دوفصلنامه مدیریت مهندسی و رایانش نرم/دوره ششم، شماره ۱، بهار و تابستان ۱۳۹۹، شماره پیاپی ۱۱
صفحات ۱۲۹- ۱۰۹ (DOI): 10.22091/JEMSC.2018.1278
JEMSC.QOM
مدیریت مهندسی و رایانش نرم ۱۳۹۴

# پیش‌بینی اهداء خون با استفاده از داده‌کاوی بر پایه الگوریتم‌های درخت تصمیم، KNN، SVM و MLP *

**آرش فهمی حسن** [۱]
**محمدرضا مغاری** [۲]
**امیدمهدی عبادتی** [۳]

## چکیده

*اهداء خون نقش حیاتی و حساسی در حفظ سلامت و بقاء زندگی انسان‌ها دارد. امروزه علیرغم تحولات عظیم علمی و پیشرفت‌های علوم پزشکی، هنوز تأمین کافی خون سالم یکی از چالش‌ها و دغدغه‌های مجامع پزشکی در جهان است. پیش‌بینی و برنامه‌ریزی اهداء خون به منظور حفظ و تأمین حجم خون مورد نیاز در بانک‌های خون با توجه به تنوع گروه‌های خونی و ارتباطات بین آن‌ها در طول زمان بسیار مهم و دشوار است. در این مقاله سعی شده است تا از تکنیک‌های داده‌کاوی و یادگیری ماشین به منظور پیش بینی اهداء خون استفاده شود تا بتوان در بازه‌های زمانی مختلف، حجم خون مورد نیاز بانک‌های خون را تخمین زده و تأمین نمائیم. در همین راستا از چند الگوریتم طبقه بندی در یادگیری با نظارت از جمله الگوریتم‌های درخت تصمیم، KNN، SVM و MLP برای پیش بینی استفاده شده و نتایج میزان دقت هر کدام ارائه شده است. در مجموع، عملکرد الگوریتم‌های KNN و MLP در پیش‌بینی اهداء خون از دقت بیشتری برخوردار است.*

***کلمات کلیدی:*** *داده کاوی، درخت تصمیم، شبکه عصبی مصنوعی، ماشین بردارپشتیبان، یادگیری ماشین، K- نزدیکترین همسایه*

---

* تاریخ دریافت: ۹۷/۸/۱۵؛ تاریخ پذیرش: ۹۷/۱۰/۲۸.

۱. دانشجوی کارشناسی ارشد تحقیق در عملیات، دانشکده مدیریت، دانشگاه خوارزمی، تهران، ایران.
std_fahmihassan@khu.ac.ir
۲. دانشجوی کارشناسی ارشد تحقیق در عملیات، دانشکده مدیریت، دانشگاه خوارزمی، تهران، ایران.
mr_moghari@yahoo.com
۳. استادیار گروه مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه خوارزمی، تهران، ایران. (نویسنده مسئول)
ebadati@khu.ac.ir

## مقدمه

اهدای خون به دلیل نقش حیاتی و حساسی که در امر حفظ سلامت و بقاء زندگی انسان دارد مورد توجه بوده و لازم است تا سازمان‌دهی‌هایی در این حوزه در سطح خرد و کلان کشور صورت بپذیرد. در جهان امروز علیرغم تحول عظیم علمی و با وجود پیشرفت‌های بزرگی که در علوم پزشکی رخ داده است، هنوز تأمین کافی خون سالم یکی از چالش‌ها و دغدغه‌های مجامع پزشکی جهان است. بشر تاکنون هیچ جایگزین مناسبی برای این ماده حیاتی نیافته است و لذا یکی از مهم‌ترین نیازهای مراکز درمانی در جهان برای نجات جان آسیب دیدگان، خون و فراورده‌های خونی سالم است.

یکی از جنبه‌های جالب در مورد خون این است که یک کالای معمولی نیست. زیرا خون ماهیت فاسد شدنی دارد (باهل و همکاران، ۲۰۱۷). بسیاری از اجزای سازنده خون عمر مفید کوتاهی دارند و نگهداری و عرضه مداوم آن‌ها همواره با مشکلاتی همراه است. با توجه به گفته صلیب‌سرخ آمریکا، عمر ماندگاری خون تقریباً ۴۲ روز است (دارویچ و همکاران، ۲۰۱۰). با این حال، چیزی که این مسئله را چالش برانگیزتر می‌سازد، رفتار تصادفی در تأمین خون است. خون اغلب به پلاکت‌ها، گلبول‌های قرمز خون و پلاسما تقسیم می‌شود که هر کدام نیازمندی‌های ذخیره و ماندگاری خود را دارند. به عنوان مثال، پلاکت باید در حدود ۲۲ درجه سلسیوس ذخیره شود در حالی که گلبول‌های قرمز خون ۴ درجه سلسیوس و پلاسما را در دمای منفی ۲۵ درجه سلسیوس نگهداری می‌کنند. علاوه بر این، پلاکت را می‌توان اغلب به مدت ۵ روز ذخیره کرد، گلبول‌های قرمز خون تا ۴۲ روز و پلاسما را تا یک سال می‌توان ذخیره نمود (باهل و همکاران، ۲۰۱۷).

امور مربوط به اهدای خون در کشورهای مختلف توسط سازمان‌های متفاوتی انجام می‌پذیرد، به عنوان مثال، در استرالیا توسط سرویس خون صلیب سرخ این کشور و در ایران توسط سازمان انتقال خون انجام می‌گیرد.

اهدای خون هنگامی رخ می‌دهد که یک فرد سالم به طور داوطلبانه مقدار مشخصی از خون خود را در یک مرکز انتقال خون هدیه می‌کند. انتقال خون در پزشکی جنبه حیاتی

دارد و در موارد خاصی توسط پزشک معالج تجویز می‌گردد. با توجه به وظایف حیاتی خون، کمبود و یا وقفه طولانی در خون‌رسانی هر فرد می‌تواند منجر به آسیب‌های وسیع در اجزای بدن شخص شود که در نهایت به مرگ یا معلولیت‌های غیرقابل برگشت منجر خواهد شد. از هر سه نفر مردم دنیا، یک نفر در طول زندگی احتیاج به تزریق خون و فرآورده‌های خونی پیدا می‌کند. بارزترین مثال برای موقعیت‌هایی که در آن نیاز واجب به خون پیدا می‌شود عبارت است از زمان بروز حوادث و سوانح گوناگونی نظیر تصادفات رانندگی، سوختگی‌ها و اعمال جراحی، همچنین خانم‌های باردار در حین زایمان، نوزادان و بخصوص نوزادان نارسی که به زردی دچار می‌شوند، بیماران سرطانی که تحت شیمی درمانی یا اشعه درمانی قرار دارند و مواردی دیگر از جمله نیازمندان به خون سالم می‌باشند.

