فصلنامه علمی پژوهشی
دانش حسابداری و حسابرسی مدیریت
سال دهم / شماره ٤٠/ زمستان ١٤٠٠

پیش‌بینی تقلب صورت‌های مالی با استفاده از رویکرد کریسپ(CRISP)

مهدی رضائی
دانشجوی‌دکتری‌حسابداری، واحد یزد ،دانشگاه آزاد اسلامی، ، یزد، ایران.
Meh.rezaie@gmail.com

مهدی ناظمی اردکانی
استادیار گروه حسابداری، واحد یزد، دانشگاه آزاداسلامی، یزد، ایران. (نویسنده مسئول)
Nazemi@yazd.ac.ir

علیرضا ناصر صدرآبادی
استادیار گروه مدیریت صنعتی، دانشگاه یزد، یزد، ایران.
alireza_naser@yazd.ac.ir

تاریخ دریافت: ۹۸/۱۱/۲۵ تاریخ پذیرش: ۹۹/۰۲/۲۲

چکیده

هدف اصلی این مقاله پیش بینی تقلب صورت‌های مالی با استفاده از رویکردکریسپ 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 می-باشد، که به روش نمونه گیری حذف سیستماتیک انتخاب 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 ۴۰ 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 جمع‌آوری شده‌اند، براساس رویکرد کریسپ، جهت تعیین وزن و ویژگی متغیرهای بااهمیت به مدل آنتروپی شانون و به منظور پیش‌بینی تقلب به چهار تکنیک برتر از بین تکنیک‌های هوش‌مصنوعی داده شده است، که این تکنیک‌ها شامل؛ درخت تصمیم، شبکه های عصبی، ماشین بردار پشتیبان و روش ترکیبی آدابوست ماشین بردار پشتیبان می باشد. با استفاده از آنتروپی شانون از بین ۴۰ متغیر پژوهش، ۲۷ متغیر برتر براساس ویژگی سود اطلاعاتی، مشخص گردید، که متغیر نسبت سود انباشته به فروش به عنوان با اهمیت ترین متغیر در زمینه پیش بینی تقلب صورت های مالی شناسایی شده است. پس از بکارگیری رویکرد کریسپ، نتایج نشان داد تمامی تکنیک ها قابلیت کشف تقلب صورت‌های مالی را در سطح نسبتا بالایی دارند و تکنیک پیشنهادی آدابوست ماشین بردار پشتیبان در مرحله آموزش با نرخ دقت ۸۱٫۶۹٪ دارای دقت و توان ارزیابی بالاتری نسبت به سایر تکنیک ها بوده و این تکنیک در مرحله آزمایش ۸۲٪ صورت‌های مالی متقلبانه و غیرمتقلبانه سال ۱۳۹۶ را بدرستی تشخیص داد.

واژه‌های کلیدی: تقلب صورت‌های مالی، رویکرد کریسپ، هوش مصنوعی. آنتروپی شانون.

۱- مقدمه

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 می‌باشد (جن چن و همکاران[۱]، 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 ها و رویکردهای کارآمدتری از جمله تکنیک های هوش مصنوعی(AI[۲]) و رویکردهای کارآمدی چون سود اطلاعاتی[۳]، (آنتروپی شانون) 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 و سود اطلاعاتی بعنوان یکی از تکنیک های نوظهور در این زمینه به بررسی کشف تقلب صورت‌های مالی، پرداخته است.

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 همکاران[۴]، ۲۰۱۸). فصل مشترک تعریف‌های موجود از تقلب نشان می‌دهد که همه تقلب‌ها به نیت گول‌زدن دیگران و به منظور رسیدن به نفع شخصی است و به این دلیل با اشتباه تفاوت دارد (رحمانی و همکاران، ۱۳۹۶) . طبق 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 بند۴ 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 صورت‌های مالی به منظور ارائه تصویر نادرست از واحد تجاری(وکیلی فرد و همکاران، ۱۳۸۸)؛ انگیزه ها و دلایل زیادی برای تقلب گزارشگری مالی وجود دارد. نمونه‌هایی 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 بینی‌های داخلی و خارجی (فاما و جنسن[۷] ،۱۹۸۳)،حداقل کردن بدهی مالیاتی (دیچو و همکاران، ۱۹۹۶)، جلوگیری از تخلف در قراردادهای بدهی (اسپاتیس[۸]، ۲۰۰۲)، تأمین مالی به ارزانترین شکلِ ممکن (کوتسیانتیس[۹] ، ۲۰۰۶) و همچنین ارتقاط سطح سرمایه فکری در شرکت ها(ممشلی و کارشناسان، ۱۳۹۸).

هوش مصنوعی(AI)

هوش مصنوعی موضوعی چند رشته ای است که علوم مختلفی نظیر کامپیوتر، روانشناسی، ریاضیات، زیست شناسی و ... در تدوین این رشته نقش دارند. در این تکنیک فرآیندی نظیر آنچه توسط هوش طبیعی و استدلال‌های مغز آدمی در حل مسائل اتفاق می‌افتد، عملی می شود و در حال حاضر توجه به تکنیک‌های مختلف هوش مصنوعی و ابزارهای مدل-سازی، تصمیم‌گیری، انتخاب مدیریت و ... در حوزه حسابداری و حسابرسی رو به گسترش است. با استفاده از هوش مصنوعی فرآیند تصمیم گیری در شرایط مختلف مالی و سرمایه‌گذاری تسهیل و غیر حرفه‌ای‌ها هم می‌توانند تصمیم‌گیری کنند. مضافا"اینکه شیوه بهینه در تصمیم‌گیری، حتی با منفک شدن اشخاص خبره از غیرخبره در شرکت همچنان در اختیار امور مدیریتی و تصمیم‌گیری باقی خواهد ماند و این مهارت نزد شرکت‌ها و موسسات بهره‌بردار حفظ می شود. هوش مصنوعی در حوزه حسابداری کاربردهای زیادی دارد. به طور عام از این تکنیک در انجام کارهای مختلفی نظیر گزارشگری و تحلیل مالی، حسابداری و حسابرسی، برنامه ریزی مالیات، تهیه برنامه-های ممیزی در ارزیابی سیستم‌های کنترل داخلی، کشف تقلب مالی و... استفاده شده است(عرب‌مازار یزدی و همکاران ، ۱۳۸۵).

در این پژوهش از چهار تکنیک معروف و از قوی ترین تکنیک های هوش مصنوعی، شامل؛ درخت تصمیم (DT[۱۰])، شبکه‌های عصبی مصنوعی (ANNs[۱۱])، ماشین بردار پشتیبان (SVM[۱۲]) و تکنیک ترکیبی آدابوست ماشین بردار پشتیبان استفاده می شود. در اینجا به تشریح چهار تکنیک هوش مصنوعی مورد استفاده در این پژوهش می پردازیم.

- درخت تصمیم یک ساختار درختی مانند فلوچارت دارد. در این تکنیک، پیش‌بینی به دست 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 می‌توان 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 مزایای این روش مستقل‌بودن آن از چگونگی توزیع داده‌ها و وابستگی متغیرهای ورودی می‌باشد. درخت‌های تصمیم به علت دقت مناسب و همچنین ساده‌بودن و فهم راحت آن، یکی از محبوب‌ترین روش‌های طبقه‌بندی هستند(حاجی حیدری و خاکباز، ۱۳۹۲). مدل‌های حاصل از درخت تصمیم دارای الگوریتم متفاوتی مثلCHAID، CART ،C4.5 می باشد(کیقبادی و همکاران، ۱۳۹۴).در این پژوهش از الگریتم C4.5 استفاده شده است.

- شبکه های عصبی مصنوعی چکیده ای از مغز انسان است؛ که 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 پرسپترون یکی از مهمترین و رایج‌ترین شبکه‌های عصبی در کاربردهای دنیای واقعی هستند(چی و هانگ[۱۳]، ۲۰۰۸).

در این پژوهش نیز از شبکه عصبی چندلایه پرسپترون 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 دقیق‌تر از رگرسیون خطی توصیف نماید. به علاوه، این روش نیازی به برقراری فروض کلاسیک رگرسیون ندارد (حجازی و همکاران ، ۱۳۹۱).

- 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[۱۴]، ۱۹۹۵). 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 نمونه‌ها دارد قابلیت تعمیم بالایی دارد.

