Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

شناسایی الگوی رفتار مردم در اهدای خون با استفاده از الگوریتم K-Means مبتنی بر تازگی، بسامد و ارزش خون

جواد رجبی[1]/ سمیه علیزاده[2]/ مریم عاشوری[3]

چکیده

مقدمه: اهدای خون در کشورهای توسعه یافته ۱۸ برابر بیشتر از کشورهای در حال توسعه است. اگر تنها پنج درصد از کل جمعیت جهان به‌طور مستمر خون اهدا کنند، ذخیره کافی برای تامین نیازهای جامعه وجود خواهد داشت. این رقم در کشورهای در حال توسعه کمتر از یک درصد است. هدف مطالعه حاضر شناخت رتبه وفاداری اهداکنندگان خون به‌منظور برنامه‌ریزی مناسب برای گسترش و تقویت رفتار اهدای خون در جامعه است.

روش ها: پژوهش حاضر به روش توصیفی- مقطعی صورت گرفته و نمونه‌گیری به روش سرشماری بوده است. جامعه پژوهش متشکل از داده‌های سازمان انتقال خون تهران از آبان ۱۳۸۳ تا خرداد ۱۳۸۸ است. برای تحلیل داده‌ها از نرم افزار Clementine12.0 استفاده شده است. برای مدل‌سازی الگوریتم K-Means در دو حالت جمعیت شناختی و مبتنی بر مقادیر تازگی، بسامد و ارزش خون (Recency-Frequency- Monetary (RFM)) اجرا گردید تا ارتباط بین متغییرهای مختلف بدست آید.

یافته‌ها: مقدار میانگین شاخص ریشه میانگین مربع انحراف از معیار برای خوشه‌بندی مبتنی بر تازگی، بسامد و ارزش خون و جمعیت شناختی ۱۰/۷۲ و ۱۱/۱٤ است و خوشه‌بندی بهتر داده‌ها توسط الگوریتم مبتنی بر تازگی، بسامد و ارزش خون را تایید کرد. این الگوریتم داده‌ها را در چهار خوشه قرار داد که خوشه چهارم شامل مردان مجرد حائز بیشترین رتبه وفاداری اهدای خون و خوشه سه شامل زنان متاهل حائز کمترین رتبه وفاداری گردید.

نتیجه‌گیری: استفاده از روش‌های نوین داده‌کاوی برای تحلیل و رده بندی اهدا کنندگان خون، شیوه نگرش سازمان‌های انتقال خون را به روند اهدای خون تغییر می‌دهد. بررسی رفتار اهداکنندگان به شناسایی سریع‌تر و دقیق‌تر وفاداری اهداکنندگان و مدیریت صحیح پایگاه خون کمک می‌کند.

واژه های کلیدی: خوشه‌بندی، تازگی، بسامد، ارزش خون، وفاداری اهداکننده

• وصول مقاله: ۹٦/۰٤/۰۵ اصلاح نهایی: ۹٦/۰۹/۲۰ پذیرش نهایی: ۹٦/۱۰/۲۰

۱. کارشناس ارشد مدیریت سیستم‌های اطلاعات، دانشکده مهندسی صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

۲. دانشکده مهندسی صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

۳. دانشکده فنی و مهندسی، مجتمع آموزش عالی سراوان، سراوان، ایران؛ نویسنده مسئول (mashoori@saravan.ac.ir)

www.SID.ir

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

مقدمه

حجم داده های پزشکی که به صورت الکترونیکی ذخیره می‌شوند روز به روز در حال افزایش است. مجموعه عظیم داده‌های خام به خودی خود، کاربردی ندارند اما دانشی که در این پایگاه‌ها وجود دارد، معمولا نهان و استفاده نشده است. برای معنی بخشیدن به این داده‌ها باید آنها را تحلیل و تبدیل به اطلاعات و یا بهتر از آن یعنی دانش کرد. از این‌رو، ابزارهای داده‌کاوی می‌توانند در به‌دست آوردن این اطلاعات نهان بکار گرفته شوند. داده‌کاوی تشخیص الگو را به صورت خودکار انجام می‌دهد و تلاش می‌کند الگوهایی را در داده‌ها کشف نماید که با روش‌های آماری مرسوم به سختی قابل کشف هستند.[۱]الگوهای کشف شده می‌تواند اطلاعات ارزشمندی را در اختیار متخصصان و دست اندرکاران حوزه سلامت قرار دهد.[۲] در حال حاضر داده‌های مراکز انتقال خون توسط برنامه‌های نرم افزاری در پایگاه‌های داده ذخیره می‌شوند و از این داده‌ها جز برای گزارش‌های ساده مدیریتی استفاده چندانی نمی‌شود؛ بنابراین، می‌توان با روش‌های داده کاوی به تحلیل داده‌های پایگاه‌های خون پرداخت، زیرا روش‌های داده‌کاوی می‌تواند به‌عنوان ابزاری مهم برای استخراج دانش از داده‌های پزشکی و سلامت استفاده شود.[۳]

افزایش شمار تصادفات و مشکلات سلامت، نیاز به خون را افزایش داده است.[۴] خون عنصر حیاتی زندگی بشر است و هیچ جایگزینی برای آن وجود ندارد.[۵] سالانه حدود ۷۵ میلیون واحد خون در سراسر دنیا اهدا [۶] و میلیون‌ها واحد خون مصرف می‌گردد.[۷] اهدای خون به ازای هر ۱۰۰هزار نفر جمعیت در کشورهای توسعه یافته ۱۸ برابر بیشتر از کشورهای در حال توسعه است. اگر تنها پنج درصد از کل جمعیت به‌طور مستمر خون اهدا کنند، ذخیره کافی برای تامین نیازهای جامعه حاصل می‌گردد. در کشور ایران بین ۲/۵ درصد تا ۲/۶درصد از کل جمعیت و ۱۰۰درصد داوطلبانه خون اهدا می‌کنند.[۶] پیشرفت فناوری و افزایش قوانین مراقبت بهداشتی بر پیچیدگی سازماندهی اطلاعات بانک خون افزوده و مدیریت داده‌های آن را مشکل‌تر کرده است. پردازش داده‌ها، مدیریت اطلاعات تولید شده و ارائه آنها به روش منظم لزوم ارائه سیستم اطلاعاتی کارآمد را مطرح می‌نماید.[۸] مطالعات صورت گرفته در زمینه استفاده از علم داده‌کاوی در انتقال خون اندک است به‌طوری که پیشرفت های اخیر داده کاوی بر ارتباط رویکردهای آماری، تجسمی و تحلیل داده در زمینه‌های مختلف مهندسی تمرکز داشته است.[۹]

