فصلنامه علمی پژوهشی دانش سرمایه‌گذاری
دوره ۱۶ / شماره ٤ (پیاپی ٦٤) / زمستان ۱٤۰۶
صفحه ۲۲۱ تا ۲٥٤

ترکیب الگوریتم‌های فراابتکاری و یادگیری ماشین برای مدیریت سبد سرمایه‌گذاری بر اساس پیش‌بینی میان‌مدت داده‌های شاخص بورس

بهناز بدائی
دانشجوی دکتری مهندسی مالی، دانشگاه آزاد اسلامی واحد تهران جنوب، تهران، ایران
حسین بدیعی
استادیار گروه حسابداری، دانشگاه آزاد اسلامی واحد تهران جنوب، تهران، ایران (نویسنده مسئول)
badiei@iau.ac.ir
محسن حمیدیان
دانشیار گروه حسابداری، دانشگاه آزاد اسلامی واحد تهران جنوب، تهران، ایران
علی امیری
استادیار گروه حسابداری، دانشگاه آزاد اسلامی واحد بندرعباس، بندرعباس، ایران
تاریخ دریافت: ۱۴۰۴/۰۱/۱۹ تاریخ پذیرش: ۱۴۰۴/۰۲/۳۰

چکیده
این مطالعه به ارائه یک چارچوب نوین برای بهینه‌سازی سبد سرمایه‌گذاری با ترکیب الگوریتم‌های یادگیری ماشین و روش‌های فراابتکاری می‌پردازد. پژوهش حاضر با استفاده از داده‌های تاریخی بازار بورس تهران (۱۳۹۰-۱۴۰۲)، سیستم هوشمندی طراحی کرده که قادر است رفتار بازار را پیش‌بینی و سبدهای بهینه سرمایه‌گذاری را تشکیل دهد. در مرحله اول، از مدل‌های یادگیری ماشین شامل رگرسیون بردار پشتیبان (SVR) و جنگل تصادفی (RF) برای پیش‌بینی بازده سهام و شاخص کل استفاده شده است. پارامترهای این مدل‌ها به کمک الگوریتم بهینه‌سازی تجمع ذرات (PSO) تنظیم گردید.در مرحله بعد، مسئله بهینه‌سازی سبد سرمایه‌گذاری به صورت یک مسئله چندهدفه فرمول‌بندی شده و با الگوریتم ژنتیک دوهدفه NSGA-II حل گردید. این الگوریتم با ایجاد جبهه پارتو، امکان دستیابی همزمان به اهداف متعارض حداکثرسازی بازده و حداقل‌سازی ریسک را فراهم آورد. نتایج تجربی نشان داد سبدهای تشکیل شده با این روش بازدهی بالاتری نسبت به شاخص بازار با سطح ریسک قابل قبول ارائه می‌دهند. این پژوهش از سه جهت حائز اهمیت است: نخست نشان می‌دهد ترکیب هوشمندانه یادگیری ماشین و روش‌های فراابتکاری می‌تواند به تصمیم‌گیری‌های سرمایه‌گذاری کارآمدتر منجر شود. دوم، کارایی الگوریتم PSO در تنظیم پارامترهای مدل‌های مالی و توانایی NSGA-II در حل مسائل پیچیده بهینه‌سازی سبد را نشان می‌دهد. سوم، چارچوبی کاربردی برای تحلیلگران مالی و مدیران پرتفوی ارائه می‌دهد که قابلیت تطبیق با شرایط مختلف بازار را دارد.

واژه‌های کلیدی: NSGAII، PSO، بهینه‌سازی پورتفولیو، یادگیری ماشین.

۱- بیان مسئله

در دنیای مالی، سرمایه‌گذاران و مدیران سبد سهام با چالش‌های چندوجهی و پیچیده‌ای مواجه‌اند که ناشی از نوسانات شدید بازار، تغییرات سریع اقتصادی و افزایش رقابت هستند. در این محیط پویا، نیاز به استراتژی‌های هوشمند و قابل انعطاف برای بهینه‌سازی عملکرد سبد سرمایه‌گذاری و مدیریت ریسک‌ها به شدت احساس می‌شود (چنگ[1] و همکاران، ۲۰۲۴). به ویژه، ذینفعان به دنبال ابزارهایی هستند که نه‌تنها بازده بالاتری را ارائه دهند، بلکه در عین حال جلوگیری از ریسک‌های غیرقابل پیش‌بینی و هزینه‌های اضافی را نیز در نظر داشته باشند.

مدل‌سازی سبد سرمایه‌گذاری به عنوان یک مسئله چند هدفه مطرح می‌شود که در آن نیاز به توازن میان اهداف متعارضی مانند حداکثر کردن بازده و حداقل کردن ریسک وجود دارد. از آنجایی که این اهداف به طور مداوم در حال تغییر و تعامل هستند، چالش اصلی در ایجاد مدلی است که بتواند به‌طور هم‌زمان و بهینه به این نیازها پاسخ دهد و در عین حال محدودیت‌های موجود را در نظر بگیرد (اودیمی[2] و همکاران، ۲۰۲۴).

گام‌های مؤثر برای پاسخ به این چالش شامل توسعه یک مدل بهینه‌سازی سبد سرمایه‌گذاری است که بتواند از تکنیک‌های پیشرفته یادگیری ماشین در راستای پیش‌بینی دقیق بازده دارایی‌ها بهره‌برداری کند. با این حال، یکی از مسائل باقی‌مانده در ادبیات موجود، پیچیدگی ادغام مدل‌های پیش‌بینی با الگوریتم‌های بهینه‌سازی چند هدفه است که می‌تواند به‌ویژه در پیش‌بینی روندهای بازار و شاخص‌های کلید اقتصادی مؤثر باشد (گوارینو[3] و همکاران، ۲۰۲۴).

مدیریت سبد سرمایه‌گذاری به دو رویکرد کلی فعال و غیرفعال تقسیم می‌شود. در روش مدیریت فعال، هدف مدیران پرتفوی کسب بازدهی بالاتر از میانگین بازار (مثلا شاخص کل) است. این در حالی است که فرض اساسی در مدیریت فعال این است که بازار به طور کامل کارا نیست و می‌توان از فرصت‌های موجود بهره‌برداری کرد (گارسیا[4] و همکاران، ۲۰۲۴). در این تحقیق، ما با استفاده از الگوریتم‌های یادگیری ماشین پیشرفته، به بررسی پیش‌بینی دقیق بازده شاخص کل و نحوه تأثیر این پیش‌بینی‌ها بر تصمیمات سرمایه‌گذاری خواهیم پرداخت. در واقع، هدف ما توسعه یک چارچوب تحلیلی است که از پیش‌بینی‌های خود برای مقایسه بازده مورد انتظار شاخص کل با اوراق بهادار با درآمد ثابت استفاده کند. با توجه به این که سرمایه‌گذاران تمایل دارند در سناریوهایی که پیش‌بینی‌ها نشان می‌دهند بازده بالاتری نسیب آن‌ها می‌شود اقدام به سرمایه‌گذاری کنند، ما این مدل را بر اساس یک الگوریتم بهینه‌سازی

[1] Cheng
[2] Odeyemi
[3] Guarino
[4] García

چندهدفه طراحی خواهیم کرد. اگر بازده پیش‌بینی شده برای شاخص کل از بازده اوراق بهادار با درآمد ثابت بیشتر باشد، بهینه‌سازی سبد با استفاده از الگوریتم‌های مناسب انجام خواهد شد؛ در غیر این صورت، سرمایه‌گذاری در اوراق با درآمد ثابت پیشنهاد می‌شود.

این تحقیق به دنبال آن است که با ارائه روشی جامع، شکاف‌های موجود در ادبیات بهینه‌سازی سبد را پر کند و با ترکیب مدل‌های پیش‌بینی پیشرفته با تکنیک‌های بهینه‌سازی چند هدفه، ابزاری جدید برای سرمایه‌گذاران ایجاد نماید که توانایی مدیریت مؤثر ریسک‌ها و بازده را بهبود بخشد.

از منظر علمی، نقطه قوت این پژوهش در این است که بر روی ادغام مدل‌سازی پیش‌بینی با بهینه‌سازی تمرکز دارد و به تحلیل دقیق‌تر الگوهای پیچیده و روابط پنهان در داده‌ها می‌پردازد. این تحقیق همچنین از منظر کاربردی به مدیران سبد سهام کمک خواهد کرد تا با استفاده از پیش‌بینی‌های بهبود یافته، تصمیمات مالی بهتری اتخاذ کنند و در چشم‌انداز نوسان‌دار بازار، به عملکرد بهتری دست یابند. با تحلیل و بررسی عمیق این موضوع، این مطالعه می‌تواند به عنوان یک پایه علمی و عملی برای نتایج بهینه‌سازی سبد سرمایه‌گذاری در دوران پرتلاطم بازار عمل کند و راهکارهای قابل اعتمادی برای ایجاد ارزش افزوده ارائه دهد.

۱-۱- اهداف تحقیق

هدف اصلی

مدیریت سبد چندهدفه بر مینای پیش‌بینی بازده سهام و شاخص کل با استفاده از مدل‌های یادگیری ماشین بهینه‌سازی شده

اهداف فرعی

۱) ارزیابی مدل‌های یادگیری ماشین در پیش‌بینی شاخص کل و بازده سهام
۲) مقایسه بازده استراتژی معاملاتی با استفاده از مدل‌های یادگیری ماشین و بهینه‌سازی چندهدفه در بورس تهران با بازدهی پورتفولیو با اوزان برابر

۲- مبانی نظری و مرور ادبیات

۲-۱- بهینه‌سازی میانگین-واریانس

بهینه‌سازی میانگین-واریانس (MVO[1])، که توسط مارکوویتز در سال ۱۹۵۲ معرفی شد، یک نظریه اساسی در امور مالی مدرن برای ساخت پرتفوی سرمایه‌گذاری بهینه است. ایده اصلی این است که با انتخاب ترکیبی از دارایی‌هایی که بازده مورد انتظار را برای سطح معینی از ریسک به حداکثر می‌رساند یا ریسک را برای سطح معینی از بازده مورد انتظار به حداقل می‌رساند، تعادل بین ریسک و بازده ایجاد شود (لوین[2] و همکاران، ۲۰۱۷). در اینجا توضیح مفصلی از MVO آورده شده است:

۱. بازده مورد انتظار:

بازده مورد انتظار یک پرتفوی مجموع وزنی از بازده مورد انتظار دارایی‌های منفرد در پرتفوی است و از نظر ریاضی به صورت زیر بیان می‌شود:

$$E(R_p) = \sum_{i=1}^{m} w_i * E(R_i) \quad (۱)$$

جایی که:

- $E(R_p)$ بازده مورد انتظار سبد است.
- w_i وزن دارایی i در پورتفولیو است.
- $E(R_i)$ بازده مورد انتظار دارایی i است.

۲. واریانس پرتفوی:

واریانس (به عنوان یک معیار ریسک) یک پرتفوی نه تنها به واریانس دارایی‌های فردی بستگی دارد، بلکه به کوواریانس بین آنها نیز بستگی دارد. واریانس پورتفولیو توسط عبارت زیر اندازه‌گیری می‌شود:

$$Var(R_p) = \sum_{i=1}^{m} \sum_{j \geq i}^{m} w_i * w_j * Cov(R_i, R_j) \quad (۲)$$

جایی که:

- $Var(R_p)$ واریانس پورتفولیو است.
- w_i و w_j به ترتیب وزن دارایی‌های i و j در پورتفولیو هستند.

[1] Mean-Variance Optimization
[2] Lwin

- $\mathrm{Cov}(\mathrm{R_i}, \mathrm{R_j})$ کوواریانس بین بازده دارایی‌های i و j است.

۳. کوواریانس:

کوواریانس نحوه حرکت دو دارایی با هم را اندازه گیری می‌کند. به صورت زیر محاسبه می‌شود:

$$Cov(R_i, R_j) = \frac{\sum_{t=1}^{N}\left[\left(R_{i,t} - E(R_i)\right) * \left(R_{j,t} - E(R_j)\right)\right]}{(N-1)} \qquad (۳)$$

جایی که:

- $R_{i,t}$ و $R_{j,t}$ بازده دارایی‌های i و j در زمان t هستند.
- $E(R_i)$ و $E(R_j)$ بازده مورد انتظار دارایی‌های i و j هستند.
- N تعداد مشاهدات است.

