فصلنامــه اقتصاد مالی
سال پانزدهم / شماره ٥٤ / بهار ١٤٠٠
صفحه ٢٦٧ تا ٢٩٢

کارایی مدل‌های آماری والگوهای یادگیری ماشین در پیش بینی گزارشگری مالی متقلبانه

حسن ملکی کاکلر[١]
جمال بحری ثالث[٢]
سعید جبارزاده کنگرلویی[٣]
علی آشتاب[٤]

تاریخ دریافت: ١٣٩٩/١٠/١٤ تاریخ پذیرش: ١٣٩٩/١٢/١٦

چکیده

وجود تقلب و تداوم آن در صورتهای مالی، آثار گسترده‌ای بر سلامت مالی شرکت‌ها و توسعه پایدار بازار سرمایه دارد. روش‌های متداول حسابرسی در پیشگیری و کشف صورت‌های مالی متقلبانه، نتوانسته اند با تقلب‌های حسابداری نوظهور به دلیل فقدان دانش مورد نیاز داده‌کاوی، پیچیدگی تقلب‌های جدید و عدم تجربه کافی حسابرسان کنار بیایند. در این پژوهش، انواع مدل‌های آماری و یادگیری ماشین در دست‌یابی به الگویی با کارایی بالا در پیش‌بینی گزارشگری مالی متقلبانه استفاده شد. از ٢٠ متغیر در قالب الگوی پنج ضلعی تقلب با تاکید بر ساختار کنترل‌های داخلی در ١۶۶ شرکت‌های فعال در بورس اوراق بهادار تهران طی سالهای ١٣٨٨ الی ١٣٩٧ و مقایسه بین مدل‌های مورد بررسی، با کمک آزمـون مقایسـهٔ نسبت‌ها، نشان می‌دهد که به لحاظ آماری مدل‌های یادگیری ماشـین در پیش‌بینی گزارشگری مالی متقلبانه نسـبت بـه مدل‌های آماری، کارایی و دقت بیشتری دارند. ترکیب الگوریتم درخت تصمیم‌گیری CHAID، C5 و C&R بالاترین دقت در پیش‌بینی گزارشگری مالی متقلبانه را با دقت بالای ٩٢/۶١ درصد در پیش‌بینی تقلب نشان می‌دهد. روش‌های داده‌کاوی بر پایه مدل‌های یادگیری ماشین و بویژه ترکیب آنها بطور موفقیت‌آمیزی در پیش‌بینی و کشف تقلب در صورت‌های مالی می تواند مورد استفاده قرار گیرد.

واژه‌های کلیدی: گزارشگری مالی متقلبانه، مدل‌های‌آماری، مدل‌های یادگیری ماشین.
طبقه بندی JEL : M41,M42 ,G32

١- گروه حسابداری، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران، assan.maleki.k@gmail.com
٢- گروه حسابداری، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران، j.bahri@iaurmia.ac.ir
٣- گروه حسابداری، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران،(نویسنده مسئول) s.jabbarzadeh@iaurmia.ac.ir
۴- گروه حسابداری، دانشگاه ارومیه، ارومیه، ایران، a.ashtab@urmia.ac.ir

۱- مقدمه

روند افزایشی تقلب در محیط کسب و کار دنیای امروزی باعث شده کشف تقلب در صورت‌های مالی به موضوعی نوظهور برای صاحبان بنگاه‌های تجاری، صنایع و همچنین تحقیقات دانشگاهی تبدیل شود (شرما و پانیگراهی[1]، ۲۰۱۳). رسوایی‌های مالی شرکت‌های بزرگی همچون انرون، ورلدکام در دنیا و آثار زیان باری که در طی یک دهه گذشته به بار آورده‌اند، باعث کاهش اعتماد عمومی و بویژه کاهش اعتماد سرمایه‌گذاران نسبت به گزارش‌های مالی و خدمات حسابرسان شده است (خواجوی و ابراهیمی، ۱۳۹۶). از این رو سازمان‌ها باید در تلاش دائمی برای پیشگیری، بازدارندگی و کشف تقلب باشند (امیدی و همکاران، ۲۰۱۹). بنا به گزارش بازرسان رسمی تقلب[2] (۲۰۱۸) انواع تقلب در سه شکل سوءاستفاده از دارایی‌ها ، فساد مالی و تقلب در صورت‌های مالی طبقه بندی شده است. در این میان، گزارشگری مالی متقلبانه اگرچه در مقایسه با سایر اشکال تقلب کمتر اتفاق می‌افتد با این‌حال از لحاظ هزینه‌های تحمیل شده به شرکت‌ها گران‌ترین نوع تقلب به شمار می‌رود.

در سال‌های اخیر در ایران نیز ارتکاب تقلب، بحث غالب مطرح بازارهای مالی و نهادهای اقتصادی- اجتماعی داخلی بوده است. بحران مالی ایران در سالهای ۱۳۹۰ و ۱۳۹۷ که همراه با کاهش شدید ارزش ریال بود پیامدهای منفی از طریق تاثیرگذاری بر فضای کسب و کار و محدود کردن منابع مالی در اختیار به دنبال داشت. نمونه بارز تقلب‌های مالی سال‌های اخیر، پرونده‌ی فساد گسترده بانک سرمایه به میزان ۱۴ هزار میلیارد تومان که از آن به عنوان ابرفساد مالی یاد می‌شود؛ تاثیر بسزای مستقیم و غیرمستقیم برروی بازارهای مالی داخلی داشته‌اند. اگرچه در ایران آمارهای رسمی از تقلب‌ها ارائه نمی‌شود ولی طبق آمارهای سازمان شفافیت بین‌الملل در سال ۲۰۱۹، ایران با نمره ۲۷، در بین ۱۸۰کشور دنیا رتبه ۱۴۶ را دارد.

کشف تقلب مدیریت با استفاده از روشهای حسابرسی عادی به دلایل زیرکار نسبتا دشواری است. اول اینکه شناخت کافی در مورد ویژگی‌های تقلب مدیریت وجود ندارد. دوم ، بیشتر حسابرسان فاقد تجربه لازم برای کشف تقلب بوده و نهایتا ، مدیران مالی و حسابداران عمدتاً تلاش می‌کنند تا کار حسابرسی را بی‌اهمیت جلوه دهند. برای چنین مدیرانی که از این محدودیت‌های حسابرسی مطلع هستند، ممکن است روش‌های استاندارد حسابرسی کافی نباشد (امیدی و همکاران، ۲۰۱۹). از این رو برای کشف مؤثر صورتهای مالی متقلبانه نیاز به روش‌های تحلیلی اضافی امری غیر قابل اجتناب می‌نماید. با توجه به تأثیر فزاینده گزارش حسابرسی شرکت‌ها بر ثبات بازارهای مالی، در سال‌های اخیر تلاش‌های گسترده‌ای برای ارتقای عملیات حسابرسی انجام شده، ابزارها و فنونی طراحی شده است که حسابرسان بتوانند عملیات حسابرسی خود را به کمک

رایانه انجام دهند، که یکی از این فنون و ابزارها، روش‌های یادگیری ماشین[۳] است. در طی دهه گذشته اگرچه از این روش‌های داده‌کاوی[۴] برای کشف تقلب‌هایی نظیر پولشویی[۵] ،تقلب در کارت اعتباری تجارت الکترونیکی[۶]، تقلب در سازمان‌های بیمه با موفقیت استفاده شده است با این حال، به‌دلیل ماهیت پیچیده گزارشگری مالی متقلبانه کشف و شناسایی آن نیز پیچیده و دشوار است (آپاراو[۷] و همکاران، ۲۰۰۹).

در بیشتر پژوهش‌ها در زمینه کشف تقلب در صورت‌های مالی، به بررسی رویکرد آمار کلاسیک از جمله رگرسیون پرداخته شده است (زلقی و اعتمادی، ۱۳۹۲). مدل‌های آماری اگرچه توانایی پیش‌بینی‌های قابل قبول در تحلیل مسایل اقتصاد و حسابداری ارائه می‌کنند با این‌حال مفروضات محدودکننده برخی از این مدل‌ها (ابهام در روابط متغیرهای ورودی و خروجی مدل، خطای ناشی از هم خطی، کمبود مشاهدات...) اثربخشی آنها را کاهش می‌دهد (اصفهانی‌پور و همکاران، ۲۰۱۵). بنابراین این سوال که آیا می‌توان مدلی ارائه داد که بتواند گزارشگری مالی متقلبانه را با دقت بالایی پیش‌بینی کند، از موضوعات بسیار مهم در این زمینه است. برای انجام پیش‌بینی معمولا با تجزیه و تحلیل داده‌های گذشته، الگویی شناسایی شده که بتوان از وقوع تقلب جلوگیری کرد. برای بررسی این امر در این پژوهش، از روش‌های یادگیری ماشین استفاده می‌شود. در روش‌های یادگیری ماشین، الگوریتم‌هایی که براساس داده‌ها توانایی یادگیری و پیش‌بینی دارند، ایجاد و مورد بررسی قرار می‌گیرد. چنین الگوریتم‌هایی صرفا از دستورهای برنامه پیروی نمی‌کنند بلکه از طریق مدل‌سازی داده‌های ورودی نمونه، پیش‌بینی یا تصمیم‌گیری می‌کنند.

از آنجا که آگاهی ضرورت انکارناپذیر تصمیم‌گیری است، پیش‌بینی تقلب در صورت‌های مالی می‌تواند از تحمیل هزینه‌های زیاد، کاهش اعتبار شرکت و در نتیجه افزایش ریسک و ناکارآمدی بازار جلوگیری کند. بنابراین ضرورت اصلی این پژوهش، ارتقای سطح آگاهی حسابرسان، نهادهای ناظر و سرمایه‌گذاران در این زمینه است. آنچه این پژوهش را از پژوهش‌های مشابه متمایز می‌کند، به شرح زیر است:

- برای پیش‌بینی گزارشگری مالی متقلبانه، از ۲۰ متغیر در قالب پنج ضلعی تقلب با تاکید بر ساختار کنترل‌های داخلی (فشار، فرصت، توجیه، قابلیت، تکبر و ساختار کنترل‌های داخلی) که شامل نسبت‌های مالی، متغیرهای غیرمالی و متغیرهای حاکمیت شرکتی می‌باشد استفاده شده است؛
- برای پیش‌بینی گزارشگری مالی متقلبانه، از ۲۷ مدل مختلف شامل ۸ مدل آماری و ۱۹ مدل یادگیری ماشین در نرم‌افزارهای داده‌کاوی(SPSS Modeler 18 و RapidMiner) و

نرم‌افزارهای آماری (نسخهٔ SPSS 22، نسخهٔ Eviews 9 ونسخه Minitab 18) به‌صورت مقایسه‌ای استفاده شده است.

