

## ارزیابی خطر تقلب مدیران با استفاده از روش داده کاوی

علیرضا رهروی دستجردی\*

دکتر داریوش فروغی\*\*

دکتر غلامحسین کیانی\*\*\*

### چکیده

بررسی، ارزیابی و تشخیص گزارش‌های متقلبانه که از آن با نام گزارش‌های دستکاری شده نیز یاد شده است، سابقه دیرینه‌ای در ادبیات حسابداری و مالی دارد. سهامداران، مدیران را به عنوان نمایندگان خود در شرکت انتخاب می‌کنند و طبیعی است که نسبت به تشخیص صداقت در گزارشگری مدیران، حساس باشند. حسابرسان نیز در راستای بررسی‌های خود ملزم به ارزیابی و برآورد خطر تقلب در روش‌های اجرا شده توسط مدیریت واحد تجاری هستند تا بر اساس این ارزیابی، حجم آزمون‌های حسابرسی و متعاقباً حق الزحمه پیشنهادی خود را تخمین بزنند. پژوهش حاضر برای ارزیابی و تشخیص خطر تقلب مدیران، از یک روش غیرمالی استفاده کرده است که مشخصاً مبتنی بر تجزیه و تحلیل متن گزارش هیئت‌مدیره به مجمع عمومی صاحبان سهام است. در این روش، واژه‌های گزارش‌های هیئت‌مدیره به مجمع بررسی شده، پس از پالایش شدن، با

---

\* دانشجوی دکتری حسابداری، دانشگاه اصفهان، اصفهان، ایران.

\*\* دانشیار حسابداری، دانشگاه اصفهان، اصفهان، ایران.

\*\*\* استادیار اقتصاد، دانشگاه اصفهان، اصفهان، ایران.

نویسنده مسئول مقاله: دکتر داریوش فروغی (Email: foroghi@ase.ui.ac.ir)

تاریخ پذیرش: ۹۶/۱۱/۸

تاریخ دریافت: ۹۶/۶/۲۵

استفاده از نوع خاصی از رگرسیون تحت عنوان رگرسیون‌های LASSO الگویی برای ارزیابی و تشخیص شاخص خطر تقلب در شرکت‌ها ارائه شده است که با دقتی بین ۸۹٪ تا ۹۱٪ قادر به تشخیص صحیح شاخص خطر بالای تقلب در شرکت‌ها است.

**واژه‌های کلیدی:** خطر تقلب، گزارشگری متقلبانه، داده‌کاوی، رگرسیون‌های LASSO

#### مقدمه

تقلب توسط مدیریت واحد تجاری، به دلیل گمراه کردن ذینفعان از طریق صورت‌های مالی متقلبانه (گمراه‌کننده) بر قیمت سهام، اثر معکوس دارد (الیت و ویلینگهام<sup>۱</sup>، ۱۹۸۰). به همین دلیل، موضوع تشخیص و پیش‌بینی خطر تقلب در گزارشگری مالی، سابقه دیرینه‌ای در ادبیات حسابداری دارد. استاندارد حسابرسی ۲۴۰ ایران تحت عنوان «مسئولیت حسابرسان در ارتباط با تقلب و اشتباه» در زمینه ارزیابی خطر تقلب مدیریت شرکت برای حسابرسان، مسئولیت قائل شده است. از سوی دیگر بر اساس نظریه نمایندگی، سهامداران شرکت نیز به دلیل وجود تضاد منافع با مدیران نسبت به عدم وجود تقلب مدیریت در اداره امور شرکت حساس هستند (جنسن و مک‌لینگک<sup>۲</sup>، ۱۹۷۶). همچنین در ایران نیز قوانین و دستورالعمل‌هایی وجود دارد که به منظور اطمینان از عملکرد صادقانه و به دور از تقلب مدیریت، ضمانت‌های اجرایی خاصی در نظر گرفته‌اند (به عنوان مثال، «فصل ششم از قانون بازار اوراق بهادار»، «دستورالعمل انضباطی ناشران پذیرفته شده در بورس» و «دستورالعمل رسیدگی به تخلفات اشخاص موضوع ماده ۳۵ قانون بازار اوراق بهادار»). این قوانین نشان می‌دهد که نهادهای ناظر نیز نسبت به صادقانه یا متقلبانه بودن گزارش‌هایی که توسط مدیران تهیه می‌شود (مانند گزارش هیئت مدیره به مجمع عمومی صاحبان سهام)، حساس هستند. چون انگیزه‌های تقلب در مدیران می‌تواند در گزارش‌های مذکور منعکس شود.

در همین راستا پژوهش‌های زیادی سعی کرده‌اند با تکیه بر داده‌های مالی، روش‌ها و الگوهایی برای تشخیص تقلب مدیران در شرکت‌ها ارائه دهند. اما پژوهش‌های بسیار اندک و انگشت‌شماری در صدد بوده‌اند که از داده‌های غیرمالی برای این منظور استفاده کنند. در پژوهش حاضر اعتقاد بر این است که با تجزیه و تحلیل متن گزارش‌های تهیه شده توسط مدیران، می‌توان به ارزیابی و تشخیص انگیزه‌های تقلب در مدیران پرداخت. مرور ادبیات، نشان می‌دهد که پژوهشگران با تمرکز بر بخش «تجزیه و تحلیل‌های مدیریت»<sup>۳</sup> از صورت‌های مالی سالیانه به ارزیابی خطر تقلب در مدیران پرداخته‌اند (به عنوان مثال، هامفریس<sup>۴</sup> و همکاران، ۲۰۱۱؛ پوردا و اسکیلیکورن<sup>۵</sup>، ۲۰۱۵؛ فلدمن<sup>۶</sup> و همکاران، ۲۰۱۰ و سچینی<sup>۷</sup> و همکاران، ۲۰۱۰). گزارش هیئت مدیره به مجمع عمومی صاحبان سهام در ایران نیز عملکردی مشابه با بخش «تجزیه و تحلیل‌های مدیریت» در کشورهای دیگر دارد. زیرا هر دوی این گزارش‌ها تصویری از عملکرد مدیران و چشم‌اندازی از فرصت‌های پیش روی شرکت ایجاد می‌کنند.

در کشور ایران و در رشته حسابداری، پژوهشی که خطر تقلب در شرکت‌ها را از طریق تجزیه و تحلیل متن گزارش‌ها ارزیابی کرده باشد، یافت نشد. لذا، سؤال اصلی پژوهش حاضر این است که آیا با تجزیه و تحلیل متن گزارش‌های هیئت مدیره به مجمع عمومی صاحبان سهام، می‌توان به ارزیابی و تشخیص خطر تقلب در شرکت‌ها پرداخت؟

### ادبیات و مبانی نظری

امروزه تجزیه و تحلیل متن<sup>۸</sup> به یکی از ابزارهای مشهور برای بررسی محتوای اطلاعاتی گزارش‌های شرکت‌ها تبدیل شده است (پوردا و اسکیلیکورن، ۲۰۱۵). یکی از مهم‌ترین منابع اطلاعاتی که می‌تواند برای تجزیه و تحلیل متن، مورد استفاده قرار گیرد، گزارش‌های هیئت مدیره است. هامفریس و همکاران (۲۰۱۱) بیان می‌کنند که بخش مذکور، به این دلیل

برای تجزیه و تحلیل متن مناسب است که «نسبت به دیگر بخش‌های گزارش‌های سالیانه، دست مدیران را برای انتخاب لحن و زبانی که مناسب می‌دانند بازتر قرار می‌دهد».

