فصلنامه مهندسی مالی و مدیریت اوراق بهادار
دوره پانزده، شماره شصت، پائیز ۱۴۰۳
نوع مقاله: علمی پژوهشی
صفحات: ۱۲۵-۱۰۷

پیش‌بینی نگهداشت وجه نقد با استفاده از الگوریتم‌های یادگیری ماشین نظارت‌شده در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران

سعید فلاح پور[۱]
رضا راعی[۲]
نگار توکلی[۳]

تاریخ دریافت مقاله : ۱۴۰۲/۰۴/۲۱ تاریخ پذیرش مقاله : ۱۴۰۲/۱۰/۲۱

چکیده

این مطالعه با توجه به ۲۲ ویژگی انتخاب شده (که در حین پژوهش بررسی می‌شوند) با روش‌های یادگیری ماشین، نگهداری وجه نقد شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را پیش‌بینی می‌کند. ۲۰۱ شرکت از سال ۱۳۹۶ تا سال ۱۴۰۰ بررسی شد. رگرسیون خطی چندگانه، کی-نزدیک‌ترین همسایه، رگرسیون بردار پشتیبان، درخت تصمیم، جنگل تصادفی، الگوریتم تقویت گرادیان شدید و شبکه‌های عصبی چندلایه برای پیش‌بینی استفاده می‌شود. نتایج نشان می‌دهد که روش‌های رگرسیون خطی چندگانه، کی-نزدیک‌ترین همسایه خطای جذر میانگین مربعات و میانگین قدرمطلق خطا بالا را ارائه می‌دهند. در همین حال، الگوریتم‌های پیچیده‌تر، به خصوص رگرسیون بردار پشتیبان، دقت بالاتری را به دست می‌آورند؛ یافته‌ها حاکی از آن بوده است که با کاهش به ۱۵ متغیر، روش‌های یادگیری ماشین به خصوص کی-نزدیک‌ترین همسایه نتایج بهتری را ارائه دادند. بر مبنای آزمون مقایسه زوجی نیز رگرسیون بردار پشتیبان عملکرد بهتری از سایر الگوریتم‌های یادگیری ماشین نظارت شده به جز درخت تصمیم دارد. همچنین مهم‌ترین متغیرها نیز اندازه شرکت و مخارج سرمایه‌ای به دست آمد. شاخص عدم قطعیت جهانی و تورم نیز از متغیرهایی با اهمیت نسبتاً بالایی بودند؛ بنابراین، با استفاده از الگوریتم رگرسیون بردار پشتیبان، ممکن است میزان وجه نقد را به میزان قابل‌توجهی پیش‌بینی کنیم.

کلمات کلیدی

رگرسیون بردار پشتیبان، الگوریتم تقویت گرادیان شدید، نگهداشت پول نقد، یادگیری ماشین، شاخص عدم قطعیت جهانی

۱-دانشیار، گروه مالی و بیمه، دانشکده مدیریت، دانشگاه تهران، تهران، ایران. falahpor@ut.ac.ir
۲-استاد، گروه مالی و بیمه، دانشکده مدیریت، دانشگاه تهران، تهران، ایران. raei@ut.ac.ir
۳-دانشجوی کارشناسی ارشد، گروه مالی و بیمه، دانشکده مدیریت، دانشگاه تهران، تهران، ایران. (نویسنده مسئول) negartavakoli0721@gmail.com

مقدمه

انگیزه شرکت‌ها برای نگهداری وجه نقد و معادل وجوه نقد چیست؟ به عبارت دیگر، چرا شرکت‌ها از پول نقد خود برای سرمایه‌گذاری مجدد استفاده نمی‌کنند؟ این سؤالات دو مورد از موضوعات مورد بحث در ادبیات مالی شرکت‌ها هستند. شرکت‌ها به طور قابل‌توجهی دارایی‌های نقدی خود را در دو دهه گذشته افزایش داده‌اند، به ویژه به این دلیل که به آن‌ها اجازه می‌دهد تا تغییرات غیرقابل پیش‌بینی جریان نقدی، عملیات تأمین مالی روزانه و تأمین مالی پروژه‌های بلندمدت را مدیریت کنند (اوپلر و همکاران،۱۹۹۹). با این حال، شرکت‌ها باید مقدار مناسبی از پول نقد را نگه دارند. نگه‌داشتن بیش‌ازحد باعث می‌شود که مدیران به دنبال منافع خود بروند و در نتیجه منجر به زیان سهامداران و شاید یک بحران مالی شود. نرخ بازده نگهداری وجه نقد شرکتی معمولاً کمتر از نرخ بهره بازار است و هزینه فرصت نگهداری وجه نقد را افزایش می‌دهد (وو و همکاران، ۲۰۲۱). بر اساس دو رویکرد مختلف، نگهداری یک مقدار نقدینگی بهینه یک موضوع ضروری در امور مالی است. (کینز،۱۹۳۶) در کتاب مشهور خود به نام نظریه عمومی اشتغال، نرخ بهره و پول، سه انگیزه را برای نگهداری وجه نقد برمی‌شمارد که عبارت‌اند از: انگیزه احتیاطی، انگیزه معاملاتی، انگیزه سفته‌بازی. در ادبیات مالی، چهار دسته از انگیزه‌ها برای شرکت‌ها برای نگهداری وجه نقد شناسایی شده است (بیتس و همکاران، ۲۰۰۹): معامله، احتیاط، هزینه نمایندگی و انگیزه مالیاتی. اولاً، شرکت‌هایی که دارای منابع مالی داخلی کافی نیستند، می‌توانند دارایی‌های غیرمالی را به پول نقد تبدیل کنند، سهام و بدهی جدید منتشر کنند یا پرداخت سود سهام را کاهش دهند. با این حال، شرکت‌ها می‌خواهند از هزینه‌های مبادله که انگیزه معامله را ایجاد می‌کند، اجتناب کنند. (میلر و اور، ۱۹۶۶) اعلام کردند که هزینه‌های مبادله می‌تواند یک شرکت را وسوسه کند که دارایی‌های نقدی بیشتری نگهداری کند. انگیزه احتیاطی به ذخایر نقدی اشاره دارد که به عنوان انگیزه در برابر کمبودهای غیرمنتظره یا به دست آوردن فرصت‌های سرمایه‌گذاری سودآور نگهداری می‌شود (بیتس و همکاران، ۲۰۰۹؛ کینز، ۱۹۳۶؛ کیم و همکاران، ۱۹۹۸). مدیران تمایل دارند به جای پرداخت سود سهام به سهامداران هنگام مواجهه با پروژه‌هایی با نرخ سرمایه‌گذاری منفی، پول نقد را حفظ کنند. به این ترتیب، آن‌ها دارایی‌های تحت کنترل خود را افزایش می‌دهند و بر تصمیمات سرمایه‌گذاری شرکت قدرت دارند (جنسن، ۱۹۸۶). هنگامی که شرکت‌ها با مالیات بر بازده بیشتر مواجه می‌شوند، آن‌ها تصمیم می‌گیرند که پول نقد زیادی را در خارج از کشور به عنوان انگیزه مالیاتی نگه دارند (فولی و همکاران، ۲۰۰۷).

برای تعیین رفتار نگهداری وجه نقد شرکت‌ها، مطالعات از متغیرهای مالی مختلف استفاده کرده‌اند.

