فصلنامه راهبرد مدیریت مالی
دانشگاه الزهرا
سال یازدهم، شماره چهل و دوم، پاییز ۱۴۰۲
صفحات ۲۸-۱

مقاله پژوهشی

مقایسه دقت مدل‌های آماری و یادگیری ماشین برای پیش‌بینی نگهداشت وجه نقد و ارائه مدل بهینه [۱]

سجاد میرزایی [۲]، مهدی محمدی [۳]، غلامرضا منصورفر [۴]

تاریخ دریافت: ۱۴۰۱/۱۱/۲۷ تاریخ پذیرش:۱۴۰۲/۰۷/۰۲

چکیده

پژوهش حاضر، مقایسه دقت مدل‌های یادگیری ماشین و آماری در پیش‌بینی نگهداشت وجه نقد را با استفاده از مجموعه متغیرهای مالی و اقتصادی مورد بررسی قرار داده است. روش‌شناسی پژوهش را می‌توان به سه مرحله گزینش مجموعه داده و متغیرها، مدل‌سازی و قیاس تقسیم‌بندی کرد. نمونه‌آماری پژوهش حاضر بورس اوراق بهادار تهران است که داده‌های ۱۷۳ شرکت در طی بازه زمانی ۱۴۰۰-۱۳۸۹ مورد بررسی قرارگرفته است. نتایج حاکی از دقت بالای مدل رگرسیون نمادین با استفاده از الگوریتم ژنتیک با ضریب دقت ۷۰ درصد در این زمینه است. بعدازآن به ترتیب مدل‌های رگرسیون مارس، رگرسیون بردار پشتیبان، تقویت گرادیان درختی، شبکه عصبی و تقویت گرادیان فوق‌العاده به‌عنوان دقیق‌ترین مدل‌ها جهت پیش‌بینی ارزیابی شدند. درنهایت مدل K نزدیک‌ترین همسایه ضعیف‌ترین دقت پیش‌بینی را از خود نشان داد. همچنین، اگرچه مدل‌های آماری دقت پیش‌بینی پایینی را نشان دادند اما بااین‌حال از برخی مدل‌های یادگیری ماشین ضریب دقت بالاتری را کسب کردند. همچنین نتایج نشان داد استفاده از رگرسیون لاسو موجب بهبود دقت مدل‌های آماری و برخی از مدل‌های یادگیری ماشین می‌گردد. این پژوهش می‌تواند زوایای جدیدی از تکنیک‌های پیش‌بینی نگهداشت وجه نقد را در مطالعات مالی بیفزاید که تاکنون در ادبیات مالی مورد بررسی قرار نگرفته است.

واژگان کلیدی: رگرسیون لاسو، پیش‌بینی نگهداشت وجه نقد، یادگیری ماشین.

طبقه‌بندی موضوعی: *C52, C53, G32*.

---

۱. کد DOI مقاله: 10.22051/JFM.2023.42943.2789

۲. کارشناسی ارشد، مدیریت مالی، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. Email: st_s.mirzaei@urmia.ac.ir

۳. کارشناسی ارشد، حسابداری مدیریت، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. Email: mehdim0719@gmail.com

۴. دانشیار، گروه حسابداری و مدیریت مالی، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. (نویسنده مسئول). Email: g.mansourfar@urmia.ac.ir

مقدمه

ادبیات آکادمیک در مورد نگهداری وجه نقد به کینز[1] (۱۹۳۶) برمی‌گردد که دو فایده اصلی از نگهداری وجه نقد را نشان می‌دهد: مقدار کافی وجه نقد در هنگام پرداخت‌های مالی باعث می‌شود که نیازی به فروش دارائی‌ها نباشد که این موجب کاهش هزینه‌های تراکنش می‌شود و همچنین یک ذخیره‌ی احتیاطی با ارزش برای برآورده کردن موارد احتمالی غیرمنتظره است.

یکی از سؤالات مورد بحث در ادبیات مالی شرکت‌ها این است که انگیزه شرکت‌ها برای انباشت وجه نقد و معادل وجوه نقد چیست؟ به‌عبارت‌دیگر، چرا شرکت‌ها از وجه نقد خود برای توزیع مجدد یا سرمایه‌گذاری مجدد استفاده نمی‌کنند؟ در پاسخ به این سؤال، اوزلم و تان[2]، (۲۰۲۲) بیان می‌دارند که در دو دهه اخیر، شرکت‌ها به‌منظور مدیریت حوادث غیرقابل پیش‌بینی در جریانات نقدی، تأمین مالی عملیات روزانه و حتی گاهی برای تأمین مالی پروژه‌های بلندمدت دارایی‌های نقدی خود را به‌طور قابل‌توجهی افزایش داده‌اند. انتظار می‌رود سطح نگهداشت وجه نقد در شرکت متأثر از عوامل متعددی مانند اندازه شرکت، مخارج سرمایه‌ای، خالص سرمایه در گردش غیر نقدی، فرصت‌های رشد شرکت و بدهی‌های کوتاه‌مدت باشد (کیم و همکاران[3]، ۲۰۱۵؛ پالازو[4]، ۲۰۱۲؛ گارسیا و همکاران[5]، ۲۰۰۹). به‌عنوان مثال، شرکت‌هایی که اندازه بزرگ‌تری دارند نسبت به شرکت‌هایی که اندازه کوچک‌تری دارند وجه نقد کمتری نگهداری می‌کنند، زیرا آن‌ها به سطح معینی از وجه نقد برای صرفه‌جویی ناشی از مقیاس در چرخه تولید نیاز دارند. از سوی دیگر انتظار می‌رود اندازه شرکت به دلیل دسترسی بهتر شرکت‌های بزرگ به منابع مالی خارجی دارای تأثیر منفی بر سطح نگهداشت وجه نقد باشد (ملکی و همکاران، ۱۴۰۱). همچنین، اوزکان و اوزکان[6] (۲۰۰۴) استدلال کردند ازآنجاکه بدهی زیاد ممکن است منجر به مشکلات مالی شود، شرکت‌ها دارایی‌های نقدی خود را افزایش می‌دهند تا احتمال مشکلات مالی را کاهش دهند. علاوه‌براین، شرکت‌های دارای محدودیت مالی، یعنی شرکت‌هایی که دسترسی کمتری به تأمین مالی خارجی دارند، باید تمایل بیشتری برای پس‌انداز وجه نقد داشته باشند و انباشت وجه نقد برای تأمین سطوح بالاتر نیازهای مالی ترجیح دهند (آچاریا و همکاران[7]، ۲۰۰۷). گیاکومینو و مایلک (۱۹۹۳) نیز نشان دادند شرکت‌هایی که میزان بالاتری از مخارج سرمایه‌ای دارند میزان بیشتری از وجه نقد نگهداری می‌کنند.

در ادبیات مالی، چهار دسته از انگیزه شامل معامله، احتیاط، هزینه نمایندگی و انگیزه مالیاتی برای شرکت‌ها در مورد نگهداری وجه نقد شناسایی شده است (بیتس و همکاران[8]، ۲۰۰۹).

---

1. Keynes
2. Ozlem & Tan
3 . Kim et al
4. Palazzo
5. García
6. Ozkan & Ozkan
7. Acharya et al
8. Bates et all

در انگیزه‌ی معاملاتی، نگهداری وجه نقد برای پرداخت صورتحساب‌ها انجام می‌شود. وجه نقد مورد نیاز مربوط به معاملات یک شرکت از پرداخت‌های عادی و دریافت مطالبات آن ناشی می‌شود. شرکت‌هایی که دارای منابع مالی داخلی کافی نیستند، می‌توانند دارایی‌های غیرمالی را به وجه نقد تبدیل کنند، سهام و بدهی جدید منتشر کنند، یا پرداخت سود سهام را کاهش دهند (اوزلم و تان، ۲۰۲۲).

انگیزه احتیاطی به ذخایر نقدی اشاره دارد که به‌عنوان انگیزه‌ای پیشگیرانه در برابر کمبود وجه نقد یا به دست آوردن فرصت‌های سرمایه‌گذاری سودآور نگهداری می‌شود. شرکت‌هایی که بودجه‌ای را برای این منظور کنار نمی‌گذارند ممکن است مجبور شوند از پروژه‌های سرمایه‌گذاری ارزشمند صرف‌نظر کنند یا با ورشکستگی مبارزه کنند (بیتس و همکاران، ۲۰۰۹).

تضاد منافع بین مدیران و سهامداران منجر به انگیزه‌های نمایندگی برای نگهداری وجه نقد می‌شود. مدیران ترجیح می‌دهند از منابع شرکت برای تأمین منافع خود استفاده کنند تا اینکه منافع سهامداران را به حداکثر برسانند. همچنین مدیران تمایل دارند به‌جای پرداخت سود سهام به سهامداران و سرمایه‌گذاری در پروژه‌های با خالص ارزش فعلی منفی، وجه نقد را حفظ کنند. به‌این‌ترتیب، آن‌ها دارایی‌های تحت کنترل خود را افزایش می‌دهند و دارای قدرت تصمیم‌گیری در سرمایه‌گذاری‌های شرکت می‌باشند (کمپیل و همکاران[1]، ۲۰۱۰). زمانی که هزینه‌های نمایندگی ناشی از تفکیک کنترل مالکیت زیاد باشد سهامداران ارزش کمتری را به دارایی‌های نقدی در شرکت‌هایی با عدم تقارن اطلاعاتی بالا و با حاکمیت شرکتی ضعیف اختصاص می‌دهند (هارفورد و همکاران[2]، ۲۰۰۸).

همچنین شرکت‌هایی که با مالیات بر بازگشت بیشتری به کشور مواجه می‌شوند تصمیم می‌گیرند که وجه نقد زیادی را در خارج از کشور به‌عنوان انگیزه مالیاتی نگه‌دارند (فولی و همکاران[3]، ۲۰۰۷).

باید توجه داشت که هرچند نگهداری مقدار مناسب وجوه نقد از الزامات مدیریت بهینه سرمایه-درگردش شرکت‌ها محسوب می‌گردد، وجود و نگهداشت بیش از حد وجوه نقد می‌تواند منجر به تأمین منافع کوتاه‌مدت توسط مدیران شود که نتیجه آن نیز ممکن است زیان وارده به سهامداران و یا بروز بحران‌های مالی برای شرکت باشد (وو و همکاران[4]، ۲۰۲۱).

به‌طورکلی، نرخ بازده حاصل از نگهداشت و استفاده از موجودی وجه نقد شرکت کمتر از نرخ بهره بازار است (وو و همکاران، ۲۰۲۱)؛ بنابراین، از یک‌سو، نگهداری وجه نقد موجب افزایش هزینه فرصت شرکت خواهد شد و از سوی دیگر، نگهداری ناکافی وجه نقد نیز ممکن است منجر به از دست دادن فرصت‌های سرمایه‌گذاری و از دست دادن گردش مالی و حتی بروز بحران مالی در شرکت شود (وو و همکاران، ۲۰۲۱).