۲-۲- الگوریتم ژنتیک با مرتب‌سازی نامغلوب

الگوریتم[1] NSGA-II برای اولین بار توسط دب[2] و همکاران (۲۰۰۲) معرفی شد و به عنوان یکی از قوی‌ترین و مؤثرترین الگوریتم‌های بهینه‌سازی چندهدفه شناخته می‌شود. این الگوریتم تلاش می‌کند تا طیف وسیعی از راه‌حل‌های بهینه را که نقاط پارتو[3] نامیده می‌شوند، تولید کند. نقاط پارتو به نقاطی گفته می‌شوند که در آن‌ها هیچ یک از اهداف نمی‌تواند بهبود یابد بدون اینکه حداقل یکی از اهداف دیگر بدتر شود. NSGA-II می‌تواند با استفاده از جمعیتی از راه‌حل‌ها فعالیت کند و از تکرارهای تکاملی (با استفاده از عملگرهایی مانند تقاطع و جهش) برای بهبود این جمعیت با گذشت زمان استفاده کند.

یکی از ویژگی‌های کلیدی الگوریتم NSGA-II روش مرتب‌سازی نامغلوب است. این روش به این صورت عمل می‌کند که ابتدا جمعیت را به زیر مجموعه‌هایی تقسیم می‌کند. هر زیر مجموعه شامل راه‌حل‌هایی است که هیچ کدام بر دیگری برتری ندارند. پس از آن، این ساختار نامغلوب به ما این امکان را می‌دهد که تعیین کنیم کدام راه‌حل‌ها در سطح اول، سطح دوم و غیره قرار دارند. به علاوه، برای حفظ تنوع در جمعیت، NSGA-II یک معیار به نام فاصله ازدحامی[4] را به کار می‌گیرد. این معیار به هر فرد

[1] Non-dominated Sorting Genetic Algorithm II
[2] Deb
[3] Pareto
[4] crowding distance

در جمعیت یک مقدار فاصله ازدحامی اختصاص می‌دهد که به الگوریتم کمک می‌کند تا از تمرکز بر روی یک ناحیه کوچک از فضای حل جلوگیری و تنوع را در نواحی مختلف حفظ کند (چن[1] و همکاران، ۲۰۲۵).

این الگوریتم نیاز به تنظیم پارامتر را کاهش می‌دهد و به جای آن بر روی انتخاب طبیعی و تنوع جمعیت تأکید می‌کند. از آنجا که NSGA-II یک روش مبتنی برایجاد جمعیت است، این الگوریتم به خوبی در مسائل با ابعاد بالا و ساختار پیچیده عمل می‌کند و به همین دلیل، در حوزه‌های گوناگونی همچون مهندسی، طراحی سیستم‌ها، مسائل مدیریت منابع و برنامه‌ریزی به طور گسترده‌ای مورد استفاده قرار می‌گیرد. به طور کلی، NSGA-II به عنوان یک ابزار بسیار کارآمد و انعطاف‌پذیر برای حل مشکلات بهینه‌سازی چندهدفه شناخته شده است و اهمیت قابل توجهی در تحقیقات علمی و کاربرد صنعتی دارد.

۲-۳- مدل‌های یادگیری ماشین

۲-۳-۱- رگرسیون بردار پشتیبان

رگرسیون بردار پشتیبان (SVR[2]) بر اساس همان اصولی است که ماشین‌های بردار پشتیبان (SVM[3]) برای طبقه‌بندی دارند. در SVM، ما سعی می‌کنیم یک ابر صفحه پیدا کنیم که نقاط داده را با حداکثر حاشیه به کلاس‌های مختلف تفکیک کند. از سوی دیگر، هدف SVR یافتن ابرصفحه ای است که بهترین تناسب با نقاط داده را در یک رگرسیون دارد (آواد[4] و همکاران، ۲۰۱۵).

ایده کلیدی پشت SVR یافتن یک ابر صفحه در فضای ویژگی است که نزدیک به نقاط داده باشد اما همچنان یک حاشیه تلورانس ε را حفظ کند. ابر صفحه با معادله $f(x) = w^T * x + b$ تعریف می‌شود که w بردار وزن، x بردار ویژگی ورودی و b عبارت بایاس است. در SVR، می‌خواهیم انحرافات خروجی پیش‌بینی‌شده $f(x)$ را از خروجی واقعی y به حداقل برسانیم، در حالی که همچنان آنها را در حاشیه تحمل ε نگه داریم. این با یافتن بردار وزن بهینه w و عبارت بایاس b که تابع هدف را به حداقل می‌رساند به دست می‌آید:

$$\left(\frac{1}{2}\right) * ||w||^2 + C * \sum_{i=1}^{m} (\max(0, |y_i - f(x_i)| - \varepsilon)) \qquad (۱۲)$$

[1] Chen
[2] Support Vector Regression
[3] Support Vector Machine
[4] Awad

که در آن $||w||^2$ نرم بردار وزن w را نشان می‌دهد، C یک ابرپارامتر است که مبادله بین پیچیدگی مدل و خطاها را کنترل می‌کند و انحراف خروجی‌های پیش‌بینی‌شده از خروجی‌های واقعی را جریمه می‌کند. عبارت $\sum_{i=1}^{m}(\max(0, |y_i - f(x_i)| - \varepsilon))$ تضمین می‌کند که خروجی پیش‌بینی شده در حاشیه تلورانس ε برای همه نقاط داده قرار دارد و نقاط داده خارج از این حاشیه به تابع زیان کمک می‌کنند و جریمه می‌شود. برای حل مسئله بهینه‌سازی، SVR از روش لاگرانژ برای تبدیل مسئله به شکل دوگان[1] آن استفاده می‌کند، که امکان بهینه‌سازی کارآمد را با استفاده از یک الگوریتم برنامه‌ریزی درجه دوم فراهم می‌کند. هنگامی که بردار وزن بهینه w و عبارت بایاس b تعیین شد، می‌توانیم از تابع $f(x)$ برای پیش‌بینی داده‌های ورودی جدید استفاده کنیم.

۲-۳-۲- جنگل تصادفی

جنگل تصادفی[2] یک الگوریتم یادگیری گروهی است که می‌تواند برای کارهای طبقه بندی و رگرسیون استفاده شود. در زمینه رگرسیون، رگرسیون جنگل تصادفی درخت‌های تصمیم گیری متعددی را در طول آموزش ایجاد می‌کند و پیش‌بینی‌های آنها را برای پیش‌بینی نهایی تجمیع می‌کند. هر درخت در جنگل تصادفی بر روی یک زیرمجموعه نمونه‌برداری شده تصادفی متفاوت از داده‌ها و ویژگی‌های آموزشی آموزش داده می‌شود و عنصر تصادفی را به مدل اضافه می‌کند (سدورسکی[3]، ۲۰۲۱).

هر درخت تصمیم در جنگل تصادفی با استفاده از زیرمجموعه ای از داده‌ها و ویژگی‌های آموزشی ساخته می‌شود. این فرآیند شامل تقسیم بازگشتی داده‌ها به زیر مجموعه‌ها بر اساس ویژگی‌ها برای ایجاد یک ساختار درختی است که متغیر هدف را پیش‌بینی می‌کند. نقاط تقسیم بر اساس معیارهایی مانند به حداقل رساندن واریانس متغیر هدف در هر زیر مجموعه انتخاب می‌شوند.

جنگل تصادفی با نمونه‌برداری از داده‌های آموزشی و ویژگی‌های هر درخت، تصادفی بودن را لحاظ می‌کند. این فرآیند به کاهش بیش برازش[4] و افزایش توانایی تعمیم مدل کمک می‌کند. به طور معمول، هر درخت بر روی یک نمونه بوت استرپ (نمونه گیری با جایگزینی) از داده‌های آموزشی و یک زیر مجموعه تصادفی از ویژگی‌ها برای تقسیم در هر گره آموزش داده می‌شود.

فرض کنید یک جنگل تصادفی با N درخت تصمیم داریم که با $f_{i(x)}$ برای $i = 1$ تا N نشان داده شده است. در این حالت برای تجمیع داریم (که در آن $f_{i(x)}$ پیش‌بینی درخت تصمیم i–ام است):

[1] Dual
[2] Random Forest
[3] Sadorsky
[4] Overfitting

$$y_{hat} = \left(\frac{1}{N}\right) * \sum_{i=1}^{N} f_{i(x)} \quad (۱۳)$$

۲-۴- الگوریتم بهینه‌سازی تجمع ذرات برای تنظیم پارامترهای مدل‌های یادگیری ماشین

الگوریتم بهینه‌سازی ازدحامی ذرات (PSO) به عنوان یک روش جستجو و بهینه‌سازی معاصر، به شکل‌گیری خود از رفتار اجتماعی موجودات زنده، نظیر پرندگان در هنگام جستجو برای غذا و یا ماهی‌های درون گروه، الهام گرفته است. این الگوریتم با ملاحظه رفتار جمعی، یک گروه از ذرات را در نظر می‌گیرد که به طور همزمان در فضای جستجو حرکت می‌کنند و با توجه به اطلاعات حاصل از بهترین موقعیت‌های خود و بهترین موقعیت گروهی، به روز رسانی می‌شوند. هر ذره در هر مرحله به سمت هدف، به روز می‌شود تا به بهینه‌ترین پاسخ ممکن برسد (پرادهان[1] و همکاران، ۲۰۲۵).

در این الگوریتم، هر ذره دارای یک موقعیت (که به عنوان یک راه‌حل کاندید شناخته می‌شود) و یک سرعت (که تغییراتی که برای حرکت به سمت موقعیت جدید باید داشته باشد را مشخص می‌کند) می‌باشد. به روز رسانی موقعیت و سرعت به صورت زیر انجام می‌شود:

۱) **به‌روزرسانی سرعت**: سرعت هر ذره به‌روز می‌شود تا به بهترین موقعیت شناسایی شده‌ی خود و بهترین موقعیت گروهی نزدیک‌تر شود. فرمول‌های اساسی برای به‌روزرسانی سرعت به صورت زیر هستند

۲)

$$v_{\{i\}(t+1)} = w \cdot v_{\{i\}(t)} + c_1 \cdot r_1 \cdot \left(p_{\{i\}} - x_{\{i\}(t)}\right) + c_2 \cdot r_2 \cdot \left(g - x_{\{i\}(t)}\right)$$

که در آن $v_{\{i\}}$ سرعت ذره i ، $x_{\{i\}}$ موقعیت ذره، $p_{\{i\}}$ بهترین موقعیت شخصی موجود، g بهترین موقعیت گروهی و c_1 و c_2 ضرایب یادگیری برای شناسایی بهترین موقعیت‌ها هستند. w یک ضریب اینرسی است که می‌تواند بر روی تنوع و تمرکز جستجو تأثیر بگذارد و (r_1) و (r_2) اعداد تصادفی بین ۰ و ۱ هستند.

۳) **به‌روزرسانی موقعیت**: پس از به‌روزرسانی سرعت، موقعیت هر ذره نیز بر اساس سرعت جدید به‌روز می‌شود

$$x_{\{i\}(t+1)} = x_{\{i\}(t)} + v_{\{i\}(t+1)}$$

[1] Pradhan

این الگوریتم به سرعت در میان جامعه تحقیقاتی به محبوبیت رسید و در بسیاری از زمینه‌ها از جمله طراحی شبکه، بهینه‌سازی فعالیت‌های تولیدی، مسائل مالی و علوم داده مورد استفاده قرار گرفت. این الگوریتم به خاطر سادگی، کارایی و قابلیت تطبیق‌پذیری‌اش در برابر مسائل مختلف به مراتب بهتر از بسیاری از روش‌های کلاسیک بهینه‌سازی عمل می‌کند.

الگوریتم PSO به عنوان یک روش موثر برای تنظیم پارامترهای مدل‌های یادگیری ماشین شناخته می‌شود. در این رویکرد، هر ذره نماینده یک مجموعه از پارامترهای مدل مثل وزن‌ها، ضریب‌های یادگیری و سایر تنظیمات است. با شروع از یک جمعیت تصادفی از این ذرات، PSO به صورت تکراری عملکرد مدل را با استفاده از پارامترهای معرفی شده توسط هر ذره ارزیابی می‌کند و تعیین می‌کند که کدام ترکیب از پارامترها بهترین عملکرد را داشته است. بهترین موقعیت شخصی هر ذره و بهترین موقعیت گروهی به‌روز می‌شود تا فرآیند جستجو برای یافتن بهترین پارامترها در فضای جستجو تسهیل شود (داوریان[1] و همکاران، ۲۰۲۵).

مزیت اصلی PSO در تنظیم پارامترهای مدل‌های یادگیری ماشین این است که می‌تواند به‌طور همزمان ناحیه‌های وسیع‌تری از فضای پارامتری را جستجو کند و از اطلاعات سایر ذرات برای ارتقاء جستجو بهره‌برداری کند. این الگوریتم به دلیل سادگی و قابلیت تطبیق‌پذیری در زمینه‌های مختلف یادگیری ماشین، از جمله یادگیری عمیق و بهینه‌سازی مدل‌های پیچیده، کاربرد زیادی دارد.