در مراکز انتقال خون بیشتر کشورها، اطلاعات افرادی که برای اهداء خون به آنجا مراجعه می‌کنند بر اساس چندین مشخصه، جمع‌آوری و در یک پایگاه داده[1] ذخیره می‌شود. دسترسی به این اطلاعات آن هم در سطح کلان و به صورت یکپارچه، فواید و استفاده‌های بسیار زیادی در سطوح مختلف از جمله تشخیص، درمان و تصمیم‌گیری‌های کلان دارد. به طوری که این اطلاعات و سوابق باعث افزایش سرعت و کیفیت در ارائه خدمات درمانی به افراد و بیماران به خصوص در مواقع حساس و اورژانسی می‌شود. در سطوح مختلف تصمیم‌گیری نیز به منظور حفظ و تأمین حجم خون مورد نیاز در بانک‌های خون هر مرکز انتقال خون در هر منطقه، با در نظر گرفتن میزان تقاضای جاری بر اساس اطلاعات گذشته و در نظر گرفتن ظرفیت احتیاطی برای مواقع بحرانی و اتفاقات غیرمترقبه، این اطلاعات به منظور پیش‌بینی و تصمیم‌گیری مورد استفاده قرار می‌گیرند. یکی دیگر از مواردی هم که این حوزه را مهم‌تر و پیچیده‌تر می‌کند گروه‌های متنوع خونی و ارتباطاتی است که بین آن‌ها وجود دارد و با فرض این که برخی گروه‌های خونی کمیاب‌تر می‌باشند، برنامه‌ریزی اهداء کنندگان در طول زمان مهم‌تر و پیچیده‌تر می‌شود.



1 Data Base

با توجه به توضیحات ارائه شده، در این مقاله سعی شده است تا در سطوح تصمیم‌گیری مربوط به حوزه مذکور، از تکنیک‌های داده‌کاوی و یادگیری ماشین برای پیش‌بینی اهداء خون استفاده کنیم. در همین راستا از چند الگوریتم طبقه‌بندی در یادگیری با نظارت از جمله الگوریتم‌های درخت تصمیم، KNN، SVM و MLP که یکی از انواع شبکه‌های مصنوعی عصبی (ANN) است، استفاده شده و نتایج هر کدام ارائه شده است.

## پیشینه تحقیق

پژوهش یو و همکاران (۲۰۰۷) با استفاده از داده‌های آماری و مدل‌سازی و آنالیز به کمک الگوریتم درخت تصمیم صورت گرفته است. مدل‌سازی انجام شده و نتیجه مدل این پژوهش، قادر به شناسایی اهداءکنندگانی است که برای اولین بار خون اهداء می‌کنند و پتانسیل تبدیل شدن به یک اهداءکننده متعهد را که دوباره برای اهداء خون مراجعه می‌کنند، دارد. این اطلاعات برای توسعه استراتژی‌های حفظ اهداءکننده به عنوان هدف مفید است. یک یافته جالب این مقاله این است که مکان اهداء خون نیز یک معیار مهم در تعیین و تشخیص اهداء مجدد خون است.

مصطفی (۲۰۰۹) به منظور کشف ابعاد متنوع الگوهای رفتاری اهداءکنندگان خون و آزمایش مؤلفه‌های مختلف جمعیت‌شناسی، شناختی و روان‌شناختی اهداء خون در مصر، از الگوریتم‌های MLP و PNN[1] استفاده نموده و با یک روش آماری استاندارد (LDA[2]) مقایسه کرده است.

مطالعه تستیک و همکاران (۲۰۱۲) از تکنیک‌های داده کاوی، روش خوشه‌ای دو مرحله‌ای و روش طبقه‌بندی درخت رگرسیون برای شناسایی الگوی ورود روزانه و ساعتی در مرکز خون یک بیمارستان استفاده کرد.

پژوهش باردواج، شارما و شریواستاوا (۲۰۱۲) با کمک روش داده کاوی و استفاده از داده‌های یک بانک خون، رسیدن به اهداف این بانک خون را تسهیل می‌کند. اهداف

1 Probabilistic Neural Network
2 Linear Discriminant Analysis

سیستم بانک اهداء خون عبارت است از: افزایش نرخ اهدای خون و موارد مرتبط، استفاده مفیدتر از خون‌های اهدا شده، سیاست‌های استفاده از اهدا کنندگان و تأسیس بانک خون‌های جدید. برای مثال تکنیک تحلیل پیش‌بینی می‌تواند برای اهداءکنندگان خون در پیش‌بینی رفتار آینده آن‌ها مورد استفاده قرار گیرد. در این مقاله اهداءکننده متغیر مستقل و خون متغیر وابسته می‌شود. سپس بر اساس داده‌های تاریخی، می‌توانیم منحنی رگرسیون متناسب را که برای پیش‌بینی رفتار اهداءکننده استفاده می‌شود، رسم کنیم. این مقاله نتیجه می‌گیرد می‌توان از تکنیک‌های داده کاوی در راستای اهداف سیستم بانک اهداء خون استفاده کرد.

عاشوری و همکاران (۲۰۱۵) در پژوهش خود با استفاده از الگوریتم‌های CHAID[1]، CART[2]، C5 و QUEST[3] به پیش‌بینی رفتار آینده اهداءکنندگان خون سالم پرداخته و عملکرد الگوریتم‌های مذکور را با یک دیگر مقایسه کرده که در بین آن‌ها الگوریتم C5 از عملکرد بهتری برخوردار بوده است.

اکبن (۲۰۱۸) در پژوهش خود با به کارگیری روش‌های داده‌کاوی بر روی داده‌های حاصل از تست خون، تست ادرار و سوابق بیماری افراد، به تشخیص بیماری مزمن کلیوی در مراحل اولیه بیماری پرداخته است، بدین صورت که ابتدا با انجام خوشه بندی به روش K-Means، بر روی داده‌ها پیش‌پردازش انجام داده است و سپس با الگوریتم‌های KNN و SVM و Naïve Bayes به شناسایی بیماری مزمن کلیوی پرداخته و عملکرد الگوریتم‌ها را مقایسه نموده است.

## داده کاوی[4]

داده‌کاوی عبارت است از فرایند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده‌های بزرگ و استفاده از آن در تصمیم‌گیری در فعالیت‌های

1 Chi-Squared Automatic Interaction Detector
2 Classification and Regression Tree
3 Quick Unbiased and Efficient Statistical Tree
4 Data Mining

تجاری مهم. اصطلاح داده‌کاوی به فرایند نیمه‌خودکار تجزیه و تحلیل پایگاه داده‌های بزرگ به منظور یافتن الگوهای مفید اطلاق می‌شود (خمری و بارانی، ۱۳۹۷).

