فصلنامه علمی پژوهشی
دانش حسابداری و حسابرسی مدیریت
دوره ۱٤/ شماره ۳ (پیاپی ۵۵)/ پائیز ۱٤۰٤
صفحه ۲۶۵ تا ۲۷۸

مدل‌یابی پیش‌بینی ورشکستگی با تاکید بر روش‌های نوین اندازه‌گیری شبکه‌های عصبی و ماشین بردار پشتیبان

محمدرضا نصرت‌نظامی
دانشجوی دوره دکترای مهندسی مالی، گروه مهندسی مالی، واحد رشت، دانشگاه آزاد اسلامی، رشت، ایران
Mrnn1350@gmail.com

فاضل محمدی نوده
استادیار، گروه مدیریت، واحد لاهیجان، دانشگاه آزاد اسلامی، لاهیجان، ایران
(نویسنده مسئول)
Mnfazel2@gmail.com

سینا خردیار
دانشیار، گروه حسابداری، واحد رشت، دانشگاه آزاد اسلامی، رشت، ایران
Kheradyar@iaurasht.ac.ir

تاریخ دریافت: ۱۴۰۲/۱۰/۱۰ تاریخ پذیرش: ۱۴۰۲/۱۲/۱۴

چکیده

هدف از انجام این تحقیق مدل‌یابی پیش‌بینی ورشکستگی با تاکید بر روش‌های نوین اندازه‌گیری شبکه‌های عصبی و ماشین بردار پشتیبان در شرکت‌های بورس اوراق بهادار تهران می‌باشد. نتایج این پژوهش نشان‌دهنده این موضوع است که استفاده از ساختارهای جدید مانند سیستم‌های هوشمند ترکیبی مبتنی بر مدل‌های داده کاوی توانایی بالایی در تشخیص ورشکستگی شرکت‌ها در سطح کشور دارد. در این تحقیق داده‌های مورد استفاده به دو صنعت مواد غذایی و نساجی تعمیم یافته و داده‌های پژوهش محدود به داده‌های شرکت‌های بورس اوراق بهادار نمی‌باشد. مشاهده می‌شود که در دو صنعت مورد بررسی نسبت مالی "سود انباشته به کل دارایی‌ها" بالاترین تکرار را در پیش‌بینی ورشکستگی به خود اختصاص داده است. به این دلیل نسبت فوق به عنوان نسبت برتر برای پیش بینی ورشکستگی انتخاب خواهد شد. نتایج، مسأله مهم دیگری را نیز نمایان می‌سازد. مشاهده می‌شود که در هر یک از صنایع نسبت‌های مالی نسبتاً متفاوتی به عنوان نسبت‌های مالی با بالاترین تکرار انتخاب شده‌اند. این مسأله نشان می‌دهد که امکان در نظر گرفتن یک مجموعه مشخص نسبت مالی به عنوان ورودی مدل همانگونه که در اکثر مطالعات انجام شده اینگونه در نظر گرفته شده است برقرار نبوده و در هر صنعت با توجه به ویژگی‌ها و ساختار، ترکیب ورودی‌ها با صنعت دیگر متفاوت می‌باشد. در نتیجه می‌توان یافته‌های این بخش را به عنوان یک نوآوری در پژوهش حاضر و یکی از دلایل مهم برتری سیستم طراحی شده نسبت به ساختارهای مختلف ارائه شده در مطالعات داخلی و بین‌المللی عنوان کرد.

واژه‌های کلیدی: ورشکستگی، روش‌های نوین اندازه‌گیری، شبکه عصبی مصنوعی و ماشین بردار پشتیبان.

۱- مقدمه

در تجزیه‌وتحلیل ورشکستگی مالی، تشخیص شرکت‌هایی که در معرض خطر ورشکستگی هستند، در آماده‌سازی برای محافظت در برابر هر گونه خسارت مالی که شرکت‌های در معرض خطر قرار دارند وارد کنند، بسیار مهم است (Kim et al، ۲۰۱۸). باتوجه‌به Rybárová و همکاران (۲۰۱۶)، مدل‌های ورشکستگی، سیستم‌های هشدار اولیه مبتنی بر تجزیه‌وتحلیل شاخص‌های انتخابی هستند که قادر به شناسایی یک موضوع برای سلامت مالی یک شرکت هستند. (Kiaupaite-Grushniene ۲۰۱۶) بیان می‌کند که ایجاد مدل‌های قابل‌اعتماد پیش‌بینی ورشکستگی برای فرایندهای مختلف تصمیم‌گیری ضروری است. به گفته موسوی و همکاران (۲۰۱۵)، مدل‌های پرکاربرد عمدتاً Altman Z-Score، Taffler Z-Score و Index IN۹۵ هستند. تعداد زیادی از محققان دانشگاهی از سراسر جهان درحال‌توسعه مدل‌های پیش‌بینی ورشکستگی شرکت‌ها بر اساس تکنیک‌های مدل‌سازی مختلف هستند. روش‌های آماری متعددی توسعه‌یافته است (Balcaen and Ooghe ۲۰۰۴). با وجود محبوبیت روش‌های آماری کلاسیک، مشکلات قابل‌توجهی دررابطه با کاربرد این روش‌ها برای پیش‌بینی ورشکستگی شرکت‌ها باقی‌مانده است. بنابراین، ایجاد تکنیک‌های ورشکستگی شرکتی قابل اعتماد در مدیریت ریسک معاصر در سطوح خرد و کلان ضروری است. باتوجه به اهمیت اقتصادی شرکت‌ها، تحقیقات مربوط به پیش‌بینی ورشکستگی شرکت‌ها برای سیاست‌گذاران برای تحریک رشد و توسعه اقتصاد کلان حیاتی است بنابراین، مدل‌های پیش‌بینی ورشکستگی به بانک‌ها در ارزیابی ریسک ورشکستگی شرکت‌ها کمک می‌کند و در واقع، پیش‌بینی به موقع ورشکستگی شرکت‌ها ممکن است به وام دهندگان کمک کند تا از برخی هزینه‌های مربوط به ورشکستگی جلوگیری کنند (Tinoco و Wilson ۲۰۱۳). همچنین، مدل‌های پیش‌بینی ورشکستگی به مؤسسات مالی کمک می‌کند تا سیاست‌های مرتبط با تأمین و هزینه اعتبار برای شرکت‌ها را طراحی کنند، و در تعیین کفایت سرمایه برای بانک‌ها که زیان‌های غیرمنتظره آتی را پوشش می‌دهند، ضروری هستند.

علاوه بر این، از آنجایی که صاحبان و مدیران شرکت‌ها از چالش‌ها و تهدیدات احتمالی برای شرکت‌های خود آگاه می‌شوند، پیش‌بینی دقیق شکست شرکت‌ها می‌تواند به مالکان و مدیران شرکت اجازه دهد تا فعالیت‌های اصلاحی انجام دهند و شرکت‌ها را از ورشکستگی مالی نجات دهند و در نتیجه ثبات در هر دو بانک را ارتقا دهند. صنعت و بخش غیرمالی شرکتی اگرچه تعداد زیادی از ادبیات پیش‌بینی ورشکستگی شرکت‌های آکادمیک و حرفه‌ای ارائه شده است، اکثریت تحقیقات عمدتاً بر روی مدل‌های تشخیص ورشکستگی برای شرکت‌های دولتی در کشورهای توسعه‌یافته تمرکز دارد، زیرا داده‌ها و اطلاعات قابل‌اعتماد و جامع ورشکستگی برای شرکت‌های سهامی عام است. به طور گسترده در دسترس است (پاپانا و اسپیریدو ۲۰۲۰؛ چارالامباکیس و گرت ۲۰۱۹؛ چارالامباکیس ۲۰۱۴). تحقیقات در مورد پیش‌بینی ورشکستگی برای شرکت‌های خصوصی به طور کلی محدود است و حتی برای شرکت‌های خصوصی در اقتصادهای درحال‌توسعه محدودتر است.

ادبیات پیش‌بینی ورشکستگی برای شرکت‌های سهامی عام در مطالعات متعدد مورد بررسی قرار گرفته است، هیچ بررسی سیستماتیک یا هر شکل دیگری از بررسی ادبیات برای شرکت‌های تولیدی در اقتصادهای در حال انحلال انجام نشده است. این مطالعه دانش جدیدی را با پیش‌بینی ورشکستگی شرکت‌های تولیدی در اقتصادهای درحال‌توسعه اضافه می‌کند. تأکید اصلی بررسی محدوده، تحلیل دلایل و انگیزه های تحقیق، روندهای نوظهور و شکاف‌های تحقیقاتی در مدل سازی ریسک ورشکستگی برای شرکت‌های خصوصی در کشورهای درحال‌توسعه است. محققان بر این باورند که این تحقیق قاله اولین کار تحقیقاتی است که چنین بررسی محدوده‌ای را انجام می‌دهد. بررسی محدوده نتیجه می‌گیرد که تخمین احتمال ورشکستگی شرکت‌های خصوصی در بازارهای توسعه نیافته یک رشته مرتبط است که به درستی مورد بررسی قرار نگرفته است و به دلیل پیچیدگی آن و اخلاق تجاری متفاوت شرکت‌های خصوصی، دارای حوزه‌های منحصربه‌فرد و ناشناخته‌ای است.

