فصلنامه علمی پژوهشی
دانش حسابداری و حسابرسی مدیریت
سال دهم / شماره ۳۷/ بهار ۱۴۰۰

کاربرد یادگیری ماشین در ارائه الگویی برای کشف تحریفات حسابداری

افشین هاشمی گل سفیدی
گروه حسابداری ،واحد بین الملل کیش ،دانشگاه آزاد اسلامی ،جزیره کیش ،ایران
afshinhashemi54@yahoo.com

زهرا لشگری
استادیار گروه حسابداری، واحد تهران مرکز دانشگاه آزاد اسلامی، تهران، ایران(نویسنده مسئول)
z_lashgari@iauctb.ac.ir

زهره حاجیها
دانشیار گروه حسابداری، واحد تهران شرق، دانشگاه آزاد اسلامی، تهران، ایران
drzhajiha@gmail.com

تاریخ دریافت: ۹۹/۰۹/۰۸ تاریخ پذیرش: ۹۹/۰۹/۱۱

چکیده

یادگیری ماشینی یک رشته وسیع است که الگوریتم های یادگیری را طراحی کرده که می تواند محرک ها را هدایت کند، زبان گفتاری را تشخیص دهد و تنظیمات پنهان در رشد حجم داده را کشف کند، که داده های مالی نیز از این قاعده مستثنی نیست. بنابراین هدف از تحقیق حاضر، بررسی کاربرد یادگیری ماشین در ارائه الگویی برای کشف تحریفات با اهمیت حسابداری در شرکت های پذیرفته شده در بورس اوراق بهادار تهران است. تحلیل های آماری پژوهش، براساس داده های استخراج شده ۳۰۸ شرکت پذیرفته شده در بورس اوراق بهادار تهران در بازه زمانی ۱۳۸۹ تا ۱۳۹۸(۳۰۸۰ سال – شرکت) انجام و به منظور نمونه گیری از روش غربالگری استفاده شد. تحریفات حسابداری متغیر وابسته که از طریق متغیر مجازی صفر و یک بدست آمده و متغیرهای اقلام تعهدی غیر اختیاری، تغییر در حسابهای دریافتنی، تغییر در موجودی کالا، دارایی های نرم، تغییر در فروش نقدی، تغییر در بازده داراییها و انتشار اوراق بهادار به عنوان متغیرهای مستقل در نظر گرفته شده و برای آزمون فرضیه ها از رگرسیون غیرخطی و به منظور پیاده‌سازی الگوریتم خوشه‌بندی K –Medians و محاسبات مربوطه از نرم افزار محاسبات اماری R استفاده شده است.

نتایج یافته ها در سطح اطمینان ۹۵٪ نشان می‌دهد از میان نسبت های مالی مشخص شده در مدل اول فقط متغیرهای دارایی های نرم، تغییر در فروش نقدی و انتشار اوراق بهادار می توانند توانایی کشف تحریفات حسابداری را بهبود ببخشند و دیگر متغیرهای اقلام تعهدی غیر اختیاری، تغییر در حسابهای دریافتنی، تغییر در موجودی کالا و تغییر در بازده داراییها توانایی بهبود کشف تحریفات حسابداری در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را دارا نمی باشند. همچنین در مدل دوم نسبت های مالی مشخص شده و با اضافه کردن متغیر DevScore برای گروه هایی که بر اساس صنعت و اندازه ساخته شده اند، مدل اصلاح شده و کشف تحریفات حسابداری را بهبود می بخشد، زیرا این متغیر توانست با اضافه شدن در مدل ضریب تشخیص مک فادن را از ۴۱,۶ درصد به ۴۹,۵ درصد افزایش دهد.

تکنیک یادگیری ماشین، طبقه بندی K-Medians را به نسبت های مالی کلیدی شرکت های مرتبط با اهداف تحقیق خاص اعمال می کند. شرکت ها بر اساس نسبت های مالی به خوشه های مختلفی تقسیم می شوند. متغیر DevScore انحراف شرکت ها از همسالان خود را در سال بعد اندازه گیری می کند. هنگامی که DevScore افزایش می یابد، احتمال ناهنجاری های حسابداری نیز افزایش می یابد. همچنین یک شرکت در صورت داشتن نسبت های مالی مرتبط با تحریفات حسابداری که پایین تر از میانگین همسالان خوشه ای خود است، احتمالاً دارای گزارشگری همراه با تحریف است. از این روش می توان برای شناسایی یا پیش بینی وقایع مهم شرکت ها، مانند سوء استفاده های مالی و کشف تحریفات حسابداری شرکت استفاده کرد.

واژه‌های کلیدی: یادگیری ماشین، الگوریتم خوشه‌بندی، کشف تحریفات حسا.

۱- مقدمه

در یادگیری ماشینی یک رشته وسیع است که الگوریتم های یادگیری را طراحی کرده که می تواند محرک ها را هدایت کند، زبان گفتاری را تشخیص دهد و تنظیمات پنهان در رشد حجم داده را کشف کند. داده های مالی نیز از این قاعده مستثنی نیست. این کار با جریان های داده ای که ویژگی های شرکت، خصوصیات حاکمیت شرکتی، گزارش های حسابرسی، داده های بازار و متغیرهای حسابداری را ضبط کرده، کار می کند. الگوریتم های یادگیری ماشینی الگوهای پیچیده ای را در این داده ها تشخیص می دهند، بهترین متغیرها را برای توضیح متغیر انتخاب و از ترکیب مناسب متغیرها برای پیش بینی دقیق نمونه استفاده می کند. آنها کلیدهای باز کردن منابع داده های بزرگ و در حال رشد هستند که می توانند پیش بینی های بهتری داشته و تصمیمات هوشمندانه تری بگیرند. یادگیری ماشینی در علوم اجتماعی مورد توجه بسیاری قرار گرفته است و بطور انکار بر رویکردهای موجود برای تحلیل داده ها می افزاید. اگرچه اصول اساسی کاهش داده ها در تحقیقات حسابداری جدید نیستند، اما یادگیری ماشین در اعداد حسابداری الگوها و رویکردهای جدیدی را ارائه می دهد و به تنظیم کنندگان کمک می کند تا شیوه های گزارش گیری را رصد کنند(رنچر، ۲۰۰۲). مطالعه دیچو و همکاران(۲۰۱۱) یک تحقیق در این زمینه است. آن ها یک مدل پیش بینی را ایجاد کرده که با استفاده از متغیرهای صورت های مالی، یک احتمال لجستیکی از بی نظمی بودن حسابداری را برای هر سال از شرکت به دست می آورد.

با پیدایش شرکت های سهامی که منجر به مشکل نمایندگی شده و همچنین طبق تئوری اثباتی هر فردی به دنبال حداکثر کردن منافع و ثروت خویش است. مدیران نیز از این قاعده مستثنا نمی باشند. در نتیجه تحریف در صورت های مالی از این طریق زاده شد و قدمت طولانی دارد. تحریفات صدماتی را در سطح خرد به سرمایه گذاران و سایر ذینفعان و در سطح کلان به اقتصاد وارد می نماید. وجود تحریف انگیزه سرمایه گذاری را از افراد گرفته و سبب کاهش فعالیت های مولد می شود. همچنین حجم تحریفات یکی از شاخص های اصلی فساد می باشد که خود مانعی برای سرمایه گذاری به حساب می آید(لمبزدورف، ۲۰۰۳). وجود تحریف در صورت های مالی باعث کاهش اعتماد عمومی می شود و برگرداندن این اعتماد مستلزم صرف هزینه و زمان زیادی است. تحریفات صورت های مالی در دو دسته اشتباهات و تقلب قرار می گیرند. گزارشگری مالی متقلبانه می تواند در راستای کاهش سود با هدف کاهش مالیات بر درآمد صورت گیرد. همچنین مدیریت شرکت ها ممکن است در زمانی که سود بالا است آن را کمتر از واقع گزارش کنند تا با شناسایی آن در دوره های آینده، سود دوره های آتی را افزایش دهند(باباجانی و تحریری، ۱۳۹۲). بنابراین تحقیق حاضر به دنبال ارائه یک الگوی پویا است که این الگو می تواند به کشف ناهنجاریهای احتمالی که ممکن است از وقوع رویدادهای مهم و قابل توجه در اینده حکایت داشته باشد، کمک کند(برای مثال کشف تحریفات با اهمیت حسابداری).

بطور کلی یادگیری ماشین می تواند در سه نوع از روش های تحقیق به کار گرفته شود که عبارتند از رگرسیون، طبقه بندی و خوشه بندی. خوشه بندی K -Medians یک تکنیک یادگیری ماشین است که در مطالعات حسابداری کاربرد منحصر به فردی دارد(چن و همکاران، ۱۹۹۶؛ هوبرگ و فیلیپر،۲۰۱۰ و ۲۰۱۶). به عبارتی تحقیق حاضر بدنبال این است که آیا خوشه بندی K -Medians می تواند کشف تحریفات با اهمیت حسابداری را بهبود بخشد یا خیر. نتایج حاصل از این پژوهش، می‌تواند در کشف تحریفات با اهمیت حسابداری شرکت ها با توجه به جریان تحقیقاتی ایجاد شده آن قطعاً مورد علاقه و توجه طیف وسیعی از اقشار از جمله مراجع گوناگون تدوین استانداردهای حسابداری، سایر نهادهای قانون گذار مثل بورس های اوراق بهادار و نهادهای نظارتی آنها، مدیران شرکت ها و سایر استفاده کنندگان از صورتهای مالی و اطلاعات حسابداری و بطور کلی ذینفعان واحدهای تجاری و شهروندان جامعه مدنی می باشد. در این پژوهش کشف تحریفات با اهمیت حسابداری با استفاده از یادگیری ماشین مورد توجه است که در بخش های بعدی مقاله به صورت مفصل تبیین می گردد.