آن طور که در پژوهش لی (۲۰۱۰) توضیح داده شده است؛ مطالعاتی که به تجزیه و تحلیل زبان به کار رفته در متن گزارش‌های شرکت‌ها می‌پردازند، معمولاً دو رویکرد دارند. رویکرد اول تا حدود زیادی از مطالعات قبلی در حوزه‌های زبان‌شناسی و روانشناسی استفاده کرده، بر لیست‌های از پیش تعریف شده از واژه‌ها اتکا می‌کند که اصطلاحاً «جعبه‌واژه‌ها»<sup>۹</sup> نامیده می‌شوند. در این رویکرد، این‌گونه تصور می‌شود که این لیست‌ها با فضاهای احساسی خاصی در متون، مانند منفی‌نگری، خوش‌بینی، فریبکاری یا عدم اطمینان در ارتباط هستند. لوران و مک‌دونالد<sup>۱۰</sup> (۲۰۱۱) معتقدند که این رویکرد در تحلیل متون تجاری و مالی، چندان مناسب نیست (به عنوان مثال: لارکر و زاکولیو کینا<sup>۱۱</sup>، ۲۰۱۲؛ ژول<sup>۱۲</sup> و همکاران، ۲۰۱۰؛ و هامفریس و همکاران، ۲۰۱۱). رویکرد دوم به جای استفاده از لیست‌های از پیش تعیین شده، به روش‌های آماری اتکا کرده و اجازه می‌دهد تا داده‌ها تعیین کنند که کدام واژه، یک واژه مهم هست یا نیست. لی (۲۰۱۰) بیان کرده است که رویکرد اول در هر حوزه نیاز به لیست واژه‌گان تخصصی همان حوزه دارد؛ در حالی که رویکرد دوم، نیازی به داشتن یک لیست منطبق بر متون تخصصی ندارد. به دلیل مزایایی که رویکرد دوم دارد، توسط پژوهشگران مختلفی برای تشخیص صورت‌های مالی متقلبانه مورد استفاده قرار گرفته است (سِچینی و همکاران، ۲۰۱۰؛ هامفریس و همکاران، ۲۰۱۱؛ گلنسی و یاداو<sup>۱۳</sup>، ۲۰۱۱؛ ژول و گانگلی<sup>۱۴</sup>، ۲۰۱۲ و پوردا و اسکیلیکورن، ۲۰۱۵). پژوهش حاضر نیز به منظور تجزیه و تحلیل متن گزارش‌های هیئت مدیره به مجمع، از رویکرد دوم استفاده کرده است.

پژوهشگران به منظور تجزیه و تحلیل متن، معمولاً از روش‌های مبتنی بر «داده‌کاوی»<sup>۱۵</sup> استفاده می‌کنند. داده‌کاوی در حال حاضر یکی از مهم‌ترین ابزارهای موجود در حوزه هوش

تجاری است. حسابداری نیز بیش از سه دهه است که از این تکنیک‌ها استفاده کرده است (بالدوین<sup>۱۶</sup> و همکاران، ۲۰۰۶). حسابداری یکی از اولین حوزه‌های تجاری بود که از روش‌های داده کاوی برای بهبود تصمیم‌گیری‌ها استفاده کرده است (امانی و فادلالا<sup>۱۷</sup>، ۲۰۱۷). داده کاوی به عنوان فرایند تشخیص الگوهای صحیح، نوین و غیرقابل درک (توسط انسان) از مجموعه بزرگی از داده‌ها (پویاری<sup>۱۸</sup>، ۲۰۰۱) و همچنین به عنوان روشی برای استخراج دانش از داخل انبوهی از داده‌ها (هان<sup>۱۹</sup> و همکاران، ۲۰۰۶) تعریف شده است. مرور ادبیات نشان می‌دهد که پژوهشگران حوزه حسابداری (خارج از کشور) به منظور تشخیص تقلب در شرکت‌ها به وفور از مزایای روش‌های مبتنی بر داده کاوی بهره جسته‌اند (به عنوان مثال، هوبرگ و لویس<sup>۲۰</sup>، ۲۰۱۷؛ تکیت<sup>۲۱</sup>، ۲۰۱۳؛ کیم و واسارلی<sup>۲۲</sup>، ۲۰۱۲؛ آلدن<sup>۲۳</sup> و همکاران، ۲۰۱۲ و کاجتواکروسکی<sup>۲۴</sup> و همکاران، ۲۰۱۱).

امانی و فادلالا (۲۰۱۷) پژوهش‌هایی که از داده کاوی در یکی از موضوعات ارزیابی، تشخیص، محاسبه یا پیش‌بینی خطر تقلب در حسابداری استفاده کرده باشند را در یکی از دو دسته «گذشته‌نگر-توصیفی» یا «آینده‌نگر-پیش‌بینی» جای داده‌اند. پژوهش حاضر با تجزیه و تحلیل متن گزارش‌های هیئت مدیره به مجمع شرکت‌ها در سال‌های گذشته، واژه‌هایی که بیشترین قدرت را در تبیین شاخص خطر تقلب در شرکت‌ها دارند شناسایی کرده است. سپس با کمک گرفتن از این واژه‌ها شرکت‌های با خطر بالای تقلب را از سایر شرکت‌ها تفکیک کرده است. بر همین اساس می‌توان این پژوهش را در دسته «گذشته‌نگر-توصیفی» طبقه‌بندی نمود. پژوهش‌های این دسته عمدتاً با بازنگری از وضعیت مالی واحد تجاری در گذشته، به تحلیل‌گران این امکان را می‌دهند که الگوها و روندها را در داده‌ها مشاهده کرده و به مشکلات و روش‌های حل کردن آنها پی ببرند (امانی و فادلالا، ۲۰۱۷).

### پیشینه پژوهش

در بین پژوهش‌های داخلی هیچ پژوهشی یافت نشد که برای ارزیابی و تشخیص خطر تقلب در شرکت‌ها از روش‌های مبتنی بر تجزیه و تحلیل متن استفاده کرده باشد. با این حال، پژوهش‌هایی با موضوع تقلب یا داده‌کاوی به صورت جداگانه انجام شده است (به عنوان مثال، خدادادی و همکاران، ۱۳۹۵؛ نمازی و ابراهیمی، ۱۳۹۵؛ فرج‌زاده دهکردی و آقایی، ۱۳۹۴؛ اعتمادی و زلّقی، ۱۳۹۲؛ رهنمای رودپشتی، ۱۳۹۱؛ صفرزاده، ۱۳۸۹؛ وحیدی و حامدیان، ۱۳۸۸؛ و مهرانی و حصارزاده، ۱۳۸۷) که در اینجا به منظور جلوگیری از اطاله کلام، از ورود به جزئیات آنها خودداری شده است.

در پژوهش‌های خارجی، برای استفاده از روش‌های مبتنی بر داده‌کاوی به منظور تشخیص تقلب مدیریت، مصداق‌های زیادی را می‌توان یافت. در ادامه به تعدادی از جدیدترین و مربوط‌ترین این پژوهش‌ها به موضوع پژوهش حاضر اشاره شده است. هوبرگ و لویس (۲۰۱۷) بخش «تجزیه و تحلیل‌های مدیریت» گزارش‌های شرکت‌ها را تجزیه و تحلیل کرده و دریافتند که لحن گزارش‌های مشکوک به تقلب، به طور بااهمیتی از حالت استاندارد فاصله داشته و به حالت متقلبانانه نزدیک است. به عنوان مثال، مدیران در گزارش‌هایی که مشکوک به تقلب بودند کمتر راجع به جزئیات عملکرد شرکت سخن گفته و سعی می‌کنند بیشتر راجع به نکات مثبت عملکرد خود توضیح دهند و راجع به تیم مدیریتی شرکت نیز جزئیات چندانی افشا نمی‌کنند. وانگ<sup>۲۵</sup> و همکاران (۲۰۱۷) نیز به این نتیجه رسیده‌اند که هم اطلاعات مالی و هم اطلاعات متنی از گزارش‌های تهیه شده توسط مدیریت می‌توانند در تشخیص تقلب، مفید واقع شوند. آنها دریافتند که روش‌های مبتنی بر آموزش ماشین، در تشخیص شرکت‌های متقلب و روش‌های مبتنی بر تحلیل بیز در تشخیص شرکت‌های غیرمتقلب، بهتر عمل می‌کنند.