۲-۵- مرور ادبیات

۲-۵-۱- پیشینه خارجی

مدل بهینه‌سازی پورتفولیو به دلیل مسائل تخمینی در داده‌های واقعی تأثیر محدودی دارد، اما بن[2] و همکاران (۲۰۱۸) دو روش منظم‌سازی و اعتبارسنجی متقابل را برای رفع این مشکل معرفی کرده‌اند. آن‌ها منظم‌سازی مبتنی بر عملکرد (PBR) را ارائه می‌دهند که واریانس‌های نمونه ریسک و بازده پرتفوی را محدود می‌کند و به حل مسائلی همچون میانگین-واریانس و میانگین CVaR-کمک می‌کند. در نهایت، نتایج نشان می‌دهد که PBR در مقایسه با دیگر معیارها بر اساس داده‌های فاما-فرنچ برتری دارد.
ادغام پیش‌بینی بازده مدل‌های سری زمانی سنتی در تشکیل پورتفولیو می‌تواند عملکرد مدل بهینه‌سازی سبد اصلی را بهبود بخشد.

[1] Daviran
[2] Ban

با توجه به برتری مدل‌های یادگیری ماشین و یادگیری عمیق نسبت به مدل‌های سری زمانی، مقاله ما[1] و همکاران (۲۰۲۱) پیش‌بینی بازده در تشکیل پورتفولیو را با استفاده از دو مدل یادگیری ماشین (جنگل تصادفی و رگرسیون برداری پشتیبان) و سه مدل یادگیری عمیق (شبکه عصبیLSTM ، پرسپترون چند لایه عمیق و شبکه عصبی کانولوشن) ترکیب می‌کند. این مقاله به‌طور خاص مدل‌های پیش‌بینی را برای پیش‌انتخاب سهام قبل از تشکیل پورتفولیو به کار می‌برد و نتایج آن‌ها را در مدل‌های بهینه‌سازی پورتفولیو امگا و میانگین واریانس ترکیب می‌کند. نتایج تجربی نشان می‌دهد که مدل‌های MVو امگا با پیش‌بینی بازده RF عملکرد بهتری نسبت به مدل‌های دیگر دارند و به سرمایه‌گذاران توصیه می‌کند که پورتفولیو MVF را با پیش‌بینی بازده RF برای سرمایه‌گذاری‌های روزانه تشکیل دهند.

در مقاله چن[2] و همکاران (۲۰۲۱)، یک رویکرد ترکیبی برای ساخت پورتفوی با استفاده از مدل یادگیری ماشین (XGBoost) و الگوریتم کرم شب تاب برای پیش‌بینی قیمت سهام و مدل میانگین واریانس برای انتخاب پورتفوی توسعه داده شده است. نتایج نشان می‌دهد که این روش پیشنهادی از نظر بازده و ریسک نسبت به روش‌های سنتی برتری دارد.

مطالعه پایوا و همکاران (۲۰۱۹) به پیش‌بینی بازده سهام در بازارهای مالی پرداخته و یک مدل تصمیم‌گیری منحصر به فرد برای سرمایه‌گذاری‌های معاملاتی روزانه پیشنهاد کرده است. این مدل با استفاده از یک رویکرد ترکیبی که شامل طبقه‌بندی‌کننده یادگیری ماشین (SVM) و مدل میانگین-واریانس (MV) برای انتخاب پورتفوی است، ارزیابی شده است. با استفاده از داده‌های بورس اوراق بهادار سائوپائولو، این مدل و دو مدل مقایسه‌ای دیگر، یعنی SVM+1/N و تصادفیMV+، تحت شرایط مختلف آزمایش شده‌اند. نتایج نشان می‌دهد که مدل پیشنهادی عملکرد قابل‌توجهی دارد و قابلیت‌های یادگیری ماشین را در پیش‌بینی قیمت سهام گسترش می‌دهد.

در مطالعه وو[3] و همکاران (۲۰۲۲) ساده‌سازی یک پورتفولیو با استفاده از روش‌ها و عناصر یادگیری ماشین، و همچنین یک استراتژی برای ایجاد پورتفولیو که بر کاهش ریسک یک پورتفولیو به تعهدات عامل ریسک متکی است، مورد بررسی قرار می‌گیرد. نتایج نشان میدهد روش‌های یادگیری ماشین ساختارهای کوواریانس و وزن پورتفولیو را ایجاد می‌کنند که ارزیابی آنها دشوار است. پرتفوی با کمترین تغییر در به حداقل رساندن ریسک از معیارهای ساده تر بهتر عمل می‌کند. در طول دوره‌های بی‌ثباتی بالا، بازدهی تعدیل‌شده بر اساس ریسک وجود دارد و این اثرات برای سرمایه‌گذارانی که حساسیت بیشتری نسبت به تغییرات شانسی در بازده دارند، تقویت می‌شود.

[1] Ma
[2] Chen
[3] Wu

مقاله سن و دوتا[1] (۲۰۲۳) یک رویکرد سیستماتیک برای طراحی سبد با استفاده از دو رویکرد بهینه‌سازی ریسک، الگوریتم برابری ریسک سلسله مراتبی و الگوریتم سهم ریسک مساوی سلسله مراتبی در هفت بخش مهم و سهام NIFTY 50 فهرست‌شده در بازار سهام هند ارائه می‌کند. پرتفوی‌ها بر اساس قیمت‌های تاریخی سهام از ۱ ژانویه ۲۰۱۶ تا ۳۱ دسامبر ۲۰۲۰ ساخته شده اند. عملکرد پرتفوی بر اساس داده‌های آزمایشی از ۱ ژانویه ۲۰۲۱ تا ۱ نوامبر ۲۰۲۱ ارزیابی می‌شود. نتایج بک تست پرتفوی‌ها نشان می‌دهد که عملکرد پورتفولیو الگوریتم برابری ریسک سلسله مراتبی نسبت به پورتفولیو الگوریتم سهم ریسک مساوی سلسله مراتبی در آموزش و داده‌های آزمون برای اکثر بخش‌های مورد مطالعه در این کار برتر است.

در مقاله بهارا[2] و همکاران (۲۰۲۳) به عنوان یک تکنیک منحصربه‌فرد ساخت پورتفولیو ارائه شده است. مدل‌های رگرسیون یادگیری ماشین مانند جنگل تصادفی، تقویت گرادیان پیشرفته (XGBoost)، تقویت تطبیقی (AdaBoost)، رگرسیون ماشین بردار پشتیبان (SVR)، k نزدیک‌ترین همسایه‌های (KNN) و شبکه عصبی مصنوعی (ANN) برای پیش‌بینی سهام استفاده می‌شوند. ارزش‌های دوره بعدی سهام با بازدهی بیشتر در مرحله اول انتخاب می‌شوند. علاوه بر این، مدل بهینه‌سازی پرتفوی میانگین-VaR برای انتخاب پورتفولیو در مرحله دوم استفاده می‌شود. مجموعه داده‌های ماهانه بورس اوراق بهادار بمبئی هند، بورس اوراق بهادار توکیو ژاپن و بورس شانگهای چین، به عنوان نمونه تحقیق استفاده می‌شود و یافته‌ها نشان می‌دهد که مدل میانگین-VaR با پیش‌بینی AdaBoost از سایر مدل‌ها بهتر عمل می‌کند.

هباب و کامپوریدیس[3] (۲۰۲۴) ارزیابی کاملی از ۵ الگوریتم یادگیری ماشین (ML) از جمله رگرسیون خطی حداقل مربعات معمولی (LR)، رگرسیون بردار پشتیبانی (SVR)، k-نزدیکترین همسایگان (KNN)، تقویت گرادیان پیشرفته (XGBoost). و شبکه‌های عصبی حافظه کوتاه‌مدت بلند (LSTM)، و همچنین سایر معیارهای مالی مانند هموارسازی نمایی هولت (HES)، مثلثاتی، تبدیل باکس-کاکس، خطاهای ARMA، روند، و مؤلفه‌های فصلی و میانگین متحرک یکپارچه رگرسیون خودکار (ARIMA) انجام داده اند. آنها این الگوریتم‌ها را برای پیش‌بینی قیمت‌های آینده برای ۳۰ REIT از ایالات متحده، بریتانیا و استرالیا و همچنین ۳۰ سهام و ۳۰ اوراق قرضه به کار برده اند. سپس دارایی‌ها به عنوان بخشی از پورتفولیو مورد استفاده قرار گرفتند که با استفاده از الگوریتم ژنتیک آن را بهینه کردند. نتایج نشان داد که استفاده از الگوریتم‌های ML برای پیش‌بینی قیمت حداقل سه برابر بازدهی نسبت به مدل‌های معیار ارائه می‌کند و ریسک را تقریباً دو برابر کاهش می‌دهد.

[1] Sen & Dutta
[2] Behera
[3] Habbab & Kampouridis

مطالعه جها[1] و همکاران (۲۰۲۵) به‌دنبال بهبود دقت مدل‌سازی مالی در بازارهای پر نوسان سهام و ارز خارجی از طریق ترکیب بهینه‌سازی ازدحامی ذرات (PSO) با مدل‌های کلیدی یادگیری ماشین است. هدف، دستیابی به تعادل و دقت بین پیچیدگی مدل و کیفیت تناسب آن است و تأثیرات مثبت تکنیک‌های بهینه‌سازی پیشرفته را در این زمینه نشان می‌دهد. آزمایش‌های تجربی بر روی مدل‌های مهمی مانند ماشین‌های بردار پشتیبان (SVM)، درخت تصمیم و جنگل تصادفی در یک چارچوب جامع با PSO انجام شد و تأثیر PSO بر معیارهای دقت، قبل و بعد از پیاده‌سازی PSO مورد بررسی قرار گرفت. نتایج نشان می‌دهد که ترکیب PSO با مدل‌های یادگیری ماشین دقت پیش‌بینی برای حجم کلی ماهانه بازارهای سهام و ارز را بهبود می‌بخشد.

۲-۵-۲- پیشینه داخلی

تهرانی و همکاران (۱۳۹۴) رویکرد جدیدی برای مدیریت فعال پرتفوی با استفاده از الگوریتم ژنتیک و ترکیب مدل‌های پیش‌بینی شبکه عصبی و نزدیک‌ترین همسایگی ارائه کردند. این رویکرد با پیش‌بینی قیمت سهام و تعریف استراتژی معاملاتی بر اساس قیمت‌های پیش‌بینی‌شده، شامل خرید، فروش، و عدم اقدام به معامله است. نتایج نشان داد که پرتفوی بهینه‌سازی‌شده با روش شبکه عصبی مبتنی بر الگوریتم ژنتیک بالاترین بازده را در مقایسه با رویکرد مدیریت غیر فعال پرتفوی داشت.

میرعلوی و همکاران (۱۳۹۸) با استفاده از روش‌های فراابتکاری و شبکه‌های عصبی مصنوعی مدلی جهت پیش‌بینی قیمت سهام ارائه دادند و نتیجه گرفتند که شبکه عصبی پرسپترون نسبت به مدل‌های دیگر در پیش‌بینی شاخص سهام خطای کمتری دارد.

شبان و همکاران (۱۳۹۹) به بررسی سرایت‌پذیری شاخص کل بورس اوراق بهادار تهران از قیمت دارایی‌های موازی با استفاده از شبکه عصبی پویا پرداختند که در آن قیمت سکه، قیمت مسکن، قیمت نفت خام و نرخ دلار به عنوان متغیرهای ورودی لحاظ شده‌اند. نتایج نشان داد که با حداکثر دو وقفه زمانی، شاخص کل بورس از این قیمت‌ها سرایت‌پذیری دارد و این امر کارایی ضعیف بازار اوراق بهادار تهران را نشان می‌دهد. همچنین، شبکه عصبی طراحی‌شده قادر است شاخص کل بورس و نوسانات آتی آن را پیش‌بینی کند.

هدف مقاله کرامتی و همکاران (۱۴۰۱) بررسی کاربرد معاملات الگوریتمی با استفاده از رویکرد یادگیری تقویتی برای بهینه‌سازی پرتفوی سهام‌های منتخب است. این پژوهش بر روی داده‌های پنج شرکت بورسی از جامعه آماری ۶۷۲ شرکت طی دوره ۱۳۹۶-۱۴۰۰ انجام شد و نتایج آن نشان داد که

[1] Jha

رویکرد یادگیری تقویتی به طور معناداری در هر دو بازار صعودی و نزولی نسبت به استراتژی خرید و نگهداری عملکرد بهتری دارد. به طور کلی، این تحقیق به تأکید بر سودآوری بالاتر رویکرد یادگیری تقویتی نسبت به روش‌های سنتی خرید و نگهداری پرداخته و پیشنهاد می‌کند که از این روش در بهینه‌سازی پرتفوی استفاده شود.

