DOI: 10.22100/jkh.v11i4.1525

مقاله پژوهشی

مجله دانش و تندرستی، دوره ۱۱، شماره ۴، زمستان ۱۳۹۵

دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شاهرود

کشف الگوی وضعیت اهداکنندگان خون از طریق خوشه‌بندی: روشی برای افزایش کیفیت خدمات در مراکز اهدای خون

مریم عاشوری*[۱]، شهریار محمدی[۲]، هدی‌سادات حسینی ایوری[۳]

۱- مجتمع آموزش عالی سراوان- دانشکده فنی و مهندسی- گروه فناوری اطلاعات- مربی.

۲- دانشگاه صنعتی خواجه نصرالدین طوسی- دانشکده مهندسی صنایع- گروه فناوری اطلاعات- استادیار.

۳- دانشگاه آزاد فردوس- گروه مهندسی کامپیوتر- کارشناس ارشد.

تاریخ دریافت: ۱۳۹۵/۴/۷، **تاریخ پذیرش:** ۱۳۹۵/۱۰/۶

چکیده

مقدمه: *نیاز فوری به خون و عدم جایگزین مناسب برای آن، ضرورت وجود الگویی برای کمک به پزشکان جهت ارایه خدمات درست به اهداکنندگان و مدیریت صحیح پایگاه خون را نشان می‌دهد. پژوهش حاضر با هدف شناسایی الگویی برای تشخیص وضعیت اهداکنندگان خون صورت گرفته است.*

مواد و روش‌ها: *پژوهش حاضر به روش پیمایشی- مقطعی و به‌صورت سرشماری صورت گرفته است. جامعه پژوهش متشکل از داده‌های سازمان انتقال خون بیرجند در ماه‌های خرداد تا شهریور ۹۲ است که با مراجعه مستقیم پژوهش‌گر به سازمان و به‌صورت فایل اکسل تهیه گردید. جهت تحلیل داده‌ها از نرم‌افزار Clementine12.0 استفاده شده است. در پژوهش حاضر ابتدا خوشه‌بندی Two-Step و سپس الگوریتم‌های C5.0، C&R Tree، CHAID و QUEST اجرا گردیدند تا بهترین نسبت بین فیلدهای مختلف به‌دست آید.*

نتایج: *مقدار صحت به‌دست آمده از اجرای الگوریتم‌های C5.0، C&R Tree، CHAID و QUEST به‌ترتیب ۰/۹۹۹۸، ۰/۹۹۶۰، ۰/۹۹۳۰، ۰/۸۹۱۳ می‌باشد. مقادیر به‌دست آمده برای شاخص‌های حساسیت، شفافیت، صحت، دقت، شاخص F، میانگین هندسی، نرخ مثبت غلط، نرخ منفی غلط و نرخ خطا برای مدل C5.0 نشان‌دهنده عملکرد بهتر این الگوریتم نسبت‌به سایرین می‌باشد. تأثیرگذارترین شاخص‌ها در تولید مدل، دسته فشارخون، وضعیت اهدای خون و دمای بدن هستند.*

نتیجه‌گیری: *مدل ارایه شده به پیش‌بینی سریع‌تر و دقیق‌تر وضعیت اهدای خون و نیز مدیریت صحیح پایگاه خون کمک می‌نماید و می‌تواند گامی مؤثر جهت استفاده کارآمد از خون اهدایی و کاهش هزینه‌های نگهداری خون محسوب گردد.*

واژه‌های کلیدی: اهداکنندگان خون، داده کاوی، درخت تصمیم، خوشه‌بندی.

***نویسنده مسئول:** سراوان، بلوار پاسداران، جنب ساختمان فرمانداری، مجتمع آموزش عالی سراوان، **تلفن:** ۰۵۴-۳۷۶۳۰۰۹۲، **نمابر:** ۰۵۴-۳۷۶۳۰۰۹۰ ،

Email: mashoori@saravan.ac.ir

ارجاع: عاشوری مریم، محمدی شهریار، حسینی ایوری هدی‌سادات. کشف الگوی وضعیت اهداکنندگان خون از طریق خوشه‌بندی: روشی برای افزایش کیفیت خدمات در مراکز اهدای خون. مجله دانش و تندرستی ۱۳۹۵؛۱۱(۴):۷۳-۸۲.

مقدمه

امروزه با افزایش شمار تصادفات و مشکلات سلامت، نیاز به خون افزایش یافته است (۱). خون عنصر حیاتی زندگی بشر است و هیچ جایگزینی برای آن موجود نیست. طبق استاندارد صلیب سرخ آمریکا، خون اهدا شده حداکثر تا ۴۲ روز قابل استفاده است (۲). هر ساله میلیون‌ها واحد خون مصرف می‌گردد و به همین علت نیاز به اهداکنندگان سالم خون مورد توجه است (۳). اهدای خون و سرویس انتقال آن جزء جدایی‌ناپذیر سیستم سلامت می‌باشد (۴) و مجموعه‌ای از عملیات وابسته بهم شامل ثبت نام اهداکننده، ارزیابی سلامت اهداکننده، جمع‌آوری خون، غربال‌گری خون، تولید فرآورده، مدیریت موجودی، و انتشار خون است (۵). در مرحله ارزیابی سلامت اهداکننده، اهداکنندگان به اتاق‌های چکاب برای سنجش دمای بدن، فشارخون، ضربان قلب به جهت ارزیابی سابقه سلامت اهداکننده هدایت می‌گردند. در صورت عدم وجود محدودیت برای اهدای خون مرحله جمع‌آوری خون یا خون‌گیری صورت می‌گیرد (۳). آلفونسو و دیگران به نقل از سازمان سلامت جهانی بیان کردند که هر کشوری با چالش جمع‌آوری خون کافی از اهداکنندگان سالم مواجه است (۶) و از طرفی اولویت سرویس انتقال خون اطمینان از سلامت، کفایت، دسترس‌پذیری و عرضه ذخیره خون در همه سطوح است (۴).

در نهایت پس از جمع‌آوری خون اهدا شده مسأله مدیریت خون اهمیت می‌یابد. مدیریت خون به‌عنوان یک وظیفه چالش‌برانگیز شناخته شده است زیرا ماهیت تهدیدکننده فرآورده‌های خونی مستلزم اداره دقیق می‌باشد در حالی که ماهیت فاسد شدنی آن مستلزم پردازش به موقع است (۷). بنابراین خون اهدا شده باید به سرعت مورد استفاده قرار گیرد در غیر اینصورت از بین خواهد رفت (۲). بنابراین جمع‌آوری و نگهداری ذخیره خون کافی و نیز انتقال آن یک چالش به‌ویژه در زمان وقوع حوادث و فجایع است. به همین علت فراهم‌آوردن خدمات درست به‌صورت فوری برای اهداکنندگان و مدیریت صحیح پایگاه‌های خون ضروری است و نیاز به استفاده از علوم جدید مورد توجه است.

