داده کاوی و کاربرد آن در سلامت

حمید مقدسی[1]، اعظم السادات حسینی[2]، فرخنده اسدی[2]، مریم جهانبخش[3]

چکیده

پایگاه داده‌ها در حوزه‌ی سلامت حاوی میزان وسیعی از داده‌های بالینی است که کشف ارتباطات و الگوها در آن می‌تواند به دانش جدید پزشکی بیانجامد. امروزه با توجه به ظهور نظام‌های اطلاعات یکپارچه و رشد فن‌آوری اطلاعات، این مهم بیش از پیش نمایان شده است. داده کاوی از جمله پیشرفت‌های فن‌آوری در راستای مدیریت داده‌ها است و استفاده‌ی گسترده از سیستم‌های اطلاعات و پایگاه‌های داده، ادغام آن را با شیوه‌های سنتی به یک الزام تبدیل کرده است. در این مقاله که از نوع مروری می‌باشد هدف آن است تا به بررسی مفاهیم مرتبط با داده کاوی و کاربرد آن در حوزه‌ی سلامت پرداخته شود.

واژه‌های کلیدی: داده کاوی؛ سلامت؛ دانش.

نوع مقاله: مروری

دریافت مقاله: ۹۰/۵/۳ اصلاح نهایی: ۹۰/۶/۱۹ پذیرش مقاله: ۹۰/۸/۱۰

ارجاع: مقدسی حمید، حسینی اعظم السادات، اسدی فرخنده، جهانبخش مریم. **داده کاوی و کاربرد آن در سلامت.** مدیریت اطلاعات سلامت ۱۳۹۱؛ ۹ (۲): ۳۰۴-۲۹۷.

مقدمه

تحول در صنعت سلامت به واسطه‌ی این هدف واحد که «چگونه سازمان‌های سلامت هزینه‌ها را کاهش و کیفیت را افزایش دهند و هم چنان رقابتی باقی بمانند؟» به پیش می‌رود و این مقوله همواره یک چالش بزرگ محسوب می‌شود. بهبود کیفیت در صنعت سلامت را می‌توان به واسطه‌ی نیروهای محرکی که بر آن تأثیرگذار است بهتر تعریف نمود و از جمله‌ی این نیروهای محرک، داده‌های سلامت است؛ به عبارت دیگر در هر نوع برنامه‌ی بهبود کیفیت متمرکز بر بیمار، داده‌ها قلب آن برنامه به حساب می‌آید (۱). داده‌ها در عصر امروزی یعنی عصر اطلاعات، عمده‌ترین دارایی برای سازمان‌های سلامت بوده (۲) و موفقیت سازمان‌های سلامت در گروی جمع‌آوری، ذخیره و تحلیل آن‌ها است (۳). با این وجود، جمع‌آوری و ذخیره‌ی میزان زیادی از داده‌ها می‌تواند یک نوع اتلاف محسوب شود؛ مگر این که داده‌ها به شکل سودمند استفاده شده و تبدیل به یک منبع مالی برای سازمان گردد. برای تبدیل این ارزش بالقوه به اطلاعات استراتژیک، بسیاری از سازمان‌ها به داده کاوی روی آورده‌اند (۱)؛ چرا که به واسطه‌ی داده کاوی امکان کشف روابط، روندها و الگوهای مخفی بین داده‌ها و دستیابی به دانش نوین در زمینه‌ی چالش‌های آشکار و نهان سازمان میسر خواهد شد.

داده کاوی در اواخر دهه‌ی ۱۹۸۰ پدیدار گشت، در دهه‌ی ۱۹۹۰ گام‌های بلندی در این شاخه از علم برداشته شد و انتظار می‌رود در قرن حاضر نیز به رشد و توسعه‌ی خود ادامه دهد و پیش‌بینی‌ها حاکی از آن است که در دهه‌های آتی با توسعه‌ای انقلابی مواجه شود. مؤسسه‌ی فن‌آوری ماساچوست (Massachusetts institute of technology) داده کاوی را یکی از ده فن‌آوری برتری می‌داند که نقش چشم‌گیری در تحول جهان خواهد داشت (۴).