هدف پژوهش اقتصاد و محمدی (۱۴۰۲) کمینه‌سازی خطای پیش‌بینی در تشکیل سبد سرمایه‌گذاری و تأکید بر اهمیت قابل پیش‌بینی بودن گزینه‌های سرمایه‌گذاری است. این مطالعه با ادغام مدل‌های یادگیری ماشین (جنگل تصادفی) و یادگیری عمیق (LSTM) نشان می‌دهد که بهینه‌سازی میانگین واریانس با پیش‌بینی بازدهی بهتر عمل می‌کند و داده‌های تاریخی ۵ ساله از صنایع مختلف مورد تحلیل قرار گرفته است.

هدف مقاله علیزاده و کیانفر (۱۴۰۲) توسعه مدل مارکوویتز به‌گونه‌ای است که با شرایط واقعی بازار بیشتر تطابق داشته باشد و از عوامل تحلیل بنیادی و محدودیت‌های بازار برای تخمین ریسک استفاده کند. این پژوهش نشان می‌دهد که استفاده از معیارهای مختلف مانند میانگین نیمه‌واریانس و الگوریتم جستجوی‌هارمونی به بهبود عملکرد مدل و انتخاب سهام در سبد سرمایه‌گذاری کمک می‌کند.

هدف مطالعه رستمی و سهیلی (۱۴۰۲) پیش‌بینی شاخص بورس با استفاده از مدل پویای میانگین‌گیری و روش مدل پویای انتخابی است؛ این تحقیق با داده‌های فصلی سال‌های ۱۳۸۰-۱۳۹۹ و نرم‌افزار متلب انجام شده است. مزیت اصلی این مدل، قابلیت ورود تعداد زیادی متغیر مستقل بدون ایجاد مشکل برازش بیش‌ازحد است. نتایج نشان می‌دهد که متغیرهای رشد حجم پول، رشد شبه پول، تورم و رشد شاخص قیمت زمین در شهرهای بزرگ تأثیر بیشتری بر تخمین بازده سهام بورس دارند.

پژوهش رضاشاطری و همکاران (۱۴۰۳) با هدف ارائه الگوی بهینه سبد سهام از طریق محدودیت تسلط تصادفی و کاهش ریسک گریزی مطلق انجام شد. نمونه آماری بورس اوراق بهادار و نوع داده های گردآوری شده از آن سری زمانی تغییرات و تغییرات تجمعی شاخص کل بورس اوراق بهادار تهران در سال ۱۳۹۰ تا ۱۴۰۰ است. سبد سهام براساس اطلاعات ۵۰ شرکت مورد بررسی قرار گرفته است. در ابتدا این داده ها مورد بررسی شده، سپس الگوریتم های مورد نظر طراحی گردید و مدل براساس شرایط و ازمون های لازم تجزیه و تحلیل شد. کارایی براساس ازمون های معیار شارپ، معیار ترینر، معیار سورتینو و معیار امگا ارزیابی شد. در ادامه سبد سهام مورد بررسی براساس فروض اولیه تعیین شده و براساس بازدهی شرکت ها مورد الگوسازی و سنجش قرار گرفت.

۲-۶- شکاف پژوهشی و نوآوری پژوهش

در این پژوهش، نوآوری اصلی در طراحی یک چارچوب ترکیبی و هوشمند برای مدیریت سبد سرمایه‌گذاری نهفته است که از تلفیق الگوریتم‌های یادگیری ماشین (مانند SVR وRF) با روش‌های بهینه‌سازی فراابتکاری (شامل PSO وNSGA-II) بهره می‌برد. برخلاف مطالعات پیشین که عمدتاً یا به پیش‌بینی بازده دارایی‌ها پرداخته‌اند یا بهینه‌سازی سبد را به صورت ایستا انجام داده‌اند، این تحقیق با تمرکز بر پیش‌بینی میان‌مدت شاخص کل و بازده سهام، داده‌های خروجی مدل‌های یادگیری ماشین را به‌صورت مستقیم در فرایند بهینه‌سازی سبد به کار می‌گیرد. این ادغام هوشمندانه، امکان تصمیم‌گیری پویا و سناریومحور را فراهم می‌سازد؛ به‌گونه‌ای که بسته به پیش‌بینی بازده شاخص کل، سرمایه‌گذاری در سهام یا اوراق با درآمد ثابت پیشنهاد می‌شود.

نوآوری دیگر این پژوهش در نحوه تنظیم و بهینه‌سازی مدل‌ها نهفته است؛ به‌طوری که الگوریتم PSOبرای انتخاب بهینه پارامترهای مدل‌های یادگیری ماشین به‌کار گرفته شده و NSGA-II نیز برای حل مسئله چندهدفه‌ی بهینه‌سازی سبد استفاده شده است. این رویکرد ترکیبی، مزایای تحلیل‌پذیری دقیق روابط پیچیده بین ریسک و بازده را با قابلیت جست‌وجوی جهانی الگوریتم‌های فراابتکاری در هم می‌آمیزد. چارچوب پیشنهادی نه‌تنها از نظر دقت پیش‌بینی و کیفیت سبدهای بهینه برتری دارد، بلکه از منظر عملی نیز قابل پیاده‌سازی در محیط واقعی بازار است و ابزار تصمیم‌سازی نوینی در اختیار تحلیلگران مالی و مدیران پرتفوی قرار می‌دهد که متناسب با تغییرات سریع بازار قابل تنظیم است.

۳- روش پژوهش

تحقیق حاضر به‌عنوان یک مطالعه توسعه‌ای و کاربردی، بر روی توسعه مدل‌ها و روش‌های جدیدی که در عمل قابل استفاده باشند، تمرکز دارد. منابع داده این تحقیق شامل پایگاه‌های اطلاعاتی مالی مانند بورس اوراق بهادار تهران، بانک مرکزی و مرکز آمار ایران است که داده‌های تاریخی جامعی در خصوص قیمت سهام و شاخص‌های اقتصادی ارائه می‌دهند. معیارهای انتخاب داده‌ها شامل پوشش دوره تاریخی (۱۳۹۰ تا ۱۴۰۲) و دسترسی به اطلاعات قیمتی برای حداقل ۸۵٪ روزهای معاملاتی است. بر این اساس، تعداد ۲۰ شرکت (نماد بورسی) به عنوان نمونه استفاده شده است که تشکیل پورتفولیو با استفاده از آن‌ها صورت می‌پذیرد. مراحل تحلیل داده‌ها به صورت زیر است:

- محاسبه بازده ماهانه هر سهم و شاخص کل
- تشکیل مجموعه داده ورودی و خروجی مدل‌های یادگیری ماشین. در این پژوهش، از الگوریتم‌های یادگیری ماشین به دو منظور استفاده شده است: پیش‌بینی شاخص و پیش‌بینی

بازده شرکت‌ها. در پیش‌بینی شاخص، متغیرهای ورودی مدل‌ها در هر ماه شامل تغییرات نرخ ارز، تغییرات نرخ تورم و نرخ بهره در ماه فعلی به همراه مقادیر بازده شاخص در ماه‌های گذشته است. متغیر خروجی نیز پیش‌بینی بازده شاخص در دوره بعد است. به منظور انتخاب تعداد ماه‌های گذشته که به عنوان ورودی انتخاب می‌شوند از تابع خودهمبستگی (ACF[1]) استفاده شده است. این تابع، مقدار همبستگی شاخص با مقادیر گذشته خود را نشان می‌دهد. بر این اساس، دوره‌های پیشین که دارای قدر مطلق خودهمبستگی بیشتر از ۰.۲ هستند به عنوان ورودی مدل در نظر گرفته می‌شوند. در پیش‌بینی بازده شرکت‌ها، متغیرهای ورودی مدل‌ها در هر ماه شامل مقادیر بازده شرکت در ماه‌های گذشته و مقادیر بازده شاخص در ماه‌های گذشته است. متغیر خروجی نیز پیش‌بینی بازده شرکت در دوره بعد است. مجددا به منظور انتخاب تعداد ماه‌های گذشته که به عنوان ورودی انتخاب می‌شوند از تابع خودهمبستگی استفاده شده است. بر این اساس، دوره‌های پیشین که دارای قدر مطلق خودهمبستگی بیشتر از ۰.۲ هستند به عنوان ورودی مدل در نظر گرفته می‌شوند.

- تنظیم پارامترهای مدل‌های یادگیری ماشین با استفاده از الگوریتم PSO: پس از ایجاد ورودی‌ها و خروجی‌های مدل‌ها، داده‌ها به سه بخش تقسیم می‌شوند: ۷۰ درصد برای آموزش، ۱۰ درصد برای اعتبارسنجی و ۲۰ درصد برای آزمون مدل‌ها. مدل‌ها با استفاده از داده‌های آموزش، آموزش داده‌شده و توسط داده‌های اعتبارسنجی، اعتبارسنجی شده‌اند. درنهایت، بهترین ترکیبات از پارامترهای مدل بر اساس معیار میانگین مربعات خطا (MSE[2]) و با استفاده از الگوریتم PSO انتخاب شده‌اند.
- پیش‌بینی بازده با استفاده از مدل‌های یادگیری ماشین: در این مرحله، پیش‌بینی بازده و شاخص توسط دو مدل SVR و RF انجام شده و مقایسه می‌شود.
- مدلسازی و حل مسئله برنامه‌ریزی دوهدفه طراحی شده، به منظور تخصیص دارایی: برای این کار، بازه‌های زمانی یک‌ماهه برای متعادل‌سازی پورتفولیو در نظر گرفته می‌شود. به همین دلیل، به صورت پنجره غلطان با طول پنجره ۳۰ روزه پیش رفته و در هر پنجره، پارامترهای مدل برنامه‌ریزی دوهدفه تخصیص دارایی با استفاده از اطلاعات همان دوره محاسبه شده و مدل بر اساس آن‌ها مجددا حل می‌شود. به این صورت، اوزان پورتفولیو به صورت پویا در هر دوره به‌روزرسانی می‌شوند. در ادامه، این گام به طور مفصل‌تر تشریح می‌شود:

[1] Auto-Correlation Function
[2] Mean Squared Error

پس از اینکه پیش‌بینی بازده دوره بعد شاخص سهام از طریق مدل‌های یادگیری ماشین انجام شد، انتخاب بین سرمایه گذاری در بازار سهام یا سرمایه گذاری در اوراق بهادار با درآمد ثابت بر اساس قائده زیر انجام می‌شود:

- چنانچه بازدهی پیش بینی شده برای شاخص بازار بیشتر یا مساوی بازدهی مورد انتظار برای اوراق بهادار با درآمد ثابت باشد، رویکرد سرمایه گذاری در بازار سهام پیشنهاد می‌گردد.
- چنانچه بازدهی پیش بینی شده برای شاخص بازار کمتر از بازدهی مورد انتظار برای اوراق بهادار با درآمد ثابت باشد، رویکرد سرمایه گذاری در اوراق بهادار با درآمد ثابت انتخاب می‌شود.

در این بخش‌مدل بهینه سازی چند هدفه به منظور تخصیص دارایی بین سهام مختلف بررسی می‌شود. چنانچه از بررسی صورت گرفته در بخش قبل تصمیم بر سرمایه گذاری در بازار سهام حاصل شود، موضوع انتخاب بهترین سهام برای تخصیص دارایی موضوعیت پیدا خواهد کرد و رویکرد مطرح شده در این بخش برای تخصیص دارایی استفاده می‌شود. مدل بهینه سازی چند هدفه به صورت زیر تعریف می‌شود:

$$Max\left(R_P = \sum_{i=1}^{n} x_i r_i\right)$$

$$Min\left(MpSn = \sum_{i=1}^{n} x_i p_i\right)$$

$subject$ to:

$$\sum_{i=1}^{i} x_i = 1$$

$$0 \le x_i \le \frac{1}{\text{mininnumber}}$$

در این فرمول n نمایانگر تعداد سهام مورد بررسی، R_p نشانگر بازدهی پرتفوی، $MsPn$ توزیع احتمال شرطی معکوس پرتفوی p، x_i میزان سرمایه گذاری شده در سهم شماره i، r_i بازدهی سهم i، p_i توزیع احتمال معکوس سهم i و $mininnumber$ حداقل تعداد سهامی است که می‌خواهیم در آن سرمایه گذاری کنیم.

