نشریـه علمـی حسابداری مدیریت
سال سیزدهم/ شماره چهل‌وپنجم/ تابستان ۱۳۹۹

کشف تقلب صورت‌های مالی با توجه به گزارش حسابرسی صورت‌های مالی

مهدی رضائی[۱]
مهدی ناظمی اردکانی[۲]
علیرضا ناصر صدرآبادی[۳]

تاریخ دریافت: ۹۸/۰۳/۲۵ تاریخ پذیرش: ۹۸/۰۷/۲۰

چکیده

هدف اصلی این مقاله کشف تقلب صورت‌های مالی با توجه به گزارش حسابرسی صورت‌های مالی است. داده های 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 نمونه گیری حذف سیستماتیک گزینش شده اند. نمونه آماری پژوهش با توجه به نوع گزارش حسابرسی به دو گروه مجزا شامل شرکت های متقلب(۱) و شرکت های غیرمتقلب(۰) دسته بندی شده اند. متغیرهای مستقل تاثیر گذار بر تقلب در این پژوهش در برگیرنده ۴۱ متغیر مالی و غیر مالی می باشد که بر اساس مبانی نظری و پیشینه پژوهش انتخاب شده اند. در نهایت داده های مربوط به متغیرها که به روش کتابخانه ای جمع آوری گردیده است، به پنج تکنیک برتر از بین تکنیک های یادگیری ماشین داده شده است، که این تکنیک ها شامل؛ شبکه های بیزین، درخت تصمیم، شبکه های عصبی ، ماشین بردار پشتیبان و روش ترکیبی می باشد. پس از بکارگیری این تکنیک ها در کشف تقلب صورت‌های مالی ، نتایج نشان داد تمامی تکنیک ها قابلیت کشف تقلب صورت‌های مالی را در سطح نسبتا بالایی دارند و تکنیک پیشنهادی ترکیبی با میزان نرخ پیش بینی ۹۶,۲٪ دارای دقت و توان ارزیابی بالاتری نسبت به سایر تکنیک ها 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 صورت‌های 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 همکاران۱، 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 همکاران۲، ۲۰۱۸). در داخل کشور به موضوع کشف تقلب صورت‌های 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 اهیمت 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 صورت‌های 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 صورت‌های 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، حیله گری و رفتار غیر منصفانه بوسیله یک شخص فریبکار و متقلب است (جاویر و همکاران۴، 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 صورت‌های مالی ، طبق الزامات افشاء استانداردهای حسابداری است. طبق بند ۴ این استاندارد؛ تقلب عبارتست از هرگونه اقدام عمدی یا فریبکارانه یک یا چندنفر 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 صورت‌های 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: پاداشِ 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] ،1983)، حداقل کردن بدهی مالیاتی (دچو و همکاران، ۱۹۹۶)، جلوگیری از تخلف در قراردادهای بدهی (اسپاتیس[۸]،

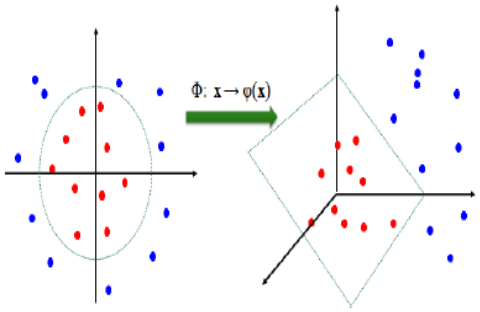
۲۰۰۲)، تأمین مالی به ارزانترین شکلِ ممکن (کوتسیانتیس[۹]، ۲۰۰۶) .

