



متن کاوی و پیش بینی درماندگی مالی

دکتر زهره حاجیها^۱

دانشیار گروه حسابداری، واحد تهران شرق، دانشگاه آزاد اسلامی، تهران، ایران

محمدرضا صابری روچی^۲

عضو انجمن حسابداری ایران، دانشجوی دکتری حسابداری، دانشگاه آزاد اسلامی، بیرجند، ایران

(تاریخ دریافت: ۱۱ دی ۱۳۹۹؛ تاریخ پذیرش: ۱۵ اردیبهشت ۱۴۰۰)

مقاله حاضر با عنوان متن کاوی و پیش بینی درماندگی مالی و با هدف پیش‌بینی درماندگی مالی در قالب عبارات کلیدی با استفاده از روش تکنیک متن کاوی انجام گرفته است. متن کاوی فرآیندی است که طی آن اطلاعات مفید از داده‌های غیر ساختار یافته (داده‌های متنی) استخراج و مورد استفاده قرار می‌گیرد. جامعه آماری گزارش‌های حسابرسی ۵۰ شرکت درمانده مالی پذیرفته شده در بورس در سال ۱۳۹۶ که برای جمع آوری داده‌ها گزارشات حسابرسی سال‌های ۱۳۹۴ و ۱۳۹۵ مورد متن کاوی قرار گرفت. یافته‌ها نشان داد که اصطلاحات پرکاربرد که در طول دوره زمانی دو ساله قبل از درماندگی مالی شرکت‌ها بطور مستمر تکرار شده عبارتند از: عدم رعایت مقررات و قوانین با ۶۲، مغایرت با ۱۲۳، عدم حضور مدیر ذینفع با ۱۲۲، کاهش سود، کاهش ارزش کالا و دارایی‌ها با ۱۱۶، عدم دسترسی به اطلاعات و اسناد با ۱۰۲، تعدیلات سنواتی با ۱۰۱ مورد و... لذا می‌توان نتیجه گرفت با توجه به واژه‌های استخراج شده روش متن کاوی می‌تواند در پیش‌بینی درماندگی مالی شرکت‌ها موثر واقع گردد.

واژه‌های کلیدی: پیش‌بینی، تکنیک متن کاوی، درماندگی مالی، گزارشات حسابرسی، ماده ۱۴۱.

¹ drzhajiha@gmail.com

² mr.saberi4@yahoo.com

مقدمه

از آنجا که بازار سرمایه در ایران جوان است و به عنوان یک بازار کارا مطرح نمی‌باشد، تحقیقات علمی می‌تواند بیان تئوریک برای این بازار پایه ریزی نماید تا اطلاعات منتشره سازمان بورس در تصمیم‌گیری صاحبان سرمایه و سرمایه‌گذاران بالقوه مفید باشد. یکی از راه‌هایی که می‌توان با استفاده از آن به بهره‌گیری مناسب از فرصت‌های سرمایه‌گذاری و تخصیص بهتر منابع کمک کند، ارزیابی درماندگی مالی است [۵]. زیرا درماندگی مالی^۱ شرکت‌ها معمولاً بر نقدینگی بازار سرمایه و توسعه اقتصاد مؤثر است. در زمان درماندگی مالی، بانک‌ها معمولاً اعتباردهی را کاهش داده و در ازای وامی که به شرکت‌ها می‌دهند، بهره بالاتری را برای جبران ریسک اضافی درخواست می‌کنند. به صورت مشابهی، مؤسسات سرمایه‌گذاری همچون صندوق‌های بازنشستگی و شرکت‌های بیمه، خرید سهام را کاهش داده و بیشتر به سراغ سرمایه‌گذاری و خرید اوراق قرضه بانک‌ها یا بازارهای مشابه به آن اقدام می‌کنند. همه این‌ها به کاهش نقدینگی در بازارهای سرمایه افزایش هزینه سرمایه شرکت‌ها و کاهش رشد اقتصادی منجر خواهد شد [۲].

با توجه به اثرات درماندگی مالی شرکت‌ها بر مدیریت، سهامداران، مشتریان، بستانکاران و سایر افراد ذینفع، همواره یافتن شاخص‌های هشدار دهنده وقوع درماندگی مالی یکی از جذاب‌ترین و با اهمیت‌ترین حوزه‌ها در تحقیقات مالی بوده است. زیرا چنانچه بتوان در مورد وقوع آن پیش از رخداد اطلاعاتی به دست آورد، می‌توان از پیامدهای اقتصادی و اجتماعی آن کاسته یا حتی جلوگیری نمود [۱۲]. بر این اساس سیاستگذاران و شرکت‌کنندگان در بازار هر روز مقدار زیادی از گزارش‌های مالی و مقالات خبری را برای وقوع درماندگی مالی استفاده می‌کنند. با این وجود افشای متنی در مدل‌های مالی و اندازه‌گیری داده‌های متنی برای شناسایی درماندگی مالی کار دشواری است [۱۹]. مطالعات اخیر شواهد محکمی را نشان می‌دهد که شرکت‌ها اطلاعات کیفی ارزشمندی را ثبت می‌کنند [۱۷].

با این حال یکی از الگوهای پیش‌بینی ورشکستگی موجود، نسبت‌های مالی می‌باشد. اما با توجه به اینکه نسبت‌های حسابداری تنها براساس اطلاعات سالانه بدست می‌آید، بنابراین، همیشه ریسک اعتماد به اطلاعات تاریخی وجود دارد. از طرف دیگر ارقام حسابداری به شدت تحت تأثیر رویه‌های حسابداری انتخابی از سوی مدیر می‌باشد، در چنین شرایطی می‌توان گفت اعداد استخراج شده از صورت‌های مالی فاقد قدرت کافی جهت پیش‌بینی بحران مالی شرکت‌ها هستند. با توجه به اینکه در حال حاضر منابع اطلاعاتی گسترده‌ای در محیط تجاری برای پیش‌بینی رخدادهای مالی آتی از جمله درماندگی مالی وجود دارد، این امر ضروری به نظر می‌رسد که جایگاه سایر منابع اطلاعاتی در این رقابت شناسایی شود تا بتوان در صورت امکان با در کنار یکدیگر قراردادن آن‌ها در یک مدل پیش‌بینی بر کاستی‌های احتمالی یک منبع به تنهایی فائق آمد [۶]. بیشتر مطالعات نشان داده‌اند که ریسک اعتباری از طریق تکنیک‌های خلاصه‌سازی متن مانند تعداد کلمه‌ها، احساسات و خوانایی و همچنین سیگنال موجود در متن مالی به مراتب فراتر اقدامات اهرم‌های هستند که برای استخراج و بهره‌برداری اطلاعات از داده‌های متنی و ارزش

