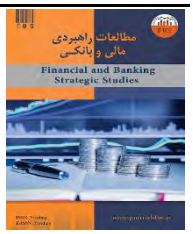
E-ISSN: 2981-1589 | P-ISSN: 2981-1570

Financial and Banking Strategic Studies

www.journal-fbs.com

Fin. Bank. Stra. Stud. Vol. 3, No. 1 (2025) 41–55.



Paper Type: Original Article

Portfolio Optimization Using Data Envelopment Analysis Integration with Multiple Data Sources with a Machine Learning Approach in Tehran Stock Exchange

Meysam Kaviani[1,*], Masumeh Jafari[1], Morteza Kaviani[2], Kaveh Kaviani[1]

[1] *Department of Management and Accounting, Karaj Branch, Islamic Azad University, Karaj, Iran; meysamkaviani@gmail.com; m.jafari8269@gmail.com; kaveh.k79@gmail.com.*

[2] *Department of Accounting, Shafaq Non-Profit Higher Education Institute, Tonekabon, Iran; morteza.kaviani41@yahoo.com.*

Citation:

Kaviani, M., Jafari, M., Kaviani, M., & Kaviani, K. (2025). Portfolio optimization using data envelopment analysis integration with multiple data sources with a machine learning approach in Tehran Stock Exchange. *Financial and banking strategic studies*, *3*(1), 41-55.

Received: 24/07/2024 Reviewed: 26/09/2024 Revised: 04/11/2024 Accepted: 03/01/2025

Abstract

Purpose: This study aims to optimize investment portfolios by combining Data Envelopment Analysis (DEA) and machine learning, using multi-source data from the Tehran Stock Exchange. The main goal is to provide an advanced model for stock selection and portfolio optimization so that investors can adopt strategies that are more efficient than the traditional methods.

Methodology: First, the DEA model was used to evaluate stock efficiency in terms of historical returns and asset correlation. Then, using a Support Vector Machine (SVM) and combining multi-source data, the stock price movement trend is predicted. To improve the accuracy of the model, random search and network methods were used to adjust the algorithm hyperparameters optimally. Finally, the resulting data are integrated into a portfolio optimization model and a proposed investment strategy is formulated.

Findings: Experimental results on the Tehran Stock Exchange data showed that the proposed model can improve the performance of investment strategies compared to traditional methods. Sharpe and Sortino ratios indicate the superiority of the proposed model over the minimum global variance strategy. It was also found that using a low-diversity investment strategy can provide greater efficiency than using fully diversified strategies.

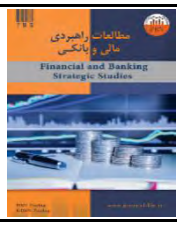
Originality/Value: This study proposes a new approach to stock selection and portfolio optimization by presenting a combined model of DEA and machine learning. The use of multi-source data and advanced machine learning methods improves the accuracy of investor forecasting and decision-making, paving the way for future research in areas such as the use of fuzzy models, meta-heuristic algorithms, and the analysis of relationships between financial indicators.

Keywords: Data envelopment analysis, Portfolio optimization, Machine learning, Multiple data sources.

Corresponding Author: meysamkaviani@gmail.com 10.22105/fbs.2025.511565.1154

Licensee. **Financial and Banking Strategic Studies**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

مطالعات راهبردی مالی و بانکی
دوره ۳، شماره (۱)، (۱۴۰۴)، ۵۵-۴۱
www.journal-fbs.com



نوع مقاله: پژوهشی

بهینه‌سازی پرتفوی با استفاده از ادغام تحلیل پوششی داده‌ها با منابع داده چندگانه با رویکرد یادگیری ماشین در بورس اوراق بهادار تهران

میثم کاویانی[۱،*]، معصومه جعفری[۱]، مرتضی کاویانی[۲]، کاوه کاویانی[۱]

[۱]گروه مالی، دانشکده مدیریت و حسابداری، واحد کرج، دانشگاه آزاد اسلامی، کرج، ایران.

[۲]گروه حسابداری، موسسه آموزش عالی و غیرانتفاعی شفق، تنکابن، ایران.

چکیده

هدف: این پژوهش با هدف بهینه‌سازی پرتفوی سرمایه‌گذاری از طریق ترکیب تحلیل پوششی داده‌ها و یادگیری ماشین با استفاده از داده‌های چندمنبعی در بورس اوراق بهادار تهران انجام شده است. هدف اصلی ارایه یک مدل پیشرفته برای انتخاب سهام و بهینه‌سازی پرتفوی به‌گونه‌ای است که سرمایه‌گذاران بتوانند استراتژی‌های کارآمدتری را نسبت به روش‌های سنتی اتخاذ کنند.

روش‌شناسی پژوهش: در این مطالعه، ابتدا از مدل تحلیل پوششی داده‌ها برای ارزیابی کارایی سهام از نظر بازده تاریخی و همبستگی دارایی استفاده شده است. سپس با استفاده از ماشین بردار پشتیبان و ترکیب داده‌های چندمنبعی، روند حرکت قیمت سهام پیش‌بینی شده است. به‌منظور بهبود دقت مدل، روش‌های جستجوی تصادفی و شبکه‌ای برای تنظیم بهینه هایپرپارامترهای الگوریتم به کار گرفته شده است. در نهایت، داده‌های حاصل در یک مدل بهینه‌سازی پرتفوی ادغام شده و استراتژی سرمایه‌گذاری پیشنهادی تدوین شده است .

یافته‌ها: نتایج تجربی بر روی داده‌های بورس اوراق بهادار تهران نشان داد که مدل پیشنهادی قادر به بهبود عملکرد استراتژی‌های سرمایه‌گذاری در مقایسه با روش‌های سنتی است. نسبت‌های شارپ و سورتینو نشان‌دهنده برتری مدل پیشنهادی نسبت به استراتژی حداقل واریانس سراسری بوده‌اند. همچنین مشخص شد که استفاده از یک استراتژی سرمایه‌گذاری کم‌تنوع می‌تواند کارایی بیشتری نسبت به استراتژی‌های کاملا متنوع ارایه دهد.

اصالت/ارزش‌افزوده علمی: این پژوهش با ارایه یک مدل ترکیبی از تحلیل پوششی داده و یادگیری ماشین، رویکردی نوین برای انتخاب سهام و بهینه‌سازی پرتفوی پیشنهاد می‌دهد. استفاده از داده‌های چندمنبعی و روش‌های پیشرفته یادگیری ماشین، دقت پیش‌بینی و تصمیم‌گیری سرمایه‌گذاران را بهبود بخشیده و زمینه را برای تحقیقات آینده در حوزه‌هایی مانند استفاده از مدل‌های فازی، الگوریتم‌های فرا ابتکاری و تحلیل روابط بین شاخص‌های مالی فراهم می‌کند.

کلیدواژه‌ها: تحلیل پوششی داده‌ها، بهینه‌سازی پرتفوی، یادگیری ماشین، منابع داده چندگانه.

۱- مقدمه

بهینه‌سازی پرتفوی یکی از کانون‌های پژوهشی مهم درزمینه مالی از زمان نظریه کلاسیک پرتفوی میانگین- واریانس ارایه‌شده توسط مارکوویتز بوده است [1]. مدل مارکوویتز به‌عنوان آغازنظریه مدرن پرتفوی، ابتدا یک راه‌حل کارآمد برای تبادل بین بیشینه‌سازی بازده مورد انتظار و کمینه‌سازی

ریسک ارایه می‌دهد. این مدل‌های کلاسیک پرتفوی معمولا از میانگین بازده تاریخی سهام به‌عنوان بازده مورد انتظار استفاده می‌کنند که برای سرمایه‌گذاری بلندمدت در بازار سهام مناسب است؛ اما این مدل‌ها ممکن است برای سرمایه‌گذاری کوتاه‌مدت مناسب نباشند. این پژوهش بر چگونگی تخصیص ثروت سرمایه‌گذاران بین دارایی‌های مختلف و توسعه یک استراتژی سرمایه‌گذاری موثر برای سرمایه‌گذاران متمرکز است.

به بیان دقیق، یک مصالح سرمایه‌گذاری کامل شامل دو فرآیند انتخاب دارایی و فرمول‌بندی وزن سرمایه‌گذاری است. اولین فرآیند شامل تجزیه‌وتحلیل عملکرد جامع دارایی‌های جایگزین (مانند چشم‌انداز صنعت، وضعیت مالی شرکت، عملکرد تاریخی، شرایط بازار و ...) سپس شناسایی دارایی‌های سرمایه‌گذاری به‌عنوان زیربنای پرتفوی‌ها است. فرآیند دوم را می‌توان به‌عنوان یک مصالح بهینه‌سازی (به‌عنوان‌مثال، به حداقل رساندن ریسک پرتفوی برای یک بازده مورد انتظار معین، یا به حداکثر رساندن بازده مورد انتظار برای یک ریسک پرتفوی معین) که هدف آن تامین سرمایه‌گذاران با وزن سرمایه‌گذاری موثر است، درنظر گرفت.

با این‌حال، مطالعات موجود در مورد بهینه‌سازی پرتفوی عمدتا نقش انتخاب دارایی را نادیده می‌گیرند و فقط مصالح انتخاب پرتفوی را به‌عنوان یک مصالح بهینه‌سازی خالص درنظر می‌گیرند. در فرآیند سرمایه‌گذاری واقعی، انتخاب دارایی‌های اساسی به همان اندازه حیاتی است و بازده پرتفوی به‌دست‌آمده از انتخاب دارایی‌های مختلف نیز تفاوت‌های زیادی دارد. از طرفی داده‌های چندمنبعی نیز نقش مهمی در تحلیل ارزش سرمایه‌گذاری سهام دارند.

