



کشف تقلب صورت‌های مالی با توجه به گزارش حسابرسی صورت‌های مالی

مهدی رضائی^۱

مهدی ناظمی اردکانی^۲

علیرضا ناصر صدرآبادی^۳

تاریخ پذیرش: ۹۸/۰۷/۲۰

تاریخ دریافت: ۹۸/۰۳/۲۵

چکیده

هدف اصلی این مقاله کشف تقلب صورت‌های مالی با توجه به گزارش حسابرسی صورت‌های مالی است. داده‌های اولیه مورد بررسی در این پژوهش، مربوط به نمونه آماری با حجم ۱۶۴ شرکت پذیرفته شده در بورس اوراق بهادار تهران طی مقطع زمانی ۱۳۹۶-۱۳۹۳ می‌باشد، که به روش نمونه‌گیری حذف سیستماتیک گزینش شده‌اند. نمونه آماری پژوهش با توجه به نوع گزارش حسابرسی به دو گروه مجزا شامل شرکت‌های متقلب (۱) و شرکت‌های غیرمتقلب (۰) دسته‌بندی شده‌اند. متغیرهای مستقل تاثیر گذار بر تقلب در این پژوهش در برگیرنده ۴۱ متغیر مالی و غیر مالی می‌باشد که بر اساس مبانی نظری و پیشینه پژوهش انتخاب شده‌اند. در نهایت داده‌های مربوط به متغیرها که به روش کتابخانه‌ای جمع‌آوری گردیده است، به پنج تکنیک برتر از بین تکنیک‌های یادگیری ماشین داده شده است، که این تکنیک‌ها شامل؛ شبکه‌های بی‌زین، درخت تصمیم، شبکه‌های عصبی، ماشین بردار پشتیبان و روش ترکیبی می‌باشد. پس از بکارگیری این تکنیک‌ها در کشف تقلب صورت‌های مالی، نتایج نشان داد تمامی تکنیک‌ها قابلیت کشف تقلب صورت‌های مالی را در سطح نسبتاً بالایی دارند و تکنیک پیشنهادی ترکیبی با میزان نرخ پیش‌بینی ۹۶٫۲٪ دارای دقت و توان ارزیابی بالاتری نسبت به سایر تکنیک‌ها است.

واژه‌های کلیدی: تقلب صورت‌های مالی، گزارش حسابرسی صورت‌های مالی، یادگیری ماشین.

۱- دانشجوی دکتری حسابداری، واحد یزد، دانشگاه آزاد اسلامی، یزد، ایران. Meh.rezaie@gmail.com

۲- استادیار گروه حسابداری، دانشگاه یزد، یزد، ایران. (نویسنده مسئول) Nazemi@yazd.ac.ir

۳- استادیار گروه مدیریت صنعتی، دانشگاه یزد، یزد، ایران. alireza_naser@yazd.ac.ir

۱- مقدمه

رشد کمی و پیچیدگی های روز افزون فعالیت های اقتصادی در دهه های اخیر موجب شده است که اطلاعات مالی در ارزیابی واحد های اقتصادی نقش به سزایی داشته باشد، به طوری که وجود اطلاعات مالی قابل اتکا لازمه بقای جامعه امروزی است (ستایش و همکاران، ۱۳۹۱). حسابرسی در این میان نقش خود را برای افزایش قابلیت اتکا اطلاعات مالی پیدا کرده است. از دیدگاه استفاده کنندگان به ویژه سهامداران، حسابرسی زمانی سودمند است که حسابرسان طی رسیدگی های خود به بررسی و ارزیابی برقراری تداوم فعالیت، عاری بودن صورت های مالی از تقلب و تخلف با اهمیت و اعمال غیر قانونی مؤثر بر فعالیت واحد مورد رسیدگی بپردازند (امینی و همکاران، ۱۳۹۰). بنابراین تقلب یکی از مصادیق با اهمیت رسیدگی حسابرس در طول فرآیند حسابرسی می باشد. تقلب مالی هم بر عملکرد صنایع مختلف و هم بر زندگی روزمره بشر تاثیرگذار است. تقلب می تواند اعتماد به صنعت و کسب و کار را کاهش داده باعث بی ثباتی در سپرده گذاری مردم و در نتیجه تحمیل هزینه اضافی بر زندگی گردد (سادگالی و همکاران، ۱، ۲۰۱۹). در محیط تجاری مدرن امروزی با پیچیده تر شدن سیستم ها و فعالیت ها و ازدیاد اطلاعات فریبکارانه، فناوری های مورد استفاده برای جلوگیری و کشف تقلب نیز به روز شده است؛ بنابراین توسعه و رشد روش های مربوط به داده های متنوع از جمله داده کاوی برای تشخیص تقلب های مالی در اولویت می باشد (جن چن و همکاران، ۲، ۲۰۱۸). در داخل کشور به موضوع کشف تقلب صورت های مالی توجه زیادی نشده است. مسئله گزارشگری مالی متقلبانه در ایران از اهمیتی ویژه برخوردار است. افزایش تعداد شرکتهای پذیرفته شده در بورس که به منظور جذب منابع مالی به انتشار اوراق بهادار اقدام می کنند، تلاش به منظور کاهش مالیات بر درآمد و ... از جمله دلایل اهمیت این موضوع است (تاراسی و همکاران، ۱۳۹۸). بنابراین در حوزه کشف تقلب هرچقدر حسابرسان به تکنیک های کارآمدتری از جمله

تکنیک های یادگیری ماشین (ML)^۳، دسترسی داشته باشند قاعدتاً، توانایی و کارایی آنها در حوزه تائیدپذیری اطلاعات صورت های مالی بالاتر خواهد رفت. در این راستا این مطالعه، با استفاده از تکنیک های یادگیری ماشین بعنوان یکی از تکنیک های نوظهور در این زمینه به بررسی کشف تقلب صورت های مالی، پرداخته است.

۲- مبانی نظری و پیشینه پژوهش

• تقلب در گزارشگری مالی

تقلب یک فعالیت بزهکارانه و فریبکارانه با قصد منفعت مالی و سایر منافع می باشد، همچنین تقلب بعنوان یک عمل بزهکارانه در برگیرنده حقه بازی، حيله گری و رفتار غیر منصفانه بوسیله یک شخص فریبکار و متقلب است (جاویر و همکاران، ۴، ۲۰۱۸). طبق بند ۴ استاندارد حسابرسی شماره ۴۵۰ ایران، تحریف شامل هرگونه مغایرت بین مبلغ، نحوه طبقه بندی، ارائه یا افشای یک قلم گزارش شده در صورت های مالی، طبق الزامات افشاء استانداردهای حسابداری است. طبق بند ۴ این استاندارد؛ تقلب عبارتست از هرگونه اقدام عمدی یا فریبکارانه یک یا چند نفر از مدیران یا اشخاص ثالث برای برخورداری از یک مزیتی ناروا یا غیر قانونی (کمیته تدوین استانداردهای حسابرسی، ۱۳۸۸).

