی وزن‌های پرتفوی است به نحوی که نسبت شارپ را حداکثر کند. برای این منظور، با روش مونت کارلو ۱۰۰,۰۰۰ بردار وزن مجاز (طولانی‌مدت، جمع وزن‌ها برابر ۱ و با سقف وزنی هر دارایی) به صورت تصادفی تولید می‌شود. برای هر بردار، بازده موردانتظار و نوسان‌پذیری برآورد شده و نسبت شارپ محاسبه می‌گردد؛ سپس برداری که بیشترین شارپ را دارد به عنوان ترکیب بهینه انتخاب می‌شود. ابر نقاط حاصل، تقریبی از «مرز کارآیی مارکویتز» را فراهم می‌کند و نقطه‌ی منتخب، پرتفوی با بهترین توازن ریسک-بازده در میان نمونه‌های شبیه‌سازی شده است. نرخ بدون ریسک سالانه ۳٪ (متناظر با اوراق خزانه‌داری ایالات متحده) اعمال و برای محاسبات روزانه به مقیاس ۲۵۲ روز معاملاتی تبدیل شده است.



شکل ۱- مقایسه بازدهی نمادهای مختلف برای دوره زمانی یک ماهه در بازه تاریخ ۲۹ آگوست ۲۰۲۲ تا ۲۹ سپتامبر ۲۰۲۲

حسابداری، امور مالی و هوش محاسباتی

در شکل مرز کارای مارکوویتز را نشان می‌دهد که با شبیه‌سازی مونت کارلو و تولید پرتفوی‌های مجاز متعدد ترسیم شده است و پرتفوی بهینه‌با بیشترین نسبت شارپ-به رنگ قرمز مشخص شده است. براساس نتایج به‌دست‌آمده، پرتفوی منتخب دارای نسبت شارپ سالانه ۳/۱۵۸۵، بازده موردانتظار سالانه ۷۹.۵۶٪ و نوسان سالانه ۲۴۲۴٪ (۲۴.۲۴٪) است. این اعداد بیانگر توازن مناسب میان ریسک و بازده در ترکیب انتخابی بوده و با قیود استاندارد (جمع وزن‌ها برابر یک و سقف وزنی هر دارایی) سازگار هستند. فایل تصویری مرز کارا نیز ذخیره شده تا جایگاه پرتفوی منتخب روی مرز به‌صورت بصری قابل مشاهده باشد.

جدول ۷ وزن‌های بهینه دارایی‌ها در پرتفوی منتخب را برای نمادهای بررسی‌شده گزارش می‌کند. بیشترین وزن‌ها به Procter & Gamble، Salesforce (CRM) و (PG) و Coca-Cola (KO) اختصاص یافته و پس از آن Apple (AAPL) قرار می‌گیرد؛ سایر نمادها سهم‌های کوچک‌تری در ترکیب دارند. جمع وزن‌ها برابر یک است.

جدول ۷- نتایج وزن بهینه دارایی از نمادهای منتخب

نام نماد	وزن بهینه دارایی
Salesforce (CRM)	۰.۲۳۲۴
Procter & Gamble (PG)	۰.۲۳۲۴
Coca-Cola (KO)	۰.۲۳۲۴
Apple (AAPL)	۰.۲۰۲۳
Ford Motor (F)	۰.۰۳۱۸
Amazon (AMZN)	۰.۰۲۶۰
Microsoft (MSFT)	۰.۰۲۲۸
Advanced Micro Devices (AMD)	۰.۰۱۸۹
Boeing (BA)	۰.۰۰۰۸

