

# شناسایی عوامل مرتبط با طول مدت اقامت بیمارستانی با استفاده از درخت تصمیم

حسام کریم<sup>۱</sup> / کبری اطمینانی<sup>۲</sup> / سید محمود تارا<sup>۳</sup> / مرضیه مردانی<sup>۴</sup>

چکیده

مقدمه: یکی از عوامل اصلی هزینه‌های بیمارستانی طول مدت اقامت می‌باشد که باعث محدودیت منابع در بیمارستان می‌گردد. شناسایی عوامل مرتبط با این شاخص جهت بهره بردای بهینه از منابع و کاهش مدت اقامت، می‌تواند مفید باشد. مطالعه حاضر با هدف بررسی عوامل مرتبط با طول مدت اقامت با استفاده از اعمال درخت تصمیم بر روی داده‌های پذیرش بیمارستان انجام شد.

روش کار: این مطالعه از نوع توصیفی گذشته نگر است. نمونه‌های مورد مطالعه از ۱۸۸۰۶۸ پرونده‌ی کامپیوتری بیمارستان بستری در تمام بخش‌های بیمارستان قائم(عج) مشهد شامل ۱۶ متغیر و در بازه‌ی زمانی ابتدای سال ۱۳۸۸ تا انتهای سال ۱۳۹۲ خورشیدی استخراج شد. پس از پاکسازی داده‌ها، جهت تعیین عوامل مرتبط با طول مدت اقامت، از تحلیل تک متغیره شامل تست‌های آماری کای دو و تحلیل واریانسیک طرفه و همچنین الگوریتم درخت تصمیم CHAID در نرم افزار SPSS نسخه ۱۹ استفاده گردید.

یافته‌ها: میانگین مدت اقامت در بیمارستان مورد مطالعه ۶/۵ با میانه چهار روز بوده است. از بین ۱۵ متغیر مستقل وارد شده به درخت تصمیم، ۱۱ متغیر شامل بخش بستری، سن، دلیل مراجعه، پزشک معالج، نحوه مراجعه، شغل، وضعیت تأهل، نوع محل زندگی، داشتن همراه، روز پذیرش در هفته، و نوع بیمه، به عنوان عوامل مرتبط با LOS شناخته شدند.

نتیجه گیری: نتایج تکنیک‌های داده کاوی می‌تواند بر اساس داده‌های ورودی متفاوت باشد، لذا جهت آگاهی از عوامل مرتبط با طول مدت اقامت لازم است داده‌های مربوط به هر بیمارستان به‌طور جداگانه وارد مدل شده و از نتایج به‌دست آمده، در همان بیمارستان استفاده شود.

کلید واژه‌ها: طول مدت اقامت، عوامل مرتبط، داده کاوی، درخت تصمیم

• وصول مقاله: ۹۳/۱۰/۱۶ • اصلاح نهایی: ۹۴/۰۴/۲۳ • پذیرش نهایی: ۹۴/۰۵/۱۴

۱. دانشجوی کارشناسی ارشد انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران؛ نویسنده مسئول (Karim.hesam@gmail.com)

۲. استادیار گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

۳. استادیار گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

۴. کارشناس مدارک پزشکی، بیمارستان شهدای دهقان، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

## مقدمه

یک عامل غیر مستقیم به کار رود و ابزار مناسبی برای برنامه‌ریزی و تخصیص بهینه‌ی منابع در مراکز ارائه خدمات بهداشتی باشد [۷]. با توجه به موارد یاد شده و علاقه‌ی زیاد بیمارستان‌ها به کنترل هزینه‌های خود، برای رسیدن به این هدف، یک راه حل، تعیین عوامل مرتبط با طول مدت اقامت و به دنبال آن تخمین مدت اقامت هر بیمار می‌باشد [۱]. تعیین عوامل مرتبط با طول مدت اقامت بیمارستانی به منظور توسعه استراتژی‌های بالقوه برای بهره بردای بهینه از ظرفیت تخت‌ها، می‌تواند مفید باشد [۸]. به عبارتی دیگر، تخمین مدت زمان اقامت بیماران، و شناخت عوامل مؤثر بر آن، برای پیش‌بینی و حل مشکلات قریب‌الوقوع عملیاتی که ممکن است منجر به فشارهای مالی و عملکردی در اورژانس یا بیمارستان شوند، حایز اهمیت است [۲].

اگر هنگام پذیرش بیماران، مدت اقامت آن‌ها درست تخمین زده نشود، ممکن است منابع مورد نیاز به طور مطلوب به هر بیمار تخصیص داده نشود، لذا جهت تخمین دقیق‌تر طول مدت اقامت هر بیمار در هنگام پذیرش، باید از عوامل بالینی و غیر بالینی مرتبط با مدت اقامت آگاهی داشت. مطالعات قبلی نشان داده‌اند که تعداد زیادی از عوامل بالینی و غیر بالینی، مدت اقامت را تحت تأثیر خود قرار می‌دهند [۱۹-۱]. این عوامل بسته به جمعیت مورد مطالعه، متفاوت می‌باشند [۹] با این وجود در حال حاضر در مورد عوامل تعیین کننده و مرتبط با طول مدت اقامت بیمار، یک توافق عمومی وجود ندارد [۱۰]. اما به‌طور کلی این عوامل را می‌توان به دو گروه اصلی تقسیم نمود: ۱- متغیرهای عرضه که مربوط به شرایط ارائه دهنده مراقبت هستند مانند فراهم بودن تخت، روش‌های پرداخت و سیاست‌های ترخیص بیمارستانی؛ ۲- متغیرهای تقاضا که مربوط به نیازها و شرایط بیماران می‌باشند مانند شدت بیماری، وجود بیماری‌های همراه و هزینه‌های مستقیم و غیر مستقیم [۴]. تعیین تمام این عوامل به‌صورت جامع، تقریباً غیر ممکن