¹ Distress

افشای این متن، نیاز به الگوریتم های مؤثر وجود دارد [۹]. بر این اساس متن کاوی، بعنوان روشی در استخراج دانش از متون، یکی از موضوعات مهم در گستره‌ای از اعمال مدیریت اطلاعات می‌باشد. در این میان آنچه از اهمیت فوق العاده ای برخوردار است ارائه راهکارهایی برای مواجهه با این حجم عظیم اطلاعاتی و استفاده بهینه از اطلاعات در جهت خلق دانش، تولید سینرژی^۱ و در نهایت افزایش خرد جمعی است. نشانه‌های پریشانی مالی خود را به سرعت نشان نمی‌دهند بلکه در میان حجم انبوهی از اطلاعات مالی و غیر مالی خود را مستتر می‌سازند [۲]. از طرفی این نشانه‌ها در اطلاعات غیر مالی بسیار زودتر از داده‌های مالی نمایان می‌شوند. همچنین اطلاعات کیفی تنها منعکس کننده عملکرد گذشته شرکت نمی‌باشد بلکه سیگنال‌هایی از عملکرد و وضعیت آینده شرکت نیز به همراه دارد. به این دلیل است که در سال‌های اخیر تجزیه و تحلیل داده‌های کیفی بسیار با اهمیت گردیده است. از آنجایی که حدود ۹۰٪ از داده‌های سازمان‌ها با روش‌های متعارف داده کاوی قابل کشف نیستند. لذا انگیزه زیادی جهت استفاده از متن کاوی وجود دارد. در این زمینه پژوهش‌های زیادی انجام شده است اما آنچه تا به امروز کمتر مورد توجه قرار گرفته است، انتخاب معیارهای مؤثر در پیش بینی درماندگی مالی در بازار مالی ایران یا همان بحث انتخاب ویژگی است [۲] به همین دلیل در این پژوهش سعی بر آن است با رویکردی جدید به مسئله انتخاب ویژگی در پیش بینی درماندگی مالی پرداخته شود. در واقعه تلاش محقق، استخراج اطلاعات مفید برای پیش بینی درماندگی مالی در قالب کلمات و عبارات کلیدی با استفاده از تکنیک متن کاوی می‌باشد.

مبانی نظری

تمایز مفهوم درماندگی مالی و ورشکستگی

نکته قابل تامل در تحقیقات انجام شده در ایران در مورد موضوع درماندگی آن است که در کلیه این تحقیقات دو مفهوم کاملاً متمایز درماندگی و ورشکستگی بصورت مترادف و با مفهوم یکسان در نظر گرفته شده است. بطوری که برای شناسایی شرکت‌های درمانده و نیز ورشکسته از معیار ماده ۱۴۱ قانون تجارت استفاده شده است [۱۰]. اما با توجه به این که شرکت‌های مشمول این ماده لزوماً اعلام انحلال و ورشکستگی نمی‌کنند در این تحقیق نیز به چنین شرکت‌هایی درمانده مالی اطلاق گردید.

پیش بینی درماندگی مالی

تا به امروز تکنیک‌های مختلفی برای طراحی مدل‌های پیش‌بینی درماندگی مالی مورد استفاده قرار گرفته است. به طور کلی مدل‌های پیش‌بینی درماندگی مالی را می‌توان در سه گروه مدل‌های آماری، هوش مصنوعی و نظری طبقه بندی کرد. مدل‌های آماری خود به دو گروه تک متغیره^۲ و چند متغیره تقسیم می‌شوند. مدل‌های آماری چند متغیره از مدل‌های تحلیل تشخیصی چندگانه^۳، احتمالی خطی، لوجیت و

^۱ Synergy

^۲ Univariate Ratio Analysis

^۳ Multiple Discriminant Analysis (MDA)

پروبیته^۱، مجموع تجمعی و فرآیندهای تعدیل ناقص تشکیل شده است. شبکه‌های عصبی مصنوعی، الگوریتم ژنتیک، افراز بازگشتی^۲، مجموعه‌های سخت، ماشین بردار پشتیبان و منطق فازی تشکیل دهنده فنون هوش مصنوعی هستند [۵] و مدل‌های نظری نیز شامل معیارهای تجزیه ترازنامه/نظریه آنتروپی، نظریه ورشکستگی سفته باز، نظریه مدیریت وجوه نقد و نظریه ریسک اعتباری است [۱۵] اما مطالعات جدید بیشتر به بررسی پیش‌بینی‌ها ورشکستگی پرداخته است. در این تحقیقات قدرت منابع داده‌ها فراتر از دو نوع متغیر است. به عنوان مثال، لیانگ و همکاران^۳ [۲۰] قدرت پیش‌بینی ورشکستگی را در مجموعه گسترده‌ای از شاخص‌های حاکمیت شرکتی همچون ساختار هیئت مدیره، ساختار مالکیت، ویژگی‌های کمیته حسابداری و حسابرسان مستقل بررسی کرده‌اند. دومپوس و همکاران^۴ [۱۸] به طراحی مدل پیش‌بینی ورشکستگی در سطح کشورها با استفاده از داده‌های اقتصادی، تجاری، سیاسی و حتی زیست محیطی پرداختند. در اکثر موارد، محققان براساس پیشینه پژوهش‌های انجام شده اقدام به انتخاب متغیرهای ساختار یافته (متغیرهای مالی و غیرمالی با اهمیت) کرده و با استفاده از تکنیک‌های آماری، هوش مصنوعی یا مدل‌های نظری اقدام به پیش‌بینی در ماندگی مالی می‌کنند [۱۶].

