

## معرفی یک چارچوب مناسب سازماندهی برای کاربردهای داده کاوی در حسابداری و حسابرسی: مروری بر تکنیک‌های پر کاربرد طبقه‌بندی داده‌های مالی

یاسر رضائی پیتته نوئی<sup>۱</sup>

مصطفی اصغری شلمانی<sup>۲</sup>

حسین دلیری دهبنه<sup>۳</sup>

تاریخ دریافت: ۱۴۰۰/۱۰/۲۸ تاریخ چاپ: ۱۴۰۰/۱۱/۲۸

### چکیده

عصر اقتصاد جهانی به سرعت در حال تغییر، سازمان‌ها و شرکتها را ملزم می‌کند برای این که بتوانند رقابت مفیدی داشته باشند، طیف گسترده‌ای از رویکردهای مدیریت نوآورانه و فناوری‌های اطلاعاتی پیشرفته را مورد توجه قرار دهند. از این رو داده کاوی از اهمیت خاصی در آینده‌ی حرفه‌ی حسابداری و حسابرسی برخوردار است زیرا داده کاوی یکی از مهم‌ترین موضوعات اخیر جهت تحلیل تجاری هوشمند و ابزار پشتیبانی تصمیمات می‌باشد. داده کاوی مزایا و توانایی‌های گسترده‌ای شامل پیش‌بینی کارآمد روند-های آتی توسعه شرکتها و افزایش رقابت کسب و کار را در اختیار سازمان‌ها برای تصمیم‌گیری‌های بهتر قرار می‌دهد. از سوی دیگر با وجود شبکه جهانی وب، هر لحظه بر حجم اطلاعات سیستم‌های یکپارچه اطلاعاتی افزوده می‌شود، بر این اساس نیاز به یک علم مناسب جهت کشف و استخراج سریع و دقیق اطلاعات از این پایگاه‌های داده ضروری است که در حال حاضر داده-کاوی مهم‌ترین فناوری برای بهره‌وری موثر، سریع و صحیح از داده‌های حجیم می‌باشد. امروزه داده کاوی و ابزارهای آن از اهمیت ویژه‌ای در امور مالی و حسابداری برخوردار است. از این رو در این مقاله به بررسی کاربرد داده کاوی در حرفه حسابداری و حسابرسی پرداخته شده است. همچنین از آنجایی که طبقه‌بندی و توانایی‌های تکنیک‌های داده کاوی باعث می‌شود، این علم روشی جهت پیش‌بینی ورشکستگی، نابسامانی مالی، تشخیص کلاهبرداری مدیریتی، برآورد ریسک و پیش‌بینی عملکرد شرکت ارائه دهد، در بخش اصلی مقاله تکنیک‌های پر کاربرد طبقه‌بندی در حسابداری و حسابرسی، معرفی و مورد بحث قرار گرفت.

### واژگان کلیدی

حسابداری، حسابرسی، داده‌های مالی، داده کاوی و طبقه‌بندی

<sup>۱</sup> استادیار حسابداری، دانشکده ادبیات و علوم انسانی، دانشگاه گیلان، رشت، ایران. ([Rezaei.yasser@gmail.com](mailto:Rezaei.yasser@gmail.com))

<sup>۲</sup> معاون مالی بیمارستان ولیعصر (عج) نداجا رشت. ([Mostafaasgharimali@gmail.com](mailto:Mostafaasgharimali@gmail.com))

<sup>۳</sup> مسئول تنظیم حساب امور مالی شهرداری رشت. ([hdhossein1961@gmail.com](mailto:hdhossein1961@gmail.com))

## ۱- مقدمه بر حسابداری و حسابرسی

مفهوم اولیه‌ی عملکردهای پیوسته حسابداری، الزاما یک ایده جدید یا مبتکرانه نیست اما با پیشرفت‌های فنی جدید در حوزه بازار، به نظر می‌رسد که این مفهوم در حال حرکت به سوی واقعیت می‌باشد. حسابداری فارغ از این که در صنایع خصوصی و یا شرکت‌های حسابداری عمومی می‌باشد، هنوز مبتنی بر مرورهای دوره‌ای و تحلیل اطلاعات مالی می‌باشد. در حالی که کاملاً مشخص است که تایید فعالیت، مالیات و خدمات مشاوره‌ای هنوز فرصت‌های طلایی برای متخصصان های حسابداری به شمار می‌آیند و مبنای عملکرد تخصصی این زمینه را فراهم می‌کنند. روش‌هایی که این خدمات را ارائه می‌دهند به نظر در حال تغییر و تکامل می‌باشد. از شواهد موجود بازار به نظر می‌رسد که رشد شرکت‌های حسابداری با تکامل و تغییر مطابق با نیروهای بازاری، می‌توانند به موفقیت دست پیدا کنند. به بیان دیگر فعالیت‌های فعلی و روندهای مرتبط با ارائه ارزش توسط متخصصان های حسابداری تقویت شده و یا حتی به صورت کلی ممکن است تحت نیروهای تکنولوژی، تغییر کند. این تغییر که از قبل در بازار ایجاد شده است، به نظر انشعاب‌هایی برای تمام ابعاد حرفه‌ی حسابداری خواهد داشت [۱]. گسترش روز افزون جوامع و پیچیدگی مسائل خاص آنها، افزایش نیاز به اطلاعات اقتصادی مربوط و در نتیجه افزایش تقاضا برای وجود سیستم‌ها و فرآیندهای فراهم کننده چنین اطاعتی را ایجاد کرده است. حسابداری مالی وظیفه انتقال این اطلاعات را بر عهده دارد. گزارشگری مالی که محصول نهایی فرایند حسابداری مالی است، در بنگاه‌های اقتصادی، بازتاب نیازهای اطلاعاتی و انتظارات گروه‌های مختلف استفاده کننده صورت‌های مالی، چون سرمایه‌گذاران، اعتباردهندگان، دولت و مدیران است.

از حسابرسی به عنوان رشته‌ای قدیمی یاد می‌شود و و متناوباً حساب‌برسان را به عنوان افرادی خشک و رسمی و ایرادگیر مورد قضاوت قرار می‌دهند. در واقع این موضوع دیگر حقیقت ندارد. در سالهای گذشته حساب‌برسان پی به افزایش شدید حجم معاملات و پیچیدگیهای داده‌های مالی و غیر مالی صاحبکارشان برده‌اند [۱۹]. در واقع حسابرسی جزء لاینفک انتقال اطلاعات اقتصادی به شمار می‌آید و بازتاب با اهمیتی در شناسایی و اندازه‌گیری اطلاعات مربوط دارد. استفاده از خدمات حساب‌برس مستقل به منظور اعتباربخشی به اطلاعات ارائه شده، امری ضروری است. هدف حسابرسی در چارچوب فرآیند آزمون محور، زدودن آلودگی‌های اطلاعات است تا از طریق افزایش توان اتکای آن بستری مساعد برای استفاده از اطلاعات در راستای تصمیم‌گیری‌های اقتصادی فراهم شود. محصول نهایی این فرایند گزارشی است که حساب‌برس در آن، نظر حرف‌های خود را در مورد مطلوب بودن صورت‌های مالی ارائه می‌کند. اظهار نظر در حسابرسی به مجموع قضاوت حرف‌ها، تصمیم‌گیری‌ها و توجه‌های حساب‌برس به صورت فشرده به عنوان یک نظر کارشناسی و حرف‌های حساب‌برس گفته می‌شود. استاندارد حسابرسی شماره ۷۰۰ بیان می‌کند که حساب‌برس باید درباره اینکه آیا صورت‌های مالی، از تمام جنبه‌های با اهمیت، طبق استانداردهای حسابداری، به نحو مطلوب تهیه شده است یا خیر، اظهار نظر کند. حساب‌برس زمانی باید اظهار نظر تعدیل نشده ارائه کند که براساس نتیجه‌گیری او، صورت‌های مالی از تمام جنبه‌های با اهمیت، طبق استانداردهای حسابداری، به نحو مطلوب تهیه شده باشد. اگر حساب‌برس براساس شواهد حسابرسی کسب شده به این نتیجه برسد که، صورت‌های مالی عاری از تحریف با اهمیت نیست؛ یا قادر به کسب شواهد حسابرسی کافی و مناسب برای نتیجه‌گیری درباره نبود تحریف با اهمیت در صورت‌های مالی نباشد، باید طبق استاندارد ۷۰۵ اظهار نظر خود را در گزارش حسابرسی تعدیل کند [۲۰].

## ۲- داده‌ها به عنوان یک دارایی رقابتی

داده‌ها تقریباً یک نقش کلیدی در روندهای مدیریت سازمان‌ها، توسعه و هدایت آنها دارند و نشان می‌دهند که این نتایج چگونه به گروه‌های سهام‌دار داخلی و خارجی منتقل می‌شود. بر این اساس و با در نظر داشتن این حقیقت که این مفهوم و ایده‌ی استفاده از اطلاعات برای تصمیم‌گیری‌های موثر موضوع جدیدی نیست، پتانسیل قوی و مقیاس‌پذیری پردازش داده‌ها در محیط فعلی، در حال توسعه می‌باشند. سیستم‌های اطلاعاتی حسابداری، جدا از ابزار و بسترهای مختلف برای پردازش داده‌ها، پست‌ها و گزارش نتایج به بازار، یکی از مهمترین موضوعات برای حرفه و تجارت در مقیاس بزرگ بوده‌اند و خواهند بود. درک علت این موضوع همراه با زمینه‌ی فعلی محیط مد نظر یک گام منطقی برای شروع تحلیل و بررسی این موضوع می‌باشد.

متخصصان حسابداری این مجموعه از مهارت‌ها و توانایی‌های ذهنی را از قبل دارند تا بتوانند داده‌های کمی را بررسی، تحلیل کرده و آن‌ها را به بخش‌های درون و بیرون سازمان انتقال و گزارش دهند. اطلاعات مالی به صورت ذاتی کمی هستند اما باید به این نکته اشاره کرد که واقعا تمام بخش‌های داده‌های عملکردی، حتی داده‌های غیر ساختاری نیز مشخصه‌های کمی دارند. با بررسی دقیق این موضوع و اعمال تکنیک‌های مختلف مورد استفاده در تحلیل داده‌های مالی به نظر می‌رسد که متخصص‌های حسابداری به صورت خاص در جایگاهی قرار دارند که باید مقادیر گسترده‌ای از اطلاعات را بررسی کنند که توسط سازمان‌ها ایجاد شده است. همچنین در واقع موضوع کاربردی‌تر برای متخصص‌های حسابداری که در فعالیت‌های صنعتی و عمومی شاغل هستند این است که بهبودهای ایجاد شده در تکنولوژی را می‌توان در بسیاری از روندهای غیر مالی صنعتی مورد استفاده کرد. ادغام این بهبودها چه از نظر توانایی‌های ذهنی و چه فعالیت‌های واقعی، به صورت مستقیم با مفاهیم و اجرای عملکردهای پیوسته‌ی حسابداری در شرکت‌های عمومی و یا شرکتهای صنعتی ارتباط دارد [۱].

