



سومین همایش ملی کامپیوتر

National Conference on Computer Science



3

rd

۱۰ اردیبهشت ۱۳۹۴

پیش‌بینی تأثیر عوامل خطرزای بیماری عروق کرونر در مردان با استفاده از تکنیک شبکه عصبی در داده کاوی

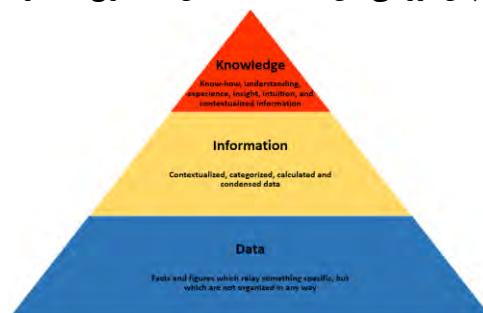
مولود آبدار^۱دانشجوی مهندسی کامپیوتر، دانشگاه دامغان، m.abdar1987@gmail.com^۱**چکیده**

در سال‌های اخیر، آمار مرگ‌ومیر در اثر ابتلا به یکی از انواع بیماری‌های قلبی حدود ۲۵ درصد گزارش شده است. همچنان این بیماری هزینه‌ی زیادی برای بیماران و دولت‌ها به وجود آورده است لذا راهکاری برای کاهش میزان ابتلا به این بیماری می‌تواند کمک زیادی به پایین آوردن این هزینه‌ها کند. در این مقاله با به کارگیری تکنیک‌های مختلف با استفاده از نرم‌افزار IBM SPSS modeler بررسی عوامل خطرزا در مردان پرداخته شده است. برای انجام این تحقیق از داده‌های موجود در اینترنت استفاده شده است. در این پژوهش از مدل Neural Network چندلایه‌ای استفاده شده است. بعد از انجام مراحل حذف فاکتورهایی که تأثیر کمتری در پیش‌بینی داشتند، مدلی با دقت٪ ۸۵.۹ به دست آمد. فاکتورهای exercise angina و chest pain نیز به عنوان مؤثرترین فاکتورهای پیش‌بینی تولید و ثبت شد. میزان درست بودن پیش‌بینی جهت طبقه‌بندی بیماران هم ٪ ۸۰.۹ به دست آمد.

واژه‌های کلیدی: داده کاوی، بیماری عروق کرونر، شبکه‌ی عصبی، نمودار پرسپترون، نرم افزار IBM SPSS Modeler

۱- مقدمه

داده به یک مجموعه خام گفته می‌شود که اگر بر روی آن‌ها مراحلی جهت پردازش تحلیل صورت نگیرد مفید نخواهد بود. داده‌ها بعد از انجام مراحلی به اطلاعات تبدیل شده و سپس بررسی این اطلاعات، دانش استخراج شده از داده‌های اولیه را نتیجه می‌دهد.



شکل(۱): روند تبدیل داده به دانش





سومین همایش ملی کامپیوتر

National Conference on Computer Science

داده‌کاوی امروزه به یکی از موضوعات مهم جهت بیرون کشیدن علم پنهان شده در وجود این داده‌های خام است. امروزه داده‌کاوی وارد پیشتر مباحث علمی شده است، که از آن جمله می‌توان به علوم پزشکی اشاره کرد.

استفاده از داده‌کاوی در پزشکی باعث شده است تا در بیماری‌ها به یکروند جهت پیشگیری و پیش‌بینی دست پیدا کنیم. این نکته قابل توجه است که هر چه پیش‌بینی و پیشگیری دقیق‌تر باشد منجر به نجات بیماران می‌شود. یکی از این نوع بیماری‌ها که بسیار مورد توجه محققان قرار گرفته است، بیماری‌های قلبی-عروقی است. طبق آمارهای ثبت شده، یکی از عوامل اصلی مرگ‌ومیر در سرتاسر جهان، ابتلا به یکی از انواع بیماری‌های قلبی است. بیماری عروق کرونر یکی از انواع بیماری‌های قلبی است که بیشترین درصد مرگ را به خود اختصاص داده است. این بیماری در اثر ایجاد یک لایه چربی میان رگ‌های خونی به وجود می‌آید.

