فصلنامه مهندسی مالی و مدیریت اوراق بهادار
دوره پانزده، شماره پنجاه و هشت، بهار ۱۴۰۳
نوع مقاله : علمی پژوهشی
صفحات : ۲۱۵-۱۹۴

ارائه الگوریتم ترکیبی یادگیری ماشین و ترکیب سنجه‌های ریسک و نظریه فازی در انتخاب سبد سرمایه‌گذاری

دانیال محمدی[۱]
سیدجعفر سجادی[۲]
عمران محمدی[۳]
نعیم شکری[۴]

تاریخ دریافت مقاله : ۱۴۰۱/۱۲/۱۰ تاریخ پذیرش مقاله : ۱۴۰۲/۰۵/۱۶

چکیده

پژوهش حاضر جهت یافتن پرتفوی بهینه برای سرمایه‌گذاری از سهام بورسی انجام گرفته و یکی‌از روش‌هایی‌که در حال حاضر محبوبیت زیادی در بین تحلیل‌گران و پژوهش‌گران این حوزه شکل گرفته، روش‌های مبتنی‌بر هوش مصنوعی و در پی آن روش‌هایی با هدف کاهش سنجه‌های ریسک می‌باشد. هدف تحقیق، تشکیل پرتفوی باا ستفاده از روش‌های یادگیری ما شین، سنجه ریسک و ترکیب آن با نظریه فازی است، که بازده‌ای بهتر از بازده میانگین بازار داشته باشد. خروجی هر روش وارد الگوریتم جنگل تصادفی شده و پیش‌بینی به‌و سیله این الگوریتم صورت می‌گیرد و در مرحله آخر، خروجی پیش‌بینی برای تشکیل سبد سرمایه وارد مدل بهینه سازی ارزش در معرض ریسک و ارزش در معرض ریسک شرطی با رویکرد نظریه فازی می‌شوند. اطلاعات سهم‌ها به‌صورت روزانه و بازه زمانی آن از ابتدای سال ۱۳۹۴ تا اواسط سال ۱۳۹۸ می‌باشد. در پایان هرکدام از این روش‌ها و مراحل با بازده واقعی بازار مقایسه گردید. بر اساس نتایج بدست آمده سنجه‌ریسک CVAR قابلیت بهتری را نسبت‌به سنجه ریسک VAR داشته است، هم‌چنین الگوریتم جنگل تصادفی در بین الگوریتم‌های یادگیری ماشین استفاده شده، نتایج بهتری را در انتخاب سبد سرمایه‌گذاری رقم زده است.

کلمات کلیدی

یادگیری ماشین، بیز ساده، ارزش در معرض ریسک شرطی (CVaR)، سبد سرمایه‌گذاری، بورس اوراق بهادار تهران. طبقه‌بندی JEL: G01، G32،E44

۱-گروه مهندسی مالی، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران. dnl_mhm@yahoo.com
۲-گروه مهندسی مالی، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران. sjsadjadi@iust.ac.ir
۳-گروه مهندسی مالی، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران. e_mohammadi@iust.ac.ir
۴-گروه توسعه و برنامه ریزی اقتصادی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران. (نویسنده مسئول)
n.shokri@modares.ac.ir

مقدمه

انتخاب سبد مطلوب سرمایه‌گذاری باعث می‌شود افراد با سرمایه‌گذاری در زمینه‌های مناسب، سود بیشتری کسب کنند. بنابراین مطالعه راه‌های تعیین سبد سهام بهینه از اهمیت و ضرورت بالایی برخوردار است. هدف از حل مساله بهینه‌سازی سبد سرمایه‌گذاری، ارائه ابزاری مناسب برای سرمایه‌گذاران جهت انتخاب سبد سهام خود است (مهرجردی، ۱۳۹۱). عملکرد آتی بازارهای سهام حیاتی ترین عامل در ایجاد پرتفوی است. همان‌طور که تکنیک یادگیری ماشین در حال پیشرفت است، امکانات جدیدی برای گنجاندن مفاهیم پیش‌بینی در انتخاب پورتفولیو باز شده است(Kumar Pasayat, 2023). از طرفی باتوجه به این‌که عدم‌قطعیت در شرایط اقتصادی آینده نقشی کلیدی را در تصمیم گیری‌های مالی به‌ویژه مسائل انتخاب سبد سهام بازی می‌کنند، باید تکنیک‌های بهینه‌سازی سبد سهام را در کنار تکنیک‌های سنجه‌ریسک و برنامه‌ریزی احتمالی مطالعه نمود. هم در قسمت کلاس‌بندی به‌وسیله الگوریتم‌های یادگیری ماشین برای تفکیک داده‌ها و هم در قسمت انتخاب سبد و پرتفوی بهینه می‌توان شکاف تحقیق را بررسی کرد. در قسمت کلاس‌بندی دیتاها تاکنون از روش‌هایی چون طبقه‌بندی رندوم یا تصادفی استفاده می‌شد که در این مطالعه از روش هوش مصنوعی استفاده شده است و از آن مهم‌تر در قسمت انتخاب و بهینه‌کردن سبد سرمایه‌گذاری. تحقیقات گذشته همگی برفرض نرمال بودن دیتاها و استفاده از مدل مارکوییتز بنا شده بودند (بیگلری، ۱۳۸۹). پیشرفت‌ها در یادگیری ماشین طیف گسترده‌ای از امکانات جدید را برای استفاده از الگوریتم‌های پیشرفته کامپیوتری مانند یادگیری تقویتی در مدیریت ریسک پورتفولیو باز کرده است(Minh Ngo, 2023). معمولا فرض بر این است که سرمایه‌گذاران ریسک را دوست ندارند و از آن گریزانند و همواره درپی آن هستند تا در اقلامی از دارایی‌ها سرمایه‌گذاری کنند که بالاترین بازده و کم‌ترین ریسک را در پی داشته باشند. به عبارت دیگر، سرمایه‌گذاران به بازده سرمایه‌گذاری به‌عنوان یک عامل مطلوب و مناسب و به واریانس بازده‌ها به عنوان یک عنصر نامطلوب و یا نامناسب می‌نگرند (فرح آبادی و همکاران، ۱۴۰۱). مدل مارکوییتز با استفاده از مدل‌های برنامه‌ریزی ریاضی قابل‌حل می‌باشد، ولی وقتی محدودیت‌های دنیای واقعی به آن افزوده می‌شود، استفاده از الگوریتم‌های دقیق ریاضی ناممکن شده و از این‌روست که الگوریتم‌های فراابتکاری جایگاه ویژه‌ای میابند. (فضل‌زاده، ۱۳۹۰). در این مطالعه سعی می‌شود تا با استفاده از توانایی شبکه عصبی (یادگیری ماشین)، پرتفویی بااستفاده از روش‌های یادگیری ماشین و هم‌چنین ارزش در معرض ریسک و ارزش در معرض ریسک شرطی و ترکیب آن با نظریه فازی تشکیل شود، که بازده‌ای بهتر از بازده میانگین بازار داشته باشد و این روش نوآوری پژوهش حاضر را در مقایسه با سایر پژوهش‌های این حوزه آشکار می‌سازد.

پرسش‌های تحقیق

۱- چه میزان بازده پرتفوی سهام ایجاد شده توسط مدل‌های ارائه شده در این تحقیق می‌تواند به‌طور قابل‌توجهی عملکرد بهتری از بازده بازار (بازده شاخص) داشته باشد؟

۲- الگوریتم‌های مورد استفاده در این پژوهش، چه عملکردی در کلاس‌بندی دیتاها و به‌تبع آن بازده پرتفوی استخراجی نسبت به هم دارند؟

۳- بکارگیری از سطوح‌اطمینان مختلف برای ایجاد پرتفوی بهینه، بااستفاده از نظریه فازی و سنجه ریسک، چه تغییری در بازده خروجی سبد سهام خواهد داشت؟

۴- به‌طور کلی علت استفاده از الگوریتم‌های یادگیری در حوزه مالی، بالاخص در مباحث بازار سرمایه و ایجاد پرتفوی سرمایه‌گذاری چیست؟

