Archive of SID

پژوهش‌های حسابداری مالی و حسابرسی
سال دهم / شماره ۳۷ / بهار ۱۳۹۷
صفحه ۱۵۹ تا ۱۹۳

پیش‌بینی ورشکستگی شرکت‌ها مبتنی بر سیستم‌های هوشمند ترکیبی

مهدی غضنفری[۱]
اقبال رحیمی کیا[۲]
علی عسکری[۳]

تاریخ دریافت: ۱۳۹۶/۱۰/۱۹ تاریخ پذیرش: ۱۳۹۶/۱۲/۲۱

چکیده

با توجه به شرایط رقابتی اقتصاد کشورها و بحران‌های اقتصادی ایجاد شده در سطح بین‌المللی و داخل کشور، نیاز به یک مدل مناسب برای پیش‌بینی ورشکستگی شرکت‌های کشور احساس می‌شود. تصمیم‌گیران کلان اقتصادی، سازمان‌های اقتصادی کشور و سیستم بانکی با استفاده از این مدل‌ها توانایی اتخاذ تصمیمات دقیق‌تر و با عوارض کمتری را دارا خواهند بود. همچنین مدل‌های فوق در سطح خرد نیز برای تصمیم‌گیری برای سرمایه‌گذاری‌های آتی قابل استفاده می‌باشد. در این تحقیق با پیاده‌سازی یک سیستم منسجم و هوشمند مبتنی بر شبکه عصبی، ماشین‌های بردار پشتیبان و یادگیری تشدید شده و در کنار آن استفاده از الگوریتم‌های بهینه‌سازی رقابت استعماری، الگوریتم فرهنگی و جستجوی هارمونی سعی شده است تا حد امکان نواقص مدل‌های پیشین در سطح بین‌الملل رفع شود. علاوه بر آن با همکاری سازمان امور مالیاتی کشور مقیاس بررسی سیستم به داده‌های کل کشور تعمیم یافته است که بررسی در ابعاد فوق در سطح بین‌الملل منحصر به فرد می‌باشد. تعداد نمونه‌های مورد بررسی در صنعت موادغذایی و نساجی به ترتیب برابر با ۵۸۲۵ و ۴۰۸۹ می‌باشد که با اعمال معیار قانونی ورشکستگی به ترتیب ۹۹۹ و ۸۴۸ نمونه شرایط ورشکستگی را در دو سال مورد بررسی دارا بوده‌اند. نتایج نشان دهنده برتری عملکرد ترکیب ماشین بردار پشتیبان با الگوریتم‌های بهینه‌سازی جستجوی هارمونی و رقابت استعماری در شرایط عدم حذف داده‌های پرت می‌باشد.

واژه‌های کلیدی: پیش بینی ورشکستگی، بحران مالی، داده کاوی، هوش مصنوعی، سیستم هوشمند ترکیبی.

۱- استاد دانشکده مهندسی صنایع، دانشگاه علم و صنعت، تهران، ایران، نویسنده اصلی.mehdi@iust.ac.ir
۲- کارشناس ارشد مهندسی صنایع، دانشگاه علم و صنعت، تهران، ایران، مسئول مکاتبات.erahimi@ut.ac.ir
۳- استادیار گروه اقتصاد ،دانشکده اقتصاد، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران. askari.azadwar@gmail.com

www.SID.ir

Archive of SID

۱- مقدمه

طی گذر تاریخ محققان بسیاری سعی در پیش‌بینی آینده داشته‌اند و این روند در علوم مختلف متبلور بوده است. به طور ویژه در مفاهیم اقتصادی طی چند دهه اخیر توجه زیادی به مفاهیم پیش‌بینی شده و بر پایه آن مطالعات مختلفی در سطوح مختلف انجام پذیرفته است. از میان مجموعه گسترده‌ای از حوزه‌های گوناگون می‌توان به پیش‌بینی قیمت سهام، پیش‌بینی روند شاخص‌های کلان اقتصادی مانند رشد اقتصادی و تورم، پیش‌بینی شاخص‌های مختلف مالی شرکت‌ها و ... اشاره کرد اما بررسی‌ها نشان می‌دهد که پرمخاطب و پرچالش‌ترین حوزه، پیش‌بینی ورشکستگی شرکت‌ها بوده است. علاوه بر آن رفع نا آگاهی از آینده از اصلی‌ترین دغدغه‌های انسان در طول تاریخ بوده است. آدمی همیشه به دنبال آن بوده که از آینده خود آگاه شده و آن را به نحویکه خود می‌خواهد سازمان دهد. در ابتدا چون توان پیش‌بینی صحیح و قابل اطمینان فراهم نبود، انسان متوسل به نیروهای فراطبیعی شد و آنگاه که توانایی عقلایی بیشتری پیدا کرد، درصدد استفاده از این توانایی برآمد و چون علم امکاناتی برای پیش‌بینی‌های او فراهم آورد، از این دستاورد بهره جست (مقدم و همکاران، ۱۳۸۸، ۱۹۵).

از دلایل گسترش توجه به این حوزه می‌توان به این مسأله اشاره کرد که پیش‌بینی ورشکستگی شرکت‌ها بر روی گروه‌های مختلف اقتصادی اثرات مستقیم و غیرمستقیمی دارد. در نگاه کلی بخش کوچکی از اثرات ورشکستگی یک شرکت بیکار شدن نیروی کار، مشکلات مالی برای صاحبان شرکت، بحران مالی بدهکاران و همچنین اثرات منفی ایجاد شده در زنجیره تأمینی[1] است که شرکت فوق بخشی از آن می‌باشد. در واقع این امکان وجود دارد که ورشکستگی به سایر شرکت‌های متصل نیز انتقال یافته و بحران فوق در زمان کوتاهی به کسب و کارهای مختلف سرایت کند. توسعه مالی و حاکم شدن شرایط رقابتی بین شرکت‌ها از طبیعی‌ترین دلایل ورشکستگی شرکت‌ها می‌باشد اما عوامل کلان اقتصادی مانند رکود اقتصادی، تورم، تحریم، تخصیص نامناسب منابع و ... نیز از دلایل برجسته آن است که کل کسب و کار کشور را در بحران فرو برده که یکی از نمودهای این بحران، ورشکستگی شرکت‌ها است.

به قطع بهبود بستر اقتصادی و فضای کسب و کار مهم‌ترین عامل جلوگیری از ورشکستگی است اما در این میان مدل‌های پیش‌بینی ورشکستگی مانند زنگ خطر عمل کرده و افراد و گروه‌های مختلف را برای تصمیم‌گیری صحیح پیش از وقوع حادثه آگاه می‌سازند. در واقع این مدل‌ها چه با استفاده از روش‌های آماری و چه با استفاده از روش‌های هوشمند بر اساس روند رفتار گذشته اقدام به پیش‌بینی ورشکستگی شرکت‌ها در آیندهمی‌کنند و برای این امر مجموعه‌ای از متغیرهای مالی، بازار و کلان اقتصادی را مورد استفاده قرار می‌دهند. با بررسی سیر مطالعات انجام

www.SID.ir

Archive of SID

شده در حوزه پیش‌بینی ورشکستگی شرکت‌ها می‌توان به این نکته پی برد که در بحران‌های اقتصادی جهانی بخصوص در بحران مالی اخیر به دلیل ورشکستگی تعداد زیادی از شرکت‌ها در سطح بین‌الملل نیاز به تجدید نظر و ارائه مدل‌های بهبود یافته احساس شد و محققان بسیاری در چند سال اخیر به توسعه مدل‌های ترکیبی و بهبود یافته پرداختند. به طور کلی می‌توان اینگونه بیان کرد که با توجه به پیشرفت روز افزون روش‌ها و الگوریتم‌های هوشمند و بهبود بستر نرم افزاری و سخت افزاری رایانه‌ای، به طور پیوسته مدل‌های بهینه‌تر و دقیق‌تری ارائه شده و انتظار می‌رود فرآیند فوق در سال‌های آتی نیز ادامه یابد.

از موارد بسیار مهم و تأثیرگذاری که در طول مطالعات انجام شده کمتر به آن توجه شده است فرآیند انتخاب متغیرهای ورودی[۲] به صورت هوشمند می‌باشد. اکثر مطالعات انجام شده فرآیند انتخاب متغیرهای ورودی و ساخت مدل را به دو مرحله مجزا تقسیم کرده‌اند (لیانگ و همکاران، ۲۰۱۵، ۲۸۹). ساختار مرحله‌ای ذکر شده موجب عدم توجه به ساختار مدل در فرآیند انتخاب متغیرهای تأثیرگذار خواهد شد. در این پژوهش با مبنا قرار دادن مدل طبقه‌بندی شبکه عصبی پرسپترون چندلایه[۳]، ماشین بردار پشتیان[۴]، یادگیری تشدید شده[۵] (بدون بهره گیری از هسته[۶]) و یادگیری تشدید شده (با بهره گیری از هسته) در کنار الگوریتم‌های بهینه سازی رقابت استعماری[۷]، جستجوی هارمونی[۸] و فرهنگی[۹] پارامترهای مدل طبقه‌بندی در کنار ترکیب مناسب متغیرهای ورودی به منظور دستیابی به دقت طبقه‌بندی حداکثری مورد استفاده قرار خواهد گرفت. علاوه بر آن بهره‌گیری از اعتبارسنجی ضربدری ۱۰ بخشی[۱۰] و تکرار آن بر روی هسته‌های منطقی[۱۱] واحد پردازشی مرکزی[۱۲] بوسیله استفاده از پردازش موازی[۱۳] از قرارگیری جواب نهایی الگوریتم بهینه سازی در بهینه محلی[۱۴] جلوگیری خواهد کرد. در این پژوهش به دلیل نیاز شبکه عصبی به سه مجموعه داده آموزش، آزمون و اعتبارسنجی از روش ویژه ای به‌منظور اعتبارسنجی ضربدری بهره گرفته شده است. جواب نهایی سیستم پس از حذف نتایج پرت[۱۵] میانگین مجموعه‌ای از پاسخ‌ها (تعداد حلقه‌های سیستم ضرب در تعداد بخش‌ها در اعتبارسنجی ضربدری) را شامل خواهد شد. همچنین در ادامه از مراحل بهینه‌سازی الگوریتم مورد نظر با نام قدم[۱۶] یاد خواهد شد. بعلاوه دو معیار برای توقف سیستم در نظر گرفته شده است. اولین معیار توقف سیستم بر مبنای حداکثر قدم‌های الگوریتم بهینه‌سازی بوده و معیار دوم توقف سیستم پس از عدم بهبود دقت کلی در تعداد قدم‌های مشخص است.