- آدابوست ماشین بردار پشتیبان از ترکیب ماشین‌های بردار پشتیبان مختلف حاصل می شود و الگویتم جهش یافته ماشین بردار پشتیبان پایه می باشد، که دقت طبقه بندی آن در بین سایر تکنیک های ماشین بردار پشتیبان نسبتا بالاتر می باشد(دمینگس[۱۵] و همکاران، ۲۰۱۹).

روش شاخص وزن دهی (سود اطلاعاتی)

این روش یکی از ساده‌ترین روش‌های انتخاب ویژگی است. عملگر «وزندهی با شاخص سود اطلاعاتی»، رابطه بین متغیرها را براساس آنتروپی یا سود اطلاعاتی روی مجموعه داده‌ها محاسبه می‌کند. این عملگر، وزن هر متغیر را با محاسبه به خروجی نرم‌افزار می‌دهد. در واقع روشی برای محاسبه ناخالصی متغیر است. متغیر نامتوازن متغیری است که کمترین میزان ناخالصی را داشته باشد. و گره متوازن بیشترین ناخالصی را دارد و بهتر می تواند تفکیک را بین دو کلاس دربخش دسته بندی داشته باشد، به عبارت دیگر بیشترین سود اطلاعاتی را دارد (ریجرام[۱۶] و همکاران، ۲۰۱۷).

پیشینه پژوهش

سادگالی و همکاران[۱۷](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 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[۱۸](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 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[۱۹] (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 گسترده‌ای برای پیش‌بینی گزارش‌های مالی متقلبانه استفاده شده است، عملکرد مناسب‌تری دارد.

چی چن و همکاران[۲۰] (۲۰۱۵) به تجزیه و تحلیل تفاوت های میان روش‌های داده کاوی و خبره‌سنجی در کشف تقلب صورت‌های مالی پرداختند. روش‌های تجزیه و تحلیل داده‌ها به کاربرده شده در این پژوهش شامل رگرسیون منطقی ، CART، شبکه های 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 شبکه-های 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، می توان شبکه عصبی طراحی نمود، که با دقت قابل قبولی بتوانند، احتمال گزارشگری مالی متقلبانه شرکت را کشف و پیش‌بینی نمایند.

عظیمی و نورحسینی (۱۳۹۶) به کشف و پیش بینی تقلب مالی در بانک ها با استفاده از روش های داده‌کاوی، پرداختند. نتایج خروجی نشان داد که داده‌کاوی در حدود ۸۳٪ پیش بینی درستی در خروجی نمایش می دهد.

حمزه زادی و اعتمادی (۱۳۹۵) به کاربرد شبکه عصبی مصنوعی در شناسایی تقلب در صورت‌های مالی پرداختند. نمونه آماری پژوهش شامل ۸۵۵ شرکت (۳۷۹ شرکت دارای نشانه های تقلب و ۴۷۶ شرکت فاقد نشانه های تقلب) پذیرفته شده در بورس اوراق بهادار تهران و در دوره زمانی ۱۳۸۸-۱۳۸۴ است. از میان ۲۹ نسبت مالی اولیه، ۱۵ نسبت به عنوان پیش‌بینی کننده‌های بالقوه برای آزمون انتخاب شده اند. چند مدل شناسایی صورت های مالی دارای تقلب بر اساس شبکه عصبی پرسپترون چند لایه با تعداد نرون و توابع فعال‌سازی مختلف ایجاد شدند. نتایج حاکی توانایی مدل‌های مبتنی بر شبکه عصبی در پیش بینی تقلب در صورت‌های مالی است.

مرادی و همکاران (۱۳۹۳) به شناسایی عوامل خطر مؤثر بر احتمال وقوع تقلب در گزارشگری مالی از دید حسابرسان و بررسی تأثیر آن‌ها بر عملکرد مالی شرکت پرداختند. هدف این پژوهش، شناسایی عوامل مؤثر بر احتمال وقوع تقلب در گزارشگری صورت‌های مالی از دید حسابرسان و مدیران است. جامعه‌ی آماری این پژوهش، حسابرسان مستقل‌مؤسسه‌های حسابرسی عضو جامعه حسابداران رسمی و مدیران مالی شرکت‌های بورسی هستند. به منظور آزمون فرضیه‌های تحقیق از آزمون تحلیل واریانس یک طرفه استفاده شده و ارزیابی عملکرد شرکت‌ها نیز برای سال‌های ۱۳۸۷-۱۳۸۹ با استفاده از مدل‌های رگرسیونی و به کمک تحلیل همبستگی صورت گرفته است. یافته های پژوهش حاکی از آن است که بین ویژگی‌های مدیریت، تبعیت مدیریت از کنترل‌های داخلی و استانداردهای لازم الاجرا، عوامل خطر مرتبط با شرایط بازار و صنعت، ویژگی های عملیاتی، نقدینگی و ثبات مالی با احتمال وقوع تقلب رابطه‌ی معناداری وجود دارد. همچنین نتایج حاکی از وجود رابطه‌ی معنادار بین عملکرد شرکت (متغیرهای نرخ بازده دارایی‌ها، جریان‌های نقدی عملیاتی، بازده سهام و بازده شرکت) با ریسک تقلب است.

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.

روش شناسی پژوهش

این پژوهش از لحاظ هدف کاربردی و از نوع پژوهش‌های تجربی است و با توجه به اینکه از در آن از اطلاعات تاریخی استفاده شده است؛ در دسته پژوهش‌های شبه آزمایشی قرار می گیرد. برای جمع‌آوری اطلاعات مربوط به مبانی نظری و پیشینه پژوهش و همچنین داده‌های 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 افزار Rapidminer 7.3 برای سنجش توانایی تکنیک‌های هوش مصنوعی در کشف تقلب صورت‌های مالی استفاده شده است.

جامعه آماری پژوهش شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران طی سال‌های ۱۳۹۳تا ۱۳۹۶ می‌باشد. در این پژوهش با توجه به بازه‌ی زمانی سال‌های ۱۳۹۳ تا ۱۳۹۶ و شرایط انتخاب متغیرها، از روش نمونه گیری حذف سیستماتیک برای رسیدن به نمونه استفاده گردیده که محدودیت‌های اعمال شده برای انتخاب نمونه عبارتند از : ۱- جزء شرکت‌های سرمایه‌گذاری ، بانک‌ها و واسطه‌گری مالی، هلدینگ‌ها، لیزینگ‌ها و بیمه ها نباشند.۲- پایان سال مالی آن‌ها پایان اسفند باشد. ۳- نماد معاملاتی شرکت فعال و سهام آن‌ها حداقل یکبار در سال معامله شده باشد. ۴- اطلاعات مالی

آن‌ها در بازه‌ی زمانی پژوهش، در دسترس باشد. با توجه به محدودیت‌های تعیین شده، ۱۶۴ شرکت بعنوان نمونه استخراج گردیده است. بعبارت دیگر ۶۵۶ سال شرکت مورد استفاده قرار گرفته است.

اندازه گیری تقلب صورت‌های مالی

متغیر وابسته در این تحقیق، تقلب در صورت‌های مالی است که از ماهیت کیفی برخوردار بوده ودارای مقیاس سنجش اسمی است. برای اندازه گیری تقلب صورت‌های مالی بعنوان متغیر وابسته پژوهش حاضر در پژوهش‌های مختلف معیارهای متفاوتی ملاک عمل قرار گرفته است. در پژوهش حاضر برای اندازگیری تقلب صورتهای مالی و تفکیک شرکتهای متقلب از غیر متقلب از شیوه اندازه گیری تقلب توسط فرقاندوست حقیقی و همکاران (۱۳۹۳) استفاده شده است. در این شیوه فرض گردیده که شرکت‌هایی که گزارش حسابرسی آن‌ها مردود، عدم اظهار نظر یا مشروط بوده‌اند با احتمال بیشتری دست به تقلب زده‌اند نسبت به شرکت‌های که گزارش مقبول دارند. بنابراین در ابتدا شرکت هایی که گزارش حسابرسی مردود، عدم اظهار نظر یا مشروط بوده شناسایی می شوند. سپس از بین موارد شناسایی شده، شرکت هایی که یکی از مصادیق تقلب ذیل به عنوان علت نوع اظهار نظرشان ذکر گردیدند، با احراز یکی از مصادیق، به عنوان شرکت های متقلب شناسایی می شوند، که مصادیق بر مبنای پژوهش های ماهر و همکاران ۲۱ (۲۰۰۷)، فرقاندوست حقیقی و همکاران (۱۳۹۳) و وکیلی فرد و همکاران (۱۳۸۸) انتخاب گردیدند؛ که عباتند از :

۱) شناسایی نادرست درآمد ها و اندازه گیری درآمدهای تحقق یافته

۲) بیش نمایی دارایی ها و موجودی های پایان دوره

۳) شناسایی نادرست هزینه ها و عدم اندازه گیری هزینه های تحقق یافته.

