مدیریت سرمایه گذاری در بورس اوراق بهادار با استفاده از رویکرد یادگیری ماشین جهت انتخاب سهام و بکارگیری آن به عنوان فین تک

کامیار منکچیان شریف آباد
دکتر فریدون رهنمای رودپشتی
دکتر نرگس یزدانیان
دکتر سیدعلیرضا میرعرب

چکیده:

این مقاله به بررسی راهکارهای نوین مدیریت سرمایه‌گذاری در بورس اوراق بهادار با تمرکز بر بهره‌گیری از رویکردهای یادگیری ماشین پرداخته تا فرآیند تصمیم‌گیری سرمایه‌گذاران بهبود بخشد. با توجه به رشد بی‌سابقه حجم داده‌های مالی و چالش‌های ناشی از کاهش کارایی روش‌های سنتی پیش‌بینی، استفاده از الگوریتم‌های هوشمند به عنوان ابزاری کارآمد جهت پیش‌بینی بازده سهام مطرح گردیده است. در این پژوهش، ابتدا به بررسی مفاهیم اساسی بورس، اهمیت تحلیل داده‌های سری زمانی مالی و نقش داده‌های حجیم در ارائه پیش‌بینی‌های دقیق پرداخته شده است؛ همچنین ضرورت به کارگیری فناوری‌های نوین مالی در بهبود عملکرد سیستم‌های سرمایه‌گذاری مورد تاکید قرار گرفته است.
سپس الگوریتم جنگل تصادفی به عنوان یکی از روش‌های برجسته یادگیری ماشین معرفی و به تفصیل روند ترکیب آن با مدل برنامه‌ریزی درجه دوم جهت تشکیل سبد سهام بهینه شرح داده می‌شود. این مدل ترکیبی، ضمن بهره‌گیری از ویژگی‌های انتخابی الگوریتم، به منظور تخمین دقیق‌تر بازده آتی سهام و تقسیم بهینه وزن‌ها در پرتفوی، رویکردی نوآورانه ارائه می‌دهد. نتایج حاصل از تحلیل‌های آماری نشان می‌دهد که رویکرد پیشنهادی در مقایسه با روش‌های کلاسیک مانند پرتفوی حداقل واریانس، عملکرد بهتری از نظر کاهش ریسک و افزایش بازده داشته و اعتبار علمی آن از طریق شاخص‌های آماری مورد قبول واقع شده است.
این پژوهش علاوه بر ارائه چارچوبی کاربردی برای تحقق سرمایه‌گذاری بهینه، می‌تواند به عنوان مرجعی مفید برای پژوهشگران و سرمایه‌گذاران جهت بهره‌مندی از فناوری‌های یادگیری ماشین و تکنیک‌های پیشرفته در حوزه مدیریت سرمایه‌گذاری و همچنین فین تک به شمار آید. ارائه این روش نوین، زمینه‌ساز توسعه سیستم‌های تصمیم‌گیری خودکار، افزایش شفافیت بازار و بهبود توان رقابتی در بازارهای مالی می‌باشد.

کلمات کلیدی به فارسی : مدیریت سرمایه گذاری، بورس اوراق بهادار، یادگیری ماشین، فین تک
کلمات کلیدی به انگلیسی **Investment Management, Stock Exchange, Machine Learning, FinTech**

مقدمه:

بورس اوراق بهادار به عنوان یکی از ارکان اصـــلی بازارهای مالی و ســـرمایه، نقش حیاتی در تأمین نیاز مالی شـــرکت‌ها و توســـعه اقتصـادی کشـور دارد. تلاش برای اسـتفاده از فناوری‌های نوین در تحلیل روندهای قیمتی و شکل‌دهی سبدهای سرمایه‌گذاری به منظور کسـب بازدهی مطلوب از دغدغه‌های اصلی سرمایه‌گذاران امروزی محسوب می‌شود. با افزایش پیچیدگی‌ها در داده‌های مالی و تأثیر عوامل بیرونی نظیر سیاست و عوامل روانشناختی سرمایه‌گذاران، رویکردهای سنتی تحلیل اغلب ناکافی به نظر می‌رسند. در همین راسـتا، اسـتفاده از رویکردهای یادگیری ماشـین به ویژه الگوریتم جنگل تصادفی، با توجه به توانمندی فراگیر آن در تحلیل داده‌های حجیم و نامنظم، زمینه ساز ایجاد سیستم‌های پیش‌بینی دقیق و ارائه راهکارهای بهینه در مدیریت سبد سهام شده است.

این مقاله با هدف ارائه چارچوبی علمی و کاربردی جهت اسـتفاده همزمان از الگوریتم‌های یادگیری ماشین و روش‌های برنامه‌ریزی درجه دوم، قابلیت پیش‌بینی قیمت و بازده سهام و بهینه‌سازی سبد سرمایه‌گذاری را مورد بررسی قرار می‌دهد. از مهمترین اهداف این پژوهش می‌توان به بهبود دقت پیش‌بینی، کاهش ریسـک سـرمایه‌گذاری و ارائه سـودمندی بالاتر نسبت به روش‌های کلاسیک مانند مدل مارکویتز اشاره نمود.

با توجه به اهمیت تشـکیل سـبد بهینه جهت کاهش ریسـک و افزایش بازده در سـرمایه‌گذاری‌های بورس، مدل‌های سـنتی مانند پرتفوی حداقل واریانس مبتنی بر داده‌های تاریخی، محدودیت‌هایی از جمله عدم انعکاس دقیق روند آینده را دارا می‌باشند. استفاده از داده‌های تاریخی به تنهایی برای پیش‌بینی تغییرات آینده سـهام ریسـک‌های قابل‌توجهی به همراه دارد؛ از این رو، رویکردهای مبتنی بر یادگیری ماشین با استفاده از داده‌های به‌روزتر و الگوریتم‌های هوشمند می‌توانند به عنوان جایگزینی کارآمد معرفی شوند. با توجه به تحولات سریع فناوری و گسترش فناوری‌های نوین مالی، ضرورت جایگزینی مدل‌های سنتی با رویکردهای دانش‌بنیان بیش از پیش احساس می‌شود. توجه به اهمیت شفافیت اطلاعات و افزایش اعتماد سرمایه‌گذاران از جمله مواردی است که می‌تواند به توسعه بازارهای سرمایه کمک شایانی نماید. استفاده از یادگیری ماشین به عنوان روشی جهت پیش‌بینی روندهای اقتصادی آینده، به سرمایه‌گذاران کمک می‌کند تا تصمیمات خود را با استناد به داده‌های علمی و دقیق اتخاذ نمایند.

اهداف پژوهش از منظر کاربردی شامل:

- پیش‌بینی بازده آتی سهام با استفاده از الگوریتم جنگل تصادفی و داده‌های قیمتی بازار
- ارزیابی عملکرد سبد سهام بهینه با بهره‌گیری از مدل برنامه‌ریزی درجه دوم
- مقایسه عملکرد سبدهای بهینه حاصل از روش ارائه شده با روش‌های سنتی مانند پرتفوی حداقل واریانس
- ارائه چارچوبی نوین جهت ترکیب فناوری‌های داده کاوی با نظریه‌های کلاسیک مدیریت سرمایه‌گذاری

این اهداف در راستای ارتقای کیفیت تصمیم‌گیری سرمایه‌گذاران و کاهش خطاهای ناشی از نوسانات بازار مورد بررسی قرار می‌گیرند.

ادبیات و پیشینه تحقیق

ادبیات موضوع

در هر پژوهش علمی، مرور ادبیات و پیشینه تحقیقات انجام‌شده ابزاری ساختاری برای تعیین چارچوب نظری ارائه می‌دهد. در این بخش به دو محور اصلی پرداخته می‌شود:

تاریخچه و تکامل بازار بورس :بورس اوراق بهادار از قرون وسطی آغاز شد و با گذر زمان به یکی از مهم‌ترین ارکان سیستم‌های مالی تبدیل گردید. نخســتین مفاهیم خرید و فروش ســهام در اروپا ایجاد شــد و ســپس با گســترش اقتصــادی و فن‌آوری‌های مدرن بهبودهایی صــورت گرفت؛ از جمله ظهور مدل‌های تحلیل بنیادی و تکنیکال که در روند تصــمیم‌گیری‌های ســرمایه‌گذاران نقش مهمی ایفا کردند.

نظریه‌های مدیریت سبد سرمایه‌گذاری: یکی از مهم‌ترین مدل‌های کلاسیک در این حوزه مدل مارکویتز است که با استفاده از میانگین و واریانس بازده، سبد بهینه‌ای را ارائه می‌دهد. علی‌رغم این که مدل مارکویتز از نظر محاسباتی کارآمد است، اما انتقاداتی از جمله اتکا به داده‌های تاریخی دارد. از سوی دیگر، مدل‌های نوین مبتنی بر فناوری‌های یادگیری ماشین قابلیت انعطاف‌پذیری و دقت بیشتری در تشخیص الگوهای پنهان در داده‌های مالی فراهم می‌کنند.