مبانی نظری تحقیق

امروزه یکی از دغدغه‌های سرمایه‌گذاران، سرمایه‌گذاری در سبد اوراقی است که پر بازده‌تر و بهینه‌تر باشد. یکی از وظایفی که مهندسین مالی با آن مواجه هستند، کنترل و مشاوره برای ایجاد یک زمینه‌ی مناسب برای سرمایه‌گذاری به‌منظور دستیابی به سبدی بهینه از اوراق بهادار می‌باشد. بنابراین امروزه کارگزاری‌ها، صندوق‌های سرمایه‌گذاری و شرکت‌های تامین‌سرمایه به‌دنبال روشی مناسب برای این منظور می‌باشند. در این رابطه، بررسی و مطالعه سرمایه‌گذاران در جهت انتخاب مناسب‌ترین سبد سرمایه‌گذاری باتوجه به میزان ریسک و بازده آن انجام خواهد شد (راعی، ۱۳۹۳). بهره‌برداری از الگوریتم‌های فراابتکاری به‌دلیل توانایی آن‌ها در کشف کارآمد فضاهای جستجوی بزرگ در مسئله بهینه‌سازی سبد سهام، در حوزه‌های تحقیقات مالی رایج و محبوب است.(Morteza, 2023) . در بهینه‌سازی پورتفوی مسئله‌ی اصلی، انتخاب بهینه‌ی دارایی‌ها و اوراق بهاداری است که با مقدار مشخصی سرمایه می‌توان ایجاد کرد. اگرچه کمینه‌کردن ریسک و بیشینه نمودن بازده سرمایه‌گذاری به‌نظر ساده باشد، اما در عمل روش‌های متعددی برای تشکیل پورتفوی بهینه به‌کار می‌رود. بیشتر مسائل بهینه‌سازی که در جهان واقعی با آن ها روبرو هستیم، بیش‌از یک هدف را در برمی‌گیرد. در این‌گونه مسائل، که با عنوان بهینه‌سازی چندهدفه می‌شناسیم ناگزیریم چندین تابع هدف یا شاخص عملکرد را تعریف نماییم و به‌طور هم‌زمان و موازی مقدار همه‌ی آن‌ها را بهینه نماییم. از آنجایی‌که روش‌های بهینه‌سازی تک‌هدفه در هر مرحله از اجرا تنها یک پاسخ را می‌توانند بدهند، نمی‌توانند برای یافتن مجموعه‌ای از پاسخ‌های بهینه مطلوب باشند، بنابراین با توجه به بالا بودن درجه‌ی پیچیدگی آن، امروزه الگوریتم‌های تکاملی چندهدفه و استفاده از هوش‌مصنوعی ابزار مطلوب برای حل مسائل بهینه‌سازی چندهدفه درنظر گرفته

می‌شوند (سینایی، ۱۳۹۳).

پیشینه پژوهش

در جدول ۱ اسامی پژوهش‌گرانی که در حوزه یادگیری ماشین و کاربرد آن در مبحث مالی و بهینه‌سازی پرتفوی فعالیتی‌را صورت داده‌اند، ارائه شده است.

جدول ۱: پیشینه پژوهش

			الگوریتم‌های هوش مصنوعی								
	نویسنده	سال	SVM	DL	ANN	FL	GA	E	B	DT	O
۱	سانته کیدیس و همکاران	۲۰۱۷	■	■	■						
۲	بوستوس و همکاران	۲۰۱۸	■		■						
۳	چاکرابورتی و همکاران	۲۰۱۸	■					■		■	
۴	کوینه و همکاران	۲۰۱۸			■						
۵	فیشر و کراوس	۲۰۱۸		■							
۶	هو و همکاران	۲۰۱۸			■						
۷	هوانگ و همکاران	۲۰۱۸						■			
۸	کیا و همکاران	۲۰۱۸						■			
۹	لیو و همکاران	۲۰۱۸		■							■
۱۰	مالاگرینو و همکاران	۲۰۱۸							■		
۱۱	رن و همکاران	۲۰۱۸	■								
۱۲	وانگ ژو و همکاران	۲۰۱۸	■	■	■			■			
۱۳	ژانگ و همکاران	۲۰۱۸						■			
۱۴	ژو و همکاران	۲۰۱۸								■	
۱۵	چاتزیس	۲۰۱۹	■								
۱۶	فیشر	۲۰۱۹								■	
۱۷	گویخمان	۲۰۱۹			■						
۱۸	یاپینگ کایا	۲۰۱۹			■						
۱۹	ایواسکو	۲۰۲۰			■						
۲۰	چادهوری	۲۰۲۰					■				
۲۱	ذوالفقاری و غلامی	۲۰۲۰	■								
۲۲	وویو	۲۰۲۰	■		■						

منبع: یافته‌های پژوهش

معرفی مدل

در این مطالعه سعی می‌شود تا بااستفاده از توانایی شبکه‌عصبی (یادگیری‌ماشین) در ایجاد رابطه بین متغیرهای مختلف، مدلی طراحی شود تا بتوان یک سبد سهام بهینه را طراحی کرد. قبل‌از ورود به بحث بهینه‌سازی سبد سهام، ابتدا داده‌ها به‌وسیله یادگیری ماشین تفکیک و کلاس‌بندی می‌شود و سپس داده‌های فیلتر شده برحسب نوع اطلاعاتی که وجود دارد، بهینه خواهد شد (بیگی، ۱۳۸۹). هدف اصلی از مطالعه حاضر ساختن یک مدل تصمیم‌گیری است که یکپارچه‌سازی آن به‌وسیله الگوریتم‌های یادگیری ماشین (جنگل تصادفی و بیز ساده) انجام شده و درنهایت از کلاس مورد نظر یک پرتفوی بهینه با ارزش درمعرض ریسک (VAR) و ارزش درمعرض ریسک شرطی (C_VAR) که با نظریه فازی ترکیب شده است، خارج خواهد شد. در این راستا بعداز دریافت اطلاعات (قیمت‌ها) از سایت رسمی سازمان بورس اوراق بهادار، آن‌ها به‌وسیله یک سری ویژگی‌ها و به‌کمک هوش مصنوعی کلاس‌بندی می‌شوند. این سهم‌ها به دو کلاس مثبت و منفی تقسیم می‌شوند و هدف آن است که از بین سهم‌های کلاس مثبت یک پرتفوی بهینه به‌کمک سنجه‌ریسک C_VAR و VaR ایجاد شود. درنهایت بااستفاده از معیار بررسی بهینگی ترینر، پرتفوهای ایجادشده با بازدهی صورت‌گرفته از طریق بازار مقایسه خواهند شد. ازطرفی CVaR معیارمنسجمی برای اندازه‌گیری ریسک است و خصوصیات جالبی را از خود بروز می‌دهد، CVaR نسبت‌به آنالیز میانگین - واریانس در مواجه با توزیع نامتقارن بازدهی سرمایه، بهتر عمل می‌کند (فیضی، ۱۳۹۱).

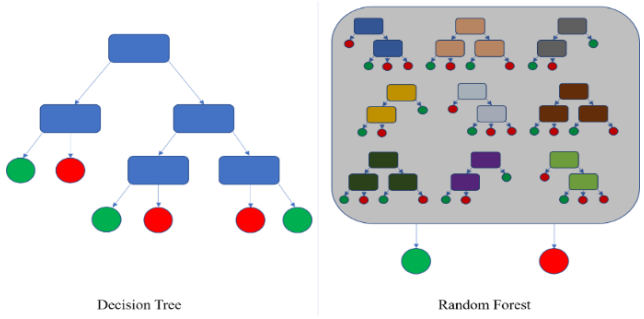
تشریح موضوع

فرآیند استخراج اطلاعات از داده‌ها، یک عمل پیچیده است که می‌توان از کارهایی که انجام می‌شود، به بحث پیش‌بینی و دسته‌بندی، کاهش پیچیدگی داده، کاهش حجم داده و استخراج مهم‌ترین ویژگی‌ها اشاره نمود، که می‌توان به‌وسیله آن‌ها حجم عظیمی از داده‌های پخش شده در اطرافمان را براساس الگوهایی که دارند، به‌درستی دسته‌بندی کرد و اطلاعات و دانش مورد نیاز خود را از آن‌ها استخراج کنیم. این کار یک مزیت دیگر نیز دارد که آن، کاهش پراکندگی داده‌ها است، که باعث خواهد شد در هنگام بهره برداری از داده‌ها، دچار سردرگمی نشویم. درحقیقت، ما به بررسی یکی از ابزارهای جدید و در عین‌حال کارامد در حوزه داده‌کاوی پرداخته‌ایم، که قابلیت زیادی دارد و می‌تواند بهترین نتیجه را نمایش دهد(البرزی، ۱۳۸۰).

