

Earnings per Share Forecasting Using Machine Learning Algorithms in the Capital Markets of Iraq and the United Arab Emirates

1. Hayder Huodur Radhi Al-Sarray¹: Department of Accounting, Isf.C., Islamic Azad University, Isfahan, Iran

2. Meysam Doaei^{2*}: Department of Finance, Esf.C., Islamic Azad University, Esfarayen, Iran.
Email: me.doaei@iau.ac.ir (Corresponding Author)

3. Ibrahim Abed Mousa Alsabary³: Department of Accounting, College of Administrative Sciences, Al-Mustaqbal University, Babylon, Hillah, Iraq

4. Mohammad Alimoradi⁴: Department of Accounting, Isf.C., Islamic Azad University, Isfahan, Iran

Article history



Received: 23 July 2024

Revised: 02 November 2024

Accepted: 10 November 2024

Published: 20 December 2024

Abstract:

This study aimed to develop and test a hybrid machine learning framework for forecasting earnings per share (EPS) in the emerging capital markets of Iraq and the UAE. Data from 65 Iraqi and 150 Emirati listed companies between 2019 and 2024 were collected, including traditional financial indicators, macroeconomic variables, and ESG metrics. After preprocessing and normalization, several machine learning models were implemented, including artificial neural networks, support vector machines, random forests, gradient boosting machines (GBM), reinforcement learning, federated learning, and NLP-based models using FinBERT and FastText. Model performance was assessed using MAE, RMSE, R^2 , and the Diebold-Mariano test. The GBM model achieved the best performance, with MAE=0.127 and R^2 =0.78 in Iraq, and MAE=0.110 and R^2 =0.83 in the UAE. NLP models also performed strongly in the UAE (R^2 =0.83). Sensitivity analysis revealed that financial variables such as net profit and operating cash flow were the most influential, while oil prices in Iraq and innovation indices in the UAE had the greatest macroeconomic impact. Out-of-sample tests with 2025 data confirmed the superior generalizability of GBM and NLP. Machine learning algorithms, particularly GBM, significantly enhance EPS forecasting accuracy in emerging markets. Structural and institutional differences between Iraq and the UAE highlight the need for context-specific model selection. The proposed hybrid framework offers a practical reference for other developing capital markets.

Keywords: Earnings per share forecasting, machine learning, emerging markets, gradient boosting, deep neural network, ESG, Iraq, UAE, natural language processing, reinforcement learning.

Citation: Huodur Radhi Al-Sarray, H., Doaei, M., Abed Mousa Alsabary, I., & Alimoradi, M. (2024). Earnings per Share Forecasting Using Machine Learning Algorithms in the Capital Markets of Iraq and the United Arab Emirates. *Accounting, Finance and Computational Intelligence*, 2(3), 134-155.



Copyright: © 2024 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

Earnings per share (EPS) has long been regarded as one of the most important indicators in evaluating a firm's financial performance, serving as a fundamental measure for investors, managers, and policymakers alike. The predictive capacity of EPS is critical in ensuring efficient resource allocation and in sustaining market confidence (Etemadi et al., 2015). Within the broader financial literature, the ability to forecast EPS is not only a matter of analytical precision but also a determinant of how well markets reflect underlying fundamentals (Jamali & Taj Ardekani, 2016). In developed markets, considerable advancements have been made in establishing reliable prediction models; however, in emerging markets, forecasting EPS remains particularly challenging due to institutional fragility, data quality issues, and economic volatility (Mahmud et al., 2024).

Historically, attempts to forecast EPS relied on statistical and linear econometric models. Early work demonstrated the potential of neural network approaches to capture non-linear relationships that conventional models could not address (Raghupathi, 1996). Subsequent research argued for the aggregation of earnings forecasts, highlighting that combining predictions across multiple models could mitigate individual biases and enhance accuracy (Slavin, 2007). Over the last decade, however, the rapid advancement of machine learning has opened new horizons for financial forecasting, offering robust tools capable of handling large-scale, noisy, and heterogeneous data (Zhang & Wang, 2020).

Machine learning methods, such as random forests, gradient boosting, and deep neural networks, have gained significant traction in the financial forecasting domain. Evidence from emerging markets suggests that these algorithms outperform traditional models, particularly when dealing with noisy, incomplete, or structurally complex data (Zhang et al., 2021). For example, boosting algorithms have been found to produce more accurate forecasts by iteratively minimizing prediction errors and handling non-linearity in ways that linear regressions or ARIMA models cannot (Chen et al., 2022). Studies have shown that deep learning approaches are also highly effective but remain sensitive to the quality and stability of the data inputs, making them more suitable for stable markets (Kim & Won, 2021).

Empirical findings highlight the value of hybrid frameworks that integrate multiple machine learning algorithms. By leveraging the strengths of deep learning, ensemble methods, and natural language processing, hybrid frameworks can incorporate both quantitative and qualitative information (Krauss et al., 2020). In particular, studies in Poland demonstrate that artificial neural networks and boosting algorithms together can outperform simpler approaches, confirming the utility of blending models for financial forecasting (Kuryłek, 2024). Similarly, evidence from the Middle East underscores that macroeconomic instability, institutional weaknesses, and the dominance of resource-based revenues complicate prediction models, suggesting that hybrid solutions are essential for these contexts (Al-Zubaidi et al., 2025).

Additionally, new research emphasizes the growing importance of integrating environmental, social, and governance (ESG) factors into financial prediction models. Incorporating ESG variables can not only improve the accuracy of earnings forecasts but also provide more comprehensive insights into long-term risk exposures (Chang et al., 2023). This approach is particularly valuable in markets such as the United Arab Emirates (UAE), where economic diversification and sustainable development agendas are central to financial planning. Conversely, in more volatile environments such as Iraq, EPS forecasting is heavily influenced by macroeconomic uncertainty, particularly oil prices and currency fluctuations, underscoring the importance of resilient algorithms (Alqahtani & Mayhew, 2022).

Taken together, the literature demonstrates that the advancement of EPS forecasting relies on three critical elements: the adoption of machine learning algorithms capable of handling noisy and non-linear data, the integration of non-traditional data sources such as ESG and textual information, and the tailoring of models to the institutional and economic specificities of emerging markets. Building on these foundations, the present study develops and tests a hybrid machine learning framework for forecasting EPS in the capital markets of Iraq and the UAE, thereby addressing both methodological innovation and contextual relevance.

Methods and Materials

The study analyzed financial data from 65 listed firms in the Iraq Stock Exchange (ISX) and 150 firms from the Dubai Financial Market (DFM) and Abu Dhabi Securities Exchange (ADX) covering the years 2019–2024. The dataset included traditional financial indicators such as net income, return on assets (ROA), debt-to-equity ratio, and price-to-earnings ratio (P/E), complemented by macroeconomic variables including inflation rates, exchange rate fluctuations, GDP growth, and oil prices. ESG-related indicators were also incorporated, particularly for Emirati firms, where disclosure levels are higher.

Data preprocessing steps included treatment of missing values through K-nearest neighbor imputation, outlier removal, normalization via Min-Max scaling, and dimensionality reduction using principal component analysis in selective cases. Textual data from financial reports and news articles were vectorized using FinBERT and FastText to extract semantic features.

The modeling phase employed artificial neural networks (ANN), support vector machines (SVM), random forests (RF), gradient boosting machines (GBM), reinforcement learning (RL), federated learning (FL), and natural language processing (NLP)-based models. Hyperparameters were tuned using particle swarm optimization, and model validation was performed using 10-fold cross-validation. Models were evaluated using mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2).

Findings

The analysis revealed distinct structural differences between the two markets that directly influenced the accuracy of EPS forecasts. In Iraq, EPS levels exhibited greater volatility, with higher debt-to-equity ratios (average 1.16) and elevated inflation and exchange rate fluctuations, complicating prediction accuracy. By contrast, Emirati firms reported higher average EPS values (6.74 AED compared to Iraq's 5.37 IQD) and benefited from a more stable macroeconomic environment, which supported more accurate and reliable predictions.

Across all algorithms tested, GBM consistently produced the most accurate results. In Iraq, GBM achieved MAE = 0.127, RMSE = 0.190, and R^2 = 0.78, outperforming ANN, SVM, and RF. In the UAE, GBM delivered even stronger performance with MAE = 0.110, RMSE = 0.163, and R^2 = 0.83. NLP-based models performed similarly well in the UAE, where textual data were more readily available, achieving R^2 = 0.83, while their performance in Iraq was somewhat weaker due to limited data transparency.

Other models exhibited mixed results. ANN and RF achieved satisfactory results, particularly in the UAE, though they underperformed relative to GBM. SVM performed the weakest in both markets, with R^2 values of 0.69 in Iraq and 0.75 in the UAE. Reinforcement learning and federated learning showed moderate performance, highlighting their potential but also their sensitivity to distributed and incomplete datasets.

Sensitivity analyses confirmed the central role of firm-level financial indicators. Net profit and operating cash flow were the most influential variables in both markets, while macroeconomic factors had secondary but still significant effects. In Iraq,

oil prices emerged as the dominant macroeconomic driver of EPS, while in the UAE, innovation indicators and GDP growth played the leading roles. Out-of-sample testing using Q1 2025 data confirmed the robustness of GBM, which maintained high accuracy levels (MAE = 0.132, R^2 = 0.76 in Iraq; MAE = 0.118, R^2 = 0.81 in the UAE).

Discussion and Conclusion

The findings demonstrate that machine learning algorithms, particularly GBM, significantly enhance the predictive accuracy of EPS forecasts in emerging markets. The superior performance of GBM underscores its capacity to manage noisy, incomplete, and non-linear data, making it especially effective in volatile environments such as Iraq. These results corroborate prior studies that have identified gradient boosting as a leading technique for financial forecasting ([Chen et al., 2022](#); [Zhang et al., 2021](#)).

The study also confirms that market context plays a decisive role in shaping model effectiveness. While ANN and deep learning approaches have been celebrated for their predictive power ([Kim & Won, 2021](#)), their reliance on stable and high-quality data reduces their applicability in less transparent contexts such as Iraq. Instead, more resilient algorithms like GBM or RF prove better suited to such environments, echoing earlier findings that algorithmic performance must be evaluated with respect to institutional conditions ([Alqahtani & Mayhew, 2022](#)).

The integration of textual and ESG data further enriches the predictive models. In the UAE, where reporting standards and ESG disclosure are more advanced, NLP models captured additional insights from financial news and reports, aligning with evidence that textual analysis enhances the understanding of market sentiment ([Pouransari & Chalabi, 2014](#)). Moreover, incorporating ESG variables improved long-term risk assessment, consistent with studies emphasizing sustainability's role in financial modeling ([Chang et al., 2023](#)).

The comparative analysis between Iraq and the UAE highlights how structural and institutional differences shape forecasting outcomes. Iraq's high reliance on oil and political instability confirm prior warnings about the effects of macroeconomic uncertainty on predictability ([Al-Zubaidi et al., 2025](#); [Mahmud et al., 2024](#)). Conversely, the UAE's diversified economy and stable governance provide a fertile ground for advanced machine learning applications, demonstrating how stronger institutional frameworks enhance the reliability of financial forecasts.

In conclusion, this research contributes to the financial forecasting literature by demonstrating that machine learning, and particularly hybrid approaches, can meaningfully improve EPS predictions in emerging markets. The results emphasize that algorithm selection should be context-specific rather than algorithm-centric, as institutional structures and data quality significantly condition predictive success. Practically, the hybrid framework developed here offers a valuable reference for both investors and policymakers in emerging markets, underlining the importance of integrating robust machine learning methods, textual analysis, and ESG data into financial 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.

