



فصلنامه مهندسی مالی و مدیریت اوراق بهادار
دوره شانزده / شماره شصت و چهار / پائیز ۱۴۰۴

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی بهینه بر پایه شاخص‌های بنیادی

هدیه غفوری وایقان^۱، رباب کلانتری^۲

۱. دانشجوی کارشناسی ارشد، مهندسی صنایع گرایش مهندسی مالی، دانشکده فنی دانشگاه خاتم، تهران، ایران.
۲. استادیار، گروه مالی، دانشگاه خاتم، تهران، ایران

نویسنده مسئول مقاله:

* E-mail: r.kalantari@khatam.ac.ir

شماره تلفن نویسنده مسئول: ۰۹۹۱۲۰۴۴۱۵۰

شناسه پژوهشگر (شماره ارجید) نویسنده مسئول:

[0000-0002-1262-8933](https://doi.org/10.22092/0000-0002-1262-8933)

نویسنده‌ی اول مقاله: هدیه غفوری وایقان

0009-0006-5225-1858

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی بهینه بر پایه شاخص‌های بنیادی

چکیده

در شرایط پرنوسان و ناپایدار بازارهای مالی، بهینه‌سازی تصمیمات سرمایه‌گذاری، به‌ویژه در بازارهای نوظهور، اهمیتی دوچندان دارد. هدف این پژوهش، طراحی و ارزیابی چارچوبی ترکیبی برای انتخاب و تخصیص پرتفوی سهام در بازار سرمایه ایران است که در آن الگوریتم‌های یادگیری ماشین برای پیش‌گزینش دارایی‌ها و مدل مارکویتز برای بهینه‌سازی پرتفوی به‌کار گرفته می‌شوند. داده‌های بنیادی و قیمتی ۱۵۷ شرکت بورس تهران طی دوره فروردین ۱۳۹۲ تا آذر ۱۴۰۳ گردآوری و در قالب ۴۲ پنجره زمانی پردازش شد. پس از پاک‌سازی و نرمال‌سازی، چهار الگوریتم طبقه‌بندی شامل درخت تصمیم، جنگل تصادفی، رگرسیون لجستیک و ماشین بردار پشتیبان آموزش داده شدند. عملکرد مدل‌ها با شاخص‌هایی مانند صحت، دقت، بازخوانی و امتیاز F1 سنجیده شد و ماشین بردار پشتیبان با صحت ۶۱٪ و امتیاز F1 برابر ۷۱.۸٪ بهترین عملکرد را داشت. سپس دو استراتژی پرتفوی شامل وزن مساوی و بهینه‌سازی مارکویتز با هدف بیشینه‌سازی نسبت شارپ پیاده‌سازی شدند. نتایج نشان داد پرتفوی مبتنی بر مدل مارکویتز نسبت به پرتفوی وزن مساوی و شاخص بازار از نظر بازده تجمعی (۴۵.۸٪)، نسبت شارپ (۰.۵۲)، و آلفای جنسن عملکرد برتری دارد و در کنترل افت سرمایه و حفظ بتای متعادل نیز موفق بود. نوآوری اصلی این پژوهش در یکپارچه‌سازی الگوریتم‌های یادگیری ماشین با مدل کلاسیک مارکویتز است که موجب ارتقای دقت ورودی‌ها و جلوگیری از ورود سهام با پتانسیل ضعیف به مرحله بهینه‌سازی می‌شود. همچنین، ارزیابی هم‌زمان چند الگوریتم و انتخاب کاراترین مدل برای پیش‌گزینش دارایی‌ها، ارزش افزوده علمی و عملی ایجاد کرده است. این چارچوب به‌عنوان یکی از نخستین نمونه‌ها در بازار سرمایه ایران، می‌تواند الگویی نوین و قابل تعمیم برای سایر بازارهای نوظهور باشد.

کلمات کلیدی

انتخاب سهام، بهینه‌سازی پرتفوی، مدل مارکویتز، الگوریتم ماشین بردار پشتیبان (SVM)، نسبت

شارپ

در سال‌های اخیر، پیشرفت فناوری‌های داده‌پردازی و رشد نمایی داده‌های مالی موجب شده است که روش‌های سنتی تحلیل و تصمیم‌گیری در بازارهای مالی، کارایی گذشته خود را از دست بدهند. نوسانات شدید قیمتی، شوک‌های اقتصادی و ساختار پیچیده بازار، نیاز به مدل‌های هوشمند را افزایش داده است (رحمانی و ابراهیمی، ۱۴۰۳). در بازارهای نوظهور مانند ایران، وجود نوسانات بالا و هزینه‌های اطلاعاتی زیاد، ضعف مدل‌های کلاسیک را بیش از پیش نمایان می‌سازد (کیانی‌زاده و همکاران، ۱۴۰۲). مدل کلاسیک میانگی واریانس مارکویتز^۱ (مارکویتز، ۱۹۵۲)، به‌شدت به صحت ورودی‌هایی چون میانگین بازده و ماتریس کوواریانس وابسته است؛ هرگونه خطا در این ورودی‌ها می‌تواند منجر به انتخاب دارایی‌های نامناسب و کاهش کارایی پرتفوی شود (کریمی و نظری، ۱۴۰۱).

در این میان، الگوریتم‌های یادگیری ماشین با توانایی شناسایی الگوهای پنهان در داده‌های مالی، می‌توانند ابزاری کارآمد برای پیش‌گزینش دارایی‌ها پیش از ورود به مدل مارکویتز باشند (دهقانی و همکاران، ۱۳۹۸؛ هاشمی و همکاران، ۱۴۰۰). بنابراین، مسئله اصلی این پژوهش آن است که آیا استفاده از مدل‌های یادگیری ماشین در مرحله پیش‌گزینش دارایی‌ها می‌تواند کیفیت ورودی‌های مدل مارکویتز را بهبود داده و منجر به ارتقای عملکرد پرتفوی از نظر بازدهی تعدیل‌شده با ریسک گردد؟

الگوریتم‌هایی مانند ماشین بردار پشتیبان (SVM)^۲ قادرند الگوهای پنهان در داده‌های بنیادی و قیمتی را شناسایی کرده و سهامی با عملکرد بالقوه بهتر از شاخص بازار را انتخاب کنند. در محیط بازار سرمایه ایران که با نوسانات شدید، عدم تقارن اطلاعاتی و محدودیت‌های ساختاری روبه‌روست، طراحی چارچوبی هوشمند و داده‌محور می‌تواند تصمیم‌گیری سرمایه‌گذاران را به‌طور معناداری بهبود بخشد. پژوهش‌ها نشان داده‌اند که به‌کارگیری الگوریتم‌های یادگیری ماشین در انتخاب سهام، می‌تواند منجر به تصمیم‌گیری‌های پایدارتر و پرتفوی‌های کاراتر شود (محمدی و جعفری، ۱۴۰۲).

این پژوهش به‌دنبال ارائه‌ی چارچوبی هوشمند برای انتخاب و بهینه‌سازی پرتفوی است که در آن سهام با پتانسیل رشد بالا به‌کمک الگوریتم یادگیری ماشین و داده‌های بنیادی شناسایی شده و سپس با مدل مارکویتز به‌منظور بهینه‌سازی نسبت شارپ بهینه می‌شود.

اهداف فرعی شامل ارزیابی عملکرد الگوریتم یادگیری ماشین در انتخاب سهام، مقایسه بهینه‌سازی پرتفوی با مدل مارکویتز با پرتفوی وزن مساوی، و تحلیل ریسک و پایداری پرتفوی‌ها با شاخص‌هایی مانند نسبت شارپ و آلفای جنسن است.

