تاریخ دریافت: ۱۳۹۰/۴/۱۰
تاریخ پذیرش: ۱۳۹۰/۵/۱۲

فناوری داده‌کاوی؛ رویکردی نوین در حوزه مالی

دکتر محسن دستگیر / استاد دانشگاه شهید چمران اهواز
مرتضی شفیعی سردشت / دانشجوی کارشناسی ارشد دانشگاه آزاد مشهد

چکیده

در طول سالیان اخیر، حجم مخازن داده‌های الکترونیکی ایجاد شده توسط بانک‌ها، شرکت‌ها و سایر موسسات تجاری، همواره در حال افزایش بوده است. بیت‌های با ارزش اطلاعات در این مخازن داده، جـای گرفته‌اند، به گونه‌ای که اندازه و حجم بسـیار زیاد ایـن منابع، تجزیه و تحلیل آن‌ها را برای انسان به منظور خلق اطلاعات یا الگوهای مفید و مناسب در جهت فرآیند تصمیم‌گیری، غیر ممکن ساخته است. فنونی نظیر داده‌کاوی می‌توانند به طورخودکار، اطلاعات را از میان این مقادیر حجیم داده اسـتخراج نمایند. داده‌کاوی، تاثیراتی عمیق بر شیوه‌های کسب و کار و مدیریت دانش در سـال‌های اخیر داشته اسـت. هوشمندی کسب و کار، مشهورترین کاربرد فنون داده‌کاوی است. قابلیت‌های طبقه‌بندی و پیش‌بینی ابزارهای داده‌کاوی، موجب کاربرد این ابزارها به منظور مقاصد پیش‌بینی ورشکستگی، وضعیت تداوم فعالیت، پیش‌بینی درماندگی مالی، تشخیص و شناسایی تقلب مدیریت، هم‌چنین تخمین ریسک اعتباری و عملکرد واحد تجاری شده و بدین‌ترتیب داده‌کاوی را تبدیل به موضوعی با درجه اهمیت زیاد، در امور مالی و حسـابداری نموده اسـت. هدف این مقاله بررسـی اجمالی این فناوری جدید و تحقیقات مرتبط با آن، به منظور نمایان سـاختن کارکردهای داده‌کاوی در امور مالی و حسابداری وفرصت‌های تحقیقاتی آن می‌باشد.

واژه‌های کلیدی: حسابداری، امور مالی، روش‌های داده‌کاوی، ابزارهای داده‌کاوی، داده‌کاوی

مقدمه

داده‌کاوی، شاخه‌ای درخشان در علوم کامپیوتری است که در اواخر دهه ۸۰ با بکارگیری مفاهیم و روش‌های مرتبط با هوش مصنوعی[۱]، شناسایی الگو[۲]،سیستم‌های پایگاه داده[۳] وعلم آمار[۴]، پا به عرصه گذاشت.

در طول سال‌های اخیر، به تدریج با این واقعیت رشد یافته‌ایم که حجم عظیمی از داده‌ها وجود دارند که کامپیوترها، شبکه‌ها و در حقیقت تمام زندگی ما را فرا گرفته‌اند. سازمان‌های دولتی، مؤسسات علمی و تجاری سرمایه هنگفتی 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 تقریبا 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 ۱۸الی ۲۴ماه دو برابر شده، در حالی که هزینه مرتبط با این توان محاسباتی در حال کاهش می‌باشد(Seifert, ۲۰۰۴).

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.

این مشکل چطور به وجود آمد؟ در پاسخ باید گفت اغلب تلاش‌ها برای ایجاد مجموعه‌ای از داده، بر روی موضوعاتی از قبیل کارآمدی و ذخیره‌سازی متمرکز می‌شوند و در حقیقت، هیچگونه طرح یا برنامه‌ای برای این وجود ندارد که در نهایت، داده‌ها چگونه استفاده وتجزیه و تحلیل شوند (علیخانزاده،۱۳۸۵).

با ورود به عصر اطلاع رسانی دیجیتال، مسئله انفجار داده‌ها، اطلاعات وضرورت نگهداری آنها هر روز بیش‌ازپیش 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 و پرداخت نشده خود 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 سرریزی وحجم زیاد اطلاعات امروز آگاه است. نتایج بررسی‌های به عمل آمده نشان می‌دهد که:

الف- ۶۱ درصد از مدیران معتقد به سرریزی اطلاعات در شرکت‌ها می‌باشند.

ب-۸۰ درصد اعتقاد به بدتر شدن وضعیت فوق دارند.

ج- بیش از ۵۰ درصد مدیران از داده‌ها در مراحل تصمیم‌گیری فعلی به واسطه سرریزی اطلاعات صرف‌نظر می‌کنند.

د- ۸۴ درصد از مدیران، این اطلاعات را برای آینده ذخیره می‌کنند و از آنها برای تحلیل‌های جاری استفاده نمی‌کنند.

هـ- ۶۰ درصد عقیده دارند که هزینه جمع‌آوری و ذخیره اطلاعات بیشتر از ارزش خود اطلاعات می‌باشد(همان).

در این شرایط، داده‌کاوی[۵] ابزاری سودمند است که در اختیار تمام شرکت‌ها می‌باشد. داده‌کاوی و کشف دانش، ابزارهایی هوشمند هستند که به جمع‌آوری و پردازش داده‌ها و بهره‌مندی از آنها، منجر می‌شوند. در واقع، داده‌کاوی شامل استفاده از ابزارهای تحلیلی پیچیده به منظور کشف الگوهای معتبر و ناشناخته و وابستگی‌های موجود در مجموعه داده‌های حجیم می‌باشد. نتیجه این که، فناوری داده‌کاوی چیزی بیش از جمع‌آوری و مدیریت داده بوده و شامل تجزیه و تحلیل و پیش‌بینی نیز می‌باشد (Seifert, ۲۰۰۴).

برخی از صاحب‌نظران براساس مفاهیم داده‌کاوی، معتقدند این فناوری تنها مرحله‌ای از فرآیندی فراگیرتر به نام کشف دانش درپایگاه داده[۶] می‌باشد. مراحل دیگر در این فرآیند فراگیر شامل پاک‌سازی[۷]، یکپارچه‌سازی[۸]،انتخاب داده‌ها[۹]، تبدیل[۱۰] و ارزیابی الگو[۱۱] است. فناوری داده‌کاوی میان بسیاری از حوزه‌های فنی شامل پایگاه داده، آمار، یادگیری ماشین[۱۲] و رابط انسان و کامپیوتر[۱۳] ارتباط برقرار می‌سازد(pechenizkiy and et.al, ۲۰۰۳).

داده‌کاوی و داده‌های مالی

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 جمع آوری و نگهداری می‌شوند. روش‌های داده‌کاوی در داده‌های مالی، می‌تواند در حل مشکلات طبقه‌بندی و پیش‌بینی و تسهیل فرآیند تصمیم‌گیری به کار رود. نمونه‌هایی از مسائل طبقه‌بندی مالی[۱۴] شامل ورشکستگی شرکت‌ها[۱۵]، تخمین ریسک اعتباری[۱۶]، گزارش تداوم

فعالیت[17]، درماندگی مالی[18] و پیش‌بینی عملکرد واحد تجاری[19] می‌باشد.[20] بدین‌ترتیب دامنه اهمیت داده‌کاوی در امور مالی و حسابداری می‌تواند طیفی گسترده از کشف تقلب تا افزایش سودآوری واحد تجاری باشد (Rumsey and et.al,۲۰۰۴).

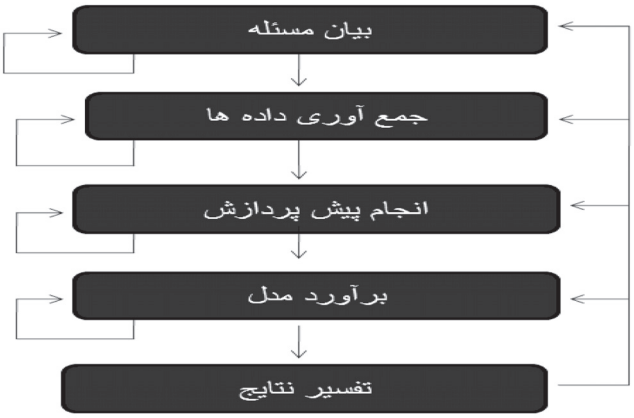
اهمیت داده‌کاوی توسط بسیاری از سازمان‌های حرفه‌ای تشخیص داده شده است. انجمن حسابداران رسمی آمریکا، داده‌کاوی را به عنوان یکی از ۱۰ فناوری برتر برای آینده معرفی کرده است. همچنین انجمن حسابرسان داخلی آمریکا نیز این فناوری را در فهرست یکی از چهار اولویت تحقیقاتی خود گنجانده است (kirkods and et.al,۲۰۰۴).

