3rd International Conference on Modern Developments in Management, Economics and Accounting 22nd August 2019 TEHRAN - IRAN
سومین کنفرانس بین المللی تحولات نوین در مدیریت، اقتصاد و حسابداری
۳۱ مردادماه ۱۳۹۸ – تهران
کد مجوز وزارت علوم، تحقیقات و فناوری : ۸۱۱۶۹۰۲
کد مجوز پایگاه استنادی علوم جهان اسلام (ISC) : ۹۸۱۹۰-۵۷۰۱۰

پیش بینی حق الزحمه حسابرسی با استفاده ار الگوریتم های داده کاوی

دکترمحمد عکافان
رضاشایان
reza.shayan83@yahoo.com

چکیده
برای پیش بینی میزان حق الزحمه حسابرسی از الگوریتم های داده کاوی و نرم افزار اورنج استفاده نمودیم داده های ۵سال متوالی ۷۲۰شرکت را، از سال ۹۲ تا ۹۷ را براساس متغیرهای مدل آقای هاسمن جهت پیش بینی میزان حق الزحمه حسابرسی مورد آزمون داده کاوی قرار دادیم .درنهایت الگوریتم شبکه عصبی دقیقترین میزان حق الزحمه حسابرسی را پیش بینی نمود.

کلمات کلیدی: حق الزحمه حسابرسی، داده کاوی،شبکه عصبی

Archive of SID

www.SID.ir

دلیل اصلی وجود حرفه حسابرسی مستقل، وظیفه اعتباردهی است. اعتبار دادن به صورتهای مالی به ایجاد اطمینان از مطلوبیت و قابلیت اتکای صورتهای مالی است .وجود اطلاعات مالی شفاف و قابل اتکا که محصول یک سیستم گزارشگری جامع و مناسب می باشدازازکان اصلی ارزیابی وضعیت عملکرد یک شرکت و تصمیم گیری در مورد مبادله اوراق بهادار منتشره از سوی آن به حساب می آید. [۱]

حق الزحمه حسابرسی ، مطالعه تأثیر شهرت و اندازه حسابرس بر حق الزحمه حسابرسی در برنامه ریزی و اجرای مناسب و باکیفیت کار حسابرسی مالی مؤثر است. کیفیت پایین حسابرسی موجب کاهش اعتماد استفاده کنندگان صورتهای مالی می شود و این امر نه تنها منجر به ناکامی از دستیابی به اهداف حسابرسی می شود، بلکه موجب کاهش اعتبار فرایند حسابرسی در ابعاد کلان خواهد شد و مانع تخصیص بهینه سرمایه در بازار اوراق بهادار و افزایش هزینه سرمایه و تأمین مالی می شود [۲]در کشور ما نحوه تعیین حق الزحمه های حسابرسی به معضل تبدیل شده و آشفتگی قیمت گذاری خدمات حسابرسی، موجب شده هیچ مبنای مشخصی برای تعیین حق الزحمه حسابرسی مالی وجود نداشته باشد و بعضاً قضاوت های حرفه ای حسابرسان منجر به پیشنهادهای ضد و نقیضی گردد که تناسبی با یکدیگر ندارد . [۳]

حسابرسی

حسابرسی یعنی بازرسی جستجو گرانه مدارک حسابداری و سایر شواهد زیر بنای صورتهای مالی .حسابرسان از راه کسب آگاهی از سیستم کنترل داخلی و بازرسی مدارک،مشاهده داراییها، پرسش از منابع داخل و شواهد لازم را برای تعیین این که صورتهای مالی، تصویری ، خارج شرکت و اجرای سایر روشهای رسیدگی مطلوب و به نسبت کامل از وضعیت مالی شرکت و فعالیتهای آن در طول دوره مورد رسیدگی ارائه می کند یا خیر، گردآوری می کنند. یا اینکه حسابرسی فرایندی است منظم و با قاعده جهت جمع آوری و ارزیابی بیطرفانه شواهد درباره ادعاهای مربوط به فعالیتها و وقایع اقتصادی، به منظور تعیین درجه انطباق این ادعاها با معیارهای از پیش تعیین شده و گزارش به افراد ذینفع. حسابرسی در معنای اعم کلمه دارای سابقه های طولانی است. حسابرسی از زمانی متولد گشت که نیاز به کنترل نسبت به عملیات مختلف مالی به خصوص وجوه دولتی احساس شد. پس از دوره رنسانس و توسعه تجارت جهانی نیاز به داشتن حساب و کتاب بیشتر شد. لذا اصول دفترداری دوطرفه در قرن ۱۵ میلادی درایتالیا متولد شد. انقلاب صنعتی کارخانه های بزرگ را به وجود آورد و از تشکیل سرمایه های کوچک و اشتراک آنها، سرمایه های بزرگ ایجاد گردیده و نهایتاً منجر به ایجاد بورس و بازار اوراق بهادار شد [4]

سازمان بورس و اوراق بهادار، بُعد تازه ای به حسابداری، از جمله لزوم ارائه اطلاعات مالی به سهامداران و سایرین برای چگونگی اداره سرمایه ها، ارزیابی عملکرد و سنجش کارایی مدیران و . . . بخشید. [۳] گزارشهای برون سازمانی حسابداری را افراد شاغل در حرفه به تنهایی نمی توانستند به طور کامل و بی عیب ارایه دهند زیرا وجود رابطه استخدامی مستقیم، آنها را ناگزیر به پذیرش نظرات مدیران واحدهای اقتصادی در تهیه صورتهای مالی می کرد. حال آنکه صورت های مالی باید نیاز گروه های مختلف استفاده کننده از صورتهای مالی را برآورده می نمود. لذا چاره این کار انتخاب حسابرس مستقل و خبره توسط مجامع عمومی صاحبان سهام بود. تجمع حسابداران ورزیده و دارای تخصص کافی که عملیات مذکور را انجام می دادند در انجمن هایی که الزام به رعایت رفتار حرفه ای شرط لازم آن بود، نظام حرفه حسابرسی را پدید آورد. نخستین نمونه آن انجمن حسابداران خبره انگلستان بود که در کشورهای دیگر نیز الگو قرار گرفت[4]

به طور کلی، سه فرضیه عمده برای نیاز به حسابرسی مطرح است [5]

این فرضیات عبارتند از:

۱- فرضیه نمایندگی

این فرضیه، به نقشی اشاره دارد که حسابرسی در کاهش مسائل و خطرات نادیده گرفتن اصول اخلاقی ایفا می کند . طبق فرضیه نمایندگی، حسابرس در چار چوب روابط بین نماینده – مالک یخش لاینفک ساز وکاری قراردادی است که برای کنترل و نظارت بر هزینه های نمایندگی مدیر ایجاد می شود. به گفته جنسن و مک ولینگ [۵]هنگامی که نماینده (مدیر) نسبت به مالک از

