KMANPUB | Decision Science and Intelligent Systems Vol. 2 No. 1 (2025): 1-25

Article history:
Received 23 December 2024
Revised 04 March 2025
Accepted 13 March 2025
Published online 09 April 2025

Evaluating The Efficiency of Iranian Listed Companies Using a Combined Method of Data Envelopment Analysis and Machine Learning

Mojtaba Ghiyasi[1*], Omid Valizadeh[2], Bahareh Joshani[3], Mohsen Lotfi[4]

[1] Associate Professor, Faculty of Industrial Engineering and Management, Shahrood University of Technology, Shahrood, Iran
[2] PhD Student in Industrial Management, Ferdowsi University of Mashhad, Mashhad, Iran
[3] Management Graduate, Shahrood University of Technology, Shahrood, Iran
[4] Assistant Professor, Faculty of Industrial Engineering and Management, Shahrood University of Technology, Shahrood, Iran

*** Corresponding author email address**: mogshu@gmail.com

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Ghiyasi, M., Valizadeh, O., Joshani, B., & Lotfi, M. (2025). Evaluating The Efficiency of Iranian Listed Companies Using a Combined Method of Data Envelopment Analysis and Machine Learning. *Decision Science and Intelligent Systems. 2(1), 1-25.*

© 2025 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

A B S T R A C T

This paper examines the efficiency of Iranian listed companies using data envelopment analysis and machine learning algorithms. First, the efficiency scores of 130 listed companies in the period from 2007 to 2013 were calculated using data envelopment analysis. Input variables included total debt and RETRSIK and output variables included annual stock return and liquidity. Then, the machine learning algorithm was used to identify and evaluate critical variables in predicting company performance. The results showed that the XGBoost algorithm had superior prediction performance. This study shows that the combination of data envelopment analysis and machine learning methods can be an effective tool for analyzing and predicting the efficiency of listed companies and help investors and managers make decisions. This research emphasizes the role of modern data analysis methods in improving economic and financial decision-making.

Keywords: *Iranian listed companies, performance evaluation, data envelopment analysis, machine learning*

Extended Abstract

Introduction

The capital market plays a pivotal role in facilitating financial resource allocation and fostering economic growth, particularly through stock exchanges that enhance transparency, attract both domestic and foreign investments, and stimulate various sectors (Birjandi et al., 2021). In Iran, listed companies form a central pillar of the Tehran Stock Exchange (TSE), contributing significantly to funding economic projects and expanding financial markets (Adabi et al., 2017). Established in 1967, the TSE has evolved under political, economic, and social influences, becoming an integral part of the country's economic framework (Keshavarz & Rezaei, 2022). Evaluating the performance of these companies is vital for optimizing resource utilization, improving competitiveness, and guiding managerial and investment decisions.

Data Envelopment Analysis (DEA) has emerged as a robust mathematical programming approach for assessing the relative efficiency of decision-making units (DMUs) without requiring predefined functional forms or parameter estimates (Panwar et al., 2022). DEA's non-parametric nature enables simultaneous handling of multiple inputs and outputs, making it ideal for comparing entities that share similar resource structures but achieve different outcomes (Fotova Čiković et al., 2022). Despite its strengths, DEA's accuracy depends heavily on input-output selection and data quality, and it faces challenges in accounting for environmental variables and standardizing benchmarks globally (Krmac & Mansouri Kaleibar, 2023). Within the capital market, DEA facilitates detailed benchmarking of firms based on profitability, return on capital, and optimal resource use (Fallah et al., 2020), allowing identification of inefficiencies and targeted improvement strategies (Li et al., 2020). Innovations in DEA have also introduced time-based performance rankings, capturing efficiency trends over multiple periods (Rahimi et al., 2022).

Parallel to DEA, machine learning (ML) has revolutionized predictive analytics by enabling systems to learn patterns from historical data and make accurate forecasts without explicit programming (Brunton et al., 2020). ML methods—including decision trees, random forests, support vector machines (SVM), and eXtreme Gradient Boosting (XGBoost)—offer advanced capabilities for modeling complex, nonlinear relationships and incorporating non-financial variables such as customer behavior, macroeconomic indicators, and market sentiment (Burkart & Huber, 2021; Bouasabah, 2024). In capital markets, ML can integrate structured and unstructured data, conduct sentiment analysis, and uncover hidden performance drivers that traditional methods often overlook (Jain et al., 2021; Jiang, 2021). Studies demonstrate ML's superior predictive accuracy for stock returns and corporate performance, especially when combined with feature selection and ensemble methods (Khan et al., 2022; Hegde & Rokseth, 2020). The integration of DEA and ML creates a hybrid analytical framework that not only benchmarks current efficiency but also forecasts future performance based on multidimensional datasets (Rouf et al., 2021).

Recent literature supports this integration. For example, Vijh et al. (2020) applied artificial neural networks and random forests to predict stock prices, achieving low RMSE and MAPE values, while Zaman et al. (2023) combined ML with sentiment and commodity price data, improving forecast accuracy to 87.2%. DEA-ML hybrids have been used to assess managerial capability impacts on firm performance (Ravanshad et al., 2020), predict financial distress (Rahimi et al., 2022), and rank sector-specific companies with multi-criteria decision-making methods (Shahriari, 2023). These studies highlight the

potential of combining DEA's efficiency measurement with ML's predictive power to create actionable insights for corporate governance, investment, and policy-making.

Against this backdrop, the present study evaluates the efficiency of 130 Iranian listed companies from 2007 to 2023 using a combined DEA–ML approach. First, DEA efficiency scores were computed based on carefully selected financial inputs and outputs. Then, multiple ML algorithms were trained to identify key predictors and assess their accuracy in forecasting efficiency. This dual approach addresses DEA's limitations by incorporating predictive modeling and provides a scalable framework for ongoing performance monitoring in volatile markets.

Methods and Materials

The study was conducted in two major phases. In the first phase, the efficiency scores of 130 Iranian listed companies were calculated annually over the 2007–2023 period using DEA. Input variables included total debt and RETRSIK (return risk), while output variables consisted of annual stock return and liquidity. DEA was implemented in both input-oriented and output-oriented models to ensure robustness. Data were collected from the Rahavard Novin software and validated through expert consultation.

In the second phase, the DEA efficiency scores served as the target variable in predictive models built with four ML algorithms: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost). Twenty-five financial and market-related indicators were used as input features, covering asset structure, capital composition, profitability, leverage, liquidity, growth, valuation ratios, and trading activity. Preprocessing steps included missing value assessment, outlier detection using the interquartile range method, and feature scaling via min–max normalization. The dataset was split into 80% training and 20% testing subsets. Model performance was evaluated using mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). All computations were conducted in Python 3.1.1.

Findings

The DEA results revealed notable fluctuations in company efficiency over the 17-year period. The average efficiency ranged from a low of 0.1269 in 2021 to a high of 0.4403 in 2019. From 2007 to 2012, scores remained relatively stable between 0.1580 and 0.2156, followed by a sharp rise in 2013 to 0.2712. A downward trend was observed between 2014 and 2017, reaching 0.1717, before surging in 2019. The steep decline in 2020 and 2021 corresponded with macroeconomic instability, including currency volatility and inflationary pressures. Efficiency partially recovered to 0.2390 in both 2022 and 2023.

In predictive modeling, XGBoost outperformed all other algorithms, achieving MAE = 0.0189, MSE = 0.0026, and RMSE = 0.0514. RF ranked second with slightly higher errors, followed by DT and SVM, the latter exhibiting the weakest performance with MAE = 0.1035, MSE = 0.0150, and RMSE = 0.1226. Scatter plots comparing predicted and actual DEA scores showed that XGBoost predictions closely aligned with the reference line, indicating strong model fit.

Feature importance analysis from XGBoost identified total assets from the previous year (16.66%) and total debt (14.88%) as the most influential predictors of efficiency. RETRSIK (7.36%), annual stock return (5.63%), and net profit (5.25%) also emerged as critical factors. Weighted average cost of capital (4.30%), market-to-book ratio (4.20%), and interest expenses (4.20%) had moderate importance. Several variables, including company size (SIZE), stock return (RETURN), and the number of shares, showed

negligible impact in this model. These results suggest that historical asset base, leverage, and market risk metrics are dominant drivers of efficiency in the Iranian capital market context.

Discussion and Conclusion

The integration of DEA and ML provided complementary strengths: DEA quantified relative efficiency among peer firms, while ML identified key determinants and delivered accurate forecasts. The superior performance of XGBoost aligns with prior findings on its robustness in handling complex, high-dimensional financial datasets (Ali et al., 2023; Bentéjac et al., 2021). The high importance of prior-year assets underscores the role of capital structure stability in sustaining operational efficiency, while the prominence of total debt reflects the balancing act between leverage and value creation in emerging markets. RETRSIK's significance highlights the influence of return volatility on investor confidence and firm performance.

The fluctuations in DEA scores mirror macroeconomic cycles and market shocks, suggesting that efficiency is sensitive to both firm-level management practices and external conditions. The steep drop during 2020–2021 coincided with heightened economic uncertainty, consistent with literature linking efficiency to systemic risk factors (Keshavarz & Rezaei, 2022). The negligible role of SIZE and RETURN in the predictive model indicates that scale and short-term market performance may be less relevant for long-term efficiency, contrasting with some traditional finance theories (Alarussi, 2021).

In conclusion, this study demonstrates that combining DEA with advanced ML algorithms, particularly XGBoost, can enhance the accuracy and interpretability of efficiency evaluations for listed companies. The methodology offers a replicable framework for investors, managers, and policymakers to monitor performance, diagnose inefficiencies, and prioritize strategic interventions. Future research could extend the model to incorporate macroeconomic indicators, sector-specific variables, and real-time market sentiment to further improve predictive capability and adaptability in dynamic market environments.

علم تصمیم گیری و سیستم های هوشمند دوره ۲، شماره ۱ (۱۴۰۴): ۲۵-۱

تاریخچه مقاله
دریافت شده در تاریخ ۳ دی ۱۴۰۳
اصلاح شده در تاریخ ۱۴ اسفند ۱۴۰۳
پذیرفته شده در تاریخ ۲۳ اسفند ۱۴۰۳
منتشر شده در تاریخ ۲۰ فروردین ۱۴۰۴

ارزیابی کارایی شرکت‌های بورسی ایران با استفاده از روش ترکیبی تحلیل پوششی داده و یادگیری ماشین

مجتبی غیاثی[1*]، امید ولیزاده[2]، بهاره جوشنی[3]، محسن لطفی[4]

۱. دانشیار، دانشکده مهندسی صنایع و مدیریت، دانشگاه صنعتی شاهرود، شاهرود، ایران
۲. دانشجوی دکتری مدیریت صنعتی، دانشگاه فردوسی مشهد، مشهد، ایران
۳. دانش آموخته مدیریت، دانشگاه صنعتی شاهرود، شاهرود، ایران
۴. استادیار، دانشکده مهندسی صنایع و مدیریت، دانشگاه صنعتی شاهرود، شاهرود، ایران

[*] **ایمیل نویسنده مسئول:** mogshu@gmail.com

اطلاعات مقاله

نوع مقاله
پژوهشی اصیل

نحوه استناد به این مقاله:
غیاثی، مجتبی.، ولیزاده، امید.، جوشنی، بهاره.، و لطفی، محسن. (۱۴۰۴). ارزیابی کارایی شرکت‌های بورسی ایران با استفاده از روش ترکیبی تحلیل پوششی داده و یادگیری ماشین. *علم تصمیم گیری و سیستم های هوشمند*، ۲(۱)، ۲۵-۱.

