مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره پنجم شماره اول. ص ۲۴-۱۲، ۱۳۹۷

برازش مدل‌های هوش مصنوعی و آمار کلاسیک سری زمانی جهت پیش‌بینی تعداد بیماران بستری بیمارستان‌ها

سمیرا رسولی[۱]، حامد تابش[۲]، کبری اطمینانی[۳]*

• دریافت مقاله: ۹۶/۷/۹ • پذیرش مقاله: ۹۷/۲/۳

مقدمه: مطالعه و تجزیه‌وتحلیل هر سیستم بهداشتی و درمانی، یک ضرورت برای بهبود عملکرد آن در طول زمان به شمار می‌آید. در این میان، مدیریت و تجزیه‌وتحلیل تعداد بیماران یک عامل مهم و تعیین‌کننده در بهبود تصمیمات مدیران این حوزه است. هدف از این پژوهش، مطالعه‌و بررسی روش‌های پیش‌بینی مبتنی بر سری‌های‌زمانی جهت پیش‌بینی ماهیانه تعداد بیماران بستری و مقایسه صحت عملکرد این روش‌ها می‌باشد.

روش: در این مطالعه مقطعی مدل‌سازی بر اساس داده‌های ماهیانه تعداد بیماران بستری ۶ بیمارستان دولتی شهر مشهد از فروردین ۱۳۸۳ تا فروردین ۱۳۹۵ انجام‌گرفت که از پایگاه داده اداره آمار دانشگاه علوم‌پزشکی‌مشهد استخراج شد. جهت پیش‌بینی تعداد بیماران بستری سه‌ماهه نخست سال ۱۳۹۵ هر یک از بیمارستان‌ها، از تکنیک‌های پیش‌بینی Holt-Winters، SARIMA، MLP و GRNN استفاده گردید. برای هر مدل، خطای مقادیر پیش‌بینی‌شده توسط معیار میانگین قدرمطلق درصد خطا (MAPE) گزارش شد.

نتایج: روش Holt-Winters با ارائه بهترین نتایج پیش‌بینی برای ۴ بیمارستان می‌تواند روش کارآمدی برای پیش‌بینی تعداد بیماران بیمارستان‌ها باشد. در مجموع، مدل‌های پیش‌بینی مورد بررسی در این مطالعه با ارائه معیار صحت MAPE در بازه ۲/۱۳ درصد تا ۴/۱۲ درصد، عملکرد قابل‌قبولی برای هر ۶ بیمارستان داشته‌اند.

نتیجه‌گیری: در این مطالعه نشان‌داده‌شد که تجزیه‌وتحلیل سری‌های‌زمانی ابزاری مناسب و کاربردی برای پیش‌بینی تعداد بیماران بستری بیمارستان‌های مورد مطالعه می‌باشد. با توجه به ویژگی‌های منحصربه‌فرد بیمارستان‌های مختلف، فرایندهای ذکرشده در این پژوهش شامل مدل‌سازی و تجزیه‌وتحلیل داده‌ها را می‌توان در دیگر بیمارستان‌ها جهت بهبود تخصیص منابع و برنامه‌ریزی‌های استراتژیک مورد استفاده قرار داد.

کلید واژه‌ها: تجزیه‌و تحلیل سری‌زمانی، تعداد بیماران، SARIMA،Holt-Winters، شبکه‌های عصبی

•ارجاع: رسولی سمیرا، تابش حامد، اطمینانی کبری. برازش مدل‌های هوش مصنوعی و آمار کلاسیک سری زمانی جهت پیش‌بینی تعداد بیماران بستری بیمارستان‌ها. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۷؛ ۵(۱): ۲۴-۱۲.

۱. دانشجوی کارشناسی ارشد انفورماتیک پزشکی، کمیته تحقیقات دانشجویی، گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران
۲. دکتری آمار زیستی، استادیار گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران
۳. دکتری کامپیوتر، استادیار گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران
* **نویسنده مسئول:** مشهد، میدان آزادی، پردیس دانشگاه، دانشکده پزشکی، گروه انفورماتیک پزشکی
• **شماره تماس:** ۰۵۱۳۸۰۰۲۴۲۹
• **Email:** Etminanik@mums.ac.ir

[Downloaded from jhbmi.com on 2024-04-24]

مقدمه

مطالعه و تجزیه‌وتحلیل هر سیستم بهداشت و درمانی تبدیل به یک ضرورت برای بهبود عملکرد آن در طول زمان شده است چراکه بایستی تعدادی اهداف اغلب متضاد مانند به حداقل رساندن هزینه بهداشت و درمان، به حداکثر رساندن استفاده از منابع فیزیکی و انسانی، بهبود کیفیت مراقبت با ارائه سیستم‌های تشخیصی کارآمد، رسیدگی مؤثر به افزایش تعداد بیماران در مدت زمان کوتاه، ساماندهی انواع امکانات بهداشت و درمان در یک محل واحد و به‌طورکلی بهبود عملکرد سیستم بهداشت و درمان را با توجه به بودجه و زمان محدود و از پیش تعیین‌شده برآورده سازد [۱]. مدیریت بیمارستان‌ها، به‌عنوان یکی از زیرسیستم‌های مهم یک سیستم بهداشت و درمانی، ممکن است در مسیر تحقق این اهداف متضاد با چالش‌های بی‌شماری مواجه گردند [۲، ۱].

یک عامل مهم و تعیین‌کننده بر عملکرد فرایندهای ارائه مراقبت‌های بهداشتی در یک سیستم بیمارستانی تعداد بیماران می‌باشد. زیرا مشکلات تصمیم‌گیری یک بیمارستان به‌طور مستقیم به تغییرات ماهیانه تعداد بیماران مرتبط بوده و تحت تأثیر آن قرار می‌گیرد. پیش‌بینی کوتاه‌مدت تعداد بیماران، مبنای اصلی تصمیمات برنامه‌ریزی کوتاه‌مدت برای تجهیزات بیمارستانی، نیروی انسانی، تجهیزات موردنیاز آزمایشگاهی، خدمات غذایی، نیازهای بهداشتی و مانند آن می‌باشد. پیش‌بینی افزایش یا کاهش تعداد بیماران در کوتاه مدت، نیاز به افزایش (کاهش) تجهیزات و پرسنل را نشان می‌دهد. و پیش‌بینی بلند مدت آن برای تصمیمات برنامه‌ریزی بلند مدت در مورد امکانات و بودجه‌بندی سرمایه، حیاتی است [۴، ۳]. همچنین تعداد بیماران بر بارکاری تأثیرگذار بوده و دوره‌های بارکاری بیش‌ازحد، ممکن است ایمنی بیمار و رضایت شغلی کارکنان بیمارستان را تحت تأثیر قرار دهد [۵]. از سوی دیگر، افزایش نیروی انسانی پرهزینه و غیردائمی است؛ بنابراین پیش‌بینی بهتری از تعداد بیماران، ممکن است بر بهبود تصمیمات مربوط به نیروی انسانی و برنامه‌ریزی، مؤثر واقع شود. همچنین این پتانسیل را دارد که ظرفیت سیستم موجود را افزایش دهد، تأخیر در مراقبت از بیمار را به حداقل برساند و منجر به بهبود کیفیت کلی مراقبت گردد [۶]. هرچند تخمین نادرست تعداد بیماران نیز منجر به گسترش بی‌مورد امکانات و درنتیجه افزایش غیرضروری هزینه‌های عملیاتی بر بیمارستان و نیز بیماران می‌گردد. این موارد، نیاز به روشی را برای مدل‌سازی برجسته می‌سازد، که بتواند تصمیم‌گیرندگان را در برآورد قابل‌اعتمادی از تعداد بیماران در آینده یاری رساند.

پیش‌بینی تعداد بیماران به کمک تکنیک‌های سری‌های زمانی به‌طور گسترده‌ای به‌منظور بررسی‌های مربوط به اشغال تخت و پذیرش بیماران در زمینه طب اورژانس مورد مطالعه قرار گرفته است [۱۱-۷]. در مطالعه‌ای مشابه که به پیش‌بینی تعداد بیماران پذیرش شده پرداخته است، از دستاورد این مطالعه به‌عنوان ابزاری در جهت کمک به مدیریت بیمارستان، ازجمله مواردی نظیر برنامه‌ریزی عمل‌های جراحی، مدیریت تخت و منابع کارکنان اشاره شده است و ادعا می‌کند که این ابزار می‌تواند تأثیر قابل‌توجهی بر مدیریت تخت و برنامه‌ریزی خدمات بهداشتی داشته باشد [۱۲].