- **یادگیری ماشین(ML)**

یادگیری ماشین زمینه تحقیقاتی نسبتاً جدیدی از هوش مصنوعی است که در حال حاضر دوران رشد و تکامل خود را می گذراند و زمینه ای بسیار فعال در علوم کامپیوتر می باشد. علوم مختلفی در ارتباط با یادگیری ماشین هستند. این تکنیک ها قابلیت به کارگیری در زمینه های مختلفی از قبیل؛ کشف تقلب، کشف ورود بی اجازه، تشخیص اشتباهات پزشکی، لایروبی داده ها ، جلوگیری از خطا و ... را دارا می باشند (دومینگس و همکاران[۱۰]، ۲۰۱۹). در این پژوهش از چهار تکنیک معروف و از قوی ترین تکنیک های یادگیری ماشین شامل شبکه های بیزین(KNNs[۱۱]) ، درخت تصمیم(TREE[۱۲])، شبکه های عصبی (ANNs[۱۳])، ماشین بردار پشتیبان (SVM[۱۴]) استفاده می شود.

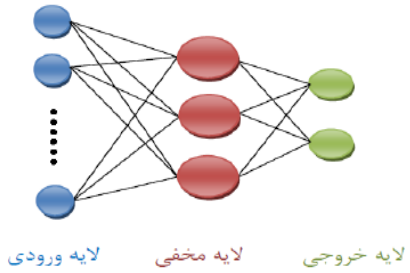
- شبکه بیزین یک مدل گرافیکی است که متغیرهای موجود در یک مجموعه داده را که معمولا گره نامیده می شوند؛ همراه با روابط احتمالی و شرطی بین آن ها نشان می دهند. روابط سببی بین گره ها می تواند به وسیله شبکه بیزین نمایش داده شود، البته پیوندهای موجود در شبکه (کمان ها) لزوماً ارتباط مستقیم علت معلولی را نشان نمی دهند. این شبکه ها در مقابل داده های مفقود بسیار مقاومند و بهترین پیش بینی را ارائه می دهند (علیزاده و ملک محمدی ، ۱۳۹۳). یکی از سودمندترین مزایای شبکه های بیزین این است که جهت آنالیز، چیزی به عنوان اندازه حداقل برای داده های نمونه، وجود ندارد و این شبکه ها در محاسبات خود از تمامی داده های موجود استفاده می کنند و حتی با وجود نمونه هایی با اندازه ای نسبتاً کوچک می توان به پیش بینی های نسبتا خوبی دست یافت (خدمتگزار سالانقوچ، ۱۳۹۲). این ساختارهای گرافیکی برای نشان دادن اطلاعات در یک حوزه دارای عدم قطعیت به کار می روند. از جمله روش های معروف بیزین روش شبکه باور بیزین و روش بیز ساده می باشد (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، پیش بینی به درست آمده از درخت در قالب یکسری قواعد توضیح داده می شود (غضنفری، علیزاده و تیمورپور، ۱۳۸۸). یکی از مزایای آن این است که بعد از ساخت درخت به راحتی می توان علت استنتاج قواعد بدست آمده را مشاهده نمود بدین معنا که درخت تصمیم همانند شبکه عصبی مانند یک جعبه سیاه عمل نمی کند؛ 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 می باشد. درختهای تصمیم به علت دقت مناسب و همچنین ساده بودن و فهم راحت آن، یکی از محبوبترین روش های طبقه بندی هستند(حاجی حیدری و خاکباز، ۱۳۹۲). مدل های حاصل از درخت تصمیم دارای الگوریتم متفاوتی مثلCHAID، CART ،C5.0 می باشد(کیقبادی و همکاران، ۱۳۹۴). در این پژوهش بدلیل توانمندتر بودن تکنیک CART از این تکنیک جهت طبقه بندی شرکت های متقلب و غیر متقلب استفاده شده است.
- ماشین بردار پشتیبان از روش های یادگیری تحت نظارت است که برای دسته بندی و رگرسیون استفاده می شود. ماشین بردار پشتیبان توسط واپنیک بر اساس تئوری یادگیری آماری معرفی شده است (واپنیک۱۹۹۵،۱۵). در این روش اساس کار بر این است که بردارهای ورودی (بردارهای

ویژگی) به فضایی با ابعاد بالاتر نگاشت می شوند، در این فضای جدید دسته بندی این بردارها کاراتر و مناسب تر انجام می شود (شکل۱) همچنین SVM به دلیل انتخاب جداکننده بهینه که بیشترین فاصله را از نمونه ها دارد قابلیت تعمیم بالایی دارد.