یافته‌ها نشان دادند که ادغام ویژگی‌های فنی و احساسی، برازش و پیش‌بینی بسیاری از نمادها را به‌طور معناداری بهبود می‌دهد؛ به‌ویژه نمادهایی مانند KO و COST با R^2 حدود ۰/۸۹ کارایی بالایی داشتند، درحالی‌که نمادهای پرنوسانی چون TSLA و BX به‌طور نسبی دقت پایین‌تری نشان دادند. جست‌وجوی فرایارامترها و انتخاب ویژگی‌های متمایز (حدود ۲۹ تا ۴۱ ویژگی بسته به نماد) ساختاری کم‌افزونگی و سازگار با رژیم‌های رفتاری سهام ایجاد کرد و جریان پیش‌پردازش متنی امن، پیوستگی پنل روزانه را حفظ نمود. در مرحله تصمیم‌گیری، بهره‌گیری از میانگین بازده‌های پیش‌بینی‌شده و بهینه‌سازی مارکوویتز به پرتفویی با نسبت شارپ سالانه ۳/۱۶، بازده موردانتظار ۷۹٪ و نوسان ۲۴٪ انجامید؛ ترکیبی که با تمرکز وزنی بر CRM، PG، KO و AAPL شکل گرفت و نشان داد پیوند پیش‌بینی داده‌محور با بهینه‌سازی پرتفوی می‌تواند به ساخت سبدی متوازن و اجرایی برای افق کوتاه

تشکیل یک پرتفوی کارا بدون پیش‌بینی قابل‌اعتماد از قیمت یا بازده آینده امکان‌پذیر نیست، درحالی‌که روش‌های سنتی به‌دلیل غیرخطی بودن رفتار بازار و حجم بالای داده‌ها کارایی محدودی دارند. به‌همین دلیل، در این تحقیق داده‌های تاریخی در کنار سیگنال‌های رفتاری بازار یعنی احساسات سرمایه‌گذاران تلفیق شده و با استفاده از روش‌های یادگیری ماشین دقت پیش‌بینی افزایش یافته است. خروجی این پیش‌بینی‌ها مستقیماً وارد چارچوب بهینه‌سازی پرتفوی شده است؛ رویکردی که هم نقدی بر دیدگاه «کفایت تاریخچه قیمت» محسوب می‌شود و هم بر مبانی نظری میانگین-واریانس در موازنه ریسک و بازده استوار است.

ادبیات مالی رفتاری نشان می‌دهد که احساسات می‌تواند بازده، نوسان و حتی نرخ‌های تنزیل ضمنی را تحت تأثیر قرار دهد؛ چه از طریق شاخص‌های پیمایشی و متغیرهای بازار و چه از راه تحلیل متن خبرها و شبکه‌های اجتماعی شواهد جدید این نقش را در بازارهای مختلف تأیید کرده‌اند، از بازار رمزارز با داده‌های تراکنشی کوین‌بیس تا بورس چین با شاخص‌های و در پژوهش‌های داخلی نیز ارتباط معنادار احساسات با بازده و نوسان گزارش شده است. در نتیجه، روشن می‌شود که افزودن سیستماتیک سیگنال‌های احساسی به داده‌های تاریخی می‌تواند قدرت تبیینی و پیش‌بینی را برای تصمیم‌گیری‌های پرتفوی افزایش دهد.

داده‌های روزانه ۱۶ نماد بورس آمریکا در بازه ۳۰ سپتامبر ۲۰۲۱ تا ۲۹ سپتامبر ۲۰۲۲ گردآوری و با تقویم معاملاتی همگام می‌شود؛ سری‌های استاندارد OHLCV پاک‌سازی و به شاخص‌های کمکی مانند بازده لگاریتمی، شکاف میانگین‌های متحرک، سنج‌های نوسان و مومنتوم بسط داده می‌شود. هم‌زمان، لایه متنی توییت‌ها برای هر نماد پیش‌پردازش می‌شود تا نویزها، ربات‌ها و موارد نامرتب حذف و زمان پیام‌ها به روز معاملاتی صحیح نگاشت شود. خروجی این مرحله ماتریس ویژگی ۱۰۵ بعدی است (۹۲ فنی و ۱۳ احساسی) که در سطح «نماد-روز» ادغام و بدون شکاف زمانی نگهداری می‌شود؛ برای روزهای فاقد متن نیز پرشدن امن شاخص‌های احساسی به کار می‌رود تا ساختار پنل منسجم باقی بماند.

