



## کاربردهای داده کاوی در حسابداری

مهديه شورئي<sup>۱</sup>. علي تقوي مقدم<sup>۲</sup>. روح اله رحمانی<sup>۳</sup>

۱= دانشجوی کارشناسی ارشد حسابداری موسسه آموزش عالی فردوس مشهد

۲= گروه حسابداری، واحد بجنورد، دانشگاه آزاد اسلامی، بجنورد

۳= گروه حسابداری، واحد موسسه آموزش عالی فردوس مشهد

### چکیده

در طول سال‌های اخیر، حجم مخازن داده‌های الکترونیکی ایجاد شده توسط بانک‌ها، شرکت‌ها و سایر مؤسسات تجاری، همواره در حال افزایش بوده است. بیت‌های با ارزش اطلاعات در این مخازن داده، جای گرفته‌اند، به گونه‌ای که اندازه و حجم بسیار زیاد این منابع، تجزیه و تحلیل آن‌ها را برای انسان به منظور خلق اطلاعات یا الگوهای مفید و مناسب در جهت فرآیند تصمیم‌گیری، غیرممکن ساخته است. فنونی نظیر داده‌کاوی می‌توانند به طور خودکار، اطلاعات را از میان این مقادیر حجیم داده استخراج نمایند. داده‌کاوی، تأثیراتی عمیق بر شیوه‌های کسب و کار و مدیریت دانش در سال‌های اخیر داشته است. هوشمندی کسب و کار، مشهورترین کاربرد فنون داده‌کاوی است. قابلیت‌های طبقه‌بندی و پیش‌بینی ابزارهای داده‌کاوی، موجب کاربرد این ابزارها در سطح وسیعی شده است و بدین ترتیب داده‌کاوی را تبدیل به موضوعی با درجه اهمیت زیاد، در امور مالی و حسابداری نموده است. داده‌کاوی فرآیندی است که با نگرشی نو به مسئله استخراج اطلاعات از داده‌های حجیم پرداخته و به کمک مجموعه‌ای از روش‌های آماری و مدل‌سازی، می‌تواند الگوها و روابط پنهان موجود در داده‌ها را در کمترین زمان ممکن و با دقتی بالا تشخیص دهد. هدف داده‌کاوی جستجو برای یافتن اطلاعاتی با ارزش تجاری در یک پایگاه داده است و از آن می‌توان برای پیش‌بینی رویه‌ها و رفتارهای آتی در بازارهای مالی استفاده کرد.

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

آکادمی داده

آکادمی داده  
dataacademy.ir



## مقدمه

در دورانی که اقتصاد جهانی به سرعت در حال تغییر و تحول است و بازارها به شدت با یکدیگر رقابت دارند، سازمان ها برای قرار گرفتن در این فضای رقابتی باید انطباق یا نهادینه سازی طیف گسترده ای از روش ها و رویکردهای نوین مدیریت و تکنولوژی اطلاعات پیشرفته را در نظر بگیرند. در این بین هوش مصنوعی یا AI برای آینده ی حرفه ی حسابداری امری بسیار مهم به شمار می رود و سیستم های هوشمند باعث پیشرفت های بسیار و چند بعدی در قدرت آنالیز و کارآمدی فرآیندهای حسابداری شده اند. به این شکل مسائل بسیار واضحی وجود دارند که نیازمند توجه بیشتری هستند و باید به وجود فرصت هایی در مقیاس شرکت های کلان اشاره کرد تا بتوانند هر چه بهتر و کاملتر از توانمندی سیستم های حسابداری شرکت خود بهره ببرند. پرسشی که بی پاسخ مانده، این است: که آیا نبود بهره وری کامل از این توانمندی های آنالیز را می توان به واسطه پیچیدگی این سیستم ها، آنگونه که کیم و همکارانش به آن اشاره کردند، توضیح داد یا می تواند به خاطر عواملی دیگر مانند ابعاد خاصی باشد که در تکنیک های داده کاوی یا سرشت کاربرد حسابداری هوشمند وجود دارند؟

داده کاوی یکی از مهم ترین پارادایم های کنونی در ابزار تحلیل پیشرفته ی تجارت های هوشمند و پشتیبانی از تصمیمات است. چنین تمایزی مورد توجه اغلب هیئت های حرفه ای حسابداری می باشد. AICPA داده کاوی را به عنوان یکی از ده تکنولوژی برتر برای آینده در نظر گرفته، IIA آن را به عنوان یکی از چهار اولویت پژوهش در نظر می گیرد و علاوه بر این CGMA گزارش داده بیش از پنجاه درصد شرکت های بزرگ داده کاوی و داده های گسترده را برزمره ی ۱۰ اولویتی می دانند که برای دوران داده - محور تجارت، امری بنیادین و اساسی به شمار می رود. داده کاوی به عنوان فرآیند شناسایی الگوهای معتبر تعریف شده که به صورت بالقوه نوین هستند و امکان درک بسیار بالایی از آنها وجود دارد. داده کاوی همچنین به عنوان فرآیند استخراج دانش از توده های عظیم داده شناخته می شود تا تصمیم گیری ها در یک رشته یا زمینه ی خاص را بهبود بخشد. در نتیجه، عمده توجه داده کاوی، کاربردی کردن دارایی های اطلاعاتی یک سازمان برای بدست آوردن مزایا و سود اقتصادی و غیر اقتصادی می باشد. در نتیجه، داده کاوی تقریباً برای تمام رشته های تجاری و غیر تجاری، شامل حسابداری کاربرد یافته است.

## داده کاوی

داده کاوی، کاربرد الگوریتم هایی خاص برای استخراج الگو از داده ها می باشد. این کار امکان اکتشاف اتوماتیک الگوهای شاخص و دانسته های جالب توجهی که در مقادیر عظیم داده پنهان شده اند را، فراهم می آورد. داده کاوی به شرکت ها کمک می کند بر مهم ترین اطلاعاتی که در دیتابیس های کنونی آنها موجود است، متمرکز شوند. اما داده کاوی فقط یک ابزار است و نیاز به شناخت تجارت، درک داده ها یا درک روش های تجزیه و تحلیل مورد نیاز را از بین نمی برد. داده کاوی سه هدف اصلی دارد: توضیح، پیش بینی و تجویز. از سوی دیگر، تجویز بر فراهم کردن بهترین راه حل برای مشکل مطرح شده متمرکز است. این اهداف را می توان با به کارگیری بسیاری از کارهای داده کاوی شامل طبقه بندی، دسته بندی، پیش بینی، شناسایی داده ی پرت، بهینه سازی و تصویر سازی دنبال کرد. این کارها بسته به نوع مشکلی که باید حل شود، به این شکل با یکدیگر فرق می کنند:



- ❖ طبقه بندی بر باز نمایی داده ها روی مجموعه ای از طبقات با صفات کیفی از پیش تعیین شده و مجزا متمرکز است که می توانند باینری یا چند دسته ای باشند.
- ❖ دسته بندی مرتبط با دسته بندی داده ها به گروه ها یا دسته های معنا دار است.
- ❖ پیش بینی بر یافتن داده های عددی آینده (پیش بینی) یا ارزش های غیر عددی (طبقه بندی) متمرکز است.
- ❖ شناسایی داده ی پرت بر یافتن داده هایی متمرکز است که به طرز قابل توجهی از سطح عادی انحراف دارند.
- ❖ بهینه سازی متمرکز بر یافتن بهترین راه حل با توجه به منابع است.
- ❖ تصویر سازی بر ارائه و درک بصری داده ها تمرکز دارد.
- ❖ رگرسیون بر تخمین متغیر وابسته از مجموعه متغیرهای غیر وابسته متمرکز است.