بیمارستان‌ها به شدت دارای محدودیت تخت و منابع برای مراقبت از بیماران می‌باشند، و بسیاری از آن‌ها با فشار مالی زیادی روبرو هستند [۱]. آن‌ها در مواجهه با چالش‌های مختلف مانند افزایش تعداد بیماران و محدودیت‌های تخت بستری، کمبود پرستار و منابع مواجه می‌شوند که باعث افزایش هزینه و فشار کاری در بخش‌های مختلف بیمارستان می‌شود [۲]. بخش‌های عمومی بیمارستان باید دارای گردش بیمار با اقامت کوتاه مدت باشند [۳] زیرا کاهش مدت اقامت در بیمارستان، جهت افزایش درآمد، کاهش هزینه‌ها، کاهش نوسانات و تفاوت‌ها در اقدامات بالینی، افزایش کیفیت و افزایش سود فرصت‌هایی را فراهم می‌آورد [۴]. می‌توان گفت طول مدت اقامت بیمارستانی، عامل اصلی هزینه‌ی بیمارستان‌ها است [۵]. اگر بتوان بیماران با اقامت طولانی مدت در بیمارستان را بلافاصله پس از پذیرش شناسایی کرد، منابع مناسب جهت تلاش برای سرعت بخشیدن به مراقبت‌های بهداشتی در زمان اولیه، می‌تواند در دسترس قرار گیرد [۳]. در دسترس بودن تخت بیمارستانی با فاکتورهای محدودی در ارتباط می‌باشد. نرخ مراجعه مجدد و طول مدت اقامت از فاکتورهای اصلی هستند [۶]. طول مدت اقامت (Length of Stay) که به اختصار LOS نامیده می‌شود به عنوان تعداد روزهایی تعریف می‌گردد که یک بیمار در بیمارستان یا هر مرکز پزشکی دیگری بستری می‌شود [۱]. طول مدت اقامت یکی از مفیدترین شاخص‌های بیمارستانی است که از آن می‌توان برای اهدافی از قبیل مدیریت مراقبت‌های بیمارستانی، کنترل کیفیت، برنامه‌ریزی بیمارستان، تعیین میزان کارایی و میزان استفاده از منابع بیمارستانی، استفاده نمود [۴]. همچنین طول مدت اقامت برای برآورد منابع مصرفی و میزان بهره‌وری در داخل بیمارستان می‌تواند به عنوان

از منابع بیمارستانی هنگام ورود بیماران با ویژگی‌های بالینی و غیربالینی مختلف به بیمارستان، مفید باشد.

### روش کار

این مطالعه از نوع توصیفی گذشته نگر است که به روش مقطعی و با هدف شناسایی عوامل بالینی و غیر بالینی مرتبط با طول مدت اقامت بیماران بستری شده در بیمارستان انجام شده است.

نمونه‌های مورد مطالعه، پرونده‌های کامپیوتری بیماران موجود در قسمت پذیرش بیمارستان آموزشی قائم (عج) مشهد از ابتدای سال ۱۳۸۸ تا انتهای سال ۱۳۹۲ خورشیدی بوده است. تعداد کل داده‌های استخراج شده اولیه ۸۳۱۴۶۷ رکورد بود که شامل تمام پرونده‌های کامپیوتری موجود در بازه زمانی مورد نظر بود و از سیستم اطلاعات بیمارستانی یا (HIS: Hospital Information System) استخراج گردید. پرونده‌های استخراج شده شامل ۱۳ متغیر بود که بر اساس تاریخ پذیرش، سه متغیر دیگر شامل ماه پذیرش، روز پذیرش در ماه و روز پذیرش در هفته نیز به این متغیرها اضافه گردید (جدول ۱). در مرحله بعد، با استفاده از نرم افزار مدیریت بانک اطلاعاتی Microsoft SQL Server نسخه ۲۰۱۲ عملیات پاک‌سازی (Cleaning) و پیش پردازش داده‌ها (DataPreprocessing) انجام شد که در ادامه بیان شده است.

قبل از هر نوع عملیات تحلیل روی داده‌ها و داده کاوی، باید عملیاتی بر روی داده‌ها انجام شود تا داده‌های، یک دست و عاری از خطا، وارد مرحله تحلیل شوند. این عملیات، مرحله پاک‌سازی داده‌ها (Data Cleaning) در داده کاوی نامیده می‌شود. بر اساس معیارهای ورود و خروج تعیین شده در این مطالعه و با توجه به اینکه هدف از این مطالعه بررسی عوامل مرتبط با اقامت بیمارانی است که پس از تکمیل دوره مراقبت خود از بیمارستان ترخیص شده‌اند، اطلاعات بیمارانی که در حین بستری فوت کرده یا بنا بر دلایلی غیر از ترخیص

است، اما می‌توان فاکتورهای بالینی و غیر بالینی در دسترس هنگام ورود بیمار به بیمارستان را بررسی کرده و از بین آن‌ها فاکتورهای مرتبط با طول مدت اقامت را تعیین نمود.

برای شناسایی عوامل مرتبط با یک یا چند متغیر وابسته، می‌توان از روش‌های مختلف داده کاوی استفاده نمود. این روش‌ها علاوه بر شناسایی و تعیین فاکتورهای مورد نظر توانایی پیش بینی مقادیر متغیر پاسخ را نیز دارند. روش‌های داده کاوی از الگوریتم‌های متعدد یادگیری برای تولید مدل‌های پیش بینی استفاده می‌کنند که با توجه به برخی عوامل نظیر ویژگی‌های داده‌ها و حیطه کاربردی مدل‌ها، برخی الگوریتم‌ها نسبت به سایر آن‌ها در اولویت قرار می‌گیرند. در حوزه علوم پزشکی الگوریتم‌های ساخت درخت تصمیم به علت سادگی و فهم آسان‌تر نسبت به سایر الگوریتم‌های یادگیری ماشین محبوبیت بیشتری دارند. از ویژگی‌های بارز درخت‌های تصمیم که آن‌ها را از سایر روش‌های یادگیری ماشین متمایز نموده است، سادگی در فهم نتایج و خروجی قابل درک برای متخصصین هر حوزه‌ای است. همچنین می‌توان نتایج درخت تصمیم را به صورت قواعد منطقی اگر-آن‌گاه (if-then) بیان کرد که برای تصمیم‌گیری در مورد یک مساله خاص می‌تواند بسیار کمک‌کننده باشد.

