

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه قم

دانشکده فنی و مهندسی

پایان نامه دوره کارشناسی ارشد رشته مهندسی فناوری اطلاعات

عنوان:

به کارگیری قوانین انجمنی برای کشف الگوهای پنهان در داده‌های سازمان انتقال خون استان قم

استاد راهنما:

دکتر بهروز مینایی بیدگلی

استاد مشاور:

دکتر فرشید عابدی

نگارنده:

مهدیه خلیلی نژاد

زمستان / ۱۳۸۸



«صورت جلسه دفاع از پایان نامه کارشناسی ارشد»

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عجل الله تعالی فرجه الشریف)

جلسه دفاعیه پایان نامه کارشناسی ارشد آقای / خانم: رشته:

تحت عنوان:

با حضور هیأت داوران در محل دانشگاه قم در تاریخ: / / تشکیل گردید.

در این جلسه، پایان نامه با نمره (به عدد ، به حروف)

بدرجه: عالی بسیار خوب خوب قابل قبول مورد دفاع قرار گرفت.

نام و نام خانوادگی	سمت	مرتبۀ علمی	امضاء
	استاد راهنما		
	استاد مشاور		
	استاد ناظر		
	استاد ناظر		
	نماینده کمیته تحصیلات تکمیلی		

مدیر امور آموزش و تحصیلات تکمیلی
نام و امضاء

معاون آموزشی و پژوهشی دانشکده
نام و امضاء

تقدیم به:

روح پدرم برای همه بزرگواری هایش

مادرم برای همه محبت هایش

به همسرم برای همه همراهی هایش

تشکر و قدردانی:

خداوند بزرگ را سپاسگزارم که بر این بنده توفیق ارزانی داشت تا انجام این تحقیق را به پایان برسانم.

هرچند واژه‌ها را یارای آن نیست که لطف و محبت کسانی را که در دوران این تحقیق، جرعه‌نوش دریای بیکران مهر و محبتشان بوده‌ام به تصویر بکشند، به رسم ادب و احترام، بر خود واجب میدانم از زحمات و ارشادات کلیه اساتیدم در دوره کارشناسی ارشد مراتب تشکر قلبی خویش را از الطاف آنها ابراز دارم. به ویژه از استاد ارجمند جناب آقای دکتر مینایی بیدگلی که با راهنمایی‌ها و نظرات ارزنده‌شان، نقش چشمگیری در به ثمر رسیدن این تحقیق داشتند صمیمانه تقدیر و تشکر می‌نمایم. همچنین از استاد بزرگوار آقای دکتر عابدی‌فر به خاطر مساعدت‌ها و ارشادات ارزشمندشان سپاسگزارم. از استاد گرامی جناب آقای دکتر فرجامی، مدیر گروه محترم‌مان، که با نظرات شیوا و راهگشای شان بر غنای علمی این تحقیق افزودند.

بی‌تردید انجام این پژوهش بدون کمک فکری، همکاری و همدلی این اساتید غیرممکن می‌نمود و همنشینی و شاگردی آنان را که از بزرگترین افتخارات و شیرین‌ترین لحظات زندگی من می‌باشد، منت‌دار محبت‌های کارسازشان هستم.

از جناب آقای دکتر ابوالقاسمی مدیر عامل محترم سازمان انتقال خون ایران، بسیار سپاسگزارم که بی‌شک بدون همکاری و زحمات ارزنده ایشان انجام این تحقیق میسر نمی‌گردید و همچنین از خانم دکتر مقصودلو و آقای دکتر صیف که در خلال انجام تحقیق در سازمان انتقال خون بسیار مرا یاری نمودند صمیمانه تشکر می‌کنم.

از آقایان احمدناده‌علی، حمید اسلامی و علی نصیری، بهزاد لاجوردی نیز که در خلال تدوین پایان‌نامه زحمات فراوانی را متحمل شده تشکر و قدردانی نموده، موفقیت و بهروزی روزافزون‌شان را آرزومندم.

