

# Forecast of Medical Expenses of Iran Health Insurance Organization Using Machine Learning Based Methods

Maryam Arab<sup>1</sup> , Mohammad Fathian<sup>1\*</sup> , Hosein Aliahmadi Jeshfaghani<sup>1</sup> 

<sup>1</sup> Faculty of Progress Engineering, University of Science and Technology, Tehran, Iran

\*Corresponding Author: Mohammad Fathian, Professor, PhD in Industrial Engineering, Department of Progress Engineering, University of Science and Technology, Tehran, Iran. Email: fathian@iust.ac.ir

Received: February 13, 2021

Revised: January 5, 2022

Accepted: January 26, 2022

Online Published: March 9, 2022

## Abstract

**Introduction:** Accurate funding in order to better manage costs is one of the main concerns of managers. The Health Insurance Organization of Iran, as one of the largest basic insurance organizations, is no exception to this and certainly needs to identify and accurately predict the costs of treatment in order to provide financial resources and obtain the necessary funds in its field of treatment. Using machine learning methods to create a model for predicting treatment costs can be a great help in accurately financing.

**Methods:** This study has provided a model and method for predicting the costs of the organization by using the cost data available in the medical documentation systems of the provinces of the organization during the years 1385 to 1398 shamsi and using the SARIMAX and LSTM methods. This method can help to more accurately predict the costs of the organization.

**Results:** Determining the method with better performance based on the MAPE index alone did not meet the desired model; therefore, by creating a combined method and using the criterion of percentage of realization of the forecast, the optimal model for cost forecasting is presented.

**Conclusion:** Due to the need for a scientific method to more accurately predict the costs of the organization, the proposed method and model was able to predict the costs of the organization with minimal errors compared to the errors accepted in manual processes.

**Keywords:** Iran Health Insurance Organization, Medical Expenditure, Forecast, Machine Learning, Time Series

## Citation:

Arab M, Fathian M, Aliahmadi Jeshfaghani H. Forecast of Medical Expenses of Iran Health Insurance Organization Using Machine Learning Based Methods. Iran J Health Insur. 2022;4(4):269-279

# پیش بینی هزینه های درمانی سازمان بیمه سلامت ایران با استفاده از روش های مبتنی بر یادگیری ماشین

مریم عرب<sup>۱</sup>، محمد فتحیان<sup>۱\*</sup>، حسین علی احمدی جشفقانی<sup>۱</sup>

<sup>۱</sup> دانشکده مهندسی پیشرفت، دانشگاه علم و صنعت، تهران، ایران

\* نویسنده مسئول: محمد فتحیان، استاد، دکترای مهندسی صنایع، دانشکده مهندسی پیشرفت، دانشگاه علم و صنعت، تهران، ایران.  
پست الکترونیک: fathian@iust.ac.ir

تاریخ انتشار آنلاین: ۱۴۰۰/۱۲/۱۸

تاریخ پذیرش: ۱۴۰۰/۱۱/۰۶

تاریخ تصحیح: ۱۴۰۰/۱۰/۱۵

تاریخ دریافت: ۱۴۰۰/۱۱/۲۵

کیده

**مقدمه:** تأمین دقیق منابع مالی به منظور مدیریت بهتر هزینه ها یکی از دغدغه های اصلی مدیران سازمان ها است. سازمان بیمه سلامت ایران با عنوان یکی از بزرگ ترین سازمان های بیمه گر پایه از این امر مستثنی نبوده و قطعاً برای تأمین منابع مالی و اخذ بودجه های لازم در حوزه درمان خود، نیازمند شناسایی و پیش بینی دقیق هزینه های درمان است. استفاده از روش های مبتنی بر یادگیری ماشین به منظور ایجاد مدل پیش بینی هزینه های درمان می تواند کمک بزرگی به تأمین دقیق تر منابع مالی کند.

**روش بررسی:** این پژوهش با استفاده از داده های هزینه ای موجود در سامانه اسناد پزشکی استان های سازمان طی سال های ۱۳۸۵ تا ۱۳۹۸ و با استفاده از روش های SARIMAX و LSTM، مدل و روشی را برای پیش بینی هزینه های سازمان ارائه داده است. این روش می تواند به پیش بینی دقیق تر هزینه های سازمان کمک کند.

**یافته ها:** مشخص کردن روش با کارایی بهتر بر اساس شاخص MAPE به تنهایی جوابگوی ایجاد مدل مطلوب نیست، بنابراین با ایجاد یک روش ترکیبی و استفاده از معیار درصد تحقق پیش بینی، مدل مطلوب برای پیش بینی هزینه ها ارائه شده است.

**نتیجه گیری:** با توجه به ضرورت داشتن روش علمی به منظور پیش بینی دقیق تر هزینه های سازمان، روش و مدل پیشنهاد شده توانست با حداقل خطا نسبت به خطاهای پذیرفته شده در فرآیندهای دستی، هزینه های سازمان را پیش بینی کند.

**واژگان کلیدی:** سازمان بیمه سلامت ایران، هزینه های درمانی، پیش بینی، یادگیری ماشین، سری زمانی

## مقدمه

سهم هزینه های خانوارها از مجموع هزینه های مصرفی بهداشت و درمان کشور، از عوامل محدودکننده در تأمین مالی هزینه های مصرفی در بخش سلامت محسوب می شود [۱]. یکی از راهکارها برای برخورد با کسری منابع مالی، مدیریت و کنترل هزینه ها بر اساس پیش بینی دقیق از هزینه ها است. در تمامی سازمان ها حجم انبوهی از داده وجود دارد که استفاده و تجزیه و تحلیل آنها می تواند کمک شایان توجهی به تصمیم گیری مدیران داشته باشد، اما از طرفی حجم بالای این داده ها و تنوع موجود در آنها و ارتباطات زیاد و ناشناخته بین آنها، باعث شده است که ابزارهای دستی و سیستمی معمولی قادر به استفاده درست از آنها نبوده و بررسی این داده ها را با روش های قدیمی، خارج از توان افراد و ناکارآمد کند. پیشرفت فناوری و ایجاد