توضیح الگوریتم جنگل تصادفی به عنوان یکی از ابزارهای یادگیری ماشین: الگوریتم جنگل تصادفی ترکیبی از چندین درخت تصمیم‌گیری است که از طریق بگینگ و انتخاب تصادفی ویژگی‌ها عملکرد بهتری در پیش‌بینی متغیرهای پیوسته و طبقه‌بندی ارائه می‌دهد. به دلیل کاهش احتمال بیش‌برازش[1] و افزایش تعمیم‌پذیری، این الگوریتم در پیش‌بینی بازده سهام قابلیت کاربرد دارد.

نظریه های متفاوتی در خصوص ارزیابی و پیشبینی بورس در بازارهای سازمان یافته مطرح شده است .در اوایل قرن بیستم، گروهی از متخصصان صاحب تجربه در ارزیابی اوراق بهادار اعتقاد راسخ بر این امر داشتند که میتوان از طریق مطالعه و تجزیه وتحلیل روند تاریخی تغییرات قیمت سهام، تصویری را برای پیشبینی قیمت آینده سهام ارائه نمود .مطالعات علمی تر با تأکید بر شناسایی دقیق رفتار قیمت سهام، گرایش به سمت مدلهای ارزشیابی قیمت سهام را به وجود آورد .در ابتدا نظریه گامهای تصادفی به عنوان یک شروع در تعیین رفتار قیمت سهام مطرح شد .سپس به ویژگیها و ساختار بازار سرمایه توجه شد که نتیجه این مطالعات و بررسی ها منجر به فرضیه ی بازار کارآی سرمایه شد .این فرضیه به دلیل ترکیب خاص آن، مورد توجه محافل علمی قرار گرفت .در بازار کارآی سرمایه، اعتقاد بر این است که قیمت سهام انعکاسی از اطلاعات جاری مربوط به آن سهم است و تغییرات قیمت سهام دارای الگوی خاص قابل پیشبینی نیست .نظریات مطرح شده تا دهه ی ۱۹۸۰ میلادی به خوبی تعیین کننده رفتار قیمت سهام در بازار بودند تا اینکه تحولات بازار سهام نیویورک درسال۱۹۸۷ میلادی، اعتبار فرضیات بازار کارآی سرمایه و مدلهایی نظیر تصادفی بودن قیمتها را به شدت زیر سؤال برد .در دهه ۱۹۹۰ میلادی و بعد از آن، بیشتر توجه متخصصان به یک رفتار آشوبگرانه همراه بانظم معطوف شد و تلاش در جهت طراحی مدلهای غیرخطی به منظور پیشبینی قیمت سهام اهمیت روزافزونی یافت .با این نظریات، ازجمله تکنیکهایی که اهمیت بالایی یافتند، سیستمهای هوشمند بودند؛ زیرا با فرض خطی بودن ساختار بازار، به آسانی میتوان بسیاری از مدلها را طراحی نمود .بااین وجود، بسیار سخت است که بتوان رفتار مجموعه های پیچیدهای نظیر بازار سرمایه در یک مجموعه اقتصادی مدرن را به طور کامل در یک مجموعه معادلات ساده و خطی نشان داد. مزیت عمده سیستمهای هوشمند مانند یادگیری ماشین که شبکه های عصبی مصنوعی و درخت تصمیم نیز از آن جمله هستند، در مدلسازی و پیشبینی مجموعه های نامنظم و غیرخطی است(کامرو, ۱۳۹۷)

پرتفوی

پرتفوی به معنی سبد سرمایه گذاری ترجمه شده، سبد سرمایه گذاری مفهومی فراتر از سبد سهام است و شامل سایر سرمایه گذاری های غیر سهم نیز می شود. به لحاظ فنی یک سبد سرمایه گذاری مجموعه کامل دارایی های حقیقی و مالی سرمایه گذاری را در بر می گیرد. لغت پرتفولیو، در عبارت ساده به ترکیبی از دارایی ها گفته می شود که توسط یک سرمایه گذار برای سرمایه گذاری تشکیل می شود. این سرمایه گذار می تواند یک فرد یا یک موسسه باشد. از نظر تکنیکی، یک پرتفولیو در برگیرنده مجموعه ای از دارایی های واقعی و مالی سرمایه گذاری شده یک سرمایه گذار است (بیات & اسدی, ۱۳۹۶)

[1] overfitting

بهینه سازی پرتفوی

مفاهیم بهینه سازی سبد سهام و تنوع بخشی به مثابه ابزاری در راستای توسعه و فهم بازارهای مالی و تصمیم گیری مالی در آمده اند .انتشار نظریه پرتفوی سهام هری مارکوویتز، اصلی ترین ومهمترین موفقیت در این راستا بود. از زمانی که مارکویتز مدل خود را منتشر کرد این مدل تغییرات و بهبودهای فراوانی را در شیوه نگرش مردم به سرمایه گذاری و سبد سهام ایجاد کرد و به عنوان ابزاری کارا برای بهینه سازی سبد سهام به کار گرفته شده است. مارکوویتز پیشنهاد کرد که سرمایه گذاران ریسک و بازده را به صورت توامان در نظر میگیرند و میزان تخصیص سرمایه بین فرصتهای سرمایه گذاری گوناگون را بر اساس تعامل بین این دو انتخاب نمایند یکی از مباحث مهمی که در بازارهای سرمایه مطرح است و باید مورد توجه سرمایه گذاران اعم از اشخاص حقیقی یا حقوقی قرار گیرد، بحث انتخاب سبد سرمایه گذاری بهینه می باشد و در این رابطه، بررسی و مطالعه سرمایه گذاران درجهت انتخاب بهترین سبد سرمایه گذاری با توجه به میزان ریسک و بازده آن انجام می شود . معمولاً فرض بر این است که سرمایه گذاران ریسک را دوست ندارند و از آن گریزانند و همواره در پی آن هستند تا در اقلامی از دارایی ها سرمایه گذاری کنند که بیشترین بازده و کمترین ریسک را داشته باشند . به عبارت دیگر، سرمایه گذاران به بازده سرمایه گذاری به عنوان یک عامل مطلوب می نگرند و به واریانس بازده ها(ریسک)به عنوان یک عنصر نامطلوب نظردارند .در بهینه سازی پرتفوی مسئله اصلی انتخاب بهینه دارایی ها واوراق بهاداری است که با مقدار مشخصی سرمایه میتوان تهیه کرد اگر چه کمینه کردن ریسک و بیشینه نمودن بازده سرمایه گذاری به نظر ساده می رسد اما در عمل روش های متعددی برای تشکیل پرتفوی بهینه به کار رفته است . مسئله بهینه سازی مارکوویتز و تعیین مرز کارای سرمایه گذاری، زمانی که تعداد دارایی های قابل سرمایه گذاری و محدودیت های موجود در بازار کم باشد، توسط مدل های ریاضیات حل شدنی است .اما هنگامی که شرایط و محدودیت های دنیای واقعی در نظرگرفته شود، مسئله پیچیده و مشکل خواهد بود، سال هاست که در حل چنین مسائل پیچیده ای ریاضیات پیشرفته و کامپیوترها به کمک انسان شتافته تا هر چه بیشتر وی را در بیرون آوردن از شرایط عدم اطمینان محیطی و ابهام یاری رساند .از جمله روش هایی که در سالهای اخیر در حل بسیاری مسائل بهینه سازی، گره گشای ابهامات بشر بوده است و در پاسخ به مسائل پیچیده رویکردی موفّق داشته است، روش ها و الگوریتم های موسوم به ابتکاری(فراابتکاری نیز نامیده میشوند)است .روش های ابتکاری که با هدف رفع کاستی های روش های کلاسیک بهینه سازی معرفی شدند با جستجویی جامع و تصادفی، احتمال دستیابی به نتایج بهتر را تا حد زیادی تضمین می کنند ازجمله ی این الگوریتم ها میتوان به الگوریتم ژنتیک ، کوچ پرندگان ، تبرید فلزات ، جستجوی ممنوعه، کولونی مورچگان، رقص زنبورها، رقابت استعماری و ... اشاره کرد (بیات & اسدی, ۱۳۹۶)

بهیه سازی کلاسیک

مارکویتز کسی بود که مفهوم تنوع بخشی در سبد سهام را معرفی کرد و آن را توسعه داد .او به طور کلی نشان داد که چگونه تنوع بخشی در سبد سرمایه، ریسک آن را برای سرمایه گذار کاهش می دهد. سرمایه گذاران می توانند سبد سهام کارا را به ازای یک بازده معین و از طریق کمینه کردن ریسک سبد سهام به دست آورند .در ادامه فرآیند فوق می تواند منجر به تشکیل سبدهای کارا شود که اصطلاحا مرز کارای میانگین واریانس نامیده می شود.

علت این که سهام یک شرکت به عنوان یک دارایی ریسکی مطرح می شود، ثابت نبودن(تصادفی بودن) نرخ بازدهی کل(هفتگی، ماهیانه، سالانه)آن است .چون این نرخ ها با توجه به زمان تغییر می کنند، لذا می توان برای آن ها تابع توزیع احتمال تشکیل داد و معیارهای مورد نیاز مدل مارکویتز نظیر، میانگین، انحراف معیار، کواریانس و .... را از آن به دست آورد.

مدل مارکویتز بر مبنای مفروضات زیر بیان شده است :

- سرمایه گذاران ریسک گریزند و دارای مطلوبیت مورد انتظار افزایشی می باشند و منحنی مطلوبیت نهایی ثروت آن ها کاهنده است.

