

## بررسی جایگاه داده‌کاوی در امور مالی و حسابداری

### شبلم طاوری

کارشناسی ارشد هوش مصنوعی و ریاتیک-دیوان محاسبات فارس

#### چکیده

در طول سالیان اخیر، با پیشرفت فناوری اطلاعات، حجم داده‌ها در پایگاه داده‌ها هر دو سال یک بار، دو برابر شد و همچنان با سرعت بیش‌تری نسبت به گذشته حجم اطلاعات ذخیره شده بیش‌تر می‌شود. با وجود شبکه جهانی وب، سیستم‌های یکپارچه اطلاعاتی، سیستم‌های یکپارچه بانکی، تجارت الکترونیکی و ... لحظه به لحظه به حجم داده‌ها در پایگاه داده‌ها اضافه شده و باعث به وجود آمدن انبارهای (توده‌های) عظیمی از داده‌ها شده است، به طوری که ضرورت کشف و استخراج سریع و دقیق دانش از این پایگاه داده‌ها را بیش‌تر از پیش نمایان کرده است. در حال حاضر، داده‌کاوی مهم‌ترین فناوری برای بهره‌وری موثر، صحیح و سریع از داده‌های حجیم است و اهمیت آن رو به فزونی است. داده‌کاوی پل ارتباطی میان علم آمار، علم کامپیوتر، هوش مصنوعی، الگوشناسی، فراگیری ماشین داده‌ها می‌باشد. داده‌کاوی فرآیندی پیچیده جهت شناسایی الگوها و مدل‌های صحیح، جدید و به صورت بالقوه مفید، در حجم وسیعی از داده‌ها می‌باشد، به طریقی که این الگوها و مدل‌ها برای انسان‌ها قابل درک باشند. از مشهورترین کاربردهای داده‌کاوی هوشمندی کسب و کار، است. قابلیت‌های ابزاری داده‌کاوی منجر به استفاده از این فناوری در امور مالی و حسابداری شده است این ابزارها به منظور پیش‌بینی ورشکستگی، وضعیت تداوم فعالیت، پیش‌بینی درماندگی مالی، تشخیص و شناسایی تقلب مدیریت، تخمین ریسک اعتباری و عملکرد واحد تجاری شده و بدین ترتیب داده‌کاوی را تبدیل به موضوعی با اهمیت زیاد، در امور مالی و حسابداری نموده است. هدف این مقاله بررسی اجمالی این فناوری جدید و تحقیقات مرتبط با آن، به منظور نمایان ساختن کارکردهای داده‌کاوی در امور مالی و فرصت‌های تحقیقاتی آن می‌باشد.

**واژه‌های کلیدی:** حسابداری، امور مالی، روش‌های داده‌کاوی، داده‌کاوی.

## مقدمه

داده کاوی فرآیندی است برای استخراج الگوهایی که دانش را بازنمایی می کند که این دانش به صورت ضمنی در پایگاه داده های عظیم، انباره داده و دیگر مخازن بزرگ اطلاعات، ذخیره شده است. داده کاوی بطور همزمان از چندین رشته علمی بهره می برد نظیر: تکنولوژی پایگاه داده، هوش مصنوعی، یادگیری ماشین، شبکه های عصبی، آمار، شناسایی الگو، سیستم های مبتنی بر دانش، حصول دانش، بازیابی اطلاعات.

با توجه به وجود اطلاعات ارزشمند در پایگاه های داده ای در اواخر دهه ۸۰ میلادی، تلاش برای استخراج و استفاده از اطلاعات پایگاه های داده ای شروع شد. داده کاوی فرآیندی است که در آغاز دهه ۹۰ پا به عرصه ظهور گذاشته و با نگرشی نو، به مسئله استخراج اطلاعات از پایگاه داده ها می پردازد. در سال ۱۹۸۹ و ۱۹۹۱ کارگاه های کشف دانش از پایگاه داده ها توسط پیاتتسکی و همکارانش و در فاصله سال های ۱۹۹۱ تا ۱۹۹۴ کارگاه های فوق، توسط فایاد و پیا تتسکی و دیگران برگزار شد. به طور رسمی اصطلاح داده کاوی برای اولین بار توسط «فیاض» در اولین کنفرانس بین المللی «کشف دانش و داده کاوی» در سال ۱۹۹۵ مطرح شد. از سال ۱۹۹۵ داده کاوی به صورت جدی وارد مباحث آمار شد؛ و در سال، اولین شماره مجله کشف دانش از پایگاه داده ها منتشر شد (پورحیدری و اعظمی، ۱۳۸۹).