طبق گزارشی که سازمان بهداشت جهانی منتشر کرده است در سال ۲۰۰۸، تعداد افرادی که براثر ابتلا به بیماری عروق کرونر جان خود را از دست داده‌اند حدود ۷/۲۵ میلیون نفر گزارش شده است [۱]. یک گزارش دیگر نیز که در سال ۱۳۸۸ منتشر شده است حدود ۰/۲۵٪ مرگ‌ومیرهای ثبت شده در ایران مختص به بیماران قلبی عروق کرونر بوده است [۲].

استفاده از تکنیک‌های مختلف داده‌کاوی از جمله درخت تصمیم، شبکه‌های عصبی، دسته‌بندی و... به ما کمک می‌کند تا روش‌های مختلف را بررسی کرده سپس با توجه به نتایج حاصل شده بهترین تکنیک را معرفی کنیم.

در این مقاله سعی شده است به بررسی بیماری قلبی در مردان و تأثیر عوامل مختلف آن موردنظری قرار گیرد. به این منظور در قسمت دوم به معرفی مختصراً از بیماری‌های قلبی و علائم **ویاکتوروگرافی** خطر آن و معرفی داده‌ها و تعداد آن‌ها می‌پردازد. در قسمت سوم به تشریح روش انجام تحقیق و مراحل انجام کار اختصاص پیدا کرده است. قسمت آخر نیز به نتیجه‌گیری پژوهش اختصاص پیدا کرده است.

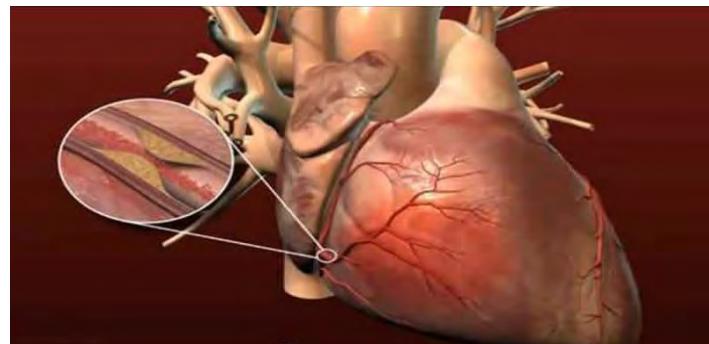
۱- بیماری عروق کرونر:

بیماری عروق کرونر شایع‌ترین نوع بیماری قلبی است. بیماری عروق زمانی اتفاق می‌افتد که تجمع کلسترول و مواد دیگر به نام پلاک در دیواره داخلی زیاد شده و این کار باعث به وجود آمدن سختی و تنگی در عضله‌ها می‌شود. همین امر می‌تواند به درد قفسه سینه (آنژین صدری) و یا حمله قلبی منجر شود. این بیماری به علت تجمع غیرطبیعی لیپید، مواد چربی و بافت چربی در جدار رگ ایجاد و موجب بلوکاژ، تنگی رگ و کاهش جریان خون به عضله قلب می‌شود. جراحی قلب و عروق مذکور بودن، سابقه ژنتیکی، افزایش چربی خون، استعمال دخانیات، افزایش فشارخون، کاهش سطح کلسترول HDL، چاقی مفرط را از جمله علل ابتلا به بیماری تنگی عروق کرونر است. کنترل منظم فشارخون، قند و چربی، کنترل بیماری دیابت در صورت ابتلا، قطع استعمال دخانیات، ورزش‌های منظم، رژیم غذایی مناسب و کاهش مصرف غذاهای آماده (غذای فوری) از راهکارهای پیشگیری از این بیماری است [۳]، [۴].



سومین همایش ملی کامپیوتر

۳rd National Conference on Computer Science



شکل(۲) : بیماری عروق کرونر

آزمایش‌های مختلفی برای تشخیص بیماران قلبی وجود دارد که از آن جمله می‌توان به موارد زیر اشاره کرد:

- ۱. تست‌های غربالگری
- ۲. الکتروکاردیوگرام
- ۳. تست ورزش
- ۴. تست تالیوم
- ۵. اکوکاردیوگرافی
- ۶. آنژیوگرافی
- ۷. ونتریوکولرگرام
- ۸. اولتراسوند در عروق کرونری :