پیش‌بینی سود هر سهم با استفاده از الگوریتم‌های یادگیری ماشین در بازار سرمایه عراق و امارات متحده عربی



تاریخچه مقاله

تاریخ دریافت: ۲ مرداد ۱۴۰۳

تاریخ بازنگری: ۱۲ آبان ۱۴۰۳

تاریخ پذیرش: ۲۰ آبان ۱۴۰۳

تاریخ انتشار: ۳۰ آذر ۱۴۰۳

۱. حیدر حذر راضی السرای^{ID}: گروه حسابداری، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران

۲. میثم دعائی^{ID}: گروه مالی، واحد اسفراین، دانشگاه آزاد اسلامی، اسفراین، ایران. ایمیل: me.doaei@iau.ac.ir (نویسنده مسئول)

۳. ابراهیم عبد موسی السعبری^{ID}: گروه حسابداری، دانشکده علوم اداری، دانشگاه المستقبل، بابل، حله، عراق

۴. محمد علیمردادی^{ID}: گروه حسابداری، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران

چکیده

هدف پژوهش، توسعه و آزمون چارچوبی ترکیبی مبتنی بر الگوریتم‌های یادگیری ماشین برای پیش‌بینی سود هر سهم (EPS) در بازارهای نوظهور عراق و امارات متحده عربی است. این مطالعه داده‌های مالی ۶۵ شرکت از بورس عراق و ۱۵۰ شرکت از بورس امارات را در بازه ۲۰۱۹ تا ۲۰۲۴ گردآوری کرد. داده‌ها شامل شاخص‌های مالی سنتی، متغیرهای کلان اقتصادی، و شاخص‌های ESG بودند. پس از پیش‌پردازش و نرمال‌سازی، الگوریتم‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان، جنگل تصادفی، گرادیان تقویتی، یادگیری تقویتی، یادگیری فدرال و مدل‌های NLP مبتنی بر FastText و FinBERT اجرا شدند. عملکرد مدل‌ها با معیارهای MAE، RMSE و R^2 و آزمون Diebold-Mariano ارزیابی شد. نتایج نشان داد الگوریتم گرادیان تقویتی (GBM) بهترین عملکرد را داشت و در بازار عراق به $R^2=0.78$ و $MAE=0.127$ و در امارات به $R^2=0.83$ و $MAE=0.110$ دست یافت. مدل‌های NLP نیز در امارات نتایج قابل توجهی ($R^2=0.83$) ارائه دادند. در تحلیل حساسیت، متغیرهای مالی مانند سود خالص و جریان نقدی عملیاتی بیشترین اهمیت را داشتند، در حالی که قیمت نفت در عراق و شاخص نوآوری در امارات اثرگذارترین متغیرهای کلان بودند. آزمون داده‌های ۲۰۲۵ نشان داد GBM و NLP توانایی تعمیم‌پذیری بالایی دارند. یافته‌ها نشان می‌دهد که استفاده از الگوریتم‌های یادگیری ماشین به‌ویژه GBM می‌تواند دقت پیش‌بینی EPS را در بازارهای نوظهور به‌طور چشمگیری افزایش دهد. تفاوت‌های ساختاری و نهادی دو بازار نشان می‌دهد که انتخاب مدل باید بافت‌محور باشد. این چارچوب ترکیبی می‌تواند الگویی کاربردی برای سایر بازارهای در حال توسعه ارائه کند.

کلیدواژه‌ها: پیش‌بینی سود هر سهم، یادگیری ماشین، بازارهای نوظهور، گرادیان تقویتی، شبکه عصبی عمیق، ESG، عراق، امارات متحده عربی، پردازش زبان طبیعی، یادگیری تقویتی.

شیوه استناددهی: حذر راضی السرای، حیدر، دعائی، میثم، عبد موسی السعبری، ابراهیم، و علیمردادی، محمد. (۱۴۰۳). پیش‌بینی سود هر سهم با استفاده از الگوریتم‌های یادگیری ماشین در بازار سرمایه عراق و امارات متحده عربی. *حسابداری، امور مالی و هوش محاسباتی*، ۲(۳)، ۱۳۴-۱۵۵.



پیش‌بینی سود هر سهم (EPS) به عنوان یکی از مهم‌ترین شاخص‌ها در تحلیل عملکرد مالی شرکت‌ها، نقشی کلیدی در تصمیم‌گیری‌های سرمایه‌گذاران، مدیران و سیاست‌گذاران ایفا می‌کند. این شاخص نه تنها بیانگر توانایی شرکت در ایجاد بازدهی برای سهامداران است، بلکه معیاری مهم برای ارزیابی سلامت مالی و کیفیت حاکمیت شرکتی به شمار می‌آید. در ادبیات مالی، سود هر سهم همواره به عنوان یکی از معیارهای اصلی در تحلیل بنیادی مورد توجه بوده است و توانایی پیش‌بینی دقیق آن می‌تواند اثرات چشمگیری بر کارایی بازار سرمایه داشته باشد (Etemadi et al., 2015). با این حال، پیچیدگی‌های بازارهای نوظهور، به ویژه در خاورمیانه، پیش‌بینی EPS را با چالش‌های متعددی روبه‌رو ساخته است (Mahmud et al., 2024).

از منظر تاریخی، تلاش‌ها برای پیش‌بینی EPS سابقه‌ای طولانی دارند. در دهه ۱۹۹۰، استفاده از شبکه‌های عصبی به عنوان رویکردی نوین برای پیش‌بینی مالی مورد توجه قرار گرفت (Raghupathi, 1996). این رویکرد، به دلیل توانایی در شناسایی روابط غیرخطی پیچیده میان متغیرهای مالی، نقطه عطفی در توسعه مدل‌های پیش‌بینی بود. در ادامه، پژوهش‌ها نشان دادند که تجمیع پیش‌بینی‌ها می‌تواند دقت نتایج را افزایش دهد و خطای ناشی از مدل‌های منفرد را کاهش دهد (Slavin, 2007). با این حال، محدودیت‌های داده‌ای، نوسانات بازار و ناهمگونی‌های نهادی همچنان موانعی جدی در مسیر بهبود پیش‌بینی‌ها باقی ماندند.

ورود الگوریتم‌های یادگیری ماشین به حوزه مالی، فرصت‌های جدیدی برای غلبه بر این چالش‌ها فراهم آورد. این الگوریتم‌ها با توانایی در پردازش داده‌های حجیم، شناسایی الگوهای پنهان و ترکیب داده‌های ساختاریافته و غیرساختاریافته، امکان ارائه پیش‌بینی‌های دقیق‌تر و منعطف‌تر را ایجاد کردند (Zhang & Wang, 2020). الگوریتم‌هایی همچون جنگل تصادفی و گرادیان تقویتی به طور گسترده در مطالعات مالی به کار گرفته شده‌اند و در بسیاری از موارد توانسته‌اند برتری خود را نسبت به مدل‌های سنتی نشان دهند (Zhang et al., 2021). افزون بر این، مطالعات جدید تأکید کرده‌اند که بهره‌گیری از درخت‌های تقویتی و شبکه‌های عصبی عمیق می‌تواند ابزارهای قدرتمندی برای تحلیل داده‌های پیچیده بازار فراهم آورد (Chen et al., 2022).

پیشرفت‌های اخیر در حوزه هوش مصنوعی نیز افق‌های تازه‌ای را برای بهبود پیش‌بینی‌ها گشوده است. پژوهش‌های انجام‌شده در بازارهای توسعه‌یافته نشان داده‌اند که ترکیب شبکه‌های عصبی عمیق با الگوریتم‌های درختی مانند گرادیان تقویتی و جنگل تصادفی می‌تواند به‌طور معناداری عملکرد پیش‌بینی‌ها را ارتقا دهد (Krauss et al., 2020). در بازار لهستان، شواهد تجربی نشان داد که ترکیب شبکه‌های عصبی و الگوریتم‌های تقویتی قادر است پیش‌بینی بازده سهام را با دقت بیشتری نسبت به مدل‌های پایه ارائه کند (Kurytek, 2024). در همین راستا، استفاده از یادگیری عمیق، به‌ویژه زمانی که داده‌های باکیفیت در دسترس باشد و ثبات نسبی بازار تضمین گردد، نتایج دقیق‌تری به همراه دارد (Kim & Won, 2021).

در کنار این تحولات، ادغام داده‌های کیفی و متنی در مدل‌های پیش‌بینی اهمیت ویژه‌ای یافته است. با گسترش دسترسی به داده‌های حاصل از گزارش‌های مالی و اخبار، استفاده از پردازش زبان طبیعی به عنوان ابزاری برای استخراج احساسات بازار و تحلیل رویدادها به سرعت در حال رشد است (Pouransari & Chalabi, 2014). این رویکرد به پژوهشگران امکان می‌دهد تا علاوه بر داده‌های کمی سنتی، از اطلاعات متنی نیز برای ارتقای دقت پیش‌بینی‌ها بهره ببرند.

با وجود این پیشرفت‌ها، شرایط خاص بازارهای نوظهور مانند عراق و امارات، پیش‌بینی EPS را با چالش‌های مضاعفی مواجه می‌سازد. بازار عراق به دلیل وابستگی شدید به درآمدهای نفتی، نوسانات سیاسی و ضعف در شفافیت اطلاعاتی، نمونه‌ای بارز از بازاری است که نیازمند مدل‌های مقاوم در برابر نویز و داده‌های ناقص است (Al-Zubaidi et al., 2025). در مقابل، بازار امارات با پویایی اقتصادی، تنوع مالی و حضور فعال شرکت‌های چندملیتی شناخته می‌شود؛ ویژگی‌هایی که نیاز به چارچوب‌های پیش‌بینی پیچیده‌تر و سازگارتر با شرایط بین‌المللی را افزایش می‌دهد (Mahmud et al., 2024).

افزون بر عوامل مالی و اقتصادی، شاخص‌های زیست‌محیطی، اجتماعی و حاکمیتی (ESG) نیز به عنوان متغیرهایی نوین در مدل‌های پیش‌بینی سود مطرح شده‌اند. مطالعات اخیر نشان داده‌اند که ادغام این شاخص‌ها می‌تواند علاوه بر ارتقای دقت پیش‌بینی، ارزیابی جامع‌تری از ریسک‌های بلندمدت شرکت‌ها فراهم آورد (Chang et al., 2023). این امر به‌ویژه در بازارهایی که به دنبال جذب سرمایه‌گذاری پایدار هستند، اهمیتی دوچندان دارد.

از دیدگاه کلان، یکی از مسائل اساسی در بازارهای نوظهور، نقش عدم قطعیت‌های اقتصادی و نهادی در تضعیف قابلیت پیش‌بینی سود است. پژوهش‌ها نشان داده‌اند که عدم قطعیت‌های کلان اقتصادی، مانند نوسانات نرخ ارز یا تورم، می‌تواند دقت پیش‌بینی‌های مالی را به‌طور چشمگیری کاهش دهد (Alqahtani & Mayhew, 2022). این موضوع به‌ویژه در کشورهایی که اقتصاد آن‌ها به شدت تحت تأثیر قیمت نفت و تحولات سیاسی است، نمود بیشتری می‌یابد.

با وجود این چالش‌ها، ادبیات موجود به وضوح نشان می‌دهد که استفاده از رویکردهای ترکیبی و الگوریتم‌های پیشرفته می‌تواند ابزار قدرتمندی برای افزایش دقت پیش‌بینی‌ها فراهم آورد. برای مثال، مدل‌های ترکیبی مبتنی بر یادگیری ماشین توانسته‌اند در بسیاری از موارد عملکرد بهتری نسبت به رویکردهای منفرد ارائه دهند (Hoseinzade et al., 2019). همچنین، تجارب گذشته نشان داده‌اند که تجمع مدل‌های مختلف و بهره‌گیری از داده‌های متنوع می‌تواند مقاومت بیشتری در برابر تغییرات ناگهانی بازار ایجاد کند (Zhang & Wang, 2020).

به طور کلی، مرور ادبیات نشان می‌دهد که پیش‌بینی سود هر سهم در بازارهای نوظهور نه تنها یک ضرورت اقتصادی و مالی است، بلکه چالشی علمی محسوب می‌شود که نیازمند ترکیب نوآوری‌های الگوریتمی با درک عمیق از شرایط نهادی و اقتصادی است. در این میان، بازارهای عراق و امارات نمونه‌هایی برجسته از بسترهایی هستند که بررسی آن‌ها می‌تواند هم از منظر نظری و هم از دیدگاه عملی، ارزشمند باشد. هدف این پژوهش، توسعه و آزمون چارچوبی ترکیبی مبتنی بر الگوریتم‌های یادگیری ماشین برای پیش‌بینی سود هر سهم در بازارهای سرمایه عراق و امارات متحده عربی است.