باتوجه به توانایی‌های این مدل‌ها، انتظار می‌رود از آن برای موضوع پیچیده کشف تقلب در صورت‌های مالی استفاده شود، زیرا گزارشگری مالی متقلبانه تأثیر منفی بر بقا و عملکردهای پایدار بازارهای مالی، جامعه و اقتصاد جهانی دارد و به توسعه پایدار شرکت‌ها آسیب می‌رساند. بنابراین نیاز به ابزار هوشمند و جدید برای فعال کردن حسابرسان در کشف داده‌های جعلی بسیار ضروری است. به‌علاوه، گزارشگری مالی متقلبانه به متغیرهای مالی و غیرمالی زیادی بستگی داشته که خود این متغیرها مستقل از هم بوده و روابط بین آن‌ها مشخص نیست. در چنین شرایطی کشف تقلب برای حسابرسان، نهادهای ناظر و سرمایه‌گذاران دشوار خواهد بود.

ساختار مقاله حاضر بدین صورت است که در بخش اول به مبانی نظری و پیشینه پژوهش پرداخته می‌شود. در ادامه، به توصیف روش پژوهش و متغیرهای پژوهش پرداخته و سپس، روش تجزیه و تحلیل داده‌ها و یافته‌های پژوهش مطرح می‌شود. در آخر، نتیجه‌گیری، پیشنهادها و محدودیت‌های پژوهش بیان می‌گردد.

مبانی نظری و پیشینه تجربی پژوهش

روش‌های حسابرسی به دلیل اینکه برای کشف تقلب در صورت‌های مالی طراحی نشده‌اند دارای نقایص بی‌شماری برای شناسایی این نوع تقلب هستند. اگرچه از نظر اخلاقی، مدیر مسئولیت کشف داده‌های تقلب مالی در یک سازمان را بر عهده دارد ولی در واقع اکثر تقلب‌ها در صورت‌های مالی با آگاهی یا رضایت مدیریت انجام می‌شود. نکته مهم اینست که هرگونه عدم موفقیت در کشف گزارشگری مالی متقلبانه می‌تواند به اعتبار حرفه حسابرسی آسیب جدی وارد کند (چوی[۸] و پیک[۹]، ۲۰۱۳) همچنین، شیوه‌های حسابرسی باید به موقع انجام شود تا از بروز موارد تقلب در صورت‌های مالی و افزایش روزافزون آن جلوگیری شود. تکنیک های جدید مانند داده‌کاوی که دارای طبقه-بندی پیشرفته و قابلیت پیش بینی است می تواند برای تسهیل نقش حسابرسان از لحاظ کشف موفقیت آمیز تقلب مورد استفاده قرار گیرد (لین[۱۰] و همکاران، ۲۰۱۵).

داده‌کاوی فرآیندی است که در آن از تکنیک‌های مختلف برای استخراج دانش از حجم انبوهی از داده‌ها استفاده می‌شود. داده‌کاوی شامل مجموعه‌ای از تکنیک‌هایی است که در حوزه‌های دیگر علمی مانند پایگاه داده‌ها، آمار، یادگیری ماشین، شبکه‌های عصبی، بازیابی اطلاعات و تشخیص الگو می‌توان آن را یافت. مراحل مختلف داده‌کاوی شامل انتخاب مناسب‌ترین روش‌ها (مانند طبقه-بندی، رگرسیون، خوشه بندی[۱۱] یا قوانین انجمنی[۱۲]) و سپس انتخاب الگوریتم مناسب متعلق به

یکی از روش‌های مذکور است. نهایتا، با انتخاب و تنظیم پارامترهای اصلی و مرحله اعتبارسنجی، از الگوریتم انتخاب شده برای حل مسئله استفاده می‌شود(ژانگ و ژو[۱۳]، ۲۰۰۴). به طور کلی، روشهای داده‌کاوی به دو نوع توصیفی و پیش بینی تقسیم می‌شوند. دسته‌ی اول به توصیف خصوصیت داده‌ها در پایگاه داده‌ها می‌پردازند و عملیات دسته‌ی دوم با مدل‌سازی داده‌های در دسترس سعی می‌کنند تا یک پیش‌بینی صحیح از داده‌های آتی و آزمایشی داشته باشند. روش‌های پیش‌بینی نیز خود به روش‌های آماری و روش‌های نمادین طبقه بندی می‌شوند. روش‌های آماری با ارائه دانش از طریق مدل‌های ریاضی بهمراه محاسبات مربوطه شناخته می‌شوند. روش‌های نمادین دانش را با استفاده از نمادها و قوانین انجمنی ارائه می‌دهند، که در نهایت مدل‌های قابل تفسیرتری ارائه می-کنند. (اسماعیلی، ۱۳۹۸)

مدل‌های رگرسیون از جمله مدل‌های کلاسیک هستند که در گروه روش‌های آماری قرار گرفته و به نوعی به معادلات مدل‌سازی نیاز دارند. رگرسیون خطی، درجه دوم و لجستیک از مشهورترین مدل‌های داده‌کاوی هستند. ویژگی اصلی این مدل‌ها اینست که الزامات اساسی را بر داده‌ها تحمیل کرده و از همه کارکرد آنها چه مفید باشد یا نباشد، استفاده می‌کنند.

افزون بر این‌ها، شبکه‌های عصبی مصنوعی، یکی از قدرتمندترین مدل‌های ریاضی است که تقریباً برای کلیه کارهای داده‌کاوی مناسب است. شبکه‌های عصبی از لحاظ یادگیری در دو دسته‌ی شبکه‌های وزن ثابت و شبکه‌هایی با وزن متغیر تقسیم می‌شوند. شبکه‌هایی با وزن متغیر که به نام شبکه‌های یادگیرنده نیز شناخته می‌شوند، خود می‌توانند به دو دسته هدایت‌شده[۱۴] و هدایت نشده[۱۵] تقسیم می‌شوند. به منظور آموزش شبکه روش‌های بسیار زیادی وجود دارند که یکی از مشهورترین آن الگوریتم پس انتشار خطا است که یکی از روش‌های هدایت شده تلقی می‌شود. واژه ی پس انتشار به معنای این است که خطاها به سمت عقب در شبکه تغذیه می‌شوند تا وزن ها را اصلاح کنند. در این الگوریتم ابتدا فرض بر این است که وزن‌های شبکه بطور تصادفی انتخاب شده-اند (اسماعیلی، ۱۳۹۸).

روش‌های پیش‌بینی کننده عمدتاً به یادگیری هدایت‌شده نسبت داده می‌شود. روش‌های هدایت شده، روابط بین صفات ورودی و یک ویژگی مشخص را در ساختاری که "مدل" نامیده می‌شود، آشکار می‌کنند. مسائل رگرسیون و طبقه‌بندی هر دو جز روش‌های هدایت شده هستند. سناریوی هدایت شده به گونه‌ای عمل می‌کند که یک مدل در یک مجموعه داده آموزشی قرار گیرد و سپس برای پیش‌بینی موارد مشاهده نشده استفاده شود. در اینجا، هدف اصلی نقشه برداری ورودی‌ها به خروجی است که مقادیر صحیح آن توسط یک ناظر تعیین می‌شود. از طرف دیگر، چنین ناظری در یادگیری بدون هدایت وجود ندارد و فقط داده‌های ورودی در دسترس است. یکی از مزایای

یادگیری بدون هدایت این است که، برخلاف یادگیری هدایت شده، مدل‌های پیشرفته‌تری می-توانند تحت آموزش قرار بگیرند. از آنجا که در یادگیری هدایت شده، هدف این است که رابطه بین دو مجموعه مشاهدات برقرار شود، با افزایش تعداد مراحل، فرایندهای یادگیری به دلیل هزینه‌های بالای محاسباتی بصورت تصاعدی رشد می‌کند و بنابراین این مدل‌ها نمی‌توانند بطور ریشه‌ای تحت فرآیند آموزش قرار بگیرند. (امیدی و همکاران، ۲۰۱۹).

مزیت بزرگ ماشین‌بردار پشتیبان این است که قدرت روش‌های آماری تئوری محور و روش‌های یادگیری ماشین داده محور را ترکیب می‌کند (مین و لی[۱۶]، ۲۰۰۵). ایده اساسی این الگوریتم آنست که بردارهای ورودی به صورت غیرخطی در فضای ویژگی که بُعد بسیار بالایی دارد، رسم شود. براساس این معیار که اگر بردارها بتوانند سخت‌ترین نقاط را مشخص کنند، دیگر نقاط به آسانی مشخص می‌شوند، ماشین بردار پشتیبان برخلاف روش‌های بی‌شماری که بر کل داده‌ها تمرکز می‌کنند به دشوارترین مسأله تشخیص داده شده توسط نقاط توجه می‌کند. بردارهایی که تشخیص آنها دشوارتر از همه است نزدیک به ابرصفحه‌ای قرار می‌گیرند، که اصطلاحا بردار پشتیبان نامیده می‌شوند. فاصله موجود بین نزدیکترین نقاط داده در هر طبقه، نسبت به ابرصفحه حاشیه نامیده می‌شود. هدف این الگوریتم حداکثر کردن این حاشیه‌ها است، بنابراین هر اندازه بردارها دورتر از ابرصفحه باشند، اطمینان از درستی طبقه‌بندی آنها بیشتر می‌شود. ماشین‌های بردار پشتیبان زمانی‌که داده‌ها به صورت خطی از هم تفکیک شوند بسیار خوب عمل می‌کنند.از سوی دیگر بر خلاف مدل‌های رگرسیون، این مدل‌ها معمولاً نیازی به ایجاد تعامل بین متغیرها ندارند، مشابه شبکه‌های عصبی مصنوعی، نسبت به پارازیت‌ها مناسب هستند (قاضی زاده احسایی و همکاران، ۱۳۹۸).

داده کاوی به طور گسترده در بسیاری از حوزه‌ها کاربرد دارد. در زیر خلاصه‌ای از کاربردهای داده کاوی آورده شده است:

الف) برنامه های داده کاوی در امور مالی: پولشویی و کشف سایر جرایم مالی، طبقه‌بندی مشتریان و بازاریابی هدف، پیش‌بینی پرداخت وام و تجزیه و تحلیل اعتبار مشتری.

ب) کاربرد داده‌کاوی در صنعت خرده فروشی: تجزیه و تحلیل اثربخشی نمایندگی فروش ، حفظ مشتری و تجزیه و تحلیل وفاداری مشتری ، سفارش محصول و.