یک سری زمانی مجموعه‌ای از مشاهدات درباره یک متغیر است که در نقاط گسسته‌ای از زمان که معمولاً فاصله‌های مساوی دارند (روز، ماه، سال) اندازه‌گیری شده و برحسب زمان مرتب شده‌اند. تفاوت سری‌های زمانی و سایر روش‌های مدل‌سازی در این است که پیش‌بینی سری‌های زمانی از مقادیر داده‌های گذشته و حال مربوط به یک یا چند سری زمانی برای پیش‌بینی مقادیر آینده آن سری استفاده می‌کنند درحالی‌که دیگر روش‌های مدل‌سازی اغلب با استفاده از متغیرهای مستقل دیگر سعی به پیش‌بینی متغیر موردنظر می‌نمایند. از ویژگی‌های اصلی بسیاری ازسری‌های زمانی می‌توان به روند و تنوعات فصلی اشاره کرد که می‌توانند به‌صورت قطعی با توابع ریاضی مربوط به زمان، مدل‌سازی شوند و همچنین ویژگی دیگر آن داده‌های تصادفی موجود در سری‌ها می‌باشد که از الگوی مشخصی پیروی نمی‌کنند؛ اما یکی دیگر از ویژگی‌های مهم سری‌های زمانی این است که مشاهدات نزدیک به هم از لحاظ زمانی، دارای همبستگی هستند. بسیاری از روش‌های تجزیه‌وتحلیل سری‌های زمانی با هدف توضیح این ارتباط و ویژگی‌های اصلی داده‌ها، از مدل‌های آماری مناسب و روش‌های توصیفی استفاده می‌کنند. هنگامی که یک مدل مناسب پیدا و بر روی داده‌ها برازش داده شد می‌توان از آن مدل برای پیش‌بینی مقادیر آینده و یا شبیه‌سازی استفاده نمود، تا به‌عنوان راهنما در تصمیمات مربوط به برنامه‌ریزی مورد استفاده قرار گیرند. مدل برازش داده شده، خود به‌عنوان پایه‌ای برای آزمون‌های آماری در آینده می‌تواند مورد استفاده قرار گیرد [۱۳]. به‌عنوان‌مثال، نوسانات ماهیانه تعداد بیماران می‌تواند شواهد برخی از تغییرات اساسی در سیاست‌گذاری‌ها را نیز مشخص نماید؛ بنابراین درنهایت، با ایجاد

[Downloaded from jhbmi.com on 2024-04-24]

یک مدل بر اساس ویژگی‌های اصلی یک سری زمانی، می‌توان زمینه‌ای برای تصمیمات ضروری مدیران و سیاست‌گذاران فراهم نمود.

هدف از انجام این پژوهش، بررسی قابلیت پیش‌بینی تعداد بیماران بستری در ۶ بیمارستان تحت پوشش دانشگاه علوم پزشکی مشهد با استفاده از تکنیک‌های پیش‌بینی شناخته‌شده مبتنی بر تئوری‌های یادگیری ماشین و آمار کلاسیک سری‌های زمانی جهت ارائه روشی قابل‌اعتماد به‌منظور پیش‌بینی تعداد بیماران این بیمارستان‌ها می‌باشد.

روش

در این مطالعه مقطعی، مدل‌سازی بر اساس مجموعه داده‌های آموزشی یعنی تعداد بیماران بستری ماه‌به‌ماه ۱۲ سال گذشته (از فروردین ۱۳۸۳ تا فروردین ۱۳۹۵) ۶ بیمارستان دولتی شهر مشهد صورت گرفت. داده‌های ماهیانه تعداد بیماران بستری این بیمارستان‌ها از پایگاه داده اداره آمار دانشگاه علوم پزشکی مشهد استخراج شدند. هدف، پیش‌بینی تعداد بیماران بستری برای سه‌ماهه نخست سال ۱۳۹۵ است. مشخصات بیمارستان‌های مورد مطالعه در جدول ۱ نشان داده شد. (به دلیل رعایت موازین اخلاقی از آوردن نام بیمارستان‌ها خودداری گردید).

جدول ۱: مشخصات بیمارستان‌های مورد مطالعه

نام بیمارستان	نوع بیمارستان	انحراف معیار± میانگین تعداد بیماران در ماه
بیمارستان ۱	بیمارستان عمومی	۴۵۱۷ ±۴۱۷
بیمارستان ۲	بیمارستان عمومی	۵۱۲۱ ± ۳۸۵
بیمارستان ۳	بیمارستان تخصصی زنان و زایمان	۱۰۷۳ ±۱۴۹
بیمارستان ۴	سرطان و انکولوژی	۵۴۶ ±۱۱۲
بیمارستان ۵	بیمارستان فوق تخصصی کودکان	۱۳۹۹ ±۱۸۳
بیمارستان ۶	بیمارستان سوانح	۱۳۵۳ ± ۱۲۰

آنالیز اولیه داده‌ها

جهت آنالیز، مدل‌سازی و پیش‌بینی از نرم‌افزار R (نسخه ۳. ۱. ۳.) و نرم‌افزار Matlab نسخه ۲۰۱۵ استفاده شد. پیش از هرگونه اقدام در جهت مدل‌سازی یا پیش‌بینی سری زمانی، تجزیه‌وتحلیل توصیفی اولیه‌ای از داده‌ها امری ضروری می‌باشد. در این مطالعه آنالیز اولیه داده‌ها بر روی مجموعه داده‌های آموزشی صورت گرفت که شامل شناسایی ویژگی‌های مهمی نظیر همبستگی، الگوهای فصلی، تغییرات دوره‌ای تودرتو در الگوهای فصلی، روند، داده‌های پرت و هرگونه نوسانات قابل‌توجه دیگر در این سری‌ها می‌باشد. همچنین ایستایی (Stationary) داده‌ها نیز بایستی مورد بررسی قرار گیرد. منظور از ایستایی این است که آیا خواص اساسی آماری مانند میانگین و واریانس سری زمانی در طول زمان ثابت باقی می‌ماند یا خیر؟ [۱۴]. اکثر مدل‌های پیش‌بینی فرض می‌کنند داده‌های سری زمانی ایستا می‌باشند چراکه با این فرض پیچیدگی ریاضی مدل‌های برازش داده‌شده کاهش می‌یابند حال‌آنکه اغلب در واقعیت چنین نیست. در این مطالعه از آزمون ریاضی Dickey-Fuller برای تشخیص ایستایی و یا غیر ایستایی داده‌ها استفاده شده است [۱۵]. آنگاه در مواردی که سری زمانی غیرایستا بود از تفاضل گیری (Differencing) و تبدیلات توان برای حذف غیر ایستایی روند و یا الگوهای فصلی استفاده شد. آنالیز اولیه داده‌ها نیز از طریق تجزیه‌وتحلیل بصری نمودارهای زمانی شامل ACF (Auto Correlation Function) و Partial PACF(Autocorrelation Function) و محاسبات آمار توصیفی پایه انجام گرفت. ضمناً جهت شناسایی الگوهای موجود در سری‌های زمانی از تابع Decompose در نرم‌افزار R استفاده گردید که این تابع از روش میانگین متحرک برای تجزیه سری زمانی به اجزای فصلی، روند و بخش نامنظم و تصادفی استفاده می‌کند.