با توجه به اینکه مطالعات مختلف فاکتورهای متفاوتی را به عنوان عوامل تأثیرگذار بر مدت اقامت بیمارستانی معرفی کرده‌اند [۱-۱۹]، هدف از مطالعه حاضر بررسی و شناسایی عوامل مرتبط با طول مدت اقامت در بیمارستان قائم (عج) مشهد بر اساس داده‌های پذیرش که در هنگام ورود بیمار به بیمارستان این داده‌ها در دسترس می‌باشد و در سیستم اطلاعات بیمارستانی ثبت می‌گردد با استفاده از تکنیک داده کاوی درخت تصمیم می‌باشد. نتایج این مطالعه می‌تواند در برنامه‌ریزی و استفاده بهینه

توسط پزشک مانند فوت یا ترخیص با میل شخصی، از بیمارستان خارج شده بودند، از نمونه کنار گذاشته شدند. در برخورد با داده‌های گم شده (Missing Values) به دلیل کیفی بودن متغیرهای مورد مطالعه و عدم امکان تخمین مقادیر آنها، فیلدهای خالی در هر رکورد بدون تغییر باقی گذاشته شد. رکوردهایی که تاریخ پذیرش یا ترخیص آنها نامعلوم بود، به طور کامل حذف گردیدند. پس از انجام عملیات فوق، با استفاده از متغیرهای تاریخ تولد و تاریخ پذیرش و ترخیص، فاکتورهای سن و طول مدت اقامت محاسبه گردیدند. شاخص طول مدت اقامت که متغیر وابسته در مطالعه می‌باشد از اختلاف تاریخ پذیرش و تاریخ ترخیص به دست آمد. با توجه به این که فراوانی بیماران با مدت اقامت بیش از ۶۰ روز بسیار پایین (کمتر از یک درصد از کل داده‌ها) بود، بازه‌ی طول مدت اقامت از صفر تا ۶۰ روز در نظر گرفته شد که طول مدت اقامت صفر، به معنای مواردی است که بیمار دقیقاً در همان تاریخ پذیرش، ترخیص شده است، به عبارتی دیگر، اقامت زیر ۲۴ ساعت به عنوان مدت اقامت صفر در نظر گرفته شد. مقادیر بعضی از متغیرهای اسمی مثل شغل و همراه که به صورت دستی وارد سیستم می‌شدند دارای ناهمگونی‌های بسیاری بودند که با بررسی تک تک مقادیر، تمام محتویات فیلدهای اطلاعاتی به صورت یک شکل در آمدند. به عنوان مثال در رکوردهای مختلف برای فیلد شغل، مقادیر «معلم، آموزگار و مدرس» ثبت شده بودند که تمام آنها با کلمه «معلم» به طور یکسان مقداردهی گردیدند. پس از عملیات پاک‌سازی داده‌ها، تعداد رکوردهای باقی مانده ۱۸۸۰۶۸ رکورد بود که برای عملیات تحلیل داده مورد استفاده قرار گرفت. مجوز استفاده از داده‌ها در این مطالعه به تایید کمیته اخلاق دانشگاه علوم پزشکی مشهد رسیده است. جدول (۱) متغیرهایی که در ایجاد درخت تصمیم به کار گرفته شدند را نشان می‌دهد.

جهت تحلیل داده‌های مورد مطالعه و تعیین ارتباط تک تک متغیرهای مستقل با متغیر وابسته از روش‌های آمار توصیفی شامل شاخص‌های مرکزی (میانگین، میانه، نما) و پراکندگی

(واریانس و انحراف معیار) و همچنین آمار استنباطی تست‌های آماري کای دو (Chi-Square) و تحلیل واریانس یک طرفه (One-way ANOVA) برای متغیرهای بیش از دو حالت استفاده گردید. برای تعیین تأثیر هم‌زمان متغیرهای مستقل بر یک متغیر وابسته، در مطالعات مختلف از روش‌های گوناگون آماري و داده کاوی از جمله رگرسیون، درخت تصمیم و شبکه‌های عصبی مصنوعی استفاده شده است. الگوریتم‌های متعددی برای ساخت درخت تصمیم وجود دارند که هر کدام با توجه به حجم و نوع داده‌ها مورد استفاده قرار می‌گیرند. دو الگوریتمی که بیشترین استفاده را داشته‌اند الگوریتم‌های (CART: Classification and Regression Tree) و (CHAID: Chi-squared Automatic Interaction Detector) می‌باشند [۲۱، ۲۰]. در این طرح جهت شناسایی عوامل مرتبط با طول مدت اقامت از الگوریتم درخت تصمیم CHAID استفاده شد. این الگوریتم قادر است درخت غیر دودویی ایجاد نماید به این مفهوم که هر نود درخت می‌تواند بیش از دو فرزند داشته باشد. همچنین متغیرهای هدف و ورودی مدل می‌تواند از نوع کمی پیوسته یا طبقه بندی شده باشد. این الگوریتم برای یافتن تقسیم بهینه هر گره (Node) در درخت از آزمون کای دو استفاده می‌نماید. مزیت الگوریتم CHAID، ایجاد درخت با بیش از دو فرزند در هر گره می‌باشد که به فهم درخت بیشتر کمک می‌کند. همچنین الگوریتم درخت تصمیم به کار رفته در مطالعه، علاوه بر نمایش متغیرهای تأثیر گذار بر مدت اقامت اهمیت و وزن هر کدام از آنها را نیز محاسبه و نمایش می‌دهد. میزان اهمیت و وزن هر کدام از متغیرهای تأثیر گذار بر پیش بینی مدت اقامت توسط الگوریتم محاسبه و در خروجی نرم افزار نشان داده می‌شود. هنگام استفاده از مدل، ۷۰ درصد از داده‌ها برای مرحله آموزش و ۳۰ درصد برای مرحله تست مورد استفاده قرار گرفت. در کلیه‌ی آزمون‌های این مطالعه سطح معنی داری کمتر از پنج درصد در نظر گرفته شد. تمام مراحل تحلیل داده‌ها به وسیله نرم افزار آماري و داده کاوی SPSS نسخه ۱۹ انجام گردید.