۱. دانشیار، مدیریت اطلاعات سلامت، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.
۲. استادیار، مدیریت اطلاعات سلامت، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.
۳. مربی، آموزش مدارک پزشکی، دانشگاه علوم پزشکی اصفهان و دانشجوی دکتری، مدیریت اطلاعات سلامت، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران. (نویسنده‌ی مسؤول)
Email: jahanbakhsh.him@gmail.com

Archive of SID
www.mui.ac.ir
www.SID.ir

با توجه به این که در حال حاضر در کشورمان روزانه حجم وسیعی از داده‌ها طی فرایندهای مراقبت و درمان حاصل می‌شود، اما تحلیل و تفسیر آن‌ها با هدف کشف الگوها و دانش جدیدی که بتواند منتج به ارتقای سلامت باشد بسیار کمرنگ است. هدف از مطالعه‌ی حاضر، بررسی مقوله‌ی داده کاوی به عنوان یکی از رویکردهای مهم در این راستا بوده است.

شرح مقاله

اصطلاح داده کاوی برگرفته از (Gold mining) یا استخراج طلا از صخره‌های سنگی است. در رابطه با استخراج و اکتشاف طلا از واژه‌ی Rock mining استفاده نشده است و بنابراین شاید نام مناسب‌تر برای Data mining نیز واژه‌ی Knowledge mining from data بود، اما از آن جا که این واژه طولانی است و ممکن است اصطلاح کوتاه‌تر Knowledge mining نیز نتواند مؤکد بر استخراج دانش از میزان وسیعی از داده‌ها باشد از واژه‌ی Data mining استفاده می‌شود.

Data archeology، Data pattern analysis و Information harvesting از دیگر نام‌های به کار گرفته شده برای داده کاوی است (۶، ۵).

داده کاوی مفهومی است که با پایگاه‌های داده‌های بزرگی مانند Data repositoryها وData warehouseها عجین شده است (۷) و به دنبال استخراج اطلاعات مفید و ناشناخته از داده‌های خام است (۹، ۸).

اگر چه داده کاوی هم چون دیگر مفاهیم فن‌آوری اطلاعات معانی متعددی را برای افراد مختلف تداعی می‌کند، اما اگر به طور دقیق استفاده شود به ابزار تحلیلی پیچیده‌ای – که به طور خودکار الگوهای مفیدی را در میان داده‌های یک مخزن داده کشف می‌کند– اشاره دارد.

در واقع داده کاوی شکل پیشرفته‌ی پشتیبانی از تصمیم است و برخلاف ابزارهای پرس و جوی غیر فعال (Passive) بدون الزام به طرح سؤال از طرف کاربر به تولید الگو، روندها و قواعد برنامه‌ریزی شده می‌پردازد (۷)، به عبارت دیگر قدرت داده کاوی در این است که می‌تواند الگوهایی را که در جستجوی کاربر مورد توجه قرار نگرفته است، افشا کند و پاسخ‌هایی را برای سؤالاتی که هرگز درخواست نشده بود، تولید نماید (۳).

بنابراین هدف داده کاوی، استخراج اطلاعات با ارزش از داده‌ها با هدف غایی کشف دانش است (۱۰–۸)، لذا برخی از افراد اصطلاح کشف دانش از داده‌ها (Knowledge discovery from data) را به جای داده کاوی به کار می‌برند و برعکس عده‌ی زیادی داده کاوی را هسته‌ی فرایند کشف دانش از داده‌ها و گامی اساسی در آن دانسته (۱۱، ۶، ۵) و آن را یکی از مهم‌ترین زیر شاخه‌های مدیریت دانش معرفی می‌کنند (۱۲).

با توجه به تعریف داده کاوی نظر دوم صحیح است و داده کاوی بخشی از کشف دانش است که شامل مراحل زیر می‌باشد:

۱. *پاک‌سازی داده‌ها (Data cleaning)*: در این مرحله داده‌های غیر همسان و زاید از میان داده‌ها خارج می‌شود.

۲. *یکپارچگی داده‌ها (Data integration)*: در این مرحله منابع متعدد داده‌ها با هم ترکیب می‌شوند.

۳. *گزینش داده‌ها (Data selection)*: در این مرحله داده‌های مربوط به هدف، از پایگاه داده‌ها بازیابی می‌شود.

۴. *تغییر شکل داده‌ها (Data transformation)*: در این مرحله داده‌ها تغییر شکل یافته و به واسطه‌ی عملیات مختلفی مثل خلاصه‌برداری، به اشکال مناسب برای استخراج تبدیل می‌شوند.

۵. *داده کاوی (Data mining)*: فرایندی ضروری است که برای استخراج الگوهای داده‌ها از شیوه‌های منطقی و خردمند استفاده می‌کند.

۶. *ارزشیابی الگو (Pattern evaluation)* (۵)

۷. *ارایه‌ی دانش (Knowledge presentation)* (۱۳)

هر یک از این مراحل پیش‌نیاز مرحله‌ی بعد بوده و فرایندی تکرارپذیر است. بدین ترتیب کل فرایند کشف دانش از داده‌ها، از مجموعه‌ی وسیعی از داده‌ها شروع شده و به محصول نهایی دانش و ایجاد مدل ختم می‌شود (۱۳).

