

کاربرد داده کاوی در مدیریت داده های بزرگ در حوزه اطلاعات سلامت با استفاده از

الگوریتم CRISP-DM

مجید رهنمایی ذکاوت

دانشجوی دکترای علم اطلاعات و دانش شناسی دانشگاه تهران

چکیده:

هدف: داده کاوی یکی از راه های تحلیل و استفاده از اطلاعات است که با استفاده از روش های تخصصی آماری و منطقی به تحلیل داده های بزرگ مقیاس می پردازد و به سازمانها در اخذ تصمیم های کلان کمک می کند. پژوهش حاضر با هدف شناسایی مهم ترین کاربردهای داده کاوی در حوزه مدیریت داده های بزرگ در حوزه اطلاعات سلامت می باشد.

روش شناسی: در پژوهش حاضر جهت شناسایی کاربردهای داده کاوی علاوه بر مطالعات کتابخانه ای با تحلیل و بررسی کاربردهای داده کاوی در حوزه های بانکداری، مدیریت ریسک، مدیریت تجارت و آمار، کاربردهای داده کاوی در حوزه سلامت و خدمات درمانی نیز ارائه می شود.

یافته ها: یافته ها نشان داد داده کاوی در بخش های مختلفی از حوزه تحلیل و مدیریت داده ها از جمله جست و جوی اطلاعات، فراهم آوری، مدیریت و خدمات درمانی، جمع آوری اطلاعات، مراجعین به مراکز درمانی، حفظ وفاداری مراجعان، تشخیص و پیش بینی بیماری ها، تحلیل دارو درمانی بیماران و رتبه بندی بیمارستانها، مراکز درمانی کاربرد دارد، همچنین مشخص شد که از داده های بخش سلامت و داده های هزینه ای جهت بهتر شدن عملکرد مراکز سلامت و بیشتر شدن هزینه-سودمندی مراکز درمانی میتوان بهره برد.

در کل هدف از بکارگیری داده کاوی در حوزه سلامت، جایگزینی آن با متخصصین و محققین این حوزه نیست. بلکه بعنوان مکمل آنها در جهت افزایش دقت، کارایی و کاهش هزینه ها و زمان رسیدن به نتایج و دانش مورد نظر می باشد.

نتیجه گیری: کاربردهای داده کاوی در حوزه مدیریت داده در حوزه اطلاعات سلامت و مراکز درمانی بیشتر مختص به بخش های تشخیص بیماری، و مدیریت هزینه ها است. اما می تواند در سایر بخش ها نیز از داده کاوی برای تحلیل داده های موجود در راستای دستیابی هر چه بهتر به اهداف حوزه سلامت استفاده کرد.

حوزه پزشکی و سلامت از بخش های مهم در جوامع امروزی می باشد. استخراج دانش از میان حجم انبوه داده های مرتبط با سوابق بیماری و پرونده های پزشکی افراد با استفاده از فرآیند داده کاوی می تواند منجر به شناسایی قوانین حاکم بر ایجاد، رشد و تسری بیماری ها گردیده و اطلاعات ارزشمندی را به منظور شناسایی علل رخداد بیماری ها، تشخیص، پیش بینی و درمان بیماری ها با توجه به عوامل محیطی حاکم در اختیار متخصصین و دست اندرکاران حوزه سلامت قرار دهد. نتیجه این مساله به معنای افزایش عمر و ایجاد آرامش برای افراد جامعه است.

کلیدواژگان:

داده کاوی، اطلاعات سلامت، کشف دانش، داده های بزرگ

مقدمه

پیشرفته، انگیزه ای شده است برای اینکه محققین پزشکی از تکنیک های داده کاوی برای استخراج دانش از این پایگاه داده ها استفاده کنند. [1] امروزه با کمک الگوریتم های داده کاوی می توان الگوهای نهان در میان داده ها را استخراج نمود و با ساخت مدل مربوط به پایگاه های داده، سیستم کمک تصمیم طراحی کرد که در زمینه تصمیم گیری به پزشکان و کارکنان دانشی حوزه سلامت کمک کند.

تحول در صنعت سلامت به واسطه ی این هدف واحد که «چگونه سازمان های سلامت هزینه ها را کاهش و کیفیت را افزایش دهند و هم چنان رقابتی باقی بمانند؟» به پیش می رود و این مقوله همواره یک چالش بزرگ محسوب می شود. بهبود کیفیت در صنعت سلامت ر می توان به واسطه ی نیروهای محرکی که بر آن تأثیر گذار است بهتر تعریف نمود و از جمله این نیروهای محرک، داده های سلامت است؛ به عبارت دیگر در هر نوع برنامه بهبود کیفیت متمرکز بر بیمار، داده ها قلب آن برنامه به حساب می آید [4]. داده ها در عصر امروزی یعنی عصر اطلاعات، عمده ترین دارایی برای سازمان های سلامت بوده و موفقیت سازمان های سلامت در گروی جمع آور، ذخیره و تحلیل آن ها است. [6]

داده کاوی فرآیندی پیچیده جهت شناسایی الگوها و مدل های صحیح و بصورت بالقوه مفید در حجم وسیعی از داده می باشد، به طریقی که این الگوها و مدل ها 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

از سال 1950 به بعد که رایانه، در تحلیل و ذخیره سازی داده ها به کار رفت؛ حجم اطلاعات ذخیره شده بر روی رایانه ها پس از دهه های مختلف رو به افزایش هست و همزمان با پیشرفت فناوری اطلاعات، حجم داده ها در پایگاه داده ها همچنان با سرعت بیش تری نسبت به گذشته افزایش می یابد و چندین برابر شده است اگر در مقطعی از زمان صحبت ذخیره سازی و نگهداری اطلاعات در حجم گیگا بایت بوده است در حال حاضر صحبت از نگهداری داده های بزرگ در حجم ترابایت و حتی پتا بایت می باشد.