به گفته برانسی و همکاران (۲۰۱۲)، اغلب عملیات و فعالیت‌های نهادهای دولتی و خصوصی در پایگاه‌داده‌های بزرگ ثبت و جمع‌آوری می‌شوند، تکنیک داده کاوی (DM) یکی از مؤثرترین گزینه‌ها برای استخراج دانش از حجم بالای داده‌ها، کشف روابط مخفی، الگوها و ایجاد قواعد برای پیش‌بینی و ارتباط دادن داده‌ها است که می‌تواند به مؤسسات در تصمیم‌گیری سریع‌تر کمک کند یا حتی به درجه بیشتری از اعتماد برساند. داده‌کاوی به معنی جستجو برای الگوهای خاص درون مجموعه داده‌های بزرگ است که بسیاری از احتمالات برای مدیران کسب وکار و تصمیم‌گیرندگان را ایجاد می‌کند. این روزها اطلاعات و دانش، امتیازات قانونی، برای شرکت‌های سلامت و کنترل اجتماعی که در جستجوی استقلال بیشتر در اقدامات خود و کاهش زمان تصمیم‌گیری هستند، استراتژیک و ضروری محسوب می‌شوند. به همین دلیل، شرکت‌های مختلف ملی و بین‌المللی در زمینه تولید، مصرف، بازار مالی، مؤسسات آموزشی و کتابخانه‌ها پیش از این در امور عادی خود، داده‌کاوی را برای نظارت بر بودجه، مصرف مشتری، جلوگیری و کشف تقلب و پیش‌بینی ریسک‌های بازار در میان دیگران، به کار گرفته‌اند (خامیس، چریوییت و کیمانی، ۲۰۱۴).

در بخش بهداشت عمدتاً بخش عمومی، کاربرد آن به عنوان روشی برای تسریع جستجوی دانش پذیرفته شده‌است. علاوه بر این، استفاده از داده‌کاوی در پایگاه‌های داده بیمارستان‌های بزرگ یا حتی در سیستم‌های اطلاعاتی سلامت عمومی به کشف روابط کمک می‌کند تا آن‌ها بتوانند بر مبنای گذشته یک پیش‌بینی از آینده داشته باشند، تا بتوانند به بهترین شکل برای کمک، تشخیص و درمان‌های پزشکی موفق بیماری‌های مختلف را شناسایی کرده و الگوهای جراحات جدید را نشان دهند (کاردوسو و مارچادو، ۲۰۰۸).

## روش‌شناسی پژوهش

این پژوهش از نوع توصیفی با رویکرد کاربردی می‌باشد. طبقه‌بندی یکی از تکنیک‌های یادگیری ماشین است که به منظور پیش‌بینی کلاس داده‌ها به کار می‌رود. پیش‌بینی یعنی، آنچه که انتظار می‌رود در آینده بر اساس دانش و تجربه اتفاق بیفتد، اما نه همیشه (علم آموز و ندیمی، ۲۰۱۲)، به عبارت دیگر طبقه‌بندی به پیاده‌سازی ساختار شناخته شده بر داده‌های آزمایشی می‌پردازد (بالاکریشنان، ۲۰۱۰). در پژوهش حاضر، از روش‌های طبقه‌بندی پیشرفته‌ی الگوریتم‌های درخت تصمیم، KNN، SVM و MLP که یکی از انواع شبکه‌های مصنوعی عصبی (ANN) می‌باشد، استفاده شده است.

در این پژوهش از زبان برنامه‌نویسی پایتون[1] استفاده شده است که یک زبان برنامه‌نویسی پویا و همه‌منظوره است و در طیف وسیعی از برنامه‌های نرم‌افزاری از جمله در توسعه برنامه‌های تحت وب و برنامه‌های با قابلیت واسط گرافیکی کاربر (GUI) قابل استفاده می‌باشد. علاوه بر این، پایتون یکی از ابزارهای اصلی برای توسعه پلتفرم‌های در مقیاس داده‌های بزرگ[2] می‌باشد.

الگوریتم‌های فوق الذکر را با استفاده از زبان پایتون بر روی مجموعه داده‌های مرکز انتقال خون که شامل ۷۴۸ نمونه و ویژگی‌های تازگی، تناوب، حجم خون اهدائی، زمان اولین مراجعه افراد و یک متغیر صفر و یک می‌باشد، پیاده‌سازی و نتایج هر کدام ارائه شده است.

## یافته‌های پژوهش

### مجموعه داده[3]

مجموعه داده مورد استفاده در این پژوهش از سایت UCI گرفته شده و مربوط به اطلاعات پایگاه‌داده یک مرکز خدمات انتقال خون در شهر Hsin-Chu در تایوان می‌باشد. مجموعه داده شامل اطلاعات ۷۴۸ اهدا کننده خون می‌باشد و شامل پنج مشخصه

1 Python
2 Big Data
3 Data Set

R(تازگی- آخرین زمان اهداء خون)، F (تناوب- تعداد دفعات اهداء خون)، M (حجم خون اهداء شده) و T (زمان- زمان اولین مراجعه فرد برای اهداء خون) و یک متغییر باینری که نشان‌دهنده آن است که هر فرد در مارس ۲۰۰۷، خون اهداء کرده یا خیر (اهداء نموده برابر مقدار یک و در صورت عدم اهداء خون، برابر صفر).

مشخصه‌های انتخاب شده براساس مدل RFM بوده که به تحلیل رفتار و بیان تفاوت مشتریان (که در اینجا اهداءکنندگان) با استفاده از سه متغیر تازگی، تکرار و مبلغ خرید (در اینجا حجم خون اهدائی) می‌پردازد که توسط هاگز (۱۹۹۴) ارائه شده است. بر طبق نظر ریناتز و کومار (۲۰۰۰)، چانگ و تسای (۲۰۰۴) مدل RFM نمی‌تواند مشتریان دارای ارتباط بلندمدت و مشتریان دارای ارتباط کوتاه‌مدت با سازمان را مشخص نماید. آن‌ها در تحقیق خود ایده طول ارتباط مشتری را پیشنهاد می‌دهند و به بررسی تأثیر آن بر وفاداری و سودآوری مشتری می‌پردازند. آن‌ها بیان می‌کنند که افزایش طول ارتباط با مشتری، وفاداری مشتری را بهبود خواهد بخشید و این متغیر را که نشان دهنده فاصله زمانی بین اولین و آخرین مراجعه مشتری در بازه مورد مشاهده است تعریف کرده‌اند. بنابراین بُعد طول ارتباط مشتری (L) به مدل RFM اضافه می‌شود که در اغلب متون RFML و در برخی از متون از آن به عنوان RFMT یاد می‌کنند. این مدل RFML یا RFMT روشی است که برای خوشه‌بندی مشتریان در مدیریت ارتباط با مشتری (CRM) استفاده می‌شود.

در این پژوهش با به کارگیری تعدادی از الگوریتم‌های یادگیری ماشین در حوزه یادگیری با ناظر، و پیاده‌سازی آن‌ها بر روی اطلاعات به دست آمده با استفاده از مدل RFML از مرکز انتقال خون، عمل اهداء خون را در افراد، بر اساس اطلاعات در هر مشخصه پیش‌بینی کنیم و دقت پیش‌بینی هر کدام را متناسب با این مجموعه داده مشخص نمائیم. الگوریتم‌های مورد استفاده در این پژوهش شامل الگوریتم درخت تصمیم، نزدیک‌ترین همسایه (KNN)، ماشین بردار پشتیبان (SVM) و همچنین پرسپترون چندلایه (MLP) است که از الگوریتم‌های پیشخور شبکه عصبی مصنوعی می‌باشد.