با روش‌های رگرسیون کلاسیک، تأثیر بسیاری از متغیرهای مالی بر رفتار نگهداری وجه نقد شرکت‌ها بررسی شده است. برخلاف ادبیات قبلی، ما سعی می‌کنیم رفتار نگهداری وجه نقد شرکت‌ها را با استفاده از رویکردهای یادگیری ماشین پیشرفته در ادبیات پیش‌بینی کنیم. تکنیک‌های یادگیری ماشین را می‌توان برای پیش‌بینی و تجزیه و تحلیل به جای گزارش صرف اعداد و آمار استفاده کرد (رافی و همکاران، ۲۰۲۰).

هدف مطالعه حاضر پیش‌بینی نگهداری وجه نقد شرکت‌های ایرانی با استفاده از روش‌های مختلف یادگیری ماشین نظارت‌شده به صورت جداگانه با شروع از روش‌های ساده، مانند رگرسیون خطی چندگانه[۱]، رگرسیون بردار پشتیبان[۲] و کی-نزدیک‌ترین همسایه[۳] و با الگوریتم‌های پیچیده‌تر مانند الگوریتم تقویت گرادیان شدید[۴] و شبکه‌های عصبی چندلایه[۵] ادامه می‌دهیم. تمام روش‌های یادگیری نظارت‌شده در روش تحقیق را بر اساس معیار خطای معیار خطای جذر میانگین مربعات[۶] و میانگین خطا مطلق[۷] ارزیابی می‌کنیم. فرضیه اصلی این پژوهش اینگونه است که مدل ارائه شده این پژوهش، رگرسیون بردار پشتیبان، در پیش‌بینی نگهداشت وجه نقد نسبت به سایر مدل‌های رقیب عملکرد بهتری دارد.

سهم عمده این مطالعه پر کردن شکاف‌های زیر در ادبیات است. اولاً، اکثر مطالعات قبلی از تحلیل رگرسیون برای پیش‌بینی موجودی وجه نقد استفاده کرده‌اند و تنها تعداد کمی از مطالعات از تکنیک‌های یادگیری ماشین استفاده می‌کنند. دوماً، این مطالعه اولین مطالعه‌ای است که نگهداشت وجه نقد را با الگوریتم‌های یادگیری ماشین در ایران پیش‌بینی می‌کند. مدل ما دارای ۱۷ نسبت مالی به اضافه سن و صنعت شرکت‌ها و شاخص عدم قطعیت جهانی[۸] و همه‌گیری[۹] ویژه کشور ایران و همچنین رشد تولید ناخالص داخلی است و برای اولین بار تأثیر نگهداشت وجه نقد شرکت‌ها را بر روی عامل کرونا و سه متغیر آخر سنجیده می‌شود.

مبانی نظری و پیشینه پژوهش

در سال‌های اخیر، الگوریتم‌های یادگیری ماشین در حوزه مالی شرکتی مورد استفاده قرار گرفته‌اند. به عنوان مثال، (وو و همکاران، ۲۰۲۱) با استفاده از درخت مدل لجستیک[۱۰]، جنگل تصادفی[۱۱]، درخت REP، نمودار ساده[۱۲]، درخت اضافی[۱۳] و درخت BF از روش‌های درخت تصمیم[۱۴] برای پیش‌بینی دارایی‌های نقدی صنعت پیشرفته در تایوان استفاده کرد. یافته‌های آن‌ها نشان داد که جنگل تصادفی بهترین نرخ پیش‌بینی را در بین تمام درخت تصمیم دارد. علاوه بر این، (وون و همکاران، ۲۰۱۲) با استفاده از فرآیند گاوسی و شبکه عصبی شعاعی، محدودیت‌های مالی را برای شرکت‌های پذیرفته‌شده

در بورس اوراق بهادار تهران پیش‌بینی کرد. آن‌ها تأیید کردند که روش‌های یادگیری ماشین برای پیش‌بینی محدودیت‌های مالی مناسب هستند. درصد مالکیت نهادی، بازده دارایی‌ها، اهرم مالی، جریان نقدی عملیاتی به دارایی‌ها و ارزش شرکت متغیرهای اصلی در پیش‌بینی محدودیت‌های مالی هستند. (غلام‌زاده و همکاران، ۲۰۲۱)

در همین حال، (موسی و همکاران، ۲۰۲۱) برای پیش‌بینی عملکرد مالی ۶۳ بانک فهرست شده در بازارهای نوظهور از سه روش یادگیری ماشینی نظارت‌شده، یعنی جنگل تصادفی، تجزیه و تحلیل تفکیک درجه دوم و تجزیه و تحلیل تفکیک خطی استفاده کرد. آن‌ها نشان دادند که روش جنگل تصادفی بهترین مدل‌های پیش‌بینی کننده را ارائه می‌دهد و اینکه ترکیب متغیرهای افشاکننده در مدل پیش‌بینی‌کننده با متغیرهای مالی، دقت و کیفیت این مدل‌ها را افزایش می‌دهد.

(پوپسکو و دراگوتا، ۲۰۱۸) بحران‌های مالی و ورشکستگی را با استفاده از مدل‌های مختلف الگوریتم یادگیری ماشین بررسی کردند. در همین حال، (کو و همکاران، ۲۰۱۴) رویکرد پیشنهادی است که از روش‌های تصمیم‌گیری چند معیارِ، k میانگین[۱۵]، انتظار-بیشینه‌سازی، رویکرد تصنیف مکرر، الگوریتم تقسیم‌بندی نمودار و روش‌های مبتنی بر چگالی برای ارزیابی کیفیت الگوریتم‌های خوشه‌بندی در حوزه تحلیل ریسک مالی استفاده می‌کند. یافته‌های آن‌ها نشان می‌دهد که رویکرد تصنیف مکرر از سایر الگوریتم‌های خوشه‌بندی منتخب بهتر عمل می‌کند. (بساک و همکاران، ۲۰۱۹) قیمت سهام را بر اساس الگوریتم تقویت گرادیان شدید پیش‌بینی کردند و نتایج دقیق‌تری پیدا کردند.

در زمینه پژوهش‌های داخلی انجام شده نیز موجودی نقد، یکی از سیال‌ترین دارایی‌های مالی شرکت‌ها می‌باشد. این ویژگی موجودی نقد، اهمیت فوق‌العاده‌ای را به آن بخشیده توانایی اخذ تصمیمات مالی بهینه و به موقع، به مقدار زیادی تحت تأثیر این ویژگی قرار دارد. در صورتی که شرکتی با کمبود وجه نقد مواجه بوده باشد، قادر نخواهد بود از فرصت‌های سرمایه‌گذاری خود استفاده کند یا دیگر نیازهای مالی خود را تأمین نماید؛ بنابراین این کمبود، تأثیر منفی بر ارزش شرکت خواهد گذاشت. هرچند وجوه موجودی نقد در تراز مالی همه شرکت‌ها مهم و ضروری است اما میزان اهمیت وجود این دارایی می‌تواند برای شرکت‌های مختلف تغییر کند (سلیمانی امیری و همکاران، ۱۴۰۰). بقای یک شرکت به توانایی عملکرد سودآورانه و مدیریت زمان دریافت و پرداخت وجوه نقد بستگی دارد، به منظور سنجش عملکرد یک شرکت اصلی‌ترین شاخص در متن صورت‌های مالی شرکت انتشار یافته و در قالب یک رقم سود یا مشتقی از آن همچون سود هر سهم خلاصه می‌گردد و سودهای جاری پیشین جریان‌های نقدی آتی بوده و این سودها نسبت به جریان‌های نقدی عملیاتی کنونی از نظر

ارزشی مرتبط‌تر می‌باشند و مرتبط بودن ارزش جریان‌های نقدی عملیاتی و سودها ممکن است بر اساس شرایط اقتصادی مختلف، متفاوت باشد (صیدخانی و همکاران، ۱۴۰۰).