کاربرد احتمالی داده‌کاوی درحسابداری مالی منجر به استفاده از صورت‌های مالی به جهت تعیین سودآوری، تطابق نسبت‌ها با میانگین صنعت و بررسی اثربخشی یک کسب و کار می‌گردد. در حسابداری مدیریت نیز داده‌های واحدهای تجاری به منظور تعیین چگونگی عملیات روزمره واحد تجاری، تجزیه و تحلیل بهای تمام شده و سودآوری بخش‌های مختلف، مورد استفاده قرار می‌گیرد. در حسابرسی، ابزارهای داده‌کاوی برای تجزیه و تحلیل داده‌های مرتبط با احتمال تقلب مدیریت می‌توانند به کار گرفته شوند. بدین‌ترتیب داده‌کاوی در تمام جنبه‌های حسابداری سودمند خواهد بود، زیرا حسابداری عبارت از جمع‌آوری، تجزیه و تحلیل واستفاده از اطلاعات مالی و غیر مالی به منظور پردازش برای اشخاص ذی‌نفع است(Rumsey and et.al,۲۰۰۴). شرکت‌های تجاری می‌توانند از داده‌کاوی به منظور شناسایی فرصت‌های مالی و عملیاتی قابل توجه برای اطمینان از کارایی فرآیند زنجیره تامین[21] بهره‌مند شوند.

مفاهیم داده‌کاوی

الف- مراحل(فرآیند) داده‌کاوی

با نگاهی اجمالی به مراحل داده‌کاوی، می‌توان نحوه انجام آن را به صورت نمودار شماره یک بیان کرد:



نمودار شماره یک : مراحل داده‌کاوی

مراحل داده‌کاوی را می‌توان به گونه‌ای جزئی تر به صورت زیر ابراز کرد:

۱- شناسایی هدف: در این مرحله مشخص می‌شود کاربر چه چیزی را می‌خواهد و تا چه سطحی از اطلاعات را در نظر دارد که از پایگاه داده، اخذ نماید.

۲- انتخاب داده‌ها: در این مرحله باید داده‌ها بر مبنای معیارهای مشخص انتخاب گردند.

۳- آماده‌سازی داده‌ها: شکل قابل استفاده داده و شناسایی متغیرهای زائد وظیفه این مرحله از فرآیند خواهد بود.

۴- ارزیابی داده‌ها: چارچوب کلی این مرحله، معیارهایی از قبیل نوع توزیع داده‌ها، ویژگی‌ها و ساختار پایگاه داده و شرایط کلی داده‌ها و... می‌باشد.

۵- قالب‌بندی پاسخ : خروجی این بخش، ارایه فرمت به شکل تصویر، نمودار، شبکه عصبی و... است.

۶- انتخاب ابزار: در این مرحله ابزارهای مناسب برای داده‌کاوی انتخاب می‌گردد.

۷- الگوسازی: فرآیند داده‌کاوی به صورت اصلی از این مرحله آغاز می‌گردد که شامل جستجوی الگوها در مجموعه داده، طبقه‌بندی و ارزشیابی داده‌ها و... می‌باشد.

۸- اعتبارسازی یافته‌ها: این مرحله، شامل آزمون کردن الگوها است.

۹- ارایه نتایج: نتیجه این بخش، گزارش نهایی برای کاربر است.

۱۰- استفاده از نتایج: هدف اصلی داده‌کاوی استفاده از نتایج کشف شده برای تصمیم‌گیری، سیاست‌گذاری و پیش‌بینی به منظور ایجاد یک موقعیت بهتر و جدید می‌باشد (نجات و اکبری، ۱۳۸۷).

ب: ابزارهای داده‌کاوی[۲۲]

ابزارهای داده‌کاوی نرم‌افزارهایی هستند که به کاربران اجازه استخراج اطلاعات از داده‌ها را می‌دهند. این ابزارها توانایی گردآوری داده‌ها و کاربرد آن‌ها به منظور تصمیم‌گیری در خصوص مصرف‌کننده خاص یا گروهی از مصرف‌کنندگان را، به سازمان‌ها و افراد می‌دهند. استخراج دستی داده‌ها از صدها سال پیش وجود داشته است. این در حالی است که ماشینی شدن فرآیند داده‌کاوی از هنگام ورود کامپیوتر شایع شده است. هدف نهایی این ابزارها ظاهر ساختن الگوهای پنهان[۲۳] می‌باشد. به هرحال ابزارهای داده‌کاوی می‌تواند شامل نرم‌افزارهای زیر باشد:

۱- Data to knowledge[۲۴]

۲- sas[۲۵]

۳- clementine[۲۶]

۴- Intelligent-miner[۲۷]

۵- Insightful miner[۲۸]

۶-weka

۷- microsoft excel

۸- ACL[۲۹]

ج- روش‌های داده‌کاوی[۳۰]

اهداف داده‌کاوی شامل پیش‌بینی[۳۱] و توصیف[۳۲] یا ترکیبی از آنهاست. هدف "پیش‌بینی" تمرکز بر روی دقت در توانایی پیش‌بینی بوده و "توصیف" بر درک فرآیند تولید داده‌ها تمرکز دارد. در پیش‌بینی تا زمانی که مدل قدرت پیش‌بینی دارد، کاربر توجهی به این ندارد که مدل انعکاس دهنده واقعیت است؛ مثلا مدلی که شاخص‌های مالی را به شکل غیر خطی ترکیب می‌کند تا نرخ تبادل ارز را پیش‌بینی نماید. از سویی دیگر مدل توصیفی به مثابه واقعیت تفسیر می‌شود. برای مثال مدلی که متغیرهای اقتصادی و جمعیتی را به پیشرفت‌های آموزشی مرتبط می‌سازد، به عنوان مبنایی برای

توصیه‌های سیاست اجتماعی توصیف می‌شود. در عمل، اغلب کاربردهای اکتشاف دانش به درجه‌ای از هر دو مدل سازی توصیفی و پیش‌بینی نیاز دارند.

به هر ترتیب، اهداف داده‌کاوی با اســتفاده از روش‌های داده‌کاوی، محقق می‌شــوند. اصطلاح روش‌های داده‌کاوی در واقع بیانگر جمع کثیری از الگوریتم‌ها[۳۳] و فنون اســت که از علومی مانند آمار، یادگیری ماشــین، پایگاه داده وتجسم‌سازی[۳۴] استنتاج شده‌اند. برخی از این روشها در مورد داده‌های مالی به کار رفته‌اند. روشــهای داده‌کاوی مشــهوری که در این مقاله معرفی خواهند شد شــامل شــبکه‌های عصبی[۳۵]، الگوریتم‌های ژنتیک[۳۶] درختان تصمیم[۳۷]، نظریــه مجموعه اولیه[۳۸]، اســتدلال مبتنی بر مورد[۳۹] (اســتدلال موردگرا) و برنامه‌نویسی ریاضی[۴۰] هستند. بدیهی است که این فهرســت از روش‌ها، فهرســتی کامل نبوده وترتیبی برای اولویت‌بندی در کاربرد این روش‌ها، پیشنهاد نمی‌شود (علیخانزاده، ۱۳۸۵).

