

مجله‌ی پیشرفت‌های حسابداری دانشگاه شیراز
دوره‌ی نهم، شماره‌ی دوم، پاییز و زمستان ۱۳۹۶، پیاپی ۷۳/۳، صفحه‌های ۱-۳۴
(مجله‌ی علوم اجتماعی و انسانی پیشین)

ارائه‌ی یک رویکرد محاسباتی نوین برای پیش‌بینی تقلب در صورت‌های مالی، با
استفاده از شیوه‌های خوشه‌بندی و طبقه‌بندی
(شواهدی از شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران)

مهرداد ابراهیمی**
دانشگاه شیراز

دکتر شکراله خواجهی*
دانشگاه شیراز

چکیده

رسوایی‌ها و شکست‌های شرکتی، اطمینان سرمایه‌گذاران به درست و منصفانه بودن حساب‌های واحدهای تجاری را مخدوش کرده است. تکنولوژی‌های مبتنی بر آمار و یادگیری ماشینی راهکاری اثربخش برای پیشگیری و کشف تقلب هستند؛ بنابراین در این پژوهش به بررسی این مسئله پرداخته می‌شود که آیا می‌توان از طریق شناسایی عوامل مرتبط با تقلب در صورت‌های مالی و با به‌کارگیری شیوه‌های داده‌کاوی، مدلی برای کشف تقلب در صورت‌های مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران ارائه کرد؟ برای پاسخ‌گویی به این سؤال از ۱۹ علائم خطر اشاره‌شده در استاندارد حسابداری ۲۴۰ به همراه شیوه‌های داده‌کاوی تحلیل مؤلفه‌های اساسی و خوشه‌بندی، برای تعیین شرکت‌های متقلب استفاده شد؛ سپس به‌منظور ارائه مدلی برای پیش‌بینی صورت‌های مالی متقلبان، از ۴۰ متغیر مالی و غیرمالی به همراه شیوه‌های درخت تصمیم، ماشین بردار پشتیبان و روش بوستینگ استفاده شد. یافته‌های پژوهش بیان‌گر وجود شواهدی دال بر عملکرد مناسب مدل‌های پیشنهادی برای پیش‌بینی تقلب در صورت‌های مالی است.

کلیدواژه‌ها: پیش‌بینی تقلب، تقلب در صورت‌های مالی، داده‌کاوی، بورس اوراق بهادار تهران.

* استاد حسابداری، ایران shkhajavi@gmail.com

** دانشجوی دکتری حسابداری، ایران mehrdadebrahimi66@gmail.com

تاریخ پذیرش: ۱۳۹۶/۲/۲

تاریخ دریافت: ۱۳۹۵/۵/۱۵

۱. مقدمه

کارایی، نقد شوندگی، امنیت و سلامت بازارهای مالی برای رشد و شکوفایی اقتصاد ملت‌ها ضروری است. سرمایه‌گذاران تا آنجایی در بازارهای سرمایه مشارکت می‌کنند که به کیفیت، قابلیت اتکا و شفافیت اطلاعات مالی منتشرشده شرکت‌ها اطمینان داشته باشند. اطلاعات مالی باکیفیت در صورت‌های مالی را که شرکت‌های سهامی عام تهیه کرده‌اند و حساب‌برسان مستقل این اطلاعات را حسابرسی کرده‌اند، تا حد زیادی بر اعتماد سرمایه‌گذاران تأثیر می‌گذارد. پاسخ‌گویی و مسئولیت حساب‌برس برای جستجو، کشف و گزارشگری تقلب در صورت‌های مالی^۱، اهمیت و توجه زیادی بر بازسازی اطمینان سرمایه‌گذاران و اعتماد عمومی دارد (رضایی، ۲۰۱۰).

متأسفانه رسوایی‌ها و شکست‌های شرکتی اخیر، اطمینان سرمایه‌گذاران به درست و منصفانه بودن حساب‌های واحدهای تجاری را مخدوش کرده است. در بیشتر موارد گزارش‌شده، مدیریت از طریق دست‌کاری صورت‌های مالی شرکت برای انعکاس نتایج دلخواه خود به عموم، به کلاهبرداری و پنهان ساختن آن مبادرت کرده است و به این دلیل امکان‌پذیر است که ماهیت گزارشگری مالی مستلزم قضاوت و برآورد در فرآیند تکمیل صورت‌های مالی است؛ همچنین در برخی از این موارد، حساب‌برسانی که انتظار می‌رفت استقلال داشته باشند و به اطلاعات افشاشده در حساب‌ها اعتبار بخشند، متهم شناخته شدند. اصلاحات حاکمیت شرکتی، شامل اصلاح قانون شرکت‌ها و استانداردهای حسابداری جدید، تلاش کرده‌اند تا برخی از راه‌های فرار شناسایی‌شده در تقلب‌های کشف‌شده را ببندند (اکیک، ۲۰۱۱).

تفاوت‌های اندکی از تعریف تقلب در صورت‌های مالی وجود دارد. جامعه‌ی ممیزان رسمی تقلب ۲، تقلب در صورت‌های مالی را اینچنین تعریف می‌کند: حذف یا ارائه نادرست عمدی و آگاهانه حقایق عمده یا داده‌های حسابداری به‌طور گمراه‌کننده که با در نظر گرفتن همه اطلاعات در دسترس سبب می‌شود خواننده قضاوت یا تصمیم خود را تغییر دهد. تقلب در صورت‌های مالی ممکن است در عمل مستلزم این عوامل باشد: ۱- دست‌کاری ثبت‌های مالی؛ ۲- حذف عمدی رویدادها، معاملات، حساب‌ها و سایر اطلاعات مهمی که صورت‌های مالی بر اساس آن‌ها تهیه می‌شوند؛ ۳- کاربرد نادرست اصول حسابداری، سیاست‌ها و رویه‌های مورد استفاده برای اندازه‌گیری، شناسایی، گزارشگری و افشای معاملات تجاری. این نوع از تقلب اغلب به ورشکستگی، زیان‌های قابل توجه برای سهامداران و مشکلات حقوقی برای مؤسسات

حسابرسی منجر شده است (ژو و کپور، ۲۰۱۱). تقلب در صورت‌های مالی نه تنها اثر نامساعد فزاینده‌ای بر افراد سرمایه‌گذار بلکه بر ثبات اقتصاد جهانی گذاشته است.

کشف تقلب در صورت‌های مالی به چند دلیل دشوار است؛ نخست، اعضای گروه مدیریت که عمدتاً مرتکبین اصلی تقلب هستند، به طور معمول تلاش زیادی برای پنهان ساختن تقلب در صورت‌های مالی انجام می‌دهند؛ دوم، به این دلیل که اعضای گروه مدیریت از اعتماد قابل توجهی برخوردار هستند، به راحتی کنترل‌های داخلی کلیدی را نقض می‌کنند و در نتیجه تقلب در صورت‌های مالی آسان‌تر و کشف آن سخت‌تر می‌شود؛ سوم، مرتکبین تقلب در صورت‌های مالی اغلب از تبانی و سندسازی برای ارتکاب و پنهان کردن تقلب استفاده می‌کنند و بیشتر رویه‌های حسابرسی مستقل برای کشف چنین طرح‌هایی از تقلب طراحی نشده‌اند؛ چهارم، داده‌های صورت‌های مالی تا حد زیادی خلاصه و تجمیع شده‌اند که پنهان کردن تقلب را ساده‌تر و کشف آن را با استفاده از روش‌های مدل‌سازی آماری و تحلیلی سخت‌تر می‌کند (وایتینگ و همکاران، ۲۰۱۲).

رویکردهای جدیدی برای کشف تقلب وجود دارد؛ اما از داده‌کاوی ۳ که هدف آن استخراج اطلاعات موردعلاقه از مجموعه گسترده‌ای از داده‌ها است، به‌طور گسترده‌ای به‌عنوان ابزار تصمیم‌گیری فعال، استفاده می‌شود. این رویکرد باعث افزایش توجه پژوهش‌گران در غنی کردن داده‌ها، باهدف تسهیل کشف الگوهای موردعلاقه شده است (وایتینگ و همکاران، ۲۰۱۲). داده‌کاوی در یک سناریوی تجزیه و تحلیل اکتشاف که هیچ ایده‌ای از پیش تعیین شده‌ای در رابطه با عوامل تشکیل‌دهنده نتیجه موردعلاقه وجود ندارد، بسیار سودمند است. به‌کارگیری شیوه‌های داده‌کاوی برای طبقه‌بندی مالی زمینه‌ای ناب برای پژوهش است. بسیاری از دستگاه‌های مجری قانون و بازرسی خاص که مأموریت آن‌ها کشف فعالیت‌های متقلبانه است، نیز به‌طور موفقیت‌آمیزی از داده‌کاوی استفاده می‌کنند؛ با این وجود، برخلاف سایر زمینه‌های پژوهش مانند پیش‌بینی ورشکستگی یا درماندگی مالی، پژوهش‌اندکی در رابطه با کاربرد شیوه‌های داده‌کاوی برای کشف تقلب مدیریت انجام شده است (کرکوز و همکاران، ۲۰۰۷)؛ بنابراین در این پژوهش به بررسی این مسئله پرداخته می‌شود که آیا می‌توان از طریق شناسایی عوامل مرتبط با تقلب در صورت‌های مالی و با به‌کارگیری شیوه‌های خوشه‌بندی و طبقه‌بندی، مدلی برای کشف تقلب در صورت‌های مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران ارائه کرد؟ حساب‌رسان با به‌کارگیری شیوه‌های داده‌کاوی می‌توانند وظیفه خود برای

کشف تقلب مدیران را تسهیل نمایند. بردار ورودی، اطلاعات استخراج شده از گزارش‌های سالانه شرکت‌ها است. شیوه‌های معرفی شده، برحسب عملکرد نیز با یکدیگر مقایسه خواهند شد. دو جریان متفاوت در متون پژوهشی، پیشینه و مبنایی برای این پژوهش فراهم می‌کند؛ یکی پژوهش‌های انجام شده در رابطه با گزارشگری مالی متقلبانه و دیگری متون مربوط به داده‌کاوی است. این دو جریان قلمروهای پژوهشی گسترده‌ای هستند که پژوهش‌های متقابل بسیار اندکی در داخل کشور، در این زمینه انجام شده است. پژوهش‌های صورت گرفته نیز با توجه به محدودیت در زمینه دسترسی به داده‌های تقلب با یکدیگر متفاوت هستند؛ از این رو، این پژوهش در تلاش است تا با ارائه رویکردی جدید برای تعیین شرکت‌های متقلب و در آخر معرفی رویکردهای داده‌کاوی بدیع برای انتخاب متغیرهای مؤثر به منظور پیش‌بینی و طبقه‌بندی تقلب در صورت‌های مالی، متون پژوهشی مربوط را بهبود بخشد.

۲. مبانی نظری

واحدهای تجاری در سرتاسر جهان برای رشد و توسعه نیازمند تأمین مالی هستند؛ باین وجود سرمایه‌گذاران بالقوه تنها در آن دسته از واحدهای تجاری سرمایه‌گذاری می‌کنند که از سلامت مالی و پتانسیل رشد و توسعه برخوردار باشند. گزارش‌ها و حساب‌های مالی سالانه و سایر اطلاعات در دسترس عموم واحدهای تجاری، درک سرمایه‌گذاران در رابطه با چگونگی اداره امور شرکت را بهبود می‌بخشد. با توجه به اینکه این گزارش‌ها تحت الزامات قانونی شدید و استانداردهای گزارشگری تجویز شده تهیه می‌شوند و همچنین موضوع حسابرسی مستقل هستند، سرمایه‌گذاران برای کسب کردن اطمینان مجدد به دوام مالی واحدهای تجاری، به این گزارش‌ها می‌نگرند (اکیک، ۲۰۱۱).

طبق استانداردهای حسابرسی، «حسابرس باید تحریف‌های شناسایی شده در جریان حسابرسی را، غیر از مواردی که مشخصاً ناچیز است، گردآوری کند» (کمیته تدوین استانداردهای حسابرسی، بخش ۴۵۰، بند ۵). «حسابرس باید درباره اینکه آیا صورت‌های مالی، از تمام جنبه‌های بااهمیت، طبق استانداردهای حسابداری، به نحو مطلوب تهیه شده است یا خیر، اظهار نظر کند» (کمیته تدوین استانداردهای حسابرسی، بخش ۷۰۰، بند ۹). «حسابرس برای تعیین نوع اظهار نظر باید نتیجه‌گیری کند که آیا نسبت به عاری بودن یا نبودن صورت‌های مالی از تحریف‌های بااهمیت ناشی از تقلب یا اشتباه، اطمینان معقول کسب کرده

است یا خیر» (کمیته تدوین استانداردهای حسابرسی، بخش ۷۰۰، بند ۱۰). «حسابرس زمانی باید اظهار نظر تعدیل نشده ارائه کند که بر اساس نتیجه‌گیری او صورت‌های مالی، از تمام جنبه‌های بااهمیت، طبق استانداردهای حسابداری، به نحو مطلوب تهیه شده است» (کمیته تدوین استانداردهای حسابرسی، بخش ۷۰۰، بند ۱۴).