رشد پرشتاب هزینه های بهداشتی و درمانی در کشور به طور عمده ناشی از عوامل مختلفی مانند رشد جمعیت، سالمند شدن جمعیت، تغییر الگوی بیماری ها و گسترش بیماری های غیرواگیر و مزمن، تغییر سطح انتظارات، گرایش روزافزون به استفاده از تجهیزات و فناوری نوین و گران قیمت درمانی است. در چنین شرایطی توجه به مکانیسم تأمین منابع مالی در بخش بهداشت و درمان بیش از پیش اهمیت یافته است. با توجه به روند اعتبارات و تنوع منابع در بخش بهداشت و درمان، بر اساس آنچه در چارچوب حساب های ملی بهداشت و درمان کشور مشاهده می شود، پایین بودن سهم اعتبارات تخصیصی از محل بودجه عمومی دولت و منابع پیش پرداخت در بیمه های اجتماعی درمانی و بالا بودن

پیش‌بینی، شناسایی و تعیین شوند که به‌طور معمول انتخاب آنها بر اساس نظریات علمی صورت می‌گیرد. برای پیش‌بینی با استفاده از مدل رگرسیونی حاوی متغیرهای توضیحی و قرار دادن آنها در مدل رگرسیونی برآورد شده، به پیش‌بینی متغیر وابسته دست‌یافت. به‌عبارت‌دیگر، حتی در صورت وجود داده‌های تاریخی کافی از کلیه متغیرهای توضیحی مدل، توانایی در پیش‌بینی صحیح متغیر وابسته، منوط به توانایی پیش‌بینی کلیه متغیرهای مستقل است که این امر سبب ضعف مدل‌های علی و معلولی در استفاده از آنها در امر پیش‌بینی می‌گردد. در مقابل مدل‌های علی و معلولی، مدل‌های سری زمانی وجود دارد که هیچ‌گونه اتکایی به نظریات و تئوری‌های علمی و اقتصادی نداشته و حتی از اطلاعات مربوط به سایر متغیرهای اثرگذار بر متغیر مورد‌نظر استفاده نمی‌کنند، بلکه در ساده‌ترین حالت تنها با اتکا به مقادیر گذشته خود متغیر اصلی و همچنین مقادیر جاری و گذشته خطای مدل‌سازی (به‌عنوان عامل دربرگیرنده تأثیر سایر عوامل اثرگذار)، سعی در پیش‌بینی آینده دارد. طبیعت یک سری زمانی، وابسته یا همبسته بودن مشاهدات آن است و از این رو برخلاف مدل‌های علی و معلولی، ترتیب مشاهدات در سری‌های زمانی دارای اهمیت است. به‌عبارت‌دیگر، در پیش‌بینی یک پدیده با استفاده از مدل‌های سری زمانی، ورودی‌های فیزیکی سیستم، تنها عامل تعیین‌کننده خروجی‌ها نبوده و خروجی سیستم متأثر از یک پارامتر مهم به نام زمان نیز هست [۲].

یک سری زمانی دنباله‌ای از مشاهدات بر روی یک متغیر مورد توجه است که در نقاط گسسته‌ای از زمان که معمولاً فاصله‌های مساوی دارند (روزانه، هفتگی، ماهانه، فصلی، سالانه)، رخ می‌دهد. تجزیه و تحلیل سری‌های زمانی، متضمن توصیف فرآیند یا پدیده‌ای است که تولید دنباله می‌کند. برای پیش‌بینی سری‌های زمانی، لازم است که رفتار فرآیند را با یک مدل ریاضی که قابل تعمیم به آینده باشد، توصیف کرد. مهم‌ترین نکته در داده‌های سری زمانی، آن است که این داده‌ها دارای همبستگی هستند. از ضریب همبستگی برای تعیین همبستگی بین مقادیر  $X$  و  $Y$  استفاده می‌شود، اما وقتی خود متغیرهای مستقل، مقادیرشان به هم مرتبط باشد، به آن خودهمبستگی می‌گویند [۳].

هر سری زمانی یکی از دو نوع زمان پیوسته یا زمان گسسته است. در واقع سری زمانی، زمان پیوسته است هرگاه مشاهدات در یک فاصله زمانی به‌طور پیوسته جمع‌آوری شده باشند. یک سری زمانی زمان گسسته است، هرگاه مشاهدات فقط در زمان‌های گسسته متعلق به مجموعه‌ای شما را جمع‌آوری کرده باشند. اکثر سری‌های زمانی که با آن سروکار داریم از نوع گسسته

فناوری‌های جدید، دغدغه‌های موجود در این زمینه را کاهش داده و با پیشنهاد به‌کارگیری ابزارها و تکنیک‌های جدید، امکان بررسی داده‌های انبوه و کشف دانش از دل این داده‌ها را امکان‌پذیر ساخته است. یکی از امروزی‌ترین و جدیدترین این تکنولوژی‌های، داده‌کاوی و استفاده از الگوریتم‌های مختلف آن است [۱]. داده‌کاوی می‌تواند بسیاری از سؤالات مدیران را پاسخ داده و فرآیند تصمیم‌گیری را بهینه، منطقی‌تر و برخاسته از شواهد داده‌ای مطمئن کند و در نتیجه باعث صرفه‌جویی فراوان در هزینه و منابع مالی و انسانی به‌کار گرفته‌شده به‌منظور تحلیل داده‌ها شود.

سازمان بیمه سلامت نیز همانند سایر سازمان‌ها و دستگاه‌ها، دارای سامانه‌های اطلاعاتی متعددی است که روزانه حجم زیادی از اطلاعات را در خود ذخیره و نگهداری می‌کند. طبیعتاً استفاده از این داده عظیم تولیدشده با استفاده از تکنیک‌های جدیدی همچون داده‌کاوی می‌تواند سازمان را به کشف دانش فراوانی از داده‌ها و موجودیت‌های آن برساند. سازمان بیمه سلامت برای مواجهه با مهم‌ترین دغدغه خود که پیش‌بینی و مدیریت هزینه‌های درمانی سازمان است، از روش‌ها و شاخص‌های سنتی و قدیمی استفاده می‌کند و موارد استفاده از تکنیک‌های داده‌کاوی در خصوص این موضوع بسیار کم و اندک است. در این پژوهش سعی شده با استفاده از الگوریتم‌ها و روش‌های آماری مبتنی بر یادگیری ماشین، مدل‌های بهینه برای پیش‌بینی هزینه‌های درمان سازمان ایجاد شود که با استفاده از آن بتوان ضمن پیش‌بینی هزینه‌های آتی سازمان بیمه سلامت، مدیریت هزینه‌های درمانی سازمان را نیز به نحو مطلوب انجام داد.