اصولاً هر جایی که داده وجود داشته باشد ، داده کاوی نیز معنا می یابد. این فرآیند، دارای مزایایی با اهمیت است از جمله:

(۱) کمک به مدیران در تصمیم گیری

(۲) یافتن مقالات خبری روزآمد

(۳) کمک به کاربر برای جستجو در وب

(۴) بهبود امر بازاریابی

(۵) کشف کلاهبرداری ها، تقلب و ناپهنجارها

اساس داده کاوی بر مبنای سه فعالیت اصلی است که عبارتند از:

الف - حذف داده ها : داده های بی ارزش و عوامل بیرونی حذف می شوند.

ب - فشرده سازی داده ها : این عمل به وسیله کد گذاری داده ها صورت می گیرد.

ج - کشف الگوها : الگوهای موجود در پایگاه داده ها از قبیل طبقه بندی، الگوهای زنجیری و .... کشف می شوند.

### مفاهیم داده کاوی

#### الف - مراحل داده کاوی

پژوهشگران مراحل مختلفی را برای داده کاوی معرفی کرده اند و در تمامی آنها داده کاوی در قالب یک فرآیند معرفی شده است. مراحل داده کاوی عبارتند از:

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



- ۲- انتخاب داده ها : در این مرحله باید داده ها بر مبنای معیارهای مشخص انتخاب گردند.
- ۳- آماده سازی داده ها : شکل قابل استفاده داده و شناسایی متغیرهای زائد وظیفه این مرحله از فرآیند خواهد بود.
- ۴- ارزیابی داده ها : چارچوب کلی این مرحله، معیارهایی از قبیل نوع توزیع داده ها، ویژگی ها و ساختار پایگاه داده و شرایط کلی داده ها و .... می باشد.
- ۵- قالب بندی پاسخ: خروجی این بخش، ارائه فرمت به شکل تصویر، نمودار، شبکه عصبی و .... است .
- ۶- انتخاب ابزار: در این مرحله ابزارهای مناسب برای داده کاوی انتخاب می گردد.
- ۷- الگوسازی: فرآیند داده کاوی به صورت اصلی از این مرحله آغاز می گردد که شامل جستجوی الگوها در مجموعه داده، طبقه بندی و ارزشیابی داده ها و .... می باشد.
- ۸- اعتبار سازی یافته ها : این مرحله، شامل آزمون کردن الگوها است.
- ۹- ارایه نتیجه : نتیجه این بخش ، گزارش نهایی برای کاربر است.
- ۱۰- استفاده از نتایج: هدف اصلی داده کاوی استفاده از نتایج کشف شده برای تصمیم گیری، سیاست گذاری و پیش بینی به منظور ایجاد یک موقعیت بهتر و جدید می باشد. (نجات و اکبری ، ۱۳۸۷)

## ب- فنون داده کاوی

فنون داده کاوی می تواند به دو دسته فنون مستقیم ( با رویکرد از بالا به پایین) و غیر مستقیم (با رویکرد از پایین به بالا) تقسیم شود. با استفاده از داده کاوی مستقیم، متغیر خاصی شناسایی می شود که مدنظر قرار گرفته شده است. فناوری داده کاوی روابط بین یک متغیر و جمعیت انتخاب شده از دیگر متغیرها را می یابد. با داده کاوی غیر مستقیم متغیر مشخصی هدف قرار نمی گیرد( متغیر وابسته) بلکه هدف یافتن روابط بین متغیرها در جمعی از اطلاعات است. راه دیگر برای توصیف این دو دسته فنون این طور است که بگوییم رویکرد از بالا به پایین ، فرضیه های خاص را آزمون می کند و رویکرد از پایین به بالا، فرضیه های جدید را تولید می کند.

نمونه هایی از داده کاوی مستقیم به این شرح است:

طبقه بندی : مجموعه از قوانین بر اساس داده های موجود (مجموعه های آموزشی) ایجاد می شود تا بر اساس آن طبقه بندی مناسبی برای موضوعی جدید در میان طبقه بندی های مختلف انجام گیرد. کیفیت طبقه بندی فنون داده کاوی می تواند از طریق صحت و درستی طبقه بندی ، تأثیر گذاری و توانایی تفسیر نتایج طبقه بندی، سنجیده شود. روش های طبقه بندی شامل درخت تصمیم گیری، شبکه های بیزی، شبکه های عصبی، ماشین های بردار پشتیبان ، الگوریتم های ژنتیکی و .... می باشد.



برآورد: طبقه بندی برای پیامدهای نسبتاً مجزا به کار می رود و در آن متغیرها مستقل هستند ولی در فنون برآورد از متغیرهای پیوسته استفاده می شود. به عنوان نمونه، به جای طبقه بندی ریسک در سطح پایین، متوسط و بالا، فنون برآورد به سطح ریسک امتیاز می دهد و ریسک را از ۰ تا ۱۰ تعیین می کنند. سپس این برآوردها می توانند از طریق فنی مشابه رگرسیون لجستیک، عمل طبقه بندی را انجام دهد و موضوعی را طبقه بندی کند. به عنوان نمونه، اگر ریسک بالا برای مشتری، ریسک بالای ۷/۵ می باشد. برآوردها می توانند برای رتبه بندی یک مجموعه و سپس برای تعیین یک آستانه نمره، مورد استفاده قرار گیرند.

پیش بینی: طبقه بندی و برآورد از جمله فنون داده کاوی هستند که برای افشای اطلاعات و ویژگی های متغیرها، مورد استفاده قرار می گیرند که قبلاً در مجموعه ای از داده ها، شناسایی شده است.

پژوهشگران داده کاوی اغلب بین فنون "پیش بینی" با طبقه بندی و برآورد تفاوت قائل می شوند. هدف از پیش بینی، آزمون سیستماتیک فرضیه و همچنین یافتن وضعیت نامطلوبی است که خارج از محدوده مورد انتظار اتفاق می افتد. این فنون پیش بینی مشابه فنون طبقه بندی، در قالب مجموعه های آموزشی برای مدلسازی اولیه مانند نمونه های آزمایشی یا روش های خود راه اندازی در مجموعه داده های نهایی استفاده می شود. (Han and Kamber 2006)

نمونه هایی از داده کاوی غیر مستقیم به این شرح است:

**dataacademy.ir**

گروه بندی ترکیبی: هدف از گروه بندی ترکیبی، یافتن روابط بین متغیرها در مجموعه داده ها می باشد. در این گروه بندی هیچ متغیر مستقل و وابسته خاصی وجود ندارد. تحلیل گر داده ها به دنبال جستجوی روابط بین متغیرها می باشد.