## ارزیابی الگوریتم‌ها

در بحث ارزیابی دقت و کارایی الگوریتم‌ها در پیش‌بینی، از جمله شـــاخص‌هایی که مورد بررسـی قرار گرفته‌اند شـامل دقت[1]، صـحت[2]، بازخوانی[3]و امتیاز F1[4] می‌باشـند که فرمول محاسبه هر کدام به شرح زیر است:

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$ رابطه ۱)

$$\text{Precision} = \frac{tp}{tp + fp}$$ رابطه ۲)

$$\text{Recall} = \frac{tp}{tp + fn}$$ رابطه ۳)

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$ رابطه ۴)

به طوری که

tp : تعداد نمونه‌های عضو کلاس و درست تشخیص داده شده

fp : تعداد نمونه‌های عضو کلاس و اشتباه تشخیص داده شده

tn : تعداد نمونه‌های غیرعضو کلاس و درست تشخیص داده شده

fn : تعداد نمونه‌های غیرعضو کلاس و اشتباه تشخیص داده شده

که در مورد مجموعه داده اهداء خون منظور از نمونه‌های عضو کلاس، افرادی هستند که خون اهداء می‌کنند.

الگوریتم‌های مذکور با زبان برنامه‌نویسی پایتون، در نرم‌افزار spyder و با مشـخصـات پردازنـده و ورژن زبـان برنـامـه‌ریزی (Python 3.6.4 – MSC v.1900 – 64 bit – 64 AMD)، پیاده‌ســـازی شـــده و در ادامه نحوه اجرای هر کدام از الگوریتم‌ها توضـــیح داده می‌شود.



1 Accuracy
2 Precision
3 Recall
4 F1-score

## الگوریتم درخت تصمیم (ID3)[1]

درخت تصمیم در داده‌کاوی مدلی است که جهت نمایش طبقه‌بندی‌ها و رگرسیون‌ها استفاده می‌شود. همان‌طور که از نام آن مشخص است، این درخت از تعدادی گره و شاخه تشکیل شده است. در درخت تصمیمی که عمل طبقه‌بندی را انجام می‌دهد، برگ‌ها بیانگر کلاس‌ها هستند. در هر یک از گره‌های دیگر (گره‌های غیربرگ) با توجه به یک یا چند صفت خاص تصمیم‌گیری صورت می‌گیرد. درخت تصمیم به دلیل سادگی و قابل فهم بودن تکنیک محبوبی در داده‌کاوی محسوب می‌شود. به عبارت دیگر درخت تصمیم خود به تنهایی همه مطالب را توصیف می‌کند و نیاز به فرد خبره‌ای نیست تا خروجی را تفسیر کند. در واقع، این یک روش گرافیکی است و بدین دلیل تفسیر آن شاید ساده‌تر از تکنیک‌های دیگر طبقه‌بندی باشد. اما به خاطر داشته باشید که داشتن تعداد گره‌های زیاد در درخت می‌تواند نمایش گرافیکی درخت تصمیم را با مشکل روبرو سازد (نوروزی طیولا، موسوی و کاظمی، ۱۳۹۶).

درخت‌های تصمیم‌گیری در بین رویکردهای یادگیری ماشین، به عنوان روش کارا و اثربخش شناخته شده‌اند و آن‌ها با موفقیت برای حل مشکلات دنیای واقعی در حوزه هوش مصنوعی به کار گرفته شده‌اند. این موفقیت به دلیل توانایی عالی آن‌ها برای حل مشکلات پیچیده از طریق نمایش‌های گرافیکی قابل خواندن توسط انسان و توسط کامپیوتر است (کوپینلان، ۱۹۸۶، ۱۹۹۳؛ بریمان و همکاران، ۱۹۸۴؛ ترابلسی، الویدی و لفوره، ۲۰۱۸).

در این الگوریتم پس از فراخوانی داده‌ها در محیط برنامه، داده‌ها را به دو بخش داده‌های آموزشی و داده‌های تست تقسیم نموده که معمولاً ۷۰٪ درصد داده‌ها را به عنوان داده‌های آموزشی و ۳۰٪ آن را به عنوان داده‌های تست در نظر می‌گیرند. در اینجا ما داده‌های تست را در چهار اندازه مختلف به کار گرفتیم و دقت پیش‌بینی را با توجه به هر کدام بدست آوردیم که در جدول (۱) قابل مشاهده است. در ادامه با فراخوانی کتابخانه scikit-learn، ساب پکیج sklearn.tree و طبقه‌بندی کننده DecisionTreeClassifier،

1 Decision tree

مدل درخت تصمیم را ساخته و داده‌های آموزشی (x_train, y_train) را وارد مدل کرده تا مدل آموزش ببیند. در ادامه برای مشخص نمودن دقت مدل، داده‌های تست (x_test) را وارد مدل کرده تا پیش‌بینی کند و در مقایسه با برچسب‌های دادهای تست (y_test) دقت پیش‌بینی را ارزیابی نماید.

این فرایند را بار دیگر انجام داده اما این بار به جای استفاده از داده‌های اصلی، برای هم مقیاس شدن داده‌ها، آن‌ها را نرمال‌سازی کرده و سپس داده‌ها را به دو دسته داده‌های آموزشی و تست تقسیم کرده و دقت مدل را برای هر یک از مقادیر داده‌های تست، ارزیابی کردیم. نتایج ارزیابی مدل دسته‌بندی کننده درخت تصمیم در جدول (۱) قابل مشاهده است.

**جدول ۱. نتایج ارزیابی مدل درخت تصمیم**

| | Data type | Test size | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|---|---|
| Decision Tree | Original | ۰/۳ | ۰/۷۵ | ۰/۷۷ | ۰/۷۶ | ۰/۷۶۸۸ |
| | | ۰/۲۵ | ۰/۷۲ | ۰/۷۵ | ۰/۷۳ | ۰/۷۵۴۰ |
| | | ۰/۲ | ۰/۷۵ | ۰/۷۸ | ۰/۷۶ | ۰/۷۸ |
| | | ۰/۱۵ | ۰/۸۱ | ۰/۸۱ | ۰/۸۱ | ۰/۸۰۵۳ |
| | Normalize | ۰/۳ | ۰/۷۳ | ۰/۷۶ | ۰/۷۴ | ۰/۷۶۴۴ |
| | | ۰/۲۵ | ۰/۷۷ | ۰/۸۰ | ۰/۷۷ | ۰/۷۹۶۷ |
| | | ۰/۲ | ۰/۷۶ | ۰/۷۹ | ۰/۷۷ | ۰/۷۸۶۶ |
| | | ۰/۱۵ | ۰/۷۷ | ۰/۷۹ | ۰/۷۸ | ۰/۷۸۷۶ |

## K- نزدیکترین همسایگی (KNN)[1]

الگوریتم KNN یکی از متداول‌ترین الگوریتم‌های طبقه‌بندی است. این الگوریتم مبتنی بر نمونه است و بر اساس k همسایه نزدیک، طبقه‌بندی را انجام می‌دهد. الگوریتم KNN به عنوان الگوریتم تنبل شناخته می‌شود، زیرا مبتنی بر تقریب محلی است و همه محاسبات تا انجام طبقه‌بندی معوق می‌ماند (عقیقی، عقیقی و عبادتی، ۱۳۹۶؛ رچاردز و

1 k-nearest neighbor

ریچاردز، ۱۹۹۹). این روش بر اساس شباهت داده‌ها طبقه‌بندی را انجام می‌دهد. در واقع برای هر داده‌ی آزمایشی جدید، فواصل k همسایه نزدیک را محاسبه کرده و برچسبی مشابه برچسب غالب این k همسایه برای نقطه مورد نظر را تعیین می‌کند (عقیقی، عقیقی و عبادتی، ۱۳۹۶). طبقه‌بندی کننده k–نزدیک‌ترین همسایه، یکی از الگوریتم‌های طبقه‌بندی شناخته‌شده و ساده می‌باشد. این اولین بار توسط فیکس و هادجس به عنوان یک الگوریتم ناپارامتری معرفی شد که هیچ فرضی بر توزیع داده‌های ورودی ایجاد نمی‌کند؛ بنابراین به طور گسترده در کاربردهای مختلف استفاده می‌شود (دودا، هارت و استورک، ۲۰۱۲).