Archive of SID
www.SID.ir

اطلاعات بیشتری راجع به عملیات داخلی شرکت برخوردار است، با فرض اینکه مدیر به دنبال منافع شخصی خود است، بیشتر از حد مطلوب از منابع شرکت استفاده کرده و برای انتقال ثروت از مالکان به نفع خود تلاش می کند. بنابراین سرمایه گذاران دارای رفتار عقلایی، چنین رفتارهایی را در قیمت گذاری اوراق بهادار شرکت لحاظ خواهند کرد و مبلغ کمتری را برای آن اوراق بهادار می پردازند [5]

حتی به عنوان آخرین گزینه ممکن است سرمایه خود را از شرکت خارج نمایند. شرکت خارج نمایند. در نتیجه سرمایه گذاران برای اجتناب از چنین پیشامدهایی سعی می کنند سازوکارهایی را برقرار سازند که بر رفتارها و انگیزه های فرصت طلبانه نظارت و کنترل داشته باشد. وجود حسابرس مستقل، از این سازوکارهاست. بدین ترتیب، در فرضیه نمایندگی، نقش حسابرس کاهش عدم تقارن اطلاعاتی و خطرات زیرپا گذاشتن اصول اخلاقی است. بدون وجود حسابرس ممکن است صورتهای مالی جانبدارانه و نادرست ارائه شوند [5]

۲ - فرضیه اعتمادسازی

بر اساس این فرضیه، حسابرس به کاهش تصمیمات نادرست در بازار سرمایه کمک می کند و سرمایه گذار سعی می کند حسابرس معتبری را برگزیند تا به بازار سرمایه نشان دهد عملیات شرکت شفاف و مناسب می باشد. سرمایه گذاران بالقوه نیز برای تصمیمات سرمایه گذاری خود از این علائم استفاده می کنند [6]

توجیه این فرضیه در مورد نحوه قیمت گذاری خدمات حسابرسی بدین صورت است که موسسات حسابرسی بزرگ و معتبر از صاحبکاران خود حق الزحمه بالاتری را طلب می کنند زیرا چنین موسساتی از کیفیت و اعتبار بالایی نزد سرمایه گذاران و جامعه برخوردارند و در نتیجه شفافیت و عملکرد مناسب تری را برای شرکت به ارمغان می آورند[7]

۳-فرضیه اعتباربخشی

بر اساس این فرضیه که نسبت به دو فرضیه قبلی، کمتر مورد توجه محققان و نظریه پردازان قرار گرفته است، ممکن است حرفه حسابرسی مدعی باشد در اعتبار بخشی به نتایج و صورتهای مالی هیچ مسئولیتی ندارد ولی در طرف دیگر، سرمایه گذاران معتقدند در صورتی که در نتیجه استفاده از صورتهای مالی گمراه کننده به آنها زیان وارد شود، حسابرسان باید تاوان آن را بپردازند و زیان آنها را جبران کنند. بنابراین ممکن است در مواردی که به سرمایه گذاران زیان وارد می شود، حسابرسان مورد بازخواست قرار گیرند. درنتیجه، حسابرسان باید در قیمت گذاری خدمات خود این ریسک را لحاظ کنند و هر چه نتایج عملیات شرکتی ضعیف تر باشد (سوددهی پایین یا زیان ده بودن) این ریسگ افزایش یافته و حق الزحمه حسابرسی نیز به تبع آن افزایش می یابد [8] در عمل یگ یا ترکیبی از این فرضیات متناسب با شرایط فرهنگی و محیط قانونی و قضایی به کار می رود و در یک کشور ممکن است یکی از آنها نسبت به بقیه، توجیه بهتری را ارائه نماید[5]

داده کاوی (Data Mining):

فناوری های نوین اطلاعاتی و ارتباطی، و همچنین تکنولوژی های پشتیبان تصمیم، با جمع آوری، ذخیره، ارزیابی، تفسیر و تحلیل، بازیابی و اشاعه اطلاعات و دانش به کاربران خاص، می توانند در اطلاع یابی به موقع، صحیح و مورد نیاز به افراد تاثیر زیادی داشته باشند. یکی از ابزارهای مورد استفاده در این فناوری ها، داده کاوی می باشد. داده کاوی شامل استفاده از ابزار های پیشرفته تحلیل داده به منظور کشف الگو های معتبر، از قبل نا شناخته و روابط در مجموعه داده های بزرگ است. این ابزار ها، مدلهای آماری، الگوریتم های ریاضی و متد های یادگیری ماشین (الگوریتم هایی که عملکرد خود را از طریق تجربه به صورت اتوماتیک بهبود می دهند) می باشد. داده کاوی فراتر از جمع آوری و مدیریت داده است، و شامل تجزیه و تحلیل و پیش گویی می شود. نام دیگر آن کشف دانش در پایگاه داده یا به اختصار KDD است[9]

Archive of SID
www.SID.ir

مفاهیم پایه در داده کاوی

در داده کاوی معمولا به کشف الگوهای مفید از میان داده ها اشاره می شود . منظور از الگوی مفید ، مدلی در داده ها است که ارتباط میان یک زیر مجموعه از داده ها را توصیف می کند و معتبر ، ساده ، قابل فهم و جدید است

تعریف داده کاوی

در متون آکادمیک تعاریف گوناگونی برای داده کاوی ارائه شده است . در برخی از این تعاریف داده کاوی در حد ابزاری که کاربران را قادر به ارتباط مستقیم با حجم عظیم داده ها می سازد معرفی گردیده است و در برخی دیگر ، تعاریف دقیقتر که درآنها به کاوش در داده ها توجه می شود موجود است . برخی از این تعاریف عبارتند از:

داده کاوی عبارت است از فرایند استخراج اطلاعات معتبر ، از پیش ناشناخته ، قابل فهم و قابل اعتماد از پایگاه داده های بزرگ و استفاده از آن در تصمیم گیری در فعالیت های تجاری مهم

تاریخچه داده کاوی

اخیرا داده کاوی موضوع بسیاری از مقالات ، کنفرانس ها و رساله ها ی عملی شده است ، اما این واژه تا اوایل دهه نود مفهومی نداشت وبه کار برده نمی شد .

در دهه شصت و پیش از آن زمینه هایی برای ایجاد سیستم ها ی جمع آوری و مدیریت داده ها ایجاد شد و تحقیقاتی در این زمینه انجام پذیرفت که منجر به معرفی و ایجاد سیستم های مدیریت پایگاه داده ها گردید .

ایجاد و توسعه مدلهای داده ای برای پایگاه سلسله مراتبی ، شبکه ای و بخصوص رابطه ای در دهه هفتاد ، منجر به معرفی مفاهیمی همچون شاخص گذاری و سازماندهی داده ها و در نهایت ایجاد زبان پرسش SQL در اوایل دهه هشتاد گردید تا کاربران بتوانند گزارشات و فرمهای اطلاعاتی مورد نظر خود را ، از این طریق ایجاد نمایند .