۴)کم نمایی در بدهی ها و استفاده نادرست و متقلبانه از حساب های ذخیره

۵) عدم تهیه صورت های مالی با فرض توقف فعالیت های شرکت هایی که فرض تداوم فعالیت آن ها به شکل اساسی زیر سوال رفته و در گزارش های حسابرسی بازیافت یافته است.

متغیر های مستقل

متغیر های مستقل این پژوهش که به عنوان متغیرهای پیش بین و تاثیرگذار بر تقلب صورت‌های مالی می باشند، شامل؛ ۴۰ 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 بشرح جدول ۱ می باشد.

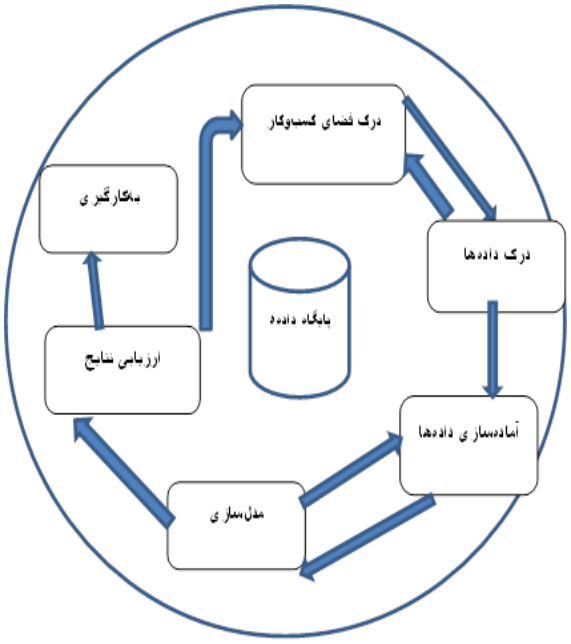
جدول ۱: متغیر های مستقل(پیش بین) پژوهش

متغیر مستقل	نحوه اندازه گیری
متغیر های غیرمالی:	
نوع عملکرد	در صورت زیان ده بودن ۱ ، در غیر این صورت ۰
دعاوی حقوقی	در صورت وجودی دعاوی حقوقی علیه شرکت ۱ ، در غیر این صورت ۰
نوع موسسه حسابرسی	سازمان حسابرسی ۱ ، در غیر این صورت ۰
متغیرهای مالی :	
سود پس از مالیات به فروش	سود خالص / فروش
سود قبل از مالیات و پس از بهره به فروش	سود عملیاتی – هزینه بهره / فروش
نسبت جاری	دارایی های جاری / بدهی های جاری
نسبت کل بدهی به دارایی ها	کل بدهی ها / کل دارایی ها
لگاریتم طبیعی کل دارایی ها (اندازه شرکت)	Ln کل دارایی ها
نسبت حساب های دریافتنی به کل دارایی ها	حسابهای دریافتی / کل دارایی ها
دفعات گردش موجودی کالا	فروش خالص/ متوسط موجودی کالا
دوره گردش موجودی کالا	۳۶۵/ دفعات گردش موجودی کالا
نسبت آنی	دارایی های جاری – موجودی کالا – پیش پرداخت / بدهی های جاری
شاخص ورشکستگی	$ZScore=3.2078X_1 + 1.803X_2 + 1.613X3 + 0.5X4 -0.169X5 – 0.397X6 -0.125X7+338X8 + 1.423X9$
لگاریتم طبیعی فروش خالص	Ln فروش خالص

متغیر مستقل	نحوه اندازه گیری
لگاریتم طبیعی تعداد کارکنان	Ln تعداد کارکنان
نسبت پرداخت بدهی ها	مجموع دارایی ها/ مجموع بدهی ها
سرمایه در گردش	دارایی های جاری - بدهی های جاری
سرانه دارایی ها	کل دارایی ها / تعداد کارکنان
سرانه فروش	فروش خالص/ تعداد کارکنان
سرانه سود خالص	سود خالص/ تعداد کارکنان
گردش حساب های دریافتنی	متوسط حسابهای دریافتنی / فروش خالص
دوره وصول مطالبات	گردش حساب های دریافتنی / ۳۶۵
گردش خالص دارایی ها	فروش خالص / متوسط مجموع دارایی ها
گردش خالص دارایی های ثابت	فروش خالص / متوسط مجموع دارایی های ثابت
سود قبل از بهره و مالیات به فروش	سود قبل از بهره و مالیات / فروش
وجه نقد ناشی از فعالیت های عملیاتی به فروش	وجه نقد ناشی از فعالیت های عملیاتی/ فروش
بازده حقوق صاحبان سهام	سود خالص/ حقوق صاحبان سهام
نسبت ارزش ویژه به بدهی بلندمدت	ارزش دفتری حقوق ص. س /حصه جاری تسهیلات دریافتی + بدهی های بلندمدت
بازده کل دارایی ها	میانگین مجموع دارایی ها / فروش خالص
بازده سرمایه بکار گرفته شده	کل دارایی ها – بدهی های جاری / حاشیه سود قبل از بهره و مالیات
ذخیره مالیات به فروش	ذخیره مالیات / فروش خالص
سرانه ذخیره مزایای پایان خدمت کارکنان	ذخیره مزایای پایان خدمت کارکنان/ فروش خالص
سود (زیان) انباشته به فروش	سود (زیان) انباشته / فروش خالص
درصد تغییرات در سودآوری	سود سال جاری – سود سال قبل/ سود سال قبل
سود ناخالص به فروش	سود ناخالص/ فروش خالص
افزایش یا کاهش در وجه نقد	وجه نقد پایان دوره – وجه نقد اول دره
دارایی های جاری	دارایی های جاری
بدهی های جاری	بدهی های جاری
بدهی بلندمدت	بدهی بلندمدت
وجه نقدناشی ازفعالیتهای سرمایه گذاری به فروش	وجه نقد ناشی از فعالیت های سرمایه گذاری / فروش خالص

رویکرد کریسپ(CRISP[۲۲])

یکی از پرکاربردترین روش‌ها در داده کاوی ، رویکرد کریسپ یا فرایند استاندارد متقاطع است(چاپمن[۲۳] و همکاران، ۲۰۰۰). این رویکرد در اواخر سال ۱۹۹۶ توسط سه سازمان تحلیل داده طرح‌ریزی شد. فرآیند داده‌کاوی در استاندارد کریسپ ، به روش کشف دانش که توسط اسامه فیاد و همکارانش (۱۹۹۶) ارائه شد، در طول شش مرحله اجرا می‌شود؛ مدل اجرایی کریسپ یک مدل حلقوی و تکرارشونده است. به این معنی که برخی از مراحل پس از اجرا، ممکن است چندین بار تکرار شوند؛ تا نتیجه موردنظر حاصل شود. در نمودار ۱ مدل فرآیندی کریسپ به‌طور حلقوی ترسیم شده است. توالی حرکت مستقیم نیست. بردارها وابستگی‌های مهم و غالب بین مراحل را مشخص می‌کند. در ادامه به جزییات مراحل اجرای این پژوهش به روش کریسپ پرداخته می شود.



نمودار ۱- مدل رویکرد کریسپ (وایرد و هیپ[۲۴]، ۲۰۰۰)

گام‌های رویکرد کریسپ

درک داده

همانطور که قبلا توضیح داده شد 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 معیارهای اعمال شده برای انتخاب نمونه ارائه شده است. با توجه به فیلترهای تعیین شده، ۱۶۴ شرکت بعنوان نمونه آماری مشخص گردیده است. که داده‌های مربوط به متغیر وابسته(تقلب صورت‌های مالی) فقط برای سال ۱۳۹۶ و داده های مربوط به متغیرهای مستقل(شامل ۴۰ متغیر پیش بین) برای سال‌های ۱۳۹۳ الی ۱۳۹۵ برای ۱۶۴ شرکت نمونه، بصورت فایل اکسل جمع‌آوری گردیده است.