چکیده

افزایش دقت در شناخت گروه‌های مختلف اهداکنندگان خون و کمک به تشخیص بهتر رفتار آنان و تعیین بهترین گروه اهداکننده به منظور جذب آنها، از ملزومات سازمان انتقال خون می‌باشد. این تحقیق با استفاده از تکنیک قوانین انجمنی در داده‌کاوی سعی در شناسایی فاکتورهای تاثیرگذار نظیر وضعیت تاهل، سن، میزان تحصیلات، نوع شغل و سطح درآمد افراد در شناسایی افراد مستمر و تشخیص سالم‌ترین گروه افراد برای اهدا خون، را دارد. الگوریتم مورد استفاده در این تحقیق **Apriori** است. در این تحقیق، بررسی قوانین و الگوهای پنهان در ۴ نوع عامل اصلی مردودی برای اهدا خون انجام شد. این ۴ عامل شامل رفتار پرخطر، فشارخون، کم‌خونی و بیماری‌های خونی است. بررسی‌های علمی بر روی اهداکنندگان نشان داده است که سالم‌ترین و قابل اعتمادترین خون‌ها از اهداکنندگان داوطلب به ویژه اهداکنندگان مستمر دریافت می‌گردد. همچنین سالم‌ترین گروه افراد در جامعه آگاه‌ترین آنها هستند. بدین معنا که افزایش سطح تحصیلات در بین افراد جامعه و اطلاع‌رسانی به آنان در جلسات مشاوره در حین اهدا خون، جهت آگاهی در حفظ سلامت اهداکنندگان باعث رشد سطح سلامت در افراد جامعه می‌شود. زیرا طبق نتایج بدست آمده از این تحقیق کمترین میزان معافیت در اهدا در بین دو گروه دیده می‌شود، یکی اهداکنندگان دارای تحصیلات بالا و دیگری افرادی که به دفعات اقدام به اهدای خون کرده‌اند. همچنین بیشترین گروهی از افراد جامعه که در معرض فشارخون، کم‌خونی، بیماری‌های خونی و رفتارهای پرخطر قرار گرفته‌اند مشخص شده و نتایج مطلوبی از آن بدست آمد.

کلمات کلیدی:

داده‌کاوی، انتقال خون، قواعد وابستگی

فهرست مطالب

صفحه

عنوان

فصل اول: کلیات تحقیق

۲	۱-۱. مقدمه.....	
۴	۲-۱. بیان مسئله تحقیق.....	
۴	۱-۲-۱. تشریح و معرفی مساله.....	
۶	۲-۲-۱. جنبه‌های مجهول و مبهم.....	
۶	۳-۲-۱. متغیرهای مسئله.....	
۷	۴-۲-۱. هدف تحقیق.....	
۷	۳-۱. اهمیت و ضرورت انجام تحقیق.....	
۷	۱-۳-۱. اختلاف نظرها و خلاءهای تحقیقاتی.....	
۸	۲-۳-۱. ضرورت انجام تحقیق.....	
۹	۳-۳-۱. فواید عملی تحقیق.....	
۹	۴-۳-۱. جنبه جدید بودن و نوآوری در تحقیق.....	
۱۰	۴-۱. سوالات تحقیق.....	
۱۰	۱-۴-۱. سوال اصلی تحقیق.....	
۱۰	۲-۴-۱. سوال فرعی تحقیق.....	
۱۱	۵-۱. فرضیه تحقیق.....	
۱۱	۶-۱. قلمرو تحقیق.....	
۱۲	۷-۱. ساختار اجرایی کار.....	
۱۲	۸-۱. ساختار پایان‌نامه.....	
۱۲	۹-۱. نتیجه‌گیری فصل.....	

فصل دوم: ادبیات تحقیق

۱۴	۱-۲. مقدمه.....	
۱۵	۲-۲. داده‌کاوی در پزشکی.....	
۱۸	۳-۲. طب انتقال خون.....	
۲۱	۱-۳-۲. مفاهیم مربوط به انتقال خون.....	
۳۰	۴-۲. داده‌کاوی و کشف دانش از پایگاه‌داده.....	

۳۱ ۲-۴-۱. مفهوم داده‌کاوی
۳۳ ۲-۴-۲. فرایند کلی داده‌کاوی
۳۵ ۲-۴-۳. روش‌ها و تکنیک‌های داده‌کاوی
۳۷ ۲-۲-۴. قوانین انجمنی در داده‌کاوی
۵۴ ۲-۵. سابقه و پیشینه تحقیق در داده‌کاوی پزشکی و انتقال خون
۶۰ ۲-۶. نتیجه‌گیری فصل
فصل سوم	
۶۳ ۳-۱-۱. مقدمه
۶۳ ۳-۲-۲. فرایند تحقیق
۶۳ ۳-۲-۱. مدل فرایند داده‌کاوی بر اساس استاندارد CRISP-DM
۶۶ ۳-۲-۲. مدل اجرایی تحقیق
۶۸ ۳-۳-۳. داده‌های تحقیق
۶۸ ۳-۳-۱. شیوه و ابزار جمع‌آوری داده‌ها
۶۸ ۳-۳-۲. نوع داده‌ها و مقیاس آنها
۶۹ ۳-۳-۳. جامعه آماری و نمونه‌گیری
۶۹ ۳-۳-۴. اعتبار و کارایی روش ارائه شده
۷۰ ۳-۳-۵. ساختار اجرایی تحقیق
۷۸ ۳-۶. نتیجه‌گیری فصل
فصل چهارم	
۸۰ ۴-۱-۱. مقدمه
۸۰ ۴-۲-۲. توصیف داده‌ها
۸۰ ۴-۲-۱. ویژگی داده‌ها
۹۴ ۴-۲-۲. آماده‌سازی داده‌ها برای مدل
۱۰۰ ۴-۳-۳. تحلیل داده‌ها
۱۰۰ ۴-۳-۱. مدل‌سازی
۱۳۰ ۴-۳-۲. ارزیابی و اعتبار مدل
۱۳۲ ۴-۴. نتیجه‌گیری فصل
فصل پنجم	
۱۳۴ ۵-۱-۱. مقدمه

