فصلنامه بورس اوراق بهادار شمارهٔ ۴۸ زمستان ۹۸ سال دوازدهم ـ تاریخ دریافت: ۹۷/۱۲/۲۶ ـ تاریخ پذیرش: ۹۸/۰۶/۰۹ ـ ص ۵-۳۸

پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار تهران با استفاده از رگرسیون لجستیک لاسو

محمد نمازی[۱]

شهلا ابراهیمی [۲]

چکیده

توانایی پیش‌بینی درماندگی مالی شرکت برای تک تک واحدهای تجاری و همچنین، برای کل اقتصاد مهم است. از این رو، هدف این مقاله، کشف درماندگی مالی بالقوه و هشدار زودهنگام درماندگی مالی قریب‌الوقوع شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار است. بدین منظور، دامنه گسترده‌ای از ویژگی‌ها از جمله متغیرهای حسابداری تعهدی، حسابداری نقدی، بازار سهام، مکانیسم‌های حاکمیت شرکتی و شاخص‌های اقتصاد کلان برای پیش‌بینی درماندگی مالی شرکت‌های نمونه شناسایی شده‌اند. نمونه نهایی شامل ۴۲۱ شرکت و در نتیجه، ۳۶۷۰ شرکت-سال مشاهده است. سپس، داده آماده شده با استفاده از نسبت ۷۰ به ۳۰ به مجموعه داده آموزشی و آزمایشی تفکیک شد.

در این پژوهش، تکنک‌های پیش پردازش داده یادگیری ماشین نظیر استانداردسازی نمره Z، وان-هات انکدینگ، اعتبارسنجی متقابل K لایه طبقه‌ای، همراه با مهندسی ویژگی برای بهبود عملکرد طبقه‌بندی کننده بکار گرفته شدند. روش اعتبارسنجی متقابل K لایه طبقه‌ای با (K=۵) برای برآورد عملکرد پیش‌بینی مدل طی مرحله آموزش استفاده شد. طی مرحله آموزش، میزان‌سازی اَبرپارامتر مدل با استفاده از جستجوی حریص انجام شد. افزون بر این، رویکرد فرایادگیری حساس به هزینه همراه با معیار مختص مسائل نامتوازن یعنی نمره $F1$ برای غلبه بر مسأله نامتوازنی افراطی کلاس‌ها استفاده شده است. بر اساس نتایج تجربی، مدل لجستیک لاسو به نمره $F1$، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۵۰٪، ۵۰٪، ۷۳٪ و ۳۸٪ بر روی مجموعه آموزشی دست یافت. سرانجام، مدل پیشنهادی بر روی مجموعه آزمایشی کنار گذاشته شده آزمون شد که به نمره $F1$، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۵۱٪، ۵۱٪، ۷۳٪ و ۳۸٪ بر روی مجموعه آزمایشی منجر شد.

واژگان کلیدی: پیش‌بینی درماندگی مالی، رگرسیون لجستیک لاسو ، یادگیری ماشین، داده‌کاوی، بورس اوراق بهادار تهران.

طبقه‌بندی موضوعی: G33، G38، G34، E44، C19، M40

---

۱. استاد حسابداری دانشگاه شیراز، ایران mnamazi@rose.shirazu.ac.ir

۲ . کارشناس ارشد مدیریت مالی دانشگاه شهید بهشتی ـ نویسنده مسئول ـ shebrahimi_3622@yahoo.com

۱- مقدمه

درماندگی اقتصادی به معنای ناتوانی شرکت در کسب منابع کافی برای داوم عملیات معمولش است. عواملی نظیر هزینه‌های زیاد، کاهش تقاضای مشتریان و مدیریت مالی ضعیف می‌تواند موجب درماندگی اقتصادی شود. ورشکستگی یکی از پیامدهای احتمالی درماندگی اقتصادی شرکت است. سایر پیامدهای درماندگی اقتصادی عبارت‌اند از انحلال، تصفیه، ادغام، تجدید ساختار یا تداوم عملیات. بنا به گفته کورمایر و همکاران (۱۹۹۵: ۲۰۳) هر یک از این رویدادها بیانگر پایان شرکت است. پایان شرکت به عنوان عدم تداوم اقتصادی تعریف می‌شود. ارائه مدلی برای پیش‌بینی هر یک از این پیامدهای بسیار متفاوت، مسأله بسیار دشواری است؛ زیرا شناسایی و تعیین زمان هر یک از این رویدادها اغلب به راحتی قابل تعیین نیست (Lensberg, Eilifsen & McKee, 2006: 677-679).

در شرایط اقتصادی کنونی، سیستم‌های هشدار ورشکستگی زودهنگام به صورت ابزارهای با اهمیت کلیدی درآمده‌اند؛ زیرا توان بالقوه آن‌ها جهت اجتناب از زیان‌رسانی به سهامداران، بستانکاران، مدیران و سایر اشخاص ذینفع موجب تضمین ثبات اقتصادی می‌شود. از اواخر دهه ۱۹۶۰، روش‌های زیادی برای ارائه سیستم‌های پیش‌بینی ورشکستگی ارائه شده است. بیشتر این روش‌ها دارای این ویژگی مشترک هستند که به تکنیک‌های آماری یا هوشمند چندمتغیره‌ای تکیه می‌کنند که متغیرهای ورودی آن عمدتاً ویژگی‌های مالی شرکت‌ها است (Andrés, Landajo & Lorca, 2012: 67).

نسبت‌های مالی به طور گسترده‌ای در متون پیشین برای پیش‌بینی ورشکستگی مورد استفاده قرار گرفته‌اند (Xu, Fu & Pan, 2019: 1). دسترسی به نسبت‌های مالی به دلیل قابلیت دسترسی به آن‌ها در صورت‌های مالی شرکت‌ها، ساده است. استفاده از نسبت‌های مالی در تمایز بین شرکت‌های درمانده و سالم از اواسط دهه ۱۹۳۰ با پژوهش ویناکار و اسمیت[3] (۱۹۳۵) درباه بکارگیری نسبت‌های مالی برای ارزیابی سلامت مالی شرکت‌ها شروع شد. اگرچه نسبت‌های مالی می‌توانند بیانگر سلامت مالی شرکت باشند، کیفیت این نسبت‌های در معرض تهدیدهای جدی در خصوص سودمندی آن‌ها است؛ زیرا این اطلاعات ممکن است دستکاری شوند (Waqas & Md-Rus, 2018: 3). نسبت‌های مالی، داده‌های تاریخی هستند که تفاوت مالی شرکت‌ها را به موقع منعکس نمی‌کنند. به ویژه، زمانی که محیط عملیات شرکت به سرعت تغییر می‌کند، نقش نسبت‌های مالی در پیش‌بینی درماندگی مالی احتمالاً کاهش خواهد یافت (Xu et al., 2019: 1).

[3] Smith & Winakor

همچنین، جریان‌های نقدی نیز برای بقای شرکت مهم هستند. تئوری جریان نقدی بیانگر این است که شرکت در صورتی از نظر مالی توانمند خواهد بود که بتواند جریان نقدی کافی از محل عملیات خود ایجاد کند و هنگامی که نتواند جریان نقدی ورودی کافی از محل عملیات خود ایجاد کند، درمانده خواهد شد (Waqas & Md-Rus, 2018: 3). از دیدگاه شهودی، شرکت‌های دارای جریان‌های نقدی کمتر و پر نوسان تر و ذخایر دارایی‌های نقد کمتر جهت پوشش کمبودهای بالقوه باید از احتمال درماندگی بیشتری برخوردار باشند (Klobucnik et al., 2017: 9).

توانایی پیش‌بینی این مدل‌ها در چند دهه گذشته تنزل یافته است. از این رو، پژوهشگران استدلال می‌کنند که متغیرهای توضیحی بیشتری در مدل‌ها گنجانده شوند. کاهش توانایی پیش‌بینی این متغیرها را می‌توان به عواملی نظیر ادراک سرمایه‌گذاران و اعتباردهندگان از افزایش اختیار مدیران در محیط‌های حسابداری ارزش منصفانه نسبت داد (Cenciarelli et al, 2018: 2). برای غلبه بر بسیاری از محدودیت‌های مدل‌های مبتنی بر حسابداری و در نظر گرفتن ماهیت ناهمگن درماندگی مالی، متون اخیر دربرگیرنده اطلاعات خاصی از بازار سهام نیز است (Klobucnik, Miersch & Sievers, 2017: 7). منطق تئوریکی معیارهای سنجش قیمت بازار، جذاب است. صورت‌های مالی دارای چندین محدودیت اساسی هستند. برای نمونه، این صورت‌ها گذشته‌نگر هستند و بر مبنای مفروضات تداوم فعالیت و محافظه‌کاری تهیه شده‌اند که سودمند بودن آن‌ها را محدود می‌کند. افزون بر این، داده‌های صورت‌های مالی دربرگیرنده اثرات نوسان نیستند. اثرات نوسان می‌تواند در مدل‌های پیش‌بینی ورشکستگی شرکت دارای ارزش پیش‌بینی باشد. مهم‌تر آنکه ارزش بازار سهام نشان‌دهنده مبلغی است که ارزش دارایی‌ها می‌تواند کاهش یابد و همچنان دارایی‌ها برای پوشش ارزش فعلی پرداخت‌های بدهی کافی باشد (Jones, 2017: 6).

افزون بر این، رسوایی‌های مالی اخیر انرون و وردکام شواهدی است دال بر اینکه ساختارهای حاکمیتی ضعیف، عامل کلیدی دستکاری داده‌های مالی و افزایش احتمال ورشکستگی است. بنابراین، حاکمیت شرکتی دارای توان بالقوه تأثیرگذاری بر عملکرد شرکت و سوق آن به درماندگی است (Shah, 2016: 25).

لازم به ذکر است که شرکت‌ها نمی‌توانند بر وضعیت کلی اقتصاد کشور و سیاست‌های اقتصاد کلان مقام‌های دولتی تأثیر بگذارند؛ اما این عوامل بر وضعیت مالی شرکت نظیر نقدینگی و توانایی پرداخت آن تأثیر می‌گذارند. در نتیجه، شاخص‌های مالی شرکت حاوی اطلاعات کافی درباره وضعیت اقتصادی پیش روی مدیریت شرکت و سرمایه گذاران نیستند. از این رو، بکارگیری

شاخص‌های اقتصاد کلان نیز توصیه می‌شود (Fontaine Rezende, Montezano, Nascimento de Oliveira, & Lameira, 2017: 394).

به دلیل عدم وجود تئوری ورشکستگی اقتصادی زیربنایی، جستجو برای مدل‌های پیش‌بینی ورشکستگی صحیح‌تر کماکان در زمینه پیش‌بینی ورشکستگی به عنوان یک آرمان مطرح است. در حقیقت، پیش‌بینی ورشکستگی را می‌توان در قالب مسأله طبقه‌بندی در نظر گرفت (Chen et al., 2011: 1348). طبقه‌بندی به معنای تفکیک مجموعه‌های متمایزی از موضوعات یا مشاهدات و تخصیص موضوعات یا مشاهدات به گروه‌های تعریف شده قبلی است (Jo & Han, 1996: 415).

تمامی پژوهش‌های داخلی و بسیاری از پژوهش‌های خارجی پیش‌بینی ورشکستگی و درماندگی مالی (برای نمونه، آلتمن[4]، ۱۹۶۸؛ پندهارکار[5]، ۲۰۰۵؛ جاردین[6]، ۲۰۱۰) از روش نمونه‌گیری جفت تطبیقی استفاده کرده‌اند. عیب رویکرد تطبیق این است که بررسی اثرات بخش صنعت، اندازه شرکت یا سال ورشکستگی بر احتمال ورشکستگی امکان‌پذیر نیست (Lennox, 1999: 348; Chancharat, 2008: 116). دست‌کم، به نظر می‌رسد که گنجاندن این متغیرها به عنوان پیش‌بینی کننده، سودمندتر از بکارگیری آن‌ها برای اهداف تطبیق باشد (Ohlson, 1980: 112). افزون بر این، استفاده از نمونه‌های با اندازه کوچک می‌تواند به برازش بیش از اندازه منجر شود (Lennox, 1999: 348; Chancharat, 2008: 116). تا کنون، هیچ کدام از پژوهش‌های داخلی و خارجی، فهرست جامعی از انواع متغیرهای پیش‌بینی کننده درمانده مالی شامل متغیرهای مبتنی بر حسابداری تعهدی، حسابداری نقدی، بازار، حاکمیت شرکتی و اقتصاد کلان را به صورت همزمان بکار نگرفته‌اند. همچنین، هیچ کدام از پژوهش‌های داخلی از نمونه آزمایشی کنار گذاشته شده استفاده نکرده‌اند. بنابراین، توازن بین واریانس و سوگیری[7] مدل مورد ارزیابی قرار نگرفته است. همچنین، هیچ یک از پژوهش‌های داخلی و خارجی پیش‌بینی درماندگی مالی به تنظیم اَبرپارامترهای[8] مدل نپرداخته‌اند و معیارهای ارزیابی عملکرد محدودی را برای ارزیابی مدل پیشنهادی خود ارائه کرده‌اند. افزون بر این، مدل‌های پیشنهادی در پژوهش‌های داخلی و برخی پژوهش‌های خارجی، به دلیل بکارگیری نمونه‌های کوچک، عدم انتخاب تصادفی نمونه و انتخاب جانبدارانه شرکت‌های دارای عملکرد بسیار خوب یا بسیار ضعیف،

4 Altman
5 Pendharkar
6 Jardin
7 Bias
8 Hyperparameters

عدم کنارگذاری بخشی از داده به عنوان داده آزمایشی و انجام پیش‌پردازش و انتخاب ویژگی اولیه بر روی کل داده، در معرض برازش بیشتر از اندازه مدل قرار دارند.