تقلب در صورت‌های مالی در طول سال‌های گذشته بیش از ۵۰۰ میلیارد دلار به سرمایه‌گذاران زیان وارد کرده است. برآورد می‌شود که تقلب در صورت‌های مالی شرکت انرون حدود ۷۰ میلیون دلار برای سرمایه‌گذاران، کارکنان و بازنشستگانی که سهام شرکت را در حساب‌های خود نگهداری می‌کردند، زیان به همراه داشته است. اداره حسابداری عمومی آمریکا در سال ۲۰۰۲ گزارشی منتشر کرد که نشان می‌داد از ژانویه ۱۹۹۷ تا ژوئن ۲۰۰۲ تعداد تجدید ارائه‌ها به دلیل ناهنجاری‌های حسابداری به‌طور معناداری افزایش یافته است. این مطالعه نشان می‌دهد که حدود ۱۰ درصد از شرکت‌های پذیرفته‌شده در بورس، حداقل یک‌بار اعلام تجدید ارائه کرده‌اند که حاکی از رشد ۱۴۵ درصدی آن در طول این دوره است و انتظار می‌رود تا پایان سال ۲۰۰۲ به ۱۷۰ درصد برسد (رضایی، ۲۰۰۵).

تقلب در صورت‌های مالی و کاهش ارزش بازار سهام نشان‌دهنده اهمیت کیفیت گزارش‌های مالی و کارکرد حسابرسی و درک عوامل بالقوه رسوایی‌های حسابداری است. سقوط شرکت‌های بسیاری مانند انرون، وردکام و گلوبال کراسینگ، اثربخشی حاکمیت شرکتی، کیفیت گزارش‌های مالی و قابلیت اعتماد کارکردهای حسابرسی را با تردید همراه ساخت. این تقلب‌ها در صورت‌های مالی، نگرانی‌های جدی را درباره این عوامل به همراه داشت: ۱- اثربخشی حاکمیت شرکتی؛ ۲- صداقت و رفتار اخلاقی مدیران ارشد، به‌ویژه زمانی که آن‌ها متهم به دست‌کاری حساب‌ها شده‌اند؛ ۳- کفایت و اثربخشی کنترل‌های داخلی؛ ۴- قابلیت اتکای گزارش‌های مالی؛ ۵- کیفیت حسابرسی؛ ۶- درستی بازارهای سهام (رضایی، ۲۰۰۵).

تقلب در صورت‌های مالی بسیار گسترده است و مورد توجه سرمایه‌گذاران، تحلیل‌گران و سایرین است؛ زیرا مستلزم فریب ذی‌نفعان توسط مدیریت از طریق افزایش مصنوعی درآمدها و بهبود شفافیت مالی شرکت است. تقلب در صورت‌های مالی مستلزم نقض اصول عمومی پذیرفته‌شده حسابداری در جهت ارائه عملکرد مالی مطلوب شرکت است. این نوع از تقلب اغلب به ورشکستگی، زیان‌های قابل توجه برای سهامداران و مشکلات حقوقی برای مؤسسات حسابرسی منجر شده است. تقلب در صورت‌های مالی، به موارد اعلام‌شده‌ای مانند انرون،

ورد کام، هلت سوئ و غیره محدود نخواهد شد. تقلب در صورت‌های مالی را می‌توان به دودسته کشف شده (گزارش شده) و کشف نشده تقسیم کرد. تنها بخش کوچکی از تقلب، در صورت‌های مالی کشف (گزارش) شده‌اند و سایر موارد تا زمانی که کشف نشوند ادامه خواهند یافت. در بسیاری از موارد تقلب را سال‌ها پس از ارتکاب آن، مدیران ارشد کشف کرده‌اند. در محیطی فعال از تقلب در صورت‌های مالی، سازوکارهای کشف تقلب با کمک رایانه بسیار مؤثرتر و کارا تر خواهد بود. تکنولوژی‌های مبتنی بر آمار و یادگیری ماشینی راهکاری اثربخش برای پیشگیری و کشف تقلب هستند؛ اما مرتکبین تقلب خود را انطباق می‌دهند و به طور معمول قادر به کشف راه‌هایی برای دور زدن این تکنولوژی‌ها هستند. شیوه‌های کنونی کشف تقلب در اکثر موقعیت‌های مرتبط با تقلب، از اصول داده‌کاوی مشابهی استفاده می‌کنند؛ اما می‌توانند از لحاظ دانش قلمرو تخصص متفاوت باشند. زمانی که مدیران مالی درگیر در تقلب، از نرم‌افزارها و شیوه‌های کشف تقلب اطلاع کافی دارند، روش‌هایی را برای ارتکاب تقلب بکار می‌گیرند که کشف آن‌ها، به‌ویژه با استفاده از شیوه‌های فعلی دشوار است. نیازی مبرم برای روش‌هایی وجود دارد که نه تنها کارا باشند، بلکه برای کشف فریبکاری‌های مالی انطباقی و نوظهور مؤثر باشند (ژو و کپور، ۲۰۱۱).

۳. پیشینه پژوهش

۳-۱. پژوهش‌های خارجی

پرسونس (۱۹۹۵) با نمونه‌ای متشکل از ۱۰۳ شرکت متقلب در سال وقوع و ۱۰۰ شرکت متقلب در سال قبل از وقوع تقلب، توانایی داده‌های صورت‌های مالی برای شناسایی عوامل مرتبط با گزارشگری مالی متقلبان را بررسی کرد. نتایج حاصل از مدل‌های لجستیک گام‌به‌گام ۴ حاکی از این بود که اهرم مالی، گردش سرمایه، ترکیب دارایی‌ها و اندازه شرکت، عوامل معنادار مرتبط با گزارشگری مالی متقلبان هستند.

فروز و همکاران (۲۰۰۰) با نمونه‌ای متشکل از ۱۳۲ شرکت شامل ۴۲ شرکت متقلب و ۹۰ شرکت غیر متقلب، سودمندی علائم مالی و غیرمالی خطر تقلب را در کشف گزارشگری مالی متقلبان بررسی کردند. در مجموع چهار متغیر مالی شامل سودآوری، دفعات پرداخت بهره، نسبت حساب‌های دریافتی به فروش، امتیاز Z آلتمن و سه متغیر غیرمالی شامل تغییر مدیر ارشد اجرایی، تغییر مدیر ارشد مالی و تغییر حسابرس، به عنوان علائم خطر تقلب در صورت‌های

مالی انتخاب شدند. نتایج حاصل از شبکه‌های عصبی مصنوعی ۵ بیان‌گر اهمیت به کارگیری روش‌های تحلیلی، شامل نسبت‌های مالی و متغیرهای غیرمالی در هر کار حسابرسی است. سپاتیس (۲۰۰۲) با نمونه‌ای متشکل از ۷۶ شرکت شامل ۳۸ شرکت متقلب و ۳۸ شرکت غیر متقلب، توانایی اطلاعات حسابداری در کشف گزارشگری مالی متقلبان را بررسی کرد. در مجموع ۱۰ متغیر مالی به‌عنوان پیش‌بینی‌کننده‌های بالقوه تقلب در صورت‌های مالی انتخاب شدند. نتایج حاصل از رگرسیون لجستیک گام‌به‌گام حاکی از این بود که شرکت‌های با درصد بالای موجودی کالا به فروش، نسبت بالای بدهی به کل دارایی‌ها، نسبت پایین سود خالص به کل دارایی‌ها، نسبت پایین سرمایه در گردش به کل دارایی‌ها و امتیاز پایین Z آلتمن، با احتمال بیشتری صورت‌های مالی خود را دست‌کاری می‌کنند.

کامینسکی و همکاران (۲۰۰۴) با نمونه‌ای متشکل از ۷۹ شرکت متقلب و ۷۹ شرکت غیر متقلب و با استفاده از ۱۶ نسبت مالی به بررسی توانایی نسبت‌های مالی در کشف گزارشگری مالی متقلبان پرداختند. مبنای آن‌ها برای تعیین شرکت‌های متقلب، نشریه‌های اجرایی کمیسیون بورس و اوراق بهادار طی سال‌های ۱۹۹۹-۱۹۸۲ بود. به‌طور کلی، نتایج حاصل از تحلیل ممیزی^۶ بیان‌گر توانایی محدود نسبت‌های مالی در کشف یا پیش‌بینی گزارشگری مالی متقلبان بود.

کرکوز (۲۰۰۷) با نمونه‌ای متشکل از ۳۸ شرکت متقلب و ۳۸ شرکت غیر متقلب و با استفاده از ۱۰ نسبت مالی به بررسی سودمندی درخت‌های تصمیم^۷، شبکه‌های عصبی مصنوعی و شبکه‌های باور بیزین^۸ در کشف صورت‌های مالی متقلبان پرداخت. نتایج پژوهش حاکی از این بود که صورت‌های مالی منتشرشده دربرگیرنده اطلاعات سودمندی برای کشف تقلب در گزارشگری مالی هستند؛ همچنین مدل شبکه باور بیزین عملکرد بهتری نسبت به مدل شبکه عصبی مصنوعی و درخت تصمیم داشت و خطای نوع اول در هر سه مدل کمینه بود.

آلدن و همکاران (۲۰۱۲) با نمونه‌ای متشکل از ۴۵۸ شرکت شامل ۲۲۹ شرکت متقلب و ۲۲۹ شرکت غیر متقلب، سودمندی طبقه‌بندی‌کننده‌های مبتنی بر قواعد فازی^۹ در کشف الگوهای گزارشگری مالی متقلبان را بررسی کردند. در مجموع ۱۸ متغیر مالی به‌عنوان پیش‌بینی‌کننده‌های بالقوه گزارشگری مالی متقلبان انتخاب شدند. نتایج حاصل از طبقه‌بندی‌کننده‌های مبتنی بر قواعد فازی با الگوریتم ژنتیک و الگوریتم برآورد توزیع یادگیری مارکو^{۱۰}،

حاکمی از سودمندی مدل‌های تکاملی برای کشف گزارشگری مالی متقلبانه بود. هوانگ و همکاران (۲۰۱۴) با نمونه‌ای متشکل از ۱۴۴ شرکت شامل ۷۲ شرکت متقلب و ۷۲ شرکت غیر متقلب، به ارائه مدلی مبتنی بر نگاه خودسازمان ده سلسله مراتبی در حال رشد ۱۱، برای کشف تقلب در گزارشگری مالی پرداختند. در مجموع ۲۴ متغیر مستقل برای انجام تحلیل ممیزی به منظور تعیین ورودی‌های مدل انتخاب شدند و نتایج حاصل از تحلیل ممیزی حاکمی از این بود که اثرات ۸ متغیر از نظر آماری معنادار است. این متغیرها وضعیت یک شرکت را از جنبه‌های سودآوری، نقدینگی، ساختار مالی، دسترسی به وجه نقد، در ماندگی مالی و حاکمیت شرکتی اندازه‌گیری می‌کنند. نتایج حاصل از پژوهش بیان‌گر عملکرد مناسب مدل پیشنهادی برای کشف گزارشگری مالی متقلبانه بود.

چن و همکاران (۲۰۱۴) با نمونه‌ای متشکل از ۹۴ شرکت شامل ۴۷ شرکت متقلب و ۴۷ شرکت غیر متقلب و با استفاده از درخت‌های تصمیم، تحلیل مجموعه راف یکپارچه ۱۲ و شبکه‌های عصبی به ارائه مدلی برای کشف گزارشگری مالی متقلبانه پرداختند. در مجموع ۲۱ متغیر مالی و ۱۱ متغیر حاکمیت شرکتی به‌عنوان پیش‌بینی‌کننده‌های بالقوه تقلب در صورت‌های مالی انتخاب شدند. نتایج حاصل از پژوهش حاکمی از سودمندی مدل پیشنهادی برای کشف تقلب در صورت‌های مالی بود.

