

The Conformity of Fraud Incentives in Managers with Cumulative Prospect Theory Pattern through Text Analysis

Alireza Rahravi Dastjerdi¹, Daryoosh Foroghi², Gholam Hosein Kiani³

Received : 2017/09/16

Approved: 2018/01/09

Abstract

Fraudulent reporting means intentional providence of false and misleading reports. Paragraph 7 of Iranian Accounting Standard No. 1 considers the board of directors as responsible for financial statements providence, however, managers may provide the firm's reports including the report of the board of directors, with fraudulent incentives. One of the theories explaining the managerial decision-making methods based on their personality traits (risk taking and risk aversion) is the Cumulative Prospect Theory (CPT). Based on this theory, fraud incentives will be formed in managers' mind when they feel their performance will go away from the reference point they have already created in their minds. This study determines an index with more than 90% accuracy for assessing and detecting the risk of fraud in the board's report, focusing the report text and using two data mining methods including decision tree and machine learning methods. Then, the index is used to examine whether the manager incentives for using high fraud risk reporting methods follow the pattern presented in cumulative prospect theory. The results indicate that managers' fraud incentives in Iran are not consistent with CPT.

Keyword: Cumulative Prospect Theory, Fraud Risk, Data Mining, Decision Tree, Machine Learning.

Jel clacification: C33,C35,D81,G32,M41, M42

DOI: 10.22051/jera.2018.17252.1791

¹ Phd. student of Accounting, Esfahan University, Esfahan, Iran, (Alireza18467@gmail.com)

² Associate professor of Accounting, Esfahan University, Esfahan, Iran, corresponding author, (foroghi@ase.ui.ac.ir)

³ Assistant professor of Economic, Esfahan University, Esfahan, Iran, (gh.kiani@ase.ui.ac.ir)

انطباق انگیزه‌های تقلب در مدیران با الگوی نظریه

چشم‌انداز تجمعی از طریق تجزیه و تحلیل متن^۱

علیرضا رهروی دستجردی*، داریوش فروغی**، غلامحسین کیانی***

تاریخ دریافت: ۹۶/۰۶/۲۵

تاریخ پذیرش: ۹۶/۱۰/۱۹

چکیده

گزارشگری متقلبانانه، به معنای اقدام عمدی در جهت تهیه گزارش‌های نادرست و گمراه‌کننده است. بند ۱۷ استاندارد حسابداری شماره ۱ ایران هیئت مدیره را مسئول تهیه صورت‌های مالی دانسته است و مدیران نیز می‌توانند با انگیزه‌های متقلبانانه اقدام به تهیه گزارش‌های شرکت، از جمله گزارش هیئت مدیره به مجمع نمایند. یکی از نظریه‌هایی که به توضیح شیوه‌ی تصمیم‌گیری مدیران بر اساس ویژگی‌های شخصیتی آن‌ها (ریسک‌پذیری و ریسک‌گریزی) پرداخته است، نظریه چشم‌انداز تجمعی (CPT) است. بر اساس این نظریه مدیران، عملکردی را بعنوان نقطه‌ی مرجع در ذهن خود ایجاد کرده و چنانچه احساس کنند که عملکردشان از نقطه‌ی مرجع فاصله خواهد گرفت، انگیزه‌های تقلب در ذهن آن‌ها شکل می‌گیرد. در این پژوهش ابتدا با تمرکز بر تجزیه و تحلیل متن گزارش‌های هیئت مدیره به مجمع و با استفاده از دو روش داده‌کاوی شامل یک روش مبتنی بر درخت تصمیم و یک روش مبتنی بر آموزش ماشین، شاخصی برای ارزیابی و تشخیص خطر تقلب در گزارش‌های هیئت مدیره تعیین شده است که دقتی بیش از ۹۰٪ دارد. سپس با استفاده از شاخص مذکور، بررسی شده است که آیا انگیزه‌های مدیران برای استفاده از روش‌هایی با خطر بالای تقلب در تهیه گزارش‌ها، از الگویی که توسط نظریه چشم‌انداز تجمعی ارائه شده است پیروی می‌کند یا خیر؟ نتایج نشان‌دهنده‌ی عدم انطباق انگیزه‌های تقلب مدیران در ایران، با نظریه مذکور است.

واژه‌های کلیدی: نظریه چشم‌انداز تجمعی، خطر تقلب، داده‌کاوی، درخت تصمیم، آموزش ماشین.

طبقه‌بندی موضوعی: C33، C35، D81، G32، M41، M42

۱. مقاله مستخرج از رساله دکتری است.

DOI: 10.22051/jera.2018.17252.1791

* دانشجوی دکتری حسابداری دانشکده علوم اداری و اقتصاد دانشگاه اصفهان، اصفهان، ایران. (Alireza18467@gmail.com)

** دانشیار گروه حسابداری دانشکده علوم اداری و اقتصاد دانشگاه اصفهان، اصفهان، ایران. (نویسنده مسئول).

(foroghi@ase.ui.ac.ir)

*** استادیار گروه اقتصاد دانشکده علوم اداری و اقتصاد دانشگاه اصفهان، اصفهان، ایران. (gh.kiani@ase.ui.ac.ir)

مقدمه

استاندارد حسابرسی ۲۴۰ ایران با عنوان مسئولیت حسابرس در ارتباط با تقلب و اشتباه، بیان می‌کند که «تقلب» عبارتست از هرگونه اقدام عمدی یا فریبکارانه یک یا چند نفر از مدیران، کارکنان یا اشخاص ثالث، برای برخورداری از یک مزیت ناروا یا غیرقانونی (کمیتة تدوین استانداردهای حسابرسی، استاندارد ۲۴۰، ۱۳۸۴: بند ۳ و ۴). تقلب توسط مدیریت واحد تجاری به دلیل گمراه کردن ذینفعان از طریق صورت‌های مالی متقلبانه (گمراه‌کننده)، بر قیمت سهام اثر معکوسی دارد (الیت و ویلینگهام، ۱۹۸۰). بر همین اساس می‌توان گفت استفاده از روش‌هایی با خطر بالای تقلب در تهیه گزارش‌ها توسط مدیریت، شیوه‌ای پرخطر برای بهبود ظاهری وضعیت و عملکرد شرکت است و لذا انتظار می‌رود که انگیزه‌های استفاده از روش‌های مذکور با تمایل مدیر به ریسک‌پذیری در ارتباط باشد (فانگ، ۲۰۱۵).

تغییر در رفتارهای مدیران که ناشی از میزان ریسک‌پذیری و ریسک‌گریزی آن‌هاست توسط نظریه چشم‌انداز تجمعی^۱ به خوبی توضیح داده شده است. در این نظریه (CPT) که توسط ورسکی و کانمن (۱۹۹۲) ارائه شد؛ اعتقاد بر این است که ریسک‌پذیری و ریسک‌گریزی افراد، یک امر مطلق و همیشگی نیست. بلکه یک امر نسبی بوده و با توجه به یک نقطه‌ی مرجع، تعیین می‌شود. در این نظریه، افراد در مواجهه با سودهایی با احتمال دستیابی پایین یا زیان‌هایی با احتمال دستیابی بالا، ریسک‌پذیر بوده و در مقابل، در مواجهه با سودهایی با احتمال دستیابی بالا یا زیان‌هایی با احتمال دستیابی پایین، ریسک‌گریز هستند. (فانگ، ۲۰۱۵: ۵۵).

در ایران نیز تشخیص صادقانه یا متقلبانه بودن گزارش‌های تهیه شده توسط مدیران برای نهادهای ناظر (مانند بورس اوراق بهادار) حائز اهمیت است. زیرا انگیزه‌های تقلب در مدیران می‌تواند در این گزارش‌ها نمود پیدا کند. از مهم‌ترین گزارش‌های تهیه شده توسط مدیران که کمتر حالت استاندارد شده داشته و اختیار مدیر در انتخاب واژه‌ها و حجم آن، بیشتر است گزارش هیئت مدیره به مجمع عمومی صاحبان سهام است. پژوهش حاضر اعتقاد دارد که با بررسی متن این گزارش‌ها، می‌توان به وجود یا عدم وجود انگیزه‌های تقلب در مدیران پی برد.

پیش از این نیز پژوهشگرانی با بررسی متن بخش تجزیه و تحلیل‌های مدیریت (MD&A)^۲ از گزارش‌های مدیریتی به ارزیابی خطر تقلب در شرکت‌ها پرداخته‌اند (به عنوان مثال: فلدمن و همکاران، ۲۰۱۰؛ سچینی و همکاران، ۲۰۱۰ و هامف‌ریس و همکاران، ۲۰۱۱). در ایران،

گزارش هیئت مدیره به مجمع عمومی صاحبان سهام، شباهت زیادی به بخش MD&A در کشورهای دیگر دارد. زیرا هر دوی این گزارش‌ها به سرمایه‌گذاران این امکان را می‌دهند که به عملکرد و فرصت‌های پیش روی شرکت از دید مدیران بنگرند.

