مقاله مروری

مجله علوم پیراپزشکی و بهداشت نظامی
سال دهم، شماره یکم، تابستان ۱۳۹۴، مسلسل ۱۷

بررسی کاربردهای داده کاوی در نظام سلامت

لیلا غلامحسینی[۱]، مصطفی دامرودی*[۲]

چکیده

مقدمه: با گسترش داده‌های ذخیره‌شده در پایگاه‌های داده پزشکی نیاز به ابزارهای تخصصی برای دسترسی به داده‌ها، تجزیه و تحلیل داده‌ها، کشف دانش، و استفاده مؤثر از داده‌ها می‌باشد. داده کاوی یکی از مهم‌ترین این روش‌ها می‌باشد. در این مقاله تکنیک‌ها و نمونه‌هایی از کاربرد داده کاوی برای تشخیص پزشکی و پیش‌بینی بیماری‌ها در نظام سلامت بیان‌شده است.

مواد و روش‌ها: مطالعه حاضر با جستجوی مروری در متون موجود در پایگاه‌های اطلاعاتی فارسی و انگلیسی شامل Magiran: Iranmedex, SID, Google Scholar, OVID, Scopus, PubMed طی سال‌های ۱۹۹۸ تا۲۰۱۳ انجام گردید.

یافته‌ها: داده کاوی به معنای جستجوی خودکار منابع داده‌ای بزرگ، جهت یافتن الگوها و وابستگی‌هایی است که تحلیل‌های ساده آماری قادر به انجام آن نیستند. یکی از زمینه‌هایی که نیازمند استفاده از این ابزارها جهت تحلیل داده‌های وسیع و مدل‌سازی پیش‌گویانه با روش‌های محاسباتی جدید است، علم پزشکی می‌باشد. در علم پزشکی کشف و تشخیص به موقع بیماری‌ها می‌تواند از ابتلا به بسیاری از بیماری‌های مهلک نظیر سرطان جلوگیری نموده و موجب نجات زندگی مردم گردد. این مطالعه نشان می‌دهد که پیشگویی‌های داده کاوی ابزارهای ضروری را برای محققان و پزشکان جهت بهبود در پیشگیری از بیماری‌ها، روش‌های تشخیصی و برنامه‌های درمانی فراهم می‌نمایند.

بحث و نتیجه‌گیری: امروزه در دانش پزشکی جمع‌آوری داده‌های بیماری‌های مختلف اهمیت فراوانی یافته است. پیشرفت‌های دهه اخیر در ارتباط با فناوری‌های اطلاعات و نرم‌افزار کمک شایانی به بررسی‌های همه‌جانبه و کامل‌تر از داده‌های حجیم تولید شده به عمل آورده و توانسته با استفاده از علوم گوناگون مثل آمار، کامپیوتر، یادگیری ماشینی و..... به جستجوی دانش نهفته در داده‌ها پرداخته وعلم نوینی را به نام داده کاوی به وجود آورد.

کلمات کلیدی: داده کاوی (data mining)، سلامت (health)، کشف دانش (knowledge discovery)

مقدمه

امروزه در دانش پزشکی جمع‌آوری داده‌های فراوان در مورد بیماری‌های مختلف دارای اهمیت فراوانی است. مراکز پزشکی با مقاصد گوناگون به جمع‌آوری این داده‌ها می‌پردازند. تحقیق روی این داده‌ها و کشف نتایج و رابطه در مورد بیماری‌ها یکی از اهداف استفاده از این داده‌ها است. امروزه با توجه به ظهور نظام‌های اطلاعات یکپارچه و رشد فناوری اطلاعات این مهم بیش از پیش نمایان شده است. داده کاوی ازجمله پیشرفت‌های فناوری در راستای مدیریت داده است.

از سال ۱۹۵۰ به بعد که رایانه، در تحلیل و ذخیره‌سازی داده‌ها به کار رفت، حجم اطلاعات ذخیره‌شده در آن پس از حدود۲۰سال دو برابر شد و همزمان با پیشرفت فناوری اطلاعات، حجم داده‌ها در پایگاه داده‌ها هر دو سال یک‌بار، دو برابر شد و همچنان با سرعت بیشتری نسبت به گذشته حجم اطلاعات ذخیره‌شده بیشتر

۱ـ گروه مدیریت اطلاعات سلامت دانشگاه علوم پزشکی ایران، دانشگاه علوم پزشکی ارتش، تهران، ایران
۲ـ دانشکده پیراپزشکی، دانشگاه علوم پزشکی ارتش، تهران، ایران (*نویسنده مسئول)
تلفن: ۰۹۱۵۴۷۶۳۸۷۸ آدرس الکترونیک: mostafadamroodi1992@yahoo.com

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

و بیشتر می‌شود.

شدت رقابت‌ها در عرصه‌های علمی ازجمله سلامت به واسطه این هدف که (چگونه سازمان‌های سلامت هزینه‌ها را کاهش و کیفیت را افزایش دهند؟) اهمیت دسترسی به داده‌ها و اطلاعات را دو چندان کرده است، به عبارت دیگر در هر نوع برنامه بهبود کیفیت متمرکز بر بیمار داده‌ها قلب آن برنامه به حساب می‌آیند. (۱)

کشف دانش (Knowledge discovery) که بیش از یک دهه وارد محیط مالی و پزشکی شده است با هدف شناسایی سوءاستفاده‌های مالی وارد عرصه سلامت شد، اما به تدریج در حوزه بالینی نیز مورد استفاده قرار گرفت. صنعت سلامت به طور مستمر در حال تولید میزان زیادی داده می‌باشد و افرادی که با این نوع داده‌ها مواجه هستند دریافته‌اند که بین جمع‌آوری تا تفسیر داده شکاف وسیعی وجود دارد و داده کاوی ازجمله شیوه‌هایی است که می‌تواند این صنعت را ازتحلیل عمیق این داده‌ها بهره‌مند سازد و به توسعه تحقیقات پزشکی و تصمیم‌گیری‌های علمی در زمینه تشخیص و درمان منتج شود. (۲)

مواد و روش‌ها

مطالعه حاضر با جستجوی مروری در متون علمی مرتبط موجود در پایگاه‌های اطلاعاتی انگلیسی و فارسی شامل Magiran, Iranmedex, SID, Google Scholar, ovid, scopus, PubMed؛ با کلیدواژه‌های data minig و knowledge health, discovery مربوط به سال‌های ۱۹۹۸ تا ۲۰۱۳ انجام شد.

