



سنجش اقلام تعهدی از طریق گزارش‌های متنی با استفاده از تکنیک رگرسیون بردار پشتیبان و رویکرد مبتنی بر لغت‌نامه

زهرة زیودار^۱

داریوش فروغی^۲

غلامحسین کیانی^۳

تاریخ پذیرش: ۹۸/۱۰/۲۸

تاریخ دریافت: ۹۸/۰۸/۲۸

چکیده

هدف این پژوهش سنجش اقلام تعهدی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران، از طریق گزارش‌های کیفی و ناساختاریافته، با استفاده از تکنیک‌های متن‌کاوی است. جامعه آماری پژوهش، شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران و نمونه مورد مطالعه شامل ۲۰۵ شرکت در بازه زمانی هفت‌ساله ۱۳۹۰ تا ۱۳۹۶ است. تکنیک آماری مورد استفاده جهت آزمون فرضیه‌ها، رگرسیون چندمتغیره از نوع داده‌های ترکیبی با استفاده از نرم‌افزار آماری استتاست. اندازه‌گیری اقلام تعهدی متنی از طریق گزارش هیات مدیره به مجمع شرکت با استفاده از دو تکنیک رگرسیون بردار پشتیبان و روش مبتنی بر لغت‌نامه در نرم‌افزار پایتون انجام گرفت. یافته‌های پژوهش نشان داد که ۱۹/۶ درصد از تغییرات اقلام تعهدی واقعی (اقلام تعهدی سطح شرکت) توسط اقلام تعهدی متنی قابل تبیین است. همچنین اقلام تعهدی لغت‌نامه‌ای (برآورد شده بر اساس فرهنگ واژگان) قادر به تبیین ۱۵/۳ درصد از تغییرات اقلام تعهدی واقعی هستند؛ در نتیجه اقلام تعهدی متنی دارای قدرت تفسیری بیشتری نسبت به اقلام تعهدی لغت‌نامه‌ای جهت توصیف اقلام تعهدی واقعی در شرکت است. در صورتی که در یک مدل، اقلام تعهدی متنی و لغت‌نامه‌ای همزمان به کار گرفته شوند، قادر به تبیین ۵۳ درصد از تغییرات اقلام تعهدی واقعی هستند.

واژه‌های کلیدی: اقلام تعهدی، داده‌های ناساختاریافته، متن‌کاوی.

۱- دانشجوی دکتری حسابداری، گروه حسابداری، واحد شهرکرد، دانشگاه آزاد اسلامی، شهرکرد، ایران. zzivdar55@yahoo.com
۲- دانشیار، دانشکده امور اداری و اقتصاد، گروه حسابداری، دانشگاه اصفهان، اصفهان، ایران. (نویسنده مسوول) foroghi@ase.ui.ac.ir
۳- استادیار، دانشکده امور اداری و اقتصاد، گروه اقتصاد، دانشگاه اصفهان، اصفهان، ایران. gh.kiani@ase.ui.ac.ir

۱- مقدمه

اطلاعات مالی شفاف و قابل مقایسه، عامل اصلی پاسخگویی و تصمیم‌گیری‌های اقتصادی و ابزار اصلی جهت ارائه اطلاعات مالی شفاف، صورت‌های مالی اساسی و گزارش‌های مکمل است. سود، بااهمیت‌ترین رقم حسابداری طی حداقل هفتادسال گذشته است که محاسبه و اندازه‌گیری آن به روش مرسوم حداکثر صدوده‌سال عمر دارد (ثقفی، ۱۳۹۳). دیچو و همکاران (۲۰۱۰) برای سنجش کیفیت سود از شاخص‌های مختلفی مانند پایداری سود، قابلیت پیش‌بینی سود، هموارسازی سود و کیفیت اقلام تعهدی استفاده نمودند. از سال‌های پیش تا کنون پژوهش‌های بسیاری درباره اقلام تعهدی با تکیه بر گزارش‌ها و داده‌های ساختاریافته^۱ و افشاهای کمی انجام گرفته است (دیچو و دیچف ۲۰۰۲؛ فرانسیس و همکاران ۲۰۰۵؛ هریبار و همکاران ۲۰۱۷؛ ثقفی و محمد ۱۳۹۱؛ فروغی و همکاران ۱۳۹۲؛ رحیمی و همکاران ۱۳۹۵). داده‌های ساختاریافته به داده‌هایی گفته می‌شود که به‌طور کاملاً مستقل از همدیگر ولی یکسان از لحاظ ساختاری در یک محل گردآوری شده‌اند (داس، ۲۰۱۴). به‌طور عمده، اطلاعات مالی در صورت‌های مالی نهفته است، اما باید توجه داشت که اطلاعات مالی نیز می‌تواند متنی باشد. پژوهشگران برای درک رویکردهای رفتاری مدیران و رفتارهای تجاری به تجزیه و تحلیل این داده‌ها و استخراج آن از متن نیاز دارند (توریگان، ۲۰۱۹). امروزه بخش زیادی از اطلاعات مالی از طریق افشاهای کیفی^۲ مانند گزارش‌های روایتی^۴ و اسناد متنی^۵ منتشر می‌شوند. اسناد متنی حسابداری شامل اصول عمومی پذیرفته شده حسابداری، صورت‌های مالی سالانه شرکت‌ها به همراه سایر گزارش‌های حسابداری ضمیمه آنها، است (فیشر و همکاران ۲۰۱۰). تحلیل گزارش‌های روایتی که بیشتر شامل داده‌های ناساختاریافته^۳ است در حسابداری و مالی رو به افزایش است (بیتی، ۲۰۱۴). داده‌های ناساختاریافته به آن دسته از داده‌هایی اشاره دارد که مدل داده‌ای و یا روش سازمان‌دهی از پیش تعیین شده‌ای ندارند. داده‌های ناساختاریافته معمولاً

متنی سنگین و مشکل برای تحلیل دارند، اما بسیاری از پژوهشگران از این داده‌ها برای استخراج احساسات، ایجاد شاخص احساسات و پیش‌بینی بازده استفاده می‌کنند. فایل‌ها و افشاهای اجباری مانند فرم‌های مختلف گزارش تفسیری مدیریت، بخش‌های گوناگون گزارش هیات مدیره به مجمع، اعلامیه‌های سود و دیگر بیانیه‌های منتشر شده، کنفرانس‌های تلفنی، مقالات چاپ شده در فضاها اجتماعی و رسانه‌های مالی، اخبار، یادداشت‌های تحلیلی، همگی نمونه‌ای از منابع داده‌های مالی ناساختاریافته می‌باشند (لی‌گو و همکاران ۲۰۱۶). آرایه ابزارهایی که با بررسی منابع مالی و تحلیل آنها بتواند از میان داده‌های خام، دانش مفید را استخراج نماید، منجر به شکل‌گیری شاخه‌ای در هوش مصنوعی گردید که به داده‌کاوی^۶ معروف است (ویتن و فرانک ۲۰۰۵). استفاده از داده‌کاوی در مورد متن، شاخه‌ای در علوم هوش مصنوعی به نام متن‌کاوی^۱ ایجاد کرد (مورو و همکاران، ۲۰۱۵). یکی از منابع داده‌ای مهم برای متن‌کاوی، گزارش تجزیه و تحلیل‌های مدیریت^۲ است که حاوی اطلاعات کیفی مهم برای تحلیل است. در ایران ارائه گزارش تفسیری مانند آن چه که از سال ۲۰۰۳ در آمریکا و دیگر کشورها ارائه می‌گردد وجود نداشت تا این که طبق اطلاعیه مورخ ۱۳۹۶/۱۰/۵ سازمان بورس اوراق بهادار، شرکت‌های پذیرفته شده در بورس و فرابورس مکلف شدند «گزارش تفسیری مدیریت» را افشا کنند. به دلیل عدم تهیه این گزارش از سوی بسیاری از شرکت‌های بورسی ایران و همپوشانی بالای سرفصل‌های مهم گزارش هیات مدیره با گزارش تفسیری مدیریت، گزارش هیات مدیره شرکت‌ها به-عنوان یک گزارش کیفی، در انجام پژوهش‌های متن‌کاوی، مورد استفاده قرار می‌گیرد. در ادبیات مالی و حسابداری، پژوهشگرانی مانند لو و همکاران (۲۰۱۷)، دایر و همکاران (۲۰۱۶)، لاندلم و همکاران (۲۰۱۴)، در پژوهش خود بر گزارش تفسیری مدیریت تمرکز نمودند. پژوهشگران حسابداری و مالی تکنیک‌های تحلیل متنی بسیاری را برای درک رابطه بین متن در افشاهای شرکتی (مانند گزارش تفسیری

ابزارهای نوین سنجش و اندازه‌گیری مانند متن‌کاوی و هوش مصنوعی به‌منظور تعیین کیفیت سود در مبحث مدیریت و دستکاری سود نقش با اهمیتی دارد. از سوی دیگر پژوهش حاضر با بررسی گزارش هیات مدیره، سعی در سنجش قابلیت و توانایی این دسته از گزارش‌های کیفی در تبیین و توصیف عناصر بنیادی شرکت‌ها و عملکرد آنها دارد، لذا نتیجه آن می‌تواند برای قانون‌گذاران، کارشناسان بورس، فعالان بازار سرمایه، حسابرسان و حسابداران، تحلیل‌گران مالی، سازمان حسابرسی و نهادهای نظارتی مفید باشد. در پژوهش حاضر اقلام تعهدی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران از طریق گزارش کیفی هیات مدیره به مجمع شرکت‌ها، با کمک دو تکنیک متن‌کاوی (رگرسیون بردار پشتیبان و روش لغت-نامه‌ای) اندازه‌گیری می‌شود تا به این پرسش پاسخ داده شود که آیا می‌توان در شرکت‌های پذیرفته شده بورس اوراق بهادار تهران با استفاده از تکنیک‌های متن‌کاوی بر توانایی گزارش‌های کیفی و ناساختاریافته‌ای چون گزارش هیات مدیره جهت تبیین اقلام تعهدی تکیه کرد؟ ساختار این پژوهش به‌صورت زیر خواهد بود: ابتدا در بخش ادبیات و مبانی نظری به مفاهیم داده‌کاوی، متن‌کاوی، داده‌های ساختاریافته و ناساختاریافته در حوزه مالی و حسابداری اشاره خواهد شد. چارچوب‌های موجود در این حوزه مورد بررسی قرار خواهد گرفت و مبانی نظری پشتیبانی‌کننده فرضیه‌های پژوهش ارائه خواهد شد. در بخش بعدی، به مرتبط‌ترین پژوهش‌های داخلی و خارجی با این پژوهش اشاره می‌شود. در ادامه، فرضیه‌ها، مدل‌ها، متغیرهای پژوهش و روش -شناسی تشریح و در پایان نتایج آمار توصیفی و استنباطی ارائه و نتیجه‌گیری انجام گرفته است.

