دانشگاه صنعتی امیرکبیر
(پلی‌تکنیک تهران)
دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان‌نامه کارشناسی
گرایش مهندسی کامپیوتر

عنوان
پیش‌بینی و تشخیص بیماری‌های قلبی با استفاده از داده‌کاوی

نگارش
ارمغان سرور

استاد راهنما
دکتر رضا صفابخش

خردادماه ۱۳۹۸


بسم الله الرحمن الرحيم

دانشگاه صنعتی امیرکبیر
(پلی‌تکنیک تهران)
دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان‌نامه کارشناسی
گرایش مهندسی کامپیوتر

عنوان
پیش‌بینی و تشخیص بیماری‌های قلبی با استفاده از داده‌کاوی

نگارش
ارمغان سرور

استاد راهنما
دکتر رضا صفابخش

خردادماه ۱۳۹۸

سپاس‌گزاری

هرگاه که نگاه نویی نسبت به زندگی و عظمت آن می‌یابیم، میل به تلاش و امید در وجودمان زنده می‌شود.

وظیفه خود می‌دانم که مراتب امتنان خود را نسبت به استاد راهنمایم **جناب آقای دکتر رضا صفابخش** که طی تدوین و پیش‌برد این پایان‌نامه مرا یاری نموده‌اند، اعلام بدارم.

ارمغان سرور
خرداد ۹۸

چکیده

امروزه کاربرد موفقیت‌آمیز داده‌کاوی درزمینه های بسیاری مانند تجارت الکترونیک، بازاریابی و صنایع دیگر قابل رؤیت است. در میان این بخش‌ها مراقبت‌های بهداشتی هنوز در مرحله کشف هستند. صنعت بهداشت و درمان به‌طورکلی شامل اطلاعات غنی است اما هنوز تمام اطلاعات برای کشف الگوهای پنهان و تصمیم‌گیری مؤثر استخراج نشده‌اند.

بیماری قلبی یکی از دلایل اصلی مرگ‌ومیر در سراسر جهان بوده و پیش‌بینی آن در مراحل اولیه ضروری است. سیستم‌های کامپیوتری می‌توانند به‌عنوان ابزار پیش‌بینی و تشخیص بیماری قلبی به پزشکان کمک کنند و با گسترش تحقیقات در این زمینه، آشنا نمودن خوانندگان با خلاصه‌ای از روش‌های فعلی پیش‌بینی و زیرمجموعه‌های آن ها اهمیت بالایی پیدا کرده است.

هدف این بررسی، شناسایی سیستم‌های پشتیبانی تصمیم برای پیش‌بینی و تشخیص این نوع بیماری‌ها بوده که از طریق داده‌کاوی و روش‌های هوشمند ترکیبی قابل بهبود است.

واژه‌های کلیدی:

داده‌کاوی، بیماری قلبی، سیستم پشتیبانی تصمیم

فهرست مطالب

فهرست اشکال

فهرست جداول و نمودارها

فصل اول

مقدمه

مقدمه

قلب مهم‌ترین عضو عضلانی بدن انسان بوده که خون را به تمامی قسمت‌های آن پمپاژ می‌کند. زندگی انسانی وابسته به عملکرد مناسب قلب است و عملکرد نامنظم آن بر سایر بخش‌های بدن انسان نظیر مغز، کلیه و... تأثیر می‌گذارد. به‌طورکلی ایست جریان خون در قلب، حمله قلبی و ایست آن در مغز، سکته مغزی به‌حساب می‌آید. اگر گردش خون در بدن ناکارآمد باشد، هم بر عملکرد قلب و هم مغز تأثیر خواهد داشت. طبق گزارش آمار بهداشت جهانی علت اصلی ابتلا به بیماری و مرگ‌ومیر در جامعه مدرن بیماری قلبی است. برای مثال در کشور خودمان این رقم نزدیک به ۳۰ درصد بوده که نگرانی پزشکان و مهندسین حوزه بیوتکنولوژی را برمی‌انگیزد.

تشخیص پزشکی امری بسیار مهم اما پیچیده است که نیازمند تجربه، اطلاعات بالای پزشک و همچنین هزینه بالا از سوی بیمار است. اگرچه پیشرفت قابل‌توجهی در تشخیص و درمان بیماری‌های قلبی حاصل شده است اما تحقیقات باید به بالاترین صحت خود برسد.

دسترسی به تعداد زیادی از داده‌های پزشکی و گسترش داده‌های ذخیره‌شده منجر به نیاز به ابزارهای قدرتمند برای تجزیه و تحلیل آنها جهت استخراج دانش مفید است. داده‌کاوی از ابزارهای تحلیل مؤثر برای کشف روابط پنهان و الگوهای میان داده‌ها محسوب می‌شود. با بکارگیری الگوریتم‌های داده‌کاوی در حوزه تشخیص و درمان بیمارهای قلبی، می‌توان سیستم‌های هوشمندی ابداع نمود که به شکل خودکار و بدون نیاز به نظارت پزشک قادر به فهم و تفسیر بیمارهای قلبی افراد باشند و یا اطلاعات مفیدی را اکتشاف نمایند که متخصصان را در قضاوت صحیح یاری رساند.

در این گزارش تلاش گردیده تا کاربرد داده‌کاوی در این حوزه نشان داده شود و الگوریتم‌های سرآمد، معرفی و بررسی گردند.