با توجه به گسترش سیستم‌های پایگاه داده و حجم بالای داده‌های ذخیره شده در این سیستم‌ها نیاز به ابزاری است تا بتوان داده‌ها را پردازش کرده و اطلاعات حاصل از این پردازش را در اختیار کاربران قرار داد. افزایش پیچیدگی روزافزون سیستم‌های اطلاعاتی، حجم عظیم داده‌های مورد نیاز و هم‌چنین پیچیدگی مبادلات و معاملات تجاری شرکت‌ها، باعث شده تا حساب‌رسان قادر به استفاده از روش‌های سنتی به منظور اظهار نظر مناسب نسبت به اطلاعات ارائه شده نباشند؛ بنابراین مسأله‌ای که در این پیچیدگی سیستم‌های اطلاعاتی و حجم داده‌های فراوان برای تمامی استفاده‌کنندگان از جمله حساب‌رسان وجود دارد این است که بتوان از میان این حجم عظیم داده خام دانش مورد نظر را استخراج نمود. در این زمینه تکنیک‌هایی برای کشف دانش به وجود آمده‌اند. یکی از ابزارهایی که در عصر اطلاعات می‌تواند حساب‌رسان را یاری کند، استفاده از روش‌ها و تکنیک‌های داده کاوی است. داده کاوی یک فرایند ثابت است که با کاربرد تکنیک‌های کامپیوتری، داده‌های مطلوب و دارای بار اطلاعاتی را استخراج و آنها را مورد پردازش قرار می‌دهد. در بخش بعدی علم داده کاوی کاملاً تحلیل و شرح شده است [۲۰].

## ۳- داده کاوی

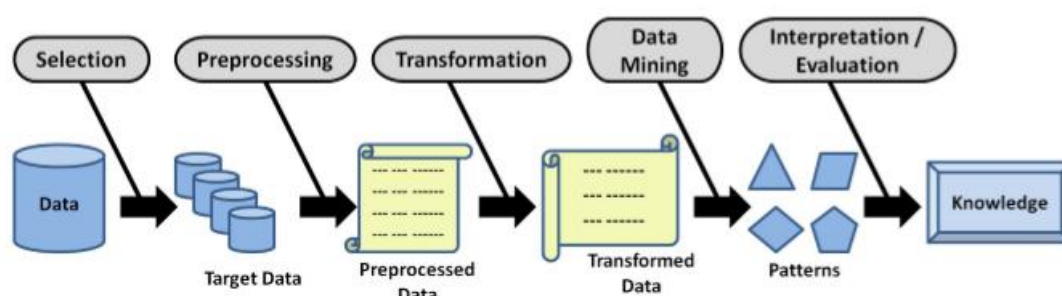
داده کاوی فرآیند اکتشاف الگوها و اطلاعات مخفی از داده‌های موجود می‌باشد. تفاوت بین داده در پایگاه‌های داده و یک انبار داده این است که در یک پایگاه داده، داده‌ها به شکل ساخت یافته نگهداری می‌شوند در حالیکه در یک انبار داده‌ای، داده‌ها ممکن است به شکل ساخت یافته ارائه گردند و یا نه. ساختار داده‌ها می‌تواند طوری تعریف گردد که

آنها را برای پردازش سازگار نماید. از این رو در داده کاوی؛ باید اساساً بر روی پاکسازی داده‌ها تمرکز کرد به طوری که پردازش بعدی آنها ممکن باشد. فرآیند پاکسازی داده‌ها به عنوان حذف نویز یا کاهش نویز نیز نامیده می‌شود [۲]. تکنیک‌های داده کاوی برای تبدیل داده‌های خام به اطلاعات یا دانش مفید استفاده می‌شود. داده‌ها به خودی خود هیچ چیز نیستند اما پردازش آنها بسیار مفید و جالب است. بسیاری از فناوری‌های پیشرفته که به طور هوشمندانه از داده‌ها به عنوان اطلاعات مفید استفاده می‌کنند، وجود دارد [۳]. امروزه داده کاوی توسط اکثر شرکت‌هایی با تمرکز بسیار قوی بر مصرف‌کننده، سازمان‌های خرده فروشی، مالی، ارتباطی و بازاریابی، برای "بررسی دقیق" داده‌های معاملاتی خود و تعیین قیمت‌گذاری، تنظیمات مشتری و اطلاعات مربوط به محصول، تأثیر بر فروش، رضایت مشتری و سود شرکت مورد استفاده قرار می‌گیرند. با کمک داده کاوی، یک خرده فروش می‌تواند از سوابق خرید و فروش مشتری برای تولید محصولات و تبلیغات برای جلب نظر بخش‌های خاص مشتری استفاده کند [۴].

داده‌ها ماده خام اطلاعاتی هستند که می‌توانند به عنوان واقعیات، اعداد یا متنی قابل پردازش توسط ماشین باشند. در واقع اطلاعات داده‌هایی‌اند که معنای جدیدی از ارتباطات رابطه‌ای به آنها داده شده‌است. برای تجزیه و تحلیل روند فروش سالهای خاص، می‌توان از داده‌های جمع‌آوری شده از معاملات فروش استفاده کرد. دانش استفاده از داده‌ها و اطلاعات است که می‌توان آن را آگاهی عمومی از اطلاعات، حقایق، ایده‌ها، حقیقت یا اصل دانست. در میان طیف گسترده‌ای از زمینه‌ها، داده‌ها با سرعت چشمگیری جمع‌آوری می‌شوند. در یک سطح انتزاعی، حوزه KDD مربوط به توسعه روش-ها و تکنیک‌ها برای ایجاد معنی داده است. حوزه KDD برای توسعه روش‌ها و تکنیک‌ها جهت ایجاد معنی (منطقی کردن) داده نگران کننده‌است. مشکل اساسی که توسط فرآیند KDD برطرف شده نقشه‌برداری از داده‌های سطح پایین به شکل‌های دیگر است که ممکن است جمع و جورتر، انتزاعی‌تر یا مفیدتر باشد. در هسته این فرآیند استفاده از روش‌های خاص داده کاوی برای کشف و استخراج الگو وجود دارد [۴]. کشف دانش و داده کاوی به دلیل تقاضای اخیر در حال افزایش برای تکنیک‌های KDD، از جمله مواردی که در کسب دانش، یادگیری ماشین، پایگاه‌های داده، آمار، تجسم داده‌ها و محاسبات با کارایی بالا مورد استفاده قرار می‌گیرد، به موضوعات دارای اهمیت در حال رشد تبدیل شده‌اند. کشف دانش و داده کاوی می‌تواند برای حوزه هوش مصنوعی در بسیاری از زمینه‌ها، به عنوان مثال آموزش، صنعت، تجارت، دولت و غیره، بسیار مفید باشد [۴].

### ۳-۱ مراحل کشف دانش

کشف دانش در پایگاه داده فرآیند استخراج خروجی مورد نیاز در قالب‌های مختلف از داده‌های خام است. همچنین KDD به عنوان فرایندی برای مشاهده الگوهای مفید در داده‌ها تعریف می‌شود [۳]. نمودار عمومی و رایج داده کاوی یا KDD در شکل ۱ نشان داده شده‌است.



شکل ۱: مراحل کشف دانش

داده کاوی آموزشی عمده ترین کاربرد تکنیک های داده کاوی است و برای تولید چنین مجموعه داده هایی که تصمیم کاربران و مدیران را به طور خودکار انجام می دهد، استفاده می شود. این برنامه به عنوان یک برنامه نوظهور در سازمان ها ذکر شده که انواع مختلفی از داده های تولید شده توسط این سازمان ها را کاوش می کند. این داده ها از طریق سیستم آموزشی تولید، تجزیه و تحلیل شده و یادگیری و اثرات آموزشی را بهبود می بخشد [۴]. روند KDD یک توالی تکراری و تعاملی از مراحل اصلی زیر است [۵]:

**انتخاب<sup>۱</sup>:** هدف اصلی آن ایجاد یک مجموعه داده هدف از داده های اصلی است. به عنوان مثال، انتخاب زیرمجموعه ای از متغیرها یا نمونه داده ها که باید کشف شوند.

**پیش پردازش<sup>۲</sup>:** هدف آن "پاک کردن" داده ها با انجام عملیات مختلف، از جمله مدل سازی و حذف نویز، تعیین استراتژی های مناسب برای مدیریت موارد از دست رفته داده، محاسبه اطلاعات توالی زمان است.

**تبدیل:** وظیفه کاهش و فرافکنی داده ها را دارد تا بتواند نمایشی مناسب برای وظیفه خاص بدست آورد. این به طور معمول با درگیر کردن تکنیک ها یا روش های تبدیل که قادر به یافتن بازنمایی های ثابت داده ها هستند، انجام می شود.

**داده کاوی:** این بخش با استخراج الگوهای جالب به وسیله انتخاب سروکار دارد. ۱: یک روش یا کار خاص داده - کاوی (به عنوان مثال، جمع بندی، طبقه بندی، خوشه بندی، رگرسیون و غیره)، ۲: الگوریتم های مناسب برای انجام وظیفه مورد نظر و ۳: نمایش مناسب نتایج خروجی.

**تفسیر/ارزیابی:** توسط کاربر برای تفسیر و استخراج دانش از الگوهای استخراج شده با تجسم الگوها مورد سو استفاده قرار می گیرد. این تفسیر به طور معمول با تجسم الگوها، مدل ها یا داده های داده شده در این مدل ها انجام می شود و در صورت تکرار مجدد مراحل قبلی فرآیند انجام می شود.

داده کاوی نشان دهنده گام "اصلی" فرآیند KDD است، به طوری که اصطلاحات "داده کاوی" و "KDD" اغلب به عنوان مترادف تلقی می شوند. چندین تعریف از داده کاوی عنوان شده است، مانند: "استخراج خودکار و در عین حال غیر پیش پا افتاده اطلاعات ضمنی، از قبل ناشناخته و بالقوه مفید از داده ها"، "اکتشاف خودکار و تجزیه و تحلیل مقدار زیادی از داده ها به منظور کشف الگوهای معنی دار". همه تعاریف تقریباً با یکدیگر برابر هستند. همه آنها در مورد جنبه های اصلی داده کاوی توافق داشته که عبارتند از: حجم عظیمی از داده ها که باید مورد تجزیه و تحلیل قرار گیرند، استخراج آنچه "دانش" یا "اطلاعات مفید" یا "الگوها" نامیده می شود، یعنی چیزی که توسط انسان قابل پردازش و بهره برداری احتمالی باشد.