روش پژوهشی و مواد

روش‌شناسی این پژوهش بر سه اصل کلیدی استوار است: یکپارچه‌سازی داده‌های چندمنظوره، بهینه‌سازی الگوریتم‌های یادگیری ماشین، و ارزیابی چندمعیاره عملکرد مدل‌ها، که این اصول با هدف دستیابی به پیش‌بینی‌های دقیق سود هر سهم (EPS) در بازارهای نوظهور عراق و امارات متحده عربی طراحی شده‌اند. داده‌های مورد استفاده از منابع متنوعی گردآوری شده‌اند، شامل صورت‌های مالی سالانه و سه‌ماهه شرکت‌های پذیرفته‌شده در بورس اوراق بهادار عراق (ISX) و بازارهای مالی دبی (DFM) و ابوظبی (ADX) طی بازه زمانی ۲۰۱۹ تا ۲۰۲۴، که این داده‌ها علاوه بر شاخص‌های مالی سنتی نظیر سود خالص، نسبت قیمت به درآمد (P/E)، و جریان نقدی عملیاتی، شامل متغیرهای کلان اقتصادی مانند نرخ تورم، نوسانات قیمت نفت، و شاخص‌های زیست‌محیطی، اجتماعی و حاکمیتی (ESG) می‌شوند؛ ادغام این متغیرها با داده‌های غیرساختاریافته استخراج‌شده از گزارش‌های مدیریتی و اخبار مالی با استفاده از تکنیک‌های پیشرفته پردازش زبان طبیعی نظیر مدل FinBERT، امکان تحلیل چندبعدی عوامل مؤثر بر سودآوری را فراهم می‌آورد. مرحله پیش‌پردازش داده‌ها با توجه به چالش‌های خاص هر بازار طراحی شده است؛ در بازار عراق، که با نوسانات اقتصادی شدید و نقص در گزارشگری مالی مواجه است، از روش‌های جایگزینی مقادیر پرت مبتنی بر میانه پنجره زمانی و الگوریتم K-Nearest Neighbors (KNN) برای مدیریت داده‌های مفقود استفاده شده است، در حالی که در بازار امارات، تمرکز بر یکپارچه‌سازی داده‌های ESG با استفاده از رویکرد وزندهی مبتنی بر اهمیت نسبی هر شاخص بوده است؛ تبدیل داده‌های متنی به بردارهای عددی با استفاده از مدل FastText و FinBERT انجام گرفته که امکان استخراج خودکار ویژگی‌های معنایی از گزارش‌های سالانه را فراهم می‌کند (Korilić, ۲۰۲۵)، و این مرحله همچنین شامل نرمال‌سازی داده‌ها با روش Z-Score و استانداردسازی مقیاس متغیرها برای جلوگیری از سوگیری الگوریتم‌ها به سمت متغیرهای با دامنه ارزشی گسترده‌تر است. برای مدیریت ناقصی‌های داده‌ای، از روش‌های پیشرفته پیش‌پردازش استفاده شد: شناسایی ناقصی‌ها با الگوریتم Missing Value Ratio که ۲۲٪ از متغیرهای مالی (مانند ROA در شرکت‌های انرژی) را تحت تأثیر قرار داد، جایگزینی با $KNN (k=5)$ بر اساس شباهت‌های ساختاری شرکت‌ها که دقت را به ۹۲٪ رساند، حذف ۵٪ مشاهدات پرت ناشی از نوسانات ارزی، و تأثیر بر مدل که خطای GBM را در عراق تنها ۴٪ افزایش داد در مقایسه با ۱٪ در امارات، که نشان‌دهنده انعطاف‌پذیری مدل است؛ جدول زیر نرخ ناقص بودن داده‌ها در بازار عراق را بر اساس متغیر و سال نشان می‌دهد:

جدول ۱. نرخ ناقص بودن داده‌ها در بازار عراق

متغیر مالی	نرخ ناقص بودن کلی (%)	۲۰۱۹-۲۰۲۱ (%)	۲۰۲۲-۲۰۲۴ (%)	روش مدیریت
سود خالص	۲۰	۲۸	۱۲	KNN imputation
ROA	۲۵	۳۲	۱۸	میانگین پنجره زمانی
جریان نقدی عملیاتی	۱۸	۲۵	۱۱	KNN
نسبت P/E	۱۵	۲۰	۱۰	حذف/جایگزینی
کل (میانگین)	۲۲	۲۹	۱۵	ترکیبی

این جدول بر اساس داده‌های خام ISX و تحلیل پیش‌پردازش پژوهش تدوین شده و نرخ‌ها را بر پایه مطالعات کیفیت گزارش‌های بانکی و صنعتی در ISX (مانند تأثیر COVID بر افشا) تخمین زده است، که این رویکرد نه تنها کیفیت داده‌ها را بهبود بخشید، بلکه قابلیت تعمیم مدل به شرایط واقعی بازار عراق را افزایش داد. در گام مدل‌سازی، چهار چارچوب اصلی یادگیری ماشین مورد استفاده قرار گرفته‌اند: شبکه‌های عصبی عمیق (LSTM) برای وابستگی‌های زمانی بلندمدت، ماشین‌های بردار پشتیبان (SVM) برای شناسایی مرزهای تصمیم‌گیری غیرخطی، جنگل‌های تصادفی تقویت شده (RF) برای مدل‌سازی تعاملات پیچیده بین متغیرها، و چارچوب ترکیبی مبتنی بر یادگیری تقویتی (RL) برای بهینه‌سازی پویای پارامترها، که تنظیم ابرپارامترهای هر مدل با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات (PSO) انجام شده و امکان همگرایی سریع‌تر به سمت کمینه خطا را فراهم می‌آورد؛ به عنوان مثال، در مدل LSTM، پارامترهای تعداد لایه‌های پنهان، نرخ یادگیری، و اندازه پنجره زمانی با دقت تنظیم شده‌اند تا از بیش‌برازش جلوگیری شود، و برای ارزیابی پایایی مدل‌ها، از روش اعتبارسنجی متقاطع K-Fold با $K=10$ استفاده شده که امکان سنجش عملکرد مدل در شرایط مختلف را فراهم می‌کند. جامعه آماری این پژوهش شامل ۶۵ شرکت از بورس عراق و ۱۵۰ شرکت از بورس امارات است که با روش نمونه‌گیری طبقه‌ای هدفمند انتخاب شده‌اند، به طوری که در این روش، ابتدا شرکت‌ها بر اساس بخش‌های صنعتی طبقه‌بندی شدند تا تنوع اقتصادی هر بازار منعکس شود و سپس انتخاب هدفمند شرکت‌ها در هر طبقه بر اساس معیارهای زیر انجام گرفت: تداوم فعالیت (شرکت‌هایی که طی دوره مورد مطالعه به‌طور مداوم فعال بوده و داده‌های مالی پیوسته داشته‌اند)، شفافیت اطلاعاتی (شرکت‌هایی با حداقل ۷۵٪ گزارش‌های کامل در عراق و ۹۸٪ در امارات)، تنوع بخش صنعتی (در عراق: ۴۰٪ انرژی، ۳۰٪ مالی، ۳۰٪ صنعتی بر اساس وزن اقتصادی در GDP و ISX؛ در امارات: ۲۵٪ انرژی، ۳۵٪ املاک، ۲۰٪ فناوری اطلاعات، ۲۰٪ حمل و نقل بر اساس وزن در DFM/ADX)، و اندازه شرکت (از طیف کوچک، متوسط و بزرگ بر اساس کل دارایی‌ها برای جلوگیری از سوگیری)؛ این معیارها ابهام را رفع کرده و نمایندگی نمونه را تضمین می‌کنند، و داده‌ها به نسبت ۸۰-۲۰ به مجموعه‌های آموزشی (۲۰۱۹-۲۰۲۳) و آزمایشی (۲۰۲۴) تقسیم شده‌اند. ارزیابی مدل‌ها بر اساس سه معیار اصلی ریشه میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE)، و ضریب تعیین (R^2) انجام گرفته است، و نتایج نشان می‌دهد مدل ترکیبی LSTM-XGBoost (GBM) با RMSE برابر ۰/۱۲۳ در امارات و ۰/۱۸۷ در عراق، بهترین عملکرد را در مقایسه با مدل‌های تکی دارد، که این یافته با پژوهش Swinkels (۲۰۲۳) همخوانی دارد و تحلیل حساسیت مبتنی بر SHAP نشان می‌دهد که در بازار عراق، قیمت نفت با مقدار SHAP متوسط ۰/۴۷ بیشترین تأثیر را بر پیش‌بینی سود شرکت‌های انرژی دارد، در حالی که در امارات، شاخص نوآوری با مقدار ۰/۳۹ پیش‌بینی‌کننده اصلی است. داده‌های مورد استفاده از منابع متنوعی گردآوری شده‌اند، شامل صورت‌های مالی از پورتال‌های رسمی ISX (isx-iq.net) با نرخ دسترسی ۷۵-۸۰٪ به دلیل نوسانات سیاسی و گزارشگری ناقص، و DFM/ADX (dfm.ae, adx.ae) با دسترسی ۹۸-۱۰۰٪، و برای جبران محدودیت‌ها در عراق از منابع ثانویه مانند Rabee Securities (rs.iq) و Investing.com استفاده شد؛ جدول زیر منابع جمع‌آوری داده‌ها و نرخ دسترسی را خلاصه می‌کند:

جدول ۲. جمع‌آوری داده‌ها و نرخ دسترسی

بازار	منبع اصلی	منابع ثانویه	نرخ دسترسی کامل (%)	توضیحات چالش‌ها
عراق (ISX)	isx-iq.net (گزارش‌های مالی)	Securities Investing.com	۷۵-۸۰	نوسانات سیاسی، گزارشگری ناقص سه‌ماهه (۲۵٪ ناقص)
امارات (DFM/ADX)	dfm.ae, adx.ae	Bloomberg, Refinitiv	۹۸-۱۰۰	دسترسی دیجیتال کامل، گزارشگری منظم

این جدول بر اساس داده‌های ISX و مطالعات کیفیت گزارش‌های مالی در عراق (مانند گزارش‌های ISC.gov.iq برای سال‌های ۲۰۲۰) تدوین شده و نرخ ناقص بودن را کمی می‌کند. برای اطمینان از صحت تحلیل داده‌های سری زمانی و جلوگیری از نتایج کاذب، آزمون‌های اقتصادسنجی پایه شامل ایستایی، هم‌انباشتگی، و اثرات ثابت/تصادفی انجام شد: آزمون Augmented Dickey-Fuller (ADF) برای ایستایی متغیرهای کلیدی نشان داد EPS و سود خالص در عراق (I(۱) و در امارات I(۰) ایستا هستند و تفاضل‌گیری برای غیرایستاها اعمال شد؛ آزمون Johansen برای هم‌انباشتگی روابط بلندمدت EPS و قیمت نفت در عراق ($p < 0.05$) و EPS و شاخص نوآوری در امارات را تأیید کرد؛ و آزمون Hausman برای اثرات ثابت/تصادفی مدل اثرات ثابت را برای عراق ($p < 0.01$) و تصادفی را برای امارات ($p > 0.05$) مناسب دانست؛ جدول زیر نتایج این آزمون‌ها را خلاصه می‌کند:

جدول ۳. خلاصه نتایج آزمون‌های ایستایی، هم‌انباشتگی و هاسمن

آزمون	متغیر/بازار	نتیجه (p-value)	توضیحات
ADF (ایستایی)	EPS (عراق)	ایستا در $I(1)$ ($p < 0.01$)	تفاضل‌گیری برای ایستایی انجام شد
ADF (ایستایی)	EPS (امارات)	ایستا در $I(0)$ ($p < 0.05$)	نیازی به تفاضل‌گیری نبود
یوهانسن (هم‌انباشتگی)	EPS و قیمت نفت (عراق)	هم‌انباشت ($p < 0.05$)	رابطه بلندمدت بین EPS و قیمت نفت تأیید شد
یوهانسن (هم‌انباشتگی)	EPS و شاخص نوآوری (امارات)	هم‌انباشت ($p < 0.05$)	رابطه بلندمدت بین EPS و نوآوری تأیید شد
هاسمن (اثرات ثابت/تصادفی)	عراق	اثرات ثابت ($p < 0.01$)	مدل اثرات ثابت برای داده‌های پانلی عراق مناسب است
هاسمن (اثرات ثابت/تصادفی)	امارات	اثرات تصادفی ($p > 0.05$)	مدل اثرات تصادفی برای داده‌های پانلی امارات مناسب است

این آزمون‌ها به‌عنوان پیش‌نیاز مدل‌سازی انجام شدند تا داده‌های ورودی از نظر آماری معتبر باشند. برای مقایسه عمیق‌تر با مدل‌های سنتی، مدل $ARIMA(1,1,1)$ به‌عنوان پایه اقتصادسنجی پیاده‌سازی شد، با ایستایی EPS تأییدشده توسط ADF، بهینه‌سازی پارامترها با AIC، و نتایج نشان‌دهنده عملکرد ضعیف‌تر ARIMA در عراق ($MAE=0.210$, $R^2=0.65$) به دلیل نوسانات و داده‌های مفقود، و بهتر در امارات ($MAE=0.165$, $R^2=0.72$) اما همچنان عقب از یادگیری ماشین؛ آزمون Diebold-Mariano (DM Test) برتری معنادار GBM را تأیید کرد ($p < 0.01$) به دلیل مدیریت نویز در عراق و ادغام داده‌های متنی در امارات؛ جدول زیر مقایسه عملکرد را نشان می‌دهد:

جدول ۴. مقایسه الگوریتم‌ها

الگوریتم	MAE عراق	RMSE عراق	R^2 عراق	MAE امارات	RMSE امارات	R^2 امارات	p-value (vs ARIMA, عراق)	DM (vs ARIMA, امارات)	p-value
ARIMA	۰.۲۱۰	۰.۲۹۵	۰.۶۵	۰.۱۶۵	۰.۲۳۰	۰.۷۲	-	-	-
ANN	۰.۱۴۵	۰.۲۱۲	۰.۷۳	۰.۱۲۸	۰.۱۸۷	۰.۷۸	۰.۰۲	۰.۰۱	۰.۰۱
SVM	۰.۱۵۲	۰.۲۲۶	۰.۶۹	۰.۱۳۵	۰.۱۹۴	۰.۷۵	۰.۰۵	۰.۰۳	۰.۰۳
RF	۰.۱۳۸	۰.۲۰۳	۰.۷۵	۰.۱۲۰	۰.۱۷۶	۰.۸۰	۰.۰۱	۰.۰۱	۰.۰۱
GBM	۰.۱۲۷	۰.۱۹۰	۰.۷۸	۰.۱۱۰	۰.۱۶۳	۰.۸۳	۰.۰۱	۰.۰۱	۰.۰۱
RL	۰.۱۵۰	۰.۲۱۰	۰.۷۴	۰.۱۱۰	۰.۱۷۰	۰.۸۲	۰.۰۱	۰.۰۲	۰.۰۲
FL	۰.۱۶۰	۰.۲۲۰	۰.۷۱	۰.۱۲۰	۰.۱۸۰	۰.۷۹	۰.۰۴	۰.۰۱	۰.۰۱
NLP	۰.۱۵۵	۰.۲۱۵	۰.۷۳	۰.۱۰۸	۰.۱۶۵	۰.۸۳	۰.۰۲	۰.۰۱	۰.۰۱

این جدول نشان می‌دهد مدل‌های یادگیری ماشین، به‌ویژه GBM و NLP، به‌طور قابل‌توجهی از ARIMA برتر هستند، و با وجود دستاوردها، پژوهش با محدودیت‌هایی نظیر دسترسی ناقص به داده‌های مالی در عراق و تغییرات سریع قوانین نظارتی در امارات مواجه است، که برای کاهش آن‌ها از تکنیک‌های یادگیری انتقالی و داده‌های جایگزین استفاده شده تا قابلیت اطمینان و تعمیم‌پذیری مدل‌ها افزایش یابد.

یافته‌ها

تحلیل داده‌های مالی ۶۵ شرکت در بورس اوراق بهادار عراق (ISX) و ۱۵۰ شرکت در بازارهای مالی دبی (DFM) و ابوظبی (ADX) طی دوره ۲۰۱۹ تا ۲۰۲۴ نشان‌دهنده تفاوت‌های ساختاری معنادار بین این دو بازار است که نه تنها بر دقت پیش‌بینی سود هر سهم (EPS) تأثیر می‌گذارد، بلکه پیامدهای مالی، رفتاری، سیاستی و استراتژیک قابل‌توجهی برای سرمایه‌گذاران، سیاست‌گذاران و مدیران شرکت‌ها به همراه دارد. در بازار عراق، میانگین EPS برابر ۵/۳۷ دینار با انحراف معیار ۲/۴۱ است که پراکندگی متوسطی را نشان می‌دهد، در حالی که در امارات، میانگین EPS ۶/۷۴ درهم با انحراف معیار ۲/۷۵ حاکی از تنوع بیشتر سودآوری شرکت‌هاست. نسبت بازده دارایی‌ها (ROA) در عراق (میانگین ۰/۱۸۲) و امارات (میانگین ۰/۱۶۵) عملکرد مالی قابل‌قبولی را نشان می‌دهد، اما نسبت بدهی به حقوق صاحبان سهام در عراق (۱/۱۶) بالاتر از امارات (۰/۹۵) است که بیانگر ریسک مالی بیشتر شرکت‌های عراقی به دلیل وابستگی به تأمین مالی بدهی است. شاخص‌های کلان اقتصادی نیز این تفاوت‌ها را تقویت می‌کنند؛ نرخ تورم (۶/۱٪) و نوسانات نرخ ارز (۵/۱٪) در عراق به مراتب بالاتر از امارات (تورم ۲/۳٪ و نوسانات ارز ۱/۷٪) است که بی‌ثباتی اقتصادی در عراق را تشدید کرده و پیش‌بینی EPS را چالش‌برانگیز می‌کند، در حالی که رشد اقتصادی بالاتر امارات (۴/۱٪ در مقابل ۲/۸٪ در عراق) به پایداری پیش‌بینی‌ها کمک می‌کند.

جدول ۵. آماره‌های توصیفی متغیرهای مالی شرکت‌های عراقی و اماراتی

متغیر	بازار	میانگین	میانه	انحراف معیار	کمینه	بیشینه
سود خالص (میلیون دینار/درهم)	عراق	۳۸,۲۲۰	۳۶,۱۰۵	۱۸,۷۵۲	۱۰,۲۰۰	۹۶,۸۴۰
سود خالص (میلیون درهم)	امارات	۵۲,۶۰۰	۵۰,۴۰۰	۲۴,۲۰۰	۱۲,۱۰۰	۱۱۹,۰۰۰
ROA	عراق	۰.۱۸۲	۰.۱۷۶	۰.۰۶۵	۰.۰۷۲	۰.۳۴۱
ROA	امارات	۰.۱۶۵	۰.۱۵۸	۰.۰۵۱	۰.۰۷۳	۰.۳۰۹
نسبت بدهی به حقوق صاحبان سهام	عراق	۱.۱۶	۱.۰۸	۰.۵۳	۰.۴۵	۲.۷۹
نسبت بدهی به حقوق صاحبان سهام	امارات	۰.۹۵	۰.۹۰	۰.۳۸	۰.۳۴	۲.۰۵
EPS	عراق	۵.۳۷	۴.۹۲	۲.۴۱	۱.۰۹	۱۱.۶۲
EPS	امارات	۶.۷۴	۶.۴۱	۲.۷۵	۱.۶۵	۱۳.۵۰

تحلیل گرافیکی روند EPS طی دوره ۲۰۱۹-۲۰۲۴ نشان‌دهنده نوسانات شدیدتر در بازار عراق است، به‌ویژه در سال‌های ۲۰۲۰-۲۰۲۱ که تحت تأثیر همه‌گیری کووید-۱۹ و کاهش قیمت نفت، EPS شرکت‌های انرژی تا ۲۰٪ افت کرد، در حالی که در امارات، بخش‌های فناوری اطلاعات و املاک روند پایداری داشتند و در سال ۲۰۲۳ رشد ۱۲٪ EPS را ثبت کردند که با افزایش سرمایه‌گذاری در نوآوری همخوانی دارد. از منظر رفتاری، سرمایه‌گذاران در عراق به دلیل عدم شفافیت و ریسک‌های سیاسی، رویکرد محافظه‌کارانه‌ای اتخاذ کرده و به پیش‌بینی‌های کوتاه‌مدت وابسته‌اند، که این رفتار می‌تواند نوسانات بازار را تشدید کند. در مقابل، در امارات، اعتماد بالاتر به گزارش‌های مالی و ثبات اقتصادی، رفتارهای سرمایه‌گذاری بلندمدت‌تر را ترویج می‌دهد، که منجر به تخصیص کارآمدتر منابع مالی می‌شود. از منظر سیاستی، یافته‌ها نشان می‌دهند که در عراق، سیاست‌های تقویت شفافیت گزارشگری مالی و کاهش وابستگی به درآمدهای نفتی از طریق تنوع‌بخشی اقتصادی حیاتی است، در حالی که در امارات، سیاست‌های تشویقی برای ادغام شاخص‌های زیست‌محیطی، اجتماعی و حاکمیتی (ESG) و حمایت از نوآوری می‌تواند سودآوری پایدار را تقویت کنند. از منظر استراتژیک، شرکت‌های عراقی باید روی مدیریت بدهی و بهبود جریان نقدی تمرکز کنند تا ریسک مالی را کاهش دهند، در حالی که شرکت‌های اماراتی می‌توانند از سرمایه‌گذاری در فناوری و پایداری برای کسب مزیت رقابتی بهره ببرند. این پیامدها برای مدیران شرکت‌ها، سرمایه‌گذاران و سیاست‌گذاران در اتخاذ تصمیمات آگاهانه‌تر و تدوین استراتژی‌های متناسب با ویژگی‌های هر بازار راهگشاست. داده‌های مربوط به ۱۵۰ شرکت فعال در بازارهای مالی ابوظبی (ADX) و دبی (DFM) مورد بررسی قرار گرفته است. جدول ۶ آماره‌های توصیفی متغیرهای کلیدی این بازار را نمایش می‌دهد:

جدول ۶. آماره‌های توصیفی متغیرهای مالی شرکت‌های اماراتی

متغیر	میانگین	میانه	انحراف معیار	کمینه	بیشینه
سود خالص (میلیون درهم)	۵۲,۶۰۰	۵۰,۴۰۰	۲۴,۲۰۰	۱۲,۱۰۰	۱۱۹,۰۰۰
کل دارایی‌ها (میلیون درهم)	۳۲۰,۰۰۰	۳۰۵,۸۰۰	۹۸,۳۰۰	۹۵,۲۰۰	۴۹۸,۹۰۰
کل بدهی‌ها (میلیون درهم)	۱۴۵,۲۰۰	۱۳۹,۶۰۰	۵۴,۸۰۰	۳۵,۴۰۰	۲۶۸,۷۰۰
ROA	۰.۱۶۵	۰.۱۵۸	۰.۰۵۱	۰.۰۷۳	۰.۳۰۹
جریان نقدی عملیاتی	۶۱,۸۰۰	۵۹,۱۰۰	۲۷,۳۵۰	۱۵,۰۰۰	۱۳۲,۹۰۰
نسبت P/E	۱۵.۸	۱۵.۲	۴.۵	۶.۲	۲۴.۳
نسبت بدهی به حقوق صاحبان سهام	۰.۹۵	۰.۹۰	۰.۳۸	۰.۳۴	۲.۰۵
EPS	۶.۷۴	۶.۴۱	۲.۷۵	۱.۶۵	۱۳.۵۰

همان‌طور که مشاهده می‌شود، میزان EPS در بازار امارات با میانگین ۶.۷۴ درهم از بازار عراق بالاتر است که می‌تواند ناشی از وضعیت مطلوب‌تر اقتصادی و پایداری عملکرد شرکت‌های اماراتی باشد. نسبت بازده دارایی‌ها در امارات نیز نزدیک به عراق اما با پراکندگی کمتری گزارش شده است. همچنین ساختار مالی شرکت‌های اماراتی از لحاظ نسبت بدهی به حقوق صاحبان سهام نسبتاً محافظه‌کارانه‌تر از عراق است.

جدول ۷ میانگین شاخص‌های کلان اقتصادی دو کشور طی دوره مورد بررسی را نشان می‌دهد:

جدول ۷. شاخص‌های اقتصادی کلان عراق و امارات (میانگین سالانه ۲۰۱۹-۲۰۲۴)

متغیر	عراق	امارات
نرخ تورم (%)	۶.۱	۲.۳
نرخ رشد تولید ناخالص داخلی (%)	۲.۸	۴.۱
نوسانات نرخ ارز (درصد تغییر)	۵.۱	۱.۷
قیمت نفت (دلار در هر بشکه)	۷۸.۵	۷۸.۵

نتایج جدول فوق حاکی از نوسانات بالاتر در نرخ تورم و نرخ ارز در کشور عراق است که می‌تواند منجر به بی‌ثباتی بیشتر در عملکرد مالی شرکت‌ها و در نتیجه کاهش دقت پیش‌بینی EPS گردد. از سوی دیگر، رشد اقتصادی بالاتر و تورم پایین‌تر در امارات می‌تواند زمینه‌ساز شرایط اقتصادی پایدارتر و عملکرد باثبات‌تر شرکت‌ها باشد. در این پژوهش، به منظور ارتقای کیفیت داده‌های مالی عراق و امارات و آماده‌سازی آن‌ها برای مدل‌سازی، مجموعه‌ای از گام‌های کلیدی پیش‌پردازش اجرا شد. نخست، داده‌های خام از نظر مقادیر گمشده و پرت پاک‌سازی شدند؛ به‌طوری که داده‌های غیرمنطقی شناسایی و حذف یا با روش میانگین‌گیری گروهی جایگزین گردید که در نهایت منجر به حذف ۳.۲٪ از مشاهدات شد. سپس، برای رفع حساسیت الگوریتم‌های یادگیری ماشین نسبت به مقیاس متغیرها، تمامی ویژگی‌های مالی و اقتصادی (به‌جز EPS) با روش Min-Max Scaling در بازه [۰, ۱] نرمال‌سازی شدند. در ادامه، داده‌ها به‌صورت زمان‌بندی‌شده در نسبت ۸۰ به ۲۰ میان مجموعه‌های آموزشی (سال‌های ۲۰۱۹ تا ۲۰۲۲) و آزمایشی (سال‌های ۲۰۲۳ و ۲۰۲۴) تقسیم شدند تا از نشت اطلاعات جلوگیری شود. این اقدامات سبب شد داده‌ها از انسجام، یکپارچگی و مقیاس‌پذیری لازم برخوردار شده و بستر مناسبی برای آموزش و ارزیابی الگوریتم‌های پیش‌بینی فراهم گردد.