ادبیات نظری و پیشینه

ورشکستگی

عدم توانایی تاجر به پرداخت دیون و بدهی‌ها، ورشکستگی نامیده می‌شود. در تعریف ورشکستگی طبق ماده ۴۱۲ قانون تجارت آمده است، که ورشکستگی تاجر یا شرکت‌تجاری، به دلیل توقف پرداخت بدهی‌ها اتفاق می‌افتد. ورشکسته به تجار و شرکت‌های تجاری و بازرگانی اطلاق می‌شود. اگر این افراد قادر به پرداخت بدهی‌ها و تعهدات خود نباشند، ورشکسته محسوب می‌شوند.

پیش‌بینی ورشکستگی

پیش‌بینی ورشکستگی شرکت یکی از علوم مدیریتی مهم است و هدف اصلی آن تمایز بین شرکتهایی با احتمال توقیف قانونی از شرکت های سالم است. علاوه بر این، اتخاذ تصمیمات نادرست در موسسات مالی می‌تواند به بروز مشکلات مالی یا توقیف شرکت بیانجامد و هزینه های اجتماعی زیادی را به مالکان یا سهامداران، مدیران، دولتها و غیره تحمیل نماید. به همین دلیل، نحوه پیش‌بینی ورشکستگی شرکتها به موضوعی پر طرفدار هم

در حوزه‌های صنعتی و هم در تحقیقات دانشگاهی بدل شده است. از آنجاییکه هنوز هیچ نظریه پخته‌ای در زمینه ورشکستگی شرکتها وجود ندارد، مطالعات انجام شده در این حوزه غالباً بر اساس فرایندهای سعی و خطای گزینش ویژگی و مدل های پیش‌بینی کننده صورت پذیرفته اند. در گذشته بسیاری از محققان نسبت به ارائه روشهایی آماری در زمینه ورشکستگی شرکتها اقدام نموده اند. اصلی ترین روش های آماری عبارتند از LDA,MDA,QDA,LRA و FA.[کریمی شهرکی، مهدی و مطلبیان، مجتبی(۱۳۹۵)

شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی یا به زبان ساده‌تر شبکه‌های عصبی سیستم‌ها و روش‌های محاسباتی نوین برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به دست آمده در جهت بیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده هستند.

ماشین بردار پشتیبان

ماشین بردار پشتیبان یا Support Vector Machine که به‌اختصار به آن SVM گفته می‌شود یک الگوریتم یادگیری ماشین با ناظر است که نمونه‌ی داده‌هایی را به‌صورتی نقاطی در فضا نشان داده شده است، با استفاده از یک خط یا هایپرپلین از هم جدا می‌کند. این جداسازی به‌گونه‌ای است که نقاط داده‌ای که در یک طرف خط هستند مشابه به‌هم و در یک گروه قرار می‌گیرند.

روش‌شناسی پژوهش

تحقیق حاضر از لحاظ هدف، از نوع تحقیقات کاربردی و از نظر نوع روش، روش تحقیق آمیخته۱ محسوب می‌گردد. جامعه آماری پژوهش حاضر در مرحله اول، جامعه آماری پژوهش تعداد ۸۰ شرکت، کارشناس و خبرگان در حوزه ساخت و ساز و کارشناسان در منطقه ۳ شهرداری تهران می‌باشد. در پژوهش حاضر برای تجزیه‌وتحلیل داده‌ها در بخش کیفی از نرم‌افزار MAX QDA استفاده شد. از روشهای آمار توصیفی و آمار استنباطی از روش‌های شبکه عصبی و ماشین بردار پشتیبان با استفاده از نرم افزار SPSS و گمز و متلب استفاده می‌شود.

یافته‌های پژوهش

نتایج دلفی فازی

راند یک : میانگین فازی و برونداد فازی زدائی شده مقادیر مربوط به شاخص‌ها آمده است. مقدار فازی زدائی شده بزرگتر از ۰/۷ مورد قبول است و هر شاخصی که امتیاز کمتر ۰/۷ داشته باشد رد می‌شود (وو و فانگ ، ۲۰۱۱، ص۷۵۱). تمامی مواردی امتیازی بیشتر از ۷ کسب کرده‌اند و پذیرش شدند.

راند دو: تحلیل دلفی فازی برای شاخص‌های باقی مانده در راند دوم ادامه پیدا کرد. نتایج حاصل از فازی‌زدایی عناصر در راند دوم در جدول زیر گزارش شده است: در این مرحله نیز تمام موارد در راند دوم نیز تمام موارد مجددا امتیازی بیشتر از ۰.۷ کسب کردند و نیاز به حرف هیچ شاخصی نبود.

پایان راندهای تکنیک دلفی: در دور دوم هیچ شاخصی حدف نشد که این خود نشانه‌ای برای پایان راندهای دلفی است. بطور کلی یک رویکرد برای پایان دلفی آن است که میانگین امتیازات سوالات راند اول و راند دوم باهم مقایسه شوند. در صورتیکه اختلاف بین دو مرحله از حد آستانه خیلی کم (۰/۸) کوچکتر باشد در این صورت فرایند نظرسنجی متوقف می‌شود (چنگ و لین، ۲۰۰۲؛ نمونه فارسی لطیفی و همکاران، ۱۳۹۷). در پژوهش حاضر برای تجزیه‌وتحلیل داده‌ها در بخش کیفی از نرم‌افزار MAX QDA استفاده شد.

ساختار هسته سیستم

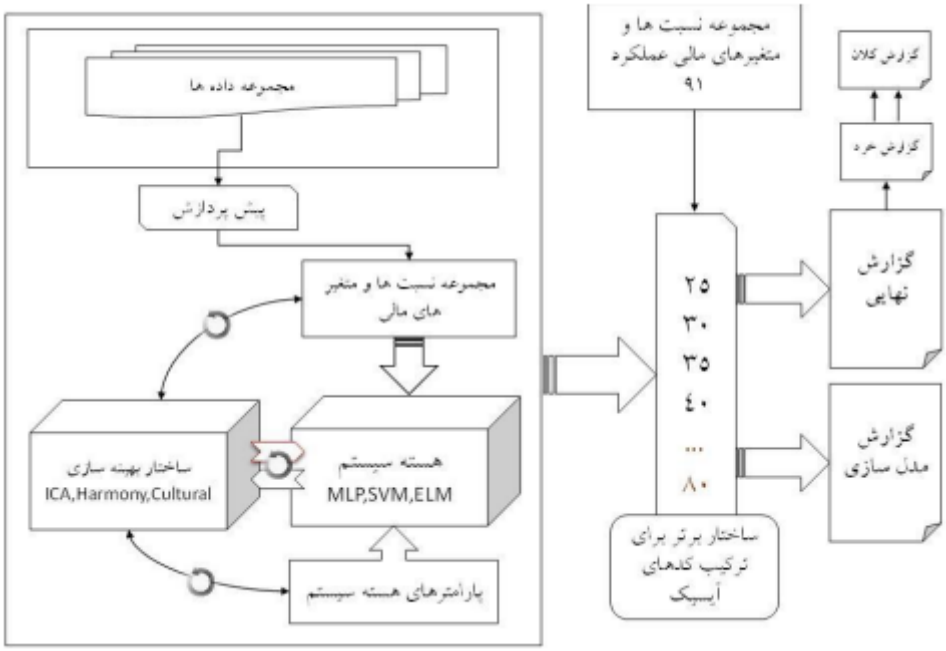
ساختار کلی هسته سیستم در شکل ۲ ارائه شده است.

مطابق با شکل ۲ الگوریتم‌های بهینه‌سازی اقدام به بهینه‌سازی ترکیب ورودی‌های سیستم (۲۱ متغیر ورودی) و پارامترهای هسته سیستم خواهند کرد. در هر قدم برای اندازه‌گیری تابع هزینه الگوریتم بهینه‌سازی از ۴۰ بار تکرار ساختار (۴ تکرار موازی اعتبارسنجی ضربدری ۱۰ بخشی) استفاده خواهد شد. بازه‌های بهینه‌سازی و مقادیر پیش‌فرض ارائه شده توسط طراحان الگوریتم‌های بهینه‌سازی در زیر ارائه شده است. لازم به ذکر است بازه‌های بهینه‌سازی مدل‌های طبقه‌بندی بر پایه نظرخواهی از خبرگان و سعی و خطا انتخاب شده و امکان بررسی سایر بازه‌ها در مطالعات آتی برقرار می‌باشد.

همچنین برای آموزش شبکه عصبی از الگوریتم trainscg با توابع انتقال تانژانت سیگموئید در لایه ورودی و میانی و تابع انتقال Softmax در لایه خروجی بهره گرفته شده است.

با استفاده از ساختارهای احتمالی شبکه عصبی و بردار پشتیبان (بهره‌گیری از برداری‌سازی امکان ارائه احتمال ورشکستگی در مدل‌های فوق برقرار می‌باشد اما در دو مدل مبتنی بر یادگیری تشدید شده خروجی‌های نهایی تنها به صورت دودویی و بدون ارائه احتمال قرارگیری در هر گروه می‌باشند.