از موارد مهم بررسی شده در این پژوهش اثر حذف داده‌های پرت بر روی دقت نهایی سیستم است. به منظور بررسی این اثر از روش حذف آماری داده‌های پرت بر مبنای ۲ انحراف معیار استفاده شده است (حذف ۲ انحراف معیار در مطالعات مرتبط به تشخیص داده‌های پرت معمول

www.SID.ir

Archive of SID

می‌باشد). علاوه بر آن فرضیه اول تحقیق جاری بهبود نتایج حاصل از ترکیبات در شرایط استفاده از فرآیندهای حذف داده‌های پرت در مقایسه با عدم استفاده از فرآیندهای حذف داده‌های پرت می‌باشد. فرضیه دوم بالاتر بودن قدرت تشخیص (بالاتر بودن دقت تشخیص کلی، دقت تشخیص شرکت‌های ورشکسته و دقت تشخیص شرکت‌های سالم) روش ماشین بردار پشتیبان را در ترکیب با الگوریتم‌های بهینه سازی مختلف نسبت به شبکه عصبی مصنوعی و مدل‌های مبتنی بر یادگیری تشدید شده مورد بررسی قرار می‌دهد. فرضیه سوم نیز وجود تمایز میان نتایج الگوریتم‌های بهینه‌سازی مختلف مورد بررسی را شامل می‌شود (مقایسه دو به دوی الگوریتم‌ها).

بخش‌های مقاله به صورت زیر تنظیم شده است. در ابتدا پیشینه تحقیق مورد بررسی قرار می‌گیرد. در بخش بعد مبانی نظری با معرفی مختصر مدل‌های طبقه بندی و الگوریتم‌های بهینه‌سازی ذکر خواهد شد. سپس هدف، داده‌ها و متغیرها و ساختار سیستم بررسی شده و در بخش پایانی بحث و نتیجه‌گیری ذکر خواهد شد.

۲- پیشینه تحقیق

در این بخش ابتدا مطالعات مبتنی بر هوش مصنوعی[۱۷] و داده کاوی[۱۸] انجام شده در حوزه پیش‌بینی ورشکستگی مورد مرور قرار خواهد گرفت. لازم به ذکر است که تعداد مطالعات انجام شده در سطح بین‌الملل در این حوزه گسترده بوده و امکان مرور تمامی مقالات میسر نمی‌باشد به این دلیل تمرکز اصلی بر مقالات منتشر شده در سال‌های اخیرمی‌باشد.

اعتمادی و همکاران (۲۰۰۹) با استفاده از برنامه‌ریزی ژنتیک[۱۹] به پیش‌بینی ورشکستگی شرکت‌ها در بورس اوراق بهادار پرداخته‌اند. آن‌ها با بررسی ۱۴۴ شرکت پذیرفته شده در بورس اوراق بهادار تهران و اعمال ماده ۱۴۱ قانون تجارت برای تشخیص شرکت‌های دارای بحران مالی به دقتی برابر با ۹۴ درصد برای نمونه‌های آموزش و ۹۰ درصد برای نمونه‌های آزمون دست یافتند. علاوه بر آن با ارائه مدل تحلیل ممیزی چندگانه[۲۰] برتری مدل خود را نسبت به مدل آماری فوق اثبات کرده‌اند. دقت روش تحلیل ممیزی چندگانه برای نمونه‌های آموزش ۷۷ درصد و برای نمونه های آزمون ۷۴ درصد بوده که پس از آزمون آماری، معنادار بودن برتری روش برنامه‌ریزی ژنتیک نسبت به تحلیل ممیزی چند گانه نتیجه شده است.

آقایی و سعیدی (۲۰۰۹) با استفاده از شبکه‌های عصبی بیزینی[۲۱] ورشکستگی شرکت‌های بورس اوراق بهادار تهران را مورد بررسی قرار داده‌اند. دوره زمانی مورد بررسی از سال ۱۹۹۷ تا سال ۲۰۰۷ و تعداد شرکت‌های ورشکسته بر اساس ماده ۱۴۱ قانون تجارت ایران ۷۲ و تعداد شرکت های سالم نیز ۷۲ مورد می‌باشد. ۲۰ نسبت مالی نیز به عنوان ورودی مدل انتخاب شده است.

www.SID.ir

Archive of SID

همچنین مدل طراحی شده با رگرسیون لجستیک[۲۲] مورد مقایسه قرار گرفته است. دقت رگرسیون لجستیک ۹۰ درصد و دقت مدل طراحی شده ۹۳ درصد می‌باشد.

مقاله بعدی توسط رافعی و همکاران (۲۰۱۱) ارائه شده است. آن‌ها با استفاده از سه روش شبکه عصبی مصنوعی، الگوریتم ژنتیک[۲۳] و تحلیل ممیزی چندگانه به بررسی ورشکستگی شرکت ها در بورس اوراق بهادار تهران بر پایه ماده ۱۴۱ قانون تجارت پرداخته‌اند. مدل شبکه عصبی دارای دقت ۹۸/۶ و ۹۶/۳ درصد در آموزش و آزمون، الگوریتم ژنتیک دارای دقت ۹۲/۵ درصد و ۹۱/۵ درصد و تحلیل ممیزیچندگانه نیز دارای دقت ۸۰/۶ و ۷۹/۹ درصد بوده است. نوع شبکه عصبی مورد استفاده پرسپترون چند لایه بوده و برتری آن نسبت به دو روش الگوریتم ژنتیک و تحلیلممیزی چندگانه قابل مشاهده است.

در مطالعه‌ای دیگر آشوری و محمدی (۲۰۱۱) با استفاده از شبکه عصبی پرسپترون چند لایه و انتخاب ویژگی با استفاده از درخت تصمیم‌گیری CART اقدام به پیش‌بینی ورشکستگی شرکت های بورس اوراق بهادار تهران کرده‌اند. بدون استفاده از انتخاب ویژگی و با استفاده از ۲۹ نسبت مالی مختلف، مدل دقت ۸۶/۰۵ در آموزش و ۸۴/۸ درصد در آزمون را دارا بوده است. پس از اعمال روش CART برای انتخاب ویژگی، ورودی‌های مدل از ۲۹ به ۱۲ مورد کاهش یافته و دقت برای مجموعه آموزش به ۸۴/۸۸ و برای مجموعه آزمون به ۸۶/۹۶ درصد رسیده است. البته باید به این نکته توجه شود که برای مقایسه دو مدل، مقایسه سایر عوامل مانند خطای نوع اول، خطای نوع دوم و توابع هزینه‌ای مورد نیازمی‌باشد که در این مقاله خطای نوع اول و دوم دو ساختار مقایسه شده است. مرادی و همکاران (۲۰۱۲) نیز دو روش ماشین‌های بردار پشتیبان و تحلیل ممیزی چندگانه را در شرکت‌های بورس اوراق بهادار تهران و با در نظر گرفتن ۲۰ متغیر ورودی مورد مقایسه قرار داده‌اند. نتایج نشان از عدم برتری آماری مدل‌ها نسبت به یکدیگر دارد.

بربوزا و همکاران (۲۰۱۷) روش‌های مختلف یادگیری ماشینی بردار پشتیبان، Bagging، Boosting و Random Forest را برای پیش‌بینی ورشکستگی یک سال قبل از وقوع مورد ارزیابی قرار داده و نتایج را با روش‌های تحلیل ممیزی خطی[۲۴]، رگرسیون لجستیک و شبکه عصبی مورد مقایسه قرار داده‌اند. داده‌های مورد استفاده شرکت‌های آمریکایی را در دوره زمانی ۱۹۸۵ تا ۲۰۱۳ در بر می‌گیرد. نتایج نشاندهنده بهبود حدودی ۱۰ درصدی دقت تشخیص در مقایسه با روش‌های سنتی می‌باشد. دقت مدل Random Forest، رگرسیون لجستیک و تحلیل ممیزی به‌ترتیب ۸۷، ۶۹ و ۵۰ درصد گزارش شده‌اند. بعلاوه نتایج نشان‌دهنده پایین‌تر بودن دقت ماشین بردار پشتیبان نسبت به سایر مدل‌ها می‌باشد. در بخش بعدی به مطالعات منتشر شده در داخل کشور پرداخته خواهد شد.

www.SID.ir

Archive of SID

راعی و فلاح پور (۱۳۸۳) دو مدل شبکه عصبی مصنوعی و تحلیل ممیزی چندگانه را برای پیش بینی ورشکستگی شرکت‌های بورس اوراق بهادار تهران مورد ارزیابی قرار داده‌اند. پنج متغیر مورد بررسی محققان نسبت جاری، نسبت سود قبل از بهره و مالیات به دارایی‌ها، حقوق صاحبان سهام به بدهی‌ها، سرمایه در گردش به کل دارایی‌ها و نسبت سود قبل از بهره و مالیات به فروش می‌باشد. بر اساس ماده ۱۴۱ قانون تجارت از میان مجموع شرکت‌ها، ۴۰ شرکت دارای شرایط بحران مالی شناخته شده و ۴۰ شرکت نیز به صورت تصادفی از میان شرکت‌هایسالم انتخاب شدند. دوره زمانی مورد بررسی مدل‌های ارائه شده از سال ۱۳۷۳ تا سال ۱۳۸۰ را در برگرفته است. در این پژوهش دقت روش تحلیل ممیزی چندگانه ۹۳/۴ و دقت شبکه عصبی مصنوعی ۹۵/۳ درصد می‌باشد. در پایان با استفاده از آزمون آماری نتیجه شده است که شبکه عصبی مصنوعی به طور معناداری نسبت به روش تحلیل ممیزی چندگانه از دقت بیشتری برخوردار است.