ج) داده کاوی برای صنعت ارتباطات از راه دور: تجزیه و تحلیل الگوهای تقلب و شناسایی الگوهای غیر معمول، و خدمات ارتباط از راه دور.

در مطالعات گسترده‌ای از روش‌های نوآورانه برای شناسایی وکشف تقلب استفاده شده است. به عنوان مثال (چاندولا[۱۷] و همکاران، ۲۰۰۹) در مورد کاربرد تکنیک‌های مختلف برای تشخیص

ناهنجاری و تقلب با استفاده از رویکردهای هدایت شده داده‌کاوی بحث کردند. آنها از تکنیک‌های نتیجه محور مانند ماشین بردار پشتیبان برای انجام طبقه‌بندی استفاده کردند و دریافتند که در بسیاری از موارد مدل‌ها در مقایسه با حسابرسان در کشف سریع بدون انجام رسیدگی کارایی بالاتری دارند. (شاو[۱۸] و همکاران، ۲۰۱۲) با مرور تکنیک‌های مورد استفاده برای کشف تقلب، کارایی رایج‌ترین تکنیک‌های خوشه بندی مورد استفاده در تقلب وکشف ناهنجاری‌ها را مورد بررسی قرار دادند. (ژو و کاپور،[۱۹] ۲۰۱۱) اثربخشی و محدودیت‌های تکنیک های داده کاوی مانند درخت تصمیم گیری و شبکه عصبی و شبکه های بیزین را بررسی کردند. آنها برای شناسایی تقلب در صورت‌های مالی یک چارچوب خود سازگار[۲۰] را بر اساس مدل سطح پاسخ[۲۱] در حوزه علم کشف کردند. (شرما و پانیگراهی[۲۲]، ۲۰۱۳) یک چارچوب داده‌کاوی برای کشف تقلب در صورت‌های مالی پیشنهاد دادند. (سچینی[۲۳] و همکاران، ۲۰۱۰) با استفاده از داده‌های مالی اولیه و عمومی که در دسترس عموم هستند، یک روش برای کمک به کشف گزارشگری مالی متقلبانه پیشنهاد دادند. (چن و روکو[۲۴]، ۲۰۰۹) نشان دادند که یک روش هوش مصنوعی در شناسایی و کشف تقلب بسیار خوب عمل می‌کند، و از این رو می‌تواند یک ابزار حمایتی برای متخصصان امر باشد. تجزیه و تحلیل نسبت‌ها معمول‌ترین رویکردی است که حسابرسان برای کشف تقلب به کار می‌برند. با این وجود، مشکل این رویکرد انتزاعی بودن انتخاب نسبت‌هایی است که به احتمال زیاد نشان‌دهنده تقلب هستند (هوگان[۲۵] و همکاران، ۲۰۰۸). تکنیک‌های داده‌کاوی برای کشف مفاهیم مذکور که قبلاً ناشناخته بودند کاملا کاربردی است، (لین و همکاران، ۲۰۱۵) با این‌حال تا به امروز، استفاده از این تکنیک‌ها در گزارشگری مالی متقلبانه محدود بوده است. بیشتر تکنیک‌های داده‌کاوی مورد استفاده در این زمینه تکنیک های هدایتی مانند رگرسیون لجستیک ، شبکه های عصبی ، درخت تصمیم گیری و استخراج متن هستند.

(ابطحی و همکاران، ۲۰۱۷) از مدل طبقه‌بندی بیزین سیستمی که در آن تقلب در معاملات ساختگی بازار قراردادهای آتی قابل تشخیص است استفاده کرده‌اند. برچسب‌گذاری اولیهٔ داده‌ها با استفاده از خوشه‌بندی کاـ میانگین انجام گرفته است و آزمون مدل، نشان داد مدل پیشنهادی می-تواند ۹۴/۵۵ درصد موارد تقلب را با دقت طبقه‌بندی کند.

(ژان[۲۶]، ۲۰۱۸) ۱۶۰شرکت (از جمله ۴۰ شرکت متقلب) را برای ارزیابی چندین تکنیک داده کاوی از جمله شبکه عصبی مصنوعی و ماشین بردار پشتیبان مورد استفاده قرار داد. سپس چهار نوع درخت تصمیم‌گیری (Cart ، CHAID، C5.0، و QUEST) استفاده شد. نتایج مطالعه نشان می دهد که ترکیب مدل ANN + Cart بهترین نتایج طبقه بندی را با دقت ۹۰/۸۳ درصد در کشف تقلب در اظهارات مالی ارائه می دهد.

امیدی[۲۷] و همکاران (۲۰۱۹)، با استفاده از ۱۸ نسبت مالی مدل مثلث تقلب به اثربخشی روشهای پیش بینی کننده تقلب در صورتهای مالی در شرکت‌های بورسی کشور چین تقلب پرداختند. این مطالعه با بررسی خواص پنج رویکرد هدایت شده یعنی، شبکه عصبی پیشخور چند لایه، شبکه عصبی احتمال محور[۲۸]، دستگاه بردار پشتیبانی[۲۹]، مدل چند جمله‌ای خطی[۳۰]، و تجزیه و تحلیل تفکیک کننده[۳۱]، نشان داد که شبکه عصبی پیشخور چند لایه بهترین نتایج را در تشخیص صورت‌های مالی متقلبانه ارائه می دهد.

در پژوهش‌های مربوط به ایران نیز تاراسی و همکاران (۱۳۹۸) توانایی شبکه های عصبی مصنوعی در پیش بینی احتمال گزارشگری مالی متقلبانه مورد بررسی قرار دادند. در آن تحقیق، شبکه های بسیاری با ترکیب‌های مختلف توابع انتقال طراحی شد و مشخص گردید که در خصوص مسئله‌ی پیش بینی احتمال وقوع گزارشگری مالی متقلبانه با ۱۷ ورودی و ۷ خروجی، شبکه‌ی عصبی دو لایه با توابع انتقال پایه شعاعی و خطی از صحت ۹۷/۴ درصد برخوردار بوده و بهترین تخمین تابع و پیش بینی را انجام می دهد.

با بررسی دقت پیش بینی مدیریت سود با استفاده از از شبکه عصبی پیشخور تعمیم یافته و درخت تصمیم‌گیری Cart ,C5.0 و مقایسه آن با مدل خطی رگرسیون کمترین مربعات، صالحی و فرخی (۱۳۹۷)دریافتند روش شبکه عصبی و درخت تصمیم گیری در پیش بینی مدیریت سود نسبت به روشهای خطی دقیق تر و دارای سطح خطای کمتری است.

در پژوهش آشتاب و همکاران (۱۳۹۶) نیز دقت مدل‌های پیش‌بینی بحران مالی و رویکردهای مدیریت سود مورد بررسی قرار گرفت. بدین منظور برای پیش‌بینی بحران مالی، مدل-های یادگیری ماشین و مدل‌های آماری با یکدیگر مقایسه شدند و به کمک آزمون مقایسهٔ میانگین، مشخص شد که از نظر پیش‌بینی بحران مالی، مدل‌های یادگیری ماشین نسبت به مدل‌های آماری دقت بیشتری دارند.

در پژوهش دیگر، خواجوی و ابراهیمی (۱۳۹۶) با به‌کارگیری تکنیک‌های داده‌کاوی مدلی را برای کشف تقلب در صورتهای مالی ارائه کردند. یافته‌های پژوهش بیانگر وجود شواهدی دال بر عملکرد مناسب مدل‌های پیشنهادی و برتری الگوریتم جنگل تصادفی و شبکه بیزین در مقایسه با شبکه عصبی مصنوعی برای پیش‌بینی تقلب در صورتهای مالی بود.

کاردان و همکاران (۱۳۹۶) با بررسی دقت الگوریتم‌های خطی و غیرخطی در پیش‌بینی مدیریت سود نشان دادند الگوریتم‌های غیرخطی از دقت بیشتری نسبت به الگوریتم‌های خطی برخوردار بوده و الگوریتم رگرسیون پشتیبان، مدیریت سود را بهتر از سایر الگوریتم‌ها پیش‌بینی می‌کند. همچنین الگوریتم خطی در پیش‌بینی سود نتایج تقریباً مشابهی را از خود نشان داد.

روش‌شناسی پژوهش

در این پژوهش برای پیش‌بینی بحران مالی، از ۲۷ مدل مختلف شامل ۸ مدل آماری و۱۹ مدل یادگیری ماشین به‌صورت مقایسه‌ای استفاده شده است. در این پژوهش برای تجزیه و تحلیل داده‌ها در و انجام آزمون‌های آماری از نرم‌افزارهایEviews 9, SPSS 22 وMiniTab 18 و همچنین نرم‌افزارهای داده‌کاوی RapidMiner وSPSS Modeler استفاده شده است. ویژگی بارز نرم-افزارSPSS Modeler در این است که پردازش داده‌های خود را با استفاده از گره‌هایی که به یکدیگر متصل شده‌اند و قالب یک جریان را به‌وجود می‌آورند، انجام می‌دهد. به‌علاوه، پس از اتمام فرایند داده‌کاوی، داده‌های مصور شده را به کاربر ارائه می‌دهد.

الگوریتم‌های یادگیری ماشین مورد استفاده در این پژوهش به‌شرح زیر است:

الگوریتم ماشین بردار پشتیبان: ماشین بردار پشتیبان، یکی از قوی‌ترین و دقیق‌ترین الگوریتم‌های یادگیری ماشین است که جهت طبقه‌بندی و تفکیک گروه‌ها به‌کار می‌رود. این الگوریتم روش‌های آماری و یادگیری ماشین را باهم ترکیب می‌کند. بنابراین، اساس نظری آن برپایه تئوری یادگیری آماری قرار دارد (باستی[۳۲] و همکاران، ۲۰۱۵). با مشخص شده داده‌های پژوهش، مدل ماشین بردار پشتیبان داده‌ها را به گروه‌های متمایزی تقسیم می‌کند. این مدل‌ها دارای خواص کلی طبقه‌بندی داده‌ها با حداکثر قابلیت تعمیم، رسیدن به نقطه بهینه تفکیک داده-ها، تعیین خودکار ساختار بهینه برای طبقه‌بندی کننده و امکان مدل کردن داده‌های غیرخطی با استفاده از تجزیه و تحلیل اجزای اصلی هستند (فلاح شمس و همکاران، ۱۳۹۱). ماشین بردار پشتیبان الگوریتمی است که نوع خاصی از مدل‌های خطی را می‌یابد که موجب حداکثر شدن تفکیک بین طبقات می‌شود. بردارهای پشتیبان در واقع نزدیک‌ترین نقاط به حاشیه ابرصفحه هستند و تنها از این بردارها (نقاط) برای مشخص کردن مرز بین طبقات استفاده می‌شود. این الگوریتم از تئوری بهینه‌سازی برای طبقه‌بندی استفاده می‌کند و براساس تئوری یادگیری آماری، خطای طبقه‌بندی را به حداقل می‌رساند (اسماعیلی و همکاران، ۱۳۹۷).