در حالت کلی می‌توان روش‌های پیش‌بینی را به دور روش کیفی و کمی تقسیم کرد. روش‌های کیفی پیش‌بینی مبتنی بر تخمین ذهنی از طریق نظرات و عقاید متخصصان در موضوع مورد‌نظر است. از این رو در روش‌های کیفی، رویه‌های سیستماتیک برای طرح، ترکیب و همگراسازی نظرات کارشناسان و صاحب‌نظران مربوطه طراحی می‌شود. یکی از روش‌های پرکاربرد در این زمینه، روش دلفی است. روش‌های کمی که بر پایه مدل‌های آماری بنا نهاده شده‌اند، سعی دارند با استفاده از داده‌های تاریخی، به شناخت فرآیند و یا به تعبیری تابع مولد این داده‌ها بپردازند تا در نهایت با فرض پایداری این فرآیند، از معلومات به‌دست‌آمده برای ترسیم وضعیت آینده آن استفاده کنند. روش‌های کمی را می‌توان به دو دسته مدل‌های علی و معلولی و مدل‌های سری زمانی تقسیم کرد. استفاده از مدل‌های علی و معلولی و یا در اصطلاح مدل‌های ساختاری در امر پیش‌بینی، مستلزم این است که در ابتدا متغیرهای مرتبط و اثرگذار بر متغیر مورد‌نظر، به‌منظور

در ادامه به منظور مشخص تر شدن روش انجام پیش‌بینی، به بررسی مطالعات انجام شده پرداخته شده است. در بررسی مطالعات سعی شده مطالعات انجام شده کاملاً مرتبط با مدل‌ها و الگوریتم‌های استفاده شده در خصوص پیش‌بینی هزینه‌های درمان در سازمان‌های بیمه پایه باشد اما به دلیل محدودیت در مطالعات موجود، با توجه به اهمیت تبیین مدل‌ها و الگوریتم‌های به کار گرفته شده در خصوص پیش‌بینی، برخی از مطالعات در خصوص سایر موضوعات مرتبط با حوزه بهداشت و درمان نیز آورده شده است.

۱. تعدد زیاد موضوعات حوزه بیمه، درمان و سلامت
  ۲. گرایش اکثر موضوعات حوزه درمان و سلامت به موضوعاتی مانند پیش‌بینی، تشخیص و جلوگیری از بیماری‌ها، پیش‌بینی هزینه‌های درمانی یک بیماری خاص و...
  ۳. گرایش اکثر موضوعات حوزه بیمه به موضوعات حوزه بیمه تجاری مانند مدیریت ارتباط با مشتریان و مدیریت ریسک و...
  ۴. موردی بودن مطالعه و فقتم امکان تعمیم و توسعه آن به موضوعات مشابه
  ۵. نبود عوامل مؤثر سیاسی و اقتصادی خاص ایران بر ایجاد هزینه‌های درمانی در مطالعات خارج از کشور
  ۶. وجود محدودیت در دسترسی به اطلاعات و دیتاهای مورد نیاز جهت انجام مطالعه
  ۷. نبود مدل علمی و دقیق به منظور پیش‌بینی هزینه‌ها در سازمان بیمه سلامت
- از این رو با توجه به دلایل مطرح شده در بندهای ۶ و ۷، با توجه به نبود محدودیت پژوهشگر در دسترسی به اطلاعات و داده‌ها و ضرورت شناسایی عوامل مؤثر بر ایجاد هزینه‌های درمانی از نگاه یک سازمان بیمه درمانی دولتی و توسعه مدلی علمی و دقیق و براساس تکنیک‌های روز داده‌کاوی برای پیش‌بینی هزینه‌ها، ضرورت انجام پژوهش واضح و مبرهن به نظر می‌رسد.
- همچنین با بررسی مطالعات فوق، می‌توان لیستی از روش‌ها و مدل‌های به کار گرفته شده برای انجام پیش‌بینی را تهیه کرد. همان‌طور که مشخص است، در اکثر مطالعات از روش‌های آماری مبتنی بر ARIMA به همراه روش‌های مبتنی بر شبکه‌های عصبی استفاده شده، بنابراین در این پژوهش با توجه اینکه داده‌های مورد پژوهش دارای ماهیت سری زمانی هستند، از روش‌های توسعه یافته کلاسیک آماری مانند SARIMAX «میانگین متحرک خود توضیح فصلی با متغیر برون‌زا» و روش‌های مبتنی بر شبکه‌های عصبی مصنوعی مانند LSTM «حافظه کوتاه‌مدت بلند» برای ارائه مدل پیش‌بینی نهایی استفاده شده است.

هستند زیرا آنها به صورت ساعتی، روزانه، ماهانه، فصلی، سالانه و... ثبت شده‌اند [۴].

اولین و معروف‌ترین مدل مورد استفاده در مدل‌سازی سری‌های زمانی، مدل ARMA است. در این مدل فرض می‌شود که فرآیند تولید داده‌های سری زمانی مبتنی بر یک الگوی خطی است و نحوه مدل‌سازی آن توسط باکس و جنکینز مطرح شده است. اگرچه در سال‌های اخیر ارائه و بسط مدل‌هایی نظیر شبکه‌های عصبی مصنوعی که قابلیت تخمین مدل پیچیده و غیرخطی را دارند گسترش یافته، اما همچنان به دو دلیل عمده استفاده از مدل‌های خطی در کارهای تجربی رواج دارد؛ اول اینکه بسیاری از اوقات تقریب خطی مدل‌های غیرخطی ساده‌تر بوده و به طور معمول اطلاعات چندانی از دست نمی‌رود. دوم اینکه با کنار گذاشتن استفاده از تقریب خطی یک فرآیند غیرخطی، طیف وسیعی از الگوهای غیرخطی مطرح می‌شوند که در بسیاری از اوقات، تعیین بهترین الگوی غیرخطی کاری بسیار دشوار بوده و گاهی استفاده از الگوی غیرخطی نامناسب، عواقب بسیار ناگوارتری از تقریب خطی فرآیند مورد نظر دارد [۲].

از طرفی ماهیت روابط، تشریح‌کننده بسیاری از فرآیندها و سری‌های زمانی اغلب غیرخطی است. به همین دلیلی پیش‌بینی رفتار چنین فرآیندها و سری‌هایی نیازمند ابزارهای پیش‌بینی دقیق و اثربخش است. در واقع، در بسیاری از سیستم‌های پیچیده و خصوصاً غیرخطی که مدل‌سازی و به دنبال آن پیش‌بینی و کنترل آنها از طریق روش‌های کلاسیک رگرسیونی امری بسیار دشوار حتی بعضاً غیرممکن است، از روش‌های غیر کلاسیک که به طور عمده برگرفته از رفتار طبیعت هستند، استفاده می‌شود. از جمله این روش‌ها، شبکه‌های عصبی مصنوعی است که یکی از روش‌های بدیع و در حال تحول محسوب می‌شود و در موضوعات متنوعی از قبیل شناخت الگو، خوشه‌بندی، پیش‌بینی و... به کار رفته و نتایج مفیدی نیز داشته است [۲].