لین و همکاران (۲۰۱۶) با نمونه‌ای متشکل از ۱۲۹ شرکت متقلب و ۴۴۷ شرکت غیر متقلب به طبقه‌بندی و رتبه‌بندی عوامل تقلب طی سال‌های ۲۰۱۰-۱۹۹۸ پرداختند. در این پژوهش ۳۲ عامل تقلب که مطابق نظر متخصصان برای کشف تقلب مناسب شناخته شدند، بکار رفته است. نتایج حاصل از شیوه‌های داده‌کاوی رگرسیون لجستیک، درخت تصمیم و شبکه‌های عصبی مصنوعی حاکمی از دقت بیشتر روش‌های درخت تصمیم و شبکه‌های عصبی مصنوعی نسبت به رگرسیون لجستیک است. در پایان نیز به منظور بهبود دستاوردهای پژوهش مقایسه‌ای بین قضاوت متخصصان و شیوه‌های داده‌کاوی صورت گرفته است.

۲-۳. پژوهش‌های داخلی

صفرزاده (۱۳۸۹) با نمونه‌ای متشکل از ۶۶ شرکت متقلب و ۱۱۲ شرکت غیر متقلب و با استفاده از ۱۰ نسبت مالی به بررسی نقش داده‌های حسابداری در ایجاد یک الگو برای کشف عوامل مرتبط با تقلب در گزارشگری مالی پرداخت. شرکت‌های متقلب بر مبنای ۱- شمول

شرکت در فهرست سازمان بورس و اوراق بهادار به دلایلی مرتبط با تحریف داده‌های مالی و ۲- انجام دادن معاملات نهانی و آرای صادر شده توسط دادگاه در خصوص تحریف در گزارشگری مالی، انتخاب شدند. نتایج پژوهش حاکی از عملکرد مناسب الگوی پیشنهادی در طبقه‌بندی شرکت‌های نمونه داشت.

اعتمادی و زلفی (۱۳۹۲) با نمونه‌ای متشکل از ۶۸ شرکت شامل ۳۴ شرکت متقلب و ۳۴ شرکت غیر متقلب و با استفاده از ۹ نسبت مالی به بررسی کاربرد رگرسیون لجستیک در شناسایی گزارشگری مالی متقلبانه پرداختند. در این پژوهش برای تفکیک شرکت‌های متقلب و غیر متقلب از این معیارها استفاده شد: ۱- اظهارنظر غیر مقبول حسابرسی؛ ۲- وجود اختلاف مالیاتی با حوزه مالیاتی؛ ۳- وجود تعدیلات سنواتی بااهمیت. نتایج پژوهش بیان‌گر سودمندی مدل پیشنهادی در شناسایی گزارشگری مالی متقلبانه بود.

فرقاندوست حقیقی و همکاران (۱۳۹۳) با نمونه‌ای متشکل از ۱۱۵ شرکت متقلب و ۱۱۵ شرکت غیر متقلب به بررسی رابطه مدیریت سود و امکان تقلب در صورت‌های مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران پرداختند. نتایج این پژوهش حاکی از این بود که در شرکت‌های با سابقه مدیریت سود، امکان ارتکاب تقلب در صورت‌های مالی وجود دارد؛ همچنین در صورت وجود سابقه مدیریت سود، وجود عوامل انگیزشی سبب افزایش احتمال ارتکاب تقلب در صورت‌های مالی می‌شود.

جهانشاد و سرداری‌زاده (۱۳۹۳) با نمونه‌ای متشکل از ۸۰ شرکت شامل ۴۰ شرکت متقلب و ۴۰ شرکت غیر متقلب، به بررسی رابطه معیار مالی (اختلاف رشد درآمد) و معیار غیرمالی (رشد تعداد کارکنان) با گزارشگری مالی متقلبانه پرداختند. شرکت‌های متقلب بر مبنای شمول شرکت در فهرست سازمان بورس اوراق بهادار تهران به دلایلی مرتبط با ۱- تحریف داده‌های مالی و ۲- انجام معاملات نهانی در خصوص تحریف در گزارشگری مالی، انتخاب شده‌اند. نتایج پژوهش بیان‌گر رابطه‌ای منفی و معنادار میان رشد درآمد، رشد کارکنان و اختلاف این دو متغیر با گزارشگری مالی متقلبانه بود.

۴. روش پژوهش

این پژوهش کاربردی و طرح آن از نوع شبه تجربی و با استفاده از رویکرد پس‌رویدادی است. طرح‌های شبه تجربی مستلزم تخصیص تصادفی واحدهای آزمون به تیمارهای تجربی یا

تخصیص تصادفی تیمار تجربی به واحدهای آزمون نیست (سریجش و همکاران، ۲۰۰۸). در این حالت، طرح‌های شبه تجربی به کنترل متغیرهای برون‌زا کمک می‌کنند؛ اما به دلیل کنترل عوامل مرتبط با کاهش روایی داخلی به‌اندازه‌ی طرح‌های تجربی واقعی سودمند نیستند (زیکموند و همکاران، ۲۰۰۸). برخی مواقع، طرح‌های شبه تجربی تنها راه برای انجام پژوهش هستند. طرح‌های سری زمانی، معروف‌ترین طرح شبه تجربی است که پژوهش‌گران استفاده می‌کنند.

۴-۱ سؤال‌های پژوهش

با توجه به مبانی نظری و پیشینه پژوهش، سؤال‌هایی به شرح زیر تدوین شد:
 آیا شیوه‌های داده‌کاوی امکان تفکیک شرکت‌های متقلب و غیرمتقلب را با استفاده از گزارش‌های حسابرسی سالانه شرکت‌ها که به سازمان بورس و اوراق بهادار ارائه می‌شود، فراهم می‌کنند؟
 آیا می‌توان با استفاده از اطلاعات مالی و غیرمالی گزارش‌های سالانه شرکت‌ها که به سازمان بورس و اوراق بهادار ارائه می‌شود، مدلی کمی برای کشف تقلب ارائه داد که بیان‌گر یک روش تحلیلی خودکار برای کشف تقلب بالقوه باشد؟

۴-۲ متغیرهای پژوهش

۴-۲-۱ متغیر پاسخ

متغیر پاسخ در این پژوهش، تقلب در صورت‌های مالی است. در این پژوهش برای تعیین شرکت‌های متقلب از علائم خطر استاندارد حسابرسی شماره ۲۴۰ با عنوان «مسئولیت حسابرس در ارتباط با تقلب و اشتباه در صورت‌های مالی» با به‌کارگیری روش‌های داده‌کاوی غیر نظارتی، به همراه شیوه‌های کاهش داده و بهینه‌سازی استفاده می‌شود. علائم نشان‌دهنده احتمال وجود تحریف‌های ناشی از گزارشگری مالی متقلبان، به این شرح ارائه می‌شود: ۱- بیش‌نمایی موجودی کالا؛ ۲- بیش‌نمایی حساب‌ها و اسناد دریافتی؛ ۳- بیش‌نمایی دارایی‌های ثابت؛ ۴- بیش‌نمایی سرمایه‌گذاری‌ها؛ ۵- کسری ذخیره مطالبات مشکوک‌الوصول؛ ۶- کسری استهلاک؛ ۷- بیش‌نمایی درآمدها؛ ۸- بیش‌نمایی سود؛ ۹- بیش‌نمایی سود انباشته؛ ۱۰- کم‌نمایی حساب‌ها و اسناد پرداختی؛ ۱۱- کسری ذخیره مالیات؛ ۱۲- بدهی‌های احتمالی؛

۱۳- کسری ذخیره مزایای پایان خدمت کارکنان؛ ۱۴- کم‌نمایی هزینه‌ها؛ ۱۵- حساب‌ها و اسناد دریافتی که مدت زیادی از سررسید آن‌ها گذشته است؛ ۱۶- موجودی راکد؛ ۱۷- دارایی راکد؛ ۱۸- اشتباه در به‌کارگیری استانداردهای حسابداری مرتبط با اندازه‌گیری، شناسایی، طبقه‌بندی، ارائه یا افشا. چنانچه هر یک از این علائم در بندهای تعدیلی گزارش‌های حسابرسی سالانه شرکت‌ها وجود داشته باشد، مقدار آن برابر با یک و در غیر این صورت صفر خواهد بود.

۲-۲-۴. متغیرهای پیش‌بینی‌کننده

در این پژوهش، عوامل مالی و غیرمالی خطر تقلب مرتبط با گزارشگری مالی متقلبانه، متغیرهای پیش‌بین هستند که با مرور پژوهش‌های پیشین از قبیل ستایس (۱۹۹۱)، سامرز و سوینی (۱۹۹۸)، پرسونس (۱۹۹۵)، سپاتیس (۲۰۰۲)، سپاتیس و همکاران (۲۰۰۲)، کامینسکی و همکاران (۲۰۰۴)، کرکوز و همکاران (۲۰۰۷)، برازل و همکاران (۲۰۰۹)، آلدن و همکاران (۲۰۱۲)، چن و همکاران (۲۰۱۴)، صفرزاده (۱۳۸۹)، اعتمادی و زلّقی (۱۳۹۲)، مرادی و همکاران (۱۳۹۳)، خواجوی و قدیریان آرانی (۱۳۹۴) و پیری و برزگری صدقیانی (۱۳۹۵) تعیین شده است. در یک دسته‌بندی کلی، این عوامل عبارت‌اند از:

اهرم مالی که برای اندازه‌گیری آن از لگاریتم کل بدهی و نسبت‌های کل بدهی‌ها به کل دارایی‌ها، بدهی‌های بلندمدت به کل دارایی‌ها، کل بدهی‌ها به حقوق صاحبان سهام، بدهی‌های بلندمدت به حقوق صاحبان سهام و نسبت پوشش بهره استفاده می‌شود؛

سودآوری که شامل متغیرهای بازده دارایی‌ها، بازده حقوق صاحبان سهام، سود قبل از بهره و مالیات، سود هر سهم و نسبت‌های سود خالص به فروش، سود ناخالص به فروش، سود عملیاتی به فروش، سود ناخالص به کل دارایی‌ها، سود انباشته به کل دارایی‌ها و نسبت سود خالص به دارایی‌های ثابت است؛

ترکیب دارایی‌ها که با استفاده از متغیرهای تغییر در حساب‌های دریافتی، تغییر در موجودی کالا و نسبت‌های جمع دارایی‌های جاری به کل دارایی‌ها، حساب‌های دریافتی به کل دارایی‌ها، موجودی کالا به فروش خالص، حساب‌های دریافتی به فروش خالص و نسبت موجودی کالا به جمع دارایی‌ها اندازه‌گیری می‌شود؛ نقدینگی که شامل نسبت‌های آبی، نقدی، جاری، وجه نقد به کل دارایی‌ها و نسبت سرمایه در گردش به کل دارایی‌ها است؛

کارایی که برای اندازه‌گیری آن از نسبت‌های فروش به کل دارایی‌ها، فروش به کل دارایی‌های ثابت، فروش به حساب‌های دریافتی و بهای تمام‌شده کالای فروش رفته به

موجودی کالا استفاده می‌شود؛ اندازه شرکت که با متغیرهای لگاریتم طبیعی کل دارایی‌ها، لگاریتم طبیعی فروش و لگاریتم طبیعی ارزش بازار شرکت اندازه‌گیری می‌شود؛ رشد شرکت و وضعیت مالی کلی آن که به ترتیب با استفاده از متغیرهای رشد فروش و Z آلتمن اندازه‌گیری می‌شود؛ عوامل حسابرسی که شامل اندازه حسابرسی، دوره تصدی حسابرسی و تغییر حسابرسی می‌شود.

۳-۴. مدل‌های پژوهش

داده‌های روش‌های گوناگونی مانند مالی، مهندسی، زیست‌پزشکی^{۱۳} و امنیت سایبری^{۱۴} استفاده می‌شود. روش‌های داده‌کاوی به دودسته تقسیم می‌شوند: نظارت‌شده^{۱۵} و غیر نظارت‌شده^{۱۶}. شیوه‌های داده‌کاوی نظارت‌شده با داده‌های آموزشی، یک تابع پنهان را پیش‌بینی می‌کنند. داده‌های آموزشی چندین جفت متغیر ورودی و برچسب‌ها یا کلاس‌های خروجی دارند. خروجی روش قادر خواهد بود برچسب کلاس، متغیرهای ورودی را پیش‌بینی کند. طبقه‌بندی و پیش‌بینی، مثال‌هایی از کاوش نظارت‌شده هستند. داده‌کاوی غیر نظارتی، تلاش برای شناسایی الگوهای مخفی با استفاده از داده‌های معین و بدون معرفی داده‌های آموزشی، مانند جفت‌های ورودی و برچسب کلاس، هستند. خوشه‌بندی^{۱۷} و کاوش قواعد وابستگی^{۱۸} نمونه‌هایی از کاوش غیر نظارتی هستند. داده‌کاوی همچنین بخشی از کشف دانش در پایگاه‌های داده، فرآیندی تکراری برای استخراج اطلاعات غیر بدیهی از داده‌ها، است. کشف شناخت در پایگاه داده دربرگیرنده چندین مرحله از جمع‌آوری داده‌های خام تا ایجاد شناخت جدید می‌شود. این فرآیند تکراری از گام‌های زیر تشکیل می‌شود: پاک‌سازی داده‌ها^{۱۹}؛ تجمیع داده‌ها^{۲۰}؛ انتخاب داده‌ها^{۲۱}؛ انتقال داده‌ها^{۲۲}؛ ارزیابی الگو^{۲۳}؛ نمایش شناخت^{۲۴} (دوا و دو، ۲۰۱۱).