1 Mixed Methods Research



شکل ۲- ساختار سیستم هوشمند پیش‌بینی ورشکستگی شرکت‌ها

منبع: یافته‌های پژوهشگر

در پایان ۴۰ مدل آموزش دیده (۴ تکرار موازی اعتبارسنجی ضربدری ۱۰ بخشی) خروجی سیستم می‌باشد. تمامی مدل‌ها از لحاظ متغیرهای مالی منتخب (از میان ۲۱ متغیر ورودی اولیه) یکسان می‌باشند. اطلاعات مالی وارد ۴۰ مدل فوق شده و ۴۰ احتمال (و یا خروجی دودویی) حاصل می‌شود. در مرحله بعد به منظور حذف خروجی‌های پرت، ۲ انحراف معیار مطابق معادله ۱ از نتایج حذف شده و سپس از باقی‌مانده خروجی‌ها میانگین گرفته خواهد شد.

- شبکه عصبی پرسپترون چندلایه: تعداد لایه‌ها (۱ یا ۲)
- تعداد نورون‌ها در هر لایه (لایه اول: ۴ تا ۲۵/لایه دوم:۰ تا ۲۵)
- ماشین بردار پشتیبان: (Cبین و) –گاما (بین و)
- ماشین با یادگیری تشدید شده (بدون هسته): تعداد نورون‌های لایه پنهان (بین ۱ تا ۵۰) – پارامتر مرتبط با کرنل (RBFبین و ۸۰۰)
- ماشین با یادگیری تشدید شده (دارای هسته): ضریب تنظیم[و پارامتر مرتبط با کرنل (RBFبین و ۸۰۰)

همچنین برای آموزش شبکه عصبی از الگوریتم trainscg با توابع انتقال تانژانت سیگموئید در لایه ورودی و میانی و تابع انتقال Softmax در لایه خروجی بهره گرفته شده است.

با استفاده از ساختارهای احتمالی شبکه عصبی و بردار پشتیبان (بهره‌گیری از برداری‌سازی) امکان ارائه احتمال ورشکستگی در مدل‌های فوق برقرار می‌باشد اما در دو مدل مبتنی بر یادگیری تشدید شده خروجی‌های نهایی تنها به صورت دودویی و بدون ارائه احتمال قرارگیری در هر گروه می‌باشند.

در پایان ۴۰ مدل آموزش دیده (۴ تکرار موازی اعتبارسنجی ضربدری ۱۰ بخشی) خروجی سیستم می‌باشد. تمامی مدل‌ها از لحاظ متغیرهای مالی منتخب (از میان ۲۱ متغیر ورودی اولیه) یکسان می‌باشند. اطلاعات مالی وارد ۴۰ مدل فوق شده و ۴۰ احتمال (و یا خروجی دودویی) حاصل می‌شود. در مرحله بعد به منظور حذف خروجی‌های پرت، ۲ انحراف معیار مطابق معادله ۱ از نتایج حذف شده و سپس از باقی‌مانده خروجی‌ها میانگین گرفته خواهد شد.

$$|X - average(X)| \geq \alpha * STD(X)$$

در معادله ۱ برابر با ۲ در نظر گرفته شده و X مجموعه احتمالات خروجی را شامل می‌شود. علاوه بر آن نشان دهنده اپراتور میانگین و نیز نشان‌دهنده اپراتور انحراف معیار می‌باشد. مقادیریکه معادله ۱ برای آن‌ها برقرار می‌باشد از مجموعه احتمالات حذف می‌شود. همچنین در حالت خروجی دودویی از رأی‌گیری حداکثری استفاده می‌شود.

علاوه بر استفاده از معادله ۱ به منظور حذف خروجی‌های پرت، از این معادله به منظور تشخیص نمونه‌های ورودی پرت نیز استفاده خواهد شد. به طور خلاصه معیار ذکر شده بر روی

هر نمونه (شرکت) اعمال شده و در صورت برقرار بودن آن بر روی هر یک از نسبت‌ها و متغیرهای مالی آن نمونه، نمونه مورد بررسی حذف خواهد شد. در این پژوهش استفاده از فرآیند حذف داده‌های پرت در مرحله پیش پردازش نیز مورد آزمون قرار خواهد گرفت.

یکی از موارد مهم در طراحی سیستم‌های نرم افزاری محاسباتی بخصوص در شرایط استفاده از پردازش موازی و یا پردازش مبتنی بر هسته‌های گرافیکی بحث توابع تولید کننده اعداد تصادفی می‌باشد. در این پژوهش برای تولید اعداد تصادفی از روش Mersenne Twister استفاده شده است. یکی از محدودیت‌های این روش عدم امکان ایجادزیرجریان ها[جریان‌های انشعابی) از جریان اصلی ایجاد شده می‌باشد. با تغییر کدهای پایه (پیش‌فرض نرم‌افزاری)، اعداد تصادفی تولید شده توسط روش Mersenne Twister تحت بسته‌های مشخصی به هسته‌های سیستم ارسال شده است. همچنین با تغییر فرآیند پیش‌فرض، رشته‌های اعداد تصادفی تولید شده ویژگی‌های مورد نیاز (از جمله استقلال) را دارا خواهند بود.

اعتبارسنجی

روش اعتبارسنجی این پژوهش ساختاری اصلاح شده از اعتبارسنجی ضربدری پایه می‌باشد. در روش اعتبارسنجی ضربدری ۱۰ بخشی پایه ابتدا مجموعه نمونه‌ها به ۱۰ بخش تقسیم می‌شود (به صورت تصادفی). سپس مدل با چهاربخش اول آموزش دیده و با بخش پنجم مورد آزمون قرار می‌گیرد. در مرحله بعد مدل با بخش دوم تا پنجم آموزش دیده و با بخش اول آزمون می‌شود. این فرآیند ۱۰ بار در کل داده‌ها تکرار خواهد شد. یکی از محدودیت‌های این روش این مسأله است که برای بخش اعتبارسنجی مورد نیاز شبکه عصبی داده‌ای تخصیص داده نمی‌شود. در روش اصلاح شده داده‌ها به ۱۰ بخش (تصادفی) تقسیم می‌شوند. سپس بخش اول به عنوان داده‌های آزمون (۲۰ درصد) و بخش دوم نیز به عنوان داده‌های اعتبارسنجی (۲۰ درصد) انتخاب شده مابقی داده‌ها (۶۰ درصد) برای آموزش سیستم مورد استفاده قرار خواهند گرفت. این فرآیند ۱۰ بار به صورت چرخشی مورد تکرار قرار می‌گیرد. مشاهده می‌شود که ۶۰ درصد نمونه‌ها به منظور آموزش، ۲۰ درصد به منظور آزمون و ۲۰ درصد نیز برای اعتبارسنجی و جلوگیری از برازش بیش از اندازه استفاده خواهند شد. همچنین برای جلوگیری از برازش بیش از اندازه از روش توقف زود هنگام در ۶ تکرار (مقدار پیش‌فرض) بهره گرفته شده است. در سایر مدل‌های طبقه‌بندی از اعتبارسنجی ضربدری ۱۰ بخشی پایه استفاده خواهد شد. برای طراحی ساختار اصلی سیستم از نرم افزار MATLAB نسخه a۲۰۱۴R و b۲۰۱۴R بهره گرفته شده است. برای گزارش‌گیری استاندارد از سیستم از [lxxviii]VBAدر محیط Microsoft ۲۰۱۳Excel استفاده شده و ارتباطات بین دو نرم افزار کدنویسی شده است. علاوه بر موارد فوق برای بهبود عملکرد و کارایی سیستم در بخش‌هایی از زبان جاوا در محیط MATLAB استفاده شده است. همچنین به منظور ارتباط با پایگاه داده و تبدیل و انتقال داده‌ها از ۲۰۱۴SQL server بهره گرفته شده است. واحد پردازش مرکزی مورد استفاده Intel Core i۷– MQ۴۷۰۲ دارای ۴ هسته فیزیکی و ۸ هسته منطقی بوده و هشت گیگابایت حافظه دسترسی تصادفی و سیستم‌عامل ۸.۱Windows bit۶۴Enterprise edition مورد استفاده قرار گرفته است. به منظور افزایش سرعت پردازش از تولباکس پردازش موازی برای انتقال تکرار اعتبارسنجی ضربدری به پردازش موازی بهره گرفته شده است. ماشین بردار پشتیبان بر پایه کتابخانه ماشین بردار پشتیبان و مدل‌های ماشین با یادگیری تشدید شده نیز بر پایه کدهای ارائه‌کنندگان این روش پیاده‌سازی شده‌اند.

نتایج اجرا و بهینه‌سازی سیستم

نتایج سیستم به شرح جدول ۴ می‌باشد. در این پژوهش حداکثر تعداد قدم‌ها ۱۰۰ و تعداد قدم‌های عدم بهبود یک پنجم این مقدار (۲۰ قدم) در نظرگرفته شده است. علاوه بر آن ترکیب‌های برتر (بر مبنای بالاترین دقت کلی) در جدول ۴ نشان داده شده است. مجموعه نتایج بر مبنای میانگین دقت‌های حاصل از داده‌های آزمون ارائه شده است.