انتخاب و برازش مدل‌ها

روش‌های بسیاری برای مدل‌سازی و پیش‌بینی سری‌های زمانی توسعه‌یافته‌اند و انتخاب یک روش مناسب، بسیار اهمیت دارد. تحقیقات گسترده‌ای انجام‌شده است که نشان می‌دهند یک روش پیش‌بینی برتر برای تمامی موارد و در حالت جامع در این زمینه

[Downloaded from jhbmi.com on 2024-04-24]

وجود ندارد [۱۶، ۱۴]؛ بنابراین، جهت به‌کارگیری یک مدل پیش‌بینی با صحت قابل‌قبول، بایستی روش‌های پیش‌بینی مختلفی مورد مقایسه قرار گرفته و توانایی هر یک در پیش‌بینی مشاهدات نمونه‌های آینده مورد ارزیابی قرار گیرد. به دلیل وجود الگوهای فصلی در داده‌های بیمارستان‌های مورد مطالعه، مدل‌های پرکاربرد و دارای صحت بالاتر در مطالعات مشابه که قابلیت پیش‌بینی الگوهای فصلی را نیز داشتند، به‌عنوان مدل‌های پیش‌بینی سری زمانی در این مطالعه در نظر گرفته شده‌اند. که مدل‌های آماری عبارت‌اند از: Holt-Winters، Seasonal Autoregressive Integrated Moving Average(SARIMA) [۱۹-۱۷] و مدل‌های یادگیری ماشین شامل دو مدل از دسته مدل‌های شبکه‌های عصبی یعنی MLP (Multi-Layer Perceptron) و Generalized Regression Neural Networks(GRNN) می‌باشد [۲۰، ۱۶]. به‌طورکلی مجموعه داده‌های هر بیمارستان به دو گروه تقسیم شد. الف) مجموعه داده آموزشی که برای ساخت مدل‌ها به کار گرفته شد و شامل داده‌ها از سال ۱۳۸۳ تا ابتدای ۱۳۹۵ بود. ب) مجموعه داده تست برای ارزیابی صحت (Accuracy) پیش‌بینی مدل‌ها که شامل داده‌های سه‌ماهه نخست سال ۱۳۹۵ بود.

مدل SARIMA

مدل SARIMA گسترش‌یافته روش (Average ARIMA(Autoregressive Integrated Moving برای داده‌های فصلی می‌باشد. منظور ازسری‌های زمانی فصلی، سری‌هایی می‌باشند که رفتار نسبتاً منظمی در فواصل زمانی با طول ثابت از خودشان نشان می‌دهند. در هنگام کار با داده‌های سری زمانی که دارای الگوهای فصلی می‌باشند، نه تنها بررسی وجود همبستگی بین مشاهدات فعلی و مشاهدات اخیر ضروری است که بایستی وجود همبستگی میان مشاهدات فعلی و مشاهدات قبلی در فصل‌های گذشته نیز مورد بررسی قرار گیرد که روشی برای شناسایی وجود یا عدم وجود الگوی فصلی در داده‌ها نیز می‌باشد این مدل فرض می‌کند که سری زمانی در نظر گرفته‌شده خطی است و از یک توزیع آماری شناخته‌شده خاص، مانند توزیع نرمال تبعیت می‌کند. محبوبیت مدل SARIMA عمدتاً به دلیل انعطاف‌پذیری آن در نمایش ساده انواع مختلفی از سری‌های زمانی برای فرایند ساخت مدل بهینه است. برای پیاده سازی مدل SARIMA از الگوریتم ارائه شده توسط Hyndman و Khandakar استفاده شده است [۲۱].

مدل Holt-Winters

Holt-Winters یکی از محبوب‌ترین روش‌های هموارسازی نمایی (Exponential Smoothing) می‌باشد. این مدل بر اساس الگوهای فصلی و روند موجود در داده‌ها ایجاد می‌شود. محبوبیت این روش به دلیل مزایایی از قبیل نیاز محاسباتی کمتر، سادگی استفاده و دقت پیش‌بینی این مدل به‌خصوص برای سری‌های دارای الگوی فصلی می‌باشد. این روش دو گونه معادله (Additive و Multiplicative) دارد ازآنجاکه نوسانات سری‌های زمانی این مطالعه نسبتاً ثابت بودند از معادله Additive استفاده شد که برای سری زمانی $\{x_t: t = 1, \dots \dots, n\}$ به‌صورت زیر می‌باشد:

$$\hat{x}_{t+h/t} = l_t + hb_t + s_{t-m+h_m^+} \quad (۱)$$

که h افق پیش‌بینی است برای مثال در این مطالعه که پیش‌بینی برای سه ماه آینده می‌باشد $h = ۳$ است و سایر پارامترها عبارت اند از:

$$\begin{aligned} \text{Level: } l_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\ \text{Trend: } b_t &= \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \\ Seasonality: s_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned} \quad (۲)$$

α، β^* و γ پارامترهایی هموارسازی هستند که $0 \le \alpha \le 1$ ، $0 \le \beta^* \le 1$ و $0 \le \gamma \le 1 - \alpha$ و m طـول دوره فصلی می‌باشد.

[Downloaded from jhbmi.com on 2024-04-24]

مدل شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی از متداول‌ترین انواع مدل‌های یادگیری ماشین برای پیش‌بینی سری‌های زمانی می‌باشند ایده اصلی شبکه عصبی بر مبنای شبیه‌سازی عملکرد مغز انسان است. از ویژگی‌های بارز این روش در پیش‌بینی سری‌های زمانی، قابلیت ذاتی در مدل‌سازی غیرخطی، بدون هیچ‌گونه پیش‌فرض در مورد توزیع‌های آماری مشاهدات است و مدلی مناسب و سازگار با داده‌ها ایجاد می‌کند. همچنین در شرایطی که در آن داده‌های ورودی نادرست، ناقص و یا فازی باشند نیز قابل اجرا می‌باشد. انواع مختلفی از شبکه‌های عصبی وجود دارد که دو مورد پرکاربرد آن در زمینه سری‌های زمانی یعنی MLP و GRNNدر این مطالعه مورد استفاده قرار گرفته‌اند. مدل GRNN، به کمک Neural Network Toolbox در نرم افزار Matlab نسخه ۲۰۱۵ پیاده‌سازی شد. به طور خاص، مشخصات معماری مدل شبکه عصبی از نظر تعداد ورودی و تعداد نورون‌های لایه پنهان کار آسانی نیست. در این مطالعه به منظور پیدا کردن بهترین معماری برای روش‌های شبکه عصبی، مجموعه داده‌های هر بیمارستان تا ابتدای سال ۱۳۹۵ به سه بخش تقسیم شد. سه ماه آخر مجموعه داده‌ها (دی ۹۴ تا ابتدای فروردین ۹۵) را جهت بررسی عملکرد به عنوان مجموعه تست و سه ماه پیش از آن (پاییز ۹۴) را به عنوان مجموعه صحت سنجی (Validation Set) و بقیه داده‌های موجود به عنوان مجموعه داده‌های آموزشی برای هر بیمارستان در نظر گرفته شد. معماری‌های مختلفی برای ساخت بهترین مدل MLP بر اساس داده‌های آموزشی مورد بررسی قرار گرفت. در نهایت مدل با کمترین مقدار AIC (Information Criterion Akaike) به عنوان مدل نهایی جهت برازش داده‌های هر بیمارستان انتخاب شد. این مدل به کمک بسته نرم‌افزاری "RMiner" در نرم‌افزار R پیاده‌سازی شد.