© ۱۴۰۴ تمامی حقوق انتشار این مقاله متعلق به نویسنده است. انتشار این مقاله به‌صورت دسترسی آزاد مطابق با گواهی (CC BY-NC 4.0) صورت گرفته است.

چکیده

این مقاله به بررسی کارایی شرکت‌های بورسی ایران با استفاده از روش تحلیل پوششی داده‌ها و الگوریتم‌های یادگیری ماشین می‌پردازد. ابتدا، نمرات کارایی ۱۳۰ شرکت بورسی در بازه زمانی ۱۳۸۶ تا ۱۴۰۲ با تحلیل پوششی داده‌ها محاسبه شده است. متغیرهای ورودی شامل جمع بدهی و RETRSIK و متغیرهای خروجی شامل بازده سالیانه سهام و نقد شوندگی انتخاب شدند. سپس، الگوریتم یادگیری ماشین برای شناسایی و ارزیابی متغیرهای حیاتی در پیش‌بینی عملکرد شرکت‌ها به کار گرفته شد. نتایج نشان داد که الگوریتم XGBoost عملکرد برتری در پیش‌بینی داشت. این مطالعه نشان می‌دهد که ترکیب روش‌های تحلیل پوششی داده‌ها و یادگیری ماشین می‌تواند ابزار مؤثری برای تحلیل و پیش‌بینی کارایی شرکت‌های بورسی باشد و به تصمیم‌گیری‌های سرمایه‌گذاران و مدیران کمک کند. این پژوهش به نقش روش‌های نوین تحلیل داده‌ها در بهبود تصمیم‌گیری‌های اقتصادی و مالی تأکید دارد.

کلیدواژگان: *شرکت‌های بورسی ایران، ارزیابی کارایی، تحلیل پوششی داده، یادگیری ماشین*

مقدمه

بورس اوراق بهادار به‌عنوان یکی از ابزارهای کلیدی تأمین مالی برای شرکت‌ها، تأثیر شگرفی در ایجاد شفافیت مالی، جذب سرمایه‌گذاری‌های داخلی و خارجی و توسعه بخش‌های مختلف اقتصادی دارد. این بازار می‌تواند با فراهم آوردن بستری برای ارتباط میان سرمایه‌گذاران و تولیدکنندگان و به‌ویژه با حمایت از کسب‌وکارهای نوپا، فرصت‌های جدیدی برای رشد و شکوفایی اقتصادی ایجاد کند (Birjandi et al., 2021). شرکت‌های بورسی در ایران نیز به‌عنوان یکی از ارکان اصلی بازار سرمایه، نقشی محوری در تأمین منابع مالی برای پروژه‌های اقتصادی و گسترش بازارهای مالی کشور ایفا می‌کنند. روند پیدایش و توسعه این شرکت‌ها در ایران، به تغییرات عمده‌ای در ساختار اقتصادی و نظام مالی کشور برمی‌گردد (Adabi et al., 2017). بورس اوراق بهادار تهران، که در سال ۱۳۴۶ تأسیس شد، در طول سال‌های فعالیت خود تحت تأثیر تحولات سیاسی، اقتصادی و اجتماعی کشور قرار گرفته و مسیر تکامل خود را پیموده است. (Keshavarz & Rezaei, 2022).

تحلیل پوششی داده (DEA[1]) یک روش ریاضیاتی و آماری است که برای ارزیابی کارایی واحدهای تصمیم‌گیری مختلف (DMUs[2]) به کار می‌رود و به‌طور خاص برای مقایسه عملکرد واحدهایی که منابع مشابهی دارند و نتایج مختلفی را تولید می‌کنند، بسیار مناسب است (Panwar et al., 2022). یکی از ویژگی‌های برجسته روش DEA این است که نیازی به فرضیات خطی و پارامتریک ندارد و قادر است بدون نیاز به مدل‌های فرضی خاص، به‌طور مستقیم کارایی واحدهای تصمیم‌گیری را ارزیابی کند. از دیگر قابلیت‌های روش DEA، مقایسه واحدهای مختلفی است که ورودی‌ها و خروجی‌های مشابه دارند اما در عملکرد خود تفاوت‌های معناداری نشان می‌دهند. همچنین، DEA توانایی بررسی هم‌زمان چندین ورودی و خروجی را دارد (Fotova Čiković et al., 2022). هرچند روش DEA ابزاری قوی برای ارزیابی کارایی است، اما برای دستیابی به نتایج دقیق و قابل‌اعتماد، باید از داده‌های صحیح استفاده کرد و محدودیت‌های این روش را در نظر گرفت. علاوه بر دقت داده‌ها، عدم توانایی روش DEA در شبیه‌سازی تأثیرات محیطی و نبود استاندارد جهانی برای ارزیابی کارایی از چالش‌های اصلی این روش محسوب می‌شود (Krmac & Mansouri Kaleibar, 2023).

تحلیل کارایی شرکت‌های بورسی با استفاده از روش DEA یکی از رویکردهای پیشرفته و کارآمد در ارزیابی عملکرد و بهره‌وری سازمان‌ها محسوب می‌شود. این روش، که کاربرد گسترده‌ای در مطالعات مالی و اقتصادی دارد، بر مبنای مقایسه ورودی‌ها و خروجی‌های مجموعه‌ای از واحدهای تصمیم‌گیرنده عمل کرده و امکان سنجش میزان کارایی نسبی آن‌ها را فراهم می‌آورد. در حوزه بازار سرمایه، روش DEA به‌عنوان ابزاری توانمند می‌تواند عملکرد شرکت‌های بورسی را بر اساس شاخص‌هایی همچون سودآوری، بازده سرمایه و میزان بهره‌گیری بهینه از منابع مالی مورد ارزیابی و مقایسه قرار دهد (Fallah et al., 2020). این روش به دلیل قابلیت‌های خاص خود در شناسایی نقاط قوت و ضعف واحدها، به‌ویژه برای شرکت‌های فعال در بورس اهمیت دارد. در محیط رقابتی بازار سرمایه، ارزیابی دقیق عملکرد شرکت‌ها می‌تواند به تصمیم‌گیری بهتر سرمایه‌گذاران و مدیران کمک کند. با استفاده از تحلیل پوششی داده‌ها، می‌توان نقاط بهبود در استفاده از منابع و تحقق اهداف مالی و اقتصادی را شناسایی کرد. این امر نه‌تنها به ارتقای کار آیی شرکت‌ها کمک می‌کند، بلکه در بلندمدت می‌تواند ارزش بازار آن‌ها را افزایش دهد. این مسئله به‌ویژه در دنیای امروز که داده‌ها در بسیاری از زمینه‌ها به‌سرعت در حال رشد هستند و نیاز به محاسبات زیادی دارند، اهمیت بیشتری پیداکرده است (Li et al., 2020). یکی از ویژگی‌های جالب این مدل، معرفی یک روش جدید برای رتبه‌بندی شرکت‌ها

[1] Data Envelopment Analysis
[2] Decision-Making Unit

است، این روش بر اساس روند ناکارآمدی در طول زمان عمل می‌کند و به تصمیم‌گیرندگان کمک می‌کند تا واحدهای ناکارآمد را در هر دوره زمانی شناسایی کنند. علاوه بر این، در این روش، امکان ارزیابی شیب کارایی شرکت‌ها در طول دوره‌های زمانی مختلف وجود دارد. این امر به شرکت‌ها این امکان را می‌دهد که بر اساس تغییرات کارایی‌شان در طول زمان رتبه‌بندی شوند (Rahimi et al., 2022a). در این مقاله، به بررسی کاربردهای تحلیل پوششی داده‌ها در ارزیابی کار آیی شرکت‌های بورسی پرداخته و نحوه استفاده از این روش برای بهبود عملکرد شرکت‌ها در بازارهای مالی و اقتصادی بررسی خواهد شد.

یادگیری ماشین[1] یکی از زیرشاخه‌های اصلی هوش مصنوعی است که در سال‌های اخیر نقش بسیار مهمی در تحول فناوری و تحلیل داده‌ها ایفا کرده است. این علم به سیستم‌ها امکان می‌دهد تا از طریق داده‌های موجود، الگوها را شناسایی کنند و بدون نیاز به دستورالعمل‌های صریح و از پیش نوشته‌شده، عملکرد خود را بهبود دهند. به‌عبارت‌دیگر، در یادگیری ماشین، سیستم‌ها قادرند از تجربه‌ها و داده‌ها درس بگیرند و این دانش را برای پیش‌بینی رویدادها، تصمیم‌گیری‌ها و حل مسائل پیچیده به‌کارگیرند .(Brunton et al., 2020) روش‌های یادگیری ماشین شامل مجموعه متنوعی از الگوریتم‌ها و تکنیک‌هاست که هر یک برای کاربردهای خاصی طراحی‌شده‌اند. این روش‌ها به سیستم‌ها کمک می‌کنند تا مشکلات مختلف را شناسایی و حل کنند و کارایی فرایندها را بهینه‌سازی نمایند. به‌طور مثال، از یادگیری ماشین می‌توان در شناسایی الگوهای رفتاری مشتریان، پیش‌بینی وضعیت بازارهای مالی، تشخیص بیماری‌ها، ترجمه زبان و حتی رانندگی خودکار استفاده کرد. یادگیری نظارت‌شده یکی از رایج‌ترین روش‌ها است که در آن مدل‌ها با استفاده از داده‌هایی که برچسب یا جواب صحیح دارند، آموزش می‌بینند. ویژگی کلیدی یادگیری ماشین این است که سیستم‌ها به‌مرورزمان و با دریافت داده‌های بیشتر، هوشمندتر و دقیق‌تر می‌شوند؛ این امر باعث شده است که این فناوری در بسیاری از زمینه‌ها مانند صنعت، پزشکی، تجارت وزندگی روزمره، جایگاه ویژه‌ای پیدا کند و به‌عنوان یکی از مهم‌ترین ابزارهای فناوری مدرن شناخته شود (Burkart & Huber, 2021).

مرور ادبیات و پیشینه تحقیق

ارزیابی کارایی شرکت‌های بورسی با روش‌های سنتی

روش‌های سنتی بر پایه بررسی اطلاعات مالی و صورت‌های مالی شرکت‌ها بنا شده و عمدتاً از داده‌های مالی گذشته برای ارزیابی عملکرد سازمان‌ها بهره می‌برند. این رویکردها نقش بسزایی در تحلیل بهره‌وری شرکت‌ها، شناسایی نقاط ضعف و قوت آن‌ها و پیش‌بینی بحران‌های مالی ایفا می‌کنند. از دیگر مزایای این روش‌ها می‌توان به سادگی و قابلیت فهم بالای آن‌ها اشاره کرد که امکان استفاده مؤثر در سطوح مختلف مدیریتی و مالی را فراهم می‌آورد (Narkunienė & Ulbinaitė, 2018). از جمله ابزارهای اصلی تحلیل مالی می‌توان به ترازنامه، صورت سود و زیان و صورت جریان وجوه نقد اشاره کرد. ترازنامه نمایانگر وضعیت مالی شرکت بوده و شامل دارایی‌ها، بدهی‌ها و حقوق صاحبان سهام است. صورت سود و زیان نیز عملکرد مالی شرکت را در یک دوره مشخص از طریق بررسی درآمدها، هزینه‌ها و میزان سود یا زیان ارزیابی می‌کند. در کنار این دو، صورت جریان وجوه نقد میزان توانایی شرکت در مدیریت نقدینگی و جریان‌های نقدی را مشخص می‌سازد.علاوه بر تحلیل صورت‌های مالی، یکی از روش‌های رایج سنتی در ارزیابی عملکرد شرکت‌ها، تحلیل نسبت‌های مالی است که در حوزه‌های گوناگونی کاربرد دارد. نسبت‌های سودآوری همچون بازده دارایی‌ها و بازده حقوق صاحبان سهام، میزان توانایی شرکت در ایجاد سود را نشان می‌دهند. نسبت‌های نقدینگی مانند نسبت جاری و نسبت سریع، توانایی شرکت را در پرداخت بدهی‌های کوتاه‌مدت ارزیابی می‌کنند. همچنین، نسبت‌های اهرمی از جمله نسبت بدهی به دارایی، میزان وابستگی شرکت به منابع مالی خارجی را مشخص می‌سازد. این نسبت‌ها

