



بهینه‌سازی سبد سهام با استفاده از روش میانگین-واریانس تکنیکالی و جنگل

تصادفی

لیلا برومند^۱، سعید آقاسی^۲ و محمدرضا شریفی قزوینی^۳

تاریخ دریافت: ۱۴۰۳/۵/۲۲ و تاریخ پذیرش: ۱۴۰۳/۸/۲۷

چکیده

هدف از انجام این پژوهش بهینه‌سازی سبد سهام با استفاده از روش میانگین-واریانس تکنیکالی و جنگل تصادفی (RF) است. روش پژوهش حاضر بر اساس هدف از نوع کاربردی و از نظر نوع داده‌ها، پس‌رویدادی می‌باشد. جامعه آماری این پژوهش شامل کلیه شرکت‌های فعال در بورس اوراق بهادار تهران طی سال‌های ۱۳۹۰ الی ۱۴۰۲ می‌باشد. در تحقیق حاضر تعداد ۴۰ سهم از سهام فعال در بازار سهام که داده‌های پیوسته و کامل داشته‌اند انتخاب شده است. برای بررسی بهینه‌سازی سبد سهام از اطلاعات قیمت سهام ۵ شرکت پذیرفته شده در بورس اوراق بهادار استفاده شد. ملاک انتخاب این ۵ شرکت بالا بودن میزان میانگین قیمت، میانگین بازده و تعداد روزهای معامله شده نسبت به سایر شرکت‌های حاضر در بورس اوراق بهادار است. این پژوهش، برای بهینه‌سازی میانگین-واریانس سبد سهام بر اساس تحلیل سیگنال‌های تکنیکی از شاخص قدرت نسبی (RSI) از نوع شاخص‌های هدایت‌گر استفاده کرده و برای مدلسازی از دو روش تحلیل میانگین-واریانس مارکوفیتز تکنیکالی و الگوریتم جنگل تصادفی استفاده شده است. برای ارزیابی عملکرد مدل پیشنهادی، میانگین خطای مطلق (MAE)، درصد میانگین مطلق خطا (MAPE) و خطای میانگین مربع ریشه (RMSE) به کار رفته است. نتایج به دست آمده نشان می‌دهد که مقادیر شاخص‌های مطروحه برای شبکه مدل میانگین-واریانس تکنیکالی نسبت به الگوریتم RF از مقادیر کمتری برخوردار هستند. و این موضوع دقت بالاتر این شبکه عصبی بازگشتی نسبت به الگوریتم جنگل تصادفی را در مدلسازی و پیش‌بینی مقادیر بهینه‌سازی سبد سهام نشان می‌دهد. نتایج به دست آمده نشان می‌دهد که نمودار R^2 برای داده‌های مربوط به بهینه‌سازی سبد سهام برای مدل میانگین-واریانس تکنیکالی دارای مقادیر نزدیکتری نسبت به ۱ دارند در حالی که الگوریتم RF دارای مقادیر کمتری از R^2 برای داده‌های مربوط به مدل میانگین-واریانس تکنیکالی می‌باشد که این موضوع نشان دهنده دقت بالای مدل میانگین-واریانس تکنیکالی است.

واژه‌های کلیدی: بورس اوراق بهادار، جنگل تصادفی، سبد سهام، میانگین-واریانس تکنیکالی.

۱. دانشجوی دکتری مدیریت صنعتی، گروه مدیریت، واحد دهقان، دانشگاه آزاد اسلامی، دهقان، ایران؛

Leilaboroomandi@gmail.com

۲. استادیار، گروه مدیریت، واحد دهقان، دانشگاه آزاد اسلامی، دهقان، ایران. (نویسنده مسئول): sae_aghahi@yahoo.com

۳. استادیار، گروه مهندسی صنایع، واحد دهقان، دانشگاه آزاد اسلامی، دهقان، ایران، sharifidocument@gmail.com

مقدمه

زمانی که سرمایه‌گذار تصمیم می‌گیرد تا سرمایه خود را بین چند دارایی مالی تقسیم کند، مسئله انتخاب سبدسهم بهینه خود را نشان می‌دهد. سوال اساسی این است که سرمایه اولیه به چه نسبتی بین دارایی‌ها تقسیم گردد تا سرمایه‌گذار بتواند به هدف خود دست یابد. این هدف غالباً رسیدن به یک حداقل سطح بازده یا ثروت مشخص در ضمن تحمل کمترین ریسک ممکن می‌باشد و مدل سبدسهم مارکوویتز^۱ (مدل میانگین-واریانس) اولین بار به مدل‌سازی چنین مسئله‌ای پرداخت (ژانگ^۲، ۲۰۲۲). مدل‌های مختلفی از انتخاب بهینه سبدسهم وجود دارند که در تابع هدف، محدودیت‌ها، تک دوره‌ای بودن یا چند دوره‌ای بودن، وجود یا عدم وجود هزینه‌های معاملاتی و ... با هم متمایز هستند. پس از بسته شدن سبد، تغییرات قیمتی دارایی‌های سبد در گذر زمان رخ می‌دهد و برای سبد در انتهای افق زمانی سرمایه‌گذاری، بازدهی را حاصل می‌کند. این بازده ممکن است به علت تغییرات شدید قیمتی تعدادی از دارایی‌ها، از بازدهی که در مرحله نظری مورد انتظار بوده، فاصله زیادی داشته باشد. در واقع آنچه رخ داده فاصله زیاد پیش‌بینی تا واقعیت است و از این رو با نوعی عدم استواری در انتخاب سبدسهم مواجه هستیم. در کشور ما همواره بازار سرمایه با ریسک‌های زیادی خصوصاً ریسک سیاسی و تغییر قوانین مواجه است که این ریسک‌ها می‌تواند دارایی‌های سبد سرمایه‌گذاران را با تغییرات قیمتی شدید پیش‌بینی نشده‌ای مواجه کرده و احیاناً منجر به افت‌های شدیدی در ارزش سبدسهم گردد. در برنامه‌ریزی ریاضی معمولاً مسائل با پیش فرض قطعی بودن پارامترهای مدل بهینه می‌گردند. حال آنکه در دنیای واقعی اکثر داده‌ها دچار عدم قطعیت‌اند. پیش فرض اصلی برنامه‌ریزی‌های ریاضی توسعه مدل بر اساس داده‌های صریحاً معین و برابر با مقداری اسمی است (جی و همکاران^۳، ۲۰۲۲). در پژوهش حاضر اقدام به استفاده از مدل میانگین-واریانس با رویکرد تحلیل تکنیکال و الگوریتم جنگل تصادفی (RF) شده است. نتایج این بهینه‌سازی سبد سهام می‌تواند به سرمایه‌گذاران کمک کند تا ریسک سبد سرمایه‌گذاری خود را کنترل کنند. با استفاده از این مدل، سرمایه‌گذاران می‌توانند ترکیبی از سهام‌ها را انتخاب کنند که ریسک کل سبد را کاهش داده و بهینه‌ترین تراز بین ریسک و عایدی را فراهم کند. بنابراین، هدف پژوهش حاضر بهینه‌سازی سبد سهام با استفاده از روش میانگین-واریانس تکنیکالی و جنگل تصادفی است.

مبانی نظری و پیشینه پژوهش

بازارهای سرمایه یکی از مهمترین و پویاترین بخش‌های اقتصاد هر کشور است. در این بازار، ارزش‌های مالی و سرمایه‌ای مانند سهام، اوراق قرضه و مشتقات مالی، خرید و فروش می‌شوند. این بازار، به سرمایه‌گذاران و شرکت‌ها امکان می‌دهد تا منابع مالی خود را تأمین کنند و سرمایه خود را در سراسر اقتصاد گسترش دهند (بلجان^۴ و همکاران، ۲۰۲۲).

^۱Markowitz

^۲ Zhang

^۳ Ji et al.

^۴ Beljan

در میان بازارهای مالی، بازار بورس اوراق بهادار نقش بسیار مهمی را ایفا می‌نماید. زیرا این بازار، به عنوان یکی از اصلی‌ترین و پررونق‌ترین بازارهای سرمایه، عملکرد و تأثیر قابل توجهی در اقتصاد و سرمایه‌گذاری دارد. بازار بورس اوراق بهادار به عنوان یک محیط معاملاتی مطمئن و نظام‌مند برای خرید و فروش اوراق بهادار عمل می‌کند. این بازار فضایی فراهم می‌کند که سرمایه‌گذاران (شامل سرمایه‌گذاران حقیقی و حقوقی) بتوانند اوراق بهادار را به صورت عمومی و مشروع معامله کنند. این بورس امکان دسترسی آسان و تسهیل فرآیند خرید و فروش را به سرمایه‌گذاران می‌دهد (ژو^۱ و همکاران، ۲۰۲۲). اما نکته بسیار مهمی که در بازارهای سرمایه و خصوصاً در بازار بورس اوراق بهادار باید به آن توجه داشت، انتخاب یک سبد سهام مفید و موثر است. این سبد سهام بر اساس ترکیبی از رویکردهای مختلف شامل استراتژی سرمایه‌گذاری، هدف‌های مالی، زمانبندی سرمایه‌گذاری و تحلیل بازار انتخاب می‌شوند. بنابراین انتخاب یک سبد سهام قوی، نقش بسیار مهمی در تثبیت، تقویت و بهینه‌سازی دارایی‌های افراد سرمایه‌گذار دارد (آراشی^۲ و همکاران، ۲۰۲۲).

انتخاب یک سبد سهام نقش مهمی در میزان حفظ و توسعه سرمایه‌های افراد سرمایه‌گذار دارد. بهینه‌سازی سبد سهام در بازار بورس اوراق بهادار یک رویکرد بسیار مهم است تا سبد سهامی با عملکرد بهتر و ریسک کمتر را تشکیل داده و به سرمایه‌گذاران کمک کند تا بازدهی بهینه را از سرمایه خود کسب کنند (فرناندز^۳ و همکاران، ۲۰۱۹).