در زمینه بررسی نگرش های مراجعه‌کنندگان به اهدای خون، کفاش پور و همکاران با استفاده از نظریه رفتار برنامه‌ریزی شده در ایران مطالعه‌ای انجام دادند. متغییر های مستقل در این تحقیق، باورهای سه‌گانه مدل رفتار برنامه‌ریزی شده (رفتاری، هنجاری و ادراکی) بود و متغییر وابسته (هدف) تمایل افراد نسبت به اهدای داوطلبانه خون بود. در این مطالعه ۱۸۵ داوطلب اهدای خون آزموده شدند و اطلاعات آنها از طریق پرسشنامه جمع‌آوری گردید. نتایج نشان داد که از میان متغییرهای مدل، میزان کنترل رفتار ادراکی افراد تاثیر بیشتری در تمایل اهدای خون آنها دارد و اهمیت باور ادراکی در تمایل به اهدای خون تایید گردید.[۱۰] رخشانی و همکاران نشان دادند که بین آگاهی برای اهدا و متغییرهای سن و سطح تحصیلات ارتباط آماری معنا داری وجود ندارد اما بین جنس و آگاهی ارتباط آماری معنا داری هست، زیرا آگاهی زنان به اهدای خون بیش از مردان بود. در این پژوهش کمک به همنوع و صاف شدن خون انگیزه‌های اصلی اهدای خون بود. همچنین، در این مطالعه بین نگرش با سن و جنس افراد مورد مطالعه ارتباط آماری معنی داری مشاهده نشد اما نگرش با سطح تحصیلات ارتباط آماری معنی داری دارد. ارتباط اهدای خون با سن و جنس افراد معنی‌دار بود در حالی که با سطح تحصیلات، مذهب و وضع تاهل تفاوتی مشاهده نشد.[۶] در پژوهشی دیگر مدلی مفهومی برای بررسی اثر انگیزه اهداکننده خون، رضایتمندی و وفاداری وی ارائه و معناداری ارتباط بین انگیزه اهداکننده، رضایتمندی و وفاداری وی تایید گردیده است.[۱۱]

در این پژوهش با پیاده سازی الگوریتم K–Means به‌صورت جمعیت شناختی و مبتنی بر PFM از مجموعه رویکردهای توصیف در داده‌کاوی به بررسی رفتار اهداکنندگان خون پرداخته

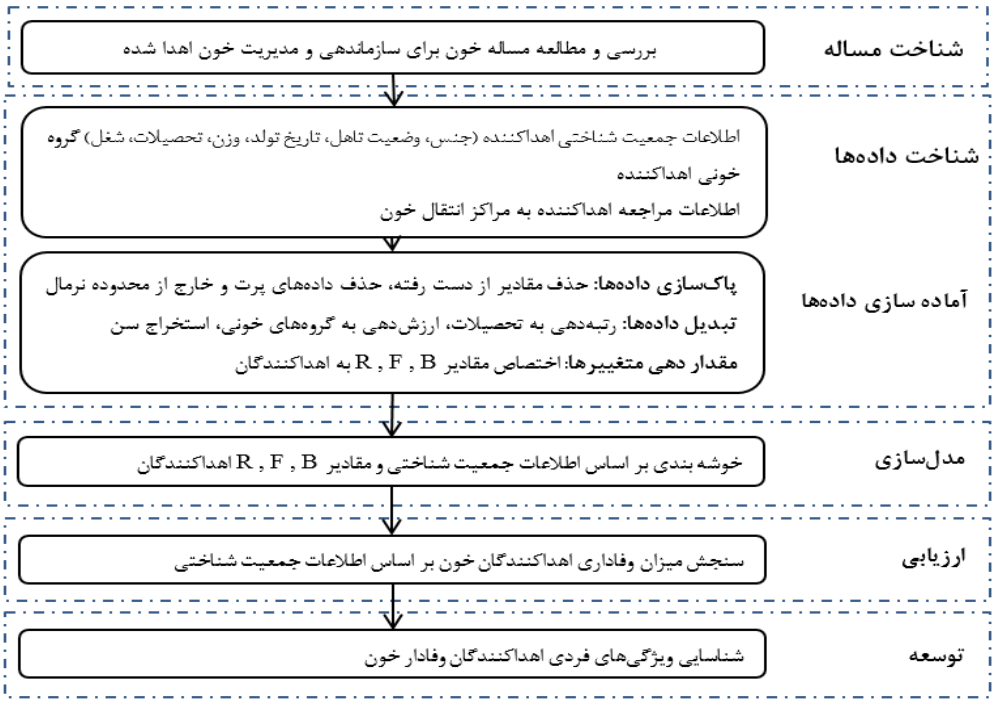
Archive of SID
www.SID.ir

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

شـد. بررسـی رفتار اهداکنندگان خون اهمیت فراوانی دارد، زیرا به شناسایی اهداکنندگان وفادار به منظور ارسال درخواست‌های بعدی اهدای خون کمک نموده و از رویگردانی آنها جلوگیری می‌نماید. رتبه بندی وفاداری اهداکنندگان خون بیمارستان‌ها و مراکز درمانی را قادر می سازد تا برنامه‌ریزی نظام مندی برای ارائه خدمات درمانی یا بهداشتی رایگان در راستای حفظ اهداکنندگان وفادار به شیوه‌ای موثر انجام دهند. هدف اصـلی این مطالعه ارائه روش و نمونه‌ای در زمینه کاربرد علم داده کاوی در انتقال خون اسـت و این که داده کاوی چگونه می‌تواند در ارائه بهتر خدمات ایفای نقش نماید.

روش‌ها

پژوهش حاضر بر اساس روش‌شناسی فرآیندهای استاندارد صنعت متقابل برای داده کاوی (CRISP:Cross Industry Process for Data Mining) انجام شـده اسـت. این روش، یکی از روش‌های بسـیار قوی برای پیاده سـازی و اجرای پروژه‌های داده کاوی و شـامل پنج مرحله اسـت [۱۲] که در شـکل شـماره یک ارائه شده است.