اهمیت نگهداری وجه نقد را نمی‌توان نادیده گرفت، زیرا نقش مهمی در امور مالی شخصی و تجاری دارد. در امور مالی شخصی، دارایی‌های نقدی به‌عنوان یک فضای ایمن در مواقع اضطراری عمل می‌کند و

1. Campello et al
2. Harford et al
3. Foley et al
4. Wu et al

داشتن مقدار کافی وجه نقد می‌تواند به کاهش استرس مالی و ایجاد آرامش در مواقع سخت کمک کند. در شرکت‌ها نیز، نگهداری وجه نقد برای مدیریت جریان نقدی و عملیات تأمین مالی ضروری است. جریان نقدی مثبت برای بقای هر کسب‌وکاری حیاتی است و داشتن وجه نقد می‌تواند این اطمینان را حاصل کند که شرکت دارای بودجه کافی برای تأمین هزینه‌های روزانه خود و پرداخت به کارکنان، تأمین‌کنندگان و وام‌دهندگان خود است(دیتمار و همکاران[1]، ۲۰۰۳). دارایی‌های نقدی همچنین به شرکت‌ها اجازه می‌دهد از فرصت‌هایی مانند سرمایه‌گذاری در تجهیزات جدید یا خرید شرکت‌های دیگر استفاده کنند. به‌عبارت‌دیگر پیش‌بینی نقدینگی به مدیران کمک می‌کند تا تعیین کنند چگونه می‌توان از وجه نقد برای ایجاد سود بیشتر استفاده کرد و چگونه مدیران می‌توانند از شرکت در برابر چالش‌های مالی محافظت کنند (دونپودی و همکاران[2]، ۲۰۲۰). بنابراین لازم است میزان نگهداشت بهینه وجه نقد سنجیده و مورد پیش‌بینی قرار گیرد.

دو روش رایج برای پیش‌بینی، استفاده از مدل‌های آماری و یادگیری ماشین است. مدل‌های آماری رایج‌ترین روش‌های قابل‌استفاده برای مدل‌سازی می‌باشند که شامل دودسته تک متغیره و چند متغیره می‌باشند. ویژگی مثبت مدل‌های آماری ساده اجراشدن و از معایب آن‌ها می‌توان به مفروضات متعدد و محدودیت‌های فراوان اشاره کرد. همچنین در مدل‌های آماری برخلاف مدل‌های یادگیری ماشین پارامترهای مدل، برای تناسب داده‌ها با ساختار از قبل مشخص‌شده تخمین زده می‌شود (داستیل و همکاران[3]، ۲۰۲۰). توسعه و ترقی‌های اخیر درزمینه یادگیری ماشینی موجب استقرار موفقیت‌آمیز الگوریتم‌های یادگیری ماشینی در امور پیش‌بینی برای انواع کاربردهای خارج از امور مالی و حسابداری شده است. پژوهشگران مالی نیز اخیراً روش‌های یادگیری ماشینی را برای مشکلات مربوط به گزارش‌های مالی مانند کشف تقلب (پرولز[4]، ۲۰۱۱)، پیش‌بینی ورشکستگی (باربوزا و همکاران[5]، ۲۰۱۷) و تحلیل متنی (لی[6]، ۲۰۱۰) اتخاذ کرده‌اند. به دلیل گستردگی اطلاعات موجود در صورت‌های مالی و وابستگی‌های بین زمانی آن‌ها، رویکردهای مبتنی بر یادگیری ماشین ممکن است به پژوهش‌های تجزیه‌وتحلیل صورت‌های مالی کمک کند (والن و ویلند[7]، ۲۰۱۱). روش‌های یادگیری ماشین احتمالاً برای حل مسئله، ایجاد پیش‌بینی نگهداشت وجه نقد خارج از نمونه مناسب‌تر هستند، زیرا ازنظر تاریخی یادگیری ماشین بر پیش‌بینی و به حداکثر رساندن دقت پیش‌بینی متمرکز بوده است (مولیناتان و اسپیس[8]، ۲۰۱۷). همچنین با گسترش اصطلاح یادگیری ماشین بسیاری از روش‌ها در یادگیری ماشین بهتر از روش‌های سنتی مبتنی بر رگرسیون می‌توانند

1. Ditmar et al
2. Donepudi et al
3. Dastile et al
4. Perols
5. Barboza et al
6. Li
7. Wahlen & Wieland
8. Mullainathan & Spiess

مشکلات هم‌خطی و غیرخطی بودن را مدیریت کنند. بسیاری از روش‌های یادگیری ماشین بیشتر یک جعبه سیاه هستند تا یک رگرسیون معمولی و این معاوضه بین استفاده از برخی روش‌های یادگیری ماشین و روش‌های سنتی است. استفاده از یادگیری ماشین به‌عنوان جایگزینی برای رگرسیون، معمولاً دقت پیش‌بینی خارج از نمونه را به‌خوبی به دست می‌آورد اما مقداری قابلیت تفسیر را از دست می‌دهد. با توجه به برخی مشکلات موجود، بهبود در دقت پیش‌بینی خارج از نمونه، این معامله می‌تواند ارزشمند باشد (آناند و همکاران[1]، ۲۰۱۹). به‌طورکلی می‌توان گفت، مدل‌های یادگیری ماشین نسبت به مدل‌های رگرسیونی انعطاف‌پذیرتر هستند، زیرا بر مفروضات آماری و اقتصادی محدودکننده تکیه نمی‌کنند و تحت تأثیر سوگیری‌های شناختی قرار نمی‌گیرند. در عوض، آن‌ها از الگوها و روندها در داده‌های تاریخی برای پیش‌بینی استفاده می‌کنند (رامناث و همکاران[2]، ۲۰۰۸).

بنابراین در این پژوهش به این مهم پرداخته شده است که کدام نوع از مدل‌ها توانایی بالاتری در پیش‌بینی نگهداشت وجه‌نقد دارند و آیا مدل‌های غیرقابل تفسیر و با پیچیدگی بیشتر عملکرد بهتری را از خود نشان می‌دهند.

در ادامه ابتدا به مبانی نظری و پیشینه پژوهش پرداخته شده است. سپس فرضیه های پژوهش ارائه شده است و پس از آن به ترتیب روش‌شناسی پژوهش، متغیر و مدل‌های پژوهش حاضر تشریح شده است. در نهایت، ابتدا داده ها مورد تجزیه و تحلیل قرارگرفته است و سپس نتایج حاصل مورد بحث و نتیجه‌گیری قرار گرفته است.

مبانی نظری و مروری بر پیشینه پژوهش

مبانی نظری

نگهداشت وجه نقد

وجه نقد به لحاظ عینیت آن به‌عنوان شاخصی از تداوم فعالیت و توانایی شرکت هنگام مواجهه با مشکلات مالی و نیز به‌عنوان ابزاری برای بهره‌برداری از فرصت‌های رشد و سرمایه‌گذاری سودآور، موردتوجه استفاده‌کنندگان از صورت‌های مالی قرار دارد. امکان سوءاستفاده از موجودی وجوه نقد، همواره نگرانی‌هایی را در خصوص میزان نگهداشت آن در شرکت به همراه دارد. به‌عنوان مثال، عدم تقارن اطلاعاتی بالا در شرکت، منجر به فزونی تمایل مدیران به پذیرش پروژه‌هایی با خالص ارزش فعلی منفی و گسترش حکمرانی آن‌ها می‌شود که این امر کاهش ارزش نگهداشت وجه نقد را به دنبال دارد. چنانچه محرک اصلی این کاهش ارزش، سرمایه‌گذاری بیش‌ازحد شرکت باشد، انتظار می‌رود محافظه‌کاری بر ارزش نهایی نگهداشت وجه نقد، تأثیری مثبت داشته باشد چراکه محافظه‌کاری، شناسایی سریع‌تر زیان‌های ناشی از عملکرد ضعیف پروژه‌ها را ایجاب می‌کند که این خود احتمال رسیدن به آستانه‌ی نقض قراردادهای بدهی و متعاقباً به

1. Anand et al
2. Ramnath et al

مخاطره افتادن امنیت شغلی مدیر را به همراه خواهد داشت. بدین ترتیب انگیزه‌ی پذیرش پروژه‌هایی با خالص ارزش منفی کاهش خواهد یافت (بال و شیواکومار[1]، ۲۰۰۵).

تصور تمامی ذینفعان، سرمایه‌گذاران، وام‌دهندگان و کارکنان از بنگاه اقتصادی منبعی از وجه نقد است که منجر به تقسیم سود، پرداخت بهره، افزایش در ارزش بازار سهام، بازپرداخت وام‌ها، پرداخت بهای کالا و خدمات یا حقوق و دستمزد می‌شود. در اهمیت نقدینگی همین بس که شرکت‌هایی با سودآوری پایین یا حتی غیر سودآور تا مدت‌ها می‌توانند به فعالیت خود ادامه دهند اما تداوم فعالیت شرکت‌های بدون نقدینگی بسیار سخت است (شکرخواه و مرتضی زاده، ۱۳۹۴).

پیشینه پژوهش

رجب‌زاده و همکاران(۱۴۰۱)، طی پژوهشی با عنوان پیش‌بینی جریان وجه نقد عملیاتی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران با استفاده از روش یادگیری ماشین به پیش‌بینی جریان نقد عملیاتی با رویکرد هوش مصنوعی PLSVM و CART در بازه زمانی ۱۳۹۰ تا ۱۳۹۷ پرداختند و نشان دادند که رویکرد هوش مصنوعی قانون‌گرا و غیرخطی پارامتریک توانایی بالایی در پیش‌بینی نقدینگی شرکت‌های بورس اوراق بهادار تهران دارد. پورقدیمی و همکاران (۱۴۰۱) به بررسی ارائه الگوی توسعه‌یافته مدل بنیش با تأکید بر ویژگی‌های خاص شرکت با استفاده از الگوریتم‌های شبکه عصبی، ماشین بردار پشتیبان و جنگل تصادفی پرداختند. نتایج نشان داد دقت ضرایب به‌دست‌آمده از مدل جنگل تصادفی ۹۹ درصد، و بیشتر از دو مدل شبکه عصبی و ماشین بردار پشتیبان با مقدار ۹۴ درصد است. همچنین مدل بسط یافته نسبت به مدل پایه دقت بیشتری دارد. افلاطونی و همکاران(۱۴۰۱) در پژوهشی به بررسی اثر رونق و رکود اقتصادی بر سرعت تعدیل نگهداشت وجه نقد با استفاده از رگرسیون خطی چندگانه پرداختند. آن‌ها برای این منظور از ۲۸۹۳ مشاهده در بازه‌ی زمانی ۱۳۸۲-۱۳۹۹ در قالب داده‌های ترکیبی نامتوازن استفاده کردند. برای برآورد مدل‌های ایستا از برآوردگر حداقل مربعات معمولی و جهت تخمین مدل‌های پویا از برآوردگر گشتاورهای تعمیم‌یافتهٔ سیستمی استفاده کردند. نتایج پژوهش آن‌ها وجود نسبت نگهداشت وجه نقد بهینه را در شرکت‌های ایرانی تأیید کرد. همچنین افزون بر آن، نتایج نشان داد که در قیاس با دوران رکود، در دوران رونق اقتصادی، شرکت‌ها نسبت نگهداشت وجه نقد واقعی خود را با سرعت بیشتری به سمت نسبت هدف، تعدیل می‌کنند. ملکی و همکاران(۱۴۰۱) در پژوهشی به بررسی رابطه بین تجربه مدیرعامل و سطح نگهداشت وجه نقد با استفاده از رگرسیون خطی چندگانه بر اساس اطلاعات ۱۴۴ شرکت پذیرفته‌شده در بورس اوراق بهادار تهران بین سال‌های، ۱۳۹۲-۱۳۹۹ که به روش حذف سیستماتیک نمونه انتخاب شده پرداختند و نتایج پژوهش نشان داد بین تجربه مدیرعامل و سطح نگهداشت وجه نقد رابطه مثبت و معنی‌داری وجود دارد. ملائی(۱۴۰۱) در پژوهشی به بررسی تأثیر کیفیت اطلاعات حسابداری بر نگهداشت وجه نقد با تأکید بر نرخ تورم با استفاده از رگرسیون خطی چندگانه بر اساس اطلاعات ۱۲۸