در این پژوهش، پس از نرمال‌سازی داده‌ها، برای تعیین شرکت‌های متقلب، از روش‌های داده‌کاوی غیر نظارتی به همراه شیوه‌های کاهش داده و بهینه‌سازی استفاده می‌شود. در ادامه به منظور پیش‌بینی صورت‌های مالی متقلبانه از شیوه‌های داده‌کاوی نظارت‌شده مانند ماشین‌های بردار پشتیبان^{۲۵} و درخت‌های تصمیم و الگوریتم بوستینگ^{۲۶} استفاده خواهد شد.

۱-۳-۴. کاهش داده

شیوه‌های کاهش داده را برای دستیابی به تصویری کاهش‌یافته از مجموعه داده می‌توان اعمال کرد؛ این شیوه حجم بسیار کمتری دارد و هم‌زمان یکپارچگی داده‌های اولیه را تا حد زیادی حفظ می‌کند؛ به بیان دیگر، کاوش داده‌های کاهش‌یافته باید کاراتر بوده و نتایج تحلیلی مشابه یا تقریباً مشابهی داشته باشد. راهبردهای کاهش داده، دربرگیرنده کاهش ابعاد^{۲۷}، کاهش تعداد^{۲۸} و فشردن داده‌ها^{۲۹} هستند (هان و همکاران، ۲۰۱۲).

کاهش ابعاد، فرآیند کاهش تعداد متغیرها یا خصیصه‌های تصادفی تحت بررسی است. روش‌های کاهش ابعاد عبارت‌اند از تبدیل موجک^{۳۰} و تحلیل مؤلفه‌های اساسی^{۳۱} که داده‌های اولیه را به فضای کوچک‌تری تبدیل می‌کنند یا نمایش می‌دهند. انتخاب زیرمجموعه‌ای از خصیصه‌ها^{۳۲}، روشی برای کاهش ابعاد است که در آن، خصیصه‌ها یا ابعاد نامرتبط، کم‌مرتبط یا تکراری کشف و حذف می‌شوند (هان و همکاران، ۲۰۱۲).

۲-۳-۴. خوشه‌بندی

خوشه‌بندی، تخصیص اشیا به گروه‌ها (خوشه‌ها) است به گونه‌ای که اشیای یک خوشه نسبت به اشیای خوشه‌های متفاوت شباهت بیشتری با یکدیگر دارند. همانندی اشیا اغلب از طریق فاصله بین اشیا بر روی ابعاد چندگانه مجموعه داده تعیین می‌شود. خوشه‌بندی به‌طور گسترده‌ای در حوزه‌های مختلفی مانند بیوانفورماتیک^{۳۳}، متن‌کاوی^{۳۴}، شناخت الگو^{۳۵} و تحلیل تصویر^{۳۶} به کار رفته است. خوشه‌بندی رویکردی از یادگیری غیر نظارتی است که مثال‌ها برچسب ندارند؛ به بیان دیگر از قبل طبقه‌بندی نشده‌اند. در این پژوهش به منظور تعیین شرکت‌های متقلب از روش خوشه‌بندی k -Means استفاده شده است (دو و دو، ۲۰۱۱).

خوشه‌بندی k -Means: این روش، نقاط داده مشخص X را به k خوشه تفکیک می‌کند که در آن هر نقطه داده بسیار مشابه با مرکز خوشه خود و نه مراکز سایر خوشه‌ها است. به‌طور کلی، الگوریتم خوشه‌بندی k -means شامل گام‌های زیر است:

گام ۱- انتخاب k مراکز خوشه اولیه، $C_1, C_2, C_3, \dots, C_k$ ؛

گام ۲- تخصیص هر نمونه x در S به خوشه‌ای که نزدیک‌ترین مرکز به x را دارد؛

گام ۳- محاسبه مجدد مرکز هر خوشه بر مبنای عناصری که در آن قرار گرفته‌اند؛

گام ۴- تکرار گام‌های ۲ و ۳ تا زمانی که همگرایی حاصل شود.

دو مسئله کلیدی در پیاده‌سازی موفق روش k -means اهمیت دارد: تعداد خوشه k برای تفکیک و معیار فاصله. فاصله اقلیدسی، مهم‌ترین مقیاس برای استفاده در خوشه‌بندی k -means است (دوا و دو، ۲۰۱۱). پیاده‌سازی این روش ساده و اغلب کاربردی است؛ اما نتایج الگوریتم k -means به طور قابل توجهی به مقادیر اولیه بستگی دارد؛ به بیان دیگر، کارایی آن به شدت به مراکز و پارامترهای خوشه اولیه وابسته است (دوا و دو، ۲۰۱۱). در این پژوهش برای حل این مشکل از الگوریتم بهینه سازی کلونی زنبور عسل مصنوعی (ABC) استفاده خواهد شد.

۳-۳-۴. انتخاب ویژگی

داده‌های نامتوازن به طور معمول با فضای ویژگی با ابعاد بالا همراه هستند. در میان ویژگی‌های با ابعاد بالا، وجود تعداد زیادی ویژگی‌های پر اختلال^{۳۷} می‌تواند عملکرد طبقه‌بندی کننده را مختل و کاهش دهد. در سال‌های اخیر، روش‌های انتخاب ویژگی و زیر فضا برای حل این مسئله معرفی و ارزیابی شدند. روش‌های انتخاب زیرمجموعه ویژگی برای انتخاب زیرمجموعه ویژگی کوچکی، از میان ویژگی‌های با ابعاد بالا مطابق با معیارهای انتخاب ویژگی به کار می‌رود (دوا و دو، ۲۰۱۱).

روش‌های انتخاب ویژگی را می‌توان به دودسته تقسیم کرد: انتخاب عددی ویژگی که ویژگی‌ها را به صورت واحد انتخاب می‌کنند؛ انتخاب برداری ویژگی که ویژگی‌ها را بر مبنای همبستگی متقابل بین ویژگی‌ها برمی‌گزینند. انتخاب عددی ویژگی دارای مزیت سادگی محاسباتی است و احتمالاً برای مجموعه داده‌ای که در آن ویژگی‌ها دارای همبستگی متقابل هستند، مؤثر نیست. روش‌های انتخاب برداری ویژگی بهترین ترکیب برداری ویژگی را انتخاب می‌کند. روش‌های انتخاب برداری ویژگی به دودسته روش‌های دسته‌بندی^{۳۸} و روش‌های مبتنی بر فیلتر^{۳۹} تقسیم می‌شوند. روش‌های دسته‌بندی با استفاده از روش‌های یادگیری ماشینی، مانند جعبه سیاه، ویژگی‌های مرتبط‌تر را انتخاب می‌کنند؛ به گونه‌ای که روش یادگیری دارای عملکرد بهینه است. راهبردهای جستجو دربرگیرنده جستجوی فراگیر، جستجوی پرتو، شاخه و حد، الگوریتم ژنتیک، روش‌های جستجوی حریم‌نا و موارد دیگر هستند. هنگام استفاده از روش دسته‌بندی، ویژگی‌های انتخاب شده مستعد بیش‌برازش داده‌ها هستند. در روش انتخاب

ویژگی مبتنی بر فیلتر، یک ویژگی با طبقه‌ای از ویژگی‌ها و زیرمجموعه ویژگی متناظر آن همبستگی دارد (دوا و دو، ۲۰۱۱).

در این پژوهش به منظور انتخاب متغیرهایی که بیشترین تأثیر را در کشف تقلب در صورت‌های مالی دارند، از روش انتخاب ویژگی مبتنی بر همبستگی که هال (هال، ۱۹۹۹) ارائه کرده است، استفاده خواهد شد. استفاده از معیار همبستگی به راه‌حل بهینه در انتخاب ویژگی منجر می‌شود؛ بنابراین این روش بر دو مسئله تمرکز دارد: معیار اندازه‌گیری همبستگی و الگوریتم انتخاب ویژگی. معیار همبستگی می‌تواند ضریب همبستگی پیرسون^{۴۰} (PCC)، اطلاعات متقابل^{۴۱} (MI) و سایر معیارهای مرتبط باشد. ضریب همبستگی پیرسون معیاری برای وابستگی خطی بین متغیرها و ویژگی‌ها است و با متغیرهای پیوسته و دوتایی قابلیت کاربرد دارد. اطلاعات متقابل توانایی اندازه‌گیری وابستگی غیرخطی را دارد که بی‌ربطی هریک از متغیرها و ویژگی‌ها را با استفاده از واگرایی کولبک-لیبلر^{۴۲} اندازه‌گیری می‌کند؛ با این حال، برآورد اطلاعات متقابل نسبت به ضریب همبستگی پیرسون، به ویژه برای داده‌های پیوسته، سخت‌تر است (دوا و دو، ۲۰۱۱).

۴-۳-۴. طبقه‌بندی

طبقه‌بندی شکلی از تحلیل داده است که به استخراج مدل‌های توضیح‌دهنده طبقات مهم داده می‌پردازد. چنین مدل‌هایی طبقه‌بندی کننده نامیده می‌شوند و برچسب طبقات رسته‌ای (گسسته، نامرتب) را پیش‌بینی می‌کنند؛ برای مثال، می‌توان از مدل‌های طبقه‌بندی برای دسته‌بندی درخواست‌های وام بانکی به بی‌خطر و پرخطر استفاده کرد (هان و همکاران، ۲۰۱۲). در این پژوهش برای دسته‌بندی شرکت‌ها به متقلب و غیر متقلب از روش‌های طبقه‌بندی درخت تصمیم، ماشین بردار پشتیبان و بوستینگ استفاده خواهد شد که در ادامه به شرح مختصر آن‌ها می‌پردازیم.

درخت‌های تصمیم: شیوه‌های درخت تصمیم مجموعه‌ای از قواعد مبتنی بر درخت برای ساخت درخت تصمیم (درخت طبقه‌بندی) ایجاد می‌کنند. درخت‌های تصمیم، داده‌ها را به گروه‌های طبقه‌بندی از پیش تعریف‌شده تسهیم می‌کنند. در پیش‌بینی درماندگی مالی شرکت‌ها، درخت تصمیم، هر شرکت را به دودسته سالم یا درمانده تقسیم می‌کند. به‌طور کلی، درخت‌های تصمیم درخت‌های دودویی هستند که از گره ریشه^{۴۳}، گره‌های غیر برگ^{۴۴} و

گره‌های برگ^{۴۵} که به وسیله شاخه‌ها^{۴۶} به هم متصل شده‌اند، تشکیل می‌گردند و هر گره غیر برگ دارای دوشاخه است که به دو گره متمایز منتهی می‌شود.

زمانی که درخت‌های تصمیم در رابطه با مسائل طبقه‌بندی مانند پیش‌بینی در ماندگی مالی، به کار می‌روند، گره‌های برگ نشان‌دهنده گروه‌های طبقه‌بندی (درمانده یا سالم) و هر گره غیر برگ شامل یک قاعده (تصمیم) تفکیک کننده است؛ بنابراین درخت تصمیم از طریق فرآیندی بازگشتی^{۴۷} برای تفکیک داده‌ها در حرکت از بالا به پایین درخت ایجاد می‌شود. قواعد تفکیک کننده، دربرگیرنده عبارتی (معمولاً یک نسبت مالی) است که برای هر مورد (شرکت) ارزیابی و با یک مقدار برش^{۴۸} مقایسه می‌شود (چپ و همکاران، ۲۰۱۰).

ماشین‌های بردار پشتیبان: با معرفی ماشین‌های بردار پشتیبان از تئوری یادگیری آماری^{۴۹} توسط وپنیک^{۵۰}، مطالعات تئوریک و کاربردی بسیاری در رابطه با آن انجام شده است. ماشین‌های بردار پشتیبان در مقایسه با دیگر شیوه‌های یادگیری، عملکرد بهتری در شناسایی الگو^{۵۱}، برآورد رگرسیون^{۵۲} و سایر زمینه‌ها از قبیل پیش‌بینی سری‌های زمانی مالی^{۵۳}، بازاریابی^{۵۴}، تخمین بازده تولید^{۵۵}، طبقه‌بندی متن^{۵۶}، تشخیص چهره^{۵۷} با تصویر، شناسایی حروف دست‌نوشته^{۵۸} و تشخیص پزشکی^{۵۹} دارند. در ادامه شرح مختصری از کاربرد ماشین‌های بردار پشتیبان برای شناسایی الگو در زمینه طبقه‌بندی ارائه می‌شود.