شکل ۱: گام‌های روش کریسپ و مدل پیشنهادی

نیاز جامعه به خون و ماهیت تهدید کننده، فاسـد شـدنی و حیاتی خون و فرآورده‌های آن، ضـرورت داشـتن الگویی برای شـناسـایی رفتار اهداکنندگان را آشکار می‌سازد. بررسی رفتار اهداکنندگان به تامین ذخیره کافی خون، کاهش چالش فقدان ذخیره احتیاطی خون در مواقع بروز بحران و شـناسـایی اهداکنندگان وفادار کمک می‌نماید. اهداکننده وفادار فردی اسـت که به‌طور پیوسـته و متناوب به سازمانی خون اهدا می‌نماید.[۱۳]

مطالعه حاضر از نوع توصیفی- مقطعی بوده و مجموعه داده‌های آن متعلق بـه مرکز انتقـال خون تهران اسـت. نمونـه‌گیری بـا روش سر شماری انجام شد و شامل ۱۳۸۴۱۳۲ نفر مراجعه کننده از آبان سـال ۱۳۸۳ تا خرداد ۱۳۸۸ بود که با مراجعه مسـتقیم پژوهش‌گر به پایگاه انتقال خون و به صـورت فایل اکسـل تهیه گردید و محتوای داده‌ها مورد تایید متخصـصـان حوزه مربوطه قرار گرفت. به منظور رعایت ملاحظات اخلاقی، نوع مطالعه و هدف آن برای کارکنان مرکز انتقال خون توضـیح و به آنها اطمینان داده شـد که اطلاعات

www.SID.ir

گردآوری شـده به صـورت محرمانه محفوظ خواهد ماند. به‌منظور آماده سازی داده‌ها، نمونه‌هایی که دارای مقادیر از دست رفته بودند یا هیچ وجه تشابهی با سایر داده‌ها نداشتند، از مطالعه حذف شدند و تعداد نمونه‌های بررسـی شـده به ۷۳۰۲۵۴ مورد رسـید. ویژگی‌های آماری اهدا کنندگان شـامل جنس، وضـعیت تاهل، گروه خونی، تحصیلات، شغل، میانگین سن و میانگین وزن آنها در جدول شماره یک نشان داده شده است.

جدول ۱: ویژگی‌های جمعیت شناختی و گروه خونی اهدا کنندگان سازمان انتقال خون از آبان ۱۳۸۳ تا خرداد ۱۳۸۸

مشخصه		تعداد	*درصد	*انحراف معیار±میانگین
جنس	زن	۱۰۵۱۹۴	۷/۶	---
	مرد	۱۲۷۸۹۳۸	۹۲/۴	---
وضعیت تاهل	مجرد	۳۱۰۰۴۶	۲۲/۴	---
	متاهل	۱۰۷۴۰۸۶	۷۷/۶	---
تحصیلات	بی‌سواد	۱۱۰۷۳	۰/۸	---
	ابتدایی	۷۸۸۹۶	۵/۷	---
	سیکل	۲۹۰۶۶۸	۲۱/۰	---
	دیپلم	۵۹۶۵۶۱	۴۳/۱	---
	فوق دیپلم	۱۰۹۳۴۶	۷/۹	---
	کارشناسی	۲۳۹۴۵۵	۱۷/۳	---
	کارشناسی ارشد	۴۲۹۰۸	۳/۱	---
	دکتری و بالاتر	۱۳۸۴۱	۱/۰	---
گروه خونی	A-	۴۷۰۶۰	۳/۴	---
	A+	۳۹۰۳۲۵	۲۸/۲	---
	AB-	۱۱۰۷۳	۰/۸	---
	AB+	۹۴۱۲۱	۶/۸	---
	B-	۳۴۶۰۳	۲/۵	---
	B+	۲۹۶۲۰۵	۲۱/۴	---
	O-	۵۶۷۴۹	۴/۱	---
	O+	۴۵۳۹۹۱	۳۲/۸	---
شغل	آزاد	۵۴۳۹۶۴	۳۹/۳	---
	خانه‌دار	۶۵۰۵۴	۴/۷	---
	اداری و دفتری	۴۷۴۷۵۷	۳۴/۳	---
	نظامی	۴۹۸۲۹	۳/۶	---
	بازنشسته	۴۴۲۹۲	۳/۲	---
	دانش آموز	۱۳۸۴۱	۱	---
	دانشجو	۷۱۹۷۵	۵/۲	---
	کارگر	۶۷۸۲۲	۴/۹	---
	سایرموارد	۵۲۵۹۸	۳/۸	---
سن		۱۳۸۴۱۳۲	---	۳۵/۳±۱۰/۵۴
وزن		۱۳۸۴۱۳۲	---	۸۰/۲±۱۳/۶۸

*متغییرهای عددی به‌صورت «انحراف معیار±میانگین» و متغییرهای اسمی به‌صورت «درصد» گزارش شده‌اند.

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

www.SID.ir

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

مدل‌سازی با استفاده از نرم افزار SPSS Clementine12.0 انجام شد. روش کار در پژوهش حاضر داده کاوی توصیف کننده است. در روش‌های توصیفی داده‌کاوی هدف یافتن الگوهایی در مورد داده‌ها ست که برای انسان قابل تفسیر باشد.[۱۴] الگوریتم K-Means با مقدار K از سه تا ۱۰ برای مدل سازی استفاده شد. الگوریتم K-Means روش خوشه بندی برای گروه‌بندی تعدادی از پیشینه‌هاست. این گروه بندی براساس شباهت مقدار ورودی برای مجموعه داده‌هاست بطوری‌که نمونه‌های هر خوشه (به تعداد k خوشه) بیشترین شباهت را به یکدیگر و بیشترین تفاوت را با اعضای خوشه‌های دیگر داشته باشند. K-means الگوریتمی تکرار شونده است، زیرا با مجموعه‌ای تصادفی از خوشه‌ها شروع می‌شود و این خوشه‌ها پیوسته بهبود می‌یابند تا جایی که بهبود بیشتر امکان‌پذیر نباشد (یا اینکه تعداد تکرارها به محدودیت مشخص شده برسد).[۱۴]