1. Ball & Shivakumar

شرکت پذیرفته شده در بورس اوراق بهادار تهران برای سال‌های ۱۳۹۲ تا ۱۳۹۹ پرداخت. نتایج نشان داد که کیفیت اطلاعات حسابداری بر نگهداشت وجه نقد تأثیر معکوس و معنی‌داری دارد. همچنین، نرخ تورم بر ارتباط بین کیفیت اطلاعات حسابداری بر نگهداشت وجه نقد تأثیر معکوس و نرخ تورم ارتباط معکوس بین کیفیت اطلاعات حسابداری و نگهداشت وجه نقد را تشدید می‌کند. اوزلم و تان(۲۰۲۲) در پژوهشی به پیش‌بینی نگهداشت وجه نقد با استفاده از اطلاعات ۲۱۱ شرکت ترکیه‌ای با استفاده از روش یادگیری ماشین برای سال‌های ۲۰۰۶ تا ۲۰۱۹ پرداختند و نتایج آن‌ها نشان داد رگرسیون خطی چندگانه[1]، kنزدیک‌ترین همسایه[2] و رگرسیون بردار پشتیبان[3] مجذور میانگین مربعات خطا[4] بالا و مقادیر ضریب دقت پایین ارائه می‌دهند. در همین حال الگوریتم‌های پیچیده‌تر مانند درخت تصمیم[5] و به‌خصوص تقویت گرادیان فوق‌العاده[6] دقت بالاتری به دست می‌آورند. همچنین وو و همکاران (۲۰۲۱) در پژوهشی به بررسی پیش‌بینی موجودی وجه نقد با استفاده از الگوریتم‌های درخت تصمیم شامل، مدل درخت لجستیک، جنگل تصادفی و درخت اضافی در صنایع الکترونیک تایوان پرداختند. یافته‌های آن‌ها نشان داد جنگل تصادفی بالاترین و بهترین نرخ پیش‌بینی را با استفاده از مدل رگرسیون لجستیک دارد. اوز و همکاران(۲۰۲۱) به بررسی نقش مؤلفه‌های سود در آشکارسازی وخامت عملکرد شرکت با استفاده از الگوریتم‌های یادگیری ماشین (مانند: درخت تصمیم، شبکه‌ی عصبی، ماشین بردار پشتیبان ، جنگل تصادفی، تقویت گرادیان درختی) و مدل آماری رگرسیون لجستیک پرداختند. نتایج نشان داد شبکه‌ی عصبی بیشترین ضریب دقت را در پیش‌بینی بحران سود ارائه می‌دهد و ماشین بردار پشتیبان بیشترین ضریب دقت را در پیش‌بینی بحران جریان وجه نقد ارائه داد. همچنین رگرسیون لجستیک ضریب دقت بیشتری را نسبت به برخی از مدل‌های یادگیری ماشین مانند جنگل تصادفی نشان داد و استفاده از مدل تقویت گرادیان موجب بهبود دقت پیش‌بینی شد. شینیو[7] و همکاران(۲۰۲۰) طی پژوهشی با عنوان استفاده از یادگیری‌ماشینی برای پیش‌بینی درآمدهای آینده به‌طور جامع امکان‌پذیری و مناسب بودن اتخاذ مدل‌های یادگیری ماشین را بر پیش‌بینی سود ارزیابی کردند و نشان دادند که در مقایسه با مدل‌های آماری سنتی قبلی مانند رگرسیون لجستیک روش آن‌ها در دقت و سرعت پیش‌بینی به پیشرفت رضایت‌بخشی دست یافته است. آناند و همکاران(۲۰۱۹) طی پژوهشی با عنوان پیش‌بینی سودآوری با استفاده از یادگیری ماشین به پیش‌بینی خارج از نمونه سودآوری و مقایسه آن با گام تصادفی در ایالت متحده آمریکا در بازه زمانی ۲۰۱۰ تا ۲۰۱۸ پرداختند و به دقت طبقه‌بندی ۵۷ تا ۶۴ درصد برای معیارهای سودآوری در مقایسه با ۵۰ درصد گام تصادفی دست‌یافتند. همچنین دریافتند که دقت طبقه‌بندی در افق‌های یک تا پنج‌ساله یکسان است و

1. Multiple Linear Regression
2. K-Nearest Neighbors
3. Support Vector Machine
4. Root-Mean-Square Error
5. Decision Tree
6. XGradient Boosting
7. Xinyue

همچنین نشان دادند که روش‌های یادگیری‌ماشین عملکرد پیش‌بینی بهتری نسبت به روش‌های سنتی مبتنی بر رگرسیون ارائه می‌دهند. جونز و همکاران(۲۰۱۷) به بررسی ورشکستگی شرکت‌های آمریکایی با استفاده از مدل‌های آماری (رگرسیون لجستیک ، پروبیت و آنالیز تشخیصی خطی) و مدل‌های یادگیری ماشین (شبکه‌های عصبی، ماشین‌های بردار پشتیبان، تقویت تعمیم‌یافته و جنگل‌های تصادفی) پرداختند. نتایج نشان داد رگرسیون لجستیک و آنالیز تشخیصی خطی، ورشکستگی را به خوبی پیش‌بینی می‌کنند. همچنین آن‌ها استفاده از مدل‌های یادگیری ماشین را در مدل‌سازی ورشکستگی شرکت‌ها توصیه کردند، زیرا مدل‌های یادگیری ماشین به‌طور قابل توجهی بهتر از همه طبقه‌بندی کننده‌های دیگر در هر دو نمونه آزمایش مقطعی و طولی پیش‌بینی می‌کنند. الیاسیانی و موقاری (۲۰۲۲) در پژوهشی به شناسایی عوامل تعیین‌کننده وجه نقدی شرکت‌های ایالات متحده با استفاده از تکنیک رگرسیون قوی لاد[1] و روش پیشرفته انتخاب متغیر لاسو پرداختند و نتایج آنان نشان داد که میانه‌ی خطاهای مطلق متغیرهای انتخاب شده توسط لادلاسو[2] به‌طور قابل‌توجهی پایین‌تر از مقادیر متناظر تعیین‌کننده‌های اساسی در هر دو طرح درون نمونه و خارج از نمونه است. همچنین اهرم مالی یک عامل تعیین‌کننده کلیدی نگهداری وجه نقد است که نشان می‌دهد سیاست‌های نقدی و بدهی ارتباط تنگاتنگی با یکدیگر دارند. الیاسیانی و همکاران (۲۰۱۹) به بررسی عوامل تعیین‌کننده پرداخت سود سهام و تمایل به سود سهام در بازارهای نوظهور با استفاده از رگرسیون لاسو پرداختند. نتایج نشان داد متغیرهای انتخاب شده توسط لاسو از متغیرهای انتخاب شده توسط مدل‌های آماری از نظر برازش مدل و دقت پیش‌بینی عملکرد بهتری دارند. همچنین در صورت ترکیب الگوریتم ماشین بردار پشتیبان با متغیرهای انتخاب شده توسط رگرسیون لاسو، بهتر می‌تواند بین شرکت‌های پرداخت‌کننده سود سهام و شرکت‌های غیرپرداخت کننده نسبت به روش‌های دیگر مانند رگرسیون لجستیک و آنالیز تفکیک خطی تمایز قائل شود.

در بسیاری از پژوهش‌های قبلی (به‌طور مثال: آچاریا و همکاران[3]، ۲۰۰۷؛ سابرامانیام و همکاران[4]، ۲۰۱۱؛ گارسیا و همکاران، ۲۰۰۹؛ اوپلیر و همکاران[5]، ۱۹۹۹) با استفاده از روش‌های رگرسیون کلاسیک، تأثیر بسیاری از متغیرهای مالی بر رفتار نگهداری وجه نقد شرکت‌ها بررسی‌شده است. اگرچه در پژوهش‌های پیشین داخلی پیش‌بینی نگهداشت وجه نقد انجام شده است اما این مورد فقط محدود به رگرسیون خطی چندگانه بوده است و در پژوهش‌های خارجی نیز تعداد معدودی از پژوهش‌ها روش‌های جدیدتر یادگیری ماشین را در این حوزه مورد بررسی قرار داده‌اند بنابراین برخلاف ادبیات قبلی، هدف پژوهش حاضر پیش‌بینی رفتار نگهداری وجه نقد شرکت‌ها با استفاده از روش‌های پیشرفته یادگیری ماشین است همچنین به این مهم پرداختیم که آیا مدل‌های پیچیده‌تر از مدل‌های ساده‌تر در پیش‌بینی نگهداشت وجوه نقد کاراتر

1. Lad Regression
2. Lad Lasso
3. Acharya
4. Subramaniam et al
5. Opler et al

هستند. علاوه بر این در پژوهش حاضر از تکنیک رگرسیون لاسو به‌منظور بهبود نتایج پیش‌بینی استفاده شده است و فهرست جامع‌تری از متغیرهای مالی و اقتصادی مورد بررسی قرارگرفته است.

فرضیه‌های پژوهش

با توجه به مطالب بیان‌شده در مقدمه و پژوهش‌های نام برده، از میان پژوهش‌های مربوط به پیش‌بینی نگهداشت وجه نقد و سایر پیش‌بینی‌های حوزه مالی، تاریخ بیانگر این موضوع است که مدل‌های یادگیری ماشین به نظر کاراتر از مدل‌های آماری است. همچنین تاریخ نشان داده است استفاده از رگرسیون لاسو موجب بهبود دقت پیش‌بینی مدل‌ها به‌خصوص در مدل‌های آماری شده است. بنابراین فرضیه‌های پژوهش حاضر به شکل زیر تدوین شده است:

فرضیه اول: دقت بهترین مدل یادگیری ماشین از بهترین مدل آماری جهت پیش‌بینی نگهداشت وجه نقد بالاتر است.

فرضیه دوم: استفاده از رگرسیون لاسو در انتخاب ویژگی موجب افزایش ضریب دقت در مدل‌های آماری و یادگیری ماشین می‌شود.

روش‌شناسی پژوهش

جامعه آماری این پژوهش کلیه شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را در دوره زمانی ۱۳۸۹ تا ۱۴۰۰ را شامل می‌شود. برای انتخاب نمونه آماری از روش حذف سیستماتیک استفاده شده است. بدین‌صورت که شرکتی که همه معیارها را احراز کرده باشد، به‌عنوان نمونه پژوهش انتخاب می‌شود. معیارهای موردنظر عبارت‌اند از: شرکت در گروه شرکت‌های سرمایه‌گذاری و یا واسطه‌گری مالی نباشد، اطلاعات مالی آن برای دوره موردنظر در اختیار باشد، در طول دوره موردنظر پژوهش تغییر سال مالی نداشته باشند، در طول دوره موردنظر پژوهش در بورس حضور داشته باشد و دوره مالی شرکت‌ها منتهی به ۱۲/۲۹ هر سال باشد. پس از اعمال محدودیت‌های مذکور ۱۷۳ شرکت به‌عنوان نمونه نهایی پژوهش در نظر گرفته شد. برای جمع‌آوری داده‌ها از بانک‌های اطلاعاتی بورس‌ویو، ره‌آورد نوین، وب‌سایت کدال، وب‌سایت مرکز آمار ایران و وب‌سایت بانک مرکزی ایران استفاده شده است.