- داپلر سونوکاروتید
- توموگرافی کامپیوتراپتیک
- تست خونی بسیار حساس سی راکتیو پروتئین

۲-۱- مجموعه داده‌ها:

برای انجام این مقاله از مجموعه داده‌های بیماران عروق کرونر مربوط به مردان موجود در منابع اینترنتی استفاده شده است [۵]. این داده‌ها شامل ۲۰۹ بیمار است که ۱۹ بیمار برای تست و ۱۹۰ بیمار برای آموزش به کار گرفته شده است. هر کدام از بیماران دارای فاکتورهای زیر می‌باشد:

- ۱: سن : از ۲۸ سال تا ۶۶ سال
- ۲: non_anginal , atyp_angina , asympt : chest_pain
- ۳: rest_bpress : از ۹۲ تا ۲۰۰
- ۴: blood_sugar : درست یا غلط (F or T)



سومین همایش ملی کامپیوتر

National Conference on Computer Science



(Null) ? و st_t_wave_abnormality , left_vent_hyper , normal : rest_electro : ۵

۶ : از ۸۲ تا ۱۸۸ : max_heart_rate

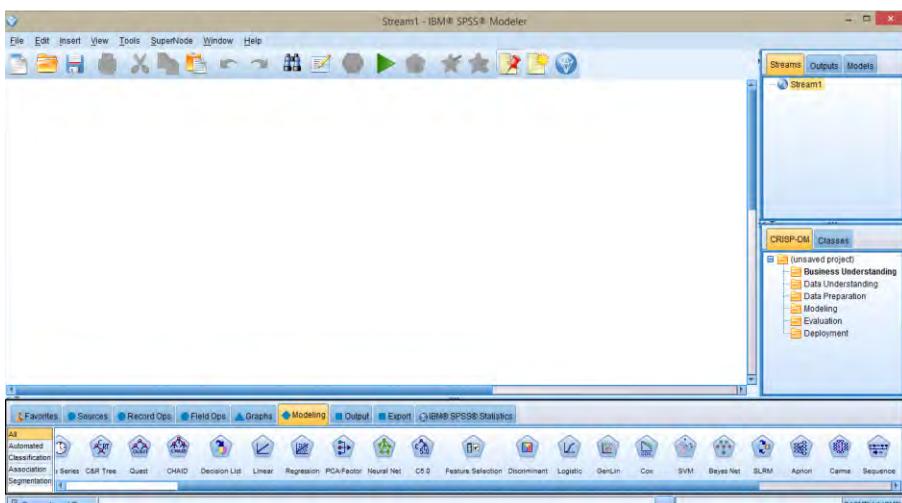
(Yes or No : بله یا خیر) exercise_angina : ۷

(Yes or No : بله یا خیر) disease : ۸

(test و train) : آموزش و آزمایش (status) : وضعیت (۹)

۱-۳- معرفی مختصر نرم افزار IBM SPSS modeler

برای انجام تحقیق در زمینه‌ی داده کاوی نرم افزارهای زیادی از جمله IBM SPSS .Microsoft Analysis Services .SAS Data Mining .Angoss Knowledge STUDIO .R .Miner را می‌توان نام برد. نرم افزار IBM SPSS Modeler قبل از ورود به بازار به اسم Clementine شناخته می‌شد و توسط شرکت SPSS طراحی شده بود. بعد از آن که شرکت IBM سهام شرکت SPSS را خرید این نرم افزار به IBM SPSS Modeler تغییر یافت. این نرم افزار قدرتمند جزو ۱۳ نرم افزار برتر داده کاوی محسوب می‌شود [۶]. زبان برنامه نویسی شده‌ی این نرم افزار Java می‌باشد. این نرم افزار برای پیش‌بینی و آنالیز داده‌ها با استفاده از مدل‌های زیادی از جمله Neural Network .KNN .SLRM .Apriori .SVM .CHAID .C&R tree .net .Bhernes می‌گیرد.