شاخص و فاداری مشتریان یکی از ابزار های رایج برای اندازه‌گیری اهمیت مشتری برای یک بنگاه کسب و کار است.[۱۵] RFM شامل مقادیری است که میزان وفاداری مشتری را به محصولات یک کسب و کار نشان می‌دهند. R نشان‌دهنده مدت زمان سپری شده از آخرین خرید مشتری، F نشان‌دهنده تعداد دفعات خرید و مراجعه مشتری برای خرید محصولات و M نشان‌دهنده پولی است که مشتری به منظور خرید کالا یا دریافت خدمات پرداخت کرده است. ارزش پولی خرید مشتری با احتساب جمع خرید مشتری در دفعات خرید می‌باشد. با وجود این سه مقدار می‌توان میزان وفاداری مشتری به کسب و کار را مشخص نمود.[۱۶] مقادیر تازگی، بسامد و ارزش پولی را می‌توان به طور مناسب و مخصوص کسب و کار تعریف نمود.[۱۷] استفاده از ابزار تازگی، بسامد و ارزش پولی و تعریف متغیرهای مربوطه در اهدای خون و خوشه‌بندی اهداکنندگان بر اساس این متغییرها می‌تواند سبب کسب نتایج مفیدی برای گروه‌بندی اهداکنندگان گردد.[۱۹، ۱۸] ارزش خونی (Blood Value(B)) بجای M استفاده گردید و میزان وفاداری اهداکنندگان با شاخص تازگی، بسامد و ارزش خون سنجیده شد. ورودی‌های تحلیل تازگی، بسامد و ارزش خون سه متغیر تازگی اهدا، بسامد اهدا و ارزش خون اهدایی است.

مسئله مهم مقداردهی این متغییرها در اهدای خون است. تازگی اهدا مدت زمان سپری شده از آخرین اهدای فرد است. در کسب و کارهای دیگر تازگی را عمدتا بر اساس روزهای سپری شده از آخرین خرید محاسبه می‌کنند اما در اهدای خون، پس از یک اهدا و قبل از مراجعه بعدی توصیه می‌شود مدت زمانی سپری شود. کارپ و کینگ این زمان را براساس جنس و ویژگی‌های جسمی بین ۵۶ تا ۱۲۰ روز اعلام کرده‌اند.[۲۰] در مطالعه حاضر عدد در نظر گرفته شده برای تازگی اهدا، میانگین این دو عدد و مقدار ۸۸ است؛ بنابراین، تازگی به عنوان تعداد دوره‌های ۸۸ روزه سپری شده از آخرین اهدا در نظر گرفته شد. بسامد با تعداد دفعات اهدای صورت گرفته، مقداردهی گردید.

متغیر M در کسب و کارهای دیگر ارزش پولی خرید مشتری نام دارد. با توجه به این‌که در اهدای خون، مشتری خریدی انجام نمی‌دهد می‌توان این متغیر را از مسئله حذف نمود و یا در حوزه انتقال خون مفهومی نزدیک به آن یافت و آن‌را کمینه کرد. فراوانی گروه‌های خونی می‌تواند مفهومی نزدیک به این متغیر به نظر برسد ولی به دلیل توزیع نسبتا یکسان گروه‌های خونی در اهداکنندگان و گیرندگان خون، استفاده از فراوانی، به عنوان متغیر M مناسب به نظر نمی‌رسد. مفهوم دیگری که بررسی شد، معیار مناسب خون‌دهی گروه‌های خونی به یکدیگر بود. از این‌رو، ارزش گذاری هر کدام از گروه‌های خونی بر اساس این که به چه تعداد گروه خونی دیگر می‌توانند خون بدهند، صورت گرفت. برای گروه خونی O- که قادر به خون‌دهی به همه گروه‌های خونیاست بالاترین ارزش و گروه خونی AB+ که تنها به یک گروه خون می‌دهد، پایین‌ترین ارزش در نظر گرفته شد. جدول شماره دو شیوه ارزش‌گذاری متغیر M که با B جایگزین شده است را نشان می‌دهد. همچنین، به متغییر اسمی میزان تحصیلات بر اساس طبقه بی‌سواد، ابتدایی، سیکل، دیپلم، فوق دیپلم،

Archive of SID
www.SID.ir

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

کارشناسی، کارشناسی ارشد، دکتری و بالاتر به ترتیب مقادیر یک تا هشت داده شد و این متغییر با رتبه دهی به یک متغییر عددی تبدیل گردید. در خوشه‌بندی جمعیت شناختی متغییرهای جنس، سن، وزن، وضعیت تاهل، میزان تحصیلات، شغل و گروه خونی و در خوشه‌بندی مبتنی بر تازگی، بسامد و ارزش خون متغییرهای جنس، سن، وزن، وضعیت تاهل، میزان تحصیلات، شغل، تازگی، بسامد و ارزش برای خوشه‌بندی استفاده شد.

جدول ۲: ارزش‌گذاری گروه‌های خونی

گروه خونی	ارزش متغیر B (M)
O-	۴
O+ , A- , B-	۳
AB- , A+ , B+	۲
AB+	۱

برای ارزیابی مدل و انتخاب عدد خوشه بهینه از شاخص دان (رابطه شماره یک) استفاده گردید. هدف شاخص دان حداکثر نمودن فاصله‌ی درون خوشه‌ای در ضمن حداقل کردن فاصله‌ی برون خوشه‌ای است[۲۱ ، ۱] تا نمونه‌های مشابه در یک خوشه قرار گیرند.

$$D = \min_{i=1\dots n_c}\left\{\min_{j=i+1\dots n_c}\left(\frac{d(c_i,c_j)}{\max_{k=1\dots n_c}(diam(c_k))}\right)\right\} \quad (۱)$$