با این‌حال، اکثر مطالعات انجام‌شده از مدل‌های اقتصادسنجی و رویکردهای داده‌کاوی برای تحلیل ارزش سرمایه‌گذاری سهام استفاده می‌کنند؛ بنابراین، نتایج این مطالعات حساسیت بیشتری نسبت به تعیین پارامترها و انتخاب مدل‌ها دارد. تحلیل پوششی داده‌ها[1]، به‌عنوان یک روش ارزیابی ناپارامتریک، می‌تواند به‌طور موثر از چنین مشکلاتی جلوگیری کند، زیرا شکل تابع تولید آن نیازی به مشخص شدن ندارد. رویکرد *DEA* به‌عنوان یک روش ارزیابی مبتنی بر داده بر اساس شاخص‌های ورودی و خروجی چندگانه، به‌طور گسترده در ارزیابی عملکرد واحدهای تصمیم‌گیرنده استفاده‌شده است. بدیهی است که از این روش می‌توان برای ارزیابی ارزش سرمایه‌گذاری سهام استفاده کرد.

همچنین نتایج پیش‌بینی قیمت سهام نیز می‌تواند به‌عنوان یک شاخص مهم برای ساخت یک پرتفوی کارآمد درنظر گرفته شود. با توجه به تعداد بالای سهام‌داران و حجم بالای نقدینگی در بازار سرمایه، آگاهی نسبت به وضعیت آینده سهام به‌منظور دانستن بهترین زمان انجام معامله، چگونگی شناسایی سهام قابل سرمایه‌گذاری در بازار سرمایه با تعداد زیادی دارایی و چگونگی ایجاد یک پرتفوی موثر برای سرمایه‌گذاران و اطمینان خاطر بیشتر از موضوعات کلیدی هستند که باید به آن‌ها پرداخته شود. با توجه به موارد فوق، هدف اصلی این پژوهش بهینه‌سازی پرتفوی با استفاده از ادغام *DEA* با منابع داده چندگانه با رویکرد یادگیری ماشین خواهد بود.

به‌منظور دستیابی به هدف تعیین‌شده، در گام نخست، چندین منبع داده نظیر داده‌های تاریخی معاملات و شاخص‌های فنی گردآوری می‌شود تا امکان تجزیه‌وتحلیل ارزش سرمایه‌گذاری سهام فراهم گردد. در ادامه، از روش *DEA* به‌منظور طراحی یک چارچوب انتخاب سهام بهره گرفته می‌شود که به ارزیابی ارزش سرمایه‌گذاری سهام بر اساس بازده تاریخی و همبستگی میان دارایی‌ها می‌پردازد. همچنین، الگوریتم ماشین بردار پشتیبان[2] با داده‌های چندمنبعی تلفیق شده و جهت پیش‌بینی حرکت قیمت سهام مورد استفاده قرار می‌گیرد. افزون بر این، طرح پیشنهادی انتخاب سهام و نوسانات پیش‌بینی‌شده قیمت سهام در مدل بهینه‌سازی پرتفوی ادغام شده و در نهایت، یک استراتژی سرمایه‌گذاری مؤثر برای سرمایه‌گذاران ارایه می‌شود.

۲- مبانی نظری و پیشینه پژوهش

مارکوویتز پیشنهاد کرد که "تخم‌مرغ‌های خود را در یک سبد قرار ندهید [1]، یعنی سرمایه‌گذاران باید سرمایه‌گذاری‌های خود را تا حد امکان متنوع کنند تا ریسک سرمایه‌گذاری را کاهش دهند. با این‌حال، در فرآیند سرمایه‌گذاری واقعی، اگر پرتفوی بیش‌ازحد متنوع باشد، سرمایه‌گذاران باید هزینه‌های تراکنش بیشتری را بپردازند و استراتژی سرمایه‌گذاری پیشنهادی نیز لزوما در عمل بهینه نیست. برای حل این مشکل، کینز و همکاران [2]

[1] Data Envelopment Analysis (DEA)

[2] Support Vector Machine (SVM)

استدلال کردند که سرمایه‌گذاران باید ثروت خود را در تعداد کمی از سهام با ارزش سرمایه‌گذاری، متمرکز کنند و درنتیجه تنوع پرتفوی‌ها را کاهش دهند. این استراتژی سرمایه‌گذاری به‌عنوان استراتژی کم‌نور نیز نامیده می‌شود. از آن زمان، مصالح عدم تنوع پرتفوی‌ها به‌طور گسترده مانند [7–3] موردمطالعه قرارگرفته است.

با این‌حال، مطالعات فوق به چگونگی انتخاب دارایی‌های با ارزش سرمایه‌گذاری نمی‌پردازند؛ بنابراین، چگونگی شناسایی سهام قابل سرمایه‌گذاری در بازار اوراق بهادار با تعداد زیادی دارایی و چگونگی ایجاد یک پرتفوی موثر برای سرمایه‌گذاران نیز، موضوعات کلیدی هستند که در این مقاله باید به آن‌ها پرداخته شود. دلایل انتخاب تحلیل پوششی داده‌ها به‌عنوان الگو بی مناسب در اندازه‌گیری کارایی فنی واحدها را می‌توان در توانایی منحصربه‌فرد این الگو در ارزیابی واقع‌بینانه، ارزیابی هم‌زمان مجموعه عوامل، عدم نیاز به وزن‌های از قبل تعیین‌شده، جبرانی بودن، ارزیابی با گرایش مرزی و تصو یر کردن بهترین وضعیت عملکردی به‌جای وضعیت مطلوب عنوان کرد [8].

علاوه بر مزایای یادشده، این روش قادر است کارایی را برای تعداد زیادی از سهام بورس اوراق‌بهادار تهران مستقل از واحد اندازه‌گیری متای ورودی و خروجی اندازه‌گیری کند؛ بنابراین، در پژوهش حاضر روش ناپارامتریک تحلیل پوششی داده‌ها استفاده‌شده است. ادیرایسینگه و ژانگ [9] یک مدل *DEA* تعمیم‌یافته را برای تجزیه‌وتحلیل اصول بنیادی شرکت‌ها، انتخاب سهام مناسب برای سرمایه‌گذاری و ارایه یک استراتژی سرمایه‌گذاری موثر به سرمایه‌گذاران پیشنهاد کردند. چن [10] از دو مدل *DEA* برای ارزیابی کارایی شرکت‌ها و ایجاد پرتفوی با انتخاب سهام با کارایی بالاتر استفاده کرد. نتایج تجربی نشان داد که پرتفوی‌های مبتنی بر *DEA* می‌توانند بازده مازاد قابل‌توجهی را به‌دست آورند.

اسکرینجاریک یک مدل پو یا *SBM DEA* را برای ارزیابی کارایی نسبی سهام پیشنهاد کرد که در میان آن بازده سه‌ماهه به‌عنوان یک خروجی، انحراف استاندارد به‌عنوان ورودی و ضرایب چولگی به‌عنوان شرایط پیوند درنظر گرفته می‌شود. یافته‌های تجربی نشان می‌دهد که استراتژی سرمایه‌گذاری مبتنی بر *DEA* پو یا می‌تواند از عملکرد بازار بهتر عمل کند [11]. هوآنگ و همکاران [12] از *DEA* برای یافتن سهام با ارزش سرمایه‌گذاری استفاده کرد و سپس یک مدل بهینه‌سازی پرتفوی چندهدفه یکپارچه را توسعه داد. گردیجان و اسکرینجار [13] یک انتخاب سهام را با استفاده از امتیازهای کارایی *DEA* انجام دادند، همچنین دو استراتژی سرمایه‌گذاری مبتنی بر *DEA*، ازجمله نسبت‌های مساوی و استراتژی‌های وزن‌دهی امتیاز، ساختند. نتایج نشان می‌دهد که استراتژی سرمایه‌گذاری مبتنی بر *DEA* می‌تواند از بازده بازار بهتر عمل کند.

شاخص‌های ورودی و خروجی طرح انتخاب سهام مبتنی بر *DEA* در موارد بالا، بیشتر از صورت‌های مالی و داده‌های معاملات سهام شرکت‌های فهرست شده مشتق شده‌اند. دوره انتشار اطلاعات مالی شرکت‌های بورسی نسبتا طولانی است که دارای تاخیر زمانی آشکار است. با توجه به شیوع روزافزون معاملات با فرکانس بالا، روش‌های انتخاب سهام و وزن‌دهی پرتفوی بر اساس *DEA* فوق به‌طور قابل‌توجهی پاسخگو نیستند. درواقع، داده‌هایی که سرمایه‌گذاران می‌توانند مستقیما مشاهده کنند، داده‌های معاملاتی سهام است (به‌عنوان‌مثال، قیمت افتتاحیه، بالاترین قیمت، پایین‌ترین قیمت، قیمت پایانی و شاخص‌های فنی و ...).

ساختار این مقاله به شرح زیر است: در بخش ۳، یک طرح انتخاب سهام به‌منظور تجزیه‌وتحلیل ارزش سرمایه‌گذاری سهام ارایه می‌شود. سپس، مدلی برای بهینه‌سازی پرتفوی بر اساس طرح انتخاب سهام و نوسانات قیمت سهام حاصل از الگوریتم ماشین بردار پشتیبان توسعه داده می‌شود. در بخش ۴، سهام تشکیل‌دهنده شاخص بورس اوراق بهادار تهران به‌عنوان گزینه‌های جایگزین برای ارزیابی طرح انتخاب سهام و تدوین استراتژی‌های سرمایه‌گذاری در چارچوب خارج از نمونه در نظر گرفته می‌شوند. در نهایت، نتایج پژوهش در بخش ۵ جمع‌بندی و ارایه می‌گردد.