وظیفه ی داده کاوی، کاویدن و استخراج از منابع عظیم داده و بانک های اطلاعاتی است تا اطلاعات گران بهایی که در حجم انبوهی از اطلاعات سطحی پنهان شده است را استخراج کند. برخی از صاحب نظران براساس مفاهیم داده کاوی، معتقدند این فناوری تنها مرحله ای از فرآیندی فراگیرتر به نام کشف دانش در پایگاه داده می باشد. مراحل دیگر در این فرآیند فراگیر شامل پاسازی، یکپارچه سازی، انتخاب داده، تبدیل و ارزیابی الگو است. فناوری داده کاوی میان بسیاری از حوزه های فنی شامل پایگاه داده، آمار، یادگیری ماشین و رابط انسان و کامپیوتر ارتباط برقرار می سازد (ستایش و جمالیانپور، ۱۳۸۸).

اطلاعات مالی توسط سازمان هایی شبیه بانک ها، تحلیل گران 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. روش های داده کاوی در داده های مالی، می تواند در حل مشکلات طبقه بندی و پیش بینی و تسهیل فرآیند تصمیم گیری به کار رود. نمونه هایی از مسائل طبقه بندی مالی شامل ورشکستگی شرکت ها، تخمین ریسک اعتباری، گزارش تداوم، فعالیت، درماندگی مالی و پیش بینی عملکرد واحد تجاری می باشد بدین ترتیب دامنه اهمیت داده کاوی در امور مالی و حسابداری می تواند طیفی گسترده از کشف تقلب تا افزایش سودآوری واحد تجاری باشد (سجادی و همکاران، ۱۳۸۶).

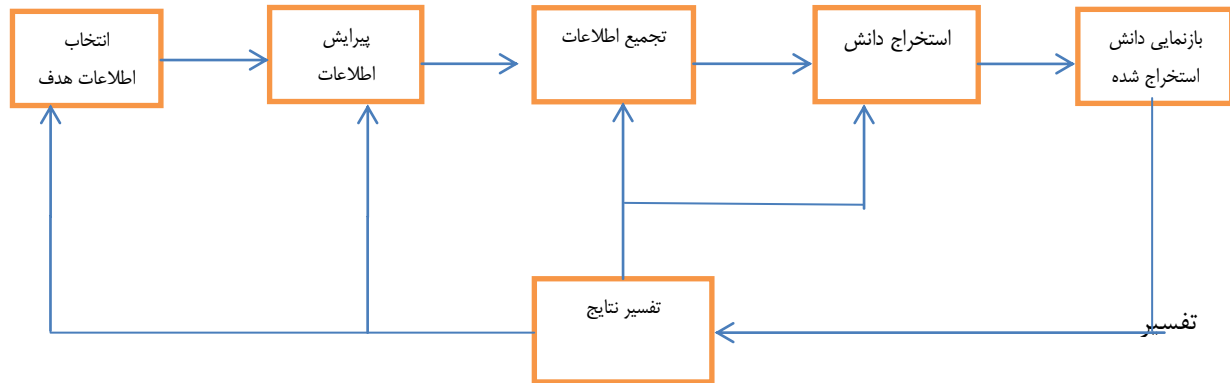
اهمیت داده کاوی در امور مالی و حسابداری در بسیاری از سازمان ها مشاهده شده است. جامعه حسابداران رسمی آمریکا داده کاوی را به عنوان یکی از ده روش برتر برای آینده تعریف کرده است؛ و انجمن حسابرسان داخلی آمریکا داده کاوی را به عنوان یکی از چهار تحقیق برتر این رشته معرفی کرده است (منهاج، ۱۳۸۷).