۲- مبانی نظری و تدوین فرضیه ها

در یک اقتصاد بازار، اطلاعات مالی نقش مهمی در تصمیم گیری کسب و کار دارد. اطلاعات نادرست و نامتوازن منجر به اشتباه در تصمیم گیری و ضرر در استفاده کنندگان صورت های مالی می شود. بنابراین، اطمینان از شفافیت و صداقت اطلاعات مالی همیشه در اولویت ذینفعان بوده است. یک سوء تفاهم عمدی هنگامی بوجود می آید که مدیریت انگیزه هایی برای مدیریت سود یا ارتکاب کلاهبرداری برای تحقق اهداف خاص، مانند حداکثر رساندن سود شخصی از طریق جبران سهام مبتنی بر سهام دارد (اریکسون و همکاران، ۲۰۰۶). آیا راهی وجود دارد که از طریق آن بتوان تحریف و تقلب گزارش های مالی را کشف کرد؟

واتز و زیمرمن(۱۹۸۶) عنوان می کنند که حسابرسی صورتهای مالی ساختاری کنترلی است که از طریق فراهم کردن اطمینان معقول از اینکه صورت های مالی فاقد تحریف های با اهمیت است به کاهش عدم تقارن اطلاعاتی و حمایت از منافع سرمایه گذاران کمک می کند؛ اما کشف تقلب های مدیریت از طریق فرایندهای معمول حسابرسی، کار چندان سادهای نیست؛ زیرا در ارتباط با ویژگی هایِ تقلبِ مدیریت شناخت زیادی وجود ندارد. در طول دو دهه گذشته، محققان حسابداری سعی در ایجاد روش هایی برای پیش بینی و کشف سوءاستفاده های صورت های مالی به دلیل خطا یا کلاهبرداری کرده اند. تحقیقات اصلی در مورد این موضوع نشان داده است که استفاده از نسبت های ساده مالی می تواند به استفاده کنندگان صورت های مالی اجازه دهد تا شرکت های با تحریف صورت های مالی را با کمی دقت پیش بینی کنند(بایلی و تیلور، ۲۰۰۷). اولین مطالعات در مورد شناسایی تحریف در صورت های مالی، تحقیقات با استفاده از مدل های اقلام تعهدی اختیاری است. پیشگامان این گروه ابتدا دی آنجلو(۱۹۸۶)، سپس فریدلن(۱۹۹۴)، هیلی(۱۹۸۵)، جونز(۱۹۹۱) هستند. گروه دیگری از تحقیقات با استفاده از تکنیک های آماری برای شناسایی کلاهبرداری ها و اشتباهات موجود در صورت های مالی به لطف دقت بالای آن، به طور فزاینده ای محبوب شدند. تحقیقات برجسته در مورد این گروه شامل بنیش(۱۹۹۷ ، ۱۹۹۹) و دیچو و همکاران(۲۰۱۱) است. بیشتر تحقیقات متغیرهای مالی را پیدا کرده که در پیش بینی شرکت های با تحریف صورت های مالی از اهمیت برخوردار است، اگرچه محاسبه متغیرها پیچیده می باشد. کامینسکی و همکاران(۲۰۰۴) با استفاده از بیست و یک نسبت برگرفته از صورت های مالی نشان دادند که از این ۲۱ نسبت، فقط سه مورد از این نسبت ها برای سه سال متوالی معنادار و بقیه به طور مداوم در طول دوره نمونه معنی دار نبود. کوتسیانس و همکاران(۲۰۰۶) نشان دادند نسبت های مالی شرکت ها از توانی بالقوه جهت کشف تقلب در گزارشگری مالی برخوردارند. نسبت های زیر بیشتر در کارهای تحقیقاتی در رابطه با کشف تقلب مورد استفاده قرار می گیرد: نسبت بدهی به کل دارایی ها(کرکوس و همکاران ، ۲۰۰۷ ؛ گگانیس ، ۲۰۰۹؛ دالنیل و همکاران ، ۲۰۱۴) یا نسبت کل بدهی به سهام(اسپاتیس و همکاران ، ۲۰۰۲؛ کرکوس و همکاران ، ۲۰۰۷؛ دالنیال و همکاران ، ۲۰۱۴). نسبت سرمایه در گردش به کل دارایی ها، نسبت دارایی های جاری به بدهی های جاری(لنارد و علم ، ۲۰۰۹ ؛ راویسانکار و همکاران ، ۲۰۱۱). طبق گفته سونگ و همکاران (۲۰۱۴) ، استایس(۱۹۹۱)، انگیزه کلاهبرداری دیگر برای مدیران این است که نشان دهند شرکت در حال رشد است. بدین منظور محققان از نسبت فعالیت، سودآوری، ترکیب دارایی ها استفاده کردند: نسبت فروش به کل دارایی ها، نسبت سود خالص به فروش، نسبت سود خالص به کل دارایی، نسبت دارایی های جاری به کل دارایی ها. کرکوس و همکاران(۲۰۰۷) ادعا می کنند که حاشیه ناخالص نیز مستعد دستکاری است. محققان نسبت های زیر را برای ردیابی کلاهبرداری استفاده کردند: نسبت سود ناخالص به فروش، نسبت ناخالص به کل دارایی. طبق گفته های استایس(۱۹۹۱) ، پرسونس(۱۹۹۵)، کامینسکی و همکاران(۲۰۰۴) ، کیرکوس و همکاران (۲۰۰۷) ، پرولز (۲۰۱۱)، موجودیها، حساب های دریافتنی متغیرهای صورت های مالی هستند که تخمین ذهنی را مجاز می دانند. بنابراین نسبت هایی که برای تعیین چنین اظهارات استفاده می شوند عبارتند از نسبت موجودی کالا به فروش، نسبت موجودی کالا به کل دارایی، نسبت حسابهای دریافتنی به فروش. مبانی نظری تحقیق نشان می دهند که استفاده از نسبت های مالی برای تعیین تحریف و کلاهبرداری صورت های مالی ابزاری مناسب و ساده هستند. با این حال، مشکل تفسیر نتیجه بوجود می آید، یعنی اینکه چه مقدار نسبت مالی نشان می دهد که صورت های مالی تحریف شده هستند. با توجه به مطالب بیان شده و ادبیات موجود فرضیه پژوهش به شرح زیر تدوین می‌گردد:

فرضیه اول: نسبت های مالی می تواند کشف تحریفات حسابداری را بهبود ببخشد.

تحریفات صورت های مالی و کلاهبرداری هم صنعت مالی و هم زندگی روزمره را تحت تأثیر قرار می دهد. تقلب می تواند اعتماد به نفس در صنعت را کاهش دهد، پس انداز را بی ثبات کند و بر هزینه زندگی تأثیر بگذارد. شرکت ها برای حل این مشکل از انواع مدل های پیشگیری استفاده می کنند. علیرغم تلاش زیاد شرکت ها، اجرای قوانین، کلاهبرداری مالی همچنان رو به رشد است. در این مقاله، سعی شده است تجزیه و تحلیل مقایسه ای از تکنیک های کشف تحریفات، مانند تکنیک های یادگیری ماشین، که نقش مهمی در کشف تحریف و تقلب دارد، انجام دهد، زیرا اغلب برای استخراج و کشف حقایق پنهان در پشت مقادیر بسیار زیاد داده ها استفاده می شود. شاغلین در حرفه و پژوهشگران اغلب یک شرکت را با محکی(معیاری) از پیش تعیین شده برای ارزیابی عملکرد آن مقایسه می کنند(کونسه و لایپ، ۲۰۱۰). با وجود جذابیت مفهومی ساده، شناسایی یک الگوی بهینه مناسب کار ساده ای نیست. اغلب از شرکتهای همسان(Peer firms) برای ساخت گروه های کنترل به عنوان الگوی بهینه یا محک استفاده می