فصل دوم

تعریف و کاربرد داده‌کاوی

تعریف و کاربرد داده‌کاوی

خوب است پیش از ورود به بحث اصلی، با داده‌کاوی بیشتر آشنا شویم. این شناخت به ما کمک می‌کند تا الگوریتم‌های داده‌کاوی را بهتر بشناسیم و برای بهبود آنها راه‌های بهتری پیشنهاد دهیم.

در ادامه بیشتر درباره‌ی داده‌کاوی خواهید خواند.

۲-۱ تعریف داده‌کاوی

به مجموعه‌ای از روش‌های قابل اعمال بر پایگاه داده‌های بزرگ و پیچیده به‌منظور کشف الگوهای پنهان و جالب توجه نهفته در میان داده‌ها، داده‌کاوی گفته می‌شود. روش‌های داده‌کاوی تقریباً همیشه به لحاظ محاسباتی پرهزینه هستند. علم میان‌رشته‌ای داده‌کاوی، پیرامون ابزارها، متدولوژی‌ها و تئوری‌هایی است که برای آشکارسازی الگوهای موجود در داده‌ها مورداستفاده قرار می‌گیرند و گامی اساسی در راستای کشف دانش محسوب می‌شود. دلایل گوناگونی پیرامون چرایی مبدل شدن داده‌کاوی به چنین حوزه مهمی از مطالعات وجود دارد. برخی از این موارد در ادامه بیان‌شده‌اند[۱].

- رشد انفجاری داده‌ها در گستره وسیعی از زمینه‌ها
- افزایش سریع قدرت پردازش کامپیوتری

روش‌های داده‌کاوی دارای انواع گوناگونی هستند و از کلاس‌بندی گرفته تا روش‌های تشخیص با الگوی پیچیده و هزینه محاسباتی بالا که ریشه در علوم کامپیوتر دارند را شامل می‌شوند. هدف اصلی این روش‌ها انجام پیش‌بینی می‌باشد اما این تنها هدف آن‌ها نیست.

۲-۲ تاریخچه

در سال ۱۹۶۰، کارشناسان آمار از اصطلاحات «صید داده[1]» و «لایروبی داده[2]» برای ارجاع به فعالیت‌های «تحلیل داده[3]» استفاده می‌کردند. اصطلاح «داده‌کاوی» در حدود سال ۱۹۹۰ در جامعه پایگاه داده مورداستفاده قرار گرفت و به محبوبیت قابل‌توجهی دست پیدا کرد. عنوان مناسب‌تر برای فرآیند داده‌کاوی، «کشف دانش از داده» [4] است.

در حال حاضر، یادگیری آماری و «علم داده[4]» از دیگر عباراتی هستند که با معنای مشابه مورد استفاده قرار می‌گیرند، حال‌آنکه گاه تفاوت‌های ظریفی میان این موارد وجود دارد.

تولید داده موضوعی، به موازات پیشرفت‌های مادی و معنوی ابعاد گسترده‌ای یافته است و روز به روز نیز بر دامنه آن افزوده می‌شود. امروزه وسعت کاربرد آن کاملا مشهود است و قرن ۲۱ را قرن تولید داده و فناوری اطلاعات می‌نامند.

[1] Data Fishing
[2] Data Dredging
[3] Data Analytics
[4] Data Science

۲-۳ اهمیت

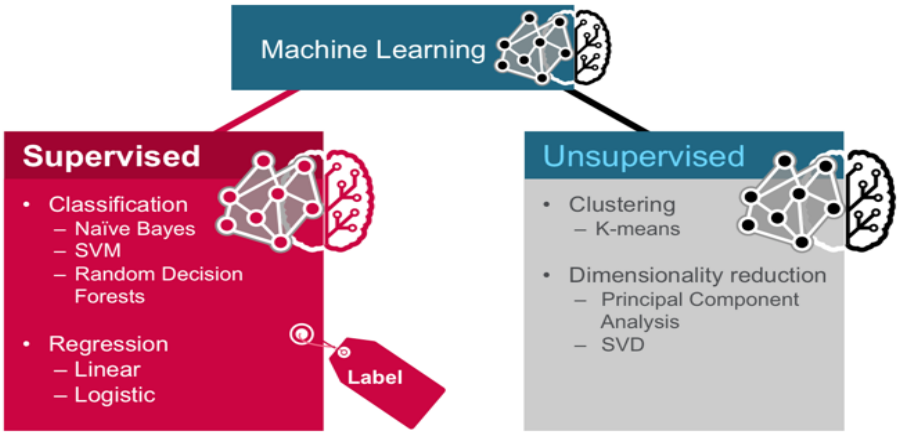
با رشد و افزایش توجه به داده‌کاوی، پرسش «چرا داده‌کاوی؟» مطرح می‌شود. در پاسخ به این پرسش باید گفت، داده‌کاوی کاربردهای زیادی دارد. بدین ترتیب، زمینه‌ای نوظهور و آینده‌دار برای نسل کنونی محسوب می‌شود. این زمینه توانسته توجهات زیادی را به صنایع و جوامع اطلاعاتی جلب کند که با وجود گستره وسیع داده‌ها، نیاز حتمی به تبدیل چنین داده‌هایی به اطلاعات و دانش وجود دارد.