مکیان و کریمی تکلو (۱۳۸۸) بر پایهیک شبکه عصبی پرسپترون چند لایه با دو لایه پنهان اقدام به پیش‌بینی ورشکستگی شرکت‌ها کرده‌اند. بررسی آن‌ها طی سال‌های ۱۳۸۳ تا ۱۳۸۶ انجام شده و ۴۰ شرکت سالم و ۴۰ شرکت ورشکسته را در برمی‌گیرد. معیار ماده ۱۴۱ قانون تجارت نیز برای تمیز دادن شرکت‌های سالم و ورشکسته مورد استفاده قرار گرفته است. همچنین متغیرهای ورودی مدل طراحی شده نسبت جاری، نسبت سود قبل از بهره و مالیات به کل دارایی‌ها، نسبت کل حقوق صاحبان سهام به کل بدهی‌ها، نسبت سرمایه در گردش به کل دارایی‌ها و نسبت سود قبل از هزینه‌های بهره و مالیات به خالص فروش بوده است. بر پایه ترکیب داده‌های ۱، ۲ و ۳ سال قبل از ورشکستگی، مدل طراحی شده به دقت پیش‌بینی ۱۰۰ درصد دست یافته است. همچنین بر پایه داده‌هاییک سال قبل از ورشکستگی نیز این دقت ۱۰۰ درصد بوده است. لازم به ذکر است دلیل دقت حداکثری مدل تعداد کم نمونه‌های آزمون (۵ شرکت سالم و ۵ شرکت ورشکسته) می‌باشد. پیش‌بینی ورشکستگیآن‌ها برای سال ۱۳۸۷ نشان داده است که در آن سال تنها یکی از شرکت‌ها دچار بحران مالی شده و بقیه شرکت‌ها در وضعیت سالم قرار دارند.

قدرتی و معنوی مقدم (۱۳۸۹) به مقایسه مجموعه روش‌های آماری و دو روش هوشمند ژنتیک فرج زاده و ژنتیک مک کی در بورس اوراق بهادار تهران طیسال‌های ۱۳۸۲ تا ۱۳۸۶ پرداخته‌اند. برای تفکیک شرکت‌های سالم و دارای بحران مالی از ماده ۱۴۱ قانون تجارت بهره گرفته شده است. نتایج نشان می‌دهد که الگوی پیش‌بینی بحران مالی زمینسکی، اسپرینگیت، سی ای اسکور، ژنتیک فرج زاده و ژنتیک مک کی توانایی پیش‌بینی تداوم فعالیت شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران را دارا می‌باشند. همچنین مدل‌هایی که بر پایه الگوریتم ژنتیک طراحی شده‌اند نسبت به روش آماری دارای دقت بیشتری می‌باشند. نتایج، دقت کلی ۹۱/۷ درصد را برای

www.SID.ir

Archive of SID

الگوریتم ژنتیک فرج زاده و همچنین دقت ۹۱/۶۵ درصد را برای الگوریتم ژنتیک مک کی نشان می‌دهد. مدل‌های آماری زمینسکی، اسپرینگیت و سی ای اسکور نیز همگی دارای توان کلی پیش‌بینی ۹۰ درصد بوده‌اند.

مکیان و همکاران (۱۳۸۹) با استفاده از داده‌های شرکت‌های استان کرمان در دوره زمانی ۱۳۷۴ تا ۱۳۸۶ اقدام به پیش‌بینی ورشکستگی با استفاده از روش‌های شبکه عصبی، رگرسیون لجستیک و تحلیل ممیزی چندگانه کرده‌اند. نمونه آن‌ها ۴۰ شرکت سالم و ۴۰ شرکت ورشکسته را بر پایه ماده ۱۴۱ قانون تجارت در برگرفته و شرکت‌های سالم براساس اندازه شرکت (ارزش کل دارایی‌ها) با شرکت‌های ورشکسته منطبق شده است. همچنین متغیرهای مورد استفاده آن‌ها نسبت جاری، نسبت سود قبل از هزینه بهره و مالیات به کل دارایی‌ها، نسبت کل حقوق صاحبان سهام به کل بدهی‌ها، نسبت سرمایه در گردش به کل دارایی‌ها و نسبت سود قبل از بهره و مالیات به خالص فروش در نظر گرفته شده است. بر پایه داده‌های ۱ سال قبل از ورشکستگی، دقت مدل شبکه عصبی، تحلیل ممیزی و رگرسیون لجستیک به ترتیب برابر با ۹۷/۵، ۹۱/۲ و ۹۲/۵ درصد می‌باشند. همچنین بر پایه داده‌های ۲ سال قبل از ورشکستگی برای مدل شبکه عصبی، تحلیل ممیزی و رگرسیون لجستیک به ترتیب به دقت‌هایی برابر با ۹۶/۲، ۸۵ و ۸۸/۷ درصد دست یافته‌اند.

اکرمی و سید حسینی (۱۳۹۱) با استفاده از الگوریتم ژنتیک، الگوهایی با ترکیبات مختلفی از متغیرهای حسابداری و بازار را طراحی و مورد آزمون قرار داده و برای اطمینان بیشتر به نتایج اعتبارسنجیضربدری را مورد استفاده قرار داده‌اند. نمونه، شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران را بین سال‌های ۱۳۸۰ تا ۱۳۸۸ در برگرفته است. نتایج نشان می‌دهد که اطلاعات حسابداری در پیش‌بینی ورشکستگی از سودمندی بیشتری نسبت به اطلاعات بازار برخوردار بوده و اطلاعات بازار نمی‌تواند مکمل خوبی برای اطلاعات حسابداری در پیش بینی ورشکستگی شرکت‌ها باشد.

مهرآذین و همکاران (۱۳۹۲) سه شبکه عصبی شعاعی[۲۵] مجزا را بر مبنای مدل‌های آلتمن، اسمایوسکی و ترکیب آن‌ها آموزش و نتایج را بر مبنای آزمون‌های آماری فیشر[۲۶] (فیش، ۱۹۲۲، ۸۷) و مکنمار[۲۷] (مکنمار، ۱۹۴۷، ۱۵۳) مورد مقایسه قرار داده‌اند. داده‌های مورد استفاده محققان فوق سال‌های ۱۳۸۳ تا ۱۳۹۰ را در برگرفته و از بورس اوراق بهادار تهران استخراج شده است. نتایج برتری متغیرهای مدل آلتمن در پیش‌بینی ورشکستگی را نسبت به دو ترکیب دیگر نشان می‌دهد.

www.SID.ir

Archive of SID

وکیلی فرد و همکاران (۱۳۹۲) با استفاده از سیستم استنتاج فازی عصبی انطباق پذیر[۲۸] شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران را بین سال‌های ۱۳۸۰ تا ۱۳۸۹ مورد بررسی قرار داده‌اند. تعداد نمونه‌های آن‌ها متشکل از ۴۰ شرکت ورشکسته و ۴۰ شرکت سالم بر مبنای ماده ۱۴۱ قانون تجارت می‌باشد. همچنین انتخاب متغیرهای ورودی بر مبنای نظر خبرگان و فرآیند تحلیل سلسله مراتبی[۲۹] می‌باشد. مدل طراحی شده ورشکستگی را با دقت ۸۳/۷۵ درصد در یکسال پیش از وقوع پیش‌بینی کرده است.

مشایخی و گنجی (۱۳۹۳) با مقایسه توان پیش‌بینی متغیرهای سودآوری شرکت‌هایی با سود با کیفیت و شرکت‌های با سود بی‌کیفیت، اثر کیفیت سود بر کارایی متغیرهای سودآوری در پیش‌بینی ورشکستگی را با استفاده از شبکه عصبی مصنوعی مورد بررسی قرار داده‌اند. بر اساس آزمون‌های انجام شده دقت پیش‌بینی شبکه عصبی مصنوعی برای شرکت‌های با سود با کیفیت به طور معنی‌داری بیشتر از شرکت‌های با سود بی‌کیفیت می‌باشد.

اقدامی و همکاران (۱۳۹۴) از روش تحلیل پوششی داده‌ها[۳۰] برای پیش‌بینی ورشکستگی شرکت‌های بورس اوراق بهادار تهران بهره برده‌اند. دوره زمانی داده‌های مورد استفاده در این پژوهش سال‌های ۱۳۸۶ تا ۱۳۸۹ را در برگرفته و مبنای متغیرهای بکار رفته نیز متغیرهای مورد استفاده در پژوهش شتی و همکاران (۲۰۱۲) می‌باشد. نتایج نشان‌دهنده این مساله است که از میان ۵۲ شرکت موجود در نمونه مورد بررسی ۹ شرکت (۱۷/۳ درصد از شرکت‌ها) کارا، ۲۱ شرکت (۴۰/۳ درصد از شرکت‌ها) در وضعیت ناکارا و تعداد ۲۲ شرکت (۴۲/۳ درصد از شرکت‌ها) در بعضی از سال‌ها کارا و در بعضی از سال‌ها ناکارا شناخته شده‌اند.

صالحی و عظیمی (۱۳۹۵) با بهره‌گیری از تابع رگرسیون لاجیت و منحنی مشخصه عملکرد سیستم قدرت پیش‌بینی ورشکستگی مدل‌های خطر مختلف را در ایران مورد بررسی قرار داده‌اند. داده‌های مورد استفاده ۲۴۲ شرکت غیرمالی بورس اوراق بهادار تهران از سال ۱۳۸۲ تا ۱۳۸۳ را در برمی‌گیرد. نتایج نشان‌دهنده کارایی مدل خطر با ترکیب اطلاعات حسابداری و بازار برای پیش‌بینی ورشکستی در محیط اقتصادی ایران است.

مشاهده می‌شود که تمامی مطالعات مبتنی بر داده کاوی داخل کشور تمرکز خود را بر روی داده‌های حاصل از بورس اوراق بهادار قرار داده‌اند. بررسی مطالعات خارجی نیز نشان دهنده استفاده از داده‌های بورس اوراق بهادار کشورهای مختلف می‌باشد که دلیل آن را می‌توان عدم امکان دسترسی به پایگاه داده‌های غنی‌تر دانست. مطالعات زیر تعدادی از پژوهش‌های انجام شده در سطع بین‌المللیمی‌باشند. این مطالعات در تلاش‌اند تا فرآیند انتخاب متغیرهای ورودی را وارد

www.SID.ir

Archive of SID

مدل‌های پیش‌بینی ورشکستگی شرکت‌ها نمایند. علاوه بر آن در بعضی از مطالعات فرآیند حذف داده‌های پرت نیز مورد ارزیابی قرار گرفته است.

مین و همکاران (۲۰۰۶) الگوریتم بهینه سازی ژنتیک را در ترکیب با ماشین بردار پشتیبان به‌منظور بهینه‌سازی همزمان متغیرهای ورودی و پارامترها مورد استفاده قرار داده و نتایج را با مدل‌های طبقه‌بندی شبکه عصبی مصنوعی و رگرسیون لجستیک مورد مقایسه قرار داده‌اند. نتایج نشان دهنده بهبود استفاده از ترکیب الگوریتم بهینه‌سازی ژنتیک با ماشین بردار پشتیبان نسبت به مدل‌های پایه می‌باشد.