شود(برای مثال کوتاری و همکاران، ۲۰۰۵؛ دی فرانکو و همکاران، ۲۰۱۵؛ یونگ و ژنگ ۲۰۱۵) از جمله اینکه بابر و همکاران(۲۰۱۱) از شرکت هایی که به لحاظ صنعت و اندازه همسان بودند را برای ساخت الگوی بهینه هرسال انتخاب کردند و یک سطح نرمال را برای خالص دارایی های عملیاتی بعنوان میانگین ارزش های شرکت همسان تعریف کردند. برای دست یافتن به بهترین نتیجه، شرکت های هم رده یا همسان باید به گونه ای قابل مقایسه باشند که بتوانند اثرات عواملی که مورد توجه پژوهشگران نیست از طرح تحقیق آنها جدا شوند(لی و همکاران، ۲۰۱۵). از آنجایی که متغیرهایی که مورد توجه استفاده کنندگان مختلف است با هم فرق می کنند تعیین شرکت های همسان باید با اهداف خاص هر پژوهش بستگی داشته باشد. سیستم های طبقه بندی صنعت پس از دهه ۱۹۷۰ در شناخت صنایع جدید و نوظهور کند بوده و همچنین، کدهای صنعت فقط گهگاه تغییر می کنند، اگرچه بعید است که همسالان یک شرکت برای مدت طولانی بدون تغییر باقی بمانند. محیط شغلی که به سرعت در حال تحول است، به دنبال روش های طبقه بندی انعطاف پذیر تری است که بتواند موجب پیشرفت تجارت و کسب و کار شود(هوبرگ و فلیپس، ۲۰۱۶). سیستم های طبقه بندی صنعت برخی موانع اجرایی را دارند که مانع کاربرد آنها را در عمل می شود. هولمز و استیونز(۲۰۰۴) استدلال می کنند که بنگاه های اقتصادی که طبق طبقه بندی های استاندارد به عنوان همسالان صنعتی شناخته می شوند، ممکن است واقعاً یکدست نباشند. از طرف دیگر، تحقیقات آکادمیک برای ایجاد روش های انتخاب همسالان مبتکرانه و هدفمند، نسبتاً کند بوده است. فان و لانگ(۲۰۰۰) از داده های جریان ورودی و خروجی کالاها استفاده می کنند که منعکس کننده رابطه دو بنگاه برای تصمیم گیری برای همسالان صنعتی است. رامنات(۲۰۰۲) شرکت ها را به عنوان همسالان در نظر می گیرد اگر حداقل پنج تحلیلگر همزمان آنها را تحت پوشش قرار دهند. هوبرگ و فیلیپس(۲۰۱۰، ۲۰۱۶) چشم انداز بازار محصول را در نظر گرفته و از رویکرد متنی بر شرکت های گروه با شرح محصول مشابه استفاده می کنند. بیسلی و همکاران(۲۰۱۰) دریافتند که سوءاستفاده ها در برخی از صنایع خوشه بندی می شوند و این نشان دهنده اهمیت در نظر گرفتن صنعتی است که یک شرکت در آن فعالیت می کند.

یکی از الگوریتم های یادگیری ماشین، الگوریتم خوشه بندی K –Medians است که مجموعه معینی از مشاهدات را به گروه های معنی دار به گونه ای تقسیم می کند که مشاهدات یک گروه مشابه یکدیگر و با مشاهدات گروه های دیگر تفاوت داشته باشند که به این ترتیب شناسایی شرکت هایی که دارای نسبت های مالی مشابه هستند و ویژگی های خاصی را نشان می دهند، را برآورده می سازد(ون لیبرگن، ۲۰۱۷). این روش در زمینه های متعددی(بطور مثال بدهی های معوق مالیاتی، تلاش حسابرسی و کیفیت حسابداری) نیز کاربرد دارد. بنابراین فرضیه دوم تحقیق نیز به شرح زیر ارائه می گردد:

فرضیه دوم: خوشه بندی K –Medians می تواند کشف تحریفات حسابداری را بهبود ببخشد.

۳- پیشینه تحقیق

توصیف و پیش بینی انواع شرکت هایی که صورت های مالی را تحریف می کنند یا مرتکب کلاهبرداری می شوند، حوزه گسترده ای از تحقیقات است. بسیاری از مطالعات از نمونه هایی استفاده کرده اند که شامل شرکت هایی است که تحت انتشار استاندارد های حسابداری و حسابرسی قرار دارند. در این بخش به برخی تحقیقات انجام شده در این زمینه اشاره می شود:

رضایی و همکاران(۱۳۹۹) به بررسی کشف تقلب صورت‌های مالی با توجه به گزارش حسابرسی صورت‌های مالی پرداختند. نمونه آماری با حجم ۱۶۴ شرکت پذیرفته شده در بورس اوراق بهادار تهران طی مقطع زمانی ۱۳۹۶-۱۳۹۳ به روش نمونه گیری حذف سیستماتیک گزینش و با توجه به نوع گزارش حسابرسی به دو گروه مجزا شامل شرکت های متقلب(۱) و شرکت های غیرمتقلب(۰) دسته بندی شدند. متغیرهای مستقل تاثیر گذار بر تقلب در این پژوهش در برگیرنده ۴۱ متغیر مالی و غیر مالی می باشد که بر اساس مبانی نظری و پیشینه پژوهش انتخاب و در نهایت داده های مربوط به متغیرها که به روش کتابخانه ای جمع آوری گردیده به پنج تکنیک برتر از بین تکنیک های یادگیری ماشین داده شد، که این تکنیک ها شامل؛ شبکه های بیزین، درخت تصمیم، شبکه های عصبی ، ماشین بردار پشتیبان و روش ترکیبی می باشد. پس از بکارگیری این تکنیک ها در کشف تقلب صورت‌های مالی، نتایج نشان داد تمامی تکنیک ها قابلیت کشف تقلب صورت‌های مالی را در سطح نسبتا بالایی دارند و تکنیک پیشنهادی ترکیبی با میزان نرخ پیش بینی ۹۶,۲٪ دارای دقت و توان ارزیابی بالاتری نسبت به سایر تکنیک ها است.

رحیمیان و حاجی حیدری(۱۳۹۸) به بررسی کشف تقلب با استفاده از مدل تعدیل شده بنیش و شناسایی نسبت‌های مالی حساس به تقلب پرداختند. ابتدا از بین ۱۵۰ شرکت با استفاده از مدل تعدیل شده بنیش، شرکت‌های متقلب

شناسایی و از شرکت‌های غیر متقلب تفکیک شده و از بین نسبت‌هایی که تحقیقات انجام شده قبلی معرفی نموده‌اند، ۲۵ نسبت مالی انتخاب شده و پس از اجرای آزمون کولموگروف-اسمیرنوف، ۱۰ نسبت مالی شرکت‌های متقلب و غیرمتقلب تعیین شد. پس از اجرای مدل رگرسیون در سه مرحله، نتایج نشان دادند که نسبت فروش به مجموع دارایی‌ها و نسبت حقوق صاحبان سهام به مجموع دارایی‌ها دو نسبت مالی حساس به تقلب هستند. این مدل در طبقه بندی نمونه موردنظر در این تحقیق از نرخ دقت کلی ۶۹/۱ درصد برخورداراست. در نتیجه این مدل نقش اثربخشی درکشف تقلب صورت‌های مالی داشته است.

تشدیدی و همکاران(۱۳۹۸) با استفاده از الگوریتم زنبورعسل به توسعه روش های کشف تقلب در صورت‌های مالی پرداختند. برای بررسی موضوع سه روش الگوریتم زنبور عسل، الگوریتم ژنتیک و رگرسیون لوجستیک به کارگرفته شد. نمونه آماری این مطالعه متشکل از ۱۲۰ شرکت پذیرفته شده در بورس اوراق بهادار (۶۰ شرکت مشکوک به تقلب و ۶۰ شرکت غیر متقلب) برای دوره زمانی ۱۳۹۶-۱۳۸۵ است. شرکت‌های مشکوک به تقلب بر مبنای ۱) اظهارنظر تعدیل شده حسابرسی، ۲) وجود تعدیلات سنواتی با اهمیت و صورت‌های مالی تجدید ارائه شده در مورد موجودی‌ها و سایر دارایی‌ها و... ۳) وجود اختلافات مالیاتی با حوزه مالیاتی طبق یادداشت ذخیره مالیات بر درآمد و پرونده مالیاتی و بند شرط گزارش حسابرسی انتخاب شدند. پس از استفاده از آنتروپی متقابل، ۱۶ نسبت مالی به عنوان پیش‌بینی کننده‌های بالقوه گزارشگری مالی متقلبانه معرفی شدند. نتایج نشان داد که روش الگوریتم زنبور عسل با دقت پیش‌بینی ۸۲,۵ درصد نسبت به دو روش الگوریتم ژنتیک با دقت ۷۷,۵ درصد و رگرسیون لوجستیک با دقت ۷۲,۵ درصد، از عملکرد بهتری جهت شناسایی شرکت‌های مشکوک به تقلب در صورت‌های مالی برخوردار است.

خواجوی و ابراهیمی(۱۳۹۶) به بررسی متغیرهای اثرگذار در کشف تقلب در صورتهای مالی و با به کارگیری تکنیک های داده کاوی مدلی برای کشف تقلب در صورتهای مالی شرکتهای پذیرفته شده در بورس اوراق بهادار تهران پرداختند. برای پاسخگویی به این سؤال از یک نمونه تطبیقی متشکل از ۶ شرکت متقلب و ۶۶ شرکت غیر متقلب(نمونه نامتوازن) به همراه تکنیک های طبقه بندی شبکه عصبی مصنوعی، شبکه بیزین و الگوریتم جنگل تصادفی طی سالهای ۱۳۹۳-۱۳۷۹ استفاده گردید. نتایج نشان داد که مدل های پیشنهادی و برتری الگوریتم جنگل تصادفی و شبکه بیزین برای پیشبینی تقلب در صورتهای مالی هستند. نتایج حاصل از انتخاب ویژگی به روش مبتنی بر همبستگی حاکی از سودمندی متغیرهای نسبت پوشش بهره، نسبت حسابهای دریافتنی به کل داراییها، نسبت موجودی کالا به فروش خالص، نسبت نقدی، لگاریتم طبیعی فروش، نسبت سود خالص به فروش و نسبت جمع داراییهای جاری به کل داراییها برای کشف تقلب بود.

