

بررسی پتانسیل داده کاوی در پیش بینی بیماری قلبی: مروری

محمد سیاوشی^۱، علی وحیدی سبزواری^۲، فریده امامی^۳

^۱گروه پرستاری، دانشگاه علوم پزشکی نیشابور، نیشابور، ایران
siavoshi_mohammad@yahoo.com

^۲دانشجوی دکتری مدیریت، دانشگاه تهران، ایران
vahidiali86@ut.ac.ir

^۳کارشناس پرستاری، بهداشت و درمان شمال شرق کشور
fa.emami03@gmail.com

چکیده

داده کاوی به فرایند محاسباتی بررسی مجموعه داده های وسیع برای استخراج بینش معنادار اشاره دارد. ابزارهای مورد استفاده در داده کاوی قادر به پیش بینی روندهای آینده هستند و سازمان ها را قادر می سازند تا در تصمیم گیری آگاهانه و فعال شرکت کنند. با استفاده از این ابزارها، شرکت ها می توانند سؤالات تجاری را که به طور سنتی نیازمند زمان و تلاش قابل توجهی است، حل کنند. حجم وسیعی از داده های تولید شده برای پیش بینی بیماری های قلبی، اغلب برای رویکردهای تحلیلی مرسوم، بسیار پیچیده و فراوان است که نمی توان آنها را به طور مؤثر مدیریت کرد. داده کاوی هم متدولوژی و هم چارچوب تکنولوژیکی لازم برای تبدیل این مجموعه داده های گسترده به اطلاعات عملی برای اهداف تصمیم گیری را ارائه می دهد. هنگام استفاده از تکنیک های داده کاوی، پیش بینی بیماری ها می تواند سریع تر و بادقت بیشتری تکمیل شود. این مقاله به بررسی مطالعات مختلفی می پردازد که یک یا چند الگوریتم داده کاوی را در پیش بینی بیماری قلبی پیاده سازی می کنند. قابل ذکر است، نتایج با استفاده از شبکه های عصبی در برخی مطالعات نزدیک به ۱۰۰٪ دقت دارند؛ بنابراین، استفاده از الگوریتم های داده کاوی نتایج کارآمد و قابل اعتمادی را در تجزیه و تحلیل پیش بینی به همراه دارد. ترکیب تکنیک های داده کاوی در تجزیه و تحلیل داده های درمان بیماری قلبی می تواند به سطوح عملکردی قابل مقایسه با آنچه در تشخیص این بیماری به دست می آید منجر شود.

کلمات کلیدی: بیماری قلبی، داده کاوی، تکنیک های داده کاوی.

۱ مقدمه

هدف اصلی مقاله ما، کشف تکنیک های مختلف داده کاوی مورد استفاده در پیش بینی بیماری قلبی از طریق استفاده از ابزارهای مختلف داده کاوی است. قلب یک عضو حیاتی در بدن انسان است و عملکرد کارآمد آن برای سلامت کلی حیاتی است. هرگونه اختلال در عملکرد آن می تواند بر سایر سیستم های بدن از جمله مغز

و کلیه‌ها تأثیر منفی بگذارد. بیماری قلبی به شرایطی اطلاق می‌شود که عملکرد قلب را مختل می‌کند و عوامل متعددی در افزایش خطر ابتلا به چنین بیماری‌هایی نقش دارند. در حال حاضر، بیماری قلبی عامل اصلی مرگ و میر در سراسر جهان است. سازمان بهداشت جهانی (WHO) پیش‌بینی کرده است که سالانه حدود ۱۲ میلیون مرگ به دلیل مشکلات قلبی رخ می‌دهد. تنها در سال ۲۰۰۸، بیماری قلبی جان ۳.۱۷ میلیون نفر را گرفت. نگران کننده است که بیش از ۸۰ درصد از مرگ و میرهای جهانی را می‌توان به بیماری قلبی نسبت داد. WHO پیش‌بینی می‌کند که تا سال ۲۰۳۰، نزدیک به ۶.۲۳ میلیون نفر به بیماری قلبی تسلیم خواهند شد، همان‌طور که در [۱۰] نشان داده شده است. استفاده از تکنیک‌های داده‌کاوی برای پیش‌بینی، بینش دقیقی از این وضعیت پزشکی به ما می‌دهد. سیستم هوشمند پیش‌بینی بیماری‌های قلبی (IHDPS) قادر به کشف و استخراج بینش‌های پنهان مربوط به بیماری‌های قلبی از یک پایگاه داده تاریخی جامع است. این سیستم می‌تواند به پرسش‌های پیچیده مربوط به تشخیص بیماری‌های قلبی رسیدگی کند و از این طریق به متخصصان مراقبت‌های بهداشتی در تصمیم‌گیری آگاهانه بالینی که فراتر از قابلیت‌های سیستم‌های پشتیبانی تصمیم‌گیری سنتی است، کمک می‌کند. این مقاله تجزیه و تحلیلی از تکنیک‌های مختلف داده‌کاوی را ارائه می‌دهد که در جداول به تفصیل شرح داده شده است، که برای تحلیلگران و پزشکان در دستیابی به تشخیص دقیق بیماری قلبی مفید بوده است.

عوامل خطر بیماری قلبی

سابقه خانوادگی در ابتلا به بیماری قلبی: اکثر مردم می‌دانند که بیماری قلبی می‌تواند در خانواده‌ها ارثی باشد. اگر فردی در خانواده‌اش سابقه‌ی بیماری‌های وجود داشته باشد، ممکن است او نیز در معرض خطر ابتلا به حمله قلبی، سکته مغزی و سایر بیماری‌های سخت باشد.