توسعه سیستم های پایگاهی پیشرفته در دهه هشتاد و ایجاد پایگاه های شی گرا ، کاربرد گرا و فعال باعث توسعه همه جانبه و کاربردی شدن این سیستم ها در سراسر جهان گردید . بدین ترتیب DBMS هایی همچون DB2، Oracle، Sybase ، ... ایجاد شدند و حجم زیادی از اطلاعات با استفاده از این سیستم ها مورد پردازش قرار گرفتند . شاید بتوان مهمترین جنبه در معرفی داده کاوی را مبحث کشف دانش از پایگاه داده ها (KDD) دانست بطوری که در بسیاری موارد DM و KDD بصورت مترادف مورد استفاده قرار می گیرند .

برای اولین بار مفهوم داده کاوی در کارگاه IJCAI در زمینه KDD توسط Shapir مطرح گردید . به دنبال آن در سالهای ۱۹۹۱ تا ۱۹۹۴ ، کارگاههای KDD مفاهیم جدیدی را در این شاخه از علم ارائه کردند بطوری که بسیاری از علوم و مفاهیم با آن مرتبط گردیدند.

برخی از کاربردهای داده کاوی در محیطهای واقعی عبارتند از:

خرده فروشی

از کاربردهای کلاسیک داده کاوی است که می توان به موارد زیر اشاره کرد:

تعیین الگوهای خرید مشتریان

تجزیه و تحلیل سبد خرید بازار

پیشگویی میزان خرید مشتریان از طریق پست (فروش الکترونیکی)

بانکداری:

پیش بینی الگوهای کلاهبرداری از طریق کارتهای اعتباری

تشخیص مشتریان ثابت

تعیین میزان استفاده از کارتهای اعتباری بر اساس گروههای اجتماعی

بیمه:

تجزیه و تحلیل دعاوی

Archive of SID
www.SID.ir

پیشگویی میزان خرید بیمه نامه های جدید توسط مشتریان

پزشکی:

تعیین نوع رفتار با بیماران و پیشگویی میزان موفقیت اعمال جراحی

تعیین میزان موفقیت روشهای درمانی در برخورد با بیماریهای سخت

مراحل فرایند کشف دانش از پایگاه داده ها

فرایند کشف دانش از پایگاه داده ها شامل پنج مرحله است که عبارتند از: انبارش داده ها

انتخاب داده ها

تبدیل داده ها

کاوش در داده ها

تفسیر نتیجه

همانگونه که مشاهده می شود داده کاوی یکی از مراحل این فرایند است که به عنوان بخش چهارم آن نقش مهمی در کشف دانش از داده ها ایفا می کند.

انبارش داده ها

وجود اطلاعات صحیح و منسجم یکی از ملزوماتی است که در داده کاوی به آن نیازمندیم . اشتباه و عدم وجود اطلاعات صحیح باعث نتیجه گیری غلط و در نتیجه اخذ تصمیمات ناصحیح در سازمانها می گردد و منتج به نتایج خطرناکی خواهد گردید که نمونه های آن کم نیستند.

اکثر سازمانها دچار یک خلا اطلاعاتی هستند . در اینگونه سازمانها معمولا سیستم های اطلاعاتی در طول زمان و با معماری و مدیریت های گوناگون ساخته شده اند ، به طوری که سازمان اطلاعاتی یکپارچه و مشخصی مشاهده نمی گردد . علاوه بر این برای فرایند داده کاوی به اطلاعات خلاصه و مهم در زمینه تصمیم گیریهای حیاتی نیازمندیم.

هدف از فرایند انبارش داده ها فراهم کردن یک محیط یکپارچه جهت پردازش اطلاعات است . در این فرایند ، اطلاعات تحلیلی و موجز در دوره های مناسب زمانی سازماندهی و ذخیره می شود تا بتوان از آنها در فرایند های تصمیم گیری که از ملزومات آن داده کاوی است ، استفاده شود . به طور کلی تعریف زیر برای انبار داده ها ارائه می گردد:

انبار داده ها ، مجموعه ای است موضوعی، مجتمع ، متغیر در زمان و پایدار از داده ها که به منظور پشتیبانی از فرایند مدیریت تصمیم گیری مورد استفاده قرار می گیرد.

انبارش داده ها خود موضوع مفصلی است که مقاله ها و رساله ها ی گوناگونی در مورد آن نگاشته شده اند . در این فصل به منظور آشنایی با این فرایند به آن اشاره ای شد.

انتخاب داده ها

انبار داده ها شامل انواع مختلف و گوناگونی از داده ها است که همه آنها در داده کاوی مورد نیاز نیستند . برای فرایند داده کاوی باید داده ها ی مورد نیاز انتخاب شوند . به عنوان مثال در یک پایگاه داده های مربوط به سیستم فروشگاهی ، اطلاعاتی در مورد خرید مشتریان ، خصوصیات آماری آنها ، تامین کنندگان ، خرید ، حسابداری و ... وجود دارند . برای تعیین نحوه چیدن قفسه ها تنها به داده ها یی در مورد خرید مشتریان و خصوصیات آماری آنها نیاز است . حتی در مواردی نیاز به کاوش در تمام محتویات پایگاه نیست بلکه ممکن است به منظور کاهش هزینه عملیات ، نمونه هایی از عناصر انتخاب و کاوش شوند .

Archive of SID
www.SID.ir

تبدیل داده ها

هنگامی که داده های مورد نیاز انتخاب شدند و داده ها ی مورد کاوش مشخص گردیدند ، معمولا به تبدیلات خاصی روی داده ها نیاز است . نوع تبدیل به عملیات و تکنیک داده کاوی مورد استفاده بستگی دارد : تبدیلاتی ساده همچون تبدیل نوع داده ای به نوع دیگر تا تبدیلات پیچیده تر همچون تعریف صفات جدید با انجام عملیاتهای ریاضی و منطقی روی صفات موجود .

کاوش در داده ها

داده های تبدیل شده با استفاده از تکنیکها و عملیاتهای داده کاوی مورد کاوش قرار می گیرند تا الگوهای مورد نظر کشف شوند .

تفسیر نتیجه

اطلاعات استخراج شده با توجه به هدف کاربر تجزیه و تحلیل و بهترین نتایج معین می گردند . هدف از این مرحله تنها ارائه نتیجه (بصورت منطقی و یا نموداری) نیست ، بلکه پالایش اطلاعات ارایه شده به کاربر نیز از اهداف مهم این مرحله است .