اهمیت داده کاوی امروزه عمدتاً توسط بسیاری از داده های جمع آوری شده و ذخیره شده توسط برنامه های برجسته امروزی ایجاد می شود. این داده ها شامل داده های وب، داده های تجارت الکترونیکی، خریدها، معاملات بانکی و غیره است. همچنین، تعداد برنامه های کاربردی با داده هایی که باید با سرعت های بسیار زیاد پردازش شوند (GB / ثانیه یا حتی بیشتر) به سرعت در حال افزایش هستند. این نمونه ها شامل سنسورهای از راه دور روی ماهواره، تلسکوپ هایی که آسمان ها را اسکن می کنند، داده های بیان ژن<sup>۳</sup> برای تولید ریزآرایه و شبیه سازی های علمی است. با توجه به ویژگی داده های اساسی، آشکار است که تجزیه و تحلیل داده ها در چنین زمینه های چالش برانگیزی را نمی توان با روش های

<sup>1</sup> Selection

<sup>2</sup> preprocessing

<sup>3</sup> gene-expression data

سنتی تجزیه و تحلیل داده‌ها، دستی یا خودکار انجام داد. داده‌کاوی با توجه به ماهیت میان رشته‌ای ذاتی خود که آن را در تقاطع تعدادی از کلاس‌های کلاسیک‌تر، مانند هوش هنری، آمار، سیستم‌های پایگاه داده، یادگیری ماشین ایجاد می‌کند، این شکاف (چالش) را برطرف می‌کند.

### ۲-۳ تکنیک‌های داده‌کاوی

یک پایگاه داده بزرگ/یا انبار داده ممکن است انواع مختلفی از الگوهای ناشناخته را در خود داشته باشد. برای استخراج این تنوع از الگوهای ناشناخته، می‌توان از انواع متمایز عملکرد داده‌کاوی، روش‌ها و تکنیک‌ها استفاده کرد. بر اساس انواع مختلف الگوها، توابع داده‌کاوی را می‌توان به صورت خلاصه‌سازی، خصوصیات و تبعیض، طبقه‌بندی، تحلیل رگرسیون و روند، خوشه‌بندی، تجزیه و تحلیل دور، ارتباط و غیره دسته‌بندی کرد [۶].

**خلاصه‌سازی:** خلاصه‌سازی مجموعه کوچکتر را نتیجه داده و خلاصه‌ای از داده‌های دقیق بر اساس سلسله مراتب مفهوم را ارائه می‌دهد. معمولاً خلاصه‌سازی با استفاده از تجمعی انجام می‌شود که می‌تواند به سطوح مختلف انتزاع گسترش یافته و از زوایای مختلف قابل مشاهده باشد. انواع مختلفی از الگوها را می‌توان بر اساس ترکیب سطوح مختلف انتزاع و ابعاد مختلف استخراج کرد. خلاصه‌سازی داده‌ها معمولاً با استفاده از رویکرد القایی وابسته و روش مکعب داده انجام می‌شود.

**خصوصیات و تبعیض:** خصوصیات در واقع خلاصه‌ای از سلسله مراتب مفهوم مبتنی بر داده است و قوانین خصوصیات را ایجاد می‌کند. از طرف دیگر، برای شناسایی انواع مختلف مجموعه داده‌ها از تبعیض استفاده می‌شود. خروجی تبعیض در قالب قوانین تبعیض آمیز تولید می‌شود.

**طبقه‌بندی:** فرآیند طبقه‌بندی مشاهدات جدید براساس کلاس‌های از پیش تعیین شده، یعنی یادگیری تحت نظارت است. از الگوریتم‌های طبقه‌بندی برای پیش‌بینی کلاس‌های داده استفاده می‌شود. مجموعه بزرگی از الگوریتم‌های طبقه‌بندی (یا طبقه‌بندی‌ها) تاکنون توسط محققان ارائه شده است.

**خوشه‌بندی (تجزیه و تحلیل خوشه‌ای):** خوشه‌بندی برای تقسیم یا بخش‌بندی اشیا داده (یا مشاهدات) به زیر مجموعه‌هایی که به عنوان گروه یا خوشه نامیده می‌شوند، استفاده می‌شود. تجزیه و تحلیل خوشه یکی از محبوب‌ترین تکنیک‌ها است که نه تنها در داده‌کاوی بلکه در دامنه‌های دیگری مانند آمار، تقسیم بندی تصویر، تشخیص الگو، تشخیص شی، بازیابی اطلاعات، بیوانفورماتیک و غیره نیز مورد استفاده قرار می‌گیرد. در دو دهه گذشته بسیاری از محققان مجموعه بزرگی از الگوریتم‌های خوشه‌بندی را پیشنهاد داده‌اند. همچنین بسیاری از محققان مجموعه‌ای از الگوریتم‌های خوشه‌بندی جدید مانند روش بدون پارامتر با استفاده از حداقل طول توصیف را ارائه/یا بحث کرده‌اند.

**تجزیه و تحلیل داده‌های خارج از محدوده<sup>۴</sup> (پرت):** اشیا داده‌ای که در رفتار عمومی داده‌ها متفاوت هستند، به عنوان دور از دسترس (خارج از محدوده) خوانده می‌شوند. اکثر روشهای داده‌کاوی، داده‌های پرت (داده‌های خارج از محدوده) را به عنوان نویز یا موارد استثنایی کنار می‌گذارند. گاهی اوقات، ممکن است داده‌های پرت در مقایسه با سایر اشیا داده‌ها اطلاعات بیشتری داشته باشند؛ بنابراین تجزیه و تحلیل داده‌های پرت برای برخی از زمینه‌های کاربردی مانند تشخیص نفوذ، تشخیص تقلب، تشخیص ناهنجاری و غیره مهم است. بسیاری از تکنیک‌های داده‌کاوی به طور کلی از خوشه‌بندی برای شناسایی داده‌های پرت به عنوان یک نویز استفاده می‌کنند. روش‌های شناسایی داده‌های پرت می‌توان

<sup>4</sup> Outlier analysis

به عنوان روش‌های طبقه‌بندی شده، روش‌های آماری، روش‌های خوشه‌بندی، روش‌های نظارت شده، نیمه نظارت شده و بدون نظارت، روش‌های مبتنی بر انحراف و روش‌های نزدیکی طبقه‌بندی کرد.

**تجزیه و تحلیل انجمن (یا استخراج انجمن):** تجزیه و تحلیل انجمن، ارتباطات (یا پیوندها) را در میان مجموعه داده‌ها کشف کرده و اشیاء داده را که می‌تواند به طور جمعی با حداقل آستانه‌های پشتیبانی و اطمینان محقق شود، مشخص می‌کند. شناسایی کلیه مجموعه‌های مکرر اقلام و به دنبال آن ایجاد قوانین قوی ارتباط در استخراج معادن انجام می‌شود. تجزیه و تحلیل انجمن شامل استخراج اقلام مکرر، متعلقات و زیر ساختها است. تجزیه و تحلیل سبد بازار عمدتاً با استفاده از تجزیه و تحلیل ارتباط است. الگوریتم Apriori به طور گسترده برای ارتباط استفاده می‌شود. الگوریتم‌های تجزیه و تحلیل انجمن را می‌توان به الگوریتم‌های کلاسیک، الگوریتم‌های نمایشی متراکم و الگوریتم‌های مجموعه ناقص طبقه‌بندی کرد.

**رگرسیون و تحلیل روند (یا تحلیل تکامل):** رگرسیون مقدار ویژگی را بر اساس تکنیک(های) رگرسیون در طول زمان پیش‌بینی می‌کند. مقادیر آینده متغیرها با کمک نمودار سری زمانی تاریخی پیش‌بینی می‌شود. تجزیه و تحلیل روند (که به آن تجزیه و تحلیل تکامل نیز گفته می‌شود) الگوهای جالبی را در تاریخ تکامل اشیاء کشف می‌کند. شناسایی الگوها در تکامل یک شی و تطبیق روند تغییر اشیاء دو جنبه اصلی تجزیه و تحلیل روند است. روند اشیاء که رفتار آنها با گذشت زمان تکامل می‌یابد، می‌تواند با استفاده از مدل تجزیه و تحلیل روند و رگرسیون توصیف شود. تجزیه و تحلیل روند متغیر زمان از اشیاء داده را در مجموعه داده ظاهر می‌کند. از تجزیه و تحلیل ارتباط می‌توان برای تحلیل تکامل نیز استفاده کرد.

#### ۴- داده‌کاوی در حسابداری و حسابرسی

در طول سالهای اخیر، به تدریج این واقعیت رشد یافته که حجم عظیمی از داده‌ها وجود دارند که کامپیوترها، شبکه‌ها و در حقیقت تمام زندگی را فرا گرفته‌اند. سازمانهای دولتی، مؤسسات علمی و تجاری سرمایه‌هنگفتی را برای گردآوری و ذخیره این داده‌ها اختصاص داده‌اند. حجم اطلاعات جمع‌آوری و ذخیره شده رشد سریعی به خود گرفته است. برخی از تحقیقات نشان می‌دهند که میزان داده‌های جهانی تقریباً هر سال دو برابر می‌شوند، در همان حال نیز هزینه ذخیره‌سازی این داده‌ها به طور قابل توجهی از یک دلار برای هر مگابایت به چند «پن» کاهش یافته است. به گونه‌ای مشابه، توان محاسباتی در هر ۱۸ الی ۲۴ ماه دو برابر شده، در حالی که هزینه مرتبط با این توان محاسباتی در حال کاهش می‌باشد. با وجود این، فقط مقدار کمی از این داده‌ها مورد استفاده قرار می‌گیرند، زیرا در بسیاری از موارد، حجم داده‌های لازم برای سازماندهی، بسیار بالا بوده یا ساختار آنها به منظور تجزیه و تحلیل مؤثر و کارا به شدت پیچیده می‌باشد. با ورود به عصر اطلاع‌رسانی دیجیتال، مسئله انفجار داده‌ها، اطلاعات و ضرورت نگهداری آنها هر روز بیش از پیش احساس می‌شود. تحلیل داده‌ها می‌تواند دانش مضاعفی در مورد تجارتی خاص ایجاد کند. این مهم با افزایش شفافیت داده‌ها به منظور استخراج دانش، دست‌یافتنی خواهد بود. مجموعه داده‌ها در شکل خام از ارزش کمی برخوردار هستند، آنچه که در حقیقت ارزشمند است دانش و شناختی است که می‌تواند از داده‌ها حاصل شود و مورد استفاده قرار گیرد. به لحاظ نظری، داده‌های حجیم و بزرگ می‌توانند به نتایج قویتری منجر شوند. جامعه تجاری به خوبی از سر ریزی و حجم زیاد اطلاعات امروز آگاه است.