ماشین بردار پشتیبان از طریق بی‌نهایت انطباق غیرخطی بردارهای ورودی به فضای ویژگی با ابعاد بالا یک طبقه‌بندی کننده دودویی^{۶۰}، ابر صفحه‌های تفکیک کننده بهینه^{۶۱}، تولید می‌کند. ماشین بردار پشتیبان از طریق مرزهای طبقه غیرخطی بر مبنای بردارهای پشتیبان، مدلی خطی برای تخمین تابع تصمیم^{۶۲} ایجاد می‌کند. چنانچه داده‌ها به صورت خطی تفکیک شده باشند، ماشین بردار پشتیبان با آموزش ماشین‌های خطی، به دنبال یافتن صفحه بهینه‌ای است که داده‌ها را بدون خطا و با حداکثر فاصله بین ابر صفحه و نزدیک‌ترین نقاط آموزشی تفکیک می‌کند. نزدیک‌ترین نقاط به ابر صفحه تفکیک کننده بهینه، بردارهای پشتیبان نامیده می‌شوند. سایر مثال‌های آموزشی^{۶۳} برای تعیین مرزهای کلاس دودویی نامربوط هستند. در موارد کلی که داده‌ها به صورت خطی تفکیک نشده‌اند، ماشین بردار پشتیبان برای یافتن صفحه‌ای که تعداد خطاها را برای مجموعه آموزشی^{۶۴}، کمینه می‌نماید، از ماشین‌های غیرخطی استفاده می‌کند (شین و همکاران، ۲۰۰۵).

الگوریتم بوستینگ: همان طور که در بخش مقدمه اشاره شد، سیستم طبقه‌بندی، مدلی

ارائه می‌کند که با فرض وجود یک مجموعه داده مشخص، قادر خواهد بود طبقه مشاهده جدید را پیش‌بینی کند. صحت^{۶۵} طبقه‌بندی کننده به کیفیت روش استفاده‌شده و سختی کاربرد مشخص بستگی دارد. چنانچه طبقه‌بندی کننده استفاده‌شده در مقایسه با قاعده پیش‌فرض^{۶۶}، به صحت بیشتری برسد، می‌توان نتیجه گرفت روش طبقه‌بندی، ساختاری در داده‌ها یافته است. بوستینگ روشی برای استفاده حداکثری از طبقه‌بندی کننده، از طریق بهبود صحت آن است؛ بنابراین روش طبقه‌بندی به‌عنوان زیر روالی^{۶۷} برای ایجاد یک طبقه‌بندی کننده دقیق در مجموعه آموزشی استفاده می‌شود.

روش بوستینگ سیستم طبقه‌بندی را به‌طور مداوم بر روی داده‌های آموزشی به کار می‌برد؛ اما در هر بار کاربرد، توجه یادگیری معطوف به مثال‌های مختلفی از این مجموعه است. زمانی که این فرآیند خاتمه یابد، هر کدام از طبقه‌بندی کننده‌های به‌دست آمده باهم ترکیب می‌شوند تا یک طبقه‌بندی کننده نهایی و بسیار دقیق بر روی مجموعه آموزشی حاصل شود؛ از این رو مطابق با پژوهش‌های پیشین، طبقه‌بندی کننده نهایی، از صحت بالایی در رابطه با داده‌های آزمون^{۶۸} برخوردار است (کرتس و همکاران، ۲۰۰۷).

۵-۳-۴. عملکرد طبقه‌بندی کننده^{۶۹}

در این پژوهش به منظور ارزیابی عملکرد هر یک از الگوریتم‌های پیشنهادی مطابق با ماتریس درهم‌ریختگی^{۷۰}، حسب مورد از معیارهای صحت کلی^{۷۱}، دقت^{۷۲}، فراخوانی^{۷۳}، F -Measure و AUC استفاده خواهد شد.

۴-۴. جامعه آماری، روش نمونه‌گیری و محدوده زمانی

جامعه آماری این تحقیق، کلیه شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران است. در انتخاب نمونه، موارد زیر در نظر گرفته خواهد شد:

پایان سال مالی شرکت‌ها ۲۹ اسفند باشد؛ شرکت‌ها در طی دوره موردنظر، سال مالی خود را تغییر نداده باشند؛ صورت‌های مالی و یادداشت‌های همراه شرکت‌ها در دوره زمانی ۱۳۷۹ الی ۱۳۹۳ به گونه کامل در سایت بورس اوراق بهادار موجود باشد؛ شرکت انتخابی، جزو شرکت‌های بیمه، بانک‌ها و مؤسسه‌های اعتباری، واسطه‌گری‌های مالی و شرکت‌های چند رشته‌ای صنعتی نباشد.

با توجه به محدودیت‌های فوق، تمام این شرکت‌ها بررسی خواهد شد و هیچ‌گونه نمونه‌گیری انجام نمی‌شود. دوره زمانی مورد مطالعه برای کلیه سؤال‌های پژوهش، سال‌های ۱۳۸۱ تا ۱۳۹۳ است.

۴-۵. روش گردآوری داده‌ها

در این پژوهش برای جمع‌آوری داده‌ها و اطلاعات، از روش آرشیوی استفاده می‌شود. برای نگارش و جمع‌آوری اطلاعات مورد نیاز بخش مبانی نظری، از مجلات تخصصی لاتین و برای گردآوری سایر داده‌ها و اطلاعات مورد نیاز، عمدتاً از بانک‌های اطلاعاتی سازمان بورس اوراق بهادار تهران، گزارش‌های روزانه و هفتگی سازمان بورس و نرم‌افزارهای رهاورد نوین، صحرا و تدبیر پرداز استفاده شده است.

۴-۶. یافته‌های پژوهش

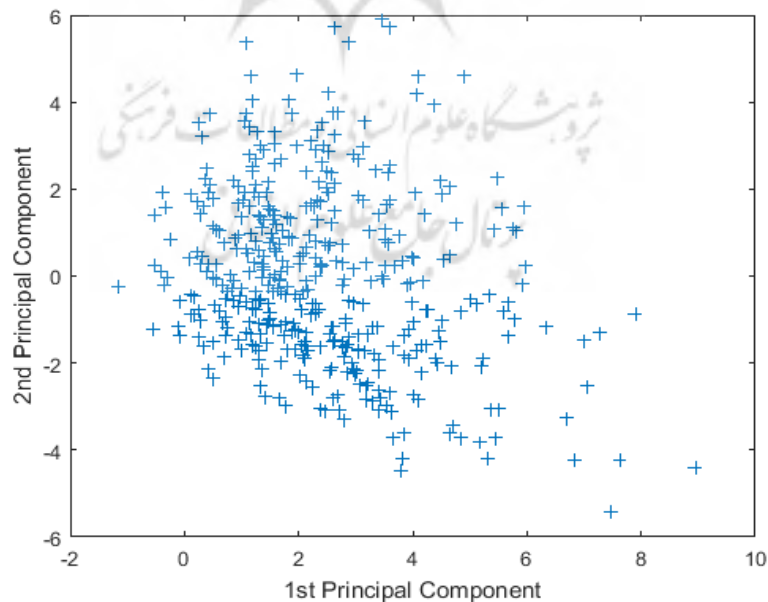
همان‌طور که پیش‌ازین اشاره شد، متغیر پاسخ در این پژوهش تقلب در صورت‌های مالی است. برای اندازه‌گیری آن مطابق با استاندارد حسابرسی شماره ۲۴۰ با عنوان «مسئولیت حسابرسان در ارتباط با تقلب و اشتباه در صورت‌های مالی» از علائم نشان‌دهنده احتمال وجود تحریف‌های ناشی از گزارشگری مالی متقلبان به شرح زیر استفاده می‌شود:

- بیش‌نمایی موجودی کالا؛
- بیش‌نمایی حساب‌ها و اسناد دریافتنی؛
- بیش‌نمایی دارایی‌های ثابت؛
- بیش‌نمایی سرمایه‌گذاری‌ها؛
- کسری ذخیره مطالبات مشکوک‌الوصول؛
- کسری استهلاک؛
- بیش‌نمایی درآمدها؛
- بیش‌نمایی سود؛
- بیش‌نمایی سود انباشته؛
- کم‌نمایی حساب‌ها و اسناد پرداختنی؛
- کسری ذخیره مالیات؛

بدهی‌های احتمالی؛
 کسری ذخیره مزایای پایان خدمت کارکنان؛
 کم‌نمایی هزینه‌ها؛
 حساب‌ها و اسناد دریافتی که مدت زیادی از سررسید آن‌ها گذشته است؛
 موجودی راکد؛
 دارایی راکد؛
 مشکلات تداوم فعالیت؛

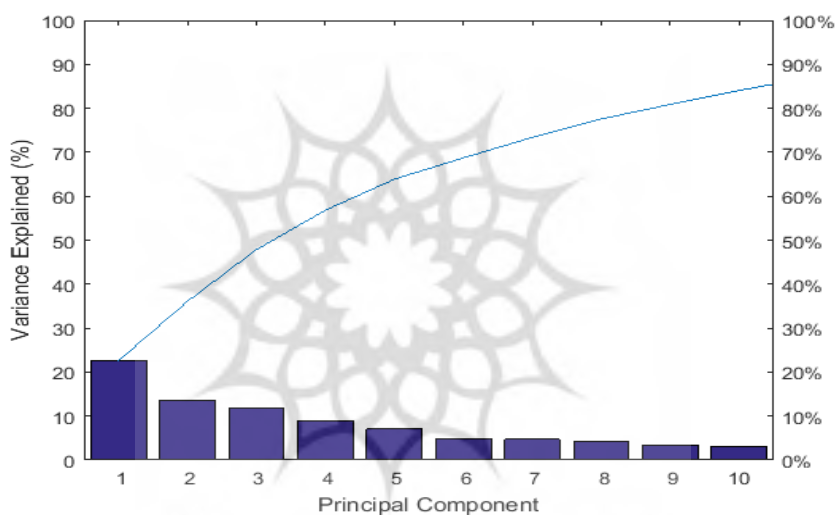
اشتباه در به‌کارگیری استانداردهای حسابداری مرتبط با اندازه‌گیری، شناسایی، طبقه‌بندی، ارائه یا افشا.

چنانچه هر یک از این علائم در بندهای تعدیلی گزارش‌های حسابرسی سالانه شرکت‌ها وجود داشته باشد، مقدار آن برابر با یک و در غیر این صورت صفر خواهد بود؛ بنابراین این مجموعه داده شامل ۱۸۰ شرکت و ۱۹ ویژگی طی سال‌های ۱۳۹۳-۱۳۸۱ خواهد بود. در ادامه از روش تحلیل مؤلفه‌های اساسی (PCA) برای کاهش ویژگی‌های اولیه و نگاشت مجموعه داده‌های اولیه در فضای کوچک‌تر، استفاده خواهد شد. نمودار شماره ۱ نگاشت حاصل از داده‌های نرمال شده بر روی دو مؤلفه اساسی اول را نشان می‌دهد.



نمودار ۱: نگاشت داده‌های نرمال روی دو مؤلفه اساسی اول

نمودار صخره‌ای شماره ۲ واریانس توضیح داده شده توسط ده مؤلفه اول را نشان می‌دهد. تنها تفاوت آشکار در واریانس احتساب شده توسط هر مؤلفه، بین مؤلفه‌های اول و دوم است؛ باین حال مؤلفه اول تنها حدود ۲۳ درصد از کل واریانس را توضیح می‌دهد و در نتیجه مؤلفه‌های بیشتری را باید در نظر گرفت. همان طور که در نمودار مشاهده می‌شود، ده مؤلفه اول (به جای کل ۱۹ مؤلفه) حدود هشتادوپنج درصد از کل واریانس مجموعه داده‌های نرمال شده را توضیح می‌دهند و این دلیلی منطقی برای کاهش ابعاد داده‌های اولیه است.