خوشه بندی: هدف از خوشه بندی این است که متغیرها را بر اساس روابط موجود در مجموعه داده ها، نسبت به زیر گروه ها دسته بندی شود. هدف از طبقه بندی روش مستقیم و خوشه بندی غیر مستقیم، مشابه می باشد. با این حال شیوه عمل و نتایج کاملاً متفاوت است. در طبقه بندی مستقیم، زیر گروه ها دارای دسته های از پیش تعریف شده بر حسب متغیرهای از پیش انتخاب شده می باشند.

### ج- روش های داده کاوی

اهداف داده کاوی شامل پیش بینی و توصیف یا ترکیبی از آنهاست. هدف "پیش بینی" تمرکز بر روی دقت در توانایی پیش بینی بوده و "توصیف" بر درک فرآیند تولید داده ها تمرکز دارد. در پیش بینی تا زمانی که مدل قدرت پیش بینی دارد، کاربر توجهی به این ندارد که مدل انعکاس دهنده واقعیت است؛ مثلاً مدلی که شاخص های مالی را به شکل غیر خطی ترکیب می کند تا نرخ تبادل ارز را پیش بینی نماید. از سویی دیگر مدل توصیفی به مثابه واقعیت تفسیر می شود. برای مثال مدلی که متغیرهای اقتصادی و جمعیتی را به پیشرفت های آموزشی مرتبط می سازد، به عنوان مبنایی برای توصیه های سیاست اجتماعی توصیف می شود. در عمل، اغلب کاربردهای اکتشاف دانش به درجه ای از هر دو مدل سازی توصیفی و پیش بینی نیاز دارند.



به هر ترتیب، اهداف داده کاوی با استفاده از روش های داده کاوی ، محقق می شوند. اصطلاح روش های داده کاوی در واقع بیانگر جمع کثیری از الگوریتم ها و فنون است که از علومی مانند آمار، یادگیری ماشین ، پایگاه داده و تجسم سازی استنتاج شده اند.

### تکنیک های داده کاوی

طیف گسترده ای از تکنیک های داده کاوی وجود دارد که از آن ما به تعریف این موارد می پردازیم شبکه های عصبی، الگوریتم های ژنتیک، ماشین های بردار پشتیبان ، نزدیکترین همسایگی.

بدیهی است که این فهرست از تکنیک ها، فهرستی کامل نبوده و ترتیبی برای اولویت بندی در کاربرد این روش ها، پیشنهاد نمی شود.

### شبکه های عصبی NNS

این شبکه ها ابزارهایی هستند که در موارد پیش بینی، خوشه بندی و تخمین کاربرد دارند. جذابیت شبکه های عصبی در این است که آن ها به وسیله مدل سازی ارتباطات عصبی مغز انسان در کامپیوترهای دیجیتالی پلی برای فاصله موجود ایجاد می کنند.

[dataacademy.ir](http://dataacademy.ir)

برای شناخت الگوهایی که در داده ها وجود دارند بسیار مفید هستند، خصوصاً در مواقعی که نوع رابطه بین هدف (به عنوان مثال، ارزیابی وضعیت اعتباری وام گیرنده) و متغیرهای ورودی (مثلاً ویژگی های جمعیتی، نسبت های مالی، روند اقتصادی، کیفیت مدیریت و ...) نامشخص و یا پیچیده باشد. یک شبکه عصبی براساس سه ویژگی داده های ورودی، وزن ها و لایه های پنهان مشخص می شود. برای ایجاد شبکه های عصبی چندین روش وجود دارد که شبکه عصبی پرسپترون، به ویژه پرسپترون چند لایه ، در زمره ی کاربردی ترین شبکه های عصبی مصنوعی می باشد. این شبکه ها می توانند با گزینش شماره لایه ها و سلول های عصبی (نرون ها) ، که اغلب زیاد نیستند، یک نگاشت غیر خطی را با دقت دلخواه انجام دهد. همچنین این روش به عنوان " طبقه بندی جامع " شناخته می شود به این دلیل که از لحاظ تئوریک قادر به طبقه بندی هر فرآیند تصمیم گیری است.

شبکه های عصبی روش محاسبه ای متفاوت با روش های متداول می باشند. محاسباتی که با روش های معمولی انجام می شود از نوع برنامه ریزی شده است و در آن ها الگوریتم ها و مجموعه هایی از قواعد به کار می روند تا مسئله را حل کنند. در این روش ها اگر الگوریتمی در دست نباشد یا الگوریتم پیچیده باشد، راه حلی برای مسئله وجود نخواهد داشت. اما در محاسباتی که در شبکه عصبی اجرا می کند به الگوریتم و مجموعه قواعد نیاز نیست.

### الگوریتم های ژنتیک GA

الگوریتم ژنتیک نخستین بار به وسیله هلند در سال ۱۹۷۵ مطرح شد و در سال های بعد توسط محققان دیگر توسعه پیدا کرد. الگوریتم ژنتیک بخشی از نظریه محاسبه تکاملی است که در حال حاضر به عنوان بخشی از هوش مصنوعی به سرعت در حال رشد است. ایده اصلی این الگوریتم در نظریه تکاملی داروین نهفته است. از نظر کاربردی، الگوریتم ژنتیک یکی از روش های



بهینه سازی مسائل است که اساس آن بر انتخاب طبیعی (عامل اصلی تکامل زیستی) و برخی از مفاهیم مهم از علم ژنتیک استوار است. در این روش برای بهینه سازی تابع هدف (تابع شایستگی مسئله)، از یک جمعیت اولیه کروموزوم ها (افراد) که در حقیقت پاسخ های اولیه مسئله هستند، به یک جمعیت جدید از کروموزوم ها یا یک نسل جدید که در حقیقت پاسخ های ثانویه مسئله مفروض است، می رسد. با تکرار این عملیات و تولید جمعیت جدید از جمعیت قبلی در هر مرحله و در نتیجه رسیدن به نسل های موفق، جمعیت به سمت یک پاسخ بهینه رشد خواهد کرد.

### ساختار درختی تصمیم گیری DT

الگوریتم درخت تصمیم یکی از عمومی ترین تکنیک های داده کاوی می باشد. این الگوریتم با ایجاد یکسری گره ها و شعبه های به هم متصل سعی در طبقه بندی متغیرهای اسمی و تخمین متغیرهای عددی دارد. گره های یک درخت تصمیم نشان دهنده نقاط آزمون متغیرها است. بر مبنای نتایج آزمون صورت گرفته درخت تصمیم به گره ها یا برگ های جدید می رسد. ریشه درخت اولین و اساسی ترین گره درخت و برگ های آن آخرین نقطه تفکیک و به عبارتی نشان دهنده طبقه بندی است. مدل های حاصل از درخت تصمیم قابلیت تفسیر بالایی دارند. تکنیک درخت تصمیم دارای الگوریتم های متفاوتی مثل متد ID3 ، C5.0 ، CART و CHAID می باشد (Ata & sey rek 2009).