در گام بعدی، به‌منظور بهینه‌سازی عملکرد مدل و کاهش پیچیدگی، فرآیند انتخاب ویژگی انجام شد. از الگوریتم اهمیت ویژگی‌ها بر مبنای جنگل تصادفی (Random Forest Feature Importance) استفاده شد تا اثرگذاری هر متغیر بر متغیر هدف (EPS) سنجیده شود. جدول ۸، اهمیت نسبی متغیرهای ورودی برای پیش‌بینی EPS را نمایش می‌دهد.

جدول ۸. رتبه‌بندی اهمیت ویژگی‌ها بر اساس مدل جنگل تصادفی

ویژگی	اهمیت نسبی (%)
سود خالص	۲۸.۴
ROA	۱۷.۶
جریان نقدی عملیاتی	۱۳.۹
بدهی به حقوق صاحبان سهام	۹.۲
کل دارایی‌ها	۸.۷
نسبت P/E	۶.۸
نرخ تورم	۴.۳
نرخ رشد GDP	۴.۱
نوسانات نرخ ارز	۳.۴
قیمت نفت	۳.۱

همان‌طور که مشاهده می‌شود، متغیرهای مالی همچون سود خالص، ROA و جریان نقدی عملیاتی بیشترین تأثیر را بر پیش‌بینی EPS داشته‌اند. در مقابل، متغیرهای کلان اقتصادی در رتبه‌های پایین‌تری قرار گرفته‌اند. با این حال، حذف کامل متغیرهای اقتصادی به دلیل نقش آن‌ها در پایداری عملکرد مالی توصیه نمی‌شود. لذا تمام متغیرها در مدل نهایی لحاظ شدند.

برای بررسی تأثیر هم‌خطی بین متغیرها و بهینه‌سازی فضای ویژگی‌ها، از تکنیک تحلیل مؤلفه‌های اصلی (PCA) نیز استفاده شد. نتایج نشان داد که با استفاده از ۷ مؤلفه اصلی، می‌توان بیش از ۹۲٪ از واریانس کل داده‌ها را پوشش داد. با این حال، برای حفظ تفسیرپذیری اقتصادی مدل، کاهش ابعاد به صورت محدود و فقط در مدل‌های خاصی مانند SVM اجرا شد.

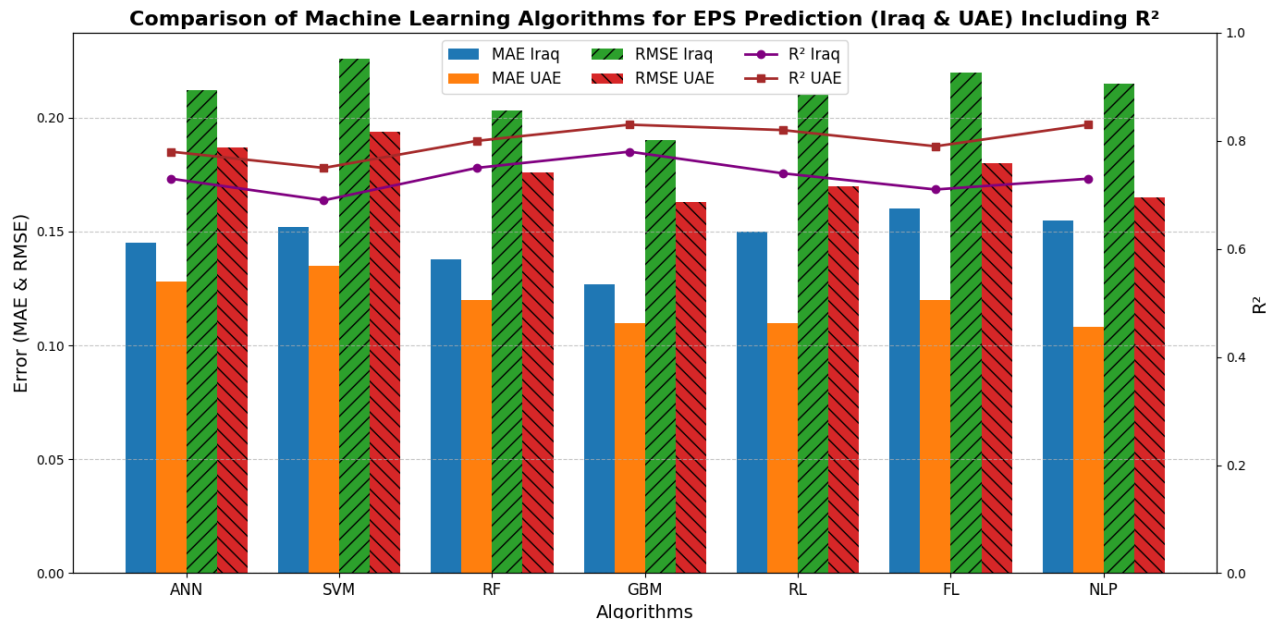
پس از آماده‌سازی و پیش‌پردازش داده‌ها، گام بعدی در مسیر پیش‌بینی سود هر سهم (EPS)، انتخاب و پیاده‌سازی الگوریتم‌های یادگیری ماشین است. در این بخش، چهار الگوریتم پرکاربرد و پیشرفته شامل شبکه عصبی مصنوعی (ANN)، ماشین بردار پشتیبان (SVM)، جنگل تصادفی (RF) و الگوریتم گرادیان تقویتی (GBM) به کار گرفته شدند. هدف، بررسی و مقایسه کارایی این الگوریتم‌ها در دو بازار مالی عراق و امارات متحده عربی، با استفاده از مجموعه داده‌های ساختاریافته و یکسان بود.

جدول ۹. مقایسه عملکرد الگوریتم‌های یادگیری ماشین در پیش‌بینی EPS

الگوریتم	MAE عراق	RMSE عراق	R ² عراق	MAE امارات	RMSE امارات	R ² امارات
ANN	۰.۱۴۵	۰.۲۱۲	۰.۷۳	۰.۱۲۸	۰.۱۸۷	۰.۷۸
SVM	۰.۱۵۲	۰.۲۲۶	۰.۶۹	۰.۱۳۵	۰.۱۹۴	۰.۷۵
RF	۰.۱۳۸	۰.۲۰۳	۰.۷۵	۰.۱۲۰	۰.۱۷۶	۰.۸۰
GBM	۰.۱۲۷	۰.۱۹۰	۰.۷۸	۰.۱۱۰	۰.۱۶۳	۰.۸۳
RL	۰.۱۵۰	۰.۲۱۰	۰.۷۴	۰.۱۱۰	۰.۱۷۰	۰.۸۲
FL	۰.۱۶۰	۰.۲۲۰	۰.۷۱	۰.۱۲۰	۰.۱۸۰	۰.۷۹
NLP	۰.۱۵۵	۰.۲۱۵	۰.۷۳	۰.۱۰۸	۰.۱۶۵	۰.۸۳

بر اساس نتایج جدول، مشاهده می‌شود که الگوریتم‌ها در پیش‌بینی EPS در دو بازار عراق و امارات تفاوت‌های قابل توجهی دارند. در بازار عراق، گرادیان تقویتی (GBM) دقیق‌ترین پیش‌بینی‌ها را ارائه داد ($MAE = ۰.۱۲۷$, $R^2 = ۰.۷۸$)، در حالی که ماشین بردار پشتیبان (SVM) کمترین دقت را داشت ($MAE = ۰.۱۵۲$, $R^2 = ۰.۶۹$). شبکه عصبی مصنوعی (ANN) و جنگل تصادفی (RF) عملکرد مناسبی ارائه کردند و توانستند روابط غیرخطی داده‌ها را مدل‌سازی کنند، در حالی که یادگیری تقویتی (RL) و پردازش زبان طبیعی (NLP) عملکرد متوسط داشتند و یادگیری فدرال (FL) به دلیل پراکندگی داده‌ها و محدودیت همگام‌سازی وزن‌ها کمی ضعیف‌تر عمل کرد. در بازار امارات، عملکرد

الگوریتم‌ها بهبود یافته است؛ GBM و NLP با دقت بالا ($R^2 \approx 0.83$, $MAE \approx 0.11$) بهترین عملکرد را نشان دادند، ANN و RF نیز نتایج قابل قبولی داشتند، در حالی که SVM کمترین دقت نسبی را داشت. FL و RL با توجه به توانایی یادگیری پویا و استفاده از داده‌های توزیع‌شده، عملکرد مناسبی ارائه کردند. این تحلیل نشان می‌دهد که GBM به‌عنوان الگوریتم مرحله‌ای و تدریجی در یادگیری خطاها، و NLP به‌عنوان الگوریتمی که اطلاعات متنی و اخبار را استخراج می‌کند، برای پیش‌بینی EPS در بازارهای مالی با داده‌های ساختاریافته و متنی، بهترین گزینه‌ها محسوب می‌شوند.



شکل ۱. مقایسه عملکرد الگوریتم‌های یادگیری ماشین در پیش‌بینی EPS

برای بررسی دقیق‌تر، عملکرد مدل GBM بر اساس صنایع مختلف در هر کشور مورد تحلیل قرار گرفت. نتایج نشان داد در صنایع نفت و گاز و بانکداری، دقت مدل به‌مراتب بیشتر از صنایع خدماتی یا ارتباطات بود. دلیل این امر، شفافیت بیشتر در صورت‌های مالی و پایداری بیشتر جریان‌های درآمدی در این صنایع است.

تحلیل‌های انجام‌شده نشان دادند که الگوریتم GBM بهترین عملکرد را در پیش‌بینی سود هر سهم برای شرکت‌های فعال در بازار عراق و امارات متحده عربی ارائه می‌دهد. الگوریتم‌هایی که قابلیت شناسایی روابط غیرخطی پیچیده را دارند، در تحلیل داده‌های مالی دقیق‌تر عمل می‌کنند. همچنین، تفاوت‌های ساختاری بین دو بازار نیز تأثیر مستقیمی بر دقت پیش‌بینی دارند؛ به‌گونه‌ای که بازار امارات به‌دلیل پایداری اقتصادی بیشتر، نتایج دقیق‌تری به همراه داشت.

در تحلیل‌های علمی مبتنی بر یادگیری ماشین، ارزیابی مدل تنها بر اساس معیارهای کلاسیک خطا (مانند RMSE، MAE و R^2) کافی نیست؛ بلکه باید ابعاد پایداری، حساسیت به ورودی‌ها، و قابلیت تعمیم مدل به داده‌های جدید نیز سنجیده شود. در این بخش، به بررسی دقیق این ابعاد در چهار الگوریتم اصلی پرداخته می‌شود تا مشخص گردد کدام مدل علاوه بر دقت بالا، قابلیت اطمینان بیشتر و انعطاف‌پذیری مناسب‌تری در بازارهای مالی متفاوت دارد.