الگوریتم درخت تصمیم‌گیری: روش مبتنی بر درخت تصمیم، یکی از ابزارهای قوی برای دسته‌بندی و اعتباردهی محسوب می‌شود. درخت تصمیم از نظریه اطلاع و مقدار آنتروپی جهت انتخاب بهترین متغیر شروع پیمایش استفاده می‌کند. هر راس درخت به‌عنوان یک کلاس یا قاعده، نمایشگر یک آزمایش یا تصمیم یکتاست. یال‌های هر راس، متناظر با احتمال‌های حاصل از آزمایش روی راس است. پیشامد داده‌ها را به چندین زیرمجموعه افراز می‌کند که توسط برگ‌های این درخت شناخته می‌شود. درخت تصمیم برخلاف سایر فنون به تولید قانون می‌پردازد و پیش‌بینی خود را در قالب قوانین توضیح می‌دهد، در حالی که در سایر فنون، پیش‌بینی نهایی بدون چگونگی

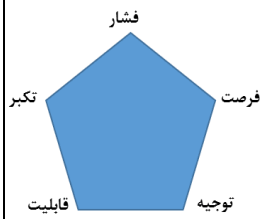
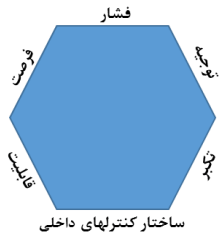
اجرا بیان می‌شود (مفاخری، ۱۳۹۷). رایج‌ترین الگوریتم‌های درخت تصمیم شامل CHAID، C4.5، C5، CART و QUEST هستند (باستی و همکاران، ۲۰۱۵). یکی از پرکاربردترین درخت‌های تصمیم، C5 است، چرا که این الگوریتم با توجه به اریب کمتر نسبت به الگوریتم‌های مشابه برای داده‌هایی که دارای نویز هستند، مناسب بوده و استفاده از روش بوستینگ[۳۳] باعث افزایش عملکرد آن می‌شود.

مدل درخت تصمیم مزایای بسیاری همانند قابل فهم بودن قوانین ایجاد شده توسط درخت تصمیم، برخورداری از کارایی بالاتر، مناسب برای نمونه‌هایی با حجم زیاد داده و دارا بودن دقت طبقه‌بندی بالاتر دارد (هان،[۳۴] ۲۰۱۶).

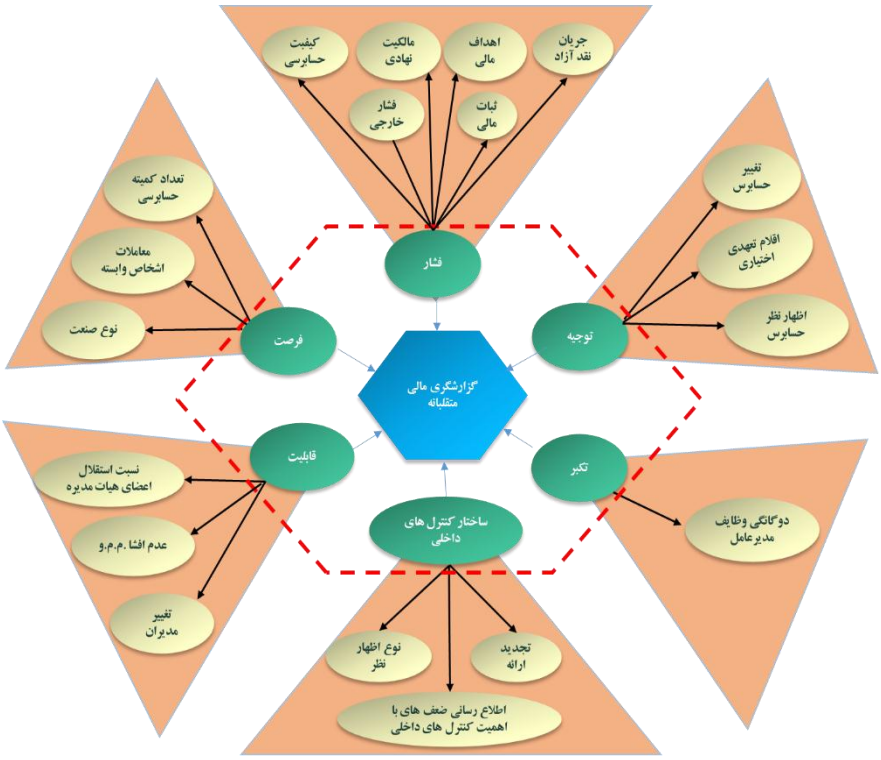
الگوریتم شبکه بیزین: شبکه‌های بیزین جزو دسته‌بندی کننده‌های آماری هستند و می‌توانند ردیف تعلق داده‌ها را مشخص کنند و در آن صفات داده‌های ورودی از هم مستقل بوده و بر روی هم تاثیر ندارند. از این روش تحت عنوان تکلیف هدایت شده یاد می‌شود. مبنای رده‌بندی در الگوریتم بیزین، احتمالات است. در واقع رده‌بندی بیزین چیزی جز احتمالات شرطی نیست. اما ویژگی بسیار مثبت الگویتم بیز این است که امکان اثبات بهینگی دارد (مفاخری، ۱۳۹۷).

مدل پژوهش

سیر تکاملی مدل‌های تقلب در صورت‌های مالی در شکل‌های ۱-۳ آورده شده است. بر این اساس، مدل اندازه‌گیری و کشف مثلث تقلب ابتدا به مدل لوزی (الماس تقلب) و نهایتا مدل پنتاگون بسط و توسعه یافته است. با توجه به مسئله اصلی این تحقیق مبنی بر تاثیر ساختار کنترل‌های داخلی در کشف تقلب، مدل بسط یافته پنتاگون در شکل ۴ نشان داده شده است. شکل ۵ نیز مدل مفهومی پژوهش را با جزئیات ارائه می‌کند.

			
شکل۱ مدل مثلث تقلب کریسی (۱۹۷۳)	**شکل ۲ مدل لوزی تقلب ولف و هرمانسون (۲۰۰۴)**	**شکل ۳ مدل پنج ضلعی تقلب کروهوروات (۲۰۱۱)**	**شکل ۴ مدل شش ضلعی تقلب**

منبع: یافته های پژوهشگر



شکل ۵. مدل مفهومی پژوهش

منبع: یافته های پژوهشگر

جامعه آماری این پژوهش کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران هستند که در بازه زمانی ۱۳۸۸ تا ۱۳۹۷ فعالیت کرده‌اند.جدول ۱ روند انتخاب نمونه آماری پژوهش را نشان می-دهد.

جدول ۱. روند انتخاب نمونه آماری پژوهش

جامعه آماری به غیر از شرکت های خارج شده از بورس	۵۲۲
کسر می‌شود: شرکت‌هایی که در طی دوره تحقیق از بورس خارج شده اند	۱۱۷
کسر می‌شود: شرکت‌هایی که عضو بانک‌ها، شرکت‌های سرمایه‌گذاری، هلدینگ‌ها و بیمه‌ها هستند	۶۵
کسر می‌شود: شرکت‌هایی که اطلاعات آنها بطور کامل در دسترس نیست	۱۲۴
کسر می‌شود: شرکت‌هایی که پایان سال مالی آنها منتهی به پایان اسفند نباشد	۵۱
نمونه آماری قابل آزمون با در نظر گرفتن پیش فرض‌ها	۱۶۶

منبع: یافته های پژوهشگر

با اعمال شرایط فوق، تعداد ۱۶۶ شرکت از شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران، جهت برآورد مدل‌ها و آزمون فرضیه‌های پژوهش انتخاب شد. فرضیه‌های تحقیق به شرح زیر در نظر گرفته شدند.

۱) بین عامل فشار و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل مالکیت نهادی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل اهداف مالی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل ثبات مالی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل فشار خارجی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل جریان نقد آزاد و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل کیفیت حسابرسی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.

۲) بین عامل فرصت و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل تعداد کمیته حسابرسی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل نوع صنعت و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل معاملات اشخاص وابسته و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.

۳) بین عامل منطق گرایی (توجیه) و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل تغییر حسابرس و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل اظهار نظر حسابرس و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل اقلام تعهدی اختیاری و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.

۴) بین عامل توانایی (قابلیت) و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین نسبت استقلال اعضای هیات مدیره و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل تغییر مدیران و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل عدم افشای مطالبات مشکوکالوصول و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.

۵) بین عامل غرور (شدت عمل) و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین عامل دوگانگی وظایف مدیر عامل و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.

۶) بین ساختار کنترلهای داخلی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین تجدید ارائه صورتهای مالی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین نوع اظهار نظر حسابرسی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.
- بین اطلاع رسانی ضعفهای با اهمیت کنترلهای داخلی و گزارشگری مالی متقلبانه رابطه معناداری وجود دارد.