یکی از کاربردهای بهینه‌سازی سبد سهام، تنوع سرمایه‌گذاری است. با بهینه‌سازی سبد سهام، سرمایه‌گذاران می‌توانند تنوع بیشتری در سبد سهام خود ایجاد کنند. این به معنای سرمایه‌گذاری در صنایع و شرکت‌های مختلف، محصولات مالی مختلف یا مناطق جغرافیایی مختلف است. با داشتن تنوع در سبد سهام، سرمایه‌گذاران می‌توانند ریسک سرمایه‌گذاری خود را کاهش داده و از فرصت‌های بازار بهره‌برداری کنند (سوکونو^۴ و همکاران، ۲۰۲۴).

همچنین با بهینه‌سازی سبد سهام، می‌توان ریسک سیستماتیک (ریسک بازار) را کاهش داد. با ترکیب سهام‌هایی که در برابر ریسک‌های مختلف مستقل هستند، می‌توان ریسک کلی سبد سهام را کاهش داده و به ازای هر سطح بازار، عملکرد بهتری را نسبت به یک سهم تکی داشت. با استفاده از روش‌های بهینه‌سازی، می‌توان سبد سهامی را تشکیل داد که در دوره‌های طولانی‌تر بازدهی بهتری داشته باشد. با توجه به تاریخچه قیمت و اطلاعات مالی شرکت‌ها، می‌توان به سرمایه‌گذاری در سهام‌هایی که انتظار می‌رود در آینده عملکرد بهتری داشته باشند، تمرکز کرد (پوروانداری^۵ و همکاران، ۲۰۲۳).

^۱ Xu^۲ Arashi^۳ Fernandez^۴ Sukono^۵ Purwandari

در کل، بهینه‌سازی سبد سهام در بازار بورس اوراق بهادار به سرمایه‌گذاران کمک می‌کند تا تنوع، کاهش ریسک، بهبود بازدهی و رعایت استراتژی‌های سرمایه‌گذاری خود را در نظر بگیرند. برای دستیابی به این هدف، معمولاً از روش‌های مختلف بهینه‌سازی مانند بهینه‌سازی مارکوویتز و استفاده از مدل‌های ریاضی و روش‌های شبیه‌سازی استفاده می‌شود (چاول و چاوزیدویا^۱، ۲۰۱۹). همچنین انتخاب سهام ناپراکنده یا در صنایع ضعیف می‌تواند منجر به کاهش بازدهی سرمایه‌گذاری شود. عدم تنوع کافی در سبد سهام ممکن است باعث افزایش ریسک و کاهش بازدهی کلی سرمایه‌گذاری شود. از سوی دیگر، انتخاب سهام بدون تحلیل کافی از عملکرد شرکت‌ها و زمینه فعالیت آنها می‌تواند به ناپایداری سبد سهام و عدم استحکام در طول زمان منجر شود (ادریس^۲ و همکاران، ۲۰۱۹). همچنین اگر سبد سهام انتخاب شده تطابق کافی با اهداف و نیازهای سرمایه‌گذار نداشته باشد، سرمایه‌گذار ممکن است به دستیابی به اهداف مالی خود نرسد و ناامید شود. عدم دقت در انتخاب سبد سهام می‌تواند آسیب‌پذیری سرمایه‌گذار را در برابر تغییرات ناگهانی در شرایط بازار و نوسانات قیمتی بورس افزایش دهد (شاه^۳ و همکاران، ۲۰۱۹).

پیشینه پژوهش

تاکنون پژوهش‌های مختلفی با هدف بررسی و بهینه‌سازی سبد سهام انجام شده است. آدینه‌وند و همکاران (۱۴۰۲) در پژوهشی دیگر، به بررسی کارآمدی مدل‌های بهینه‌سازی فرا ابتکاری الگوریتم ژنتیک چندهدفه تحت معیار ریسک *MSV* و الگوریتم ازدحام ذرات تحت معیار ریسک *CVaR* در تعیین سبد سهام شرکت‌های پذیرفته شده در سازمان بورس اوراق بهادار اقدام کردند. نتایج نشان می‌دهد که مدل الگوریتم ژنتیک چند هدفه تحت معیار ریسک *MSV* دارای بازده بیشتر و ریسک کمتری می‌باشد، در نتیجه مدل الگوریتم ژنتیک تحت معیار ریسک *MSV* از مدل الگوریتم ازدحام ذرات تحت معیار ریسک *CVaR* کارآمدتر می‌باشد. همچنین جعفری و همکاران (۱۴۰۲) بهینه‌سازی سبد سهام با استفاده از مدل ارزش در معرض خطر شرطی با رویکرد شبیه‌سازی را بررسی کردند. با بررسی نتایج تخمین‌های بدست آمده از مدل‌ها با داده‌های بازار این نتیجه حاصل می‌شود که این مدل نتایجی نزدیکتر به واقعیت برای سرمایه‌گذاران ایجاد می‌کند.

آنونو^۴ و همکاران (۲۰۲۴) در پژوهشی آزمایش بهینه‌سازی پرتفوی توسط استراتژی سرمایه‌گذاری پرتفوی در بازار سهام را بررسی کردند. ارزیابی پرتفوی بهینه با استفاده از پارامترهای متعدد نشان داد که بیشترین بازده مورد انتظار و کمترین ریسک به ترتیب ۱/۲۲ و ۳/۱۲ درصد بوده است. جینگ^۵ و همکاران (۲۰۲۳) در پژوهشی دیگر، انتخاب بهینه پرتفوی سهام با استفاده از روش‌های تصمیم‌گیری چند معیاره را مورد بررسی قرار دادند. از شاخص موجود در بورس اوراق بهادار تهران می‌توان برای ارائه مدلی جامع و بهینه برای سبد سهام استفاده کرد. روش‌های مختلف تصمیم‌گیری چند شاخصی همگی می‌توانند سبد سهام

^۱ Chavalle and Chavez-Bedoya

^۲ Idrees

^۳ Shah

^۴ Anuno

^۵ Jing

بهینه و بهترین سبد سهام را برای بالاترین بازدهی شناسایی کنند. بریوآگا^۱ و همکاران (۲۰۲۳) در پژوهشی بهینه‌سازی پورتفولیو با استفاده از مدل درخت پوشای حداقل در بازار بورس مراکش را ارزیابی کردند. یافته‌ها نشان می‌دهد که روش پیشنهادی می‌تواند عملکرد پرتفوی را افزایش دهند، شواهدی از بازار سرمایه که می‌تواند به سرمایه‌گذاران یا مدیران فعال در هنگام بهینه‌سازی پرتفوی کمک کند.

ژو^۲ و همکاران (۲۰۲۳) در پژوهش خود به این نتیجه رسیدند که مدل بهینه‌سازی میانگین - واریانس از ریاضیات پایه استفاده می‌کند و محدودیت‌های ریاضی را در نظر می‌گیرد. این مدل قادر است به طور دقیق، محدودیت‌های ریاضی را در رابطه با ترکیب بهینه از دارایی‌ها براساس اهداف ترسیمی در نظر بگیرد. پارک^۳ (۲۰۲۱) بیان داشت که تحلیل سیگنال‌های تکنیکی امکان تجزیه و تحلیل دقیق‌تر بازار را در زمان واقعی فراهم می‌کند. با استفاده از این تحلیل، می‌توان الگوها، روندها و سیگنال‌های بازار را تشخیص داد و در مدل بهینه‌سازی میانگین - واریانس لحاظ کرد. تحلیل سیگنال‌های تکنیکی در این مدل می‌تواند الگوها و سیگنال‌های قیمتی را در داده‌های سهام تشخیص دهد. این الگوها ممکن است شامل سیگنال‌های خرید یا فروش، تغییر روندهای قیمتی، سطوح پشتیبانی و مقاومت و سایر الگوهای فنی باشند. با شناخت این الگوها، می‌توان سبد سهام را براساس سیگنال‌های قوی و قابل اعتماد بهبود داد. همچنین فبریانتي^۴ و همکاران (۲۰۲۳) در یافتند که تحلیل سیگنال‌های تکنیکی در این مدل می‌تواند به تعیین زمان ورود و خروج از سهام کمک کند. با تحلیل نمودارهای قیمتی، می‌توان نقاط ورود و خروج استراتژی‌های سرمایه‌گذاری را تعیین کرد. برای مثال، می‌توان نقاط ورود را در زمانی تعیین کرد که سیگنال‌های خرید قویتری وجود دارد و نقاط خروج را در زمانی که سیگنال‌های فروش قویتری مشاهده می‌شود. با استفاده از تحلیل سیگنال‌های تکنیکی در این مدل، می‌توان ریسک سبد سرمایه را پایش کرده و مدیریت کرد. مورتاس^۵ و همکاران (۲۰۲۲) در پژوهش خود به این نتیجه رسیدند که با توجه به سیگنال‌های تکنیکی، می‌توان میزان واریانس و نوسانات بازدهی سبد سهام را پیش‌بینی کرد و بر اساس آن، ترکیب بهینه از دارایی‌ها را تعیین کرد.

استفاده از ابزارهای تحلیل تکنیکال در بازار سرمایه به سرعت در حال گسترش است و این ابزارها امکان سنجش روزانه بازار را فراهم می‌کنند. تحلیل تکنیکال به سرمایه‌گذاران کمک می‌کند تا زمان ورود و خروج خود را مشخص کنند و با بررسی الگوها و شاخص‌ها، نقاط ورود مطمئن‌تری را شناسایی کنند. همچنین، این تحلیل به شناسایی روندهای بازار (صعودی، نزولی یا جانبی) کمک کرده و ریسک‌ها را کاهش می‌دهد (نور و همکاران^۶، ۲۰۲۳). در این پژوهش، بهینه‌سازی سبد سهام با استفاده از الگوریتم جنگل تصادفی (RF) انجام شده است. این الگوریتم یک روش یادگیری ماشینی نظارت شده است که از ترکیب چندین مدل

^۱ Berouaga

^۲ Xu

^۳ Park

^۴ Febrianti

^۵ Mourtas

^۶ Nor

تصمیم‌گیری ساده (درخت تصمیم) برای پیش‌بینی استفاده می‌کند. هر درخت تصمیم به صورت مستقل از داده‌ها آموزش دیده و در نهایت نتایج این درختان با هم ترکیب می‌شوند تا پیش‌بینی نهایی صورت گیرد (سه و همکاران^۱، ۲۰۱۸).