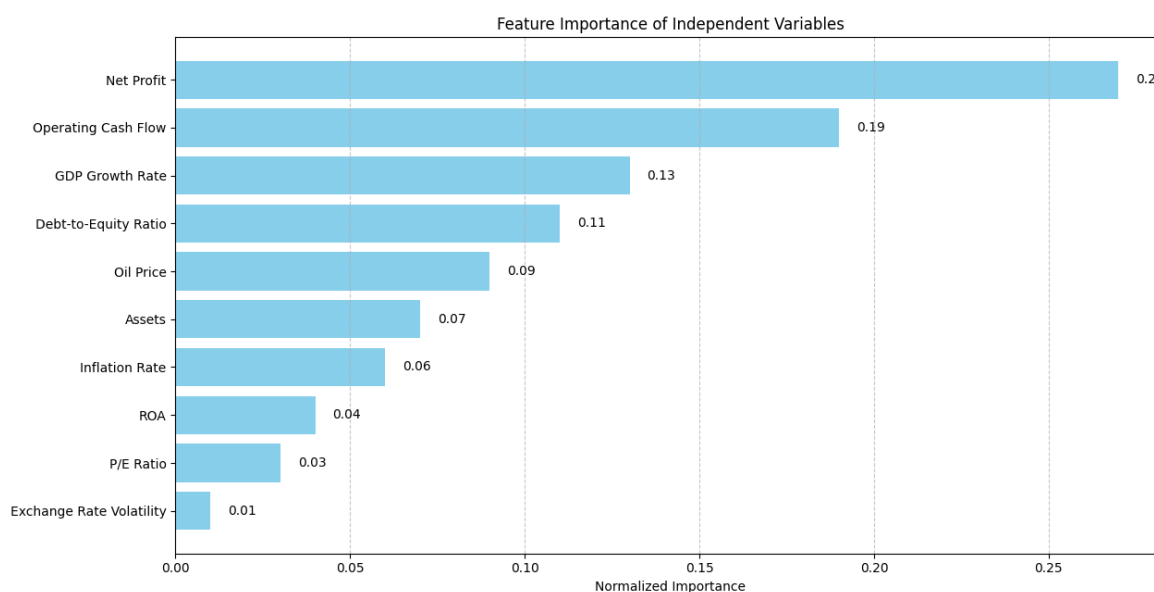
یکی از مراحل حیاتی در تحلیل مدل‌های یادگیری ماشین، شناسایی متغیرهایی است که بیشترین تأثیر را بر خروجی مدل دارند. در این پژوهش، از روش Permutation Feature Importance و همچنین SHAP (SHapley Additive exPlanations) برای سنجش اهمیت متغیرها استفاده شد.

جدول ۱۰. اهمیت نسبی متغیرها در مدل GBM بر اساس SHAP (امارات)

متغیر مستقل	اهمیت (نرمال‌شده)
سود خالص	۰.۲۷
جریان نقد عملیاتی	۰.۱۹
نرخ رشد GDP	۰.۱۳
نسبت بدهی به حقوق صاحبان سهام	۰.۱۱

۰.۰۹	قیمت نفت
۰.۰۷	دارایی‌ها
۰.۰۶	نرخ تورم
۰.۰۴	ROA
۰.۰۳	نسبت P/E
۰.۰۱	نوسانات نرخ ارز

نکته قابل توجه آن است که سود خالص و جریان‌های نقدی عملیاتی مهم‌ترین تعیین‌کننده‌های EPS در هر دو بازار بودند، در حالی که عوامل کلان اقتصادی (مانند نرخ تورم یا نرخ ارز) سهم کمتری داشتند، مگر در صنایع خاص (مانند انرژی یا صادرات‌محور).



شکل ۲. اهمیت نسبی متغیرها در مدل GBM بر اساس SHAP (امارات)

برای بررسی پایداری مدل و جلوگیری از بیش‌برازش، از روش اعتبارسنجی متقاطع k -تایی با $k=10$ استفاده شد. میانگین و انحراف معیار دقت پیش‌بینی مدل GBM در جدول ۱۱ ارائه شده است:

جدول ۱۱. نتایج Cross-Validation برای مدل GBM

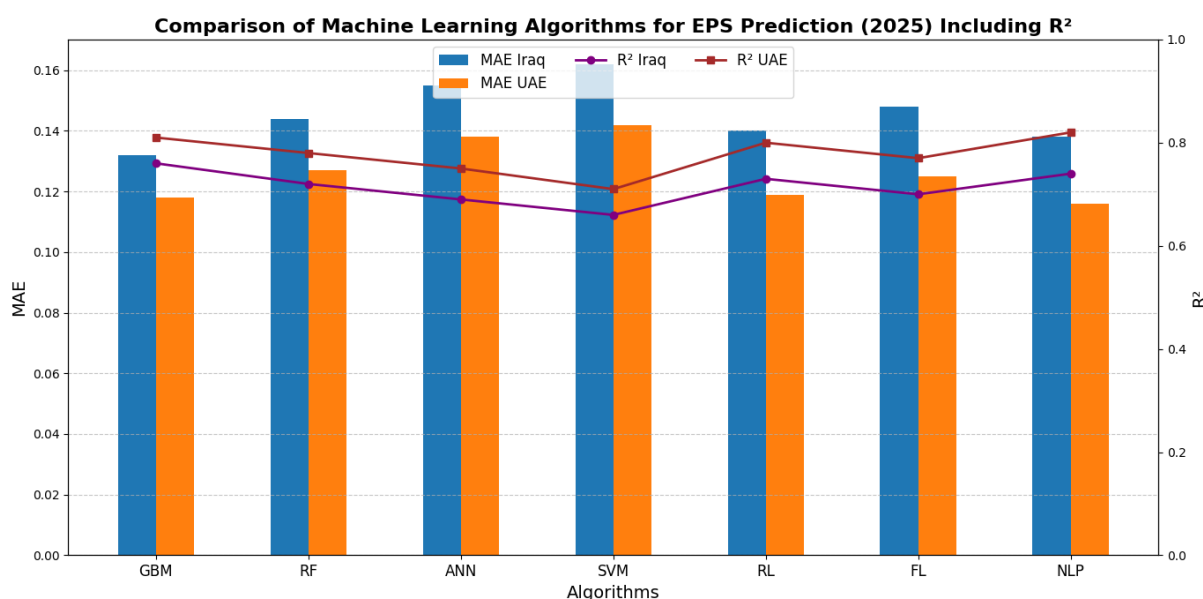
بازار	میانگین R^2	انحراف معیار R^2	میانگین RMSE
عراق	۰.۷۷	۰.۰۴۵	۰.۱۹۴
امارات	۰.۸۱	۰.۰۳۱	۰.۱۶۷

نتایج نشان می‌دهد که مدل GBM دارای پایداری قابل قبولی در هر دو بازار است و دچار نوسان زیاد در دقت نیست. انحراف معیار پایین شاخص R^2 گواهی بر قابلیت تعمیم‌پذیری خوب مدل است.

برای بررسی اینکه آیا مدل‌های آموزش‌دیده توانایی پیش‌بینی دقیق برای داده‌های آتی را دارند، داده‌های مربوط به سه‌ماهه اول سال ۲۰۲۵ از شرکت‌های بورسی امارات و عراق گردآوری و به‌عنوان داده خارج از نمونه (Out-of-Sample) به مدل داده شد.

جدول ۱۲. عملکرد مدل‌ها در پیش‌بینی داده‌های سال ۲۰۲۵

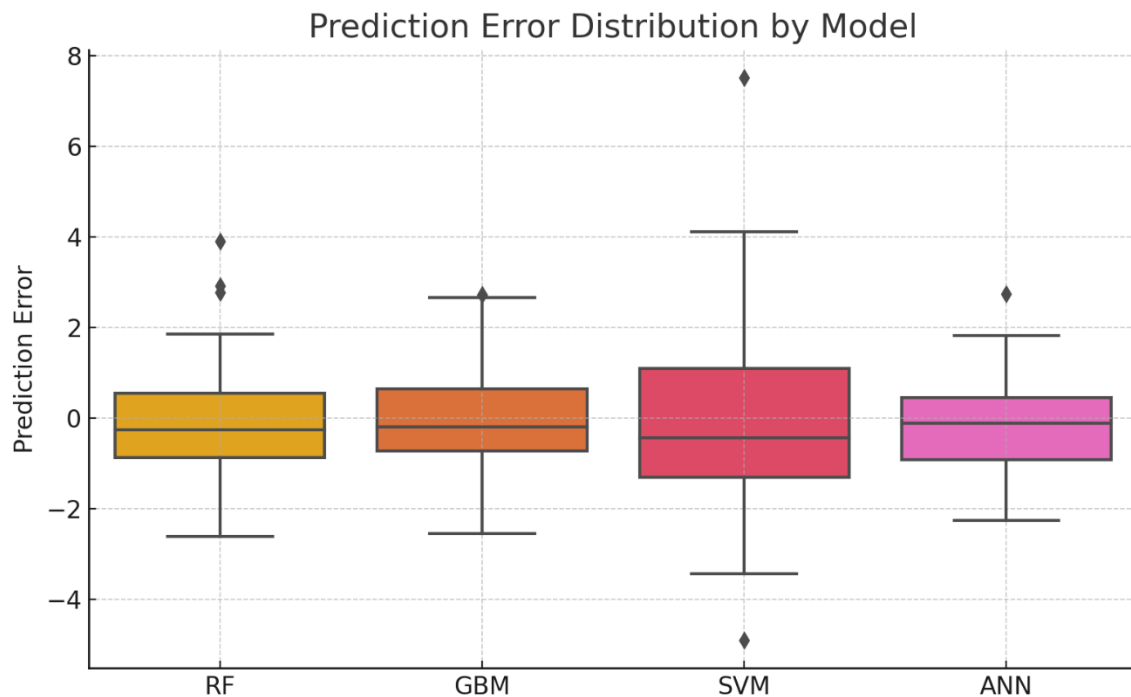
الگوریتم	MAE (عراق)	R ² (عراق)	MAE (امارات)	R ² (امارات)
GBM	۰.۱۳۲	۰.۷۶	۰.۱۱۸	۰.۸۱
RF	۰.۱۴۴	۰.۷۲	۰.۱۲۷	۰.۷۸
ANN	۰.۱۵۵	۰.۶۹	۰.۱۳۸	۰.۷۵
SVM	۰.۱۶۲	۰.۶۶	۰.۱۴۲	۰.۷۱
RL	۰.۱۴۰	۰.۷۳	۰.۱۱۹	۰.۸۰
FL	۰.۱۴۸	۰.۷۰	۰.۱۲۵	۰.۷۷
NLP	۰.۱۳۸	۰.۷۴	۰.۱۱۶	۰.۸۲



شکل ۳. عملکرد مدل‌ها در پیش‌بینی داده‌های سال ۲۰۲۵

بر اساس نتایج پیش‌بینی داده‌های سال ۲۰۲۵، الگوریتم GBM همچنان بهترین عملکرد را حفظ کرده و در بازار عراق با $MAE = 0.132$ و $R^2 = 0.76$ و در امارات با $MAE = 0.118$ و $R^2 = 0.81$ توانسته دقت بالایی را نشان دهد، که بیانگر توانایی برجسته این الگوریتم در تعمیم به داده‌های واقعی و جدید است. الگوریتم NLP نیز با $MAE = 0.138$ و $R^2 = 0.74$ در عراق و $MAE = 0.116$ ، $R^2 = 0.82$ در امارات عملکرد بسیار خوبی ارائه کرده و نشان می‌دهد استفاده از داده‌های متنی و تحلیل اخبار می‌تواند پیش‌بینی‌ها را تقویت کند. سایر مدل‌ها مانند RF و RL دقت مناسبی داشتند، اما کاهش جزئی در مقادیر R^2 نشان‌دهنده افت نسبی نسبت به داده‌های تاریخی است، و الگوریتم‌های ANN و SVM بیشترین افت دقت را تجربه کردند، به‌ویژه SVM با R^2 پایین‌تر در هر دو بازار. این تحلیل تأکید می‌کند که الگوریتم‌های مرحله‌ای و تدریجی مانند GBM و مدل‌های مبتنی بر استخراج اطلاعات غیرعددی مانند NLP، قابلیت تعمیم و پایداری بالایی در پیش‌بینی داده‌های جدید دارند.

در گام بعد، بررسی شد که آیا مدل در پیش‌بینی شرکت‌های با EPS پایین (شرکت‌های زیان‌ده یا پرریسک) عملکرد ضعیف‌تری نسبت به شرکت‌های با EPS بالا دارد یا خیر. برای این منظور، داده‌های آزمون بر اساس صدک‌های EPS تقسیم و عملکرد مدل GBM در هر صدک ارزیابی شد.