که $d(c_i, c_j)$ و $diam(c_i)$ طبق روابط شماره دو و سه محاسبه می‌گردند.[۲۲]

$$d(c_i, c_j) = \min_{x\in c_i, y\in c_j}\{d(x,y)\} \quad (۲)$$

$$diam(c_i) = \max_{x,y\in c_i}\{d(x,y)\} \quad (۳)$$

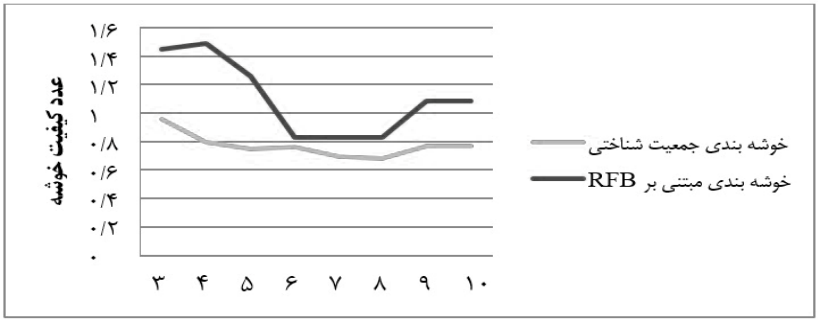
برای ارزیابی مدل‌ها و انتخاب مدل بهتر از شاخص ریشه میانگین مربع انحراف از معیار استفاده شده است (رابطه شماره چهار). این معیار دقت مدل را بر اساس تفاضل بین ارزش واقعی و ارزش برآورد شده ارزیابی می‌کند و هرچه کمتر باشد نشان‌دهنده خوشه‌بندی بهتر داده‌هاست.[۲۳ ، ۲۲]

$$RMSSDT = \sqrt{\frac{\sum_{i=1\dots n_c}\sum_{\substack{j=1\dots d}}^{n_{ij}}(x-\overline{x_j})^2}{\sum_{i=1\dots n_c}\sum_{j=1\dots d}(n_{ij}-1)}} \quad (۴)$$

به‌طور کلی وجود رویکردی سازمان دهی شده در توصیف میزان وفاداری اهداکنندگان خون برای کمک به کارکنان سازمان انتقال خون به‌منظورایجاد ذخیره مناسب خون و کاهش چالش‌های فقدان ذخیره کافی خون ضروری است. نقطه قابل بهبود در حیطه مورد بررسی ایجاد سیستمی تصمیم‌یار به‌منظور کمک به کارکنان سازمان برای ارائه خدمات مناسب و سریع به نیازمندان است.

یافته ها

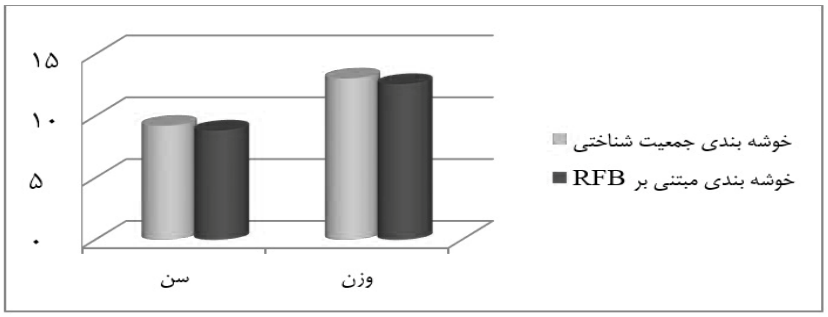
شاخص دان (نمودار شماره یک) نشان داد که عدد خوشه بهینه برای خوشه‌بندی جمعیت شناختی سه و برای خوشه‌بندی مبتنی بر تازگی، بسامد و ارزش خون، چهار است. هر چه مقدار به‌دست آمده از شاخص دان بزرگتر باشد بهتر است و تعداد خوشه‌ای که مقدار این شاخص را زیادتر نماید، عدد بهینه تعداد خوشه‌هاست.



نمودار ۱: تعیین تعداد خوشه بهینه

Archive of SID
www.SID.ir

مقادیر شـاخص ریشـه میانگین مربع انحراف از معیار برای دو متغییر عددی و مشترک در هر دو مدل (سن و وزن) در نمودار شماره دو نشان داد که خوشه‌بندی مبتنی بر تازگی، بسامد و ارزش خون بهترین مدل را تولید نموده است.



نمودار ۲:نتایج محاسبه RMSSDT

الگوریتم K-Means مبتنی بر تازگی، بسـامد و ارزش خون داده ها را در چهار خوشـه قرار داد بهطوریکه ۷۱درصـد اهداکنندگان در خوشـه یک، دو درصـد در خوشـه دو، هفت درصد در خوشه سه و ۲۰ درصد در خوشه چهار قرار گرفتند. جدول شماره سه نتایج حاصل از خوشه‌بندی مبتنی بر تازگی، بسامد و ارزش خون را نشان می‌دهد.

جدول ۳: نتایج حاصل از خوشه‌بندی مبتنی بر RFB

مشخصه		خوشه ۱ ۵۱۹۶۱۰ عضو	خوشه ۲ ۱٤۷۷۸ عضو	خوشه ۳ ۵۲۵۳۸ عضو	خوشه ٤ ۱٤۳۳۲۸ عضو
جنس	زن		۱۰۰	۱۰۰	
	مرد	۱۰۰			۱۰۰
سن		۹/۵۹ ± ۳۷/۶*	۹/۰۸ ± ۲۶/۹	۱۰/۶۵ ± ۳۹/۴	۵/۰۸ ± ۲۴/۵
وزن		۱۳/۵۷ ± ۸۰/۵	۱۱/۴۶ ± ۶۸/۶	۱۱/۹۷ ± ۷۳/۳	۱۳/۲۳ ± ۷۸/۳
وضعیت تاهل	مجرد		۱۰۰		۱۰۰
	متاهل	۱۰۰		۱۰۰	
میزان تحصیلات		۱/۳۳ ± ۴/۱	۱/۱۸ ± ۴/۸	۱/۲۵ ± ۳/۷	۱/۱۵ ± ۴/۵
شغل	آزاد	۴۱/۰۶	۸/۳	۳	۴۱/۱۳
	خانه دار	۰/۱۱	۲۲/۱۴	۷۴/۴۴	۰/۰۵
	اداری و دفتری	۴۰/۱۴	۲۸/۰۶	۱۵/۸	۲۰/۵۷
	بازنشسته	۳/۷۳	۰/۸۹	۱/۷۲	۰/۰۸
	نظامی	۴/۲۳	۰/۲۸	۰/۱	۴/۲۹
	دانش آموز	۰/۲۲	۵/۴۹	۰/۴	۳/۹۱
	دانشجو	۱/۳۵	۲۷/۰۷	۲/۷	۱۸/۱۴
	کارگر	۶/۴۵	۰/۴۸	۰/۳۳	۳/۶
	سایر موارد	۲/۷۱	۷/۲۹	۱/۵۱	۸/۲۳
تازگی		۵/۴ ± ۱۲/۸	۵/۰۴ ± ۱۳/۲	۵/۳۹ ± ۱۳/۸	۵/۱۵ ± ۱۲/۳
روز		۴۵۷ ± ۱۱۲۸	۴۴۳ ± ۱۱۶۵	۴۷۵ ± ۱۲۱۶	۴۵۳ ± ۱۰۸۴
بسامد		۱/۵۳ ± ۱/۹	۱/۲۰ ± ۱/۷	۱/۰۴ ± ۱/۵	۱/۶۱ ± ۲/۰
ارزش		۳/۹۹ ± ۴/۶	۳/۲۷ ± ۴/۰	۲/۸۶ ± ۳/۷	۴/۲۵ ± ۴/۸

*متغییرهای کمی به‌صورت «انحراف معیار±میانگین» و متغییرهای اسمی به‌صورت «درصد» گزارش شده‌اند.