شکل ۱ : نگاشت نمونه ها به فضایی با ابعاد بالاتر در فرآیند تولید مدل در روش SVM

- شبکه های عصبی مصنوعی چکیده ای از مغز انسان است. شبکه عصبی مصنوعی، مشابه مغز، از نورون های مصنوعی و اتصالات ساخته شده است. ورودی شبکه عصبی بردارهای ویژگی و خروجی شبکه معرف طبقات مختلف خواهد بود. شبکه های چندلایه پیش خور یکی از مهمترین و رایج ترین شبکه های عصبی در کاربردهای دنیای واقعی هستند(چی و هانگ۱۶، ۲۰۰۸).



شکل۲: نمونه ای از ANN با یک لایه مخفی

در این پژوهش نیز از شبکه عصبی چندلایه پیش خور با الگوریتم پیش انتشار استفاده شده است. علت استفاده از تکنیک شبکه عصبی به این روش ، توان بالای آن در انجام پردازش موازی و غیر خطی برروی دادها است که می تواند شدت ارتباط متغیرهای مستقل و وابسته را دقیق تر 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 مطالعه ای به پیشنهاد تکنیک های برتر یادگیری ماشین در زمینه کشف و پیشگیری از تقلب مالی پرداختند. در این پژوهش پس از شناسایی تکنیک ها و روش های مختلف یادگیری ماشین که در کشف تقلب صورت‌های مالی در پژوهش های گذشته استفاده شده بودند؛ آنها را از نظر شاخص هایی چون دقت، حساسیت، بلادرنگ بودن، بانک اطلاعاتی و مشاهدات مورد مقایسه قرار دادند. نتایج این مطالعه نشان داد که تکنیک های کشف تقلب دورگه (هایبردی) نسبت به سایر تکنیک ها، عملکرد بهتری دارند بویژه زمانیکه آن ها چندین روش کشف تقلب سنتی را ترکیب می نمایند.

جن چن وهمکاران (۲۰۱۸)در مطالعه ای رویکردهای کشف تقلب صورت‌های مالی گروه های 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 مکانیسم های کشف تقلبی که دارای دقت و توانایی بالایی در کشف تقلب بودند شناسایی نمودند و توسعه دادند که در بین این تکنیک های 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 پیش بینی 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

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

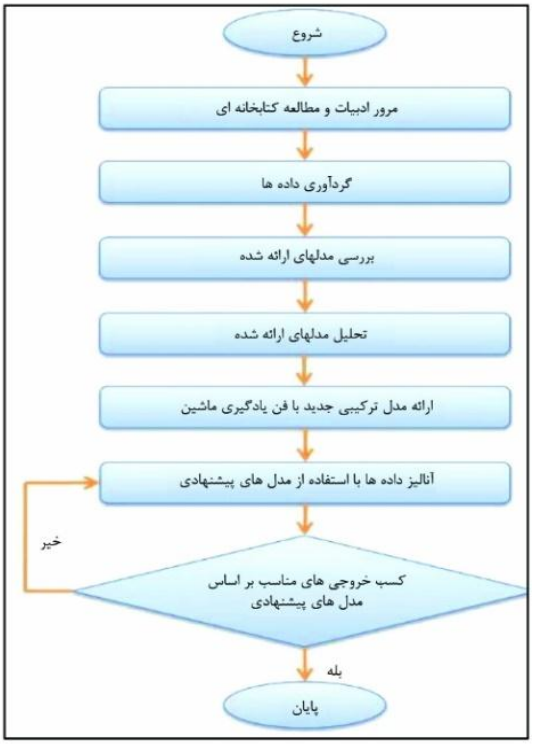
چی چن و همکاران[18] (۲۰۱۵) به تجزیه و تحلیل تفاوت های میان تکنیک های داده کاوی و خبره سنجی در کشف تقلب صورت‌های مالی پرداختند. روش های تجزیه و تحلیل داده ها به کاربرده شده در این پژوهش شامل رگرسیون منطقی ، CART، شبکه های ANNS بوده است. نتایج این پژوهش بیانگر کارایی و توانایی بالاتر تکنیک های داده کاوی نسبت به خبره سنجی است.

