

تجزیه و تحلیل کلان داده‌ها و کاربرد آن در حسابرسی صورت‌های مالی

کیمیا نوش گنجی
کارشناسی ارشد حسابداری
دانشگاه تهران

تصمیم‌گیری است. در راستای دستیابی به اهداف مطالعه حاضر، فرض بر آن است که حسابرسان بر روی معاملات، مانده‌ها و افشای اطلاعات زیربنای صورت‌های مالی متمرکز هستند که به ادعاهای مدیریت مربوط است. در حسابرسی صورت‌های مالی در هماهنگی با استانداردهای بین‌المللی حسابرسی، فرصت‌های بالقوه بی‌شماری برای تجزیه و تحلیل کلان داده وجود دارد. به عنوان نمونه، فعالیت‌های حسابرسی زیر احتمالاً از تجزیه و تحلیل کلان داده‌ها منتفع خواهند شد:

- شناسایی و ارزیابی ریسک وابسته به پذیرش یا ادامه تعاملات حسابرسی، برای مثال، ریسک‌های ورشکستگی یا تقلب مدیران ارشد.
 - شناسایی و ارزیابی ریسک‌های تحریف‌های بااهمیت صورت‌های مالی با توجه به تقلب، و آزمون تقلب با توجه به ریسک‌های ارزیابی شد.
 - شناسایی و ارزیابی تحریف‌های بااهمیت از طریق درک واحد تجاری و محیط آن. این مورد شامل روش‌های تحلیلی مقدماتی و به همین ترتیب ارزیابی طراحی و پیاده‌سازی کنترل‌های داخلی و آزمون اثربخشی‌های عملیاتی آنها است.
 - اجرای روش‌های تحلیلی محتوایی در پاسخ به ارزیابی حسابرسان از ریسک‌های تحریف‌های بااهمیت.
 - اجرای روش‌های تحلیلی در نزدیکی پایان کار حسابرسی، به حسابرسان کمک می‌کند تا یک نتیجه‌گیری کلی درباره وضعیت تطابق صورت‌های مالی با درک حسابرسان از واحد تجاری کند.
- در این مقاله، فرض بر آن است که حسابرسان صورت‌های مالی می‌توانند به صورت بالقوه با روش‌های تحلیلی بهبود یابند که از کلان داده استفاده می‌کنند. در این نوع حسابرسی‌ها، داده‌ها شامل معاملات و مانده‌هایی است که در حالت عادی از برنامه‌ریزی منابع سازمانی^(۵) واحد تجاری و سیستم پایگاه داده آن واحدها به

کلان داده‌ها، به مجموعه داده‌هایی اطلاق می‌شود که برای دستکاری یا تحقیق با استفاده از روش‌ها و ابزارهای استاندارد، پیچیده و بسیار بزرگ هستند. کلان داده‌ها شامل سه مولفه حجم^(۱)، نرخ تولید^(۲) و تنوع^(۳) هستند که تحت عنوان "سه V" مطرح شده‌اند. حجم به حجم گسترده مجموعه داده‌ها، نرخ تولید به سرعت تولید داده و تنوع به تعدد منابع داده اشاره دارد، هر سه V باهم در ارتباط هستند. کلان داده‌ها، پدیده نسبتاً جدیدی است، محصول محیط فناورانه است که در آن محیط، تقریباً همه چیز می‌تواند به صورت دیجیتال ثبت شود، اندازه گرفته شود و در نتیجه، به داده تبدیل شود. این موضوع همان مفهوم داده‌ای شدن^(۴) را مورد اشاره قرار می‌دهد. داده‌ای شدن می‌تواند باعث شود هزاران رخداد را که شامل تعدادی عدد، متن، تصویر، صدا و ویدیو است و نیازمند ظرفیت حافظه پتابایتی (هزاران ترابایتی) است، به صورت همزمان ردیابی و به صورت بلادرنگ اجرا شوند. به عنوان نمونه‌ای از کلان داده‌ها می‌توان به انجام بیش از یک میلیون تراکنش مشتریان در هر ساعت در فروشگاه‌های مارت، بیش از ۵۰ میلیارد عکس در فیس‌بوک و ۲۰۰ گیگابایت داده‌های نجومی اشاره کرد که هر شب جمع‌آوری می‌شوند. کلان داده‌ها در بازاریابی برای جذب مشتریان بالقوه، در مبارزه سیاسی برای مطالعه جمعیتی رای دهندگان، در ورزش برای ارزیابی تیم و بازیکنان، در امنیت ملی برای شناسایی تهدیدها، در بیولوژی برای مطالعه DNA و در اجرای قانون برای شناسایی مظنونان به جرم و جنایات استفاده می‌شود.