عملیاتهای داده کاوی

در داده کاوی چهار عمل اصلی انجام می شود که عبارتند از:

مدلسازی پیشگویی کننده

تقطیع پایگاه داده ها

تحلیل پیوند

تشخیص انحراف

از عملیاتهای اصلی مذکور ، یک یا بیش از یکی از آنها در پیاده سازی کاربرد های گوناگون داده کاوی استفاده می شوند . به عنوان مثال برای کاربرد های خرده فروشی معمولا از عملیات تقطیع و تحلیل پیوند استفاده می شود در حالی که برای تشخیص کلاهبرداری ، می توان از هر یک از چهار عملیات مذکور استفاده نمود . علاوه برا ین می توان از دنباله ای از عملیاتها برای یک منظور خاص استفاده کرد . مثلا برای شناسایی مشتریان ، ابتدا پایگاه تقطیع می شود و سپس مدلسازی پیشگویی کننده در قطعات ایجاد شده اعمال می گردد .

تکنیکها ، روشها و الگوریتمهای داده کاوی ، راههای پیاده سازی عملیاتهای داده کاوی هستند . اگر چه هر عملیات نقاط ضعف و قوت خود را دارد ، ابزارهای گوناگون داده کاوی عملیاتها را بر اساس معیارهای خاصی ، انتخاب می کنند . این معیارها عبارتند از:

تناسب با نوع داده های ورودی

شفافیت خروجی داده کاوی

مقاومت در مقابل اشتباه در مقادیر داده ها

میزان صحت خروجی

توانایی کار کردن با حجم بالای داده ها

مدلسازی پیشگویی کننده

مدلسازی پیشگویی کننده ، شبیه تجربه یادگیری انسان در به کار بردن مشاهدات برای ایجاد یک مدل از خصوصیات مهم پدیده ها است . در این روش از تعمیم دنیای واقعی و قابلیت تطبیق داده های جدید با یک قالب کلی ، استفاده می شود.

Archive of SID
www.SID.ir

در این مدل می توان با تحلیل یک پایگاه داده های موجود خصوصیات مجموعه های داده را تعیین کرد . این مدل با استفاده از روش یادگیری نظارت شده شامل دو فاز آموزش و آزمایش ایجاد شده است . در فاز آموزش با استفاده از نمونه های عظیمی از داده های سابقه ای مدلی ساخته می شود که به آن مجموعه آموزشی می گویند . در فاز آزمایش این مدل روی داده هایی که در مجموعه آموزشی قرار ندارند اعمال می شود تا صحت و خصو صیات آن تایید گردد.

از کاربردهای عمده این مدل می توان به مدیریت مشتریان تصویب اعتبار بازاریابی مستقیم در خرده فروشی و ... اشاره کرد

کاوی به فرایند نیم خودکار تجزیه و تحلیل پایگاه داده های بزرگ به منظور یافتن الگوهای مفید اطلاق می شود.

داده کاوی یعنی جستجو در یک پایگاه داده ها برای یافتن الگوهایی میان داده ها .

داده کاوی یعنی استخراج دانش کلان ، قابل استناد و جدید از پایگاه داده ها ی بزرگ .

داده کاوی یعنی تجزیه و تحلیل مجموعه داده های قابل مشاهده برای یافتن روابط مطمئن بین داده ها .

همانگونه که در تعاریف گوناگون داده کاوی مشاهده می شود ، تقریبا در تمامی تعاریف به مفاهیمی چون استخراج دانش ، تحلیل و یافتن الگوی بین داده ها اشاره شده است .

داده کاوی داده های مالی

در دنیای تجارت داده های مالی به عنوان سرمابه راهبردی مطرح هستند داده های مالی توسط موسساتی مانند بانکها،بورس اوراق بهادار،سازمان مالیاتی،و پایگاههای داده ویژه حسابرسان و حسابداران .عیره جمع آوری و نگهداری می شوند.روشهای داده کاوی در داده های مالی می توانددر حل مشکلات ظبقه بندی و پیش بینی و سهولت فرآیند تصمیم گیری کاربرد داشنه باشد.نمونه ای از مسایل ظیقه یندی مالی،شامل ورشکستگی شرکت ها ،نخمین ریسک اعتیاری،درماندگی مالی وپیش بینی عملکرد واحد تجاری می باشد. [۱۰]

اهمیت داده کاوی توسط بسیاری از سازمانهای حرفه ای تشخیص داده شده است.انجمن حسایداران رسمی آمریکا داده کاوی را به عنوان یکی از ۱۰فناوری برتر برای آینده معرفی نموده است.همچنین انجمن حسابرسان داخلی آمریکا این فناوری رادر فهرست یکی از چهار اولویت تحفیفاتی خود قرار داده است [10].

کاربرد احنمالی داده کاوی در حسایداری مالی منجر به استفاده از صورتهای مالی به جهت تعیین سودآوری، تظابق نسیت ها با میانگین صنعت و بررسی اثربخشی بک کسب و کار می گردد.همچنین در حسابداری مدیریت نیز کاربردهای بسیاری نطیرنجزیه و نحلیل بهای تمام شده و سود آوری بخشهای مختلف دارددر حسابرسی ابزارهای داده کاوی برای تجزیه و تحلیل داده های مرتبط با احتمال تقلب مدیریت می توانند مورد استفاده قرار بگیرند به این ترنیب داده کاوی در تمام جنبه های حسابداری سودمند خواهد بودزیرا حسابداری عبارت است ازجمع آوری ،تجزیه و تحلیل و استغاده از اطلاعات مالی و غیرمالی به منظورپردازشبرای اشخاص ذینفع است[11] شرکتهای تجاری می توانناز داده کاوی به منظورشناسایی فرصت های مالی و عملیاتی قابل توجه برای اظمینان از کارایی فرآبند زنجیره تامین بهره ببرند.

شبکه عصبی

شبکه عصبی یکی از الگوریتم های داده کاوی است.شبکه عصبی روشهای محاسیانی هستند که بر پایه انصال چندین واحد پردازشی ابجاد شده اند این شبکه از تعداد زیادی نرون تشکیل شده است ومججوعه ورودی را به خروجی ارتباط می دهددرواقع شیکه های عصبی از طبیعت الهام گرفته اند ودر آنها یک مجموعه پیچیده از نرونهای به هم متصل در کار یادگیری دخیل هستندبه همراه هر اتصال بک ارزش عددی که وزن نامیده می شود وجود دارد.هر نرون سیگنالهای از نرونهای متصل دریافت می نمایداگر جمع سیگنالهای ورودی از ورودی آستانه تجاوز یابد نرون از کار می افتدنرون ها بصورت گزوهی لایه بندی می شوند