جریان نقد عملیاتی، وجوه نقدی است که شرکت از طریق انجام فعالیت‌های تجاری خود ایجاد می‌کند جریان نقد عملیاتی را می‌توان برای کنترل کیفیت سود سهام شرکت‌ها مورد استفاده قرار داد. نسبت‌های مالی به منظور کمک به ارزیابی صورت‌های مالی طراحی شده‌اند در واقع نسبت‌های مالی واقعیت‌های مهمی را در ارتباط با عملیات و وضعیت مالی یک شرکت آشکار می‌سازند (نصیرزاده و رستمی، ۱۳۹۱). برای ارزیابی (تمری نیا و همکاران، ۱۴۰۱) به بررسی اعتبار جریان‌های نقد آزاد و کاربرد عملی آن در پیش‌بینی بحران‌های مالی بر پایه استانداردهای بین‌المللی حسابداری می‌پردازند. داده‌های پژوهش با استفاده از نمونه‌ای شامل ۱۵۶۰ مشاهده از ۲۶۰ شرکت طی سال‌های ۱۳۸۷ تا ۱۳۹۶ به دست آمده است. یافته‌های پژوهش نشان می‌دهد که مفروضات جریانات نقد آزاد در بازار سرمایه ایران دارای اعتبار بالایی می‌باشد؛ بنابراین تئوری جنسن در ایران دارای کاربرد عملی نیز می‌باشد علاوه بر این مدل نهایی پژوهش بحران‌های مالی شرکت‌ها در بازار سرمایه ایران را به نحو مناسبی شناسایی و در مقایسه با مدل رایج آلتمن دقت بالاتری دارد. با توجه به نتایج این پژوهش می‌توان گفت که در بازار سرمایه ایران مدل‌های مبتنی بر جریانات نقد آزاد قدرت تبیین بیشتری در ارتباط با پیش‌بینی بحران‌های مالی دارند.

تأثیر محافظه‌کاری بر رابطه بین جریان نقد عملیاتی و ارزش وجه نقد را مورد مطالعه قرار دادند نتایج به دست آمده حاکی از آن است که وجه نقد عملیاتی و ارزش وجه نقد شرکت مرتبط هستند و رابطه مثبت و معناداری دارند (سلیمانی امیری و همکاران، ۱۴۰۰). به بررسی توانایی سود و جریان نقد عملیاتی در توضیح ارزش ذاتی تحقق‌یافته شرکت پرداختند. جامعه آماری تحقیق کلیه شرکت‌های فعال پذیرفته‌شده در بورس بود که تعداد ۱۱۸ شرکت به عنوان نمونه انتخاب شدند. نتایج پژوهش نشان داد توانایی سود و جریان نقد عملیاتی در توضیح ارزش ذاتی تحقق‌یافته شرکت تأثیر معنی‌داری دارد. (انورخطیبی و همکاران، ۱۴۰۰)

تحقیق با عنوان وجه نقد مازاد ارزش شرکت و ریسک نقدشوندگی سهام در شرکت‌های دارای فرصت رشد یا محدودیت مالی انجام دادند. نتایج آزمون پژوهش نشان می‌دهد که وجه نقد مازاد معیار ارزشمندی برای ریسک نقدشوندگی شرکت‌ها به شمار می‌رود وجوه نقد، مازاد تأثیر منفی و معناداری بر ارزش شرکت دارد، همچنین در بررسی شرکت‌های دارای فرصت رشد یا محدودیت، مالی مشخص شد برای شرکت‌هایی که فرصت رشد یا محدودیت مالی دارند رابطه بین وجه نقد مازاد و نقد شوندگی

سهام شرکت شدیدتر است. (اصولیان و همکاران، ۱۴۰۰)

روش‌شناسی پژوهش

اخیراً الگوریتم‌های یادگیری ماشین اغلب به‌عنوان ابزار پیش‌بینی حتی در امور مالی، به‌ویژه برای پیش‌بینی قیمت، مدیریت ریسک مالی، خدمات مالی و تصمیم‌گیری مورد استفاده قرار گرفته‌اند (شائو و کی، ۲۰۲۱). برای پیش‌بینی وام‌دهی بانکی، از الگوریتم‌های مختلف یادگیری ماشینی مانند رگرسیون چندگانه، رگرسیون درختی، جنگل تصادفی و الگوریتم تقویت گرادیان شدید استفاده کردند (اوزگور و همکاران، ۲۰۲۱). علاوه بر این، رویکردهای نظارت در محل و خود نظارتی با استفاده از رویکردهای یادگیری ماشینی مانند الگوریتم جنگل تصادفی مقایسه می‌شوند (آنتونس، ۲۰۲۱). در زمینه ارزهای دیجیتال، رویکردهای مبتنی بر یادگیری ماشین، مانند به عنوان رگرسیون بردار پشتیبان و جنگل تصادفی برای استراتژی‌های معاملاتی استفاده می‌شود (سباستیا و گودینیو، ۲۰۲۱). جنگل تصادفی و حافظه کوتاه‌مدت که یک روش یادگیری عمیق است، برای تجزیه‌وتحلیل اثر کووید-۱۹ بر مقررات بانکی ترکیب شده‌اند (پولیزوس و همکاران، ۲۰۲۱). روش‌های مختلف رگرسیون یادگیری ماشینی مورد استفاده در این مطالعه را در ادامه توضیح دادیم.

رگرسیون خطی چندگانه

این روش نسخه توسعه یافته رگرسیون خطی ساده با رابطه (۱) نشان داده شده است:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \qquad (۱)$$

این فرمول شکل برداری شده برای n مقدار داده است که در آن:

Y: متغیر پاسخ (هدف) به عنوان بردار n مقدار، X_k : k تا متغیر توضیحی (هر عنصر k به عنوان بردار n مقدار)، β_0: ثابت (مقدار برای y)، β_k: ضریب شیب برای k متغیر توضیحی، ε: عبارت خطای مدل.

برای اعمال یک مدل رگرسیون چندگانه باید پنج فرض زیر رعایت شود. پس از بررسی این مفروضات، مدل را بر اساس برخی معیارهای عملکرد، مانند RMSE و MAEاجرا و ارزیابی کردیم.

کی- نزدیک‌ترین همسایه

الگوریتم کی-نزدیک‌ترین همسایه بیشتر برای طبقه‌بندی استفاده می‌شود، اما می‌تواند مشکلات رگرسیون را نیز حل کند. الگوریتم رگرسیون کی-نزدیک‌ترین همسایه با تعریف فواصل بین هر مقدار داده مشاهده شده (با ویژگی‌های داده شده) و مقدار داده جدید با هدف ناشناخته شروع می‌شود.

معیارهای فاصله یا توابع فاصله اقلیدسی یا منهتن هستند (ژانگ، ۲۰۱۶). در فضای n بعدی، فاصله اقلیدسی بین دو نقطه $p(p_1, \ldots, p_n)$ و $q(q_1, \ldots, q_n)$ با استفاده از رابطه (۲) زیر محاسبه می‌شود:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + \cdots + (p_n - q_n)^2} \quad (۲)$$

علاوه بر این، تابع فاصله منهتن در مورد تفاوت مطلق نقاط است:

$$d(p,q) = \sum_{i=1}^{n} |p_i - q_i| \quad (۳)$$

اعتبارسنجی متقابل جستجوی شبکه[۱۶] که تکنیکی برای تعیین فراپارامترهای بهینه در مدل انتخاب شده است، اغلب برای یافتن بهترین مقدار k استفاده می‌شود. گام بعدی یافتن تابع زیان بین مقدار وابسته اختصاص داده شده و مقدار متغیر وابسته واقعی مربوطه (مقادیر وجه نقد برای مشاهدات مختلف) است. عملکرد کلی ضرر در مرحله آموزش به حداقل می‌رسد و نتیجه در تنظیمات مدل منعکس می‌شود.