با این وجود در ۹۹٪ شرکت‌ها، علت در ماندگی مالی و در نهایت ورشکستگی‌ها مربوط به ضعف مدیریت سازمان در مواجهه با تغییرات شرایط محیطی و عدم توانایی در اخذ تصمیمات مناسب در سازمان می‌باشد [۶]. همچنین مدیریت ممکن است انگیزه‌ای برای مخفی کردن اطلاعات نزولی یا استفاده از زبان مبهم در افشای آن‌ها داشته باشد [۲۱].

متن کاوی

با گسترش شگرف اینترنت و استفاده روز افزون از آن در جهت ارائه و یا کسب اطلاعات، شاهد حجم انبوهی از اسناد و مقالات برخط هستیم که بعنوان یکی از مشخصات بارز زندگی مدرن امروزی، تحت عنوان افزونگی اطلاعاتی مطرح می‌گردد. در این میان دسترسی سریع و صحیح به منابع مهم و مورد علاقه، یکی از دغدغه‌های استفاده‌کنندگان از این منبع اطلاعاتی بسیار بزرگ است. در دنیای کنونی این کمبود اطلاعات نیست که مسئله است بلکه کمبود دانشی است که از این اطلاعات می‌توان حاصل کرد و همچنین کمبود روش‌هایی در جهت یافت و بهره برداری از اطلاعات در دسترس به نحوی بهینه است. دانش خلاصه‌ی اطلاعات است و نیز نتیجه‌گیری و حاصل فکر و تحلیل بر روی اطلاعات. ارائه ابزارهایی که با بررسی متون بتواند تحلیلی روی آن‌ها انجام دهند منجر به شکل‌گیری این زمینه در هوش مصنوعی شده که به متن کاوی معروف است. این حوزه تمام فعالیت‌هایی که به نوعی به دنبال کسب دانش از متن هستند را شامل می‌گردد [۲]. آنالیز داده‌های متنی توسط تکنیک‌های یادگیری ماشین، بازیابی اطلاعات هوشمند، پردازش زبان طبیعی یا روش‌های مرتبط دیگر همگی در زمره مقوله یادگیری متن قرار می‌گیرند.

¹ Logit & Probit Analysis

² Recursive Partitioning Algorithm (RPA)

³ Liang et al

⁴ Doumpos et al

این تکنیک‌ها در ابتدا در مورد داده‌های ساخت یافته به کار گرفته شدند و علمی به نام داده کاوی را بوجود آوردند. داده‌های ساخت یافته به داده‌هایی گفته می‌شود که بطور کاملاً مستقل از همدیگر ولی یکسان از لحاظ ساختاری در یک محل گردآوری شده‌اند. انواع بانک‌های اطلاعاتی را می‌توان نمونه‌هایی از این دسته اطلاعات نام برد. در اینصورت مسئله داده کاوی عبارت است از کسب اطلاعات و دانش از این مجموعه ساخت یافته. اما در مورد متون که عمدتاً غیر ساخت یافته یا نیمه ساخت یافته هستند ابتدا باید توسط روش‌هایی، آن‌ها را ساختارمند نمود و سپس از این روش‌ها برای استخراج اطلاعات و دانش از آن‌ها استفاده کرد. متن کاوی به دنبال استخراج اطلاعات مفید از داده‌های متنی غیر ساخت یافته از طریق تشخیص و نمایش الگوها است یا به عبارت دیگر متن کاوی روشی برای استخراج دانش از متون است. متن کاوی یکی از زمینه‌هایی است که به دنبال استخراج اطلاعات مفید، از داده‌های متنی بدون ساختار، به وسیله شناسایی و اکتشاف الگوها می‌باشد. ایده اصلی متن کاوی، یافتن قطعات کوچک اطلاعات از حجم زیاد داده‌های متنی، بدون نیاز به خواندن تمام آن است [۱۳].

متن کاوی در پیش بینی درماندگی مالی

با توجه به شرایط متغیر اقتصادی و نوسانات شدید در محیط‌های تجاری تصمیم‌گیری مالی نسبت به گذشته راهبردی‌تر شده و همواره با ریسک و عدم اطمینان همراه می‌باشد و واکنش سریع و درست مقابل شرایط بسیار متغیر بازار، در موقعیت بنگاه‌ها نقش بسزایی دارد. از این رو، یکی از راه‌های کمک به سرمایه‌گذاران به منظور ارائه ی اطلاعات به آن‌ها، ارائه الگوهای مناسب پیش‌بینی دور‌نمای شرکت است [۷]. با توسعه‌ی بازارهای پولی و مالی و متعاقب آن حاکم شدن وضعیت رقابتی، بسیاری از شرکت‌های درمانده و ورشکسته از گردونه رقابت خارج می‌شوند. این امر موجب نگرانی صاحبان سرمایه شده لذا برای اینکه از سوخت شدن سرمایه‌های خود جلوگیری نمایند به دنبال روش‌هایی هستند که بحران مالی شرکت‌ها را پیش‌بینی نمایند. بحران مالی و نهایتاً ورشکستگی واحدهای اقتصادی می‌تواند زبان‌های هنگفتی را در سطح خرد و کلان جامعه وارد نماید [۸].