متغیرهای پژوهش

گزارش مالی متقلبانه به عنوان متغیر وابسته که با استفاده از مدل(۱) F-Score مورد سنجش قرار گرفت (دیچو و همکاران،[۳۵] ۲۰۱۱) که به شرح زیر است:

Predicted Value = +7.893 – 0.790*Rsst-acc + 2.518*Chrec + 1.191*Chinv + 1.979*Softassets+ 0.171*Chcs - 0.932*Chroa + 1.029*Issue + ε+

(متوسط مجموع دارایی ها) / (*تغییرات سرمایه در گردش + تغییرات تعهدات عملیاتی غیرجاری + تغییرات تعهدات مالی*)= کیفیت اقلام تعهدی

بدهی‌های جاری – دارایی‌های جاری = *سرمایه در گردش*

(بدهی‌های بلند مدت-بدهی‌های جاری-کل بدهی‌ها) – (سرمایه‌گذاری و توسعه - دارایی‌های جاری -کل داریی‌ها) =*تعهدات عملیاتی غیرجاری*

(سهام ممتاز+ بدهی‌های کوتاه مدت + بدهی‌های بلند مدت) –(سرمایه‌گذاری کوتاه مدت + سرمایه‌گذاری بلند مدت)= *تعهدات مالی*

متوسط مجموع دارایی‌ها / تغییرات در حسابهای دریافتنی = تغییر در حسابهای دریافتنی

متوسط مجموع دارایی‌ها / تغییرات موجودی‌های کالا = تغییر درموجودی‌های کالا

کل دارایی‌ها / (وجه نقد و معادل نقد – دارایی‌های ثابت مشهود – کل دارایی‌ها) = درصد نقدینگی دارایی

تغییرات در حساب‌های دریافتنی – درآمد فروش = تغییر در فروش‌های نقدی

(متوسط مجموع دارایی‌ها در ابتدای سال/ سود سال قبل) – (متوسط مجموع دارایی‌ها در انتهای سال / سود سال جاری) = تغییرات در سود

متغییر موهومی، عدد یک در صورتیکه شرکت در سال جاری اوراق بدهی یا سهام منتشر کند و درغیر اینصورت عدد صفر = انتشار سهام

برای محاسبه F-Score، احتمال پیش‌بینی شده از طریق تقسیم e(VALUE) / (1+ e(VALUE)) بر احتمال غیر شرطی تقلب (۰,۰۰۳۷) بدست می‌آید که در آن PV ارزش پیش‌بینی

بدست آمده از الگوی (۱) می‌باشد. بر اساس پژوهش دیچو و همکاران (۲۰۱۱)، در این پژوهش مشاهدات دارای F-Score بالاتر از ۱/۸۵ به عنوان شرکت‌های با ریسک بالای گزارشگری متقلبانه شناسایی می‌شوند.

در این مطالعه متغیرهای خارجی یا مستقل شامل شش متغیر پنهان یعنی فشار، فرصت، عقلانی، قابلیت و تکبر و ساختار کنترل داخلی است. هر یک از فرایندها به شرح زیر است(توفیق اکبر ، ۲۰۱۷):

جدول ۲. تعریف عملیاتی متغیرهای پژوهش

متغیرها		تعریف عملیاتی
فشار	اهداف مالی	اندازه گیری با استفاده از بازده دارایی ها (تیفانی و مارفو[۳۶]، ۲۰۱۵) سود خالص پس از کسر مالیات تقسیم بر جمع دارایی‌ها
	ثبات مالی	اندازه گیری با استفاده از درصد کل تغییرات در دارایی‌ها (تیفانی و مارفو، ۲۰۱۵) جمع دارایی‌های سال جاری منهای دارایی‌های سال قبل تقسیم بر دارایی‌های سال جاری
	فشار خارجی	اندازه گیری با استفاده از اهرم مالی (تیفانی و مارفو، ۲۰۱۵) جمع بدهی‌ها تقسیم بر جمع دارایی‌ها
	مالکیت نهادی	اندازه گیری با استفاده از درصد مالکیت نهادی (تسا و هارتو[۳۷]، ۲۰۱۶) متغیر ساختگی عدد '۱' بیشتر از میانه نشان دهنده مالکیت نهادی و کد '۰' در غیر این صورت.
	جریان نقد آزاد	اندازه‌گیری با استفاده از جریان نقد آزاد (چانگ[۳۸] و همکاران، ۲۰۰۵) (سود عملیاتی قبل از کسر هزینه استهلاک – سود پرداختی به سهامداران – مالیات بر درآمد)/ ارزش دفتری دارایی‌ها
	کیفیت حسابرسی	اندازه‌گیری با استفاده از رتبه بندی شرکت‌های حسابرسی (اوکیف[۳۹] و همکاران، ۱۹۹۴ و لونسوهن[۴۰] و رک[۴۱]، ۲۰۰۴) متغیر ساختگی عدد '۱' نشان دهنده رتبه الف و ب شرکت‌های حسابرسی و کد '۰' در غیر این صورت.
فرصت	تعداد اعضای کمیته حسابرسی	با استفاده از تعداد کمیته های حسابرسی اندازه گیری شد (امالیا و همکاران[۴۲]، ۲۰۱۵) تعداد کل اعضای کمیته های حسابرسی
	نوع صنعت	اندازه گیری با استفاده از نسبت کل موجودی (آنیسیا و همکاران، ۲۰۱۶) (موجودی کالای طی سال/ فروش طی سال) - (موجودی کالای سال قبل/ فروش سال قبل)
	معاملات اشخاص وابسته	متغیر ساختگی عدد '۱' نشان دهنده معاملات اشخاص وابسته و کد '۰' در غیر این صورت.
توجیه	تغییر حسابرسان	اندازه گیری با استفاده از تغییر در حسابرس (تیفانی و مارفو، ۲۰۱۵) متغیر ساختگی عدد '۱' نشان دهنده تغییر در حسابرس و کد '۰' در غیر این صورت.
	نظر حسابرس	اندازه گیری با استفاده از نظر که در گزارش حسابرسی نشان داده شده است (امالیا و همکاران، ۲۰۱۵) متغیر ساختگی عدد '۱' نشان دهنده اظهار نظر مقبول و کد '۰' در غیر این صورت.
	اقلام تعهدی اختیاری	اندازه‌گیری با استفاده از مدل تعدیل شده جونز

متغیرها		تعریف عملیاتی
		مقادیر خطای مدل به صورت بهتری اقلام تعهدی اختیاری را نشان می‌دهد
قابلیت	تغییر مدیران	اندازه گیری با استفاده از تغییر در ساختار مدیران (آنیسیا و همکاران، ۲۰۱۶) متغیر ساختگی عدد '۱' نشان دهنده تغییر در ساختار مدیران و کد '۰' در غیر این صورت.
	نسبت استقلال اعضای هیئت مدیره	اندازه گیری شده با استفاده از نسبت استقلال اعضای هیئت مدیره (ایندارتو و گوزالی[۴۳]، ۲۰۱۶) تعداد اعضای مستقل هیئت مدیره تقسیم بر تعداد کل اعضای هیئت مدیره
	عدم افشاء مطالبات مشکوک الوصول	متغیر ساختگی عدد'۱' نشان دهنده افشاء مطالبات مشکوک الوصول و کد '۰' در غیر این صورت.
تکبر	دوگانگی وظایف مدیر عامل	اندازه گیری با استفاده از دوگانگی وظایف مدیر عامل (یوسف و همکاران، ۲۰۱۵) متغیر ساختگی عدد'۱' نشان دهنده اینست که مدیر عامل شرکت دارای وظایف دوگانه است و کد '۰' در غیر این صورت.
ساختار کنترل داخلی	نوع اظهارنظر حسابرس	نوع اظهار نظر حسابرس (کرمی و همکاران، ۱۳۹۷) متغیر ساختگی عدد'۱' نشان دهنده اینست که اظهارنظر حسابرس مقبول باشد و کد '۰' در غیر این صورت.
	عدم تجدید ارائه صورت-های مالی	عدم تجدید ارائه صورت‌های مالی(کرمی و همکاران، ۱۳۹۷) متغیر ساختگی عدد'۱' نشان دهنده اینست که عدم تجدید ارائه صورت‌های مالی باشد و کد '۰' در غیر این صورت.
	اطلاع رسانی ضعف های بااهمیت ساختار کنترل داخلی	اطلاع رسانی ضعف های بااهمیت ساختار کنترل داخلی(کرمی و همکاران، ۱۳۹۷) متغیر ساختگی عدد'۱' نشان دهنده اینست که عدم تجدید ارائه صورت‌های مالی باشد و کد '۰' در غیر این صورت.

منبع: یافته های پژوهشگر

تحلیل یافته‌ها

آمار توصیفی متغیرهای پژوهش در جدول زیر آورده شده است.

جدول ۳. آمار توصیفی متغیرهای پژوهش

متغیرهای اصلی	متغیرهای فرعی	میانگین	میانه	انحراف معیار	کمینه	بیشینه
فشار	اهداف مالی	۰.۰۹۴	۰.۰۸۷	۰.۱۷۱	-۱.۱۰۰	۲.۱۰۱
	ثبات مالی	۰.۱۰۹	۰.۱۰۲	۰.۱۸۵	-۱.۹۴۰	۰.۸۷۷
	فشار خارجی	۰.۶۸۳	۰.۶۴۵	۰.۴۸۳	۰.۰۴۷	۹.۰۰۳
	جریان نقد آزاد	۰.۰۷۳	۰.۰۵۸	۰.۱۴۹	-۰.۷۴۵	۱.۲۲۶
فرصت	تعداد اعضای کمیته حسابرسی	۱.۳۱۹	۰.۰۰۰	۱.۵۹۸	۰.۰۰۰	۵.۰۰۰
	نوع صنعت	۰.۳۵۵	-۰.۰۰۱	۳۸.۲۸۳	-۷۸۹.۲۶۸	۱۳۴۴.۴۳۵

متغیرهای اصلی	متغیرهای فرعی	میانگین	میانه	انحراف معیار	کمینه	بیشینه
توجیه	اقلام تعهدی اختیاری	۰.۰۰۰	-۰.۰۰۵	۰.۱۳۷	-۰.۹۴۶	۰.۸۶۲
قابلیت	تغییر مدیران	۰.۶۲۷	۱.۰۰۰	۰.۴۸۸	۰.۰۰۰	۲.۰۰۰
متغیرهای صفر و یک		صفر			یک	
		تعداد	درصد		تعداد	درصد
F-Score	گزارشگری مالی متقلبانه	۱۳۱۶	۷۹		۳۴۴	۲۱
فشار	کیفیت حسابرسی	۱۳۷	۸		۱۵۲۳	۹۲
	مالکیت نهادی	۲۴۱	۱۴,۵		۱۴۱۹	۸۵,۵
فرصت	معاملات اشخاص وابسته	۱۲	۰/۸		۱۶۴۸	۹۹,۲
توجیه	تغییر حسابرس	۶۴۶	۳۹		۱۰۱۴	۶۱
	اظهار نظر حسابرس	۹۲۵	۵۶		۷۳۵	۴۴
قابلیت	استقلال اعضای هیئت مدیره	۳۶۰	۲۲		۱۳۰۰	۷۸
	عدم افشاءم.م.و	۵۱۲	۳۱		۱۱۴۸	۶۹
تکبر	دوگانگی وظایف مدیرعامل	۱۰	۰/۶		۱۶۵۰	۹۹,۴
ساختار کنترل‌های داخلی	تجدید ارائه	۳۲۳	۱۹,۵		۱۳۳۷	۸۰,۵
	اطلاع رسانی ضعف	۵۲۰	۳۱		۱۱۴۰	۶۹
	نوع اظهار نظر	۹۲۵	۵۶		۷۳۵	۴۴

منبع: یافته های پژوهشگر

آمار استنباطی: این تحقیق از روش F-score در داده‌های پانل به منظور شناسایی شرکت‌ها به دو گروه متقلب و غیرمتقلب استفاده می‌کند. جهت بررسی روابط میان متغیرها از مدل‌های لاجیت، پروبیت و مدل ارزش حدی استفاده شده است. بدین‌ترتیب که با استفاده مدل‌های مذکور یک الگوی رگرسیون تخمین زده شده و مورد تجزیه و تحلیل قرار می‌گیرد (جدول ۴).