شبکه‌های عصبی

شــبکه‌های عصبی روشــی محاسباتی هســتند که بر پایه اتصال چندین واحد پردازشی، ایجاد می‌شــوند. شبکه مورد نظر از تعداد دلخواهی سلول یا گره[۴۱] یا نرون[۴۲] تشکیل می‌شود که مجموعه ورودی را به مجموعه خروجی ارتباط می‌دهند. در واقع شــبکه‌های عصبی مصنوعی، تا حد زیادی الهام یافته از سیســتم‌های یادگیری طبیعی هســتند که درآنها یک مجموعه پیچیده از نرون‌های بــه هم متصل، در کار یادگیری دخیل می‌باشــند. به همراه هر اتصال یــک ارزش عددی که وزن[۴۳] نامیده می‌شــود وجود دارد. هر نرون ســیگنال‌هایی[۴۴] از نرون‌های متصل دریافت می‌کند. اگر جمع سیگنال‌های ورودی ازآستانه[۴۵] تجاوز یابد در این صورت نرون از کار می‌افتد. نرون‌ها به صورت گروهی لایه‌بندی می‌شوند.

یک شبکه لایه‌ای، لااقل شامل یک لایه ورودی[۴۶] و یک لایه خروجی[۴۷] می‌باشد. بین این لایه‌های ورودی و خروجی ممکن است لایه‌های پنهان[۴۸] نیز وجود داشته باشد. وقتی سیگنال یا پالسی به یک لایه ارســال می‌شود، این سیگنال از لایه بالایی شروع به فعالیت می‌کند و توسط نرون‌های آن لایه بررسی و اصلاح می‌شود. در حقیقت هر نرون قدرت سیگنال را بالا می‌برد و آن پالس را به لایه بعدی انتقال می‌دهد (صفایی، ۱۳۸۵).

انواع مختلف شبکه‌های عصبی تعداد متناوبی از لایه‌ها را دارند. نقشه‌های خود سازمان یافته[۴۹]

تنها یک لایه ورودی و خروجی دارند در حالی که شبکه عصبی پس انتشار خطا[۵۰] یک یا چند لایه پنهان اضافی دارد.برای اینکه به شبکه عصبی موجود توانایی آموختن داده شود، بعد از اینکه سیگنال از لایه اول شـــبکه به لایه پایینی می‌رود، بایســـتی اطلاعات هر نرون که روی سیگنال اثر می‌گذارد، به‌روزآوری و اصلاح شود، این رویه را به اصطلاح BP یا پس انتشار خطا گویند.

در شبکه‌های پس انتشار خطا، خروجی با نتیجه دلخواه، مقایسه می‌شود و خطاهای ایجاد شده در لایه‌های قبلی در شبکه عصبی با میزان‌سازی وزن‌های اتصال‌ها، مشخص می‌شوند. این فرآیند تا زمانی که یک نرخ اشتباه قابل قبول بدست آید، تکرار خواهد شد. شبکه‌های عصبی پس انتشار خطا، به منظور پیش‌بینی و طبقه‌بندی مسائل از انواع معتبر تلقی می‌شوند(KIRKDOS and et al, ۲۰۰۴).

نقشه‌های خود سازمان یافته، یک روش تصویرسازی و خوشه‌بندی از نوع آموزش بدون مربی[۵۱] می‌باشند. دو کاربرد معمول توپولوژی نقشه‌های خود سازمان یافته شامل لاتیس (شبکه) مستطیل[۵۲] جائی که هر نرون دارای ۴ مجاور است و لاتیس شش ضلعی[۵۳] جائی که هر نرون ۶ مجاور دارد، می‌باشد.

از معایب شبکه‌های عصبی این است که تفسیر روشی که این شبکه‌ها به تصمیمات خود می‌رسند، بســـیار دشوار است. از دیگر انتقادهای وارده بر شبکه‌های عصبی می‌توان به تعدادی از عوامل نظیر توپولوژی شبکه اشاره کرد که باید به طور تجربی، تعریف شوند. شبکه‌های عصبی مصنوعی بر خلاف روش‌های آماری کلاســـیک، مفروضاتی در خصوص ویژگی‌های توزیع داده‌ها و متغیرهای مســـتقل ورودی ندارند.

الگوریتم ژنتیک

این الگوریتم بر پایه فرضیه تکامل داروین بوده و کاربرد آن بر ژنتیک طبیعی استوار است. اصول اولیه الگوریتم ژنتیک توسط هلند[۵۴] و همکاران در سال ۱۹۶۲ ارایه شده است. بر این اساس موجودات دارای سازگاری بیشتر با محیط، با میزان بالاتری زنده مانده و تولید مثل می‌کنند، در نتیجه شانس حضور آنها در نسل‌های بعدی بیشتر است (موسی زادگان و ذکری، ۱۳۸۷).

کاربرد الگوریتم ژنتیک، مستلزم بیان هر یک از پاسخ‌های یک مسئله خاص، در قالب یک رشته از اعداد می‌باشد. به هر یک از این رشته‌ها که در واقع نشان دهنده یکی از جواب‌های مسئله خواهد بود؛ کروموزم و به اعداد تشکیل دهنده آن نیز ژن[۵۵] گویند. الگوریتم ژنتیک با تعدادی از جواب‌های گوناگون آغاز می‌شـــود که جمعیت نســـل اولیه را تشـــکیل می‌دهند. تعداد این جوابها در تکرارها و

نسل‌های بعدی ثابت هستند، ولی کیفیت آن‌ها به تدریج بهبود می‌یابد. در هر نسل جواب‌ها به کمک تابع هدف[۵۶] سنجیده شده و بر آن اساس، نسل بعد شکل می‌گیرد. بدین ترتیب برای تولید نسل بعد، تعدادی از جواب‌ها مستقیماً به نسل بعدی منتقل می‌شود و برخی دیگر از جواب‌ها با ایجاد جهش در کروموزوم آن‌ها، به جواب‌های دیگری تبدیل شده و به نسل بعد خود، منتقل می‌گردند. در واقع الگوریتم ژنتیک، جستجو در علم رایانه برای یافتن راه‌حل بهینه و مسائل مرتبط با جستجو است.

درخت تصمیم‌گیری

درخت تصمیم‌گیری، ساختاری باز رخداد نگر (بازگشتی) برای بیان یک فرآیند طبقه بندی متناوب می‌باشد که به وسیله مجموعه‌ای از صفات[۵۷] تشریح گردیده و یک وضعیت را به مجموعه‌ای گسسته از طبقات تخصیص می‌دهد (QUINLAN, ۱۹۸۷).

هر برگ درخت تصمیم‌گیری، نماینده یک طبقه می‌باشد. درخت تصمیم روش کارآمد ویژه‌ای برای ایجاد دسته‌بندی کننده[۵۸]ها از داده‌ها است. مهم‌ترین خصوصیت درخت‌های تصمیم، قابلیت آنها در شکستن فرآیند پیچیده تصمیم‌گیری به مجموعه‌ای از تصمیمات ساده‌تر است که براحتی قابل تفسیر هستند (یزدانی و کنگاوری، ۱۳۸۳).

نواحی تصمیم پیچیده سراسری (خصوصاً در فضاهایی با ابعاد زیاد) می‌توانند با اجتماع نواحی تصمیم محلی ساده‌تر در سطوح مختلف درخت تقریب زده شوند. برخلاف دسته‌بندی کننده‌های تک مرحله‌ای رایج که هر نمونه، روی تمام دسته‌ها امتحان می‌شود، در یک دسته‌بندی کننده درخت، یک نمونه روی زیر مجموعه‌های خاصی از دسته‌ها امتحان شده و محاسبات غیر لازم حذف می‌شوند. در دسته‌بندی کننده تک مرحله‌ای، فقط از زیر مجموعه‌ای از صفات، برای روش بین دسته‌ها استفاده می‌شود که معمولاً با یک معیار بهینه سراسری انتخاب می‌شود. در دسته‌بندی کننده درخت، انعطاف‌پذیری انتخاب زیر مجموعه‌های مختلفی از صفات در گروه‌های داخلی مختلف درخت وجود دارد، به شکلی که زیر مجموعه انتخاب شده به شکل بهینه بین دسته‌های این گروه را تفکیک می‌کند. این انعطاف‌پذیری ممکن است بهبودی در کارایی را نسبت به دسته‌بندی‌کننده‌های تک مرحله‌ای ایجاد کند.

نظریه مجموعه اولیه

تئوری مجموعه اولیه در سال ۱۹۹۱ بوسیله پالاک[۵۹] ارایه گردید. نظریه مجموعه اولیه تعمیمی از نظریه دوتایی[۶۰] می‌باشد که از نظریه‌های معروف در ریاضیات بازه هست (عربانی، کلانتری، ۱۳۸۱).