رگرسیون بردار پشتیبانی

این روش یکی دیگر از الگوریتم‌های کاربردی ساده است که توسط (واپنیک، ۱۹۹۵) طراحی شده است. برخلاف روش رگرسیون چندگانه که سعی می‌کند خطا بین مقدار واقعی هدف و مقدار هدف پیش‌بینی‌شده را به حداقل برساند، رگرسیون بردار پشتیبان بهترین مرز تصمیم‌گیری را در یک مقدار آستانه پیدا می‌کند. این فاصله هر مقدار هدف تا یک مقدار اپسیلون یا حداکثر خطا است:

$$|y_i - wx_i| \leq \varepsilon \quad (۴)$$

در این فرمول y مقدار واقعی وابسته و wx_i مقدار مدل برازش شده است؛ بنابراین، این روش در مقایسه با رگرسیون خطی انعطاف‌پذیر است (انعطاف‌پذیری در تعیین مقدار آستانه). یکی از فراپارامترهای حیاتی در این روش، تنظیم (به عنوان مثال، تکنیک برای به حداقل رساندن اضافه برازش) پارامتر C است. اعتبارسنجی متقابل جستجوی شبکه اغلب برای یافتن بهترین مقدار C استفاده می‌شود.

درخت تصمیم

درخت تصمیم یک روش درختی است که برای طبقه‌بندی و مشکلات رگرسیون استفاده می‌شود. این روش در حین توسعه یک درخت تصمیم مرتبط، یک مجموعه داده را به قطعات کوچک‌تر کاهش می‌دهد. تعیین اصطلاحات آنتروپی و بهره اطلاعاتی برای کاربردهای درخت تصمیم بسیار مهم است.

آنتروپی H یک متریک برای عدم قطعیت توزیع احتمال p است که در رابطه (۵) نشان داده شده است:

$$H(p) = H(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i * log_2 p_i \qquad (۵)$$

که سعی شده به حداقل برسد (ارتل، ۲۰۱۷). در همین حال، بهره اطلاعاتی[۱۷] (IG) متریکی است که کاهش (بهبود) آنتروپی در X را پس از تقسیم مجموعه داده در مورد ویژگی (متغیر) Y نشان می‌دهد. به صورت رابطه (۶) محاسبه می‌شود:

$$IG(X;Y) = H(X) - H(X|Y) \qquad (۶)$$

مجموعه داده با توجه به بالاترین IG تقسیم شده است؛ بنابراین، الگوریتم‌های درخت تصمیم از بالا به پایین کار می‌کنند و متغیری را انتخاب می‌کنند که به طور بهینه مجموعه اشیاء را در هر مرحله از هم جدا کند. به جای یک درخت، برخی از تکنیک‌ها که اغلب روش‌های گروهی نامیده می‌شوند، بیش از یک درخت تصمیم می‌سازند. آن‌ها درختان تقویت‌شده و درخت تصمیم کیسه‌ای نامیده می‌شوند (بریمن، ۱۹۹۶؛ فریدمن، ۱۹۹۹). هدف درختان تقویت‌شده کاهش بایس است، درحالی‌که هدف از بسته‌بندی درختان کاهش واریانس است (روکاچ و میمون، ۲۰۰۵).

جنگل تصادفی

این روش که یک تکنیک مجموعه کیسه‌ای است، پیش‌بینی‌های چند درخت تصمیم (نتیجه) را کنار هم می‌آورد و بر اساس مقادیر میانگین پیش‌بینی‌های این درختان پیش‌بینی می‌کند. اولین مرحله انتخاب زیرمجموعه‌ای از مجموعه داده است و سپس درخت تصمیم جداگانه با زیرمجموعه‌ای از ویژگی‌ها به طور تصادفی انتخاب شده به صورت موازی ساخته می‌شود. برخلاف درخت تصمیم، گره‌های ریشه و جدا شده به طور تصادفی در اینجا انتخاب می‌شوند. همان‌طور که انتظار می‌رود، با افزایش تعداد درختان، دقت بهبود می‌یابد.

الگوریتم افزایش گرادیان شدید

یکی دیگر از روش‌های یادگیری ماشینی تحت نظارت، تقویت گرادیان است که توسط (چن و گسترین، ۲۰۱۶) توسعه یافته است. این یک الگوریتم سریع و کارآمد است و محبوبیت بسیار بالایی در زمینه یادگیری ماشین پیدا می‌کند. برخلاف الگوریتم‌های جنگل تصادفی، در تقویت گرادیان شدید، درخت تصمیم های متنوع به صورت متوالی اجرا می‌شوند، نه به صورت موازی. در این الگوریتم درختان به صورت جداگانه به گروه اضافه می‌شوند و اشتباهات پیش‌بینی مدل‌های گذشته تصحیح

می‌شوند. در اینجا از الگوریتم نزول گرادیان برای به حداقل رساندن گرادیان تلفات استفاده می‌شود.

شبکه‌های عصبی چندلایه

این روش توسط (روملهارت و همکاران،۱۹۸۶) توسعه داده شده است و اساس مطالعات یادگیری عمیق را تشکیل می‌دهد. این شبکه‌ها از یک لایه ورودی، حداقل یک لایه پنهان و یک لایه خروجی تشکیل شده‌اند و هر لایه از مجموعه‌ای از واحدها (نرون‌ها) تشکیل شده است. لایه‌ها کاملاً متصل هستند (متراکم)، به این معنی که تمام واحدهای ورودی از یک لایه به هر واحد فعال‌سازی لایه بعدی متصل هستند. شبکه پیش‌بینی را از طریق انتشار رو به جلو با چندین تابع فعال‌سازی محاسبه می‌کند و با اصلاح وزن‌ها و بایاس‌های شبکه برای تنظیم پارامترهای بهینه برای پیش‌بینی، خطا را از طریق انتشار به عقب به حداقل می‌رساند. لایه ورودی متشکل از متغیرهای توضیحی به نام ویژگی‌ها است و اطلاعات از این لایه به لایه‌های پنهان ارسال می‌شود. در قوس‌های لایه‌های پنهان، پارامترهایی به نام وزن و بایاس وجود دارد. هدف شبکه یافتن تنظیمات پارامتر بهینه است که خطا بین مقدار تخمینی و هدف واقعی را به حداقل برساند.

جامعه آماری

قلمرو زمانی تحقیق از ابتدای سال ۱۳۹۶ تا پایان سال ۱۴۰۰ را شامل می‌شود و قلمرو مکانی بازار سرمایه کشور (مشخصاً بورس اوراق بهادار تهران) است که بر اساس داده‌های مورد نیاز از گزارش‌های فشرده سازمان بورس اوراق بهادار استخراج و دسته‌بندی‌شده است. شرکت‌های مورد مطالعه سابقه عضویت در بورس را برای مدت ۵ سال قبل از آغاز دوره تحقیق داشته باشند. شرکت‌های مورد نظر جزو بانک‌ها، واسطه‌گری‌های مالی، سرمایه‌گذاری‌ها، لیزینگ‌ها و شرکت‌های بیمه صندوق بازنشستگی نباشند. شرکت‌هایی که پایان سال مالی آن‌ها ۲۹ یا ۳۰ اسفند هرسال باشد. شرکت‌هایی که در طول دوره زمانی تحقیق، سال مالی خود را تغییر نداده باشند. اطلاعات مورد نیاز جهت محاسبه متغیرهای تحقیق در سال‌های مورد بررسی در دسترس باشد. شرکت‌ها سهامی عام بورسی و فعال باشند. با مدنظر قرار دادن این محدودیت‌ها تعداد ۲۰۱ شرکت، نمونه این پژوهش را تشکیل می‌دهند.