تعریف داده کاوی (Data Mining)

در دهه‌ی اخیر با تسهیل جمع‌آوری داده‌های پزشکی و دسترس‌پذیری حجم زیاد داده‌های الکترونیکی، اکتشاف دانش توسعه‌یافته است. امروزه پژوهشگران از طریق فرایند اکتشاف دانش، اقدام به شناسایی روابط بین متغیرها، شناسایی الگوها و تولید مدل‌های پیش‌بینی پیامد بیماری‌ها می‌کنند. الگوریتم‌های یادگیری ماشینی از ابزارهای مرسوم در فرایند اکتشاف دانش بوده که با استفاده از آن‌ها مدل‌های پیش‌بینی با دقت بالایی ایجاد می‌شوند. به استفاده از الگوریتم‌های یادگیری ماشین برای اکتشاف دانش در حجم عظیمی از داده‌ها، داده کاوی گفته می‌شود. (۳)

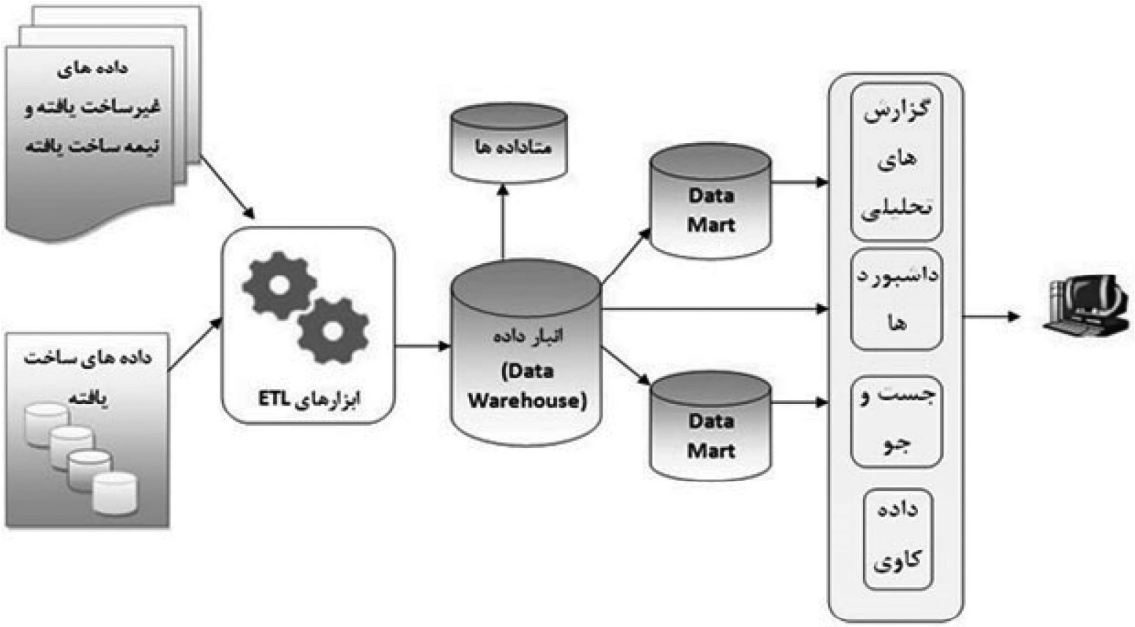
نگاهی به ترجمه لغوی داده کاوی به ما در درک بهتر این واژه کمک می‌کند. واژه لاتین Mine به معنای استخراج از منابع نهفته و با ارزش اطلاق می‌شود. ادغام این کلمه با Data به معنی داده، بر جستجویی عمیق از داده‌های قابل دسترس با حجم زیاد برای یافتن اطلاعات مفید که قبلاً نهفته بودند تأکید دارد. به عبارت دیگر داده کاوی استخراج اطلاعات مفید، مفهومی و ناشناخته از پایگاه داده می‌باشد. به لحاظ فنی، داده کاوی عبارت از فرآیندی است که در میان حوزه‌های گوناگون بانک‌های اطلاعاتی ارتباطی بزرگ، همبستگی‌ها یا الگوهایی را پیدا می‌کند. داده کاوی در حقیقت یکی از ابزارهای علم آمار به حساب می‌آید، اما گاهی دیده می‌شود که در برخی موارد آن را به عنوان شاخه‌ای جداگانه در نظر می‌گیرند و یا حتی سعی در بیان تفاوت‌های آن با آمار دارند. (۴)

روش‌های استخراج داده‌های سلامت عبارت‌اند از: تصویرسازی اطلاعات، تجزیه و تحلیل اطلاعات مرتبط، تجزیه و تحلیل منابع داده‌ها و طبقه‌بندی شبکه‌ها که تمام آن‌ها در پایگاه اطلاعاتی بیماران به کار گرفته می‌شود. این روش‌ها می‌توانند به شناسایی بیماران کمک کرده و فاکتورهای مهم موجود در بیماری را مشخص نمایند. در این ساختار فنون ساده تصویرسازی بیشترین استفاده را دارد که با دانش روز پزشکی سازگار است. (۵)

بیشتر افرادی که علم داده کاوی را مورد استفاده قرار می‌دهند افراد متخصصی در یک زمینه خاص علمی (پزشک، رادیولوژیست، مدیر فروش،...) بوده که نه تنها به داده‌های خاص خود دسترسی دارند بلکه به جمع‌آوری آن نیز اقدام می‌ورزند. این افراد ترجیح می‌دهند داده‌های خود را بهتر بشناسند و همچنین مایل‌اند دانش جدیدی را در رابطه با زمینه فعالیت خود کشف نمایند.