جدول ۴- نتایج سیستم هوشمند پیش‌بینی ورشکستگی شرکت‌ها

		عدم استفاده از حذف داده‌های پرت					حذف داده‌های پرت – ۲ انحراف معیار				
	روش	تعداد سالم ورشکسته	دقت کلی	دقت تشخیص سالم	دقت تشخیص ورشکسته	سطح زیر ROCنمودار	تعداد سالم ورشکسته	دقت کلی	دقت تشخیص سالم	دقت تشخیص ورشکسته	سطح زیر ROCنمودار
صنعت ۱	شبکه عصبی مصنوعی	۴۸۲۶ ۹۹۹	۷۴/۸۴	۷۹/۷۲	۶۹/۹۵	۰/۷۴۸۳	۴۳۶۸ ۸۷۴	۸۵/۰۷	۸۴/۴۶	۸۵/۶۷	۰/۸۵۱۷

		عدم استفاده از حذف داده‌های پرت					حذف داده‌های پرت – ۲ انحراف معیار				
	روش	تعداد سالم ورشکسته	دقت کلی	دقت تشخیص سالم	دقت تشخیص ورشکسته	سطح زیر ROCنمودار	تعداد سالم ورشکسته	دقت کلی	دقت تشخیص سالم	دقت تشخیص ورشکسته	سطح زیر ROCنمودار
	ماشین بردار پشتیبان		۸۳/۸۲	۸۱/۷۳	۸۵/۹۰	۰/۸۳۸۲		۸۳/۴۰	۸۳/۶۲	۸۳/۱۹	۰/۸۳۴۴
	(بدون ELM هسته)		۸۰/۹۴	۷۹/۱۸	۸۲/۷۱	۰/۸۱۰۱		۸۳/۱۴	۸۴/۱۶	۸۲/۱۲	۰/۸۳۲۰
	(با ELM بهره‌گیری از هسته)		۷۸/۶۷	۸۱/۲۸	۷۶/۰۶	۰/۷۸۷۴		۸۲/۳۶	۸۲/۱۲	۸۲/۶۰	۰/۸۲۳۷
صنعت۲	شبکه عصبی مصنوعی	۳۲۴۱ ۸۴۸	۶۷/۱۰	۶۵/۹۰	۶۸/۳۰	۰/۶۷۱۱	۲۹۰۰ ۷۸۰	۸۳/۹۲	۸۱/۸۴	۸۶/۰۰	۰/۸۳۹۵
	ماشین بردار پشتیبان		۸۰/۹۶	۸۲/۰۹	۷۹/۸۳	۰/۸۰۹۵		۸۷/۰۹	۹۰/۹۴	۸۳/۲۲	۰/۸۷۱۰
	(بدون ELM هسته)		۸۳/۰۰	۷۸/۶۶	۸۷/۳۵	۰/۸۳۰۲		۸۴/۴۶	۸۱/۴۸	۸۷/۴۵	۰/۸۴۴۶
	(با ELM بهره‌گیری از هسته)		۷۰/۳۹	۶۸/۱۴	۷۲/۶۳	۰/۷۰۴۲		۸۳/۸۲	۸۴/۵۷	۸۳/۰۶	۰/۸۳۸۲
صنعت۳	شبکه عصبی مصنوعی	۴۸۲۶ ۹۹۹	۶۷/۱۰	۶۵/۹۰	۶۸/۳۰	۰/۶۷۱۱	۴۳۶۸ ۸۷۴	۸۳/۹۲	۸۱/۸۴	۸۶/۰۰	۰/۸۳۹۵
	ماشین بردار پشتیبان		۸۰/۹۶	۸۲/۰۹	۷۹/۸۳	۰/۸۰۹۵		۸۷/۰۹	۹۰/۹۴	۸۳/۲۲	۰/۸۷۱۰
	(بدون ELM هسته)		۸۳/۰۰	۷۸/۶۶	۸۷/۳۵	۰/۸۳۰۲		۸۴/۴۶	۸۱/۴۸	۸۷/۴۵	۰/۸۴۴۶
	(با ELM بهره‌گیری از هسته)		۷۰/۳۹	۶۸/۱۴	۷۲/۶۳	۰/۷۰۴۲		۸۳/۸۲	۸۴/۵۷	۸۳/۰۶	۰/۸۳۸۲
صنعت ۴	شبکه عصبی مصنوعی	۳۲۴۱ ۸۴۸	۷۳/۵۵	۷۶/۰۷	۷۱/۰۴	۰/۷۳۶۶	۲۹۰۰ ۷۸۰	۸۵/۰۹	۸۵/۶۴	۸۴/۵۵	۰/۸۵۱۲
	ماشین بردار پشتیبان		۸۳/۹۵	۸۳/۴۸	۸۴/۴۱	۰/۸۳۹۶		۸۴/۷۲	۸۴/۶۷	۸۴/۷۷	۰/۸۴۷۳
	(بدون ELM هسته)		۸۲/۶۱	۷۸/۵۴	۸۶/۶۷	۰/۸۲۶۴		۸۳/۵۳	۷۹/۷۹	۸۷/۲۶	۰/۸۳۵۶
	(با ELM بهره‌گیری از هسته)		۸۰/۱۲	۷۵/۲۵	۸۴/۹۹	۰/۸۰۱۷		۸۵/۱۹	۸۸/۷۷	۸۱/۶۲	۰/۸۵۲۰
صنعت ۵	شبکه عصبی مصنوعی	۴۸۲۶ ۹۹۹	۶۶/۹۹	۶۴/۷۳	۶۹/۲۵	۰/۶۷	۴۳۶۸ ۸۷۴	۸۳/۰۴	۸۳/۴۴	۸۲/۶۳	۰/۸۳۰۶
	ماشین بردار پشتیبان		۸۰/۸۹	۸۲/۰۶	۷۹/۷۱	۰/۸۰۹۰		۸۷/۱۵	۸۹/۵۹	۸۴/۷۰	۰/۸۷۲۰
	(بدون ELM هسته)		۸۰/۳۴	۷۶/۴۲	۸۴/۲۶	۰/۸۰۴۰		۸۵/۷۳	۸۶/۴۷	۸۴/۹۹	۰/۸۵۷۴
	(با ELM بهره‌گیری از هسته)		۸۰/۰۹	۸۳/۶۲	۷۶/۵۶	۰/۸۰۱۱		۸۴/۱۹	۸۲/۷۴	۸۵/۶۳	۰/۸۴۲۷
صنعت ۶	شبکه عصبی مصنوعی	۳۲۴۱ ۸۴۸	۷۴	۷۹/۴۵	۶۸/۵۴	۰۰/۷۴	۲۹۰۰ ۷۸۰	۸۲/۹۸	۸۲/۰۱	۸۳/۹۶	۰/۸۲۹۷

		عدم استفاده از حذف داده‌های پرت					حذف داده‌های پرت – ۲ انحراف معیار				
	روش	تعداد سالم ورشکسته	دقت کلی	دقت تشخیص سالم	دقت تشخیص ورشکسته	سطح زیر ROCنمودار	تعداد سالم ورشکسته	دقت کلی	دقت تشخیص سالم	دقت تشخیص ورشکسته	سطح زیر ROCنمودار
	ماشین بردار پشتیبان		۷۲/۳۱	۷۸/۰۹	۶۶/۵۲	۰/۷۲۳۱		۸۳/۱۷	۸۳/۶۶	۸۲/۶۸	۰/۸۳۲۱
	(بدون ELM هسته)		۷۹/۸۶	۷۴	۸۵/۷۰	۰/۷۹۸۹		۸۰/۶۲	۸۱/۶۱	۷۹/۶۲	۰/۸۰۶۶
	(با ELM بهره‌گیری از هسته)		۷۲/۶۶	۷۸/۱۵	۶۷/۱۷	۰/۷۲۶۷		۸۳/۳۰	۸۷/۱۲	۷۹/۴۷	۰/۸۳۳۳

منبع: یافته‌های پژوهشگر

اثر حذف داده‌های پرت در فرآیند پیش‌پردازش

برای بررسی میزان اثر حذف داده‌های پرت، کلیه داده‌های صنعت مورد نظر قبل از بالانس داده را وارد مجموعه مدل‌ها کرده و تعداد نمونه‌ها (شرکت‌های) حذفی اندازه‌گیری شد. نتایج در جدول ۵ ارائه شده است.

جدول ۵- اثر حذف داده‌های پرت

تعداد کل شرکت‌ها	تعداد حذفی	تعداد کل شرکت‌ها	تعداد حذفی
۵۸۲۸	۱۲۳۰	۴۰۹۲	۸۶۷

منبع: یافته‌های پژوهشگر

مشاهده می‌شود که اعمال فرآیند حذف داده‌های پرت موجب حذف تعداد گسترده‌ای از شرکت‌ها در هر صنعت شده است. دلیل این مسأله را می‌توان در ساختار داده‌های ورودی به سیستم بررسی کرد. همان‌طور که ذکر شد در اکثر مطالعات انجام شده داده‌های مورد استفاده از بورس اوراق بهادار استخراج شده است. شرکت‌های بورسی اغلب دارای اندازه‌های مشابه (و بزرگ) بوده و تنوع اندازه‌ای در میان آن‌ها نسبت به مقیاس کل کشور کمتر می‌باشد. با حرکت به سمت استفاده از داده‌های کل کشور در یک صنعت مشخص که در این پژوهش مورد توجه قرار گرفته است تعداد و تنوع شرکت‌ها به شکل قابل توجهی افزایش یافته و همین امر موجب می‌شود تا تعداد شرکت‌های حذفی نیز بالا باشد. از سوی دیگر این امکان وجود دارد که نمونه‌های حذفی دارای ویژگی‌های خلاف قاعده (در این پژوهش ورشکسته) بوده و به اشتباه به عنوان داده پرت در نظر گرفته شده باشند. از طرف دیگر با توجه به اعمال فرآیند حذف داده‌های پرت بر روی داده‌های خارجی (داده‌های جدید وارد شده به سیستم) امکان تصمیم‌گیری و سیاست‌گذاری در رابطه با بخش زیادی از شرکت‌ها از دست خواهد رفت. با توجه به دلایل ذکر شده در این مرحله "عدم استفاده از فرآیند حذف داده‌های پرت" به عنوان روش برتر انتخاب شده و فرضیه اول تحقیق مبنی بر برتری استفاده از فرآیندهای حذف داده‌های پرت نسبت به عدم استفاده از فرآیندهای حذف داده‌های پرترد خواهد شد.