بر این بستر داده، چارچوب BGWO-XGBoost برای انتخاب ویژگی و یادگیری به کار گرفته می‌شود. در بخش بهینه‌سازی، پارامترهای BGWO با جمعیت ۲۰، بعد ۱۰۵ و ۵۰ تکرار تنظیم می‌شود و دامنه فرارپارامترهای XGBoost (عمق ۴ تا ۱۲، نرخ یادگیری ۰/۰۱ تا ۰/۱، نمونه‌برداری سطر و ستونی ۰/۶ تا ۱، و منظم‌سازی‌های گوناگون) جست‌وجو می‌شود تا برای هر نماد به ترکیب پیکربندی مناسبی برسد. ارزیابی با چند شاخص خطا و برازش نشان می‌دهد ادغام سیگنال‌های فنی و احساسی بهبود معناداری ایجاد می‌کند؛ برای نمونه، KO و COST به R^2 حدود ۰/۸۹ می‌رسند و در مقابل نمادهای پرنوسانی مانند TSLA و BX دقت پایین‌تری ثبت می‌کنند. سید ویژگی‌های منتخب، بسته به نماد، بین ۲۹ تا ۴۱ متغیر متغیر است و معمولاً ترکیبی از روند، مومنتوم، نوسان، شاخص‌های مبتنی بر حجم و چند نمایه احساسی (میانگین/میان‌ه احساس، نسبت مثبت به منفی و شدت گفتگو) را دربر می‌گیرد؛ انتخابی که افزونگی را مهار و پوشش اطلاعاتی را حفظ می‌کند.

در گام تصمیم‌سازی، پیش‌بینی بازده ۳۱ روزه مبنا قرار می‌گیرد. ابتدا نمادهایی که در بخش آزمون به R^2 بالاتر از ۰/۷۰ می‌رسند به‌عنوان کاندیدای پرتفوی برگزیده می‌شود و از میان آن‌ها، ۹ نماد با میانگین بازده پیش‌بینی شده مثبت به مرحله بهینه‌سازی میانگین-واریانس وارد می‌شود. سپس با شبیه‌سازی مونت کارلو ۱۰۰ هزار بردار وزن (جمع وزن‌ها برابر یک و سقف وزنی برای هر دارایی) مرز کارای مارکوویتز ترسیم و نقطه با بیشترین نسبت شارپ تعیین می‌شود؛ نرخ بدون ریسک سالانه ۳ درصد و مقیاس‌گذاری روزانه نیز لحاظ می‌شود. نتیجه نهایی پرتفوی است با نسبت شارپ سالانه ۳/۱۵۸۵، بازده موردانتظار ۷۹۵۶٪ و نوسان ۲۴۲۴٪ که تمرکز وزنی آن بر CRM، PG و KO و سپس AAPL قرار می‌گیرد و سایر نمادها وزن‌های کوچک‌تری می‌گیرند. این جمع‌بندی نشان می‌دهد پیوند پیش‌بینی داده‌محور (با ترکیب ویژگی‌های فنی و احساسی) و بهینه‌سازی کلاسیک پرتفوی می‌تواند در افق کوتاه‌مدت به سببی متوازن، اجرایی و از نظر ریسک-بازده رقابتی منتهی شود.

بحث و نتیجه‌گیری

هدف اصلی این پژوهش، ارائه یک مدل یکپارچه برای پیش‌بینی بازده سهام و بهینه‌سازی پرتفوی با در نظرگیری احساسات سرمایه‌گذاران بود. نتایج نشان داد که ترکیب هم‌زمان ویژگی‌های فنی، داده‌های تاریخی و شاخص‌های احساسی استخراج‌شده از شبکه‌های اجتماعی توانست عملکرد مدل پیش‌بینی را به‌طور معناداری بهبود بخشد. مقادیر بالای ضریب تعیین برای برخی نمادها، به‌ویژه KO و COST، بیانگر آن است که مدل BGWO-XGBoost توانسته است روابط غیرخطی و پیچیده میان متغیرهای بازار را با دقت مناسبی شناسایی کند. این یافته با مطالعاتی که بر توانایی الگوریتم‌های یادگیری ماشین در مدل‌سازی رفتار بازار تأکید کرده‌اند همسو است (Lin & Marques, 2024; Qiu & Song, 2016). لین و مارکز بیان کردند که مدل‌های مبتنی بر هوش مصنوعی در مقایسه با رویکردهای سنتی، عملکرد بهتری در استخراج الگوهای پنهان بازار دارند و می‌توانند وابستگی‌های پیچیده میان متغیرها را شناسایی کنند (Lin & Marques, 2024). همچنین نتایج پژوهش یون و همکاران نشان داد که ترکیب XGBoost با الگوریتم‌های بهینه‌سازی می‌تواند دقت پیش‌بینی جهت حرکت قیمت سهام را افزایش دهد (Yun et al., 2021). بنابراین، یافته‌های پژوهش حاضر تأیید می‌کند که استفاده از الگوریتم‌های ترکیبی مبتنی بر یادگیری ماشین، رویکردی مؤثر برای تحلیل بازارهای مالی پیچیده محسوب می‌شود.