روش شناسی پژوهش

روش پژوهش حاضر بر اساس هدف از نوع کاربردی و از نظر نوع داده‌ها، پس رویدادی می‌باشد. جامعه آماری این پژوهش شامل کلیه شرکت‌های فعال در بورس اوراق بهادار تهران طی سال‌های ۱۳۹۰ الی ۱۴۰۲ می‌باشد. در تحقیق حاضر تعداد ۴۰ سهم از سهام فعال در بازار سهام که داده‌های پیوسته و کامل داشته‌اند انتخاب شده است. بازه زمانی مورد نظر از تاریخ ۱ فروردین ۱۳۹۰ تا ۲۹ اسفند ۱۴۰۲ بوده است. به منظور دستیابی به نوسانات قابل توجه بازه زمانی برای محاسبه ریسک و بازه دوره‌های سه روز کاری در نظر گرفته شده است. بر این اساس ۴۱۰ مشاهده برای هر سهم که هر ده روز کاری را می‌توان برای ایجاد یک سبد سهام بکار گرفت. در واقع میانگین و واریانس یک ماهه برای تشکیل هر پرتفوی بکار گرفته می‌شود. داده یک ماه آخر یعنی پرتفوی آخر بر خلاف دیگر مشاهدات تنها به منظور محاسبه بازده یک ماه پس از پرتفوی ۴۰-ام بکار گرفته شد و در تابع بهینه‌یابی وارد نمی‌شود. برای بررسی بهینه‌سازی سبد سهام از اطلاعات قیمت سهام ۵ شرکت پذیرفته شده در بورس اوراق بهادار استفاده شد. ملاک انتخاب این ۵ شرکت بالا بودن میزان میانگین قیمت، میانگین بازده و تعداد روزهای معامله شده نسبت به سایر شرکت‌های حاضر در بورس اوراق بهادار است. اطلاعات مربوط به این شرکت‌ها در جدول ۱ ارائه شده است.

جدول ۱- اطلاعات قیمت ۵ سهم پذیرفته شده در بورس اوراق بهادار تهران

شماره سهم	میانگین قیمت	میانگین بازده	انحراف معیار بازده (6)	نیم‌انحراف معیار بازده (Σ)	تعداد روزهای معامله شده
۱	۴۹۱۲/۳	-۰/۰۰۰۹	۰/۰۲۶۱	۰/۰۲۰۴	۲۴۵۰
۲	۱۳۵۲/۵	-۰/۰۰۱۱	۰/۰۳۵۲	۰/۰۲۱۹	۲۱۳۶
۳	۲۵۵۳/۶	-۰/۰۰۰۸	۰/۰۱۱۲	۰/۰۰۹۸	۲۵۱۱
۴	۴۳۸۲/۱	-۰/۰۰۱۰	۰/۰۲۴۶	۰/۰۱۶۸	۲۰۶۴
۵	۱۹۸۲/۵	-۰/۰۰۰۵	۰/۰۱۱۹	۰/۰۱۰۱	۲۲۷۸

ماخذ: یافته‌های تحقیق

^۱ Seh

متغیر مورد بررسی

در این پژوهش، برای بهینه‌سازی میانگین - واریانس سید سهام بر اساس تحلیل سیگنال‌های تکنیکی از شاخص قدرت نسبی^۱ (RSI) از نوع شاخص‌های هدایت‌گر استفاده شده است. $RSI_{t,p}$ که همان شاخص قدرت نسبی در زمان t و مربوط به دوره p می‌باشد تنها با استفاده از قیمت‌های بسته شدن بدست می‌آید:

الف) ابتدا تغییر قیمت روزانه را حساب کرده و اگر مقدار آن مثبت بود آن را در ستون افزایش‌ها نوشته و اگر مقدار آن منفی بود، در ستون کاهش‌ها نوشته می‌شود.

ب) سپس میانگین افزایش‌ها برای یک دوره ۱۴ روزه حساب می‌گردد. در محاسبات روزانه از میانگین‌های ۱۴ روزه و در محاسبات هفتگی از میانگین‌های ۱۴ هفته‌ای استفاده می‌شود، در این پژوهش نیز از میانگین‌های ۱۴ روزه استفاده شده است. برای به دست آوردن مقدار میانگین افزایش میزان تغییرات ($U_{t,p}$) در ۱۴ روز گذشته روزهای مثبت قیمت ($up-Closes$) یا (U_i) را با هم جمع کرده (منظور از روزهای مثبت روزهایی است که قیمت سهام افزایش داشته است و حاصل بر ۱۴ تقسیم می‌شود. همچنین میانگین کاهش‌ها را برای همان دوره محاسبه کرده برای محاسبه میانگین کاهش میزان تغییرات ($D_{t,p}$) در ۱۴ روز گذشته روزهای منفی قیمت ($down Closes$ یا D_i) را با هم جمع کرده و حاصل بر عدد ۱۴ تقسیم می‌شود. بدین ترتیب اولین عددی که بدست می‌آید برای روز چهاردهم است و برای روز پانزدهم به بعد از فرمول دیگری استفاده می‌شود که بر خلاف یک میانگین ساده وزن آخرین عدد را بیشتر منظور می‌کند. بدین ترتیب به قیمت آخرین روز بیشتر تاکید شده و باعث بالا رفتن حساسیت شاخص می‌شود. لازم به ذکر است که این فرآیند برای همه روزها انجام می‌شود و بعد از روز چهاردهم برای کلیه روزها دارای میانگین افزایش‌ها و میانگین کاهش‌ها خواهد بود.

RSI نوسان‌نمایی به صورت $RSI_t = \{i: t - p \leq i \leq t\}$ می‌باشد.

$$D_i = \begin{cases} C_{i-1} - C_i & \text{if } C_{i-1} > C_i \\ 0 & \text{otherwise} \end{cases} \text{ and } U_i = \begin{cases} C_i - C_{i-1} & \text{if } C_i > C_{i-1} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

که برای هر $i \in I_{t,p}$ یک قیمت بسته شدن برای دوره i است.

$$D_{t,p} = I_{t,p} \text{ و } U_{t,p} = I_{t,p} \text{ میانگین } D_i \text{ بالاتر از } I_{t,p} \text{ و میانگین } U_i \text{ بالاتر از } I_{t,p} \quad (2)$$

ج) حال باید قدرت نسبی را حساب کرد برای این کار باید برای هر روز میانگین افزایش‌ها را بر میانگین کاهش‌های آن روز تقسیم کرد (بدون در نظر گرفتن علامت). بنابراین قدرت نسبی (RS) عبارت است از حاصل تقسیم میانگین افزایش بر میانگین کاهش.

$$RS_{t,p} = \frac{U_{t,p}}{D_{t,p}} = \text{قدرت نسبی روز } t \text{ ام} \quad (3)$$

د) با داشتن قدرت نسبی، شاخص قدرت نسبی را حساب کرده و عدد به دست آمده برای RS در فرمول RSI قرار داده شد:

^۱ Relative Strength Index

$$RSI_{t,p} = 100 - \frac{100}{1 + RSI_{t,p}} \quad (4)$$

مقدار $RSI=100$ دلالت بر حرکت روند قیمت به طرف بالا دارد یعنی بازار با حالت خرید بیش از حد مواجه است و زمان مناسبی برای فروش است و مقدار $RSI=0$ دلالت بر یک کاهش خالص در حرکات قیمت دارد یعنی بازار با حالت فروش بیش از حد مواجه است و زمان مناسبی برای خرید است (دیازا و همکاران، ۲۰۲۲).

روش تحلیل میانگین-واریانس مارکویتز تکنیکالی

روش تحلیل میانگین-واریانس مارکویتز تکنیکالی بر اساس دو مفهوم اصلی میانگین و واریانس قیمت‌ها، قصد دارد با مقایسه تغییرات و ارتباط بین آن‌ها در بازه‌های زمانی مختلف، الگوها و سیگنال‌هایی را برای تصمیم‌گیری در معاملات مالی ارائه دهد. با توجه به میانگین و واریانس، الگوها و تغییراتی در رفتار قیمت‌ها شناسایی می‌شوند که می‌تواند به عنوان سیستمی برای تصمیم‌گیری در معاملات مورد استفاده قرار گیرد. برای به دست آوردن انتخاب سبد سرمایه بهینه در روش میانگین-واریانس مارکویتز تکنیکالی که حداقل واریانس برای سطح خاصی از بازده است، مدل برنامه‌ریزی خطی زیر مورد استفاده قرار گرفته است (چاووی و انسون و چایسیری، ۲۰۲۲):

$$\begin{aligned} \text{Min } z &= \delta_p^2 \\ \text{St : } r\bar{p} &= \sum_{j=1}^n w_j \cdot r_j \\ \sum_{j=1}^n w_j &= 1 \\ w_j &> 0 \end{aligned} \quad (5)$$

در رابطه بالا داریم:

w_i = وزن مربوط به سهم i ام در سبد سهام؛

$r\bar{p}$ = بازدهی مورد انتظار سبد سهام؛

r_i = بازدهی سهم i ام؛

δ_p^2 = واریانس بازده سبد سهام

واریانس بازده سبد سهام طبق رابطه زیر محاسبه می‌شود:

$$\delta_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j cov(r_i, r_j) \quad (6)$$

^۱ Díaz

^۲ Chaweeewanchon and Chaysiri

الگوریتم جنگل تصادفی

جنگل تصادفی یکی از روش‌های درخت تصمیم است. RF مجموعه از روش‌های یادگیری است که می‌تواند به منظور رگرسیون و طبقه‌بندی داده‌ها از آن استفاده کرد. اصل اساسی روش شناسی در جنگل تصادفی، ساختن جنگلی از درختان تصادفی است که از این طریق در هر گره C and یک درخت تصمیم ایجاد می‌شود. RF استحکام چندین درخت منفرد را با هم ادغامی کند تا با استفاده از رویکرد بدست آمده و جدیی، مدل دقیق تری را بسازد. مطالعات بسیاری در مهندسی کاربرد RF را مورد بررسی قرار دادند. همچنین در فرآیند پیش‌بینی، کاربرد و قابلیت روش جنگل تصادفی را مورد پژوهش قرار دادند. تحت روش بوت استرپینگ^۱، داده‌ها به طور مستقل و تصادفی برای مرحله آموزش انتخاب می‌شوند. در فرآیند انتخاب متغیر، مدل داده‌های آزمایشی را کنار گذاشته که آن‌ها را به عنوان نمونه‌های خارج از کیسه (OOB)^۲ نامیده می‌شوند. با توجه به طبقه‌بندی جنگل تصادفی که شامل چندین پارامتر مانند تعداد درخت، حداقل بهره و حداکثر عمق درخت می‌بایست این روش بهینه شود (سکولیچ^۳ و همکاران، ۲۰۲۰).