در این شرایط، داده کاوی ابزاری سودمند است که در اختیار تمام شرکتها می باشد. داده کاوی و کشف دانش، ابزارهایی هوشمند هستند که به جمع آوری و پردازش داده ها و بهره مندی از آنها منجر می شوند. در واقع، داده کاوی شامل استفاده از ابزارهای تحلیلی پیچیده به منظور کشف الگوهای معتبر و ناشناخته و وابستگیهای موجود در مجموعه داده های حجیم می باشد. نتیجه اینکه، فناوری داده کاوی چیزی بیش از جمع آوری و مدیریت داده بوده و شامل تجزیه و تحلیل و پیش بینی نیز می باشد. اهمیت داده کاوی توسط بسیاری از سازمان های حرفه ای تشخیص داده شده است. انجمن حسابداران رسمی آمریکا، داده کاوی را به عنوان یکی از ده فناوری برتر برای آینده معرفی کرده است. همچنین انجمن حسابسان داخلی آمریکا نیز این فناوری را در فهرست یکی از چهار اولویت تحقیقاتی خود گنجانده است.

کاربرد احتمالی داده کاوی در حسابداری مالی منجر به استفاده از صورتهای مالی به جهت تعیین سودآوری، تطابق نسبت ها با میانگین صنعت و بررسی اثربخشی یک کسب و کار می گردد. در حسابداری مدیریت نیز داده های واحدهای تجاری به منظور تعیین چگونگی عملیات روزمره واحد تجاری، تجزیه و تحلیل بهای تمام شده و سودآوری بخش های مختلف، مورد استفاده قرار می گیرد. در حسابرسی، ابزارهای داده کاوی برای تجزیه و تحلیل داده های مرتبط با احتمال تقلب مدیریت می توانند به کار گرفته شوند. بدین ترتیب داده کاوی در تمام جنبه های حسابداری سودمند خواهد بود، زیرا حسابداری عبارت از جمع آوری، تجزیه و تحلیل و استفاده از اطلاعات مالی و غیرمالی به منظور پردازش برای اشخاص ذینفع است. شرکت های تجاری می توانند از داده کاوی به منظور شناسایی فرصت های مالی و عملیاتی قابل توجه برای اطمینان از کارایی فرآیند زنجیره تامین بهره مند شوند [۱۹].

داده کاوی یکی از شاخه های علم کامپیوتر محسوب می شود که از مفاهیم و روش های مرتبط با هوش مصنوعی، شناسایی الگو، سیستم های پایگاه داده و علم آمار، پا به عرصه گذاشت. اصلی ترین دلیلی که باعث شد داده کاوی کانون توجهات در صنعت اطلاعات قرار بگیرد، مساله در دسترس بودن حجم وسیعی از داده ها و نیاز شدید به این است که از این داده ها اطلاعات و دانش سودمند استخراج شود. اطلاعات و دانش بدست آمده در کاربردهای وسیعی از مدیریت کسب و کار و کنترل تولید و تحلیل بازار تا طراحی مهندسی و تحقیقات علمی مورد استفاده قرار می گیرد. داده کاوی را می توان حاصل سیر تکاملی طبیعی تکنولوژی اطلاعات دانست که این سیر تکاملی ناشی از یک سیر تکاملی در صنعت پایگاه داده نظیر عملیات: جمع آوری داده ها و ایجاد پایگاه داده، مدیریت داده و تحلیل و فهم داده ها می باشد. تکامل تکنولوژی پایگاه داده و استفاده فراوان آن در کاربردهای مختلف سبب جمع آوری حجم فراوانی داده شده است [۲۱].

#### ۴-۱ بررسی یک چارچوب سازماندهی برای کاربردهای داده کاوی در حسابداری

حسابداری سنگ بستر هر کاری بوده و طیف گسترده ای از وظایف از جمله گزارش دهی داخلی و خارجی، هزینه یابی، تخمین، ارزشیابی، تجزیه و تحلیل و حسابرسی را شامل می شود. بیشتر این وظایف مستلزم حجم بالایی از عدم قطعیت و پیچیدگی های خطر هستند. حسابداری دارای تاریخچه ی کاربرد های هوشمندی است که به بیشتر از سه دهه گذشته باز می گردد و در واقع یکی از ابتدایی ترین رشته های بازرگانی است که روش داده کاوی را برای مطرح کردن بهتر این خطر و پیچیدگی ها بکار می گیرد [۱۱]. در این بخش یک روش اثباتی، چارچوب سازماندهی برای کاربردهای داده کاوی در حسابداری معرفی شده است. چارچوب هایی که ادبیات موضوع را بطور شفاف سازماندهی می کنند توپولوژی ادبیات را خلاصه کرده، باعث درک بهتر روابط پیچیده شده و مکانیزم مناسبی برای تحقیق در نقشه سازی در یک حوزه ی خاصی ارائه می کنند. در این قسمت متدولوژی همجواری سازی مولفه های هویت های مختلف برای ساختار چارچوب مشابه با

رویکرد مورد استفاده توسط پریس و رومباخ و ریچارد سون و همکارانش اقتباس شده است. پریس و رومباخ چارچوبی را با ترکیب رویکردهای سنجش رشته‌های مهندسی نرم‌افزار و تعامل انسان و کامپیوتر خلق کردند، در حالی که ریچارد سون و همکارانش این چارچوب را با پیوند دادن مولفه‌های یک ماهیت حرفه‌ای (حسابدار) با مولفه‌های ماهیت تخصص (حسابداری) تشکیل دادند. چارچوب معرفی شده این قسمت ویژگی‌های رشته حسابداری را با ویژگی‌های رشته داده-کاوی ادغام می‌کند. این چارچوب، مخصوصاً دو دیدگاه گزارش دهی (آینده نگرانه و گذشته نگرانه) و سه هدف اصلی داده کاوی (توصیف، پیش بینی و تجویز) را با هم ادغام می‌کند.

دوگانگی گذشته نگرانه-آینده نگرانه‌ی گزارش دهی در حسابداری به خوبی در آثار برینبرگ، کارنگی و اوون به چشم می‌خورد. گزارش گذشته نگرانه عمدتاً به ارائه گزارش انعکاسی از وضعیت مالی تاریخی سازمان عمدتاً از لحاظ ارزش-گذاری مالی، تصمیم‌گیری و یا اهداف قبول اجابت می‌پردازد. به فرض مثال، تهیه صورتحساب‌های مالی خلاصه گذشته نگرانه‌ای از موقعیت مالی یک سازمان در نقطه‌ای از زمان (تراز نامه) یا سود و زیان برای یک دوره زمانی (صورتحساب درآمد) فراهم می‌کند. از طرف دیگر، گزارش آینده نگرانه همسو با آینده و شامل چشم اندازهای مالی آتی، تخمین‌ها و پیش بینی‌ها می‌باشد. بطور مثال، هرگاه اطلاعات تاریخی برای پیش‌بینی بعضی ابعاد آتی سازمان مورد استفاده قرار بگیرند، از قبیل جهت یا دامنه عملکرد آتی، رشد یا عملکرد مالی و یا شاخص سلامت، کانون گزارش حسابداری آینده نگرانه می‌شود [۷، ۱۲].

#### ۴-۲ سه هدف اصلی داده کاوی در چارچوب معرفی شده

داده کاوی دارای سه هدف اصلی توصیف، پیش بینی و تجویز است. هدف اصلی داده کاویی توصیفی درک تجارت و داده‌ها و آن چه که اتفاق افتاده است، می‌باشد [۷].

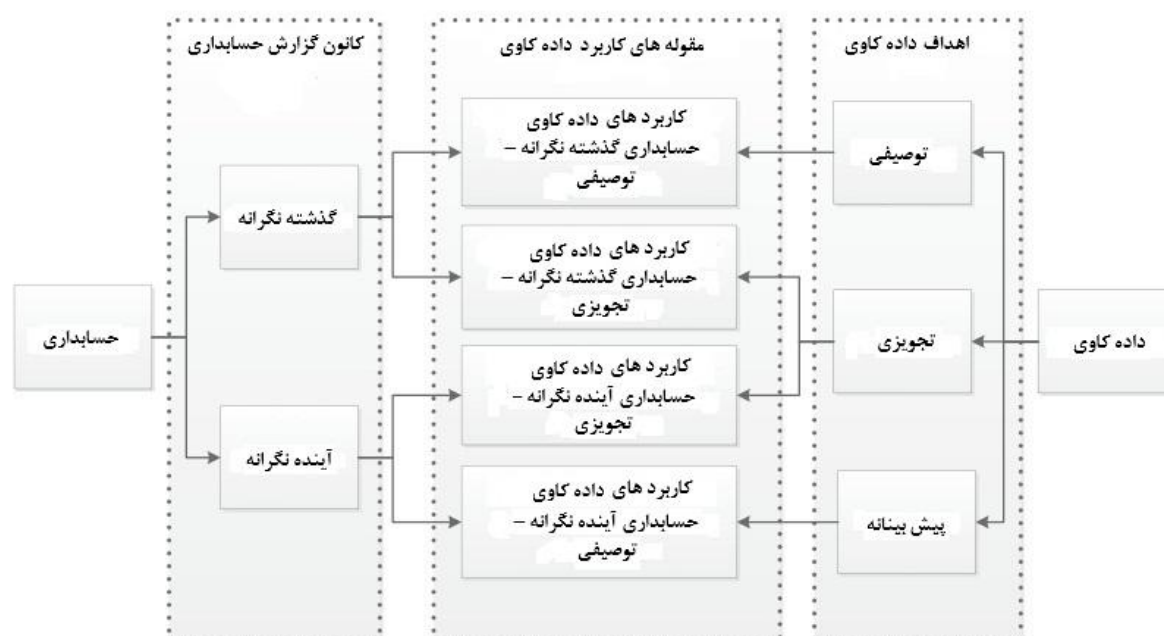
هدف داده کاویی پیش‌بینانه استفاده از زمان گذشته برای درک آینده است (آن چه که ممکن است اتفاق بیفتد). هدف داده کاوی تجویزی دستیابی به بهترین نتیجه (آن چه که باید اتفاق بیفتد) می‌باشد. داده کاویی توصیفی متداول‌ترین و قابل درک‌ترین نوع استفاده از داده‌ها برای درک گذشته و حال و متعاقباً اتخاذ تصمیمات آگاهانه می‌باشد. از تکنیک‌هایی برای مقوله‌بندی، تعیین ویژگی‌ها، منسجم‌سازی و تجسم داده‌ها و تبدیل آن‌ها به اطلاعات مفیدی در جهت درک بهتر داده‌ها و تجارت استفاده می‌کند. داده کاویی توصیفی کاربران را قادر می‌سازد تا الگوها در روندهای مربوط داده‌ها را شناسایی و مسائل و یا زمینه‌های فرصت را اکتشاف کنند. از طرف دیگر، داده کاوی پیش‌بینانه گذشته را با هدف پیش‌بینی آینده از طریق بررسی داده‌های تاریخی، آشکارسازی الگوهای روابط در این داده‌ها و استنباط روابط در آینده تحلیل می‌کند. بطور مثال، نظام بانکی با استفاده از این روش به مشتریان کارت‌های اعتباری در رابطه با شارژ احتمالاً جعلی هشدار می‌دهد. این روش در وهله اول اطلاعاتی در زمینه تصمیم‌گیری همسو با آینده ارائه می‌کند. نهایتاً داده-کاوی تجویزی از تکنیک‌های بهینه‌سازی برای شناسایی بهترین شیوه‌ها برای کمینه‌سازی یا بیشینه‌سازی توابع هدفمند استفاده می‌کند [۷، ۱۳].