فرآیند کاوش و تحلیل حجم بزرگی از داده‌ها به کمک کامپیوتر به‌منظور استخراج دانش معنادار، داده کاوی یا کشف دانش نام دارد. داده کاوی برای برآوردن اهداف زیر به‌وجود آمده است: پیش‌بینی ناشناخته‌ها یا مقادیر آینده متغیرها با استفاده از بعضی مقادیر شناخته شده با روش‌های پیش‌بینی و شناسایی الگوهای توصیف داده به روش قابل فهم برای انسان با روش‌های توصیفی. وظایف داده کاوی عبارتند از: دسته‌بندی یا طبقه‌بندی، خوشه‌بندی، کشف قواعد تلازمی، کشف الگوی ترتیبی، رگرسیون و آشکارسازی انحراف (۸). تکنیک‌های داده کاوی مشکلات مربوطبه ابعاد بزرگ مجموعه داده را حل می‌نماید (۹) و در ساخت مدل‌های پیش‌بینی کننده مورد استفاده قرار می‌گیرند (۱۰) درخت‌های رگرسیون و طبقه‌بند برای مدل‌های پیش‌بینی‌کننده مورد استفاده قرار می‌گیرند (۱۱) که این درختان با ترکیب تعاملات و خصوصیات انتخابی و نمایش تفسیر شهودی، کارآمد هستند و صحت پیش بینی بالایی دارند (۱۲). مدل‌ها با افراز بازگشتی داده‌ها به‌طور بهینه و انتخاب یک مدل پیش‌بینی ساده در هر افراز به‌دست می‌آیند. نتایج افرازها به‌صورت گرافیکی با درخت تصمیم نشان داده می‌شود. درخت‌های تصمیم برای متغیرهای وابسته مورد استفاده قرار می‌گیرند (۱۱) نتایج درخت تصمیم قابل درک، کوتاه و شامل قوانین طبقه‌بندی کارآمد است و نتایج آن اغلب قابل افزودن به پرونده بیماران می‌باشد (۱۳)

سانتهانام و ساندرام (۲۰۱۰) از الگوریتم CART برای استخراج مجموعه‌ای از قواعد جهت شناسایی رفتار اهداکنندگان خون استفاده نمودند (۱۴). فعالیت‌های انجام شده در حوزه خون و در سال ۲۰۱۱ عبارتند از: توسعه سیستمی برای تعیین عدم شباهت‌ها در رفتار اهداکنندگان خون فعلی و تمایل آنها به اهدای خون داوطلبانه با استفاده از الگوریتم خوشه‌بندی K-Means و طبقه‌بندی با الگوریتم‌های Bayes Naive و NB tree و Decision Tree (۱۵)، ارایه سیستمی برای تحلیل دقیق داده‌های اهداکنندگان خون با استفاده از الگوریتم طبقه‌بند J48 (۱۶)، توسعه سیستمی برای مدیریت اهدای خون آنی و آزمودن نتایج دسته‌بندی اهداکنندگان خون با الگوریتم RVD (۱۷) و استفاده از الگوریتم‌های طبقه‌بند CART-RVD و CART-DB2k7 به‌منظور شناسایی داوطلبان اهدای خون بر اساس الگوهای اهدای خون (۱۸).

تستیک و همکاران با دسته‌بندی از طریق خوشه‌بندی و با استفاده از الگوریتم‌های Two-step و CART به شناسایی الگوهای ورود ساعتی و روزانه اهداکنندگان خون پرداختند (۳). رموا و همکاران (۲۰۱۲) با الگوریتم‌های طبقه‌بند C4.5 و J48 به ارایه چارچوبی برای تصمیم‌گیری در پایگاه خون پرداختند (۱۹). پیش‌بینی تعداد اهداکنندگان خون در یک سن و گروه خونی خاص با طبقه‌بند J48 نیز توسط شارما و گوپتا (۲۰۱۲) صورت گرفت (۲۰). در سال ۲۰۱۳ ونکاتسوارلو و پراساد با K-Means بهبودیافته به استخراج اطلاعات اهداکنندگان در زمان محاسباتی کمتر دست یافتند (۱).

در سال ۲۰۱۴، مقایسه الگوریتم‌های داده کاوی برای انتقال پلاکت خون با استفاده از طبقه‌بندهای Decision Table, Naïve Bayes و J48 (۲۱)، ارایه سیستمی برای حل مشکلات انتقال خون با کمترین زمان توسط طبقه‌بندهای Naive Bayes، J48 و Random Tree (۲۲) و پیش‌بینی رفتار اهداکنندگان خون با طبقه‌بندهای J48، Bayes net، ZeroR، Ridor، PART، Prism، Naïve Bayes و CBA (۲۳) صورت گرفت. عاشوری و همکاران (۲۰۱۵) با طبقه‌بندهای C5.0، QUEST، CART و CHAID به پیش‌بینی رفتار مستمر اهداکنندگان سالم خون پرداختند (۲۴). همچنین طبقه‌بندی با الگوریتم‌های Naive Bayes، J48، CBA (Classification Based Association) و Random Tree و خوشه‌بندی با الگوریتم K-Means منجر به ارایه تکنیکی برای توسعه

سیستم تحلیل‌کننده پایگاه اهدای خون براساس الگوریتم‌های طبقه‌بندی و سپس خوشه‌بندی نتایج برای به‌دست آوردن گروه خونی اهداکنندگان گردید (۲۵).

بررسی مطالعات انجام شده از سال ۲۰۰۰ تاکنون در حوزه خون و داده کاوی نشان می‌دهد که تلاش‌های صورت گرفته در حوزه خون با علم داده کاوی مربوطبه سال‌های اخیر و یک حوزه مطالعاتی جدید می‌باشد. از طرفی در تلاش‌های صورت گرفته، ارایه الگویی برای پیش‌بینی وضعیت اهدای خون کمتر مورد توجه قرار گرفته است. در پژوهش حاضر باتوجه به شاخص‌های مهم در اهدای خون به کشف الگویی برای پیش‌بینی وضعیت اهدای خون با درخت تصمیم پرداخته شده است. با کمک این الگو می‌توان به پزشکان جهت پیش‌بینی وضعیت اهداکنندگان خون (اولین بار– مستمر و با سابقه) به‌منظور ارایه خدمات درست آنی و نیز مدیریت صحیح پایگاه خون کمک نمود. بنابراین می‌توان امیدوار بود که گامی مؤثر در جهت استفاده کارآمد از خون اهدایی برداشته شود و هزینه‌های مربوطبه نگهداری خون کاهش یابد. از این رو در پژوهش حاضر از درخت‌های تصمیم تقسیم و رگرسیون (Classification and Regression Tree: C&R Tree or CART)، درخت آماری کارا و بی طرف سریع (Quick Unbiased and Efficient Statistical Tree: QUEST)، آشکارساز تعامل خودکار مجذور کای (Chi-Squared Automatic Interaction Detector: CHAID) و درخت C5.0 برای ساخت مدل پیش‌بینی در حوزه خون استفاده گردیده است.