پیاده‌سازی مدل کدنویسی در نرم‌افزار متلب انجام شده است.

یافته‌های پژوهشی

در ارائه مدل برای بهینه‌سازی سبد سهام با رویکرد میانگین-واریانس تکنیکالی باید به این نکته توجه داشت که این مدل در مورد انتخاب وزن پرتفوی است که مبادله بهینه بین میانگین (به عنوان معیار سود) و واریانس (به عنوان معیار ریسک) بازده پرتفوی را برای یک دوره آتی فراهم می‌کند. برای پورتفولیوی متشکل از m دارایی از سهام‌های مورد بررسی با بازده مورد انتظار μ_i ، باید مقدار w_i وزن ارزش سبد سرمایه گذاری شده در دارایی i باشد به طوری که رابطه $\sum_{i=1}^m w_i = 1$ و $w = (w_1, \dots, w_m)^T$ ، $\mu = (\mu_1, \dots, \mu_m)$ در طول مدلسازی برقرار باشد. بازده پرتفوی دارای میانگین $w^T \mu$ و واریانس $w^T \Sigma w$ است که ماتریس کوواریانس بازده دارایی را تعیین می‌کند. در این مدل با توجه به مقدار هدف μ برای میانگین بازده یک سبد، یک پورتفولیوی کارآمد با وزن بردار w_{eff} مشخص می‌کند که چگونه روند بهینه‌سازی سبد سهام طی شود که مطابق رابطه زیر داریم:

$$w_{eff} = \arg \min w^T \Sigma w ; w^T \mu = \mu^*, w^T 1 = 1, w \geq 0 \quad (7)$$

هنگامی که فروش کوتاه مجاز است، محدودیت $w \geq 0$ (یعنی $w \geq 0$ برای همه i) در رابطه بالا را می‌توان حذف کرد و مشکل زیر را ایجاد کرد که یک راه حل صریح دارد:

^۱ Bootstrapping method

^۲ Out Of Bag

^۳ Sekulić

$$w_{eff} = \arg_{w: w^T 1 = 1, w^T \mu = \mu^*} \min w^T \Sigma w = \{B \Sigma_{m=1}^{-1} 1 - A \Sigma_{m=1}^{-1} \mu + \mu^* (C \Sigma_{m=1}^{-1} \mu - A \Sigma_{m=1}^{-1} 1) / D\} \quad (8)$$

که در رابطه بالا:

$$A = \mu^T \Sigma_{m=1}^{-1} 1 = 1^T \Sigma_{m=1}^{-1} \mu \quad (9)$$

$$B = \mu^T \Sigma_{m=1}^{-1} \mu \quad (10)$$

$$C = 1^T \Sigma_{m=1}^{-1} 1 \quad (11)$$

$$D = BC - A^2 \quad (12)$$

در مدل مارکویتز دو پارامتر μ و Σ به عنوان پارامترهای شناخته شده فرض می‌شوند. از آنجایی که در عمل μ و Σ ناشناخته هستند، یک رویکرد رایج مورد استفاده تخمین μ و Σ از روی داده‌های تاریخی است. مقدار A یک مدل استاندارد برای قیمت P_{it} از دارایی i و در زمان t است که در تئوری مالی، حرکت هندسی براونی $dP_{it}/P_{it} = \theta_i dt + \sigma_i dB_t^{(i)}$ است، که در آن $\{B_t^{(i)} \geq 0\}$ حرکت براونی استاندارد است.

برای محاسبه بردار وزن بهینه در رویکرد میانگین-واریانس تکنیکالی فرض شده است که μ_n و V_n ماتریس میانگین خلفی و گشتاور دوم با توجه به مجموعه R_n بازده فعلی و مجموعه $r_1, \dots, r_n$ باشد. از آنجایی که w بر اساس R_n است، از $E(r_{n+1}|R_n) = \mu_n$ و $E(r_{n+1} + r_{n+1}^T | R_n) = V_n$ نتیجه می‌شود که:

$$E(W^T r_{n+1}) = E(W^T \mu_n) \quad (13)$$

$$E\{(W^T r_{n+1})^2\} = E(W^T V_n W) \quad (14)$$

بدون فروش کوتاه، بردار وزن $w(\eta)$ توسط آنالوگ زیر ارائه شده است.

$$w(\eta) = \arg_{w: w^T 1 = 1, w > 0} \min \{ \lambda W^T V_n w - \eta W^T \mu_n \} \quad (15)$$

که این رابطه با برنامه نویسی درجه دوم محاسبه شده است (توسط *quadprog* در متلب). هنگامی که فروش کوتاه مجاز است، محدودیت $w \geq 0$ را می‌توان حذف کرد و $w(\eta)$ را به صراحت توسط رابطه زیر محاسبه کرد:

$$w(\eta) = \arg_{w: w^T 1 = 1} \min \{ \lambda W^T V_n w - \eta W^T \mu_n \} = \frac{1}{C_n} V_n^{-1} 1 + \frac{\eta}{2\lambda} V_n^{-1} (\mu_n - \frac{A_n}{A_c} 1) \quad (16)$$

که در آن برابری دوم را می‌توان با استفاده از ضریب لاگرانژ و رابطه زیر محاسبه می‌شود:

$$A_n = \mu_n^T V_n^{-1} 1 = 1^T V_n^{-1} \mu_n; \quad B_n = \mu_n^T V_n^{-1} \mu_n; \quad C_n = 1^T V_n^{-1} 1 \quad (17)$$

لازم به ذکر است که در روابط ۱۵، ۱۶ و ۱۷ اساساً تخمین‌های بایس از μ و $V = \Sigma + \mu \mu^T$ را به بردار وزن بهینه که μ و Σ را فرض می‌کند و باید شناخته شود متصل می‌کند. با این حال، بر خلاف مرز کاراً، ابتدا مسئله بهینه‌سازی نمونه کارها با میانگین واریانس اصلی به یک مسئله بهینه‌سازی "میانگین در مقابل لحظه دوم" تبدیل شد که دارای یک پارامتر اضافی η است.

$$C(\eta) = E\{W^T(\eta) \mu_n\} + \lambda E\{(W^T(\eta) \mu_n)^2\} - \lambda E\{W^T(\eta) V_n w(\eta)\} \quad (18)$$

که برابر با $E[w^T(\eta)r] - \lambda \text{Var}[w^T(\eta)r]$ در ۲۰ است، می‌توان از روش برنت^۱ برای به حداکثر رساندن $C(\eta)$ استفاده کرد. لازم به ذکر است که این استدلال به طور ضمنی فرض می‌کند که ماکزیمم ۱۷ با مقداری w بدست می‌آید و متناهی است. این فرض زمانی به آن استناد می‌شود که عدم وجود فروش کوتاه مدت برآورده شده باشد و زمانی که فروش کوتاه مدت مجاز باشد، به این فرض استناد نمی‌شود. بنابراین حداکثر مقدار بازدهی از رابطه زیر به دست می‌آید:

$$E\left\{\left(Bn - \frac{A_n^2}{c_n}\right)\left(Bn - \frac{A_n^2}{c_n} - 1\right)\right\} < 0 \quad (۱۹)$$

در مورد $0 > E\left\{\left(Bn - \frac{A_n^2}{c_n}\right)\left(Bn - \frac{A_n^2}{c_n} - 1\right)\right\}$ ، به جای آن یک حداقل دارد و به ۱ به عنوان $|\eta|$ نزدیک می‌شود. ۱ در این مورد، دارای یک مقدار بی نهایت است و باید به جای حداکثر به عنوان یک مقدار برتر (که به دست نمی‌آید) تعریف شود. Σn نشان دهنده ماتریس کوواریانس خلفی داده شده Rn باشد. باید توجه داشت که قانون انتظارات مشروط تکرار شده، که در رابطه ۱۸ و ۱۹ آمده است، آنالوگ زیر را برای $\text{Var}(W)$ دارد:

$$\text{Var}(w) = E\{\text{Var}(W|R_n)\} + \text{Var}\{E(W|R_n)\} = E(w^T \Sigma w) + \text{Var}(w^T \mu_n) \quad (۲۰)$$

استفاده از n برای جایگزینی در بردار وزن بهینه که μ را فرض می‌کند و شناخته می‌شود. بنابراین، واریانس $w^T \mu_n$ در رابطه فوق را نادیده می‌گیرد، و این حذف یک دلیل اصلی مهم برای معمای بهینه‌سازی مارکوفیتز مربوط به مرزهای کارآمد است. برای مدل‌سازی انعطاف‌پذیرتر، می‌توان به توزیع قبلی در رویکرد بیزی اجازه داد که شامل فرآیندهای نامشخص باشد، که می‌تواند از نمونه آموزشی با حداکثر احتمال، با روش لحظه‌ها یا روش‌های دیگر تخمین زده شود. به عنوان مثال، برای مزدوج قبلی بیز و برآوردهای انقباضی، می‌توان v و φ را توابعی از فرآیندهای خاص فرض کرد که با یک مدل چند عاملی مرتبط هستند. این به معنای استفاده از یک مدل تجربی بیز برای (μ, Σ) در مسئله بهینه‌سازی تصادفی است. برای ارزیابی (w^T) می‌توان از یک روش ناپارامتریک برای ارزیابی این مقادیر استفاده کرد که در زیر توضیح داده شده است.