- سـرمایه گذاران سـبد سرمایه خود را بر مبنای میانگین واریانس مورد انتظار بازدهی انتخاب می کنند .بنابراین منحنی های بی تفاوتی آن ها تابعی از نرخ بازده و واریانس مورد انتظار می باشد.

- هر گزینه سـرمایه گذاری، تا بی نهایت بار قابل تقسـیم اسـت .سـرمایه گذاران افق زمانی(یک دوره ای) داشته و این برای همه سـرمایه گذاران مشابه است .سرمایه گذاران در یک سطح مشخصی از ریسک، بازده بالاتری را ترجیح می دهند و بالعکس، سرمایه گذاران در انتخاب خود به دو عامل توجه می کنند.

الف("بازده مورد انتظار بالا" که عامل مطلوب است.

ب("عدم اطمینان بازده" که عامل نامطلوب است. (بیات & اسدی, ۱۳۹۶)

بهیه سازی نوین

مبنای این روش بر پایه این استدلال استوار است که احتمال خطر از دست دادن سرمایه یا سود یک نوع سهام در بازار بسیاربیشتر از مجموعه یا ترکیب سـهام است. لذا قاعدتاً سرمایه گذار حرفه ای نباید تمام سرمایه خود رادر یک قلم دارایی سرمایه گذاری کند. بلکه بایستی آن را درمجموعهای ازسهام سرمایه گذاری کند که این مجموعه به پرتفوی معروف است. پرتفوی یا سبد سهام متشکل از سهام متعدد شرکتهایی است که در شرایط عادی احتمال کاهش بازده همه داراییها تاحدودی به صفر نزدیک است

با توجه به پیچیده شدن و سرعت عوامل تأثیرگذار در تشکیل پرتفوی بهینه، استفاده از روشهای سنتی تشکیل پرتفوی کار دشواری میباشد. رفتار سهام در بازار، رفتار غیرخطی است بنابراین روشهای خطی نسبت به تشخیص صحیح رفتار غیرخطی ناتوان هستند و استفاده از الگوهای غیرخطی تأثیر به سزایی در اتخاذ تصمیمات بهینه در بازار سهام دارد.

پس از مدلهای کلاسیک، مدلهای زیادی برای تشکیل پرتفوی بهینه سهام ارائه شده است؛ که اکثر مدلها در کنار روشهای گذشته، عوامل متعددی را بر تشـکیل پرتفوی بهینه سـهام اثرگذار میدانند و به دنبال راهی برای حداکثر سـاختن سـود در بازارهای مالی هستند. ایده اساسی نظریه مدرن پرتفوی این است که اگر در داراییهایی که بهطور کامل باهم همبستگی ندارند سرمایهگذاری شود؛ ریسک آن داراییها یکدیگر را خنثی کرده و میتوان یک بازده ثابت با ریسک کمتر به دست آورد.

در زمینه برنامه نویسـی نرم افزار بر پایه الگوی تعیین ارزش سـهام، اوراق قرضـه یا برگهای اختیار معامله سـازمان های تخصـصی متعددی به وجود آمده و راه و روشهای زیادی برای ترکیب مطلوب سرمایه گذاریها ارائه شده است.

یکی از فناوری های نوین که اخیر توجه دانش پژوهان قرار گرفته اسـت مدلها و رویکردهای مبتنی بر یادگیری ماشـین و هوش مصـنوعی میباشـد،این مدلها می توانند روشـهای متعددی را برای انتخاب پرتفوی بهینه سـهام مورد ارائه نمایند. (سـرچمی, پور, محمدی, & زینلی, ۱۳۹۹)

بازده سهام:

بازده سـهام عبارت اسـت از نسبت کل درآمد (یا زیان) حاصل از سرمایه گذاری و میزان سرمایه ای که برای کسب درآمد طی یک دوره معین اسـتفاده شـده است. این دوره می تواند یک روز، یک ماه، یک سال و غیره باشد. برای ارزیابی عملکرد واحدهای تجاری تاکنون ملاک های مختلفی ارائه شـده اسـت که از جمله آنها مهمترین ملاک برای ارزیابی عملکرد و سودآوری موسسات نرخ بازده سهام است که نقش اساسی در سرمایه گذاری دارد. این معیار به تنهایی دارای محتوای اطلاعاتی برای سرمایه گذاران است و از آن برای ارزیابی عملکرد و عملکرد یک واحد تجاری استفاده می شود. وقتی این معیار کاهش یابد، این یک علامت هشدار دهنده برای شـرکت اسـت و نشـانه ای از عدم عملکرد شـرکت اسـت. شـاید این معیار نسبت به معیارهای عملکرد مبتنی بر حسابداری دارای محتوای اطلاعات بیشتری باشد زیرا ارزیابی عملکرد مبتنی بر بازار منعکس کننده اطلاعات سرمایه گذاران نیست(راستین, ۱۳۹۸)

یادگیری ماشین

در موارد زیر به یادگیری ماشین احتیاج داریم:

- در مواردی که انسان خبره در دسترس نیست؛
- زمانی که مسئله ی موردنظر در طول زمان تغییر میکند و به شرایط محیط وابسته است؛
- حالاتی که حل مسئله به تطبیق با شرایط خاصی وابسته است.

هرچند ممکن است یک مدل دقیق ارائه نگردد، اما میتوان یک تقریب خوب و مفید به دست آورد. مدل به دست آمده میتواند برای پیشبینی مورد استفاده قرار گیرد و یا به منظور استخراج دانش از داده ها به کار آید. روشهای مختلف یادگیری ماشین به شرح زیر است:

- روشهای یادگیری نظارت شده

در این الگوریتمها باید تابعی ایجاد شود که ورودیها (متغیرهای مستقل) را گرفته و خروجی موردنظر(متغیر وابسته یا هدف) را تولید کند. نمونه هایی از این الگوریتمها عبارتاند از رگرسیون، درختهای تصمیم، جنگلهای تصادفی،N نزدیکترین همسایه و رگرسیون لجستیک.

- روشهای یادگیری بدون نظارت

یادگیری بدون ناظر متغیر هدف ندارد و خروجی الگوریتم نامشخص است. بهترین مثالی که برای این نوع از الگوریتمها میتوان زد، گروهبندی خودکار (خوشه بندی) یک جمعیت است مثلاً با داشتن اطلاعات شخصی و خریدهای مشتریان، به صورت خودکار آنها را به گروههای همسان و همارز تقسیم کنیم. الگوریتم k-medoid و K-Means از این دسته هستند.

- روشهای تقویت شده

یادگیری تقویت شونده نوع سوم از الگوریتم ها که شاید بتوان آنها را در زمره الگوریتم های بدون ناظر هم دستهبندی کرد. در این نوع از الگوریتم ها؛ یک ماشین برای گرفتن یک تصمیم خاص، آموزش داده میشود و ماشین بر اساس موقعیت فعلی (مجموعه متغیرهای موجود) و عمل های مجاز (مثلاً حرکت به جلو، حرکت به عقب و غیره)، یک تصمیم را میگیرد که در دفعات اول، این تصمیم میتواند کاملاً تصادفی باشد و به ازای هر عمل یا رفتاری که بروز میدهد، سیستم یک بازخورد یا امتیاز به او میدهد و از روی این بازخورد، ماشین متوجه میشود که تصمیم درست را اتخاذ کرده است یا نه که در دفعات بعد در آن موقعیت، همان عمل را تکرار کند یا عمل و رفتار دیگری را امتحان کند. با توجه به وابسته بودن حالت و رفتار فعلی به حالات و رفتارهای قبلی، فرآیند تصمیم گیری مارکوف، یکی از مثالهای این گروه از الگوریتمها میتواند باشد. الگوریتمهای شبکه های عصبی هم میتوانند در این دسته به حساب آیند. منظور از کلمه تقویت شونده در نامگذاری این الگوریتمها هم اشاره به مرحله بازخورد است که باعث تقویت و بهبود عملکرد برنامه و الگوریتم میشود.