[1] Machine Learning

ابزاری کارآمد برای تحلیل عملکرد شرکت در زمینه‌های مختلفی همچون سودآوری، توانایی بازپرداخت بدهی‌ها و میزان اتکا به تأمین مالی خارجی محسوب می‌شوند (Alarussi, 2021). روش دیگری که برای ارزیابی عملکرد مالی شرکت‌ها به کار می‌رود، تحلیل روند است؛ این روش به بررسی تغییرات مالی یک شرکت در طول زمان می‌پردازد تا مشخص شود که آیا عملکرد آن در حال بهبود است یا دچار افت شده است. تحلیل روند، تصویری روشن از وضعیت شرکت ارائه می‌دهد و نشان می‌دهد که آیا شرکت در مسیر رشد و توسعه قرار دارد یا با چالش‌ها و مشکلاتی مواجه است (Nti et al., 2020). یکی از روش‌های اساسی در ارزیابی عملکرد شرکت‌ها، مقایسه آن‌ها با شرکت‌های مشابه در همان صنعت است که از آن با عنوان بنچمارکینگ (Benchmarking) یاد می‌شود. در این رویکرد، شاخص‌های مالی یک شرکت با میانگین صنعت یا عملکرد رقبا سنجیده می‌شود تا میزان موفقیت و جایگاه آن در مقایسه با سایر شرکت‌ها مشخص گردد. این مقایسه امکان شناسایی نقاط قوت و ضعف را فراهم کرده و به مدیران در اتخاذ تصمیم‌های راهبردی برای بهبود عملکرد شرکت کمک می‌کند (Ruiz & Sirvent, 2022). روش دقیق‌تری که برای ارزیابی عملکرد مالی شرکت‌ها مورد استفاده قرار می‌گیرد، تحلیل دوپونت نام دارد؛ این روش به شفاف‌سازی منابع سودآوری یک شرکت کمک کرده و نشان می‌دهد که بازده حقوق صاحبان سهام (ROE) از چه عواملی تأثیر می‌پذیرد. در این تحلیل، سودآوری شرکت بر اساس سه مؤلفه اساسی بررسی می‌شود: میزان سود حاصل از فروش، میزان کارایی و بهره‌وری در استفاده از دارایی‌های شرکت و میزان استفاده از بدهی‌ها برای تأمین مالی. تحلیل دوپونت با تفکیک این عوامل، دیدگاهی جامع از وضعیت مالی شرکت ارائه می‌دهد و به سرمایه‌گذاران و مدیران کمک می‌کند تا تصمیمات آگاهانه‌تری در زمینه راهبردهای مالی اتخاذ کنند (Arsad et al., 2022). علاوه بر این، برخی از شاخص‌های بازار نقش مهمی در ارزیابی عملکرد شرکت‌ها ایفا می‌کنند. برای نمونه، نسبت قیمت به درآمد (P/E) بیانگر آن است که قیمت سهام شرکت در مقایسه با سودآوری آن تا چه میزان بالا یا پایین ارزیابی می‌شود. نسبت ارزش دفتری به ارزش بازار (P/B) نیز وضعیت ارزش بازاری شرکت را در مقایسه با ارزش حسابداری آن مشخص می‌سازد. از سوی دیگر، بازده سود نقدی نشان‌دهنده درصدی از سود شرکت است که به سهامداران پرداخت می‌شود. این شاخص‌ها ابزارهایی کلیدی برای سرمایه‌گذاران به شمار می‌روند و آن‌ها را در تصمیم‌گیری درباره خرید یا فروش سهام یاری می‌کنند.با این حال، روش‌های سنتی ارزیابی بازار دارای محدودیت‌هایی هستند. نخست آنکه این روش‌ها عمدتاً بر داده‌های تاریخی متکی‌اند و ممکن است در پیش‌بینی تغییرات ناگهانی بازار یا شرایط اقتصادی ناتوان باشند. علاوه بر این، تمرکز اصلی این روش‌ها بر داده‌های مالی است، درحالی‌که عوامل غیرمالی نظیر نوآوری، کیفیت مدیریت و میزان رضایت مشتریان را در نظر نمی‌گیرند. ازاین‌رو، در شرایط پیچیده یا هنگام وقوع تغییرات سریع اقتصادی، ممکن است این روش‌ها به نتایج نادرستی منجر شوند. با وجود این محدودیت‌ها، همچنان این شاخص‌ها به‌عنوان ابزارهای اصلی در ارزیابی عملکرد شرکت‌های بورسی مورد استفاده قرار می‌گیرند و نقشی اساسی در تصمیم‌گیری‌های سرمایه‌گذاری ایفا می‌کنند (Zhang et al., 2022).

نقش یادگیری ماشین در ارزیابی کارایی شرکت‌های بورسی

یادگیری ماشین به‌عنوان یک شاخه از هوش مصنوعی، توانسته است تأثیر زیادی در تحلیل داده‌های مالی و ارزیابی کارایی شرکت‌های بورسی بگذارد. این روش با استفاده از الگوریتم‌های پیچیده قادر است پیش‌بینی‌های دقیق‌تری نسبت به روش‌های سنتی انجام دهد و به تحلیل‌گران کمک می‌کند تا با استفاده از داده‌های تاریخی شرکت‌ها و بازار، روندهای آینده را شبیه‌سازی کرده و تصمیمات بهتری در زمینه سرمایه‌گذاری اتخاذ کنند. استفاده از یادگیری ماشین در ارزیابی کارایی شرکت‌ها به‌ویژه در بورس و بازارهای مالی، این امکان را فراهم می‌آورد که تحلیل‌های دقیق‌تری انجام شود و نقص‌های روش‌های سنتی که معمولاً بر داده‌های گذشته تکیه دارند، جبران گردد. یادگیری ماشین با استفاده از الگوریتم‌های پیشرفته قادر به تحلیل داده‌های غیرمالی مانند رفتار مشتری، وضعیت بازار و عوامل اقتصادی خارجی است (Bouasabah, 2024). مدل‌های یادگیری ماشین مانند درخت تصمیم، شبکه‌های عصبی و ماشین‌های بردار پشتیبان می‌توانند پیش‌بینی‌هایی

دقیق‌تر و شخصی‌سازی‌شده برای ارزیابی کارایی ارائه دهند. این مدل‌ها با استفاده از داده‌های بزرگ قادر به تحلیل و پیش‌بینی روند بازار و رفتار سهام به صورت دقیق و کارآمد هستند؛ درواقع یادگیری ماشین قادر است حجم بالایی از داده‌ها را تجزیه و تحلیل کرده و الگوهای پنهانی را شبیه‌سازی کند که در روش‌های سنتی قابل شناسایی نیستند. این قابلیت باعث می‌شود که تحلیل‌گران بتوانند به نتایج دقیق‌تری دست یابند و در تصمیمات سرمایه‌گذاری خود بهتر عمل کنند (Jiang, 2021). یادگیری ماشین می‌تواند برای پیش‌بینی بازده سهام یا سودآوری شرکت‌ها از داده‌های مالی مختلفی مانند قیمت سهام، سودآوری، حجم معاملات و دیگر شاخص‌های مالی استفاده کند (Khan et al., 2022). علاوه بر این، یادگیری ماشین می‌تواند در تحلیل ریسک‌ها نیز مؤثر باشد؛ مدل‌های یادگیری ماشین قادر به شبیه‌سازی و مدل‌سازی ریسک‌های مختلف اقتصادی و بازار هستند و تأثیر عواملی مانند تغییرات در نرخ بهره، بحران‌های اقتصادی یا نوسانات بازار را بر عملکرد شرکت‌ها تحلیل می‌کنند، این امر به سرمایه‌گذاران کمک می‌کند تا استراتژی‌های بهتری برای مدیریت ریسک اتخاذ کنند (Hegde & Rokseth, 2020). همچنین یادگیری ماشین می‌تواند در تحلیل داده‌های غیرساختاری مانند اخبار بازار، نظرات مشتریان، رسانه‌های اجتماعی و تحلیل‌های اقتصادی کمک کند. با استفاده از تحلیل احساسات (Sentiment Analysis) از این داده‌ها، می‌توان پیش‌بینی‌های دقیقی از تغییرات قیمت سهام و بازده شرکت‌ها در آینده انجام داد (Jain et al., 2021). همینطور، یادگیری ماشین امکان شناسایی شاخص‌های جدید و پنهان را فراهم می‌آورد که ممکن است در مدل‌های سنتی نادیده گرفته شوند. از دیگر مزایای این روش‌ها می‌توان به کاهش خطاهای انسانی و بهبود تصمیم‌گیری‌های سرمایه‌گذاری اشاره کرد. در نتیجه، یادگیری ماشین توانسته است به ابزاری ارزشمند در ارزیابی کارایی شرکت‌های بورسی تبدیل شود که علاوه بر افزایش دقت، توانایی پیش‌بینی آینده را نیز دارد (Rouf et al., 2021).

در ادامه به برخی از مطالعات در زمینه تحلیل پوششی داده و هوش مصنوعی نیز اشاره شده است:

مهار ویج و همکارانش در این مطالعه، از تکنیک‌های شبکه عصبی مصنوعی و جنگل تصادفی برای پیش‌بینی قیمت بسته شدن سهام در روز بعد برای پنج شرکت از بخش‌های مختلف اقتصادی استفاده کرده‌اند. داده‌های مالی شامل قیمت‌های باز، بالا، پایین و بسته سهام هستند که برای ایجاد متغیرهای جدید و ورودی به مدل‌ها به کار می‌روند. مدل‌ها با استفاده از دو شاخص استاندارد برای ارزیابی عملکرد خود، یعنی RMSE و MAPE، سنجیده می‌شوند. مقادیر پایین این دو شاخص نشان‌دهنده کارآمد بودن مدل‌ها در پیش‌بینی دقیق قیمت بسته شدن سهام است (Vijh et al., 2020).

وصیت خان و همکارانش در این تحقیق به بررسی تأثیر داده‌های رسانه‌های اجتماعی و اخبار مالی بر دقت پیش‌بینی بازار سهام برای ده روز آینده پرداخته و با استفاده از الگوریتم‌های مختلف یادگیری ماشین، به بهبود عملکرد مدل‌ها از طریق انتخاب ویژگی و حذف توییت‌های هرزنامه پرداخته است. نتایج نشان می‌دهند که بالاترین دقت پیش‌بینی با استفاده از رسانه‌های اجتماعی ۸۰.۵۳٪ و با اخبار مالی ۷۵.۱۶٪ به‌دست‌آمده است. همچنین، پیش‌بینی بازارهای نیویورک و ردهت دشوارتر است، به‌طوری‌که سهام نیویورک و IBM بیشتر تحت تأثیر رسانه‌های اجتماعی و سهام لندن و مایکروسافت بیشتر از اخبار مالی تأثیر می‌پذیرند. طبقه‌بندی جنگل تصادفی بهترین عملکرد را با دقت ۸۳.۲۲٪ داشته است (Khan et al., 2022).