باید توجه داشت که می‌توان عملکرد مکرر بیز یا سایر قوانین تخصیص دارایی را با استفاده از روش بوت استرپ ارزیابی کرد. نمونه‌های بوت استرپ $\{r_{b1}^*, \dots, r_{bn}^*\}$ با جایگزینی از نمونه مشاهده شده $\{r_1, \dots, r_n\}$ ، $1 \leq b \leq B$ ترسیم شده است، می‌تواند برای تخمین $E_{\mu, \Sigma}(w^T r_{n+1}) = E_{\mu, \Sigma}(w_n^T \mu) + \text{Var}_{\mu, \Sigma}(w_n^T \mu)$ و $E_{\mu, \Sigma}(w_n^T r_{n+1}) = E_{\mu, \Sigma}(w_n^T \mu)$ از پورتفولیوهای مختلف که بردارهای وزن آنها مختلف است، استفاده کند.

بسط مدل‌های سری زمانی بازده در رویکرد میانگین-واریانس تکنیکالی

یک فرض مهم در اصلاح رویکرد میانگین-واریانس تکنیکالی این است که r_t با میانگین μ و ماتریس کوواریانس Σ دارای ارتباط است. بررسی‌های تشخیصی میزان نقض این فرض باید در عمل انجام شود. تئوری بهینه‌سازی تصادفی در بخش قبل

^۱ Brent

در واقع به این فرض نیاز ندارد و فقط به میانگین خلفی و ماتریس گشتاور دوم بردار بازگشتی برای دوره بعدی نیاز دارد. بنابراین بر این اساس بررسی های تشخیصی نشان می دهد چنین تغییراتی مورد نیاز است. یک روش ساده برای معرفی چنین اصلاحی، استفاده از مدل رگرسیون تصادفی فرم است که در ادامه عبارت است از:

$$r_{it} = \beta_i^T x_{it} + \varepsilon_{it} \quad (21)$$

که در آن اجزای x_{it} شامل ۱، متغیرهای عاملی مانند بازده پرتفوی بازار مانند **S&P500** در زمان $t-1$ ، و متغیرهای با تاخیر $r_{i,t-1}, r_{i,t-2}, \dots$ است. ایده اصلی زیربنای رابطه بالا این است که متغیرهای کمکی را معرفی کند (از جمله متغیرهای تاخیری برای در نظر گرفتن اثرات سری زمانی) به طوری که خطاهای ε_{it} را بتوان به عنوان یک رویه تکنیکالی در نظر گرفت. پارامتر رگرسیون β_i^T را می توان با روش گشتاورها که معادل حداقل مربعات است تخمین زد. همچنین می توانیم با فرض اینکه $\varepsilon_{it} = s_{it}(y_i)z_{it}$ با میانگین ۰ و واریانس ۱ ناهمگونی را رفع کرد. y_i بردار پارامتری است که می توان آن را با حداکثر احتمال یا روش تعمیم یافته گشتاورها تخمین زد و s_{it} یک تابع معین است که به $r_{i,t-1}, r_{i,t-2}, \dots$ بستگی دارد. مثال شناخته شده مدل **GARCH(1,1)** است که عبارت است از:

$$\varepsilon_{it} = s_{it}z_{it} \quad (22)$$

$$s_{it}^2 = \omega_i + a_i s_{it}^2 + b_i r_{i,t-1}^2 \quad (23)$$

که در این رابطه $y_i = (\omega_i, a_i, b_i)$ است.

چنانچه مدل رگرسیون تصادفی ۲۳ در نظر گرفته شود. همانطور که اشاره شد، یک عنصر کلیدی در بردار وزن بهینه که مسئله بهینه سازی را حل می کند (μ_n, V_n) است، که در آن $\mu_n = E(r_{n+1}|R_n)$ و $V_n = E(r_{n+1}r_{n+1}^T | R_n)$ است. به جای مدل کلاسیک بازده، می توان دانش دامنه m دارایی ها را با مدل سازی سری های زمانی ترکیب کرد تا پیش بینی کننده های بهتری از بازده های آتی از طریق μ_n و V_n به دست آورد.

رگرسیون های خروجی در رابطه ۲۱ را می توان برای ساخت یک مدل تجربی-ماهوی ترکیبی برای پیش بینی انتخاب کرد. باید توجه داشت که روابط ۲۱، ۲۲ و ۲۳ بازده دارایی را به طور جداگانه مدل می کند. این روابط به جای اینکه به طور مشترک در یک مدل رگرسیون چند متغیره یا چند متغیره **GARCH** که مشکل داشتن پارامترهای بیش از حد برای برآورد را دارد، بازده را جداگانه بررسی می کنند. در حالی که بردارهای ε_{it} (یا z_{it}) در روابط ۲۲ و ۲۳ اجزای آن ها را نامرتبب فرض نمی کند، زیرا مؤلفه ها را به جای مشترکاً، به طور مجزا بررسی می کند. کوواریانس مقطعی مشروط بین بازده دارایی های i و j با R_n توسط رابطه زیر به دست می آید:

$$\text{Cov}(r_{i,n+1}, r_{j,n+1} | R_n) = s_{i,n+1}(y_i) s_{j,n+1}(y_j) \text{Cov}(z_{i,n+1}, z_{j,n+1} | R_n) \quad (24)$$

برای مدل ۲۱، ۲۲ و ۲۳ باید توجه داشت که در روابط مذکور، $s_{i,n+1}^2$ را به صورت بازگشتی از R_n تعیین می‌کند و z_{n+1} مستقل از R_n است و بنابراین ماتریس کوواریانس آن را می‌توان به طور پیوسته از باقیمانده‌های z_t تخمین زد. در روابط فوق رویکرد تکنیکالی از فرمول‌های زیر برای μ_n و V_n در روابط ۲۲ و ۲۳ استفاده می‌کند:

$$\mu_n = (\beta_1^T x_{1,n+1}, \dots, \beta_m^T x_{m,n+1})^T \quad (25)$$

$$V_n = \mu_n + \mu_n^T (s_{i,n+1}, s_{j,n+1}, \sigma_{i,j})_{1 \leq i,j \leq n} \quad (26)$$

که در آن β_i حداقل مربعات تخمین β_i است، و $s_{i,n+1}$ و σ_{ij} برآوردهای معمول $s_{i,n+1}$ و $Cov(z_{i,1}, z_{j,1})$ بر اساس R_n هستند.

مدلسازی داده‌های بهینه‌سازی سبد سهام با الگوریتم جنگل تصادفی

الگوریتم جنگل تصادفی یک الگوریتم گروهی با مجموعه‌های از درختان تصمیم است. دقت طبقه‌بندی روش جنگل تصادفی با ساخت مجموعه‌ای از درختان و رأی‌گیری بین آنها برای به دست آوردن رده‌ای با بیشترین تعداد رأی، پیشرفت‌های قابل توجهی داشته است. دو ویژگی مهم در ساخت جنگل‌های تصادفی، روش بگینگ^۱ و انتخاب تصادفی در هر گره است. در ادامه، ویژگی‌ها و خصوصیات جنگل تصادفی مورد استفاده در این پژوهش توضیح داده شده است.

روش بگینگ

روش بگینگ توسط لئو بریمن در سال ۱۹۹۶ مطرح شد. این روش یک فرا الگوریتم بر مبنای مفاهیم خود راه انداز و ترکیب، برای بهبود یادگیری ماشین است. الگوریتم‌های گروهی در یادگیری ماشین، چند یادگیرنده ضعیف را ترکیب می‌کنند تا به یک یادگیرنده قوی دست یابند. این روش از بیش برآزش داده‌ها جلوگیری می‌کند. در بگینگ نتایج خوب زمانی تولید می‌شود که طبقه‌بندهای پایه جزء الگوریتم‌های یادگیری ناپایدار باشند (مانند درخت تصمیم‌گیری یا شبکه عصبی)، به طوری که تغییرات کوچک در داده‌های آموزشی منجر به تغییرات عمده‌ای در مدل ساخته شده توسط آن الگوریتم شود. فرآیند الگوریتم بگینگ در پژوهش حاضر بدین شرح می‌باشد:

یک مجموعه آموزش از داده‌های سبد سهام با نام D و به اندازه $m=50$ تشکیل شد. بگینگ با نمونه‌گیری یکنواخت و با جایگزینی نمونه‌ها از D ، مجموعه آموزشی جدید با اندازه اولیه m تولید می‌کند. نمونه‌گیری با جایگزینی این امکان را می‌دهد که در هر Di بعضی از نمونه‌ها امکان تکرار داشته باشند. این نوع نمونه‌گیری به عنوان نمونه‌گیری خود راه انداز شناخته می‌شود. خروجی ترکیب n مدل با میانگین‌گیری برای رگرسیون و رأی‌گیری برای طبقه‌بندی به دست می‌آید. بنابراین با استفاده از نمونه‌گیری دوباره و تولید مجموعه داده‌های مختلف، تنوع مورد نیاز حاصل خواهد شد.

^۱ Bagging

در مرحله دوم اقدام به تشکیل و بررسی ویژگی‌های تصادفی شده است. خصوصیت ویژگی‌های تصادفی بدین صورت است که در هر گره از ساخت هر درخت به تصادف یک گروه کوچک از ویژگی‌های ورودی انتخاب می‌شود و برای تقسیم گره به جای جستجو از میان همه ویژگی‌ها، از میان ویژگی‌های این زیرگروه بهترین ویژگی با بیشترین بهره اطلاعاتی برای رشد درخت انتخاب می‌شود. تعداد این ویژگی‌ها کمتر از تعداد ویژگی‌های اصلی است. هر درخت در جنگل تصادفی با استفاده از الگوریتم درخت تصمیم کارت و با حداکثر اندازه و بدون هرس رشد می‌کند. در این پژوهش در هر گره از تعداد $[\log_2 M] + 1$ استفاده کرده است که در آن M تعداد کل ویژگی‌های ورودی است.