یک شبکه عصبی شامل لایه ورودی ، لایه خروجی و لایه پنهان است.انواع مختلف شیکه عصبی تعداد متناوبی از لابه ها دارند ذر نقشه های خود سازمان یافته تنها یک ورودی و یک خروجی دارنددرحالی که شبکه عصبی پس از انتشار خطا یک یا چند لایه پنهان اضافی دارد برای اینکه به شیکه عصبی موجود توانایی آموختن داده شود.بعد از اینکه سیگنال از لایه اول شبکه به لابه

Archive of SID
www.SID.ir

پایینی می رود اطلاعات هر نرون روی سیگنال اثر می گذارد به روز آوری و اصلاح می شوداین روبه را به اصطلاحBP یا پس از انتشار خطا می گویند

در شبکه های پس از انتشار خطا خروجی با نتیجه دلخواه مقایسه می شودوخطاهای ایجاد شده در لایه قبلی در شیکه عصبی با میران سازی وزن های اتصال ها مشخص می شونداین فرآیند تا زمانی که یک نرخ اشتباه قابل قبول بدست آیدتکرار خواهد شدشبکه های عصبی پس از انتشار خطا به منظور پیش بینی و طبقه بندی مسایل از انواع مغتبر تلقی می شوند [10]

KNN الگوریتم نزدیکترین همسایه

بخاطر سادگی و موثر بودن در طبقه بندی متون بسیار کاربرد دارد.مبنای کار این الگوریتم به این صورت هست که اسنادی از مجموعه آموزش که مشابه به سند جاری است. انتخاب می شوند وکلاس سند جاری برابربا کلاسی می شود که اکثریت اسناد مشابه دارند.در روش رده بندی kتا همسایه نزدیکترk تا سند از مجموعه آموزش که بیشترین شباهت(براساس یک معیار شباهت تعریف شده)را به سند جاری دارد به عنوان همسایگان آن سند انتخاب می نماید.هنگامی که متن ناشناخته ای را الگوریتم دریافت می کند.فضای الگو را برای یافتن متون آموزشی که شبیه متن ناشناخته باشند جستجو می کند.دراینجا از فاصله اقلیدوسی به عنوان معیار شباهت استفاده شده است

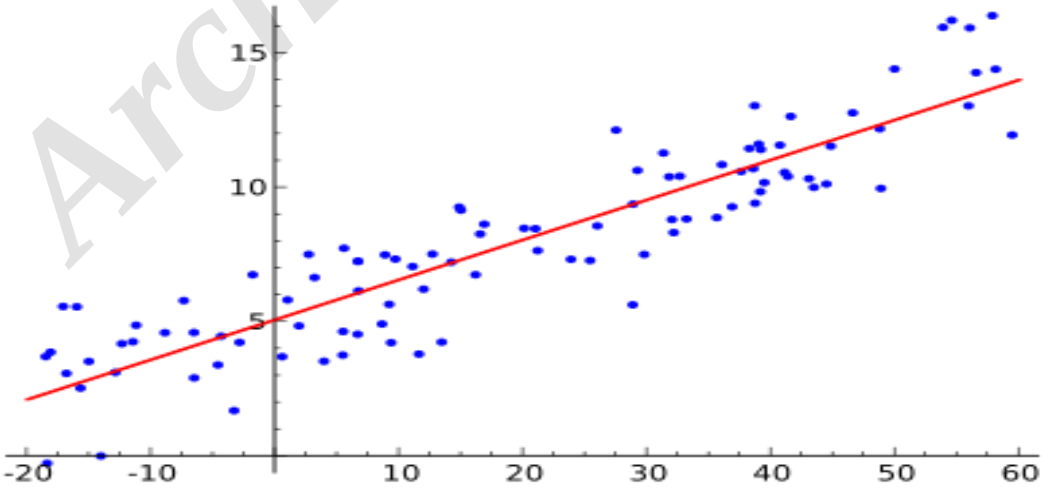
رگرسیون خطی (Linear Regression) :

الگوریتم رگرسیون خطی ، از نقاط داده برای یافتن بهترین خط مناسب جهت مدل سازی داده ها استفاده می کند. یک خط را می توان با معادله y = m * x + c نشان داد که در آن y متغیر وابسته و x متغیر مستقل است. برای پیدا کردن مقادیر m و c با استفاده از دیتاست داده شده، از اصول محاسبات ریاضی پایه (عمومی) استفاده می شود.

رگرسیون خطی دارای ۲ نوع است که عبارتند از :

رگرسیون خطی ساده که تنها از یک متغیر مستقل استفاده می کند.

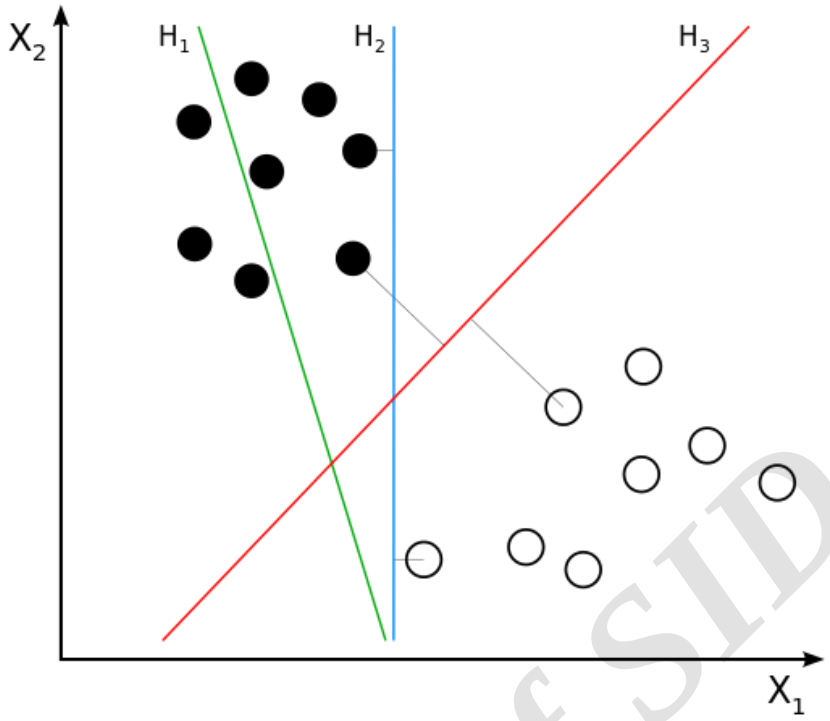
رگرسیون خطی چندگانه که متغیرهای مستقل مختلف برای آن تعریف می شود



ماشین بردار پشتیبان (SVM (Support Vector Machine :