تقلب در گزارشگری مالی: عبارت است از تحریف عمدی در نتایج صورت های مالی به منظور ارائه تصویر نادرست از واحد تجاری (وکیلی فرد و همکاران، ۱۳۸۸)؛

انگیزه های زیادی برای تقلب گزارشگری مالی وجود دارد. نمونه هایی از آن بدین شرح می باشد: پاداش مبتنی بر سود گزارش شده (واتز و زیمرمن^۵، ۱۹۸۶)، حفظ یا افزایش قیمت بازار سهام (دیچو و همکاران^۶، ۱۹۹۵)، دستیابی به پیش بینی های داخلی و خارجی (فاما و جنسن^۷، ۱۹۸۳)، حداقل کردن بدهی مالیاتی (دچو و همکاران، ۱۹۹۶)، جلوگیری از تخلف در قراردادهای بدهی (اسپاتیس^۸،

۲۰۰۲)، تأمین مالی به ارزانترین شکل ممکن (کوتسیان‌تیس^۹، ۲۰۰۶).

• یادگیری ماشین (ML)

یادگیری ماشین زمینه تحقیقاتی نسبتاً جدیدی از هوش مصنوعی است که در حال حاضر دوران رشد و تکامل خود را می‌گذراند و زمینه‌ای بسیار فعال در علوم کامپیوتر می‌باشد. علوم مختلفی در ارتباط با یادگیری ماشین هستند. این تکنیک‌ها قابلیت به کارگیری در زمینه‌های مختلفی از قبیل؛ کشف تقلب، کشف ورود بی‌اجازه، تشخیص اشتباهات پزشکی، لایروبی داده‌ها، جلوگیری از خطا و ... را دارا می‌باشند (دومینگس و همکاران^{۱۰}، ۲۰۱۹). در این پژوهش از چهار تکنیک معروف و از قوی‌ترین تکنیک‌های یادگیری ماشین شامل شبکه‌های بیزین (KNNs)^{۱۱}، درخت تصمیم (TREE)^{۱۲}، شبکه‌های عصبی (ANNs)^{۱۳}، ماشین بردار پشتیبان (SVM)^{۱۴} استفاده می‌شود.

- شبکه بیزین یک مدل گرافیکی است که متغیرهای موجود در یک مجموعه داده را که معمولاً گره نامیده می‌شوند؛ همراه با روابط احتمالی و شرطی بین آن‌ها نشان می‌دهند. روابط سببی بین گره‌ها می‌تواند به وسیله شبکه بیزین نمایش داده شود، البته پیوندهای موجود در شبکه (کمان‌ها) لزوماً ارتباط مستقیم علت معلولی را نشان نمی‌دهند. این شبکه‌ها در مقابل داده‌های مفقود بسیار مقاومند و بهترین پیش‌بینی را ارائه می‌دهند (علیزاده و ملک محمدی، ۱۳۹۳). یکی از سودمندترین مزایای شبکه‌های بیزین این است که جهت آنالیز، چیزی به عنوان اندازه حداقل برای داده‌های نمونه، وجود ندارد و این شبکه‌ها در محاسبات خود از تمامی داده‌های موجود استفاده می‌کنند و حتی با وجود نمونه‌هایی با اندازه‌ای نسبتاً کوچک می‌توان به پیش‌بینی‌های نسبتاً خوبی دست یافت (خدمتگزار سالانقوچ، ۱۳۹۲). این ساختارهای

گرافیکی برای نشان دادن اطلاعات در یک حوزه دارای عدم قطعیت به کار می‌روند. از جمله روش‌های معروف بیزین روش شبکه باور بیزین و روش بیز ساده می‌باشد (نعمتی و همکاران، ۱۳۹۲). لازم به ذکر است در این پژوهش از روش باور بیزین استفاده شده است. - درخت تصمیم یک ساختار درختی مانند فلوچارت دارد. در این تکنیک، پیش‌بینی به درست آمده از درخت در قالب یکسری قواعد توضیح داده می‌شود (غضنفری، علیزاده و تیمورپور، ۱۳۸۸). یکی از مزایای آن این است که بعد از ساخت درخت به راحتی می‌توان علت استنتاج قواعد بدست آمده را مشاهده نمود بدین معنا که درخت تصمیم همانند شبکه عصبی مانند یک جعبه سیاه عمل نمی‌کند؛ و منطق کاری آن روشن و آشکار است (علیزاده و ملک محمدی، ۱۳۹۳). از جمله مزیت‌های این روش مستقل بودن آن از چگونگی توزیع داده‌ها و وابستگی متغیرهای ورودی می‌باشد. درخت‌های تصمیم به علت دقت مناسب و همچنین ساده بودن و فهم راحت آن، یکی از محبوبترین روش‌های طبقه بندی هستند (حاجی حیدری و خاکباز، ۱۳۹۲). مدل‌های حاصل از درخت تصمیم دارای الگوریتم متفاوتی مثل CHAID، CART، C5.0 می‌باشد (کیقبادی و همکاران، ۱۳۹۴). در این پژوهش بدلیل توانمندتر بودن تکنیک CART از این تکنیک جهت طبقه بندی شرکت‌های متقلب و غیر متقلب استفاده شده است.

- ماشین بردار پشتیبان از روش‌های یادگیری تحت نظارت است که برای دسته بندی و رگرسیون استفاده می‌شود. ماشین بردار پشتیبان توسط واپنیک بر اساس تئوری یادگیری آماری معرفی شده است (واپنیک، ۱۹۹۵). در این روش اساس کار بر این است که بردارهای ورودی (بردارهای

بالای آن در انجام پردازش موازی و غیر خطی بر روی داده‌ها است که می‌تواند شدت ارتباط متغیرهای مستقل و وابسته را دقیق‌تر از رگرسیون خطی توصیف نماید. به علاوه، این روش نیازی به برقراری فروض کلاسیک رگرسیون ندارد (حجازی و همکاران، ۱۳۹۱).

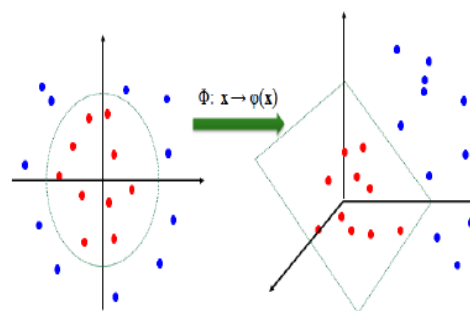
۲-۱- پیشینه پژوهش

سادگالی و همکاران (۲۰۱۹) در مطالعه‌ای به پیشنهاد تکنیک‌های برتر یادگیری ماشین در زمینه کشف و پیشگیری از تقلب مالی پرداختند. در این پژوهش پس از شناسایی تکنیک‌ها و روش‌های مختلف یادگیری ماشین که در کشف تقلب صورت‌های مالی در پژوهش‌های گذشته استفاده شده بودند؛ آنها را از نظر شاخص‌هایی چون دقت، حساسیت، بلادرنگ بودن، بانک اطلاعاتی و مشاهدات مورد مقایسه قرار دادند. نتایج این مطالعه نشان داد که تکنیک‌های کشف تقلب دورگه (هایبردی) نسبت به سایر تکنیک‌ها، عملکرد بهتری دارند بویژه زمانیکه آن‌ها چندین روش کشف تقلب سنتی را ترکیب می‌نمایند.

جن چن و همکاران (۲۰۱۸) در مطالعه‌ای رویکردهای کشف تقلب صورت‌های مالی گروه‌های تجاری را توسعه دادند. در این پژوهش با بکارگیری تکنیک TREE، ANNs، KNNs و SVM به کشف تقلب صورت‌های مالی پرداخته شد و در نهایت مکانیسم‌های کشف تقلبی که دارای دقت و توانایی بالایی در کشف تقلب بودند شناسایی نمودند و توسعه دادند که در بین این تکنیک‌های معرفی شده، تکنیک ماشین بردار پشتیبان با دقت ۹۴٪ و توانایی ۹۴٫۷٪ قابلیت بالاتری در کشف تقلب صورت‌های مالی داشت.