یکی از مراکز مهمی که حجم زیادی از داده ها در آن نگهداری و آرشیو می شود مربوط به سیستم های اطلاعات سلامت در حوزه بهداشت و درمان است. با وجود شبکه جهانی وب، سیستم های یکپارچه اطلاعاتی، تجارت الکترونیکی و موارد مشابه لحظه به لحظه به حجم داده ها در پایگاه داده های حوزه سلامت اضافه شده است و باعث بوجود آمدن انبارهای بزرگی از داده هادر این حوزه شده است. به طوری که ضرورت کشف و استخراج سریع و دقیق دانش از این پایگاه داده ها را بیشتر از قبل نمایان کرده است.

تعریف مساله

اخیرا استفاده از داده کاوی در حوزه پزشکی مورد توجه زیادی قرار گرفته است. معمولا محیط پزشکی غنی از اطلاعات است در حالی که نیازمند دانش می باشد. تکنیک های داده کاوی می توانند برای کشف دانش مورد نیاز استفاده شوند. رشد سریع پایگاه های داده پزشکی در کشورهای

استفاده نمیشود به این دلیل نیاز به تبدیل اطلاعات به دانش در بسیاری از زمینه‌ها آشکار گردیده است.

پیشینه و سوابق تحقیق

در مورد کاربردهای داده کاوی در حوزه اطلاعات سلامت و خدمات درمانی پژوهشهایی انجام گرفته است که در زیر به برخی از آنها اشاره میشود.

Thangavel و همکاران (به نقل از Canlas) از طریق الگوریتم های خوشه بندی به تحلیل بیماران مبتلا به سرطان گردن رحم پرداختند و نتایج پیش گیری کننده تری را نسبت به عقاید پزشکی کنونی پیدا کردند [18]. همچنین Gorunescu (به نقل از Canlas) نشان داد که داده کاوی میتواند با ادغام CAD (تشخیص با کمک کامپیوتر) و اندوسکوپی اولتراسونوگرافی در شناسایی غیر تهاجمی سرطان به عنوان یک شیوه ی جدید به کار رود [18].

سپهری و همکاران در تحقیقی در به کارگیری داده کاوی بر روی داده های پزشکی دستاوردهای حیاتی و اثرگذاری را در انتخاب نوع درمان مناسب و نجات جان انسان ها به ارمغان آورده است . به عنوان مثال در بیمارستان شهید هاشمی نژاد تهران برای تعیین نوع درمان سنگ حالب از راهکار داده کاوی استفاده می شود . در این بیمارستان یک الگوریتم درختی وجود دارد که پزشک بر اساس آن درمانی را که میزان موفقیت بالاتری برای بیمار دارد انتخاب می کند و بررسی ها نشان داده است که میزان عملکرد این الگوریتم 77 درصد بوده و بسیار بهتر از عملکرد الگوریتم بیمارستانی و مدل ذهنی پزشکان است [19].

Obenshain در تحقیقی در سال 2004 به بررسی داده کاوی و کاربرد آن در رتبه بندی بیمارستانها پرداخته است. رتبه بندی بیمارستان ها

شود. داده ها اغلب حجیم می باشند و به تنهایی قابل استفاده نیستند اما دانش نهفته در داده ها قابل استفاده می باشد.

هدف تحقیق

در این تحقیق به بیان کاربردهای داده کاوی در حوزه سلامت و درمانی می پردازیم همچنین با توجه به ماهیت اطلاعات سلامت به چالش های موجود در داده کاوی در حوزه سلامت و درمان اشاره داریم و با همین رویکرد به انتخاب الگوریتم مناسب داده کاوی مانند پیشگویانه در حوزه های مدیریت درمانی و سیستم های اطلاعات سلامت پرداخته و بعنوان جنبه نوآورانه تحقیق به ارزیابی پیشنهاداتی در استفاده از الگوریتم های داده کاوی مانند CRISP-DM در حوزه سلامت با توجه به چالش ها یی مانند حفظ محرمانگی داده ها و ناهمگونی داده های پزشکی و درمانی می پردازیم.

دلایل استفاده از داده کاوی

1. توسعه تکنولوژیهای ذخیره و بازیابی اطلاعات امکانی است برای محقق شدن داده کاوی
2. افزایش روزافزون حجم اطلاعات ذخیره شده
3. تنوع بسیار زیاد در اطلاعات موجود
4. بانکهای اطلاعاتی
5. فایل‌های چندرسانه‌ای (تصاویر متحرک، فایل‌های صوتی)
6. اطلاعات متنی و فاقد ساختار
7. آرشيوهای اطلاعاتی، به دلیل حجم بسیار زیاد، غالباً به مقبره های اطلاعات تبدیل میشوند. علیرغم هزینه های سنگین در بخش تکنولوژی اطلاعات، بسیاری از تصمیمها همچنان در فقر اطلاعاتی اتخاذ میگردند. از قابلیت‌های بالقوه اطلاعات ذخیره شده

زمانی به استخراج قوانین و روابطی پردازند که با اعمال آنها بتوان سطح اثربخشی و کفایت بیماران را بهبود بخشید. نکته قابل ذکر در خصوص این تحقیق این است که در آن به فاکتورهای آزمایشگاهی توجه نشده و تنها مواردی که به طور اتوماتیک توسط دستگاه دیالیز ثبت و نگهداری می شوند مورد بررسی قرار گرفته اند. این فاکتورها شامل شدت جریان خون، میزان کاهش وزن، طول جلسه دیالیز، نوع دیالیزور، فشارخروجی، کاهش آب بدن، هموگلوبین و حجم پلاسما می باشد [21].

