فصلنامه حسابداری مالی/سال نهم/ شماره ۳۳ /بهار ۹۶/صفحات ۵۰- ۲۳

مدل‌سازی متغیرهای اثرگذار برای کشف تقلب در صورت‌های مالی با استفاده از تکنیک‌های داده‌کاوی

شکراله خواجوی *
مهرداد ابراهیمی**

تاریخ دریافت: ۱۳۹۵/۰۷/۱۷ تاریخ پذیرش: ۱۳۹۶/۰۱/۱۹

چکیده

در این پژوهش به بررسی این مسئله پرداخته می‌شود که آیا می‌توان از طریق شناسایی و انتخاب متغیرهای اثرگذار در کشف تقلب در صورت‌های مالی و با به‌کارگیری تکنیک‌های داده‌کاوی مدلی برای کشف تقلب در صورت‌های مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران ارائه کرد؟ برای پاسخگویی به این سؤال از یک نمونه تطبیقی متشکل از ۶ شرکت متقلب و ۶۴ شرکت غیر متقلب (نمونه نامتوازن) به همراه تکنیک‌های طبقه‌بندی شبکه عصبی مصنوعی، شبکه بیزین و الگوریتم جنگل تصادفی طی سال‌های ۱۳۹۳-۱۳۷۹ استفاده گردید. برای حل مشکل نمونه نامتوازن نیز رویکرد معرفی‌شده توسط هی و همکاران (۲۰۰۸) بکار رفت. متغیرهای پیش‌بین، عوامل مالی و غیرمالی خطر تقلب مرتبط با گزارشگری مالی متقلبانه هستند که با مرور پژوهش‌های انجام‌شده در این زمینه تعیین خواهند شد. یافته‌های پژوهش بیان‌گر وجود شواهدی دال بر عملکرد مناسب مدل‌های پیشنهادی و برتری الگوریتم جنگل تصادفی و شبکه بیزین برای پیش‌بینی تقلب در صورت‌های مالی است. نتایج حاصل از انتخاب ویژگی به روش مبتنی بر همبستگی حاکی از سودمندی متغیرهای نسبت پوشش بهره، نسبت حساب‌های دریافتنی به کل دارایی‌ها، نسبت موجودی کالا به فروش خالص، نسبت نقدی، لگاریتم طبیعی فروش، نسبت سود خالص به فروش و نسبت جمع دارایی‌های جاری به کل دارایی‌ها برای کشف تقلب بود.

---

* استاد گروه حسابداری، دانشگاه شیراز، شیراز، ایران (نویسنده مسئول)
Email: shkhajavi@gmail.com

** دانشجوی دکترای حسابداری، دانشگاه شیراز، شیراز، ایران
Email: mehrdadebrahimi66@gmail.com

واژه‌های کلیدی: کشف تقلب، تقلب در صورت‌های مالی، داده‌کاوی، بورس اوراق بهادار تهران.

۱- مقدمه

حسابداری یک فعالیت خدماتی است که هدف آن ارائه اطلاعات مفید مرتبط با واحدهای اقتصادی به تصمیم‌گیرندگان درون‌سازمانی و برون‌سازمانی است. ازاین‌رو، حسابداری باید فعالیت‌های یک واحد اقتصادی را به‌منظور گزارشگری مؤثر به ذینفعان، شناسایی، تجزیه‌وتحلیل و به شکل مناسبی ثبت کند (سگال[1]، ۲۰۱۶). اطلاعات ارائه‌شده درباره وضعیت مالی و عملکرد یک شرکت، اهمیت زیادی برای استفاده‌کنندگان از صورت‌های مالی دارد زیرا مبنایی برای تصمیم‌گیری و تخصیص سرمایه است؛ بنابراین، قابلیت اتکا و شفافیت در گزارشگری مالی به ارائه درست و صادقانه دستاوردهای مالی شرکت منجر خواهد شد و نقش مهمی در پایداری سیستم مالی دارد. بااین‌حال، سقوط شرکت‌های بسیاری (مانند انرون[2]، وردکام[3] و گلوبال کراسینگ[4]) اثربخشی حاکمیت شرکتی، کیفیت گزارش‌های مالی و قابلیت اعتماد کارکردهای حسابرسی را با تردید همراه ساخت (رضایی، ۲۰۰۵). صورت‌های مالی متقلبانه اثرات منفی بر اقتصادهای دنیا داشته و به زیان‌های مالی قابل‌توجهی برای افراد و شرکت‌ها منجر شده است. ازاین‌رو، در یک محیط تجاری مبتنی بر فناوری که با تغییرات سریع همراه است، نیاز مبرمی به روش‌های مؤثر برای پیشگیری و کشف تقلب وجود دارد. روش‌های داده‌کاوی[5] با به‌کارگیری موارد قبلی تقلب، مدلی برای شناسایی و تشخیص خطر تقلب ارائه می‌دهد که قادر به پیشگیری و کشف تقلب در صورت‌های مالی است (سگال، ۲۰۱۶).

به‌عبارت‌دیگر، دانشگاهیان و موسسه‌های حسابرسی به‌منظور بهبود کیفیت، صداقت و قابلیت اتکای فرآیند گزارشگری مالی به دنبال راه‌هایی برای کشف تقلب در صورت‌های مالی هستند. درحالی‌که پژوهش‌های دانشگاهی بسیاری برای استفاده از رویکردهای کمی در سایر زمینه‌های تقلب مانند بیمه[6]، کارت اعتباری[7] و ارتباطات[8]، انجام‌شده، تلاش اندکی برای استفاده از تکنیک‌های داده‌کاوی جهت کشف تقلب در صورت‌های مالی صورت گرفته است؛ بنابراین، در

1 Segal
2 Enron
3 WorldCom
4 Global Crossing
5 Data Mining
6 Insurance Fraud
7 Credit Card Fraud
8 Telecommunication Fraud

این پژوهش به بررسی این مسئله پرداخته می‌شود که آیا می‌توان با به‌کارگیری تکنیک‌های داده‌کاوی و از طریق شناسایی متغیرهای مؤثر در کشف تقلب در صورت‌های مالی، مدلی برای کشف تقلب در صورت‌های مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران ارائه کرد؟ این پژوهش با معرفی متغیرهای پیشنهادی و تعیین متغیرهایی که شرکت‌های متقلب را به بهترین شکل از شرکت‌های غیر متقلب تفکیک می‌کند، متون حسابداری و حسابرسی را بهبود می‌بخشد.

ازاین‌رو، پس از تعیین شرکت‌های متقلب، مطابق با مبانی نظری و پیشینه پژوهش از تکنیک‌های انتخاب ویژگی و متغیرهای مرتبط با اهرم مالی، سودآوری، ترکیب دارایی‌ها، نقدینگی، کارایی، اندازه، رشد، وضعیت مالی کلی، اندازه حسابرس، دوره تصدی حسابرس و تغییر حسابرس برای انتخاب متغیرهای اثرگذار جهت پیش‌بینی تقلب در صورت‌های مالی استفاده می‌گردد. در ادامه به‌منظور پیش‌بینی تقلب در صورت‌های مالی از تکنیک‌های داده‌کاوی شبکه عصبی مصنوعی، شبکه بیزین و الگوریتم جنگل تصادفی استفاده خواهد شد.

۲- مبانی نظری

طی ۱۵۰ سال گذشته یا فراتر از آن، سازمان‌های تجاری از مالک - مدیر یا واحدهای تجاری خانوادگی، با تعداد اندک کارکنان به شرکت‌های چندملیتی گسترده‌ای که هزاران نفر را در استخدام دارند، تغییر شکل داده‌اند. این رشد در سایه وجوهی که از بازارهای سرمایه تأمین مالی گردیده امکان‌پذیر و به گذر مدیریت از شکل مالکان- سهامدار به گروه کوچکی از مدیران حرفه‌ای منجر شده است. ازاین‌رو، مدیران می‌بایست نتایج عملکرد خود را به مالکان واحد تجاری و سایر تأمین‌کنندگان وجوه مانند بانک‌ها و اعتباردهندگان گزارش کنند. بااین‌حال، گزارشگری مالی شامل طیف وسیعی از اعداد حسابداری برای نشان دادن سود، جریان نقدی و وضعیت مالی واحد تجاری است؛ بنابراین، مدیریت قادر خواهد بود با استفاده نادرست از حسابداری ساختگی یا متهورانه[1] وضعیت مالی شرکت را بهتر از واقعیت نشان دهد و از اثرات آن در کوتاه‌مدت بهره ببرد. هرچند ممکن است عملکرد مالی یک شرکت در سال مالی جاری به دلیل استفاده از حسابداری ساختگی قابل‌تحسین باشد، اما درنهایت اخبار بد آشکار خواهند شد. از این قبیل می‌توان به برخی رسوایی‌های اخیر مانند انرون، وردکام، آدلفیا[2]، آهولد[3] و پارمالات[4] و سایر موارد

1 Creative Or Aggressive Accounting
2 Adelphia
3 Ahold
4 Parmalat

اشاره کرد. ارائه عمدی اطلاعات مالی ناصحیح و انتشار صورت‌های مالی نادرست در بازارهای سرمایه توسط مدیران این شرکت‌ها، پیامدهای مالی مهمی مانند زیان سرمایه‌گذاری‌ها، از دست دادن شغل و از بین رفتن اعتماد در بازارهای سرمایه را به همراه داشته است (اکیک[1]، ۲۰۱۱).