داده کاوی با سرنام SEMMA مشخص می‌شود. این سرنام متشکل از Sample (نمونه) که بیانگر یک نمونه‌ی آماری از داده‌ها است، Explore (اکتشاف) که به استفاده از

Archive of SID
www.mui.ac.ir
www.SID.ir

تکنیک‌های اکتشافی و آماری اشاره دارد، Modify یا Manipulation (اصلاح یا دستکاری) که به گزینش و تغییر شکل مهم‌ترین متغیرهای قابل پیش‌بینی اشاره دارد، Model (نمونه) به معنای مدلی از متغیرها برای پیش‌بینی پیامدها یا الگوها می‌باشد و Assess (ارزیابی) به معنای تأیید صحت مدل ارایه شده است. SEMMA خود یک چرخه است که در صورت نیاز می‌تواند به طور متوالی انجام شود (۱۴).

مهم‌ترین دلایل استفاده از داده کاوی

Wickramasinghe و همکاران، مهم‌ترین علل استفاده از داده کاوی را به شرح زیر بیان کرده‌اند:

۱. دلیل اصلی استفاده از داده کاوی، افزایش حجم داده‌های موجود و آتی سازمان‌ها است که نیاز به پردازشی فراتر از رویکردهای سنتی دارد. برای انسان کنار آمدن با این حجم انبوه و در حال رشد داده‌ها با اتکای محض به شیوه‌های سنتی غیر ممکن است.

۲. ذهن انسان به هنگام تحلیل انبوه داده‌ها ممکن است دچار خطا شود و علاوه بر این، همیشه سعی بر این دارد که نتایج را مبنی بر تجربیات و آزمون‌های قبلی بنا نهد؛ اما داده کاوی بدون تکیه بر فرضیه‌های از پیش پنداشته به تحلیل داده‌ها می‌پردازد.

۳. یکی از مهم‌ترین مزایای داده کاوی به ویژه در حجم وسیع داده‌ها، این است که داده کاوی نسبت به استخدام تیمی از افراد خبره بسیار کم هزینه‌تر است. اگر چه داده کاوی مانند سایر تکنیک‌ها بدون مداخله‌ی انسان میسر نیست، اما به راحتی اجازه می‌دهد تحلیلگری که در آمار یا برنامه‌نویسی زیاد حرفه‌ای نیست بتواند به مدیریت فرایند استخراج دانش از داده‌ها بپردازد (۱۵).

داده کاوی در عرصه‌ی سلامت

مفهوم کشف دانش از داده‌ها بیش از یک دهه است که در محیط‌های مالی- تجاری در حال استفاده می‌باشد (۱۳) و در علوم مدیریت ارتباطات، مهندسی، وب‌کاوی، تحلیل جرایم و پزشکی جای خود را باز کرده است (۱۲). اگر چه کشف دانش با هدف شناسایی اختلاس مالی وارد عرصه‌ی سلامت شد، اما به تدریج در حوزه‌ی بالینی نیز مورد استفاده قرار گرفت. این مهم ناشی از تغییر سریع هوشیاری نسبت به اطلاعات در حوزه‌ی سلامت است (۱۳).

صنعت سلامت به طور مستمر در حال تولید میزان زیادی از داده‌ها می‌باشد (۱۷، ۱۶) و افرادی که با این نوع داده‌ها مواجه هستند، دریافته‌اند که بین جمع‌آوری تا تفسیر آن‌ها شکاف وسیعی وجود دارد؛ حوزه‌ی به نسبت جوان و در حال رشد داده کاوی در سلامت (۱۸) از جمله شیوه‌هایی است که می‌تواند این صنعت را از تحلیل عمیق این داده‌ها بهره‌مند سازد و به توسعه‌ی تحقیقات پزشکی و تصمیم‌گیری‌های علمی در زمینه‌ی تشخیص و درمان منتج شود (۱۸-۱۶).

داده کاوی در پزشکی و بیولوژی بخش مهمی از انفورماتیک زیست- پزشکی است و یکی از کاربردی‌ترین علوم کامپیوتر در این علم بوده که در بیمارستان‌ها، کلینیک‌ها، آزمایشگاه‌ها و مراکز تحقیقاتی به کار گرفته شده است (۱۶).