اگرچه مسئله کمیت اندازه‌گیری، کیفیت یادگیری و کمال دانش ناظران در مواجهه با اطلاعات ناقص هنوز به طور کامل حل نشده است، اما در هر حال نظریه مجموعه اولیه، ابزار قدرتمندی برای استخراج دانش از اطلاعات پر ابهام و ناقص می‌باشد. این نظریه می‌تواند برای شرح وابستگی بین صفات و به منظور سنجش میزان اهمیت صفات و ارتباط با داده‌های متناقض نیز به کار برده شود (kirkdos and et.al, ۲۰۰۴).

استدلال مبتنی بر مورد (استدلال مورد گرا)

استدلال مبتنی بر مورد، حل مسئله به روش استنتاجی[۶۱] است. این روش براساس استفاده از پاسخ مسائل قبلی، برای حل مسائل مشابه جدید، شکل گرفته است. استدلال مبتنی بر مورد، به عنوان روشی شناخته می‌شود که از نحوه رفتار انسان‌ها در برخورد با مسائل جدید الگوبرداری کرده است (فائز و قدسی‌پور و غضنفری، ۱۳۸۵) به این ترتیب که از تجربیات کسب شده در حل مسائل گذشته به عنوان راهنمایی برای حل مسئله جدید بهره می‌برد.

حل مسئله به روش CBR در برگیرنده ۴ عمل عمده است:

۱- بازیابی مورد[۶۲] مشابه بازیابی شده با مسئله جدید.

۲- استفاده از پاسخ مسئله مشابه بازیابی برای تهیه پاسخ پیشنهادی برای مسئله جدید.

۳- بازبینی در پاسخ پیشنهادی در صورت وجود مغایرت در شرایط مسئله جدید و مسائل بازیابی شده.

۴- نگهداری مورد جدید (مسئله جدید و پاسخ آن) برای استفاده در آینده. (همان، پیشین: ۵۷۱)

حوزه‌های کاربردی داده‌کاوی

به واسطه قابلیت‌های طبقه‌بندی و پیش‌بینی روش‌های داده‌کاوی، این فنون به منظور تسهیل فرآیند حسابرسی، پیش‌بینی عملکرد شرکت و کمک به ارزیابی ریسک اعتباری مورد استفاده واقع می‌شوند.

از حسابرسی به عنوان رشته‌ای قدیمی یاد می‌شود و متناوباً حسابرسان را به عنوان افرادی خشک

و رسمی و ایرادگیر مورد قضاوت قرار می‌دهند. در واقع این موضوع دیگر حقیقت ندارد. در سال‌های گذشته حسابرسان پی به افزایش شدید حجم معاملات و پیچیدگی‌های داده‌های مالی و غیر مالی صاحبکارشان برده‌اند (sirikulvaadhana, ۲۰۰۲).

به منظور حسابرسی این شرکت‌ها، حسابرسان پی‌درپی باید با داده‌های حجیم و ساختار پیچیده این داده‌ها سر و 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، اگر چه تطابق فنون داده‌کاوی در فرآیند حسابرسی در واقع زمینه‌ای نسبتاً جدید می‌باشد، با وجود این، داده‌کاوی نشان داده است که در بسیاری از کاربردهای تجاری مرتبط با حسابرسی نظیر کشف تقلب، حسابداری دادگاهی[۶۳] دارای مزیت‌های فراوانی خواهد بود. از طرفی دیگر رویدادهای اخیر نیز نشان دهنده مشکلات قابل توجهی در فرآیند حسابرسی است. سقوط انرون[۶۴] و آرتور اندرسن[۶۵] و سایر موارد مشابه، همچنین کاربرد رو به گسترش پختن دفاتر[۶۶] در حرفه حسابداری شواهد دیگری برای تغییر در فرآیند حسابرسی ایجاد نموده است.

بر اساس استاندارد حسابرسی شماره ۵۶ منتشر شده توسط انجمن حسابداران رسمی امریکا، حسابرس باید انتظارات خود را گسترش داده و آنها را با مبالغ ثبت شده یا نسبت‌ها مقایسه کند. رویه‌های تحلیلی اجازه بررسی صحت مانده یک حساب را بدون آزمایش و بررسی معاملات انفرادی مرتبط می‌دهد (Kirkdos and et.al, ۲۰۰۴).

گرایش جدید حسابرسی شامل مفاهیم ریسک اقتصادی[۶۷] بوده که بر اهداف راهبردی واحد تجاری تاکید دارد. در این رویکرد از بالا به پایین[۶۸] حسابرس این اهداف راهبردی را شناخته و در راستای فرآیند تجاری انجام وظیفه می‌کند. روش‌های داده‌کاوی مانند CBR، GA، NN و منطق فازی[۶۹] می‌تواند این رویکرد جدید حسابرسی مبتنی بر ریسک را تسهیل سازند.

در هر حال مقالات مرتبط با حوزه‌های کاربردی در حسابرسی غالباً اشاره بر پیش‌بینی ورشکستگی، پیش‌بینی تداوم فعالیت، مضیقه مالی و تقلب مدیریت دارند. از سویی دیگر، هسته مرکزی سایر مقالات شامل پیش‌بینی عملکرد شرکت و تخمین و ارزیابی ریسک اعتباری واحدهای تجاری است.

پیش‌بینی ورشکستگی

به نظر می‌رسد پیش‌بینی ورشکستگی مشهورترین عنوان برای کاربرد روش‌های داده‌کاوی در داده‌های مالی است. ورشکستگی شرکت‌ها باعث خسارات اقتصادی فراوانی برای مدیریت، سرمایه‌گذاران، اعتباردهندگان و کارمندان شده و هزینه‌های اجتماعی[۷۰] بسیاری نیز در برخواهد داشت. رویکردهای آماری مانند تحلیل تشخیصی و تحلیل لاجیت[۷۱] و پروبیت[۷۲] و همچنین رویکرد فنون هوشمند محاسباتی شامل سیستم‌های خبره[۷۳]، شبکه‌های عصبی مصنوعی و... در فرآیند پیش‌بینی ورشکستگی شرکت‌ها قابل استفاده می‌باشند.

اغلب تحقیقات پیش‌بینی ورشکستگی، عملکرد هر یک روش‌های فوق را مقایسه می‌کنند، اما به طور قطعی نمی‌توان اظهار داشت که کدام یک از این روشها، از دقت کلی بیشتری در مقایسه با سایر روش‌ها برخوردار است (Wermter and et.al,۲۰۰۷). پیش‌بینی ورشکستگی به وسیله داده‌های صورت‌های مالی نخستین بار از آثار آلتمن (۱۹۶۸) سرچشمه گرفته است. آلتمن استدلال می‌کند ورشکستگی فرآیندی بلندمدت است، از این‌رو صورت‌های مالی، می‌تواند حاوی هشدارهایی از رخداد یک ورشکستگی قریب‌الوقوع باشد. وی با به کارگیری تحلیل تشخیصی چندگانه[۷۴] مدلی را برای پیش‌بینی ورشکستگی شرکت‌ها ایجاد کرد. در همین حال بسیاری از محققان نیز مدل‌های جایگزین را با استفاده از فنون آماری گسترش می‌دادند.