یادگیری ماشین، کاربرد الگوریتم‌هایی است که برای ما حجم عظیمی از داده‌ها را به دانش تبدیل می‌کنند. یادگیری ماشین تلاش می‌کند برنامه‌هایی را طراحی کند که از داده‌ها آموزش ببینند. به عبارت دیگر، الگوریتم‌ها این قابلیت را به ماشین می‌دهند تا رفتارشان را متناسب با داده یا به عبارتی الگوها تغییر دهد. ما در عصری زندگی می‌کنیم که به دلیل دسترسی بودن کتابخانه‌های متن‌باز قدرتمند، ورود به عرصه یادگیری ماشین و استفاده از الگوریتم‌های قدرتمند برای مشاهده الگوهای موجود در داده‌ها و پیش‌بینی‌ها درباره وقایع آینده برای ما امکان‌پذیر و آسان است [۵].

## روش بررسی

### ابزار

ابزار استفاده‌شده برای ایجاد مدل‌ها، زبان برنامه‌نویسی پایتون و کتابخانه‌های قدرتمند موجود در این زبان برنامه‌نویسی بوده و برای جمع‌بندی داده‌های خروجی مدل‌ها و نتایج پژوهش از ویرایشگر پرس‌وجو SQL Server استفاده شده است.

### آماده‌سازی داده‌ها

داده‌های مورد استفاده در این پژوهش، داده‌های هزینه‌های درمان سازمان بیمه سلامت ایران مربوط به سال‌های ۱۳۸۵-۱۳۹۸ است. هزینه‌ها در سامانه اسناد پزشکی استان‌ها به تفکیک پزشک و مؤسسات طرف قرارداد و به صورت ماهانه ثبت می‌شوند. هزینه‌هایی که در این پژوهش مورد استفاده قرار گرفته، سرجمع هزینه‌های ثبت‌شده به تفکیک مؤسسات درمانی طرف

قرارداد است که بر اساس مشخصه‌های هزینه‌ای شامل استان، محل ارائه خدمات درمانی، نوع مالکیت مؤسسات طرف قرارداد، نوع صندوق بیمه‌ای و نوع خدمت درمانی ارائه‌شده (۸ مورد پرهزینه) و به صورت ماهانه گروه‌بندی شده است. هزینه‌های دیگر، شامل خدمات متفرقه، خدمات بستری موقت، خدمات ارائه‌شده به بیماران خاص و خسارت متفرقه در این پژوهش نیامده است.

تعداد کل رکوردهای داده با ساختار و گروه‌بندی فوق ۲۱۳۶۷۶ رکورد است که برای ایجاد هر مدل، حداکثر ۱۶۸ رکورد (۱۲ ماه در ۱۴ سال) استفاده شده است. آماده‌سازی این داده‌ها در محیط SQL Server انجام شده و برای ایجاد مدل‌ها و استفاده در محیط برنامه‌نویسی پایتون، در قالب فایل CSV درآمده است.

## جدول ۱ | خلاصه مطالعات انجام شده

ردیف	مرجع	داده‌های مورد آزمایش	مدل استفاده‌شده به منظور پیش‌بینی
۱	کو و دیگران (۲۰۱۲) [۶]	۱۲ معیار ماهانه هزینه‌های درمانی شامل کل هزینه‌های مراقبت‌های بهداشتی و ۱۱ زیربخش از هزینه‌های درمانی	مدل خطی ARMA و مدل غیرخطی شبکه عصبی
۲	چابونی و ابدنادر (۲۰۱۳) [۷]	متغیرهای اجتماعی، اقتصادی و جمعیتی هزینه‌های سلامت کشور تونس	ARDL و ANN
۳	خو و دیگران (۲۰۱۶) [۸]	دو دیتاست واقعی بیمارستان‌های دالیان استان لیائونینگ چین	ARIMA, LR, GLM, ARIMAX, ARIMAX-ANN, ARIMA-LR
۴	چلوهان و سنق (۲۰۱۷) [۹]	داده‌های مربوط به مواد زاید بهداشتی و درمانی تولیدشده در بیمارستان‌های منطقه گرهووال در اوتاراخند هند	مدل ARIMA
۵	ژو و دیگران (۲۰۱۷) [۱۰]	داده‌های مربوط به ظرفیت تخت روز در بیمارستان	مدل SRARIMA, مدل MSARIMA و مدل ترکیبی از MSARIMA و مدل‌های زنجیره‌ای مارکوف وزن‌دار
۶	لنگت و دیگران (۲۰۱۷) [۱۱]	موارد سرطانی تمامی افراد در هر رده سنی در مراکز درمانی مختلف در سراسر کشور کنیا	مدل ARMA
۷	مورد و دیگران (۲۰۱۷) [۱۲]	داده‌های طرح‌های بهداشتی دانشگاه بوئا از اکتبر ۲۰۱۳ تا اکتبر ۲۰۱۶ شامل حدود ۹۰۰۰۰ بیمه‌شده، ۶۳ میلیون مطالبات درمانی و ۱۲ میلیون مطالبات دارویی	روش تقویت گرادیان، ANN و مدل رگرسیون ریج
۸	ریچاردسون و هارتمن (۲۰۱۸) [۱۳]	دیتاستی از مطالبات به‌وسیله ETG	رگرسیون غیر پارامتری بیزی
۹	سیامی‌نمینی و دیگران (۲۰۱۸) [۱۴]	بر اساس داده‌های مالی	LSTM و AR, MA, SES, ARIMA
۱۰	کلازولگو و درینسکیس (۲۰۱۸) [۱۵]	داده‌های سالانه هزینه‌های سلامت آمریکا در مقطع ۱۹۷۰ تا ۲۰۱۵	ARMA
۱۱	زینوری و دیگران (۲۰۱۸) [۱۶]	داده‌های چندساله مربوط به تقاضای انجام عمل جراحی	SARIMA
۱۲	کوی و دیگران (۲۰۱۸) [۱۷]	دیتاست پرونده الکترونیک سلامت با حدود ۷۵۰۰۰۰ رکورد	۴ مدل یادگیری ماشین چند خروجی شامل چندکاره لاسو، درخت تصمیم، جنگل تصادفی و شبکه عصبی
۱۳	وانگ و دیگران (۲۰۱۸) [۱۸]	داده‌های مطالبات ۱۴۴۴۹۱۲ بیمه‌شده شامل اطلاعات مربوط به ۶۲۴۰۶۳۷۹ پرونده طی یک دوره تقریباً ۷ ساله	یادگیری عمیق، رگرسیون خطی، رگرسیون لاسو، رگرسیون ریج و جنگل تصادفی
۱۴	یوسفی (۲۰۱۹) [۱۹]	داده‌های چند سال قبل موجود در RTNH واقع در بلو هوریزنته پایتخت ایالت میناس گرایس و ششمین شهر بزرگ برزیل	SARIMA-LR و ARMA, MLR, SARIMA, GEE, GLM, SVR
۱۵	دنگ و دیگران (۲۰۲۰) [۲۰]	داده‌های مربوط به بارمراجعة ویزیت‌های سرپایی	ARMA (به‌تنهایی)، LSTM (به‌تنهایی) و مدل ترکیبی ARMA-LSTM
۱۶	کوشک و دیگران (۲۰۲۰) [۲۱]	هزینه‌های هفتگی داروهای خاص بیماران	دو مدل آماری پایه و ARIMA, یک مدل پرسپترون چندلایه (MLP), یک مدل LSTM و مدل ترکیبی از هر سه مدل MLP, ARIMA و LSTM
۱۷	یوکاسان و دیگران (۲۰۲۰) [۲۲]	داده‌های پذیرش بیماران در بخش اورژانس یک بیمارستان خصوصی در ترکیه	LR, ARMA, ANN, ARMA-LR و ANN

