مجله مدیریت اطلاعات سلامت
سال بیستم/ شماره چهارم/ زمستان ۱۴۰۲

Published by *Isfahan University of Medical Sciences*
DOI: 10.48305/him.2024.42198.1151

پیش‌بینی پذیرش بخش اورژانس با استفاده از داده‌کاوی (مطالعه موردی: بیمارستان امام علی شهرکرد)

صبا پایدار [۱]، غلامعلی رئیسی اردلی [۲]، حسین رئیسی دزکی [۳]

مقاله پژوهشی

چکیده

مقدمه: امروزه، بیمارستان‌ها با چالش‌هایی مانند ازدحام در بخش اورژانس و افزایش بی‌نظمی و اختلال در کار پرسنل مواجه هستند که سبب افزایش نارضایتی بیماران می‌شود. با پیشرفت هوش مصنوعی و گسترش علم داده‌کاوی، پیش‌بینی پذیرش بیماران اهمیت زیادی پیدا کرده است. هدف این پژوهش، پیش‌بینی پذیرش بیماران بخش اورژانس در بیمارستان امام‌علی (ع) شهرکرد است.

روش بررسی: در این پژوهش، ۲۱۸۰ پرونده بیماران بخش اورژانس بیمارستان مورد بررسی قرار گرفت و اطلاعات اولیه بیماران شامل مشخصات فردی، علائم حیاتی بیمار و سطح تریاژ که توسط پرستاران در فرم تریاژ ثبت شده بود، استخراج شد. با استفاده از ماتریس مقایسات زوجی، ویژگی‌های مؤثر توسط خبرگان انتخاب شدند. سپس با استفاده از الگوریتم‌های بیز ساده، درخت تصمیم، جنگل تصادفی و ماشین بردار پشتیبان، داده‌ها طبقه‌بندی شدند.

یافته‌ها: از بین ۱۴ ویژگی جمع‌آوری شده، ۹ ویژگی منتخب توسط خبرگان انتخاب شد و نتایج نشان داد که الگوریتم جنگل تصادفی با دقت ۹۲/۲ درصد و حساسیت ۹۶ درصد و صحت ۸۶/۳ درصد نسبت به الگوریتم‌های بیز ساده ، ماشین بردار پشتیبان و درخت تصمیم بهترین عملکرد را در پیش‌بینی پذیرش بیماران این مطالعه موردی داشته است.

نتیجه‌گیری: نتایج نشان می‌دهد که دقت مدل‌های یادگیری ماشین با استفاده از نظرات خبرگان در انتخاب ویژگی‌ها بیشتر می‌شود و و با توجه به حوزه مورد مطالعه، الگوریتم‌های یادگیری ماشین نتایج متفاوتی دارند که در این مطالعه الگوریتم جنگل تصادفی بهترین عملکرد را در پیش‌بینی پذیرش بیماران داشته است.

واژه‌های کلیدی: داده‌کاوی؛ پیش‌بینی؛ پذیرش بیماران؛ بخش اورژانس.

پیام کلیدی: استفاده از الگوریتم‌های یادگیری ماشین می‌تواند به عنوان ابزار کارآمد برای کمک به کادر درمان به منظور پیش‌بینی پذیرش بیماران در بخش اورژانس مورد استفاده قرار بگیرد.

دریافت مقاله:۱۴۰۲/۵/۲۳ پذیرش مقاله:۱۴۰۲/۹/۱ تاریخ انتشار:۱۴۰۲/۱۰/۱۵

ارجاع: پایدار صبا، رئیسی اردلی غلامعلی، رئیسی دزکی حسین. **پیش‌بینی پذیرش بخش اورژانس با استفاده از داده‌کاوی (مطالعه موردی: بیمارستان امام علی شهرکرد).** مدیریت اطلاعات سلامت۱۴۰۲؛۲۰ (۴):۱۹۷-۱۹۰.

مقدمه

هزینه‌های سلامت و درمان در کشورها روز به روز در حال افزایش است و عواملی مانند افزایش جمعیت، سالمندی جمعیت، بیماری‌های واگیردار و انتظارات بالای بیماران از خدمات درمانی باعث افزایش تقاضا برای این خدمات شده است. بنابراین، تصمیم‌گیرندگان و مسئولین حوزه سلامت همواره سعی دارند که سیستم‌های درمانی را توسعه و بهبود داده و هزینه‌های آن را کاهش دهند (۱, ۲). بیمارستان، بازوی مهم ارائه‌ی خدمات درمانی می‌باشد که قسمت عمده منابع نظام سلامت را به خود اختصاص داده است و بخش اورژانس یکی از حیاتی‌ترین بخش‌های بیمارستان می‌باشد. این بخش، اولین مکان ارائه خدمات تشخیصی–درمانی به بیماران بوده که از واحدهای حساس و پیچیده بیمارستان‌ها به‌حساب می‌آید و به هماهنگی در سطح بالا بین منابع و تجهیزات نیاز دارد. از این‌رو رسیدگی سریع و مطلوب به بیماران اورژانسی حائز اهمیت است (۳). از طرفی رسالت اصلی بخش اورژانس ارائه مناسب‌ترین مراقبت در کوتاه‌ترین زمان ممکن است. با توجه به اینکه حدود ۷۸ درصد از مراجعین به بیمارستان‌ها را مراجعه‌کنندگان به اورژانس‌ها تشکیل می‌دهند، دقایق و حتی ثانیه‌ها هم دارای اهمیت هستند؛ چراکه ۸۰ درصد مرگ و میرها در ۲۰ دقیقه اول بعد از حادثه رخ می‌دهد و بیشتر حوادث در ۱۰ دقیقه اول، زمانی که تصمیمات مهم گرفته می‌شود، پیشرفت کرده یا مهار می‌شود (۴). یکی از قسمت‌های کلیدی در بخش اورژانس، قسمت پذیرش بیماران است که تصمیمات در مورد پذیرش بیماران را بر عهده داشته و یکی از واحدهای پر هزینه در بیمارستان می‌باشد و به‌طور مثال، پذیرش بیماران بخش اورژانس حدود ۸/۳ درصد از هزینه‌های بهداشت ملی ایالات‌متحده آمریکا را تشکیل می‌دهد (۵).