از سوی دیگر مالکان، مدیران، سرمایه‌گذاران، اعتباردهندگان، شرکت‌ها تجاری و همچنین موسسات دولتی تمایل زیادی برای پیش‌بینی بحران‌های مالی بنگاه‌ها دارند، زیرا در صورت درماندگی مالی هزینه‌های زیادی به آن‌ها تحمیل می‌گردد. هر کدام از این الگوها نقاط قوت و ضعف خاص خود را دارند [۱۵] و از آنجا که در سال‌های اخیر حجم بانک‌های اطلاعاتی ایجاد شده توسط شرکت‌ها، بانک‌ها و خصوصاً سازمان بورس و اوراق بهادار افزایش یافته است. بخش عمده‌ای از این اطلاعات، به صورت داده‌های متنی می‌باشند؛ که به دلیل حجم بسیار زیاد آن‌ها و عدم شناخت مجامع علمی و محققان در استفاده از اینگونه داده‌ها، کمتر مورد توجه بوده است. به منظور استفاده از این حجم داده‌ها و هوشمند شدن فرایند استخراج اطلاعات مفید از آن‌ها تکنولوژی متن کاوی^۱ به وجود آمد. متن کاوی فرآیندی

^۱ Text Mining

است که طی آن اطلاعات مفید از داده‌های غیر ساختار یافته (داده‌های متنی) استخراج و مورد استفاده قرار می‌گیرد [۴].

سوالات تحقیق

۱. آیا می‌توان تکنیک متن کاوی را روشی مناسب برای استخراج اطلاعات مفید از متون معرفی کرد؟
۲. آیا می‌توان ادعا کرد با استفاده از تکنیک متن کاوی در اطلاعات استخراج شده از گزارش‌های حسابرسی پیش‌بینی وضعیت مالی آتی شرکت‌ها دقیق‌تر انجام می‌گردد؟

روش تحقیق

در واقع تکنیک‌های تولید و جمع‌آوری پایگاه داده‌ها بسیار سریع‌تر از توانایی ما در درک و استفاده از آن‌ها رشد کرده است اغلب پژوهش‌های موجود با استفاده از یکسری داده‌های مالی (نسبت‌های مالی) و غیر مالی طبقه بندی شده (معیارهای حاکمیت شرکتی، گزارش حسابرسی و غیره) اقدام به پیش‌بینی درماندگی مالی شرکت‌ها می‌کنند [۲۲] که معمولاً با گذشت زمان و تغییر شرایط، کارایی خود را از دست می‌دهند [۱۸]. نگاهی به ترجمه لغوی داده کاوی، به ما در درک بهتر این واژه کمک می‌کند. داده کاوی به معنای استخراج از منابع نهفته و با ارزش است که در پیوند کلمه به کلمه داده و جستجوی عمیق آن‌ها جهت پیدا کردن اطلاعات اضافی مفید که قبلاً در متن نهفته بوده است، یاری می‌رساند [۱] داده کاوی یک رشته نسبتاً جدید علمی می‌باشد که از انجام تحقیقات حداقل در رشته‌های آمار، یادگیری ماشین، علوم رایانه خصوصاً مدیریت پایگاه داده‌ها شکل گرفته است [۲۳].^۱ البته مرزهای این رشته‌ها در داده کاوی مبهم نیز می‌باشد [۱۴]. بنابراین یکی از رویکردهای جدید مورد استفاده در پیش‌بینی درماندگی مالی شرکت‌ها، استفاده از تحلیل محتوای متون، اسناد و گزارش‌های مالی و غیر مالی می‌باشد. [۲]

در اولین گام می‌بایست لیست شرکت‌های درمانده مالی را از بانک‌های اطلاعاتی معتبر استخراج و گزارش‌های حسابرسی مربوط به دو دوره قبل از درماندگی آن‌ها جمع‌آوری گردد. همانطور که گفته شد ملاک درماندگی و انتخاب شرکت‌های درمانده مالی مشمول بودن ماده ۱۴۱ قانون تجارت می‌باشد. طبق این قانون «اگر بر اثر زیان‌های وارد شده حداقل نصف سرمایه شرکت از میان برود، هیأت مدیره مکلف است بلافاصله مجمع عمومی فوق العاده صاحبان سهام را دعوت کند تا موضوع انحلال یا بقای شرکت مورد شور و رأی واقع شود. هرگاه مجمع مزبور رأی به انحلال شرکت ندهد، باید در همان جلسه و با رعایت مقررات ماده ۶ قانون سرمایه شرکت را به مبلغ سرمایه موجود کاهش دهد. در صورتی که هیئت مدیره بر خلاف این ماده به دعوت مجمع عمومی فوق العاده مبادرت نکند یا مجمعی که دعوت می‌شود نتواند مطابق مقررات قانونی منعقد شود هر ذینفعی می‌تواند انحلال شرکت را از دادگاه صلاحیت دار درخواست کند».

¹ Monroe, W., and Jurafsky, D

در گام دوم باید کلمات کلیدی را از محتوای متن گزارش‌های حسابرسی شرکت‌های درمانده مالی استخراج کرد. برای این امر ابتدا به پیش پردازش متن پرداخته می‌شود. بدین منظور در ابتدا قیود و کلمه‌های توقف (کلمه‌هایی مثل: "از"، "به"، "در"، "برای" و ...) را از محتوای گزارش‌های حسابرسی حذف می‌شود. سپس کلمه‌های موجود را با ریشه‌ی آن جایگزین می‌گردد (مثلاً: کلمه "ذخیره" را جایگزین کلمه‌های "ذخایر" و "ذخیره‌ها" می‌گردد). در پایان کلمه‌های مترادف را با توجه به محتوای آن متن، در قالب یک کلمه مناسب با هم ترکیب می‌شود (مثلاً: زمانی که کلمه‌های "سهام" و "اوراق بهادار" در متن وجود دارد صرفاً از کلمه اوراق بهادار برای توصیف اینگونه کلمه‌ها استفاده شود).

همچنین در این مرحله باید کلمات را رفع ابهام نمود زیرا ممکن است یک کلمه دارای چند معنی باشد که باید به مفهوم اصلی آن اشاره و ابهامات در معنی یک کلمه یا عبارت رفع گردد. این امر یکی از چالش‌های تکنیک متن کاوی به شمار می‌رود زیرا ممکن است جملات به طور مستقیم مفهوم خود را بروز ندهند و یا بصورت استعاره به کار روند و همچنین این که ممکن است ساختار جمله چنان پیچیده باشد که به دلیل رسمی نبودن الزاماً در چارچوب کلی زبان قرار نگیرند؛ بدین دلایل بررسی تنها لغت به لغت می‌تواند باعث ایجاد نتایج غلط گردد (برای مثال: کلمه‌ی سود در گزارش حسابرسی هر دو گروه شرکت درمانده و سالم به کار می‌رود ولی در شرکت‌های سالم به معنی "کسب سود" و "سود تقسیمی" ولی در شرکت‌های درمانده در جملاتی نظیر "عدم پرداخت به موقع سود سهام" و ... به کار می‌رود).