از این رو، هدف این پژوهش، ارائه هشدار زودهنگام درماندگی مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از داده‌های مبتنی بر ترازنامه و صورت سود (زیان)، صورت گردش وجوه نقد، بازار سهام و نقدشوندگی، حاکمیت شرکتی و متغیرهای کلان اقتصادی به ذینفعان است تا آن‌ها بتوانند اقدامات مقتضی جهت پیشگیری از درماندگی مالی و همچنین، کاهش هزینه‌ها و زیان‌های ناشی از درماندگی مالی انجام دهند. پرسش‌های مهمی که در اینجا مطرح می‌شوند، این است که کدام یک از انواع متغیرهای حسابداری، جریان وجوه نقد، اطلاعات بازار و نقدشوندگی، حاکمیت شرکتی و اقتصاد کلان، بهترین متغیرهای پیش‌بینی کننده درماندگی مالی هستند؟ در حقیقت، باید تعیین نمود که اثربخش‌ترین مدل یادگیری ماشین جهت پیش‌بینی درماندگی مالی چیست؟ انگیزه اصلی این پژوهش، پاسخ به پرسش‌های بالا است.

۲- مروری بر پیشینه پژوهش

۲-۱- پیشینه پژوهش‌های داخلی

اعتمادی، انواری رستمی و فرج زاده دهکردی (۲۰۰۹) با بکارگیری برنامه‌نویسی ژنتیک و تحلیل تشخیصی چندمتغیره به ارائه مدلی برای پیش‌بینی ورشکستگی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران پرداختند. نمونه مورد بررسی آن‌ها شامل ۷۲ شرکت ورشکسته و ۷۲ شرکت غیرورشکسته طی دوره زمانی ۱۹۹۸–۲۰۰۸ بود. مدل برنامه‌نویسی ژنتیک آن‌ها به نرخ صحتی برابر با ۹۴٪ و ۹۰٪ به ترتیب در خصوص نمونه‌های آموزشی و آزمایشی دست یافت در حالی که نرخ صحت مدل تحلیل تشخیصی چندمتغیره در نمونه‌های آموزشی و آزمایشی به ترتیب برابر با ۷۷٪ و ۷۳٪ بود. آزمون مک‌نمار نیز نشان داد که عملکرد رویکرد برنامه‌نویسی ژنتیک در خصوص مسأاله ورشکستگی شرکت، بهتر از عملکرد تحلیل تشخیصی چندمتغیره است.

مخاطب رفیعی، منظری و بوستانیان (2011) به پیش‌بینی ورشکستگی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از شبکه عصبی، الگوریتم ژنتیک و تحلیل تشخیصی چند متغیره پرداختند. نتایج حاصل از آزمون این مدل‌ها بر روی نمونه‌ای شامل ۵۸ شرکت ورشکسته و ۱۲۲ شرکت غیرورشکسته طی سال مالی منتهی به ۲۰۰۸ بیانگر این است که شبکه عصبی، الگوریتم ژنتیک و تحلیل تشخیصی چند متغیره به ترتیب دارای بیشترین صحت پیش‌بینی بودند.

آشوری و محمدی (2011) با استفاده از تکنیک‌های درخت طبقه‌بندی و رگرسیون و شبکه عصبی پرسپترون چند لایه به ارائه مدلی برای پیش‌بینی ورشکستگی شرکت‌های پذیرفته شده در بورس سهام تهران پرداختند. نمونه مورد بررسی آن‌ها شامل ۶۶ شرکت ورشکسته و ۶۶ شرکت غیرورشکسته طی دوره زمانی ۲۰۰۸-۱۹۹۸ بود. آن‌ها عملکرد شبکه عصبی پرسپترون چند لایه در صورت عدم بکارگیری درخت طبقه‌بندی و رگرسیون برای انتخاب متغیرهای ورودی را با عملکرد شبکه عصبی پرسپترون چند لایه در صورت بکارگیری درخت طبقه‌بندی و رگرسیون برای انتخاب متغیرهای ورودی مقایسه نمودند. آن‌ها دریافتند که بکارگیری درخت طبقه‌بندی و رگرسیون تنها موجب کاهش خطای نوع ۲ شده و تأثیر مثبتی بر خطای نوع ۱ و صحت پیش‌بینی مدل ندارند.

خواجوی و قدیریان آرانی (۱۳۹۶) به بررسی نقش توانایی مدیریت در بهبود عملکرد مدل‌های پیش‌بینی بحران مالی مبتنی بر نسبت‌های حسابداری پرداختند. نتایج حاصل از بکارگیری الگوریتم‌های طبقه‌بندی کننده بوستینگ، بگینگ و جنگل‌های چرخشی بر روی نمونه‌ای شامل ۲۰۱ سال-شرکت بحران‌زده و ۲۰۱ سال-شرکت سالم پذیرفته شده در بورس اوراق بهادار تهران در دوره زمانی ۱۳۹۴-۱۳۸۵ بیانگر این است که افزودن متغیر توانایی مدیریت به نسبت‌های حسابداری موجب بهبود معنادار عملکرد الگوریتم‌های بکار گرفته شده می‌شود. به ویژه، هر چقدر که فاصله دوره زمانی پیش‌بینی با وقوع بحران مالی بیشتر باشد، گنجاندن توانایی مدیریت در مدل‌های پیش‌بینی بحران مالی به میزان بیشتری عملکرد آن‌ها را بهبود می‌دهد.

۲-۲- پیشینه پژوهش‌های خارجی

اُهلسن (۱۹۸۰) با استفاده از تحلیل لوجیت به محاسبه احتمال ورشکستگی ۱۰۵ شرکت ورشکسته و ۲۰۵۸ شرکت غیرورشکسته طی دوره ۱۹۷۶-۱۹۷۰ پرداخت. نتایج حاصل از آزمون مدل وی بیانگر این بود که صحت عملکرد مدل پیشنهادی در یک سال قبل از ورشکستگی برابر با ۹۶/۱۲٪ است. افزون بر این، چهار عامل اندازه شرکت، ساختار مالی بر حسب اهرم، معیار عملکرد و نقدینگی جاری برای ارزیابی احتمال ورشکستگی معنادار هستند.

زمیجوئسکی (Zmijewski, 1984) تأثیر جانبداری نمونه‌گیری جفت تطبیقی و جانبداری انتخاب نمونه‌های دارای داده‌های کامل بر مدل پیش‌بینی ورشکستگی را مورد بررسی قرار داد. نتایج حاصل از آزمون مدل پروبیت پیشنهادی وی بر روی نمونه‌های مختلف نشان‌دهنده این بود که پیش‌بینی مدل دارای نمونه جفت تطبیقی جانبدارانه است، این جانبداری با افزایش احتمال ورشکستگی نمونه به

احتمال ورشکستگی جامعه کاهش می‌یابد و جانبداری در صورت بکارگیری تکنیک‌های تخمین مناسب کاهش می‌یابد. همچنین، نتایج وی موید وجود جانبداری در مدل‌های پیش‌بینی ورشکستگی است که شرکت‌های دارای داده‌های ناقص را کنار می‌گذارند و صرفاً شرکت‌های دارای داده‌های کامل را بررسی می‌کند. با این وجود، در هر دو مورد، نتایج بیانگر تغییر معنادار در نرخ طبقه‌بندی و پیش‌بینی کلی نبود و صرفاً نرخ پیش‌بینی و طبقه‌بندی هر یک از گروه‌های ورشکسته و غیرورشکسته به گونه معناداری تحت تأثیر قرار می‌گرفت.

تینُکو و ویلسان (Tinoco & Wilson, 2013) به بررسی مطلوبیت ترکیب داده‌های حسابداری، داده‌های مبتنی بر بازار و داده‌های اقتصاد کلان برای پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در بورس سهام بریتانیا پرداختند. نتایج مدل‌های برآوردی آن‌ها با استفاده از رگرسیون لجستیک بر روی نمونه ای شامل ۲۳۲۱۸ شرکت-سال طی دوره زمانی ۱۹۸۰-۲۰۱۱ بیانگر این است که گنجاندن متغیرهای بازار و اقتصاد کلان در مدل داده‌های حسابداری موجب بهبود عملکرد مدل می‌شود.

مای، تیان، لی و ما (Tian et al., 2019) توانایی مدل‌های یادگیری عمیق برای پیش‌بینی ورشکستگی با استفاده از داده‌های مبتنی بر حسابداری، داده‌های مبتنی بر بازار و داده‌های ساختار نیافته برگرفته از تجزیه و تحلیل و بحث مدیریت در فرم 10-K را مورد بررسی قرار دادند. مجموعه داده مورد بررسی آن‌ها شامل ۱۱۸۲۷ شرکت ایالات متحده آمریکا یا ۹۴۹۹۴ سال-شرکت طی دوره زمانی ۱۹۹۴-۲۰۱۴ است که از این بین ۴۷۷ شرکت-سال ورشکسته و مابقی شرکت- سال‌ها سالم است. یافته‌های آن‌ها بیانگر این است که هنگام بکارگیری داده‌های عددی به عنوان پیش‌بینی کننده، هیچ دلیل متقاعدکننده‌ای برای استفاده از یادگیری عمیق وجود ندارد و مدل‌های داده‌کاوی سنتی به ویژه، مدل‌های غیرخطی نظیر ماشین بردار پشتیبان و جنگل تصادفی از توانایی مشابهی برخوردارند. افزون بر این، در صورت بکارگیری داده‌های متنی به عنوان پیش‌بینی کننده، عملکرد مدل یادگیری عمیق با استفاده از لایه‌های تعبیه‌ای[9] بهتر از رگرسیون لجستیک، ماشین بردار پشتیبان و جنگل تصادفی است. مهم‌تر اینکه آن‌ها دریافتند که هنگام پیش‌بینی ورشکستگی، داده‌های متنی می‌تواند مکمل داده‌های سنتی مبتنی بر حسابداری و داده‌های مبتنی بر بازار باشد و صحت پیش‌بینی مدل یادگیری عمیق در صورت بکارگیری ورودی‌های عددی همراه با ورودی‌های متنی بیشتر از سایر مدل‌ها است.

9 Embedding Layers

۳- پرسش‌های پژوهش

با توجه به پیشینه و مبانی نظری، این پژوهش به بررسی این می‌پردازد که:

۱- آیا متغیرهای حسابداری، جریان وجوه نقد، بازار و نقدشوندگی، حاکمیت شرکتی و اقتصاد کلان می‌توانند بین شرکت‌های درمانده مالی و شرکت‌های موفق مالی تمایز ایجاد کنند؟ به بیان دقیق‌تر، آیا این متغیرها، پیش‌بینی کننده درماندگی مالی هستند؟

۲- کدام یک از متغیرهای حسابداری، جریان وجوه نقد، بازار و نقدشوندگی، حاکمیت شرکتی و اقتصاد کلان، برای استفاده‌کنندگان در پیش‌بینی درماندگی مالی مفید هستند؟

٤- روش پژوهش

٤-١- نوع مطالعه و روش بررسی پرسش‌ها

از آنجا که این پژوهش، مبتنی بر داده است و نتایج آن از طریق مشاهده قابل تأیید است، در قالب پژوهش تجربی قرار می‌گیرد (Kothari, 2004: 4). افزون بر این، با توجه به اینکه درصدد پاسخگویی به سؤال‌های دنیای واقعی است، کاربردی (Johnson & Christensen, 2014: 52-53) و به دلیل بکارگیری داده‌های کمّی، در زمره پژوهش کمّی محض جای می‌گیرد (Johnson & Christensen, 2014: 81). همچنین، این پژوهش، پژوهش غیرآزمایشی طولی (آینده‌نگر) پیش‌بینی کننده است. غیرآزمایشی از این جهت که هیچ گونه دست‌کاری یا تخصیص تصادفی وجود ندارد (Belli, 2009: 60). به دلیل جمع‌آوری داده‌های هر شرکت در چندین دوره زمانی، نیز پژوهش طولی از نوع پنل است (Johnson & Christensen, 2014: 546). پیش‌بینی کننده به این دلیل که درصدد پیش‌بینی متغیر هدف بر مبنای چندین متغیر پیش‌بینی کننده است (Johnson & Christensen, 2014: 548-549).