پوردا و اسکیلیکورن (۲۰۱۵) برای تشخیص تقلب در گزارشگری، متن بخش «تجزیه و تحلیل‌های مدیریت» گزارش‌های سالیانه شرکت‌ها را بررسی کرده و با استفاده از داده کاوی، الگویی با دقت بیش از ۸۲ درصد ارائه دادند. آنها همچنین توانایی این الگو را با هشت روش دیگر که قبلاً در این حوزه کار شده بود مقایسه کرده و دریافتند که روش جدید نسبت به روش‌های قبلی، عملکرد بهتری دارد. ژول و گانگلی (۲۰۱۲) نشان دادند که لحن نوشتن گزارش‌های مدیران و شیوه ارائه این گزارش‌ها قادر است که وجود تقلب را نشان دهد. آنها ویژگی‌های زبانی را بین گزارش‌های متقلبانه و گزارش‌های غیرمتقلبانه بررسی کرده و از این لحاظ، تفاوت معناداری را بین این دو دسته از گزارش‌ها مشاهده کردند.

کیم و واسارلی (۲۰۱۲) از داده کاوی به منظور تشخیص تقلب در صنعت بیمه استفاده کرده‌اند. الگوی آنها مبتنی بر شاخص‌های وجود تقلب در تراکنش‌های الکترونیکی یکی از بزرگترین کمپانی‌های صنعت بیمه در آمریکا بود. پژوهش آنها نشان دهنده کارایی داده کاوی در تشخیص تقلب بود. هامفریس و همکاران (۲۰۱۱) زبان و لحن گزارش‌های متقلبانه را بررسی کرده‌اند. آنها برای این منظور، ۲۴ متغیر را که نشان دهنده لحن گزارش‌ها بود طبقه‌بندی کرده و به این نتیجه رسیدند که گزارش‌های متقلبانه نسبت به گزارش‌های غیرمتقلبانه، بیشتر از لحن فعال (لفاظی) استفاده کرده، تعداد واژگان بیشتر اما تنوع واژگانی کمتری داشته و پیچیده‌تر هستند.

پژوهش‌های زیاد دیگری را نیز می‌توان یافت که برای تشخیص تقلب از روش‌های مبتنی بر داده کاوی استفاده کرده‌اند که در اینجا جزئیات آنها ارائه نشده است (به عنوان مثال، برای مطالعه بیشتر می‌توان به پای<sup>۲۶</sup> و همکاران، ۲۰۱۱؛ ژول و همکاران، ۲۰۱۰؛ سچینی و همکاران، ۲۰۱۰؛ و لئو<sup>۲۷</sup>، ۲۰۰۸ مراجعه کرد).

## سؤال پژوهش

این پژوهش به دنبال تأیید یا رد فرضیه خاصی نیست. هدف اصلی پژوهش حاضر، تمرکز بر متن گزارش‌های هیئت‌مدیره به مجمع عمومی صاحبان سهام، پالایش واژه‌های این گزارش‌ها و در نهایت، ارزیابی و تشخیص شاخص خطر تقلب در شرکت‌ها به کمک واژه‌های مذکور است. بر همین اساس، سؤال پژوهش حاضر به این صورت تدوین شده است: «آیا با استفاده از الگوی مبتنی بر تجزیه و تحلیل متن گزارش‌های هیئت‌مدیره به مجمع، می‌توان بالا بودن شاخص خطر تقلب مدیران در شرکت‌ها را تشخیص داد؟».

## روش نمونه‌گیری و دوره زمانی

به دلیل ماهیت روش پژوهش، شرکت‌ها برای وارد شدن در نمونه این پژوهش باید دو شرط را هم‌زمان داشته باشند. شرط اول، قابلیت تبدیل گزارش‌های هیئت‌مدیره از PDF به فایل Word با کیفیتی قابل قبول و شرط دوم افشا شدن حق‌الزحمه حسابرسی در یادداشت‌های توضیحی صورت‌های مالی است.

پس از بررسی گزارش‌های هیئت‌مدیره به مجمع عمومی تمامی شرکت‌های بورسی طی دوره زمانی پژوهش، تعداد ۵۶ شرکت یافت شد که گزارش‌های آنها حداقل برای چهار سال، قابل تبدیل به فایل Word (به صورت خوانا و قابل استفاده) بودند. معیار «حداقل چهار سال» با روش آزمون و خطا مشخص شده است. افزایش این معیار باعث اندک شدن حجم نمونه و کاهش این معیار نیز باعث کاهش قابلیت اتکای نتایج پژوهش می‌شد. همچنین بررسی یادداشت‌های توضیحی کلیه شرکت‌های بورسی نشان داد که ۸۵ شرکت، حق‌الزحمه حسابرسی را در یک یا تعدادی از سال‌های دوره زمانی پژوهش افشا کرده‌اند.

بنابراین، نمونه پژوهش حاضر را کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران تشکیل می‌دهند که بین دو گروه مذکور (۵۶ شرکتی و ۸۵ شرکتی) مشترک باشند. این

اشتراک، فقط شامل ۳۰ شرکت بود. هیچ محدودیت دیگری (مانند منتهی شدن سال مالی شرکت به پایان اسفند، تولیدی بودن شرکت، عدم توقف نماد شرکت و ...) بر نمونه اعمال نشده است تا حجم نمونه بیش از این کاهش پیدا نکند. کوچک شدن حجم نمونه، محدودیت اصلی پژوهش حاضر بود. با این حال در رویکرد اقتصادسنجی پیشرفته، اعتقاد بر این است که وقتی تعداد کل مشاهدات از ۳۰ مشاهده فراتر رود (مثلاً در داده‌های ترکیبی) بنا بر قضیه حد مرکزی، توزیع ضرایب به سمت نرمال میل کرده و می‌توان ضرایب را سازگار و بدون تورش در نظر گرفت (گرین<sup>۲۸</sup>، ۲۰۱۲؛ باوم<sup>۲۹</sup>، ۲۰۰۶؛ بروکس<sup>۳۰</sup>، ۱۹۷۱ و افلاطونی، ۱۳۹۴).

دوره زمانی پژوهش حاضر نیز سال‌های ۱۳۸۶ تا ۱۳۹۴ است. ارائه گزارش فعالیت هیئت‌مدیره به مجمع عمومی صاحبان سهام از سال ۱۳۸۶ برای شرکت‌های بورسی اجباری شده است. به همین دلیل از سال ۱۳۸۷ به تعداد بسیار کم و از سال ۱۳۸۸ می‌توان تعداد قابل قبولی از گزارش مذکور را در شرکت‌های بورسی یافت. از سوی دیگر چون در برخی الگوها و محاسبه بعضی از متغیرهای پژوهش به داده‌های یک سال قبل و دو سال قبل هم نیاز بوده است دوره تخمین الگوها به سال‌های ۱۳۸۸ تا ۱۳۹۴ محدود شده است. نهایتاً داده‌های مورد نیاز برای اجرای روش‌های آماری این پژوهش در زمان اجرای این روش‌ها حداکثر تا سال ۱۳۹۴ در دسترس بوده است.