۱۳۴	۲-۵. نتایج و یافته‌های تحقیق
۱۳۴	۱-۲-۵. پاسخ به سوال تحقیق و بررسی فرضیه
۱۳۶	۲-۲-۵. نتایج مستقیم حاصل از اجرای تحقیق
۱۳۸	۳-۵. مزایای تحقیق
۱۳۸	۴-۵. پیشنهادات
۱۳۹	۱-۴-۵. پیشنهادات برای تحقیقات آتی
۱۴۰	۲-۴-۵. پیشنهادات برای سازمان انتقال خون
۱۴۱	۵-۵. محدودیت‌های تحقیق
۱۴۲	فهرست منابع و مآخذ

فهرست جداول

صفحه	عنوان
۸۱	جدول ۴-۱. چارچوب کلی پایگاه داده مورد بررسی.....
۸۲	جدول ۴-۲. عناوین فیلدها مربوط به خصوصیات اهداکننده در پایگاه انتقال خون استان قم.
۸۲	جدول ۴-۳. عناوین فیلدها مربوط به اهدا (خصوصیات خون فرد در زمان اهدا).....
۸۳	جدول ۴-۴. فیلدهای ورودی به الگوریتم پس از انجام عملیات پیش پردازش.....
۸۳	جدول ۴-۵. عناوین فیلدها مربوط به خصوصیات اهداکننده و شرایط اهدا (خون) در پایگاه داده استان هرمزگان.....
۸۵	جدول ۴-۶. دسته بندی فیلد شغل در داده های استان قم و هرمزگان.....
۸۵	جدول ۴-۷. دسته بندی فیلد تحصیلات و تبدیل به گروه تحصیلی.....
۸۶	جدول ۴-۸. دسته بندی فیلد سال.....
۸۶	جدول ۴-۹. تقسیم بندی فیلد سن به گروه سنی.....
۸۷	جدول ۴-۱۰. انواع متغیرهای اسمی در پایگاه داده استان قم و هرمزگان.....
۹۱	جدول ۴-۱۱. اطلاعات مربوط به افراد مستمر در پایگاه داده قم و هرمزگان.....
۹۲	جدول ۴-۱۲. ویژگی های خبرگان.....
۹۸	جدول ۴-۱۳. یک شکل نمودن داده ها.....
۹۹	جدول ۴-۱۴. ساخت فیلد سطح درآمد از روی فیلد گروه شغلی.....
۱۰۰	جدول ۴-۱۵. نتایج حاصل از الگوریتم Apriori بر روی داده های استان قم.....
۱۰۱	جدول ۴-۱۶. نتایج حاصل از الگوریتم Apriori بر روی داده های استان هرمزگان.....
۱۰۲	جدول ۴-۱۷. اعمال الگوریتم Apriori بر روی کل داده ها با ضریب اطمینان و پشتیبانی متفاوت استان قم.....
۱۰۲	جدول ۴-۱۸. اعمال الگوریتم Apriori بر روی کل داده ها با ضریب اطمینان و پشتیبانی های متفاوت استان هرمزگان.....
۱۰۳	جدول ۴-۱۹. گروه های مختلف اهداکننده در استان قم.....
۱۰۳	جدول ۴-۲۰. گروه های مختلف اهداکننده در استان هرمزگان.....
۱۰۴	جدول ۴-۲۱. بررسی وضعیت سلامت اهداکنندگان مستمر در استان قم.....
۱۰۴	جدول ۴-۲۲. بررسی وضعیت سلامت اهداکنندگان مستمر در استان هرمزگان.....
۱۰۵	جدول ۴-۲۳. تاثیر تحصیلات بر سلامت اهداکنندگان مستمر در استان قم.....