۲- ادبیات و مبانی نظری

با پیدایش روزافزون انواع کسب و کارهای نوین که منجر به افزایش پیچیدگی محیط تجاری گردیده است، مدیران، سرمایه‌گذاران و بسیاری از فعالان بازار سرمایه همواره در جستجوی ابزارهایی به‌منظور پیش-

مدیریت) و عناصر بنیادی شرکت مانند سود، اقلام تعهدی جریان‌های نقدی، به‌کار برده‌اند (لی ۲۰۰۸؛ کوتاری و همکاران ۲۰۰۹؛ بال و همکاران ۲۰۱۵؛ فرانکل و همکاران ۲۰۱۶). هرچند تاریخچه طولانی از پژوهش‌هایی که اقلام تعهدی را بررسی و توصیف می‌کند وجود دارد؛ با این حال، توانایی تشریح اقلام تعهدی به مرحله کاملی نرسیده است و هنوز برای رسیدن به سطح آرمانی فاصله بسیاری است (هوبرگ، ۲۰۱۶). بررسی پژوهش‌های پیشین نشان می‌دهد که سنجش اقلام تعهدی اغلب از طریق افشاهای کمی و داده‌های ساختاریافته (مانند اطلاعات صورت‌های مالی اساسی) و با استفاده از مدل‌های رایج (دیچو و دیچف ۲۰۰۲، مدل تعدیل‌شده مک‌نیکولز ۲۰۰۲، بال و شیواکومار ۲۰۰۵) است. اما پژوهش‌های متکی بر افشاهای کیفی و ناساختاریافته، بسیار معدود و انگشت‌شمار است.

پژوهش حاضر از دیدگاه نظری، کاربردی و روش-شناسی دارای اهمیت است. استفاده‌کنندگان گزارش-های مالی، تحلیل‌گران اوراق بهادار، مدیران شرکت‌ها و سرمایه‌گذاران، اطلاعات مورد نیاز خود را از منابع مختلفی فراهم می‌کنند و از آنجایی که سود حسابداری و اجزای آن به‌عنوان برجسته‌ترین متغیرهای حسابداری نیز یکی از اطلاعات مهم است، بیشترین توجه خود را به رقم سود خالص به‌عنوان آخرین رقم اطلاعاتی صورت سود و زیان معطوف می‌دارند. سود حسابداری بر مبنای فرض تعهدی و در چارچوب اصول حسابداری اندازه‌گیری و ارائه می‌شود و بر قابلیت اتکای اقلام گزارش‌شده تاکید دارد. کیفیت اقلام تعهدی به علت تأثیری که بر درک سرمایه‌گذاران از کیفیت سود دارد، همواره مورد توجه مدیریت است به عبارت دیگر کیفیت اقلام تعهدی شاخصی برای سنجش کیفیت گزارشگری مالی است (فرانسیس و همکاران، ۲۰۰۵). هرچه کیفیت اقلام تعهدی بیشتر باشد، کیفیت سود نیز بیشتر خواهد بود، لذا اندازه‌گیری دقیق و بهتر اقلام تعهدی و سنجش کیفیت این اقلام با استفاده از تمامی گزارش‌های مالی کمی و کیفی و بهره‌گیری از

فریزر (۱۹۸۰) اولین مقاله را برای معرفی تحلیل محتوا برای ارزیابی کمی داده‌های روایتی بر اساس تکرار کلمات منحصربه‌فرد ارائه کردند. بعد از پژوهش فوق، مقالات حسابداری و مالی برای کمی کردن اطلاعات متنی، فعال شدند. اکثر پژوهشگران در حیطه حسابداری و حسابرسی، اغلب تاثیر اطلاعات کیفی را بر ارزش‌گذاری حقوق صاحبان سهام بررسی کرده‌اند. سرمایه‌گذاران بازار سهام، برای ارزشیابی‌های خود بیشتر از داده‌های کمی استفاده می‌کنند اما نظام حسابداری و مالی روش‌ها و فن‌آوری‌های جدید را همواره با آغوش باز می‌پذیرد (لوگران و مکدونالد، ۲۰۱۶). از آنجایی که متن‌کاوی یکی از حوزه‌های نوظهور در حسابداری و مالی است، الگوریتم طبقه‌بندی آن هنوز به خوبی بسط و توسعه نیافته است. دیدگاه‌های مختلفی در این حوزه وجود دارد. دیدگاه اول متعلق به امانی و فادالالا (۲۰۱۷) است که به منظور ارائه چارچوبی جامع در زمینه کاربرد داده-کاوی در حسابداری، زیرساخت‌های «حسابداری» و «داده‌کاوی» را به‌طور همزمان بررسی کردند. آنها با توجه به وظایف اصلی داده‌کاوی که شامل توصیف^{۱۱}، پیش‌بینی^{۱۲} و طبقه‌بندی^{۱۳} است به معرفی چارچوبی پرداختند که در آن، دو دیدگاه شناخته‌شده برای گزارشگری در «حسابداری» شامل دیدگاه گذشته-نگر^{۱۴} و دیدگاه آینده‌نگر^{۱۵} را با سه هدف شناخته‌شده برای «داده‌کاوی» شامل توصیف، پیش‌بینی و طبقه‌بندی، ترکیب کردند و شش حالت ترکیبی که فقط چهار حالت آن به‌طور منطقی امکان‌پذیر است شناسایی نمودند: (۱) داده‌کاوی توصیفی در گزارشگری گذشته‌نگر، (۲) داده‌کاوی طبقه‌بندی در گزارشگری گذشته‌نگر، (۳) داده‌کاوی طبقه‌بندی در گزارشگری آینده‌نگر، (۴) داده‌کاوی پیش‌بینی‌کننده در گزارشگری آینده‌نگر (رهروی، ۱۳۹۶). پژوهش حاضر از گزارش‌های هیات 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 اقلام تعهدی در شرکت‌ها دارند. بر همین اساس می‌توان پژوهش حاضر را در

بینی دقیق‌تر عملکرد مالی شرکت‌ها و حداکثر نمودن سود حاصل از سرمایه‌گذاری می‌باشند. بسیاری از وظایفی که متخصصان حسابداری انجام می‌دهند برای ایجاد حس پیش‌بینی در مشتریان و دنیای تجارت است (درو و تیسلیاک، ۲۰۱۹). داده‌کاوی فرآیند بررسی و تجزیه و تحلیل مقادیر عظیمی از داده‌ها جهت استخراج اطلاعات، دانش، کشف الگوهای پنهان، یادگیری الگوها و قوانین معنی‌دار نهفته در انبار داده است و در حال حاضر یکی از مهم‌ترین الگوهای موجود در حوزه هوش تجاری و ابزارهای تصمیم‌گیری است (هان و کامبر، ۲۰۱۲). با این که تکنیک‌های داده-کاوی را می‌توان بر روی داده‌های ناساختاریافته (مانند متون) نیمه‌ساختاریافته^{۱۶} (مانند اسناد) و ساختاریافته (مانند جداول و گزارش‌های عددی) اعمال نمود؛ اما اغلب داده‌کاوی از میان داده‌های ساختاریافته به استخراج اطلاعات می‌پردازد (داس، ۲۰۱۴). متن-کاوی یک حوزه بین‌رشته‌ای است که از مفاهیم داده-کاوی، آموزش ماشین، آمار و زبان برای تجزیه و تحلیل متون تهیه شده توسط انسان مانند مقالات، گزارش‌ها، کتاب‌ها، وبسایت‌ها و مواردی از این قبیل استفاده می‌کند. هدف نهایی از متن‌کاوی، استخراج اطلاعات باکیفیت از داخل متن است (رهروی، ۱۳۹۶). تفاوت متن‌کاوی و داده‌کاوی این است که داده‌کاوی بر روی داده‌های ساختاریافته پایگاه داده و متن‌کاوی، بر روی داده‌های ناساختاریافته و نیمه‌ساختاریافته مانند رایانامه و مستندات تمام متنی کار می‌کند (اسماعیلی، ۱۳۹۱). نسبت به روش‌های کمی که به‌صورت سنتی در حسابداری و مالی استفاده می‌شوند، تجزیه و تحلیل‌های متنی، از دقت کمتری نسبت به تحلیل‌های کمی برخوردارند، اما شناخت این روش به همان اندازه روش‌های کمی دارای اهمیت است. در حسابداری و مالی، دسترسی آنلاین به مقالات به‌روز و جدید، اسناد و مدارک منتشرشده از سوی شرکت‌ها و نهادهای قانون‌گذار، کنفرانس‌های خبری و متون موجود در شبکه‌های اجتماعی، خوراک زیادی برای استفاده از این فن‌آوری فراهم می‌کند (لوگران و مکدونالد، ۲۰۱۶). اینگرام و

ای با استفاده از فرهنگ واژگان تخصصی حسابداری (در پژوهش حاضر فرهنگ واژگان تخصصی حسابداری نوروش) به شیوه‌ای که در بخش دوم گام ۵ به تفصیل شرح گردیده محاسبه می‌شود. در ادامه مبانی نظری این دو تکنیک مورد بررسی قرار می‌گیرد. ماشین بردار پشتیبان یکی از فنون دسته‌بندی است که بر اساس تئوری یادگیری آماری پایه‌گذاری و معرفی شده است (وپنیک، ۱۹۹۸). این تکنیک به دنبال یافتن ابرصفحه‌ای است که با استفاده از آن داده‌ها را دسته‌بندی کند. این ابرصفحه به گونه‌ای عمل می‌کند که حاشیه دسته‌بندی را حداکثر نماید.