لین[۷۵] و مک لین[۷۶] (۲۰۰۱) تلاش کردند تا با استفاده از ۴ روش متفاوت، ورشکستگی شرکتها را پیش‌بینی نمایند. ۲ مورد از این روشها، روشهای آماری (تحلیل تشخیصی ورگرسیون لجستیک[۷۷]) و دو مورد دیگر نیز از فنون یادگیری ماشین (درختان تصمیم‌گیری و شبکه‌های عصبی) بوده‌اند. همچنین آنها براساس روش‌های پیش گفته، الگورتیمی ترکیبی[۷۸] را نیز پیشنهاد کردند. نمونه مورد استفاده شامل ۱۱۳۳ شرکت تجاری انگلیسی بوده است. ۶۹۰ شرکت موفق و ۱۰۶ شرکت ناموفق به عنوان مجموعه آموزشی[۷۹] و ۲۸۹ شرکت ناموفق و ۴۸ شرکت موفق به عنوان مجموعه آزمایشی[۸۰] استفاده شدند. تلاشی برای تطبیق کمپانی‌های موفق و ناموفق صورت نگرفت.۳۷ مورد از نسبت‌های مالی، برگرفته از تراز نامه و صورت سود و زیان به عنوان متغیرهای ورودی[۸۱] انتخاب شد. دو روش انتخاب ویژگی[۸۲] بکار رفته، متغیرهای ورودی را با اعمال قضاوت انسان[۸۳] به ۴ و با اعمال تحلیل واریانس[۸۴] به ۱۵ مورد کاهش داده است. نتایج حاصل از روشهای شبکه‌های عصبی مصنوعی و درختان تصمیم برای هر دو انتخاب ویژگی مبتنی بر قضاوت انسان و تحلیل واریانس، قابل قبول‌تر

از دیگر روش‌ها بوده اســت. سرانجام نویسندگان مقاله، الگوریتمی ترکیبی را پیشنهاد می‌کنند که رای‌گیری وزن‌دار[۸۵] طبقه‌بندی‌کننده‌های مختلف را به کار می‌برد. تانگ[۸۶] و همکاران (۲۰۰۴) یک روش ترکیبی را با یکپارچه‌سازی شبکه‌های عصبی مصنوعی و سیستم‌های فازی به کار برده‌اند. این مدل که شبکه عصبی فازی خود سازمانده عمومی نامیده می‌شود، یک سیستم مبتنی بر قاعده است کــه مبتنی بر قانون فازی «اگر- آنگاه[۸۷]» بــوده که نهایتا منجر به خود تعدیلی(اصلاح) پارامترهای قوانین فازی با استفاده از الگوریتم‌های یادگیری می‌گردد. این الگوریتم‌ها از الگوی شبکه عصبی به دست می‌آیند. مزیت اصلی شبکه‌های عصبیِ فازی، توانایی در مدل‌سازی مسئله‌ای خاص با استفاده از یک مدل زبانی سطح بالای قابل فهم، بر خلاف عبارات ریاضی پیچیده می‌باشد. این مدل به منظور پیش‌بینی ورشکســتگی بانک به کار رفته اســت. متغیرهای ورودی شامل ۹ متغیر مالی بوده‌اند که در بررسی‌های قبلی اهمیت آن‌ها محرز شده است. نمونه مورد استفاده شامل داده‌هایی در خصوص ۲۵۵۵۵ بانک موفق و ۵۴۸ بانک ناموفق بوده‌اند. ۲۰ درصد از این داده‌ها به عنوان مجموعه آموزشی و ۸۰ درصد مابقی به عنوان مجموعه آزمایش انتخاب شدند. همچنین برای کاهش خطای نوع اول[۸۸]، نمونه نسبت به تعداد بانک‌های موفق و ناموفق، به میزان مساوی تعدیل گردید. نویسندگان هنگام اســتخراج داده‌ها از آخرین صورت‌های مالی، عملکرد مدل را ۹۳ درصد، هنگامی که این داده‌ها از صورت‌های مالی ســال قبل بدست آمد میزان عملکرد را ۸۵ درصد و با دریافت داده‌ها از صورتهای مالی دو ســال قبل عملکرد را ۷۵ درصد، گزارش نمودند. مدل مورد نظر، در حدود ۵۰ قاعده فازی اگر- آنگاه، تولید کرد که تعاملات بین ۹ متغیر ورودی منتخب و اثرشان بر سلامت مالی بانک‌های مورد بررسی را، تشریح نموده است.

شین[۸۹] و لی[۹۰] (۲۰۰۲) نیز مدلی مبتنی بر الگوریتم ژنتیک پیشنهاد دادند. آنها بر این واقعیت تاکید کردند که برخلاف شــبکه‌های عصبی مصنوعی، الگوریتم‌های ژنتیک می‌توانند قوانینی قابل فهم ایجاد کنند. در این مورد، الگوریتم‌های ژنتیک به منظور یافتن آســتانه‌های یک یا چند متغیر در بالا یا پایین ســطحی به کار رفتند که موقعیت یک واحد تجاری را بحرانی می‌سازد. این مدل از ساختاری قاعده‌مند بهره می‌برد که دارای ۵ شرط است که هر کدام به یک متغیر خارج از ۹ نسبت مالی، اشاره دارد. این شرایط با عملکرد عطف منطقی[۹۱]، ترکیب خواهند شد. مجموعه داده‌ها مرکب از ۲۶۴ شرکت موفق و ۲۶۴ شرکت ناموفق بوده و ۹ نسبت مالی به عنوان متغیرهای ورودی انتخاب گردیدند. ۹۰ درصد نمونه به عنوان مجموعه آموزشی و ۱۰ درصد نیز به منظور اعتبار سنجی به کار

رفته‌اند. دقت کلی روش مزبور در حدود ۸۰ درصد گزارش شد.

دیمیتراس[۹۲] و همکاران (۱۹۹۸) به بررسی پیش‌بینی ورشکستگی با استفاده از نظریه مجموعه اولیه پرداختند. مجموعه آموزشی مورد استفاده متشکل از ۴۰ شرکت ناموفق و ۴۰ شرکت موفق از میان واحدهای تجاری یونانی در یک دوره ۵ ساله و مجموعه آزمایش نیز از ۱۹ شرکت ناموفق و ۱۹ شرکت موفق تشکیل شده بود. ۱۲ نسبت مالی بر اساس نظر مدیر اعتباری بانکی یونانی به منظور ورود به جدول اطلاعات انتخاب گردید. سرانجام قواعد تصمیم[۹۳]، استنتاج شد و نتایج این روش با نتایج تحلیل تشخیصی و تحلیل لاجیت مقایسه و برتری این روش نسبت به سایر روش‌ها مشخص گردید.

پارک[۹۴] وهان[۹۵] نیز در سال ۲۰۰۲ در یک تحقیق براساس روش استدلال مبتنی بر مورد، مدلی برای پیش‌بینی ورشکستگی بانک‌ها طراحی کردند.

تداوم فعالیت و مضیقه مالی

رسوایی‌هایی حسابرسی در سال‌های اخیر، اهمیت تجزیه و تحلیل‌های حسابرسی‌های کامل را دو چندان ساخته است. براساس استاندارد حسابرسی شماره ۵۹ [۹۶] حسابرس باید توانایی صاحبکار خود را در مورد امکان ادامه فعالیت حداقل یک سال بعد از تاریخ ترازنامه ارزیابی کند.چنان چه نشانه‌هایی از وجود مشکلات مالی دیده شود که احتمالا منجر به ورشکستگی خواهد شد، حسابرس می‌باید در خصوص تداوم فعالیت صاحبکار اظهارنظر کند. ارزیابی وضعیت تداوم فعالیت شرکت‌ها کارچندان ساده‌ای نیست. مطالعات نشان می‌دهند بخشی نسبتاً کوچک از واحدهای تجاری ورشکسته، واجد شرایط مبنای تداوم فعالیت هستند (KIRKDOS and et.al, ۲۰۰۴). به هرحال جهت تسهیل وظیفه حسابرسان و وظایف آنها در مورد وضعیت تداوم فعالیت شرکت‌ها، روش‌های آماری و یادگیری ماشین پیشنهاد شده‌اند.

کاه[۹۷] (۲۰۰۴) در بررسی موردی[۹۸]، پیش‌بینی تداوم فعالیت روش‌هایی نظیر شبکه‌های عصبی مصنوعی، درختان تصمیم و رگرسیون لجستیک را بایکدیگر مقایسه کرده است. نمونه مورد استفاده شامل ۱۶۵ شرکت دارای فعالیت و ۱۶۵ شرکت فاقد فعالیت بوده است. ۶ نسبت مالی برگزیده به عنوان متغیرهای ورودی انتخاب شدند. نهایتاً، نتایج نشان می‌دهد روش درختان تصمیم نسبت به دو روش دیگر از عملکرد بهتری برخوردار بوده است.