مواد و روش‌ها

مطالعه حاضر از نوع پیمایشی– مقطعی بوده و مجموعه داده‌های آن متعلق به سازمان انتقال خون بیرجند است که به‌صورت سرشماری صورت گرفته است. ۱۶۲۲ نفر در فاصله خرداد تا شهریور ۹۲ را در برمی‌گیرد که به‌صورت فایل اکسل تهیه گردید. یکی از روش‌های بسیار قوی برای پیاده‌سازی و اجرای پروژه‌های داده کاوی متدولوژی CRISP (Cross industry process for data mining) است (۲۶). در پژوهش حاضر مدل پیشنهادی براساس CRISP ارایه شده است که شامل پنج فاز است. هر یک از این فازها خود شامل زیر بخش‌هایی می‌شوند. حرکت رو به جلو و عقب بین فازهای مختلف نیاز است، زیرا ورودی هر فاز به خروجی فاز مرحله قبل وابسته است (۲۷). هر یک از این پنج فاز در شکل ۱ نشان داده شده‌اند.

الف) شناخت سیستم

باتوجه به ماهیت تهدیدکننده، فاسد شدنی و حیاتی خون و فرآورده‌های آن، بررسی داده‌های جمع‌آوری شده مرتبط با خون می‌تواند مفید باشد. با استفاده از الگوی وضعیت اهدای خون می‌توان وضعیت اهداکنندگان آتی را مشخص نموده و از بروز چالش‌های ذکر شده در مواقع بروز بحران جلوگیری به‌عمل آورد. همچنین می‌توان به مدیریت صحیح پایگاه خون پرداخت.

ب) شناخت داده‌ها و آماده‌سازی آنها

در مرحله شناخت داده‌ها ویژگی‌های آزمایشگاهی بیماران بررسی و شناسایی گردید و در مرحله آماده‌سازی داده‌ها، برای پاکسازی مجموعه داده با نظر افراد خبره اعمال زیر انجام گرفت. رکوردهایی که در این مرحله دارای مقادیر از دست رفته بودند یا مقادیر دور افتاده از سایر داده‌ها داشتند، حذف گردیدند. در مرحله آماده‌سازی داده‌ها تعداد اهداکنندگان به ۱۶۲۲ مورد رسید. داده‌ها با مراجعه مستقیم پژوهشگر به پایگاه انتقال خون و به‌صورت فایل اکسل تهیه گردید که محتوای داده‌ها مورد تأیید متخصصان حوزه مربوطه می‌باشد. آزمون طبیعی‌بودن داده‌ها به‌روش کولموگرف اسمیرنف و با فرض P>۰/۰۵ روی متغیرهای عددی صورت گرفت.

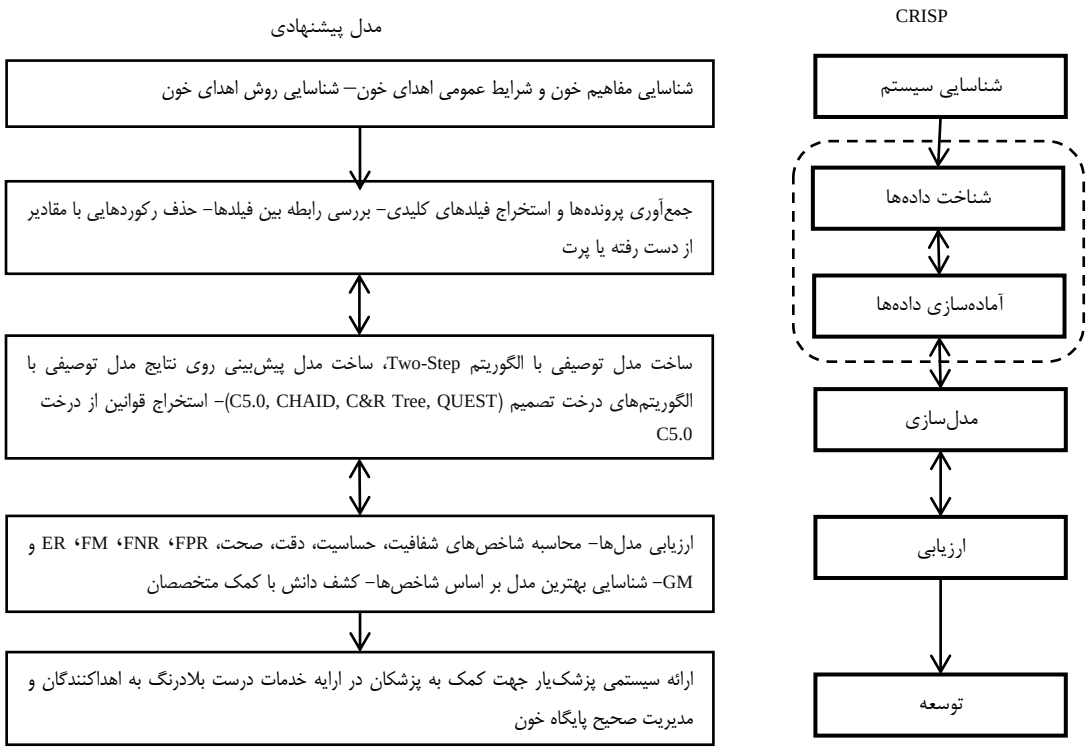
ج) مدل‌سازی

مدل‌سازی با استفاده از نرم‌افزار SPSS Clementine12.0 انجام شده است. روش کار در پژوهش حاضر داده‌کاوی پیش‌بینانه از طریق داده کاوی توصیف‌کننده می‌باشد. خوشه‌بندی یک فرآیند غیرنظارتی برای گروه‌بندی عناصر شبیه در خوشه‌ها می‌باشد. دسته‌بندی می‌تواند مبتنی‌بر خوشه‌بندی اجرا گردد در صورتی‌که اطلاعات دسته یا کلاس برای ارزیابی خوشه‌های به‌دست آمده استفاده شود. این رویکرد مبتنی‌بر روال ارزیابی "خوشه به دسته" است و یک نگاشت با حداقل خطا از خوشه‌ها به کلاس‌ها را می‌یابد (۲۸). در پژوهش حاضر ابتدا از خوشه‌بندی Two-Step استفاده گردیده است و سپس از الگوریتم‌های درخت تصمیم استفاده شده است تا بهترین نسبت بین فیلدهای مختلف به‌دست آید (۲۷). خوشه‌بندی Two-Step به‌طور اتوماتیک تعداد خوشه‌ها را تشخیص می‌دهد. این الگوریتم ابتدا نمونه‌ها را در خوشه‌ها قرار داده و سپس توسط یک الگوریتم سلسله مراتبی خوشه‌های حاصل را با یکدیگر ترکیب می‌نماید (۲۹). برای آموزش درخت تصمیم یک متغیر طبقه‌ای باید فیلد خروجی باشد و یک یا تعداد بیشتری فیلد ورودی وجود داشته باشد. فیلدهای ورودی متغیرها و نتیجه خوشه‌بندی با الگوریتم Two-Step به‌عنوان متغیر خروجی و هدف پیش‌بینی در نظر گرفته شد. اجرای الگوریتم‌های درخت تصمیم شامل C5.0, CHAID, C&R Tree, QUEST روی داده‌های موجود با هدف پیش‌بینی وضعیت اهدای خون اهداکنندگان صورت گرفت که تحقق این امر با بررسی شاخص‌های درجه حرارت بدن، نبض، فشارخون، هموگلوبین، وزن و سن صورت گرفت.