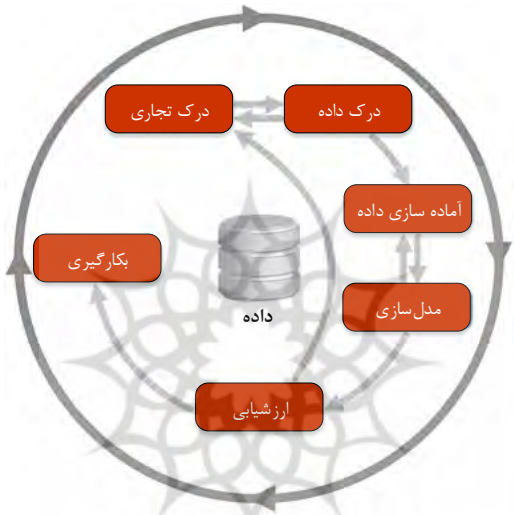
٤-٢- مدل پژوهش

در این پژوهش، یادگیری ماشین[10] جهت ارائه مدلی برای پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار تهران بکار گرفته می‌شود. برای ارائه یک مدل پیش‌بینی کننده معتبر، چندین گام می‌بایست اجرا شود. به منظور مدل‌سازی این گام‌ها، از مدل فرآیند استاندارد صنعت متقابل برای داده‌کاوی $(CRISP-DM)$[11] استفاده خواهد شد که در شکل ۱

[10] Machine Learning

[11] Cross Industry Standard Process-Data Mining (CRISP-DM)

نمایش داده شده است. بر اساس نظرسنجی آنلاین صورت گرفته توسط جامعه داده‌کاوی بین‌المللی KDNuggets در سال ۲۰۱۴، $CRISP-DM$ با ۴۳٪ سهم مشهورترین روش‌شناسی برای پروژه‌های تحلیلی، داده‌کاوی و علم داده و متداول‌ترین فرآیند داده‌کاوی مورد ارجاع و مورد استفاده در عمل است (Niaksu, 2015: 52). این مدل، گام‌های متداول مورد استفاده در فرآیند داده‌کاوی را شرح می‌دهد و به سازمان‌دهی پروژه کمک می‌کند. مدل ($CRISP-DM$) شامل شش مرحله است: درک تجاری[۱۲]، درک داده[۱۳]، آماده‌سازی داده[۱۴]، مدل‌سازی[۱۵]، ارزیابی[۱۶] و بکارگیری[۱۷] (Shearer, 2000: 14; Kwitek, 2016: 32).



شکل ۱: مراحل مدل $CRISP-DM$

برگرفته از: شرِر، ۲۰۰۰: ۱۴؛ نیاکسو، ۲۰۱۵: ۵۳

٤-۲-۱- متغیرهای پژوهش

در این پژوهش، از دو دسته متغیر استفاده می‌شود:

۱- متغیر مستقل: اطلاعات قابل اندازه‌گیری مورد استفاده برای پیش‌بینی متغیر هدف است که پیش‌بینی کننده، ویژگی، خصوصیت، متغیر ورودی، متغیر توضیحی یا از دید پایگاه داده، فیلد نیز نامیده می‌شود

12 Business Understanding
13 Data Understanding
14 Data Preparation
15 Modeling
16 Evaluation
17 Deployment

(University of Notre Dame, n.d.). متون پیش‌بینی درماندگی بدون تئوری مشخصی تکامل یافته است که تعیین کند کدام نسبت‌های مالی یا چند نسبت یا چه رویکرد وزن‌دهی به بهترین نحو امکان ارزیابی احتمال درماندگی را فراهم می‌کند (Beaver, Correia & McNichols, 2010: 113). از این رو، رویکرد معمول برای پیش‌بینی ورشکستگی این است که برای شناسایی مجموعه بزرگی از متغیرهای مستقل مالی یا غیرمالی بالقوه به مرور متون پرداخت و سپس، از طریق ترکیب تحلیل‌های ریاضیاتی و قضاوتی به مجموعه کوچک‌تری از متغیرهای مستقل دست یافت که ورشکستگی را پیش‌بینی می‌کنند. برخی از این مدل‌ها از صحت پیش‌بینی حدود ۹۰ درصد برخوردارند. اما، مسأله مطرح این است که مدل‌های ارائه شده معمولاً از متغیرهای مختلف و شکل‌های مختلفی برای تعیین روابط بین این متغیرها استفاده می‌کنند (McKee & Lensberg, 2002: 436-437). بنابراین، پس از ۳۰ سال پژوهش درباره این موضوع، هیچ مدل پذیرفته شده کلی برای پیش‌بینی ورشکستگی وجود ندارد که اساس آن مبتنی بر تعیین روابط علی عوامل اقتصادی زیربنایی باشد (McKee & Lensberg, 2002: 436-437; Lin, Wang, Wu & Chuang, 2009: 1593). در این پژوهش، سه معیار جهت انتخاب نسبت‌های مالی برای پیش‌بینی بحران مالی استفاده شده است:

۱- نسبت‌های مالی باید به صورت متداول در متون پیش‌بینی بحران مالی مورد استفاده قرار گرفته باشند: بیش از ۸۵۲ متغیر یا نسبت مختلف از ۱۲۸ مقاله معتبر (برای نمونه، تینُکو و ویلسان (۲۰۱۳)، ژو و همکاران (۲۰۱۲)، چن (۲۰۱۱)، اسکلنگر و کراس (۱۹۹۴) و اُهلسان (۱۹۸۰)) استخراج گردید. انتخاب تمامی این متغیرها برای ایجاد مدل غیرممکن است. از این رو، نسبت‌ها یا عواملی که بیشتر از سه دفعه در ۱۲۸ مقاله مورد استفاده قرار گرفته‌اند، انتخاب خواهند شد.

۲- اطلاعات مورد نیاز برای محاسبه این نسبت‌های مالی باید در دسترس باشد: کشورهای مختلف ممکن است سیاست‌های حسابداری مختلفی داشته باشند و شرکت‌های مختلف متعلق به صنایع مختلف ممکن است انواع مختلفی از صورت‌های مالی داشته باشند. از این رو، معمولاً تمامی ۸۵۲ متغیر برای همه شرکت‌ها در همه پایگاه داده‌ها در دسترس نیست (Zhou, Lai & Yen, 2012: 2486).

۳- تصمیم پژوهشگر بر مبنای تجربه قبلی خویش در پژوهش‌های پیشین یا بر مبنای آزمایش‌های مقدماتی

با در نظر گرفتن این سه معیار، ۱۲۴ متغیر پیش‌بین به شرح مندرج در جدول ۱ انتخاب شده است.

جدول (۱): متغیرهای پیش‌بین (ورودی‌ها یا ویژگی‌ها)	
طبقه	متغیر
نسبت‌های مالی مبتنی بر ترازنامه و صورت سود (زیان)	۱- نسبت جاری = دارایی‌های جاری / بدهی‌های جاری، ۲- بازده دارایی‌ها = سود خالص / کل دارایی‌ها، ۳- نسبت بدهی = کل بدهی‌ها / کل دارایی‌ها، ۴- گردش کل دارایی‌ها = فروش / کل دارایی‌ها، ۵- بازده عملیاتی دارایی‌ها = سود عملیاتی / کل دارایی‌ها، ۶- سرمایه در گردش به کل دارایی‌ها = سرمایه در گردش / کل دارایی‌ها، ۷- نسبت سریع = دارایی‌های سریع / بدهی‌های جاری، ۸- حاشیه سود خالص = سود خالص / فروش، ۹- بازده حقوق صاحبان سهام = سود خالص / حقوق صاحبان سهام، ۱۰- سود انباشته به کل دارایی‌ها = سود انباشته / کل دارایی‌ها، ۱۱- دارایی‌های جاری به کل دارایی‌ها = دارایی‌های جاری / کل دارایی‌ها، ۱۲- نسبت پوشش بهره = سود عملیاتی / هزینه بهره، ۱۳- حقوق صاحبان سهام به کل دارایی‌ها = حقوق صاحبان سهام / کل دارایی‌ها، ۱۴- کل بدهی‌ها به حقوق صاحبان سهام = کل بدهی‌ها / حقوق صاحبان سهام، ۱۵- گردش حساب‌های دریافتنی = فروش / حساب‌های دریافتنی، ۱۶- حاشیه عملیاتی = سود عملیاتی / فروش، ۱۷- گردش موجودی کالا = بهای تمام شده کالای فروش رفته / موجودی کالا، ۱۸- وجه نقد به کل دارایی‌ها = وجه نقد / کل دارایی‌ها، ۱۹- هزینه بهره به فروش = هزینه بهره / فروش، ۲۰- وجه نقد به بدهی‌های جاری = وجه نقد / بدهی‌های جاری، ۲۱- گردش دارایی‌های جاری = فروش / دارایی‌های جاری، ۲۲- گردش دارایی‌های ثابت = فروش / دارایی‌های ثابت، ۲۳- حاشیه سود ناخالص = سود ناخالص / فروش، ۲۴- بدهی‌های جاری به کل دارایی‌ها = بدهی‌های جاری / کل دارایی‌ها، ۲۵- سود هر سهم = سود خالص / تعداد سهام عادی، ۲۶- اندازه شرکت (لگاریتم کل دارایی‌ها)، ۲۷- گردش سرمایه در گردش = فروش / سرمایه در گردش، ۲۸- رشد کل دارایی‌ها، ۲۹- بدهی بلندمدت به کل دارایی‌ها = بدهی بلندمدت / کل دارایی‌ها، ۳۰- دارایی‌های سریع به کل دارایی‌ها = دارایی‌های سریع / کل دارایی‌ها، ۳۱- گردش حساب‌های پرداختنی = بهای تمام شده کالای فروش رفته / حساب‌های پرداختنی تجاری، ۳۲- دارایی‌های سریع به فروش = دارایی‌های سریع / فروش، ۳۳- بدهی‌های جاری به کل بدهی‌ها = بدهی‌های جاری / کل بدهی‌ها، ۳۴- فروش به موجودی کالا = فروش / موجودی کالا، ۳۵- دوره حساب‌های دریافتنی = حساب‌های دریافتنی × ۳۶۰ / فروش، ۳۶- سود خالص به دارایی‌های جاری = سود خالص / دارایی‌های جاری، ۳۷- بدهی‌های بلندمدت به حقوق صاحبان سهام = بدهی‌های بلندمدت / حقوق صاحبان سهام، ۳۸- وجه نقد به فروش = وجه نقد / فروش، ۳۹- رشد فروش، ۴۰- دارایی‌های

جدول (۱): متغیرهای پیش‌بین (ورودی‌ها یا ویژگی‌ها)	
طبقه	متغیر
	ثابت به کل دارایی‌ها = $\frac{\text{دارایی‌های ثابت}}{\text{کل دارایی‌ها}}$، ۴۱- گردش حقوق صاحبان سهام = $\frac{\text{فروش}}{\text{حقوق صاحبان سهام}}$، ۴۲- رشد سود خالص (در صورت وجود سود خالص در سال گذشته: $\frac{\text{سود (زیان) خالص سال جاری - سود خالص سال گذشته}}{\text{سود خالص سال گذشته}}$، در صورت وجود زیان خالص در سال گذشته: $\frac{\text{سود (زیان) خالص سال جاری} + \left\lvert\text{زیان خالص سال گذشته}\right\rvert}{\left\lvert\text{زیان خالص سال گذشته}\right\rvert}$)، ۴۳- حقوق صاحبان سهام به دارایی‌های ثابت = $\frac{\text{حقوق صاحبان سهام}}{\text{دارایی‌های ثابت}}$، ۴۴- بازده دارایی‌های ثابت = $\frac{\text{سود خالص}}{\text{دارایی‌های ثابت}}$، ۴۵- سرمایه در گردش به حقوق صاحبان سهام = $\frac{\text{سرمایه در گردش}}{\text{حقوق صاحبان سهام}}$، ۴۶- سود خالص به کل بدهی‌ها = $\frac{\text{سود خالص}}{\text{کل بدهی‌ها}}$، ۴۷- بدهی‌ها به خالص دارایی‌های مشهود = $\frac{\text{بدهی‌ها}}{\text{کل دارایی‌ها - دارایی‌های نامشهود - کل بدهی‌ها}}$، ۴۸- خالص دارایی‌ها به ازای هر سهم = $\frac{\text{خالص دارایی‌ها}}{\text{تعداد سهام}}$، ۴۹- بدهی‌های جاری به حقوق صاحبان سهام = $\frac{\text{بدهی‌های جاری}}{\text{حقوق صاحبان سهام}}$، ۵۰- بهای تمام شده کالای فروش رفته به فروش = $\frac{\text{بهای تمام شده کالای فروش رفته}}{\text{فروش}}$، ۵۱- اندازه شرکت (لگاریتم فروش)، ۵۲- حاشیه سود عملیاتی قبل از استهلاک = $\frac{\text{سود عملیاتی قبل از استهلاک}}{\text{فروش}}$، ۵۳- بازده ناخالص کل دارایی‌ها = $\frac{\text{سود ناخالص}}{\text{کل دارایی‌ها}}$، ۵۴- فروش به ازای کارکنان = $\frac{\text{فروش}}{\text{کل تعداد کارکنان}}$، ۵۵- تغییر در سود خالص = $\frac{(NI_t - NI_{t-1})}{(\lvert NI_t\rvert + \lvert NI_{t-1}\rvert)}$، ۵۶- موجودی کالا به کل دارایی‌ها = $\frac{\text{موجودی کالا}}{\text{کل دارایی‌ها}}$، ۵۷- رشد دارایی‌های ثابت، ۵۸- تعدیلات سنواتی به کل دارایی‌ها = $\frac{\text{تعدیلات سنواتی}}{\text{کل دارایی‌ها}}$، ۵۹- تعدیلات سنواتی به فروش = $\frac{\text{تعدیلات سنواتی}}{\text{فروش}}$، ۶۰- هزینه مالیات به سود قبل از مالیات = $\frac{\text{هزینه مالیات}}{\text{سود قبل از مالیات}}$.
نسبت‌های جریان نقدی	۶۱- بازده نقدی دارایی‌ها = $\frac{\text{جریان نقدی عملیاتی}}{\text{کل دارایی‌ها}}$، ۶۲= جریان نقدی عملیاتی به کل بدهی‌ها = $\frac{\text{جریان نقدی عملیاتی}}{\text{کل بدهی‌ها}}$، ۶۳- جریان نقدی عملیاتی به بدهی‌های جاری = $\frac{\text{جریان نقدی عملیاتی}}{\text{بدهی‌های جاری}}$، ۶۴- جریان نقدی عملیاتی به فروش = $\frac{\text{جریان نقدی عملیاتی}}{\text{فروش}}$، ۶۵- جریان نقدی عملیاتی هر سهم = $\frac{\text{جریان نقدی عملیاتی}}{\text{تعداد سهام عادی}}$، ۶۶- بازده نقدی حقوق صاحبان سهام = $\frac{\text{جریان نقدی عملیاتی}}{\text{حقوق صاحبان سهام}}$، ۶۷- پوشش نقدی بهره = $\frac{\text{جریان نقدی عملیاتی}}{\text{بهره پرداخت شده}}$، ۶۸- پوشش سرمایه‌گذاری مجدد = $\frac{\text{جریان نقدی عملیاتی}}{\text{وجه نقد پرداختی بابت دارایی‌های بلندمدت}}$، ۶۹- پوشش پرداخت بدهی = $\frac{\text{جریان نقدی عملیاتی}}{\text{وجه نقد پرداختی بابت بازپرداخت بدهی‌های بلندمدت}}$،