تاراسی و همکاران (۱۳۹۸) با استفاده از تکنیک شبکه های عصبی به کشف تقلب صورت‌های مالی پرداختند. نتایج این پژوهش،نشان داد که تکنیک شبکه عصبی و مدل سازی مبتنی بر شبکه عصبی از صحت ۹۷,۴ درصد برخوردار است و با طراحی و آموزش دقیق، می توان شبکه عصبی طراحی نمودکه با دقت قابل قبولی بتوانند احتمال گزارشگری مالی متقلبانه شرکت را کشف و پیش بینی نمایند. عظیمی و نورحسینی (۱۳۹۶) به کشف و پیش بینی تقلب مالی در بانک ها با استفاده ازروش های داده کاوی، پرداختند. نتایج خروجی نشان داد که داده کاوی در حدود ۸۳٪ پیش بینی درستی در خروجی نمایش می دهد.

رهنمای رودپشتی (۱۳۹۱) اثربخشی تکنیک های داده کاوی در تشخیص رفتارهای متقلبانه شرکت هایی که صورتهای مالی متقلبانه گزارش نموده اند را بررسی کرده تا عوامل موثر بر اینگونه رفتارها را شناسایی نماید. نتایج مطالعه نشان می دهد که اولاً ، تکنیک های داده کاوی، در شناسایی صورتهای مالی متقلبانه سودمند هستند. ثانیاً داده کاوی، به عنوان کانون هدایت فکر در مدیریت تجارت جهت کشف تقلب می تواند مورد توجه قرار گیرد.

۳- روش شناسی پژوهش

این پژوهش از لحاظ هدف کاربردی و از نوع پژوهش های تجربی است و با توجه به اینکه از اطلاعات تاریخی استفاده شده است در دسته پژوهش های شبه آزمایشی قرار می گیرد. برای جمع آوری اطلاعات مربوط به مبانی نظری و پیشینه پژوهش و همچنین داده های لازم برای اجرای تکنیک ها از روش گردآوری کتابخانه ای و ابزارگردآوری اسنادکاوی استفاده شده است. داده های جمع آوری شده با استفاده از نرم افزار اکسل تلخیص و سپس متغیرهای مورد نظر محاسبه شده اند. در نهایت از نرم افزارهای Rapidminer 7.3 و matlaبرای سنجش توانایی تکنیک های 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 می شود.

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

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 ویژگی‌ها، 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 شکل‌گیری شاخصی شد که نشان‌دهنده رفتار تقلب و فاقد تقلب در طول چندین دوره بوده که بر اساس سابقه عملکرد، ویژگی‌های 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 مدل‌های رفتاری، لازم بود آن‌ها را تجزیه‌وتحلیل نموده تا به متغیرهای تأثیرگذار تقلیل یابد. درنهایت، مؤلفه‌های اصلی شرکت‌های انتخاب شده به ترتیب شامل ساختار سرمایه، سودآوری ، عملکرد، نقدینگی، جریان نقد عملیاتی و معیارهای غیرمالی را در برمی‌گیرد.

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 اندازگیری تقلب صورت‌های مالی و تفکیک شرکت های متقلب از غیر متقلب از شیوه اندازه گیری تقلب توسط فرقاندوست 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 دریافتنی به کل دارایی ها ۷- دفعات گردش موجودی کالا۸-نسبت آنی۹-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

۱۵- سرانه فروش ۱۶- سرانه سودخالص ۱۷- گردش حساب های دریافتنی ۱۸-دوره وصول مطالبات۱۹ - 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 حسابرسی۲- 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 ذکرشده 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 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 داده‌کاوی 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 مدل ها 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 تکنیک ها پرداخته شده است.