د) ارزیابی

برای بررسی مدل، ابتدا داده‌های تحت بررسی به دو بخش آموزش و آزمایش تقسیم گردید. داده‌های بخش آموزش (۷۰ درصد) درخت را تولید می‌نمایند و داده‌های بخش آزمایش (۳۰ درصد) درخت تولید شده را آزمایش و برچسب مربوطبه رکوردهای مذکور را تعیین می‌نمایند. شاخص‌های مختلفی مانند شفافیت (Specificity)، حساسیت (Sensitivity)، دقت

(Precision) و صحت (Accuracy) برای ارزیابی روش‌های دسته‌بندی وجود دارند. همچنین می‌توان نرخ خطا (ER: Error rate) یا دسته‌بندی نادرست را بر اساس شاخص صحت محاسبه کرد. معیار خطای دسته‌بندی یا نرخ خطا دقیقاً برعکس معیار صحت بوده و کمترین مقدار آن (صفر) زمانی است که بهترین کارایی حاصل گردد. همچنین بیشترین مقدار آن (یک) زمانی است که کمترین کارایی حاصل گردد (۲۶ و ۳۰). در مسائل واقعی، معیار صحت دسته‌بندی به هیچ عنوان معیار مناسبی برای ارزیابی کارایی الگوریتم‌های دسته‌بندی نمی‌باشد، به این دلیل که در رابطه صحت دسته‌بندی، ارزش رکوردهای دسته‌های مختلف یکسان در نظر گرفته می‌شوند. بنابراین مسائلی که با دسته‌های نامتعادل سروکار دارند و یا در مسائلی که ارزش دسته‌ای در مقایسه با دسته دیگر متفاوت است، از معیارهای دیگری استفاده می‌شود. در مسائل واقعی معیارهای دیگری نظیر نرخ مثبت درست (TPR: True positive rate) و نرخ مثبت غلط (FPR: False positive rate) اهمیت ویژه‌ای دارند (۱۰).



شکل ۱– گام‌های روش CRISP و مدل پیشنهادی

و) توسعه

وجود رویکردی سازمان‌دهی شده جهت پیش‌بینی وضعیت اهدای خون برای کمک به پزشکان جهت کاهش چالش‌های مرتبط با جمع‌آوری، نگهداری و انتقال خون ضروری است. نقطه قابل بهبود در حیطه مورد بررسی ایجاد سیستمی پزشک‌یار جهت کمک به پزشکان در ارایه خدمات بلادرنگ به اهداکنندگان و مدیریت پایگاه خون می‌باشد.

نتایج

یافته‌ها نشان می‌دهد که میانگین سن بیماران ۹/۶۳۸±۳۴/۰۴۶ سال و ۹۰ درصد آنها مرد و مابقی زن هستند. ۹۷ درصد اهداکنندگان کمتر از ۴۱ سال سن دارند. ۰/۳ درصد دارای هموگلوبین کمتر از ۱۲/۵ هستند. الگوریتم خوشه‌بندی داده‌های تحت بررسی را در ۶ خوشه قرار داد. جدول ۱ فراوانی متغیرها در خوشه‌بندی را نشان می‌دهد. برچسب تعیین شده برای هر خوشه با نظر پزشک انتقال خون به‌صورت زیر تعریف گردید: خوشه یک (اهداکنندگان دارای فشار خون بالا): اهداکنندگان این خوشه در مرحله اول یا دوم پرفشاری خون قرار دارند. خوشه دو (اهداکنندگان مرد بار اول): اهداکنندگان این خوشه مردان با وضعیت اهدای خون اولین بار هستند. خوشه سه (اهداکنندگان مرحله پیش پرفشاری خون): اهداکنندگان این خوشه در مرحله پیش پرفشاری خون قرار دارند. خوشه چهار (اهداکنندگان مرد باسابقه): وضعیت اهدای خون مردان اهداکننده در این خوشه با سابقه است. خوشه پنج (اهداکنندگان مرد مستمر): وضعیت اهدای خون مردان اهداکننده در

این خوشه مستمر است. خوشه شش (اهداکنندگان بدون فشار خون): اهداکنندگان این خوشه فشار خون طبیعی دارند. جدول ۲ مقادیر به‌دست آمده برای شاخص‌های نرخ مثبت درست، نرخ منفی درست، نرخ مثبت غلط، نرخ منفی غلط، دقت، صحت، شاخص F، نرخ خطا و میانگین هندسی را نشان می‌دهد.

جدول ۱- نتایج حاصل از خوشه‌بندی

مشخصه	خوشه ۱ اهداکنندگان دارای فشار خون بالا	خوشه ۲ اهداکنندگان مرد بار اول	خوشه ۳ اهداکنندگان مرحله پیش پرفشاری خون	خوشه ۴ اهداکنندگان مرد باسابقه	خوشه ۵ اهداکنندگان مرد مستمر	خوشه ۶ اهداکنندگان بدون فشار خون
جنسیت						
مرد	۹۱/۵۲	۱۰۰	۳۰/۹۴	۱۰۰	۱۰۰	۸۶/۳۵
زن	۸/۴۸	۰	۶۹/۰۶	۰	۰	۱۳/۶۵
سن	۳۷/۰۲۲±۱۰/۴۰۶	۳۲/۳۱۱±۹/۵۷۶	۳۵/۳۶۹±۱۰/۷۳۴	۳۴/۹۲۷±۹/۰۸۲	۳۴/۱۸۳±۹/۹۹۷	۳۰/۴۶۷±۸/۰۳۴
فشار خون کمینه	۸۳/۶۳۸±۵/۱۷	۷۸/۸۴۶±۳/۳۵۵	۷۸/۷۴۱±۳/۶۱۶	۷۸/۹۵۳±۳/۰۷	۷۹/۲۳۸±۲/۹۶	۶۷/۷۶۲±۵/۰۹۴
فشار خون بیشینه	۱۴۲/۸۳۵±۹/۰۷۷	۱۱۶/۶۳±۶/۷۷۹	۱۱۷/۰۵±۷/۶۵۷	۱۱۷/۲۲۵±۶/۱۶۸	۱۱۷/۵۵۲±۶/۶۱۳	۱۰۷/۰۰۶±۵/۳۹۴
هموگلوبین	۱۵/۰۸۵±۱/۰۱	۱۵/۰۹۹±۰/۶۷۱	۱۴/۰۹۴±۱/۱۴۱	۱۵/۱۹۹±۰/۷۶۹	۱۵/۲۱۲±۰/۷۰۸	۱۴/۸۴۴±۱/۰۰۵
نبض	۷۸/۳۳۹±۵/۹۷۷	۷۳/۸۱۳±۷/۵۴۲	۷۴/۰۰۷±۷/۱۲۳	۷۵/۶۳۴±۶/۷۴۲	۷۵/۷۰۸±۷/۰۴۴	۷۱/۱۱۱±۸/۰۵۷
درجه حرارت	۳۶/۰۵۵±۰/۲۱۴	۳۶/۰۰۸±۰/۰۶۴	۳۶/۳۵۷±۰/۴۶۶	۳۶/۰۰۱±۰/۰۷۷	۳۶/۰۱۴±۰/۰۷۸	۳۶/۱۹۶±۰/۳۷۵
وزن	۸۴/۶۷۴±۱۴/۶۴۹	۷۹/۲۲۷±۱۲/۰۰۳	۷۵/۵۴۷±۱۱/۷۹۷	۸۰/۸۲۷±۱۱/۲۲۴	۸۲/۸۳۵±۱۲/۲۳۵	۷۵/۴۱۳±۱۲/۹۶۳
وضعیت اهدای خون						
اولین بار	۲۲/۳۲	۱۰۰	۳۶/۶۹	۰	۰	۴۰/۳۲
باسابقه	۲۱/۴۳	۰	۲۰/۸۶	۱۰۰	۰	۲۱/۵۹
مستمر	۵۶/۲۵	۰	۴۲/۴۵	۰	۱۰۰	۳۸/۱
دسته‌بندی فشار خون						
پیش پرفشاری خون	۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۰
مرحله اول پرفشاری خون	۸۸/۸۴	۰	۰	۰	۰	۰
مرحله دوم پرفشاری	۱۱/۱۶	۰	۰	۰	۰	۰
فشار خون طبیعی	۰	۰	۰	۰	۰	۱۰۰