تاسی (۲۰۰۹) با اعمال روش‌های انتخاب ویژگی آزمون t[۳۱]، ماتریس همبستگی[۳۲]، رگرسیون پله‌ای[۳۳]، تحلیل مؤلفه‌های اصلی[۳۴] و تحلیل عاملی[۳۵] بر روی داده‌های مورد بررسی و سپس اعمال شبکه عصبی مصنوعی بر روی داده‌های حاصل نتیجه گرفت که دقت طبقه‌بندی حاصل از آزمون t از سایر روش‌های انتخاب ویژگی بالاتر می‌باشد.

تاسی و چنگ (۲۰۱۲) اثر حذف داده‌های پرت مبتنی بر فاصله[۳۶] را بر روی دقت مدل‌های شبکه عصبی مصنوعی، درخت تصمیم‌گیری، رگرسیون لجستیک و ماشین بردار پشتیبان در اطلاعات مالی کشورهای استرالیا، آلمان، ژاپن و پایگاه داده UCC[۳۷] مورد بررسی قرار داده‌اند. نتایج نشان دهنده متفاوت بودن دقت‌های طبقه‌بندی حاصل در کشورهای مختلف می‌باشد اما به طور کلی حذف ۵۰ درصد داده‌ها بهبود دقت را در پی داشته است. علاوه بر آن نتایج نشان دهنده عملکرد مناسب و پایدارتر ماشین بردار پشتیبان در شرایط استفاده از حذف داده‌های پرت می‌باشد. علاوه بر آن دقت طبقه‌بندی ماشین بردار پشتیبان در شرایط عدم استفاده و استفاده از حذف داده‌های پرت نزدیک می‌باشد.

ونگ و همکاران (۲۰۱۴) روش FS-Boosting را به‌منظور انتخاب متغیرهای اثرگذار بر روی پیش‌بینی ورشکستگی شرکت‌ها در دو پایگاه داده بین‌المللی مورد ارزیابی قرار داده‌اند. نتایج نشان دهنده بهبود دقت مدل در شرایط استفاده از FS-Boosting نسبت به عدم استفاده از آن می‌باشد. نتایج حاصل در این پژوهش از جهت دقت کلی طبقه‌بندی، خطای نوع اول و خطای نوع دوم مورد ارزیابی قرار گرفته است.

اما در جدیدترین مقاله لیانگ و همکاران (۲۰۱۵) دقت حاصل از انتخاب متغیرهای ورودی را بر پایه روش‌های آماری و هوشمند مورد ارزیابی قرار داده‌اند. مدل‌های طبقه‌بندی مورد بررسی ماشین بردار پشتیبان خطی، ماشین بردار پشتیبان تابع شعاعی پایه[۳۸]، نزدیک‌ترین k همسایه[۳۹]، روش Naïve Bayes و شبکه عصبی مصنوعی را شامل می‌شود. روش‌های انتخاب ویژگی آماری مورد بررسی آزمون t، تحلیلممیزی خطی و رگرسیون لجستیک و روش‌های هوشمند استفاده از

www.SID.ir

Archive of SID

الگوریتم بهینه‌سازی ژنتیک و الگوریتم بهینه‌سازی ازدحام ذرات[۴۰] را شامل می‌شود. نتایج نشان دهنده عدم امکان تصمیم‌گیری به منظور تعیین بهترین ترکیب به دلیل متفاوت بودن خروجی بر روی پایگاه داده‌های مختلف می‌باشد اما به طور کلی دقت استفاده از الگوریتم ژنتیک و رگرسیون لجستیک به منظور انتخاب ویژگی از سایر ترکیبات بالاتر می‌باشد. لازم به ذکر است که بهینه‌سازی بررسی شده در این مقاله تنها بهینه‌سازی ترکیب متغیرهای ورودی می‌باشد.

بررسی مقالات خارجی نشان دهنده جهت‌گیری جریان کلی به سمت ورود انتخاب متغیرهای ورودی در فرآیند طراحی مدل‌های پیش‌بینی ورشکستگی است. علاوه بر آن مجموعه مقالات تنها چند پایگاه داده مبنا را مورد بررسی قرار داده و به منظور افزایش اطمینان به مدل‌ها لازم است تا مدل طراحی شده بر روی پایگاه داده‌های وسیع‌تری مورد ارزیابی و آزمون قرار گیرد.

۳- مبانی نظری

با توجه به اهمیت موضوع، مدل‌های طبقه‌بندی به تفکیک مورد بررسی قرار گرفته و به منظور جلوگیری از طولانی شدن بحث مجموعه الگوریتم‌های بهینه‌سازی به صورت خلاصه در یک بخشارائه شده است.

۳-۱- شبکه عصبی پرسپترون چند لایه و ماشین بایادگیری تشدید شده

یک شبکه عصبی مصنوعی، مجموعه‌ای از واحدهای محاسباتی مصنوعی به نام نورون[۴۱] و نحوه اتصالات بین آن واحدها می‌باشد. هر نورون دارای یک یا چند ورودی و تنها یک خروجی می‌باشد. در واقع می‌توان شبکه عصبی را به صورت یک گراف جهت‌دار و وزن‌دار معرفی کرد که در آن نورون‌ها رئوس گراف و یال‌های وزن‌دار ارتباطات بین آن‌ها می‌باشند (جین و همکاران،۱۹۹۶، ۳۱).

هر نورون یک واحد محاسباتی است که تابعی مشخص را بر روی ورودی‌های خود محاسبه کرده و خروجی متناظر آن را به نورون دیگری (یا به خروجی) منتقل می‌کند. توابع در نورون‌ها به صورت تابعی یک متغیره تعریف می‌شوند که ورودی این تابع برابر ضرب داخلی بردار ورودی در بردار وزن‌ها به علاوه مقدار ثابتی به نام بایاس[۴۲] می‌باشد.

از متداول‌ترین توابع فعال‌سازی که در شبکه‌های عصبی استفاده می‌شود می‌توان به تابع خطی[۴۳]، تابع پله‌ای[۴۴]، تابع خطی محدود[۴۵] و تابع سیگموید[۴۶] اشاره کرد. شبکه‌های عصبی بر اساس معماری شبکه دارای لایه‌های متفاوتی از نورون‌ها می‌باشند. در شبکه‌های چند لایه، لایه‌های مختلفی از نورون‌ها (لایه‌های مخفی) بین ورودی و لایه خروجی قرار می‌گیرند. یکی از ویژگی‌های جدایی ناپذیر هوشمندی، توانایی یادگیری می‌باشد. به طور کلی فرآیند یادگیری در

www.SID.ir

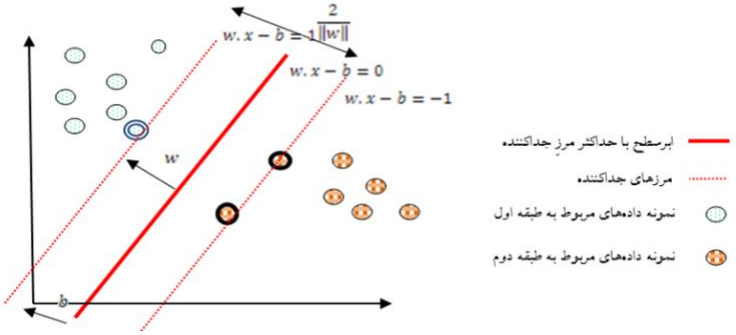
Archive of SID

شبکه عصبی مصنوعی به معنی تنظیم و به روز کردن معماری شبکه و وزن‌های آن است به طوری که شبکه یک مسأله را به نحو کارا حل نماید.

ماشین با یادگیری تشدید شده از ساختارهای شبکه عصبی جدید معرفی شده می‌باشد که علاوه بر سرعت بسیار بالای آموزش دارای دقت نسبتاً مناسبی نیز می‌باشد. این شبکه عصبی توسط هوانگ (۲۰۰۴) معرفی شده است. در الگوریتم پیشنهادی برخلاف تئوری‌های معمول برای بهینه‌سازیوزن‌ها و بایاس‌ها، وزن‌ها و بایاس‌های فوق به صورت تصادفی مقداردهی می‌شوند (در صورتی که تابع فعال‌سازی در بی‌نهایت مشتق‌پذیر باشد). به دلیل کاهش حجم محاسباتی سرعت آموزش شبکه فوق به صورت چشمگیری افزایش می‌یابد. در مقابل همان طور که ذکر شد این نوع از شبکه دقت نسبتاً قابل قبولی را نیز دارا می‌باشد. در حالت با هسته از توابع فعال‌سازی غیر خطی مانند RBF[۴۷] و ... با کمی تغییرات در ساختار شبکه بهره گرفته می‌شود.

۳-۲- ماشین بردار پشتیبان

ماشین بردار پشتیبان در ابتدا توسط وپنیک با نام شبکه‌های بردار پشتیبان[۴۸](۱۹۹۵) ارائه شده است. روش‌های طبقه‌بندی خطی تلاش می‌کنند تا با ساختن یک ابرسطح[۴۹] (که عبارت است از یک معادله خطی)، داده‌ها را از هم تفکیک کنند. روش طبقه‌بندی ماشین بردار پشتیبانیکی از روش‌های طبقه‌بندی خطی است که طی فرآیندی بهترین ابرسطحی را یافته که در آن با حداکثر فاصله[۵۰]، داده‌های مربوط به دو طبقه ازهم تفکیکمی‌شوند. به منظور درک بهتر مطلب، در شکل ۱ تصویری از یک مجموعه داده متعلق به دو کلاس نشان داده شده که روش ماشین بردار پشتیبان بهترین ابرسطح را برای جداسازی آن‌ها انتخاب می‌کند.



شکل ۱- ساختار مدل طبقه‌بندی ماشین بردار پشتیبان

منبع: یافته‌های پژوهشگر

www.SID.ir

Archive of SID

ایده اصلی این است که یک جداکننده مناسب انتخاب شود. منظور، جداکننده‌ای است که بیشترین فاصله را با نقاط همسایه از هر دو طبقه دارد. این جواب در واقع بیشترین مرز را با نقاط مربوط به دو طبقه مختلف دارد و می‌تواند با دو ابرسطح موازی که حداقل از یکی از نقاط دو طبقه عبور می‌کنند، کران‌دار شود. این بردارها، بردارهای پشتیبان نام دارند. پس از محاسبات پیشنهادی وپنیک فرآیند به یک مسأله بهینه سازی تبدیل شده که با تکنیک‌های برنامه ریزی درجه دوم قابل حل می‌باشد.