مقاله حاضر به بررسی رویکرد داده کاوی در امور مالی در ۴ بخش می پردازد: در بخش ۱ به مفاهیم داده کاوی می پردازیم، در بخش ۲ مروری بر تکنیک های رایج در داده کاوی خواهیم داشت، در بخش ۳ حوزه کاربردی داده کاوی در امور مالی را بررسی نموده در بخش پایانی نتیجه گیری را ارائه می دهیم.

## ۱- مفاهیم داده کاوی:

الف) مراحل داده کاوی

مراحل داده کاوی را می توان بصورت نمودار زیر نمایش داد:



### نمودار ۱: مراحل داده کاوی

- ۱- انتخاب اطلاعات: در این مرحله باید داده‌ها بر مبنای معیارهای مشخص انتخاب گردند.
- ۲- پیرایش داده‌ها: داده‌ها به قالبی قابل استفاده برای داده‌کاوی در می‌آیند. از عملی که در این مرحله صورت می‌گیرد، میتوان به خلاصه سازی و یا محاسبه مقادیر تجمعی اشاره کرد.
- ۳- تجمیع اطلاعات: بخش اصلی فرایند که در آن با استفاده از روش‌ها و تکنیک‌های خاص، استخراج الگوهای دانش صورت می‌گیرد.
- ۴- استخراج دانش: تشخیص الگوهای صحیح مورد نظر، از سایر الگوها در این مرحله انجام می‌شود. صحت الگوها بر اساس یک سری معیارهای جذابیت سنجیده می‌شود.
- ۵- بازنمایی دانش: در این بخش به منظور ارائه دانش استخراج شده به کاربر، از یک سری ابزارهای بصری سازی استفاده می‌گردد.
- ۶- تفسیر نتایج: نتیجه این بخش، گزارش نهایی برای کاربر است. هدف اصلی داده‌کاوی استفاده از نتایج کشف شده برای تصمیم‌گیری، سیاست‌گذاری و پیش‌بینی به منظور ایجاد یک موقعیت بهتر و جدید می‌باشد (تونگ و همکاران<sup>۱</sup>، ۲۰۱۰).

#### ب) ابزارهای داده‌کاوی

با توجه به اهمیت داده‌کاوی در دنیای کسب و کارهای بزرگ و حرفه‌ای، ابزارها و نرم‌افزارهای متعددی برای این منظور طراحی و توسعه داده شده‌اند. برخی از این ابزارها، رایگان و متن‌باز<sup>۲</sup> هستند و برخی دیگر به صورت بسته‌های نرم‌افزاری تجاری ارائه شده‌اند. از این میان، تعدادی از مواردی که بیشتر مورد استفاده هستند، در ادامه آمده است.

<sup>۱</sup> Tung et al.

<sup>۲</sup> Open Source

- 1- Data to knowledge
- 2- sas
- 3- clementine
- 4- Intelligent-miner
- 5- Insightful miner
- 6-weka

## ۲- تکنیک‌های رایج در داده کاوی:

واژه داده کاوی بر پایه تعداد زیادی الگوریتم‌ها قرار گرفته است. مدل‌ها و روش‌های گرفته شده از نفوذ آمار، ماشین‌های یادگیری، پایگاه داده‌ها و شبکه‌های عصبی و... برخی از این روش‌ها برای آزمایش و آزمون اطلاعات مالی به کار می‌روند. عمده روش‌های داده کاوی که در این مقاله مورد بررسی قرار خواهند گرفت عبارتند از: شبکه‌های عصبی، الگوریتم‌های ژنتیک، درخت تصمیم، استدلال مبتنی بر موضوع، نظریه مجموعه اولیه.

### ۲-۱ شبکه‌های عصبی

شبکه عصبی مصنوعی یک مدل ریاضی گسترش یافته از طبیعت عصبی انسان است که مشابه مغز انسان اطلاعات را پردازش می‌کند. یک مدل شبکه عصبی مصنوعی می‌تواند با داشتن چندین متغیر ورودی، چندین متغیر خروجی را پیش‌بینی کند و شامل تعداد زیادی عناصر پردازنده ساده به نام نرون است (جیمز و همکاران<sup>۱</sup>، ۲۰۰۳).