متغیرهای کمی به‌صورت "انحراف معیار± میانگین" و متغیرهای اسمی به‌صورت "درصد" گزارش شده‌اند.

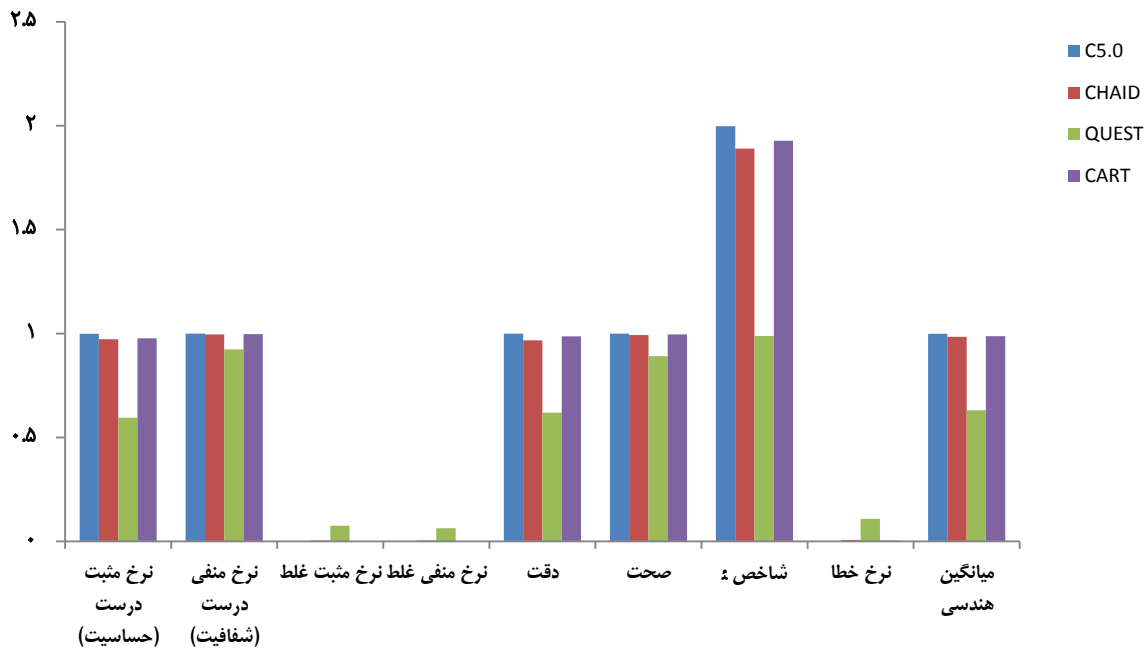
جدول ۲- مقادیر شاخص‌ها برای مدل‌های تولید شده

	درخت C5.0	آشکارساز تعامل خودکار مجذور کای	درخت آماری کارا و بی طرف سریع	درخت تقسیم و رگرسیون
نرخ مثبت درست (حساسیت)	۰/۹۹۸۸	۰/۹۷۲۸	۰/۵۹۵۱	۰/۹۷۷۲
نرخ منفی درست (شفافیت)	۱	۰/۹۹۶۰	۰/۹۲۴۴	۰/۹۹۷۶
نرخ مثبت غلط	۰	۰/۰۰۳۹	۰/۰۷۵۴	۰/۰۰۲۱
نرخ منفی غلط	۰	۰/۰۰۴۳	۰/۰۶۳۷	۰/۰۰۲۱
دقت	۱	۰/۹۶۷۴	۰/۶۱۹۲	۰/۹۸۶۶
صحت	۰/۹۹۹۸	۰/۹۹۳۰	۰/۸۹۱۳	۰/۹۹۶۰
شاخص F	۱/۹۹۷۶	۱/۸۸۹۶	۰/۹۸۸۲	۱/۹۲۷۶
نرخ خطا	۰/۰۰۰۲	۰/۰۰۷۰	۰/۱۰۸۷	۰/۰۰۴۰
میانگین هندسی	۰/۹۹۹۳	۰/۹۸۴۰	۰/۶۳۰۵	۰/۹۸۷۰

مقادیر شاخص‌های ارایه شده در جدول ۲ نشان می‌دهد که گره C5.0 بهترین مدل را تولید نموده است. شاخص‌های حساسیت، شفافیت، صحت، دقت، شاخص F و میانگین هندسی برای این مدل دارای بیشترین مقدار است. مقادیر این شاخص‌ها هرچه بیشتر باشد نشان‌دهنده این است که طبقه‌بند مورد استفاده نمونه‌های بیشتری را در جای درست خود طبقه‌بندی کرده است. شاخص‌های نرخ مثبت غلط، نرخ منفی غلط و نرخ خطا برای این مدل کمترین مقدار را دارند. مقادیر کم این شاخص‌ها تأییدکننده وقوع خطای کمتر در طبقه‌بندی نمونه‌ها است. نمودار ۱ شاخص‌ها را جهت مقایسه بهتر نشان می‌دهد. از بین الگوریتم‌های مورد استفاده بهترین نتایج مربوطبه الگوریتم C5.0 با دقت ۱ و صحت ۰/۹۹۹۸ است. جدول ۳ قوانین ایجاد شده توسط درخت تصمیم C5.0 را نشان می‌دهد.