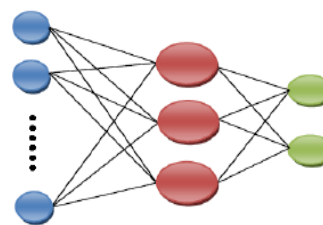
نورماه و همکاران^{۱۷} (۲۰۱۷) در پژوهشی به بررسی پیش‌بینی تقلب در گزارش‌های مالی با استفاده از شبکه عصبی مصنوعی، پرداختند. نتایج نشان می‌دهد که روش پیشنهادی شبکه‌های عصبی مصنوعی، از سایر تکنیک‌های آماری که به طور

ویژگی) به فضایی با ابعاد بالاتر نگاشت می‌شوند، در این فضای جدید دسته‌بندی این بردارها کارتر و مناسب‌تر انجام می‌شود (شکل ۱) همچنین SVM به دلیل انتخاب جداکننده بهینه که بیشترین فاصله را از نمونه‌ها دارد قابلیت تعمیم بالایی دارد.



شکل ۱: نگاشت نمونه‌ها به فضای با ابعاد بالاتر در فرآیند تولید مدل در روش SVM

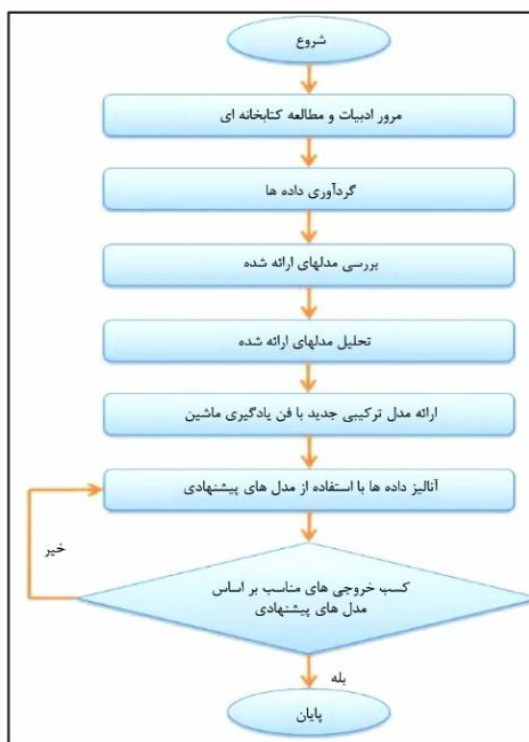
- شبکه‌های عصبی مصنوعی چکیده‌ای از مغز انسان است. شبکه عصبی مصنوعی، مشابه مغز، از نورون‌های مصنوعی و اتصالات ساخته شده است. ورودی شبکه عصبی بردارهای ویژگی و خروجی شبکه معرف طبقات مختلف خواهد بود. شبکه‌های چندلایه پیش‌خور یکی از مهمترین و رایج‌ترین شبکه‌های عصبی در کاربردهای دنیای واقعی هستند (چی و هانگ، ۱۶، ۲۰۰۸).



شکل ۲: نمونه‌ای از ANN با یک لایه مخفی
لایه ورودی لایه مخفی لایه خروجی

در این پژوهش نیز از شبکه عصبی چندلایه پیش‌خور با الگوریتم پیش‌انتشار استفاده شده است. علت استفاده از تکنیک شبکه عصبی به این روش، توان

های شبه آزمایشی قرار می‌گیرد. برای جمع آوری اطلاعات مربوط به مبانی نظری و پیشینه پژوهش و همچنین داده‌های لازم برای اجرای تکنیک‌ها از روش گردآوری کتابخانه‌ای و ابزارگردآوری اسنادکاوی استفاده شده است. داده‌های جمع‌آوری شده با استفاده از نرم‌افزار اکسل تلخیص و سپس متغیرهای مورد نظر محاسبه شده‌اند. در نهایت از نرم‌افزارهای Rapidminer 7.3 و matlab برای سنجش توانایی تکنیک‌های یادگیری ماشین در کشف تقلب صورت‌های مالی استفاده شده است. طرح کلی پژوهش بدین شرح تدوین شده است:



شکل ۳- مراحل انجام این پژوهش

در ابتدای امر هدف، پیدا کردن متغیرهای تأثیرگذار در خصوص سنوات گذشته شرکت‌ها است که منجر به رسیدگی و تنظیم صورت مالی آن‌ها می‌شود.

جامعه آماری پژوهش شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران طی سال‌های ۱۳۹۳ تا ۱۳۹۶ می‌باشد. در این پژوهش با توجه به بازه زمانی سال ۱۳۹۳ تا ۱۳۹۶ و شرایط انتخاب متغیرها، از

گسترده‌ای برای پیش‌بینی گزارش‌های مالی متقلبانه استفاده شده است، عملکرد مناسب‌تری دارد.

چی‌چن و همکاران^{۱۸} (۲۰۱۵) به تجزیه و تحلیل تفاوت‌های میان تکنیک‌های داده‌کاوی و خبره‌سنجی در کشف تقلب صورت‌های مالی پرداختند. روش‌های تجزیه و تحلیل داده‌ها به کار برده شده در این پژوهش شامل رگرسیون منطقی، CART، شبکه‌های ANNs بوده است. نتایج این پژوهش بیانگر کارایی و توانایی بالاتر تکنیک‌های داده‌کاوی نسبت به خبره‌سنجی است.

تاراسی و همکاران (۱۳۹۸) با استفاده از تکنیک شبکه‌های عصبی به کشف تقلب صورت‌های مالی پرداختند. نتایج این پژوهش، نشان داد که تکنیک شبکه عصبی و مدل‌سازی مبتنی بر شبکه عصبی از صحت ۹۷٫۴ درصد برخوردار است و با طراحی و آموزش دقیق، می‌توان شبکه عصبی طراحی‌شده را با دقت قابل قبولی بتواند احتمال گزارشگری مالی متقلبانه شرکت را کشف و پیش‌بینی نمایند. عظیمی و نورحسینی (۱۳۹۶) به کشف و پیش‌بینی تقلب مالی در بانک‌ها با استفاده از روش‌های داده‌کاوی، پرداختند. نتایج خروجی نشان داد که داده‌کاوی در حدود ۸۳٪ پیش‌بینی درستی در خروجی نمایش می‌دهد.

رهنمای رودپشتی (۱۳۹۱) اثربخشی تکنیک‌های داده‌کاوی در تشخیص رفتارهای متقلبانه شرکت‌هایی که صورت‌های مالی متقلبانه گزارش نموده‌اند را بررسی کرده تا عوامل موثر بر اینگونه رفتارها را شناسایی نماید. نتایج مطالعه نشان می‌دهد که اولاً، تکنیک‌های داده‌کاوی، در شناسایی صورت‌های مالی متقلبانه سودمند هستند. ثانیاً داده‌کاوی، به عنوان کانون هدایت فکر در مدیریت تجارت جهت کشف تقلب می‌تواند مورد توجه قرار گیرد.