داده کاوی به کندی اما به طور فزاینده‌ای برای رفع مشکلات متعدد در کشف دانش و در بخش سلامت به کار گرفته شده است. ۴ مورد از مهم‌ترین دلایل رشد کند این علم در سلامت؛ حساسیت علم پزشکی و گره خوردن آن با جان انسان‌ها (تفاوت جزیی در الگوهای داده کاوی می‌تواند به تغییر تعادل بین مرگ و زندگی بیانجامد)، سردرگمی در تعریف داده کاوی (گاه ایجاد یک طرح ساده از پایگاه داده‌های پزشکی به غلط به عنوان الگوی حاصل از داده کاوی مطرح می‌شود)، حریم شخصی و محرمانگی داده‌های سلامت و در نهایت مهم‌ترین چالش این است که اگر فرض بر این باشد که نتایج داده کاوی به طور کامل قابل اعتماد است؛ تغییر عادت ارایه دهندگان مراقبت از پزشکی سنتی به پزشکی مبتنی بر شواهد (Evidence based medicine) دشوار است (۱۹).

با این وجود، امروزه بخش سلامت بیش‌ترین نیاز را به داده کاوی پیدا کرده است و حرکت از پزشکی سنتی به سمت پزشکی مبتنی بر شواهد از جمله مواردی است که می‌تواند مؤکد این امر باشد. در ادامه رایج‌ترین استراتژی‌های داده کاوی، تکنیک‌های داده کاوی و نمونه‌هایی از کاربردهای آن

www.mui.ac.ir
www.SID.ir
Archive of SID

در سلامت بیان می‌گردد.

استراتژی‌های داده کاوی

به طور کلی هدف داده کاوی، یادگیری و آموختن از داده‌ها است و بر این اساس دو دسته‌ی کلی از استراتژی‌های داده کاوی شامل یادگیری نظارت شده و یادگیری فاقد نظارت وجود دارد. شیوه‌های یادگیری نظارت شده زمانی به کار می‌رود که ارزش متغیرهای ورودی (درون‌دادها) برای ما شناخته شده باشد. یافتن مدل‌های پیش‌بینی خطا در مطالبات بیمه‌ی یک مؤسسه‌ی سلامت، نمونه‌ای از استراتژی یادگیری نظارت شده است؛ در این استراتژی مدل‌ها و ویژگی‌ها برای ما شناخته شده و با هدف پیش‌بینی داده‌ها و کشف اطلاعات به کار می‌رود اما در شیوه‌ی یادگیری فاقد نظارت، بر روی داده‌هایی که ارزش آن‌ها شناخته شده نیست تمرکز می‌شود. به عنوان مثال در مدل‌های فاقد نظارت، ویژگی‌ها و مدل‌های خطاهای مطالبات شناخته شده نیست، اما الگوها و خوشه‌های حاصل از داده کاوی منجر به کشفیات جدید می‌شود (۱۴).

تکنیک‌های داده کاوی

داده کاوی از تکنیک‌هایی هم چون مدل‌های Bayesian، درخت تصمیم‌گیری، شبکه‌های عصبی مصنوعی، قواعد ارتباطات و الگوریتم‌های ژنتیک برای کشف الگوها استفاده می‌کند (۱۲) که برخی از آن‌ها بیان می‌شود.

درخت تصمیم‌گیری (Decision tree): این تکنیک، ساختاری شبیه به درخت است که مجموعه قواعدی که منجر به تصمیم‌گیری شده است را توصیف می‌کند و سهولت تفسیر از ویژگی‌های مهم آن است. برای مثال درخت تصمیم‌گیری می‌تواند عوامل مؤثر بر میزان بقای پیوند کلیه را تعیین کند، همچنین استفاده از الگوریتم DRG (کدهای تشخیصی وابسته) در بازپرداخت هزینه‌های بیمه‌ی پزشکی سالمندان در ایالات متحده، مثال کلاسیکی از روش شناسی این تکنیک محسوب می‌شود.

شبکه‌های عصبی (Neural networks): این تکنیک مدل‌های پیش‌بینی غیر خطی تولید می‌کند که یاد می‌دهد چگونه یک الگو با یک پروفایل خاص قابل تطبیق است، اما درباره‌ی علت رسیدن به این نتیجه‌ی خاص توضیحی ارایه نمی‌کند. برای مثال شبکه‌های عصبی می‌تواند مشخص کند که چه نوع بیماری‌هایی احتمال دارد با یک بیماری همراه شود (۳) و با تحلیل تصاویر، نوارهای قلبی و سایر مشاهدات بالینی به تشخیص، درمان و تولید دارو کمک می‌کند (۲۰).