جدول ۳- نمونه‌ای از قواعد استخراج شده از درخت تصمیم C5.0

ردیف	قوانین
۱	اگر دسته فشارخون، مرحله اول پرفشاری خون/ مرحله دوم پرفشاری خون باشد آنگاه برچسب دسته " اهداکنندگان دارای فشار خون بالا" می‌شود.
۲	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، اولین بار؛ جنسیت، مرد و دمای بدن کمتر مساوی ۳۶/۳ باشد آنگاه برچسب دسته "اهداکنندگان مرد بار اول" می‌شود.
۳	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، اولین بار؛ جنسیت، مرد؛ دمای بدن بیشتر از ۳۶/۳ و کمتر مساوی ۳۶/۷ باشد آنگاه برچسب دسته " اهداکنندگان مرد بار اول" می‌شود.
۴	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، اولین بار؛ جنسیت، مرد و دمای بدن بیشتر از ۳۶/۷ باشد آنگاه برچسب دسته " اهداکنندگان مرحله پیش پرفشاری خون" می‌شود.
۵	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، اولین بار و جنسیت، زن باشد آنگاه برچسب دسته "اهداکنندگان مرحله پیش پرفشاری خون" می‌شود.
۶	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، باسابقه؛ جنسیت، مرد و دمای بدن بیشتر از ۳۶/۶ باشد آنگاه برچسب دسته "اهداکنندگان مرحله پیش پرفشاری خون" می‌شود.
۷	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، باسابقه و جنسیت، زن باشد آنگاه برچسب دسته "اهداکنندگان مرحله پیش پرفشاری خون" می‌شود.
۸	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، مستمر؛ جنسیت، مرد و دمای بدن بیشتر از ۳۶/۶ باشد آنگاه برچسب دسته " اهداکنندگان مرحله پیش پرفشاری خون" می‌شود.
۹	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، مستمر و جنسیت، زن باشد آنگاه برچسب دسته "اهداکنندگان مرحله پیش پرفشاری خون" می‌شود.
۱۰	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، باسابقه؛ جنسیت، مرد و دمای بدن کمتر مساوی ۳۶/۶ باشد آنگاه برچسب دسته "اهداکنندگان مرد باسابقه" می‌شود.
۱۱	اگر دسته فشارخون، پیش پرفشاری خون؛ وضعیت اهدای خون، مستمر؛ جنسیت، مرد و دمای بدن کمتر مساوی ۳۶/۶ باشد آنگاه برچسب دسته "اهداکنندگان مرد مستمر" می‌شود.
۱۲	اگر دسته فشارخون، فشارخون طبیعی باشد آنگاه برچسب دسته "اهداکنندگان بدون فشار خون" می‌شود.



نمودار ۱- نمودار شاخص‌ها برای مدل‌های تولید شده

بحث

یافته‌ها نشان می‌دهد که از بین الگوریتم‌های مورد استفاده، بهترین نتایج از الگوریتم درخت C5.0 با دقت ۱ و صحت ۰/۹۹۹۸ جهت کشف الگوی وضعیت اهدای خون حاصل گردید. استفاده از قوانین ایجاد شده برای یک نمونه جدید با ویژگی‌های مشخص، می‌توان پیش‌بینی کرد که فرد در کدام دسته قرار خواهد گرفت. نتایج نشان می‌دهد که اگر اهداکننده در مرحله اول یا دوم پرفشاری خون باشد برچسب دسته "اهداکنندگان دارای فشار خون بالا" و اگر فشارخون وی طبیعی باشد برچسب دسته "اهداکنندگان بدون فشار خون" و

متناسب با کلاسی است که نمونه در آن قرار خواهد گرفت. در صورتی که اهداکننده پیش پرفشاری خون داشته باشد، وضعیت اهدای خونش اولین بار و جنسیت مرد باشد براساس دمای بدن در دو خوشه جایابی خواهد شد. به همین دلیل دمای بدن یکی از شاخص‌های تأثیرگذار مدل C5.0 محسوب می‌شود. اگر دمای بدن کمتر یا مساوی ۳۶/۷ باشد، فرد در خوشه "اهداکنندگان مرد بار اول" و اگر بیشتر از ۳۶/۷ باشد فرد در خوشه "اهداکنندگان مرحله پیش پرفشاری خون" قرار می‌گیرد. مردان با پیش پرفشاری خون و وضعیت اهدای خون باسابقه یا مستمر و دمای بدن بیشتر از ۳۶/۶ و تمامی زنان دارای پیش پرفشاری خون، در خوشه "اهداکنندگان مرحله پیش پرفشاری خون" قرار می‌گیرند.

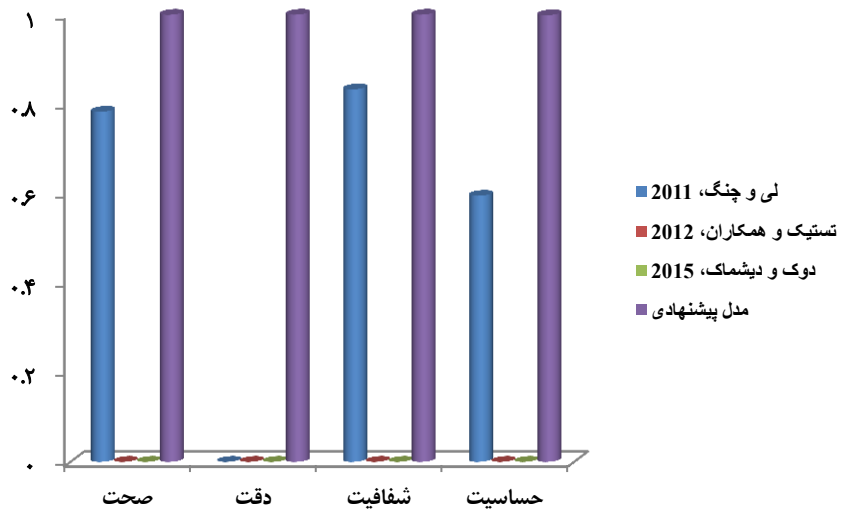
مردان دارای پیش پرفشاری خون و دمای بدن کمتر یا مساوی ۳۶/۶ براساس وضعیت اهدای خون در دو خوشه "اهداکنندگان مرد باسابقه" و "اهداکنندگان مرد مستمر" جایابی می‌شوند. به همین دلیل وضعیت اهدای خون یکی از شاخص‌های تأثیرگذار محسوب می‌شود. در صورتی که وضعیت اهدای خون باسابقه باشد برچسب دسته "اهداکنندگان مرد باسابقه" و در صورتی‌که وضعیت اهدای خون مستمر باشد برچسب دسته "اهداکنندگان مرد مستمر" می‌شود. کارهای مشابه که به‌طور همزمان از روش‌های نظارتی و غیرنظارتی استفاده نموده‌اند، اهداف مختلفی را دنبال نموده‌اند و نتایج به‌دست آمده از آنها با نتایج مدل ارایه شده در پژوهش حاضر متفاوت است.