در این پژوهش برای رفع مشکل فوق، کلمات کلیدی در یک فرهنگ لغت ثبت شده و برای هر یک از آن‌ها تعریفی ارائه گردیده است، تعریفی که بار مثبت یا منفی آن کلمه را مشخص می‌کند. از آنجایی که تحقیق با پیدا کردن کلمات کلیدی به دنبال پیش‌بینی درماندگی مالی شرکت‌ها می‌باشد، لذا بار منفی و بدبینانه لغات در نظر می‌باشد (برای مثال: کلمه "مطالبات" در گزارش‌های حسابرسی شرکت‌های درمانده با عبارات زیر تعریف می‌شود: "اقلام راكد و سنواتی، ذخیره مطالبات مشکوک الوصول، چک‌های لاصول، تعهدات سایر شرکت‌ها و غیره" و یا کلمه "هزینه‌های جذب نشده" با عبارات "ظرفیت بلااستفاده ماشین آلات، اوقات تلف شده نیروی انسانی، ضایعات غیر عادی تولید، توقف تولید به دلیل کمبود مواد اولیه و غیره" تعریف می‌شود).

در گام سوم وزن‌های ریشه‌های باقیمانده محاسبه می‌گردد. رایج‌ترین روش، محاسبه فرکانسی است که کلمه با آن پدیدار می‌شود. پس به محاسبه فراوانی عبارات و کلمه‌های مشخص شده در گام قبلی، به تفکیک هر گزارش حسابرسی و کل گزارش‌های حسابرسی پرداخته می‌شود [۲۰].

در گام چهارم عبارات و کلماتی که به لحاظ فراوانی (در سطح یک گزارش و کل گزارش‌های حسابرسی) در شرکت‌های درمانده بسیار تکرار شده است (بر اساس آزمون‌های آماری) مشخص می‌گردد.

و نهایتاً در گام پنجم، با تفسیر عبارات و کلمه‌های کلیدی استخراج شده در گام قبل، دانش موجود در گزارش‌های حسابرسی (مربوط به درماندگی مالی شرکت‌ها) به دست می‌آید.

با توجه به گام‌های ذکر شده در مرحله اول گزارشات حسابرسی سال‌های ۹۴ و ۹۵ شرکت‌های درمانده مالی سال ۹۶ استخراج شدند. در مرحله دوم حدود ۱۰۰ کلمه از گزارشات حسابرسی متن کاوی شد که

این اصطلاحات و عبارات دارای بار معنای منفی بودند. در مرحله سوم با توجه به تکرار و امتداد بعضی واژگان در نهایت ۲۱ عبارت یا اصطلاح که در اکثر شرکت در مانده وجود داشت انتخاب گردید و در مرحله نهمی با کسر تعداد داده ها از میانگین جملاتی ۱۰ واژه که پرکاربردتر و تکرار بیشتری صورت گرفته بودند انتخاب شده‌اند.

جامعه آماری، حجم نمونه و روش نمونه‌گیری مورد مطالعه

جامعه آماری این تحقیق ۵۰ شرکت در مانده مالی سال ۹۶ بورس اوراق بهادار تهران می‌باشد که برای جمع آوری داده‌های گزارش‌های حسابرسی این شرکت‌ها طی دو سال قبل مورد متن کاوی قرار گرفت. به عبارت دیگر این پژوهش به دنبال آزمون تجربی گزارش حسابرسی شرکت‌ها در مانده مالی سال ۱۳۹۵ و ۱۳۹۴ می‌باشد، بر این اساس بازه زمانی این پژوهش بین سال‌های ۹۶ تا ۹۴ می‌باشد. انتخاب شرکت‌های در مانده با توجه به مفاد ماده ۱۴۱ قانون تجارت انجام می‌گیرد.

یافته‌های تحقیق

پس از شناسایی ۵۰ شرکت در مانده مالی در سال ۱۳۹۶ گزارشات حسابرسی این شرکت‌ها برای سال‌های ۱۳۹۵ و ۱۳۹۴ استخراج گردید. بر این اساس داده‌های غیر مالی این تحقیق برابر با ۱۰۰ سال شرکتی می‌باشد که متن کاوی از این نمونه‌ها صورت گرفته است. ابتدا حدود ۱۰۰ کلمه که دارای معانی منفی بودن از گزارشات حسابرسی متن کاوی شد. سپس کلمات کلیدی که در اکثر گزارشات حسابرسی شرکت‌ها قبل از درماندگی بطور مکرر تکرار شده بودند در حدود ۲۱ واژه انتخاب گردید که در جدول شماره یک آورده شده است.

جدول ۱ کلمات کلیدی

ردیف	کلید واژه	تعداد کلید واژه در شرکت‌ها در مانده مالی	درصد فراوانی
۱	زیان انباشته	۸۵	۵,۳۱۶
۲	اقلام راکد	۷۶	۴,۷۵۳
۳	کاهش سود، کاهش ارزش کالا و دارایی‌ها	۱۱۶	۷,۲۵۵
۴	عدم رعایت مقررات و قوانین	۱۶۲	۱۰,۱۳۱
۵	عدم ایفای تعهدات	۴۹	۳,۰۶۴
۶	عدم ارائه برنامه‌های آتی	۴۶	۲,۸۷۷
۷	عدم دسترسی به اطلاعات و اسناد	۱۰۲	۶,۳۷۹
۸	عدم تاییدیه درخواستی	۲۰	۱,۲۵۱
۹	عدم احتسابات (ذخیره‌ها، مخارج و...)	۲۲	۱,۳۷۶