یافته‌های پژوهش نشان داد که ادغام شاخص‌های احساسی با داده‌های تکنیکال و تاریخی موجب بهبود عملکرد مدل شده است. این نتیجه بیانگر آن است که رفتار سرمایه‌گذاران و احساسات منتشرشده در شبکه‌های اجتماعی، دارای اطلاعات ارزشمندی درباره روندهای آینده بازار هستند. این یافته با مبانی نظری مالی رفتاری و پژوهش‌های انجام‌شده در این حوزه همخوانی دارد (Pompian, 2009; Raei & Fallahpour, 2004). بر اساس دیدگاه مالی رفتاری، سرمایه‌گذاران تحت تأثیر هیجانات، سوگیری‌های شناختی و واکنش‌های روان‌شناختی تصمیم‌گیری می‌کنند و همین مسئله موجب فاصله گرفتن قیمت‌ها از ارزش‌های بنیادی می‌شود (Eslami Bidgoli & Kordlouie, 2010).

همین راستا، پژوهش‌های فسلر و همکاران و گامبتی و جیوسبرتی نشان دادند که احساساتی مانند ترس، خشم و اضطراب می‌تواند رفتارهای ریسک‌پذیری و انتخاب‌داری‌ها را تغییر دهند (Fessler et al., 2004; Gambetti & Giusberti, 2012). بنابراین، نتایج این پژوهش نشان می‌دهد که ورود داده‌های احساسی به مدل‌های پیش‌بینی مالی می‌تواند بخشی از اطلاعات رفتاری پنهان بازار را آشکار سازد.

از دیگر یافته‌های مهم پژوهش، عملکرد مناسب داده‌های متنی و تحلیل احساسات شبکه‌های اجتماعی در پیش‌بینی بازده سهام بود. شاخص‌های مبتنی بر احساسات توانستند در کنار ویژگی‌های تکنیکال، نقش مهمی در افزایش دقت پیش‌بینی ایفا کنند. این نتیجه با مطالعات جدید حوزه پردازش زبان طبیعی و تحلیل احساسات مالی همسو است (Gu et al., 2025; Saravanos & Kanavos, 2025). گو و همکاران نشان دادند که استفاده از مدل‌های FinBERT-LSTM در تحلیل اخبار مالی و احساسات بازار می‌تواند دقت پیش‌بینی قیمت سهام را به‌طور معناداری افزایش دهد (Gu et al., 2024). همچنین ساراوانوس و کاناووس گزارش کردند که تحلیل احساسات رسانه‌های اجتماعی، ابزار مؤثری برای پیش‌بینی نوسانات بازار سهام است (Saravanos & Kanavos, 2025). در پژوهش حاضر نیز مشاهده شد که نمادهایی که حجم بیشتری از داده‌های متنی و تعاملات اجتماعی را در اختیار داشتند، الگوهای رفتاری واضح‌تری ارائه کردند و مدل توانست پیش‌بینی دقیق‌تری برای آن‌ها ارائه دهد. این مسئله نشان می‌دهد که داده‌های متنی و شبکه‌های اجتماعی به منبعی مهم برای تحلیل مالی تبدیل شده‌اند.