بنابراین، بشر از اطلاعات برای دامنه‌ی وسیعی از کاربردها، از تحلیل بازار گرفته تا تشخیص بیماری‌ها، کشف کلاه‌برداری و پیش‌بینی قیمت سهام استفاده می‌کند. درمجموع، باید گفت ضرب‌المثل انگلیسی «نیاز، مادر همه ابداعات بشر است»، پاسخی کوتاه و گویا به پرسش مطرح شده است.

۲-۴ روش‌های داده‌کاوی

در داده‌کاوی عمدتاً دو روش اصلی یادگیری وجود دارد. این دو رویکرد یادگیری نظارت‌شده[۱] و یادگیری بدون نظارت[۲] نام دارند که در ادامه به توضیح آن‌ها خواهیم پرداخت[۱].

- یادگیری نظارت‌شده: یادگیری تحت نظارت رایج‌ترین روش برای یادگیری است.در این روش از مدل‌های از پیش تعریف‌شده، با هدف آموزش استفاده می‌شود. مدل موردنظر با استفاده از داده‌های آموزش‌داده‌شده ساخته می‌شود. داده‌های ورودی جدید با مدل آموزش‌دیده مقایسه می‌شوند و برچسب کلاس داده‌های جدید تعیین خواهد شد. روش‌های طبقه‌بندی و رگرسیون زیرمجموعه‌ی این دسته هستند.
- یادگیری بدون نظارت: برخلاف روش‌های داده‌کاوی تحت نظارت، در این روش نتیجه‌ای از محیط اطراف دریافت نمی‌شود. اگرچه تجسم چگونگی آموزش یک دستگاه بدون پاسخ از محیط اطرافش دشوار است اما این روش‌ها به‌خوبی کار می‌کنند. به احتمال زیاد، یک مدل مناسب برای روش‌های یادگیری بدون نظارت ایجاد می‌شود که هدف آن استفاده از ویژگی‌های ورودی فعلی برای پیش‌بینی ورودی‌های آینده، تصمیم‌گیری و انتقال مؤثر به یک الگوی دیگر است. خوشه‌بندی[۳] و روش کاهش ابعاد[۴] در این دسته قرار می‌گیرند.



شکل شماره ۲-۱ طبقه‌بندی الگوریتم‌های یادگیری

[1] Supervised learning
[2] Unsupervised learning
[3] Clustering
[4] Dimensionality reduction

۲ـ۵ کاربردها

داده‌کاوی در بسیاری از صنایع برای بهبود تجربه و رضایت مشتری و افزایش ایمنی و قابلیت استفاده محصول استفاده شده است. در مراقبت‌های بهداشتی نیز، استفاده از داده‌کاوی درزمینه‌هایی مانند پیش‌بینی بیماری، تشخیص تقلب و سوءاستفاده، مدیریت مراقبت‌های بهداشتی و اندازه‌گیری اثربخشی درمان‌های خاص اثبات شده است.

در ادامه به توضیح مختصری از دو مورد از این کاربردها می‌پردازیم.

- اندازه‌گیری اثربخشی درمان: این کاربرد از داده‌کاوی شامل مقایسه علائم، علل و دوره‌های درمان برای پیدا کردن مؤثرترین روش درمان یک بیماری خاص است. به‌عنوان مثال، گروه‌های بیمارانی که با رژیم‌های دارویی مختلف درمان می‌شوند، می‌توانند برای تعیین بهترین برنامه‌های درمان و صرفه‌جویی در هزینه‌ها، مقایسه شوند.
- تشخیص تقلب و سوءاستفاده: این مورد شامل ایجاد الگوهای طبیعی و سپس شناسایی الگوهای غیرمعمول ادعاهای پزشکی توسط کلینیک‌ها، پزشکان و یا آزمایشگاه‌ها است. این روش همچنین می‌تواند برای شناسایی ارجاع‌های نامناسب، نسخه‌های تقلبی و مدارک پزشکی جعلی مورداستفاده قرار گیرد.

فصل سوم

درباره‌ی بیماری‌های قلبی

درباره‌ی بیماری‌های قلبی

بیماری‌های قلبی و عروقی به معضلات زندگی انسان مدرن تبدیل شده‌اند. درنتیجه بخش مهمی از دانش بشری به این موضوع اختصاص یافته است. در این تحقیق نیز تمرکز ما بر بیماری قلبی است. لذا پیش از شروع بحث اصلی، ابتدا بهتر است با ابعاد و گسترش آن آشنا شویم.

۳-۱ تعریف

واژه بیماری قلبی معمولاً برای اشاره به حمله قلبی اشاره می‌شود، درحالی‌که شامل سایر مشکلات احتمالی قلب ازجمله بیماری عروق کرونر، نارسایی قلبی و سکته قلبی می‌باشد. چند نمونه از انواع مختلف بیماری‌های قلبی همراه با شرح در جدول زیر آورده شده است.

