

کاربرد تکنیک های داده کاوی در حسابداری: بازنگری تحقیقات اخیر و چشم انداز تحقیقات آتی

مجتبی ابراهیمی رومنجان

کارشناس ارشد حسابداری و مدیر گروه حسابداری علمی کاربردی جهاد دانشگاهی خراسان جنوبی-

Ebrahimi362@yahoo.com

محمد خراشادی

کارشناس ارشد حسابداری و مدیر گروه حسابداری دانشگاه فنی و حرفه ای (آموزشکده فنی پسران بیرجند)

عاطفه پیروزمند

دانشجوی کارشناسی حسابداری دانشگاه علمی کاربردی جهاد دانشگاهی خراسان جنوبی

فاطمه ایزدمهر

دانشجوی کارشناسی حسابداری دانشگاه علمی کاربردی جهاد دانشگاهی خراسان جنوبی

چکیده

امروزه با پیچیده تر شدن مسائل مربوط به حوزه مالی، علوم دیگر به مدد این حوزه آمده اند. کاربرد ابزار های داده کاوی در امور مالی و حسابداری از آن جمله هستند که در پیش بینی ورشکستگی، وضعیت تداوم فعالیت و نیز پیش بینی بحران مالی و کشف تقلبات مدیریت، برآورد ریسک اعتباری و تخمین عملکرد شرکت، توسط محققین بکار رفته اند. این مقاله مقصد دارد ضمن معرفی ابزارهای داده کاوی و کاربرد آنها در تحقیقات مالی و حسابداری، مطالعات انجام گرفته مربوط به این حوزه را به صورت اجمالی بررسی کرده و برای تحقیقات آتی نیز ارائه دیدگاه نماید.

dataacademy.ir

واژه های کلیدی: تکنیک های داده کاوی، حسابداری، مالی، حسابرسی.

مقدمه

داده کاوی یکی از شاخه های علم کامپیوتر محسوب می شود که بعد از دهه ۸۰ ایجاد شده واز مفاهیم و روش های هوش مصنوعی، تشخیص روند (الگوریتم مورچگان)، پایگاه داده ها و آمار استفاده می کند. داده کاوی در حوزه های پیچیده اطلاعات بکار می رود و به این دلیل، داده کاوی را معادل (KDD) یا کشف دانش از پایگاه داده ها قرار می دهند.

اطلاعات مالی توسط سازمان هایی شبیه بانک ها، تحلیلگران بورس اوراق بهادار، ممیزان مالیاتی، حسابداران بزرگ و حسابرسان دفاتر تخصصی اطلاعات غیره جمع آوری می شوند. در برخی موارد نیز به صورت عمومی در دست می باشند. داده کاوی در اطلاعات مالی، در حل مسائل تسهیل فرآیند تصمیم گیری مدیران کاربرد دارد. مواردی نظیر ورشکستگی شرکت، برآورد ریسک اعتباری، گزارشگری تداوم فعالیت، پریشانی مالی و توصیف عملکرد شرکت و...

اهمیت داده کاوی در امور مالی و حسابداری در بسیاری از سازمانها مشاهده شده است. جامعه حسابداران رسمی آمریکا داده کاوی را به عنوان یکی از ده روش برتر برای آینده تعریف کرده است. و انجمن حسابرسان داخلی آمریکا داده کاوی را به عنوان یکی از چهار تحقیق برتر این رشته معرفی کرده است.

تحقیق در داده کاوی در امور مالی و حسابداری و کاربرد آن، به طور نسبی یک رشته تحقیقی جدید را به وجود آورده است. لذا در تحقیق حاضر به بازنگری تحقیقات انجام شده در این حوزه پرداخته شده است. و در ابتدا انواع خاص تحقیقات فعلی انجام شده و روش های مربوط تشریح می شود. اهمیت مقاله حاضر در چهار مورد خلاصه می شود:

- ۱- معرفی کاربردهای خاص مالی در مورد روش های داده کاوی
- ۲- علت توسعه و کاربرد روشهای داده کاوی در چیست؟ آیا این روشها، مدل های سنتی را نقض می نمایند؟
- ۳- روی چه نوع از داده هایی این روش ها عمل می کنند؟
- ۴- روش های انتخاب شده که برای آینده تحقیقات به کاربرده می شوند چه هستند؟

از طرف دیگر این مقاله نقاط مفید برای انجام کار تحقیقی در این حوزه را معرفی می نماید. لذا در بخش بعدی ادبیات موضوع مطرح و سپس خلاصه ای از تحقیقات توصیفی ارائه می شود و در ادامه زمینه های کاربردی داده کاوی معرفی شده اند. در انتهای مقاله نیز چشم انداز تحقیقات آتی ارائه شده است.

ادبیات تحقیق

واژه داده کاوی بر پایه تعداد زیادی الگوریتم ها قرار گرفته است. مدل ها و روش های گرفته شده از نفوذ آمار، ماشین های یاد گیری، پایگاه داده ها و شبکه های عصبی و برخی از این روش ها برای آزمایش و آزمون اطلاعات مالی به کار می روند. عمده روش های داده کاوی که در این مقاله مورد بررسی قرار خواهند گرفت عبارتند از: شبکه های عصبی، الگوریتم های ژنتیک، دخت تصمیم، برنامه ریزی خطی و غیره.