بررسی شهودی و آماری الگوریتم‌های بهینه‌سازی

به صورت شهودی در جدول ۴ مشاهده می‌شود که الگوریتم فرهنگی در اکثر حالات دقت کمتری را از دو الگوریتم دیگر دارا می‌باشد. در صورت معنادار بودن تفاوت نتایج الگوریتم بهینه‌سازی فرهنگی با الگوریتم رقابت استعماری و جستجوی هارمونی الگوریتم فوق از فرآیند حذف خواهد شد. برای بررسی آماری تفاوت ذکر شده از آزمون مکنمار بهره گرفته شده است. آزمون مکنمار یک آزمون ناپارامتریک است که اغلب در مورد داده‌های اسمی دو مقوله‌ای یا دوپاسخی مربوط به دو نمونه مرتبط یا همبسته به کار می‌رود. این آزمون به طور ویژه در مواردی به کار می‌رود که می‌خواهیم عملکردهای قبلی یا بعدی موردها (در اینجا نتایج حاصل از دو مدل) را با هم مقایسه کنیم. آزمون مورد استفاده در این پژوهش از نوع اصلاح شده ییتس(ییتس، ۱۹۳۴، ۲۱۷) بوده و آماره آزمون چی اسکوئر می‌باشد. با توجه به محدودیت‌های موجود در ساختار سیستم (وجود اعتبارسنجی ضربدری و تکرار آن در حلقه‌های متوالی) از تمامی داده‌های ورودی به سیستم (آموزش+آزمون+اعتبارسنجی قبل از بالانس داده‌ها) برای اعمال آزمون بهره گرفته شده است. جدول ۶ به صورت دو به دو، معنادار بودن تفاوت الگوریتم‌ها را مورد بررسی قرار می‌دهد. سطح اطمینان در نظر گرفته شده برای بررسی نتایج ۹۵ درصد می‌باشد.

جدول ۶- اثر معنادار بودن تفاوت الگوریتم‌ها

مقایسه الگوریتم‌ها/ صنعت	P-value McNemar's test output	شبکه عصبی	ماشین بردار پشتیبان	یادگیری تشدید شده (بدون هسته)	یادگیری تشدید شده (با بهره‌گیری از هسته)
صنعت ۱	بدون حذف داده‌های پرت	۰/۰۰۰ ۱۶/۲۰۷**	۰/۰۰۰ ۱۱/۵۱	۰/۰۰۰ ۱۲۹/۰۲۱	۰/۰۰۰ ۳۲/۳۲۵
	حذف داده‌های پرت	۰/۱۴۹ ۲/۰۸۱	۰/۰۰۴ ۷/۹۵۹	۰/۰۰۰ ۸۶/۵۶	۰/۲۰۲ ۱/۶۲۶
صنعت ۲	بدون حذف داده‌های پرت	۰/۰۰۰ ۱۵/۰۷۷	۰/۳۶۲۲ ۰/۸۳	۰/۰۰۰ ۶۸/۳۱۲	۰/۰۰۰ ۱۲/۲۰۲
	حذف داده‌های پرت	۰/۰۰۰ ۱۱/۶۳۲	۰/۰۰۰ ۱۴/۶۴۷	۰/۰۱۵ ۵/۸۸۹	۰/۰۰۰ ۲۷/۲۳۲
صنعت ۳	بدون حذف داده‌های پرت	۰/۰۰۰ ۲۳/۹۵۲	۰/۰۰۰ ۵۶/۶۴۶	۰/۰۰۰ ۱۷/۸۹۸	۰/۰۰۰ ۱۲۱/۹۶۱
	حذف داده‌های پرت	۰/۰۰۰ ۷۵/۸۱۳	۰/۰۰۰ ۲۸/۸۱۹	۰/۰۰۰ ۶۴/۵۱	۰/۰۰۰ ۵۱/۳۶
صنعت ۴	بدون حذف داده‌های پرت	۰/۱۴۶ ۲/۱۰۶	۰/۰۰۰ ۵۵/۹۷۸	۰/۰۰۰ ۹۸/۵۱۲	۰/۰۰۰ ۱۱/۱۲۸
	حذف داده‌های پرت	۰/۰۰۰ ۳۹/۳۱۳	۰/۰۰۰ ۲۰/۴۳۲	۰/۰۰۰ ۴۱/۸۵۴	۰/۰۰۰ ۱۲۸/۲۵۴

منبع: یافته‌های پژوهشگر

تنها در صنعت مواد غذایی در شرایط استفاده از شبکه عصبی، استفاده از الگوریتم فرهنگی دقت بالاتری را در مقایسه با استفاده از الگوریتم هارمونی نتیجه داده است (نتیجه فوق با ستاره نشان داده شده است). به طور کلی از مجموعه ۳۲ حالت مقایسه شده در جدول ۶ اختلاف (ضعیف‌تر بودن الگوریتم فرهنگی) ۲۸ مورد معنادار و تنها اختلاف ۴ مورد بی‌معنی می‌باشد. با توجه به نتایج می‌توان نتیجه گرفت که در مجموع استفاده از الگوریتم فرهنگی نسبت به الگوریتم‌های رقابت استعماری و جستجوی هارمونی از دقت نهایی کمتری برخوردار بوده و این تفاوت دقت از لحاظ آماری نیز معنادار می‌باشد. به این دلیل الگوریتم بهینه‌سازی فرهنگی از ادامه بررسی‌ها حذف خواهد شد. لذا در مقایسه میان الگوریتم فرهنگی با الگوریتم‌های رقابت استعماری و جستجوی هارمونی فرضیه سوم مورد قبول قرار می‌گیرد.

بررسی حساسیت مدل‌های طبقه‌بندی به داده‌های پرت

جدول ۷ میزان بهبود حاصل شده در استفاده از فرآیند حذف داده‌های پرت در مقایسه با عدم استفاده از حذف داده‌های پرت را ارائه می‌دهد. لازم به ذکر است هدف یافتن مدل (یا مجموعه مدل‌هایی) می‌باشد که کمترین حساسیت را به حذف داده‌های پرت و بالاترین دقت را در حالت عدم حذف داده‌های پرت دارا می‌باشند. جدول ۷ موارد فوق را در دو صنعت موادغذایی و نساجی مورد بررسی قرار می‌دهد.

اولین مسأله مشهود در جدول ۷ حساسیت بالای شبکه عصبی به داده‌های پرت می‌باشد. علاوه بر آن شبکه عصبی با توجه به ستون دقت کلی (قبل از حذف) در شرایط عدم استفاده از حذف داده‌های پرت دارای پایین‌ترین دقت میانگین می‌باشد. با توجه به موارد ذکر شده استفاده از شبکه عصبی به عنوان هسته سیستم از فرآیند بررسی حذف شده و انتخاب به "ماشین بردار پشتیبان"، "یادگیری تشدید شده (بدون هسته)" و "یادگیری تشدید شده (با هسته)" تقلیل می‌یابد. پس از شبکه عصبی، یادگیری تشدید شده (با هسته) دارای بیشترین افزایش دقت پس از حذف داده‌های پرت می‌باشد. علاوه بر آن نتایج نشان می‌دهد که در شرایط عدم استفاده از فرآیند حذف داده‌های پرت، یادگیری تشدید شده (با هسته) پس از شبکه عصبی دارای کمترین میانگین دقت کل می‌باشد. می‌توان نتیجه گرفت که از لحاظ کمترین حساسیت به داده‌های پرت مدل‌های طبقه‌بندی "ماشین بردار پشتیبان" و "یادگیری تشدید شده (بدون هسته)" مناسب‌ترین مدل‌های طبقه‌بندی قابل استفاده می‌باشند. نتایج فوق در سه ستون دقت کلی (قبل از حذف)، دقت کلی (پس از حذف) و میانگین بهبود کل در جدول ۷ قابل

مشاهده است. شبکه عصبی و یادگیری تشدید شده دارای پایین‌ترین دقت کلی قبل از حذف بوده و با حذف داده‌های پرت دقت به طور چشمگیری بهبود یافته است. در مقابل ماشین بردار پشتیبان و یادگیری تشدید شده (بدون هسته) با شرایط عکس روبرو بوده و این دو مدل طبقه‌بندی به عنوان مدل‌های برتر در این بخش انتخاب می‌شوند.

