فصلنامه مهندسی مالی و مدیریت اوراق بهادار
دوره چهارده، شماره پنجاه و پنج، تابستان ۱۴۰۲
نوع مقاله : علمی پژوهشی
صفحات : ۱۸۳-۱۶۵

حل مساله توزیع داده‌های نامتوازن در پیش‌بینی ورشکستگی به روش یادگیری حساس به هزینه

تاریخ دریافت مقاله : ۱۴۰۱/۰۳/۱۱ تاریخ پذیرش مقاله : ۱۴۰۱/۰۵/۱۶

سیدبهروز رضوی‌قمی[۱]
ابراهیم عباسی[۲]

چکیده

هدف از پژوهش حاضر افزودن تکنیک یادگیری حساس به هزینه به مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن به منظورکاهش خطای نوع اول و افزایش معیار میانگین هندسی دقت است تا بتوان هزینه‌های طبقه‌بندی اشتباه شرکت‌های ورشکسته را بر ذینفعان کاهش داد. بدین منظور خطای نوع اول، خطای نوع دوم و معیار میانگین هندسی دقت مدل‌های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه با مدل‌های پیش‌بینی ورشکستگی با داده به شدت نامتوازن مقایسه شدند، نمونه آماری شامل ۱۲۰۰ سال-شرکت در بازه زمانی سال ۱۳۸۰ الی سال ۱۳۹۹ شامل ٪۹۰ شرکت‌های سالم و ٪۱۰ شرکت‌های ورشکسته است، نتایج آزمون فرضیه‌ها بیانگر آن است که افزودن تکنیک یادگیری حساس به هزینه به مدل‌های پیش‌بینی ورشکستگی منجر به کاهش معنادار خطای نوع اول، افزایش معنادار خطای نوع دوم و افزایش معنادار معیار میانگین هندسی دقت مدل‌های مبتنی بر داده‌های نامتوازن در سطح اطمینان ٪۹۵ شد، همچنین با افزایش مقدار هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، خطای نوع اول روند نزولی، خطای نوع دوم روند صعودی و میانگین هندسی دقت روند صعودی دارد.

کلمات کلیدی

پیش‌بینی ورشکستگی، داده‌های نامتوازن، یادگیری حساس به هزینه، نسبت‌های مالی، بهینه‌سازی جستجو شبکه‌ای. طبقه بندی موضوعی: G33، G01

۱-گروه حسابداری، دانشکده حسابداری، موسسه آموزش عالی سناباد گلبهار، گلبهار، ایران. (نویسنده مسئول)
behroooz.razavi0065@gmail.com
۲-گروه مدیریت، دانشکده علوم اجتماعی و اقتصادی، دانشگاه الزهرا، تهران، ایران. abbasiebrahim2000@Alzahra.ac.ir

مقدمه

عمدتا مدل‌های پیش‌بینی ورشکستگی پیشنهادی با استفاده از پارادایم سنتی یا نمونه‌گیری جفتی طراحی شده‌اند (اولسن و همکاران ۲۰۱۲)، به این معنا که داده‌ها شامل تعداد برابری از شرکت‌های سالم و شرکت‌های درمانده مالی بودند، اما چنین نمونه‌گیری مطابق با واقعیت جامعه آماری شرکت‌ها نیست. در مواردی که تعداد شرکت‌های درمانده مالی نادر و تعداد شرکت‌های سالم خیلی زیاد است و پژوهشگران با داده‌های نامتوازن روبرو خواهند شد (اندرسون، ۲۰۰۷). زمیوسکی (۱۹۸۴) نشان داد که اگر نسبت یا تعداد شرکت های سالم به نسبت یا تعداد شرکت‌های درمانده مالی با واقعیت انطباق نداشته باشد یا داده‌های متوازن باشند، ممکن سوگیری در انتخاب نمونه منجر به کم‌برآوردی خطای نوع اول و بیش‌برآوردی خطای نوع دوم مدل‌های پیش‌بینی درماندگی مالی شود. مدل‌های پیش‌بینی درماندگی مالی مبتنی بر داده‌های نامتوازن در مقایسه با داده‌های متوازن با مساله بالاتر بودن خطای نوع اول روبرو خواهند شد زیرا طراحی مدل‌های پیش‌بینی ورشکستگی به گونه‌ای است که هزینه طبقه‌بندی شرکت‌های سالم برابر با هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته باشد. مدل‌های پیش‌بینی به گونه‌ای طراحی شده اند که دقت کلی بهینه شود و توزیع نسبی هر طبقه را مورد توجه قرار نمی‌دهند (لوپز و همکاران، ۲۰۱۳) اما در واقعیت هزینه خطای نوع اول در مقایسه با هزینه خطای نوع دوم مدل‌های پیش‌بینی ورشکستگی برای سرمایه‌گذاران، اعتباردهندگان و سایر ذینفعان بیشتر است. در پژوهش‌های قبلی به منظور کاهش خطای نوع اول (حل مساله داده‌های نامتوازن) از ۳ راه حل در سطح داده‌ها، در سطح الگوریتم و یادگیری حساس به هزینه[1] استفاده شده است (چاولا و همکاران، ۲۰۰۴). نوع آوری پژوهش حاضر، پیشنهاد مدل‌های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه به منظور کاهش خطای نوع اول و افزایش معیار میانگین هندسی دقت کلی است.

خطای نوع اول و خطای نوع دوم مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های متوازن در مقایسه با داده‌های نامتوازن به ترتیب کمتر از واقع و بیشتر از واقع است، از آنجایی که عمدتا در پژوهش‌های قبلی از مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های متوازن استفاده شده است لذا در پژوهش حاضر به منظور پر شدن خلا مذکور، از مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های به شدت نامتوازن با درجه ٪۹۰ به ٪۱۰ استفاده شد تا از کم‌برآوردی خطای نوع اول و بیش‌برآوردی خطای نوع دوم مدل‌های مبتنی بر داده‌های متوازن اجتناب شود، اگرچه خطای نوع اول، خطای نوع دوم و معیار دقت مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن با واقعیت جامعه آماری شرکت‌ها انطباق بیشتری دارد اما از نقطه نظر ماشین یادگیری استفاده از داده‌های نامتوازن در پیش‌بینی

ورشکستگی چالش برانگیز است، اصطلاحا به چالش مذکور، مساله داده‌های نامتوازن در پیش‌بینی ورشکستگی گفته می شود که باید حل شود، مساله مذکور اشاره به آن دارد که خطای نوع اول مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن در مقایسه با مدل‌های مبتنی بر داده‌های متوازن بیشتر است و هزینه خطای نوع اول در مقایسه با هزینه خطای نوع دوم مدل‌های مذکور برای ذینفعان بیشتر است لذا ضرورت دارد به منظور حل مساله داده‌های نامتوارن در پیش‌بینی ورشکستگی، خطای نوع اول مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن کاهش یابد تا هزینه های تحمیل شده به ذینفعان کاهش یابد. در پژوهش‌های قبلی به حل مساله مذکور توجه چندانی نشده است، بنابراین در پژوهش حاضر به منظور حل مساله مذکور، مدل های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه پیشنهاد شده است، دلیل استفاده از تکنیک یادگیری حساس به هزینه در مقایسه با روش های بازنمونه‌گیری به منظور حل مساله مذکور، اجتناب از مشکلات تغییر در توزیع داده‌های آموزشی نامتوازن است، انتظار می‌رود با افزودن تکنیک یادگیری حساس به هزینه به مدل‌های مبتنی بر داده‌های نامتوازن و تنظیم مقادیر بیشتر برای هزینه پیش‌بینی اشتباه شرکت‌های ورشکسته در مقایسه با شرکت‌های سالم، مدل پیش‌ینی مبتنی بر داده‌های نامتوازن در مرحله آموزشی به سمت شرکت‌های ورشکسته متمایل شوند و خطای نوع اول مدل‌های مبتنی بر داده‌های نامتوارن کاهش و در نتیجه هزینه‌های تحمیل شده برذینفعان کاهش و معیارمیانگین هندسی دقت افزایش یابد. از آنجایی که مقادیر متفاوت هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، اثرات متفاوتی بر تمایل مدل‌های پیش‌بینی بر شرکت‌های ورشکسته دارند، لذا در پژوهش حاضر اثر طیفی از مقادیر هزینه‌های طبقه‌بندی اشتباه مذکور بر معیارهای خطای نوع اول، خطای نوع دوم و میانگین هندسی دقت بررسی شده است.