۳-۳- الگوریتم‌های بهینه‌سازی

به منظور جلوگیری از طولانی‌شدن بحث در این بخش مفهوم اصلی الگوریتم‌های بهینه‌سازی ارائه شده و از بیان جزئیات الگوریتم‌های مورد استفاده صرف‌نظر شده است. به طور کلی از الگوریتم‌های بهینه‌سازی به منظور حرکت از یک جواب اولیه تصادفی به سمت جواب بهینه با توجه به تابع هزینه تعریف‌شده در الگوریتم بهره گرفته می‌شود. هر یک از الگوریتم‌های بهینه‌سازی با توجه با ساختار خود که اکثراً الهام گرفته از طبیعت و منطبق بر واقعیت می‌باشند فرآیند فوق را به شکلی خاص دنبال می‌کنند.

الگوریتم رقابت استعماری یکی از الگوریتم‌های نسبتاً جدید در حوزه محاسبات تکاملی می‌باشد که توسط آتشپز و کارولوکس[۵۱](۲۰۰۷) با دیدی اجتماعی سیاسی بنیانگذاری شده است. ساختار تمامی الگوریتم‌های بهینه‌سازی به این شکل است که جمعیت اولیه‌ای (جواب‌های قابل قبول مسأله) به صورت اتفاقی ایجاد خواهند شد. به هر یک از اجزای جمعیت اولیه فوق در الگوریتم رقابت استعماری کشور[۵۲] گفته می‌شود. در فرآیند اعمال شده بر روی این مجموعه جواب اولیه تعدادی از بهترین جواب‌ها (کشورها) به عنوان استعمارگر[۵۳] و باقی به عنوان مستعمره[۵۴] انتخاب می‌شوند و الگوریتم طی فرآیندی به سمت مقدار بهینه حرکت خواهد کرد.

الگوریتم جستجوی هارمونی[۵۵] ابداع شده توسط گیم (۲۰۰۱) به دلیل کاربردی بودن برای مسائل بهینه‌سازی گسسته و پیوسته، محاسبات ریاضیات یکم، مفهوم ساده، پارامترهای کم و اجرای آسان به یکی از پرکاربردترین الگوریتم‌های بهینه‌سازی در سال های اخیر در مسائل مختلف تبدیل شده است.

این الگوریتم در مقایسه با سایر الگوریتم‌های فراابتکاری الزامات ریاضیات یکم تری دارد و می‌توان آن را با مسائل مختلف مهندسی با تغییر در پارامترها و عملگرها منطبق نمود. از مزیت‌های دیگر این الگو ریتم نسبت به سایر الگوریتم‌ها این است که برای ایجاد حل جدید بر خلاف سایر روش‌ها که از دو بردار حل در هر نسل استفاده می‌کند، این الگوریتم از همه حل‌های موجود در

www.SID.ir

Archive of SID

حافظه اش استفاده می‌کند. این ویژگی موجب افزایش انعطاف و جستجوی بهتر فضاهای گسترده‌تر جواب می‌شود. از ویژگی‌های دیگر الگوریتم جستجوی هارمونی این است که در مدت زمان مناسبی فضاهای حل با محدوده عملکرد بهتر را شناسایی می‌کند.

الگوریتم فرهنگی توسط رینولد (۱۹۹۴) ابداع شده است. در این الگوریتم برخلاف اکثر الگوریتم‌های بهینه‌سازی دو فضای جستجو برای یافتن جواب‌های بهینه برقرار می‌باشد. یکی فضای جمعیت[۵۶] که بر مبنای نظریه ژنتیکی داروین است و دیگری فضای باور[۵۷]که قسمتی از فرهنگ را ارائه می‌کند که فضای دوم وجه تمایز ساختار فوق با الگوریتم‌های مشابه الگوریتم ژنتیک (تنها مبتنی بر فضای جمعیت) می‌باشد (اندرس، ۲۰۰۷، ۲۶۲). فضای باور، در واقع اطلاعات فرهنگ افراد را مدل می‌کند. به طور خلاصه می‌توان اینگونه بیان کرد که هر دو فضا بصورت موازی با هم کار کرده و بر روی هم تأثیر می‌گذارند.

۴- هدف، داده‌ها و متغیرها و ساختار سیستم

۴-۱- هدف و روش تحقیق

هدف این پژوهش آزمون ساختارهای مختلف یک سیستم هوشمند ترکیبی برای پیش‌بینی ورشکستگی شرکت‌ها با در نظر گرفتن داده‌هایی فراتر از داده‌های شرکت‌های بورسی در ایران می‌باشد. مهم‌ترین سؤال کارایی سیستم‌های فوق در پیش‌بینی ورشکستگی شرکت‌ها در این مقیاس وسیع است. سؤال دوم به بررسی این مساله می‌پردازد که کدام ترکیب نتایج بهتری را در هر صنعت مورد بررسی در پی داشته و آیا امکان رسیدن به یک مدل واحد برای پیش‌بینی ورشکستگی برقرار می‌باشد یا خیر. سؤال سوم نیز امکان استخراج متغیرهای اثرگذار را در سیستم‌های فوق مورد ارزیابی قرار می‌دهد. روش تحقیق مورد استفاده بر مبنای هدف یک تحقیق کاربردی و روش کلی مورد استفاده نیز مبتنی بر بهره‌گیری از داده‌ها[۵۸] می‌باشد.

۴-۲- معیار تعیین وضعیت ورشکستگی شرکت‌های کشور

معیار مورد استفاده به منظور تعیین وضعیت ورشکستگی یک شرکت مشخص ماده ۱۴۱ قانون تجارت می‌باشد. بر اساس ماده ۱۴۱ قانون تجارت ایران شرکتی به عنوان ورشکسته در نظر گرفته می‌شود که بر اساس زیان‌های وارده، حداقل نیمی از سرمایه شرکت از میان برود. در شرایط فوق طبق قانون ذکر شده هیئت مدیره مکلف است بلافاصله طی مجمع عمومی فوق‌العلاده، صاحبان سهام کسب و کار را دعوت کرده تا موضوع انحلال یا بقای شرکت به رأی گذاشته شود. در صورتی که مجموع شرکت به انحلال رأی مثبت ندهد در همان جلسه و با رعایت مقررات ماده ۶ این قانون،

www.SID.ir

Archive of SID

سرمایه شرکت به مقدار سرمایه موجود کاهش خواهد یافت (ناصرزاده، ۱۳۷۴، ۱۹). بر پایه این قانون شرکت‌های ورودی به منظور آموزش سیستم به دو گروه شرکت‌های سالم و شرکت‌های ورشکسته تقسیم شده‌اند.

۴-۳- پیش پردازش

با توجه به برابر نبودن تعداد شرکت‌های سالم و ورشکسته (به طور معمول تعداد شرکت‌های ورشکسته به صورت معناداری از تعداد شرکت‌های سالم کمتر می‌باشد) نیاز است تا با اعمال یک ساختار بالانس کننده 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.

فرآیند فوق در حالت کمتر بودن شرکت‌های ورشکسته نسبت به شرکت‌های سالم مورد بررسی قرار می‌گیرد (حالت معکوس نیز قابل اعمال می‌باشد). در این حالت شرکت‌های سالم بر مبنای لگاریتم مجموع دارایی‌ها (یکی از پراکسی‌های[۵۹] مشخص کننده اندازه شرکت) به ۵ بخش تقسیم می‌شود سپس معادل با مقدار لگاریتم مجموع دارایی‌های شرکت ورشکسته بخش متناسب از میان شرکت‌های سالم انتخاب شده و از میان نمونه‌های موجود در آن بخش به صورت تصادفی شرکتی انتخاب خواهد شد (در صورت خالی بودن بخش مورد نظر اولویت به ترتیب با بخش‌های کوچک‌تر و در غیر این صورت اولویت با شرکت‌های بزرگ‌تر می‌باشد). فرآیند فوق تا نمونه‌گیری کامل ادامه خواهد یافت.

۴-۴- جامعه آماری

این پژوهش صنایع مواد غذایی و نساجی را با مجموعه کدهای آیسیک[۶۰] جدول ۱ در بر می‌گیرد.

جدول ۱- کدهای آیسیک و صنایع مورد بررسی

کدهای آیسیک (سطح ۲)	صنعت
۹۱۵۱،۹۱۵۲،۹۱۵۳،۹۱۵۴،۹۱۵۵	موادغذایی
۹۱۷۱،۹۱۷۲،۹۱۷۳،۹۱۸۱،۹۱۸۲،۹۱۹۱،۹۱۹۲	نساجی

منبع: لیست کدهای آیسیک مورد استفاده سازمان امور مالیاتی کشور

www.SID.ir

Archive of SID

اطلاعات ورودی به سیستم با همکاری سازمان امور مالیاتی کشور از اظهارنامه عملکرد اشخاص حقوقی حاصل شده است. همچنین دوره زمانی مورد بررسی سال‌های ۱۳۸۹، ۱۳۹۰ و ۱۳۹۱ را در بر می‌گیرد. اطلاعات ورودی بر مبنای ماده ۱۴۱ قانون تجارت به دو گروه شرکت‌های ورشکسته و سالم تقسیم شده‌اند. در واقع ورودی سیستم بر مبنای سال‌های ذکر شده نسبت‌های مالی در سال ۱۳۸۹ (ورودی)، وضعیت ورشکستگی در سال ۱۳۹۰ (خروجی) و نسبت‌های مالی در سال ۱۳۹۰ (ورودی) و وضعیت ورشکستگی در سال ۱۳۹۱ (خروجی) را شامل می‌شود. تعداد نمونه‌های مورد بررسی در صنعت موادغذایی و نساجی به ترتیب برابر با ۵۸۲۵ و ۴۰۸۹ می‌باشد که با اعمال معیار قانونی ورشکستگی به ترتیب ۹۹۹ و ۸۴۸ نمونه شرایط ورشکستگی را در دو سال مورد بررسی دارا بوده‌اند.

۴-۵- نسبت‌های مالی منتخب اولیه و صنایع مورد بررسی

به منظور انتخاب مجموعه متغیر و نسبت‌های مالی ورودی اولیه ۱۱۷ مقاله بین سال‌های ۱۹۶۸ تا ۲۰۱۳ مورد بررسی قرار گرفت. سپس ۳۰ متغیر (برابر با ۲۵ درصد متغیرهای دارای بیشترین استفاده) انتخاب و بر داده‌های کشور اعمال شد. با توجه به نتایج حاصل بخشی از متغیرها و نسبت‌های مالی با توجه به غیرقابل محاسبه بودن بر مبنای داده‌های موجود (نبود اطلاعات مورد نیاز محاسبه متغیر برای تعداد زیادی از شرکت‌های ورودی) حذف و در پایان متغیرها و نسبت‌های حاصل بر مبنای نظر خبرگان مورد ارزیابی و انتخاب قرار گرفته‌اند.۲۱ متغیر منتخب نهایی ورودی سیستم در جدول ۲ ارائه شده است.