۳- روش‌شناسی پژوهش

۳-۱- گردآوری منابع داده

در فرآیند انتخاب پرتفوی، روش موثر انتخاب سهام تا حد زیادی تعیین‌کننده موفقیت استراتژی سرمایه‌گذاری خواهد بود. در عمل، سرمایه‌گذاران معمولا سهام را بر اساس عملکرد تاریخی آن‌ها بررسی می‌کنند. با این‌حال، برخی از پژوهشگران معتقدند که یک اوراق بهادار که درگذشته خوب عمل می‌کند لزوما عملکرد خوب آن را در آینده تضمین نمی‌کند. درواقع، همبستگی بین عملکرد گذشته یک دارایی و عملکرد آتی آن به اثربخشی بازارهای مالی بستگی دارد. واضح است که داده‌های تاریخی معاملات سهام نمی‌توانند به‌طور کامل ارزش سرمایه‌گذاری آتی سهام را منعکس کنند.

به این دلایل، برای ارزیابی ارزش سرمایه‌گذاری سهام، داده‌های چندگانه از سهام، ازجمله داده‌های معاملاتی تاریخی و شاخص فنی را جمع‌آوری خواهیم کرد.

داده‌های تاریخی معاملات

این مقاله داده‌های معاملات تاریخی روزانه سهام را جمع‌آوری می‌کند. اطلاعات جمع‌آوری‌شده شامل قیمت افتتاحیه، بالاترین قیمت، پایین‌ترین قیمت، قیمت بسته شدن و حجم معاملات می‌باشد. داده‌های معاملات تاریخی فوق می‌توانند عملکرد تاریخی سهام را منعکس کنند. همچنین می‌توانند راهنمایی استراتژیک متناظری را برای انتخاب سهام سرمایه‌گذاران تا حد معینی ارایه دهند.

شاخص‌های فنی

علاوه بر داده‌های تاریخی معاملات، داده‌های شاخص فنی سهام نیز می‌تواند مرجعی برای انتخاب سهام باشد. با اشاره به کار [14] و [15]، ما شاخص تصادفی[1]، شاخص قدرت نسبی[2] و میانگین متحرک همگرایی و واگرایی[3] را به‌عنوان شاخص‌های ارزیابی فنی انتخاب می‌کنیم. از فرمول محاسبه شاخص‌های فنی [16] می‌توانیم ارزش شاخص‌های فنی را بر اساس داده‌های معاملات تاریخی به‌دست آوریم. فرمول‌های محاسباتی به شرح زیر است:

میانگین متحرک همگرا - واگرا[4]

این شاخص ارتباط بین دو میانگین متحرک نمایی[5] *EMA* قیمت‌ها را نشان می‌دهد. این شاخص از تفاوت میانگین متحرک نمایی ۲۶ روزه سهم از میانگین متحرک نمایی ۱۲ روزه قیمت آن برآورد می‌شود. خط سیگنال که همان *EMA* ۹ روزه است، معیاری است که از مقایسه آن با *MACD* فرصت‌های خریدوفروش شناسایی می‌گردند. بر این اساس حالت فروش زمانی است که *MACD* زیرخط سیگنال قرار می‌گیرد. به‌طور مشابه، یک سیگنال خرید زمانی روی می‌دهد که *MACD* بالای خط سیگنال قرار بگیرد. رابطه آن به شرح **معادله (۱)** است.

$$MACD = EMA[Stock\ Prices, 12] - EMA[Stock\ Prices, 26]. \quad (۱)$$

شاخص قدرت نسبی[6]

شاخص *RSI* مقدار بازده‌های اخیر سهام با زیان‌های اخیر آن را مورد مقایسه قرار می‌دهد و این اطلاعات را به عددی بین ۰ تا ۱۰۰ تبدیل می‌کند. این شاخص را می‌توان از **معادله (۲)** محاسبه نمود. اگر *RSI* از ۳۰ بیشتر شود سیگنال فروش و اگر از ۷۰ کمتر شود سیگنال خرید صادر می‌شود.

$$RSI = 100 - \frac{100}{1+RS}. \quad (۲)$$

$$RS = \frac{\text{متوسط بازده ها در } n \text{ روزی که قیمت پایانی بیشتر از قیمت پایانی روز قبل بوده است}}{\text{متوسط زیان ها در } n \text{ روزی که قیمت پایانی کمتر از قیمت پایانی روز قبل بوده است}}. \quad (۳)$$

شاخص تصادفی (KDJ)

معمولا برای تجزیه‌وتحلیل روندهای کوتاه‌مدت یا میان‌مدت بازار سهام استفاده می‌شود. از طریق تجزیه‌وتحلیل شاخص KDJ، پژوهشگران می‌توانند روند قیمت سهام را پیش‌بینی کنند و از قدرت بازار احتیاط کنند. برای پیش‌بینی حرکت قیمت سهام، شاخص سریع K را در شاخص KDJ انتخاب می‌کنیم. شاخص KDJ (%K) دارای مقداری بین ۰ تا ۱۰۰ است. به‌طورکلی، مقدار بالای ۹۰ بیش‌ازحد

[1] Random Index (KDJ)
[2] Relative Strength Index (RSI)
[3] Moving Average Convergence Divergence (MACD)
[4] Moving Average Convergence Divergence (MACD)
[5] Exponential Moving Average (EMA)
[6] Relative Strength Index (RSI)

خرید و زیر ۱۰ بیش‌ازحد فروخته می‌شود. فرمول محاسبه به شرح معادله (۴) است.

$$\%K = \left[\frac{\left(\text{کمترین قیمت} - \text{قیمت بسته شدن}\right)}{\left(\text{کمترین قیمت} - \text{بالاترین قیمت}\right)}\right] \times 100. \qquad (۴)$$

۲-۳- طراحی طرح انتخاب سهام

ارزش سرمایه‌گذاری سهام از دو جنبه زیر ارزیابی می‌شود:

۱. عملکرد بازده تاریخی سهام در حال ارزیابی. بسیاری از پژوهشگران بر این باورند که اثر حرکت خاصی در بازار سهام وجود دارد، یعنی سهامی که درگذشته عملکرد خوبی داشته‌اند، در آینده نیز عملکرد خوبی خواهند داشت؛ بنابراین، این مقاله عمدتا بر تحلیل میانگین و واریانس بازده تاریخی سهام متمرکز است. اگر یک سهم بازده مورد انتظار تاریخی بالاتر و واریانس کمتری داشته باشد، نشان می‌دهد که سهام ممکن است دارای ارزش سرمایه‌گذاری مشخصی باشد.

۲. میزان همبستگی بین سهام در حال ارزیابی. هنگام ساخت یک سبد، سرمایه‌گذاران ترجیح می‌دهند سهامی را با همبستگی کمتر انتخاب کنند تا بتوانند ریسک‌های سرمایه‌گذاری را به‌طور موثرتری پوشش دهند.

در پژوهش حاضر از مدل تابع فاصله جهت‌دار برای ارزیابی ارزش سرمایه‌گذاری سهام استفاده می‌شود. در مدل *DDF* برای تغییرات ورودی و خروجی جهتی انتخاب می‌شود و واحدها در آن جهت حرکت می‌کنند. ماتریس قیمت‌های تاریخی m سهام درنظر گرفته‌شده A است که A یک ماتریس $T\times m$ است و T نشان‌دهنده تعداد روزهای معاملاتی است. بر اساس قیمت‌های تاریخی جمع‌آوری‌شده (توجه داشته باشید که قیمت‌های بسته‌شدن روزانه سهام در این پژوهش استفاده‌شده است)، بازده ناخالص روزانه سهام را می‌توان به‌صورت $\Delta p_{t+1,j} = 1 + \left(p_{t+1,j} - p_{t,j}\right)/p_{t,j}$ بیان کرد که در آن $p_{t,j}$ و $p_{t+1,j}$ به‌ترتیب نشان‌دهنده قیمت بسته شدن سهام j در بازه‌های زمانی t و $t+1$ است $t = 1,2,\dots,T-1$ و $j = 1,2,\dots,m$. علاوه بر این، فرض می‌کنیم که r_j و σ_j به‌ترتیب میانگین و واریانس نمونه بردار بازده تاریخی $\Delta p_j = \left[\Delta p_{2,j}, \Delta p_{3,j}, \dots, \Delta p_{T,j}\right]'$ را نشان می‌دهند که در آن $j = 1,2,\dots,m$ است. ما از فرض بازده به مقیاس متغیر[1] برای ایجاد **امکان پرتفوی (۵)** که تحت چارچوب میانگین- واریانس کلاسیک تنظیم شده است، استفاده می‌کنیم.

$$P_1 = \left\{ r,\sigma) | r \le \sum_{j=1}^{m} \lambda_j r_j, \sigma \ge \sum_{j=1}^{m} \lambda_j \sigma_j, \sum_{j=1}^{m} \lambda_j = 1, \lambda_j \ge 0, j = 1,2,\dots,m \right\}. \qquad (۵)$$

مجموعه امکان پرتفوی (۵) از کار مارکوویتز مشتق شده است که می‌تواند به‌عنوان تقریب خطی جزئی آن در مارکوویتز درنظر گرفته شود [17]. ساخت **مجموعه امکان (۵)** توسط سایر پژوهشگران [23–18] در ارزیابی پرتفوی به رسمیت شناخته شده است. بازده به مقیاس بیانگر پیوند بین تغییرات ورودی‌ها و خروجی‌های یک سیستم می‌باشد. درواقع افزایش متناسب بیشینه همه خروجی‌ها به سبب افزایش متناسب معین همه ورودی‌ها است [24].