## ماشین های بردار پشتیبان SVM

این متد یک متد ریاضی قوی در دسته بندی است. متدی شبیه به شبکه های عصبی که به جای خط جدا کننده به دنبال بهترین خط جدا کننده است که دارای حداکثر حاشیه باشد. یعنی بهترین خط جدا کننده ای که با نزدیکترین نقطه کم ترین فاصله را داشته باشد. در ماشین بردار پشتیبان انتخاب بهینه ویژگی ها در دقت مدل تأثیر بسزایی دارد از این رو در سال های گذشته بسیاری از تحقیقات بر روی ماشین بردار پشتیبان ، بر روی انتخاب بهینه پارامترها متمرکز می باشد.

تکنیک های یادگیری ماشین بخشی از علم هوش مصنوعی به شمار می آیند. هوش مصنوعی به طور مؤثر و موفق در گستره وسیعی از شاخه های تجاری همچون بازاریابی ، حسابداری ، سیستم های اطلاعات مدیریت و مدیریت تولید به کار گرفته می شود. در اکثر تحقیقات از روش های یادگیری ماشینی برای پیش بینی تغییرات آتی سهام، بحران های مالی، ورشکستگی، شناسایی کلاهبرداری کارت های اعتباری به کار گرفته می شود.

### نزدیکترین همسایگی K

نزدیکترین همسایگی یک روش طبقه بندی اشیاء بر اساس نزدیکترین نمونه های آموزشی در فضای ویژگی است . هنگام تلاش برای حل مسائل جدید ، افراد معمولاً به راه حل های مسائل مشابه پیشین ، مراجعه می کنند. الگوریتم K – نزدیکترین همسایگی یک تکنیک دسته بندی است که از نسخه ای از این روش استفاده می کند. در این روش تصمیم گیری، تعیین اینکه یک مورد جدید در کدام دسته قرار گیرد، با بررسی تعدادی (k) از شبیه ترین موارد یا همسایه ها انجام می شود. تعداد موارد هر دسته شمرده شده و مورد جدید به دسته ای که تعداد بیشتری از همسایه ها به آن تعلق دارند، نسبت داده می شود. اولین کاربرد k – نزدیکترین همسایگی یافتن معیاری برای محاسبه فاصله بین ویژگی ها در داده ها است. در حالی که این عمل برای داده های



عددی آسان است، متغیرهای دسته ای نیاز به برخورد خاصی دارند. هنگامی که فاصله بین موارد مختلف اندازه گیری شد، می توان از مجموعه مواردی که قبلاً دسته بندی شده اند، به عنوان پایه و دسته بندی موارد جدید استفاده کرد و پس از تعیین فاصله همسایگی، چگونگی شمارش همسایه را مشخص کرد.

**مطالعه تحقیقات کاربرد داده کاوی در الف.** ژورنال های حسابداری مطرح ، ب. تمام ژورنال هایی که با مجلات گروه الف همپوشانی نداشته باشند و توسط انجمن حسابداری آمریکا (AAA) منتشر شده باشد، ج. منابع الکترونیک کتابخانه دانشگاه که شامل بیش از هشتاد درصد توضیحات جنبی درباره ی دیتابیس های الکترونیک باشند، د. ohio LINK که دیتابیس الکترونیک حاوی بیش از ۹۰۰۰ ژورنال مربوط به ۱۰۱ انتشارات است و Google Scholar.

نمونه مورد استفاده شامل ۲۰۹ کاربر داده کاوی در حسابداری بین سال های ۱۹۸۹ تا ۲۰۱۴ را شناسایی کرد (۲۳ مورد از آنها در مقالات قبل از کنفرانس و ۱۸۶ مورد در مقالات ژورنال ها مورد بررسی قرار گرفته اند). یک گرایش تصاعدی و روبه رشد کاربرد داده کاوی بین سال های ۱۹۹۵ و ۲۰۰۱ و ۲۰۰۴ تا ۲۰۱۴ مشاهده می شود که بیشترین تعداد کاربرد مربوط به سال ۲۰۱۳ است این طور به نظر می رسد که پژوهشگران و متخصصان حسابداری به مزایای استفاده از داده کاوی در حسابداری پی برده و در نتیجه با گذر زمان ، تمایل بیشتری به استفاده از آن پیدا کرده اند. علاقه ی اصلی ما به جهش کوانتومی در تعداد چنین کاربردهایی، از سال ۲۰۱۰ به بعد است. یکی از دلایل اصلی برای چنین جهشی می تواند نیاز به پیراستگی بیشتر در مدلسازی حسابداری باشد که از بحران مالی ۲۰۰۸ در سرتاسر جهان و فروپاشی بنگاری متعاقب آن ناشی می شود.

مرور این مقالات نشان می دهد که وضعیت کنونی و درهم آمیخته ی حسابداری که یکی از رشته های اصلی تجارت به شمار می رود، با داده کاوی که یکی از ده تکنولوژی برتر سیستم اطلاعاتی در آینده است ، به هیچ وجه مورد کاربرد گسترده نیستند. کاربردهای داده کاوی در حسابداری عمدتاً و با درجات مختلف، به سه شاخه از حسابداری متمرکز هستند: بیمه و حسابداری مدیریت و حسابداری مالی. داده کاوی در رشته ی بیمه و حسابرسی بیشترین کاربرد را دارد و پس از آن حسابداری مدیریتی و مالی قرار می گیرند. اغلب این کاربردها به طرز شگفت آوری بر پیش بینی به عنوان یک هدف، طبقه بندی به عنوان یک کار و شبکه های عصبی به عنوان تکنیک متمرکز هستند. دو هدف دیگر داده کاوی، یعنی توصیف و تجویز مورد توجه شایانی در این زمینه قرار نگرفته اند. هر چند پیش بینی در مورد اغلب مسائل حسابداری مناسب به شمار می رود و در برگیرنده ی چشم انداز آینده، جهت دهی های استراتژیک، راهنمایی ها و موضع گیری است، اما نمی توان اینگونه مطرح کرد که دیگر اهداف داده کاوی که توجه کمتری به آنها شده، برای تصمیم گیری های مرتبط با حسابداری از ارزش استراتژیک کمتری برخوردار هستند. استفاده ی کم از کاربردهای تجویزی، شکافی آشکار است که باید با پویایی بیشتری مورد توجه قرار گیرد تا مزایای بهینه سازی در تصمیم گیری آشکار شود.

توجه بیش از حد به طبقه بندی، فرصت بالقوه برای استفاده از دیگر کارهای داده کاوی مانند آنالیز الگو و ارتباط را آشکار می کند که در رشته های دیگر، مزایای بسیاری داشته اند. هر چند غلبه ی استفاده از تکنیک شبکه های عصبی بازتاب دهنده قدرت این تکنیک به عنوان یک ابزار عمومی برای مدل سازی است، اما فقط تکیه بر این تکنیک که به خاطر جعبه سیاه بودنش مشهور است، به ایجاد مدل های متعدد دیگری می انجامد که فاقد قابلیت توضیح کافی هستند. به طور خلاصه باید گفت، کاربردهای





کنونی داده کاوی در حسابداری عمدتاً در رشته بیمه و حسابرسی هستند و پیش بینی را به عنوان هدف اصلی، طبقه بندی را به عنوان کار اصلی و شبکه های عصبی را به عنوان تکنیک اصلی خود اتخاذ کرده اند. تحت چنین شرایطی نمی توان گفت که حسابداری به راستی و به اندازه کافی از قدرت و توانمندی های داده کاوی بهره برده است.