تکنیک‌های آماری و ریاضیاتی این روش را می‌توان با بهینه‌سازی ادغام کرد تا تصمیماتی در رابطه با عدم قطعیت داده‌ها گرفته شود. هدف خواه گزارش گذشته‌نگرانه و خواه گزارش آینده‌نگرانه باشد، می‌توان روش تجویز (بهینه‌سازی) را برای پشتیبانی از کمینه یا بیشینه‌سازی این اهداف به روشی به کار برد که از لحاظ استفاده از این منابع بسیار کارآمد باشد. در حقیقت داده کاوی تکنیک‌های پیشرفته‌ای ارائه می‌کند تا مدلسازی توصیفی، پیش‌گویانه و تجویزی تسهیل

شود، بنابراین تا حد زیادی از دیدگاه‌های اصلی گزارش‌دهی حسابداری حمایت مستقیمی به عمل می‌آورد. با کنار هم قرار دادن دو دیدگاه اصلی گزارش حسابداری و سه هدف اصلی داده کاوی، شش ترکیب حاصل می‌شود. تنها چهار مورد از این ترکیبات از لحاظ منطقی عملی هستند.

- داده کاویی توصیفی در گزارش گذشته نگرانه
- داده کاویی تجویزی در گزارش گذشته نگرانه
- داده کاویی تجویزی در گزارش آینده نگرانه
- داده کاویی پیش بینانه در گزارش آینده نگرانه.

این چهار ترکیب نماد گروه‌بندی‌های اصلی چارچوب پیشنهادی بوده و برای سازماندهی تحقیقات انتشار یافته در کاربردهای حسابداری داده کاوی مورد استفاده قرار می‌گیرد. در این قسمت مقوله‌های شکل گرفته با این ترکیبات را به ترتیب گذشته نگرانه - توصیفی، گذشته نگرانه - تجویزی، آینده نگرانه - تجویزی و آینده نگرانه - پیش بینانه نامیده شده است (شکل ۲).

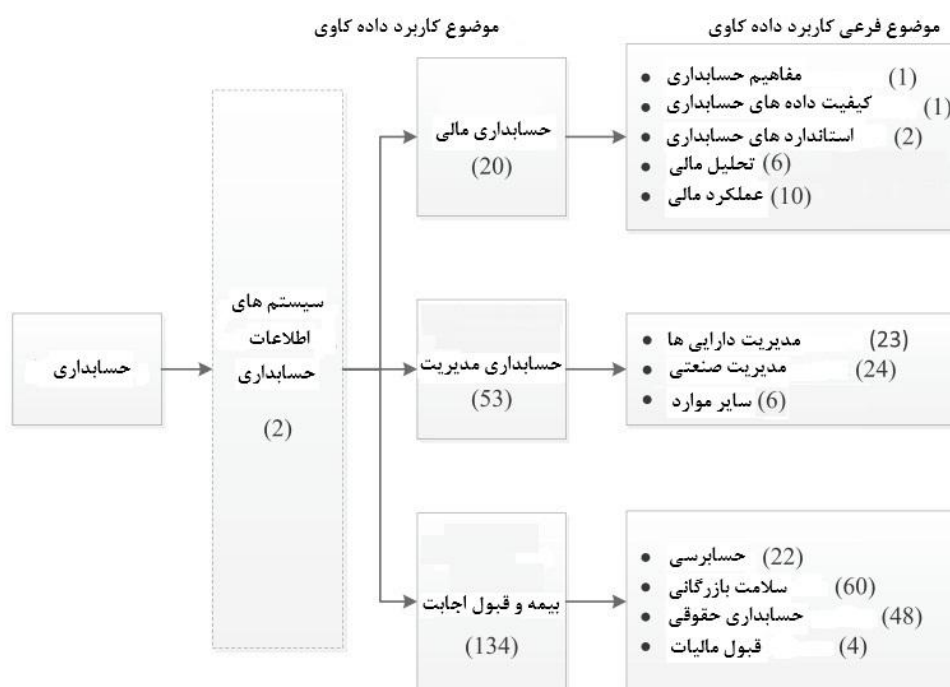


شکل ۲. چارچوب سازماندهی برای کاربردهای داده کاوی در حسابداری [۷].

کاربردهای گذشته نگرانه - توصیفی بر ادراک تجارت و داده‌ها از دیدگاه تاریخی متمرکز هستند. وظایف اصلی داده - کاوی که توسط این گروه از کاربردها اجرا می‌شوند شامل اکتشاف، خوشه‌بندی، تحلیل تداعی، تجسم‌سازی، تقسیم‌سازی و تشخیص الگو هستند. کاربردهای گذشته نگرانه - تجویزی و آینده نگرانه - تجویزی بر کارایی تاکید دارند، بنابراین بهینه‌سازی و تخمین را به عنوان وظایف مربوط به داده کاوی به کار می‌گیرند، با این حال از لحاظ جهت‌گیری زمانی متفاوت از هم می‌باشند. در این جا کاربردهای گذشته نگرانه - توصیفی بر گذشته و حال و کاربردهای آینده نگرانه - توصیفی بر آینده متمرکز هستند. نهایتاً کاربردهای آینده نگرانه - پیش بینانه بر بعد آتی تجارت با استفاده از داده - های تاریخی متمرکز بوده و وظایفی از قبیل طبقه‌بندی، پیش بینی و تخمین را اجرا می‌کنند. چارچوب پیشنهادی جامع، ساده، درک آن آسان و از لحاظ تجربی قابل تایید می‌باشد [۷].

### ۳-۴ معرفی کاربردها در موضوعات اصلی و فرعی حسابداری

شکل ۳ تعداد کاربردها در موضوعات اصلی و فرعی حسابداری و در بخش‌هایی را خلاصه می‌کند که جزئیات بیشتری در مورد هر کدام از زمینه‌های کاربرد ارائه می‌کنند. شدت متغیر استفاده از داده کاوی در شاخه‌های مختلف حسابداری منعکس کننده شدت نیاز به تحلیل پیشرفته در هر کدام از این شاخه‌ها می‌باشد. ناکامی‌های حسابرسی که به اطلاع عموم رسیده‌اند، ورشکستگی‌ها و وضع جدید قوانین نظارتی و نظارت و بازرسی باعث شده است تا جستجوی فناوری پیشرفته در حوزه بیمه و قبول اجابت ضروری به نظر برسد. فشارهای رقابتی و پیگیری کارآمدی‌های شرکتی باعث ضرورت بیشتر داده کاوی در حسابداری مدیریت در مقایسه با حسابداری مالی شده‌است. در شکل ۳ ارقام داخل پرانتز نشان دهنده‌ی تعداد کاربردهای داده کاوی در حوزه مرتبط می‌باشند [۷].



### تصویر ۳: کاربرد داده کاوی در موضوعات اصلی و فرعی حسابداری

**داده کاوی در حسابداری مالی:** کاربردهای حسابداری مالی عمدتاً تحلیل و عملکرد مالی را بررسی می‌کنند. داده-کاوی در حسابداری مالی در وهله اول بر عملکرد مالی و تحلیل نسبت، پیش بینی درآمد سه ماهه، مقایسه ارزش اطلاعاتی در مقابل داده‌های عددی و متنی برای سنجش عملکرد، شاخص گذاری عملکرد مالی، شناسایی عوامل خطر در گزارش‌های سالیانه، تاثیر اظهاریه‌های مدیریت و نحوه بیان آن‌ها در پاسخ به بازار و ... متمرکز می‌باشد. این کاربردها عمدتاً بر توصیف و پیش‌بینی به عنوان اهداف متمرکز بوده و از خوشه‌بندی و طبقه‌بندی به عنوان وظایف داده-کاوی داده‌ها استفاده می‌کنند. شبکه‌های عصبی و متن کاوی متداول‌ترین تکنیک‌های این کاربردها هستند. فرصت‌های تحقیقات آتی شامل بکارگیری مولفه‌های متنی گزارش‌های مالی در پیش‌بینی عملکرد مالی، اهمیت تخصص رشته‌ای در کاربردهای داده‌ها، توجه بیش‌تر به مسائل کیفیت داده‌ها، اهمیت شاخص گذاری کاربردها، فراتر رفتن از نسبت‌های مالی برای پی بردن به ورودی‌های متناسب برای پیش‌بینی بهتر عملکرد مالی، تحلیل حساسیت مدل‌های بدست آمده برای ویژگی‌های داده‌ها و اهمیت بررسی روابط متغیر در این تحلیل هستند [۷].

**داده‌کاوی در حسابداری مدیریت:** کاربردهای حسابداری مدیریت بر حوزه‌های عمده‌ای از قبیل مدیریت صنفی، مدیریت دارایی، تنظیم بودجه و مدیریت قیمت‌گذاری متمرکز هستند [۷].