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول (۱): متغیرهای پیش‌بین (ورودی‌ها یا ویژگی‌ها)	
طبقه	متغیر
	۷۰- پوشش پرداخت سود سهام = $\frac{\text{جریان نقدی عملیاتی}}{\text{سود سهام پرداخت شده}}$، ۷۱- پوشش خروجی سرمایه‌گذاری و تأمین مالی = $\frac{\text{جریان نقدی عملیاتی}}{\text{وجه نقد خروجی بابت فعالیت‌های سرمایه‌گذاری و تأمین مالی}}$، ۷۲- مالیات نقدی پرداختی به سود قبل از مالیات = $\frac{\text{مالیات پرداخت شده}}{\text{سود قبل از مالیات}}$، ۷۳- سود تقسیمی هر سهم = $\frac{\text{سود خالص تقسیم شده}}{\text{تعداد سهام عادی}}$.
متغیرهای بازار سهام و نقدشوندگی	۷۴- ارزش بازار سهام به ارزش دفتری بدهی‌ها = $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری بدهی‌ها}}$، ۷۵- ارزش بازار سهام به ارزش دفتری سهام = $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}}$، ۷۶- اندازه شرکت (لگاریتم ارزش بازار شرکت). ۷۷- بازدهی سهام، ۷۸- رتبه نقدشوندگی، ۷۹- قیمت سهام، ۸۰- ریسک سیستماتیک ($\beta_i = \frac{Cov(R_i,R_M)}{Var(R_M)}$)، ۸۱- تعداد معامله‌های سهام، ۸۲- حجم معامله‌های سهام، ۸۳- ارزش معامله‌های سهام، ۸۴- تعداد روزهای معامله سهام، ۸۵- آمیوست ($Amivest_i = \frac{1}{N_i}\sum_{t=1}^{N_i}\frac{Vol_{i,t}}{\left\|R_{i,t}\right\|}$ که در این رابطه، N_i: تعداد روزهای با بازده غیرصفر سهام i، $Vol_{i,t}$: حجم معامله سهام i بر حسب ریال در روز t و $\left\|R_{i,t}\right\|$: قدرمطلق بازده سهام i در روز t)، ۸۶- آمیهود ($Amihud_i = \frac{1}{N_i}\sum_{t=1}^{N_i}\frac{\left\|R_{i,t}\right\|}{Vol_{i,t}}$ که در این رابطه، N_i: تعداد روزهای با حجم معامله غیرصفر سهام i، $Vol_{i,t}$: حجم معامله سهام i بر حسب ریال در روز t و $\left\|R_{i,t}\right\|$: قدرمطلق بازده سهام i در روز t)، ۸۷- سهام شناور آزاد (حاصل‌ضرب ضریب شناوری در مجموع سهام سهامداران غیرمدیریتی).
متغیرهای حاکمیت شرکتی	۸۸- تمرکز مالکیت ($HHI = \sum_{i=1}^{N_j}\left(SHARE_{i,j}\right)^2$ که در این رابطه، $SHARE_{i,j}$: درصد سهام تحت تملک سهامدار i در شرکت j)، ۸۹- سهامداران عمده (مجموع درصد سهام تحت تملک سهامداران دارای مالکیت بیشتر از ۵٪)، ۹۰- نخستین سهامدار بزرگ (درصد سهام تحت تملک نخستین سهامدار بزرگ دارای مالکیت بیشتر از ۵٪)، ۹۱- دومین سهامدار بزرگ (درصد سهام تحت تملک دومین سهامدار بزرگ دارای مالکیت بیشتر از ۵٪)، ۹۲- مالکیت مدیریتی (مجموع درصد سهام تحت تملک اعضای هیأت مدیره و مدیر عامل دارای مالکیت بیشتر از ۵٪)، ۹۳- مالکیت خانوادگی (مجموع درصد سهام تحت تملک اعضای خانواده‌های دارای مالکیت بیشتر از ۵٪)، ۹۴- مالکیت نهادی (مجموع درصد سهام تحت تملک سهامداران نهادی (بانک‌ها، شرکت‌های بیمه، صندوق‌های بازنشستگی، شرکت‌های سرمایه‌گذاری و سایر مؤسسه‌هایی که به خرید و فروش حجم بالایی از اوراق بهادار می‌پردازند) دارای مالکیت بیشتر از ۵٪)، ۹۵- اندازه مؤسسه حسابرسی (بزرگ (سازمان حسابرسی) و کوچک (سایر مؤسسه‌های حسابرسی))، ۹۶- نوع گزارش حسابرس (مقبول، مشروط، عدم اظهارنظر، مردود)، ۹۷- تغییر حسابرس ((عدم) تغییر حسابرس شرکت نسبت به سال گذشته)، ۹۸- اندازه هیأت مدیره (تعداد اعضای هیأت مدیره)، ۹۹- دوگانگی مدیر

جدول (۱): متغیرهای پیش‌بین (ورودی‌ها یا ویژگی‌ها)	
طبقه	متغیر
	عامل (بکارگیری همزمان یک مدیر به عنوان مدیر عامل و رئیس هیأت مدیره)، ۱۰۰- استقلال رئیس هیأت مدیره (مؤظف نبودن رئیس هیات مدیره)، ۱۰۱- استقلال هیأت مدیره = $\frac{\text{تعداد اعضای غیرمؤظف هیأت مدیره}}{\text{تعداد کل اعضای هیأت مدیره}}$.
متغیرهای اقتصاد کلان	۱۰۲- رشد تولید ناخالص داخلی به قیمت پایه، ۱۰۳- رشد تولید ناخالص ملی به قیمت پایه-ثابت، ۱۰۴- رشد صادرات کالاها و خدمات-ثابت، ۱۰۵- رشد واردات کالاها و خدمات-ثابت، ۱۰۶- شاخص کل قیمت بورس، ۱۰۷- رشد نقدینگی (پول + شبه پول)، ۱۰۸- نرخ بیکاری، ۱۰۹- نرخ بازار دلار، ۱۱۰- رشد سرمایه‌گذاری خارجی، ۱۱۱- نرخ تورم، ۱۱۲- رشد مخارج دولت (کل پرداخت‌ها)، ۱۱۳- چرخه تجاری (رونق)، ۱۱۴- رتبه صندوق بین‌المللی پول، ۱۱۵- شاخص کل تولید کارگاه‌های بزرگ صنعتی، ۱۱۶- شاخص بهای کالاها و خدمات مصرفی، ۱۱۷- شاخص بهای تولیدکننده، ۱۱۸- نرخ اوراق مشارکت، ۱۱۹- نرخ وام - صنعت و معدن، ۱۲۰- نرخ بهره واقعی.
سایر متغیرها	۱۲۱- صنعت (طبقه‌بندی شرکت‌ها در ۳۷ صنعت بر اساس آیسیک[18])، ۱۲۲- سود انباشته به سرمایه، ۱۲۳- عمر شرکت بر حسب مدت پذیرش در بورس، ۱۲۴- عمر شرکت بر حسب زمان تأسیس.

۲- متغیر وابسته: متغیر مورد پیش‌بینی در یادگیری نظارت شده است که متغیر پاسخ، متغیر خروجی، متغیر هدف، متغیر پیامد یا برچسب نیز نامیده می‌شود (University of Notre Dame, n.d.). در مطالعات تجربی مندرج در متون، ورشکستگی و درماندگی مالی به جای یکدیگر به کار گرفته شده‌اند. استفاده از درماندگی مالی موجب انعطاف‌پذیری در مرحله پژوهش می‌شود. درماندگی مالی در مقایسه با ورشکستگی تعریف متداول‌تری است و برای افزایش اندازه نمونه به پژوهش محدودیت‌هایی در پژوهش و کاهش اندازه نمونه می‌شود. استفاده از درماندگی مالی نه تنها در عمل بلکه در تئوری نیز دارای برتری است؛ زیرا همه شرکت‌های درمانده مالی ورشکسته نمی‌شوند. ورشکستگی آخرین گزینه برای شرکت‌هایی است که نمی‌توانند مشکلات مالی خودشان را حل کنند. به گونه خلاصه، استفاده از ورشکستگی تنها موجب محدود شدن به یک جنبه از درماندگی مالی می‌شود (Aktan, 2011: 23). از این رو، در این پژوهش، برای تشریح وضعیت شرکت‌های روبه رو با دشواری مالی از عبارت درماندگی مالی استفاده می‌شود. افزون بر این، برای تشخیص شرکت‌های درمانده مالی نیز درماندگی مالی بر اساس ماده ۱۴۱ قانون تجارت ایران تعریف می‌شود (کاتبی، ۱۳۸۰: ۹۵):

18 International Standard Industrial Classification (ISIC)

اگر بر اثر زیان‌های وارده حداقل نصف سرمایه شرکت از میان برود، هیأت مدیره مکلف است بلافاصله مجمع عمومی فوق‌العاده صاحبان سهام را دعوت نماید تا موضوع انحلال یا بقاء شرکت مورد شور و رأی واقع شود.

٤-۳- جامعهٔ آماری، روش نمونه‌گیری و حجم نمونه

جامعه آماری این پژوهش، شامل تمامی شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار تهران است. برای اجتناب از انتقادهای وارد بر جانبداری نمونه جفت تطبیقی، در این پژوهش از نمونه‌ای استفاده خواهد شد که نشان‌دهنده نرخ درماندگی واقعی است. داده‌های ۴۲۱ شرکت پذیرفته شده در فرابورس و بورس اوراق بهادار تهران طی دوره ۱۶ ساله از سال ۱۳۹۵-۱۳۸۰ در تمامی صنایع به جز صنعت مالی گنجانده خواهد شد. این روش نمونه‌گیری، امکان ارزیابی تأثیر اندازه شرکت و صنعت بر احتمال درماندگی مالی را فراهم می‌کند. بر این اساس شرکت‌های خدمات مالی نظیر بانک‌ها، بیمه‌ها، کارگزاری‌ها، لیزینگ‌ها و غیره در تجزیه و تحلیل گنجانده نشده‌اند (Ohlson, 1980: 114; Lee, Booth & Alam, 2005: 5; Pendharkar, 2005: 2569; Chancharat, 2008: 116; Li & Miu, 2010: 824)؛ زیرا شرکت‌های عضو این صنایع از دیدگاه ساختاری متفاوت هستند، محیط ورشکستگی متفاوتی دارند و در برخی موارد، بدست آوردن داده‌های مربوطه آن‌ها دشوار است (Ohlson, 1980: 114). افزون بر این، مؤسسه‌های مالی نظیر بانک‌های تجاری یا بانک‌های سرمایه‌گذاری به این دلیل در این مجموعه داده گنجانده نشده‌اند که در اقتصاد ایران، سرنوشت چنین واسطه‌گری‌های مالی به نظر می‌رسد که بیشتر تحت تأثیر سیاست‌های دولت و نه وضعیت مالی خودشان باشد (Lee, Booth & Alam, 2005: 5). فهرست اسامی صنایع مورد بررسی در این پژوهش، در جدول ۲ نمایش داده شده است. انتظار بر این است که بکارگیری کل صنایع شرکت‌های پذیرفته شده به ارائه مدل پیش‌بینی درماندگی مالی اثربخشی در بورس اوراق بهادار تهران منجر شود؛ زیرا این نمونه می‌بایست معرف کل اقتصاد ایران باشد.