هنگامی‌که فقط به رابطه بین معیارهای میانگین و واریانس نگاه می‌کنیم، فرض *VRS* ممکن است به دلیل رابطه غیرخطی بین این دو کاربرد بیشتری داشته باشد. تنها زمانی که میانگین و انحراف استاندارد به‌عنوان شاخص‌های ارزیابی استفاده می‌شود، می‌توان فرض بازگشت ثابت به مقیاس[2] (که معادل خطی بودن است) اتخاذ شود، زیرا میانگین و انحراف استاندارد یک رابطه خطی دارند [25]. علاوه بر این، ازنظر تئوری قیمت لذت‌گرا، با توجه به ماهیت غیرخطی عموما در توابع قیمت ضمنی برای ویژگی‌ها، نیاز به تحمیل *VRS* به‌جای *CRS* است [26].

CRS ممکن است در ارزیابی پرتفوی قابل‌اجرا نباشد، زیرا اجازه می‌دهد بازدهی بیشتر از هر پرتفوی تولید کند [27]؛ بنابراین، در مقایسه با مدل *CRS*، معتقدیم که مدل *VRS* ممکن است برای تحلیل ارزش سهام در این پژوهش مناسب‌تر باشد. اجازه دهید O پانو یس سهام مورد ارزیابی را مشخص کند. در ادامه، از تابع فاصله جهت‌دار کلاسیک برای ارزیابی عملکرد سهام استفاده می‌کنیم که می‌تواند ریسک را کاهش و بازده را افزایش

[1] Variable Return to Scale (VRS)

[2] Constant Return to Scale (CRS)

دهد. برای یک‌جهت طرح‌ریزی داده شده $(g_r, -g_\sigma)'$، با فرض $(r_0 + \eta g_r, \sigma_0 - \theta g_\sigma) \in P_1$، می‌توانیم مدل ارزیابی *DEA* زیر را بسازیم: $\hat{\theta}_1 = \min_{\theta,\eta}\{\frac{1-\theta}{1+\eta} | (r_0 + \eta g_r, \sigma_0 - \theta g_\sigma) \in P_1\}$. به‌طور خاص، می‌توان آن را به‌صورت **مدل (۶)** بیان کرد:

$$\theta_1 = min \frac{1-\theta}{1+\eta}.$$

$$s.t. \begin{cases} \sum_{j=1}^{m} \lambda_j r_j \geq r_o + \eta g_r, \\ \sum_{j=1}^{m} \lambda_j \sigma_j \leq \sigma_o - \theta g_\sigma, \\ \sum_{j=1}^{m} \lambda_j = 1, \theta \geq 0, \eta \geq 0, \\ \lambda_j \geq 0, j = 1,2, \dots, m. \end{cases} \quad (۶)$$

توجه داشته باشید که، در اینجا فرض می‌کنیم که جهت $g_r = r_0$ و $g_\sigma = \sigma_0$ را برآورده می‌کند. ازآنجایی‌که تابع هدف **مدل (۶)** یک تابع غیرخطی است، **مدل (۶)** را با تبدیل زیرخطی می‌کنیم. اجازه دهید $1/\tau = 1 + \eta, \tilde{\lambda}_j = \tau\lambda_j, \tilde{\theta} = \tau\theta \; and \; \tilde{\eta} = \tau\eta$، سپس **مدل (۶)** را می‌توان به‌صورت **مدل (۷)** بازنویسی کرد.

$$\theta_1 = min \, \tau - \widetilde{\theta}.$$

$$s.t. \begin{cases} \sum_{j=1}^{m} \lambda_j r_j \geq \tau r_o + \eta g_r \, , \\ \sum_{j=1}^{m} \lambda_j \sigma_j \leq \tau\sigma_o - \theta g_\sigma \, , \\ \tau + \eta = 1, \eta \geq 0, \tau \geq 0, \theta \geq 0, \\ \sum_{j=1}^{m} \lambda_j = \tau, \lambda_j \geq 0, j = 1,2, \dots m \, . \end{cases} \quad (۷)$$

مدل‌های (۶) و **(۷)** عمدتا کارایی سهام را از منظر میانگین و واریانس بازده ارزیابی می‌کنند. با این‌حال، کار پیشگام مارکوویتز نشان داد، که هر چه همبستگی بین سهام کمتر باشد، پرتفوی بیشتر می‌تواند ریسک را کاهش دهد [1]. بر اساس بازده سهام تحت ارزیابی، ماتریس ضریب همبستگی را می‌توان محاسبه کرد که با $\Theta = \left[\rho_{i,j}\right]_{m\times m}$ مشخص می‌شود. به این معنا که، برای دارایی j می‌توانیم دنباله‌ای از ضرایب همبستگی بین بازده این

$$P_2 = \left\{ \rho, \sigma) | \rho \geq \sum_{j=1}^{m} \mu_j \rho_j \, , \sigma \geq \sum_{j=1}^{m} \mu_j \sigma_j \, , \sum_{j=1}^{m} \mu_j = 1, \mu_j \geq 0, j = 1,2, \dots, m \right\}. \quad (۸)$$

دارایی و سایر دارایی‌های جایگزین را به‌دست آوریم. علاوه بر این، باید دنباله ضریب همبستگی بالا را اندازه‌گیری کنیم تا همبستگی جامع بین دارایی j و سایر دارایی‌ها را توصیف کنیم. به‌منظور اندازه‌گیری بیشتر همبستگی بین سهام انتخاب‌شده، این پژوهش میانگین و واریانس ضرایب همبستگی در ستون j را به ترتیب ρ_j و σ_j درنظر می‌گیرد که $j = 1,2, \dots, m$ است. سرمایه‌گذاران می‌خواهند ρ_j و σ_j تا حد امکان کوچک باشند، زیرا این می‌تواند اطمینان حاصل کند که ضریب همبستگی بین سهام انتخاب‌شده و سایر سهام دارای میانگین و نوسان کم است. به‌طور مشابه، با استفاده از فرض *VRS* می‌توانیم **مجموعه امکان پرتفوی (۸)** را بسازیم.

هر دو میانگین و واریانس در **مجموعه (۸)** دو شاخص ارزیابی هستند که از ماتریس ضریب همبستگی Θ مشتق شده‌اند. در اینجا، معتقدیم که هر چه دو شاخص فوق کوچک‌تر باشند، بهتر است. بر اساس مجموعه بالا، برای یک‌جهت طرح‌ریزی داده شده $\left(-g_\rho, -g_\sigma\right)'$، با فرض اینکه $\left(\rho_0 - \alpha g_\rho, \sigma_0 - \beta g_\sigma\right) \in P_2$، می‌توانیم مدل ارزیابی *DEA* زیر را بسازیم: $\theta_2 = \min_{\alpha,\beta}\left\{1 - \frac{\alpha+\beta}{2} | \left(\rho_0 - \alpha g_\rho, \sigma_0 - \beta g_\sigma\right) \in P_2\right\}$. شکل خاص را می‌توان به‌صورت **معادله (۹)** بیان کرد.

$$\theta_2 = min \, 1 - \; \alpha + \beta)/2.$$

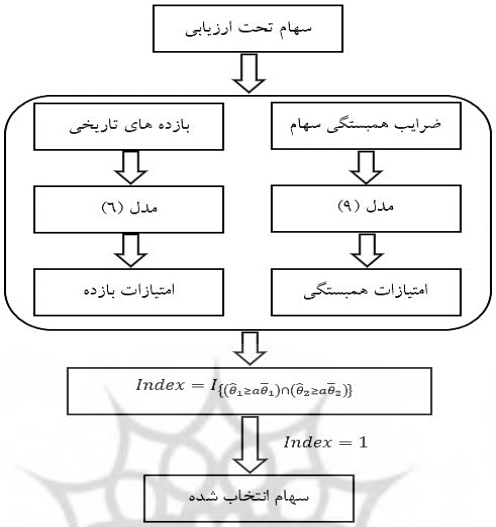
$$s.t. \begin{cases} \sum_{j=1}^{m} \mu_j \rho_j \leq \rho_o - \alpha g_\rho, \\ \sum_{j=1}^{m} \mu_j \sigma_j \leq \sigma_o - \beta g_\sigma, \\ \sum_{j=1}^{m} \mu_j = 1 \, , \alpha \geq 0, \beta \geq 0, \\ \mu_j \geq 0, j = 1,2, \dots, m. \end{cases} \quad (۹)$$

در اینجا، جهت بالا فرض می‌شود که، $g_\rho = \rho_0$ و $g_\sigma = \sigma_0$ را برآورده می‌کند زیرا داده‌های مطالعه تجربی همه بزرگ‌تر از ۰ هستند. بر اساس مدل‌های *DEA* فوق، می‌توانیم کارایی متناظر سهام را تحت معیارهای ارزیابی مختلف یعنی، $\hat{\theta}_1$ و $\hat{\theta}_2$ به‌دست آوریم. به‌منظور تسهیل طراحی طرح

انتخاب سهام، میانگین کارایی تمام سهام تحت معیارهای ارزیابی مختلف به ترتیب $\bar{\theta}_1$ و $\bar{\theta}_2$ تعریف می‌شود. سپس، می‌توانیم شاخص انتخاب سهام زیر را بسازیم:

$$Index = I_{\{(\theta_1 \geq a\theta_1)\cap(\theta_2 \geq a\theta_2)\}}, \qquad (۱۰)$$

که در آن $I_{\{.\}}$ یک تابع نشان‌گر است و ضریب a یک سطح از پیش تعیین‌شده بین ۰ و ۱ است. در اینجا a را می‌توان به‌عنوان اولویت انتخاب سهام تصمیم‌گیرندگان درنظر گرفت. بدیهی است که هر چه ارزش a بزرگ‌تر باشد، الزامات تصمیم‌گیرندگان سخت‌گیرانه‌تر است و بنابراین سهام کمتری انتخاب می‌شود. مراحل خاص انتخاب سهام در **شکل ۱** نشان داده شده است.



شکل ۱- جریان انتخاب سهام بر اساس DEA با منابع داده چندگانه.

Figure 1- Stock selection flow based on DEA with multiple data sources.