سیگار کشیدن: سیگار کشیدن دلیل اصلی حمله قلبی، سکته مغزی و سایر بیماری‌های شریانی محیطی (تنگی عروق) است. تقریباً ۴۰٪ از افرادی که دخانیات مصرف می‌کنند، به دلیل بیماری‌های قلبی و تنفسی، فوت می‌کنند. در افراد سیگاری، خطر ابتلا به حمله قلبی پس از یک سال سیگار نکشیدن، به سرعت کاهش می‌یابد.

کلسترول: سطوح غیر طبیعی لیپیدها (چربی‌ها) در خون، فاکتور خطر بیماری‌های قلبی هستند. کلسترول یک ماده‌ی نرم و مومی شکل است که در میان لیپیدهای موجود در جریان خون و در تمام سلول‌های بدن یافت می‌شود. سطح بالای تری‌گلیسیرید (شایع‌ترین نوع چربی در بدن)، به همراه سطح بالای LDL (لیپوپروتئین کم چگالی) کلسترول را افزایش می‌دهد و افزایش کلسترول باعث تصلب شریان می‌شود که خطر ابتلا به بیماری‌های قلبی را افزایش می‌دهد.

فشار خون بالا: فشار خون بالا که تحت عنوان HBP و یا فشار خون شناخته شده است، یک وضعیت پزشکی است که غالباً نیز اشتباه اندازه‌گیری می‌شود. فشار خون بالا خطر آسیب به دیواره‌های عروق خونی

ما را افزایش می‌دهد؛ همچنین خطر ابتلا به حمله قلبی یا سکته مغزی، نارسایی قلبی، نارسایی کلیه و بیماری‌های عروقی محیطی را نیز افزایش می‌دهد.

چاقی: اصطلاح چاقی برای توصیف وضعیت سلامتی هر فردی استفاده می‌شود که وزن او به میزان قابل توجهی بالاتر از وزن ایده‌آل خود است. چاق شدن برای هر فردی، خطر بروز مشکلات سلامتی مانند بیماری قلبی، سکته مغزی، فشار خون بالا، دیابت و موارد دیگر را افزایش می‌دهد.

عدم انجام تمرینات ورزشی: عدم انجام تمرینات ورزشی، یک عامل خطر برای ایجاد بیماری عروق کرونر (CAD) است. عدم انجام ورزش بدنی، خطر ابتلا به CAD را افزایش می‌دهد، زیرا خطر ابتلا به دیابت و فشار خون بالا را نیز افزایش می‌دهد.

۲ بررسی ادبیات پژوهش

بیماری قلبی یک اصطلاح است که به تعداد بسیاری از شرایط پزشکی مرتبط با قلب اختصاص داده می‌شود. این شرایط پزشکی، شرایط غیرطبیعی سلامتی را توصیف می‌کنند که به‌طور مستقیم بر روی قلب و تمام قسمت‌های آن تاثیر می‌گذارد. امروزه، بیماری‌های قلبی یک مشکل مهم از نظر سلامتی هستند. هدف این مقاله بررسی تکنیک‌های مختلف داده-کاوی است، تکنیک‌هایی که در سال‌های اخیر برای پیش‌بینی بیماری‌های قلبی ارائه شده‌اند. جدول ۱ روش‌های مختلف داده‌کاوی که برای تشخیص بیماری قلبی مورد استفاده قرار گرفته‌اند، و بر روی مجموعه داده‌های مختلف از بیماری‌های قلبی اعمال شده‌اند را نشان می‌دهد. در برخی از مقالات گفته شده است که آنها تنها از یک روش برای تشخیص بیماری‌های قلبی استفاده می‌کنند که در پژوهش شهاب و همکارانش [۱۲] و کارلوس و همکارانش [۵] و غیره نیز چنین است. اما در پژوهش‌های دیگر، از بیش از یک تکنیک داده‌کاوی برای تشخیص بیماری‌های قلبی استفاده می‌شود؛ که در پژوهش خانم اسحاق و همکارانش [۳]، MA.JABBAR و همکارانش [۲] و Shantakumar و همکارانش [۷] نیز چنین است.

داده‌کاوی

تمرکز اصلی داده‌کاوی بر روی تحلیل داده‌ها می‌باشد؛ ابزارها و تکنیک‌های داده‌کاوی، برای یافتن الگوهای موجود در مجموعه داده‌ها مورد استفاده قرار می‌گیرند. هدف اصلی داده‌کاوی این است که به‌طور خودکار الگوهای موجود در داده‌ها را با حداقل داده‌های ورودی توسط کاربر و حداقل تلاش بیابد. داده‌کاوی یک ابزار قدرتمند است که می‌تواند در مورد روند آینده بازار تصمیم‌گیری کرده و آن را پیش‌بینی نماید. ابزار و تکنیک‌های داده‌کاوی را می‌توان در زمینه‌های مختلف و در اشکال مختلف، به‌طور موفقیت‌آمیزی مورد استفاده قرار داد. اکنون، بسیاری از سازمان‌ها از داده‌کاوی به عنوان ابزاری برای مقابله با محیط‌های رقابتی و برای تحلیل داده‌ها استفاده می‌کنند. با استفاده از ابزار و تکنیک‌های کاوش^۱، می‌توان با ارزیابی روندهای

¹Mining

جدول ۱: این جدول تکنیک‌های مختلف داده کاوی را که برای تشخیص بیماری‌های قلبی مورد استفاده قرار گرفته‌اند را نشان می‌دهد؛ این تکنیک‌ها بر روی مجموعه داده‌های مختلف بیماری‌های قلبی، اعمال شده‌اند.