جنگل تصادفی

جنگل تصادفی یک روش یادگیری گره برای طبقه‌بندی و رگرسیون است که با ساخت بسیاری از

درخت‌های تصمیم‌گیری در زمان آموزش و خروجی طبقات با درختان مجزا اداره می‌شود. جنگل تصادفی متناسب با درختان تصمیم متعدد بر روی داده‌ها و با انتخاب N نمونه به‌صورت تصادفی، مجموعه داده‌های آموزشی می‌شود. به‌منظور برازش درختان متعدد، جنگل تصادفی به‌طور تصادفی F ویژگی ورودی برای انشعاب در هر گره از درخت تصمیم‌گیری انتخاب می‌کند. مقدار F معمولا برابر با $\log2(|D|)$می‌باشد که $|D|$تعداد کل نمونه‌ها است (راوی ، ۲۰۱۷). مقدار کمی از F عملکرد پایین‌تری نسبت‌به همبستگی میان درختان تصمیم‌گیری می‌دهد، درحالی‌که مقدار زیادی از F طبقه‌بندی با دقت بهتری می‌دهد. در روش جنگل تصادفی عدم‌تشابه مشاهدات به‌طریقی کاملا متفاوت از توابع فاصله متداول و براساس ساختار این مدل تعیین می‌گردد. تشابه مشاهدات در این روش برمبنای قرارگرفتن آن‌ها در برگ‌های یکسان اندازه‌گیری می‌شود. در شکل ۱ نحوه تبدیل الگوریتم درخت تصمیم به الگوریتم جنگل تصادفی ارایه شده است.



شکل ۱: نحوه تبدیل الگوریتم درخت تصمیم به الگوریتم جنگل تصادفی

منبع: یافته‌های پژوهش

- ✓ هر گره داخلی یا غیربرگ با یک ویژگی مشخص می‌شود. این ویژگی سوالی را در رابطه با مثال ورودی مطرح می‌کند.
- ✓ در هر گره داخلی به تعداد جواب‌های ممکن برای این سوال، شاخه وجود دارد که هریک از شاخه ها با مقدار آن جواب مشخص می‌شوند.
- ✓ برگ‌های این درخت با یک کلاس که به آن برچسب گفته می‌شود و یا یک دسته از جواب‌ها مشخص می‌شوند.
- ✓ بالاترین گره را در درخت تصمیم، گره ریشه می‌نامند.

بیز ساده

در یادگیری ماشین، گاهی سعی داریم که از میان فضای فرضیه‌های H بهترین فرضیه‌ی سازگار با نمونه‌های آموزشی D را پیدا کنیم. چندین راه برای تعریف "بهترین" در این جمله وجود دارد، یکی از این تعاریف "محتمل‌ترین" است، با در دست داشتن داده‌های D بدون نیاز به هیچ اطلاعات اولیه‌ی دیگر نمی‌توان محتمل‌ترین فرضیه را انتخاب کرد. قضیه‌ی بیز متدی مستقیم برای محاسبه ی احتمالات فرضیه‌های موجود در H ارائه می‌کند. به عبارت دیگر، قضیه‌ی بیز روشی برای محاسبه‌ی احتمال یک فرضیه بر اساس احتمال قبلی‌اش، احتمال مشاهده‌ی داده‌های سازگار با فرض درستی این فرضیه و احتمال خود داده‌های مشاهده‌شده ارائه می‌کند. برای تعریف دقیق قضیه‌ی بیز، ابتدا بیاییدنشانه‌گذاری‌ها را معرفی کنیم. برای نشان دادن احتمال اولیه‌ی فرضیه‌ی h احتمال قبل از مشاهده‌ی داده‌های آموزشی، از P(h) استفاده می‌کنیم. به P(h) احتمال اولیه h نیز می‌گویند، این احتمال از اطلاعات قبلی ای که در مورد احتمال درستی فرضیه‌ی h داریم تأثیر می‌پذیرد. به‌طور مشابه از P(D) برای نمایش احتمال اولیه‌ی مشاهده‌ی نمونه‌های آموزشی D استفاده می‌کنیم (مثلاً احتمال مشاهده‌ی D بدون داشتن هیچ اطلاعات قبلی در مورد اینکه با چه فرضیه‌هایی سازگار است). برای نشان دادن احتمال مشاهده‌ی D در جایی که فرضیه‌ی h درست است از P(D|h) استفاده می‌کنیم. در حالت کلی، از p(x|y) برای نشان دادن احتمال x با فرض وقوع y استفاده می‌کنیم. در مسائل یادگیری ماشین، علاقه‌ی ما به احتمال P(h|D) است که در آن h یک فرضیه و D نمونه‌های آموزشی مشاهده شده هستند. به P(h|D) احتمال ثانویه h نیز می‌گویند، زیرا که اطمینان ما به فرضیه‌ی h بعد از مشاهده‌ی نمونه‌های آموزشی D را نشان می‌دهد. توجه داشته باشید که احتمال ثانویه P(h|D) بر خلاف احتمال اولیه P(h) که از نمونه‌های آموزشی مستقل است، از نمونه‌های آموزشی D تأثیر می‌پذیرد.

قضیه بیز، اساس متدهای یادگیری بیز است، زیرا که راهی برای محاسبه‌ی احتمال ثانویه P(h|D) از P(h)، P(D) و P(D|h) طبق رابطه (۱) می‌باشد:

$$P(h|D) = \frac{p(D|h)P(h)}{p(D)} \qquad (1)$$

همان‌طور که انتظار می‌رود، بر اساس قضیه‌ی بیز P(h|D) با افزایش P(h) و P(D|h) افزایش میابد. هم‌چنین منطقی است که P(h|D) با افزایش P(h) کاهش می‌یابد، زیرا که هر چه که احتمال مشاهده‌ی D به‌طور مستقل از h بالاتر رود دیگر D مدرکی برای درستی h نخواهد بود. در بسیاری از مسائل یادگیری، یادگیر مجموعه‌ی فرضیه‌هایی مثل H را در نظر می‌گیرد و در بین آن‌ها به‌دنبال محتمل‌ترین

فرضیه‌ی hϵH با توجه به نمونه‌های آموزشی D می‌گردد (یا حداقل یکی از محتمل‌ترین فرضیه‌ها). هر کدام از این محتمل‌ترین، فرضیه با حداکثر احتمال ثانویه یا (MAP) نامیده می‌شود. فرضیه‌های MAP را می‌توان با استفاده از قضیه‌ی بیز برای محاسبه‌ی احتمال ثانویه هر فرضیه مشخص کرد. در بعضی موارد، فرض می‌کنیم که هر فرضیه در H احتمال اولیه‌ی مساوی ای دارد (برای هر hj و hi در H داریم که P(hi)= P(hj) در این شرایط می‌توان رابطه ی ۲ را بیشتر ساده کرد و کافی است که فقط عبارت P(D|h) را برای پیدا کردن محتمل‌ترین فرضیه در نظر بگیریم. P(D|h) گاهی محتمل بودن داده‌های D برای h نیز نامیده می‌شود و هر فرضیه‌ای که P(D|h) را ماکزیمم کند (ML)، hML نامیده می‌شود. برای مشخص شدن رابطه با مسائل یادگیری ماشین، ابتدا قضیه‌ی بیز را با توجه به نمونه‌های D و فضای فرضیه‌ای H معرفی کردیم. در واقع قضیه‌ی بیز کلی‌تر از آنچه در بالا گفته شد است. از قضیه‌ی بیز می‌توان برای هر زیرمجموعه ی H که ناسازگارند (اشتراک ندارند) استفاده کرد (مثل "آسمان آبی است" و "آسمان آبی نیست"). در این فصل، در اکثر موارد فرض خواهیم کرد که H فضای فرضیه‌ای که تابع هدف را شامل می‌شود است و D نمونه‌های آموزشی هستند. در مواقع دیگر فرض می‌کنیم که H مجموعه‌ی دیگر ناسازگاری با یکدیگر از فرضیه‌هاست و D نیز مجموعه‌ی دیگری از داده‌هاست.

