سال دهم، شماره یکم، تابستان ۱۳۹۴، مسلسل ۱۷



بررسی کاربردهای داده کاوی در نظام سلامت

ليلا غلامحسيني'، مصطفى دامرودى*'

چکیدہ

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 روش ها: 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.

یافته ها: داده کاوی به معنای جستجوی خودکار منابع داده ای بزرگ، جهت یافتن الگوها و وابستگی هایی است که تحلیل های ساده آماری قادر به انجام آن نیستند. یکی از زمینه هایی که نیازمند استفاده از این ابزار ها جهت تحلیل داده های وسیع و مدل سازی پیش گویانه با روش های محاسباتی جدید است، علم پزشکی می باشد. در علم پزشکی کشف و تشخیص به موقع بیماری ها می تواند از ابتلا به بسیاری از بیماری های مهلک نظیر سرطان جلوگیری نموده و موجب نجات زندگی مردم گردد. این مطالعه نشان می دهد که پیشگویی های داده کاوی ابزار های ضروری را برای محققان و پزشکان جهت بهبود در پیشگیری از بیماری ها، روش های تشخیصی و برنامه های درمانی فراهم می نمایند.

بحث و نتیجه گیری: امروزه در دانش پزشکی جمع آوری دادههای بیماریهای مختلف اهمیت فراوانی یافته است. پیشرفتهای دهه اخیر در ارتباط با فناوریهای اطلاعات و نرمافزار کمک شایانی به بررسیهای همهجانبه و کامل تر 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.

كلمات كليدى: داده كاوى (data mining)، سلامت (health)، كشف دانش (knowledge discovery)

مقدمه

امروزه در دانش پزشکی جمع آوری دادههای فراوان در مورد بیماریهای مختلف دارای اهمیت فراوانی است. مراکز پزشکی با مقاصد گوناگون به جمع آوری این دادهها می پردازند. 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 ذخیر مسازی دادهها به کار رفت، حجم اطلاعات ذخیر مشده در آن پس از حدود ۲۰سال 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 ذخیره شده بیشتر

۱_ گروه مدیریت اطلاعات سلامت دانشگاه علوم پزشکی ایران، دانشگاه علوم پزشکی ارتش، تهران، ایران

[.] دانشكده پيرايزشكى، دانشگاه علوم پزشكى ارتش، تهران، ايران («نويسنده مسئول) تلفن: ٩١۵۴۷۶۳۸۷۹ آدرس الكترونيك: mostafadamroodi1992@yahoo.com

و بیشتر می شود.

شدت رقابتها در عرصههای علمی از جمله 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 می آیند. (۱)

بر بیمتر عند عنب تا بر عنب تا بر عنی بیمتری یک رو از دمحیط کشف دانش (Knowledge discovery) که بیش از یک دهه وار د محیط 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 می باشد و افرادی که با این نوع داده ها مواجه هستند دریافته اند که بین جمع آوری تا تفسیر داده شکاف وسیعی وجود دارد و داده کاوی از جمله شیوه هایی است که می تواند این صنعت را از تحلیل عمیق این داده ها به رهمند سازد و به توسعه تحقیقات پزشکی و تصمیم گیری های 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, data داره ازههای SID, Google Scholar, ovid, scopus, PubMed مربوط به سالهای ۱۹۹۸ تا minig ند.

تعريف داده كاوى (Data Mining)

دردههی اخیرباتسهیل جمع آوری داده های پزشکی و دسترس پذیری حجم زیاد داده های الکترونیکی، اکتشاف دانش تو سعه یافته است. امروزه پژو هشگر ان از طریق فر ایند اکتشاف دانش، اقدام به شناسایی روابط بین متغیرها، شناسایی الگوها و تولید مدل های پیش بینی پیامد بیماری ها می کنند. الگوریتم های یادگیری ماشینی از ابزارهای مرسوم در فر ایند اکتشاف دانش بوده که با استفاده از آن ها مدل های پیش بینی با دقت بالایی ایجاد می شوند. به استفاده از الگوریتم های یادگیری ماشین برای اکتشاف دانش در حجم عظیمی از داده ها، داده کاوی گفته می شود. (۳)