### الگوی پژوهش

در اجرای روش مبتنی بر رگرسیون‌های LASSO که در پژوهش حاضر برای تشخیص

شاخص خطر تقلب به کار رفته از رابطه ۱ استفاده شده است:

$$\begin{bmatrix} Y1 \\ Y2 \\ \vdots \\ Ym \end{bmatrix} = \begin{bmatrix} X11 & X12 & \dots & X1n \\ X21 & X22 & \dots & X2n \\ \vdots & \vdots & \vdots & \vdots \\ Xm1 & Xm2 & \dots & Xmn \end{bmatrix} \times \begin{bmatrix} \beta1 \\ \beta2 \\ \vdots \\ \betan \end{bmatrix} \quad (1)$$

متغیر وابسته (Y) به صورت یک ماتریس  $m \times 1$  است که اعضای آن، برای شرکت‌هایی که شاخص خطر تقلب در «تقسیم‌بندی اولیه» برای آنها بالا تشخیص داده شده است معادل یک و برای سایر شرکت‌ها معادل صفر است. همچنین ماتریس متغیرهای توضیحی (X) یک ماتریس  $m \times n$  است که در آن m معادل تعداد شرکت‌سال‌ها و n معادل تعداد واژه‌هایی است که به عنوان متغیرهای توضیح دهنده شاخص خطر تقلب، استفاده شده و باید از بین آنها واژه‌هایی که بیشترین قدرت را در تبیین شاخص خطر تقلب دارند، مشخص شوند. هر عضو این ماتریس که محل تلاقی یک سطر با یک ستون است، فراوانی نسبی یک واژه را نشان می‌دهد که از تقسیم تعداد تکرارهای یک واژه خاص (ستون) به کل واژه‌های یک گزارش خاص (سطر) به دست می‌آید. علت این تقسیم، نوعی مقیاس زدایی است؛ به نحوی که اثر تفاوت بین طول گزارش‌های شرکت‌ها خنثی شود (پوردا و اسکیلیکورن، ۲۰۱۵). ماتریس ضرایب ( $\beta$ ) نیز ضریب (اهمیت) هر واژه را در تبیین شاخص خطر تقلب نشان می‌دهد. در ادامه جزئیات محاسبه ماتریس‌های Y و X توضیح داده شده است.

#### ماتریس متغیر وابسته: شاخص خطر تقلب (Y)

در کشورهای خارجی معمولاً به طور رسمی، لیستی منتشر می‌شود که نام شرکت‌هایی که مشکوک به تقلب هستند را نشان می‌دهد. لذا، پژوهشگر قادر است مستقیماً شرکت‌های موجود در آن لیست را در گروه متقلب (یک) و مابقی شرکت‌ها را در گروه غیرمتقلب (صفر) قرار دهد. به عنوان مثال، پوردا و اسکیلیکورن (۲۰۱۵) و دجو<sup>۳۱</sup> و همکاران (۲۰۱۱) به منظور تفکیک شرکت‌های با خطر بالای تقلب از سایر شرکت‌ها به لیست AAER<sup>۳۲</sup> که در آمریکا با همین مضمون توسط کمیسیون بورس اوراق بهادار منتشر می‌شود اتکا کرده‌اند.

در ایران چنین لیستی با ویژگی فوق به طور رسمی وجود ندارد و در صورت وجود نیز در اختیار عموم پژوهشگران قرار نمی‌گیرد. به همین دلیل نیاز به ساختن شاخصی وجود دارد که

بیشترین همبستگی را با خطر تقلب در شرکت‌ها داشته باشد. در همین راستا و با توجه به اینکه در استاندارد حسابداری ۲۴۰ ایران ارزیابی خطر تقلب جزو مسئولیت‌های حسابرسان دانسته شده است، در پژوهش حاضر از رویکردی برای ساخت این شاخص استفاده شد که مبتنی بر رسیدگی‌های حسابرسان مستقل است.

برای این منظور از رویکرد ارائه شده توسط ریبار<sup>۳۳</sup> و همکاران (۲۰۱۴) استفاده شده است. آنها اثبات کردند که پسماندهایی که از اجرای رگرسیون از حق الزحمه حسابداری بر روی متغیرهای اثرگذار بر این حق الزحمه حاصل می‌شود، قادرند «تقلب» را در سطح اطمینان ۹۹٪ پیش‌بینی نمایند. آنها شاخص مذکور را شاخص UAF<sup>۳۴</sup> نامیدند. در پژوهش حاضر نیز اعتقاد بر این است که شاخص UAF قادر است تخمین زنده مناسبی از خطر تقلب در شرکت باشد. چون از برآورد خطر تقلب توسط یک شخص حرفه‌ای (حسابرس) در شرکت حاصل شده است. بنابراین، با استناد به الگوی ریبار و همکاران (۲۰۱۴) از بخش توضیح داده نشده از حق الزحمه حسابداری (UAF) به عنوان شاخصی از خطر تقلب در هر شرکت سال استفاده شد. الگوی مذکور به شرح رابطه ۲ است:

$$\text{Ln (AUDIT FEE)}_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + \dots + \beta_n X_{ni,t} + e_{i,t} \quad (2)$$

که در آن:

Ln (AUDIT FEE): لگاریتم حق الزحمه حسابداری و  $X_1$  تا  $X_n$ : متغیرهای اثرگذار بر حق الزحمه حسابداری هستند. ریبار و همکاران (۲۰۱۴)  $X_i$ ها را به صورت زیر ارائه داده‌اند:  
 BIG: در صورتی که حسابرس شرکت در لیست گروه الف از حسابرسان اعلام شده توسط جامعه حسابداران رسمی ایران حضور داشته باشد، مقدار یک و در غیر این صورت مقدار صفر را اختیار خواهد کرد. Ln ASSETS: لگاریتم ارزش دفتری کل دارایی‌ها. INV: نسبت موجودی مواد و کالا به کل دارایی‌های سال قبل. REC: نسبت حساب‌های دریافتی به

کل دارایی‌های سال قبل. CR: نسبت دارایی‌های جاری به بدهی‌های جاری. BTM: نسبت ارزش دفتری به ارزش بازار سرمایه. LEV: نسبت کل بدهی‌ها به کل دارایی‌های سال قبل. EMPLS: تعداد کارکنان شرکت. ROA: نسبت سود خالص عملیاتی به کل دارایی‌های سال قبل. LOSS: در صورتی که شرکت در سال قبل یا دو سال قبل زیان داشته باشد برابر یک و در غیر این صورت برابر صفر. AUD OPIN: در صورتی که گزارش حسابرسی غیرمقبول باشد برابر یک و در غیر این صورت برابر صفر. CLIENT: تعداد سال‌هایی که شرکت، مشتری یک حسابرس خاص بوده است. IPO: در صورتی که سال اول ورود شرکت به بورس باشد برابر یک و در غیر این صورت برابر صفر. ISSUE: در صورتی که شرکت در سال جاری، افزایش سرمایه از محل انتشار سهام داشته باشد برابر یک و در غیر این صورت برابر صفر.

پس از اجرای رابطه ۲ و استخراج پسماندها (UAF) برای همه شرکت‌سال‌ها، متغیر UAF دهک‌بندی شده و شرکت‌سال‌هایی که در بالاترین دهک قرار دارند (با هدف اعمال حداکثر احتیاط) به عنوان شرکت سال‌های با خطر بالای تقلب (یک) و شرکت سال‌های موجود در مابقی دهک‌ها به عنوان سایر شرکت‌سال‌ها (صفر) در نظر گرفته شده است.