در طبقه‌بندی‌کننده KNN، یک نمونه ناشناخته براساس شباهت بین نمونه‌های شناخته‌شده آموزش‌دیده یا برچسب‌دار برمبنای محاسبه فاصله بین نمونه‌های ناشناس با نمونه‌های برچسب‌دار، شناخته می‌شود. سپس k نزدیک‌ترین نمونه‌ها به عنوان پایه برای طبقه‌بندی انتخاب می‌شوند و نمونه نامشخص (xtest) به کلاسی اختصاص می‌یابد که بیشترین نمونه‌ها را در میان نزدیک‌ترین نمونه‌ها دارد. به همین منظور، الگوریتم طبقه‌بندی کننده KNN بستگی دارد به: (۱) تعداد k همسایه عدد صحیح و تغییر مقدار پارامتر k که ممکن است نتایج طبقه‌بندی را تغییر دهد. (۲) مجموعه داده‌های برچسب‌دار؛ بنابراین اضافه کردن یا حذف هر گونه نمونه به نمونه‌های آموزشی، بر تصمیم نهایی طبقه‌بندی کننده KNN، تأثیر می‌گذارد و (۳) معیار فاصله. در KNN، از فاصله اقلیدسی معمولاً به عنوان معیار فاصله برای اندازه‌گیری فاصله بین دو نمونه استفاده می‌شود. طبقه‌بندی کننده KNN به صورت تحلیلی قابل ردیابی است و به سادگی پیاده‌سازی می‌شود، اما یکی از مشکلات اصلی الگوریتم KNN این است که به همه نمونه‌های آموزشی نیاز دارد که در زمان اجرا در حافظه باشند؛ به همین دلیل، طبقه‌بندی مبتنی بر حافظه نامیده می‌شود (دودا، هارت و استورک، ۲۰۱۲؛ تاروات، قانم و حسنین، ۲۰۱۳).

این الگوریتم نیز همانند الگوریتم درخت تصمیم، پس از فراخوانی داده‌ها در محیط برنامه، داده‌ها را به دو بخش داده‌های آموزشی و داده‌های تست تقسیم نموده، در ادامه با فراخوانی کتابخانه scikit-learn، ساب پکیج sklearn.neighbors و طبقه‌بندی کننده

KNeighborsClassifier، مدل K- نزدیکترین همسایه را ساخته و داده‌های آموزشی (x_train, y_train) را وارد مدل کرده تا مدل آموزش ببیند. در ادامه برای مشخص نمودن دقت مدل، داده‌های تست (x_test) را وارد مدل کرده تا پیش‌بینی کند و در مقایسه با برچسب‌های داده‌های تست (y_test) دقت پیش‌بینی را ارزیابی نماید. این فرایند را بار دیگر انجام داده اما این بار به جای استفاده از داده‌های اصلی، برای هم مقیاس شدن داده‌ها، آن‌ها را نرمال‌سازی کرده و سپس داده‌ها را به دو دسته داده‌های آموزشی و تست تقسیم کرده و دقت مدل را برای هر یک از مقادیر داده‌های تست، ارزیابی کردیم. نتایج ارزیابی مدل دسته‌بندی کننده KNN در جدول (۲) قابل مشاهده است.

**جدول ۲. نتایج ارزیابی مدل *K*- نزدیکترین همسایه**

| | Data type | Test size | k- nearest neighbor | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|---|---|---|
| KNN | Original | ۰/۳ | ۱۳ | ۰/۸۳ | ۰/۸۵ | ۰/۸۲ | ۰/۸۴۸۸ |
| | | ۰/۲۵ | ۱۳ | ۰/۸۶ | ۰/۸۵ | ۰/۸۲ | ۰/۸۵۰۲ |
| | | ۰/۲ | ۱۳ | ۰/۸۶ | ۰/۸۷ | ۰/۸۵ | ۰/۸۶۶۶ |
| | | ۰/۱۵ | ۱۴-۱۳-۶ | ۰/۸۹ | ۰/۸۸ | ۰/۸۷ | ۰/۸۸۴۹ |
| | Normalize | ۰/۳ | ۲۰-۱۹ | ۰/۸۵ | ۰/۸۶ | ۰/۸۵ | ۰/۸۶۲۲ |
| | | ۰/۲۵ | ۲۵ | ۰/۸۶ | ۰/۸۶ | ۰/۸۴ | ۰/۸۶۰۹ |
| | | ۰/۲ | ۲۱ | ۰/۸۶ | ۰/۸۷ | ۰/۸۶ | ۰/۸۶۶۶ |
| | | ۰/۱۵ | ۲۵ | ۰/۸۹ | ۰/۸۹ | ۰/۸۹ | ۰/۸۹۳۸ |

## ماشین بردارپشتیبان (SVM)[1]

SVM یک ابزار ریاضی است که مبتنی بر اصل حداقل‌سازی خطای عملیاتی است و سابقه آن به سال ۱۹۶۰ بر می‌گردد. SVM بر اساس نظریه یادگیری آماری بنا نهاده شده و یک روش آماری غیرپارامتریک نظارت شده است (لامبدا و کومار، ۲۰۱۶).

در کاربردهای امروزی یادگیری ماشین، ماشین بردار پشتیبان به عنوان یکی از قدیمی‌ترین و دقیق‌ترین روش‌ها در میان الگوریتم‌های معروف شناخته می‌شود.