زارع و ملکیان(۱۳۹۴) به بررسی پیش بینی تقلب در صورتهای مالی با استفاده از نسبت های مالی در شرکت های پذیرفته شده در بورس اوراق بهادار تهران از سال ۱۳۸۵-۱۳۹۳ پرداختند. تعداد ۱۸۱ شرکت متناسب با محدودیت ها انتخاب و برای برآورد تقلب، بندهای موجود در اظهار نظر حسابرس مستقل شرکت ها طبقه بندی و با نظرسنجی از خبرگان این بندها با استفاده از مدل AHP فازی رتبه بندی و در نهایت درصدهایی برای هر سال شرکت اختصاص داده شد. علاوه بر این، متناسب با مبانی نظری نسبت های مالی استخراج شد که در نهایت با استفاده از آزمونهای آماری نسبت سرمایه در گردش به دارایی، حساب دریافتنی به فروش، وجه نقد به بدهی جاری، موجودی به دارایی جاری، بدهی به حقوق صاحبان سهام، سود ناخالص به دارایی و قدرمطلق تغییرات در نسبت جاری انتخاب شدند.

صفرزاده(۱۳۸۹) با استفاده از تحلیل لاجیت در داده های مقطعی به بررسی نقش دادههای حسابداری در ایجاد یک الگو برای کشف عوامل مرتبط با تقلب در گزارشگری مالی پرداخت. نمونه آماری متشکل از ۱۷۸ شرکت(۶۶ شرکت متقلب و ۱۱۲ شرکت غیرمتقلب) پذیرفته شده در بورس اوراق بهادار برای دوره زمانی ۱۳۸۳-۱۳۸۶ است. پس از تحلیل، ده نسبت مالی به عنوان پیش بینی کننده های بالقوه گزارشگری مالی متقلبانه معرفی شد. نتایج نشان داد که الگوی تحقیق، توانایی کشف تقلب در گزارشگری مالی را دارد و الگوی پیشنهادی می تواند به گروه های مختلف استفاده کننده همچون حسابرسان، مقامات مجاز مالیاتی، سیستم بانکی و ... در تمایز شرکتهای متقلب از غیرمتقلب کمک کنند.

سادگالی و همکاران(۲۰۱۹) به بررسی تکنیک های برتر یادگیری ماشین در زمینه کشف و پیشگیری از تقلب مالی پرداختند. در این پژوهش پس از شناسایی تکنیک ها و روش های مختلف یادگیری ماشین که در کشف تقلب صورت های مالی در پژوهش های گذشته استفاده شده بودند؛ آنها را از نظر شاخص هایی چون دقت، حساسیت، بلادرنگ بودن، بانک اطلاعاتی و مشاهدات مورد مقایسه قرار دادند. نتایج نشان داد که تکنیک های کشف تقلب دورگه(هایبردی) نسبت به سایر

تکنیک ها، عملکرد بهتری دارند بویژه زمانیکه آن ها چندین روش کشف تقلب سنتی را ترکیب می نمایند.

جن چن و همکاران(۲۰۱۸) به بررسی رویکردهای کشف تقلب صورت های مالی گروه های تجاری پرداختند. در این پژوهش با بکارگیری تکنیک SVM، KNNs، ANNs، TREE به کشف تقلب صورت های مالی پرداخته شد و در نهایت مکانیسم های کشف تقلبی که دارای دقت و توانایی بالایی در کشف تقلب بودند شناسایی نمودند و توسعه دادند که در بین این تکنیک های معرفی شده، تکنیک ماشین بردار پشتیبان با دقت۹۴ درصد و توانایی ۹۴,۷ درصد قابلیت بالاتری در کشف تقلب صورت های مالی داشت.

تران و همکاران(۲۰۱۷) به بررسی شرایط کنونی تحریف های بااهمیت از جمله تقلب و اشتباه در صورت های مالی شرکت های پذیرفته شده هوشی مین پرداختند. این تحقیق بر اساس مدل تحقیقاتی بنیش(۱۹۹۹) با متغیرهای کنترل اضافی مانند بازده دارایی، اندازه شرکت و اهرم مالی انجام گردید. داده های تحقیق شامل ۲۰۸ شرکت در بازه زمانی سال ۲۰۱۴ تا ۲۰۱۶ با ۶۲۴ مشاهده بود. نتایج نشان داد که ۳ متغیر شاخص ناخالص ناخالص، شاخص استهلاک و شاخص فروش و هزینه های عمومی و اداری بر احتمال کلاهبرداری و خطا در صورت های مالی تأثیر می گذارد. همچنین سایر متغیرهای کنترل اضافه شده در مدل ها، همه دارای تأثیراتی با معنای آماری هستند.

لین و همکاران (۲۰۱۶) با نمونه ای متشکل از ۱۲۹ شرکت متقلب و ۴۴۷ شرکت غیر متقلب به طبقه بندی و رتبه بندی عوامل تقلب طی سالهای ۱۹۹۸- ۲۰۱۰ پرداختند. در این پژوهش ۳۲ عامل تقلب که مطابق نظر متخصصان برای کشف تقلب مناسب شناخته شدند بکار رفته است. نتایج حاصل از تکنیک های داده کاوی رگرسیون لجستیک، درخت تصمیم و شبکه های عصبی مصنوعی حاکی از دقت بیشتر روشهای درخت تصمیم و شبکه های عصبی مصنوعی نسبت به رگرسیون لجستیک است

کاناپیکن و همکاران (۲۰۱۵) در یک تحقیق تجربی با عنوان مدل کشف تقلب صورت های مالی بر مبنای نسبتهای مالی،۴۰ مجموعه از صورت های مالی متقلبانه و ۱۲۵ مجموعه از صورت های مالی غیرمتقلبانه را به منظور مشخص کردن حساس ترین نسبت های مالی نسبت به تقلب و ارزش هر یک در نشان دادن صورت های مالی متقلبانه را مورد بررسی قرار دادند. آنها از رگرسیون لجستیک استفاده و نتایج نشان داد که در بیشتر موارد تقلب به این دلیل انجام می شود که نشان دهند شرکت در حال رشد است و همچنین به منظور انجام تعهدات اجباری در این مطالعه ۵۱ نسبت مالی بررسی شد.

۴- روش تحقیق

این این پژوهش از نوع تجربی در حوزه تحقیقات اثباتی حسابداری می‌باشد. در این پژوهش اطلاعات مالی از صورت‌های مالی و یادداشت‌های مربوط به شرکت‌های مورد مطالعه و با کمک لوح‌های فشرده سازمان بورس و اوراق بهادار تهران استفاده می‌شود. روش پژوهش فوق استقرائی و کمی است، بدین معنی که مبانی نظری و پیشینه پژوهش از راه مطالعات کتابخانه ای، مقالات و مسیر سایت ها به صورت قیاسی و گردآوری اطلاعات و داده‌ها برای آزمون فرضیه‌ها در قالب استقرائی انجام گرفته است. از آنجایی که نتایج حاصل از این پژوهش می‌تواند در فرآیند تصمیم گیری استفاده شود، این پژوهش از لحاظ هدف تحقیق، کاربردی می باشد. داده‌های مالی مورد نیاز با مراجعه به سازمان بورس اوراق بهادار تهران، گزارش‌های هفتگی و ماهنامه بورس اوراق بهادار و هم‌چنین سایت‌های رایانه‌ای مربوط حاصل شده است. جامعه آماری تحقیق، شرکت های پذیرفته شده در بورس اوراق بهادار تهران در بازه زمانی ۱۳۸۸ تا ۱۳۹۷ است. در این تحقیق به منظور نمونه گیری از روش غربالگری استفاده شده است. بدین ترتیب که ابتدا تمامی شرکت های پذیرفته شده در بورس تهران انتخاب گردید و در نهایت شرکت هایی انتخاب شدند که شرایط زیر را دارا باشند:

۱) ۱)شرکت‌هایی که در بازه زمانی تحقیق (سال ۱۳۸۹ الی ۱۳۹۸) در بورس اوراق بهادار حضورداشته باشند.
۲) شرکت‌هایی که سال مالی آن‌ها منتهی به ۲۹ اسفند باشد.
۳) شرکت‌هایی که جزء شرکت‌های لیزینگ و واسطه‌گری مالی نباشند.
۴) اطلاعات لازم در خصوص آن‌ها در بازه زمانی تحقیق (سال ۱۳۸۹ الی ۱۳۹۸) در دسترس باشد.

در نهایت با توجه به موارد فوق ۳۰۸ شرکت در دوره تحقیق به عنوان نمونه آماری مورد مطالعه انتخاب گردید.