جدول شماره ۳-۱: انواع بیماری‌های قلبی

بیماری قلبی-عروقی	توصیف
سندرم حاد کرونری[1]	تأمین خون به عضله قلب به‌سرعت از بین می‌رود
آنژین[2]	درد قفسه سینه به علت کمبود خون ارسالی به عضله قلب
بیماری قلبی عروقی[3]	شریان‌هایی که خون را به عضله قلب می‌رسانند، بسته می‌شوند
بیماری‌های التهاب قلب	رطوبت عضله یا بافت قلب

مأخذ: سایت وزارت بهداشت و درمان، ۱۳۹۸

[1] Acute coronary syndrome

[2] Angina

[3] Coronary heart disease

۳-۲ وسعت بیماری قلبی

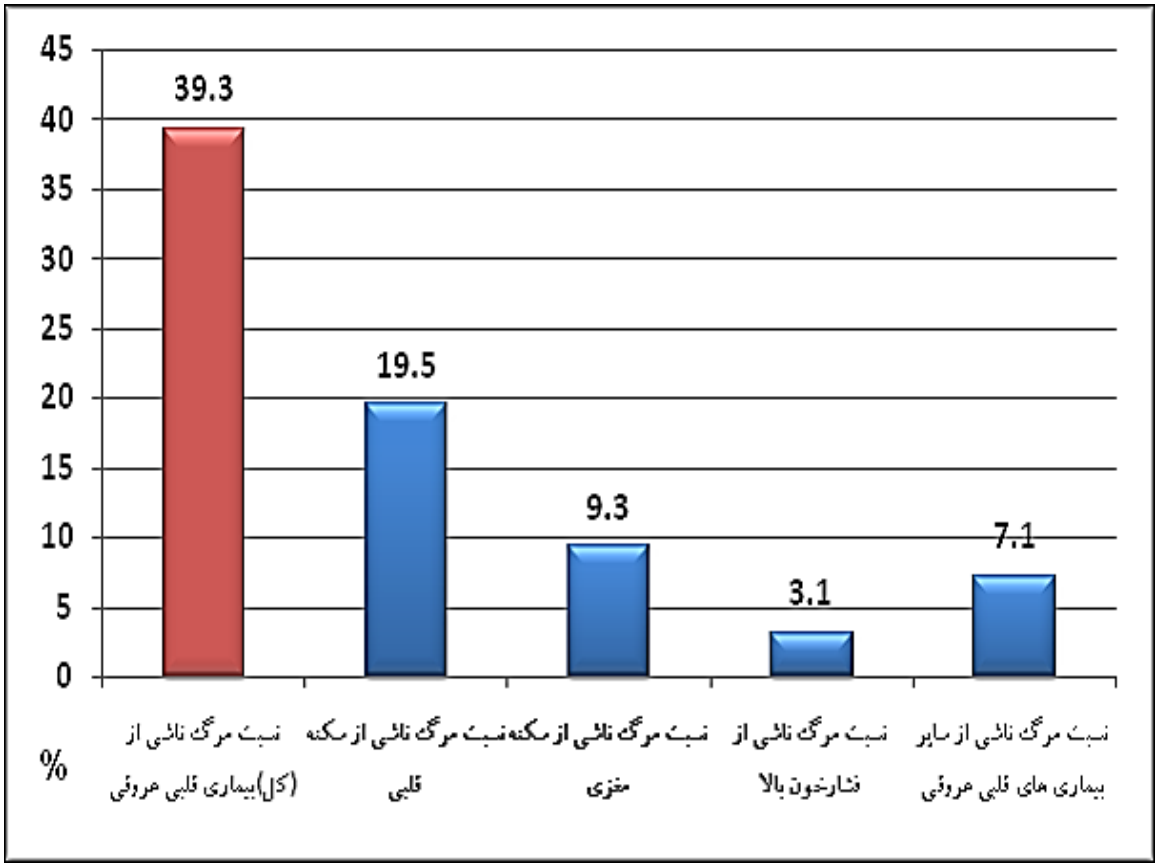
طبق برآورد صورت گرفته، ۱۷,۵ میلیون نفر در جهان در سال ۲۰۱۲ به علت بیماری‌های قلبی عروقی فوت نموده‌اند که ۳۱ درصد از کل موارد مرگ‌ومیرها را شامل می‌شود، از این مرگ‌ها حدود ۷,٤ میلیون به علت بیماری کرونر قلب و ٦,۷ میلیون ناشی از سکته‌ها بوده‌است. بیش از سه‌چهارم مرگ‌ومیر قلبی و عروقی در کشورهای کم‌درآمد و با درآمد متوسط رخ می‌دهد. از ١٦ میلیون مرگ‌ومیر زیر ۷۰ سال به علت بیماری‌های غیر واگیر، ۸۲ درصد آن در کشورهای کم‌درآمد و با درآمد متوسط ایجاد می‌شود و ۳۷ درصد آن به علت بیماری‌های قلبی و عروقی است[۲].

متأسفانه آمار مرگ‌ومیر ناشی از بیماری‌های قلبی در کشور ما بالاتر از کشورهای پیشرفته است، بنابراین پیشگیری از این بیماری ضروری به نظر می‌رسد.

طبق آخرین آمارهای ارائه‌شده، ۵۰ درصد جامعه شهری ایران افزایش وزن دارند، این در حالی است که تحقیقات نشان می‌دهد چاقی شدید به‌تنهایی می‌تواند خطر نارسایی قلبی و مرگ‌ومیر ناشی از بیماری‌های قلبی را افزایش دهد.

بنابر آمارهای رسمی، مساله بیماری‌های قلبی یک مساله جهانی است و عدم تشخیص به موقع نارسایی در تشخیص، پیامدهای انسانی و مالی زیادی را متوجه بودجه‌های عمومی کشورها و خانوارها می‌سازد.