۱- دانشجوی کارشناسی ارشد، مهندسی صنایع، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه صنعتی اصفهان،اصفهان، ایران
۲- دانشیار، مهندسی صنایع، دانشکده مهندسی صنایع و سیسم‌ها دانشگاه صنعتی اصفهان، اصفهان، ایران
۳- دانشجوی کارشناسی ارشد، مهندسی صنایع، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه صنعتی اصفهان، اصفهان، ایران
نویسنده طرف مکاتبه: صبا پایدار؛ دانشجوی کارشناسی ارشد، مهندسی صنایع، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه صنعتی اصفهان،اصفهان، ایران
Email: saba.p1374@gmail.com

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

مطالعات و تحقیقات برای بهبود عملکرد فرایند پذیرش بیماران و همچنین کاهش تعداد پذیرش‌های غیرضروری می‌تواند به صرفه‌جویی قابل‌توجهی در هزینه و افزایش سطح رضایت بیماران منجر شود.

یکی از موضوعات تحقیقاتی در این راستا که توجه ویژه‌ای به آن شده است، پیش‌بینی پذیرش بیماران می‌باشد. پیش‌بینی پذیرش بیمار در بخش اورژانس به بیمارستان‌ها کمک می‌کند تا بیماران را به‌طور کارآمد پذیرش کنند و سطح رضایت بیماران را به‌طور چشمگیری افزایش دهند. یکی از تکنیک‌ها و ابزارهای مؤثر برای توسعه و بهبود خدمات درمانی در بخش اورژانس به‌خصوص در مورد موضوع پیش‌بینی پذیرش بیماران، داده‌کاوی و استفاده از داده‌ها می‌باشد. داده کاوی فرایندی است که حجم قابل‌توجهی از داده‌ها را به اطلاعات معنی‌دار برای پشتیبانی کارآمد از تصمیم‌گیری تبدیل می‌کند. به‌عبارت دیگر به مجموعه‌ای از روش‌های قابل اعمال بر پایگاه داده‌های بزرگ و پیچیده به‌منظور کشف الگوهای پنهان و جالب توجه نهفته در بین داده‌ها، داده‌کاوی گویند (۶).

روش پیشنهادی این پژوهش، استفاده از داده‌کاوی برای پیش‌بینی پذیرش بیمار در بخش اورژانس است. با توجه به شناسایی عوامل مؤثر در پذیرش بیمار می‌توان وضعیت بیماران را با دقت خوبی پیش‌بینی کرد که منجر به کاهش زمان پذیرش بیمار و ممانعت از ازدحام و شلوغی در بخش اورژانس می‌شود. در این پژوهش ابتدا با استفاده از نظرات متخصصان حوزه سلامت، عوامل مؤثر از بین کلیه عوامل شناسایی شده انتخاب و در نهایت با استفاده از الگوریتم‌های یادگیری ماشین و مقایسه‌ی عملکرد آنها با یکدیگر به پیش‌بینی پذیرش بخش اورژانس پرداخته شد.

Lucini و همکاران (۷) به منظور پیش‌بینی پذیرش بیماران از الگوریتم‌های یادگیری عمیق و مدل‌های گرافیکی استفاده کرده‌اند. ویژگی‌های مهم شامل سن بیمار، تاریخچه بیماری، شدت علایم، داده‌های آزمایشگاهی و تصویربرداری استخراج شدند. Graham و همکاران (۸)، با استفاده از داده‌های دو بیمارستان در ایرلند شمالی، پذیرش بخش اورژانس را پیش‌بینی کردند. متغیرهایی مانند جنسیت، سن، روز، ساعت و ماه مراجعه، نحوه ورود، سطح تریاژ، گروه درمان و مراجعه قبلی را در نظر گرفتند. الگوریتم گرادیان تقویتی با دقت ۸۱/۳۱ درصد بهترین نتیجه را ارائه داد. Parker و همکاران (۹)، با استفاده از الگوریتم رگرسیون لجستیک، احتمال بستری شدن بیماران بخش اورژانس را پیش‌بینی کردند. ۸ متغیر در نظر گرفته شده شامل گروه سنی، ملیت، کد پستی، روز و زمان مراجعه، سطح تریاژ، نحوه ورود و وضعیت تب می‌باشند. این مدل با سطح زیر منحنی ۰/۸۲۵ به‌خوبی عمل کرده است. Mowbray و همکاران (۱۰)، با استفاده از الگوریتم‌های یادگیری ماشین، پذیرش بخش اورژانس بیمارستان را برای بیماران بالای ۷۵ سال پیش‌بینی کردند. متغیرهای پیش‌بینی کننده شامل سندرم‌های سالمندی، ارزیابی‌های عملکردی و نیازهای مراقبت‌های اولیه بودند. گرادیان تقویتی بهترین الگوریتم برای پیش‌بینی پذیرش بوده است. De Hond و همکاران (۱۱)، با استفاده از الگوریتم‌های یادگیری ماشین، پذیرش بخش اورژانس بیمارستان را برای بیماران اورژانسی سه بیمارستان بررسی کردند. متغیرهای پیش‌بینی کننده شامل جمعیت‌شناسی، فوریت، شکایات، شدت بیماری، اختلالات هم‌زمان و پیچیدگی بودند. درخت تصمیم تقویت شده با گرادیان بهترین عملکرد را داشته است. Menditto و همکاران (۱۲)، با استفاده از الگوریتم‌های یادگیری ماشین، بستری مجدد بخش اورژانس بیمارستان را برای بیماران مبتلا به کرونا پیش‌بینی کردند. نتایج نشان داد که ۱۱ درصد از آن‌ها پس از مرخصی دوباره بستری شدند .