شبکه های عصبی: شبکه های عصبی یک تکنولوژی کامل با تئوری شناخته شده است. شبکه های عصبی یک تعداد عصب به هم پیوسته هستند. پیوستگی با هر اتصال، یک عدد ارزشی است که «وزن یا اهمیت» نامیده می شود. هر عصب سیگنالهایی از اتصال اعصاب دریافت می کند.

اگر ورودی ها از یک حد خاصی تجاوز کند آنگاه اعصاب به حرکت در می آید. ارزش ورودی توسط تابع انتقال اعصاب اندازه گیری می شود.

اعصاب در لایه هایی طبقه بندی شده اند. لایه های شبکه ای حداقل یک لایه ورودی و خروجی دارد. بین لایه ورودی و خروجی ممکن است یک لایه پنهان یا بیشتر وجود داشته باشد. تفاوت انواع شبکه های عصبی تفاوت در تعداد لایه های عصبی می باشد .

نقشه های سازمانی تنها یک ورودی و یک خروجی دارند. هرچند یک شبکه عصبی یک یا چند لایه پنهان دارد . بعد از تعریف معماری شبکه، بایستی مرتب شوند. در شبکه الگویی به کاربرده می شود که لایه ورودی و خروجی نهایی محاسبه می شوند در لایه خروجی، خروجی با نتایج پیش بینی شده مقایسه می شود و انحرافات با برگشت به عقب در اتصالات هم وزن می شوند. این نتایج تکرار می شود تا زمانی که یک نرخ اشتباه قابل قبول به دست آید. از این روش برای اثبات و طبقه بندی مشکلات استفاده می شود .

نقشه های سازمانی یک روش خوشه بندی و بصری برای یادگیری هدایت شده می باشند . برای هر بخش ورودی تنها یک عصب خروجی فعال خواهد بود . آن بخش موفق تر می باشد که مطابق با بخش های ورودی بروز رسانی شود. بنابراین ورودی های مشابه نگاشته می شوند در نزدیکی یا شبیه عصب های خروجی به شکل خوشه در استفاده عادی از نقشه های سازمانی هستند شبکه های خوشه دار جایی که هر نرون چهار گوشه دارد و شبکه شش گوشه جایی که هر نرون شش گوشه دارد .

یک اشکال عمده شبکه های عصبی این است که آنها به عنوان باکسهای خالی عمل می کنند و درک این تغییر که چرا شبکه های عصبی و چگونه به تصمیمات آنها می رسند برای انسان ها مشکل است.

به هر حال الگوریتم ها برای استخراج نقشه های دریافتنی از شبکه های عصبی پیشنهاد می شوند. انتقاد دیگر شبکه های عصبی این است که یک عده پارامترهایی شبکه، شبکه های جغرافیایی باید به صورت آزمایشی تعریف شوند.

به نظر می رسد که شبکه های عصبی تلاش می کنند که محققان را در حوزه های مربوط به ساختار و اصول کاری آنها جلب کنند. آنها محقق را قادر می سازند که مشکلاتی که یک الگوریتم توان حل آن را ندارد به راحتی مرتفع نمایند زیرا آنها از مثال ها استفاده می کنند و به مشاهدات جدید تعمیم می دهند. آنها می توانند الگوهای مشاهده نشده قبلی را طبقه بندی کنند. آنها این ویژگی را دارند که سروکار داشته باشند با اطلاعات ناقص، مبهم و بهم ریخته، بر خلاف روش های سنتی آماری، در توزیع موسسات اطلاعات قدیمی استفاده نمی کنند.

الگوریتم ژنتیک

این الگوریتم ایده هایی را از سیر تکاملی طبیعی بکار می برد که به افراد در ادامه نسل آنها کمک می کند. الگوریتم ژنتیک در باقی ماندن و توسعه مجدد جامعه کمک می کند.

کروموزوم یک رشته خاصی است که یک نقطه در زمان خاص رانشان می دهد. جامعه، مجموعه ای از کروموزوم هاست. بعد از تعیین تصادفی اولیه جامعه هر کروموزوم به عنوان یک تابع کمکی ارزیابی می شود. نقش تابع کمکی این است که عملکرد کروموزوم را ارزیابی می کند.

dataacademy.ir

سه عملکرد کروموزوم ها را به کار می برد: تکثیر: تکثیر افراد توسط تکرار خودشان با یک احتمال مشابه.

دورگه: هنگامی که دو کروموزوم بطور متقابل برخی از بیت ها تغییر می دهند و یک کروموزوم جدید ایجاد می کنند.

تغییر (دگرگونی): وقتی که تغییر (دگرگونی) در یک کروموزوم ساده با تغییر یک یا چند بیت اتفاق می افتد.

درخت تصمیم

درخت تصمیم یک روش پیش بینی و طبقه بندی است که به طور متوالی مشاهدات را در یک زیر گروه خاص به طور متقابل تحریک می کند. این روش برای یافتن نسبتی که بهترین تفکیک را برای نمونه طبقات افراد جستجو می کند. این زیر گروه ها به طور متوالی تقسیم می شوند تا زمانی که زیر گروه ها کوچک شوند یا همچنین تفاوت های آماری بی اهمیت موجود بین زیر مجموعه مشاهده شود. اگر درخت تصمیم بزرگ باشد در نهایت با این اقدام به اصطلاح هرس خواهد شد.