### کاربرد داده کاوی در بررسی عوامل مالی و غیر مالی مؤثر برگریز مالیاتی

مالیات به عنوان یکی از مهمترین منابع مالی دولت، نقش بسزایی در رشد اقتصادی و توزیع درآمد ایفا می کند. سازمان امور مالیاتی به عنوان متولی وصول قانونی درآمدهای مالیاتی کشور، با موضوع گریز مالیاتی مودیان مواجه می باشد. انحراف مالیات ابرازی مودیان از مالیات قطعی شده آنان، می تواند مصداقی از گریز مالیاتی باشد که به نوبه خود تأثیرات مخربی بر اقتصاد کشور وارد می کند. رویکرد سازمان امور مالیاتی جهت رویارویی با این چالش، استفاده از روش های سنتی ممیزی و زمانبر می باشد. استفاده از این رویکرد در بلند مدت موجب کاهش درآمدهای مالیاتی دولت و افزایش هزینه های آن می شود. داده کاوی از جمله ابزارهای مفید جهت افزایش کارایی و اثر بخشی کشف فرارهای مالیاتی بوده و در حداکثر کشورها به این منظور استفاده می شود. (Micc – Barrera 2004)

به طور کلی، هرگونه تلاش غیر قانونی برای پرداخت نمودن تمام یا قسمتی از مالیات (مثل ندادن اطلاعات در خصوص عواید و

منافع)، فرار مالیاتی خوانده می شود (شفیعی و صبوری ۱۳۸۸، جهرمی و دیگران، ۱۳۸۸)

وجود نارسایی در قوانین موجود، کمبود ممیزان و مأمورین مالیاتی مجرب، نزولی بودن نرخ های مالیاتی در عمل، برخوردهای ناعادلانه مالیاتی، غیر عادلانه بودن نرخ های مالیاتی، کندی سیستم مالیاتی، تبانی مودی و ممیز، حساب سازی های مالیاتی و فقدان یک مرکز قوی اطلاعات مالیاتی از عوامل بروز فساد مالی و اقتصادی در حوزه مالیاتی می باشد (عزیز خانی و افشاری، ۱۳۸۴). این موضوعات باعث می شود که هر ساله درصدی از تولید ناخالص داخلی کشور خارج از شمول مالیات قرار گرفته و موجب عدم تحقق اهداف پیش بینی شده در مورد درآمدهای مالیاتی و در نتیجه عدم تحقق عدالت مالیاتی می شود. همان طور که ذکر شد، عمده عوامل بروز فساد مالی و اقتصادی در حوزه مالیاتی عمدتاً مربوط به دو حوزه عالی (سیاستگذاری و قانونگذاری) و عملیاتی (رسیدگی و ممیزی) می باشند.

داده کاوی می تواند عوامل مؤثر برگریز مالیاتی را شناسایی و مدل هایی را جهت کشف میزان احتمالی گریز مالیاتی مودیان ارائه دهد. بنابراین، دانش ایجاد شده از فرایند داده کاوی راهکارهایی را به سیاستگذاران حوزه مالیاتی در خصوص قانون گذاری و چارچوبی را به ممیزان مالیاتی جهت رسیدگی کارا و اثر بخش ارائه می دهد. در حوزه عملیاتی (رسیدگی و ممیزی)، داده کاوی پا را فراتر گذاشته و می تواند مبنایی جهت پیاده سازی سیستم حسابرسی مبتنی بر ریسک در سازمان امور مالیاتی فراهم آورد. براین اساس، ممیزان مالیاتی می توانند قبل از نمونه گیری و انجام رسیدگی، کلید مودیان مالیاتی را در بازه ای از ریسک پایین تا بالا (به لحاظ احتمال گریز مالیاتی) دسته بندی و برنامه های رسیدگی خود را براساس این دسته بندی تدوین و اجرا نمایند. به این ترتیب ممیزان مالیاتی می توانند در خصوص این سؤال که کدامیک از مودیان نیاز به بررسی بیشتری دارند، تصمیمات مناسبی را اتخاذ نمایند.



این رویه از یک طرف موجب افزایش کارایی و اثر بخشی رسیدگی ها شده و از طرف دیگر موجب کاهش هزینه های رسیدگی و افزایش درآمدهای مالیاتی سازمان امور مالیاتی می شود. به طور کل، مزایای استفاده از فرآیند داده کاوی در کشف گریز مالیاتی مودیان به شرح زیر است:

۱- افزایش کارایی و اثربخشی در رسیدگی ها از طریق ارائه الگویی جامع در جهت پیاده سازی حسابرسی مبتنی بر ریسک در سازمان امور مالیاتی

۲- استفاده از فناوری های جدید در رسیدگی ها و افزایش دقت عمل و همچنین تسهیل و یا حذف عملیات مکانیکی در رسیدگی ها

۳- شناسایی عوامل مؤثر برگریز مالیاتی مودیان و ارائه راهکارها قانونی و عملی جهت جلوگیری از آن

۴- افزایش سلامت نظام مالیاتی و تبع آن رشد و شکوفایی اقتصادی و بهبود رفاه و عدالت اجتماعی

۵- افزایش درآمد مالیاتی سازمان امور مالیاتی

ایالت تگزاس آمریکا یکی از اولین استفاده کنندگان از تکنیک های داده کاوی جهت کشف فرارهای مالیاتی می باشد. مدل های ایجاد شده در این طرح، به این سؤال مبنیان مالیاتی که آیا مودی مورد نظر نیاز به بررسی بیشتر دارد یا خیر، جواب می دادند. نتایج اجرای طرح نشان داد که استفاده از تکنیک های داده کاوی منجر به کشف و وصول ۳۶۲ میلیون دلار مالیات از مودیان که فرار مالیاتی داشتند، شده است (Hoover, 2009). تحقیق دیگری در سال ۲۰۱۲ با استفاده از تکنیک های قواعد انجمنی داده کاوی و به کارگیری متغیرهای مالی (همچون نسبت فرار مالیاتی به مالیات، نوع صنعت، میزان سرمایه، میزان فروش، نسبت فروش معاف از مالیات به فروش، نسبت فروش به سرمایه) قواعدی را برای کشف فرار از مالیات بر ارزش افزوده ارائه داده است. فرار مالیاتی مودیان را با دقت های ۷۳/۳۵ درصد و ۸۴/۱۷ درصد کشف می کردند. (wu & et al, 2012)

### کاربرد داده کاوی در پیش بینی گزارش حسابرس مستقل

حسابرسان پی در پی باید با داده های حجیم و ساختار پیچیده این داده ها سروکار داشته باشند. در نتیجه، حسابرسان بیش از این نمی توانند تنها به ابزارهای گزارشگری یا خلاصه کردن در فرآیند حسابرسی متکی باشند. ابزارهای دیگری نظیر داده کاوی که به طور خودکار اطلاعات را از میان حجم داده های زیاد، استخراج می کنند، می توانند مفید باشند. اگر چه تطابق فنون داده کاوی در فرآیند حسابرسی در واقع زمینه ای نسبتاً جدید می باشد، با وجود این، داده کاوی نشان داده است که در بسیاری از کاربردهای تجاری مرتبط با حسابرسی نظیر کشف تقلب، حسابداری داده کاوی دارای مزیت های فراوانی خواهد بود. از طرفی دیگر رویدادهای اخیر نیز نشان دهنده مشکلات قابل توجهی در فرآیند حسابرسی است. سقوط انرون و آرتور اندرسن و سایر موارد مشابه، همچنین کاربرد رو به گسترش پختن دفاتر در حرفه حسابداری شواهد دیگری برای تغییر در فرآیند حسابرسی ایجاد نموده است.