پرسش اصلی این پژوهش این است که آیا به کارگیری مدل‌های یادگیری ماشین در مرحله‌ی پیش‌گزینش دارایی‌ها می‌تواند عملکرد پرتفوی نهایی را از نظر بازدهی تعدیل‌شده با ریسک بهبود بخشد. بر اساس این پرسش، سه فرضیه مطرح می‌شود: نخست، استفاده از الگوریتم‌های یادگیری ماشین موجب افزایش دقت در انتخاب سهام نسبت به روش‌های سنتی خواهد شد. دوم، ترکیب الگوریتم یادگیری ماشین با مدل مارکویتز می‌تواند نسبت شارپ پرتفوی را در مقایسه با پرتفوی وزن مساوی بهبود دهد. سوم، پرتفوی حاصل از این رویکرد ترکیبی، از پایداری بالاتر و ریسک کمتری نسبت به پرتفوی سنتی برخوردار خواهد بود.

نوآوری این تحقیق در توسعه‌ی چارچوبی ترکیبی و بومی است که از الگوریتم‌های یادگیری ماشین برای پیش‌گزینش دارایی‌ها و مدل مارکویتز برای بهینه‌سازی پرتفوی استفاده می‌کند. این رویکرد، برای نخستین‌بار در بازار سرمایه ایران پیاده‌سازی شده و نشان داده که الگوریتم ماشین بردار پشتیبان توانایی شناسایی سهام برتر با دقت بالاتر نسبت به سایر روش‌ها را دارد. ادغام داده‌های بنیادی با الگوریتم‌های هوشمند، ورودی‌های مدل‌های کلاسیک را تقویت کرده و پرتفوی‌های مقاوم‌تری در برابر نوسانات و ریسک‌ها می‌سازد. این رویکرد قابلیت تعمیم به سایر بازارهای نوظهور و محیط‌های مالی با داده‌های محدود را دارد و به‌عنوان الگویی عملی برای سرمایه‌گذاران و مؤسسات مالی مفید است.

مبانی نظری پژوهش

نظریه نوین پرتفوی برای نخستین بار توسط هری مارکویتز معرفی شد و نقطه عطفی در علم مالی و مدیریت سرمایه‌گذاری محسوب می‌شود. مدل پیشنهادی مارکویتز با هدف دستیابی به ترکیبی بهینه از دارایی‌ها، تلاش می‌کند بازده مورد انتظار پرتفوی را در ازای سطح معینی از ریسک حداکثر کرده یا ریسک را برای یک سطح معین از بازدهی به حداقل برساند (مارکویتز، ۱۹۵۲، ۸۰).

در این مدل، بازده پرتفوی به‌صورت میانگین وزنی بازده دارایی‌ها تعریف می‌شود و ریسک پرتفوی نیز از طریق انحراف معیار بازده پرتفوی یا به‌طور دقیق‌تر، واریانس ترکیبی دارایی‌ها با لحاظ کوواریانس بین آن‌ها محاسبه می‌گردد. فرمول بازده مورد انتظار با رابطه ۱ و واریانس پرتفوی با رابطه ۲ تعریف می‌شود.

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی... /

$$E(R_p) = \sum_{i=1}^n w_i E(R_i) \quad \text{رابطه (۱)}$$

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \quad \text{رابطه (۲)}$$

در این روابط، w_i وزن دارایی i در پرتفوی را نشان می‌دهد، $E(R_i)$ بازده مورد انتظار دارایی i است و σ_{ij} کوواریانس بین بازده‌های دارایی i و j را مشخص می‌کند، که میزان همبستگی و پراکندگی بازده‌ها نسبت به یکدیگر را نشان می‌دهد.

در چارچوب نظریه مارکویتز، بهینه‌سازی پرتفوی به صورت یک مسئله برنامه‌ریزی درجه دوم تعریف می‌شود که هدف آن کمینه‌سازی واریانس پرتفوی با رعایت محدودیت‌هایی مانند جمع وزن دارایی‌ها برابر با یک و دستیابی به سطح مشخصی از بازده است. این روابط در معادلات ۳ و ۴ نشان داده شده‌اند.

$$\text{Min}_w w^T \Sigma w \quad \text{رابطه (۳)}$$

$$\sum_{i=1}^n w_i = 1, \quad w^T \mu = \mu_p \quad \text{رابطه (۴)}$$

در این مدل، w نشان‌دهنده بردار وزن دارایی‌ها در پرتفوی است، Σ ماتریس کوواریانس بازده‌های دارایی‌ها را نشان می‌دهد که میزان همبستگی و پراکندگی بین بازده‌ها را منعکس می‌کند، و μ بردار بازده مورد انتظار هر دارایی را شامل می‌شود. همچنین، μ_p بیانگر بازده هدف پرتفوی است که سرمایه‌گذار قصد دارد با ترکیب بهینه دارایی‌ها به آن دست یابد.

نسبت شارپ که نخستین بار توسط ویلیام شارپ^۲ (۱۹۶۶) معرفی شد، معیاری برای اندازه‌گیری بازده مازاد پرتفوی نسبت به نرخ بازده بدون ریسک، در ازای واحدی از نوسان‌پذیری (انحراف معیار بازده) است (شارپ، ۱۹۶۶، ۱۳۸). در ادبیات مالی معیارهای متعددی برای تعدیل بازده بر اساس ریسک ارائه شده‌اند که مهم‌ترین آن‌ها در جدول ۱ مقایسه شده است.

جدول ۱- مقایسه معیارهای ارزیابی عملکرد پرتفوی

Table 1 – Comparison of Portfolio Performance Evaluation Metrics

ردیف	معیار	نوع ریسک	مزایا	محدودیت‌ها	کاربرد رایج
۱	نسبت شارپ	انحراف معیار کل	ساده و رایج، مناسب برای مقایسه	حساس به نوسانات مثبت و منفی	ارزیابی عملکرد تعدیل‌شده با ریسک کل
۲	نسبت ترینور	بتا (ریسک سیستماتیک)	مناسب در چارچوب CAPM	نیازمند برآورد دقیق بتا	پرتفوی‌های متنوع‌سازی‌شده
۳	نسبت سورتینو	انحراف معیار منفی	تمرکز بر نوسانات منفی	پیچیده‌تر از شارپ، کمتر رایج	استراتژی‌های محافظه‌کارانه
۴	آلفای جنسن	ریسک سیستماتیک	ارزیابی عملکرد مطلق نسبت به بازار	حساس به دقت مدل CAPM	تحلیل بازده اضافی فعال
۵	نسبت کالمار	افت سرمایه ^۴	مناسب برای تحلیل ریسک سقوط شدید	نیازمند داده‌های بلندمدت و پیوسته	استراتژی‌های پرریسک یا پرتلاطم

تحلیل بنیادی^۵ یکی از رویکردهای اصلی در ارزیابی ارزش واقعی اوراق بهادار به شمار می‌رود. در این نوع تحلیل، تمرکز اصلی بر بررسی وضعیت مالی، عملکرد عملیاتی، و توان سودآوری شرکت‌ها است. شاخص‌های بنیادی، مقادیر عددی استخراج‌شده از صورت‌های مالی شرکت‌ها هستند که معیارهایی برای ارزیابی سودآوری، کارایی عملیاتی، نقدینگی، اهرم مالی، و رشد آتی شرکت ارائه می‌دهند.

هدف استفاده از این شاخص‌ها، شناسایی شرکت‌های با پتانسیل رشد پایدار و وضعیت مالی سالم است. برخلاف تحلیل تکنیکال که بر رفتار گذشته قیمت تمرکز دارد، تحلیل بنیادی بر «ارزش ذاتی» و

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی... /

«درون‌مایه اقتصادی» دارایی تأکید دارد. در این پژوهش، تمرکز بر پرکاربردترین شاخص‌های بنیادی است که در جدول ۲ به آن‌ها اشاره شده است.