سیدرا مهتاب و همکارش در این مقاله چارچوبی قوی و دقیق برای پیش‌بینی قیمت سهام ارائه می‌دهند که از مدل‌های آماری، یادگیری ماشینی و یادگیری عمیق بهره می‌برد. داده‌های روزانه قیمت سهام یک شرکت مشهور در بورس ملی هند (NSE) که در فواصل پنج‌دقیقه‌ای جمع‌آوری‌شده‌اند، برای ساخت مدل‌های پیش‌بینی استفاده می‌شود. نتایج عملکرد این مدل‌ها به‌طور گسترده بررسی‌شده است (Mehtab & Sen, 2020).

دیپاک کومار و همکارانش در این مقاله به بررسی ۳۰ تحقیق می‌پردازند که روش‌ها، الگوریتم‌های یادگیری ماشینی، پارامترهای عملکرد و مجلات معتبر را معرفی می‌کنند. مطالعات انتخاب‌شده به شناسایی تکنیک‌های یادگیری ماشینی و مجموعه داده‌های مربوطه برای پیش‌بینی بازار سهام کمک می‌کنند. بیشتر تکنیک‌ها مانند ANN و NN برای دستیابی به پیش‌بینی دقیق‌تر استفاده می‌شوند، هرچند که روش‌های جدید همچنان با محدودیت‌هایی مواجه‌اند. درنهایت، پیش‌بینی بازار سهام باید فرآیندی یکپارچه باشد که در آن پارامترهای دقیق‌تری برای پیش‌بینی در نظر گرفته شوند (Kumar et al., 2022).

درون آوراموف و همکارانش در این مقاله نشان می‌دهند که سرمایه‌گذاری بر اساس سیگنال‌های یادگیری عمیق می‌تواند از سهامی که آربیتراژ آن‌ها دشوار است، سود خوبی کسب کند، مخصوصاً در شرایطی که بازار محدودیت‌های زیادی دارد. با این حال، این سود در شرکت‌های کوچک، سهام نوسانی یا دوره‌های پرنوسان بازار کاهش پیدا می‌کند. عملکرد روش‌های یادگیری ماشین زمانی که هزینه‌های معاملاتی در حد معقول باشد، ضعیف‌تر می‌شود، زیرا گردش مالی بالا و موقعیت‌های پرریسک در سبد سهام به دلیل نوسانات قیمت ایجاد می‌شود. با وجود پیچیدگی، این روش‌ها می‌توانند سهامی را که قیمت‌گذاری درستی ندارد، شناسایی کنند. در نهایت، سیگنال‌های یادگیری عمیق در بلندمدت سودآور بوده و ریسک پایینی دارند (Avramov et al., 2023).

نور زمان و همکارانش در این مطالعه مدلی مبتنی بر یادگیری ماشین پیشنهاد داده‌اند که داده‌های تاریخی را با عواملی بیرونی مانند احساسات موجود در رسانه‌های اجتماعی، روند نوسان نفت و طلا و داده‌های اخبار مالی ترکیب می‌کند تا دقت پیش‌بینی‌ها را به طرز چشم‌گیری افزایش دهد. برای ارزیابی اثربخشی این مدل، از داده‌های بازار بورس شرکت‌های HPQ، IBM، ORCL و MSFT بهره برده اند و تأثیرات شیوع ویروس کووید-۱۹ بر عملکرد شرکت‌ها را نیز مورد تحلیل قرار داده اند. نتایج تجربی نشان داد که دقت پیش‌بینی مدل ما با استفاده از داده‌های نفت و احساسات به ۸۷.۲٪ رسید. علاوه بر این، مشاهده شد که رسانه‌های اجتماعی تأثیر قابل‌توجهی بر سهام شرکت IBM دارند و الگوریتم GBM توانست نتایج دقیق و پایداری را ارائه دهد (Zaman et al., 2023).

عامر قاسم و همکارانش در این مقاله، به‌ضرورت توسعه چارچوبی برای کمک به سرمایه‌گذاران در تصمیم‌گیری‌های روزمره خود با استفاده از اطلاعات صورت‌های مالی به شیوه‌ای خودکار پرداخته‌اند. آنها در این مقاله یک سیستم هوش مصنوعی مبتنی بر یادگیری ماشین پیشنهاد می‌دهند که اطلاعات تاریخی قیمت و همچنین اقلام مختلف ترازنامه، سود و زیان و جریان نقدی شرکت‌ها را وارد کرده و احتمال کاهش قیمت سهام آن‌ها را پیش‌بینی می‌کند. این پیش‌بینی بر انتخاب خودکار ویژگی‌های مهم حسابداری و تاریخی متکی است و می‌تواند مسائل خاص شرکت‌ها را که برای تصمیم‌گیری سرمایه‌گذاران مهم است، روشن کند (Bogdanova & Stancheva-Todorova, 2021).

هدف محمدرضا روان شاد و همکارانش از این مطالعه بررسی رابطه بین توانمندی مدیریتی و عملکرد شرکت‌ها است. ابتدا یک مدل جدید DEA دو مرحله‌ای با رویکرد برنامه‌ریزی چندهدفه فازی برای ارزیابی عملکرد شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران معرفی می‌شود. نتایج نشان می‌دهد که رابطه‌ای مثبت بین توانمندی مدیریتی و عملکرد شرکت وجود دارد؛ به این معنا که توانمندی مدیریتی به طور قابل توجهی با عملکرد شرکت مرتبط است به عبارت دیگر، با افزایش توانمندی مدیریتی، عملکرد شرکت بهبود یافته و منابع بهتر استفاده شده و در نتیجه کارایی کلی افزایش می‌یابد (Ravanshad et al., 2020).

فلاح و همکارانش در پژوهش حاضر به استفاده از تکنیک‌های تحلیل تمایزی و تحلیل پوششی داده‌ها می‌پردازند. هدف اصلی این تکنیک‌ها طراحی مدل‌هایی است که بتوانند شرکت‌ها را بر اساس عملکردشان به دو گروه موفق و ناموفق تفکیک کنند. تحلیل پوششی داده‌ها روشی است که با اندازه‌گیری کارایی، واحدهای کارآمد را از ناکارآمد جدا می‌کند. در این پژوهش، مدل‌سازی بر روی شرکت‌های نفتی پذیرفته‌شده در بورس ایران انجام شده است. این پژوهش با ترکیب تحلیل پوششی داده‌ها و تحلیل تمایزی، یک مدل جدید ارائه کرده است

که ابرصفحه‌هایی برای جداسازی دو گروه (موفق و ناموفق) طراحی می‌کند. نتایج نشان می‌دهد این روش می‌تواند به پیش‌بینی وضعیت شرکت‌ها کمک کند و راهکاری برای شناسایی مسیرهای موفقیت یا جلوگیری از شکست ارائه دهد (Fallah et al., 2020).

رحیمی و همکارانش این مطالعه را با هدف شناسایی معیارهای مالی مناسب برای تحلیل و پیش‌بینی وضعیت ورشکستگی مالی شرکت‌های فعال در بورس تهران انجام داده اند. مدل تحلیل پوششی داده‌ها - تحلیل تمایز (DEA-DA) به‌عنوان مدل پیشنهادی این تحقیق اجرا شده است. نتایج نشان داد که مدل پیشنهادی توانایی بالایی در پیش‌بینی صحیح وضعیت مالی شرکت‌های جدید دارد. این روش، ابزاری پویا برای مدیران و سرمایه‌گذاران فراهم می‌کند تا با استفاده از تحلیل‌های مالی جامع، تصمیمات بهتری در زمینه مدیریت و سرمایه‌گذاری اتخاذ کنند و از خطرات احتمالی بحران مالی جلوگیری کنند (Rahimi et al., 2022b).

هدف شهریاری و همکارانش در این مطالعه، بررسی و رتبه‌بندی عملکرد شرکت‌های دارویی فعال در بورس تهران با استفاده از روش‌های DEA و تصمیم‌گیری چند معیاره (MCDM) است. روش‌هایی همچون DEA و MCDM به دلیل قابلیت بررسی هم‌زمان چندین عامل و شاخص، می‌توانند ابزاری مناسب برای ارزیابی عملکرد شرکت‌ها و به‌دست آوردن شاخص‌های عملکرد قابل اعتماد باشند. استفاده از این روش‌ها می‌تواند تحلیلی جامع از عملکرد شرکت‌ها ارائه دهد و ابزاری مؤثر برای کمک به تصمیم‌گیری‌های مدیریتی و سرمایه‌گذاری در اختیار سرمایه‌گذاران و مدیران قرار دهد. این رویکردها می‌توانند به شناخت بهتر نقاط قوت و ضعف شرکت‌ها کمک کنند و از این طریق، به بهبود فرآیندهای تصمیم‌گیری و تخصیص منابع در بازار بورس منجر شوند (Shahriari, 2023).

روش تحقیق

تحلیل پوششی داده (Data envelopment analysis)

تحلیل پوششی داده‌ها روشی مبتنی بر برنامه‌ریزی ریاضی جهت تعیین کارایی مجموعه‌ای از واحدهای تصمیم‌گیرنده‌ی با چندین ورودی و چندین خروجی می‌باشد. با استفاده از تحلیل پوششی داده‌ها واحدهای کارا و ناکارا مشخص می‌شوند. از مدل‌های شناخته‌شده DEA می‌توان به دو نمونه ورودی گرا و خروجی گرا اشاره کرد. در رویکرد ورودی گرا؛ هدف حداقل ورودی‌های لازم جهت تولید سطح فعلی خروجی‌ها می‌باشد. این مدل با ثابت نگه‌داشتن خروجی‌ها، ورودی‌ها را کاهش می‌دهد تا واحد موردنظر به مرز کارایی برسد. در رویکرد خروجی گرا؛ هدف یافتن حداکثر خروجی‌های ممکن با استفاده از سطح فعلی ورودی‌ها می‌باشد. این مدل با ثابت نگه‌داشتن ورودی‌ها، خروجی‌ها را افزایش می‌دهد تا واحد موردنظر به مرز کارایی برسد. درروش DEA، DMU ها در حالی ارزیابی می‌شوند که هرکدام از m ورودی متمایز استفاده می‌کنند و s خروجی متنوع را تولید می‌کنند.

$$\theta_j = \frac{\sum_{r=1}^{s} v_r y_{rj}}{\sum_{i=1}^{m} u_i x_{ij}} \qquad (1)$$

در فرمول شماره ۱، DMU_j مقادیر ورودی $X_{ij}\ (i=1,\ldots,m)$ را جهت تولید مقادیر خروجی $y_{rj}\ (r=1,\ldots,s)$ مصرف می‌کند. $v_r = (r=1,\ldots,s)$ به‌عنوان اوزان خروجی و همچنین $u_i (i=1,\ldots,m)$ به‌عنوان اوزان ورودی در نظر گرفته می‌شوند. چارنز و همکاران مدل معروف CCR را توسعه دادند که کارایی DMU ها را به‌صورت زیر محاسبه می‌کند:

$$\max \theta_o = \frac{\sum_{r=1}^{s} v_r y_{ro}}{\sum_{i=1}^{m} u_i x_{io}}$$

$$s.t. \theta_j = \frac{\sum_{r=1}^{s} v_r y_{rj}}{\sum_{i=1}^{m} u_i x_{ij}} \leq 1; \quad j = 1,...,n \qquad (2)$$

$$v_r, u_i \geq \varepsilon, r = 1,...,s \qquad i = 1,...,m$$

که در آن DMUo نشان‌دهنده DMU تحت ارزیابی می‌باشد، $u_i (i = 1,...,m)$ و $v_r (r = 1,...,s)$ به‌عنوان متغیرهای تصمیم شناخته می‌شوند. مدل برنامه‌نویسی کسری فوق را می‌توان به مسئله برنامه‌ریزی خطی (LP) زیر تبدیل کرد:

$$\max \theta_o = \sum_{r=1}^{s} v_r y_{ro}$$

$$s.t$$

$$\sum_{r=1}^{s} v_r y_{rj} - \sum_{i=1}^{m} u_i x_{ij} \leq 0$$

$$\sum_{i=1}^{m} u_i x_{io} = 1$$

$$v_r, u_i \geq \varepsilon$$

$$i = 1,...,m$$

$$r = 1,...,s \qquad (3)$$

مدل‌های یادگیری ماشینی (ML)

یادگیری ماشین به‌عنوان ابزاری قدرتمند برای شناسایی و مدل‌سازی روابط پیچیده میان متغیرهای ورودی و خروجی به کار گرفته می‌شود. در این فرآیند، مدل‌ها به‌گونه‌ای آموزش می‌بینند که توانایی پیش‌بینی دقیق و ارائه تحلیل‌های معتبر بر روی داده‌های جدید را داشته باشند. استفاده از این روش‌ها، به دلیل توانایی آن‌ها در شناسایی تغییرات الگوها و روابط داده‌ها، کمک می‌کند تا تحلیل دقیق‌تری از عملکرد ارائه شود و از طریق بهبود مدل‌سازی، نتایج قابل‌اعتمادی به دست آید. در این تحقیق جهت تحلیل و پیش‌بینی کارایی شرکت‌های بورسی، داده‌ها به نسبت ۸۰ درصد برای آموزش و ۲۰ درصد برای آزمایش تقسیم شدند. هدف از این فرآیند، شناسایی روابط غیرخطی و پیچیده میان متغیرهای ورودی و خروجی و بهبود دقت پیش‌بینی‌ها بود (Rouf et al., 2021).