یافته‌های پژوهش

در این بخش، ما سعی می‌کنیم با استفاده از چندین تکنیک یادگیری ماشینی تحت نظارت، میزان دارایی‌های نقدی شرکت‌ها را پیش‌بینی کنیم. برای پیش‌بینی وجه نقد با استفاده از نرم‌افزار پایتون، تمام روش‌های رگرسیون یادگیری تحت نظارت را که در بخش قبل بر اساس معیار خطای RMSE و MAE مورد بحث قرار گرفته‌اند، ارزیابی کردیم.

RMSE تابعی از تفاوت بین مقادیر مشاهده شده و پیش‌بینی شده است؛ بنابراین، مقادیر RMSE پایین‌تری از مدل‌های رگرسیونی انتظار می‌رود. معیار MAE از لحاظ مشخصات شبیه MSE هست، با این تفاوت که در MAE به جای محاسبه میانگین مربعات خطا (اختلاف بین خروجی تخمین زده شده و خروجی واقعی)، قدرمطلق خطا محاسبه می‌شود. در مقایسه با MSE، معیار MAE ارائه مستقیم‌تری از مجموع خطا دارد. چراکه MSE با خطاها متفاوت برخورد می‌کند. برای مثال اگر خطا زیاد باشد، موقع به توان رسیدن مقدار خطا خیلی زیادتر می‌شود درحالی‌که اگر مقدار خطا کم باشد، با توان دو رسیدن خطا، مقدار خیلی تغییر نمی‌کند. این در حالی هست که در MAE تمامی خطا یک‌جور برخورد می‌شود و قدرمطلق خطا (اختلاف) محاسبه می‌شود، پس مانند RMSE مقادیر پایین‌تری از آن انتظار می‌رود.

ابتدا، ضریب همبستگی بین وجه نقد و سایر متغیرها را حساب کردیم. ضریب همبستگی بین متغیر نسبت سود تقسیمی و وجه نقد صفر است. صفر بودن ضریب همبستگی به این معناست که دو پارامتر مستقل از یکدیگر بوده‌اند و بر اساس اطلاعات موجود از کاهش یا افزایش یکی، نمی‌توان در مورد کاهش یا افزایش دیگری اظهار نظر کرد؛ بنابراین ۲۱ متغیر باقی می‌ماند. همان‌گونه که در جدول ۱ مشاهده می‌کنید، انتخاب متغیر انجام شده است و متغیرهایی که قدرمطلق آن‌ها با هم همبستگی بالاتری داشتند را در یک لیست قرار دادیم و در نهایت ۶ متغیر که بیشترین همبستگی را داشته‌اند، باقی مانده است.

جدول ۱- انتخاب متغیر

اسم مدل	پیش‌بینی کننده‌های مدل
۱۵ ویژگی برتر	{اندازه شرکت، مخارج سرمایه‌ای، بازده سرمایه، بازده دارایی، جریان نقدی، شاخص عدم قطعیت جهانی، EPS، نرخ تورم، نسبت بدهی کوتاه مدت، نسبت دارایی ثابت، تولید ناخالص داخلی ایران، بازده حقوق صاحبان سهام، شاخص عدم قطعیت همه‌گیری جهانی، رشد فروش، سرمایه در گردش خالص}
۸ ویژگی برتر	{اندازه شرکت، مخارج سرمایه‌ای، بازده سرمایه، بازده دارایی، جریان نقدی، شاخص عدم قطعیت جهانی، EPS، نرخ تورم}
۶ ویژگی برتر	{اندازه شرکت، مخارج سرمایه‌ای، بازده سرمایه، بازده دارایی، جریان نقدی، نرخ تورم}

ابتدا الگوریتم رگرسیون خطی چندگانه برای پیش‌بینی وجه نقد استفاده می‌شود. برای اعمال رگرسیون چندگانه، فرضیات را در بالا بررسی کردیم. معیارهای عملکرد پس از اعمال الگوریتم رگرسیون خطی چندگانه در جدول ۲ نشان داده شده است.

جدول ۲- معیارهای عملکرد الگوریتم رگرسیون خطی چندگانه

رگرسیون خطی چندگانه	۲۱ متغیر	۱۵ متغیر	۸ متغیر	۶ متغیر
RMSE	۰,۷۴۹۷	۰,۷۶۳۰	۰,۷۴۸۵	۰,۷۴۷۰
MAE	۰,۳۴۶۹	۰,۳۵۶۲	۰,۳۲۸۶	۰,۳۲۹۲

همان‌گونه که در جدول ۲ مشاهده می‌کنید مقدار RMSE و MAE بالا است. معیارهای عملکرد با آن ۲۱، ۱۵، ۸ و ۶ متغیر درمدل درجدول ۲ نشان داده‌شده‌است که فرق چندانی دردسته‌بندی‌های متفاوت ندارد. نتایج هنوز ناموفق هستند؛ بنابراین، می‌توان نتیجه گرفت که رگرسیون خطی چندگانه در پیش‌بینی مقادیر وجه نقد خوب نیست.

الگوریتم دیگر، کی-نزدیک‌ترین همسایه است که برای پیش‌بینی مقدار وجه نقد با چندین متغیر پیش‌بینی کننده استفاده می‌شود. برای یافتن بهترین مقدار k که خطای مدل را به حداقل می‌رساند، اعتبارسنجی متقابل جستجوی شبکه را اعمال کردیم و ۹ را به عنوان k بهینه انتخاب کردیم. مقدار k بهینه با مدل ۱۵ ویژگی انتخاب‌شده برابر با ۵۰ است.

همان‌طور که در جدول ۳ نشان داده شده است، نتایج RMSE و MAE برای هر دو مدل همچنان بالا است ولی به مرور با تعداد ویژگی‌های کمتر، بهتر می‌شود؛ بنابراین، کی-نزدیک‌ترین همسایه در پیش‌بینی مقادیر وجه نقد ناموفق است، اگرچه نتایج بهتری در مقایسه با مدل رگرسیون خطی چندگانه ارائه می‌دهد.

جدول ۳- معیارهای عملکرد الگوریتم کی-نزدیک‌ترین همسایه

کی-نزدیک‌ترین همسایه	۲۱ متغیر	۱۵ متغیر	۸ متغیر	۶ متغیر
k	۹	۵۰	۵۰	۵۰
RMSE	۰,۵۳۳۱	۰,۴۴۲۱	۰,۳۹۳۴	۰,۴۰۰۹
MAE	۰,۱۸۱۴	۰,۱۳۹۷	۰,۱۱۲۴	۰,۱۱۵۷

الگوریتم رگرسیون بردار پشتیبان سومین الگوریتم یادگیری ماشینی تحت نظارت برای پیش‌بینی وجه نقد است. با اعتبارسنجی متقابل جستجوی شبکه، مقدار فراپارامتر C برابر ۱۰ است.