جنگل تصادفی یک طبقه بند مجموعه‌ای متشکل از طبقه‌بندهای درخت تصمیم است. هر طبقه‌بند برای هر نمونه ورودی به صورت $h(x, \theta_k)$ است که x یک نمونه ورودی و θ_k مجموعه آموزش برای درخت k ام است. θ ها مستقل از یکدیگر ولی با توزیع یکسان هستند. برای هر نمونه x هر درخت یک پیش‌بینی را برای رده نمونه x ارائه می‌دهد و در نهایت رده‌ای با بیشترین تعداد رأی درختان روی ورودی x به عنوان رده نمونه انتخاب می‌شود. با تشکیل طبقه‌بندی‌های از داده‌ها اقدام به تشکیل یک فرآیند تخمین داده‌ها از خارج کیسه (OOB) شده است. نمونه‌های آموزش در مجموعه داده آموزش اصلی که در مجموعه آموزش طبقه‌بند k نیست،

نمونه‌های خارج از کیسه طبقه‌بند k -ام نامیده می‌شود که با استفاده از رابطه زیر حاصل شده است:

$$y(x) = \arg \max_c \left(\frac{1}{K} \sum_{k=1}^K I(h_k(x) = c) \cdot x \in OOB_k \right) \quad (27)$$

$$I(h_k(x) = c) = \begin{cases} 1 & \text{if } h_k(x) = c \\ 0 & \text{otherwise} \end{cases} \quad x \in OOB_k$$

که K تعداد درختان، c نشان دهنده رده $h_k(x)$ پیش‌بینی درخت k -ام روی نمونه x را نشان می‌دهند و OOB_k مجموعه نمونه‌های OOB درخت k -ام می‌باشند. رابطه بالا نشان می‌دهد که مقدار تابع شاخص ۱ یک خواهد بود اگر x در مجموعه نمونه‌های درخت k -ام قرار گیرد (عضو مجموعه آموزش درخت k -ام نیست) و همچنین درخت k -ام نمونه x را به رده c طبقه‌بندی کند. در غیر این صورت، مقدار تابع شاخص، صفر می‌شود.

برای به دست آوردن تخمین نمونه‌های OOB روی جنگل از $error_k(OOB)$ در رابطه زیر استفاده می‌شود که خطای طبقه‌بندی جنگل روی نمونه‌های OOB درخت k -ام می‌باشد:

$$error_k(OOB) = \frac{\sum_{(x_i, y_i) \in OOB_k} I(y_i \neq h_k(x_i))}{\sum_{(x_i, y_i) \in OOB_k} 1} \quad (28)$$

$$error(OOB) = \frac{\sum_{i=1}^N I(y_i \neq h_i(x_i))}{\sum_{i=1}^N I((x_i, y_i) \in OOB_k)} \quad (29)$$

N تعداد همه نمونه‌های مجموعه آموزش اصلی، x_i نمونه i -ام روی مجموعه آموزش اصلی، y_i رده واقعی $y(x_i)$ ، y_i رده پیش‌بینی شده برای x_i است. در رابطه ۲۸ مقدار تابع I یک خواهد بود اگر نمونه (x_i, y_i) متعلق به مجموعه OOB درخت k باشد و در غیر این صورت، صفر است.

در جنگل تصادفی برای خطا، یک کران بالا در نظر گرفته می‌شود. دو پارامتر برای این کران بالا اندازه‌گیری می‌شود. پارامتر اول قدرت طبقه‌بندهای فردی و پارامتر دوم، وابستگی بین درختان جنگل است.

$$PE^* \leq \rho (1-S^2)/S^2 \quad (30)$$

در رابطه S قدرت طبقه‌بندهای فردی در جنگل و ρ وابستگی بین درختان جنگل را نشان می‌دهد. بر اساس این رابطه برای خطای PE^* هرچه مقدار S بیشتر و مقدار ρ کمتر باشد میزان خطا نیز کمتر خواهد بود. نسبت $\frac{\rho}{S^2}$ برای جنگل تصادفی در درک عملکرد آن یک راهنمای مفید است. این نسبت تقسیم وابستگی بر مربع قدرت است و هر چه کوچک تر باشد عملکرد جنگل بهتر و خطا نیز کمتر است.

در ساخت الگوریتم جنگل تصادفی در پژوهش حاضر از دو معیار بهره اطلاعات و مرحله آموزش استفاده شده است که روابط به دست آمده در ادامه تشریح شده است.

بهره اطلاعات $Information\ Gain(S,A)$ برای یک ویژگی نظیر A نسبت به مجموعه نمونه‌های S در رابطه ۳۲ تعریف می‌شود که در آن $Values(A)$ مجموعه همه مقدار ویژگی‌های A بوده و S_v زیرمجموعه‌ای از S است که برای آن ویژگی A دارای مقدار v است. $|S_v|$ تعداد نمونه‌ها با مقدار v روی ویژگی A در مجموعه S و $|S|$ تعداد کل نمونه‌ها روی مجموعه S است.

$$Information\ Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \left[\frac{|S_v|}{|S|} Entropy(S_v) \right] \quad (31)$$

$$Entropy(S) = - \sum_{i=1}^{numclass} p_i \log_2 p_i \quad (32)$$

p_i نسبت نمونه‌ها در S که به رده i -ام تعلق دارد، می‌باشد.

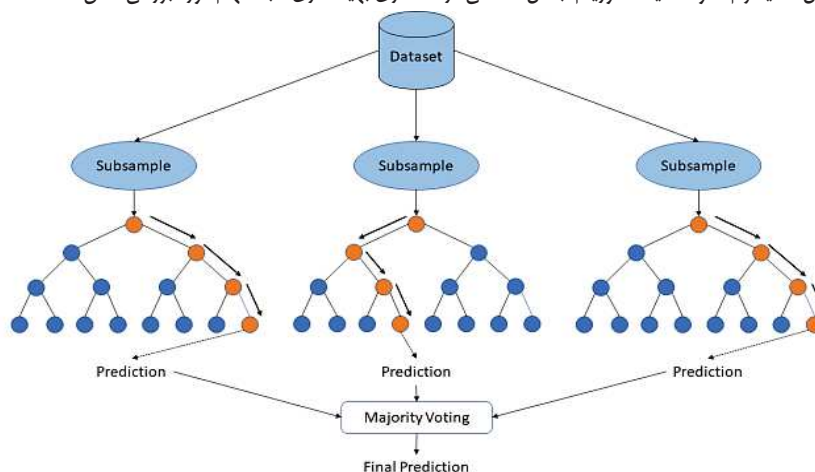
مرحله آموزش (IRF): در فرآیند ایجاد جنگل، در ساخت هر درخت و برای تقسیم هر گره، یک زیرمجموعه از ویژگی‌های ورودی را به تصادف انتخاب کرده و برای به دست آوردن بهترین ویژگی در گره برای ایجاد شاخه، از معیار بهره اطلاعات استفاده شده است. هر درخت روی جنگل مجموعه نمونه OOB مخصوص به خود را دارد. دقت هر درخت را با استفاده از نمونه‌های OOB آن درخت به دست آورده و آن را به عنوان وزن آن درخت در نظر گرفت. وزن هر درخت از رابطه زیر به دست می‌آید:

$$W_k = \frac{\sum_{i=1}^N I(h_k(x_i) = y_i, (x_i, y_i) \in OOB_k)}{\sum_{i=1}^N I(x_i, y_i) \in OOB_k} \quad (33)$$

مرحله آخر در استفاده از الگوریتم جنگل تصادفی، مرحله آزمایش IRF است. در این مرحله برای ارزیابی الگوریتم از مجموعه داده آزمایشی استفاده شده است که رده نمونه‌های آزمایشی بر اساس وزن درختان جنگل پیش‌بینی می‌شود. برای به دست آوردن رده هر نمونه همه درختان جنگل شرکت دارند و از رابطه ۳۴ استفاده شده است که براساس وزن درختان به دست آمده در مرحله آموزش است. برای هر رده C مجموعه وزن درختانی که رده C را برای نمونه x پیش‌بینی می‌کنند به دست می‌آید و سپس رده‌ای با بیشترین میانگین وزن روی درختان به عنوان رده آزمایشی در نظر گرفته شده است. در رابطه زیر $y(x)$ رده پیش‌بینی شده توسط جنگل تصادفی بهبودیافته برای نمونه x است، K تعداد درختان و W_k وزن درخت k -ام است.

$$y(x) = \arg \max_c \left\{ \frac{1}{K} \sum_{k=1}^K W_k \cdot I(h_k(x) = c) \right\} \quad (34)$$

در شکل ۱ دیاگرام نحوه فعالیت الگوریتم جنگل تصادفی در مدلسازی بهینه‌سازی سبد سهام مورد بررسی نشان داده شده است.



شکل ۱- دیاگرام نحوه فعالیت الگوریتم جنگل تصادفی در داده‌های بهینه‌سازی سبد سهام

ارزیابی نتایج به دست آمده توسط مدل‌ها

برای ارزیابی عملکرد مدل پیشنهادی، میانگین خطای مطلق (MAE)^۱، درصد میانگین مطلق خطا^۲ ($MAPE$) و خطای میانگین مربع ریشه^۳ ($RMSE$) به کار رفته است. این خطاها در ادامه تشریح شده‌اند.

$$MAE = \frac{1}{N} \sum_{L=1}^N |y_L - \bar{y}_L| \quad (35)$$

$$MAPE = \frac{1}{N} \sum_{L=1}^N \left| \frac{y_L - \bar{y}_L}{y_L} \right| \quad (36)$$

$$RMSE = \frac{1}{N} \sum_{L=1}^N (y_L - \bar{y}_L)^2 \quad (37)$$

که در آن N اندازه آموزش یا نمونه‌های آزمون است و y_L و $\bar{y}_L$ به ترتیب مقدار پیش‌بینی شده و مقدار واقعی هستند. در این روابط، MAE میانگین خطاهای مطلق بین مقادیر پیش‌بینی شده و مقادیر واقعی است که نشان دهنده خطای واقعی مقدار پیش‌بینی شده است. $MAPE$ همچنین نسبت بین خطا و مقدار واقعی را در نظر می‌گیرد. $RMSE$ نشان دهنده انحراف استاندارد

^۱ Mean Absolute Error

^۲ Mean Absolute Percentage Error

^۳ Root Mean Square Error

نمونه از تفاوت بین مقادیر پیش‌بینی شده و مقادیر مشاهده شده واقعی است. هرچه مقادیر MAE ، $MAPE$ و $RMSE$ کوچکتر باشد، عملکرد مدل‌ها در پیش‌بینی بهتر است.