ردیف	کلید واژه	تعداد کلید واژه در شرکت ها درمانده مالی	درصد فراوانی
۱۰	عدم حضور مدیر ذینفع	۱۲۲	۷,۶۳۰
۱۱	عدم پرداخت مالیات	۴۸	۳,۰۰۲
۱۲	عدم ارائه به موقع و صورت های مورد نیاز (تلفیقی، میان دوره، هیئت مدیره و...)	۴۹	۳,۰۶۴
۱۳	عدم استفاده از ظرفیت های تولید	۳۱	۱,۹۳۹
۱۴	مغایرت	۱۲۳	۷,۶۹۲
۱۵	سهام وثیقه مدیران	۹۳	۵,۸۱۶
۱۶	مجوز هیات مدیره	۸۶	۵,۳۷۸
۱۷	تعدیلات سنواتی	۱۰۱	۶,۳۱۶
۱۸	تجدید ارائه	۸۵	۵,۳۱۶
۱۹	پایان خدمت	۴۳	۲,۶۸۹
۲۰	تسهیلات مالی	۷۶	۴,۷۵۳
۲۱	چک تضمینی	۶۴	۴,۰۰۳
	جمع	۱۵۹۹	۱۰۰,۰۰۰

آمار توصیفی

همان طور که در جدول دو مشاهده می گردد از ۱۰۰ سال شرکت انتخاب شده کمترین عبارت یا اصطلاح ۲۰ دفعه و بیشترین عبارت یا اصطلاح ۱۶۲ بار تکرار گردیده است. بیشترین تکرار واژه عدد ۷۶ بوده است که براساس کل کلمات بدست آمده به ترتیب میانگین ۷۱,۸۵ و انحراف معیار ۱۶۲۰ و واریانس ۴۰,۲۵ محاسبه شده است. در بخش بعدی (شکل یک) نمودار فراوانی عبارت، اصطلاحات و یا کلمات نیز ارائه شده است.

جدول ۲ آمار توصیفی

شرح	تعداد	حداقل	حداکثر	میانگین	مد	انحراف استاندارد	واریانس
شرکت درمانده مالی	۱۰۰	۲۰	۱۶۲	۷۱,۸۵	۷۶	۱۶۲۰,۷۲	۴۰,۲۵

شکل یک نمودار فراوانی اصلاحات



در نهایت با بدست آمدن میانگین داده‌ها و کسر نمودن از تعداد واژگان عبارات پر کاربرد از واژگان کم کاربرد جدا گردید. براین اساس اصطلاحات پر کاربرد که در طول دوره زمانی دو ساله قبل از درماندگی مالی شرکت‌ها بطور مستمر تکرار شده در جدول دو ارائه شده است. همانطور که در جدول دو مشاهده می‌گردد این اصطلاحات عبارتند از: عدم رعایت مقررات و قوانین (۱۶۲)، مغایرت (۱۲۳)، عدم حضور مدیر ذینفع (۱۲۲)، کاهش سود، کاهش ارزش کالا و دارایی‌ها (۱۱۶)، عدم دسترسی به اطلاعات و اسناد (۱۰۲)، تعدیلات سنواتی (۱۰۱)، سهم وثیقه مدیران (۹۳)، مجوز هیات مدیره (۸۶) و در نهایت زیان انباشته و تجدید ارائه (۸۵).

جدول ۲ اصلاحات پر کاربرد

ردیف	کلید واژه	تعداد کلید واژه در شرکت‌ها درمانده مالی	فراوانی
۱	زیان انباشته	۸۵	۷,۹۰۷
۲	کاهش سود، کاهش ارزش کالا و دارایی‌ها	۱۱۶	۱۰,۷۹۱
۳	عدم رعایت مقررات و قوانین	۱۶۲	۱۵,۰۷۰
۴	عدم دسترسی به اطلاعات و اسناد	۱۰۲	۹,۴۸۸
۵	عدم حضور مدیر ذینفع	۱۲۲	۱۱,۳۴۹
۶	مغایرت	۱۲۳	۱۱,۴۴۲
۷	سهم وثیقه مدیران	۹۳	۸,۶۵۱
۸	مجوز هیات مدیره	۸۶	۸,۰۰۰
۹	تعدیلات سنواتی	۱۰۱	۹,۳۹۵
۱۰	تجدید ارائه	۸۵	۷,۹۰۷

ردیف	کلید واژه	تعداد کلید واژه در شرکت ها درمانده مالی	فراوانی
جمع		۱۰۷۵	۱۰۰

نتیجه گیری و پیشنهاد

هدف از انجام این تحقیق، فراهم آوردن نگاهی جدید در پیش بینی ورشکستگی شرکت ها با استفاده از روش متن کاوی می باشد. در این تحقیق ابتدا ۵۰ شرکت درمانده مالی مشمول ماده ۱۴۱ پذیرفته شده در بورس اوراق بهادار سال ۹۶ انتخاب گردید سپس گزارشات حسابرسی سال های ۱۳۹۴ و ۱۳۹۵ آن ها مورد متن کاوی قرار گرفت. علت انتخاب شرکت های درمانده مالی این است که این شرکت ها می توانند نماینده های خوبی از متون مالی برای پیش بینی ورشکستگی باشند. با تاکید بر مطالب عنوان شده، متن کاوی عموماً به فرایند استخراج اطلاعات و دانش مورد نظر از متون ساختار یافته و نیافته اشاره می کند و از طرفی متن کاوی یک حوزه تحقیقاتی در حال رشد بوده که مانند هر تکنولوژی در حال ظهور، پیچیدگی های ذاتی خود را دارد. حجم عظیم اطلاعات موجود در سیستم های مالی شرکت ها باعث ایجاد چالش ها و فرصت های بسیاری شده است که نیازمند تحقیقات و مطالعات بیشتری در این زمینه می باشد. یافته های تحقیق حاکی از این است که اصطلاحات و واژگان عدم رعایت مقررات و قوانین با ۱۶۲ مورد، مغایرت با ۱۲۳ مورد، عدم حضور مدیر ذینفع با ۱۲۲ مورد، کاهش سود، کاهش ارزش کالا و دارایی ها با ۱۱۶ مورد، عدم دسترسی به اطلاعات و اسناد با ۱۰۲ مورد، تعدیلات سنواتی با ۱۰۱ مورد، سهام وثیقه مدیران با ۹۳ مورد، مجوز هیات مدیره با ۸۶ مشاهده و در نهایت زیان انباشته و تجدید ارائه با ۸۵ مشاهده در گزارشات حسابرسی سال های ۹۴ و ۹۵ شرکت های مذکور تکرار شده اند. با توجه به عبارات و واژه های استخراج شده، می توان چنین استنباط نمود که روش متن کاوی در پیش بینی درماندگی مالی شرکت ها موثر بوده و به عنوان روشی مناسب برای استخراج اطلاعات مفید از گزارشات حسابرسی قابل استفاده می باشد. اما برای اثبات برتری نیاز به انجام تحقیقات بیشتری در این زمینه است. نتایج این تحقیق با مطالعات دستگیر و شفیع سردشت [۱۱]، حاتمی ناغانی و عابسی [۳] و مای و همکاران [۲۱] همسو می باشد. آن ها نشان دادند که تکنیک داده کاوی جزوه فناوری جدیدی است که می تواند در پیش بینی درماندگی و ورشکستگی شرکت ها در امور مالی و حسابداری مورد استفاده قرار گیرد.