جدول ۲ - شاخص‌های بنیادی پرکاربرد در ادبیات

Table 2 – Common Fundamental Indicators in the Literature

شاخص	تعریف	کاربرد
P/E ⁶	نسبت قیمت بازار به سود هر سهم	سنجش ارزشگذاری سهم
ROE ⁷	سود خالص / حقوق صاحبان سهام	سنجش توان سودآوری نسبت به سرمایه
ROA ⁸	سود خالص / کل دارایی‌ها	ارزیابی کارایی استفاده از منابع در تولید سود
P/B ⁹	قیمت بازار / ارزش دفتری	مقایسه قیمت سهم با دارایی خالص شرکت
نسبت جاری	دارایی جاری / بدهی جاری	ارزیابی توانایی شرکت در ایفای تعهدات کوتاه‌مدت
حاشیه سود خالص	سود خالص / فروش خالص	سنجش کارایی عملیاتی و کنترل هزینه‌ها

مرور پیشینه پژوهش

یکی از مهم‌ترین حوزه‌های کاربردی یادگیری ماشین در امور مالی، انتخاب و بهینه‌سازی پرتفوی سرمایه‌گذاری است. هدف اصلی در این حوزه، یافتن ترکیب بهینه‌ای از دارایی‌هاست که ضمن دستیابی به بازده مناسب، ریسک کلی سرمایه‌گذاری را به حداقل برساند. در این زمینه، مدل کلاسیک میانگین واریانس مارکویتز (مارکویتز، ۱۹۵۲)، نقش تعیین‌کننده‌ای داشته است. با این حال، فرضیات محدودکننده این مدل، مانند توزیع نرمال بازده‌ها و ثبات ماتریس کوواریانس، سبب شده است تا پژوهشگران به‌سوی مدل‌های ترکیبی و داده‌محور گرایش پیدا کنند (اولیویرا^{۱۰} و همکاران، ۲۰۲۰).

سیلوا^{۱۱} و همکاران (۲۰۲۴) در مطالعه خود با بهره‌گیری از الگوریتم SVM برای پیش‌گزینش سهام و استفاده از مدل مارکویتز برای تخصیص بهینه وزن دارایی‌ها، توانستند پرتفویی طراحی کنند که نسبت شارپ بالاتر، آلفای مثبت و عملکرد تجمعی بهتری نسبت به شاخص بازار و سایر استراتژی‌های پایه

داشت. این مقاله به عنوان یکی از پایه‌های نظری و عملی پژوهش حاضر در نظر گرفته شده است و از پیشنهادات ارائه شده در بخش پایانی آن نیز برای طراحی چارچوب تجربی تحقیق حاضر بهره‌گیری شده است.

در راستای بهره‌گیری از تکنیک‌های نوین تحلیل داده‌ها در حوزه مالی، مطالعات متعددی به بررسی عملکرد مدل‌های یادگیری ماشین در حوزه‌هایی نظیر پیش‌بینی بازار، رتبه‌بندی سهام، تحلیل ریسک، کشف تقلب و بهینه‌سازی پرتفوی پرداخته‌اند. برای نمونه، فیشر و کراوس (۲۰۱۸) با استفاده از شبکه‌های عصبی حافظه بلندمدت ¹²(LSTM) به پیش‌بینی روند شاخص S&P500 پرداخته و گزارش کردند که مدل LSTM توانسته است با دقتی بین ۶۰ تا ۶۵ درصد، عملکرد بهتری نسبت به روش‌های سنتی داشته باشد، امری که بیانگر ظرفیت بالای شبکه‌های عمیق در شناسایی الگوهای زمانی پیچیده در داده‌های مالی است.

در مطالعه‌ای دیگر، لی و همکاران (۲۰۲۱) به منظور رتبه‌بندی و پیش‌گزینش سهام، الگوریتم‌های جنگل تصادفی ¹³(RF)، SVM و شبکه عصبی مصنوعی را بر پایه داده‌های بنیادی شرکت‌ها نظیر نسبت قیمت به درآمد (P/E)، بازده حقوق صاحبان سهام (ROE)، بازده دارایی‌ها (ROA) و نسبت قیمت به ارزش دفتری (P/B) به کار گرفتند. نتایج آن‌ها نشان داد که ترکیب الگوریتم RF با شاخص‌های بنیادی توانسته است دقتی معادل ۷۳٫۱ درصد در شناسایی سهام پربازده در بورس چین ارائه دهد.

در حوزه تحلیل ریسک اعتباری، لسمن^{۱۴} و همکاران (۲۰۱۵) به مقایسه الگوریتم‌های متداول از جمله رگرسیون لجستیک، درخت تصمیم و گرادیان بوستینگ پرداختند. داده‌های ورودی این پژوهش شامل اطلاعاتی از قبیل درآمد مشتری، بدهی‌ها و سابقه بانکی بود. یافته‌ها حاکی از آن بود که مدل گرادیان بوستینگ با دقت ۸۹ درصد عملکرد بسیار قابل قبولی در پیش‌بینی نکول وام‌ها از خود نشان داده است.

یکی از مهم‌ترین حوزه‌های کاربردی در استفاده از روش‌های نوین داده‌محور، انتخاب و بهینه‌سازی پرتفوی سرمایه‌گذاری است. در این زمینه، پژوهش‌های متعددی در داخل کشور انجام شده که به‌طور قابل توجهی بر بهبود عملکرد پرتفوی از طریق تلفیق مدل‌های آماری و یادگیری ماشین تمرکز داشته‌اند. برای نمونه، نوراحمدی و همکاران (۲۰۲۱) با به کارگیری روش خوشه‌بندی سلسله‌مراتبی پرتفوی (HRP)¹⁵ و مقایسه آن با مدل حداقل واریانس مارکویتز در بورس اوراق بهادار تهران نشان دادند که

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی... /

پرتفوی حاصل از HRP نسبت شارپ بالاتری داشته و در عین حال از پایداری بیشتری در برابر نوسانات بازار برخوردار است.

همچنین، در پژوهش باقری مزرعه و همکاران (۲۰۲۲)، ترکیب مرحله‌ی پیش‌گزینش سهام بر پایه الگوریتم‌های یادگیری ماشین از SVM و RF با الگوریتم بهینه‌سازی چندهدفه گرگ خاکستری (MOGWO)¹⁶ به کار گرفته شد. نتایج این مطالعه نشان داد که چارچوب ترکیبی یادشده می‌تواند منجر به افزایش بازده پرتفوی و کاهش ریسک سیستماتیک در مقایسه با مدل‌های سنتی گردد.