منطق فازی (Fuzzy logic): منطق فازی نسبت به دیگر تکنیک‌ها انعطاف‌پذیرتر می‌باشد و مفاهیم مبهم و پیچیده را مدیریت می‌کند. شناسایی بیمارانی که شاید حاضر هستند آگاهانه به ابتکار بیمارستان‌ها در پیش‌گیری از ایدز به عنوان نمونه‌ی مورد مطالعه پاسخ دهند از موارد کاربرد این تکنیک است.

Nearest neighbor: تکنیکی قدیمی برای کشف ارتباطات در مواردی که داده‌ها عددی هستند، می‌باشد.

الگوریتم‌های ژنتیک (Genetic algorithms): تکنیک‌های بهینه‌ای برای ارتقای سایر الگوریتم‌های داده کاوی، به گونه‌ای که از بهترین مدل بر روی مجموعه‌ی داده‌ها استفاده کنند و می‌تواند برای یک بیماری خاص بهترین برنامه‌ی درمانی را تعیین کند (۳).

نمونه‌هایی از کاربردهای داده کاوی در سلامت

– *داده کاوی در تشخیص‌های غیر تهاجمی:* برخی از اقدامات تشخیصی و آزمایشگاهی برای بیماران، تهاجمی و هزینه‌بر و در عین حال رنج‌آور هستند. به عنوان مثال بافت برداری از گردن رحم به منظور تشخیص سرطان گردن رحم از جمله این موارد است.

Thangavel و همکاران (به نقل از Canlas) از طریق الگوریتم‌های خوشه‌بندی به تحلیل بیماران مبتلا به سرطان گردن رحم پرداختند و نتایج پیش‌گیری کننده‌تری را نسبت به عقاید پزشکی کنونی پیدا کردند (۱۹). 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 شیوه‌ی جدید به کار رود (۱۹).

– *داده کاوی در تعیین نوع درمان:* به کارگیری داده کاوی

Archive of SID
www.mui.ac.ir
www.SID.ir

بر روی داده‌های پزشکی دستاوردهای حیاتی و اثرگذاری را در انتخاب نوع درمان مناسب و نجات جان انسان‌ها به ارمغان آورده است. به عنوان مثال در بیمارستان شهید هاشمی‌نژاد تهران برای تعیین نوع درمان سنگ حالب از راهکار داده کاوی استفاده می‌شود. در این بیمارستان یک الگوریتم درختی وجود دارد که پزشک بر اساس آن درمانی را که میزان موفقیت بالاتری برای بیمار دارد انتخاب می‌کند و بررسی‌ها نشان داده است که میزان عملکرد این الگوریتم ۷۷ درصد بوده و بسیار بهتر از عملکرد الگوریتم بیمارستانی و مدل ذهنی پزشکان است (۲۱).

– *داده کاوی در شناسایی عوارض جانبی داروها:* برخی از داروها که در ابتدا به عنوان داروهای غیر مضر برای انسان به تأیید رسیده‌اند، بعد از استفاده‌ی عموم در طولانی مدت اثرات زیان‌بار خود را نشان می‌دهند.

اداره‌ی دارو و غذای آمریکا (FDA یا Food and drudge administration) با استفاده از داده کاوی به کشف دانش درباره‌ی عوارض جانبی داروها در پایگاه داده‌ی خود پرداخته است. الگوریتم مورد استفاده در این روش (Multi-item gamma poisson shrinker یا MGPS) نام دارد که توانسته است با موفقیت ۶۷ درصد عوارض جانبی داروها را ۵ سال زودتر از شیوه‌ی سنتی شناسایی کند (۱۹).

– *داده کاوی در پرونده‌ی الکترونیک سلامت:* در حال حاضر مطالعات متعددی مؤکد این است که تکنیک‌های داده کاوی ابزار مؤثری را برای شناسایی الگوهای مهم سلامت از درون پرونده‌های پزشکی فراهم می‌کند (۲).

پرونده‌های سلامت کامپیوتری به واسطه‌ی دربرداشتن مجموعه‌ای از داده‌ها درباره‌ی تشخیص، درمان، اقدامات آزمایشگاهی و دارویی (۲۰) به طور بالقوه منبع غنی از دانش هستند (۲۱، ۱۹). اگر چه کشف دانش از انبوه داده‌ها در آن‌ها برای انسان غیر ممکن نیست، اما امری دشوار است و داده کاوی بهترین شیوه برای حل این چالش می‌باشد (۱۹).