تجزیه و تحلیل کلان داده‌ها متشکل از فرآیند بازرسی، شفاف‌سازی، تبدیل و مدل‌سازی برای کشف و برقراری ارتباط الگوها و اطلاعات سودمند، بیان نتیجه‌گیری و پشتیبانی از

خود را بهتر مدیریت کند. داده‌های جمعیت‌شناختی و جغرافیایی به صورت بالقوه می‌توانند پیش‌بینی کننده‌ای منطقی برای درآمد و فروش در موقعیت و مکان کسب و کارهای فردی باشند. تخمین‌های نتیجه شده ممکن است به عنوان معیاری برای ارزیابی مقدار فروش با توجه به موقعیت و مکان استفاده شوند. علاوه بر این، معیارهای مبتنی بر گروه هم‌پیشگان می‌تواند برای جلب توجه معیارهای مشکل زای ممکن استفاده شوند. تجزیه و تحلیل‌های مشابه ممکن است فرآیندهای حسابرسی را با تاکید بر روی منابع در قسمت‌های مخاطره آمیزتر کسب و کار بهبود بخشند.

سوم، تجزیه و تحلیل کلان داده معمولاً شامل چند منبع داده است که برخی از آن‌ها ساختاریافته‌اند و برخی دیگر ساختاریافته، از جمله اعداد، متن، تصاویر، صدا، ویدیو که پردازش هر یک از آنها شامل ترکیب‌هایی از روش‌های تحلیلی متفاوت از رشته‌های متفاوت هستند. یک نمونه‌ای از تجزیه و تحلیل چشم انداز^(۱۰) آیاتا^(۱۱) است که در اکتشافات نفت و گاز برای پیش‌بینی محل‌های حفاری بهینه از داده‌هایی مانند تصویر از چاه‌ها، ویدیو جریان سیالات از شکستگی‌های هیدرولیکی، صدای عملیات حفاری، متون یادداشت‌های حفاری، و اعداد گزارش‌های تولید استفاده می‌شود. در راستای چالش یکپارچه‌سازی منابع متفاوت کلان داده‌ها از جمله خبرها، جریان‌های صوتی و تصویری، موارد ثبت شده در تلفن‌های سلولی، نظرات در رسانه اجتماعی و استفاده از آن‌ها برای اهداف حسابرسی توسط مافیت و واسرلی (۲۰۱۳) بحث شد که این افراد استفاده از چنین داده‌هایی را برای به‌دست آوردن فرم جدیدی از شواهد، تایید وجود رخداد، و اعتبارسنجی اجزای گزارش را پیشنهاد دادند.

چهارم، پلیس لس‌آنجلس داده‌های صحنه جرم، از جمله زمان، محل، موقعیت، ماهیت و عوامل را به منظور پیش‌بینی محتمل‌ترین زمان بندی و محتمل‌ترین محل جرم و مستقر کردن موثر نیروها تجزیه و تحلیل می‌کند. نتایج حاصل از این تجزیه و تحلیل، بهبود قابل توجهی در توانایی پلیس برای پیشگیری از فعالیت‌های جنایی و خنثی کردن عوامل بالقوه مانند اعضای باند در نواحی پیش‌بینی شده ایجاد کرده است. تجزیه و تحلیل‌های مشابه که مبتنی بر اطلاعاتی در مورد فعالیت‌های گذشته یا نتایج حسابرسی گذشته شرکت است می‌توانند توسط حسابرسان برای شناسایی ریسک‌های تقلب و هدایت تلاش‌های حسابرسی با هدف کشف تقلب استفاده شوند.