درک کامل از ماهیت، اهمیت و پیامدهای فعالیت‌های مرتبط با گزارشگری مالی متقلبانه مستلزم تعریف مناسبی از تقلب در صورت‌های مالی است. عمدتاً به این دلیل که حرفه حسابداری تنها در طول یک دهه گذشته از واژه تقلب در اعلامیه‌های حرفه‌ای خود استفاده کرده است، یافتن تعریف دقیقی از تقلب صورت‌های مالی در اعلامیه‌ها و بیانیه‌های الزام‌آور دشوار است. انجمن حسابداران رسمی آمریکا[2] در بیانیه شماره ۸۲ استانداردهای حسابرسی، تقلب در صورت‌های مالی را به‌صورت حذف یا ارائه نادرست در صورت‌های مالی به‌صورت عمد تعریف می‌کند (رضایی و ریلی، ۲۰۱۰). مطابق با استاندارد حسابرسی ۲۴۰، تقلب عبارت است از هرگونه اقدام عمدی یا فریبکارانه یک یا چند نفر از مدیران، کارکنان یا اشخاص ثالث، برای برخورداری از یک مزیتی ناروا یا غیرقانونی. هرچند تقلب یک مفهوم گسترده‌ای دارد، اما آنچه به‌حسابرس مربوط می‌شود، اقدامات متقلبانه‌ای است که به تحریف بااهمیت در صورت‌های مالی می‌انجامد. هدف برخی تقلبات ممکن است تحریف صورت‌های مالی نباشد (کمیته تدوین استانداردهای حسابرسی، بخش ۲۴۰: بند ۴).

طرح‌های تقلب در صورت‌های مالی به‌طورمعمول به‌صورت زیر است:

- بیش نمایی دارایی‌ها یا درآمدها
- کم نمایی بدهی‌ها یا هزینه‌ها

بیش نمایی دارایی‌ها و درآمدها با احتساب مبالغ ساختگی بهای تمام‌شده به‌حساب دارایی‌ها یا از طریق شناسایی درآمدهای مصنوعی، وضعیت مالی یک شرکت را بهتر نشان می‌دهد. کم نمایی بدهی‌ها و هزینه‌ها با عدم شناسایی هزینه‌ها یا تعهدات مالی صورت می‌گیرد. هر دو این روش‌ها به افزایش حقوق مالکانه و خالص ثروت شرکت منجر می‌شوند. دست‌کاری اعداد و ارقام صورت‌های مالی افزایش سود هر سهم یا منافع در سود تضامنی را به همراه دارد یا تصویر پایدارتری از وضعیت واقعی شرکت را نشان می‌دهد (راهنمای ممیزان تقلب[3]، ۲۰۱۱).

به‌منظور درک کامل بیش نمایی دارایی‌ها و درآمدها و کم نمایی بدهی‌ها و هزینه‌ها، طرح‌های تقلب مورداستفاده برای بهبود صورت‌های مالی به پنج طبقه دسته‌بندی می‌شوند. با توجه به

1 Okike
2 American Institute of Certified Public Accountants
3 Fraud Examiners Manual

این‌که نگهداری سوابق مالی مستلزم استفاده از سیستم حسابداری دوطرفه است، ثبت‌های حسابداری متقلبانه به‌طورمعمول حداقل بر دو حساب و بنابراین بر دودسته از صورت‌های مالی اثر خواهد گذاشت؛ بنابراین، یک معامله متقلبانه همیشه طرف دیگری خواهد داشت. طرح‌های تقلب می‌توانند دربرگیرنده ترکیبی از چندین روش باشند. طرح‌های تقلب در صورت‌های مالی به پنج طبقه به شرح زیر دسته‌بندی می‌شوند (راهنمای ممیزان تقلب، ۲۰۱۱).

- درآمدهای ساختگی
- تفاوت‌های زمانی
- ارزشیابی نادرست دارایی‌ها
- بدهی‌ها و هزینه‌های پنهان
- افشای نامناسب

اگرچه مسئولیت کشف تقلب و اشتباه با مدیریت و افراد عهده‌دار راهبری شرکت است اما استانداردهای جدید حسابداری بین‌الملل (استاندارد حسابرسی ۲۴۰) از حسابرسان انتظار دارد که تا حدی مسئولیت ثانویه کشف تقلب و اشتباه در صورت‌های مالی را بپذیرند. حسابرسان ملزم هستند که حسابرسی خود را با نگرش تردید حرفه‌ای و ذهنی پرسشگر برنامه‌ریزی و انجام دهند، یعنی ممکن است شرایطی وجود داشته باشد که می‌تواند به ارائه نادرست بااهمیت در صورت‌های مالی منجر گردد. ناکامی در کشف گزارشگری متقلبانه در انجام کار حسابرسی می‌تواند حسابرسان را در معرض پیامدهای حقوقی و قانونی نامساعدی قرار دهد. این امر به هزینه‌های دعاوی حقوقی قابل‌توجه برای حسابرسان و صدمات جبران‌ناپذیر به شهرت آن‌ها منجر می‌شود.

امروزه با تکامل تکنولوژی و شبکه‌های ارتباطی پرسرعت و جهانی، ارتکاب تقلب آسان‌تر و درنتیجه کشف آن دشوارتر گردیده است. اعمال متقلبانه به زیان‌های مالی بزرگی برای ایالت‌ها و واحدهای تجاری در سطح جهان منجر شده است. بنابراین در زمان کنونی، کشف و مبارزه با تقلب جهت جبران زیان‌های وارده بسیار بااهمیت است. علاوه بر این، با تکامل تکنولوژی شکل‌های جدیدی از تقلب مانند تقلب ارتباطات نفوذ کامپیوتری[1]، بیمه و کارت‌های اعتباری به وجود آمده است. هر واحد تجاری ویژگی‌های منحصربه‌فرد خود را دارد و درنتیجه روش‌های کشف و مبارزه در واحدهای تجاری مختلف، متفاوت خواهد بود. در حالت پیچیده‌تر، هر ساله

1 Computer Intrusion

داده‌های بسیاری تولید می‌شوند و کشف اطلاعات متقلبانه مستلزم تکنیک‌های کاراتری جهت کاوش داده‌ها می‌باشد (آلمیدا[1]، ۲۰۰۹).

رویکردهای جدیدی برای کشف تقلب وجود دارد، اما داده‌کاوی که هدف آن استخراج اطلاعات موردعلاقه از مجموعه گسترده‌ای از داده‌ها است به‌طور گسترده‌ای به‌عنوان یک ابزار تصمیم‌گیری فعال مورداستفاده قرارگرفته است. این امر باعث افزایش توجه پژوهشگران در غنی کردن داده‌ها باهدف تسهیل کشف الگوهای موردعلاقه شده است (وایتینگ و همکاران[2]، ۲۰۱۲). داده‌کاوی، استخراج یا کاوش دانش از حجم زیادی اطلاعات است. الگوها یا قواعد قوی کشف‌شده از طریق تکنیک‌های داده‌کاوی را می‌توان برای پیش‌بینی غیر بدیهی[3] داده‌های جدید استفاده کرد. ازاین‌رو، داده‌کاوی یک حوزه میان‌رشته‌ای است که از ابزارهای تجزیه‌وتحلیل مدل‌های آماری، الگوریتم‌های ریاضیاتی و روش‌های یادگیری ماشینی[4] برای کشف الگوها و روابط معتبر قبلاً ناشناخته در مجموعه داده‌های بزرگ استفاده می‌کند (دوا و دو، ۲۰۱۱).

بنابراین، در محیطی فعال از تقلب در صورت‌های مالی، مکانیزم‌های کشف تقلب با کمک رایانه بسیار مؤثرتر و کاراتر خواهد بود. تکنولوژی‌های مبتنی بر آمار و یادگیری ماشینی یک راهکار اثربخش برای پیشگیری و کشف تقلب هستند اما مرتکبین تقلب خود را انطباق می‌دهند و به‌طورمعمول قادر به کشف راه‌هایی برای دور زدن این تکنولوژی‌ها هستند. تکنیک‌های 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 تکنیک‌های کشف تقلب اطلاع کافی دارند، روش‌هایی را برای ارتکاب تقلب بکار می‌گیرند که کشف آن‌ها به‌ویژه با استفاده از تکنیک‌های فعلی دشوار است. ازاین‌رو، یک نیاز مبرم برای روش‌هایی که نه تنها کارا باشند بلکه برای کشف فریبکاری‌های مالی انطباقی و نوظهور مؤثر هستند، وجود دارد (ژو و کپور[5]، ۲۰۱۱).