بر اساس فرآیند انتخاب سهام نشان داده شده در **شکل ۱**، سهام مربوطه را می‌توان به‌عنوان دارایی‌های اساسی برای ساخت پرتفوی انتخاب کرد.

۳-۳- پیش‌بینی حرکات آتی قیمت سهام

داده‌های چند منبع فوق نه‌تنها می‌توانند به‌عنوان مبنایی برای انتخاب سهام، بلکه برای پیش‌بینی روند آتی قیمت سهام نیز مورداستفاده قرار گیرند. فراز و نشیب‌های آتی سهام نیز می‌تواند حمایت تصمیم‌گیری مربوطه را برای سرمایه‌گذاران برای ایجاد یک سبد سرمایه‌گذاری موثر فراهم کند.

ایده الگوریتم طبقه‌بندی *SVM* را به‌طور خلاصه به‌صورت زیر شرح می‌دهیم:

برای مجموعه آموزشی داده شده $A = \{(\hat{x}_1, \hat{y}_1), (\hat{x}_2, \hat{y}_2), \dots, (\hat{x}_N, \hat{y}_N)\}$، مجموعه ویژگی $\hat{x}_i \in R^{\hat{m}}$ است و آیتم دسته‌بندی $\hat{y}_i = \{0,1\}$ را برآورده می‌کند، جایی که $i = 1,2,\dots,N$ است. در این پژوهش، N نشان‌دهنده طول دوره نمونه (یعنی تعداد روزهای معاملاتی) و $\hat{m}$ نشان‌دهنده مقدار داده‌های آموزشی (یعنی داده‌های معاملات تاریخی و شاخص‌های فنی) در روز معاملاتی i که $i = 1,2,\dots,N$، برای هر سهم در حال ارزیابی است.

تابع نگاشت $\emptyset(\hat{x})$ و همچنین پارامتر جریمه C را انتخاب کنید و جواب‌های بهینه w^*, b^*, ε^* را با توجه به **مصالح بهینه‌سازی (۱۱)** به‌دست آورید.

$$\min_{w,b,\varepsilon} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\varepsilon_i \qquad (۱۱)$$
$$s.t. \quad \hat{y}_i\, w^T \emptyset\, \hat{x}_i) + b) + \varepsilon_i \geq 1$$
$$\varepsilon_i \geq 0, i = 1,2,\dots,N.$$

در اینجا، ε_i عبارت ریسک تجربی است.

تابع کرنل $K(\hat{x}_i,\hat{x}_j) = \varnothing\ \hat{x}_i).\varnothing(\hat{x}_j)$ در مدل‌سازی از تابع کرنل پایه شعاعی استفاده می‌شود.

ابر صفحه متمایز شده $w^*)^T\hat{x} + b^* = 0$ را بر اساس راه‌حل‌های بهینه w^*، b^*، ε^* بسازید **تابع تصمیم (۱۲)** را به‌دست آورید:

$$f\ \hat{x}) = sgn\quad w^*)^T\varnothing\ \hat{x}) + b^*). \qquad (۱۲)$$

توجه داشته باشید که $sgn\ \ z)$ نشان‌دهنده تابع علامت است، زمانی که $sgn\ \ z) = 1$ ،$z \geq 0$، در غیر این صورت $sgn\ \ z) = 0$.

علاوه بر این، شکل دوگان **مدل (۱۱)** را می‌توان به‌صورت **مدل (۱۳)** بیان کرد.

$$\min_a \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{y}_i\hat{y}_j a_i a_j K(\hat{x}_i,\hat{x}_j) - \sum_{i=1}^{N} a_i.$$
$$s.t.\begin{cases}\sum_{i=1}^{N}\hat{y}_i a_i = 0,\\ 0 \leq a_i \leq C, i = 1,2,\dots,N.\end{cases} \qquad (۱۳)$$

جواب بهینه مدل‌بالا به‌صورت $a^* = \ a_1^*, a_2^*, \dots, a_N^*)^T$ ثبت می‌شود، سپس تابع تصمیم نیز می‌تواند به‌صورت **معادله (۱۴)** بیان شود:

$$f\ \hat{x}) = sgn\left(\sum_{i=1}^{N}\hat{y}_i a_i^* K\ \hat{x}_i,\hat{x}) + b^*\right). \qquad (۱۴)$$

۴-۳- ارایه و ارزیابی مدل بهینه‌سازی پیشنهادی

بر اساس طرح انتخاب سهام پیشنهادی، می‌توانیم گروهی از سهام را از سهام درحال ارزیابی غربال کنیم که می‌توان آن‌ها را به‌عنوان دارایی‌های زیربنایی پرتفوی درنظر گرفت. برای راحتی، فرض می‌کنیم که تعداد سهام انتخاب‌شده n است و ماتریس بردار میانگین و کوواریانس بازده تاریخی به ترتیب $R_n = \ r_1, r_2, \dots, r_n)'$ و Ω_n نامیده می‌شوند. درادامه، با ادغام طرح انتخاب سهام فوق، نتایج پیش‌بینی قیمت سهام و نظریه پرتفوی، یک استراتژی سرمایه‌گذاری ترکیبی برای سرمایه‌گذاران ارایه خواهیم کرد. در این پژوهش، ما نتایج پیش‌بینی قیمت سهام را با تئوری پرتفوی حداقل واریانس سراسری ترکیب می‌کنیم تا برای سرمایه‌گذاران یک استراتژی سرمایه‌گذاری ترکیبی فراهم کنیم. به‌طورخاص، مدل بهینه‌سازی پرتفوی پیشنهادی را می‌توان به‌صورت **مدل (۱۵)** بیان کرد.

$$\min_x x'\Omega_n x - \phi \times x' f_n\ ,$$
$$s.t.\sum_{i=1}^{n} z_i = k,$$
$$\sum_{i=1}^{n} x_i = 1, \qquad (۱۵)$$
$$l_i z_i \leq x_i \leq u_i z_i \quad i = 1,2,\dots,n),$$
$$z_i \in [0,1] \quad i = 1,2,\dots,n),$$
$$x_i \geq 0 \quad i = 1,2,\dots,n).$$

در اینجا، $f_n = \left(f_n^1, f_n^2, \dots, f_n^n\right)'$ بردار پیش‌بینی‌شده قیمت سهام را نشان می‌دهد. علاوه بر این، f_n^i یک متغیر ۰-۱ است، که در آن $f_n^i = 1$ نشان می‌دهد که قیمت سهام در آینده کاهش نخواهد یافت و $f_n^i = 0$ نشان می‌دهد که قیمت سهام در آینده کاهش می‌یابد، $i = 1,2,\dots,n$. ضریب ϕ یک پارامتر از پیش تعیین‌شده است که می‌تواند به‌عنوان پارامتر معاوضه بین به حداکثر رساندن شاخص پیش‌بینی قیمت سهام و به حداقل رساندن واریانس بازده پرتفوی درنظر گرفته شود. عدد k تعداد سهام موجود در پرتفوی می‌باشد که بیانگر محدودیت کاردینال است. پارامترهای l_i و u_i به ترتیب کف و سقف هر سهم در پرتفوی می‌باشد.

مدل‌های بهینه‌سازی پرتفوی موجود معمولا بر اساس بازده دارایی برای ایجاد هدف بهینه‌سازی هستند و سپس از بازده تاریخی دارایی‌های اساسی برای تخمین پارامترهای ورودی موردنیاز برای هدف بهینه‌سازی استفاده می‌کنند. با این‌حال، درواقع، ویژگی‌های آماری بازده‌های تاریخی دارایی ممکن است در آینده نتوانند آن‌ها را به‌دست آورند که باعث می‌شود عملکرد استراتژی سرمایه‌گذاری حاصل رضایت‌بخش نباشد به‌عنوان‌مثال، [28–30]. ازجمله محدودیت‌های کاربردی در بهینه‌سازی سبد سرمایه‌گذاری می‌توان به حداقل یا حداکثر کردن میزان سرمایه‌گذاری در یک دارایی اشاره نمود که توسط بیل و فارست به مدل اولیه مارکوویتز اضافه شد [31]. با توجه به تمایل سرمایه‌گذاران به در اختیار داشتن و مدیریت سبدی از سهام که تعداد محدودتری از دارایی‌ها را در خود جای‌داده باشد، محدودیت دیگری تحت عنوان محدودیت کاردینالیتی به مدل اضافه شد. برخلاف مدل‌های موجود، مدل پیشنهادی دارای دو ویژگی زیر است: ۱- دارایی‌های اساسی در **مدل (۱۵)** با استفاده از روش انتخاب سهام در بخش ۳-۲ انتخاب می‌شوند و ۲- **مدل (۱۵)** شاخص‌های پیش‌بینی قیمت سهام را دربر می‌گیرد که به یک مبادله بین عملکرد گذشته و آینده پرتفوی می‌رسد و همچنین وابستگی بیش‌ازحد استراتژی سرمایه‌گذاری منتج به عملکرد تاریخی دارایی‌های اساسی را تا حد زیادی تعدیل می‌کند.