توزیع احتمال شرطی معکوس یک صندوق یا یک پرتفوی به عنوان عامل ریسک مشخص شده و مورد استفاده قرار می‌گیرد. این متغیر میزان احتمال آنکه یک صندوق سرمایه گذاری در زمانی که بازدهی

شاخص بازار بیشتر از اوراق بهادار با درآمد ثابت بوده است و بازدهی صندوق کمتر از میزان شاخص کل بوده است را مورد ارزیابی قرار می‌دهد.

در این فرمول، P_i به صورت زیر محاسبه می‌شود:

$$p_i = \frac{p(MpSn)i}{p(Mp)}$$

در این فرمول صورت کسر توزیع احتمال مشترک آن است که شاخص، بازدهی بیشتر از اوراق بهادار با درآمد ثابت داشته باشد و سهم i موفق به کسب بازدهی حداقل به اندازه شاخص نشده باشد و مخرج کسر احتمال کسب بازدهی بیشتر از اوراق بهادار با درآمد ثابت توسط شاخص است. مدل فوق از طریق دو رویکرد حل می‌شود. در رویکرد اول، مدل دو هدفه فوق با استفاده از یکی از دو روش مجموع وزنی ساده یا اپسیلون محدودیت به یک مدل تک‌هدفه تبدیل شده و حل خواهد شد. در رویکرد دوم، از الگوریتم NSGAII استفاده شده و یک جبهه پارتو به دست می‌آید. در نهایت، متوسط جواب‌های جبهه پارتو به عنوان یک جواب انتخاب می‌شود. مراحل پیاده‌سازی الگوریتم NSGAII در پژوهش حاضر به صورت زیر است (تا زمان رسیدن به شرط خاتمه، در اینجا ۱۰۰ تکرار):

۱) تولید جواب اولیه به صورت بردار تصادفی ۲۰ عنصری به تعداد مشخص (pop_size)
۲) ارزیابی جواب‌ها با استفاده از تابع هدف بازده و ریسک
۳) تولید جواب‌های جدید با استفاده از عملگرهای تقاطع و جهش به ترتیب به تعداد nc و nm (در عملگر جهش یک جواب جدید به صورت تصادفی ایجاد می‌شود و در عملگر تقاطع از طریق دو جواب مثل a و b یک جواب جدید به صورت c=(a+b)/2 یا c=(a-b)/2 ایجاد می‌شود).
۴) ارزیابی جواب‌های جدید با استفاده از تابع هدف بازده و ریسک
۵) جبهه‌بندی و محاسبه فاصله ازدحامی (در این مرحله، جواب‌هایی که توسط هیچ جواب دیگری مغلوب نمی‌شوند (یعنی هیچ جواب دیگر نیست که بدون بدتر کردن حداقل یک هدف بتواند در حداقل یک هدف بهبود ایجاد کند) در جبهه اول (جبهه پارتو) قرار می‌گیرند. سپس، این جواب‌ها از جمعیت کنار گذاشته شده و همین منطق اعمال می‌شود. این روند تا زمانی که هر جواب به یک جبهه تخصیص یابد ادامه می‌یابد.).
۶) انتخاب pop_size جواب برتر و رفتن به گام III

۴- یافته‌های پژوهش

در این فصل، نتایج محاسباتی حاصل از اجرای الگوریتم‌های پیشرفته یادگیری ماشین و تکنیک‌های بهینه‌سازی بر روی مدل پرتفوی پیشنهادی ارائه شده است. جدول ۱ آمار مربوط به بازده ماهانه ۲۰ شرکت منتخب و شاخص کل (TEDPIX) را نشان می‌دهد و شامل پنج ستون اصلی است: میانگین، انحراف معیار، چارک اول، میانه و چارک سوم. در این تحلیل، هدف بررسی و مقایسه عملکرد مالی این شرکت‌ها بر اساس بازده ماهانه آنهاست. با توجه به این که میانگین بازده مثبت است، به نظر می‌رسد که شرکت‌ها به طور کلی عملکرد مناسبی دارند، اگرچه پراکندگی در بازده‌ها متفاوت است.

پراکندگی آماری میان شرکت‌ها می‌تواند به‌عنوان یکی از شاخص‌های مهم در ارزیابی ریسک سرمایه‌گذاری مورد توجه قرار گیرد. شرکت‌هایی با انحراف معیار بالا معمولاً دارای نوسانات بیشتری در بازده هستند و سرمایه‌گذاری در آن‌ها با ریسک بالاتری همراه است، در حالی که شرکت‌هایی با انحراف معیار پایین‌تر، ثبات بیشتری در عملکرد مالی خود نشان می‌دهند. همچنین، مقایسه چارک‌ها و میانه در هر شرکت می‌تواند اطلاعات مفیدی در مورد توزیع بازده‌ها و وجود داده‌های پرت یا رفتارهای نامتقارن فراهم کند. این شاخص‌های آماری در کنار نتایج پیش‌بینی مدل‌های یادگیری ماشین، نقش مهمی در تشکیل سبدهای بهینه ایفا می‌کنند و به تصمیم‌گیرندگان کمک می‌کنند تا ترکیبی از دارایی‌ها را با در نظر گرفتن بازده مورد انتظار و سطح ریسک انتخاب نمایند.

جدول ۱: آماره‌های توصیفی بازده ماهانه

	میانگین	انحراف معیار	چارک اول	میانه	چارک سوم
نماد ۱	3.71	14.11	-5.11	0.5	9.81
نماد ۲	2.82	14.83	-3.9	0	7.22
نماد ۳	2.5	13.14	-4.21	0	8.28
نماد ۴	2.33	14.11	-3.95	0	7.84
نماد ۵	3.89	12.73	-4.81	0.61	10.54
نماد ۶	2.97	13.76	-4.83	0	9.01
نماد ۷	3.69	20.09	-8.75	0.22	10.25
نماد ۸	4.01	15.16	-3.51	0.53	9.3
نماد ۹	3.88	12.45	-3.95	1.62	10.94
نماد ۱۰	4.14	19.29	-3.43	0.35	7.08
نماد ۱۱	2.88	16.82	-5.78	0	9.08
نماد ۱۲	5.4	31.23	-5.42	0	9.29

	میانگین	انحراف معیار	چارک اول	میانه	چارک سوم
نماد ۱۳	5.56	21.9	-5.78	0	11.71
نماد ۱۴	3.76	21.79	-8.65	-0.89	9.17
نماد ۱۵	3.61	21.99	-7.57	0	10.87
نماد ۱۶	3.32	21.53	-8.07	-1.16	9.57
نماد ۱۷	4.01	14.46	-4.93	-0.36	10.4
نماد ۱۸	5.94	36.42	-4.18	1.36	10.13
نماد ۱۹	3.49	16.36	-7.49	0.07	12.59
نماد ۲۰	2.13	14.28	-4.31	0.17	5.09
شاخص کل	0.14	1.12	-0.33	0.05	0.55
تغییرات نرخ ارز	8.72	16.43	-0.90	4.10	13.32
تغییرات نرخ تورم	3.99	17.02	-6.16	-0.11	13.67
نرخ بهره سالانه	17.02	2.27	15.00	17.00	18.00

جدول ۲ پارامترهای بهینه الگوریتم بهینه‌سازی تجمع ذرات (PSO) را برای تنظیم مدل‌های یادگیری ماشین (جنگل تصادفی و رگرسیون بردار پشتیبان) نمایش می‌دهد. برای تنظیم این پارامترها از رویکرد جستجوی تصادفی استفاده شده است. این جدول شامل سه پارامتر کلیدی PSO برای هر مدل است: c1 (ضریب یادگیری شناختی)، c2 (ضریب یادگیری اجتماعی) و w (ضریب اینرسی). مقادیر ارائه شده نشان می‌دهد که پارامترهای بهینه برای هر نماد بورسی متفاوت است، که بیانگر نیاز به تنظیم خاص برای هر دارایی می‌باشد. به عنوان مثال، برای نماد ۱ در مدل جنگل تصادفی، مقادیر c1=1.49، c2=1.05 و w=0.97 به دست آمده، در حالی که برای همان نماد در مدل رگرسیون بردار پشتیبان، این مقادیر به ترتیب ۱.۵۱، ۰.۷۳ و ۱.۳۵ هستند. این تفاوت‌ها نشان می‌دهد که هر مدل یادگیری ماشین به تنظیمات خاصی از الگوریتم PSO نیاز دارد تا به بهترین عملکرد دست یابد. به طور کلی، مقادیر c1 و c2 عموماً در محدوده ۰.۲ تا ۲ و w در بازه ۰.۳ تا ۱.۹ قرار دارند که نشان‌دهنده تنوع قابل توجه در پارامترهای بهینه برای نمادهای مختلف است.

این یافته‌ها اهمیت به‌کارگیری رویکرد تنظیم پارامتر تطبیقی را برجسته می‌سازد، چراکه عملکرد مدل‌های یادگیری ماشین به‌شدت وابسته به انتخاب مناسب این پارامترهاست. در واقع، استفاده از الگوریتم PSO به‌عنوان یک روش جستجوی هوشمند برای بهینه‌سازی پارامترها، باعث شده است تا مدل‌ها برای هر نماد بورسی به صورت خاص تنظیم شوند و از رفتار منحصر به‌فرد هر سهم در داده‌های تاریخی پیروی کنند. این موضوع نه‌تنها دقت پیش‌بینی را بهبود می‌بخشد، بلکه از بیش‌برازش مدل‌ها

جلوگیری کرده و قابلیت تعمیم آن‌ها را در مواجهه با داده‌های جدید افزایش می‌دهد. به‌کارگیری این فرآیند تنظیم دقیق، گامی مؤثر در جهت ساخت مدل‌های پیش‌بینی مقاوم و کارآمد برای استفاده در چارچوب بهینه‌سازی سبد سرمایه‌گذاری محسوب می‌شود.

جدول ۲: پارامترهای منتخب الگوریتم‌های PSO برای تنظیم پارامترهای هر الگوریتم یادگیری ماشین

نماد	RF			SVR		
	c1	c2	w	c1	c2	w
نماد ۱	1.49	1.05	0.97	1.51	0.73	1.35
نماد ۲	0.64	0.69	0.65	1.13	0.44	1.75
نماد ۳	1.26	1.12	0.76	1.36	0.63	1.98
نماد ۴	1.00	1.97	1.58	1.11	0.67	1.71
نماد ۵	1.75	0.52	0.89	0.47	1.81	1.66
نماد ۶	0.82	1.09	0.30	1.19	0.20	0.35
نماد ۷	1.46	1.24	1.10	1.73	0.89	1.22
نماد ۸	1.45	1.98	0.40	1.74	0.19	1.81
نماد ۹	1.85	1.12	0.91	0.92	1.35	1.47
نماد ۱۰	1.83	0.28	0.33	0.83	1.33	1.52
نماد ۱۱	1.58	0.95	1.51	1.06	1.19	1.50
نماد ۱۲	1.39	1.73	1.75	0.35	1.31	1.24
نماد ۱۳	1.90	1.77	0.64	1.27	0.37	1.65
نماد ۱۴	0.35	0.19	1.06	0.79	0.11	1.76
نماد ۱۵	0.59	0.78	1.88	0.71	1.03	1.91
نماد ۱۶	1.59	1.04	0.99	1.36	1.04	0.94
نماد ۱۷	0.27	0.15	0.86	1.22	0.11	1.36
نماد ۱۸	1.79	0.65	1.14	0.73	0.71	1.72
نماد ۱۹	0.72	1.08	0.89	1.54	0.17	1.70
نماد ۲۰	1.69	1.31	1.04	0.99	0.98	1.89

جدول ۳ پارامترهای بهینه الگوریتم‌های یادگیری ماشین (جنگل تصادفی و رگرسیون بردار پشتیبان) را که از طریق الگوریتم بهینه‌سازی تجمع ذرات (PSO) تعیین شده‌اند، نمایش می‌دهد. برای مدل جنگل تصادفی (RF) پارامترهای اصلی شامل تعداد تخمین‌گرها (درختان) و حداکثر عمق درختان است که به ترتیب بین ۱ تا ۵ تخمین‌گر و عمق ثابت ۱ متغیر بوده‌اند. در مورد مدل رگرسیون بردار پشتیبان (SVR)،

پارامترهای بهینه شامل پارامتر تنظیم (C) با مقادیری از ۰٫۰۰۱ تا ۴۲٫۱۸، نوع کرنل (شامل linear، rbf، poly و sigmoid) و پارامتر اپسیلون (ε) با دامنه‌ای از ۰٫۰۰۱ تا ۵۶٫۴۴ است. این جدول به وضوح نشان می‌دهد که پارامترهای بهینه برای هر نماد بورسی متفاوت است؛ به عنوان مثال، نماد ۲ به کرنل خطی با C=42.18 و ε=56.44 نیاز دارد، در حالی که نماد ۴ از کرنل rbf با C=7.40 و ε=22.61 استفاده می‌کند. این تنوع در پارامترهای بهینه نشان‌دهنده اهمیت تنظیم خاص مدل برای هر دارایی و پیچیدگی رفتارهای بازار سهام است که الگوریتم PSO موفق به کشف این تنظیمات بهینه برای هر مورد شده است.