1 Almeida
2 Whiting, et al.
3 Nontrivial
4 Machine Learning Methods
5 Zhou & Kapoor

۳- پیشینه پژوهش

۳-۱- پژوهش‌های خارجی

سپاتیس[1] (۲۰۰۲)، با نمونه‌ای متشکل از ۷۶ شرکت شامل ۳۸ شرکت متقلب و ۳۸ شرکت غیر متقلب، توانایی اطلاعات حسابداری در کشف گزارشگری مالی متقلبانه را موردبررسی قرار داد. در مجموع، ده متغیرمالی به‌عنوان پیش‌بینی کننده‌های 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، با احتمال بیش‌تری صورت‌های مالی خود را دست‌کاری می‌کنند.

هوگس و همکاران[2] (۲۰۰۷)، با نمونه‌ای متشکل از ۳۹۰ شرکت شامل ۵۱ شرکت متقلب و ۳۳۹ شرکت غیر متقلب طی سال‌های ۲۰۰۴-۱۹۹۸، به معرفی رویکردی تکاملی برای کشف گزارشگری مالی متقلبانه پرداختند. در مجموع، ۸۵ متغیرمالی و غیرمالی به‌عنوان پیش‌بینی کننده‌های بالقوه خطر تقلب در صورت‌های مالی انتخاب شدند. نتایج حاصل از الگوریتم ژنتیک[3] بیان‌گر سودمندی این تکنیک و همچنین اهمیت اطلاعات مالی و غیرمالی در کشف گزارشگری مالی متقلبانه است.

کرکوز و همکارن[4] (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[5]، شبکه‌های عصبی مصنوعی و شبکه‌های باور بیزین[6] در کشف صورت‌های مالی متقلبانه پرداخت. نتایج پژوهش حاکی از این بود که صورت‌های مالی منتشره دربرگیرنده اطلاعات سودمندی برای کشف تقلب در گزارشگری مالی می‌باشند. 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[7] (۲۰۱۲)، با نمونه‌ای متشکل از ۴۵۸ شرکت شامل ۲۲۹ شرکت متقلب و ۲۲۹ شرکت غیر متقلب، سودمندی طبقه‌بندی کننده‌های مبتنی بر قواعد فازی[8] در کشف

1 Spathis
2 Hoogs, et al.
3 Genetic Algorithm
4 Kirkos, et al.
5 Decision Trees
6 Bayesian Belief Networks
7 Alden et al.
8 Fuzzy Rule-Based Classifiers

الگوهای گزارشگری مالی متقلبانه را موردبررسی قرار داد. در مجموع، ۱۸ متغیر مالی به‌عنوان پیش‌بینی کننده‌های بالقوه گزارشگری مالی متقلبانه انتخاب شدند. نتایج حاصل از طبقه‌بندی کننده‌های مبتنی بر قواعد فازی با الگوریتم ژنتیک و الگوریتم برآورد توزیع یادگیری مارکو[1] حاکی از سودمندی مدل‌های تکاملی برای کشف گزارشگری مالی متقلبانه بود.

راویسنکار و همکاران[2] (۲۰۱۱)، با نمونه‌ای متشکل از ۲۰۲ شرکت شامل ۱۰۱ شرکت متقلب و ۱۰۱ شرکت غیر متقلب به بررسی سودمندی تکنیک‌های داده‌کاوی در کشف تقلب در صورت‌های مالی پرداختند. نتایج حاصل از به‌کارگیری تکنیک‌های داده‌کاوی مانند شبکه‌های عصبی چندلایه پیش‌خور[3]، ماشین‌های بردار پشتیبان[4]، برنامه‌ریزی ژنتیک[5]، روش گروهی مدل‌سازی داده‌ها[6]، رگرسیون لجستیک[7] و شبکه عصبی احتمالی[8]، حاکی از برتری شبکه‌های عصبی احتمالی بدون انتخاب ویژگی نسبت به سایر تکنیک‌ها در کشف صورت‌های مالی متقلبانه بود. در صورت به‌کارگیری انتخاب ویژگی، برنامه‌ریزی ژنتیک و شبکه‌های عصبی احتمالی با صحت تقریباً یکسانی عملکردی بهتر از سایر روش‌ها داشتند.

هوانگ و همکاران[9] (۲۰۱۴)، با نمونه‌ای متشکل از ۱۴۴ شرکت شامل ۷۲ شرکت متقلب و ۷۲ شرکت غیر متقلب، به ارائه مدلی مبتنی بر نگاشت خودسازمان‌ده سلسله مراتبی در حال رشد[10] برای کشف تقلب در گزارشگری مالی پرداختند. در مجموع ۲۴ متغیر مستقل برای انجام تحلیل ممیزی جهت تعیین ورودی‌های مدل انتخاب شدند و نتایج حاصل از تحلیل ممیزی حاکی 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، درماندگی مالی و حاکمیت شرکتی اندازه‌گیری می‌کنند. نتایج حاصل از پژوهش بیان‌گر عملکرد مناسب مدل پیشنهادی برای کشف گزارشگری مالی متقلبانه بود.

1 Markovian Learning Estimation of Distribution Algorithm
2 Ravisankar et al.
3 Multilayer Feed Forward Neural Network
4 Support Vector Machines
5 Genetic Programming
6 Group Method of Data Handling
7 Logistic Regression
8 Probabilistic Neural Network
9 Huang et al.
10 Growing Hierarchical Self-Organizing Map

دالنیال و همکاران[1] (۲۰۱۴)، با نمونه‌ای متشکل از ۶۵ شرکت متقلب و ۶۵ شرکت غیر متقلب پذیرفته‌شده در بورس اوراق بهادار مالزی به بررسی سودمندی متغیرهای مالی در کشف شرکت‌های متقلب طی سال‌های ۲۰۱۱-۲۰۰۰ پرداختند. نتایج پژوهش حاکی از این بود که تفاوت معناداری بین میانگین نسبت‌های کل بدهی به کل حقوق صاحبان سهام و حساب‌های دریافتنی به فروش در شرکت‌های متقلب و شرکت‌های غیر متقلب وجود دارد. همچنین، امتیاز Z آلتمن که احتمال ورشکستگی را اندازه‌گیری می‌کند از اهمیت زیادی در کشف گزارشگری مالی متقلبانه برخوردار است.

لین و همکاران[2] (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 تکنیک‌های داده‌کاوی رگرسیون لجستیک، درخت تصمیم و شبکه‌های عصبی مصنوعی حاکی از دقت بیش‌تر روش‌های درخت تصمیم و شبکه‌های عصبی مصنوعی نسبت به رگرسیون لجستیک است. در پایان نیز به‌منظور بهبود دستاوردهای پژوهش مقایسه‌ای بین قضاوت متخصصان و تکنیک‌های داده‌کاوی صورت گرفته است.

۳-۲- 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 صادرشده 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 زلقی (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 معیارهای؛ ۱. اظهارنظر غیر مقبول حسابرسی، ۲. وجود اختلاف مالیاتی با حوزه مالیاتی و ۳.

1 Dalnial et al.
2 Lin et al.

وجود تعدیلات سنواتی بااهمیت استفاده‌شده است. 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 پذیرفته‌شده در بورس اوراق بهادار تهران پرداختند. نتایج این پژوهش حاکی از این بود که در شرکت‌های باسابقه 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 می‌گردد.

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 انتخاب‌شده‌اند. نتایج پژوهش بیان‌گر یک رابطه منفی و معنادار میان رشد درآمد، رشد کارکنان و اختلاف این دو متغیر با گزارشگری مالی متقلبانه بود.

مشایخی و حسین‌پور (۱۳۹۵)، با نمونه‌ای متشکل از ۱۰۷ شرکت مشکوک به تقلب به بررسی رابطه مدیریت سود واقعی و مدیریت سود تعهدی در شرکت‌های مشکوک به تقلب بورس اوراق بهادار تهران پرداختند. شرکت‌های مشکوک به تقلب بر مبنای یکسری عوامل مرتبط با ارائه نادرست و گمراه‌کننده اطلاعات حسابداری در گزارشگری مالی گزینش‌شده‌اند. نتایج حاصل از پژوهش حاکی این است که در شرکت‌های بورسی مشکوک به تقلب، مدیریت سود واقعی بر مدیریت سود تعهدی تأثیر منفی و معنی‌داری دارد.