بررسی بهینگی پرتفوی

روش‌های متنوعی برای بررسی بهینگی پرتفوی وجود دارد. در این پژوهش از معیار ترینر که توسط فرمول‌های (۲) تا (۵) معرفی شده است، جهت بررسی بهینگی پرتفوی استفاده شده است. این روش که به معیار بازده به نوسان‌پذیری نیز معروف است، رابطه را برای پرتفوی موردنظر و بازار محاسبه کرده و به مقایسه آن‌ها می‌پردازد اگر این مقدار برای پرتفوی موردنظر از مقدار محاسبه شده برای بازار بیشتر باشد، حاکی از بهینگی نسبی پرتفوی است:

$$\frac{R_p - R_f}{\beta_p} \qquad (2)$$

$$R_i = R_f + \beta_i(R_m - R_f) \qquad (3)$$

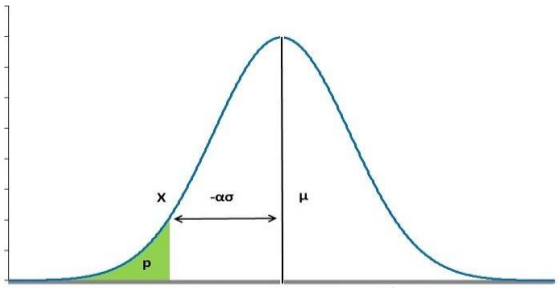
$$\beta_i = \frac{R_i - R_f}{R_m - R_f} \qquad (4)$$

$$R_p = \frac{\sum_{i=1}^{n} R_i}{n} \qquad (5)$$

ارزش در معرض ریسک و ارزش درمعرض ریسک شرطی تحت نظریه اعتبار

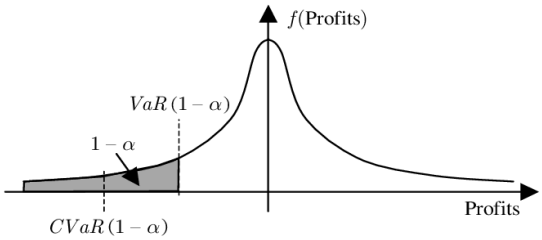
در نظریه مدرن پرتفوی، براساس رابطه بازهی و ریسک نامطلوب به تبیین رفتار سرمایه‌گذار و انتخاب

سبد سرمایه بهینه پرداخته خواهد شد. ارزش در معرض ریسک معروف‌ترین سنجه موجود در خانواده سنجه‌های ریسک نامطلوب است. اعتقاد عمومی در ادبیات مالی بر این است که VaR رویکردی جدید برای اداره و کنترل ریسک است. به عبارت ساده‌تر داریم باتوجه به شکل ۲، ما X درصد اطمینان داریم که طی N روز آینده، قطعاً بیش از مبلغ V متحمل زیان نخواهیم شد.



شکل ۲: ارزش درمعرض ریسک (منبع: یافته‌های پژوهش)

ارزش در معرض ریسک شرطی(CVaR) سنجه‌ای است که از ویژگی انسجام برخوردار بوده و بنابراین از اعتبار بیش‌تری نسبت به VaR برخوردار می‌باشد. ارزش در معرض ریسک مشروط طبق شکل ۳، به ما می‌گوید که در حالت‌های بد، چه انتظاری داشته باشیم. C-VaR بیان‌گر مقدار زیان در طی یک دورهٔ N روزه است مشروط به این که ما به اندازه $(X - 100)$ درصد در قسمت برآمدگی چپ منحنی توزیع قرار داریم.



شکل ۳: ارزش درمعرض ریسک شرطی (منبع: یافته‌های پژوهش)

در این بخش که نوآوری پژوهش نیز می‌باشد، سنجه‌ریسک‌های VAR و CVAR با نظریه فازی ادغام شده و برای عمل بهینه‌سازی و ساخت پرتفو مورداستفاده قرار گرفته است. باتوجه به غیرقطعی بودن بازده‌های پیش‌بینی شده از الگوریتم جنگل تصادفی استفاده، از روش‌های فازی باعث دقت بیشتری در مدل‌سازی می‌شود. هم‌چنین استفاده از سنجه‌ی ریسک ارزش در معرض خطر مشروط به این‌که اندازه ضرر را به سرمایه‌گذار نشان می‌دهد، به تصمیم‌گیری بهتر کمک می‌کند. در این مقاله با استفاده

از سنجه ارزش در معرض ریسک مشروط و تخمین آن به‌وسیله نظریه اعتبار فازی اقدام به بهینه‌سازی سبدسرمایه‌گذاری شده‌است. به‌این‌منظور، بازده انتظاری‌پرتفوی به‌وسیله‌میانگین اعتبارفازی‌به‌دست آمده و سپس ارزش در معرض ریسک مشروط به وسیله همین نظریه تخمین زده شده است. محدودیت‌های درنظر گرفته شده برای مدل بهینه سازی، شامل محدودیت سقف و کف نسبت خرید هر سهم، که حداقل آن صفر و حداکثر آن ۰,۷ است و هم‌چنین محدودیت حداقلی از بازده می‌باشد.

$$\mathrm{Var}(\alpha) = -\varphi^{-1}(\alpha) = \begin{cases} 2(r_1 - r_2)\alpha - r_2 & \alpha \le 0.5 \\ 2(r_2 - r_3)\alpha + r_3 - 2r_2 & \alpha > 0.5 \end{cases} \quad (6)$$

$$\mathrm{Min} - \sum_{i=1}^{n} [\mathrm{VaR}(\alpha)] x_i$$

$$\text{s.t: } \frac{1}{4}\sum_{i=1}^{n} [r_1 + 2r_2 + r_3] x_i \ge \mu_p$$

$$\sum_{j=1}^{n} x_j = 1 \qquad j = 1,2,3,\dots,n$$

$$l_i y_i \le x_i \le u_i y_i \qquad y_i = \{0,1\}$$

$$x_i \ge 0$$

$$CVar(\alpha) = \frac{1}{1-\alpha}\int_{\alpha}^{1} Var(r)\,dr = \begin{cases} \alpha r_1 - (1+\alpha) r_2 & \alpha \le 0.5 \\ (\alpha - 1) r_2 - \alpha r_3 & \alpha > 0.5 \end{cases} \quad (7)$$

$$\mathrm{Min} - \sum_{i=1}^{n} [\mathrm{CVaR}(\alpha)] x_i \quad (8)$$

$$\text{s.t: } \frac{1}{4}\sum_{i=1}^{n} [r_1 + 2r_2 + r_3] x_i \ge \mu_p$$

$$\sum_{j=1}^{n} x_j = 1 \qquad j = 1,2,3,\dots,n$$

$$l_i y_i \le x_i \le u_i y_i \qquad y_i = \{0,1\}$$

$$x_i \ge 0$$

تحلیل‌های تجربی

در این پژوهش ابتدا ۳۰ نماد بورسی، براساس عمق بازار بالا و هم‌چنین تنوع صنایع، انتخاب شده، سپس سهم‌های انتخابی وارد ۲ الگوریتم یادگیری ماشین (جنگل تصادفی و بیز ساده) شده و مرحله کلاس‌بندی سهم‌ها اتفاق می‌افتد. این کلاس‌بندی طبق ۱۴ ویژگی تکنیکال و نرخ بازده می‌باشد. بعداز

این مرحله کلاس مثبت هر الگوریتم، به‌وسیله الگوریتم جنگل تصادفی پیش‌بینی می‌شود. پیش‌بینی برای هر سهم و هر خروجی الگوریتم به تعداد ۳ مرحله صورت گرفته و علت آن نیز استفاده از نظریه فازی و ترکیب آن با سنجه‌های ریسک VAR و CVAR برای تشکیل پرتفوی می‌باشد. درنهایت نیز خروجی‌های هر الگوریتم یک‌مرتبه به‌وسیله سنجه‌ریسک VAR و یک مرتبه نیز به‌وسیله سنجه‌ریسک CVAR بهینه خواهد شد و در پایان خروجی‌ها با یکدیگر مقایسه می‌شوند. برای بهینه‌سازی سبد سرمایه‌گذاری به‌وسیله دو سنجه‌ریسک مطرح شده، از دو سطح اطمینان ۹۰ و ۹۹ درصد استفاده شده و در مرحله آخر پرتفوهای تشکیل شده با بازده واقعی بازار و استفاده از معیار ترینر برای مقایسه آن، استفاده می‌شوند.

شرکت‌های مورد بررسی در پژوهش

فرآیند انتخاب شرکت‌های مورد بررسی در این تحقیق با رویکرد توجه به صنایع مختلف بورس و هم‌چنین عمق بازار این سهم‌ها می‌باشد. در این انتخاب سعی شده که اکثریت صنایع مهم بورس در فرآیند، مورد بررسی قرار گیرند. در جدول ۲ شرکت‌های مورد بررسی در پژوهش ذکر شده است.