مروری بر مبانی نظری و پیشینه پژوهش

به داده‌هایی نامتوازن[۲] گفته می شود اگر حداقل یکی از مقادیر متغیر وابسته به گونه‌ای بااهمیت، تعداد نمونه‌های کمتری از سایر مقادیر متغیر وابسته داشته باشد (ثابتاه و همکارن، ۲۰۱۸). درحوزه پیش‌بینی ورشکستگی، سناریوی مجموعه داده‌های نامتوازن، ناشی از نمونه‌های محدود در طبقه اقلیت (شرکت‌های ورشکسته) است. از آنجایی که تعداد شرکت‌های ورشکسته در مقایسه با تعداد شرکت‌های سالم کمتر است، لذا شرکت‌های ورشکسته نماینده طبقه اقلیت[۳] و شرکت‌های سالم نماینده طبقه اکثریت[۴] در داده‌های نامتوازن هستند. ضعف مدل‌های پیش‌بینی مبتنی بر داده‌های نامتوازن، ناشی از مرحله یادگیری[۵] مدل‌های پیش‌بینی است، به طوری که در طی مرحله یادگیری، مدل‌های پیش‌بینی بر دقت طبقه اکثریت (شرکت‌های سالم) متمایل می‌شوند و دقت طبقه اقلیت (شرکت‌های ورشکسته) را

نادیده می‌گیرند (کیم و آهن، ۲۰۱۵).

زمیوسکی (۱۹۸۴) نشان داد که اگر نسبت شرکت‌های سالم به شرکت‌های درمانده مالی نماینگر جامعه آماری شرکت‌ها در واقعیت نباشد، خطا یا سوگیری در انتخاب نمونه شرکت‌ها رخ می دهد که منجر به برآورد کمتر از واقع خطای نوع اول و برآورد بیشتر از واقع خطای نوع دوم خواهد شد و به طور خاص بین تعداد یا نسبت شرکت‌های درمانده مالی با نرخ خطای نوع اول رابطه منفی و بین تعداد یا نسبت شرکت های درمانده مالی با نرخ خطای نوع دوم رابطه مثبت وجود دارد. خطای نوع اول مدل‌های پیش‌بینی مبتنی بر داده‌های نامتوازن در مقایسه با داده‌های متوازن بیشتر است، از آنجایی که خطای نوع اول در مقایسه با خطای نوع دوم، هزینه بیشتری را به ذینفعان تحمیل می کند، لازم است که خطای نوع اول مدل‌های مبتنی بر داده‌های نامتوازن کاهش یابد تا هزینه‌های تحمیل شده بر ذینفعان کاهش یابد، اصطلاحا به چگونگی کاهش خطای نوع اول مدل‌های پیش بینی ورشکستگی مبتنی بر داده‌های نامتوازن، حل مساله داده‌های نامتوازن در پیش‌بینی ورشکستگی گفته می‌شود. در پژوهش‌های قبلی به منظور کاهش خطای نوع اول مدل‌های مبتنی برداده‌های نامتوازن (حل مساله داده‌های نامتوازن) از ۳ رویکرد الف: درسطح داده، ب: در سطح الگوریتم و ج: یادگیری حساس به هزینه استفاده شده است، راه‌حل در سطح داده شامل روش‌های بازنمونه‌گیری است، مهمترین روش‌های بازنمونه‌گیری شامل کم‌نمونه‌گیری، بیش‌نمونه‌گیری و ترکیبی است، راه حل در سطح الگوریتم شامل تعدیل یا اصلاح الگوریتم و یادگیری یک طبقه‌ای (چاولا و همکاران، ۲۰۰۴) و راه‌حل حساس به هزینه، متمرکز بر هزینه طبقه‌بندی اشتباه متفاوت طبقه‌های اکثریت و اقلیت است، لذا ماتریسی برای هزینه طبقه‌بندی اشتباه نیز می‌توان ایجاد کرد (لانگادج ، ۲۰۱۳).

تکنیک‌های بازنمونه‌گیری منجر به تغییر در توزیع طبقه‌های نامتوازن می شوند و مهمترین ایرادات تکنیک‌های بازنمونه‌گیری شامل کم‌نمونه‌گیری[۶] و بیش‌نمونه‌گیری[۷] به ترتیب حذف نمونه‌های بااهمیت از فرایند یادگیری مدل و بیش‌برازشی (افزایش سربار اطلاعاتی) است، محققان به منظور غلبه بر چالش مجموعه داده‌های نامتوازن و اجتناب از ایرادات روش‌های بازنمونه‌گیری، روش یادگیری حساس به هزینه را پیشنهاد دادند زیرا روش یادگیری حساس به هزینه، تغییری در توزیع نامتوازن طبقه‌ها ایجاد نمی‌کند، به دلیل آنکه در واقعیت هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته بیشتر از شرکت‌های سالم است، درحوزه پیش‌بینی ورشکستگی شرکت‌ها، ممکن است اثرات تجاری هر طبقه برای ذینفعان یکسان نباشد، در چنین مواردی، مدل طبقه‌بندی کننده باید برای هر طبقه به گونه‌ای متفاوت و بر اساس وزن همان طبقه عمل کنند، وزن متفاوت طبقه‌ها و در نتیجه رویکرد متفاوت به طبقه‌ها، طبقه‌بندی‌کننده را به

الگوریتم‌های حساس به هزینه تبدیل کرده است (ژو، ۲۰۱۱، دومینگوس، ۱۹۹۹).

