

ارتباط بین شاخص‌های پروگنوستیک سرطان سینه با استفاده از تکنیک‌های داده‌کاوی

www.ketab.ir

مؤلف

حسن سالار آبادی

انتشارات ناران

سرشناسه	: سالارآبادی، حسن، ۱۳۶۱-
عنوان و نام پدیدآور	: ارتباط بین شاخص‌های پروگنوستیک سرطان سینه با استفاده از تکنیک‌های داده‌کاوی / مؤلف حسن سالارآبادی؛ ویراستار منیرالسادات موسویان.
مشخصات نشر	: اصفهان: ناران، ۱۴۰۳.
مشخصات ظاهری	: ۱۰۳ص: مصور، جدول، نمودار.
شابک	: ۹۷۸-۶۲۲-۸۴۸۵-۰۴-۱
وضعیت فهرست نویسی	: فیبا
یادداشت	: کتابنامه: ص. ۱۰۱ - ۱۰۳.
موضوع	: پستان -- سرطان -- ایران -- گزارش‌های بالینی Breast-- Cancer-- Iran-- Case studies پستان -- سرطان -- ایران -- تشخیص -- گزارش‌های بالینی Breast -- Cancer -- Iran -- Diagnosis -- Case studies
رده بندی کنگره	: RC۲۸۰/پ۵
رده بندی دیویی	: ۶۱۶/۹۹۴۴۹۰۹۵۵
شماره کتابشناسی ملی	: ۹۶۵۵۰۲۱
اطلاعات رکورد کتابشناسی	: فیبا

ارتباط بین شاخص‌های پروگنوستیک سرطان سینه با استفاده از تکنیک‌های داده‌کاوی

مؤلف: حسن سالارآبادی

ویراستار: منیرالسادات موسویان

طراح: سیده لیلا کشمیری

انتشارات: ناران

شابک: ۹۷۸-۶۲۲-۸۴۸۵-۰۴-۱

نوبت چاپ: اول - ۱۴۰۳

شمارگان: ۱۰۰۰ جلد

قیمت: ۱,۰۰۰,۰۰۰ ریال

آدرس: اصفهان، خیابان آپادانا اول، کوچه باغ توت، کوچه باغ کاج، کوچه هاشم‌زاده

تماس: ۰۹۱۳۲۰۱۸۰۸۹، پیام‌رسان: ۰۹۳۶۹۴۵۹۱۲۴

*اینستاگرام: Nashrnaran

فهرست مطالب

پیش گفتار.....	۹
بخش اول: مباحث مقدماتی	
فصل اول: کلیات	۱۵
بیان مسئله	۱۵
علت انتخاب موضوع و اهمیت آن	۱۸
فصل دوم: مفهوم شناسی	۲۱
تعاریف مفاهیم و اصطلاحات	۲۲
بخش دوم: پیشینه پژوهش	
فصل اول: داده کاوی و تکنیک های آن	۳۵
قوانین انجمنی	۳۵
طبقه بندی	۳۸
خوشه بندی	۴۴
الگوریتم های ژنتیک	۴۵
معیارهای ارزیابی	۴۵
نرم افزارهای داده کاوی	۴۷
فصل دوم: پیشینه داده کاوی در پزشکی	۴۹
فصل سوم: کاربرد الگوریتم های داده کاوی در سرطان پستان	۵۳
طبقه بندی بیماران سرطان پستان با الگوریتم های داده کاوی	۵۴
تعیین نوع درمان بیماری با الگوریتم های داده کاوی	۵۴
پیش بینی عود مجدد بیماری با داده کاوی	۵۴
پیش بینی و تشخیص بیماری ها	۵۵
بررسی روابط معنادار فاکتورهای مؤثر سرطان پستان با تکنیک های داده کاوی	۵۶
بخش سوم: تحقیق و بررسی	
فصل اول: معرفی الگوریتم پیشنهادی	۶۱
دسته بندی بیماران سرطان پستان با استفاده از الگوریتم های داده کاوی	۶۱
مرحله بندی سرطان پستان	۶۲
مرحله بندی جدید سرطان پستان	۶۸
نتایج دست آمده از تکنیک های داده کاوی جهت مرحله بندی بیماری	۶۹
بررسی وضعیت غدد لنفاوی در سرطان پستان با تکنیک های داده کاوی	۷۱
خروجی تکنیک های داده کاوی	۷۱
بررسی میزان تأثیر گیرنده های استروژن و پروژسترون در بیماری سرطان پستان توسط ماشین	۷۳