تان[۹۹] و دیهاردجو[۱۰۰] (۲۰۰۱) براساس تحقیقی از تان، سعی در پیش‌بینی مضیقه مالی

اتحادیه‌های اعتباری استرالیا[۱۰۱] براساس روش شبکه‌های عصبی مصنوعی داشته اند. در تحقیق قبلی از تان، داده‌های مالی سه ماهه استفاده شد و تلاش شد که درماندگی مالی براساس رویکردی سه ماهه پیش‌بینی گردد. آنها مدل مزبور را با وارد کردن مفهومی به نام پیش یاب[۱۰۲] ارتقاء دادند. بدین ترتیب وقتی مدل، پیش‌بینی می کند یک اتحادیه اعتباری در سه ماهه‌ای خاص دچار درماندگی مالی خواهد شد و واقعاً آن اتفاق در سه ماهه بعدی می‌افتد، به آن سه ماهه پیش یاب گفته می‌شود. این روش نسبت به روش قبلی به لحاظ میزان خطای نوع دوم[۱۰۳] از عملکرد بهتری برخوردار بوده است. در این مطالعه ۱۳ نسبت مالی به عنوان متغیر ورودی، مورد استفاده واقع شده و مشاهدات نمونه شامل ۲۱۴۴ مورد بوده است. نتایج بدست آمده با نتایج مدل پروبیت مقایسه و به طور نامحسوسی قابلیت آن برای میزان خطای نوع اول بیشتر تخمین زده شد. کنو[۱۰۴] و کوباشی[۱۰۵] (۲۰۰۰) روشی را برای رتبه‌بندی واحدهای تجاری براساس فنون برنامه‌نویسی ریاضی ارایه کردند. این مدل هیچ فرضی در خصوص توزیع تحمیل نمی کند.

سه روش جایگزین شامل تشخیص مبتنی بر تقابل مربعی[۱۰۶] (رویه درجه دوم)، تشخیص مبتنی برابر صفحه[۱۰۷] و تشخیص مبتنی بر رویه بیضوی[۱۰۸] بوده‌اند. ۶ نسبت مالی بدست آمده از صورت‌های مالی به عنوان متغیرهای ورودی و داده‌های نمونه شامل ۴۵۵ واحد تجاری بوده‌اند. در نهایت مدل برای هر واحد تجاری امتیازی خاص محاسبه کرده است.

تقلب مدیریت

تحقیقات نشان داده‌اند که پیش‌بینی تقلب، تقریباً غیر ممکن است. در واقع بسیاری از خصوصیات و ویژگی‌های مرتبط با جرم کارمندی[۱۰۹]، دقیقاً همان‌هایی هستند که سازمان‌ها به هنگام استخدام کارمندان خود در جستجوی آنها هستند (ردگر و همکاران، ۲۰۰۳). تقلب مدیریت، تقلبی فراگیر است که مدیریت بواسطه تحریف صورت‌های مالی مرتکب می‌شود. تقلب مدیریت می‌تواند منافع مسئولین مالیاتی، سهامداران و اعتباردهندگان را با مخاطره مواجه کند.

اسپاتیس[۱۱۰] (۲۰۰۲) از دو مدل برای شناسایی تحریف صورت‌های مالی از طریق داده‌های قابل دسترس عمومی استفاده کرد. متغیرهای ورودی برای نخستین مدل شامل ۹ نسبت مالی بوده و برای دومین مدل امتیاز z[۱۱۱] به عنوان متغیرهای ورودی به منظور تطبیق رابطه میان درماندگی مالی و دستکاری ارقام مالی، اضافه گردید. این روش از رگرسیون لجستیکی بهره می‌برد. داده‌های نمونه

مرکب از ۳۸ مورد تحریف صورت‌های مالی و ۳۸ مورد عدم تحریف بوده است. نتایج هر دو مدل نشان داد سه متغیر، با ضریبِ معنا داری[۱۱۲] وارد مدل شده‌اند.

پیش‌بینی عملکرد واحد تجاری

با وجود چالش‌های ناشی از تفکیک مالکیت از مدیریت و تضاد منافع بین گروههای ذی‌نفع، شاهد هستیم از معیارهای حسابداری نظیر سود هر سهم، نسبت بازدهی حقوق صاحبان سهام و یا بازدهی سرمایه‌گذاری به منظور ارزیابی عملکرد واحدهای تجاری استفاده می‌شود (ایزدی نیا، ۱۳۸۴).

لم[۱۱۳] (۲۰۰۳) مدلی را به منظور پیش‌بینی نرخ بازده حقوق سهامداران عادی بکار گرفته است. وی از شبکه‌های عصبی پس انتشار خطا در این خصوص استفاده کرد. بردار ورودی شامل ۱۵ نسبت مالی و یک متغیر از نوع تحلیل تکنیکال[۱۱۴] بود.

بک و همکاران[۱۱۵] (۲۰۰۱) اقدام به ارایه دو مدل به منظور گروه بندی شرکت‌ها براساس عملکرد نمودند. در هر دو مدل از نقشه‌های خود سازمان یافته استفاده شده بود. نخستین مدل با داده‌های مالی ۱۶۰ شرکت سر و کار داشته و بر اساس به کارگیری روش متن کاوی[۱۱۶] انجام یافت. دومین مدل نیز به تحلیل و بررسی گزارش‌های سالانه ارایه شده به مدیران ارشد اجرایی می‌پردازد. نویسندگان تفاوت‌هایی میان نتایج گروه بندی در دو روش مشاهده می‌کنند.

ارزیابی ریسک اعتباری

تجزیه و تحلیل ریسک اعتباری به واسطه افزایش تعداد شرکت‌های ورشکسته و پیشنهادهای رقابتی اعتبار دهندگان، تبدیل به یکی از محبوب‌ترین حوزه‌های مالی شده است. روش‌های داده‌کاوی به منظور تسهیل پیش‌بینی ریسک اعتباری می‌تواند به کار برده شود. هانگ و همکاران[۱۱۷] (۲۰۰۳) تجزیه و تحلیل رتبه‌بندی ریسک اعتباری[۱۱۸] را با استفاده از ماشین بردار پشتیبان[۱۱۹] که یک روش یادگیری ماشین است، اجرا کردند. دو مجموعه داده مورد استفاده در این تحقیق، شامل ۷۴ شرکت کره‌ای و دیگری شامل ۲۶۵ شرکت آمریکایی بوده است. برای هر مجموعه داده، ۵ نوع رتبه‌بندی تعریف شد. همچنین، دو مدل برای داده‌های شرکت‌های کره‌ای و دو مدل نیز برای شرکت‌های آمریکایی هر کدام با یک بردار ورودی به کار رفت. ماشین بردار پشتیبان و شبکه‌های عصبی پس از انتشار خطا،برای اهداف پیش‌بینی رتبه‌بندی اعتباری استفاده شدند. مدل گارسون[۱۲۰] نیز برای

اندازه‌گیری اهمیت نسبی ارزش‌های ورودی استفاده گردید.

مـــوس و همکاران[۱۲۱] (۲۰۰۴) نیز به منظور بررســـی قواعد ارزیابی ریســـک اعتباری، از نمودار تصمیم[۱۲۲] اســـتفاده کردند. نمودارهای تصمیم دارای مزیت‌های نظری نســـبت به درختان تصمیم می‌باشند. این نمودارها از تکرار زیر درخت‌های متناظر جلوگیری می‌کنند. مجموعه داده‌های مورد استفاده در این تحقیق، شامل داده‌های کشور آلمان و داده‌های اتحادیه بنولکس[۱۲۳] بوده‌اند. شبکه‌های عصبی نیز به منظور طبقه‌بندی به کار گرفته شد. همچنین با روش‌های استخراج قواعد[۱۲۴] به بررسی اقتباس قواعد از شـــبکه پرداخته شد. نتایج بررسی مورد نظر نشان داد، عملکرد روش‌های استخراج قواعد در مقایسه با روش شبکه‌های عصبی یکسان و نسبت به سایر روش‌ها برتری دارد. در آخر نیز، قواعد به شکل نمودار تصمیم نشان داده شد.

