مطالعات حسابداری و حسابرسی، بهار ۱۴۰۲، شماره ۴۵، از صفحه ۱۲۵ تا ۱۴۶

پیش‌بینی بازده سهام با تاکید بر نقش معیارهای مالی و نظارتی با استفاده از روش‌های یادگیری ماشین

دکتر علی جعفری[1]©

استادیار حسابداری، دانشگاه آزاد اسلامی واحد قائم‌شهر، ایران

مصطفی منصوری‌خواه

دانشجوی دکتری حسابداری، دانشگاه آزاد اسلامی واحد قائم‌شهر، ایران

دکتر عباسعلی پورآقاجان

استادیار حسابداری، دانشگاه آزاد اسلامی واحد قائم‌شهر، ایران

(تاریخ دریافت: ۱۹ اردیبهشت ۱۴۰۱؛ تاریخ پذیرش: ۱۶ آبان ۱۴۰۱)

مقدمه: پیش‌بینی بازده سهام اگرچه پیچیده است ولی همواره مورد علاقه سرمایه‌گذاران می‌باشد. سرمایه‌گذاران و تصمیم‌گیران جهت پیش‌بینی بازده سهام نیازمند اطلاعات هستند.

هدف: این پژوهش با هدف پیش‌بینی بازده سهام در شرکت‌های پذیرفته‌شده بورس اوراق بهادار تهران با تاکید بر نقش معیارهای مالی و نظارتی با استفاده از روش‌های یادگیری ماشین انجام شد.

روش: این پژوهش از لحاظ هدف کاربردی و به لحاظ نوع مطالعه میدانی- کتابخانه‌ای است. جامعه آماری را شرکت‌های پذیرفته‌شده بورس اوراق بهادار تهران طی سال‌های ۱۳۹۰ تا ۱۳۹۶ (یک دوره هفت‌ساله) تشکیل دادند و آزمون فرضیه‌ها با استفاده از روش‌های یادگیری ماشین (رویکرد خطی و غیرخطی و قانون‌گرا) انجام گرفت.

یافته‌ها: نتایج آزمون متغیرگزینی با استفاده از روش Relief-F حاکی از ارتباط معنادار متغیرهای بازده دارایی، نسبت قیمت به سود هر سهم،کیوتوبین، نسبت وجه‌نقد عملیاتی، اهرم مالی و نسبت مالکان نهادی با بازده سهام است. همچنین نتایج نشان داد که روش‌های یادگیری ماشین PINSVR در حالت خطی و غیرخطی و CART توانایی مناسبی (بیش از ۹۰ درصد) جهت پیش‌بینی بازده سهام دارند اما روش غیرخطی PINSVR و قانون‌گرای CART نسبت به روش خطی قدرت پیش‌بینی بالاتری را نشان داد.

نتیجه‌گیری: با توجه به نتایج به‌دست آمده، به صاحبان سرمایه و تصمیم‌گیران شرکت‌ها توصیه می‌شود که در تصمیم‌گیری‌های خود پیرامون سرمایه‌گذاری در بورس اوراق بهادار تهران از قدرت پیش‌بینی الگوریتم‌های هوش مصنوعی به‌ویژه روش‌های غیرخطی استفاده کنند.

واژه‌های کلیدی: بازده سهام، روش‌های یادگیری ماشین، معیارهای مالی و نظارتی.

[1] jafarilarijani@gmail.com

© (نویسندهٔ مسئول)

مقاله علمی - پژوهشی

مقدمه

بازده سهام عبارت است از نسبت میان کل عایدی (زیان) حاصل از سرمایه‌گذاری بر میزان سرمایه‌ای که به‌منظور کسب عایدی در یک دوره معین (یک روز، یک ماه، یک سال وغیره) مصرف گردیده است [۴]. مفهوم بازده دارای اهمیت بسیار است، زیرا سرعتی را اندازه می‌گیرد که طی آن ثروت سرمایه‌گذار افزایش می‌یابد یا اگر زیانی متوجه اوست، ثروت وی کاهش می‌یابد [۱۶]. نرخ بازده سهام یک سرمایه‌گذاری مفهومی است که از نظر سرمایه‌گذاران مختلف، معانی متفاوتی دارد. در حال حاضر معمولی‌ترین و مهم‌ترین معیار ارزیابی عملکرد و سودآوری مؤسسات که نقش کلیدی در سرمایه‌گذاری دارد، بازده سهام است. این معیار شاید دارای محتوای اطلاعاتی بیشتری در مقایسه با معیارهای عملکرد بر مبنای حسابداری باشد چون ارزیابی عملکرد بر مبنای بازار، اطلاعات سرمایه‌گذاران را به‌خوبی منعکس نمی‌سازد [۵ , ۴]. با گسترش روزافزون و اهمیت بازارهای سرمایه در تجهیز و گردآوری سرمایه‌های کوچک فردی به سمت فعالیت‌های تولیدی، شناسایی رفتار سرمایه‌گذاران و متغیرهای تأثیرگذار بر قیمت و بازده سهام در این بازارها اهمیت زیادی پیدا کرده است. بی‌تردید سرمایه‌گذاری در بورس، بخش مهمی از اقتصاد کشور را تشکیل می‌دهد و بیشترین میزان سرمایه از طریق بازارهای سرمایه در سرتاسر جهان مبادله می‌شود و اقتصاد ملی به‌شدت متأثر از عملکرد بازار سرمایه است. اگرچه این بازار هم برای سرمایه‌گذاران حرفه‌ای و هم برای عموم مردم به‌عنوان یک ابزار سرمایه‌گذاری در دسترس است، لیکن بازارهای سرمایه از پارامترهای کلان اقتصادی و غیراقتصادی و بسیاری از متغیرهای دیگر تأثیر می‌پذیرد که این تعدد عوامل مؤثر بر بازارهای سرمایه و ناشناخته بودن آن‌ها، موجب عدم اطمینان در زمینه سرمایه‌گذاری است [۱۴]. نظر به اینکه بازده سهام یکی از فاکتورهای مهم در انتخاب بهترین سرمایه‌گذاری است از این رو پژوهش‌گران و تحلیل‌گران حوزه مالی در جستجوی یافتن سنجه عملکردی هستند که از سوی آن بتوانند با نزدیک‌ترین تقریب، بازده و تغییرات آن را برای یک شرکت پیش‌بینی کنند. لذا شناسایی و معرفی اطلاعات حسابداری مرتبط جهت پیش‌بینی بازده و استفاده از روش‌های کارای پیش‌بینی به منظور راهنمایی سرمایه‌گذاران و تحلیل‌گران مالی امری ضروری به نظر می‌رسد [۸]. از جنبه نوآوری این پژوهش می‌توان به موارد زیر اشاره کرد:

۱. تحلیل بازده سهام به‌طور همزمان با معیارهای نظارتی[1] و مالی با رویکرد **Relief-F** که ایده اصلی آن متغیر گزینی بر اساس فاصله اقلیدسی و مستقل از متغیرهای انتخاب شده قبلی است.
۲. پیش‌بینی بازده سهام با متغیرهای منتخب با رویکرد خطی، غیرخطی و قانون‌گرا و معرفی روش برتر در پیش‌بینی بازده سهام.