مجموعه های سخت

این تئوری توسط پادلاک در ۱۹۸۲ مطرح شد. توسعه تئوری تنظیم با اندیشه عضویت احتمالی اقلام در یک مجموعه رخ داد. تعیین یک طبقه C که پایین ترین تقریب C شامل نمونه هایی است که به طور معین به C تعلق دارند. تخمین بالای C شامل نمونه هایی که نمی توانند به عنوان آنها بی که تعلق ندارند به C تعریف شوند. این نوع تئوری ممکن است در تشریح ارتباط بین ویژگی ها استفاده شود، در حدی که با اطلاعات ناسازگار و ابهام سروکار دارد.

استدلال مبتنی بر موضوع

این روش یک روش حل مسئله با دلیل یابی از روش های مسائل حل شده مشابه برای حل یک مشکل است. این روش تلاش می کند برای بازیابی یک مورد مشابه در مورد مینا، کلیه موضوعات در این روش نوعی مشابه سازی در اندازه گیری بوده و بازیافتن یک مورد مشابه می باشد. برای شبیه سازی بین دو مورد با استفاده از محاسبات و تفاوت هندسی بین آنها تلاش می شود. این روش فرض می کند که تمام ترکیب ها (خصوصیات) به طور مساوی مربوط هستند.

حوزه های کاربرد و مطالعاتی تحقیق خاص

طبقه بندی بدهی ها و توان پرداخت آنها، در روش های داده کاوی با تحلیل فرآیند حسابرس در پیش بینی عملکرد شرکت و تحلیل ریسک اعتباری بر آوردی گسترش یافته اند.

در رشته حسابرسی، روش های داده کاوی در حل این قبیل مسائل کمک می کند. مطالعات اخیر نشان می دهد که این مسائل در حسابرسی قابل مشاهده می باشند. انحلال انرون و ارتور اندرسون و دیگر شرکت های مشابه در این حوزه قرار می گیرند.

بر اساس استاندارد حسابرسی ۵۶ آمریکا، حسابرس نظر خود را در خصوص انتظارات مالکان ارائه می دهد و این انتظارات در مبالغ ثبت شده و یافته های کشف شده نشان داده می شوند. در انجام این وظیفه حسابرسی بررسی های تحلیلی رابه کار می برد که ارتباط بین مبالغ پیش بینی شده را با مشاهدات واقعی و اقلام اطلاعات در واقعیت را مقایسه می کند. بررسی های تحلیلی این اجازه را می دهد که دقت یک مانده حساب را بدون آزمون کردن معاملات اشخاص و وابسته آزمون کند فرارز روش های مشاهدات تحلیلی به عنوان روش های غیر کمی به عنوان یک بررسی اجمالی راطبقه بندی می کند. روش مقداری ساده مثل بررسی روند یا نرخ آزمون منطقی و روش مقداری پیشرفته مثل تحلیل رگرسیون و شبکه های عصبی .

جهت گیری امروزی در حسابرسی مفهوم ریسک تجاری که تأکید دارد به مشاهدات استراتژیک یک واحد تجاری رادر بر می گیرد. در این روش از بالا به پایین حسابرس می فهمد که مشاهدات استراتژیک به سمت پایین در فرآیند موسسه کار می کند.

مقالات مرتبط با حوزه های کاربردی خاص در رشته حسابرسی اشاره دارد به پیش بینی ورشکستگی، پیش بینی تداوم فعالیت و پریشانی مالی و تقلبات مدیریت .

پیش بینی ورشکستگی

این موضوع به نظر می رسد که یکی از پر کاربرد ترین موارد استفاده داده کاوی در اطلاعات مالی می باشد. ورشکستگی شرکت باعث خسارات اقتصادی به مدیران، سرمایه گذاران، اعتبار دهندگان و کارکنان و دیگر افراد شده و نیز باعث هزینه های اجتماعی زیادی می شود. به این دلایل پیش بینی ورشکستگی یکی از مسائل مهم در مسائل مالی به شمار می آید. در پیش بینی ورشکستگی استفاده از اطلاعات صورتهای مالی تلاش های عمده از سال (۱۹۶۸) کارهای توماس آلتمن شروع شد. آلتمن پیشنهاد می کند که شکست یک شرکت یک فرآیند بلند مدت است و اینکه اطلاعات صورتهای مالی باید سیگنالهای خطر را برای ورشکستگی قریب الوقوع صادر کند. با کاربرد روش های تحلیلی چند متغیره او یک مدلی را برای پیش بینی ورشکستگی مطرح کرد. بعد از او کار آلتمن را بسیاری از محققان دیگر با استفاده از روش های آماری توسعه دادند. در سال های اخیر محقق ها تلاش می کنند که مدل های ساختاری مبتنی بر داده کاوی برای این موضوع بسازند.

لین و مک کلین تلاش کردند که پیش بینی کنند ورشکستگی شرکت را با استفاده از چهار روش مختلف انجام دهند. دو تا از این روش ها آماری هستند (رگرسیون لجستیک و تحلیل مشخص) دو روش دیگر تکنیک های محاسباتی یاد گیری ماشینی و محاسباتی هستند (درخت تصمیم، شبکه های عصبی). مضافاً اینکه آنها یک الگوریتم دورگه پیشنهاد کردند. نمونه آنها شامل اطلاعات ۱۱۳۳ شرکت بود. ۶۹۰ شرکت غیر ورشکسته و ۴۸ شرکت ورشکسته که در آزمون مورد استفاده قرار گرفته بودند. هیچ تلاشی در جهت مقایسه شرکت های ورشکسته و غیر ورشکسته انجام نشده است. ۳۷ نسبت مالی اصلی از ترازنامه و سود و زیان به عنوان متغیر های ورودی انتخاب شد. دوروش انتخابی خاص برای کاهش متغیر های ورودی بکار برده شدند. با ۴ مورد قضاوت انسانی و ۱۵ مورد آنالیز واریانس، تدوین کنندگان گزارش نتایج بهتری برای شبکه های عصبی و مدل های درخت تصمیم برای هر روی قضاوت های انسانی انتخاب و استفاده می شد. در نهایت محققان پیشنهاد میکنند یک الگوریتم دورگه بهترین عملکرد گزارش شده برای مدل دورگه می باشد. سود مندی اصلی شبکه عصبی فازی به توان مدل برای حل یک مشکل با استفاده از درک ساده ای از مدل زبان شناسی به جای ارائه محاسبات پیچیده ریاضی اشاره دارد.