جدول ۲: شرکت‌های مورد بررسی در پژوهش

ردیف	نماد	شرکت	ردیف	نماد	شرکت
۱	اخابر	مخابرات ایران	۱۶	زاگرس	پتروشیمی زاگرس
۲	همراه	ارتباطات سیار ایران	۱۷	شبریز	پالایشگاه نفت تبریز
۳	وسینا	بانک سینا	۱۸	شبندر	پالایشگاه نفت بندرعباس
۴	وبملت	بانک ملت	۱۹	ولساپا	لیزینگ رایان سایپا
۵	پارسیان	بیمه پارسیان	۲۰	وایران	لیزینگ ایرانیان
۶	آسیا	بیمه اسیا	۲۱	وصنعت	سرمایه‌گذاری توسعه صنعت و تجارت
۷	خبهمن	گروه بهمن	۲۲	وخارزم	سرمایه‌گذاری خوارزمی
۸	خرینگ	رینگ‌سازی مشهد	۲۳	ومعادن	توسعه معادن و فلزات
۹	خزامیا	زامیاد	۲۴	ثفارس	عمران و توسعه فارس
۱۰	فولاد	فولاد مبارکه اصفهان	۲۵	رکیش	کارت اعتباری ایران کیش
۱۱	فملی	ملی صنایع مس ایران	۲۶	فاراک	ماشین‌سازی اراک
۱۲	سغرب	سیمان غرب	۲۷	دارو	کارخانجات داروپخش
۱۳	سکرد	سیمان کردستان	۲۸	قزوین	کارخانجات قند قزوین
۱۴	شفن	پتروشیمی فناوران	۲۹	غشاذر	پگاه آذربایجان غربی
۱۵	کرماشا	صنایع پتروشیمی کرمانشاه	۳۰	حکشتی	کشتی‌رانی جمهوری اسلامی ایران

منبع: یافته‌های پژوهش

ویژگی‌های مورد بررسی در الگوریتم‌های مورد استفاده

در این قسمت از پژوهش، ویژگی‌ها (فیچرهای) استفاده شده در الگوریتم‌های مورد بررسی ارایه شده است. این ویژگی‌ها ترکیبی از شاخص‌های تکنیکال و بازدهی های مختلف است. در انتخاب اندیکاتورها/اسیلاتورها در تحلیل تکنیکال، معیاری‌که برای ما در این مقاله مهم بوده سازگاری آن با بازار سرمایه ایران است. این هدف زمانی محقق می‌شود که از بین تعداد زیادی از اندیکاتور/اسیلاتور، آن‌هایی برگزیده می‌شوند که حتی‌المقدور پارامتر میانگین را در فرمول خود دارا باشند، زیرا این پارامتر نسبت به باقی پارامترها بیشترین کمک را در مدل‌سازی و پیش‌بینی قیمت‌ها به محقق (معامله‌گران) خواهد کرد.

در جدول ۳ ویژگی‌های مورد بررسی در الگوریتم‌ها ذکر شده است.

جدول ۳: ویژگی‌های مورد بررسی در الگوریتم‌ها

ردیف	نام ویژگی (فیچر)	ادیکاتور / اسیلاتور	پارامترهای مورد استفاده	فرمول
۱	EMA	Exponential Moving Average	close (i,i-1,i-2,…)	$\left(\text{Value}_i * (\frac{\text{smoothing}}{1+\text{days}})\right) + \text{EMA}_{i-1}(1-(\frac{\text{smoothing}}{1+\text{days}}))$
۲	ADX	Average Directional Index	DI^+, DI^-, ADX	$\text{TR}_{i+1} = \text{TR}_{i-1} - \frac{\text{TR}_{i-1}}{14} + \text{TR}_i$
۳	RSI	Relative Strength Index	Profit , loss	$100 - \frac{100}{1+\text{RS}}$
٤	Stoch	Stochastic	Close , low , high	$100 * \frac{C - L_{14}}{H_{14} - L_{14}}$
٥	MACD	Moving Average Convergence Divergence	EMA	$\text{EMA}(10) - \text{EMA}(20)$
٦	CCI	Commodity Channel Index	High , low , close	$\frac{\text{price} - \text{ma}}{0.015 * d}$
۷	ATR	Average True Range	High , low	$H - L$
۸	OBV	On Balanced Volume	Volume	$\text{OBV}(t) = \text{OBV}(t-1) \pm \text{Vol}$
۹	TRIX	Triple Exponential Average	EMA	$\text{EMA}(a, b, c)$
۱۰	R(-1)	Return(-1)	Close i , close i-1	$\ln(\frac{\text{close i}}{\text{close i}-1})$
۱۱	R(-2)	Return(-2)	Close i-1,close i-2	$\ln(\frac{\text{close i}-1}{\text{close i}-2})$
۱۲	R(-3)	Return(-3)	Close i-2,close i-3	$\ln(\frac{\text{close i}-2}{\text{close i}-3})$

۱۳	R(-4)	Return(-4)	Close i-3,close i-4	$\ln(\frac{\text{close i}-3}{\text{close i}-4})$
۱٤	R(-5)	Return(-5)	Close i-4,close i-5	$\ln(\frac{\text{close i}-4}{\text{close i}-5})$

منبع: یافته‌های پژوهش

در این راستا در جدول ۴ ویژگی های مدل مورد استفاده به‌صورت خلاصه قابل ملاحظه می باشد.

جدول ۴: ویژگی‌های مدل مورد استفاده

۱	الگوریتم مورد استفاده برای کلاس‌بندی	جنگل تصادفی – بیز ساده
۲	پیش‌بینی	جنگل تصادفی
۳	بهینه‌سازی(فازی)	VaR – CvaR
۴	سطح اطمینان(درصد)	۰,۹ و ۰,۹۹

جنگل تصادفی

در این بخش، داده‌های سری‌زمانی سهم‌ها وارد ماشین RF می‌شوند. بعداز فیلترشدن آن‌ها از طریق ویژگی‌های وارده در جدول ۳، خروجی موردنظر براساس دو کلاس مثبت و منفی در قالب جدول ۵ ارائه می‌گردد. ملاحظه می‌شود ۹ سهم طبق این الگوریتم، مجوز حضور در کلاس مثبت را ازآن خود کرده‌اند.

جدول ۵: خروجی الگوریتم جنگل تصادفی

	اخابر	همراه	وِسینا	وبملت	پارسیان	آسیا	خبهمن	خرینگ	خزامیا	فولاد
کلاس	–	–	–	–	+	+	–	–	–	–
دقت	٪۵۵	٪۵۸	٪۵۷	٪۵۶	٪۶۰	٪۵۹	٪۵۳	٪۴۸	٪۵۰	٪۶۲
	فملی	سغرب	سکرد	شفن	کرماشا	زاگرس	شبریز	شبندر	ولساپا	وایران
کلاس	+	–	–	+	–	–	–	–	–	–
دقت	٪۵۳	٪۵۷	٪۵۵	٪۶۵	٪۵۸	٪۵۱	٪۵۴	٪۵۳	٪۵۸	٪۵۴
	وصنعت	وخارزم	ومعادن	ثفارس	رکیش	فاراک	دارو	قزوین	غشاذر	حکشتی
کلاس	+	–	+	+	–	+	–	–	+	–
دقت	٪۵۴	٪۴۸	٪۴۸	٪۵۳	٪۴۹	٪۵۶	٪۶۰	٪۵۹	٪۴۸	٪۵۷

منبع: یافته‌های پژوهش

شرکت‌هایی که وارد کلاس مثبت شده‌اند (هم در روش جنگل تصادفی و هم در روش بیز ساده)، برای پیش‌بینی به‌وسیله الگوریتم جنگل تصادفی، وارد مرحله بعد می‌شوند.

بیز ساده

در این بخش (جدول ۶)، داده‌های سری‌زمانی سهم‌ها وارد ماشین NB می‌شوند. بعداز فیلترشدن آن‌ها از طریق ویژگی‌های وارده، خروجی موردنظر براساس دو کلاس مثبت و منفی در قالب جدول زیر ارائه می‌گردد. ملاحظه می‌شود ۱۸ سهم طبق این الگوریتم، مجوز حضور در کلاس مثبت را ازآن خود کرده‌اند.