درخت تصمیم (Decision Tree)

درخت تصمیم یکی از روش‌های یادگیری ماشین تحت نظارت است که به دلیل سادگی و کارایی، به‌طور گسترده در مسائل دسته‌بندی و رگرسیون مورداستفاده قرار می‌گیرد. این الگوریتم با ارائه ساختاری گرافیکی، تصمیم‌گیری‌ها و نتایج ممکن را به‌صورت شاخه‌های یک درخت نمایش می‌دهد (Bansal et al., 2022). فرآیند ساخت درخت تصمیم با تقسیم‌بندی داده‌ها بر اساس ویژگی‌های کلیدی آغاز می‌شود. در هر مرحله، یک گره تصمیم، شرایط خاصی را تعیین می‌کند و داده‌ها را به شاخه‌های مرتبط هدایت می‌کند. این روند تا زمانی ادامه می‌یابد که به گره‌های برگ برسیم، جایی که نتایج یا پیش‌بینی‌های نهایی ارائه می‌شوند (Charbuty & Abdulazeez, 2021). ازجمله ویژگی‌های بارز درخت تصمیم می‌توان به‌سادگی و شفافیت آن اشاره کرد. ساختار این الگوریتم به‌گونه‌ای است که حتی افراد بدون دانش تخصصی نیز می‌توانند عملکرد و نتایج آن را به‌راحتی درک کنند. همچنین انعطاف‌پذیری درخت تصمیم در کار با داده‌های عددی و طبقه‌بندی‌شده، آن را به ابزاری کارآمد برای تحلیل انواع داده‌ها تبدیل کرده است (Zhou et al., 2021). کاربردهای درخت تصمیم بسیار متنوع است؛ این الگوریتم در طبقه‌بندی داده‌ها، مانند پیش‌بینی نوع مشتری یا تشخیص بیماری و همچنین در پیش‌بینی متغیرهای عددی و روندها استفاده می‌شود. علاوه بر این، در مسائل پیچیده تصمیم‌گیری نیز می‌تواند راه‌حلی شفاف و مؤثر ارائه دهد (Tangirala, 2020). درمجموع، درخت تصمیم به‌عنوان یکی از ابزارهای ساده و کارآمد در تحلیل داده‌ها و یادگیری ماشین شناخته می‌شود؛ با استفاده از ساختار گرافیکی و توانایی تقسیم‌بندی دقیق داده‌ها، این الگوریتم امکان تصمیم‌گیری بهینه و پیش‌بینی‌های دقیق را فراهم می‌کند (Costa & Pedreira, 2023).

جنگل تصادفی (Random Forest)

جنگل تصادفی یکی از الگوریتم‌های قدرتمند و پرکاربرد در یادگیری ماشین است که با استفاده از ترکیب چندین درخت تصمیم‌گیری برای بهبود دقت پیش‌بینی‌ها در مسائل مختلف، ازجمله مسائل دسته‌بندی و رگرسیون، عمل می‌کند. در این روش، تعدادی درخت تصمیم‌گیری به‌طور تصادفی ساخته می‌شوند و هر درخت از زیرمجموعه‌ای تصادفی از داده‌ها آموزش می‌بینند. علاوه بر این، در هنگام تقسیم داده‌ها در هر گره، تنها یک زیرمجموعه تصادفی از ویژگی‌ها برای انتخاب بهترین ویژگی در نظر گرفته می‌شود. این نیز به ایجاد تنوع بیشتر و کاهش هم‌خطی میان ویژگی‌ها کمک می‌کند و از بروز مشکلاتی مانند Overfitting جلوگیری می‌کند. در مسائل دسته‌بندی، نتیجه نهایی بر اساس رأی اکثریت درخت‌ها تعیین می‌شود؛ به‌عبارت‌دیگر، کلاس پیش‌بینی‌شده آن کلاسی است که بیشترین رأی را از سوی درخت‌ها کسب کرده است. در مسائل رگرسیون، پیش‌بینی نهایی از میانگین پیش‌بینی‌ها توسط تمام درخت‌ها به‌دست می‌آید (Salman et al., 2024). جنگل تصادفی به دلیل استفاده از مجموعه‌ای از مدل‌های ضعیف (درخت‌های تصمیم‌گیری) که به‌صورت تصادفی ساخته می‌شوند، معمولاً دقت بالایی در پیش‌بینی‌ها دارد. این الگوریتم در برابر نویز موجود در داده‌ها نیز مقاوم است؛ این ویژگی باعث می‌شود که مدل جنگل تصادفی در شرایط مختلف و با داده‌های پیچیده عملکرد خوبی داشته باشد. اگرچه جنگل تصادفی دقت بالایی دارد، اما زمان پیش‌بینی در این الگوریتم ممکن است کندتر از مدل‌های ساده‌تر باشد و نیاز به حافظه بیشتری داشته باشد، زیرا به‌طور هم‌زمان چندین درخت تصمیم‌گیری را نگه‌داری و پردازش می‌کند. بااین‌حال، این مشکلات به‌راحتی با افزایش قدرت محاسباتی یا بهینه‌سازی‌های خاص قابل مدیریت است (Talekar & Agrawal, 2020). به‌طورکلی، جنگل تصادفی به‌عنوان یک روش ترکیبی کارآمد و قدرتمند، در بسیاری از کاربردهای عملی یادگیری ماشین ازجمله تحلیل داده‌های پزشکی، تشخیص تقلب، تحلیل بازارهای مالی، پردازش تصویر و... مورداستفاده قرار می‌گیرد و توانسته است اعتبار و شهرت بالایی در میان پژوهشگران و متخصصان یادگیری ماشین پیدا کند (Hu & Szymczak, 2023).

ماشین‌های بردار پشتیبانی (Support Vector Machine)

ماشین بردار پشتیبان (SVM) یکی از الگوریتم‌های یادگیری ماشین نظارتی است که برای حل مسائل طبقه‌بندی و رگرسیون به کار می‌رود. هدف اصلی این الگوریتم یافتن یک ابر صفحه (Hyperplane) مناسب است که بتواند داده‌های متعلق به کلاس‌های مختلف را در یک فضای چندبعدی از یکدیگر جدا کند. این الگوریتم با تعیین مرزی که بیشترین فاصله را از نزدیک‌ترین داده‌های هر کلاس دارد، تصمیم‌گیری بهینه‌ای ارائه می‌دهد (Otchere et al., 2021). SVM برای دستیابی به مرزهای جداکننده دقیق، داده‌ها را به فضایی با ابعاد بالاتر منتقل می‌کند. این انتقال با استفاده از توابعی به نام کرنل (Kernel Function) صورت می‌گیرد که امکان انجام محاسبات فضای بالاتر را بدون نیاز به مشخص کردن مختصات آن فراهم می‌کنند. ازجمله مهم‌ترین توابع کرنل می‌توان به کرنل چندجمله‌ای، کرنل شعاع پایه (RBF) و کرنل سیگموید اشاره کرد. این روش به الگوریتم SVM این قابلیت را می‌دهد که حتی داده‌های غیرخطی و پیچیده را نیز به شکلی مؤثر طبقه‌بندی کند (Ahsaan et al., 2022). یکی از مزایای برجسته SVM توانایی آن در کار با داده‌هایی است که ویژگی‌های زیادی دارند، اما تعداد نمونه‌ها محدود است. این الگوریتم با استفاده از اصول بهینه‌سازی، به‌دقت بالایی در تحلیل داده‌های جدید و دیده‌نشده دست می‌یابد. همچنین، با بهره‌گیری از توابع کرنل، SVM قادر است مرزهای پیچیده و غیرخطی را مدیریت کرده و داده‌های متنوع را به‌درستی طبقه‌بندی کند (Sha'Abani et al., 2020). SVM به دلیل دقت و انعطاف‌پذیری خود، در حوزه‌های مختلف کاربردهای گسترده‌ای دارد. این الگوریتم در طبقه‌بندی تصاویر، مانند شناسایی و دسته‌بندی انواع تصاویر و طبقه‌بندی متون، مانند تشخیص هرزنامه‌ها از ایمیل‌های عادی، نقش بسزایی ایفا می‌کند. همچنین در زیست اطلاعات برای تحلیل داده‌های ژنتیکی و پزشکی و در پیش‌بینی‌های مالی، ازجمله پیش‌بینی قیمت سهام و تحلیل روندهای بازار، مورداستفاده قرار می‌گیرد (Chandra & Bedi, 2021). درمجموع، SVM به‌عنوان یکی از الگوریتم‌های برجسته یادگیری ماشین، به دلیل دقت بالا، توانایی مدیریت داده‌های غیرخطی و قابلیت تعمیم‌پذیری، ابزار ارزشمندی در حل مسائل پیچیده محسوب می‌شود. این ویژگی‌ها آن را به یکی از محبوب‌ترین انتخاب‌ها در میان الگوریتم‌های یادگیری ماشینی تبدیل کرده است (Shahin et al., 2023).