نمودار ۲: واریانس توضیح داده شده ده مؤلفه اساسی اول

هر نوزده متغیر در نمودار دوتایی شماره ۳ از طریق یک بردار نشان داده شده است. در این نمودار علائم و نوشته‌ها بدین قرارند:

InvOE، بیش‌نمایی موجودی کالا؛

RecOE، بیش‌نمایی حساب‌ها و اسناد دریافتی؛

AssOE، بیش‌نمایی دارایی‌های ثابت؛

InvestOE، بیش‌نمایی سرمایه‌گذاری‌ها؛

BDR، کسری ذخیره مطالبات مشکوک‌الوصول؛

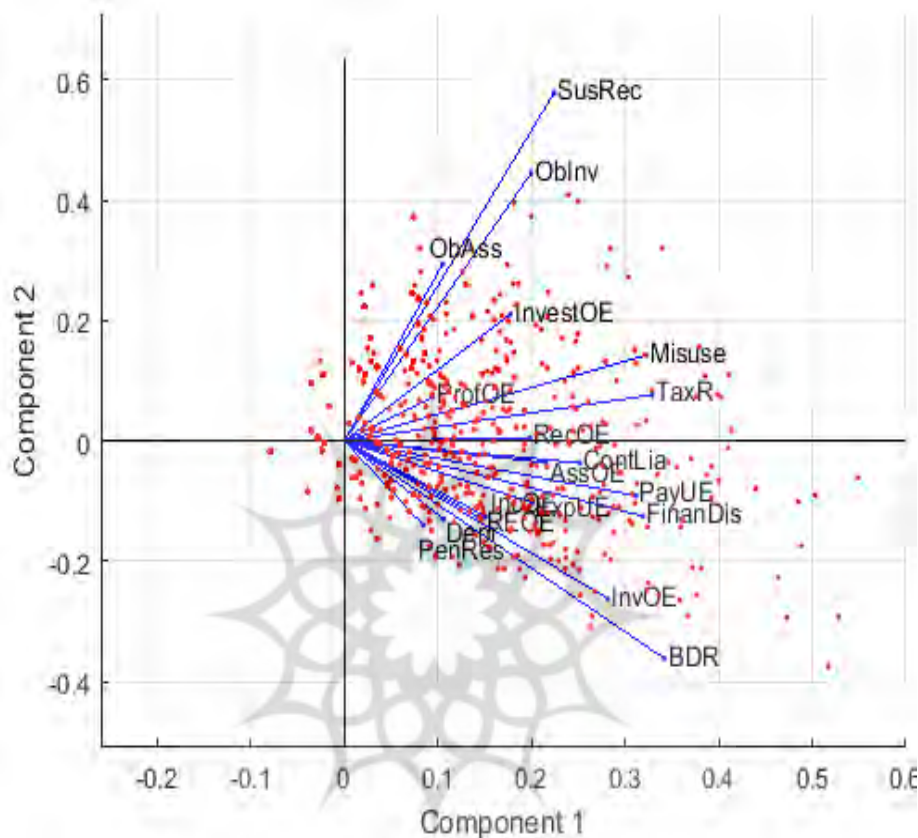
Depr، کسری استهلاک؛

IncOE، بیش‌نمایی درآمدها؛

ProfOE، بیش‌نمایی سود؛
 REOE، بیش‌نمایی سود انباشته؛
 PayUE، کم‌نمایی حساب‌ها و اسناد پرداختی؛
 TaxR، کسری ذخیره مالیات؛
 ContLia، بدهی‌های احتمالی؛
 PenRes، کسری ذخیره مزایای پایان خدمت کارکنان؛
 ExpUE، کم‌نمایی هزینه‌ها؛
 SusRec، حساب‌ها و اسناد دریافتی که مدت زیادی از سررسید آن‌ها گذشته است؛
 ObInv، موجودی راکد؛
 ObAss، دارایی راکد؛
 FinanDis، مشکلات تداوم فعالیت؛
 Misuse، اشتباه در به‌کارگیری استانداردهای حسابداری مرتبط با اندازه‌گیری، شناسایی، طبقه‌بندی، ارائه یا افشا.

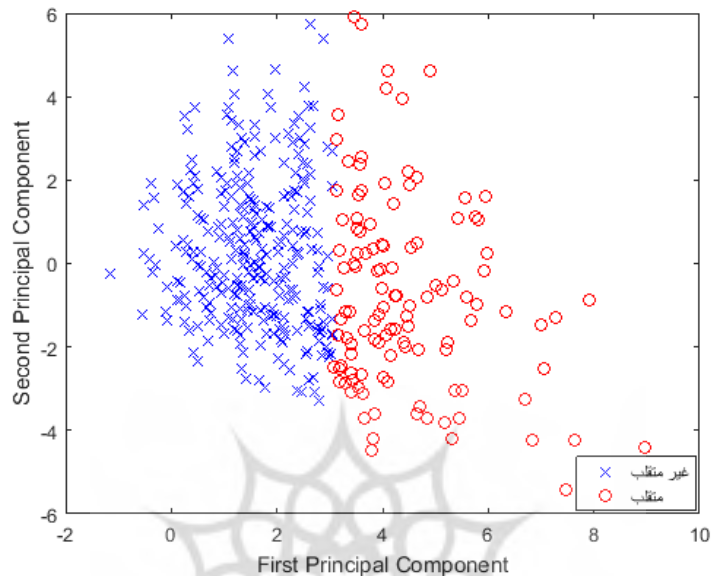
جهت و طول به‌کارگیری، بیان‌گر تأثیری است که هر متغیر بر دو مؤلفه اساسی نمودار می‌گذارد؛ برای مثال، مؤلفه اساسی اول، محور افقی، به‌ازای تمامی ۱۹ متغیر دارای ضریب مثبت است؛ به همین دلیل، هر ۱۹ بردار نیز در نیمه سمت راست نمودار قرار گرفته‌اند. بیشترین ضرایب در مؤلفه اساسی اول متعلق به متغیرهای کسری ذخیره مطالبات مشکوک‌الوصول، کسری ذخیره مالیات بر درآمد و اشتباه در به‌کارگیری استانداردهای حسابداری مرتبط با اندازه‌گیری، شناسایی، طبقه‌بندی، ارائه یا افشا است.

مؤلفه اساسی دوم، محور عمودی، برای هفت متغیر حساب‌ها و اسناد دریافتی که مدت زیادی از سررسید آن‌ها گذشته است، موجودی راکد، دارایی راکد، بیش‌نمایی سرمایه‌گذاری‌ها، بیش‌نمایی سود، اشتباه در به‌کارگیری استانداردهای حسابداری مرتبط با اندازه‌گیری، شناسایی، طبقه‌بندی، ارائه یا افشا و کسری ذخیره مالیات دارای ضرایب مثبت و به‌ازای سایر متغیرها داری ضرایب منفی است؛ بنابراین مؤلفه اساسی دوم بین شرکت‌هایی که هفت متغیر اول در آن‌ها مقادیر بالایی دارد و سایر متغیرها مقادیر پایین دارند، با شرکت‌هایی که عکس این حالت هستند، تمایز می‌گذارد.



نمودار ۳: نمودار دوتایی متغیرهای پژوهش

پس از آنکه مؤلفه‌های اساسی به شرح بالا انتخاب شد، از روش k -means تعدیل شده با الگوریتم کلونی زنبور عسل مصنوعی (ABC)، برای تفکیک شرکت‌های نمونه به متقلب و غیر متقلب استفاده خواهد شد. نمودار شماره ۴، نتایج حاصل از تفکیک شرکت‌های متقلب و غیر متقلب را بر روی مؤلفه‌های اساسی اول و دوم به تصویر می‌کشد. به‌کارگیری روش k -means تعدیل شده، نشان‌دهنده وجود ۱۲۲ شرکت متقلب و ۲۲۱۸ شرکت غیر متقلب (در مجموع ۲۳۴۰ شرکت - سال) طی سال‌های ۱۳۹۳-۱۳۸۱ است.



نمودار ۴: تفکیک شرکت‌ها به متقلب و غیر متقلب

رویکرد ارائه‌شده در بالا با نمونه‌ای متشکل از ۷ شرکت متقلب و ۱۰ شرکت غیر متقلب آزمون شد. شرکت‌های متقلب با مراجعه به نشریات ویژه سازمان بورس و اوراق بهادار که در وبسایت www.seo.ir نمایه شده‌اند، انتخاب گردیدند. بررسی کامل نشریات ویژه نشان می‌دهد نشریات شماره ۳، ۴ و ۷ دربرگیرنده چهارده رأی قطعی صادره در محاکم دادگستری هستند. هر یک از این چهارده رأی در راستای اجرای ماده ۵۲ قانون بازار اوراق بهادار مصوب آذرماه ۱۳۸۴، به دلایل مختلفی صادرشده‌اند، از قبیل ایجاد ظاهری گمراه‌کننده از روند معاملات اوراق بهادار، مبادرت به معاملات اوراق بهادار با استفاده از اطلاعات نهانی، خودداری از ارائه اطلاعات و اسناد و مدارک مهم به سازمان بورس و اوراق بهادار، تخلف از مقررات قانون بازار در تهیه اسناد و مدارک، تقسیم منافع موهوم به استناد صورت‌داری و ترازنامه مزور و موارد دیگر.

موضوع پژوهش حاضر به‌طور مشخص تقلب در صورت‌های مالی است؛ پس فقط شرکت‌هایی به‌عنوان شرکت‌های متقلب در نظر گرفته شدند که با تعریف زیر سازگار بودند: شرکت‌هایی که فقط به دلایل مرتبط با تقلب در صورت‌های مالی که تعریف آن در بخش مقدمه به‌طور کامل ارائه شد، علیه مدیران آن‌ها در محاکم دادگستری آرای قطعی صادرشده است و با بررسی اطلاعات ارائه‌شده در نشریات ویژه سازمان بورس و اوراق بهادار قابل تشخیص

بودند.

با اعمال این رویکرد، تنها ۷ شرکت از ۱۴ شرکت مزبور به‌عنوان شرکت‌هایی که صورت‌های مالی متقلبانه داشته‌اند، انتخاب گردیدند؛ سپس علائم خطر تقلب هر یک از این شرکت‌ها با بررسی گزارش حسابرسی آن‌ها و مطابق با رویه اشاره شده در بالا تعیین شدند. علاوه بر این ۷ شرکت، ۱۰ شرکت غیر متقلب نیز با رعایت دقت لازم به‌صورت تصادفی و به‌گونه‌ای انتخاب شدند که هیچ‌کدام از علائم خطر تقلب در گزارش حسابرسی آن‌ها وجود نداشته باشد. توضیح آنکه هیچ‌یک از این ۱۷ شرکت در نمونه اولیه، برای تفکیک شرکت‌های متقلب و غیر متقلب قرار نداشته‌اند. در مورد شرکت‌های متقلب از گزارش حسابرسی یک سال قبل از وقوع تقلب که در اینجا سال مالی قبل از تاریخ اعلام جرم در محاکم دادگستری است، برای تعیین علائم خطر استفاده شد. نتایج حاصل از به‌کارگیری الگوریتم *k-means* تعدیل‌شده بر روی این نمونه آزمون، حاکی از این است که الگوریتم پیشنهادی، شش شرکت از هفت شرکت متقلب را به‌درستی در خوشه متقلب قرار می‌دهد و از ده شرکت غیر متقلب نیز همگی در خوشه غیر متقلب قرار می‌گیرند. بررسی‌های بیشتر در رابطه با تنها شرکت متقلبی که به‌اشتباه در خوشه غیر متقلب قرار گرفته است، نیز نشان می‌دهد که با به‌کارگیری علائم خطر تقلب ارائه‌شده در گزارش حسابرسی دو سال قبل شرکت مذکور، این شرکت در زمره شرکت‌های متقلب جای می‌گیرد و به احتمال زیاد اعلام‌جرم با فاصله بیش از یک سال از وقوع تقلب در صورت‌های مالی صورت گرفته است. نگاره شماره ۲ نتایج حاصل از آزمون رویکرد *k-means* تعدیل‌شده بر روی داده‌های آزمون را نشان می‌دهد.

نگاره ۲: نتایج حاصل از آزمون روش *k-means* تعدیل‌شده

معیار عملکرد	درصد
صحت کلی	۹۴/۱۱٪
دقت	۱۰۰٪
فراخوانی	۸۵/۷۱٪
<i>F – Measure</i>	۹۲/۳۱٪
<i>G – mean</i>	۹۲/۵۸٪

نتایج پژوهش‌های پیشین حاکی از این است که میانگین دوره زمانی وقوع تقلب قبل از کشف آن ۱۸ ماه است (راهنمای ممیزان تقلب، ۲۰۱۲). از این‌رو، پس از آنکه شرکت‌های

متقلب به شرح بالا تعیین شد، مطابق با مبانی نظری پژوهش از ۴۰ متغیر مالی و غیرمالی مرتبط با اهرم مالی، سودآوری، ترکیب دارایی‌ها، نقدینگی، کارایی، اندازه، رشد، وضعیت مالی کلی، اندازه حسابرس، دوره تصدی حسابرس و تغییر حسابرس به همراه روش‌های درخت تصمیم، ماشین بردار پشتیبان و الگوریتم بوستینگ برای پیش‌بینی تقلب در صورت‌های مالی به ترتیب برای هر یک از سال‌های یک سال قبل از وقوع تقلب و دو سال قبل از وقوع آن استفاده شده است.