رگرسیون بردار پشتیبان، پژوهشگر را ملزم به تعیین شکل واژه‌نامه‌های از پیش تعیین شده، موضوعات و افشاهای طبقه‌بندی شده دستی نمی‌کند و به قضاوت پژوهشگران برای ارزیابی محتوای افشاهای متنی برای توصیف عناصر بنیادی شرکت متکی نیست. این تکنیک به پژوهشگران اجازه می‌دهد لغت‌نامه‌ای پویا از لغات کلیدی و وزن آن‌ها تعیین نمایند که با تغییر اوضاع اقتصادی در نوسان و تغییر باشند؛ لذا رگرسیون بردار پشتیبان دو نتیجه مهم برای تمام روش‌های آماری که به‌طور وسیع در پژوهش‌های حسابداری انجام می‌شوند به همراه دارد: اول این که می‌تواند برای انواعی از افشاهای، زبان‌ها و متون محتوایی با هزینه نسبی پایین به کار برده شود؛ دوم رگرسیون بردار پشتیبان می‌تواند در شرایطی به کار گرفته شود که پژوهشگران از قبل لغاتی را که با عناصر بنیادی قابل‌سنجش در ارتباط هستند، مشخص کرده باشند و می‌تواند به آسانی برای نمونه‌های بزرگ به کار گرفته شود (لی گو و همکاران، ۲۰۱۶).

رویکرد مبتنی بر فرهنگ لغت بر استخراج و تفسیر معانی از پیام‌ها و متون بر اساس روش‌های رایانه‌ای متمرکز است. کیف لغات^{۲۰} یکی از ساده‌ترین و عمومی‌ترین روش‌های این رویکرد است. این روش به‌طور وسیعی توسط پژوهشگران حسابداری و مالی مورد استفاده قرار گرفته است (تتلاک و همکاران ۲۰۰۷؛ مک‌اسکیسی ۲۰۰۸؛ کوتاری ۲۰۰۹؛ هانلی و همکاران ۲۰۱۰؛ کاهان و همکاران ۲۰۱۳؛ رانجان و

دسته «گذشته‌نگر-توصیفی» طبقه‌بندی نمود. دیدگاه دوم متعلق به فرانکل و همکاران (۲۰۱۶) است. آنها پژوهش‌هایی را که با استفاده از اطلاعات کیفی و تکنیک متن‌کاوی به تحلیل متون مالی و حسابداری می‌پردازند به سه گروه تقسیم کردند: گروه اول مطالعات، از فرهنگ واژگان برای ارزیابی خصوصیات افشاء، مانند عدم اطمینان، لحن^{۱۶} و رقابت، استفاده کرده‌اند (آلی و دی‌آنجلیس ۲۰۱۵؛ موسلو و همکاران ۲۰۱۴؛ لی و همکاران ۲۰۱۳؛ کراوت و موسلو ۲۰۱۳؛ روجرز و همکاران ۲۰۱۱). گروه دوم پژوهش‌ها به بررسی و سنجش ارتباط بین خوانایی^{۱۷} (قابلیت فهم)، میزان شباهت افشاء و طول متن افشاء شده با عناصر بنیادی شرکت پرداخته‌اند (لوگران و مکدونالد ۲۰۱۶؛ بنسال و همکاران ۲۰۱۵؛ پترسون و همکاران ۲۰۱۵؛ لی و همکاران ۲۰۱۴؛ و براون و تاکرا ۲۰۱۱). گروه سوم از پژوهشگران معتقد هستند که پژوهشگران می‌توانند برای کاهش شدت کار در حیطه‌های کاری و وظیفه‌ای خود، از تکنیک‌های پردازش زبان طبیعی که از سایر حوزه‌های دیگر وارد شده است استفاده کنند (بائو و داتا ۲۰۱۴؛ لانگ و لارنس ۲۰۱۵؛ کمیل و همکاران ۲۰۱۴؛ کراوت و موسلو ۲۰۱۳؛ لی ۲۰۱۰). در پژوهش حاضر برای فرضیه اول تکنیک ماشین بردار پشتیبان^{۱۸} استفاده شده و از این لحاظ، پژوهش در گروه سوم طبقه‌بندی فوق قرار می‌گیرد و برای فرضیه دوم از رویکرد مبتنی بر فرهنگ لغت^{۱۹} برای توصیف اقلام تعهدی استفاده شده است لذا در گروه اول قرار می‌گیرد. با استفاده از دو تکنیک متن‌کاوی اشاره شده، اقلام تعهدی متنی و اقلام تعهدی لغت‌نامه‌ای سنجش و در فرضیه‌های پژوهش برای تبیین اقلام تعهدی واقعی به کار گرفته می‌شوند. اقلام تعهدی واقعی که "اقلام تعهدی کل در سطح شرکت" نیز نامیده می‌شود بر اساس مدل سرمایه در گردش دیچو و دیچف (۲۰۰۲) و مدل اسلوان (۱۹۹۶) و مطابق رویکرد ترازنامه‌ای و از تفاوت دارایی‌های جاری غیرنقد و بدهی‌های جاری با توجه به جمع کل دارایی‌ها در ابتدای سال محاسبه می‌شود (فرانکل و همکاران، ۲۰۱۶) اقلام تعهدی لغت‌نامه-

هستون ۲۰۱۵) از آنجایی که رویکرد مبتنی بر واژه-نامه برای برنامه‌نویسی بسیار آسان است، لذا عمومی‌ترین رویکردی است که در پژوهش‌های حسابداری استفاده می‌شود. از نظر فنی، یک لغت‌نامه مجموعه‌ای از واژگان طبقه‌بندی شده‌ای است که با ویژگی‌ها و معیارهای مشخصی مرتبط است. بنابراین متن باید به فهرستی از لغات و کلماتی تبدیل شود که به‌واسطه آنها ویژگی‌های خاص و معینی از اسناد نشان داده شود. بر اساس همین فهرست، پژوهشگر می‌تواند لغات مرتبط با هر ویژگی و خاصه را شمارش کند و یک معیار مقایسه‌ای فراهم نماید (لی گو و همکاران، ۲۰۱۶).

۳- پیشینه پژوهش

پژوهش‌های مبتنی بر افشاهای کیفی و گزارش‌های ناساختاریافته که به بررسی اقلام تعهدی و ارتباط آن با عناصر بنیادی شرکت مانند سود، جریان‌های نقدی آتی و... می‌پردازند در دنیا بسیار کم و در ایران تاکنون انجام نشده است؛ در ایران پژوهش‌هایی در حیطه متن‌کاوی و به کارگیری تکنیک‌های آن انجام شده است که در ادامه به تعدادی از مربوط‌ترین و جدیدترین پژوهش‌های خارجی و داخلی مرتبط با این پژوهش اشاره می‌شود.

چو و همکاران (۲۰۱۸) در مقاله‌ای با عنوان "اندازه‌گیری ثبات اطلاعات کمی و کیفی در گزارش‌های مالی: یک رویکرد علمی" به بررسی همخوانی بین نسبت‌های مالی کمی و افشاهای روایتی کیفی در گزارش‌های سالانه با استفاده از یک رویکرد علمی پرداختند. آنها برای استخراج اطلاعات در مورد لحن داده‌های کیفی متنی ناساختاریافته، ابتدا از روش تکرار واژگان در متن‌کاوی استفاده کردند. سپس افشاهای روایتی هر شرکت با توجه به لحن به "مثبت" یا "منفی" طبقه‌بندی گردید. نتایج نشان داد که شرکت‌های آمریکایی نسبت به شرکت‌های چینی و تایوانی، کمتر در افشاهای روایتی و گزارش تفسیری مدیریت خود مبالغه می‌کنند.

چن و لی (۲۰۱۷) در پژوهشی با عنوان "تخمین میزان برآورد در اقلام تعهدی" نقش تخمین را در بخش تعهدی سود بررسی کردند. برای این کار، گزارش تفسیری مدیریت و یادداشت‌های همراه شرکت‌های بورس اوراق بهادار آمریکا را مورد بررسی قراردادند. آنها دریافتند، اقلام تعهدی که مستلزم ارزیابی بیشتری هستند، پایداری کمتری دارند و در پیش‌بینی جریان‌های نقدی آتی نقش کمتری ایفا می‌کنند.

آتورلی (۲۰۱۷) در مقاله‌ای با عنوان "مقایسه‌ای از به کارگیری تحلیل محتوا و متن‌کاوی در افشای گزارشگری مسوولیت‌های اجتماعی شرکت" دو تکنیک متن‌کاوی و تحلیل محتوا را در انجام تحلیل متنی گزارش‌های اجتماعی و محیطی شرکت‌ها، مقایسه نمود. بررسی وی نشان داد که هر دو روش تحلیل محتوا و متن‌کاوی می‌تواند به عنوان روش‌های کیفی برای حل مسایل پژوهشی در حیطه گزارشگری مسوولیت‌های اجتماعی شرکت‌ها به کار روند و مفید واقع شوند.

لو و همکاران (۲۰۱۷) در پژوهشی با عنوان "مدیریت سود و خوانایی (قابل فهم بودن) گزارش‌های مالی سالیانه" به بررسی تاثیر مدیریت سود بر خوانایی گزارش‌های مالی پرداختند. آنها برای اندازه‌گیری خوانایی گزارش‌های مالی سالیانه از شاخص فاگ استفاده نمودند و خوانایی اطلاعات مندرج در گزارش تفسیری مدیریت را در شرکت‌های بورس اوراق بهادار آمریکا اندازه‌گیری کردند. یافته‌های پژوهش نشان داد که شرکت‌ها برای مدیریت سود اقدام به پیچیده‌تر کردن گزارش‌های مالی می‌نمایند.