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018
Archive of SID
www.SID.ir

جدول شماره چهار رتبه وفاداری خوشه‌ها را با توجه به مقادیر تازگی اهدا، بسامد اهدا و ارزش خون اهدایی و پس از بررسی رفتار اهدایی خوشه‌ها نشان می‌دهد.

جدول ٤: رتبه بندی وفاداری اهداکنندگان براساس مقادیر RFB

نام خوشه	تازگی اهدا	بسامد اهدا	ارزش خون اهدایی	رتبه وفاداری
خوشه ۱	۱۲/۸	۱/۹	۴/۶	۲
خوشه ۲	۱۳/۲	۱/۷	۴/۰	۳
خوشه ۳	۱۳/۸	۱/۵	۳/۷	۴
خوشه ٤	۱۲/۳	۲/۰	۴/۸	۱

بحث

استفاده از روش‌های نوین داده‌کاوی برای تحلیل و رده بندی اهدا کنندگان خون، نحوه‌ی نگرش سازمان‌های انتقال خون را به روند اهدای خون تغییر می‌دهد به‌طوری که تحلیل و رده بندی مناسب می‌تواند به مدیران مراکز انتقال خون در اتخاذ راهبردهای مناسب برای گسترش و تقویت رفتار اهدای خون در جامعه کمک نماید. پژوهش حاضر رویکردی برای سازماندهی اطلاعات ارائه نموده است که پردازش داده‌های سیستم با رتبه بندی وفاداری اهداکنندگاه خون انجام گرفته است. نتایج حاصل بیمارستان‌ها و مراکز درمانی را قادر می‌سازد تا برنامه‌ریزی نظام مند برای ارائه خدمات درمانی یا بهداشتی رایگان به‌منظور حفظ اهداکنندگان وفادار به شیوه‌ای موثر انجام دهند. شناخت رتبه وفاداری اهداکنندگان خون به‌منظور تامین ذخیره احتیاطی خون و جلوگیری از عوارض شدید و خطرناک فقدان ذخیره کافی، ضروری است. یافته‌ها نشان داد که خوشه‌بندی بر اساس مقادیر تازگی، بسامد و ارزش خون مدل بهتر را تولید نمود و این مدل افراد را در چهار خوشه قرار داد. برای سنجش میزان وفاداری اهداکنندگان از ابزار تازگی، بسامد و ارزش خون استفاده گردید به‌طوری که هرچه مقدار تازگی اهدا کمتر و بسامد اهدا و ارزش خون اهدایی بیشتر باشد رتبه وفاداری بهتر می شود. خوشه چهار و خوشه سه به ترتیب با مقادیر ۱۲/۳ و ۱۳/۸ بهترین و بدترین حالت برای تازگی اهدا دارند. همچنین خوشه چهار و خوشه سه با مقادیر ۲ و ۱/۵ بهترین و بدترین حالت برای بسامد اهدا دارند. این خوشه‌ها با مقادیر ۴/۸ و ۳/۷ بهترین و بدترین حالت ارزش خون اهدایی را دارند. اعضای خوشه چهار با جمعیت ۲۰ در صد، وفادارترین اهداکنندگان خون هستند و مولفه‌های میزان تحصیلات بالاتر و سن کمتر می‌توانند دلایل اصلی حمایت اعضای این خوشه از برنامه‌های سازمان انتقال خون باشد. خوشه یک با جمعیت ۷۱ درصد دومین رتبه وفاداری را دارد. با توجه به جمعیت بالای خوشه یک و چهارنسبت به سایر خوشه‌ها لازم است سازمان انتقال خون با همکاری سایر موسسات بهداشت و درمان، برنامه‌هایی جهت حمایت و حفظ اهداکنندگان وفادار تنظیم و اجرا نماید. راهکارهای پیشنهادی برای حمایت از اهداکنندگان وفادار می‌تواند شامل ارائه برخی خدمات درمانی بصورت رایگان و ارتباط مستمر و مستقیم سازمان با اهداکننده باشد.

پژوهش حاضر خوشه‌بندی جمعیت شناختی و خوشه‌بندی مبتنی بر تازگی، بسامد و ارزش خون را بررسی کرده است. محمد مصطفی نیز از شبکه‌های عصبی برای بررسی تاثیر اطلاعات جمعیت شناسی، شناختی و روان شناسی در اهدا خون در مصر بهره برده است. وی از شیوه‌های مدل‌سازی هوشمند برای پرونده اهدا کنندگان خون از غیر اهداکنندگان استفاده کرد. [۲۴] متغییرهای مورد مطالعه در خوشه‌بندی جنس، سن، وزن، وضعیت تاهل، میزان تحصیلات، شغل و گروه خونی بودند که در خوشه بندی مبتنی بر تازگی، بسامد و ارزش خون متغییر های تازگی، بسامد و ارزش به‌عنوان

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018
Archive of SID
www.SID.ir

مقادیر تازگی، بسامد و ارزش خون به مجموعه متغییرهای قبلی افزوده شدند. متغییرهای مورد استفاده در خوشه‌بندی کسب نتیجه مناسب را تضمین می‌نماید به‌طوری‌که راماچاندرن متغییرهای جنس، گروه خونی، وزن و سن را برای خوشه‌بندی داده‌های اهدای خون استفاده کردند. ایشان در مورد تاثیر استاندارد بودن مجموعه داده بر موفقیت خوشه‌بندی بحث نمودند.[۲۵] در پژوهشی دیگر علی وفایی نجار و همکاران در یک مطالعه آماری توزیع فراوانی سن، جنس، شغل و وضعیت تاهل اهداکنندگان را برر سی کردند و از این معیارها به منظور پیش‌بینی احتمال معافیت مراجعان از اهدای خون بهره بردند.[۲۶]