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 و۵ 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

در ادامه جداول ارائه شده بیانگر ماتریس در هم ریختگی یا اغتشاش می باشد که تعداد سطر ها و ستون های این ماتریس به تعداد طبقات(class) بستگی دارد؛ که در این پژوهش داده ها در برگیرنده دو طبقه شامل شرکت های متقلب و شرکت های غیر متقلب می باشد ؛ بنابراین این ماتریس برای تمامی تکنیک های مورد استفاده دو در دو ترسیم می شود. بعد از ساخت یک مدل طبقه بندی (یعنی یافتن الگوریتمی که با مشاهده یک داده جدید، طبقه آن را مشخص کند)؛ برای سنجش میزان کارآیی و دقت مدل پیشنهادی آن را بر روی داده های آموزشی (یعنی داده هایی که از قبل طبقه بندی آن ها مشخص شده است)، اعمال می کنیم و نهایتاً در مورد داده های آزمایشی (test) نیز اعمال می شود. جدول شماره ۵ نشاندهنده حالت های مختلفی است که در نتیجه مقایسه مقادیر پیش بینی شده و واقعی در ماتریس درهم ریختگی پیش می آید.

جدول ۵- ماتریس در هم ریختگی خام

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

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

مالی شرکت عاری از تقلب بوده است. بنابراین تحلیل ماتریس درهم ریختگی برای پژوهش عبارت است از:

- درست مثبت(TP): اگر صورت‌های مالی شرکت در واقع متقلبانه باشد و مقدار پیش بینی شده، نشانگر متقلبانه بودن صورت‌های مالی شرکت باشد.
- نادرست مثبت(FP): اگرصورت‌های مالی شرکت در واقعیت متقلبانه نباشد اما مقدار پیش بینی شده، نشانگر متقلبانه بودن صورت‌های مالی شرکت باشد (خطای نوع اول).
- درست منفی(TN): اگرصورت‌های مالی شرکت در واقع متقلبانه نباشد و مقدار پیش بینی شده نیز، نشانگر عاری از تقلب بودن گزارشات باشد.
- نادرست منفی(FN): اگرصورت‌های مالی شرکت در واقعیت متقلبانه باشد اما مقدار پیش بینی شده، نشانگر متقلبانه بودن صورت‌های مالی شرکت نباشد(خطای نوع دوم). که این موضوع بااهمیت تر بودن خطای نوع دوم را نشان می دهد.

جدول ۶- نتیجه ماتریس درهم ریختگی تکنیک SVM

طبقه بندی عملی	مدل طبقه بندی		
	داده های تقلب	داده های بدون تقلب	نرخ پیش بینی (٪)
مجموعه داده‌های آموزشی(train):			
داده های تقلب	۶۳	۲۳	۷۳,۵۸
داده های بدون تقلب	۱۲	۶۶	۸۴,۶
نرخ دقت کل	-	-	۹۳,۴
مجموعه داده‌های آزمایشی(test):			
داده های تقلب	۵۶	۳۰	۶۵,۲۹
داده های بدون تقلب	۹	۶۹	۸۸,۶
نرخ دقت کل	-	-	۸۹,۶۸

تکنیک SVM همانطور که در جدول شماره ۶ آمده است؛ براساس داده های 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 داده های آموزشی ٪۷۳,۵۸ و طبق داده های آزمایشی ٪۶۵,۲۹ از شرکت های متقلب را پیش بینی کند. خطای نوع دوم این تکنیک حدود ٪۱۱ می باشد. بنابراین تکنیک SVM با توجه به نرخ دقت ٪۹۳,۴ و قدرت پیش بینی و کشف ٪۸۹,۶۸ در تشخیص تقلب قابلیت کشف تقلب در سطح نسبتا خوبی را داراست.

جدول ۷- نتیجه ماتریس درهم ریختگی تکنیک ANNs

طبقه بندی عملی	مدل طبقه بندی		
	داده های تقلب	داده های بدون تقلب	نرخ پیش بینی (٪)
مجموعه داده‌های آموزشی(train):			
داده های تقلب	۷۶	۱۰	۸۸,۶
داده های بدون تقلب	۲	۷۶	۹۷,۳
نرخ دقت کل	-	-	۹۷,۵
مجموعه داده‌های آزمایشی(test)			
داده های تقلب	۷۴	۱۲	۸۶,۹۶
داده های بدون تقلب	۲	۷۶	۹۷,۴۱
نرخ دقت کل	-	-	۹۶,۳