در بازار سرمایه ایران نیز مطالعات متعددی نقش احساسات سرمایه‌گذاران را در رفتار بازار تأیید کرده‌اند و یافته‌های پژوهش حاضر با این مطالعات همسو است. آقابابائی نشان داد که احساسات سرمایه‌گذاران می‌تواند نقدشوندگی و نوسانات بازار را تحت تأثیر قرار دهد (Aghababaei, 2022). حسینی و مرشدی نیز گزارش کردند که احساسات سرمایه‌گذاران بر پویایی معاملات بورس تهران اثرگذار است (Hosseini & Morshedi, 2019). همچنین دادگر و همکاران بیان کردند که احساسات بازار و رفتار دولت می‌تواند موجب تشدید نوسانات بورس شود (Dadgar et al., 2023). نتایج پژوهش حاضر نیز نشان داد که متغیرهای احساسی نقش مهمی در افزایش قدرت تبیین مدل دارند و حذف آن‌ها موجب افت عملکرد پیش‌بینی می‌شود. این موضوع بیانگر آن است که در بازارهای مالی، رفتار سرمایه‌گذاران تنها بر مبنای اطلاعات بنیادی شکل نمی‌گیرد، بلکه هیجانات جمعی و واکنش‌های رفتاری نیز بخش مهمی از ساختار تصمیم‌گیری را تشکیل می‌دهند.

نتایج مربوط به انتخاب ویژگی نیز از دیگر بخش‌های مهم پژوهش بود. الگوریتم گرگ خاکستری دودویی توانست از میان ۱۰۵ ویژگی اولیه، مجموعه‌ای بهینه و کم‌افزونگی از متغیرها را انتخاب کند و در نتیجه عملکرد مدل را بهبود بخشد. این یافته با مطالعات مرتبط با انتخاب ویژگی و بهینه‌سازی داده‌ها همخوانی دارد (Li et al., 2017; Miao & Niu, 2016). لی و همکاران بیان کردند که انتخاب ویژگی در داده‌های مالی با ابعاد بالا، نقش تعیین‌کننده‌ای در کاهش نویز و افزایش دقت مدل دارد (Li et al., 2017). همچنین میائو و نیو تأکید کردند که روش‌های انتخاب ویژگی می‌توانند پیچیدگی مدل را کاهش داده و از بیش‌برازش جلوگیری کنند (Miao & Niu, 2016). در همین راستا، استفاده از الگوریتم گرگ خاکستری در پژوهش حاضر باعث شد تا مدل بتواند متغیرهای مؤثرتر را شناسایی کرده و ساختاری پایدارتر ایجاد کند. این نتیجه با یافته‌های اماری و همکاران نیز همسو است که کارایی الگوریتم گرگ خاکستری را در مسائل انتخاب ویژگی تأیید کرده‌اند (Emary et al., 2016). همچنین نسخه‌های توسعه‌یافته این الگوریتم در پژوهش لیو و همکاران نشان داده‌اند که توانایی بالایی در جست‌وجوی فضای پیچیده و جلوگیری از گیر افتادن در بهینه‌های محلی دارند (Liu et al., 2023).

نتایج مربوط به مرحله بهینه‌سازی پرتفوی نیز نشان داد که استفاده از پیش‌بینی‌های مبتنی بر یادگیری ماشین می‌تواند به تشکیل سبدی با نسبت شارپ بالا و تعادل مناسب میان ریسک و بازده منجر شود. پرتفوی نهایی که تمرکز بیشتری بر نمادهای AAPL، KO، PG، CRM داشت، توانست بازده موردانتظار بالا را در کنار سطح مناسبی از نوسان ارائه دهد. این یافته نشان می‌دهد که ادغام پیش‌بینی داده‌محور با مدل‌های کلاسیک مالی می‌تواند عملکرد تصمیم‌گیری سرمایه‌گذاری را بهبود بخشد. این نتیجه با پژوهش شینگ و همکاران همسو است که نشان دادند استفاده از احساسات بازار در تخصیص دارایی می‌تواند منجر به تشکیل پرتفوی‌های هوشمندتر و کارا تر شود (Xing et al., 2018). همچنین شیانگ گزارش کرد که احساسات سرمایه‌گذاران بر عملکرد شرکت‌ها و رفتار سرمایه‌گذاری اثر مستقیم دارد و می‌تواند در تخصیص بهینه دارایی‌ها مورد استفاده قرار گیرد (Xiang, 2022). بنابراین، یافته‌های پژوهش حاضر نشان می‌دهد که ترکیب داده‌های احساسی با چارچوب میانگین-واریانس مارکویتز، رویکردی مؤثر برای ساخت پرتفوی‌های رقابتی در بازارهای پرنوسان است.