- روشهای یادگیری عمیق

نوع جدیدی از یادگیری ماشین برگرفته از نحوه عملکرد مغز هست، مغز انسان ابتدا لبه ها را استخراج میکند، سپس لکه ها، سپس سطوح، سپس اشیاء و غیره، این مشاهدات الهام بخش روند اخیری در یادگیری ماشین شناخته شده به عنوان یادگیری عمیق است که تلاش میکند تا این نوع معماری را در یک کامپیوتر تکثیر کند. (سرچمی ۱۳۹۹ ,. et al)

جنگل تصادفی[2] یکی از محبوبترین و پرکاربردترین الگوریتمهای یادگیری ماشین از نوع نظارتشده و غیر پارامتریک است که برای پیشبرد امور طبقهبندی و رگرسیون از آن استفاده میشود و همچنین یکی از الگوریتمهای مورد توجه در دوره یادگیری ماشین است. جنگل تصادفی از مجموعهای از درختهای تصمیم چندگانه برای تولید پیشبینی یا طبقهبندی استفاده میکند. با ترکیب خروجیهای درختان، الگوریتم جنگل تصادفی نتیجه تلفیقی و دقیقتری ارائه میدهد.این الگوریتم قادر است تا بهطور

[2] Random Forest

موثری مشـــکلات طبقه‌بندی و رگرســـیون را حل نماید، زیرا قدرت این الگوریتم در توانایی مدیریت مجموعه داده‌های پیچیده و کاهش بیش‌برازش نهفته است. الگوریتم جنگل تصادفی در سال ۲۲۲۳ توسط لیو بریمان در دانشگاه اکسفورد ساخته شده است.RF چندین درخت تصـــمیم مختلف به عنوان طبقه بندی های پایه تولید میکند و رأی اکثریت را برای ترکیب با نتایج درختان اصـــلی اعمال میکند.(سعادت،فرساد،۱۳۹۵) جنگل تصادفی هم توانایی رگرسیون، و هم توانایی کلاسبندی را با هم دارد. روش کار در یک جنگل تصـادفی چنین اسـت که، داده های نمونه ای با عمل جایگذاری به تعداد n مجموعه نمونه تقسـیم میشـود و از هر دسته، نمونه ای برای آموزش یک درخت اسـتفاده می شود. تمام درختها عمق مشخصی دارند و در هر گره تقسیم یک ویژگی یا فیچر به تصـادف از بین مجموعه ویژگیها انتخاب میشـود و تقسـیم یا شـاخه بندی بر اساس آن صورت میگیرد. به علت استفاده از چندین دسته نمونه این روش مشکل داده های پرت و دادهه ای گم شده را ندارد.(غلامیان & داودی, ۱۳۹۷)

فینتک

فینتک در لغت به معنای فن آوری مالی اسـت که در حال تبدیل شـدن به بازوی سیستم مالی فعلی می باشد. البته استفاده از فن آوری در امور مالی چندان جدید نیسـت ولی استفاده از بسترهای ترکیبی این فن آوری ها، نظیر بلاک چین، اینترنت اشیا و هوش مصنوعی این حوزه را تحت تغییرِ عمده قرار خواهد داد. (صندوق-بازنشستگی-کشوری, ۱۳۹۶)

تکنولوژی ها و فناوری های نوآورآنه در حال ایجاد تغییرات و تحولات مهمی در عرصـه خدمات مالی از جمله سـپرده گذاری و وام دهی، سـرمایه گذاری، بیمه، خدمات پرداخت و خدمات توسـعه بازار سرمایه می باشد که استفاده از این فناوری ها موجب کاهش زمان و هزینه و ارتقاء کیفیت خدمات این حوزه ها شده که آثار مثبت و مهمی را نیز در رشد و توسعه اقتصادی، افزایش داده ها و اطلاعات جهت تحلیل ها و بروز شـــفافیت بیشـــتر و همچنین افزایش تاب آوری اقتصـــادی خواهد داشـــت.(Monokchian-sharifabad, Mohamadi, & Bigvand, 2021)

فناوری مالی (به اختصـار fintech یاFinTech) فناوری و نوآوری ای اسـت که هدف آن رقابت با روشـهای مالی سـنتی در ارائه خدمات مالی اسـت. این صـنعت در حال ظهور اسـت که از فناوری برای بهبود فعالیت ها در امور مالی استفاده می کند. استفاده از تلفن های هوشـمند برای بانکداری همراه ، سـرمایه گذاری ، خدمات قرض گیری و ارزهای رمزنگاری شـده نمونه هایی از فناوری هایی اسـت که هدف آن دسـترسـی بیشـتر خدمات مالی به عموم مردم است. شرکت های فن آوری مالی شامل هر دو استارتاپ و موسـسـات مالی تأسیس شده و شرکت های فناوری است که سعی می کنند استفاده از خدمات مالی ارائه شده توسط شرکت های مالی موجود را جایگزین یا افزایش دهند.

فناوری های مالی جدید با اسـتفاده از علوم مختلف مانند فناوری بلاک چین ، یادگیری ماشـینی ، بیومتریک و محاسـبات ابری تغییرات بزرگی را در زمینه خدمات مالی ایجاد خواهد کرد. این فناوری ها می توانند تغییرات قابل توجهی در بازارهای مالی و رفتارکرنشگران در این بازارها ایجاد کنند. بهره مندی از فرصت های فناوری های مالی جدید منجر به رشد بیشتر و توسعه اقتصادی بیشتر کشور می شود.

در شـرایط کنونی رشـد بیشـتر و توسعه اقتصادی بیشتر کشور ضروری است. افزایش کارایی و اثربخشی اکثر فعالیتها ، به ویژه در زمینه مالی می تواند زمینه مناسبی برای این رشد و توسعه باشد Fintechs .در توسعه و ارتقاء خدمات مالی می تواند هزینه ها را کاهش دهد ، سـرعت را افزایش دهد ، تصمیمات آگاهانه تری بگیرد ، به داده های جدید دسترسی پیدا کند و درک عمیق تری از مشـتریان و بازارها به دسـت آورد. در این مطالعه ارزیابی میزان آگاهی جامعه از فناوری های جدید مالی (fintech) و ارزیابی انتظارات در این زمینه پرداخته شده است. نتایج این تحقیق می تواند به عنوان مبنایی برای مسائل مربوط به فناوری های نوین و فناوری های مربوط به خدمات مالی استفاده شود. (Monokchian-sharifabad, 2021).

بکارگیری فن آوری های مالی(FinTechs) در راستای دستیابی به راه حل بهینه برای تصمیم گیری های کارآمدتر درخصوص سرمایه گذاری در بازار سهام در عصر حاضر ضرورت داشته تا در کنار مدل های کلاسیک(سنتی) مانند تئوری های مدرن و فرامدرن پورتفوی از یادگیری ماشینی یا هوش مصنوعی برای کمک به تصمیمات سرمایه گذاری استفاده نماییم.
با توجه به نقش مهمی که صنعت خدمات مالی در زندگی روزمره افراد ایفا می کند، داشتن چشم انداز از تغییرات بالقوه ای که می تواند این صنعت را دست خوش تغییر کند- به خصوص برای فعالان این صنعت -امری ضروری می باشد.

در حالیکه فینتک از بسیاری جهات، مانند ارائه ی راه حل های مشتری محور و هزینه های کمتر، بهتر از روش های سنتی عمل می کنند اما هنوز هم سازمان های سنتی مالی در برخی از عرصه های مهم هم چون اعتمادسازی موفق تر از رقبای فینتکی خود ظاهر می شوند با این وجود با توجه به موفقیت های و ارزش افزوده احتمالی پیش بینی می گردد سهم بازار فینتک از بازار خدمات مالی در آینده افزایش یابد. (Monokchian-sharifabad, 2021)

پیشینه پژوهش

مطالعات خارجی

در تحقیقی که توسط شونبرگ در سال ۱۹۹۰ با عنوان پیشبینی قیمت سهام با استفاده از شبکه عصبی ارائه شد، امکان پیشبینی قیمت سهام بهصورت کوتاهمدت، روزبهروز به کمک شبکههای عصبی با مطالعه بر روی سه سهام مهم آلمان که بهطور تصادفی انتخاب شدند، مورد بررسی قرار داده شد. در این تحقیق شبکه های پرسپترون، آدالین، مادالین و پس انتشار خطا مورد بررسی قرار داده شد. در مدت زمان ۱۰ روزه به درجه بسیار بالای دقت تا ۹۰ ٪ رسیدند (کامرو, ۱۳۹۷)

در تحقیقی که توسط کوهارا ۱ و همکاران در سال ۱۹۹۷ با عنوان پیشبینی سهام با استفاده از دانش قبلی و شبکه های عصبی انجام شد، روشی برای استفاده همزمان از دانش قبلی و شبکه های عصبی برای بهبود توانایی پیشبینی ارائه شد .در این پژوهش از دانش قبلی درباره پیشبینی قیمت سهام و اطلاعات روزنامه داخلی و خارجی استفاده شده است .چند شاخص اقتصادی نیز بر اساس دانش پیشین انتخاب شده و با دانش موجود به شبکههای عصبی وارد شده است .نتایج نشان داد که خطای پیشبینی رویکرد ارائه شده کوچکتر از تجزیه وتحلیل رگرسیون چندگانه در سطح ۵ ٪ اهمیت است(کامرو, ۱۳۹۷)

در تحقیقی که توسط بالینگز و همکاران در سال ۲۰۱۵ با عنوان ارزیابی چند دسته بندی برای پیشبینی جهت روند قیمت سهام انجام شد، هدف از انجام آن، ارزیابی روشهای اجماعی نظیر جنگل تصادفی، AdaBoost و Kernel Factory در برابر مدلهای طبقه بندی واحد نظیر شبکه های عصبی، رگرسیون منطقی، ماشینهای بردار پشتیبانی وK- همسایه نزدیکتر عنوان شد. اطلاعات مورداستفاده از ۵۷۶۷ شرکت اروپایی که به طور عمومی فهرست شده اند جمع آوری شده است. نتایج نشان داد که جنگل تصادفی، بهترین الگوریتم است که به دنبال آن ماشین بردار پشتیبان قرار دارد، Kernel Factory ، AdaBoost ، شبکه عصبی،K همسایه نزدیک و رگرسیون منطقی قرار دارد (کامرو, ۱۳۹۷)