۳- روش شناسایی پژوهش

این پژوهش از لحاظ هدف کاربردی و از نوع پژوهش‌های تجربی است و با توجه به اینکه از اطلاعات تاریخی استفاده شده است در دسته پژوهش

۳-۱- اندازه گیری تقلب

تقلب صورت‌های مالی بعنوان متغیر وابسته (هدف) این پژوهش می باشد، که برای اندازه گیری آن در پژوهش های مختلف معیارهای متفاوتی ملاک عمل قرار گرفته است. در این پژوهش برای اندازه گیری تقلب صورت‌های مالی و تفکیک شرکت های متقلب از غیر متقلب از شیوه اندازه گیری تقلب توسط فرکاندوست حقیقی و همکاران (۱۳۹۳)، اعتمادی و عبدلی (۱۳۹۶) غلامی مقدم و همکاران (۱۳۹۷)، استفاده شده است. در این شیوه فرض گردیده که شرکت هایی که گزارش حسابرسی آن ها مردود، عدم اظهار نظر یا مشروط بوده، با احتمال بیشتری، نسبت به شرکت هایی که گزارش مقبول دارند، مرتکب تقلب شده اند. بنابراین در این پژوهش، گزارش مقبول حسابرسی نسبت به صورت‌های مالی به منزله غیر متقلب بودن شرکت و گزارش غیر مقبول نسبت به صورت‌های مالی به منزله متقلب بودن شرکت می باشد.

۳-۲- متغیرهای مستقل

متغیر های مستقل این پژوهش که به عنوان متغیرهای پیش بین و تاثیرگذار بر تقلب صورت‌های مالی می باشند، شامل؛ ۴۱ نسبت و معیار مالی و غیر مالی بوده که بر اساس مطالعه مبانی نظری و پیشینه پژوهش گزینش شده اند. بررسی ها و تجربه نشان داده که این نسبت ها و معیارها رابطه معنادارتری با تقلب دارند. متغیرهای مستقل پژوهش در دو دسته کلی مالی و غیر مالی عبارتند از:

متغیرهای مالی

- ۱- سود پس از مالیات به فروش ۲- سود قبل از مالیات و پس از بهره به فروش ۳- نسبت جاری ۴- نسبت کل بدهی به دارایی ها ۵- اندازه شرکت ۶- نسبت حساب های دریافتی به کل دارایی ها ۷- دفعات گردش موجودی کالا ۸-نسبت آئی ۹-شاخص ورشکستگی ۱۰- لگاریتم طبیعی فروش خالص ۱۱- لگاریتم طبیعی تعداد کارکنان ۱۲- نسبت پرداخت بدهی ها ۱۳- سرمایه در گردش ۱۴- سرانه دارایی ها

روش نمونه گیری حذف سیستماتیک برای رسیدن به نمونه استفاده گردیده که محدودیت های اعمال شده برای انتخاب نمونه عبارتند از :

- ۱) جزء شرکت های سرمایه گذاری ، بانک ها و واسطه گری مالی، هلدینگ ها، لیزینگ ها و بیمه ها نباشند.
- ۲) پایان سال مالی آن ها پایان اسفند باشد.
- ۳) نماد معاملاتی شرکت فعال و سهام آن ها حداقل یکبار در سال معامله شده باشد.
- ۴) اطلاعات مالی آنها در بازه زمانی پژوهش، در دسترس باشد.

با توجه به محدودیت های تعیین شده، ۱۶۴ شرکت بعنوان نمونه استخراج گردیده است. بعبارت دیگر ۶۵۶ سال شرکت مورد آزمون قرار گرفت.

در تکنیک های طبقه بندی یادگیری ماشین جهت ساخت بردار ویژگی ها، اطلاعات صورت‌های مالی شرکت ها و فاکتورهای با اهمیت غیر مالی مرتبط با تقلب در نظر گرفته شده است. بر اساس ویژگی های مالی و غیر مالی شرکت ها، این رویکرد منجر به شکل گیری شاخصی شد که نشان دهنده رفتار تقلب و فاقد تقلب در طول چندین دوره بوده که بر اساس سابقه عملکرد، ویژگی های خاص شرکت ها و اطلاعات تولید شده در مراحل مختلف می باشد. جهت نرمال نمودن پراکندگی داده ها طی یک فرآیند کلی، مواردی از داده ها که بالاتر از پنج برابر میانگین انحراف معیار بودند، به عنوان مقادیر خارج از محدوده شناخته شدند و از نمونه آماری حذف و الباقی، جهت مطالعه باقی ماندند. علاوه بر این، قبل از انتخاب متغیرها جهت استفاده در مدل های رفتاری، لازم بود آن ها را تجزیه و تحلیل نموده تا به متغیرهای تأثیرگذار تقلیل یابد. درنهایت، مؤلفه های اصلی شرکت های انتخاب شده به ترتیب شامل ساختار سرمایه، سودآوری ، عملکرد، نقدینگی، جریان نقد عملیاتی و معیارهای غیرمالی را در برمی گیرد.

صورت‌های مالی در سال ۱۳۹۶ با استفاده از تکنیک های یادگیری ماشین پرداخته شده است. انتخاب تکنیک های برتر یادگیری ماشین شامل: ماشین بردار پشتیبان، شبکه های عصبی، درخت تصمیم، شبکه های بیزین و روش ترکیبی (combine) براساس مبانی نظری و پیشینه صورت گرفته است.

نتایج اجرای تکنیک های یادگیری ماشین به منظور کشف تقلب صورت‌های مالی

برای انجام داده کاوی ابتدا تکنیک‌های یادگیری ماشین آموزش داده شده است. به این صورت که یک بخش از داده‌ها به‌عنوان داده آموزشی (test) در نرم افزار های داده کاوی به تکنیک‌های مورد استفاده در این پژوهش داده شده است؛ تا میزان درصد آموزش مدل‌ها مشخص شود؛ در نهایت داده‌های جدید بعنوان داده های آزمایشی (train) در سیستم نرم افزار اعمال گردیده تا ارزیابی اصلی تکنیک‌های یادگیری ماشین صورت گرفته و میزان تقلب صورت‌های مالی بررسی و ارزیابی گردد، که در ادامه نشان داده شده است. برای مقایسه و ارزیابی تکنیک‌ها از شش شاخص ۱-صحت ۲-دقت ۳-حساسیت ۴-تشخیص ۵- نرخ درست مثبت (TP) ۶- بازخوانی، استفاده شده تا توانایی روش‌ها به‌خوبی نمایش داده شود در ادامه به ارائه نتایج حاصل از اجرای تکنیک‌ها پرداخته شده است.

۱۵- سرانه فروش ۱۶- سرانه سودخالص ۱۷- گردش حساب های دریافتی ۱۸- دوره وصول مطالبات ۱۹- گردش خالص دارایی ها ۲۰- گردش خالص دارایی های ثابت ۲۱- سود قبل از بهره و مالیات ۲۲- وجه نقد ناشی از فعالیت های عملیاتی به فروش ۲۳- بازده حقوق صاحبان سهام (ROE) ۲۴- نسبت ارزش ویژه به بدهی بلندمدت ۲۵- بازده کل دارایی ها ۲۶- ROI ۲۷- دوره گردش موجودی کالا ۲۸- ذخیره مالیات به فروش ۲۹- سرانه ذخیره مزایای پایان خدمت کارکنان ۳۰- سود (زیان) انباشته به فروش ۳۱- درصد تغییرات در سودآوری ۳۲- سود ناخالص به فروش ۳۳- افزایش یا کاهش در وجه نقد ۳۴- دارایی های جاری ۳۵- بدهی های جاری ۳۶- بدهی های بلندمدت ۳۷- وجه نقد ناشی از فعالیت های سرمایه گذاری به فروش.