ایجاد شده است.

نکته مهم در ایجاد هر مدل در هر روش این است که با توجه به اینکه از توابع خودکار پایتون برای ایجاد مدل‌ها استفاده شده، بر اساس مقادیر در نظر گرفته شده برای پارامترهای توابع ایجاد مدل، مدل ایجاد شده بهترین مدل بر اساس آزمون‌های مربوطه و شاخص‌های ارزیابی مدل برای آن گروه از اطلاعات است.

برای ایجاد این مدل‌ها تکه کدهای مربوط به ایجاد مدل با هر دو روش در حلقه‌های تودرتو اجرا شده و همزمان با ایجاد، برازش و پیش‌بینی مدل، شاخص ارزیابی مدل نیز محاسبه و تمامی مقادیر لازم در جدولی به منظور بررسی‌های بعدی ذخیره شده است.

#### برازش، پیش‌بینی و استخراج نتایج

پس از ایجاد مدل، عملیات برازش و عملیات پیش‌بینی بر اساس مدل ایجاد شده، بر روی داده‌های آزمون انجام می‌شود و در نهایت خروجی‌های مدل استخراج و جمع‌آوری می‌شود.

#### شاخص‌های ارزیابی مدل

برای انتخاب مدل با کارایی بهتر هم از شاخص MAPE «درصد خطای مطلق میانگین» و هم از شاخص‌های تجربی مانند درصد تحقق پیش‌بینی استفاده شده است.

#### انتخاب مدل بهینه

پس از جمع‌آوری نتایج و خروجی‌های مدل‌های ایجاد شده، برای پردازش و استخراج آمارهای مورد نیاز، اطلاعات به محیط SQL Server منتقل شده و در آنجا با توجه به حالت‌های مختلفی که برای انجام پیش‌بینی در نظر گرفته شده و وضعیتی که شاخص‌های ارزیابی مدل در هر مدل دارند، گزارش‌ها و آمارهای مختلف ساخته شده است.

در انجام عملیات پیش‌بینی دستی برای پیش‌بینی هزینه‌های کلی، مطابق جدول ۳ حالت‌های مختلفی از جزئیات هزینه‌ها را به تفکیک مشخصه‌های هزینه‌ای، می‌توان در نظر گرفت. در این پژوهش نیز برای ارزیابی و انتخاب مدل بهینه هم از این حالت‌ها استفاده می‌شود. مطابق با جدول زیر چنانچه برای هر یک از مشخصه‌های هزینه‌ای دو حالت سرجمع و تفکیک در نظر گرفته شود، جمعاً ۱۶ حالت به وجود می‌آید که از همین ۱۶ حالت برای رسیدن به مدل بهینه استفاده خواهد شد.

با توجه به ایجاد دو مدل برای هر سری از اطلاعات بر اساس روش‌های SARIMAX و LSTM و در نظر گرفتن مقدار شاخص ارزیابی MAPE، برای هر مدل دو وضعیت به شرح ذیل به وجود می‌آید:

◀ در روش LSTM نسبت به SARIMAX مقدار شاخص MAPE کمتر است.

برای ایجاد مدل پیش‌بینی از زبان برنامه‌نویسی پایتون نسخه ۳٫۹ استفاده شده است. کتابخانه‌های مهم استفاده شده پایتون برای ایجاد مدل پیش‌بینی به شرح ذیل است:

pandas, numpy, math, matplotlib, scikit-learn, pmdarima, statsmodels, keras

#### روش ایجاد مدل

تقسیم داده‌ها بین دو دسته آزمون و آزمایش داده‌ها به منظور ایجاد مدل به شرح ذیل به دو دسته آموزش و آزمون تقسیم شده‌اند:

برای آموزش ۱۵۶ رکورد که مربوط به سال‌های ۱۳۸۵-۱۳۹۷ است.

برای آزمون ۱۲ رکورد آخر که مربوط به سال ۱۳۹۸ است.

#### ایجاد مدل با توابع پایتون

در این مرحله با استفاده از توابع موجود در پایتون و تعیین مقادیر مناسب برای پارامترهای این توابع و داده‌های آموزش، مدل‌های مورد نظر با استفاده از هر دو روش SARIMAX و LSTM ایجاد می‌شود.