رضوی و همکاران (۱۴۰۱)، اقدام به مقایسه مدل‌های پیش بینی درماندگی مالی مبتنی بر داده‌های نامتوازن با ۴ درجه متفاوت کردند، نتایج حاصل از پژوهش مذکور حاکی از آن بود بهترین عملکرد برای داده‌های نامتوازن با درجه‌های کمتر مربوط به جنگل تصادفی و بهترین عملکرد برای داده‌های نامتوازن با درجه‌های بیشتر مربوط به ماشین بردار پشتیبان است، عملکرد الگوریتم نزدیکترین همسایگی تفاوت با اهمیتی با سایر مدل ها ندارد و عملکرد شبکه عصبی مصنوعی متوسط یا مناسب در پیش بینی درماندگی مالی است. حق پرست و همکاران (۱۴۰۰) اقدام به پیش‌بینی ورشکستگی شرکت‌ها با استفاده از شبکه عصبی کانولوشن و نسبت‌های مالی تصویری کردند، نتایج حاصل از تحلیل یافته‌ها نشان داد که مدل شبکه عصبی کانولوشن در پیش‌بینی و شناخت از روی تصاویر، توانایی پیش‌بینی با دقت ۵۰٪ را دارد. حیدری و همکاران (۱۳۹۹) اقدام به پیش‌بینی ورشکستگی شرکت‌ها با استفاده از مدل شبکه عصبی مصنوعی و الگوریتم کرم شب تاب کردند، نتایج حاکی از آن است که عملکرد شبکه عصبی بهینه شده به وسیله الگوریتم کرم شب تاب در مقایسه به شبکه عصبی مصنوعی پس انتشار بهتر است. قلی زاده سالطه و همکاران (۱۳۹۸) اقدام پیش‌بینی ورشکستگی با مدل یادگیری ماشین مبتنی برکرنل بهینه شده با الگوریتم گرگ خاکستری کردند، نتایج حاصل از آزمون فرضیه‌ها حاکی از آن است که الگوریتم گرگ خاکستری عملکرد بسیار خوبی دارد و در مقایسه با الگوریتم ژنتیک عملکرد بهتری دارد. خردیار و همکاران (۱۳۹۷) با استفاده از سیستم استنتاج فازی عصبی مبتنی بر تحلیل مولفه‌های اصلی اقدام به پیش‌بینی ورشکستگی شرکت‌ها کردند، نتایج آزمون فرضیه‌ها حاکی از آن است که سیستم استنتاج فازی عصبی مبتنی بر تحلیل مولفه اصلی قابلیت پیش‌بینی ورشکستگی شرکت‌ها را دارد و زمانی که مدل مذکور با الگوریتم فراابتکاری ازدحام کبوتر ترکیب می شود، با کاهش مقدار خطا، دقت افزایش می‌یابد. بیات و همکاران (۱۳۹۷) اقدام به پیش‌بینی ورشکستگی شرکت‌ها با استفاده از الگوریتم کرم شب تاب کردند، نتایج حاکی از آزمون فرضیه‌ها بیانگر پیش‌بینی صحیح ۹۵,۱۲ ٪ در سال اول، ۸۵,۳۶٪ در سال دوم و ۸۰,۴۸٪ در سال سوم است.

گارسیا (۲۰۲۲) اقدام به پیش‌بینی ورشکستگی شرکت‌ها مبتنی بر داده‌های به شدت نامتوازن کرد، نمونه شامل ۱۸۲۴ سال-شرکت بود، نتایج مبین آن بود که اعمال تکنیک بیش‌نمونه‌گیری مصنوعی طبقه اقلیت بر داده‌های آموزشی و متوازن‌سازی داده‌های آموزشی منجر به بهود بااهمیت دقت مدل‌های طبقه‌بندی می‌شود و بهترین عملکرد مربوط به مدل‌های طبقه‌بندی مبتنی بر داده‌های آموزشی متوازن شده به روش ترکیبی بیش‌نمونه‌گیری مصنوعی طبقه اقلیت با کم‌نمونه‌گیری مبتنی بر خوشه‌بندی است.

فاریس و همکاران(۲۰۲۰) برای مقابله با مساله داده‌های نامتوازن در پیش‌بینی ورشکستگی، روش ترکیبی بیش‌نمونه‌گیری مصنوعی طبقه اقلیت با روش‌های یادگیری تجمیعی را پیشنهاد کردند، نتایج حاصل نشان داد که رویکرد پیشنهادی، راه‌حلی کارآمد برای مواجه با داده های به شدت نامتوازن است. زوریکاک و همکاران (۲۰۲۰) اقدام به پیش‌بینی ورشکستگی شرکت‌های کوچک و متوسط با داده‌های نامتوازن (با درجه‌های متفاوت) کردند، داده‌ها نشان داد که نسبت واقعی بین شرکت‌های سالم با شرکت‌های ورشکسته به شدت نامتوازن است، راه‌حل پژوهش مذکور برای غلبه بر مساله داده‌های نامتوازن، استفاده از مدل‌های پیش‌بینی یک طبقه‌ای است، نتایج نشان داد که بهترین نمره میانگین هندسی دقت کلی ٪۹۱ است. وگانزونس و سورین (۲۰۱۸) مطالعه ترکیبی پیرامون کاهش عملکرد مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن با درجه‌های متفاوت و روش‌های بهبود عملکرد (حل مساله داده های نامتوازن) مدل‌های پیش‌بینی ورشکستگی شرکت‌ها را انجام دادند، نتایج پژوهش مذکور نشان داد که توزیع نامتوازنی که در آن طبقه اقلیت۲۰٪ تعداد کل شرکت‌های منتخب باشد به طور قابل توجهی عملکرد مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن را مختل می‌کند، علاوه بر این روش‌های نمونه‌گیری می‌توانند منجر به بهبود عملکرد مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن شده است. ژو و همکاران (۲۰۱۳) اقدام به پیش‌بینی ورشکستگی شرکت‌ها با داده‌های نامتوازن کردند، بدین منظور اثر ۷ روش نمونه‌گیری بر عملکرد مدل‌های پیش‌بینی ورشگستگی بررسی شد، نتایج نشان داد که روش نمونه‌گیری مطلوب در پیش‌بینی ورشکستگی تاحد زیادی وابسته به تعداد شرکت‌های ورشکسته در نمونه داده‌های آموزشی است.

فرضیه‌های پژوهش

با توجه به مبانی نظری پژوهش، فرضیه‌ها زیر تدوین شد:

۱. معیارمیانگین‌هندسی دقت‌کلی مدل‌های‌پیش‌بینی ورشکستگی مبتنی بریادگیری‌حساس به هزینه در مقایسه با مدل‌های پیش‌بینی ورشکستگی (غیر حساس به هزینه ها) به شکل معناداری بیشتر است.

۲. خطای نوع اول (شناسایی شرکت ورشکسته به اشتباه به عنوان شرکت سالم) مدل‌های پیش بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه در مقایسه با مدل‌های پیش‌بینی ورشکستگی (غیر حساس به هزینه) به شکل معناداری کمتر است.

۳. خطای نوع دوم (شناسایی شرکت سالم به اشتباه به عنوان شرکت ورشکسته) مدل‌های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه در مقایسه با مدل‌های پیش بینی ورشکستگی (غیر حساس به هزینه) به شکل معناداری بیشتر است.

روش‌شناسی پژوهش

پژوهش حاضر از نوع کاربردی است، طرح پژوهش از نوع شبه تجربی و با استفاده از رویکرد پس رویدادی (از طریق اطلاعات گذشته) است. در پژوهش حاضر به منظور آزمون فرضیه‌ها از اطلاعات ۱۲۰۰ سال-شرکت شامل ۱۰۸۰ شرکت سالم و ۱۲۰ شرکت ورشکسته در بازه زمانی سال ۱۳۸۰ الی سال ۱۳۹۹ استفاده شده است. جامعهٔ آماری پژوهش حاضر کلیهٔ شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران به جز شرکت‌های سرمایه گذاری و بانک ها است، متغیر وابسته پژوهش بر مبنای ماده ۱۴۱ قانون تجارت تعیین شده به طوری که اگر زیان انباشته شرکتی بیشتر از ٪۵۰ سرمایه شرکت باشد، شرکت ورشکسته خواهد بود و مقدار متغیر وابسته برابر با یک است در غیر این صورت شرکت سالم خواهد بود و مقدار متغیر وابسته برابر با صفر خواهد بود. پس از مطالعه ادبیات پژوهش‌های قبلی در ارتباط با متغیرهای پیش‌بینی ورشکستگی شرکت‌ها، ۶۴ متغیر پیش‌بینی انتخاب گردید. جدول ۱ شامل فهرست و نحوه محاسبه ۶۴ متغیر پیش‌بین است.