جمع‌بندی و شکاف تحقیق

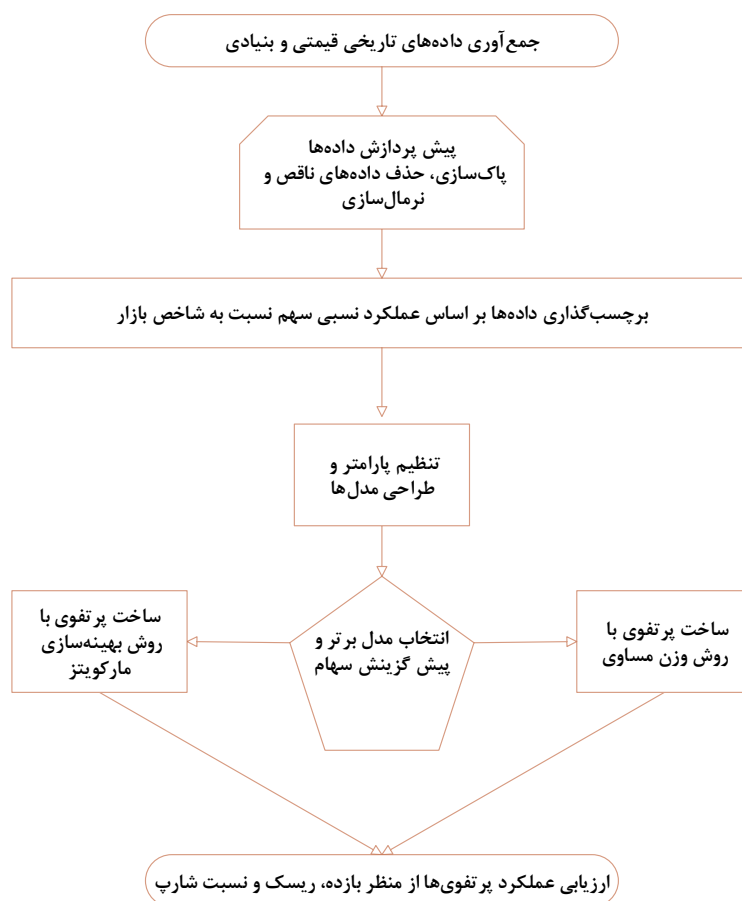
مرور مطالعات داخلی و خارجی نشان می‌دهد که با وجود رشد چشمگیر استفاده از الگوریتم‌های یادگیری ماشین در تحلیل داده‌های مالی، پیش‌بینی بازده سهام و بهینه‌سازی پرتفوی، هنوز شکاف‌های مهمی در این حوزه وجود دارد. نخست، اغلب پژوهش‌های بین‌المللی مانند مطالعات سیلوا و همکاران (۲۰۲۴)، نگوین (۲۰۲۵) و چن و همکاران (۲۰۲۱) بر بازارهای توسعه‌یافته متمرکز بوده و به بازارهای نوظهور همچون بورس تهران کمتر پرداخته‌اند. ویژگی‌هایی مانند نوسان بالا، عدم تقارن اطلاعاتی و رفتار غیرخطی سرمایه‌گذاران در این بازارها، تعمیم نتایج آن پژوهش‌ها را دشوار می‌سازد.

دوم، بیشتر تحقیقات بر پیش‌بینی بازده یا رتبه‌بندی سهام تمرکز داشته‌اند و مرحله‌ی پیش‌گزینش سهام مبتنی بر یادگیری ماشین را با مدل‌های بهینه‌سازی پرتفوی مانند مدل مارکویتز به‌صورت یکپارچه ترکیب نکرده‌اند. مطالعاتی نظیر فیشر و کراوس (۲۰۱۸) یا لی و همکاران (۲۰۲۱) صرفاً به ارزیابی عملکرد الگوریتم‌های پیش‌بینی پرداخته و تأثیر آن‌ها بر بهینه‌سازی پرتفوی را تجربی بررسی نکرده‌اند. سوم، بسیاری از مدل‌های موجود فاقد توضیح‌پذیری کافی هستند. در بازارهای مالی حساس، علاوه بر دقت، قابلیت تفسیر نتایج و شناسایی عوامل مؤثر بر انتخاب دارایی ضروری است. این جنبه در بیشتر پژوهش‌های داخلی و حتی برخی مطالعات خارجی نادیده گرفته شده است.

در مجموع، پیشینه پژوهش‌ها نشان می‌دهد که تاکنون مدلی ترکیبی، داده‌محور، بومی و توضیح‌پذیر که در آن از الگوریتم‌های یادگیری ماشین برای پیش‌گزینش سهام و از مدل‌های بهینه‌سازی چندهدفه برای تخصیص دارایی‌ها در بازار سرمایه ایران استفاده شود، ارائه نشده است. از این رو، پژوهش حاضر با طراحی چارچوبی ترکیبی مبتنی بر یادگیری ماشین و مدل میانگین-واریانس مارکویتز، و با بهره‌گیری از داده‌های بنیادی و تاریخی شرکت‌های بورسی، درصدد پر کردن این شکاف علمی و کاربردی است.

روش‌شناسی پژوهش

رویکرد پژوهش مبتنی بر تحلیل‌های کمی و عددی است. داده‌های قیمتی و بنیادی از منابع رسمی نظیر سامانه کدال و TSETMC استخراج شده‌اند. پس از گردآوری، این داده‌ها مورد پیش‌پردازش قرار گرفته و در چارچوب الگوریتم‌های طبقه‌بندی یادگیری ماشین و مدل‌های بهینه‌سازی کلاسیک، تحلیل شده‌اند. در انتها، عملکرد مدل ترکیبی پیشنهادی در مقایسه با رویکردهای مرجع و شاخص بازار مورد ارزیابی تجربی قرار گرفته است. فلوچارت مراحل در شکل ۱ قابل نمایش است.



شکل ۱ - فلوچارت مراحل روش‌شناسی پژوهش

Figure 1 – Flowchart of the Research Methodology Steps

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی... /

جامعه آماری این پژوهش شامل تمامی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران است که طی بازه زمانی فروردین ۱۳۹۲ تا آذر ۱۴۰۳ به‌طور مستمر در ترکیب شاخص کل حضور داشته‌اند. شاخص کل به‌عنوان مبنای ارزیابی انتخاب شده است؛ زیرا با پوشش گسترده صنایع مختلف، نمایانگر عملکرد کلی بازار و معیاری مناسب برای مقایسه بازده نسبی سهام محسوب می‌شود.

از میان شرکت‌های فعال در این دوره، تنها شرکت‌هایی انتخاب شدند که داده‌های قیمتی روزانه برای محاسبه بازده و صورت‌های مالی فصلی کامل برای استخراج شاخص‌های بنیادی در دسترس داشتند. پس از پالایش و حذف داده‌های ناقص، در نهایت ۱۵۷ شرکت به‌عنوان نمونه آماری نهایی برای تحلیل انتخاب شدند.

در این پژوهش، مرحله پیش‌گزینش سهام با هدف شناسایی کاراتر سهام پربازده پیش از بهینه‌سازی پرتفوی طراحی شده است. برای این منظور، داده‌های بنیادی شرکت‌ها با استفاده از چهار الگوریتم یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی، رگرسیون لجستیک و ماشین بردار پشتیبان تحلیل و مقایسه شدند تا مدلی با بالاترین دقت در شناسایی سهام برتر انتخاب شود.

داده‌ها به‌صورت فصلی و بر اساس توالی زمانی تقسیم‌بندی شدند تا از نشت اطلاعات جلوگیری شود و پارامترهای مدل‌ها با جستجوی شبکه‌ای و اعتبارسنجی متقاطع پنج مرحله‌ای بهینه گردیدند. در نهایت، مدلی که بالاترین دقت، تعمیم‌پذیری و توان مدل‌سازی روابط غیرخطی را داشت، به‌عنوان مدل نهایی انتخاب شد. این مرحله به‌عنوان یک لایه تحلیلی هوشمند عمل کرده و با حذف سهام کم‌پتانسیل، دقت و کارایی فرآیند تشکیل پرتفوی را به‌طور محسوسی افزایش داده است.

پس از اتمام مرحله پیش‌گزینش سهام، مجموعه‌ای از سهام مشخص شد که از نظر عملکرد پیش‌بینی‌شده در فصل آینده، دارای پتانسیل بازدهی بالاتری نسبت به شاخص بازار بودند. در ادامه، با هدف ارزیابی کارایی مدل در ساخت پرتفوی‌های سرمایه‌گذاری، دو رویکرد متفاوت برای تخصیص وزن به این دارایی‌ها مورد استفاده قرار گرفت.