[1] Regulatory criteria

چارچوب نظری و پیشینه پژوهش

پیش‌بینی بازده سهام اگرچه پیچیده است ولی همواره مورد علاقه سرمایه‌گذاران می‌باشد. سرمایه‌گذاران و تصمیم‌گیران جهت پیش‌بینی بازده سهام نیازمند اطلاعات هستند. بخشی از این اطلاعات از طریق گزارش‌های مالی شرکت‌ها تأمین می‌شود.گزارش‌های مالی محصول حسابداری است، لذا سودمندی آن را می‌توان از طریق بررسی نقش این اطلاعات در پیش‌بینی بازده سهام بررسی کرد. یکی از موضوعات مهم در این پیش‌بینی‌ها انتخاب مدل‌های مناسب می‌باشد [۹]. حاکمیت شرکتی می‌تواند با ارتقای عملکرد شرکت و افزایش دسترسی آن‌ها به منابع مالی موجبات رشد و توسعه اقتصادی پایدار را فراهم کند، همچنین می‌تواند میزان آسیب‌پذیری شرکت را در مواقع بروز بحران مالی و هزینه معاملات و سرمایه را کاهش داده و به گسترش بازار سرمایه کمک کند [۱۱]. صاحب‌نظران و دست‌اندرکاران بازار معتقدند تعداد اعضای هیات مدیره شرکت یکی از عوامل مهم تعیین‌کننده ارزش شرکت و بازده سهام آن است. اعضای زیاد مانع از برقراری ارتباطات شده و تصمیم‌گیری را مشکل‌تر کرده واثربخشی نظارت را از بین می‌برد [۲]، یک هیات مدیره در تنظیم اهداف شرکت که همانا ارزش سهامداران در بلندمدت است و نیز در ارزیابی مناسب بودن شیوه‌ها و استراتژی‌هایی که مدیریت جهت رسیدن به اهداف شرکت مورد استفاده قرار می‌دهد، مسئول است. هیات مدیره برای اینکه از بکارگیری موثر یک استراتژی اطمینان حاصل کند، عملکرد مدیریت را به دقت مورد بازنگری قرار می‌دهد و این کار را با هدف اعطای پاداش یا تنبیه مدیریت انجام می‌دهد. حاکمیت شرکتی به‌عنوان یک محرک عملکرد و یک معیار حیاتی برای ارزشیابی واحد تجاری و تضمینی برای اعتبار گزارش‌های مالی و حسابداری تعبیر شده است. بنابراین حاکمیت شرکتی می‌تواند هم اطلاعات حسابداری و هم ارزش بازار را تحت تاثیر قرار دهد [۶]. به بیان بوهل[1] و همکاران، مالکان نهادی یکی از بازیگران اصلی در بازار سرمایه هستند که استفاده آن‌ها از توانایی‌هایشان تابعی از میزان سرمایه‌گذاری آن‌هاست. لذا میزان مالکیت نهادی این سرمایه‌گذاران می‌تواند با نوسان‌پذیری بازده سهام در ارتباط باشد و مالکان نهادی با پیاده‌سازی حاکمیت شرکتی، شفافیت اطلاعاتی را افزایش داده، از عدم تقارن اطلاعاتی می‌کاهند و این عمل موجب تقویت کارایی بازار سرمایه می‌شود؛ به‌طوری‌که انتظار می‌رود نوسان‌پذیری بازده سهام کاهش یافته، سبب ایجاد بازاری جذاب و مطمئن برای سرمایه‌گذاران جدید شود [۲۰]. از سوی دیگر تمرکز مالکیت به سه دلیل باید تأثیر مثبتی بر ارزش شرکت داشته باشد: نخست، سهامداران عمده از فعالیت‌های یک جانبه بزرگ‌ترین سهامدار جلوگیری می‌کنند. چون این سهامداران انگیزه زیادی دارند تا کنترل شرکت را در دست گیرند و بزرگ‌ترین سهامدار را محدود سازند. دوم، کارایی بازار برای کنترل شرکت افزایش می‌یابد چون این سهامداران عمده می‌توانند مبارزه‌ای را برای کنترل شرکت به راه اندازند و یا زمانی که مدیریت موجود عملکرد ضعیفی دارد به شروع یک نزاع برون سازمانی بر سر کنترل کمک کنند. سوم، سهامداران عمده انگیزه دارند تا مدیریت را مستقیماً کنترل

[1] Bohl, M.T.

کنند [۲۳]. در همین راستا، اسوالد[1] و جاهرا [2] از وجود ارتباط مستقیم بین سهام تحت تملک مدیران با بازده سهام خبر دادند [۲۲]. هادسان[3] و همکاران نیز ارتباط مستقیم و معناداری بین مالکیت درون سازمانی و بازده غیرعادی سهام یافتند. افزایش درصد سهام تحت تملک مدیران در کنار نظارت اعضای هیات مدیره و سهامداران برون سازمانی بر عملکرد مدیریت به منظور جلوگیری از عملکرد ضعیف وی، می‌تواند این نتیجه را تعدیل کند. لذا افزایش مالکیت مدیریت موجب تقویت هرچه بیشتر بازده سهام شرکت می‌شود [۲۱].

مطالعه‌های بسیاری در زمینه پیش‌بینی بازده سهام و عوامل موثر بر آن با استفاده از روش‌های گوناگون انجام گرفته است. رجبی و همکاران [۸] به پیش‌بینی بازده سهام با استفاده از تکنیک‌های داده‌کاوی (تحلیل تفکیکی) پرداختند. عوامل سود (زیان) عملیاتی، سود هر سهم، نسبت جاری و نسبت آنی به‌عنوان ورودی‌های پژوهش انتخاب شده‌اند. برازش مدل بر مبنای فرم خطی تابع کاب – داگلاس است. یافته‌ها حاکی از عملکرد مناسب تحلیل تفکیکی (روش گام به گام) در پیش‌بینی بازده سهام، تنها با استفاده از اثرگذارترین متغیر مستقل بر متغیر وابسته می‌باشد. عسگرنژاد نوری [۱۲] به بررسی عوامل مؤثر در بازده سهام شرکت‌ها با به‌کارگیری روش فراتحلیل پرداخته است. نتایج نشان داد نسبت‌های نقدینگی، اهرمی، فعالیت، مدیریت سود و ویژگی‌های شرکت در بازده سهام شرکت‌ها تأثیر نداشته است؛ اما تأثیر مثبت سایر عوامل یعنی نسبت‌های سودآوری شامل نرخ بازده حقوق صاحبان سهام، میزان سود و حاشیهٔ سود، نسبت‌های بازار شامل بازده بازار، نسبت‌های جریان‌های نقدی شامل جریان‌های نقدی ناشی از عملیات، شاخص‌های ریسک شامل صرف ریسک، شاخص‌های پیش‌بینی سود شامل افق زمانی پیش‌بینی سود و درنهایت، سرمایه‌گذاری واقعی در بازده سهام تأیید شد. یافته‌های پژوهش طبیبی [۱۱] نشان داد که با تقویت عوامل مربوط به حاکمیت شرکتی و افزایش میزان تمرکز مالکیت، کیفیت افشا و استقلال هیات‌مدیره، از میزان نوسان‌پذیری بازدهی سهام و همچنین بازده غیرعادی، کاسته می‌شود. به‌عبارتی نشان‌دهنده وجود رابطه معکوس بین عوامل حاکمیت شرکتی و ریسک سهام می‌باشد. انصاری و همکاران [۱] با بررسی هم زمان عوامل موثر بر ساختار سرمایه و بازده سهام دریافتند که ساختار سرمایه، سودآوری، تکانه قیمت سهام و ارزش شرکت عوامل موثر بر بازده سهام است. همچنین ارتباط متقابل بین ساختار سرمایه و بازده سهام وجود دارد؛ به‌گونه‌ای که بازده سهام طبق تئوری زمان‌بندی بازار بر ساختار سرمایه تأثیر منفی و ساختار سرمایه طبق رابطه مستقیم ریسک و بازده بر بازده سهام تأثیر مثبت داشته است. علی‌محمدی و همکاران [۱۳] با پیش‌بینی بازده سهام شرکت‌ها با استفاده از نسبت‌های مالی تحت رویکرد درخت تصمیم نشان دادند که الگوریتم‌های CRT و ECHAID در تبیین بازده جاری بهترین عملکرد را دارند. همچنین قدرت مدل‌ها در تبیین بازده جاری بیشتر از پیش‌بینی بازده آتی است. اما به دلیل

[1] Oswald, S.L.

[2] Jahera, J.S.

[3] Hudson, C.D.

عدم قابل اتکا بودن مدل‌ها از نظر آماری، فرضیه برقراری ارتباط تبیینی بین نسبت‌های مالی و تغییرات بازده جاری و آتی سهام رد شد. مدرس، کهنسال [۱۷] با ارزیابی ساختارهای خطی و غیرخطی در پیش‌بینی بازده سهام نشان دادند، شبکه‌های عصبی نسبت به مدل‌های ساختاری از انعطاف‌پذیری بالاتری در حل مسایل برخوردار می باشند. می‌توان ادعا کرد که این شبکه‌های عصبی به خصوص زمانی که امکان بکارگیری روش‌های پارامتری به دلیل غیرخطی بودن روند تغییرات وجود نداشته باشد، ابزاری سودمند در پیش‌بینی‌های مالی به حساب آمده و دقت بالاتری بدست می‌دهند. راعی و چاوشی [۷] به پیش‌بینی‌پذیری رفتار بازده سهام در بورس بوسیله مدل خطی عاملی و شبکه‌های عصبی مصنوعی پرداختند. نتایج حاصله حاکی از موفقیت این دو مدل در پیش‌بینی رفتار بازده سهام مورد نظر و همچنین برتری عملکرد شبکه‌های عصبی مصنوعی بر مدل چند عاملی است. بالکیلار[1] و همکاران [۱۹] با بررسی نقش عدم قطعیت سیاست اقتصادی در پیش‌بینی بازده سهام و نوسانات آن دریافتند که شواهد آماری غیرخطی در مجموعه داده‌های مورد بررسی نسبت به آزمون خطی توانایی بالاتری جهت پیش‌بینی بازده سهام دارد.

با توجه به مبانی نظری بیان‌شده و هدف پژوهش، فرضیه‌های زیر تدوین‌شده است.

۱) قدرت تبیین بازده سهام توسط معیارهای مالی نسبت به معیارهای نظارتی با استفاده از رویکرد متغیر گزینی Relief-F بالاتر است.

۲) روش یادگیری ماشین PINSVR[2] در حالت غیرخطی نسبت به روش PINSVR در حالت خطی توانایی بالاتری جهت پیش‌بینی بازده سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را دارد.