در این تحقیقی که توسط میلوسویچ در سال ۲۰۱۶ با عنوان پیشبینی سود: پیشبینی حرکت بلندمدت قیمت سهام با استفاده از یادگیری ماشین انجام شد از یک روش یادگیری ماشین برای ارزیابی قیمت آتی سهام در طول زمان استفاده شد. این روش قادر است در ۷۶,۵ درصد موارد بهدرستی پیشبینی کند که آیا ارزش یک شرکت میتواند ۱۰ درصد بالاتر یا بیشتر در طول یک سال بالا رود (کامرو, ۱۳۹۷)

تاکور و همکاران (۲۰۱۷) انتخاب پرتفوی سهام را با استفاده از تئوری شواهد دمپستر شیفر در بورس اوراق بهادار بمبئی با استفاده از متغیرهای نسبت قیمت به سود، نسبت قیمت به ارزش دفتری، نسبت قیمت به فروش، نسبت بدهیهای بلندمدت به حقوق صاحبان سهام، سود هر سهم و نسبت پرداخت مورد بررسی قرار دادهاند. از مدل دلفی-فازی برای شناسایی عوامل ضروری استفادهشده سپس عوامل ضروری و دادههای تاریخی با استفاده از تئوری شواهد دمپستر شیفر رتبه بندی سهام انجام شده و

سـرانجام یک مدل پرتفوی که سـهام با رتبه بالا را پیشنهاد میدهد انتخاب گردیده است. نتیجه عملکرد در مقایسه با عملکرد اخیر دارایی ها بسیار رضایت بخش میباشد. (سرچمی ۱۳۹۹ ., et al)

پاتاری و همکاران (۲۰۱۸) مقایسـه روشهای تصمیمگیری چند معیاره برای انتخاب پرتفوی سهام در بورس نزدک آمریکا طی یک دوره چهل و سـه سـاله از ۱۹۶۹ تا ۲۰۱۱ را با مورد بررسی قرار دادند. از چهار روش مقیاس متوسط، روند سلسله مراتب تحلیلی، تکنیک برای ترجیح سـفارش با شباهت به راه حل ایده آل و تجزیه وتحلیل پوشش افزایشی داده ها استفاده شده است. متغیرهای تحقیق شـامل ارزش بازار سهام، ارزش شرکت، سود، ارزش دفتری سهام، فروش، سود سهام عادی، جریان نقد عملیاتی، جریان نقد آزاد، سـود قبل از بهره، مالیات و اسـتهلاک و سـود قبل از بهره و مالیات هسـتند. نتایج حاکی از این اسـت که سرمایه گذارانی که سـبک های سـرمایه گذاری مشخص را دنبال میکنند، این سبک های متفاوت را هنگام انتخاب روشهای تصمیم گیری چند معیاره میتوانند داشته باشند که بهترین سازگاری را با هدفهای انتخاب پرتفوی آنها به همراه دارد. (سرچمی ۱۳۹۹ ., et al)

جین و همکاران (۲۰۱۹) به بررسـی رویکرد برابری ریسـک سلسله مراتبی (HRP) تخصیص سبد سهام ، معرفی شده توسط لوپز د پرادو (۲۰۱۶) پرداختند و از نظریه نمودار(گراف) و یادگیری ماشینی برای ایجاد یک پرتفوی متنوع استفاده نمودند. مانند روشهای سـنتی تخصـیص مبتنی بر ریسک ، HRP نیز تابعی از برآورد ماتریس کوواریانس است ، با این حال ، به برگشت ناپذیری[3] آن نیاز ندارد. در آن مقاله ، ابتدا تأثیر نادرست تشخیص کوواریانس بر عملکرد روشهای مختلف تخصیص را مطالعه شده و در مرحله بعد ، تحت یک مدل پیش بینی کوواریانس مناسـب مطالعه صـورت گرفته که آیا HRP مبتنی بر یادگیری ماشـین از سـبدهای سـنتی مبتنی بر ریسک بهتر عمل می کند یا خیر. نتایج پژوهش حاکی از این بوده است که وقتی برآورد کوواریانس خام[4] است ، پرتفوی های معیار نوسانات معکوس[5] قوی تر هستند ، و پس از آن پرتفوی های مبتنی بر یادگیری ماشین. (Jain & Jain, 2019)

کازمارک و همکاران(۲۰۲۱) نشـان دادند که تکنیک های بهینه سـازی سـبد دارایی که توسـط میانگین های واریانس مارکوویتز و برابری ریسـک سـلسـله مراتبی (HRP) ارائه می شـوند ، بازدهی تعدیل شده سهام با استفاده از سهام از پیش انتخاب شده با ابزار یادگیری ماشین را افزایش می دهد. آنها از روش جنگل تصادفی برای پیش بینی مقطع بازده اضافی مورد انتظار استفاده نموده و n سـهام با بیشترین پیش بینی ماهانه را انتخاب می کردند با سه تکنیک مختلف-میانگین واریانس ، HRP و را برای ایجاد وزن سبد با استفاده از بازده سهام از S & P500 و STOXX 600 برای استحکام مقایسه نمودند. نتایج خارج از نمونه[6] نشان می دهد که هر دو بهینه کننده میانگین واریانس و HRP از قانون 1/N بهتر عمل می کنند. این نتیجه گیری در تقابل با انتقاد رایج از کارآیی بهینه سازان است و روش جدیدی را در استفاده عملی بالقوه آنها ارائه می دهد. (Kaczmarek & Perez, 2021)

مطالعات داخلی

مدرس و اسـتخری (۱۳۸۶) سـبد سـهام در بورس اوراق بهادار تهران را با الگوریتم ژنتیکی انتخاب می کنند. نتایج تحقیقات آنها همچنین نشان می دهد که بهینه سازی پرتفوی با استفاده از ژنتیک نسبت به راه حل آن با مدل های کلاسیک کارآمدتر است و با این رویکرد ابتکاری می توان حدود عملکرد بهتری نسبت به راه حل مدل سنتی مارکویتز به دست آورد. (راستین, ۱۳۹۸)

رهنمای رود پشـتی و همکاران (۱۳۸۹) در پژوهش حاضـر به معرفی مدل نوینی از مدل های قیمتگذاری دارایی سرمایهای به نام پرداخته می شـود . ب رای ای ن (R-CAPM) مدل قیمت گذاری دارایی سـرمایه ای تجدید نظرشـده منظور، ابتدا درجه اهرم اقتصادی، یعنی اثرات اختلالات اقتصادی که شامل پنج متغیر کلان اقتصادی نرخ تورم، نرخ ارز، نرخ بیکاری، صادرات و هزینه مالی اسـت، بر روی فروش شرکت ها اندازه گیری و از این طریق ضریب حساسیت بتا را محاسبه کرده و بهدنبال آن با دستیابی به مدل

[3] invertibility
[4] crude
[5] inverse volatility
[6] out-of-sample

قیمت گذاری دارایی سرمایهای تجدید نظرشده، بازده مورد انتظار بهصورت واقعی محاسبه ، سپس پیشبینی میشود و جهت سنجش توان تبیین این مدل را با سایر مد لهای قیمت گذاری اعم از مدل قیمت گذاری دارایی سرمایهای، مدل قیمت گذاری دارایی سرمایهای کاهشی و مدل قیمت گذاری دارایی سرمایهای تعدیل شده برای ۶۷ شرکت پذیرفته شده در بورس اوراق بهادار تهران در دوره زمانی ۸ ساله مورد مقایسه قرار داده و با استفاده از تحلیل همبستگی و رگرسیون و آزمون هایی نظیر فریدمن، ویلکاکسون دادهها مورد تجزیه و تحلیل قرار می گیرد . نتیجه پژ وهش بیانگر آن است که مدل قیمت گذاری دارایی سرمایهای تجدید نظرشده با توجه به شرایط حاکم بر جامعه ما در مقایسه با سایر روشهای قیمت گذاری در پیشبینی ریسک و بازده دارای توان تبیین بالاتری است. (رهنمای-رودپشتی & حسینی, ۱۳۸۹)

رهنمای رود پشتی و همکاران (۱۳۹۳) بهینه سازی پرتفوی متشکل از سهام صندوقهای سرمایه گذاری مشترک بورس اوراق بهادار تهران را با رویکرد الگوریتم ژنتیک در سی صندوق سرمایه گذاری مشترک فعال در بازار سرمایه ایران در طی سالهای ۱۳۸۹ الی ۱۳۹۱ با سه متغیر بازده مورد انتظار، ریسک پرتفوی و کوواریانس مورد بررسی قرار دادند. نتایج حاکی از آن است که الگوریتم ژنتیک میتواند جهت انتخاب سبد متشکل از سهام صندوقهای مشترک به کار رود و با استفاده از آزمون زوجی مشخص شد که سبدهای تشکیل شده با استفاده از الگوریتم ژنتیک نسبت به روش سنتی، مطلوبتر میباشند. همچنین اندازه سبد تأثیر چندانی بر نتایج نداشته و در تمام سطوح، الگوریتم ژنتیک دارای عملکرد بهتری است و هر چه تنوع سبد بیشتر و بزرگتر باشد، برتری الگوریتم ژنتیک بر روش خطی قابل ملاحظه تر میشود. (سرچمی et al., 1399)