- جدول ۴-۲۴ تاثیر تحصیلات بر سلامت اهداکنندگان مستمر در استان هرمزگان..... ۱۰۶
- جدول ۴-۲۵ بررسی میزان سلامت گروه‌های مختلف سنی اهداکنندگان مستمر استان قم..... ۱۰۷
- جدول ۴-۲۶ بررسی میزان سلامت گروه‌های مختلف سنی اهداکنندگان مستمر استان هرمزگان..... ۱۰۸
- جدول ۴-۲۷ میزان سلامت اهداکنندگان اولین بار در دو استان قم و هرمزگان به طوری که مجرد هستند..... ۱۰۸
- جدول ۴-۲۸ میزان سلامت اهداکنندگان اولین بار جوان، در دو استان قم و هرمزگان به طوری که مجرد هستند..... ۱۰۸
- جدول ۴-۲۹ سالم‌ترین گروه از نظر شغلی و سطح درآمد، در بین گروه‌های مختلف شغلی افراد مستمر در هر دو استان قم و هرمزگان..... ۱۰۹
- جدول ۴-۳۰ دومین گروه سالم از نظر شغلی و سطح درآمد در بین گروه‌های مختلف شغلی افراد مستمر در هر دو استان قم و هرمزگان..... ۱۰۹
- جدول ۴-۳۱ میزان اهداکنندگان مستمر در دو استان قم و هرمزگان مورد مقایسه قرار می‌گیرد..... ۱۱۰
- جدول ۴-۳۲ بیشترین گروهی که تمایل به اهدا خون به صورت مستمر را دارند در دو استان نشان می‌دهند..... ۱۱۱
- جدول ۴-۳۳ تاثیر سن بر روی مستمر شدن اهداکنندگان در استان قم..... ۱۱۱
- جدول ۴-۳۴ تاثیر سن بر روی مستمر شدن اهداکنندگان در استان هرمزگان..... ۱۱۲
- جدول ۴-۳۵ تاثیر میزان تحصیلات در اهدای خون مردان به صورت مستمر در استان قم..... ۱۱۲
- جدول ۴-۳۶ تاثیر میزان تحصیلات در اهدای خون مردان به صورت مستمر در استان هرمزگان..... ۱۱۳
- جدول ۴-۳۷ تاثیر فیلد تحصیلات در مستمر شدن زنان در استان قم..... ۱۱۳
- جدول ۴-۳۸ تاثیر فیلد تحصیلات در مستمر شدن زنان در استان هرمزگان..... ۱۱۴
- جدول ۴-۳۹ تاثیر اشتغال در اهدا خون به صورت مستمر در خانم‌های استان قم..... ۱۱۴
- جدول ۴-۴۰ تاثیر اشتغال در اهدا خون به صورت مستمر در خانم‌های استان هرمزگان..... ۱۱۵
- جدول ۴-۴۱ تاثیر فیلد سن در مردان متاهل استان قم..... ۱۱۵
- جدول ۴-۴۲ تاثیر فیلد سن در مردان متاهل استان هرمزگان..... ۱۱۵
- جدول ۴-۴۳ بررسی وضعیت اهداکنندگان گروه‌های مختلف با نتیجه=" رفتار مخاطره- آمیز"..... ۱۱۷
- جدول ۴-۴۴ بررسی تاثیر ازدواج در کاهش رفتارهای مخاطره‌آمیز..... ۱۱۸
- جدول ۴-۴۵ تاثیر فیلد تحصیلات بر رفتارهای مخاطره- آمیز..... ۱۱۸