این فرآیند تنظیم پارامترها با استفاده از PSO نه تنها دقت مدل‌های پیش‌بینی را افزایش داده، بلکه از افت کارایی ناشی از تنظیمات عمومی و غیراختصاصی جلوگیری کرده است. به عبارت دیگر، با جستجوی هدفمند در فضای پارامترها، PSO توانسته است ترکیب‌هایی از پارامترها را شناسایی کند که بیشترین تطابق را با ویژگی‌های رفتاری هر نماد بورسی دارند. این موضوع به‌ویژه در بازارهایی با رفتارهای نوسانی و غیرخطی مانند بورس تهران اهمیت بالایی دارد، چرا که استفاده از تنظیمات یکنواخت می‌تواند منجر به پیش‌بینی‌های نادرست و تصمیم‌گیری‌های زیان‌بار شود. بنابراین، استفاده از PSO در این پژوهش نه‌فقط به عنوان یک ابزار کمکی، بلکه به عنوان یک مؤلفه کلیدی در افزایش اثربخشی مدل‌های یادگیری ماشین برای تحلیل‌های مالی تلقی می‌شود.

جدول ۴ به مقایسه عملکرد دو مدل پیش‌بینی یادگیری ماشین - رگرسیون بردار پشتیبان (SVR) و جنگل تصادفی (RF) - بر اساس سه معیار ارزیابی اصلی می‌پردازد. این معیارها شامل جذر میانگین مربعات خطا (RMSE) میانگین خطای مطلق (MAE) و میانگین خطای پیش‌بینی (MFE) است که همگی بر روی داده‌های آزمون محاسبه شده‌اند. نتایج نشان می‌دهد که در اکثر نمادها (۱۳ نماد از ۲۰ نماد)، مدل SVR عملکرد بهتری از نظر RMSE داشته است که نشان‌دهنده دقت بالاتر این مدل در پیش‌بینی بازده سهام است. همچنین در معیار MAE که اندازه‌گیری میانگین قدرمطلق خطاهاست، مدل SVR در ۱۴ نماد از ۲۰ نماد برتری داشته است. در مقابل، مدل RF تنها در برخی نمادهای خاص مانند نماد ۱ و ۹ عملکرد نسبتاً مشابهی با SVR نشان داد. معیار MFE که جهت خطای پیش‌بینی را نیز در نظر می‌گیرد، نشان می‌دهد که هر دو مدل تمایل به کم‌بینی یا بیش‌بینی دارند، اما میزان این خطاها در مدل SVR عموماً کمتر بوده است. به طور کلی، این جدول مؤید برتری نسبی مدل SVR در پیش‌بینی بازده سهام در بازار بورس تهران است، اگرچه در برخی موارد خاص، مدل RF نیز نتایج قابل قبولی ارائه کرده است.

افزون بر نتایج کمی ارائه‌شده در جدول ۴، این مقایسه نشان‌دهنده اهمیت انتخاب مدل مناسب در مرحله پیش‌بینی برای بهینه‌سازی سبد سرمایه‌گذاری است. عملکرد بهتر SVR در معیارهای RMSE و MAEبه‌ویژه در نمادهایی با نوسانات بالا، نشان می‌دهد که این مدل توانایی بیشتری در یادگیری الگوهای

پیچیده بازده سهام دارد. همچنین، مقدار کمتر MFE در SVR بیانگر این است که این مدل نه‌تنها از نظر بزرگی خطا بلکه از نظر جهت پیش‌بینی نیز دقت بیشتری دارد. این برتری در پیش‌بینی‌های دقیق‌تر، مستقیماً به بهبود کیفیت ورودی مدل بهینه‌سازی منجر می‌شود و در نتیجه، تشکیل سبدهایی با ریسک پایین‌تر و بازده بالاتر را تسهیل می‌کند. بنابراین، انتخاب مدل SVR به عنوان ابزار پیش‌بینی، نقش مهمی در موفقیت چارچوب ترکیبی ارائه‌شده در این پژوهش ایفا کرده است.

جدول ۳: پارامترهای بهینه الگوریتم‌های یادگیری ماشین حاصل شده از طریق PSO

نماد	RF		SVR		
	تعداد	حداکثر عمق	C	Kernel	Epsilon
نماد ۱	1	1	0.01	sigmoid	21.99
نماد ۲	1	1	42.18	linear	56.44
نماد ۳	1	1	0.01	sigmoid	14.41
نماد ۴	2	1	7.40	rbf	22.61
نماد ۵	1	1	0.01	sigmoid	24.86
نماد ۶	1	1	0.59	linear	0.55
نماد ۷	1	1	3.22	linear	3.07
نماد ۸	1	1	16.55	sigmoid	50.35
نماد ۹	1	1	4.78	linear	22.29
نماد ۱۰	1	1	0.01	sigmoid	11.91
نماد ۱۱	1	1	0.01	sigmoid	17.63
نماد ۱۲	1	1	4.44	sigmoid	11.86
نماد ۱۳	1	1	0.01	sigmoid	24.57
نماد ۱۴	2	1	16.26	sigmoid	42.93
نماد ۱۵	5	1	0.01	linear	32.00
نماد ۱۶	1	1	0.01	poly	0.01
نماد ۱۷	1	1	2.53	rbf	2.50
نماد ۱۸	2	1	2.82	sigmoid	2.50
نماد ۱۹	2	1	0.01	poly	14.23
نماد ۲۰	2	1	0.01	sigmoid	19.07

جدول ۴: مقایسه عملکرد مدل‌های پیش‌بینی

نماد	SVR				RF			
	MSE	RMSE	MAE	MFE	MSE	RMSE	MAE	MFE
نماد ۱	156.85	12.52	11.35	-2.16	156.96	12.53	11.00	-1.35
نماد ۲	176.94	13.30	10.52	-1.21	173.22	13.16	10.64	-1.64
نماد ۳	146.63	12.11	10.86	-0.10	220.92	14.86	11.19	6.88
نماد ۴	185.50	13.62	11.78	-1.20	204.65	14.31	12.74	1.58
نماد ۵	152.99	12.37	10.64	-2.22	302.48	17.39	12.28	-0.10
نماد ۶	194.22	13.94	12.45	-3.16	315.43	17.76	14.86	-7.08
نماد ۷	238.15	15.43	14.29	-0.72	460.73	21.46	17.35	3.47
نماد ۸	208.93	14.45	12.09	-1.11	220.96	14.86	13.63	-4.40
نماد ۹	186.67	13.66	11.66	-1.11	211.75	14.55	11.77	-0.32
نماد ۱۰	190.27	13.79	11.33	2.16	242.32	15.57	12.75	0.36
نماد ۱۱	221.45	14.88	12.74	-1.24	290.20	17.04	15.07	4.57
نماد ۱۲	246.83	15.71	13.83	1.03	426.16	20.64	17.46	0.95
نماد ۱۳	143.55	11.98	9.38	1.19	244.76	15.64	10.50	4.22
نماد ۱۴	227.03	15.07	13.52	0.90	249.07	15.78	14.96	2.76
نماد ۱۵	145.38	12.06	9.75	-0.55	320.35	17.90	13.26	3.86
نماد ۱۶	180.84	13.45	11.55	2.09	206.89	14.38	11.57	-0.32
نماد ۱۷	138.52	11.77	10.24	0.78	178.02	13.34	10.69	0.50
نماد ۱۸	182.39	13.51	12.16	-1.32	396.68	19.92	14.17	-5.79
نماد ۱۹	95.94	9.80	7.72	2.45	236.55	15.38	10.46	4.11
نماد ۲۰	177.45	13.32	10.62	0.79	348.95	18.68	14.73	5.69

به منظور حل مسئله دو هدفه بهینه‌سازی پورتفولیو از الگوریتم NSGAII استفاده شده است. فرض شده است که ترکیب پورتفولیو هر ماه به‌روزرسانی می‌شود. بنابراین، به ازای هر دوره آزمون یک مسئله (در مجموع ۳۳ مسئله) حل و تحلیل می‌شود. در هر دوره، پیش‌بینی بازده به دست آمده توسط الگوریتم‌های مختلف (SVR و RF) به عنوان بازده مورد انتظار هر سهم در تابع هدف استفاده می‌شود. توجه داشته باشید انتخاب پارامتر حداقل تعداد سهام در پورتفولیو ($mininnumber$) بستگی به ترجیحات سرمایه‌گذار دارد و نمی‌توان مقدار مشخصی برای آن در نظر گرفت. به همین منظور، مسئله با درنظر گرفتن مقادیر مختلف برای این پارامتر حل شده است (مقادیر ۱ تا ۲۰ برای حداقل تعداد سهام در پورتفولیو). شکل‌های ۱ و ۲ جبهه‌های پارتو را برای برخی از حالات نشان می‌دهد. از این شکل‌ها می‌توان

دریافت که (همانگونه که انتظار می‌رود) ارتباط بین ریسک و بازده پوتفولیو مستقیم است. به عبارتی، برای ایجاد بازد مورد انتظار بیشتر، ریسک بیشتری باید تحمل کرد.

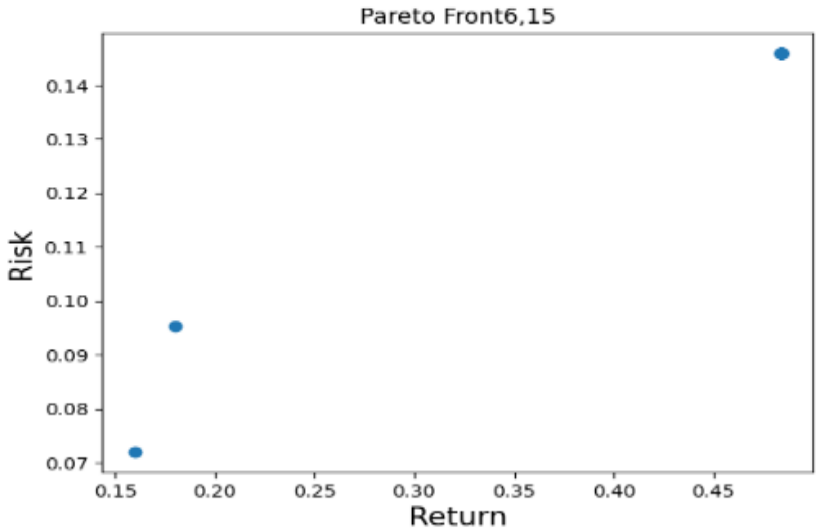
استفاده از الگوریتم NSGA-II این امکان را فراهم می‌آورد که مجموعه‌ای از راه‌حل‌های پارتو به‌دست آید که سرمایه‌گذار بتواند بسته به میزان ریسک‌پذیری خود، از میان آن‌ها انتخاب بهینه‌تری داشته باشد. به‌کارگیری مدل‌های یادگیری ماشین برای پیش‌بینی بازده سهام در هر دوره، منجر به پویایی تصمیم‌گیری و انطباق بهتر با شرایط واقعی بازار شده است. به‌علاوه، اجرای الگوریتم در بازه‌های ماهانه و تحلیل سبدهای بهینه حاصل از هر مرحله، تصویر روشن‌تری از عملکرد مدل در طول زمان فراهم می‌کند. بررسی نتایج برای مقادیر مختلف حداقل تعداد سهام در پورتفولیو (minnumber) نیز نشان‌دهنده تأثیر قابل توجه این پارامتر بر ساختار ریسک-بازده سبد است و این نکته را تقویت می‌کند که چارچوب پیشنهادی به گونه‌ای طراحی شده که قابلیت انطباق با ترجیحات متنوع سرمایه‌گذاران را داشته باشد. چنین رویکردی، نه‌تنها از لحاظ علمی، بلکه در عمل نیز گامی مؤثر در جهت توسعه ابزارهای تصمیم‌یار در مدیریت پورتفولیو محسوب می‌شود.