بیات، شکری (۱۳۹۴)، در پژوهشی به بررسی فرایند انتخاب پرتفوی بهینه به روش ارزش در معرض ریسک پرداختند ودر آن به انواع روش های بهینه سازی به روش ارزش درمعرض ریسک مانند روش واریانس کواریانس، شبیه سازی تاریخی وروش شبیه سازی مونت کارلو اشاره کردند و در نهایت به این – نتیجه رسیدند که در انتخاب پرتفوی بهینه مدل ارزش در معرض ریسک بهترین مدل پیشنهادی می باشد (بیات & اسدی, ۱۳۹۶)

کوهبنانی نژاد و همکاران (۱۳۹۷) انتخاب پرتفوی بهینه با استفاده از سیستم خبره در محیط فازی ممدانی را در صنایع شیمیایی، فراورده نفتی، کانی فلزی و فلزات اساسی بورس اوراق بهادار تهران در بازه زمانی ۱۳۸۷ تا ۱۳۹۰ انجام دادند. با در نظر گرفتن متغیرهای حاشیه سود خالص، سود قبل از کسر مالیات، بازده حقوق صاحبان سهام، بدهی به حقوق صاحبان سهام، ارزش بازار، نسبت قیمت به سود، سود تقسیمی، نسبت ارزش بازار به ارزش دفتری، مومنتوم قیمت و شاخص فنی، شرکتها را رتبهبندی کردند و پس از بهینه سازی پرتفوی با معیارهای بازده، بازده مرکب، ترینر و جنسن عملکرد پرتفوی تشکیلشده با استفاده از سیستم خبره در محیط فازی ممدانی را مورد سنجش قرار دادند. نتایج ارزیابی عملکرد پرتفوی تشکیل شده برای سه حالت سرمایه گذار ریسک گریز، ریسک خنثی و ریسک پذیر نشان میدهد که عملکرد پرتفوی پیشنهادی مثبت بوده؛ اما در مقیاسی دقیقتر پرتفوی تشکیل شده برای سرمایه گذار ریسک گریز در وضعیت مطلوبتری قرار دارد. (سرچمی et al., 1399)

(اقتصاد & محمدی, ۱۴۰۲)پیشبینی بازده در تشکیل پورتفولیو را با مدل یادگیری ماشین، یعنی جنگل تصادفی و مدل یادگیری عمیق حافظه کوتاه مدت طولانی ترکیب میکند. به منظور ارزیابی عملکرد مدل پیشنهادی، داده های تاریخی ۵ ساله از سال ۱۳۹۶ تا ۱۴۰۱ از شاخص ۵ صنعت بانکی، خودرویی، دارویی، فلزی و نفتی است. نتایج تجربی نشان میدهد که مدلهای بهینه سازی میانگین واریانس با پیشبینی بازدهی بهو سیله جنگل تصادفی ، بهتر عمل میکنند.

روش‌شناسی پژوهش

پژوهش حاضر از نظر نوع داده کمی بوده و با استفاده از داده‌های واقعی از شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران انجام شده است. در ادامه، گام‌های اجرایی و فرآیند تحلیل داده‌ها تشریح می‌شود.

- جامعه آماری و نمونه پژوهش جامعه آماری شامل تمامی شرکت‌های پذیرفته‌شده در بورس تهران است که پس از اعمال معیارهای انتخاب، نمونه نهایی شامل شرکت‌هایی می‌شود که:

شرکت‌های هلدینگ و سرمایه‌گذاری نباشند
بانک‌ها به دلیل ماهیت خاص فعالیت حذف شوند
تغییر سال مالی یا توقف عملیات نداشته باشند
اطلاعات مورد نیاز به صورت کامل در دسترس باشند دوره زمانی مورد بررسی از سال ۱۳۹۵ تا ۱۴۰۳ بوده و داده‌های مربوط به قیمت پایان روز، حجم معاملات و مبلغ معاملات روزانه ثبت شده‌اند.
- فرآیند پیش‌بینی قیمت سهام
الف) گردآوری داده‌ها:داده‌های جمع‌آوری شده از منابع معتبر مانند نرم‌افزار دریافت اطلاعات بورس و بانک‌های اطلاعاتی سازمان بورس، ابتدا در نرم‌افزار Excel جهت تحلیل اولیه ذخیره و بررسی شدند.
ب) پیاده‌سازی الگوریتم جنگل تصادفی:
• الگوریتم یادگیری ماشین جنگل تصادفی با استفاده از نرم‌افزار R پیاده‌سازی شده است.
• فرآیند برگزاری آموزش مدل شامل انتخاب تصادفی زیرمجموعه‌های داده (با تکنیک Bootstrap Sampling) و ساخت مجموعه‌ای از درختان تصمیم‌گیری انجام شده و برای هر نمونه جدید پیش‌بینی بر مبنای میانگین خروجی درختان صورت می‌گیرد.
• معیارهای ارزیابی عملکرد مدل شامل میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و میانگین نسبی خطای مطلق برای سنجش دقت پیش‌بینی به کار گرفته شده‌اند.
- تشکیل سبد بهینه سهام:
پس از پیش‌بینی بازده آتی سهام، فرآیند بهینه‌سازی سبد سرمایه‌گذاری با استفاده از روش برنامه‌ریزی درجه دوم انجام می‌گیرد. در این بخش مدل برنامه‌ریزی درجه دوم جهت تعیین وزن هر سهم در سبد بهینه به کار رفته است و هدف ارائه سبدی است که برای سطح معینی از بازده، حداقل ریسک محتمله را تامین کند.
با ترکیب خروجی‌های الگوریتم جنگل تصادفی و مدل برنامه‌ریزی درجه دوم، سبد بهینه سهام شناخته شده و از لحاظ عملکردی از طریق معیارهایی نظیر نسبت شارپ ارزیابی شده است.

الگوریتم جنگل تصادفی[۷] یکی از قدرتمندترین الگوریتم‌های یادگیری ماشین است که برای مسائل طبقه‌بندی[۸] و رگرسیون[۹] مورد استفاده قرار می‌گیرد. این الگوریتم بر پایه‌ی درختان تصمیم‌گیری[۱۰] ساخته شده است و با ایجاد مجموعه‌ای از درختان تصمیم، پیش‌بینی‌های دقیقی ارائه می‌دهد .جنگل تصادفی به زبان ساده از ترکیب چندین درخت تصمیم‌گیری استفاده می‌کند تا دقت و پایداری مدل را بهبود ببخشد. این رویکرد بر اساس دو تکنیک اصلی عمل می‌کند:

۱. بگینگ[۱۱] (تقویت داده با مجموع‌گیری بوت‌استراپ): در جنگل تصادفی، داده‌های آموزشی به صورت تصادفی و با جایگذاری[۱۲] انتخاب می‌شوند و برای ساخت درختان مختلف مورد استفاده قرار می‌گیرند. به این ترتیب، هر درخت تصمیم روی یک زیرمجموعه تصادفی از داده‌ها آموزش داده می‌شود.

---

[7] Random Forest
[8] Classification
[9] Regression
[10] Decision Trees
[11] Bagging
[12] Sampling with Replacement

۲. انتخاب تصادفی ویژگی‌ها: در هر گره[۱۳] از هر درخت، تنها یک زیرمجموعه تصادفی از ویژگی‌ها[۱۴] برای تقسیم‌بندی داده‌ها انتخاب می‌شود. این امر باعث تنوع بیشتر میان درختان می‌شود و مشکلاتی نظیر تسلط ویژگی‌های قوی[۱۵] بر مدل را کاهش می‌دهد.

به دلیل استفاده از بگینگ و انتخاب تصادفی ویژگی‌ها، احتمال بیش‌برازش[۱۶] مدل کاهش یافته و توان تعمیم مدل افزایش می یابد. مراحل کار الگوریتم جنگل تصادفی بشرح زیر می باشد :

۱-انتخاب زیرمجموعه‌ای از داده‌ها[۱۷]:

به صورت پیش‌فرض، برای هر درخت یک زیرمجموعه تصادفی از داده‌ها با جایگذاری [۱۸] انتخاب می‌شود.

داده‌هایی که در زیرمجموعه آموزشی یک درخت گنجانده نشده‌اند، به عنوان داده‌های خارج از بسته[۱۹] عمل می‌کنند. این داده‌ها برای ارزیابی عملکرد مدل استفاده می‌شوند.

۲-ایجاد درختان تصمیم‌گیری:برای هر زیرمجموعه، یک درخت تصمیم‌گیری جداگانه ساخته می‌شود. در هر گره درخت، زیرمجموعه تصادفی از ویژگی‌ها برای یافتن بهترین تقسیم‌بندی بررسی می‌شود، این ویژگی باعث کاهش وابستگی به یک ویژگی خاص و افزایش تنوع میان درخت‌ها می‌شود.

۳-پیش‌بینی با هر درخت:هر درخت ساخته‌شده، پیش‌بینی خود را بر روی داده‌های جدید ارائه می‌دهد.

۴- در مسائل رگرسیون: از میانگین مقادیر پیش‌بینی‌شده توسط درختان بشرح زیر استفاده می‌شود.