در پژوهش حاضر، مدل‌سازی پیش‌بینی نگهداشت وجه نقد در مرحله اول با کل مجموعه متغیرهای پژوهش و در مرحله دوم و به‌منظور بررسی تأثیر رویکرد انتخاب ویژگی در نتایج پیش‌بینی و همچنین کاهش وجود هم خطی احتمالی با استفاده از مجموعه متغیرهای پیشنهاد شده توسط رگرسیون لاسو انجام شده است. در این پژوهش از ۲ مدل آماری، رگرسیون چند متغیره خطی و مدل خطی تعمیم‌یافته و ۱۰ مدل یادگیری ماشین شامل روش‌های یادگیری عمیق، درخت تصمیم، جنگل تصادفی، تقویت گرادیان

درختی، تقویت گرادیان فوق‌العاده، رگرسیون بردار پشتیبان، k نزدیک‌ترین همسایه، رگرسیون نمادین، رگرسیون مارس و شبکه عصبی استفاده شده است. داده‌های گردآوری‌شده با استفاده از الگوریتم نرم‌افزارهای داده‌کاوی SPM، RapidMiner، Eureqa و نرم‌افزار اقتصادسنجی و آماری Stata مورد تجزیه‌وتحلیل قرارگرفته و همچنین برای مرتب‌سازی و دسته‌بندی داده‌ها از Excel استفاده شده است.

متغیرها و مدل‌ها

در پژوهش حاضر از داده‌های ۲۵ متغیر مالی و اقتصادی استفاده شده است که تعریف عملیاتی متغیرها در جدول ۱ بیان شده است. در این پژوهش مدل ها از طریق رویکرد اعتبارسنجی متقابل با k=10 مورد ارزیابی قرارگرفته است. مدل‌های پژوهش با توجه به مدل‌های استفاده شده در پیش‌بینی نگهداشت وجه نقد برای مثال اوزلم و تان (۲۰۲۲)، یا سایر پیش‌بینی‌های حوزه مالی که عملکرد مطلوبی داشته است مانند اوز و همکاران(۲۰۲۱) و جونز و همکاران(۲۰۱۷) انتخاب شده است. اگرچه فهرست جامعی نیست، اما ۱۲ مدل انتخاب‌شده برای این پژوهش به‌طور گسترده‌ای نمایانگر پرکاربردترین و مورد استنادترین مدل‌های رگرسیونی در ادبیات هستند(برای مروری بر کاربردهای مدرن به هستی[1] و همکاران(۲۰۰۹) مراجعه کنید). در یک طرف طیف، مدل‌های خطی نسبتاً ساده‌ای مانند رگرسیون خطی چندگانه و مدل خطی تعمیم‌یافته را داریم. این مدل‌ها در ادبیات رایج هستند، اما ظرفیت محدودی برای مدل‌سازی روابط غیرخطی و ناهمسانی در مجموعه داده‌ها دارند. بااین‌حال، آن‌ها به‌طورکلی از نظر درک رابطه عملکردی بین متغیرهای مستقل و وابسته قابل تفسیر هستند. در وسط طیف، از جمله رویکردهای رگرسیون کمانک تطبیقی چند متغیره و رگرسیون نمادین مورد استفاده قرارگرفته که برای کنترل غیرخطی بودن و ناهمسانی، مجهزتر هستند. انعطاف‌پذیری بیشتر این مدل‌ها معمولاً به تناسب بهتر و عملکرد پیش‌بینی بهبود یافته، به قیمت تفسیرپذیری کمتر، ختم می‌شود. در انتهای طیف، مدل‌های غیرخطی و کاملاً کلی داریم که به گونه‌ای طراحی شده‌اند که تمام روابط و تعاملات غیرخطی در مجموعه داده را به تصویر بکشند. این مدل‌ها شامل شبکه عصبی، یادگیری عمیق، رگرسیون بردار پشتیبان، مدل‌های تقویت گرادیان درختی و فوق‌العاده، درخت تصمیم و جنگل‌های تصادفی هستند. درحالی‌که الگوریتم‌های پیچیده زیربنای بسیاری از این مدل‌ها برای افزایش دقت پیش‌بینی طراحی شده‌اند، آن‌ها نیز موانع بزرگی برای تفسیر ایجاد می‌کنند، زیرا رابطه بین متغیرهای مستقل و متغیر وابسته تا حد زیادی در ریاضیات داخلی سیستم مدل پنهان است (جونز و همکاران، ۲۰۱۷). مزیت استفاده از مدل غیرخطی پیچیده‌تر (مانند شبکه‌های عصبی یا تقویت گرادیان) باید از موفقیت پیش‌بینی خارج از نمونه حاصل شود. اگر دو مدل عملکرد پیش‌بینی قابل مقایسه‌ای داشته باشند، مدل ساده‌تر و قابل تفسیرتر به مدل کمتر تفسیرپذیر ترجیح داده می‌شود، به‌ویژه اگر استنتاج آماری و نه فقط پیش‌بینی هدف اصلی مدل‌سازی باشد (جیمز[4] و همکاران ۲۰۱۳). بنابراین مدل‌های مختلف مورد بررسی قرارگرفته تا مشخص شود کدام یک توانایی بالاتری

1. Hastie

در پیش‌بینی نگهداشت وجه نقد دارد. همچنین پژوهش حاضر این مهم را ارزیابی می‌کند که آیا مدل‌های پیچیده‌تر در واقع منجر به موفقیت بهتر پیش‌بینی خارج از نمونه به‌ویژه در مقایسه با مدل‌های ساده‌تر و قابل تفسیر می‌شوند.

جدول ۱. تعریف عملیاتی متغیرهای پژوهش

متغیر	تعریف عملیاتی	منبع
وجه نقد	$\frac{\text{دارایی کوتاه مدت + وجه نقد}}{\text{کل دارایی ها}}$	اوزلم و تان (۲۰۲۲)
سود تقسیمی پرداختی	$\frac{\text{کل سود سهام پرداختی}}{\text{کل دارایی ها}}$	سانگ و لی[1] (۲۰۱۲)
نرخ رشد فروش	$\frac{\text{فروش ابتدای سال − فروش پایان سال}}{\text{فروش ابتدای سال}}$	بوباکری و همکاران[2] (۲۰۱۳)
اندازه شرکت	لگاریتم طبیعی دارایی‌ها	اوزکان و اوزکان (۲۰۰۴)
مخارج سرمایه‌ای	$\frac{\text{استهلاک + دارائی ثابت مشهود ابتدای دوره − دارائی ثابت مشهود انتهای دوره}}{\text{کل دارایی ها}}$	بوباکری و همکاران (۲۰۱۳)
درآمد قبل از مالیات به علاوه استهلاک	$\frac{\text{استهلاک + سود قبل مالیات}}{\text{کل دارایی ها}}$	لوزانو و یامان (۲۰۲۰)
هزینه‌ی مالی	$\frac{\text{هزینه بهره}}{\text{کل دارایی ها}}$	شاوتن و همکاران[3] (۲۰۱۱)
سرمایه در گردش غیر نقدی	$\frac{\text{سرمایه در گردش غیر نقدی}}{\text{کل دارایی ها}}$	دیاو[4] (۲۰۲۱)
خالص دارایی‌های ثابت	$\frac{\text{خالص دارایی های ثابت}}{\text{کل دارایی ها}}$	بهویان و هوکس[5] (۲۰۱۹)
بدهی کوتاه‌مدت	$\frac{\text{بدهی های کوتاه مدت}}{\text{کل دارایی ها}}$	لوزانو و یامان (۲۰۲۰)
بازده دارایی	$\frac{\text{سود خالص}}{\text{کل دارایی ها}}$	سارفریز و همکاران[6] (۲۰۲۰)
بازده حقوق صاحبان سهام	$\frac{\text{سود خالص}}{\text{حقوق صاحبان سهام}}$	مانوئل و همکاران[7] (۲۰۱۸)

1. Song & Lee
2. Boubakri et al
3. Schauten et al
4. Diaw
5. Bhuiyan & Hook
6. Sarfraz et al
7. Manoel et al

متغیر	تعریف عملیاتی	منبع
حساب های دریافتنی	حسابهای دریافتنی / کل دارایی ها	وو و همکاران (۲۰۱۲)
حساب‌های پرداختنی	حساب‌های پرداختنی / کل دارایی ها	چن و همکاران[1] (۲۰۱۴)
نسبت جاری	دارایی های جاری / بدهی های جاری	مانوئل و همکاران (۲۰۱۸)
سود هر سهم	سود خالص / تعداد سهام منتشره	اوزلم و تان (۲۰۲۲)
سود عملیاتی پس از کسر مالیات	سود عملیاتی پس از کسر مالیات / کل دارایی ها	سارفریز و همکاران (۲۰۲۰)
حاشیه سود خالص	سود خالص / فروش خالص	آنجلوسکا و والنتینچیچ[2] (۲۰۱۹)
سن شرکت	لگاریتم سال تاسیس تا سال ورود به بورس	اوزلم و تان (۲۰۲۲)
شاخص عدم قطعیت ایران	میانگین سالانه داده‌های فصلی وب سایت شاخص عدم قطعیت جهانی	اوزلم وتان (۲۰۲۱)
نرخ تورم	استخراج از وب سایت بانک داده‌های اقتصادی و مالی ایران	ملائی (۱۴۰۱)
رشد تولید ناخالص داخلی	استخراج از وب سایت بانک داده‌های اقتصادی و مالی ایران	هزبرکیانی و همکاران (۱۴۰۰)
نرخ رشد قیمت نفت	قیمت نفت ابتدای سال − قیمت نفت پایان سال / قیمت نفت ابتدای سال	وو و همکاران (۲۰۲۱)
رشد اقتصادی	استخراج از وب سایت بانک داده‌های اقتصادی و مالی ایران	اصغر پور و همکاران (۱۴۰۱)
نرخ بهره	استخراج از وب سایت بانک داده‌های اقتصادی و مالی ایران	لطفعلی پور و کریمی علویجه(۱۳۹۸)

منبع: یافته‌های پژوهش

رگرسیون لاسو

رگرسیون لاسو[3] نوعی تکنیک انتخاب متغیر غالب است و دارای چندین ویژگی مطلوب است که آن را برای کاربرد در امور مالی تجربی مناسب می‌کند (الیاسیانی و همکاران[4]، ۲۰۱۹). همچنین، رگرسیون لاسو نوعی تکنیک انتخاب متغیر پایدار است، به این معنا که معادلات رگرسیون تولید شده با تغییرات کوچک در داده‌ها تغییر قابل‌توجهی نمی‌کند (بریمان[5]، ۱۹۹۶). رگرسیون لاسو اغلب دقت پیش‌بینی را با حذف پیش‌بینی‌کننده‌های نامربوط و اجتناب از برازش بیش‌ازحد بهبود می‌بخشد و به‌طور طبیعی مشکل هم‌خطی را حل می‌کند (تیان و همکاران[6]، ۲۰۱۵).

1. Chen et al
2. Angelovska & Valentinčič
3. Lasso regression
4. Elyasiani & Movaghari
5. Breiman
6. Tian et all

رگرسیون نمادین با الگوریتم ژنتیک

رگرسیون نمادین یک مدل رگرسیون عملکردی است که الگوریتم ژنتیک را بهبود می‌بخشد. رگرسیون نمادین می‌تواند شبیه به یک دانشمند روباتیک، به‌طور خودکار روابطی را بر اساس ویژگی‌های داخلی داده‌ها برقرار کند. مدل، عملکردی با بالاترین برازش را برای نشان دادن این رابطه پیدا می‌کند و پارامترها و ساختار هر مدل رگرسیونی را تعیین می‌کند (کلاوریا و همکاران[1]، ۲۰۱۷). هسته رگرسیون نمادین بر اساس نظریه تکامل داروین است که عوامل مهم را انتخاب می‌کند تا به‌تدریج یک مدل را شکل دهد و به‌طور خودکار عوامل غیر مهم را از مدل جدا کند. (لیو و ژانگ[2]، ۲۰۲۲).