لی و چنگ الگوریتم K-Means و الگوریتم‌های Naïve Bayes و NB tree و Decision Tree را روی شاخص‌های تازگی اهدا (Recency) (تعداد ماه‌های گذشته از آخرین اهدا)، تعداد تکرار اهدا (Frequency)، ارزش پولی اهدا (Monetary)، زمان (تعداد ماه‌ها از اولین اهدا) و متغیر اهدای خون اجرا نمودند و تازگی اهدا، تعداد تکرار اهدا و ارزش پولی اهدا به‌عنوان تأثیرگذارترین شاخص‌ها شناخته شدند (۱۵). تستیک و همکاران (۲۰۱۲) از الگوریتم Two-step و الگوریتم CART در روش انتخابی خود بهره بردند و روز هفته و ساعت در روز از میان شاخص‌های مورد بررسی شامل سال (۳–۱)، ماه (۱۲–۱)، روز ماه (۳۱–۱)، روز هفته (۷–۱) و نرخ ورود ساعتی برای دوره زمانی ۸ تا ۲۴، به‌عنوان تأثیرگذارترین شاخص‌ها انتخاب گردیدند (۳). الگوریتم‌های Naive Bayes، J48، CBA و Random Tree و الگوریتم K-Means روی شاخص‌های گروه خونی، سن، قد، جنسیت، آخرین زمان اهدای خون، تحصیلات، وضعیت تأهل اجرا شدند (۲۵). نمودار ۲ نتایج پژوهش حاضر با کارهای مشابه را مقایسه نموده است. در مدل ارایه شده همه شاخص‌های سنجش مدل محاسبه گردیده و مقادیر به‌دست آمده برای آنها بالا بودن کیفیت مدل تولید شده را تأیید می‌نماید.

سایر مطالعات انجام شده بهره گرفته از طبقه‌بندی در حوزه خون شامل موارد زیر می‌باشند. سانتهانام و ساندرام (۲۰۱۰) از الگوریتم CART برای استخراج مجموعه‌ای از قواعد جهت شناسایی رفتار اهداکنندگان خون استفاده نمودند (۱۴). ارایه سیستمی برای تحلیل دقیق داده‌های اهدا کنندگان خون با استفاده از الگوریتم طبقه‌بند J48 (۱۶)، توسعه سیستمی برای مدیریت اهدای خون آنی و آزمودن نتایج دسته‌بندی اهداکنندگان خون با الگوریتم RVD (۱۷) و استفاده از الگوریتم‌های طبقه‌بند CART-RVD و CART-DB2k7 به‌منظور شناسایی داوطلبان اهدای خون براساس الگوهای اهدای خون (۱۸) نیز از دستاوردهای سال ۲۰۱۱ محسوب می‌گردد. رموا و همکاران (۲۰۱۲) با الگوریتم‌های طبقه‌بند C4.5 و J48 به ارایه چارچوبی برای تصمیم‌گیری در پایگاه خون پرداختند (۱۹). پیش‌بینی تعداد اهداکنندگان خون در یک سن و گروه خونی خاص با طبقه‌بند J48 نیز توسط شارما و گوپتا (۲۰۱۲) صورت گرفت (۲۰). در سال ۲۰۱۳ ونکاتسوارلو و پراساد با K-Means بهبود یافته به استخراج اطلاعات اهداکنندگان در زمان محاسباتی کمتر دست یافتند (۱).

در سال ۲۰۱۴، مقایسه الگوریتم‌های داده کاوی برای انتقال پلاکت خون با استفاده از طبقه‌بندهای Naïve Bayes,Decision Table و J48 (۲۱)، ارایه سیستمی برای حل مشکلات انتقال خون با کمترین زمان توسط طبقه‌بندهای Naive Bayes، J48 و Random Tree (۲۲) و پیش‌بینی رفتار اهدا کنندگان خون با طبقه‌بندهای Ridor، ZeroR، PART، Prism، J48، Bayes net، Naïve Bayes صورت گرفت (۲۳). عاشوری و همکاران با طبقه‌بندهای C5.0، QUEST، CART و CHAID به پیش‌بینی رفتار مستمر اهداکنندگان سالم خون پرداختند (۲۴). این مطالعات فقط از روش‌های نظارتی برای مدل‌سازی بهره برده‌اند در حالی که مطالعه حاضر به‌طور همزمان از روش‌های نظارتی و غیرنظارتی بهره گرفته است. الگوریتم Two-step و الگوریتم‌های CART، C5.0، QUEST و CHAID در پژوهش حاضر روی شاخص‌های جنسیت، وضعیت اهدای خون، سن، وزن، هموگلوبین، فشار خون کمینه، فشار خون بیشینه، دسته‌بندی فشار خون، دمای بدن و نبض اجرا گردیده و شاخص‌های دسته فشارخون، وضعیت اهدای خون و دمای بدن به‌عنوان تأثیرگذارترین شاخص‌ها توسط مدل C5.0 تعیین گردیدند.

با کمک الگوی ارایه شده می‌توان به پزشکان جهت پیش‌بینی وضعیت اهداکنندگان خون (اولین بار– مستمر و با سابقه) به‌منظور ارایه خدمات درست آنی و نیز مدیریت صحیح پایگاه خون کمک نمود. نقطه قابل بهبود در حیطه مورد بررسی ایجاد سیستمی پزشک‌یار بوده که می‌تواند جهت استفاده کارآمد از خون اهدایی و کاهش هزینه‌های



نمودار ۲- مقایسه نتایج کار با کارهای انجام شده قبلی

مربوطبه نگهداری خون مفید واقع گردد. همچنین با توسعه سیستم فعلی جهت ایجاد سیستمی برای پیش‌بینی میزان پلاکت اهدایی می‌توان گامی مؤثر جهت کمک به بیماران نیازمند برداشت. ۹۰ درصد جنسیت مرد در اهدای خون و تعداد کم قوانین ایجاد شده برای اهداکنندگان زن نیاز به نمونه‌های بیشتر با جنسیت زن جهت پیش‌بینی صحیح در پژوهش‌های آتی را آشکار می‌سازد.

References

1. Venkateswarlu B, Prasad Raju GSV. Mine Blood Donors Information through Improved K-Means Clustring. International Journal of Computational Science and Information Technology 2013;1:9-15.

2. Darwiche M, Feuilloy M, Bousaleh G, Schang D. Prediction of blood transfusion donation. In Fourth International Conference on Research Challenges in Information Science 2010;51-6. doi: 10.1109/RCIS.2010.5507363

3. Testik MC, Ozkaya BY, Aksu S, Ozcebe OI. Discovering blood donor arrival patterns using data mining: a method to investigate service quality at blood centers. Journal of Medical Systems 2012; 36:579-94. doi: 10.1007/s10916-010-9519-7

4. Saiful Islam AHM, Ahmed N, Hasan K, Jubayer M. mHealth: Blood Donation Service in Bangladesh. In International Conference on Informatics, Electronics & Vision 2013;1-6. doi: 10.1109/ICIEV.2013.6572594

5. Li BN, Dong MC. Banking on blood [electronic donor card system]. Computing & Control Engineering Journal 2006;17:22-5.

6. Alfonso E, Xie X, Augusto V, Garraud O. Modeling and simulation of blood collection systems. Health Care Manag Sci 2012;15:63-78. doi: 10.1007/s10729-011-9181-8

7. Li BN, Dong MC, Chao S. On decision making support in blood bank information systems. Expert Systems with Applications 2008;34:1522-32. doi: 10.1016/j.eswa.2007.01.016

8. Shmiel O, Shmiel T, Dagan Y, Teicher M. Processing of Multichannel Recordings for Data-Mining Algorithms. IEEE Transactions on Biomedical Engineering 2007;54:444-53.