متغیرهای غیر مالی

۱- نوع موسسه حسابرسی ۲- دعاوی حقوقی ۳- نوع عملکرد ۴- غیرمقبول بودن گزارش حسابرسی سال قبل

۳-۳- شیوه اجرا پژوهش

در این پژوهش برای سال ۱۳۹۶ از ۱۶۴ شرکت نمونه (بر اساس شیوه اندازه گیری ذکر شده در قسمت اندازه گیری متغیر تقلب) ۸۶ شرکت متقلب و ۷۸ شرکت غیرمتقلب بصورت متغیر دامی (صفر و ۱) مشخص شده است. سپس با جمع اوری دیتاهای مربوط به متغیرهای پیش بین تقلب، برای سال های ۱۳۹۳ الی ۱۳۹۵ به کشف و پیش بینی تقلب

جدول ۱- خروجی درصد ارزیابی‌ها برای داده‌های آموزش به ازای ۶ ارزیابی و ۵ روش

Classifier	صحت	دقت	حساسیت	تشخیص	نرخ TP	بازخوانی
combine	۹۸,۵	۹۷,۵	۹۸,۶	۹۸,۷	۹۸,۸۹	۹۶,۶
ANN	۹۷,۵	۹۶,۴	۹۷,۴	۹۷,۶	۹۷,۸۵	۹۵,۵
TREE	۹۶,۱	۹۵,۱	۹۶,۴	۹۶,۱	۹۶,۴۱	۹۴,۳
KNN	۹۴,۶	۹۲,۴	۹۳,۲	۹۲,۳	۹۴,۵۶	۹۱,۳
svm	۹۳,۴	۹۱,۳	۹۰,۳	۸۹,۹	۹۲,۲۲	۹۰,۱۲

جدول ۲- خروجی درصد ارزیابی‌ها برای داده‌های تست به ازای ۶ ارزیابی و ۵ روش

Classifier	صحت	دقت	حساسیت	تشخیص	نرخ TP	بازخوانی
combine	۹۸,۸	۹۷,۳	۹۹,۶	۹۷,۶	۹۹,۸۹	۹۴,۳
ANN	۹۶,۳	۹۵,۶	۹۶,۴	۹۵,۳۶	۹۶,۹۵	۹۰,۴
TREE	۹۳,۳	۹۰,۳	۹۴,۶	۹۴,۳	۹۴,۲۳	۸۹,۶
KNN	۹۰,۳۶	۸۹,۵۶	۸۸,۵۶	۸۸,۳۶	۸۹,۸۸	۸۵,۳
svm	۸۹,۶۸	۸۵,۵	۸۹,۳۶	۸۶,۳	۸۶,۵۴	۸۰,۲

جدول ۳- خروجی درصد ارزیابی‌ها برای داده‌های آموزش به ازای ۶ ارزیابی و ۵ روش با ویژگی‌های کم شده

Classifier	صحت	دقت	حساسیت	تشخیص	نرخ TP	بازخوانی
combine	۹۰,۳۶	۹۰,۱	۹۰,۴	۹۰,۴	۹۰,۴۵	۸۹,۳
ANN	۸۸,۳	۸۷,۶	۸۱,۳	۸۷,۳	۸۱,۴۵	۸۱,۳
TREE	۸۱,۲	۷۸,۲	۸۱,۲	۸۰,۲	۷۸,۸۹	۷۸,۳
KNN	۷۵,۴	۷۵,۲	۷۱,۲	۷۰,۳	۷۱,۲۵	۷۲,۳
svm	۷۰,۴	۶۹,۶	۶۶,۴	۶۹,۵	۶۹,۸۱	۶۸,۶

جدول ۴- خروجی درصد ارزیابی‌ها برای داده‌های تست به ازای ۶ ارزیابی و ۵ روش با ویژگی‌های کم شده

Classifier	صحت	دقت	حساسیت	تشخیص	نرخ TP	بازخوانی
combine	۸۴,۶	۸۱,۶	۸۴,۹	۸۳,۵	۸۴,۴۱	۸۱,۳
ANN	۷۸,۳۶	۷۲,۳	۷۴,۳	۷۵,۶	۷۵,۴۵	۷۶,۳
TREE	۷۰,۳	۶۹,۵	۷۱,۲	۷۰,۸	۷۰,۶۹	۷۱,۳
KNN	۶۷,۶	۶۲,۲	۶۹,۵	۶۴,۶	۶۴,۸۷	۶۴,۶
svm	۶۰,۶	۵۹,۵	۶۰,۳	۵۸,۶	۵۹,۸۵	۵۷,۳

نتیجه مقایسه مقادیر پیش بینی شده و واقعی در ماتریس درهم ریختگی پیش می‌آید.

جدول ۵- ماتریس درهم ریختگی خام

مقادیر واقعی		N	
منفی	مثبت	مثبت	منفی
نادرست مثبت FP خطای نوع اول	درست مثبت TP	مثبت	مقادیر پیش بینی شده
درست منفی TN	نادرست منفی FN خطای نوع دوم	منفی	

در حالت کلی به دنبال این هستیم که در این ماتریس پیش بینی‌های نادرست، به حداقل برسد. با توجه به اینکه هدف این پژوهش پیش بینی تقلب می‌باشد؛ اگر پیش بینی مثبت باشد بدین معناست که صورت‌های مالی شرکت متقلبانه بوده است، و اگر پیش بینی منفی باشد، بدین معناست که صورت‌های

در ادامه جداول ارائه شده بیانگر ماتریس درهم ریختگی یا اغتشاش می‌باشد که تعداد سطرها و ستون‌های این ماتریس به تعداد طبقات (class) بستگی دارد؛ که در این پژوهش داده‌ها در برگیرنده دو طبقه شامل شرکت‌های متقلب و شرکت‌های غیر متقلب می‌باشد؛ بنابراین این ماتریس برای تمامی تکنیک‌های مورد استفاده دو در دو ترسیم می‌شود. بعد از ساخت یک مدل طبقه بندی (یعنی یافتن الگوریتمی که با مشاهده یک داده جدید، طبقه آن را مشخص کند)؛ برای سنجش میزان کارایی و دقت مدل پیشنهادی آن را بر روی داده‌های آموزشی (یعنی داده‌هایی که از قبل طبقه بندی آن‌ها مشخص شده است)، اعمال می‌کنیم و نهایتاً در مورد داده‌های آزمایشی (test) نیز اعمال می‌شود. جدول شماره ۵ نشان‌دهنده حالت‌های مختلفی است که در

تکنیک SVM همانطور که در جدول شماره ۶ آمده است؛ براساس داده‌های آزمایشی (test) از بین ۸۶ شرکت متقلب ۵۶ شرکت متقلب و از بین ۷۸ شرکت غیرمتقلب ۶۹ شرکت را بدرستی تشخیص داده است؛ که این موضوع بیانگر این است که تکنیک SVM قادر بوده طبق داده‌های آموزشی ۷۳,۵۸٪ و طبق داده‌های آزمایشی ۶۵,۲۹٪ از شرکت‌های متقلب را پیش‌بینی کند. خطای نوع دوم این تکنیک حدود ۱۱٪ می‌باشد. بنابراین تکنیک SVM با توجه به نرخ دقت ۹۳,۴٪ و قدرت پیش‌بینی و کشف ۸۹,۶۸٪ در تشخیص تقلب قابلیت کشف تقلب در سطح نسبتاً خوبی را داراست.