با توجه به بررسی‌های صورت گرفته، در کشور ایران و در رشته حسابداری تا کنون پژوهشی که با استفاده از تجزیه و تحلیل متن گزارش‌ها به ارزیابی و تشخیص خطر تقلب در شرکت‌ها پرداخته باشد انجام نشده است. در پژوهش حاضر ابتدا با استفاده از دو روش مبتنی بر داده کاوی (شامل درخت تصمیم و آموزش ماشین)، متن گزارش‌های هیئت مدیره به مجمع، تجزیه و تحلیل شده است. در نتیجه این تجزیه و تحلیل، واژه‌هایی از متن گزارش‌های مذکور که بیشترین قدرت را در تبیین خطر تقلب در شرکت‌ها دارند استخراج شده و با استفاده از آن‌ها مدلی برای تشخیص خطر تقلب در شرکت‌ها ارائه شده است. سپس با استفاده از این مدل، انطباق انگیزه‌های تقلب در مدیران با الگوی نظریه چشم انداز تجمعی (CPT) بررسی شده است. لذا سؤال اصلی پژوهش حاضر این است که آیا انگیزه‌های تقلب مدیران در شرکت‌های ایرانی، از الگوی ارائه شده توسط نظریه چشم انداز تجمعی (CPT) پیروی می‌کند؟

ادبیات و مبانی نظری

برای توضیح شیوه‌ی تصمیم‌گیری^۳ اشخاص در شرایطی که با ریسک مواجه هستند، نظریه‌های متعددی ارائه شده است. نظریه مطلوبیت مورد انتظار (EUT)^۴ (وان‌نیومن و مورگنسترن، ۱۹۴۴) و نظریه مطلوبیت ذهنی مورد انتظار (SEU)^۵ (ادواردز، ۱۹۵۵) دو نمونه از مهمترین این نظریه‌ها هستند. پس از ارائه این نظریه‌ها پژوهشگران زیادی به روش‌های نوینی سعی کردند درستی آن‌ها را آزمون کنند. یکی از پذیرفته شده‌ترین نظریه‌های نوین در این زمینه نظریه چشم انداز تجمعی (CPT) است که توسط ورسکی و کانمن (۱۹۹۲) ارائه شد. CPT نگرش افراد راجع به ریسک را به صورت الگوی چهاربعدی زیر ارائه می‌دهد: افراد در مواجهه با سودهایی با احتمال دستیابی پایین یا زیان‌هایی با احتمال دستیابی بالا، ریسک‌پذیر بوده و در مقابل، در مواجهه با سودهایی با احتمال دستیابی بالا یا زیان‌هایی با احتمال دستیابی پایین، ریسک‌گریز هستند. در نظریه CPT سودها و زیان‌ها نسبت به یک نقطه مرجع^۶ سنجیده می‌شوند (نظریه چشم‌انداز تجمعی، خود زاییده نظریه چشم‌انداز^۷ است). ایرادی که ورسکی و

کانمن (۱۹۹۲) به نظریه مطلوبیت گرفتند این بود که تابع مطلوبیت نهایی همواره نزولی و مقعر می‌باشد و این بدان معنی است که سرمایه‌گذاران همواره ریسک‌گریز هستند. در حالی که از نظر ورسکی و کانمن سرمایه‌گذاران تا نقطه خاصی ریسک‌پذیر بوده و از آن نقطه به بعد ریسک‌گریز خواهند بود. لذا تابع ارزش در این دو حالت به ترتیب محدب و مقعر خواهد بود. آن‌ها این نقطه عطف را اصطلاحاً «نقطه مرجع» نامیدند. نقطه مرجع لزوماً معادل صفر ریال نیست، اما از دیدگاه سرمایه‌گذاران، این نقطه معادل مبلغی است که هر سودی کمتر از آن به معنای زیان واقعی و هر سودی بیشتر از آن به معنای سود واقعی خواهد بود.

پژوهش حاضر با الهام گرفتن از نظریه CPT، احتمال می‌دهد که انگیزه‌های مدیران برای استفاده از روش‌هایی با خطر بالای تقلب در تهیه گزارش‌ها به عملکرد رقبای آن‌ها (به عنوان یک نقطه مرجع) مرتبط باشد. اگر هدف این باشد که با بینش نظریه چشم‌انداز تجمعی راجع به موضوع خطر تقلب در گزارشگری شرکت‌ها صحبت کرد می‌توان الگوی چهاربعدی CPT را به این صورت بیان کرد که مدیران زمانی که با احتمال بالایی (پایینی) در خصوص عملکرد ضعیف‌تر (قوی‌تر) نسبت به رقبای خود مواجه باشند بیشتر احتمال دارد که ریسک‌پذیرانه عمل نمایند (یعنی به میزان بیشتری از روش‌هایی با خطر بالای تقلب در تهیه گزارش‌ها استفاده کنند). همچنین زمانی که با احتمال بالایی (پایینی) در خصوص عملکرد قوی‌تر (ضعیف‌تر) نسبت به رقبای خود مواجه باشند کمتر احتمال دارد که ریسک‌پذیرانه عمل نمایند (یعنی به میزان کمتری از روش‌هایی با خطر بالای تقلب در تهیه گزارش‌ها استفاده کنند). پس در پژوهش حاضر، استفاده از روش‌هایی با خطر بالای تقلب در گزارشگری توسط مدیران، شاخصی برای ریسک‌پذیرانه عمل کردن آن‌ها است (با استناد به قوانین و آیین‌نامه‌هایی مانند «فصل ششم از قانون بازار اوراق بهادار جمهوری اسلامی ایران» و «دستورالعمل رسیدگی به تخلفات اشخاص موضوع ماده ۳۵ قانون بازار اوراق بهادار» و مواردی از این قبیل).

پژوهش‌های زیادی به بررسی و ارائه راهکارهایی برای تشخیص تقلب در گزارشگری پرداخته‌اند که در ادامه به برخی از آن‌ها اشاره شده است. در میان این روش‌ها «داده‌کاوی»^۸ در حال حاضر یکی از مهم‌ترین پارادایم‌های موجود در حوزه‌ی هوش تجاری و ابزارهای تصمیم‌گیری است. مرور ادبیات نشان می‌دهد که استفاده از روش‌های مبتنی بر داده‌کاوی برای تشخیص تقلب، امری است که بین پژوهشگران و متخصصان حسابداری (در خارج از

کشور) بسیار رایج است (بعنوان مثال، تکت (۲۰۱۳)؛ کیم و واسارلی (۲۰۱۲)؛ آلدن و همکاران (۲۰۱۲)؛ کاجتواکروسکی و همکاران (۲۰۱۱)؛ دیرنسی و گری (۲۰۱۰)؛ لیو (۲۰۰۸) و لین و همکاران (۲۰۰۳)).

پژوهش حاضر از گزارش‌های هیئت مدیره به مجمع شرکت‌ها در سال‌های گذشته استفاده کرده و با تجزیه و تحلیل متن این گزارش‌ها، به شناسایی واژه‌هایی پرداخته است که بیشترین قدرت را در تبیین خطر تقلب در شرکت‌ها دارند. سپس بر اساس این واژه‌ها به تفکیک شرکت‌های با خطر بالای تقلب از سایر شرکت‌ها پرداخته است. بر همین اساس می‌توان پژوهش حاضر را در بین پژوهش‌هایی که از داده‌کاوی استفاده کرده‌اند، در دسته‌ی «گذشته‌نگر^۹-توصیفی^{۱۰}» طبقه‌بندی نمود. این نوع از داده‌کاوی به استفاده‌کنندگان، این امکان را می‌دهد که الگوها و روندها را در داده‌ها مشاهده کرده و مشکلات و فرصت‌های حل کردن آن‌ها را کشف کنند (امانی و فادلالا، ۲۰۱۷: ۵۰-۵۱).

پیشینه‌ی پژوهش

در حوزه‌ی پژوهش‌های داخلی هیچ پژوهشی یافت نشد که از روش‌های مبتنی بر تجزیه و تحلیل متن، به ارزیابی و تشخیص خطر تقلب در شرکت‌ها پرداخته یا نقش نظریه چشم‌انداز تجمعی را در توضیح انگیزه‌های تقلب مدیران بررسی کرده باشد. با این وجود، پژوهش‌هایی با محوریت نظریه چشم‌انداز، ریسک‌گریزی در افراد، تقلب یا داده‌کاوی به صورت جداگانه انجام شده است (بعنوان مثال: نمازی و ابراهیمی (۱۳۹۵)، فرج زاده دهکردی و آقایی (۱۳۹۴)، اعتمادی و زلفی (۱۳۹۲)، مصلح شیرازی و همکاران (۱۳۹۲)، پورحیدری و بذرافشان (۱۳۹۱)، رهنمای رودپشتی (۱۳۹۱)، وحیدی و حامدیان (۱۳۸۸)، مهرانی و حصارزاده (۱۳۸۷)، مظلومی و همکاران (۱۳۸۶)) که در اینجا به منظور جلوگیری از اطاله‌ی کلام، از توضیح جزئیات آن‌ها خودداری شده است.

در حوزه‌ی پژوهش‌های خارجی، مصداق استفاده از نظریه چشم‌انداز تجمعی در توضیح تقلب، بیشتر به چشم می‌خورد. همچنین مجموعه وسیعی از پژوهش‌های خارجی را می‌توان یافت که برای تشخیص تقلب مدیریت، از روش‌های مبتنی بر داده‌کاوی استفاده کرده‌اند. در اینجا به تعدادی از جدیدترین و مربوط‌ترین این پژوهش‌ها به موضوع پژوهش حاضر اشاره

شده است. تروگسا و آرونل (۲۰۱۷) از منطق فازی استفاده کردند تا درستی این باور که در بخش‌های دولتی، این ریسک‌گریزی مدیران است که باعث کاهش اثربخشی آنان می‌شود را مورد بررسی بیشتری قرار دهند. آن‌ها دریافتند که هم مدیران با ریسک‌گریزی بالا و هم مدیران با ریسک‌گریزی پایین، قادر بوده‌اند که با ابتکارات خود منافع سرشاری را عاید سازمان نمایند. این فقط استراتژی‌های آن‌هاست که با هم تفاوت دارد.

فانگ (۲۰۱۵) به بررسی این موضوع پرداخت که آیا انگیزه‌های گزارشگری متقلبان در مدیران از الگوی نظریه چشم‌انداز تجمعی پیروی می‌کند یا خیر؟ نتایج پژوهش او نشان داد که احتمال روی آوردن مدیران به گزارشگری متقلبان با احتمال دستیابی به سودی کمتر (بیشتر) از عملکرد مرجع (میانگین سود رقبا) به صورت مثبت (منفی) و معناداری در ارتباط است. پوردا و اسکیلیکورن (۲۰۱۵) به منظور کشف تقلب در گزارشگری، تلاش کردند تا متن بخش MD&A از گزارش‌های سالیانه شرکت‌ها را تجزیه و تحلیل کرده و با استفاده از داده‌های کاوی، مدلی ارائه دادند که دقتی بیش از ۸۲ درصد داشت. آن‌ها قدرت روش مذکور را با هشت روش دیگر که در این حوزه ارائه شده بود مقایسه کرده و دریافتند که روش جدید به نسبت، بهتر عمل می‌کند.