جدول ۲- متغیرهای منتخب اولیه سیستم هوشمند پیش‌بینی ورشکستگی شرکت‌ها

سود (زیان) قبل از بهره و مالیات به کل دارایی‌ها	فروش خالص به کل دارایی‌ها
سود (زیان) پس از کسر مالیات به کل دارایی‌ها *(ROA)*	سود (زیان) پس از کسر مالیات به جمع حقوق صاحبان سهام
سود انباشته به کل دارایی‌ها	کل بدهی‌ها به کل دارایی‌ها
دارایی‌های جاری به کل دارایی‌ها	دارایی‌های آنی[۶۱] به کل دارایی‌ها
کل دارایی‌ها	کل دارایی‌ها به حقوق صاحبان سهام
سود (زیان) عملیاتی به کل دارایی‌ها	دارایی‌های ثابت[۶۲] به کل دارایی‌ها
موجودی انبار به کل دارایی‌ها	کل بدهی‌ها به حقوق صاحبان سهام
کل بدهی‌ها	اسناد دریافتنی[۶۳]
اسناد دریافتنی به کل دارایی‌ها	سود (زیان) ناخالص فروش به کل دارایی‌ها
کل بدهی‌ها به (کل بدهی‌ها + کل حقوق صاحبان سهام)	سود (زیان) ناخالص فروش
سود (زیان) پس از کسر مالیات	

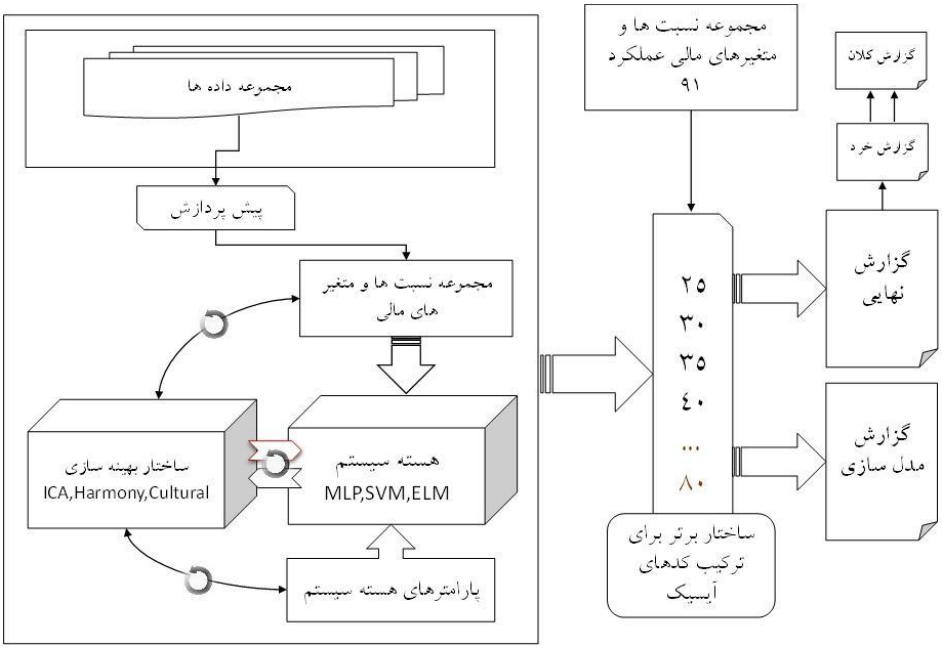
منبع: یافته‌های پژوهشگر

www.SID.ir

Archive of SID

۴-۶- ساختار هسته سیستم

ساختار کلی هسته سیستم در شکل ۲ ارائه شده است.



شکل ۲- ساختار سیستم هوشمند پیش‌بینی ورشکستگی شرکت‌ها

منبع: یافته‌های پژوهشگر

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 بهینه‌سازیدر 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.

www.SID.ir

Archive of SID

- 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: C(بین 2^{-7}و2^{7}) –گاما (بین 10^{-3} و 10^{3})
- ماشین با یادگیری تشدید شده (بدون هسته): تعداد نورون‌های لایه پنهان (بین ۱ تا ۵۰) – پارامتر σ مرتبط با کرنل RBF (بین 10^{-6} و 800)
- ماشین با یادگیری تشدید شده (دارای هسته): ضریب تنظیم[۶۴] و پارامتر σ مرتبط با کرنل RBF (بین 10^{-6} و 800)

جدول ۳- پارامترهای پیش‌فرض الگوریتم‌های بهینه‌سازی

پارامترهای الگوریتم فرهنگی	پارامترهایالگوریتم جستجوی هارمونی	پارامترهای الگوریتم رقابت استعماری
Population size = 40	Harmony memory size = 100	Number of initial countries = 100
Acceptance ratio = 0/35	Number of new harmonies = 100	Number of initial imperialists = 8
Alpha lower bound = 0/20	Fret width coefficient = 0/01	Revolution rate = 0/3
Alpha upper bound = 0/40	Pitch adjustment rate = 0/3	Percentage of search space size = 0/02
	Fret width damp ratio = 0/999	Assimilation coefficient = 2
	Consideration rate = 0/9	Assimilation angle coefficient = 0/5
	Harmony memory size = 100	Cost coefficient (Zeta) = 0/02
		Dump ratio = 0/99

منبع: یافته‌های پژوهشگر

همچنین برای آموزش شبکه عصبی از الگوریتم trainscg[۶۵]با توابع انتقال تانژانت سیگموئید[۶۶] 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[۶۷]) 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)

www.SID.ir

Archive of SID

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.

(۱)

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

در معادله ۱α برابر با ۲ در نظر گرفته شده و X مجموعه احتمالات خروجی را شامل می‌شود. علاوه بر آن $average$ نشان دهنده اپراتور میانگین و STD 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[۶۸] 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 پژوهشاستفاده 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[۶۹] 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[۷۰](جریان‌های انشعابی) از جریان اصلی[۷۱] 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[۷۲].

۴-۷- 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[۷۳] دیده و با بخش پنجم مورد آزمون[۷۴] 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

www.SID.ir

Archive of SID

برای بخش اعتبارسنجی[۷۵] 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[۷۶] استفاده خواهند شد. همچنین برای جلوگیری از برازش بیش از اندازه از روش توقف زود هنگام[۷۷] در ۶ تکرار (مقدار پیش‌فرض) بهره گرفته شده است. در سایر مدل‌های طبقه‌بندی از اعتبارسنجی ضربدری ۱۰ بخشی پایه استفاده خواهد شد.

۴-۸- ابزار

برای طراحی ساختار اصلی سیستم از نرم افزار MATLAB نسخه R2014a و R2014b بهره گرفته شده است. برای گزارش‌گیری استاندارد از سیستم از VBA[۷۸] در محیط Microsoft Excel 2013 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[۷۹] در محیط MATLAB استفاده شده است. همچنین به منظور ارتباط با پایگاه داده و تبدیل و انتقال داده‌ها از SQL server 2014 بهره گرفته شده است. واحد پردازش مرکزی مورد استفاده *Intel Core i7–4702MQ* دارای ۴ هسته فیزیکی[۸۰] و ۸ هسته منطقی[۸۱] بوده و هشت گیگابایت حافظه دسترسی تصادفی[۸۲] و سیستم‌عامل *Windows* 8.1 Enterprise edition 64bit مورد استفاده قرار گرفته است. به منظور افزایش سرعت پردازش از تولباکس پردازش موازی MATLAB[۸۳] برای انتقال تکرار اعتبارسنجی ضربدری به پردازش موازی بهره گرفته شده است. ماشین بردار پشتیبان بر پایه کتابخانه ماشین بردار پشتیبان[۸۴] و مدل‌های ماشین با یادگیری تشدید شده نیز بر پایه کدهای ارائه‌کنندگان این روش[۸۵] پیاده‌سازی شده‌اند.

۵- نتایج

۵-۱- 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

www.SID.ir

Archive of SID

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 نمودار ROC[۸۶]	تعداد سالم ورشکسته	دقت کلی	دقت تشخیص سالم	دقت تشخیص ورشکسته	سطح زیر نمودار ROC
مواد غذایی- ICA[۸۷]	شبکه عصبی مصنوعی	۴۸۲۶ ۹۹۹	۷۴/۸۴	۷۹/۷۲	۶۹/۹۵	۰/۷۴۸۳	۴۳۶۸ ۸۷۴	۸۵/۰۷	۸۴/۴۶	۸۵/۶۷	۰/۸۵۱۷
	ماشین بردار پشتیبان		۸۳/۸۲	۸۱/۷۳	۸۵/۹۰	۰/۸۳۸۲		۸۳/۴۰	۸۳/۶۲	۸۳/۱۹	۰/۸۳۴۴
	ELM (بدون هسته)		۸۰/۹۴	۷۹/۱۸	۸۲/۷۱	۰/۸۱۰۱		۸۳/۱۴	۸۴/۱۶	۸۲/۱۲	۰/۸۳۲۰
	ELM (با بهره‌گیری از هسته)		۷۸/۶۷	۸۱/۲۸	۷۶/۰۶	۰/۷۸۷۴		۸۲/۳۶	۸۲/۱۲	۸۲/۶۰	۰/۸۲۳۷
نساجی- ICA	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
	ELM (بدون هسته)		۸۳/۰۰	۷۸/۶۶	۸۷/۳۵	۰/۸۳۰۲		۸۴/۴۶	۸۱/۴۸	۸۷/۴۵	۰/۸۴۴۶
	ELM (با بهره‌گیری از هسته)		۷۰/۳۹	۶۸/۱۴	۷۲/۶۳	۰/۷۰۴۲		۸۳/۸۲	۸۴/۵۷	۸۳/۰۶	۰/۸۳۸۲
مواد غذایی- HS[۸۸]	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
	ELM (بدون هسته)		۸۳/۰۰	۷۸/۶۶	۸۷/۳۵	۰/۸۳۰۲		۸۴/۴۶	۸۱/۴۸	۸۷/۴۵	۰/۸۴۴۶
	ELM (با بهره‌گیری از هسته)		۷۰/۳۹	۶۸/۱۴	۷۲/۶۳	۰/۷۰۴۲		۸۳/۸۲	۸۴/۵۷	۸۳/۰۶	۰/۸۳۸۲
نساجی- HS	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
	ELM (بدون هسته)		۸۲/۶۱	۷۸/۵۴	۸۶/۶۷	۰/۸۲۶۴		۸۳/۵۳	۷۹/۷۹	۸۷/۲۶	۰/۸۳۵۶
	ELM (با بهره‌گیری از هسته)		۸۰/۱۲	۷۵/۲۵	۸۴/۹۹	۰/۸۰۱۷		۸۵/۱۹	۸۸/۷۷	۸۱/۶۲	۰/۸۵۲۰