۳) روش یادگیری ماشین PINSVR در حالت غیرخطی نسبت به روش قانون‌گرای CART توانایی بالاتری جهت پیش‌بینی بازده سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را دارد.

۴) روش یادگیری ماشین PINSVR در حالت خطی نسبت به روش قانون‌گرای CART توانایی بالاتری جهت پیش‌بینی بازده سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را دارد.

روش‌شناسی پژوهش

این پژوهش از لحاظ هدف کاربردی و به لحاظ نوع مطالعه میدانی- کتابخانه‌ای، با استفاده از اطلاعات تاریخی به‌صورت پس رویدادی است. جامعه آماری این تحقیق شامل تمامی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران که شرایط زیر را دارا باشند، است.

۱. در دوره مورد بررسی تغییر دوره مالی نداشته باشند.
۲. شرکت‌های سرمایه‌گذاری، واسطه‌گری‌های مالی، بانک، بیمه و لیزینگ نباشند.
۳. داده‌های مورد نظر آن‌ها در دسترس باشد.

[1] Balcilar, M.

[2] Parametric-insensitive nonparallel support vector regression

در نهایت با توجه به محدودیت‌های ذکرشده تعداد ۲۰۲ شرکت طی سال‌های ۱۳۹۰ تا ۱۳۹۶ به‌عنوان جامعه آماری انتخاب شده‌اند که با توجه به در دسترس بودن اطلاعات، تمام شرکت‌ها به‌عنوان نمونه آماری مورد بررسی قرارگرفته است.

متغیرهای پژوهش

با توجه به اهداف پژوهش، معیارهای نظارتی منتخب مطابق با پژوهش‌های (رهنمای رودپشتی و محسنی [۱۰]؛ کیان و فقیه [۱۵]؛ طبیبی [۱۱]) و معیارهای مالی منتخب مطابق با پژوهش‌های، (عسگرنژاد نوری [۱۲]؛ انصاری و همکاران [۱]) انتخاب شده‌اند. متغیرهای استفاده شده در این پژوهش به شرح جدول شماره ۱ می‌باشند.

جدول ۱. متغیرهای پژوهش

نام متغیر کمی	تعریف عملیاتی
بازده دارایی	از تقسیم سود خالص بر کل دارایی‌ها
نسبت دارایی جاری	از تقسیم دارایی جاری بر کل دارایی‌ها
بازده حقوق صاحبان سهام	از تقسیم سود خالص بر حقوق صاحبان سهام
اهرم مالی	از تقسیم کل بدهی ها بر جمع دارایی‌ها
بازده فروش	از تقسیم سود خالص بر کل فروش
نسبت سود تقسیمی	از تقسیم سود تقسیمی بر دارایی شرکت
بازده فروش	از تقسیم سود خالص بر کل فروش شرکت
نسبت وجه نقد نگه‌داری شده در شرکت به دارایی‌ها	از تقسیم حاصل جمع وجه نقد و سرمایه‌گذاری کوتاه مدت بر جمع دارایی‌ها
اندازه شرکت	از لگاریتم طبیعی کل دارایی‌ها
حاشیه سود عملیاتی	از تقسیم سود عملیاتی بر فروش شرکت
نسبت جاری	از تقسیم دارایی جاری بر بدهی جاری
نسبت فروش به کل دارایی‌ها	از تقسیم فروش بر دارایی‌ها
نسبت ارزش افزوده اقتصادی	$EVA = NOPATt - (WACCt \times Capitalt - 1)$ EVA ارزش‌افزوده اقتصادی، NOPAT سود خالص عملیاتی پس از کسر مالیات، Capital سرمایه بکار گرفته‌شده در شرکت، WACC نرخ متوسط هزینه سرمایه است، در نهایت بر جمع دارایی‌های شرکت تقسیم شده است.

نسبت ارزش افزوده بازار	از حاصل (ارزش بازار سهام – حقوق صاحبان سهام) تقسیم بر کل دارایی‌ها
نسبت وجه نقد عملیاتی	از تقسیم وجه نقد عملیاتی بر کل دارایی‌ها
نقش دوگانه مدیر عامل	اگر مدیر عامل شرکت، رئیس یا نایب رئیس هیات مدیره باشد از متغیر مصنوعی ۱ و در غیر این صورت از متغیر مصنوعی صفر
نسبت کیوتوبین	از حاصل (ارزش بازار سهام +ارزش دفتری بدهی ها) ÷ ارزش دفتری دارایی‌ها
تغییر مدیرعامل	اگر مدیر عامل نسبت به سال قبل تغییر کرده باشد از متغیر مصنوعی ۱ و در غیر این صورت صفر
ریسک سیستماتیک	ریسک سیستماتیک درجه تغییرات بازده یک سرمایه‌گذاری خاص نسبت به تغییرات بازده مجموعه سرمایه‌گذاری بازار است و با شاخص β اندازه‌گیری می‌شود. $\beta = \frac{Cov(R_m, R_i)}{\delta^2 R_m}$
اندازه هیات‌مدیره	تعداد اعضای هیات‌مدیره شرکت
نسبت مدیران غیرموظف	از تقسیم تعداد مدیران غیرموظف به‌کل اعضای هیات‌مدیره
نسبت مالکان نهادی	مطابق تعریف بند ۲۷ ماده ۱ قانون بازار اوراق بهادار، بانک‌ها، شرکت‌ها و هر شخصیت که بیش از ۵ درصد سهام منتشرشده را در دست داشته باشد به‌عنوان معیار محاسبه سهامدار نهادی در نظر گرفته‌شده است.
تمرکز مالکیت	از شاخص هریفیندال – هریشمن
ارتباطات سیاسی	درصد سهام متعلق به دولت
تخصص کمیته حسابرسی	از نسبت اعضای کمیته حسابرسی با تخصص مالی تقسیم بر کل اعضا
استقلال کمیته حسابرسی	از نسبت اعضای مستقل کمیته حسابرسی نسبت به کل اعضا
متغیرهای وابسته	
بازده سهام	بازده سهام سالانه شرکت.
روش پژوهش	با استفاده از الگوریتم رگرسیون بردار پشتیبان غیرحساس-پارامتریک در حالت خطی و غیرخطی
دوره پژوهش	۱۳۹۶-۱۳۹۰ (یک دوره ۷ ساله)
طرح پژوهش	رویکرد دو مرحله‌ای، ۱- روش انتخاب ویژگی مبتنی برRelief-F، ۲- پیش‌بینی بازده سهام با استفاده از روش یادگیری ماشین PINSVR و CART

یافته‌های پژوهش

روش‌های گوناگونی برای پیش‌بینی بازده سهام وجود دارد. روش‌های یادگیری ماشین[1] ممکن است بهترین روش برای پیش‌بینی بازار سهام باشد، زیرا بر اساس تجربیات یاد می‌گیرد و در واقع به بن‌بست نمی‌رسد. این الگوریتم‌ها از ابزارهای ایده آلی هستند که علاوه بر بهره‌برداری از آمار، به جنبه‌های ذهنی نیز توجه می‌نماید؛ بنابراین امکان پیش‌بینی بازده سهام با این روش‌ها وجود دارد. برای پیاده‌سازی الگوریتم‌ها از نرم‌افزار متلب[2] نسخه **۲۰۱۲a** استفاده شد.

متغیر گزینی (روش انتخاب ویژگی مبتنی بر Relief-F):

این روش از یک راه‌حل آماری برای انتخاب ویژگی استفاده می‌کند. این روش یک الگوریتم مبتنی بر وزن‌دهی به متغیرهای مستقل است که ایده آن از الگوریتم‌های مبتنی بر نمونه الهام گرفته‌شده است. این الگوریتم از میان مجموعه D نمونه آموزشی یک زیرمجموعه شرکت را انتخاب می‌کند. الگوریتم به‌صورت تصادفی یک شرکت-سال از این زیرمجموعه را به‌عنوان یک نمونه انتخاب می‌کند، سپس مبتنی بر ویژگی‌های (متغیرهای مستقل) این نمونه، نزدیک‌ترین برخورد[3] و نزدیک‌ترین شکست[4] را بر اساس تابع ارزیابی فاصله اقلیدسی پیدا می‌کند. نزدیک‌ترین برخورد، نمونه‌ای (شرکت-سالی) است که کمترین فاصله اقلیدسی را در میان سایر نمونه‌های هم‌کلاس با نمونه انتخاب‌شده دارد. ایده اصلی در این الگوریتم این است که هر چه اختلاف بین اندازه یک ویژگی در شرکت- سال انتخاب‌شده و نزدیک‌ترین برخورد کمتر باشد، این ویژگی بهتر است. بعلاوه یک ویژگی خوب آن است که اختلاف بین اندازه آن ویژگی و نزدیک‌ترین شکست آن بیشتر باشد. در این الگوریتم هرکدام از متغیرهای مستقل در ابتدا دارای یک وزن W هستند، که در شروع الگوریتم مقدار آن برابر صفر است. الگوریتم پس از تعیین نزدیک‌ترین برخورد و نزدیک‌ترین شکست، وزن‌های ویژگی‌ها را به‌روزرسانی می‌کند. بعد از تعیین فاصله برای تمام شرکت-سال‌های موجود در مجموعه نمونه‌ها، الگوریتم، ویژگی‌هایی (f) را که وزن آن‌ها کمتر یا مساوی با یک حد آستانه ($Threshold$) و منفی است را حذف می‌کند و سایر ویژگی‌ها به‌عنوان زیرمجموعه ویژگی جواب (T)، باز می‌گردند. مقدار حد آستانه توسط کاربر تعیین می‌گردد، Relief برای ویژگی‌های نویزی و همبسته خوب عمل می‌کند و پیچیدگی زمانی آن به‌صورت تابعی خطی از تعداد ویژگی‌های داده‌شده و $NoSample$ است. یکی از محدودیت‌های اساسی این الگوریتم این است که ویژگی‌هایی که دارای افزونگی[5] باشند را پیدا نمی‌کند. این مشکل را می‌توان با یک جستجوی تعیین جامعیت[6] برای