- **مدیریت صنعتی:** داده‌کاوی در حوزه مدیریت صنعتی در سطوح مختلف هزینه‌یابی شامل: تجهیزات، فرایند، ساخت، محصول و پروژه در سطح تجهیزات، داده‌کاوی برای برآورد هزینه تولید تجهیزات، ارتقا دقت بازرسی و تعمیر تجهیزات و هم چنین برای پیگیری هزینه تعویض تجهیزات کاربرد دارد. کاربرد داده‌کاوی برای تحقیق در این رشته اثبات‌کننده بسیاری از نقاط قوت و ضعف آن می‌باشد. نقاط قوت شامل سنجش‌های دقت خوب تعریف شده، بکارگیری رویکردهای تحلیل تسلسلی و شفاف‌سازی اهمیت ادراک داده‌ها و نقاط ضعف شامل محدوده و اندازه بسیار محدود مطالعه و استفاده از متغیرهای محدود با داده‌های شبیه‌سازی شده می‌باشد. در سطح فرایند تجاری، داده‌کاوی برای تعریف محرک‌های هزینه در هزینه‌یابی مبتنی بر فعالیت و ارتقا مسیریابی فرایند تولید و هم چنین برای تصمیم‌گیری در مورد قیمت انتقال هوشمند مورد استفاده قرار می‌گیرد. در سطح ساخت، مولفان بر کاربرد داده‌کاوی برای مدیریت هزینه ساخت کاربرد داشته و سیستم شبکه‌های عصبی برای برآورد سریع و دقیق هزینه کل یا واحد ساخت، عملیات و نگهداری ایجاد می‌شود، بعضی از این مطالعات قابل تکرار نیستند، چرا که متغیرهای مورد استفاده خاص نبوده و تعدادی در سیستم‌های عملیاتی کنونی اجرا نمی‌شوند. در سطح پروژه، داده-کاوی برای توسعه سیستم کنترل هزینه و توسعه سیستم برآورد هزینه بکار گرفته می‌شود. برنامه‌های کاربردی برآورد هزینه به محصولات یا پروژه‌های محسوس محدود نشده بلکه برای برآورد هزینه برای پروژه‌های نامحسوس از قبیل پروژه‌های نرم افزار گسترش می‌یابند [۱۴].

- **مدیریت دارایی:** مدیریت موجودی، از جمله طبقه‌بندی موجودی، هزینه‌یابی، بهینه‌سازی و کنترل عوامل کلیدی هستند که بر رقابت شرکت تاثیر می‌گذارند. در حوزه کنترل موجودی، شبکه‌های عصبی برای بهینه‌سازی سطح موجودی بکار رفته و کاهش ۵۰٪ در هزینه موجودی (از بیش از یک میلیارد دلار به حدود نیم میلیون دلار) گزارش می‌شود. در حالی که احتمال در این سطح باقی می‌ماند که تقاضای مشتری خاصی بر آورده خواهد شد. بعلاوه این مقالات استفاده از تکنیک‌های آماری سنتی برای کمک به تعیین بهترین نوع شبکه عصبی برای یک کاربرد خاص را توصیف می‌کنند. برای مدیریت بهتر موجودی، تکنیک‌های داده‌کاوی زیادی شامل درختی‌های تصمیم، شبکه-های عصبی فازی و الگوریتم‌های ژنتیک مورد استفاده قرار گرفتند. در این تحقیقات ارتقا فرایندهای مدیریت و در نتیجه کاهش هزینه موجودی با استفاده از مدل‌های داده‌کاوی دو گانه از جمله ترکیب سیستم‌های فازی و شبکه‌های عصبی، ادغام شبکه‌های عصبی، فرایند سلسله مراتب تحلیل گرایانه و الگوریتم‌های ژنتیک گزارش دادند [۷، ۱۵].

- سایر موارد: این مطالعه، علاوه بر مقوله‌های کلی مدیریت دارایی و هزینه، کاربرد داده‌کاوی برای سایر زمینه‌های حسابداری مدیریت نشان می‌دهد. برای تنظیم بودجه، یک سیستم استدلال اینترنتی مبتنی بر مورد برای تنظیم بودجه اولیه توسعه یافت تا در غربالگری پروژه به تصمیم‌گیرندگان کمک شود، در اینجا فرایند سلسله مراتب تحلیل فازی برای تصمیم‌گیری چند معیاری به کار گرفته شد تا تصمیمات تخصیص بودجه ارتقا یابند. کاربردهای داده‌کاوی در حسابداری مدیریت عمدتاً مدیریت هزینه را در سطوح مختلف (در سطوح محصول با تجهیزات، فرایند و پروژه) به مدیریت دارایی (عمدتاً موجودی) و همچنین در سطوح دیگری مطرح می‌کنند که تاکید کمتری بر آنها می‌شود. این کاربردها به اشکال مختلفی از قبیل طبقه‌بندی، انتخاب، پیش‌بینی و بهینه‌سازی مدیریت موجودی، تعریف

محرك‌های هزینه، برآورد و پیش‌بینی هزینه محصول و پروژه، توسعه مدل‌های برآورد هزینه برای محصول، تجهیزات و پروژه، توسعه دادن سیستم‌های تنظیم بودجه، پیش‌بینی جزئیات نقدی و ... اجرا می‌شوند [۷، ۱۶، ۱۷].

**داده کاوی در بیمه و قبول اجابت:** کاربردهای داده کاوی در بیمه و قبول اجابت عمدتاً بر سه موضوع اصلی: حسابرسی (از جمله تعهد، برنامه‌ریزی، هدایت و مراحل پس از حسابرسی)، سلامت تجاری (از جمله قابلیت رشد مالی، ورشکستگی و دغدغه اخیر) و حسابداری حقوقی (از جمله آشکارسازی جعل و مدیریت درآمد) متمرکز هستند [۷].

- **حسابرسی:** اولین مورد در این زمینه حسابرسی است. تعاملات حسابداری پیچیده‌تر شده و با افزایش استفاده از سیستم‌های آنلاین و تکثیر وسایل هوشمند و اینترنت دستکاری آنها آسان‌تر می‌شود. این کار مستلزم حسابرسی تخصصی و از جمله افزایش استفاده از تکنیک‌های پیشرفته داده کاوی می‌باشد. نمی‌توان نقش مهمی که فناوری اطلاعات در ارتقا کارایی فرایند پایش و کنترل ایفا می‌کند را نادیده گرفت. داده کاوی در سراسر چرخه حسابرسی کاربرد دارد: برنامه‌ریزی (از قبیل تعهد، ارزشیابی، خطر، طراحی طرح حسابرسی)، هدایت (عمدتاً انجام آزمون‌های حسابرسی اثباتی) و گزارش‌دهی (گزارش حسابرسی) داده کاوی پس از تکمیل چرخه حسابرسی نیز کاربرد داشته و بر عقیده حساب‌رسان و نتایج کار آن‌ها تاثیر می‌گذارد [۷].

- **سلامت تجاری:** در سلامت تجاری، محققان بر سه حوزه اصلی تداوم‌پذیری مالی، ورشکستگی و دغدغه کنونی در کاربرد داده کاوی متمرکز هستند. تداوم‌پذیری مالی یا شکست تجاری را می‌توان به عنوان موقعیتی تعریف کرد که شرکت قادر به پرداخت حق وام دهندگان، سهامداران ممتاز، تامین کنندگان یا سایر بستانکاران نیست و یا شرکت بر طبق قانون در آستانه ورشکستگی است. داده کاوی در زمینه‌های متعددی مانند پشتیبانی از قضاوت حساب‌رس در مورد تداوم‌پذیری مالی مشتری، پیش‌بینی ناکامی تجاری، طبقه‌بندی، پیش‌بینی و جلوگیری از ورشکستگی‌های بانکی کاربرد دارد. ویژگی معمول این تحقیقات این است که تقریباً همه‌ی آنها از رویکرد مدل‌سازی داده کاوی دو گانه استفاده می‌کنند [۷].

- **حسابداری حقوقی:** AICPA با صراحت تمام مسئولیت حساب‌رسان در آشکارسازی جعل را تصدیق می‌کند. آشکارسازی صورتی بهای مالی دستکاری شده با استفاده از روش‌های حسابرسی عادی به وظیفه‌ای دشوار تبدیل شده است. ارزشیابی خطر جعل، فرایندی بسیار پیچیده و بخشی از مسئولیت کلی حسابرسی می‌باشد. با گذشت زمان، شرایط قانونگذاری مدت زمان و تلاش حساب‌رسان برای ارزشیابی جعل را افزایش داده است؛ بنابراین می‌توان گفت که ارزشیابی خطر جعل فرصت مطلوبی برای مساعدت در زمینه فناوری می‌باشد. بررسی ادبیات موضوع نشان دهنده‌ی کاربرد رایج داده کاوی توسط محققان و متخصصان در جهت آشکارسازی جعل می‌باشد. محققان سطوح و زمینه‌های مختلف جعل را پیگیری می‌کنند [۷، ۱۸].

- **قبول اجابت مالیات:** داده کاوی در تعیین مالیات، به فرض در مطالعه دنتون و همکارانش کاربرد دارد. این مولفان از شبکه‌های عصبی برای طبقه‌بندی کارکنان به منظور اهداف مالیاتی استفاده کردند. برای قبول اجابت مالیات در سطح شرکتی، وو از شبکه‌های عصبی و کالو و باک از نقشه خود سازمان یافته استفاده کردند تا بتوان شرکت‌های نیازمند پژوهش بیشتر در حسابرسی مالیات را به نحو موفقیت آمیزی شناسایی کرد. وو و همکارانش تکنیک‌های داده کاوی را برای ارتقا عملکرد آشکارسازی فرار مالیاتی بکار گرفتند تا چارچوب غربالگری برای فیلتر کردن گزارش‌های مالیاتی احتمالاً غیر پرداختی و متحمل حسابرسی بیشتر توسعه یابد. نتایج وو و همکارانش

نشان می‌دهند که تکنیک‌های داده‌کاوی پیشنهادی آنها آشکارسازی فرار مالیاتی را ارتقا می‌دهد، بنابراین، می‌توان آن را برای کاهش یا به حداقل رساندن کارآمد زیان حاصل از فرار مالیاتی به کار برد [۷].

### ۵- معرفی تکنیک‌های طبقه‌بندی پر کاربرد داده‌کاوی در حسابداری و حسابرسی

به منظور حسابرسی شرکتها، حسابرسان پی‌درپی باید با داده‌های حجیم و ساختار پیچیده این داده‌ها سرو کار داشته باشند. در نتیجه، حسابرسان بیش از این نمی‌توانند تنها به ابزارهای گزارشگری یا خلاصه کردن در فرآیند حسابرسی متکی باشند. ابزارهای دیگری نظیر داده‌کاوی که به طور خودکار اطلاعات را از میان حجم داده‌های زیاد، استخراج می‌کنند، می‌توانند مفید باشند، نشان داده شده است که در بسیاری از کاربردهای تجاری مرتبط با حسابرسی نظیر کشف تقلب، حسابداری دادگاهی دارای مزیت‌های فراوانی خواهد بود از این رو در این بخش تکنیک‌های پر کاربرد داده‌کاوی که برای استخراج داده‌های مالی و طبقه‌بندی نتایج حسابداری و حسابرسی بسیار مفید است، معرفی شده است [۱۹].