نرون‌ها با اتصالاتی که به آن‌ها وزن گفته می‌شود، به یکدیگر مرتبط شده‌اند هر نرون، یک تابع فعال‌سازی دارد که خروجی آن را تعیین می‌کند. برای تعیین وزن‌ها و آستانه نرون‌ها، بدین منظور که شبکه عصبی مصنوعی خروجی‌های نزدیک به مقادیر هدف را تخمین بزند، از آموزش استفاده می‌شود. داده‌های موجود به دو قسمت تقسیم می‌گردد که قسمتی برای آموزش و قسمتی برای ارزیابی مدل در نظر گرفته می‌شود. داده‌های آموزشی باید دارای الگوی کافی باشند، به طوری که شبکه بتواند رابطه بین ورودی و خروجی را به دقت یاد

بگیرد. یک شبکه عصبی مصنوعی که به خوبی آموزش دیده باشد، در پیش‌بینی بهتر از یک مدل ریاضی و یا برنامه شبیه‌سازی عمل می‌کند. شبکه پسخور چند سطحی نوعی از معماری شبکه عصبی مصنوعی است و کاربرد زیادی در پیش‌بینی دارد. یک شبکه عصبی مصنوعی چند لایه شامل یک لایه ورودی، یک لایه خروجی و یک یا چند لایه مخفی است (کوینلان<sup>۲</sup>، ۱۹۸۷).

یادگیری در شبکه عصبی مصنوعی چند لایه شامل وارد کردن بخشی از داده‌ها با خروجی مشخص است. سپس الگوریتم یادگیری وزن‌های سیستم را به گونه‌ای تعیین می‌کند که خطای بین خروجی واقعی و خروجی شبکه حداقل شود. فرایند یادگیری تا زمانی که خروجی شبکه وهدف یکی شود، ادامه می‌یابد. به اختلاف بین خروجی شبکه و هدف، خطا گفته می‌شود. این خطا می‌تواند از طریق تغییر و اصلاح وزن‌ها کاهش یابد. فرایند آموزش زمانی پایان می‌پذیرد که خطا کمتر از مقدار معین شود یا تعداد تکرار آموزش به میزان مشخصی برسد. روش پس انتشار خطا یکی از الگوریتم‌های یادگیری است که خطا را محاسبه کرده و وزن‌های لایه مخفی را بر اساس آن دست‌کاری می‌کند.

<sup>1</sup> James et al.

<sup>2</sup> Quinlan

## ۲-۲ الگوریتم ژنتیک

این الگوریتم بر پایه فرضیه تکامل داروین بوده و کاربرد آن بر ژنتیک طبیعی استوار است. اصول اولیه الگوریتم ژنتیک توسط هلند و همکاران در سال ۱۹۶۲ ارائه 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. 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؛ کروموزوم و به اعداد تشکیل دهنده آن نیز ژن گویند. الگوریتم ژنتیک با تعدادی از جواب‌های گوناگون آغاز می‌شود که جمعیت نسل اولیه را تشکیل 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 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 (کونو<sup>۱</sup>، ۲۰۱۳).

## ۲-۳ درخت تصمیم

درخت تصمیم در حقیقت یک گراف با ساختاردرخت است که هر رأس آن نشان دهنده یک آزمون یا مقایسه مقدار یک متغیر می‌باشد و یال‌هایی که از آن رأس خارج می‌شود، نشان دهنده تصمیمی است که در مقابل هر نتیجه بدست آمده از آزمون گرفته می‌شود. در این روش تلاش می‌شود تا مشاهدات به زیرگروه‌هایی تقسیم شوند. از جمله مزایای این روش مستقل بودن آن از چگونگی توزیع داده‌ها و وابستگی متغیرهای ورودی می‌باشد. در حقیقت، درخت تصمیم یک مدل مفهومی ساده را با استفاده از تعدادی تصمیم ساده ایجاد می‌کند. الگوریتم یادگیری در این روش بسیار سریع است (کسکیوآرا<sup>۲</sup>، ۲۰۰۴).

## ۲-۴ استدلال مبتنی بر موضوع

این روش یک روش حل مسئله با دلیل یابی از روش‌های مسائل حل شده مشابه برای حل یک مشکل است. این روش تلاش می‌کند برای بازیابی یک مورد مشابه در مورد مبنا، کلیه موضوعات در این روش نوعی مشابه سازی در اندازه‌گیری بوده و بازیافتن یک مورد مشابه می‌باشد. برای شبیه سازی بین دو مورد با استفاده از محاسبات و تفاوت هندسی بین آنها تلاش می‌شود. این روش فرض می‌کند که تمام ترکیب‌ها (خصوصیات) به طور مساوی مربوط هستند (لم<sup>۳</sup>، ۲۰۱۲).