الگوریتم (Extreme Gradient Boosting)

در دنیای یادگیری ماشین، انتخاب الگوریتم مناسب برای حل مسائل یکی از مهم‌ترین چالش‌ها است. در این راستا، الگوریتم XGBoost به‌عنوان یکی از پرکاربردترین و مؤثرترین الگوریتم‌ها شناخته می‌شود. این الگوریتم به‌ویژه در حل مسائل رگرسیون، طبقه‌بندی و رتبه‌بندی مورداستفاده قرار می‌گیرد و نسخه‌ای پیشرفته از روش تقویت گرادیان (Gradient Boosting) است. XGBoost به دلیل ویژگی‌هایی همچون دقت بالا و سرعت محاسباتی، در سال‌های اخیر به یکی از محبوب‌ترین ابزارها درزمینه یادگیری ماشین تبدیل‌شده است (Ali et al., 2023). XGBoost با استفاده از مجموعه‌ای از درخت‌های تصمیم (Decision Trees) عمل می‌کند. هر درخت به‌منظور اصلاح خطاهای درخت قبلی ساخته می‌شود و این فرآیند به‌صورت تدریجی ادامه پیدا می‌کند تا زمانی که مدل به عملکرد مطلوبی برسد. الگوریتم از روش بهینه‌سازی گرادیان نزولی (Gradient Descent) برای کمینه کردن خطا استفاده می‌کند که بسته به نوع مسئله، این خطا ممکن است میانگین مربعات خطا (MSE) یا لگاریتم خطا (Log Loss) باشد (Bentéjac et al., 2021). یکی از ویژگی‌های کلیدی XGBoost، استفاده از تکنیک‌های منظم‌سازی است که برای جلوگیری از بیش برازش (Overfitting) طراحی‌شده‌اند. این ویژگی، با افزودن جریمه‌ای به پیچیدگی مدل، به بهبود تعمیم‌پذیری آن کمک می‌کند. همچنین، فرآیند هرس درخت‌ها (Tree Pruning) در این الگوریتم به حذف شاخه‌های غیرضروری کمک می‌کند که نتیجه آن‌یک مدل ساده‌تر و کارآمدتر است. علاوه بر این، XGBoost از پردازش موازی (Parallel Processing) بهره می‌برد که با تقسیم‌کار بین پردازنده‌ها، سرعت محاسبات را به‌طور قابل‌توجهی افزایش می‌دهد (Asselman

XGBoost (et al., 2023). در پروژه‌های مختلف یادگیری ماشین کاربردهای گسترده‌ای دارد. این الگوریتم به‌ویژه در حوزه‌هایی همچون تشخیص تصاویر، طبقه‌بندی متون و پردازش زبان طبیعی به کار گرفته می‌شود و نتایج چشمگیری در این زمینه‌ها ارائه داده است (Bentéjac et al., 2021).

نتایج و یافته‌های تحقیق

در این قسمت ابتدا با استفاده از روش DEA کارایی شرکت‌های بورسی موردبررسی قرارگرفته و سپس با استفاده از روش‌های یادگیری ماشین به بررسی متغیرهای حیاتی در پیش‌بینی عملکرد شرکت‌ها پرداخته می‌شود.

محاسبه کارایی شرکت‌های بورسی با استفاده از DEA

در مرحله نخست با استفاده از روش DEA نمرات کارایی شرکت‌های بورسی از سال ۱۳۸۶ تا ۱۴۰۲(۱۳۰ شرکت در هرسال مالی) مورد تجزیه‌وتحلیل قرار می‌گیرد. اطلاعات مربوط به این شرکت‌ها از نرم افزار ره آورد نوین و محاسبات مویسندگان جمع‌آوری‌شده است. ازآنجایی‌که روش DEA به انتخاب ورودی و خروجی بسیار حساس می‌باشد و توافق کلی در این زمینه جهت انتخاب این متغیرها وجود ندارد، در این مطالعه با بررسی ادبیات مربوطه و مصاحبه با کارشناسان دو متغیر جمع بدهی و RETRSIK به‌عنوان متغیر ورودی و بازده سالیانه سهام و نقد شوندگی به‌عنوان خروجی انتخاب می‌شوند. در جدول شماره (۱) خلاصه آماری شرکت‌ها بیان‌شده است.

جدول ۱

خلاصه آماری ورودی‌ها و خروجی‌های تحلیل پوششی داده

خروجی‌ها		ورودی‌ها		شاخص‌های آماری
نقد شوندگی	بازده سالیانه سهام	RetRisk	جمع بدهی	
۰.۳	۰.۱	۰.۱۰۰۳	۱۲۷۹۴۴	حداقل
۴.۰۵۵۷	۲۵.۱۶	۰.۵۰۴۴	۱۸۰۲۵۸۶۶۴۶	حداکثر
۰.۰۸۴۸	۱.۷۰۶۷	۰.۰۳۶۳	۷۷۵۳۹۴۳۶	انحراف معیار
۰.۱۰۳	۱.۶۴۹	۰.۱۳۶۴	۱۱۶۳۵۸۵۳	میانگین

همچنین در جدول شماره (۲) میانگین کارایی شرکت‌های بورسی ذکرشده است:

جدول ۲

میانگین کارایی ۱۳۰ شرکت بورسی با استفاده از DEA

میانگین کارایی ۱۳۰ شرکت	سال
۰.۱۸۱۷	۱۳۸۶
۰.۱۵۸۰	۱۳۸۷
۰.۲۱۵۵	۱۳۸۸
۰.۲۱۳۲	۱۳۸۹
۰.۱۸۷۴	۱۳۹۰
۰.۱۹۲۱	۱۳۹۱
۰.۲۷۱۱	۱۳۹۲
۰.۱۶۱۵	۱۳۹۳

۰.۱۸۸۹	۱۳۹۴
۰.۱۸۵۵	۱۳۹۵
۰.۱۷۱۷	۱۳۹۶
۰.۲۲۷۴	۱۳۹۷
۰.۴۴۰۲	۱۳۹۸
۰.۲۱۷۱	۱۳۹۹
۰.۱۲۶۸	۱۴۰۰
۰.۲۳۹۰	۱۴۰۱
۰.۲۳۹۰	۱۴۰۲
۰.۲۱۲۷	مجموع سال‌ها

با توجه به جدول (۲)، نمرات کارایی ۱۳۰ شرکت بورسی ایران نوسانات قابل‌توجهی را نشان می‌دهد. بالاترین نمره میانگین کارایی مربوط به سال ۱۳۹۸ با مقدار ۰.۴۴۰۳ و کمترین آن مربوط به سال ۱۴۰۰ با مقدار ۰.۱۲۶۹ است. در بازه سال‌های ۱۳۸۶ تا ۱۳۹۱ میانگین کارایی نسبتاً باثبات و در محدوده ۰.۱۵۸۰ تا ۰.۲۱۵۶ بوده است، اما در سال ۱۳۹۲ با افزایش به ۰.۲۷۱۲ جهش قابل‌توجهی تجربه کرد. از سال ۱۳۹۳ تا ۱۳۹۶ کارایی روند کاهشی داشته و به ۰.۱۶۱۶ در سال ۱۳۹۳ و ۰.۱۷۱۷ در سال ۱۳۹۶ رسیده است. پس از آن، در سال ۱۳۹۸ شاهد بهبود چشمگیر به ۰.۴۴۰۳ بودیم، اما این بهبود موقتی بود و از سال ۱۳۹۹ تا ۱۴۰۰ کاهش شدیدی رخ داد و به ۰.۱۲۶۹ در سال ۱۴۰۰ رسید. در سال‌های ۱۴۰۱ و ۱۴۰۲ نیز کارایی با مقدار ۰.۲۳۹۰ تثبیت شده است. این نوسانات می‌تواند تحت تأثیر عوامل مختلفی از جمله نوسانات ارزی، تورم، و شرایط اقتصادی کلان کشور باشد.

رویکرد ترکیبی تحلیل پوششی داده و یادگیری ماشین

در این مرحله ابتدا باید ورودی ها و خروجی‌های مدل‌های یادگیری ماشین مشخص شوند. در جدول شماره (۳) معیارهای به‌کاررفته به همراه تعاریف عملیاتی آن‌ها ذکرشده است.

جدول ۳

متغیرهای به کار رفته در مدل‌های یادگیری ماشین

متغیرهای به کار رفته	تعاریف عملیاتی	نقش متغیر در مدل‌های یادگیری ماشین
جمع دارایی‌ها	منابع مولد ایجاد جریان‌های نقدینگی در سال جاری	ورودی
جمع دارایی‌های سال گذشته	منابع مولد ایجاد جریان‌های نقدینگی در سال گذشته	ورودی
جمع بدهی	تعهدات مالی شرکت به افراد خارج از شرکت	ورودی
سرمایه	منابع اولیه تامین شده شرکت	ورودی
حقوق صاحبان سهام	جمع منابع تامین شده توسط صاحبان سرمایه	ورودی
فروش	میزان کالا و خدمات ارائه شده به مشتریان در سال جاری	ورودی
فروش سال گذشته	میزان کالا و خدمات ارائه شده به مشتریان در سال گذشته	ورودی
سود عملیاتی	سود فعالیت‌های مستمر	ورودی
هزینه بهره	مخارج تامین مالی خارج از شرکت	ورودی
مالیات	سهم دولت از سود کسب شده شرکت‌ها	ورودی
سود خالص	هزینه‌ها - جمع درآمد	ورودی

بازده سالیانه سهام	مجموع حقوق و مزایای تعلق گرفته به جمع سهام	ورودی
ارزش بازار سهام	تعداد سهام × قیمت سهام	ورودی
حجم معاملات سهام	میزان سهام معامله شده	ورودی
تعداد سهام شرکت	تعداد برگه سهام‌های منتشر شده	ورودی
ROA	نسبت سود خالص به جمع دارایی‌های شرکت	ورودی
ROE	نسبت سود خالص به جمع حقوق صاحبان سهام	ورودی
MTB	نسبت ارزش بازار شرکت (ارزش بازار سهام به‌علاوه ارزش دفتری بدهی‌ها) به ارزش دفتری دارایی‌های شرکت	ورودی
GROWTH	درصد تغییرات درفروش سال جاری (t) نسبت به سال ماقبل آن(t-۱)	ورودی
ILLIQ	اندازه‌گیری میزان نقد شوندگی سهام از معیار آمیهود مبتنی بر معیار معامله	ورودی
RETRISK	ریسک بازده سهام با استفاده از انحراف معیار مقادیر باقیمانده مدل سه عاملی فاما و فرنچ	ورودی
SIZE	لگاریتم طبیعی جمع دارایی‌های شرکت	ورودی
LEV	نسبت بدهی به جمع دارایی‌ها	ورودی
WACC	میانگین موزون هزینه سرمایه	ورودی
VOL	نسبت تعداد سهام معامله‌شده شرکت طی سال مالی به جمع کل تعداد سهام شرکت	ورودی
RETURN	مزایایی که در طول دوره موردنظر به سهم تعلق می‌گیرد نسبت به قیمت ابتدای دوره	ورودی
FVALUE	لگاریتم ارزش بازار شرکت	ورودی
EVA	ارزش خلق‌شده پس از هزینه‌ها	ورودی
نمرات کارایی تحلیل پوششی داده	----------	هدف (تارگت)

آماده‌سازی اطلاعات

در این مرحله قبل از اجرای مدل‌های یادگیری ماشین کیفیت داده‌های تحقیق موردبررسی قرار گرفت. برای انجام این کار، مراحلی ازجمله ارزیابی کیفیت داده‌ها، مدیریت مقادیر ازدست‌رفته و بررسی داده‌های پرت انجام شد. مرحله اولیه تجزیه‌وتحلیل داده‌های ازدست‌رفته و مقادیر پرت بود. در این تحقیق هیچ مقدار ازدست‌رفته‌ای وجود نداشت و جهت مدیریت مقادیر پرت از روش IQR[1] استفاده‌شده است. در این روش مقادیری که خارج از محدوده ۱.۵ برابر فاصله بین چارک اول و سوم قرار داشتند، به‌عنوان داده پرت شناسایی شدند. در مرحله بعدی، تمام متغیرها را با استفاده از رویکرد حداقل حداکثر و فرمول شماره (۴) استاندارد کردیم تا اختلاف مقیاس بین متغیرها را کاهش یابد.