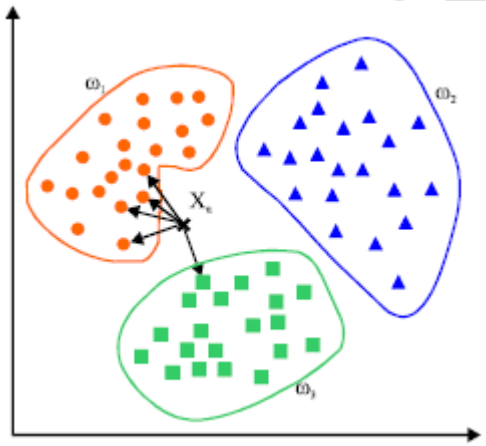
این مورد به نوعی از الگوریتم ها که جهت طبقه بندی (classification) استفاده میشوند تعلق دارد. الگوریتم نقاطی از داده را با استفاده از یک خط جدا خواهد کرد. این خط جدا کننده بطوری انتخاب شده است که نزدیک ترین خطی باشدکه بین دو دسته انتخاب شده اند

Archive of SID
www.SID.ir



الگوریتم k نزدیک ترین همسایه (KNN (K-Nearest Neighbors :

این یک الگوریتم ساده است که نقاط داده ای ناشناخته را با نزدیک ترین k همسایگان خود پیش بینی می کند. مقدار k در اینجا در مورد دقت پیش بینی ها ، یک عامل حیاتی است. با محاسبه فاصله از طریق توابع پایه ای از قبیل هندسه اقلیدسی، نزدیکترین فاصله را تعیین می کند



www.SID.ir

درخت تصمیم

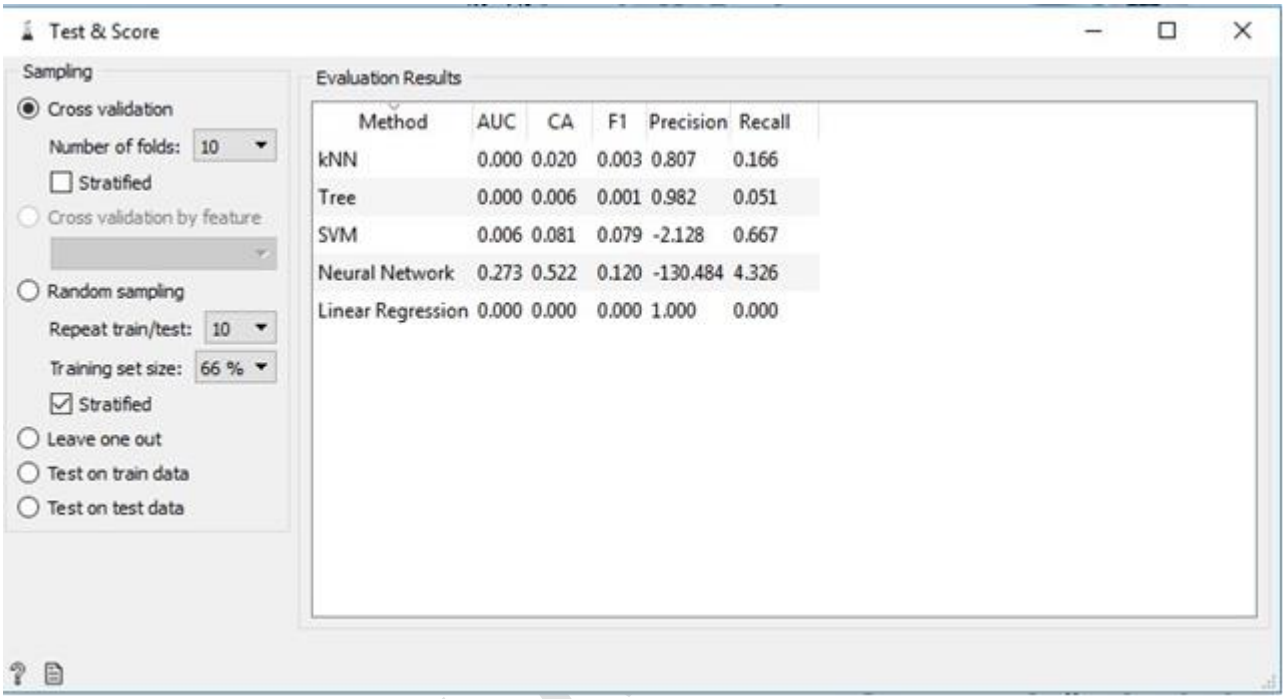
درخت تصمیم ساختاری مشابه فلوچارت ارایه می دهدکه درآن نودهای میانی یک تست بر روی ویژگی ها بوده و برگ ها و نودهای پایانی برچسب کلاس هدف را نشان می دهند. برای ساخت درخت،معیاری جهت انتخاب ویژگی در نودهای میانی در نظر گرفته می شود ویژگی منتخب باید بگونه ای باشد که عناصر رابه بهترین وجه به کلاس مجزا تقسیم کند[12]