در علم پزشکی کشف و تشخیص به موقع بیماری‌ها می‌تواند از ابتلا به بسیاری از بیماری‌های مهلک نظیر سرطان جلوگیری نموده و باعث نجات زندگی مردم گردد. امروزه با پیشرفت‌های بیولوژیکی، توسعه تکنولوژی و استفاده از فناوری‌های روز و تجهیزات مدرن پزشکی، متخصصین قادرند تا به جمع‌آوری اطلاعات دقیق‌تری در مورد بیماران بپردازند که تحلیل آن‌ها به دلیل حجم بالای اطلاعات مشکل بوده و نیازمند فناوری‌های نوین می‌باشد که تکنولوژی‌های داده کاوی به کمک الگوریتم‌های قدرتمند خود به این مهم دست یافته است. هدف از روش‌های پیشگویی داده کاوی در پزشکی

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

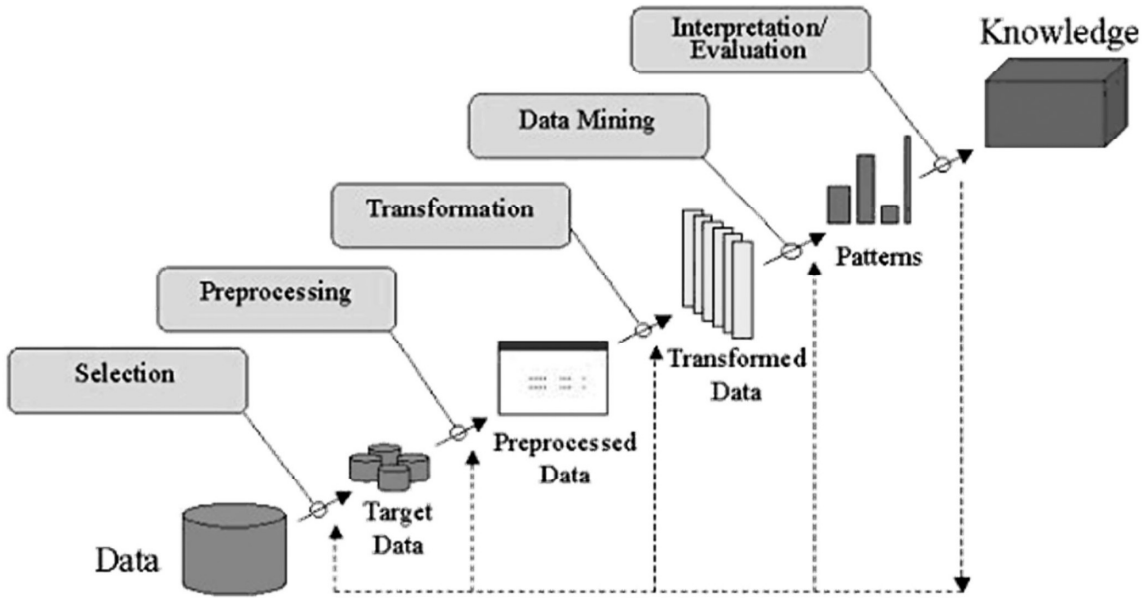


بالینی ساخت یک مدل پیش‌گویانه است که به پزشکان کمک می‌کند تا روش‌های پیشگیری، تشخیص و برنامه‌های درمانی خود را بهبود بخشند. (۶)

مراحل داده کاوی

به صورت ساده می‌توان بیان نمود که داده کاوی به استخراج دانش از حجم انبوهی از داده‌ها اطلاق می‌شود. بسیاری از افراد این واژه را مترادفی برای واژه‌ی کشف دانش می‌دانند، اما با توجه به شکل زیر داده کاوی در واقع مرحله‌ای از فرآیند کشف دانش تلقی می‌شود. مراحل موجود در فرآیند استخراج دانش در شکل یک نشان داده شده است و شامل مراحل زیر است:

- **جمع‌آوری داده‌ها:** در این مرحله پس از پالایش داده‌ها، چندین منبع داده‌ای در یک انبار داده یکپارچه قرار می‌گیرند.
- **انتخاب و آماده‌سازی داده‌ها:** در این قسمت داده‌های مرتبط انتخاب می‌گردد و به شکل مناسبی برای داده کاوی تبدیل می‌شوند.



شکل ۱- مراحل داده کاوی و کشف دانش از داده

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

- **داده کاوی:** فرآیندی که با به خدمت گرفتن روش‌های هوشمند در میان داده‌ها به دنبال الگوهای خاصی می‌گردد.
- **تفسیر و ارزشیابی الگوها:** از میان انبوهی از الگوها، با تعریف معیارهای متنوع الگوهای محدودی برای تفسیر و تحلیل انتخاب می‌شوند.
- **ارائه دانش:** در این مرحله با کمک ابزار بصری سازی و فن‌های مختلف داده کاوی، دانش کشف‌شده به کاربر و یا تحلیل‌گر ارائه می‌شود. (۷)

مهم‌ترین دلایل استفاده از داده کاوی (Wickramasinghe) و همکاران

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

۲ـ 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 آزمون قبلی بنا نهد، اما داده کاوی بدون تکیه بر فرضیه‌های پیشین به تحلیل داده‌ها می‌پردازد.

۳ـ 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.