نمونه دیگری از تحقیقات این زمینه در سال 2007 در آمریکا به عنوان پایان نامه کارشناسی ارشد توسط Bries انجام شده است. در این تحقیق مجموعه داده پیچیده ای که مربوط به بیش از 3000 بیمار دیالیزی در بیمارستان های متعدد در سطح کشور آمریکا می باشند مورد استفاده قرار گرفته و شامل اطلاعات اجرای دیالیز، اطلاعات دموگرافیک قابل اجرا، داروها و اطلاعات جلسات دیالیز بیمار می شود. در مراحل مختلف تحقیق ابتدا الگوریتم های حداکثر رسانی انتظارات EM و K-Means برای خوشه بندی داده ها مورد استفاده قرار گرفته اند که پس از بررسی هایی که بر روی نتایج این دو الگوریتم انجام شده است تعداد بهینه خوشه ها براساس الگوریتم EM برابر 6 خوشه در نظر گرفته شده است. سپس در ادامه محقق تهیه مدل بر اساس سری های زمانی می پردازد تا نتیجه تست هموگلوبین بیماران دیالیزی را بر استاس مدل ارائه شده پیش بینی نماید. بدین منظور الگوریتم انتخاب ویژگی با استفاده از الگوریتم ژنتیک مورد استفاده قرار گرفته و درخت تصمیم 48 لایه شبکه های عصبی برای ارزیابی زیرمجموعه های بالقوه انتخاب شده اند. در انتهای

و برنامه های بهداشتی می تواند بر مبنای اطلاعات گزارش شده توسط ارائه دهندگان مراقبت باشد، بنابراین گزارش دهی استاندارد برای مقایسه ی معنی دار بیمارستان ها و رتبه بندی آن ها بسیار مهم است. از جمله شیوه های استاندارد کردن این گزارشات استفاده از تکنیک های داده کاوی میباشد، به عنوان مثال اگر کدهای ICD (کدهای عوامل خطر) اختصاص یافته به پرونده های بیمار با تکنیک های داده کاوی مانند خوشه بندی و ارتباطات همراه شود، می تواند به ایجاد گزارشاتییانجامد که با تعیین نرخ واقعی میزان ناخوشی ها، مرگ و میر و سایر شاخص های کیفیت در رتبه بندی بیمارستان ها مورد استفاده واقع شود [20].

طاها صمد سلطانی در تحقیقی در سال 1394 به کاربرد داده کاوی در بیماران مبتلا به بیماری آسم پرداخته است در این تحقیق با انتخاب جامعه پژوهش از بیماران بیمارستان امام خمینی و دسیخ دانشوری به تحلیل داده های گرد آوری شده از پژوهش ذوالنوری و همکاران با استفاده از تکنیک های داده کاوی در نرم افزار اکسل پرداخته است.

مکفی (1392) در پژوهشی تحت عنوان "کاربرد داده کاوی در پیش بینی هزینه های درمانی" به دنبال یافتن راه حلی برای تخمین هزینه های بیمارستانی با استفاده از اطلاعات موجود و الگوریتمهای داده کاوی بود. به طور کلی در این طرح پیشنهاد یک راه حل دقیق و علمی برای تعیین تعرفه های پزشکی را مطرح نمود.

Bellazzi در تحقیق دیگری در سال 2005 بر روی داده های 5800 جلسه دیالیز مربوط به 43 بیمار در طول 19 ماه در ایتالیا انجام شده است، محققین تلاش می کنند با استفاده از تکنیک های Temporal Data mining و نیز تکنیکهای Rule Discovery در سری های

خدمات اینترنت بانک ارائه انجام crisp-dm کردند. مدل آنها بر اساس استاندارد گرفته بود و داده های مورد نیاز از پایگاه داده مشتریان اینترنت بانک سینا استخراج شده بود.

اصلی ترین دلیلی که باعث شده داده کاوی کانون توجهات در علوم پزشکی قرار بگیرد، مسئله در دسترس بودن حجم وسیعی از داده‌ها و نیاز شدید به اینکه از این داده‌ها، اطلاعات و دانش سودمند استخراج می‌کنند. داده کاوی عبارت از اقتباس یا استخراج دانش از مجموعه‌ای از داده‌ها است.

بررسی روشهای داده کاوی

از داده کاوی برای شناسایی روابط و الگوهای نو، صحیح، قابل فهم و به صورت بالقوه مفید در درون داده‌ها، با استفاده از ترکیب مجموعه داده‌ها، و استخراج الگوهای پیچیده برای انسان استفاده می‌شود [8]. داده کاوی، گامی مهم در کشف و استخراج دانش می باشد؛ این اصطلاح به معنای کاوش مجموعه داده های بزرگ برای استخراج الگوهای ناشناخته بین داده ها است [9]. از آنجایی که کشف روابط بین داده ها در روش های سنتی آماری بسیار مشکل است [10]، لذا کاربرد داده کاوی به سرعت در بخشهای وسیعی از قبیل سازمان های ارائه خدمات بهداشتی، پیش بینی مالی و پیش بینی هوا گسترش یافته است [11].

روش های داده کاوی به دو دسته توصیفی و پیش بینانه تقسیم می شوند. روش های توصیفی خواص عمومی داده ها را آشکار می کنند. هدف از توصیف یافتن الگوهایی از داده هاست که برای انسان قابل تفسیر باشد و ظایف پیش بینانه به منظور پیش بینی رفتار های آینده داده ها استفاده می شوند. منظور از پیش بینی بکار گیری چند متغیر برای

دقت مدل پیشنهادی بررسی شده و پیشنهادهای برای توسعه مدل ارائه شده است [23].

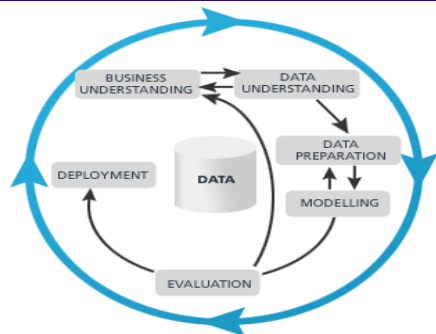
داده کاوی در همه رشته های که با آمار سروکار دارند کاربرد دارد بنابراین در زیر تعدادی کاربردهای داده کاوی در سایر رشته ها بیان میگردد:

- منصوری و همکاران (1386) در پژوهشی تحت عنوان "داده کاوی و کاربرد آن در تصمیم گیری ها" به تشریح تکنولوژی داده کاوی در حوزه هوش محاسباتی و متدولوژی استاندارد CRISP Data Mining پرداخته و یک طبقه بندی از کاربردهای این تکنولوژی در حوزه تصمیم گیری های کسب و کار ارائه دادند. در ادامه یک مطالعه موردی بر روی پایگاه داده مشتریان یکی از بانکها با استفاده از داده کاوی را ارائه کردند.

- قره خانی و ابوالقاسمی (1389) در پژوهشی تحت عنوان "کاربردهای داده کاوی در صنعت بیمه" به بیان کاربردهای داده کاوی در صنعت بیمه پرداختند آنها ضمن بررسی اجمالی داده کاوی، به نقش آن در کشف دانش موجود در پایگاههای داده و بهبود امور مرتبط با صنعت بیمه پرداختند.