#### ماتریس متغیرهای توضیحی: واژه‌ها (X)

همانطور که قبلاً اشاره شد متغیرهای توضیحی (ماتریس X) در این پژوهش، همان واژه‌هایی هستند که در گزارش‌های هیئت‌مدیره به مجمع از آنها استفاده شده است. در ادامه، مراحلی که برای تشکیل این ماتریس طی شده است به ترتیب و به صورت خلاصه توضیح داده شده است. در ابتدا گزارش‌های هیئت‌مدیره به مجمع برای کلیه شرکت‌های بورسی بین سال‌های ۱۳۸۷ تا ۱۳۹۴ از سایت‌های «سامانه جامع اطلاع‌رسانی ناشران»<sup>۳۵</sup> و «مرکز پژوهش، توسعه و مطالعات اسلامی»<sup>۳۶</sup> دانلود شده و شرکت‌هایی که کمتر از چهار گزارش هیئت‌مدیره به مجمع (قابل تبدیل به فایل Word با کیفیت قابل قبول) داشتند حذف شدند. سپس

شرکت‌هایی که حصولتی بودند، یعنی سهامداران آنها دارای ویژگی‌های خاصی بوده و ممکن است مدیران آنها احساس مصونیت شغلی بیشتری داشته و در استفاده از روش‌هایی با خطر بالای تقلب، ریسک کمتری احساس نمایند، مشخص شدند. در مجموع ۱۴ شرکت با سهامدارانی با ویژگی‌های فوق پیدا شده و از نمونه حذف شد. در اینجا به منظور حفظ حریم حقوقی، از اشاره به نام این شرکت‌ها و سهامداران عمده آنها پرهیز شده است.

در مرحله بعد، گزارش‌های شرکت‌های عضو نمونه از PDF به Word تبدیل شده است. برای این تبدیل، نرم افزار «پارسی پی دی اف» از سایت شرکت سازنده<sup>۳۷</sup> خریداری و فرایند تبدیل انجام شد. سپس شرکت‌هایی که گزارش‌های آنها پس از تبدیل شدن به Word به دلایل مختلف از کیفیت لازم برخوردار نبودند از نمونه حذف شدند (به عنوان مثال، واژه‌ها پس از تبدیل شدن به هم چسبیده و قابل خواندن نبوده و یا به جای کاراکترهای فارسی، به کاراکترها و سمبل‌های ناخوانا و غیر قابل تشخیص در Word تبدیل شده بود). حال برای تشکیل ماتریس X به لیستی نیاز است که هم واژه‌های تشکیل دهنده هر گزارش و هم تعداد دفعات تکرار هر واژه در آن گزارش را نشان دهد. برای تهیه این لیست از سایت [www.textfixer.com](http://www.textfixer.com) استفاده شد. سپس خروجی مربوطه از این سایت به Excel منتقل شده و درصد مربوط به هر واژه از تقسیم فراوانی آن واژه به جمع فراوانی همه واژه‌ها به دست آمده است. از آنجا که واژه‌های تشکیل دهنده ماتریس X که از آنها به منظور تبیین شاخص خطر تقلب استفاده خواهد شد، باید دارای منطق و پشتوانه باشند و همچنین به منظور دوری جستن از اتکای صرف به نرم‌افزار (جکسن<sup>۳۸</sup>، ۲۰۰۲)، در چندین مرحله، پالایشی روی این واژه‌ها انجام شده و واژه‌هایی که فاقد ماهیت یک واژه با مفهوم بودند حذف شدند. در نهایت ماتریس واژه‌ها (X) با استفاده از لیست حاصل از ۷/۱۱۷ واژه پالایش شده (که فقط شامل واژه‌های بامفهوم است) ساخته شد.

### رگرسیون‌های LASSO

رگرسیون‌های LASSO<sup>۳۹</sup> نوع خاصی از رگرسیون‌های Sparse (در لغت به معنای کم‌پشت و تُنک) است که توسط تیشیرانی<sup>۴۰</sup> (۱۹۹۶) معرفی شده و می‌تواند برای انتخاب بهترین زیرمجموعه از یک مجموعه مرجع، مورد استفاده قرار بگیرد. این روش زمانی استفاده می‌شود که در رگرسیون، تعداد زیادی متغیر توضیحی وجود داشته باشد و هدف این باشد که از بین این تعداد زیاد، فقط تعداد محدودی از آنها برای منظوری خاص انتخاب شوند. بنابراین، در این رگرسیون‌ها یک قید تعریف می‌شود مبنی بر اینکه رگرسیون با این شرط اجرا شود که تعداد زیادی از ضرایب، معادل صفر باشند (یا ضرایب غیر صفر، بیش از تعداد خاصی نشوند). در این پژوهش نیز تعداد زیادی واژه وجود دارند که باید شاخص خطر تقلب را در گزارش‌ها تبیین نمایند. منطقی نیست که همه واژه‌ها چنین قدرتی داشته باشند. بنابراین، کاملاً هم منطقی خواهد بود اگر قیدی قائل شویم مبنی بر اینکه رگرسیون با این شرط اجرا شود که قدرت تبیین تعداد زیادی از واژه‌ها معادل صفر باشند.

روش مذکور توسط درصدی از مشاهدات به عنوان «نمونه آموزش»، آموزش داده شده و ماتریس ضرایب (ضریب هر واژه) به دست می‌آید. پس از به دست آمدن ماتریس ضرایب، روش LASSO قادر خواهد بود که با استفاده از ماتریس مذکور (که اهمیت هر واژه را در تبیین شاخص خطر تقلب نشان می‌دهد) شاخص خطر تقلب را در «نمونه آزمون» تشخیص دهد. پس از تخصیص صفر و یک‌ها به شرکت‌سال‌های گروه آزمون توسط روش LASSO، این مقادیر با مقادیر صفر و یک واقعی (حاصل از الگوی حق‌الزحمه حسابرسی (UAF)) مقایسه می‌شود. دقت الگو در این حالت از حاصل جمع تعداد کل تشخیص‌های صحیح بر کل مشاهدات در نمونه آزمون به دست می‌آید. به این معنی که اگر برای یک شرکت‌سال، شاخص

خطر تقلب در دنیای واقعی (همان UAF) بالا (پایین) تشخیص داده شده است، روش LASSO نیز شاخص خطر تقلب را برای این شرکت سال بالا (پایین) تشخیص داده باشد. در این پژوهش برای اجرای روش LASSO از بسته‌ای که برای همین منظور (با عنوان GLMNET) برای نسخه ۳-۳-۲ از نرم‌افزار R نوشته شده استفاده شده است. بسته مذکور مخصوص اجرای رگرسیون‌های LASSO در نمونه‌هایی است که متغیر پاسخ در آنها (در اینجا شاخص خطر تقلب در گزارش‌ها) از نوع دو وجهی (صفر و یک) باشد (فریدمن<sup>۴۱</sup> و همکاران، ۲۰۱۰).

### نتایج تجزیه و تحلیل‌ها

در ابتدا برای محاسبه متغیر وابسته، رابطه ۲ به صورت ترکیبی اجرا شد. برای اجرای رابطه ۲ از آزمون F لیمر و هاسمن برای تشخیص نوع داده‌های ترکیبی و از آزمون‌های والد تعدیل شده و ولدریج به ترتیب به منظور آزمون همسانی واریانس و خودهمبستگی مرتبه اول بین پسماندها استفاده شده و سپس الگو با استفاده از روش تابلویی با اثرات ثابت و با وزن مقطعی<sup>۴۲</sup> به منظور تخفیف مسأله ناهمسانی واریانس اجرا شد. در اینجا از ارائه نتایج مربوط به ضرایب الگو پرهیز شده و تنها ضرایب تعیین، آماره F فیشر، نتایج آزمون‌های F لیمر، هاسمن، والد تعدیل شده و ولدریج در جدول شماره ۱ ارائه شده است.