شکل (۳) : نمایی از نرم افزار IBM SPSS Modeler

۱-۴- معرفی بر کارهای صورت گرفته بر روی بیماران عروق کرونر با استفاده از داده کاوی

در مقاله‌ی Roohallah and Jafar .Bagging .SMO از الگوریتم‌های Neural Network استفاده شده است. آن‌ها از داده های Z-Alizadeh Sani که شامل ۳۰۳ بیمار با ۵۴ ویژگی می‌باشد، را به کار گرفته‌اند. سپس به این نکته اشاره شده که



سومین همایش ملی کامپیوتر

National Conference on Computer Science



۳

rd

۱۰ اردیبهشت ۱۳۹۴

و

تر

بیشترین تاثیر بهره وری را نسبت به بقیه‌ی ویژگی‌ها به خود اختصاص داده‌اند، که دقت گزارش شده آن‌ها ۹۴.۰۸٪ بوده است [۷]. در مقاله‌ی دیگری که توسط Mohammad and Assal Mohammad انجام شده است، به مقایسه‌ی ANN و ANFIS با به کارگیری نرم افزار MATLAB می‌پردازد. دقت داده‌های آموزشی برای ANFIS برابر ۱۰۰٪ و برای ANN برابر ۷۵.۹۳٪ محاسبه شده است. این در حالی است که دقت داده‌های آزمایشی برای ANN برابر ۸۷.۰۴٪ و برای ANFIS برابر ۹۰.۷۴٪ محاسبه شده است. این در دست آمده است [۸].

یکی از تحقیقات که توسط Sumit and Parveen and Pillai صورت گرفته است، به بررسی تکنیک SVM پرداخته شده است که دقت زمانی که از پنج کلاس استفاده شود حدود ۵۵.۷٪ به دست آمده است در عوض وقتی آن را با ۲ کلاس افراد مريض و افراد سالم بررسی نموده اند دقت ۹۰.۵۷٪ را ثبت کرده اند [۹].

در تحقیق با ارزش دیگری که توسط Jasmin and Tasaduq and Kevin and Yi-Ping صورت گرفته، به نقش ارتباط بین عوامل بیماری‌های قلبی برای شناسایی در زنان و مردان می‌پردازد و اشاره به این دارد احتمال این که زنان در معرض ابتلاء به بیماری کرونر قلبی (Coronary heart disease) قرار بگیرند نسبت به مردان کمتر است. با انجام ورزش در زنان و مردان می‌توان بر Chest pain غلبه کرد. یکی از نکات استخراج شده در مقاله این است که فاکتورهای Rest ECG به دو صورت normal و hyper و Slope being flat فقط به عنوان فاکتورهای خطر معرفی می‌شوند، این در حالی است که برای مردان Rest ECG مقایسه شده اند که نتیجه آن به شرح زیر گزارش شده است: با نقش میزان اطمینان بالای ۹۰٪، میزان دقت بالای ۹۹٪ و میزان تصدیق بالای ۷۹٪ به ترتیب برای مدل‌های Predictive Apriori و Tertius [۱۰].

در یک بررسی که بر روی داده‌های جمع آوری شده مابین سال‌های ۱۹۹۹ تا ۲۰۱۳ بوده است، به کشف دانش در پزشکی می‌پردازد. هر بخش از مقاله را به یکی از ۶ وظایف (task) پزشکی اختصاص داده‌اند، که شامل موارد زیر است: screening, diagnosis, treatment, prognosis, monitoring and classification, regression, management clustering, association and hybrid و همچنین ادغام آن‌ها جهت استخراج اطلاعات پزشکی و داده‌کاوی میان ۲۹۱ مقاله چاپ شده در سال‌های مذکور می‌پردازد [۱۱].

۲- روش کار

برای انجام این پژوهش از مدل‌سازی سه لایه‌ای شبکه عصبی MLP استفاده شده است. برای شروع کار، همه‌ی فاکتورهای موجود در قسمت ۴ در مدل شرکت داده شد. برای این‌که عامل‌هایی که تأثیر کمتری دارند را حذف کنیم، از تکنیک گزینش و حذف با توجه به مدل C&R Tree استفاده شد. برای این کار در هر مرحله فاکتوری که کمترین تأثیر را داشت، حذف شد. پس از انجام ۵ مرحله حذف به فاکتورهایی که بیشترین تأثیر را در پیش‌بینی داشتند، به دست آمد.