تکنیک متن کاوی از یک سو می تواند روش جدیدی برای تجزیه و تحلیل داده های متنی باشد و از سوی دیگر در صورت ادغام با سایر روش های پیش بینی ورشکستگی شرکت ها می تواند تاثیر بیشتری بگذارد. زیرا اولاً اطلاعات به دست آمده از متن کاوی به همراه روش های سنتی پیش بینی ورشکستگی (نسبت های مالی) می تواند اطلاعات کافی و دقیق تری را در اختیار استفاده کنندگان قرار دهد. دوماً اطلاعات به دست آمده از متن کاوی می تواند نشانه های بهنگامی قبل از وقوع ورشکستگی را نمایان سازد. علی الخصوص هنگامی که افق زمانی طولانی تر است، ارزش پیش بینی با استفاده از متن کاوی بدون توجه به ساختار متن، قابلیت مقایسه بیشتری دارد. لذا با توجه به نتایج این پژوهش به سرمایه گذاران و اعتبار دهندگان پیشنهاد می شود قبل از سرمایه گذاری یا اعطای اعتبار با استفاده از تکنیک متن کاوی و مرور ضمنی گزارشات حسابرسی سال های گذشته نسبت به واژگان و عبارت دارای بار منفی حساسیت بیشتری نشان دهند. همچنین یافته های تحقیق و این مدل ارائه شده در محیط اقتصادی کشور می تواند برای پی بینی درماندگی مالی شرکت ها مورد توجه سرمایه گذاران و تحلیلگران مالی قرار گیرد و علاقه مندان می توانند

توان پیش‌بینی مدل‌های تعدیل شده را در تحقیقات بعدی برای موسسات مالی، بانک‌ها و در سطح صنعت ارزیابی نمایند.

فهرست منابع

۱. احمدی، سید مهدی؛ پژویان، جمشید؛ غلامی، ابهام. (۱۳۹۳). "هدفمند کردن قیمت حامل‌های انرژی و رفتار مصرفی خانوارهای شهری". **مدلسازی اقتصادی**، شماره ۲۷، ۱-۱۶.
۲. بت شکن، محمدهاشم؛ پیمانی، مسلم؛ صدرالدین کرمی، محمدمسعود. (۱۳۹۷). "برآورد و ارزیابی ارزش در معرض ریسک و ریزش مورد انتظار ناپارامتریک بر مبنای تحلیل مولفه‌های اساسی در بورس اوراق بهادار تهران". **چشم انداز مدیریت مالی**، سال هشتم، شماره ۲۴، ۱۰۲-۷۹.
۳. حاتمی ناغانی، بهمن؛ عابسی، مسعود. (۱۳۹۵). "تحلیل محتوای مقالات علمی با استفاده از متن کاوی". **مدیریت کسب و کار هوشمند**، دوره ۵، شماره ۱۸، ۱۶۷-۱۳۷.
۴. دادرسی، کیوان و غریب، ایمان. (۱۳۹۴). "مقایسه تکنیک‌های داده کاوی در پیش‌بینی بحران مالی (مطالعه موردی: شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران)". **فصلنامه توسعه مدیریت پولی و بانکی**، سال ۳، شماره ۹، صص ۳۳-۵۳.
۵. رستمی، محمدرضا؛ فلاح شمس، میرفیض؛ اسکندری، فرزانه. (۱۳۹۰). "ارزیابی درماندگی مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران: مطالعه مقایسه‌ای بین تحلیل پوششی داده‌ها و رگرسیون لجستیک". **فصلنامه پژوهش‌های مدیریت در ایران - مدرس علوم انسانی**، دوره ۱۵، شماره ۳، صص ۱۲۹-۱۴۷.
۶. راهنمای رودپشتی، فریدون؛ فیروزیان، محمود؛ محمدی، لیلا. (۱۳۹۰). "گروه بندی پرتفوی شرکت سرمایه‌گذاری گروه توسعه ملی بر اساس ماتریس شبکه و مقایسه عملکرد پرتفوی‌های حاصل از این روش، با استفاده از نسبت پتانسیل مطلوب". **تحقیقات مالی**. دانشگاه تهران، دوره ۱۳، شماره ۳۲، ۳۴-۱۵.
۷. بیگی، ریحانه؛ صالحی، مهدی. (۱۳۹۲). "کاربرد الگوریتم فاخته در پیش‌بینی ورشکستگی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران"، یازدهمین همایش ملی حسابداری ایران.
۸. عثمانی، محمد قسیم، جاوید، داریوش و رحیمی، سعید. (۱۳۹۰). "بررسی تأثیر بازدارندگی مکانیزم‌های راهبری شرکت‌ها از درماندگی مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران"، **تحقیقات حسابداری و حسابرسی**، انجمن حسابداری ایران. سال ۱۲، شماره ۱۵، صص ۳-۱۲.
۹. رحیم نیا، فریبرز؛ کفاش پور، آذر؛ فیض محمدی، شیرین. (۱۳۹۳). "بررسی اثر تمایز و پرستیژ برند بر وفاداری مشتری به واسطه هویت یابی مشتری با برند (مورد مطالعه مشتریان نمایندگی‌های تویوتا)". **تحقیقات بازاریابی نوین**، دوره ۴، شماره ۳، ۱-۱۶.
۱۰. کردستانی، غلامرضا؛ تاتلی، رشید و کوثری فر، حمید. (۱۳۹۳). "ارزیابی توان پیش‌بینی مدل تعدیل شده آلتمن از مراحل درماندگی مالی نیوتن و ورشکستگی شرکت‌ها". **فصلنامه علمی پژوهش دانش سرمایه‌گذاری**، سال ۳، شماره ۹، صص ۸۳-۹۹.