جدول ۷- بررسی حساسیت مدل‌های طبقه‌بندی به داده‌های پرت

روش / الگوریتم (درصد)	صنعت موادغذایی			صنعت نساجی			میانگین بهبود کل	دقت کلی (قبل از حذف)	دقت کلی (پس از حذف)
	الگوریتم رقابت استعماری	جستجوی هارمونی	میانگین بهبود در صنعت	الگوریتم رقابت استعماری	جستجوی هارمونی	میانگین بهبود در صنعت			
شبکه عصبی	+۱۰/۲۳	+۱۱/۵۴	+۱۰/۸۸	+۱۶/۸۲	+۱۶/۰۵	+۱۶/۴۳	+۱۳/۶۵۵	۷۰/۶۲	۸۴/۲۸
ماشین بردار پشتیبان	-۰/۴۲	+۰/۷۷	+۰/۱۷	+۶/۱۳	+۶/۲۶	+۶/۱۹۵	+۳/۱۸۲	۸۲/۴۰	۸۵/۵۹
یادگیری تشدید شده (بدون هسته)	+۲/۲	+۰/۹۲	+۱/۵۶	+۱/۴۶	+۵/۳۹	+۳/۴۲۵	+۲/۴۹۲	۸۱/۷۲	۸۴/۲۱
یادگیری تشدید شده (با هسته)	+۳/۶۹	+۵/۰۷	+۴/۳۸	+۱۳/۴۳	+۴/۱	+۸/۷۶۵	+۶/۵۷۲	۷۷/۳۱	۸۳/۸۹

منبع: یافته‌های پژوهشگر

انتخاب مدل طبقه‌بندی برتر

در ابتدا ذکر این نکته ضروری به نظر می‌رسد که مدل طبقه‌بندی ماشین بردار پشتیبان دارای طولانی‌ترین زمان و مدل طبقه‌بندی یادگیری تشدیدشده (بدون هسته) دارای کمترین زمان فرآیند بهینه‌سازی می‌باشند (به طور میانگین ماشین بردار پشتیبان در حدود ۱۴ تا ۲۰ ساعت و یادگیری تشدید شده-بدون هسته ۰/۵ تا ۱/۵ ساعت). از لحاظ زمان پردازش برنده یادگیری تشدید شده (بدون هسته) می‌باشد. اما در مقایسه دقت‌های حاصل در الگوریتم‌های بهینه‌سازی جستجوی هارمونی و رقابت استعماری در اکثر موارد ماشین بردار پشتیبان دارای دقت کل نهایی بالاتری از یادگیری تشدید شده (بدون هسته) می‌باشد. بررسی تفاوت‌های ذکر شده از لحاظ آماری در جدول ۸ ارائه شده است. نتایج بالای قطر (دو مربع خالی در هر یک از صنایع به عنوان قطر در نظر گرفته شده است) مقایسه آزمون آماری در حالت استفاده از الگوریتم بهینه‌سازی هارمونی و نتایج پایین قطر، آزمون آماری در حالت استفاده از الگوریتم رقابت استعماری را نشان می‌دهد.

از میان چهار ترکیب بررسی شده در جدول ۸، تنها در یک حالت (در جدول با ستاره نشان داده شده است) دقت‌های حاصل از مدل طبقه‌بندی یادگیری تشدید شده (بدون هسته) از ماشین بردار پشتیبان بیشتر بوده (با توجه به نتیجه آزمون آماری مکنمار تفاوت فوق معنادار نیز می‌باشد) و در سایر حالات دقت تشخیص ماشین بردار پشتیبان بالاتر بوده و این تفاوت‌ها نیز معنادار می‌باشند. در بخش‌های پیشین به این مسأله اشاره شد که ماشین بردار پشتیبان دارای خروجی‌های احتمالی بوده و با بهره‌گیری از این مدل می‌توان احتمال ورشکستگی هر یک از نمونه‌ها (شرکت‌ها) را در سال‌های آتی تعیین کرد. در مقابل مدل طبقه‌بندی یادگیری تشدید شده بررسی شده در این پژوهش تنها دارای نتایج دودویی (ورشکسته یا سالم) می‌باشد.

جدول ۸- مقایسه آماری مدل‌های طبقه‌بندی ماشین بردار پشتیبان و یادگیری تشدید شده (بدون هسته)

P-value McNemar's test output	موادغذایی		نساجی	
	ماشین بردار پشتیبان	یادگیری تشدید شده (بدون هسته)	ماشین بردار پشتیبان	یادگیری تشدید شده (بدون هسته)
ماشین بردار پشتیبان		۰/۰۰۰ ۱۰۴/۲۵۴۸		۰/۰۰۰ ۱۲۷/۲۹۱۲
یادگیری تشدید شده (بدون هسته)	۰/۰۰۰ ۱۲/۸۸۰۰		۰/۰۰۰ ۳۹/۹۴۲۴**	

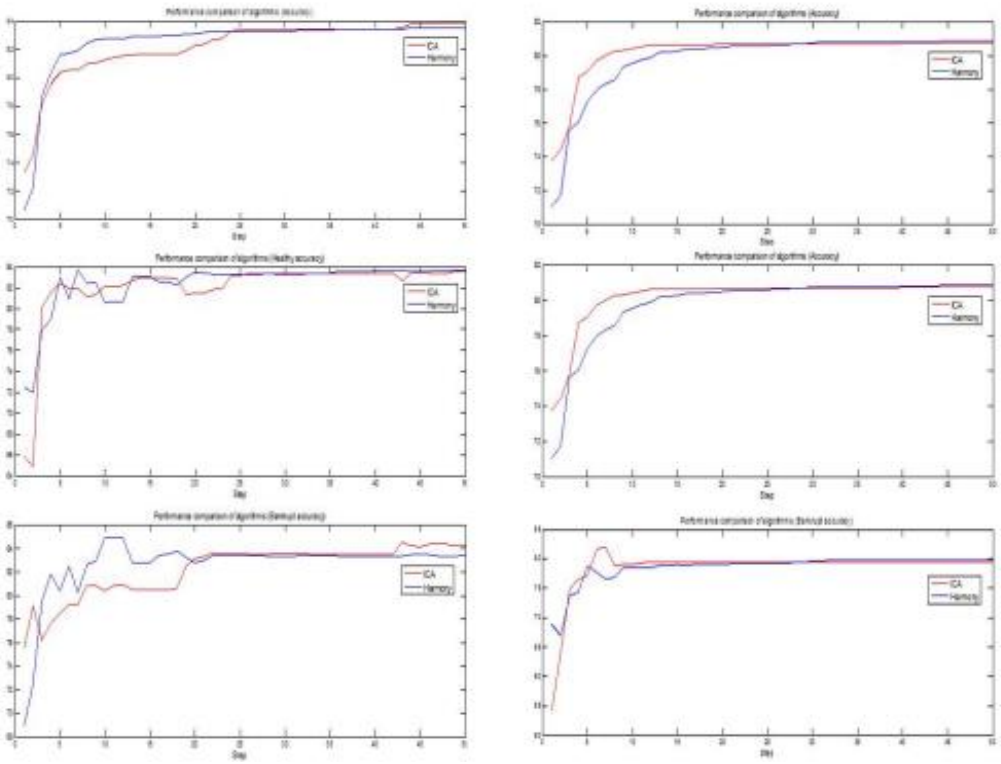
منبع: یافته‌های پژوهشگر

در استفاده اجرایی از سیستم طراحی شده تعیین احتمالات ورشکستگی از اهمیت بالایی برخوردار می‌باشد. به این دلیل علی‌رغم سرعت بالای مدل طبقه‌بندی یادگیری تشدید شده (بدون هسته) نسبت به ماشین بردار پشتیبان، با توجه به دقت‌های بالاتر حاصل از ماشین بردار پشتیبان و همچنین احتمالی بودن ساختار آن، مدل طبقه‌بندی ماشین بردار پشتیبان به عنوان مدل برتر در این بخش انتخاب خواهد شد. تا این مرحله از بررسی، مدل طبقه بندی ماشین بردار پشتیبان در شرایط عدم استفاده از فرآیند حذف داده‌های پرت با بهره‌گیری از الگوریتم‌های بهینه سازی جستجوی هارمونی و رقابت استعماری به عنوان ترکیب بهینه انتخاب شده‌اند. مطابق با نتایج فوق فرضیه دوم مبنی بر برتری ماشین بردار پشتیبان نسبت به سایر مدل‌های طبقه بندی در ترکیب با الگوریتم‌های بهینه سازی مختلف مورد تأیید قرار گرفته است.

انتخاب الگوریتم بهینه‌سازی برتر

تمرکز این بخش بر روی انتخاب الگوریتم بهینه‌سازی برتر از میان الگوریتم‌های بهینه سازی جستجوی هارمونی و رقابت استعماری می‌باشد. بدین‌منظور در ابتدا الگوریتم‌های فوق در ترکیب با مدل طبقه‌بندی ماشین بردار پشتیبان در شرایط عدم حذف داده‌های پرت ۳ بار تکرار شده و از نتایج میانگین گرفته شده است. علاوه بر آن برای این آزمون تعداد قدم‌های سیستم برابر با ۵۰ قدم ثابت (بدون استفاده از ویژگی توقف) در نظر گرفته شده است.