از آنجایی‌که موضوع مورد بحث، مرتبط با جان انسان‌ها بوده و از این نظر دارای حساسیت بالایی می‌باشد. همچنین، بر اساس موارد مشابه، مشاهده شده است که قانع کردن مدیریت بیمارستان‌ها جهت استفاده از این رویکردها و الگوریتم‌ها با توجه به نتایج به‌دست‌آمده کار دشواری است. به همین جهت نیاز است که پژوهش‌های بیشتر برای افزایش دقت عملکرد الگوریتم‌ها و کم کردن خطای آنها انجام شود. در این پژوهش نیز، با بررسی میزان رضایتمندی مراجعین از خدمات بیمارستان اهمیت این موضوع متذکر شده و همچنان فقدان عملکرد مناسب در این بخش حس می‌شود. علاوه بر این، در حوزه سلامت به‌خصوص در بخش پذیرش اورژانس کمتر مطالعه‌ای جهت انتخاب ویژگی از نظر خبرگان حوزه سلامت استفاده کرده است. به دلیل اهمیت و حساسیت مسئله این حوزه، استفاده از نظر متخصصین باتجربه، می‌تواند کمک شایانی در جهت تعیین ویژگی‌های مناسب کند. بنابراین در این پژوهش با بررسی ویژگی‌ها توسط متخصصین این حوزه به انتخاب ویژگی‌های مناسب پرداخته شده است. لازم به توضیح است، به‌منظور تعیین سنجش مدل خود از داده‌های واقعی استفاده شده و داده‌های جمع‌آوری شده مربوط به بیمارستان امام علی (ع) شهرکرد می‌باشد.

روش بررسی

این پژوهش از نوع توصیفی بوده و مجموعه داده‌ی مورد استفاده در این پژوهش مربوط به مراجعه‌کنندگان به بخش اورژانش بیمارستان امام علی شهرکرد در بازه زمانی اردیبهشت ماه تا شهریور ماه سال ۱۴۰۰ می‌باشد. تعداد ۲۱۸۰ پرونده با توجه به مقالات مشابه و با الگوگیری از آنها به عنوان نمونه از جامعه بیماران مراجعه کننده به بخش اورژانس به صورت تصادفی انتخاب شده است و اطلاعات مورد نیاز از فرم‌های تریاژ بیماران استخراج شده است و با جمع‌آوری این اطلاعات، پایگاه داده‌ای در صحفه گسترده Excel ایجاد شده است. با استفاده از ادبیات موضوع و همچنین اطلاعات ثبت شده در فرم‌های تریاژ بیماران مورد مطالعه، لیست اولیه متغیرها، تهیه گردیده که در جدول ۱ قابل مشاهده است. در مجموع تعداد ۱۴ ویژگی دسته‌ای و عددی جمع‌آوری شده که به عنوان ویژگی‌های مستقل شناخته می‌شوند و همچنین ویژگی خروجی در دو مقدار پذیرش شده و پذیرش نشده می‌باشد. لازم به ذکر است که بخشی از پرونده‌ها انتخاب شده، بیماران بوده‌اند که در برگه تریاژ آنها پذیرش نشده ثبت شده است که کلیه پرونده‌ها در بخش مدارک پزشکی بیمارستان بایگانی شده است.

http://him.mui.ac.ir

جدول ۱: ویژگی‌های اولیه

متغیر	عبارت لاتین	متغیر	عبارت لاتین
سن	Age	جنسیت	Sex
فشارخون سیستولیک	Systolic blood pressure	اکسیژن خون	Saturation of Peripheral Oxygen
فشارخون دیاستولیک	diastolic blood pressure	مراجعه قبلی	Admitted in past day
ضربان نبض	Pulse Rate	ماه مراجعه	Arrival month
تعداد تنفس	Respiratory Rate	دلیل مراجعه	Chief complaints
دمای بدن	Temperature	نحوه مراجعه	Arrival mode
سطح تریاژ	Triage	بخش اورژانس	Emergency department

این پژوهش از نظرات و دانش متخصصان خبره در زمینه بالینی با سابقه کاری حداقل ۷ سال که تیمی متشکل از ۳ نفر (یک پزشک و دو پرستار) است بهره‌مند شده است. سپس با استفاده از نظر خبرگان و ابزار طیف ساعتی که یک ماتریس می‌باشد و اهمیت نسبی هر عنصر نسبت به عناصر دیگر در رابطه با آن خصوصیت تعیین می‌شود، به تشکیل ماتریس مقایسه زوجی پرداخته شده است (۱۳) و امتیاز هر متغیر محاسبه شده و متغیرها رتبه بندی شده‌اند.