معیار ارزیابی مدل‌ها

با توجه به اهمیت اساسی پیش‌بینی سری‌های زمانی در بسیاری از موارد عملی، انتخاب یک مدل خاص که از صحت و دقت مطلوب برخوردار باشد بسیار تأثیرگذار است. به منظور پیش‌بینی تعداد بیماران بستری سه ماهه نخست سال ۱۳۹۵ هر یک از بیمارستان‌ها، هر چهار مدل ذکر شده، برای سری زمانی تعداد بیماران بستری هر بیمارستان به‌طور مجزا پیاده‌سازی شدند و پس از حصول نتایج واقعی تعداد بیماران بستری در این سه ماه و مقایسه آن‌ها با نتایج حاصل از پیش‌بینی، خطای پیش‌بینی، جهت بررسی کارایی هر یک از مدل‌ها بر اساس معیار صحت میانگین قدرمطلق درصد خطا The Mean Absolute Percentage Error (MAPE) محاسبه گردید که تفاوت نسبی بین مقادیر واقعی (y_j) و پیش‌بینی‌شده ($\hat{y}_j$) توسط یک مدل را برآورد می‌کند و عبارت است از:

$$MAPE = \frac{100}{N}\sum_{j=1}^{N}\left|(y_j - \hat{y}_j)/y_j\right| \qquad (۳)$$

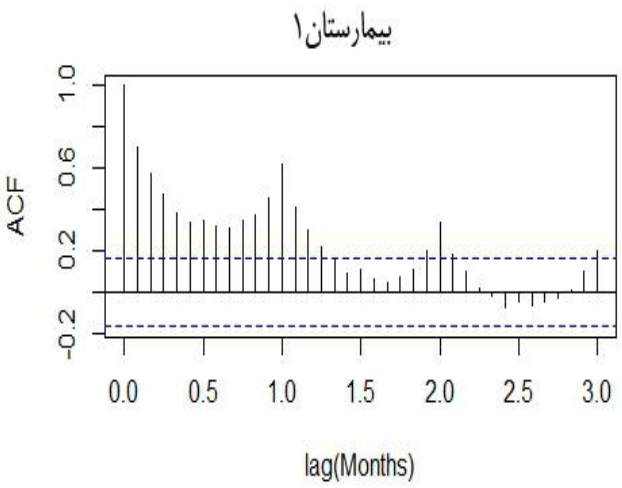
هرچه قدر مقدار MAPE کوچک‌تر باشد پیش‌بینی حاصل از مدل دقیق‌تر خواهد بود. مزیت این روش در این است که مستقل از مقیاس می‌باشد و برای مقایسه عملکرد پیش‌بینی مدل‌های سری زمانی مختلف کاربرد دارد. در نهایت بر اساس این معیار برای هر یک از بیمارستان‌ها به صورت مجزا، مدلی که نتایج بهتری نسبت به سایرین داشته است به عنوان بهترین مدل پیش‌بینی تعداد بیماران بستری آن بیمارستان گزارش شد.

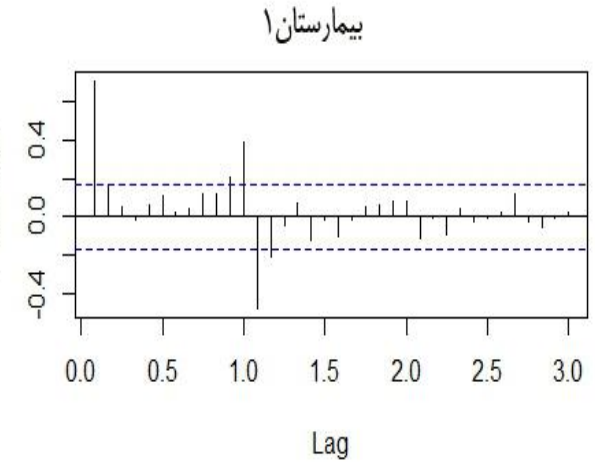
نتایج

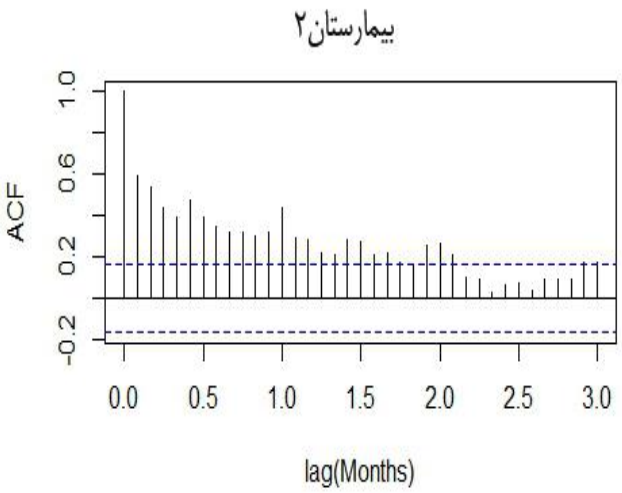
اولین قدم در پیش‌بینی سری‌های زمانی، شناسایی ویژگی‌های مهمی، نظیر همبستگی، الگوهای فصلی، تغییرات دوره‌ای در الگوهای فصلی، روند و ایستایی یا غیر ایستایی داده‌ها می‌باشد. وجود همبستگی در میان داده‌های سری زمانی، امکان انجام مدل سازی و پیش‌بینی مقادیر بعدی سری زمانی را فراهم می‌کند. در این مطالعه وجود همبستگی با رسم نمودار ACF و PACF برای هر بیمارستان با فاصله اطمینان ۹۵ درصد (خطوط نقطه چین) بررسی گردید (شکل ۱). اگرمقدار همبستگی مشاهده شده درخارج از این محدوده واقع شود، می‌گوییم مقدار همبستگی در سطح پنج درصد به طور معنی‌داری با صفر اختلاف دارد. همچنین آزمون Dickey-Fuller برای تشخیص ایستایی یا ناایستایی سری‌های زمانی بر روی مجموعه داده‌های آموزشی هر بیمارستان انجام گرفت (جدول ۲).

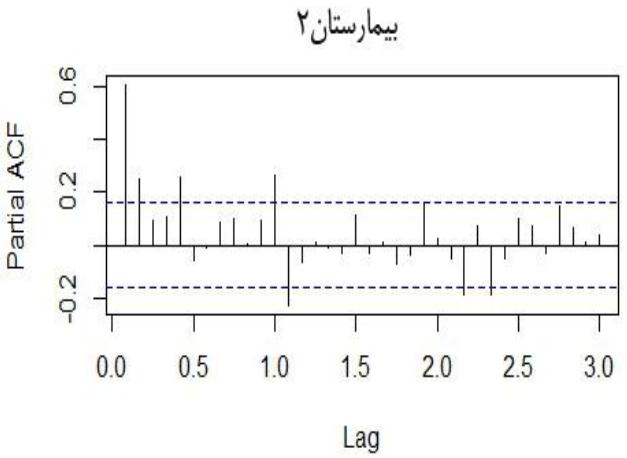
همان‌طور که در نمودار ACF و PACF بیمارستان‌ها مشاهده می‌شود وقفه‌های زمانی معنادار نشان دهنده وجود همبستگی بین مقادیر سری زمانی است که انجام مدل‌سازی و پیش‌بینی مقادیر آینده را برای هر ۶ بیمارستان ممکن می‌سازد.

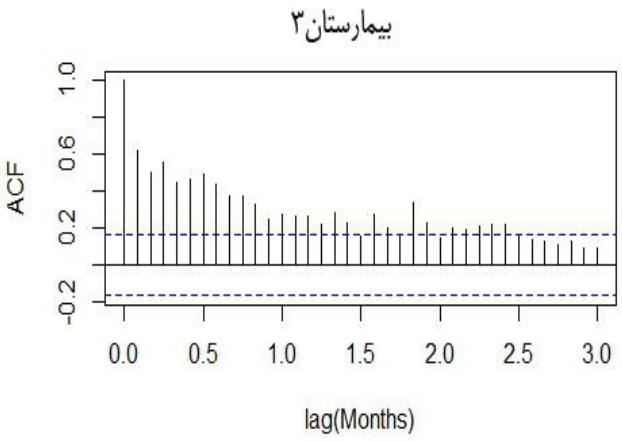
[Downloaded from jhbmi.com on 2024-04-24]