متغییرهای تازگی، بسامد و ارزش خون اهدایی در سنجش وفاداری اهداکنندگان مورد ا ستفاده قرار گرفته ا ست. سانتانام و ساندارام از معیارهای تازگی، دفعات اهدای خون و مقدار خون اهدایی برای دسته‌بندی اهدا کنندگان خون و تفکیک اهداکنندگان دائمی از اهداکنندگان غیر دائمی و مطالعه رفتار آنها بهره بردند.[۲۷] عاشوری و همکاران به ارائه مدلی برای پیش‌بینی رفتار مستمر اهداکنندگان سالم خون پرداختند. ایشان از الگوریتم‌های درخت تصمیم برای پیش‌بینی ا ستفاده نمودند و سپس از الگوریتم K-Means برای تایید نتایج حاصل از مدل بهتر استفاده نمودند.[۳] همچنین عاشوری و همکاران در پژوهشی دیگر به کشف الگوری وضعیت اهدای خون از طریق خوشه‌بندی پرداختند و نشان دادند که الگوریتم دسته‌بندی اهداکنندگان را متنا سب با الگوریتم خو شه‌بندی K-Means جایابی می‌نماید.[۲۸] بررسی پژوهش‌های انجام شده در حوزه داده کاوی و خون نبود تمرکز پژوهش‌گران بر استفاده از روش‌های توصیفی بصورت مستقل بر داده‌های اهدای خون را نشان می‌دهد.

با توجه به تمرکز کمتر مطالعات پیشین بر خو شه‌بندی مبتنی بر RFM و نیز تفسیرپذیری و قابل فهم بودن الگوهای کشف شده با شیوه‌های توصیف کننده، پژوهش حاضر از روش خوشه‌بندی برای مدل‌سازی استفاده کرد. از نظر هدف نیز شناسایی الگوی وفاداری اهداکنندگان خون با الگوریتم‌های داده‌کاوی کمتر مورد توجه قرار گرفته است در حالی‌که در پژوهشی دیگر مدلی مفهومی برای بررسی اثر انگیزه اهداکننده خون، رضایتمندی و وفاداری ارائه و معناداری ارتباط بین این عوامل تایید شد.[۱۱]

با توجه به نتایج پژوهش، در پایگاه‌های انتقال خون؛ وجود رویکردی سازمان دهی شده برای شناسایی رتبه وفاداری اهداکنندگان خون به‌منظور کمک به کارکنان سازمان انتقال خون برای تامین ذخیره احتیاطی خون و جلوگیری از عوارض شدید و خطرناک فقدان ذخیره منا سب خون ضروری ا ست. نقطه قابل بهبود در این حیطه تعیین دقیق میزان وفاداری اهداکنندگان به‌منظور حفظ رفاه حال مراجعان با ارائه خدمات منا سب و ایجاد سیستم‌های تصمیم‌یار برای کمک به کارکنان در تعیین گروه‌های خونی پراهداست. راهکارهای پیشنهادی برای حمایت از اهداکنندگان وفادار می‌تواند شامل ارائه برخی خدمات درمانی اهداکنندگان به‌صورت رایگان و ارتباط مستمر و مستقیم سازمان با اهداکننده با شد. پایگاه‌های خون با کمک مراکز درمانی و با اتخاذ راهبردهای مناسب می‌توانند برای حفظ اهداکنندگان وفادار و تبدیل آنها به اهداکنندگان ماهانه در صورت امکان، تلاش نماید. همچنین می‌توانند در زمینه تعداد درخواست‌ها و زمان ارسال درخواست برای اهداکنندگان وفادار تصمیم‌گیری نماید. با انتخاب راهبردهای مناسبی برای تشویق اهداکنندگان با رتبه وفاداری پایین برای تبدیل شدن به اهداکنندگان دائمی می‌توانند در را ستای تامین ذخیره مناسب خون گام‌های موثری بردارند. انتخاب راهبرد نامنا سب منجر به از د ست رفتن اهداکننده وفادار می‌گردد که این تهدیدی جدی برای حوزه بهداشت، درمان و سلامت محسوب می‌گردد، زیرا اهدا کنندگان خون به‌عنوان تامین کنندگان اولیه فرآورده های خونی برای پایگاه های خون از اهمیت ویژه‌ای برخوردارند.

فقدان صفاتی چون نبض، درجه حرارت بدن، فشار خون دیاستولیک و سیستولیک به‌منظور رسیدن به الگویی با جزئیات بیشتر از محدودیت‌های پژوهش حا ضر محسوب می‌گردد. از آنجا که نتایج این تحقیق وابسته به داده‌های یک پایگاه خون

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018
Archive of SID
www.SID.ir

است، پیشنهاد می‌شود برای بررسی بیشتر در این زمینه، در مطالعات بعدی از داده‌های سایر پایگاه‌ها و یا الگوریتم‌های دیگر استفاده گردد و بدین طریق امکان ارز یابی عملکرد پایگاه‌های خون مختلف میسر گردد.

تشکر و قدردانی

از کارکنان محترم سازمان انتقال خون تهران جهت همکاری در تهیه داده قدردانی می گردد. همچنین لازم به ذکر است که در این مطالعه تمامی ملاحظات اخلاقی از جمله شرط امانت و صداقت مورد توجه قرار گرفته است.

References

1. Jooriyan N, and Ashoori M. [Predicting the effectiveness of preeclampsia medications based on dose and method of drug consumption using data mining], Iranian Journal of Obstetrics, Gynecology and Infertility 2014; 17(123): 13-22.[In Persian]

2. Ashoori M. [A model to predict hemodialysis buffer type using data mining techniques], Journal of Health Administration 2017; 20(67): 99-110. [In Persian]

3. Ashoori M, Alizade S, Hosseiny Eivary H. S., Rastad S and Hosseiny Eivary S. S. [A model to predict the sequential behavior of healthy blood donors using data mining], Journal of Research & Health 2015; 5(2): 141-148, 2015. [In Persian]

4. Venkateswarlu B, and Prasad Raju G. Mine blood donors information through Improved K-Means clustring, International Journal of Computational Science and Information Technology 2013; 1(3): 9-15.

5. Darwiche M, Feuilloy M, Bousaleh G and Schang D. Prediction of blood transfusion donation, In Fourth International Conference on Research Challenges in Information Science; 2010.