رگرسیون کمانک تطبیقی چند متغیره

رگرسیون کمانک تطبیقی چند متغیره یا مارس، روشی استاندارد برای گسترش توابع رگرسیون در روابط غیرخطی و جایگزینی برای یک مدل خطی با یک تابع چندجمله‌ای است. در حقیقت مارس الگوریتمی برای مسائل پیچیده رگرسیون غیرخطی است (چن و همکاران[3]، ۲۰۱۹). این الگوریتم شامل یافتن مجموعه‌ای از توابع خطی ساده است که درمجموع بهترین عملکرد پیش‌بینی را به همراه دارد. این مدل در مقایسه با مدل‌های رگرسیون خطی سنتی حساسیت کمتری به نقاط پرت دارد (جونز و همکاران[4]، ۲۰۱۷).

شبکه عصبی

شبکه‌های عصبی گاهی اوقات به‌عنوان مدل‌های متمایز غیرخطی توصیف می‌شوند. اساساً شبکه‌های عصبی یک مدل رگرسیون یا طبقه‌بندی دو مرحله‌ای هستند. شبکه عصبی مجموعه‌ای از الگوریتم‌ها است که تلاش می‌کند تا روابط زیربنایی را در مجموعه‌ای از داده‌ها از طریق فرآیندی که نحوه عملکرد مغز انسان را تقلید می‌کند، تشخیص دهد (اوز و همکاران، ۲۰۲۱). شبکه عصبی از شبکه‌ای از گره‌های به هم پیوسته برای یادگیری الگوها و روابط در داده‌ها استفاده می‌کند. گره‌ها یا نورون‌های مصنوعی، داده‌های ورودی را پردازش می‌کنند و آن‌ها را از طریق اتصالات وزنی برای تولید خروجی تبدیل می‌کنند. وزن این اتصالات در طول فرآیند آموزش برای به حداقل رساندن یک تابع هزینه تنظیم می‌شود، که تفاوت بین پیش‌بینی‌های شبکه و برچسب‌های واقعی در داده‌های آموزشی را اندازه‌گیری می‌کند. (جونز و همکاران، ۲۰۱۷).

یادگیری عمیق[5]

یادگیری عمیق نوع خاصی از یادگیری ماشین است که از چندین لایه شبکه عصبی تشکیل شده است. به عبارتی، مدل‌های این دو حوزه، با استفاده از اطلاعات آماری داده‌های ورودی خود به پیش‌بینی

1. Claveria et al
2. Liu & Zhang
3. Chen et al
4. Jones et al
5. Deep Learning

مقداری در خروجی می‌پردازند. یادگیری عمیق بر اساس یک شبکه عصبی مصنوعی پیش‌خور چندلایه است که با نزول گرادیان تصادفی با استفاده از انتشار پس از آموزش، آموزش داده شده است (سزر و همکاران[1]، ۲۰۲۰). ویژگی‌های پیشرفته‌ای مانند نرخ یادگیری تطبیقی، تکرار نرخ، آموزش تکانه و منظم‌سازی، دقت پیش‌بینی بالایی را ممکن می‌سازد (اوزبای اوغلو و همکاران[2]، ۲۰۲۰).

درخت تصمیم

درخت تصمیم ، می‌تواند برای کارهای طبقه‌بندی و پیوسته اعمال شود. درخت تصمیم یک الگوریتم یادگیری نظارت‌شده ناپارامتریک است و ساختار درختی سلسله مراتبی دارد که از یک ریشه تشکیل شده است (بریمن و همکاران[3]، ۱۹۸۴). علیرغم اینکه درخت تصمیم یک طبقه‌بندی موفق است، به‌خصوص با افزایش پیچیدگی مدل پیش‌بینی ممکن است مستعد به‌تناسب بیش‌ازحد داده‌ها باشد. درخت تصمیم از یک ساختار درخت مانند برای تصمیم‌گیری بر اساس ویژگی‌های ورودی استفاده می‌کند. این مدل به دلیل قابلیت تفسیر، توانایی مدیریت داده‌های دسته‌بندی و عددی و هزینه محاسباتی کم برای آموزش و پیش‌بینی شناخته شده است (اوز و همکاران[4]،۲۰۲۱).

جنگل تصادفی

جنگل‌های تصادفی یک روش یادگیری مجموعه‌ای برای طبقه‌بندی، رگرسیون و سایر وظایف است که با ساختن تعداد زیادی درخت تصمیم در زمان آموزش عمل می‌کند. یک جنگل از درختان ساخته شده است و درختان بیشتر به معنای جنگل مقاوم‌تر است. به‌طور مشابه، الگوریتم درخت تصادفی روی نمونه‌های داده، درختان تصمیم‌گیری می‌سازد و سپس از هر کدام از آن‌ها پیش‌بینی می‌گیرد و درنهایت، به‌واسطه رأی‌گیری، بهترین راه‌حل را انتخاب می‌کند (آتول و مونوگان[5]، ۲۰۱۵). این یک روش گروهی است که از یک درخت تصمیم‌گیری مجزا بهتر است، زیرا با میانگین‌گیری، بیش برازش را کاهش می‌دهد (جونز و همکاران، ۲۰۱۷).

تقویت گرادیان

هدف تقویت گرادیان تنظیم مجموعه‌ای از یادگیرندگان ضعیف به یک یادگیرنده قوی و درنتیجه، بهبود نتایج پیش‌بینی است. با آموزش یادگیرندگان ضعیف با خطاهای بین مشاهدات پیش‌بینی‌شده و واقعی، یک تابع ضرر را بهینه می‌کند، سپس با ایجاد یک سری، یادگیرندگان را به‌صورت افزایشی گسترش

1. sezer et al
2. Ozbayoglu et al
3. Breiman et al
4. oz et al
5. Attewell & Monaghan

می‌دهد که در آن هر یادگیرنده، آموزش‌دیده و خطای سری یادگیرنده قبلی را تنظیم می‌کند. مجموعه یادگیرنده ضعیف گسترش می‌یابد تا زمانی که پیش‌بینی رضایت‌بخش حاصل شود؛ بنابراین، کیفیت طبقه‌بندی ضعیف برای عملکرد پیش‌بینی تعیین‌کننده است (شاپیره و فروند[1]، ۲۰۱۲). تقویت گرادیان به دلیل تکرار تصادفی زیرنمونه از نمونه آموزشی برای تناسب با یادگیرنده، بررسی خوبی در مورد مشکل بیش برازش ارائه می‌دهد (انوراوز و همکاران،۲۰۲۱).

تقویت گرادیان فوق‌العاده

تقویت گرادیان فوق‌العاده یک یادگیری تحت نظارت است که می‌تواند برای مشکلات رگرسیون و طبقه‌بندی استفاده شود. مفهوم تقویت گرادیان فوق‌العاده، ترکیبی از نزول گرادیان و یادگیری گروه درختی است. از طریق آموزش افزودنی، هر بار یک درخت جدید به مدل قبلی تقویت گرادیان فوق‌العاده اضافه می‌شود تا عملکرد پیش‌بینی آن را بهبود بخشد و مقدار یک تابع هدف از پیش تعیین‌شده را بهینه کند (شیمین و همکاران[2]، ۲۰۲۰). درمجموع ۵ فرا پارامتر برای مدل تقویت گرادیان فوق‌العاده وجود دارد: نرخ یادگیری، حداکثر عمق، درصد نمونه‌برداری فرعی برای نمونه‌ها، درصد نمونه‌برداری فرعی برای ویژگی‌ها و حداکثر تعداد تکرار (هانگ و ین، ۲۰۱۹).

رگرسیون بردار پشتیبان

در یادگیری ماشینی، ماشین‌های بردار پشتیبان، مدل‌های یادگیری تحت نظارت با الگوریتم‌های یادگیری مرتبط هستند که داده‌ها را برای طبقه‌بندی و تحلیل رگرسیون تجزیه‌وتحلیل می‌کنند. ماشین بردار پشتیبان همچنین می‌تواند به‌عنوان یک روش رگرسیون استفاده شود و تمام ویژگی‌های اصلی را که الگوریتم را مشخص می‌کند (حداکثر حاشیه) حفظ می‌کند (کورتس و واپنیک[3]، ۱۹۹۵). رگرسیون بردار پشتیبان (SVR) از همان اصول ماشین بردار پشتیبان (SVM) برای طبقه‌بندی استفاده می‌کند، تنها با چند تفاوت جزئی. هدف آن یافتن بهترین مرز یا ابر صفحه است که داده‌ها را به کلاس‌های مختلف جدا می‌کند (آتول و مونوگان، ۲۰۱۵). در رگرسیون بردار پشتیبان، هدف به حداقل رساندن خطا بین مقادیر پیش‌بینی‌شده و واقعی متغیر هدف است (جونز و همکاران،۲۰۱۷).

K نزدیک‌ترین همسایه

الگوریتم k-نزدیک‌ترین همسایه یک الگوریتم یادگیری ماشین نظارت‌شده ساده و آسان است که می‌تواند برای حل مسائل رگرسیون و طبقه‌بندی استفاده شود. الگوریتم k نزدیک‌ترین همسایه فرض می‌کند که چیزهای مشابه در نزدیکی هم وجود دارند. k-نزدیک‌ترین همسایه با محاسبه فاصله بین یک نقطه داده

1. Schapire & Freund
2. shimin et al
3. Cortes & Vapnik

جدید و تمام نقاط داده در مجموعه داده آموزشی کار می‌کند. متریک فاصله مورد استفاده در این محاسبه می‌تواند فاصله اقلیدسی یا منهتن باشد. (اولوداگ و گورسوی[1]، ۲۰۲۰). برای رگرسیون، پیش‌بینی میانگین K نزدیک‌ترین همسایه‌ها است. k-نزدیک‌ترین همسایه یک الگوریتم ساده و سریع است، اما می‌تواند به ویژگی‌های نامربوط حساس باشد و ممکن است در داده‌های با ابعاد بالا عملکرد خوبی نداشته باشد (جونز و همکاران، ۲۰۱۷).

رگرسیون خطی چند متغیره

این روش نسخه توسعه‌یافته رگرسیون خطی ساده است. رگرسیون خطی چندگانه یک مدل رگرسیونی است که رابطه بین یک متغیر وابسته کمی و دو یا چند متغیر مستقل را با استفاده از یک خط مستقیم تخمین می‌زند. این شامل یافتن بهترین خط برازش (خط رگرسیون) است که بتواند متغیر وابسته را بر اساس مقادیر متغیرهای مستقل پیش‌بینی کند. (اوزلم و تان، ۲۰۲۲).

مدل خطی تعمیم‌یافته

در آمار، یک مدل خطی تعمیم‌یافته یک تعمیم انعطاف‌پذیر از رگرسیون خطی معمولی است. مدل خطی تعمیم‌یافته، تعمیم رگرسیون خطی است برای داده‌هایی که توزیع نرمال ندارند. ایده اصلی پشت یک مدل خطی تعمیم‌یافته، مدل‌سازی میانگین متغیر وابسته به‌عنوان ترکیبی خطی از متغیرهای مستقل است، اما به جای فرض توزیع نرمال، توزیع متفاوتی از خانواده نمایی مانند دوجمله‌ای، پواسون یا نمایی را در نظر می‌گیرد (دیوید[2]، ۲۰۱۵). تابع پیوند، که رابطه بین میانگین متغیر وابسته و ترکیب خطی متغیرهای مستقل را مشخص می‌کند، جزء مهم مدل خطی تعمیم‌یافته است. انتخاب تابع پیوند به توزیع متغیر وابسته بستگی دارد. پارامترهای مدل با تخمین حداکثر درستنمایی تخمین زده می‌شوند (جونز و همکاران، ۲۰۱۷).