در پژوهش حاضر از مدل دیچو و همکاران استفاده شده است.دیچوو همکاران (۲۰۱۱) مدل تشخیص تحریف اشتباه یک مدل رگرسیون لجستیک با متغیر وابسته برابر با یک برای سال های شرکت شامل تحریف های حسابداری و در غیر این صورت صفر است. متغیرهای مستقل شامل اقلام تعهدی ، تغییر در حسابهای دریافتنی ، تغییر در موجودی کالا ، دارایی

نرم، تغییر در فروش نقدی ، تغییر در بازده دارایی و صدور اوراق بهادار) می باشد. مدل به شرح زیر مشخص شده است:
برای آزمون و محاسبه متغیرها از مدل های ارائه شده در تحقیق دینگ و همکاران (۲۰۱۹) استفاده و به شرح زیر می باشد:

1) $Misstate_{it}=\beta_0 + \beta_1 RSST_{it}+ \beta_2 Change\ in\ receivables_{it} + \beta_3 Change\ in\ inventory_{it}+ \beta_4\ Soft\ assets_{it}+ \beta_5\ Change\ in\ cash\ sale_{it}+ \beta_6 Change\ in\ ROA_{it}+ \beta_7\ Issue_{it}+ \varepsilon$

2) $Misstate_{it}=\beta_0 + \beta_1 RSST_{it}+ \beta_2 Change\ in\ receivables_{it}+ \beta_3 Change\ in\ inventory_{it}+ \beta_4\ Soft\ assets_{it}+ \beta_5 Change\ in\ cash\ sale_{it}+ \beta_6 Change\ in\ ROA_{it}+ \beta_7\ Issue_{it}+ \beta_8\ DevScore_{it+}\ \varepsilon$

تعریف عملیاتی و طبقه‌بندی متغیرهای تحقیق حاضر به شرح زیر است:

تحریفات حسابداری (Misstate): متغیر وابسته تحقیق بوده که با بهره گیری از الگوی مطرح شده دیچو(۲۰۱۱) با استفاده از نسبت های مالی، حساب های دریافتنی به فروش، حاشیه سود ناخالص، کیفیت دارایی، نسبت فروش، نسبت استهلاک، نسبت فروش، هزینه عمومی و تبلیغات به فروش، نسبت بدهی، نسبت اقلام تعهدی در صورتی که نسبت محاسبه شده بالاتر از میانه باشد مقدار یک و در غیر این صورت صفر می باشد.

متغیر های مستقل شامل:

اقلام تعهدی غیر اختیاری(RSST): از تقسیم (تغییر در سرمایه در گردش غیرنقد + تغییر در خالص داراییهای عملیاتی غیرجاری + تغییر در خالص داراییهای مالی) به میانگین مجموع دارایی ها بدست می آید.

تغییر در حسابهای دریافتنی (Change in receivables): از تقسیم تغییر در حسابهای دریافتنی به میانگین مجموع داراییها بدست می آید.

تغییر در موجودی کالا (Change in inventory): از تقسیم تغییر در موجودی به میانگین مجموع دارایی ها بدست می آید.

دارایی های نرم(Soft assets): از تقسیم (مجموع داراییها – داراییهای ثابت – وجوه نقد و معادل آن) به مجموع دارایی ها بدست می آید.

تغییر در فروش نقدی(Change in cash sale): از درصد تغییر در فروش نقدی بدست می آید.

تغییر در بازده داراییها(Change in ROA): از رابطه زیر بدست می آید:

(سود ناخالص سال جاری / میانگین مجموع داراییهای سال جاری) - (سود ناخالص سال قبل / میانگین مجموع داراییهای سال قبل)

انتشار اوراق بهادار(Issue): متغیر مجازی در صورتی که شرکت در سال جاری اوراق بهادار منتشر کند مقدار ۱ و در غیر این صورت صفر می باشد.

سپس مدل(۱) با روش های گروه بندی بر اساس طبقه بندی صنعت و اندازه شرکت مقایسه می شود. تعداد صنایع در بورس اوراق بهادار تهران شرکت های موجود در هر خوشه در سال گذشته(t-1) بعنوان شرکت های همسان در نظر گرفته می شوند. در سال جاری(t) نسبت های مالی هر شرکت با میانگین شرکت های همسان برای محاسبه DevScore مقایسه می شود. با ادامه این گام ها پس از محاسبه DevScore برای گروه هایی که بر اساس صنعت و اندازه ساخته شده اند، به منظور ارزیابی اینکه آیا این مدل اصلاح شده با DevScore کشف تحریفات حسابداری را بهبود می بخشد یا خیر، بعنوان یک متغیر مستقل وارد مدل می شود.

DevScore: متغیر مجازی بوده که تعداد نسبت های بالاتر از میانگین همسالان مشخص شده با روش خوشه بندی K-medians می باشد.

برای خوشه‌بندی شرکت‌ها در سال tام از شاخص‌های اقلام تعهدی غیر اختیاری(RSST)، تغییر در حسابهای دریافتنی (Change in receivables)، تغییر در موجودی کالا (Change in inventory)، دارایی های نرم(Soft assets)، تغییر در فروش نقدی(Change in cash sale)، تغییر در بازده داراییها(Change in ROA) و انتشار اوراق بهادار(Issue) استفاده شده است. به این ترتیب هر نقطه در الگوریتم خوشه‌بندی K-medians دارای ۷ ویژگی یا متغیر است پس یک نقطه هفت بعدی در نظر گرفته می‌شود. الگوریتم خوشه‌بندی برای هر سال اجرا شده و با خوشه‌های همسان به تفکیک سال، استخراج و نتیجه حاصل از خوشه‌بندی برای محاسبه شاخص DevScore به کار برده می شوند. این شاخص به عنوان یک تابع نشانگر عمل کرده و برای هر ویژگی یا شاخص مالی، مقایسه‌ای بین شرکت‌هایی هر خوشه و میانگین آن ویژگی در خوشه صورت می‌گیرد. که مقدار هر یک از ویژگی آن‌ها از میانگین خوشه همسانشان بیشتر است مقدار ۱ در نظر گرفته می‌شود. در غیر اینصورت مقدار صفر لحاظ خواهد شد. مجموع این مقادیر شاخص DevScore هر شرکت را در سال جاری(t) مشخص می‌کند.

۶- نتایج تحقیق

نخستین مرحله در هر تحلیل آماری و تجزیه و تحلیل اطلاعات، محاسبه‌ی توزیع فراوانی و شاخص‌های توصیفی می‌باشد. بنابراین برای ورود به مرحله‌ی تجزیه و تحلیل، آماره-ی توصیفی داده‌ها که شامل شاخص‌های مرکزی، پراکندگی و انحراف از قرینگی بوده و همچنین آزمون جارگ‌برا که توزیع نرمال پسماندها را بررسی می‌کند، محاسبه گردیده و نتایج در جدول (۱) و (۲) درج شده است.

جدول۱. توزیع فراوانی متغیرهای پژوهش

متغیر		فراوانی	درصد فراوانی
تحریفات حسابداری	بدون تحریف	947	30.7
	با تحریف	2133	69.3
	جمع	3080	100
انتشار اوراق بهادار	بدون انتشار	2243	72.8
	با انتشار	837	27.2
	جمع	**3080**	**100**

جدول(۱) توزیع فراوانی نمونه تحقیق را برای متغیرهای تحریفات حسابداری و انتشار اوراق بهادار را نشان می دهد. با توجه به نتایج بدست آمده ۶۹,۳ درصد شرکت های نمونه دارای صورت های مالی با تحریفات حسابداری هستند و همچنین فقط ۸۳۷ نمونه(سال – شرکت) اوراق بهادار منتشر نموده اند.

مهمترین شاخص مرکزی میانگین است که نشان دهنده نقطه تعادل و مرکز ثقل توزیع بوده، برای مثال میانگین نسبت اقلام تعهدی غیر اختیاری(RSST) برابر با ۰,۲۲۳ می باشد، که نشان می دهد بیشتر داده های مربوط به این متغیر حول این نقطه تمرکز یافته اند. پارامترهای پراکندگی، به طور کلی معیاری برای تعیین میزان پراکندگی داده ها از یکدیگر یا میزان پراکندگی آنها نسبت به میانگین است. از جمله مهمترین پارامترهای پراکندگی انحراف معیار است. مقدار این پارامتر برای متغیر مزبور برابر ۰,۶۵۱ است که میزان پراکندگی متغیر مزبور را در اطراف میانگین نشان می دهد. میزان عدم تقارن منحنی فراوانی را چولگی می نامند. اگر ضریب چولگی صفر باشد، جامعه کاملاً متقارن است و چنانچه ضریب مثبت باشد، چولگی به راست و اگر منفی باشد، چولگی به چپ وجود خواهد داشت. به عنوان مثال ضریب چولگی متغیر اقلام تعهدی غیر اختیاری(RSST) ۵,۷۳۷ می باشد، یعنی این متغیر چولگی به راست دارد. میزان کشیدگی اگر حدود ۳ باشد، منحنی فراوانی از لحاظ کشیدگی وضعیت متعادل و نرمال خواهد داشت، اگر این مقدار بزرگتر از ۳ باشد منحنی برجسته و اگر کوچکتر از ۳ باشد منحنی پهن می باشد. کشیدگی اکثر متغیر ها در این پژوهش بیشتر از ۳ است. نتایج حاصل از آزمون جارک - برا مبنی بر نرمال بودن متغیرهای پژوهش نشان می دهد، با توجه به اینکه سطح معنی داری کلیه متغیرها کمتر از ۰,۰۵ است، در نتیجه فرضیه صفر مبنی بر نرمال بودن متغیرها رد می شود به عبارتی توزیع داده ها نرمال نمی باشد. همچنین با توجه به اینکه متغیر وابسته تحقیق تحریفات حسابداری بصورت متغیر مجازی صفر و یک می باشد بنابراین در این پژوهش در ارتباط با آزمون فرضیه های تحقیق از رگرسیون غیر خطی(لاجیت) استفاده می شود. با توجه به نتایج جدول(۳) ارزیابی همبستگی بین متغیرها بر اساس ضرایب همبستگی، شواهد نشان دادند بیشترین میزان همبستگی بین اقلام تعهدی غیر اختیاری(RSST) با متغیر تغییر در حسابهای دریافتنی (CIR)، میزان آن۶۳,۲ درصد و کمترین میزان همبستگی بین تغییر در بازده داراییها(CIROA) با شاخص DevScore به میزان ۰,۰۶۰- قابل مشاهده است. طبق نتایج حاصل شده در آزمون همبستگی، در این پژوهش همخطی به معنی همبستگی قوی بین متغیرها وجود ندارد و به نظر می رسد مشکل خاصی از این بابت در تحلیل مدل های رگرسیون وجود نداشته باشد.