## ۲-۵ نظریه مجموعه اولیه

تئوری مجموعه اولیه در سال ۱۹۹۱ بوسیله پالاک ارائه گردید نظریه مجموعه اولیه تعمیمی از نظریه دوتایی می‌باشد که از نظریه‌های معروف در ریاضیات بازه هست. (پچنیزکی و همکاران<sup>۴</sup>، ۲۰۰۶).

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 پرابهام 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.

<sup>1</sup> konno

<sup>2</sup> koskivaara

<sup>3</sup> lam

<sup>4</sup> Pechenizkiy et al.

**۳- حوزه کاربردی داده کاوی در امور مالی:**

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.

بر اساس استاندارد حسابرسی ۵۶ آمریکا، حسابرس نظر خود را در خصوص انتظارات مالکان ارائه می‌دهد و این انتظارات در مبالغ ثبت شده و یافته‌های کشف شده نشان داده می‌شوند. در انجام این وظیفه حسابرس بررسی‌های تحلیلی را به کار می‌برد که ارتباط بین مبالغ پیش‌بینی شده را با مشاهدات واقعی و اقلام اطلاعات در واقعیت را مقایسه می‌کند. بررسی‌های تحلیلی این اجازه را می‌دهد که دقت یک مانده حساب را بدون آزمون کردن معاملات اشخاص وابسته آزمون، روش مقداری ساده مثل بررسی روند یا نرخ آزمون منطقی و روش مقداری پیشرفته مثل تحلیل رگرسیون و شبکه‌های عصبی (جیمز و همکاران<sup>۱</sup>، ۲۰۰۳).

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

مقالات مرتبط با حوزه‌های کاربردی خاص در رشته حسابرسی اشاره دارد به پیش‌بینی ورشکستگی، پیش‌بینی تداوم فعالیت و پرباشانی مالی و تقلبات مدیریت.

**۳-۱ پیش‌بینی ورشکستگی**

این موضوع به نظر می‌رسد که یکی از پر کاربردترین موارد استفاده داده کاوی در اطلاعات مالی می‌باشد. ورشکستگی شرکت باعث خسارات اقتصادی به مدیران، سرمایه‌گذاران، اعتبار دهندگان و کارکنان و دیگر افراد شده و نیز باعث هزینه‌های اجتماعی زیادی می‌شود. به این دلایل پیش‌بینی ورشکستگی یکی از مسائل مهم در مسائل مالی به شمار می‌آید. در پیش‌بینی ورشکستگی استفاده از اطلاعات صورت‌های مالی تلاش‌های عمده از سال (۱۹۶۸) کارهای توماس آلتمن شروع شد. آلتمن پیشنهاد می‌کند که شرکت یک فرآیند بلند مدت است و این که اطلاعات صورتهای مالی باید سیگنالهای خطر را برای ورشکستگی قریب الوقوع صادر کند. با کاربرد روش‌های تحلیلی چند متغیره او یک مدلی را برای پیش‌بینی ورشکستگی مطرح کرد. بعد از او کار آلتمن را بسیاری از محققان دیگر با استفاده از روش‌های آماری توسعه دادند. در سال‌های اخیر محققان تلاش می‌کنند که مدل‌های ساختاری مبتنی بر داده کاوی برای این موضوع بسازند.

لین و مک لین (۲۰۰۱) تلاش کردند تا با استفاده از ۴ روش متفاوت، ورشکستگی شرکت‌ها مورد از این روش‌ها، (روش‌های آماری تحلیل تشخیصی و رگرسیون لجستیک) و دو مورد دیگر نیز از فنون یادگیری ماشین درختان تصمیم‌گیری و شبکه‌های عصبی بوده‌اند.