در مسائل رگرسیون ، مدل به صورت زیر پیش‌بینی انجام می‌دهد:

۱. مدل تعداد زیادی درخت تصمیم‌گیری می‌سازد. هر درخت به صورت جداگانه روی یک زیرمجموعه تصادفی از داده‌ها، که از طریق بوت‌استراپینگ انتخاب شده‌اند، آموزش داده می‌شود. هر زیرمجموعه شامل نمونه‌هایی از داده‌های اصلی و معمولاً کمتر از کل داده‌ها است.
۲. وقتی می‌خواهیم مقدار یک نمونه جدید را پیش‌بینی کنیم:
   - هر درخت تصمیم‌گیری یک پیش‌بینی مستقل می‌دهد (مقداری را که برای آن نمونه مناسب می‌داند انتخاب می‌کند). این پیش‌بینی بر اساس اصول تقسیم‌بندی است که در طول ساخت درخت‌ها آموخته شده است.
۳. مقدار پیش‌بینی‌شده نهایی برابر با میانگین مقادیر پیش‌بینی‌شده توسط همه درخت‌ها خواهد بود:

---

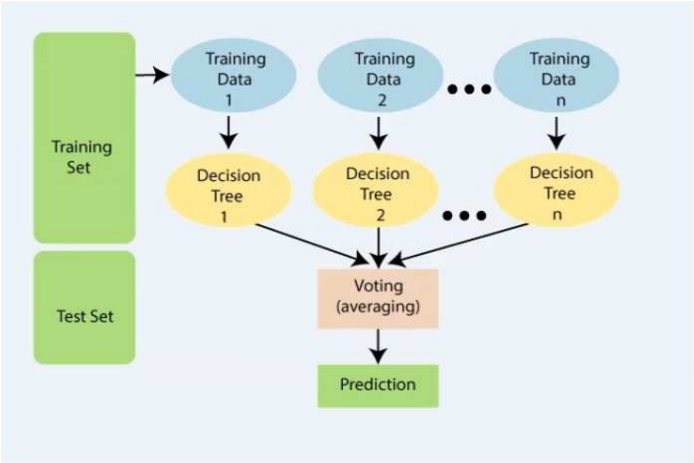
[13] Node
[14] Features
[15] Dominant Features
[16] Overfitting
[17] Bootstrap Sampling
[18] Sampling with Replacement
[19] Out-of-Bag - OOB

فرمول

$$\hat{y} = \frac{1}{n}\sum y_n$$

که در آن n تعداد درخت‌ها و yi مقدار بازگشتی پیش‌بینی‌شده توسط درخت iام است.)

که شماتیک آن نیز بشرح شکل زیر می باشد :



جنگل تصادفی مورد استفاده در این پژوهش بر مبنای الگوریتم جنگل تصادفی Breiman (بر اساس کد فرترن اصلی بریمن و کاتلر) که برای طبقه‌بندی و رگرسیون پیاده‌سازی می شود همچنین می تواند برای پیش بینی و در حالت بدون نظارت برای ارزیابی نزدیکی بین نقاط داده استفاده شود.(L Breiman, 2001; L Breiman, 2002)

و برای سنجش عملکرد مدل در پیش بینی از دو معیار میانگین مربعات خطا[۲۰](RMSE)، میانگین خطای مطلق[۲۱] (MAE) و میانگین نسبی خطای مطلق بشرح زیر استفاده شده است :

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{t=1}^{N}(r_t - \hat{r_{t-1}})\text{^}2}$$

$$\text{MAE}=\frac{1}{N}\sum_{t=1}^{N}|r_t - \hat{r_{t-1}}|$$

که در آن r و $\hat{r}$ به ترتیب بازده واقعی و بازده پیش بینی شده را در زمان t نشان می دهد.

بهینه‌سازی سبد سهام در حالت عمومی که شامل برخی محدودیت ها روی وزن ها است یک مساله برنامه ریزی درجه دوم است و تکنیک های حل مسائل برنامه‌ریزی درجه دو[۲۲] (QP) یک روش قدرتمند برای مدیریت ریسک و بازده در سرمایه‌گذاری‌های

[20]Root Mean Squared Error
[21] Mean Absolute Error
22 Quadratic Programming

مالی اســت. بهینه‌سازی سبد سهام با استفاده از برنامه‌ریزی درجه دو به ما این امکان را می‌دهد که با در نظر گرفتن ریسک و بازده مورد نظر، ســبد سهام بهینه‌ای را برای سرمایه‌گذاری مشخص کنیم. در این روش، تابع هدف معمولاً به شکل یک تابع درجه دو که شامل ریسک (واریانس یا انحراف معیار) و بازده مورد انتظار سبد است، تعریف می‌شود.

توابع نرم‌افزار R، برای حل مسائل بهینه‌سازی برنامه‌ریزی درجه دو یکی از ابزارهای پرکاربرد برای بهینه‌سازی پرتفوی است که در آن هدف کمینه کردن ریسک (با استفاده از ماتریس واریانس و کواریانس بازده‌ها) است در حالی که بازده مورد انتظار مشخص شده است.

الگوریتم و تابع مورد استفاده برای حل مسائل بهینه‌سازی درجه دو با استفاده از روش‌های عددی پیشرفته طراحی شده است.

مسائل بهینه‌سازی درجه دو به طور کلی به شکل زیر هستند:

$$\min_X \left(\frac{1}{2} X^T Q X + C^T X\right)$$

در آن:

X : یک بردار از متغیرهای تصمیم است (وزن‌های دارایی‌های مختلف در یک پرتفوی)، بردار وزن‌های سهام در سبد سرمایه گذاری، که نشان‌دهنده درصد سرمایه‌گذاری در هر سهم است. به عنوان مثال، اگر x_1 نشان‌دهنده وزن سهم اول باشد، x_2 وزن سهم دوم و غیره، باید مجموع وزن‌ها برابر با ۱ باشد:

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

Q : یک ماتریس مربعی مثبت (ماتریس واریانس و کواریانس است)، که این ماتریس واریانس و کوواریانس بین سهام‌ها را شامل می‌شود و ریسک سبد سهام را به تصویر می‌کشد. هر عنصر σij در ماتریس نشان‌دهنده کوواریانس بین بازده‌های سهام i و j است. این ماتریس به ما کمک می‌کند تا چگونگی تداخل (ارتباط) ریسک بین سهام‌ها را بفهمیم.

C : یک بردار از مقادیر خطی است (برای بازده مورد انتظار دارایی‌ها تعریف می‌شود)و بردار شامل بازده‌های مورد انتظار برای هر سهم است.

در کنار این تابع هدف، محدودیت‌های خطی اعمال می شوند:

$$AX \leq b$$

که در اینجا:

A ماتریسی است که محدودیت‌های غیرمنفی و مساوی را بیان می‌کنند.

b برداری است که محدودیت‌های مورد نظر را تعریف می‌کنند.

روش عددی به کار رفته برای حل این نوع مسائله برنامه ریزی درجه دوم، روش عددی پایدار الگوریتم‌های بهینه سازی محدود (Constrained Optimization Algorithms) بر مبنای تکنیک متغیرهای دوگان (Dual Variable Methods) برای حل مسائل برنامه‌ریزی درجه دو می باشد. الگوریتم پیشنهادی بر اساس روش دوگان عمل می‌کند و با استفاده از تکنیک‌های عددی پایدار برای حل مسائل بهینه سازی طراحی شده است.

در روش دوگان (Dual Methods) تابع بهینه‌سازی به دو بخش تقسیم می‌شود، تابع هدف اصلی و تابع هدف دوگان که محدودیت‌ها را در نظر می‌گیرد. با حل دو مسئله‌ی جداگانه و تطبیق مقادیر دوگانه، می‌توان به حل بهینه دست یافت که فرآیند حل آن بشرح ذیل می باشد :

۱- تعریف مسئله: ابتدا مسئله‌ی بهینه‌سازی را از طریق تابع مربوطه و در قالب تابع هدف و محدودیت‌ها تعریف می نماییم.

۲- محدودیت‌ها: محدودیت‌ها به صورت ماتریس‌ها و بردارهایی برای تضمین شرایط سرمایه‌گذاری مانند محدودیت‌های غیرمنفی بودن وزن‌ها (در صورتی که خرید کوتاه (short selling) مجاز نباشد) یا محدودیت‌هایی مثل مجموع وزن‌ها برابر یک باشد، وارد می‌شوند.
۳- بهینه‌سازی: الگوریتم پس از دریافت داده‌های ورودی، شروع به بهینه‌سازی تابع هدف می‌کند. این کار با استفاده از یک روش عددی انجام می‌شود که از روش‌های دوگان برای یافتن پاسخ بهینه استفاده می‌کند.
۴- توقف الگوریتم: فرآیند بهینه‌سازی زمانی متوقف می‌شود که تغییرات بین تکرارهای متوالی کمتر از یک آستانه معین باشد یا تعداد تکرارها به حداکثر برسد.

بهینه سازی سهام بر مبنای تکنیک برنامه ریزی درجه دوم روش دوگانه گلدفارب و ایدنانی (۱۹۸۲، ۱۹۸۳) برای حل مسائل برنامه ریزی درجه دوم به شکل $\min(-d^T b + \frac{1}{2} b^T D b)$ با محدودیت های $A^T b \geq b_0$ پیاده سازی شده است. (Goldfarb & Idnani, 1982, 1983)

در این روش ابتدا بردار بازدهی هدف را به عنوان قید مساوی در بهینه سازی مارکویتز اسفتاده نموده و سپس ماتریس واریانس و کواریانس دارایی ها مشخص می گردد و سپس با استفاده از نرم افزار مدل سازی صورت گرفته و وزنهای بهینه و مقدار تابع هدف(واریانس بازده پرتفوی) در قالب لیست به عنوان خروجی داده می شود.