نمودار شماره ۳-۱ نسبت مرگ‌ومیر ناشی از انواع بیماری‌های قلبی در ایران در سال ۱۳۹۵



مأخذ: سایت وزارت بهداشت و درمان

فصل چهارم

الگوریتم‌های مورداستفاده

الگوریتم‌های مورداستفاده

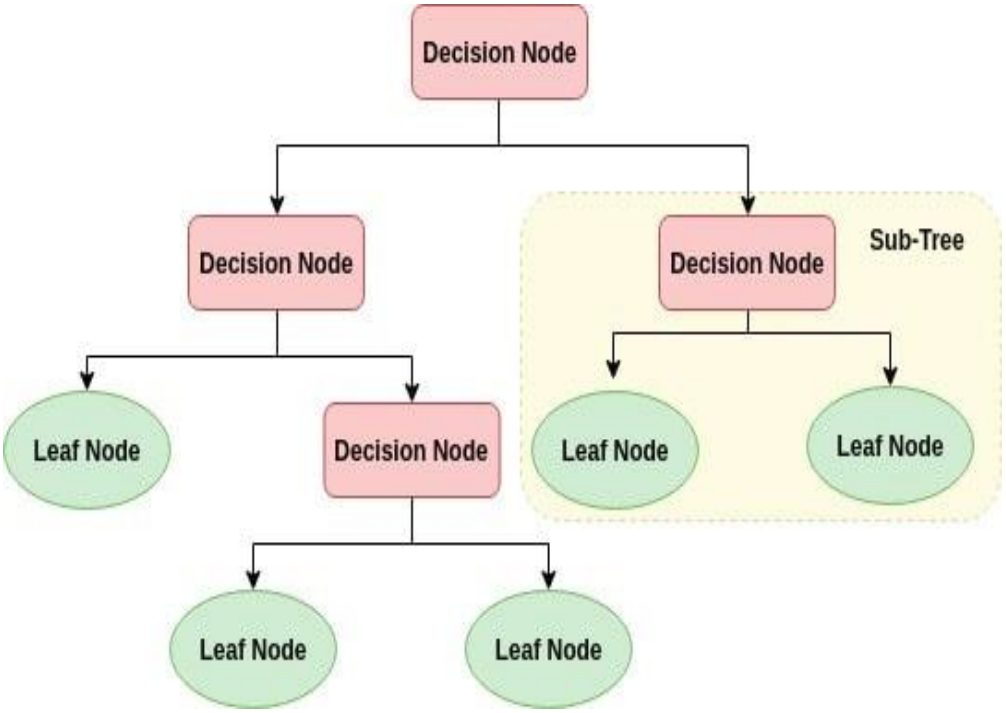
ابتدا نیاز داریم با روش‌های مختلف داده‌کاوی که تاکنون بر داده‌های بیماران اعمال گشته و آزمایش و بررسی‌شده‌اند آشنا شویم. اولین الگوریتم، درخت تصمیم[1]می‌باشد که در ادامه به معرفی آن می‌پردازیم.

٤-١ درخت تصمیم

درخت تصمیم یک طبقه‌بندی محبوب و ساده و آسان برای پیاده‌سازی محسوب می‌شود. در این روش هیچ نیازی به دانستن دامنه یا تنظیم پارامتر وجود ندارد و اجازه می‌دهد بتوانیم داده‌های با ابعاد بزرگ را مدیریت کنیم و نتایجی را که برای خواندن و تفسیر آسان هستند به دست آوریم. دسترسی به پروفایل دقیق بیماران تنها در درخت تصمیم امکان‌پذیر است.

یک درخت تصمیم‌گیری یک ساختار درخت مانند نمودار جریان است که هر گره داخلی نشان‌دهنده تست یک ویژگی و هر شاخه نشان‌دهنده نتیجه آزمایش است. همچنین گره‌های برگ نشان‌دهنده کلاس‌ها یا توزیع‌های کلاس هستند. گره بالاتر در یک درخت، گره ریشه است. گره‌های داخلی توسط مستطیل و گره‌های برگ توسط بیضی نشان داده می‌شوند. برای طبقه‌بندی یک نمونه ناشناخته، مقادیر ویژگی نمونه در درخت تصمیم‌گیری مورد آزمایش قرار می‌گیرند. مسیری از ریشه به گره برگ پیش‌بینی کلاس برای هر نمونه را نگه می‌دارد. بنابراین درخت‌های تصمیم می‌توانند به‌راحتی به قوانین طبقه‌بندی تبدیل شوند[٣].

[1] Decision Tree



شکل شماره ٤-١ شمای کلی یک درخت تصمیم

انواع پیاده‌سازی‌های درخت تصمیم در موارد زیر متفاوت می‌شوند[۴]:

۱. معیار تقسیم (چگونگی محاسبه واریانس[1])

۲. قابلیت ساخت مدل برای رگرسیون[2]

۳. نحوه مدیریت داده‌ها و اطلاعات ناقص

بر اساس این تفاوت‌ها الگوریتم‌های مختلفی شکل گرفتند که از میان آنها به موارد زیر اشاره می‌کنیم.

[1] Variance

[2] Regression

٤-١-١ روش ID3[1]

این روش از اولین پیاده‌سازی‌های درخت تصمیم به شمار می‌رود که طی زمان اجرا، درخت تصمیم‌گیری برای طبقه‌بندی نمونه‌های جدید از پیش مشاهده نشده با کمک مقادیر گره‌های درخت مورد استفاده قرار گرفته و به شما کلاس توزیعی نمونه را برمی‌گرداند.