بال و تابور (۱۹۹۱) در پژوهش خود به این نتیجه رسیدند که نرخ بازده سرمایه، نسبتاً گردش موجودی کالا، نسبت گردش حساب های دریافتی، نسبت جاری و اهرم مالی توانایی پیش بینی گزارش مشروط حسابرسی را دارند.

اسپاتیس (۲۰۰۳) در پژوهشی با استفاده از نمونه ای شامل ۱۰۰ شرکت یونانی و با بهره گیری از روش های رگرسیون حداقل مربعات معمولی و لجستیک به بررسی تأثیر دعاوی حقوقی و اطلاعات مالی بر گزارش مشروط حسابرسی پرداخت. وی به این نتیجه رسید که دعاوی حقوقی و بحران مالی از عوامل مؤثر بر صدور گزارش مشروط می باشند. صحت مدل وی در پیش بینی نوع اظهار نظر حسابرسی به میزان هشتاد و هفت درصد بود.

دومپوس و همکاران (۲۰۰۵) با استفاده از نمونه ای شامل ۱۷۵۴ شرکت انگلیسی طی دوره زمانی ۱۹۹۸ تا ۲۰۰۳ از ماشین بردار پشتیبان برای ایجاد مدل های خطی و غیر خطی برای پیش بینی نوع اظهار نظر حسابرس استفاده کردند. آن ها به این نتیجه رسیدند که روش غیر خطی در مقایسه با روش خطی نتایج مطلوب تری ایجاد نمی کند.

پاسیوراس و همکاران (۲۰۰۶) با استفاده از مدل متمایز سلسله مراتبی چند گروهی به پیش بینی نوع اظهار نظر حسابرسان پرداختند. نمونه آن ها شامل ۸۲۳ شرکت طی سال های ۱۹۹۸ تا ۲۰۰۳ بوده آن ها از تجزیه و تحلیل تمایزی و لجستیک به منظور مقایسه استفاده نمودند و به این نتیجه رسیدند که مدل متمایز سلسله مراتبی چند گروهی در مقایسه با مدل های ذکر شده دارای توان پیش بینی بیشتری است.

گاگانیس و همکاران (۲۰۰۷) در پژوهشی با استفاده از نمونه ای شامل ۸۸۱ شرکت انگلیسی طی سال های ۱۹۹۷ تا ۲۰۰۴، پتانسیل روش شبکه عصبی احتمالی را در ایجاد مدلی برای پیش بینی نوع اظهار نظر حسابرسی مورد بررسی قرار دادند. نتایج، بیانگر قدرت مدل بدست آمده با استفاده از شبکه عصبی احتمالی بود. همچنین یافته ها نشان داد که این روش در مقایسه با روش شبکه عصبی مصنوعی و رگرسیون لجستیک روش مناسبتری است.

گاگانیس و همکاران (۲۰۰۷) در پژوهشی کارایی روش نزدیک ترین همسایگی، برای ایجاد مدل هایی برای برآورد نوع اظهار نظر حسابرسی، را در مقایسه با روش های لجستیک و تجزیه و تحلیل تمایزی خطی مورد بررسی قرار دادند. نمونه مورد بررسی آن ها متشکل از ۵۲۷۶ را مشاهده بود. مقایسه روش های مختلف نشان داد که روش نزدیک ترین همسایگی در مقابل سایر روش ها مورد بررسی می تواند اثر بخش تر باشد.

افستاتیوس و همکاران (۲۰۰۷) با استفاده از نمونه ای شامل ۴۵۰ شرکت ایرلندی و انگلیسی از سه روش داده کاوی شامل روش پرسپترون چند لایه، درخت تصمیم و شبکه بیزین برای طبقه بندی اظهار نظر حسابرسان استفاده نمودند. نتایج این پژوهش بیانگر عملکرد کلی بالاتر شبکه بیزین نسبت به سایر روش ها است.

زدولسک و جاگریک (۲۰۱۱) با استفاده از نمونه ای شامل ۵۳۰ شرکت انگلیسی و ایرلندی مدلی برای تعیین اظهار نظر حسابرسی مشروط ارائه دادند و آن را به وسیله منحنی ویژگی های عملیاتی دریافت کننده (ROC) و هزینه های طبقه بندی اشتباه مورد ارزیابی قرار دادند. آن ها به این نتیجه رسیدند که استفاده از ترجیحات ذهنی حسابرسان برای ارزیابی عملکرد مدل، بر مبنای این معیارهای عملکرد، کافی است.



سجادی و همکارانش (۱۳۸۶) در پژوهشی به بررسی عوامل مؤثر بر گزارش مشروط حسابرسی با استفاده از نمونه ای شامل ۱۴۴ شرکت طی سال های ۱۳۸۱ تا ۱۳۸۳ پرداختند. نتایج حاصل از پژوهش آن ها براساس رگرسیون لجستیک و آزمون استقلال کای مربع حاکی از این بود که از بین متغیرهای مورد بررسی ، نسبت جاری و نسبت حساب های دریافتی به دارایی بر گزارش مشروط حسابرسی مؤثر هستند. همچنین ، بین گزارش مشروط حسابرسی سال قبل و نوع مؤسسه حسابرسی با گزارش حسابرسی سال جاری رابطه معناداری وجود دارد.

ستایش و جمالیان پور (۱۳۸۸) با استفاده از نمونه ای شامل ۸۹ شرکت طی سال های ۱۳۸۱ تا ۱۳۸۵ به بررسی رابطه میان ۱۴ نسبت مالی و دو متغیر غیر مالی با نوع اظهارنظر حسابرسان پرداختند. آن ها با بهره گیری از تحلیل رگرسیون لجستیک به مدلی شامل ۷ متغیر مالی و غیر مالی با ۷۶/۲ درصد موفقیت در پیش بینی اظهارنظر مقبول دست یافتند.

پورحیدری و اعظمی (۱۳۸۹) در پژوهشی به شناسایی نوع اظهارنظر حسابرسان با استفاده از شبکه عصبی و رگرسیون لجستیک پرداختند. نتایج این پژوهش با استفاده از نمونه ای شامل ۱۰۱۸ مشاهده طی سال های ۱۳۷۹ تا ۱۳۸۶ حاکی از توان بالای شبکه عصبی در شناسایی و پیش بینی نوع اظهارنظر حسابرسان بود. این روش با میزان صحت ۸۷/۷۵ درصد عملکرد بهتری را نسبت به رگرسیون لجستیک در شناسایی نوع اظهار نظر حسابرسان داشت.