جدول شماره ۱. نتایج اجرای رابطه ۲ برای استخراج پسماندها (UAF)

آزمون	فرض صفر	آماره آزمون	معناداری	نتیجه
$R^2$	۰/۸۷	دوربین واتسون:	آماره F	۲۴/۴۴۴
$R^2$ تعدیل شده	۰/۸۴	۱/۷۸	معناداری آماره F	۰/۰۰۰
F لیمر	داده‌های تلفیقی	۳/۰۶۸	۰/۰۰۰	داده‌های تابلویی
هاسمن	اثرات تصادفی	۲۴/۳۷۶	۰/۰۴۱	اثرات ثابت
والد تعدیل شده	همسانی واریانس	۱۷۹۶۲	۰/۰۰۰	ناهمسانی واریانس
ولدریج	عدم خودهمبستگی	۲/۶۱۰	۰/۱۱۷	عدم خودهمبستگی

پسماندهای حاصل از اجرای الگوی فوق، به ده قسمت تقسیم شده و شرکت سال‌های موجود در بالاترین دهک، به عنوان شرکت سال‌های با خطر بالای تقلب (یکک) و مابقی دهک‌ها به عنوان سایر شرکت سال‌ها در نظر گرفته شده است.

جدول شماره ۲. حالت‌های مختلف برای آزمون دقت الگو

ترکیب	گروه آموزش	گروه آزمون	دقت روش LASSO	
			S=۰/۰۵	S=۰/۰۱
ترکیب ۱	آموزش ۲۴ شرکت از بالا طی ۷ سال	آزمون ۶ شرکت از پایین طی ۷ سال	دقت الگو در حالت پنجم	دقت الگو در حالت اول
ترکیب ۲	آموزش ۲۴ شرکت از پایین طی ۷ سال	آزمون ۶ شرکت از بالا طی ۷ سال	دقت الگو در حالت ششم	دقت الگو در حالت دوم
ترکیب ۳	آموزش ۲۴ شرکت از بالا و پایین طی ۷ سال	آزمون ۶ شرکت از وسط طی ۷ سال	دقت الگو در حالت هفتم	دقت الگو در حالت سوم
ترکیب ۴	آموزش ۳۰ شرکت طی سال‌های ماقبل آخر	آزمون ۳۰ شرکت برای سال آخر	دقت الگو در حالت هشتم	دقت الگو در حالت چهارم

پس از استخراج متغیر وابسته (ماتریس  $Y$ ) و با داشتن متغیرهای توضیحی (ماتریس  $X$ ) اقدام به اجرای رگرسیون LASSO در محیط نرم‌افزار R و محاسبه دقت الگو شد. در این مرحله و برای افزایش استحکام و قابلیت اتکای نتایج، دو عامل، تغییر داده شده‌اند و پس از هر تغییر، یکبار رگرسیون LASSO اجرا شده است تا قابلیت مقایسه نتایج حاصل شود. اولین عامل، مقدار تابع هدف (در جدول با  $S$  نشان داده شده است) برای بهینه‌سازی است که هرچه کوچکتر شود تعداد قابل قبول برای متغیرهای معنادار (واژه‌های مهم) کاهش پیدا کرده و رگرسیون با سخت‌گیری بیشتری اجرا می‌شود. مقدار پیش‌فرض برای این پارامتر معمولاً ۰/۰۵ در نظر گرفته می‌شود. ولی در این پژوهش، با مقدار ۰/۰۱ نیز اجرا شده و نتایج با هم مقایسه شده است. دومین عامل، شیوه «ترکیب‌بندی نمونه‌های آموزش و آزمون» است که به چهار

حالت مختلف انجام شده تا قابلیت مقایسه فراهم شود. این رویکرد اصطلاحاً اعتبارسنجی متقابل چندوجهی<sup>۴۳</sup> نامیده می‌شود (هستی<sup>۴۴</sup> و همکاران، ۲۰۰۸). جدول شماره ۲ ترکیب‌های مختلف حاصل از تغییر این دو عامل را نشان می‌دهد. در این جدول منظور از واژه‌های «بالا»، «پایین و وسط» محل قرار گرفتن شرکت‌ها در نمونه ۳۰ شرکتی است که برای حفظ حالت تصادفی، بر اساس حروف الفبا مرتب شده‌اند.

پس از اجرای روش LASSO دقت الگو در حالت‌های یک تا چهار ( $S=0/01$ ) به ترتیب معادل ۸۵/۳۶٪، ۹۴/۵۹٪، ۸۷/۱۸٪ و ۹۲٪ با میانگین ۸۹/۷۸٪ و در حالت‌های پنج تا هشت ( $S=0/05$ ) نیز به ترتیب معادل ۹۲/۶۸٪، ۹۴/۵۹٪، ۸۹/۷۴٪ و ۸۸٪ با میانگین ۹۱/۲۵٪ بوده است. یعنی روش رگرسیون LASSO در حالت‌های ۰/۰۱ و ۰/۰۵ برای پارامتر S به طور میانگین توانسته است به ترتیب با دقت‌های ۸۹/۷۸٪ و ۹۱/۲۵٪ بالا یا پایین بودن شاخص خطر تقلب در شرکت‌ها را به درستی تشخیص دهد. با استناد به این نتایج، می‌توان با استفاده از تجزیه و تحلیل متن گزارش‌های هیئت‌مدیره به مجمع عمومی صاحبان سهام، بالا بودن شاخص خطر تقلب مدیران را تشخیص داد.

جدول شماره ۳. جزئیات اجرای روش LASSO در هشت حالت مختلف

ترکیب	حالت	دقت الگو	خطا در تشخیص	
			خطر بالای تقلب	خطر پایین تقلب
۱	اول	۹۲/۶۸٪	۰	۷/۳۲٪
	پنجم	۸۵/۳۶٪	۷/۳۲٪	۷/۳۲٪
۲	دوم	۹۴/۵۹٪	۰	۵/۴۱٪
	ششم	۹۴/۵۹٪	۰	۵/۴۱٪
۳	سوم	۸۹/۷۴٪	۰	۱۰/۲۶٪
	هفتم	۸۷/۱۸٪	۲/۵۶٪	۱۰/۲۶٪
۴	چهارم	۸۸٪	۰	۱۲٪
	هشتم	۹۲٪	۰	۸٪

همچنین به منظور بررسی بیشتر، دو نوع خطا نیز برای این حالت‌های هشت‌گانه محاسبه شده است که عبارتند از: (۱) خطا در تشخیص خطر بالای تقلب: شاخص خطر تقلب در یک شرکت سال، بالا تشخیص داده شود در حالی که در واقع پایین است. (۲) خطا در تشخیص خطر پایین تقلب: شاخص خطر تقلب در یک شرکت سال، پایین تشخیص داده شود در حالی که در واقع بالا است. جدول شماره ۳ جزئیات این دو خطا را برای این هشت حالت نشان داده است.

### نتیجه‌گیری، دانش‌افزایی و محدودیت‌ها

ایده اصلی داده‌کاوی این است که داده‌ها می‌توانند سخن بگویند. به این معنی که در درون حجم انبوهی از داده‌ها، می‌توان روندها و الگوهای ارزشمندی یافته و از آنها برای تصمیم‌گیری و سیاست‌گذاری استفاده کرد. تشخیص این الگوها و روندها نیاز به روش‌های خاصی دارد که داده‌کاوی نامیده می‌شوند. یکی از انواع داده‌ها که در حجم وسیعی توسط انسان تولید می‌شود، «متن» است و همانطور که ژول و گانگلی (۲۰۱۲) بیان کرده‌اند، متن گزارش‌های هیئت‌مدیره به مجمع، حاوی سرنخ‌های ارزشمندی است که همانند داده‌های مالی، قادرند بالا بودن شاخص خطر تقلب در شرکت‌ها را تشخیص دهد. سؤال اصلی پژوهش حاضر این بود که آیا با تجزیه و تحلیل متن گزارش‌های هیئت‌مدیره به مجمع عمومی صاحبان سهام، می‌توان به ارزیابی و تشخیص شاخص خطر تقلب در شرکت‌ها پرداخت؟ در راستای پاسخ به این سؤال، گزارش‌های هیئت‌مدیره به مجمع به کمک یکی از الگوریتم‌های داده‌کاوی (رگرسیون‌های LASSO) تجزیه و تحلیل شد تا به ارزیابی و تشخیص شاخص خطر تقلب پرداخته شود. نتایج نشان داد که داده‌کاوی قادر است با تجزیه و تحلیل متن گزارش‌های هیئت‌مدیره به مجمع، بالا یا پایین بودن شاخص خطر تقلب در شرکت‌ها را در مجموع با دقتی بین ۸۹ تا ۹۱ درصد به درستی تشخیص دهد.