- آذر و همکاران (1388)، در مقالهای با عنوان "طراحی مدل انتخاب نیروی انسانی با رویکرد داده کاوی" به شناسایی شاخصهای نیروی انسانی که بر عملکرد یا ارتقا مؤثر هستند، پرداخته و در نهایت، مدلی برای انتخاب متغیرهای تأثیرگذار، متغیر هدف و الگوریتمهای مناسب، از بین قواعد به دست آمده، ارائه کرده اند.

- رادفر و همکاران (1393) در پژوهشی تحت عنوان "طبقه بندی مشتریان اینترنت بانک" با کمک الگوریتمهای داده کاوی یک مدل مناسب برای طبقه بندی مشتریان بر مبنای بهره گیری از



مراحل انجام فرآیند داده کاوی

منبع (Chapman et al., 2000)

ابزارهای داده کاوی روش‌ها و رفتارهای آینده و امکان ایجاد کسب و کار فعال و تصمیمات دانش محور را پیش بینی می‌کنند. تکنیک‌های داده کاوی می‌توانند به سرعت روی پلت فرم‌های نرم افزار و سخت افزار موجود به منظور افزایش ارزش منابع اطلاعاتی موجود، اجرا شوند.

روش تحقیق

روش پژوهش حاضر به صورت توصیفی تحلیلی بود به گونه‌ای که ابتدا به بررسی کاربردهای داده کاوری در حوزه‌های بانکداری، مدیریت ریسک، هتلداری، مدیریت، تجارت، آمار و صنعت پرداخته شد، سپس براساس رسالت حوزه سلامت به کاربردهای داده کاوی در حوزه مدیریت داده‌ها در حوزه اطلاعات سلامت و خدمات درمانی‌ها پرداختیم.

تکنیک Crisp داده کاوی

داده کاوی به عنوان مهمترین کاربرد data warehouse یا انبارهای داده شناخته می‌شود. به وسیله داده کاوی داده‌های موجود مورد تحلیل قرار می‌گیرند تا روندهای احتمالی، ارتباط‌های

پیش بینی مقادیر آینده یا ناشناخته سایر متغیرهای مورد علاقه است.

بطور کلی گام‌های اجرای داده کاوی در تمام منابع یکسان نیستند ولی آنچه تقریباً در بین همه آنها بصورت مشترک وجود دارد، کشف دانش را دارای مراحل زیر می‌دانند:

1. پاکسازی داده‌ها (از بین بردن نویز و ناسازگاری داده‌ها)
2. یکپارچه سازی داده‌ها (ترکیب چندین منبع داده)
3. انتخاب داده‌ها (بازیابی داده‌های مرتبط با آنالیز از پایگاه داده)
4. تبدیل کردن داده‌ها (تبدیل داده‌ها به شکلی که مناسب داده کاوی باشند مثل خلاصه سازی و همسان سازی)
5. داده کاوی (فرآیند اصلی که روال‌های هوشمند برای استخراج الگوها از داده‌ها به کار گرفته می‌شوند)
6. ارزیابی الگو (مشخص کردن الگوهای صحیح و مورد نظر توسط معیارهای اندازه گیری)
7. ارایه دانش (نمایش بصری تکنیک‌های بازنمایی دانش برای ارایه دانش کشف شده به کاربر)

داده کاوی فقط یک ابزار است و نه یک عصای جادویی. داده کاوی به این معنی نیست که شما راحت به کنار بنشینیم و ابزارهای داده کاوی همه کار را انجام دهد.

در این مقاله به بررسی تکنیک‌های همچون شبکه های عصبی مصنوعی و درخت تصمیم گیری و کلاس بندی و خوشه بندی در حوزه داده های سلامت می پردازیم.

تکنیک های داده کاوی در حوزه اطلاعات سلامت

در این مقاله تکنیک‌های داده کاوی مانند شبکه های عصبی مصنوعی (Neural Net) و درخت تصمیم گیری و کلاس بندی و خوشه بندی با رویکرد کاربرد در حوزه درمانی و سلامت بررسی می شوند.

تکنیک کلاس بندی، یک نمونه داده را به یکی از کلاس های از قبل تعریف شده انتساب می کند . در این روش یک مجموعه از قوانین کلاس بندی بر مبنای ویژگی های مجموعه داده ی آموزشی تولید می شوند که می توانند برای کلاس بندی نمونه داده ی جدید استفاده شوند بتوان مثال قوانین کلاس بندی درباره یک بیماری می توانند از روی علائم و مشخصه های بیمارهای شناخته شده ی فعلی کشف و برای تشخیص آن بیماری در بیمارهای جدید با توجه به علائم بیماری آنها استفاده شوند. تشخیص پزشکی یکی از مهمترین کاربردهای کلاس بندی می باشد.

تکنیک خوشه بندی، شناسایی و گروه بندی مجموعه ای از داده ها یا اشیاء به زیر کلاس هایی با مفهوم خوشه، بطوری که یک خوشه شامل یک سری داده های مشابه می باشد که همانند یک گروه رفتار می کنند . خوشه بندی همان کلاس بندی است با این تفاوت که کلاس ها از پیش تعریف شده و معین نمی باشند و عمل گروه بندی داده ها بدون نظارت انجام می گیرد.

غیر محسوس و الگوهای مخفی از بین انبوه داده ها شناسایی شوند.

ما در این تحقیق سعی داریم تا با استفاده از تکنیک های داده کاوی به بررسی بازیابی اطلاعات و دانش از سیستم های اطلاعات سلامت بپردازیم . بدین منظور نیاز به تشریح و به کارگیری فرآیند داده کاوی داریم، یکی از متدولوژی های انجام پروژه های داده کاوی متدولوژی Crisp است که در این تحقیق جهت مراحل کشف دانش از این تکنیک استفاده شده است.