پنجم، کمیسیون بورس و اوراق بهادار (SEC) بر روی برنامه‌های کاربردی تجزیه و تحلیل داده‌ها برای نظارت بر رویدادهای بازار، کشف تقلب در صورت‌های مالی و شناسایی شکست‌های حسابرسی سرمایه‌گذاری می‌کند. برای نمونه، سیستم تجزیه و تحلیل داده‌ها اطلاعات بازار^(۱۲) توسط کمیسیون بورس و اوراق بهادار در ژانویه ۲۰۱۳ توسعه یافت و حدوداً یک میلیارد رکورد را روزانه از مبادلات بازار سرمایه جمع‌آوری می‌کند. در داده بسیار حجیم، پردازش‌های بزرگ چالش برانگیز هستند و نیاز به متخصص برای داده‌های تخصصی پیش می‌آید. در جولای سال ۲۰۱۳، نهاد گزارشگری مالی و نیروی کار حسابرسی^(۱۳) برای تقویت تلاش برای شناسایی نقض قوانین اوراق بهادار در ارتباط با آماده‌سازی صورت‌های مالی، افشای اطلاعات و گزارش دهی؛ و شکست‌های

دست آمده‌اند. این داده‌ها به خودی خود کلان داده نیستند مگر اینکه در دوره زمانی مهمی جمع‌آوری شوند و با حقایق اضافی تکمیل شوند. بنابراین، بیشتر فرصت‌های کلان داده‌ها که در این مقاله بر آن‌ها بحث می‌شود از داده‌های کمکی می‌آیند که بعد از پردازش ممکن است توجه حسابرس را به موارد بااهمیت جلب کند. کلان داده مورد نظر بالقوه برای حسابرسی شامل اطلاعات رسانه اجتماعی، سیستم‌های نظارت تصویری، و داده‌های معاملات بازار سهام است.

نمونه‌هایی از تجزیه و تحلیل کلان داده‌ها

از آنجایی که چند برنامه کاربردی فعلی از تجزیه و تحلیل کلان داده‌ها در حسابرسی مستقل وجود دارد، که ما از آن‌ها اطلاعی نداریم، نمونه‌هایی از دیگر رشته‌ها و فرضیه‌ها در راستای این شرح داده می‌شود که چگونه برنامه‌های کاربردهای مشابه می‌توانند در حسابرسی مستقل پیاده‌سازی شوند.

ابتدا، در دسترس بودن منابع داده حجیم و آگاهی‌دهنده در رویکردهای جدید برای پیش‌بینی میانگین قیمت سهام نتیجه‌بخش بوده است. برای نمونه، بال، ماو و ژنگ (۲۰۱۱) خلق و خوی عمومی کلی را براساس داده‌های توییتر سنجیدند و به طور موفقیت آمیزی از آن برای پیش‌بینی نوسانات روزانه شاخص میانگین صنعتی داو جونز^(۶) استفاده کردند. آنان از نمایه وضعیت خلقی گوگل و ابزارهای دانشگاهی نظریاب استفاده کردند که برای تولید سری زمانی داده‌های روزانه از تغییر حالات براساس حدوداً ۱۰ میلیون توییت پست شده توسط تقریباً ۲,۷ میلیون کاربر توسعه یافته‌اند. با انجام این کار، مولفان قادر به پیش‌بینی تغییرات در شاخص داو جونز برای سه تا چهار روز آینده شدند. علاوه بر رسانه اجتماعی، مقالات جدید به عنوان پیش‌بین برای حرکات قیمت سهام شناخته شدند. با این اوصاف می‌توان استدلال کرد که منابع داده مشابهی می‌توانند برای پیش‌بینی ورشکستگی یا ارزیابی کلی وضعیت مالی شرکت مورد استفاده قرار گیرد. اینگونه ابزارها می‌توانند برای شناخت بهتر و ارزیابی پذیرش ریسک و همچنین کاهش ریسک دعاوی حقوقی مورد استفاده قرار گیرد.