جدول ۴ نشان می‌دهد که مقدار RMSE خیلی بالا نیست. رگرسیون بردار پشتیبان معیارهای عملکرد بسیار بهتری را در مقایسه با الگوریتم‌های رگرسیون خطی چندگانه و کی-نزدیک‌ترین همسایه برای پیش‌بینی وجه نقد ارائه می‌کند.

جدول ۴- معیارهای عملکرد الگوریتم رگرسیون بردار پشتیبان

رگرسیون بردار پشتیبان	۲۱ متغیر	۱۵ متغیر	۸ متغیر	۶ متغیر
RMSE	۰,۳۸۰۶	۰,۳۷۷۷	۰,۳۹۷۴	۰,۳۹۴۸
MAE	۰,۱۴۰۲	۰,۱۳۸۰	۰,۱۴۸۵	۰,۱۴۲۳

پس از آن، الگوریتم درخت تصمیم برای پیش‌بینی وجه نقد اعمال می‌شود. برای این الگوریتم، پارامتر حداکثر بهینه عمق درخت ۲ است. تعداد ویژگی‌های استفاده شده در این الگوریتم بر اساس نمرات همبستگی نزولی کاهش می‌یابد و آن مدل‌های جدید نیز اجرا می‌شوند. بر اساس جدول ۵ مقادیر RMSE بزرگ‌تر از خروجی‌های الگوریتم رگرسیون بردار پشتیبان هستند؛ بنابراین، الگوریتم درخت تصمیم نیز در پیش‌بینی متغیر وجه نقد خوب نیست.

جدول ۵- معیارهای عملکرد الگوریتم درخت تصمیم

درخت تصمیم	۲۱ متغیر	۱۵ متغیر	۸ متغیر	۶ متغیر
RMSE	۰,۴۵۷۹	۰,۴۵۷۹	۰,۸۱۲۶	۰,۸۱۲۶
MAE	۰,۱۷۵۲	۰,۱۷۵۲	۰,۲۶۱۳	۰,۲۶۱۳

برای الگوریتم جنگل تصادفی نیز همان‌طور که در جدول ۶ نشان داده شده است، مقادیر RMSE در مقایسه با الگوریتم‌های قبلی بیشتر است. علاوه بر این، با کاهش تعداد ویژگی‌ها، این دو معیار بهبود نمی‌یابند و بدتر نیز می‌شوند.

جدول ۶- معیارهای عملکرد الگوریتم جنگل تصادفی

جنگل تصادفی	۲۱ متغیر	۱۵ متغیر	۸ متغیر	۶ متغیر
RMSE	۰,۵۷۰۴	۰,۵۶۸۱	۰,۶۰۵۴	۰,۶۲۹۲
MAE	۰,۱۷۵۸	۰,۱۷۵۷	۰,۱۷۹۸	۰,۱۹۰۴

سپس الگوریتم تقویت گرادیان شدید برای پیش‌بینی وجه نقد اعمال می‌شود. جدول ۷ نشان می‌دهد که الگوریتم تقویت گرادیان شدید نیز RMSE و MAE بالا را به دست آورده است. مدل ۷۳ درصد از تنوع مشاهده شده در مقادیر وجه نقد را به تصویر می‌کشد. هنگامی‌که تعداد ویژگی‌های مورد استفاده در مدل کاهش می‌یابد، مقادیر نتیجه مدل به طور قابل‌توجهی بدتر می‌شود؛ ولی برای ۱۵ متغیره نتایج بهتر می‌شود؛ بنابراین، مدل با تمام ویژگی‌های گنجانده شده به عنوان بهترین مدل برای پیش‌بینی متغیر پاسخ وجه نقد انتخاب نمی‌شود.

جدول ۷- معیارهای عملکرد الگوریتم تقویت گرادیان شدید

تقویت گرادیان شدید	۲۱ متغیر	۱۵ متغیر	۸ متغیر	۶ متغیر
RMSE	۰,۴۴۰۹	۰,۴۳۹۱	۰,۴۹۷۱	۰,۵۳۷۸
MAE	۰,۲۰۶۵	۰,۱۸۱۱	۰,۲۱۴۷	۰,۲۱۲۸

در نهایت، الگوریتم یادگیری عمیق شبکه عصبی چندلایه (شبکه‌های عصبی چندلایه) برای پیش‌بینی وجه نقد استفاده می‌شود. بهترین تنظیمات فراپارامتر این الگوریتم شامل دو تا پنج لایه پنهان متراکم است. برای مدل با ۱۵ متغیر با ۵ لایه پنهان نیز امتحان شد که همان‌گونه که می‌بینید نتایج درخشانی نداشت ولی ۱۵ متغیره با ۲ لایه پنهان نتایج بهتری داشت. خروجی مدل با RMSE و MAE بالا نشان می‌دهد که این مدل در پیش‌بینی مقادیر وجه نقد موفق عمل کرده است.

جدول ۸- معیارهای عملکرد الگوریتم شبکه‌های عصبی چندلایه

شبکه‌های عصبی چندلایه	۲۱ متغیر	۱۵ متغیر	۱۵ متغیر	۸ متغیر	۶ متغیر
Hidden layers	۲	۲	۵	۲	۲
RMSE	۰,۳۹۸۷	۰,۳۹۹۸	۰,۵۳۴۸	۰,۵۷۹۱	۰,۶۳۴۲
MAE	۰,۱۱۵۳	۰,۱۲۴۸	۰,۱۲۹۰	۰,۱۸۸۹	۰,۱۹۴۷

آزمون مقایسه زوجی

آزمون t زوجی[18] برای تعیین وجود تفاوت آماری معنی‌داری در میانگین متغیر وابسته بین دو گروه مرتبط استفاده می‌شود. این آزمون همچنین به عنوان آزمون t نمونه زوجی و یا آزمون t وابسته نیز شناخته می‌شود.

مانند بسیاری از روش‌های آماری، آزمون t زوجی دارای دو فرضیه است، فرضیه صفر و فرضیه یک. فرضیه صفر فرض می‌کند که تفاوت میانگین واقعی بین نمونه‌های زوج صفر است. تحت این مدل، همه تفاوت‌های قابل مشاهده با تغییرات تصادفی توضیح داده می‌شوند. برعکس، فرضیه یک فرض می‌کند که تفاوت میانگین واقعی بین نمونه‌های زوجی برابر با صفر نیست. فرضیه یک بسته به نتیجه مورد انتظار می‌تواند یکی از چندین شکل را داشته باشد. اگر جهت تفاوت مهم نباشد، از فرضیه دو دنباله استفاده می‌شود. در غیر این صورت می‌توان از فرضیه دم بالا یا پایین برای افزایش قدرت آزمون استفاده کرد. فرضیه صفر برای هر نوع فرضیه یک یکسان باقی می‌ماند. مهم است که به یاد داشته باشید که فرضیه‌ها هرگز در مورد داده‌ها نیستند، بلکه در مورد فرآیندهایی هستند که داده‌ها را تولید می‌کنند. هدف از آزمون فرضیه تعیین فرضیه یک است که داده‌ها با آن سازگارتر هستند. آزمون مقایسه زوجی حتی زمانی که بخواهیم شباهت میانگین دو جامعه را مورد بررسی قرار دهیم نیز به کار

گرفته می‌شود. البته به شرطی که هردو جامعه از بقیه جهات کاملاً شبیه به یکدیگر باشند.

آزمون t نمونه زوجی دارای چهار فرض اصلی است. متغیر وابسته باید پیوسته (فاصله/نسبت) باشد. مشاهدات مستقل از یکدیگر هستند. متغیر وابسته باید تقریباً به طور نرمال توزیع شود. متغیر وابسته نباید حاوی هیچ نقطه پرت باشد.

دراین پژوهش، ما از این آزمون برای مقایسه الگوریتم رگرسیون بردارپشتیبان با سایر الگوریتم‌های یادگیری ماشین نظارت شده استفاده کردیم که داده‌های مورد استفاده، مقدار RMSE در ۲۱، ۱۵، ۸ و ۶ متغیره برای هر دو الگوریتم است.

اگر مقدار احتمال از میزان خطایی که در نظر می‌گیریم بیشتر باشد، به نظر می‌رسد که نمونه تصادفی از فرض صفر پشتیبانی می‌کند در نتیجه دلیلی بر رد فرض صفر وجود ندارد. برعکس اگر مقدار احتمال از خطای مورد نظر کوچک‌تر باشد، فرض صفر توسط نمونه تصادفی پشتیبانی نمی‌شود در نتیجه به کمک این نمونه تصادفی، فرض صفر را رد می‌کنیم. در فاصله اطمینان ۹۵ درصد در مقایسه الگوریتم رگرسیون بردار پشتیبان با الگوریتم درخت تصمیم عملکرد رگرسیون بردار پشتیبان خوب نیست ولی در مقایسه با سایر الگوریتم‌ها، رگرسیون بردار پشتیبان از الگوریتم مورد مقایسه بهتر عمل کرده است.

نتیجه گیری و بحث

به طور خلاصه، ابتدا روش‌های یادگیری ماشین ساده‌تر برای مجموعه داده اعمال می‌شوند که با رگرسیون خطی چندگانه شروع می‌شود. مفروضات بررسی می‌شوند و نتایج رگرسیون خطی چندگانه معیارهای عملکرد ضعیفی را نشان می‌دهد (مقادیر RMSE و MAE بالا). بهترین نتایج با استفاده از الگوریتم رگرسیون بردار پشتیبان (مقادیر RMSE ۰,۳۸ و MAE ۰,۱۴) به دست می‌آید. در مقایسه با رگرسیون خطی چندگانه که بدترین الگوریتم نتیجه دهنده است، رگرسیون بردار پشتیبان مقدار RMSE کمتر ارائه می‌کند.

نتایج با ۱۵ متغیر به طرز قابل توجهی بهبود بخشید. رگرسیون خطی چندگانه همچنان الگوریتم خوبی برای پیش‌بینی نیست اما کی-نزدیک‌ترین همسایه به طرز قابل توجهی بهبود یافت. الگوریتم‌های درخت تصمیم، رگرسیون بردار پشتیبان و جنگل تصادفی تقریباً مانند قبل هستند اما الگوریتم‌های تقویت گرادیان شدید و شبکه‌های عصبی چندلایه در مقادیر MAE بهبود یافته‌اند. همچنین بر مبنای آزمون مقایسه زوجی رگرسیون بردار پشتیبان عملکرد بهتری از سایر الگوریتم‌های یادگیری ماشین

نظارت شده به جز درخت تصمیم دارد.

برخی از الگوریتم‌های یادگیری ماشین، غالب‌ترین (مهم‌ترین) ویژگی‌ها را با استفاده از نمودارهای میله‌ای ارائه می‌کنند. به طور کلی در همه‌ی نمودارها، ویژگی‌های مشترک مهم برای هر یک از این چهار الگوریتم عبارت‌اند از: سایز شرکت (SIZE)، مخارج سرمایه‌ای (CapEx) و با تفاوت قابل توجهی TANG(نسبت دارایی ثابت) و WUI_IRN (شاخص عدم قطعیت جهانی برای کشور ایران)

بر اساس یافته‌های ما، این مطالعه پیامدهای قابل‌توجهی برای مدیران و محققان شرکت‌ها دارد. مدیران می‌توانند از این اطلاعات برای تعیین میزان نگهداری وجه نقد شرکت‌ها برای ایجاد سیاست‌های شرکتی استفاده کنند. در همین حال، محققان می‌توانند از اطلاعات برای ایجاد مدل‌های رگرسیون بهتر و یافتن رفتار نگهداری وجه نقد شرکت‌ها استفاده کنند. با توجه به نتایج به‌دست‌آمده پیشنهاد می‌گردد که از الگوریتم رگرسیون بردار پشتیبان برای پیش‌بینی نگهداری وجه نقد استفاده کنند. تأثیرگذارترین متغیر مستقل در پیش‌بینی نگهداری وجه نقد اندازه شرکت‌ها است که پیشنهاد می‌گردد تحلیلگران به آن توجه کنند. همچنین پیشنهاد می‌گردد به شاخص عدم قطعیت جهانی که یکی از متغیرهای مستقل است و با توجه به نتایج به‌دست‌آمده می‌تواند برای پیش‌بینی نگهداری وجه نقد مؤثر باشد، توجه شود.

ما عمدتاً بر شرکت‌های ایرانی و ویژگی‌های آن‌ها تمرکز می‌کنیم و دوره مورد مطالعه بین سال‌های ۱۳۹۶ تا ۱۴۰۰ است. در مطالعات بعدی می‌توان دوره را گسترش داد. البته باید در نظر داشت به دلیل افزایش تعدادی از شرکت‌ها از ۱۳۹۶ به بعد تعداد شرکت‌ها کاهش می‌یابد ولی با در نظر گرفتن سال ۱۴۰۰ به بعد می‌توان دوره پساکووید را نیز در نظر گرفت. متغیرهای کلان اقتصادی بیشتری مانند قیمت نفت را به مطالعات اضافه کرد. علاوه بر بازه زمانی، تعداد کشورها را می‌توان افزایش داد. مطالعات آینده می‌تواند یک تحلیل بین کشوری را در نظر بگیرد. برای مثال، محققان می‌توانند میزان نگهداری وجه نقد را برای بازارهای توسعه‌یافته و نوظهور پیش‌بینی کنند تا مشخص کنند که آیا تفاوت‌هایی در سطوح نگهداری وجه نقد بین بازارها وجود دارد یا خیر. آن‌ها همچنین می‌توانند شرکت‌ها را در قاره‌های مختلف مقایسه کنند تا تفاوت‌های منطقه‌ای را در تأثیرات روی سطوح نگهداری وجه نقد بیابند. همچنین می‌توان از روش‌های دیگر یادگیری ماشین نیز استفاده کرد تا پژوهش کامل‌تری ارائه شود.

منابع

۱) اصولیان، محمد؛ تجویدی، الناز و پازوکی، یاسمن (۱۴۰۰). وجه نقد مازاد، ارزش شرکت و ریسک نقدشوندگی سهام در شرکت‌های دارای فرصت رشد یا محدودیت مالی. بررسی‌های حسابداری و حسابرسی،۲۸(۲): ۲۴۸-۲۷۳.

۲) انورخطیبی، سعید؛ سیفی گواهر، لیلا و انصاری، فریبا (۱۴۰۰). توانایی سود و جریان نقد عملیاتی در توضیح ارزش ذاتی تحقق یافته شرکت. چشم‌انداز حسابداری و مدیریت، ۴(۴۵): ۹۷-۱۱۲.

۳) تمری نیا، آیت اله؛ نظری، رضا و مرادزاده فرد، مهدی (۱۴۰۱). بررسی اعتبار جریان‌های نقد آزاد و کاربرد عملی آن در پیش‌بینی بحران‌های مالی بر پایه استانداردهای بین‌المللی حسابداری (IFRS): شواهدی از بازار سرمایه ایران. دانش حسابداری و حسابرسی مدیریت، ۱۱(۴۱): ۲۵۹-۲۷۰.

۴) سلیمانی امیری، غلامرضا؛ جمشیدی، طیبه و عنبری، حمزه (۱۴۰۰). بررسی تأثیر محافظه‌کاری بر رابطه بین جریان نقد عملیاتی و ارزش وجه نقد. فصلنامه علمی کارافن، ۲(۱۸): ۷۳-۸۸.

۵) صیدخانی، رضا؛ محمدی ملقرنی، عطا و امینی، پیمان (۱۴۰۰). بررسی توانمندی جریان‌های نقدی عملیاتی در ارزیابی عملکرد شرکت‌ها با تأکید بر کیفیت افشا در دوره‌های بحران مالی. پژوهش‌های حسابداری مالی و حسابرسی، ۴۹(۱۳): ۱۴۷-۱۷۶.

۶) نصیرزاده، فرزانه و رستمی، امین (۱۳۹۱). بررسی رابطه‌ی بین شاخص‌های نقدینگی نوین و مبتنی بر صورت جریان وجه نقد با سودآوری شرکت‌ها (معیارهای مالی و مبتنی بر بازار). اقتصاد پولی، مالی، ۳(۱۹): ۲۸-۵۲.

7) Antunes, J. A. P. (2021). “To supervise or to self-supervise: A machine learning based comparison on credit supervision.” Financial Innovation, 7(1): 1-21.

8) Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). “Predicting the direction of stock market prices using tree-based classifiers.” The North American Journal of Economics and Finance, 47: 552-567.

9) Bates, T. W., Kahle, K. M., & Stulz, R. M. (2009). “Why do US firms hold so much more cash than they used to?.” The journal of finance, 64(5): 1985-2021.

10) Breiman, L. (1996). “Bagging predictors.” Machine learning, 24: 123-140.

11) Chen, T., & Guestrin, C. (2016). “Xgboost: A scalable tree boosting system.” In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794.

12) Ertel, W. (2018). “Introduction to artificial intelligence.” Springer.

13) Foley, C. F., Hartzell, J. C., Titman, S., & Twite, G. (2007). “Why do firms hold so much cash? A tax-based explanation.” Journal of financial economics,

86(3): 579-607.
14) Friedman, J.H. (2002). "Stochastic gradient boosting." Computational statistics & data analysis, 38(4): 367-378.
15) Gholamzadeh, M., Faghani, M., & Pifeh, A. (2021). "Implementing machine learning methods in the prediction of the financial constraints of the companies listed on Tehran's stock exchange." International Journal of Finance & Managerial Accounting, 6(20): 131-144.
16) Jensen, M. C. (1986). "Agency costs of free cash flow, corporate finance, and takeovers." The American economic review, 76(2): 323-329.
17) Keynes, J. M. (1936). The general theory of unemployment. Interest and Money. Harcourt Brace, London.
18) Kim, C. S., Mauer, D. C., & Sherman, A. E. (1998). "The determinants of corporate liquidity: Theory and evidence." Journal of financial and quantitative analysis, 33(3): 335-359.
19) Kou, G., Peng, Y., & Wang, G. (2014). "Evaluation of clustering algorithms for financial risk analysis using MCDM methods." Information sciences, 275: 1-12.
20) Jensen, M. C., & Meckling, W. H. (2019). "Theory of the firm: Managerial behavior, agency costs and ownership structure." In Corporate governance, 77-132.
21) Miller, M. H. and D. Orr (1966). "A model of the demand for money by firms." The Quarterly journal of economics, 80(3): 413-435.
22) Mousa, G. A., Elamir, E. A., & Hussainey, K. (2022). "Using machine learning methods to predict financial performance: Does disclosure tone matter?." International Journal of Disclosure and Governance, 1-20.
23) 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.
24) Polyzos, S., Samitas, A., & Kampouris, I. (2021). "Economic stimulus through bank regulation: Government responses to the COVID-19 crisis." Journal of International Financial Markets, Institutions and Money, 75: 101444.
25) Popescu, M. E. and V. Dragotă (2018). "What do post-communist countries have in common when predicting financial distress?" Prague Economic Papers, 27(6): 637-653.
26) 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), 1-6
27) Rokach, L., & Maimon, O. (2005). "Top-down induction of decision trees classifiers-a survey." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 35(4): 476-487.
28) Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). "Learning representations by back-propagating errors." nature, 323(6088): 533-536.
29) Sebastião, H., & Godinho, P. (2021). "Forecasting and trading cryptocurrencies

with machine learning under changing market conditions." Financial Innovation, 7(1): 1-30.
30) Vapnik, V. (1999). "The nature of statistical learning theory." Springer science & business media.
31) Won, C., Kim, J., & Bae, J. K. (2012). "Using genetic algorithm based knowledge refinement model for dividend policy forecasting." Expert Systems with Applications, 39(18): 13472-13479.
32) 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.
33) Xiao, F., & Ke, J. (2021). "Pricing, management and decision-making of financial markets with artificial intelligence: introduction to the issue." Financial Innovation, 7: 1-3.
34) Zhang, Z. (2016). "Introduction to machine learning: k-nearest neighbors." Annals of translational medicine, 4(11).

یادداشت‌ها:

1 Multiple Linear regression
2 Support vector regressor (SVR)
3 K-Nearest Neighbors
4 Extreme gradient boosting algorithm (XGBoost)
5 Multi-layer neural networks (MLNN)
6 Root Mean Squared Error (RMSE)
7 Mean Absolute Error (MAE)
8 World Uncertainty Index (WUI)
9 World Pandemic Uncertainty Index (WPUI)
10 Logistic Model Tree (LMT)
11 Random Forest (RF)
12 Simple CHART
13 Extra Tree
14 Decision Tree
15 k-means
16 Grid search cross-validation
17 Information Gain
18 Paired Sample t-test (PSPT)

Predicting cash holdings using supervised machine learning algorithms in companies listed on the Tehran Stock Exchange (TSE)

Saeid Fallahpour[1]
Reza Raei[2]
Negar Tavakoli[3]

Receipt: 12/07/2023 Acceptance: 11/01/2024

Abstract

According to the 22 selected features (which are checked during the research) with machine learning methods, this study predicts the cash holding of companies admitted to the Tehran Stock Exchange. 201 companies were investigated from 1396 to 1400. Multiple linear regression, K-nearest neighbor, support vector regression, decision tree, random forest, extreme gradient boosting algorithm and multilayer neural networks are used for prediction. The results show that the multiple linear regression methods provide the k-nearest neighbor of the root mean square error (RMSE) and the mean absolute error (MAE) of the high error. Meanwhile, more complex algorithms, especially support vector regression, achieve higher accuracy; The findings indicated that by reducing to 15 variables, machine learning methods, especially K-nearest neighbor, provided better results. Based on the paired sample t-test, support vector regression has a better performance than other supervised machine learning algorithms except decision tree. Also, the most important variables were company size and capital expenditures (CapEx). The World Uncertainty Index and inflation were also relatively important variables; Therefore, by using the support vector regression algorithm, we may predict the amount of cash to a significant extent.

Keywords

SVR, XGBoost, Cash holdings, Machine learning, World Uncertainty Index.

1-Associate Professor, Department of Finance, Faculty of Management, University of Tehran, Tehran, Iran. falahpor@ut.ac.ir
2- Professor, Department of Finance, Faculty of Management, University of Tehran, Tehran, Iran. raei@ut.ac.ir
3-MSc. Student, Department of Finance and Insurance, Faculty of Management, University of Tehran, Tehran, Iran. (Corresponding Author) negartavakoli0721@gmail.com