ژول و گانگلی (۲۰۱۲) نشان دادند که متن گزارش‌های مدیران شامل لحن نوشتن و شیوهی ارائه‌ی آن حاوی نشانه‌های ارزشمندی است که می‌تواند وجود تقلب را نشان دهد. آن‌ها با استفاده از نسبت کای دو نشان دادند که ویژگی‌های زبانی بین گزارش‌های متقلبان و گزارش‌های غیرمتقلبان به صورت معناداری متفاوت است. هامفریس و همکاران (۲۰۱۱) سعی کردند تا زبان و لحن گزارش‌های متقلبان را بررسی نمایند. آن‌ها ۲۴ متغیر را که نشان دهنده لحن گزارش‌ها بودند بررسی کرده و به این نتیجه رسیدند که گزارش‌های متقلبان نسبت به گزارش‌های غیرمتقلبان به نسبت بیشتری از لحن فعال (لفاظی) استفاده می‌کنند، تعداد واژگان بیشتر اما تنوع واژگانی کمتری داشته و پیچیده‌تر هستند. هارباو و همکاران (۲۰۱۰) دریافتند که در مجموع، افراد در مواجهه با سودهای با احتمال کم و زیان‌های با احتمال زیاد، ریسک‌پذیرانه و در مواجهه با سودهای با احتمال زیاد و زیان‌های با احتمال کم، ریسک‌گریزانه عمل می‌کنند. مشاهده این نتایج نشان داد که افزایش در میزان شناختی که افراد از پدیده‌ها دارند، می‌تواند منجر به بروز نابهنجاری‌های زیادی در رفتار مورد انتظار آن‌ها شود.

پژوهش‌های بسیار دیگری نیز در این حوزه یافت می‌شوند که در اینجا از ذکر جزئیات آن‌ها خودداری شده است (بعنوان مثال می‌توانید به ژول و همکاران (۲۰۱۰)، سیچینی و همکاران (۲۰۱۰)، فرناندس و گودس (۲۰۱۰)، پینلو (۲۰۰۸)، برنز و کدیا (۲۰۰۶) و اوکانر و همکاران (۲۰۰۶) مراجعه نمایید).

فرضیه‌های پژوهش

در راستای پاسخ به سوال پژوهش، دو فرضیه زیر تدوین شده و مورد آزمون قرار گرفته‌اند:
فرضیه اول) انتظار مدیریت از مثبت بودن انحراف از عملکرد مرجع در سال بعد، خطر تقلب مدیر را کاهش می‌دهد.

فرضیه دوم) انتظار مدیریت از منفی بودن انحراف از عملکرد مرجع در سال بعد، خطر تقلب مدیر را افزایش می‌دهد.

جامعه آماری و روش نمونه‌گیری

با توجه به نحوه‌ی محاسبه‌ی متغیر وابسته در پژوهش حاضر (بالا بودن خطر تقلب در شرکت)، شرکت‌ها برای قرار گرفتن در نمونه‌ی این پژوهش باید دو شرط را همزمان احراز کنند. این دو شرط عبارتند از اینکه هم گزارش‌های هیئت‌مدیره به مجمع آن‌ها موجود و قابل تبدیل از PDF به فایل Word بوده و هم حق الزحمه حسابرس را در یادداشت‌های توضیحی افشا کرده باشند.

برهمن اساس، گزارش‌های هیئت‌مدیره به مجمع عمومی تمامی شرکت‌های بورسی طی دوره زمانی پژوهش بررسی شده و نهایتاً تعداد ۵۶ شرکت یافت شد که گزارش‌های آن‌ها حداقل برای چهار سال، قابل تبدیل به Word بوده و پس از تبدیل نیز، خوانا و قابل استفاده باشند. قائل شدن معیار «حداقل چهار سال» کاملاً اختیاری است. اما افزایش این معیار باعث ریزش شدید شرکت‌های عضو نمونه و کاهش این معیار نیز باعث کاهش قابلیت اتکای روش‌های اجرایی می‌شد. همچنین با بررسی یادداشت‌های توضیحی کلیه شرکت‌های بورسی مشخص شد که ۸۵ شرکت وجود دارند که حق الزحمه حسابرسی را در یک یا تعدادی از سال‌های دوره زمانی پژوهش افشا کرده‌اند.

بنابراین نمونه‌ی این پژوهش از بین کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران انتخاب شده است که بین دو گروه فوق (۵۶ شرکتی و ۸۵ شرکتی) مشترک باشند. این اشتراک، تنها شامل ۳۰ شرکت بود. هیچ محدودیت دیگری (شامل منتهی شدن سال مالی شرکت به ۲۹ اسفند، تولیدی بودن شرکت، عدم توقف نماد شرکت و شرط‌های مرسوم دیگر) بر نمونه اعمال نشده است تا از کاهش حجم نمونه بیش از این جلوگیری شود. اگرچه کوچک شدن حجم نمونه، بزرگترین محدودیت پژوهش حاضر است اما در رویکرد اقتصادسنجی پیشرفته اعتقاد بر این است که زمانی که حجم کل مشاهدات از ۳۰ مورد بیشتر می‌شود (مانند آنچه در داده‌های ترکیبی اتفاق می‌افتد) بنابر قضیه حد مرکزی، توزیع ضرایب به سمت نرمال میل کرده و می‌توان ضرایب را سازگار و بدون تورش در نظر گرفت (گرین، ۲۰۱۲: ۲۴؛ باوم، ۲۰۰۶: ۷۳؛ بروکس، ۱۹۷۱: ۱۶۷ و افلاطونی، ۱۳۹۴: ۱۲۱).

دوره زمانی پژوهش

دوره زمانی پژوهش حاضر سال‌های ۱۳۸۶ تا ۱۳۹۴ است. علت انتخاب سال ۱۳۸۶ بعنوان سال ابتدایی نمونه این است که شرکت‌ها از سال ۱۳۸۶ ملزم به ارائه‌ی گزارش هیئت‌مدیره به مجمع عمومی صاحبان سهام شده‌اند. لذا برای اولین بار از سال ۱۳۸۷ به تعداد بسیار اندک و از سال ۱۳۸۸ می‌توان این گزارش‌ها را به تعداد کافی در شرکت‌های بورسی یافت. همچنین در برخی تحلیل‌ها و محاسبه‌ی بعضی از متغیرها به داده‌های یک سال قبل و دو سال قبل هم نیاز بوده است که به همین دلیل برای تخمین مدل‌ها از داده‌های سال‌های ۱۳۸۸ تا ۱۳۹۴ استفاده شده است. همچنین در زمان اجرای روش‌های آماری این پژوهش، گزارش‌ها و اطلاعات مالی شرکت‌ها حداکثر تا سال ۱۳۹۴ به تعداد قابل قبول، در دسترس بوده است.

مدل پژوهش برای آزمون فرضیه‌ها

برای آزمون فرضیه‌های پژوهش با استناد به پژوهش فانگ (۲۰۱۵) رابطه (۱) به روش لاجیت اجرا شده است:

$$R_{i,t} = \alpha + \beta_1 P(I_{i,t+1}=1) + \beta_2 P(I_{i,t+1}=3) + NOA_{i,t} + LQD_{i,t} + EXF_{i,t} + e_i \quad (1) \text{ رابطه}$$

که در آن:

$R_{i,t}$ = یک متغیر دو وجهی است. مقدار این متغیر برابر یک برای شرکت‌هایی که خطر تقلب در آن‌ها بالا تشخیص داده شده (طبق روش‌هایی که در ادامه توضیح داده خواهد شد) و صفر برای سایر شرکت‌هاست، $P(I_{i,t+1}=1)$ = احتمال اینکه انحراف از عملکرد مرجع در سال آینده منفی باشد، $P(I_{i,t+1}=3)$ = احتمال اینکه انحراف از عملکرد مرجع در سال آینده مثبت باشد، $NOA_{i,t}$ = دارایی‌های غیرعملیاتی است که از جمع بدهی‌ها و حقوق صاحبان سهام منهای وجوه نقد و سرمایه‌گذاری‌های کوتاه‌مدت تقسیم بر جمع دارایی‌های ابتدای دوره بدست می‌آید، $LQD_{i,t}$ = شاخص نقدینگی است که از تقسیم وجوه نقد و سرمایه‌گذاری‌های کوتاه‌مدت بر جمع بدهی‌ها بدست می‌آید، $EXF_{i,t}$ = تأمین مالی خارجی است که از جمع حساب‌های پرداختی بلندمدت و تسهیلات مالی بلندمدت تقسیم بر جمع دارایی‌های ابتدای دوره بدست می‌آید. در ادامه در خصوص متغیرهای PI و R در رابطه‌ی (۱) به تفصیل، توضیحاتی ارائه شده است.

متغیرهای توضیحی (احتمالات مربوط به انحراف از عملکرد مرجع در سال بعد)

نظریه CPT قائل به وجود رابطه‌ای بین ویژگی‌های ریسک‌پذیری از یک طرف و انحراف از سود یا زیان مرجع از سوی دیگر است. منظور از سود و زیان مرجع در اینجا، میانگین عملکرد صنعت، بدون احتساب عملکرد خود شرکت است (فانگ، ۲۰۱۵). بنابراین می‌توان یک متغیر سه وجهی به نام $I_{i,t+1}$ ساخت که شاخصی برای حالت‌های مختلف انحراف از عملکرد مرجع در سال آینده (منفی، صفر یا مثبت) باشد. این متغیر برای هر یک از حالت‌های مذکور، به ترتیب مقادیر ۱، ۲ و ۳ را اختیار خواهد کرد. فرض کنید انحراف از عملکرد مرجع در سال آینده را با y_{t+1} نشان دهیم. یعنی:

$$y_{t+1} = \text{عملکرد مرجع در سال بعد} - \text{عملکرد شرکت در سال بعد} \quad (۲)$$

که در آن:

عملکرد: نسبت ROE شرکت‌ها که از تقسیم سود خالص در هر سال بر ارزش بازار حقوق صاحبان سهام شرکت در آن سال به دست می‌آید.

در نتیجه خواهیم داشت:

اگر $y_{t+1} < 0$ آنگاه $I_{i,t+1} = 1$ پس عملکرد مرجع در سال آینده منفی باشد.	$P(I_{i,t+1}=1)$: احتمال اینکه انحراف از
اگر $y_{t+1} = 0$ آنگاه $I_{i,t+1} = 2$ پس عملکرد مرجع در سال آینده صفر باشد.	$P(I_{i,t+1}=2)$: احتمال اینکه انحراف از
اگر $y_{t+1} > 0$ آنگاه $I_{i,t+1} = 3$ پس عملکرد مرجع در سال آینده مثبت باشد.	$P(I_{i,t+1}=3)$: احتمال اینکه انحراف از

حالت دوم ($y_{t+1} = 0$) به ندرت اتفاق می‌افتد. زیرا به ندرت مشاهده می‌شود که عملکرد شرکت در یک سال، دقیقاً منطبق بر عملکرد مرجع باشد. لذا در این پژوهش به پیروی از فانگ (۲۰۱۵) صرفاً حالت‌های اول ($y_{t+1} < 0$) و سوم ($y_{t+1} > 0$) بررسی شده است. بر این اساس می‌توان فرضیه‌های ۱ و ۲ را با استفاده از رابطه (۱) آزمون کرد. رابطه (۱) با استفاده از مدل لاجیت اجرا می‌شود. برای هر شرکت در هر سال، دو احتمال شامل $P(I_{i,t+1}=1)$ و دیگری $P(I_{i,t+1}=3)$ محاسبه می‌شود. پژوهش حاضر برای محاسبه این احتمال‌ها از روش لاجیت رتبه‌ای^{۱۱} استفاده کرده که در ادامه توضیح داده شده است.