9. Altiparmak F, Ferhatosmanoglu H, Erdal S, Trost DC. Information mining over heterogeneous and high-dimensional time-series data in clinical trials databases. IEEE Trans Inf Technol Biomed 2006;10: 254-63.

10. Seliya N, Khoshgoftaar TM. The use of decision trees for cost-sensitive classification: an empirical study in software quality prediction. WIREs Data Mining and Knowledge Discovery 2011; 1: 448-59. doi: 10.1002/widm.38

11. Loh WY. Classification and regression trees. WIREs Data Mining and Knowledge Discovery 2011;1:14-23. doi: 10.1002/widm.8

12. Chen X, Wang M, Zhang H. The use of classification trees for bioinformatics. Wiley Interdiscip Rev Data Min Knowl Discov 2011;1:55-63. doi: 10.1002/widm.14

13. Kokol P, Pohorec S, ˇStiglic G, Podgorelec V. Evolutionary design of decision trees for medical application. WIREs Data Mining and Knowledge Discovery 2012;2:237-54. doi: 10.1002/widm.1056

14. Santhanam T, Sundaram S. Application of CART algorithm in blood donors classification. Journal of Computer Science 2010;6: 548-52. doi : 10.3844/jcssp.2010.548.552

15. Lee WC, Cheng BW. An intelligent system for improving performance of blood donation. Journal of Quality 2011;18: 173-85.

16. Ramachandran P, Girija N, Bhuvaneswari T. Classifying blood donors using data mining techniques. IJCSET 2011;1:10-3.

17. Sundaram S, Santhanam T. Real-time blood donor management using dashboards based on data mining models. International Journal of Computer Science 2011;8:159-63.

18. Sundaram S, Santhanam T. A comparison of blood donor classification data mining models. Journal of Theoretical and Applied Information Technology 2011;30:98-101.

19. Ramoa A, Maia S, Lourenço A. A rational framework for production decision making in blood establishments. J Integr Bioinform 2012;9:1-11. doi: 10.2390/biecoll-jib-2012-204

20. Sharma A, Gupta PC. Predicting the number of blood donors through their age and blood group by using data mining tool. International Journal of Communication and Computer Technologies 2012;1:6-10.

21. Hari Ganesh S, Vanitha K. Comparative study of data mining approaches for blood platelet transfusion. International Journal of Advanced Research in Computer Engineering & Technology 2014; 3:3069-74.

22. Asha Rani S, Hari Ganesh S. A comparative study of classification algorithm on blood transfusion. International Journal of Advancements in Research & Technology 2014;3:57-60.

23. Ritika , Paul A. Prediction of blood donors‟ population using data mining classification technique. International Journal of Advanced Research in Computer Science and Software Engineering 2014;4:634-8.

24. Ashoori M, Alizade S, Hossieny H, Hossieny S. A model to predict the sequential behavior of healthy blood donors using data mining. Journal of Research & Health 2015; 5:141-8.

25. Dhoke NW, Deshmukh SS. To improve blood donation process using data mining techniques. International Journal of Innovative Research in Computer and Communication Engineering 2015;3: 4834-40. doi: 10.15680/ijircce.2015.0305166

26. Alizadeh S, Ghazanfari M, Teimorpour B. Data mining and knowledge discovery. 2nd ed. Tehran: Publication of Iran University of Science and Technology;2011.[Persian].

27. Ameri H, Alizadeh S, Barzegari A. Knowledge extraction of diabetics' data by decision tree method. Health Management 2013;16:58-72.[Persian].

28. López MI, Luna JM, Romero C, Ventura S. Classification via clustering for predicting final marks based on student participation in forum. Proceeding of 5th International Conference on Educational Data Mining; 2012 Jun 19-21; Greece,China.p.148-51.

29. Ashoori M, Taheri Z. Using clustering methods for identifying blood donors behavior. Proceeding of 5th Iranian Conference on Electrical and Electronics Engineering 2013; Gonabad, Iran.p.4055-77.

30. Han J, Kamber M. Data Mining: Concepts and Techniques. 2nd ed. Morgan Kaufman;2006.

31. Chen G, Asterbro T. How to deal with missing categorical data: test of a simple Bayesian Method. Organizational Research Methods 2003;6:309-27.

32. Papagiannis D, Rachiotis G, Symvoulakis EK, Anyfantakis D, Douvlataniotis K, Zilidis C, et al. Blood donation knowledge and attitudes among undergraduate health science students: A cross-sectional study. Transfus Apher Sci 2016;54:303-8. doi: 10.1016/j.transci.2015.11.001

Journal of Knowledge & Health 2017; Vol 11, No 4
Shahroud University of Medical Sciences

DOI: 10.22100/jkh.v11i4.1525
Original Article

Exploring Blood Donors' Status Through Clustering: A Method to Improve the Quality of Services in Blood Transfusion Centers

Maryam Ashoori (M.Sc.)[1*], Shahriar Mohammadi (Ph.D.)[2], Hoda Sadat Hossieny Eivary (M.Sc.)[3]

1- Dept. of Information Technology Engineering, School of Technical and Engineering, Higher Educational Complex of Saravan, Saravan, Iran.
2- Dept. of Information Technology Engineering, School of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran.
3- Dept. of Computer Engineering, Azad University, Ferdos, Iran.

Received: 27 June 2016**, Accepted:** 26 December 2016

Abstract:

Introduction: *Urgent need for blood and lack of an alternative for it necessitates the presence of a model to assist doctors in providing the proper services for the donors and also the right management of blood transfusion centers. The present study is aimed at determining blood donors' status.*

Methods: *Cross-sectional survey was applied in the present study through census. The population included the data extracted from blood transfusion center of Birjand from Khordad to Shahrivar 1392 which was provided as an Excel file by the direct visit of the researcher from the blood transfusion organization. In the present study, first, two-step clustering and then C50, C&R TREE, CHAID, and QUEST algorithms were executed to obtain the best ratio among different fields. Analysis was performed using Clementine12.0 software.*

Results: *The obtained accuracy for executing C50, C&R Tree, CHAID, and QUEST equals 0.9998, 0.9960, 0.9930, and 0.8913, respectively. The results of indices including sensitivity, Specificity, accuracy, precision, FM, GM, FPR, FNR, and ER for C50 are indicators of better performance of this algorithm compared to the other ones. Important variables in modeling are blood pressure label, blood donation status and temperature.*

Conclusion: *The proposed model helps us in predicting faster and more precise status of blood donation and also the proper management of the blood transfusion centers and it canbe and effive step for efficient usage of blood donation and decreasing the blood maintenance costs.*

Keywords: Blood donors, Data mining, Decision tree, Clustering.

Conflict of Interest: No

***Corresponding author:** M. Ashoori**, Email:** mashoori@saravan.ac.ir*

Citation: Ashoori M, Mohammadi Sh, Hossieny Eivary H.S. Exploring blood donors' status through clustering: A method to improve the quality of services in blood transfusion centers. Journal of Knowledge & Health 2017;11(4):73-82.