خصوصیت	تکنیک مورد استفاده	سال	نویسنده
۲۵	قوانین انجمنی	۲۰۰۱	کارلوس و همکاران
۱۳	طبقه‌بندی	۲۰۱۱	K. Usha Rani
	شبکه‌های عصبی		
۱۴	اپریوری	۲۰۱۳	Jesmin Nahar و همکاران
	اپریوری پیش بینی شده		
	ترتیبوس		
۱۴	الگوریتم ژنتیک	۲۰۰۸	Latha و همکاران
	CANFIS		
۱۴	خوشه‌بندی	۲۰۱۱	Majabbar و همکاران
	کاوش قواعد انجمنی		
	شماره ترتیب		
۱۵	درخت تصمیم	۲۰۱۳	خانم اسحاق و همکاران
	شبکه‌های عصبی		
	بیزین ساده		
	(EVAR)	۲۰۱۲	Nan-Chen و همکاران
	یادگیری ماشین		
	مارکوف بلنکت		
	شبکه عصبی مصنوعی	۲۰۱۲	Oleg و همکاران
	پلیمورفسم ژنتیکی		
۱۵	بیزین ساده	۲۰۱۲	Shadab و همکاران
۱۳	MAFIA	۲۰۰۹	Shantakumar و همکاران
	خوشه‌بندی		
	K-Means		

مختلف و الگوی بازار و تولید سریع و تحلیل موثر روند بازار، زمینه‌های مختلف کسب و کار را به راحتی کشف کرد. داده کاوی ابزار بسیار مفیدی برای تشخیص بیماری‌ها می باشد.

تکنیک‌های مورد استفاده در داده کاوی

وابستگی. وابستگی^۲، یکی از شناخته شده ترین تکنیک‌های داده کاوی است. در این تکنیک، یک الگوی براساس رابطه میان یک آیتم خاص و سایر آیتم‌های دیگر، در همان تراکنش، کشف می شود. به عنوان مثال، تکنیک وابستگی در پیش بینی بیماری‌های قلبی مورد استفاده قرار می گیرد؛ زیرا این تکنیک از ارتباط میان صفات مختلف، برای تحلیل و دسته بندی بیماری‌ها استفاده می کند؛ برای پیش بینی بیماری، می بایست تمام فاکتورهای خطر در نظر گرفته شوند.

طبقه بندی. طبقه بندی یک تکنیک کلاسیک داده کاوی است که براساس یادگیری ماشین عمل می کند. اساسا از این تکنیک برای طبقه بندی هر یک از اقلام موجود در یک مجموعه داده، در مجموعه ها، کلاس ها یا گروه های از پیش تعریف شده، استفاده می شود. تکنیک طبقه بندی از روش های ریاضیاتی مانند درخت تصمیم، برنامه نویسی خطی، شبکه عصبی و آمار استفاده می کند.

خوشه بندی. خوشه بندی یک روش داده کاوی است که با استفاده از روش های اتوماتیک، از میان اشیائی با ویژگی های مشابه، خوشه ی معنادار یا مفیدی را ایجاد می کند. تکنیک های خوشه بندی، متفاوت از تکنیک های طبقه بندی هستند؛ تکنیک های خوشه بندی، کلاس هایی را تعریف می کنند و اشیاء را در آنها قرار می دهند؛ در حالی که، در تکنیک های طبقه بندی، اشیاء به کلاس هایی که از پیش تعریف شده اند، تخصیص داده می شوند. به عنوان مثال، برای پیش بینی بیماری قلبی با استفاده از خوشه بندی، ما یک خوشه را در نظر می گیریم یا اینکه می گوئیم این فهرست از بیماران دارای یک عامل خطر مشابه هستند. این بدین معنی است که لیست جداگانه ای نیز از بیماران مبتلا به قند خون بالا و عوامل مرتبط با آن به وجود آورده ایم.

پیش بینی. پیش بینی، یک عنوان ضمنی است که به یکی از تکنیک های داده کاوی نسبت داده شده است؛ این تکنیک رابطه ی میان متغیرهای مستقل و رابطه ی میان متغیرهای وابسته و مستقل را کشف می کند. به عنوان مثال، تکنیک تحلیل پیش بینی می تواند در بازار فروش برای پیش بینی سود فروش در آینده استفاده شود؛ اگر ما فروش را یک متغیر مستقل در نظر بگیریم، سود می تواند یک متغیر وابسته باشد. سپس براساس تاریخچه ی داده های فروش و سود، می توان یک منحنی رگرسیون مناسب رسم کرد و از آن برای پیش بینی سود استفاده کرد.

داده کاوی در زمینه های مختلفی مورد استفاده قرار گرفته است از جمله: **هوش تجاری:** مجموعه ای از نظریه ها، روش ها، معماری ها و فناوری هایی است که داده های خام را به اطلاعات معنی دار و مفید، برای اهداف تجاری، تبدیل می کند. هوش تجاری می تواند حجم عظیمی از داده های بدون ساختار را برای کمک

²Association

جدول ۲: جدول ارائه شده، مقایسه‌ای میان روند استخراج داده‌های مختلف از گذشته تا آینده را ارائه می‌دهد.