نتیجه گیری

فناوری داده‌کاوی از قابلیت‌های پیش‌بینی و طبقه‌بندی فراوانی برخوردار است. بنابراین می‌تواند فرآیند تصمیم‌گیری در مسایل مالی را تسهیل کند. کاربرد روش‌های داده‌کاوی با توجه به مطالعات مرتبط و ماهیت آنها، می‌تواند طیف گسترده‌ای شامل پیش‌بینی ورشکستگی، تخمین ریسک اعتباری، وضعیت تداوم فعالیت، مضیقه مالی، پیش‌بینی عملکرد واحد تجاری و تقلب مدیریت را در برگیرد. در این بین، پیش‌بینی ورشکســـتگی مشهورترین این عناوین است. همچنین، روش‌های داده‌کاوی به کار رفته در تحقیقات مختلف نیز شـــامل شـــبکه‌های عصبی مصنوعی، الگوریتم ژنتیک، درخت تصمیم‌گیری، نظریه مجموعه اولیه، استدلال مبتنی بر مورد و برنامه‌نویسی ریاضی هستند. بر اساس بررسی‌های صورت گرفته، روش شبکه‌های عصبی مصنوعی تاکنون یکی از محبوب‌ترین روش‌های داده‌کاوی بوده است. اگرچه تعداد زیادی از تحقیقات در ارتباط با کاربرد روش‌های داده‌کاوی در امور مالی بوده‌اند، لیکن حوزه‌های مستعد دیگری به منظور گسترش دامنه تحقیقات وجود دارند. معرفی مدل‌های ترکیبی، بهبود مدل‌های کنونی، اســـتخراج قواعدی جامع از شبکه‌های عصبی مصنوعی، بهسازی و یکپارچه‌سازی سیستم‌های ERP با ابزارهای داده‌کاوی می‌توانند گستره‌ای نوین و متفاوت از تحقیقات داده‌کاوی باشـــند. ارزیابی هزینه‌های منتســـب به خطاهای نوع اول و دوم، توسعه بردار ورودی بوســـیله اطلاعات کیفی وهمچنین بهره‌برداری از روش‌های معتبر انتخاب ویژگی، می‌توانند از جمله سایر روش‌های تحقیق در حوزه داده‌کاوی در آینده باشند. بدین ترتیب با ارتقای مدل‌ها و

روش‌های مورد استفاده، می‌توان تبدیل شدن داده‌کاوی به ابزاری مفیدتر در فرآیند تصمیم‌گیری به خصوص در حوزه‌های مالی را، به نظاره نشست.

پی‌نوشت‌ها:

1- Artificial Intelligence
2- Pattern Recognition
3- Database System
4- Statistics
5- Data Mining(DM)
6- Knowledge Discovery In Database (KDD)
7- Data Cleaning
8- Data Integration
9- Data Selection
10- Data Transformation
11- Pattern Evaluation
12- Machine Learning
13- Human-Computer Interface
14- Financial Classification Problem
15- Corporate Bankruptcy
16- Credit Risk Estimation
17- Going Concern Reporting
18- Financial Distress
19- Corporate Performance Prediction

۲۰- در این مقاله سعی شده به بیان کاربردهایی از داده‌کاوی بپردازیم که غالبا هسته مرکزی بیشتر پژوهش‌ها بوده‌اند. در هر حال سایر کاربردهای داده‌کاوی مطابق مقالات موجود می‌تواند شامل موارد زیر باشد:
- کاربرد داده‌کاوی در تحلیل‌های مرتبط با پولشویی
-کاربرد داده کاوی در حسابداری دادگاهی
-استفاده از تکنیک‌های داده کاوی در مدیریت ارتباط با مشتری
-استفاده از داده کاوی در بخش عمومی

21- Supply Chain
22- Data Ming Tool
23- Hidden Pattern
24- Www.Ncsa.Uiuc.Edu/Divisions/Dmv/Alg/D2k
25- Www.Sas.Com/
26- Www.Spss.Com/Spssbi/Clementine/
27- Www.3.Ibm.Com/Software/Data/Miner/
28- Www.Insightful.Com/Products/Product.Asp?Pid=26
29- Www.Acl.Com/Product/
30- Data Ming Technique
31- Prediction
32- Description

۳۳- مجموعه‌ای از قواعد و مراحل منظم که در حل مسائل ریاضی به صورت ثابت و متحد الشکل بکار می‌روند

34- Visualization
35- Neural Network (NN)

36- Genetic Algorithms(GA)
37- Decision Tree
38- Rough Set Theory(RST)
39- Case Base Reasoning(CBR)
40- Mathematical Programming
41- Node
42- Neuron
43- Weight
44-Signal
45-Threshold
46- Input Layer
47- Output Layer
48- Hidden Layer
49- Self-Organizing Map(SOM)
50- Back Propagation

۵۱- درشـــبکه خود سازمان یافته، از روش یادگیری رقابتی برای آموزش استفاده می‌شود و مبتنی برمشخصه‌های خاصی از مغز انسان توسعه یافته است. سلول‌ها در مغز انسان در نواحی مختلف طوری سازمان دهی شده‌اند که در نواحی حسی مختلف، با نقشه‌های محاسباتی مرتب و معنی‌دار ارایه می‌شوند. برای نمونه ورودی‌های حسی لامسه شـــنوایی، و... با یک ترتیب هندســـی معنی دار به نواحی مختلف مرتبط هستند. در یک شبکه خود سازمان یافته، واحدهای پردازشـــگر در گره‌های یک شبکه یک بعدی، دو بعدی یا بیشتر قرارداده می‌شوند. واحدها در یک فرآیند یادگیری رقابتی نســـبت به الگوهای ورودی، منظم می‌شوند. محل واحدهای تنظیم شده در شبکه به گونه‌ای نظم می‌یابد که برای ویژگی‌های ورودی، یکدستگاه مختصات معنی‌دار روی شبکه ایجاد شود. لذا نقشه خود سازمان یافته یک نقشه توپوگرافیک از الگوهای ورودی را تشکیل می‌دهد که در آن محل قرار گرفتن واحدها، متناظر ویژگیهای ذاتی الگوهای ورودی اســـت. یادگیری رقابتی که در این قبیل شبکه‌های بکار گرفته می‌شود بدین صورت است که در هرقدم یادگیری، واحدها برای فعال شـــدن با یکدیگر به رقابت می پردازند. در پایان یک مرحله رقابت تنها یک واحد برنده می‌شود، که وزن‌های آن نسبت به وزن‌های سایر واحدها به شکل متفاوتی تغییر داده می‌شود. این نوع از یادگیری را یادگیری بدون مربی یا بی‌نظارت گویند.

52- Rectangular Lattice
53- Hexagonal Lattice
54- Holland
55- Gene

۵۶- تابع هدف جهت تعیین اینکه افراد چگونه در محدوده یک مساله ایفای نقش می‌کنند، استفاده می‌شود.

57- Attribute
58- Classifier
59- Pawlak
60- Twin

۶۱- به عنوان زیر مجموعه‌ای گسترده از هوش مصنوعی، یادگیری ماشین وظیفه طراحی و توسعه الگوریتم‌هایی را به عهده دارد که به کامپیوتر اجازه یادگیری می‌دهند. به طور کلی همواره با دو نوع یادگیری استنتاجی و استقرایی روبرو هستیم.

62- case
63- Forensic Accounting
64-Enron
65- Arthur Andersen

۶۶-به معنی یکدست نشان دادن سود می‌باشد. (Book Cooking)
۶۷- احتمال خطر اینکه واحد اقتصادی از پوشش هزینه‌های عملیاتی خود عاجز باشد.