الگوریتم‌های بیز ساده، ماشین بردار پشتیبان، درخت تصمیم و جنگل تصادفی، الگوریتم‌هایی هستند که در مسائل پیش‌بینی در حوزه سلامت مورد توجه می‌باشند و با توجه ادبیات موضوع مسئله نیز مشاهده می‌شود که این الگوریتم‌ها در مسئله پذیرش بیماران نیز بسیار مورد استفاده قرار گرفته‌اند و در این پژوهش نیز مورد استفاده قرار گرفته‌اند. در ادامه مزایای این الگوریتم‌ها در جدول ۲ قابل مشاهده می‌باشد.

جدول ۲: مزایای الگوریتم‌های یادگیری ماشین

الگوریتم	مزایا
بیز ساده	سریع – فهم آسان – تاثیر نگرفتن از ترتیب داده‌ها
ماشین بردار پشتیبان	کاهش خطا – مورد استفاده در مسائل پیچیده
درخت تصمیم	سریع – فهم آسان
جنگل تصادفی	بهرمندی از چند درخت تصمیم – بسیار دقیق

در این مطالعه داده‌ها با نسبت ۸۰ درصد به ۲۰ درصد برای داده‌های آموزشی و آزمایشی تقسیم‌بندی شده است. ارزیابی عملکرد الگوریتم‌ها یک فرایند است که با استفاده از معیارها و روش‌های مختلف، میزان دقت، صحت و حساسیت یک الگوریتم را بررسی می‌کند که برای محاسبه‌ی مقادیر این معیارها از پارامترهای ماتریس درهم ریختگی استفاده می‌شود که در شکل ۱ قابل مشاهده می‌باشد.

مقادیر واقعی		مقادیر پیش‌بینی شده: پذیرش‌نشده	مقادیر پیش‌بینی شده: پذیرش‌شده
	پذیرش‌نشده	TN (صحیح منفی)	FP (اشتباه مثبت)
	پذیرش‌شده	FN (اشتباه منفی)	TP (صحیح مثبت)

شکل ۱: ماتریس درهم‌ریختگی

صحیح مثبت (TP): بیمارانی که پذیرش شده بوده‌اند و الگوریتم نیز به‌درستی آنها را پذیرش شده تشخیص می‌دهد.

صحیح منفی (TN): بیمارانی که پذیرش نشده بوده‌اند و الگوریتم نیز به‌درستی آنها را پذیرش نشده تشخیص می‌دهد.

اشتباه مثبت (FP): بیمارانی که پذیرش نشده بوده‌اند و الگوریتم به اشتباه آنها را پذیرش شده تشخیص می‌دهد.

اشتباه منفی (FN): بیمارانی که پذیرش شده بوده‌اند و الگوریتم به اشتباه آنها را پذیرش نشده تشخیص می‌دهد.

با استفاده از این ماتریس می‌توان معیارهای دقت (Accuracy)، صحت (Precision) و حساسیت (Sensitivity) را محاسبه کرد که در ادامه تعریف شده است:

دقت: دقت هر مدل که معمول‌ترین معیار مدل‌های دسته‌بندی به شمار می‌رود با استفاده از رابطه (۱) قابل‌محاسبه است. این معیار، میزان تشخیص صحیح در کل داده‌ها را نشان می‌دهد.

$$\text{دقت (Accuracy)}: \frac{TP+TN}{TP+TN+FP+FN} \quad (۱)$$

حساسیت: حساسیت به‌عنوان نرخ مثبت درست شناخته می‌شود که مدل با این احتمال برای بیماری که پذیرش خواهد شد پیش‌بینی پذیرش شده می‌کند. نحوه محاسبه به شرح رابطه (۲) می‌باشد.

$$\text{حساسیت (Sensitivity)} \quad \frac{TP}{TP+FN} \quad (۲)$$

http://him.mui.ac.ir

صحت: وضوح به‌عنوان نرخ منفی درست می‌باشد به‌عبارت‌دیگر مدل با این احتمال برای بیماری که پذیرش نخواهد شد پیش‌بینی پذیرش نشده می‌کند. طبق رابطه (۳) به دست می‌آید.

$$\text{صحت (Precision)}: \frac{TP}{TP+FP} \quad (۳)$$

منحنی مشخصه‌ی عملکرد پذیرش :(ROC) پایه این روش را ماتریس درهم ریختگی شکل می‌دهد به صورتی که نرخ مثبت واقعی (TPR) روی محور Y و نرخ مثبت کاذب(FPR) ، بر روی محور X ترسیم می‌شود. روش محاسبه نرخ مثبت کاذب، با استفاده از رابطه ۴ می‌باشد.

$$FPR=1-Specificity \quad (۴)$$

در واقع یک منحنی ROC، نمایانگر توازن نسبی منافع و هزینه‌های مدل دسته‌بندی می‌باشد. مساحت زیر نمودار منحنی را AUC می‌نامند که همواره مقداری بین صفر تا یک می‌گیرد. هر چقدر این مساحت به یک نزدیک‌تر بوده، مدل قدرت بیشتری برای تمایز بین دو دسته دارد (۱۴).

یافته ها:

نتایج حاصل از اجرای الگوریتم‌های یادگیری ماشین با ۱۴ ویژگی در جدول ۳ آورده شده است. بهبود مقدار AUC مدل‌ها نسبت به ادبیات موضوع مسئله مدنظر این پژوهش می‌باشد که به این منظور از نظر خبرگان برای انتخاب ویژگی‌های موثرتر استفاده شده است.

در ادامه به به منظور تعیین رتبه‌بندی متغیرها از ماتریس مقایسات زوجی استفاده شده است که در جدول ۴ قابل مشاهده می‌باشد.

