Archive of SID

پیش‌بینی مدت اقامت بیماران در بخش مراقبت‌های ویژه مغز و اعصاب با تکنیک‌های داده‌کاوی

دکتر سعیده ناطقی‌نیا
دکترای مدیریت صنعتی و پژوهشگر گروه پژوهشی مدیریت سیستم های سلامت، بیمارستان لقمان حکیم، دانشگاه علوم پزشکی شهید بهشتی
دکتر محمدرضا حاجی اسماعیلی
استادیار بیهوشی، بیمارستان لقمان حکیم، دانشگاه علوم پزشکی شهید بهشتی
امید شفق سرخ[1]
دانشجوی دکترای مدیریت فناوری اطلاعات و پژوهشگر گروه پژوهشی مدیریت سیستم های سلامت، بیمارستان لقمان حکیم، دانشگاه علوم پزشکی شهید بهشتی
دکتر رضا گوهرانی
استادیار بیهوشی، بیمارستان لقمان حکیم، دانشگاه علوم پزشکی شهید بهشتی
دکتر مسعود زنگی
متخصص بیهوشی، بیمارستان لقمان حکیم، دانشگاه علوم پزشکی شهید بهشتی
دکتر حسنعلی احمدی
دستیار بیهوشی، بیمارستان لقمان حکیم، دانشگاه علوم پزشکی شهید بهشتی

Prediction of the Length of Stay of Patients in the Neuro-Critical Care Unit Using Data Mining Techniques

Saeedeh **Nateghinia,** PhD
Mohammadreza **Hajiesmaeili,** MD
Omid **Shafagh Sorkh,** PhD candidate
Reza **Goharani,** MD
Masoud **Zangi,** MD
Hassanali **Ahmadi,** MD

ABSTRACT

Introduction: Today, cost reduction and resource planning play an important role in managing hospitals. Hospital's intensive care departments have the most expensive beds. The prediction model for length of stay is a tool for optimal management of beds and scarce resources in intensive care units. Our goal in this study is to provide models for predicting the patient length of stay in neuro-critical care unit using data mining techniques.

Materials and methods: In this study, RapidMiner data mining software was used for modeling in order to classify and construct the prediction model for patients admitted to the neuro-critical care unit of Loghman Hakim Hospital in Tehran. The data of this study were extracted from 592 patients admitted to the intensive care unit between 94-97. Artificial neural network, K-nearest neighbors, decision tree and random forest algorithms used to classify the patients. Finally, the confusion matrix was obtained to calculate accuracy.

[1]. نویسنده مسؤول: omidshafagh@gmail.com

www.SID.ir

Archive of SID

Results: The findings of this study indicate that the type of surgery and the pneumonia as a complication have the greatest impact on the length of ICU stay. Also, the accuracy of the algorithms used to construct the prediction model was: decision tree 84.28%, random forest 83.96%, artificial neural network 83.79%, and the K-nearest neighbors, 77.90%.

Conclusion: Models of four techniques used in this study were able to predict the length of ICU stay. But the findings from the confusion matrix showed that the Decision trees with accuracy 84.28% had a relatively better performance among the techniques studied and the rules extracted from the decision tree could serve as a model for predicting the patient's stay in the neuro-critical care unit to be used.

Keywords: length of stay, data mining, intensive care unit, prediction model

چکیده

مقدمه: امروزه کاهش هزینه‌ها و برنامه‌ریزی منابع نقش بسیار مهمی در مدیریت بیمارستان‌ها ایفا می‌کنند. بخش مراقبت‌های ویژه بیمارستان‌ها دارای پرهزینه‌ترین تخت‌ها است. مدل‌های پیش‌بینی مدت اقامت بیماران در بخش مراقبت‌های ویژه ابزاری برای مدیریت بهینه تخت‌ها و منابع کمیاب بخش مراقبت‌های ویژه است. هدف ما در این پژوهش ارائه مدل‌هایی جهت پیش‌بینی مدت اقامت بیماران در بخش مراقبت‌های ویژه مغز و اعصاب با استفاده از تکنیک‌های داده‌کاوی است.

روش پژوهش: در این مطالعه از نرم‌افزار داده‌کاوی رپیدماینر برای مدلسازی به منظور طبقه‌بندی و ساخت مدل پیش‌بینی برای بیماران بستری در بخش مراقبت‌های ویژه مغز و اعصاب بیمارستان لقمان حکیم تهران استفاده شده است. داده‌های مورد نیاز این پژوهش از ۵۹۲ بیمار بستری در بخش مراقبت‌های ویژه بین سال های ۹۴ تا ۹۷ استخراج گردید. الگوریتم‌هایی که برای طبقه‌بندی بیماران استفاده گردید شامل شبکه عصبی، نزدیک‌ترین همسایه، درخت تصمیم و جنگل تصادفی بود. در نهایت ماتریس اغتشاش برای محاسبه دقت به دست آمد.

نتایج: یافته‌های این پژوهش نشان می‌دهد که متغیر نوع جراحی و عارضه پنومونی بیشترین تأثیر را بر مدت اقامت بیماران دارند. همچنین دقت الگوریتم‌های مورد استفاده برای ساخت مدل پیش‌بینی به ترتیب درخت تصمیم ۸۴,۲۸٪ ، جنگل تصادفی ۸۳,۹۶٪ ، شبکه عصبی ۸۳,۷۹٪ و نزدیک‌ترین همسایه ۷۷,۹۰٪ به دست آمد.