68- Top Down Approach
69- Fuzzy Logic
70- Social Cost
۷۱- مدل لاجیت با اختصاص وزن‌هایی به متغیرهای مستقل، رتبه هریک از شرکتهای نمونه را پیش‌بینی می‌کند. از این رتبه برای تعیین احتمال عضویت در یک گروه معین(برای مثال موفق یا ناموفق)استفاده می‌شود.
72- Probit
73- Expert System
74- Multiple Discriminant Analysis
75- lin
76- mcclean
۷۷- رگرسیون لوجستیک یک مدل آماری رگرسیون برای متغیرهای وابسته دودویی است. این مدل را می‌توان به عنوان مدل خطی تعمیم یافته‌ای که از تابع لوجیت به عنوان تابع پیوند اســتفاده می‌کند و خطایش از توزیع چند جمله‌ای پیروی می‌کند، به حساب آورد.
78- Hybrid Algorithm
79- Training Set
80- Testing Set
81- Input Variable
82- Feature Selection
83- Human Judgment
84- Analysis Of Variance(ANOVA)
۸۵- یک سیستم رای‌گیری است بر این پایه که تمام رای‌دهندگان لزوما با یک دیگر یکسان نمی‌باشند.
86- Tung
87- If-Then
۸۸- رد کردن فرضیه صفر که در واقع درست است.
89- shin
90- Lee
91- Logical AND Operator
92- Dimitras
93- Decision Rule
94- Park
95- Han
96- Statement Of Auditing Standards No.59:"The Auditors Consideration Of An Entity's Ability To Continue As A Going Concern"
97- KOH
98- Case study
99- Tan
100- Dihardjo
101- Australian Credit Union
102- Early Detector
۱۰۳- عدم رد فرضیه صفر که در واقع غلط است.
104- konno
105- Kobashi
106- Discrimination By Hyperplane
107- Discrimination By Quadratic Surface
108- Discrimination By Elliptic Surface
109- White –Collar Criminal
110- Spathis

111- Z Score
112- Significant Coefficient
113- Lam
114- Technical Analysis Variable
115- Back and et.al
۱۱۶- امروزه بانک‌های متنی بواسطه افزایش اطلاعات به شکل الکترونیکی مانند نشر الکترونیکی، کتابخانه‌های دیجیتال، پست الکترونیکی و وب جهانی به سرعت در حال رشد می‌باشند. متن کاوی در واقع نوعی از روش‌های داده کاوی است که به جستجو و استخراج اطلاعات جدید در مجموعه بزرگ داده‌های متنی، می‌پردازد.
117- Hung and et.al
118- Credit Rating Analyzing
119- Support Vector Machine(SVM)
120- Garson Model
121- Hung and et.al
122- Decision Diagram
۱۲۳- اتحادیه گمرکی بلژیک و هلند و لوکزامبوزگ که به تدریج معنی شخصیتهای حقوقی این سه کشور را نیز به خود گرفته است.
124- Rule Extraction

منابع و ماخذ:

۱- اصغری اسکویی، محمد رضا(۱۳۸۱) کاربرد شبکه‌های عصبی در پیش‌بینی سریهای زمانی، فصلنامه پژوهش‌های اقتصادی ایران، شماره ۱۲.
۲- ایزدی‌نیا، ناصر(۱۳۸۴) نقدی بر معیارهای حسابداری ارزیابی عملکرد و پیشنهاد معیارهای ارزش افزوده اقتصادی و جریان‌های آزاد نقدی برای گزارشگری ارزش‌های واحد تجاری، مجله دانشکده علوم اداری و اقتصاد دانشگاه اصفهان.
۳- بهرامفر، نقی و جواد ساعی (۱۳۸۵) ارایه مدل برای پیش‌بینی عملکرد شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از اطلاعات مالی منتشره، بررسی‌های حسابداری و حسابرسی، شماره ۴۵.
۴- صادقی مقدم، محمدرضا، امیر افسر و بابک سهرابی (۱۳۸۴) مدلسازی جریان مواد زنجیره تأمین با رویکرد الگوریتم ژنتیک، مجله علمی پژوهشی مدرس (علوم انسانی).
۵- صفایی، امین(۱۳۸۵) چگونه یک شبکه عصبی هوشمند بسازیم؟ - مثالی از برنامه‌نویسی شیءگرا در شبکه‌های عصبی و هوش مصنوعی، ماهنامه شبکه، شماره ۷۱.
۶- صورتگر، ابوالفضل و علیرضا داوودآبادی (۱۳۸۵) تشخیص اشغال کاذب در تجهیزات مخابراتی توسط شبکه‌های عصبی، مجله الکترونیکی پژوهشگاه اطلاعات و مدارک علمی ایران، شماره اول دوره ششم، قابل دسترس به صورت آنلاین در:
davoodabadi_abs.htm/http://www.irandoc.ac.ir/data/E_J/vol۶
۷- عربانی، مهیار و فرزین کلانتری (۱۳۸۱) ارزیابی استفاده از تئوری Rough Set در سنجش عوامل ناپایداری شیب‌ها، سومین همایش بین‌المللی مهندسی ژئوتکنیک و مهندسی خاک ایران.
۸- فائز، فرهاد، حسن قدسی‌پور و مهدی غضنفری (۱۳۸۵) ارایه یک مدل تصمیم یار برای انتخاب فروشنده با استفاده از روش استدلال مبتنی بر مورد در محیط فازی، نشریه دانشکده فنی، شماره ۴.
۹- مهمد کانتاردزیک (۲۰۰۲) داده‌کاوی، ترجمه امیر علیخانزاده. چاپ اول، بابل: انتشارات علوم رایانه.
۱۰- موسی زادگان، حسنعلی و حسام‌الدین ذگردی (۱۳۸۷) مدلی جدید برای حل مسئله موازنه خط مونتاژ هزینه گرا، نشریه بین‌المللی علوم مهندسی دانشگاه علم و صنعت ایران، شماره ۱۹.
۱۱- نجات، امیر رضا و آرش علی اکبری(۱۳۸۷) داده‌کاوی راهی به سوی ناشناخته‌ها. دو ماهنامه توسعه انسانی پلیس، شماره ۱۸.

۱۲- همایون‌پور، مهدی و داریوش حکیم‌زاده (۱۳۸۱) سنجش وضعیت در ماشین‌های الکتریکی با استفاده از شبکه‌های عصبی، هشتمین کنفرانس انجمن کامپیوتر ایران.
13- Abd Rahman (2008). Utilization Of Data Mining Technology Within The Accounting Information System In The Public Sector: A Country Study – Malaysia. Degree of Doctor of Philosophy, University of Tasmania.

14- Dass.R(2006).Data Mining In Banking And Finance: A Note For Bankers. Available on line at: Http://Citeseerx.Ist.Psu.Edu/Viewdoc/Download?Doi=10.1.1.102.7502&Rep=Rep1&Type=Pdf
15- Kirkos S. and Manolopoulos Y., (2004), 'Data Mining in Finance and Accounting: A Review of Current Research Trends', Proceedings of the 1st International Conference on Enterprise Systems and Accounting (ICESAcc), Thessaloniki,Greece, pp. 63-78.
16- Rui. Hort, Beatriz De S. L. Pires De Lima And Carlos Cristiano H. Borges (2009).Data Preprocessing Of Bankruptcy Prediction Models Using Data Mining Techniques. Available on line At: Www.Blog.Campe.Com.Br/Wp-Content/Uploads/2009/03/Witpress_Conf-2.Pdf.
17- James A. Rodger, Parag C. Pendharka, David J. Pape And George Letcher(2003). Utilization Of Data Mining Techniques To Detect And Predict Accounting Fraud: A Comparison Of Neural Networks And Discriminant Analysis. Available on line At: Http://Portal.Acm.Org/Citation. Cfm?Id=778616.
18- Gary M. Weiss, Maytal Saar-Tsechansky And Bianca Zadrozny(2005). Utility-Based Data Mining. Available on line At: Www.Springer.Com/Journal/10618.
19- Gersper.m(2007). Data Mining Helps. Available on line At:
www.Gdmllc.Com/Pdf/Data_Mining_Help_Financial_Executives.Pdf.
20- Hung,C , Chen,J, Wermter, S. (2007). Hybrid Probability-Based Ensembles For Bankruptcy Prediction. International Conference on Business And Information, July 11-13, 2007, Tokyo, Japan.
21- Pechenizkiy, S. Puuronen And A. Tsymbal(2006). On The Use of Information Systems Research Methods in Data Mining. Information Systems Development. 487-499
22-Quinlan.J. Generating Production Rules from Decision Trees (1987). IJCAI'87 Proceedings of the 10th international joint conference on Artificial intelligence - Volume 1
23- Sirikulvadhana. S (2002). Data Mining As A Financial Auditing Tool. Available on line At: Pafis.Shh.Fi/Graduates/Supsir01.Pdf.
24- Jeffrey W. Seifert (2004). Data Mining: An Overview. Available on line At: Www.Fas.Org/Irp/Crs/Rl31798.Pdf.