جدول ۴. نتایج آزمون فرضیات اصلی و فرعی

نام متغیرها	لاجیت		پروبیت		مقدار حدی	
	ضریب	آماره Z	ضریب	آماره Z	ضریب	آماره Z
فشار	۵۶۹۰۹.۰	***۸۴۷۷۸.۷	۳۳۶۹۴.۰	***۱۱۲۸۷.۸	۳۱۷۷۹.۰	***۰۲۷۳۳.۸
فرصت	۱۵۰۴۵.۰	*۹۱۷۳۰.۱	۰۸۱۵۴.۰	*۹۲۵۹۶.۱	۰۶۵۶۵.۰	**۹۵۹۹۹.۱
توجیه	۱۸۰۴۲.۰	***۰۳۴۳۴.۳	۱۰۴۳۵.۰	***۰۳۶۰۳.۳	۰۹۲۴۳.۰	***۰۳۳۲۶.۳
قابلیت	۱۲۹۵۲.۰	**۱۱۴۸۷.۲	۰۷۴۵۳.۰	**۱۲۶۱۹.۲	۰۶۵۴۹.۰	**۱۴۲۳۴.۲
تکبر	۰۴۴۸۵.۰	۰۵۶۵۶.۰	۰۲۵۷۰.۰	۰۵۶۷۱.۰	۰۲۲۴۴.۰	۰۵۶۹۳.۰
ساختار کنترل‌های داخلی	۱۶۷۶۱.۰	***۸۶۶۸۷.۲	۰۹۷۳۷.۰	***۸۵۴۲۴.۲	۰۸۶۹۳.۰	***۸۲۵۷۹.۲

نام متغیرها		لاجیت		پروبیت		مقدار حدی	
		ضریب	آماره Z	ضریب	آماره Z	ضریب	آماره Z
فشار	مالکیت نهادی	۰۵۱۴۰.۰	۲۵۵۷۳.۰	۰۲۲۷۱.۰	۲۰۰۵۸.۰	۰۰۸۷۶.۰	۰۸۸۸۶.۰
	اهداف مالی	۱۷۱۵۱.۱	۴۸۹۴۰.۱	۶۸۷۷۱.۰	۴۹۶۵۰.۱	۶۴۰۵۳.۰	۴۸۷۱۹.۱
	ثبات مالی	۷۲۳۳۵.۴	***۶۹۳۱۰.۸	۷۱۱۸۴.۲	***۶۵۳۲۰.۸	۳۸۰۵۶.۲	***۲۴۷۴۵.۸
	فشار خارجی	۸۹۸۶۹.۰-	**۰۹۹۱۰.۲-	۵۰۷۹۴.۰-	**۱۱۴۵۲.۲-	۴۲۹۵۷.۰-	**۰۸۱۴۱.۲-
	جریان نقد آزاد	۹۶۵۷۱.۱-	**۲۵۱۱۰.۲-	۱۶۰۳۱.۱-	**۳۰۲۶۰.۲-	۰۷۶۱۸.۱-	**۳۷۰۳۸.۲-
	کیفیت حسابرسی	۰۴۶۰۰.۰	۱۸۹۰۰.۰	۰۲۶۱۶.۰	۱۸۸۴۶.۰	۰۲۰۱۸.۰	۱۶۵۰۴.۰
فرصت	اعضای کمیته حسابرسی	۰۹۳۰۱.۰	**۵۵۸۲۷.۲	۰۵۴۲۷.۰	**۵۵۳۷۳.۲	۰۴۸۸۶.۰	**۵۳۶۸۲.۲
	نوع صنعت	۰۳۴۹۶.۰	۰۵۱۶۳.۰	۰۱۱۱۷.۰	۰۲۸۳۶.۰	۰۰۳۳۷.۰-	۰۰۹۵۳.۰-
	معاملات اشخاص وابسته	۳۰۵۶۰.۰	۳۹۹۹۶.۰	۱۷۶۳۴.۰	۴۱۹۳۷.۰	۱۵۴۵۳.۰	۴۴۷۴۲.۰
توجیه	تغییر حسابرس	۰۱۲۹۱.۰	۰۹۶۹۷.۰	۰۱۲۰۶.۰	۱۵۸۳۱.۰	۰۱۰۱۵.۰	۱۵۲۰۵.۰
	اظهارنظر حسابرس	۰۰۸۹۵.۰	۰۷۲۱۶.۰	۰۱۹۸۵.۰	۲۷۹۸۶.۰	۰۲۶۹۹.۰	۴۳۳۴۱.۰
	اقلام تعهدی اختیاری	۱۵۹۰۰.۳	***۳۰۰۴۶.۶	۲۸۰۱۵.۲	***۵۹۱۸۲.۶	۹۴۸۴۵.۱	***۵۷۴۶۵.۶
قابلیت	استقلال هیئت مدیره	۲۶۹۴۹.۰	*۷۲۳۳۴.۱	۱۵۳۵۷.۰	*۷۴۶۳۷.۱	۱۳۳۰۴.۰	*۷۷۸۶۵.۱
	تغییر مدیران	۱۳۴۴۰.۰	۰۵۱۳۶.۱	۰۷۷۴۴.۰	۰۶۲۱۱.۱	۰۶۸۱۹.۰	۰۷۷۷۲.۱
	عدم افشاء م. م. و	۰۲۸۴۷.۰	۲۱۵۲۳.۰	۰۱۷۲۷.۰	۲۲۷۷۷.۰	۰۱۶۳۷.۰	۲۴۶۹۵.۰
تکبر	دوگانگی وظایف مدیرعامل	۰۴۴۸۵.۰	۰۵۶۵۶.۰	۰۲۵۷۰۰.۰	۰۵۶۷۱۴.۰	۰۲۲۴۴.۰	۰۵۶۹۴.۰
ساختار کنترل‌های داخلی	تجدید ارائه	۱۹۶۰۳.۰-	۳۱۷۷۲.۱-	۱۱۳۱۲.۰-	۳۰۲۵۹.۱-	۰۹۹۸۱.۰-	۲۷۶۰۳.۱-
	نوع اظهار نظر	۰۰۵۰۰.۰-	۰۴۰۶۴.۰-	۰۰۰۶۱.۰-	۰۰۸۶۳.۰-	۰۰۲۵۸.۰	۰۴۱۷۵.۰
	اطلاع‌رسانی ضعف‌ها	۴۰۴۵۵.۰	***۳۰۴۷۰.۳	۲۳۲۳۸.۰	***۳۰۷۶۰.۳	۲۰۴۰۲.۰	***۲۹۸۸۵.۳

*- تایید فرضیه با احتمال ۹۰٪ ، **- تایید فرضیه با احتمال ۹۵٪ و ***- تایید فرضیه با احتمال ۹۹٪

منبع: یافته های پژوهشگر

نتایج بدست آمده از هر سه مدل مورد مطالعه برای متغیرهای اصلی پژوهش، نشان داد همه عوامل مدل توسعه یافته‌ی پنتاگون به جز تکبر تاثیر مثبت ومعناداری بر گزارشگری مالی متقلبانه در شرکتهای بورسی دارند. عوامل فشار، توجیه و ساختار کنترل‌های داخلی با ۹۹ ٪ اطمینان، متغیر مستقل قابلیت با ۹۵ ٪ اطمینان و متغیر فرصت با ۹۰ ٪ اطمینان، تاثیر موثر و معنادار بر روی گزارشگری مالی متقلبانه دارند. با استفاده از مدل مقدار حدی تاثیر عامل فرصت نسبت به دو مدل دیگر با درصد اطمینان بیشتری (٪۹۵) تایید می‌شود (جدول ۴).

همچنین نتایج مدل نشان دهنده آن است که بین متغیرهای فرعی فشار شامل ثبات مالی با ۹۹ ٪ اطمینان و فشار خارجی و جریان نقدآزاد با ۹۵ ٪ اطمینان با احتمال تقلب رابطه معناداری وجود دارد. همچنین با توجه به ضرایب منفی بدست آمده برای متغیرهای فشار خارجی و جریان نقد آزاد می‌توان نتیجه گرفت که یک رابطه منفی بین این متغیرها و گزارشگری مالی متقلبانه وجود دارد. علاوه بر اینها تعداد اعضای کمیته حسابرسی با ۹۵ ٪ اطمینان، اقلام تعهدی اختیاری با ۹۹ ٪ اطمینان و استقلال اعضای هیئت مدیره با ۹۰ ٪ اطمینان و نهایتا فرضیه اطلاع‌رسانی ضعف‌های ساختار کنترل‌های داخلی با ۹۹ ٪ اطمینان بر روی گزارشگری مالی متقلبانه تاثیر دارند.از سوی دیگر ارتباطی بین فرضیه فرعی عامل تکبر (دوگانگی وظایف مدیر عامل) و گزارشگری مالی متقلبانه مشاهده نشد (جدول ۴).

از آنجایی‌که همه متغیرهای فرعی مدل نهایی باید معناداری لازم را داشته باشـــند بنابراین متغیرهایی که معنادار نبوده از مدل حذف می‌شوند. در مرحله بعد، برای بررسی دقت پیش‌بینی مدل‌های طبقه‌بندی، داده‌های شرکت‌های نمونه آماری به‌عنوان ورودی وارد نرم‌افزار و نتایج پیش-بینی برای مدل‌ها محاسبه شد. برای ارزیابی همه‌جانبه توانمندی مدل‌ها معیار دقت کل (به‌صورت درصدی از اطلاعاتی که به‌طور صحیح توسط مدل پیش‌بینی می‌شوند) محاسبه شد که این امر برای هریک از مدل‌ها به‌طور جداگانه محاسبه و در جدول ۵ آورده شده است. بر این اساس، الگوریتم‌های مدل درخت CHAID، درخت تصمیم‌گیری C5 و مدل درخت C&R به ترتیب دارای بالاترین دقت پیش‌بینی ۸۰/۳۱، ۸۰/۳۰ و ۸۰/۲۴ ٪ هستند. از سوی دیگر همانطور که در جدول ۵ نشان داده شده است مدل‌های آماری پروبیت و گامبیت هم دقت بالایی در حدود ۸۰٪ داشتند که بیانگر کارایی نسبتا بالای این مدل‌های آماری در پیش‌بینی گزارشگری مالی متقلبانه می‌باشد.