جدول ۷- نتیجه ماتریس درهم‌ریختگی تکنیک ANNs

مدل طبقه‌بندی			طبقه‌بندی عملی
نرخ پیش‌بینی (%)	داده‌های بدون تقلب	داده‌های تقلب	
مجموعه داده‌های آموزشی (train):			
۸۸,۶	۱۰	۷۶	داده‌های تقلب
۹۷,۳	۷۶	۲	داده‌های بدون تقلب
۹۷,۵	-	-	نرخ دقت کل
مجموعه داده‌های آزمایشی (test):			
۸۶,۹۶	۱۲	۷۴	داده‌های تقلب
۹۷,۴۱	۷۶	۲	داده‌های بدون تقلب
۹۶,۳	-	-	نرخ دقت کل

تکنیک ANNs همانطور که در جدول شماره ۷ آمده است؛ براساس داده‌های آزمایشی (test)، از بین ۸۶ شرکت متقلب ۷۴ شرکت متقلب و از بین ۷۸ شرکت غیرمتقلب ۷۶ شرکت را بدرستی تشخیص داده است؛ که این موضوع بیانگر این است که تکنیک ANNs قادر بوده طبق داده‌های آموزشی ۸۸,۶٪ و طبق داده‌های آزمایشی ۸۶,۹۶٪ از شرکت‌های متقلب را پیش‌بینی کند. خطای نوع دوم این تکنیک حدود ۲,۵٪ می‌باشد. تکنیک ANNs با توجه به نرخ دقت ۹۷,۵٪ و قدرت کشف ۹۶,۳٪ در تشخیص تقلب

مالی شرکت‌های عاری از تقلب بوده است. بنابراین تحلیل ماتریس درهم‌ریختگی برای پژوهش عبارت است از:

- درست مثبت (TP): اگر صورت‌های مالی شرکت در واقع متقلبان باشد و مقدار پیش‌بینی شده، نشانگر متقلبان بودن صورت‌های مالی شرکت باشد.
- نادرست مثبت (FP): اگر صورت‌های مالی شرکت در واقعیت متقلبان نباشد اما مقدار پیش‌بینی شده، نشانگر متقلبان بودن صورت‌های مالی شرکت باشد (خطای نوع اول).
- درست منفی (TN): اگر صورت‌های مالی شرکت در واقع متقلبان نباشد و مقدار پیش‌بینی شده نیز، نشانگر عاری از تقلب بودن گزارشات باشد.
- نادرست منفی (FN): اگر صورت‌های مالی شرکت در واقعیت متقلبان باشد اما مقدار پیش‌بینی شده، نشانگر متقلبان بودن صورت‌های مالی شرکت نباشد (خطای نوع دوم). که این موضوع بااهمیت‌تر بودن خطای نوع دوم را نشان می‌دهد.

جدول ۶- نتیجه ماتریس درهم‌ریختگی تکنیک SVM

مدل طبقه‌بندی			طبقه‌بندی عملی
نرخ پیش‌بینی (%)	داده‌های بدون تقلب	داده‌های تقلب	
مجموعه داده‌های آموزشی (train):			
۷۳,۵۸	۲۳	۶۳	داده‌های تقلب
۸۴,۶	۶۶	۱۲	داده‌های بدون تقلب
۹۳,۴	-	-	نرخ دقت کل
مجموعه داده‌های آزمایشی (test):			
۶۵,۲۹	۳۰	۵۶	داده‌های تقلب
۸۸,۶	۶۹	۹	داده‌های بدون تقلب
۸۹,۶۸	-	-	نرخ دقت کل

خطای نوع دوم این تکنیک حدود ۲٪ می باشد. بنابراین این تکنیک با توجه به نرخ دقت ۹۸٫۵٪ و قدرت کشف ۹۸٫۸٪ در تشخیص تقلب دارای قدرت و قابلیت بالایی می باشد و اینکه تکنیک combine در کشف و پیش بینی تقلب دارای قدرت و قابلیت بالاتری نسبت به سایر تکنیک های مورد استفاده در این پژوهش می باشد.

جدول ۸- نتیجه ماتریس درهم ریختگی تکنیک TREE

مدل طبقه بندی			طبقه بندی عملی
نرخ پیش بینی (%)	داده های بدون تقلب	داده های تقلب	
مجموعه داده های آموزشی (train):			
۸۳٫۹۶	۱۴	۷۲	داده های تقلب
۹۳٫۳	۷۳	۵	داده های بدون تقلب
۹۶٫۱	-	-	نرخ دقت کل
مجموعه داده های آزمایشی (test):			
۸۰٫۴۳	۱۷	۶۹	داده های تقلب
۹۳٫۷۸	۷۳	۵	داده های بدون تقلب
۹۳٫۳	-	-	نرخ دقت کل

جدول ۹- نتیجه ماتریس درهم ریختگی تکنیک KNN

مدل طبقه بندی			طبقه بندی عملی
نرخ پیش بینی (%)	داده های بدون تقلب	داده های تقلب	
مجموعه داده های آموزشی (train):			
۷۵٫۴۳	۲۲	۶۴	داده های تقلب
۹۰٫۴	۷۰	۸	داده های بدون تقلب
۹۴٫۶	-	-	نرخ دقت کل
مجموعه داده های آزمایشی (test):			
۷۱٫۷۴	۲۵	۶۱	داده های تقلب
۹۱٫۱۹	۷۱	۷	داده های بدون تقلب
۹۰٫۳۶	-	-	نرخ دقت کل

دارای قدرت و قابلیت بالایی می باشد و اینکه توان ANNs در کشف تقلب از تکنیک SVM بالاتر است.

تکنیک درخت تصمیم (TREE) همانطور که در جدول شماره ۸ آمده است؛ براساس داده های آزمایشی (test)، از بین ۸۶ شرکت متقلب ۶۹ شرکت متقلب و از بین ۷۸ شرکت غیرمتقلب ۷۳ شرکت را بدرستی تشخیص دهد؛ که این موضوع بیانگر این است که تکنیک TREE قادر بوده طبق داده های آموزشی ۸۳٫۹۶٪ و طبق داده های آزمایشی ۸۰٫۴۳٪ از شرکت های متقلب را پیش بینی کند. خطای نوع دوم این تکنیک حدود ۶٪ می باشد. تکنیک TREE با توجه به نرخ دقت ۹۶٫۱٪ و قدرت کشف ۹۳٫۳٪ در تشخیص تقلب دارای قدرت و قابلیت بالایی می باشد و اینکه توان تکنیک TREE در کشف تقلب از تکنیک ANNs پایین تر ولی از تکنیک SVM بالاتر است.

تکنیک شبکه بیزین (KNN) همانطور که در جدول شماره ۹ آمده است؛ براساس داده های آزمایشی (test)، از بین ۸۶ شرکت متقلب ۶۱ شرکت متقلب و از بین ۷۸ شرکت غیرمتقلب ۷۱ شرکت را بدرستی تشخیص داده است؛ که این موضوع بیانگر این است که تکنیک KNN قادر بوده طبق داده های آموزشی ۷۵٫۴۳٪ و طبق داده های آزمایشی ۷۱٫۷۴٪ از شرکت های متقلب را پیش بینی کند. خطای نوع دوم این تکنیک حدود ۹٪ می باشد. بنابراین این تکنیک با توجه به نرخ دقت ۹۴٫۶٪ و قدرت کشف ۹۰٫۳۶٪ در تشخیص تقلب دارای قدرت و قابلیت بالایی می باشد و اینکه توان تکنیک KNN در کشف تقلب از تکنیک های ANNs و TREE پایین تر ولی از تکنیک SVM بالاتر است.