### شکل ۴: نمایی از تکنیک‌های طبقه‌بندی پر کاربرد داده‌کاوی در حسابداری و حسابرسی

**درخت تصمیم‌گیری:** درخت تصمیم‌گیری در داده‌کاوی، معمولاً جهت بررسی داده‌ها و استنتاج درخت و قوانین آن که برای انجام پیش‌بینی از آن‌ها استفاده می‌شود، بکار گرفته می‌شوند. هنگامی که نمونه‌ها باید در دسته‌ها یا کلاس‌های مختلف قرار گیرند، پیش‌بینی می‌تواند شامل پیش‌بینی مقادیر دسته‌ها (درخت‌های کلاس‌بندی) باشد. درخت تصمیم‌گیری، یک طبقه‌بند به فرم ساختار درختی است که در آن، هر گره یا به صورت برگ و یا به صورت گره تصمیم‌گیری است. گره برگ نشان‌دهنده مقدار مشخصه هدف یا کلاس نمونه‌ها است، درحالی که گره تصمیم‌گیری به روی تک‌تک مقادیر مشخصه آزمایش انجام داده و می‌تواند در ازای نتیجه هر آزمون، شامل یک شاخه یا زیر درخت شود. از درخت تصمیم می‌توان در کلاس‌بندی یک نمونه استفاده کرد؛ بدین صورت که ابتدا از ریشه درخت آغاز شده و به درون آن حرکت کرده تا به یک گره برگ برسد، به نحوی که این گره، کلاس‌بندی نمونه را مهیا سازد [۸]. درخت تصمیم را می‌توان مشابه با فلوچارتی دانست که در آن، هر گره‌ی غیر برگ بیانگر یک تست بر روی یک صفت خاص بوده و هر شاخه بیانگر نتیجه‌ی آن تست بوده و هر گره‌ی برگ دارای یک برچسب کلاسی می‌باشد. گره‌ای که دارای بالاترین برچسب‌ها در درخت می‌باشد گره‌ی ریشه نام دارد. با استفاده از درخت تصمیم، تصمیم‌گیران می‌توانند بهترین راهکار را برای حرکت از ریشه به برگ که بیانگر تفکیک کلاس انحصاری بر مبنای بهره‌ی اطلاعاتی حداکثری می‌باشد انتخاب نمایند [۱۰].

**شبکه‌های عصبی:** شبکه‌های عصبی می‌توانند توابع بسیار پیچیده، به‌ویژه غیرخطی را مدل‌سازی کنند. ساختار شبکه‌های عصبی از واحدهای متعدد به هم متصل یا همان نورون‌های مصنوعی ساخته شده است. هر کدام از این واحدها شامل مشخصات ورودی/خروجی است که محاسبه یا تابعی محلی را پیاده‌سازی و اجرا می‌کنند. تابع می‌تواند محاسبه حاصل جمع‌های وزن‌های ورودی‌ها باشد که در صورت رسیدن به یک حد آستانه، خروجی را تولید می‌کند. خروجی

(یا همان نتیجه) می تواند به عنوان ورودی برای سایر نوروهای موجود در شبکه در نظر گرفته شود. این فرایند تا زمانی که یک خروجی نهایی به وجود بیاید، ادامه پیدا می کند [۹،۸]. در واقع شبکه‌ی عصبی را می توان الگوریتمی برای دسته بندی دانست که از روش گرادیان نزولی بر مبنای سیستم عصب ییولوژیکی که دارای چندین المان پردازشی به هم مربوطه تحت عنوان نرون هست استفاده کرده و برای حل مسائلی خاص کاربرد دارد. قواعد از شبکه‌ی عصبی آموزش دیده شده استخراج شده تا بتواند عملکرد متقابل شبکه‌ی یادگیری شده را بهبود دهد. به منظور حل یک مسئله‌ی خاص، شبکه عصبی از عصب‌هایی استفاده کرده که شامل المان‌های پردازشی سازمان‌دهی شده می‌باشند. از شبکه‌ی عصبی برای دسته‌بندی و تشخیص الگو استفاده می‌شود. یک شبکه عصبی دارای ماهیتی انطباقی می‌باشد چرا که می‌تواند ساختار خود را تغییر داده و وزن خود را به منظور کاهش خطا تطبیق دهد. تطبیق وزن بر مبنای اطلاعاتی که به صورت داخلی و خارجی در شبکه و آن‌هم در طول فاز یادگیری جریان دارد صورت می‌گیرد. در شبکه عصبی چند کلاسه، ممکن است مسئله با استفاده از تکنیک feedforward<sup>۶</sup> چند لایه‌ای حل شود که در آن، عصب‌ها در لایه‌ی خروجی بکار گرفته می‌شوند و نه اینکه تنها بر روی یک لایه بکار گرفته شوند [۱۰]. با توجه به اینکه این تکنیک در روش پیشنهادی مورد استفاده قرار می‌گیرد در بخش بعدی بسیار کامل به آن پرداخته شده است.

**بیز ساده<sup>۷</sup>:** بیز ساده یک راه ساده برای ایجاد مدل‌های پیشگوی آماری است که بر مبنای تئوری بیز است. در این روش کلاس‌های مختلف، هر کدام به شکل یک فرضیه دارای احتمال در نظر گرفته می‌شوند. هر رکورد آموزشی جدید، احتمال درست بودن فرضیه‌های پیشین را افزایش و یا کاهش می‌دهد و در نهایت، فرضیاتی که دارای بالاترین احتمال شوند، به عنوان یک کلاس در نظر گرفته شده و بر چسبی بر آن‌ها زده می‌شود. این تکنیک با ترکیب تئوری بیز و رابطه سببی بین داده‌ها، به طبقه بندی می‌پردازد. در این روش احتمال اولیه وجود دارد که آن را با  $P(C = c)$  نشان می‌دهند. علاوه بر احتمال اولیه، الگوریتم بیز همچنین احتمال مربوط به اینکه نمونه مشخص  $X$ ، با فرض اینکه مشخصه‌ها مستقل هستند مقدار  $C$  را باز می‌گرداند را نیز محاسبه می‌کند. این احتمال تبدیل به حاصل ضرب احتمالات هر مشخصه جداگانه می‌شود. بدین ترتیب، احتمالات می‌توانند از تکرار نمونه‌ها در مجموعه آموزشی، تخمین زده شوند [۹،۸]. در واقع دسته‌بندی مبتنی بر تئوری بیزی را به عنوان دسته‌بندی بیزی می‌شناسند. این دسته بند، یک دسته بند ساده‌ای بوده که می‌تواند با استفاده از الگوریتم دسته‌بندی به دست آید. قضیه‌ی بیزی می‌تواند مبنایی را برای دسته‌بندی نایو بیز و شبکه‌ی باور بیزی فراهم سازد [۱۰].

**رگرسیون لجستیک<sup>۷</sup>:** از رگرسیون برای پیدا کردن توابعی که به تشریح همبستگی‌های بین متغیرهای مختلف می‌پردازند استفاده می‌شود. یک مدل ریاضی با استفاده از مجموعه داده‌ای آموزشی ایجاد می‌شود. در مدل‌سازی آماری، دو نوع متغیر بکار گرفته شده که یکی متغیر وابسته بوده و دیگری متغیر مستقل نام داشته و به صورت  $Y$  و  $X$  نشان داده می‌شود. همچنین یک متغیر وابسته وجود دارد در حالی که ممکن است دو یا چند متغیر مستقل وجود داشته باشد. رگرسیون را می‌توان یک روش آماری دانست که به بررسی رابطه‌ی بین متغیرها می‌پردازد. با استفاده از رگرسیون، وابستگی‌های یک متغیر به سایر متغیرها ایجاد می‌شود. بر مبنای تعداد متغیرهای مستقل، رگرسیون دارای دو

<sup>۶</sup> یک شبکه عصبی مصنوعی است که در آن اتصال میان واحد های تشکیل دهنده آن یک چرخه را تشکیل نمی‌دهند. در واقع این شبکه متفاوت از شبکه های عصبی بازگشتی می‌باشد.

<sup>۶</sup> Naive Bayes

<sup>۷</sup> Logistic Regression (LR)

نوع رگرسیون خطی و رگرسیون غیر خطی می‌باشد. رگرسیون خطی می‌تواند به تشخیص رابطه‌ی یک متغیر وابسته و یک یا چند متغیر مستقل بپردازد. این روش بر مبنای یک مدلی بوده که از تابع خطی برای ساخت خود استفاده می‌کند. رگرسیون خطی می‌تواند یک خط را پیدا کرده و فاصله‌ی عمومی نقاط از این خط را محاسبه نماید و جمع ریشه‌ی فاصله‌ی عمومی را به حداقل سطح ممکن برساند. در این روش، متغیرهای وابسته و مستقل شناخته شده می‌باشند و هدف این بوده که یک خطی پیدا شود که بین این متغیرها یک ارتباط داشته باشد. ولی رگرسیون خطی محدود به داده‌های عددی بوده و نمی‌تواند برای داده‌های قابل دسته‌بندی بکار گرفته شود. رگرسیون لجستیکی نیز یک نوع رگرسیون غیرخطی بوده که داده‌های قابل دسته‌بندی را پذیرفته و می‌تواند احتمال رخداد را با استفاده از یک تابع لجیت پیش‌بینی نماید. رگرسیون لجستیکی دارای دو نوع دو جمله‌ای و چند جمله‌ای می‌باشد [۱۰].

**ماشین بردار پشتیبان:** ماشین بردار پشتیبان عبارت از مجموعه‌ای از روش‌های یادگیری با نظارت مرتبط هستند که داده‌ها را مورد تحلیل قرار داده و الگوها را شناسایی می‌کنند. از ماشین‌های بردار پشتیبان برای کلاس‌بندی و تحلیل‌های رگرسیون استفاده می‌شود. ماشین بردار پشتیبان یک الگوریتم است که تلاش می‌کند تا یک جداکننده خطی (آبر صفحه) را بین نقاط داده‌ای مربوط به دو کلاس در فضای چندبعدی پیدا کند. ماشین بردار پشتیبان یک روش یادگیری را ارائه می‌دهد که در آن از اصول و قواعد تئوری یادگیری آماری پیروی می‌شود. به‌طور کلی، ایده اصلی بردار پشتیبان برگرفته شده از کلاس‌بندی دودویی است، یعنی یافتن یک ابر صفحه به‌نحوی که به‌منظور به حداقل رساندن خطای کلاس‌بندی، دو کلاس را از هم جدا کند. ماشین‌های بردار پشتیبان بر پایه اصل حداقل سازی خطای ساختاری از نظریه-های یادگیری آماری (محاسباتی) توسعه یافته‌اند. این دسته‌بندی کننده خود کارایی بسیار بهتری را نسبت به روش‌های دیگر در دسته‌بندی متن از خود نشان داده است. یکی از فاکتورهای مهم در یادگیری متن این است که، یادگیرنده (دسته‌بندی کننده) قابلیت تعمیم مناسب با استفاده از تعداد معدودی داده آموزشی را داشته باشد. در اغلب کاربردهای واقعی و مخصوصاً در فضای وب تعداد داده‌های آموزشی کافی و یا زیادی برای یادگیرنده وجود ندارند و یا فراهم نمودن آن زمان‌بر و مشکل است [۱۰].