جدول ۵. نتایج دقت پیش‌بینی برای مدل‌های آماری و یادگیری ماشین

مدل طبقه بندی	دقت پیش بینی (٪)	AUC
درخت CHAID	۸۰/۳۱	۰/۷۰۶
مدل درخت تصمیم گیری C5	۸۰/۳۰	۰/۵۷۵
مدل درخت R&C	۸۰/۲۴	۰/۶۴
پروبیت	۷۹/۷۲	۰/۰۰۰
گامپیت	۷۹/۵۹	۰/۰۰۰
تحلیل ممیزی خطی	۷۹/۵۲	۰/۰۰۰
رگرسیون لجستیک بر مبنای SVM	۷۹/۵۲	۰/۵۶۸
شبکه عصبی تعدیل شده بر مبنای ترکیب الگوریتم ژنتیک و بهینه سازی تصادفی	۷۹/۵۰	۰/۶۲۳
ماشین بردار پشتیبان خطی	۷۹/۲۸	۰/۴۴۶
درخت تصادفی	۷۹/۲۸	۰/۵
مدل درخت تصمیم‌گیری Quest	۷۹/۲۸	۰/۵
درخت تصمیم گیری	۷۸/۱۳	۰/۶۱۶
جنگل تصادفی	۷۷/۷۱	۰/۶۳۹
رگرسیون لجستیک	۷۷/۳۱	۰/۶۴۶
شبکه عصبی	۷۶/۵۱	۰/۶۳۷
KNN	۷۴/۰۴	۰/۸۲۵
مدل ماشین بردار پشتیبان کتابخانه ای	۷۳/۸۶	۰/۶۲۳
مدل شبکه بیزین (Bayesian Network)	۷۲/۹۳	۰/۷۰۳
مدل ماشین بردار پشتیبان (SVM)	۷۲/۴۱	۰/۶۶۹
تحلیل ممیزی تعمیم یافته	۷۱/۹۹	۰/۰۰۰
مدل Decision List	۶۸/۶۸	۰/۶۵۹
ماشین بردار پشتیبان با بهینه سازی توده ای ذرات	۶۲/۵۳	۰/۵۵۹
یادگیری عمیق	۶۲/۴۵	۰/۶۱۲
شبکه عصبی پرسپترون	۶۲/۲۵	۰/۴۷۸
ماشین بردار پشتیبان تکاملی	۵۹/۶۴	۰/۶۰۵
رگرسیون لجستیک تکاملی	۵۰/۲۰	۰/۵۱۸
تحلیل ممیزی درجه دوم	۴۷/۲۳	۰/۰۰۰

منبع: یافته های پژوهشگر

با توجه به این موضوع، مقایسه کلی بین مدل‌های آماری و همچنین مدل‌های یادگیری ماشین انجام شد. بر این اساس متوسط دقت پیش بینی مدل های آماری ۷۰/۶۴ ٪ و برای مدل های یادگیری ماشین ۷۳/۶۵ ٪ بدست آمد (جدول ۶). همچنین نتایج آزمون مقایسه نسبت‌ها برای مدل‌ها با استفاده از مقدار آماره **z** (۴/۷۵-) نشان می‌دهد که دقت پیش بینی مدل‌های آماری کمتر از مدل‌های یادگیری ماشین بوده و با توجه به سطح معنی‌داری ۰/۰۰۰ این تفاوت از نظر آماری معنی‌دار است. به عبارت دیگر، دقت پیش بینی مدل‌های یادگیری ماشین به صورت معنی-داری از مدل‌های آماری بیشتر است.

جدول ۶. مقایسه دقت پیش‌بینی مدل‌های آماری و یادگیری ماشین

نوع مدل	دقت پیش‌بینی (٪)	آزمون مقایسه نسبت‌ها (آماره z)	سطح معنی-داری	آماره F
مدل‌های آماری	۷۰/۶۴	- ۴/۷۵	۰/۰۰۰	۰/۰۰۰
مدل‌های یادگیری ماشین	۷۳/۶۵			

منبع: یافته های پژوهشگر

در ادامه برای بهینه‌سازی پیش‌بینی تقلب در صورت‌های مالی، ترکیب الگوریتم‌های مدل درخت CHAID ، درخت تصمیم‌گیری C5 و مدل درخت C&R که بالاترین دقت پیش‌بینی را داشتند انجام شد. براساس نتایج طی فرآیند آموزش و آزمایش ، ترکیب این مدل‌ها به ترتیب دقت پیش-بینی ۹۴/۷۶٪ و ۹۲/۶۱ ٪ بدست آمد (جدول ۷).

جدول۷. بهینه سازی سه مدل برتر یادگیری ماشین

پیش بینی / طبقه بندی	آموزش	آزمایش
تعداد پیش بینی صحیح	۱۰۸۶	۴۷۶
تعداد پیش بینی غلط	۶۰	۳۸
دقت کل پیش بینی (٪)	۹۴/۷۶	۹۲/۶۱

منبع: یافته های پژوهشگر

بحث و نتیجه‌گیری

گزارش‌های مالی اطلاعات مفیدی در بازارهای مالی در مورد وضعیت فعلی و چشم انداز آتی شرکت‌ها برای ذینفعان ارائه می‌دهد. صورت‌های مالی متقلبانه، جعل عمدی صورت‌های مالی با حذف ارزش عناصر معین در این صورت‌ها به منظور گمراه کردن استفاده‌کنندگان است. اگر پاداش هیئت مدیره با درآمد و سود گزارش شده مرتبط باشد محرک‌های انگیزشی کافی در دستیابی به اهداف سود و منافع شخصی مانند ارتقاء، افزایش حقوق و دستمزد برای گزارش‌های مالی متقلبانه ایجاد می‌شود. گاهی اوقات شرکت‌ها صورت‌های مالی دستکاری شده را برای دسترسی به تأمین اعتبار بدهی بلند مدت یا افزایش قیمت سهم ارائه می‌دهند. هر مورد تقلب در صورت‌های مالی شرکت‌ها، به ویژه شرکت‌های بورسی، نه تنها ذینفعان، مانند سهامداران، سرمایه‌گذاران، سرمایه-گذاران بالقوه، اعتباردهندگان، مشتریان را متضرر می‌کند، بلکه به بقا و عملکردهای پایدار و همچنین سلامت بازارهای مالی و توسعه پایدار شرکت‌ها آسیب می‌رساند. حتی اگر حسابداران و حسابرسان صورت‌های مالی شرکت را با رعایت قوانین و مقررات مربوطه در شناسایی و کشف رفتارهای تقلب آمیز مدیران مورد بررسی قرار دهند، همواره ایراداتی در کشف تقلب وجود خواهد داشت. ایجاد یک مدل دقیق و مؤثر برای کشف تقلب در صورت‌های مالی شرکت‌ها می‌تواند تا حد زیادی تقلب در صورت‌های مالی شرکت‌ها و ریسک حسابرسی را به میزان زیادی کاهش دهد. این امر از یک‌سو مانعی اثرگذار در مدیریت سود شرکت‌ها شده و از سوی دیگر به بنگاه‌ها و بازارهای مالی کمک می‌کند تا توسعه پایدار را حفظ کنند.

روش‌های آماری و داده‌کاوی می‌تواند تا حد زیادی یک سیستم پشتیبانی برای تصمیم‌گیری مدیران در کشف گزارشگری مالی متقلبانه ارائه دهد. از این رو، در این مطالعه انواع مختلف مدل-های آماری و یادگیری ماشین برای کشف تقلب توسعه داده شد. از میان این مدل‌ها، مدلی با بالاترین کارایی و ضریب تاثیر انتخاب شد. برای این منظور از ۲۰ متغیر در قالب پنج ضلعی تقلب با تاکید ساختار کنترل‌های داخلی (فشار، فرصت، توجیه، قابلیت، تکبر و ساختار کنترل‌های داخلی) در ۱۶۶شرکت طی سالهای ۱۳۸۸ الی ۱۳۹۷ در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران استفاده گردید. در مرحله اول مدل سازی، ۲۷ مدل اعم از مدل‌های آماری و یادگیری برای بررسی دقت پیش‌بینی تقلب در صورت‌های مالی مورد مطالعه قرار گرفت که نتایج نشان-دهنده دقت بالای الگوریتم‌های درخت تصمیم‌گیری C5، مدل درخت CHAID و مدل درخت C&R از زیرمجموعه مدل‌های یادگیری ماشین و همچنین مدل‌های پروبیت و گامبیت از زیر مجموعه مدل‌های آماری بود (جدول ۴). در ادامه مقایسه کلی و تحلیل صحت پیش‌بینی گزارشگری مالی متقلبانه توسط مدل‌های یادگیری ماشین و همچنین مدل‌های آماری صورت

گرفت. نتایج بدست آمده از طریق آزمون مقایسه میانگین‌ها برای مدل‌ها نشان داد که دقت پیش بینی مدل های یادگیری ماشین به صورت معنی‌داری از مدل‌های آماری بیشتر است. نتایج بدست آمده از این تحقیق با نتایج تحقیقات (آشتاب و همکاران، ۱۳۹۶)، (صالحی و فرخی، ۱۳۹۷) و (امیدی و همکاران، ۲۰۱۹) مبنی بر این‌که مدل‌های یادگیری ماشین کارایی بالاتری نسبت به مدل‌های آماری دارند، مطابقت دارد. همچنین دقت بالای پیش‌بینی (۹۲/۶۱ ٪) حاصل از ترکیب الگوریتم‌های مدل درخت تصمیم‌گیری CHAID، C5 و C&R نشان می‌دهد روش‌های داده‌کاوی بر پایه مدل‌های یادگیری ماشین و بویژه ترکیب آنها بطور موفقیت‌آمیزی می‌تواند در پیش‌بینی و کشف تقلب در در صورت‌های مالی مورد استفاده قرار گیرد. این مطالعه همچنین پیشنهاداتی برای پیشگیری از تقلب در صورت‌های مالی ارائه می‌دهد. اگرچه تکنیک ها و استراتژی های مختلف مورد استفاده در حسابداری برای تشخیص تقلب می توانند تأثیر مثبتی در پیشگیری و کنترل این پدیده داشته باشند، با این وجود، هوشیاری و احتیاط از طریق اطلاع‌رسانی حتما ضرورت دارد تا در کنار استفاده از برنامه های مختلف ضد تقلب، برای تحریک مسئولیت پذیری مدیران و کارمندان در مبارزه با این پدیده مخرب بکار گرفته شود. از سوی دیگر باید الزامات قانونی ناظران در مورد ایجاد سیستم‌های کنترل داخلی و حسابرسی قوی توسط شرکت‌ها و کمک به حسابداران و حسابرسان در اجرای آن اعمال شود. همچنین مطابق قوانین و مقررات باید شفافیت اطلاعات مالی توسط شرکت-های پذیرفته شده وجود داشته باشد تا مردم بتوانند به عنوان مکانیسم نظارت عمل کنند. شایان ذکر است مهمترین محدودیت در انجام این پژوهش مانند اغلب تحقیقات مربوط به حوزه تقلب در صورت‌های مالی عدم وجود آرشیو غنی از اطلاعات معتبر و قابل اتکا در شناسایی و گزارش شرکت‌های متقلب در بازار بورس و اوراق بهادار تهران بوده است.