پس از اجرای رابطه (۱) به منظور تصمیم‌گیری در خصوص رد یا عدم رد فرضیه‌های ۱ و ۲ به صورت زیر عمل خواهد شد: اگر احتمال مثبت بودن انحراف از عملکرد مرجع در سال آینده بالا (پایین) باشد، آنگاه $P(I_{i,t+1}=3)$ افزایش (کاهش) می‌یابد. بنابراین مدیر با احتمال کمتری (بیشتری) به استفاده از روش‌هایی با خطر بالای تقلب در تهیه گزارش‌ها روی خواهد آورد. لذا $\beta_2 < 0$ و فرضیه ۱ رد نمی‌شود. اگر احتمال منفی بودن انحراف از عملکرد مرجع در سال آینده بالا (پایین) باشد، آنگاه $P(I_{i,t+1}=1)$ افزایش (کاهش) می‌یابد. بنابراین مدیر با احتمال بیشتری (کمتری) به استفاده از روش‌هایی با خطر بالای تقلب در تهیه گزارش‌ها روی خواهد آورد. لذا $\beta_1 > 0$ و فرضیه ۲ رد نمی‌شود.

روش رگرسیون‌های رتبه‌ای زمانی استفاده می‌شود که متغیر وابسته، دارای نوعی ترتیب باشد (مثلاً مدرک تحصیلی یا سن افراد). در این پژوهش نیز متغیر $I_{i,t+1}$ دارای نوعی ترتیب است. لذا به منظور محاسبه $P(I_{i,t+1}=1)$ و $P(I_{i,t+1}=3)$ به پیروی از فانگ (۲۰۱۵) رابطه (۳) به روش لاجیت رتبه‌ای اجرا شده است:

$$I_{i,t+1} = \beta_1 y_{i,t} + \beta_2 \text{Growth}_{i,t} + e_{i,t+1} \quad \text{رابطه (۳)}$$

در این رابطه، $y_{i,t}$ = عملکرد شرکت در سال جاری و $\text{Growth}_{i,t}$ = فرصت‌های رشد شرکت در سال جاری است که از نسبت ارزش بازار به ارزش دفتری سهام شرکت به دست می‌آید. پس از اجرای رابطه (۳) به روش لاجیت رتبه‌ای، روش مذکور قادر خواهد بود که P ($I_{i,t+1}=1$) و همچنین P ($I_{i,t+1}=3$) را برای هر شرکت سال به صورت جداگانه محاسبه نماید.

متغیر وابسته (بالا بودن خطر تقلب در شرکت)

متغیر وابسته در رابطه (۱) «بالا بودن خطر تقلب» است. این یک متغیر دو وجهی است و مقدار آن برای شرکت‌هایی که خطر تقلب در آن‌ها بالا باشد معادل یک و برای شرکت‌های دیگر معادل صفر است. پژوهش حاضر برای تشخیص خطر تقلب (محاسبه متغیر وابسته در رابطه (۱))، به جای استفاده از اعداد و ارقام مالی، از تجزیه و تحلیل «متن گزارش‌های هیئت‌مدیره به مجمع» استفاده کرده است. برای این تجزیه و تحلیل، با استفاده از دو روش شامل یک روش مبتنی بر درخت تصمیم به نام روش «جنگل‌های تصادفی» و یک روش مبتنی بر آموزش ماشین به نام روش «ماشین بردار پشتیبان» مدلی برازش شده است که قادر به ارزیابی خطر تقلب و تشخیص بالا بودن آن است. این دو روش در ادامه توضیح داده شده‌اند.

پیش از توضیح دو روش مذکور، یادآوری می‌شود که اجرای هر دو روش نیازمند دو ماتریس تحت عنوان ماتریس واژه‌ها (X) و ماتریس متغیر پاسخ (Y) است. ماتریس واژه‌ها (X) یک ماتریس $m \times n$ است که در آن m معادل تعداد شرکت‌سال‌ها و n معادل تعداد واژه‌هایی است که بعنوان متغیرهای توضیح دهنده‌ی خطر تقلب استفاده شده و پژوهش حاضر در صدد آن است که از بین آن‌ها تعدادی از بهترین واژه‌ها که بیشترین قدرت را در پیش‌بینی خطر تقلب دارند، مشخص کند. هر عضو این ماتریس که محل تلاقی یک سطر با یک ستون است، نشان دهنده «درصد» حاصل از تقسیم تکرارهای یک واژه خاص (ستون) به کل واژه‌های یک گزارش خاص (سطر) است. علت تقسیم شدن فراوانی هر واژه از هر گزارش بر جمع تعداد کل واژه‌های آن گزارش، نوعی مقیاس زدایی است؛ به نحوی که اثر طول گزارش (که برای شرکت‌های مختلف یکسان نیست) خنثی شود (پوردا و اسکلیکورن، ۲۰۱۵).

همچنین متغیر پاسخ (Y) به صورت یک ماتریس $m \times 1$ است که اعضای آن، شاخصی از خطر تقلب در شرکت‌هاست. یعنی اعضای این ماتریس، معادل یک برای شرکت‌هایی که «در تقسیم‌بندی اولیه» در دسته‌ی با ریسک بالای تقلب، و صفر برای شرکت‌هایی است که «در تقسیم‌بندی اولیه» در دسته‌ی با ریسک پایین تقلب قرار گرفته باشند. در ادامه جزئیات محاسبه‌ی این دو ماتریس توضیح داده شده است.

ماتریس متغیر پاسخ: خطر تقلب (Y)

در پژوهش‌های خارجی عمدتاً لیستی رسمی از شرکت‌هایی که مشکوک به تقلب هستند وجود دارد و پژوهشگر بدون دخالت در آن لیست، شرکت‌های موجود در آن را در دسته متقلب (یک) و مابقی شرکت‌ها را در دسته غیرمتقلب (صفر) قرار می‌دهد (مثلاً لیست AAER^{۱۲} در آمریکا). بعنوان مثال پوردا و اسکلیکورن (۲۰۱۵) و دجو و همکاران (۲۰۱۱) دقیقاً از همین رویکرد برای تفکیک شرکت‌های با خطر بالای تقلب از سایر شرکت‌ها استفاده کرده‌اند.

در ایران هیچ سازمان یا نهادی متولی اجرای این بررسی و انتشار چنین لیستی با ویژگی فوق نیست. یعنی چنین تفکیکی به طور رسمی وجود ندارد و در صورت وجود نیز در اختیار عموم پژوهشگران قرار نمی‌گیرد. برای رفع این مشکل باید به دنبال ساختن یک «شاخص» بود که بیشترین نزدیکی را با خطر تقلب در شرکت‌ها داشته باشد. برای این منظور و با استناد به استاندارد حسابرسی ۲۴۰ ایران که ارزیابی خطر تقلب را جزو وظایف و مسئولیت‌های حسابرسان دانسته است، در پژوهش حاضر از روشی برای ساخت شاخص مذکور استفاده شد که مبتنی بر رسیدگی‌های افراد حرفه‌ای (حسابرسان مستقل) باشد.

برای این منظور در این پژوهش از روش ارائه شده توسط ریبار و همکاران (۲۰۱۴) استفاده شده است. آن‌ها نشان دادند که اگر رگرسیون از حق الزحمه حسابرسی بر روی عواملی که انتظار می‌رود بر این حق الزحمه اثرگذار باشند اجرا شود، بخش توضیح داده نشده از این رگرسیون (شاخص UAF)^{۱۳} قادر است «تقلب» را در سطح اطمینان ۹۹٪ پیش‌بینی نماید. بر همین اساس، در پژوهش حاضر اعتقاد بر این است که شاخص مذکور، قادر است تخمین‌زننده (شاخص) خوبی از میزان خطر تقلب در شرکت باشد. چرا که به طور غیرمستقیم، ناشی

از برآورد خطر تقلب توسط یک شخص حرفه‌ای (حسابرس) در شرکت‌هاست. بنابراین با استناد به مدل ریبار و همکاران (۲۰۱۴) از بخش توضیح داده نشده از حق الزحمه حسابرسی (UAF) به عنوان شاخصی از خطر تقلب در هر شرکت سال استفاده شد. مدل مذکور به شرح رابطه (۴) است:

$$\text{Ln (AUDIT FEE)}_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + \dots + \beta_n X_{ni,t} + e_{i,t} \quad \text{رابطه (۴)}$$

که در آن:

Ln (AUDIT FEE) : لگاریتم حق الزحمه حسابرسی و X_1 تا X_n : متغیرهای اثرگذار بر حق الزحمه حسابرسی هستند. ریبار و همکاران (۲۰۱۴) X_i ها را از ادبیات مربوطه و پژوهش‌های قبلی استخراج کرده و آن‌ها را به صورت زیر ارائه داده اند:

BIG: در صورتی که حسابرس شرکت در لیست گروه الف از حسابسان 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 رابطه‌ی (۴) و استخراج پسماندها (UAF) برای همه شرکت‌ها، اقدام به دهک‌بندی متغیر UAF شده و شرکت‌ها را در بالاترین دهک (به منظور اعمال

حداکثر احتیاط) بعنوان شرکت‌سال‌های با خطر بالای تقلب (یک) و مابقی دهک‌ها بعنوان سایر شرکت‌سال‌ها (صفر) در نظر گرفته شده است.