از آنجایی که تعداد شرکت‌های متقلب (۱۲۲ شرکت) در مقایسه با تعداد شرکت‌های غیر متقلب (۲۲۱۸ شرکت) بسیار کمتر است، با یک مجموعه داده نامتوازن سروکار داریم؛ از این رو برای حل این مسئله از رویکردی که هی و همکاران (۲۰۰۸) معرفی کردند، استفاده خواهد شد. در این رویکرد که آدیسین^{۷۴} نام دارد، ایده اصلی به‌کارگیری توزیع موزون برای نمونه‌های طبقات اقلیت، مطابق با سطح دشواری آن‌ها در یادگیری است؛ بنابراین داده‌های ترکیبی بیشتری برای نمونه‌های طبقات اقلیتی که یادگیری آن‌ها سخت‌تر است، در مقایسه با نمونه‌های طبقات اقلیتی که یادگیری آن‌ها آسان‌تر است، ایجاد می‌شود.

به‌کارگیری داده‌های آموزشی برای برآورد عملکرد مدل، ممکن است با سوپره همراه باشد. در بیشتر موارد مدل‌ها به‌جای یادگیری، تمایل به حفظ کردن نمونه دارند (بیش‌برازش^{۷۵} داده‌ها). در این پژوهش برای اجتناب از این مشکل از اعتبارسنجی متقابل با ۱۰ لایه^{۷۶} استفاده شده است؛ بنابراین برای هر لایه، مدل با استفاده از ۹ لایه باقیمانده آموزش داده می‌شود و با لایه جداشده آزمون می‌گردد. در نهایت میانگین عملکرد محاسبه خواهد شد. نگاره شماره ۳ نتایج حاصل از برآورد مدل‌های پیشنهادی برای یک سال قبل از تقلب ($t-1$) را نشان می‌دهد.

نگاره ۳: نتایج حاصل از مدل‌های پیشنهادی یک سال قبل از تقلب ($t-1$)

الگوریتم بوستینگ	ماشین بردار پشتیبان	درخت تصمیم	
٪۶۹/۸	٪۷۴/۸	٪۹۱/۸	صحت کلی
٪۷۰/۵	٪۷۵/۷	٪۹۱/۹	دقت
٪۶۹/۸	٪۷۴/۸	٪۹۱/۸	فراخوانی
٪۶۹/۶	٪۷۴/۷	٪۹۱/۸	<i>F – Measure</i>
۰/۷۷۷	۰/۷۵۱	۰/۹۲۷	<i>AUC</i>

همان‌طور که در نگاره بالا مشاهده می‌شود، نتایج معیارهای عملکرد مدل‌های پیشنهادی یک سال قبل از وقوع تقلب، حاکی از مناسب بودن مدل‌های درخت تصمیم، ماشین بردار پشتیبان و الگوریتم بوستینگ برای پیش‌بینی تقلب است. به‌منظور مقایسه عملکرد مدل‌های پیشنهادی در کشف تقلب از آزمون‌های آماری استفاده خواهد شد؛ بنابراین برای بررسی فرض نرمال بودن توزیع معیارهای عملکرد محاسبه شده از آزمون کولموگوروف-اسمیرنوف (KS) استفاده شد. نتایج حاصل از این آزمون در سطح معناداری ۵ درصد حاکی از رد فرض H_0 مبنی بر نرمال بودن توزیع معیارهای ارزیابی عملکرد است؛ بنابراین از آزمون رتبه علامت دار ویلکاکسون (Wilcoxon) برای مقایسه معیارهای ارزیابی عملکرد استفاده خواهد شد. نگاره شماره ۴ نتایج حاصل از آزمون رتبه علامت دار ویلکاکسون (Wilcoxon) را برای معیار ارزیابی عملکرد $F - Measure$ یک سال قبل از وقوع تقلب، نشان می‌دهد. در حالت کلی تفاوت معناداری بین عملکرد درخت تصمیم با ماشین بردار پشتیبان و الگوریتم بوستینگ وجود دارد؛ از این‌رو الگوریتم درخت تصمیم روش مناسب‌تری برای طبقه‌بندی شرکت‌های متقلب و غیر متقلب است.

نگاره ۴: نتایج حاصل از آزمون رتبه علامت دار ویلکاکسون (Wilcoxon)

طبقه‌بندی کننده‌ها	آماره آزمون	آماره Z	احتمال آماره Z
درخت تصمیم و ماشین بردار پشتیبان	۰/۰۰۰	-۸/۶۸۲	۰/۰۰۰
درخت تصمیم و الگوریتم بوستینگ	۰/۰۰۰	-۸/۶۸۲	۰/۰۰۰
ماشین بردار پشتیبان و الگوریتم بوستینگ	۲۴/۰۰۰	-۸/۵۹۹	۰/۰۰۰

نگاره شماره ۵ نتایج حاصل از برآورد مدل‌های پیشنهادی برای دو سال قبل از تقلب ($t-2$) را نشان می‌دهد.

نگاره ۵: نتایج حاصل از مدل‌های پیشنهادی دو سال قبل از تقلب ($t-2$)

صحت کلی	درخت تصمیم	ماشین بردار پشتیبان	الگوریتم بوستینگ
٪۹۰/۴	٪۷۲/۹	٪۷۰/۴	
دقت	٪۹۰/۶	٪۷۴/۳	٪۷۰/۸

الگوریتم بوستینگ	ماشین بردار پشتیبان	درخت تصمیم	
۰/۷۰/۴	۰/۷۲/۹	۰/۹۰/۳	فراخوانی
۰/۷۰/۴	۰/۷۲/۷	۰/۹۰/۴	<i>F-Measure</i>
۰/۷۸۳	۰/۷۳۲	۰/۹۱۵	<i>AUC</i>

چنانکه در نگاره شماره ۵ مشاهده می‌شود، عملکرد مدل‌های پیشنهادی در دو سال قبل از وقوع تقلب نیز با یکدیگر متفاوت است. بررسی مقایسه‌ای عملکرد مدل‌های پیشنهادی حاکی از بهتر بودن معیارهای عملکرد درخت تصمیم نسبت به ماشین بردار پشتیبان و الگوریتم بوستینگ، برای پیش‌بینی تقلب است. نگاره شماره ۶ نتایج حاصل از آزمون رتبه‌ علامت‌دار ویلکاکسون (Wilcoxon) برای معیار ارزیابی عملکرد *F-Measure* را برای دو سال قبل از وقوع تقلب، نشان می‌دهد؛ از این‌رو، نتایج حاصل از آزمون رتبه‌ علامت‌دار ویلکاکسون (Wilcoxon) نیز حاکی از برتری درخت تصمیم در مقایسه با ماشین بردار پشتیبان و الگوریتم بوستینگ، برای کشف تقلب در صورت‌های مالی است.

نگاره ۶: نتایج حاصل از آزمون رتبه‌ علامت‌دار ویلکاکسون (Wilcoxon)

طبقه‌بندی کننده‌ها	آماره آزمون	آماره Z	احتمال آماره Z
درخت تصمیم و ماشین بردار پشتیبان	۰/۰۰۰	-۸/۶۸۲	۰/۰۰۰
درخت تصمیم و الگوریتم بوستینگ	۰/۰۰۰	-۸/۶۸۲	۰/۰۰۰
ماشین بردار پشتیبان و الگوریتم بوستینگ	۳۰۵۱/۰۰۰	۱/۸۰۹	۰/۰۷۱

همچنین در این پژوهش از آزمون من ویتنی (U آزمون) برای مقایسه معیارهای ارزیابی عملکرد یک سال قبل از وقوع تقلب ($t-1$) با دو سال قبل از وقوع آن ($t-2$)، استفاده خواهد شد. نگاره شماره ۷ نتایج حاصل از این آزمون را برای هر یک از مدل‌های پیشنهادی نشان می‌دهد.

نگاره ۷: نتایج حاصل از آزمون من‌ویتنی (آزمون U)

طبقه‌بندی کننده‌ها	آماره آزمون	آماره Z	احتمال آماره Z
درخت تصمیم	۳۵۲۰/۰۰۰	-۳/۶۱۶	۰/۰۰۰
ماشین بردار پشتیبان	۲۰۹۶/۰۰۰	-۷/۰۹۶	۰/۰۰۰
الگوریتم بوستینگ	۷۶۸۱/۰۰۰	۶/۵۵۱	۰/۰۰۰

همان‌طور که در نگاره شماره ۷ مشاهده می‌شود، تفاوت معناداری بین عملکرد مدل‌های پیشنهادی در یک سال قبل از وقوع تقلب ($t-1$) و دو سال قبل از وقوع آن ($t-2$) وجود دارد؛ از این‌رو می‌توان نتیجه گرفت استفاده از داده‌های یک سال قبل، برای پیش‌بینی تقلب در صورت‌های مالی عملکرد بهتری خواهد داشت؛ همچنین نتایج حاصل از انتخاب ویژگی به روش مبتنی بر همبستگی، حاکی از سودمندی متغیرهای کل بدهی‌ها به حقوق صاحبان سهام، سود ناخالص به کل دارایی‌ها، حساب‌های دریافتی به کل دارایی‌ها، موجودی کالا به جمع دارایی‌ها، نسبت آبی، نسبت وجه نقد به کل دارایی‌ها، لگاریتم طبیعی کل دارایی‌ها، اندازه حسابرس و دوره تصدی حسابرس برای کشف تقلب داشت.

۵. نتیجه‌گیری و پیشنهادها

رسوایی‌ها و شکست‌های شرکتی، اطمینان سرمایه‌گذاران نسبت به درست و منصفانه بودن حساب‌های واحدهای تجاری را مخدوش کرده است. در بیشتر موارد گزارش شده، مدیریت از طریق دست‌کاری صورت‌های مالی شرکت برای انعکاس نتایج دلخواه خود به عموم، به کلاه‌برداری و پنهان ساختن آن مبادرت کرده است؛ از این‌رو در محیطی فعال از تقلب در صورت‌های مالی، سازوکارهای کشف تقلب با کمک رایانه بسیار مؤثرتر و کاراتر خواهد بود. تکنولوژی‌های مبتنی بر آمار و یادگیری ماشینی راهکاری اثربخش برای پیشگیری و کشف تقلب هستند؛ از این‌رو در این مقاله نخست به بررسی این مسئله پرداخته شد که آیا شیوه‌های داده‌کاوی امکان تفکیک شرکت‌های متقلب و غیرمتقلب را با استفاده از گزارش‌های حسابرسی سالانه شرکت‌ها فراهم می‌کنند؟ برای پاسخ‌گویی به این سؤال از ۱۹ علائم خطر اشاره‌شده در استاندارد حسابرسی ۲۴۰، به همراه شیوه‌های داده‌کاوی تحلیل مؤلفه‌اساسی و خوشه‌بندی استفاده شد. یافته‌های پژوهش حاکی از این بود که رویکرد ارائه‌شده امکان تفکیک شرکت‌های

نمونه به دودسته متقلب و غیر متقلب را فراهم می‌کند؛ سپس این سؤال مطرح شد که آیا می‌توان با استفاده از اطلاعات مالی و غیرمالی گزارش‌های سالانه شرکت‌ها، مدلی کمی برای کشف تقلب ارائه داد که بیان‌گر یک روش تحلیلی خودکار برای کشف تقلب بالقوه باشد؟ برای پاسخ‌گویی به این سؤال نیز مطابق با مبانی نظری و پیشینه پژوهش از ۴۰ متغیر مالی و غیرمالی به همراه شیوه‌های درخت تصمیم، ماشین بردار پشتیبان و روش بوستینگ استفاده شد. یافته‌های پژوهش بیان‌گر وجود شواهدی دال بر عملکرد مناسب مدل‌های پیشنهادی، برای پیش‌بینی تقلب در صورت‌های مالی است. نتایج حاصل از آزمون رتبه علامت‌دار ویلکاکسون (Wilcoxon) نیز حاکی از برتری درخت تصمیم در مقایسه با ماشین بردار پشتیبان و الگوریتم بوستینگ، برای کشف تقلب در صورت‌های مالی بود؛ همچنین نتایج آزمون من‌ویتنی (آزمون U) نشان می‌دهد داده‌های یک سال قبل برای پیش‌بینی تقلب در صورت‌های مالی مناسب‌تر از داده‌های دو سال قبل است. به‌طور کلی نتایج این پژوهش با نتایج پژوهش‌های پرسونس (۱۹۹۵)، فروز و همکاران (۲۰۰۰)، سپاتیس (۲۰۰۲)، کامینسکی و همکاران (۲۰۰۴)، کرکوز و همکاران (۲۰۰۷)، آلدن و همکاران (۲۰۱۲)، هوانگ و همکاران (۲۰۱۴)، چن و همکاران (۲۰۱۴) و لین و همکاران (۲۰۱۶) مبنی بر سودمندی شیوه‌های داده‌کاوی برای پیش‌بینی تقلب در صورت‌های مالی مطابقت دارد.