از منظر روش‌شناختی، یافته‌های پژوهش حاضر اهمیت استفاده از مدل‌های ترکیبی و چندلایه در تحلیل مالی را نشان می‌دهد. مطالعات سنتی معمولاً تنها بر داده‌های قیمتی یا شاخص‌های تکنیکال تمرکز می‌کردند، در حالی که پژوهش حاضر نشان داد ترکیب هم‌زمان داده‌های تاریخی، ویژگی‌های تکنیکال و احساسات بازار می‌تواند دید جامع‌تری از رفتار بازار ارائه دهد. این نتیجه با یافته‌های رضاییان و همکاران و اصغری و همکاران همسو است که بر لزوم استفاده از عوامل چندگانه در مدل‌سازی قیمت سهام تأکید کرده‌اند (Asghari et al., 2024; Rezaeian et al., 2024). همچنین پژوهش صالحی چگنی و همکاران نشان داد که یادگیری عمیق در تحلیل احساسات کاربران شبکه‌های سرمایه‌گذاری هوشمند، می‌تواند در استخراج الگوهای رفتاری بازار بسیار مؤثر باشد (Salehi Chegni et al., 2025). در مجموع، یافته‌های پژوهش حاضر بیانگر آن است که آینده تحلیل مالی به سمت مدل‌های هوشمند، داده‌محور و مبتنی بر تحلیل رفتار بازار حرکت می‌کند.

به‌طور کلی، نتایج پژوهش حاضر نشان داد که ترکیب الگوریتم‌های یادگیری ماشین، تحلیل احساسات و بهینه‌سازی پرتفوی می‌تواند شکاف میان مدل‌های نظری و رفتار واقعی بازار را کاهش دهد. مدل پیشنهادی توانست با استفاده از داده‌های چندمنبعی و تحلیل رفتار سرمایه‌گذاران، پرتفوی متوازن و دارای عملکرد مطلوب ارائه کند. این یافته نشان می‌دهد که بازارهای مالی امروز دیگر صرفاً بر مبنای داده‌های اقتصادی قابل تحلیل نیستند و متغیرهای رفتاری و احساسی به بخشی جدایی‌ناپذیر از فرآیند تصمیم‌گیری مالی تبدیل شده‌اند. از این‌رو، استفاده از مدل‌های هوشمند مبتنی بر تحلیل احساسات می‌تواند به توسعه نسل جدیدی از سیستم‌های مدیریت سرمایه‌گذاری منجر شود که قدرت سازگاری بیشتری با شرایط متغیر بازار دارند.

این پژوهش با وجود نتایج ارزشمند، دارای محدودیت‌هایی بود که باید در تفسیر یافته‌ها مورد توجه قرار گیرد. نخست، داده‌های مورد استفاده تنها مربوط به ۱۶ نماد منتخب بورس آمریکا در یک بازه زمانی مشخص بود و ممکن است تعمیم نتایج به سایر بازارها یا دوره‌های زمانی با احتیاط همراه باشد. دوم، داده‌های احساسی صرفاً از شبکه‌های اجتماعی استخراج شد و سایر منابع اطلاعاتی مانند اخبار رسمی، گزارش‌های تحلیلی و رسانه‌های تخصصی در مدل لحاظ نشدند. همچنین، رفتار کاربران شبکه‌های اجتماعی ممکن است تحت تأثیر ربات‌ها، اطلاعات نادرست یا رفتارهای هیجانی کوتاه‌مدت قرار گیرد که می‌تواند بر کیفیت شاخص‌های احساسی اثر بگذارد. علاوه بر این، اگرچه الگوریتم‌های یادگیری ماشین قدرت بالایی در مدل‌سازی روابط پیچیده دارند، اما همچنان با چالش‌هایی مانند بیش‌برازش و حساسیت به تنظیم فرآیندها مواجه هستند.