دوم، داده‌های جمعیت شناختی و وضعیت جوی برای پیش‌بینی رفتار مشتری استفاده شدند. شرکت آفیس مکس^(۷)، یک خرده فروش بزرگ لوازم اداری، از لایو پردیکت^(۸)، سیستمی که ارائه دهنده فناوری برخط است که توسط شرکت مونیت^(۹) ارائه می‌گردد، برای شخصی‌سازی صفحات اصلی آنلاین براساس ویژگی جمعیت‌شناختی مشتریان ایجاد کرد. جالب توجه است که تمرکز اصلی این سیستم، تلاش برای پیش‌بینی دید سیاسی مشتریان، و تعدیل بر طبق آن است. سیستم از آدرس IP برای شناسایی موقعیت مشتریان و از داده‌های سرشماری ایالت متحده برای ایجاد نمایه جمعیت شناختی استفاده کرد. در یک برنامه کاربردی مربوط به آب و هوا، شرکت وال مارت، با تجزیه و تحلیل داده‌های مربوط به ترابایت‌ها تراکنش مشخص شد که در زمان تهدید طوفان، مشتریان نه تنها چراغ قوه اضافی می‌خرند، بلکه فروش پاپ تارت توت فرنگی (یک اسنک محبوب صبحانه) هفت برابر افزایش می‌یابد. این یافته و یافته‌های مشابه از تجزیه و تحلیل کلان داده‌ها به شرکت وال مارت کمک می‌کند که موجودی‌های



حسابرسی تشکیل شد. نیروی کار از یک مدل کیفیت حسابداری تحلیلی، برای اسکن بایگانی‌های نظارتی معمولی و فعالیت‌ها با ریسک بالا استفاده می‌کند که با بازرسی دقیق‌تر توسط تیم اجرایی کمیسیون بورس و اوراق بهادار تضمین می‌شوند. در همان زمان، کمیسیون بورس و اوراق بهادار تشکیل نیروی کار تقلب میکروکاپ را برای بررسی تقلب در صدور، بازاریابی، و مبادله اوراق بهادار را اعلام کرد که وب سایت و رسانه‌های اجتماعی توسط آن نظارت خواهد شد زیرا کلاهبرداران میکروکامپ^(۱۴)، مکرراً آن‌ها را برای شکار سرمایه داران ساده لوح بکار می‌گیرند. تجزیه و تحلیل‌های مشابه می‌توانند توسط حسابرسان برای شناسایی تقلب یا فعالیت‌های با ریسک بالا استفاده شوند.

لازم به یادآوری است که گروه حسابرسی داخلی در برخی از شرکت‌های بزرگ از کلان داده‌ها در سازمان استفاده می‌کنند. برای نمونه، تیم حسابرسی داخلی در شرکت‌های بلوکراس^(۱۵) و بلوشیلدز^(۱۶) کارولینای شمالی از تجزیه و تحلیل کلان داده‌ها برای شناسایی ادعاها و مطالبات مضاعف بیمه از بین میلیون‌ها ادعا در هر ماه استفاده می‌کنند. آنها بررسی می‌کنند که مبادا بیمه‌گذاران برای یک ادعا، دو بار یا بیشتر درخواست به بیمه‌گر ارائه داده باشند. مؤسسات حسابرسی بزرگ دنیا، از جمله مؤسسات کی‌پی‌ام‌جی^(۱۷)، دیلویت^(۱۸) و پی

دیلیو سی^(۱۹) نشریاتی در وب سایت‌های خود منتشر کرده‌اند دال بر اینکه چگونه حسابرسان داخلی می‌توانند از تجزیه و تحلیل داده‌ها برای بهبود اثربخشی و کارایی استفاده کنند. برای نمونه، مؤسسه کی‌پی‌ام‌جی بیان می‌کند که "با تجزیه و تحلیل داده‌ها، سازمان‌ها، توانایی بررسی هر تراکنشی را- نه تنها تراکنش نمونه را- پیدا می‌کنند که این مسئله تحلیل کارآمدتر در مقیاس بزرگتر را موجب می‌شود". بسیاری از فعالیت‌های حسابرسی داخلی، حسابرسی مستقل صورت‌های

مالی را منعکس می‌کنند تجزیه و تحلیل کلان داده‌های مشابهی می‌تواند اعمال شود.