جدول ۱: متغیرهای مورد استفاده در ایجاد درخت تصمیم

ردیف	نام متغیر	مقیاس	دامنه مقادیر متغیر
۱	مدت اقامت (متغیر وابسته)	فاصله ای	از ۰ تا ۶۰ روز (اختلاف تاریخ پذیرش و ترخیص)
۲	سن	فاصله ای	از ۰ تا ۹۰ (اختلاف تاریخ تولد بیمار و تاریخ پذیرش)
۳	جنس	اسمی	مونث، مذکر
۴	وضعیت تأهل	اسمی	متأهل، مجرد، همسر فوت شده، مطلقه
۵	نوع محل زندگی	اسمی	روستایی، شهری، کارتن خواب، اتباع خارجی
۶	شغل	اسمی	شغل به صورت رشته ای ثبت شده است
۷	داشتن همراه	اسمی	نسبت همراه با بیمار مثل مادر، پدر، خواهر و غیره
۸	دلیل مراجعه	اسمی	تصادفی، سوختگی، دیالیز، اورژانس، پیوند کلیه، زایمان، روتین، سنگ شکن، اعزامی، آنژیوگرافی
۹	نحوه مراجعه	اسمی	خود بیمار، توسط همراه، آمبولانس، توسط پلیس، ارجاع بیمارستانی
۱۰	بخش بستری	اسمی	سوختگی، اورژانس، گوش و حلق و بینی، قلب، اطفال، داخلی، زنان و زایمان، جراحی و اتاق عمل، عروق، CCU، ICU، سایر بخش‌ها.
۱۱	پزشک معالج	اسمی	رزیدنت، متخصص
۱۲	نحوه پرداخت	اسمی	نقدی، چک، قسط، مددکاری
۱۳	نوع بیمه	اسمی	آزاد، خدمات درمانی، تامین اجتماعی، نیروهای مسلح و غیره
۱۴	ماه پذیرش	اسمی	از فروردین تا اسفند
۱۵	روز پذیرش در ماه	اسمی	[۱-۳۱]
۱۶	روز پذیرش در هفته	اسمی	از شنبه تا جمعه

### یافته ها

بر طبق آمار توصیفی، میانگین طول مدت اقامت در بیمارستان مورد مطالعه ۶/۵ با انحراف معیار ۷/۵ روز و میانه چهار روز بوده است. کل متغیرهای مورد مطالعه شامل ۱۵ متغیر مستقل و یک متغیر وابسته (طول مدت اقامت) بود که در تحلیل تک متغیره و درخت تصمیم مورد استفاده قرار گرفتند. بر اساس تحلیل تک متغیره از بین ۱۵ متغیر مستقل، ۱۳ متغیر به طور مجزا با طول مدت اقامت ارتباط معنی دار آماری داشتند؛ نتایج تحلیل تک متغیره که توسط آزمون کای-دو انجام شد در جدول (۲) نشان داده شده است. پس از دوره سنی کودکی و نوجوانی با بالا رفتن سن طول مدت اقامت نیز افزایش یافته است به عبارتی دیگر متغیر سن با مدت اقامت ارتباط مستقیم داشته است که از لحاظ آماری این ارتباط معنی دار بود ( $P < 0.001$ ). از نظر جنس تقریباً حدود ۵۱ درصد از بیماران مؤنث و ۴۹ درصد مذکر بودند. بر اساس نتایج آزمون کای دو، ارتباط

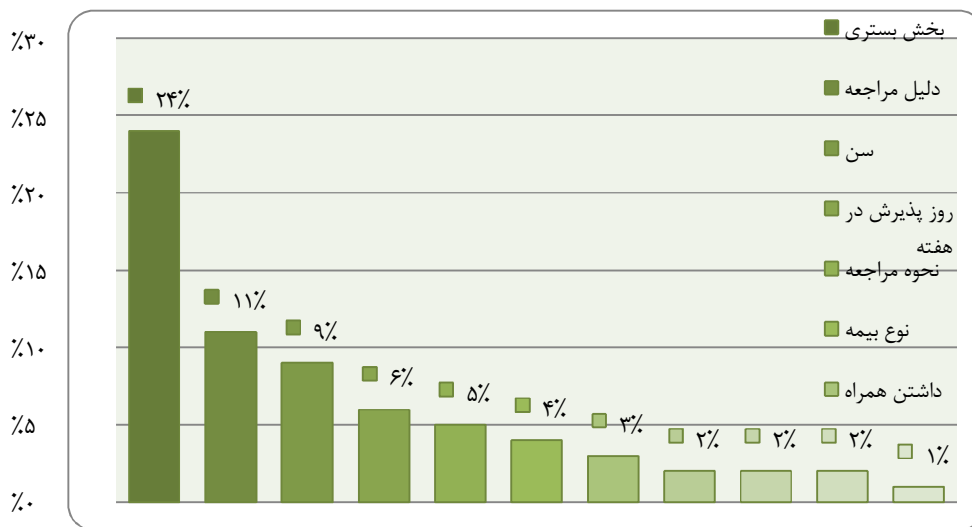
جنس با مدت اقامت از لحاظ آماری معنی دار بود ( $P < 0.001$ ). بر اساس یافته‌های مطالعه، بین دلیل مراجعه و مدت اقامت ارتباط معنی دار وجود داشت ( $P < 0.001$ ). مدت اقامت طولانی‌تر مربوط به پذیرش به دلیل سوختگی و پس از آن پیوند کلیه بود. بیشترین دلیل مراجعه (۵۴/۱ درصد) مربوط به موارد اورژانسی و کم‌ترین دلیل مراجعه (کمتر از ۰/۱ درصد) مربوط به دیالیز و پیوند کلیه بود. در تحلیل تک متغیره، ارتباط بین متغیر پذیرش بر اساس روز در هفته با طول مدت اقامت از لحاظ آماری معنی دار بود ( $P < 0.001$ ). بین طول مدت اقامت و نحوه‌ی پرداخت ارتباط معنی داری مشاهده نشد ( $P > 0.05$ ). در تحلیل تک متغیره، متغیرهای نوع بیمه، همراه، پزشک معالج، نوع محل زندگی، نوع ارجاع، شغل، بخش بستری، وضعیت تأهل و ماه پذیرش در سال با طول مدت اقامت ارتباط معنی داری داشتند ( $P < 0.001$ ). در مورد تأثیر هم زمان متغیرهای مستقل بر روی متغیر مدت اقامت، نتایج درخت تصمیم نشان داد که از بین ۱۵ متغیر

پزشک معالج، شغل، وضعیت تأهل و نوع محل زندگی بیمار به ترتیب بیشترین اهمیت را در تعیین طول مدت اقامت بیمار در بیمارستان داشته‌اند. همچنین متغیرهای جنس، ماه پذیرش، روز پذیرش در ماه و نحوه پرداخت تأثیری بر روی طول مدت اقامت نداشته‌اند.