همچنین آنها براساس روش‌های پیش‌گفته، الگوریتمی ترکیبی 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 و ۲۸۹ شرکت ناموفق و ۴۸ شرکت موفق به عنوان مجموعه آزمایشی 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 انتخاب شد. دو روش انتخاب ویژگی بکار رفته، متغیرهای ورودی را با اعمال قضاوت انسان به ۴ و با اعمال تحلیل واریانس به مورد کاهش داده است. نتایج حاصل از روش‌های شبکه‌های عصبی مصنوعی و درختان تصمیم‌گیری برای هر دو انتخاب ویژگی مبتنی بر قضاوت انسان و تحلیل

<sup>1</sup> James et al.



اسپیناس 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 ۱۱۱ به عنوان متغیرهای ورودی به منظور تطبیق رابطه میان درماندگی مالی Z دومین مدل امتیاز 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 ضریب معنا داری ۱۱۲ وارد مدل شده‌اند.

### ۳-۳ 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 (ساندگرن<sup>۱</sup>، ۲۰۱۴).

لم 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 بود.

یک و همکاران اقدام به آرایه دو مدل به منظور گروه‌بندی شرکت‌ها براساس عملکرد نمودند. در هر دو مدل از نقشه‌های خود سازمان یافته استفاده شده بود نخستین مدل با داده‌های مالی ۱۶۰ شرکت سروکار داشته و بر اساس به کارگیری روش متن کاوی انجام یافت. دومین مدل نیز به تحلیل و بررسی گزارش‌های سالانه آرایه شده به مدیران ارشد اجرایی می‌پردازد. نویسندگان تفاوت‌هایی میان نتایج گروه بندی در دو روش مشاهده می‌کنند.

### ۳-۴ 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 داده کاوی به منظور تسهیل پیش‌بینی ریسک اعتباری می‌تواند به کار برده شود.

هانگ و همکاران تجزیه و تحلیل رتبه‌بندی ریسک اعتباری را با استفاده از ماشین بردار پشتیبان که یک روش یادگیری ماشین است، اجرا کردند. دو مجموعه داده مورد استفاده در این تحقیق، شامل ۷۴ شرکت کره ای و دیگری شامل ۲۶۵ شرکت آمریکایی بوده است. برای هر مجموعه داده، ۵ نوع رتبه‌بندی تعریف شد. همچنین، دو مدل برای داده‌های شرکت‌های کره ای 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 نیز برای اندازه گیری اهمیت نسبی ارزش‌های ورودی استفاده گردید.

### ۴- نتیجه‌گیری:

رشد روزافزون، اهمیت مسائل مالی و گسترش جوامع همراه با پیچیدگی محیط‌های اقتصادی مستلزم آن است که حسابداری و امور مالی از سایر علوم برای حل مسائل و مشکلات خود کمک بگیرد. فنون مدیریت داده‌هایکی از روش‌های نوظهور در این حوزه بوده که روز به روز بر کاربرد و اهمیت آن در حل مشکلات مالی افزوده می‌شود. این فنون دارای انواع مختلفی می‌باشند و قابلیت‌های پیش‌بینی دارند و می‌توانند فرایند تصمیم‌گیری را در مسایل مالی تسهیل کنند. وظایف مالی و غیر مالی در

<sup>1</sup> Sundgren



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 جامع از 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.