نگاهی به ترجمه لغوی داده کاوی به ما در درک بهتر این واژه کمک می کند. واژه لاتین Mine به معنای استخراج از منابع نهفته و با ارزش اطلاق می شود. ادغام این کلمه با Data به معنی داده، بر جستجویی عمیق از داده های 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 می باشد. به لحاظ فنی، داده کاوی عبارت از فرآیندی است که در میان حوزه های گوناگون بانکهای اطلاعاتی ار تباطی بزرگ، همبستگی ها علم آمار به حساب می آید، اما گاهی دیده می شود که در برخی موارد آن را به عنوان شاخهای جداگانه در نظر می گیرند و یا حتی سعی در بیان تفاوت های آن با آمار دارند. (۴)

روش های استخراج داده های سلامت عبارت اند از: تصویر سازی اطلاعات، تجزیه و تحلیل اطلاعات مرتبط، تجزیه و تحلیل منابع داده ها و طبقه بندی شبکه ها که تمام آن ها در پایگاه اطلاعاتی بیماران به کار گرفته می شود. این روش ها می توانند به شناسایی بیماران کمک کرده و فاکتور های مهم موجود در بیماری را مشخص نمایند. در این ساختار فنون ساده تصویر سازی بیشترین استفاده را دارد که با دانش روز پزشکی ساز گار است. (۵)

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

در علم پزشکی کشف و تشخیص به موقع بیماری ها می تواند از ابتلا به بسیاری از بیماری های مهلک نظیر سرطان جلوگیری نموده و باعث نجات زندگی مردم گردد. امروزه با پیشرفت های بیولوژیکی، توسعه تکنولوژی و استفاده از فناوری های روز و تجهیزات مدرن پزشکی، متخصصین قادرند تا به جمع آوری اطلاعات دقیق تری در مورد بیماران بپردازند که تحلیل آن ها به دلیل حجم بالای اطلاعات مشکل بوده و نیازمند فناوری های نوین می باشد که تکنولوژی های داده کاوی به کمک الگوریتم های قدر تمند خود به این مهم دست یافته است. هدف از روش های پیشگویی داده کاوی در پزشکی





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

مراحل داده کاوی

به صورت ساده می توان بیان نمود که داده کاوی به استخراج دانش از حجم انبوهی از دادهها اطلاق می شود. بسیاری از افراد این واژه را مترادفی برای واژهی کشف دانش می دانند، اما با توجه به شکل زیر

داده کاوی در واقع مرحلهای از فرآیند کشف دانش تلقی می شود. مراحل موجود در فرآیند استخراج دانش در شکل یک نشان داده شده است و شامل مراحل زیر است:

- جمع آوری داده ها: در این مرحله پس از پالایش داده ها، چندین
 منبع داده ای در یک انبار داده یکپارچه قرار می گیرند.
- انتخاب و آمادهسازی دادهها: در این قسمت دادههای مرتبط انتخاب می گردد و به شکل مناسبی برای داده کاوی تبدیل می شوند.





- تفسیر و ارزشیابی الگوها: از میان انبوهی از الگوها، با تعریف معیارهای متنوع الگوهای محدودی برای تفسیر و تحلیل انتخاب می شوند.
- ارائه دانش: در این مرحله با کمک ابزار بصری سازی و فن های مختلف داده کاوی، دانش کشف شده به کاربر و یا تحلیل گر ارائه می شود. (۷)

مهمترین دلایل استفاده از داده کاوی (Wickramasinghe) و همکاران

- ۱- دلیل اصلی استفاده از داده کاوی، افزایش حجم داده های موجود
 و آتی سازمان ها است که نیاز به پردازش فراتر از رویکردهای
 سنتی دارد. برای انسان کنار آمدن با این حجم انبوه و در حال
 رشد داده ها با اتکای محض به شیوه های سنتی غیرممکن است.
- ۲_ ذهن انسان به هنگام تحلیل انبوه داده ها 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 فرضیه های پیشین به تحلیل داده ها می پردازد.
- ۳_ یکی از مهمترین مزایای داده کاوی به ویژه در حجم وسیع داده ها این است که داده کاوی نسبت به استخدام تیمی از افراد خبره بسیار کم هزینه تر است.