مستقل وارد شده به مدل، ۱۱ متغیر بر طول مدت اقامت تأثیر دارند و می‌توانند برای پیش بینی این شاخص استفاده شوند. میزان اهمیت و وزن هر کدام از متغیرهای تأثیرگذار بر پیش بینی مدت اقامت در نمودار (۱) نشان داده است. بر اساس این نمودار متغیرهای بخش بستری، دلیل مراجعه، سن، روز پذیرش در هفته، نحوه مراجعه، نوع بیمه، داشتن همراه،

جدول ۲: نتایج آزمون کای-دو در تحلیل تک متغیره جهت تعیین فاکتورهای مرتبط با مدت اقامت

متغیر	مقدار کای دو	p-Value
جنس	۰/۲۵۷	۰/۷۹۸
نحوه پرداخت	۱/۴۳۴	۰/۱۵۲
روز پذیرش در ماه	۲/۶۲۰	۰/۰۴۹
داشتن همراه	۸/۳۰۵	<۰/۰۰۱
ماه پذیرش	۱۰/۳۷۸	<۰/۰۰۱
پزشک معالج	۲۰/۶۷۲	<۰/۰۰۱
نوع محل زندگی	۲۰/۸۷۰	<۰/۰۰۱
روز پذیرش در هفته	۳۸/۰۵۴	<۰/۰۰۱
شغل	۱۱۱/۱۱۱	<۰/۰۰۱
دلیل مراجعه	۱۱۳/۲۸۸	<۰/۰۰۱
وضعیت تأهل	۳۸۸/۷۸۹	<۰/۰۰۱
نوع بیمه	۴۱۷/۵۵۴	<۰/۰۰۱
نوع مراجعه	۵۵۸/۵۱۶	<۰/۰۰۱
سن	۵۹۱/۹۳۸	<۰/۰۰۱
بخش بستری	۳۹۸۴/۸۸۴	<۰/۰۰۱



نمودار ۱: میزان اهمیت متغیرهای تأثیرگذار در پیش بینی طول مدت اقامت بر اساس درخت تصمیم

معنادر گزارش شده است [۱۸-۱۲، ۵-۳، ۱]. یافته‌های مطالعه حاضر نشان داد مدت بستری کودکان تا سنین بین ۱۰ تا ۲۰ سال افزایش یافته، از سنین ۲۰ تا ۳۰ سالگی کاهش و مجدداً تا سنین پیری افزایش پیدا کرده است. این رابطه می‌تواند به دلیل این باشد که افراد مسن مستعد بیماری‌های مزمن هستند و نیز ممکن است مبتلا به بیماری‌های هم‌زمان باشند؛ همچنین در سنین کودکی و میان‌سالی به بالا، افراد در مراقبت از خود ناتوان هستند [۱۴]. در بین عوامل مورد بررسی، بخش بستری از ارتباط معنی‌دار زیادی با طول مدت اقامت بیماران بود. این ارتباط در مطالعات دیگر نیز به اثبات رسیده است [۴، ۱۰]. در این مطالعه بیشترین طول مدت اقامت بر اساس دلیل مراجعه، مربوط به سوختگی و پس از آن حوادث ناشی از تصادفات بوده است. در مطالعات مختلف عامل دلیل مراجعه به عنوان یکی از اصلی‌ترین عوامل مرتبط با مدت اقامت گزارش شده است [۳، ۴، ۱۰، ۱۲، ۱۳، ۱۵، ۱۹]. از نظر نحوه‌ی مراجعه، بیشترین مدت اقامت مربوط به بیمارانی مشاهده شد که توسط آمبولانس به بیمارستان منتقل شده بودند. در مطالعات دیگر نیز نحوه مراجعه با طول مدت اقامت ارتباط معنی‌داری داشته است [۴، ۱۰، ۱۴، ۱۵، ۱۸]. در مورد متغیر محل زندگی، علیرغم نتایج این مطالعه و همچنین نتایج مطالعات دیگر که ارتباط معنادر این عامل با مدت بستری را اثبات می‌کنند [۱۰، ۱۵، ۱۶، ۱۸]، در آمار توصیفی اختلاف چندانی بین میانگین طول مدت اقامت بیماران شهرنشین و روستایی ملاحظه نشد. با این وجود نتایج آمار توصیفی نشان می‌دهد بیمارانی که کارتن خواب بوده‌اند مدت اقامت بیشتری نسبت به سایرین داشته‌اند که این امر می‌تواند ناشی از عدم توانایی بیمار در پرداخت هزینه‌های بیمارستانی و همچنین، نداشتن انگیزه کافی جهت بازگشت به خانه باشد. همان‌طور که در مطالعات قبلی نیز به وجود ارتباط بین وضعیت تأهل و طول مدت اقامت اذعان شده است [۱۰، ۱۵، ۱۶، ۱۸] نتایج این مطالعه نیز وجود ارتباط معنی‌دار بین این دو عامل را نشان می‌دهد. بر اساس این نتایج، بیماران مجرد و متأهل نسبت به افراد مطلقه یا بیمارانی که همسرانشان فوت کرده‌اند، دارای طول مدت