فرآیند داده کاوی با الگوریتم Crisp (Cross Industry Standard Process) شامل مراحل ذیل است:

Business Understanding درک کسب و کار

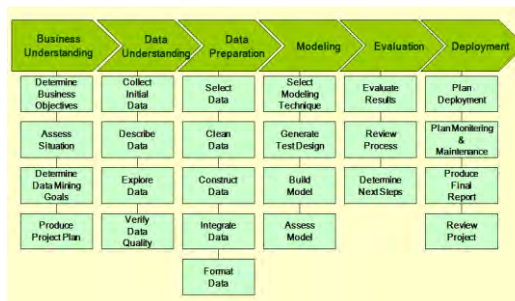
Data Understanding فهم داده

Data Preparation آماده سازی داده ها

Modeling مدل سازی

Evaluation ارزیابی

فازهای متدولوژی CRISP



فازها و فعالیت های متدولوژی CRISP DM

منبع (Chapman et al., 2000)

کلاهبرداری و سوء استفاده ه ها را شناسایی نمایند؛ سازمان های بهداشتی درمانی تصمیمات مربوط به مدیریت ارتباط با مراجعه کنندگان را اتخاذ نمایند؛ پزشکان، درمانهای موثر و بهترین شیوه های درمانی را شناسایی نمایند، و بیماران، خدمات مراقبتی بهتر و رضایت بخش تر را دریافت نمایند [8]. شناخت روشهای پیشگیرانه و درمانی در بیماری های مزمن از قبیل آسم از طریق داده کاوی میسر می شود و این روش نقش بارزی در شناخت الگوها دارد [15].

□ کاربرد داده کاوی در مدیریت خدمات درمانی

و سلامت

داده کاوی با گسترش شناسایی و پیدا کردن بهتر حالات بیماری مزمن و بیماران پرخطر، برنامه ریزی اقدامات مناسب و کاهش تعداد درخواستها و مراجعین به مراکز درمانی می تواند به مدیریت خدمات بهداشتی درمانی و پزشکی کمک موثری نماید [8]. بعنوان مثال با دسته بندی بیماران با توجه به ویژگی های دموگرافی و وضعیت پزشکی آنها می توان گروه هایی که منابع درمانی بیشتری را مصرف می کنند شناسایی و برنامه هایی برای آموزش آنها تدوین و وضعیت پزشکی آنها را مدیریت کرد.

□ کاربرد داده کاوی در پیش بینی و تشخیص

بیماری

➤ انتخاب درمان نامناسب برای بیمار، علاوه بر هدر دادن وقت و پول، می تواند اثرات زیانباری را نیز به همراه داشته باشد. حتی در برخی از موارد این انتخاب نادرست می تواند منجر به مرگ بیمار شود. همه ی پزشکان در تمام زمینه های شغلی خود مهارت ندارند [16].

✓ بعنوان مثال یک مجموعه از بیماری های جدید می توانند بر مبنای شباهت در علائم شان به چندین دسته گروه بندی شده و علائم مشترک بیماری های یک دسته، برای توصیف بیماری های آن دسته استفاده شوند.

❖ درخت تصمیم، درختی است که در آن نمونه ها را به نحوی دسته بندی می کنند که از ریشه به سمت پایین رشد می کنند و در نهایت به گره های برگ می رسد. هر گره داخلی یا غیر برگ با یک ویژگی مشخص می شود؛ این ویژگی سوالی را در رابطه با مثال ورودی مطرح می کند.

✓ برای مثال درخت تصمیم گیری می تواند عوامل مؤثر بر میزان بقای پیوند کلیه را تعیین کند، همچنین استفاده از الگوریتم DRG (کدهای تشخیصی وابسته) در بازپرداخت هزینه های بیمه پزشکی سالمندان در ایالات متحده، مثال کلاسیکی از روش شناسایی این تکنیک محسوب می شود.

❖ شبکه های عصبی مصنوعی که معمولا به عنوان شبکه های عصبی نام برده می شوند یک الگوی ریاضی مبنی بر سیستم زیستی است.

✓ برای مثال شبکه های عصبی می تواند مشخص کند که چه نوع بیماری هایی احتمال دارد با یک بیماری همراه شود و با تحلیل تصاویر، نوارهای قلبی و سایر مشاهدات بالینی به تشخیص، درمان و تولید دارو کمک می کند.

کاربردهای داده کاوی در حوزه سلامت و خدمات درمانی

کاربردهای داده کاوی می توانند در تمام بخشهای صنعت مراقبت بهداشتی درمانی به صورت جدی مثر ثمر باشند. برای مثال، سازمان های ارائه دهنده خدمات بیمه بهداشتی درمانی می توانند

تجویزی به بیماران می توان با داده کاوی نسبت به تعیین بازار هدف برای شرکت های داروسازی اقدام نمود و همچنین ارزانتترین و موثر ترین روش درمان برای یک بیماری را شناسایی کرد.

□ کاربرد داده کاوی در اثر بخشی معالجات

➤ داده کاوی با مقایسه ی عوامل، علائم و سیر درمان یک بیماری تحلیلی از سیر عملیات موثر و کارا ارائه می کند و می تواند برای ارزیابی میزان اثربخشی معالجات پزشکی استفاده شود. بعنوان مثال مقایسه نتایج گروه بندی بیماران درمان شده با رده های دارویی متفاوت برای یک بیماری خاص می تواند بهترین روش معالجه با کمترین هزینه و عرضه را تعیین کند [8]. و یا داده کاوی می تواند به شناسایی معالجات موفق برای یک بیماری خاص کمک کند.

➤ در حوزه درمانی و سلامت حجم انبومی از اطلاعات ذخیره شده در پایگاه داده ها وجود دارد که با داده کاوی می توان کاربردهای زیاد و متنوعی در حوزه بهداشت و درمان همچون پیش بینی بیماری ها و کنترل روند درمان جهت کیفیت سرویس دهی به بیماران و غربال گری دوره ای بیماران و کنترل بالینی علایم آنها در راستای مدیریت سلامت کلان جامعه انجام داد.

محدودیت های داده کاوی در حوزه اطلاعات سلامت

داده کاوی در حوزه سلامت ویژگیهای مربوط به خود را دارد که آنرا از سایر زمینه هایی که داده کاوی انجام می دهیم مجزا می کند. در داده کاوی حوزه سلامت و پزشکی درمانی ما با محدودیت هایی از قبیل داده های ناهمگن ، محرمانه یا خصوصی و حجیم و توزیع شده برخورد داریم که به برخی از این محدودیت ها اشاره می کنیم .