۷۳	خروجی تکنیک‌های داده‌کاوی
۷۵	تشخیص میزان رابطه واسکولار با هرتونیو با الگوریتم‌های دسته‌بندی
۷۷	اثبات رابطه غدد لنفاوی با گیرنده هورمونی هرتونیو با تکنیک‌های داده‌کاوی
۸۱	فصل دوم: ارزیابی الگوریتم پیشنهادی
۸۱	محیط پیاده‌سازی و اجرای الگوریتم
۸۳	داده‌های مورد استفاده
۸۴	تنظیم پارامترها
۸۴	روش‌های مورد استفاده به منظور مقایسه
۸۵	معیارهای ارزیابی
۸۶	ارزیابی کارایی الگوریتم پیشنهادی
۸۶	ارزیابی کارایی الگوریتم‌های دسته‌بندی جهت مرحله‌بندی بیماری
۸۸	ارزیابی کارایی الگوریتم‌های دسته‌بندی درمورد وضعیت غدد لنفاوی
۸۹	ارزیابی کارایی الگوریتم‌های دسته‌بندی درمورد تأثیر گیرنده‌های استروژن و پروژسترون
۹۱	ارزیابی کارایی الگوریتم‌های دسته‌بندی درمورد رابطه واسکولار با هرتونیو
۹۲	ارزیابی کارایی الگوریتم‌های دسته‌بندی مربوط به رابطه غدد لنفاوی با گیرنده هورمونی هرتونیو
۹۵	فصل سوم: نتیجه‌گیری و پیشنهادها
۹۵	نتایج حاصل از تحقیق
۹۶	پیشنهادها
۹۹	کتابنامه

پیش گفتار

علم داده‌کاوی به کشف الگوهای پنهان و ناشناخته در میان حجم عظیمی از داده‌ها می‌پردازد که گاه از دید متخصصین پنهان می‌مانند. اصلی‌ترین دلیلی که باعث شده داده‌کاوی کانون توجهات در صنعت اطلاعات قرار بگیرد، مسئله در دسترس بودن حجم وسیعی از داده‌ها است و نیاز شدید به این‌که از این داده‌ها، اطلاعات و دانش سودمند استخراج کنیم.

پیشرفت‌های به‌وجود آمده در جمع‌آوری داده‌ها و قابلیت‌های ذخیره‌سازی در طی دهه‌های اخیر باعث شده‌است تا در بسیاری از علوم با حجم بزرگی از اطلاعات روبه‌رو شویم. محققان در زمینه‌های مختلف مانند مهندسی، اقتصاد، ستاره‌شناسی و زیست‌شناسی هر روز با مشاهدات بیشتر و بیشتری روبه‌رو می‌شوند. در مقایسه با بسترهای داده‌ای قدیمی و کوچک‌تر، بسترهای داده‌ای امروزی چالش‌های جدیدی در تحلیل داده‌ها به‌وجود آورده‌اند. روش‌های آماری سنتی به دو دلیل امروزه کارایی خود را از دست داده‌اند؛ علت اول، افزایش تعداد مشاهدات و علت دوم که از اهمیت بالاتری برخوردار است، افزایش تعداد متغیرهای مربوط به یک مشاهده می‌باشد. (اسماعیلی، ۱۳۹۱)

داده‌کاوی به فرآیند کاوش و تحلیل یک مجموعه داده بزرگ به‌منظور یافتن الگوهای منظم، استخراج دانش مرتبط و به‌دست آوردن قوانین تکراری گفته می‌شود. امروزه داده‌کاوی نقش مهمی در مطالعات نظری و علمی مختلف ایفا می‌کند. (ورسلیس، ۱۳۹۲)

یادگیری ماشین، زیرمجموعه‌ای از هوش مصنوعی است. با استفاده از تکنیک‌های یادگیری ماشین، کامپیوتر، الگوهای موجود در داده‌ها را یاد گرفته و می‌تواند از آن استفاده کند. الگوریتم‌های یادگیری ماشین به سه نوع اصلی تقسیم می‌شوند.

یادگیری نظارت شده^۱: در این نوع از الگوریتم‌ها که بار اصلی یادگیری ماشین را بر دوش می‌کشند، با دو نوع از متغیرها سروکار داریم. نوع اول که متغیرهای مستقل نامیده می‌شوند، یک یا چند متغیر هستند که قرار است براساس مقادیر آن‌ها، یک متغیر دیگر را پیش‌بینی کنیم. نوع دوم هم متغیرهای وابسته یا هدف یا خروجی هستند و قرار است مقادیر آن‌ها را به کمک این

الگوریتم‌ها پیش‌بینی کنیم.

یادگیری بدون ناظر^۱: در این نوع از الگوریتم‌ها، متغیر هدف نداریم و خروجی الگوریتم، نامشخص است. بهترین مثالی که برای این نوع از الگوریتم‌ها می‌توان زد، گروه‌بندی خودکار (خوشه‌بندی)^۲ یک جمعیت است.