فرانکل و همکاران (۲۰۱۶) در پژوهشی با عنوان "استفاده از افشای کیفی و ناساختاریافته در توصیف اقلام تعهدی" به بررسی سودمندی رگرسیون‌های بردار پشتیبان در ارزیابی محتوای اطلاعاتی افشای کیفی و ناساختاریافته در تشریح اقلام تعهدی پرداختند. یافته‌های پژوهش نشان داد زمانی که برای تخمین اقلام تعهدی از اطلاعات مندرج در گزارش

دارابی و مشایخی (۱۳۹۵) در پژوهشی با عنوان "تاثیر هوش مالی در پیش‌بینی ریسک اعتباری با استفاده از مدل ماشین بردار پشتیبان" به بررسی تاثیر هوش مالی در پیش‌بینی ریسک اعتباری شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران پرداختند. نتایج حاصل از پژوهش نشان داد که هوش مالی توانایی پیش‌بینی ریسک اعتباری را دارد. به عبارت دیگر هوش مالی توانایی آن را دارد که با بهره‌گیری از مدل آماری ماشین بردار پشتیبان ریسک پرداخت تعهدات و ریسک سودآوری را به درستی پیش‌بینی کند اما در پیش‌بینی ریسک بهره‌وری ناتوان عمل می‌کند.

۴- فرضیه‌های پژوهش

این پژوهش دارای دو فرضیه به شرح زیر است:
فرضیه اول: اقلام تعهدی متنی (برآورد شده از متن گزارش هیات مدیره) دارای قدرت تفسیر و تبیین اقلام تعهدی واقعی (اقلام تعهدی سرمایه در گردش) در شرکت‌های پذیرفته‌شده بورس اوراق بهادار تهران است.

فرضیه دوم: اقلام تعهدی متنی دارای قدرت تفسیری بیشتری نسبت به اقلام تعهدی لغت‌نامه‌ای (برآورد شده براساس فرهنگ واژگان) جهت تبیین و توصیف اقلام تعهدی واقعی در شرکت‌های پذیرفته شده بورس اوراق بهادار تهران است.

۵- روش‌شناسی پژوهش

جامعه آماری پژوهش، شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران و قلمرو زمانی پژوهش از سال ۱۳۹۰ الی ۱۳۹۶ است. در این پژوهش برای انتخاب نمونه آماری از روش حذف سامانمند استفاده شد، بدین‌منظور کلیه شرکت‌هایی که اطلاعات و صورت‌های مالی، به ویژه، گزارش هیات مدیره آنها، به صورت کامل در دوره مورد بررسی در دسترس بودند، جزء شرکت‌های فعال بورس بوده یا حداقل در دوره زمانی مورد بررسی فعال باشند، پایان سال مالی آنها ۲۹ اسفند هر سال باشد و جزء شرکت‌های سرمایه-

تفسیری مدیریت و تکنیک‌های متن‌کاوی استفاده شود، بهتر می‌توان اقلام تعهدی شرکت را برآورد نمود. هنری و لئون (۲۰۱۶) در پژوهشی با عنوان "اندازه‌گیری اطلاعات کیفی در پژوهش‌های بازار سرمایه: مقایسه روش‌شناسی‌های جایگزین برای اندازه‌گیری لحن افشاء" به بررسی و ارزیابی شاخص‌های جایگزین لحن افشاء مالی پرداختند. آنها از تکنیک متن‌کاوی فراوانی لغت و فراوانی متن و تکنیک دسته‌بندی مبتنی بر نظریه بیز جهت سنجش لحن افشاء استفاده نمودند. نتایج نشان داد که لحن افشای اندازه‌گیری شده با استفاده از تکنیک فراوانی کلمه دارای محتوای اطلاعاتی همانند تکنیک دسته‌بندی مبتنی بر نظریه بیز است.

ظفری و همکاران (۱۳۹۸) در پژوهشی با عنوان "تاثیر قابلیت مقایسه و ثبات رویه حسابداری بر عدم تقارن اطلاعاتی: رویکرد متن‌کاوی" به بررسی تاثیر قابلیت مقایسه و ثبات رویه حسابداری بر عدم تقارن اطلاعاتی در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران با استفاده از تکنیک متن‌کاوی پرداختند. نتایج و یافته‌های پژوهش نشان داد که قابلیت مقایسه حسابداری بر عدم تقارن اطلاعاتی تاثیر منفی دارد و با افزایش ثبات رویه شرکت، عدم تقارن اطلاعاتی کاهش می‌یابد.

رهروی دستجردی (۱۳۹۶) در رساله دکترای خود با عنوان "برآورد میزان احتمال خطر تقلب در گزارشگری مالی از طریق تجزیه و تحلیل متن و تطبیق انگیزه‌های آن با الگوی نظریه چشم‌انداز تجمعی" به بررسی وجود تاثیر و شیوه‌های تاثیرگذاری انحراف از عملکرد مرجع بر انگیزه‌های ارتکاب به تقلب مدیران و تطبیق آن با الگوی نظریه چشم‌انداز طبیعی بین شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران پرداخت. به همین منظور با تمرکز بر تجزیه و تحلیل متن گزارش‌های هیات مدیره از سه روش مبتنی بر داده‌کاوی شامل درخت تصمیم، ماشین بردار پشتیبان و بهینه‌سازی محدب استفاده نمود و شاخصی برای ارزیابی و تشخیص خطر تقلب در گزارش‌های هیات مدیره به دست آورد.

گزارش‌ها پیش‌پردازش گردید؛ یعنی عبارت‌های توقف^{۲۲} یا همان کلمات ایست(کلمات ایست لغاتی هستند که در جملات کاربرد ربطی دارند و زیاد استعمال می‌شوند مثل «گر»، «و»، «با»، این کلمات برخلاف این که بسیار استفاده می‌شوند اما از لحاظ معنایی دارای اهمیت کمی بوده و به همین دلیل اغلب در فعالیت‌های مربوط به حوزه پردازش زبان طبیعی در فاز پیش‌پردازش حذف می‌شوند. همچنین عمل ریشه‌یابی لغات^{۲۳} نیز انجام شد به گونه‌ای که کلماتی که دارای ریشه مشابهی هستند به عنوان یک لغت در نظر گرفته شد. به عنوان مثال دو عبارت "فروش" و "فروش‌ها" به عنوان یک عبارت مشابه تلقی گردید. پس از فرآیند پیش‌پردازش متن، مجموعه‌ای از داده‌ها به صورت سال-شرکت که شامل تعداد کل عبارت‌های یک و دو کلمه‌ای در گزارش هیات مدیره هر شرکت است، ایجاد گردید. تعداد کل کلمات و عبارات، با جمع زدن تعداد کلمه و عبارت در هر گزارش هیات مدیره محاسبه گردید. از آنجا که تعداد مشاهدات کمتر از تعداد کلمات و عبارات منحصر به فرد است، نمی‌توان از رگرسیون حداقل مربعات معمولی استفاده کرد. این مجموعه داده برای توضیح اقلام تعهدی(واقعی) با استفاده از رگرسیون بردار پشتیبان به کار گرفته شد؛ که در آن اقلام تعهدی واقعی همان اقلام تعهدی سرمایه در گردش، بر اساس مدل دیچو و دیچف(۲۰۰۲) است. در ادامه، با استفاده از اقلام تعهدی سرمایه در گردش(اقلام تعهدی واقعی) که در این‌جا متغیر وابسته است و تعداد کل کلمات و عبارات یک و دو کلمه‌ای در گزارش هیات مدیره که در این‌جا متغیر مستقل است، رگرسیون‌های بردار پشتیبان تخمین زده شد و اقلام تعهدی متنی محاسبه گردید.

گام دوم: سنجش توانایی گزارش هیات مدیره شرکت در توضیح اقلام تعهدی واقعی

این گام دو فرضیه دارد که در ادامه الگوی هر فرضیه و متغیرهای مربوط به آن شرح داده خواهد شد. جهت آزمون فرضیه اول پژوهش از مدل رگرسیون داده‌های ترکیبی شماره (۱) برگرفته از

گذاری، واسطه‌گری مالی و لیزینگ نباشند؛ به عنوان نمونه انتخاب و بقیه حذف شدند. با استفاده از روش فوق، تعداد ۲۰۵ شرکت به عنوان نمونه مورد مطالعه انتخاب شد. برای جمع‌آوری داده‌های موردنیاز جهت آزمون فرضیه‌ها از روش اسنادکاوی استفاده شد و داده‌های مالی شرکت‌ها برای تجزیه و تحلیل‌های رگرسیونی از نرم‌افزار «رهاورد نوین» استخراج و گزارش‌های هیات مدیره به مجمع عمومی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران نیز از «سامانه جامع اطلاع‌رسانی ناشران» و «مرکز پژوهش، توسعه و مطالعات اسلامی سازمان بورس اوراق بهادار تهران» دانلود و بررسی شده است. در این پژوهش از ابزارهای مختلفی برای تجزیه و تحلیل استفاده گردید. به منظور ویرایش‌های متنی، تصویری محتویات فایل‌ها، استخراج متن و تصاویر درون فایل‌ها و آماده‌سازی فایل‌های پی‌دی‌اف گزارش هیات مدیره که اغلب به-درستی تهیه و تنظیم نشده بودند از نرم‌افزار نیترو پی‌دی‌اف و برای تبدیل فایل پی‌دی‌اف گزارش‌های هیات مدیره به مجمع عمومی صاحبان سهام به ورد از سایت نرم‌افزاری ایبو^{۲۱} استفاده شد. تجزیه و تحلیل و آزمون فرضیه‌های پژوهش، به همراه نتایج آن نیز با توجه به خروجی‌های به دست آمده با استفاده از نرم‌افزار استنا انجام گرفت. متن‌کاوی گزارش‌های هیات مدیره، با استفاده از تکنیک ماشین‌های بردار پشتیبان با نرم‌افزار پایتون انجام گرفت. با توجه به هدف کلی پژوهش که سنجش توانایی افشاهای کیفی و ناساختاریافته در توصیف اقلام تعهدی با کمک روش‌های متن‌کاوی است؛ این پژوهش دربرگیرنده دو گام اصلی بود. گام اول محاسبه اقلام تعهدی متنی گزارش هیات مدیره، با استفاده از رگرسیون بردار پشتیبان و گام دوم سنجش توانایی این گزارش در تبیین اقلام تعهدی، از طریق آزمون فرضیه‌های اول و دوم است.