نتیجه‌گیری: مدل‌های ساخته شده بر اساس هر چهار تکنیک مورد استفاده در این پژوهش قادر به پیش‌بینی مدت اقامت بیمار بودند. اما یافته‌های حاصل از ماتریس اغتشاش نشان داد که تکنیک درخت تصمیم با دقت ۸۴,۲۸٪ عملکرد نسبتاً بهتری را در بین تکنیک‌های مورد مطالعه داشته و قوانین استخراج شده از درخت تصمیم می‌تواند به عنوان الگویی برای پیش‌بینی وضعیت مدت اقامت بیماران در بخش مراقبت‌های ویژه مورد استفاده قرار گیرد.

گل‌واژگان: مدت اقامت، داده‌کاوی، بخش مراقبت‌های ویژه، مدل پیش‌بینی

دکتر سعیده ناطقی‌نیا و همکاران

www.SID.ir

Archive of SID

مقدمه

مخارج صنعت بهداشت و درمان بخش قابل توجهی از تولید ناخالص ملی GDP در بسیاری از کشورها را تشکیل می دهد.(۱) این سهم بر طبق آمار بانک جهانی و سازمان بهداشت جهانی در طول سال‌های ۱۹۹۵ تا ۲۰۱۴ پیشرفت قابل توجهی داشته و به طور مثال اتحادیه اروپا ۱۰ درصد، خاورمیانه ۶,۴ درصد و در ایران ۶,۹ درصد تولید ناخالص ملی به صنعت بهداشت و درمان اختصاص یافته است.(۲، ۳) با وجود این بودجه دولتی قادر به تأمین تمام هزینه‌های بهداشت و درمان نبوده و از طرفی مؤسسات مراقبت بهداشتی با رشد تعداد بیماران مواجهند.(۴) از این رو برنامه‌ریزی کارا برای منابع و هزینه‌ها در بیمارستان‌ها نقشی حیاتی در مدیریت بیمارستان و بخش‌های آن ایفا می‌کند(۵) و همچنین این حوزه بسیار مورد توجه مدیران در سال های گذشته قرار گرفته است.(۶)

از جمله بخش‌های پرهزینه در بیمارستان بخش مراقبت‌های ویژه است. بخش مراقبت‌های ویژه کمتر از ۱۰ درصد تخت‌های بیمارستان را تشکیل می‌دهد در حالی که در حدود ۳۰ درصد از منابع و هزینه های بیمارستان را به خود اختصاص می‌دهد.(۷) هزینه‌های بالا به علت چندین فاکتور از جمله تعداد زیاد تجهیزات مورد استفاده، تعداد زیاد پزشک و پرستار، تنوع زیاد روش‌های درمانی و داروها و همچنین مدت اقامت بیمار است. همه این فاکتورها منجر به افزایش هزینه‌های روزانه در بخش مراقبت‌ها ی ویژه می‌شود.(۵)

با وجود اهمیت همه فاکتورهای مذکور مدت اقامت به عنوان یک شاخص سنجش میزان مصرف منابع در بخش مراقبت‌های ویژه به طور مکرر و روتین مورد استفاده قرار می‌گیرد هر چند تعریفی واحد از میزان مدت اقامت نرمال وجود ندارد اما مطالعات نشان می‌دهد که درصد کمی، در حدود ۷ تا ۱۱ درصد بیمارانی که در بخش مراقبت‌های ویژه بستری می‌شوند و اقامت طولانی دارند بخش اعظم منابع در حدود ۴۰ تا ۵۰ درصد را به خود اختصاص می‌دهند.(۸-۱۰)

شناسایی زودهنگام این بخش کوچک از بیماران و تخمین دقیق مدت اقامت آنها می‌تواند به مدیران در مدیریت بهتر تخت‌ها، اجتناب از هدر رفت زمان، سازمان‌دهی پرسنل مورد نیاز و کاهش هزینه‌ها کمک کند(۱۱, ۱۲) از این رو پیش‌بینی مدت اقامت بیماران امری دشوار اما ضروری است. با توجه به حجم زیاد داده‌هایی که هر روزه برای هر بیمار در بخش مراقبت‌های ویژه تولید می‌شود و با در نظر گرفتن اینکه داده‌های تولید شده در بخش مراقبت‌های ویژه بیش از ۲۵۰ پارامتر مختلف هستند، ساخت مدلی برای پیش‌بینی مدت اقامت نیازمند روش‌ها و تکنیک‌های به روزتری است.(۶، ۱۳-۱۵)

روش آماری استانداردی چون رگرسیون لجستیک از سوی کارشناسان مراقبت‌های ویژه برای پیش‌بینی ریسک مورتالیتی یا رویدادهای نامطلوب برای بیماران دچار ناراحتی‌ها یا آسیب‌های شدید که در بخش مراقبت‌های ویژه پذیرش شده‌اند به خوبی مورد استقبال قرار گرفته است. اما این پیش‌بینی‌ها به اندازه کافی برای یک بیمار دقیق نیستند و ابزاری برای پیش‌بینی مطمئن و به موقع برای یک بیمار خاص در شرایط مراقبت‌های ویژه وجود ندارد. در حال حاضر رویکردهای پیش‌بینی نوین با استفاده از الگوریتم‌های یادگیری ماشین

www.SID.ir

Archive of SID

مثل شبکه‌های عصبی مصنوعی و درختان تصمیم منتهی به ظهور تعدادی از مدل‌های پیش‌بینی در محیط‌های مختلف مراقبت بالینی شده است(۱۶).