جدول ۶: خروجی الگوریتم بیز ساده

	اخابر	همراه	وسینا	وبملت	پارسیان	آسیا	خبهمن	خرینگ	خزامیا	فولاد
کلاس	–	+	+	+	–	+	+	–	+	+
دقت	٪۴۸	٪۴۱	٪۵۶	٪۵۴	٪۵۴	٪۵۷	٪۴۸	٪۴۰	٪۵۴	٪۴۶
	فملی	سغرب	سکرد	شفن	کرماشا	زاگرس	شبریز	شبندر	ولساپا	وایران
کلاس	+	–	+	+	+	+	–	–	+	–
دقت	٪۴۷	٪۵۷	٪۵۹	٪۵۷	٪۴۹	٪۵۲	٪۴۷	٪۴۶	٪۴۴	٪۵۱
	وصنعت	وخارزم	ومعادن	ثفارس	رکیش	فاراک	دارو	قزوین	غشاذر	حکشتی
کلاس	–	+	+	–	+	+	–	–	–	+
دقت	٪۴۹	٪۵۴	٪۴۱	٪۴۸	٪۵۰	٪۶۱	٪۴۶	٪۴۷	٪۴۷	٪۴۶

منبع: یافته‌های پژوهش

پیش‌بینی با استفاده از الگوریتم جنگل تصادفی

در این بخش کلیه خروجی‌های الگوریتم‌ها، که برای عمل کلاس‌بندی استفاده شده‌اند، به‌وسیله الگوریتم جنگل تصادفی در جداول ۷ و ۸ پیش‌بینی می شوند. استفاده از نظریه فازی و سه پیش‌بینی خوش‌بینانه (r1)، محتمل (r2) و بدبینانه (r3) باعث نوآوری جدیدی در این پژوهش شده است. برای هر سهم که در مرحله قبل مجوز حضور در کلاس مثبت را گرفته است (به‌تفکیک دو روش جنگل تصادفی و بیز ساده)، تعداد سه پیش‌بینی (۳ عدد تصادفی) برای روز بعد شناسایی می‌شود، که این ۳ عدد ورودی‌های بهینه‌سازی پرتفوی سرمایه‌گذاری به روش سنجه ریسک و نظریه فازی است.

جدول ۷: پیش‌بینی براساس خروجی الگوریتم جنگل تصادفی

ردیف	سهم	r_1	r_2	r_3	ردیف	سهم	r_1	r_2	r_3
۱	آسیا	٪۰,۹	٪۱,۱۵	٪۱,۱۶	۶	سفارس	٪۰,۴	٪۰,۵	٪۱,۸
۲	فاراک	٪۱,۱	٪۱,۳	٪۲,۱	۷	شفن	٪۰,۰۸	٪۰,۱	٪۱,۹
۳	فملی	-٪۰,۱	۰	٪۰,۹۲	۸	ومعادن	-٪۰,۸	٪۰,۰۱	٪۳,۲۴

۴	غشاذر	٪۰,۵	٪۰,۷	٪۰,۸	۹	وصنعت	٪۰,۷	٪۰,۸	٪۱,۲
۵	پارسیان	٪۰,۷	٪۰,۸	٪۱,۱					

منبع: یافته‌های پژوهش

جدول ۸: پیش‌بینی براساس خروجی الگوریتم بیز ساده

ردیف	سهم	r_1	r_2	r_3	ردیف	سهم	r_1	r_2	r_3
۱	آسیا	٪۰,۹	٪۱,۱۵	٪۱,۱۶	۱۰	رکیش	-٪۰,۲	۰	٪۱,۱۲
۲	فاراک	٪۱,۱	٪۱,۳	٪۲,۱	۱۱	سکرد	-٪۰,۵	٪۰,۲	٪۱,۸
۳	فملی	-٪۰,۱	۰	٪۰,۹۲	۱۲	شفن	٪۰,۰۸	٪۰,۱	٪۱,۹
۴	فولاد	٪۰,۱	٪۰,۲	٪۲,۲۸	۱۳	وبملت	-٪۰,۳	٪۰,۶	٪۱,۷۳
۵	همراه	٪۰,۰۶	٪۰,۰۹	٪۰,۳	۱۴	وخارزم	-٪۱,۸	-٪۰,۲	٪۲,۵
۶	حکشتی	٪۱,۲	٪۲,۲۷	٪۲,۹۲	۱۵	ولساپا	٪۰,۴	٪۱,۲۴	٪۴,۲
۷	کرماشا	٪۰,۹	٪۱,۰۵	٪۲,۴	۱۶	ومعادن	-٪۰,۸	٪۰,۰۱	٪۳,۲۴
۸	خبهمن	-٪۱,۴	-٪۰,۲	٪۰,۳۴	۱۷	وسینا	٪۰,۳	٪۰,۴	٪۱,۲
۹	خزامیا	٪۱,۶	٪۱,۹۲	٪۱,۹۵	۱۸	زاگرس	-٪۱,۲	٪۰,۳	٪۲,۵۸

منبع: یافته‌های پژوهش

جداول بهینه‌سازی خروجی الگوریتم‌ها

در جداول ۹ و ۱۰ مقادیر بهینه‌سازی خروجی الگوریتم جنگل تصادفی بااستفاده از ارزش درمعرض ریسک، ارایه شده است. در مرحله سوم ، بهینه‌سازی سبد سهام با استفاده از دو روش VaR و Cvar انجام شده که در هریک از این روش‌ها، از دو سطح‌اطمینان ۹۰ و ۹۹ درصد بهره می‌گیریم و خروجی هریک را به‌تفیفک ذکر کرده‌ایم. در مرحله چهارم و آخر، باتوجه به نسبت پیشنهادی خرید هر سهم، بازده پرتفوی با استفاده از فرمول R_{index} محاسبه خواهد شد.

جدول ۹: جدول بهینه‌سازی خروجی الگوریتم جنگل تصادفی

ردیف	متغیر	نماد	$W(\alpha_1 = 0.1)$	$W(\alpha_2 = 0.01)$	ردیف	متغیر	نماد	$W(\alpha_1 = 0.1)$	$W(\alpha_2 = 0.01)$
۱	X_1	آسیا	۰,۲۲	۰	۶	X_6	سفارس	۰	۰,۰۰۸۷
۲	X_2	فاراک	۰	۰,۰۰۹۵	۷	X_7	شفن	۰,۱۱	۰
۳	X_3	فملی	۰	۰,۲۷	۸	X_8	ومعادن	۰	۰,۱۶۹
۴	X_4	غشاذر	۰	۰,۵۲	۹	X_9	وصنعت	۰,۳۳۳	۰
۵	X_5	پارسیان	۰,۳۴۷	۰,۰۹۶					

منبع: یافته‌های پژوهش

جدول ۱۰: مقایسه سطح‌اطمینان‌های متفاوت با بازده واقعی درصورت وجود وزن‌دهی پیشنهادی

$\alpha_2 = 0.01$	$\alpha_1 = 0.1$	
٪۳,۰۵	٪۲,۶۱	R_{Index}

منبع: یافته‌های پژوهش

در جداول ۱۱ و ۱۲ مقادیر بهینه‌سازی خروجی الگوریتم بیز ساده بااستفاده از ارزش درمعرض ریسک، ارایه شده است.

جدول ۱۱: جدول بهینه‌سازی خروجی الگوریتم بیز ساده

ردیف	متغیر	نماد	$W(\alpha_1 = 0.1)$	$W(\alpha_2 = 0.01)$	ردیف	متغیر	نماد	$W(\alpha_1 = 0.1)$	$W(\alpha_2 = 0.01)$
۱	X_1	آسیا	۰,۰۲۱	۰,۱۲۲	۱۰	X_{10}	رکیش	۰	۰,۱۷۸
۲	X_2	فاراک	۰۱۳۴	۰,۰۸	۱۱	X_{11}	سکرد	۰,۰۹۹۲	۰
۳	X_3	فملی	۰	۰	۱۲	X_{12}	شفن	۰	۰
۴	X_4	فولاد	۰	۰	۱۳	X_{13}	وبملت	۰	۰
۵	X_5	همراه	۰	۰	۱۴	X_{14}	وخارزم	۰	۰,۶۱
۶	X_6	حکشتی	۰,۲۸۷	۰	۱۵	X_{15}	ولساپا	۰	۰
۷	X_7	کرماشا	۰	۰	۱۶	X_{16}	ومعادن	۰,۱۸۲	۰
۸	X_8	خبهمن	۰	۰	۱۷	X_{17}	وسینا	۰,۲	۰
۹	X_9	خزامیا	۰,۰۶۹۲	۰	۱۸	X_{18}	زاگرس	۰	۰

منبع: یافته‌های پژوهش

جدول ۱۲: مقایسه سطح‌اطمینان‌های متفاوت با بازده واقعی در صورت وجود وزن‌دهی پیشنهادی

$\alpha_2 = 0.01$	$\alpha_1 = 0.1$	
٪۷,۷۶	٪۶,۳۲	R_{Index}

منبع: یافته‌های پژوهش

در جداول ۱۳ و ۱۴ مقادیر بهینه‌سازی خروجی الگوریتم جنگل تصادفی بااستفاده از ارزش درمعرض ریسک شرطی، ارایه شده است.