افزایش جمعیت بیماران، کمبود منابع بیمارستانی و بالا بودن هزینه‌های درمانی باعث شده است هم بیماران و هم مدیران بیمارستانی در جهت کاهش مدت اقامت بیمارستانی تلاش نمایند. شاخص طول مدت اقامت که امروزه در بیمارستان‌ها به شکل قابل توجهی از آن استفاده می‌گردد، بیانگر میزان کارایی و عملکرد بیمارستانی می‌باشد. بنابراین، می‌تواند به عنوان یکی از معیارهای اندازه‌گیری و تحلیل عملکرد هر بیمارستان به کار رود [۱۱]. طول مدت اقامت متأثر از عوامل بالینی و غیر بالینی مختلفی است که می‌توان آن‌ها را به چهار دسته تقسیم کرد: ۱- مربوط به بیمار مانند سن، جنس، برخورداری از وضعیت اقتصادی-اجتماعی، وضعیت تغذیه‌ای و عفونت‌های بیمارستانی بعد از عمل؛ ۲- مربوط به بیمارستان مانند اندازه بیمارستان (تعداد تخت)، نسبت پرستار به بیمار و منطقه‌ای که بیمارستان در آن قرار دارد؛ ۳- منبع و نوع پرداخت مانند دارا بودن یا فقدان بیمه و نحوه پرداخت به بیمارستان؛ ۴- مربوط به پزشک مانند نوع فعالیت پزشک به صورت پزشک خانواده، پزشک خصوصی و یا پزشک مقیم بیمارستان [۱۰]. علاوه بر این، مدت اقامت می‌تواند تحت تأثیر فاکتورهای مختلف دیگری مثل سبک عملکرد کارکنان بهداشتی، در دسترس بودن تخت و فراهم بودن خدمات اجتماعی مانند سرویس‌های مراقبت طولانی مدت قرار گیرد. به هر حال درک بهتر از عوامل مؤثر بر مدت اقامت در بیمارستان باید به اقدامات لازم برای کاهش طول مدت اقامت منجر شود [۹].

بر اساس داده‌های بیمارستان مورد بررسی و یافته‌های این مطالعه، فاکتورهایی که به عنوان عوامل مرتبط با مدت اقامت تعیین شدند عبارتند از بخش بستری، دلیل مراجعه، سن، روز پذیرش در هفته، نحوه‌ی مراجعه، نوع بیمه، داشتن همراه، پزشک معالج، شغل، وضعیت تأهل و نوع محل زندگی بیمار. متغیر سن یکی از اصلی‌ترین فاکتورهای مربوط به بیمار است که در اغلب مطالعات ارتباط آن با مدت اقامت مستقیم و



در این مطالعه علاوه بر سن بیمار، بین فاکتورهای بخش بستری، علت مراجعه و نوع مراجعه و طول مدت اقامت ارتباط معنی‌دار مشاهده شد. تصمیم‌گیری در مورد این سه فاکتور برای کاهش طول مدت اقامت برعهده برنامه‌ریزان و مدیران بیمارستانی می‌باشد. آن‌ها می‌توانند با برنامه‌ریزی‌های مناسب مثل افزایش پرسنل درمانی و آماده‌سازی تجهیزات مورد نیاز در بخش‌های دارای ازدحام بیشتر، در جهت کاهش طول 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، می‌تواند به کاهش طول مدت اقامت و در پی آن کاهش ازدحام و هزینه‌های درمانی کمک کند. با توجه به تأثیر نوع بیمه بر طول مدت اقامت، بازنگری مدیران بیمه‌های درمانی در امر بازپرداخت هزینه‌ها با هدف کاهش مدت اقامت در بیمارستان، ضروری به نظر می‌رسد.

تکنیک‌های داده کاوی می‌توانند علاوه بر تعیین متغیرهای مرتبط با متغیر وابسته، برای ایجاد مدل پیش‌گویی آن متغیر نیز مورد استفاده قرار گیرند [۱]. به عبارت دیگر تکنیک‌های داده کاوی مورد استفاده علاوه بر تعیین فاکتورهای مرتبط با مدت اقامت توانایی پیش‌بینی این شاخص را نیز دارند اما این نوع از تحقیقات ذاتا دشوار هستند، و نتایج آن‌ها باید با احتیاط تفسیر شود [۹]. در مطالعه حاضر، از تکنیک درخت تصمیم برای بررسی عوامل مرتبط با طول مدت اقامت در بیمارستان مورد مطالعه، استفاده گردید. یکی از مشکلات عمده در مورد درختان تصمیم، واریانس بالا و تغییرات زیاد در آن‌هاست. غالبا یک تغییر کوچک در داده‌ها می‌تواند منجر به انشعابات بسیار متفاوت در شاخه‌های درخت شود که تفسیر نتایج را تا حدودی مخاطره آمیز می‌کند [۲۲]. با این وجود استفاده از درخت تصمیم به دلیل سادگی در فهم و نمایش نتایج، می‌تواند روش مناسبی برای چنین مطالعاتی باشد اما باید توجه داشت که بنا به دلایل ذکر شده، نتایج چنین مطالعاتی بسته به داده‌های مورد استفاده از هر بیمارستان ممکن است با بیمارستان‌های دیگر متفاوت باشد.