جدول ۱ : نسبت های مالی مورد استفاده در پیش بینی درماندگی مالی (منبع: یافته‌های پژوهشگر)

ردیف	نام متغیر	ردیف	نام متغیر	ردیف	نام متغیر	ردیف	نام متغیر
۱	NI/SE	۲۰	S/FA	۳۹	CL/TL	۵۷	RE/Inv
۲	NI/TA	۲۱	S/SE	۴۰	D/NI	۵۸	RE/SC
۳	OCF	۲۲	S/TA	۴۱	EPS	۵۹	RE/TA
۴	OCF/SE	۲۳	SE/TA	۴۲	EBIT/IE	۶۰	S/Ca
۵	OCF/CL	۲۴	SE/TL	۴۳	EBIT/S	۶۱	CA/S
۶	OCF/IE	۲۵	Size(log TA)	۴۴	EBIT/TA	۶۲	CA/TA
۷	OCF/S	۲۶	TIBL/TL	۴۵	FA/(SE+LTD)	۶۳	CL/SE
۸	OCF/TA	۲۷	TL/TA	۴۶	FA/TA	۶۴	CL/TA
۹	OCF/TL	۲۸	WC/S	۴۷	GP/S		
۱۰	OCF/NI	۲۹	WC/TA	۴۸	IE/GP		
۱۱	OCF/OI	۳۰	(Ca+STI)/CL	۴۹	IE/S		
۱۲	NI/GP	۳۱	R+Inv)/TA	۵۰	Inv/WC		
۱۳	OI/S	۳۲	P/S	۵۱	Inv/S		
۱۴	OI/TA	۳۳	R/S	۵۲	LTD/SE		
۱۵	PIC/SE	۳۴	Ca/CL	۵۳	LTD/TA		
۱۶	QA/CL	۳۵	Ca/TA	۵۴	MVE/TA		
۱۷	QA/Inv	۳۶	NI/S	۵۵	MVE/TL		
۱۸	QA/TA	۳۷	CA/CL	۵۶	MVE/SE		

CA: دارایی جاری، NI : سودخالص، C: موجودی نقدی، OI: سود عملیاتی، CL: بدهی جاری، QA: دارایی‌های آنی، PIC: سرمایه پرداخت شده، R: حساب‌ها و اسناد دریافتنی، EBIT: سود قبل از بهره و مالیات، RE: سود انباشته، FA: دارایی‌های ثابت، S:درآمدها، GP= سود ناخالص، SC=سرمایه سهام، IE= هزینه‌های مالی، SE= حقوق صاحبان سهام،INV= موجودی مواد و کالا، STI= سرمایه گذاری‌های کوتاه مدت، TA= مجموع دارایی ها، LTD=بدهی‌های بلند مدت، TL=مجموع بدهی ها، MVE= ارزش بازار سهام، WC= سرمایه درگردش، OCF= جریان نقدی عملیاتی، D= سود تقسیمی، TIBL= مجموع بدهی‌های بهره.

در پژوهش جاری به منظور اجرای روش یادگیری حساس به هزینه ها از نرم افزار رپیدماینر۸ و اپراتور فراهزینه۹ استفاده شده است، دراپراتور فراهزینه، طبقه بندی کننده حساس به هزینه با استفاده از ماتریس هزینه C ایجاد می‌شود، ماتریس هزینه C دارای پارامترهای هزینه‌های طبقه‌بندی اشتباه شرکت ورشکسته و شرکت سالم است، روش استفاده شده توسط اپراتور فراهزینه مشابه با روش توصیف شده توسط دومینگوس (۱۹۹۹) است. مطابق رابطه ۱، هزینه‌های طبقه‌بندی اشتباه به ترتیب با لاندا و میو نمایش داده شده است.

$$C = \begin{bmatrix} 0 & \lambda \\ \mu & 0 \end{bmatrix}$$ رابطه ۱

به منظور تعیین پارامتر بهینه هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، از معیار میانگین هندسی دقت استفاده شد و معیار مذکور با مقادیر متفاوت هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته در بازه ۱ الی ۵ با فاصله‌های ۰,۵تایی، محاسبه شد. مقدار بهینه هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، مقداری هزینه‌ای است که به ازای آن معیار میانگین هندسی دقت مدل پیش‌بینی ورشکستگی بیشترین باشد. در ادامه به منظور مقایسه معناداری معیارهای ارزیابی عملکرد مدل‌های مبتنی بر یادگیری حساس به هزینه در مقایسه با مدل‌های پیش‌بینی ورشکستگی (غیر حساس به هزینه) در سطح اطمینان ٪۹۵، از مقدار بهینه هزینه استفاده شد. علاوه بر این در پژوهش حاضر از تابع شعاعی ماشین بردار پشتیبان استفاده شد که شامل دو پارامتر c و gamma است، پارامترهای gamma و c نقشی حیاتی در عملکرد مدل پیش‌بینی ماشین بردار پشتیبان دارند (هسو و همکاران، ۲۰۰۴). بنابراین انتخاب نامناسب دو پارامتر مذکور می تواند منجر به مساله بیش‌برازشی و کم‌برازشی شود، هدف از تنظیم پارامترهای c و gamma آن است که بتوان دقت مدل پیش‌بینی را برای داده‌های دیده نشده به حداکثر شود، رشد نمایی از دو پارامتر c و gamma روشی کاربردی در تنظیم پارامترهای بهینه است. در این پژوهش از

مقادیر 2^{-5}، 2^{-3}، 2^{-1}، 2^{1}، 2^{3}، 2^{5}، 2^{7}، 2^{9}، 2^{11}، 2^{13}، 2^{15}برای پارامتر C و از مقادیر 2^{-15}، 2^{-13}، 2^{-11}، 2^{-9}، 2^{-7}، 2^{-5}، 2^{-3}، 2^{1}، 2^{3}، 2^{5}برای پارامتر gamma استفاده شده است، لذا در کل ۱۲۱ ترکیب از پارامترها تیونیگ شده است. در پژوهش حاضر از مدل پیش‌بینی شبکه عصبی پرسپترون یک لایه استفاده شد، مقادیر نرخ یادگیری و نرخ مومنتوم برابر با ۰,۱، ۰,۲، ۰,۳، تنظیم شده است، در مجموع ۹ ترکیب از پارامترها برای شبکه عصبی مصنوعی تنظیم شده است. به منظور تعیین پارامتر بهینه مدل پیش‌بینی جنگل تصادفی، ۴معیارGain_ Ratio، Information Gain، Gini Index و Accuracy و ۲۸ مقدار از پارامتر تعداد درختان در دامنه ۱۰ الی ۱۰۰۰ و در مجموع ۱۱۲ ترکیب از دو پارامتر مذکور تیونیگ شده است، به منظور تعیین پارامترهای بهینه مدل پیش‌بینی الگوریتم k نزدیکترین همسایگی، ۲۸ مقدار از پارامتر k در دامنه ۱ الی ۱۰۰۰ تیونیگ شده است و به منظور ارزیابی عملکرد مدل‌های پیش‌بینی ورشکستگی، از معیارهای میانگین هندسی دقت، خطای نوع اول و خطای نوع دوم استفاده شده است، جدول ۲، توصیفی از معیارهای مذکور در ماتریس درهم ریختگی است:

جدول ۲ : ماتریس درهم ریختگی

ماتریس درهم ریختگی	مثبت پیش‌بینی شده	منفی پیش‌بینی شده
مثبت واقعی	مثبت صحیح (TP)	منفی اشتباه (FN)
منفی واقعی	مثبت اشتباه (FP)	منفی صحیح (TN)

اولین معیار ارزیابی عملکرد مدل‌های پیش‌بینی ورشکستگی، معیار میانگین هندسی دقت است، معیار میانگین هندسی دقت معیاری قابل اتکا در ارزیابی توان پیش‌بینی مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن است مطابق با رابطه ۲، معیار میانگین هندسی دقت شامل دقت پیش‌بینی صحیح شرکت‌های ورشکسته (حساسیت) و دقت پیش‌بینی صحیح شرکت‌های سالم (خاصیت) است، TPRدرصد نمونه‌های ورشکسته است که مدل به درستی در طبقه شرکت‌های ورشکسته و TNR درصد نمونه‌های سالم است که مدل به درستی در طبقه سالم قرار داده است (کوتسیانتیس و پینتلاس، ۲۰۰۳).

رابطه ۲ $$G - mean = \sqrt{TPR * TNR}$$

دومین معیار ارزیابی عملکرد خطای نوع اول است، رابطه ۳ حساسیت و رابطه ۴ خطای نوع اول را نمایش می‌دهد، (ونگازونس و سورین، ۲۰۱۸)، خطای نوع اول، مکمل حساسیت است.

رابطه ۳ $$\text{حساسیت} = \frac{TP}{TP + FN}$$

رابطه ۴ $$\text{خطای نوع اول} = 1 - \text{حساسیت}$$

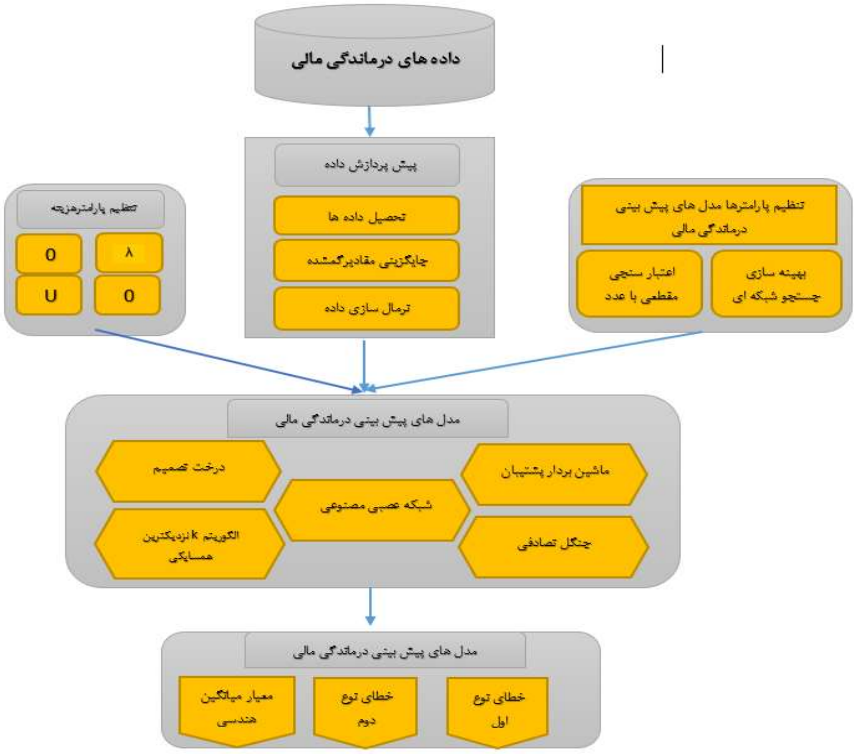
سومین معیار ارزیابی عملکرد خطای نوع دوم است، رابطه ۵ خاصیت و رابطه ۶ خطای نوع دوم را نمایش می دهد، (ونگازونس و سورین ،۲۰۱۸)، خطای نوع دوم، مکمل خاصیت است.

رابطه ۵ $$\text{خاصیت} = \frac{TN}{TN + FP}$$

رابطه ۶ $$\text{خطا نوع دوم} = 1 - \text{خاصیت}$$

مدل پیشنهادی و چارچوب تجربی آزمون پرسش پژوهش

هدف از پژوهش حاضر، پیشنهاد مدل پیش‌بینی ورشکستگی است که بتواند اثر توزیع داده‌های به شدت نامتوازن را کاهش دهد، انتظار می رود با افزودن تکنیک یادگیری حساس به هزینه‌ها، به مدل‌های پیش‌بینی ورشکستگی، عملکرد مدل‌ها بهبود یابد، مراحل اصلی روش پیشنهادی پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه به شرح نمودار ۱ است که شامل: الف) داده‌ها پیش‌پردازش خواهند شد، ب) با ایجاد ماتریس هزینه، هزینه طبقه‌بندی شرکت‌های ورشکسته در مقایسه با شرکت‌های سالم افزایش داده شد. ج) تکنیک یادگیری حساس به هزینه اشاره شده در مرحله قبلی، به مدل‌های پیش‌بینی ورشکستگی شامل شبکه عصبی مصنوعی، ماشین بردار پشتیبان، جنگل تصادفی، الگوریتم K نزدیکترین همسایگی و درخت تصمیم اضافه شد. در این مرحله به منظور تنظیم پارامترهای بهینه مدل‌ها از روش ترکیبی اعتبارسنجی مقطعی با عدد ۱۰ و بهینه‌سازی جستجو شبکه‌ای استفاده شد، د) به منظور مقایسه مدل‌های مبتنی بر یادگیری حساس به هزینه در با مدل‌های پیش‌بینی ورشکستگی، از معیارهای ارزیابی عملکرد شامل میانگین هندسی دقت کلی، خطای نوع اول و خطای نوع دوم مدل‌های پیش‌بینی ورشکستگی استفاده شده است.



نمودار ۱: مراحل روش پیشنهادی پیش بینی ورشکستگی مبتنی بریادگیری حساس به هزینه

نتایج آزمون فرضیه‌های پژوهش

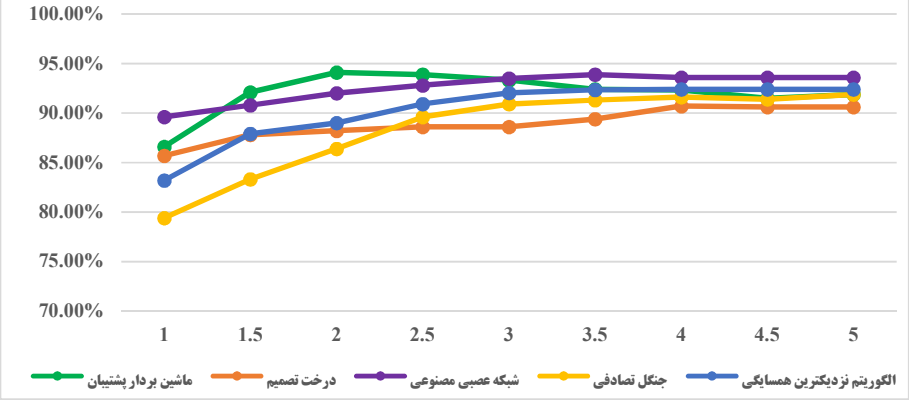
نتیجه آزمون فرضیه اول: مطابق نمودار ۲ با افزایش مقدار هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، معیار میانگین هندسی دقت کلی مدل‌های پیش‌بینی ورشکستگی به تدریج افزایش یافته است و بیشینه مقدار آن برابر با ۹۴,۱۰٪ است، با افزایش هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته به خصوص به ازای مقادیر بیشتر هزینه، معیار میانگین هندسی دقت مدل‌های پیش‌بینی ورشکستگی به تدریج همگراتر و پایدارتر شده است. مطابق نمودار ۲، در اکثر موارد با افزایش مقدار هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، میانگین هندسی دقت مدل‌های پیش‌بینی ورشکستگی افزایش داشته است مطابق جدول ۳، معیار میانگین هندسی دقت مدل‌های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه (به ازای مقدار هزینه منتخب) در مقایسه با مدل های پیش‌بینی ورشکستگی (مقدار هزینه برابر