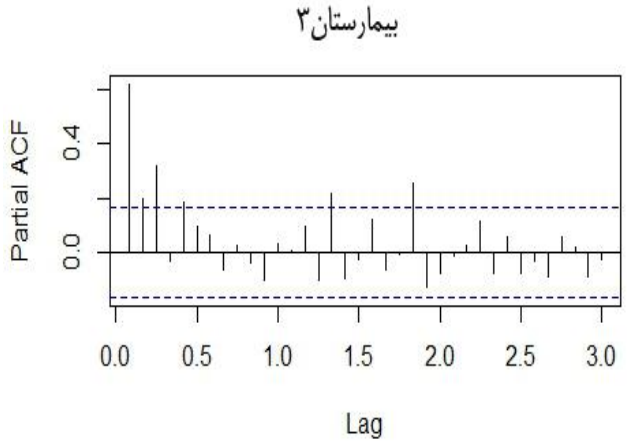




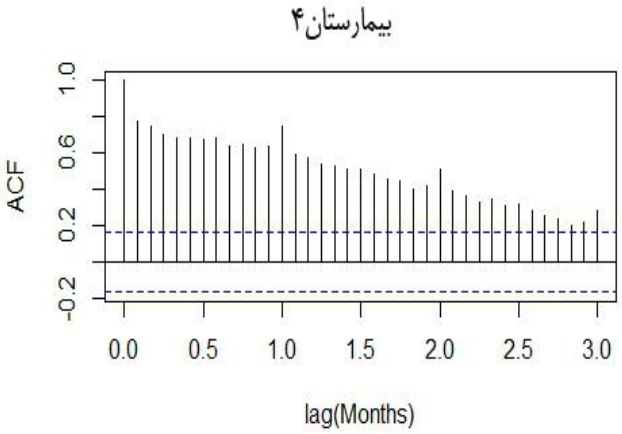


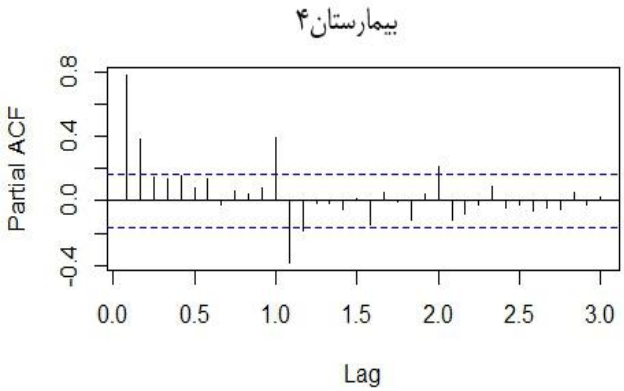


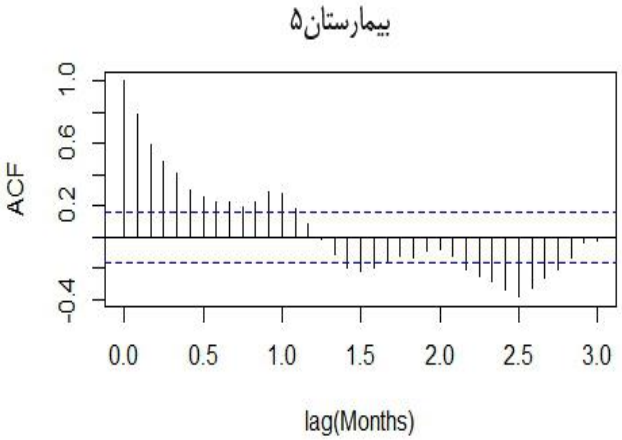


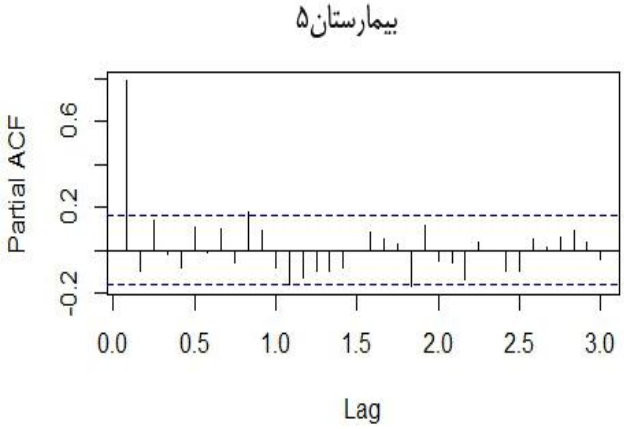


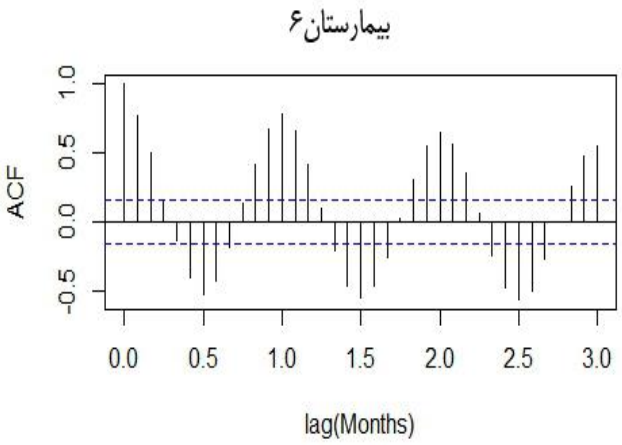
[Downloaded from jhbmi.com on 2024-04-24]

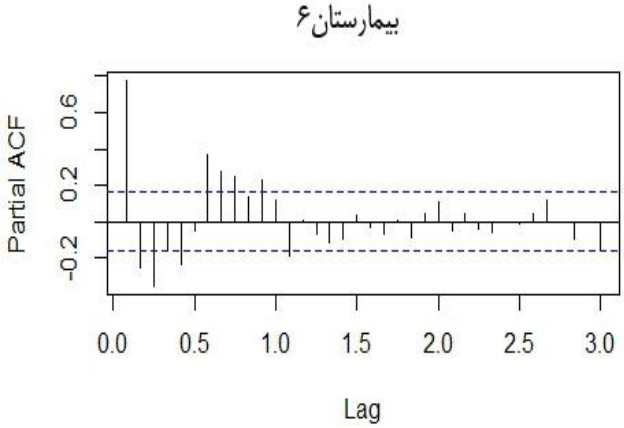












شکل۱: نمودار ACF و PACF تعداد بیماران بستری بیمارستان‌ها

[Downloaded from jhbmi.com on 2024-04-24]

جدول ۲: نتایج آزمون Dickey-Fuller

آزمون Dickey-Fuller	
بیمارستان	P-value
بیمارستان ۱	۰/۰۸۹
بیمارستان ۲	۰/۰۱
بیمارستان ۳	۰/۳۹
بیمارستان ۴	۰/۳۵
بیمارستان ۵	۰/۱۲
بیمارستان ۶	۰/۰۱

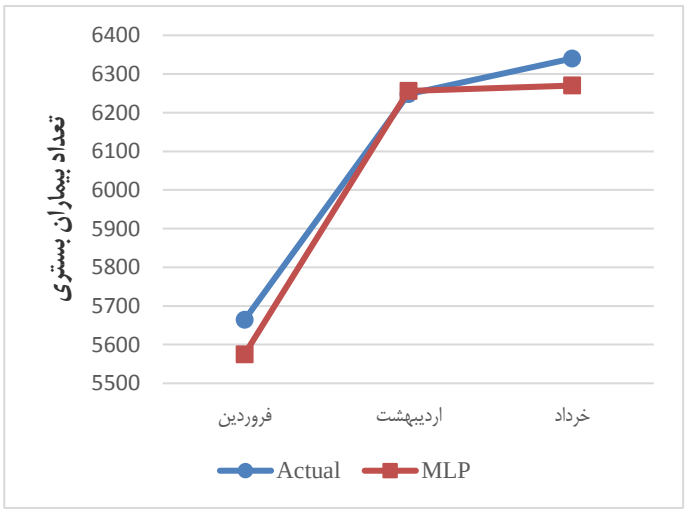
در آزمون Dickey-Fuller، سطح معناداری برای احتمال (P-Value<۰/۰۵) در نظر گرفته شد. فرض H_0 این آزمون نا ایستایی سری را بررسی می‌کند. بر اساس این آزمون، سری زمانی بیمارستان‌های شماره ۲ و ۶ ایستا می‌باشد.