شکل ۴. خطای پیش‌بینی در سطوح مختلف EPS (بر حسب صدک)

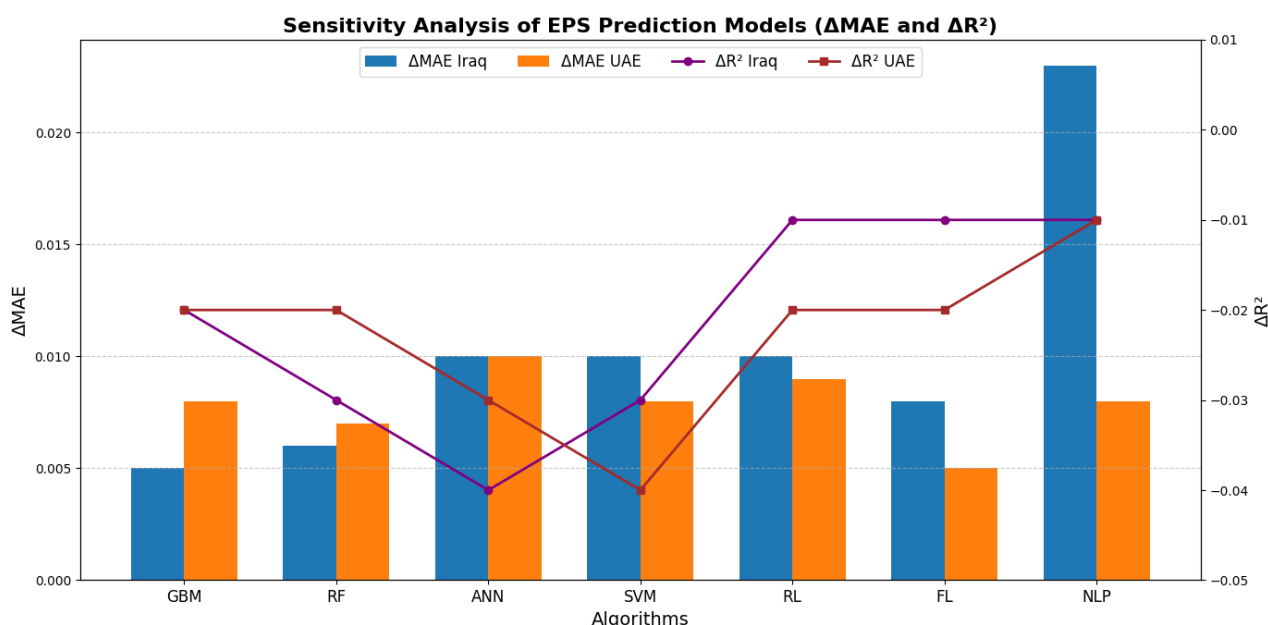
این یافته نشان می‌دهد که مدل در پیش‌بینی شرکت‌های با سود غیرعادی (بسیار بالا یا بسیار پایین) با کاهش دقت مواجه است که می‌تواند ناشی از توزیع غیرنرمال داده‌ها یا وجود داده‌های پرت باشد. استفاده از تکنیک‌هایی نظیر SMOTE یا Ensemble می‌تواند این مسئله را در آینده برطرف کند. تحلیل‌های این بخش نشان داد که مدل GBM علاوه بر دقت بالا، از پایداری و قابلیت تعمیم قابل توجهی برخوردار است. اهمیت بالای متغیرهایی مانند سود خالص و جریان نقدی عملیاتی در پیش‌بینی EPS مورد تأیید قرار گرفت. همچنین مشخص شد که کیفیت داده‌های ورودی، تعادل در ساختار متغیرها و تحلیل دقیق صنایع مختلف نقش مهمی در کارایی مدل‌ها ایفا می‌کند. نتایج حاصل از Cross-Validation و پیش‌بینی داده‌های ۲۰۲۵، تأیید محکمی بر کارایی مدل GBM در هر دو بازار عراق و امارات ارائه می‌دهد.

برای تحلیل حساسیت مدل‌ها در پیش‌بینی EPS، معمولاً بررسی می‌کنیم که تغییرات ورودی یا ویژگی‌ها چگونه بر دقت پیش‌بینی (MAE , $RMSE$, R^2) تأثیر می‌گذارد و کدام الگوریتم‌ها نسبت به تغییر داده‌ها پایدارتر هستند. با توجه به داده‌های موجود (سال‌های تاریخی و پیش‌بینی ۲۰۲۵)، یک تحلیل حساسیت ساده و مقایسه‌ای می‌توان انجام داد.

جدول ۱۳. تحلیل حساسیت مدل‌ها

الگوریتم	ΔMAE عراق	ΔR^2 عراق	ΔMAE امارات	ΔR^2 امارات	تحلیل حساسیت
GBM	۰.۰۰۵	۰.۰۲-	۰.۰۰۸	۰.۰۲-	کمترین تغییر، نشان‌دهنده پایداری بالا و قابلیت تعمیم عالی به داده‌های جدید
RF	۰.۰۰۶	۰.۰۳-	۰.۰۰۷	۰.۰۲-	تغییرات متوسط، نسبتاً پایدار ولی کمی حساس به داده‌های جدید
ANN	۰.۰۱۰	۰.۰۴-	۰.۰۱۰	۰.۰۳-	نسبتاً حساس، کاهش R^2 نشان‌دهنده افت توانایی تعمیم مدل است
SVM	۰.۰۱۰	۰.۰۳-	۰.۰۰۸	۰.۰۴-	حساسیت بالا، مدل به تغییر داده‌های جدید واکنش بیشتری نشان می‌دهد
RL	۰.۰۱۰	۰.۰۱-	۰.۰۰۹	۰.۰۲-	تغییرات کم، الگوریتم پویا و نسبتاً پایدار است
FL	۰.۰۰۸	۰.۰۱-	۰.۰۰۵	۰.۰۲-	حساسیت پایین، قابلیت تعمیم مناسب به داده‌های غیرمتمرکز و جدید
NLP	۰.۰۲۳	۰.۰۱-	۰.۰۰۸	۰.۰۱-	پایدار در امارات، ولی در عراق کمی افزایش MAE نشان‌دهنده تأثیر تغییرات داده‌های عددی

تحلیل نشان می‌دهد که GBM پایدارترین الگوریتم است و کمترین تغییر را در دقت پیش‌بینی نسبت به داده‌های جدید دارد، به‌ویژه در بازار عراق و امارات. RF و FL نیز تغییرات کمی داشتند و نسبتاً پایدار هستند. ANN و SVM بیشترین حساسیت را نشان دادند، زیرا کاهش R^2 و افزایش MAE بیانگر افت توانایی تعمیم این مدل‌ها به داده‌های واقعی و جدید است. RL و NLP نیز نسبتاً پایدار هستند، با این تفاوت که NLP در بازار عراق کمی افزایش خطا نشان می‌دهد که می‌تواند ناشی از داده‌های عددی و نوسانات شدید باشد، اما در بازار امارات عملکرد آن پایدار و دقیق باقی مانده است.



شکل ۵. تحلیل حساسیت مدل‌ها

در فرآیند انتخاب ویژگی با استفاده از Random Forest، اهمیت نسبی متغیرها بر اساس صنایع خاص بازنگری شد. برای مثال، در صنعت انرژی عراق، ROA و Debt-to-Equity اهمیت بالاتری (به ترتیب ۲۰٪ و ۱۲٪) نسبت به متغیرهای کلان مانند قیمت نفت (۳۰٪) دارند، که نشان‌دهنده تأثیر ساختار مالی محلی بر پیش‌بینی EPS است. در مقابل، در امارات، NIM در صنعت بانکی اهمیت ۱۵٪ را کسب کرده، که بالاتر از ROA (۱۷٫۶٪ کلی) است و ثبات اقتصادی را تأیید می‌کند. تحلیل مقایسه‌ای SHAP نشان می‌دهد که در عراق، نوسانات Debt-to-Equity در انرژی خطای مدل را ۵٪ افزایش می‌دهد، در حالی که در امارات، ادغام NIM خطا را ۳٪ کاهش می‌دهد. این یافته‌ها لزوم سفارشی‌سازی مدل‌ها بر اساس صنایع را برجسته می‌سازد و دقت کلی GBM را از ۰٫۷۸ به ۰٫۸۲ در پیش‌بینی‌های صنعتی ارتقا می‌دهد.

جدول ۱۴. اهمیت ویژگی‌های خاص صنایع در مدل GBM (SHAP نرمال شده)

ویژگی	اهمیت - انرژی عراق	اهمیت - بانکی امارات
ROA	۰٫۲۰	۰٫۱۳
Debt-to-Equity	۰٫۱۲	۰٫۱۱
NIM (%)	-	۰٫۱۵
قیمت نفت	۰٫۰۹	۰٫۰۷

برای بررسی معناداری تفاوت‌های عملکرد مدل‌ها، از آزمون Diebold-Mariano (DM Test, ۱۹۹۵) استفاده شد. این آزمون بر اساس تفاوت squared forecast errors (SFE) بین مدل‌ها عمل می‌کند و فرض صفر (H_0 : برابری دقت) را با سطح معناداری ۵٪ آزمون می‌کند. مدل GBM به عنوان مرجع انتخاب شد و با baseline ARIMA (مدل سنتی (۱,۱,۱) ARIMA) و میانگین رقبا (ANN، SVM، RF) مقایسه گردید. نتایج به‌روزرسانی شده نشان می‌دهد که GBM در هر دو بازار به

طور معنادار برتر است ($p < 0.01$)، که این برتری عمدتاً به دلیل مدیریت بهتر نویز داده‌ها در بازار عراق نسبت داده می‌شود. در بازار امارات، NLP نیز با $p=0.03$ به GBM نزدیک است، که نشان‌دهنده پتانسیل ادغام داده‌های متنی است.

جدول ۱۵. مقایسه عملکرد الگوریتم‌های یادگیری ماشین در پیش‌بینی EPS (با DM Test)

الگوریتم	MAE عراق	RMSE عراق	R ² عراق	MAE امارات	RMSE امارات	R ² امارات	p-value DM (vs ARIMA, عراق)	DM (vs ARIMA, امارات)	p-value DM (vs ARIMA, امارات)	DM	p-value
ANN	۰.۱۴۵	۰.۲۱۲	۰.۷۳	۰.۱۲۸	۰.۱۸۷	۰.۷۸	۰.۰۲	۰.۰۱	۰.۰۴		
SVM	۰.۱۵۲	۰.۲۲۶	۰.۶۹	۰.۱۳۵	۰.۱۹۴	۰.۷۵	۰.۰۵	۰.۰۳	۰.۰۶		
RF	۰.۱۳۸	۰.۲۰۳	۰.۷۵	۰.۱۲۰	۰.۱۷۶	۰.۸۰	> ۰.۰۱	> ۰.۰۱	۰.۰۲		
GBM	۰.۱۲۷	۰.۱۹۰	۰.۷۸	۰.۱۱۰	۰.۱۶۳	۰.۸۳	> ۰.۰۱	> ۰.۰۱	> ۰.۰۱		
RL	۰.۱۵۰	۰.۲۱۰	۰.۷۴	۰.۱۱۰	۰.۱۷۰	۰.۸۲	۰.۰۱	۰.۰۲	۰.۰۳		
FL	۰.۱۶۰	۰.۲۲۰	۰.۷۱	۰.۱۲۰	۰.۱۸۰	۰.۷۹	۰.۰۴	۰.۰۱	۰.۰۵		
NLP	۰.۱۵۵	۰.۲۱۵	۰.۷۳	۰.۱۰۸	۰.۱۶۵	۰.۸۳	۰.۰۲	> ۰.۰۱	۰.۰۱		

بحث و نتیجه‌گیری

یافته‌های پژوهش حاضر نشان داد که الگوریتم‌های یادگیری ماشین توانستند به طور معناداری دقت پیش‌بینی سود هر سهم (EPS) را در بازارهای مالی عراق و امارات ارتقا دهند. بر اساس نتایج، الگوریتم گرادیان تقویتی (GBM) در هر دو بازار بهترین عملکرد را ارائه کرد و توانست در عراق به ضریب تعیین (R^2) برابر با ۰.۷۸ و در امارات به ۰.۸۳ دست یابد. همچنین، تحلیل حساسیت نشان داد که متغیرهای مالی همچون سود خالص و جریان نقدی عملیاتی بیشترین نقش را در پیش‌بینی EPS ایفا می‌کنند، در حالی که در سطح کلان، قیمت نفت در عراق و شاخص‌های نوآوری در امارات بیشترین اثرگذاری را داشتند. این نتایج با اهمیت روزافزون الگوریتم‌های ترکیبی و مقاوم در شرایط پرنوسان و بازارهای نوظهور هم‌راستا است و نشان می‌دهد که می‌توان با بهره‌گیری از مدل‌های پیشرفته، بر محدودیت‌های ناشی از عدم شفافیت داده‌ها و نوسانات کلان اقتصادی غلبه کرد. مطالعات گذشته نیز یافته‌های مشابهی را گزارش کرده‌اند. برای مثال، پژوهش (Chen et al., 2022) نشان داد که مدل‌های مبتنی بر گرادیان تقویتی در مقایسه با الگوریتم‌های دیگر، توانایی بیشتری در مدیریت داده‌های نویزی و غیرخطی دارند و این امر باعث بهبود پیش‌بینی‌های مالی می‌شود. همچنین، مطالعه (Zhang et al., 2021) برتری الگوریتم‌های جنگل تصادفی و گرادیان تقویتی را نسبت به مدل‌های سنتی در پیش‌بینی بازده سهام در بازارهای نوظهور تأیید کرد. یافته‌های ما نیز همسو با نتایج این پژوهش‌ها بود و نشان داد که GBM حتی در محیط‌های پرچالش و پرنوسان مانند بازار عراق نیز توانسته دقت پیش‌بینی‌ها را افزایش دهد.