Archive of SID

www.SID.ir

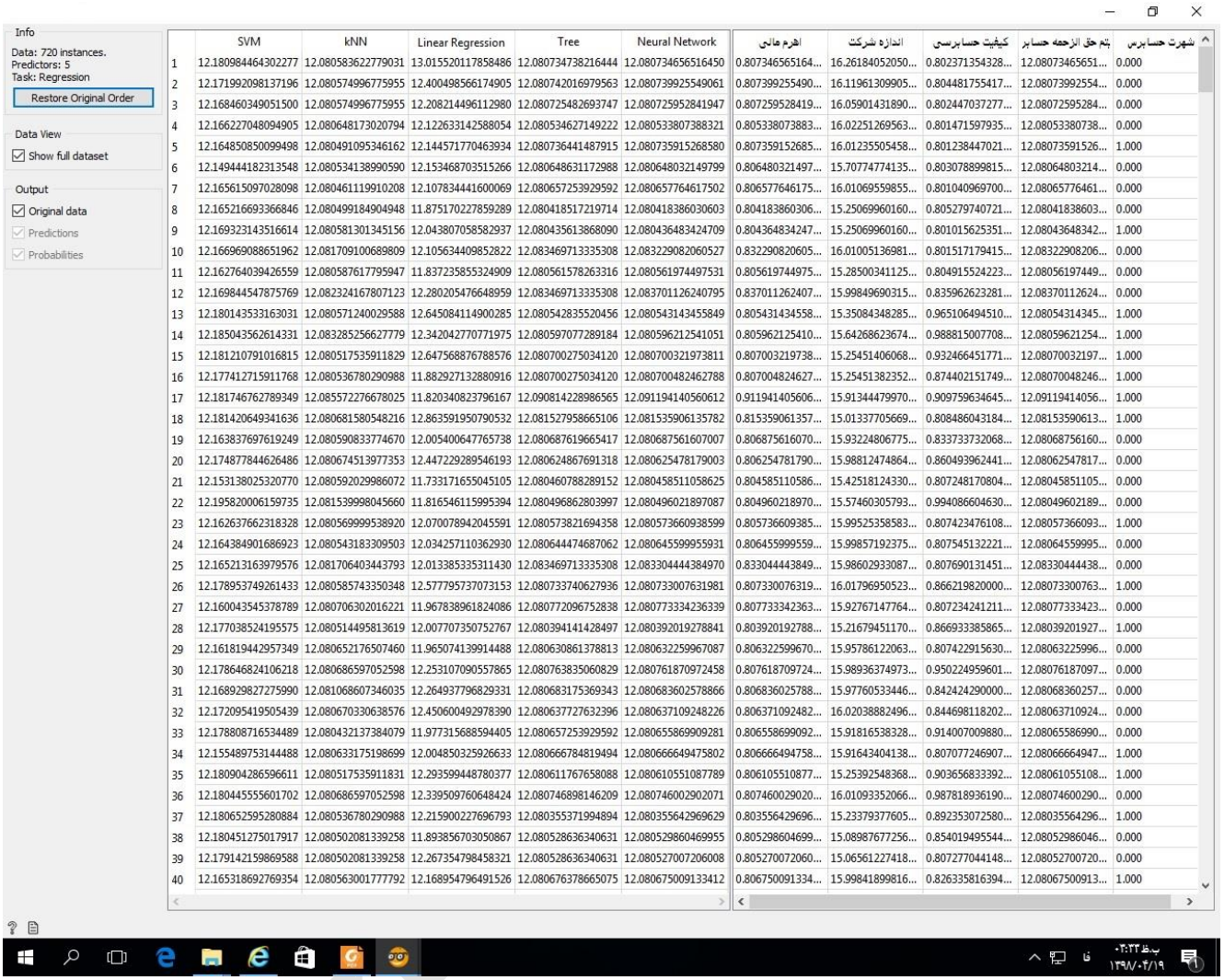
نتیجه گیری

درنهایت با استفاده از نرم افزار اورنج که یک جعبه ابزار متن باز تصویرسازی اطلاعات ،یادگیری ماشین و داده کاوی است .داده ها را مورد آزمون قرار دادیم و دربخش ارزیابی الگوریتم ها نرم افزار اورنج ، الگوریتم شبکه عصبی بهترین الگوریتم تشخیص داده شد



همچنین جهت پیش بینی حق الزحمه حسابرسی مقادیر پیش بینی شده توسط الگوریتم شبکه عصبی نزدیکترین مقادیر به مقادیر واقعی مبالغ حق الزحمه حسابرسی می باشد

www.SID.ir



	SVM	kNN	Linear Regression	Tree	Neural Network	اهرم مالی	اندازه شرکت	کیفیت حسابرسی	بهم حق الزحمه حسابر	شهرت حسابرس
1	12.180984464302277	12.080583622779031	13.015520117858486	12.080734738216444	12.080734656516450	0.807346565164...	16.26184052050...	0.802371354328...	12.08073465651...	0.000
2	12.171992098137196	12.080574996775955	12.400498566174905	12.080742016979563	12.080739925549061	0.807399255490...	16.11961309905...	0.804481755417...	12.08073992554...	0.000
3	12.168460349051500	12.080574996775955	12.208214496112980	12.080725482693747	12.080725952841947	0.807259528419...	16.05901431890...	0.802447037277...	12.08072595284...	0.000
4	12.166227048094905	12.080648173020794	12.122633142588054	12.080534627149222	12.080533807388321	0.805338073883...	16.02251269563...	0.801471597935...	12.08053380738...	0.000
5	12.164850850099498	12.080491095346162	12.144571770463934	12.080736441487915	12.080735915268580	0.807359152685...	16.01235505458...	0.801238447021...	12.08073591526...	1.000
6	12.149444182313548	12.080534138990590	12.153468703515266	12.080648631172988	12.080648032149799	0.806480321497...	15.70774774135...	0.803078898815...	12.08064803214...	0.000
7	12.165615097028098	12.080461119910208	12.107834441600069	12.080657253929592	12.080657764617502	0.806577646175...	16.01069559855...	0.801040969700...	12.08065776461...	0.000
8	12.165216693366846	12.080499184904948	11.875170227859289	12.080418517219714	12.080418386030603	0.804183860306...	15.25069960160...	0.805279740721...	12.08041838603...	0.000
9	12.169323143516614	12.080581301345156	12.043807058582937	12.080435613868090	12.080436483424709	0.804364834247...	15.25069960160...	0.801015625351...	12.08043648342...	1.000
10	12.166969088651962	12.081709100689809	12.105634409852822	12.083469713335308	12.083229082060527	0.832290820605...	16.01005136981...	0.801517179415...	12.08322908206...	0.000
11	12.162764039426559	12.080587617795947	11.837235855324909	12.080561578263316	12.080561974497531	0.805619744975...	15.28500341125...	0.804915524223...	12.08056197449...	0.000
12	12.169844547875769	12.082324167807123	12.280205476648959	12.083469713335308	12.083701126240795	0.837011262407...	15.99849690315...	0.835962623281...	12.08370112624...	0.000
13	12.180143533163031	12.080571240029588	12.645084114900285	12.080542835520456	12.080543143455849	0.805431434558...	15.35084348285...	0.965106494510...	12.08054314345...	1.000
14	12.185043562614331	12.083285256627779	12.342042770771975	12.080597077289184	12.080596212541051	0.805962125410...	15.64268623674...	0.988815007708...	12.08059621254...	1.000
15	12.181210791016815	12.080517535911829	12.647568876788576	12.080700275034120	12.080700321973811	0.807003219738...	15.25451406068...	0.932466451771...	12.08070032197...	1.000
16	12.177412715911768	12.080536780290988	11.882927132880916	12.080700275034120	12.080700482462788	0.807004824627...	15.25451382352...	0.874402151749...	12.08070048246...	1.000
17	12.181746762789349	12.085572276678025	11.820340823796167	12.090814228986565	12.091194140560612	0.911941405606...	15.91344479970...	0.909759634645...	12.09119414056...	1.000
18	12.181420649341636	12.080681580548216	12.863591950790532	12.081527958665106	12.081535906135782	0.815359061357...	15.01337705669...	0.808486043184...	12.08153590613...	1.000
19	12.163837697619249	12.080590833774670	12.005400647765738	12.080687619665417	12.080687561607007	0.806875616070...	15.93224806775...	0.833733732068...	12.08068756160...	0.000
20	12.174877844626486	12.080674513977353	12.447229289546193	12.080624867691318	12.080625478179003	0.806254781790...	15.98812474864...	0.860493962441...	12.08062547817...	0.000
21	12.153138025320770	12.080592029986072	11.733171655045105	12.080460788289152	12.080458511058625	0.804585110586...	15.42518124330...	0.807248170804...	12.08045851105...	0.000
22	12.195820006159735	12.081539998045660	11.816546115995394	12.080496862803997	12.080496021897087	0.804960218970...	15.57460305793...	0.994086604630...	12.08049602189...	0.000
23	12.162637662318328	12.080569999538920	12.070078942045591	12.080573821694358	12.080573660938599	0.805736609385...	15.99525358583...	0.807423476108...	12.08057366093...	1.000
24	12.164384901686923	12.080543183309503	12.034257110362930	12.080644474687062	12.080645599955931	0.806455999559...	15.99857192375...	0.807545132221...	12.08064559995...	0.000
25	12.165213163979576	12.081706403443793	12.013385335311430	12.083469713335308	12.083304444384970	0.833044443849...	15.98602933087...	0.807690131451...	12.08330444438...	0.000
26	12.178953749261433	12.080585743350348	12.577795737073153	12.080733740627936	12.080733007631981	0.807330076319...	16.01796950523...	0.866219820000...	12.08073300763...	1.000
27	12.160043545378789	12.080706302016221	11.967838961824086	12.080772096752838	12.080773334236339	0.807733342363...	15.92767147764...	0.807234241211...	12.08077333423...	0.000
28	12.177038524195575	12.080514495813619	12.007707350752767	12.080394141428497	12.080392019278841	0.803920192788...	15.21679451170...	0.866933385865...	12.08039201927...	1.000
29	12.161819442957349	12.080652176507460	11.965074139914488	12.080630861378813	12.080632259967087	0.806322599670...	15.95786122063...	0.807422915630...	12.08063225996...	0.000
30	12.178646824106218	12.080686597052598	12.253107090557865	12.080763835060829	12.080761870972458	0.807618709724...	15.98936374973...	0.950224959601...	12.08076187097...	0.000
31	12.168929827275990	12.081068607346035	12.264937796829331	12.080683175369343	12.080683602578866	0.806836025788...	15.97760533446...	0.842424290000...	12.08068360257...	0.000
32	12.172095419505439	12.080670330638576	12.450600492978390	12.080637727632396	12.080637109248226	0.806371092482...	16.02038882496...	0.844698118202...	12.08063710924...	0.000
33	12.178808716534489	12.080432137384079	11.977315688594405	12.080657253929592	12.080655869909281	0.806558699092...	15.91816538328...	0.914007009880...	12.08065586990...	0.000
34	12.155489753144488	12.080633175198699	12.004850325926633	12.080666784819494	12.080666649475802	0.806666494758...	15.91643404138...	0.807077246907...	12.08066664947...	1.000
35	12.180904286596611	12.080517535911831	12.293599448780377	12.080611767658088	12.080610551087789	0.806105510877...	15.25392548368...	0.903656833392...	12.08061055108...	1.000
36	12.180445555601702	12.080686597052598	12.339509760648424	12.080746898146209	12.080746002902071	0.807460029020...	16.01093352066...	0.987818936190...	12.08074600290...	0.000
37	12.180652595280884	12.080536780290988	12.215900227696793	12.080355371994894	12.080355642969629	0.803556429696...	15.23379377605...	0.892353072580...	12.08035564296...	1.000
38	12.180451275017917	12.080502081339258	11.893856703050867	12.080528636340631	12.080529860469955	0.805298604699...	15.08987677256...	0.854019495544...	12.08052986046...	0.000
39	12.179142159869588	12.080502081339258	12.267354798458321	12.080528636340631	12.080527007206008	0.805270072060...	15.06561227418...	0.807277044148...	12.08052700720...	0.000
40	12.165318692769354	12.080563001777792	12.168954796491526	12.080676378665075	12.080675009133412	0.806750091334...	15.99841899816...	0.826335816394...	12.08067500913...	1.000