منابع محاسباتی	فرمت‌های داده	الگوریتم‌ها یا تکنیک‌های استفاده شده	روندهای داده‌کاوی
تکامل 4G PL و انواع تکنیک‌های مرتبط با آن	داده‌های عددی و داده‌های ساختاریافته‌ی ذخیره‌شده در پایگاه‌داده‌های سنتی	تکنیک‌های آماری، تکنیک‌های یادگیری ماشین	گذشته
شبکه‌های دارای سرعت بالا، دستگاه‌های ذخیره‌سازی بالا-پایین ^۱ و دستگاه‌های ذخیره‌سازی موازی، رایانش توزیع‌شده و غیره	فرمت‌های داده‌ای ناهمگن؛ شامل: داده‌های ساختاریافته، داده‌های نیمه‌ساختاریافته و داده‌های بدون ساختار	تکنیک‌های آماری، تکنیک‌های یادگیری ماشین، هوش مصنوعی، تکنیک‌های شناسایی الگو	حال
تکنیک‌های چندعاملی و رایانش ابری	اشیاء داده‌ای پیچیده؛ شامل: داده‌های دارای ابعاد بالا، جریان‌های داده‌ای دارای سرعت بالا، دنباله، نویز در سری‌های زمانی، گراف، اشیاء چند نمونه‌ای و غیره	تکنیک‌های محاسبات نرم، شامل: منطق فازی، شبکه‌های عصبی و برنامه‌نویسی ژنتیک	آینده

به شناسایی، توسعه و ایجاد فرصت‌های جدید مدیریت کند. به عبارت ساده، هوش تجاری، حجم عظیمی از اطلاعات را تفسیر می‌کند. استفاده از فرصت‌های جدید و اجرای یک راهبرد موثر می‌تواند در درازمدت، مزیت رقابتی بازار و ثبات بازار را فراهم آورد. فناوری‌های هوش تجاری، امکان ارائه‌ی دیدگاه‌های تاریخی، دیدگاه‌های جاری و پیش‌بینی دیدگاه‌های آینده را برای عملیات تجاری فراهم می‌آورد. **ورزش:** ورزش، زمینه‌ی ایده‌آلی برای استفاده از ابزارها و تکنیک‌های داده‌کاوی است. در دنیای ورزش، آمار زیادی برای هر بازیکن، تیم، بازی و فصل جمع‌آوری می‌شود. سازمان‌های ورزشی می‌توانند از داده‌کاوی برای تحلیل آماری، کشف الگو و نیز پیش‌بینی نتایج استفاده کنند. الگوهای موجود در داده‌ها اغلب برای پیش‌بینی رویدادهای آینده مفید هستند. از داده‌کاوی می‌توان برای کشف، پیش‌بینی عملکرد، انتخاب بازیکنان، مربی‌گری و آموزش و برنامه‌ریزی استراتژی استفاده کرد. **تحلیل عملکرد دانشجویان:** برای ارزیابی عملکرد دانشجویان، از طبقه‌بندی استفاده می‌شود و با توجه به رویکردهای بسیاری که برای طبقه‌بندی داده‌ها وجود دارد، ما در اینجا از روش درخت تصمیم استفاده می‌کنیم. اطلاعاتی مانند حضور در کلاس، آزمون کلاسی، سمینار و واحدهای انتخاب شده، همگی اطلاعاتی هستند که از سیستم مدیریت دانشجویان جمع‌آوری شده‌اند و برای پیش‌بینی عملکرد دانشجویان در پایان ترم، مورد استفاده قرار می‌گیرند. این مقاله به بررسی دقت تکنیک‌های درخت تصمیم برای پیش‌بینی عملکرد دانشجویان می‌پردازد. **صنعت مخابرات:** خدمات مخابراتی، از خدمات ارتباطات صوتی محلی و رادور، به فاکس، پیجر، تلفن‌های همراه و پست الکترونیکی توسعه یافته‌اند. در حال حاضر خدمات مخابراتی با رایانه، اینترنت، شبکه و دیگر فناوری‌های ارتباطی ادغام شده است. باتوجه به

پیشرفت‌های فناوری‌های مخابراتی و به‌کارگیری این فناوری‌ها، تکنیک‌های داده‌کاوی می‌توانند با استفاده از این فناوری‌ها، نتایج موثری را تولید کنند. داده‌کاوی به شناسایی الگوهای ارتباطات مخابراتی و شناسایی فعالیت‌های تقلبی کمک می‌کند؛ همچنین به استفاده بهتر از منابع و بهبود کیفیت خدمات، نیز کمک می‌کند. **صنعت خرده‌فروشی:** داده‌کاوی نیز نقش مهمی در صنعت خرده‌فروشی بازی می‌کند. صنایع خرده‌فروشی شامل حجم زیادی از اطلاعات هستند، که شامل اطلاعات حمل‌ونقل، فروش و مصرف کالا و خدمات است. با توجه به افزایش خرید و فروش در کسب و کار، این داده‌ها به سرعت در حال افزایش هستند. امروزه با رشد شرکت‌ها و همچنین بهبود تجربه فروش آنلاین، تجارت الکترونیک توسعه می‌یابد. تجارت الکترونیک، خرید و فروش محصولات، خدمات و اطلاعات را از طریق شبکه‌های کامپیوتری از جمله اینترنت توصیف می‌کند.