آماده سازی داده ها

- **حذف داده از دست رفته** : در این مرحله با درک داده و نمودار آماری تشخیص داده می‌شود که برخی از ستون ها دارای مقدار از دست رفته می باشند. برای حذف مقدارهای از دست رفته یکی از راه ها حذف سطرهایی است که مقدار از دسته رفته دارند؛ که این راه منطقی نیست، زیرا ممکن است نمونه های مهم که در ستون‌های دیگر مقدارهای با ارزشی دارند حذف شده و الگوی داده از دست رود. یک راه بهتر جانشینی مقدارهای ازدست رفته در ستون مربوط به میانگین ستون است.

- **حذف داده تکراری**: از آنجا که سطر تکراری در داده‌ها نباید وجود داشته باشد، نیازی به اجرای این مرحله در این پژوهش نیست زیرا داده ها با تحلیل نرم‌افزار دارای سطر تکراری نیستند.

- **انتخاب ویژگی**

در این بخش روش انتخاب ویژگی شاخص آنتروپی یا سود اطلاعاتی جهت ارزیابی و انتخاب متغیرهای تأثیرگذار انتخاب شد. فرمول محاسبه سود اطلاعاتی که بر پایه آنتروپی شانون تعریف می‌شود به صورت فرمول۱ تعریف می شود.

$$informationGain(A) = Entropy(D) - Entropy_A^D \quad (۱)$$

که در آن A ویژگی و D دلالت بر مجموعه داده‌های آموزشی دارد. و فرمول آنتروپی شانون به صورت فرمول ۲ تعریف می شود. و در آن c تعداد برچسب کلاس‌های موجود در داده های آموزشی، p_i احتمال اینکه نمونه ای از داده ها متعلق به کلاس i ام باشد، v تعداد اعضای دامنه صفت خاصه A و D_j قسمتی از داده های اولیه است . در ضمن $|D|$ دلالت بر اندازه مجموعه داده دارد.

$$Entopy(D) = -\sum_{i=1}^{c} P_i \times \log(pi) \quad (۲)$$

$$Entopy(D)_A = \sum_{j=1}^{v} \frac{|Dj|}{|D|} \times Entopy(D)$$

مدلسازی

در این پژوهش یکی از اهداف اصلی پیش‌بینی تقلب سال ۱۳۹۶ بعنوان سال مبنا می‌باشد، که در این راستا از ۱۶۴ شرکت نمونه بر اساس شیوه اندازه‌گیری تقلب ذکر شده در بالا ۸۶ شرکت متقلب و ۷۸ شرکت غیرمتقلب تشخیص داده شده و به شکل متغیر دامی به شرکت‌های متقلب عدد یک (۱) و به شرکت‌های غیرمتقلب عدد صفر (۰) اختصاص داده می شود. سپس با جمع اوری داده های مربوط به متغیرهای مستقل برای سال‌های ۱۳۹۳ الی ۱۳۹۵ و تعیین مؤثرترین این متغیرها بر تقلب صورت‌های مالی با استفاده از رویکرد سود اطلاعاتی، (در مرحله آموزش) به کشف تقلب صورت‌های مالی سال‌های ۱۳۹۳ تا۱۳۹۵ و پیش‌بینی تقلب صورت‌های مالی (در مرحله آزمایش) برای سال ۱۳۹۶ با استفاده از تکنیک‌های هوش مصنوعی پرداخته می شود .

همچنین با ساخت مدل طبقه‌بندی و پیش‌بینی تقلب در سال ۱۳۹۶ الگوریتم‌های دسته‌بندی به‌منظور بیان دانش و استخراج قوانین به کار می‌روند. در این تحقیق دسته-بند(تکنیک)های مختلفی مورد استفاده قرار گرفته‌اند؛ که

شامل: درخت تصمیم C4.5، شبکه عصبی پرسپترون (ANNs)، ماشین بردار پشتیبان (SVM) و یک الگوریتم ترکیبی آدابوست ماشین بردار پشتیبان (AdaBoost SVM) می‌باشد؛ لازم به ذکر است؛ که انتخاب این تکنیک‌ها بر اساس مطالعه مبانی نظری و پیشینه صورت گرفته است.

اعتبارسنجی کراس یا ضربدری

با استفاده از روش اعتبارسنجی ضربدری مجموعه داده ها را به دو دسته آموزشی و تست تقسیم شد. در این نوع اعتبارسنجی، داده‌ها به K زیرمجموعه افراز می‌شوند؛ از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش بکار می‌روند. این روال K بار تکرار می‌شود و همه داده‌ها دقیقا یک‌بار برای آموزش و یک‌بار برای اعتبارسنجی بکار می‌روند. در نهایت میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می‌شود.بنابراین در روش اعتبارسنجی ضربدری، روی مجموعه داده آموزشی هر دسته بند آموزش دیده و با داده تست دقت مدل ارزیابی می‌شود. مدلی که خطای پیش بینی کمتری داشته باشد نتایج دقیق تری دارد. که در این پژوهش تعداد K برای داده های آموزشی(Train) برابر ۴۹۲ صورت مالی بوده و برای داده های آزمایشی(Test) تعداد ۱۶۴ صورت مالی بوده است.

ارزیابی

– پارامترهای ارزیابی

قبل از آنکه به بررسی انواع مهم معیارهای دسته‌بندی پرداخته شود، لازم است که مفهوم ماتریس درهم‌ریختگی که از در این بخش بهره خواهیم برد، تشریح شود. این ماتریس چگونگی عملکرد الگوریتم دسته‌بندی را با توجه به مجموعه داده ورودی به تفکیک انواع دسته‌های مساله نشان می‌دهد. با توجه به جدول ۲ ، مفاهیم مثبت درست(TP)[۲۵]، مثبت غلط(TF)[۲۶]، منفی درست(TN)[۲۷] و منفی غلط(FN)[۲۸] به شرح ذیل می باشد.

جدول ۲- ماتریس درهم‌ریختگی

		نمونه های تخمینی	
		دسته منفی (–)	دسته مثبت (+)
نمونه های واقعی	دسته منفی (–)	TN	FP
	دسته مثبت (+)	FN	TP

- ✓ مثبت درست(TP): این مقدار بیانگر تعداد نمونه‌هایی است که دسته واقعی آن‌ها مثبت بوده و دسته‌بند نیز دسته آن‌ها را به درستی مثبت تشخیص داده است.
- ✓ مثبت غلط(FP): این مقدار بیانگر تعداد نمونه‌هایی است که دسته واقعی آن‌ها منفی بوده و دسته‌بند دسته آن‌ها را به نادرستی مثبت تشخیص داده است.
- ✓ منفی درست(TN): این مقدار بیانگر تعداد نمونه‌هایی است که دسته واقعی آن‌ها منفی بوده و دسته‌بند نیز دسته آن‌ها را به‌درستی منفی تشخیص داده است.
- ✓ منفی غلط(FN): این مقدار بیانگر تعداد نمونه‌هایی است که دسته واقعی آنها مثبت بوده و دسته‌بند دسته آن‌ها را به نادرستی منفی تشخیص داده است.

در ادامه به تشریح انواع معیارهای مهم ارزیابی تکنیک‌های دسته بندی پرداخته می شود.

– معیارهای ارزیابی براساس مفاهیم ماتریس درهم ریختگی:

- معیار دقت مدل دسته بندی

$$Accuracy = \frac{Tp+TN}{Tp+TN+Fp+FN} \quad (۱)$$

- معیار حساسیت مدل دسته بندی

$$Sensitivity = \frac{TP}{FN+TP} \quad (۲)$$

- معیار مختصات مدل دسته بندی

$$Specificity = \frac{TN}{TN+FP} \quad (۳)$$

- معیارسطح زیر نمودار تحلیل ویژگی عامل‌ها

$$FAR = \frac{FP}{TN+FP} \quad (۴)$$

نتایج اجرای مدل‌های سود اطلاعاتی و هوش مصـنوعی براساس رویکرد کریسپ

- **انتخاب ویژگی**

در این مرحله برای انتخاب ویژگی و ارزیابی متغیرهای تأثیرگذار بر تقلب صورت‌های مالی، از شاخص آنتروپی شانون یا روش سود اطلاعاتی، استفاده شده است؛ از این رو در این قسمت به تشریح نتایج روش آنتروپی شانون، که خروجی حاصل از اجرای این مرحله در نرم‌افزار رپیدماینر (Rapidminer) می‌باشد، پرداخته شده است. ارزیابی الگوریتم انتخاب ویژگی در مرحله مدل‌سازی براساس مقدار پارامترهای دقت و وزن، بررسی و تحلیل خواهد شد. نتیجه اجرای الگوریتم آنتروپی شانون در جدول ۳ ارائه شده است.