۵- 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 (سریجیش و همکاران[1]، ۲۰۱۴). در این حالت، طرح‌های شبه تجربی به کنترل متغیرهای برون‌زا کمک می‌کنند اما به دلیل کنترل عوامل مرتبط با کاهش روایی داخلی به‌اندازه طرح‌های تجربی واقعی سودمند نیستند (زیکموند و همکارن[2]،

1 Sreejesh et al.
2 Zikmund et al.

۲۰۱۰). برخی مواقع، طرح‌های شبه تجربی تنها راه برای انجام پژوهش هستند. طرح‌های سری زمانی، معروف‌ترین طرح شبه تجربی مورداستفاده به‌وسیله پژوهشگران است.

۴-۱- سؤال پژوهش

با توجه به مبانی نظری و پیشینه پژوهش، سؤال پژوهش به شرح زیر تدوین شد:

آیا می‌توان با استفاده از اطلاعات مالی و غیرمالی 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 تعیین شرکت‌های متقلب باشد؟

۴-۲- متغیرهای پژوهش

متغیر پاسخ

متغیر پاسخ در این پژوهش، تقلب در صورت‌های مالی است که مقدار آن برای شرکت‌های متقلب عدد یک و برای شرکت‌های غیر متقلب عدد صفر است. شرکت‌های متقلب با مراجعه به نشریات ویژه سازمان بورس و اوراق بهادار که در وب‌سایت سازمان بورس و اوراق بهادار نمایه شده‌اند انتخاب گردیدند. ازاین‌رو بررسی کامل نشریات ویژه نشان می‌دهد نشریات شماره ۳، ۴ و ۷ دربرگیرنده چهارده رأی قطعی صادره در محاکم دادگستری می‌باشند. هر یک از این چهارده رأی در راستای اجرای ماده ۵۲ قانون بازار اوراق بهادار مصوب آذرماه ۱۳۸۴ به دلایل مختلفی از قبیل ایجاد ظاهری گمراه‌کننده از روند معاملات اوراق بهادار، مبادرت به معاملات اوراق بهادار با استفاده از اطلاعات نهانی، خودداری از ارائه اطلاعات، 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 صورت دارایی و ترازنامه مزور و موارد دیگر صادرشده‌اند. با توجه به این‌که موضوع پژوهش حاضر به‌طور مشخص تقلب در صورت‌های مالی است، شرکت‌هایی که فقط به دلایل مرتبط با تقلب در صورت‌های مالی علیه مدیران آن‌ها در محاکم دادگستری آرای قطعی صادرشده است و با بررسی اطلاعات ارائه‌شده در نشریات ویژه سازمان بورس و اوراق بهادار قابل‌تشخیص بودند، به‌عنوان شرکت‌های متقلب در نظر گرفته شدند.

متغیرهای پیش‌بینی کننده

در این پژوهش، عوامل مالی و غیرمالی خطر تقلب مرتبط با گزارشگری مالی متقلبانه متغیرهای پیش‌بین هستند که با مرور پژوهش‌های انجام‌شده در این زمینه تعیین خواهند شد. در یک دسته‌بندی کلی، این عوامل عبارت‌اند از: اهرم مالی، سودآوری، ترکیب دارایی‌ها، نقدینگی، کارایی،

اندازه، رشد، وضعیت مالی کلی و کیفیت حسابرسی. جدول (۱) خلاصه متغیرهای مرتبط با هر یک از این عوامل را نشان می‌دهد.

جدول(۱). خلاصه متغیرهای پیش‌بینی کننده

عامل	متغیرها	نام پژوهش
اهرم مالی	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.	پرسونس[1] (۱۹۹۵)، سپاتیس (۲۰۰۲)، سپاتیس و همکاران[2] (۲۰۰۲)، کامینسکی و همکاران[3] (۲۰۰٤)، کرکوز و همکاران (۲۰۰۷)، برازل و همکاران[4] (۲۰۰۹)، آلدن و همکاران (۲۰۱۲)، چن و همکاران[5] (۲۰۱٤).
سودآوری	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.	پرسونس (۱۹۹۵)، سپاتیس (۲۰۰۲)، سپاتیس و همکاران (۲۰۰۲)، کامینسکی و همکاران (۲۰۰٤)، کرکوز و همکاران (۲۰۰۷)، برازل و همکاران (۲۰۰۹)، آلدن و همکاران (۲۰۱۲)، چن و همکاران (۲۰۱٤).
ترکیب دارایی‌ها	تغییر در حساب‌های دریافتنی، تغییر در موجودی کالا و نسبت‌های جمع دارایی‌های جاری به کل دارایی‌ها، حساب‌های دریافتنی به کل دارایی‌ها، موجودی کالا به فروش خالص، حساب‌های دریافتنی به فروش خالص و نسبت موجودی کالا به جمع دارایی‌ها.	پرسونس (۱۹۹۵)، سپاتیس (۲۰۰۲)، سپاتیس و همکاران (۲۰۰۲)، کرکوز و همکاران (۲۰۰۷)، برازل و همکاران (۲۰۰۹)، آلدن و همکاران (۲۰۱۲)، چن و همکاران (۲۰۱٤)، راعی و همکاران (۱۳۹۳).
نقدینگی	نسبت‌های آنی، نقدی، جاری، وجه نقد به کل دارایی‌ها و نسبت سرمایه در گردش به کل دارایی‌ها.	پرسونس (۱۹۹۵)، سپاتیس (۲۰۰۲)، سپاتیس و همکاران (۲۰۰۲)، کامینسکی و همکاران (۲۰۰٤)، کرکوز و همکاران (۲۰۰۷)، برازل و همکاران (۲۰۰۹)، آلدن و همکاران (۲۰۱۲)، چن و همکاران (۲۰۱٤).
کارایی	نسبت‌های فروش به کل دارایی‌ها، فروش به کل دارایی‌های ثابت، فروش به حساب‌های دریافتنی و بهای تمام‌شده کالای فروش رفته به موجودی کالا.	پرسونس (۱۹۹۵)، سپاتیس (۲۰۰۲)، سپاتیس و همکاران (۲۰۰۲)، کامینسکی و همکاران (۲۰۰٤)، کرکوز و همکاران (۲۰۰۷)، برازل و همکاران (۲۰۰۹)، آلدن و همکاران (۲۰۱۲)، چن و همکاران (۲۰۱٤).

1 Persons
2 Spathis et al.
3 Kaminski et al.
4 Brazel et al.
5 Chen et al.

ادامه جدول (۱). خلاصه متغیرهای پیش‌بینی کننده

عامل	متغیرها	نام پژوهش
اندازه	لگاریتم طبیعی کل دارایی‌ها، لگاریتم طبیعی کل فروش و لگاریتم طبیعی ارزش بازار شرکت.	پرسونس (۱۹۹۵)، سپاتیس (۲۰۰۲)، سپاتیس و همکاران (۲۰۰۲)، کامینسکی و همکاران (۲۰۰٤)، کرکوز و همکاران (۲۰۰۷)، برازل و همکاران (۲۰۰۹)، آلدن و همکاران (۲۰۱۲)، چن و همکاران (۲۰۱٤).
وضعیت مالی کلی	از Z آلتمن برای اندازه‌گیری سلامت مالی یک شرکت استفاده می‌شود.	ستایس[1] (۱۹۹۱)، پرسونس (۱۹۹۵)، سامرز و سوینی[2] (۱۹۹۸)، سپاتیس (۲۰۰۲)، سپاتیس و همکاران (۲۰۰۲)، کرکوز و همکاران (۲۰۰۷).
رشد	برای اندازه‌گیری رشد شرکت از متغیر رشد فروش استفاده می‌شود.	ستایس (۱۹۹۱)، سامرز و سوینی (۱۹۹۸)، کرکوز و همکاران (۲۰۰۷).
کیفیت حسابرسی	اندازه حسابرس یک متغیر موهومی است که اگر حسابرس شرکت سازمان حسابرسی باشد، مقدار آن یک و در غیر این صورت صفر می‌باشد. تعداد سال‌های پیاپی که حسابرس در استخدام یک شرکت است، دوره تصدی حسابرس نامیده می‌شود. متغیر تغییر حسابرس، به‌صورت یک متغیر دوتایی تعریف می‌شود که در آن عدد یک نشان‌دهنده صاحبکار جدید و عدد صفر نشان‌دهنده صاحبکار ثابت است.	پیر و اندرسن[3] (۱۹۸٤)، تیو و وانگ (۱۹۹۳)، ستایس (۱۹۹۱)، کارسلو و ناجی (۲۰۰٤)، پرسونس (۱۹۹۵)، سامرز و سوینی (۱۹۹۸)، برازل و همکاران (۲۰۰۹)، پورحیدری و همکاران (۱۳۹٤).