تکنیک ترکیبی (combine) همانطور که در جدول شماره ۱۰ آمده است؛ براساس داده های آزمایشی (test)، این از بین ۸۶ شرکت متقلب ۷۹ شرکت متقلب و از بین ۷۸ شرکت غیرمتقلب ۷۶ شرکت را بدرستی تشخیص داده است؛ که این موضوع بیانگر این است که تکنیک combine قادر بوده طبق داده های آموزشی ۹۶٫۲٪ و طبق داده های آزمایشی ۹۱٫۸٪ از شرکت های متقلب را پیش بینی کند.

یافته‌های پژوهش حاکی از آن است که مولفه‌های نسبت‌های مالی و سایر متغیرهای مالی و غیر مالی استفاده شده در این پژوهش تخمین زنده‌های توانایی در کشف تقلب صورت‌های مالی می‌باشند. نتایج بدست آمده از این پژوهش با نتایج پژوهش‌های چی چن و همکاران (۲۰۱۵)، تاراسی و همکاران (۱۳۹۷) و جن چن و همکاران (۲۰۱۸) همخوانی دارد به گونه‌ای که تکنیک‌های مشترک استفاده شده در اینگونه مطالعات با این پژوهش از جمله درخت تصمیم، شبکه‌های عصبی و ماشین بردار پشتیبان در حد بالایی توان کشف تقلب صورت‌های مالی را دارد و می‌تواند بوسیله استفاده کنندگان از صورت‌های مالی مورد استفاده قرار گیرد. با توجه به نتایج پژوهش، پیشنهادات کاربردی زیر ارائه می‌گردد:

- ۱) با توجه به واقعی بودن داده‌های مورد استفاده در این پژوهش تکنیک‌های بکارگرفته شده می‌تواند بر روی داده‌های دیگر صورت‌های مالی نیز برای پیش‌بینی تقلب اعمال شده و جلوی خسارهای ناشی از تقلب گرفته شود.
- ۲) با نظر به اینکه در این پژوهش رویکرد ترکیبی پیشنهادی از بین همه تکنیک‌ها توانایی بالاتری در پیش‌بینی و کشف تقلب دارد؛ حساب‌برسان، اعتباردهندگان، مدیران، سرمایه‌گذاران و سایر ذینفعان می‌توانند با استفاده از رویکرد ترکیبی پیشنهادی شرکت‌های متقلب و غیر متقلب را به منظور تصمیمات مالی خود تشخیص داده و از زیان سرمایه‌گذاری در شرکت‌های متقلب جلوگیری کنند.
- ۳) در کشور ایران علی‌رغم توانایی تکنیک‌های یادگیری ماشین در مباحث پیش‌بینی و ارزیابی در موضوعات مالی، عملاً کمتر مورد استفاده قرار گرفته‌اند؛ که یکی از دلایل اصلی این امر، عدم آشنایی با نحوه انجام پژوهش‌ها در این حوزه و نحوه اعمال این تکنیک‌ها می‌باشد؛ که با آموزش بیشتر این گونه پژوهش‌ها و نحوه کارکرد این تکنیک‌ها از طریق برگزاری کارگاه‌های آموزشی درآینده؛ می‌

جدول ۱۰- نتیجه ماتریس درهم‌ریختگی تکنیک combine

مدل طبقه بندی			طبقه بندی عملی
نرخ پیش‌بینی (%)	داده‌های بدون تقلب	داده‌های تقلب	
مجموعه داده‌های آموزشی (train):			
۹۶,۲	۴	۸۲	داده‌های تقلب
۹۸,۸	۷۷	۱	داده‌های بدون تقلب
۹۸,۵	-	-	نرخ دقت کل
مجموعه داده‌های آزمایشی (test):			
۹۱,۸	۷	۷۹	داده‌های تقلب
۹۸,۴۵	۷۶	۲	داده‌های بدون تقلب
۹۸,۸			نرخ دقت کل

۴- نتیجه‌گیری

با عنایت به مطالب مطرح‌شده مبرهن است که مدیریت، انگیزه‌هایی دارد که می‌تواند منجر بروز تقلب در صورت‌های مالی شود؛ از این رو کشف تقلب در زمره موضوعات مهم حوزه حسابداری و مالی به شمار می‌رود. پژوهش پیش‌روی، قصد دارد با تکنیک‌های یادگیری ماشین به کشف تقلب بپردازد. در این مقاله به بررسی تکنیک‌های زیرمجموعه‌ی یادگیری ماشین شامل؛ ماشین بردار پشتیبان، درخت تصمیم، شبکه عصبی، شبکه بیزین و روش ترکیبی پرداخته‌شده و نتایج آن‌ها مورد بررسی قرار گرفته است.

طبق جدول ۶ نتیجه پیش‌بینی SVM، جدول ۷ نتیجه پیش‌بینی ANNs، جدول ۸ نتیجه پیش‌بینی TREE، جدول ۹ نتیجه پیش‌بینی KNN و جدول ۱۰ نتیجه پیش‌بینی combine، نرخ دقت کلی برای مجموعه داده‌های آموزشی ۹۸,۵٪ و میزان دقت کلی برای مجموعه داده‌های آزمایشی ۹۸,۸٪ می‌باشد. با توجه به قدرت تشخیص شرکت‌های متقلب با داده‌های آموزشی، میزان نرخ پیش‌بینی ۹۶,۲٪ است، در حالیکه قدرت تشخیص با داده‌های آزمایشی برابر با ۹۱,۳٪ می‌باشد. با بررسی نتایج می‌توان مشاهده کرد که روش پیشنهادی ترکیبی (combine) دارای دقت و توان ارزیابی بالاتری است و به خوبی موارد تقلب صورت‌های مالی شرکت‌ها را استخراج می‌نماید.

- * ستایش، محمد حسین، ابراهیمی، فهیمه، سیف، سید مجتبی، ساریخانی، مهدی، (۱۳۹۱)، "پیش بینی نوع اظهار نظر حسابرسان با رویکردی بر روش های داده کاوی"، فصلنامه علمی پژوهشی حسابداری مدیریت، ۵(۱۱)، ۶۹-۸۲.
- * عظیمی، افسانه، نورحسینی، سیدمجید، (۱۳۹۶)، "کشف و پیش بینی تقلبات مالی در بانک ها با استفاده از روش های داده کاوی"، دومین کنفرانس پژوهش های دانش بینان در مهندسی کامپیوتر فناوری اطلاعات، تهران، دانشگاه مجلسی.
- * غلامی زاده، فائزه، حصارزاده، رضا، ساعی، محمد، (۱۳۹۷)، معیارهای واژه محور و معیارهای سنتی تقلب در گزارش های مالی، راهبرد مدیریت مالی، ۶(۲۰)، ۷۳-۹۲.
- * کمیته تدوین استانداردهای حسابرسی (۱۳۸۸)، "استانداردهای حسابرسی"، چاپ دهم، تهران، انتشارات سازمان حسابرسی.
- * علیزاده، سمیه و ملک محمدی، سمیرا، (۱۳۹۳)، "داده کاوی و کشف دانش گام به گام با نرم افزار کلمنتین"، تهران: انتشارات دانشگاه خواجه نصیرالدین طوسی.
- * غضنفری، مهدی، علیزاده، سمیه، تیمورپور، بابک، (۱۳۸۷)، "داده کاوی و کشف دانش، انتشارات دانشگاه علم و صنعت.
- * فرقاندوست حقیقی، کامبیز، بروآری، فرید و فروغی دهکردی، امین، (۱۳۹۳)، "مطالعه رابطه مدیریت سود و امکان تقلب در صورت های مالی"، فصلنامه دانش حسابرسی، ۱۴(۵۶)، ۶۷-۶۸.
- * کیقبادی، امیررضا، فتحی، سمیه، سیف، سمیرا، (۱۳۹۴). رتبه بندی میزان تاثیر اقلام کلیدی ترازنامه ای و نسبت های سودآوری در انتخاب پرتفوی بهینه (با استفاده از داده کاوی)، پژوهش های حسابداری مالی و حسابرسی، دوره ۷(۲۸)، ۷۵-۸۶.
- * نعمتی، مجتبی، سینه سپهر، مهرنوش و احمدی ترشیزی، حسن، (۱۳۹۲)، "تشخیص تقلب بر توان از تکنیک های یادگیری ماشین بیشتر در مباحث مالی استفاده کرد.
- ۴) اکثر متغیرهای مستقل (پیش بین) تقلب در این پژوهش می توانند در کشف تقلب صورت های شرکت های سرمایه گذاری، بیمه ای و واسطه ای موجود در بورس که از نمونه این پژوهش فاکتور گرفته شده اند نیز استفاده شوند.