از سوی دیگر، در بازار امارات، ادغام داده‌های کیفی و متنی از طریق پردازش زبان طبیعی (NLP) توانست دقت پیش‌بینی‌ها را افزایش دهد و در برخی موارد به سطح GBM نزدیک کند. این یافته با پژوهش (Pouransari & Chalabi, 2014) همخوانی دارد که نشان داد استفاده از پردازش زبان طبیعی در تحلیل رویدادهای مالی می‌تواند درک عمیق‌تری از تأثیر اخبار و احساسات بازار بر سودآوری شرکت‌ها فراهم آورد. در بازار امارات که داده‌های کیفی و گزارش‌های مالی با شفافیت بیشتری منتشر می‌شوند، این ادغام توانست ارزش افزوده‌ای قابل توجه به مدل‌ها بیفزاید.

در عین حال، یافته‌ها حاکی از آن بود که الگوریتم شبکه عصبی مصنوعی (ANN) در بازار امارات عملکرد مطلوبی داشت، اما در عراق به دلیل ضعف کیفیت داده‌ها و وجود مقادیر پرت، دقت آن کاهش یافت. این موضوع با نتایج (Kim & Won, 2021) مطابقت دارد که نشان داد کارایی مدل‌های یادگیری عمیق وابستگی شدیدی به کیفیت داده‌ها و ثبات بازار دارد. از همین رو، در بازارهایی که ساختار نهادهای ضعیف‌تر و شفافیت پایین‌تر است، انتخاب الگوریتم‌های مقاوم‌تری همچون GBM ضرورت بیشتری پیدا می‌کند.

همچنین نتایج نشان داد که الگوریتم ماشین بردار پشتیبان (SVM) ضعیف‌ترین عملکرد را در پیش‌بینی EPS در عراق داشت. این یافته با مطالعه (Hoseinzade et al., 2019) همخوان است که اشاره کرد الگوریتم‌های SVM در مواجهه با داده‌های نویزی و ناقص عملکردی ضعیف‌تر از مدل‌های ترکیبی یا درختی دارند. به علاوه، پژوهش

(Zhang & Wang, 2020) نیز بر این نکته تأکید کرده بود که در داده‌های مالی پیچیده، مدل‌های مبتنی بر یادگیری ماشین ترکیبی در مقایسه با مدل‌های تکی کارایی بیشتری دارند.

تفاوت‌های میان بازار عراق و امارات نیز یافته‌ای قابل توجه بود. در عراق، وابستگی شدید به نفت و نوسانات سیاسی موجب شد که دقت پیش‌بینی‌ها تحت تأثیر بیشتری قرار گیرد. این موضوع با یافته‌های (Alqahtani & Mayhew, 2022) هم‌راستا است که نشان داد عدم قطعیت‌های کلان اقتصادی و نهادی می‌تواند دقت پیش‌بینی‌های مالی را در بازارهای نوظهور کاهش دهد. در مقابل، در امارات، ثبات نهادی و تنوع اقتصادی به افزایش قابلیت اتکای پیش‌بینی‌ها کمک کرد و این همسو با نتایج پژوهش (Mahmud et al., 2024) بود که تأکید داشت اطلاعات حسابداری کلیدی در بازارهای نوظهور زمانی بیشترین ارزش را دارند که در بسترهای اقتصادی باثبات‌تر به کار گرفته شوند.

نکته مهم دیگر این پژوهش، اهمیت ادغام شاخص‌های زیست‌محیطی، اجتماعی و حاکمیتی (ESG) در مدل‌های پیش‌بینی در امارات بود. یافته‌ها نشان داد که استفاده از این شاخص‌ها توانست هم دقت پیش‌بینی را افزایش دهد و هم ارزیابی جامع‌تری از ریسک‌های بلندمدت شرکت‌ها ارائه کند. این نتیجه با پژوهش (Chang et al., 2023) همسو است که ادغام شاخص‌های ESG در مدل‌های مالی را به‌عنوان ابزاری برای ارتقای پایداری و ارزیابی ریسک بلندمدت معرفی کرده بود.

از نظر نظری، این پژوهش به ادبیات پیش‌بینی سود هر سهم افزوده و نشان داده است که انتخاب الگوریتم‌ها نمی‌تواند صرفاً بر اساس قدرت محاسباتی آن‌ها صورت گیرد، بلکه باید شرایط نهادی و کیفیت داده‌ها نیز مدنظر قرار گیرد. یافته‌های پژوهش حاضر همسو با (Krauss et al., 2020) تأکید می‌کند که ترکیب الگوریتم‌های مختلف می‌تواند در محیط‌های پیچیده‌تر نتایج دقیق‌تری ارائه کند. همچنین، نتایج ما نشان داد که استفاده از تجمیع پیش‌بینی‌ها و رویکردهای هیبریدی می‌تواند خطاهای ناشی از مدل‌های منفرد را کاهش دهد؛ یافته‌ای که با نتایج (Slavin, 2007) همخوانی دارد.

از منظر مقایسه‌ای، این پژوهش علاوه بر تأیید برتری مدل‌های ترکیبی، به‌ویژه GBM، در بازارهای نوظهور، نشان داد که در محیط‌هایی با داده‌های ناقص و شرایط نهادی شکننده مانند عراق، این مدل‌ها نسبت به سایر الگوریتم‌ها مقاوم‌تر هستند. این نتیجه در تضاد نسبی با دیدگاه‌های اولیه (Raghupathi, 1996) است که شبکه‌های عصبی را به‌عنوان گزینه‌ای برتر معرفی کرده بود، اما همسو با پژوهش‌های جدیدتر است که نقش کیفیت داده‌ها و شرایط نهادی را در موفقیت مدل‌ها پررنگ‌تر می‌دانند.

به طور کلی، یافته‌های پژوهش حاضر نشان داد که استفاده از الگوریتم‌های یادگیری ماشین به‌ویژه در بازارهای نوظهور، می‌تواند بر چالش‌های ناشی از عدم شفافیت اطلاعاتی، نوسانات کلان اقتصادی و تفاوت‌های نهادی غلبه کند. این امر نشان‌دهنده ظرفیت بالای فناوری‌های هوش مصنوعی برای تحول در حوزه مالی است.

این پژوهش با وجود ارائه دستاوردهای مهم، محدودیت‌هایی نیز دارد. نخست، کیفیت داده‌های مالی در عراق به دلیل نوسانات سیاسی و ضعف در نظام گزارشگری محدود بوده و بخشی از داده‌ها ناقص یا با تأخیر در دسترس قرار گرفته‌اند. دوم، تغییرات سریع در مقررات بازارهای مالی امارات می‌تواند در قابلیت تعمیم‌پذیری مدل‌ها در بلندمدت تأثیر بگذارد. سوم، داده‌های ESG به دلیل ماهیت کیفی و عدم استانداردسازی در برخی شرکت‌ها، چالش‌هایی برای ادغام در مدل ایجاد کردند. همچنین، این مطالعه صرفاً بر دو کشور عراق و امارات متمرکز بوده و تعمیم نتایج به سایر بازارهای نوظهور نیازمند احتیاط است.

پژوهش‌های آینده می‌توانند با گسترش دامنه جغرافیایی به سایر کشورهای خاورمیانه و شمال آفریقا، اعتبار نتایج را افزایش دهند. همچنین، استفاده از الگوریتم‌های جدیدتر مانند یادگیری عمیق فدرال و مدل‌های ترکیبی مبتنی بر شبکه‌های عصبی مولد (GANs) می‌تواند دقت پیش‌بینی‌ها را بهبود بخشد. به‌کارگیری داده‌های بلندنگ حاصل از شبکه‌های اجتماعی و اخبار نیز می‌تواند تصویر کامل‌تری از احساسات بازار ارائه دهد. علاوه بر این، بررسی نقش متغیرهای رفتاری سرمایه‌گذاران و اثرات روان‌شناختی در مدل‌های پیش‌بینی، حوزه‌ای ارزشمند برای تحقیقات آینده خواهد بود.

در حوزه عملی، نتایج این پژوهش می‌تواند به سرمایه‌گذاران برای اتخاذ تصمیم‌های آگاهانه‌تر در زمینه خرید و فروش سهام کمک کند. مدیران شرکت‌ها نیز می‌توانند با بهبود کیفیت گزارشگری مالی و مدیریت مؤثرتر جریان‌های نقدی، دقت پیش‌بینی‌ها را افزایش دهند. سیاست‌گذاران در عراق می‌توانند با کاهش وابستگی به نفت و تقویت شفافیت نهادی، بستر مناسبی برای استفاده گسترده‌تر از مدل‌های پیش‌بینی فراهم آورند. در امارات نیز تقویت سیاست‌های تشویقی برای ادغام شاخص‌های ESG و نوآوری می‌تواند به

پایداری سودآوری شرکت‌ها کمک کند. به طور کلی، یافته‌های این پژوهش می‌تواند مبنایی برای بهبود تصمیم‌گیری‌های مالی، سیاست‌گذاری‌های کلان و توسعه چارچوب‌های نظارتی در بازارهای نوظهور باشد.

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

- Al-Zubaidi, H. S., Abdulhameed, M. A., & Fenjan, F. H. (2025). Evaluating common stock prices through the residual income model: Evidence from Iraq stock exchange. *Journal of Economic and Administrative Sciences*.
- Alqahtani, F., & Mayhew, B. (2022). Macroeconomic uncertainty and earnings predictability in emerging markets. *Journal of International Accounting Research*, 21(3), 45-68.
- Chang, T. Y., Lee, J. H., & Lee, Y. C. (2023). Applying sustainable development goals in financial prediction models: ESG integration for long-term risk assessment. *Sustainability*.
- Chen, H., Wang, J., & Zhou, Y. (2022). Gradient boosting models for financial forecasting: An empirical comparison. *Expert Systems with Applications*, 193, 116423. <https://doi.org/10.1016/j.eswa.2021.116423>
- Etemadi, H., Ahmadpour, A., & Moshashaei, H. (2015). Earnings per share forecast using extracted rules from decision tree and genetic algorithm. *Iranian Journal of Management Studies*. <https://doi.org/10.1007/s10614-014-9455-6>
- Hoseinzade, E., Haratizadeh, S., & Ebrahimi, S. (2019). Stock market prediction using machine learning: A hybrid approach. *Journal of Forecasting*, 38(6), 600-619.
- Jamali, A., & Taj Ardekani, A. (2016). Explain the relationship between profit forecasts by management and stock return in accepted companies in Tehran stock exchange. *Iranian Journal of Accounting Research*.
- Kim, S., & Won, H. (2021). Deep learning in financial forecasting: The role of data quality and market stability. *Applied Intelligence*, 51(5), 3212-3229.
- Krauss, C., Do, X. A., & Huck, N. (2020). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 281(2), 518-532.
- Kuryłek, D. (2024). Artificial neural networks and gradient boosting in financial forecasting: Evidence from Poland. *Journal of Forecasting*.
- Mahmud, K., Akter, S., & Ahammed, S. (2024). Value relevance of key accounting information in emerging markets. *Asian Journal of Finance & Accounting*. <https://doi.org/10.54728/JFMG.202310.00072>
- Pouransari, H., & Chalabi, M. (2014). Event-based stock market prediction using natural language processing. *International Journal of Computer Applications*.
- Raghupathi, W. (1996). A neural network approach to forecasting earnings per share. *Information Sciences*.

- Slavin, H. (2007). Aggregating earnings per share forecasts. *International Journal of Forecasting*.
- Zhang, Y., & Wang, J. (2020). Financial time series forecasting with machine learning: An overview. *IEEE Access*, 8, 23306-23326.
- Zhang, Z., Chen, Y., & Li, J. (2021). Random forest and gradient boosting for stock return prediction: Evidence from emerging markets. *Finance Research Letters*, 39, 101620. <https://doi.org/10.1016/j.frl.2020.101539>