گام اول: محاسبه اقلام تعهدی متنی با تکنیک رگرسیون بردار پشتیبان

ابتدا گزارش هیات مدیره ۲۰۵ شرکت از طریق نرم‌افزار موجود در سایت ایبو از قالب پی‌دی‌اف و عکس به فرمت ورد تبدیل شد. در مرحله بعد متن

گردید. به‌طور خاص با استفاده از مشاهدات سال قبل، یک رگرسیون حداقل مربعات معمولی از اقلام تعهدی روی تعداد لغات حسابداری مرتبط، تخمین زده شد. سپس ضرایب تخمینی از این برآورد برای شمارش-های مربوط به خودشان در سال t به‌کار رفت. تخمین ایجاد شده از اقلام تعهدی به این روش، "اقلام تعهدی لغت‌نامه‌ای" (استخراج شده از فرهنگ لغات حسابداری) نامیده می‌شود. برای ایجاد مدل مربوط به آزمون فرضیه دوم، به جای اقلام تعهدی متنی در مدل (۱) اقلام تعهدی لغت‌نامه‌ای اضافه و مدل (۲) ایجاد می‌گردد.

مدل (۲)

$$\text{Accruals}_{i,t} = \alpha_0 + \alpha_1 \text{Dict MDA} \text{ Accruals}_{i,t} + \alpha_2 \text{IND} + \alpha_3 \text{YEAR} + \epsilon_{i,t}$$

سپس برای آزمون فرضیه دوم به مدل (۲) اقلام تعهدی متنی برگرفته از گزارش هیات مدیره نیز اضافه و مدل (۳) به شرح زیر به دست آمد:

مدل (۳)

$$\text{Accruals}_{i,t} = \alpha_0 + \alpha_1 \text{MDA} \text{ Accruals}_{i,t} + \alpha_2 \text{Dict MDA} \text{ Accruals}_{i,t} + \alpha_3 \text{IND} + \alpha_4 \text{YEAR} + \epsilon_{i,t}$$

اقلام تعهدی واقعی شرکت t : $\text{Accruals}_{i,t}$
 اقلام تعهدی متنی شرکت t : $\text{MDA} \text{ Accruals}_{i,t}$
 اقلام تعهدی متنی لغت‌نامه-ای شرکت t : $\text{Dict MDA} \text{ Accruals}_{i,t}$

IND: متغیر مجازی برای کنترل اثر صنعت

YEAR: متغیر مجازی برای کنترل اثر سال

$\epsilon_{i,t}$: پسماند (جزء خطای مدل)

۶- متغیرهای پژوهش

متغیرهای پژوهش شامل سه دسته متغیرهای وابسته، مستقل و مجازی است:

۶-۱- متغیر وابسته

متغیر وابسته این پژوهش، اقلام تعهدی واقعی است که "اقلام تعهدی کل در سطح شرکت" نیز نامیده می‌شود. براساس مدل سرمایه در گردش دیچو و دیچف (۲۰۰۲) و مدل اسلون (۱۹۹۶) و مطابق رویکرد ترازنامه‌ای و از تفاوت دارایی‌های جاری غیرنقد

پژوهش فرانکل و همکاران (۲۰۱۶)، استفاده شده است همان‌طور که در گام یک توضیح داده شد اقلام تعهدی متنی (متغیر مستقل) در این مدل با کمک تکنیک رگرسیون بردار پشتیبان محاسبه شده و اقلام تعهدی واقعی (متغیر وابسته) بر اساس مدل دیچو و دیچف (۲۰۰۲) و مدل اسلون (۱۹۹۶) و مطابق با رویکرد ترازنامه‌ای محاسبه می‌شود:

مدل (۱)

$$\text{Accruals}_{i,t} = \alpha_0 + \alpha_1 \text{MDA} \text{ Accruals}_{i,t} + \alpha_2 \text{IND} + \alpha_3 \text{YEAR} + \epsilon_{i,t}$$

اقلام تعهدی واقعی شرکت t : $\text{Accruals}_{i,t}$

اقلام تعهدی متنی شرکت t : $\text{MDA} \text{ Accruals}_{i,t}$

IND: متغیر مجازی برای کنترل اثر صنعت

YEAR: متغیر مجازی برای کنترل اثر سال

$\epsilon_{i,t}$: پسماند (جزء خطای مدل)

فرضیه دوم پژوهش، قدرت تبیین اضافی تر اقلام تعهدی متنی را نسبت به روش لغت‌نامه‌ای بررسی می‌کند؛ به همین منظور برآورد دیگری از اقلام تعهدی با استفاده از تعدادی از لغات و اصطلاحات مرتبط با اقلام تعهدی که از فرهنگ لغت حسابداری استخراج شده است انجام می‌گردد. روش این فرضیه برای سنجش اقلام تعهدی لغت‌نامه‌ای از الگوی لغت‌نامه‌ای پژوهش‌های فرانکل و همکاران (۲۰۱۶)، آلی و دی‌آنجلیس (۲۰۱۵)، لوگران و مکدونالد (۲۰۱۴) برگرفته شده است. برای پژوهش حاضر از لغت‌نامه تخصصی واژگان حسابداری نوروش (۱۳۹۴) که دارای ۵۳۹۲ واژه است استفاده شد. بسیاری از واژگان لغت-نامه تخصصی حسابداری در گزارش‌های هیات مدیره شرکت‌ها به‌کار گرفته نشده‌اند یا بسیار کم تکرار بوده-اند؛ لذا با استفاده از ۲۷۱۹ واژه پرتکرار در گزارش هیات مدیره شرکت‌ها به مجامع که توسط نرم‌افزار پایتون از طریق متن‌کاوی به دست آمده است ۱۳۹۷ واژه تخصصی لغت‌نامه که بیشترین تکرار را در گزارش‌های مذکور داشتند انتخاب گردید و در رویکرد مبتنی بر لغت‌نامه از آن استفاده شد. سپس با روش-های مشابهی که قبلاً برای تخمین اقلام تعهدی متنی استفاده شده است، تخمینی برای اقلام تعهدی انجام

و بدهی‌های جاری با توجه به جمع کل دارایی‌ها در ابتدای سال محاسبه می‌شود (فرانکل و همکاران، ۲۰۱۶).

$$\text{Accruals}_{i,t} = (\Delta \text{CA}_{i,t} - \Delta \text{Cash}_{i,t}) - (\Delta \text{CL}_{i,t} - \Delta \text{STD}_{i,t}) + \Delta \text{TAX}_{i,t} - \text{Dep}_{i,t}$$

$\text{Accruals}_{i,t}$: اقلام تعهدی واقعی شرکت در سال t
 $\Delta \text{CA}_{i,t}$: تغییرات دارایی‌های جاری شرکت در سال جاری نسبت به سال قبل
 $\Delta \text{Cash}_{i,t}$: تغییرات وجوه نقد شرکت در سال جاری نسبت به سال قبل
 $\Delta \text{CL}_{i,t}$: تغییرات بدهی‌های جاری شرکت در سال جاری نسبت به سال قبل
 $\Delta \text{STD}_{i,t}$: تغییرات بدهی‌های بلندمدت شرکت در سال جاری نسبت به سال قبل
 $\Delta \text{TAX}_{i,t}$: تغییرات مالیات‌های پرداختی شرکت در سال جاری نسبت به سال قبل
 $\text{Dep}_{i,t}$: هزینه استهلاک شرکت در سال t

۶-۲- متغیر مستقل

متغیرهای مستقل پژوهش براساس پژوهش فرانکل و همکاران (۲۰۱۶) شامل اقلام تعهدی متنی و اقلام تعهدی لغت‌نامه‌ای استخراج شده از فرهنگ لغات حسابداری است:

الف) اقلام تعهدی متنی: همان‌طور که پیش‌تر در گام اول بخش ۵ اشاره شد عبارتست از میانگین ارزش پیش‌بینی شده حاصل از دو تخمین جداگانه (صنعت و سال) که با کمک رگرسیون بردار پشتیبان محاسبه می‌شود. رگرسیون بردار پشتیبان اقلام تعهدی واقعی را به عنوان متغیر وابسته در مدل رگرسیون وارد کرده، تعداد کل لغات یک کلمه‌ای و دو کلمه‌ای مستخرج از گزارش هیات مدیره را به عنوان متغیرهای مستقل در نظر می‌گیرد. رگرسیون بردار پشتیبان ضرایب تخمینی را روی متغیرهای مستقل در سرتاسر زیرپنجره‌های تخمین از سال‌های قبل تا سال t به کار می‌گیرد سپس ضرایب تخمینی را برای ارزش‌های سال t به منظور استخراج ارزش‌های پیش‌بینی شده

خارج از نمونه برای محاسبه اقلام تعهدی متنی در سال t به کار می‌بندد (فرانکل و همکاران، ۲۰۱۶: ۳۸).
ب) اقلام تعهدی لغت‌نامه‌ای: با توجه به موارد مطرح شده در گام دوم بخش ۵ که به تفصیل در بالا به آن اشاره شده؛ اقلام تعهدی لغت‌نامه‌ای ارزشی از اقلام تعهدی حاصل از به‌کارگیری رویکرد لغت‌نامه‌ای است. با استفاده از مشاهدات سال قبل، ابتدا رگرسیون حداقل مربعات معمولی از اقلام تعهدی روی تعداد لغات مرتبط با حسابداری که از فرهنگ اصطلاحات حسابداری برگرفته شده است و در متن گزارش هیات مدیره وجود دارد، تخمین زده می‌شود. سپس ضرایب تخمین‌زده شده از این برآورد را برای تعداد لغات مربوط به خودشان در سال t به کار گرفته خواهد شد، تا ارزش خارج از نمونه تخمینی، برای اقلام تعهدی در سال t محاسبه شود (فرانکل و همکاران، ۲۰۱۶: ۳۸).