این مدل برای پیش بینی ورشکستگی بانک مورد استفاده قرار گرفت. متغیر های ورودی شامل ۹ متغیر مالی بوده که از مطالعات قبلی و مشابه به طور خاص استخراج شده است. نمونه شامل اطلاعات درباره ۲۵۵۵ بانک غیر ورشکسته و ۵۴۸ بانک ورشکسته می باشد. ۲۰٪ از اطلاعات به عنوان آزمون های ثابت و ۸۰٪ به عنوان متغیر مورد استفاده بودند. برای کاهش خطای نوع اول نمونه تراز شده شامل تعداد برابر ورشکسته و غیر ورشکسته بود.

شین ولی (۲۰۰۲) مدلی را مبتنی بر الگوریتم ژنتیک پیشنهاد کردند. آنها اعتقاد داشتند که واقعیاتی که در شبکه های عصبی و الگوریتم ژنتیک پیشنهاد می شد، می تواند قابلیت درک قواعد را داشته باشد. الگوریتم ژنتیک برای یافتن حدی برای یک یا بیشتر از متغیر ها بالاتر یا پایین تر که یک شرکت در معرض خطر می باشد به کار می رود. مدل از یک قاعده ساختاری استفاده می کند که شامل ۵ شرط است که هر کدام بر یک متغیر خروجی اشاره دارد.

کیم وهان (۲۰۰۳) یک مدل کیفی براساس دانش حل مسائل ایجاد کردند. متخصصان با دانش موضوعی شان حقایق کمی و کیفی را بررسی می کردند. مدل از یک روش الگوریتم ژنتیک برای استخراج قوانین تصمیم گیری از پیش بینی های کیفی استفاده می کرد. مدل از روش متخصصان بانک تجاری کره تبعیت می کرد. به منظور پیش بینی، متخصصان ۶ عامل ریسک اصلی را بررسی کردند. در مدل یک کروموزوم شامل ۶ بخش ارائه دهنده ی یک طبقه از یک شرکت براساس ۶ عامل ریسک بود. یک بخش هفتم در کروموزوم شرکت را به عنوان ورشکستگی و غیر ورشکستگی طبقه بندی کرد. داده های نمونه شامل ۷۷۲ شرکت بود که نیمی از آنها ورشکسته بودند. متخصصان ۶ عامل ریسک برای شرکت هایشان را بررسی کردند. فرآیند تکامل ژنتیک ۱۱ قانون ورشکستگی را استخراج کرد. قوانین دیگر پس از رواج یافتن شبکه های عصبی و یادگیری استخراج شدند. قوانین استخراج شده با الگوریتم ژنتیک گزارش شدند.

دیمیتراس و همکاران (۱۹۹۸)، مجموعه های ساختار برای هدف پیش بینی ورشکستگی اعمال کردند. مجموعه آموزش شامل داده هایی برای ۴۰ شرکت ورشکسته و ۴۰ مورد غیر ورشکسته بود. مجموعه آزمایش شامل ۱۹ شرکت ورشکسته و ۱۹ غیر ورشکسته بود. یک مدیر معتبر از بانک گریک ۱۲ نسبت مالی برای وارد کردن جدول اطلاعات انتخاب کرد که هر یک شامل ۵ تا ۷ ویژگی بود. مدیر بانک یک نتیجه را انتخاب کرد و ویژگی هایی را که حذف نشده بودند را نگه داشت. در نهایت قوانین تصمیم مشتق شدند. نتایج روش با نتایج تحلیل مقایسه شدند. مک کی (۲۰۰۳) نتایج به دست آمده را با استفاده از مجموعه های ساختار نظر بازرس واقعی برای هدف پیش بینی مقایسه کرد. داده های نمونه شامل ۱۴۶ ورشکسته و ۱۴۵ غیر ورشکسته شرکت (US) مقایسه شده بودند. ۱۱ عامل پیش بینی کننده انتخاب

شدند. ۱۰ مورد نسبت های مالی بودند و ۱ مورد نظر بررسی قبلی بود. مجموعه ۸۷ نتیجه ای ایجاد می کرد که هر کدام ۴ تا ۶ متغیر را به کار می گرفتند و ۲ نتیجه انتخاب شدند. ۲ مدل قواعد تصمیم توسعه یافتند. نتایج مدل ها با حسابرسان واقعی مقایسه شدند و تقریباً با هم مساوی یافت شدند. نویسنده نتیجه می گیرد که مدل های توسعه یافته در این تحقیق هیچ مزیت پیش بینی قابل توجیه در روش شناسی کنونی حسابرسان ندارند.