جدول ۳: وزن متغیرها در شاخص سود اطلاعات

رتبه	متغیرهای مستقل(پیش بین)	وزن
۱	سود وزیان انباشته به فروش	۱,۰
۲	شاخص ورشکستگی (zscore)	۰,۹۴۱۰
۳	گردش خالص دارایی	۰,۸۰۹۸
۴	دوره وصول مطالبات	۰,۷۹۲۱
۵	بازده کل دارایی	۰,۷۵۶۱
۶	نسبت پرداخت بدهی	۰,۶۳۹۰
۷	نسبت کل بدهی به کل دارایی	۰,۶۳۹۰
۸	سرانه فروش	۰,۶۲۷۳
۹	سود قبل مالیات و پس از بهره به فروش	۰,۵۶۶۸
۱۰	سود پس از مالیات به فروش	۰,۵۶۶۸
۱۱	سرانه سود خالص	۰,۵۶۶۶
۱۲	گردش حسابهای دریافتنی	۰,۵۲۲
۱۳	بازده حقوق صاحبان سهام	۰,۴۵۶
۱۴	ذخیره مالیات به فروش	۰,۴۵۵
۱۵	سرمایه در گردش	۰,۴۴۴
۱۶	سود ناخالص به فروش	۰,۴۳۷
۱۷	سرانه ذخیره مزایای پایان خدمت کارکنان	۰,۴۲۰
۱۸	لگاریتم طبیعی تعداد کارکنان	۰,۴۱۷
۱۹	سود قبل از بهره و مالیات به فروش	۰,۴۱۶۷
۲۰	نسبت ارزش ویژه به بدهی بلند مدت	۰,۴۱۶۰
۲۱	لگاریتم طبیعی فروش	۰,۴۱۳۲
۲۲	نسبت جاری	۰,۴۰
۲۳	بازده سرمایه بکار گرفته	۰,۳۶
۲۴	نوع عملکرد	۰,۳۳
۲۵	گردش خالص دارایی ثابت	۰,۳۲
۲۶	سرانه دارایی	۰,۳۱
۲۷	اندازه شرکت (لگاریتم کل دارایی)	۰,۲۸
۲۸	دوره گردش موجودی کالا	۰,۲۸
۲۹	دفعات گردش موجودی کالا	۰,۲۸
۳۰	جمع دارایی‌های جاری	۰,۲۷
۳۱	نسبت آنی	۰,۲۷
۳۲	جمع بدهی‌های جاری	۰,۲۶
۳۳	وجه نقد ناشی از فعالیتهای سرمایه گذاری به فروش	۰,۲۵
۳۴	وجه ناشی از جریانهای عملیاتی به فروش	۰,۲۴

رتبه	متغیرهای مستقل(پیش بین)	وزن
۳۵	بدهی بلند مدت	۰,۲۳
۳۶	افزایش یا کاهش وجه نقد	۰,۲۰
۳۷	درصد تغیرات سود آوری	۰,۲۰
۳۸	نسبت حسابهای دریافتنی به کل دارایی	۰,۱۹
۳۹	نوع موسسه حسابرسی	۰,۱۶
۴۰	دعاویی حقوقی	۰

- ✓ لازم بذکر است انتخاب ۲۷ متغیر در انتخاب ویژگی بر اساس آزمون و خطا در نرم افزار Rapidminer است؛ که مؤثرترین و کاراترین تعداد متغیر در پیش‌بینی تقلب بر اساس روش آنتروپی می‌باشد.
- ✓ تاثیرگذارترین متغیر سود وزیان انباشته به فروش است. که با وزن ۱ به طور مستقیم در متقلبانه بودن و غیرمتقلبانه بودن صورت‌های مالی دلالت دارد.
- ✓ سایر متغیرهای تاثیرگذار در متقلبانه بودن یا غیرمتقلبانه بودن صورت‌های مالی، شاخص ورشکستگی، گردش خالص دارایی، دوره وصول مطالبات و بازده کل دارایی‌هاست.
- ✓ دعاویی حقوقی با وزن صفر کم اثرترین ویژگی یا متغیر بر تقلب صورت‌های مالی است.
- ✓ طبق روش آنتروپی شانون نوع موسسه حسابداری، نسبت حسابهای دریافتی به کل دارایی و افزایش و کاهش نقد نیز از متغیرهای کم تأثیر ارزیابی و شناسی شد.

اجرای مدل‌سازی

در این پژوهش برای ساخت مدل‌ها از دسته‌بند(تکنیک)های برتر استفاده شده است. دسته‌بندهای برتر مورد استفاده، همانطور که قبلا ذکر شد عبارتند از: درخت تصمیم، شبکه عصبی، ماشین بردار پشتیبان پایه و آدابوست ماشین بردار پشتیبان. هرکدام از این دسته‌بندها، دارای پارامترهایی جهت ارزیابی هستند. برای مدل‌سازی بر اساس هریک از دسته‌بندها، در وهله اول، بهینه‌ترین پارامترهای مجموعه داده‌ها جهت تنظیم پارامترهای دسته‌بندها، انتخاب شده‌اند و سپس طی چندین مرحله اجرا و طبق جدول۴ دسته‌بندها تنظیم گردیده‌اند. تا در اجرای دسته‌بندها از آن‌ها استفاده شود.

جدول ۴: پارامترهای دسته بندها

پارامترهای تنظیم	نام مدل	مخفف نام
• شاخص جینی به منظور انشعاب • حداکثر عمق:۲۰ • حداقل سایز برگ:۳ • سطح اطمینان برای هرس:۰,۲۵ • حداقل تعداد رکودستها جهت خاتمه:۴	درخت تصمیم	DT
• نوع: شبکه‌های عصبی بازگشتی • سیکل آموزش:۵۰۰ • ضریب مومنتوم (حرکت آنی): ۰,۷ • نرخ یادگیری: ۰,۳ • تعداد لایه‌ها:۲ لایه اول ۵ لایه دوم ۶ • **Decay**: فعال	شبکه عصبی پرسپترون	ANN
• نوع هسته: تابع کرنل خطی • ثابت **C**: ۰,۲ • تعداد تکرارها: ۱۰۰۰۰	ماشین بردار پشتیبان	Svm

هر یک از تکنیک‌های فوق براساس تنظیمات جدول ۴ ابتدا با استفاده از داده‌های سال‌های ۱۳۹۳ تا ۱۳۹۵ آموزش(Train) داده شده است؛ سپس با انتخاب قویترین تکنیک از بین آن‌ها مرحله آزمایش(Test) برای پیش‌بینی تقلب سال ۱۳۹۶ صورت گرفته است.

نتایج درخت تصمیم

در این مرحله نتایج درخت تصمیم را یکبار بدون انتخاب ویژگی و یکبار با انتخاب ویژگی سود اطلاعاتی روی داده متغیرهای صورت‌های مالی در سال‌های ۹۳ تا ۹۵ گزارش می‌دهیم. و نتایج را مقایسه می‌کنیم.

جدول ۵: نتایج درخت تصمیم

Model	Acc	Sensitivity	specificity	AUC
DT	81.08%	80.222%	82.08%	0.791
DT+ Antropy	81.28%	80.22%	82.50%	0.796

نتایج نشان می‌دهد که درخت تصمیم با انتخاب ویژگی 27 متغیر تاثیرگذاراول از مجموع 40 متغیر بهبود دقت داشته است.

نتایج شبکه عصبی

در این مرحله نتایج شبکه عصبی پرسپترون را یکبار بدون انتخاب ویژگی و یکبار با انتخاب ویژگی سود اطلاعاتی روی داده متغیرهای صورت‌های مالی در سال‌های ۹۳ تا ۹۵ گزارش می‌دهیم. و نتایج را مقایسه می‌کنیم.