جدول ۳: جمع‌بندی نتایج الگوریتم‌های مورداستفاده با تمام ویژگی‌ها

مدل	شاخص ارزیابی			
	دقت	حساسیت	صحت	سطح زیر منحنی ROC
بیز ساده	۰/۷۴۵	۰/۷۹۴	۰/۶۷۶	۰/۸۰
ماشین بردار پشتیبان	۰/۷۹۴	۰/۸۶۸	۰/۶۸۹	۰/۸۶۷
درخت تصمیم	۰/۷۴۲	۰/۸۳۷	۰/۶۰۹	۰/۸۱۶
جنگل تصادفی	۰/۸۲۴	۰/۸۱۷	۰/۸۳۵	۰/۸۷۸

جدول ۴: جدول ماتریس مقایسات زوجی

	نحوه مراجعه	سطح تریاژ	ماه مراجعه	دمای بدن	تعداد تنفس	درصد اشباع اکسیژن	جنسیت	فشارخون دیاستولیک	دلیل مراجعه	سن	بخش اورژانس	ضربان نبض	فشارخون سیستولیک	مراجعه قبلی
نحوه مراجعه	۱	۰/۲	۳	۰/۲	۰/۱۴۲۸	۰/۱۴۲۸	۳	۰/۳۳۳	۰/۲	۰/۳۳۳	۰/۲	۰/۱۴۲۸	۰/۲	۱
سطح تریاژ	۵	۱	۷	۳	۱	۱	۷	۵	۳	۵	۵	۳	۵	۵
ماه مراجعه	۰/۳۳۳	۰/۱۴۲۸	۱	۰/۱۴۲۸	۰/۱۴۲۸	۰/۱۴۲۸	۱	۰/۳۳۳	۰/۲	۰/۳۳۳	۰/۲	۰/۱۴۲۸	۰/۲	۰/۳۳۳
دمای بدن	۵	۰/۳۳۳	۷	۱	۰/۳۳۳	۰/۳۳۳	۵	۳	۱	۳	۳	۱	۱	۳
تعداد تنفس	۷	۱	۷	۳	۱	۱	۷	۵	۵	۵	۳	۱	۳	۵
درصد اشباع اکسیژن	۷	۱	۷	۳	۱	۱	۷	۵	۵	۵	۳	۱	۳	۵
جنسیت	۰/۳۳۳	۰/۱۴۲۸	۱	۰/۲	۰/۱۴۲۸	۰/۱۴۲۸	۱	۰/۲	۰/۱۴۲۸	۰/۲	۰/۲	۰/۱۴۲۸	۰/۱۴۲۸	۰/۲
فشارخون دیاستولیک	۳	۰/۲	۳	۰/۳۳۳	۰/۲	۰/۲	۵	۱	۰/۲	۰/۳۳۳	۰/۲	۰/۲	۰/۲	۳
دلیل مراجعه	۵	۰/۳۳۳	۵	۱	۰/۲	۰/۲	۷	۵	۱	۳	۰/۳۳۳	۰/۲	۰/۳۳۳	۵
سن	۳	۰/۲	۳	۰/۳۳۳	۰/۲	۰/۲	۵	۳	۰/۳۳۳	۱	۰/۲	۰/۲	۰/۲	۳
بخش اورژانس	۵	۰/۲	۵	۰/۳۳۳	۰/۳۳۳	۰/۳۳۳	۵	۵	۳	۵	۱	۰/۳۳۳	۱	۵
ضربان نبض	۷	۰/۳۳۳	۷	۱	۱	۱	۷	۵	۵	۵	۳	۱	۵	۵
فشارخون سیستولیک	۵	۰/۲	۵	۱	۰/۳۳۳	۰/۳۳۳	۷	۵	۳	۵	۱	۰/۲	۱	۵
مراجعه قبلی	۱	۰/۲	۳	۰/۳۳۳	۰/۲	۰/۲	۵	۰/۳۳۳	۰/۲	۰/۳۳۳	۰/۲	۰/۲	۰/۲	۱

پس از محاسبات انجام شده، مقدار ویژه ماتریس مقایسات زوجی برابر با ۱۵/۵۷۴، شاخص سازگاری برابر با ۰/۱۲۱ و نسبت سازگاری برابر ۰/۰۷۷ می‌باشد که نشان دهنده سازگار بودن ماتریس می‌باشد. با محاسبات انجام شده، مقدار اوزان هر ویژگی، به‌دست آمده که در جدول ۵ آورده شده و همچنین رتبه‌بندی نهایی نیز قابل مشاهده می‌باشد.