بینون و پیل (۲۰۰۱) یک مدل توسعه یافته مجموعه های سخت رابکار گرفتند (VPRST) به قوانین تصمیم قابل پیش بینی کمک کرد و به طبقه بندی بخش با معرفی درجه ی اطمینان در طبقه بندی اجازه داد. نویسنده روش (FVSINTER) را برای هدف بکار گرفت. داده های نمونه شامل ۴۵ شرکت ورشکسته و ۴۵ مورد غیر ورشکسته در انگلستان بود. ۱۲ متغیر، ۸ متغیر مالی و ۴ متغیر کیفی برای ایجاد قوانین انتخاب شدند. نتایج (VPRST) با نتایج تجزیه و تحلیل چندگانه مقایسه شده و مطابقت داشتند.

پارک و هان (۲۰۰۲) در یک مطالعه ی استدلال مبتنی بر موضوع یک مدل برای پیش بینی ورشکستگی بانک ها از معیار فاصله از خصوصیات وزنی استفاده کردند. وزن ها با استفاده از روش فرآیند تحلیل سلسله مراتبی محاسبه شدند. نمونه شامل ۱۰۷۲ شرکت ورشکسته و ۱۰۷۲ شرکت غیر ورشکسته بود. ۱۳ متغیر مالی و ۱۵ متغیر غیر مالی برای ورودی انتخاب شدند. نویسنده بیان کرد که استدلال مبتنی بر موضوع / تحلیل سلسله مراتبی بهتر از استدلال مبتنی بر موضوع خاص عمل می کند.

بسوی تمرکز وعدم تمرکز مالی

مدیر باید توانایی مشتریانش را از نظر اعتباری برای حداقل یک سال با استفاده از ترازنامه بسنجد اگر شاخص هایی وجود دارند که مشتری شرکت با مشکلات مالی روبرو خواهد شد، که ممکن است منجر به ورشکستگی شوند، باید گزارش شود. بررسی وضعیت کنونی کار آسانی نیست. مطالعات نشان می دهند که تنها یک نسبت کوچک مبنی از شرکت های ورشکسته در یک مبنای مد نظر واجد شرایط بوده اند.

کو (۲۰۰۴)، شبکه های عصبی، درخت های تصمیم و روش های رگرسیون لجستیک را در مطالعه پیش بینی مقایسه کرد. داده های نمونه شامل ۱۶۵ شرکت بود. ۶ نسبت مالی انتخاب شده به عنوان متغیر ورودی انتخاب شدند. نویسنده بیان کرد که درخت های تصمیم متفاوت از دو روش دیگر عمل می کنند. تن و دیماردجو (۲۰۰۱) روی مطالعه قبلی تن کار کردند، که سعی کرد وضعیت مالی را برای استرالیا با استفاده از شبکه های عصبی پیش بینی کنند. در مطالعه قبلی اش تن، از داده های مالی استفاده کرد و سعی کرد بر مبنای وضعیت کنونی پیش بینی کند. موقعی که مدل پیش بینی میکند این روش بهتر از مورد قبلی از نظر نوع خطا عمل می کند. ۱۳ نسبت مالی به عنوان متغیر ورودی انتخاب شد و یک نمونه از ۲۱۴۴ مشاهده استفاده شد. نتایج با موارد مدل پروبیت مقایسه شدند و به خصوص برای نوع خطای ۱ بهتر یافت شدند. کونو کوبایاش (۲۰۰۰) یک روش برای رتبه بندی شرکت با استفاده از روش های برنامه ریزی ریاضی پیشنهاد دادند. روش هیچ فرض پشتیبان در مورد داده ها نداشت. سه گزینه بکار گرفته شد. ۶ نسبت مالی از بیانیه های مالی به عنوان متغیر ورودی استفاده شدند. داده های نمونه شامل ۴۵۵ شرکت بود. این روش یک نمره برای هر شرکت محاسبه کرد.

تقلب در مدیریت

تقلب مدیریت قصد جعل اسناد مالی توسط مدیریت است. تقلب مدیریت به مقامات مالیاتی، سهامداران و اظهارات مالی ضربه می زند. اسپایتس (۲۰۰۲) دو مدل را برای شناسایی بیانیه هایی مالی اشتباه از داده های موجود پیشنهاد می کند. متغیرهای ورودی برای مدل نخست شامل ۹ نسبت مالی است. برای مدل دوم نمره و امتیاز ۲ به عنوان متغیر ورودی به رابطه ی بین توزیع مالی و بیانیه های مالی اضافه شده است. روش

استفاده شده رگرسیون لجستیک است و داده های نمونه شامل ۳۸ شرکت ورشکسته و ۳۸ شرکت غیر ورشکسته است. برای هر دومدل نتایج نشان می دهد که ۳متغیر با ضرایب معنی دار در مدل وارد شده است.