جدول (۲): صنایع مورد بررسی

ردیف	صنعت	ردیف	صنعت
۱	ابزار پزشکی، اپتیکی و اندازه‌گیری	۲۰	سایر محصولات کانی غیرفلزی
۲	استخراج زغال سنگ	۲۱	سیمان، آهک و گچ
۳	استخراج سایر معادن	۲۲	عرضه برق، گاز، بخار و آب گرم
۴	استخراج کانه‌های فلزی	۲۳	فرآورده‌های نفتی، کک و سوخت هسته‌ای
۵	استخراج نفت گاز و خدمات جنبی به جز اکتشاف	۲۴	فلزات اساسی
۶	انبوه‌سازی، املاک و مستغلات	۲۵	قند و شکر
۷	انتشار، چاپ و تکثیر	۲۶	کاشی و سرامیک
۸	پیمانکاری صنعتی	۲۷	لاستیک و پلاستیک
۹	تجارت عمده‌فروشی به جز وسایل نقلیه موتور	۲۸	ماشین‌آلات و تجهیزات
۱۰	تجارت عمده وخرده‌فروشی وسائط نقلیه موتور	۲۹	ماشین‌آلات و دستگاه‌های برقی
۱۱	حمل و نقل، انبارداری و ارتباطات	۳۰	محصولات چوبی
۱۲	خدمات فنی و مهندسی	۳۱	محصولات شیمیایی
۱۳	خرده‌فروشی، به استثنای وسایل نقلیه موتوری	۳۲	محصولات غذایی و آشامیدنی به جز قند و شکر
۱۴	خودرو و ساخت قطعات	۳۳	محصولات کاغذی
۱۵	رایانه و فعالیت‌های وابسته به آن	۳۴	مخابرات
۱۶	زراعت و خدمات وابسته	۳۵	منسوجات
۱۷	ساخت دستگاه‌ها و وسایل ارتباطی	۳۶	مواد و محصولات دارویی
۱۸	ساخت محصولات فلزی	۳۷	هتل و رستوران
۱۹	سایر تجهیزات حمل و نقل		

٤-٤- ابزار گردآوری داده‌ها

در این پژوهش، برای نگارش و جمع‌آوری اطلاعات مورد نیاز بخش مبانی نظری و پیشینه پژوهش از مجله‌ها و کتاب‌های تخصصی لاتین، و برای گردآوری سایر داده‌های کمّی و اطلاعات مورد نیاز از سایت کدال، سایت بورس اوراق بهادار تهران، بانک‌های اطلاعاتی سازمان بورس اوراق بهادار تهران، سایت بانک مرکزی جمهوری اسلامی ایران، سایت صندوق بین‌المللی پول، گزارش کلان اقتصادی

پژوهشکده پولی و بانکی بانک مرکزی جمهوری اسلامی ایران، نرم افزار ره آورد نوین، نرم افزار تدبیر پرداز، و نرم افزار تی اس ای کلاینت[19] استفاده خواهد شد. بنابراین، اطلاعات مورد نیاز بخش مبانی نظری و پیشینه پژوهش، داده ثانویه از نوع مستندات رسمی و داده‌های مورد نیاز بخش مدل‌سازی، داده ثانویه از نوع داده‌های پژوهشی آرشیوی است (Johnson & Christensen, 2014).

٤-٥- روش تجزیه و تحلیل داده‌ها

٤-٥-١- پیش پردازش (تکمیل مقادیر گمشده و استخراج ویژگی‌های اولیه)

پیش پردازش داده‌ها یک گام مهم برای داده‌کاوی با کیفیت خوب در فرآیندهای داده‌کاوی یا کشف دانش در پایگاه داده است (Tsai & Cheng, 2012: 333). تکنیک‌های پیش پردازش مورد استفاده در این پژوهش عبارت‌اند از:

پاک‌سازی داده: پاک‌سازی داده به معنای تکمیل مقادیر نامشخص، هموارسازی داده‌های نویزی، شناسایی داده‌های پرت و رفع داده‌های ناسازگار است (Li & Sun, 2011: 464). در این فرآیند، نخست رکوردهای تکراری حذف گردیدند اما به علت حجم اندک داده‌های در دسترس و زیاد بودن رکوردهای حاوی مقادیر گمشده، رکوردهای حاوی مقادیر گمشده حذف نشدند و بر حسب مورد به صورت دستی از طریق گزارش‌های مالی مندرج در سایت کدال و سایت بورس اوراق بهادار تهران، میانگین مقدار آن متغیر در سال قبل و بعد همان شرکت، یا مقدار آن متغیر در سال بعد همان شرکت تکمیل شده‌اند.

ترکیب داده: در این فرآیند، چندین منبع داده شامل ترازنامه، صورت سود و زیان، صورت جریان وجوه نقد، حاکمیت شرکتی، داده‌های بازار سهام و نقدشوندگی بر اساس دو شناسه نام (نماد) شرکت و سال و متغیرهای کلان اقتصادی بر اساس شناسه سال ادغام شده‌اند.

استخراج ویژگی‌های اولیه: استخراج ویژگی‌های مورد نیاز جهت پیش‌بینی درماندگی مالی بر اساس مرور ۱۲۸ مقاله پیش‌بینی درماندگی مالی، انتخاب نسبت‌ها یا ویژگی‌های مورد استفاده در بیشتر از ۳ مقاله از ۱۲۸ مقاله، امکان‌پذیری محاسبه آن‌ها و یا سودمندی آن‌ها بر اساس دانش و تجربه پژوهشگر انجام شده است. ویژگی‌های استخراجی اولیه در جدول ۱، ارائه شده‌اند. با توجه به اینکه از یک سو، برای محاسبه برخی نسبت‌های مالی به اطلاعات سال گذشته نیاز است و از سوی دیگر، برای پیش‌بینی

19 TSEClient

درماندگی مالی نیز از ویژگی‌های یک سال قبل استفاده می‌شود، نمونه انتخابی شامل ۳۶۷۰ سال-شرکت و ۱۲۳ ویژگی است.

وان هات انکدینگ: با توجه به اینکه ویژگی صنعت، یک متغیر طبقه‌ای اسمی هست، بر روی این ویژگی، وان هات انکدینگ اجرا شده است. پس از اجرای وان هات انکدینگ، به مجموعه داده، ۳۷ ستون جدید با مقادیر صرفاً ۰-۱ افزوده شده است. اما از آنجا که ویژگی نوع اظهار نظر حسابرس، متغیر طبقه‌ای ترتیبی است، با استفاده از اعداد ۱ تا ۴ به مقادیر عددی تبدیل شده است.

ساخت ویژگی: گاهی اوقات جهت دستیابی به عملکرد بهتر به ویژگی‌های اضافی نیاز است؛ زیرا ویژگی‌های اولیه مندرج در داده خام به منظور توصیف خصوصیات نمونه‌ها ممکن است اضافی باشند یا معنادار نباشند (Fan, 2016: 10-11). در نتیجه، ویژگی جدیدی تحت عنوان نسبت سود (زیان) انباشته به سرمایه سال قبل برای پیش‌بینی درماندگی مالی در سال بعد ساخته شده است.

٤-٥-٢- تفکیک نمونه

سپس، داده آماده شده به مجموعه داده آموزشی و آزمایشی تفکیک می‌شود. در این پژوهش، از تفکیک ۷۰ به ۳۰ استفاده می‌شود. بدین معنا که ۷۰٪ داده به صورت تصادفی انتخاب می‌شود تا برای آموزش مدل استفاده شود در حالی که ۳۰٪ داده به عنوان مجموعه آزمایشی کنار گذاشته شده در نظر گرفته می‌شود تا در ارزیابی نهایی مورد استفاده قرار گیرد. مقدار بنیادی مولد تصادفی برای این تفکیک ثابت نگه داشته می‌شود تا مدل قابلیت ایجاد دوباره را داشته باشد (Oleksy, 2017: 51). به دلیل زیادی تعداد متغیرهای پیش‌بینی کننده، صرفاً آمار توصیفی برخی از متغیرهای پیش‌بینی کننده مجموعه داده آموزشی در جدول ۳ نمایش داده شده است.

٤-٥-٣- اعتبارسنجی متقابل

در این پژوهش، برای ارزیابی ترکیبات مختلف اَبرپارامترهای الگوریتم یادگیری از اعتبارسنجی متقابل k لایه طبقه‌ای (با $k = ۵$) استفاده شده است. اعتبارسنجی متقابل طبقه‌ای به ویژه، در خصوص مسائل کلاس نامتوازن می‌تواند برآورد بهتری از واریانس و سوگیری ارائه کند. در اعتبارسنجی متقابل k لایه طبقه‌ای، نسبت کلاس‌ها در هر لایه ثابت نگه داشته می‌شود تا اطمینان حاصل شود که هر لایه معرف نسبت کلاس‌ها در مجموعه داده آموزشی است.

٤-٥-٤- پیش پردازش (مقیاس‌دهی ویژگی و انتخاب ویژگی)

مقیاس دهی ویژگی‌ها: در این پژوهش، برای مقیاس‌دهی ویژگی‌ها از استانداردسازی استفاده شده است. رویه استانداردسازی را می‌توان با معادله زیر بیان کرد:

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x} \qquad (۱)$$

که در این رابطه: μ_x میانگین نمونه یک ستون ویژگی خاص و σ_x انحراف معیار مربوط به آن است (Raschka, 2015: 111).

جدول (۳): آمار توصیفی مجموعه داده آموزشی

پیش‌بینی کننده	هدف	تعداد	میانگین	انحراف معیار	چولگی	کشیدگی	پیش‌بینی کننده	هدف	تعداد	میانگین	انحراف معیار	چولگی	کشیدگی
۳۹x	۰	۲۴۷۴	۶۱/۱۴	۲۲۴۷/۷۲	۴۱/۱۳	۱۷۷۷/۵۴	۵۸x	۰	۲۴۷۴	-۰/۰۱	۰/۰۳	-۲/۵۹	۱۸/۴۲
	۱	۹۵	۰/۰۱	۰/۳۹	۰/۶۶	۱/۰۳		۱	۹۵	-۰/۰۲	۰/۰۴	-۱/۸۷	۱۰/۴۷
۴۰x	۰	۲۴۷۴	۰/۲۶	۰/۱۸	۰/۹	۰/۳۶	۶۰x	۰	۲۴۷۴	۰/۱۲	۰/۰۹	-۰/۰۳	-۱/۱۸
	۱	۹۵	۰/۲۲	۰/۱۷	۱/۲۲	۱/۱۸		۱	۹۵	۰/۰۴	۰/۰۹	۰/۴۶	۴/۰۲
۴۲x	۰	۲۴۷۴	۰/۹۷	۱۲/۹۸	۲۶/۵۲	۸۷۶/۸۲	۶۳x	۰	۲۴۷۴	۰/۳۲	۰/۹	-۲۱/۷۴	۹۰۸/۳۹
	۱	۹۵	-۴/۸۷	۲۶/۰۸	-۶/۳۵	۴۵/۹۸		۱	۹۵	۰/۰۳	۰/۲۱	-۰/۲۳	۲/۴۹
۴۵x	۰	۲۴۷۴	۰/۱۷	۰/۷۳	-۳/۴۵	۲۹/۴۵	۷۹x	۰	۲۴۷۴	۶۸۲۵/۰۵	۸۸۲۳/۷	۳/۸	۲۰/۰۳
	۱	۹۵	-۰/۱۷	۱/۴۸	-۰/۸۶	۶/۶		۱	۹۵	۲۸۵۹/۸۹	۲۵۸۸/۸۳	۲/۳۹	۶/۵۶
۴۶x	۰	۲۴۷۴	۰/۳۱	۰/۴۶	۵/۹۱	۶۴/۴۷	۸۲x	۰	۲۴۷۴	۸۹۳۸۴۳۴۱	۰۸E+۴/۵۳	۱۶/۱۶	۳۸۵/۶۲
	۱	۹۵	-۰/۰۳	۰/۱	-۰/۱۹	۳/۲۵		۱	۹۵	۷۲۴۱۰۰۵۹	۰۸E+۴/۶۲	۹/۳۱	۸۶/۲۵
۴۷x	۰	۲۴۷۴	۱۱۵۲/۳۸	۴۰۸۲۳/۷۴	۳۶/۶۳	۱۳۷۱/۷۳	۸۸x	۰	۲۴۷۴	۰/۳۴	۰/۲۲	۰/۷۸	۰/۱
	۱	۹۵	۱۵۹۰۲/۰۷	۱۵۴۹۲۲	۹/۵۹	۹۰/۰۱		۱	۹۵	۰/۳۷	۰/۲۴	۰/۹۲	۰/۱۲
۴۸x	۰	۲۴۷۴	۲۴۸۴/۴۳	۱۷۶۴/۶۴	۳/۲	۱۵/۵۸	۹۲x	۰	۲۴۷۴	۱/۳۳	۸/۰۷	۸/۰۷	۷۲/۲۶
	۱	۹۵	۱۳۶۹/۶۲	۱۰۲۷/۱۶	۲/۸۴	۹/۴۶		۱	۹۵	۱/۶۸	۷/۲۶	۵/۴۷	۳۲/۴۱
۵۰x	۰	۲۴۷۴	۰/۷۲	۰/۱۷	-۰/۷۱	۱/۷۷	۹۳x	۰	۲۴۷۴	۵/۵۵	۱۷/۵۹	۳/۳۷	۱۰/۶۹
	۱	۹۵	۰/۸۹	۰/۱۳	-۰/۲	۲/۲۸		۱	۹۵	۴/۰۷	۱۷/۲	۴/۵۵	۱۹/۷۷
۵۱x	۰	۲۴۷۴	۱۲/۹۶	۱/۷۳	-۰/۰۸	۵/۷۴	۹۵x	۰	۲۴۷۴	۰/۲۴	۰/۴۳	۱/۲	-۰/۵۶
	۱	۹۵	۱۲/۰۷	۲/۰۵	-۱/۷۳	۱۱/۵۱		۱	۹۵	۰/۲۴	۰/۴۳	۱/۲	-۰/۵۵

انتخاب ویژگی: انتخاب زیرمجموعه ویژگی به معنای یافتن زیرمجموعه ویژگی I به گونه‌ای است که ریسک مورد انتظار طبقه کننده در صورت آموزش با I حداقل شود (Pajkossy, 2013: 6). در این پژوهش، انتخاب ویژگی به روش ترکیبی با استفاده از عبارت پراکندگی λ_1 انجام شده است. تکنیک انتخاب ویژگی حاصل از بکارگیری λ_1 را لاسو می‌نامند. لاسو روش قدرتمندی است که دو کار اصلی را انجام می‌دهد: تنظیم و انتخاب ویژگی. روش لاسو، محدودیتی بر مجموع قدرمطلق پارامترهای مدل وضع می‌کند مبنی بر اینکه این مجموع باید کمتر از یک مقدار ثابت (حد بالا) باشد. بدین منظور، این روش فرآیند انقباض (تنظیم) را به کار می‌گیرد که به موجب آن، ضرایب متغیرهای رگرسیون را با انقباض برخی از آن‌ها به صفر، جریمه می‌کند. طی فرآیند انتخاب ویژگی، متغیرهایی که بعد از فرآیند انقباض همچنان دارای ضریب غیرصفر هستند، به عنوان بخشی از مدل انتخاب می‌شوند. هدف این فرآیند، حداقل کردن خطای پیش‌بینی است (Fonti, 2017: 5). مقدار بهینه λ، تعداد ویژگی‌های انتخاب شده را کنترل می‌کند و می‌تواند با استفاده از اعتبارسنجی متقابل تعیین شود (Pajkossy, 2013: 6).