جدول ۲ نتایج تجربی به دست آمده میانگین خطای مطلق (MAE) در مورد بهینه‌سازی سید سهام برای شبکه مورد بررسی را نشان داده است.

جدول ۲- نتایج تجربی به دست آمده میانگین خطای مطلق (MAE) برای بهینه‌سازی سید سهام

شماره سهام	<i>Mean-variance Technical model</i>	<i>RF</i>
۱	۲۲۳/۰۲۵	۲۹۶/۱۱۹
۲	۳۱۹/۱۱۴	۳۶۱/۳۴۴
۳	۲۶۷/۴۴۲	۳۲۸/۰۶۴
۴	۳۰۱/۰۰۳	۳۶۷/۱۱۲
۵	۳۱۸/۶۴۹	۴۰۱/۰۹۶

ماخذ: یافته‌های تحقیق

جدول ۳ نتایج تجربی به دست آمده میانگین خطای مطلق ($MAPE$) در مورد بهینه‌سازی سید سهام برای شبکه مورد بررسی را نشان داده است.

جدول ۳- نتایج تجربی به دست آمده میانگین خطای مطلق ($MAPE$) برای بهینه‌سازی سید سهام

شماره سهام	<i>Mean-variance Technical model</i>	<i>RF</i>
۱	۰/۰۶۲۰	۰/۰۸۶۴
۲	۰/۰۷۱۱	۰/۰۹۶۶
۳	۰/۰۵۹۷	۰/۰۷۷۲
۴	۰/۰۷۹۵	۰/۰۹۲۷
۵	۰/۰۶۶۴	۰/۰۸۱۱

ماخذ: یافته‌های تحقیق

جدول ۴ نتایج تجربی به دست آمده میانگین خطای مطلق ($RMSE$) در مورد بهینه‌سازی سید سهام برای شبکه مورد بررسی را نشان داده است.

جدول ۴- نتایج تجربی به دست آمده میانگین خطای مطلق ($RMSE$) برای بهینه‌سازی سید سهام

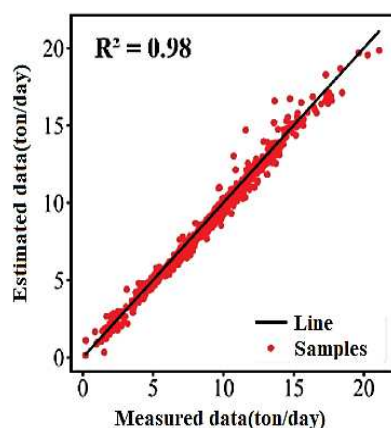
شماره سهام	<i>Mean-variance Technical model</i>	<i>RF</i>
۱	۱۶۸۵/۵۱۹	۲۰۱۶/۰۷۸۴
۲	۱۵۵۶/۰۳۴	۱۹۶۴/۸۶۴
۳	۱۸۲۶/۱۶۴	۱۹۶۳/۰۷۹

۴	۱۶۳۷/۰۰۶	۱۸۳۴/۱۷۷
۵	۱۷۳۳/۲۷۴	۱۹۸۳/۰۱۹

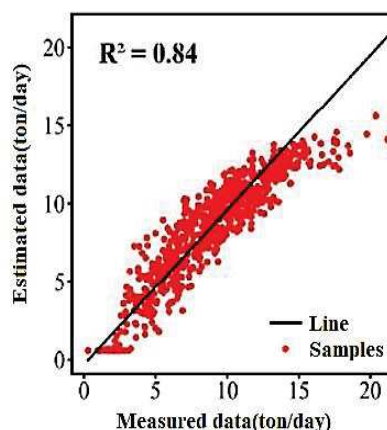
ماخذ: یافته‌های تحقیق

نتایج به دست آمده نشان می‌دهد که مقادیر مقادیر MAE ، $MAPE$ و $RMSE$ برای شبکه مدل میانگین-واریانس تکنیکالی نسبت به الگوریتم RF از مقادیر کمتری برخوردار هستند که این موضوع دقت بالاتر این شبکه عصبی بازگشتی نسبت به الگوریتم جنگل تصادفی را در مدلسازی و پیش‌بینی مقادیر بهینه‌سازی سبد سهام نشان می‌دهد. در شکل ۲ نمودارهای به دست آمده برای مقادیر R^2 در داده‌های بهینه‌سازی سبد سهام نشان داده شده است.

نمودار R^2 برای داده‌های بهینه‌سازی سبد سهام میانگین-واریانس تکنیکالی



نمودار R^2 برای داده‌های بهینه‌سازی سبد سهام شبکه RF



شکل ۲- نمودارهای به دست آمده برای مقادیر R^2 در داده‌های بهینه‌سازی سبد سهام

نتایج به دست آمده نشان داده است که نمودار R^2 برای داده‌های مربوط به بهینه‌سازی سبد سهام برای مدل میانگین-واریانس تکنیکالی دارای مقادیر نزدیکتری نسبت به ۱ دارند در حالی که الگوریتم RF دارای مقادیر کمتری از R^2 برای داده‌های مربوط به مدل میانگین-واریانس تکنیکالی می‌باشد که این موضوع نشان دهنده دقت بالای مدل میانگین-واریانس تکنیکالی است. این نتایج نشان داده است که مدل میانگین-واریانس تکنیکالی دارای دقت بالاتری در بهینه‌سازی سبد سهام است.

نتیجه‌گیری و پیشنهاد

در این پژوهش، مدل بهینه‌سازی میانگین-واریانس برای سبد سهام بر اساس تحلیل سیگنال‌های تکنیکی ارائه شده است. براساس نتایج به دست آمده از پژوهش، در بهینه‌سازی سبد سهام با رویکرد میانگین - واریانس تکنیکالی باید به این نکته توجه داشت که این مدل در مورد انتخاب وزن پرتفوی است که مبادله بهینه بین میانگین (به عنوان معیار سود) و واریانس (به عنوان معیار ریسک) بازده پرتفوی را برای یک دوره آتی فراهم می‌کند. رویکرد تکنیکالی استفاده شده در این پژوهش، میزان توزیع را برای (μ, Σ) فرض می‌کند و بهینه‌سازی پورتفولیوی میانگین واریانس را به عنوان یک مسئله بهینه‌سازی تصادفی فرموله می‌کند. این مشکل بهینه‌سازی زمانی که توزیع قبلی منحط است به مشکل مارکوویتز کاهش می‌یابد. از همین رو، رویکرد تکنیکالی در رویکرد میانگین-واریانس از توزیع پسین مشاهدات فعلی و گذشته استفاده می‌کند تا عدم قطعیت‌های μ و Σ در واریانس بازده پورتفولیو $W^T r_{n+1}$ را وارد کند. برای به کارگیری این مدل در عمل، سرمایه گذار باید ضریب ریسک گریزی خود را مشخص کند که ممکن است کار دشواری باشد. نظریه مارکوویتز با در نظر گرفتن مرز کارآمد، که منحنی (μ, σ) پرتفوی کارآمد است، دور می‌زند، زیرا λ در تمام مقادیر ممکن تغییر می‌کند، که μ میانگین و σ^2 واریانس بازده پرتفوی است. با این حال، سرمایه گذاران اغلب ترجیح می‌دهند از نسبت شارپ $(\mu - \mu_0)/\sigma$ به عنوان معیاری برای عملکرد پرتفوی استفاده کنند، جایی که μ_0 نرخ بهره بدون ریسک یا بازده مورد انتظار پرتفوی بازار است. باید توجه داشت که نسبت شارپ با $\mu - \mu_0$ متناسب و با σ نسبت معکوس دارد و می‌تواند به عنوان بازده اضافی در واحد ریسک در نظر گرفته شود. یک فرض مهم در اصلاح رویکرد میانگین - واریانس تکنیکالی این است که r_t با میانگین μ و ماتریس کوواریانس Σ دارای ارتباط است. بررسی‌های تشخیصی میزان نقض این فرض باید در عمل انجام شود. تئوری بهینه‌سازی تصادفی در بخش قبل در واقع به این فرض نیاز ندارد و فقط به میانگین خلفی و ماتریس گشتاور دوم بردار بازگشتی برای دوره بعدی نیاز دارد. بنابراین بر این اساس بررسی‌های تشخیصی نشان می‌دهد چنین تغییراتی مورد نیاز است. یک روش ساده برای معرفی چنین اصلاحی، استفاده از مدل رگرسیون تصادفی فرم است.

نتایج به دست آمده نشان می‌دهد که مقادیر شاخص‌های مطروحه برای شبکه مدل میانگین-واریانس تکنیکالی نسبت به الگوریتم RF از مقادیر کمتری برخوردار هستند. و این موضوع دقت بالاتر این شبکه عصبی بازگشتی نسبت به الگوریتم جنگل تصادفی را در مدلسازی و پیش‌بینی مقادیر بهینه‌سازی سبد سهام نشان می‌دهد. نتایج به دست آمده نشان می‌دهد که نمودار R^2 برای داده‌های مربوط به بهینه‌سازی سبد سهام برای مدل میانگین-واریانس تکنیکالی دارای مقادیر نزدیکتری نسبت به ۱ دارند در حالی که الگوریتم RF دارای مقادیر کمتری از R^2 برای داده‌های مربوط به مدل میانگین-واریانس تکنیکالی می‌باشد که این موضوع نشان دهنده دقت بالای مدل میانگین-واریانس تکنیکالی است.

پژوهش حاضر با پژوهش‌های ژو و همکاران (۲۰۲۳)، پارک و همکاران (۲۰۲۱)، فیریانتی و همکاران (۲۰۲۳)، مورتاس (۲۰۲۲) همراستا می‌باشند.