داده کاوی از ساخت مدل‌های تحلیلی، دسته‌بندی و پیش‌بینی اطلاعات و ارائه نتایج با استفاده از ابزارهای مرتبط استفاده می‌کند. برای اینکه الگوریتم داده کاوی بتواند عمل استخراج دانش را به‌خوبی انجام دهد، نیاز به یک سری پیش‌پردازش‌ها بر روی داده‌ها و یک سری پس پردازش‌ها بر روی الگوهای استخراج شده دارد. (۸)

استراتژی‌های داده کاوی

به طور کلی هدف داده کاوی، یادگیری و آموختن از داده‌ها است و بر این اساس دو دسته 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 برای ما شناخته شده باشد. یافتن مدل‌های پیش‌بینی خطا در مطالبات بیمه‌ای یک موسسه 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 شیوه یادگیری فاقد نظارت، ویژگی‌ها و مدل‌های خطاهای مطالبات شناخته شده نیست، اما الگوها و خوشه‌های حاصل از داده کاوی منجر به کشفیات جدید می‌شود. (۹)

رایج‌ترین تکنیک‌های داده کاوی و کاربرد آن در حوزه سلامت

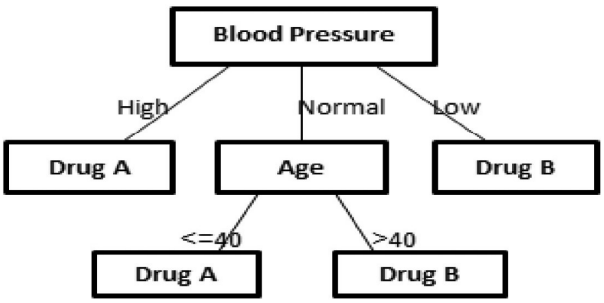
تکنیک‌های داده کاوی

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

- **درخت تصمیم‌گیری (Decisions Tree):** این تکنیک ساختاری شبیه به درخت دارد و مجموعه قواعدی که منجر به تصمیم‌گیری شده 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 و همچنین استفاده از درخت الگوریتم DRG (Diagnosis Related Groups)

(کدهای تشخیصی وابسته) در بازپرداخت هزینه‌های بیمه پزشکی سالمندان در ایالات‌متحده مثال کلاسیکی از این تکنیک است. شکل ۲–نمونه‌ای از درخت تصمیم‌گیری را برای مثالی از داده‌های پزشکی نشان می‌دهد.

در درخت بالا می‌بینیم که چگونه یک پزشک می‌تواند بر اساس صفات خاصه‌ی فشار خون و سن بیمار داروی مناسب جهت مداوای



شکل۲- درخت تصمیم‌گیری

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

او را تجویز کند. این مثال به خوبی نشان می‌دهد که چگونه یک درخت تصمیم برای نمایش یک مدل طبقه‌بندی استفاده می‌شود. (۷)

- **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 درباره علت رسیدن به این نتیجه خاص توضیحی ارائه نمی‌کند. برای مثال شبکه‌های عصبی قادرند مشخص کنند که چه نوع بیماری‌هایی احتمال دارد با یک بیماری همراه شود و با تحلیل تصاویر، نوارهای قلبی و سایر مشاهدات بالینی به تشخیص، درمان و تولید دارو کمک نمایند.
- **الگوریتم‌های ژنتیک (Genetic Algorithms):** تکنیک‌های بهینه‌ای برای ارتقای سایر الگوریتم‌های داده کاوی، به گونه‌ای که از بهترین مدل بر روی مجموعه داده‌ها استفاده کنند و می‌تواند برای یک بیماری خاص بهترین برنامه درمانی را تعیین کند.

نمونه‌هایی از کاربرد داده کاوی در پزشکی

۱ـ **کاربرد داده کاوی در بیماری‌های قلبی و عروقی:** رشد چشمگیر بیماری‌های قلبی و عروقی، اثرات و عوارض آن‌ها و هزینه‌های بالایی که بر جامعه وارد می‌کند، باعث شده که جامعه پزشکی به دنبال برنامه‌هایی جهت بررسی بیشتر، پیشگیری، شناسایی زود هنگام و درمان مؤثر آن باشند. از این‌رو با استفاده از داده کاوی و کشف دانش در سیستم مراکز قلب و عروق می‌توان دانش ارزشمندی را ایجاد کرد که این دانش کشف شده می‌تواند باعث بهبود کیفیت سرویس به وسیله مدیران مرکز شود و همچنین می‌تواند به وسیله پزشکان استفاده شود تا رفتار آینده بیماران قلبی و عروقی را از روی سابقه داده شده پیش‌بینی کند. تشخیص بیماری قلبی از روی ویژگی‌های گوناگون و نشانه‌ها، ارزیابی فاکتورهای ریسکی که باعث افزایش حمله قلبی می‌شود از مهم‌ترین کاربردهای داده کاوی و کشف دانش در سیستم بیماران قلبی و عروقی است. (۱۰)

۲ـ **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 بافت‌برداری از گردن رحم به منظور تشخیص سرطان گردن رحم از جمله این موارد است. گورانسکو (Gorunescu) به نقل از کانلاس (Canlas) نشان داد که داده کاوی می‌تواند به ادغام CAD (Computer Assisted Diagnosis) (تشخیص به کمک رایانه) و اندوسکوپی اولتراسونوگرافی در تشخیص غیر تهاجمی سرطان به عنوان یک شیوه جدید به کار رود.

۳ـ **داده کاوی در کنترل عفونت بیمارستانی:** در ایالات‌متحده سالانه دو میلیون نفر به عفونت‌های بیمارستانی مبتلا می‌شوند، لذا تمرکز زیادی برای شناسایی این بیماران صورت گرفته است. به عنوان مثال در ایالت آلباما 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 (LIS)) 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. (۱۱)

۴ـ **داده کاوی در تعیین نوع درمان:** به‌کارگیری داده کاوی 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 هاشمی نژاد 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 که میزان موفقیت بالاتری برای بیمار دارد را انتخاب می‌کند.