برای نشان دادن اعتبار استراتژی سرمایه‌گذاری پیشنهادی، استراتژی حداقل واریانس سراسری کلاسیک را برای مقایسه ارایه می‌کنیم. برای این منظور، ابتدا فرض می‌کنیم که سرمایه‌گذاران شاخص‌های پیش‌بینی قیمت سهام را درنظر نمی‌گیرند و فقط نگران واریانس بازده پرتفوی هستند و سپس می‌توان مدل بهینه‌سازی پرتفوی حداقل واریانس سراسری متناظر را به‌صورت **مدل (۱۶)** ساخت.

$$
\begin{aligned}
&\min_{x} x'\Omega_n x,\\
&s.t. \sum_{i=1}^{n} z_i = k,,\\
&\sum_{i=1}^{n} x_i = 1,,\\
&l_i z_i \le x_i \le u_i z_i, \quad i = 1,2,\dots,n),,\\
&z_i \in [0,1], \quad i = 1,2,\dots,n),,\\
&x_i \ge 0, \quad i = 1,2,\dots,n).
\end{aligned}
\qquad (۱۶)
$$

برخی از شاخص‌های ارزیابی کلاسیک (مانند نسبت شارپ و نسبت سورتینو) را برای اندازه‌گیری بازده پرتفوی اعمال می‌کنیم و سپس تفاوت در عملکرد استراتژی‌های سرمایه‌گذاری مختلف را بررسی می‌کنیم. اجازه دهید دنباله بازده پرتفوی به‌دست‌آمده از یک استراتژی سرمایه‌گذاری مشخص، یک متغیر تصادفی r_p باشد، سپس نسبت شارپ و نسبت سورتینو مربوط به بازده پرتفوی را می‌توان به ترتیب به‌صورت **معادله‌های (۱۷)** و **(۱۸)** بیان کرد.

$$\theta^{Sharpe} = \frac{\bar{E}(r_p) - r_f}{\sqrt{\overline{Var}(r_p)}}. \qquad (۱۷)$$

$$\theta^{Sortino} = \frac{\bar{E}(r_p) - r_f}{\sqrt{\bar{E}\left[\left(r_p - \bar{E}(r_p)\right)^2\right]}}. \qquad (۱۸)$$

در اینجا $\bar{E}(.)$ و $\overline{Var}(.)$ به ترتیب میانگین نمونه و واریانس نمونه را نشان می‌دهند. توجه داشته باشید ازآنجایی‌که افق سرمایه‌گذاری کوتاه است، نرخ بدون ریسک را ۰ فرض می‌کنیم.

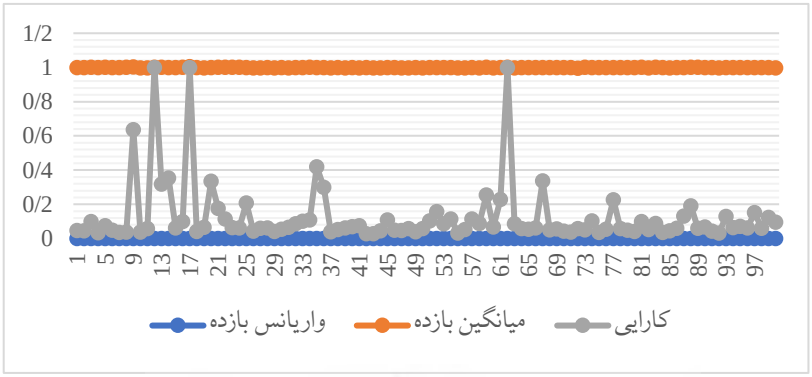
۴- تجزیه‌وتحلیل داده‌ها

در این پژوهش داده‌های جمع‌آوری‌شده، متعلق به ۱۰۰ شرکت پذیرفته‌شده در سازمان بورس و اوراق بهادار تهران از ۲۴ صنعت مختلف است که در حدفاصل شهریورماه ۱۴۰۰ تا آبان‌ماه ۱۴۰۱ اخذوتحلیل شده‌اند. برای انتخاب مجموعه سهام، معیارهای زیر، اعمال شده است:

۱. شرکت قبل از سال ۱۴۰۰ در بورس اوراق بهادار تهران پذیرفته‌شده باشد و نام آن تا پایان آبان ماه ۱۴۰۱ از فهرست شرکت‌های یادشده حذف نشده باشد.
۲. سال مالی آن‌ها به پایان اسفندماه هرسال منتهی شود.

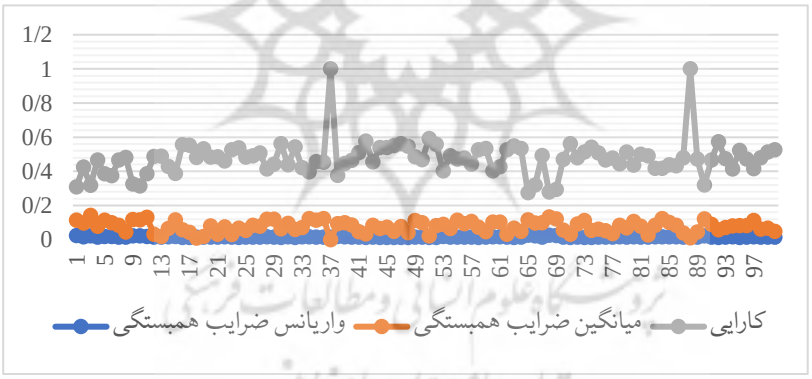
۳. در دوره موردبررسی توقف نماد بیش از ۲ ماه نداشته باشند.
۴. اطلاعات آن‌ها کامل و در دسترس باشد.

سیستم‌عامل مربوط برای اجرای نرم‌افزارها ویندوز نسخه ۱۱ بوده و محاسبات توسط پردازشگر کامپیوتر قابل‌حمل ۸ هسته‌ای با فرکانس ۱٫۸ گیگاهرتز و حافظه موقت ۸ گیگابایت صورت گرفته است. درابتدا کارایی دارایی‌ها ازنظر بازده و همبستگی با استفاده از روش تابع فاصله جهت‌دار معرفی‌شده، ارزیابی می‌شود و دارایی‌های زیربنایی پرتفوی انتخاب می‌گردد.



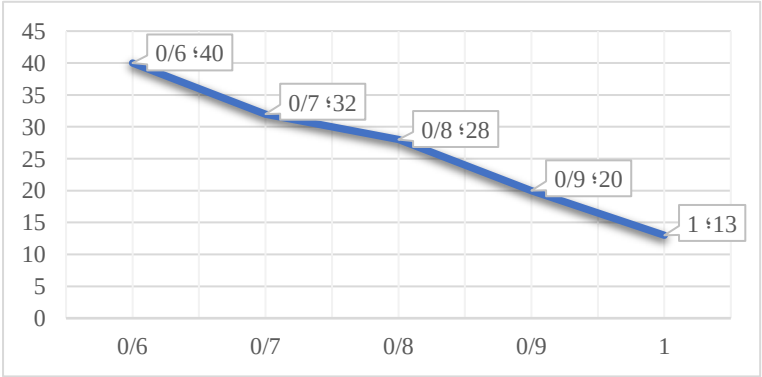
شکل ۲- امتیازات کارایی دارایی‌ها ازنظر میانگین-واریانس بازده.
Figure 2- Asset efficiency scores in terms of mean-variance of returns.

در **شکل ۲** امتیازات کارایی از منظر میانگین- واریانس بازده دارایی‌ها نمایش داده شده است. مطابق شکل سهم‌های ۱۲، ۱۷ و ۶۲ کارا هستند.به‌منظور اندازه‌گیری همبستگی بین داده‌ها از ضریب همبستگی پیرسون استفاده می‌شود.



شکل ۳- امتیازات کارایی دارایی‌ها ازنظر همبستگی بازده.
Figure 3- Asset efficiency scores in terms of return correlation.

در **شکل ۳** امتیازات کارایی از منظر همبستگی بازده دارایی‌ها نمایش داده شده است. مطابق شکل سهم‌های ۳۷ و ۸۸ کارا هستند.



شکل ۴- تعداد سهام انتخابی در سطح اولویت انتخاب سهام a.
Figure 4- Number of selected stocks at stock selection priority level a.

هر چه ارزش a بزرگ‌تر باشد، الزامات تصمیم‌گیرندگان سخت‌گیرانه‌تر است و بنابراین سهام کمتری انتخاب می‌شود. **شکل ۴** تعداد سهام انتخابی در سطوح مختلف اولویت انتخاب سهام را نمایش می‌دهد. با درنظر گرفتن سطح اولویت انتخاب سهام ۰/۶ دارایی‌های زیربنایی پرتفوی به شرح **جدول ۱** انتخاب می‌شوند.

جدول ۱- سهام منتخب در سطح اولویت انتخاب سهام $a = 0.6$.

Table 1- Selected stocks at stock selection priority level a=0.6.

Name	j	Name	j	Name	j	Name	j
فولاد	74	دالبر	52	شپنا	22	ختوقا	3
کنور	77	دتماد	53	شنفت	25	خگستر	5
غشهد	81	دفارا	54	شستا	32	خزامیا	9
غگل	83	دپارس	57	وغدیر	33	مارون	12
پارسیان	87	درازک	58	وبانک	34	بوعلی	13
دانا	88	دسینا	59	وامید	35	جم	14
قشهد	93	وخاور	61	وصندوق	36	حکشتی	16
سیتا	97	وکار	62	فاما	41	حفارس	17
سرود	99	وپارس	63	واعتبار	45	همراه	20
سدشت	100	وپاسار	67	دابور	51	شبندر	21

در گام بعد از ماشین بردار پشتیبان[1] ترکیب‌شده با داده‌های چندمنبعی برای پیش‌بینی حرکات قیمت سهام استفاده می‌کنیم. ابتدا از روش *z_score* برای نرمال‌سازی مجموعه داده‌های‌ورودی استفاده می‌کنیم. سپس از *SVM* برای تنظیم، آموزش و پیش‌بینی داده‌های‌ورودی استفاده می‌شود. ما از روش *Cross-validation* با K برابر ۵ برای یادگیری‌مدل استفاده می‌کنیم. سپس، می‌توانیم هایپرپارامتر[2]ها را با توجه به نتایج‌مجموعه تست تنظیم کنیم. به‌طور کلی، زمانی‌که تابع‌کرنل پایه شعاعی[3] انتخاب می‌شود، عملکرد مدل نسبتا پایدار است؛ بنابراین دراین‌پژوهش ابتدا تابع کرنل *RBF* انتخاب‌شده، سپس، با استفاده از جستجوی تصادفی بازه کوچکتری از مقادیر هایپرپارامترها را فراهم می‌کنیم و فضای‌جستجو را کوچکتر می‌سازیم و در آخر ازروش جستجوی شبکه‌ای برای تنظیم هایپرپارامترهای C و γ استفاده می‌شود. محدوده مقدار هایپرپارامتر جریمه C را روی [1,10] قرار می‌دهیم و طول گام را، ۱ تعیین می‌کنیم. محدوده مقدار γ را روی [0.001,0.1] قرار می‌دهیم و طول گام را، ۰/۰۰۱ تعیین می‌کنیم. با تنظیم هایپرپارامترها می‌توان هایپرپارامترهای دارای بالاترین دقت را استخراج کرد. تنظیمات هایپرپارامترهای الگوریتم *SVM* در **جدول ۲** نشان داده شده است.