## کاربرد داده کاوی در پیش بینی قیمت سهام

# dataacademy.ir

یکی از راه های کمک به سرمایه گذاران ، شرکت ها و سایر افراد و یا نهادهایی که به نوعی با بازار پول و سرمایه درگیر هستند ارائه الگوهای پیش بینی درباره دورنمای کلی شرکت است تا از این طریق سرمایه گذاران قادر به تصمیم گیری شوند.

هدف، پیش بینی ریسک تصمیم گیری است. با وجود آنکه پیش بینی ها معمولاً دقیق نیست و میزان خطای پیش بینی به سیستم مورد استفاده برای پیش بینی بستگی دارد، بنابراین با صرف منابع بیشتر برای پیش بینی می توان دقت آن را افزایش داد و برخی از زیان های ناشی از عدم اطمینان را حذف یا کاهش داد (چن و همکاران ، ۲۰۰۵)

قبل از وجود رایانه ها و استفاده از آن ها برای پیش بینی در بورس اوراق بهادار کار پیش بینی با روش های دیگری انجام می گرفت. سرمایه گذاران از شیوه های مختلف پیش بینی برای حداکثر کردن بازده و حداقل کردن ریسک استفاده می کرده اند. روش های پیش بینی که در بورس کاربرد داشتند به روش های سنتی پیش بینی معروف هستند. این روش ها عبارتند از تجزیه و تحلیل تکنیکی و تجزیه و تحلیل اساسی (متوسلی و کاشفی ، ۱۳۸۵)

با افزایش روز افزون قدرت رایانه ها استفاده از شبکه عصبی مصنوعی به شدت در حال افزایش است. شبکه های عصبی ابزار محاسباتی ساده ای برای آزمون داده ها و ایجاد مدل ها از ساختار داده ها است.

بروکس (۲۰۰۱) در مطالعاتی که بر روی بازار اوراق بهادار کانادا انجام داد ، داده های هجده سال را مورد بررسی قرار داد و به این نتیجه رسید که شبکه عصبی نتایج بهتری را در مقایسه با روش های رگرسیون بدست می آورد.



خالو زاده (۱۳۸۳) با استفاده از اطلاعات سری زمانی قیمت و بازده سهام چند شرکت در بازار سهام و اوراق بهادار تهران به پیش بینی قیمت سهام و نیز ارائه مدل بهینه پرداخت. پیش بینی براساس سه روش خطی و غیر خطی و شبکه عصبی با ساختار پیشنهادی انجام شد. برای مدل سازی خطی از مدل های ARIMA، برای مدل سازی غیر خطی از شبکه عصبی پرسپترون سه لایه با الگوریتم پس انتشار خطا استفاده شد. نتایج به دست آمده حاکی از آن است که استفاده از انواع مختلف روش های خطی به دلیل وجود ساختار آشوب گونه سری ها درست نیست و همچنین استفاده از روش های غیر خطی شبکه های عصبی به شکل متعارف نتایج قابل توجهی به دنبال ندارد. با این که مدل های شبکه عصبی تک خروجی نتایج بسیار خوبی را برای پیش بینی قیمت سهام داشته و خطای نسبی این مدل ها کم بوده است اما به دلیل حساسیت بسیار زیاد این مدل ها نسبت به شرایط اولیه و اختلاف اندک در مقادیر پیش بینی روز بعد موجب خطای زیاد شده است. بنابراین این شبکه برای پیش بینی دراز مدت کارایی لازمه را ندارد. بنابراین با استفاده از ساختار پیشنهادی می توان پیش بینی خوبی از قیمت سهام و بازده داشت.

تهرانی و عباسیون (۱۳۸۷) قابلیت شبکه های عصبی مصنوعی را در پیش بینی روند کوتاه مدت قیمت سهام در بورس اوراق بهادار تهران با استفاده از شاخص های تکنیکی مورد بررسی قرار داده اند. نتایج پژوهش نشان می دهد، شبکه عصبی از قابلیت پیش بینی روند کوتاه مدت قیمت سهام در بورس اوراق بهادار تهران چه در بازار صعودی و چه در بازار نزولی برخوردار است.

## dataacademy.ir

نمازی و کیامهر (۱۳۸۶) در مطالعه خود به بررسی پیش بینی رفتار بازده سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران و همچنین انجام عمل پیش بینی بازده با استفاده از شبکه های عصبی مصنوعی پرداخته اند. به منظور انجام عمل پیش بینی بازده، در مرحله اول روند گذشته سری زمانی مربوط به شرکت ها و همچنین سه متغیر از متغیرهای تحلیل تکنیکی (شاخص سهام، حجم سهام مبادله شده و آخرین نرخ سهام در روز (برای مدت ۵ سال)) مورد استفاده قرار گرفت و با تغییر پارامترهای شبکه عصبی مصنوعی مدل بهینه جهت پیش بینی بازده روزانه سهام هر شرکت طراحی گردید. در این پژوهش از شبکه با توابع یادگیری متفاوت استفاده گردیده است. نتایج تحقیق تأیید می نماید شبکه های عصبی مصنوعی توانایی پیش بینی بازده روزانه را با میزان خطای نسبتاً مناسبی دارند.

عادل آذر و امیر افسر اقدام به مدل سازی پیش بینی قیمت سهام با استفاده از رویکرد شبکه های عصبی فاز نموده و مدل طراحی شده را از لحاظ شش معیار عملکرد با روش ARIMA مقایسه شده است، نتایج تحقیق بیانگر این است که شبکه های عصبی فازی در تمام شش معیار ارزیابی عملکرد بر روش ARIMA برتری داشته است (آذر، ۱۳۸۵)

### نتیجه گیری

فناوری داده کاوی از قابلیت های پیش بینی و طبقه بندی فراوانی برخوردار است. بنابراین می تواند فرآیند تصمیم گیری در مسائل مالی را تسهیل کند. کاربرد روش های داده کاوی با توجه به مطالعات مرتبط و ماهیت آنها می تواند طیف گسترده ای را در برگیرد. همچنین، روش های داده کاوی به کار رفته در تحقیقات مختلف نیز شامل شبکه های عصبی مصنوعی، الگوریتم ژنتیک، درخت تصمیم گیری، ماشین بردار پشتیبان، نزدیکترین همسایگی هستند. براساس بررسی های صورت گرفته، روش



شبکه های عصبی مصنوعی تاکنون یکی از محبوب ترین روش های داده کاوی بوده است. اگر چه تعداد زیادی از تحقیقات در ارتباط با کاربرد روش های داده کاوی در امور مالی بوده اند. لیکن حوزه های مستعد دیگری به منظور گسترش دامنه تحقیقات وجود دارند. معرفی مدل های ترکیبی، بهبود مدل های کنونی، استخراج قواعدی جامع از شبکه های عصبی مصنوعی و بهسازی با ابزارهای داده کاوی می توانند گستره ای نوین و متفاوت از تحقیقات داده کاوی باشند.