این الگوریتم جواب بهینه را تضمین نمی‌کند و ممکن است در بهینه‌های محلی به اتمام برسد اما به‌طورکلی درخت‌های کوچک تصمیم تولید می‌کند که استفاده از آن‌ها باعث عدم روی‌هم افتادگی کامل[2] با داده‌ها خواهد شد[۵].

٤-١-٢ روش CART[3]

این روش، اگرچه عملکرد بخصوصی دارد اغلب به‌عنوان حالت کلی درخت تصمیم شناخته می‌شود. نمایش این مدل با استفاده از درخت دودویی[4] انجام می‌شود. در این درخت هر گره ریشه بیان‌گر یک متغیر ورودی است و گره‌های برگ برای نمایش متغیر خروجی استفاده می‌شوند. در این الگوریتم ما از روابط زیر استفاده می‌کنیم:

- کسب اطلاعات[5]

$$Information\ Gain = I(p,n) \qquad (٤\text{-}١)$$

$$I(p,n) = \frac{-p}{p+n}\log_2\left(\frac{p}{p+n}\right) - \left(\frac{n}{p+n}\right)\log_2\left(\frac{n}{p+n}\right)$$

اما P و n در این رابطه بیانگر چه هستند؟ برای یافتن آن‌ها باید ویژگی یا نتیجه کلاس را مشخص کنیم که آن نیز دودویی است. برای p مقدار واقعی یک و برای n مقدار کاذب صفر را در نظر می‌گیریم.

[1] Iterative Dichotomiser 3

[2] Overfitting

[3] Classification and Regression Tree

[4] Binary

[5] Information Gain

- بی‌نظمی[1]

$$Entropy(A) = \sum_{i=0}^{v} \frac{\mathrm{p_i}+n_i}{p+n} \big(I(p.n)\big) \qquad (٤\text{-}٢)$$

در حقیقت از بی‌نظمی برای ساخت درخت تصمیم استفاده می‌شود.

- بهره[2]

$$Gain = Information\ Gain - Entropy \qquad (٤\text{-}٣)$$
$$Gain = I(p,n) - E(A)$$

و درنهایت، از بهره برای یافتن یکی از ویژگی‌های مجموعه آموزشی استفاده می‌کنیم.

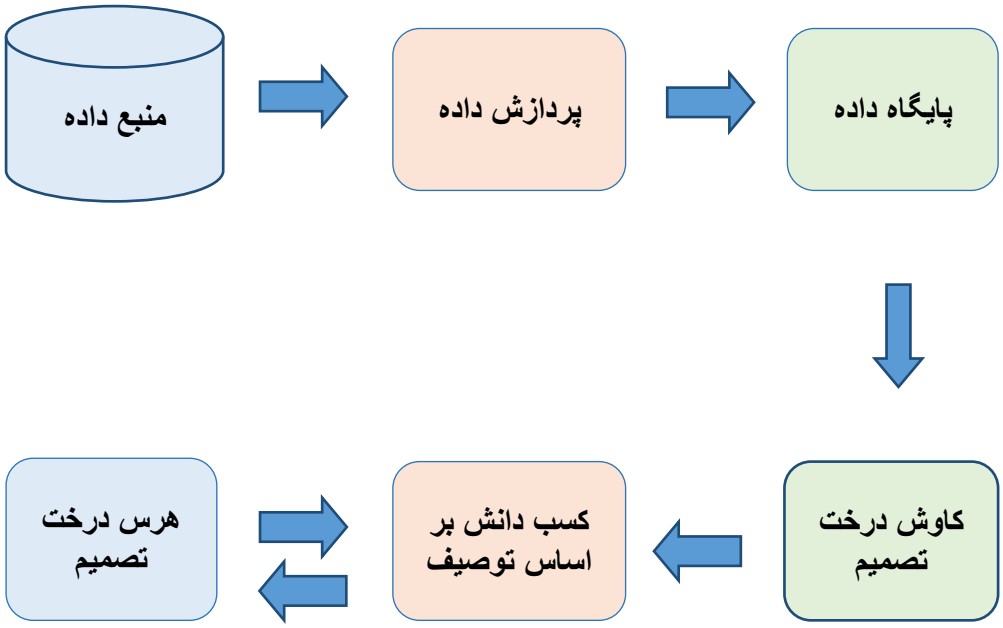
٤-١-٣ روش J48

تصور کنید که شما یک مجموعه داده، لیست پیش‌بینی کننده‌ها یا متغیرهای مستقل و یک لیست از اهداف یا متغیرهای وابسته دارید. یک درخت تصمیم‌گیری مانند J48، به شما امکان می‌دهد که متغیر هدف یک نمونه داده جدید را پیش‌بینی کنید.

درواقع درخت تصمیم‌گیری J48 اجرای الگوریتم ID3 می‌باشد که توسط تیم پروژه WEKA توسعه یافته است. اجرای بازگشتی این الگوریتم و پردازش داده در مراحل اولیه، موجب تسریع درروند عملکرد آن شده است. در شکل زیر مراحل دقیق اجرا قابل مشاهده است[۶].