جدول۲. آمار توصیفی متغیرهای پژوهش

متغیر	RSST	CIROA	CICS	SOFT	CII	CIR	DevScore	DEV(SIC)
میانگین	0.223	0.064	0.349	0.704	0.034	0.080	6.850	۱,۵۶۲
میانه	0.036	0.043	0.191	0.739	0.005	0.007	7	۲
انحراف معیار	0.651	0.070	0.992	0.189	0.130	0.309	2.014	0.868
چولگی	5.737	2.623	10.904	-0.785	8.412	7.255	0.044	-0.036
کشیدگی	42.23	13.55	144.03	3.081	90.75	66.01	2.75	2.50
جارک - برا	214477.4	17821.12	2613631	317.5	1024617	536595.2	8.823	32.607
سطح معنی داری	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000
مشاهده ها	۳۰۸۰	۳۰۸۰	۳۰۸۰	۳۰۸۰	۳۰۸۰	۳۰۸۰	۳۰۸۰	۳۰۸۰

جدول۳. همبستگی متغیرهای پژوهش

متغیر	CICS	CII	CIR	CIROA	DEVSCORE	DEVSIC	ISSVE	MISSTATE	RSST	SOFT
CICS	1									
CII	0.052	1								
CIR	0.038	0.577	1							
CIROA	0.139	0.068	0.032	1						
DEVSCORE	-0.040	0.072	0.109	-0.060	1					
DEVSIC	-0.006	-0.035	-0.034	0.008	0.039	1				
ISSVE	0.010	0.023	0.008	0.018	-0.037	-0.014	1			
MISSTATE	0.035	0.027	0.038	0.025	-0.023	0.0004	0.353	1		
RSST	0.013	0.578	0.632	0.044	0.072	-0.056	-0.012	0.018	1	
SOFT	0.058	-0.007	0.012	-0.058	0.332	0.155	-0.016	0.037	-0.043	1

نتایج آزمون و پیش فرض های مدل های رگرسیونی در جدول(۴) ارائه شده اند. با توجه به جدول(۴) طبق آماره اف آزمون هاسمر–لمشو و معناداری آن که کمتر از سطح خطای ۵ درصد است، همسانی واریانس خطاها وجود دارد و همچنین طبق آماره اف لیمر آزمون چاو و معناداری آماره آن که کمتر از سطح خطای ۵ درصد است، آزمون مدل ها باید به روش ترکیبی انجام شود. طبق آماره کایدو آزمون هاسمن و معناداری آماره آن که کمتر از سطح خطای ۵ درصد است، آزمون مدل ها باید به روش اثرات ثابت انجام شود.

هدف ما در مدل اول، نسبت های مالی می تواند کشف تحریفات حسابداری در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را بهبود ببخشد. لذا می‌توان مدل را به شکل زیر تصریح نمود:

$$Misstate_{it}=\beta_0+\beta_1 RSST_{it}+\beta_2 Change\ in\ receivables_{it}+\beta_3 Change\ in\ inventory_{it}+\beta_4\ Soft\ assets_{it}+\beta_5\ Change\ in\ cash\ sale_{it}+\beta_6 Change\ in\ ROA_{it}+\beta_7\ Issue_{it}+\varepsilon$$

پس از تبیین مدل و انتخاب بهترین روش، نتایج تخمین برای شرکت‌های منتخب به شرح جدول ۵ است.

جدول۴. نتایج آزمون انتخاب الگوی مدل های رگرسیونی جهت آزمون فرضیه های پژوهش

مدل های پژوهش	آزمون	آماره	مقدار	معناداری	نتیجه آزمون
مدل اول	چاو	اف لیمر	۲,۷۸۳	0.00۰	داده های ترکیبی
	هاسمن	کای دو	۱۱,۴۹۷	۰,۱۷۵	اثرات ثابت
	هاسمر– لمشو	اف	۳۴,۰۸۹	0.0۰۰	همسانی واریانس خطاها
مدل دوم	چاو	اف لیمر	۲,۸۸۴	۰,۰۰۰	داده های ترکیبی
	هاسمن	کای دو	۱۱,۷,۱۳	۰,۱۶۴	اثرات تصادفی
	هاسمر– لمشو	کای دو	۲۸,۸۶۲	0.0۰۰	همسانی واریانس خطاها

جدول۵. نتایج آزمون مدل اول

متغیر	ضریب برآورد شده	خطای استاندارد	آماره Z	سطح معنی داری
مقدار ثابت	-1.019242	0.185474	-5.495339	0.0000
CICS	0.052514	0.039352	1.334476	0.1820
CII	-0.068636	0.394903	-0.173804	0.8620
CIR	0.260415	0.177224	1.469414	0.1417
CIROA	0.525867	0.556429	0.945074	0.3446
ISSVE	1.674929	0.089738	18.66466	0.0000
RSST	0.025860	0.082993	0.311595	0.7553
SOFT	0.646172	0.221181	2.921461	0.0035
DevScore	-0.033847	0.010756	-3.146801	0.0009
آماره نسبت درست نمایی (احتمال)	۴۰۵,۹۴(۰,۰۰۰)	ضریب تشخیص مک فادن		۰,۴۱۶

نتایج جدول (۵) نشان می دهد مقدار عددی ضریب تشخیص مک فادن تحریفات حسابداری با متغیرهای مستقل و توضیحی ۰,۴۱۶ می‌باشد. یعنی ۴۱,۶ درصد تغییرات متغیر وابسته تحریفات حسابداری توسط متغیرهای اقلام تعهدی غیر اختیاری(RSST)، تغییر در حسابهای دریافتنی (Change in receivables)، تغییر در موجودی کالا (Change in inventory)، دارایی های نرم(Soft assets)، تغییر در فروش نقدی(Change in cash sale)، تغییر در بازده داراییها(Change in ROA) و انتشار اوراق بهادار(Issue) قابل توضیح است. همچنین مقدار احتمال برآوردشده برای آماره نسبت درست نمایی(LR) برابر با ۰,۰۰۰ بوده لذا معنی‌داری کل رگرسیون قابل تأیید است.

نتایج آزمون فرضیه اول نشان داد متغیرهای دارایی های نرم با آماره Z و سطح معنی داری(۲,۹۲۱)، (۰,۰۰۳۵)، انتشار اوراق بهادار(۱۸,۶۶۴) (۰,۰۰۰) و DevScore (-۳,۱۴۶) ، (۰,۰۰۰) معنادار می باشد.

همچنین اقلام تعهدی غیر اختیاری با آماره Z و سطح معنی داری(۰,۳۱۱) ، (۰,۷۵۵)، تغییر در حسابهای دریافتنی (۱,۴۶۹) ، (۰,۱۴۱)، تغییر در موجودی کالا (-۰,۱۷۳) ، (۰,۸۶۲)، تغییر در فروش نقدی(۱,۳۳۴) ، (۰,۱۸۲)، تغییر در بازده داراییها(۰,۹۴۵) ، (۰,۳۴۴) که بیشتر از سطح خطای ۵ درصد است، معنادار نمی باشد. به عبارتی در سطح اطمینان ۹۵٪ می‌توان گفت از میان نسبت های مالی مشخص شده در مدل فقط متغیرهای دارایی های نرم، انتشار اوراق بهادار و متغیر DevScore می توانند توانایی کشف تحریفات حسابداری را بهبود ببخشند و دیگر متغیرهای اقلام تعهدی غیر اختیاری، تغییر در حسابهای دریافتنی، تغییر در موجودی کالا، تغییر در فروش نقدی، تغییر در بازده داراییها توانایی بهبود مدل کشف تحریفات حسابداری در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را دارا نمی باشند.