داده‌کاوی[1] ما را به آخرین تکنیک‌ها و روش‌ها مجهز کرده که می‌تواند کمک هوشمندانه‌ای در انتقال حجم وسیعی از داده‌ها به اطلاعات و دانش مفید به ما انجام دهد. اطلاعات حاصل از داده‌کاوی به عنوان یک منبع مضاعفی از دانش برای پزشکان و مدیران بیمارستان‌ها در کشف خطاها و سوء استفاده‌های پزشکی، در شناسایی درمان‌های مؤثر و روش‌های بهتر و همچنین پیش‌بینی مدت اقامت بیماران مورد استفاده قرار گیرد.(۶)

در این پژوهش ما به دنبال ارائه مدل پیش‌بینی به عنوان راه حلی برای کاهش هزینه کرد منابع کمیاب برای بستری‌های غیر ضروری و اقامت طولانی مدت در بخش مراقبت‌های ویژه با استفاده از تکنیک‌های داده‌کاوی و مدلسازی جهت پیش‌بینی مدت اقامت بیماران در زمان پذیرش در بخش مراقبت‌های ویژه هستیم.

روش پژوهش

داده‌های این مطالعه از ۵۹۲ بیمار بستری در بخش مراقبت‌های ویژه مغز و اعصاب بیمارستان لقمان حکیم دانشگاه علوم پزشکی شهید بهشتی، در بازه زمانی مهر ۹۴ تا تیر ۹۷ جمع‌آوری شده است. این بیمارستان ۴۰۰ تخت خوابی دولتی، آموزشی و ریفرال است و بخش مراقبت‌های ویژه مغز و اعصاب دارای ۱۴ تخت اختصاصی برای بیماران مغز و اعصاب است. داده‌های جمع‌آوری شده از پرونده بیماران، متشکل از داده‌های دموگرافیک، مدت اقامت، بیماری‌های همزمان، نوع جراحی و عوارض انجام گرفته است.

در این پژوهش از نرم‌افزار رپیدماینر جهت داده‌کاوی استفاده گردید و فرایند داده‌کاوی شامل پیش پردازش داده، مدلسازی و تست مدل است.

پیش‌پردازش داده

در مرحله پیش‌پردازش داده‌ها، داده‌های ورودی جهت عملیات داده‌کاوی آماده‌سازی می‌شوند تا دقت مدل ساخته شده افزایش یابد. در این پژوهش نیز جهت آماده‌سازی داده‌ها ابتدا متغیر هدف تعیین و گسسته‌سازی گردید و داده‌های مفقوده مدیریت شدند. متغیر هدف در این پژوهش مدت اقامت بیماران در نظر گرفته شد و مدلسازی بر اساس این متغیر انجام گرفت و از آنجا که متغیر مدت اقامت بیماران در بخش مراقبت‌های ویژه مغز و اعصاب به صورت عددی ذخیره شده بود و متغیر کمی محسوب می‌شد ابتدا این متغیر گسسته‌سازی شد و بر این اساس به سه دسته مدت اقامت کوتاه، متوسط و بلندمدت تقسیم شد. سپس رکوردهایی که به صورت تکراری ثبت شده بودند شناسایی و حذف گردیدند و از آنجا که برخی از متغیرهای این پژوهش شامل مقادیر مفقوده بودند این مقادیر با روش‌های جایگزینی مدیریت شد.

مدلسازی:

پس از عملیات پیش‌پردازش داده‌ها، به منظور پیش‌بینی مدت اقامت بیماران از چهار الگوریتم شناخته شده طبقه‌بندی در حوزه سلامت که جزء روش‌های پیش‌بینی کننده داده‌کاوی هستند شامل

[1] . Data mining

www.SID.ir

Archive of SID

درخت تصمیم، نزدیک‌ترین همسایه، شبکه عصبی مصنوعی و جنگل تصادفی استفاده گردید تا مدل پیش‌بینی مدت اقامت بیماران در بخش مراقبت‌های ویژه مغز و اعصاب از بهترین الگوریتم و دقت بالاتری برخوردار باشد.

- الگوریتم درخت تصمیم[1]

درخت تصمیم یکی از مشهورترین و قدیمی‌ترین روش‌های ساخت مدل طبقه‌بندی است. در الگوریتم‌های طبقه‌بندی مبتنی بر درخت تصمیم، دانش خروجی به صورت یک درخت از حالات مختلف مقادیر و ویژگی‌ها ارائه می‌شود. نمایش دانش به شکل درخت سبب شده است که رده‌های مبتنی بر درخت تصمیم کاملاً قابل تفسیر باشند. لذا درخت‌های تصمیم‌گیری را می‌توان به صورت قواعد بیان کرد.(۱۷، ۱۸)

در این پژوهش وزن هر متغیر بر اساس متغیر هدف که مدت اقامت بیماران است با روش سود اطلاعاتی سنجیده شد تا تأثیر متغیرها روی متغیر هدف مشخص گردد. با این روش متغیری که بیشترین وزن را در ارتباط با متغیر هدف دارد در ریشه مدل قرار می‌گیرد و درخت تصمیم از این متغیر شروع می‌شود. مدل درخت تصمیم با حداکثر عمق ۲۰ ساخته شد.