۵ـ **داده کاوی در پرونده الکترونیک سلامت:** در حال حاضر مطالعات متعددی مؤکد این مطلب است که تکنیک‌های داده کاوی ابزار مؤثری را برای شناسایی الگوهای مهم سلامت از درون پرونده‌های پزشکی فراهم می‌کنند.

پرونده‌های سلامت رایانه‌ای به واسطه دربرداشتن مجموعه‌ای از داده‌ها درباره تشخیص، درمان، اقدامات آزمایشگاهی و دارویی به طور بالقوه منبع غنی از دانش هستند و کشف دانش از انبوه داده به کمک داده کاوی صورت می‌گیرد.

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

۶ـ داده کاوی در رتبه‌بندی بیمارستان‌ها: رتبه‌بندی بیمارستان‌ها و برنامه‌های بهداشتی می‌تواند بر مبنای اطلاعات گزارش شده توسط ارائه دهندگان مراقبت باشد، بنابراین گزارش دهی استاندارد برای مقایسه معنی‌دار بیمارستان‌ها و رتبه‌بندی آن‌ها بسیار مهم است. ازجمله شیوه‌های استاندارد کردن این گزارشات استفاده از داده کاوی است. به عنوان مثال اگر کدهای ICD با تکنیک‌های داده کاوی مانند خوشه‌بندی همراه شود، می‌تواند به ایجاد گزارشاتی که با نرخ واقعی میزان ناخوشی‌ها، مرگ‌ومیر و سایر شاخص‌های کیفیت در رتبه‌بندی بیمارستان‌ها بیانجامد. (۸)

۷ـ تعیین نوع رفتار با بیماران و پیشگویی میزان موفقیت اعمال جراحی (۱۲)

۸ـ مدیریت درمان: با استفاده از داده کاوی می‌توان بیماری‌های مزمن و خطرناک را شناسایی کرد و نتیجه‌ی این کار مدیریت بهتر بر روی بیماران می‌باشد که می‌توانیم به این بیماری‌ها اولویت بیشتری اختصاص داده و در اورژانس‌ها و موقعیت‌هایی که زمان و منابع مورد نیاز در برخورد با تمام بیماران محدود می‌باشد، این بیماران را در اولویت اول قرار داد. حتی می‌توان تعداد موارد بستری در بیمارستان‌ها را کاهش داد و از پذیرش‌های بی جهت جلوگیری کرد. همچنین می‌توان عوارض دارویی و تداخلات دارویی در برخی از درمان‌ها را شناسایی کرد و برای درمان دیگر بیماران از نتیجه این الگوها استفاده نمود. مانند:

- **ObamaCare:** یک طرح برای اصلاح نظام خدمات درمانی و بیمه، که یکی از اصلاحات آن بهبود کیفیت درمانی بیمارستان‌ها، صرفه‌جویی در هزینه درمان و... است.
- **Arkansas Data Network:** به دنبال مراجعات و بستری مجدد، بهره‌برداری از منابع و مقایسه داده‌ها با انتشارات علمی پژوهشی در جهت تعیین بهترین گزینه درمان.
- **Group Cooperative Health:** انواع بیماران را طبق وضعیت پزشکی و دسترسی به منابع گروه‌بندی می‌کند و در نتیجه برای آموزش این گروه‌ها و جلوگیری یا مدیریت وضعیت بیماری بیماران، برنامه‌ریزی می‌کند.

۹ـ تشخیص تقلب و سوءاستفاده: برخی از افراد سودجو در خرید دارو از بیمه سوءاستفاده می‌کنند و داروهایی که خود نیاز ندارند را تحت پوشش بیمه تهیه می‌کنند. یکی از کاربردهای داده کاوی تشخیص این گونه سوءاستفاده‌ها می‌باشد.

۱۰ـ مطالعه انجام شده در مورد پیش‌بینی عوارض دیابت: با توجه به اینکه بیماری دیابت، به عنوان یک بیماری بسیار مزمن شناخته شده است و آسیب‌های جبران ناپذیری به اندام‌ها و اعضاء حیاتی بدن وارد می‌کند، استفاده از ابزارهای هوشمند داده کاوی می‌تواند برای بهبود روش‌های شناسایی و کنترل بیماری به پزشکان کمک بزرگی باشد. در تحقیق انجام شده با استفاده از الگوریتم‌های داده کاوی به دسته‌بندی بیماران دیابتی بر اساس عارضه مشاهده شده در آن‌ها پرداخته شد. عوارض این بیماری را بر اساس دو دسته میکرواسکولار و ماکرواسکولار دسته‌بندی شد. از بین الگوریتم‌های داده کاوی بهترین نتایج از الگوریتم درخت تصمیم به دست آمد که دقت مدل آن برابر ۸۹,۰۶ درصد و صحت مدل ۸۹,۷۴ درصد است. بیشترین پارامترهای تأثیرگذار بر روی عوارض بیماری میزان فشار سیستولیک، سن، سابقه خانوادگی و چربی مضر شناخته شد. با استفاده از قوانین ایجاد شده بر روی یک نمونه جدید با ویژگی‌های مشخص، می‌توان پیش‌بینی کرد که این فرد احتمالاً دچار چه نوع عارضه‌ای خواهد شد. با کنترل عوامل تأثیرگذار بر بروز عارضه درتر بیمار، می‌توان امیدوار بود از بروز عارضه تا حدی اجتناب شود و یا آن را به تعویق انداخت. (۱۳)