ابزارهای رایگان برای داده‌کاوی

ابزار وکا: وکا^۳، یک سیستم داده‌کاوی است که توسط دانشگاه وایکاتو در نیوزیلند توسعه داده شده است و الگوریتم‌های داده‌کاوی را با استفاده از زبان جاوا پیاده‌سازی می‌کند. وکا، پیشرفته‌ترین ابزار برای توسعه تکنیک‌های یادگیری ماشین و استفاده از آنها در زمینه‌های داده‌کاوی در دنیای واقعی می‌باشد. وکا، مجموعه‌ای از الگوریتم‌های یادگیری ماشین برای کارهای داده‌کاوی است. الگوریتم‌ها مستقیماً به یک مجموعه داده اعمال می‌شوند. وکا، الگوریتمی را برای پیش‌پردازش داده‌ها، طبقه‌بندی، رگرسیون، خوشه‌بندی و قوانین انجمنی، پیاده‌سازی می‌کند. همچنین شامل ابزارهای تجسم داده نیز می‌باشد. طرح‌های جدید یادگیری ماشین نیز می‌تواند با استفاده از وکا توسعه یابند. وکا یک نرم‌افزار منبع باز (رایگان) است. وکا، به‌طور معمول از فایل‌های داده با فرمت ARFF استفاده می‌کند که این فایل‌ها شامل برچسب‌هایی ویژه برای نشان دادن اقسام مختلف موجود در فایل داده می‌باشند.

TANAGRA: Tanagra یک نرم‌افزار رایگان داده‌کاوی است که برای اهداف علمی و پژوهشی از آن استفاده می‌شود. در این نرم‌افزار، چندین روش داده‌کاوی از تحلیل داده‌های اکتشافی، یادگیری آماری، یادگیری ماشین و پایگاه داده، ارائه می‌شود. Tanagra یک پروژه منبع باز است زیرا هر پژوهشگری می‌تواند به کد منبع آن دسترسی پیدا کند و الگوریتم‌های خود را به آن اضافه کند (البته تا آنجاکه موافق و مطابق با مجوز^۴ توزیع نرم‌افزار است). هدف اصلی پروژه Tanagra این است که پژوهشگران و دانشجویان، به راحتی و مطابق با هنجارهای موجود در حوزه توسعه نرم‌افزار، از نرم‌افزار داده‌کاوی استفاده کنند و امکان تحلیل اطلاعات واقعی یا مصنوعی را برای آنها فراهم می‌آورد.

متلب: متلب، یک زبان و محیطی تعاملی برای محاسبات عددی، تجسم^۵ و برنامه‌نویسی است. با استفاده از متلب، ما می‌توانیم داده‌ها را تحلیل کنیم، الگوریتم‌ها را توسعه دهیم و مدل‌ها و برنامه‌های کاربردی را

³WEKA

⁴license

⁵visualization

جدول ۳: ابزارهای مختلف داده کاوی، برای پیش بینی دقیق بیماری های قلبی

نویسنده	تکنیک استفاده شده	ابزار داده کاوی	دقت	هدف
Abhi shek و همکارانش (۲۰۱۳)	J48	وکا 3.6.4	۹۵/۵۶٪	سیستم پیش بینی بیماری قلبی، که از تکنیک های داده-کاوری استفاده می کند
	بیزین ساده		٪۹۲/۴۲	
	J48		٪۹۴/۸۵	
Chai و همکارانش (۲۰۱۲)	شبکه عصبی	وکا 3.6.6	۱۰۰٪	پیش بینی بیماری قلبی
Mon ali Et و همکارانش	C4.5	وکا		مطالعه و تحلیل الگوریتم های داده کاوی برای پشتیبانی از سیستم مراقبت بهداشتی الکترونیکی و سیستم پشتیبانی از تصمیم
	پرسپترون چند لایه			
	بیزین ساده			
Nidh و همکاران (۲۰۱۲)	بیزین ساده	وکا 3.6.6	٪۹۰/۷۴ ٪۹۹/۶۲ ٪۱۰۰	تحلیل سیستم پیش بینی بیماری های قلبی، با استفاده از تکنیک های مختلف داده کاوی
	درخت تصمیم	TANAGRA	٪۵۲/۳۳ ٪۵۲ ٪۴۵/۶۷	
		وکا 3.6.0	٪۸۶/۵۳ ٪۸۹ ٪۸۵/۵۳	
	شبکه های عصبی	پلت فرم دات نت	٪۹۶/۵ ٪۹۹/۲ ٪۸۸/۳	
Resul و همکاران (۲۰۰۹)	شبکه های عصبی	نرم افزار الکترونیکی 9.1.3 مبتنی بر SAS	٪۹۷/۴	تشخیص سلول های بیماری قلبی
Rash eDur و همکاران (۲۰۱۳)	شبکه عصبی	وکا	٪۷۹/۱۹	مقایسه ی میان تکنیک های مختلف طبقه بندی
	منطق فازی	TANAGRA	٪۸۳/۸۵	
	درخت تصمیم	متلب	-	
Resu و همکاران (۲۰۰۹)	شبکه های عصبی	نرم افزار الکترونیکی 9.1.3 مبتنی بر SAS	٪۸۹/۰۱	تشخیص بیماری قلبی

ایجاد کنیم. این زبان، با ابزار و توابع ریاضی آماده‌ی خود، ما را قادر به کشف رویکردهای متعدد و دستیابی به یک راه‌حل سریع‌تر (در مقایسه با صفحات گسترده و زبان‌های برنامه‌نویسی سنتی، مانند جاوا، C++ یا C) می‌سازد.