٤-٥-٥- آموزش الگوریتم یادگیری

الگوریتم یادگیری: رگرسیون لجستیک بسیار کارا است، نیازمند منابع محاسباتی خیلی زیادی نیست، از قابلیت تفسیر بالایی برخوردار است، تنظیم آن ساده است و احتمال پیش‌بینی‌ها را نیز ارائه می‌کند. افزون بر این، الگوریتم‌های ساده نظیر مدل رگرسیون لجستیک برای مجموعه داده‌های کوچک دارای همبستگی زیاد بین متغیر پاسخ و ویژگی‌ها مناسب است. از این رو، در این پژوهش از رگرسیون لجستیک به عنوان الگوریتم یادگیری استفاده شده است. رگرسیون لجستیک، نوعی تجزیه و تحلیل رگرسیونی است که در آن متغیر پاسخ، باینری است بدین معنا که صرفاً مقادیر ۰ یا ۱ را می‌تواند بگیرد. متغیرهای توضیحی می‌توانند گسسته یا پیوسته باشند. اجزای اصلی این مدل عبارت‌اند از (Fonti, 2017: 18):

۱- **جزء تصادفی:** توزیع احتمال f متغیر پاسخ، دوجمله‌ای است:

$$Y_i \sim Binomial(n_i, \pi_i) \qquad (۲)$$

که در این رابطه، n_i مخرج دوجمله‌ای و π_i احتمال است.

۲- **جزء سیستماتیک:** ترکیب خطی از متغیرهای توضیحی است:

$$\eta_i = x_i^T \beta \qquad (۳)$$

۳- **تابع رابط:** تابع رابط در واقع تابع لوجیت است:

$$\eta_i = logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = \sum_{j=0}^{p} \beta_j x^{(j)} \quad (۴)$$

بکارگیری تعداد زیادی متغیر برای مدل‌سازی به صورت غیرضروری تفسیر تجزیه و تحلیل را پیچیده می‌کند، ممکن است به برازش بیشتر از اندازه منجر شود و همچنین، اصل امساک را نقض کند. افزون بر این، به احتمال زیاد ساختار همبستگی بین مجموعه متغیرهای بزرگ وجود دارد که به بی ثباتی فضای پاسخ و نتایج نامنسجم منجر می‌شود. برای نمونه، در رگرسیون چندگانه، همخطی چندگانه بین مجموعه متغیرها می‌تواند به رگرسیونی منجر شود که حتی هنگامی که هیچ کدام از متغیرها معنادار نیستند کل رگرسیون معنادار است. یک شیوه غلبه بر این مشکلات، بکارگیری روش لاسو است (Grosswindhager, 2009: 12).

رگرسیون لجستیک لاسو: روش لاسو، حد بالای ثابتی را بر مجموع قدر مطلق پارامترهای مدل وضع می‌کند. در خصوص مدل‌های خطی تعمیم یافته، این محدودیت را می‌توان از طریق جریمه کردن منفی لگاریتم احتمال با نرم L_1 بیان کرد. در مدل رگرسیون لجستیک، منفی لگاریتم احتمال برابر است با (Fonti, 2017: 18-19):

$$-\sum_{i=1}^{n} log\left(P_\beta(Y_i|X_i)\right) =$$
$$\sum_{i=1}^{n}\left\{-Y_i\left(\sum_{j=0}^{p}\beta_j x^{(j)}\right) + log\left(1 + exp\left(\sum_{j=0}^{p}\beta_j x^{(j)}\right)\right)\right\} \quad (۵)$$

که می‌توان آن را بر حسب تابع زیان ρ نوشت:

$$\rho_{(\beta)}(x,y) = -y\left(\sum_{j=0}^{p}\beta_j x^{(j)}\right) + log\left(1 + exp\left(\sum_{j=0}^{p}\beta_j x^{(j)}\right)\right) \quad (۶)$$

برآوردکننده لاسو برای مدل رگرسیون لجستیک بدین گونه تعریف می‌شود:

$$\hat{\beta}(\lambda) = \underset{\beta}{\operatorname{argmin}}\left(n^{-1}\sum_{i=1}^{n}\rho(\beta)\,(X_i, Y_i)\right) + \lambda\|\beta\|_1 \quad (۷)$$

رگرسیون لجستیک لاسو می‌تواند برای ترکیب انتخاب ویژگی و آموزش طبقه‌بندی کننده بکار گرفته شود. مزیت این کار این است که مجموعه ویژگی‌ها برای مدل خاصی بهینه می‌شوند و در نتیجه، عملکرد مدل بهبود می‌یابد. افزون بر این، ترکیب انتخاب ویژگی و برآورد ضرایب موجب کاهش بار محاسباتی می‌شود؛ زیرا نیازی به هیچ گونه الگوریتم انتخاب ویژگی صریحی نیست. به

دلیل سادگی مدل، ضرایب مدل به سادگی تفسیرپذیر هستند و دانش بیشتری را ارائه می‌کنند (Manninen, 2014: 9).

یادگیری نامتوازن: در حوزه مسائل طبقه‌بندی، اغلب سناریوی داده‌های نامتوازن مطرح می‌شود. ویژگی اصلی این نوع مسأله طبقه‌بندی این است که تعداد نمونه‌های یک نوع کلاس به میزان قابل توجهی بیشتر از تعداد نمونه‌های کلاس دیگر است (López, 2013: 114). هنگامی که یک طبقه‌بندی کننده استاندارد بر روی یک داده نامتوازن بکار گرفته می‌شود، مدلی که کورکورانه تمامی نمونه‌ها را به عنوان کلاس اکثریت طبقه‌بندی می‌کند، کماکان می‌تواند به نرخ طبقه‌بندی نادرست اندکی دست یابد. بنابراین، نه تنها به چارچوب مناسب‌تری برای مسائل داده نامتوازن نیاز است بلکه به سیستمی نیاز است که مدل‌های دارای نرخ طبقه‌بندی نادرست متوازن‌تر در بین کلاس‌ها را انتخاب کند. مسأله نخست، معمولاً با اصلاح‌های مبتنی بر داده یا الگوریتم روش‌های کلاسیک قابل حل و فصل است. مسأله بعدی به این موضوع می‌پردازد که چه معیاری برای ارزیابی مسائل کلاس نامتوازن مناسب‌تر است (Ren, 2014: 20)؟

در این پژوهش، با توجه به اینکه مجموعه داده درماندگی مالی، نامتوازن است، برای پاسخ به مسأله نخست از روش یادگیری حساس به هزینه استفاده شده است. این روش مستلزم اختصاص هزینه‌های مختلف به نمونه‌های دارای طبقه‌بندی نادرست، از طریق ماتریس هزینه است. برای نمونه، فرض کنید $C(i,j)$ هزینه طبقه‌بندی نمونه کلاس i اُم به عنوان کلاس j اُم است. آنگاه، معمولاً، $C(i,i) = ۰$ و $C(i,j) \geq C(j,i)$ در نظر گرفته خواهد شد چنانچه فراوانی کلاس i اُم کمتر باشد. هدف از این کار، حداقل نمودن هزینه کل بعد از اجرای ماتریس هزینه هست (Ren, 2014: 24).

افزون بر این، برای پاسخ به مسأله بعدی نیز نمره $F_۱$ به عنوان معیار ارزیابی عملکرد انتخاب می‌شود؛ زیرا نمره $F_۱$ هنگام بررسی مسأله کلاس نامتوازن مناسب‌تر از AUC است. افزون بر این، AUC معیاری مبتنی بر منحنی جهت نمایش عملکرد میانگین یا کلی یک طبقه‌بندی کننده است در حالی که نمره $F_۱$ معیاری نقطه‌ای است که در برخی موارد در عمل معنادارتر است (Fan, 2016: 19).

٤-٥-٦- بهینه‌سازی اَبرپارامتر

اَبرپارامترها، پارامترهای مختص مدل هستند که برای بهینه‌سازی مدل مورد استفاده قرار می‌گیرند. معمولاً، به منظور بهینه‌سازی عملکرد مدل و کاهش واریانس و سوگیری مدل باید اَبرپارامترها را تنظیم کرد (Bonnes, 2017: 27). تنظیم اَبرپارامترهای مدل با استفاده از کاوش کامل زیرمجموعه‌هایی از

فضای ترکیبات احتمالی اَبرپارامتر انجام می‌شود. هدف، یافتن مقادیری برای اَبرپارامترها است که به عملکرد طبقه‌بندی بهینه منجر شود. این نوع جستجو را جستجوی شبکه‌ای نیز می‌نامند؛ زیرا آرایه‌های پارامترها یک شبکه چندبعدی را تشکیل می‌دهند (Oleksy, 2017: 51).

بکارگیری اعتبارسنجی متقابل k لایه همراه با جستجوی شبکه‌ای، رویکرد مفیدی برای میزان‌سازی دقیق عملکرد مدل یادگیری ماشین از طریق تغییر مقادیر اَبرپارامترهای آن هست (Raschka, 2015: 187). اما بهینه‌سازی اَبرپارامترها با استفاده از جستجوی شبکه‌ای به روش اعتبارسنجی متقابل، فرآیندی مستلزم صرف منابع و زمان زیاد است. از این رو، در این پژوهش، فضای جستجوی اَبرپارامترها و همچنین، تعداد تکرارهای اعتبارسنجی متقابل محدود شده است (Oleksy, 2017: 52).

مجموعه اَبرپارامترهای ارائه شده به الگوریتم رگرسیون لجستیک در جدول ۴ نمایش داده شده است.

جدول (۴): فضای جستجوی اَبرپارامترها	
اَبرپارامترها	مقادیر
C	[۰/۰۰۱, ۰/۰۱, ۰/۰۰۴, ۰/۰۰۲, ۰/۰۲, ۰/۰۰۵, ۰/۰۰۰۷, ۰/۰۰۰۶, ۰/۰۰۰۵, ۰/۰۰۶, ۰/۰۳, ۰/۰۲۵]
$class_weight$	[{۰:۱, ۱:۱۱}, {۰:۱, ۱:۱۲}, {۰:۱, ۱:۸}, {۰:۱, ۱:۹}, {۰:۱, ۱:۱۰}, {۰:۱, ۱:۷/۵}, {۰:۱, ۱:۸/۵}, {۰:۱, ۱:۱۳}]

٤-٥-٧- پس پردازش (معیارهای سنجش عملکرد و انتخاب مدل)

پس از آموزش رگرسیون لجستیک لاسو با مقادیر مختلف اَبرپارامترها و ارزیابی عملکرد آن، اَبرپارامترهای مندرج در جدول ۵، به عنوان اَبرپارامترهای بهینه مدل انتخاب شده‌اند که به مدل دارای بهترین عملکرد از نظر توان پیش‌بینی منجر می‌شوند.

جدول(۵): اَبرپارامترهای بهینه	
اَبرپارامترها	مقادیر
$penalty$	$L1$
$solver$	$saga$
C	۰/۰۲
$class_weight$	{۰:۱, ۱:۸}

٤-٥-٨- مدل طبقه‌بندی نهایی

پس از اجرای رگرسیون لجستیک لاسو با اَبرپارامترهای بهینه بر روی داده‌های آموزشی، ضرایب ویژگی‌های نامربوط و اضافی در مدل، صفر و نسبت شانس آن‌ها برابر با ۱ می‌شود. این بدان معناست که این ویژگی‌ها تأثیری بر شانس وقوع درماندگی مالی ندراند. ضرایب سایر ویژگی‌های مربوط و مهم به شرح مندرج در جدول ۶ است.

