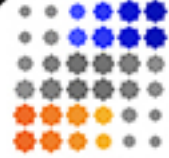




آکادمی داده

dataacademy.ir



داده کاوی و کاربرد آن در پزشکی

سولماز ایزدپناه

دستیار آموزشی، گروه فناوری اطلاعات، دانشگاه پیام نور صندوق پستی ۳۶۹۷-۱۹۳۹۵ تهران

Sizadpanah90@gmail.com

زهرا مکرمی

دانشجوی کارشناسی دانشگاه پیام نور

zahramokarrami1373@gmail.com

سعیده سالاری

دانشجوی کارشناسی دانشگاه پیام نور

Saidehsalari.s@gmail.com

رضا مظفریان

دانشجوی کارشناسی دانشگاه پیام نور

null@gmail.com

چکیده

امروزه در دانش پزشکی جمع آوری داده های فراوان در مورد بیماری های مختلف از اهمیت فراوانی برخوردار است. مراکز پزشکی با مقاصد گوناگونی به جمع آوری این داده ها می پردازند. تحقیق روی این داده ها و به دست آوردن نتایج و الگوهای مفید در رابطه با بیماری ها، یکی از اهداف استفاده از این داده ها است. حجم زیاد این داده ها و سردرگمی حاصل از آن مشکلی است که مانع رسیدن به نتایج قابل توجه می شود. بنابراین از داده کاوی برای غلبه بر این مشکل و به دست آوردن روابط مفید بین عوامل خطر زا در بیماری ها استفاده می شود. این مقاله به معرفی داده کاوی و کاربرد آن در صنعت پزشکی (پیش بینی بیماری) با استفاده از الگوریتم های داده کاوی به همراه نرم افزارهای مرتبط با آن پرداخته است.

واژگان کلیدی: داده کاوی، درخت تصمیم، پیش بینی بیماری، دیابت



مقدمه

ذخیره سازی داده ها به کار رفت ، حجم اطلاعات ذخیره شده در آن پس از حدود ۲۰ سال دو برابر شد و همزمان با پیشرفت فناوری اطلاعات ، حجم داده ها در پایگاه داده ها هر دو سال یک بار ، دو برابر شد و همچنان با سرعت بیش تری نسبت به گذشته حجم اطلاعات ذخیره شده بیش تر و بیش تر می شود. با وجود شبکه جهانی وب ، سیستم های یکپارچه اطلاعاتی ، سیستم های یکپارچه بانکی ، تجارت الکترونیکی و ... لحظه به لحظه به حجم داده ها در پایگاه داده ها اضافه شده و باعث به وجود آمدن انبارهای (توده های) عظیمی از داده ها شده است ، به طوری که ضرورت کشف و استخراج سریع و دقیق دانش از این پایگاه داده ها را بیش از پیش نمایان کرده است. شدت رقابت ها در عرصه های علمی ، اجتماعی ، اقتصادی ، سیاسی و نظامی نیز اهمیت سرعت یا زمان دسترسی به اطلاعات را دو چندان کرده است. بنا براین نیاز به طراحی سیستم هایی که قادر به اکتشاف سریع اطلاعات مورد علاقه کاربران با تاکید بر به حداقل مداخله انسانی باشند از یک سو و روی آوردن به روش های تحلیل متناسب با حجم داده های حجیم از سوی دیگر ، به خوبی احساس می شود. در حال حاضر ، داده کاوی مهم ترین فناوری برای بهره وری موثر ، صحیح و سریع از داده های حجیم است و اهمیت آن رو به فزونی است فراگیری ماشین داده می باشد ، الگوشناسی ، هوش مصنوعی ، علم کامپیوتر ، داده کاوی پل ارتباطی میان علم آمار باشد. داده کاوی فرآیندی پیچیده جهت شناسایی الگوها و مدل های صحیح، جدید و به صورت بالقوه مفید، در حجم وسیعی از داده می باشد، به طریقی که این الگو ها و مدلها برای انسانها قابل درک باشند. داده کاوی به صورت یک محصول قابل خریداری نمی باشد، بلکه یک رشته علمی و فرآیندی است که بایستی به صورت یک پروژه پیاده سازی شود داده ها اغلب حجیم می باشند و به تنهایی قابل استفاده نیستند، اما دانش نهفته در داده ها قابل استفاده می باشد. بنابراین بهره گیری از قدرت فرآیند داده کاوی جهت شناسایی الگوها و مدلها و نیز ارتباط عناصر مختلف در پایگاه داده جهت کشف دانش نهفته در داده ها و نهایتا تبدیل داده به اطلاعات، روز به روز ضروری تر می شود. 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.