داده کاوی از ساخت مدلهای تحلیلی، دستهبندی و پیش بینی اطلاعات و ارائه نتایج با استفاده از ابزارهای مرتبط استفاده می کند. برای اینکه الگوریتم داده کاوی بتواند عمل استخراج دانش را به خوبی انجام دهد، نیاز به یک سری پیش پر دازش ها بر روی داده ها و یک سری پس پر دازش ها 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 استر اتژی های داده کاوی شامل یادگیری نظارت شده و یادگیری فاقد نظارت وجود دارد. شیوه های یادگیری نظارت شده زمانی به کار می رود که ارزش متغیر های ورودی برای ما شناخته شده باشد. یافتن مدل های پیش بینی 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 می رود، 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 می شود. (۹)

رایج ترین تکنیکهای داده کاوی و کاربرد آن در حوزه سلامت

تکنیکهای داده کاوی

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

- درخت تصمیم گیری (Decisions Tree): این تکنیک ساختاری شبیه به درخت دارد و مجموعه قواعدی که منجر به تصمیم گیری شده را توصیف می کند و سهولت تفسیر از ویژگی های مهم آن است. برای مثال درخت تصمیم گیری می تواند عوامل مؤثر بر میزان بقای پیوند کلیه را تعیین کند و محوینین استفاده از درخت الگوریتم DRG (Diagnosis Related)
- (کدهای تشخیصی وابسته) در بازپرداخت هزینههای بیمه پزشکی سالمندان در ایالاتمتحده مثال کلاسیکی از این تکنیک است. شکل۲-نمونهای از درخت تصمیمگیری را برای مثالی از دادههای پزشکی نشان میدهد.

در درخت بالا می بینیم که چگونه یک پزشک می تواند بر اساس صفات خاصه ی فشار خون و سن بیمار داروی مناسب جهت مداوای



او را تجویز کند. این مثال به خوبی نشان میدهد که چگونه یک درخت تصمیم برای نمایش یک مدل طبقه بندی استفاده می شود. (۷)

- شبکههای عصبی (Neural Networks): این تکنیک مدلهای پیشبینی غیر خطی تولید می کند که یاد می دهد چگونه یک الگو با یک پروفایل خاص قابل تطبیق است، اما درباره علت رسیدن به این نتیجه خاص توضیحی ارائه نمی کند. برای مثال شبکههای عصبی قادرند مشخص کنند که چه نوع بیماریهایی شبکههای عصبی قادرند مشخص کنند که چه نوع بیماریهایی 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 گونه ای که از بهترین مدل بر روی مجموعه داده ها استفاده کنند و می تواند برای یک بیماری خاص بهترین برنامه درمانی را تعیین کند.

نمونههایی از کاربرد داده کاوی در پزشکی

- ۱- 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 می کند، باعث شده که جامعه پزشکی به دنبال برنامههایی جهت بررسی بیشتر، پیشگیری، شناسایی زود هنگام و درمان مؤثر آن باشند. از این رو 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 میتوان دانش ارزشمندی را ایجاد کرد که این دانش کشف شده 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 میتران قلبی و عروقی را از روی سابقه داده شده پیش بینی 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)
- ۲_ داده کاوی در تشخیص های غیر تهاجمی: برخی از اقدامات تشخیصی و آزمایشگاهی برای بیماران، تهاجمی و هزینهبر و در عین حال رنج آور هستند. به عنوان مثال بافت برداری 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) (CAD (Computer) (میتواند به کمک رایانه) و اندوسکوپی اولتراسونو گرافی در تشخیص غیر تهاجمی سرطان به عنوان یک شیوه جدید به کار رود.