جدول ۵: وزن‌های به‌دست‌آمده

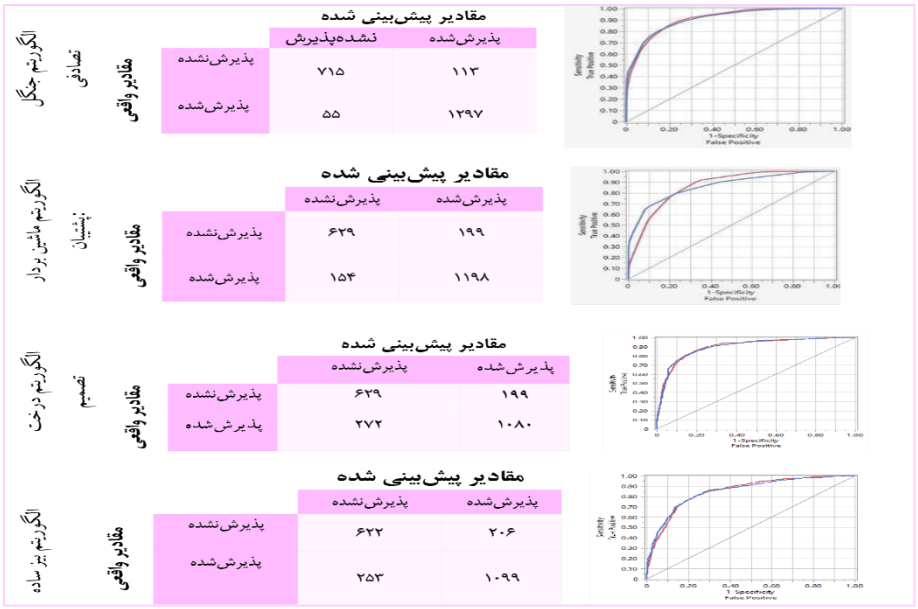
متغیر	وزن	رتبه
سطح تریاژ	۰/۱۶۶	۱
تعداد تنفس	۰/۱۵۳	۲
درصد اشباع اکسیژن	۰/۱۵۳	۲
ضربان نبض	۰/۱۱۸	۳
دمای بدن	۰/۰۷۷	۴
فشارخون سیستولیک	۰/۰۵۸	۵
بخش اورژانس	۰/۰۵۶	۶
دلیل مراجعه	۰/۰۵۵	۷
سن	۰/۰۴۱	۸
فشارخون دیاستولیک	۰/۰۳۶	۹
نحوه مراجعه	۰/۰۲۹	۱۰
مراجعه قبلی	۰/۰۲۳	۱۱
جنسیت	۰/۰۱۸	۱۲
ماه مراجعه	۰/۰۱۷	۱۳

از بین ۱۴ ویژگی موجود با مشورت خبرگان ۹ ویژگی اول به‌عنوان ویژگی‌های منتخب در نظر گرفته شده‌اند که در جدول ۶ لیست نهایی ویژگی‌های انتخابی همراه با ارزش متغیرها (بازه مقداری نرمال شده آن‌ها را نشان می‌دهد)، مشاهده می‌شود.

جدول ۶: جدول ویژگی‌های مدل‌های پیش‌بینی

متغیرها	ارزش متغیرها	نوع متغیر
پذیرش در بیمارستان[1]	۱ ۰	نامینال(دسته‌ای)
سن	[۰-۱]	نومریک(عددی)
سطح تریاژ	۲-۵	نامینال(دسته‌ای)
بخش اورژانس	۰ ۱	نامینال(دسته‌ای)
تعداد تنفس	[۰-۱]	نومریک(عددی)
ضربان نبض	[۰-۱]	نومریک(عددی)
دمای بدن	[۰-۱]	نومریک(عددی)
فشارخون سیستولیک	[۰-۱]	نومریک(عددی)
درصد اشباع اکسیژن	[۰-۱]	نومریک(عددی)
دلیل مراجعه	۲-۵	نامینال(دسته‌ای)

در ادامه در شکل ۲ مقادیر ماتریس درهم ریختگی و منحنی ROC را برای ۹ ویژگی منتخب نشان می‌دهد.



شکل ۲: مقادیر ماتریس درهم‌ریختگی و منحنی ROC

الگوریتم‌های مختلف، مورد بررسی قرار گرفته است و با توجه به داده‌های موجود و شاخص‌های ارزیابی مدل‌ها، نتایج مدل‌های استفاده شده برای ویژگی‌های منتخب در جدول ۷ قابل مشاهده است. که همانطور که ملاحظه می‌کنید الگوریتم جنگل تصادفی با دقت ۹۲/۲ درصد بهترین عملکرد را در بین الگوریتم‌ها داشته است.

جدول ۷ : جمع‌بندی نتایج الگوریتم‌های مورداستفاده با ویژگی‌های منتخب

مدل	شاخص ارزیابی			
	دقت	حساسیت	صحت	سطح زیر منحنی ROC
بیز ساده	۰/۷۸۹	۰/۸۱۲	۰/۷۵۱	۰/۸۵۱
ماشین بردار پشتیبان	۰/۸۳۸	۰/۸۸۶	۰/۷۵۹	۰/۸۹۹
درخت تصمیم	۰/۷۸۳	۰/۷۹۸	۰/۷۵۹	۰/۸۶۳
جنگل تصادفی	۰/۹۲۲	۰/۹۵۹	۰/۸۶۳	۰/۹۱۰

بحث

در این پژوهش با استفاده از روش‌های داده‌کاوی مدلی جهت پیش‌بینی پذیرش بیماران توسعه داده شده است .

داده‌های بالینی و غیربالینی بیماران، به عنوان ورودی مدل‌ها قرار گرفته‌اند و خروجی به صورت دسته‌بندی دو سطحی (پذیرش یا عدم پذیرش) مشخص شده است که الگوریتم جنگل تصادفی بیش‌ترین دقت را داشته است.

ضرورت شناسایی متغیرهای موثر در مدل‌های پیش‌بینی: با توجه به نتایج موجود در بخش یافته‌ها، استفاده از تکنیک مقایسات زوجی که از نظر خبرگان استفاده می‌کند منجر به استخراج ویژگی‌های منتخب شده و دقت مدل‌ها را افزایش داده است که مطالعه مروری Fernandes و همکاران (۱۵)، تایید کننده ویژگی‌های انتخاب شده توسط نظر خبرگان در این پژوهش بوده است. بنابراین بهتر است متخصصین بالینی نیز توجه بیشتری به این ویژگی‌ها داشته باشند.