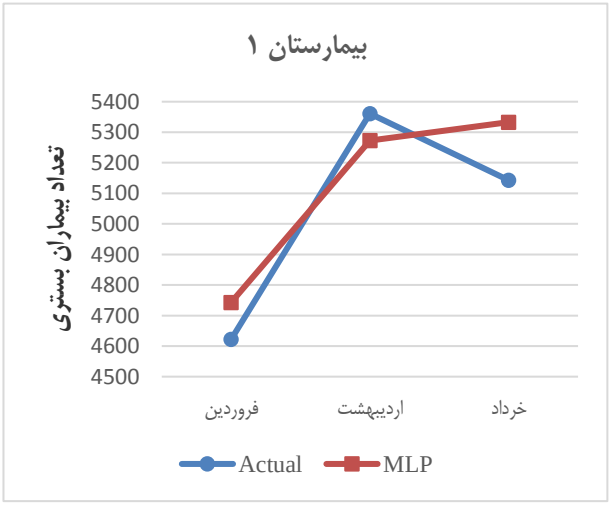
خطای حاصل از برازش مدل‌ها برای هر یک از بیمارستان‌ها در جدول ۳ آورده شد. که مدل با کمترین معیار خطا MAPE به‌عنوان بهترین مدل برای هر بیمارستان انتخاب شد. نتایج حاصل از برازش مدل‌ها برای هر بیمارستان عبارت‌اند از: مدل MLP بهترین مدل برای پیش‌بینی تعداد بیماران بستری بیمارستان‌های شماره ۱ و ۲ با خطای به ترتیب ۲/۲۲ درصد و ۲/۲۶ درصد بود. مدل Holt-Winters بهترین مدل برای پیش‌بینی تعداد بیماران بستری بیمارستان‌های شماره ۳ تا ۶ با خطای به ترتیب ۳/۲۲ درصد، ۴/۱۲ درصد، ۲/۴۳ درصد و ۲/۱۳ درصد می‌باشد. نتایج حاصل از پیش‌بینی بهترین مدل برای هر یک از بیمارستان‌ها برای سه‌ماهه نخست سال ۹۵ در شکل ۲ نشان داده شد.

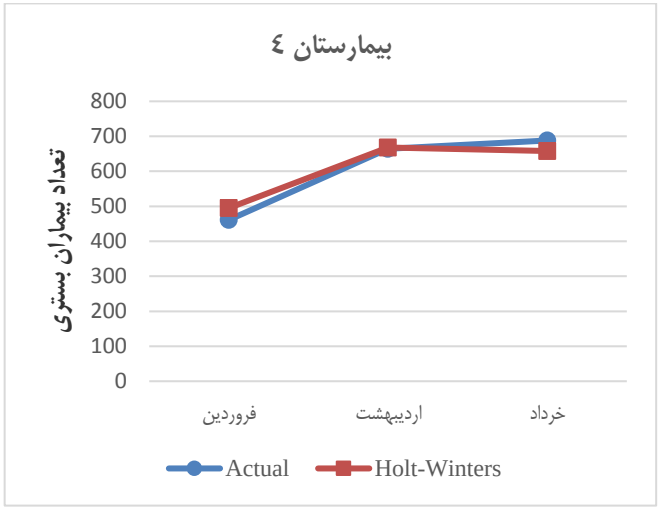
جدول ۳: خطای حاصل از برازش مدل‌ها برای هر یک از بیمارستان‌های مورد مطالعه

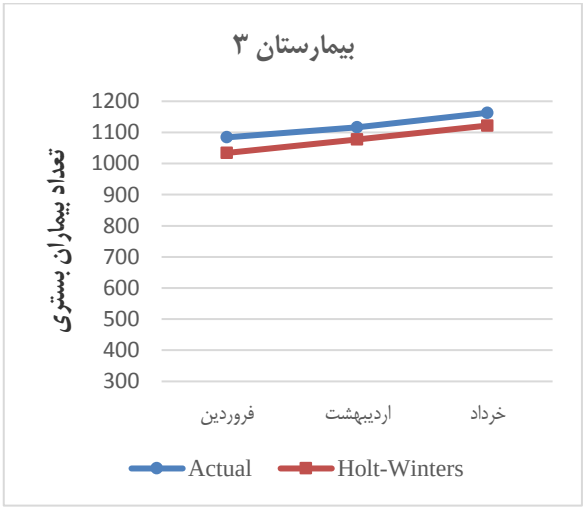
خطای MAPE برای هر مدل (٪)				
بیمارستان	SARIMA	Holt-Winters	MLP	GRNN
بیمارستان ۱	۴/۲۸	۲/۶۳	۲/۲۲	۲/۹
بیمارستان ۲	۴/۹۵	۳/۴	۲/۲۶	۸/۵۸
بیمارستان ۳	۵/۱۷	۳/۲۲	۶/۶۷	۵/۹۹
بیمارستان ۴	۶/۱۴	۴/۱۲	۷/۷۹	۱۳/۹۱
بیمارستان ۵	۲/۹۳	۲/۴۳	۹/۶۷	۵/۶۲
بیمارستان ۶	۳/۵۳	۲/۱۳	۵/۱۴	۲/۴۸

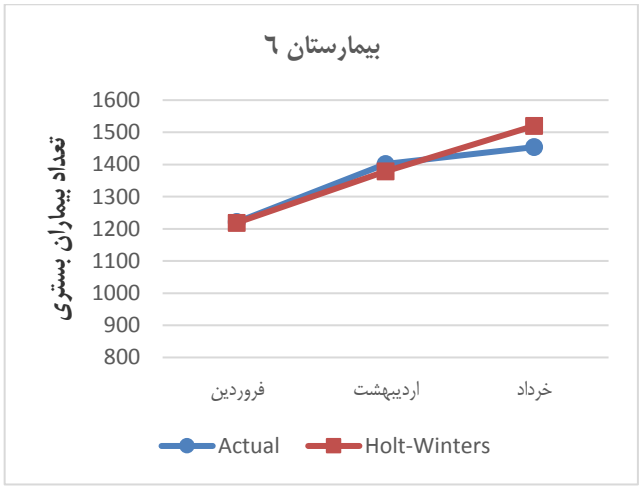
[Downloaded from jhbmi.com on 2024-04-24]

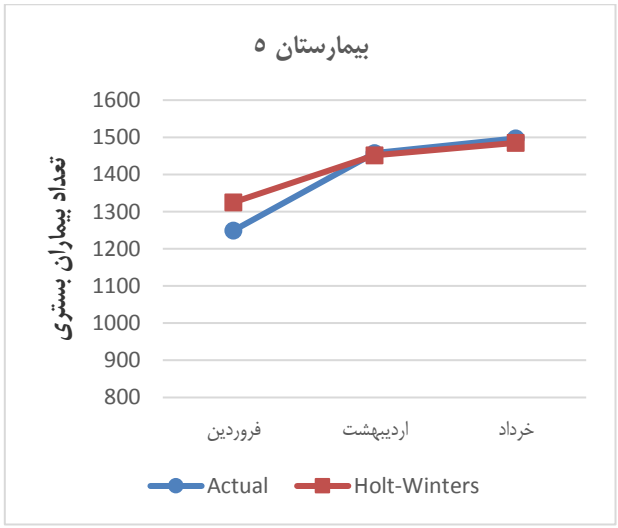












شکل ۲: پیش‌بینی تعداد بیماران بستری بهترین مدل هر بیمارستان برای سه‌ماهه نخست سال ۱۳۹۵

[Downloaded from jhbmi.com on 2024-04-24]

بحث و نتیجه‌گیری

نیاز به پیش‌بینی در سیستم‌های مدیریت بیمارستانی امری ضروری است، زیرا داشتن دانشی حتی جزئی در نتیجه مدل‌سازی و پیش‌بینی، بهتر از عدم داشتن دانش برای گرفتن تصمیمات مدیریتی و انجام برنامه‌ریزی است؛ بنابراین به هر میزان که یک سیستم بیمارستانی قابلیت تخمین و برآورد بهتری نسبت به وضعیت آینده داشته باشد، بهتر می‌تواند برای آن آماده گردد. استفاده از پیش‌بینی ممکن است در مورد برنامه‌ریزی استفاده از تخت‌های بیمارستانی و بهبود تخصیص منابع به کار گرفته شود تا تعداد بحران‌های تعداد تخت کاهش یابد یا در نهایت به طور کامل برطرف شود. به منظور ایجاد یک پیش‌بینی از نیازهای تخت در آینده، نیاز به درک استفاده فعلی از منابع وجود دارد و این امر می‌تواند با استفاده از مدل‌سازی میسر گردد. هدف از مدل سازی امکان استنتاج یک فرایند از داده‌های مشاهده شده است [۲۲].