تعاریف داده کاوی

به معنای Mine نگاهی به ترجمه لغوی داده کاوی به ما در درک بهتر این واژه کمک می کند. واژه لاتین به معنی داده بر جستجویی Data استخراج از منابع نهفته و با ارزش زمین اطلاق می شود. ادغام این کلمه با عمیق از داده های قابل دسترس با حجم زیاد برای یافتن اطلاعات مفید که قبلا نهفته بودند، تاکید دارد داده کای دارای تعریف های مختلفی است این تعریف ها به مقدار زیادی به پیش زمینه ها و نقطه نظرهای افراد بستگی دارد. هر نویسنده، محقق و کاربر با توجه به پیش زمینه ها و نقطه نظر های افراد بستگی دارد. هر نویسنده ، محقق و کاربر با توجه به دیدگاه و نوع نگرش خود تعریف های مختلفی از داده کاوی ارائه کرده اند به عنوان مثال می توان به چند تعریف داده کاوی که در ادامه آمده است اشاره کرد:

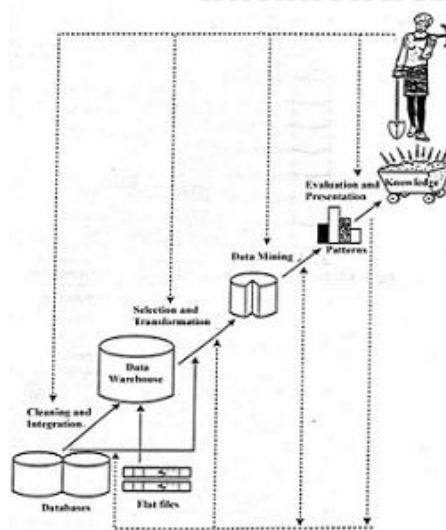
-داده کاوی استخراج اطلاعات مفهومی، ناشناخته و به صورت بالقوه مفید از پایگاه داده می باشد.

-داده کاوی علم استخراج اطلاعات مفید از پایگاه های داده یا مجموعه داده ای می باشد.

-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.

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 ۱۹۹۶، اولین شماره مجله کشف دانش از پایگاه داده ها منتشر شد.

امروزه کنفرانس های مختلفی در این زمینه در سراسر دنیا برگزار می شود. داده کاوی حاصل تحول تدریجی در طول تاریخ بوده و از اوایل دهه 90 همزمان با همه گیر شدن استفاده از پایگاه های داده ای به عنوان یک علم مطرح شده است.



شکل ۱-۲: داده کاوی به عنوان یک مرحله از فرآیند کشف دانش

کاربردهای داده کاوی

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

تجزیه و تحلیل سبد خرید بازار
پیشگویی میزان خرید مشتریان از طریق
پست (فروش الکترونیکی)
بیمه:

خرده فروشی:

تعیین الگوهای خرید مشتریان



پیش بینی الگوهای کلاهبرداری از طریق
 کارتهای اعتباری
 تشخیص مشتریان ثابت
 تعیین میزان استفاده از کارتهای اعتباری
 براساس گروههای اجتماعی

تجزیه و تحلیل دعاوی
 پیشگویی میزان خرید بیمه نامه های
 جدید توسط مشتریان
 پزشکی:

حوزه کاربردی فضاوسفرهای فضایی:

حجم بسیار زیادی از اطلاعات
 نوین بسیار بالا
 ارزش بسیار زیاد دانش قابل استخراج
 پردازش اطلاعات جمع آوری شده از فضا
 پردازش اطلاعات مربوط به سفینه های
 فضایی
 ارایه دانش مفید برای اتخاذ تصمیم
 نهایی جهت پرتاب یا عدم پرتاب یک
 سفینه به فضا

تعیین نوع رفتار با بیماران و پیشگویی
 میزان موفقیت اعمال جراحی
 تعیین میزان موفقیت روشهای درمانی
 دربرخورد با بیمارپهای سخت
 تشخیص بیماریها براساس انواع اطلاعات
 (تصاویر پزشکی، مشخصات بیمار
 احتمالی)
 تشخیص ناهنجاری هایی که توسط
 انسان به سختی قابل تشخیص خواهند
 بود

بانکداری:

درخت تصمیم گیری

درخت تصمیم درختی است که در آن نمونه ها را به نحوی دسته بندی می کند که از ریشه به سمت پائین رشد میکنند و در نهایت به گره های برگ میرسند. هر گره داخلی یا غیر برگ با یک ویژگی مشخص می شود، این ویژگی سوالی را در رابطه با مثال ورودی مطرح می کند. در هر گره داخلی به تعداد جواب های ممکن با این سوال شاخه وجود دارد که هر یک با مقدار آن جواب مشخص می شود. برگ های این درخت با یک کلاس و یا یک دسته از جواب ها مشخص میشوند. علت نامگذاری آن با درخت تصمیم این است که این درخت فرآیند تصمیم گیری برای تعیین دسته یک مثال ورودی را نشان می دهد.

داده کاوی در عرصه پزشکی

تحول در صنعت سلامت به واسطه این هدف واحد که « چگونه سازمان های سلامت هزینه ها را کاهش و کیفیت را افزایش دهند و همچنان رقابتی باقی بمانند؟ » به پیش می رود و این مقوله همواره یک چالش بزرگ محسوب می شود. بهبود کیفیت در صنعت سلامت را می توان به واسطه ی نیروهای محرکی که بر آن تاثیرگذار است بهتر تعریف نمود و از جمله این نیروهای محرک داده های سلامت است؛ به عبارت دیگر در هر نوع برنامه ی بهبود کیفیت متمرکز بر بیمار، داده ها قلب آن برنامه به حساب می آید. داده ها در عصر امروزی یعنی عصر اطلاعات، عمده ترین دارایی برای سازمان های سلامت بوده و موفقیت سازمان های سلامت در گروهی جمع آوری، ذخیره و تحلیل آن هاست. با این وجود، جمع آوری و ذخیره ی میزان زیادی از داده ها می تواند یک نوع اتلاف محسوب شود؛ مگر این که داده ها به شکل سودمند استفاده شده و تبدیل به یک منبع مالی برای سازمان گردد. برای تبدیل این ارزش بالقوه به اطلاعات استراتژیک، بسیاری از سازمان



ها به داده کاوی روی آورده اند؛ چرا که به واسطه ی داده کاوی امکان کشف روابط، روندها و الگوهای مخفی بین داده ها و دستیابی به دانش نوین در زمینه ی چالش های آشکار و نهان سازمان میسر خواهد شد .

مفهوم کشف دانش از داده ها بیش از یک دهه است که در محیط های مالی-تجاری در حال استفاده می باشد و در علوم مدیریت ارتباطات، مهندسی، وب کاوی، تحلیل جرایم و پزشکی جای خود را باز کرده است. اگرچه کشف دانش باهدف شناسایی اختلاس مالی وارد عرصه ی سلامت شد، اما به تدریج در حوزه ی بالینی نیز مورد استفاده قرار گرفت. این مهم ناشی از تغییر سریع هوشیاری نسبت به اطلاعات در حوزه ی سلامت است