1 Support Vector Machine

الگوریتم SVM جزء الگوریتم‌های تشـخیص الگوی دسـته‌بندی می‌باشـد. از الگوریتم SVM، در هر جایی که نیاز به تشخیص الگو یا دسته‌بندی اشیاء در کلاس‌های خاص باشد می‌توان استفاده کرد. همچنین ماشین بردار پشتیبان یکی از روش‌های یادگیری باناظر است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. مبنای کاری دسته‌بندی کننده ایـن مدل، دسـته‌بندی خطی داده‌ها می‌باشـد و در تقسـیم خطی داده‌ها سـعی بر آن اسـت خطی انتخاب شـود که حاشـیه اطمینان بیشـتری را داشـته باشـد. البته ماشـین بردار پشـتیبان در دسـته‌بندی غیرخطی هم کاربرد دارد. به طور کلی این الگوریتم از یک نگاشـت غیرخطی برای تبدیل داده‌های اصلی به ابعاد بالاتر استفاده می‌کند و سپس در این بعد جدید به دنبال ابرصفحه‌ای است که نمونه‌های یک کلاس را از کلاس‌های دیگر جدا کند. با یک نگاشت غیرخطی مناسب، مجموعه داده‌های دو کلاسی می‌توانند توسط یک ابرصفحه جدا شوند. در واقع ایده اصلی ماشین بردار پشتیبان رسم ابرصفحه‌هایی در فضا است که عمل تمایز نمونه‌های مختلف داده‌ها را به طور بهینه انجام می‌دهند و ابرصـفحه هایی را که بیشـترین حاشـیه جداسـازی را دارند پیدا می‌کند و نزدیک‌ترین داده‌های آموزشـی به ابرصـفحه، جداکننده بردارهای پشـتیبان نامیده می‌شـوند. این روش تا حدودی پیچیده اسـت و ویژگی مثبت آن در این است که به تعداد نمونه‌های آموزش وابسته نمی‌باشد و با تعداد ویژگی‌های بالا و تعداد نمونه‌های کم می‌تواند به خوبی کار کند. از جمله محدودیت‌های این الگوریتم این اسـت که فقط روی داده‌هایی با مقدار واقعی کار می‌کند و انواع دیگر داده‌ها باید به داده‌های عددی تبدیل شوند (فضلی و مومنی، ۲۰۱۳). در حقیقت تابع هسته از شباهت بین داده‌ها در فضـای اولیه برای یافتن شـباهت بین بردارها در فضـایی با ابعاد بالاتر اسـتفاده می‌کند. تابع هسـته[1] می‌تواند تابع چندجمله‌ای، تابع RBF، تابع تانژانت هایپربولیک یا توابع مناسب دیگری انتخاب شود (شکیبا، خدری و فقیه موسوی، ۱۳۹۶).

این الگوریتم نیز همانند الگوریتم‌های قبلی، پس از فراخوانی داده‌ها در محیط برنامه، داده‌ها را به دو بخش داده‌های آموزشی و داده‌های تست تقسیم نموده، در ادامه با فراخوانی

1 Kernel

کتابخانه scikit-learn، ساب پکیج sklearn.svm و مدل SVC، مدل ماشین بردار پشتیبان را ساخته و داده‌های آموزشی (x_train, y_train) را وارد مدل کرده تا مدل آموزش ببیند. همان‌طور که گفته شد، تابع کرنل یا هسته انواع مختلفی دارد که در این تحقیق از یک تابع پایه‌ای شعاعی گوسی (RBF)[1] استفاده شده است. در ادامه برای مشخص نمودن دقت مدل، داده‌های تست (x_test) را وارد مدل کرده تا پیش‌بینی کند و در مقایسه با برچسب‌های داده‌های تست (y_test) دقت پیش‌بینی را ارزیابی نماید.

این فرایند را بار دیگر انجام داده اما این بار به جای استفاده از داده‌های اصلی، برای هم‌مقیاس شدن داده‌ها، آن‌ها را نرمال‌سازی کرده و سپس داده‌ها را به دو دسته داده‌های آموزشی و تست تقسیم کرده و دقت مدل را برای هر یک از مقادیر داده‌های تست، ارزیابی کردیم. نتایج ارزیابی مدل SVM در جدول (۳) قابل مشاهده است.



1 Gaussian Radial Basis Function

**جدول ۳. نتایج ارزیابی مدل SVM**

| | Data type | Test size | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|---|---|
| SVM | Original | ۰/۳ | ۰/۸۰ | ۰/۸۰ | ۰/۷۵ | ۰/۸۰۴۴ |
| | | ۰/۲۵ | ۰/۷۹ | ۰/۸۰ | ۰/۷۴ | ۰/۸۰۷۴ |
| | | ۰/۲ | ۰/۷۸ | ۰/۸۱ | ۰/۷۶ | ۰/۸۱۳۳ |
| | | ۰/۱۵ | ۰/۷۸ | ۰/۸۲ | ۰/۷۹ | ۰/۸۲۳۰ |
| | Normalize | ۰/۳ | ۰/۷۹ | ۰/۸۲ | ۰/۷۵ | ۰/۸۱۷۷ |
| | | ۰/۲۵ | ۰/۷۸ | ۰/۸۱ | ۰/۷۴ | ۰/۸۰۷۴ |
| | | ۰/۲ | ۰/۷۹ | ۰/۸۳ | ۰/۸۰ | ۰/۸۲۶۶ |
| | | ۰/۱۵ | ۰/۸۲ | ۰/۸۴ | ۰/۷۹ | ۰/۸۴۰۷ |

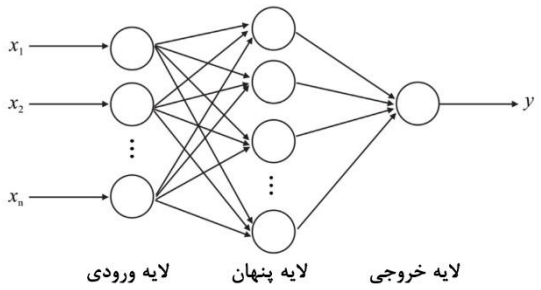
## پرسپترون چندلایه (MLP)[1]

شبکه عصبی مصنوعی[2] مشابه مغز انسان است زیرا هر دوی آن‌ها شامل تعداد زیادی پردازش و واحدهای هوشمند هستند که نورون‌ها یا سلول‌های مغزی نامیده می‌شوند. هدف توسعه شبکه عصبی مصنوعی، یافتن رابطه بین داده‌های ورودی و داده‌های خروجی است. نورون‌ها مانند سلول‌های مغزی بیولوژیکی عمل می‌کنند تا لایه‌هایی را بسازند که عملکرد مدل را ارزیابی می‌کنند. شبکه عصبی مصنوعی به عنوان یک سیستم توزیع‌شده موازی شناخته می‌شود که شامل نورون‌های محاسباتی ساده است. با استفاده از آزمون و خطا، تعداد نورون‌ها در لایه‌های پنهان و تعداد لایه‌های پنهان را می‌توان محاسبه کرد (اسپارکس، هرناندز و استوز، ۲۰۰۸).

مزیت اصلی روش شبکه عصبی مصنوعی رسیدن به راه حل مشکلات پیچیده است که حل آن با سایر تکنیک‌های متعارف دشوار است و سرعت پردازش آن بسیار سریع است (قریتلار، کومار و کریشنا، ۲۰۱۸). شبکه عصبی تکنیکی است که توانایی ضبط و نمایش روابط پیچیده ورودی/ خروجی را دارد (بیشاپ، ۱۹۹۵). یکی از متداول‌ترین مدل‌های شبکه عصبی، شبکه عصبی پیشخور (MLP) نامیده می‌شود (ون اک و ون وزل، ۲۰۰۸).