در این پژوهش برای ارزیابی دقیق مدل و مشخص کردن تأثیری که هر یک از مشخصه‌های هزینه‌ای محل ارائه خدمات درمانی، نوع مالکیت مؤسسات طرف قرارداد، نوع صندوق بیمه‌ای و نوع خدمات ارائه‌شده بر ایجاد هزینه‌های درمانی دارند، برای هر سری از اطلاعات بر اساس گروه‌بندی صورت گرفته بر اساس عناوین مشخصه‌های هزینه‌ای، یک مدل با هر دو روش SARIMAX و LSTM ایجاد می‌شود. حداکثر تعداد مدل در صورت وجود داده در هر گروه، برابر ۵۷۶۰ مدل است اما به دلیل فقدان داده در برخی از گروه‌ها، تعداد ۴۹۷۳ مدل در هر روش

#### جدول ۲ | ساختار داده‌های پژوهش

تاریخ	مقادیر ۲۰۰۶-۰۴-۰۱ لغایت ۲۰۲۰-۰۳-۰۱ (۱۳۸۵-۱۳۹۸)
استان ارائه خدمات درمانی	مقادیر: - کلیه استان‌ها و ۱ تا ۳۱ برای ۳۱ استان کشور
نوع اسناد درمانی	مقادیر: ۱- غیر خسارت متفرقه
نوع مالکیت مؤسسات درمانی	مقادیر: - کلیه مالکیت‌ها - ۱ دولتی دانشگاهی - ۲ دولتی غیردانشگاهی - ۳ خصوصی و خیریه
صندوق بیمه‌ای	مقادیر: ۰- کلیه صندوق‌ها ۱- کارکنان دولت ۲- خویش‌فرما ۳- سایر اقشار ۴- روستائیان
هزینه پرداختی پزشک عمومی	هزینه پرداختی پزشک متخصص و فوق تخصص
هزینه پرداختی دندانپزشک	هزینه پرداختی داروخانه
هزینه پرداختی آزمایشگاه	هزینه پرداختی مؤسسات پرتوی پزشکی
هزینه پرداختی مؤسسات توان‌بخشی	هزینه پرداختی خدمات بستری
کل هزینه پرداختی	

پیش‌بینی صورت گرفته به مقدار واقعی است که در بهترین حالت می‌بایست صددرصد باشد.

#### یافته‌ها

در این قسمت، ابتدا برای هر ۴۹۷۳ مدل بر اساس مقدار شاخص MAPE در هر روش (LSTM یا SARIMAX)، مدل با کارایی بهتر مشخص می‌شود. سپس برای پیش‌بینی هزینه‌های کلی و مشخص کردن روش با کارایی بهتر، تعداد مدل‌ها در هر روش شمارش می‌شود. برای جمع‌بندی، بر اساس حالت‌هایی که ویژگی‌های مختلف مشخصه‌های هزینه‌ای می‌توانند با هم داشته باشند (حداکثر ۱۶ حالت)، تعداد مدل‌ها در هر حالت شمارش می‌شود. در جدول ۴، رتبه‌بندی وضعیت شاخص MAPE بر اساس بالاترین تعداد مدل در ۱۰ حالت مختلف مشخصه‌های هزینه‌ای آورده شده است.

تفسیر سطر شماره ۱ جدول: در حالتی که هر ۴ مشخصه هزینه‌ای به‌صورت سرجمع در نظر گرفته شوند، فقط یک مدل در هر روش ایجاد می‌شود که در این حالت، به‌دلیل اینکه مقدار شاخص MAPE در روش LSTM از روش SARIMAX کمتر است، بنابراین روش کارآتر، روش LSTM است.

◀ در روش SARIMAX نسبت به LSTM مقدار شاخص MAPE کمتر است.

شاخص فوق‌کارایی بیشتر یک مدل را در یکی از روش‌ها مشخص می‌کند و برای رسیدن به مدلی که پیش‌بینی هزینه کلی را به بهترین شکل انجام دهد کافی نیست، بنابراین در کنار استفاده از این شاخص، شاخص دیگری که در فرآیندهای دستی مورد استفاده قرار می‌گیرد، در اینجا نیز استفاده شده است. همان‌طور که بیان شد، پیش‌بینی انجام‌شده برای هزینه‌ها در سازمان، زمانی ارزش پیدا می‌کند که هزینه‌های ایجاد شده پس از انجام پیش‌بینی، انحراف عددی زیادی با سقف هزینه‌های پیش‌بینی‌شده که منجر به دریافت اعتبارات بودجه‌ای می‌شود، نداشته باشد. شاخصی که در اینجا مطرح می‌شود، شاخص درصد تحقق پیش‌بینی است، بنابراین در این پژوهش نیز از این شاخص برای رسیدن به روش و مدل پیش‌بینی بهینه که دارای بالاترین درصد تحقق باشد، استفاده می‌شود.

شاید در فرآیند واقعی، برای صرفه‌جویی در هزینه‌ها، هرچه درصد تحقق پایین‌تر باشد بهتر باشد اما در اینجا چون مقادیر پیش‌بینی‌شده، این مقادیر نیز می‌بایست با خطای حداقلی به مقادیر واقعی نزدیک باشد. درصد تحقق پیش‌بینی نسبت مقدار

جدول ۳ | حالات مختلف در نظر گرفتن مشخصه‌های هزینه‌ای به‌منظور انجام پیش‌بینی

توضیحات	نوع خدمت درمانی	نوع صندوق بیمه‌ای	نوع مالکیت موسسه درمانی طرف قرارداد	استان ارائه خدمت درمانی	
هر ۴ مشخصه به‌صورت سرجمع	سرجمع	سرجمع	سرجمع	سرجمع	
نوع خدمت تفکیک و ۳ مشخصه دیگر سرجمع	تفکیک				
نوع صندوق تفکیک و ۳ مشخصه دیگر سرجمع	سرجمع	تفکیک	تفکیک		
نوع صندوق و خدمت تفکیک و ۲ مشخصه دیگر سرجمع	تفکیک				
نوع مالکیت تفکیک و ۳ مشخصه دیگر سرجمع	سرجمع	سرجمع			تفکیک
نوع مالکیت و خدمت تفکیک و ۲ مشخصه دیگر سرجمع	تفکیک				
نوع مالکیت و صندوق تفکیک و ۲ مشخصه دیگر سرجمع	سرجمع	تفکیک		سرجمع	
استان سرجمع و ۳ مشخصه دیگر تفکیک	تفکیک				
استان تفکیک و ۳ مشخصه دیگر سرجمع	سرجمع	سرجمع	تفکیک		
نوع مالکیت و صندوق سرجمع و ۲ مشخصه دیگر تفکیک	تفکیک				
نوع مالکیت و خدمت سرجمع و ۲ مشخصه دیگر تفکیک	سرجمع	تفکیک			تفکیک
نوع مالکیت سرجمع و ۳ مشخصه دیگر تفکیک	تفکیک				
نوع صندوق و خدمت سرجمع و ۲ مشخصه دیگر تفکیک	سرجمع	سرجمع		تفکیک	
نوع صندوق سرجمع و ۳ مشخصه دیگر تفکیک	تفکیک				
نوع خدمت سرجمع و ۳ مشخصه دیگر تفکیک	سرجمع	تفکیک	تفکیک		
هر ۴ مشخصه به‌صورت تفکیک	تفکیک				

مشخصه هزینه‌های دیگر به صورت سرجمع و تفکیک در نظر گرفته شوند، ۱۳۵۷ مدل در هر روش ایجاد می‌شود که در این حالت، روش SARIMAX با ۷۵۱ مدل دارای کارایی بهتری است.