**K نزدیک ترین همسایه:** کلاس‌بندی K-NN نمونه‌ها را بر اساس شباهت آن‌ها دسته‌بندی می‌کند. یک شیء با استفاده از اکثریت همسایه‌های خود کلاس‌بندی می‌شوند. مقدار K همیشه یک عدد صحیح مثبت است. همسایه‌ها از مجموعه‌ای از اشیاء انتخاب می‌شوند، به‌نحوی که کلاس آن‌ها مشخص شده باشد. نمونه‌های آموزشی به‌وسیله مشخصه‌های عددی  $n$  بعدی توصیف می‌شوند. هر نمونه نشان‌دهنده یک نقطه در فضای  $n$  بعدی است. به‌این ترتیب، تمام نمونه‌های آموزشی در فضای الگوهای  $n$  بعدی ذخیره شده‌اند. در زمان مواجه با یک نمونه ناشناخته، کلاسیفایر K-NN در فضای الگوها به دنبال نمونه‌های آموزشی می‌گردد که بیشترین تشابه و نزدیکی را به نمونه ناشناخته داشته باشند. "نزدیکی" با استفاده از فاصله اقلیدسی تعریف می‌شود. نمونه ناشناخته به کلاسی نسبت داده می‌شود که متعلق به اکثر  $k$  همسایه نزدیک‌تر به آن است. اگر  $k = 1$  باشد، آنگاه نمونه ناشناخته به کلاس متعلق به نمونه آموزشی که دارای نزدیک‌ترین فاصله به آن در فضای الگو باشد، نسبت داده می‌شود. در وکا به این کلاسیفایر IBK<sup>۸</sup> (روش نمونه بر اساس بر مبنای K) گفته می‌شود [۹، ۸].

<sup>8</sup> Instance-Based method based on k

## ۶- نتیجه گیری

در دنیای تجارت نیز داده‌های مالی به عنوان سرمایه راهبرد مطرح هستند. داده‌های مالی توسط موسساتی مثل بانک‌ها، بورس اوراق بهادار، سازمانهای مالیاتی، پایگاههای داده ویژه حسابرسان و حسابداران و غیره جمع‌آوری و نگهداری می‌شوند. روش‌های داده کاوی در داده‌های مالی، می‌تواند در حل مشکلات طبقه‌بندی و پیش‌بینی و تسهیل فرآیند تصمیم‌گیری به کار رود. نمونه‌هایی از مسائل طبقه‌بندی مالی شامل ورشکستگی شرکت‌ها تخمین ریسک اعتباری، گزارش تداوم، فعالیت، درماندگی مالی و پیش‌بینی عملکرد واحد تجاری می‌باشد. بدین ترتیب دامنه اهمیت داده کاوی در امور مالی و حسابداری می‌تواند طیفی گسترده از کشف تقلب تا افزایش سودآوری واحد تجاری باشد. از این رو طبقه‌بندی مسائل مهم مالی از جمله ورشکستگی شرکت‌ها، تخمین ریسک اعتباری، گزارش تداوم فعالیت، درماندگی مالی و پیش‌بینی عملکرد واحد تجاری بسیار حائز اهمیت است، زیرا داده کاوی فرآیند استخراج دانش پنهان شده در حجم زیادی از داده‌ها و یکی از مهم‌ترین نمونه‌های فعلی تجزیه و تحلیل پیشرفته تجارت و ابزار پشتیبانی تصمیم‌گیری است. در سال‌های اخیر، استفاده از تکنیک‌های داده کاوی در حسابرسی صورت‌های مالی افزایش یافته است. در این مقاله به بررسی کاربرد داده کاوی در حرفه حسابداری و حسابرسی پرداخته شد و در بخش آخر تکنیک‌های پر کاربرد طبقه‌بندی در حسابداری و حسابرسی، معرفی شد تا برای کاربران این حوزه جهت ایده نوین در موضوع مرتبط و تحقیقات آتی مثمر ثمر واقع شود.

## منابع

- [1] Smith, S. Stein. "Digitization and financial reporting—how technology innovation may drive the shift toward continuous accounting." *Accounting and Finance Research* 7, no. 3 (2018): 240-250.
- [2] PhridviRaj, M. S. B., and C. V. GuruRao. "Data mining—past, present and future—a typical survey on data streams." *Procedia Technology* 12 (2014): 255-263.
- [3] Madni, Hussain Ahmad, Zahid Anwar, and Munam Ali Shah. "Data mining techniques and applications—A decade review." In *2017 23rd International Conference on Automation and Computing (ICAC)*, pp. 1-7. IEEE, 2017.
- [4] Kumar, Nitin, Sumika Jain, and Kuldeep Chauhan. "Knowledge Discovery from Data Mining Techniques." *International Journal of Engineering and Technology*, ISSN: 2278-0181.2019.
- [5] Gullo, Francesco. "From patterns in data to knowledge discovery: What data mining can do." *Physics Procedia* 62 (2015): 18-22.
- [6] Gupta, Manoj Kumar, and Pravin Chandra. "A comprehensive survey of data mining." *International Journal of Information Technology* (2020): 1-15.
- [7] Amani, Farzaneh A., and Adam M. Fadlalla. "Data mining applications in accounting: A review of the literature and organizing framework." *International Journal of Accounting Information Systems* 24 (2017): 32-58.
- [8] Abd AL-Nabi, Delveen Luqman, and Shereen Shukri Ahmed. "Survey on classification algorithms for data mining: (comparison and evaluation)." *International Journal of Computer Engineering and Intelligent Systems* 4, no. 8 (2013): 18-27.
- [9] Fdez-Glez, Jorge, David Ruano-Ordas, José Ramón Méndez, Florentino Fdez-Riverola, Rosalía Laza, and Reyes Pavón. "A dynamic model for integrating simple web spam classification techniques." *Expert Systems with Applications* 42, no. 21 (2015): 7969-7978.

- [10] Tomar, Divya, and Sonali Agarwal. "A survey on Data Mining approaches for Healthcare." *International Journal of Bio-Science and Bio-Technology* 5, no. 5 (2013): 241-266.
- [11] Baldwin, Amelia A., Carol E. Brown, and Brad S. Trinkle. "Opportunities for artificial intelligence development in the accounting domain: the case for auditing." *Intelligent Systems in Accounting, Finance & Management: International Journal* 14, no. 3 (2006): 77-86.
- [12] Owen, Gareth. "Integrated reporting: A review of developments and their implications for the accounting curriculum." *Accounting Education* 22, no. 4 (2013): 340-356.
- [13] James, R. E. "Business analytics: Methods, models and decisions." (2013).
- [14] Khalifelu, Zeynab Abbasi, and Farhad Soleimanian Gharehchopogh. "Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation." *Procedia Technology* 1 (2012): 65-71.
- [15] Braglia, Marcello, Andrea Grassi, and Roberto Montanari. "Multi-attribute classification method for spare parts inventory management." *Journal of quality in maintenance engineering* (2004).
- [16] Ragothaman, Srinivasan, and Angeline Lavin. "Restatements due to improper revenue recognition: a neural networks perspective." *Journal of Emerging Technologies in Accounting* 5, no. 1 (2008): 129-142.
- [17] Chou, Jui-Sheng. "Web-based CBR system applied to early cost budgeting for pavement maintenance project." *Expert Systems with Applications* 36, no. 2 (2009): 2947-2960.
- [18] Dikmen, Burcu, and Güray Küçükkoçaoğlu. "The detection of earnings manipulation: the three-phase cutting plane algorithm using mathematical programming." *Journal of Forecasting* 29, no. 5 (2010): 442-466.

[۱۹] دستگیر محسن، شفیعی سردشت مرتضی. فناوری داده کاوی، رویکردی نوین در حوزه مالی. دانش حسابرسی. ۱۳۹۰؛ ۱۱ (۴).

[۲۰] ستایش محمدحسین، فتاحی نافچی حسن، عباسپور سمیره، روستایی میثم. ارائه رویکردی نوین در صدور گزارش حسابرسی با استفاده از داده کاوی (مطالعه موردی: شرکت های پذیرفته شده در بورس اوراق بهادار تهران). دانش حسابرسی.

[۲۱] حمیدیان محسن، لبافی معصومه. بررسی نقش سازمانی دادهکاوی در حوزههای حسابداری و مالی. ماهنامه پژوهشهای مدیریت و حسابداری، ماهنامه پژوهشهای مدیریت و حسابداری.

# Introducing an Appropriate Organizational Framework for Data Mining Applications in Accounting and Auditing: An Overview of Common Financial Data Classification Techniques

Yaser Rezaei Pita Noei<sup>1</sup>  
Mustafa Asghari Shalmani<sup>2</sup>  
Hossein Daliri Dehbaneh<sup>3</sup>

Date of Receipt: 2022/01/16 Date of Issue: 2022/02/17

## Abstract

The rapidly changing age of the global economy requires organizations and companies to consider a wide range of innovative management approaches and advanced information technologies in order to compete effectively. Therefore, data mining is of particular importance in the future of the accounting and auditing profession because data mining is one of the most important recent topics for smart business analysis and decision support tools. Data mining provides organizations with a wide range of benefits and capabilities, including efficient forecasting of future corporate development trends and increased business competitiveness, for better decision making. On the other hand, with the existence of the World Wide Web, the volume of information in integrated information systems is increasing every moment, so the need for a proper science to discover and extract information quickly and accurately from these databases is essential, which is now data mining. The most important technology for efficient, fast and accurate utilization of large data. Today, data mining and its tools are of particular importance in finance and accounting. Therefore, this article examines the application of data mining in the accounting and auditing profession. Also, because the classification and capabilities of data mining techniques provide a way to predict bankruptcy, financial turmoil, managerial fraud detection, risk estimation, and performance forecasting, the main part of the article is Techniques. The application of classification in accounting and auditing was introduced and discussed.

## Keyword

Accounting, auditing, financial data, data mining and classification

1. Assistant Professor of Accounting, Faculty of Literature and Humanities, University of Guilan, Rasht, Iran. (Rezaei.yasser@gmail.com)

2. Deputy Finance Minister of Valiasr Hospital in Rasht. (Mostafaasgharimali@gmail.com)

3. Responsible for setting up Rasht Municipality financial accounts. (hdhossein1961@gmail.com)