جدول ۵ و ۶ به بررسی عملکرد سبدهای تشکیل شده با استفاده از nsaii مبتنی بر RF و SVR در دوره های زمانی مختلف و به ازای تعداد سهام سبد برابر ۱۰ (به عنوان نمونه) می‌پردازند. محاسبات مشابه برای تعداد سهام (MinNum) مختلف انجام شده است. در جدول ۵، که از الگوریتم RF برای پیش‌بینی بازده استفاده می‌شود، تعداد نقاط پارتو حاصل شده در هر دوره متفاوت است. توجه داشته باشیدکه تعداد نقاط پارتو بیشتر نشان دهنده تولید جوابهای متنوع‌تر و عملکرد بهتر مدل است. برای مثال، در دوره زمانی ۱۱، تعداد نقاط پارتو تنها ۱ نقطه است اما بازده سبد میانگین به ۱۰.۶ درصد می‌رسد. نکته قابل توجه در این جدول این است که در برخی دوره‌ها (دوره‌های ۲، ۳، ۶، و غیره)، بازده سبد میانگین برابر بازده اوراق درآمد ثابت است و در این موارد، ستاره (*) نشان‌دهنده استفاده نشدن از الگوریتم nsgaii است زیرا بازده پیش‌بینی شده شاخص کمتر از بازده درآمد ثابت است. در مجموع، جدول ۵ ترکیبی از بازده‌های متوسط و ریسک‌های متعادل را نشان می‌دهد .از طرف دیگر، جدول ۶ عملکرد استراتژی معاملاتی را با الگوریتم SVR بررسی می‌کند. در این جدول، بازده سبد میانگین در برخی دوره ها بیشتر از جدول اول است، مثلا در دوره‌های ۱ و ۴ که بازده‌هایی برابر با ۳.۴۶ و ۴.۷۲ درصد را نشان می‌دهد. همچنین در این جدول نیز تعداد نقاط پارتو متغیر است. به عنوان مثال، در دوره ۵، تعداد نقاط پارتو ۴۴ و بازده سبد ۲.۳۵ درصد است، که نشان‌دهنده عملکرد قابل قبول است. در هر دو جدول، دوره‌هایی وجود دارند که الگوریتم nsgaii به دلیل پیش‌بینی بازده ضعیف شاخص استفاده نشده است.

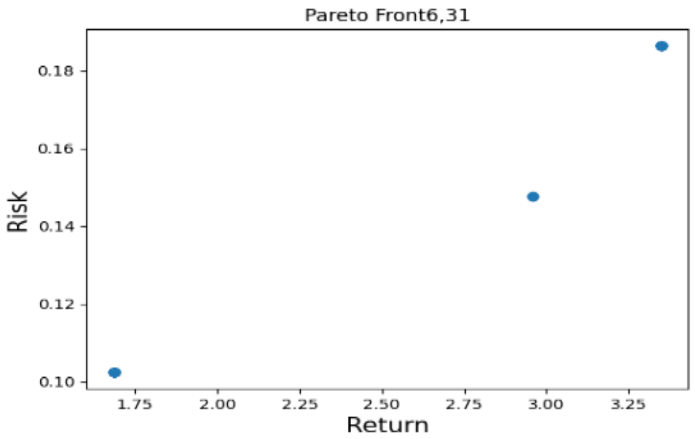
جدول ۷ و جدول ۸ نتایج آزمون t زوجی را برای مقایسه بازدهی سبدهای بهینه حاصل از ترکیب الگوریتم‌های NSGAII و RF و همچنین NSGAII و SVR نسبت به بازدهی شاخص ارائه می‌دهند. در

جدول ۷، با توجه به p-value برابر با ۰.۰۰۰ برای تمامی مقادیر تعداد سهام(MinNum)، می‌توان نتیجه گرفت که تفاوت‌های مشاهده شده در میانگین بازدهی سبد بهینه نسبت به بازدهی شاخص به‌طور معناداری وجود دارد. همچنین، مقادیر آماره t که در محدوده‌ای بالاتر از ۹ قرار دارند، و میانگین تفاضل‌ها که همیشه مثبت است، نشان‌دهنده این است که بازدهی سبد بهینه حاصل از ترکیب NSGAII و RF در تمامی حالات بیشتر از بازدهی شاخص است.

در جدول ۸، نتایج مشابهی برای مقایسه بازدهی سبد بهینه حاصل از ترکیب NSGAII و SVR ارائه شده است. مانند جدول قبل، p-value برابر با ۰.۰۰۰ برای تمامی مقادیر نشان می‌دهد که تفاوت‌های معنادار در میانگین‌ها وجود دارد. اگرچه در این جدول نیز میانگین تفاضل‌ها مثبت هستند، به طور کلی مقادیر آماره t و میانگین تفاضل‌ها نسبت به جدول اول کمی پایین‌تر هستند، که ممکن است نشان‌دهنده این باشد که عملکرد سبد حاصل از ترکیب NSGAII و SVR به اندازه NSGAII و RF قوی نیست، اما همچنان به‌طور معناداری بهتر از بازدهی شاخص است. این دو جدول به مقایسه کارایی دو الگوریتم مختلف در تولید سبدهای بهینه و تأثیر آنها بر بازدهی سرمایه‌گذاری کمک می‌کنند.



شکل ۱: جبهه پارتو برای پورتفولیو دوره ۲۱ با حداقل تعداد سهام در پورتفولیو برابر ۶ و استفاده از الگوریتم RF برای پیش‌بینی بازده



شکل ۲: جبهه پارتو برای پورتفولیو دوره ۳۱ با حداقل تعداد سهام در پورتفولیو برابر ۶ و استفاده از الگوریتم RF برای پیش‌بینی بازده

جدول ۵: عملکرد استراتژی معاملاتی پیشنهادی با تعداد سهام سبد (MinNum) برابر ۱۰ و الگوریتم RF برای پیش‌بینی بازده

دوره زمانی	تعداد نقاط پارتو	بازده سبد میانگین	ریسک سبد میانگین
1	16	1.5	0.2
2	*	1.3	0.0
3	*	1.3	0.0
4	2	2.6	0.2
5	1	2.0	0.2
6	*	1.3	0.0
7	2	3.1	0.2
8	*	1.3	0.0
9	*	1.3	0.0
10	*	1.3	0.0
11	1	10.6	0.5
12	*	1.3	0.0
13	33	0.5	0.2
14	2	9.9	0.5
15	*	1.3	0.0
16	2	0.0	0.0
17	*	1.3	0.0

دوره زمانی	تعداد نقاط پارتو	بازده سبد میانگین	ریسک سبد میانگین
18	*	1.3	0.0
19	*	1.3	0.0
20	*	1.3	0.0
21	*	1.3	0.0
22	*	1.3	0.0
23	5	5.0	0.6
24	19	10.6	0.5
25	*	1.5	0.0
26	*	1.5	0.0
27	3	12.6	0.5
28	8	7.5	0.5
29	2	1.5	0.2
30	*	1.5	0.0
31	11	3.5	0.3
32	6	2.4	0.2
33	*	1.5	0.0

جدول ۹ نتایج آزمون t زوجی را برای مقایسه عملکرد الگوریتم NSGAII مبتنی بر مدل‌های RF و SVR ارائه می‌دهد. این جدول شامل سه عامل مقایسه است: بازده سبد میانگین، ریسک سبد میانگین، و تعداد نقاط پارتو. برای هر یک از عوامل، p-value و میانگین تفاضل ارائه شده است که به مقایسه این دو الگوریتم در زمینه‌های مختلف کمک می‌کند.

جدول ۶: عملکرد استراتژی معاملاتی پیشنهادی با تعداد سهام سبد (MinNum) برابر ۱۰ و الگوریتم SVR برای پیش‌بینی بازده

دوره زمانی	تعداد نقاط پارتو	بازده سبد میانگین	ریسک سبد میانگین
1	3	3.46	0.54
2	*	1.33	0.00
3	*	1.33	0.00
4	1	4.72	0.54
5	44	2.35	0.55
6	*	1.33	0.00
7	29	0.53	0.26
8	*	1.33	0.00

دوره زمانی	تعداد نقاط پارتو	بازده سبد میانگین	ریسک سبد میانگین
9	*	1.33	0.00
10	*	1.33	0.00
11	22	4.97	0.55
12	*	1.33	0.00
13	6	6.32	0.56
14	23	2.47	0.54
15	*	1.33	0.00
16	1	1.46	0.56
17	*	1.33	0.00
18	*	1.33	0.00
19	*	1.33	0.00
20	*	1.33	0.00
21	*	1.33	0.00
22	*	1.33	0.00
23	3	5.11	0.54
24	1	8.00	0.53
25	*	1.50	0.00
26	*	1.50	0.00
27	8	2.04	0.28
28	16	0.84	0.22
29	4	3.21	0.54
30	*	1.50	0.00
31	15	0.83	0.30
32	42	3.30	0.58
33	*	1.50	0.00

برای عامل بازده سبد میانگین، p-value برابر با ۰.۰۰۰ است که نشان‌دهنده وجود یک تفاوت معنادار بین دو الگوریتم در این زمینه است. میانگین تفاضل ۰.۹۵۶ نشان می‌دهد که الگوریتم NSGAII مبتنی بر RF به‌طور میانگین بازدهی بالاتری نسبت به NSGAII مبتنی بر SVR دارد. این امر حاکی از این است که استفاده از RF به‌عنوان مدل پیش‌بینی می‌تواند عملکرد بهتری در تولید سبدهای سرمایه‌گذاری داشته باشد.

در مورد ریسک سبد میانگین، p-value برابر با ۰.۰۰۴ به معنای وجود یک تفاوت معنادار بین دو الگوریتم در این زمینه است. با این حال، میانگین تفاضل منفی (۰.۰۱۸-) به این معناست که ریسک سبد بهینه حاصل از NSGAII مبتنی بر RF کمی کمتر از NSGAII مبتنی بر SVR است. این بدان معناست که بازده سبد با استفاده از RF بالاتر بوده و با ریسک کمتری همراه است.

برای عامل تعداد نقاط پارتو، p-value برابر با ۰.۰۰۱ به وجود تفاوت معنادار در این زمینه اشاره می‌کند. میانگین تفاضل ۱.۹۶۸- نشان می‌دهد که تعداد نقاط پارتو حاصل از NSGAII مبتنی بر SVR بیشتر از NSGAII مبتنی بر RF است. این عامل می‌تواند نشان‌دهنده این باشد که NSGAII مبتنی بر SVRاز لحاظ تنوع بهتر عمل کرده و نقاط بهینه بیشتری را شناسایی کرده است که می‌تواند به دست‌یابی به سبدهای متنوع‌تر کمک کند. در مجموع، این نتایج تحلیلی ارزشمند از عملکرد دو الگوریتم مختلف در زمینه‌های کلیدی ارائه می‌دهند و بر اهمیت انتخاب مدل مناسب برای بازده و ریسک‌های سرمایه‌گذاری تأکید می‌کنند.

جدول ۷: نتایج آزمون t زوجی برای مقایسه بازدهی سبد بهینه حاصل از ترکیب NSGAII و RF نسبت به بازدهی شاخص به ازای تعداد سهام سبد (MinNum) مختلف

تعداد سهام سبد (MinNum)	pvalue	آماره	میانگین تفاضل
1	0.000	13.118	0.798
2	0.000	10.156	0.742
3	0.000	9.346	1.143
4	0.000	11.619	1.170
5	0.000	13.158	1.169
6	0.000	11.353	0.768
7	0.000	12.623	0.861
8	0.000	11.159	0.986
9	0.000	12.051	0.751
10	0.000	10.635	0.867
11	0.000	12.122	0.819
12	0.000	11.512	0.861
13	0.000	12.145	0.853
14	0.000	11.656	0.904
15	0.000	10.982	0.968
16	0.000	12.343	0.923

تعداد سهام سبد (MinNum)	pvalue	آماره	میانگین تفاضل
17	0.000	10.631	0.948
18	0.000	11.131	0.951
19	0.000	11.417	0.867
20	0.000	10.874	0.970

جدول ۸: نتایج آزمون t زوجی برای مقایسه بازدهی سبد بهینه حاصل از ترکیب NSGAII و SVR نسبت به بازدهی شاخص به ازای تعداد سهام سبد (MinNum) مختلف

تعداد سهام سبد (MinNum)	pvalue	آماره	میانگین تفاضل
1	0.000	12.792	0.805
2	0.000	12.354	0.854
3	0.000	12.389	0.773
4	0.000	11.577	0.752
5	0.000	11.858	0.846
6	0.000	12.547	0.567
7	0.000	12.473	0.738
8	0.000	12.748	0.792
9	0.000	11.953	0.693
10	0.000	12.391	0.749
11	0.000	12.288	0.672
12	0.000	12.819	0.687
13	0.000	12.320	0.615
14	0.000	12.521	0.719
15	0.000	12.902	0.620
16	0.000	12.652	0.698
17	0.000	12.761	0.665
18	0.000	12.475	0.704
19	0.000	12.530	0.677
20	0.000	12.453	0.706

جدول ۹: آزمون t زوجی برای مقایسه عملکرد الگوریتم NSGAII مبتنی بر RF نسبت به NSGAII مبتنی بر SVR از لحاظ عوامل مختلف

عامل مقایسه	pvalue	میانگین تفاضل
بازده سبد میانگین	0.000	0.956
ریسک سبد میانگین	0.004	-0.018
تعداد نقاط پارتو	0.001	-1.968

بحث و نتیجه‌گیری

این پژوهش به ارائه یک چارچوب ترکیبی نوآورانه برای مدیریت سبد سرمایه‌گذاری می‌پردازد که در آن از مدل‌های پیشرفته یادگیری ماشین و الگوریتم‌های فراابتکاری بهره گرفته شده است. مطالعه حاضر با تمرکز بر داده‌های بازار بورس تهران، رویکردی نظام‌مند را در پیش‌بینی مالی و بهینه‌سازی پرتفوی معرفی می‌کند.