جدول (۶): ضرایب رگرسیون لجستیک لاسو

ویژگی	ضرایب	نسبت شانس	ویژگی	ضرایب	نسبت شانس
Intercept	-۲/۸۵۴۹۶	۰/۰۵۷۵۵۸	x۸۸	-۰/۰۳۹۳۷	۰/۹۶۱۳۹۷
x۳۹	-۰/۰۲۷۳۱	۰/۹۷۳۰۶۲	x۹۲	-۰/۰۴۴۷۵	۰/۹۵۶۲۳۲
x۴۰	۰/۲۴۲۷۷۷	۱/۲۷۴۷۸۴	x۹۳	۰/۰۶۱۲۷۵	۱/۰۶۳۱۹۱
x۴۲	-۰/۵۳۵۰۶	۰/۵۸۵۶۳۲	x۹۵	-۰/۱۱۰۲۱	۰/۸۹۵۶۴۴
x۴۵	۰/۰۷۷۳۰۵	۱/۰۸۰۳۷۱	x۹۷	-۰/۰۹۰۲۷	۰/۹۱۳۶۸۳
x۴۶	-۰/۳۴۴۷۳	۰/۷۰۸۴۱۱	x۹۸	-۰/۰۴۳۸	۰/۹۵۷۱۴۹
x۴۷	-۰/۷۸۷۶۱	۰/۴۵۴۹۳۱	x۱۰۲	-۱/۵۴E-۰۶	۰/۹۹۹۹۹۸
x۴۸	۰/۰۰۰۵۷	۱/۰۰۰۵۷۱	x۱۰۸	۰/۰۲۴۵۵۱	۱/۰۲۴۸۵۵
x۵۰	-۰/۲۴۲۷۴	۰/۷۸۴۴۷۷	x۱۰۹	-۰/۰۱۵۱۲	۰/۹۸۴۹۹۷
x۵۱	۰/۱۶۹۸۹۳	۱/۱۸۵۱۷۸	x۱۱۳	-۰/۰۸۴۲	۰/۹۱۹۲۵۱
x۵۸	-۰/۰۴۷۲۱	۰/۹۵۳۸۸۹	x۱۲۲	-۰/۰۳۲۲	۰/۹۶۸۳۱۱
x۶۰	-۰/۰۴۹۲۴	۰/۹۵۱۹۵	x۱۲۱_۵	۰/۰۲۷۲۲۳	۱/۰۲۷۵۹۷
x۶۳	-۰/۰۲۲۷۷	۰/۹۷۷۴۸۹	x۱۲۱_۸	۰/۳۰۳۶۸۸	۱/۳۵۴۸۴۶
x۷۹	-۰/۰۰۳۱۸	۰/۹۹۶۸۲۳	x۱۲۱_۱۰	۰/۲۱۴۴۷۴	۱/۲۳۹۲۰۹
x۸۲	۰/۰۱۷۱۷۴	۱/۰۱۷۳۲۲	x۱۲۱_۱۵	۰/۰۶۶۸۸۲	۱/۰۶۹۱۷
x۸۴	۰/۰۰۶۸۴۸	۱/۰۰۶۸۷۲	x۱۲۱_۲۰	۰/۰۲۳۰۴۳	۱/۰۲۳۳۱
x۸۶	۰/۰۲۷۴۲۶	۱/۰۲۷۸۰۶	x۱۲۱_۲۱	-۰/۱۱۳۲۲	۰/۸۹۲۹۵۶
x۸۷	۰/۰۴۹۱۸۷	۱/۰۵۰۴۱۶	x۱۲۱_۲۷	-۰/۰۰۱۷۲	۰/۹۹۸۲۸۴

بر اساس نتایج مندرج در جدول ۶، یک سال قبل از درماندگی مالی، ویژگی‌های رشد فروش، رشد سود خالص، سود خالص به کل بدهی‌ها، بدهی‌ها به خالص دارایی‌های مشهود، بهای تمام شده

کالای فروش رفته به فروش، تعدیلات سنواتی به کل دارایی‌ها، هزینه مالیات به سود قبل از مالیات، جریان نقدی عملیاتی به بدهی‌های جاری، قیمت پایانی سهام، تمرکز مالکیت، مالکیت مدیریتی، اندازه موسسه حسابرسی، تغییر حسابرس، اندازه هیات مدیره، رشد تولید ناخالص داخلی به قیمت پایه، نرخ بازار دلار، چرخه تجاری (رونق)، سود انباشته به سرمایه، صنایع سایر محصولات کانی غیرفلزی، لاستیک و پلاستیک دارای تأثیر منفی بر وقوع درماندگی مالی هستند. در واقع، همان گونه که از نسبت شانس آن‌ها پیداست، با افزایش این ویژگی‌ها یا با فعالیت در این صنایع، شانس وقوع درماندگی مالی (به اندازه ۱ منهای نسبت شانس) کاهش می‌یابد. همچنین، ویژگی‌های دارایی‌های ثابت به کل دارایی‌ها، سرمایه در گردش به حقوق صاحبان سهام، خالص دارایی‌ها به ازای هر سهم، لگاریتم فروش، حجم معاملات، تعداد روز معاملاتی، آمیهود، سهام شناور آزاد، مالکیت خانوادگی، نرخ بیکاری، صنایع استخراج نفت گاز و خدمات جنبی به جز اکتشاف، پیمانکاری صنعتی، تجارت عمده و خرده فروشی وسائط نقلیه موتور و رایانه و فعالیت‌های وابسته به آن دارای تأثیر مثبت بر وقوع درماندگی مالی هستند. در حقیقت، همان گونه که از نسبت شانس آن‌ها پیداست، با افزایش این ویژگی‌ها یا با فعالیت در این صنایع، شانس وقوع درماندگی مالی (به اندازه ۱ منهای نسبت شانس) افزایش می‌یابد.

سپس، مدل انتخابی، بر روی کل ۷۰٪ داده آموزشی تفکیک شده، آموزش داده می‌شود. آموزش مدل با استفاده از کل مجموعه داده آموزشی، به ماتریس درهم ریختگی مندرج در جدول ۷، منجر می‌شود.

جدول(۷): ماتریس درهم ریختگی مجموعه داده آموزشی

وضعیت		واقعی	
		تعداد شرکت سالم	تعداد شرکت درمانده
پیش‌بینی شده	تعداد شرکت سالم	۲۳۶۰	۲۶
	تعداد شرکت درمانده	۱۱۴	۶۹
جمع		۲۴۷۴	۹۵

معیارهای ارزیابی عملکرد طبقه‌بندی کننده انتخابی بر روی کل مجموعه داده آموزشی در جدول ۸ نشان داده شده است.

جدول (۸): معیارهای ارزیابی مجموعه داده آموزشی			
معیار	**مقدار**	**معیار**	**مقدار**
صحت	٪۹۵	ضریب همبستگی متیوز	٪۵۰
دقت	٪۳۸	میانگین هندسی	٪۸۳
فراخوانی	٪۷۳	شاخص صحت متوازن	٪۵۱
معیار F	٪۵۰	خطای نوع ۱	٪۲۷
ناحیه زیر منحنی مشخصه عملیاتی دریافت کننده	٪۹۵	خطای نوع ۲	٪۰۵
کاپا کوهن	٪۴۷		

٤-٥-٩- ارزیابی مدل نهایی

ارزیابی مدل نهایی به معنای برآورد عملکرد مدل انتخابی بر روی داده جدید است. از این رو، عملکرد مدل انتخابی در قسمت قبل، بر روی ۳۰٪ داده آزمایشی کنار گذاشته شده، مورد آزمون قرار می‌گیرد.

پس از ارائه مجموعه داده آزمایشی به مدل انتخابی، ماتریس درهم ریختگی مندرج در جدول ۹ بدست می‌آید.

جدول(۹): ماتریس درهم ریختگی مجموعه داده آزمایشی			
وضعیت		واقعی	
		تعداد شرکت سالم	تعداد شرکت درمانده
پیش‌بینی شده	تعداد شرکت سالم	۱۰۱۶	۱۱
	تعداد شرکت درمانده	۴۵	۲۹
		۱۰۶۱	۴۰

معیارهای ارزیابی حاصل از ماتریس درهم ریختگی مجموعه داده آزمایشی به شرح مندرج در جدول ۱۰ است.

جدول (۱۰): معیارهای ارزیابی مجموعه داده آزمایشی			
معیار	**مقدار**	**معیار**	**مقدار**
صحت	٪۹۵	ضریب همبستگی متیوز	٪۵۱
دقت	٪۳۹	میانگین هندسی	٪۸۳
فراخوانی	٪۷۳	شاخص صحت متوازن	٪۵۱
معیار F	٪۵۱	خطای نوع ۱	٪۲۸
ناحیه زیر منحنی مشخصه عملیاتی دریافت کننده	٪۹۶	خطای نوع ۲	٪۰۴
کاپا کوهن	٪۴۸		

۵- نتیجه‌گیری و ارایه پیشنهادها

پیش‌بینی درماندگی مالی کماکان مسأله بسیار بااهمیتی و مورد توجه گروه‌های درون سازمانی و برون سازمانی نظیر مدیران شرکت، سرمایه‌گذاران، بستانکاران، حسابرسان، قانون‌گذاران دولت و سایر ذینفعان است (Aktan, 2011: 62). از این رو، هدف پژوهش حاضر، ارائه مدلی برای پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار تهران است. بر اساس نتایج این پژوهش، اطلاعات صورت‌های مالی، بازار سهام، حاکمیت شرکتی و متغیرهای اقتصاد کلان می‌تواند به عنوان ابزاری برای علامت‌دهی درماندگی مالی مورد استفاده قرار گیرد. افزون بر این، پژوهش بیور و همکاران (۲۰۰۵) نیز بیانگر این است که محتوای اطلاعاتی داده‌های صورت‌های مالی برای هشدار درماندگی مالی در طی زمان کاسته نشده است. از این رو، یافته‌های این پژوهش با یافته‌های آن‌ها سازگار است اگر چه این پژوهش‌ها در محیط متفاوت و با استفاده از تکنیک‌های تحلیلی متفاوتی در دوره‌های زمانی مختلفی انجام شده‌اند. همچنین، نتایج تجربی بیانگر این است که مدل لجستیک لاسو به نمره F1، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۵۰٪، ۵۰٪، ۷۳٪ و ۳۸٪ بر روی مجموعه آموزشی منجر می‌شود. افزون بر این، آزمون مدل پیشنهادی بر روی مجموعه آزمایشی کنار گذاشته شده حاکی از این است که نمره F1، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۵۱٪، ۵۱٪، ۷۳٪ و ۳۸٪ بر روی مجموعه آزمایشی بدست می‌آید.

در انجام این پژوهش محدودیت‌هایی به شرح زیر وجود داشت: مدل ارائه شده صرفاً داده‌های کمّی را بکار خواهد گرفت. نمونه شرکت‌های مورد بررسی در این پژوهش، محدود به شرکت‌های سهامی عام پذیرفته شده در فرابورس و بورس اوراق بهادار تهران است. این سوگیری مهم است؛ زیرا

شرکت‌های کوچک بیشتر در معرض درماندگی مالی هستند. از این رو، این محدودیت ممکن است تعمیم نتایج تجربی را محدود نماید. حجم اندک داده در دسترس، محدودیت‌هایی را در خصوص بکارگیری الگوریتم‌های پیچیده ایجاد می‌کند. از یک سو، پذیرش تعریف قانونی از درماندگی مالی طبق ماده ۱۴۱ قانون تجارت، و از سوی دیگر، احتمال دست‌کاری حساب‌ها، ابهاماتی را در خصوص تمایز دقیق شرکت‌های درمانده از غیردرمانده مطرح می‌کند. به بیان دقیق‌تر، نه تنها خود شرکت‌ها تمایل دارند که به منظور عدم شمول ماده ۱۴۱ قانون تجارت، صورت‌های مالی خود را دست‌کاری نمایند بلکه قانون‌گذار نیز انتقال مازاد تجدید ارزیابی شرکت‌ها به حساب افزایش سرمایه را برای شرکت‌های مشمول ماده ۱۴۱ قانون تجارت که به وسیله این تجدید ارزیابی از ماده ۱۴۱ قانون تجارت خارج شوند، را مجاز می‌داند هر چند که از لحاظ محتوایی و بنیادی، تغییر اساسی در ساختار مالی شرکت‌ها ایجاد نمی‌شود و صرفاً نوعی حساب آرایی است. در واقع، پتانسیل و نقدینگی جدیدی وارد شرکت نمی‌شود بلکه صرفاً ارزش دارایی‌های فعلی شرکت‌ها به روزرسانی می‌شود و تنها ترازنامه شرکت‌ها تغییر می‌کند. بدیهی است که عدم تفکیک دقیق شرکت‌های درمانده از غیردرمانده تأثیر منفی بر عملکرد مدل می‌گذارد. افزون بر این، ماهیت نویزی داده‌های حسابداری و مالی، دست‌کاری حساب‌ها، تجدید ارائه مکرر صورت‌های مالی و در نتیجه تعدیلات سنواتی زیاد، عدم رعایت یکنواختی و ثبات رویه در تهیه گزارش‌های مالی و همچنین، دانش و دقت ناکافی دست‌اندرکاران تهیه و ارائه صورت‌های مالی در طبقه‌بندی صحیح حساب‌ها در سرفصل‌های مربوطه موجب کاهش محتوای اطلاعاتی داده‌های پیش‌بینی کننده و در نتیجه، تخریب عملکرد مدل می‌شود.