الف) پرتفوی با وزن مساوی: در این استراتژی پایه، به تمامی سهام منتخب توسط SVM وزن مساوی اختصاص داده شد. در این صورت، اگر N سهم توسط مدل به‌عنوان برتر انتخاب شده باشد، وزن هر سهم به‌صورت رابطه ۵ تعریف می‌شود. در این پژوهش این پرتفوی تحت عنوان SVM-1/n شناخته می‌شود.

$$w_i = \frac{1}{N}, i = 1, 2, \dots, N \quad \text{(رابطه ۵)}$$

ب) پرتفوی بهینه‌شده با مدل مارکوویتز: این مدل بر اساس مفروضات سرمایه‌گذار به‌دنبال بیشینه‌سازی بازده با حداقل ریسک است. بازده دارایی‌ها قابل پیش‌بینی از طریق میانگین تاریخی است. ریسک پرتفوی از طریق واریانس و کوواریانس بین بازدهی دارایی‌ها اندازه‌گیری می‌شود. در این بخش، از مدل میانگین واریانس مارکوویتز با هدف بیشینه‌سازی نسبت شارپ^{۱۷} استفاده شده است. مدل ریاضی این بهینه‌سازی به صورت رابطه ۶ و قیود آن به صورت رابطه ۷ تعریف می‌شوند. که در آن μ بردار بازده مورد انتظار دارایی‌ها، Σ ماتریس کوواریانس بازدهی، R_f نرخ بازده بدون ریسک، و W بردار وزن‌های پرتفوی است.

$$\max_w \left(\frac{w^T \mu - R_f}{\sqrt{w^T \Sigma w}} \right) \quad \text{رابطه ۶}$$

$$\sum w = 1, \quad w \geq 0 \quad \text{رابطه ۷}$$

جهت ارزیابی قابلیت تعمیم‌پذیری و پایداری مدل‌ها، از تکنیک اعتبارسنجی متقابل چندبخشی استفاده می‌شود. معیارهای مختلفی برای ارزیابی عملکرد مدل‌ها تعریف می‌شوند. جدول ۳ روابط ریاضی و تعریف این معیارها را ارائه می‌دهد.

جدول ۳ - معیارهای ارزیابی مورد استفاده جهت سنجش عملکرد مدل‌های یادگیری ماشین

Table 3 – Evaluation Metrics Used to Assess the Performance of Machine Learning Models

فرمول ریاضی	تعریف	ردیف
$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$	صحت ^{۱۸}	رابطه ۸
$\text{Sensitivity (TPR)} = \frac{TP}{TP + FN}$	حساسیت ^{۱۹}	رابطه ۹
$\text{Precision} = \frac{TP}{TP + FP}$	دقت ^{۲۰}	رابطه ۱۰
$\text{f1_Score} = \frac{2 * \text{precision} * \text{Sensitivity}}{\text{precision} + \text{Sensitivity}}$	امتیاز F1	رابطه ۱۱

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی... /

برای مقایسه عملکرد پرتفوی‌ها با یکدیگر و با شاخص بازار، از شاخص‌های مالی استاندارد استفاده شده است. بازده تجمعی کل بازده دوره را نشان می‌دهد و بازده سالانه نرخ رشد متوسط سالانه را محاسبه می‌کند. نسبت شارپ بازده اضافی نسبت به کل ریسک و نسبت سورتینو بازده نسبت به ریسک منفی را می‌سنجد. آلفای جنسن بیانگر توان مدیریت فعال بر اساس مدل CAPM و بتا نشان‌دهنده حساسیت پرتفوی نسبت به نوسانات بازار است. همچنین، حداکثر افت سرمایه میزان بیشترین کاهش ارزش پرتفوی را اندازه‌گیری کرده و شاخصی کلیدی برای ارزیابی ریسک نزولی است.

یافته‌های پژوهش

برای افزایش دقت و جلوگیری از بیش‌برازش، تنظیم پارامترهای الگوریتم‌های یادگیری ماشین با روش جستجوی شبکه‌ای و اعتبارسنجی متقاطع پنج‌بخشی انجام شد. این فرآیند تنها بر روی داده‌های آموزشی (۸۰٪ ابتدایی داده‌ها) اجرا گردید تا از نشت اطلاعات جلوگیری شود. سپس عملکرد الگوریتم‌های تنظیم‌شده بر روی داده‌های آزمایشی (۲۰٪ پایانی داده‌ها) ارزیابی گردید تا دقیق‌ترین و پایدارترین مدل برای پیش‌گزینش سهام با بازدهی بالاتر از شاخص بازار شناسایی شود. نتایج مقایسه مدل‌های گزارش‌شده در جدول ۴ مبنای انتخاب مدل نهایی پژوهش قرار گرفتند.

جدول ۴ - مقایسه عملکرد الگوریتم‌های طبقه‌بندی در پیش‌گزینش سهام

Table 4 – Comparison of Classification Algorithms' Performance in Stock Pre-selection

الگوریتم طبقه‌بندی	صحت	دقت	بازخوانی	امتیاز F1
درخت تصمیم	۵۸٪	۴۹٪	۸۷٪	۶۲٪
جنگل تصادفی	۶۰٪	۵۳٪	۹۲٪	۶۷٪
رگرسیون لجستیک	۵۹٪	۵۱٪	۸۹٪	۶۵٪
ماشین بردار پشتیبان	۶۱٪	۵۶٪	۱۰۰٪	۷۱.۸٪

نتایج جدول فوق بیانگر آن است که الگوریتم SVM در تمامی معیارهای اصلی، به‌ویژه در بازخوانی و امتیاز F1، عملکرد بهتری نسبت به سایر الگوریتم‌ها داشته است. بازخوانی ۱۰۰٪ به این معناست که

هیچ یک از سهام با بازدهی بالاتر از شاخص بازار در فصل آزمون از دید مدل پنهان نمانده است؛ این ویژگی برای اهداف این پژوهش، که تمرکز اصلی آن بر شناسایی فرصت‌های سودآور سرمایه‌گذاری است، از اهمیت بسیار بالایی برخوردار است.

هرچند دقت در مدل SVM برابر با ۵۶٪ بوده و نشان می‌دهد برخی پیش‌بینی‌ها شامل سهام کمتر سودآور نیز بوده است، اما این امر در چارچوب استراتژی‌های سرمایه‌گذاری قابل‌قبول و حتی مطلوب ارزیابی می‌شود. با توجه به تحلیل‌های فوق، ماشین بردار پشتیبان به‌عنوان مدل نهایی برای ادامه فرآیند پیش‌گزینش سهام انتخاب گردید و به‌عنوان مبنای تشکیل پرتفوی‌های سرمایه‌گذاری در مراحل بعدی استفاده شده است.

در مجموع، ۶۹ نماد بورسی توسط مدل به‌عنوان سهام با پتانسیل عملکرد بهتر از شاخص کل در فصل پاییز ۱۴۰۳ پیش‌بینی شدند. این سهام از صنایع متنوعی انتخاب شده‌اند که نشان‌دهنده کارکرد متعادل مدل SVM در پوشش بخش‌های مختلف بازار سرمایه است. این انتخاب، مبنای تشکیل دو پرتفوی سرمایه‌گذاری با رویکردهای متفاوت قرار گرفته است که در ادامه شرح داده می‌شوند. نمونه‌ای از نمادهای منتخب در جدول ۵ آمده است.

جدول ۵ - سهام منتخب در مرحله پیش‌گزینش سهام

Table 5– Selected Stocks in the Stock Pre-selection Stage

ردیف	نماد	سال	فصل
۱	اخابر	۱۴۰۳	Q3
۲	مارون	۱۴۰۳	Q3
۳	کگل	۱۴۰۳	Q3
...	...	...	...
۶۹	شپاکسا	۱۴۰۳	Q3

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی... /

پس از اجرای مدل ماشین بردار پشتیبان و ارزیابی عملکرد آن در طبقه‌بندی سهام، فرآیند پیش‌گزینش نهایی در فصل سوم سال ۱۴۰۳ انجام شد. در این مرحله، ۶۹ نماد بورسی به‌عنوان سهام با پتانسیل عملکرد بهتر از شاخص کل بازار شناسایی شدند که از صنایع مختلف انتخاب شده‌اند.