[1] Entropy

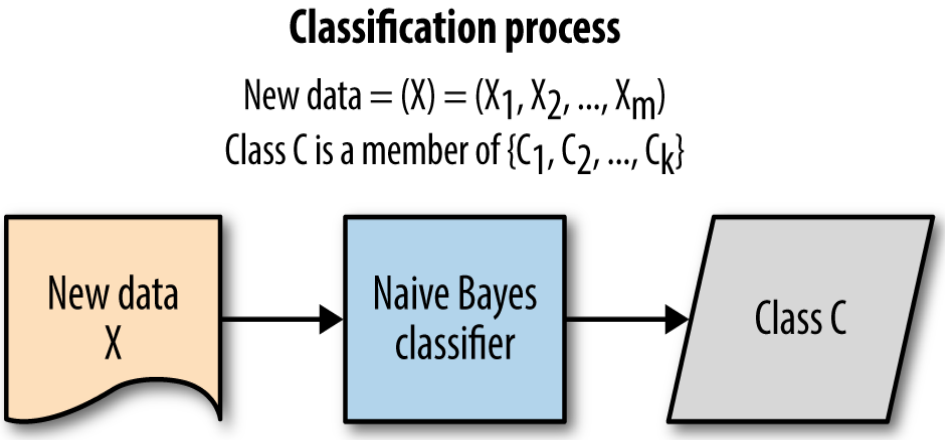
[2] Gain



شکل شماره ٤-٢ مراحل الگوریتم J48

٤-٢ الگوریتم Naïve Bayes

این الگوریتم یکی دیگر از سری الگوریتم‌های داده‌کاوی بوده که بر اساس اسم ابداع‌کننده‌اش نام‌گذاری شده است و الگوریتمی بانظارت محسوب می‌شود. این روش به طور کلی یک طبقه بندی آماری است که وابستگی بین صفات را دربرمی‌گیرد و در عین حال از استقلال مشروط استفاده می‌کند. لذا هر داده در آن کاملاً مستقل از سایر داده‌ها مورد بررسی قرار می‌گیرد. در شکل زیر حالت کلی اعمال این روش بر داده‌ها قابل مشاهده است[۷].

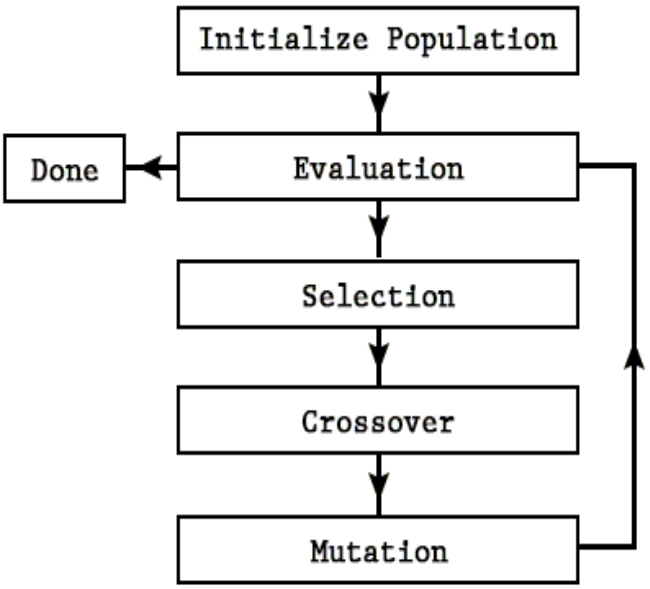


شکل شماره ۴-۳ اعمال الگوریتم Naïve Bayes به داده‌های جدید

٤-٣ اعمال الگوریتم ژنتیک

الگوریتم ژنتیک در هوش مصنوعی یک تکنیک جستجو است که از فرایند انتخاب طبیعی استفاده می‌کند. الگوریتم‌های تکاملی راه‌حل‌هایی برای بهینه‌سازی و جستجوی مشکلات با استفاده از روش‌هایی مانند جهش، انتخاب و بازترکیبی ارائه می‌کنند. الگوریتم ژنتیک معمولی نیاز به نمایش ژنتیکی دامنه راه‌حل و یک تابع تناسب[1] برای ارزیابی دامنه راه‌حل دارد. پس از انجام ارزیابی می‌توانیم اعضایی که شایستگی بالایی دارند انتخاب کرده و با اعمال ترکیب و جهش بر این اعضا تعادلی میان تکامل و تنوع که دو اصل مهم در الگوریتم‌های تکاملی محسوب می‌شوند برقرار کنیم[۸].

طبق آزمایش‌های اخیر، اگر الگوریتم‌های تکاملی مانند ژنتیک و یا الگوریتم‌های فازی را با سایر تکنیک‌های داده‌کاوی ترکیب کنیم، به جواب بهینه در مدت زمان کمتر و با سرعت بیشتری دست پیدا می‌کنیم.