در بخش مدل‌سازی پیش‌بینی، از رگرسیون بردار پشتیبان و جنگل تصادفی برای پیش‌بینی بازده سهام و شاخص کل استفاده شد. الگوریتم بهینه‌سازی تجمع ذرات نقش محوری در تنظیم پارامترهای این مدل‌ها ایفا کرد و با جستجوی کارآمد در فضای پارامترها، مقادیر بهینه را برای هر مدل تعیین نمود.

در گام بعدی، الگوریتم ژنتیک با مرتب‌سازی نامغلوب به عنوان هسته اصلی سیستم بهینه‌سازی چندهدفه مورد استفاده قرار گرفت. این الگوریتم با ایجاد جبهه پارتو، امکان دستیابی همزمان به اهداف متعارض حداکثرسازی بازده و حداقل‌سازی ریسک را فراهم آورد. نتایج تجربی نشان داد که این چارچوب ترکیبی قادر است پرتفوی‌هایی با عملکرد برتر نسبت به شاخص بازار تولید کند.

مقایسه عملکرد مدل‌های یادگیری ماشین نشان داد که هر یک از این مدل‌ها در شرایط مختلف مزایای خاص خود را دارند. به طکه ور کلی، سبد حاصل از ترکیب NSGAII و RF بازدهی بالاتر و ریسک کمتری ارائه می کند، در حالی جوابهای حاصل از ترکیب NSGAII و SVR متنوع‌تر هستند.

یافته‌های این پژوهش حاکی از آن است که ترکیب تکنیک‌های یادگیری ماشین با الگوریتم‌های فراابتکاری می‌تواند به تصمیم‌گیری‌های سرمایه‌گذاری کارآمدتر منجر شود. همچنین، الگوریتم بهینه‌سازی تجمع ذرات ابزاری قدرتمند برای تنظیم پارامترهای مدل‌های پیچیده مالی است و الگوریتم ژنتیک با مرتب‌سازی نامغلوب نیز توانایی خود را در حل مسائل بهینه‌سازی چندهدفه مالی به خوبی نشان داد.

این مطالعه گامی مهم در جهت تلفیق روش‌های هوش محاسباتی با حوزه مالی محسوب می‌شود و چارچوبی را ارائه می‌دهد که می‌تواند مبنایی برای توسعه سیستم‌های پشتیبان تصمیم‌گیری مالی

هوشمند باشد. پیشنهاد می‌شود در تحقیقات آتی این چارچوب با الگوریتم‌های ترکیبی جدیدتر توسعه یابد و بر روی بازارهای مالی مختلف اعتبارسنجی شود. به طور دقیق‌تر، دو پیشنهاد پژوهشی زیر ارائه می‌شود:

توسعه چارچوب ترکیبی با الگوریتم‌های یادگیری عمیق

پیشنهاد می‌شود مدل‌های یادگیری عمیق مانند LSTM و Transformer به چارچوب موجود اضافه شوند. این مدل‌ها می‌توانند الگوهای پیچیده‌تر و روابط غیرخطی در داده‌های مالی را شناسایی کنند. همچنین می‌توان از معماری‌های هیبریدی که قابلیت پردازش همزمان داده‌های ساختاریافته و غیرساختاریافته (مانند اخبار و تحلیل‌های متنی) را دارند استفاده نمود. این توسعه می‌تواند دقت پیش‌بینی‌ها را بهبود بخشد.

ادغام معیارهای ریسک پیشرفته در بهینه‌سازی

پیشنهاد می‌شود معیارهای ریسک دیگر مانند Conditional Value-at-Risk (CVaR) و ریسک سیستماتیک در الگوریتم NSGA-II گنجانده شود. این کار مستلزم توسعه فرمول‌بندی جدیدی برای مسئله بهینه‌سازی چندهدفه است که بتواند علاوه بر بازده و معیار ریسک فعلی، این معیارهای ریسک را نیز در نظر بگیرد.

فهرست منابع

اقتصاد, امیرعلی؛ محمدی,عمران. (1402). بهینه سازی سبد سرمایه گذاری به کمک پیش بینی بازده مورد انتظار با استفاده از روش های شبکه عصبی LSTM، جنگل تصادفی و ARIMA. چشم انداز مدیریت مالی.

تهرانی رضا؛ محمد هندیجانی زاده، و عیسی نوروزیان لکوان ،. (۱۳۹۴). ارائه رویکردی جدید برای مدیریت فعال پرتفوی وانجام معاملات هوشمند سهام با تاکید بر نگرش انتخاب ویژگی. دانش سرمایه‌گذاری، ۴(۱۳): ۱۰۷۱۲۶.

رستمی, ژیلا, فتاحی, سهیلی. (۱۴۰۲). مدل‌سازی و تخمین بازده بورس اوراق بهادار تهران با استفاده از مدل‌های پویا. اقتصاد مالی, ۱۷(۶۲), ۱۸۵۲۱۶.

رضا شاطری مریم، نصیری محمد، زینالی مهدی. ارائه الگوی بهینه سبد سهام از طریق محدودیت تسلط تصادفی و کاهش ریسک گریزی مطلق. نشریه اقتصاد و بانکداری اسلامی. ۱۴۰۳; ۱۳ (۴۶) :۲۲۷-۲۵۲

شبان مهدی ؛ اله نخعی حبیب؛ الله طالب نیاقدرت و نازنین بشیری منش. (۱۳۹۹). طراحی الگوی غیرخطی سرایت‌پذیری شاخص کل بورس اوراق بهادار تهران از بازار دارایی‌های فیزیکی (کاربردی از مدل شبکه عصبی مصنوعی .(NARXمهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۶): ۴۷۶۴۹۴.

علیزاده, حسین؛ کیانفر, کامران. (۱۴۰۲). توسعهٔ مدل مارکوویتز در بهینه‌سازی سبد سهام با در نظر گرفتن محدودیت‌های واقع‌گرایانه. مدیریت دارایی و تامین مالی, ۱۱(۴), ۶۵۹۲.

کرامتی, اسفندیار؛ غلامی جمکرانی, و کاشفی نیشابوری. (۱۴۰۱). بهینه‌سازی پرتفوی سهام در بورس اوراق بهادار تهران (کاربرد رهیافت یادگیری تقویتی). مدلسازی اقتصادی, ۱۶(۵۸), ۵۱۶۶.

میرعلوی سید حسن؛ زهرا پورزمانی و آزیتا جهانشاد. (۱۳۹۸). ارائه مدلی مبتنی بر رفتار مالی سرمایه گذاران جهت پیش بینی قیمت سهام با استفاده از روش های فرا ابتکاری شبکه های عصبی. فصلنامه بورس اوراق بهادار، ۴۷(۱۲): ۱۰۹۱۷۶.

Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support vector regression. Efficient learning machines: Theories, concepts, and applications for engineers and system designers, 67-80.

Ban, G. Y., El Karoui, N., & Lim, A. E. (2018). Machine learning and portfolio optimization. Management Science, 64(3), 1136-1154.

Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. Engineering Applications of Artificial Intelligence, 120, 105843.

Chen, H., Wang, P., Li, J., Zhang, G., & Zhang, Y. (2025). An improved NSGAII-SA algorithm for the cell manufacturing system layout optimization problem. Operational Research, 25(1), 22.

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. Applied Soft Computing, 100, 106943.

Cheng, Q., Yang, L., Zheng, J., Tian, M., & Xin, D. (2024). Optimizing Portfolio Management and Risk Assessment in Digital Assets Using Deep Learning for Predictive Analysis. arXiv preprint arXiv:2402.15994.

Daviran, M., Maghsoudi, A., & Ghezelbash, R. (2025). Optimized AI-MPM: Application of PSO for tuning the hyperparameters of SVM and RF algorithms. Computers & Geosciences, 195, 105785.

García-Gonzalo, E., García-Nieto, P. J., Fidalgo Valverde, G., Riesgo Fernández, P., Sánchez Lasheras, F., & Suárez Gómez, S. L. (2024). Hybrid DE-Optimized GPR and NARX/SVR Models for Forecasting Gold Spot Prices: A Case Study of the Global Commodities Market. Mathematics, 12(7), 1039.

Guarino, A., Santoro, D., Grilli, L., Zaccagnino, R., & Balbi, M. (2024). EvoFolio: a portfolio optimization method based on multi-objective evolutionary algorithms. Neural Computing and Applications, 36(13), 7221-7243.

Habbab, F. Z., & Kampouridis, M. (2024). An in-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs. Expert Systems with Applications, 235, 121102.

Lwin, K. T., Qu, R., & MacCarthy, B. L. (2017). Mean-VaR portfolio optimization: A nonparametric approach. European Journal of Operational Research, 260(2), 751-766.

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications, 165, 113973.

Odeyemi, O., Mhlongo, N. Z., Nwankwo, E. E., Scholatica, U. C., & Okoye, C. C. (2024). Big data applications in portfolio management: A review of techniques and strategies. International Journal of Science and Research Archive, 11(1), 1984-1999.

Pradhan, A., Das, A., & Bisoy, S. K. (2025). Modified parallel PSO algorithm in cloud computing for performance improvement. Cluster Computing, 28(2), 131 Sadorsky, P. (2021). A random forests approach to predicting clean energy stock prices. Journal of Risk and Financial Management, 14(2), 48.

Sen, J., & Dutta, A. (2023). Portfolio optimization for the Indian stock market. In Encyclopedia of Data Science and Machine Learning (pp. 1904-1951). IGI Global.

Wu, L., Ahmad, M., Qureshi, S. A., Raza, K., & Khan, Y. A. (2022). An analysis of machine learning risk factors and risk parity portfolio optimization. Plos one, 17(9), e0272521.

Combining Metaheuristic Algorithms and Machine Learning for Investment Portfolio Management Based on Medium-Term Forecasting of Stock Market Index Data

Behnaz Badaei
PhD Student of Financial Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

Hossein Badiei
Assistant Professor of Accounting Department, South Tehran Branch, Islamic Azad University, Tehran, Iran
(Corresponding Author)
badiei@iau.ac.ir

Mohsen Hamidian
Associate Professor of Accounting Department, South Tehran Branch, Islamic Azad University, Tehran, Iran

Ali Amiri
Assistant Professor of Accounting department, Bandar Abbas Branch, Islamic Azad University, Bandar Abbas, Iran

Abstract
This study presents a new framework for investment portfolio optimization by combining machine learning algorithms and metaheuristic methods. Using historical data of the Tehran Stock Exchange (2011-2023), the present study designs an intelligent system that is capable of predicting market behavior and forming optimal investment portfolios.
In the first stage, machine learning models including support vector regression (SVR) and random forest (RF) are used to predict stock returns and the total index. The parameters of these models are adjusted using the particle swarm optimization (PSO) algorithm.
In the next stage, the investment portfolio optimization problem is formulated as a multi-objective problem and solved with the two-objective genetic algorithm NSGA-II. By creating a Pareto front, this algorithm enables the simultaneous achievement of the conflicting goals of maximizing returns and minimizing risks. Experimental results show that portfolios constructed using this method provide higher returns than the market index with an acceptable level of risk.
This research is important in three ways: first, it shows that the intelligent combination of machine learning and meta-heuristic methods can lead to more efficient investment decisions. Second, it demonstrates the efficiency of the PSO algorithm in tuning the parameters of financial models and the ability of NSGA-II to solve complex portfolio optimization problems. Third, it provides a practical framework for financial analysts and portfolio managers that can adapt to different market conditions.
Keywords: NSGAII, PSO, Portfolio Optimization, Machine Learning