- جدول ۴-۴۶ بررسی فیلد تحصیلات با وضعیت اهداکننده..... ۱۱۹
- جدول ۴-۴۷ بررسی فیلد سن و رفتار مخاطره‌آمیز..... ۱۱۹
- جدول ۴-۴۸ تاثیر ازدواج در رفتارهای مخاطره‌آمیز در گروه پرخطر(دیپلم و زیردیپلم، اولین بار)..... ۱۲۰
- جدول ۴-۴۹ تاثیر ازدواج در رفتارهای مخاطره‌آمیز در گروه پرخطر (دیپلم و زیردیپلم، مستمر)..... ۱۲۱
- جدول ۴-۵۰ میزان ابتلا به بیماری فشارخون در زنان و مردان استان هرمزگان..... ۱۲۱
- جدول ۴-۵۱ بیشترین گروه سنی که خطر ابتلا به بیماری فشارخون را در بین زنان و مردان دارند..... ۱۲۲
- جدول ۴-۵۲ بررسی فشارخون و وضعیت تاهل در خانم-ها..... ۱۲۲
- جدول ۴-۵۳ بررسی فشارخون و وضعیت تاهل در آقایان..... ۱۲۳
- جدول ۴-۵۴ بیشترین میزان گروهی که با توجه به دو عامل سن و وضعیت تاهل دچار فشارخون شده‌اند..... ۱۲۳
- جدول ۴-۵۵ رابطه شغل و فشارخون در آقایان..... ۱۲۴
- جدول ۴-۵۶ تاثیر اشتغال و ابتلا به فشارخون در خانم‌ها..... ۱۲۴
- جدول ۴-۵۷ بررسی بیماری کم‌خونی در آقایان و خانم‌ها..... ۱۲۵
- جدول ۴-۵۸ بررسی وضعیت تاهل و کم‌خونی در خانم‌ها..... ۱۲۵
- جدول ۴-۵۹ مقایسه گروه‌های سنی مختلف در بین آقایان و خانم‌ها در ابتلا به بیماری کم-خونی..... ۱۲۶
- جدول ۴-۶۰ بیشترین افراد در بین خانم‌ها با توجه به شرایط سنی و وضعیت تاهل که در معرض ابتلا به کم‌خونی هستند..... ۱۲۶
- جدول ۴-۶۱ مقایسه گروه‌های شغلی و سطح درآمد مختلف در بین آقایان و ارتباط آن با ابتلا به بیماری کم‌خونی..... ۱۲۷
- جدول ۴-۶۲ بررسی فیلد سن و بیماری خونی..... ۱۲۸
- جدول ۴-۶۳ بیشترین مشاغلی که در سنین مختلف در آقایان با بیماری‌های خونی مواجه‌اند..... ۱۲۸
- جدول ۴-۶۴ بیشترین گروهی که در معاینه پزشک معاف موقت شناخته شدند..... ۱۲۹
- جدول ۴-۶۵ اعمال الگوریتم Apriori بر روی ۵۰٪ از رکوردهای پایگاه داده استان قم..... ۱۳۰
- جدول ۴-۶۶ اعمال الگوریتم Apriori بر روی ۵۰٪ از رکوردهای پایگاه داده استان هرمزگان..... ۱۳۰

فهرست شکل ها

صفحه	عنوان
۳۴	شکل ۱-۲. مدل داده کاوی به صورت جعبه سیاه.....
۳۹	شکل ۲-۲. مراحل کار الگوریتم AIS.....
۴۰	شکل ۳-۲. مراحل کار الگوریتم SETM.....
۴۴	شکل ۴-۲. مراحل کار الگوریتم Apriori.....
	شکل ۵-۲. مثالی از سلسله مراتب مفهومی برای کاویدن مجموعه اقلام پرتکرارهای مکرر چند
۴۹	سطحی.....
۵۰	شکل ۶-۲. یک نمونه از مجموعه داده.....
۵۱	شکل ۷-۲. درخت الگوی دنباله ای برای پایگاه داده T.....
۶۴	شکل ۱-۳. مراحل مدل فرایندی داده کاوی بر اساس استاندارد CRISP-DM.....
۶۷	شکل ۲-۳. مدل اجرای تحقیق.....
۹۴	شکل ۱-۴. تقسیم بندی بر مبنای نتیجه معاینه پزشک که روی پایگاه داده دوستان، انجام شد..
۹۵	شکل ۲-۴. تقسیم بندی پایگاه داده دو استان بر اساس وضعیت اهداکننده و نتیجه معاینه پزشک
۹۷	شکل ۳-۴. حذف برخی فیلدها.....

فهرست علائم و اختصارات (Abbreviations)

NIS	National Institute of Health
NFS	National Science Foundation
NLM	National Library of Medicine
NCI	National Cancer Institute
WHO	World Health Organization
FP	Frequent Pattern
MRI	Magnetic Resonance Imaging
fMRI	functional-Magnetic Resonance Imaging
CART	Classification And Regression Tree
MLP-NN	Multi-Layer Perception Neural Network
PNN	Probabilistic Neural Network
LDA	Linear Discriminant Analysis
BRAMS	Blood Reserve Availability Assessment, tracking and Management System
CRISP-DM	Cross Industry Standard Process for Data Mining
FP growth	Frequent Pattern growth