مقایسه سطح زیر منحنی مدل‌ها: همانطور که بیان شد، هرچه سطح زیر منحنی ROC بیشتر باشد دقت مدل در تفکیک مقادیر بالاتر خواهد بود. با توجه به جدول ۶ مدل جنگل تصادفی با AUC برابر با ۰/۹۱۰ بالاترین قدرت تفکیک‌پذیری را در بین الگوریتم‌ها دارد. در مقاله Mowbray و همکاران (۱۰)، الگوریتم جنگل تصادفی دارای AUC 0.72 بوده است که کم بودن دقت مدل می‌تواند به دلیل تمرکز بیشتر مقاله Mowbray و همکاران (۱۰) بر پذیرش بیماران مسن باشد. در مقاله De Hond و همکاران (۱۱)، الگوریتم جنگل تصادفی با AUC 0.85 بوده که دلیل آن می‌تواند حاکی از آن باشد که داده‌های جمع‌آوری شده مربوط به سه بیمارستان می‌باشد که هر کدام از بیمارستان بار کاری و فرایند پذیرش متفاوتی دارند.

مقایسه معیار دقت مدل‌ها: دقت مدل جنگل تصادفی در این پژوهش ۰/۹۲۲ می‌باشد که به طور کلی نمایانگر میزان پیش‌بینی درست در بین کل داده‌ها است. که با توجه به ادبیات موضوع، پژوهش Boyle و همکاران (۱۶) دارای دقت‌ها ۰/۹۳ می‌باشند که دلیل بالاتر بودن آن را می‌توان تعداد بسیار کم ویژگی‌های (۳ عدد) این پژوهش دانست و یا در پژوهش Mowbray و همکاران (۱۱) دقت مدل جنگل تصادفی ۰/۷۹ می‌باشد که دلیل آن می‌تواند تعداد بسیار زیاد ویژگی‌های (۱۵ عدد) باشد و در مقاله نیز نشانه‌هایی از بیش برازش وجود دارد. بنابراین می‌توان نتیجه گرفت رابطه معنی‌داری بین تعداد ویژگی و دقت مدل‌ها وجود دارد و به طور کلی هر چقدر تعداد ویژگی‌ها بیشتر شود، دقت مدل کاهش می‌یابد و این نشان می‌دهد که تعداد ویژگی‌ها از اهمیت بالایی برخوردار است و دقت بالا مدل‌های این پژوهش نشان می‌دهد که نظر خبرگان جهت انتخاب ویژگی‌های برتر موثر واقع شده است .

مقایسه سایر معیارهای ارزیابی: جدول ۶ نشان می‌دهد که سایر معیارهای ارزیابی مدل جنگل تصادفی نسبت به سه مدل دیگر بیشتر می‌باشد. میزان حساسیت۰/۹۵۹ می‌باشد یعنی این مدل با احتمال ۹۶ درصد پیش‌بینی بیمارانی که پذیرش شده‌اند را به درستی انجام داده است و همچنین صحت ۰/۸۶۳ بدین معناست که مدل با احتمال ۸۶ درصد بیمارانی که پذیرش نشده‌اند را درست تشخیص داده است. که با توجه به جدول ۶ کلیه مدل‌ها دارای مقدار صحت پایینی می‌باشند که می‌تواند به دلیل کم بودن پرونده بیماران پذیرش نشده باشد .

اعتبارسنجی متقابل داده‌ها: یکی از مشکلات رایج در مدل‌های یادگیری ماشین بیش برازش داده می‌باشد که این مسئله در مقالات Graham و همکاران (۸) و Mowbray و همکاران (۱۰) نیز اشاره شده است و به منظور رفع آن از تکنیک اعتبارسنجی متقابل (k-fold) با ۱۰ لایه استفاده شده است که در این مطالعه نیز مانند مطالعات پیشین این تکنیک به کار گرفته شده است تا از مشکل بیش برازش جلوگیری شود.

عدم دسترسی به تمام داده‌های بیماران به دلیل محرمانه بودن آنها، ثبت نشدن اطلاعات بیماران به طور کامل در فرم‌های تریاژ، سخت وقت‌گیر بودن جمع‌آوری داده‌ها به دلیل الکترونیکی نبودن و همچنین ثبت نشدن ساعت مراجعه بیماران که الگوی کاری بخش اورژانس را نشان می‌دهد از محدودیت‌های این پژوهش به شمار می‌رود.

نتیجه‌گیری

نتایج نشان می‌دهد که با استفاده از الگوریتم‌های یادگیری ماشین و ویژگی‌هایی که توسط خبرگان انتخاب شده‌اند، مدل‌های با دقت بالاتری بدست می‌آید. در این مطالعه دقت الگوریتم جنگل تصادفی بیشتر از سایر الگوریتم‌ها است. در مقایسه با

مطالعات انجام شده، بر اساس نوع و ماهیت متغیرها و همچنین مطالعه موردی، الگوریتم‌های مختلف می‌تواند رفتارهای متفاوتی نشان دهند. بنابراین بهتر است که با توجه به جامعه و هدفی که مد نظر است به بررسی الگوریتم‌های مختلف پرداخته شود تا بهترین و دقیق‌ترین نتایج گرفته شود.

پیشنهادها

در تعیین پارامترهای مربوط به مدل‌ها می‌توان از مباحث ارائه‌شده در موضوعات عدم قطعیت و فازی نیز استفاده برد و همچنین می‌وان با در نظر گرفتن زمان مراجعه بیماران که الگوی بار کاری بخش اورژانس را نشان می‌دهد، مدل را در قالب یک مسئله سری زمانی تعریف و مورد توجه قرار داد.