۱۱ـ مطالعه انجام شده در مورد پیش‌بینی عوارض عروق کرونر: یکی از شایع‌ترین بیماری‌های قلبی ـ عروقی، بیماری عروق کرونر است که بر اساس اطلاعات حاصل از یک مطالعه اپیدیمیولوژیک که به بررسی علت مرگ‌ومیر در سال ۱۳۸۸ پرداخته است، ۲۵,۶ درصد عامل منجر به مرگ محسوب می‌شود. بهترین روش ارزیابی بیماری عروق کرونر، آنژیوگرافی می‌باشد. با این وجود آنژیوگرافی یک روش گران و تهاجمی (Invasive) بوده و همراه با ریسک‌هایی همچون مرگ، سکته قلبی و مغزی می‌باشد. از این رو وجود نظام‌های پشتیبان تصمیم‌گیری در کنار روش‌های قبل از آنژیوگرافی برای کم کردن نتایج کاذب لازم به نظر می‌رسد. این سیستم‌ها با استفاده از تکنیک‌های داده کاوی می‌توانند به کشف الگوها در داده‌های پزشکی پرداخته و فرآیند تصمیم‌گیری را بهبود بخشند. در مطالعه انجام

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

شده برای پیش‌بینی بیماری عروق کرونر از تکنیک قدرتمند شبکه‌های عصبی استفاده گردید که مدل نهایی به دست آمده دارای دقت ۷۴/۱۹ درصد، ویژگی ۳۳/۲۵ درصد و حساسیت ۹۲/۴۱ درصد بود و توانست علاوه بر توانایی بالا در تشخیص افراد بیمار، تعداد قابل قبولی از افرادی که فاقد بیماری عروق کرونر بودند را نیز شناسایی کند. (۱۴)

۱۲ـ مطالعه انجام شده در مورد کاربرد داده کاوی در پیش‌بینی مرگ بیماران سوختگی: پیش‌بینی پیامد بیماری سوختگی بر اساس شرایط بیمار از اهمیت بالایی برخوردار است. از آنجایی که پیش‌بینی صحیح پیامد بیماری‌ها ابزار مناسبی برای متخصصین جهت تصمیم‌گیری بالینی آگاهانه به حساب می‌آید، پژوهشگران پزشکی همیشه به دنبال مدل‌هایی جهت پیش‌بینی پیامد بیماری‌ها و تعیین اهمیت فاکتورهای خطر بوده‌اند. در این مطالعه با استفاده از ابزارها و الگوریتم‌های داده کاوی از جمله مدل شبکه عصبی مصنوعی که دقت بالایی در پیش‌بینی پیامد بیماران سوختگی داشت، استفاده شد. درنهایت می‌توان گفت پیش‌بینی صحیح، تأثیر قابل توجهی در تخصیص بهینه‌ی منابع بیمارستانی داشته و در نتیجه باعث ارتقای کیفیت ارائه‌ی خدمات مراقبت می‌گردد. (۱۵)

چالش‌های داده کاوی در پزشکی: اکثر روش‌های داده کاوی استاندارد، می‌توانند تنها الگوها را نمایش دهند، و قادر نیستند الگوها و تمایلات را شرح دهند. در مقابل، پزشکی نیاز به توصیف الگوها دارد زیرا یک تفاوت اندک می‌تواند نتیجه (مرگ یا ادامه زندگی) را تغییر دهد. مانند تشخیص بیماری آنفلوانزا و آنتراکس (Anthrax) که نشانه‌های هر دو بیماری شبیه به هم هستند.

آماده کردن داده‌ها یا همان مرحله پیش‌پردازش (Preprocessing) داده‌ها یک مرحله بسیار دشوار است، چون داده‌ها باید کامل و بی‌نقص باشند تا داده کاوی بر روی داده‌ها درست انجام شود. چالش مهمی که با آن روبرو هستیم، کامل نبودن داده‌های پزشکی، استفاده از منابع مختلف داده‌ای و همچنین تلفیق اطلاعات و آماده‌سازی داده‌ها که امری بسیار دشوار است.

حفظ حریم خصوصی در پزشکی: موضوع حفظ حریم خصوصی موضوعی بسیار مهم در امر انتشار و تحویل داده‌ها پیش از داده کاوی می‌باشد. داده‌های پزشکی نیز شامل بیماری‌ها و اطلاعات حساس افراد می‌باشند که این گونه اطلاعات در فرآیند داده کاوی نباید فاش شوند. مکانیزم‌ها و مدل‌هایی برای حفظ حریم خصوصی در داده کاوی طراحی شده که می‌توانند از فاش شدن اطلاعات حساس جلوگیری به عمل آورند. این مدل‌ها و روش‌ها شامل دستکاری داده‌ها، تصادفی سازیِ (Randomization) نمونه گیری و حفظ گمنامی (Maintain Anonymity) می‌باشد. اما اگر کارایی داده کاوی از اهمیت بیشتری برخوردار باشد، سیاست‌های امنیتی در جهت حفظ حریم خصوصی باید به همراه داده کاوی انجام شود که شامل دستکاری در الگوریتم‌های داده کاوی برای محدودسازی داده کاوی و همچنین جلوگیری از خرابکاری از طریق حسابرسی پرس و جو (query auditing) می‌باشد. (۶۱)

جایگاه داده کاوی در ایران و سایر کشورهای جهان

در سال‌های اخیر علم داده کاوی محبوبیت خاصی در میهن عزیزمان ایران پیدا کرده است. حال می‌توان به برخی موارد از پروژه‌های اجرا شده و قابل اجرا در زمینه داده کاوی و دستاوردهای حاصل از آن در سازمان بیمه درمانی تأمین اجتماعی اشاره نمود:

- استفاده از تکنیک خوشه‌بندی و دسته‌بندی برای تعیین ویژگی مؤسسات با بدهی بالا، شناخت عملکرد کارفرمایان؛ استراتژی و تصمیمات آگاهانه با درجه اطمینان بالا جهت وصول بدهی‌ها که توسط مدیران ارشد سازمان صورت می‌پذیرد.
- پیش‌بینی و تخمین در داده کاوی به منظور پیش‌بینی الگوی رشد بدهی کارفرمایان، میزان درآمد، وصولی و هزینه‌های مرتبط با مدیریت امور مالی و ذخایر سازمان و همچنین در برآورد بودجه‌های مالی بسیار اثربخش خواهد بود.
- ارائه خدمات کوتاه مدت و بلندمدت سازمان با در نظر داشتن اصول مدیریت ارتباط با مشتریان (CRM) (Customer Relationship Management) با استفاده از فن‌آوری داده کاوی جهت جلب رضایت بیشتر مخاطبین (بیمه‌شدگان، مستمری بگیران و کارفرمایان) با شناخت بیشتر از داده‌های موجود میسر خواهد بود.
- کشف جرم و تقلب در سطح نظام‌های اطلاعاتی در دو حوزه بیمه‌ای و درمان، با استفاده از کاربردهای مؤثر داده‌کاوی و

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

تکنیک‌های کشف ناهنجاری‌ها و بررسی موارد پرت(Noise) قابل انجام می‌باشد. برای نمونه، می‌توان به استخراج الگوی مصرف داروهای خاص بر اساس سوابق داده‌های قبلی پرداخت و موارد ناهمگون و پرت را شناسایی نمود. علم داده کاوی در کشورهای توسعه‌یافته پیشرفت بیشتری داشته و به دنبال آن کارهای مرتبط با داده کاوی نیز گسترش بسیار زیادی پیدا کرده‌اند.

سایت Kdnugget یکی از مشهورترین پایگاه‌های اینترنتی است که در خصوص داده کاوی فعالیت می‌نماید. طی نظرسنجی صورت گرفته از این سایت در سال ۲۰۱۲، مشهورترین و پرکاربردترین حوزه‌های داده کاوی مشخص شد که حوزه بهداشت با۱۷ درصد در جایگاه سوم قرار داشت. (۱۷)

بحث

امروزه جمع‌آوری داده‌های فراوان در مورد بیماری‌های مختلف در علم پزشکی اهمیت بسیاری یافته و مراکز پزشکی با مقاصد گوناگون این داده‌ها را گردآوری می‌کنند. مطالعه بر روی این داده‌ها و کشف نتایج و روابط بیماری‌ها یکی از اهداف جمع‌آوری این داده‌ها است. در علم پزشکی، کشف و تشخیص به موقع بیماری‌ها می‌تواند از ابتلا به بسیاری از بیماری‌های مهلک نظیر سرطان جلوگیری نموده و باعث نجات زندگی مردم گردد.

با پیشرفت‌های بیولوژیکی، تکنولوژی و به‌کارگیری فناوری‌های نوین و تجهیزات مدرن پزشکی، متخصصین قادرند تا به جمع‌آوری داده‌های دقیق‌تر بیماران بپردازند. به دلیل حجم بالای داده‌ها تحلیل آن دشوار بوده و نیازمند فناوری‌های جدید می‌باشد، که تکنولوژی‌های داده کاوی با کمک الگوریتم‌های قدرتمند خود به این مهم دست یافته است.

کشف دانش از پایگاه‌های داده پزشکی به منظور تشخیص مؤثر در پزشکی بسیار مهم است. هدف از داده کاوی استخراج دانش از اطلاعات ذخیره شده در پایگاه داده و ایجاد شرح روشن و قابل فهم از الگوها است. از بررسی و مقایسه مطالعات صورت گرفته، به صورت کلی می‌توان نتیجه گرفت که هدف از روش‌های پیشگویی داده کاوی در پزشکی بالینی ساخت یک مدل پیش‌گویانه است که به پزشکان کمک می‌کند تا روش‌های پیشگیری، تشخیص و برنامه‌های درمانی خود را بهبود بخشند. نکته آخر اینکه استفاده از مدل‌های پیش‌بینی، هیچ‌گاه جایگزینی برای تصمیم‌گیری‌های پزشکان نمی‌باشد، بلکه هدف، فراهم‌سازی اطلاعات تکمیلی و پشتیبان برای ایشان در اتخاذ تصمیمات آگاهانه‌تر است.

نتیجه‌گیری

با توجه به اهمیت و حساسیت داده کاوی در پزشکی و همچنین نیاز به حرکت از پزشکی سنتی به سمت پزشکی مبتنی بر شواهد، در این مطالعه کاربردهای داده کاوی در حوزه سلامت مورد بررسی قرار گرفت.

به دنبال تولید مستمر و انبوه داده در سازمان‌های مراقبت سلامت و توسعه پایگاه‌های داده نسبت به دهه‌های گذشته نیازهای جدیدی را مانند خلاصه‌سازی خودکار داده و استخراج اطلاعات ذخیره‌شده از داده‌های خام به وجود آورده است که داده کاوی پزشکی نمونه‌ای از آن می‌باشد. داده‌ها در عصر امروزی عمده‌ترین دارایی سازمان‌های سلامت هستند و موفقیت آن‌ها منوط به جمع‌آوری، ذخیره و تحلیل این داده‌ها است. از این‌رو بسیاری از سازمان‌ها به داده کاوی روی آورده‌اند، چرا که به‌واسطه داده کاوی امکان کشف روابط و الگوهای مخفی بین داده‌ها و دستیابی به دانش نوین میسر خواهد شد.

داده کاوی در پزشکی و بیولوژی بخش مهمی از انفورماتیک زیست–پزشکی است ویکی از کاربردی‌ترین علوم رایانه در این علم بوده که در بیمارستان‌ها، کلینیک‌ها، آزمایشگاه‌ها و مراکز تحقیقاتی به کار گرفته شده است. امروزه سنجش سلامت بیشترین نیاز را به داده کاوی پیدا کرده است و حرکت از پزشکی سنتی به سمت پزشکی مبتنی بر شواهد ازجمله مواردی است که می‌تواند مؤکد این امر باشد.