اقامت کمتری بوده‌اند. شغل از دیگر عواملی است که در این مطالعه به عنوان فاکتور مرتبط با مدت بستری شناسایی شد. معنادار بودن ارتباط این متغیر با مدت اقامت در دیگر مطالعات نیز به اثبات رسیده است [۱۰، ۱۴، ۱۵]. در این مطالعه بیشترین مدت زمان بستری مربوط به بیماران بازنشسته یا از کار افتاده، معلولین، زندانیان و افراد بی‌کار است. توجه این واقعیت احتمالا می‌تواند بر اساس دو علت باشد، نخست وجود کسالت و بیماری‌های همراه برای این گروه از افراد و دوم عدم انگیزه لازم برای بهبودی و برگشت به جامعه که این دو دلیل روند مراقبت و بهبودی بیماران را با تاخیر مواجه می‌سازد. بر طبق نتایج به دست آمده در این مطالعه، متغیر نوع بیمه با مدت اقامت افراد ارتباط معنی‌داری داشته است. به طوری که بیماران فاقد بیمه، کم‌ترین مدت اقامت را داشته‌اند و این امر شاید به دلیل عدم توانایی آنان در پرداخت هزینه‌های زیاد بیمارستان بوده است که علیرغم دستور پزشک مبنی بر نیاز به اقامت بیشتر بیمار در بیمارستان، اصرار بر ترخیص و ترک بیمارستان داشته‌اند. همچنین بیمارانی که بیمه نداشته‌اند ولی از طرف سازمان‌های خیریه مانند کمیته امداد معرفی‌نامه جهت تخفیف ۱۰۰ درصد ارائه داده‌اند، بیشترین مدت اقامت را نسبت به سایرین داشته‌اند که شاید دلیل این افزایش مدت اقامت به خاطر عدم توجه بیمار به هزینه‌های مراقبتی بوده است. وحیدی و همکارانش نیز در مطالعه خود به این واقعیت اشاره نموده‌اند [۱۴]. فاکتور دیگری که به عنوان عوامل مرتبط با طول مدت اقامت در این مطالعه تشخیص داده شد، روز پذیرش در هفته می‌باشد. آمار توصیفی مطالعه نشان می‌دهد بیمارانی که در روز شنبه پذیرش شده‌اند طول مدت اقامت بیشتری نسبت به سایر بیمارانی داشته‌اند که در روزهای دیگر هفته پذیرش شده‌اند. بررسی علت این واقعیت خود نیازمند مطالعات دیگری است. متغیرهای نوع پزشک معالج و داشتن همراه نیز بر اساس خروجی درخت تصمیم، با طول مدت اقامت در ارتباط بودند اما در آمار توصیفی تفاوتی بین میانگین طول مدت اقامت در گروه‌های مختلف این دو متغیر مشاهده نشد.



به هر حال انجام چنین مطالعاتی می‌تواند با استفاده از داده‌های موجود در پرونده‌های بیماران و به کارگیری تکنیک‌های داده کاوی مثل درخت تصمیم، شبکه‌های عصبی، شبکه‌های بیزی و غیره انجام گیرد. نتایج این مطالعات می‌تواند برای برنامه‌ریزی و تخصیص بهینه‌ی منابع در مراکز ارائه خدمات بهداشتی ابزار مناسبی باشد.

### تشکر و قدردانی

این مقاله حاصل (بخشی از) پایان نامه تحت عنوان «شناسایی عوامل مرتبط با طول مدت اقامت بیمارستانی بر مبنای اعمال تکنیک‌های داده کاوی بر داده‌های پذیرش» در مقطع کارشناسی ارشد در سال ۱۳۹۳ کد ۶۴۹-آ می‌باشد که با حمایت دانشگاه علوم پزشکی مشهد اجرا شده است. بر خود لازم می‌دانیم از مسئولین و پرسنل قسمت فناوری اطلاعات این دانشگاه و همچنین پرسنل بیمارستان آموزشی قائم (عج) مشهد که ما را در انجام این مطالعه یاری نمودند، قدردانی نماییم.

یکی از عواملی که می‌تواند در میزان کارایی مدل‌های پیش بینی در داده کاوی تأثیر داشته باشد، حجم نمونه‌های مورد استفاده است. هر چه تعداد رکوردهای وارد شده به مدل کمتر باشد، دقت مدل پیش بینی دست خوش تغییرات می‌شود [۲۰]. حجم بالای نمونه‌های مورد استفاده در این پژوهش که در نوع خود بی نظیر است، یکی از امتیازات این طرح نسبت به مطالعات مشابه می‌باشد. همچنین در اکثر مطالعات قبلی، عوامل مرتبط با طول مدت اقامت در یک بخش خاص از بیمارستان انجام شده است [۱-۱۹] در حالی که در پژوهش حاضر بررسی این موضوع در تمام بخش‌های بیمارستان، مورد مطالعه قرار گرفت که این نیز یکی دیگر از ویژگی‌های ممتاز این طرح می‌باشد. به هر حال از یافته‌های این مطالعه می‌توان نتیجه گرفت که عوامل مختلفی مانند حجم داده، نوع داده، کیفیت داده و وجود داده‌های گم شده در بین داده‌های مورد مطالعه می‌تواند بر نتایج تکنیک‌های داده کاوی از جمله درخت تصمیم تأثیر گذار باشد.

از محدودیت‌هایی که پژوهشگران این مطالعه با آن مواجه بودند، پایین بودن کیفیت و کمیت داده‌های موجود در سیستم پرونده‌ی کامپیوتری بیماران بیمارستان‌های تحت مطالعه بود. در مورد کیفیت داده‌ها به دلیل این که مقادیر اکثر متغیرها به صورت دستی وارد سیستم می‌شوند، خطاهای انسانی هنگام ثبت، از صحت و دقت داده‌ها می‌کاهند که تأثیر زیادی بر نتایج کار خواهند داشت. از لحاظ کمیت نیز بسیاری از فیله‌های داده‌ای موجود در پرونده کامپیوتری مثل «دلیل اصلی مراجعه»، «تشخیص اولیه و نهایی»، «بیماری‌های همراه» و غیره، تکمیل نمی‌شود که این امر نیز باعث عدم کارایی داده‌های سیستم‌های اطلاعات بیمارستانی در حوزه‌های پژوهشی می‌شود. به طور یقین اگر داده‌های مورد بررسی از کیفیت و کمیت کافی برخوردار باشند، نتایج به دست آمده در مطالعات مختلف غنی‌تر خواهد بود و قابلیت تعمیم بیشتری بر نمونه‌های مشابه خواهند داشت. پیش بینی طول مدت اقامت و شناسایی فاکتورهای تأثیر گذار بر آن، ممکن است بر اساس محل، داده‌ها، و تکنیک داده کاوی مورد استفاده، متفاوت باشد اما