با یک) با داده های نامتوازن در سطح اطمینان ٪۹۵ به شکل معناداری بیشتر است، لذا فرضیه اول پژوهش برای ۵ مدل پیش‌بینی ورشکستگی تایید شد.

جدول۳ : مقایسه دقت کلی مدل‌های مبتنی بر یادگیری حساس به هزینه با مدل‌های پیش‌بینی ورشکستگی

نوع مدل پیش بینی	هزینه بهینه	معیار میانگین هندسی دقت کلی		مقدار P-Value	نتیجه
		مدل‌های مبتنی بر یادگیری حساس به هزینه	مدل‌های پیش بینی ورشگستگی		
ماشین بردار پشتیبان	۲	٪۹۴,۱۰	٪۸۶,۶۰	۰,۰۰۷۸	تایید فرضیه
درخت تصمیم	۴	٪۹۰,۷۰	٪۸۵,۷۰	۰,۰۱۲۰	تایید فرضیه
شبکه عصبی مصنوعی	۳,۵	٪۹۳,۹۰	٪۸۹,۶۰	۰,۰۰۳۲	تایید فرضیه
جنگل تصادفی	۵	٪۹۱,۹۰	٪۷۹,۴۰	۰,۰۰۳۴	تایید فرضیه
الگوریتم k نزدیکترین همسایگی	۴	٪۹۲,۴۰	٪۸۳,۲۰	۰,۰۰۰۱	تایید فرضیه

منبع: یافته‌های پژوهشگر



نمودار۲ : اثر افزایش هزینه طبقه بندی اشتباه شرکت‌های ورشکسته بر معیار میانگین هندسی دقت

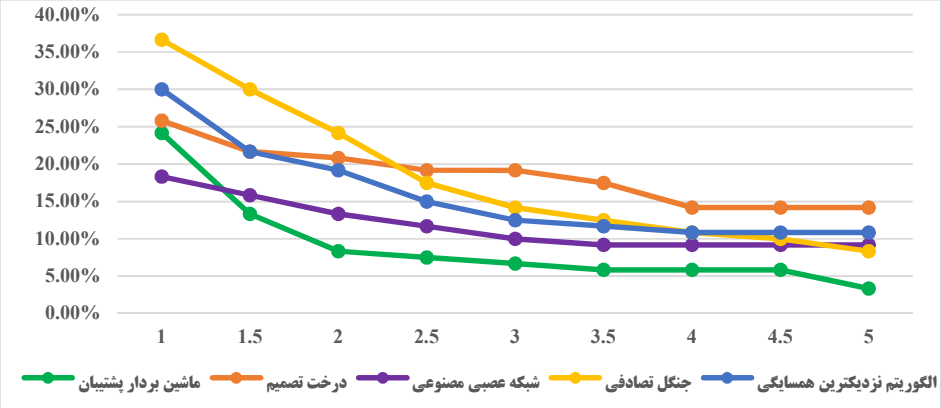
نتیجه آزمون فرضیه دوم: مطابق نمودار ۳، با افزایش مقدار هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، معیار خطای نوع اول تمامی مدل‌های پیش‌بینی ورشکستگی به تدریج کاهش یافته است و مقدار کمینه آن برابر با ٪۳,۳۳ است، همچنین با افزایش هزینه طبقه‌بندی اشتباه مذکور، به خصوص برای مقادیر هزینه ۴، ۴,۵ و ۵ معیار خطای نوع اول مدل‌های پیش‌بینی ورشکستگی به تدریج همگرا و پایدارتر شده است. مطابق نمودار ۳، در اکثر موارد با افزایش مقدار هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، خطای نوع اول مدل‌های پیش‌بینی ورشکستگی کاهش داشته است و مطابق جدول ۴، همواره

خطای نوع اول مدل‌های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه (با مقدار هزینه بهینه) در مقایسه با مدل‌های پیش‌بینی ورشکستگی (با مقدار هزینه برابر با ۱) با داده‌های به شدت نامتوازن در سطح اطمینان ٪۹۵ به شکل معناداری کمتر است، لذا فرضیه دوم پژوهش برای ۵ مدل پیش‌بینی ورشکستگی تایید شد

جدول ۴ : مقایسه خطای اول مدل‌های مبتنی بریادگیری حساس به هزینه با مدل‌های پیش‌بینی ورشکستگی

نوع مدل پیش‌بینی	هزینه بهینه	خطای نوع اول		مقدار P-Value	نتیجه
		مدل‌های مبتنی بر یادگیری حساس به هزینه	مدل‌های پیش بینی ورشگستگی		
ماشین بردار پشتیبان	۲	٪۸,۳۳	٪۲۴,۱۷	۰,۰۰۲۴۹۸	تایید فرضیه
درخت تصمیم	۴	٪۱۴,۱۷	٪۲۵,۸۳	۰,۰۰۱۳۲۲	تایید فرضیه
شبکه عصبی مصنوعی	۳,۵	٪۹,۱۷	٪۱۸,۳۳	۰,۰۰۱۰۹	تایید فرضیه
جنگل تصادفی	۵	٪۸,۳۳	٪۳۶,۶۷	۰,۰۰۰۳۸	تایید فرضیه
الگوریتم نزدیکترین همسایگی	۴	٪۱۰,۸۳	٪۳۰	۰,۰۰۰۰۳	تایید فرضیه

منبع: یافته‌های پژوهشگر



نمودار۳ : اثر افزایش هزینه طبقه بندی اشتباه شرکت‌های ورشکسته بر معیار خطای نوع اول مدل‌ها

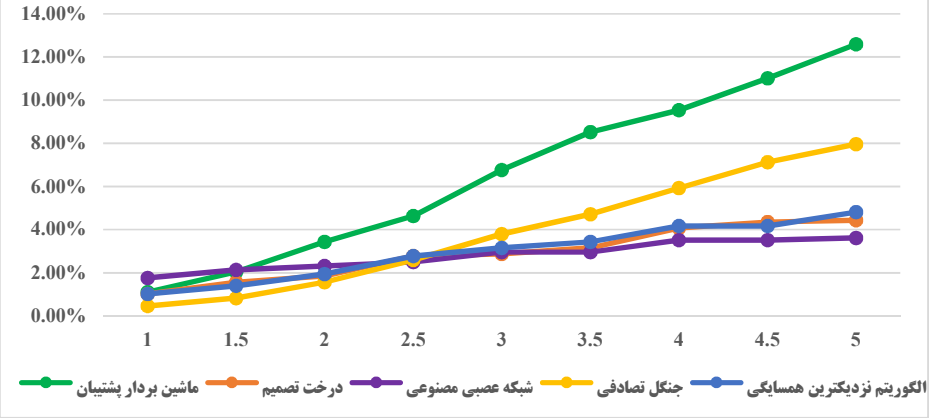
نتیجه آزمون فرضیه سوم: مطابق نمودار ۴، با افزایش مقدار هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، معیار خطای نوع دوم مدل‌های پیش‌بینی ورشکستگی به تدریج افزایش یافته است و مقدار بیشینه آن٪۱۲,۵۹رسیده است، همچنین با افزایش هزینه طبقه‌بندی اشتباه به خصوص با مقادیر ۴، ۴,۵ و ۵، معیار خطای نوع دوم مدل‌های پیش‌بینی ورشکستگی به تدریج همگرا و پایدارتر شده است. مطابق