ماتریس واژه‌ها (X)

برای تشکیل این ماتریس، فرایندی اجرا شده است که مراحل این فرایند در ادامه به ترتیبی که در اجرای پژوهش طی شده و به صورت بسیار خلاصه توضیح داده شده است. در ابتدا گزارش‌های هیئت مدیره به مجمع برای کلیه شرکت‌های بورسی بین سال‌های ۱۳۸۷ تا ۱۳۹۴ از سایت‌های codal.ir (برای بعد از سال ۸۸) و rdis.ir (برای سال ۸۸ و قبل از آن) دانلود و بررسی شده و از شرکت‌هایی که کمتر از چهار گزارش هیئت مدیره به مجمع (قابل تبدیل به فایل Word با کیفیت قابل قبول) داشتند صرف نظر شد. پس از آن لیستی از سهامداران عمده هر شرکت به همراه درصد مالکیت هر کدام تهیه شد تا شرکت‌هایی که خصوصیتی هستند؛ یعنی سهامداران آن‌ها دارای ویژگی‌های خاصی بوده و شاید مدیران در این شرکت‌ها در استفاده از روش‌هایی با خطر بالای تقلب، احساس مصونیت شغلی بیشتری داشته و ریسک کمتری احساس نمایند، مشخص شوند. در مجموع ۱۴ شرکت با سهامدارانی با ویژگی‌های فوق پیدا شده و از نمونه حذف شد. در اینجا به منظور احترام به حریم حقوقی، از ذکر نام این شرکت‌ها و سهامداران عمده آن‌ها خودداری شده است.

در مرحله بعد، گزارش‌های شرکت‌های عضو نمونه از PDF به Word تبدیل شده است. برای این تبدیل، نسخه اصلی نرم افزار «پارسی پی دی اف» از سایت شرکت سازنده (www.parsipdf.ir) خریداری و فرایند تبدیل انجام شد. سپس شرکت‌هایی که گزارش‌های آن‌ها پس از تبدیل شدن به Word به دلایل متعددی از کیفیت لازم برخوردار نبود از نمونه حذف شد (بعنوان مثال، همه واژه‌ها پس از تبدیل، به هم چسبیده و قابل خواندن نبوده و یا بجای کاراکترهای فارسی، به کاراکترها و سمبل‌های ناخوانا و غیر قابل تشخیص در Word تبدیل شده بود). پس از طی شدن این مراحل، به لیستی نیاز است که اولاً مشخص نماید که در یک گزارش، چند واژه استفاده شده است و ثانیاً تعداد دفعات تکرار هر واژه چقدر بوده است؟ برای این منظور از سایت www.textfixer.com استفاده شد. سپس خروجی مربوطه از سایت مذکور به نرم افزار Excel منتقل شده و در صد مربوط به هر واژه از تقسیم فراوانی آن واژه به جمع فراوانی همه‌ی واژه‌ها به دست آمده است. از آنجا که واژه‌هایی که وارد ماتریس

واژه‌ها می‌شوند باید دارای منطق و پشتوانه باشند و به منظور دوری جستن از اتکای صرف به نرم‌افزار (جکسین، ۲۰۰۲)، تصمیم گرفته شد که در چندین مرحله، پالایشی روی این واژه‌ها انجام شده و واژه‌های نامفهوم، حذف شوند. سپس لیست حاصل از واژه‌های پالایش شده (که فقط شامل واژه‌های بامفهوم است) بعنوان ماتریس واژه‌ها (ماتریس X) در نظر گرفته شد. پس از اجرای این مراحل و با استفاده از ۱۱۷.۷ واژه، ماتریس واژه‌ها (ماتریس X) ساخته شد.

شیوه عملکرد روش «جنگل‌های تصادفی» و روش «ماشین بردار پشتیبان»

داده کاوی مورد استفاده در پژوهش حاضر از لحاظ «هدف» در دسته‌ی «بهینه‌سازی» قرار دارد. زیرا قصد دارد از بین مجموعه وسیعی از داده‌ها، تعداد بهینه‌ای از آن‌ها را که بیشترین قدرت را در تبیین خطر تقلب دارند انتخاب کرده و با استفاده از آن‌ها مدلی برای تشخیص بالا بودن خطر تقلب، برازش نماید. همچنین از لحاظ «تکنیک» نیز از روش‌های «درخت تصمیم» و «آموزش ماشین» استفاده شده است.

روش مبتنی بر درخت تصمیم در این پژوهش، روش جنگل‌های تصادفی (RF) ^{۱۴} نام دارد که توسط یرمن (۲۰۰۱) ارائه شده و در اینجا به منظور تهیه لیستی مرتب شده از واژه‌هایی که بیشترین قدرت را در تشخیص خطر تقلب داشته باشند از آن استفاده شده است. در پژوهش حاضر، روش جنگل‌های تصادفی، اقدام به انتخاب زیرنمونه‌های تصادفی به روش خودراه‌انداز^{۱۵}، از نمونه اصلی (شرکت سال‌های مختلف) کرده و در هر زیرنمونه، یک درخت تصمیم از ماتریس واژه‌ها به ماتریس پاسخ (خطر تقلب) می‌سازد. روش مذکور در هر درخت تصمیم، تشخیص می‌دهد که تعدادی از واژه‌ها در تبیین خطر تقلب، مفیدتر هستند. در نهایت و پس از رشد دادن درخت‌های متعدد، روش جنگل‌های تصادفی، لیستی از واژه‌ها به همراه میزان اهمیت آن‌ها در تشخیص خطر تقلب، ارائه خواهد داد. در این پژوهش برای اجرای روش جنگل‌های تصادفی از بسته‌ای که با همین عنوان (Random Forests) برای نرم‌افزار R (نسخه ۳.۳.۲) نوشته شده استفاده شده است.

سپس باید با استفاده از این لیست رتبه‌بندی شده از واژه‌های با اهمیت، به تشخیص خطر تقلب (بصورت صفر و یک) در گزارش‌ها پرداخت. برای این منظور در این پژوهش از روشی موسوم به ماشین بردار پشتیبان (SVM) ^{۱۶} استفاده شده است. پوردا و اسکلیکورن (۲۰۱۵) بیان

کرده‌اند که استفاده از روش SVM به منظور تشخیص موضوع خطر تقلب در گزارش‌ها نسبت به روش‌های دیگر برتری داشته و همچنین توسط پژوهشگران قبلی به منظور طبقه‌بندی متن استفاده شده است (به عنوان مثال: جو آچیمز، ۱۹۹۸؛ آنتویلر و فرانک، ۲۰۰۴). روش SVM روشی مبتنی بر آموزش ماشین است و به منظور طبقه‌بندی و تفکیک بین دو گروه (مثلاً متقلب و غیر متقلب)، مورد استفاده قرار می‌گیرد. این روش در بسیاری از حوزه‌ها مانند تشخیص افراد از روی صدا و لحن صحبت^{۱۷}، تشخیص چهره از روی عکس^{۱۸} و طبقه‌بندی متن^{۱۹} کاربرد دارد (بارگس، ۱۹۹۸).

آموزش ماشین (Machine Learning)، زیرشاخه‌ای از علوم کامپیوتر است که با استفاده از الگوریتم‌هایی در کامپیوتر، قابلیت آموختن الگوها و روندها بدون برنامه‌نویسی‌های عمیق و پیچیده را فراهم می‌آورد (ساموئل، ۱۹۵۹). این الگوریتم‌ها به کامپیوتر کمک می‌کنند که الگوهای را از داخل داده‌ها بیاموزد و سپس با استفاده از این الگوها، به پیش‌بینی پردازد (کوه‌وی و پرووست، ۱۹۹۸). در واقع با استفاده از آموزش ماشین در کامپیوترها این قابلیت ایجاد می‌شود که همان کاری را انجام دهند که انسان‌ها از طریق اندیشیدن و کسب تجربه انجام می‌دهند (هارناد، ۲۰۰۶). آموزش ماشین از روش‌های مختلفی مانند رگرسیون، طبقه‌بندی، خوشه‌بندی و... قابل اجراست و کاربرد هر یک بستگی به نوع مسأله و هدف پژوهشگر دارد. آنان‌داراجان و آنان‌داراجان (۱۹۹۹) در زمینه حسابرسی، لی (۲۰۱۰) در زمینه محتوای اطلاعاتی گزارش‌های حسابداری، پرولز (۲۰۱۱) در زمینه تشخیص تقلب و یه و دنگ (۲۰۱۲) در زمینه حسابداری مدیریت، چند نمونه از پژوهشگرانی هستند که در حوزه‌ی حسابداری از آموزش ماشین استفاده کرده‌اند.

در این پژوهش ابتدا نمونه به دو بخش شامل «نمونه‌ی آموزش» و «نمونه‌ی آزمون» تقسیم شده و با استفاده از لیست واژه‌های با اهمیتی که از روش جنگل‌های تصادفی استخراج شد به آموزش روش SVM در «نمونه‌ی آموزش» برای تفکیک دو گروه از هم پرداخته شده است. پس از آن، روش مذکور با استفاده از این واژه‌ها و با استناد به آموزشی که دیده است به اختصاص خطر تقلب به هر گزارش در «گروه آزمون» خواهد پرداخت. در این پژوهش برای اجرای روش SVM از بسته‌ای که برای همین منظور (با عنوان e1071) برای نرم افزار R (نسخه ۳.۳.۲) نوشته شده استفاده شده است.

نتایج تجزیه و تحلیل‌ها

در این بخش، نتایج حاصل از اجرای روش‌های آماری و آزمون فرضیه‌های پژوهش ارائه شده است. در ابتدا به منظور محاسبه متغیر وابسته (پاسخ) که همان ماتریس Y است رابطه (۴) بصورت ترکیبی اجرا شد. برای اجرای رابطه (۴) از آزمون F لیمر و هاسمن برای تشخیص نوع داده‌های ترکیبی و از آزمون‌های والد تعدیل شده و وُلدریج به ترتیب به منظور آزمون همسانی واریانس و عدم وجود خودهمبستگی مرتبه اول بین پسماندها در نرم‌افزار Eviews (نسخه ۹) و نرم‌افزار Stata (نسخه ۱۴) استفاده شد. در نهایت مدل با استفاده از روش تابلویی با اثرات ثابت و با وزن مقطعی^{۲۰} به منظور تخفیف مسأله ناهمسانی واریانس اجرا شد. در اینجا به دلیل زیاد بودن تعداد متغیرهای توضیحی در رابطه (۴) و برای پرهیز از حجیم شدن مقاله، از ارائه‌ی نتایج مربوط به ضرایب مدل پرهیز شده و تنها ضرایب تعیین، آماره‌ی F فیشر، نتایج آزمون‌های F لیمر، هاسمن، والد تعدیل شده و وُلدریج در نگاره (۱) ارائه شده است.