۱۱. مرادی، محسن؛ شفیعی سردشت، مرتضی؛ ابراهیم پور، ملیحه. (۱۳۹۱)، "پیش بینی درماندگی مالی شرکت های بوسیله مدل های ماشین بردار پشتیبان و تحلیل ممیزی چندگانه"، **فصلنامه بورس اوراق بهادار** شماره ۱۸، صص ۱۱۳-۱۳۶
۱۲. موسوی، سید محمود و محمد رضا طبرستانی. (۱۳۸۸). "پیش بینی بحران مالی با استفاده از تحلیل پوششی داده ها"، **پژوهشات حسابداری و حسابرسی**، ش ۲، صص ۱۸۷-۱۵۸
۱۳. مشکی میاوقی، مهدیمش و هاشمی سعادت، مریم. (۱۳۹۴). "بررسی رابطه حاکمیت شرکتی با احتمال وقوع ورشکستگی در شرکت های پذیرفته شده در بورس اوراق بهادار تهران"، **پژوهش حسابداری**، شماره ۱۷، صص ۳۷-۵۷
۱۴. علی و ناظمی، عبدالرضا. **مقدمه ای بر داده کاوی**. چاپ ۱۳۹۵، انتشارات دانشگاه آزاد اسلامی
15. Bozanic, Z., and Thevenot, M. 2015. "Qualitative Disclosure and Changes in Sell Side Financial Analysts' Information Environment," **Contemporary Accounting Research**, (32:4), Wiley Online Library, pp. 1595-1616.
16. Calabrese, R., Degl'Innocenti, MY and Osmetti, S A. 2017. "The effectiveness of TARP-CPP on the US banking industry: A new copula-based approach," **European Journal of Operational Research**, (256:3), Elsevier, pp. 1029-1037.
17. Campbell, J.L., Chen, H., Dhaliwal, D. S., Lu, H., and Steele, L. B. 2014. "The information content of mandatory risk factor disclosures in corporate filings," **Review of Accounting Studies**, (19:1), Springer, pp. 396-455.
18. Doumpos, M., Andriosopoulos, K., Galariotis, E., Makridou, G., and Zopounidis, C. 2017+ "Corporate Failure Prediction in the European Energy Sector: A Multicriteria Approach and the Effect of Country Characteristics," **European Journal of Operational Research**, Elsevier.
19. Lang, M., and Stice-Lawrence, L. 2015. "Textual analysis and international financial reporting: Large sample evidence," **Journal of Accounting and Economics**, (60:2), Elsevier, pp. 110-135.
20. Liang, D., Lu, C.-C., Tsai, C.-F., and Shih, G.-A-2016. "Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study," **European Journal of Operational Research**, (252:2), Elsevier, pp. 561-572.
21. Mayew, W. J., Sethuraman, M., and Venkatachalam, M. 2015. "MD&A disclosure and the firm's ability to continue as a going concern," **Accounting Review**, (90:4), pp. 1621-1651 (doi: 10.2308/accr- 50983).
22. Maia, Feng; Shaonan Tianb, Chihoon Lee, Ling M (2018). Deep Learning Models for Bankruptcy Prediction using Textual Disclosures. **European Journal of Operational Research** (2018), doi: <https://doi.org/10.1016/j.ejor.2018.10.024>
23. J Li, W Monroe, D Jurafsky; " A simple, fast diverse decoding algorithm for neural generation" **Computer Science Department, Stanford University**, Stanford, CA, USA, Submitted on 25 Nov 2016 (v1), last revised 22 Dec 2016 (this version, v2).



Text-mining technique and prediction of financial distress

Zohreh Hajiha (Ph.D) ¹©

Associate Prof., Department of accounting, East Tehran Branch, Islamic Azad University of Tehran, Iran

Mohamad Reza Saberi Rochi²

Instructure, Department of Accounting, Phd candidate in accounting, Islamic Azad University of Birjand, Iran

(Received: 31 December 2020; Accepted: 5 May 2021)

The main goal of this research is to predict financial distress in terms of key words and phrases using the text mining technique. The statistical population of this research is the audit reports of 50 financially disadvantaged companies listed in Tehran Stock Exchange in 2017. Therefore, these companies were subjected to text mining in order to collect the required data for the research reports from the years 2015 and 2016. Research findings indicate that the term non-observance of regulations and rules with 162 cases, the word conflicts with 123 cases, the absence of the presence of the beneficiary's manager with 122, loss of profit, decrease in the value of goods and assets with 116 items, lack of access to information And documents with 102 cases, adjustments to 101 cases, bonds of directors with 93 cases, board approvals of 86, and ultimately accumulated losses and renewals with 85 views in audit reports. Given the extracted terms and expressions, the method of text mining can be used to predict corporate financial distress. It can also be used as a method to extract useful information from audit reports.

Keywords: Prediction, Text Mining Technique, Financial Distress, Audit Reports, Article 141.

¹ drzhajiha@gmail.com ©(Corresponding Author)

² 2mr.saberi4@yahoo.com