در این مدل ۳ هدف (target) مختلف یعنی سن، rest_bpress .max_heart_rate موردنرسی قرار گرفت. در قسمت ورودی (Input) بهجز فاکتور هدف بقیه‌ی فاکتورها دخالت داده شد. با قرارگیری سن در هدف، فاکتور rest_bpress با ۰.۵۳ و

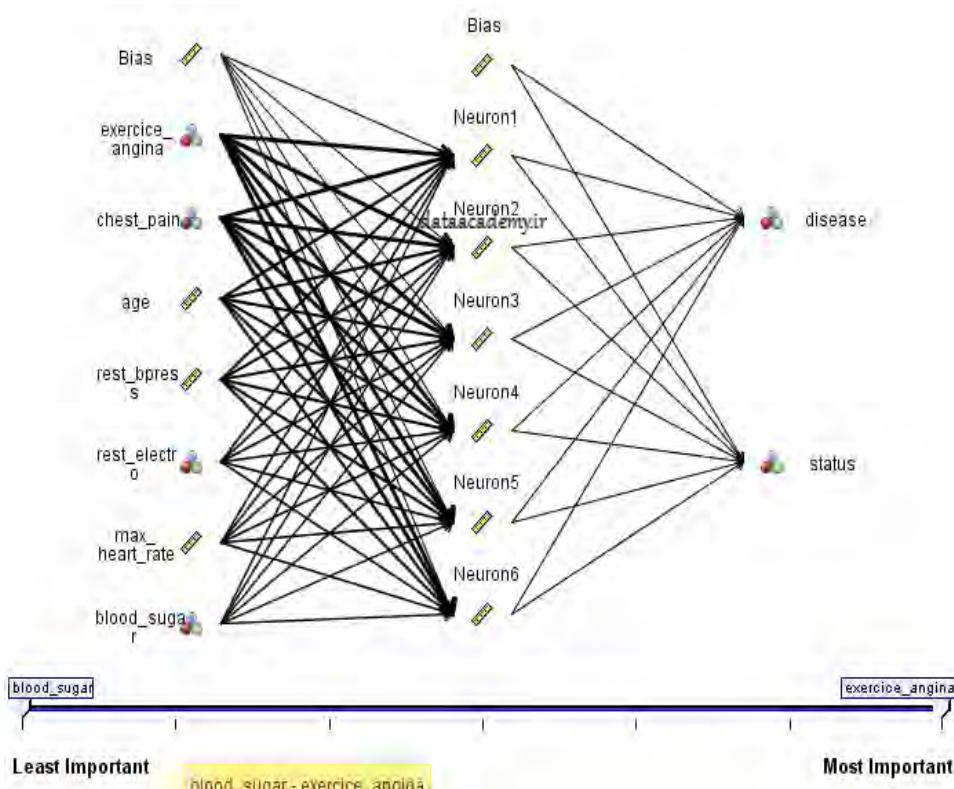




سومین همایش ملی کامپیوتر

National Conference on Computer Science

با ۰.۴۷٪ مؤثرین فاکتورها جهت پیش‌بینی بودند. در مورد rest_bpress این فاکتورها به ترتیب با ۰.۶۱٪ و سن با ۰.۳۹٪ مؤثرین فاکتورها برای پیش‌بینی به دست آمد. اما نکته‌ی جالب در مورد max_heart_rate این بود که در اولین قدم به هنگام ورود تمامی فاکتورها، chest pain به عنوان max_heart_rate مؤثرترین فاکتور بود، ولی بعد از انجام ۵ مرحله حذف عوامل کم تأثیر، exercise_angina با ۰.۵۳٪ و chest pain با ۰.۴۷٪ مؤثرترین فاکتورها بودند. در مرحله‌ی بعدی مدل neural network انتخاب شد. برای قسمت هدف دو فاکتور disease و status و همچنین برای قسمت ورودی سن و rest_electro و blood_sugar و chest_pain و rest_bpress در نظر گرفته شد. شبکه عصبی مدل موردنظر با سه لایه‌ی اول ورودی و لایه‌ی وسط لایه‌ی پنهان و لایه‌ی سوم، خروجی نامیده می‌شود به شکل زیر حاصل گردید:



شکل(۴) : نمودار پرسپترون حاصل از شبکه عصبی پیش‌بینی بیماری عروق کرونر

همان‌گونه که در شکل ۴ مشاهده می‌شود لایه‌ی ورودی دارای ۸ نرون و لایه‌ی خروجی ۲ نرون که یکی بیماری و دیگری وضعیت را نشان می‌دهد. دقت بدست‌آمده در این مدل ۸۵.۹٪ بوده است که در شکل زیر نتیجه‌ی حاصل قابل مشاهده است. با توجه به شکل می‌توان دریافت که برای در نظر گرفتن هدف (targets)، دو فاکتور در نظر گرفته شده است. دلیل این کار این است، که با در نظر گرفتن یک هدف میزان دقت ثبت‌شده کمتر بود لذا برای بالا بردن دقت تحقیق، حالتهای مختلف برای هدف در نظر گرفته شد. پس از امتحان حالتهای مختلف، حالت نهایی که بیشترین دقت را داشت، انتخاب گردید.

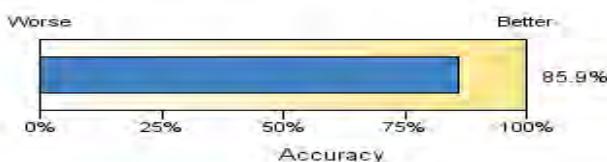


سومین همایش ملی کامپیوتر

National Conference on Computer Science

Model Summary

Targets	disease status
Model	Multilayer Perceptron
Stopping Rule Used	Error cannot be further decreased
Hidden Layer 1 Neurons	6

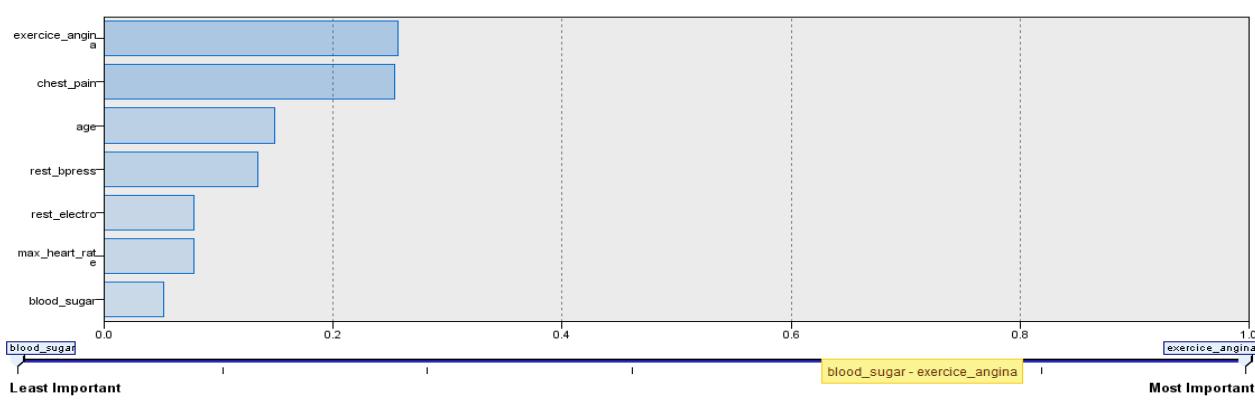


شکل (۵) : دقت محاسبه شده توسط شبکه عصبی

در ادامهی تحقیق به تقسیم فاکتورهای موردنظر بر اساس اهمیت در پیش‌بینی برای فاکتورهای هدف پرداخته شد. بیشترین تأثیر را در پیش‌بینی به خود اختصاص دادند. شکل زیر ترتیب این تأثیر را نشان می‌دهد:

Predictor Importance

Targets: disease, status



شکل (۶) : میزان تأثیر برگ فاکتورها در پیش‌بینی

بعد از مشاهدهی ترتیب تأثیر فاکتورها، طبقه‌بندی برای بیماران را نیز که میزان درست بودن درصد آن در کل مجموعه برابر ۸۰.۹٪ بود، به دست آمد. شکل ۶ این نتیجه را نشان می‌دهد:



سومین همایش ملی کامپیوتر

۳rd National Conference on Computer Science



Classification for disease

Overall Percent Correct = 80.9%

Observed	Predicted		Row Percent
	negative	positive	
negative	85.5%	14.5%	
positive	25.0%	75.0%	

شکل (۷) : طبقه‌بندی بیماران

تحلیل این نتیجه که در شکل ۶ آورده شده است، بدین صورت است که، زمانی که تعداد افرادی که بیماری ندارند به عنوان ورودی مدل پیشنهادی قرار گیرند ۸۵.۵٪ به صورت صحیح آن‌ها را پیش‌بینی کرده و سپس طبق‌بندی می‌کند. در مورد تعداد افرادی که بیمار هستند، پیش‌بینی مثبت بودن بیماری ۷۵٪ به دست آمده است و این بدین معناست که ۷۵٪ افرادی را که بیمار هستند، را درست پیش‌بینی کرده و سپس طبقه‌بندی را انجام می‌دهد.