ارزش مدل‌سازی در این است که فهمیدن را آسان می‌کند و امکان پیش‌بینی‌های آینده را فراهم می‌آورد. بدین منظور، ضروری است که انتخاب مدل مورد استفاده به طور مناسبی صورت گیرد. در این مطالعه، چهار روش مختلف مدل‌سازی و پیش‌بینی سری زمانی، به‌منظور پیش‌بینی ماهیانه تعداد بیماران بستری سه ماهه نخست سال ۱۳۹۵ برای هر یک از ۶ بیمارستان تحت پوشش دانشگاه علوم پزشکی مشهد توسعه داده شد. مدل‌های انتخابی، از دو دسته تکنیک‌های یادگیری ماشین و آمار کلاسیک سری‌های زمانی بودند. برای ارزیابی و مقایسه مدل‌ها، نتایج حاصل از پیش‌بینی، با مقادیر واقعی (مجموعه تست) به کمک معیار MAPE مقایسه شدند که مقادیر مجموعه تست در فرآیند برازش مدل در نظر گرفته نشده بودند.

بر اساس نتایج این مطالعه، روش Holt-Winters با ارائه بهترین نتایج پیش‌بینی برای ۴ بیمارستان می‌تواند روش‌های کارآمدی برای پیش‌بینی تعداد بیماران بیمارستان‌ها باشند. روش های شبکه‌های عصبی نیز برای ۲ بیمارستان بهترین نتیجه را داشته است؛ اما بایستی در نظر داشت که طراحی این مدل‌ها از نظر تعداد ورودی و نورون‌های لایه پنهان و سایر پارامترها کار چندان آسانی نیست. درمجموع، بهترین مدل‌های پیش‌بینی برای هر ۶ بیمارستان، عملکرد قابل قبولی با ارائه معیار صحت (MAPE) در بازه ۲/۱۳ درصد تا ۴/۱۲ درصد داشته‌اند. بر اساس نتایج این مطالعه، می‌توان روش Holt-Winters را به علت سادگی پیاده‌سازی در عین محاسبات کمتر و دقت پیش‌بینی، به‌عنوان روشی مناسب پیشنهاد نمود. به‌عنوان یک استراتژی مهم پیش‌بینی، بایستی در نظر داشت که پارامترهای برآورد شده هر یک از مدل‌ها، منعکس‌کننده رفتار یک بیمارستان خاص در یک مدت زمان خاص می‌باشد؛ بنابراین با افزایش داده‌ها بایستی پارامترهای هر مدل مجدداً برآورد و بر روی داده ها برازش داده شوند.

در تئوری، روش‌های یادگیری ماشین مورد استفاده در این مطالعه، با ارائه قابلیت‌های یادگیری غیرخطی، عدم نیاز به پیش فرض اولیه از توزیع داده‌ها و غیر پارامتریک بودن در مقایسه با روش‌های سری زمانی آمار کلاسیک منعطف‌تر هستند و به همین دلیل، الگوریتم‌های یادگیری ماشین به سرعت در کاربردهای مربوط به پیش‌بینی‌های سری زمانی مورد استفاده قرار گرفتند. مطالعات نشان داده‌اند که انتخاب پارامترهای روش های یادگیری ماشین، تأثیر قابل توجهی در دقت پیش‌بینی این روش‌ها دارد. مدل‌سازی شبکه‌های عصبی با چالش‌های متعددی در تعیین معماری شبکه نظیر تعداد گره‌های ورودی، نورون‌های پنهان و خروجی و پارامترهای یادگیری مرتبط با ساختار داده‌های سری زمانی رو به رو است. خوشبختانه، مطالعات بسیاری در زمینه انتخاب تعداد نورون‌های لایه پنهان صورت گرفته است [۲۳]. در حالی که، شناسایی و انتخاب متغیرهای ورودی همچنان یک چالش مهم در مدل‌سازی شبکه‌های عصبی برای پیش‌بینی سری‌های زمانی می‌باشد [۲۴]؛ بنابراین بررسی و انتخاب متغیرهای ورودی برای روش‌های یادگیری ماشین سری زمانی نیاز به مطالعه و بررسی بیشتری دارد.

از محدودیت‌های مطالعه حاضر می‌توان به این مورد اشاره کرد که تعداد داده‌های آموزشی برای ایجاد مدل و داده‌های مجموعه تست که به کمک مدل برازش داده‌شده پیش‌بینی می‌شوند مورد بررسی قرار نگرفتند. که انتظار می‌رود در پژوهش‌های آینده جهت بررسی کارایی مدل‌ها در پیش‌بینی‌های بلندمدت موردبررسی قرار گیرند. روش‌های مدل‌سازی مورد استفاده در این مطالعه تنها از الگوهای موجود در داده‌های تاریخی تعداد بیماران ماهیانه بیمارستان جهت پیش‌بینی مقادیر آینده استفاده کرده‌اند و تأثیر احتمالی متغیرهای آب و هوا و یا دیگر عوامل زیست‌محیطی را در نظر نگرفته‌اند. هرچند نتایج مختلفی در مورد اثرات این متغیرها وجود دارد و بیان می‌شود که تأثیر این متغیرها در دقت پیش‌بینی ناچیز می‌باشد [۸]. همچنین تنها متغیر مورد استفاده در این مطالعه فاکتور زمان بوده است در حالی که سایر متغیرها و همچنین اطلاعات تعداد بیماران بستری به تفکیک بخش‌ها موجود نبوده است.

[Downloaded from jhbmi.com on 2024-04-24]

اما نقاط قوت این مطالعه نسبت به مطالعات مشابه بررسی انواع مختلفی از روش‌های تجزیه و تحلیل سری زمانی شامل روش های آماری و روش‌های قدرتمند و انعطاف‌پذیر یادگیری ماشین، نظیر شبکه‌های عصبی می‌باشد. دوم اینکه در این مطالعه یک بررسی خارج از نمونه (post sample analysis) واقعی انجام شد. پس از برازش مدل بر روی داده‌های موجود و پس از پیش بینی مقادیر سه ماهه نخست سال ۱۳۹۵، مقادیر واقعی این سه ماه صرفاً جهت محاسبه خطای مدل‌ها مورد بررسی قرار گرفت. همچنین طیف وسیعی از بیمارستان‌ها با تخصص‌های مختلف مورد بررسی قرار گرفته اند که امکان مقایسه بهتری از روش‌ها را فراهم نمود.

کارایی روش‌های سری زمانی در زمینه‌های مختلف به‌ویژه در حوزه اقتصاد به اثبات رسیده است و به‌طور گسترده‌ای در نرم‌افزارهای آماری استاندارد و جعبه افزارهای‌های مالی قابل‌دسترس هستند. بااین‌حال، مطالعات اندکی در مورد به‌کارگیری روش‌های سری زمانی در حوزه سلامت وجود دارد. در این مطالعه، سعی شد با در نظر گرفتن چندین روش مختلف پیش‌بینی سری زمانی شامل روش‌های آماری و یادگیری ماشین، در بیمارستان‌های مختلف علاوه بر بررسی موضوع قابلیت پیش‌بینی تعداد بیماران، عملکرد روش‌های سری زمانی را نیز در حوزه سلامت مورد بررسی قرار داده شد و در نهایت مدل‌های پیش‌بینی مناسب و قابل قبولی برای هر یک از بیمارستان‌ها ارائه شد. پیش‌بینی دقیق این مجموعه‌ها، برنامه‌ریزی مربوط به پرسنل پرستاری و تخصیص کارکنان در بخش‌ها را تسهیل می کند و به طور بالقوه می‌تواند در پیش‌بینی اشغال تخت نیز مؤثر واقع شود. در یک مطالعه اثرات مثبت بر رضایت بیمار را به عنوان نتیجه استفاده از یک مدل پیش‌بینی برای بهینه‌سازی الگوهای کارکنان نشان می‌دهد [۲۵]. خروجی مدل‌سازی و پیش‌بینی تقاضا، پایه‌ای برای برنامه‌ریزی‌های استراتژیک خدمات بهداشتی و درمانی است که توسط برنامه‌ریزان خدمات درمانی جهت مدل سازی نیروی کار، زیرساخت و موارد مالی استفاده می‌شود.