Orange: به‌صورت رایگان به تجسم و تحلیل داده‌ها برای افراد تازه‌کار و کارشناسان، می‌پردازد. این ابزار داده-کاوی، از برنامه‌نویسی و ویژوال اسکریپت پایتون و مولفه‌های یادگیری ماشین استفاده می‌کند. افزونه‌های این ابزار برای بیوانفورماتیک و استخراج متن، مورد استفاده قرار می‌گیرند. همچنین این ابزار، دارای ویژگی‌هایی برای تحلیل داده می‌باشد.

چارچوب داتانت: چارچوب داتانت، چارچوبی نرم‌افزاری است که توسط شرکت مایکروسافت طراحی شده است عمدتاً بر روی سیستم‌عامل ویندوز اجرا می‌شود و قابلیت برنامه‌نویسی، با چند زبان برنامه‌نویسی را فراهم می‌آورد. چارچوب داتانت، برای توسعه‌دهندگان نرم‌افزار، چارچوبی جامع و سازگار فراهم می‌آورد که تجربیات بصری خیره‌کننده و همچنین امکان برقراری ارتباطاتی بی‌نظیر و امن را برای کاربر فراهم می‌آورد.

رپیدماینر: بدون شک، رپیدماینر^۶، دنیای پیش‌روی سیستم منبع‌باز برای داده‌کاوی است. این برنامه، به‌عنوان یک برنامه‌ی مستقل برای تحلیل داده‌ها و به‌عنوان یک موتور داده‌کاوی، برای ادغام در محصولات خود، در دسترس است. در بیش از ۴۰ کشور، هزاران برنامه کاربردی رپیدماینر وجود دارد؛ در نتیجه یک مزیت رقابتی برای کاربران خود فراهم می‌آورد.

روش‌شناسی مورد استفاده در داده‌کاوی

داده‌کاوی، هسته‌ی پایگاه‌داده‌ی کشف دانش (KDD^۷) است. بسیاری از افراد داده‌کاوی را به‌عنوان اصطلاحی که به پایگاه‌داده‌ی کشف دانش اشاره دارد، می‌شناسند؛ زیرا داده‌کاوی یک بخش کلیدی از روند کشف دانش است. کشف دانش به‌عنوان یک فرایند در شکل ۱ نشان داده شده است و شامل یک توالی تکراری از مراحل زیر می‌باشد:

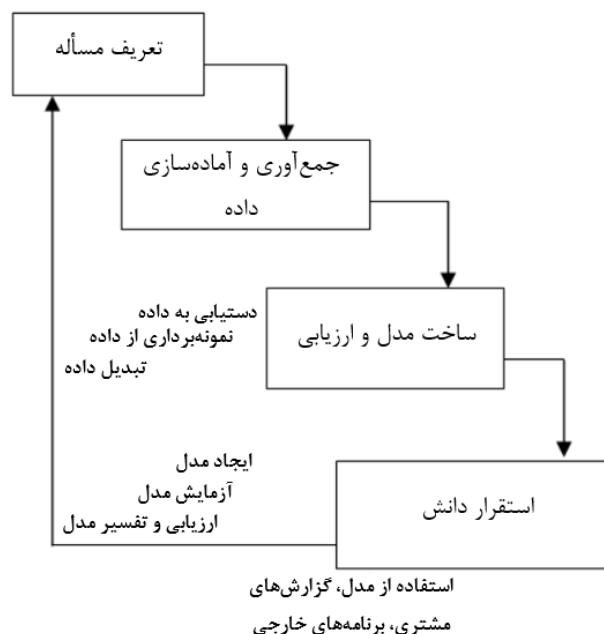
- پاک‌سازی داده: برای حذف نویز یا داده‌های نامناسب
- ادغام داده: فرایندی که طی آن داده‌های چندگانه ممکن است با یکدیگر ترکیب شوند.
- انتخاب داده‌ها: فرایندی که طی آن، داده‌های مرتبط با وظیفه تحلیل، از پایگاه‌داده بازیابی می‌شوند.
- تبدیل داده: فرایندی که طی آن، داده‌ها با انجام عملیات خلاصه‌سازی یا تجمیع، به شکل‌های مناسب برای کاوش از طریق عملیات خلاصه‌سازی یا تجمیع تبدیل می‌شوند.

^۶RapidMiner

^۷Knowledge Discovery Database

جدول ۴: این جدول، تکنیک‌های مختلف داده‌کاوی را نشان می‌دهد که بر روی مجموعه داده‌ی بیماری‌های قلبی اعمال شده‌اند.

نویسنده	سال	تکنیک	دقت
Chaitrali و همکاران	۲۰۱۲	بیزین ساده	٪۹۰/۷۴
		تبدیل داده	٪۹۹/۶۲
		شبکه عصبی	٪۱۰۰
Indira S. Fal Dessai	۲۰۱۳	شبکه عصبی احتمالاتی	٪۹۴/۶
		تبدیل داده	٪۸۴/۲
		بیزین ساده	٪۸۴
		شبکه عصبی پس انتشار خطا	٪۸۰/۴
Jesmin و همکاران	۲۰۱۳	بیزین ساده	٪۹۲/۰۸
		SMO	٪۹۶/۰۴
		IBK	٪۹۵/۰۵
		AdaBoostM1	٪۹۶/۰۴
		J48	٪۹۶/۰۴
		PART	٪۹۶/۰۴
M. anbarasi و همکاران	۱۹۹۹	بیزین ساده	٪۹۶/۵
		درخت تصمیم	٪۹۹/۲
		طبقه‌بندی از طریق خوشه‌بندی بیزین ساده	٪۸۸/۳
Matjaz و همکاران	۱۹۹۹	تمرین ECG(NN)	٪۷۴
		تمرین ECG(NN)	
		اسکن قلب (NN)	٪۸۵
N. Aditya Sundar و همکاران	۲۰۱۲	WAC	٪۸۴
		بیزین ساده	٪۷۸
T. John و همکاران	۲۰۱۲	بیزین ساده	٪۸۵/۱۸
		چندلایه	٪۷۸/۸۸
		J48	٪۸۵/۱۸
		KNN	٪۸۵/۵۵
Tanawut و همکاران	۲۰۰۸	شبکه عصبی پس انتشار خطا	٪۷۴/۵
		DK-SOM	٪۸۰/۴