این پژوهش از چند جهت، برای حوزه حسابداری و حوزه داده کاوی، حائز دانش افزایی است: اول، این پژوهش به منظور تشخیص تقلب در گزارشگری، به جای تکیه بر «اعداد و ارقام» و نسبت‌های مشتق شده از این اعداد و ارقام، بر تجزیه و تحلیل «متن گزارش‌ها» تکیه کرده و قابلیت این روش را نشان داده است. دوم، پژوهش حاضر سعی کرده است از «صورت‌های مالی» فاصله گرفته و برای بخش عمده‌ای از تحلیل‌ها بر «گزارش‌های هیئت مدیره به مجمع» تکیه کند. سوم، این پژوهش برای تعیین میزان شاخص خطر تقلب در گزارش‌های تهیه شده توسط مدیران از روش‌های قبلی که عمدتاً مبتنی بر رگرسیون‌های معمولی هستند فاصله گرفته و نوع خاصی از رگرسیون تحت عنوان رگرسیون‌های LASSO را در پژوهش‌های حسابداری و مالی در کشور، معرفی و از آن استفاده کرده است.

این پژوهش با محدودیت‌های زیادی نیز مواجه بود که در اینجا فقط چهار مورد از مهم‌ترین آنها بیان شده است: اول، این پژوهش به دلیل عدم دسترسی به یک لیست رسمی به منظور تفکیک اولیه شرکت‌ها به دو گروه با خطر بالای تقلب و با خطر پایین تقلب (ماتریس Y) ناچار به استفاده از یک روش غیر مستقیم شد که مبتنی بر حق‌الزحمه حسابرسی است. دوم، عدم دسترسی به فایل Word گزارش‌های هیئت مدیره است که نویسندگان مقاله را ناچار به تبدیل گزارش‌های هیئت مدیره کرده و منجر به ریزش شدید در حجم نمونه شد. سوم، افشا نشدن حق‌الزحمه حسابرسی در یادداشت‌های توضیحی شرکت‌های بورسی بود که دسترسی به آن به دلیل رعایت اصل رازداری از آیین رفتار حرفه‌ای، از طریق مؤسسات حسابرسی نیز ممکن نشد و حجم نمونه را کاهش داد. چهارم، عدم وجود نرم افزارهای نشانه‌گذاری متن (مانند QTagger) برای متون فارسی بود که منجر به پالایش واژه‌های گزارش‌های هیئت مدیره به صورت دستی شد.

### پیشنهادها

در اینجا با توجه به نتایج و محدودیت‌های ذکر شده در بالا پیشنهادهایی ارائه شده است. اول به نهادهای ناظر مانند سازمان بورس اوراق بهادار، سازمان بازرسی، دیوان محاسبات و حتی سازمان بازرسی کل کشور پیشنهاد می‌شود که روش معرفی شده در این پژوهش را به عنوان روشی که قادر به تشخیص شاخص خطر تقلب در گزارش‌های یک شخصیت گزارش‌دهنده است، مدنظر قرار داده، به شیوه مقتضی در رسیدگی‌های خود از آن استفاده نمایند. این روش می‌تواند برای تجزیه و تحلیل سخنرانی‌های مدیران (که به متن تبدیل شده است) نیز بسیار کارآمد باشد. دوم به حساب‌برسان مستقل پیشنهاد می‌شود که در ارزیابی خطر تقلب در رسیدگی‌های خود، متن گزارش‌های هیئت مدیره به مجمع را نیز مدنظر قرار داده و از روشی که در پژوهش حاضر استفاده شد هم برای ارزیابی خطر تقلب استفاده کنند. سوم به سازمان بورس اوراق بهادار پیشنهاد می‌شود که به صورت مستقل، شرکت‌ها را از لحاظ خطر تقلب در گزارش‌ها مورد رسیدگی قرار داده و لیست شرکت‌هایی که تذکر دریافت کرده‌اند را به صورت رسمی منتشر نماید تا پژوهشگران حوزه تقلب در آینده بتوانند به این لیست اتکا کنند. همچنین به سازمان مذکور پیشنهاد می‌شود که شرکت‌های بورسی را ملزم نماید که فایل Word صورت‌های مالی و گزارش‌های هیئت مدیره را نیز بر روی سایت گُدال بارگزاری نمایند. این الزام، نه برای شرکت‌ها و نه برای سازمان بورس اوراق بهادار، هزینه‌ای در پی ندارد، اما بسیاری از تجزیه و تحلیل‌های بعدی را برای پژوهشگران تسهیل می‌نماید.

### یادداشت‌ها

- |                                       |                           |
|---------------------------------------|---------------------------|
| 1. Elliott and Willingham             | 2. Jensen and Meckling    |
| 3. Management Discussion and Analysis | 4. Humpherys              |
| 5. Purda and Skillicorn               | 6. Feldman                |
| 7. Cecchini                           | 8. Textual Analysis       |
| 9. Bag of the Words                   | 10. Loughran and McDonald |
| 11. Larcker and Zakolyukina           | 12. Goel                  |

- |                                                     |                        |
|-----------------------------------------------------|------------------------|
| 13. Glancy and Yadav                                | 14. Goel and Gangolly  |
| 15. Data Mining                                     | 16. Baldwin            |
| 17. Amani and Fadlalla                              | 18. Pujari             |
| 19. Han                                             | 20. Hoberg and Lewis   |
| 21. Tackett                                         | 22. Kim and Vasarhelyi |
| 23. Alden                                           | 24. Kochetovakozloski  |
| 25. Wang                                            | 26. Pai                |
| 27. Liou                                            | 28. Green              |
| 29. Baum                                            | 30. Brooks             |
| 31. Dechow                                          |                        |
| 32. Accounting and Auditing Enforcement Releases    |                        |
| 33. Hribar                                          |                        |
| 34. Unexplained Audit Fee (UAF)                     |                        |
| 35. Codal.ir                                        | 36. Rdis.ir            |
| 37. www.parsipdf.ir                                 | 38. Jackson            |
| 39. Least Absolute Shrinkage and Selection Operator |                        |
| 40. Tibshirani                                      |                        |
| 41. Friedman                                        | 42. Cross Section      |
| 43. K-Fold Cross Validation                         | 44. Hastie             |

## منابع

- اعتمادی، حسین؛ زلفی، حسن. (۱۳۹۲). کاربرد رگرسیون لجستیک در شناسایی گزارشگری مالی متقلبانه. *دانش حسابرسی*، ۱۳(۵۱)، ۱۶۳-۱۴۵.
- افلاطونی، عباس. (۱۳۹۴). *تجزیه و تحلیل آماری با Eviews در تحقیقات حسابداری و مدیریت مالی*. تهران: ترمه.
- بروکس، کریس. (۱۹۷۱). *مقدمه‌ای بر اقتصادسنجی مالی*، ترجمه احمد بدری و عبدالمجید عبدالباقی، تهران: نص، (۱۳۸۹).
- خدادادی، ولی الله؛ ویسی، سجاد؛ چراغی نیا، علی. (۱۳۹۵). عدم قطعیت مدیریت و حق الزحمه حسابرسی. *دانش حسابداری*، ۷(۲۷)، ۱۳۳-۱۰۹.
- رهنمای رودپشتی، فریدون. (۱۳۹۱). داده کاوی و کشف تقلب‌های مالی. *دانش حسابداری و حسابرسی مدیریت*، ۱(۳)، ۱۷-۳۳.
- صفرزاده، محمدحسین. (۱۳۸۹). توانایی نسبت‌های مالی در کشف تقلب در گزارشگری مالی: تحلیل لاجیت. *دانش حسابداری*، ۱(۱)، صص. ۱۶۳-۱۳۷.
- فرج‌زاده دهکردی، حسن؛ آقایی، لیلا. (۱۳۹۴). سیاست تقسیم سود و گزارشگری مالی متقلبانه. *مطالعات تجربی حسابداری مالی*، ۱۳(۴۵)، صص. ۹۷-۱۱۴.