- الگوریتم K نزدیکترین همسایه[2]

روش K نزدیک‌ترین همسایه یک گروه شامل K رکورد از مجموعه رکوردهای آموزشی[3] را که نزدیک‌ترین رکوردها به رکورد آزمایشی[4] باشند انتخاب کرده و بر اساس برتری رده یا برچسب مربوط به آن‌ها در مورد دسته رکورد آزمایشی مزبور تصمیم‌گیری می‌نماید. به عبارت ساده‌تر این روش رده‌ای را انتخاب می‌کند که در همسایگی انتخاب شده بیشترین تعداد رکورد منتسب به آن دسته باشند. بنابراین رده‌ای که از همه رده‌ها بیشتر در بین K نزدیک‌ترین همسایه مشاهده شود، به عنوان رده رکورد جدید در نظر گرفته می‌شود. (۱۹)

در این بخش مدلسازی با الگوریتم KNN تعداد همسایه‌ها ۱۵ همسایه در نظر گرفته شد و از روش رأی‌گیری وزنی و از فاصله اقلیدسی به عنوان معیار فاصله استفاده گردید.

- الگوریتم شبکه عصبی مصنوعی[5]

شبکه عصبی عبارت است از یک سیستم انطباقی که تعدادی عناصر پردازش ساده را شامل می‌شود و از شبکه اعصاب مغز الگوبرداری شده است. عناصر پردازش که در واقع همان نرون‌ها هستند، به هم می پیوندند تا یک مسیر پردازش را کامل کنند. این عناصر پردازش معمولاً در لایه‌ها با صفحات منظمی قرار گرفته‌اند، به طوری که بین لایه‌ها، ارتباطات کامل یا تصادفی وجود دارد. لایه ورودی به عنوان پردازشگر عمل می‌کند که پس از پردازش داده‌های ورودی، آنها را به شبکه ارائه می‌دهد. لایه ورودی لایه عصبی محاسباتی نیست، زیرا گره‌های آن نه وزن ورودی دارند نه تابع فعالسازی. لایه آخر، لایه خروجی است که خروجی شبکه را در پاسخ به یک ورودی مشخص نشان می‌دهد. لایه‌های قرار گرفته در وسط لایه پنهان نامیده می‌شود. این لایه‌ها به

[1] . Decision tree (DT)
[2] . k-nearest neighbors (K-NN)
[3] . Train
[4] . Test
[5] . Artificial neural networks (ANN)

www.SID.ir

Archive of SID

این دلیل پنهان نامیده می‌شود، چون بین آن و دنیای خروجی ارتباطی وجود ندارد.(۲۰)

در این بخش برای ساخت مدل با الگوریتم شبکه عصبی مصنوعی (ANN) یک لایه پنهان با ۱۲ نرون برای شبکه عصبی در نظر گرفته شد. در قسمت تنظیمات مدل، تعداد تکرار ۶۰۰، نرخ یادگیری ۰,۳ و مومنتوم ۰,۶ تنظیم شد. همچنین از روش Decay در این مدلسازی استفاده گردید که بدین معنی است که نرخ یادگیری به ازاء هر تکرار کمتر می‌شود.

- الگوریتم جنگل تصادفی[1]

جنگل‌های تصادفی یک نوع مدرن از روش‌های درخت پایه هستند که شامل انبوهی از درخت‌های کلاس‌بندی و رگرسیونی هستند. مهم‌ترین ویژگی جنگل‌های تصادفی عملکرد بالای آنها در اندازه گیری اهمیت متغیرها برای مشخص کردن اینکه هر متغیر چه نقشی در پیش‌بینی پاسخ دارد.(۲۱)

در این مطالعه مدلسازی با الگوریتم جنگل تصادفی با ۱۰ درخت ساخته می‌شود و درصد ویژگی‌های استفاده شده در هر درخت به صورت تصادفی برای هر درخت انتخاب می‌گردد.

اعتبارسنجی و تست مدل:

در این مرحله مدل‌های ساخته شده با روش 10-Fold Cross Validation اعتبارسنجی و ارزیابی شدند. این روش بدین معنی است که پس از ساخت مدل با داده‌های آموزشی برای سنجش اعتبار مدل از داده‌های آزمایشی یا و بدون برچسب استفاده می‌شود، در این روش ده بار و هر بار ده درصد از داده‌های آموزشی به صورت بدون برچسب به عنوان داده‌های آزمایشی مورد استفاده قرار می‌گیرند. در نهایت مدل تست شده و بر اساس نتایج به دست آمده از ماتریس اغتشاش دقت[2]، بازخوانی[3] و صحت[4] مدل محاسبه شد.

نتایج

داده‌های مورد استفاده جهت ساخت مدل پیش‌بینی باید در لحظه ورود بیمار یا درخواست پذیرش ثبت شده باشند، لذا در این پژوهش متغیرهای جدول ۲ برای ساخت مدل پیش‌بینی انتخاب گردیدند.

میانگین سن بیماران در این مطالعه ۴۳,۲ سال بود و از نظر جنسیت، ۵۸,۱٪ مرد و ۴۱,۹٪ زن بودند. اقامت کوتاه مدت با ۴۷,۸٪ بیشترین فراوانی را دارا بود و اقامت طولانی مدت با ۱۳٪ کمترین فراوانی را داشت. وزن‌دهی به متغیرها بر اساس متغیر هدف که مدت اقامت است با روش سود اطلاعاتی انجام گرفت که نوع جراحی و عارضه پنومونی بیشترین تأثیر را بر مدت اقامت بیماران دارند. (شکل ۱) همچنین فراوانی انواع اعمال جراحی در شکل ۲ آورده شده که تومور مغزی بیشترین تعداد عمل جراحی را به خود اختصاص داده است.