نمودار ۴، در اکثر موارد با افزایش مقدار هزینه طبقه بندی اشتباه مذکور، معیار خطای نوع دوم مدل های پیش‌بینی ورشکستگی افزایش یافته است. مطابق جدول ۵ خطای نوع دوم مدل های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه (با مقدار هزینه منتخب) در مقایسه با مدل‌های پیش-بینی ورشگستگی (با مقدار هزینه برابر با ۱) با داده‌های نامتوازن درسطح اطمینان ٪۹۵ به شکل معناداری بیشتر است لذا فرضیه سوم پژوهش برای ۵ مدل پیش‌بینی ورشکستگی تایید شد.

جدول ۵ : مقایسه خطای دوم مدل‌های مبتنی بریادگیری حساس به هزینه بامدل‌های پیش‌بینی ورشکستگی

نوع مدل پیش‌بینی	هزینه بهینه	خطای نوع دوم		مقدار P-Value	نتیجه
		مدل‌های مبتنی بر یادگیری حساس به هزینه	مدل‌های پیش بینی ورشگستگی		
ماشین بردار پشتیبان	۲	٪۳,۴۳	٪۱,۱۱	۰,۰۰۰۰۱۹	تایید فرضیه
درخت تصمیم	۴	٪۴,۰۷	٪۱,۰۲	۰,۰۰۰۰۹۲	تایید فرضیه
شبکه عصبی مصنوعی	۳,۵	٪۲,۹۶	٪۱,۷۶	۰,۰۰۱۸۹۵	تایید فرضیه
جنگل تصادفی	۵	٪۷,۹۶	٪۰,۴۶	۰,۰۰۰۰۰۱	تایید فرضیه
الگوریتم نزدیکترین همسایگی	۴	٪۴,۱۷	٪۱,۰۲	۰,۰۰۰۰۷۷	تایید فرضیه

منبع: یافته‌های پژوهشگر



نمودار۴ : اثر افزایش هزینه طبقه بندی اشتباه شرکت های ورشکسته بر معیار خطای نوع دوم مدل‌ها

بحث و نتیجه گیری

مطابق با مبانی نظری پژوهش، اگر نسبت یا تعداد شرکت‌های سالم با نسبت یا تعداد شرکت‌های ورشکسته با واقعیت انطباق نداشته باشد یا داده‌های متوازن باشند، ممکن سوگیری در انتخاب نمونه منجر به کم‌برآوردی خطای نوع اول و بیش‌برآوردی خطای نوع دوم مدل‌های پیش‌بینی ورشکستگی مالی شود. خطای نوع اول مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن در مقایسه با داده‌های متوازن بیشتر است، از آنجایی که هزینه خطای نوع اول در مقایسه با هزینه خطای نوع دوم مدل‌های پیش‌بینی ورشکستگی به ذینفعان بیشتر است. بنابراین لازم است خطای نوع اول مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده‌های نامتوازن کاهش یابد تا هزینه‌های تحمیل شده بر ذینفعان کاهش یابد، در پژوهش حاضر به منظور کاهش خطای نوع اول مدل‌های پیش‌بینی ورشکستگی مبتنی بر داده های نامتوازن (حل مساله داده‌های نامتوازن در پیش‌بینی ورشگستگی) از راه‌حل یادگیری حساس به هزینه استفاده شده است.

نتایج حاصل از آزمون فرضیه اول حاکی از آن است که مطابق با انتظار، با افزایش هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، تمایل مدل‌های پیش‌بینی ورشکستگی در مرحله آموزش به شرکت‌های سالم کاهش و به شرکت‌های ورشکسته افزایش یافته است و در نتیجه به طور همزمان خطای نوع اول کاهش و خطای نوع دوم افزایش یافته است و در نتیجه معیار میانگین هندسی دقت به شکل معناداری در سطح اطمینان ٪۹۵ افزایش یافته است. نتایج حاصل از آزمون فرضیه اول با نتایج پژوهش قاتاشاه و همکاران (۲۰۲۰) همخوانی دارد، نتایج حاصل از آزمون فرضیه دوم حاکی از آن است مطابق با انتظار با افزایش هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، تمایل مدل‌های پیش‌بینی ورشکستگی در مرحله آموزش به شرکت‌های ورشکسته، افزایش یافته است لذا معیار خطای نوع اول مدل‌های پیش بینی ورشگستگی مبتنی بر یادگیری حساس به هزینه (با مقدار هزینه منتخب) در مقایسه مدل‌های پیش‌بینی ورشکستگی (با مقدار هزینه برابر با ۱) با داده‌های نامتوازن به شکل معناداری در سطح اطمینان ٪۹۵ کمتر است، لذا فرضیه دوم پژوهش تایید شد. نتایج حاصل از آزمون فرضیه دوم با نتایج پژوهش قاتاشاه و همکاران (۲۰۲۰) همخوانی دارد، نتایج حاصل از آزمون فرضیه سوم حاکی از آن است که مطابق با انتظار، با افزایش هزینه طبقه‌بندی اشتباه شرکت‌های ورشکسته، تمایل مدل‌های پیش‌بینی ورشکستگی در مرحله آموزش به شرکت‌های سالم کاهش یافته است و خطای نوع دوم مدل‌های مبتنی بر یادگیری حساس به هزینه (با مقدار هزینه بهینه) در مقایسه با مدل‌های پیش‌بینی ورشکستگی (با مقدار هزینه برابر با ۱) با داده‌های نامتوازن به شکل معناداری در سطح اطمینان ٪۹۵ بیشتر است، بنابراین فرضیه

سوم پژوهش تایید شد. نتایج حاصل از آزمون فرضیه سوم پژوهش با نتایج پژوهش قاتاشاه و همکاران (۲۰۲۰) همخوانی دارد، بنابراین با توجه به نتایج پژوهش می‌توان نتیجه گرفت که افزودن تکنیک یادگیری حساس به هزینه منجر به اصلاح فرایند آموزش و یادگیری مدل‌های پیش بینی ورشکستگی مبتنی بر داده های نامتوازن می شود بدون آنکه نیازی به تغییر در توزیع شرکت‌های سالم و ورشکسته باشد و اصلاح فرایند یادگیری مبتنی بر تکنیک یادگیری هزینه منجر به کاهش هزینه‌های تحمیل شده بر ذینفعان میشود، همچنین نتایج حاصل از پژوهش مبین آن است که با افزایش مقدار هزینه پیش‌بینی اشتباه شرکت‌های ورشکسته، معیار خطای نوع اول، روند نزولی و معیارخطای نوع دوم روند صعودی و معیار میانگین هندسی دقت، روند صعودی دارند، علاوه بر آن دراکثر موارد، بهترین مقدار هزینه پیش‌بینی اشتباه شرکت‌های ورشکسته در بازه ۳,۵ الی ۵ است. با توجه به موارد مذکور استفاده از مدل‌های پیش بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه با داده‌های نامتوازن به عنوان سیستم هشداردهنده زودهنگام، ابزار تخصیص بهینه منابع، ابزار ارزیابی تداوم فعالیت شرکت‌ها به ذینفعان پیشنهاد می شود.