www.SID.ir

Archive of SID

		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 نمودار ROC[۸۶]	تعداد سالم ورشکسته	دقت کلی	دقت تشخیص سالم	دقت تشخیص ورشکسته	سطح زیر نمودار ROC
موادغذایی – CA[۸۹]	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
	ELM (بدون هسته)		۸۰/۳۴	۷۶/۴۲	۸۴/۲۶	۰/۸۰۴۰		۸۵/۷۳	۸۶/۴۷	۸۴/۹۹	۰/۸۵۷۴
	ELM (با بهره‌گیری از هسته)		۸۰/۰۹	۸۳/۶۲	۷۶/۵۶	۰/۸۰۱۱		۸۴/۱۹	۸۲/۷۴	۸۵/۶۳	۰/۸۴۲۷
نساجی – CA	شبکه عصبی مصنوعی	۳۲۴۱ ۸۴۸	۷۴	۷۹/۴۵	۶۸/۵۴	۰۰/۷۴	۲۹۰۰ ۷۸۰	۸۲/۹۸	۸۲/۰۱	۸۳/۹۶	۰/۸۲۹۷
	ماشین بردار پشتیبان		۷۲/۳۱	۷۸/۰۹	۶۶/۵۲	۰/۷۲۳۱		۸۳/۱۷	۸۳/۶۶	۸۲/۶۸	۰/۸۳۲۱
	ELM (بدون هسته)		۷۹/۸۶	۷۴	۸۵/۷۰	۰/۷۹۸۹		۸۰/۶۲	۸۱/۶۱	۷۹/۶۲	۰/۸۰۶۶
	ELM (با بهره‌گیری از هسته)		۷۲/۶۶	۷۸/۱۵	۶۷/۱۷	۰/۷۲۶۷		۸۳/۳۰	۸۷/۱۲	۷۹/۴۷	۰/۸۳۳۳

منبع: یافته‌های پژوهشگر

۵-۲- 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

صنعت موادغذایی		صنعت نساجی	
تعداد کل شرکت‌ها	تعداد حذفی	تعداد کل شرکت‌ها	تعداد حذفی
۵۸۲۸	۱۲۳۰	۴۰۹۲	۸۶۷

منبع: یافته‌های پژوهشگر

www.SID.ir

Archive of SID

مشاهده می‌شود که اعمال فرآیند حذف داده‌های پرت موجب حذف تعداد گسترده‌ای از شرکت‌ها در هر دو صنعت موادغذایی و نساجی 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 افزایش‌یافته 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[۹۰] (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.

۵-۳- 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 بهینه سازی فرهنگی با الگوریتم رقابت استعماری و جستجوی هارمونی الگوریتم فوق از فرآیند حذف خواهد شد. برای بررسی آماری تفاوت ذکر شده از آزمون مکنماربهره 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 می رود 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[۹۱] (ییتس، ۱۹۳۴، ۲۱۷) بوده و آماره آزمون چی اسکوئر[۹۲] می‌باشد.

www.SID.ir

Archive of SID

جدول ۶- مقایسه آماری الگوریتم‌های بهینه‌سازی

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)
CA-HS موادغذایی	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
CA-ICA موادغذایی	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
CA-HS نساجی	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
CA-ICA نساجی	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

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.

تنها در صنعت موادغذایی 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

www.SID.ir

Archive of SID

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.

۵-۴- 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.

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

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"،

www.SID.ir

Archive of SID

"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 کمترینمیانگین 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.

۵-۵- 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 فرآیندبهینه‌سازی 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 تفاوت‌هایذکر 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.

www.SID.ir

Archive of SID

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

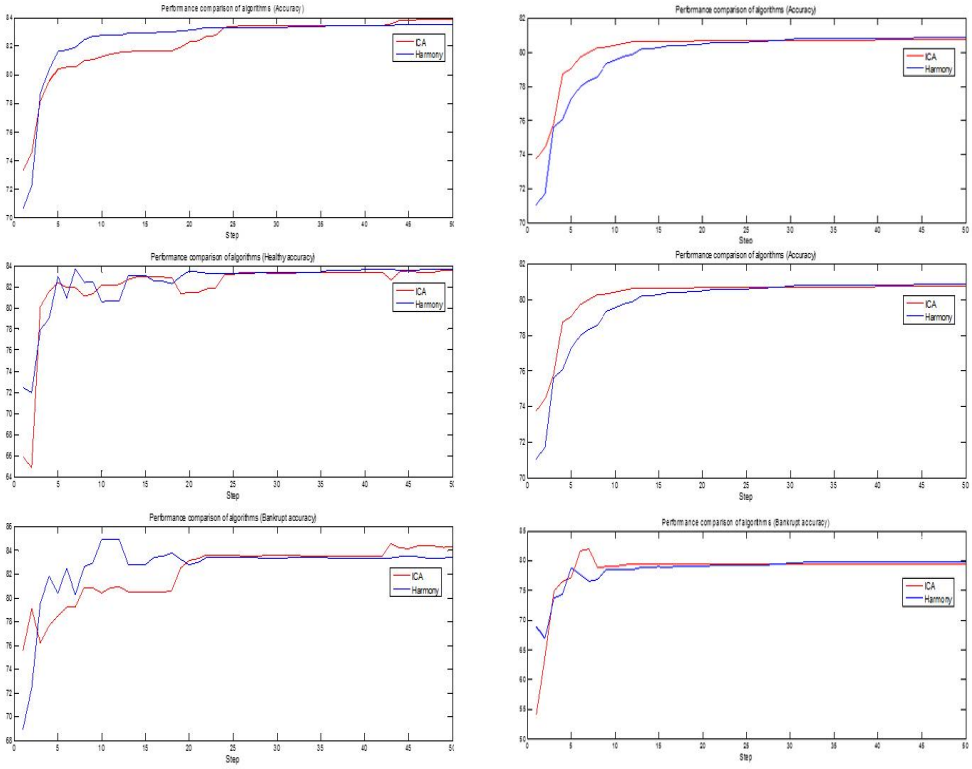
از میان چهار ترکیب بررسی شده در جدول ۸، تنها در یک حالت (در جدول با ستاره نشان داده شده است) دقت‌های حاصل از مدل طبقه‌بندییادگیری تشدید شده (بدون هسته) از ماشین بردار پشتیبان بیشتر بوده (با توجه به نتیجه آزمون آماری مکنمار تفاوت فوق معنادار نیزمی‌باشد) و در سایر حالات دقت تشخیص ماشین بردار پشتیبان بالاتر 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 طبقه‌بندییادگیری تشدید شده بررسی شده در این پژوهش تنها دارای نتایج دودویی (ورشکسته یا سالم) می‌باشد. در استفاده اجرایی از سیستم طراحی شده تعیین احتمالات ورشکستگی از اهمیت بالایی برخوردار می‌باشد. به این دلیل علارغم سرعت بالای مدل طبقه بندییادگیری 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.

www.SID.ir

Archive of SID

۵-۶- 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.



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

منبع: یافته‌های پژوهشگر

www.SID.ir

Archive of SID

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 موادغذایی 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 خروجیمدل‌ها 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 می باشد.

با توجه به نمودارها می‌توان نتیجه گرفت که علارغم 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.

۵-۷- نسبت‌ها و متغیرهای مالی برتر

برای ارائه نتایج این قسمت از مجموع نتایج مدل‌های ارائه شده در بخش (۵-۶) بهره گرفته شده است. 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.

www.SID.ir

Archive of SID

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.

www.SID.ir

Archive of SID

۶- بحث و نتیجه‌گیری

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 نتایجاین 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[۹۳] 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[۹۴]و سایر روش‌های حذف داده‌های پرت توجه ویژه‌ای شود. از نظر پردازش و ابزارهای مرتبط با آن انتقال

www.SID.ir

Archive of SID

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

۱) اقدامی، اسماعیل، سهراب کرد رستمی، مجتبی ملکی و حبیبه آزماینده، (۱۳۹۴)، "ارزیابی ورشکستگی در بورس اوراق بهادارتهران با بکارگیری مدل پویایی شبکه: روشی بر پایه تحلیل پوششی داده‌ها"، مجله مهندسی مالی و مدیریت اوراق بهادار، ۶(۲۲)، صص ۱۶۵-۱۸۰.

۲) اکرمی، غلامرضا و سید مصطفی سید حسینی، (۱۳۹۱)، "سودمندی اطلاعات حسابداری نسبت به اطلاعات بازار در پیش‌بینی ورشکستگی"، مجله دانش حسابداری، ۳(۱۰)، صص ۹۳-۱۱۶.

۳) راعی، رضا و سعید فلاح‌پور، (۱۳۸۳)، "پیش‌بینی درماندگیشرکت‌ها با استفاده از شبکه‌های عصبی مصنوعی"، تحقیقات مالی،۶(۱)، صص ۳۳-۴۶.

۴) صالحی، نازنین و مجید عظیمی یانچشمه، (۱۳۹۵)، "بررسی تطبیقی مدل خطر و مدل‌های سنتی برای پیش‌بینی ورشکستگی"، حسابداری مالی، ۸(۳۰)، صص ۹۴-۱۲۱.

۵) قدرتی، حسن و امیرهادی معنوی مقدم، (۱۳۸۹)، "بررسی دقت مدل‌های پیش‌بینی ورشکستگی (مدل‌های آلتمن، شیراتا، اهلسون، زمیسکی، اسپرینگیت، سی ای اسکور، فولمر، ژنتیک فرجزاده و ژنتیک مک‌کی) در بورس اوراق بهادار تهران"، تحقیقات حسابداری و حسابرسی، ۲(۷).