[1] Machine learning methods

[2] Matlab

[3] Near hit

[4] Near miss

[5] Redundant

[6] Subsequent exhaustive search

زیرمجموعه‌های انتخاب‌شده توسط الگوریتم حل کرد. علاوه بر این، مشکل دیگر الگوریتم این است که با مسائل دو کلاسه خوب کار می‌کند. این محدودیت نیز با الگوریتم Relief-F مرتفع شده و با الگوریتم جدید مشکل داده‌های غیرکامل (نمونه‌های آموزشی غیرکامل) نیز حل‌شده است. همچنین نسخه دیگری از این الگوریتم بانام RRelief-F برای مسائل رگرسیون نیز وجود دارد.

داده‌های شرکت-سال جمع‌آوری شده برای متغیر وابسته بازده سهام به الگوریتم انتخاب ویژگی RRelief-F داده شدند و برای پیدا کردن نزدیک‌ترین برخوردها و نزدیک‌ترین شکست‌ها از الگوریتم KNN[1] با مقدار $K = 200$ استفاده شد. در جدول ۲ متغیرهای مستقل انتخابی جهت ورود به روش‌های یادگیری ماشین ارائه‌شده است.

جدول ۲ : متغیرهای مستقل انتخابی برای پیش‌بینی بازده سهام

Article I. متغیرهای مستقل	**Article II**. وزن
Article III. بازده دارایی	**Article IV**. ۰/۰۰۷۱۳
Article V. نسبت قیمت به سود هر سهم	**Article VI**. ۰/۰۰۷۶۷
Article VII. کیوتوبین	**Article VIII**. ۰/۰۰۴۶۸
Article IX. نسبت وجه نقد عملیاتی	**Article X**. ۰/۰۰۴۶۳
Article XI. اهرم مالی	**Article XII**. ۰/۰۰۴۴۸
Article XIII. نسبت مالکان نهادی	**Article XIV**. ۰/۰۰۴۴۲

پس از انتخاب متغیرهای مستقل مسئله، این متغیرهای مستقل جهت ساخت مدل به روش‌های یادگیری ماشین داده شده‌اند.

رگرسیون بردار پشتیبان غیرموازی غیر حساس-پارامتریک (PINSVR):

در این بخش، مدل **PINSVR** ارائه‌شده است. هدف **PINSVR** ارزیابی مدل رگرسیون از طریق تطبیق خودکار ناحیه غیرحساس-پارامتریک یک‌شکل دلخواه و با سایز حداقل است. به‌طوری‌که دربرگیرنده داده‌های معلوم جهت ذخیره سایز داده و اطلاعات مرزی با دقت بالاتری باشد [۱۸].

PINSVR خطی:

فرض کنید مجموعه مشاهدات (شرکت-سال) $\boldsymbol{N}$تایی به همراه بردار ورودی (متغیرهای مستقل) $\boldsymbol{x}$ موجود باشد که تمام این شرکت- سال‌ها توسط یک ماتریس داده $\boldsymbol{X}$ نشان داده شود، به‌طوری‌که $\boldsymbol{n}$امین ردیف آن را با $\boldsymbol{x^T}$ نشان دهیم و بیانگر متغیرهای مستقل شرکت $\boldsymbol{n}$ام باشد و $\boldsymbol{n = 1, 2, \dots, N}$ باشد و $\boldsymbol{y}$ نشان‌دهنده متغیر وابسته یعنی قیمت سهام باشد. الگوریتم **PINSVR** به دنبال یافتن توابع خطی

[1] K-nearest neighbor

پروکسیمال غیر موازی $\boldsymbol{f_1(x)}$ و $\boldsymbol{f_2(x)}$ به‌طور هم‌زمان و همچنین دو تابع خطی پروکسیمال غیر موازی متفاوت $\boldsymbol{g_1(x)}$ و $\boldsymbol{g_2(x)}$ است. این توابع در زیر نشان داده شده‌اند.

$$f_1(x) = w_1^T x + b_1 \text{ و } f_2(x) = w_2^T x + b_2 \quad (۱)$$

$$g_1(x) = w_3^T x + b_3 \text{ و } g_{2(x)} = w_4^T x + b_4 \quad (۲)$$

که در آن $\boldsymbol{g_1(x) \geq 0}$ و $\boldsymbol{g_2(x) \geq 0}$ است. دو تابع ضرر غیرحساس-پارامتریک یک‌طرفه[1] به‌صورت زیر تعریف می‌شود:

$$\boldsymbol{L^{g_1}(x, y, f_1) = \sum_{i=1}^{m} max\{0, -(y_i - f_1(x_i) + g_1(x_i))\}}$$

$$\boldsymbol{L^{g_2}(x, y, f_2) = \sum_{i=1}^{m} max\{0, -(f_2(x_i) - y_i + g_2(x_i))\}}$$

بنابراین، ریسک تجربی در این الگوریتم به‌صورت زیر تعریف می‌شود:

$$R_{emp}^{g_1}[f_1] = \sum_{i=1}^{m} max\left\{0, \left(y_i - f_1(x_i)\right)^2\right\} + c_1 \sum_{i=1}^{m} max\{0, -(y_i - f_1(x_i) + g_1(x_i))\} \quad (۳)$$

$$R_{emp}^{g_2}[f_2] = \sum_{i=1}^{m} max\{0, (f_2(x_i) - y_i)^2\} + c_2 \sum_{i=1}^{m} max\{0, -(f_2(x_i) - y_i + g_2(x_i))\} \quad (۴)$$

که در آن، $\boldsymbol{c_1 > 0}$ و $\boldsymbol{c_2 > 0}$ پارامترهای آن هستند. مسئله‌های بهینه‌سازی اولیه **PINSVR** به‌صورت زیر تعریف می‌شود:

$$\min_{w_1, w_3, b_1, b_3, \xi} \frac{1}{2} c_3 (w_1^T w_1 + b_1^2 + w_3^T w_3 + b_3^2) + \frac{1}{2} \xi^{*T} \xi^* + c_1 e^T \xi \quad (۵)$$

$$s.t. \begin{cases} Y - (Aw_1 + eb_1) = \xi^* \\ Aw_3 + eb_3 \geq 0 \\ Y - (Aw_1 + eb_1) \geq -(Aw_3 + eb_3) - \xi, \xi \geq 0 \end{cases}$$

[1] One-side parametric-insensitive loss function

$$\min_{w_2,w_4,b_2,b_4,\eta} \frac{1}{2}c_4(w_2^T w_2 + b_2^2 + w_4^T w_4 + b_4^2) + \frac{1}{2}\eta^{*T}\eta^* + c_2 e^T \eta \quad (۶)$$

$$s.t.\begin{cases} (Aw_2 + eb_2) - Y = \eta^* \\ Aw_4 + eb_4 \geq 0 \\ (Aw_2 + eb_2) - Y \geq -(Aw_4 + eb_4) - \eta, \eta \geq 0 \end{cases}$$

که در آن $\boldsymbol{c_1}$، $\boldsymbol{c_2}$، $\boldsymbol{c_3}$ و $\boldsymbol{c_4}$ پارامترهای ورودی مسئله هستند.