## References

1. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients. *Healthcare Inform Res* 2013; 19(2):121–9.
2. Wrenn J, Jones I, Lanaghan K, Congdon CB, Aronsky D. Estimating Patient's Length of Stay in the Emergency Department with an Artificial Neural Network. *AMIA Annual Symposium Proceedings. American Medical Informatics Association*; 2005. p. 1155.
3. Maguire PA, Taylor IC, Stout RW. Elderly Patients in Acute Medical Wards: Factors Predicting Length of Stay in Hospital. *BMJ* 1986; 292(6530):1251-1253.
4. Ravangard R, Arab M, Rashidian A, Akbarisari A, Zarei A, Salesi M, et al. [Length of Stay in Gynecological Hospital in Tehran University of Medical Sciences and its Influencing Factors Using Survival Analysis]. *J Sch Public Health Inst Public Health Res* 2010; 8(3):25–35. [Persian]
5. Formiga F, Chivite D, Manito N, Mestre AR, Llopis F, Pujol R. Admission Characteristics Predicting Longer Length of Stay Among Elderly Patients Hospitalized for Decompensated Heart Failure. *Eur J Intern Med* 2008;19(3):198–202.
6. Zhang J, Harvey C, Andrew C. Factors Associated with Length of Stay and the Risk of Readmission in an Acute Psychiatric Inpatient Facility: a Retrospective Study. *Aust N Z J Psychiatry* 2011; 45(7): 578–85.
7. Soyiri IN, Reidpath DD, Sarran C. Asthma Length of Stay in Hospitals in London 2001–2006: Demographic, Diagnostic and Temporal Factors. *PloS One* 2011; 6(11): 27184.
8. Moloney ED, Bennett K, Silke B. Length of Hospital Stay, Diagnoses and Pattern of Investigation Following Emergency Admission to an Irish Teaching Hospital. *Ir Med J* 2004; 97(6):170–2.
9. Huang JQ, Hooper PM, Marrie TJ. Factors Associated with Length of Stay in Hospital For Suspected Community-Acquired Pneumonia. *Can Respir J J Can ThoracSoc* 2006;13(6):317.
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]. *Hakim Res J* 2010;13(2):129–36. [Persian]
11. Xiao J, Douglas D, Lee AH, Vemuri SR. A Delphi Evaluation of the Factors Influencing Length of Stay in Australian Hospitals. *Int J Health Plann Manage* 1997;12(3):207–18.
12. Appelros P. Prediction of Length of Stay for Stroke Patients. *ActaNeurolScand* 2007;116(1):15–9.
13. Clark DE, Ryan LM. Concurrent Prediction of Hospital Mortality and Length of Stay from Risk Factors on Admission. *Health Serv Res* 2002; 37(3): 631–45.
14. Vahidi R, Kushavar H, Khodayari R. [Factors Affecting Coronary Artery Patients Length of

- Stay on ShahidMadani Hospital of Tabriz-2006]. *J Health Adm* 2006; 9(25): 63–8. [Persian]
15. Arab M, Zarei A, Rahimi A, Rezaiean F, Akbari F. [Analysis of Factors Affecting Length of Stay in Public Hospitals in Lorestan Province, Iran]. *Hakim Res J* 2010;12(4): 27–32. [Persian]
16. Rafiei M, Ayatollahi SMT. [Modeling Length of Stay in Vali-Asr and Tamin-Ejtemaei Hospitals of Arak on Some Demographic Factors]. *Arak Med Univ J*. 5(4): 28–35. [Persian]
17. Gohari MR, Vahabi N, Moghadamifard Z. [Semi-parametric Cox Regression for Factors Affecting Hospitalization Length]. *Daneshvar* 2012;19(99):23-30. [Persian]
18. Rafiei M, Ayatollahi SMT, Behboodian J. [Comparison of Different Modeling Strategies of Length of Stay of Patients Hospitalized in Teaching Hospitals in Shiraz]. *Med J Tabriz Univ Med Sci Health Serv* 2007; 29(2): 43–51. [Persian]
19. Chang K-C, Tseng M-C, Weng H-H, Lin Y-H, Liou C-W, Tan T-Y. Prediction of Length of Stay of First-Ever Ischemic Stroke. *Stroke* 2002; 33(11): 2670–4.
20. Nabovati E, Azizi AA, Abbasi E, Vakili-Arki H, Zarei J, Razavi AR. [Using Data Mining to Predict Outcome in Burn Patients: A Comparison between Several Algorithms]. *HealthInf Manage* 2014; 10(6):799. [Persian]
21. Azizi AA, Zarei J, Nabovati E, Vakili-Arki H, Abbasi E, Razavi AR. [Determining of the Factors Affecting Mortality in Burn Patients Using a Decision Tree Data Mining Algorithm]. *Health Management* 2013; 16(54):34. [Persian]
22. Hastie T, Tibshirani R, Friedman J, Franklin J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. *Math Intell*; 27(2):83–5.



# Identifying Factors Associated with Length of Hospital Stay Using Decision Tree

Karim H<sup>1</sup>/Etminani K<sup>2</sup>/ Tara SM<sup>3</sup>/ Mardani M<sup>4</sup>

## Abstract

**Introduction:** One of the main factors of hospital costs resulting in hospital resources constraints is Length of Stay (LOS). Recognizing factors associated with this index can result in optimized utilization of resources and reducing LOS. The present study has been conducted with the aim of investigating LOS-related factors using decision tree applied on hospital admission data.

**Methods:** In this descriptive-retrospective study, 188068 patient data were extracted from computerized patient records including 16 variables in all wards in Ghaem Hospital in Mashhad, Iran, during 2009-2014. After data cleaning, in order to determine the factors associated with LOS, statistical procedures including univariate analysis, chi-Square statistical test, one-way ANOVA, and CHAID algorithm were applied on our data using SPSS Statistical Software version 19.

**Results:** The mean and median of LOS in Ghaem hospital were 6.5 and 4 days, respectively. Out of 15 independent variables entered the model, 11 variables including the ward, the patient's age, the admission reason, the doctor, the referral status, the patient's job, the marital status, the status of the residential place, the caregiver(s), admission day of the week, and the insurance type were associated with LOS.

**Conclusion:** The results of data mining techniques can be different based on the input data. Therefore, to identify the factors related to LOS, it is necessary to enter the related data of individual hospitals into the model and to use the obtained results at the same hospital.

**Keywords:** Length of Stay, Associated Factors, Data Mining, Decision Tree

• Received: 6/Jan/2015 • Modified: 14/July/2015 • Accepted: 5/Aug/2015

1. MSc Student of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; Corresponding Author(Karim. Hesam@gmail. com)

2. Assistant Professor of Medical Informatics Department, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

3. Assistant Professor of Medical Informatics Department, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

4. BSc of Medical Record, ShohadayeDehaghan Hospital, Isfahan University of Medical Sciences, Isfahan, Iran