در بیمارستان Brigham and Women تحقیقی با فرضیه‌ای مبنی بر این که "می‌توان عناصر اطلاعاتی موجود در پرونده‌ی الکترونیک سلامت را با تکنیک داده کاوی شناسایی و روابط بالینی معنی‌دار و درستی را کشف کرد" انجام شد؛ در این مطالعه با اجرای داده کاوی بر روی ۱۰۰۰۰۰ پرونده و تمرکز بر مجموعه داده‌های تشخیصی (۲۷۲۷۴۹ مورد)، دارویی (۴۴۲۶۵۸ مورد) و نتایج آزمایشگاهی (۱۱۸۰۱۰۶۸ مورد) روابط علمی مهمی بین مشکلات بیمار، داروها و نتایج آزمایشگاهی آن‌ها کشف و توصیف شد (۲۱).

– *داده کاوی در کنترل عفونت بیمارستانی:* در ایالات متحده سالانه ۲ میلیون نفر به عفونت‌های بیمارستانی مبتلا می‌شوند، لذا تمرکز زیادی برای شناسایی این بیماران صورت گرفته است. به عنوان مثال در ایالت آلاباما نوعی سیستم نظارتی وجود دارد که از تکنیک‌های داده کاوی استفاده می‌کند. این سیستم با استفاده از قوانین و روابط داده کاوی بر روی کشت خون بیمار و داده‌های بالینی به دست آمده از سیستم اطلاعات آزمایشگاه (Laboratory information system) الگوهای جدید و جالب توجهی را مشخص می‌سازد و ماهانه الگوهایی که توسط کارشناسان کنترل عفونت مورد بررسی قرار می‌گیرد را تهیه می‌کند. سازندگان این سیستم دریافته‌اند که ارتقای کنترل عفونت با سیستم داده کاوی حساس‌تر از سیستم کنترل عفونت سنتی عمل می‌کند (۱۴).

– *داده کاوی در رتبه‌بندی بیمارستان‌ها:* رتبه‌بندی بیمارستان‌ها و برنامه‌های بهداشتی می‌تواند بر مبنای اطلاعات گزارش شده توسط ارایه دهندگان مراقبت باشد، بنابراین گزارش‌دهی استاندارد برای مقایسه‌ی معنی‌دار بیمارستان‌ها و رتبه‌بندی آن‌ها بسیار مهم است. از جمله شیوه‌های استاندارد کردن این گزارشات استفاده از تکنیک‌های داده کاوی می‌باشد، به عنوان مثال اگر کدهای ICD (کدهای عوامل خطر) اختصاص یافته به پرونده‌های بیمار با تکنیک‌های داده کاوی مانند خوشه‌بندی و ارتباطات همراه شود، می‌تواند به ایجاد گزارشاتی بیانجامد که با تعیین نرخ واقعی میزان ناخوشی‌ها، مرگ و میر و سایر شاخص‌های کیفیت در رتبه‌بندی بیمارستان‌ها مورد استفاده واقع شود (۱۴).

– *داده کاوی در بهره‌برداری از خدمات سلامت:* در برنامه‌ی سلامت اکسفورد (Oxford health plan)، داده کاوی به منظور پاسخ به موضوعات پیچیده‌ای پیرامون

www.mui.ac.ir
www.SID.ir
Archive of SID

بهره‌برداری خدمات ارایه شده به بیماران می‌پردازد و در اصل به دنبال درک بهتر نحوه‌ی بهره‌برداری از مراقبت‌ها و خدمات ارایه شده به بیماران می‌باشد.

به واسطه‌ی داده کاوی در این برنامه، بیمارستان‌ها در دستیابی به متغیرهای کلیدی در پیش‌بینی بهره‌برداری از خدمات، بهبود پیامدهای کیفیت، پیش‌بینی رفتارهای آتی بیماران و بهبود برنامه‌های درمانی آن‌ها، شناسایی بیماران پرخطر و مدیریت مراقبت آن‌ها توانمند می‌شوند (۱).

با استفاده از داده کاوی و مدل‌سازی داده‌ها می‌توان بیماران با شرایط پرخطر را شناسایی کرد. در واقع داده کاوی با ارایه‌ی اطلاعات به ارایه دهندگان مراقبت، آن‌ها را در شناسایی بیماران پرخطر به گونه‌ای که بتوان کیفیت مراقبت آن‌ها را بهبود داد و از مشکلات آتی آن‌ها جلوگیری کرد، کمک می‌کند (۱۴) و با طراحی مداخله‌ی مناسب منجر به کاهش پذیرش‌های بیمارستانی می‌شود (۱۸). به عنوان مثال تکنیک‌های مدل‌سازی پیش‌بینی کننده‌ی داده کاوی در رابطه با مدیریت بیماری دیابت منجر به ارتقای کیفیت و کاهش هزینه‌ی بیماران مبتلا به دیابت می‌شود (۱۴).