با استفاده از تابع لاگرانژ و بررسی شرایط **KKT**[1] تابع دوگان دو مسئله بهینه‌سازی و به‌صورت مسئله دوگان زیر به دست می‌آید.

$$\min_{\alpha,\beta} \frac{1}{2c_3}\alpha^T GG^T\alpha + \frac{1}{2}\beta^T\left(G(G^TG + c_3 I_1)^{-1}G^T + \frac{1}{c_3}G^TG\right)\beta + \frac{1}{c_3}\alpha^T GG^T\beta - Y^T(G(G^TG + c_3 I_1)^{-1}G^T - I_2)\beta \quad (۷)$$

$$s.t.\begin{cases} \alpha \geq 0 \\ 0 \leq \beta \leq c_1 e \end{cases}$$

$$\min_{\alpha^*,\beta^*} \frac{1}{2c_4}\alpha^{*T} GG^T\alpha^* + \frac{1}{2}\beta^{*T}\left(G(G^TG + c_4 I_1)^{-1}G^T + \frac{1}{c_4}G^TG\right)\beta^* + \frac{1}{c_4}\alpha^{*T} GG^T\beta^* - Y^T(G(G^TG + c_4 I_1)^{-1}G^T - I_2)\beta^* \quad (۸)$$

$$s.t.\begin{cases} \alpha^* \geq 0 \\ 0 \leq \beta^* \leq c_2 e \end{cases}$$

[1] Karush–kuhn–tucker conditions

که در آن $\boldsymbol{G} = [\boldsymbol{A} \quad \boldsymbol{e}]$ است. از حل رابطه (۷) ضرایب لاگرانژ $\boldsymbol{\alpha}$ و $\boldsymbol{\beta}$ به دست می‌آید و از قرار دادن آن در روابط زیر مقدار پارامترهای خط $\boldsymbol{f_1(x)}$ و $\boldsymbol{g_1(x)}$ به دست می‌آید:

$$u_1 = (G^T G + c_3 I_1)^{-1} G^T (Y - \beta) \quad (۹)$$

$$u_3 = \frac{1}{c_3} G^T (\alpha + \beta) \quad (۱۰)$$

که در آن $\boldsymbol{u_1} = \left[\boldsymbol{w_1^T b_1^T}\right]^{\boldsymbol{T}}$ و $\boldsymbol{u_3} = \left[\boldsymbol{w_3^T b_3^T}\right]^{\boldsymbol{T}}$ است. از حل بهینه‌سازی (۸) ضرایب لاگرانژ $\boldsymbol{\alpha}^*$ و $\boldsymbol{\beta}^*$ به دست می‌آید و از قرار دادن آن در روابط زیر مقدار پارامترهای خط $\boldsymbol{f_2(x)}$ و $\boldsymbol{g_2(x)}$ به دست می‌آید:

$$u_2 = (G^T G + c_4 I_1)^{-1} G^T (Y - \beta^*) \quad (۱۱)$$

$$u_4 = \frac{1}{c_4} G^T (\alpha^* + \beta^*) \quad (۱۲)$$

حال تابع تصمیم خطی این الگوریتم به‌صورت زیر به دست می‌آید:

$$f(x) = \frac{1}{2}\left(f_1(x) + f_2(x)\right) = \frac{1}{2}(w_1 + w_2)^T x + \frac{1}{2}(b_1 + b_2) \quad (۱۳)$$

کران بالا و پایین مدل رگرسیون به‌صورت زیر محاسبه می‌شود:

$$f_1(x) - g_1(x) = (w_1 - w_3)^T x + b_1 - b_3 \quad (۱۴)$$

$$f_2(x) + g_2(x) = (w_2 + w_4)^T x + b_2 + b_4 \quad (۱۵)$$

PINSVR کرنلی:

در این بخش **PINSVR** خطی برای حالت غیرخطی با استفاده از حقه کرنل بسط داده می‌شود. داده ورودی به فضای ویژگی با ابعاد بالا با استفاده از توابع کرنل غیرخطی نگاشت داده می‌شود. در فضای ویژگی، یک تابع رگرسیون خطی متناظر با تابع رگرسیونی غیرخطی در فضای ورودی است. مشابه حالت خطی، توابع غیرخطی پروکسیمال غیرموازی$\boldsymbol{f_1(x)}$و $\boldsymbol{f_2(x)}$ و دو تابع مختلف غیرخطی پروکسیمال غیرموازی$\boldsymbol{g_1(x)}$ و $\boldsymbol{g_2(x)}$ به‌صورت زیر در نظر گرفته‌شده است.

$$f_1(x) = k(x^T, A^T)w_1 + b_1 \text{و} f_2(x) = k(x^T, A^T)w_2 + b_2 \quad (۱۶)$$

$$g_1(x) = k(x^T, A^T)w_3 + b_3 \text{ و } g_{2(x)} = k(x^T, A^T)w_4 + b_4 \quad (۱۷)$$

که در آن $\boldsymbol{k}$ تابع کرنل است و $\boldsymbol{g_1(x) \geq 0}$ و $\boldsymbol{g_2(x) \geq 0}$ است. مسئله اولیه **PINSVR** غیرخطی به‌صورت زیر تعریف می‌شود:

$$\min_{w_1,w_3,b_1,b_3,\xi} \frac{1}{2}c_3(w_1^T w_1 + b_1^2 + w_3^T w_3 + b_3^2) + \frac{1}{2}\xi^{*T}\xi^* + c_1 e^T \xi$$

$$s.t. \begin{cases} Y - (k(A, A^T)w_1 + eb_1) = \xi^* \\ k(A, A^T)w_3 + eb_3 \geq 0 \\ Y - (k(A, A^T)w_1 + eb_1) \geq -(k(A, A^T)w_3 + eb_3) - \xi, \xi \geq 0 \end{cases} \quad (۱۸)$$

$$\min_{w_2,w_4,b_2,b_4,\eta} \frac{1}{2}c_4(w_2^T w_2 + b_2^2 + w_4^T w_4 + b_4^2) + \frac{1}{2}\eta^{*T}\eta^* + c_2 e^T \eta$$

$$s.t. \begin{cases} (k(A, A^T)w_2 + eb_2) - Y = \eta^* \\ k(A, A^T)w_4 + eb_4 \geq 0 \\ (k(A, A^T)w_2 + eb_2) - Y \geq -(k(A, A^T)w_4 + eb_4) - \eta, \eta \geq 0 \end{cases} \quad (۱۹)$$

که در آن $\boldsymbol{c_1}$، $\boldsymbol{c_2}$، $\boldsymbol{c_3}$ و $\boldsymbol{c_4}$ پارامترهای ورودی مسئله هستند. با استفاده از تابع لاگرانژ و بررسی شرایط **KKT** تابع دوگان دو مسئله بهینه‌سازی رابطه (۱۸) و (۱۹) به‌صورت مسئله دوگان (یانگ[1] و همکاران، ۲۰۱۴) زیر به دست می‌آید.

[1] Yang

$$\min_{\alpha,\beta} \frac{1}{2c_3}\alpha^T HH^T\alpha + \frac{1}{2}\beta^T\left(H(H^TH + c_3I_1)^{-1}H^T + \frac{1}{c_3}H^TH\right)\beta + \frac{1}{c_3}\alpha^T HH^T\beta - Y^T(H(H^TH + c_3I_1)^{-1}H^T - I_2)\beta \quad (۲۰)$$

$$s.t.\begin{cases} \alpha \geq 0 \\ 0 \leq \beta \leq c_1 e \end{cases}$$

$$\min_{\alpha^*,\beta^*} \frac{1}{2c_4}\alpha^{*T} HH^T\alpha^* + \frac{1}{2}\beta^{*T}\left(H(H^TH + c_4I_1)^{-1}H^T + \frac{1}{c_4}H^TH\right)\beta^* + \frac{1}{c_4}\alpha^{*T} HH^T\beta^* - Y^T(H(H^TH + c_4I_1)^{-1}H^T - I_2)\beta^* \quad (۲۱)$$

$$s.t.\begin{cases} \alpha^* \geq 0 \\ 0 \leq \beta^* \leq c_2 e \end{cases}$$

که در آن $\boldsymbol{H} = \left[\boldsymbol{k}(\boldsymbol{A}, \boldsymbol{A^T}) \quad \boldsymbol{e}\right]$ است. از حل رابطه (۲۱) ضرایب لاگرانژ $\boldsymbol{\alpha}$ و $\boldsymbol{\beta}$ به دست می‌آید و از قرار دادن آن در روابط زیر مقدار پارامترهای خط $\boldsymbol{f_1(x)}$ و $\boldsymbol{g_1(x)}$ به دست می‌آید:

$$u_1 = (H^TH + c_3I_1)^{-1}H^T(Y - \beta) \quad (۲۲)$$

$$u_3 = \frac{1}{c_3}H^T(\alpha + \beta) \quad (۲۳)$$

که در آن $\boldsymbol{u_1} = \left[\boldsymbol{w_1^T b_1^T}\right]^{\boldsymbol{T}}$ و $\boldsymbol{u_3} = \left[\boldsymbol{w_3^T b_3^T}\right]^{\boldsymbol{T}}$ است. از حل بهینه‌سازی لاگرانژ $\boldsymbol{\alpha}^*$ و $\boldsymbol{\beta}^*$ به دست می‌آید و از قرار دادن آن در روابط زیر مقدار پارامترهای خط $\boldsymbol{f_2(x)}$ و $\boldsymbol{g_2(x)}$ به دست می‌آید:

$$u_2 = (H^TH + c_4I_1)^{-1}H^T(Y - \beta^*) \quad (۲۴)$$

$$u_4 = \frac{1}{c_4}H^T(\alpha^* + \beta^*) \quad (۲۵)$$

حال تابع تصمیم خطی این الگوریتم به‌صورت زیر به دست می‌آید:

$$f(x) = \frac{1}{2}\left(f_1(x) + f_2(x)\right) = \frac{1}{2}(w_1 + w_2)^T k(A, x) + \frac{1}{2}(b_1 + b_2) \quad (۲۶)$$