از این رو، پیشنهاد می‌شود که با افزودن داده‌های سال‌های آتی شرکت‌های فعلی و همچنین، شرکت‌های جدیدی که سهام آن‌ها در آینده در بورس اوراق بهادار تهران معامله خواهد شد، حجم داده‌های مورد استفاده را افزایش دهند تا بتوانند با بکارگیری الگوریتم‌های غیرخطی نظیر مدل‌های یادگیری عمیق، عملکرد مدل را بهبود دهند. همچنین، یافته‌های این پژوهش دارای مفاهیم ضمنی برای تحلیل‌گران و مدیران بخش دولتی و خصوصی است. با بکارگیری مدل پیش‌بینی درماندگی شرکت‌ها، قانون‌گذران می‌توانند به راحتی ثبات اقتصادی را نظارت و تضمین کنند، سرمایه‌گذاران می‌توانند تصمیم‌های بهتری در خصوص زمان خرید سهام جدید یا حتی فروش سهام بگیرند، بستانکاران نیز از ارزش اعتباری شرکت‌ها آگاهی می‌یابند و می‌توانند بهتر در خصوص اعطای اعتبار، شرایط و مدت آن قضاوت نمایند، مدیریت شرکت می‌تواند از از دست دادن مدیران و کارگران

کلیدی، عرضه‌کنندگان عمده، مشتریان بزرگ و همچنین، از بین رفتن اعتماد بستانکاران، سهامداران و سرمایه‌گذاران پیشگیری نماید.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع و مآخذ

۱. خواجوی، شکرالله و محمدحسین قدیریان آرانی (۱۳۹۶). نقش توانایی مدیریت در پیش‌بینی بحران مالی، پژوهش‌های حسابداری مالی، سال نهم، شماره ۴، پیاپی (۳۴)، ۸۳-۱۰۱.

۲. کاتبی، حسینقلی (۱۳۸۰). حقوق تجارت. چاپ هفتم، تهران: انتشارات گنج دانش.

3. Aktan, S. (2011). Early warning system for bankruptcy: Bankruptcy prediction (Doctoral dissertation, Karlsruhe Institute of Technology, KIT). Retrieved from https://d-nb.info/1019790032/34.
4. Altman, E. I. (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", The Journal of Finance, 23(4), 589–609.
5. Andrés, J., Landajo, M., & Lorca, P. (2012). "Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios", Knowledge-Based Systems, 30, 67–77.
6. Ashoori, S., & Mohammadi, S. (2011). "Compare failure prediction models based on feature selection technique: Empirical case from Iran", Procedia Computer Science, 3, 568–573.
7. Beaver, W. H., Correia, M., & McNichols M. F. (2010). "Financial statement analysis and the prediction of financial distress", Foundations and Trends in Accounting, 5(2), 99–173.
8. Belli, G. (2009). Nonexperimental quantitative research. In S. D. Lapan & M. T. Quartaroli (Eds.), Research essentials: An introduction to designs and practices. (pp. 59-77). Jossey-Bass Publications.
9. Bonnes, K. (2017). Predicting mortgage demand using machine learning techniques (Master Thesis, University of Twente). Retrieved from https://essay.utwente.nl/ 73640/7/Bonnes_MA_EEMCS.pdf.
10. Cenciarelli, V. G., Greco, G., & Allegrini, M. (2018). "External audit and bankruptcy prediction", Journal of Management and Governance, 22(4), 863-890.
11. Chancharat, N. (2008). An empirical analysis of financially distressed Australian companies: The application of survival analysis (Doctoral dissertation, University of Wollongong). Retrieved from https://ro.uow.edu.au/theses/ 401/.
12. Chen, M. Y. (2011). "Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches", Computers and Mathematics with Applications, 62(12), 4514–4524.
13. Chen, H. L., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S. J., & Liu, D. Y. (2011). "A novel bankruptcy prediction model based on an adaptive Fuzzy K-nearest neighbor method." Knowledge-Based Systems, 24, pp. 1348–1359.
14. Etemadi, H., Anvary Rostamy, A. A., & Farajzadeh Dehkordi, H. (2009). "A genetic programming model for bankruptcy prediction: Empirical evidence from Iran", Expert Systems with Applications, 36(2), 3199–3207.

15. Fan, X. (2016). An adaptive and diversity-based ensemble method for binary classification (Master Thesis, Carleton University). Retrieved from https://curve.carleton.ca/system/files/etd/e2e72ad2-2da5-47b8-ab6a28810 d5eb197/etd_pdf/81f038dc409086dafcc1cced7b46be03/fan-anadaptiveand diversitybasedensemblemethod.pdf.
16. Fontaine Rezende, F., Montezano, R., Nascimento de Oliveira, F., & Lameira, V. (2017). "Predicting financial distress in publicly-traded companies", Revista Contabilidade & Finanças, 28(75), 390-406.
17. Fonti, V. (2017). Feature selection using LASSO (Research paper in Business Analytics, Vrije Universiteit). Retrieved from https://beta.vu.nl/nl/ Images/werkstuk-fonti_tcm235-836234.pdf.
18. Grosswindhager, S. (2009). Using penelized logistic regression models for predicting the effects of advertising material (Master Thesis, Vienna University of Technology). Retrieved from https://publik.tuwien.ac.at/files/PubDat_179921 .pdf
19. Jardin, P. (2010). "Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy", Neurocomputing, 73(10-12), 2047–2060.
20. Jo, H., & Han, I. (1996). "Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction", Expert Systems with Applications, 11(4), 415–422.
21. Johnson, R. B., & Christensen, L. (2014). Educational research: quantitative, qualitative, and mixed approaches (5th ed.). London: SAGE Publications, Inc.
22. Jones, S. (2017). "Corporate bankruptcy prediction: a high dimensional analysis", Review of Accounting Studies, 22(3), 1366-1422.
23. Klobucnik, J., Miersch, D., & Sievers, S. (2017). "Predicting early warning signals of financial distress: Theory and empirical evidence", Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id= 2237757.
24. Kothari, C. R. (2004). Research methodology, methods and techniques. New Delhi: New Age International (P) Ltd. Publishers.
25. Kwitek, M. (2016). A feasibility study of AZURE machine learning for sheet metal fabrication (Master Thesis, University of Vaasa). Retrieved from https://www.tritonia.fi/fi/e-opinnaytteet/tiivistelma/7268/A+Feasibility +Study+of+Azure+Machine+Learning+for+Sheet+Metal+Fabrication
26. Lee, K., Booth, D., & Alam, P. (2005). "A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms", Expert Systems with Applications, 29(1), 1–16.
27. Lensberg, T., Eilifsen, A., & McKee, T. E. (2006). "Bankruptcy theory development and classification via genetic programming", European Journal of Operational Research, 169(2), 677–697.
28. Li, H., & Sun, J. (2011). "On performance of Case-Based reasoning in Chinese business failure prediction from sensitivity, specificity, positive and negative values", Applied Soft Computing, 11(1), 460–467.

29. Li, M. Y. L., & Miu, P. (2010). "A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach", Journal of Empirical Finance, 17(4), 818–833.
30. Lin, R. H., Wang, Y. T., Wu, C. H., & Chuang, C. L. (2009). "Developing a business failure prediction model via RST, GRA and CBR", Expert Systems with Applications, 36(2), 1593–1600.
31. López, V., Fernández, A., García, S., Palade, V. & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences, 250, 113-141.
32. Mai, F., Tian, S., Lee, C., & Ma, L. (2019). "Deep learning models for bankruptcy prediction using textual disclosures", European Journal of Operational Research, 274(2), 743-758.
33. Manninen, T. (2014). Predictive modeling using sparse logistic regression with applications (Doctoral Dissertation, Tampere University of Technology). Retrieved from https://tutcris.tut.fi/portal/files/1788046/manninen.pdf.
34. McKee, T. E., & Lensberg, T. (2002). "Genetic programming and rough sets: A hybrid approach to bankruptcy classification", European Journal of Operational Research, 138(2), 436–451.
35. Mokhatab Rafiei, F., Manzari, S. M., & Bostanian S. (2011). "Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian Evidence", Expert Systems with Applications, 38(8), 10210–10217.
36. Niakšu, O. (2015). Development and application of data mining methods in medical diagnostics and healthcare management (Doctoral Dissertation, Vilnius University). Retrieved from https://old.mii.lt/files/mii_dis_2015_ niaksu.pdf.
37. Ohlson, J. A. (1980). "Financial ratios and the probabilistic prediction of bankruptcy", Journal of Accounting Research, 18(1), 109 –131.
38. Oleksy, T. A. (2017). Machine learning methods for mood disorder decision support (Master Thesis, University of Bergen). Retrieved from http://bora.uib.no/ bitstream/handle/1956/16259/actigraphdataformoods.pdf?sequence=4&isAllowed=y.
39. Pajkossy, K. (2013). Studying feature selection methods applied to classification tasks in natural language processing (Master Thesis, Eotvos Lorand University of Sciences). Retrieved from https://web.cs.elte.hu/blobs/ diplomamunkak/ msc_alkmat/2013/pajkossy_katalin.pdf
40. Pendharkar, P. C. (2005). "A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem", Computers & Operations Research, 32(10), 2561–2582.
41. Raschka, S. (2015). Python machine learning. Birmingham: Packt Publishing Ltd..
42. Ren, J. (2014). Robust feature selection with penalized regression in imbalanced high dimensional data (Doctoral Dissertation, University of Southern

California). Retrieved from http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll3/ id/443080.
43. Shah, S. B. H. (2016). "The impact of corporate governance on financial distress; Evidence from Pakistan", (Master Thesis, Capital University of Science and Technology). Retrieved from https://thesis.cust.edu.pk/UploadedFiles/Syed%20Basharat%20Hussain%20Shah%20-MMS151042.pdf
44. Shearer, C. (2000). "The CRISP-DM model: The new blueprint for data mining", Journal of Data Warehousing, 5(4), 13-22.
45. Smith, R. F., & Winakor, A. H. (1935). Changes in the financial structure of unsuccessful industrial corporations. Urbana: University of Illinois.
46. Suntraruk, Phassawan (2009). Predicting Financial Distress: Evidence from Thailand. http://www. efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2009-milan/phd/phassawan. pdf
47. Tinoco, M. H., & Wilson, N. (2013). "Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables", International Review of Financial Analysis, 30, 394–419.
48. Tsai, C. F., & Cheng, K. C. (2012). "Simple instance selection for bankruptcy prediction", Knowledge-Based Systems, 27, 333–342.
49. University of Notre Dame (n.d.). Retrieved January 18, 2019, from https://www3.nd.edu/~busiforc/handouts/DataMining/dataminingdefinitions.html
50. Waqas, H., & Md-Rus, R. (2018). "Predicting financial distress: Importance of accounting and firm-specific market variables for Pakistan's listed firms", Cogent Economics & Finance, 6, 1-16.
51. Wu, Y., Gaunt, C., & Gray, S. (2010). "A comparison of alternative bankruptcy prediction models", Journal of Contemporary Accounting & Economics, 6(1), 34–45.
52. Xu, W., Fu, H., & Pan, Y. (2019). "A novel soft ensemble model for financial distress prediction with different sample sizes", Mathematical Problems in Engineering, 2019, 1-12.
53. Zhou, L., Lai, K. K., & Yen, J. (2012). "Empirical models based on features ranking techniques for corporate financial distress prediction", Computers and Mathematics with Applications, 64(8), 2484–2496.
54. Zmijewski, M. E. (1984). "Methodological issues related to the estimation of financial distress prediction models", Journal of Accounting Research, 22, 59-82.

Financial Distress Prediction of the Listed Companies on Tehran Stock Exchange (TSE) and Iran Fara Bourse (IFB) Using Lasso Logistic Regression

Mohammad Namazi[۲۰]

Shahla Ebrahimi[۲۱]

Abstract:

The ability to predict corporate financial distress is important to business individuals as well as to the economy in general. Therefore, the purpose of this article is the detection of potential financial distress and early warnings of impending financial distress among the listed companies on Tehran Stock Exchange (TSE) and Iran Fara Bourse (IFB). To do so, a wide range of features including accrual accounting variables, cash-based accounting variables, market-based variables, corporate governance mechanisms, and macroeconomic indicators have been identified to prospectively predict the financial distress in the companies.

The final sample includes 421 firms leading to 3,670 firm-year observations. The prepared data, was then split into a train and test data set using a 70/30 ratio.

In this research, various data pre-processing machine learning techniques i.e., Z-score standardization, one-hot encoding, stratified K-fold validation combined with feature engineering are applied to improve classifier performance. Stratified K-fold cross validation method, (with $k = 5$) was used for estimation of model prediction performance during training phase. During the training phase, hyper-parameter tuning of a model was carried out using a grid-search. Furthermore, a cost-sensitive meta-learning approach in conjunction with the proposed imbalance-oriented metric i.e., F1 score were used to overcome the extreme class imbalance issue.

Based on the experimental results, the tuned LASSO logistic model achieved a f1-score, MCC, recall and precision of respectively, 50%, 50%, 73% and 38% on the training set. Finally, the proposed model was tested on the hold-out test set which resulted in a f1-score, MCC, recall and precision of 51%, 51%, 73% and 39%, respectively.

Keywords: Financial Distress Prediction, Lasso Logistic Regression, Machine Learning, Data Mining, Tehran Stock Exchange.

JEL Classification: G33, G38, G34, E44, C19, M40

۲۰. Professor of Accounting Shiraz University, Iran. mnamazi@rose.shirazu.ac.ir

۲۱. PhD Student of Accounting Shiraz University, Iran. (Corresponding Author).
shebrahimi_3622@yahoo.com