نگاره (۱): نتایج اجرای رابطه (۴) برای استخراج پسماندها (UAF)

۲۴/۴۴۴۲	آماره F	آماره دورین واتسون:	۰/۸۷	R2
۰/۰۰۰۰	معناداری آماره F	۱/۷۸	۰/۸۴	R2 تعدیل شده
نتیجه	معناداری	آماره آزمون	فرض صفر	آزمون
داده‌های تابلویی	۰/۰۰۰۰	۳/۰۶۸۸	داده‌های تلفیقی	F لیمر
اثرات ثابت	۰/۰۴۱۲	۲۴/۳۷۶۴	اثرات تصادفی	هاسمن
ناهمسانی واریانس	۰/۰۰۰۰	۱۷۹۶۲	همسانی واریانس	والد تعدیل شده
عدم خودهمبستگی	۰/۱۱۷۰	۲/۶۱۰	عدم خودهمبستگی	وُلدریج

پس از اجرا شدن این مدل به منظور استخراج متغیر پاسخ (ماتریس Y)، پسماندها دهک‌بندی شده و شرکت‌های موجود در بالاترین دهک، بعنوان شرکت‌های با خطر بالای تقلب (یک) و مابقی دهک‌ها بعنوان سایر شرکت‌ها (صفر) در نظر گرفته شده است. پس از آن و با داشتن ماتریس واژه‌ها (ماتریس X) اقدام به اجرای روش‌های «جنگل‌های تصادفی» و «ماشین‌بردار پشتیبان» در محیط نرم‌افزار R و محاسبه‌ی دقت مدل شد. در این مرحله و برای افزایش قابلیت اتکای نتایج، دو پارامتر تغییر داده شده‌اند و در هر تغییر، یکبار روش‌های مذکور اجرا شده‌اند تا قابلیت مقایسه بین دقت مدل در حالت‌های مختلف حاصل شود.

اولین پارامتر، «تعداد درخت‌های تصمیم» در روش جنگل‌های تصادفی است. مقدار پیش فرض برای این پارامتر معمولاً ۵۰۰ درخت در نظر گرفته می‌شود. ولی در این پژوهش، با مقدار ۱۰۰۰ درخت نیز اجرا شده و نتایج با هم مقایسه شده است. دومین پارامتر، «ترکیب‌بندی نمونه‌های آموزش و آزمون» است که به چهار حالت مختلف انجام شده تا قابلیت مقایسه فراهم شود. به این روش اصطلاحاً اعتبارسنجی متقابل چندوجهی^{۲۱} گفته می‌شود (هستی و همکاران، ۲۰۰۸). نگاره (۲) ترکیب‌های مختلفی که از تغییر این دو پارامتر حاصل شده است را نشان می‌دهد. در این نگاره منظور از واژه‌های «بالا، پایین و وسط» محل قرار گرفتن شرکت‌ها در نمونه‌ی ۳۰ شرکتی است که برای حفظ حالت تصادفی، بر اساس حروف الفبا مرتب شده‌اند.

نگاره (۲): هشت حالت مختلف برای آزمون دقت مدل

ردیف	گروه آموزش	گروه آزمون	دقت مدل	
			با ۵۰۰ درخت	با ۱۰۰۰ درخت
۱	آموزش ۲۴ شرکت از بالا طی ۷ سال	آزمون ۶ شرکت از پایین طی ۷ سال	دقت مدل در حالت اول	دقت مدل در حالت پنجم
۲	آموزش ۲۴ شرکت از پایین طی ۷ سال	آزمون ۶ شرکت از بالا طی ۷ سال	دقت مدل در حالت دوم	دقت مدل در حالت ششم
۳	آموزش ۲۴ شرکت از بالا و پایین طی ۷ سال	آزمون ۶ شرکت از وسط طی ۷ سال	دقت مدل در حالت سوم	دقت مدل در حالت هفتم
۴	آموزش ۳۰ شرکت طی سال‌های ماقبل آخر	آزمون ۳۰ شرکت برای سال آخر	دقت مدل در حالت چهارم	دقت مدل در حالت هشتم

پس از تکمیل این نگاره، بین دقت مدل در حالت‌های اول تا چهارم، یک میانگین ساده و بین دقت مدل در حالت‌های پنجم تا هشتم نیز یک میانگین ساده گرفته می‌شود تا امکان مقایسه دقت مدل در دو حالت ۵۰۰ و ۱۰۰۰ برای تعداد درخت‌ها فراهم شود.

در این مرحله، دقت مدل از طریق تقسیم تعداد مشاهدات گروه آزمون که توسط مدل به درستی تشخیص داده شده‌اند بر جمع مشاهدات گروه آزمون، محاسبه شده است. به این معنی که اگر برای یک شرکت سال، خطر تقلب در دنیای واقعی (همان UAF) بالا (پایین) تشخیص

داده شده است، ترکیب روش‌های RF و SVM نیز خطر تقلب را برای این شرکت سال بالا (پایین) تشخیص داده باشد. نتایج این مرحله در ادامه توضیح داده شده است.

دقت مدل در حالت‌های یک تا چهار (۵۰۰ درخت) به ترتیب معادل ۹۲/۶۸٪، ۹۴/۵۹٪، ۸۹/۷۴٪ و ۸۴٪ بوده است که میانگین این چهار حالت معادل ۹۰/۲۵٪ خواهد شد. این دقت‌ها در حالت‌های پنج تا هشت (۱۰۰۰ درخت) نیز به ترتیب معادل ۹۲/۶۸٪، ۹۴/۵۹٪، ۸۹/۷۴٪ و ۸۸٪ بوده است که میانگین این چهار حالت نیز معادل ۹۱/۲۵٪ خواهد شد. یعنی هر دو حالت (۵۰۰ درخت و ۱۰۰۰ درخت) به طور میانگین توانسته‌اند به ترتیب با دقت ۹۰/۲۵٪ و ۹۱/۲۵٪ بالا یا پایین بودن خطر تقلب در شرکت‌ها را به درستی و منطبق با شاخص UAF تشخیص دهند. همچنین به منظور بررسی بیشتر، دو نوع خطا نیز در این هشت حالت محاسبه شده است که عبارتند از الف) خطا در تشخیص خطر بالای تقلب: خطر تقلب در یک شرکت سال، بالا تشخیص داده شود در حالی که در واقع پایین است و ب) خطا در تشخیص خطر پایین تقلب: خطر تقلب در یک شرکت سال، پایین تشخیص داده شود در حالی که در واقع بالا است.

در حالت‌های یک تا چهار (۵۰۰ درخت) خطا در تشخیص خطر بالای تقلب، معادل صفر و خطا در تشخیص خطر پایین تقلب به ترتیب معادل ۷/۳۲٪، ۵/۴۱٪، ۱۰/۲۶٪ و ۱۶٪ بوده است (میانگین: ۹/۷۵٪). در حالت‌های پنج تا هشت (۱۰۰۰ درخت) نیز خطاهای مذکور، به ترتیب معادل ۷/۳۲٪، ۵/۴۱٪، ۱۰/۲۶٪ و ۱۲٪ بوده است (میانگین: ۸/۷۵٪). یعنی ترکیب روش‌های RF و SVM در پژوهش حاضر در تشخیص خطر بالای تقلب، خطایی نداشته و در تشخیص خطر پایین تقلب به ترتیب به طور میانگین ۹/۷۵٪ و ۸/۷۵٪ خطا داشته است.

نگاره (۳) نتایج اجرای رابطه (۳) به روش لاجیت رتبه‌ای را نشان می‌دهد. در اینجا یادآوری می‌شود که مفهوم عرض از مبدأ در رگرسیون‌های لاجیت رتبه‌ای وجود نداشته و آماره $Pseudo R^2$ نیز در رگرسیون‌های خانواده‌ی لاجیت به دلیل گسسته بودن متغیر وابسته و عدم امکان ایجاد تناظر یک به یک بین مشاهدات و احتمال وقوع متغیر وابسته، مقدار پایینی دارد (گرین، ۲۰۱۲: ۵۳۳).

پس از اجرای رابطه (۳) به روش لاجیت رتبه‌ای، نرم افزار Eviews قادر است PIها را برای هر شرکت سال محاسبه کند. پس از استخراج PIهای حاصل از اجرای رابطه (۳)، می‌توان این

مقادیر را در رابطه (۱) قرار داده و فرضیه‌های پژوهش را آزمون کرد. یادآوری می‌شود که در اجرای رابطه (۱) مقدار متغیر $R_{i,t}$ همان مقدار شاخص تقلب (صفر و یک) است که توسط ترکیب روش‌های RF و SVM به هر شرکت سال اختصاص داده شده است.

نگاره (۳): نتایج اجرای رابطه (۳) برای محاسبه‌ی PIها

متغیر	ضریب	خطای استاندارد	آماره Z	معناداری
$Y_{i,t}$	۱/۴۳۰۶	۰/۴۴۸۰	۳/۱۹۲۹	۰/۰۰۱۴
$Growth_{i,t}$	-۰/۰۴۵۱	۰/۰۲۰۸	-۲/۱۶۰۷	۰/۰۳۰۷
آماره LR	۲۵/۵۲۹۰	مقدار آستانه‌ای C1	۰/۰۹۶۹	آماره‌ی Pseudo R ²
معناداری آماره LR	۰/۰۰۰۰	مقدار آستانه‌ای C2	۰/۱۲۳۲	۰/۰۸۲۲

$$R_{i,t} = \alpha + \beta_1 P(I_{i,t+1}=1) + \beta_2 P(I_{i,t+1}=3) + NOA_{i,t} + LQD_{i,t} + EXF_{i,t} + e_i \quad \text{رابطه (۱)}$$

رابطه (۱) در Eviews با استفاده از روش لاجیت ساده تخمین زده شد. نگاره (۴) نتایج حاصل از این تخمین را به همراه نتایج آزمون نیکویی برازش (هاسمر-لمشو)^{۲۲} مخصوص مدل‌های لاجیت نشان می‌دهد:

نگاره (۴): نتایج اجرای رابطه (۱) برای آزمون فرضیه‌های پژوهش

متغیر	ضریب	خطای استاندارد	آماره Z	معناداری
عرض از مبدأ	-۱۸۱/۵۶۷۸	۱۶۴/۰۹۹۴	-۱/۱۰۶۴	۰/۲۶۸۵
$P(I_{i,t+1}=1)$	۱۸۰/۴۶۷۳	۱۶۴/۵۲۳۷	۱/۰۹۶۹	۰/۲۷۲۷
$P(I_{i,t+1}=3)$	۱۷۸/۶۶۹۳	۱۶۶/۴۰۴۷	۱/۰۷۳۷	۰/۲۸۳۰
NOA	۱/۱۶۰۷	۱/۳۴۹۶	۰/۸۶۰۰	۰/۳۸۹۸
LQD	-۳/۴۶۴۵	۴/۶۶۶۲	-۰/۷۴۲۵	-۰/۷۴۲۵
EXF	۵/۷۸۰۸	۳/۵۳۸۳	۱/۶۳۳۸	۰/۱۰۲۳
R ² می‌مک فادن	۰/۱۲۰۰	آماره LR (معناداری)	۱۱/۸۸۴۸	(۰/۰۳۶۴)
آزمون	فرض صفر	آماره	معناداری	نتیجه
هاسمر - لمشو	نیکویی برازش	۵/۴۵۴۳	۰/۷۰۸۱	نیکویی برازش

برای اثبات انطباق انگیزه‌های تقلب در مدیران با الگوی نظریه CPT لازم است ضریب $P(I_{i,t+1}=1)$ مثبت و معنادار و ضریب $P(I_{i,t+1}=3)$ منفی و معنادار شوند تا به ترتیب، فرضیه‌های دوم و اول پژوهش، رد نشوند. همانطور که نگاره (۴) نشان می‌دهد با وجود رد نشدن نیکویی

برازش در اجرای رابطه (۱)، باز هم هیچ کدام از دو ضریب معنادار نشده‌اند و این به آن معناست که احتمال مثبت یا منفی بودن انحراف عملکرد شرکت از عملکرد مرجع (متوسط صنعت) در سال بعد، بر انگیزه‌های تقلب مدیران در سال جاری تأثیری ندارد. در واقع انگیزه‌های تقلب در مدیران از الگوی نظریه CPT تبعیت نمی‌کند. این نتیجه در بخش بعد، تحلیل و تفسیر شده است.