نتایج و تجزیه و تحلیل داده‌ها و ارزیابی عملکرد سبد سهام

پس از اجرای مدل‌های پیشنهادی در بازه زمانی مورد بررسی، نتایج به دست آمده نشان می‌دهد که دقت پیش‌بینی بازده سهام با استفاده از الگوریتم جنگل تصادفی از اطمینان قابل قبولی برخوردار بوده و همچنین معیارهای ارزیابی مدل (مانند RMSE و MAE) نشانگر همگرایی مناسب مدل به داده‌های واقعی می‌باشد.

	RMSE	MAE	میانگین نسبی خطای مطلق
Average	۵۹۸,۸	۵۲۵,۶	۰,۰۵۲

	RMSE	MAE
STDEV	۴۸۷,۲	۴۵۷,۶

سبد سهامی بهینه که از طریق ترکیب خروجی پیش‌بینی بازده و برنامه‌ریزی درجه دوم تشکیل شده، نسبت به سبدهای حاصل از رویکردهای کلاسیک دارای نسبت شارپ بهتری بوده و ریسک ناشی از نوسانات بازار را کاهش داده است.
نتایج حاصل از تحلیل‌های آماری اعتبار رویکرد ارائه شده را در شرایط واقعی بازار تأیید می‌کند و نشان می‌دهد رویکرد یادگیری ماشین در کنار مدل‌های بهینه‌سازی می‌تواند ابزار قدرتمندی برای سرمایه‌گذاران فراهم آورد.

به طور کلی عملکرد سبد سهام از جنبه‌های مختلفی از جمله نسبت شارپ و شاخص‌های بازدهی ارزیابی شده است. در این پژوهش، نسبت شارپ برای سنجش رابطه بین بازده اضافی و ریسک استفاده شده است. مدل پیشنهادی نسبت به سبدهای حاصل از روش‌های کلاسیک دارای نسبت شارپ بهینه‌تر و ریسک کمتری بوده است.

• شـــاخص‌هـای عملکرد از جمله معیارهای میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE) نیز در ارزیابی دقت پیش‌بینی به کار رفته و به نفع مدل هوشمند عمل کرده‌اند.
نتایج ارزیابی عملکرد نشـان می‌دهد که در شـرایط متغیر و نوسانی بازار، استفاده از رویکرد ترکیبی یادگیری ماشین و برنامه‌ریزی درجه دوم، عملکرد مطلوب‌تری را نسبت به روش‌های کلاسیک ارائه می‌دهد.

بحث و نتیجه گیری

نتایج به دست آمده از این پژوهش حاکی از آن است که رویکرد ترکیبی مبتنی بر الگوریتم جنگل تصادفی و برنامه‌ریزی درجه دوم، نسـبت به روش‌های کلاسیک مانند پرتفوی حداقل واریانس عملکرد بهتری از نظر کاهش ریسک و افزایش بازده دارد. نکات مهمی که در این بخش مورد بحث قرار می‌گیرند عبارتند از:

• اهـمـیـت اســـتـفـاده از داده‌هـای بـه‌روز و رویـکـردهـای نـویـن در پـیـش‌بـیـنـی رونـدهـای آتـی ســـهـام
• کاهش وابستگی به داده‌های تاریخی که در مدل‌های سنتی به چشم می‌خورد و افزایش انعطاف‌پذیری در مواجهه با نوسانات بازار
• اســـتـفـاده از الـگـوریـتـم‌هـای هـوشـــمـنـد جـهـت اســـتـخـراج الـگـوهـای پـنـهـان و بـهـبـود دقـت پـیـش‌بـینی
• اهمیت تعیین وزن بهینه در ترکیب سهام منتخب جهت کاهش ریسک غیر سیستماتیک سبد و افزایش پایداری بازدهی

با توجه به نتایج به دســت آمده، می‌توان نتیجه گرفت که رویکرد ارائه‌شـده نه تنها از نظر فنی بلکه از منظر کاربرد عملی نیز نقش مهمی در تسهیل فرآیندهای تصمیم‌گیری سرمایه‌گذاران ایفا خواهد کرد.
نتایج پژوهش حاضـر نشـان‌دهنده اهمیت و کارایی اسـتفاده از فناوری‌های نوین مالی به ویژه الگوریتم‌های یادگیری ماشـین در مدیریت سرمایه‌گذاری است. نتیجه‌گیری کلی به شرح زیر بیان می‌شود:
الگوریتم جنگل تصـادفی با اسـتفاده از ویژگی‌های انتخابی و تکنیک‌های نمونه‌گیری، توانسـته است پیش‌بینی دقیقی از تغییرات و بازده سهام ارائه دهد.
ترکیب خروجی مدل پیش‌بینی با برنامه‌ریزی درجه دوم در تعیین وزن بهینه سـهام، منجر به کاهش ریسک و بهبود عملکرد سبد سرمایه‌گذاری گردیده است.
مقایسـه با روش‌های سـنتی نشان می‌دهد که رویکرد هوشمند پیشنهادی از نظر سوددهی و کاهش ریسک، عملکرد بهتری از خود به جای گذاشته است.
با توجه به نتایج حاصـل، اسـتفاده از این رویکرد می‌تواند به عنوان یک ابزار تصـمیم‌گیری علمی برای سـرمایه‌گذاران در بازارهای بورس به کار رود.
با وجود محدودیت‌های موجود، پژوهش ارائه شـده زمینه‌های توسـعه بیشـتر و کاربرد گسـترده‌تر این روش‌ها را فراهم می‌آورد و می‌تواند گامی نوین در توسعه سیستم‌های مدیریت سرمایه‌گذاری باشد.

منابع

Uncategorized References

Breiman, L. (2001). Random Forests. *Machine Learning, 45* .

Breiman, L. (2002). Manual On Setting Up, Using, And Understanding Random Forests V3.1 .

Goldfarb, D., & Idnani, A. (1982). Dual and Primal-Dual Methods for Solving Strictly Convex Quadratic Programs. *Numerical Analysis, Springer-Verlag, Berlin*, 226–239 .

Goldfarb, D., & Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming, 27* .

Jain, P., & Jain, S. (2019). Can Machine Learning-Based Portfolios Outperform Traditional Risk-Based Portfolios? The Need to Account for Covariance Misspecification *Risks* .

Kaczmarek, T., & Perez, K. (2021). Building portfolios based on machine learning predictions. *Economic Research-Ekonomska Istraživanja* .

Monokchian-sharifabad, k. (2021). Survey and analysis of the public knowledge of new financial technologies (fintechs) and expectations in this area. *International Journal of Financial Technology (IJFT), 2* .

Monokchian-sharifabad, K., Mohamadi, A & ,.Bigvand, A. (2021). A review and analysis of conducted researches in the field of economic resilience and the role of Fin-Tech in this field. *International Journal of Financial Technology (IJFT), 1* .

اقتصاد, ا. ع., & محمدی, ع. (۱۴۰۲). بهینه سازی سبد سرمایه گذاری به کمک پیش بینی بازده مورد انتظار با استفاده از روشهای شبکه عصبیLSTM، جنگل تصادفی و ARIMA چشم انداز مدیریت مالی, ۴۳.

بیات, ع., & اسدی, . (۱۳۹۶). بهینه سازی پرتفوی سها : سودمندی الگوریتم پرندگان و مد مارکویتز. مهندسي مالي و مدیریت اوراق بهادار, ۳۲.

راستین, ف. (۱۳۹۸). مقایسه مد های انتخاب سبد بهینه سها مارکوئیتز و مد محدودیت ا با روش یادگیری ماشین. (کارشناسی ارشد), موسسه آموزش عالی حکیم طوس,

رهنماي-رودپشتي, ف., & حسینی, ز. ا. (۱۳۸۹). تبیین قیمت گذاري دارايي سرمايه اي: مقايسه تطبيقي مد لها. بررسی های حسابداری و حسابرسی, ۶۲.

سرچمی, ., پور, ا. خ., محمدی, ., & زینلی, ح. (۱۳۹۹). بهکارگیری مدلهای یادگیری ماشین در تشکیل پرتفوی بهینه سها و مقایسه کارایی آنها. مهندسی مالی و مدیریت اوراق بهادار, ۴۵.

صندوق-بازنشستگی-کشوری. (۱۳۹۶). آینده بازارهای خدمات مالی: هلدینگ بانک،بیمه و سها صندوق بازنشستگی کشوری.

غلامیان, ا., & داودی, س. ر. (۱۳۹۷). پیش بینی روند قیمت در بازار سها با استفاده از الگوریتم جنگل تصادفی. مهندسي مالي و مدیریت اوراق بهادار, ۳۵.

کامرو, س. (۱۳۹۷). ارائه استراتژی ترکیبی معاملات سها و تشکیل سبد سها در بورس اوراق بهادار با استفاده از ترکیب شیوه های نوین داده کاوی و تحلیل فنی. (کارشناسی ارشد), دانشگاه صنعتی شریف