جدول ۶: نتایج شبکه عصبی

Model	Acc	Sensitivity	specificity	AUC
ANN	80.07%	80.25%	79.95%	0.835
ANN+ Antropy	80.47%	81.39%	79.51%	0.843

شبکه عصبی نیز با انتخاب متغیرهای تاثیرگذار بهبود دقت دارد. در ضمن معیار دقت (ACC)این تکنیک پایین تر از تکنیک درخت تصمیم است.

نتایج ماشین بردار پشتیبان

در این مرحله نتایج svm را یکبار بدون انتخاب ویژگی و یکبار با انتخاب ویژگی سود اطلاعاتی روی داده متغیرهای صورتهای مالی درر سالهای ۹۳ تا ۹۵ گزارش می‌دهیم. و نتایج را مقایسه می‌کنیم.

جدول۷: نتایج ماشین بردار پشتیبان

Model	Acc	Sensitivity	Specificity	AUC
SVM	81.08%	81.60%	80.25%	0.825
SVM+ Antropy	81.30%	82.96%	80.12%	0.826

ماشین بردار نتایج بهتری از شبکه عصبی و درخت دارد و با دقت ۸۱,۳۰٪ دقت بهتری دارد. در این روش نیز با انتخاب ویژگی تا میانگین ۰,۲۲ درصد افزایش دقت داریم. همچنین دقت این دسته بند نسبت به دو دسته بند قبلی (DT , ANN) دقت بالاتری دارد.

نتایج آدابوست ماشین بردار پشتیبان

در این مرحله نتایج روش ترکیبی Adaboost Svm را یک‌دفعه بدون انتخاب ویژگی و یک‌دفعه با انتخاب ویژگی سود اطلاعاتی روی داده متغیرهای صورت‌های مالی در سال‌های ۹۳ تا ۹۵ گزارش می‌دهیم. و نتایج را مقایسه می‌کنیم.

جدول ۸: ماتریس درهم‌ریختگی آدابوست ماشین بردار با انتخاب ویژگی (داده های آموزشی)

		نمونه‌های تخمینی	
		دسته منفی (-)	دسته مثبت (+)
نمونه‌های واقعی	دسته منفی (-)	۱۸۸	۴۴
	دسته مثبت (+)	۴۸	۲۱۴

تعداد صورت‌های مالی که دسته واقعی آن‌ها مثبت (متقلبانه) بوده و دسته‌بند نیز دسته آن‌ها را به درستی مثبت تشخیص داده به تعداد ۲۱۴ صورت مالی بوده است.

تعداد صورت‌های مالی که دسته واقعی آن‌ها منفی (غیر متقلبانه) بوده و دسته‌بند، دسته آن‌ها را به نادرستی مثبت تشخیص داده به ۴۴ صورت مالی بوده است.

تعداد صورت‌های مالی که دسته واقعی آن‌ها منفی بوده و دسته‌بند نیز دسته آن‌ها را به‌درستی منفی تشخیص داده ۱۸۸ صورت مالی بوده است.

تعداد صورت‌های مالی که دسته واقعی آنها مثبت بوده و دسته بند دسته آن‌ها را به نادرستی منفی تشخیص داده ۴۸ صورت مالی است.

جدول۹: نتایج آدابوست ماشین بردار پشتیبان

Model	Acc	Sensitivity	Specificity	AUC
Adaboost SVM	81.28%	81.79%	80.36%	0.830
SVM Adaboost + Antropy	81.69%	82.96%	80.36%	0.826

نتایج روش ترکیبی آدابوست ماشین بردار پشتیبان نسبت به تکنیک پایه ماشین بردار پشتیبان بهبود دقت دارد و با انتخاب ویژگی نیز همانطور که در جدول بالا مشخص است این تکنیک بهبود دقت دارد .

جمع بندی نتایج حاصل از اجرای تکنیک‌ها

نتایج دسته بندهای مختلف براساس دیتای سال‌های ۹۳ تا ۹۵ را در جدول ۱۰ بشرح زیر است:

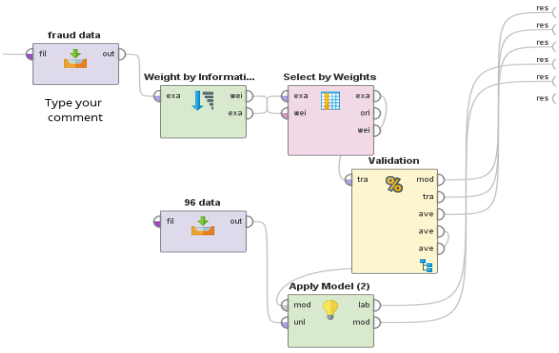
جدول ۱۰: جمع بندی نتایج و شاخص های هر تکنیک در کشف تقلب صورت های مالی با استفاده از داده های سه ساله ۹۳ الی ۹۵

Model	Acc	sensitivity	Specificity	AUC
SVM	٪ ۸۱,۰۸	٪ ۸۱,۶۰	٪ ۸۰,۲۵	۰,۸۲۵
SVM+Antropy	٪ ۸۱,۳۰	٪ ۸۲,۹۶	٪ ۸۰,۱۲	۰,۸۲۶
Adaboost Svm	٪ ۸۱,۲۸	٪ ۸۱,۷۹	٪ ۸۰,۳۶	۰,۸۳۰
Adaboost Svm +Antropy	٪ ۸۱,۶۹	٪ ۸۲,۹۶	٪ ۸۰,۳۶	۰,۸۲۶
Ann	٪ ۸۰,۰۷	٪ ۸۰,۲۵	٪ ۷۹,۹۵	۰,۸۳۵
Ann+Antropy	٪ ۸۰,۴۷	٪ ۸۱,۳۹	٪ ۷۹,۵۱	۰,۸۴۳
DT	٪ ۸۱,۰۸	٪ ۸۰,۲۲۲	٪ ۸۲,۰۸	۰,۷۹۱
DT+Antropy	٪ ۸۱,۲۸	٪ ۸۰,۲۲	٪ ۸۲,۵۰	۰,۷۹۶

پیش‌بینی و اعمال مدل تشخیص تقلب صورت‌های مالی سال ۱۳۹۶

در این گام هدف این است که قوی‌ترین مدل ساخته شده روی صورت‌های مالی سال‌های ۱۳۹۳ تا ۱۳۹۵ یعنی آدابوست ماشین بردار با دقت ۸۱,۶۹ درصد در شرایط با انتخاب ویژگی برای پیش بینی متقلبانه بودن یا نبودن صورت‌های مالی سال ۱۳۹۶ اعمال شود. اطلاعات صورت‌های مالی سال ۱۳۹۶ شامل ۱۶۴ شرکت است که مدل زیر برای پیش بینی تقلب آن‌ها استفاده شده است.

در الگوریتم ساخته شده در شکل ۱ که در نرم‌افزار رپیدماینر انجام می‌شود؛ هر یک از عملگرها که بصورت جعبه ای آورده شده وظیفه ای انجام می‌دهند؛ بدین صورت که؛عملگر ریداکسر (fraud data) داده‌های مورد استفاده جهت آموزش مدل‌ها را از فایل اکسل می خواند و برای اجرای تکنیک آنتروپی شانون به عملگر Weight By Information داده و عملگر Select By Weightes خروجی‌های وزن‌دهی بر اساس آنتروپی را ارائه می دهد(طبق جدول ۳)؛ که در عملگر Validation که اجرا کننده هر یک از تکنیک‌های هوش مصنوعی می باشد قابل استفاده است؛ و عملگرهای ریداکسر (96 data) و Apply Model جهت اعمال داده‌های آزمایشی و پیش بینی تقلب صورت‌های مالی سال ۱۳۹۶ مورد استفاده قرار گرفته است.



شکل ۱: اعمال مدل برای پیش بینی تقلب صورت‌های مالی ۱۳۹۶

بعد از اعمال این داده‌های نمونه در مدل تشخیص تقلب فوق ، نتایج پیش‌بینی به شرح زیر است

تکنیک آدابست ماشین بردار پشتیبان از ۱۶۴ صورت مالی شرکت‌های سال ۱۳۹۶ ،تعداد۷۹ صورت‌مالی متقلبانه و تعداد ۸۵ صورت مالی را غیرمتقلبانه تشخیص داده است که از ۷۹ صورت‌مالی متقلبانه، ۶۵ صورت‌مالی و از ۸۵ صورت‌مالی غیر-متقلبانه، ۷۱ صورت‌مالی را به درستی تشخیص داده است. بنابراین معیار دقت این مدل برای پیش‌بینی تقلب صورت‌های مالی ٪۸۲,۲۷ می باشد.