۶-۳- متغیر مجازی

بر اساس پژوهش فرانکل و همکاران (۲۰۱۶)، دیچو و دیچف (۲۰۰۲)، چاوداری (۲۰۰۶) جهت تصریح مدل‌ها رگرسیونی پژوهش؛ متغیرهای مجازی به شرح ذیل در مدل‌های وارد می‌شود:

الف) صنعت: در هر صنعتی نوع فعالیت، دارایی‌ها، اندازه شرکت‌ها، اندازه سرمایه، نوع مشتری و سودآوری شرکت‌ها متفاوت است و به دلیل سیاست‌ها و خط‌مشی‌های دولتی، نوع صنعت نیز می‌تواند متغیری اثرگذار بر اقلام تعهدی شرکت باشد و چون شرکت‌ها از نظر کیفیت سود، نحوه دست‌کاری سود و حجم اقلام تعهدی در صنایع مختلف، به دلیل نوع فعالیت آنها کاملاً با یکدیگر متفاوت هستند، به منظور افزایش قابلیت مقایسه داده‌ها و اندازه‌گیری صحیح مقادیر خطا و جهت تصریح بهتر مدل رگرسیونی مورد استفاده پژوهش؛ متغیر صنعت به عنوان متغیر مجازی برای کنترل اثرات صنعت وارد مدل گردید (فرانکل و همکاران، ۲۰۱۶).

ب) سال: از آن‌جا که در طول سال‌ها اتفاقات سیاسی و اقتصادی متفاوتی روی می‌دهد، در نظر نگرفتن متغیر سال و تصور ثابت بودن شرایط در طول

دوره زمانی پژوهش منجر به نتایج غلط خواهد شد. بر اساس پژوهش فرانکل و همکاران (۲۰۱۶)، دیچو و دیچف (۲۰۰۲)، چاوداری (۲۰۰۶) متغیر سال به عنوان متغیر مجازی برای کنترل اثرات دوره زمانی وارد مدل گردید (فرانکل و همکاران، ۲۰۱۶).

۷- تجزیه تحلیل داده‌های پژوهش

در جدول (۱) آماره‌های توصیفی مربوط به متغیرهای پژوهش، شامل شاخص‌های مرکزی و پراکندگی، نشان داده شده است. اصلی‌ترین شاخص مرکزی، میانگین است. مقدار میانگین متغیر «اقلام تعهدی واقعی» برابر با ۰/۰۲۱- است؛ یعنی بیش‌تر داده‌ها حول این نقطه تمرکز یافته‌اند. از مهم‌ترین شاخص‌های پراکندگی، انحراف معیار است. مقدار این شاخص برای متغیر «هزینه استهلاک» برابر با ۰/۱۰۲ و برای متغیر «تغییرات مالیات پرداختنی» برابر با ۱/۱۰۷ است؛ یعنی در بین متغیرهای پژوهش، این دو متغیر به ترتیب دارای کم‌ترین و بیش‌ترین میزان پراکندگی است. علاوه بر این، بررسی سطح چولگی و کشیدگی متغیرهای پژوهش نشان می‌دهد که متغیرهای پژوهش از توزیع نزدیک به نرمالی برخوردار هستند. در این پژوهش جهت بررسی نرمال بودن توزیع متغیرها، از آزمون شاپیرو-ویلک استفاده شد. نتایج حاصل از این آزمون نشان داد که سطح معناداری به دست آمده برای کلیه متغیرها، کوچک‌تر

از سطح خطای موردانتظار ۰/۰۵ است، لذا فرض صفر آزمون شاپیرو-ویلک مبنی بر نرمال بودن متغیرها رد می‌شود؛ اما با توجه به قضیه حد مرکزی، چنانچه تعداد مشاهدات در نمونه مورد بررسی به میزان کافی زیاد باشد، ضرورت ندارد که متغیرها دارای توزیع نرمال باشند. تعداد مشاهدات در پژوهش حاضر به میزان کافی بزرگ است (تعداد ۱۴۳۵ مشاهده)، لذا ضرورتی جهت نرمال‌سازی متغیرهای پژوهش وجود ندارد. برای کشف همخطی بین متغیرهای توضیحی از معیار عامل تورم واریانس (VIF) استفاده شده است. نتایج نشان داد مقادیر عامل تورم واریانس در همه مدل‌های مربوط به فرضیه‌های پژوهش، کمتر از ۱۰ است، لذا میزان همخطی در بین متغیرهای توضیحی، مطلوب است.

در این پژوهش به منظور بررسی خودهمبستگی سریالی بین باقیمانده‌ها، از آزمون وولدریج استفاده شده است و از آنجا که سطح معناداری به دست آمده آماره آزمون در کلیه مدل‌های مربوط به آزمون فرضیه‌های پژوهش از سطح خطای موردانتظار ۰/۰۵ کوچک‌تر است، فرض صفر این آزمون مبنی بر عدم وجود خودهمبستگی بین باقیمانده‌ها رد می‌شود.

به منظور رفع مشکل خودهمبستگی سریالی بین باقیمانده‌ها، از روش رگرسیون پرایس-وینستن جهت آزمون فرضیه‌های پژوهش، استفاده شده است.

جدول (۱) آماره‌های توصیفی متغیرهای پژوهش

نام متغیر	تعداد	میانگین	انحراف معیار	کمینه	بیشینه	چولگی	کشیدگی
اقلام تعهدی واقعی	۱۴۳۵	-۰/۰۲۱	۰/۳۰۷	-۳/۲۱۸	۱/۹۱۶	-۲/۴۱۷	۲۴/۸۶۵
اقلام تعهدی متنی	۱۴۳۵	-۰/۰۱۵	۰/۲۵۸	-۱/۰۱۵	۰/۹۸۴	-۰/۵۳۴	۶/۲۷۲
تغییرات حسابهای دریافتنی	۱۴۳۵	۰/۳۵۷	۰/۹۶۵	-۰/۹۸۸	۴/۹۰۳	۳/۳۲۶	۱۵/۴۸۸
تغییرات موجودی‌ها	۱۴۳۵	۰/۲۷۸	۰/۶۶۹	-۰/۹۸۲	۲/۸۶۴	۲/۴۳۶	۹/۸۰۸
تغییرات حسابهای پرداختنی	۱۴۳۵	۰/۲۲۲	۰/۴۳۳	-۰/۸۶۵	۰/۹۷۸	۰/۱۹۸	۲/۴۵۳
تغییرات مالیات پرداختنی	۱۴۳۵	۰/۴۰۱	۱/۱۰۷	-۰/۹۷۸	۵/۱۸۶	۲/۰۸۴	۹/۰۴۰
هزینه استهلاک	۱۴۳۵	۰/۱۱۰	۰/۱۰۲	۰/۰۰۰	۰/۹۲۴	۴/۰۶۸	۲۹/۶۵۸
اقلام تعهدی لغت‌نامه‌ای	۱۴۳۵	-۰/۰۲۰	۰/۲۲۸	-۱/۰۱۹	۰/۹۸۰	-۰/۷۰۸	۸/۴۶۳

منبع: یافته‌های پژوهش

نتایج حاصل آزمون فرضیه اول پژوهش در جدول (۲) ارائه شده است.

در فرضیه اول پژوهش، نقش اقلام تعهدی متنی در تبیین اقلام تعهدی واقعی مورد بررسی قرار گرفت. با توجه به نتایج جدول (۲)، سطح معناداری به دست آمده برای آماره z متغیر «اقلام تعهدی متنی» برابر با ۰/۰۰۰ و کوچکتر از سطح خطای مورد انتظار ۰/۰۵ است؛ و حاکی از این موضوع است که اقلام تعهدی متنی تأثیر معناداری بر اقلام تعهدی واقعی شرکت‌های مورد مطالعه دارد.

لذا با توجه به ضریب مثبت به دست آمده برای این متغیر (۰/۲۸۵) و ضریب تعیین حاصل شده، این گونه نتیجه‌گیری می‌شود که ۹/۶ درصد از تغییرات اقلام تعهدی واقعی توسط اقلام تعهدی متنی قابل توصیف و تبیین است لذا فرضیه اول تأیید می‌گردد. در فرضیه دوم پژوهش، قدرت تبیین‌کنندگی اقلام تعهدی متنی و اقلام تعهدی لغت‌نامه‌ای در توصیف اقلام تعهدی واقعی مورد مقایسه واقع شده است.