فهرست منابع

۱) آشتاب علی, حقیقت حمید و کردستانی غلامرضا. (۱۳۹۶) "مقایسه ی دقت مدل های پیش بینی بحران مالی و تأثیر آن بر ابزارهای مدیریت سود." بررسی‌های حسابداری و حسابرسی، دوره ۱۴ شماره ۲ صفحه ۱۷۲-۱۴۷

۲) اعتمادی حسین و زلقی حسن (۱۳۹۲). کاربرد رگرسیون لجستیک در شناسایی گزارشگری مالی متقلبانه. دانش حسابرسی، دوره چهارده، شماره ۵۱ ; ۲۳ - ۵

۳) تاراسی, بنی طالبی دهکردی و زمانی. (۱۳۹۸). پیش بینی گزارشگری مالی متقلبانه از طریق شبکه عصبی مصنوعی (ANN). حسابداری مدیریت، دوره ۱۲ شماره ۴۰ صفحه ۶۳-۷۹

۴) خواجوی شکراله، ابراهیمی مهرداد. (۱۳۹۶). مدل‌سازی متغیرهای اثرگذار برکشف تقلب در صورت‌های مالی با استفاده از تکنیک‌های داده‌کاوی . فصلنامه حسابداری مالی. ۱۳۹۶; ۹ (۳۳) :۲۳-۵۰

۵) قاضی زاده احسایی, نفیسه؛ محمدرضا عباس زاده؛ مهدی صالحی و مهدی جباری نوقابی، (۱۳۹۸) مقایسه دقت فناوری های نوین مدل های آماری و مدل های یادگیری ماشین به منظور پیش بینی ورشکستگی بانک ها، هفدهمین همایش ملی حسابداری ایران، قم، پردیس فارابی دانشگاه تهران.

۶) صالحی مهدی و فرخی پیله رود لاله. (۱۳۹۷). پیش بینی مدیریت سود با استفاده از شبکه عصبی و درخت تصمیم. پژوهشهای حسابداری مالی وحسابرسی, دوره ۱۰، شماره ۳۷ صفحه ۲۴-۱

۷) کاردان، قره خانی، صالحی و منصوری. (۱۳۹۶). بررسی دقت الگوریتم های خطی-تکاملی BBO و ICDE و الگوریتم های غیرخطی SVR و CART در پیش بینی مدیریت سود. *پژوهش های حسابداری مالی*، دوره ۳۱ شماره ۱ صفحه ۹۵-۷۷.

۸) کرمی غلامرضا، داداشی ایمان، فیروزنیا امیر، کلهرنیا حمید. (۱۳۹۸). بررسی تأثیر کیفیت کنترل‌های داخلی بر حداقل سازی مالیات در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران در طی سالهای ۱۳۹۴-۱۳۸۸ .دانش حسابرسی. دوره ۱۸ شماره ۷۲ صفحه ۲۳-۵۵.

1) Abtahi, Amir-Reza, Fatemeh Elahi, and Reza Yousefi-Zenouz. "An intelligent system for fraud detection in coin futures market's transactions of Iran mercantile exchange based on Bayesian network." *Journal of Information Technology Management* 9, no. 1 (2017): 1-20

2) Amaliah, B. N., Januarsi, Y., & Ibrani, E. Y. (2015). Perspektif Fraud Diamond Theory dalam Menjelaskan Earnings Management Non-GAAP pada Perusahaan Terpublikasi di Indonesia. *Jurnal Akuntansi dan Auditing Indonesia*, 19(1),51-67. Annisya, M., Lindrianasari., & Asmaranti, Y. (2016). Pendeteksian Kecurangan

Laporan Keuangan Menggunakan Fraudو Diamond. *Jurnal Bisnis dan Ekonomi*, 23(1), 72-89.

3) Apparao, G., Arun Singh, G. S. Rao, B. Lalitha Bhavani, K. Eswar, and D. Rajani. "Financial statement fraud detection by data mining." *Corporate governance* 3, no. 1 (2009): 159-163.
4) Chui, Lawrence, and Byron Pike. "Auditors' responsibility for fraud detection: New wine in old bottles?." *Journal of Forensic and Investigative Accounting* (2013).
5) Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." *ACM computing surveys (CSUR)* 41, no. 3 (2009): 1-58.
6) Chen, Hsinchun, and Mihail C. Roco. "Mapping Nanotechnology Knowledge Via Literature Database: A Longitudinal Study, 1976-2004." In *Mapping Nanotechnology Innovations and Knowledge*, pp. 1-22. Springer, Boston, MA, 2009.
7) Cecchini, Mark, Haldun Aytug, Gary J. Koehler, and Praveen Pathak. "Detecting management fraud in public companies." *Management Science* 56, no. 7 (2010): 1146-1160.
8) Dechow, P. M., & Skinner, D. J. (2000). Earning Mangement: Reconcillling the review of accounting Academics, Practitioners,and Reguators. *Accounting Horizontal*, 14(2), 235-250
9) Esfahanipour, Akbar, Milad Goodarzi, and Reza Jahanbin. "Analysis and forecasting of IPO underpricing." *Neural Computing and Applications* 27, no. 3 (2016): 651-658.
10) Hogan, Chris E., Zabihollah Rezaee, Richard A. Riley Jr, and Uma K. Velury. "Financial statement fraud: Insights from the academic literature." *Auditing: A Journal of Practice & Theory* 27, no. 2 (2008): 231-252.
11) Indarto, S. L., & Ghozali, I. (2016). Fraud Diamond: Detection Analysis on The Fraudulent *Financial Reporting. Risk Governance & Control: Financial Markets & Institution*, 6(4), 116-123.
12) Khajavi S, Ebrahimi M. Modelling The Effective Variables for of Financial Statements Fraud Detection using Data Mining Techniques . quarterly financial accounting journal. 2017; 9 (33) :23-50
13) Lin, Chi-Chen, An-An Chiu, Shaio Yan Huang, and David C. Yen. "Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments." *Knowledge-Based Systems* 89 (2015): 459-470.
14) Min, Jae H., and Young-Chan Lee. "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters." *Expert systems with applications* 28, no. 4 (2005): 603-614.
15) Omidi, Mahdi, Qingfei Min, Vahab Moradinaftchali, and Muhammad Piri. "The Efficacy of Predictive Methods in Financial Statement Fraud." *Discrete Dynamics in Nature and Society* 2019.
16) Sabau, Andrei Sorin. "Survey of clustering based financial fraud detection research." *Informatica Economica* 16, no. 1 (2012): 110.

17) Sharma, Anuj, and Prabin Kumar Panigrahi. "A review of financial accounting fraud detection based on data mining techniques." *arXiv preprint arXiv:1309.3944* (2013).
18) Jan, Chyan-long. "An effective financial statements fraud detection model for the sustainable development of financial markets: Evidence from Taiwan." *Sustainability* 10, no. 2 (2018): 513.
19) Zhou, Wei, and Gaurav Kapoor. "Detecting evolutionary financial statement fraud." *Decision support systems* 50, no. 3 (2011): 570-575.
20) Zhang, Dongsong, and Lina Zhou. "Discovering golden nuggets: data mining in financial application." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34, no. 4 (2004): 513-522.
21) Akbar, T. (2017). The determination of fraudulent financial reporting causes by using pentagon theory on manufacturing companies in indonesia. *International Journal of Business, Economics and Law, 14*(5), 106-133.
22) Lowensohn, S., and Reck, J., (2004). Longitudinal analysis of local government audit
quality. *Research in Governmental and NonProfit Accounting,* 11, 213–228.
23) Marsellisa ,N (2018). Financial Statement Fraud: Perspective of the Pentagon Fraud Model in Indonesia. *Academy of Accounting and Financial Studies Journal*.
24) O'Keefe, T.B., King, R.D., and Gaver, K.M., (1994). Audit fees, industry specialization,and compliance with GAAS reporting standards. *Auditing: A Journal of Practice and Theory,* 13(2),41–55.
25) Yusof, M.K., Ahmad, K.A.H & .Simon, J. (2015). Fraudulent Financial Reporting: An Application of Fraud Models to Malaysian Public Listed Companies. *The Macrotheme Review, 4*(3), 126–145.
26) Tiffani, L., & Marfuah. (2015). Deteksi Financial Statement Fraud dengan Analisis Fraud Triangle pada Perusahaan Manufaktur yang Terdaftar di Bursa Efek Indonesia. *Jurnal Akuntansi dan Auditing Indonesia,* 19(2), 112-125.
27) Tessa, C. G., & Harto, P. (2016). Fraudulent Financial Reporting: Pengujian Teori Fraud Pentagon Pada Sektor Keuangan dan Perbankan di Indonesia. Paper presented at Simposium Nasional Akuntansi XIX, held at Universitas Lampung,Lampung, 24-27 Agustus (1-21).

یادداشت‌ها

1. Sharma & Panigrahi
2. Association of Certified Fraud Examiners (ACFE)
3. Machine Learning
4 . Data mining
5 . Money laundering
6. E-commerce
7. Apparao
8 . Chui
9 . Pike
10 . Lin
11 . clustering
12 . association
13 . Zhang & Zhou
14. Supervised
15.Unsupervised
16. Min & Lee
17. Chandola
18. Sabau
19. Zhou & Kapoor
20. Self-adaptive framework
21. Response surface model
22. Sherma & Panigrahi
23. Cecchini
24. Chen & Roco
25. Hogan
26. Jan
27. Omidi
28. Probabilistic neural network
29. Support vector machine
30. Multinomial log-linear model
31. Discriminant analysis
32. Basti
33. Boosting
34. Han
35. Dechow and et al
36. Tiffani & Marfuah
37. Tessa & Harto
38. chung
39. O'Keefe
40. Lowensohn
41. Reck
42. Amaliah et al
43. Indarto & Ghozali