تکنیک ANNs همانطور که در در جدول شماره ۷ آمده است؛ براساس داده های 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 داده های آموزشی ٪۸۸,۶ و طبق داده های آزمایشی ٪۸۶,۹۶ از شرکت های متقلب را پیش بینی کند. خطای نوع دوم این تکنیک حدود ٪۲,۵ می باشد. تکنیک ANNs با توجه به نرخ دقت ٪۹۷,۵ و قدرت کشف ٪۹۶,۳ در تشخیص تقلب

دارای قدرت و قابلیت بالایی مـی باشـد و اینکـه تـوان ANNs در کشف تقلب از تکنیک SVM بالاتراست.

تکنیک درخت تصمیم(TREE) همـانطور کـه در در جدول شـماره ۸ آمـده اسـت؛ براسـاس داده هـای آزمایشی**(test)** ، از بین ۸۶ شرکت متقلب ۶۹ شـرکت متقلب و از بین ۷۸ شـرکت غیرمتقلـب ۷۳ شـرکت را بدرستی تشخیص دهد؛ کـه ایـن موضـوع بیـانگر ایـن است که تکنیک TREE قـادر بـوده طبـق داده هـای آموزشی ۸۳,۹۶٪ و طبق داده های آزمایشی ۸۰,۴۳٪ از شرکت های متقلب را پیش بینی کند. خطـای نـوع دوم این تکنیک حدود 6 ٪ می باشـد. تکنیـک TREE با توجه به نرخ دقت ٪۹۶,۱ و قدرت کشف ۹۳,۳٪ در تشخیص تقلب دارای قدرت و قابلیت بالایی می باشـد و اینکه توان تکنیک TREE در کشف تقلب از تکنیک ANNs پایین تر ولی از تکنیک SVM بالاتر است.

تکنیـک شـبکه بیـزین (KNN) همـانطور کـه در در جـدول شـماره ۹ آمـده اسـت؛ براسـاس داده هـای آزمایشی**(test)** ،از بین ۸۶ شرکت متقلب ۶۱ شـرکت متقلب و از بین ۷۸ شـرکت غیرمتقلـب ۷۱ شـرکت را بدرستی تشخیص داده است؛ که ایـن موضـوع بیـانگر این است که تکنیک KNN قادر بوده طبـق داده هـای آموزشی ۷۵,۴۳٪ و طبق داده های آزمایشـی ۷۱,۷۴٪ از شرکت های متقلب را پیش بینی کند. خطـای نـوع دوم این تکنیـک حـدود ۹٪ مـی باشـد. بنـابراین ایـن تکنیک با توجه به نـرخ دقـت ۹۴,۶٪ و قـدرت کشـف ۹۰,۳۶٪ در تشـخیص تقلـب دارای قـدرت و قابلیـت بالایی می باشد و اینکه توان تکنیـک KNN در کشـف تقلب از تکنیک های ANNs و TREE پایین تر ولـی از تکنیک SVM بالاتر است.

تکنیک ترکیبی (combine) همانطور که در در جدول شماره ۱۰ آمده است؛ براساس داده های آزمایشی **(test)**، این از بین ۸۶ شرکت متقلب 79 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 91.8٪ از شرکت های متقلب را پیش بینی کند. 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 واینکه 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 تکنیکTREE

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

جدول ۹- نتیجه ماتریس درهم ریختگی تکنیک KNN

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

جدول ۱۰- نتیجه ماتریس درهم ریختگی تکنیک combine

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

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 پیش روی، 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 موردبررسی قرارگرفته است.

طبق جدول ۶ نتیجه پیش بینی SVM ، جدول ۷ نتیجه پیش بینی ANNs، جدول ۸ نتیجه پیش بینی TREE ، جدول ۹ نتیجه پیش بینی KNN و جدول ۱۰ نتیجه پیش بینی combine ، نرخ دقت کلی برای مجموعه داده های آموزشی ۹۸,۵٪ و میزان دقت کلی برای مجموعه داده های آزمایشی ۹۸,۸٪ می باشد. با توجه به قدرت تشخیص شرکت های متقلب با داده های آموزشی، میزان نرخ پیش بینی ۹۶,۲٪ است، در حالیکه قدرت تشخیص با داده های آزمایشی برابر با ۹۱,۳٪ می باشد. 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.