به منظور بررسی ناهمسانی واریانس باقیمانده‌ها، از آزمون والد تعدیل شده استفاده گردید. نتایج حاکی از وجود مشکل ناهمسانی واریانس در مدل‌های مذکور بود و جهت رفع مشکل ناهمسانی واریانس، از روش رگرسیون پرایس-وینستن برای آزمون فرضیه‌های پژوهش، استفاده شده است. برای بررسی وابستگی مقطعی بین باقیمانده‌ها، از آزمون پسران استفاده شده است و چون سطح معناداری به دست آمده برای آماره آزمون پسران در کلیه مدل‌های پژوهش، بزرگتر از سطح خطای مورد انتظار ۰/۰۵ بود، عدم وجود همبستگی مقطعی بین باقیمانده‌ها تأیید گردید. از آنجایی که با کنترل اثرات سال و صنعت، دیگر نیازی به اجرای آزمون‌های انتخاب الگوی مناسب برآورد (شامل چاو، بروش پاگان، هاسمن) نیست و اجرای این آزمون‌ها نتایج متناقضی ارائه می‌کند (بنی-مهد و همکاران ۱۳۹۶؛ افلاطونی ۱۳۹۵؛ فرانکل و همکاران ۲۰۱۶) لذا در این پژوهش جهت مرتب نمودن داده‌ها در ساختار ترکیبی، روش داده‌های پانل (تابلویی) با اثرات ثابت مورد استفاده قرار گرفت

جدول (۲) نتایج آزمون فرضیه‌ی اول پژوهش

Accruals _{i,t} = α ₀ + α ₁ MDA Accruals _{i,t} + α ₂ IND + α ₃ YEAR + ε _{i,t}				
متغیر وابسته: اقلام تعهدی واقعی				
روش برآورد: رگرسیون پرایس-وینستن (جهت رفع مشکل ناهمسانی واریانس و خودهمبستگی سریالی)				
نام متغیر	ضریب	انحراف معیار	آماره z	سطح معناداری
عرض از مبدأ	۰/۳۹۸	۰/۷۱۹	۰/۵۵	۰/۵۷۹
اقلام تعهدی متنی	۰/۲۸۵	۰/۴۴۹	۶/۳۵	۰/۰۰۰
کنترل اثرات صنعت:	بلی	کنترل اثرات سال:		
ضریب تعیین تعدیل شده:	۰/۱۹۶	آماره والد:		
		سطح معناداری:		
		۰/۰۰۰		

منبع: یافته‌های پژوهش

جدول (۳) نتایج آزمون مدل پایه فرضیه دوم پژوهش

Accruals _{i,t} = α ₀ + α ₁ Dict MDA Accruals _{i,t} + α ₂ IND + α ₃ YEAR + ε _{i,t}				
متغیر وابسته: اقلام تعهدی واقعی				
روش برآورد: رگرسیون پرایس-وینستن (جهت رفع مشکل ناهمسانی واریانس و خودهمبستگی سریالی)				
نام متغیر	ضریب	انحراف معیار	آماره z	سطح معناداری
عرض از مبدأ	-۰/۰۱۹	۰/۰۶۵	-۰/۲۹	۰/۷۶۸
اقلام تعهدی لغت نامه ای	۰/۳۱۲	۰/۰۵۶	۵/۵۶	۰/۰۰۰
کنترل اثرات صنعت:	بلی	کنترل اثرات سال:		
ضریب تعیین تعدیل شده:	۰/۱۵۳	آماره والد:		
		سطح معناداری:		
		۰/۰۰۰		

منبع: یافته‌های پژوهش

آن در جدول (۴) نشان داده شده است. با وارد شدن اقلام تعهدی متنی در مدل اولی (مدل پایه)، ضریب تعیین تعدیل‌شده به عدد ۰/۵۳ ارتقاء پیدا نمود؛ یعنی اقلام تعهدی متنی و اقلام تعهدی لغت‌نامه‌ای توما قادر به تبیین ۵۳ درصد از تغییرات اقلام تعهدی واقعی می‌باشند.

لذا با عنایت به نتایج آزمون ریست رمزی ارائه شده در جدول (۵)، چون ضریب تعیین تعدیل‌شده مدل دوم بزرگ‌تر از ضریب تعیین تعدیل‌شده مدل پایه است، این‌گونه نتیجه‌گیری می‌شود که به دلیل بزرگ‌تر بودن سطح معناداری آماره $F(۰/۳۷۸)$ از سطح خطای مورد انتظار (۰/۰۵) فرضیه صفر این آزمون مبنی بر این که متغیر اضافه شده به مدل (اقلام تعهدی متنی) تاثیر معناداری روی متغیر وابسته ندارد رد می‌گردد و این‌گونه نتیجه‌گیری می‌شود که مدل دوم (با اضافه شدن اقلام تعهدی متنی) قدرت تبیین-کنندگی بالاتری در توصیف اقلام تعهدی واقعی دارد.

در وهله اول، قدرت تبیین اقلام تعهدی لغت‌نامه-ای در توصیف اقلام تعهدی واقعی بررسی شد. نتایج این آزمون در جدول (۳) نشان داده است که سطح معناداری آماره والد، برابر با ۰/۰۰۰ است یعنی مدل رگرسیونی به‌طور مناسب برازش شده است. از طرفی ضریب تعیین تعدیل‌شده به دست آمده گویای این موضوع است که اقلام تعهدی لغت‌نامه‌ای قادر به تبیین ۱۵/۳ درصد از تغییرات اقلام تعهدی واقعی هستند، درحالی‌که نتیجه آزمون فرضیه اول پژوهش نشان داد که ۱۹/۶ درصد از تغییرات اقلام تعهدی واقعی توسط اقلام تعهدی متنی قابل تبیین است. لذا اقلام تعهدی متنی دارای قدرت تفسیری بیشتری نسبت به اقلام تعهدی لغت‌نامه‌ای جهت توصیف اقلام تعهدی واقعی هستند.

در ادامه با وارد نمودن اقلام تعهدی متنی در مدل بالا، نقش اقلام تعهدی لغت‌نامه‌ای و متنی به‌منظور تبیین اقلام تعهدی واقعی، توما بررسی شد که نتایج

جدول (۴) نتایج آزمون مدل نهایی فرضیه دوم پژوهش

Accruals _{i,t} = $\alpha_0 + \alpha_1 MDA Accruals_{i,t} + \alpha_2 Dict MDA Accruals_{i,t} + \alpha_3 IND + \alpha_4 YEAR + \epsilon_{i,t}$				
متغیر وابسته: اقلام تعهدی واقعی				
روش برآورد: رگرسیون پرایس-وینستن (جهت رفع مشکل ناهمسانی واریانس و خودهمبستگی سریالی)				
نام متغیر	ضریب	انحراف معیار	آماره z	سطح معناداری
عرض از مبدأ	-۰/۰۰۴	۰/۰۵۹	-۰/۰۶	۰/۹۴۸
اقلام تعهدی متنی	۰/۵۶۴	۰/۰۹۵	۵/۹۴	۰/۰۰۰
اقلام تعهدی لغت‌نامه‌ای	۰/۲۳۰	۰/۰۴۸	۴/۷۹	۰/۰۰۰
کنترل اثرات صنعت:	بلی	کنترل اثرات سال:		
ضریب تعیین تعدیل‌شده:	۰/۵۳	آماره والد		
		سطح معناداری:		
		۳۰/۱۱۴		
		۰/۰۰۰۰		

منبع: یافته‌های پژوهش

جدول (۵) نتایج آزمون ریست رمزی در فرضیه دوم پژوهش

نام آزمون	آماره F	سطح معناداری
ریست رمزی	۰/۷۸	۰/۳۷۸

منبع: یافته‌های پژوهش

۸- نتیجه‌گیری و بحث

پژوهش‌های بسیاری درباره اقلام تعهدی با تکیه بر گزارش‌ها و داده‌های ساختاریافته و افشاهای کمی انجام گرفته است. اما پژوهش حاضر با سنجش اقلام

تعهدی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران، سعی کرده است جهت اندازه‌گیری اقلام تعهدی از اعداد و ارقام صورت‌های مالی فاصله گرفته و برای بخش عمده‌ای از تحلیل‌ها بر گزارهای هیات مدیره به

به تنهایی ارقام تعهدی لغت‌نامه ای در توصیف ارقام تعهدی واقعی مورد بررسی واقع گردید ضریب تعیین ۱۵/۳ درصد بود که در مقایسه با ضریب تعیین به دست آمده از آزمون فرضیه اول که نشان داد ۱۹/۶ درصد از تغییرات ارقام تعهدی واقعی توسط ارقام تعهدی متنی قابل توصیف و تبیین است؛ می‌توان نتیجه گرفت که تکنیک رگرسیون بردار پشتیبان مزیتی بیش از یک فرهنگ لغت حسابداری دارد و برتر است؛ ثانیاً قدرت تفسیر ارقام تعهدی متنی برای توصیف سطوح ارقام تعهدی در سطح شرکت، مستقیماً از عبارات و لغاتی که در متن گزارش هیات مدیره به مجمع شرکت‌ها، ناشی می‌شود. نتیجه به دست آمده از این پژوهش با نتیجه حاصل از پژوهش‌های فرانکل و همکاران (۲۰۱۶)، چن و لی (۲۰۱۷) همسو است. نتیجه کاربردی این پژوهش پیشنهاداتی به تحلیل‌گران و پژوهشگران علاقه‌مند به کار در حوزه متن‌کاوی است، که از متغیر ارقام تعهدی متنی در پژوهش‌ها و تجزیه و تحلیل‌های مرتبط با مدل‌های کیفیت سود با بهره‌گیری از تکنیک‌های یادگیری ماشین (به‌طور خاص رگرسیون بردار پشتیبان) استفاده نمایند. به نهادهایی مانند سازمان بورس اوراق بهادار، سازمان امور مالیاتی، سازمان بازرسی و دیوان محاسبات پیشنهاد می‌شود که روش استفاده شده در این پژوهش را به عنوان یکی از روش‌های برآورد دقیق‌تر ارقام تعهدی و در نتیجه تعیین میزان کیفیت سود ابراز شده در گزارش‌های یک شخصیت گزارش‌دهنده، مدنظر قرار داده و به شیوه مقتضی در رسیدگی‌های خود از آن استفاده نمایند. این روش می‌تواند برای تجزیه و تحلیل سخنرانی مدیران که به متن تبدیل شده باشند) نیز کارآمد باشد.

فهرست منابع

- * اسماعیلی، مهدی (۱۳۹۱). داده کاوی مفاهیم و تکنیکها. انتشارات: نیاز دانش، چاپ سوم، ص ۳۸
- * افلاطونی، ع. (۱۳۹۵). تحلیل آماری در پژوهشهای مالی و حسابداری بانرم افزار استتا. انتشارات ترمه. چاپ اول.