پیشنهاد می‌شود در پژوهش‌های آینده، مدل حاضر در بازارهای مالی مختلف و در دوره‌های زمانی طولانی‌تر مورد آزمون قرار گیرد تا پایداری عملکرد آن بررسی شود. همچنین استفاده از داده‌های چندمنبعی شامل اخبار مالی، گزارش‌های تحلیلی، داده‌های اقتصاد کلان و داده‌های رفتاری سرمایه‌گذاران می‌تواند به غنای مدل کمک کند. توسعه مدل‌های عمیق‌تر مبتنی بر ترنسفورمرها، شبکه‌های عصبی گراف و یادگیری تقویتی نیز می‌تواند مسیر مناسبی برای افزایش دقت پیش‌بینی باشد. علاوه بر این، بررسی اثر متغیرهای فرهنگی و روان‌شناختی سرمایه‌گذاران در بازارهای مختلف می‌تواند به توسعه مدل‌های بومی‌شده مدیریت سرمایه‌گذاری کمک کند. همچنین پیشنهاد می‌شود پژوهش‌های آینده، مقایسه‌ای میان الگوریتم‌های فراابتکاری مختلف در انتخاب ویژگی و بهینه‌سازی پرتفوی انجام دهند تا کاراترین روش‌ها شناسایی شوند.

نتایج این پژوهش می‌تواند برای مدیران سرمایه‌گذاری، شرکت‌های مشاوره مالی، صندوق‌های سرمایه‌گذاری و معامله‌گران حرفه‌ای کاربرد عملی داشته باشد. استفاده از سیستم‌های مبتنی بر تحلیل احساسات و یادگیری ماشین می‌تواند دقت تصمیم‌گیری مالی را افزایش داده و مدیریت ریسک را بهبود بخشد. همچنین شرکت‌های سرمایه‌گذاری می‌توانند با طراحی داشبوردهای هوشمند مبتنی بر تحلیل داده‌های شبکه‌های اجتماعی، تغییرات رفتاری بازار را سریع‌تر شناسایی کنند. به‌کارگیری مدل‌های ترکیبی مشابه پژوهش حاضر در سامانه‌های معاملات الگوریتمی نیز می‌تواند موجب واکنش سریع‌تر به تغییرات بازار و افزایش کارایی سرمایه‌گذاری شود. افزون بر این، نهادهای مالی و آموزشی می‌توانند با توسعه آموزش‌های مرتبط با مالی رفتاری، تحلیل احساسات و هوش مصنوعی، زمینه لازم برای استفاده گسترده‌تر از فناوری‌های نوین در بازار سرمایه را فراهم سازند.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

تشکر و قدردانی

از تمامی کسانی که در طی مراحل این پژوهش به ما یاری رساندند تشکر و قدردانی می‌گردد.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافی وجود ندارد.

حمایت مالی

این پژوهش حامی مالی نداشته است.

موازن اخلاقی

در انجام این پژوهش تمامی موازن و اصول اخلاقی رعایت گردیده است.