۴-۳- مدل‌های پژوهش

داده‌کاوی در رشته‌های گوناگونی مانند مالی، مهندسی، زیست پزشکی[4] و امنیت سایبری[5] استفاده می‌شود. روش‌های داده‌کاوی به دودسته تقسیم می‌شوند: نظارت‌شده[6] و غیر نظارت‌شده[7]. تکنیک‌های 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

1 Stice
2 Summers & Sweeny
3 Pierre & Anderson
4 Biomedicine
5 Cybersecurity
6 Supervised
7 Unsupervised

مخفی با استفاده از داده‌های معین و بدون معرفی داده‌های آموزشی (مانند جفت‌های ورودی و برچسب کلاس) هستند. خوشه‌بندی[1] و کاوش قواعد وابستگی[2] نمونه‌هایی از کاوش غیر نظارتی هستند.

داده‌کاوی همچنین بخشی از کشف دانش در پایگاه‌های داده، یک فرآیند 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 می‌گردد. این فرآیند تکراری از گام‌های زیر تشکیل می‌گردد: پاک‌سازی داده‌ها[3]، تجمیع داده‌ها[4]، انتخاب داده‌ها[5]، انتقال داده‌ها[6]، داده‌کاوی، ارزیابی الگو[7] و نمایش شناخت[8] (دوا و دو[9]، ۲۰۱۱).

انتخاب ویژگی

داده‌های نامتوازن به‌طورمعمول با فضای ویژگی با ابعاد بالا همراه هستند. در میان ویژگی‌های با ابعاد بالا، وجود تعداد زیادی ویژگی‌های پر اختلال[10] می‌تواند عملکرد طبقه‌بندی کننده را مختل و کاهش دهد. در سال‌های اخیر، روش‌های انتخاب ویژگی و زیر فضا[11] 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 به‌صورت واحد انتخاب می‌کنند و 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 ویژگی‌ها دارای همبستگی متقابل هستند مؤثر نیست. روش‌های انتخاب برداری ویژگی بهترین ترکیب برداری ویژگی را انتخاب می‌کند. روش‌های

1 Clustering
2 Associative Rule Mining
3 Data Cleaning
4 Data Integration
5 Data Selection
6 Data Transformation
7 Pattern Evaluation
8 Knowledge Representation
9 Dua & Du
10 Noisy
11 Subspace

انتخاب برداری ویژگی را می‌توان به دودسته روش‌های دسته‌بندی[1] و روش‌های مبتنی بر فیلتر[2] تقسیم می‌شوند. روش‌های دسته‌بندی با استفاده از روش‌های یادگیری ماشینی، مانند جعبه سیاه، ویژگی‌های مرتبط‌تر را انتخاب می‌کنند به‌گونه‌ای که روش یادگیری دارای عملکرد بهینه است. استراتژی‌های جستجو دربرگیرنده جستجوی فراگیر، جستجوی پرتو، شاخه و حد، الگوریتم ژنتیک، روش‌های جستجوی حریصانه و موارد دیگر هستند. هنگام استفاده از روش دسته‌بندی، ویژگی‌های انتخاب‌شده مستعد بیش برازش 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 به‌منظور انتخاب متغیرهایی که بیشترین تأثیر را در کشف تقلب در صورت‌های مالی دارند، از روش انتخاب ویژگی مبتنی بر همبستگی ارائه‌شده توسط هال[3] (۱۹۹۹) 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[4] (PCC)، اطلاعات متقابل[5] (MI) و سایر معیارهای مرتبط باشد. ضریب همبستگی پیرسون معیاری برای وابستگی خطی بین متغیرها و ویژگی‌ها می‌باشد 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[6] اندازه‌گیری می‌کند. بااین‌حال، برآورد اطلاعات متقابل نسبت به ضریب همبستگی پیرسون، به‌ویژه برای داده‌های پیوسته، سخت‌تر است (دوا و دو، ۲۰۱۱).

طبقه‌بندی

طبقه‌بندی شکلی از تحلیل داده است که به استخراج مدل‌های توضیح‌دهنده طبقات مهم داده می‌پردازد. در این پژوهش برای دسته‌بندی شرکت‌ها به متقلب و غیر متقلب از روش‌های شبکه عصبی مصنوعی، شبکه بیزین[7] و الگوریتم جنگل تصادفی[8] استفاده خواهد شد که در ادامه به شرح مختصر آن‌ها پرداخته می‌شود.

1 Wrapper
2 Filter
3 Hall
4 Pearson’s Correlation Coefficient
5 Mutual Information
6 Kullback-Leibler
7 Bayesian Network
8 Random Forest

شبکه عصبی مصنوعی

شبکه عصبی مصنوعی یک مدل یادگیری ماشینی است که ورودی‌ها را از طریق پردازش اطلاعات غیرخطی به خروجی‌هایی که با اهداف مطابقت داده می‌شوند در یک گروه متصل از نرون‌های مصنوعی که لایه‌های واحدهای پنهان را تشکیل می‌دهند، متصل می‌کند. فعالیت هر واحد پنهان و خروجی $\hat{Y}$ از ترکیب ورودی X و یک مجموعه از وزن‌های نرون W تعیین می‌شود:

$$\hat{Y} = f(X, W) \qquad (۱)$$

که در این رابطه به ماتریس بردارهای وزن لایه‌های پنهان اشاره دارد (دوا و دو، ۲۰۱۱). زمانی که شبکه‌های عصبی مصنوعی به‌عنوان یک روش یادگیری ماشینی مورداستفاده قرار می‌گیرد، تلاش‌هایی برای تعیین مجموعه وزن‌هایی جهت کمینه کردن خطای طبقه‌بندی صورت می‌گیرد. همگرایی حداقل میانگین مربعات یک روش شناخته‌شده متداول برای بسیاری از پارادایم‌های یادگیری است. هدف شبکه‌های عصبی مصنوعی کمینه کردن خطاهای بین واقعیت زمینی Y و خروجی مورد انتظار $f(X, W)$ یک شبکه عصبی مصنوعی است به‌طوری که:

$$E(X) = ((f(X, W) - Y)^2 \qquad (۲)$$

روش‌های شبکه عصبی مصنوعی برای طبقه‌بندی یا پیش‌بینی متغیرهای پنهان که اندازه‌گیری آن‌ها دشوار است و حل مسائل طبقه‌بندی غیرخطی به‌خوبی عمل می‌کنند و نسبت به داده‌های پرت حساسیت ندارند. مدل‌های شبکه عصبی مصنوعی رابطه بین ورودی و خروجی را به‌صورت ضمنی تعریف می‌کنند و ازاین‌رو راه‌حل‌هایی برای مسائل ملال‌آور شناخت الگو به‌ویژه زمانی که استفاده‌کنندگان هیچ ایده‌ای درباره رابطه بین متغیرها ندارند، ارائه می‌کند (دوا و دو، ۲۰۱۱).

شبکه بیزین

شبکه بیزین که شبکه باور نیز نامیده می‌شود از توزیع احتمال مشترک عاملی در یک مدل گرافیکی برای تصمیم‌گیری در رابطه با متغیرهای نامطمئن استفاده می‌کند. طبقه‌بندی کننده شبکه بیزین بر مبنای قاعده بیز که ارائه‌دهنده فرضیه H از طبقات و داده x است، قرار دارد، بنابراین؛

$$P(H|x) = \frac{P(x|H)P(H)}{P(x)} \qquad (۳)$$

که در این رابطه:

$P(H)$ نشان‌دهنده احتمالات پیشین هر طبقه بدون اطلاعاتی درباره متغیر x است.

$P(H|x)$ بیان‌گر احتمالات پسین متغیر x بر طبقات ممکن است.
$P(x|H)$ بیان‌گر احتمالات مشروط متغیر x با احتمال H است.

نایو بیز شکل ساده‌ای از مدل شبکه بیزین است که فرض می‌کند همه متغیرها از یکدیگر مستقل هستند. جهت به‌کارگیری قاعده بیز برای طبقه‌بندی نایو بیز یافتن فرضیه حداکثر احتمال که برچسب طبقه را برای هر داده آزمون x تعیین می‌کند ضروری است. با مشاهدات x و یک گروه از برچسب طبقات $C = \{c_j\}$، طبقه‌بندی کننده نایو بیز از طریق بیشینه کردن فرضیه احتمال پسین (MAP) برای داده‌ها به شرح زیر حل شود.

$$\underset{c_j \in C}{argmax}\, P(x|c_j)P(c_j) \qquad (۴)$$

نایو بیز به‌منظور انجام وظایف استنباطی کارا است. بااین‌حال، به‌شدت بر مبنای فرض مستقل بودن متغیرهای درگیر قرار دارد. باکمال شگفتی، این روش حتی در صورت نقض شرط مستقل بودن متغیرها نیز نتایج خوبی به همراه دارد (دوا و دو، ۲۰۱۱).