تفسیر سطر شماره ۱۰ و ۱۱ جدول: چنانچه مشخصه هزینه‌ای «مالکیت موسسه درمانی طرف قرارداد» به صورت تفکیک و ۳ مشخصه هزینه‌ای دیگر به صورت سرجمع و تفکیک در نظر گرفته شوند، ۳۶۱۶ مدل در هر روش ایجاد می‌شود که در این حالت، روش LSTM با ۱۹۶۳ مدل دارای کارایی بهتری است.

تفسیر سطر شماره ۱۲ و ۱۳ جدول: چنانچه مشخصه هزینه‌ای «صندوق بیمه ای» به صورت سرجمع و ۳ مشخصه هزینه‌ای دیگر به صورت سرجمع و تفکیک در نظر گرفته شوند، ۱۱۱۲ مدل در هر روش ایجاد می‌شود که در این حالت، روش SARIMAX با ۵۹۵ مدل دارای کارایی بهتری است.

تفسیر سطر شماره ۱۴ و ۱۵ جدول: چنانچه مشخصه هزینه‌ای «صندوق بیمه ای» به صورت تفکیک و ۳ مشخصه

تفسیر سطر شماره ۲ و ۳ جدول: در حالی که هر ۴ مشخصه هزینه‌ای به صورت تفکیک در نظر گرفته شوند، ۲۳۱۷ مدل در هر روش ایجاد می‌شود که در این حالت، روش LSTM با ۱۳۳۲ دارای کارایی بهتری است.

تفسیر سطر شماره ۴ و ۵ جدول: چنانچه مشخصه هزینه‌ای «استان ارائه خدمت درمانی» به صورت سرجمع و ۳ مشخصه هزینه‌ای دیگر به صورت سرجمع و تفکیک در نظر گرفته شوند، ۱۷۹ مدل در هر روش ایجاد می‌شود که در این حالت، روش SARIMAX با ۱۱۱ مدل دارای کارایی بهتری است.

تفسیر سطر شماره ۶ و ۷ جدول: چنانچه مشخصه هزینه‌ای «استان ارائه خدمت درمانی» به صورت تفکیک و ۳ مشخصه هزینه‌ای دیگر به صورت سرجمع و تفکیک در نظر گرفته شوند، ۴۷۹۴ مدل در هر روش ایجاد می‌شود که در این حالت، روش LSTM با ۲۵۰۱ مدل دارای کارایی بهتری است.

تفسیر سطر شماره ۸ و ۹ جدول: چنانچه مشخصه هزینه‌ای «مالکیت موسسه درمانی طرف قرارداد» به صورت سرجمع و ۳

جدول ۴ | رتبه‌بندی وضعیت شاخص MAPE بر اساس بالاترین تعداد مدل در ۱۰ حالت مختلف مشخصه‌های هزینه‌ای

ردیف	نام مشخصه هزینه‌ای	ویژگی مشخصه هزینه‌ای	روش پیش‌بینی کارا تر	تعداد کل مدل ایجاد شده	تعداد مدل در هر روش	
۱	تمامی مشخصه‌ها	سرجمع	LSTM	۱	۱	
۲			LSTM		۲۳۱۷	۱۳۳۲
۳			SARIMA			۹۸۵
۴	استان ارائه خدمت درمانی	سرجمع	SARIMA	۱۷۹	۱۱۱	
۵			LSTM		۶۸	
۶			LSTM		۴۷۹۴	۲۵۰۱
۷	SARIMA	۲۲۹۳				
۸	مالکیت موسسه درمانی طرف قرارداد	سرجمع	SARIMA	۱۳۵۷	۷۵۱	
۹			LSTM		۶۰۶	
۱۰			LSTM		۳۶۱۶	۱۹۶۳
۱۱	SARIMA	۱۶۵۳				
۱۲	صندوق بیمه‌ای	سرجمع	SARIMA	۱۱۲	۵۹۵	
۱۳			LSTM		۵۱۷	
۱۴			LSTM		۳۸۶۱	۲۰۵۲
۱۵	SARIMA	۱۸۰۹				
۱۶	خدمت درمانی	سرجمع	SARIMA	۶۳۳	۳۳۰	
۱۷			LSTM		۳۰۳	
۱۸			LSTM		۴۳۴۰	۲۲۶۶
۱۹	SARIMA	۲۰۷۴				



هزینه‌های دیگر به صورت سرجمع و تفکیک در نظر گرفته شوند، ۳۸۶۱ مدل در هر روش ایجاد می‌شود که در این حالت، روش LSTM با ۲۰۵۲ مدل دارای کارایی بهتری است.

تفسیر سطر شماره ۱۶ و ۱۷ جدول: چنانچه مشخصه هزینه‌های «خدمت درمانی» به صورت سرجمع و ۳ مشخصه هزینه‌های دیگر به صورت سرجمع و تفکیک در نظر گرفته شوند، ۶۳۳ مدل در هر روش ایجاد می‌شود که در این حالت، روش SARIMAX با ۳۳۰ مدل دارای کارایی بهتری است.

تفسیر سطر شماره ۱۸ و ۱۹ جدول: چنانچه مشخصه هزینه‌های «خدمت درمانی» به صورت تفکیک و ۳ مشخصه هزینه‌های دیگر به صورت سرجمع و تفکیک در نظر گرفته شوند، ۴۳۴۰ مدل در هر روش ایجاد می‌شود که در این حالت، روش LSTM با ۲۲۶۶ مدل دارای کارایی بهتری است.

#### بحث

همان‌طور که در بررسی‌ها و ارزیابی‌های انجام‌شده تاکنون مشاهده شده، حالت‌های متنوعی برای انجام پیش‌بینی هزینه‌ها در سطح سازمانی وجود دارد. به همین دلیل بر اساس تحلیل‌های فوق به راحتی نمی‌توان یک روش قطعی برای ارائه مدل مطلوب پیش‌بینی هزینه‌ها تعیین کرد و قواعدی که از تحلیل‌های فوق به دست آمده، فقط یک معیار نسبی برای انتخاب مدل را می‌دهد.