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \qquad (4)$$

آموزش و اعتبارسنجی مدل‌ها

در این مرحله داده‌ها جهت آموزش و اعتبار سنجی به دو مجموعه آموزش و آزمایش تقسیم شدند. در این تقسیم، ۸۰٪ برای آموزش و ۲۰٪ برای آزمایش اختصاص داده شد که مدل‌های یادگیری ماشین را قادر می‌سازد تا به‌طور مؤثر همبستگی بین ویژگی‌های ورودی و هدف

[1] Interquartile Range

(نمرات کارایی) را بررسی کنند. پس از آموزش و آزمایش جهت ارزیابی مدل‌های یادگیری ماشین از سه معیار ارزیابی شامل MAE،MSE2[1] و RMSE[3] استفاده‌شده است.(کلیه تحلیل‌ها با استفاده از نرم افزار python3.1.1 انجام شده است)

MAE میانگین قدر مطلق اختلاف بین مقادیر واقعی y_i و پیش‌بینی‌شده $\hat{y_i}$ را بررسی می‌کند. مقدار کمتر MAE نشان‌دهنده دقت بالای مدل می‌باشد.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y_i}\right| \qquad (5)$$

در مدل شماره (۵) n تعداد حجم نمونه است.

MSE میانگین مربعات اختلاف بین مقادیر واقعی و پیش‌بینی‌شده را محاسبه می‌کند. رابطه MSE برابر است با:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \qquad (6)$$

همچنین معیار RMSE برابر باریشه مربع MSE می‌باشد. این معیار حساسیت بیشتری به مقادیر خطای بزرگ دارد و مقادیر کمتر RMSE به معنای عملکرد بهتر مدل است.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})} \qquad (7)$$

جدول ۴

ارزیابی مدل‌های یادگیری ماشین

معیارهای ارزیابی	Xgbbost	Svm	DT	RF
MAE	۰.۰۱۸۹	۰.۱۰۳۵	۰.۰۳۱۹	۰.۰۱۹۹
MSE	۰.۰۰۲۶	۰.۰۱۵۰	۰.۰۰۳۸	۰.۰۰۲۹
RMSE	۰.۰۵۱۴	۰.۱۲۲۶	۰.۰۶۱۸	۰.۰۵۴۳

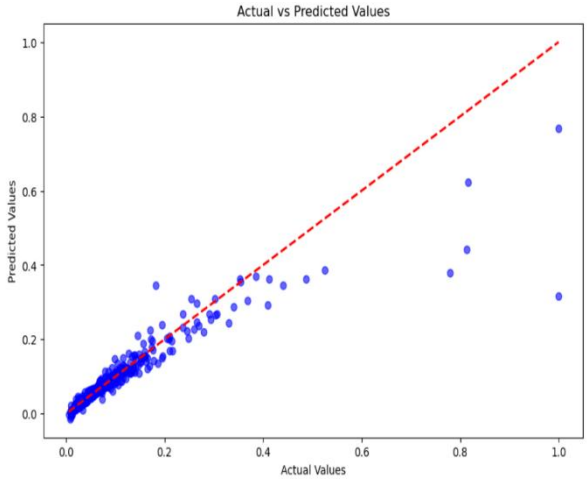
با توجه به جدول(۴) الگوریتم XGBoost بهترین عملکرد را در پیش‌بینی کارایی شرکت‌ها دارد. این الگوریتم با MAE= 0.0189، MSE= 0.0026 و RMSE =0.0514 نسبت به سایر مدل‌ها، دقت بالاتری را نشان داده است. دقت بالای این الگوریتم ناشی از تکنیک‌های پردازش موازی و بهینه‌سازی گرادیان نزولی است که به کاهش خطا و بهبود تعمیم‌پذیری مدل کمک می‌کند. در رتبه دوم الگوریتم RF پس ازXGBoost قرار می‌گیرد و نشان می‌دهد که ترکیب درخت‌های تصمیم و انتخاب تصادفی ویژگی‌ها به عملکرد قابل قبول این مدل منجر شده است. الگوریتمDT نیز با وجود سادگی و شفافیت ساختاری، دقت کمتری داشته نسبت به Xgboost و RF داشته است و معیارهای ارزیابی آن نسبت به سایر دو الگوریتم دیگر نشان‌دهنده ضعف در شناسایی روابط پیچیده بین داده‌ها است. در نهایت، الگوریتم SVM ضعیف‌ترین عملکرد را به ثبت رسانده است که این نتیجه ناشی از تنظیمات نادرست کرنل یا محدودیت در مدیریت داده‌های پیچیده باشد. همچنین نمودارهای پراکندگی جهت مقایسه مقادیر واقعی و پیش‌بینی‌شده توسط مدل استفاده می‌شود. نقاط نزدیک به خط قرمز نشان‌دهنده

[1] Mean Absolute Error
[2] Mean Squared Error
[3] Root Mean Squared Error

دقت بالای پیش‌بینی مدل هستند. فاصله نقاط از این خط بیانگر میزان خطای مدل است؛ هرچه نقاط به خط نزدیک‌تر باشند، مدل عملکرد بهتری داشته است. همچنین با توجه به شکل‌های (۱) الی (۴) الگوریتم xgboost عملکرد بهتری نسبت به ۳ مدل دیگر دارد. نقاط در نمودار پراکندگی این الگوریتم بسیار نزدیک به خط مرجع (خط قرمز) قرار گرفته‌اند که نشان‌دهنده دقت بالای پیش‌بینی و همبستگی قوی بین مقادیر واقعی و پیش‌بینی‌شده است. در مقایسه، نمودار مربوط به RF نیز نقاطی نزدیک به خط مرجع دارد، اما پراکندگی بیشتری نسبت به xgboostمشاهده می‌شود که نشان‌دهنده عملکرد ضعیف تر این مدل است. در نمودارDT نقاط پراکندگی بیشتری از خط مرجع فاصله دارند که بیانگر خطای بالاتر و توانایی کمتر این مدل در پیش‌بینی مقادیر واقعی است. در نهایت، نمودار SVM بیشترین فاصله نقاط از خط مرجع را نشان می‌دهد که بیانگر خطاهای بالا و دقت پایین این الگوریتم در پیش‌بینی کارایی شرکت‌ها است.

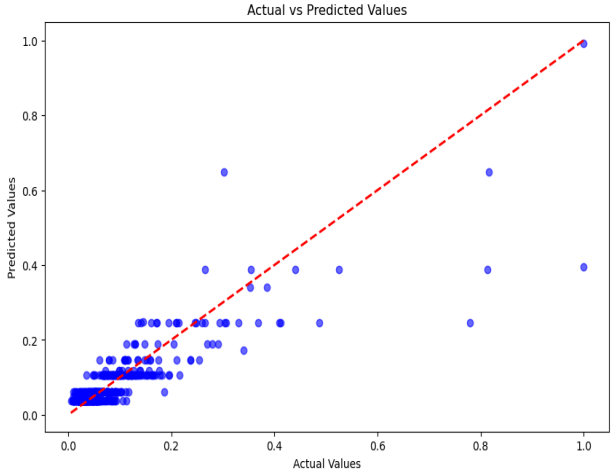
شکل ۱

نمودار مقایسه مقادیر واقعی و پیش‌بینی‌شده در روش Xgboost



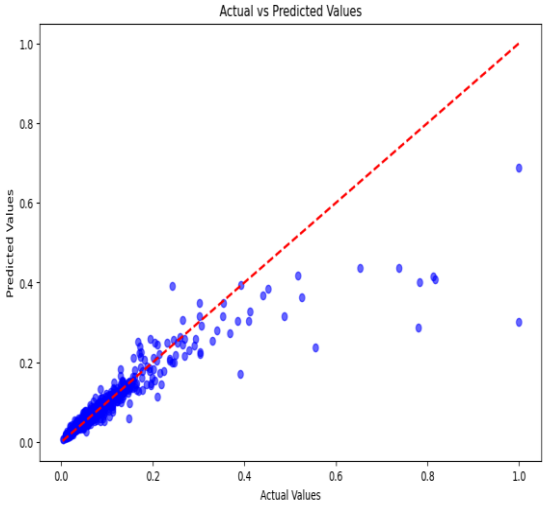
شکل ۲

نمودار مقایسه مقادیر واقعی و پیش‌بینی‌شده در روش DT



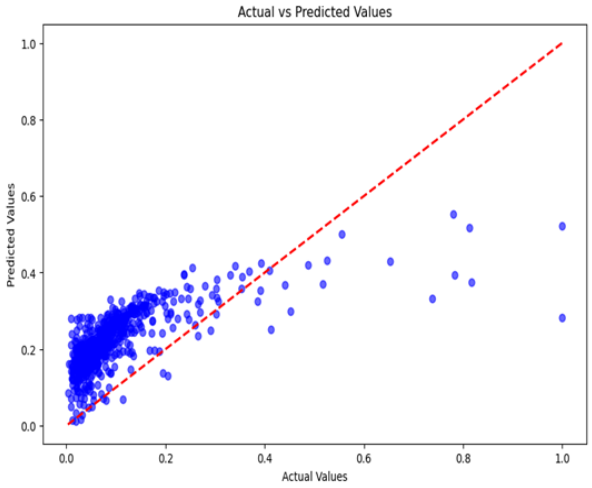
شکل ۳

نمودار مقایسه مقادیر واقعی و پیش‌بینی‌شده در روش Random forest



شکل ۴

نمودار مقایسه مقادیر واقعی و پیش‌بینی‌شده در روش SVM



در جدول شماره (۵) درصد اهمیت متغیرهای مختلف جهت پیش‌بینی کارایی شرکت‌های بورسی ایران نمایش داده شده است. درصدهای ذکرشده در این جدول، میزان تأثیرگذاری هر متغیر را در تعیین نمره نهایی کارایی مشخص می‌کنند. هدف از ارائه این جدول، شناسایی ویژگی‌های حیاتی و مهم است که نقش کلیدی در پیش‌بینی عملکرد شرکت‌ها ایفا می‌کنند و می‌توانند به بهبود تحلیل و تصمیم‌گیری در این حوزه کمک کنند.

جدول ۵

اهمیت متغییرهای مدل یادگیری ماشین

متغیرها	درصد اهمیت متغیرها
جمع دارایی‌های سال گذشته	۵۱۶.۶۵۵
جمع بدهی	۱۴.۸۸۴۰
RETRISK	۷.۳۶۱۷۰
بازده سالیانه سهام	۵.۶۳۳۹
سود خالص	۵.۲۵۷۴
WACC	۴.۲۹۷۰
MTB	۴.۲۰۰۱
هزینه بهره	۴.۱۹۵۴
جمع دارایی‌ها	۳.۶۹۷۵
VOL	۳.۲۸۳۳
LEV	۳.۰۲۵۸۷
سود عملیاتی	۲.۹۹۵۳
GROWTH	۲.۷۸۰۴
فروش	۲.۷۵۵۵
ROA	۲.۶۳۷۶
EVA	۲.۴۴۶۱
حجم معاملات سهام	۲.۳۳۸۹
ROE	۲.۰۴۷۹
ILLIQ	۲.۰۰۳۳
مالیات	۱.۸۵۸۰
سرمایه	۱.۴۷۶۶۹
حقوق صاحبان سهام	۱.۴۴۰۵
ارزش بازار سهام	۱.۰۹۵۲
FVALUE	۰.۹۲۹۶
فروش سال گذشته	۰.۷۰۲۲
SIZE	۰
RETURN	۰
تعداد سهام شرکت	۰

طبق جدول شماره (۵)، جمع دارایی‌های سال گذشته با اهمیت ۱۶.۶۵ درصد، بیشترین تأثیر را در پیش‌بینی کارایی شرکت‌ها داشته است. این امر نشان‌دهنده تأثیر مهم و مستقیم وضعیت مالی شرکت‌ها در سال گذشته بر عملکرد فعلی آن‌ها می‌باشد و بر اهمیت مدیریت دارایی‌ها برای بهبود کارایی تأکید ویژه دارد. متغیر جمع بدهی با اهمیت ۱۴.۸۸ درصد نیز در رتبه دوم قرار دارد که نشان می‌دهد میزان بدهی‌ها نقش کلیدی در توانایی شرکت برای ایجاد بازدهی و مدیریت منابع دارد. همچنین، متغیر ریسک بازدهی (RETRISK) با اهمیت ۷.۳۶ درصد، بیانگر آن است که شرکت‌هایی با نوسانات کمتر در بازدهی، عملکرد بهتر و پایدارتری دارند. علاوه بر این، متغیرهای بازده سالیانه سهام (۵.۶۳ درصد) و سود خالص (۵.۲۵ درصد) نقش مهمی در تحلیل عملکرد مالی شرکت‌ها دارند و نشان می‌دهند که بازدهی سرمایه و سودآوری، از عوامل کلیدی در کارایی محسوب می‌شوند. از سوی دیگر، متغیرهایی مانند هزینه بهره و WACC که اهمیت کمتری دارند

(حدود ۴ درصد)، به تأثیر هزینه‌های سرمایه‌گذاری و تأمین مالی در عملکرد شرکت‌ها اشاره می‌کنند.در مقابل، متغیرهایی مانند ILLIQ وROE با اهمیت حدود ۲ درصد، تأثیر کمتری در پیش‌بینی کارایی داشته‌اند، اما همچنان می‌توانند در مدل‌های ترکیبی برای بهبود نتایج مؤثر باشند. در نهایت، متغیرهایی نظیر SIZE، RETURN و تعداد سهام شرکت با اهمیت صفر، در این تحلیل خاص تأثیر معناداری نداشته‌اند و می‌توان از آن‌ها در این مدل صرف‌نظر کرد. این نتایج به شناسایی نقاط کلیدی برای بهبود عملکرد شرکت‌ها و تمرکز بر عوامل تأثیرگذار کمک می‌کند.