جنگل تصادفی

الگوریتم جنگل تصادفی، معروف‌ترین طبقه‌بندی کننده Bagging Ensemble می‌باشد. جنگل تصادفی دربرگیرنده تعداد بسیاری درخت تصمیم می‌باشد. خروجی جنگل تصادفی بر مبنای آرای مشخص هر یک از درخت‌ها تعیین می‌شود. هر درخت تصمیم با طبقه‌بندی نمونه‌های Bootstrap داده‌های ورودی از طریق الگوریتم درخت ساخته می‌شود. سپس، هر درخت برای طبقه‌بندی داده‌های آزمون استفاده خواهد شد. هر درخت یک تصمیم برای برچسب‌گذاری داده‌های آزمون دارد. این برچسب، یک رأی نام دارد. درنهایت جنگل با جمع‌آوری بیش‌ترین آرای درخت‌ها نتیجه طبقه‌بندی را قطعی خواهد کرد. در ادامه با برخی تعاریف ارائه‌شده توسط بریمن[1] (۲۰۰۱) به بررسی جنگل‌های تصادفی پرداخته می‌شود. با فرض این‌که یک جنگل متشکل از K درخت $\{T_1, \dots, T_k\}$، بردار تصادفی θ_k برای درخت kام ایجاد می‌شود، $k = 1, \dots, K$. بردارهای $\{\theta_k\}$ مستقل و دارای توزیع مشابه برای مدل‌سازی درخت هستند. این بردارها در ساخت درخت تعریف‌شده‌اند. برای نمونه در انتخاب تصادفی، این بردارها شامل اعداد صحیح تصادفی هستند که به‌طور تصادفی از $\{1, \dots, N\}$ انتخاب‌شده‌اند که N عدد جداکننده است. با استفاده از مجموعه داده آموزشی و بردارهای $\{\theta_k\}$، درخت رشد می‌کند و یک رأی واحد برای معروف‌ترین طبقه در ورودی x ارائه می‌نماید. طبقه‌بندی کننده درخت kام به‌صورت

1 Breiman

$f(x, \theta_k)$ نشان داده می‌شود و حاصل یک جنگل تصادفی متشکل از مجموعه آن درخت‌ها خواهد بود.

$$\left\{f\{x, \theta_k\}\right\}, k = 1, \dots, K \qquad (۵)$$

صحت جنگل تصادفی به قوت هر یک از درخت‌ها و معیار وابستگی بین درخت‌ها بستگی دارد. علاوه بر این، الگوریتم جنگل تصادفی برای اجتناب از سویه در مدل‌سازی درخت از بوت‌استرپ استفاده می‌کند و درنتیجه اعتبارسنجی متقابل (CV) در آموزش و آزمون موردنیاز نیست. بااین‌حال، به دلیل حداکثر سازی صحت پیش‌بینی در الگوریتم جنگل تصادفی، این طبقه‌بندی کننده با مشکل طبقات نامتوازن روبرو خواهد بود. روش‌های مبتنی بر درخت پراکندگی بالایی دارند. ساختار سلسله مراتبی درخت‌ها می‌تواند به نتیجه ناپایدار منجر شود. میانگین تعداد بسیاری درخت، مانند استفاده از بگینگ، می‌تواند پایداری الگوریتم‌های یادگیری انسمبل را بهبود دهد (دوا و دو، ۲۰۱۱).

عملکرد طبقه‌بندی کننده[1]

به‌منظور ارزیابی عملکرد هر یک از طبقه‌بندی کننده‌های شبکه عصبی مصنوعی، شبکه بیزین و الگوریتم جنگل تصادفی، مطابق با ماتریس درهم‌ریختگی[2] زیر از معیارهای صحت کلی[3]، دقت[4]، فراخوانی[5]، $F - Measure$ و AUC استفاده خواهد شد. با یک طبقه‌بندی کننده و یک نمونه مشخص، چهار خروجی احتمالی وجود دارد. چنانچه نمونه مثبت باشد و به‌عنوان مثبت نیز طبقه‌بندی شده باشد، مثبت حقیقی به‌حساب خواهد آمد و چنانچه این نمونه به‌صورت منفی طبقه‌بندی شود، منفی کاذب است. درصورتی‌که نمونه منفی بوده و به‌عنوان منفی طبقه‌بندی شده باشد، منفی حقیقی به‌حساب می‌آید و چنانچه این نمونه به‌صورت مثبت طبقه‌بندی شود، مثبت کاذب خواهد بود. نمودار شماره یک نشان‌دهنده ماتریس درهم‌ریختگی و معادلات مربوط به معیارهای متفاوتی است که با استفاده از آن محاسبه می‌شوند.

1 Classifier Performance
2 Confusion Matrix
3 Overall Accuracy
4 Precision
5 Recall

جدول (۲). ماتریس درهم‌ریختگی و محاسبه برخی معیارهای عملکرد متداول

طبقه‌های پیش‌بینی‌شده	طبقه‌های واقعی: p	طبقه‌های واقعی: N
Y	مثبت‌های واقعی (True Positives)	مثبت‌های کاذب (False Positives)
N	منفی‌های کاذب (False Negatives)	منفی‌های واقعی (True Negatives)
جمع ستون‌ها	P	N

درصد مثبت کاذب: $fp\ rate = \frac{FP}{P}$ (۶)

درصد مثبت واقعی: $tp\ rate = \frac{TP}{N}$ (۷)

دقت: $precision = \frac{TP}{TP + FP}$ (۸)

فراخوانی: $Recall = \frac{TP}{P}$ (۹)

صحت کلی: $Overall\ Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$ (۱۰)

$$F - Measure = \frac{2}{{}^{1}/_{precision} + {}^{1}/_{Recall}} \quad (۱۱)$$

فضای زیر منحنی ROC (AUC)

منحنی ROC ترسیم دوبعدی عملکرد طبقه‌بندی کننده است که در آن درصد مثبت حقیقی بر روی منحنی Y و درصد مثبت کاذب بر روی منحنی X ترسیم می‌شود. به‌منظور مقایسه طبقه‌بندی کننده‌ها ممکن است به دنبال کاهش عملکرد ROC به یک مقدار برداری واحد که بیان‌گر عملکرد مورد انتظار است باشیم. یک معیار متداول، محاسبه مساحت زیر نمودار ROC است که به اختصار AUC نامیده می‌شود (فاوست[1]، ۲۰۰۶).

که در این روابط؛ Y، شرکت‌هایی که در واقعیت متقلب هستند؛ N، شرکت‌هایی که در واقعیت غیر متقلب هستند؛ p، شرکت‌هایی که مدل به‌عنوان متقلب پیش‌بینی می‌کند؛ n، شرکت‌هایی که توسط مدل به‌عنوان غیر متقلب پیش‌بینی می‌شود؛ TP، مثبت‌های واقعی؛ FP، مثبت‌های کاذب؛ FN، منفی‌های کاذب؛ TN، منفی‌های واقعی و $F-Measure$، معیار عملکرد F است.

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

جامعه آماری این پژوهش، کلیه شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران است. در انتخاب نمونه مواردی به این شرح در نظر گرفته خواهد شد که صورت‌های مالی و یادداشت‌های همراه شرکت‌ها در دوره زمانی ۱۳۷۹ الی ۱۳۹۳ به گونه کامل در سایت بورس اوراق بهادار موجود باشد و شرکت انتخابی جزء شرکت‌های بیمه، بانک‌ها و موسسه‌های اعتباری، واسطه‌گری‌های مالی و شرکت‌های چند رشته‌ای صنعتی نباشد. همچنین، بررسی کامل نشریات ویژه سازمان بورس و اوراق بهادار حاکی از چهارده رأی قطعی صادره در محاکم دادگستری است که با توجه به تعریف تقلب در صورت‌های مالی، تنها ۶ شرکت از ۱۴ شرکت مزبور به‌عنوان شرکت‌هایی که صورت‌های مالی متقلبانه داشته‌اند انتخاب گردیدند. علاوه بر این ۶ شرکت، ۶۴ شرکت غیر متقلب نیز با رعایت دقت لازم و به‌گونه‌ای انتخاب گردیدند که جزء شرکت‌های بیمه، بانک‌ها و موسسه‌های اعتباری، واسطه‌گری‌های مالی و شرکت‌های چند رشته‌ای صنعتی نباشد، اندازه مشابهی با شرکت‌های متقلب داشته باشند و در دوره زمانی یکسانی با شرکت‌های متقلب قرار گیرند و هیچ‌کدام از علائم خطر تقلب در گزارش حسابرسی آن‌ها وجود نداشته باشد.

ازآنجایی‌که تعداد شرکت‌های متقلب در مقایسه با تعداد شرکت‌های غیر متقلب بسیار کم‌تر است، یک مجموعه داده نامتوازن وجود دارد. ازاین‌رو برای حل این مسئله از رویکرد معرفی‌شده

1 Fawcett

توسط هی و همکاران[1] (۲۰۰۸) استفاده خواهد شد. در این رویکرد که آدیسین[2] 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 می‌باشد، ایجاد می‌شود.