- **۳** داده کاوی در کنترل عفونت بیمارستانی: در ایالات متحده سالانه دو میلیون نفر به عفونت های بیمارستانی مبتلا می شوند، 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 می کند 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 (Laborator) مشخص می ساز دو ماهانه الگوهای جدید و جالب توجهی را مشخص می ساز دو ماهانه الگوهایی که تو سط کار شناسان کنتر ل عفونت مورد بررسی قرار می گیرد را تهیه می کند. سازندگان این سیستم دریافتهاند که ارتقای کنترل عفونت با سیستم داده کاوی حساس تر از سیستم کنترل عفونت سنتی عمل می کند. (۱۱)
- ۴- داده کاوی در تعیین نوع درمان: به کارگیری داده کاوی بر روی داده های 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 می شود. در این بیمارستان یک الگوریتم درختی وجود دارد که پزشک بر اساس آن درمانی که میزان موفقیت بالاتری برای بیمار دارد را انتخاب می کند.
- ۵- 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 می گیرد.

- ۹- 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 معنی دار بیمارستانها و رتبهبندی آنها بسیار مهم است. از جمله شیوههای استاندارد کردن این گزارشات استفاده از داده کاوی است. به عنوان مثال اگر کدهای ODI 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 سانجامد. (۸)
 - ۷_ تعیین نوع رفتار با بیماران و پیشگویی میزان موفقیت اعمال جراحی (۱۲)
- ۸ مدیریت درمان: با استفاده از داده کاوی می توان بیماریهای مزمن و خطرناک راشناسایی کرد و نتیجهی این کار مدیریت بهتر بر روی بیماران می باشد که می توانیم به این بیماری ها اولویت بیشتری اختصاص داده و در اورژانس ها و موقعیت هایی که زمان و منابع مورد نیاز در برخورد با تمام بیماران محدود می باشد، این بیماران را در اولویت اول قرار داد. حتی می توان تعداد موارد بستری در بیمارستان ها را کاهش داد و از پذیرش های بی جهت جلوگیری کرد. همچنین می توان عوارض دارویی و تداخلات دارویی در برخی از درمان ها را شناسایی کرد و برای درمان دیگر بیماران از نتیجه این الگوها استفاده نمود. مانند:
- DbamaCare: یک طرح برای اصلاح نظام خدمات درمانی و بیمه، که یکی از اصلاحات آن بهبود کیفیت درمانی بیمارستان ها، صرفه جویی 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.
- ۹_ 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.

- ۱۰ مطالعه انجام شده در مورد پیش بینی عوارض دیابت: با توجه به اینکه بیماری دیابت، به عنوان یک بیماری بسیار مزمن شناخته شده است و آسیبهای جبران نایذیری به اندامها و اعضاء حیاتی بدن وارد میکند، استفاده از ابزارهای هوشمند داده کاوی می تواند برای بهبود روش های 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 ماكرواسكولار دستهبندي شد. از بين الگوريتمهاي داده كاوي بهترين نتايج از الگوريتم درخت تصميم به دست آمد كه 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 ویژگی های مشخص، می توان پیش بینی کرد که این فرد احتمالاً دچار چه نوع عارضهای خواهد شد. با کنترل عوامل تأثیر گذار بر بروز عارضه درتر بيمار، مي توان اميدوار بود از بروز عارضه تا حدى اجتناب شود و يا أن را به تعويق انداخت. (١٣)
- ۱۱_مطالعه انجام شده در مورد پیش بینی عوارض عروق کرونر یکیازشایع ترین بیماری های قلبی –عروقی ، بیماری عروق کرونر است که بر اساس اطلاعات حاصل از یک مطالعه اپیدیمیولوژیک که به بررسی علت مرگومیر در سال ۱۳۸۸ پرداخته است، که به بررسی علت مرگومیر در سال ۱۳۸۸ پرداخته است، روش ارزیابی بیماری عروق کرونر، آنژیو گرافی می باشد. با روش ارزیابی بیماری عروق کرونر، آنژیو گرافی می باشد. با این وجود آنژیو گرافی یک روش گران و تهاجمی (Invasive) بوده و همراه با ریسکهایی همچون مرگ، سکته قلبی و مغزی می باشد. از این رو وجود نظام های پشتیبان تصمیم گیری در کنار روش های قبل از آنژیو گرافی برای کم کردن نتایج کاذب کنار روش های قبل از آنژیو گرافی برای کم کردن نتایج کاذب داده کاوی می توانند به کشف الگوها در داده های پزشکی پرداخته و فرآیند تصمیم گیری را بهبود بخشند. در مطالعه انجام