کران بالا و پایین مدل رگرسیون به‌صورت زیر محاسبه می‌شود:

$$f_1(x) - g_1(x) = (w_1 - w_3)^T k(A, x) + b_1 - b_3 \quad (۲۷)$$

$$f_2(x) + g_2(x) = (w_2 + w_4)^T k(A, x) + b_2 + b_4 \quad (۲۸)$$

درخت تصمیم طبقه‌بندی و رگرسیون – CART:

درخت تصمیم یکی از پرکاربردترین و سودمندترین روش‌هایی است که برای استنتاج استقرایی[1] مورد استفاده قرار می‌گیرد. درخت تصمیم، یک روش غیرپارامتریک با ساختار سلسله مراتبی داده‌ها و یادگیری با نظارت است که با استفاده از استراتژی تقسیم و حل پیاده‌سازی می‌گردد. دراین روش تابع یاد گرفته‌شده به‌صورت یک درخت تصمیم نمایش داده می‌شود که گاهی اوقات برای افزایش درجه خوانایی آن برای انسان، درخت را به‌صورت مجموعه‌ای از قوانین اگر-آنگاه[2] درمی‌آورند. این الگوریتم‌ها بر اساس داده‌های مجموعه یادگیری، یک درخت ایجاد می‌کنند که در این درخت هر گره داخلی یک آزمون بر روی یک صفت را نشان می‌دهد، هر شاخه نتیجه‌ای از ارزیابی را نشان می‌دهد و هر برگ برچسب یک کلاس را نگهداری می‌کند.

درخت تصمیم **CART**[3] از ضریب جینی[4] استفاده می‌کند. رابطه ضریب جینی به‌صورت زیر تعریف می‌گردد.

$$I_{gini} = 1 - \sum_j p(c_j)^2 \quad (۲۹)$$

که در آن $p(c_j)$ نسبت داده‌های متعلق به‌کلاس C را نشان می‌دهد. این الگوریتم ابتدا برای تمامی ویژگی‌های داده‌های اولیه ضریب جینی را با استفاده از رابطه (۲۹) محاسبه می‌کند. سپس، مقدار سودمندی اطلاعات[5] هر یک از ویژگی‌ها از رابطه زیر به دست می‌آید.

$$Gain(A) = I_{gini} - I_{res_{gini}}(A) \quad (۳۰)$$

[1] Inductive inference

[2] If-then

[3] Classification and regression tree

[4] Gini index

[5] Information gain

که $I_{res_{gini}}(A)$ از رابطه (۳۱) محاسبه می‌گردد که در آن $I_{res_{gini}}$ میزان بی‌نظمی باقیمانده در دسته‌ها به‌واسطه استفاده از ویژگی A است که با کمک مجموع احتمال وقوع هر یک از تقسیمات قابل حصول است. سپس ویژگی F که دارای بیشترین سودمندی است، به‌عنوان ریشه[1] ویژگی جداساز[2] انتخاب می‌شود.

$$I_{res_{gini}}(A)=\sum_j\left(p(a)\times\left(1-\sum_j p(c_j\mid a)^2\right)\right) \quad (۳۱)$$

که در آن a بیانگر زیرشاخه به وجود آمده با انتخاب ویژگی A ام به‌عنوان ویژگی جداساز است.

اما چیزی که باید نگران آن باشیم، اتفاق افتادن پدیده‌ای به نام بیش برازش[3] است. به همین علت برای بررسی عمومیت[4] مدل‌های ارائه‌شده، میزان خطا برای پیش‌بینی متغیر وابسته قیمت سهام در هر سه سال برای شرکت-سال‌های تست (شرکت-سال‌هایی که توسط روش **10-Fold Cross-Validation** در هر تکرار کنار گذاشته شده‌اند و الگوریتم آن‌ها را تاکنون ندیده است) به دست آورده شده است. به ازای هر معیار خطا، ۱۰ خطا که هرکدام توسط روش **10-Fold Cross-Validation** گزارش‌شده‌اند، به دست می‌آید که این خطا به همراه میانگین برای هرسال در جدول ۳ نشان داده‌شده است. نتیجه گرفته می‌شود که مدل‌های به‌دست‌آمده دارای عمومیت هستند، یعنی برای شرکت-سال‌هایی که تابه‌حال ندیده‌اند، هم خوب عمل می‌کنند و همچنین مشکل بیش برازش هم اتفاق نیفتاده است.

جدول ۳: میانگین معیارهای خطا برای ارزیابی میزان آموزش الگوریتم‌ها در سال جاری و آتی

۱) الگوریتم	۲) Linear PINSVR	۳) Non - Linear PINSVR	۴) CART
۵) سال جاری			
۶) **Fold**	۷) MAE	۸) MAE	۹) MAE
۱۰) ۱	۱۱) ۹/۶۶	۱۲) ۲/۵۹	۱۳) ۴/۳۶
۱۴) ۲	۱۵) ۹/۷۷	۱۶) ۲/۶۴	۱۷) ۴/۴۳
۱۸) ۳	۱۹) ۹/۸۶	۲۰) ۲/۶۷	۲۱) ۴/۴۳
۲۲) ۴	۲۳) ۹/۱۵	۲۴) ۲/۵۹	۲۵) ۴/۳۶
۲۶) ۵	۲۷) ۹/۶۳	۲۸) ۲/۶۵	۲۹) ۴/۲۹
۳۰) ۶	۳۱) ۹/۴۷	۳۲) ۲/۵۴	۳۳) ۴/۳۳
۳۴) ۷	۳۵) ۹/۹۶	۳۶) ۲/۷۰	۳۷) ۴/۵۳

[1] Root

[2] Splitting feature

[3] overfitting

[4] generalitty

۸ (۳۸	۹/۸۳ (۳۹	۲/۶۴ (۴۰	۴/۴۱ (۴۱
۹ (۴۲	۹/۹۶ (۴۳	۲/۶۳ (۴۴	۴/۴۸ (۴۵
۱۰ (۴۶	۹/۹۱ (۴۷	۲/۶۵ (۴۸	۴/۳۵ (۴۹
۵۰) میانگین	۹/۷۸ (۵۱	۲/۶۳ (۵۲	۴/۳۹ (۵۳
۵۴) الگوریتم	**۵۵)** Linear PINSVR	**۵۶)** Non - Linear PINSVR	**۵۷)** CART
۵۸) سال آتی			
Fold (۵۹	**۶۰)** MAE	**۶۱)** MAE	**۶۲)** MAE
۱ (۶۳	۹/۵۶ (۶۴	۲/۷۴ (۶۵	۴/۶۱ (۶۶
۲ (۶۷	۹/۸۰ (۶۸	۲/۷۱ (۶۹	۴/۸۰ (۷۰
۳ (۷۱	۹/۷۸ (۷۲	۲/۶۸ (۷۳	۴/۷۶ (۷۴
۴ (۷۵	۹/۹۳ (۷۶	۲/۷۱ (۷۷	۴/۸۳ (۷۸
۵ (۷۹	۹/۷۸ (۸۰	۲/۸۱ (۸۱	۴/۸۰ (۸۲
۶ (۸۳	۹/۸۶ (۸۴	۲/۷۷ (۸۵	۴/۸۱ (۸۶
۷ (۸۷	۹/۹۰ (۸۸	۲/۷۹ (۸۹	۴/۷۰ (۹۰
۸ (۹۱	۹/۳۷ (۹۲	۲/۷۲ (۹۳	۴/۴۶ (۹۴
۹ (۹۵	۹/۷۶ (۹۶	۲/۷۴ (۹۷	۴/۷۷ (۹۸
۱۰ (۹۹	۹/۸۵ (۱۰۰	۲/۶۳ (۱۰۱	۴/۷۳ (۱۰۲
۱۰۳) میانگین	۹/۷۵ (۱۰۴	۲/۷۳ (۱۰۵	۴/۷۳ (۱۰۶

جدول ۴ : میانگین معیارهای خطا برای ارزیابی کارایی با داده‌های تست در سال جاری و آتی