۴-۵- روش گردآوری داده‌ها

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

۵-یافته‌های پژوهش

نتایج پژوهش‌های پیشین حاکی از این است که میانگین دوره زمانی وقوع تقلب قبل از کشف آن ۱۸ ماه می‌باشد (جامعه ممیزان رسمی تقلب[3]، ۲۰۱۲). ازاین‌رو در ادامه 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 انتخاب ویژگی مبتنی بر همبستگی و روش‌های شبکه عصبی مصنوعی، شبکه بیزین و الگوریتم جنگل تصادفی برای پیش‌بینی تقلب در صورت‌های مالی برای به ترتیب برای هر یک از سال‌های یک سال قبل از وقوع تقلب و دو سال قبل از وقوع آن ارائه‌شده است. به‌کارگیری داده‌های آموزشی جهت برآورد عملکرد مدل ممکن است با سویه همراه باشد. در بیش‌تر موارد مدل‌ها به‌جای یادگیر تمایل به حفظ کردن نمونه دارند (بیش برازش[4] داده‌ها). در این پژوهش برای اجتناب از این مشکل از اعتبارسنجی متقابل با ۱۰ لایه[5] استفاده‌شده است؛ بنابراین برای هر لایه، مدل با استفاده از ۹ لایه باقیمانده آموزش داده می‌شود و با لایه جداشده آزمون می‌گردد. درنهایت میانگین عملکرد محاسبه خواهد گردید.

---

1 He et al.
2 ADASYN
3 Association of Certified Fraud Examiners
4 Overfitting
5 10-Fold Cross Validation

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

	شبکه عصبی مصنوعی	شبکه بیزین	جنگل تصادفی
صحت کلی	٪۹۲/۷٤	٪۹۱/۱۳	٪۹٤/۳۵
دقت	٪۸٦/۹٦	٪۸٤/۵۱	٪۹۰/۷۷
فراخوانی	٪۱۰۰/۰۰	٪۱۰۰/۰۰	٪۹۸/۳۳
$F-Measure$	٪۹۳/۰۲	٪۹۱/٦۰	٪۹٤/٤۰
AUC	۰/۹۱۳	۰/۹۷۱	۰/۹۹٦

همان‌طور که در جدول بالا مشاهده می‌شود نتایج معیارهای عملکرد مدل‌های پیشنهادی یک سال قبل از وقوع تقلب حاکی از مناسب بودن مدل‌های شبکه عصبی مصنوعی، شبکه بیزین و الگوریتم جنگل تصادفی برای پیش‌بینی تقلب است. به‌منظور مقایسه عملکرد مدل‌های پیشنهادی در کشف تقلب از آزمون‌های آماری استفاده خواهد شد. همان‌طور که پیش‌ازاین اشاره شد در این پژوهش برای برآورد معیارهای ارزیابی عملکرد از رویکرد اعتبار سنجی متقابل با ۱۰ لایه و در ۱۰ تکرار استفاده‌شده است؛ بنابراین برای بررسی فرض نرمال بودن توزیع معیارهای عملکرد محاسبه‌شده از آزمون کولموگوروف- اسمیرنوف (KS) استفاده گردید. نتایج حاصل از این آزمون در سطح معناداری ۵ درصد حاکی از رد فرض H_0 مبنی بر نرمال بودن توزیع معیارهای ارزیابی عملکرد می‌باشد. بنابراین از آزمون رتبه علامت‌دار ویلکاکسون (Wilcoxon) برای مقایسه معیارهای ارزیابی عملکرد استفاده خواهد شد. جدول شماره (۴) نتایج حاصل از آزمون رتبه علامت‌دار ویلکاکسون (Wilcoxon) را برای معیار ارزیابی عملکرد AUC یک سال قبل از وقوع تقلب نشان می‌دهد. در حالت کلی تفاوت معناداری بین عملکرد الگوریتم جنگل تصادفی با شبکه عصبی مصنوعی و شبکه بیزین وجود دارد. ازاین‌رو، الگوریتم جنگل تصادفی روش مناسب‌تری برای طبقه‌بندی شرکت‌های متقلب و غیر متقلب است.

جدول (٤). نتایج حاصل از آزمون رتبه علامت‌دار ویلکاکسون (Wilcoxon)

طبقه‌بندی کننده‌ها	آماره آزمون	آماره Z	احتمال آماره Z
شبکه عصبی مصنوعی و شبکه بیزین	٤۹۹/۵۰۰	۳/۰۰۰	۰/۰۰۳
شبکه عصبی مصنوعی و جنگل تصادفی	۱۱۰۲/۵۰۰	٦/۰٦۲	۰/۰۰۰
شبکه بیزین و جنگل تصادفی	٤۱۸/۵۰۰	٤/۳۸۲	۰/۰۰۰

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

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

	شبکه عصبی مصنوعی	شبکه بیزین	جنگل تصادفی
صحت کلی	٪۹۵/۹۷	٪۹۷/۵۸	٪۹۸/۳۹
دقت	٪۹۲/۳۱	٪۹۵/۲٤	٪۹٦/۷۷
فراخوانی	٪۱۰۰/۰۰	٪۱۰۰/۰۰	٪۱۰۰/۰۰
$F-Measure$	٪۹٦/۰۰	٪۹۷/۵٦	٪۹۸/۳٦
AUC	۰/۹۵۹	۱/۰۰۰	۱/۰۰۰

چنانکه در جدول 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 الگوریتم جنگل تصادفی و شبکه بیزین نسبت به شبکه عصبی مصنوعی برای پیش‌بینی تقلب است.

جدول (٦). نتایج حاصل از آزمون رتبه علامت‌دار ویلکاکسون (Wilcoxon)

طبقه‌بندی کننده‌ها	آماره آزمون	آماره Z	احتمال آماره Z
شبکه عصبی مصنوعی و شبکه بیزین	۸۲۰/۰۰۰	۵/۵٦٤	۰/۰۰۰
شبکه عصبی مصنوعی و جنگل تصادفی	۸۲۰/۰۰۰	۵/۵٦٤	۰/۰۰۰
شبکه بیزین و جنگل تصادفی	۰/۰۰۰	NAN	۱/۰۰۰

جدول شماره (۶) نتایج حاصل از آزمون رتبه علامت‌دار ویلکاکسون (Wilcoxon) برای معیار ارزیابی عملکرد AUC را برای دو سال قبل از وقوع تقلب نشان می‌دهد. ازاین‌رو، تفاوت معناداری بین عملکرد الگوریتم جنگل تصادفی و شبکه بیزین با شبکه عصبی مصنوعی وجود دارد. اما بین عملکرد الگوریتم جنگل تصادفی و شبکه بیزین، تفاوت معناداری مشاهده نشد. بنابراین، الگوریتم جنگل تصادفی و شبکه بیزین روش مناسب‌تری برای طبقه‌بندی شرکت‌های متقلب و غیر متقلب است.

همچنین در این پژوهش از آزمون من‌ویتنی (آزمون U) 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. جدول شماره (۷) نتایج حاصل این آزمون را برای هر یک از مدل‌های پیشنهادی نشان می‌دهد.

جدول (۷). نتایج حاصل از آزمون من‌ویتنی (آزمون U)

طبقه‌بندی کننده‌ها	آماره آزمون	آماره Z	احتمال آماره Z
شبکه عصبی مصنوعی	٦٦٤٠/٠٠٠	٤/٤١١	٠/٠٠٠
شبکه بیزین	٦٧٠٠/٠٠٠	٦/٣٨٤	٠/٠٠٠
جنگل تصادفی	٥٥٠٠/٠٠٠	٣/٢٣٦	٠/٠٠١

همان‌طور که در جدول 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. ازاین‌رو می‌توان نتیجه گرفت استفاده از داده‌های دو سال قبل برای پیش‌بینی تقلب در صورت‌های مالی عملکرد بهتری خواهد داشت.

به‌طورکلی، نتایج این پژوهش با نتایج پژوهش‌های پرسونس (۱۹۹۵)، فروز و همکاران (۲۰۰۰)، سپاتیس و همکاران (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. همچنین، نتایج حاصل از انتخاب ویژگی به روش مبتنی بر همبستگی برای دو سال قبل از وقوع تقلب حاکی از سودمندی متغیرهای نسبت پوشش بهره، نسبت حساب‌های دریافتنی به کل دارایی‌ها، نسبت موجودی کالا به فروش خالص، نسبت نقدی، لگاریتم طبیعی فروش، نسبت سود خالص به فروش و نسبت جمع دارایی‌های جاری به کل دارایی‌ها برای کشف تقلب داشت. بنابراین می‌توان نتیجه گرفت که عوامل خطر تقلب مرتبط با اهرم مالی، سودآوری، ترکیب دارایی‌ها، نقدینگی و اندازه شرکت از اهمیت قابل‌توجهی در پیش‌بینی تقلب در صورت‌های مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران برخوردار هستند.