تجزیه‌وتحلیل داده‌ها و آزمون فرضیه‌ها

آمار توصیفی

برای مقابله با اثرات داده‌های پرت از تکنیک وینسورایز (در سطح ۱ و ۹۹ درصد) استفاده شده است. یکی از کاربردهای مهم جدول آمار توصیفی بیان نوع توزیع داده‌ها است. با توجه به آماره آزمون جارک‌برا هیچ‌یک از متغیرهای این پژوهش دارای توزیع نرمال نمی‌باشند. نسبت جاری دارای مقدار میانه ۱/۳۱ است که نشان می‌دهد دارایی جاری نسبت به بدهی جاری در نیمی از شرکت‌های مورد بررسی بورس تهران ۳۱ درصد بیشتر است که ممکن است بتوان گفت شرکت‌های مورد بررسی با ریسک نقدینگی مواجهه می‌باشند. مقدار میانه در بازده دارایی‌ها و حاشیه سود خالص به ترتیب برابر ۰/۱۰ و ۰/۱۲ است که نشان می‌دهد

1. Uludag & Gursoy
2. David

شرکت‌های بورس تهران درنهایت سود ساز هستند. همچنین انحراف معیار بازده دارایی‌ها ۰/۵۹ و حاشیه سود خالص ۰/۵۵ است که نشان می‌دهد نوسانات دارایی شرکت‌ها از فروش بیشتر است. نتایج حاصل از آمار توصیفی متغیرهای پژوهش در جدول شماره ۲ گزارش شده است.

جدول ۲. تحلیل توصیفی متغیرهای پژوهش

متغیر	میانگین	میانه	بیشینه	کمینه	انحراف معیار	چولگی	کشیدگی	آماره جارک برا
شاخص عدم قطعیت ایران	۰/۲۲۴	۰/۲۲۸	۰/۴۱۹	۰/۰۰۰	۰/۱۳۳	-۰/۲۷۶	۱/۹۲۹	۱۲۵/۷۰۲
رشد تولید ناخالص داخلی	۰/۰۱۳	۰/۰۱۶	۰/۱۴۲	-۰/۰۸۶	۰/۰۵۵	۰/۴۴۵	۳/۵۲۵	۹۲/۴۵۲
نرخ بهره	۰/۱۶۸	۰/۱۶۰	۰/۲۳۰	۰/۱۴۰	۰/۰۲۴	۱/۲۹۲	۳/۹۵۷	۶۵۶/۸۶۷
نرخ تورم	۰/۲۲۹	۰/۲۴۲	۰/۴۰۲	۰/۰۶۹	۰/۱۱۵	-۰/۰۰۹	۱/۴۷۲	۲۰۲/۰۳۲
نرخ رشد نفت	۰/۰۶۰	-۰/۰۰۳	۰/۸۳۱	-۰/۴۶۵	۰/۳۲۷	۰/۶۴۱	۳/۳۷۱	۱۵۴/۱۲۷
سن شرکت	۳/۶۱۸	۳/۷۱۴	۴/۲۶۳	۲/۱۹۷	۰/۳۸۳	-۰/۷۰۵	۲/۷۴۳	۱۷۷/۸۸۴
وجه نقد	۰/۰۷۱	۰/۰۴۲	۰/۴۳۷	۰/۰۰۲	۰/۰۸۴	۲/۲۸۶	۸/۶۲۵	۴۵۴۴/۲۹۸
سود تقسیمی پرداختی	۰/۰۵۵	۰/۰۲۴	۰/۳۷۸	۰/۰۰۰	۰/۰۷۶	۲/۰۷۸	۷/۵۵۲	۳۲۸۷/۰۲۲
مخارج سرمایه ای	۰/۰۶۴	۰/۰۲۹	۰/۵۵۵	-۰/۰۶۲	۰/۱۰۲	۲/۸۵۰	۱۱/۹۴۷	۹۷۳۳/۵۴۶
درآمد قبل از مالیات بعلاوه استهلاک	۰/۱۷۹	۰/۱۵۶	۰/۵۶۹	-۰/۱۸۷	۰/۱۴۶	۰/۴۱۰	۳/۰۷۹	۵۸/۸۲۶
هزینه مالی	۰/۰۳۴	۰/۰۲۶	۰/۱۴۳	۰/۰۰۰	۰/۰۳۳	۱/۲۴۱	۴/۱۹۰	۶۵۵/۷۵۸
سرمایه در گردش غیر نقد	۰/۱۰۴	۰/۱۱۳	۰/۶۴۱	-۰/۵۳۰	۰/۲۲۶	-۰/۲۸۴	۳/۲۱۴	۳۱/۹۴۲
خالص دارایی ثابت مشهود	۰/۲۶۴	۰/۲۲۲	۰/۷۹۳	۰/۰۲۰	۰/۱۸۲	۰/۹۰۸	۳/۲۲۷	۲۸۹/۶۳۴
بدهی کوتاه مدت	۰/۵۰۸	۰/۵۰۴	۱/۰۲۵	۰/۰۹۹	۰/۱۹۷	۰/۱۴۸	۲/۵۷۲	۲۳/۴۱۴
بازده دارایی ها	۰/۱۳۲	۰/۱۰۷	۰/۵۵۹	-۰/۲۳۰	۰/۱۴۸	۰/۵۹۳	۳/۴۵۱	۱۳۹/۳۰۱
حساب دریافتنی	۰/۲۹۱	۰/۲۶۱	۰/۷۵۱	۰/۰۱۷	۰/۱۷۵	۰/۶۰۷	۲/۷۲۱	۱۳۴/۴۱۴
حساب پرداختنی	۰/۲۱۴	۰/۱۷۳	۰/۷۰۰	۰/۰۲۱	۰/۱۵۰	۱/۱۱۴	۳/۷۹۱	۴۸۳/۵۵۰

متغیر	میانگین	میانه	بیشینه	کمینه	انحراف معیار	چولگی	کشیدگی	آماره جارک برا
سود هر سهم	۹۷۸/۷۴	۴۹۳/۶۴	۸۱۰۵	-۱۹۳۱	۱۵۲۵/۰	۲/۲۹۴	۹/۸۳۷	۵۸۶۵/۲۷۱
سود عملیاتی پس از کسر استهلاک	۰/۱۳۶	۰/۱۱۶	۰/۴۸۶	-۰/۱۹۹	۰/۱۲۸	۰/۴۰۸	۳/۲۵۵	۶۳/۰۸۶
سود قبل از مالیات	۰/۱۸۲	۰/۱۵۶	۰/۷۰۷	-۰/۴۸۳	۰/۱۹۹	-۰/۰۹۴	۴/۱۷۴	۱۲۲/۲۳۸
نرخ رشد فروش	۰/۳۱۰	۰/۲۴۵	۱/۱۵۲	-۰/۲۶۹	۰/۳۷۸	۰/۵۸۴	۲/۶۵۹	۱۲۸/۰۵۸
بازده حقوق صاحبان سهام	۰/۲۹۷	۰/۲۸۷	۰/۷۵۸	-۰/۲۱۱	۰/۲۵۶	-۰/۰۳۰	۲/۲۶۹	۴۶/۴۷۰
نسبت جاری	۱/۴۷۰	۱/۳۱۲	۳/۱۹۶	۰/۵۹۲	۰/۶۶۶	۱/۰۸۳	۳/۶۴۰	۴۴۱/۲۵۷
حاشیه سود خالص	۰/۱۶۱	۰/۱۲۳	۰/۵۴۸	-۰/۱۴۱	۰/۱۷۶	۰/۵۵۱	۲/۶۷۰	۱۱۴/۳۴۸
اندازه شرکت	۱۴/۵۵۳	۱۴/۳۳۴	۲۱/۳۰۴	۱۰/۰۳۱	۱/۷۰۵	۰/۶۸۴	۳/۷۵۷	۲۱۱/۴۳۳

منبع: یافته‌های پژوهش

برآورد مدل‌های پژوهش

به‌منظور برازش مدل‌های پژوهش در مرحله اول از کل مجموعه متغیرهای پژوهش و سپس، با استفاده از معیار اطلاعاتی بیز از مجموعه متغیرهای انتخاب شده توسط رگرسیون لاسو استفاده شده است. به‌منظور مقایسه مدل‌ها نیز معیارهای آماری مجذور میانگین مربعات خطا و ضریب دقت محاسبه و گزارش شده است.

مطابق با جدول شماره ۳ نتایج حاصل از برازش مدل‌ها بدون استفاده از رویکرد انتخاب ویژگی رگرسیون لاسو نشان داد بیشترین ضریب دقت به ترتیب مربوط به مدل‌های رگرسیون نمادین با استفاده از الگوریتم ژنتیک، رگرسیون مارس، رگرسیون بردار پشتیبان، تقویت گرادیان درختی، شبکه عصبی و تقویت گرادیان فوق‌العاده است. سایر مدل‌ها دارای ضریب دقت کمتر از ۵۰ درصد هستند که عملکرد ضعیفی را نشان داده و در رتبه‌های بعدی قرار گرفته‌اند. همچنین کمترین مربعات خطا به ترتیب مربوط به رگرسیون نمادین، تقویت گرادیان درختی و شبکه عصبی است سایر مدل‌ها در رتبه‌های بعدی قرار می‌گیرند. متغیرهای بااهمیت از منظر مدل رگرسیون نمادین نسبت جاری، سرمایه در گردش غیر نقدی و بدهی‌های کوتاه‌مدت است. متغیرهای با اهمیت از منظر تقویت گرادیان درختی به ترتیب عبارت است از بازده دارایی، نسبت جاری، حساب دریافتنی، سرمایه در گردش غیر نقدی، نسبت دارایی ثابت به کل دارایی، سود تقسیمی پرداختی، حاشیه سود خالص، حساب پرداختنی، سن شرکت و بدهی کوتاه‌مدت. متغیرهای بااهمیت ازنظر رگرسیون مارس، سرمایه در گردش غیرنقدی، نسبت جاری، بدهی کوتاه‌مدت، حساب

دریافتنی، بازده دارایی، حاشیه سود خالص، سود عملیاتی پس از کسر مالیات، سود قبل از مالیات بعلاوه استهلاک و حساب پرداختنی است.

جدول ۳. نتایج تخمین مدل‌ها بدون انتخاب ویژگی

مدل	ضریب دقت (R^2)	رتبه دقت	مجذور میانگین مربعات خطا (RMSE)	رتبه خطا
رگرسیون خطی چندگانه	۰/۳۸۱۰	۱۰	۰/۰۶۳۰	۷
مدل خطی تعمیم‌یافته	۰/۴۳۵۰	۸	۰/۰۶۸۲	۹
یادگیری عمیق	۰/۴۴۷۰	۷	۰/۰۶۵۰	۸
درخت تصمیم	۰/۳۰۲۰	۱۱	۰/۰۷۵۴	۱۱
جنگل تصادفی	۰/۴۱۷۰	۹	۰/۰۷۱۰	۱۰
تقویت گرادیان درختی	۰/۵۶۲۰	۴	۰/۰۵۵۷	۲
تقویت گرادیان فوق‌العاده	۰/۵۰۱۰	۶	۰/۰۵۷۰	۴
رگرسیون بردار پشتیبان	۰/۵۸۸۰	۳	۰/۰۵۸۱	۵
K نزدیک‌ترین همسایه	۰/۲۰۱۰	۱۲	۰/۰۸۶۰	۱۲
رگرسیون نمادین	۰/۷۰۹۶	۱	۰/۰۲۱۴	۱
رگرسیون مارس	۰/۵۹۲۰	۲	۰/۰۵۸۸	۶
شبکه عصبی	۰/۵۲۴۰	۵	۰/۰۵۶۰	۳

منبع: یافته‌های پژوهش

روش لاسو پارامترهای مدل را با کوچک کردن ضرایب رگرسیون منظم می‌کند و برخی از آن‌ها را به صفر می‌رساند. مرحله انتخاب ویژگی پس از انقباض رخ می‌دهد، جایی که هر مقدار غیر صفر برای استفاده در مدل انتخاب شده است (الینگتون[1] و همکاران، ۲۰۲۲). نتایج مجموعه متغیرهای انتخاب شده توسط رگرسیون لاسو در جدول ۴ گزارش شده است.