Article XV.	الگوری تم	Article XVI.	Linear PINSVR	Article XVII.	Non - Linear PINSVR
Article XIX.	سال جاری				
Article XX.	Fold	Article XXI.	MAE	Article XXII.	MAE
Article XXIV.	۱	Article XXV.	۱۰/۸۴	Article XXVI.	۲/۹۷
Article XXVIII.	۲	Article XXIX.	۹/۸۶	Article XXX.	۲/۵۶
Article XXXII.	۳	Article XXXIII.	۹/۰۸	Article XXXIV.	۲/۲۴
Article XXXVI.	۴	Article XXXVII.	۱۰/۰۲	Article XXXVIII.	۳/۰۱
Article XL.	۵	Article XLI.	۱۱/۱۳	Article XLII.	۲/۴۱
Article XLIV.	۶	Article XLV.	۱۲/۵۶	Article XLVI.	۳/۴۴
Article XLVIII.	۷	Article XLIX.	۸/۱۵	Article L.	۴/۰۴
Article LII.	۸	Article LIII.	۱۰/۳۷	Article LIV.	۲/۵۵
Article LVI.	۹	Article LVII.	۸/۱۷	Article LVIII.	۲/۶۶
Article LX.	۱۰	Article LXI.	۸/۶۳	Article LXII.	۲/۴۱
Article LXIV.	میانگین	Article LXV.	۹/۸۸	Article LXVI.	۲/۸۲

CART	Article XVIII.
MAE	Article XXIII.
۴/۷۳	Article XXVII.
۴/۱۵	Article XXXI.
۴/۰۸	Article XXXV.
۴/۷۶	Article XXXIX.
۵/۳۳	Article XLIII.
۵/۰۱	Article XLVII.
۳/۱۸	Article LI.
۴/۳۰	Article LV.
۳/۹۷	Article LIX.
۴/۸۰	Article LXIII.
۴/۴۳	Article LXVII.

Article LXVIII.	الگوری تم	Article LXIX.	Linea r PINSVR	Article LXX.	Non - Linear PINSVR	Article LXXI.	CART
Article LXXII.	سال آتی						
Article LXXIII.	Fold	Article LXXIV.	MAE	Article LXXV.	MA E	Article LXXVI.	MAE
Article LXXVII.	۱	Article LXXVIII.	۹/۹۱	Article LXXIX.	۲/۶۶	Article LXXX.	۴/۷۶
Article LXXXI.	۲	Article LXXXII.	۱۰/۳۸	Article LXXXIII.	۲/۹۱	Article LXXXIV.	۴/۰۷
Article LXXXV.	۳	Article LXXXVI.	۹/۵۲	Article LXXXVII.	۳/۱۹	Article LXXXVIII.	/۸۵ ۴
Article LXXXIX.	۴	Article XC.	۹/۱۹	Article XCI.	۲/۸۹	Article XCII.	۴/۸۲
Article XCIII.	۵	Article XCIV.	۱۰/۱۷	Article XCV.	۱/۹۸	Article XCVI.	۵/۰۶
Article XCVII.	۶	Article XCVIII.	۱۰/۱۳	Article XCIX.	۲/۳۳	Article C.	۵/۹۶
Article CI.	۷	Article CII.	۹/۴۳	Article CIII.	۴/۱۸	Article CIV.	۴/۹۸
Article CV.	۸	Article CVI.	۱۰/۹۲	Article CVII.	۲/۸۴	Article CVIII.	۴/۳۵
Article CIX.	۹	Article CX.	۹/۹۵	Article CXI.	۲/۶۸	Article CXII.	۴/۸۹
Article CXIII.	۱۰	Article CXIV.	۹/۹۸	Article CXV.	۳/۶۴	Article CXVI.	۴/۸۷
Article CXVII.	میانگین	Article CXVIII.	۹/۹۵	Article CXIX.	۲/۹۳	Article CXX.	۴/۸۶

پس از تقسیم شرکت-سال‌ها به دودسته داده‌های یادگیری و تست با استفاده از روش **10-Fold Cross-Validation** برای ارزیابی مدل‌های رگرسیون خطی و غیرخطی و قانون‌گرا از معیار ارزیابی میانگین قدر مطلق خطا[1] (**MAE**) استفاده شده است که با استفاده از رابطه (۳۲) محاسبه می‌گردد:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - d_i| \qquad (۳۲)$$

که در روابط بالا y_i و d_i به ترتیب متغیر وابسته واقعی و متغیر وابسته پیش‌بینی شده توسط الگوریتم‌های پژوهش برای شرکت-سال **i** ام است و n تعداد شرکت-سال‌ها (در مرحله آموزش یا مرحله تست) و $\overline{y}$ و $\overline{d}$ میانگین متغیر وابسته واقعی و پیش‌بینی شده را به ترتیب نشان می‌دهد.

نتیجه‌گیری

یکی از الزامات برای سرمایه‌گذاران در بازار سهام و اوراق بهادار، داشتن اطلاعات کافی در رابطه با روش مطلوب خرید و فروش سهام است. در این راستا مهم‌ترین نکته انتخاب سهام مناسب (پرتفوی) برای خرید و فروش و انتخاب بهترین روش برای اطمینان از بیشترین بازده سهام است. فقدان دانش مناسب در خصوص روش‌های مطلوب در بازار سهام مهم‌ترین نگرانی در بین سرمایه‌گذاران در بازار سهام می‌باشد. در این راستا استفاده از تکنیک‌های علمی اثبات شده به دو هدف اساسی در بورس کمک می‌نماید. اولاً بکارگیری این تکنیک‌ها سبب افزایش میزان بهره‌وری و سودآوری سرمایه‌گذاران می‌گردد و دوماً این امر خود در بهبود عملکرد بازار سرمایه و سوق دادن بازار به سمت کارائی، مؤثر واقع می‌گردد [۳].

[1] Mean absolute error

هدف از این پژوهش پیش‌بینی بازده سهام در شرکت‌های پذیرفته‌شده بورس اوراق بهادار تهران با تاکید بر نقش معیارهای مالی و نظارتی با استفاده از روش‌های یادگیری ماشین (رویکرد خطی و غیرخطی و قانون‌گرا) می باشد. نتایج اولیه حاصل از تحلیل عوامل موثر بر بازده سهام از بین معیارهای مالی و نظارتی با روش متغیرگزینی **Relief-F** نشان داد که در مجموع معیارهای مالی جهت تبیین بازده سهام نسبت به معیارهای نظارتی توانایی و کاربرد بالاتری دارند. شاید دلیل این نتایج را بتوان بر نوپا بودن قوانین و مقررات مربوط به کمیته حسابرسی و همچنین ناکارآمد بودن قوانین حاکمیت شرکتی به‌عنوان معیارهای نظارتی دانست زیرا که نتایج نشان داد که از بین معیارهای مالی و نظارتی متغیرهای بازده دارایی، نسبت قیمت به سود هر سهم،کیوتوبین، نسبت وجه نقد عملیاتی، اهرم مالی و نسبت مالکان نهادی بر بازده سهام تأثیرگذار است که این نتایج با پژوهش‌های انصاری و همکاران [۱]؛ عسگرنژاد نوری [۱۲] و طبیبی [۱۱] مطابقت دارد. روش‌های یادگیری ماشین قدرت مناسبی جهت پیش‌بینی بازده سهام خواهند داشت که نتایج این پژوهش نیز تاییدی بر این فرضیه‌ها می‌باشد، هرچند که روش غیرخطی **PINSVR** و قانون‌گرای **CART** نسبت به روش خطی قدرت پیش‌بینی بالاتری را نشان داد. این نتایج با پژوهش‌های رجبی و همکاران [۸]، راعی و چاوشی [۷]، علی‌محمدی و همکاران [۱۳] و مدرس، کهنسال [۱۷] نیز مطابقت دارد.

نتایج اولیه این پژوهش نشان می‌دهد متغیرهای نسبت قیمت به سود هر سهم، نسبت سود تقسیمی، اندازه شرکت، ارزش شرکت و ریسک سیستماتیک دارای بیشترین اهمیت در پیش‌بینی بازده سهام می‌باشند لذا به مدیران شرکت‌های فعال در بازار سرمایه توصیه می‌شود متغیرهای ذکرشده را جهت تصمیم‌گیری در زمینه برآورد بازده سهام شرکت‌ها مدنظر قرار دهند.

نتایج ثانویه پژوهش نیز نشان می‌دهد الگوریتم غیرخطی و قانون‌گرا جهت پیش‌بینی بازده سهام شرکت‌ها نسبت به الگوریتم خطی توانایی بالاتری دارد لذا به صاحبان سرمایه و تصمیم‌گیران شرکت‌ها توصیه می‌شود که در تصمیم‌گیری‌های خود پیرامون سرمایه‌گذاری در بازار سرمایه از قدرت پیش‌بینی الگوریتم‌های هوش مصنوعی به‌ویژه روش الگوریتم غیرخطی استفاده کنند.