به‌طورکلی با توجه به ویژگی‌های منحصربه‌فرد بیمارستان‌های مختلف، مدل‌های پیش‌بینی بهینه می‌توانند از بیمارستانی به بیمارستان دیگر متفاوت باشند. بااین‌حال، فرایندهای ذکرشده در این مقاله شامل مدل‌سازی و تجزیه‌وتحلیل داده‌ها را می‌توان در دیگر بیمارستان‌ها و اورژانس‌ها جهت بهبود تخصیص منابع و برنامه‌ریزی‌های استراتژیک مورد استفاده قرار داد.

تشکر و قدردانی

این مقاله برگرفته از بخشی از پایان‌نامه کارشناسی ارشد رشته انفورماتیک پزشکی می‌باشد که از معاونت محترم پژوهشی دانشگاه علوم پزشکی مشهد به خاطر حمایت مالی این طرح و همچنین مدیریت محترم آمار و فناوری اطلاعات و ارتباطات دانشگاه به دلیل همکاری ایشان در گردآوری داده‌ها سپاسگزاری و قدردانی می‌گردد.

این مقاله حاصل طرح تحقیقاتی با شماره ۹۵۰۰۸۰ می‌باشد که با حمایت معاونت پژوهش و فناوری دانشگاه علوم پزشکی مشهد انجام شده است.

References

1. Bhattacharjee P, Ray PK. Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections. Computers & Industrial Engineering 2014; 78: 299-312.

2. Brailsford S, Vissers J. OR in healthcare: A European perspective. European Journal of Operational Research 2011;212(2):223-34.

3. Zhu T, Luo L, Zhang X, Shi Y, Shen W. Time-Series approaches for forecasting the number of hospital daily discharged inpatients. IEEE J Biomed Health Inform 2017;21(2):515-26.

4. Lin WT. Modeling and forecasting hospital patient movements: Univariate and multiple time series approaches. International Journal of Forecasting 1989;5(2):195-208.

5. Sheard S. Hospitalist recruitment and retention: building a hospital medicine programme. Occupational Medicine 2010; 60, (7): 579–80.

6. Institute of Medicine, Board on Health Care Services, Committee on the Future of Emergency Care in the United States Health System. Hospital-Based Emergency Care: At the Breaking Point (Future of Emergency Care). Washington, DC: National Academies Press; 2006. Available from: http://www.nationalacademies.org/hmd/Reports/2006/Hospital-Based-Emergency-Care-At-the-Breaking-Point.aspx.

7. Boyle J, Jessup M, Crilly J, Green D, Lind J, Wallis M, et al. Predicting emergency department admissions. Emerg Med J 2012;29(5):358-65.

8. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. Acad Emerg Med 2008;15(2):159-70.

9. Abraham G, Byrnes GB, Bain CA. Short-term forecasting of emergency inpatient flow. IEEE

[Downloaded from jhbmi.com on 2024-04-24]

Transactions on Information Technology in Biomedicine 2009; 13(3): 380-8.
10. Littig SJ, Isken MW. Short term hospital occupancy prediction. Health Care Manag Sci 2007;10(1):47-66.
11. Bergs J, Heerinckx P, Verelst S. Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis. Int Emerg Nurs 2014;22(2):112-5.
12. Boyle J. PAPT—patient admissions prediction tool. Healthcare IT Manage, 2009 [cited 2016 Oct 20]. Available from: https://healthmanagement.org/c/icu/issuearticle/papt-patient-admissions-prediction-tool
13. Shumway RH, Stoffer DS. Time Series Analysis and Its Applications: With R Examples. 3th ed. Springer; 2010.
14. Asplin BR, Flottemesch TJ, Gordon BD. Developing models for patient flow and daily surge capacity research. Acad Emerg Med 2006;13(11):1109-13.
15. Cochrane JH. Time series for macroeconomics and finance. Manuscript, University of Chicago, 2005. [cited 2016 Sep 28] Available from: http://econ.lse.ac.uk/staff/wdenhaan/teach/cochrane.pdf
16. Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H. An empirical comparison of machine learning models for time series forecasting. Econometric Reviews 2010;29(5-6):594-621.
17. Abraham G, Byrnes GB, Bain CA. Short-term forecasting of emergency inpatient flow. IEEE Trans Inf Technol Biomed 2009;13(3):380-8.
18. Schweigler LM, Desmond JS, McCarthy ML, Bukowski KJ, Ionides EL, Younger JG. Forecasting models of emergency department crowding. Acad Emerg Med 2009;16(4):301-8.
19. Aboagye-Sarfo P, Mai Q, Sanfilippo FM, Preen DB, Stewart LM, Fatovich DM. A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. J Biomed Inform 2015;57:62-73.
20. Singh A, Mishra GC. Application of Box-Jenkins method and Artificial Neural Network procedure for time series forecasting of prices. Statistics in Transition new series 2015; 1(16): 83–96.
21. Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 2008; 27(3): 1-22.
22. Myung IJ, Pitt MA. Applying Occam's razor in modeling cognition: A Bayesian approach. Psychonomic Bulletin & Review 1997;4(1):79-95.
23. Sheela KG, Deepa SN. Review on methods to fix number of hidden neurons in neural networks. Mathematical Problems in Engineering 2013; 2013: 1-11.
24. Curry B, Morgan PH. Model selection in neural networks: some difficulties. European Journal of Operational Research 2006; 170 (2): 567-77.
25. Batal H, Tench J, McMillan S, Adams J, Mehler PS. Predicting patient visits to an urgent care clinic using calendar variables. Acad Emerg Med 2001;8(1):48-53.

[Downloaded from jhbmi.com on 2024-04-24]

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2018; 5(1): 12-24

Evaluation of Artificial Intelligence Models and Classical Statistics Models of Time Series in Forecasting the Number of Hospital Inpatient Admissions

Rasouli Samira[1], Tabesh Hamed[2], Etminani Kobra[3*]

• **Received:** 1 Oct, 2017 • **Accepted:** 23 April, 2018

Introduction: The study and analysis of each health system has become a necessity for its performance improvement through time. In this context, management and analysis of the number of patients is an important factor in the process of improving managers' decisions. The aims of this study were to explore and evaluate the use of multiple time series forecasting methods to predict monthly hospital inpatient admissions at six public hospitals in Mashhad city and to compare the accuracy performance of these methods.

Methods: This cross-sectional modeling study was performed based on monthly data of inpatient admissions at six public hospitals in Mashhad from March 2004 through March 2016. Data were extracted from database of the Statistics Office of Mashhad University of Medical Sciences. Holt-winters, Seasonal Autoregressive Integrated Moving Average (SARIMA), Multilayer Perceptron (MLP) and Generalized Regression Neural Networks (GRNN) models were applied to forecast monthly inpatient numbers at each hospital. The error of the models in regard to the predicted values was reported through Mean Absolute Percentage Error (MAPE).

Results: Holt-Winters method, due to providing the optimal forecasting performance in four hospitals, could be an efficient method for predicting the number of inpatients in hospitals. Totally, the studied models with a MAPE from 2.13% to 4.12% showed acceptable performance in all six hospitals.

Conclusion: Time series analysis is an adequate practical tool for predicting the number of hospital inpatient admissions. Given the unique characteristics of different hospitals, applied methods in this study, including modeling and data analysis can be used in other hospitals to improve their resource allocation and strategic planning.

Keywords: Time Series Analysis, Inpatient Numbers, Holt-Winters, SARIMA, Neural Networks

• **Citation:** Rasouli S, Tabesh H, Etminani K. Evaluation of Artificial Intelligence Models and Classical Statistics Models of Time Series in Forecasting the Number of Hospital Inpatient Admissions. Journal of Health and Biomedical Informatics 2018; 5(1): 12-24.

1. MSc Student in Medical Informatics, Student Research Committee, Medical Informatics Dept., Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
2. Ph.D in Biostatistics, Assistant Professor, Medical Informatics Dept., Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
3. Ph.D in Computer, Assistant Professor, Medical Informatics Dept., Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

*__Correspondence__: Dept., Medical Informatics, Campus of University, School of Medicine, Azadi Square, Mashhad, Iran.
• **Tel:** 05138002429 • **Email:** Etminanik@mums.ac.ir

[Downloaded from jhbmi.com on 2024-04-24]