نتیجه‌گیری

در این پژوهش، کارایی ۱۳۰ شرکت بورسی ایران طی سال‌های ۱۳۸۶ تا ۱۴۰۲ با استفاده از روش DEA و الگوریتم‌های یادگیری ماشین مورد بررسی قرار گرفت. ابتدا، نمرات کارایی شرکت‌ها بر اساس متغیرهای ورودی و خروجی مشخص محاسبه شد و سپس، الگوریتم‌های یادگیری ماشین برای شناسایی متغیرهای حیاتی و ارزیابی عملکرد شرکت‌ها استفاده شدند. نتایج نشان داد که الگوریتم XGBoost بهترین عملکرد را در پیش‌بینی داشته است. تحلیل نتایج نشان داد که ترکیب روش DEA با الگوریتم‌های یادگیری ماشین می‌تواند ابزاری موثر برای تحلیل و پیش‌بینی عملکرد شرکت‌های بورسی باشد. این یافته‌ها نشان داد که استفاده از الگوریتم‌های پیشرفته می‌تواند دقت و سرعت تحلیل را به طور چشمگیری بهبود بخشد. طبق تحلیل نتایج الگوریتم‌های یادگیری ماشین جمع دارایی‌های سال گذشته و جمع بدهی به عنوان اثر گذار ترین متغیرها بر نمرات کارایی شناخته شدند و همچنین سه متغیر SIZE، RETURN و تعداد سهام شرکت اثری بر نمرات کارایی ندارند. یکی از مزایای اصلی این مطالعه، انتخاب و ترکیب متغیرهای کلیدی در چارچوب DEA و یادگیری ماشین بوده است که دقت پیش‌بینی را بهبود بخشیده است. علاوه بر این، استفاده از مدل‌های مختلف یادگیری ماشین برای مقایسه عملکرد آن‌ها و تعیین بهترین الگوریتم، نشان‌دهنده ارتقای رویکردهای موجود در ارزیابی کارایی شرکت‌ها است. مدیران شرکت‌های بورسی می‌توانند با استفاده از نتایج این پژوهش، نقاط ضعف و قوت عملکرد مالی خود را شناسایی کرده و استراتژی‌های بهبود را تدوین کنند.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافعی وجود ندارد.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

موازین اخلاقی

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

شفافیت داده‌ها

داده‌ها و مآخذ پژوهش حاضر در صورت درخواست از نویسنده مسئول و ضمن رعایت اصول کپی رایت ارسال خواهد شد.

حامی مالی

این پژوهش حامی مالی نداشته است.

References

Adabi, G. B., Moghaddam, S. R. H., Fouladi, A., & Samanta, N. (2017). An overview of corporate governance practices in Iran: a quantitative survey of listed companies in Tehran Stock Exchange. *International Company and Commercial Law Review*, *28*(2), 33-48.

Ahsaan, S. U., Kaur, H., Mourya, A. K., & Naaz, S. (2022). A hybrid support vector machine algorithm for big data heterogeneity using machine learning. *Symmetry*, *14*(11), 2344.

Alarussi, A. S. A. (2021). Financial ratios and efficiency in Malaysian listed companies. *Asian Journal of Economics and Banking*, *5*(2), 116-135.

Ali, Z. A., Abduljabbar, Z. H., Taher, H. A., Sallow, A. B., & Almufti, S. M. (2023). Exploring the power of eXtreme gradient boosting algorithm in machine learning: A review. *Academic Journal of Nawroz University*, *12*(2), 320-334.

Arsad, R., Isa, Z., Abidin, N., & Shaari, S. (2022). Stock Selection using the Data Envelopment Analysis Models and Dupont Analysis. *International Journal of Academic Research in Business and Social Sciences*, *12*(8), 1770-1787.

Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, *31*(6), 3360-3379.

Avramov, D., Cheng, S., & Metzker, L. (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science*, *69*(5), 2587-2619.

Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decision Analytics Journal*, *3*, 100071.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, 1937-1967.

Birjandi, A. K., Dehmolaee, S., Sheikh, R., & Sana, S. S. (2021). Analysis and classification of companies on tehran stock exchange with incomplete information. *RAIRO-Operations Research*, *55*, S2709-S2726.

Bogdanova, B., & Stancheva-Todorova, E. (2021). ML-based predictive modelling of stock market returns. AIP Conference Proceedings,

Bouasabah, M. (2024). A Performance Analysis of Machine Learning Algorithms in Stock Market Prediction, Compared to Traditional Indicators. International Conference on Digital Age & Technological Advances for Sustainable Development,

Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual review of fluid mechanics*, *52*(1), 477-508.

Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, *70*, 245-317.

Chandra, M. A., & Bedi, S. (2021). Survey on SVM and their application in image classification. *International Journal of Information Technology*, *13*(5), 1-11.

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20-28.

Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, *56*(5), 4765-4800.

Fallah, M., Hosseinzadeh Lotfi, F., & Hosseinzadeh, M. M. (2020). Discriminant analysis and data envelopment analysis with specific data and Its application for companies in the Iranian stock exchange. *Iranian Journal of Operations Research*, *11*(1), 144-156.

Fotova Čiković, K., Martinčević, I., & Lozić, J. (2022). Application of data envelopment analysis (DEA) in the selection of sustainable suppliers: A review and bibliometric analysis. *Sustainability*, *14*(11), 6672.

Hegde, J., & Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment–A review. *Safety science*, *122*, 104492.

Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in Bioinformatics*, *24*(2), bbad002.

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer science review*, *41*, 100413.

Jiang, W. (2021). Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications*, *184*, 115537.

Keshavarz, H., & Rezaei, M. (2022). The Effect of Economic, Financial and Political Risk on the Risk and Return of Tehran Stock Exchange. *Monetary & Financial Economics*, *28*(22), 127-152.

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 1-24.

Krmac, E., & Mansouri Kaleibar, M. (2023). A comprehensive review of data envelopment analysis (DEA) methodology in port efficiency evaluation. *Maritime Economics & Logistics*, *25*(4), 817-881.

Kumar, D., Sarangi, P. K., & Verma, R. (2022). A systematic review of stock market prediction using machine learning and statistical techniques. *Materials Today: Proceedings*, *49*, 3187-3191.

Li, D., Hou, R., & Sun, Q. (2020). The business performance evaluation index method for the high-tech enterprises based on the DEA model. *Journal of Intelligent & Fuzzy Systems*, *38*(6), 6853-6861.

Mehtab, S., & Sen, J. (2020). A time series analysis-based stock price prediction using machine learning and deep learning models. *International Journal of Business Forecasting and Marketing Intelligence*, *6*(4), 272-335.

Narkunienė, J., & Ulbinaitė, A. (2018). Comparative analysis of company performance evaluation methods. *Entrepreneurship and sustainability issues*, *6*(1), 125-138.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, *53*(4), 3007-3057.

Otchere, D. A., Ganat, T. O. A., Gholami, R., & Ridha, S. (2021). Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *Journal of Petroleum Science and Engineering*, *200*, 108182.

Panwar, A., Olfati, M., Pant, M., & Snasel, V. (2022). A review on the 40 years of existence of data envelopment analysis models: historic development and current trends. *Archives of Computational Methods in Engineering*, *29*(7), 5397-5426.

Rahimi, H., Minouei, M., & Fathi, M. (2022a). Financial distress of companies listed on the Tehran stock exchange using the dynamic worst practice frontier-based DEA model. *Advances in mathematical finance and applications*, *7*(2), 507-525.

Rahimi, H., Minouei, M., & Fathi, M. R. (2022b). Predicting the Financial Distress of Companies Listed on the Tehran Stock Exchange Using DEA-DA Technique and Artificial Neural Network. *Karafan Journal*, *19*(2), 593-621.

Ravanshad, M. R., Amiri, A., Salari, H., & Khodadadi, D. (2020). Application of the two-stage DEA model for evaluating the efficiency and investigating the relationship between managerial ability and firm performance. *Advances in Mathematical Finance and Applications*, *5*(2), 229-245.

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H.-C. (2021). Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics*, *10*(21), 2717.

Ruiz, J. L., & Sirvent, I. (2022). Benchmarking within a DEA framework: setting the closest targets and identifying peer groups with the most similar performances. *International Transactions in Operational Research*, *29*(1), 554-573.

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, *2024*, 69-79.

Sha'Abani, M., Fuad, N., Jamal, N., & Ismail, M. (2020). kNN and SVM classification for EEG: a review. InECCE2019: Proceedings of the 5th International Conference on Electrical, Control & Computer Engineering, Kuantan, Pahang, Malaysia, 29th July 2019,

Shahin, M., Chen, F. F., Hosseinzadeh, A., & Zand, N. (2023). Using machine learning and deep learning algorithms for downtime minimization in manufacturing systems: An early failure detection diagnostic service. *The International Journal of Advanced Manufacturing Technology*, *128*(9-10), 3857-3883.

Shahriari, M. (2023). Ranking and evaluation of financial efficiency of pharmaceutical companies accepted in Tehran Stock Exchange with the approach of data envelopment analysis and multi-criteria decision making. *Advances in Mathematical Finance and Applications*, *1*(1), 219.

Talekar, B., & Agrawal, S. (2020). A detailed review on decision tree and random forest. *Biosci. Biotechnol. Res. Commun*, *13*(14), 245-248.

Tangirala, S. (2020). Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. *International Journal of Advanced Computer Science and Applications*, *11*(2), 612-619.

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, *167*, 599-606.

Zaman, N., Ghazanfar, M. A., Anwar, M., Lee, S. W., Qazi, N., Karimi, A., & Javed, A. (2023). Stock market prediction based on machine learning and social sentiment analysis. *Authorea Preprints*.

Zhang, Z., Xiao, Y., Fu, Z., Zhong, K., & Niu, H. (2022). A study on early warnings of financial crisis of Chinese listed companies based on DEA–SVM model. *Mathematics*, *10*(12), 2142.

Zhou, H., Zhang, J., Zhou, Y., Guo, X., & Ma, Y. (2021). A feature selection algorithm of decision tree based on feature weight. *Expert Systems with Applications*, *164*, 113842.