فصل اول

کلیات تحقیق

۱-۱. مقدمه

دنیای امروز در حال تغییر از یک پارادایم^۱ از مدل‌سازی کلاسیک و تحلیل‌های مبتنی بر مدل اساسی اولیه به توسعه مدل‌هایی که مستقیماً از خود داده‌ها است. با وجود اینکه با استفاده از کامپیوتر مقادیر زیادی از داده‌ها توسط نهادهای دولتی، بنگاه‌های تجاری بزرگ و موسسات علمی تولید می‌شود، اما درصد بسیار اندکی از آنها واقعاً مورد استفاده قرار می‌گیرد، زیرا در بسیاری از موارد، حجم داده‌ها بزرگتر از آن است که بتوان آن را مدیریت کرد و یا پیچیدگی آنها بیش از آن است که بتوان تحلیل نمود. دلیل اصلی این امر توجه صرف به انبار داده‌ها در هنگام جمع‌آوری و عدم توجه به روش تحلیل آنها است. در همین راستا دستاوردهای قابل ملاحظه در قابلیت‌های پردازشی کامپیوتری دری را برای یک میزبان نیرومند و کاربردهای نرم افزاری جدید گشوده است. در چنین شرایطی است که باید از رشد تکنولوژی برای استفاده موثر از این ثروت بالقوه سود جست و داده‌کاوی^۲ یک جواب مناسب برای استخراج این ثروت است.^۳

اصولاً تکنولوژی داده‌کاوی، پایگاه‌های داده‌ای بزرگ را بعنوان منبع بالقوه‌ای از دانش ارزشمند برای تصمیم‌گیری در نظر می‌گیرد. در فرایند داده‌کاوی بهترین نتیجه زمانی حاصل می‌شود که دانش یک فرد خبره در خصوص یک مساله، با توانایی‌های کامپیوتری ترکیب شود و فقط در اینصورت است که می‌توان بدون نفی توانایی‌های فرد خبره، سیستم‌های کامپیوتری را در خدمت نیرومندتر ساختن وی قرار داد.

^۱ Paradigm shift

^۲ Data mining

^۳ علی موسوی، "بخش بندی مشتریان خودروهای سواری شهر تهران با استفاده از داده کاوی (مورد کاوی شرکت ایران خودرو)"، پایان نامه کارشناسی ارشد، دانشگاه تربیت مدرس، (۱۳۸۶).

یکی از مهمترین رشته‌هایی که داده‌کاوی در آن مورد استفاده قرار می‌گیرد، علم پزشکی است. داده-کاوی در زمینه پزشکی دارای کاربردهای بسیار وسیع و در عین حال حساس و حیاتی است. با توجه به این که داده‌های پزشکی انسانی با ارزش‌ترین و حساس‌ترین داده‌ها برای کاوش و تحلیل هستند، تحلیل و کسب دانش از آنها می‌بایست با درجه بالایی از دقت و حساسیت صورت گیرد. از همین رو نمی‌توان بسیاری از روش‌ها را بدون اطمینان از نتایج حاصل از آنها به کار گرفت.

با این دیدگاه، سازمان انتقال خون به عنوان یکی از سازمان‌های وابسته به سرویس‌های مختلف پزشکی، نقش ویژه‌ای را در ارائه سرویس به بیماران، ایفا می‌کند. انتقال خون رویه‌ای است که می‌بایست تحت شرایط کاملاً ایمن و بی‌خطر انجام شود. در سازمان انتقال خون، اهداءکنندگان به مثابه مشتریان این سازمان محسوب می‌شوند و شناسایی مشتریان وفادار (اهداءکنندگان مستمر)^۴ برای بقای این سازمان حائز اهمیت است. به عبارت دیگر، خون اهداءکنندگان مستمر نسبت به اهداءکنندگانی که برای نخستین بار اقدام به اهدای خون می‌نمایند به خاطر غربالگری‌های مکرر بالینی و آزمایشگاهی از نظر سلامتی قابل اعتمادتر است. با افزایش اهدای مستمر بین دریافت خون از اهداءکنندگان و عرضه آن به مراکز درمانی تعادل برقرار می‌گردد و هزینه‌های مربوط به ذخیره‌سازی خون و فرآورده‌های خونی کاهش می‌یابد.^۵ لازم به ذکر است، در ایران تنها ۴۱٪ از خون افراد اهداءکننده مستمر در چرخه خون این سازمان وجود دارد (طبق آمار دفتر پژوهش‌های سازمان انتقال خون ایران) و مابقی آن مربوط به افرادی است که برای اولین بار و یا به صورت نامرتب خون خود را اهداء می‌کنند. هدف ما استخراج فاکتورهای فردی تاثیرگذار (مانند سن، میزان تحصیلات، وضعیت تاهل و...) در میزان اهداء خون از روی پایگاه داده‌ها می‌باشد.

۲-۱. بیان مسأله تحقیق

این بخش شامل تشریح مسأله و معرفی آن، بیان جنبه‌های مجهول و مبهم، بیان متغیرهای مربوطه و منظور از تحقیق می‌باشد؛ که در ادامه تشریح می‌گردد.