## منابع

۱. پورحیدری، امید و زینب اعظمی (۱۳۸۹) شناسایی نوع اظهارنظر حسابرسان با استفاده از شبکه های عصبی، مجله دانش حسابداری، سال اول، شماره ۳، صص ۹۷-۷۷
۲. ستایش، محمدحسین و مظفر جمالیانپور ۱۳۸۸ بررسی رابطه بین نسبتهای مالی و متغیرهای غیرمالی با اظهار نظر حسابرس، فصلنامه تحقیقات حسابداری، سال اول، شماره دوم، صص ۱۵۷-۱۳۰
۳. سجادی، سیدحسین، فرازمنند، حسن و محسن دستگیر (۱۳۸۶). "عوامل مؤثر بر گزارش مشروط حسابرسی"، فصلنامه مطالعات حسابداری، شماره ۱۸، صص ۱۴۰-۱۳۰
۴. منهاج، محمد باقر (۱۳۸۷). مبانی شبکه عصبی، تهران، دانشگاه صنعتی امیرکبیر
5. E.koskivaara "artificial neural networks in analytical review procedures".managerial auditing journal, volume 19,no 2,2004,pp.191-223
6. Gaganis Ch.; Pasiouras F.; and M.Doumpos (2007). "Probabilistic Neural Networks for the Identification of Qualified Audit Opinions", Expert Systems with Applications, Vol. 32, pp. 114-124.
7. H.konno and H. kobayashi."failure discrimination and rating of enterprises by semi-define programming ",asia-pacific financial markets volume 7,issue 3,september, 2013,pp.261-273
8. Hagan, H.; and M. Demuth (1999). Neural Network Design.
9. Ireland J. (2003). "An Empirical Investigation of Determinants of Audit Reports in the UK", Journal of Business Finance and Accounting, Vol. 30, No. 78,pp. 975-1015
10. 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](http://Portal.Acm.Org/Citation.Cfm?Id=778616)
11. 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](http://Portal.Acm.Org/Citation.Cfm?Id=778616).
12. M.lam:"neural network techniques for financial performance prediction:integrating fundamental and technical analysis:decision supportsystems,in press 2012
13. Outecheva, (2007), "corporate financial distress: An empirical analyze of distress risk", PhD's dissertation, University of st. Gallen, Switzerland
14. Palmrose, Z. V. (1986). "Audit Fees and Auditor Size: Further Evidence", Journal of Accounting Research, Vol. 24, pp.97-110.
15. Pasiouras F.;Gaganis, Ch.; and C.Zopounidis (2007). "Multicriteria Decision Support Methodologies for Auditing Decisions: the Case of Qualified Audit Reports in the UK", European Journal of Operational Research, Vol. 180, No.3,

16. Pechenizkiy, S. Puuronen And A. Tsymbal(2006). On The Use of Information Systems Research Methods in Data Mining. Information Systems Development. 487-499
17. pp.1317-1330.
18. Quinlan.J. Generating Production Rules from Decision Trees (1987). IJCAI'87 Proceedings of the 10th international joint conference on Artificial intelligence - Volume 1
19. Spathis, Ch., (2003). "Audit Qualification,Firm Litigation, and Financial Information:An Empirical Analysis in Greece",International Journal of Auditing, Vol.7,pp. 71-85.
20. Sun, Jie, Li Hui, (2008), "Data mining method for listed companies' financial distress prediction", Knowledge-Based Systems, Vol. 21, PP. 1-5.
21. Sundgren, Stefan, (2014), "Auditor choices and auditor reporting practices: evidence from Finnish small firms", The European Accounting Review, Vol. 73, PP. 441-465
22. Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer-Verlag, New York
23. 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,2010.pp.567-587
24. [Www.Blog.Campe.Com.Br/Wp-Content/Uploads/2009/03/Witpress\\_Conf-2.Pdf](http://www.Blog.Campe.Com.Br/Wp-Content/Uploads/2009/03/Witpress_Conf-2.Pdf).

# Evaluation the Impact of Data Mining in Finance and Auditing

Shabnam Tavari

*Master of AI and Robotics*

---

## Abstract

Over the past few years, with the advancement of information technology, the volume of data in databases has doubled every two years, and still more than the past volume of stored information is further tightened. With the World Wide Web, integrated information systems, integrated banking systems, e-commerce, and ... instantaneously added to the amount of data in databases and created massive amounts of data. So that the need to quickly and accurately extract and extract knowledge from these databases has become more evident. Nowadays, data mining is the most important technology for efficient, correct and fast use of bulk data, and its importance is increasing. Data mining is a bridge between statistical science, computer science, artificial intelligence, modeling, machine learning. Data mining is a complex process for identifying patterns and models are correct, new and potentially useful, in a large amount of data, in such a way that these patterns and models can be understood by humans. The most advanced application of business intelligence is data mining. Data mining tools have led to the use of this technology in finance and accounting. These tools are used to predict bankruptcy, continuity of activities, predict financial distress, identify and identify management fraud. The estimation of the credit risk and the operation of the commercial unit, and thus the data mining of Rattbadil has become an issue of great importance in financial affairs and accounting. The purpose of this article is to provide a brief overview of this new technology and its related research, in order to illustrate the effectiveness of the technology in its financial affairs and its research activities

**Keyword:** accounting, financial affairs, data mining methods, data mining

---