نمونه های از کاربرد داده کاوی در پزشکی

- داده کاوی در تشخیص های غیرتهاجمی

- داده کاوی در تعیین نوع درمان

- داده کاوی در شناسایی عوارض جانبی داروها

- داده کاوی در پرونده ی الکترونیک سلامت

- داده کاوی در رتبه بندی بیمارستان ها

- داده کاوی در بهره برداری از خدمات سلامت

مقایسه الگوریتمهای هوشمند در شناسایی بیماری دیابت

بیماری دیابت یکی از شایعترین بیماریهای حاضر دنیا شناخته شده است که علیرغم گستردگی شیوع این بیماری هنوز روشی به منظور ریشه کن کردن و از بین بردن آن در دنیا شناخته نشده است هرچند که روشهای مختلفی جهت تشخیص و کنترل آن در حال حاضر مورد استفاده قرار میگیرد. از جمله عوارضی که به دنبال مبتلا شدن افراد به این بیماری گریبانگیر آنها خواهد شد می توان به گرفتگی عروق قلبی و در نوع پیشرفته آن به نابینایی، قطع اعضای بدن، اختلالات فکری و غیره اشاره نمود. مشکل عمدهای که در رابطه با بیماری دیابت وجود دارد عدم تشخیص به موقع و یا به طور کلی ضعف در تشخیص این بیماری است که این ضعف نیز بهدلیل عدم انتخاب الگوی مناسب توسط پزشک و یا عدم استفاده مناسب از الگوهای استاندارد است. بنابراین پیاده سازی روشی که بتواند هر فرد را در تشخیص صحیح ابتلا یا عدم ابتلا به این بیماری یاری رساند میتواند گام مهمی در جهت پیشگیری و کنترل این بیماری بهخصوص در مراحل ابتدایی آن باشد.

نتیجه گیری

در این مقاله مقوله داده کاوی و تکنیک های آن مورد بررسی قرار گرفت. همچنین از میان روش های دسته بندی و پیش بینی ، درخت تصمیم به عنوان یکی از ابزار های قوی و متداول در داده کاوی که درک، پیاده سازی و کاربرد آسان داشته و از نظر محاسباتی ارزان می باشد مورد بحث قرار گرفت . با توجه به اهمیت و حساسیت داده کاوی در پزشکی و همچنین نیاز مبرم این صنعت به حرکت از پزشکی سنتی به سمت پزشکی مبتنی بر شواهد لذا در این مقاله کاربرد داده کاوی در عرصه سلامت



چهارمین کنفرانس ملی

فناوری اطلاعات، کامپیوتر و مخابرات

4th National Conference On Information Technology, Computer & Telecommunication

July 13 2017

۲۲ تیر ۱۳۹۶



مورد بررسی قرار گرفت. درخت، نتایج به دست آمده روی مجموعه بیماران دیابتی نشان می دهد که در میان روش های طبقه بندی تصمیم نتیجه ی بهتری را به دست می آورد و الگوریتم نزدیک ترین همسایه نیز دقت بالاتری نسبت به روش طبقه بندی دارد اما با توجه به مطالعه سایر مقالات در زمینه تشخیص دیابت نتایج حاکی از آن است که هرگز نمی توان الگوریتمی را به عنوان الگوریتم بهینه معرفی کرد. در نتیجه برای هر کاربرد با توجه به مجموعه داده مورد استفاده می توان الگوریتمی را به عنوان الگوریتم بهینه معرفی نمود.

dataacademy.ir



مراجع

- ۱] اشکاند راد، ایمان، داده کاوی،
- ۲] مقدسی، حمید-حسینی، اعظم السادات-اسدی، فرخنده-جهانبخش، مریم (۱۳۹۰) داده کاوی و کاربرد آن در سلامت
- ۳] صادق زاده، مهدی-عشیر، امین-فروتن، فراز-سیفی پور، مرضیه، (۱۳۹۰) مقایسه الگوریتم های هوشمند در شناسایی بیماری دیابت
- ۴] عزیزی، (۱۳۸۵) یادگیری درخت های تصمیم