بهینه‌سازی سبد سهام با استفاده از روش میانگین-واریانس و جنگل تصادفی به تحلیل دقیق داده‌های تاریخی و پیش‌بینی بازده سهام کمک می‌کند. در ابتدا، با جمع‌آوری داده‌های قیمتی و محاسبه میانگین و واریانس بازده هر سهام، می‌توان ریسک و بازده مورد انتظار را تعیین کرد. سپس با استفاده از روش میانگین-واریانس، سبد سهام به گونه‌ای بهینه می‌شود که حداکثر بازده و حداقل ریسک را فراهم کند. این فرآیند شامل تنظیم وزن‌های هر سهام در سبد است تا تعادل مناسبی بین ریسک و بازده ایجاد شود. در مرحله بعد، تکنیک جنگل تصادفی به عنوان یک ابزار پیشرفته یادگیری ماشین به کار گرفته می‌شود تا الگوهای پیچیده و غیرخطی در داده‌های بازده سهام شناسایی شود. این مدل می‌تواند به شناسایی ویژگی‌های کلیدی که بر بازده تأثیر دارند، کمک کند و به شما امکان می‌دهد تا سبد بهینه‌تری طراحی کنید. با ترکیب نتایج به‌دست‌آمده از جنگل تصادفی با روش میانگین-واریانس، می‌توان وزن‌های بهینه‌ای برای هر سهام تعیین کرد و سبد سهام را به‌طور مداوم نظارت و به‌روزرسانی کرد تا در شرایط متغیر بازار عملکرد بهتری داشته باشد.

منابع و مأخذ

- Adinevand, Dariush, Razini, Ebrahim Ali, Khodam, Mahmoud, Ohadi, Fereydown, Hashemizadeh, Elham Elsadat. (۲۰۲۳). Examining the Efficiency Models, Genetic Algorithm under MSV Risk and Particle Swarm Optimization Algorithm under CVAR Risk Criterion in Selection Optimal Portfolio Shares Listed Firms on Stock Exchange. *Financial Economics*, ۶۵(۴), ۳۰۷-۳۲۲. (Persian)
- Anuno, F.; Madaleno, M.; Vieira, E. Testing of Portfolio Optimization by Timor-Leste Portfolio Investment Strategy on the Stock Market. *J. Risk Financial Manag.* ۲۰۲۴, ۱۷, ۷۸.
- Arashi, M.; Mahdi, M. Analysis of market efficiency and fractal feature of NASDAQ stock exchange: Time series modelling and forecasting of stock index using ARMA-GARCH model. *Future Bus. J.* ۲۰۲۲, ۸, ۱۴.
- Beljan, K.; Brenner, M.; Dolinar, D. The Long-Term Capital-Market Performance of the Forestry Sector: An Investors' Perspective. *Forests* ۲۰۲۲, ۱۳, ۱۳۲۹. <https://doi.org/10.3390/f13081329>
- Berouaga, Y.; El Msiyah, C.; Madkour, J. Portfolio Optimization Using Minimum Spanning Tree Model in the Moroccan Stock Exchange Market. *Int. J. Financial Stud.* ۲۰۲۳, ۱۱, ۵۳. <https://doi.org/10.3390/ijfs11020053>
- Čeh, M.; Kilibarda, M.; Lisec, A.; Bajat, B. Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments. *ISPRS Int. J. Geo-Inf.* ۲۰۱۸, ۷, ۱۶۸.
- Chavalle, L.; Chavez-Bedoya, L. The Impact of Transaction Costs in Portfolio Optimization. *J. Econ. Financ. Adm. Sci.* ۲۰۱۹, ۲۴, ۲۸۸-۳۱۱.
- Chaweewanchon, A.; Chaysiri, R. Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning. *Int. J. Financial Stud.* ۲۰۲۲, ۱۰, ۶۴. <https://doi.org/10.3390/ijfs10030064>

- Díaz, R.; Solares, E.; de-León-Gómez, V.; Salas, F.G. Stock Portfolio Management in the Presence of Down Trends Using Computational Intelligence. *Appl. Sci.* ۲۰۲۲, ۱۲, ۴۰۶۷. <https://doi.org/10.3390/app12084067>
- Febrianti, W.; Sidarto, K.A.; Sumarti, N. Solving Constrained Mean-Variance Portfolio Optimization Problems Using Spiral Optimization Algorithm. *Int. J. Financial Stud.* ۲۰۲۳, ۱۱, ۱. <https://doi.org/10.3390/ijfs11010001>
- Fernandez, E.; Navarro, J.; Solares, E.; Coello, C.C. A novel approach to select the best portfolio considering the preferences of the decision maker. *Swarm Evol. Comput.* ۲۰۱۹, ۴۶, ۱۴۰–۱۵۳.
- Idrees, S.M.; Alam, M.A.; Agarwal, P. A Prediction Approach for Stock Market Volatility Based on Time Series Data. *IEEE Access* ۲۰۱۹, ۷, ۱۷۲۸۷–۱۷۲۹۸.
- Jafari, Fatemeh, Ali Beigi Bani, Nazanin, Karimkhani, Mehdi, Izdiyar, Amin. (۲۰۲۳). Optimizing the stock portfolio using the conditional value-at-risk model with a simulation approach, ۸th National Conference on Management and Economics Studies in Human Sciences, Tehran. (Persian)
- Ji, R., Lejeune, M. A., & Fan, Z. (۲۰۲۲). Distributionally robust portfolio optimization with linearized STARR performance measure. *Quantitative Finance*, ۱–۱۵. <https://doi.org/10.1080/14697688.2021.1993623>
- Jing, D.; Imeni, M.; Edalatpanah, S.A.; Alburaikan, A.; Khalifa, H.A.E.-W. Optimal Selection of Stock Portfolios Using Multi-Criteria Decision-Making Methods. *Mathematics* ۲۰۲۳, ۱۱, ۴۱۵. <https://doi.org/10.3390/math11020415>
- Mourtas, S.D.; Kasimis, C. Exploiting Mean-Variance Portfolio Optimization Problems through Zeroing Neural Networks. *Mathematics* ۲۰۲۲, ۱۰, ۳۰۷۹. <https://doi.org/10.3390/math10173079>
- Nor, S.M.; Zawawi, N.H.M.; Wickremasinghe, G.; Halim, Z.A. Is Technical Analysis Profitable on Renewable Energy Stocks? Evidence from Trend-Reinforcing, Mean-Reverting and Hybrid Fractal Trading Systems. *Axioms* ۲۰۲۳, ۱۲, ۱۲۷. <https://doi.org/10.3390/axioms12020127>
- Park, H. Modified Mean-Variance Risk Measures for Long-Term Portfolios. *Mathematics* ۲۰۲۱, ۹, ۱۱۱. <https://doi.org/10.3390/math9020111>
- Purwandari, T.; Riaman; Hidayat, Y.; Sukono; Ibrahim, R.A.; Hidayana, R.A. Selecting and Weighting Mechanisms in Stock Portfolio Design Based on Clustering Algorithm and Price Movement Analysis. *Mathematics* ۲۰۲۳, ۱۱, ۴۱۵۱. <https://doi.org/10.3390/math11094151>
- Sekulić, A.; Kilibarda, M.; Heuvelink, G.B.M.; Nikolić, M.; Bajat, B. Random Forest Spatial Interpolation. *Remote Sens.* ۲۰۲۰, ۱۲, ۱۶۸۷. <https://doi.org/10.3390/rs12101687>
- Shah, D.; Isah, H.; Zulkernine, F. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *Int. J. Financ. Stud.* ۲۰۱۹, ۷, ۲۶.
- Sukono; Lesmana, E.; Napitupulu, H.; Hidayat, Y.; Saputra, J.; Ghazali, P.L.B. Mean-VaR Portfolio Optimisations: An Application of Multiple Index Models with Non-Constant Volatility and Long Memory Effects. *Int. J. Innov. Creat. Chang.* ۲۰۱۹, ۹, ۳۶۴–۳۸۱.

- Xu, S.; Shen, J.; Hua, H.; Li, F.; Yu, K.; Li, Z.; Gao, X.; Dong, X. Trading Portfolio Strategy Optimization via Mean-Variance Model Considering Multiple Energy Derivatives. *Processes* ۲۰۲۳, ۱۱, ۵۳۲. <https://doi.org/10.3390/pr11020532>
- Zhang, X. (۲۰۲۲). Distributional Robust Portfolio Construction based on Investor Aversion. *arXiv preprint arXiv:2203.13999*.
<https://doi.org/10.48550/arXiv.2203.13999>

Optimizing the stock portfolio using the technical mean-variance method and random forest

Leila Boroomandi^۱, Saeed Aghasi^۲ and Mohammad Reza Sharifi Qazvini^۳

Abstract

purpose of this research is to optimize the stock portfolio using the technical mean-variance method and random forest. The method of the present research is based on the purpose of the applied type and in terms of the type of data, it is post-event. The statistical population of this research includes all companies active in the Tehran Stock Exchange during the years ۱۳۹۰ to ۱۴۰۲. In this research, ۴۰ active stocks in the stock market that have continuous and complete data have been selected. To check the optimization of the stock portfolio, the stock price information of ۹۰ companies admitted to the stock exchange was used. The criteria for selecting these ۹۰ companies is the high average price, average yield and number of trading days compared to other companies in the stock exchange. This research has been used to optimize the mean-variance of the stock portfolio based on the analysis of technical signals from the relative strength index (RSI) of the type of guiding indicators. Two methods of mean-variance technical Markowitz analysis and random forest algorithm were used for modeling. To evaluate the performance of the proposed model, mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) have been used. The obtained results show that the values of the proposed indices for the technical mean-variance model network have lower values compared to the RF algorithm, which shows the higher accuracy of this recurrent neural network than the random forest algorithm in modeling and predicting optimal values. It shows the creation of the stock portfolio. The obtained results have shown that the R^2 chart for the data related to stock portfolio optimization for the technical mean-variance model has values closer to ۱, while the RF algorithm has lower values than R^2 for the data related to the mean-variance model. It is technical, which shows the high accuracy of technical mean-variance model.

Keywords: Stock exchange, Random forest, Stock portfolio, Technical mean-variance.

^۱ PhD Candidate of Industrial Management, Department of Management, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran. Email: Leilaboroomandi@gmail.com.

^۲ Assistant Professor, Department of Management, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran. (Corresponding Author), Email: sae_aghasi@yahoo.com.

^۳ Assistant Professor, Department of Industrial Engineering, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran. Email: sharifidocument@gmail.com