## منابع

۱. دستگیر، محسن و شفیعی سردشت، مرتضی (۱۳۹۰). فناوری داده کاوی، رویکردی نوین در حوزه مالی. حسابرسی. تابستان ۹۰، دوره جدید، ص ۲۷-۶
۲. رهنمای رودپشتی، فریدون (۱۳۹۱). داده کاوی و کشف تقلب های مالی. فصلنامه علمی پژوهشی دانش حسابداری و حسابرسی مدیریت - انجمن حسابداری مدیریت ایران. پائیز ۹۱، شماره ۳، ص ۳۳-۱۷
۳. انواری رستمی، علی اصغر و خادمی زارع، حسن و علی حیدری بیوکی، طاهره و نشاط، نجمه (۱۳۹۰). تعیین محرک هزینه در سیستم های هزینه یابی بر مبنای فعالیت با استفاده از روش های داده کاوی و تحلیل عاملی. بررسی های حسابداری و حسابرسی. بهار ۹۰، شماره ۶۳، ص ۳۸-۲۱
۴. نجات، امیر رضا و آرش علی اکبری (۱۳۸۷). داده کاوی ره‌ی به سوی نامناخته ها. دو ماهنامه توسعه انسانی پلیس، شماره ۱۸
۵. مسلم زاده گتایی، آزاده (۱۳۹۴). روش های داده کاوی برای کشف تقلب در حسابرسی صورت های مالی. حسابدار رسمی. بهار ۹۴، شماره ۴۱، ص ۳۲-۲۲
۶. کیقابدی، امیر رضا و خادمی، وحید (۱۳۹۲). داده کاوی صورت های مالی جهت اعطای تسهیلات مالی. پژوهش های حسابداری مالی و حسابرسی. بهار ۹۲، شماره ۱۷، ص ۲۱۱-۱۷۹
۷. میر فخرالدینی، سید حیدر و بابایی میبیدی، حمید و مروتی شریف آبادی، علی (۱۳۹۲). پیش بینی مصرف انرژی ایران با استفاده از مدل ترکیبی الگوریتم ژنتیک - شبکه عصبی مصنوعی و مقایسه آن با الگوهای سنتی. سامانه مدیریت نشریات علمی دانشگاه تربیت مدرس. تابستان ۹۲، شماره ۲، ص ۲۲۲-۱۹۶
۸. باقر پور و لاشانی، محمد علی و باقری، مصطفی و خادم، حمید و حسینی پور، رضا (۱۳۹۱). بررسی عوامل مالی و غیر مالی مؤثر بر گریز مالیاتی با استفاده از تکنیک های داده کاوی: صنعت خودرو و ساخت قطعات. فصلنامه مطالعات تجربی حسابداری مالی. تابستان ۹۱، شماره ۳۴، ص ۱۲۸-۱۰۳
۹. ستایش، محمد حسین و ابراهیمی، فهیمه و سیف، سید مجتبی و ساریخانی، مهدی (۱۳۹۱). پیش بینی نوع اظهارنظر حسابرسان با رویکردی بر روش های داده کاوی. فصلنامه علمی و پژوهشی حسابداری مدیریت. زمستان ۹۱، شماره ۱۵، ص ۸۲-۶۹
۱۰. شفیعی، سعید و صبوری دیلمی، محمد حسن (۱۳۸۸). بررسی شیوه های مبارزه با پدیده پولشویی با تأکید بر راهکارهای مالیاتی. فصلنامه تخصصی مالیات، شماره ۵، ص ۱۶۹-۱۳۷



۱۱. آذر، عادل و امیرافسر ، (۱۳۸۵) ، مدل سازی پیش بینی قیمت سهام با رویکرد شبکه های عصبی فازی . فصلنامه پژوهشنامه بازرگانی ، ص ۳۳-۵۲ .
۱۲. دارابی، رویا و کریمی راسته کناری ، ربابه ، (۱۳۹۳) موانع موجود در تعیین قیمت سهام به روش شبکه عصبی مصنوعی (مطالعه موردی صنایع فلزی و کانی) . پژوهش های حسابداری مالی و حسابرسی، شماره ۲۲ ، ص ۲۹-۶۵
۱۳. متوسلی ، محمود و بیژن طالب کاشفی (۱۳۸۵). بررسی مقایسه ای توان شبکه های عصبی با ورود شاخص های تحلیلی تکنیکی برای پیش بینی قیمت سهام . نشریه نامه اقتصادی ، شماره ۲-۱-۶۰ ، ص ۵۷-۸۵
۱۴. خالوزاده ، حمید و خاکی، علی (۱۳۸۲) . ارزیابی روش های پیش بینی قیمت سهام و ارائه مدل غیر خطی بر اساس شبکه های عصبی . مجله تحقیقات اقتصادی، شماره ۶۳ ، ص ۴۳-۸۵
۱۵. تهرانی ، رضا و عباسیون ، حمید (۱۳۸۷) . کاربرد شبکه های عصبی مصنوعی در زمان بندی معاملات سهام با رویکرد تحلیل تکنیکی . فصلنامه پروژه های اقتصادی ، شماره ۱ ، ص ۱۷۷-۱۵۱
۱۶. نمازی، محمد و کیامهر، محمد مهدی (۱۳۸۶). پیش بینی بازده روزانه سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از شبکه های عصبی مصنوعی. تحقیقات مالی ، شماره ۲۴ ، ص ۱۱۵-۱۳۵

dataacademy.ir

17- Farzaneh A. Amani , Adam M. Fadlalla . (2017), "Data mining applications in accounting : A review of the literature and organizing framework" . International Journal of Accounting Information Systems , pp.32-58 .

18- Kirkos S. and Manolopoulos Y., (2004), "Data Mining in Finance and Accounting : A Review of Current Research Trends", Proceeding of the 1st International Conference on Enterprise System and Accounting (ICES Acc) , Thessaloniki , Greece , pp. 63-78 .

19- Ata Ali , Seyrek . H. (2009) . " The use of data mining techniques in detecting fraudulent financial statements: an application on manufacturing firms " . The Journal of Faculty of Economics and Administrative Sciences, vol 142 , pp. 157-170 .

20- Micci – Barrera , D. , & Ramachandran , s. (2004). "Improving tax administration with data mining"- Online – Available at : www. Spss.com

21- Wu , Rong – Shium and Ou , c.s and Chang She- I and Yen David C. (2012) . using Data Mining Technique to Enhance Tax Evasion Detection Performance " . Expert System with Applications, vol . 39, pp. 8769-8777.



22- Chen S & leung T,(2005) , " Regression Neural Network of Error correction In Foreign Exchange and Fore casting and Trading " , Elsevier , pp . 1049-1068

23- Han J, Kamber M. Data mining: Concepts and techniques. 2nd ed. San Francisco : Morgan Kaufmann ; 2006 .

24- Gaganis ch ; pasiouras F ; and M.Doumpos (2007) . " probabilistic Neural Networks for the Identification of Qualified Audit Opinions " Expert System with Applications , vol . 32 , pp. 114-124

25-Gaganis , ch ; pasiouras F ; Spathis , ch and C. Zopounidis (2007). A comparison of Nearest Neighbors , Discriminant and Logit Models for Auditing Decissions, Intelligent System in Accounting , Finance and (Management) , vol . 15. Pp. 23-40

dataacademy.ir