دو نوع پرتفوی برای ارزیابی اثربخشی پیش‌گزینش و بهینه‌سازی طراحی شد که نتایج عملکرد این دو استراتژی در جدول ۶ ارائه شده‌اند تا با روش‌های مرجع مقایسه شوند.

۱. پرتفوی وزن مساوی (SVM-1/n): در این رویکرد ساده، تمامی ۶۹ سهم منتخب وزن یکسان (۱/۶۹) داشتند و به‌عنوان مرجع پایه برای ارزیابی اثر مدل SVM استفاده شد.

۲. پرتفوی بهینه‌شده با مدل مارکویتز (SVM-MaxSharpe): سهام منتخب، وارد فرآیند بهینه‌سازی مارکویتز شدند تا ترکیب وزنی بهینه برای حداکثر کردن نسبت شارپ تعیین شود.

جدول ۶ - مقایسه عملکرد سه استراتژی سرمایه‌گذاری در دوره آزمون

Table 6 – Comparison of the Performance of Three Investment Strategies During the Test Period

SVM-MaxSharpe	SVM-1/n	شاخص بازار	شاخص عملکرد
<u>۲۶.۹۴٪</u>	۴.۸۷٪	۳.۱۲٪ -	بازده سالانه
<u>۴۵.۸۰٪</u>	۳.۶۵٪	۶.۰۲٪ -	بازده تجمعی
<u>۰.۵۲</u>	۰.۰۷	۰.۴۲ -	نسبت شارپ
<u>۰.۶۸</u>	۰.۱۱	۰.۵۵ -	نسبت سورتینو
<u>۰.۲۹</u>	۰.۰۴	—	آلفای جنسن
<u>۱.۲۷</u>	۱.۰۲	۱.۰۰	بتا
<u>۱۳.۴٪</u>	۱۸.۷٪	۲۲.۱٪	حداکثر افت سرمایه

نتایج تحلیل داده‌ها و مقایسه عملکرد پرتفوی‌ها نشان می‌دهد که استراتژی پیشنهادی در تمامی شاخص‌های کلیدی عملکرد، برتری معناداری نسبت به شاخص بازار و پرتفوی وزن مساوی دارد. این پرتفوی با بازده سالانه ۲۶.۹۴ درصد و بازده تجمعی ۴۵.۸ درصد توانسته است بازدهی پایدار و پیوسته‌ای در طول دوره ارزیابی ایجاد کند. همچنین، نسبت‌های شارپ (۰.۵۲) و سورتینو (۰.۶۸) نشان‌دهنده آن است که پرتفوی بهینه نه تنها بازده بالایی دارد، بلکه در تولید بازده تعدیل‌شده با ریسک نیز عملکرد مطلوبی از خود نشان داده است. نسبت سورتینو که تمرکز ویژه‌ای بر ریسک‌های منفی دارد، نشان‌دهنده کنترل مناسب نوسانات نامطلوب است.

علاوه بر این، آلفای مثبت جنسن (۰.۲۹) بیانگر ایجاد بازدهی مازاد نسبت به بازده مورد انتظار بر اساس مدل CAPM است که نشان‌دهنده خلق ارزش افزوده واقعی توسط استراتژی ترکیبی است. هرچند بتای پرتفوی (۱.۲۷) وابستگی بالاتر به نوسانات بازار را نشان می‌دهد، اما این ریسک سیستماتیک به خوبی مدیریت شده و به بازدهی پایدار تبدیل گردیده است.

در مقایسه، پرتفوی SVM-1/n که صرفاً بر اساس انتخاب اولیه سهام توسط الگوریتم SVM و بدون مرحله بهینه‌سازی تشکیل شده است، عملکرد ضعیف‌تری نشان داد. اگرچه بازده سالانه آن (۴.۸۷ درصد) اندکی بالاتر از شاخص بازار بود، اما نسبت شارپ (۰.۰۷) و سایر شاخص‌های تعدیل‌شده با ریسک در سطح پایینی قرار داشتند که بیانگر ضعف در مدیریت بهینه ریسک است.

از منظر پایداری عملکرد و کنترل ریسک نیز پرتفوی SVM-MaxSharpe عملکرد بهتری داشته است. حداکثر افت سرمایه این پرتفوی ۱۳.۴ درصد گزارش شد، در حالی که شاخص بازار و پرتفوی وزن مساوی به ترتیب افت‌هایی معادل ۲۲.۱ و ۱۸.۷ درصد را تجربه کردند. این نتایج نشان‌دهنده مقاومت بالاتر پرتفوی بهینه‌شده در برابر نوسانات بازار و کاهش زیان در دوره‌های نزولی است. این ویژگی از منظر مالی رفتاری نیز اهمیت دارد، زیرا کاهش نوسانات و افت سرمایه موجب افزایش اعتماد و پایداری تصمیم‌گیری سرمایه‌گذار در افق‌های زمانی بلندمدت می‌شود.

نتیجه‌گیری و بحث

پژوهش حاضر با هدف طراحی، پیاده‌سازی و ارزیابی یک مدل ترکیبی برای انتخاب و بهینه‌سازی پرتفوی سهام در بازار سرمایه ایران انجام شد. دغدغه اصلی تحقیق، ارائه چارچوبی علمی و عملیاتی برای بهبود فرآیند تصمیم‌گیری سرمایه‌گذاران در محیطی با نوسانات بالا، اطلاعات نامتقارن و رفتارهای غیرمنطقی بود. رویکرد پیشنهادی از دو بخش مکمل تشکیل شده است:

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی... /

۱. پیش‌گزینش سهام به کمک یادگیری ماشین: در این بخش، با تکیه بر داده‌های بنیادی فصلی شرکت‌ها (شامل نسبت‌های مالی و شاخص‌های عملکرد)، مدل‌های مختلف یادگیری ماشین برای طبقه‌بندی سهام به دو دسته "با عملکرد بهتر از بازار" و "کمتر از بازار" توسعه داده شد. خروجی این مرحله، انتخاب مدل ماشین بردار پشتیبان به عنوان مدل برتر برای گزینش لیستی از نمادهای بورسی با احتمال بالای بازدهی مثبت بود.
۲. بهینه‌سازی وزن سهام منتخب با مدل میانگین واریانس مارکویتز: سهام شناسایی شده توسط مدل SVM وارد فرآیند تخصیص وزنی شدند، به‌طوری که با استفاده از تابع هدف بهینه‌سازی نسبت شارپ، ترکیبی از دارایی‌ها حاصل شود که بازده تعدیل‌شده با ریسک را به حداکثر برساند.

برای اجرای مدل، داده‌های تحقیق شامل بازده روزانه و صورت‌های مالی فصلی ۱۵۷ شرکت پذیرفته‌شده در بورس اوراق بهادار تهران طی بازه زمانی فروردین ۱۳۹۲ تا آذر ۱۴۰۳ گردآوری شد. پس از پردازش و نرمال‌سازی داده‌ها، ۴۲ پنجره زمانی فصلی ایجاد و مدل بر اساس آن‌ها آموزش داده شد. در نهایت، عملکرد دو پرتفوی طراحی شده در آخرین پنجره زمانی با شاخص کل بازار مقایسه شد و با استفاده از شاخص‌های مالی نظیر بازده سالانه، نسبت شارپ، نسبت سورتینو، آلفای جنسن، و حداکثر افت سرمایه ارزیابی گردید. نتایج نشان داد که مدل ترکیبی پیشنهادی، نسبت به شاخص بازار و پرتفوی وزن مساوی عملکرد بهتری داشته و می‌تواند در تصمیم‌گیری‌های سرمایه‌گذاری واقعی نیز به‌کار گرفته شود. مجموع نتایج به‌دست‌آمده بیانگر آن است که چارچوب ترکیبی ارائه‌شده در این پژوهش، علاوه بر پشتیبانی نظری از ادبیات علمی، در عمل نیز توانسته کارایی مناسبی در محیط واقعی بازار سرمایه ایران ارائه دهد. استفاده هم‌زمان از ابزارهای یادگیری ماشین برای انتخاب دارایی و مدل‌های کلاسیک مالی برای بهینه‌سازی، رویکردی کارآمد برای ارتقاء تصمیم‌گیری سرمایه‌گذاری فراهم آورده است.