بحث

یکی از مهم‌ترین موضوعات چالش برانگیز در مراقبت سلامت، تغییر شکل داده‌های بالینی خام به اطلاعات معنی‌دار (۲) به دنبال تولید مستمر انبوهی از داده‌ها است (۱۷)؛ چرا که بسیاری از سازمان‌های مراقبت سلامت با وجود غنای داده با فقر دانش روبرو شده‌اند (۲۲). به عبارت دیگر رویارویی با مجموعه‌های عظیم داده‌ها (۲۳) و توسعه‌ی پایگاه‌های داده‌ها نسبت به دهه‌های گذشته (۲۴) نیازهای جدیدی را مانند خلاصه‌سازی خودکار داده‌ها، استخراج اطلاعات ذخیره شده و کشف الگوها از داده‌های خام به وجود آورده است که داده کاوی نمونه‌ای از آن‌ها می‌باشد (۲۳).

استخراج اطلاعات و دانش از داده‌ها مفهومی دیرینه در مطالعات علمی و پزشکی می‌باشد و آن چه که جدید است هم‌گرایی و اشتراک چندین رشته و فن‌آوری‌های متناظر آن‌ها است که فرصت منحصر به فردی را برای داده کاوی ایجاد کرده است (۲۵).

داده کاوی با ایجاد پزشکی مبتنی بر شواهد نقش حیاتی در سلامت دارد و منجر به کشف دانش جدید، سودمند و ماندگار در پایگاه‌های داده‌ای سازمان‌های سلامت می‌شود (۱۹)؛ چرا که برای دستیابی به پزشکی مبتنی بر شواهد باید از شناسایی شکاف و خلاء دانش در فرایندهای مراقبت سلامت کنونی شروع کرد و سپس به دنبال بهترین ادله بود، در قدم بعدی باید به بررسی صحیح بودن و معتبر بودن اقدامات شناسایی شده در بهترین ادله پرداخت و در قدم آخر این ادله را بر روی بیماران اجرا کرد (۲). داده کاوی دستیابی به اولین گام در این زمینه را هموار می‌سازد.

در محیط رقابتی امروز، سازمان‌های سلامتی که به واسطه‌ی فن‌آوری‌هایی هم چون داده کاوی بتوانند داده‌ها را در راستای بهبود کیفیت سلامت به کار برند سریع‌تر به قله‌ی موفقیت خواهند رسید، لذا لازم است سازمان‌های سلامت ما نیز با استفاده از تخصص صاحبان این فن از این عرصه‌ی رقابت باز نمانند.

بسیاری از مراکز تحقیقاتی کشورمان مانند مراکز تحقیقات غدد و متابولیسم، سرطان، تروما، پوست و غیره و همچنین نظام‌های اطلاعات بیمارستان‌ها (Hospital information systems) دارای حجم زیادی از داده‌ها هستند که یا هرگز تحلیل نمی‌شوند و یا اگر هم تحلیل و به دانش منتج شوند به واسطه‌ی استفاده از شیوه‌های سنتی، امری مقطعی و زمان‌بر است؛ حال آن که با روی آوردن به داده کاوی و اجرای آن می‌توانند داده‌ها را به ابزاری نیرومند و رقابتی تبدیل نموده و گام‌های جدیدی را در پیش‌گیری، تشخیص، درمان و ارایه‌ی خدمات با کیفیت به مشتریان سلامت بردارند.

پیشنهادها

۱.داده کاوی بر روی دفاتر ثبت بیماری‌ها و حوادث

۲. داده کاوی بر روی زیر سیستم‌های HIS مانند LIS، PACS و غیره

۳. داده کاوی بر روی پایگاه داده‌های کدگذاری بیماری‌ها و اقدامات

۴. مقایسه‌ی اثربخشی مراقبت و درمان قبل از اجرای داده کاوی و بعد از اجرای آن در مراکز سلامت.