جدول ۲- هایپرپارامترهای تنظیم‌شده الگوریتم به ازای تابع کرنل RBF.

Table 2- Tuned hyperparameters of the algorithm for the RBF kernel function.

Gamma	C	j	Gamma	C	j
0.01	1	52	0.02	3	3
0.014	7	53	0.006	1	5
0.008	9	54	0.007	1	9
0.002	8	57	0.039	9	12
0.005	2	58	0.009	5	13
0.019	3	59	0.001	1	14
0.015	1	61	0.005	1	16
0.011	9	62	0.011	1	17
0.008	3	63	0.003	9	20
0.005	1	67	0.031	5	21
0.002	5	74	0.003	3	22
0.002	9	77	0.01	2	25
0.004	3	81	0.016	4	32
0.01	1	83	0.02	1	33
0.008	9	87	0.008	9	34
0.006	7	88	0.001	7	35
0.001	7	93	0.022	7	36
0.019	8	97	0.006	3	41
0.049	8	99	0.001	9	45
0.147	4	100	0.002	8	51

[1] Support Vector Machine (SVM)
[2] Hyperparameter optimization
[3] Radial Basis Function (RBF)

در ادامه، اثرات پیش‌بینی مدل *SVM* را بر اساس شاخص‌های ارزیابی دقت، صحت، فراخوان، امتیاز *F1* ارزیابی می‌کنیم. دراینجا، دقت به نسبت نمونه‌های پیش‌بینی‌شده درست به همه نمونه‌ها اشاره دارد. صحت نشان‌دهنده نسبت نمونه‌های مثبت پیش‌بینی‌شده صحیح به نمونه‌های مثبت پیش‌بینی‌شده است. فراخوان نشان‌دهنده نسبت نمونه‌های مثبت پیش‌بینی‌شده صحیح به نمونه‌های مثبت واقعی است. امتیاز *F1* میانگین هارمونیک صحت و فراخوان است. فرمول‌های محاسباتی به شرح زیر است:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (۱۹)$$

$$Precision = \frac{TP}{TP + FP}, \quad (۲۰)$$

$$Recall = \frac{TP}{TP + FN}, \quad (۲۱)$$

$$F1_{score} = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (۲۲)$$

TP نشان‌دهنده مثبت واقعی، *FP* نشان‌دهنده مثبت کاذب، *TN* نشان‌دهنده منفی واقعی و *FN* نشان‌دهنده منفی کاذب است. جزییات *TP*، *FP*، *TN* و *FN* در **جدول ۳** نشان داده شده است.

جدول۳- معانی TP، FP، TN و FN.

Table 3- Meanings of TP, FP, TN and FN.

مقدار واقعی/مقدار پیش‌بینی‌شده	1	0
1	TP	FN
0	FP	TN

چهار شاخص ارزیابی (دقت، صحت، فراخوان و امتیاز *F1*) بر اساس روش *Cross-validation* برای ارزیابی عملکرد مدل استفاده می‌شود. عملکرد پیش‌بینی قیمت سهام در **جدول ۴** نشان داده شده است. می‌توان دریافت که میانگین دقت، صحت، فراخوان و امتیاز *F1* مجموعه تست همه سهام به ترتیب ۶۷%، ۶۹%، ۵۹% و ۶۱% است. علاوه بر این، ۳۶ سهام با دقت بیش از ۶۰% و ۴۰% سهام با دقت بیش از ۵۰% وجود دارد. ۳۸ سهام با صحت بیش از ۶۰% و ۴۰ سهام با صحت بیش از ۵۰% وجود دارد. ۱۹ سهام با فراخوان بالای ۶۰% و ۲۹ سهام با فراخوان بالای ۵۰% وجود دارد. ۲۱ سهام با امتیاز *F1* بالای ۶۰% و ۳۳ سهام با امتیاز *F1* بالای ۵۰% وجود دارد.

جدول ۴- ارزیابی عملکرد پیش‌بینی قیمت سهام.

Table 4- Stock price forecast performance evaluation.

Cross-Validation					
>0.5	>0.6	متوسط	حداقل	حداکثر	شاخص‌ها
40	36	0.67	0.55	0.83	Accuracy
40	38	0.69	0.54	0.87	Precision
29	19	0.59	0.37	0.97	Recall
33	21	0.61	0.37	0.89	F1-Score

جدول ۴ نشان می‌دهد که نتایج پیش‌بینی اکثر سهام در محدوده قابل قبولی است و بنابراین می‌توان از این نتایج به‌عنوان مرجعی برای سرمایه‌گذاران به‌منظور انتخاب دارایی‌های اساسی یک پرتفوی استفاده کرد. نتایج پیش‌بینی حرکات قیمت سهام مطابق **جدول ۵** است.

جدول ۵ - نتایج پیش‌بینی قیمت سهام.

Table 5- Stock price prediction results.

Predict	j	Predict	j	Predict	j	Predict	j
0	74	0	52	0	22	0	3
1	77	1	53	1	25	1	5
1	81	0	54	0	32	0	9
0	83	0	57	1	33	1	12
0	87	0	58	1	34	0	13
1	88	1	59	0	35	0	14
0	93	0	61	1	36	0	16
1	97	1	62	0	41	0	17
1	99	1	63	0	45	1	20
0	100	1	67	1	51	0	21

درآخر، نتایج پیش‌بینی قیمت سهام را با نظریه پرتفوی حداقل واریانس سراسری ترکیب می‌کنیم تا برای سرمایه‌گذاران یک استراتژی سرمایه‌گذاری ترکیبی ۱ فراهم کنیم. برای نشان دادن اعتبار استراتژی سرمایه‌گذاری پیشنهادی، استراتژی سرمایه‌گذاری حداقل واریانس سراسری کلاسیک ۲ را برای مقایسه ارایه می‌کنیم. افزایش تعداد دارایی‌ها در فرآیند تشکیل سبد سرمایه‌گذاری، درنظرگرفتن محدودیت سرمایه‌گذاری در هر یک از دارایی‌ها و نیز لحاظ نمودن محدودیت کاردینالی موضوع بهینه‌سازی سبد سرمایه‌گذاری را به یک مصالح *NP-Hard* تبدیل می‌نماید. درنتیجه برای حل مدل از الگوریتم ژنتیک استفاده‌شده است. نتایج حاصل از اجرای استراتژی‌های سرمایه‌گذاری ۱ و ۲ در **جدول ۶** آورده شده است.

جدول ۶- پرتفوی بهینه استراتژی‌های سرمایه‌گذاری.
Table 6- Optimal portfolio of investment strategies.

استراتژی سرمایه‌گذاری ۱				
a = 0.9				
12	36	59	62	77
0.3	0.172318	0.148537	0.236434	0.142711
a = 0.6				
12	20	36	51	81
0.3	0.255804	0.244196	0.1	0.1
استراتژی سرمایه‌گذاری ۲				
a = 0.9				
12	13	20	22	36
0.3	0.1	0.281168	0.1	0.218832
a = 0.6				
9	12	13	14	100
0.1	0.3	0.152031	0.3	0.147969

شاخص‌های ارزیابی نسبت شارپ و نسبت سورتینو را برای اندازه‌گیری بازده پرتفوی اعمال می‌کنیم و سپس تفاوت در عملکرد استراتژی‌های سرمایه‌گذاری مختلف را بررسی می‌کنیم. زمانی که کارایی با اولویت انتخاب سهام به ترتیب در ۰/۹ و ۰/۶ ارزیابی شود، نسبت شارپ و نسبت سورتینو بازده پرتفوی در استراتژی‌های ۱ و ۲ **در جدول ۷** نشان داده شده است.

جدول ۷- عملکرد استراتژی‌های سرمایه‌گذاری.
Table 7- Performance of investment strategies.

استراتژی سرمایه‌گذاری	**a = 0.9**	**a = 0.6**
پنل A: نسبت شارپ بازده پرتفوی		
1	0.846691	0.373045
2	0.666201	0.243218
پنل B: نسبت سورتینو بازده پرتفوی		
1	0.946629	0.417076
2	0.744835	0.271926

برای یک ضریب معین *a*، استراتژی سرمایه‌گذاری پیشنهادی ۱ نسبت شارپ و سورتینو بالاتری نسبت به استراتژی سرمایه‌گذاری ۲ دارد. برای یک استراتژی سرمایه‌گذاری معین، نسبت شارپ و سورتینو قبل از کاهش با افزایش ضریب انتخاب سهام *a* تمایل به افزایش دارد. این همچنین نشان می‌دهد که در سرمایه‌گذاری واقعی، دارایی‌های زیربنایی پرتفوی به‌اندازه بهتر نیست و سرمایه‌گذاران باید تعداد کمی سهام با ارزش سرمایه‌گذاری را از تعداد زیادی دارایی انتخاب کنند. بدیهی است که نتایج فوق همچنین تأیید می‌کند که طرح انتخاب سهام پیشنهادی و مدل بهینه‌سازی پرتفوی بر اساس منابع داده‌های چندگانه می‌تواند عملکرد پرتفوی را بهبود بخشد.