یادگیری نیمه نظارت شده: نیز نوعی از یادگیری است که هم از داده‌های برچسب خورده و هم از داده‌های غیر و برچسب نخورده به‌صورت هم‌زمان استفاده می‌شود تا دقت یادگیری بهبود یابد.

درخت تصمیم^۳ یکی از پرکاربردترین الگوریتم‌های طبقه‌بندی داده‌ها می‌باشد. درخت تصمیم با استخراج تعدادی قانون از داده، قادر به توصیف الگوهای پنهان در داده می‌باشد. این روش هم‌چنین با ارائه درخت حاصل از مدل ساخته شده به روش گرافیکی، الگوها را توصیف می‌کند.

(Han, Cao, Molberg, Sarode, Rao, Sutton & Peng, 2011)

روش دیگری که برای دسته‌بندی داده‌ها براساس یادگیری از قوانین در این تحقیق استفاده شده‌است، الگوریتم **Jrip** می‌باشد. به‌عبارت دیگر می‌توان شروط و قوانین را به‌طور مستقیم از مجموعه داده‌های آموزشی استخراج کرد. (صنّعی، باده، محمودی و طاهرپور، ۱۳۹۱)

امروزه در حوزه پزشکی، جمع‌آوری داده‌ها درمورد بیماری‌های مختلف از اهمیت زیادی برخوردار است. حجم داده‌های جمع‌آوری شده بسیار بالاست و برای این که بتوان از بین این حجم انبوه داده‌ها، الگوها و نتایج موردنظر را به‌دست آورد، باید از تکنیک‌های داده‌کاوی استفاده کرد و نیز با به‌کارگیری الگوریتم‌های دسته‌بندی^۴ می‌توان تشخیص داد که یک شخص موردنظر با شرایط خاص ممکن است چه بیماری داشته باشد. البته باید از روش‌هایی استفاده شود که از درصد اطمینان بالایی برخوردار باشند. الگوریتم‌های مختلف در ارتباط با داده‌کاوی به‌طور قابل‌توجه‌ای به

1. Unsupervised Learning.

2. Clustering.

3. Decision Trees.

4. Classification.

درک صحیح داده‌های پزشکی کمک می‌کنند و همچنین به شناسایی روابط پیچیده و پنهان بین ویژگی‌های گروه‌های مختلف بیماران برای درمان بهتر و سریع‌تر آنان کمک می‌نمایند. (Sharma, Kumar & Gupta, 2011)

طبق آمار به‌دست‌آمده یکی از سرطان‌های شایع در زنان، سرطان پستان^۱ می‌باشد که تشخیص به‌موقع آن در ادامه حیات بیمار، نقش مهمی را ایفا می‌کند. در کشورهای توسعه یافته از هر هشت زن، یک زن درطول دوران حیات خود، سرطان سینه را تجربه می‌کند. این بیماری یکی از سرطان‌های شایع در دنیاست و آمار جهانی نشان می‌دهد که شیوع این بیماری درحال افزایش است. در دنیا سالیانه بیش از هشت میلیون نفر به انواع سرطان مبتلا می‌شوند که درحدود یک میلیون آن‌ها مربوط به سرطان پستان است. (Parkin, 2005)

هدف از اعمال الگوریتم‌های داده‌کاوی بر روی فاکتورهای کلینیکی مهم در بقاء بیماران سرطان پستان این است که تشخیص بهتری درمورد وضعیت پیش‌آگاهی بیماران به‌دست آوریم. با استفاده از الگوریتم‌های دسته‌بندی که یکی از مؤثرترین تکنیک‌های مختلف داده‌کاوی است، می‌توان بیماران مبتلا به سرطان پستان را بادقت بالا و درصد خطای کم طبقه‌بندی کرد. (Lakshmi, Veera Krishna & Prem Kumar, 2013)

می‌توان از تکنیک رگرسیون چند متغیره و تکنیک‌های داده‌کاوی جهت نشان دادن وابستگی بین ویژگی‌های تومور و فاکتور هرتونیو مثبت، استفاده کرد. (Pau, Monges, Doucet, Terris, Chenard, Bibeau, Penault-Liorce, Martin, Rabut & Leroux, 2016)

در این کتاب در ابتدا به تعریف بیان مسئله و ساختارهای کلی مورد پژوهش می‌پردازیم و سپس نکات اصلی این تحقیق و پژوهش مطرح می‌شود و در گام بعدی اهمیت ضرورت تحقیق مطرح می‌شود و نیز کاربردهای متصور موردبررسی قرار می‌گیرد و مختصری درمورد پیشینه آن صحبت می‌شود.