بعد از پیش‌پردازش داده‌ها و انجام مقدمات، مدلسازی برای چهار الگوریتم درخت تصمیم، K نزدیک‌ترین همسایه، شبکه عصبی مصنوعی و جنگل تصادفی انجام گرفت و عملکرد آن‌ها مورد ارزیابی قرار گرفت. سپس تمامی مدل‌ها با 10-Fold Cross Validation اعتبارسنجی و ارزیابی شد.

[1] . Random forest (RF)
[2] . Accuracy
[3] . Recall
[4] . Precision

www.SID.ir

Archive of SID

همچنین برای هر مدل تنظیم پارامتر با دقت بالایی انجام گرفت و ترکیب بهترین مقادیر پارامترها برای آن‌ها مشخص و تنظیم شد.

مقایسه صحت و بازخوانی هر یک از الگوریتم‌های چهارگانه در جدول ۳ نشان می‌دهد که مدلسازی با استفاده از الگوریتم درخت تصمیم دقت بالاتری داشته و داده‌ها را با کیفیت بهتری نسبت به سایر روش‌ها طبقه‌بندی کرده است. لذا الگوریتم درخت تصمیم با دقت ۸۴,۲۷٪ عملکرد بهتری برای ساخت مدل پیش‌بینی مدت مدت اقامت بیماران دارد.

جدول۱: دسته بندی مدت اقامت بیماران در بخش مراقبت‌های ویژه مغز و اعصاب

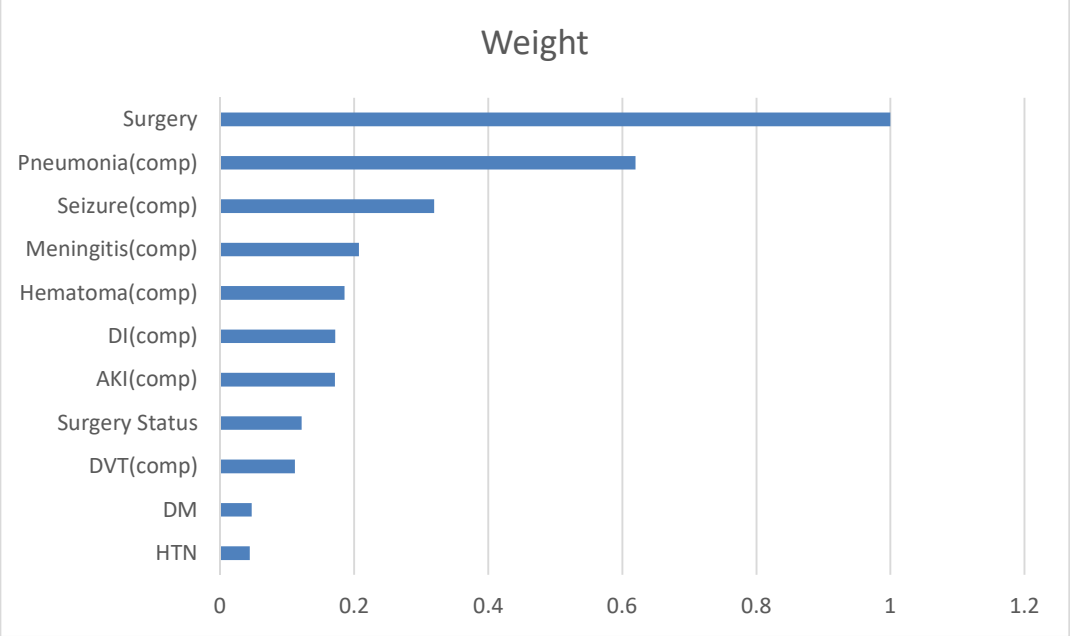
Group 1	<Mean	Short ICU Stay	<4 Day
Group 2	Mean + SD	Normal ICU Stay	4-7 Day
Group 3	>Mean + 2SD	Prolonged ICU Stay	>7 Day

جدول ۲: متغیرهای مورد استفاده در ساخت مدل پیش‌بینی

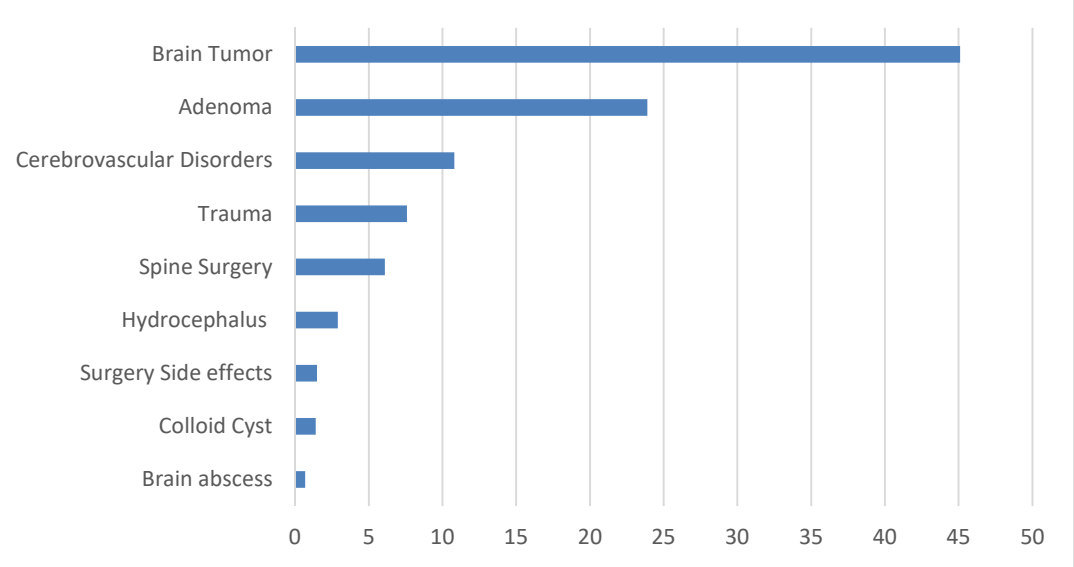
Demographic	**Coagulopathy (Y/N)**	**Complications**
Age (Y)	Corticosteroids (Y/N)	Pneumonia (Y/N)
Sex (M/F)	HIV+ (Y/N)	MI (Y/N)
Weight (Kg)	Immunodeficiency (Y/N)	Infection (Y/N)
Comorbidity	**Other**	Seizure (Y/N)
DM (Y/N)	ICU Stay (Days)	Meningitis (Y/N)
HTN (Y/N)	ICU Re-Ad (Y/N)	DVT (Y/N)
IHD (Y/N)	Re-Operation (Y/N)	AKI (Y/N)
Pneumonia (Y/N)	Cigarette/Alcohol / Drugs (Y/N)	DI (Y/N)
Asthma (Y/N)	Radio-Chemotherapy (Y/N)	PTE (Y/N)
Chronic Renal Failure (Y/N)	**Surgery**	Hematoma (Y/N)
PTE/DVT (Y/N)	Surgery Status (Elective/Emergency)	CSF Leak (Y/N)
COPD (Y/N)	Surgery (Type of Surgery)	