نتیجه‌گیری

در این پژوهش از یکی از منابع اطلاعاتی حسابداری (گزارش‌های هیئت مدیره به مجمع) در قالب دو مورد از تکنیک‌های داده کاوی (درخت تصمیم و آموزش ماشین) استفاده شد تا به ارزیابی و تشخیص یکی از موضوعات مورد توجه در این حوزه (خطر تقلب) پرداخته شود. نتایج نشان داد که این روش در مجموع قادر است بالا یا پایین بودن خطر تقلب در شرکت‌ها را با دقتی بیش از ۹۰ درصد به درستی تشخیص دهد. یعنی داده کاوی قادر است با تجزیه و تحلیل متن گزارش‌های هیئت مدیره به مجمع، خطر تقلب در شرکت‌ها را با دقت قابل قبولی ارزیابی نماید.

پس از آن و با استفاده از خروجی مدل مذکور، تلاش شد به بررسی انطباق انگیزه‌های تقلب در مدیران با الگوی پیشنهاد شده توسط نظریه چشم‌انداز تجمعی (CPT) پرداخته شود. بر اساس نظریه مذکور، احتمال بالاتر یا پایین تر بودن عملکرد مدیران در سال آینده نسبت به متوسط عملکرد صنعت، بر انگیزه تقلب در آن‌ها (به صورت یک الگوی چهاربعدی) تأثیرگذار است. اما نتایج نشان داد که چنین الگویی در مدیران شرکت‌های ایرانی (طی دوره زمانی و در نمونه‌ی این پژوهش) مشاهده نمی‌شود. این نتیجه را می‌توان به دو صورت تفسیر کرد.

تفسیر اول این است که انگیزه‌های تقلب در مدیران از الگوی نظریه چشم‌انداز تجمعی پیروی نمی‌کند. یعنی ویژگی‌های ریسک‌پذیری و ریسک‌گریزی در مدیران در شرکت‌های ایرانی به گونه‌ای است که در تصمیم‌گیری‌های خود، نقطه‌ی مرجعی را مدنظر قرار نمی‌دهند و احتمالاً عوامل دیگری (غیر از یک نقطه‌ی مرجع) بر تشدید یا تضعیف این انگیزه در مدیران مؤثر است. ممکن است عواملی مانند عوامل سیاسی (مانند توافق برجام)، زیست محیطی (مانند

خشک‌سالی)، اجتماعی (مانند ایجاد مصداق‌های مسئولیت‌های اجتماعی در شرکت)، فرهنگی (مانند اتحاد مردم برای نخریدن محصولات بی کیفیت) و مواردی از این دست، نسبت به یک نقطه‌ی مرجع (مانند متوسط صنعت) تأثیر بسیار بزرگتری بر شکل‌گیری انگیزه‌های تقلب در مدیران داشته باشند. پی بردن به عوامل مذکور، خارج از حوزه‌ی پژوهش حاضر بوده و نیازمند تدوین پژوهش مستقلی است.

تفسیر دوم این است که در صورت پیروی کردن انگیزه‌های تقلب در مدیران ایرانی از الگوی نظریه چشم‌انداز تجمعی نیز به احتمال زیاد نقطه‌ی مرجعی که در ذهن مدیران تعریف شده است، متوسط عملکرد صنعت (که در این پژوهش به پیروی از فانگ (۲۰۱۵) بعنوان شاخصی از نقطه‌ی مرجع در نظر گرفته شد) نیست. ممکن است عوامل دیگری مانند میزان سود مبنای پاداش مدیران و یا انگیزه‌های شخصی آن‌ها (مانند انگیزه‌های سیاسی یا حتی انگیزه‌های مجرمانه مانند اختلاس) نقش نقطه‌ی مرجع را در ذهن آن‌ها ایفا کند. فانگ (۲۰۱۵) از متوسط عملکرد صنعت بعنوان نقطه مرجع استفاده کرده و الگوی نظریه چشم‌انداز تجمعی را در مدیران شرکت‌های آمریکایی مشاهده نمود. قبلاً نیز کیم (۱۹۹۲) به این نتیجه رسیده بود که اگر مدیر احساس کند که تیم تحت مدیریت او به عملکرد سال گذشته بعنوان یک نقطه مرجع نگاه کرده و فکر می‌کنند که قادر به دستیابی به آن در بودجه نیستند، با وحشت از اینکه در سال آینده قادر به دستیابی به ظرفیت بودجه شده‌ی شرکت نباشد، این ظرفیت را در بودجه، اندک پیش‌بینی کرده و بر همین اساس، منابع لازم برای تولید شرکت در سال آتی نیز، کمتر خریداری یا تولید می‌شود. همین امر باعث می‌شود که برای عملکرد شرکت در سال آتی، به اندازه کافی منابع وجود نداشته و همین منجر به کسری بودجه خواهد شد.

این پژوهش با محدودیت‌های زیادی نیز مواجه بود که در صورت رفع شدن در آینده می‌توان پژوهش‌هایی از این دست را با دقت و سرعت بالاتری سازماندهی کرد. در اینجا به منظور دوری از اطاله کلام، تنها به چهار مورد از مهم‌ترین این محدودیت‌ها اشاره شده است. اول، محدودیت مربوط به تفکیک اولیه شرکت‌ها به دو گروه شامل شرکت‌های با خطر بالای تقلب و سایر شرکت‌ها بود (ماتریس Y) که این پژوهش به دلیل عدم دسترسی به یک لیست از پیش تعیین شده برای این تفکیک، ناچار به استفاده از یک روش غیر مستقیم شد که مبتنی بر حق الزحمه حسابرسی است. دوم، محدودیت مربوط به دسترسی به فایل Word گزارش‌های

هیئت مدیره است که نویسندگان مقاله حاضر را ناچار به تبدیل گزارش‌های هیئت مدیره کرده و منجر به کاهش شدید در حجم نمونه شد. سوم، محدودیت مربوط به افشا نشدن حق الزحمه حسابرسی در یادداشت‌های توضیحی شرکت‌های بورسی بود که از هیچ منبع معتبر دیگری نیز قابل دستیابی نبوده و حجم نمونه را به اندازه ۲۶ شرکت کاهش داد. چهارم، محدودیت مربوط به عدم وجود الگوریتم‌ها و نرم افزارهای نشانه‌گذاری (Qtagger) برای متون فارسی بود، که ناچاراً منجر به تجزیه و تحلیل گزارش‌های هیئت مدیره به منظور پالایش واژه‌ها بصورت دستی شد.

پیشنهادات

در اینجا با استناد به محدودیت‌ها و نتایج پژوهش، پیشنهادهایی به شرح زیر ارائه شده است. اول به سازمان بورس اوراق بهادار پیشنهاد می‌شود که بصورت مستقل و نه با اتکا به حساب‌برسان، رسیدگی‌هایی مبنی بر یافتن نواقص و خطرات گزارش‌گری در شرکت‌ها را در دستور کار قرار داده و در صورت نیاز به ارائه‌ی تذکر به شرکت‌ها، لیست شرکت‌هایی که تذکر دریافت کرده‌اند را به صورت رسمی منتشر نماید تا پژوهشگران حوزه‌ی تقلب و دستکاری صورت‌های مالی در آینده به لیست مذکور اتکا کنند. همچنین به سازمان مذکور پیشنهاد می‌شود که شرکت‌های بورسی را ملزم نماید که علاوه بر فایل PDF صورت‌های مالی و گزارش‌های هیئت‌مدیره، فایل Word این گزارش‌ها را نیز بر روی سایت گُذال بارگذاری نمایند. این الزام، نه برای شرکت‌ها و نه برای سازمان بورس اوراق بهادار، هیچ هزینه‌ای به دنبال ندارد. اما امکان بسیاری از تجزیه و تحلیل‌های بعدی را برای پژوهشگران فراهم آورده و منجر به غنای ادبیات حسابداری و مالی کشور در این حوزه خواهد شد. دوم به پژوهشگران آتی پیشنهاد می‌شود که با توجه به روند رشد روزافزون استفاده از رویکردهای جدید در پژوهش‌های حوزه‌ی حسابداری و مالی، از روش‌های معمول و مرسوم در این حوزه فاصله گرفته و از روش‌های جدیدتر مانند داده کاوی نیز استفاده کنند. همچنین می‌توان عوامل دیگری را که در بخش نتیجه‌گیری، به عنوان عوامل احتمالی اثرگذار بر شکل‌گیری انگیزه‌های تقلب در مدیران، پیشنهاد شد مورد بررسی قرار داده و حتی پژوهش‌هایی بین رشته‌ای (با رشته‌های روانشناسی، جامعه‌شناسی و حقوق) را تدوین نمود.