پیش بینی عملکرد شرکت

لام (۲۰۰۳) یک مدل برای پیش بینی نرخ بازگشت در دارایی های سهامداران معرفی نمود. اواز شبکه های عصبی وقوانین از وزن های ارتباطات با بکاربردن الگوریتم (GLARE) استفاده کرد. بردار ورودی شامل ۱۵نسبت مالی و ۱متغیر تحلیل فنی بود. دریک آزمایش دیگر ۱۱متغیر اقتصادی کلان نیز جای گرفتند. داده های نمونه شامل ۳۶۴ شرکت بود. پک وهمکاران (۲۰۰۱) دومدل برای طبقه بندی شرکت ها براساس عملکرد شان توسعه داد. هر دو روش از (SOMS) استفاده کردند. مدل نخست از داده های مالی ۱۶۰شرکت عمل کرد. مدل دومی گزارش سالانه ی مدیر عامل، شرکت ها را تحلیل کرد. نویسندگان نتیجه گرفتند که تفاوت هایی بین نتایج خوشه بندی ۲روش وجود دارند. کلویچنکو وهمکاران (۲۰۰۴) روش ذکر شده قبلی را ایجاد کردند. دو مدل توسعه یافتند. یکی نسبت های مالی را تحلیل می کرد و دیگری گزارش های مدیر عامل شرکت ها را تحلیل می کرد. در این مطالعه، یک روش متفاوت خوشه بندی متن، برای تحلیل گزارشات استفاده شد. با مقایسه نتایج روش های کیفی و کمی نویسندگان نتیجه گرفتند که گزارشات متن تمایل به تغییراتی در وضعیت مالی دارند قبل از این که این تغییرات بر نسبت های مالی تاثیر بگذارند.

تخمین ریسک

وظیفه ی تحلیل ریسک تقاضای بیشتری بخاطر تعداد افزایش یافته ی ورشکستگی بانک ها و پیشنهادات رقابتهی می طلبد. روش های مدیریت داده ها برای تخمین ریسک بکار گرفته اند. هوآنگ وهمکاران (۲۰۰۳) تحلیل رتبه بندی را با استفاده از (پشتیبانی ماشین های بردار) انجام دادند، که یک روش یادگیری است. دو مجموعه داده استفاده شدند. یکی شامل ۷۴ شرکت کره ای و دیگری شامل ۲۶۵ شرکت آمریکایی بود. برای هر دو مجموعه ی داده ۵طبقه ی رتبه بندی تعریف شد. ۲مدل برای داده های کره ای و ۲مدل برای مجموعه داده های آمریکایی استفاده شد. هر کدام یک بردار ورودی متفاوت داشتند. پشتیبانی ماشین های بردار و شبکه های عصبی برای پیش بینی رتبه بندی استفاده شدند. پشتیبانی ماشین های بردار از ۳ مورد از ۴ مدل بهتر عمل کرد. موز وهمکاران (۲۰۰۴) از دیاگرام تصمیم برای قابل رویت ساختن قوانین بررسی ریسک استفاده کردند.

دیاگرام های تصمیم مزیت های تئوریدر درخت های تصمیم دارند که آنها را برتر می کنند. دو مجموعه داده، یکی شامل داده های آلمان، و دیگری شامل داده های بنلوکس استفاده شدند. یک شبکه عصبی برای طبقه بندی اعمال شد. روش های استخراج قوانین و ترپان برای استخراج قوانین از شبکه اعمال شد. بعلاوه قوانین C4.5 و C4.5 و روش گراف های تصمیم برای ایجاد درخت های تصمیم وقوانین آن اعمال شدند. عملکرد قانون نرو و ترپان با عملکرد شبکه های عصبی قابل مقایسه بود. در نهایت قوانین به شکل دیاگرام های تصمیم قابل رویت شدند.

ارزیابی و جهت دهی تحقیقات آینده در حوزه داده کاوی

امور مالی وحسابداری رشته های کاربردی معمول برای مدیریت داده ها هستند. توانایی های پیش بینی و طبقه بندی روش های مدیریت داده ها آنها را قادر می سازد تا برای بیان ورشکستگی های بانک ها استفاده شوند. وضعیت کنونی و پیش بینی مالی، شناسایی تقلب مدیریت، تخمین ریسک و پیش بینی عملکرد شرکت است. حسابرسان، تحلیلگران وسرمایه گذاران می توانند در کارشان تسهیل شوند وزمان

وهزینه رادر فرآیند تصمیم گیری شان به دست آورند .پیش بینی ورشکستگی به نظر می رسد نظر اکثر محققان را جلب نماید زیرا بیشتر مقالات به این موضوع برمیگردد. حوزه های کاربرد ادبیات موضوع بررسی شده در جدول ۱ نشان داده شده است.بررسی ادبیات جمع آوری شده ایجاد بحث هایی را به منظور روش های به کارگرفته شده وهمچنین داده های استفاده شده وموضوعات ماتریس عملکرد رانشان می دهد.

جدول شماره ۱:حوزه های کاربرد ادبیات موضوع داده کاوی

تعداد کاربرد	زمینه های کاربرد
۱	ورشکستگی
۳	به سوی تمرکز وعدم تمرکز
۳	پیش بینی خوشه وعملکرد شرکت
۲	تخمین ریسک
۱	تقلب در مدیریت

روش ها ومدل ها

واژه روش های مدیریت داده ها شامل یک طیف وسیع از روش های نشأت گرفته از روش های آماری، هوش مصنوعی و پایگاه های اطلاعاتی است که در ادبیات تحقیق شبکه های عصبی پر کاربرد ترین مدل ها هستند. جدول ۲ را نشان می دهد .