^۴ طبق تعریف سازمان انتقال خون ایران، اهداءکننده مستمر به فردی گفته می‌شود که فاصله بین دو اهدای خون توسط وی حداکثر ۶ ماه باشد.

^۵ Linden Vj.Bianc. c." *Blood safety and surveillance*", Marcel Dekker, Inc. , New York, (2001).

۱-۲-۱. تشریح و معرفی مساله

خون در طول تاریخ حیات بشر بخشی از زندگی، تفکر و اندیشه انسانی بوده است. با نگاهی به شواهد تاریخی در مسیر تکامل طب انتقال خون، بشر همیشه خون را مترادف با مرگ و زندگی می‌دیده است. امروزه نیز طب انتقال خون مجموعه‌ای ارزشمند را از کاربرد خون به عنوان ابزار حیات و زندگی تا توسعه روش‌های خاص برای جداسازی اجزای مختلف خون توسعه داده که این مواد با کاربردهای مشخص مورد استفاده قرار می‌گیرد.^۶

امروزه گفته می‌شود مناسب‌ترین و سالم‌ترین روش اهدای خون، اهدای داوطلبانه با انگیزه خیرخواهانه، بشر دوستانه بدون دریافت پول، پاداش، نفع شخصی، اجبار و اضطراب است. همچنین اهداکنندگان داوطلبی که بدون پاداش مادی خون خود را به طور مستمر اهدا می‌نمایند، نسبت به دیگر اهداکنندگان در معرض خطر کمتری قرار دارند زیرا خون آنان به طور مکرر تحت آزمایش قرار می‌گیرد و نسبت به سایر اهداکنندگان سالم‌تر است.^۷ با وجود افزایش آمار اهدای خون، هنوز کمبود خون احساس می‌شود. در دنیا با این که سالانه میلیون‌ها واحد خون تهیه می‌شود اما به دلیل مدت کوتاه نگهداری فرآورده‌های خونی، تنها چند روز خونگیری ناکافی، اثرات سوئی بر میزان ذخایر خون دارد.^۸ رشد معافیت اهداکنندگان از اهداء خون، نقش فزاینده‌ای در کمبود خون داشته است. در دهه‌های اخیر با استفاده از آزمایش‌های غربالگری جدید تعداد زیادی از اهداکنندگان معاف شده‌اند. هر سال تخمین زده می‌شود حدود ۱۴۰۰۰ اهداکننده در امریکا به دلیل نتایج تکرارپذیر آزمایش‌های غربالگری HIV یا HCV برای همیشه معاف می‌شوند.^۹

مهمترین هدف سازمان انتقال خون، تهیه خون و فرآورده‌های خونی سالم و کافی، و کاهش خطر عفونت‌های قابل سرایت از راه انتقال خون می‌باشد که بدین منظور سطوح مختلف ایمنی و سلامت

^۶ علی اکبر پورفتح اله، "طب انتقال خون گذشته، حال و آینده"، فصلنامه خون، دوره ۲، ۷(۱۳۸۴)، ۲۸۷-۲۸۹

^۷ Rahgozar S et al., "Education, Absorption, Selection and blood donor Protection", Tehran:IBTO, Research center publications, (2003):22-29.

^۸ Blaina D et al., "Vasovagal symptoms mediate the relationship between predonation anxiety and subsequent blood donation in female volunteers", Transfusion(2006)46:1006-1010.

^۹ Zillmer VA et al., "Distraction reduce self-reported physiological reactions to blood donation in novice donors with ablating coping style", Psychom Med(2001)63; 447-452.

مورد نیاز است و یکی از راه‌های رسیدن به سلامت خون، انتخاب اهداکننده به روش صحیح است.^{۱۰} با وجود اینکه همه خون‌های اهدایی از نظر وجود عفونت‌های قابل انتقال توسط خون و فرآورده‌های خونی ایجاد کننده بیماری‌های ایدز، هپاتیت B، هپاتیت C و سیفلیس مورد آزمایش قرار می‌گیرند، اما از آنجایی که ممکن است اهداکننده خون در زمان اهدا در فاز پنجره^{۱۱} بوده و در نتیجه پاسخ آزمایش‌های مربوطه منفی گردد، مصاحبه و اخذ شرح حال دقیق از اهداکننده در زمان خون و انتخاب اهداکنندگان از میان جمعیت‌های کم‌خطر حائز اهمیت است به طوری که یکی از استراتژی‌های سازمان جهانی بهداشت جهت افزایش سلامت خون، انتخاب اهداکننده از جمعیت‌های کم‌خطر می‌باشد. جمعیت کم‌خطر شامل افرادی است که سابقه اعتیاد تزریقی، تماس جنسی نامطمئن (و احياناً با شرکای جنسی متعدد)، خالکوبی و حجامت را ذکر نمی‌نمایند. از جمله شرایط انتخاب اهداکننده سن بین ۱۷-۶۵ سال، وزن بالای ۵۰ کیلوگرم، رعایت فاصله زمانی حداقل ۸ هفته بین دو نوبت اهدا خون، میزان هموگلوبین حداقل ۱۲/۵ mg/dl، عدم بارداری، عدم ابتلا به بیماری‌هایی مانند فشار خون و دیابت، عدم سوء سابقه اعتیاد تزریقی و عدم سابقه تماس جنسی نامطمئن قابل ذکر می‌باشد.^{[۱۲][۱۳][۱۴]}