6. Rakhshani F, Sanei Moghaddam E, Soltani Sabi M and Rakhshani T. [Knowledge, attitude and practice of the population of Zahedan province about blood donation], The Scientific Journal of Iranian Blood Transfus Organization 2010; 7(1): 9-16. [In Persian]

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

Archive of SID

www.SID.ir

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

7. Testik M, Ozkaya B, Aksu S. and Ozcebe O. Discovering blood donor arrival patterns using data mining: a method to investigate service quality at blood centers, Journal of Medical Systems 2012; 36(2): 579-94.

8. Asadi F, Moghaddasi H and Mastaneh Z. [Situation analysis of hematology information systems in educationaltherapeutic hospital laboratories of shaheed beheshti university of medical sciences], Health Information Management 2009; 6(1): 11-21. [In Persian]

9. Bellazzi R and Zupan B. Predictive data mining in clinical medicine: current issues and guidelines, International Journal of Medical Informatics 2008; 77: 81-97.

10. Kafashpoor A, Mortazavi S, and Hasani Moghadam S. [Application of social marketing concept in encouraging voluntary blood donors using theory of planned behavior], The Scientific Journal of Iranian Blood Transfus Organization 2012; 9(1): 44-53. [In Persian]

11. Žemgulienė J. Managing bood bonorship: the effect of motivation on donor satisfaction and loyality, Regional Formation and Development Studies 2016; 2(19): 153-64.

12. Mazaheri S, Ashoori M and Bechari Z. [A model to predict heart disease treatment using data mining], Journal of Payavard Salamat 2017; 11(3): 287-96. [In Persian]

13. Research into donor loyalty shows importance of personalization, engagement opportunities, Nonprofit Business Advisor 2016; 321: 1-3.

14. Han J and Kamber M. Data Mining: Concepts and Techniques. 2nd ed. USA: Elsevier Inc; 2006. p.383-464.

15. Tsai C and Chiu C. A purchase-based market segmentation methodology, Expert Systems with Applications 2004; 27(2): 265-76.

16. Chen Y, Kuo M, Wu S and Tang K. Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data, Electronic Commerce Research and Applications 2009; 8(5): 241-51.

17. Miglautsch J. Thoughts on RFM scoring, Journal of Database Marketing 2000; 8(1): 67-72.

18. Paul Harper R. A review and comparison of classification algorithms for medical decision making, Health Policy 2005; 71(3): 315-31.

19. Seyed Hosseini S, Maleki A and Gholamian M. Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, Expert Systems with Applications 2010; 37(7): 5259-64.

20. Karp J and King K. International variation in volunteer whole blood donor eligibility criteria, Transfusion 2010; 50(2): 507-13.

21. Ashoori M, NajiMoghadam V, Alizadeh S and Safi M. [Classification and clustering algorithm application for prediction of tablet numbers: case study diabetes disease], Health Information Management 2013; 10(5): 739-49. [In Persian]

22. Walther B and Moore J. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators,

Archive of SID

www.SID.ir

with a literature review of estimator performance, Ecography 2005; 28(6): 815-29.

23. Aggarwal C. Data Mining: The Textbook, New York: Springer, 415; 2015.

24. Mohamed Mostafa M. Profiling blood donors in Egypt: A neural network analysis, Expert Systems with Applications 2009; 36: 5031-8.

25 Ramachandran P, Girija N and Bhuvaneswari T. Classifying blood donors using data mining techniques, International Journal of Computer Science Engineering and Technology 2011; 1(1): 10-13.

26. Vafaee Najar A, Saeidi Nejat S, Esmaeili H, Sayadpour Zanjani D and Bazargani R. [Correlation between demographic factors and deferral rate of voluntary blood donors in Mashad], The Scientific Journal of Iranian Blood Transfus Organization 2011; 7(4): 266-71. [In Persian]

27. Santhanam T and Sundaram S. Application of CART algorithm in blood donors classification, Journal of Computer Science 2010; 6(5): 548-52.

28. Ashoori M, Mohammadi S and Hossieny Eivary H. [Exploring blood donors' status through clustering: a method to improve the quality of services in blood transfusion centers], Journal of Knowledge & Health 2016; 11(4): 73-82. [In Persian].

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

Archive of SID

www.SID.ir

Downloaded from jha.iums.ac.ir at 11:39 IRDT on Sunday May 6th 2018

Identifying Behavioral Patterns in Blood Donation Using K-Means Algorithm Based on Recency, Frequency and Blood Value

Rajabi J[1]/ Alizade S[2]/ Ashoori M[3]

Abstract

Introduction: Blood donation rate in developed countries is 18 times higher than developing countries. It is estimated that if only five percent of Iran population embark on blood donation, it will be adequate to meet the needs of the community. The aim of this paper is to identify the blood donators' loyalty behavior for proper planning to extend and enhance blood donation habits among the community.

Methods: A cross-sectional survey was applied through census in the present study. The data extracted from blood transfusion center of Tehran from 2005 to 2010 was used in this study. Clementine 12.0 was used for data analysis. K-Means clustering is based on demographic data and RFM values modes which were applied to obtain the best ratio among different fields.

Results: The mean value of root mean square standard deviation for RFB-based clustering and demographic were 10.7194 and 11.1411,respectively. This finding confirmed better data clustering by RFB-based algorithm. The data were placed by RFB-based algorithm in four clusters. The fourth cluster consisted of single males who obtained the best loyalty rank and the third cluster consisted of married females who obtained the least loyalty rank.

Conclusion: Applying data mining methods for analysis and classification of blood donors changes current attitude towards blood donation procedure. Survey of donor behavior helps us to identify faster and more precise donor loyalty as well as having proper management of the blood transfusion centers.

Keywords: Clustering, Recency, Frequency, Blood Value, Donor Loyalty

● **Received:** 26/June/2017 ● **Modified:** 11/Dec/2017 ● **Accepted:** 10/Jan/2018

1. M.Sc of Information System Management, School of Industrial Engineering, K.N.Tossi University of Technology, Tehran, Iran.
2. School of Industrial Engineering, K.N.Tossi University of Technology, Tehran, Iran.
3. School of Technical and Engineering, Higher Educational Complex of Saravan, Saravan, Iran; Corresponding author (mashoori@saravan.ac.ir)

Archive of SID
www.SID.ir