جدول ۱۳: جدول بهینه‌سازی خروجی الگوریتم جنگل تصادفی

ردیف	متغیر	نماد	$W(\alpha_1 = 0.1)$	$W(\alpha_2 = 0.01)$	ردیف	متغیر	نماد	$W(\alpha_1 = 0.1)$	$W(\alpha_2 = 0.01)$
۱	X_1	آسیا	۰	۰,۲	۶	X_6	سفارس	۰	۰
۲	X_2	فاراک	۰,۲۹	۰	۷	X_7	شفن	۰,۲۵	۰

۳	X_3	فملی	۰,۰۸	۰,۰۸	۸	X_8	ومعادن	۰	۰
۴	X_4	غشاذر	۰	۰	۹	X_9	وصنعت	۰	۰
۵	X_5	پارسیان	۰,۷۲	۰,۳۸					

منبع: یافته‌های پژوهش

جدول ۱۴: مقایسه سطح‌اطمینان‌های متفاوت با بازده واقعی درصورت وجود وزن‌دهی پیشنهادی

$\alpha_2 = 0.01$	$\alpha_1 = 0.1$	
٪۱۱,۴۳	٪۷,۷۳	R_{Index}

منبع: یافته‌های پژوهش

در جداول ۱۵ و ۱۶ مقادیر بهینه‌سازی خروجی الگوریتم بیز ساده بااستفاده از ارزش درمعرض ریسک شرطی، ارایه شده است.

جدول ۱۵: جدول بهینه‌سازی خروجی الگوریتم بیز ساده

ردیف	متغیر	نماد	$W(\alpha_1 = 0.1)$	$W(\alpha_2 = 0.01)$	ردیف	متغیر	نماد	$W(\alpha_1 = 0.1)$	$W(\alpha_2 = 0.01)$
۱	X_1	آسیا	۰,۰۸	۰	۱۰	X_{10}	رکیش	۰,۱۶	۰
۲	X_2	فاراک	۰	۰	۱۱	X_{11}	سکرد	۰	۰,۲۳
۳	X_3	فملی	۰	۰,۲۴	۱۲	X_{12}	شفن	۰	۰
۴	X_4	فولاد	۰,۳۳	۰,۳۷	۱۳	X_{13}	وبملت	۰,۲	۰
۵	X_5	همراه	۰	۰	۱۴	X_{14}	وخارزم	۰,۱۵	۰
۶	X_6	حکشتی	۰	۰	۱۵	X_{15}	ولساپا	۰	۰
۷	X_7	کرماشا	۰	۰,۱۶	۱۶	X_{16}	ومعادن	۰,۰۸	۰
۸	X_8	خبهمن	۰	۰	۱۷	X_{17}	وسینا	۰	۰
۹	X_9	خزامیا	۰	۰	۱۸	X_{18}	زاگرس	۰	۰

منبع: یافته‌های پژوهش

جدول ۱۶: مقایسه سطح‌اطمینان‌های متفاوت با بازده واقعی درصورت وجود وزن‌دهی پیشنهادی

$\alpha_2 = 0.01$	$\alpha_1 = 0.1$	
٪۱۰,۹۱	٪۸,۷۷	R_{Index}

منبع: یافته‌های پژوهش

نتیجه‌گیری

در ادامه به بررسی نتایج حاصل از بکارگیری مدل‌های بررسی‌شده و محدودیت‌های ناشی از آن‌ها

خواهیم پرداخت. در این تحقیق بازده بدون‌ریسک سالانه ۱۸ درصد درنظر گرفته شده است که برای یک ماه ۱,۵ درصد برآورد شده و در مقایسات استفاده می‌شود. ازطرفی بازده ماهانه بازار در سال ۱۳۹۸ بین ماه‌های شهریور و مهر، ۲,۸۴٪ می‌باشد. در جدول ۱۷ خروجی‌های موردنظر از الگوریتم‌ها و هم‌چنین سنجه‌های ریسک به‌همراه سطوح‌اطمینان مختلف رتبه‌بندی شده تا دیدی واضح و مشخص را به خوانندگان انتقال دهد. پر واضح است که سنجه‌ریسک CVAR قابلیت و نتیجه بهتری را نسبت به سنجه ریسک VAR داشته است. هم‌چنین الگوریتم جنگل تصادفی، نتایج بهتری را رقم زده است. در تحلیل نتایج، بازدهی واقعی سهام موردبررسی با وزن خروجی از الگوریتم ترکیب شده و رتبه‌بندی صورت گرفته است. درآخر بازدهی هرکدام از روش‌ها با معیار ترینر نسبت به بازده بازار (شاخص کل) قیاس شده و به‌ترتیب آورده شده است.

جدول ۱۷: رتبه‌بندی الگوریتم‌ها و سطح‌اطمینان‌های مختلف به تفکیک بازدهی

ردیف	بازدهی	الگوریتم	سنجه‌ریسک	سطح‌اطمینان	$R_p - R_f$	β_p	مقدار ترینر برای پرتفوی	مقدار ترینر برای بازار
۱	٪۱۱,۴۳	جنگل تصادفی	CVAR	α_2	٪۹,۹۳	۱,۱۶	٪۸,۰۹	٪ ۱,۳۴
۲	٪۱۰,۹۱	بیز ساده	CVAR	α_2	٪۹,۴۱	۱,۹۴	٪۴,۸۵	٪ ۱,۳۴
۳	٪۸,۷۷	بیز ساده	CVAR	α_1	٪۷,۲۷	۱,۹۴	٪۳,۷۴	٪ ۱,۳۴
۴	٪۷,۷۶	بیز ساده	VAR	α_2	٪۶,۲۶	۱,۹۴	٪۳,۲۲	٪ ۱,۳۴
۵	٪۷,۷۳	جنگل تصادفی	CVAR	α_1	٪۶,۲۳	۱,۱۶	٪۵,۳۷	٪ ۱,۳۴
۶	٪۶,۳۲	بیز ساده	VAR	α_1	٪۴,۸۲	۱,۹۴	٪۲,۴۸	٪ ۱,۳۴
۷	٪۳,۰۵	جنگل تصادفی	VAR	α_2	٪۱,۵۵	۱,۱۶	٪۱,۳۳	٪ ۱,۳۴
۸	٪۲,۶۱	جنگل تصادفی	VAR	α_1	۱,۱۱	۱,۱۶	٪۰,۹۵	٪ ۱,۳۴

منبع: یافته‌های پژوهش

در پایان به این نتیجه خواهیم رسید که اگر از الگوریتم جنگل تصادفی برای کلاس‌بندی سهام (یافتن سهام رو به رشد)، سنجه ریسک CvaR و سطح اطمینان ۹۹ درصد برای آن استفاده کنیم، بازدهی که به معامله‌گر می‌دهد برابر با ۱۱,۴۳ درصد است. به کمک روش ترینر این عدد را استاندارد کرده که بتوان حالت مقایسه‌ای را بین بازده پرتفو و بازده شاخص ایجاد کرد. در نهایت باتوجه به برتری عدد ۸,۰۹ درصد در مثابل ۱,۳۴ درصد بازار، به این نتجه می‌رسیم به که سبد حاصله به‌اندازه ۶,۷۵ درصد نسبت به شاخص کل بازدهی بهتری را داشته است. بقیه سطوح جدول نیز به همین روش تحلیل می‌شوند.

منابع

۱) ابزری، م. (۱۳۸۴). بهینه‌سازی سبد سرمایه‌گذاری بااستفاده از روش‌های برنامه‌ریزی خطی و ارائه یک مدل کاربردی . مجله علوم اجتماعی و انسانی دانشگاه شیراز.