۳- نتیجه گیری

در این مقاله برای پیش‌بینی عوامل مهم و خطرزا در مردان با توجه به مدل شبکه‌ی عصبی موردنظری قرار گرفت. با توجه به نتایج حاصل از به کار گیری مدل C&R Tree زمانی که هدف فاکتور max_heart_rate بود، فاکتورهای exercise_angina و chest_pain به عنوان مؤثرترین فاکتورها در پیش‌بینی به دست آمد. بعد از انجام مراحل با به کار گیری مدل شبکه‌ی عصبی که دقیق آن ۸۵.۹٪ محاسبه شد، همچنین فاکتورهای ذکر شده بیشترین تأثیر را در پیش‌بینی به خود اختصاص دادند. این نتیجه نشان می‌دهد در نظر گرفتن این دو فاکتور با هم دیگر کمک زیادی جهت تشخیص بیماری عروق کرونر در مردان می‌کند. هدف از این تحقیق این بود تا راهی برای پیش‌بینی به موقع برای پیشگیری از به وجود آمدن بیماری‌های قلبی عروق کرونر در مردان به دست آید. با توجه به این هدف، مدل پیشنهادی با فاکتورهایی که از داده‌ها در اختیار بود، معرفی گردید.

۴- مراجع

- Fact sheet No. 310 : *The top ten causes of death*. Geneva: World Health Organization; 2011.
- محمودی، عیسی؛ عسکری مقدم، رضا؛ معظم، محمد هادی؛ صادقیان، سعید. مدل پیش‌بینی بیماری عروق کرونر با استفاده از شبکه‌های عصبی و گزینش متغیر مبتنی بر درخت رگرسیون و طبقه‌بندی. مجله دانشگاه علوم پزشکی شهر کرد، دوره ۱۵، شماره ۵، آذر و دی ۹۲-۵۶.
- Homoud, M. K; *Coronary Artery Disease*, Tufts-New England Medical Center Spring 2008.
- Deckelbaum, L; *HEART ATTACKS AND CORONARY ARTERY DISEASE, MAJOR CARDIOVASCULAR DISORDERS*.



سومین همایش ملی کامپیوتر National Conference on Computer Science

5. informatykplus.edu.pl/upload/list/czytelnia/heart_disease_male.xls.
6. <http://www.predictiveanalyticstoday.com/top-data-mining-software3.html>.
7. Alizadehsani R; Habibi, J; Hosseini, M. J; Mashayekhi, H; Boghrati, R; Ghandeharioun, A; Bahadorian, B; Alizadeh Sani, Z; *A data mining approach for diagnosis of coronary artery disease*, computer methods and programs in biomedicine 111 (2013) 52–61.
8. Abushariah, M. A. M; Alqudah, A. A. M; Adwan, O, A; *Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches*, Journal of Software Engineering and Applications, 2014, 7, 1055-1064.
9. Bhatia S; Prakash P; Pillai, G. N; *SVM Based Decision Support System for Heart Disease Classification with Integer-Coded Genetic Algorithm to Select Critical Features*, Proceedings of the World Congress on Engineering and Computer Science 2008, WCECS 2008, October 22 - 24, 2008, San Francisco, USA.
10. Nahar, J; Imam, T; Tickle, K. S; Phoebe Chen, Y; *Association rule mining to detect factors which contribute to heart disease in males and females*, Expert Systems with Applications 40 (2013) 1086–1093.
11. Esfandiari, N; Babavalian, M. R; Eftekhari Moghadam, A. M; Kashani Tabar, V; *Knowledge discovery in medicine: Current issue and future trend*, Expert Systems with Applications 41 (2014) 4434–4463.

dataacademy.ir