لیلا غلامحسینی و مصطفی دامرودی

شده برای پیش بینی بیماری عروق کرونر از تکنیک قدر تمند شبکه های 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)

۱۲ مطالعه انجام شده در مورد کاربرد داده کاوی در پیشبینی مرگ بیماران سوختگی: پیش بینی پیامد بیماری سوختگی بر اساس شرایط بیمار از اهمیت بالایی برخوردار است. از آنجایی که پیش بینی صحیح پیامد بیماریها ابزار مناسبی برای متخصصین جهت تصمیم گیری بالینی آگاهانه به حساب می آید، پژوه شگران پزشکی همیشه به دنبال مدل هایی جهت پیش بینی پیامد بیماری ها و تعیین اهمیت فاکتورهای خطر بودهاند. در این مطالعه با استفاده از ابزارها و الگوریتم های داده کاوی از جمله مدل شبکه عصبی مصنوعی که دقت بالایی در پیش بینی پیامد بیماران سوختگی داشت، استفاده شد. درنهایت می توان گفت پیش بینی صحیح، تأثیر قابل توجهی در تخصیص بهینه ی منابع بیمارستانی داشته و در نتیجه باعث ارتقای کیفیت ارائه ی خدمات مراقبت می گردد. (۱۵)

چالش های داده کاوی در پزشکی: اکثر روش های داده کاوی استاندارد، می توانند 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 می تواند 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 داده ها یا همان مرحله پیش پردازش (Preprocessing) داده ها یک مرحله بسیار دشوار است، چون داده ها باید کامل و بی نقص باشند تا داده کاوی بر روی داده ها درست انجام شود. چالش مهمی که با آن روبرو هستیم، کامل نبودن داده های پزشکی، استفاده از منابع مختلف داده ای و همچنین تلفیق اطلاعات و آماده سازی داده ها 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. داده های پزشکی نیز شامل بیماری ها و اطلاعات حساس افراد میباشند که این گونه اطلاعات در فرآیند داده کاوی نباید فاش شوند. مکانیزم ها و مدل هایی برای حفظ حریم خصوصی در داده کاوی طراحی شده که میتوانند از فاش شدن اطلاعات حساس جلوگیری به عمل آورند. این مدل ها و روش ها شامل دستکاری داده ها، تصادفی سازی (Randomization) نمونه گیری و حفظ گمنامی (Maintain Anonymity) میباشد. اما اگر کارایی داده کاوی از اهمیت بیشتری بر خور دار باشد، سیاست های امنیتی در جهت حفظ حریم خصوصی باید به همراه داده کاوی انجام شود که شامل دستکاری در الگوریتم های داده کاوی برای محدودسازی داده کاوی و همچنین جلوگیری از خرابکاری از طریق حسابر سی پر س 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 می توان 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 تكنيك خوشهبندى و دستهبندى براى تعيين ويژگى
 مؤسسات با بدهى بالا، شناخت عملكرد كارفرمايان؛ استراتژى
 و تصميمات آگاهانه با درجه اطمينان بالا جهت وصول بدهى ها
 كه توسط مديران ارشد سازمان صورت مى پذيرد.
- پیش بینی و تخمین در داده کاوی به منظور پیش بینی الگوی رشد
 بدهی کارفرمایان، میزان در آمد، وصولی و هزینه های مرتبط
 با مدیریت امور مالی و ذخایر سازمان و همچنین در بر آورد
 بودجه های 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 فن آوری داده کاوی جهت جلب رضایت بیشتر مخاطبین (بیمه شدگان، مستمری بگیران و کارفرمایان) با شناخت بیشتر از داده های 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

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 کر دهاند.