مهرانی، کاوه؛ حصارزاده، رضا. (۱۳۸۷). مروری بر تئوری‌ها و الگوهای کشف تقلب. *دانش و پژوهش حسابداری*، ۱۵، صص. ۶-۱۱.

نمازی، محمد؛ ابراهیمی، فهیمه. (۱۳۹۵). الگوبندی و تعیین اولویت عوامل مؤثر بر قصد گزارش تقلب‌های مالی توسط حسابداران. *مطالعات تجربی حسابداری مالی*، ۱۲(۴۹)، صص. ۱-۲۸.

وحیدی الیزی، ابراهیم؛ حامدیان، حامد. (۱۳۸۸). برداشت حساب‌رسان ایران از کارایی علائم خطر در کشف گزارشگری مالی متقلبانه. *تحقیقات حسابداری*، ۳، صص. ۱۶۲-۱۹۷.

Aflatooni, A. (2015). *Statistical Analysis in Accounting and Financial Management by Eviews* (Vol. 2). Tehran: Termeh [In Persian].

Alden, M.E., Bryan, D.M., Lessley, B.J., Tripathy, A. (2012). Detection of financial statement fraud using evolutionary algorithms. *Journal of Emerging Technologies in Accounting*, 9(1), 71-94.

Amani, F.A., Fadlalla, A.M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, 24(2017), 32-58.

Baldwin, A.A., Brown, C.E., Trinkle, B.S. (2006). Opportunities for artificial intelligence development in the accounting domain: The case for auditing. *Intelligent Systems in Accounting, Finance and Management*, 14(3), 77-86.

Baum, C.F. (2006). *An Introduction to Modern Econometrics Using Stata*. College Station, Texas: Stata Press.

Brooks, C. (1971). *Introductory Econometrics for Finance* (A. Badri & A. Abdolbaghi, Trans. Vol. 1). Tehran: Nass [In Persian].

Cecchini, M., Aytug, H., Koehler, G., Pathak, P. (2010). Making words work: Using financial text as a predictor of financial events. *Decision Support Systems*, 50(1), 164-175.

Dechow, P., Ge, W., Larson, C., Sloan, R. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research*, 28(1), 17-82.

Elliott, R.K., Willingham, J.J. (1980). *Management Fraud: Detection and Deterrence*. New York: Petrocelli Books, 35-46.

Etemadi, H., Zalaghi, H. (2013). Application of logistic regression in identifying fraudulent financial reporting. *Journal of Audit Science*, 13(51), 5-23 [In Persian].

Farajzadeh, H., Aghaei, L. (2015). Dividend policy and fraudulent financial reporting. *Empirical Studies in Financial Accounting*, 12(45), 97-114 [In Persian].

Feldman, R., Govindaraj, S., Livnat, J., Segal, B. (2010). Management's tone change,

- post earnings announcement drift and accruals. *Review of Accounting Studies*, 15(4), 915-953.
- Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1-22.
- Glancy, F., Yadav, S. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems*, 50(3), 595-601.
- Goel, S., Gangolly, J. (2012). Beyond the numbers: mining the annual reports for hidden cues indicative of financial statement fraud. *Intelligent Systems in Accounting, Finance and Management*, 19(2), 75-89.
- Goel, S., Gangolly, J., Faerman, S.R., Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting*, 7(1), 25-46.
- Greene, W.H. (2012). *Econometric Analysis* (Vol. 7): Prentice Hall.
- Han, J., Kamber, M., Pei, J. (2006). *Data Mining: Concepts and Techniques*: Morgan Kaufmann.
- Hastie, T., Tibshirani, R., Friedman, J. (2008). *The Elements of Statistical Learning Data Mining, Inference, and Prediction* (Vol. 2). Stanford, California: Springer.
- Hoberg, G., Lewis, C. (2017). Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance*, 43, 58-85.
- Hribar, P., Kravet, T., Wilson, R. (2014). A new measure of accounting quality. *Review of Accounting Studies*, 19, 506-538.
- Humphreys, S., K. Moffit, M. Burns, J. Burgoon, Felix, W. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50(3), 585-594.
- Jackson, J.M. (2002). Data mining: A conceptual overview. *Communications of the Association for Information Systems*, 8(1), 267-296.
- Jensen, M.C., Meckling, W.H. (1976). Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305-360.
- Khodadadi, V., Veisi, S., Cheraghinia, A. (2016). Management uncertainty and audit fees. *Journal of Accounting Knowledge*, 7(27), 109-133 [In Persian].
- Kim, Y., Vasarhelyi, M.A. (2012). A model to detect potentially fraudulent/ abnormal wires of an insurance company: An unsupervised rule-based approach. *Journal of Emerging Technologies in Accounting*, 9(1), 95-110.
- Kochetova-kozloski, N., Messier Jr., W.F., Eilifsen, A. (2011). Improving auditors' fraud judgments using a frequency response mode. *Contemporary Accounting Research*,

- 28(3), 837-858.
- Larcker, D., Zakolyukina, A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2), 495-540.
- Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. *Accounting Literature*, 29(1), 143-165.
- Liou, F.M. (2008). Fraudulent financial reporting detection and business failure prediction models: A comparison. *Managerial Auditing Journal*, 23(7), 650-662.
- Loughran, T., McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35-65.
- Mehrani, K., Hesarzadeh, R. (2008). A review of theories and models of fraud detection. *Accounting Knowledge and Research*, 15 (Winter), 6-11 [In Persian].
- Namazi, M., Ebrahimi, F. (2016). Modeling and identifying effective factors affecting the intention of reporting financial fraudulent by accountant. *Empirical Studies in Financial Accounting*, 13(49), 1-28 [In Persian].
- Pai, P.F., Hsu, M.F., Wang, M.C. (2011). A support vector machine-based model for detecting top management fraud. *Knowledge-Based Systems*, 24(2), 314-321.
- Pujari, A.K. (2001). *Data Mining Techniques*: Universities press.
- Purda, L., Skillicorn, D. (2015). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research*, 32(3), 1193-1223.
- Rahnamay Roodposhti, F. (2012). Data mining & financial fraud. *Journal of Management Accounting and Auditing Knowledge*, 1(3), 17-34 [In Persian].
- Safarzadeh, M. (2010). The ability of financial ratios in detecting fraudulent financial reporting: Logit analysis. *Journal of Accounting Knowledge*, 1(1), 137-163 [In Persian].
- Tackett, J.A. (2013). Association rules for fraud detection. *Journal of Corporate Accounting and Finance*, 24(4), 15-22.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1), 267-288.
- Vahidi, E., Hamedian, H. (2009). Iranian auditors' perceptions of the effectiveness of risk signs in detecting fraudulent financial reporting. *Accounting Researches*, Fall (3), 162-197 [In Persian].
- Wang, Z., Chen, M.H., Chin, C.L., Zheng, Q. (2017). Managerial ability, political connections, and fraudulent financial reporting in China. *Journal of Accounting and Public Policy*, 36(2), 141-162.