جدول شماره ۲:مدل های به کاررفته در تحقیقات که در حوزه شبکه های عصبی قرار دارند

تعداد	مدل
۱	شبکه های عصبی
۳	مجموعه ای سخت
۲	درخت های تصمیم گیری
۲	الگوریتم های ژنتیکی
۲	پیوندی
۱	استدلال مبتنی بر موضوع
۱	برنامه نویسی ریاضی
۱	رگرسیون لجستیک
۱	پشتیبانی ماشین های بردار

اگر چه بسیاری از تحقیقات بر این حقیقت تأکید می کنند که مدل هایی که ویژگی ها ومزیت های مدل های ویژه را ترکیب می کنند، ممکن است عملکرد را بهبود دهند اما مدل های ترکیبی تنها در ۲ مورد استفاده شده اند. گرایش تحقیقات آتی می تواند برای توسعه مدلهای موجود با الگوریتم های توسعه یافته باشد. متغیرمجموعه های سخت، فرآیند تحلیلسلسله مراتبی استدلال مبتنی بر موضوع والگوریتم ژنتیک که از روش نیکینگ استفاده

میکنند نمونه هایی از این مواردن. طراحی معماری شبکه عصبی هنوز یک مبحث هنری است. روش هایی که یک معماری شبکه عصبی بهینه را برای موارد خاص بیانمی کنند می توانند توسعه یابند. اگرچه انتقاد اصلی از شبکه های عصبی این استکه تنها به عنوان با کس های سیاه عمل می کنند. اما رویکرد تحقیقات آتی می تواند به سوی تغییر الگوی تصمیم گیری شبکه های عصبی جهت یابند. تمام روش ها مزیت های ثنوری دارند و تصوراتی در متغیرهای ورودی تحمیل نمیکنند. به هرحال نتایج گزارش شده از روش های مدیریت هوش مصنوعی تنها نتایج روش های آماری اند .

داده ها

داده های مورد استفاده در ادبیات اساساً نسبت های مالی مشتق شده از بیانیه های مالی اند. در ۸ مورد بردار ورودی شامل نسبت های مالی است. تنها در یک مورد ، نسبت های مالی در بردار ورودی استفاده نشده اند. بسیاری از نویسندگان به نیاز برای تقویت بردار ورودی با اطلاعات بیشتر اشاره دارند. متغیر های اقتصاد کلان نیز می توانند شامل شوند. اطلاعات کیفی از قبیل دستیابی به اهداف استراتژیک شرکت، نظر بررسی کننده قبلی، تجربه شرکت ، اطلاعات بازار وبسیاری از موارد دیگر می توانند برای فاکتورهای فنی واجتماعی ، سیاسی واقتصادی استفاده شوند. در ۲ مقاله فنون داده کاوی برای طبقه بندی وپیش بینی عملکرد شرکت های استفاده شده است که در برخی از مقالات بررسی شده، اندازه نمونه از نظر آماری بزرگ نیست. نمونه های کوچک ممکن است نتایج ضعیفی ارائه دهند. بعلاوه تفاوت های مهمی در اندازه نمونه های وجود دارد. جدول ۳ ، اندازه نمونه ها در این تحقیقات را نشان می دهد.

جدول شماره ۳: اندازه نمونه ها در تحقیقات شبکه های عصبی

مقاله ها	سایز نمونه
۴	> ۱۰۰۰
۲	> ۵۰۰
۵	> ۲۰۰
۵	<= ۲۰۰

داده های مالی در بسیاری از موارد حاوی یک تعداد مهمی از نسبت های مالی اند. بسیاری از این نسبت ها شامل اطلاعات اند. بعلاوه تحقیقات نشان داده است که یک تعداد نسبتا کوچک از نسبت ها برای اهداف پیش بینی و طبقه بندی کافی است. به این دلایل انتخاب ویژگی مورد نیاز است. در ۷ مورد، محققان بر مطالعات قبلی برای انتخاب متغیرهای ورودی تکیه می کنند. در ۴ مورد انتخاب بر اساس قضاوت افراد است. معرفی روش های رسمی از قبیل (ANOVA) ممکن است انجام انتخاب ویژگی را بهبود دهد. وجود مقادیر پرت در داده های مالی معمول است. استراتژی هایی برای مدیریت داده های مالی پرت از قبیل استفاده از مقادیر میانگین داده شده یا استفاده از مقادیر احتمالی ممکن است ارزیابی و پیشنهاد شوند. توزیع داده ها بحث دیگری برای توجه است.

ماتریس های عملکرد

بحث مهم دیگر ماتریس های عملکرد است، عملکرد با آزمون مدل ها بر خلاف یک تست و احتمال یک اعتبار نمونه ارزیابی شده است. در برخی موارد هیچ اعتبار نمونه ای وجود ندارد و آزمون الگو برای اندازه گیری عملکرد مدل استفاده شده است. برخی الگوریتم ها از آزمون مدل برای آموزش استفاده می کنند. دقت اساسی به عنوان

یک ملاحظه ای دیگر با توجه به خطایی نوع ۱ و نوع ۲ وجود دارد. یک خطای نوع ۱ اتفاق می افتد موقعی که مدل هیچ ورشکستگی برای شرکت پیش بینی نمی کند و شرکت واقعا ورشکسته می شود. یک خطای نوع ۲ زمانی اتفاق می افتد که بررسی های اضافی وجود داشته باشد، بنابراین خطای نوع ۱ هزینه بیشتری نسبت به خطای نوع ۲ دارد. هزینه های نسبی خطاهای نوع ۱ و ۲ باید در ماتریس عملکرد مورد توجه قرار گیرند.