منابع

- * امینی، پیمان، محمدی، کامران، عباسی، شعیب، (۱۳۹۰)، "بررسی عوامل مؤثر بر صدور گزارش مشروط حسابرسی: کاربرد روش شبکه عصبی"، فصلنامه حسابداری مدیریت، ۴(۱۱)، ۲۵-۳۹.
- * اعتمادی، حسین، عبدلی، لیلا، (۱۳۹۶)، "کیفیت حسابرسی و تقلب در صورت های مالی"، مجله علمی پژوهشی دانش حسابداری مالی، ۴(۴)، ۲۳-۴۳.
- * تاراسی، مجتبی، بنی طالبی دهکردی، بهاره، زمانی، بهزاد، (۱۳۹۸)، "پیش بینی گزارشگری مالی متقلبانانه از طریق شبکه عصبی مصنوعی (ANN)، فصلنامه علمی پژوهشی حسابداری مدیریت، ۱۲(۴۰)، ۶۳-۷۹.
- * حجازی، رضوان، محمدی، شاهپور، اصلانی، زهرا، آقاجانی، مجید(۱۳۹۱). "پیش بینی مدیریت سود با استفاده از شبکه عصبی و درخت تصمیم در شرکت های پذیرفته شده در بورس اوراق بهادار تهران"، بررسی های حسابداری و حسابرسی، ۱۹(۲)، ۳۱-۴۶.
- * خدمتگزار سالانقوچ، یاسر، (۱۳۹۲)، "بررسی خوشه بندی در شبکه های حسگر بی سیم و ارائه یک متد جدید در این رابطه" پایان نامه کارشناسی ارشد کامپیوتر، دانشگاه مهندسی فناوری نوین قوچان.
- * رهنمای رودپشتی. فریدون، (۱۳۹۱)، "داده کاوی و کشف تقلب"، فصلنامه دانش حسابداری و حسابرسی مدیریت، ۱(۳)، ۱۷-۳۳.

- (2018), "Fraud Ditection –Oriented Opretors In a Data Warehouse Base On Forensic Accounting Techniques", COMPUTER FRAUD & SECURITY, 10(2018),PP:13-19.
- * Kotsiantis, S., (2006), "Method Of Financing", Australian Accounting Review, Pp: 538-542 (Springer-Verlag Berlin, Heidelberg@ 2006).
- * Normah. O. Zulaikha, Amirah., J. Malcolm, S., (2017), "Predicting Fraudulent Financial Reporting Using Artificial Neural Network", Journal Of Financial Crime, 24 (2), Pp.362-387.
- * Sadgali. I, Sael. N & Benabbou. F,(2019), "Performance of machine learning techniques in the detection of financial frauds" Procedia Computer Science, 148(2019), Pp:45-54.
- * Spathis, C. , (2002). Detecting False Financial Statements Using Published Data: Some Evidence from Greece. Managerial Auditing Journal, Vol. 17(4), pp. 179-191.
- * Vapnik ,V. , (1995), "The Nature of Statistical Learning Theory", Springer-Verleg.
- * Watts, Rl. , & Zimerman, J. L., (1986), Pozetive Accounting Theory " , 2(124)
- اساس تکنیک های داده کاوی"، نخستین همایش منطقه ای فناوری اطلاعات، چالوس، موسسه آموزش عالی طبرستان.
- * وکیلی فرد، حمیدرضا، جبارزاده، سعید و سلطان احمدی، اکبر، (۱۳۸۸)، "بررسی ویژگی‌های تقلب در صورت‌های مالی"، حسابدار، ۲۴(۲۱۰)، ۳۶-۴۱.
- * هان، ژیاوی، میشلین، کامبر و ژان پی، (۱۳۹۲)، "داده کاوی-مفاهیم و روشها"، جلد دوم، ترجمه نسترن حاجی حیدری و سید بهنام خاکباز، تهران، انتشارات دانشگاه تهران.
- * Chi-Chen, Lin , An-An Chiu , Shaio Yan Huang , David C. Yen,(2015) ,” Detecting The Financial Statement Fraud: The Analysis Of The Differences Between Data Mining Techniques And Experts Judgments”, Knowledge-Based Systems, 89(15), Pp.459-470.
- * Che. Tsai, Hung. C, (2008), “Automatically Annotating Images With Keywords: A Review Of Image Annotation Systems” , Recent Patents On Computer Science, Vol.1, Pp. 55-68.
- * Dechow, M. P. , Sloan, R. G. , & Sweeney, A. P. (1996). “ Coues And Consequences Of Earnings Manipulation: An Analysis Of Firms Subject To Enforcement Actions By The Sec” Contemporary Accounting Research , vol.13 NO.1, PP.1-36.
- * Dechow, M. P. , Sloan, R. G. , & Sweeney, A. P. (1995). " Detecting earnings management " The Accounting Review, 70 (2) , 193.
- * Domingues, D., Filippone, M., Michiardi, P., (2019) “ Probabilistic Modeling For Novelty Detection With Applications To Fraud Identification,Phd Thesis of Paris Sorbonne Universite.
- * Fama, E. F. and Jensen, M. C. (1983). “Sparation of ownership and control” ,Journal of Law and Economics, Vol. 26. PP: 327-349.
- * Jen Chen, Yuh, Ching Liou, Wan, Min Chen, Jyun, Han Wu(2018), “Fraud Detection For Financial Statements Of Business Groups” ,International Journal Of Accountion Information Systems” 7(15), Pp10-26.
- * Javier Moreno Arboleda, Francisco, Guzman –Luna, Alberto, Durely Torres,

یادداشت‌ها

- ¹ Sadgali, et al
² Jen Chen, et al
³ Machine learning
⁴ Javier, et al
⁵ Watts & Zimerman
⁶ Dechow, et al
⁷ Fama & Jensen
⁸ Spathis
⁹ Kotsiantis
¹⁰ Domingues, et al
¹¹ Bayesian networks
¹² decision tree
¹³ artificial neural networks
¹⁴ support vector machine
¹⁵ Vapnik
¹⁶ Che and Hung
¹⁷ Normah, et al
¹⁸ Chi-Chen, et al