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 جمع آوری داده های فراوان در مورد بیماری های مختلف در علم پزشکی اهمیت بسیاری یافته و مراکز پزشکی با مقاصد گوناگون این داده ها را گردآوری می کنند. مطالعه بر روی این داده ها و کشف نتایج و روابط بیماری ها یکی از اهداف جمع آوری این داده ها است. در علم پزشکی، کشف و تشخیص به موقع بیماری ها می تواند از ابتلا به بسیاری از بیماری های مهلک نظیر سرطان جلو گیری نموده و باعث نجات زندگی مردم گردد.

با پیشرفتهای بیولوژیکی، تکنولوژی و بهکارگیری فناوریهای نوین و تجهیزات مدرن پزشکی، متخصصین قادرند تا به جمع آوری دادههای دقیق تر بیماران بپردازند. به دلیل حجم بالای دادهها تحلیل آن دشوار بوده و نیازمند فناوریهای جدید می باشد، 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 می توان نتیجه گرفت که هدف از روش های پیشگویی داده کاوی در پزشکی بالینی ساخت یک مدل پیش گویانه است که به پزشکان کمک می کند تا روش های پیشگیری، تشخیص و

برنامههای درمانی خود را بهبود بخشند. نکته آخر اینکه استفاده از مدلهای پیش بینی، هیچگاه جایگزینی برای تصمیم گیریهای پزشکان نمی باشد، بلکه هدف، فراهم سازی اطلاعات تکمیلی و پشتیبان برای ایشان در اتخاذ تصمیمات آگاهانه تر است.

نتيجه گيري

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 ذخیره شده از دادههای خام به وجود آورده است که داده کاوی پزشکی نمونهای از آن میباشد. دادهها در عصر امروزی عمده ترین دارایی سازمانهای سلامت هستند و موفقیت آنها منوط به جمع آوری، ذخیره و تحلیل این دادهها است. از این رو بسیاری از سازمانها به داده کاوی روی آوردهاند، چرا که به واسطه 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 کاربردی ترین علوم رایانه در این علم بوده که در بیمارستان ها، کلینیک ها، آزمایشگاه ها 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 از جمله مواردی است که می تواند مؤکد این امر باشد.

پیشنهادات

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

بسیاری از مراکز تحقیقاتی کشورمان دارای حجم زیادی از دادهها هستند که یا هرگز تحلیل نمی شوند و یا اگر هم تحلیل و به دانش لیلا غلامحسینی و مصطفی دامرودی

۴١

گامهای جدیدی را در پیش گیری، تشخیص، درمان و ارائه خدمات با کیفیت به مشتریان سلامت بر دارند.

References

- 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/
- Englebardt SP, Nelson R. Health care informatics: an interdisciplinary approach. Philadelphia: Mosby; 2002. pp125.
- 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.
- Hassanzadeh M, Razavi Ebrahimi SL. Comparison Classification of Data Mining Algorithms in Medical Sciences.
- 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.

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

- 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.
- 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 داده های بیماران دیابتی با استفاده از روش درخت تصمیم. مجله مدیریت سلامت. ۱۳۹۲، ۱۶ (۵۳)، ۸۸–۷۲.

- ۱۴ محمودیع، عسکری مقدم ر، صادقیان س. مدل پیش بینی بیماری عروق کرونر با استفاده از شبکههای عصبی و گزینش متغییر مبتنی بر درخت رگرسیون و طبقهبندی. مجله دانشگاه علوم پزشکی شهر کرد. ۱۳۹۲، ۵ (۱۵)، ۴۷–۵۶.
- ۱۵ نبوتی الف، عزیزی الف، عباسی الف، وکیلی ارکی ح، زارعی ح، رضوی الف. کاربرد داده کاوی در پیش بینی 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.

Evaluation of Data Mining Applications in the Health System

Leila Gholamhosseini1, 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