نمودارهای سمت چپ مربوط به صنعت موادغذایی و نمودارهای سمت راست مربوط به صنعت نساجی می‌باشند. از بالا به پایین نیز نمودارها نشان دهنده دقت کل، دقت تشخیص شرکت‌های سالم و دقت تشخیص شرکت‌های ورشکسته می‌باشند. مشاهده می‌شود که در صنعت مواد غذایی تا حدودی الگوریتم بهینه‌سازی جستجوی هارمونی برتری داشته اما الگوریتم رقابت استعماری در پایان به دقت بالاتری دست یافته است. در مقابل در صنعت نساجی فرآیند بهینه‌سازی الگوریتم رقابت استعماری نسبت به الگوریتم جستجوی هارمونی مناسب‌تر بوده اما الگوریتم جستجوی هارمونی در مراحل پایانی به دقت نسبتاً بهتری دست یافته است. نتایج نهایی هر دو صنعت با میانگین‌گیری نتایج خروجی مدل‌ها مورد آزمون آماری مکنمار قرار گرفته است. در صنعت موادغذایی p-value آزمون مکنمار برابر با ۰/۷۲۵ و در صنعت نساجی p-value برابر با ۰/۰۰ می‌باشد.



شکل ۳- مقایسه فرآیند بهینه‌سازی الگوریتم‌های رقابت استعماری و جستجوی هارمونی در صنایع موادغذایی و نساجی

منبع: یافته‌های پژوهشگر

با توجه به نمودارها می‌توان نتیجه گرفت که علی‌رغم کم بودن مقدار تفاوت دقت در صنعت نساجی، این تفاوت در سطح اطمینان ۹۵ درصد معنادار اما در صنعت موادغذایی با وجود زیاد بودن تفاوت، این تفاوت از لحاظ آماری بی‌معنی می‌باشد. از لحاظ سرعت رسیدن به مقدار بهینه و به طور کلی فرآیند بهینه‌سازی نیز امکان ارائه نظر قطعی برقرار نمی‌باشد. در صنعت موادغذایی سرعت رسیدن به مقدار بهینه الگوریتم جستجوی هارمونی بیشتر اما در صنعت نساجی سرعت رسیدن به مقدار بهینه الگوریتم رقابت استعماری بالاتر است. علاوه بر موارد فوق ذکر این نکته ضروری به نظر می‌رسد که در شرایط یکسان زمان بهینه‌سازی توسط الگوریتم رقابت استعماری به طور میانگین ۱۲ ساعت و زمان بهینه‌سازی توسط الگوریتم جستجوی هارمونی به طور میانگین ۹ ساعت بوده است. با توجه به توضیحات ذکر شده امکان نتیجه‌گیری نهایی برقرار نبوده و به طور کلی دو الگوریتم ارائه شده نتایج نسبتاً مشابهی را دارا می‌باشند. برای دستیابی به نتایجی با قابلیت اطمینان بیشتر در این بخش لازم است در توسعه مطالعه جاری مدل طراحی شده بر روی داده‌های متفاوت و با حجم نمونه‌ای بالاتر مورد آزمون قرار گیرد. مطابق با نتایج این بخش فرضیه سوم مبنی بر وجود تمایز میان دو الگوریتم بهینه سازی رقابت استعماری و جستجوی هارمونی رد می‌شود.

نسبت‌ها و متغیرهای مالی برتر

به دلیل نزدیک بودن نتایج الگوریتم‌های رقابت استعماری و جستجوی هارمونی، از مجموعه نتایج و تکرارهای دو الگوریتم (در مجموع ۶ تکرار) برای تهیه جدول ۹ استفاده شده است.

جدول ۹- متغیرهای منتخب خروجی سیستم در صنعت موادغذایی و نساجی

صنعت موادغذایی		صنعت نساجی	
نسبت مالی	تعداد تکرار	نسبت مالی	تعداد تکرار
سود انباشته به کل دارایی‌ها	۶	سود انباشته به کل دارایی‌ها	۶
سود (زیان) عملیاتی به کل دارایی‌ها	۵	سود (زیان) پس از کسر مالیات به جمع حقوق صاحبان سهام	۶
کل بدهی‌ها به کل بدهی‌ها+کل حقوق صاحبان سهام	۴	سود (زیان) پس از کسر مالیات به کل دارایی‌ها	۵
سود (زیان) قبل از بهره و مالیات به کل دارایی‌ها	۴	کل دارایی‌ها به حقوق صاحبان سهام	۵
سود (زیان) پس از کسر مالیات به کل دارایی‌ها	۳	سود (زیان) قبل از بهره و مالیات به کل دارایی‌ها	۲
کل بدهی‌ها به حقوق صاحبان سهام	۳	کل بدهی‌ها به کل بدهی‌ها+کل حقوق صاحبان سهام	۲
کل دارایی‌ها به حقوق صاحبان سهام	۳	کل بدهی‌ها به کل دارایی‌ها	۱
کل بدهی‌ها به کل دارایی‌ها	۲	سود (زیان) ناخالص فروش به کل دارایی‌ها	۱
سود (زیان) پس از کسر مالیات به جمع حقوق صاحبان سهام	۱	سود (زیان) عملیاتی به کل دارایی‌ها	۱

منبع: یافته‌های پژوهشگر

مشاهده می‌شود که در دو صنعت مورد بررسی نسبت مالی "سود انباشته به کل دارایی‌ها" بالاترین تکرار را در پیش‌بینی ورشکستگی به خود اختصاص داده است. به این دلیل نسبت فوق به عنوان نسبت برتر برای پیش بینی ورشکستگی انتخاب خواهد شد. لازم به ذکر است تکرارهای بیشتر سیستم اطمینان به نتایج ذکر شده را افزایش خواهد داد. سایر نسبت‌های مالی منتخب سیستم نیز در جدول ۹ به تفکیک صنعت ارائه شده است.

نتایج مسأله مهم دیگری را نیز نمایان می‌سازد. مشاهده می‌شود که در هر یک از صنایع نسبت‌های مالی نسبتاً متفاوتی به عنوان نسبت‌های مالی با بالاترین تکرار انتخاب شده‌اند. این مسأله نشان می‌دهد که امکان در نظر گرفتن یک مجموعه مشخص نسبت مالی به عنوان ورودی مدل همانگونه که در اکثر مطالعات انجام شده اینگونه در نظر گرفته شده است برقرار نبوده و در هر صنعت با توجه به ویژگی‌ها و ساختار، ترکیب ورودی‌ها با صنعت دیگر متفاوت می‌باشد. در نتیجه می‌توان یافته‌های این بخش را به عنوان یکی از دلایل مهم برتری سیستم طراحی شده در این پژوهش نسبت به ساختارهای مختلف ارائه شده در مطالعات داخلی و بین‌المللی عنوان کرد.

نتیجه‌گیری و بحث

در پژوهش حاضر برای تجزیه‌وتحلیل داده‌ها در بخش کیفی از نرم‌افزار MAX QDA استفاده شد. از روش‌های آمار توصیفی و آمار استنباطی با استفاده از نرم‌افزار SPSS و نرم‌افزار Smart PLS استفاده می‌شود. در مرحله اول نتایج کدگذاری حاصل از مصاحبه‌های تحقیق مشخص شدند دراین بین نزدیک به ۳۰ کد استخراج شدند. در کدگذاری محوری، مقوله‌های جدا از هم در چهارچوبی معنادار در کنار یکدیگر قرار می‌گیرند و روابط میان آن‌ها، به‌ویژه رابطه مقوله محوری با سایر مقوله‌ها، مشخص می‌شود. کدگذاری محوری، منجر به ایجاد گروه‌ها و مقوله‌ها می‌شود. تمامی کدهای مشابه در گروه خاص خود قرار می‌گیرند. در این راستا، تمامی کدهای ایجاد شده دوباره بازبینی شده و با متون مقایسه می‌شود تا مطلبی از قلم نیفتد. نتایج حاصل از کدگذاری محوری در جدول بالا نشان‌داده شده است. مشاهده می‌گردد ۳۰ کد اولیه در قالب ۶ مقوله دسته‌بندی شده‌اند. در مرحله بعد شاخص‌های شناسایی شده با روش دلفی فازی صحت سنجی شدند. نتایج حاصل از دلفی نشان داد که همهٔ شاخص‌ها و کدهای مربوطه به تأیید خبرگان رسیدند.

نتایج این پژوهش نشان‌دهنده این موضوع است که استفاده از ساختارهای جدید مانند سیستم‌های هوشمند ترکیبی مبتنی بر مدل‌های داده کاوی توانایی بالایی در تشخیص ورشکستگی شرکت‌ها در سطح کشور دارد. در این تحقیق داده‌های مورد استفاده به دو صنعت مواد غذایی و نساجی تعمیم یافته و داده‌های پژوهش محدود به داده‌های شرکت‌های بورس اوراق بهادار نمی‌باشد. این مساله باعث می‌شود تا میزان اتکا به نتایج و امکان بهره‌گیری از مدل‌های طراحی شده در سطح کشور به صورت عملی افزایش یابد.

از دیدگاه نتایج، بررسی‌ها نشان‌دهنده برتری عملکرد ترکیب ماشین بردار پشتیبان با الگوریتم‌های بهینه‌سازی جستجوی هارمونی و رقابت استعماری در شرایط عدم حذف داده‌های پرت می‌باشد. بعلاوه مشاهده شد که در دو صنعت موادغذایی و نساجی متغیرهای مالی مختلفی به عنوان متغیرهای اثرگذار بر فرآیند پیش‌بینی ورشکستگی استخراج شده‌اند که این مساله نشان از یکسان نبودن رفتار و دلایل ورشکستگی در دو صنعت فوق دارد. همچنین می‌توان اینگونه نتیجه گرفت که لازم است در بعد عملی مدل طراحی شده در هر صنعت به صورت مجزا مورد آموزش قرار گیرد.