مشخصه‌های تجزیه و تحلیل کلان داده‌ها

ویژگی‌های اصلی تجزیه و تحلیل کلان داده‌ها باعث می‌شود که کاربر مجدداً فکر کند که چگونه از داده‌ها استفاده کند. ابتدا، تجزیه و تحلیل همه داده‌ها یا اکثر داده‌ها به جای نمونه یا زیر مجموعه دقیق و کوچک ممکن است. این روش می‌تواند به مدل‌هایی منجر شود که نسبت به قبل مقاوم ترند. برای نمونه، اگر حسابرس بخواهد تعیین کند که چه ویژگی از ثبت‌های روزنامه، شاخصی برای ریسک خطاها یا تقلب هستند در حالی که تجزیه و تحلیل همه ثبت‌های روزنامه برای رکوردهای طولانی حفظ شده و از این اطلاعات برای شناسایی ثبت‌های روزنامه استفاده می‌شود که واقعاً غیر معمول است، ممکن است در حالی که در گذشته باید در مورد حذف داده‌های آلوده شده باید بسیار دقت می‌شد، زمانی که همه داده‌های در دسترس هستند، درجه‌ای تحریف قابل پذیرش است.

تغییر دوم در تفکر از علیت به همبستگی است. به جای تلاش برای درک علت‌های اساسی پدیده‌های پیچیده به طور فزاینده‌ای

شناسایی و استفاده از همبستگی ممکن است. برای نمونه، میر و کوکر (۲۰۱۳) گزارش دادند که پژوهشگران در مؤسسه فناوری دانشگاه اونتاریو^(۲۰) با تعدادی بیمارستان بر سر نرم‌افزاری برای این کار کردند که به دکترها کمک کنند که در زمان مراقبت از نوزادان نارس با تشخیص بهتری تصمیم بگیرند. نرم افزار، داده‌های بیماران را به صورت بلادرنگ اتخاذ و پردازش می‌کرد و از ۱۶ جریان داده متفاوت، مانند ضربان قلب، دمای بدن، فشار خون، سطح اکسیژن خون اطلاعاتی جمع‌آوری می‌کرد که همه با هم ۱۲۶۰ نقطه داده به ازای هر ثانیه هستند. در حالی که این مشاهدات ممکن است اجازه دهند که دکتر در نهایت دلایل اصلی را درک کنند، به سادگی می‌دانیم که چیزی احتمال دارد وقوع یابد که مسلماً از درک دقیق علت مهم‌تر است. تصور برنامه حسابرسی مشابه سخت نیست که در آن تجدید ارائه یا سایر رویدادهای نامساعد با شاخص‌های حذف شده از سوابق شرکت و دیگر اطلاعات همبستگی دارند.

توانایی برای استفاده از مدل‌های همبستگی با مقدار وسیعی از داده‌ها با نرخ تولید بالا به منظور تراکنش‌های دقیق یا رخداد‌های مورد نظر حسابرسی در زمانی به طور قابل توجهی سودمندتر می‌شوند که به طور پیوسته بکار می‌روند. حسابرسی مستمر و سیستم‌های نظارتی احتمالاً به طور ویژه‌ای به عصر پتابایتی در تبدیل رویه حسابرسی مرتبط هستند. به عنوان نمونه، روابط آماری بین عناصر و فرآیندهای کسب و کار متفاوت ممکن است به صورت مستمر برای کشف رویدادهای نامنظم نظارت شوند.

اجرای تجزیه و تحلیل کلان داده‌ها در حسابرسی

اجرای تجزیه و تحلیل کلان داده‌ها تلاش با اهمیتی وافر است که نیازمند افرادی با تخصص در تجزیه و تحلیل داده‌ها و همچنین منابع سخت افزاری و نرم افزاری مناسب است. در نتیجه، بسیاری از واحدهای تجاری، برنامه کاربردی کلان داده‌های خود را به شرکت‌های ارائه دهنده خدمات تخصصی (مانند آی‌بی‌ام (IBM) و ترادا) برون سپاری می‌کنند تا راه حل‌های مناسبی را دریافت کنند. به طور مشابه، آموزش‌های مربوط به تجزیه و تحلیل کلان داده‌ها ممکن است در ورای محدوده سطح علمی و حرفه‌ای یک حسابرس باشد. حرفه حسابرسی یا باید یک فرد حرفه‌ای آموزش دیده برای تجزیه و تحلیل را استخدام کند یا احتمال دارد که از خدمات قابل ارائه توسط شخص ثالث برای تجزیه و تحلیل کلان داده‌ها استفاده کند. برون‌سپاری کردن تجزیه و تحلیل کلان داده‌ها، نگرانی‌های حفظ حریم خصوصی را افزایش می‌دهد، اما این مسئله تازه‌ای نیست- حرفه در حال حاضر در زمان انجام حسابرسی به طرف سومی مانند بانک تکیه می‌کند.