یافته‌های پژوهش حاکی از آن است که مولفه های نسبت های مالی و سایر متغیرهای مالی و غیر مالی استفاده شده در این پژوهش تخمین زننده های توانایی در کشف تقلب صورت‌های مالی می باشند.

نتایج بدست آمده از این پژوهش با نتایج پژوهش های چی چن و همکاران (۲۰۱۵) ، تاراسی و همکاران (۱۳۹۷) و جن چن و همکاران (۲۰۱۸) همخوانی دارد به گونه ای که تکنیک های مشترک استفاده شده در اینگونه مطالعات با این پژوهش از جمله درخت تصمیم ، شبکه های عصبی و ماشین بردار پشتیبان در حد بالایی توان کشف تقلب صورت‌های مالی را دارد و می تواند بوسیله استفاده کنندگان از صورت‌های مالی مورد استفاده قرار گیرد. با توجه به نتایج پژوهش، پیشنهادات کاربردی زیر ارائه می گردد:

۱) با توجه به واقعی بودن داده‌های مورد استفاده در این پژوهش تکنیک های بکارگرفته شده می‌تواند بر روی‌داده‌های دیگر صورت‌های مالی نیز برای پیش بینی 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 تکنیک ها توانایی بالاتری در پیش بینی و کشف تقلب دارد؛ حسابرسان، اعتباردهندگان، مدیران، سرمایه گذاران و سایر ذینفعان می توانند با استفاده از رویکرد ترکیبی پیشنهادی شرکت های متقلب و غیر متقلب را به منظور تصمیمات مالی خود تشخیص داده و از زیان سرمایه گذاری در شرکت های متقلب جلوگیری کنند.

۳) در کشور ایران علی رغم توانایی تکنیک های یادگیری ماشین در مباحث پیش بینی و ارزیابی در موضوعات مالی، عملا کمتر مورد استفاده قرار گرفته اند؛ که یکی از دلایل اصلی این امر، عدم آشنایی با نحوه انجام پژوهش ها در این حوزه و نحوه اعمال این تکنیک ها می باشد؛ که با آموزش بیشتر این گونه پژوهش ها و نحوه کارکرد این تکنیک ها از طریق برگزاری کارگاه های آموزشی درآینده؛ می

توان از تکنیک های یادگیری ماشین بیشتر در مباحث مالی استفاده کرد.

۴) اکثر متغیرهای مستقل (پیش بین) تقلب در این پژ وهش می توانند در کشف تقلب صورت‌های 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 صورت‌های مالی"، مجله علمی پژوهشی دانش حسابداری مالی، ۴(۴)، ۲۳-۴۳.
* تاراسی، مجتبی، بنی طالبی دهکردی، بهاره، زمانی، بهزاد، (۱۳۹۸)،" پیش بینی گزارشگری مالی متقلبانه 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 وکشف تقلب"،فصلنامه دانش حسابداری وحسابرسی مدیریت، ۱(۳)،۱۷-۳۳.
* ستایش، محمد حسین، ابراهیمی، فهیمه، سیف، سید مجتبی، ساریخانی، مهدی، (۱۳۹۱)، "پیش بینی نوع اظهار نظر حسابرسان با رویکری بر روش های داده کاوی"، فصنامه 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 صورت‌های 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

اساس تکنیک های داده کاوی"، نخستین همایش منطقه ای فناوری اطلاعات، چالوس، موسسه آموزش عالی طبرستان.

* وکیلی‌فرد، 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). " Couses 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) " Probabilitic Modeling For Novelty Detection With Applications To Fraud Identification,Phd Thesis of Paris Sorbonne Universite.
* Fama, E. F. and Jensen, M. C. (1983). "Scparation 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, (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)

یادداشت‌ها

[1] Sadgali, et al
[2] Jen Chen, et al
[3] Machine learnining
[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