تحقیقات حسابداری و حسابرسی، زمستان ۱۴۰۱، شماره ۵۶، از صفحه ۱۷۱ تا ۱۹۰

کاربرد یادگیری ماشین در ارائه الگویی برای پیش بینی ورشکستگی

افشین هاشمی گل سفیدی [۱]

گروه حسابداری ، واحد بین الملل کیش ، دانشگاه آزاد اسلامی ، جزیره کیش ، ایران

دکتر زهرا لشگری [۲]©

استادیار گروه حسابداری، واحد تهران مرکز دانشگاه آزاد اسلامی ، تهران ، ایران

دکتر زهره حاجیها [۳]

دانشیار گروه حسابداری ، واحد تهران شرق ، دانشگاه آزاد اسلامی ، تهران ، ایران

(تاریخ دریافت: ۲۶ آذر ۱۳۹۹؛ تاریخ پذیرش: ۱۱ اردیبهشت ۱۴۰۰)

استفاده از ابزارها و روش های پیش بینی سنتی خطای بالایی داشته و در مقایسه با روش های جدیدتر و مدل های غیرخطی عملکرد ضعیف تری دارند. یکی از روش ها و الگوریتم های پرکاربرد در پیش بینی استفاده از یادگیری ماشین است. هدف اصلی این پژوهش بررسی کاربرد یادگیری ماشین در ارائه الگویی برای پیش بینی ورشکستگی ۳۰۸ شرکت پذیرفته شده در بورس اوراق بهادار تهران در بازه زمانی ۱۳۸۹ تا ۱۳۹۸(۳۰۸۰ سال – شرکت) می‌باشد که برای آزمون فرضیه‌ها از رگرسیون چند گانه داده‌های ترکیبی و به منظور پیاده‌سازی الگوریتم خوشه‌بندی K-Medians و محاسبات مربوطه از نرم افزار محاسبات اماری R استفاده گردید. نتایج حاصل از پژوهش نشان می‌دهد از میان نسبت های مالی مشخص شده در مدل اول فقط نسبت درآمد خالص به کل دارایی و نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار موجب می‌توانند توانایی مدل پیش بینی ورشکستگی را بهبود ببخشند. همچنین در مدل دوم نسبت های مالی مشخص شده توانایی بهبود مدل پیش بینی ورشکستگی را داشته و با اضافه کردن متغیر Devscore برای گروه‌هایی که بر اساس صنعت و اندازه ساخته شده اند، مدل اصلاح شده و پیش بینی ورشکستگی را بهبود می بخشد. نتایج حاکی از آن است که یک شرکت در صورت داشتن نسبت های مالی مرتبط با ورشکستگی که پایین تر از میانگین همسانان خوشه‌ای خود است، احتمالاً ورشکسته می‌شود.

واژه‌های کلیدی: یادگیری ماشین، الگوریتم خوشه‌بندی، پیش بینی ورشکستگی.

[1] afshinhashemi54@yahoo.com

[2] z_lashgari@iauctb.ac.ir © (نویسندهٔ مسئول)

[3] drzhajiha@gmail.com

مقاله علمی - پژوهشی

مقدمه

پیش بینی ورشکستگی به نظر می‌رسد پر طرفدارترین موضوعات در استفاده از تکنیک‌های تصمیم‌گیری در داده‌های مالی است. ورشکستگی شرکت‌ها منجر به خسارات اقتصادی برای مدیران، سرمایه‌گذاران، بستانکاران و کارکنان به همراه هزینه‌های اجتماعی می‌شود. به‌این دلایل پیش‌بینی ورشکستگی یک مسئله مهم در امور مالی است. رقابت روز افزون بنگاه‌های اقتصادی، دست‌یابی به منابع را محدود کرده است و احتمال ورشکستگی را افزایش داده است. تحقیقات انجام شده قبلی نشان داده است که شرکت‌ها ورشکستگی خود را پنهان می‌کنند و زمانی رسماً اعلام ورشکستگی می‌کنند که درآن زمان تلاش برای جلوگیری از ورشکستگی بیهوده و بسیار دیر است[۳۰]. شاغلین در حرفه و پژوهشگران اغلب یک شرکت را با محکی (معیاری) از پیش تعیین شده برای ارزیابی عملکرد آن مقایسه می‌کنند. با وجود جذابیت مفهومی ساده، شناسایی یک الگوی بهینه مناسب کار ساده‌ای نیست. اغلب از شرکتهای همسان برای ساخت گروه‌های کنترل به عنوان الگوی بهینه یا محک استفاده می‌شود[۲۴،۱۳،۳۲].

روشی که برای انتخاب شرکت های هم رده در یک شرایط خاص انتخاب می‌شود ممکن است برای شرایط دیگر مناسب نباشد. از اینرو انتخاب و دسته بندی شرکت های هم رده تنها به یک روش برای همه شرایط مناسب نیست. این تحقیق به دنبال ارائه یک الگوی پویای انتخاب شرکت های هم رده است که از این طریق گروه شرکت های هم همسان برای اهداف تحقیق خاص مشخص می‌شود. این الگو می‌تواند به کشف ناهنجاریهای احتمالی که ممکن است از وقوع رویدادهای مهم و قابل توجه در اینده حکایت داشته باشد، کمک کند(برای مثال پیش بینی ورشکستگی) یک روش معمول برای کشف یا پیش بینی رویدادهای مهم شرکت ها جستجوی شرکت هایی است که انحرافی غیرعادی از شرکت های همسان خود را نشان می دهد. به عبارتی پرسش اصلی تحقیق نیز به شرح زیر است: آیا خوشه بندی Medians–K می‌تواند مدل پیش بینی ورشکستگی را بهبود بخشد یا خیر. نتایج حاصل از این پژوهش، می‌تواند در پیش بینی ورشکستگی شرکت ها با توجه به جریان تحقیقاتی ایجاد شده آن قطعاً مورد علاقه و توجه طیف وسیعی از اقشار از جمله مراجع گوناگون تدوین استانداردهای حسابداری، سایر نهادهای قانون گذار مثل بورس های اوراق بهادار و نهادهای نظارتی آنها، مدیران شرکت ها و سایر استفاده کنندگان از صورتهای مالی و اطلاعات حسابداری و بطور کلی ذینفعان واحدهای تجاری و شهروندان جامعه مدنی می‌باشد. در این پژوهش پیش بینی ورشکستگی با استفاده از یادگیری ماشین مورد توجه است که در بخش های بعدی مقاله به صورت مفصل تبیین می گردد.

مبانی نظری و تدوین فرضیه‌ها

این روزها شاهد وقوع بحران های تجاری در سراسر جهان هستیم. شدت گرفتن رقابت در عرصه صنایع باعث شده است، بسیاری از شرکت ها ورشکست شده و از گردونه رقابت خارج شوند. این امر موجبات نگرانی صاحبان سرمایه و به طور کلی جامعه را فراهم آورده است. همچنین بحران مالی و در نهایت ورشکستگی می‌تواند زیان های هنگفتی را برای سهامداران، سرمایه گذاران، اعتباردهندگان، مدیران،

کارکنان، عرضه کنندگان مواد اولیه و مشتریان در پی داشته باشد. یکی از عوامل اصلی بحران مالی و به دنبال آن، ورشکستگی شرکت ها، اعمال نکردن کنترل از طرف ذینفعان مختلف شامل سهامداران، اعتباردهندگان، مدیران، کارگران و عرضه کنندگان است[۴].

سرمایه گذاران همواره می خواهند با پیش بینی امکان ورشکستگی یک شرکت، از سوخت شدن اصل و سود تضمین شده سرمایه خود، جلوگیری کنند. از این رو؛ آنها، در پی روشهایی هستند که بتوانند به وسیله آن مانع از ورشکستگی مالی شرکت ها شوند؛ زیرا در صورت ورشکستگی، قیمت سهام شرکت ها به شدت کاهش می یابد. یک روش معمول برای کشف یا پیش بینی رویدادهای مهم جستجوی شرکت هایی است که انحرافی غیرعادی از شرکت های همسان خود را نشان می دهد. در حال حاضر شناسایی شرکت های هم رده به طور کلی به سیستم های سنتی طبقه بندی صنعت مانند طرح طبقه بندی استاندارد صنعت(SIC) و در آمریکا سیستم طبقه بندی صنعت آمریکا(NAICS) که در آن شرکت ها در گروه‌های از پیش تعریف شده قرار داده می‌شوند. اما به هر حال این سیستم های طبقه بندی اساسا بر خروجی عملیات صنعت مثل خدمات یا محصولات فروخته شده مبتنی هستند و در شناسایی صنایع جدید کند هستند[۱۱،۱۰،۱۵،۲۳]. سیستم های طبقه بندی صنعت پس از دهه ۱۹۷۰ در شناخت صنایع جدید و نوظهور کند بوده و همچنین، کدهای صنعت فقط گهگاه تغییر می‌کنند، اگرچه بعید است که همسانان یک شرکت برای مدت طولانی بدون تغییر باقی بمانند. محیط شغلی که به سرعت در حال تحول است، به دنبال روش های طبقه بندی انعطاف پذیر تری است که بتواند موجب پیشرفت تجارت و کسب و کار شود[۲۰].

سیستم های طبقه بندی صنعت برخی موانع اجرایی را دارند که مانع کاربرد آنها را در عمل می‌شود. مطالعات اولیه اختلاف بزرگ کدهای طبقه بندی صنعت را در بانک های اطلاعاتی مختلف ثبت کرده است. هولمز و استیونز[1](۲۰۰۴) استدلال می‌کنند که بنگاه‌های اقتصادی که طبق طبقه بندی های استاندارد به عنوان همسانان صنعتی شناخته می‌شوند، ممکن است واقعاً یکدست نباشند. از طرف دیگر، تحقیقات آکادمیک برای ایجاد روش های انتخاب همسانان مبتکرانه و هدفمند، نسبتاً کند بوده است. فان و لانگ[2](۲۰۰۰) از داده‌های جریان ورودی و خروجی کالاها استفاده می‌کنند که منعکس کننده رابطه دو بنگاه برای تصمیم گیری برای همسانان صنعتی است. رامنات(۲۰۰۲) شرکت ها را به عنوان همسانان در نظر می گیرد اگر حداقل پنج تحلیلگر همزمان آنها را تحت پوشش قرار دهند. در این تحقیق از الگوریتم یادگیری ماشین به نام خوشه بندی K-Medians روی نسبت های مالی برای شناسایی ناهنجاریهای حسابداری مرتبط با رویدادهای مهم شرکت استفاده می‌شود. در همین رابطه‌هابرگ و فیلیپس

(۲۰۱۰ و ۲۰۱۶) از الگوریتم خوشه بندی از طبقه بندی صنعت بازار محصول مبتنی بر متن استفاده

[1] Holmes & Stevens

[2] Fan & Lang

کردند برای متنهای توصیف محصولات در فایل بندی K-۱۰ برای شناسایی رقبای بازار محصول در حالی که در این تحقیق شرکتهای دارای نسبتهای مالی مشابه انتخاب می‌شود. و معیارهای برای کشف ناهنجاریهای حسابداری ساخته می‌شود این الگو برای پیش بینی سایر رویدادهای مهم حسابداری نیز کاربرد دارد.

این تحقیق حیطه کاربرد نسبتهای مالی را به انتخاب معیار در حوزه تحلیل نسبتها گسترش می دهد. تحقیقات قبلی نشان داده اند که نسبتهای مالی می‌توانند اهمیت ویژگی های شرکت را در رابطه با رویدادهای مهم توصیف کنند. (گوپتا و هیومنگر ۱۹۷۲، دیچارد و همکاران ۲۰۱۱، اک و همکاران ۲۰۱۳) در این تحقیق نسبت های مالی با ابزار یادگیری ماشین، خوشه بندی K-Medians ترکیب می‌شود که به منظور شناسایی شرکتهای هم رده با ویژگی های مشابه و هدف این تحقیق این است که نشان دهد مقایسه نسبت بین یک شرکت و هم رده‌هایش که با خوشه بندی K-Medians گروه بندی شده اند یک متغیر قدرتمند در مدلها برای پیش بینی ورشکستگی است.

این تحقیق با اضافه کردن متغیر انحراف شرکت از شرکتهای هم رده خود (DevScore) درمدلهای پیش بینی ورشکستگی کشف تحریفات با اهمیت حسابداری این مدل ها ارتقا می بخشد. هدف تحقیق این است که روشی را تدوین نماید که بتواند شرکت ها را به گروه‌های مختلفی طبقه بندی کند تا شرکت های یک گروه در مقایسه با شرکت های گروه‌های دیگر بیشتر از یکدیگر باشند. این کار با استفاده از الگوریتم خوشه‌ای، یک ابزار یادگیری بدون نظارت ماشین انجام می‌شود که با به حداقل رساندن ناهمگونی درون گروه‌ها یک مجموعه داده را به گروه‌های مختلف تقسیم کند. با توجه به مطالب بیان شده فرضیه‌های پژوهش به شرح زیر تدوین می‌گردد:

فرضیه اول: نسبت های مالی می‌تواند توانایی مدل پیش بینی ورشکستگی را بهبود ببخشند.

تحلیل خوشه‌بندی یا بطور خلاصه خوشه‌بندی، فرآیندی است که در آن می‌توان مجموعه‌ای از اشیاء را براساس یک تابع مشابهت یا فاصله به گروه‌های مجزا افراز کرد. هر افراز در این حالت، یک خوشه نامیده می‌شود. معمولا ۴ گروه اصلی برای الگوریتم‌های خوشه‌بندی وجود دارد. الگوریتم‌های خوشه‌بندی تفکیکی، الگوریتم‌های خوشه‌بندی سلسله مراتبی، الگوریتم‌های خوشه‌بندی برمبنای چگالی و الگوریتم‌های خوشه‌بندی برمبنای مدل. فرض بر این است که خوشه‌بندی مناسب است که اعضاء هر خوشه با توجه به ویژگی‌هایی که دارند به یکدیگر بسیار شبیه بوده و در عوض میزان شباهت بین خوشه‌ها کمترین مقدار ممکن باشد. متاسفانه بیشینه‌سازی شباهت درون خوشه‌ها و کمینه‌سازی شباهت بین خوشه‌ها امکان‌پذیر نیست یا محاسبات بسیار پیچده‌ای دارد. به همین دلیل یکی از حالت‌ها را انتخاب می‌کنند. در خوشه‌بندی افرازی اغلب موضوع و هدف اصلی، بیشینه‌سازی (کمینه‌سازی) میزان شباهت (فاصله) مشاهدات هر خوشه است[۳۱،۹]. تکنیک های یادگیری ماشین توانایی استخراج اطلاعات معنی دار از داده‌های بدون ساختار را دارند و در عین حال به طور موثر با غیر خطی برخورد می‌کنند. با این حال، استفاده از تکنیک های پیشرفته یادگیری ماشین در پیش بینی مالی هنوز یک منطقه نسبتاً جدید برای کشف محققان است[۲۵].

خوشه بندی K–Medians یک تکنیک یادگیری ماشین است که در مطالعات حسابداری کاربرد منحصر به فردی دارد[۱۲،۱۹،۲۰]. الگوریتم های یادگیری ماشین می‌تواند به دو طبقه الگوریتم یادگیری نظارتی و الگوریتم یادگیری غیر نظارتی طبقه بندی شود بطور کلی یادگیری ماشین می‌تواند در سه نوع از روش های تحقیق به کار گرفته شود. که عبارتند از رگرسیون، طبقه بندی و خوشه بندی. ون لیبرگن[1](۲۰۱۷) بیان می دارد الگوریتم خوشه بندی K–Medians مجموعه معینی از مشاهدات را به گروه‌های معنی دار به گونه‌ای تقسیم می‌کند که مشاهدات یک گروه مشابه یکدیگر و با مشاهدات گروه‌های دیگر تفاوت داشته باشند که به‌این ترتیب هدف ما که عبارتست از شناسایی شرکت هایی که دارای نسبتهای مالی مشابه هستند و ویژگی های خاصی را نشان می دهند، را برآورده می سازد. سپس مدل پیش بینی ورشکستگی برای آزمون سودمندی الگوی ارائه شده انتخاب می‌شود که رویداد مهم هستند که می‌توانند زیان های سرمایه گذاری زیادی را به بار بیاورند. هرچند این تحقیق سودمندی طبقه بندی شرکت ها به کمک خوشه بندی K–Medians را برای بهبود و ارتقای مدل ک پیش بینی ورشکستگی بررسی می‌کند. اما می‌تواند برای سایر مدل های حسابداری نظیر ارزیابی ریسک و عملیات حسابرسی تعمیم داده شود. به عبارت دیگر فرضیه دوم تحقیق نیز به شرح زیر است:

فرضیه دوم: خوشه بندی K–Medians می‌تواند توانایی مدل پیش بینی ورشکستگی را بهبود ببخشند.

پیشینه پژوهش

پیش بینی ورشکستگی حسابداری همواره یکی از موضوعات بحث برانگیز در متون و نوشتارهای مرتبط با حسابداری بوده است و در این بخش به برخی تحقیقات انجام شده در این زمینه اشاره می شود:

قلی زاده و همکاران(۱۳۹۸) به پیش‌بینی ورشکستگی با مدل یادگیری ماشین سریع مبتنی بر کرنل بهینه‌شده با الگوریتم گرگ خاکستری پرداختند. در این مدل یاد شده روی داده‌های ۱۳۶ نمونه از شرکت‌های بورسی در بازه زمانی ۱۳۹۴ تا پایان خرداد ۱۳۹۷، پیاده‌سازی شد و در تمامی معیارهای ارزیابی، مدل‌های طبقه‌بندی، دقت، خطاهای نوع اول و دوم و ناحیه تحت منحنی ROC، در مقایسه با الگوریتم ژنتیک، کارایی بهتری ارائه کرد و معناداری آن نیز به تأیید رسید. با توجه به دقت بسیار خوب الگوریتم گرگ خاکستری و همچنین عملکرد بهتر آن در مقایسه با الگوریتم ژنتیک، می‌بایست برای پیش‌بینی ورشکستگی شرکت‌های تولیدی ایران، چه برای اهداف سرمایه‌گذاری و اعتباردهی و چه به‌منظور استفاده مدیریت داخلی شرکت، از الگوریتم گرگ خاکستری بهره برد.

اعتبار و همکاران(۱۳۹۸) به بررسی مقایسه توانایی الگوریتم‌های یادگیری ماشین آدابوست و طبقه‌بندی احتمالی بیزین در پیش‌بینی بیش‌اطمینانی مدیران شرکت‌های بازار سرمایه‌ایران طی سال‌های ۱۳۸۷ تا ۱۳۹۵پرداختند. در این راستا الگوریتم هوش مصنوعی آدابوست و احتمالی بیزین به‌منظور ارائه مدل جهت پیش‌بینی بیش‌اطمینانی مدیریت در شرکت‌های پذیرفته شده بورس اوراق بهادار تهران، برای سال جاری و یک سال آتی اعتبارسنجی شدند. درمجموع، تعداد ۳۴ صنعت مختلف به‌عنوان نمونه آماری

[1] van Liebergen

انتخاب شد (اطلاعات مالی تعداد ۳۱۴۵ شرکت-سال در مجموع). نتایج نشان داد که، مدل‌های غیرخطی پیش‌بینی الگوریتم آدابوست و الگوریتم احتمالی بیزین به دست آمده، هر دو توانایی پیش‌بینی بیش‌اطمینانی مدیریت برای سال جاری و یک سال بعد را دارند؛ اما مدل پیش‌بینی آدابوست در مقایسه با مدل پیش‌بینی بیزین نتایج بهتری را برای پیش‌بینی اعتماد به نفس کاذب مدیریت دارد که نشان‌دهنده قدرت بیشتر در یادگیری و کارآیی بهتر این مدل به‌منظور پیش‌بینی بیش‌اطمینانی مدیریت می‌باشد.

اژدری و همکاران(۱۳۹۸) به بررسی انتخاب پرتفوی سهام جهت شناسایی شرکت های برتر جهت سرمایه گذاری با روش محدودیت ال با استفاده از روش یادگیری ماشین پرداختند. بدین منظور شرکت هایی که بصورت سالاته در سبد بهینه سهام جهت سرمایه گذاری قرارگرفته اند به عنوان شرکت های برتر جهت سرمایه گذاری،معرفی شدند. نمونه آماری تحقیق شامل داده‌های مالی ۲۵۱ شرکت پذیرفته شده در بورس اوراق بهادار تهران طی سال‌های ۱۳۹۱ تا ۱۳۹۶ و نتایج نشان داد که الگوریتم بهینه‌سازی ازدحام ذرات به‌صورت سالانه و همچنین به‌صورت یکجا در دوره مورد بررسی قادر به انتخاب شرکت‌های برتر با استفاده از مدل حداقل واریانس MVP با محدودیت_l-1_l∞ است.

شاهوردیانی و خواجه زاده(۱۳۹۷) به بررسی تحلیل نوسان قیمت سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از روش یادگیری ماشین پرداختند. بدین منظور از اطلاعات ۱۷۵۰ سال شرکت با نرم افزار متلب استفاده شده و نتایج نشان داد که الگوریتم غیر خطی ارتباطی در پیش بینی نوسان قیمت سهام با استفاده از متغیر های حسابداری نسبت به الگوریتم ارتباطی خطی توانایی بالاتری دارد. لذا به صاحبان سرمایه و تصمیم گیران شرکت توصیه می‌شود در تصمیم گیری‌های خود پیرامون سرمایه‌گذاری در بازار سرمایه از قدرت پیش‌بینی الگوریتم‌های هوش مصنوعی بویژه حالت غیر خطی الگوریتم ارتباطی استفاده کنند.

آشتاب و همکاران(۱۳۹۶) به بررسی مقایسهٔ دقت مدل‌های پیش‌بینی بحران مالی و تأثیر آن بر ابزارهای مدیریت سود پرداختند. برای پیش‌بینی بحران مالی، مدل‌های یادگیری ماشین و مدل‌های آماری ۳۱۲ شرکت پذیرفته شده در بورس اوراق بهادار تهران طی سال‌های ۱۳۸۵ تا ۱۳۹۴ با یکدیگر مقایسه شدند و به کمک آزمون مقایسهٔ میانگین، مشخص شد که از نظر پیش‌بینی بحران مالی، مدل‌های یادگیری ماشین نسبت به مدل‌های آماری دقت بیشتری دارند. سپس رابطهٔ بهترین مدل پیش‌بینی بحران مالی به‌دست آمده از مرحلهٔ قبل و ابزارهای مدیریت سود با استفاده از رگرسیون چندگانهٔ خطی بررسی شد و مشخص گردید که بحران مالی پیش‌بینی شده با مدیریت سود جریان‌های نقدی عملیاتی، رابطهٔ معکوس و معنادار دارد و با مدیریت سود هزینه‌های تولیدی و مدیریت سود اقلام تعهدی، رابطهٔ مستقیم و معناداری برقرار می‌کند.

هانگ و ین[1](۲۰۱۹) به مقایسه عملکرد در بین الگوریتم های یادگیری ماشین برای پیش بینی بحران مالی پرداختند. چهار مدل یادگیری ماشینی از جمله دستگاه بردار پشتیبانی سنتی، توسعه حافظه انجمنی

[1] Huang & Yen

ترکیبی با ترجمه، خوشه بندی ترکیبی فازی و افزایش شیب شدید در عملکرد پیش بینی مقایسه شدند. شانزده متغیر مالی از صورت های مالی شرکت های تایوانی به عنوان ورودی به شش رویکرد انتخاب و یافته‌های تجربی در دوره ۲۰۱۰-۲۰۱۶، نشان داد که از بین چهار الگوریتم تحت نظارت، افزایش شیب شدید دقیق ترین پیش بینی بحران مالی را ارائه می دهد. علاوه بر این، مدل توسعه حافظه انجمنی ترکیبی با ترجمه هیبریدی قادر به پیش بینی دقیق تر از استفاده از دستگاه بردار پشتیبانی سنتی است. چن و همکاران[1](۲۰۱۸) به بررسی قدرت پیش بینی کننده شخصیت سرمایه گذار در پیش بینی عملکرد سرمایه گذاری با استفاده از مدل های یادگیری ماشین پرداختند. سرمایه گذاران از طیف گسترده‌ای از استراتژی های سرمایه گذاری استفاده می‌کنند و استراتژی انتخاب شده تأثیر مستقیمی بر عملکرد فردی سرمایه گذار دارد. نتایج نشان داد که شخصیت سرمایه گذاران بر عملکرد معاملات کوتاه مدت و بلند مدت تأثیر می گذارد. با استفاده از مدلهای آماری، سرمایه گذاران با ویژگیهای شخصیتی مانند وجدان، موافقت، برونگرایی و باز بودن، بر خلاف عصبی بودن، در دراز مدت عملکرد بهتری دارند. استفاده از الگوریتم های یادگیری ماشینی همچنین نشان داد که سرمایه گذاران با ویژگی های برونگرایی و باز بودن احتمالاً در شرایط بلند مدت سرمایه گذاری بیشتری دارند. این نتایج بینشی را در اختیار سرمایه گذاران قرار می دهد که می‌تواند به آنها در انتخاب استراتژی های سرمایه گذاری در آینده کمک کند.

پردومو و همکاران[2](۲۰۱۸) به بررسی یک مدل قابلیت اطمینان برای ارزیابی حاکمیت شرکتی با استفاده از تکنیک های یادگیری ماشین پرداختند. مدیریت شرکت، کارایی و اثربخشی عملکرد و تصمیمات شرکتها را برای اطمینان از ایجاد ارزش سهامداران و ریسک بهینه ارزیابی می‌کند. از آنجا که فرایند تصمیم گیری سرمایه گذاران تا حد زیادی به اطلاعات مالی و گزارش های شرکت بستگی دارد، شفافیت سرمایه‌ای برای ثبات یک شرکت یا حتی ثبات یک کشور از طریق بخش شرکت است. این تحقیق، تئوری قابلیت اطمینان سیستم را برای الگوبرداری از رفتار شرکتها در رابطه با مکانیسم های مدیریت شرکت آنها ارائه داد. ارزیابی چارچوب حاکمیت شرکتی را با استفاده از نمونه‌ای از ۱،۱۰۹ شرکت ذکر شده در ایالات متحده در طی دوره ۲۰۰۲-۲۰۱۴، گزارش اطلاعات مالی و غیر مالی به عنوان مؤلفه‌های سیستم حاکمیت شرکت ها و بازده دارایی ها به عنوان خروجی سیستم نشان داده شد. روش پیشنهادی برای مدل سازی سایر زیر سیستم های غیر مهندسی نیز مفید است. اگر این سیستم ها به درستی کار نکنند، شرکت ها، بازارهای مالی یا حتی اقتصاد در معرض خطر قابل توجهی قرار می گیرند.

دیاس پیوا و همکاران[3](۲۰۱۸) به بررسی تصمیم گیری برای تجارت در بازار مالی: روش تلفیقی از یادگیری ماشین و انتخاب پرتفوی پرداختند. این مطالعه یک الگوی تصمیم گیری منحصر به فرد برای سرمایه گذاری های معاملات روزانه در بورس اوراق بهادار ارائه می دهد. در این راستا، این مدل با استفاده از روش همجوشی یک طبقه بندی کننده مبتنی بر یادگیری ماشین، با روش ماشین بردار پشتیبانی و روش

[1].Chen et al.
[2].Perdomo et al.
[3].Dias Paiva et al.

میانگین واریانس برای انتخاب پرتفوی تهیه شده است. ارزیابی تجربی مدل براساس دارایی های شاخص بورس اوراق بهادار سائوپائولو (ایبووپسا) بود. در کل ۸۱ ترتیب پارامتر فرموله شد. این آزمایشات با استفاده از داده‌های تاریخی برای ۳،۷۱۶ روز معاملاتی برای تجزیه و تحلیل خارج از نمونه انجام شد. شبیه سازی ها بدون احتساب هزینه‌های معامله و همچنین با درج بخشی از این هزینه‌ها انجام شد. مدل اصلی پیشنهادی نتایج قابل توجهی را نشان داد، اگرچه تقاضا برای ارزش معاملات می‌تواند یک عامل محدود کننده برای اجرای آن باشد. با این وجود، این مطالعه کاربرد تئوریک یادگیری ماشین را گسترش داده و یک رویکرد بالقوه عملی برای پیش بینی قیمت سهام ارائه می دهد.

روش شناسی پژوهش

خوشه بندی

هدف تحقیق حاضر تدوین الگویی است که بتواند شرکت ها را به گروه‌های مختلف بگونه‌ای طبقه بندی کندکه شرکت های هر گروه نسبت به شرکت های سایر گروه‌ها به لحاظ همسانی در نسبت های مالی مشابهت بیشتری داشته باشند. این کار را می‌توان با الگوریتم خوشه بندی، یک ابزار غیرنظارتی یادگیری ماشین انجام داد که با به حداقل رساندن نا همسانی در گروه‌ها و به حداکثر رساندن ناهمسانی در گروهی، یک مجموعه داده شده را به گروه‌های مختلف تقسیم می‌کند[۳۱]،]۹]. الگوریتم خوشه بندی کاربردهای گسترده‌ای در تحقیقات حسابداری و حسابرسی دارد از دهه ۱۹۷۰، پژوهشگران حسابداری از جمله جنسن[1](۱۹۷۱)و گاپتا و همکاران[2](۱۹۷۲) از خوشه بندی برای شناسایی نهادهایی با خصوصیات مشابه و تقسیم یک نمونه داده شده به زیرمجموعه‌های معنی دار استفاده کرده اند تا اقلام موجود در خوشه همگن هستند. انواع مختلفی از الگوریتم های خوشه بندی وجود دارد، از بین آنها کای میانگین (K-Means) و کای میانه (K-Medians) دو تا از پرطرفدارترین الگوها هستند. در این تحقیق از الگوی کای میانه (K-Medians) استفاده شده است.چون خوشه بندی کای میانه نسبت به خوشه بندی کای میانگین پایداری بیشتری در برابر نقاط پرت دارد(گوها و همکاران ۲۰۰۳). الگوریتم کای میانه بصورت زیر عمل می‌کند:

هر مشاهده سال -شرکت با نسبت های مالی می‌تواند بعنوان نقطه (Xt) در یک فضای چند بعدی مشاهده شود. خوشه بندی کای میانه روی مجموعه‌ای (X) از نقطه n کار می‌کند. و تعداد K مرکز از X را انتخاب می‌کند { c_k... و c_2 و c1} و تعداد K خوشه {C_K... و C_2 و C1} تشکیل میدهد. بطوریکه مجموعه فاصله‌ها(S) از هر x_t به مرکز خوشه خود حداقل باشد. در واقع هدف کای میانه به حداقل رساندن تابع زیر است:

Min:

$$S = \sum_{I=1}^{K} \quad \sum_{X \in Ck} L_I \ (x, c_i)$$

[1] Jensen
[2] Gupta et al.

که در آن L1 عبارتست از فاصله منتهتن (Manhattan Distance) بعنوان مقادیر مطلق تفاوت مختصات به عبارتی فاصله منهتن بین هر دو نقطه p و q در یک فضای چند بعدی عبارتست از :

$$L_1(x,y) = \sum_{i}^{m} = I[p_i, q_i]$$

الگوریتم خوشه بندی کای میانه ابتدا K نقطه تصادفی در X را انتخاب می‌کنند و بعنوان مراکز اولیه بعد با تخصیص هر (n– K) نقطه به نزدیکترین مرکز خوشه‌ها را ایجاد می‌کند و سپس نقاط میانه خوشه‌ها را محاسبه می‌کند و مراکز را با نقاط میانه جایگزین می‌کند. این الگوریتم گامهای دوم و سوم را تکرار می‌کند تا به کوواریانس دست یافته شود. از نرم افزار های مختلفی برای تحلیل های خوشه بندی استفاده می‌شود. در این تحقیق از نرم افزار Stata استفاده خواهد شد.

نسبت های مالی

پژوهشگران می‌توانند مجموعه‌ای از نسبت های مالی را بر اساس یک هدف تحقیقاتی خاص انتخاب کنند. به عنوان مثال، اگر هدف از تحقیق پیش بینی یک رویداد شرکتی خاص باشد، نسبت های مالی انتخاب شده باید مربوط به‌این رویداد باشد. این مطالعه برای نشان دادن کاربرد روش پیشنهادی: پیش بینی ورشکستگی شرکت ها را مورد بررسی قرار می دهد. برای جلوگیری از تغییر نسبت ها از سطوح نسبت استفاده شده است که از کاهش چشمگیر در اندازه نمونه ناشی از مخرج های با ارزش صفر یا داده‌های مفقود شده جلوگیری میکند..در تجزیه و تحلیل پیش بینی ورشکستگی، خوشه بندی بر اساس پنج نسبت مالی معرفی بکارگرفته شده در مدل پیش بینی ورشکستگی التمن[۱] استفاده شده است. این نسبت ها عبارتند از:

نسبت خالص سرمایه در گردش به مجموع داراییها (WC/TA)، نسبت سود انباشته به مجموع دارائیها (RE/TA)، نسبت سود قبل از بهره و مالیات به مجوع دارئیها(EBIT/TA)، نسبت ارزش بازار حقوق سهام به ارزش دفتری بدهی(ME/TA) و نسبت فروش به مجموع دارائیها(SALE/TA). لازم به ذکر است که در اینجا همه نسبت های مالی ممکن در مدل لحاظ نمی‌شود زیرا زیاد بودن متغیرها به معنی افزایش تعداد ابعاد در تجزیه تحلیل کای میانه است که ممکن است اثر بخشی خوشه بندی را بدلیل وجود نقاط پرت و داده‌های مفقود شده، کاهش دهد. بنابراین، برای اینکه یک نمونه کاملاً منطقی برای انجام تجزیه و تحلیل خود داشته باشیم، همه نسبت های ممکن را در مدل خود لحاظ نمی‌کنیم بلکه در عوض مجموعه‌ای از نسبت ها را انتخاب می‌کنیم که مربوط به رویداد مورد علاقه شرکت باشند.

جامعه آماری مورد بررسی در این تحقیق، شرکت های پذیرفته شده در بورس اوراق بهادار تهران در بازه زمانی ۱۳۸۹ تا ۱۳۹۸ است که به روش حذفی سیستماتیک انتخاب شده اند. بدین ترتیب که ابتدا تمامی شرکت های پذیرفته شده در بورس تهران انتخاب و در نهایت شرکت هایی مورد بررسی قرار گرفتند که شرایط زیر را دارا بودند:

۱. شرکت‌هایی که در بازه زمانی تحقیق (سال ۱۳۸۹ الی ۱۳۹۸) در بورس اوراق بهادار حضورداشته باشند.

۲. شرکت‌هایی که سال مالی آن‌ها منتهی به ۲۹ اسفند باشد.
۳. شرکت‌هایی که جزء شرکت‌های لیزینگ و واسطه‌گری مالی نباشند.
۴. اطلاعات لازم در خصوص آن‌ها در بازه زمانی تحقیق (۱۳۸۹ الی ۱۳۹۸) در دسترس باشد.

در نهایت با توجه به موارد یادشده ۳۰۸ شرکت در دوره تحقیق برای مطالعه انتخاب گردید.

به منظور آزمون این فرضیه از مدل پیش بینی ورشکستگی شام وی[1](۲۰۰۱) که دینگ و همکاران نیز در سال ۲۰۱۹ بکار بردند، استفاده می‌شود. در این مدل از متغیر های حسابداری و متغیرهای بازار استفاده میشود.

$$1)\ Bankrupt_{it+1} = \beta_0 + \beta_1 {NI}/{TA_{it}} + \beta_2 {TL}/{TA_{it}} + \beta_3 RSize_{it} + \beta_4 ERet_{it} + \beta_5 Sigma_{it} + \varepsilon$$

$$2)\ Bankrupt_{it+1} = \beta_0 + \beta_1 {NI}/{TA_{it}} + \beta_2 {TL}/{TA_{it}} + \beta_3 RSize_{it} + \beta_4 ERet_{it} + \beta_5 Sigma_{it} + \beta_6 DevScore_{it} + \varepsilon$$

تعریف عملیاتی و طبقه‌بندی متغیرهای تحقیق حاضر به شرح زیر است:

پیش بینی ورشکستگی(Bankrupt): متغیر وابسته تحقیق بوده که با بهره گیری از الگوی مطرح شده شاموی(۲۰۰۱) با استفاده از نسبت های مالی، نسبت سرمایه در گردش خالص به مجموع دارائی ها، نسبت سود انباشته به مجموع دارائیها، نسبت سود قبل از بهره و مالیات به مجموع دارائیها، نسبت ارزش بازار حقوق صاحبان سهام به ارزش دفتر مجموع بدهی ها و نسبت فروش به دارائیها در صورتی که نسبت محاسبه شده بالاتر از میانه باشد مقدار یک و در غیر این صورت صفر می‌باشد.

متغیر های مستقل شامل:

نسبت سود خالص به مجموع دارایی ها(NI/TA): از تقسیم سود خالص به مجموع دارایی بدست می آید.

نسبت مجموع بدهی ها به مجموع دارایی ها(TL/TA): از تقسیم مجموع بدهی ها به مجموع دارایی ها بدست می آید.

نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار(RSIZE) که یک معیاری از سرمایه گذاری در بازار است که از لگاریتم ارزش بازار هر شرکت تقسیم بر کل ارزش بازار است بدست می آید.

ExRET: بازده مازاد سهام که از تفاضل بازده سهام سال قبل و بازده شاخص کل (هم وزن) سال قبل بدست می آید.

سیگما(Sigma): انحراف استاندارد باقیمانده رگرسیون بازده ماهانه سهام برروی بازده بازار در سال قبل.

تعداد صنایع در بورس اوراق بهادار تهران شرکت های موجود در هر صنعت در سال t-1 بعنوان شرکت های هم رده در نظر گرفته می‌شوند. در سال t. نسبت های مالی هر شرکت با میانگین شرکت های هم رده برای محاسبه Devscore مقایسه می‌شود. با ادامه‌این گام ها پس از محاسبه Devscore برای گروه‌هایی که بر اساس صنعت تشکیل شده اند، به منظور ارزیابی اینکه‌ایا این مدل اصلاح شده با

[1] Shumway

Devscore مدل پیش بینی ورشکستگی را بهبود می بخشد یا خیر، بعنوان یک متغیر مستقل وارد مدل پیش بینی ورشکستگی می‌شود.

DevScore : متغیر مجازی بوده که تعداد نسبت های بالاتر از میانگین هم رده مشخص شده با روش خوشه بندی K-medians می‌باشد.

DevScore(ISC): متغیر مجازی بوده که تعداد نسبت های بالاتر از میانگین هم رده مشخص شده در هر صنعت می‌باشد.

برای خوشه‌بندی شرکت‌ها در سال t ام از شاخص‌های NI/TA، TL/TA، RSIZE، ERET و Sigma استفاده شده است. به‌این ترتیب هر نقطه در الگوریتم خوشه‌بندی کای-میانه دارای ۵ ویژگی یا متغیر است پس یک نقطه پنج بعدی در نظر گرفته می‌شود. الگوریتم خوشه‌بندی برای هر سال اجرا شده و با خوشه‌های هم رده به تفکیک سال، استخراج می‌شوند. نتیجه حاصل از خوشه‌بندی برای محاسبه شاخص DevScore به کار برده شده است. این شاخص به عنوان یک تابع نشانگر عمل می‌کند و برای هر ویژگی یا شاخص مالی، مقایسه‌ای بین شرکت‌هایی هر خوشه و میانگین آن ویژگی در خوشه صورت می‌گیرد. که مقدار هر یک از ویژگی آن‌ها از میانگین خوشه هم رده شان بیشتر باشد عدد ۱ در نظر گرفته می‌شود. در غیر اینصورت عدد صفر لحاظ خواهد شد. مجموع این مقادیر شاخص DevScore هر شرکت را در سال t مشخص می‌کند.

یافته‌های پژوهش

نخستین مرحله در هر تحلیل آماری و تجزیه و تحلیل اطلاعات، محاسبه‌ی شاخص‌های توصیفی می‌باشد. بنابراین برای ورود به مرحله‌ی تجزیه و تحلیل، آماره‌ی توصیفی داده‌ها که شامل شاخص‌های مرکزی، پراکندگی و انحراف از قرینگی بوده و همچنین آزمون جارگ‌برا که توزیع نرمال پسماندها را بررسی می‌کند، محاسبه گردیده و نتایج در جدول (۱) درج شده است.

جدول(۱): آمار توصیفی و فراوانی متغیرهای پژوهش

متغیر	NI/TA	TL/TA	RSIZE	ERET	Sigma	DevScore	DEVSIC
میانگین	0.113	0.486	10.21-	3.655	$1.37E^{-16}$	918.1	1.787
میانه	0.109	0.509	10.27-	0.549	0.294-	2	2
انحراف معیار	0.204	0.191	1.981	13.378	1.194	0.881	0.938
چولگی	0.095-	0.326-	0.348	10.317	10.400	-0.072	0.031
کشیدگی	140.576	1.873	3.090	241.38	224.68	495.2	2.154
جارک - برا	2429019	217.73	63.31	7347576	6362475	40.10	92.30
سطح معنی داری	0.000	0.000	0.000	0.000	0.000	0.000	0.000

متغیر	NI/TA	TL/TA	RSIZE	ERET	Sigma	DevScore	DEVSIC
مشاهده‌ها	3080	3080	3080	3080	3080	3080	3080
شرکت های بدون ورشکستگی		فراوانی	درصد فراوانی	شرکت های با ورشکستگی		فراوانی	درصد فراوانی
		2102	68.2			978	31.8

مهمترین شاخص مرکزی میانگین است که نشان دهنده نقطه تعادل و مرکز ثقل توزیع بوده، برای مثال میانگین نسبت نسبت درآمد خالص به کل دارایی(NI/TA) برابر با ۰,۱۱۳ می‌باشد، که نشان می دهد بیشتر داده‌های مربوط به‌این متغیر حول این نقطه تمرکز یافته اند. پارامترهای پراکندگی، به طور کلی معیاری برای تعیین میزان پراکندگی داده‌ها از یکدیگر یا میزان پراکندگی آنها نسبت به میانگین است. از جمله مهمترین پارامترهای پراکندگی انحراف معیار است. مقدار این پارامتر برای متغیر نسبت درآمد خالص به کل دارایی(NI/TA) برابر ۰,۲۰۴ است که میزان پراکندگی متغیر مزبور را در اطراف میانگین نشان می دهد. میزان عدم تقارن منحنی فراوانی را چولگی می نامند. اگر ضریب چولگی صفر باشد، جامعه کاملاً متقارن است و چنانچه ضریب مثبت باشد، چولگی به راست و اگر منفی باشد، چولگی به چپ وجود خواهد داشت. به عنوان مثال ضریب چولگی متغیر مزبور ۰,۰۹۵- می‌باشد، یعنی این متغیر چولگی به چپ دارد. میزان کشیدگی اگر حدود ۳ باشد، منحنی فراوانی از لحاظ کشیدگی وضعیت متعادل و نرمال خواهد داشت، اگر این مقدار بزرگتر از ۳ باشد منحنی برجسته و اگر کوچکتر از ۳ باشد منحنی پهن می‌باشد. کشیدگی اکثر متغیر ها در این پژوهش بیشتر از ۳ است. نتایج حاصل از آزمون جارک - برا مبنی بر نرمال بودن متغیرهای پژوهش نشان می دهد، با توجه به‌اینکه سطح معنی داری کلیه متغیرها کمتر از ۰,۰۵ است، در نتیجه فرضیه صفر مبنی بر نرمال بودن متغیرها رد می‌شود به عبارتی توزیع داده‌ها نرمال نمی‌باشد. همچنین در این پژوهش با توجه به‌اینکه متغیر وابسته تحقیق یعنی ورشکستگی بصورت صفر و یک می‌باشد از آزمون های ناپارامتریک استفاده می گردد. با توجه به نتایج جدول(۲) ارزیابی همبستگی بین متغیرها بر اساس ضرایب همبستگی، شواهد نشان دادند بیشترین میزان همبستگی بین نسبت درآمد خالص به کل دارایی(NI/TA) با متغیر خوشه بندی براساس صنعت (DEVSIC) میزان آن ۳۴,۸ درصد و کمترین میزان همبستگی بین سیگما با ERET به میزان ۰,۰۷۳- قابل مشاهده است. طبق نتایج حاصل شده در آزمون همبستگی، در این پژوهش همخطی به معنی همبستگی قوی بین متغیرها وجود ندارد و به نظر می رسد مشکل خاصی از این بابت در تحلیل مدل های رگرسیون وجود نداشته باشد.

جدول(۲): همبستگی متغیرهای پژوهش

متغیر	NI_TA	TL_TA	RSIZE	ERET	SIGMA	DEVSCORE	DEVSIC
NI_TA	1						
TL_TA	0.118	1					
RSIZE	0.270	0.061	1				
ERET	0.065	0.007	0.014	1			
SIGMA	0.065	0.003-	0.015	0.073-	1		
DEVSCORE	0.209	0.013	-0.050	0.026	0.004	1	
DEVSIC	0.348	0.048	0.105	0.018	0.029	0.285	1

نتایج آزمون و پیش فرض های مدل های رگرسیونی در جدول ۳ ارائه شده اند. با توجه به جدول ۳ طبق آماره اف آزمون هاسمر– لمشو و معناداری آن که کمتر از سطح خطای ۵ درصد است، همسانی واریانس خطاها وجود دارد و همچنین طبق آماره اف لیمر آزمون چاو و معناداری آماره آن که کمتر از سطح خطای ۵ درصد است، آزمون مدل ها باید به روش ترکیبی انجام شود. طبق آماره کایدو آزمون هاسمن و معناداری آماره آن که کمتر از سطح خطای ۵ درصد است، آزمون مدل ها باید به روش اثرات ثابت انجام شود.

جدول(۳): نتایج آزمون انتخاب الگوی مدل های رگرسیونی جهت آزمون فرضیه‌های پژوهش

مدل های پژوهش	آزمون	آماره	مقدار	معناداری	نتیجه آزمون
مدل اول	چاو	اف لیمر	۶,۴۸۸	۰,۰۰۰	داده‌های ترکیبی
	هاسمن	کای دو	۲۵,۲۹۲	۰,۰۰۰	اثرات ثابت
	هاسمر– لمشو	اف	۸۱,۶۸	۰,۰۰۰	همسانی واریانس خطاها
مدل دوم	چاو	اف لیمر	۶,۷۲۴	۰,۰۰۰	داده‌های ترکیبی
	هاسمن	کای دو	۳۵,۷۰۶	۰,۰۰۰	اثرات تصادفی
	هاسمر– لمشو	کای دو	۷۶,۳۹	۰,۰۰۰	همسانی واریانس خطاها

هدف ما در مدل اول، نسبت های مالی می‌تواند توانایی مدل پیش بینی ورشکستگی در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را بهبود ببخشد. لذا می‌توان مدل را به شکل زیر تصریح نمود:

$$\text{Bankrupt}_{it+1} = \beta_0 + \beta_1 {}^{\text{NI}}/_{\text{TA}_{it}} + \beta_2 {}^{\text{TL}}/_{\text{TA}_{it}} + \beta_3 \text{RSize}_{it} + \beta_4 \text{ERet}_{it} + \beta_5 \text{Sigma}_{it} + \varepsilon$$

پس از تبیین مدل و انتخاب بهترین روش، نتایج تخمین برای شرکت‌های منتخب به شرح جدول زیر است:

جدول(۴): نتایج آزمون مدل اول

متغیر	ضریب برآورد شده	خطای استاندارد	آماره Z	سطح معنی داری
مقدار ثابت	1.331980-	0.262078	5.082379-	0.0000
TA/NI	3.405685	0.305631	11.14311	0.0000
TA/TL	0.613820	0.218118	2.814168	0.0049
RSIZE	0.050964	0.021759	2.342206	0.0192
ERET	0.005919	0.003059	1.935164	0.0530
SIGMA	0.075222	0.032890	2.287059	0.0222
DEVSCORE	0.158409	0.047744	3.317852	0.0009
آماره LR (احتمال)	264.25(0.000)	ضریب تعیین مک فادن		0.468

نتایج جدول ۴ نشان می دهد مقدار عددی ضریب تعیین مک فادن متغیر ورشکستگی با متغیرهای مستقل و توضیحی ۰,۴۶۸ می‌باشد. یعنی ۴۶,۸ درصد تغییرات متغیر وابسته ورشکستگی توسط متغیرهای متغیرهای نسبت درآمد خالص به کل دارایی، نسبت کل بدهی به کل دارایی، نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار، شاخص بازده سهام در سال قبل، سیگما قابل توضیح است. از سوی دیگر مقدار احتمال برآوردشده برای آماره‌ی LR برابر با ۰,۰۰۰ بوده لذا معنی‌داری کل رگرسیون قابل تأیید است. نتایج آزمون فرضیه اول نشان داد تاثیر ورشکستگی بر متغیرهای نسبت درآمد خالص به کل دارایی(TA / NI) آماره Z (۱۱,۱۴۳) و معناداری (۰,۰۰۰)، بدهی به کل دارایی (TA/TL) (۲,۸۱۴) و (۰,۰۰۴)، نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار (RSIZE) (۲,۳۴۲) و (۰,۰۱۹)، سیگما (SIGMA) (۲,۲۸۷) و (۰,۰۲۲)، DEVSCORE (۳,۳۱۷) و (۰,۰۰۰) معنادار می‌باشد. همچنین شاخص بازده سهام در سال قبل (ERET) با توجه به آماره Z (۱,۹۳۵) و معناداری (۰,۰۵۳) که بیشتر از سطح خطای ۵ درصد است، معنادار نمی‌باشد. به عبارتی در سطح اطمینان ۹۵٪ می‌توان گفت از میان نسبت های مالی مشخص شده در مدل فقط نسبت درآمد خالص به کل دارایی (TA/NI)، کل بدهی به کل دارایی (TA/TL)، سیگما (SIGMA)، نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار (RSIZE) و DEVSCORE موجب می‌توانند توانایی مدل پیش بینی ورشکستگی را بهبود ببخشند و نسبت شاخص بازده سهام در سال قبل (ERET) توانایی بهبود مدل پیش بینی ورشکستگی در شرکت های پذیرفته‌شده در بورس اوراق بهادار تهران را دارا نمی‌باشند.

در مدل دوم هدف، خوشه بندی K–Medians براساس صنعت می‌تواند توانایی مدل پیش بینی ورشکستگی در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را بهبود ببخشد. لذا می‌توان مدل را به شکل زیر تصریح نمود:

$$Bankrupt_{it+1} = \beta_0 + \beta_1 {}^{NI}/_{TA_{it}} + \beta_2 {}^{TL}/_{TA_{it}} + \beta_3 RSize_{it} + \beta_4 ERet_{it} + \beta_5 Sigma_{it} + \beta_6 DevScore_{it} + \varepsilon$$

پس از تبیین مدل و انتخاب بهترین روش، نتایج تخمین برای شرکت‌های منتخب به شرح جدول زیر است:

جدول(۵): نتایج آزمون مدل دوم

متغیر	ضریب برآورد شده	خطای استاندارد	آماره Z	سطح معنی داری
مقدار ثابت	1.671441-	0.269349	6.205491-	0.0000
TA/NI	2.561549	0.321747	7.961376	0.0000
TA/TL	0.684934	0.222588	3.077144	0.0021
RSIZE	0.047784	0.021719	2.200121	0.0278
ERET	0.006690	0.003132	2.136419	0.0326
SIGMA	0.078786	0.033292	2.366492	0.0180
DEVSIC	0.366529	0.049193	7.450819	0.0000
آماره LR (احتمال)	309.83(0.000)	ضریب تعیین مک فادن		0.480

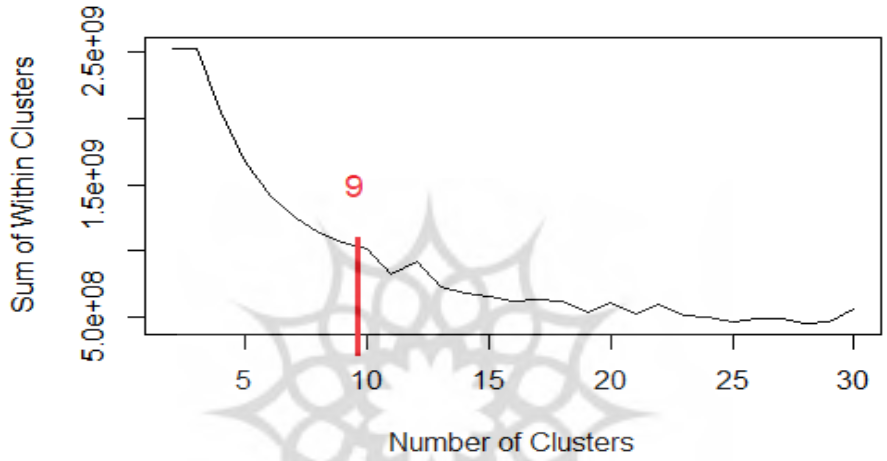
نتایج جدول ۵ نشان می دهد مقدار عددی ضریب تعیین مک فادن متغیر ورشکستگی با متغیرهای مستقل و توضیحی ۰,۴۸۰ می‌باشد. یعنی ۴۶,۸ درصد تغییرات متغیر وابسته ورشکستگی توسط متغیرهای متغیرهای نسبت درآمد خالص به کل دارایی، نسبت کل بدهی به کل دارایی، نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار، شاخص بازده سهام در سال قبل، سیگما قابل توضیح است. از سوی دیگر مقدار احتمال برآوردشده برای آماره‌ی LR برابر با ۰,۰۰۰ بوده لذا معنی‌داری کل رگرسیون قابل تأیید است.

نتایج آزمون فرضیه دوم نشان داد تاثیر ورشکستگی بر متغیرهای نسبت درآمد خالص به کل دارایی (TA/NI) آماره Z (۷,۹۶۱) و معناداری (۰,۰۰۰)، بدهی به کل دارایی (TA/TL) (۳,۰۷۷) و (۰,۰۰۲)، نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار (RSIZE) (۲,۲۰۰) و (۰,۰۲۷)، شاخص بازده سهام در سال قبل (ERET) (۲,۱۳۶) و (۰,۰۳۲)، سیگما (SIGMA) (۲,۳۶۶) و (۰,۰۱۸)، خوشه بندی براساس صنعت (DEVSIC) (۷,۴۵۰) و (۰,۰۰۰) معنادار می‌باشد. به عبارتی در سطح اطمینان ۹۵٪ می‌توان گفت خوشه بندی K-Medians براساس صنعت می‌تواند توانایی مدل پیش بینی ورشکستگی آلتمن در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را بهبود ببخشد.

برای تعیین بهترین تعداد خوشه در روش خوشه‌بندی K-Medians، یک نمودار به صورت زیر ترسیم شده که در نقطه‌ای از محور افقی شیب منحنی دارای تغییر محسوسی باشد، مناسب‌ترین خوشه‌بندی

صورت گرفته است، زیرا بعد از این شکستگی، با افزایش تعداد خوشه‌ها، تغییر زیادی در شاخص ریشه میانگین مربعات انحراف استاندارد رخ نمی‌دهد. برای داده‌های مربوط به ورشکستی شرکت‌ها این نمودار به صورت زیر درآمده است. بنابراین تعداد خوشه‌های مناسب برای اجرای الگوریتم K– Medians، ۹ خوشه است. به عبارتی شرکت هایی که پایین تر از این خوشه قرار می گیرند در معرض ورشکستگی می‌باشند.

شکل۱. خوشه‌بندی K–Medians مدل پیش بینی ورشکستگی



بحث و نتیجه گیری پژوهش

محققان غالباً از شرکت های همکار به عنوان معیار برای ارزیابی ویژگی های شرکت های هدف استفاده می‌کنند. بیشتر تلاش ها به سیستم های طبقه بندی از پیش تعریف شده صنعت مانند طبقه بندی استاندارد صنعت(SIC) و سیستم طبقه بندی صنعت آمریکا (NAICS) متکی است. این مطالعه یک روش جدید برای انتخاب همسانان را ارائه می دهد که یک تکنیک یادگیری ماشین، طبقه بندی K-Medians را به نسبت های مالی کلیدی شرکت های مرتبط با اهداف تحقیق خاص اعمال می‌کند. هر ساله، شرکت ها بر اساس نسبت های مالی به خوشه‌های مختلفی تقسیم می‌شوند. متغیر DevScore انحراف شرکت ها از همسانان خود را در سال بعد اندازه گیری می‌کند. هنگامی که DevScore افزایش می یابد، احتمال ناهنجاری های حسابداری نیز افزایش می یابد. از این روش می‌توان برای شناسایی یا پیش بینی وقایع مهم شرکت ها، مانند سوء استفاده‌های مالی مالی و ورشکستگی های شرکت استفاده کرد.

در این تحقیق به کاربرد یادگیری ماشین در ارائه الگویی برای پیش بینی ورشکستگی شرکت ها در قلمرو زمانی ۱۳۸۹ تا ۱۳۹۸ مورد آزمون و تجزیه و تحلیل قرار گرفت، نتایج و یافته‌ها نشان داد کل مدل های رگرسیونی معنی داری بود. همچنین نتایج تجزیه و تحلیل رگرسیون نشان داد(مطابق با مطالعات قبلی، شام وی، ۲۰۰۱؛ وو و همکاران، ۲۰۱۰)، در می یابیم که اگر یک شرکت دارای اهرم مالی بالاتر، اندازه

شرکت کوچکتر باشد، بازده اضافی پایین تر و انحراف استاندارد بالاتر بازده سهام احتمال ورشکستگی بیشتر است. نتایج حاکی از آن است که یک شرکت در صورت داشتن نسبت های مالی مرتبط با ورشکستگی که پایین تر از میانگین همسانان خوشه‌ای خود است، احتمالاً ورشکسته می‌شود. یافته‌های تحقیق حاضر تا حدودی با نتایج حاصل از تحقیقات قلی زاده و همکاران(۱۳۹۸)، اعتبار و همکاران(۱۳۹۸)، آشتاب و همکاران(۱۳۹۶)، هانگ و ین(۲۰۱۹)، پردومو و همکاران(۲۰۱۸) همسو می-باشد.

یافته‌های این تحقیق دلالت بر مطالعات نسبت مالی و همچنین کاربردهای عملی دارد. نتایج این استدلال را تأیید می‌کند که نسبت های مالی می‌تواند به سرمایه گذاران برای ارزیابی احتمال تجربه رویدادهای مهم شرکت ها کمک کند. تکنیک یادگیری ماشین با مدل های موجود در مورد تشخیص نادرست موارد و پیش بینی ورشکستگی ترکیب و مشاهده می‌شود که اگر اطلاعات مربوط به شرکت های همکار مبتنی بر خوشه بندی K-Medians شامل شود، این مدل ها افزایش می یابد. همچنین این رویکرد جدید که روی مجموعه‌ای از شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران آزمون شده است، نشان می دهد که چگونه یک چشم انداز کامل شفافیت (گزارش و افشای) قابلیت اطمینان سیستم را تقویت کرده و به ارزیابی چگونگی کنترل و مدیریت شرکت ها کمک می‌کند. در نتیجه، این می‌تواند به عنوان یک روش عملی تر برای طبقه بندی شرکت هایی که مطابق با منافع سهامداران فعالیت می‌کنند و کمک به آنها در فرآیند تصمیم گیری، به حساب آید. تحقیقات آینده همچنین می‌تواند این روش خوشه بندی را در زمینه‌های دیگر پیاده سازی کند. به عنوان مثال، برای حسابرسی، طرح انتخاب همسانان پیشنهادی می‌تواند ابزاری برای روش های تحلیلی در ارزیابی ریسک باشد.

فهرست منابع

۱. آشتاب، علی، حقیقت، حمید، کردستانی، غلامرضا. (۱۳۹۶). "مقایسهٔ دقت مدل‌های پیش‌بینی بحران مالی و تأثیر آن بر ابزارهای مدیریت سود". **بررسی‌های حسابداری و حسابرسی**، دوره ۲۴، شماره۲، ص ص ۱۴۷-۱۷۲.

۲. اعتبار، شکوفه، دارابی، رویا، حمیدیان، محسن، جعفری، سیده محبوبه. (۱۳۹۸). "مقایسه توانایی الگوریتم‌های یادگیری ماشین آدابوست و طبقه‌بندی احتمالی بیزین در پیش‌بینی بیش‌اطمینانی مدیران شرکت‌های بازار سرمایه ایران". **مهندسی مالی و مدیریت اوراق بهادار**. دوره ۱۰، شماره ۳۹، ص ص ۱-۲۵.

۳. اژدری، فاطمه، رهنما رودپشتی، فریدون، حمیدیان، محسن، جعفری، سیده محبوبه، باغانی، علی. (۱۳۹۸). "انتخاب پرتفوی سهام جهت سرمایه گذاری و شناسایی شرکت های برتر با روش محدودیت ال و با استفاده از روش یادگیری ماشین". **راهبرد مدیریت مالی**، (در حال انتشار). doi: 10.22051/jfm.2019.24505.1968

٤. سلیمانی امیری، غلامرضا (۱۳۸۹). "بررسی شاخص های پیش بینی کننده ورشکستگی در شرایط محیطی ایران"، **پایان نامه دکتری**، رشته حسابداری، دانشگاه تهران.

.. شاهوردیانی، شادی، خواجه زاده، سامیران. (۱۳۹۷). "تحلیل نوسان قیمت سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از روش یادگیری ماشین". **سیاست گذاری پیشرفت اقتصادی**، دوره ، شماره ۱، ص ص ۶۹-۹۱.

۶. قلی زاده سالطه، توحید؛ اقبال‌نیا، محمد؛ آقابابائی، محمد ابراهیم. (۱۳۹۸). "پیش‌بینی ورشکستگی با مدل یادگیری ماشین سریع مبتنی بر کرنل بهینه‌شده با الگوریتم گرگ خاکستری"، **تحقیقات مالی**، دوره ۲۱، شماره ۲، ص ص ۱۸۷-۲۱۲.

7. Altman, Edward I. 1968. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." **The journal of finance** 23 (4):589-609.Alpaydin, Ethem. 2009. Introduction to machine learning: **MIT press.**
8. Amit, Raphael, and Joshua Livnat. 1990. "Grouping of Conglomerates by Their Segments' Economic Attributes: Towards a More Meaningful Ratio Analysis." **Journal of Business Finance & Accounting** 17 (1):85-100.
9. Clarke, Richard N. 1989. "SICs as delineators of economic markets." **Journal of Business**:17-31.
10. Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. 1996. "Data mining: an overview from a database perspective." **IEEE Transactions on Knowledge and data Engineering** 8 (6):866-883.
11. De Franco, Gus, Ole-Kristian Hope, and Stephannie Larocque. 2015. "Analysts' choice of peer companies." **Review of Accounting Studies** 20 (1):82-109.
12. Elvis Hernandez-Perdomo , Yilmaz Guney , Claudio M. Rocco **.** (2018)**.** A reliability model for assessing corporate governance using machine learning techniques, **Reliability Engineering and System Safety**, doi: https://doi.org/10.1016/j.ress.2018.12.027
13. Fan, Joseph PH, and Larry HP Lang. 2000. "The measurement of relatedness: An application to corporate diversification**."** **The Journal of Business** 73 (4):629-660.
14. Felipe Dias Paiva , Rodrigo Tom«as Nogueira Cardoso , Gustavo Peixoto Hanaoka , Wendel Moreira Duarte **.** (2018)**.** Decision-Making for Financial Trading: A Fusion Approach of Machine Learning and Portfolio Selection, **Expert Systems With Applications**, doi: https://doi.org/10.1016/j.eswa.2018.08.003**.**
15. Guha, Sudipto, Adam Meyerson, Nina Mishra, Rajeev Motwani, and Liadan O'Callaghan. 2003."Clustering data streams: Theory and practice." **IEEE transactions on knowledge and data engineering** 15(3): 515-528.
16. Gupta, Manak C, and Ronald J Huefner. 1972. "A cluster analysis study of financial ratios and industry characteristics." **Journal of Accounting Research**:77-95.
17. Hoberg, Gerard, and Gordon Phillips. 2010. "Product market synergies and competition in mergers and acquisitions: A text-based analysis." **The Review of Financial Studies** 23 (10):3773-3811.
18. Hoberg, Gerard, and Gordon Phillips. 2016. "Text-based network industries and endogenous product differentiation." **Journal of Political Economy** 124 (5):1423-1465.

19. Holmes, Thomas J, and John J Stevens. 2004. "Spatial distribution of economic activities in North America." **In Handbook of regional and urban economics**, 2797-2843. Elsevier.
20. Jensen, Robert E. 1971. "A cluster analysis study of financial performance of selected business firms." **The Accounting Review** 46 (1):36-56.
21. Krishnan, Jayanthi, and Eric Press. 2003. "The north american industry classification system and its implications for accounting research." **Contemporary Accounting Research** 20 (4):685-717.
22. Kothari, Sagar P, Andrew J Leone, and Charles E Wasley. 2005. "Performance matched discretionary accrual measures." **Journal of Accounting and Economics** 39 (1):163-197.
23. R.C. Cavalcante, R.C. Brasileiro, V.L.F. Souza, J.P. Nobrega, A.L.I. Oliveira. (2016), Computational intelligence and financial markets: A survey and future directions, **Expert Systems with Applications**, 55 , pp 194-211.
24. Ramnath, Sundaresh. 2002. "Investor and analyst reactions to earnings announcements of related firms: An empirical analysis." **Journal of Accounting Research** 40 (5):1351-1376.
25. Shumway, Tyler. 2001. "Forecasting bankruptcy more accurately: A simple hazard model." **The journal of business** 74 (1):101-124.
26. Ting-Hsuan Chen, Ruey-Jenn Ho, Yi-Wei Liu .(2018). Investigating the Predictive Power of Investor Personality in Forecasting Investment Performance using machine learning models, **Computers in Human Behavior**, doi: 10.1016/j.chb.2018.09.027
27. van Liebergen, Bart. 2017. Machine learning: A revolution in risk management and compliance? **Journal of Financial Transformation** 45: 60-67.
28. Vlad, M., Tulvinschi, M., & Chiriță, I. (2011). The consequences of fraudulent financial reporting. **The USV Annals of Economics and Public Administration**, 11(1), 264-268.
29. Wu, Xindong, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, and S Yu Philip. 2008. "Top 10 algorithms in data mining." **Knowledge and information systems** 14 (1):1-37.
30. Young, Steven, and Yachang Zeng. 2015. "Accounting comparability and the accuracy of peerbased valuation models." **The Accounting Review** 90 (6):2571-2601.
31. Y.-P. Huang and M.-F. Yen. (2019). A new perspective of performance comparison among machine learning algorithms for financial distress prediction, **Applied Soft Computing Journal**, doi: https://doi.org/10.1016/j.asoc.2019.105663.

Accounting & Auditing Researches, Winter 2023, No. 56. pp. 171-190

Using Machine Learning to Provide a Model for Predicting Bankruptcy

Afshin Hashemi Gelsefidi[1]

Department of Accounting, Kish International Branch, Islamic Azad University, Kish Island, Iran

Zahra Lashgari (PhD)[2]©

Department of Accounting, Central Tehran Branch, Islamic Azad University, Tehran, Iran

Zohreh Hajiha (PhD)[3]

Department of Accounting, EastTehran Branch, Islamic Azad University, Tehran, Iran

(Received: 16 December 2020; Accepted: 1 May 2021)

The use of traditional forecasting tools and methods has a high error and has a poorer performance compared to newer methods and nonlinear models. One of the most widely used methods and algorithms in predicting the use of machine learning. The main purpose of this study is to investigate the application of machine learning in providing a model for predicting the bankruptcy of 308 companies listed on the Tehran Stock Exchange in the period 1389 to 1398 (3080 years - company) to test the hypotheses of multiple regression of composite data. In order to implement the Medians-K clustering algorithm and related calculations, R statistical calculation software was used. The results show that among the financial ratios identified in the first model, only the ratio of net income to total assets and the ratio of market value of equity to total market value can improve the ability of the Altman bankruptcy prediction model. Also, in the second model, the specified financial ratios have the ability to improve the bankruptcy forecast model, and by adding the Devscore variable for groups based on industry and size, the modified model improves the bankruptcy forecast, The results shows that a company is more likely to go bankrupt if it has bankruptcy-related financial ratios that are lower than the average of its cluster peers..

Keywords: Machine Learning, Clustering Algorithm, Bankruptcy Prediction.

[1] afshinhashemi54@yahoo.com
[2] z_lashgari@iauctb.ac.ir © (Corresponding Author)
[3] drzhajiha@gmail.com