www.SID.ir

Archive of SID



شکل ۱: وزن‌دهی متغیرها بر اساس متغیر هدف با روش سود اطلاعاتی



شکل ۲: نمودار درصد فراوانی انواع اعمال جراحی

www.SID.ir

Archive of SID

جدول ۳: مقایسه دقت الگوریتم‌های مورد استفاده

الگوریتم	دقت	بازخوانی	صحت
درخت تصمیم	۸۴,۲۷	۸۰,۷۹	۸۵,۴۵
K نزدیک‌ترین همسایه	۷۷,۹۰	۷۳,۳۲	۸۱,۵۲
شبکه عصبی	۸۳,۷۹	۸۰,۸۰	۸۴,۹۷
جنگل تصادفی	۸۳,۹۶	۸۰,۱۲	۸۵,۲۲

بحث و نتیجه‌گیری

با افزایش مدوام هزینه‌های مراقبت بهداشتی و بدتر شدن وضعیت اقتصاد جهانی، مدیریت هزینه‌ها و منابع در مراقبت بهداشتی نقش مهمی ایفا می‌کنند. و این حقیقت در بخش مراقبت‌های ویژه به علت هزینه‌های هر بیمار نمود بیشتری دارد.

بیماران با مدت اقامت بیش از حد در بخش مراقبت‌های ویژه نشان دهنده مصرف حجم وسیعی از منابع هستند و پیش‌بینی قابل اطمینان مدت اقامت بیماران می‌تواند منجر به تخصیص تخت بطور منطقی و همچنین بیشینه سازی منابع گردد(۵)

از آنجا که مدل‌های رایج برای پیش‌بینی ریسک در بخش مراقبت‌های ویژه که برای مقایسه ریسک مرگ و میر یا شدت بیماری در بیماران استفاده می شوند دقت کافی برای پیش‌بینی مدت اقامت و شانس زنده ماندن بیماران را ندارند و همچنین ابزار قابل اطمینانی برای پیش‌بینی عوارض بیماری نیز محسوب نمی‌شوند،(۱۴) تکنیک‌های داده‌کاوی می توانند علاوه بر تعیین متغیرهای مرتبط با متغیر وابسته، برای ایجاد مدل پیش‌بینی آن متغیر نیز مورد استفاده قرار گیرند(۱۳ و ۲۲)

در این مطالعه از چهار الگوریتم پیش‌بینی کننده در داده‌کاوی شامل الگوریتم‌های درخت تصمیم، نزدیک‌ترین همسایه، شبکه عصبی مصنوعی و جنگل تصادفی استفاده شد که الگوریتم درخت تصمیم با دقت ۸۴,۲۸٪ عملکرد بهتری را در پیش‌بینی مدت اقامت بیماران در بخش مراقبت‌های ویژه داشت.

در داده‌کاوی با استفاده از الگوریتم درخت تصمیم متغیرها بر اساس وزن‌دهی نسبت به متغیر هدف با روش سود اطلاعاتی در درخت جای گرفتند که متغیر نوع جراحی بیشترین ارتباط را با مدت اقامت بیمار داشت و در ریشه درخت قرار گرفت.

یکی از مشکلات عمده در مورد درختان تصمیم، واریانس بالا و تغییرات زیاد در آنها است. غالباً یک تغییر کوچک در داده‌ها می‌تواند منجر به انشعابات بسیار متفاوت در شاخه‌های درخت شود که تفسیر نتایج را تا حدودی مخاطره‌آمیز می‌کند. با وجود این استفاده از درخت تصمیم به دلیل سادگی در فهم و نمایش نتایج، می‌تواند روش مناسبی برای پیش‌بینی مدت اقامت بیماران باشد اما باید توجه داشت، نتایج چنین مطالعاتی بسته به داده‌های مورد استفاده از هر بیمارستان ممکن است با بیمارستان های دیگر متفاوت باشد. (۲۳) .