1. Ellington

جدول ۴. نتایج انتخاب ویژگی با استفاده از رگرسیون لاسو

شاخص عدم قطعیت	نرخ رشد نفت	درآمد قبل از مالیات بعلاوه استهلاک	بدهی کوتاه‌مدت	سود هر سهم
رشد تولید ناخالص داخلی	سن شرکت	هزینه مالی	بازده دارایی‌ها	بازده حقوق صاحبان سهام
نرخ بهره	سود تقسیمی پرداختی	سرمایه در گردش غیر نقد	حساب دریافتنی	نسبت جاری
تغییرات نرخ رشد فروش	مخارج سرمایه‌ای	خالص دارایی ثابت مشهود	حساب پرداختنی	اندازه شرکت

منبع: یافته‌های پژوهش

مطابق با جدول شماره ۵ نتایج حاصل از تخمین مدل‌ها با استفاده از رویکرد انتخاب ویژگی رگرسیون لاسو نشان داد که بیشترین ضریب دقت مربوط به مدل‌های رگرسیون نمادین، رگرسیون مارس و تقویت گرادیان درختی است و سایر مدل‌ها در رتبه بعدی قرار می‌گیرند. همچنین کمترین مربعات خطا به ترتیب مربوط به رگرسیون نمادین، شبکه عصبی، تقویت گرادیان درختی و رگرسیون مارس است و سایر مدل‌ها در رتبه بعدی قرار می‌گیرند. متغیرهای بااهمیت از منظر رگرسیون نمادین عبارت است از مخارج سرمایه‌ای، نسبت دارایی ثابت به کل دارایی، نسبت جاری، بدهی کوتاه‌مدت و سرمایه در گردش غیر نقدی. متغیرهای بااهمیت رگرسیون مارس عبارت است از سرمایه در گردش غیر نقدی، نسبت جاری، حساب دریافتنی، بدهی کوتاه‌مدت، بازده دارایی‌ها، نسبت دارایی ثابت به کل دارایی، اندازه شرکت، تورم، سود تقسیمی پرداختی و سود قبل از مالیات بعلاوه استهلاک. علاوه بر این نتایج این پژوهش نشان داد که استفاده از رگرسیون لاسو موجب افزایش ضریب دقت در مدل‌های رگرسیون خطی چندگانه، مدل خطی تعمیم‌یافته و یادگیری عمیق شده است. همچنین استفاده از رگرسیون لاسو موجب کاهش مجذور میانگین مربعات خطا در جنگل تصادفی، رگرسیون بردار پشتیبان، تقویت گرادیان درختی و تقویت گرادیان فوق العاده شده است.

جدول ۵. نتایج تخمین مدل‌ها با استفاده از انتخاب ویژگی رگرسیون لاسو

مدل	ضریب دقت (R^2)	رتبه دقت	مجذور میانگین مربعات خطا (RMSE)	رتبه خطا
رگرسیون خطی چندگانه	۰/۳۸۲۰	۱۰	۰/۰۶۳۰	۸
مدل خطی تعمیم‌یافته	۰/۴۵۳۹	۸	۰/۰۶۳۵	۹
یادگیری عمیق	۰/۴۶۵۵	۷	۰/۰۶۱۰	۷
درخت تصمیم	۰/۲۹۹۳	۱۱	۰/۰۷۴۷	۱۱

مدل	ضریب دقت (R^2)	رتبه دقت	مجذور میانگین مربعات خطا (RMSE)	رتبه خطا
جنگل تصادفی	۰/۳۸۲۳	۹	۰/۰۷۰۹	۱۰
تقویت گرادیان درختی	۰/۵۱۴۵	۳	۰/۰۵۸۴	۳
تقویت گرادیان فوق‌العاده	۰/۴۷۱۰	۶	۰/۰۵۹۱	۵
رگرسیون بردار پشتیبان	۰/۵۰۹۰	۴	۰/۰۵۹۲	۶
K نزدیک‌ترین همسایه	۰/۲۰۱۰	۱۲	۰/۰۸۶۰	۱۲
رگرسیون نمادین	۰/۶۸۵۹	۱	۰/۰۲۲۳	۱
رگرسیون مارس	۰/۵۲۲۷	۲	۰/۰۵۸۶	۴
شبکه عصبی	۰/۴۹۶۰	۵	۰/۰۵۷۰	۲

منبع: یافته‌های پژوهش

بحث و نتیجه‌گیری

تعیین میزان نگهداری وجه نقد یکی از موضوعات مهم در ادبیات مالی شرکتی است. به‌خصوص پس از مقاله اصلی (اوپلیر و همکاران، ۱۹۹۹)، حجم وسیعی از ادبیات تلاش می‌کند تا عوامل تعیین‌کننده نگهداری وجه نقد را شناسایی کند. اگرچه وجه نقد یک دارایی بسیارمهم یک شرکت است، اما کمترین بازدهی را در بین سایر دارایی‌ها ارائه می‌دهد. بااین‌وجود، شرکت‌ها کل وجوه نقد را در دارایی‌های با بازده بالاتر سرمایه‌گذاری نمی‌کنند (وو و همکاران، ۲۰۲۱). بنابراین، پیش‌بینی مقدار نگهداشت وجه نقد از مسائلی است که ذی‌نفعان یک بنگاه تجاری با آن مواجهه هستند و به هر میزان که این پیش‌بینی با دقت بیشتری انجام شود می‌تواند موجب تصمیمات بهینه از سمت مدیریت و سایر ذی‌نفعان شود.

با توجه به فرضیه اول پژوهش که بیانگر این موضوع است که مدل یادگیری ماشین دقت بیشتری نسبت به مدل آماری در پیش‌بینی نگهداشت وجه نقد دارد، نتایج برازش مدل‌ها نشان داد که مدل یادگیری ماشین رگرسیون نمادین با استفاده از الگوریتم ژنتیک با مقدار ضریب دقت ۷۰درصد بهترین عملکرد را در میان مدل‌های آماری و یادگیری ماشین دارد که ممکن است ناشی از این باشد که برای کاوش در فضای وسیعی از مدل‌های ممکن طراحی شده است، بنابراین احتمال یافتن مدلی که به‌خوبی با داده‌ها مطابقت داشته باشد بسیار زیاد است. الگوریتم یادگیری ماشین رگرسیون مارس در رتبه دوم قرار می‌گیرد که می‌تواند به علت انعطاف‌پذیری بالای این مدل در مقابل داده‌های ورودی باشد. الگوریتم‌های جدیدتر و تقویتی یادگیری ماشین یعنی تقویت گرادیان درختی و فوق‌العاده نیز همان‌طور که با توجه به فرآیند

تقویتی این مدل‌ها و پیشینه پژوهش مورد انتظار بود از اکثریت مدل‌های یادگیری ماشین دقت بیشتری را از خود نشان دادند اما پیچیدگی بالای این مدل‌ها منجر به دقت بیشتر چشمگیری نسبت به مدل‌های با پیچیدگی متوسط که در وسط طیف مدل‌های مورد بررسی پژوهش حاضر قرار می‌گیرند نشد و قرارگرفتن آن‌ها در رتبه‌های بعد از رگرسیون نمادین و مارس ممکن است به دلیل حجم کم داده‌ها در پژوهش حاضر باشد. در میان دو مدل آماری مورد بررسی در پژوهش حاضر، مدل خطی تعمیم‌یافته ضریب دقت بیشتری را نسبت به مدل رگرسیون خطی به خود اختصاص داد که احتمالاً به دلیل تناسب بیشتر این مدل با داده‌های دارای توزیع غیرنرمال است. همچنین، اگرچه اکثریت مدل‌های یادگیری ماشین از مدل‌های آماری عملکرد بالاتری داشتند، اما برخی الگوریتم‌های یادگیری ماشین از مدل‌های آماری نیز دقت کمتری را به خود اختصاص دادند. ساختار ساده مدل‌های پیش‌بینی آماری ممکن است به این روش‌های پارامتری اجازه دهد تا به دقت پیش‌بینی مناسب و بالاتری در مقایسه با برخی مدل‌های یادگیری ماشین دست یابند(جونز و همکاران، ۲۰۱۷). مدل k نزدیک‌ترین همسایه کمترین ضریب دقت را در میان مدل‌های آماری و یادگیری ماشین به خود اختصاص داد، این می‌تواند به این دلیل باشد که حوزه مالی و حسابداری و به‌طور خاص پیش‌بینی نگهداشت وجه نقد با داده‌های با ابعاد بالا روبه‌رو است و این الگوریتم در برابر داده‌های با ابعاد بالا، با ضعف عملکرد روبه‌رو است. نتایج فرضیه اول با پژوهش‌های (اوزلم و تان، ۲۰۲۲) و (وو و همکاران، ۲۰۲۱) هم‌راستا بود. در رابطه با فرضیه دوم پژوهش که بیانگر این موضوع است که استفاده از رویکرد انتخاب ویژگی رگرسیون لاسو موجب افزایش دقت در مدل‌های آماری و یادگیری ماشین می‌شود، نتایج نشان داد که استفاده از رویکرد رگرسیون لاسو موجب افزایش دقت مدل‌های آماری رگرسیون خطی و مدل خطی تعمیم‌یافته می‌شود که با پژوهش(الیاسیانی و موقاری، ۲۰۲۲) هم‌راستا بود. همچنین رگرسیون لاسو باعث افزایش دقت در مدل یادگیری عمیق از مدل‌های یادگیری ماشین شد.

بنابراین با توجه به نتایج فرضیه اول پژوهش که مدل‌های یادگیری ماشین توانستند با دقت بیشتری نگهداشت وجه نقد را پیش‌بینی کنند، به مدیران و سایر ذی‌نفعان بنگاه‌های اقتصادی پیشنهاد می‌شود که از مدل‌های یادگیری ماشین و به‌طور خاص مدل رگرسیون نمادین، رگرسیون مارس، رگرسیون بردار پشتیبان و الگوریتم‌های تقویت گرادیان درختی و فوق‌العاده در جهت پیش‌بینی نگهداشت وجه نقد استفاده کنند. همچنین با توجه به بهبود نتایج با استفاده از رگرسیون لاسو در مدل‌های آماری و برخی از مدل‌های یادگیری ماشین و به‌ویژه مدل یادگیری عمیق پیشنهاد می‌شود به کاهش تعداد متغیرها از طریق این رویکرد بپردازند.

در پایان به پژوهشگران آتی پیشنهاد می‌شود که مدل‌سازی را با استفاده از تحلیل مؤلفه‌های اصلی و سایر تکنیک‌های انتخاب ویژگی جهت بهبود نتایج مورد بررسی و مقایسه قرار دهند. همچنین می‌توان عملکرد مدل‌ها را در حجم نمونه‌های مختلف مورد مقایسه قرارداد تا تأثیر حجم داده‌ها در عملکرد مدل‌ها مشخص شود.

ملاحظات اخلاقی

حامی مالی: مقاله حامی مالی ندارد.
مشارکت نویسندگان: تمام نویسندگان در آماده‌سازی مقاله مشارکت داشته‌اند.
تعارض منافع: بنا بر اظهار نویسندگان در این مقاله هیچ‌گونه تعارض منافعی وجود ندارد.
تعهد کپی‌رایت: طبق تعهد نویسندگان حق کپی‌رایت رعایت شده‌است.

References

Acharya, V. V; Almeida, H; & Campello, M. (2007). Is cash negative debt? A hedging perspective on corporate financial policies. *Journal of financial intermediation*, 16(4), 515-554.