۲) البرزی، م. (۱۳۸۰). آشنای با شبکه‌های عصبی مصنوعی. تهران: موسسه انتشارات علمی.

۳) بیگلری، ب. (۱۳۸۹). مقایسه مدل‌های انتخاب سهام جهت تشکیل پرتفو از نظر برآ.رد بازده مورد انتظار، بازده واقعی آتی و ریسک آن‌ها در بورس اوراق بهادار تهران. پنجمین کنفرانس توسعه نظام مالی در ایران.

۴) بیگی، ع. (۱۳۸۹). بهینه‌سازی پرتفوی سهام بااستفاده از روش حرکت تجمعی ذرات. تحقیقات مالی دانشکده مدیریت دانشگاه تهران .

۵) ۵ پیکارجو، ک. (۱۳۸۸). اندازه‌گیری ریسک دارایی شرکت‌ها و موسسات مالی بااستفاده از روش ارزش در معرض ریسک. نشریه پژوهش‌نامه اقتصادی, ۹۵-۲۲۱.

۶) راعی، ر. (۱۳۹۳). مدیریت سرمایه‌گذاری پیشرفته. تهران: انتشارات سمت.

۷) سینایی، ح. (۱۳۹۳). تصمیم‌گیری برای انتخاب سبد سهام، مقایسه الگوریتم‌های ژنتیک و زنبورعسل. پژوهش‌نامه مدیریت اجرایی علمی پژوهشی.

۸) فضل‌زاده، ع. (۱۳۹۰). بررسی توانایی مدلهای تک شاخص شارپ و تحلیل پوششی داده‌ها در انتخاب پرتفوی کارا در بورس اوراق بهادار تهران. فصل‌نامه بورس اوراق بهادار .

۹) فیضی، ژ. (۱۳۹۱). بررسی روش‌های مونت کارلو برای تقریب کارایی در معرض خطر و ارزش در معرض خطر شرطی. سومین کنفرانس ریاضیات مالی و کاربردی. سمنان: دانشگاه سمنان.

۱۰) فرح آبادی، مهرداد، عیوض لو، رضا، صفری، حسین. (۱۴۰۱). بکارگیری شبکهٔ هوش مصنوعی و مدل شبکهٔ بیـزین برای پیش بینی ریسک نقدینگی در صنعت بانکـداری. فصلنامه بورس اوراق بهادار،(۵۹) ۱۵، ۱۰۰-۸۱.

۱۱) مهرجردی، ز. (۱۳۹۱). الگوریتم هوشمند ترکیبی برپایه مدل فازی میانگین-واریانس چولگی برای انتخاب پرتفولیو. نشریه بین‌المللی مهندسی صنایع و مدیریت تولید.

12) Ajit Kumar Pasayat. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets.

13) Bishop, T. (2000). variational relevance vector machines. proceedings of the sixteenth conference on uncertainty in artificial intelligence. Morgan kaufmann publishers .

14) Bustos. (2018). Multiobjective Genetic Programming, Redusing Bloat by using SPEA2. In Congress on Evolutionary Computation.

15) Chakraborty et al. (2018). A hybrid stock selection model using Genetic Algorithms and Support Vector Regression. Department of Computer Science and Information Engineerin.

16) Coyne. (2018). An Effective Decision Basic Genetic Algorithm Approach to Multiobjective Portfolio Optimization Problem. Applied Mathematical sciences.

17) Fischer and krauss. (2018). Robust Portfolio Optimization. john wiley.

18) Hu et al .(2018) .the elements of statistical learning .soringer new york.

19) Hung et al. (2018). the elements of statistical learning, data mining. inference and prediction.

20) Hakimeh Morteza. (2023). An improved learning automata based multi-objective whale optimization approach for multi-objective portfolio optimization in financial markets.

21) Kia. (2018). A Double-Stage Genetic Optimization Algorithm for Portfolio Selection. ICONIP 2006.

22) Liu. (2018) .data mining techniques .wiley.

23) Malagrino et al. (2018). Machine learning in sentiment reconstruction of the simulated stock market. statistical mechanics and its applications.

24) Mikhail Goykhman. (2019). A Novel Automatic satire and irony detection using ensembled feature selection and data mining. Elsevier.

25) Vu Minh Ngo. (2023). Does reinforcement learning outperform deep learning and traditional portfolio optimization models in frontier and developed financial markets?

26) Ren. (2018). Predicting the stock price of frontier markets using machine learning and modified Black–Scholes Option pricing model. statistical mechanics and its applications.

27) Reaz Chowdhury. (2001). A hybrid Genetic Quantitative Method for Risk-Return Optimization for Credit Portfolio Institute AIFB. Faculty of Economics.

28) Sotirios P. (2006). Portfolio Performance Evaluation in a mean-variance-skewness framework. European Journal of Operational Research.

29) Thomas Fischer. (2017). Deep adaptive group-based input normalization for financial trading. pattern recogniting letters.

30) Wang. (2018). Mean-Variance-skewness Model for Portfolio Selection With Transaction Costs. Information Journal of ystems Science.

31) wuyu, w. (2020). Portfolio formation with preselection using deep learning from long-term financial data. expert system with applications.

32) Zhang. (2018). Comparative analysis of expected shortfall and value at risk . institute for monetary and economic studies .

33) Zhou. (2018). support vector classification with input data uncertainty. advances in neural information processing systems.

34) Yaping Cai (2019).A comprehensive cluster and classification mining procedure for daily stock market return forecasting .neurokomputing

35) Zolfaghari, M., & Gholami, S. (2020). A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction. Expert Systems with Applications, 182, 115149

یادداشت‌ها:

1	Investment	15	Bustos et al	29	Thomas Fischer
2	Portfolio	16	Chakraborty et al	30	Mikhail Goykhman
3	Markowitz	17	Coyne et al	31	Yaping Caia
4	Machine Learning	18	Fischer and Krauss	32	Codruț-Florin Ivașcu
5	Other category	19	Hu et al	33	Reaz Chowdhury
6	Decision tree	20	Huang et al	34	Zolfaghari & Gholami
7	Bayasian models	21	Kia et al	35	Wuyu Wang
8	Ensemble models	22	Liu et al	36	Classification
9	Genetic algorithms	23	Malagrino et al	37	Random Forest
10	Fuzzy logic	24	Ren et al	38	Ravi
11	Artificial neutral networks	25	Wang. Xu. et al	39	Naïve Bayes ۱
12	Deep learning	26	Zhang et al	40	Treyner
13	Support vector machine	27	Zhou et al	41	Feature
14	Tsantekidis et al	28	Sotirios P. Chatzis		

Presenting the combined algorithm of machine learning and the combination of risk metrics and fuzzy theory in choosing an investment portfolio

Danial Mohammadi[1]
Sayyed jafar Sajadi[2]
Emran Mohammadi[3]
Naeim Shokri[4]

Receipt: 01/03/2023 Acceptance: 07/08/2023

Abstract

The current research was conducted to find the optimal portfolio for investing in exchange stocks, and one of the methods that is currently very popular among analysts and researchers in this field is methods based on artificial intelligence,followed by methods aimed at reducing risk metrics. The purpose of the research is to form a portfolio using machine learning methods, measure risk and combine it with fuzzy theory, which has a better return than the average return of the market.The output of each method is entered into the random forest algorithm and prediction is made by this algorithm, and in the last step, the prediction output is entered into the value-at-risk and value-at-risk optimization model with the fuzzy theory approach to form the capital portfolio. Shares information is daily and its time period is from the beginning of 2014 to the middle of 2018.At the end of each of these methods and steps, it was compared with the real return of the market.Based on the obtained results, the CVAR risk measure has a better capability than the VAR risk measure,and the random forest algorithm among the used machine learning algorithms has achieved better results in choosing the investment portfolio.

Keywords

Machine Learning, Naïve Bayes (NB), Conditional Value at Risk (CVaR), Investment Portfolio, Tehran Stock Exchange. **Jel Classification:** G01, G32, E44

1-Department of Financial Engineering, Faculty of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran. dnl_mhm@yahoo.com
2-Department of Financial Engineering, Faculty of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran. sjsadjadi@iust.ac.ir
3-Department of Financial Engineering, Faculty of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran. e_mohammadi@iust.ac.ir
4-Department of Development and Economic Planning, Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran. (Corresponding Author) n.shokri@modares.ac.ir