1 Multi-Layer Perceptron
2 Artificial Neural Network

ساختار یک شبکه عصبی مصنوعی دارای سه نوع لایه بوده و لایه‌های ورودی، پنهان و خروجی لایه‌های مذکور هستند. تعداد لایه‌های پنهان و نورون‌های هر لایه، با اعمال الگوریتم‌های بهینه‌سازی، قابل محاسبه است. جزئیات ساختار شبکه عصبی MLP و اتصالات، بسیار وابسته به متغیرهای مسئله هستند و برای ایجاد ارتباطات، گام آموزشی به کار گرفته می‌شود. به منظور دستیابی به بهترین مدل، ساختار بهینه باید با استفاده از الگوریتم‌های بهینه‌سازی مناسب انتخاب شود (مندز سانتیاگو و تجا، ۲۰۰۰).




**شکل ۱. معماری شبکه عصبی پرسپترون چندلایه**

در این الگوریتم نیز همانند الگوریتم‌های قبلی، پس از فراخوانی داده‌ها در محیط برنامه، داده‌ها را به دو بخش داده‌های آموزشی و داده‌های تست تقسیم نموده، سپس x_train و x_test را نرمال‌سازی می‌نماییم.

در ادامه با فراخوانی کتابخانه scikit-learn، ساب پکیج sklearn.neural_network و طبقه‌بندی کننده MLPClassifier، مدل MLP را ساخته و داده‌های آموزشی (x_train, y_train) را وارد مدل کرده (که x_train نرمال‌سازی شده هستند) تا مدل آموزش ببیند. در این تحقیق تعداد لایه‌های پنهان این مدل سه لایه در نظر گرفته شده و در هر لایه تعداد ۱۰ نورون را قرار دادیم (۱۰، ۱۰،۱۰). در ادامه برای مشخص نمودن دقت مدل، داده‌های تست (x_test) را که نرمال‌سازی شده هستند وارد مدل کرده تا پیش‌بینی کند و در مقایسه با برچسب‌های داده‌های تست (y_test) دقت پیش‌بینی را ارزیابی نماید. نتایج ارزیابی مدل MLP در جدول (۴) قابل مشاهده است.

**جدول ٤. نتایج ارزیابی مدل MLP**

| | Test size | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|---|
| MLP | ۰/۳ | ۰/۸۶ | ۰/۸۷ | ۰/۸۵ | ۰/۸۶۶۶ |
| | ۰/۲۵ | ۰/۸۴ | ۰/۸۶ | ۰/۸۴ | ۰/۸۵۵۶ |
| | ۰/۲ | ۰/۸۱ | ۰/۸۳ | ۰/۸۱ | ۰/۸۳۳۳ |
| | ۰/۱۵ | ۰/۸۸ | ۰/۸۸ | ۰/۸۸ | ۰/۸۸۴۹ |

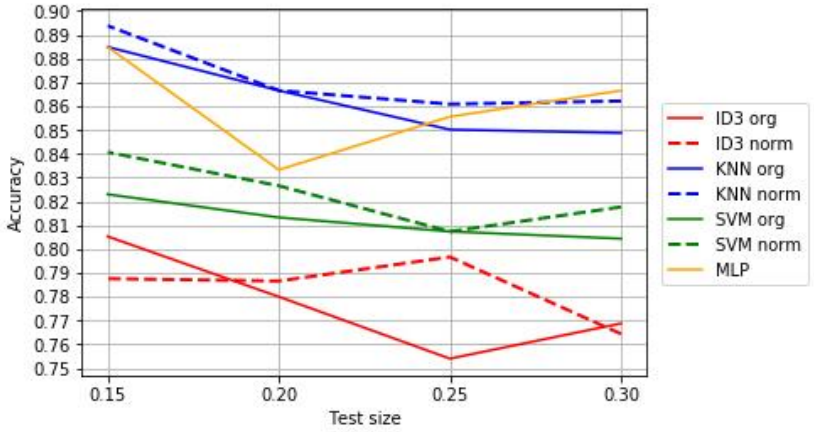
## نتیجه گیری و پیشنهادها

اهدای خون به دلیل نقش حیاتی و حساسی که در امر حفظ سلامت و بقاء زندگی انسان دارد مورد توجه می‌باشد. در جهان امروز علیرغم تحول عظیم علمی و با وجود پیشرفت‌های بزرگی که در علوم پزشکی رخ داده است، هنوز تأمین کافی خون سالم یکی از چالش‌ها و دغدغه‌های مجامع پزشکی جهان است. در این مقاله سعی شد تا از تکنیک‌های داده‌کاوی و یادگیری ماشین برای پیش‌بینی اهداء خون استفاده کنیم تا با استفاده از این مکانیزم بتوانیم پیش‌بینی کنیم که در بازه‌های زمانی مختلف، چه میزان خون به بانک‌ها و مراکز انتقال خون اهداء خواهد شد که در این صورت بتوانیم حجم مورد نیاز بانک‌های خون مناطق مختلف را تخمین و تأمین نماییم. در همین راستا از چند الگوریتم طبقه‌بندی در یادگیری با نظارت از جمله الگوریتم‌های درخت تصمیم، KNN، SVM و MLP برای پیش‌بینی اهداء خون استفاده شد و نتایج میزان دقت هر کدام در قالب جداول مختلف ارائه شد.

در اجرای الگوریتم درخت تصمیم، بیشترین میزان دقت در داده اصلی با اندازه داده‌های تست ۰/۱۵ برابر ۰/۸۰۵۳ بوده و زمانی که داده‌ها نرمال‌سازی شدند، با اندازه داده‌های تست ۰/۲۵، دقت برابر ۰/۷۹۶۷ بوده است. در پیاده‌سازی الگوریتم KNN، بیشترین دقت در داده‌های اصلی با اندازه نمونه ۰/۱۵ و با تعداد همسایه (۶-۱۳-۱۴)، برابر ۰/۸۸۴۹ بوده و زمانی که داده‌ها نرمال‌سازی شدند، با اندازه داده تست ۰/۱۵، با تعداد همسایه ۲۵، دقت برابر با ۰/۸۹۳۸ بوده است. در اجرای الگوریتم SVM با تابع کرنل RBF، در داده‌های اصلی و در اندازه داده تست ۰/۱۵، دقت برابر ۰/۸۲۳۰ ثبت شده و در

داده‌های نرمال با اندازه داده تست ۰/۱۵، دقت ۰/۸۴۰۷ بوده است. در آخر با پیاده‌سازی الگوریتم MLP، بیشترین میزان دقت با داده‌های تست ۰/۱۵، برابر ۰/۸۸۴۹ بوده است.

در شکل (۲) نیز نتایج ارزیابی هر کدام از الگوریتم‌ها جهت مقایسه بصری با توجه به اندازه داده‌های تست هرکدام، با استفاده از کتابخانه matplotlib.pyplot، رسم شده و به نمایش درآمده است. همان‌طور که ملاحظه می‌شود در کل، الگوریتم KNN و MLP در ارزیابی‌ها از دقت بیشتری برخوردار هستند.




**شکل ۲. مقایسه نتایج ارزیابی الگوریتم ها**

در کارهای آتی می‌توان با بدست آوردن ویژگی‌های دیگر در مجموعه داده و به کارگیری روش‌های ترکیبی در پیاده‌سازی الگوریتم‌ها، میزان دقت پیش‌بینی را افزایش داد.