www.mui.ac.ir

www.SID.ir

Archive of SID

References

1. 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/
2. Balib RK. Clinical Knowledge Management: Opportunities and Challenges. Hershey: Idea Group Inc (IGI); 2005.
3. LaTour KM, Eichenwald S. Health Information Management: Concepts, Principles, and Practice. Chicago: AHIMA; 2002. p. 478-80.
4. Adel A, Ahmadi P, Sebt M. Desining Model for Choosing human resources with data mining approach. Journal of Iranian Technology 2010; 2(4): 5. [In Persian].
5. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. Philadelphia: Elsevier; 2011.
6. Fayyad U, Shapiro G, Smyth P. Knowledge Discovery and Data Mining [Online]. 2011 [Cited 2011 Aug 8]; Available from: URl: http://www. Aaai.org/
7. Wager KA, Lee FW, Glaser JP. Managing Health Care Information Systems: A Practical Approach for Health Care Executives. New Jersey: John Wiley & Sons; 2005.
8. Tan J. Medical Informatics: Concepts, Methodologies, Tools, and Applications. Hershey: IGI Global snippet; 2008.
9. Koh HC, Tan G. Data mining applications in healthcare. J Healthc Inf Manag 2005; 19(2): 64-72.
10. Chakrabarti S, Cox E. Data Mining: Know It All. Amsterdam: Morgan Kaufmann p. 7; 2009.
11. Maimon OZ, Rokach L. Data Mining And Knowledge Discovery Handbook. New York: Springer Science & Business; 2010. p. 1.
12. Chen H, Fuller SS, Friedman C, Hersh W. Medical Informatics: Knowledge Management And Data Mining in Biomedicine. New York: Springer; 2005.
13. Englebardt SP, Nelson R. Health care informatics: an interdisciplinary approach. Philadelphia: Mosby; 2002. p. 125.
14. Obenshain MK. Application of data mining techniques to healthcare data. Infect Control Hosp Epidemiol 2004; 25(8): 690-5.
15. Wickramasinghe N, Gupta JN, Sharma SK. Creating Knowledge-Based Healthcare Organizations. Hershey: Idea Group Inc (IGI); 2005.
16. Berka P, Rauch J, Zighed DA. Data Mining and Medical Knowledge Management: Cases and Applications. Hershey: Idea Group Inc (IGI); 2009.
17. Zhu L, Wu B, Cao C. [Introduction to medical data mining]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2003; 20(3): 559-62.
18. Cios KJ. Medical data mining and knowledge discovery. IEEE Eng Med Biol Mag 2000; 19(4): 15-6.
19. Canlas RD. Data Mining in Healthcare: Current Applications and Issues [Online]. 2009 [cited 2009 Aug 9]; Available from: URL: http://mines.humanoriented.com/classes/2010/fall/csci568/papers/Data_Mining_Health.pdf. 2012.
20. Sepehri MM, Rahnama P, Shadpour P, Teimourpour B. A data mining based model for selecting type of treatment for kidney stone patients. Tehran University Medical Journal 2009; 67(6): 421-7. [In Persian].
21. Tavakoli N, Jahanbakhsh M. Opportunities and Challenges of EHR Implementation in Isfahan [Project]. Isfahan: School of Informatics and Management, The university of Isfahan; 2010. p. 3. [In Persian].
22. Jahanbakhsh M, Tavakoli N, Mokhtari H. Challenges of EHR implementation and related guidelines in Isfahan. Procedia Computer Science 2011; 3(0): 1199-204.
23. Wright A, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. J Biomed Inform 2010; 43(6): 891-901.
24. Kaur H, Wasan SK. Empirical Study on Applications of Data Mining Techniques in Healthcare. Journal of Computer Science 2006; 2(2): 194-200.
25. Moghaddasi H. Information Quality in Health care. Tehran: Word Processing Publishing; 2005. p. 41. [In Persian].

Archive of SID
www.mui.ac.ir
www.SID.ir

Application of Data Mining in Health

Hamid Moghaddassi, PhD[1]; Azamosadat Hoseini, PhD[2]; Farkhonde Asadi, PhD[2]; Maryam Jahanbakhsh[3]

Abstract

Health databases contain a wide scope of clinical data to explore relationships and patterns that can lead to new medical knowledge.Today, the emergence of integrated information systems and growth of information technologies have better highlighted the importance of such databases. Data mining is among the technological advances toward data management whose integration with traditional methods has become a necessity due to the widespread use of information systems and databases. This article reviews concepts related to data mining and its application in the field of health.

Keywords: Data Mining; Health; Knowledge.

Type of article: Review Article

Received: 24 Jul, 2011 *Accepted: 31 Oct, 2011*

Citation: Moghaddassi H, Hoseini A, Asadi F, Jahanbakhsh M. **Application of Data Mining.** Health Information Management 2012; 9(2): 304.

Archive of SID

1. Associate Professor, Health Information Management, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
2. Assistant Professor, Health Information Management, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
3. Lecturer, Medical Records, Isfahan University of Medical Sciences, Isfahan, Iran And PhD Student, Health Information Management, Shahid Beheshti University of Medical Sciences, Tehran, Iran. (Corresponding Author) Email: jahanbakhsh.him@gmail.com

www.mui.ac.ir www.SID.ir