Aflatooni, Abbas, Kazemi, Periyosh, Khatiri, Mohammad. (2022). Comparing the Cash Holdings Speed of Adjustment during Economic Prosperities and Recessions. *Financial Management Strategy*, 10(3), 141-160. (In Persian).

Anand, V; Brunner, R; Ikegwu, K; & Sougiannis, T. (2019). Predicting profitability using machine learning. Available at SSRN 3466478.

Angelovska M, Valentinčič A (2019) Determinants of cash holdings in private firms: the case of the Slovenian SMEs. *Econ Bus Rev* 22(1):5–36.

Asgharpour, Hossein, Rezaei, Sadegh, Hamidi Rozi, Daud, Heydari, Mansour. (2022). Investigating the Interaction Effects of Exchange Rate Regimes and Inflation on Iran's Economic Growth. *Business Journal*, 26(104), 47-74. (In Persian).

Attewell, P; & Monaghan, D. (2015). Data mining for the social sciences: An introduction. Univ of California Press.

Ball, R; & Shivakumar, L. (2005). Earnings quality in UK private firms: comparative loss recognition timeliness. *Journal of accounting and economics*, *39*(1), 83-128.

Barboza, F; Kimura, H. and Altman, E. (2017). Machine learning models and bankruptcy prediction, *Expert Systems with Applications* 83: 405–417.

Bates TW, Kahle KM, Stulz RM (2009) why do U.S. firms hold so much more cash than they used to? *J Finance* 64(5):1985–2021.

Bhuiyan MBU, Hooks J (2019). Cash holding and over-investment behavior in firms with problem directors. *Int Rev Econ Financial* 61:35–51.

Bigelli, M; & Sánchez-Vidal, J. (2012). Cash holdings in private firms. *Journal of Banking & Finance*, *36*(1), 26-35.

Boubakri N, Ghoul S, Saffar W (2013). Cash holdings of politically connected firms. *J Multinatl Finance Manag* 23(4):338–355.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24(6), 2350-2383.

Breiman, L; Friedman, J; Olshen, R; & Stone, C. (1984). Classification and regression trees. Wadsworth Int. Group, 37(15), 237-251.

Campello M, Graham JR, Harvey CR (2010). The real effects of financial constraints: evidence from a financial crisis. J Financial Econ 97(3):470–487.

Chen D, Li S, Xiao JZ, Zou H (2014). The effect of government quality on corporate cash holdings. *J Corp Finance* 27:384–400.

Chen, Y. J; Lin, J. A; Chen, Y. M; & Wu, J. H. (2019). Financial forecasting with multivariate adaptive regression splines and queen genetic algorithm-support vector regression. IEEE Access, 7, 112931-112938.

Claveria, O; Monte, E; & Torra, S. (2017). Assessment of the effect of the financial crisis on agents' expectations through symbolic regression. Applied Economics Letters, 24(9), 648-652.

Cortes, C; & Vapnik, V. (1995). Support-vector networks. Machine learning, 20, 273-297.

Dastile, X; Celik, T; & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263.

David, M. (2015). Auto insurance premium calculation using generalized linear models. Procedia Economics and Finance, 20, 147-156.

Diaw A (2021). Corporate cash holdings in emerging markets. *Borsa Istanbul Rev* 21(2) 139–148.

Ditmar, A.; Mahrt-smith, j. & servaes, H. (2003). International Corporate Governance and Corporate Cash Holdings. *Journal of Financial and Quantitative Analysis*, 38(1), pp:111-133.

Donepudi PK, Banu MH, Khan W, Neogy TP, Asadullah ABM, Ahmed AAA (2020) Artifical intelligence and machine learning in treasury management: *a systematic literature review*. Int J Manag 11(11):13–22.

Ellington, M; Stamatogiannis, M. P; & Zheng, Y. (2022). A study of cross-industry return predictability in the Chinese stock market. *International Review of Financial Analysis*, 83, 102249.

Elyasiani, E; & Movaghari, H. (2022). Determinants of corporate cash holdings: An application of a robust variable selection technique. *International Review of Economics & Finance*, *80*, 967-993.

Elyasiani, E; Jia, J; & Movaghari, H. (2019). Determinants of dividend payout and dividend propensity in an emerging market, Iran: an application of the LASSO. *Applied Economics*, 51(42), 4576-4596.

Faraji Tabrizi, Arshiya, Hejbar Kiani, Kambyz, Mimar Nejad, Abbas, Ghafari, Farhad (2021). Investigation of the Affecting on the Gross Domestic Product of Selected Countries with Emphasis on the Role of Exchange Rate; ARDL-PMG Approach. *Economic growth and development research*. (In Persian).

Foley CF, Hartzell JC, Titman S, Twite G (2007) Why do firms hold so much cash? A tax-based explanation. *J Financial Econ* 86(3):579–607.

García-Teruel, P. J; Martínez-Solano, P; & Sánchez-Ballesta, J. P. (2009). Accrual's quality and corporate cash holdings. *Accounting & Finance*, 49(1), 95-115.

Harford, J; Mansi, S. A; & Maxwell, W. F. (2008). Corporate governance and firm cash holdings in the US. *Journal of financial economics*, 87(3), 535-555.

Hastie, T; Tibshirani, R; Friedman, J. H; & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

James, G; Witten, D; Hastie, T; & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

Jones, S; Johnstone, D; & Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting*, 44(1-2), 3-34.

Kim, J. B; Lee, J. J; & Park, J. C. (2015). Audit quality and the market value of cash holdings: The case of office-level auditor industry specialization. Auditing: *A Journal of Practice & Theory*, 34(2), 27-57.

Lerner, A. P. (1936). Mr. Keynes general theory of employment, interest and money. *Int'l Lab. Rev*; 34, 435.

Li, F. (2010). The information content of forward-looking statements in corporate filingsa na¨ıve bayesian machine learning approach, *Journal of Accounting Research* 48(5): 1049– 1102.

Liu, H; & Zhang, Z. (2022). Probing the carbon emissions in 30 regions of China based on symbolic regression and Tapio decoupling. *Environmental Science and Pollution Research*, 29(2), 2650-2663.

Lozano MB, Yaman S (2020) The European financial crisis and firms' cash holding policy: an analysis of the precautionary motive. Glob Pol 11(S1):84–94.

Maleki, Atefe, Jalalinia, Saeed, Hamzaei, Asghar (2022). The relationship between CEO experience and cash holding levels. *Accounting and Management Perspective*, 5(67), 10-1. (In Persian).

Manoel AAS, Moraes MBC, Santos DFL, Neves MF (2018) Determinants of corporate cash holdings in times of crisis: insights from Brazilian sugarcane industry private firms. *Int Food Agribus Manag Rev* 21(2):201–217.

Miller MH, Orr D (1966). A model of the demand for money by firms. *Q J Econ* 80(3):413–435.

Mulai, Rahim (2022). The effect of accounting information quality on cash retention with emphasis on inflation. *Accounting and Management Perspective*, 5(61), 103-114. (In Persian).

Mullainathan, S; & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.

Opler, T; Pinkowitz, L; Stulz, R; & Williamson, R. (1999). The determinants and implications of corporate cash holdings. *Journal of financial economics*, 52(1), 3-46.

Oz, I. O; Yelkenci, T; & Meral, G. (2021). The role of earnings components and machine learning on the revelation of deteriorating firm performance. *International Review of Financial Analysis*, 77, 101797.

Ozbayoglu, A. M; Gudelek, M. U; & Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, 93, 106384.

Ozkan A, Ozkan N (2004). Corporate cash holdings: an empirical investigation of UK companies. *J Bank Finance* 28(9):2103–2134.

Özlem, Ş; & Tan, O. F. (2022). Predicting cash holdings using supervised machine learning algorithms. *Financial Innovation*, 8(1), 1-19.

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics*, 104(1), 162-185.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. Auditing: *A Journal of Practice & Theory*, 30(2), 19-50.

Pourgadimi, K; Bahri Sales, J; Jabbarzadeh Kangarliue, S; & ZavarRezaee, A. (2022). Presenting the developed model of Benish model with emphasis on audit quality features using neural network, vector machine and random forest. *Advances in Finance and Investment*, 3(6), 30-1. (In Persian).

Rafi, M; Wahab, M. T; Khan, M. B; & Raza, H. (2020, January). ATM cash prediction using time series approach. In 2020 3rd International Conference on Computing, *Mathematics and Engineering Technologies (iCoMET)* (pp. 1-6). IEEE.

Rajabzadeh, H; Gorganli‎ davaji, J; naderian, A; & ashrafi, M. (2022). Forecast the operating cash flow of accepted companies In Tehran Stock Exchange using machine learning method. *Management Accounting*. (In Persian).

Ramnath, S; Rock, S; & Shane, P. B. (2008). Financial analysts' forecasts and stock recommendations: A review of the research. *Foundations and Trends® in Finance*, 2(4), 311-421.

Sarfraz M, Shah SGM, Ivascu M, Quereshi MAA (2020). Explicating the impact of hierarchical CEO Sucession on small-medium enterprises' performance and cash holdings. *Int J Financial Econ*.

Schapire, R. E; & Freund, Y. (2012). Boosting: Foundations and Algorithms. 1621 Cambridge, MA.

Schauten MB, Dijk D, van der Wall JP (2011). Corporate governance and the value of excess cash holdings of large European firms. *Eur Financial Manag* 19(5):991–1016.

Sezer, O. B; Gudelek, M. U; & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied soft computing, 90, 106181.

Sezer, O. B; Gudelek, M. U; & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied soft computing, 90, 106181.

Shekarkhah, Javad, Mortezazadeh, Mojtabi (2015). Comparison of cash holding determinants in different industries. *Planning and Budget Quarterly*, 20(1), 67-86. (in persian).

Shimin, L. E. I; Ke, X. U; Huang, Y; & Xinye, S. H. A. (2020). An Xgboost based system for financial fraud detection. In E3S Web of Conferences (Vol. 214, p. 02042). EDP Sciences.

Song, K. R; & Lee, Y. (2012). Long-term effects of a financial crisis: Evidence from cash holdings of East Asian firms. *Journal of Financial and Quantitative analysis*, 47(3), 617-641.

Subramaniam, V; Tang, T. T; Yue, H; & Zhou, X. (2011). Firm structure and corporate cash holdings. *Journal of Corporate Finance*, 17(3), 759-773.

Tian, S; Yu, Y; Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89-100.

ULUDAĞ, O; & GÜRSOY, A. (2020). On the financial situation analysis with KNN and naive Bayes classification algorithms. Journal of the Institute of Science and Technology, 10(4), 2881-2888.

Wahlen, J. M. and Wieland, M. M. (2011). Can financial statement analysis beat consensus analysts' recommendations? *Review of Accounting Studies* 16(1): 89–115.

Wu, H. C; Chen, J. H; & Wang, P. W. (2021). Cash Holdings Prediction Using Decision Tree Algorithms and Comparison with Logistic Regression Model. *Cybernetics and Systems*, 52(8), 689-704.

Wu, X; Wang, Y; & Tong, X. (2021). Cash holdings and oil price uncertainty exposures. *Energy Economics*, *99*, 105303.

Xinyue, C; Zhaoyu, X; & Yue, Z. (2020). Using Machine Learning to Forecast Future Earnings. *Atlantic Economic Journal*, 48(4), 543-545.

Xinyue, C; Zhaoyu, X; & Yue, Z. (2020). Using Machine Learning to Forecast Future Earnings. *Atlantic Economic Journal*, 48(4), 543-545.

Zhang, X; & Zhou, H. (2022). The effect of market competition on corporate cash holdings: An analysis of corporate innovation and financial constraint. *International Review of Financial Analysis*, *82*, 102163.

COPYRIGHTS

This license allows others to download the works and share them with others as long as they credit them, but they can’t change them in any way or use them commercially.