جدول ۱۱: ماتریس درهم‌ریختگی آدابوست ماشین بردار با انتخاب ویژگی (داده های آزمایشی)

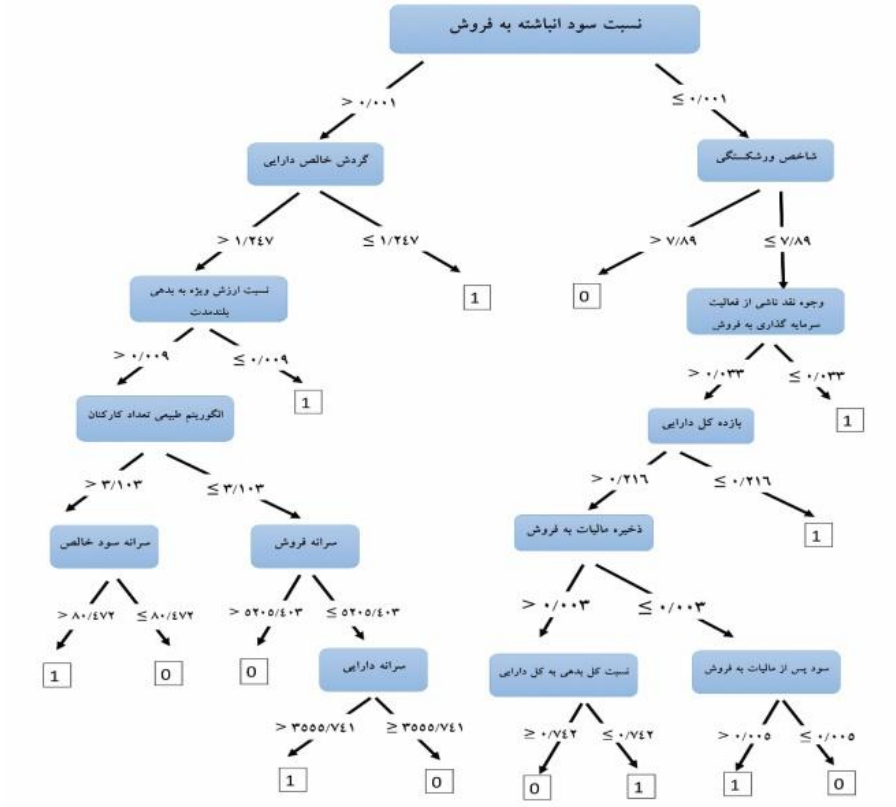
		نمونه‌های تخمینی	
		دسته منفی (−)	دسته مثبت (+)
نمونه‌های واقعی	دسته منفی (−)	۷۱	۱۴
	دسته مثبت (+)	۴	۶۵

- ✓ طبق جدول ۱۱، تعداد صورت‌های مالی که دسته واقعی آن‌ها مثبت (متقلبانه) بوده و دسته‌بند نیز دسته آن‌ها را به درستی مثبت تشخیص داده به تعداد ۶۵ صورت‌مالی بوده است.
- ✓ تعداد صورت‌های مالی که دسته واقعی آن‌ها منفی (غیر متقلبانه) بوده و دسته‌بند، دسته آن‌ها را به نادرستی مثبت تشخیص داده به تعداد ۱۴ صورت‌مالی بوده است.
- ✓ تعداد صورت‌های مالی که دسته واقعی آن‌ها منفی بوده و دسته‌بند نیز دسته آن‌ها را به‌درستی منفی تشخیص داده ۷۱صورت مالی بوده است.
- ✓ تعدادصورت‌های مالی که دسته واقعی آنها مثبت بوده و دسته بند دسته آن‌ها را به نادرستی منفی تشخیص داده ۴ صورت مالی است.

مدل درخت تصمیم و قوانین

درخت تصمیم از جمله روش هایی است که خروجی آن به صورت قوانین قابل تحلیل بررسی می‌شود، از این رو جزو روش‌های جعبه سفید است و در مقابل الگوریتم‌های جعبه سیاه مانند شبکه عصبی و ماشین بردار پشتیبان نتایج آن را می‌توان به صورت قوانین "اگر آنگاه" بررسی کرد و ارتباط بین فاکتورها را در حوزه‌ای از داده کاوی تحلیل کرد.

در این پژوهش نیز در بحث کشف و پیش بینی تقلب صورت-های مالی، شمایی از قوانین مدل درخت تصمیم با دقت خوب به شرح زیر ارائه شده است:



شکل۲: شمای مدل درخت تصمیم

نتیجه‌گیری وپیشنهادات

با عنایت به مطالب مطرح شده 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. بنابراین پژوهش پیش روی، با استفاده از رویکرد کریسپ و مدل‌های سود اطلاعاتی و هوش مصنوعی به کشف تقلب صورت‌های مالی پرداخته است.در این مقاله به بررسی تکنیک‌های زیرمجموعه‌ی هوش مصنوعی شامل؛ ماشین بردار پشتیبان ، درخت تصمیم، شبکه عصبی، و تکنیک ترکیبی آدابوست ماشین بردار پشتیبان در دو حالت با انتخاب ویژگی و بدون انتخاب ویژگی پرداخته‌شده و نتایج آن‌ها مورد بررسی قرارگرفته است. طبق جدول ۱۰ همانطور

که از مرحله مدلسازی و ستونACC مشخص است؛ دقت مدل‌های تکنیک‌های انتخابی در این پژوهش بسیار خوب بوده این مدل ها توانسته‌اند بطور میانگین ۸۱,۰۲٪ از صورت‌های مالی متقلبانه و غیرمتقلبانه را به‌درستی تشخیص دهند و اینکه، انتخاب ویژگی (Antropy) در تمامی تکنیک‌ها باعث افزایش دقت کشف(ACC) شده است؛ در ضمن، براساس داده های آموزشی‌تکنیک‌ترکیبی آدابوست ماشین بردار پشتیبان دقت مدل پایه ماشین بردار پشتیبان را افزایش داده است و هم اینکه در حالت با انتخاب ویژگی بالاترین میزان دقت (۸۱,۶۹٪) را در بین تمامی تکنیک ها داشته است. همچنین این مدل براساس الگوریتم نهایی (طبق داده های آزمایشی براساس شکل ۲ و جدول ۱۱) امکان پیش بینی صورت های مالی متقلبانه با دقت ۸۲,۲۷٪ را فراهم آورد.

یافته‌های پژوهش حاکی از آن است که مولفه‌های 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 و اینکه رویکرد سود اطلاعاتی کمک شایانی در شناسایی متغیرهایی که بیشترین تأثیر را در کشف تقلب صورت‌های مالی دارند می نماید؛ در ضمن با بررسی ۲۷ متغیر انتخابی براساس آنتروپی شانون به این نکته پی‌برده می‌شود که نسبت‌های سودآوری (مانند، نسبت سود و زیان انباشته به فروش) و نسبت‌های فعالیت (مانند، نسبت گردش خالص دارایی‌ها) نسبتاً نقش مؤثرتری در کشف و پیش بینی تقلب صورت‌های مالی ایفا می‌کنند. علاوه بر نتایج فوق می توان به تحلیل نتایج قوانین استخراج شده از مدل درخت تصمیم اشاره نمود؛ بعنوان نمونه برخی از تحلیل های درخت تصمیم طبق شکل ۲بدین شرح است:

- در صورتی سود و زیان انباشته به فروش بزرگتر از ۰,۰۰۱ و گردش خالص دارایی کمتر مساوی ۱,۲۴۷ باشد، آنگاه با احتمال زیاد صورت مالی شرکت متقلبانه است.
- در صورتی سود و زیان انباشته به فروش کوچکتر از ۰,۰۰۱ باشد و شاخص ورشکستگی کمتر از ۷,۸۹ و وجه نقد ناشی از سرمایه گذاری کمتر از ۰,۰۳۳ باشد، آنگاه به احتمال زیاد صورت‌های مالی شرکت متقلبانه است.
- در صورتی که سود انباشته به فروش کوچکتر از ۰,۰۰۱ و شاخص ورشکستگی کمتر از ۷,۸۹ باشد، به احتمال زیاد صورت مالی شرکت غیر متقلبانه است.
- در صورتی که نسبت سود انباشته به فروش بزرگتر از ۰,۰۰۱ و گردش خالص دارایی بیشتر از ۱,۲۴۷ و نسبت ارزش ویژه بیشتر از ۰,۰۰۹ و سرانه فروش بیشتر از ۵۲۰۵,۴۰۳ ارزیابی شود، به احتمال زیاد صورت مالی شرکت غیر متقلبانه است.