www.SID.ir

Estimate Audit Fees Using Data Mining Algorithms

Abstract
To forecast the amount of audit fees, we used data mining algorithms and Orange software to test the data for 5 consecutive years of 720 companies, from 92 to 97, based on the variables of Mr. Hausman's model to predict the amount of audit fees we tested for data mining. Finally, the neural network algorithm predicts the most accurate audit fee.

Keywords: *Audit Fees, Data Mining, Neural Network*

Archive of SID

www.SID.ir

منابع

۱-(حساس یگانه، یحیی و جعفری، ولی اله) ۱۳۸۹(. "بررسی تأثیر چرخش مؤسسات حسابرسی برکیفیت گزارش حسابرسی شرکت های پذیرفته دشده در بورس اوراق بهادار تهران "، فصل نامه بورس اوراق بهادار، شماره ۹، صفحه ۲۵)

۲-(رجبی، روح الله) .۱۳۸۳(. عوامل موثر بر قدرت حرفه ای حسابرسی مستقل، پایان نامه دکتری، دانشگاه علامه طباطبایی، دانشکده مدیریت و حسابداری.

۳-تنانی، محسن و نیکبخت، محمد رضا) .۱۳۸۹(. آزمون عوامل موثر بر حق الزحمه حسابرسی صورت های مالی. پژوهش های حسابداری مالی. تابستان ۱۳۸۹. سال دوم، شماره دوم، شماره پیاپی) ۴(.
ص ص ۱۱۱ تا ۱۳۲.

۴-(سازمان حسابرسی.)۱۳۸۸(. اصول حسابرسی، جلد اول. ترجمه ارباب سلیمانی و نفری. چاپ بیست و ششم

۵)Lam, K.C. (1998). Risk Adjusted Audit Pricing: Theory and Empirical Evidence. Ph.D thesis. University of Toronto. Pp .5

۶)Jensen,M & W, Meckling .(1976). "Theory of the Firm : Managerial Behavior, Agency Costs, and Ownership Structure". Journal of Financial Economics. VOL 3.no4. oct. pp 305-360.

۷)Francis, J. and D, Simon . (1987). "A test of audit pricing in the small-client segment of the US market". The Accounting Review, Vol. 62 No.1, pp. 145-57

8)Simunic, D. (1980). "The pricing of audit services: theory and evidence". Journal of Accounting Research, Vol. 18 No.1, pp. 161-190

۹)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 .(7

10)Kirkos S. and Manolopoulos Y., (2004), 'Data Mining in Finance and Accounting: (۱۰ -۱۵ A Review of Current Research Trends', Proceedings of the 1st International Conference on Enterprise Systems and Accounting (ICESAcc), Thessaloniki,Greece, pp. 63-78

11)Rumsey et al. (2004) Rumsey N, Clarke A, White P, Wyn-Williams M, Garlick W. Altered body image: auditing the appearance related concerns of people with visible disfigurement. Journal of Advanced Nursing. 2004;48:443–453. doi: 10.1111/j.1365-2648.2004.03227.x

12) Han, J. and Kamber M.(2001). "Data Mining: Concepts and Techniques", San Diego Academic Press

Archive of SID
www.SID.ir