شکل ۱: داده کاوی

- داده کاوی: فرایندی ضروری، که در آن برای استخراج الگوهای داده از روش های هوشمند استفاده می شود.
- ارزیابی الگو: برای شناسایی الگوهای واقعا جالبِ بازنمایی دانش، بر اساس برخی معیارهای جالب، از فرایند ارزیابی الگو استفاده می شود.
- ارائه دانش: به منظور ارائه دانش کاوش شده به کاربر، از تکنیک های بازنمایی دانش^۸ استفاده می شود.

مرحله ی داده کاوی ممکن است با کاربر یا پایگاه دانش تعامل داشته باشد. الگوهای جالب به کاربر ارائه می شوند و ممکن است به عنوان دانش جدید، در پایگاه دانش ذخیره شوند. داده کاوی فرآیند کشف دانش از حجم عظیمی از داده های ذخیره شده در پایگاه داده ها، انبارهای داده یا سایر مخازن اطلاعاتی می باشد.

۳ تکنیک های داده کاوی مورد استفاده برای تشخیص بیماری های دیگر

^۸knowledge representation

جدول ۵: این جدول فهرستی از مقاله‌های پژوهشی را نشان می‌دهد که در آنها از تکنیک‌های داده‌کاوی برای تشخیص بیماری‌های مختلف استفاده شده است.

فناوری	بیماری	سال	نویسنده
طبقه‌بندی، پس‌انتشار خطا، شبکه‌ی عصبی فازی	دیابت	۲۰۰۸	هومار و همکارانش
	بیماری قلبی		
طبقه‌بندی بیزین	بیماری قلبی کارسینوئید	۲۰۰۷	مارسل و همکارانش
C4.5, C5.0	سرطان پستان	۲۰۱۲	محمد و همکارانش
	بیماری قلبی		
طبقه‌بندی وابسته و الگوریتم ژنتیک	دیابت هندی پیم	۲۰۱۲	M.Akhil و همکارانش
	سرطان پستان		
	بیماری قلبی		

تکنیک‌های داده‌کاوی، در زمینه‌های مختلفی از جمله موارد اشاره شده در بالا، مورد استفاده قرار می‌گیرند؛ برخی از این زمینه‌ها عبارتند از: عملکرد دانشجویان، کسب و کار، تشخیص تقلب، بانکداری، بازاریابی، قانون، بیمه و غیره. در مقاله‌های ارائه شده، برای پیش‌بینی بیش از یک بیماری، از تکنیک‌های داده‌کاوی استفاده می‌شود؛ در مقاله‌ای که توسط محمد و همکارانش [۱] ارائه شده است، از الگوریتم‌های داده‌کاوی C4.5 و C5.0 برای پیش‌بینی بیماری‌های قلبی و همچنین برای پیش‌بینی سرطان پستان استفاده می‌شود. در مقاله هومار^۹ و همکارانش [۲۲] از تکنیک‌های طبقه‌بندی، انتشار خطا به عقب و شبکه‌ی عصبی فازی، برای پیش‌بینی بیماری دیابت و بیماری‌های قلبی استفاده می‌شود. در مقاله‌ای که توسط M. Akhil و همکارانش [۲۹] ارائه شده است، برای پیش‌بینی سه بیماری مختلف، یعنی دیابت هندی پیم^{۱۰}، سرطان پستان و بیماری‌های قلبی، از تکنیک طبقه‌بندی وابسته^{۱۱} و الگوریتم ژنتیک استفاده می‌شود.

۴ نتیجه‌گیری

هدف ما از انجام این پژوهش، مطالعه بر روی تکنیک‌های مختلف داده‌کاوی است، تکنیک‌هایی که در سیستم‌های پیش‌بینی خودکار بیماری‌های قلبی، قابل استفاده هستند. در این پژوهش، به تعریف آن دسته از تکنیک‌های مختلف و طبقه‌بندی‌کننده‌های داده‌کاوی پرداخته شد، که در سال‌های اخیر برای تشخیص بیماری‌های مختلف، عملکرد موثری داشته‌اند. تحلیل‌های انجام شده نشان می‌دهد که در تمام مقالات، فناوری‌های مختلف از تعداد مختلفی از خصوصیت‌ها استفاده می‌کنند. بنابراین، نشان داده شد که تکنولوژی‌های مختلف، دقت‌های مختلفی دارند. در برخی از مقالات نشان داده شده است که شبکه‌های عصبی در پیش‌بینی بیماری‌های قلبی به دقت ۱۰۰٪ دست یافته‌اند. از سوی دیگر، باید در نظر گرفته