نتیجه گیری

رشد روزافزون اهمیت مسائل مالی و گسترش جوامع همراه با پیچیدگی محیط های اقتصادی این را می طلبد که حسابداری و امور مالی از سایر علوم برای حل مسائل و مشکلات خود کمک بگیرد. فنون مدیریت داده های یکی از روشهای نوظهور در این حوزه بوده که روز به روز بر کاربرد و اهمیت آن در حل مشکلات مالی افزوده می شود. این فنون دارای انواع مختلفی می باشند و قابلیت های پیش بینی دارند و می توانند فرایند تصمیم گیری را در مسایل مالی تسهیل کنند. وظایف مالی و غیر مالی در ادبیات مربوط، به موضوعات پیش بینی ورشکستگی، تخمین ریسک نقدینگی، گزارشات وضعیت کنونی توجه دارند. پیش بینی ورشکستگی به نظر می رسد، مهم ترین حوزه مورد بررسی باشد.

روش های داده کاوی به کار رفته در ادبیات مربوط شامل شبکه های عصبی، الگوریتم ژنتیک، درخت های تصمیم و برنامه ریزی ریاضی اند. اگر چه مقدار قابل توجهی از تلاش های کاربرد فنون مدیریت داده هارا در مورد امور مالی مورد توجه قرار می دهد. معرفی مدل های هیبریدی، بهبود مدل های موجود، استخراج قوانین جامعاز شبکه های عصبی، بهبود عملکرد و یکپارچه سازی سیستم های مدل رابطه با ابزار مدیریت داده ها برخی از جهت گیری های تحقیقات آینده اند.

تلاش های تحقیقات آینده با بکارگیری مدل های روش های مدیریت داده ها برای امری را در حسابداری و امور مالی بهبود خواهد داد.

منابع:

1. B Back ,J,Toivonen , H.vanhatanta and A,visa : "Comparing Numerical Data and Text information from Annual Reports using self-organizing maps". international Journal of Accounting information systems, volume 2, Issue4, December ,2001. pp,249-269
2. M.J.Beynon And M.J .peel: "Variable Precision rough Set Theory and Data Discretisation : an Application to Corporate Failure prediction ". Omega The International Journal of Management science . Volume 29, Issue6, December ,2001. pp.561-576.
3. T.G. Calderon and J,J.Cheh : "A Roadmap For Future neural networks Research in Auditing and Risk Assessment". International Journal Of Accounting Information Systems . Volume 3, Issue 4 , December ,2002 ,pp.203-236
4. A,I. Dimitras ,R, Slowinski , R Susmaga and . zopounidis : "Business Failure Prediction using Rough Sets", European Journal of Operational Research , Volume 114, Issue 2 April ,1998, pp263-280
5. I,A,m, Araser , D,J, Hatherly .K,Z,Lin: "An empirical investigation of the use of analytical review by external auditors", the british accounting review volume 29, issue 1, march ,1997, pp.35-47
6. z.huang , h.chen, c.j.hsu, w.h.chen and s. wu: "credit rating analysis with support vector machines and neural networks: a market comparative study", decision support systems. in press, 2003
7. m.j.kim and i.han: 'the discover of expert 'decision rules from qualitative bankruptcy data using genetic algorithms ", expert systems with applications, volume 15, issue 4, November .2003, pp.637-646
8. A.kloptchenko ,T.eklund , j.karisson , b.back, h.vanharanta and a.visa : "combinig data and text minig techniques for analyzing financial reports". intelligent systems in accounting , finance and management , volume 12, issue 1 , January/march 2004, pp.29-41..
9. H.C.koh, "going concern prediction using data mining techniques", managerial auditing journal , volume 19, no 3, 2004, pp,462-476

10. H.konno and H. kobayashi."failure discrimination and rating of enterprises by semi-define programming",asia-pacific financial markets volume 7,issue 3,september , 2000,pp.261-273
11. e.koskivaara "artificial neural networks in analytical review procedures".managerial auditing journal, volume 19,no 2,2004,pp.191-223
12. M.lam:"neural network techniques for financial performance prediction:integrating fundamental and technical analysis :,decision supportsystems ,in press 2003
13. F.Y.lin and s .mc Clean:'a data minig approach to the orediction of corporate failure",knowledge –based systems,volume 14 , issues 3-4,june 2001 , pp.189-195.
14. T.mckee,"rough sets bankruptcyprediction models vs.auditor signaling rates ",journal of forecasting. Volume 22.issue 8,December ,2003,pp.569-586
15. C.mues , B.baesens,C.M files and j.vanthenen:"decision diagramsin machine learning :an empirical study on real-life credit-risk data ",expert systems with applications , in press ,2004
16. C.S.park and I:"A case reasoning with the feature weights derive by analytic hierarchy process for bankruptcy prediction",expert systems with application, volum 23,issue 3, October , 2002,pp.225-264
17. K.S.shin and Y.J.lee:"lee : "a genetic algorithm application in bankruptcy prediction modeling ",expert systems with applications,volume 23.issue 3,October,2002,pp.321-328
18. C,spathis:"detecting false finance statement using published data:some evidence from breece ". managerial auditing journal ,volume 17, no 4, 2002,pp.179-191
19. C.N.W tanand H.dihardjo:"A study on using artificial neural networks to develop an early warning predictor for credit union financial distress with comparison to the probit model",managerial finance,volume 27,no4,2204,pp.56-77
20. W.L.tung ,C.quek and p.cheng : "gensoews:a novel neural fuzzy based early warning systems for predicting bank failures",neural networks, volume 17,issue 4,may ,2004 .pp.567-587

dataacademy.ir