➤ امروزه پایگاه داده های بالینی حجم زیادی از اطلاعات درباره ی بیماران و وضعیت پزشکی آنها را در خود ذخیره می کنند. استخراج دانش مفید و تصمیم سازی های علمی برای تشخیص و بررسی بیماری به کمک این مجموعه داده ها می تواند بسیار مفید واقع شود. تشخیص بیماری و تعیین درمان مناسب برای بیماران در علم پزشکی از اهمیت بالایی برخوردار است .

□ کاربرد داده کاوی در مدیریت ارتباط با مشتری در بخش سلامت

➤ مدیریت ارتباط با مشتری یک راهکار مؤثر برقراری ارتباط موفق میان شرکت های تجاری از قبیل بانک ها و فروشگاه ها با مشتریان خود می باشد. این موضوع در حوزه خدمات بهداشتی و درمانی نیز کم اهمیت نمی باشد [8].

➤ داده کاری می تواند برای شناسایی و تعیین اولویت ها، الگوهای مفید، نیازهای فعلی و آینده، به منظور بهبود سطح رضایمندی در ارائه خدمات بهداشتی و درمانی استفاده شود.

➤ ارتباط با مشتری ممکن است در مطب پزشک اتفاق بیافتد یا در بخش بستری بیمار باید توجه داشت که داده کاوی با بررسی و تحلیل داده های بیماران می تواند به ارایه اطلاعاتی درباره مدت زمان انتظار بیمار ، راه کار های ممکن برای بهبود خدمات به او و ارایه دانش در باره آنچه که بیماران از مراکز درمانی می خواهند کمک کند ؛ بطور مثال با تحلیل علاقه مندی بیماران به شبکه های اجتماعی و دسته بندی آنها می تواند اطلاعات مربوط به دوران بعد از بستری و یا پیشگیری از بیماری را در شبکه های اجتماعی در بستر موبایل که قابل دسترس آسان است را در اختیار همه گروه های مورد نظر قرار دهد. و یا با تحلیل نسخ پزشکان و داروهای

و کاهش ابعاد داده ها شامل نمونه برداری تصادفی از بین حجم انبوه رکوردها و یا انتخاب ویژگی ها و صفات درخور توجه و مرتبط با هدف داده کاوی مفید خواهد بود. همچنین تکنیک های مصورسازی نیز می توانند نقش موثری در صورت مواجه با حجم زیاد داده ها داشته باشند. زیرا اطلاعات بسیار زیادی را بصورت یک تصویر لحظه ای از نتایج ارائه می کنند.

► محرمانگی داده های سلامت

با توجه به اینکه رازداری پزشکان و مراکز درمانی، و اعتماد بیمار به آنان، اصل بسیار مهمی در پزشکی بشمار می آید، تحلیلگران می بایست توجه ویژه ای به پیاده سازی محرمانگی و امنیت داده های پزشکی داشته باشند. حفظ محرمانگی اطلاعات اجازه دسترسی مستقیم به تمامی داده ها را ممکن نمی سازد. برخی دولت ها قوانین و مقررات خاصی به منظور حمایت از محرمانگی داده های پزشکی وضع نموده اند [17]. بعنوان مثال توصیه هایی به منظور پنهان کردن داده های مربوط به شناسایی بیمار از قبیل نام، آدرس، شماره تلفن بیمار کرده اند. یکی از این توصیه ها اینست که داده های مربوط به شناسایی بیمار را طوری پنهان کرد که به هیچ وجه قابل دستیافتنی نباشد. مثلا فقط دسترسی به اطلاعات مربوط به تشخیص بیماری از روی داده های بیمار امکانپذیر باشد. [17] در بعضی موارد ممکن است این رویکرد مورد پذیرش واقع نشود. زیرا نیازمند بررسی صحت تکنیک داده کاوی اعمال شده و

► ناهمگونی و حجیم بودن اطلاعات سلامت و درمانی

صنعت پزشکی جزء چند صنعتی می باشد که با حجم زیاد داده و اطلاعات، درگیر هستند. این اطلاعات و داده ها هر روزه در حال افزایش هستند. بعنوان مثال روش های عکس برداری از قبیل PET، MRI، SPECT و مجموعه سیگنال های ECG و EEG روزانه تا چندین گیگابایت داده تولید می کنند [17]. وجود ساختارهای مختلف داده باعث نوعی ناهمگونی و پیچیدگی در مجموعه داده های پزشکی شده است. این داده ها ممکن است از تصاویر و سیگنال های مختلف، ویزیت بیماران، نتایج آزمایشگاهی و بالینی، یادداشت ها، مشاهدات و تفسیرهای پزشکان و بخش های حسابداری و پذیرش بیمار بدست آمده باشد. همین این موارد می توانند بر پیشگویی، تشخیص، و درمان تاثیر گذار باشند و نمیتوان از آنها چشم پوشی کرد [17]. این حجم انبوه داده ها و ناهمگونی نیاز به دستگاه های ذخیره سازی داده با ظرفیت بالا و ابزارهای داده کاوی قوی در جهت تحلیل و کاوش در داده ها دارد. با وجود آن که تعداد سیستم ها و ابزارهای داده کاوی زیاد است، آنچه این سیستم ها را از یکدیگر متمایز می کند، توانایی آنها در مدیریت و استفاده از مقادیر حجیم و ناهمگون داده ها می باشد. توانایی دراستفاده و استخراج دانش از این داده ها بسیار مهم و تعیین کننده می باشد. اگر اطلاعات ذخیره شده در فرمت قابل درک نباشند، کمتر مفید خواهند بود. در این وضعیت، تبدیل داده ها به فرمت قابل درک و همگون تر