در شناسایی ناهنجاری‌ها و استثنائات برای بررسی‌های بیشتر حسابرسی، پیاده‌سازی‌های فعلی روش‌های تحلیلی گاهی اوقات، نسبت به آنچه که به صورت عملی توسط تیم‌های حسابرسی بررسی می‌شود، مثبت کاذب^(۲۱) بیشتری تولید می‌کند و منجر به اضافه بار اطلاعاتی می‌شود. از فرصت‌های تجزیه و تحلیل کلان داده‌ها احتمال کاهش چشمگیر خطای مثبت‌های کاذب از طریق

تجزیه و تحلیل کلان داده‌ها

متشکل از فرآیند بازرسی،

شفاف‌سازی، تبدیل و

مدل‌سازی برای کشف

و برقراری ارتباط الگوها و

اطلاعات سودمند، بیان

نتیجه‌گیری و پشتیبانی از

تصمیم‌گیری است

نشده باشد. در آخر، حرکت به سمت تجزیه و تحلیل کلان داده‌ها می‌تواند تقلب یا خطاهایی را شناسایی کند که در گذشته وجود داشته اما دیده نشده است. این نیز مسئله جدیدی نیست، اما مسئله‌ای است که حساب‌رسان پذیرفته‌اند کلان داده‌ها احتمالاً با آن سروکار دارد. در کنار استفاده از تجزیه و تحلیل کلان داده‌ها برای اجرای حسابرسی، مؤسسات حسابرسی می‌توانند به صورت بالقوه از آن برای اهداف داخلی استفاده کنند. به عنوان نمونه، از آنجایی که بیشتر مقالات کارهای حسابرسی، الکترونیکی هستند، فرصتی برای شرکت وجود دارد که حسابرسی را در سراسر نمونه کارها در جستجوی ناهنجاری‌ها و مسائل کیفیتی بالقوه تجزیه و تحلیل کند.

خلاصه و نتیجه‌گیری

کلان داده‌ها در بسیاری از رشته‌ها انقلابی به پا کردند. در این نوشته، نمونه‌هایی از تجزیه و تحلیل کلان داده‌ها ارائه شد و کاربردهای مشابه آن در حسابرسی ارائه گردید. همچنین به صورت مختصر در ارتباط با ویژگی‌های کلان داده‌ها مطالبی مطرح شد که با حسابرسی و شناخت برخی فرصت‌ها و چالش‌ها و بکارگیری آنها در ارتباط بود.

پی‌نوشت‌ها:

1. Volume
2. velocity
3. variety
4. datafication
5. Enterprise Resource Planning (ERP)
6. Dow Jones Industrial Average
7. OfficeMax
8. LivePredict
9. Monetate
10. Prescriptive Analytics
11. Ayata
12. Market Information Data Analysis System (MIDAS)
13. Financial Reporting and Audit Task Force
14. Microcap
15. Blue Cross
16. Blue Shield
17. KPMG
18. Deloitte
19. PWC
20. Ontario
21. false positives
22. noise

منابع و مآخذ:

- این مقاله ترجمه آزادی از مقاله زیر است:
Cao, M., Chychyla, R., & Stewart, T. (2015). Big Data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423-429.
- سایر منابع و مآخذ:
Bollen, J., H. Mao, and X. Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2 (1): 1-8.
Mayer-Schonberger, V., and K. Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Eamon Dolan/Houghton Mifflin Harcourt.
Moffitt, K., and M. Vasarhelyi. 2013. AIS in an age of Big Data. *Journal of Information Systems* 27 (2): 1-19.

شناسایی دقیق ناهنجاری واقعی و استثنائات با سیستم‌های بهتر اولویت بندی شده است.