با توجه به آن‌چه که ذکر شد، به‌نظر می‌رسد نتایج این پژوهش بتواند به صورت کاربردی مورد توجه مدیران بازار سرمایه ایران قرارگیرد به‌طوری‌که با پیش‌بینی بازده سهام در شرکت‌ها و بررسی عوامل مؤثر بر آن، نسبت به مدیریت کردن سرمایه سهامداران، کاهش ریسک بحران‌های مالی و کمک به سرمایه‌گذاران جهت اجتناب از زیان‌های بزرگ در بازار سهام، اقدام نمایند.

فهرست منابع

۱. انصاری، عبدالمهدی؛ یوسف‌زاده، نسرین و زارع، زهرا، (۱۳۹۴)." بررسی هم زمان عوامل موثر بر ساختار سرمایه و بازده سهام در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران". **فصلنامه تحقیقات حسابداری و حسابرسی**، انجمن حسابداری ایران، شماره ۲۵.

۲. بهمنی‌راد، مسعود، (۱۳۹۰). "رابطه عوامل حاکمیت شرکتی با ارزش و بازده سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران"، پایان نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.

۳. پورزمانی، زهرا؛ محمدی، محمدرضا، (۱۳۹۱)." مقایسه راهبردهای خرید و فروش سهام جهت محاسبه بازده سهام در سرمایه‌گذاری‌های کوتاه مدت و بلندمدت". **فصلنامه علمی پژوهشی دانش مالی تحلیل اوراق بهادار**، دوره ۵، شماره۱۴، صص ۱-۱۱.

۴. جهانخانی، علی و پارسائیان، علی، (۱۳۷۶). **مدیریت سرمایه‌گذاری و ارزیابی اوراق بهادار**. انتشارات دانشکده مدیریت دانشگاه تهران، چاپ اول.

۵. جهانگیرنیا، حسین؛ شیرمحمدپورگرجانی، عبدا... و خیراندیش، صادق، (۱۳۹۶)."بررسی رابطه بین سه متغیر سودآوری، بازده دارایی و بازده حقوق صاحبان سهام هلدینگ‌های اقتصادی با بازده شرکت‌های فرعی (مطالعه موردی هلدینگ شستا)". **فصلنامه تحقیقات حسابداری و حسابرسی**، انجمن حسابداری ایران، شماره ۳۳.

۶. حساس یگانه، یحیی و مولودی، عبدالله. (۱۳۹۰)." رابطه حاکمیت شرکتی و ارزش ایجاد شده برای سهامداران". **فصلنامه علمی ـ پژوهشی مطالعات مدیریت صنعتی**، دوره ۹، شماره ۲۳، صص ۲۳۳-۲۶۱.

۷. راعی، رضا و چاوشی، کاظم، (۱۳۸۲)."پیش‌بینی بازده سهام در بورس اوراق بهادار تهران: مدل شبکه‌های عصبی مصنوعی و مدل چند عاملی". **نشریه تحقیقات مالی**، دوره ۵، شماره ۱، صص ۹۷-۱۲۰.

۸. رجبی، راضیه؛ عسگری، محمدرضا و دهقان، عبدالمجید، (۱۳۹۸)."پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران با استفاده از تکنیک‌های داده کاوی (تحلیل تفکیکی)". **فصلنامه پژوهش‌های جدید در مدیریت و حسابداری**، دوره ۵، شماره ۵۴، صص ۷۳ -۸۸.

۹. رحمانی، علی و سعیدی، فرشته، (۱۳۹۰)."ارزیابی عملکرد مدل‌های لاجیت در پیش‌بینی بازده سهام". **علوم و فناوری اطلاعات**، شماره۳، صص ۲۲-۳۵.

۱۰. رهنمای رودپشتی، فریدون و محسنی، عبدالرضا، (۱۳۹۷)."ارتباطات سیاسی، سود نقدی و بازده سهام در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران". **دانش مالی تحلیل اوراق بهادار**، سال ۱۱، شماره ۳۸، صص ۱۲۹-۱۴۴.

۱۱. طبیبی، صدیقه، (۱۳۹۶). "تأثیر حاکمیت شرکتی بر نوسانات بازده سهام در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران". پایان نامه کارشناسی ارشد، دانشگاه پیام نور استان گیلان، مرکز پیام نور رشت.

۱۲. عسگرنژاد نوری، باقر، (۱۳۹۷)."عوامل مؤثر بر بازدهی سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران: رویکرد فراتحلیل". **نشریه مدیریت دارایی و تأمین مالی**، سال ششم، دوره ۲۰، شماره ۱، صص ۲۹-۵۰.

۱۳. علی‌محمدی، علی‌محمد؛ عباسی‌مهر، محمد حسین و جواهری، احمد، (۱۳۹۴)."پیش‌بینی بازده سهام شرکت‌ها با استفاده از نسبت‌های مالی تحت رویکرد درخت تصمیم". **راهبرد مدیریت مالی**، دوره ۳، شماره ۴، صص ۱۲۵-۱۴۶.

۱۴. عمادزاده، محمدکاظم؛ زارعی، فاطمه و طوروسیان، آرینه، (۱۳۹۰). "شاخص‌های خرد و کلان مؤثر بر بازده سهام". **مجله اقتصادی**، شماره ۳، صص ۳۱-۴۴.

۱۵. کیان، علیرضا ، فقیه، محسن، (۱۳۹۷)."تأثیر ویژگی‌های کمیته حسابرسی بر ریسک شرکت". **دانش حسابداری**، سال نهم، شماره ۳۴.

۱۶. محمدی، ناهید، (۱۳۸۷). رابطه اطلاعات ترازنامه‌ای و بازده آتی سهام. پایان‌نامه کارشناسی ارشد، دانشکده مدیریت و حسابداری علامه طباطبایی.

۱۷. مدرس، احمد، کهنسال، محمدباقر، (۱۳۹۲)."ساختارهای خطی و غیرخطی در پیش‌بینی بازده سهام". **پژوهش حسابداری**، شماره ۱۱، صص ۱-۱۴.

18. Alpaydin, E. (2014). **Introduction to machine learning**, Bogazici University, Istanbul, The MIT Press.
19. Balcilar, M., Gupta, R., Kim, Won. J., Kyei, C. (2019). "The role of economic policy uncertainties in predicting stock returns and their volatility for Hong Kong, Malaysia and South Korea". **International Review of Economics & Finance**, **Elsevier**, vol.59:150-163.
20. Bohl, M. T., Brzeszczy´nski, J., & Wilfling, B. (2009). "Institutional investors and stock returns volatility: Empirical evidence from a natural experiment". **Journal of Financial Stability**, Vol.5(2):170-182.
21. Hudson, C.D., Jahera, J.S. & Lloyd, W.P. (1992). "Further evidence on the relationship between ownership and performance". **The Financial Review**, vol.27: 227–239.
22. Oswald, S.L. & Jahera, J.S. (1991). "The influence of ownership on erformance: An empirical study". **Strategic Management Journal**, vol.12: 321–326.
23. Zsolt, B., Barnabas, A. (2007)."The impact of ownership concentration, and identity on company performance in the US and in Central and Eastern Europe". **Baltic Journal of Management**, Vol. 2(2):125-139.

Accounting & Auditing Studies, Spring 2023, No. 45. pp. 125-146

Predicting Stock Returns with Emphasis on the Role of Financial and Regulatory Criteria Using Machine Learning Methods

Ali Jafari (PhD) [1]©

Assistant professor of Accounting, Islamic Azad University, Qaemshahr Branch, Iran

Mostafa Mansourikhah

PhD candidate of Accounting, Islamic Azad University, Qaemshahr Branch, Iran

Abbas Ali Pour Aghajan (PhD)

Assistant professor of Accounting, Islamic Azad University, Qaemshahr Branch, Iran

(Received: May 9, 2022; Accepted: November 7, 2022)

Introduction: Predicting stock returns, although complex, has always been of interest to investors. Investors and decision makers need information to predict stock returns. **Aim:** The aim of this study was to predict stock returns in companies listed on the Tehran Stock Exchange by emphasizing the role of financial and regulatory criteria using machine learning methods. **Method:** This research is applied in terms of purpose and in terms of the type of field-library study. The statistical population consisted of companies listed on the Tehran Stock Exchange during the years 1390 to 1396 (a seven-year period) and the hypotheses were tested using machine learning methods (linear, nonlinear and cart). **Results:** The results of variable selection test using Relief-F method indicate a significant relationship between return on assets, price-to-earnings-per-share ratio, quota, operating cash ratio, financial leverage and institutional owners' ratio to stock returns. The results also showed that PINSVR machine learning methods in linear and nonlinear mode and CART have a good ability (more than 90%) to predict stock returns, but the nonlinear PINSVR method and CART showed higher predictive power than the linear method.
Conclusion: According to the results, capital owners and company decision makers are advised to use the predictive power of artificial intelligence algorithms, especially nonlinear methods, in their decisions about investing in the Tehran Stock Exchange.

Keywords: Stock Returns, Machine Learning Methods, Financial and Regulatory Criteria.

[1] jafarilarijani@gmail.com © (Corresponding Author)