پیشنهادها:

مطابق با مبانی نظری و نتایج پژوهش، استفاده از مدل‌های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه‌ها به ذینفعان پیشنهاد می شود زیرا کاهش خطای نوع اول می تواند منجر به کاهش هزینه مطالبات مشکوک الوصول بانک‌ها و موسسات مالی اعتباری و کاهش هزینه کاهش ارزش سرمایه‌گذاری شود. از آنجایی که نسبت شرکت‌های سالم به شرکت‌های ورشکسته (عدم توازن داده ها) در طی دوره های زمانی متفاوت و نوع صنعت متغیر است و به دلیل آن معیارهای ارزیابی عملکرد مدل های پیش‌بینی درماندگی مالی نامتوازن وابسته به درجه های عدم توازن داده‌ها است، ذینفعان باید مدل‌های پیش‌بینی ورشکستگی را متناسب با درجه عدم توازن داده‌ها در دوره زمانی مربوطه یا صنعت مورد نظر انتخاب کنند لذا پیشنهاد می‌شود معیارهای ارزیابی عملکرد مدل های پیش‌بینی ورشکستگی مبتنی بر یادگیری حساس به هزینه ها با سایر مدل‌های پیش‌بینی ورشکستگی برای طیفی از داده‌های نامتوازن با درجه‌ها از ٪۶۰ به ٪۴۰ الی ٪۹۵ به ٪۵ بررسی شود.

منابع

۱) بیات، علی، احمدی، سید علیرضا و محمدی، مجید. (۱۳۹۷)، پیش‌بینی ورشگستگی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از الگوریتم کرم شب تاب، مهندسی مالی و مدیریت اوراق بهادار، ۹(۳۷)، صص ۲۳۴-۲۶۲.

۲) حق پرست، عباسعلی، مومنی، علیرضا، گرد، عزیز و منصوری، فردین، (۱۴۰۰) نسبت‌های مالی تصویری و پیش‌بینی ورشکستگی شرکت‌ها با استفاده از مدل شبکه‌های عصبی کانولوشن، مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۶)، صص۵۵۸-۵۷۵.

۳) حیدری، مهدی، زیاری، شکراله، شایان نیا، سید احمد و رشیدی کمیجان، علیرضا. (۱۴۰۰)، پیش‌بینی ورشکستگی شرکت‌های بورس اوراق بهادار با استفاده از شبکه عصبی مصنوعی و الگوریتم کرم شب تاب، مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۶)، صص ۶۹۱- ۷۱۶.

۴) خردیار، سینا، قلیزاده، محمدحسین و لطفی، فروغ. (۱۳۹۷)، پیش‌بینی درماندگی مالی بااستفاده از روش ترکیبی PCA-ANFIS و الگوریتم فراابتکاری بهینه سازی ازدحام کبوتر، مهندسی مالی و مدیریت اوراق بهادار، ۹(۳۷)، صص ۱۳۳-۱۵۷.

۵) قلی زاده، سالطه، توحید، اقبال‌نیا، محمدو آقابابایی، محمد ابراهیم. (۱۳۹۸)، پیش‌بینی ورشکستگی با مدل یادگیری ماشین سریع مبتنی بر کرنل بهینه شده با الگوریتم گرگ خاکستری، تحقیقات مالی، صص، ۲۱(۲)، صص ۱۸۷-۲۱۲.

6) Chawla, Nitesh V., Japkowicz, Nathalie, Kotcz, Aleksandr., (2004), Editorial: Special issue on learning from imbalanced data sets. ACM Sigkdd Explorations Newsletter, 6 (1), 16.

7) Domingos, Pedro ., (1999), Meta cost a general method for making classifiers cost-sensitive. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD, 99, pp. 155–164.

8) Faris, Hossam., Abukhurma, Ruba., Waref, Almanasee,. Saadeh, Mohammed., Mora, Antonio M., Castillo, Pedro A., Aljarah, Ibrahim (2020), improving financial bankruptcy prediction in a highly imbalanced class distribution using oversampling and ensemble learning: a case from the Spanish market, Artificial Intelligence, P.31-53.

9) Garcia, John., (2022), Bankruptcy prediction using synthetic sampling. Machin learning with Applications, 9, pp. 1-11.

10) Ghatasheh, Nazeeh ., Hossam, Faris., Abukhurma, Ruba ., Castillo, Pedro A., Al-Madi, Nailah ., Mora, Antonio M., Hassanat, Ahmad., (2020), Cost-sensitive ensemble methods for bankruptcy prediction in a highly imbalanced data distribution: a real case from the Spanish market, Progress in Artificial Intelligence. 10(7), pp. 25-81

11) Hsu, Chih-Wei., Chang, Chih-Chung., Lin, Chih-Jen., (2004), A Practical Guide to Support Vector Classification. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University.

12) Kim, Taehoon., Ahn, Hyunchul., (2015). A hybrid undersampling approach for better bankruptcy prediction, Journal of Intelligence and Information Systems, 21(2), pp.173-190.

13) Kotsiantis Sotiris., Pintelas, P., (2003), Mixture of expert agents for handling imbalanced data sets, Annals of Mathematics, Computing & Tele Informatics 1. pp. 46–55.

14) Lopez, Victoria., Fernández, Alberto., García, Salvador., Palade, Vasile., Herrera, Francisco., (2013), An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences, 250, pp. 113-141

15) Olson, David L., Delen, Dursun., Meng, Yanyan., (2012), Comparative analysis of data mining methods for bankruptcy prediction, Decision Support Systems, 52(2), pp. 464-473.

16) Razavi Ghomi, S., Mehrazin, A., Shoorvarzi, M., Masih Abadi., A, (2022), Experimental Comparison of Financial Distress Prediction Models Using Imbalanced Datasets, Advances in Mathematical Finance and Applications, 7(3), pp.811-835.

17) Thabtah, Fadi., Kamalov, Firuz ., Rajab, Khairan., (2018), A new computational intelligence approach to detect autistic features for autism screening, International Journal of Medical Infromatics, 117, pp.112-124.

18) Veganzones, David., Severin, Eric ., (2018), An investigation of bankruptcy prediction in imbalanced datasets. Decision Support System, 112, pp.111-124.

19) Zhou, Ligang., (2013). Performance of corporate bankruptcy prediction models in imbalanced dataset: The effect of sampling methods, Knowledge-Based Systems, 41, pp. 16-25.

20) Zhou, Zhi-Hua., (2011), Cost-sensitive learning. In: International Conference on Modeling Decisions for Artificial Intelligence, pp. 17–18.
21) Zmijewski, M, E., (1984), Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research, 22, pp.59–82
22) Zoricák, Martin., Gnip, Peter., Drotár, Peter., Gazda, Vladimír., (2020), Bankruptcy prediction for small and medium-sized companies using severely imbalanced datasets. Economic Modelling. 84, pp. 165-180.

یادداشت‌ها :

1 Cost-Sensitive Learning
2 Imbalanced Dataset
3 Minority Class
4 Majority Class
5 Learning Phase
6 Undersampling
7 Oversampling
8 Rapid Miner
9 Meta Cost