۶) قدیریمقدم، ابوالفضل، محمد مسعود غلامپور فرد و فرزانه نصیرزاده، (۱۳۸۸)، "بررسی تواناییمدل‌های پیش بینی ورشکستگی آلتمن و اهلسون در

۷) پیش‌بینی ورشکستگیشرکت‌های پذیرفته شده در بورس اوراق بهادار"، مجله دانش و توسعه، شماره ۲۵، صص ۱۹۳-۲۲۰.

۸) مشایخی، بیتا و حمیدرضاگنجی، (۱۳۹۳)، "تأثیر کیفیت سود بر پیش‌بینی ورشکستگی با استفاده از شبکه عصبی مصنوعی"، پژوهش‌های حسابداری مالی و حسابرسی (پژوهشنامه حسابداری مالی و حسابرسی)، ۶(۲۲)، صص ۱۴۷-۱۷۳.

۹) مکیان، سیدنظام الدین و سلیم کریمی تکلو، (۱۳۸۸)، "پیش‌بینی ورشکستگی شرکت‌های تولیدی با استفاده از شبکه‌های عصبی مصنوعی (مطالعه موردی: شرکت‌های تولیدی استان کرمان)"، فصلنامه اقتصاد مقداری (فصلنامه بررسی‌های اقتصادی)، ۶(۱)، صص ۱۲۹-۱۴۴.

www.SID.ir

Archive of SID

۱۰) مکیان، سید نظام الدین، سید محمد تقی المدرسی و سلیم کریمی تکلو، (۱۳۸۹)، "مقایسه مدل شبکه‌های عصبی مصنوعی با روش‌های رگرسیون لجستیک و تحلیل ممیزی در پیش بینی ورشکستگی شرکت‌ها"، پژوهش‌های اقتصادی، ۱۰(۲)، صص ۱۴۱-۱۶۱.

۱۱) مهرآذین، علیرضا، احمد زنده دل، محمد تقی‌پور و امید فروتن، (۱۳۹۲)، "شبکه‌های عصبی شعاعی آموزش یافته بر پایه متغیرهای مدل‌های آماری و مقایسه آن‌ها در پیش بینی ورشکستگی"، فصلنامه علمی پژوهشی دانش سرمایه‌گذاری، ۲(۷)، صص ۱۴۹-۱۶۶.

۱۲) ناصرزاده، هوشنگ، (۱۳۷۴)، "قانون تجارت"، تهران، نشر دیدار.

۱۳) وکیلی‌فرد، حمیدرضا، نازنین پیله وری و سیده سمانه زیدی، (۱۳۹۳)، "ارائه مدلی جهت پیش‌بینی ورشکستگی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از سیستم استنتاج فازی عصبی انطباق‌پذیر"، مجله مهندسی مالی و مدیریت اوراق بهادار، ۵(۱۸)، صص ۱۷-۳۰.

14) Aghaie, A., & Saeedi, A, (2009),“Using Bayesian Networks for Bankruptcy Prediction: Empirical Evidence from Iranian Companies”,Information Management and Engineering, IEEE, PP. 450-455.

15) Andries, E. P, (2007),“Computational Intelligence: An Introduction”,Wiley, Second ed.

16) Ashoori, S., & Mohammadi, S, (2011),“Compare Failure Prediction Models based on Feature Selection Technique: Empirical Case from Iran. Rocedia Computer Science, No.3, PP.568-573.

17) Atashpaz gargari, E., & Lucas, C, (2007),“Imperialist Competitive Algorithm: An Algorithm for Optimization Inspired by Imperialistic Competition”, Evolutionary Computation,Singapore, IEEE Congress on, PP. 4661-4667.

18) Barboza, F., Kimura, H., & Altman, E. (2017), “Machine learning models and bankruptcy prediction”, Expert Systems with Applications, No. 83, PP. 405-417.

19) Etemadi, H., Anvary Rostamy, A., & Farajzadeh Dehkordi, H, (2009),“A Genetic Programming Model for Bankruptcy Prediction: Empirical Evidence from Iran”,Expert Systems with Applications, 36(2), PP.3199-3207.

20) Fisher, R. A, (1922),“On the Interpretation of $\chi 2$ from Contingency Tables, and the Calculation of P”,Journal of the Royal Statistical Society”, PP.87-94.

21) Geem, Z., Kim, J.-H., & Loganathan, G, (2001),“A New Heuristic Optimization Algorithm: Harmony Search. SIMULATION: Transactions of The Society for Modeling and Simulation International”, PP.60-68. doi:10.1177/003754970107600201

22) Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K, (2004),“Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks”, Neural Networks,Proceedings, IEEE International Joint Conference No.2, PP. 985-990.

23) Jain, A. K., Mao, J., & Mohiuddin, K. M, (1996),“Artificial Neural Networks: A Tutorial”,Computer - Special issue: Neural Computing, 29(3), PP.31-44.

www.SID.ir

Archive of SID

24) Liang, D., Tsai, C.-F., & Wu, H.-T, (2015),"The Effect of Feature Selection on Financial Distress Prediction", Knowledge-Based Systems, No.73, PP.289–297.
25) McNemar, Q, (1947),"Note on the Sampling Error of the Difference between Correlated Proportions or Percentages",Psychometrika, 12(2), PP.153-157.
26) Min, S.-H., Lee, J., & Han, I, (2006),"Hybrid Genetic Algorithms and Support Vector Machines for Bankruptcy Prediction",Expert Systems with Applications, 31(3), PP.652-660.
27) Mokhatab Rafei, F., Montazeri, S., & Boostanian, S, (2011),"Financial Health Prediction Models Using Artificial Neural Networks, Genetic Algorithm and Multivariate Discriminant Analysis: Iranian Evidence",Expert Systems with Applications, 38(8), PP.10210-10217.
28) Moradi, M., Shafiee Sardasht, M., & Ebrahimpour, M, (2012),"An Application of Support Vector Machines in Bankruptcy Prediction; Evidence from Iran",World Applied Sciences Journal, PP.710-717.
29) Reynolds, R. G, (1994),"An Introduction to Cultural Algorithms",Proceedings of the Third Annual Conference on Evolutionary Programming", Singapore, PP.131–139.
30) Shetty, U., Pakkala, T. P. M., & Mallikarjunappa, T, (2012),"A Modified Directional Distance Formulation of DEA to Assess Bankruptcy: An Application to IT/ITES Companies in India", Expert Systems with Applications,39(2), PP.1988-1997.
31) Tsai, C.-F, (2009),"Feature Selection in Bankruptcy Prediction",Knowledge-Based Systems, 22(2), PP.120–127.
32) Tsai, C.-F., & Cheng, K.-C, (2012),"Simple Instance Selection for Bankruptcy Prediction",Knowledge-Based Systems, No.27, PP.333–342.
33) Vapnik, V, (1995),"Support-Vector Networks",Machine learning, 20(3), PP.273-297.
34) Wang, G., Ma, J., & Yang, S, (2014),"An Improved Boosting Based on Feature Selection for Corporate Bankruptcy Prediction",Expert Systems with Applications, 41(5), PP.2353–2361.
35) Yates, F, (1934), "Contingency Tables Involving Small Numbers and the $\chi 2$ Test", Supplement to the Journal of the Royal Statistical Society, 1(2), PP.217-235.

یادداشت‌ها

[1] Supply chain
[2] Feature selection
[3] Multilayer perceptron neural network
[4] Support vector machine
[5] Extreme learning machine(ELM)
[6] Kernel trick
[7] Imperialist competitive algorithm
[8] Harmony search
[9] Cultural algorithm

www.SID.ir

Archive of SID

[10] fold cross validation
[11] Logical processor
[12] Central processing unit(CPU)
[13] Parallel computing
[14] Local optimum
[15] Outlier
[16] Step
[17] Artificial intelligence(AI)
[18] Data mining
[19] Genetic programming(GP)
[20] Multiple discriminant analysis(MDA)
[21] Bayesian neural networks(BNN)
[22] Logistic regression
[23] Genetic algorithm
[24] Linear discriminant analysis(LDA)
[25] Radial basis function network
[26] Fisher's exact test
[27] McNemar's test
[28] Adaptive neuro fuzzy inference system(ANFIS)
[29] Analytic hierarchy process(AHP)
[30] Data envelopment analysis(DEA)
[31] t-test
[32] Correlatin matrix
[33] Stepwise regression
[34] Principal component analysis(PCA)
[35] Factor analysis(FA)
[36] Distance-based outlier detection
[37] www.mill.ucsd.edu
[38] Radial basis function(RBF)
[39] K-nearest neighbor(KNN)
[40] Particle swarm optimization(PSO)
[41] Neuron
[42] Bias
[43] Linear function
[44] Step function
[45] Symmetric saturating linear transfer function
[46] Sigmoid function
[47] Radial basis function
[48] Support-vector networks
[49] Hyperplane
[50] Maximum margin
[51] Caro lucas
[52] Country
[53] Imperialist
[54] Colony
[55] Harmony search algorithm
[56] Population space
[57] Belief space
[58] Data-driven procedure
[59] Proxy
[60] International standard industrial classification(ISIC)
[۶۱] موجودی نقد بانک + سرمایه‌گذاری کوتاه‌مدت + حساب‌ها و اسناد دریافتنی + جاری شرکا، سهامداران.
[۶۲] دارایی ثابت، مشهود + دارایی‌های نامشهود.
[۶۳] حساب‌ها و اسناد دریافتنی تجاری + سایر حساب‌ها و اسناد دریافتنی.
[64] Regularization coefficient

www.SID.ir

Archive of SID

[65] Scaled conjugate gradient backpropagation
[66] Tansig
[67] Vectorization
[68] Majority voting
[69] GPU computing/CUDA
[70] Sub-stream
[71] Stream
[72] استفاده از تابع مولد Combined multiple recursive کاهش دقت سیستم را در پی داشت. استفاده از این تابع نیازمند بررسی بیشتر اثرات آن در مطالعات آتی می‌باشد.
[73] Training
[74] Testing
[75] Validation data
[76] Overfitting
[77] Early stopping
[78] Visual basic for applications(VBA)
[79] Java
[80] Physical processor
[81] Logical processor
[82] Random access memory(RAM)
[83] MATLAB parallel computing toolbox
[84] www.csie.ntu.edu.tw/~cjlin/libsvm/
[85] www.ntu.edu.sg/home/egbhuang/
[86] Receiver operating characteristic
[87] Imperialist competitive algorithm
[88] Harmony search
[89] Cultural algorithm
[90] Anomaly
[91] Yates's correction
[92] Chi-square statistics
[93] Rule-based models
[94] Dimension reduction

www.SID.ir