۶- نتیجه‌گیری و پیشنهادها

دو جریان متفاوت در متون پژوهشی، پیشینه و مبنایی برای این پژوهش فراهم می‌کند؛ یکی پژوهش‌های انجام‌شده 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 انجام‌شده است. ازاین‌رو، این پژوهش در تلاش است تا با معرفی رویکردهای داده‌کاوی کاراتر برای طبقه‌بندی صورت‌های مالی متقلبانه، متون پژوهشی مربوطه را بهبود بخشد. بنابراین، 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 تکنیک‌های شبکه عصبی مصنوعی، شبکه بیزین و الگوریتم جنگل تصادفی استفاده گردید. 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 الگوریتم جنگل تصادفی و شبکه بیزین در مقایسه با شبکه عصبی مصنوعی برای کشف تقلب در صورت‌های مالی بود. همچنین، نتایج آزمون من‌ویتنی (آزمون U) نشان می‌دهد داده‌های دو سال قبل برای پیش‌بینی تقلب در صورت‌های مالی مناسب‌تر از داده‌های یک سال قبل است. با توجه به ماهیت متفاوت فعالیت شرکت‌های بیمه، بانک‌ها و موسسه‌های اعتباری و واسطه‌گری‌های مالی این‌گونه شرکت‌ها از جامعه موردمطالعه حذف‌شده‌اند و بنابراین نتایج پژوهش قابل‌تعمیم به‌تمامی شرکت‌ها نیست. بنابراین، بررسی علائم خطر مرتبط با تقلب در صورت‌های مالی و همچنین ارائه مدلی برای کشف صورت‌های مالی متقلبانه این شرکت‌ها و مؤسسات می‌تواند یک موضوع پژوهشی بالقوه باشد. همچنین، به پژوهشگران آینده پیشنهاد می‌شود با استفاده از سایر تکنیک‌های داده‌کاوی برای خوشه‌بندی، انتخاب ویژگی و طبقه‌بندی، به بررسی و تائید 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 می‌باشد.

عدم دسترسی به داده‌ها و اطلاعات مالی شرکت‌ها در بلندمدت سبب گردید تا نتایج پژوهش تنها محدود به دوره زمانی کوتاهی باشد و بنابراین نتایج آن قابل‌تعمیم به دوره‌های زمانی بلندمدت نخواهد بود.

۷- منابع

اعتمادی، حسین و حسن زلقی. (۱۳۹۲). کاربرد رگرسیون لجستیک در شناسایی گزارشگری مالی متقلبانه، **دانش حسابرسی** ۱۳(۵۱)، ۱۶۳-۱۴۵.

پورحیدری، امید؛ مجتبی صفی‌پور افشار؛ علی گودرز تله جردی و معصومه صفی‌پور افشار. (۱۳۹۴). بررسی تاثیر کیفیت حسابرسی بر هزینه حسابرسی و قیمت‌گذاری کم‌تر از واقع در عرضه‌های اولیه، **فصلنامه حسابداری مالی** ۷(۲۶)، ۵۱-۳۱.

جهانشاد، آزیتا و سپیده سرداری‌زاده. (۱۳۹۳). رابطه معیار مالی (اختلاف رشد درآمد) و معیار غیر مالی (رشد تعداد کارکنان) با گزارشگری مالی متقلبانه، **پژوهش حسابداری**. ۴(۱۳)، ۱۹۸-۱۸۱.

راعی، رضا؛ پرویز حسن‌زاده و محسن حمشی. (۱۳۹۳). بررسی تاثیر نسبت‌های مالی بر چسبندگی هزینه‌ها، **فصلنامه حسابداری مالی** ۶(۲۲)، ۴۴-۲۶.

صفرزاده، محمدحسین. (۱۳۸۹). توانایی نسبت‌های مالی در کشف تقلب در گزار شگری مالی: تحلیل لاجیت، **دانش حسابداری** ۱(۱)، ۱۶۳-۱۳۷.

فرقاندوست حقیقی، کامبیز؛ سید عباس هاشمی و امین فروغی دهکردی. (۱۳۹۳). مطالعه رابطه بین مدیریت سود و امکان تقلب در صورت‌های مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران، **دانش حسابرسی** ۱۴(۵۶)، ۶۸-۴۷.

کمیته تدوین استانداردهای حسابرسی. (۱۳۹۴). **اصول و ضوابط حسابداری و حسابرسی: استانداردهای حسابرسی**. چاپ بیست و پنجم، تهران: سازمان حسابرسی.

مشایخی، بیتا و امیرحسین حسین‌پور. (۱۳۹۵). بررسی رابطه بین مدیریت سود واقعی و مدیریت سود تعهدی در شرکت‌های مشکوک به تقلب بورس اوراق بهادار تهران، **مطالعات تجربی حسابداری مالی** ۱۲(۴۹)، ۵۲-۲۹.

Alden, M.E., D.M. Bryan, B.J. Lessley, and A. Tripathy. (2012). Detection of financial statement fraud using evolutionary algorithms. **Journal of Emerging Technologies in Accounting**, (9): 71-94.

Almeida, M.P.S.B. (2010). **Classification for fraud detection with social network analysis**. Dissertation, Engenharia Informática e de Computadores.

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., K.L. Jones, and M.F. Zimbelman. (2009). Using nonfinancial measures to assess fraud risk. **Journal of Accounting Research**, (47): 1135-1166.

Carcello, J.V., and A.L. Nagy. (2004). Audit firm tenure and fraudulent financial reporting. **Auditing: A Journal of Practice & Theory**, (23): 55-69.

Chen, F.H., D.J. Chi, and J.Y. Zhu. (2014). Application of random forest, rough set theory, decision tree and neural network to detect financial statement fraud – taking corporate governance into consideration. **Intelligent Computing Theory**, (8588): 221-234.

Dalnial, H., A. Kamaluddin, Z.M. Sanusi, and K.S. Khairuddin, (2014). Accountability in financial reporting: Detecting fraudulent firms. **Procedia- Social and Behavioral Sciences**, (145): pp. 61-69.

Dua, S., and X. Du. (2011). **Data mining and machine learning in cybersecurity**. 1st edition, Taylor and Francis Group.

Fawcett, T., (2006). An introduction to ROC analysis. **Pattern Recognition Letters**, (27): 861-874.

He, H., Y. Bai, E.A. Garcia, and S. Li. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. **IEE Neural Networks**, pp. 1322-1328.

Hoogs, B., T. Kiehl, C. Lacomb, and D. Senturk. (2007). A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud: Research articles. **International Journal of Intelligent Systems in Accounting, Finance and Management**, (15): 41–56.

Huang, S.H., R.H. Tsaih, and F. Yu. (2014). Topological pattern discovery and feature extraction for fraudulent financial Reporting. **Expert Systems with Applications**, (41): pp. 4360–4372.

Kaminski, K.A., T.S. Wetzel, and L. Guan. (2004). Can financial ratios detect fraudulent financial reporting? **Managerial Auditing Journal**, (19): 15–28.

Kirkos, E., C. Spathis, and Y. Manolopoulos. (2007). Data mining techniques for the detection of fraudulent financial statements. **Expert Systems with Applications**, (32): 995–1003.

Lin, C-H., A-A. Chiu, S.Y. Huang, and D.C. Yen. (2016). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. **Knowledge-Based Systems**, Article in Press.

Okike, E., (2011). T**heory and practice of corporate social responsibility**. 1st edition, Springer Berlin Heidelberg.

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., and J.A. Anderson, (1984). An analysis of the factors

associated with lawsuits against public accountants. **The Accounting Review**, (59): 242-263.

Ravisankar, P., V. Ravi, G.R. Rao, and I. Bose. (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., and R. Riley, (2010). **Financial statement fraud-prevention and detection**. 2nd edition, John Wiley & Sons, Inc.

Segal, S.Y., (2016). Accounting frauds – review of advanced technologies to detect and prevent frauds. **Economics and Business Review**, 16 (4): 45–64.

Spathis, C.T., (2002). Detecting false financial statements using published data: Some evidence from Greece. **Managerial Auditing Journal**, (17): 179–191.

Spathis, C., M. Doumpos, and C. Zopounidis. (2002). Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. **European Accounting Review**, 11(3): 509-535.

Sreejesh, S., M.R. Anusree, and S. Mohapatra. (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., and J.T. Sweeney. (1998) Fraudulently misstated financial statements and insider trading: An empirical analysis. **Accounting Review**, (73): 131–46.

Whiting, D,G., J.V. Hansen, J,B. Mcdonald, C. Albrecht, and W. S. Albrecht. (2012). Machine learning methods for detecting patterns of management fraud. **Computational Intelligence**, (28): 505-527.

Zhou, W., and G. Kapoor. (2011). Detecting evolutionary financial statement fraud. **Decision Support Systems**, (50): 570-575.

Zikmund, W.G., B.J. Babin, J.C. Car, and M. Griffin. (2010). **Business research methods**. 8th edition, Cengage Learning.