سیستم های ترجمه ی موجود جملات با بیشتر از ده کلمه را نمی توانند ترجمه کنند. حتی انسان نیز معنی یک جمله ی طولانی را در اولین بار خواندن نمی تواند متوجه شود. همچنین برای یک جمله ورودی معمولاً یک ترجمه ی منحصر بفرد وجود ندارد. بعلاوه ممکن است اصطلاحات عامیانه ی موجود در جمله ی مبدا در فرهنگ لغات مترجم وجود نداشته باشد و نیاز به بروز رسانی مداوم آن باشد. استفاده از متن کاوی در دست نوشته ها، یادداشت ها و نسخه های تجویز شده توسط پزشکان نیز می تواند در این زمینه مفید باشد. اگر چه متدهای داده کاوی قادر به کشف الگوها و رابطه پنهان موجود در مجموعه داده های بزرگ پزشکی می باشند، ولی نتایج آنها بدون همکاری و بازخورد متخصصین و خبرگان حوزه سلامت در مال درمالی فایده خواهد بود. همه ی الگوهایی که از طریق داده کاوی کشف می شوند مفید و درخور توجه نمی باشند [8]. بنابراین الگوهای پیدا شده توسط روش های داده کاوی، می بایست توسط متخصصین حوزه سلامت و پزشکی که سال ها تجربه در آن زمینه ی خاص را دارند، از نظر منطقی بودن، قابل عملیاتی بودن و جدید بودن مورد ارزیابی قرار بگیرند.

نتیجه گیری

▶ امروزه تولید داده های بزرگ، بخشی از فرآیند درمان در حوزه سلامت و پزشکی شده است و از طرفی یک نمونه مورد بررسی در این حوزه شامل تعداد زیادی متغیر می باشد که این خود در افزایش حجم اطلاعات نقش عمده ای دارد. افزایش حجم

یا اطلاعات بیشتری از یک بیمار خاص باشد. بعنوان مثال پس از اعمال تکنیک های داده کاوی و تشخیص یک بیماری و روش درمان آن، بخواهیم به آن بیماران خاص درباره تشخیص و روش درمان مناسب، اطلاع رسانی کرده و یا آنها را تحت معالجه قرار داد. مساله امنیت داده ها در استفاده و بخصوص انتقال داده ها نیز از اهمیت خاصی برخوردار است. قبل از اینکه داده های مربوط به شناسایی بیماران پنهان یا محرمانه شوند، فقط افراد مجاز می بایست به داده ها دسترسی داشته باشند.

▶ محدودیت اهمیت تفسیر و تحلیل پزشکان از داده ها

معمولاً تفسیر پزشکان از تصاویر، سیگنال ها و هر داده ی بالینی دیگر، بصورت دست نوشته های بی نظم و ترتیب و به زبان انگلیسی می باشد که استاندارد سازی آنها و در نتیجه داده کاوی بسیار مشکل می باشد. حتی متخصصین یک زمینه پزشکی نیز بر روی اصطلاحات استفاده شده برای توصیف وضعیت یک بیمار توافق ندارند و ممکن است از نام های متفاوتی برای توصیف یک بیماری استفاده کنند [17]. استفاده از یک کامپیوتر مترجم را می توان بعنوان راه حل بخشی از مشکل فوق پیشنهاد نمود که عمل ترجمه در آن شامل سه مرحله می باشد: تحلیل زبان منبع، ترجمه از یک زبان به زبان دیگر و سپس تولید جملات به زبان هدف.

▶ در کل هدف از بکارگیری داده کاوی در حوزه سلامت ، جایگزینی آن با متخصصین و محققین این حوزه نیست. بلکه بعنوان مکمل آنها در جهت افزایش دقت ، کارایی و کاهش هزینه ها و زمان رسیدن به نتایج و دانش مورد نظر می باشد.

پیشنهادات

در تحقیقات آینده با توجه به ماهیت اطلاعات سلامت که بخشی از آنها متنی می باشد باید به استفاده از ابزارهای متن کاوی و معرفی نرم افزارهای مناسب در این زمینه پرداخت. لذا پیشنهاداتی از قبیل موارد زیر قابل طرح است:

✓ استفاده از متن کاوی در خواندن نسخ درمانی پزشکان و ایجاد پایگاه داده تشخیص های در مانی و دارویی.

✓ غربال گری و خوشه بندی بیماران و دسته بندی بیماری ها با توجه به مراجعه به مراکز د مانی محلی و استانی در راستای مدیریت کلان سلامت و رتبه بندی خدمات مراکز درمانی .

✓ استفاده از کانال های اجتماعی در بستر رسانه موبایل جهت آموزش بیماران و اطلاع رسانی بهینه به مشتریان و پیگیری خدمات درمانی مراجعان.

✓ بررسی کاربرد داده کاوی در تشخیص بیماری های پوستی و پیشنهاد استفاده از تکنیک های ترکیبی شبکه عصبی و الگوریتم ژنتیک برای تحلیل داده ها با توجه به المان های سابقه بیماری در والدین و حساسیت به آلرژن ها، در معرض آلودگی محیطی بودن، واکنش به هیجانان و آلرژن به غذا و در معرض دود سیگار و حیوانات خانگی بودن با رویکرد Crisp در نرم افزار کلمنتاین برای تشخیص و ارائه بهترین برنامه درمانی برای بیماری پروانه ای .

داده های حوزه سلامت از یک سو و نیاز روز افزون به افزایش فهم و دانش در حوزه های مختلف سلامت و درمانی از جمله بیولوژی، بیوشیمی، پاتولوژی، روانشناسی و موارد مشابه باعث شده است که داده کاوی بتواند کاربرد فراوانی در حوزه اطلاعات سلامت و درمانی داشته باشد.

▶ الگوریتم های استفاده شده در تکنیک های داده کاوی سعی میکنند که نزدیکترین مدل به ویژگی های داده های مورد نظر را پیدا و ارائه نمایند. مدل ها می توانند پیشگویانه یا توصیفی باشند. مدل های پیشگویانه برای اهداف پیشگویی مثلا تشخیص یک بیماری خاص استفاده می شوند. درمان یک بیماری ممکن است فقط بر اساس پیشینه و وضعیت آن بیمار خاص نباشد و نتایج حاصل از درمان سایر بیماران با علائم مشابه نیز مورد توجه قرار گیرد. در این روش ها تعیین وضیت داده های آینده، بر مبنای متادیر گذشته و فعلی بیمار صورت می گیرد. مدل های توصیفی برای شناسایی الگوها در داده ها استفاده می شوند. کلاسه بندی، رگرسیون و تحلیل سری های زمانی معروفترین روشها در مدل پیشگویی بوده و خوشه بندی ، قوانین وابستگی و مصور سازی معروفترین روش های مدل توصیفی می باشند [12].