۵- بحث و نتیجه‌گیری

هدف این پژوهش، بهینه‌سازی پرتفوی با استفاده از ادغام *DEA* با منابع داده چندگانه با رویکرد یادگیری ماشین بوده است. به‌منظور دستیابی به این هدف، ابتدا مدل *DEA* را جهت ارزیابی کارایی سهام تحت بررسی ازنظر بازده تاریخی و همبستگی دارایی و انتخاب دارایی‌های زیربنایی پرتفوی طراحی نموده‌ایم. پس از مشخص شدن سهام دارای ارزش سرمایه‌گذاری، از ماشین بردار پشتیبان *SVM* ترکیب‌شده با داده‌های چند منبعی برای پیش‌بینی حرکات قیمت سهام استفاده‌شده است. همچنین روش جستجوی تصادفی و شبکه‌ای را برای تنظیم هایپرپارامترهای الگوریتم ارایه‌شده

مدنظر قرار داده‌ایم. سپس حرکات قیمت سهام و طرح انتخاب سهام پیشنهادی را برای ساخت مدل بهینه‌سازی پرتفوی ترکیب کرده و استراتژی سرمایه‌گذاری مربوطه را برای سرمایه‌گذاران ارایه نمودیم.

در مطالعه تجربی، این پژوهش دارایی‌های زیربنایی شاخص بورس اوراق‌بهادار تهران را به‌عنوان نمونه‌های آزمایشی انتخاب نموده و سپس یک آزمون را در مورد طرح انتخاب سهام پیشنهادی و استراتژی‌های سرمایه‌گذاری انجام داده است. نتایج تجربی نشان‌دهنده این بود که طرح انتخاب سهام پیشنهادی می‌تواند عملکرد استراتژی‌های سرمایه‌گذاری مختلف را بهبود بخشد. علاوه بر آن، نسبت شارپ و سورتینو بازده استراتژی پیشنهادی نیز از استراتژی سرمایه‌گذاری حداقل واریانس سراسری به‌مراتب بهتر است. این نتایج نشان داد که سرمایه‌گذاران باید یک استراتژی سرمایه‌گذاری کم‌تنوع را به‌جای استراتژی سنتی کاملا متنوع اتخاذ کنند. در پایان و به‌عنوان پیشنهادها جهت پژوهش‌های آتی، گفتنی است که به‌کارگیری روش‌های دیگر همچون استفاده از اوزان فازی و به‌دست آوردن داده‌های فازی در مدل *FUZZY DEA*، بررسی روابط بین شاخص‌ها و مقایسه نتایج حاصل از به‌کارگیری این روش‌ها، استفاده از ماشین بردار پشتیبان با چند خروجی جهت پیش‌بینی و مقایسه نتایج آن با مدل ارایه‌شده و چگونگی یافتن اندازه بهینه موجودی سهام که طی آن می‌توان از سایر نظریه‌های پرتفوی و روش‌های حل از قبیل فراابتکاری‌ها استفاده کرد؛ می‌تواند عملکرد استراتژی سرمایه‌گذاری را بهتر نموده و در راستای نزدیکی هر چه بیشتر به شرایط واقعی بازار گام برداشت.

منابع

[1] Markowitz, H. (1952). Portfolio selection. *JSTOR*, *7*(1), 77–91. https://doi.org/10.2307/2975974

[2] Keynes, J. M., Johnson, E. & Moggridge, D. (2012). *The collected writings of John Maynard Keynes*. New York: Cambridge University Press. https://www.amazon.co.za/Collected-Writings-John-Maynard-Keynes/dp/110761046X

[3] Uppal, R., & Wang, T. (2003). Model misspecification and underdiversification. *The journal of finance*, *58*(6), 2465–2486. https://doi.org/10.1046/j.1540-6261.2003.00612.x

[4] Boyle, P., Garlappi, L., Uppal, R., & Wang, T. (2012). Keynes meets Markowitz: The trade-off between familiarity and diversification. *Management science*, *58*(2), 253–272. https://doi.org/10.1287/mnsc.1110.1349

[5] Liu, H. (2014). Solvency constraint, underdiversification, and idiosyncratic risks. *Journal of financial and quantitative analysis*, *49*(2), 409–430. https://doi.org/10.1017/S0022109014000271

[6] Guidolin, M., & Liu, H. (2016). Ambiguity aversion and underdiversification. *Journal of financial and quantitative analysis*, *51*(4), 1297–1323. https://doi.org/10.1017/S0022109016000466

[7] Florentsen, B., Nielsson, U., Raahauge, P., & Rangvid, J. (2019). The aggregate cost of equity underdiversification. *Financial review*, *54*(4), 833–856. https://doi.org/10.1111/fire.12212

[8] Amiri, M., Darestani Farahani, A., & Mahboob-Ghodsi, M. (2017). *Multi-criteria decision-making*. Kian University Press. (**In Persian**). https://B2n.ir/z91916

[9] Edirisinghe, N. C. P., & Zhang, X. (2007). Generalized DEA model of fundamental analysis and its application to portfolio optimization. *Journal of banking & finance*, *31*(11), 3311–3335. https://doi.org/10.1016/j.jbankfin.2007.04.008

[10] Chen, H. H. (2008). Stock selection using data envelopment analysis. *Industrial management & data systems*, *108*(9), 1255–1268. https://doi.org/10.1108/02635570810914928

[11] Skrinjaric, T. (2014). Investment strategy on the zagreb stock exchange based on dynamic DEA. *The institute of economics*, *16*(1), 129–160. https://ideas.repec.org/a/iez/survey/ces-v16_04-2014_skrinjaric.html

[12] Huang, C. Y., Chiou, C. C., Wu, T. H., & Yang, S. C. (2015). An integrated DEA-MODM methodology for portfolio optimization. *Operational research*, *15*, 115–136. https://doi.org/10.1007/s12351-014-0164-7

[13] M Gardijan., & Škrinjarić, T. (2015). Equity portfolio optimization: A DEA based methodology applied to the Zagreb Stock Exchange. *ResearchGate*, *16*(2), 405–417. https://doi.org/10.17535/crorr.2015.0031

[14] Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert systems with applications*, *79*, 153–163. https://doi.org/10.1016/j.eswa.2017.02.041

[15] Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The north American journal of economics and finance*, *47*, 552–567. https://doi.org/10.1016/j.najef.2018.06.013

[16] Mohammadi, S. (2004). Technical analysis in Tehran stock exchange. *Financial research journal*, *6*(1), 97–129. (**In Persian**). https://civilica.com/doc/1395623/

[17] Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the royal statistical society: Series C (applied statistics)*, *29*(2), 119–127. https://doi.org/10.2307/2986296

[18] Branda, M. (2015). Diversification-consistent data envelopment analysis based on directional-distance measures. *Omega*, *52*, 65–76. https://doi.org/10.1016/j.omega.2014.11.004

[19] Choi, H. S., & Min, D. (2017). Efficiency of well-diversified portfolios: Evidence from data envelopment analysis. *Omega*, *73*, 104–113. https://doi.org/10.1016/j.omega.2016.12.008

[20] Lin, R., & Li, Z. (2020). Directional distance based diversification super-efficiency DEA models for mutual funds. *Omega*, *97*, 102096. https://doi.org/10.1016/j.omega.2019.08.003

[21] Xiao, H., Ren, T., & Ren, T. (2020). Estimation of fuzzy portfolio efficiency via an improved DEA approach. *INFOR: information systems and operational research*, *58*(3), 478–510. https://doi.org/10.1080/03155986.2020.1734904

[22] Liu, W., Zhou, Z., Liu, D., & Xiao, H. (2015). Estimation of portfolio efficiency via DEA. *Omega*, *52*, 107–118. https://doi.org/10.1016/j.omega.2014.11.006

[23] Zhou, Z., Jin, Q., Xiao, H., Wu, Q., & Liu, W. (2018). Estimation of cardinality constrained portfolio efficiency VIA segmented DEA. *Omega*, *76*, 28–37. https://doi.org/10.1016/j.omega.2017.03.006

[24] Hosseinzadeh Lotfi, F., Jahanshahloo, G. R., Khodabakhshi, M., Rostamy-Malkhlifeh, M., Moghaddas, Z., & Vaez-Ghasemi, M. (2013). A review of ranking models in data envelopment analysis. *Journal of applied mathematics*, *2013*(1), 492421. https://doi.org/10.1155/2013/492421

[25] Zhou, Z., Xiao, H., Jin, Q., & Liu, W. (2018). DEA frontier improvement and portfolio rebalancing: An application of China mutual funds on considering sustainability information disclosure. *European journal of operational research, elsevier*, *269*(1), 111-131. https://doi.org/10.1016/j.ejor.2017.07.010

[26] Kerstens, K., Mounir, A., & de Woestyne, I. (2011). Non-parametric frontier estimates of mutual fund performance using C-and L-moments: some specification tests. *Journal of banking & finance*, *35*(5), 1190–1201. https://doi.org/10.1016/j.jbankfin.2010.09.030

[27] Lamb, J. D., & Tee, K. H. (2012). Data envelopment analysis models of investment funds. *European journal of operational research*, *216*(3), 687–696. https://doi.org/10.1016/j.ejor.2011.08.019

[28] Michaud, R. O. (1989). The Markowitz optimization enigma: is 'optimized'optimal? *Financial analysts journal*, *45*(1), 31–42. https://doi.org/10.2469/faj.v45.n1.31

[29] DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The review of financial studies*, *22*(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

[30] Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of financial economics*, *99*(1), 204–215. https://doi.org/10.1016/j.jfineco.2010.08.013

[31] Beale, E. M. L., & Forrest, J. J. H. (1976). Global optimization using special ordered sets. *Mathematical programming*, *10*, 52–69. https://doi.org/10.1007/BF01580653

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی