فصلنامه مهندسی مالی و مدیریت اوراق بهادار
دوره پانزده، شماره پنجاه و هشت، بهار ۱۴۰۳
نوع مقاله : علمی پژوهشی
صفحات : ۴۰-۲۱

پیش‌بینی ورشکستگی شرکت‌ها با استفاده از ترکیب مدل‌های داده‌کاوی مبتنی بر جریمه دسته‌بندی نادرست

عطیه ترکمن[۱]
امیرعباس نجفی[۲]

تاریخ دریافت مقاله : ۱۴۰۱/۱۲/۲۱ تاریخ پذیرش مقاله : ۱۴۰۲/۰۴/۳۱

چکیده

یکی از ابزارهای قدرتمند در مسائل پیش‌بینی ورشکستگی که در دهه‌های اخیر مورد توجه بسیاری از سرمایه‌گذاران، مدیران و محققان قرارگرفته است؛ داده‌کاوی و به طور خاص ماشین بردار پشتیبان است. اما مطالعات نشان می‌دهد این روش نسبت به انتخاب پارامترها و متغیرهای ورودی از حساسیت بالایی برخوردار است. لذا هدف از تحقیق حاضر ترکیب مدل توسعه یافته ماشین بردار پشتیبان و k-نزدیک‌ترین همسایه جهت حذف ورودی‌های دارای خطا و متعاقبا افزایش دقت پیش‌بینی ورشکستگی است. به این منظور ابتدا با استفاده از ۵ نسبت مالی شامل نسبت جاری، حاشیه سود خالص، نسبت بدهی، بازده دارایی‌ها و بازده سرمایه مرتبط به ۱۵۰ شرکت حاضر در بورس اوراق بهادار تهران در بازه ۱۰ ساله ۱۳۸۹-۱۳۹۸و الگوریتم k-نزدیکترین همسایگی داده‌های آموزش پالایش شده و سپس با تکیه بر ماشین بردار پشتیبان مبتنی بر جریمه دسته‌بندی، جهت ساخت مدل پیش‌بینی به کار گرفته می‌شوند. پس از برآورد پارامترهای بهینه، اعتبارسنجی مدل با استفاده از داده‌های آزمایش صورت خواهد گرفت. در نهایت نتایج بدست آمده از مدل پیشنهادی و مدل‌های کلاسیک مورد مقایسه قرار خواهدگرفت. نتایج نشان می‌دهد با ترکیب مدل‌های k-نزدیکترین همسایه و ماشین بردار پشتیبان خطای کلی پیش‌بینی کاهش یافته و ضرایب جریمه ماشین بردار پشتیبان با سطح احتمال بالایی معنادار هستند.

کلمات کلیدی

پیش‌بینی ورشکستگی، داده‌کاوی، ماشین بردار پشتیبان، k-نزدیکترین همسایه، ضرایب جریمه

۱-دانشجوی کارشناسی ارشد گروه مهندسی مالی، دانشکده مهندسی صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.
atiye.torkaman@gmail.com
۲-استاد، گروه مهندسی مالی، دانشکده مهندسی صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران. (نویسنده مسئول)
aanajafi@kntu.ac.ir

مقدمه

ورشکستگی مفاهیم و معانی متعددی دارد. طبق تعریف آلتمن ورشکستگی زمانی رخ می‌دهد که شرکت قادر به پرداخت بدهی‌های خود نیست بنابراین از ادامه فعالیت‌های تجاری باز می‌ماند. ویتاکر (۱۹۹۹) بحران مالی را وضعیتی در نظر می‌گیرد که در آن جریان‌های نقدی ورودی شرکت از مجموع هزینه‌های بهره مربوط به بدهی‌های بلندمدت کمتر است. از نقطه نظر اقتصادی، بحران مالی را می‌توان به زیان‌ده بودن شرکت تعبیر کرد که در این حالت شرکت دچار زیان‌های سنگین و متوالی شده است. در واقع، در این حالت نرخ بازده داخلی شرکت کمتر از نرخ هزینه سرمایه است. حالت دیگری از بحران مالی زمانی رخ می‌دهد که شرکت موفق به رعایت یک یا تعداد بیشتری از بندهای مربوط به قراردادهای تسهیلات مالی دریافتی خود نمی‌شود که بـه این حالت نکول تکنیکی گفته می‌شود.

اگرچه ورشکستگی همواره بعنوان یک پدیده نامطلوب مالی مهم بوده است. از اواسط قرن بیستم، همزمان با رشد سریع تکنولوژی، تغییرات محیطی و افزایش رقابت، احتمال ورشکستگی شرکت‌ها نیز افزایش یافت. از سوی دیگر با ظهور شرکت‌های سهامی و افزایش تقاضا برای تامین مالی از منابع خارجی، نیاز به ارزیابی شرکت‌ها و کسب اطمینان از وضعیت مالی آن‌ها توسط سرمایه‌گذاران و وام دهندگان بیشتر احساس شد. بنابراین در دهه‌های اخیر پیش‌بینی ورشکستگی و توسعه مدل‌های آن بعنوان موضوعی مهم و به‌طور گسترده مورد توجه جامعه دانشگاهی و سرمایه‌گذاران قرار گرفته است.

مطالعات اولیه در حوزه پیش‌بینی ورشکستگی با استفاده از تکنیک‌های آماری مانند تحلیل ممیزی چندگانه[۱] (آلتمن، ۱۹۶۸)، لاجیت[۲] (اولسن، ۱۹۸۰) و پروبیت[۳] (زمیجسکی، ۱۹۸۴) انجام گرفت. سال‌ها بعد تعداد زیادی از مطالعات نشان دادند که روش‌های هوش مصنوعی و به طور خاص شبکه عصبی مصنوعی[۴]، می‌توانند در حل مسائل طبقه‌بندی جایگزین مناسبی برای روش‌های آماری و سنتی باشند. با این‌حال این تحقیقات محدودیت‌هایی را برای استفاده از شبکه عصبی از جمله انتخاب مدل مناسب، همگرایی مدل به بهینه محلی و عدم تعمیم‌دهی کارآمد گزارش کردند. به مرور ماشین بردار پشتیبان[۵] که یکی دیگر از مدل‌های قدرتمند داده‌کاوی[۶] است و عملکرد قابل قبولی را در حل مسائل طبقه‌بندی از جمله تشخیص چهره و پیش‌بینی ورشکستگی ارائه داده بود، جایگزین روش‌های سابق از جمله شبکه عصبی شد. SVM روشی است که با پیدا کردن ابرصفحه‌ای[۷] میان نمونه‌ها سعی در حداکثر کردن فاصله میان ابرصفحه و نمونه‌ها و درنهایت نمونه‌ها از یکدیگر دارد. درواقع روش کار SVM معادل حل کردن مساله‌ای کوادراتیک است. این روش از قوانین حداقل‌سازی ریسک ساختاری[۸] بهره می‌گیرد و افزایش قدرت تعمیم‌دهی را به همراه دارد. به همین دلیل جواب حاصل از آن برای مسائل محدب، یکتا و کلی

می‌باشد (برگس، ۱۹۹۸). از آن‌جاییکه روش ماشین بردار پشتیبان به کیفیت داده‌ها و نمونه‌های ورودی حساسیت نشان داده و نمونه‌های دارای خطا اثر ملموسی بر دقت مدل می‌گذارند، به ذهن می‌رسد می‌توان با استفاده از روشی دیگر مانند روش ساده و در عین حال قدرتمند k-نزدیکترین همسایه[۹] و ترکیب آن‌ها با یکدیگر به کارایی بیشتری دست یافت. KNN نیز مانند SVM الگوریتمی تحت نظارت است که هدف از آن دسته‌بندی یک عضو جدید براساس ویژگی نمونه‌های آموزشی می‌باشد. بدین صورت که نمونه جدید بر اساس اکثریت K نمونه که نزدیکترین همسایگی را با آن داشته باشند، تقسیم‌بندی می‌شود. بنابراین در تحقیق حاضر سعی بر این بوده که علاوه بر توسعه رویکرد ماشین بردار پشتیبان در تشخیص الگوی پنهان داده‌ها جهت پیش‌بینی ورشکستگی شرکت‌ها، مدل توسعه یافته با استفاده از مدل k-نزدیکترین همسایه بهبود داده شود. بخش‌های پژوهش بدین صورت است که ابتدا پیشینه تحقیق مورد بررسی قرار می‌گیرد. سپس مبانی نظری و معرفی مدل‌های طبقه‌بندی مورد استفاده و الگوریتم پیشنهادی بیان می‌شود. در بخش بعد نتایج تحقیق و مقایسه مدل پیشنهادی و مدل‌های کلاسیک ارائه شده و در نهایت، بخش پایانی و نتیجه‌گیری ذکر خواهد شد.

پیشینه تحقیق

پیشینه نظری

در ۵۰ سال اخیر موضوع " پیش‌بینی ورشکستگی شرکت‌ها" به یکی از موضوعات عمده پژوهش در ادبیات مالی تبدیل شده است. به‌طور کلی، مدل‌های پیشنهاد شده برای پیش‌بینی ورشکستگی را می‌توان به دو دسته رویکردهای آماری و روش‌های هوش مصنوعی طبقه‌بندی کرد. روش‌های آماری از اولین روش‌های پرکاربرد در مسائل طبقه‌بندی بودند که غالبا فرضیات محدودکننده‌ای از جمله خطی و نرمال بودن و استقلال متغیرهای ورودی را به مدل تحمیل می‌کردند. به‌همین دلیل اثربخشی روش را کاهش می‌دادند (دیکن، ۱۹۷۶). روش‌های آماری که برای پیش‌بینی ورشکستگی استفاده می‌شوند، تجزیه و تحلیل تک‌متغیره ساده (بیور ، ۱۹۶۶)، تجزیه‌وتحلیل چندمتغیره (آلتمن ، ۱۹۶۸)، رگرسیون لجستیک (السون ،۱۹۸۰) و تحلیل عاملی (وِست ، ۱۹۸۵) را شامل می‌شوند. اما پس از سال‌ها مطالعات بعدی نشان داد که روش‌های هوش مصنوعی نسبت به این فرضیات آسیب پذیری کمتری دارند. به عبارت دیگر روش‌های هوش مصنوعی بر استخراج اطلاعات از نمونه‌های آموزشی تکیه می‌کنند، درحالیکه روش‌های آماری بر بهینه‌سازی احتمال طبقه‌بندی صحیح تمرکز دارند (لیانگ و همکاران، ۱۹۹۰). شبکه عصبی، درخت تصمیم[۱۰]، ماشین بردار پشتیبان و k-نزدیکترین همسایه از مهمترین روش‌های یادگیری ماشین بودند که در طبقه‌بندی جایگزین مدل‌های آماری شدند. و در مسائل مالی از جمله پیش‌بینی ورشکستگی

به صورت گسترده مورد مطالعه و مقایسه با روش‌های آماری قرار گرفتند (کارالامبوس و همکاران، ۲۰۰۰؛ آتیا، ۲۰۰۱؛ شی و همکاران، ۲۰۰۹؛ باربزا، ۲۰۱۷؛ مای و همکاران، ۲۰۱۸، پتاک، ۲۰۲۱). بنابراین همزمان با نوآوری ایجاد شده در مدل‌های پیش‌بینی ورشکستگی و از زمانی که وپینگ ماشین بردار پشتیبان را ارائه داد، این روش مورد توجه و پایه بسیاری از تحقیقات قرار گرفت. اساس کار ماشین بردار پشتیبان نظریه یادگیری آماری[11] و حداقل کردن ریسک ساختاری است، و برخلاف سایر روش‌های هوش مصنوعی سعی در حداقل کردن خطای تعمیم‌پذیری دارد. طی تحقیقات انجام شده برخی مزایای SVM نسبت به سایر روش‌های یادگیری ماشین بدین ترتیب ذکر شده است: (الف) با مجموعه داده‌های بزرگ سازگار است. (ب) برخلاف روش‌های سنتی یادگیری ماشین، SVM به دنبال حداقل کردن ریسک ساختاری است. (ج) مبنای این روش مساله بهینه سازی کوادراتیک و محدبی است که بر خلاف سایر روش‌ها به بهینه کلی دست می‌یابد. (د) SVM برای حل مسائل غیرخطی از توابع کرنل استفاده می‌کند، از این رو تعمیم‌پذیری و حل داده‌ها با ابعاد بزرگ مانند عکس و متن را تضمین می‌نماید. (ه) مدل یادگیرنده آن تنها با بردارهای پشتیبان ساخته می‌شود نه تمام نمونه‌ها. به همین دلیل مدلی ساده، سازگار و در عین حال قوی می‌باشد (تیان و کی، ۲۰۱۴؛ تومار و آگاروال، ۲۰۱۵).

پیشینه تجربی

همانگونه که پیش‌تر ذکر شد، تاکنون مطالعات بسیاری در خصوص پیش‌بینی درماندگی مالی صورت گرفته است. اگرچه شاید بتوان ادعا کرد قدیمی‌ترین پژوهش در این حوزه مربوط به استفاده از نسبت جاری جهت ارزیابی وضعیت اعتباری در سال ۱۸۷۰ بوده است. اما تحقیق بیور (۱۹۶۶) که همراه با یک نمونه از ۱۵۸ شرکت انجام شد و نشان داد نسبت‌های مالی می‌توانند بهترین معیار سلامت مالی شرکت باشند، پایه‌گذار مطالعه در مسائل پیش‌بینی ورشکستگی محسوب می‌شود. درواقع تحقیق او نشان می‌داد نسبت‌های مالی در شرکت‌های سالم و ورشکسته متفاوت است. در این تحقیق، بیور ۳۰ نسبت مالی که تصور می‌کرد بهترین معیار برای سنجش سلامت مالی یک شرکت هستند را انتخاب کرد. سپس این نسبت‌ها را به شش گروه تقسیم کرد. این شش گروه عبارت بودند از : نسبت‌های مربوط به جریان نقدی، بدهی به کل دارایی‌ها، دارایی‌های نقدشونده به کل دارایی‌ها، دارایی‌های نقدشونده به بدهی‌های جاری، فعالیت (گردش) و سود خالص. آلتمن (۱۹۶۸) با استفاده از ۵ نسبت مالی در حالی‌که نمونه‌های ورودی را بر اساس اندازه و صنعت به چند دسته تقسیم کرده بود، مدلی را بر مبنای MDA ارائه داد که به مدل z-score معروف است. و ادعا کرد این مدل می‌تواند در تخصیص وام‌های بانکی و ارزیابی مشخصات سرمایه‌گذاری استفاده شود. مدل او می‌توانست یک سال قبل از ورشکستگی این رویداد را با دقت ۹۶

درصد پیش‌بینی کند. اهلسون (۱۹۸۰) نیز مدلی را با استفاده از تکنیک لاجیت توسعه داد. وی در این پژوهش از ۱۰۵ شرکت ورشکسته و ۲۰۵ شرکت غیر ورشکسته بین سال‌های ۱۹۷۰ تا ۱۹۷۶ استفاده کرد و نه نسبت مالی را بعنوان متغیر مستقل بکار برد؛ از این نه نسبت مالی، پنج تای آن‌ها در مطالعات قبلی نیز استفاده شده بودند (چن، ۲۰۰۹) .شاه و مرتزا (۲۰۰۰) مدلی را با استفاده از شبکه عصبی مصنوعی برای پیش بینی ورشکستگی ارائه دادند. در این مطالعه از اطلاعات ۶۰ شرکت ورشکسته و ۵۴ شرکت غیر ورشکسته بین سال‌های ۱۹۹۲ تا ۱۹۹۴ استفاده شده و دقت پیش بینی مدل مورد استفاده ۷۳ درصد ثبت شده است. سال‌ها بعد نیز به‌دلیل عملکرد قابل قبول شبکه عصبی بسیاری از مطالعات و تحقیقات بر اساس این روش انجام می‌گرفت. اما در سال ۲۰۰۳ زمانی که ماشین بردار پشتیبان برای نخستین بار توسط هاردل و همکارانش در پیش‌بینی ورشکستگی مورد استفاده قرار گرفت و با روش‌های سابق از جمله ANN و MDA مقایسه شد، با کسب ۷۰درصد دقت و برتری نسبت به سایر روش‌ها توجه محققان را جلب کرده و بعنوان روشی پیشرو در حوزه‌های مختلف از جمله تشخیص چهره، شناسایی دستخط و متن‌کاوی بطور گسترده مورد مطالعه قرار گرفت؛ آن‌ها مجددا ثابت کردند SVM در مقایسه با مدل‌های آماری نیز کارایی بهتری دارد (ون و همکاران، ۲۰۰۳) . چن در سال ۲۰۱۱ طی تحقیقی تجربی شامل۳۳ نسبت مالی و ۸ نسبت غیرمالی و یکی از شاخص‌های کلان اقتصادی نشان داد نسبت‌های مالی در تشخیص شرکت‌های ورشکسته از کارایی بالاتری برخوردارند. پس از آن کیم و کانگ با انتخاب ۱۲۰۰ شرکت و به‌کارگیری ۷ نسبت مالی (درآمد به کل دارایی، سود ناخالص به هزینه بهره، سود انباشته به کل دارایی، بازده نقدی، کل بدهی به کل دارایی، موجودی کالا به فروش و موجودی کالا به کل دارایی) نشان دادند ماشین بردار پشتیبان با کسب دقت ۷۲/۴۵ درصد در مقایسه با سایر الگوریتم‌ها از جمله الگوریتم ژنتیک قوی‌تر ظاهر شده و خطای پیش‌بینی را به میزان قابل توجهی کاهش می‌دهد (کیم و کانگ، ۲۰۱۲). ژو (۲۰۱۳) در تحقیق خود با بکارگیری ۷ روش نمونه‌برداری و ۵ مدل پیش‌بینی و با استفاده از دو دیتابیس واقعی و نامتوازن، تاثیر روش‌های نمونه‌برداری بر عملکرد مدل‌های استفاده شده را بررسی کرد. او ثابت کرد علاوه بر اینکه تاثیر روش‌های نمونه‌برداری به تعداد شرکت‌های ورشکسته وابسته است، به عنوان مثال برای مسائل با تعداد نمونه کم روش بیش‌نمونه‌برداری مصنوعی کلاس اقلیت بهینه است. از میان ۵ روش متفاوت پیش‌بینی، ماشین بردار پشتیبان بهترین کارایی را دارد. در تحقیق دیگری نیز اثرعدم توازن داده‌های موردمطالعه بررسی‌شد. طبق یافته‌های تحقیق، اگر نمونه‌های ورشکسته کمتر از ۲۰ درصد داده‌های آموزش را تشکیل دهند، توانایی پیش‌بینی به میزان قابل توجهی کاهش می‌یابد. بنابراین برای حل این مشکل روشی ترکیبی مبتنی بر کم‌نمونه‌گیری[۱۲] و بیش‌نمونه‌گیری[۱۳]

پیشنهاد گردید. نتایج نشان می‌داد SVM تنها روشی است که کمترین میزان حساسیت را نسبت به داده‌های نامتوزان داراست (سورین، ۲۰۱۸). کیو نیز در مقایسه‌ای میان ماشین بردار پشتیبان و شبکه عصبی بیان کرد که استفاده از فضای ویژگی، SVM را قادر می‌سازد تا با استفاده از نمونه‌های آموزشی‌های کمتر و به طور کاراتری به جواب بهینه دست یابد (کیو و همکاران، ۲۰۱۹). در برخی مطالعات دیگر نیز ماشین بردار پشتیبان با سایر روش‌ها مقایسه شده و نتایج تمامی آن‌ها حاکی از برتری SVM نسبت به سایر روش‌ها بوده است (مین و همکارانش، ۲۰۰۶؛ دینگ و همکاران، ۲۰۰۸؛ چادهوری و همکاران، ۲۰۱۱؛ وانگ و ما ، ۲۰۱۲؛ لی و همکاران، ۲۰۱۸؛ آلاکا و همکاران، ۲۰۱۸).

در ایران نیز از سال‌ها قبل پیش‌بینی ورشکستگی با استفاده از روش‌های گوناگون از جمله ماشین بردار پشتیبان مورد بررسی قرار گرفته است. راعی و فلاح‌پور در سال ۱۳۸۷ از این روش به همراه یکی از پرکاربردترین مدل‌های آماری، یعنی رگرسیون لجستیک[۱۴]، جهت پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران بهره گرفتند. یافته‌های تحقیق حاکی از آن بود که نه تنها SVM نسبت به رگرسیون خطی[۱۵] از دقت کلی بهتری برخوردار است، بلکه توانایی بالاتری نیز در تعمیم‌پذیری دارد. قدرتی و معنوی مقدم نیز به مقایسه مجموعه روش‌های آماری و دو روش هوشمند ژنتیک فرج زاده و ژنتیک مک کی در بورس اوراق بهادار تهران طی سال‌های ۱۳۸۲ تا ۱۳۸۹ پرداختند. در تحقیق آن‌ها برای تفکیک شرکت‌های سالم و دارای بحران مالی از ماده ۱۴۱ قانون تجارت بهره گرفته شده بود. نتایج نشان می‌داد الگوی پیش‌بینی بحران مالی زمیسکی[۱۶]، اسپرینگیت[۱۷]، سی ای اسکور، ژنتیک فرج زاده و ژنتیک مک‌کی[۱۸] توانایی پیش‌بینی تداوم فعالیت شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران را دارند. همچنین مدل‌هایی که بر پایه الگوریتم ژنتیک طراحی شده‌اند نسبت به روش آماری دارای دقت بیشتری می‌باشند. در این تحقیق دقت کلی ۹۱,۷ درصد برای الگوریتم ژنتیک فرج زاده و ۹۱,۶۵ درصدی برای الگوریتم ژنتیک مک‌کی گزارش شده است. مدل‌های آماری زمیسکی، اسپرینگیت و سی ای اسکور نیز همگی دارای توان پیش‌بینی ۹۰ درصد بودند (غضنفری و همکاران، ۱۳۹۷). مرادی و همکارانش (۱۳۹۱) دو روش ماشین بردار پشتیبان و تحلیل ممیزی چندگانه را در شرکت‌های بورس اوراق بهادار تهران و با در نظر گرفتن ۸۸ متغیر ورودی مورد مقایسه قرار دادند. و نتایج نشان از عدم برتری آماری مدل‌ها نسبت به یکدیگر داشت. اصغری و اصفهانی‌پور در مطالعه‌ای که اخیرا صورت گرفته است، مدل ترکیبی الگوریتم بهینه‌سازی ازدحام ذرات و ماشین بردار پشتیبان جهت پیش‌بینی ورشکستگی شرکت‌ها را ارائه داده‌اند. هدف اصلی آنها ارائه یک مدل پیش‌بینی کننده با عملکرد بالا و مقایسه نتایج آن با سایر مدل‌های رایج بوده، و به همین منظور از الگوریتم بهینه‌سازی ذرات به

عنوان روش انتخاب ویژگی در ترکیب با ماشین بردار پشتیبان به عنوان الگوریتم طبقه‌بندی کننده استفاده کرده‌اند. نتایج نشان می‌دهد مدل پیشنهادی پژوهش در پیش‌بینی ورشکستگی نسبت به روش پرسپترون چند لایه از عملکرد بهتری برخوردار است (اصغری و اصفهانی‌پور، ۱۳۹۸). همانطور که مشاهده می‌شود، تمامی مطالعات داخلی مبتنی بر استفاده از داده‌کاوی در پیش‌بینی ورشکستگی بر داده‌های حاصل از بورس اوراق بهادار متمرکز بوده‌اند. بررسی مطالعات خارجی در این حوزه نیز نشان‌دهنده استفاده از نسبت‌های مالی شرکت‌ها و داده‌های بورس و اوراق بهادار کشورهای مختلف به عنوان ورودی مدل‌های پیش‌بینی می‌باشد. که دلیل آن را می‌توان عدم امکان دسترسی به پایگاه داده‌های غنیتر دانست. لذا باتوجه به عدم تغییر داده‌های ورودی جهت پیش‌بینی ورشکستگی، سوالاتی مطرح می‌گردد از جمله اینکه آیا می‌توان با توسعه روش کارایی مثل ماشین بردار پشتیبان و لحاظ جریمه برای نتایج اشتباه به دقت بیشتری در امر پیش‌بینی دست یافت؟ آیا استفاده از یک روش داده‌کاوی دیگر در نقش پالایش‌کننده ویژگی‌ها به عنوان متغیرهای ورودی اثری در دقت مدل دارد؟

روش شناسی پژوهش

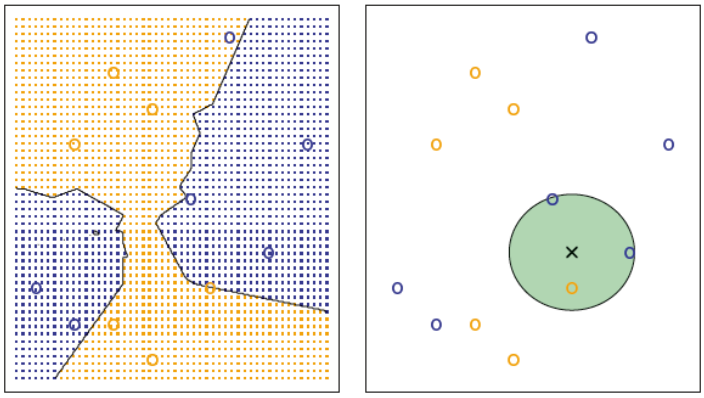
ماشین بردار پشتیبان که در حالت کلی برای دسته‌بندی داده‌ها استفاده می‌شود نسبت به داده‌های آموزشی حساسیت قابل توجهی دارد. از این رو در این تحقیق به منظور حذف اثر داده‌های دارای خطا و افزایش دقت پیش‌بینی از مدلی پیشنهادی حاصل از ترکیب روش نزدیکترین همسایگی و ماشین بردار پشتیبان استفاده شده است. بدین صورت که ابتدا بوسیله‌ی داده‌های آموزشی و با بهره‌گیری از روش نزدیکترین همسایگی مدلی جهت پیش‌بینی ورشکستگی شرکت‌ها ساخته شده و نتیجه با داده‌های واقعی مقایسه می‌گردد. داده‌هایی که توسط مدل به اشتباه برچسب‌گذاری شده‌اند حذف شده و سایر داده‌ها به‌عنوان ورودی مدل ماشین بردار پشتیبان در نظر گرفته می‌شوند. در ادامه و با در نظر گرفتن ضریب جریمه برای دسته بندی نادرست کلاس‌های مختلف، کارایی ماشین بردار پشتیبان در پیش‌بینی وضعیت شرکت‌ها با استفاده از داده‌های آزمایش مورد سنجش قرار گرفته و با مدل‌های کلاسیک مقایسه می‌گردد.

الگوریتم‌های هر مرحله

روش k-نزدیکترین همسایه

روش k-نزدیکترین همسایه، به عنوان یک رویکرد طبقه‌بندی ناپارامتریک، علیرغم سادگی توانسته است نتایج قابل قبولی را در حل مسائل دسته‌بندی نشان دهد. در این روش اگر نقطه‌ی فرضی x_0 و n نقطه آموزشی وجود داشته باشد، الگوریتم تمامی x_n هایی که در نزدیک‌ترین فاصله به x_0 باشند را پیدا

کرده و براساس رای‌گیری اکثریت درمیان k همسایه، x_0 را دسته‌بندی می‌کند. درواقع در این روش به هر نمونه دسته‌بندی نشده کلاسی اختصاص داده می شود که مشتمل بر k عدد از نزدیکترین همسایه‌های آن باشد. به همین دلیل است که این روش داده‌کاوی، k-نزدیکترین همسایه نامیده می‌شود. شکل ۱ روند تصمیم‌گیری روش مذکور با مقدار۳=k برای مجموعه‌ای از نمونه‌ها که در دو دسته تقسیم‌بندی شده‌اند را نشان می‌دهد.



شکل ۱: روش نزدیکترین همسایگی برای k=3

مهم‌ترین فاکتور در این الگوریتم که نقش مهمی در میزان دقت و موفقیت آن دارد، تابع فاصله است. در پژوهش پیش رو معیار فاصله (D) براساس فاصله اقلیدسی در نظر گرفته شده است.

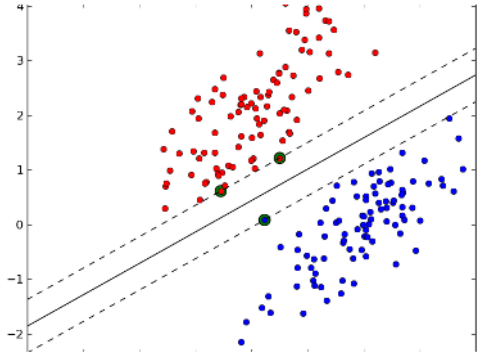
$$D = \left(\sum_{i=1}^{p}(x_i - y_i)^2\right)^{\frac{1}{2}} \qquad (۱)$$

فاکتور مهم دیگر انتخاب k یعنی تعداد نمونه‌های همسایه معیار جهت دسته‌بندی نمونه جدید است. اگرk خیلی بزرگ باشد، کلاس‌هایی که تعداد نمونه‌های آن‌ها زیاد است کلاس‌ها با تعداد نمونه‌های کم را در خود جای داده و نتایج را دچار انحراف می‌نمایند. و اگر خیلی کوچک باشد، فرآیند آموزش با مشکل مواجه خواهد شد. اما در حالت کلی مقادیر بزرگتر k نسبت به داده‌های دارای خطا ایمن‌ترند.

روش ماشین بردار پشتیبان

ماشین بردار پشتیبان یکی از روش‌های یادگیری ماشینی با نظارت است که در سال ۱۹۹۵ توسط وپنیک[۱۹] و همکارانش ارائه گردید. این روش در واقع یک طبقه بندی‌کننده دودویی[۲۰] است که سعی دارد میان داده‌های دو کلاس ابرصفحه‌ای ایجاد نماید؛ به‌طوریکه حاشیه میان داده‌های هر کلاس و ابرصفحه حداکثر گردد. داده‌هایی که در نزدیکترین فاصله نسبت به ابر صفحه قرار دارند و بردار پشتیبان

نامیده می‌شوند، برای محاسبه این فاصله بکار می‌روند (وپینک ۱۹۹۵). مهم‌ترین ویژگی ماشین بردار پشتیبان که آن را از سایر روش‌های داده‌کاوی متمایز می‌کند قدرت تعمیم‌دهی و رسیدن به نقطه بهینه کلی در مقابل نقطه بهینه محلی است. به منظور درک بهتر مطلب، در شکل ۲ تصویری از یک مجموعه داده متعلق به دو کلاس نشان داده شده که ماشین بردار پشتیبان بهترین ابرصفحه را برای جداسازی آن ها انتخاب کرده است.



شکل۲- ساختار مدل طبقه‌بندی ماشین بردار پشتیبان

مدل کلاسیک ماشین بردار پشتیبان به‌صورت زیر می‌باشد :

$$Min \quad ½ \sum_{j=1}^{m} \omega_j^2 + p \sum_{i=1}^{n} \varepsilon_i$$
$$y_i(\omega^T x + b) \geq 1 - \varepsilon_i \qquad i = 1,2,\dots,n$$
$$\varepsilon_i \geq 0 \qquad (۲)$$

n و m به‌ترتیب تعداد نمونه‌ها و ویژگی‌های مساله هستند. $\omega = (\omega_1, \omega_2, \dots, \omega_m)$ بردار نرمال و b انحراف از ابر صفحه جداکننده است. y_i نیز برچسب i امین نمونه را نشان می‌دهد. هم‌چنین p و ε_i به‌ترتیب معرف پارامتر جریمه و متغیر کمبود مدل هستند. طبق برخی تحقیقات انجام شده، در بسیاری از مجموعه داده‌ها خروجی ماشین بردار پشتیبان نسبت به سایر روش‌ها بهینه‌تر است. با این‌حال ممکن است داده‌ها و مسائلی وجود داشته باشند که ماشین بردار کلاسیک برای آن‌ها مناسب نیست. از این رو به‌منظور افزایش دقت مدل از تابع کرنل استفاده می‌شود. اضافه کردن تابع کرنل به مدل اصلی

$x \rightarrow \chi \phi$: داده‌ها را از فضای x به فضای χ که فضای ویژگی[۲۱] نامیده می‌شود منتقل می‌کند. ترکیب تابع کرنل و مدل کلاسیک، مدل غیرخطی نامیده می‌شود و به صورت زیر تعریف می‌گردد :

$$Min \quad ½ \sum_{j=1}^{m} \omega_j^2 + p \sum_{i=1}^{n} \varepsilon_i$$
$$y_i(\omega^T \phi(x) + b) \geq 1 - \varepsilon_i \qquad i = 1,2,\dots,n \qquad (۳)$$
$$\varepsilon_i \geq 0$$

تابع کرنل استفاده شده در پژوهش حاضر نیز تابع پایه شعاعی[۲۲] و به صورت زیر است :

$$K(x_i, x_j) = \exp(-\gamma \, ||x_i - x_j||^2) \qquad \gamma > 0 \qquad (۴)$$

در بسیاری از مسائل حساسیت پیش‌بینی و برچسب‌گذاری اشتباه برای کلاس‌های متفاوت یکسان نیست. به‌طور مثال دسته‌بندی شرکتی که ورشکسته است به‌عنوان شرکت سالم به مراتب تبعات بیشتر و سنگین‌تری در مقایسه با دسته‌بندی نادرست شرکتی سالم دارد. بنابراین در پژوهش حاضر سعی شده است با توسعه مدل کلاسیک ماشین بردار پشتیبان، جریمه‌های متفاوتی برای دسته‌بندی غلط مدل در نظر گرفته شود. مدل پیشنهادی بدین صورت می‌باشد :

$$Min \quad ½ \sum_{j=1}^{m} \omega_j^2 + p_1 \sum_{i \in class\ 1}^{n} \varepsilon_i + p_2 \sum_{i \in class\ 2}^{n} \varepsilon_i$$
$$y_i(\omega^T \phi(x) + b) \geq 1 - \varepsilon_i \qquad i = 1,2,\dots,n \qquad (۵)$$
$$\varepsilon_i \geq 0$$

p_1و p_2 به‌ترتیب جریمه دسته‌بندی غلط کلاس اول (شرکت‌های سالم) و جریمه دسته‌بندی غلط کلاس دوم (شرکت‌های ورشکسته) می‌باشند. سایر پارامترهای تاثیرگذار بر عملکرد ماشین بردار پشتیبان، پارامتر تعیین کننده بردارهای پشتیبان و پارامتر تابع کرنل می‌باشند که مشابه با پارامترهای مدل کلاسیک بوده و با نمادهای C و γ نمایش داده می‌شوند.

داده‌ها

معیار تعیین وضعیت ورشکستگی شرکت‌ها

معیار مورد استفاده به منظور تعیین وضعیت ورشکستگی یک شرکت مشخص، ماده ۱۴۱ قانون تجارت می‌باشد. بر اساس این ماده قانون تجارت ایران شرکتی به عنوان ورشکسته در نظرگرفته می‌شود که بر اساس زیان‌های وارده، حداقل نیمی از سرمایه خود را از دست داده باشد. در شرایط فوق طبق

قانون ذکر شده هیئت مدیره مکلف است بلافاصله طی مجمع عمومی فوق‌العاده، صاحبان سهام کسب و کار را دعوت کرده تا موضوع انحلال یا بقای شرکت به رأی گذاشته شود. در صورتیکه مجموع شرکت به انحلال رأی مثبت ندهد در همان جلسه و با رعایت مقررات ماده ۶ این قانون، درمورد افزایش و یا کاهش سرمایه تصمیم گرفته می‌شود (ناصرزاده، ۱۳۷۴، ۱۹).

پیش‌پردازش

با توجه به برابر نبودن تعداد شرکت‌های سالم و ورشکسته (به طور معمول تعداد شرکت‌های ورشکسته به صورت معناداری از تعداد شرکت های سالم کمتر می‌باشد)، نیاز است به تعداد شرکت‌های ورشکسته از میان شرکت‌های سالم، نمونه‌هایی انتخاب شوند. در صورت ورود داده‌ها بدون بهره‌گیری از فرآیند بالانس داده‌ها، سیستم طراحی شده گروهی را با دقت بالا و گروه مقابل را به دلیل پایین بودن تعداد نمونه‌های ورودی با دقت پایین‌تری مورد تشخیص قرار خواهد داد. از طرفی الگوریتم‌هایی مانند شبکه عصبی، ماشین بردار پشتیبان و بسیاری از الگوریتم‌های طبقه‌بندی جهت عملکرد مطلوب به داده‌های ورودی نرمال شده نیاز دارند؛ بنابراین تمامی داده‌های ورودی همگن و نرمال شده و سپس مورد استفاده قرار گرفتند.

جامعه مورد بررسی و متغیرهای ورودی

به منظور حذف اثر همه‌گیری ویروس کرونا بر کسب و کارها، جامعه مورد مطالعه در این پژوهش، شامل شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران[۲۳] است که تا پیش از سال ۱۳۹۹ صورت مالی خود را به بورس ارائه داده‌اند، به‌طوریکه دارای شرایط زیر باشند :

۱. صورت مالی شرکت در دسترس باشد.

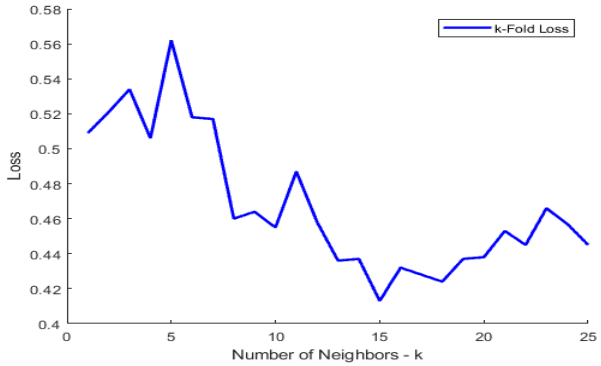
۲. سال مالی شرکت منتهی به پایان اسفند ماه هر سال باشد.

از میان شرکت‌های دارای شرایط، ۷۵ شرکت ورشکسته و ۷۵ شرکت سالم به‌صورت تصادفی انتخاب شدند. همچنین نمونه‌های موجود به دودسته داده‌های آموزشی برای برآورد پارامترهای بهینه و ساخت مدل و آزمایشی به منظور اعتبار سنجی تقسیم می‌شوند. متغیرهای پژوهش نیز در دو دسته‌ی مستقل و وابسته تعریف می‌گردند. پس از مطالعه و بررسی پژوهش‌های مشابه که در محیط اقتصادی ایران انجام شده‌اند، ۵ نسبت مالی شامل نسبت جاری، حاشیه سود خالص، نسبت بدهی، بازده دارایی‌ها و بازه سرمایه به‌عنوان متغیرهای مستقل و پیش‌بینی‌کننده شناسایی و انتخاب شدند. متغیر وابسته نیز شرکت‌های سالم و ورشکسته هستند که به‌صورت (۱-) برای شرکت‌های ورشکسته و (۱) برای شرکت‌های سالم

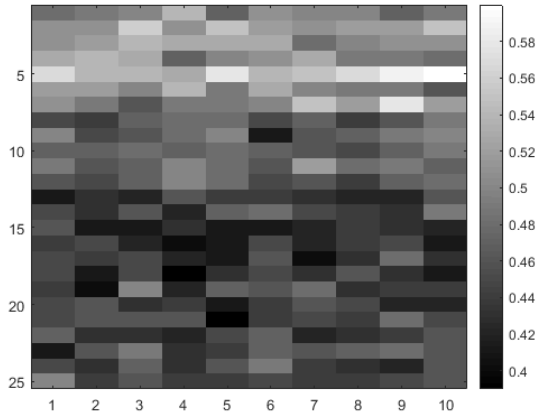
برچسب‌گذاری شده‌اند. تمامی اطلاعات مالی شرکت‌ها از صورت‌های مالی آن‌ها و با استفاده از نرم افزار ره‌آورد نوین گردآوری شده‌اند. ابزار مورد استفاده جهت پالایش داده‌ها و پیش‌بینی نیز نرم‌افزار متلب می‌باشد.

نتایج

همانطور که پیش‌تر ذکر شد، مقدار پارامتر k برای نزدیکترین همسایگی و پارامترهای C و γ برای SVM در دقت و عملکرد این روش‌ها تاثیر بسزایی دارند؛ از این رو الگوریتم‌های به کار رفته با مقادیر مختلف پارامترها و به دفعات مورد تست قرار گرفته و مدل با کمترین مقدار خطای اعتبار سنجی متقاطع[۲۴] به عنوان مدل بهینه انتخاب شده است، نتایج بدین صورت است :



شکل ۳- میانگین خطای مدل KNN به ازای مقادیر مختلف k



شکل۴- نمای شماتیک میزان خطای مدل KNN به ازای k ها و تکرارهای متفاوت

طبق شکل‌های ۳ و ۴ بالا با اجرای مدل KNN و در نظر گرفتن مقادیر ۲ تا ۲۵ برای پارامتر k در ۱۰ تکرار، بهترین مدل دارای k=۱۵ می‌باشد. جدول ۱ نیز نشان‌دهنده میزان خطا به ازای ۶ مقدار مختلف γ و ۹ مقدار متفاوت C است که با استفاده از روش خطای اعتبار سنجی متقاطع (k-fold =۳۰) بدست آمده است.

جدول۱- خطای اعتبار سنجی به ازای مقادیر مختلف C و γ

	C =۰,۰۰۱	C =۰,۰۱	C =۰,۱	C =۱	C =۱۰	C =۱۰۰	C =۱۰۰۰	C =۱۰۰۰۰	C =۱۰۰۰۰۰
۰,۱= γ	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۸۲	۰,۸۲	۰,۸۲
۱= γ	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۰۱۴۲	۰,۰۰۰۱۰۷	۰,۰۰۰۳۵۶	۰,۰۷۱۳
۱۰= γ	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۰۵۳۵	۰,۰۰۰۱۰۷	۰,۰۰۰۱۰۷
۱۰۰= γ	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۰۷	۰,۰۰۰۵۳۵
۱۰۰۰= γ	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲
۱۰۰۰۰= γ	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲	۰,۰۰۱۲

بنابراین مقادیر بهینه C و γ نیز به ترتیب ۱۰۰۰۰ و ۱ می‌باشد. هچنین طبق فرضیات این پژوهش جریمه طبقه‌بندی نادرست برای کلاس مربوط به شرکت‌های ورشکسته در بازه یک تا ۱۰ برابر جریمه کلاس مقابل در نظر گرفته شده است. واضح است که جریمه‌های مربوط به مدل با کمترین خطا به عنوان جریمه‌های بهینه انتخاب می‌گردد. جریمه‌های انتخابی بدین صورت بوده است:

جدول۲- مقادیر بهینه ضرایب جریمه دسته‌بندی نادرست

پارامتر	مقدار
p_1	۱
p_2	۹,۳۵

با توجه به نتایج جدول ۲ بنظر می‌رسد که مدل پیشنهادی نه تنها تفاوت میان جریمه‌ها را مثبت تلقی کرده بلکه سعی در حداکثر کردن اختلاف میان جریمه دو کلاس دارد. پس از تعیین مقادیر بهینه پارامترهای هر مدل و ساخت مدل پیشنهادی، نتایج بدست آمده بر اساس مدل‌های گوناگون مورد مقایسه و بررسی قرار گرفت. در جدول ۳ نتایج حاصل از پیش‌بینی با استفاده از KNN، SVM و مدل‌های ترکیبی و پیشنهادی آورده شده است:

جدول ۳- نتایج آزمون مدل‌های مورد استفاده در پیش‌بینی وضعیت شرکت‌ها

نام مدل	دقت کلی	میزان خطای جانشینی مجدد۲۵	میزان خطای اعتبار سنجی متقاطع۲۶
KNN	۶۳٪	۳۸٪	٪۴۱
SVM	۹۱٪	۱۱٪	٪۴۰
KNN-SVM	۹۴,۲٪	۱۰٪	٪۲۰
KNN-CSSVM	۹۸,۶٪	۰٪	۷٪

اعداد ذکر شده در ستون اول درصد پیش‌بینی صحیح را نشان می‌دهند. مجددا یادآوری می‌گردد تعداد شرکت‌های هر گروه ۷۵ شرکت است. بنابراین با توجه به کوچک بودن جامعه مورد بررسی دقت پایین در مدل نزدیکترین k-همسایه قابل تامل است. همانطور که جدول ۳ نشان می‌دهد، اگرچه مدل کلاسیک SVM و مدل ترکیبی KNN-SVM نیز از دقت خوبی جهت پیش‌بینی برخوردارند اما مدل پیشنهادی بالاترین دقت را در پیش بینی ورشکستگی با استفاده از نسبت‌های مالی در نظر گرفته شده داشته است. جهت تایید قدرت و اعتبار مدل بدست آمده خطای هر مدل به دوروش مجزا محاسبه شد و مطابق با انتظارات، KNN-CSSVM با خطایی نزدیک به صفر در هر دو روش کمترین میزان خطا را ثبت نمود.

برای بررسی فرضیه پژوهشی، مبنی بر عملکرد بهتر مدل پیشنهادی KNN-CSSVM در مقایسه با مدل SVM کلاسیک، از آزمون t مستقل استفاده شده است. با استفاده از این آزمون، عملکرد بهتر مدل پیشنهادی از نظر آماری بررسی می‌شود. نتایج نشان می‌دهد میانگین دقت در مدل SVM ساده ۹۲ درصد و انحراف معیار آن برابر با ۶,۴ است، در حالیکه این مقادیر برای مدل پیشنهادی به ترتیب ۹۷,۳ درصد و ۵,۳ می‌باشد. و می‌توان ادعا کرد دقت کلی مدل KNN-CSSVM، با سطح اطمینان ۹۵ درصد از مدل SVM کلاسیک بیشتر است. نتایج آزمون بدین صورت است:

جدول ۴- نتایج آزمون t در مساله مورد بررسی

الگوریتم	KNN-CSSVM	SVM
میانگین دقت کلی	۹۷,۳	۹۲
انحراف معیار	۵,۳	۶,۴
آماره t	۲,۳۶	
p-value	۰,۷	

نتیجه‌گیری

پیش‌بینی درماندگی مالی و ورشکستگی شرکت‌ها یکی از پژوهش‌های با اهمیت در حوزه مالی است. معمولاً بانک‌ها و موسسات رتبه‌بندی اعتباری از این مدل‌ها برای اتخاذ تصمیمات اعتباری و اختصاص رتبه استفاده می‌نمایند. از طرفی با پیش‌بینی وضعیت شرکت درآینده و پس از آن ریشه‌یابی و حل مسأله، می‌توان به نتایج بسیار رضایت‌بخشی در حوزه مدیریتی دست یافت. در این پژوهش، پیش‌بینی ورشکستگی شرکت‌ها با استفاده از مدل‌های KNN، SVM و مدل‌های توسعه یافته انجام گرفت. و نتایج نشان داد مدل پیشنهادی KNN-CSSVM می‌تواند در امر پیش‌بینی جانشین مناسبی برای مدل ساده SVM باشد. درواقع، یافته‌های تحقیق حاکی از این است که دقت کلی و قدرت تعمیم‌پذیری KNN-CSSVM از مدل‌های SVM و KNN-SVM به طور معناداری بیشتر است. و با اطمینان بیشتری می‌توان از آن برای پیش‌بینی یا رتبه‌بندی اعتباری شرکت‌ها در بازه زمانی یک سال قبل از ورشکستگی استفاده نمود. همچنین، نتایج حاصل از این پژوهش بار دیگر ثابت می‌کند که نسبت‌های مالی می‌توانند پیش‌بینی‌کننده مناسبی برای ورشکستگی شرکت‌ها باشند. در مجموع می‌توان ادعا نمود که با استفاده از نتایج این پژوهش بعنوان اولین گام، می توان از مبتلا شدن شرکت‌ها به درماندگی مالی و ورشکستگی و پیامدهای آن بطورمطلوبی جلوگیری کرد. البته در صورتیکه پس از پیش‌بینی به ریشه‌یابی مساله و ردیابی علل پرداخته شود. همچنین قابل ذکر است صنعت بانکداری ایران به دلیل عملکرد سال‌های اخیر، تحریم‌های بانکی و بی‌توجهی‌های مدیریتی، فاصله قابل توجهی با بانکداری جهانی دارد و این موضوع اخیرا مورد توجه مسوولان بانک مرکزی و سیستم بانکی جهان قرار گرفته است؛ از این رو، بانک‌ها و سایر موسسات مالی می‌توانند برای اتخاذ تصمیمات اعتباری دقیق‌تر و علمی‌تر از این مدل‌ها استفاده نمایند؛ بویژه بانک‌هایی که قصد دارند رویکرد رتبه‌بندی داخلی از دستورالعمل بال ۲[۲۷]را پیاده‌سازی نمایند.

منابع

۱) اصغری، زهرا و اصفهانی پور (۱۳۹۸) ارائه مدل پیش بینی ورشکستگی شرکتها با ترکیب الگوریتم بهینه سازی ازدحام ذرات و ماشین بردار پشتیبان، چهارمین کنفرانس ملی در مدیریت، حسابداری و اقتصاد با تاکید بر بازاریابی منطقه‌ای و جهانی، تهران.

۲) آلتمن، ادوارد (۱۹۶۸) نسبت‌ها مالی، آنالیز تشخیصی و پیش‌بینی ورشکستگی شرکت‌ها، فایننس، ۲۳، ۴، ۵۸۹-۶۰۹.

۳) اوهلسون، جیمز (۱۹۸۰) نسبت‌های مالی و پیش‌بینی احتمالی ورشکستگی، تحقیقات حسابداری، ۱۰۹-۱۳۱.

۴) بابوزا، کیمورا و آلتمن (۲۰۱۷) مدل‌های یادگیری ماشین و پیش‌بینی ورشکستگی، سیستم‌های خبره و کاربردها، ۸۳، ۴۱۷-۴۰۵.

۵) برگس، کریستوفر (۱۹۹۸) آموزش ماشین‌های بردار پشتیبان برای تشخیص الگو، داده‌کاوی و کشف دانش، ۲، ۹۷۴-۹۵۵.

۶) بیور، ویلیام اچ (۱۹۶۶) نسبت‌های مالی به عنوان پیش‌بینی‌کننده‌های شکست، تحقیقات حسابداری، ۷۱-۱۱۱.

۷) تومار، دیویا و اگروال (۲۰۱۵) ماشین بردار پشتیبان دوقلو : مروری از سال ۲۰۰۷ تا ۲۰۱۴، انفورماتیک مصر، ۱۶، ۱، ۶۹-۵۵.

۸) تیان، ییننگجی و کی (۲۰۱۴) بررسی ماشین بردار پشیتبان دوقلو، سالنامه علوم داده، ۲،۱، ۲۷۷-۲۵۳.

۹) چن، نینگ و ویرا (۲۰۰۹) پیش‌بینی ورشکستگی براساس تحلیل مولفه‌های مستقل، ICAART، ۱۵۰-۱۵۵.

۱۰) دیکین، ادوارد (۱۹۷۶) آنالیز تشخیصی پیش‌بینی‌کننده‌های شکست کسب و کار، تحقیق حسابداری، ۱۶۷-۱۷۹.

۱۱) غضنفری، مهدی. رحیمی کیا، عسکری (۱۳۹۷) پیش‌بینی ورشکستگی شرکت‌ها مبتنی بر سیستم‌های هوشمند ترکیبی، پژوهش‌های حسابداری مالی و حسابرسی، ۱۰، ۳۷، ۱۹۴-۱۵۹.

۱۲) کیم، کانگ (۲۰۱۲) انتخاب طبقه‌بندی‌کننده‌ها در مجموعه‌ها با استفاده از الگوریتم‌های ژنتیک برای پیش‌بینی ورشکستگی، سیستم‌های خبره و کاربردها، ۳۹، ۹۳۱۴-۹۳۰۸.

۱۳) کیو، ئی، کوان، لی و شی (۲۰۱۹) بررسی پیش‌بینی ورشکستگی با استفاده از یادگیری ماشین و تکنیک‌های یادگیری عمیق، علوم کامپیوتر پروسیدیا، ۱۶۲، ۸۹۹-۸۹۵.

۱۴) لی، شیاهوی. یانگ، فن. یو و چن (۲۰۱۸) تشخیص تمایز دستمال‌های نرم با استفاده از طیف سنجی لیزری در ترکیب k نزدیکترین همسایگی و ماشین بردار پشتیبان، فناوری اپتیک و لیزر، ۱۰۲، ۲۳۹-۲۳۳.

۱۵) لیانگ، تینگ-پنگ. چندلر و هان (۱۹۹۰) ادغام روش‌های یادگیری آماری و استقرائی برای کسب دانش، سیستم‌های خبره و کاربردها، ۱، ۴، ۳۹۱-۴۰۱.

۱۶) ناصرزاده، هوشنگ (۱۳۷۴). قانون تجارت. تهران، نشردیدار.

۱۷) وپینک، ولادیمیر (۱۹۹۵) ماهیت نظریه یادگیری آماری، رسانه‌های علمی و تجاری اسپرینگر، نیویورک.

۱۸) ون گستل، تونی و دیگران (۲۰۰۳) پیش‌بینی ورشکستگی با استفاده از طبقه‌کننده‌های ماشین بردار پشتیبان حداقل مربعات، کنفرانس بین‌المللی IEEE در زمینه هوش محاسباتی برای مهندسی مالی، ۸-۱.

19) Asghari, Zahra & Esfahanipour (2019) Introducing a model for forecasting corporate bankruptcy by integrating the particle swarm optimization algorithm with support vector machine. The fourth national conference in management, Tehran.

20) Altman, Edward (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance, 23,4, 589-609.

21) Barboza, Kimura, & Altman (2017) Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, 405-417.

22) Beaver, William H (1966) Financial ratios as predictors of failure. Journal of accounting research, 71- 111.

23) Burges, Christopher (1998) A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2, 955–974.

24) Chen, Ning, & Vieira (2009) Bankruptcy Prediction based on Independent Component Analysis. In ICAART (pp. 150-155).

25) Deakin, Edward (1976) A discriminant analysis of predictors of business failure. Journal of Accounting Research, 167–179.

26) Ghazanfari, Mahdi, Rahimikia, Askari (2018) bankruptcy prediction of companies based on hybrid intelligent systems. Financial accounting and audit researches, 10, 37, 159-194.

27) Kim, Kang (2012) classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. Expert systems with Applications,39,9308-9314.

28) Li, Xiaohui, Yang, Fan, Yu & Chen (2018) Discrimination of soft tissues using laser-induced breakdown spectroscopy in combination with k nearest neighbors (KNN) and support vector machine (SVM) classifiers. Optics & Laser Technology, 102, 233-239.
29) Liang, Ting-peng, Chandler, & Han (1990) Integrating statistical and inductive learning methods for knowledge acquisition. Expert Systems with Applications, 1, 4, 391-401.
30) Naserzade, Houshang (1995) Business Law. Didar publication, Tehran.
31) Ohlson, James A (1980) Financial ratios and the probabilistic prediction of bankruptcy. Journal of accounting research, 109-131.
32) Qu, Yi, Quan, Lei & Shi (2019) Review of bankruptcy prediction using machine learning and deep learning techniques. Procedia Computer Science, 162, 895-899.
33) Tian, Yingjie, Qi (2014) Review on: twin support vector machines. Annals of Data Science, 1, 2, 253-277.
34) Tomar, Divya, Agarwal (2015) Twin support vector machine: a review from 2007 to 2014. Egyptian Informatics Journal, 16, 1, 55-69
35) Van Gestel, Tony et al (2003) Bankruptcy prediction with least squares support vector machine classifiers, IEEE International Conference on Computational Intelligence for Financial Engineering, 1-8.
36) Vapnik, Vladimir N (1995) The nature of statistical learning theory. Springer science & business media, New York.

یادداشت‌ها:

1 Multiple Discriminant Analysis (MDA)
2 Logit
3 Probit
4 Artificial Neural Network (ANN)
5 Support Vector Machine (SVM)
6 Data Mining
7 Hyperplane
8 Structural risk minimization principle (Srm)
9 K-Nearest Neighbors (KNN)
10 Decision Tree
11 Statistical learning theory (slt)
12 Under Sampling
13 Over Sampling
14 Logistic Regression
15 Linear regression (LR)
16 Zemijewski
17 Springate
18 Mackay
19 Vladimir N. Vapnik
20 Binary classifier
21 Feature space
22 Radial basis function (RBF)
23 Tehran Stock Exchange (TSE)
24 Cross-validation
25 Resubstitution error
26 Cross-Validation error

۲۷ پیشنویس توافقنامه بال دو، نخستین بار در ژوئن سال ۱۹۹۹ میلادی تنظیم و پس از چند مرحله انتشار مقدماتی و اخذ نظرات اصلاحی و اجرای آزمایشی، سرانجام از ابتدای سال ۲۰۰۸ میلادی به مرحله اجرا درآمد و دارای سه بخش (رکن) اصلی شامل: حداقل سرمایه لازم بانکها برای پوشش ریسک‌های مترتب، فرآیند بررسی نظارتی و انضباط بازارمی‌باشد که البته این عوامل به صورت متقابل بر یکدیگر تاثیرگذار هستند.

Bankruptcy prediction using hybrid data mining models based on misclassification penalty

Receipt: 12/03/2023 **Acceptance: 22/07/2023**

Atiye Torkaman[1]
Amir Abbas Najafi, Professor[2]

Abstract

In recent years, data mining, particularly the support vector machine, has gained considerable interest among investors, managers, and researchers as an effective means of bankruptcy prediction. However, studies indicate that it is highly sensitive to the selection of parameters and input variables. Hence, the aim of this research is to improve bankruptcy prediction accuracy by combining an advanced support vector machine model with the k-nearest neighbor approach to eliminate erroneous entries. To achieve this, first, by using five financial ratios: current ratio, net profit margin, debt ratio, return on assets, and return of investment from 150 companies listed on the Tehran Stock Exchange during the 10-year period (2010-2019), and k-nearest neighbor algorithm, the training data will be refined. Then, relying on a support vector machine based on classification penalty, a prediction model will be constructed. The parameters will be estimated, and its validity will be assessed using test data. Finally, a comparison will be made between the outcomes of the proposed model and traditional models.The findings demonstrate that the combination of the k-nearest neighbor models and support vector machine reduces the overall prediction error, and the penalty coefficients of the support vector machine exhibit a high level of statistical significance.

Keywords

Bankruptcy Prediction, Data Mining, Support Vector Machine, K-Nearest Neighbor, Penalty Coefficients.

1-Financial Engineering Department, Faculty of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran. atiye.torkaman@gmail.com
2-Financial Engineering Department, Faculty of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran. (Corresponding Author) aanajafi@kntu.ac.ir