پی نوشت

۱ Cumulative Prospect Theory (CPT)	۱۳ Unexplained Audit Fee (UAF)
۲ Management Discussion and Analysis	۱۴ Random Forests (RF)
۳ Decision Making	۱۵ Bootstrap
۴ Expected Utility Theory	۱۶ Support Vector Machine
۵ Subjective Expected Utility	۱۷ Speaker Identification
۶ Reference Point	۱۸ Face Detection in Images
۷ Prospect Theory	۱۹ Text Categorization
۸ Data Mining	۲۰ Cross Section
۹ Retrospection	۲۱ K-Fold Cross Validation
۱۰ Description	۲۲ Goodness-of-Fit (Hosmer-Lemeshow)
۱۱ Ordered Logit	
۱۲ Accounting and Auditing Enforcement Releases	

منابع

- اعتمادی، حسین و حسن زلّقی. (۱۳۹۲). کاربرد رگرسیون لجستیک در شناسایی گزارشگری مالی متقلبان. دانش حسابرسی، سال سیزدهم، شماره ۵۱، تابستان ۱۳۹۲.
- افلاطونی، عباس. (۱۳۹۴). تجزیه و تحلیل آماری با Eviews در تحقیقات حسابداری و مدیریت مالی. تهران: انتشارات ترمه.
- بروکس، کریس. (۱۹۷۱). مقدمه ای بر اقتصادسنجی مالی، احمد بدری و عبدالمجید عبدالباقی، تهران: انتشارات نص، (۱۳۸۹).
- پورحیدری، امید و سعید بذرافشان. (۱۳۹۱). بررسی سودمندی استفاده از چک لیست راهنمای کشف تقلب در ارزیابی خطر تقلب مدیریت. پژوهش‌های تجربی حسابداری: ۱ (۳) ۶۹-۸۶.
- رهنمای رودپشتی، فریدون. (۱۳۹۱). داده کاوی و کشف تقلب‌های مالی. دانش حسابداری و حسابرسی مدیریت: ۱ (۳) ۱۷-۳۳.
- زارع بهنمیری، محمدجواد و اسفندیار ملکیان کله بستی. (۱۳۹۵). رتبه بندی عوامل مؤثر بر احتمال تقلب مالی با توجه به گزارش حسابرسی صورت‌های مالی. پژوهش‌های تجربی حسابداری: ۶ (۲۱)، ۱۷-

- فرج‌زاده دهکردی، حسن و لیلا آقایی. (۱۳۹۴). سیاست تقسیم سود و گزارشگری مالی متقالبانه. مطالعات تجربی حسابداری مالی: ۱۳ (۴۵): ۹۷-۱۱۴.
- کمیته تدوین استانداردهای حسابرسی. (۱۳۸۴). استاندارد حسابرسی ۲۴۰. مسئولیت حسابرس در ارتباط با تقلب و اشتباه، بند ۴.
- مصلح شیرازی، علی نقی، محمد، نمازی، علی، محمدی و احمد رجبی. (۱۳۹۲). تئوری چشم انداز و مدل‌سازی الگوی تصمیم‌گیری مدیران در بخش صنعت. چشم انداز مدیریت صنعتی: ۶ (۱۰): ۳۳-۹.
- مظلومی، نادر، فریبا، لطیفی و هیوا آسایی. (۱۳۸۶). بررسی رابطه ریسک‌پذیری مدیران با عملکرد سازمان‌ها در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. فصلنامه مطالعات مدیریت: ۱۴ (۵۶): ۷۱-۹۲.
- مهرانی، کاوه و رضا حصارزاده. (۱۳۸۷). مروری بر تئوری‌ها و مدل‌های کشف تقلب. دانش و پژوهش حسابداری، زمستان ۱۳۸۷: (۱۵): ۶-۱۱.
- نمازی، محمد و فهیمه ابراهیمی. (۱۳۹۵). مدل و تعیین اولویت عوامل مؤثر بر قصد گزارش تقلب‌های مالی توسط حسابداران. مطالعات تجربی حسابداری مالی، سال ۱۲، شماره ۴۹، بهار ۱۳۹۵، ۲۸-۱.
- وحیدی الیزی، ابراهیم و حامد حامدیان. (۱۳۸۸). برداشت حسابرسان ایران از کارایی علایم خطر در کشف گزارشگری مالی متقالبانه. تحقیقات حسابداری، شماره ۳، پاییز ۱۳۸۸: ۱۹۷-۱۶۲.
- 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, 32-58 .
- Anandarajan, M. , & Anandarajan, A. (1999). A comparison of machine learning techniques with a qualitative response model for auditor's going concern reporting. *Expert Systems with Applications*, 16 (4) , 385-392 .
- Antweiler, W, & Frank, M. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *Journal of Finance*, 59 (3) , 1259-1294.
- Baum, C. F. (2006). *An Introduction to Modern Econometrics Using Stata*. College Station, Texas: Stata Press .
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32 .

- Brooks, C. (1971). *Introductory Econometrics for Finance* (A. Badri & A. Abdolbaghi, Trans. Vol. 1). Tehran: Nass. [In Persian]
- Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2, 121-167.
- Burns, N, & Kedia, S. (2006). The Impact of Performance-based Compensation on Misreporting. *Journal of Financial Economics*, 79 (1) , 35–67.
- Cecchini, M. , H. Aytug, G. Koehler, & Pathak. , P. (2010). Making words work: Using financial text as a predictor of financial events. *Decision Support Systems*, 50 (1) , 164–175 .
- Debrecey, R. S. , & Gray, G. L. (2010). Data mining journal entries for fraud detection: an exploratory study. *International Journal of Accounting Information Systems*, 11 (3) , 157–181 .
- Dechow, P. , Ge, W. , Larson, C. , & Sloan, R. (2011). Predicting Material Accounting Misstatements. *Contemporary Accounting Research*, 28 (1) , 17-82 .
- Edwards, W. (1955). The Prediction of Decisions among Bets. *Journal of Experimental Psychology*, 50 (3) , 201–214.
- Elliott, R. K. , & Willingham, J. J. (1980). *Management Fraud: Detection and Deterrence*. In (35–46). New York: Petrocelli Books .
- 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]
- Fernandes, N, & Guedes, J. (2010). Keeping Up with The Joneses: A Model And A Test Of Collective Accounting Fraud. *European Financial Management*, 16 (1) , 72–93.
- Fung, M. K. (2015). Cumulative Prospect Theory and Managerial Incentives for Fraudulent Financial Reporting. *Contemporary Accounting Research*, 32 (1) , 55–75 .
- 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 .
- Greene, W. H. (2012). *Econometric Analysis* (Vol. 7): Prentice Hall .

- Harbaugh, W. T, K. Krause, & Vesterlund, L. (2010). The Fourfold Pattern of Risk Attitudes in Choice and Pricing Tasks. *The Economic Journal*, 120 (545) , 595–611.
- Harnad, S. (2006). The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence. In *Philosophical and Methodological Issues in the Quest for the Thinking Computer*: Kluwer .
- Hastie, T. , Tibshirani, R. , & Friedman, J. (2008). *The Elements of Statistical Learning Data Mining, Inference, and Prediction (Vol. 2)*. Stanford, California: Springer .
- 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 .
- Iran Audit Standards Committee. (2005). Auditing Standard Number 240: Auditor's Responsibility Regarding Fraud and Misconduct: Iranian Audit Organization. [In Persian]
- Jackson, J. m. (2002). Data mining: a conceptual overview. *Communications of the Association for Information Systems*, 8 (1) , Article 19 .
- Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Paper Presented at The European Conference On Machine Learning .
- Kim, D. C. (1992). Risk Preferences in Participative Budgeting. *The Accounting Review*, 67 (2) , 303–318 .
- 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 .
- Kohavi, R. , & Provost, F. (1998). Glossary of Terms. *Machine Learning*, 30, 271–274 .
- Li, F. (2010). The information content of forward-looking statements in corporate filings—a naïve Bayesian machine learning approach. *Journal of Accounting Research*, 48 (5) , 1049–1102 .

- Lin, J. W. , Hwang, M. I. , & Becker, J. D. (2003). A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal*, 18 (8) , 657–665 .
- Liou, F. M. (2008). Fraudulent financial reporting detection and business failure prediction models: a comparison. *Managerial Auditing Journal*, 23 (7) , 650–662 .
- Mazlumi, N. , Latifi, F. , & Aasaai, H. (2008). Risk Taking Behavior of CEO'S and Firm Performance (Companies Registered with Tehran Stock Exchange). *Management Studies in Development and Evolution*, 14 (56) , 71-92. [In Persian]
- Mehrani, K. , & Hesarzadeh, R. (2008). A Review of theories and models of fraud detection. *Accounting Knowledge and Research*, winter (15) , 6-11 .[In Persian]
- Moslehshirazi, A. , Namazi, M. , Mohammadi, A. , & Rajabi, A. (2013). Prospect Theory and Modeling Managers Decision Making in the Industrial Sector. *Journal of Industrial Management Perspective*, (10) , Summer, 9-33. [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]
- O'Connor, J. P, R. L. Priem, J. E. Coombs, & Gilley, K. M. (2006). Do CEO Stock Options Prevent or Promote Fraudulent Financial Reporting? *Academy Of Management Journal*, 49 (3) , 483–500.
- Perols, J. (2011). Financial statement fraud detection: an analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice and Theory* 30 (2) , 19-50 .
- Pinello, A. S. (2008). Investors' Differential Reaction to Positive versus Negative Earnings Surprises. *Contemporary Accounting Research*, 25 (3) , 891–920.
- Pourheydari, O. , & Bazrafshan, S. (2012). An Examination of the Usefulness of fraud detection Decision Aid in Assessment of Management Fraud Risk. *Journal of Empirical Researches in Accounting*, 1 (3) , 67-84. [In Persian]
- 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 .
- RahnamayRoodposhti, F. (2012). Data mining & Financial Fraud. *Journal of Management Accounting and Auditing Knowledge*, 1 (3) ,17-34. [In Persian]

- Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3 (3) , 210 - 229 .
- Tackett, J. A. (2013). Association rules for fraud detection. *Journal of Corporate Accounting and Finance*, 24 (4) , 15–22 .
- Torugsa, N & Arundel, A. (2017). Rethinking the effect of risk aversion on the benefits of service innovations in public administration agencies. *Research Policy*, 46 (5) , 900-910 .
- Tversky, A, & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5 (4) , 297–323 .
- 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]
- Von Neumann, J, & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press .
- Yeh, T. H. , Deng, S. ., (2012). Application of machine learning methods to cost estimation of product life cycle. *International Journal of Computer Integrated Manufacturing* 25 (4) , 340–352 .
- ZareBahnamiri, M. , & Malekian, E. (2016). Ranking the Factors Affecting Financial Fraud Probability, According to Audited Financial Statements. *Journal of Empirical Researches in Accounting*, 6 (1) , 1-18. [In Persian]