با افزایش شمار مبتلایان به عفونت‌های قابل سرایت از طریق انتقال خون، فراخوانی اهداکنندگان و ارزیابی و بررسی دقیق آنان قبل از اهدای خون ضروری است.^{۱۵} با توجه به اینکه هدف اصلی و نهایی سازمان خون، تامین خون سالم است، این امر فقط با جذب اهداکننده مستمر، امکان پذیر خواهد بود و هدایت اهداکنندگان سالم به سمت استمرار اهدای خون یکی از اهداف اصلی سازمان می‌باشد.

اهداکنندگان مستمر در طول سال حداقل دو بار در مراکز اهدای خون تحت معاینات پزشکی قرار می‌گیرند و آزمایش‌های غربالگری روی نمونه خون آن‌ها مکرراً صورت می‌گیرد. این معاینه‌ها و آزمایش‌های دوره‌ای می‌تواند باعث احساس سلامت و به طبع آن آرامش روحی در این افراد شود.

¹⁰ Linden Vj. Bianc. c." *Blood safety and surveillance*", Marcel Dekker, Inc. , New York, (2001).

¹¹ Window period

¹² Waller, Richard H., "Technical Manual of the American Association of Blood Banks", AABB, 13th.

¹³ Hillier CD, "Silberstein and Transfusion medicine", 1st ed, Churchill living stone, New York (2003).

¹⁴ Elghetany MT et al, "Clinical Diagnosis & management by laboratory methods".

^{۱۵} گروه مولفین سازمان انتقال خون، "درسنامه اصول انتقال خون"، ایران.

از طرفی حجم بسیار بالای داده‌ها در سازمان انتقال خون نیاز به، به کارگیری روش‌های داده‌کاوی را می‌طلبد. در این پژوهش قصد بر آن است که از تکنیک‌های داده‌کاوی برای تشخیص اهداکنندگان مستمر و با پیدا کردن الگوهای مفید راه را برای شناسایی اهداکنندگان سالم و کمک به متخصصان انتقال خون هموار نماییم. تا بتوان با استفاده از نتایج حاصل از این تحقیق، به منظور استفاده از سالم-ترین خون‌ها برای اهدا به بیماران نیازمند استفاده کرد و کاهش در هزینه‌های سازمان انتقال خون، در شناسایی کم‌خطرترین افراد برای اهدای خون داشت.

۱-۲-۲. جنبه های مجهول و مبهم

- ۱- ابزار داده‌کاوی وسیله مناسبی برای شناسایی سالم‌ترین گروه از اهداکنندگان هست؟
- ۲- در صورتی که داده‌کاوی ابزار مناسبی برای پیدا کردن روابط پنهان موجود در پایگاه‌داده‌های اهداکنندگان می‌باشد، آیا با استفاده از تکنیک قوانین انجمنی می‌توان الگوهای خاصی برای گروه‌های مختلف اهداکننده پیدا کرد؟
- ۳- آیا الگوهای پنهان در پایگاه‌داده استان قم و استان هرمزگان یکسان است و اهداکنندگان در این دو استان از رفتار مشابه برخوردارند؟

۱-۲-۳. متغیرهای مسئله

متغیر مستقل: "عوامل تاثیر گذار بر رفتار اهداکنندگان" که از بررسی پایگاه‌داده‌های اهداکنندگان خون در سازمان انتقال خون، مورد مطالعه استخراج می‌گردد.

متغیر وابسته: کلاس‌های مربوط به وضعیت اهداء (مستمر، اولین بار، باسابقه) و همچنین کلاس‌های مربوط به تشخیص پزشک (سالم و معاف موقت، معاف دائم، فصدخون، انصراف)

۱-۲-۴. هدف تحقیق

اهداف کلی:

به کارگیری قوانین انجمنی برای کشف الگوهای پنهان در دادگان سازمان انتقال خون استان قم