پیشنهادهای پژوهش‌های آینده

با توجه به نتایج حاصل از این تحقیق و همچنین محدودیت‌هایی که در فرآیند اجرا وجود داشت، چندین مسیر پیشنهادی برای ادامه یا گسترش پژوهش در این حوزه قابل طرح است. این پیشنهادات می‌توانند دقت، کارایی، و تعمیم‌پذیری مدل ترکیبی ارائه‌شده را افزایش دهند.

۱. استفاده از الگوریتم‌های یادگیری عمیق و مدل‌های پیشرفته‌تر: در این پژوهش، پس از آزمون اولیه چند الگوریتم رایج یادگیری ماشین نظیر درخت تصمیم، جنگل تصادفی و رگرسیون لجستیک، در نهایت

الگوریتم ماشین بردار پشتیبان به عنوان مدل منتخب برای مرحله پیش‌گزینش سهام انتخاب شد. با وجود نتایج رضایت‌بخش این مدل، پیشنهاد می‌شود در پژوهش‌های آتی از طیف گسترده‌تری از مدل‌های پیشرفته‌تر یادگیری ماشین و یادگیری عمیق نیز استفاده شود. به عنوان نمونه، مدل‌هایی نظیر شبکه‌های عصبی مصنوعی می‌توانند برای کشف الگوهای غیرخطی پیچیده در داده‌های بنیادی و رفتاری سهام مفید باشند.

۲. در نظر گرفتن هزینه‌های تراکنش و محدودیت‌های عملیاتی بازار: مدل طراحی شده در این تحقیق از دیدگاه تئوریک و بدون در نظر گرفتن هزینه‌های معاملاتی (مانند کارمزد، مالیات، اسپرد خرید و فروش) و محدودیت‌هایی نظیر نقدشوندگی سهام‌ها حداکثر درصد مجاز هر سهم در سبد محدودیت‌های معاملاتی نهادهای مالی یا صندوق‌ها ساخته شده است. پیشنهاد می‌شود در پژوهش‌های آینده، این قیود واقعی وارد مدل شوند تا پرتفوی نهایی از نظر عملیاتی نیز قابل پیاده‌سازی باشد.

محدودیت‌های تحقیق

با وجود تلاش فراوان برای طراحی دقیق مدل و استفاده از داده‌های قابل اطمینان، این پژوهش با چند محدودیت اساسی همراه بوده است. نخست، محدودیت در حجم داده‌های آموزشی و آزمون به دلیل محدودیت تاریخی اطلاعات مالی شرکت‌ها و حذف داده‌های ناقص، موجب کاهش اندازه نمونه و در نتیجه، محدود شدن توان تعمیم مدل‌های یادگیری ماشین شده است. دوم، ترکیب متغیرهای ورودی به طور کامل متنوع نبوده و هرچند از شاخص‌های بنیادی معتبر استفاده شده است، اما در نظر نگرفتن سایر عوامل مالی و اقتصادی، به ویژه متغیرهای کلان همچون نرخ بهره، تورم و نرخ ارز، می‌تواند موجب کاهش دقت پیش‌بینی مدل شده باشد. سوم، نتایج حاصل تا حدی به تنظیمات اولیه مدل‌های یادگیری وابسته است؛ به این معنا که انتخاب پارامترها، ساختار داده‌ها و نحوه پیش‌پردازش می‌تواند بر خروجی نهایی تأثیر بگذارد. اگرچه از روش اعتبارسنجی متقاطع برای کاهش سوگیری استفاده شده است، اما امکان بروز اثرات ناشی از تنظیمات اولیه همچنان وجود دارد. چهارم، مدل پیشنهادی تنها در شرایط نسبتاً پایدار اقتصادی آموزش دیده و عملکرد آن در مواجهه با شوک‌های اقتصادی یا تغییرات ساختاری بازار مورد بررسی قرار نگرفته است. در نهایت، ارزیابی عملکرد مدل صرفاً در یک الگوی بازار خاص انجام شده و آزمون آن در سناریوهای متفاوت همچون بازارهای صعودی، نزولی یا رکودی می‌تواند دیدگاه جامع‌تری از پایداری و کارایی مدل ارائه دهد.

- 1) Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 36(3), 5932–5941. <https://doi.org/10.1016/j.eswa.2008.05.027>.
- 2) Bagheri Mazraeh, M., Hosseini, S., & Ahmadi, N. (2022). Combining machine learning algorithms and multi-objective grey wolf optimizer in optimal portfolio selection in Tehran Stock Exchange. *Financial Research Quarterly*, 24(2), 75–96. [In Persian] <https://doi.org/10.22067/frq.v24i2.94512>.
- 3) Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments* (10th ed.). New York: McGraw-Hill Education.
- 4) Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>.
- 5) Dehghani, A., Fazeli, M., & Rahimi, H. (2019). Stock market index prediction using support vector machine algorithm. *Economic Research*, 18(4), 121–144. [In Persian] <https://doi.org/10.22059/jed.2019.278905.6542>.
- 6) Dutta, A., & Das, D. (2021). Machine learning in finance: A comprehensive survey. *Journal of Economic*, 35(1), 186–234. <https://doi.org/10.1111/joes.12389>.
- 7) Elton, E. J., & Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of Banking & Finance*, 21(11), 1743–1759. [https://doi.org/10.1016/S0378-4266\(97\)00025-8](https://doi.org/10.1016/S0378-4266(97)00025-8).
- 8) Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, 18(3), 25–4. <https://doi.org/10.1257/0895330042162430>.
- 9) Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10), 2513–2522. <https://doi.org/10.1016/j.cor.2004.03.016>.
- 10) Islamic Azad University of Esfarayen. (2024). Stock price prediction using LSTM model and technical indicators [Master's thesis, Islamic Azad University of Esfarayen, Iran]. [In Persian]. <https://thesis.iau-esfarayen.ac.ir>.
- 11) Kogan, S., Levin, D., Routledge, B. R., Sagi, J. S., & Smith, N. A. (2009). Predicting risk and return of equity portfolios using text. *Proceedings of the NAACL*. <https://aclanthology.org/N09-1008>.
- 12) Lee, T. S., & Choi, I. (2013). A multi-level modeling approach to predict stock returns. *Expert Systems with Applications*, 40(11), 4583–4590. <https://doi.org/10.1016/j.eswa.2013.01.010>.

- 13) Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking classification algorithms for credit scoring. *European Journal of Operational Research*, 247(1), 124–136. <https://doi.org/10.1016/j.ejor.2015.05.030>.
- 14) Li, X., Zhang, Y., Wang, S., & Li, L. (2021). Empirical analysis of stock price prediction via machine learning. *Soft Computing*, 25, 481–495. <https://doi.org/10.1007/s00500-020-05242-1>.
- 15) Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525>.
- 16) Norahmadi, F., Ghanbari, S., & Karimi, R. (2021). A comparison of Hierarchical Risk Parity method in Tehran Stock Exchange. *Journal of Financial Management*, 6(3), 33–54. [In Persian] <https://doi.org/10.22034/jfmis.2021.332422.1194>.
- 17) Oliveira, L. B., Silva, T. S., Lima, A. C., & Almeida, J. M. (2020). A hybrid model for portfolio optimization. *Expert Systems with Applications*, 158, 113509. <https://doi.org/10.1016/j.eswa.2020.113509>.
- 18) Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(4), 2162–2172. <https://doi.org/10.1016/j.eswa.2014.09.043>.
- 19) Sayyadi, S., & Omid, M. (2019). Predicting industrial index of stock exchange using artificial neural networks. *Journal of Financial and Economic Research*, 13(1), 45–66. [In Persian] <https://doi.org/10.22067/jfer.v13i1.83624>.
- 20) Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1), 119–138. <https://doi.org/10.1086/294846>.
- 21) Silva, N. F., de Andrade, L. P., da Silva, W. S., de Melo, M. K., & Tonelli, A. O. (2024). Portfolio optimization based on the pre-selection of stocks. *Finance Research Letters*, 61, 105014. <https://doi.org/10.1016/j.frl.2024.105014>.