در مدل دوم هدف، خوشه بندی K –Medians می تواند کشف تحریفات حسابداری در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران را بهبود ببخشد. لذا می‌توان مدل را به شکل زیر تصریح نمود:

$$Misstate_{it}=\beta_0+\beta_1 RSST_{it}+\beta_2 Change\ in\ receivables_{it}+\beta_3 Change\ in\ inventory_{it}+\beta_4 Soft\ assets_{it}+\beta_5 Change\ in\ cash\ sale_{it}+\beta_6 Change\ in\ ROA_{it}+\beta_7 Issue_{it}+\beta_8 DevScore_{it+}\varepsilon$$

پس از تبیین مدل و انتخاب بهترین روش، نتایج تخمین برای شرکت‌های منتخب به شرح جدول ۶ است.

نتایج جدول (۶) نشان می دهد مقدار عددی ضریب تشخیص مک فادن تحریفات حسابداری با متغیرهای مستقل و توضیحی ۰,۴۹۵ می‌باشد. یعنی ۴۹,۵ درصد تغییرات متغیر وابسته تحریفات حسابداری توسط متغیرهای اقلام تعهدی غیر اختیاری(RSST)، تغییر در حسابهای دریافتنی (Change in receivables)، تغییر در موجودی کالا (Change in inventory)، دارایی های نرم(Soft assets)، تغییر در فروش نقدی(Change in cash sale)، تغییر در بازده داراییها(Change in ROA) و انتشار اوراق بهادار(Issue) قابل توضیح است. همچنین مقدار احتمال برآوردشده برای آماره نسبت درست نمایی(LR) برابر با ۰,۰۰۰ بوده لذا معنی‌داری کل رگرسیون قابل تأیید است.

جدول۶. نتایج آزمون مدل دوم

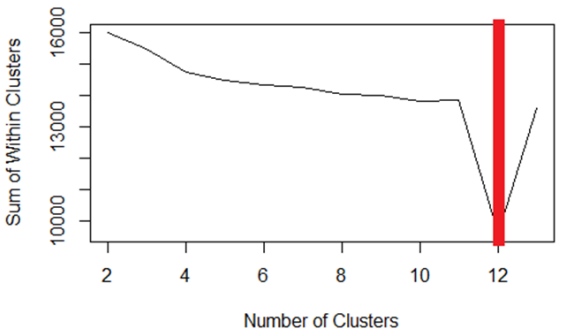
متغیر	ضریب برآورد شده	خطای استاندارد	آماره Z	سطح معنی داری
مقدار ثابت	-1.168496	0.169133	-6.908752	0.0000
CICS	0.056591	0.039250	1.441821	0.1494
CII	-0.078271	0.395204	-0.198051	0.8430
CIR	0.238974	0.176020	1.357651	0.1746
CIROA	0.559669	0.555582	1.007356	0.3138
ISSVE	1.678083	0.089701	18.70758	0.0000
RSST	0.023574	0.082981	0.284092	0.7763
SOFT	0.525169	0.210806	2.491240	0.0127
DEVSIC	0.051574	0.015445	3.339203	0.0004
آماره نسبت درست نمایی (احتمال)	۴۰۵,۹۴(۰,۰۰۰)	ضریب تشخیص مک فادن		۰,۴۹۵

نتایج آزمون فرضیه دوم نشان داد متغیرهای دارایی های نرم با آماره **Z** و سطح معنی داری(۲,۴۹۱)، (۰,۰۱۲۷)، انتشار اوراق بهادار(۱۸,۷۰۷) (۰,۰۰۰) وDEVSIC (۳,۳۳۹)، (۰,۰۰۰) معنادار می باشد.

همچنین اقلام تعهدی غیر اختیاری با آماره Z و سطح معنی داری(۰,۲۸۴) ، (۰,۷۷۶)، تغییر در حسابهای دریافتنی (۱,۳۵۷) ، (۰,۱۷۴)، تغییر در موجودی کالا (۰,۱۹۸-) ، (۰,۸۴۳)، تغییر در فروش نقدی(۱,۴۴۱) ، (۰,۱۴۹)، تغییر در بازده داراییها(۱,۰۰۷) ، (۰,۳۱۳) که بیشتر از سطح خطای ۵ درصد است، معنادار نمی باشد. به عبارتی در سطح اطمینان ۹۵٪ می‌توان گفت از میان نسبت های مالی مشخص شده در مدل فقط متغیرهای دارایی های نرم، انتشار اوراق بهادار و متغیر(DEV(SIC می توانند توانایی کشف تحریفات حسابداری را بهبود ببخشند و دیگر متغیرهای اقلام تعهدی غیر اختیاری، تغییر در حسابهای دریافتنی، تغییر در موجودی کالا، تغییر در فروش نقدی، تغییر در بازده داراییها توانایی بهبود مدل کشف تحریفات حسابداری در شرکت‌های پذیرفته-شده در بورس اوراق بهادار تهران را دارا نمی باشند.

همچنین با اضافه کردن متغیر DEV(SIC) برای گروه هایی که بر اساس صنعت و اندازه ساخته شده اند، مدل اصلاح شده و کشف تحریفات حسابداری را بهبود بخشد، زیرا این متغیر توانست با اضافه شدن در مدل ضریب تشخیص را از ۴۱,۶ درصد به ۴۹,۵ درصد افزایش دهد.

برای تعیین بهترین تعداد خوشه در روش خوشه‌بندی K –Medians، یک نمودار به صورت زیر ترسیم شده که در نقطه‌ای از محور افقی شیب منحنی دارای تغییر محسوسی باشد، مناسب‌ترین خوشه‌بندی صورت گرفته است، زیرا بعد از این شکستگی، با افزایش تعداد خوشه‌ها، تغییر زیادی در شاخص ریشه میانگین مربعات انحراف استاندارد رخ نمی‌دهد. برای داده‌های مربوط به کشف تحریفات حسابداری شرکت‌ها این نمودار به صورت زیر درآمده است. بنابراین تعداد خوشه‌های مناسب برای اجرای الگوریتم K –Medians، حداکثر ۱۲خوشه است. به عبارتی شرکت هایی که پایین تر از این خوشه قرار می گیرند در معرض تحریفات حسابداری می باشند.



شکل۱. خوشه‌بندی K –Medians مدل کشف تحریفات حسابداری

۶- بحث و نتیجه گیری

پژوهشگران غالباً از شرکت های همکار به عنوان معیار برای ارزیابی ویژگی های شرکت های هدف استفاده می کنند. بیشتر تلاش ها به سیستم های طبقه بندی از پیش تعریف شده صنعت مانند طبقه بندی استاندارد صنعت(SIC) و سیستم طبقه بندی صنعت آمریکا (NAICS) متکی است. این مطالعه یک روش جدید برای انتخاب همسالان را ارائه می دهد که یک تکنیک یادگیری ماشین، طبقه بندی K-Medians را به نسبت های مالی کلیدی شرکت های مرتبط با اهداف تحقیق خاص اعمال می کند. هر ساله ، شرکت ها بر اساس نسبت های مالی به خوشه های مختلفی تقسیم می شوند. متغیر DevScore انحراف شرکت ها از همسالان خود را در سال بعد اندازه گیری می کند. هنگامی که DevScore افزایش می یابد، احتمال ناهنجاری های حسابداری نیز افزایش می یابد. از این روش می توان برای شناسایی یا پیش بینی وقایع مهم شرکت ها، مانند سوء استفاده های مالی و تحریفات حسابداری شرکت استفاده کرد.

در این تحقیق به کاربرد یادگیری ماشین در ارائه الگویی برای تحریفات حسابداری شرکت ها در قلمرو زمانی ۱۳۸۹ تا ۱۳۹۸ مورد آزمون و تجزیه و تحلیل قرار گرفت، نتایج و یافته ها نشان داد کل مدل های رگرسیونی معنی داری بود. همچنین نتایج تجزیه و تحلیل رگرسیون نشان داد(مطابق با مطالعات قبلی، بنیش(۱۹۹۷ ، ۱۹۹۹)، کامینسکی و همکاران(۲۰۰۴)، کوتسیانس و همکاران(۲۰۰۶) و دیچو و همکاران(۲۰۱۱)) نشان دادند نسبت های مالی شرکت ها از توانی بالقوه جهت کشف تقلب در گزارشگری مالی برخوردارند. در می یابیم که نسبت های دارایی های نرم، انتشار اوراق بهادار و خوشه بندی براساس صنعت قادر به کشف تحریفات حسابداری یک شرکت هستند. همچنین نتایج حاکی از آن است که یک شرکت در صورت داشتن نسبت های مالی مرتبط با تحریفات حسابداری که پایین تر از میانگین همسالان خوشه ای خود(۱۲) است، احتمالاً دارای گزارشگری با تحریف حسابداری می باشند. یافته‌های تحقیق حاضر تا حدودی با نتایج حاصل از تحقیقات قبلی زاده و همکاران(۱۳۹۸)، اعتبار و همکاران(۱۳۹۸)، آشتاب و همکاران(۱۳۹۶)، هانگ و ین(۲۰۱۹)، پردومو و همکاران(۲۰۱۸) همسو می‌باشد.