در این مورد تحقیقات بسیاری برای یافتن الگوریتم‌های مناسب داده‌کاوی انجام شده که می

www.SID.ir

Archive of SID

توانند اطلاعات مفید و الگوهای پنهان را از مجموعه داده‌های بزرگ به طور خودکار استخراج کنند. توو و همکاران (۲۴) در سال ۱۹۹۳ با استفاده از الگوریتم شبکه عصبی مصنوعی به عنوان یک ابزار پیش‌بینی در شناسایی بیمارانی که به دنبال جراحی قلب مدت اقامت طولانی‌تری در بیمارستان داشتند مدل پیش‌بینی با دقت ۶۹٪ ارائه کردند. آذری و همکاران (۱۱) در سال ۲۰۱۲ از یک رویکرد داده‌کاوی چند لایه برای پیش‌بینی مدت اقامت در بیمارستان با تکنیک‌های Bayesnet, SVM, JRIP, J48 استفاده کردند که تکنیک JRIP با دقت ۷۴٪ بهترین عملکرد را داشت و آنها نتیجه‌گیری کردند که استفاده از خوشه‌بندی در مرحله آموزش می‌تواند عملکرد تکنیک‌های طبقه‌بندی را تقویت کند. همچنین شیخ‌نیا (۲۵) در سال ۲۰۱۲ با مقایسه چهار الگوریتم نزدیک‌ترین همسایه، بیز ساده، درخت تصمیم و شبکه عصبی مصنوعی به این نتیجه رسید که هیچ‌کدام از این تکنیک‌ها به تنهایی توانایی پیش‌بینی معنادار مدت اقامت در بیمارستان را ندارند. لذا از یک الگوریتم ترکیبی برای پیش‌بینی مدت اقامت بیماران بر اساس سابقه پزشکی بیماران استفاده کرد. در سال ۲۰۱۳، رضایی و همکاران (۱۳) به بررسی عوامل مؤثر بر مدت اقامت در بیماران دچار بیماری شریان کرونری پرداختند. آنها از تکنیک‌های طبقه بندی شامل سه الگوریتم درخت تصمیم، ماشین بردار پشتیبان و شبکه عصبی مصنوعی استفاده کردند. یافته‌های آنها نشان داد که مدل ماشین بردار پشتیبان با دقت ۹۶,۴٪ بهترین مدل پیش‌بینی در مطالعه بود. ژنگ و همکاران (۲۶) در سال ۲۰۱۵ پذیرش مجدد در یک بخش مراقبت‌های ویژه را با استفاده از چهار الگوریتم شامل مدل شبکه عصبی، الگوریتم جنگل تصادفی و مدل ترکیبی و ماشین بردار پشتیبان پیش‌بینی کردند. نتایج آنها نشان داد که مدل پیش‌بینی ماشین بردار پشتیبان بهتر از سایر الگوریتم‌ها بود و از نظر دقت پیش‌بینی به ۷۸,۴٪ رسید. جیانگ و همکاران (۲۷) در سال ۲۰۱۰ با استفاده از چهار روش داده‌کاوی رگرسیون لجستیک، شبکه عصبی، درخت تصمیم و مدلسازی ترکیبی، تجزیه و تحلیل داده‌های بیماران ترخیص شده را برای میانگین مدت اقامت بر اساس متغیرهای ورودی مطالعه کردند. این مطالعه نشان داد که مدل ترکیبی بهترین یا مناسب‌ترین مدل بود و سن و بیماری‌های مزمن، پیش‌بینی کننده‌ترین شاخصه‌ها هستند.

با توجه به مطالعات انجام شده، محدودیت‌هایی که پژوهشگران در این گونه مطالعه‌ها با آن مواجه هستند پایین بودن کیفیت و کمیت داده‌های موجود در سیستم پرونده کامپیوتری بیماران بیمارستان‌های تحت مطالعه است. در مورد کیفیت داده‌ها به دلیل اینکه مقادیر اکثر متغیرها به صورت دستی وارد سیستم می‌شوند، خطاهای انسانی هنگام ثبت، از صحت و دقت داده‌ها می کاهند که تأثیر زیادی بر نتایج کار خواهند داشت. از لحاظ کمیت نیز بسیاری از فیلدهای داده‌ای موجود در پرونده کامپیوتری مثل «دلیل اصلی مراجعه»، «تشخیص اولیه و نهایی»، «بیماری های همراه» و غیره، تکمیل نمی‌شود که این امر نیز باعث عدم کارآیی داده‌های سیستم‌های اطلاعات بیمارستانی در حوزه‌های پژوهشی می‌شود ولی در این پژوهش داده‌ها به طور دقیق ثبت شده‌اند و پرونده‌خوانی توسط پزشک مراقبت‌های ویژه انجام

www.SID.ir

Archive of SID

شده است. از آنجا که داده‌های مـورد بررسـی از کیفیـت و کمیت کافی برخوردار هستند، نتایج به دست آمده نسبت به مطالعات مختلف غنی‌تر بوده و قابلیت تعمیم بیشتری بر نمونه‌های مشابه خود خواهد داشت. پیش‌بینی مدت اقامـت و شناسـایی فاکتورهای تأثیرگذار بر آن، ممکـن اسـت بـر اسـاس محـل، داده‌ها، و تکنیک داده‌کاوی مورد استفاده، متفاوت باشـد امـا به هرحال انجام چنین مطالعاتی می‌تواند با استفاده از داده‌های موجود در پرونده‌های بیماران و به کارگیری تکنیک‌های داده کاوی مثل درخت تصمیم، شبکه‌های عصبی، شبکه های بیـزی و غیره انجام گیرد. نتایج این مطالعات می‌تواند بـرای برنامه‌ریزی و تخصیص بهینه منابع در مراکز ارائه خدمات بهداشتی ابزار مناسبی باشد.