نتایج بدست آمده از این پژوهش با نتایج پژوهش‌های چی چن و همکاران (۲۰۱۵)، حمزه زادی و اعتمادی (۱۳۹۵)، تاراسی و همکاران (۱۳۹۷) و جن چن و همکاران (۲۰۱۸) همخوانی دارد. به گونه ای 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؛ اما در هیچ یک از این پژوهش‌ها از تکنیک‌های ترکیبی مانند آدابوست ماشین بردار پشتیبان استفاده نشده است وهمچنین از رویکرد سود اطلاعاتی بصورت توأمان با این تکنیک‌ها کمتر مورد استفاده قرار گرفته است. با توجه به نتایج پژوهش، پیشنهادات کاربردی زیر ارائه می‌گردد:

۱) با نظر به اینکه در این پژوهش رویکرد ترکیبی آدابوست ماشین بردار پشتیبان از بین همه تکنیک‌ها توانایی بالاتری در پیش بینی و کشف تقلب دارد؛ حسابرسان می توانند پس از انجام فرآیند حسابرسی و قبل از اظهار نظر با اجرای تکنیک پیشنهادی این پژوهش نسبت به عدم وجود تقلب در صورت‌هالی مالی اطمینان بالایی حاصل نمایند.

۲) همچنین سرمایه گذاران می توانند با استفاده از تکنیک آدابوست ماشین بردار پشتیبان شرکت‌های متقلب وغیر 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 تکنیک‌های هوش مصنوعی در مباحث پیش بینی و ارزیابی در موضوعات مالی، عملا 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 می باشد؛ که با آموزش بیشتر این گونه پژوهش‌ها و نحوه کارکرد این تکنیک‌ها از طریق برگزاری کارگاه-های آموزشی درآینده؛ می توان از تکنیک‌های هوش مصنوعی بیشتر در مباحث مالی استفاده کرد.

۴) تعداد ۲۷ متغیر پیش بین تقلب که دراین پژوهش بر اساس رویکرد سود اطلاعاتی شناسایی شده است؛ از آنجائیکه جز متغیرهای با اهمیت در کشف تقلب صورت‌های مالی تشخیص داده شده‌اند، می توانند در پژوهش‌های این حوزه مورد استفاده قرار گیرند.

فهرست منابع

* آذرنوش، سید ایرج و همکاران. (۱۳۹۴). پیش بینی ریسک تقلب در صورت‌های مالی با استفاده از شبکه عصبی مصنوعی ، اولین کنفرانس بین المللی حسابداری، مدیریت و نوآوری در کسب و کار، صص۱۴-۲۸.
* امینی، پیمان، محمدی، کامران، عباسی، شعیب، (۱۳۹۰)، "بررسی عوامل مؤثر بر صدور گزارش مشروط حسابرسی:کاربردروش 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 شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران" ، بررسی های حسابداری و حسابرسی، ۱۹(۲)، ۳۱-۴۶.
* حمزه زادی, کبری و حسین اعتمادی، (۱۳۹۵). کاربرد شبکه عصبی مصنوعی در شناسایی تقلب در صورت‌های مالی، چهارمین کنفرانس بین المللی پژوهش‌های کاربردی در مدیریت و حسابداری، تهران، دانشگاه شهید بهشتی.
* خدمتگزار سالانقوچ، یاسر، (۱۳۹۲)، "بررسی خوشه‌بندی در شبکه‌های حسگر بی‌سیم و ارائه یک متد جدید در این رابطه"پایان نامه کارشناسی ارشد کامپیوتر، دانشگاه مهندسی فناوری نوین قوچان.
* رحمانی، حلیمه، رجب‌دُری، حسین و خرمین، منوچهر، (۱۳۹۶)، بررسی تاثیر نظریه عمل منطقی بر تمایل به بروز تقلب در گزارشگری مالی، فصنامه علمی پژوهشی دانش حسابداری و حسابرسی مدیریت، ۶(۲۴)، ۱۳۵-۱۴۳.
* رهنمای رودپشتی. فریدون، (۱۳۹۱)،" داده کاوی وکشف تقلب"،فصلنامه دانش حسابداری وحسابرسی مدیریت، ۱(۳)،۱۷-۳۴.
* 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 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، ۷(۲۸)، ۷۵-۸۶.
*مرادی، جواد؛ رستمی، راحله؛ زارع، رضا. (۱۳۹۳). شناسایی عوامل خطر مؤثر بر احتمال وقوع تقلب در گزارشگری مالی از دید حسابرسان و بررسی تأثیر آنها بر عملکرد مالی شرکت، مجله‌ی پیشرفت‌های حسابداری دانشگاه شیراز، دوره‌ی ششم، شماره‌ی اول، بهار و تابستان ۳۱، صص ۱۷۳-۱۴۱.
* ممشلی، رضا و کارشناسان، علی، (۱۳۹۸)، بهره هوشی شرکت‌ها ابزاری برای بهبود مالی یا گزارشگری خدعه آمیز، فصنامه علمی پژوهشی دانش حسابداری و حسابرسی مدیریت، ۸(۳۱)، ۱۱۷-۱۳۲.
* 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 آموزشی 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.
* Chapman, P., et al. (2000). CRISP-DM 1.0 Step-by-step data mining guide.
* 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.
* 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",Knowledge-Based Systems, 89(15), Pp.459-470.
* Dechow, M. P. , Sloan, R. G. , & Sweeney, A.P.(1995)."Detecting earning management " The Accounting Review, 70 (2) , 193.
* Dechow, M. P. , Sloan, R. G. , & Sweeney, A. P. (1996). " Couses And Consequences Of Earnings Manipulation: An Analysis Of Firms" Contemporary Accounting Research , vol.13 NO.1, PP.1-36.
* Domingues, D., Filippone, M., Michiardi, P., (2019) " Probabilitic Modeling For Novelty Detection With Applications To Fraud Identification,Phd Thesis of Paris Sorbonne University.
* Fama, E. F. and Jensen, M. C. (1983). "Scparation of ownership and control" ,Journal of Law and Economics, Vol. 26. PP: 327-349.
* Fayyad, U., G. Piatetsky-Shapiro, and P. (1996). Smyth, From data mining to knowledge discovery in databases. AI magazine. 17(3): p. 37.
* 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.
* Jen Chen, Yuh, Ching Liou, Wan, Min Chen, (2018), "Fraud Detection For Financial Statements Of Business Groups",International Journal Of Accountion Information Systems" 7(15), Pp10-26.
* Kotsiantis, S., (2006), "Method Of Financing", Australian Accounting Review, Pp: 538-542 (Springer-Verlag Berlin, Heidelberg@ 2006).
* Maher, w. Michael. Stickney, p. clyde. Weil, l. roman,(2007). managerial accounting: an introduction to concepts, methods, and use, 10th ed. Mason.
* 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.
* Rajaram, R, Castellani, B and Wilson, A. N. (2017), "Advancing Shannon Entropy For Measuring Diversity In Systems" Vol. 2017, ID: 8715605,P:10.
* 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.
* Watts, Rl. , &, Zimerman, J. L., (1986), Pozetive Accounting Theory ", 2(124).
* Wirth, R. and J. Hipp. (2000). CRISP-DM: Towards a standard process model for data mining. in Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. Citeseer.

یادداشت‌ها

[1] Jen Chen, et al
[2] Artificial Intelligence
[3] Weight by information gain
[4] Javier, et al
[5] Watts & Zimerman
[6] Dechow, et al
[7] Fama & Jensen
[8] Spathis
[9] Kotsiantis
[10] decision tree
[11] artificial neural networks
[12] support vector machine
[13] Che and Hung
[14] Vapnik
[15] Domingues
[16] Rajaram
[17] Sadgali, et al.
[18] Jen Chen, et al
[19] Normah, et al.
[20] Chi-Chen, et al.
[21] Maher, et al.
[22] Cross-Industry Standard Process-CRISP
[23] Chapman
[24] Wirth, R. and J. Hipp
[25] True positive
[26] False Positive
[27] True Negative
[28] False Negative