## منابع


- Aghighi, Farzaneh; Hossein Aghighi and Omid Mehdi ebabati. (2017). " Evaluation of the efficiency of SVM and KNN Classification algorithms to extract urban effects from LiDAR cloud points", Second International Conference on Knowledge-based Research in Computer Engineering & Information Technology, Tehran, Majlisi University. (in persian)
- Akben, S. B. (2018). Early Stage Chronic Kidney Disease Diagnosis by Applying Data Mining Methods to Urinalysis, Blood Analysis and Disease History. *IRBM*, *39*(5), 353-358.
- Ashoori, M., Alizade, S., Eivary, H. S. H., Rastad, S., & Eivary, S. S. H. (2015). A model to predict the sequential behavior of healthy blood donors using data mining. *Journal of Research & Health*, *5*(2), 141-148.
- Bahel, D., Ghosh, P., Sarkar, A., & Lanham, M. A. (2017). Predicting Blood Donations Using Machine Learning Techniques. In *CONFERENCE PROCEEDINGS BY TRACK* (p. 323).
- Balakrishnan, J. M. D. (2010). Significance of classification Techniques in prediction of Learning disabilities. *arXiv preprint arXiv:1011.0628*.
- Bhardwaj, A., Sharma, A., & Shrivastava, V. K. (2012). Data mining techniques and their implementation in blood bank sector–a review. *International Journal of Engineering Research and Applications (IJERA)*, *2*(4), 1303-1309.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth and Brooks. *Cole Statistics/Probability Series*.
- Brunassi, L. D. A., Moura, D. J. D., Nääs, I. D. A., Vale, M. M. D., Souza, S. R. L. D., Lima, K. A. O. D., ... & Bueno, L. G. D. F. (2010). Improving detection of dairy cow estrus using fuzzy logic. *Scientia Agricola*, *67*(5), 503-509.
- Cardoso, H. F. (2008). Sample-specific (universal) metric approaches for determining the sex of immature human skeletal remains using permanent tooth dimensions. *Journal of Archaeological Science*, *35*(1), 158-168.
- Chang, H. H., & Tsay, S. F. (2004). Integrating of SOM and K-mean in data mining clustering: An empirical study of CRM and profitability evaluation.
- Darwiche, M., Feuilloy, M., Bousaleh, G., & Schang, D. (2010, May). Prediction of blood transfusion donation. In *2010 Fourth International Conference on Research Challenges in Information Science (RCIS)* (pp. 51-56). IEEE.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.
- Elmamouz, G. and M. Nadimi. (2012). A review of methods for prediction of type 2 diabetes based on Bayesian theory. National Conference on Science and Computer Engineering.
- Fazli H, Momeni H. (2013). Comparison and evaluation of data mining algorithms, decision tree and SVM application for intrusion detection. In: Proceedings of 8th Symposium progress in science and technology 2013, Mashhad. Iran.
- Ghritlahre, H. K., & Prasad, R. K. (2018). Exergetic performance prediction of solar air heater using MLP, GRNN and RBF models of artificial neural network technique. *Journal of environmental management*, *223*, 566-575.
- Goldschmidt, R., & Passos, E. (2005). *Data mining: um guia prático*. Gulf Professional Publishing.
- Grilli, E., Menna, F., & Remondino, F. (2017). A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 339.
- Hughes, A. M. (1994). Strategic database marketing. IL: Probus Publishing Company.
- Khamis, H. S., Cheruiyot, K. W., & Kimani, S. (2014). Application of k-nearest neighbour classification in medical data mining. *International Journal of Information and Communication Technology Research*, *4*(4).
- Khomri, Neda and Hadi Rainani. (2018). "Data mining, concepts and applications (Electronic City)", Second International Conference on Electrical Engineering, Computer Science and Information Technology, Hamedan. (in persian)
- Lambda, A., & Kumar, D. (2016). Survey on KNN and Its Variants. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(5).

- Mendez-Santiago, J., & Teja, A. S. (2000). Solubility of solids in supercritical fluids: consistency of data and a new model for cosolvent systems. *Industrial & Engineering Chemistry Research*, *39*(12), 4767-4771.
- Mostafa, M. M. (2009). Profiling blood donors in Egypt: A neural network analysis. *Expert Systems with Applications*, *36*(3), 5031-5038.
- Nowruzi Tiolla, Sare; Morteza Mousavi and Manouchehr Kazemi. (2017). "Intrusion Detection Using Combined Clustering and Knn Algorithm", Fourth National Conference on Information Technology, Computer and Telecommunications, Mashhad, Torbat Heydarieh University. (in persian)
- Quinlan, J. R. (1993). Program for machine learning. *C4. 5*.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81-106.
- Reinartz, W. J., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of marketing*, *64*(4), 17-35.
- Richards, J. A., & Richards, J. A. (1999). *Remote sensing digital image analysis* (Vol. 3, pp. 10-38). Berlin et al.: Springer.
- Shakiba, Zeinab; Mahdieh Khedri and Faeghe Faghih Mousavi. (2017) "The performance Comparison of KNN and SVM Algorithms in Categorization of Texts", Fourth International Conference on Knowledge Based Research in Computer Engineering and Information Technology, Tehran, University of Abrar. (in persian)
- Sparks, D. L., Hernandez, R., & Estévez, L. A. (2008). Evaluation of density-based models for the solubility of solids in supercritical carbon dioxide and formulation of a new model. *Chemical Engineering Science*, *63*(17), 4292-4301.
- Testik, M. C., Ozkaya, B. Y., Aksu, S., & Ozcebe, O. I. (2012). Discovering blood donor arrival patterns using data mining: A method to investigate service quality at blood centers. *Journal of medical systems*, *36*(2), 579-594.
- Tharwat, A., Ghanem, A. M., & Hassanien, A. E. (2013, December). Three different classifiers for facial age estimation based on k-nearest neighbor. In *2013 9th International Computer Engineering Conference (ICENCO)* (pp. 55-60). IEEE.
- Trabelsi, A., Elouedi, Z., & Lefevre, E. (2018). Decision tree classifiers for evidential attribute values and class labels. *Fuzzy Sets and Systems*.
- van Eck, N. J., & van Wezel, M. (2008). Application of reinforcement learning to the game of Othello. *Computers & Operations Research*, *35*(6), 1999-2017.
- Yeh, I. C., Yang, K. J., & Ting, T. M. (2009). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, *36*(3), 5866-5871.
- Yu, P. L. H., Chung, K. H., Lin, C. K., Chan, J. S. K., & Lee, C. K. (2007). Predicting potential drop-out and future commitment for first-time donors based on first 1· 5-year donation patterns: the case in Hong Kong Chinese donors. *Vox sanguinis*, *93*(1), 57-63.


**استناد به این مقاله:** شناسه دیجیتال (DOI): 10.22091/JEMSC.2018.1278

فهمی حسن، آرش، مغاری، محمدرضا، عبادتی، امیدمهدی. (۱۳۹۸). «پیش‌بینی اهداء خون با استفاده از داده‌کاوی بر پایه الگوریتم‌های درخت تصمیم، KNN، SVM و MLP». *مدیریت مهندسی و رایانش نرم*، ۶ (۱)، ۱۲۹-۱۰۹.