مجمع تکیه کند و قابلیت این روش را نشان داده است. از سوی دیگر این پژوهش برای توصیف و تبیین ارقام تعهدی که نقش مهمی در سنجش کیفیت سود شرکت‌ها دارند از ابزار رگرسیون بردار پشتیبان که یکی از موثرترین تکنیک‌های مبتنی بر یادگیری ماشین است بهره گرفت و مفید بودن آن را نسبت به تکنیک فرهنگ واژگان نشان داد. نتیجه حاصل از آزمون فرضیه اول پژوهش حاکی از آن است که ۱۹/۶ درصد از تغییرات ارقام تعهدی واقعی توسط ارقام تعهدی متنی قابل توصیف است یعنی ارقام تعهدی متنی دارای قدرت تبیین ارقام تعهدی واقعی در شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران هستند. نتیجه به دست آمده از آزمون این فرضیه نشان داد که گزارش‌های متنی چون گزارش هیات مدیره، دارای بار اطلاعاتی برای تصمیم‌گیرندگان است. استفاده‌کنندگان مالی در صورت وجود ابزارهای مناسب برای تحلیل گزارش‌های ناساختاریافته و روایتی قادر خواهند بود از اطلاعات موجود در این گزارش‌ها بهره گیرند. نتیجه حاصل از این پژوهش با نتیجه به دست آمده در پژوهش فرانکل و همکاران (۲۰۱۶) مطابقت دارد. رویکرد مبتنی بر واژه‌نامه وسیع‌ترین و عمومی‌ترین رویکردی است که در مطالعات حسابداری و مالی استفاده می‌شود. در فرضیه دوم قدرت تبیین اضافی تر ارقام تعهدی متنی نسبت به روش لغت‌نامه‌ای (روشی مبتنی بر تکنیک لغت‌نامه) بررسی گردید. در وهله اول، قدرت تبیین ارقام تعهدی لغت‌نامه‌ای در توصیف ارقام تعهدی واقعی مورد بررسی واقع گردید. نتیجه به دست آمده گویای این موضوع بود که ارقام تعهدی لغت‌نامه‌ای قادر به تبیین ۱۵/۳ درصد از تغییرات ارقام تعهدی واقعی می‌باشند. در ادامه با وارد نمودن ارقام تعهدی متنی در مدل اولیه، نقش ارقام تعهدی لغت‌نامه‌ای در مقایسه با ارقام تعهدی متنی به‌منظور تبیین ارقام تعهدی واقعی بررسی شد؛ نتیجه نشان داد که ارقام تعهدی متنی و ارقام تعهدی لغت‌نامه‌ای، توأمآ قادر به تبیین ۵۳ درصد از تغییرات ارقام تعهدی واقعی می‌باشند. نتیجه دو تفسیر مهم به همراه دارد: اولاً وقتی

- * Amami, 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.
- * Aureli, Selena. (2017). A comparison of content analysis usage and text mining in corporate disclosure. *The International Journal of Digital Accounting Research*, 17: 1-32.
- * Ball, C., Hoberg, G., Maksimovic, V., (2015). Disclosure, Business Change, and Earnings Quality. Available at SSRN.
- * Barth ME, Cram D, Nelson K (2001). Accruals and the prediction of future cash flows. *The Accounting Review*, 76(January): 27-58.
- * Beattie, V. (2014). Accounting narratives and the narrative turn in accounting research: issues, theory, methodology, methods and a research framework. *British Accounting Review*, 46(2), 111-134
- * Chen, J. V., and F. Li. (2017). Estimating the Amount of Estimation in Accruals. Available at SSRN: <https://ssrn.com/abstract=2738842>
- * Chen, Y.; Lingling, Zh.; and L. Zhang (2013). "Financial Distress Prediction for Chinese Listed Manufacturing Companies", *Procedia Computer Science*. Vol. 17, pp. 678-686.
- * Chou, C. Janie Chang, Chen, Cand W. Chiang (2018) Measuring the Consistency of Quantitative and Qualitative Information in Financial Reports: A Design Science Approach. *Journal of Emerging Technologies in Accounting*, Vol. 15, No. 2, pp. 93-109.
- * Dechow, P.M., W. Ge, and K.M. Schrand (2010): Understanding earnings quality: A review of the proxies, their determinants and their consequences, *Journal of Accounting and Economics*, 50: 344-401.
- * Dechow, P. M., and I. D. Dichev. (2002). The quality of accruals and earnings: role of accrual estimation errors. *The Accounting Review*, 77: 3-35
- * Drew, J., K. Tysiac. (2019). What to expect in 2020? *Journal of Accountancy*, December 1.
- * Dyer, T., Lang M., and Stice-Lawrence, L. (2016). The Ever-Expanding 10-K: Why are 10-Ks Getting so Much Longer (and Does it Matter)? Working paper.
- * Frankel, R., Jennings, J and J. Lee. (2016). Using Unstructured and Qualitative Disclosure to Explain Accruals.
- * بنی مهد، ب.، م. عربی، ش. حسن‌پور. (۱۳۹۶). پژوهش‌های تجربی و روش‌شناسی در حسابداری. انتشارات ترمه، چاپ سوم.
- * ثقفی، ع. (۱۳۹۳). نظریه‌های حسابداری. نشر: انجمن حسابداری ایران با همکاری انتشارات ترمه. جلد اول. چاپ سوم. ص ۱۴۳.
- * دارابی، ر.وغ. مشایخی. (۱۳۹۵). تاثیر هوش مالی در پیش‌بینی ریسک اعتباری با استفاده از مدل ماشین بردار پشتیبان. تحقیقات حسابداری و حسابرسی (۳۰): ۱۰۲-۱۲۵
- * رحیمی دستجردی، م، اخدومی پور، وم. بهار مقدم. (۱۳۹۵). بررسی قیمت‌گذاری اقلام عادی و غیر عادی اجزای نقدی و تعهدی سود. فصلنامه حسابداری مالی ۸ (۳۱): ۱۶۱-۱۳۳
- * رهروی دستجردی، ع. (۱۳۹۶). برآورد میزان احتمال خطر تقلب در گزارشگری مالی از طریق تجزیه و تحلیل متن و تطبیق انگیزه‌های آن با الگوی نظریه چشم‌انداز تجمعی. رساله جهت اخذ درجه دکترا، تخصصی حسابداری، دانشگاه اصفهان.
- * ظفری، س.د. فروغی، غ. کیانی. (۱۳۹۸). تاثیر قابلیت مقایسه و ثبات رویه بر عدم تقارن اطلاعاتی: رویکرد متن کاوی، نشریه علمی حسابداری مدیریت، سال ۱۲، شماره ۴۱، ۱۳۳-۱۵۰
- * فروغی، د.ه. امیری، وه. شیخی. (۱۳۹۲). تاثیر کیفیت اقلام تعهدی بر صرف ریسک شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. پژوهش‌های حسابداری مالی ۵ (۱): ۱۳-۲۸.
- * نوروش، ا، ور. شیروانی. (۱۳۹۴). فرهنگ حسابداری نوروش: اصطلاحات حسابداری، مالی و مدیریت، چاپ هشتم، انتشارات اشراقی.
- * Ajina, A., M Laouiti, B Msolli. (2016). Guiding through the Fog: Does annual report readability reveal earnings management? *Research in International Business and Finance*. 38, 509-516.
- * Allee, K., DeAngelis, M., (2015). The structure of voluntary disclosure narratives: evidence from tone dispersion. *Journal of Accounting Research*, 53(2), 241-274.

- * Tetlock PC.(2007). Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance*,62(3):1139-1168.
- * Turegun,N.,(2019). Text Mining in Financial Information, *Current Analysis on Economics & Finance*,Vol 1:18-26
- * Vapnik,V.,(1998), *Statistical Learning Theory*.Wiley, New York
- * Data mining: Practical machine learning tools and techniques. Morgan Kaufmann.
- * Wooldridge, J. M. (2001). "Applications of Generalized Method of Moments Estimation. *Journal of Economic Perspectives*,15(4)
- * Zhang, Y.(2012). "The empirical study of earnings management based on Chinese listed companies," *Lingnan Journal of Banking, Finance and Economics: Vol. 3, Article 2*
- * Han, I., Kim, B. Y., Lee, J., & Park, S. H.(2013). "Information Asymmetry and the Accrual Anomaly". KAIST College of Business Working Paper Series, (2013034)
- * Healy, P. & Wahlen, J. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*,13 (4): 365-383.
- * Hribar,P. Samuel J. Melessa, R. Christopher Small and Jaron H. Wilde.(2017). Does Managerial Sentiment Affect Accrual Estimates? Evidence from the Banking Industry, *Journal of Accounting and Economics*,63(1): 26-50
- * Lang, M., and Stice-Lawrence, L., (2015). Textual Analysis and International Financial Reporting: Large Sample Evidence, *Journal of Accounting and Economics*,110-135
- * Leuz, C., and, P. Wysocki. (2016). The Economics of Disclosure:Evidence and Suggestions for Future Research. *Journal of Accounting Research*, 54:525-622.
- * Li F.(2008). Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics*, 45(2):221-247.
- * Li Guo, Feng Shi, Jun Tu.,(2016). Textual analysis and machine leaning: Crack unstructured data in finance and accounting. *The Journal of Finance and Data Science*, 2 : 153-170
- * Lo.,K.,Ramos.,F. and R.Rogo.,(2017). Earnings management and annual report readability,*Journal of Accounting and Economics*, (in Press).
- * Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. *The Journal of Financ.* 69, 1643-1671.
- * Mayew.W.J., Sethuraman,M and M. Venkatachalam.(2015). MD&A Disclosure and the Firm's Ability to Continue as a Going Concern, *The Accounting Review*90(4): 1621-1651
- * Muslu,V.,Radhakrishnan,S.,Subramanyam, K.R.,Lim,D.,(2014).Forward-lookingMD&A disclosures and the information environment.,*Management Science* ,61(5):931-948.
- * Sloan, R., (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*,71:289-316.

يادداشت‌ها

1. Structured data
2. quantitative disclosures
3. qualitative disclosures
4. Narratives
5. Textual documents
6. Unstructured data
7. Data mining
8. Text mining
9. MD&A
10. Semi structured
11. Description
12. Prediction
13. Classification
14. Retrospection
15. Prospection
16. Tone
17. Readability
18. Support vector mashin
19. Dictionary-based approach
20. Bag-of-words
21. . www.eboo.ir
22. stop words
23. stemming word