مدیران می‌توانند با برنامه‌ریزی‌های مناسب مثل افزایش پرسنل درمانی و آماده‌سازی تجهیزات مورد نیاز در بخش‌های دارای ازدحام بیشتر، در جهت کاهش مدت اقامت گام بردارند که باعث بهره‌وری بهینه از منابع موجود بیمارستانی خواهد شد. همچنین اقداماتی در جهت کاهش حضور بیماران مسن در بیمارستان مثل توسعه خدمات مراقبت در منزل و برنامه‌های خودمراقبتی می‌تواند به کاهش مدت اقامت و در پی آن کاهش ازدحام و هزینه‌های درمانی کمک کند. (۲۸)

تقدیر و تشکر: نویسندگان مراتب تقدیر و تشکر خود را از واحد توسعه تحقیقات بالینی بیمارستان لقمان حکیم، دانشگاه علوم پزشکی شهید بهشتی جهت پشتیبانی و همکاری در طول دوره مطالعه اعلام می دارند.

REFERENCE

1. Awad A, Bader–El–Den M, McNicholas J. Patient length of stay and mortality prediction: A survey. Health Services Management Research. 2017:0951484817696212.
2. World-Bank. World Bank 2017. Available from: https://data.worldbank.org/indicator/SH.XPD.TOTL.ZS.
3. WHO. Wolrd Health Organization 2017. Available from: http://www .who.int/gho/health_financing/total_ expenditure/en/.
4. Awad A, Bader-El-Den M, McNicholas J. Modeling and Predicting Patient Length of Stay: A Survey. 2016.
5. Veloso R, Portela F, Santos M, Machado JM, Abelha A, Silva Á, et al., editors. Real-time data mining models for predicting length of stay in intensive care units. KMIS 2014-International Conference on Knowledge Management and Information Sharing; 2014.
6. Seemab S, Qamar U. Pedicting patient's length of stay by mining hospital data. 2015.
7. Hunter A, Johnson L, Coustasse A. Reduction of intensive care unit length of stay: the case of early mobilization. The health care manager. 2014;33(2):128-35.
8. Kramer AA, Zimmerman JE. A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. BMC medical informatics and decision making. 2010;10(1):27.
9. Xiao J, Douglas D, Lee AH, Vemuri SR. A Delphi evaluation of the factors influencing length of stay in Australian hospitals. The International journal of health planning and management. 1997;12(3):207-18.
10. Ravangard R, Arab M, Zeraati H, Rashidian A, Akbarisari A, Niroomand N, et al. A study of patient length of stay in Tehran University of Medical Sciences' Obstetrics and Gynecology Specialty Hospital and its associated clinical and nonclinical factors . 2010.

www.SID.ir

Archive of SID

11. Azari A, Janeja VP, Mohseni A, editors. Predicting hospital length of stay (phlos): A multi-tiered data mining approach. Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on; 2012: IEEE.
12. Khajehali N, Alizadeh S. Extract critical factors affecting the length of hospital stay of pneumonia patient by data mining (case study: an Iranian hospital). Artificial Intelligence in Medicine. 2017.
13. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. Healthcare informatics research. 2013;19(2):121-9.
14. Guiza Grandas F, Fierens D, Ramon J, Blockeel H, Meyfroidt G, Bruynooghe M, et al., editors. Predictive data mining in intensive care. Proceedings of the 15th Annual Machine Learning Conference of Belgium and The Netherlands (BENELEARN); 2006.
15. Navaz AN, Mohammed E, Serhani MA, Zaki N, editors. The use of data mining techniques to predict mortality and length of stay in an ICU. Innovations in Information Technology (IIT), 2016 12th International Conference on; 2016: IEEE.
16. Kim S, Kim W, Park RW. A comparison of intensive care unit mortality prediction models through the use of data mining techniques. Healthcare informatics research. 2011;17(4):232-43.
17. Friedman JH, Kohavi R, Yun Y, editors. Lazy decision trees. AAAI/IAAI, Vol 1; 1996.
18. Rokach L, Maimon O. Data mining with decision trees: theory and applications: World Scientific; 2008.
19. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowledge and information systems. 2008;14(1):1-37.
20. Chou S-M, Lee T-S, Shao YE, Chen I-F. Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications. 2004;27(1):133-42.
21. Akar Ö, Güngör O. Classification of multispectral images using Random Forest algorithm. Journal of Geodesy and Geoinformation. 2013;1(2).
22. Huang JQ, Hooper PM, Marrie TJ. Factors associated with length of stay in hospital for suspected community-acquired pneumonia. Canadian respiratory journal. 2006;13(6):317-24.
23. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Biometrics. 2002.
24. Tu JV, Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Computers and biomedical research. 1993;26(3):220-9.
25. Sheikh-Nia S. An Investigation of Standard and Ensemble Based Classification Techniques for the Prediction of Hospitalization Duration 2012.
26. Zheng B, Zhang J, Yoon SW, Lam SS, Khasawneh M, Poranki S. Predictive modeling of hospital readmissions using metaheuristics and data mining. Expert Systems with Applications. 2015;42(20):7110-20.
27. Jiang X, Qu X, Davis LB, editors. Using Data Mining to Analyze Patient Discharge Data for an Urban Hospital. DMIN; 2010.
28. Mohammadebrahimi S, Bayati S, Mardani M, Karim H. Factors Associated with Patient Length of Stay, According to Sina Hospital's Admission Data-Mashhad. Iranian Journal of Medical Informatics. 2015;4(4).

www.SID.ir