References

- Aghababaei, M. A. (2022). Investigating the Effect of Investor Sentiment on Market Liquidity and Its Volatility in the Tehran Stock Exchange. *Financial Research*, 25(2), 357-383.
- Asghari, M., Yazdani, N., Tabrizian, B., & Rahnamay Roudposhti, F. (2024). Stock Price Prediction Based on Fundamental, Technical, and Economic Factors. *Investment Knowledge*, 13(51), 1-22.
- Dadgar, Y., Dargahi, H., & Gholizadeh, S. (2023). The Role of Investor Sentiment and Government Behavior in Tehran Stock Exchange Volatility: A Behavioral Economics Approach. *Applied Theories of Economics*, 10(1), 191-214.
- Emary, E., Zawbaa, H. M., & Hassani, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, 172, 371-381.
- Eslami Bidgoli, G., & Kordlouie, H. (2010). Behavioral Finance: The Transition Stage from Standard Finance to Neurofinance. *Financial Engineering and Securities Management*, 1(1), 19-36.
- Fessler, D. M. T., Pillsworth, E. G., & Flansburg, T. J. (2004). Angry men and disgusted women: An evolutionary approach to the influence of emotions on risk taking. *Organizational Behavior and Human Decision Processes*, 95(1), 107-123. <https://doi.org/10.1016/j.obhdp.2004.06.006>
- Gambetti, E., & Giusberti, F. (2012). The effect of anger and anxiety traits on investment decisions. *Journal of Economic Psychology*, 33(6), 1059-1069. <https://doi.org/10.1016/j.joep.2012.07.001>
- Gu, W. J., Zhong, Y. H., Li, S. Z., Wei, C. S., Dong, L. T., Wang, Z. Y., & Yan, C. (2024). Predicting stock prices with FinBERT-LSTM: Integrating news sentiment analysis. Proceedings of the 2024 8th International Conference on Cloud and Big Data Computing.
- Hosseini, A., & Morshedi, F. (2019). The Effect of Investor Sentiment on Trading Dynamics of the Tehran Stock Exchange. *Financial Accounting and Auditing Research*, 11(44), 1-22.
- Li, J., Cheng, K., Wang, S., & Liu, H. (2017). Feature selection: A data perspective. *Acm Computing Surveys*, 50(6), 1-45.
- Lin, C. Y., & Marques, J. A. L. (2024). Stock market prediction using artificial intelligence: A systematic review of systematic reviews. *Social Sciences and Humanities Open*, 9, 100864.
- Liu, L., Li, L., Nian, H., Lu, Y., Zhao, H., & Chen, Y. (2023). Enhanced grey wolf optimization algorithm for mobile robot path planning. *Electronics*, 12(19), 4026.
- Miao, J., & Niu, L. (2016). A survey on feature selection. *Procedia Computer Science*, 91, 919-926.
- Pompian, M. M. (2009). *Behavioral Finance and Wealth Management* (A. Badri, Trans.). Kayhan.
- Qiu, M., & Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PLoS One*, 11(5), e0155133.
- Raei, R., & Fallahpour, S. (2004). Behavioral Finance: A Different Approach in Finance. *Financial Research*, 6(2), 77-106.
- Ramirez-Gallego, S., Lastra, I., & Alonso-Betanzos, A. (2017). Fast-mRMR. *International Journal of Intelligent Systems*, 32(2), 134-152.
- Rezaei-Kalidbari, H. R., Davari, A., Pournaserani, A., & Mohammadi-Almani, A. (2013). The Effect of Corporate Social Responsibility and Environmental Awareness on Enhancing Green Intellectual Capital: A Case Study of Industrial Units in Rasht Industrial City. *Public Management Perspective*(15), 119-138.

- Rezaeian, S., Taleghani, M., & Sharj Sharifi, A. (2024). Developing a Comprehensive Model for Stock Price Prediction in the Stock Exchange Market Using Interpretive Structural Modeling. *Asset Management and Financing*, 12(2), 39-58. <https://doi.org/10.22108/amf.2024.138983.1821>
- Salehi Chegini, J., Saba, & Salehi. (2025). Deep Learning in Sentiment Analysis of Users of Smart Investment Networks. *Intelligent Information Systems Quarterly*, 3(6), 21-36.
- Saravanos, C., & Kanavos, A. (2025). Forecasting stock market volatility using social media sentiment analysis. *Neural Computing and Applications*, 37(17), 10771-10794.
- Vakilifard, H. R., Saeedi, A., & Eftekhari-Aliabadi, A. (2013). Investigation and Analysis of Behavioral Reactions in the Tehran Stock Exchange. *Investment Knowledge*, 3(9), 223-244.
- Xiang, X. (2022). Investor sentiment, R&D spending and firm performance. *Economic Research-Ekonomska Istraživanja*, 35(1), 6257-6278. <https://doi.org/10.1080/1331677X.2022.2048193>
- Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Intelligent asset allocation via market sentiment views. *IEEE Computational intelligence magazine*, 13(4), 25-34.
- Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data. ICML,
- Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost. *Expert Systems with Applications*, 186, 115716.
- Zare Bahman-Miri, M. J., Nazari Shams-Abadi, M. J., & Beikri, A. (2022). Social Responsibility and Investor Sentiment with a Simultaneous Equations Approach. *Financial Accounting Research*, 12(1), 61-80.