یافته های این تحقیق دلالت بر مطالعات نسبت مالی و همچنین کاربردهای عملی دارد. نتایج این استدلال را تأیید می کند که نسبت های مالی می تواند به سرمایه گذاران برای ارزیابی احتمال تجربه رویدادهای مهم شرکت ها کمک کند. تکنیک یادگیری ماشین با مدل های موجود در مورد تشخیص

موارد نادرست و کشف تحریفات ترکیب و مشاهده می شود که اگر اطلاعات مربوط به شرکت های همکار مبتنی بر خوشه بندی K-Medians شامل شود، توانایی این مدل ها افزایش می یابد. همچنین این رویکرد جدید که روی مجموعه ای از شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران آزمون شده است، نشان می دهد که چگونه یک چشم انداز کامل شفافیت (گزارش و افشای) قابلیت اطمینان سیستم را تقویت کرده و به ارزیابی چگونگی کنترل و مدیریت شرکت ها کمک می کند. در نتیجه، این می تواند به عنوان یک روش عملی تر برای طبقه بندی شرکت هایی که مطابق با منافع سهامداران فعالیت می کنند و کمک به آنها در فرآیند تصمیم گیری، به حساب آید. تحقیقات آینده همچنین می تواند این روش خوشه بندی را در زمینه های دیگر پیاده سازی کند. به عنوان مثال، برای حسابرسی، طرح انتخاب همسالان پیشنهادی می تواند ابزاری برای روش های تحلیلی در ارزیابی ریسک باشد. محدودیت های تحقیق علاوه بر محدودیت های اعمال شده توسط محقق مثل انتخاب جامعه، نمونه و دوره های زمانی موارد دیگر عبارتند از :

در روش شناسی این تحقیق از روش های کمی برای کشف تحریفات حسابداری استفاده شده است. در واقع تحلیل تحریفات حسابداری شرکت ها تنها به تعدادی نسبتها مالی شرکت ها محدود نمی گردد و عوامل محیطی مختلفی نیز در آن نقش دارد. بر این اساس، استفاده از روش ها و تکنیک های مفهومی تر و کیفی استفاده نشده لذا تعمیم نتایج این تحقیق باید محتاطانه صورت گیرد.

فهرست منابع

* تشدیدی. الهه، سپاسی. سحر، اعتمادی. حسین، آذر. عادل، (۱۳۹۸)، ارائه رویکردی نوین در پیش بینی و کشف تقلب صورت های مالی با استفاده از الگوریتم زنبور عسل، فصلنامه دانش حسابداری، سال ۱۰، شماره ۳۸، ص ص ۱۶۷-۱۳۹.
* خواجوی. شکراله، ابراهیمی. مهرداد، (۱۳۹۶)، مدل‌سازی متغیرهای اثرگذار برکشف تقلب در صورت‌های مالی با استفاده از تکنیک‌های داده‌کاوی، فصلنامه حسابداری مالی، سال ۹، شماره ۳۳، ص ص ۲۳-۵۰.
* رحیمیان. نظام الدین، حاجی حیدری. راضیه، (۱۳۹۸)،کشف تقلب با استفاده از مدل تعدیل شده بنیش و نسبت‌های مالی. پژوهش های تجربی حسابداری، سال ۸، شماره ۳، ص ص ۴۷-۷۰.
* رضائی. مهدی، ناظمی اردکانی. مهدی، ناصر صدرآبادی. علیرضا، (۱۳۹۹)، کشف تقلب صورت‌های مالی با توجه به گزارش حسابرسی صورت‌های مالی، حسابداری مدیریت، سال ۱۳، شماره ۴۵، ص ص ۱۴۱-۱۵۳.
* صفرزاده. محمد حسین، (۱۳۸۹)، توانایی نسبتهای مالی در کشف تقلب در گزارشگری مالی: تحلیل لاجیت، مجله دانش حسابداری، سال ۱، شماره ۱، ص ص ۱۳۷-۱۶۳.

* Baber, William R, Sok-Hyon Kang, and Ying Li. 2011. "Modeling discretionary accrual reversal and the balance sheet as an earnings management constraint." The Accounting Review 86 (4):1189-1212.
* Bayley, L., & S.L. Taylor. 2007. Identifying Earnings Overstatements: A Practical Test. Working paper.
* Beasley, M.S., J.V. Carcello, D.R. Hermanson, and T.L. Neal. 2010. Fraudulent Financial Reporting: 1998-2007. An Analysis of US Public Companies. Committee of Sponsoring Organizations of the Treadway Commission (COSO).
* Beneish, M., 1997, Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance, Journal of Accounting and Public Policy, 16(3), 271–309.
* Beneish, M., 1999, Incentives and penalties related to earnings overstatements that violate GAAP, The Accounting Review, 74(4), 425–457.
* Dalnial, H., Kamaluddin, A., Sanusi, Z. M., & Khairuddin, K. S. (2014). Accountability in financial reporting: detecting fraudulent firms. Procedia - Social and Behavioral Sciences, 145, 61–69.
* DeAngelo, L.E., 1986, Accounting Numbers as Market Valuation Substitutes: A Study of Management Buyouts of Public Stockholders, The Accounting Review, 61 (3), 400-420.
* Dechow, P.M., W. G., C.R. Larson and R.G.Sloan .2011."Predicting Material Accounting Misstatements", Contemporary Accounting Research, 28 (1), 17-82.
* De Franco, Gus, Ole-Kristian Hope, and Stephannie Larocque. 2015. "Analysts' choice of peer companies." Review of Accounting Studies 20 (1):82-109.
* Erickson, M., Hanlon, M., & Maydew, E. L. 2006. Is there a link between executive equity incentives and accounting fraud? Journal of Accounting Research, (pp. 113–143).
* Friedlan, J. M, 1994, Accounting choices of Issuers of Initial Public Offerings, Contemporary Accounting Research, Volume 11, Issue 1, pages 1–31, USA.
* Gaganis, C. (2009). Classification techniques for the identification of falsified financial statements: a comparative analysis. Intelligent Systems in Accounting, Finance & Management, 16, 207–229.

Learning Algorithms. Auditing: A Journal of Practice & Theory, 30, 19–50.
* Persons, O. (1995). Using financial statement data to identify factors associated with fraudulent financing reporting. Journal of Applied Business Research, 11, 38–46.
* Ravisankar, P., Ravi, V., Raghava R. G., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. Decision Support Systems, 50, 491–500.
* 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.
* Song X. P., Hu Z. H., Du J. G., & Sheng Z. H. (2014). Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China. Journal of Forecasting, 33, 611–626.
* Stice, J. D. (1991). Using Financial and Market Information to Identify Pre-Engagement Factors Associated with Lawsuits against Auditors. The Accounting Review, 66, 516-533.
* Watts, R. L. and Zimmerman, J. L. 1986. Positive Accounting Theory. NJ: Prentice-Hall.
* Young, Steven, and Yachang Zeng. 2015. "Accounting comparability and the accuracy of peerbased valuation models." The Accounting Review 90 (6):2571-2601.
* Healy, P. M, 1985, The effect of bonus schemes on accounting decisions, Journal of Accounting and Economics 7 (1985), 85-107.
* Jen Chen, Yuh, Ching Liou, Wan, Min Chen, Jyun, Han Wu(2018), "Fraud Detection For Financial Statements Of Business Groups" ,International Journal Of Accountion Information Systems" 7(15), Pp10-26.
* Jones, K., 1991, Earnings Management during Import Relief Investigation, Journal of Accounting Research, 29 (2), 193-228.
* Kaminski, K.A., T.S. Wetzel, and L. Guan. 2004. Can Financial Ratios Detect Fraudulent Financial Reporting? Managerial Auditing Journal 19 (1): 15-28.
* Kanapickiene; Rasa; Grundiene; Zivile (2015). The Model of Fraud Detection in Financial Statements by Means of Financial Ratios, International Scientific Conference Economics and Manangement.
* Kirkos, E., Spathis, Ch., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications, 32, 995–1003.
* Koonce, Lisa, and Marlys Gascho Lipe. 2010. "Earnings trend and performance relative to benchmarks: How consistency influences their joint use." Journal of Accounting Research 48 (4):859-884.
* Kotsiantis, S., Koumanakos, E., Tzelepis, D., and Tampakas, V. 2006. Forecasting Fradulent Financial Statements Using Data Mining. Procedings of World Academy of Science, Engineering and Technology, Vol.12, pp. 284-289.
* Kothari, Sagar P, Andrew J Leone, and Charles E Wasley. 2005. "Performance matched discretionary accrual measures." Journal of Accounting and Economics 39 (1):163-197.
* Lambsdorff, J. G. 2003. How corruption affects persistent capital flows. Economics of Governance, 4(3), 229-243.
* Lee, Charles MC, Paul Ma, and Charles CY Wang. 2015. "Search-based peer firms: Aggregating investor perceptions through internet co-searches." Journal of Financial Economics 116 (2):410-431.
* Lenard, M. J., & Alam, P. (2009). An historical perspective on fraud detection: from bankruptcy models to most effective indicators of fraud in recent incidents. Journal of Forensic & Investigative Accounting, 1, 1–27.
* Lin, C-H., A-A. Chiu, S.Y. Huang, and D.C. Yen. (2016). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. Knowledge-Based Systems, Article in Press.
* Manh Dung Tran, Ngoc Hung Dang, Thi Viet Ha Hoang. 2017. Research on Misstatements in Financial Statements: The Case of Listed Firms on Ho Chi Minh City Stock Exchange." International Journal of Applied Business and Economic Research 15 (23):499-518.
* Perols, J. (2011), Financial Statement Fraud Detection: An Analysis of Statistical and Machine