تشکر و قدردانی

این مقاله حاصل پایان‌نامه دوره کارشناسی ارشد با شماره ۲۹۱۹۵۰۰ می‌باشد که با حمایت دانشگاه صنعتی اصفهان انجام شده است. نویسندگان بر خود الزم لازم می‌دانند ، از پرسنل بخش مدارک پزشکی بیمارستان امام علی شهرکرد، به خاطر مشارکت درگردآوری داده‌ها تشکر و قدردانی نمایند.

تضاد منافع

در انجام پژوهش حاضر، نویسندگان هیچ‌گونه تضاد منافعی نداشتند.

References

1. .Reinhardt UE. Does the aging of the population really drive the demand for health care? Health Affairs. 2003;22(6):27-39.
2. Grossman M. The demand for health: a theoretical and empirical investigation: Columbia university press; 2017.
3. Wilber ST, Gerson LW, Terrell KM, Carpenter CR, Shah MN, Heard K, et al. Geriatric emergency medicine and the 2006 Institute of Medicine reports from the Committee on the Future of Emergency Care in the US health system. Academic Emergency Medicine. 2006;13(12):1345-51.
4. Cameron P, Schull M, Cooke M. A framework for measuring quality in the emergency department. British Association for Accident and Emergency Medicine; 2011. p. 735-40.
5. Galarraga JE, Pines JM. Costs of ED episodes of care in the United States. The American journal of emergency medicine. 2016;34(3):357-65.
6. Hand DJ. Principles of data mining. Drug safety. 2007;30(7):621-2.
7. Lucini FR, Fogliatto FS, da Silveira GJ, Neyeloff JL, Anzanello MJ, Kuchenbecker RS, et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. International journal of medical informatics. 2017;100:1-8.
8. Graham B, Bond R, Quinn M, Mulvenna M. Using data mining to predict hospital admissions from the emergency department. IEEE Access. 2018 Feb 22;6:10458-69.
9. Parker CA, Liu N, Wu SX, Shen Y, Lam SSW, Ong MEH. Predicting hospital admission at the emergency department triage: A novel prediction model. The American journal of emergency medicine. 2019;37(8):1498-504.
10. Mowbray F, Zargoush M, Jones A, de Wit K, Costa A. Predicting hospital admission for older emergency department patients: Insights from machine learning. International Journal of Medical Informatics. 2020;140:104163.
11. De Hond A, Raven W, Schinkelshoek L, Gaakeer M, Ter Avest E, Sir O, et al. Machine learning for developing a prediction model of hospital admission of emergency department patients: Hype or hope? International journal of medical informatics. 2021;152:104496.
12. Menditto VG, Fulgenzi F, Bonifazi M, Gnudi U, Gennarini S, Mei F, et al. Predictors of readmission requiring hospitalization after discharge from emergency departments in patients with COVID-19. The American journal of emergency medicine. 2021;46:146-9.
13. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: Springer; 2013.
14. Larose DT. Data Mining Meyhods and Models. John Willey & Sons. Inc New Jersey. 2006.
15. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JM. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artificial Intelligence in Medicine. 2020;102:101762.
16. Boyle J, Jessup M, Crilly J, Green D, Lind J, Wallis M, et al. Predicting emergency department admissions. Emergency Medicine Journal. 2012;29(5):358-65.

Health Information Management

Vol 20/ No 4/ Winter 2023

Published by *Isfahan University of Medical Sciences*

DOI: 10.48305/him.2024.42198.1151

Predicting Emergency Department Admission Using Data Mining: Imam Ali Hospital in Shahrekord in Focus

Saba Paydar[1], GholamAli Reisi Ardali[2], Hossein Raeisi Dezaki[3]

Original Article

Abstract

Introduction: Today, hospitals face challenges such as overcrowding in the emergency ward and increased chaos and disruption in staff work, which leads to increased patient dissatisfaction. With the advancement of artificial intelligence and the expansion of data mining, predicting patient admission has become important. This study endeavors to predict patient admission in the emergency department of Imam Ali Hospital in Shahrekord.

Methods: In this study, 2180 patient records from the emergency ward of the hospital were examined. Initial patient information, including personal details, vital signs, and triage level which were recorded by nurses, were extracted. Using pairwise comparison matrix, the effective features were selected by experts. Then, using naive Bayes, decision tree, random forest, and support vector machine algorithms, the data were classified.

Results: Out of the 15 collected features, 9 features were selected by experts, and the results revealed that the random forest algorithm had the good performance in predicting patient admission in this case study, with an accuracy of 92/2%.

Conclusion: The results indicated that the accuracy of machine learning models increases with the use of expert opinions, and the random forest algorithm can predict patients' admission with high accuracy in this case study.

Keywords: Data Mining; Forecasting; Patient Admission; Emergency Service, Hospital

Received: Accepted: Published:

Citation: Paydar S, Reisi Ardali GA, Raeisi Dezaki H. **Predicting Emergency Department Admission Using Data Mining: Imam Ali Hospital in Shahrekord in Focus.** Health Inf Manage 2023; 20(4):190-197.

Article resulted from MSc thesis with No 2919500 support of Isfahan University of Technology.

1- Msc, Industrial Engineering, Department of Industrial and Systems Engineering Isfahan University of Technology, Isfahan. Iran

2- Associate Professor, Industrial Engineering , Department of Industrial and Systems Engineering, Isfahan University of Technology, Isfahan, Iran

3- MSc Student, Industrial Engineering, Department of Industrial and Systems Engineering, Isfahan University of Technology, Isfahan, Iran

Corresponding author: Saba Paydar, Msc, Industrial Engineering, Department of Industrial and Systems Engineering Isfahan University of Technology, Isfahan. Iran. Email: saba.p1374@gmail.com