مدل ترکیبی SVM-مارکویتز برای انتخاب سهام و تشکیل پرتفوی... /

یادداشت‌ها

-
- ¹ Harry Markowitz
 - ² Support vector machines
 - ³ Sharpe
 - ⁴ Drawdown
 - ⁵ Fundamental analysis
 - ⁶ Price to earnings
 - ⁷ Return on equity
 - ⁸ Return on assets
 - ⁹ Price to book value
 - ¹⁰ Oliveira
 - ¹¹ Silva
 - ¹² Long Short-Term Memory
 - ¹³ Random forest
 - ¹⁴ Lessmann
 - ¹⁵ Hierarchical risk parity
 - ¹⁶ Multi-objective grey wolf optimizer
 - ¹⁷ Sharpe ratio
 - ¹⁸ Accuracy
 - ¹⁹ Sensitivity
 - ²⁰ Precision

A Hybrid SVM–Markowitz Model for Stock Selection and Optimal Portfolio Formation Based on Fundamental Indicators

Abstract

In the volatile and unstable conditions of financial markets, optimizing investment decisions—particularly in emerging markets—has become increasingly vital. The purpose of this study is to design and evaluate a hybrid framework for stock selection and portfolio allocation in the Iranian capital market, integrating machine learning algorithms for asset pre-selection and the Markowitz model for portfolio optimization. Fundamental and price data of 157 companies listed on the Tehran Stock Exchange during the period from April 2013 to December 2024 were collected and processed in 42 rolling time windows. After data cleaning and normalization, four classification algorithms—Decision Tree, Random Forest, Logistic Regression, and Support Vector Machine (SVM)—were trained. The models' performance was evaluated using accuracy, precision, recall, and F1-score metrics. The SVM model achieved the best results, with an accuracy of 61% and an F1-score of 71.8%. Subsequently, two portfolio strategies, equal-weight and Markowitz optimization (aimed at maximizing the Sharpe ratio), were implemented. The results indicated that the Markowitz-based portfolio outperformed both the equal-weighted portfolio and the market index in terms of cumulative return (45.8%), Sharpe ratio (0.52), and Jensen's alpha. Moreover, it demonstrated superior performance in drawdown control and maintaining a balanced beta. The main innovation of this research lies in integrating machine learning algorithms with the classical Markowitz model, thereby improving input precision and preventing the inclusion of low-potential stocks in the optimization process. Additionally, the simultaneous evaluation of multiple algorithms and the selection of the most efficient model for asset pre-screening add both scientific and practical value. As one of the pioneering studies in the Iranian capital market, this hybrid framework provides a novel and adaptable model applicable to other emerging markets.

Keywords

Stock selection, Portfolio optimization, Markowitz model, Support Vector Machine, Sharpe ratio

Introduction

In recent years, advancements in data processing technologies and the exponential growth of financial data have made traditional methods of analysis and decision-making in financial markets less effective. The high price volatility, economic shocks, and complex market structures have highlighted the increasing need for intelligent, data-driven models. In emerging markets such as Iran, the lack of transparency, high volatility, and significant informational costs exacerbate the limitations of classical models. The classical mean-variance model of Markowitz (1952), though foundational in portfolio theory, is highly dependent on the accuracy of inputs like expected returns and covariance matrices. Any error in these inputs can result in suboptimal asset selections and reduced portfolio efficiency.

This study explores the integration of machine learning algorithms, specifically Support Vector Machines (SVM), into the asset pre-selection phase, before applying the Markowitz model for portfolio optimization. The research investigates whether machine learning can improve the quality of inputs for the Markowitz model and subsequently enhance portfolio performance in terms of risk-adjusted returns.

Methods and Tools

The proposed approach in this study utilizes machine learning algorithms for asset pre-selection based on fundamental financial data. The machine learning model is trained using features such as Price-to-Earnings (P/E) ratio, Return on Equity (ROE), and other key financial indicators to classify stocks with high growth potential. After selecting the top-performing stocks, the Markowitz model is applied for portfolio optimization, focusing on maximizing the Sharpe ratio. The performance of the portfolios is evaluated by comparing risk-adjusted returns to a benchmark portfolio (equal-weighted portfolio).

The methodology involves the following key steps:

1. **Data Collection:** Gathering fundamental financial data from publicly available sources such as financial statements of listed companies.
2. **Preprocessing:** Normalizing and cleaning the data to ensure its usability for machine learning algorithms.
3. **Model Training:** Using SVM to classify stocks based on their potential for high returns and low risk.
4. **Portfolio Construction:** Optimizing the selected assets using the Markowitz model, with a focus on maximizing the Sharpe ratio.
5. **Performance Evaluation:** Comparing the optimized portfolio's performance with a baseline equal-weighted portfolio using key financial metrics such as the Sharpe ratio, Jensen's alpha, beta, and maximum drawdown.

Results

The results of the study show that the machine learning model significantly outperforms traditional methods of stock selection in terms of prediction accuracy and classification precision. The SVM algorithm, when trained on fundamental data, is able to identify stocks that exhibit better potential for high returns compared to the market index. When the selected stocks are optimized using the Markowitz model, the resulting portfolio outperforms the equal-weighted portfolio with higher risk-adjusted returns, as measured by the Sharpe ratio. Furthermore, the optimized portfolio exhibits greater stability and lower risk, as indicated by lower values of maximum drawdown and a more favorable Jensen's alpha.

Discussion and Analysis

The findings indicate that integrating machine learning algorithms, particularly SVM, into the asset pre-selection phase enhances the accuracy of stock selection and leads to better portfolio optimization. The Markowitz model, which traditionally relies heavily on precise input data, benefits from the improved quality of inputs generated by the machine learning model. This hybrid approach not only increases the overall return of the portfolio but also provides better risk management by selecting stocks with lower volatility and higher stability.

Moreover, the study confirms that the use of fundamental data in machine learning models improves the robustness of portfolio optimization, particularly in emerging markets like Iran. Given the high volatility and limited information in such markets, this approach offers a practical and effective solution for investors looking to enhance their decision-making processes.

The implications of these results suggest that machine learning algorithms can be a valuable tool for pre-selecting assets in portfolio optimization. The approach developed in this study is scalable and can be applied to other emerging markets facing similar challenges, providing a data-driven and intelligent framework for investment management.

Conclusion

This research contributes to the field of intelligent investment management by introducing a novel hybrid approach that combines machine learning algorithms with traditional financial models like the Markowitz model. The study demonstrates that machine learning can improve asset selection and portfolio optimization, resulting in enhanced risk-adjusted returns and lower volatility. This hybrid framework holds significant potential for improving investment decision-making, particularly in markets characterized by high volatility and informational constraints.

A Hybrid SVM–Markowitz Model for Stock Selection and Optimal Portfolio Formation Based on Fundamental Indicators

Hediyeh Ghafouri vaighan¹ , Robab Kalantari²

-
1. *MSc. Student, Industrial engineering, Financial Engineering Major, Khatam University, Tehran, Iran*
2. *Assistant Prof, Finance department, Khatam university, Tehran , Iran*
-

Corresponding Author: Robab Kalantari
Email: r.kalantari@khatam.ac.ir