▶ در این مقاله کاربردهای داده کاوی در حوزه مدیریت سلامت و خدمات بهداشتی و درمانی مانند پیش بینی بیماری و مدیریت ارتباط با مشتری بررسی شد و برخی از محدودیت های داده کاوی در حوزه سلامت مانند ناهمگونی و حجیم بودن داده ها و محرمانگی داده ها و راهکار های رفع یا کنترل بهینه آنها مطرح گردید. همچنین تکنیک های مفید و الگوریتم های مناسب در حوزه داده کاوی اطلاعات سلامت بیان شد.

- 7) Samad Soltani Heris T, Lagarizadeh M , Mahmoodvand Z & Zolnoori M. Intelligent diagnosis of asthma using machine learning algorithms. International Research Journal of Applied and Basic Sciences 2013; 4(12): 4041-6.
- 8) Koh HC & Tan G. Data mining applications in healthcare. Journal of Healthcare Information Management 2005; 19(2): 64-72.
- 9) Han J, Kamber M & Pei J. Data mining: Concepts and techniques. USA: Morgan Kaufmann Publishers Inc; 2005: 5-6.
- 10) Lee IN, Liao SC & Embrechts M. Data mining techniques applied to medical information. Medical Informatics and the Internet in Medicine 2000; 25(2): 81-102.
- 11) Obenshain MK. Application of data mining techniques to healthcare data. Infection Control and Hospital Epidemiology 2004; 25(8): 690-5.
- 12) Harleen Kaur , Sri Krishan Wasan and Vasudha Bhatnagar , THE IMPACT OF DATA MINING TECHNIQUES ON MEDICAL DIAGNOSTICS, Data Science Journal, Volume 5, pp119-126, 19 October 2006
- 13) Neural networks in data mining dr yashpal singh, aloksingh chauhan reader, bundelkhand institute of engineering & technology, jhansi, india lecturer, united institute of management, allahabad, india
- 14) Use of data mining & neural network in medical industry, ina kapok sharma lecture B.N college of engineering & technology current development in artificial intelligence ISSN 0976-5832 Volume 3, Number 1(2012), international research publication house pp.1-8
- 15) Bereznicki BJ, Peterson GM, Jackson SL, Walters EH, Fitzmaurice KD & Gee PR. Data-mining of medication records to

حوزه پزشکی و سلامت از بخش‌های مهم در جوامع امروزی می‌باشد. استخراج دانش از میان حجم انبوه داده‌های مرتبط با سوابق بیماری و پرونده‌های پزشکی افراد با استفاده از فرآیند داده‌کاوی می‌تواند منجر به شناسایی قوانین حاکم بر ایجاد، رشد و تسری بیماری‌ها گردیده و اطلاعات ارزشمندی را به منظور شناسایی علل رخداد بیماری‌ها، تشخیص، پیش‌بینی و درمان بیماری‌ها با توجه به عوامل محیطی حاکم در اختیار متخصصین و دست‌اندرکاران حوزه سلامت قرار دهد. نتیجه این مساله به معنای افزایش عمر و ایجاد آرامش برای افراد جامعه است.

منابع

- 1) جمال شهبازی، ابوالفضل زارع، " داده‌کاوی با کلمنتاین "، تهران، جهاد دانشگاهی، واحد صنعتی امیر کبیر 1392.
- 2) طهماسبی، حمیدرضا، " ارزیابی کارایی تکنیک‌های کلاسه‌بندی در داده‌های پزشکی"، مجموعه مقالات چهارمین کنفرانس ملی داده‌کاوی ایران، آذر 1389.
- 3) مقدسی، حمید، جهانبخش، مریم، " داده‌کاوی و کاربرد آن در سلامت"، مدیریت اطلاعات سلامت، تیر 1391.
- 4) Rogers G, Joyner E. Mining Your Data for Healthcare Quality Improvement [Online]. 2011 [cited 2011 Aug 8]; Available from: URL: <http://www2.sas.com/proceedings/sugi22/EMERGING/PAPER139.PDF/>
- 5) Balib RK. Clinical Knowledge Management: Opportunities and Challenges. Hershey: Idea Group Inc (IGI); 2005.
- 6) LaTour KM, Eichenwald S. Health Information Management: Concepts, Principles, and Practice. Chicago: AHIMA; 2002. p. 478-80.

- Medical Journal 2009; 67(6): 421-7. [In Persian].
- 20) Obenshain MK. Application of data mining techniques to healthcare data. Infect Control Hosp Epidemiol 2004;25(8): 690-5.
- 21) Bellazi, Riccardo . Larizza, Cristiana . Magni, Paolo. Bellazi, Roberto. (2005). "Temporal data mining for the quality assessment of hemodialysis services", Elsevier Journal of Computers in Biology and Medicine. , No. 34: 25-39.
- 22) Yeh, Jin-Yi . Wu, Tai-His . (2009). "A Decision Support System for Predicting Hospitalization of Hemodialysis Patients", World Academy of Science, Engineering and Technology. , No. 53.
- 23) Bries, Michael Francis. (2007). "Modeling of hemodialysis patient hemoglobin: a data mining exploration", Iowa Research Online. , USA.
- improve asthma management. Medical Journal of Australia 2008; 189(1): 21-5.
- 16) Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. International Journal of Computer Applications 17(8):43-48, March 2011.
- 17) Krzysztof J. Cios, William, Uniqueness of medical data mining, Artificial Intelligence in Medicine, Vol. 26, No. 1-2), pp. 1-24, 2002.
- 18) Canlas RD. Data Mining in Healthcare: Current Applications and Issues [Online]. 2009 [cited 2009 Aug 9]; Available from: URL: http://mines.humanoriented.com/ciasses/2010/fall/csci568/papers/Data_Mining_Health.pdf. 2012.
- 19) Sepehri MM, Rahnama P, Shadpour P, Teimourpour B. A data mining based model for selecting type of treatment for kidney stone patients Tehran University

dataacademy.ir