⁹Humar

¹⁰Pima Indian Diabetes

¹¹Associative Classification

شود که درخت تصمیم نیز با استفاده از ۱۵ ویژگی، به دقت ۹۹/۶۲٪ دست یافته و عملکرد خوبی داشته است [۶]. بنابراین، با استفاده از تکنولوژی‌های مختلف، نشان داده شده است که میزان دقت، به تعداد ویژگی‌های در نظر گرفته شده و ابزار مورد استفاده برای پیاده‌سازی، بستگی دارد. محققان بر این باورند که با توجه به افزایش سالانه‌ی مرگ و میر ناشی از بیماری‌های قلبی در سراسر جهان، و در دسترس بودن حجم عظیمی از اطلاعات، می‌توان از روش‌های داده‌کاوی برای تشخیص بیماری‌های قلبی استفاده کرد. با وجود اینکه تکنیک‌های داده‌کاوی در کمک به متخصصان مراقبت‌های بهداشتی برای تشخیص بیماری‌های قلبی موفق بوده‌اند، اما استفاده از تکنیک‌های داده‌کاوی برای شناسایی روش درمان مناسب برای بیماران مبتلا به بیماری‌های قلبی کمتر مورد توجه قرار گرفته است.

مراجع

- [1] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining, International Journal of Applied Engineering Research, 2012.
- [2] Ma.jabbar, Dr.priiti Chandra, B.L.Deekshatulu, cluster based association rule mining for heart attack prediction, Journal of Theoretical and Applied Information Technology, 2011.
- [3] Ms. Ishtake S.H ,Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J. of Healthcare & Biomedical Research, 2013.
- [4] Dr. K. Usha Rani, analysis of heart diseases dataset using neural network approach, International Journal of Data Mining & Knowledge Management Process, 2011.
- [5] Carlos Ordonez, Edward Omiecinski, Mining Constrained Association Rules to Predict Heart Disease, IEEE. Published in International Conference on Data Mining (ICDM), p. 433- 440, 2001.
- [6] Nidhi Bhatla Kiran Jyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT), 2012.
- [7] Shantakumar B.Patil, Dr.Y.S. Kumaraswamy, Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction, (IJCSNS) International Journal of Computer Science and Network 228 Security, 2009.
- [8] Abhishek taneja, Heart Disease Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co., India, 2013.
- [9] M. Anbarasi, E. Anupriya, N.ch.s.n.Iyengar, Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, International Journal of Engineering Science and Technology, 2010.
- [10] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology, 2012.

- [11] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra, performance analysis of classification data mining techniques over heart diseases data base, international journal of engineering science and advanced technology, 2012.
- [12] Shadab Adam Pattekari and Asma Parveen, prediction system for heart disease using naïve bayes, International Journal of Advanced Computer and Mathematical Sciences, 2012.
- [13] Latha Parthiban and R.Subramanian, Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm, International Journal of Biological and Medical Sciences, 2008.
- [14] Jesmin Nahar, Tasadduq Imama, Kevin S. Tickle, Yi-Ping Phoebe Chen, Association rule mining to detect factors which contribute to heart disease in males and females, Elsevier, 2013.
- [15] 15. Nada Lavrac, Selected techniques for data mining in medicine, Elsevier, 1999.
- [16] Tanawut Tantimongcolwat, Thanakorn Naenna, Identification of ischemic heart disease via machine learning analysis on Magnetocardiograms, Elsevier, 2008.
- [17] Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur, Effective diagnosis of heart disease through neural networks ensembles, Elsevier, 2009.
- [18] Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur Diagnosis of valvular heart disease through neural networks ensembles, Elsevier, 2009.
- [19] Oleg Yu. Atkov, Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters, Elsevier, 2012.
- [20] Marcel A.J. van Gerven, Predicting carcinoid heart disease with the noisy-threshold classifier, Elsevier, 2007.
- [21] Matjaz' Kukar, Analysing and improving the diagnosis of ischaemic heart disease with machine learning, Elsevier, 1999.
- [22] Humar Kahramanli, Novruz Allahverdi, Design of a hybrid system for the diabetes and heart diseases, Elsevier, 2008.
- [23] Jesmin Nahar, Tasadduq Imam, Computational intelligence for heart disease diagnosis: A medical knowledge driven approach, Elsevier, 2013.
- [24] Nan-Chen Hsieh, Lun-Ping Hung, Chun-Che Shih, Intelligent Postoperative Morbidity Prediction of Heart Disease Using Artificial Intelligence Techniques, J Med Syst, 2012.
- [25] Adebayo Peter Idowu, Data Mining Techniques for Predicting Immunize-able Diseases: Nigeria as a Case Study, International Journal of Applied Information Systems, 2013.
- [26] Monali Dey, Siddharth Swarup Rautaray, Study and Analysis of Data mining Algorithms for Healthcare Decision Support System, International Journal of Computer Science and Information Technologies, 2014.
- [27] T. John Peter, K. Somasundaram, study and development of novel feature selection framework for Heart disease prediction, International Journal of Scientific and Research Publications, 2012.

- [28] Indira S. Fal Dessai, Intelligent Heart Disease Prediction System Using Probabilistic Neural Network, International Journal on Advanced Computer Theory and Engineering, 2013.
- [29] M.Akhil jabbar, Dr.Priti Chandra, Dr.B.L Deekshatulu, Heart Disease Prediction System using Associative Classification and Genetic Algorithm, International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies, 2012.
- [30] Chaitrali S. Dangare, Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques, International Journal of Computer Applications, 2012.
- [31] Rashedur M. Rahman, Farhana Afroz, Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis, Journal of Software Engineering and Applications, 2013.
- [32] Venkatadri.M, Dr. Lokanatha C. Reddy a review on data mining from past to the future. International Journal of Computer Applications, 2011.