چند مسئله وجود دارد که حرفه حسابرسی در رابطه با تجزیه و تحلیل کلان داده‌ها نیاز دارد که با آن‌ها سروکار داشته باشد. در درجه اول، استفاده موفقیت آمیز از کلان داده‌ها نیازمند تغییرات پارادایمی است. به جای استفاده از برخی داده‌ها در مجموعه داده شفاف و تمیز کوچک و تمرکز بر روی علت و معلول (روابط قابل قبولی در استانداردهای حسابرسی)، حساب‌رسان از کلان داده‌ای استفاده می‌کنند که به استفاده از "همه" داده‌ها در مجموعه داده آلوده نسبتاً بزرگ تمایل خواهد داشت، و بیشتر بر روی همبستگی تاکید می‌کند تا علت و معلولی. درجه‌ای که این رویکرد در حسابرسی استفاده می‌شود بسته به مرحله‌ای تغییر می‌کند که حسابرسی در آن قرار دارد: استفاده از داده‌های آلوده برای برنامه‌ریزی و ارزیابی ریسک برخلاف رویه‌های محتوایی قابل تحمل‌تر است. به عنوان نمونه، تجزیه و تحلیل کلان داده‌ها می‌تواند برای شناسایی روندها و الگوهای کسب و کار استفاده شود، تجزیه و تحلیل حسابرسی سنتی و روش‌های حسابرسی به کمک کامپیوتر می‌تواند برای انجام تجزیه و تحلیل دقیق‌تر مسائل بالقوه استفاده شود و قضاوت بر روی حسابرسی معمولی می‌تواند برای تعیین تاثیرات یافته‌ها بر روی گزارش‌های مالی استفاده شود. علاوه بر این، داده‌های آلوده ممکن است برای رویه تحلیلی که به اختلال (۲۲) حساس است مناسب نباشند. باین وجود، این تفکر تا حدودی از این حرفه بیگانه است. این فکر اساساً نیازمند راهنما و آموزش جدید قابل توجه است، و ممکن است حتی نیازمند تعدیل استانداردهای حسابرسی باشد.

ثانیاً، حجم کلان داده‌ها چالش‌های محاسباتی قابل توجهی را معرفی می‌کند. بسیاری از روش‌های تحلیلی متداول استفاده شده در حسابرسی نمی‌توانند در کلان داده‌ها مورد استفاده قرار بگیرند. راه حل استفاده از روش تحلیلی ساده است که نیازمند منابع محاسباتی کمتر است، یا انتخاب زیرمجموعه‌ای از داده است که می‌تواند با ابزارهای تحلیلی پیچیده‌تر مدیریت شوند. مورد بعدی استفاده از کلان داده برای انتخاب دقیق زیر مجموعه‌ای است که برای حسابرسی ارزشمندتر است. به عنوان نمونه، روش‌هایی برای انتخاب زیر مجموعه‌ای از داده وجود دارد که در مدل‌های تحلیلی دقیق‌تر نتیجه می‌دهند.

ثالثاً، زمانی که کلان داده استفاده می‌شود، حفظ حریم شخصی نگرانی بالقوه‌ای است. برخی از تجزیه و تحلیل‌ها ممکن است نیازمند اطلاعات خصوصی صاحبکار، فراتر از آنچه باشند که برای حسابرسی منتشر می‌شود. اگرچه این نگرانی مختص حسابرسی نیست اما استفاده از چنین اطلاعات حساسی در برنامه کاربردی کلان داده، چالش‌هایی را نشان می‌دهند. به عنوان نمونه، اتحادیه اروپا گوگل را به عنوان نمونه‌ای از نگرانی‌های حریم شخصی و آنتی تراست مربوط به استفاده از کلان داده مورد نقد قرار می‌دهد. سرانجام، زمانی که "همه" داده‌ها از طریق سیستم‌های تجزیه و تحلیل حسابرسی پردازش شدند و شکستی در شناسایی تقلب یا خطا وجود داشت، ریسکی وجود دارد که حساب‌رسان آن را حدس می‌زنند. این مسئله کاملاً جدیدی نیست، اما حساب‌رسان به صورت سنتی مبتنی بر نمونه‌ها کار می‌کنند و پذیرفته‌اند ریسک‌های آماری وجود دارند که در آن‌ها تقلب یا خطا شناسایی