از دیدگاه توسعه پژوهش جاری با توجه به اینکه بیشتر بخش‌ها و ترکیبات سیستم طراحی شده جدید و ابتکاری می‌باشد امکان بررسی بیشتر پارامترها و ارائه ساختارهای متفاوت در سیستم فوق امکان‌پذیر است. همچنین ایجاد گزارش‌های ملموس بررسی ساختار پیش‌بینی ورشکستگی و استخراج اثر هر یک از مجموعه متغیرهای جدول ۹ بر آن (اثر مثبت، منفی و همچنین شدت اثر) به منظور شناخت دقیق‌تر این ساختار به تفکیک صنایع از مسائل مهم قابل بررسی در آینده است. بعلاوه از دیدگاه نرم افزاری، طراحی و پیاده‌سازی یک چارچوب نرم‌افزاری جامع به منظور پیش‌بینی ورشکستگی شرکت‌ها جهت سیاست‌گذاری‌های خرد و کلان بر مبنای نتایج این پژوهش نیز قابل بررسی می‌باشد.

از دیدگاه مدل‌سازی نیز می‌توان به موارد مختلفی اشاره کرد. یکی از موارد گسترش مجموعه مدل‌ها به بررسی سایر تقسیم‌بندی‌های داده مانند شرکت‌های موجود در استان‌ها، شهرستان‌ها و سایر تقسیم‌بندی‌های جغرافیایی می‌باشد. همچنین افزودن الگوریتم‌های بهینه‌سازی جدید به سیستم و بررسی آن‌ها و افزودن مدل‌های طبقه‌بندی دیگر به سیستم از جمله مدل‌های قاعده‌محور و مجموعه مدل‌های اقتصادسنجی باید مورد توجه قرار گیرد.

از دیدگاه پیش‌پردازش نیز لازم است به مبحث افزودن فرآیندهای پیش‌پردازش مبتنی بر کاهش ابعاد داده‌ها و سایر روش‌های حذف داده‌های پرت توجه ویژه‌ای شود. از نظر پردازش و ابزارهای مرتبط با آن انتقال محاسبات الگوریتم‌های بهینه‌سازی به پردازش موازی به منظور افزایش سرعت سیستم و بهره‌گیری از پردازش مبتنی بر هسته‌های گرافیکی از نکات مهم قابل بررسی می‌باشد.

فهرست منابع

ابزری، مهدی، منجمی، امیرحسن و مروه بخشایش. (۱۳۹۰)پیش بینی ورشکستگی شرکتهای پذیرفته شده در بورس اوراق بهادارتهران با استفاده از شبکههای عصبی باز جریانی. پایاننامه کارشناسی ارشد، دانشکده مدیریت دانشگاه اصفهان، اصفهان .

ابزری، مهدی، میرشمس شهشهانی، مرتضی و مهدی رسولزاده. (۱۳۸۰)بررسی کاربرد مدل آلتمن برای تعیین وضعیت ورشکستگی شرکتهای پذیرفته شده در بورس اوراق بهادار تهران)صنایع نساجی و ساخت فلزات اساسی(.پایاننامه کارشناسی ارشد، دانشکده مدیریت دانشگاه اصفهان، اصفهان .

امینی، پیمان. (۱۳۸۵) بررسی امکان سنجی استفاده از مدل فولمر برای تخمین ورشکستگی شرکتهای پذیرفته شده در بورس اوراق بهادار تهران. پایاننامه کارشناسی ارشد، رشته حسابداری دانشگاه تربیت مدرس، تهران .

پورحیدری، امید و مهدی کوپائی حاجی. (۱۳۸۹)پیش‌بینی بحران مالی شرکتها با استفاده از مدل مبتنی بر تابع تفکیکی خطی.مجله پژوهشهای حسابداری مالی، دوره ۲ ،ش ۱ ،صص ۴۶-۳۳ .

پورزمانی، زهرا، توانگر حمزه کالیی، افسانه و آوا کیارسی. (۱۳۸۹) بررسی کارایی الگوی لوجیت و تحلیل تمایزی چند متغیره در پیش‌بینی وضعیت مالی شرکتهای بورس اوراق بهادار تهران. پژوهشنامه حسابداری مالی و حسابرسی، دوره ۲ ،ش ۵ :۱۲۲-۹۴ .

پورزمانی، زهرا، جهانشاد، آزیتا و شهرام عین قالیی. (۱۳۸۸) پیش‌بینی وضعیت مالی و اقتصادی شرکتها با استفاده از نسبتهای مالی مبتنی بر سودآوری، جریانهای نقدی و رشد. پژوهشنامه حسابداری مالی و حسابرسی، دوره ۱، ش ۳ ، صص ۱۱۵- ۹۳ .

حاجی هاشم، مسعود. امیرحسینی، زهرا. (۱۳۹۸)پیش‌بینی ورشکستگی و راهبری شرکت ها: دیدگاه نسبت های مالی. دانش حسابداری و حسابرسی مدیریت، ۸(۳۰ ،۱)۲۰۱-۲۳۰

Bruynseels, L., & Willekens, M. (2012). The effect of strategic and operating turn around initiatives on audit reporting for distressed companies. Accounting, Organizations and Society, 27(4), 223-241.

Elsevier (2015). Retrieved from: www.elsevier.com (Accessed 20 March 2018).

Hillegeist, S., Keating, E., Cram, D., & Lundstedt, K. (2004). Assessing the Probability of Bankruptcy. Review of

Accounting Studies, 9(1), 5-34.

Jackson, R. & Wood, A. (2013). The performance of insolvency prediction and credit risk models in the UK: A comparative study. The British Accounting Review, 45(3), 183-202.

Kim, S. (2011). Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. The Service Industries Journal, 31(3), 441-468.

Lepetit, L., & Strobel, F. (2013). Bank insolvency risk and time-varying Z-score measures. Journal of International

Financial Markets, Institutions and Money, 25, 73-87.

Li, H., & Sun, J. (2012). Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples – Evidence from the Chinese hotel industry. Tourism Management, 33(3), 622-634.

Lin, F., Yeh, C., & Lee, M. (2011). The use of hybrid manifold learning and support vector machines in the prediction of business failure. Knowledge-Based Systems, 24(1), 95-101.

Pan, W. (2012). A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example.

Knowledge-Based Systems, 26, 69-74.

Sun, J., Li, H., Huang, Q., & He, K. (2014). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. Knowledge-Based Systems, 57, 41-56.

Tserng, H., Chen, P., Huang, W., Lei, M., & Tran, Q. (2014). Prediction of default probability for construction firms using the logit model. Journal of Civil Engineering and Management, 20(2), 247-255.

Wang, L., & Wu, C. (2017). Business failure prediction based on two-stage selective ensemble with manifold learning algorithm and kernel-based fuzzy self-organizing map. Knowledge-Based Systems, 121, 99-110.

Xiao, Z., Yang, X., Pang, Y., & Dang, X. (2012). The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster–Shafer evidence theory. Knowledge-Based Systems, 26, 196-206

Accounting Knowledge & Management Auditing
Vol. 14/ No. 55/ Autumn 2024

Predictive modeling of bankruptcy with an emphasis on modern measurement methods such as Neural Networks and Support Vector Machines

Mohammadreza Nosrat Nezemi
PhD candidate in Financial Engineering, Financial Engineering Department, Rasht branch, Islamic Azad University, Rasht, Iran.
Mrnn1350@gmail.com

Fazel Mohammadi Nodeh
Assistant Professor, Department of Management, lahijan Branch, Islamic Azad University, Lahijan, Iran
(Corresponding author).
Mnfazel2@gmail.com

Sina Kherad Yar
Associate professor, Department of Accounting, Rasht branch, Islamic Azad University, Rasht, Iran.
Kheradyar@iaurasht.ac.ir

Abstract
The purpose of this study is to develop a predictive modeling of bankruptcy, focusing on modern measurement methods such as Neural Networks and Support Vector Machines, specifically for listed companies in the Tehran Stock Exchange Market. The findings of this research indicate that the utilization of new structures, such as hybrid intelligent systems based on data mining models, holds significant potential in detecting company bankruptcies at the national level. It should be noted that the data used in this research was generalized to two industries, namely food and textile, and was not limited to data solely from stock-exchange-listed companies. It is observed that in the two industries under study, the financial ratio of "Accumulated Profits to Total Assets" has the highest recurrence in predicting bankruptcies. For this reason, the aforementioned ratio will be chosen as the top ratio for predicting bankruptcy. The results reveal another important issue. It is observed that in each industry, relatively different financial ratios have been selected as the financial ratios with the highest recurrence. This demonstrates that the possibility of considering a specific set of financial ratios as model inputs, as commonly done in most studies, is not valid. In each industry, the combination of inputs varies based on its characteristics and structure. Therefore, the findings of this section can be considered one of the important reasons for the superiority of the designed system in this research compared to various structures presented in domestic and international studies.
Keywords: Bankruptcy, New methods of measurement, Artificial Neural Network, Support Vector Machine