پیشنهادات

در محیط رقابتی امروز، سازمان‌هایی که به‌واسطه استفاده از فن‌آوری‌های نوین همچون داده کاوی بتوانند داده‌ها را در راستای بهبود کیفیت سلامت به کار ببرند سریع‌تر به قله‌ی موفقیت خواهند رسید، لذا پیشنهاد می‌گردد تا سازمان‌های سلامت از این عرصه بازنمانند.

بسیاری از مراکز تحقیقاتی کشورمان دارای حجم زیادی از داده‌ها هستند که یا هرگز تحلیل نمی‌شوند و یا اگر هم تحلیل و به دانش

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

منتج شوند، به‌واسطه استفاده از شیوه‌های سنتی، امری مقطعی و زمان‌بر هستند؛ حال آنکه 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 ارائه خدمات با کیفیت به مشتریان سلامت بردارند.

References

1- Rogers G, Joyner E. Mining Your Data for Healthcare Quality Improvement [Online]. 2011 [cited 2011 Aug 8]; Availablefrom: URL: http: //www2.sas.com/proceedings/sugi22/EMERGING/PAPER139.PDf/

2- Englebardt SP, Nelson R. Health care informatics: an interdisciplinary approach. Philadelphia: Mosby; 2002. pp125.

3- Koh HC, Tan G. Data mining applications in healthcare. J Health In Manage 2005; 19 (2): 64-72.

4- Basic Epidemiology-file type ppt –Payman Salamati, MD, MPH, Assistant Professor of community Medicine Tehran University ofMedical Sciences 1387.

5- Medical Informatics, The internet Medicine ISSN 1436-9238, Print /ISSN 1464-5238online c 2000 Taylor& Francis Ltd. Available from: http: //www.tandf.co.uk /journals.

6- Hassanzadeh M, Razavi Ebrahimi SL. Comparison Classification of Data Mining Algorithms in Medical Sciences.

7- Esmaeili M. Concepts & Techniques datamining, Tehran: Niasedanesh Publisher;1391.

۸- مقدسی ح، حسینیان ف، اسدی ف، جهانبخش م. داده کاوی و کاربرد آن در سلامت. مجله مدیریت اطلاعات سلامت.۱۳۹۱، ۹ (۲): ۳۰۴-۲۹۷.

9- Obenshain MK. Application of data mining techniques to healthcare data. Infect Control. HospEpidemiol. 2004;25 (8): 690-5.

۱۰- کاج آبادی س، سرایی م، عسگری ص. داده کاوی پزشکی: راهکاری جهت کشف روابط بین عوامل خطرزای قلبی.

۱۱- نوروزی ف، طائفی همراه ع. مروری بر داده کاوی و بررسی شبکه‌های عصبی در چند زمینه‌ی مختلف.

12- Wickramasinghe N, Gupta JN, Sharma SK. Creating Knowledge-Based Healthcare Organizations. Hershey: IdeaGroup Inc (IGI). 2005.

۱۳- عامری ح، علیزاده س، برزگری الف. استخراج دانش از داده‌های بیماران دیابتی با استفاده از روش درخت تصمیم. مجله مدیریت سلامت. ۱۳۹۲، ۱۶ (۵۳)، ۵۸-۷۲.

۱۴- محمودی ع، عسکری مقدم ر، صادقیان س. مدل پیش‌بینی بیماری عروق کرونر با استفاده از شبکه‌های عصبی و گزینش متغییر مبتنی بر درخت رگرسیون و طبقه‌بندی. مجله دانشگاه علوم پزشکی شهرکرد. ۱۳۹۲، ۵ (۱۵)، ۴۷-۵۶.

۱۵- نبوتی الف، عزیزی الف، عباسی الف، وکیلی ارکی ح، زارعی ح، رضوی الف. کاربرد داده کاوی در پیش‌بینی مرگ بیماران سوختگی. مجله مدیریت اطلاعات سلامت. ۱۳۹۲، ۶ (۱۰)، ۷۸۹-۷۹۹.

16- Hian ChyeKoh, Gerald Tan. Data Mining Applications in Healthcare. Journal of Healthcare Information Management. Vol.19, No.2.

17- http: //www.tamin.ir/News/Item.

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]

Evaluation of Data Mining Applications in the Health System

Leila Gholamhosseini[1], Mostafa Damroodi*[1]

Abstract

Introduction: Extensive amounts of data stored in medical databases require the development of specialized tools for accessing the data, data analysis, knowledge discovery, and the effective use of the data. Data mining is one of the most important methods. The article sketches the used Data Mining techniques, and illustrates their applicability to medical diagnostic and prognostic problems.

Materials and Methods: The current study were searched English and Persian databases including Magiran, Iranmedex, SID, Google Scholar, OVID, Scopus, and PubMed by using keywords such as “Data Mining”, “Knowledge Discovery” and “Health”. Related articles were published and assessed during 1998-2013.

Findings: Data mining is a science that is searched automatically in the large amount of data for finding models and association rules in them where other statistical analysis cannot do that. The medical science is one of sciences that need to use of these tools for analyzing the large amount of data and creating predictive model with the new computation ways. In medical sciences, discovery and early diagnosis of the diseases can restrict the fatal diseases such as cancer and they save a people’s life. This research is shown that the data mining prediction provide necessary tools for the researcher and the physician to improve in the prevention of disease, diagnosis ways and their treatment programs.

Discussion and Conclusion: Nowadays, in the medical sciences, data collection of different diseases is very important. Recent development related to information technology & software has helped to have the better survey from producing massive data and could search the hidden knowledge in the data and create a new science by using different sciences including statistics, computers, and machine learning, and etc.

Keywords: data mining, health, knowledge discovery

1- (*Corresponding author) Department of Health Information Technology, Faculty of Paramedicine, AjA University of Medical Sciences, Tehran, Iran. E-mail: mostafadamroodi1992@yahoo.com

[Downloaded from jps.ajaums.ac.ir on 2025-11-12]