شکل شماره ۴-۴ مراحل الگوریتم ژنتیک

[1] Fitness function

فصل پنجم

جمع‌بندی و نتیجه‌گیری و پیشنهادها

جمع‌بندی و نتیجه‌گیری و پیشنهادها

۵-۱ جمع‌بندی

امروزه تحقیقات زیادی درزمینه‌ی داده‌کاوی در جریان است. ما نیز سعی کردیم در این تحقیق به بررسی چند مورد از این الگوریتم‌ها بپردازیم. هدف اصلی ساخت یک سیستم تشخیص خودکار پزشکی برای کمک به پزشکان و همچنین بیماران بود تا بتوانیم هزینه‌های درمان و همچنین خطای تشخیص بیماری را کاهش دهیم. از آنجا که بیماری‌های قلبی نیازمند تشخیص زودرس هستند، هدفمان ترکیب تکنیک‌های مطرح در حوزه پردازش داده و اعمال آن ها بر ویژگی‌های فیزیکی بیمار مانند وزن، جنس، فشارخون و... بود تا با جاسازی یک سیستم تشخیص خودکار در سازمان‌های بهداشتی و درمانی قادر باشیم پس از ورود یک بیمار و دریافت ویژگی‌های موردنظر، این داده‌ها را پردازش کرده و بیان کنیم بیمار با چه احتمالی به مشکلات قلبی دچاراست. درنهایت با بررسی این الگوریتم‌ها کارایی آن‌ها را سنجیدیم و دیدیم که می‌توان با ترکیبشان با الگوریتم‌های فازی و تکاملی نتیجه را بهبود بخشید.

دیدیم که داده‌کاوی جستجوی خودکار منابع داده‌ای بزرگ، جهت یافتن الگوها و وابستگی‌هایی است که تحلیل‌های ساده آماری قادر به انجام آن‌ها نیستند. در علم پزشکی کشف و تشخیص به‌موقع بیماری‌ها می‌تواند از ابتلای افراد به بسیاری از بیماری‌های مهلک جلوگیری نموده و موجب نجات زندگی مردم گردد. این مطالعه نشان می‌دهد پیشگویی‌های داده‌کاوی ابزارهای ضروری را برای محققان، پزشکان و متخصصان حوزه بیماری‌های قلبی جهت بهبود در روش‌های تشخیصی و برنامه‌های درمانی فراهم می‌نمایند.

در سال‌های اخیر، جمع‌آوری داده‌های بیماری‌های مختلف اهمیت فراوانی یافته است. پیشرفت‌های مرتبط با فناوری اطلاعات کمک شایانی به بررسی‌های همه‌جانبه داده‌های حجیم به عمل‌آورده و توانسته است به جستجوی دانش نهفته در آن‌ها پرداخته و علم نوین داده‌کاوی را به وجود آورد.

۵-۲ نتیجه‌گیری

در جهان صنعتی امروز، سرعت به بخش لاینفک زنگی انسان‌ها تبدیل شده است. سرعت در تولید داده و شکل‌گیری بانک‌های عظیمی از آن‌ها، سبب شده است که روش‌ها و مدل‌های ساده دیگر کاربرد چندانی در تحلیل و استخراج دانش نداشته باشند. لذا داده‌کاوی برای جستجو، تحلیل و نتیجه‌گیری از بانک-داده‌های عظیم به روشی کارآمد و مؤثر تبدیل شده‌است.

یکی از زمینه‌هایی که نیازمند استفاده از ابزارهای داده‌کاوی جهت مدل‌سازی پیش‌گویانه با روش‌های محاسباتی جدید است، علم پزشکی می‌باشد که کنترل و تشخیص بخش بزرگی از بیماری‌ها را شامل می‌شود. بیماری‌های قلبی به یکی از علل اصلی مرگ‌ومیر در جهان و ایران تبدیل شده اند و بخش بزرگی از داده‌های مؤثر از تشخیص، پیشگیری و درمان در این حوزه تولید می‌شود که الگوریتم‌های داده‌کاوی نقش محوری در مدیریت و تحلیل صحیح آنها خواهد داشت.

۵-۳ پیشنهادها

در انتها پیشنهاد می‌شود درزمینه‌ی بهبود روش‌ها و الگوریتم‌های داده‌کاوی، نحوه استفاده از آن‌ها در سیستم پشتیبانی تصمیم و همچنین انواع دسته‌بندی ویژگی‌های بیماران مطالعه بیشتری صورت گیرد. این بهبود می‌تواند موجب افزایش دقت این سیستم شده و فرآیند تشخیص و پیش‌بینی بیماری را سرعت ببخشد.

منابع و مراجع

[1] Tan steinbach and kumar Addison, 2nd ed. , “Introduction to Data Mining”, Wesley, 2006.

[2] S. Vijayarani and S. Sudha, “A Study of Heart Disease Prediction in Data Mining”, *International Journal of Computer Science and Information Technology & Security (IJCSITS)* , vol. 2, no. 5, 2012.

[3] Jyoti Soni, “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”, *International Journal of Computer Applications*, vol. 17, no. 8, 2011.

[4] Chandna Deepali, “Diagnosis of heart disease using data mining algorithm” *International Journal of Computer Science and Information Technologies*, 5.2, 2015.

[5] A. Wilson, G. Wilson and J. Likhiya, “Heart Disease Prediction using the Data Mining Techniques”, *International Journal of Computer Science Trends and Technology (IJCST)*, vol. 2, no. 1, 2014.

[6] Vijayashree, J., and N. Ch SrimanNarayanaIyengar. “Heart disease prediction system using data mining and hybrid intelligent techniques: A review”, *International Journal of Bio-Science and Bio-Technology*, 2016.

[7] Bhatla, Nidhi, and Kiran Jyoti, “An analysis of heart disease prediction using different data mining techniques”, *International Journal of Engineering*, 2012.

[8] Shouman, M., T. Turner, and R. Stocker, “Using decision tree for diagnosing heart disease patients”, *9th Australasian Data Mining Conference*, 2011.