نتایج این پژوهش برای حسابرسان مستقل و داخلی، نهادهای مالیاتی و سایر دستگاه‌های دولتی، سرمایه‌گذاران و تحلیل‌گران مالی، مؤسسات رتبه‌بندی اعتباری، بانک‌ها و سایر اعتباردهندگان قابل‌استفاده است. امروزه رویه‌های حسابرسی باید به موارد فزاینده‌ای از تقلب مدیران رسیدگی کنند. قابلیت پیشرفته طبقه‌بندی و پیش‌بینی شیوه‌های داده‌کاوی، به حسابرسان مستقل و داخلی در کشف گزارشگری مالی متقلبانه کمک می‌کند.

با توجه به دلیل ماهیت متفاوت فعالیت شرکت‌های بیمه، بانک‌ها و مؤسسه‌های اعتباری و واسطه‌گری‌های مالی، این‌گونه شرکت‌ها از جامعه مورد مطالعه حذف شده‌اند؛ بنابراین نتایج پژوهش قابل‌تعمیم به تمامی شرکت‌ها نیست؛ پس بررسی علائم خطر مرتبط با تقلب در صورت‌های مالی و همچنین ارائه مدلی برای کشف صورت‌های مالی متقلبانه این شرکت‌ها و مؤسسات، می‌تواند موضوع پژوهشی بالقوه باشد؛ همچنین به پژوهش‌گران آینده پیشنهاد می‌شود با استفاده از سایر شیوه‌های داده‌کاوی برای خوشه‌بندی، انتخاب ویژگی و طبقه‌بندی، به بررسی و تأیید الگو و چارچوب پیشنهادی اقدام کنند. در انجام این پژوهش محدودیت‌هایی

به شرح زیر وجود داشت:

ارائه مجدد اکثر صورت‌های مالی در سال بعد که خود نیازمند پژوهشی در مورد علل آن و همچنین معنی‌دار بودن یا نبودن تفاوت بین اطلاعات تجدید ارائه‌شده و تجدید ارائه نشده است.

دسترسی نداشتن به داده‌ها و اطلاعات مالی شرکت‌ها در بلندمدت، سبب شد تا نتایج پژوهش تنها محدود به دوره زمانی کوتاهی باشد؛ بنابراین نتایج آن قابل تعمیم به دوره‌های زمانی بلندمدت نخواهد بود.

یادداشت‌ها

- | | |
|--|---|
| 1- Financial Statements Fraud | 2- Association of Certified Fraud Examiners |
| 3- Data Mining | 4- Stepwise-Logistic Models |
| 5- Artificial Neural Networks | 6- Discriminant Analysis |
| 7- Decision Trees | 8- Bayesian Belief Networks |
| 9- Fuzzy Rule-Based Classifiers | 10- Markovian Learning Estimation of Distribution Algorithm |
| 11- Growing Hierarchical Self-Organizing Map | 12- Integrated Rough Set Analysis |
| 13- Biomedicine | 14- Cybersecurity |
| 15- Supervised | 16- Unsupervised |
| 17- Clustering | 18- Associative Rule Mining |
| 19- Data Cleaning | 20- Data Integration |
| 21- Data Selection | 22- Data Transformation |
| 23- Pattern Evaluation | 24- Knowledge Representation |
| 25- Support Vector Machines | 26- Boosting |
| 27- Dimensionality Reduction | 28- Numerosity Reduction |
| 29- Data Compression | 30- Wavelet Transforms |
| 31- Principal Components Analysis | 32- Attribute Subset Selection |
| 33- Bioinformatics | 34- Text Mining |
| 35- Pattern Recognition | 36- Image Analysis |
| 37- Noisy | 38- Wrapper |
| 39- Filter | 40- Pearson's Correlation Coefficient |
| 41- Mutual Information | 42- Kullback-leibler |
| 43- Root Node | 44- Non-Leaf Nodes |
| 45- Leaf Nodes | 46- Branches |
| 47- Recursive Process | 48- Cut-off value |

یادداشت‌ها

- | | |
|---------------------------------------|------------------------------------|
| 49- Statistical Learning Theory | 50- Vapnik |
| 51- Pattern recognition | 52- Regression Estimation |
| 53- Financial Time-Series Forecasting | 54- Marketing |
| 55- Estimating Manufacturing Yields | 56- Text Categorization |
| 57- Face Detection | 58- Hand Written Digit Recognition |
| 59- Medical Diagnosis | 60- Binary Classifier |
| 61- Optimal Separating Hyperplanes | 62- Decision Function |
| 63- Training Examples | 64- Training Set |
| 65- Accuracy | 66- Default Rule |
| 67- Subroutine | 68- Test Set |
| 69- Classifier Performance | 70- Confusion Matrix |
| 71- Overall Accuracy | 72- Precision |
| 73- Recall | 74- ADASYN |
| 75- Overfitting | 76- 10-Fold Cross Validation |

منابع

الف. فارسی

- اعتمادی، حسین و زلفی، حسن (۱۳۹۲). کاربرد رگرسیون لجستیک در شناسایی گزارشگری مالی متقلبانة. *دانش حسابرسی*، ۱۳(۵۱)، ۱۴۵-۱۶۳.
- پیری، پرویز و برزگری صدقیانی، سمانه (۱۳۹۵). بررسی رابطه بین دوره تصدی حسابرس و تغییر حسابرس با کیفیت سود شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران بر اساس مدل کاسنیک. *پیشرفت‌های حسابداری*، ۸(۲)، ۶۵-۹۱.
- جهانشاد، آریتا و سرداری زاده، سپیده (۱۳۹۳). رابطه معیار مالی (اختلاف رشد درآمد) و معیار غیرمالی (رشد تعداد کارکنان) با گزارشگری مالی متقلبانة. *پژوهش حسابداری*، ۴(۱۳)، ۱۸۱-۱۹۸.
- خواجوی، شکراله و قدیریان آرانی، محمدحسین (۱۳۹۴). بررسی تأثیر کیفیت سود بر تجدید ارائه صورت‌های مالی. *پیشرفت‌های حسابداری*، ۷(۲)، ۵۹-۸۴.
- صفرزاده، محمدحسین (۱۳۸۹). توانایی نسبت‌های مالی در کشف تقلب در گزارشگری مالی: تحلیل لاجیت. *دانش حسابداری*، ۱(۱)، ۱۳۷-۱۶۳.
- فرقاندوست حقیقی، کامبیز؛ هاشمی، سید عباس و فروغی دهکردی، امین (۱۳۹۳). مطالعه رابطه بین مدیریت سود و امکان تقلب در صورت‌های مالی شرکت‌های پذیرفته‌شده در

بورس اوراق بهادار تهران. دانش حسابداری، ۱۴(۵۶)، ۴۷-۶۸.
 کمیته تدوین استانداردهای حسابداری (۱۳۹۴). اصول و ضوابط حسابداری و حسابداری: استانداردهای حسابداری. چاپ بیست و پنجم، تهران، سازمان حسابداری.
 مرادی، جواد؛ رستمی، راحله و زارع، رضا (۱۳۹۳). شناسایی عوامل خطر مؤثر بر احتمال وقوع تقلب در گزارشگری مالی از دید حسابرسان و بررسی تأثیر آن‌ها بر عملکرد مالی شرکت. پیشرفت‌های حسابداری، ۶(۱)، ۱۴۱-۱۷۳.

ب. انگلیسی

- Alden, M. E., Bryan, D. M., Lessley, B., J., & Tripathy, A. (2012). Detection of financial statement fraud using evolutionary algorithms. *Journal of Emerging Technologies in Accounting*, 9, 71-94.
- Altman, E. I., (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance*, 23, 589-609.
- Association of Certified Fraud Examiners (ACFE) (2011). *Fraud Examiners Manual*. Austin, Texas: ACFE.
- Association of Certified Fraud Examiners (ACFE) (2012). *Report to the Nations on Occupational Fraud and Abuse*. Austin, Texas: ACFE.
- Brazel, J. F., Jones, K., L., & Zimbelman, M. F. (2009). Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research*, 47, 1135-1166.
- Carcello, J. V., & Nagy, A. L. (2004). Audit firm tenure and fraudulent financial Reporting. *Auditing: A Journal of Practice & Theory*, 23, 55-69.
- Chen, F. H., Chi, D. J., & Zhu, J. Y. (2014). Application of random forest, rough set theory, decision tree and neural network to detect financial statement fraud° Taking torporate governance into consideration. *Intelligent Computing Theory*, 8588, 221-234.
- Cort'es, E. A., Martnez M. G., & Rubio, N. G. (2007). A boosting approach for corporate failure prediction. *Applied Intelligence*, 27, 29-37.
- Dua, S., & Du, X. (2011). *Data Mining and Machine Learning in Cybersecurity*. 1st edition. Milton Park, UK: Taylor and Francis Group.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874.

- Feroz, E. H., Kwon, T. M., Pastena, V. S., & Park, K. (2002). The efficacy of red flags in predicting the SEC targets: An artificial neural networks approach. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 9, 145° 157.
- Gepp, A., Kumar, K., & Bhattacharya, S. (2010). Business failure prediction using decision trees. *Journal of Forecasting*, 29, 536-555.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. 3rd Edition, Morgan Kaufmann Publications.
- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *IEEE Neural Networks*, 1322-1328.
- Hoogs, B., Kiehl, T., Lacombe, C., & Senturk, D. (2007). A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 15, 41-56.
- Huang, S. H., Tsaih, R. H., & Yu, F. (2014). Topological pattern discovery and feature extraction for fraudulent financial reporting. *Expert Systems with Applications*, 41, 4360° 4372.
- Kaminski, K. A., Wetzel, T. S., & Guan, L. (2004). Can financial ratios detect fraudulent financial reporting? *Managerial Auditing Journal*, 19, 15° 28.
- Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial Bee Colony (ABC) algorithm. *Journal of Global Optimization*, 39, 459-471.
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32, 995° 1003.
- Lin, C-H., Chiu, A-A., Huang, S. Y., & Yen, D. C. (2016). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts judgments. *Knowledge-Based Systems*, 89, 459-470.
- Persons, O. (1995). Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research*, 11, 38° 46.
- Pierre, K. S., & Anderson, J. A. (1984). An analysis of the factors associated with lawsuits against public accountants. *The Accounting Review*, 59, 242-263.

- Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 50, 491-500.
- Rezaee, Z. (2005). Causes, consequences, and deterrence of financial statement fraud. *Critical Perspective on Accounting*, 16, 277-298.
- Rezaee, Z., & Riley, R. (2010). *Financial Statement Fraud-prevention and Detection*. 2nd edition, John Wiley & Sons, Inc.
- Shin, K. S., Lee, T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28, 127-135.
- Spathis, C., Doumpos, M., & Zopounidis, C. (2002). Detecting falsified financial statements: A comparative study using multicriteria analysis and multivariate statistical techniques. *European Accounting Review*, 11(3), 509-535.
- Spathis, C. T. (2002). Detecting false financial statements using published data: Some evidence from Greece. *Managerial Auditing Journal*, 17, 179-191.
- Sreejesh, S., Anusree, M. R., & Mohapatra, S. (2014). *Business Research Methods: An Applied Orientation*. 1st edition, Springer International Publishing.
- Stice, D. J. (1991) Using financial and market information to identify pre-engagement market factors associated with lawsuits against auditors. *Accounting Review*, 66, 516° 33.
- Summers, S. L., & Sweeney, J. T. (1998). Fraudulently misstated financial statements and insider trading: An empirical analysis. *Accounting Review*, 73, 131° 46.
- Whiting, D., G., Hansen, J. V., Mcdonald, J., B., Albrecht, C., & Albrecht, W. S. (2012). Machine learning methods for detecting patterns of management fraud. *Computational Intelligence*, 28, 505-527.
- Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. *Decision Support Systems*, 50, 570-575.
- Zikmund, W. G., Babin, B. J., Car, J. C., & Griffin, M. (2010). *Business Research Methods*. 8th edition, Cengage Learning.