انتخاب مدل بهینه بر اساس شاخص MAPE

بر اساس آمار ارائه‌شده فوق به منظور انتخاب مدل بهینه بر اساس شاخص MAPE می‌توان قواعد زیر را استخراج کرد:

- چنانچه در انجام عملیات پیش‌بینی هر ۴ مشخصه

جدول ۵ | درصد تحقق پیش‌بینی بر اساس ویژگی‌های

وضعیت استان ارائه خدمت	وضعیت نوع مالکیت مؤسسات درمانی طرف قرارداد	وضعیت نوع صندوق بیمه‌ای	وضعیت نوع خدمت	تعداد کل مدل‌های ایجاد شده در این حالت	تعداد دارای کارایی بیشتر با روش SARIMAX	تعداد دارای کارایی بیشتر با روش LSTM	درصد تحقق پیش‌بینی
تفکیک	سرجمع	تفکیک	سرجمع	۱۲۴	۶۳	۶۱	۹۹.۳
تفکیک	تفکیک	تفکیک	سرجمع	۳۶۵	۱۸۵	۱۸۰	۹۸.۲
سرجمع	تفکیک	سرجمع	تفکیک	۲۴	۱۸	۶	۹۸.۱
تفکیک	سرجمع	سرجمع	سرجمع	۳۱	۱۷	۱۴	۹۷
تفکیک	تفکیک	سرجمع	سرجمع	۹۳	۵۳	۴۰	۹۶.۸
سرجمع	تفکیک	سرجمع	سرجمع	۳	۱	۲	۹۶.۷
تفکیک	تفکیک	سرجمع	تفکیک	۷۰۷	۳۴۵	۳۶۲	۹۶.۵
سرجمع	سرجمع	تفکیک	تفکیک	۳۲	۱۷	۱۵	۹۶.۴
تفکیک	سرجمع	سرجمع	تفکیک	۲۴۵	۱۵۵	۹۰	۹۶.۳
سرجمع	سرجمع	سرجمع	سرجمع	۱	۰	۱	۹۶.۱
سرجمع	تفکیک	تفکیک	تفکیک	۹۵	۵۸	۳۷	۹۶
سرجمع	سرجمع	سرجمع	تفکیک	۸	۶	۲	۹۵.۸
تفکیک	تفکیک	تفکیک	تفکیک	۲۳۱۷	۹۸۵	۱۳۳۲	۹۵.۲
سرجمع	تفکیک	تفکیک	سرجمع	۱۲	۸	۴	۹۴.۵
سرجمع	سرجمع	تفکیک	سرجمع	۴	۳	۱	۹۴.۵
تفکیک	سرجمع	تفکیک	تفکیک	۹۱۲	۴۹۰	۴۲۲	۹۱.۷

وجود دارند اما قابل تعمیم به کل هزینه‌های سازمان نیستند.

### نتیجه‌گیری

همان‌طور که مشاهده شد، در روش ترکیبی با استفاده از شاخص MAPE بالاترین میزان «درصد تحقق پیش‌بینی» به دست آمده است. بنابراین در اینجا می‌توان مدل و روش مطلوب و بهینه برای انجام پیش‌بینی هزینه‌ها را به شرح زیر ارائه کرد: «برای انجام عملیات پیش‌بینی با مطلوب‌ترین روش و مدل می‌بایست مقادیر هزینه استفاده شده برای انجام پیش‌بینی در استان ارائه خدمت و نوع صندوق به صورت تفکیک و در نوع مالکیت مؤسسه درمانی و نوع خدمت درمانی به صورت سرجمع باشد و مقادیر پیش‌بینی شده در مدل‌ها (۱۲۴ مدل تولید شده) از روشی (LSTM یا SARIMAX) که بر اساس شاخص MAPE دارای کارایی بهتر است، استفاده شود، به نحوی که در این حالت و بر اساس مقطع زمانی داده‌های استفاده شده در این پژوهش، خطای پیش‌بینی فقط ۰,۷ درصد است.»

مهم‌ترین موضوع در مدیریت هزینه‌ها، استفاده از روشی است که بتواند پیش‌بینی هزینه‌ها را با کمترین خطا و انحراف انجام دهد. در این پژوهش با استفاده از روش ترکیبی، میزان این انحراف در زمان استفاده از بهینه‌ترین روش به ۰,۷ درصد رسیده است. در فرآیندهای دستی پیش‌بینی هزینه‌ها که معمولاً برای تأمین منابع مالی و اعتبارات انجام می‌شود، مقدار پیش‌بینی با یک فاصله اطمینان بالاتر در نظر گرفته می‌شود و همین بالاتر دیدن منابع و اعتبارات، خود می‌تواند یکی از علل ایجاد درخواست‌های کاذب باشد. در سازمان بیمه سلامت ایران میزان انحراف از پیش‌بینی در مقادیر حدود ۵ درصد قابل قبول است. اما در این روش این مقدار به حداقل مقدار خود یعنی ۰,۷ درصد رسیده است.

**تشکر و قدردانی:** نویسندگان این مقاله از کلیه کارکنان معاونت IT تشکر می‌کنند. همچنین پژوهش حاضر حاصل پایان‌نامه در مقطع کارشناسی ارشد در دانشگاه علم و صنعت است.

**سهم نویسندگان:** موردی گزارش نشده است.

**تأیید به اخلاقی:** موردی گزارش نشده است.

**تعارض منافع:** نویسندگان این منافع هیچ‌گونه تعارضی اعلام نکرده‌اند.

**منابع مالی:** پژوهش حاضر مورد حمایت مالی قرار نگرفته است.

### References

- Masoudi Asl Y, Akhavan Behbahani A, Rahbari M. Comparative health care systems. Tehran: Satel; 2015,

آنچه اهمیت دارد رسیدن به روش یا مدلی است که بتواند پیش‌بینی هزینه‌های کلی را بالاترین دقت (انحراف حداقلی) مقادیر پیش‌بینی با مقادیر واقعی) انجام دهد. بنابراین با توجه به اینکه هر مدل برای گروه مشخصه‌های هزینه‌ای مربوطه، بهترین مدل بر اساس هر دو روش است، پیشنهاد این پژوهش استفاده از روش ترکیبی برای داشتن دقیق‌ترین مقادیر پیش‌بینی است. بدین صورت که در هر گروه از مشخصه‌هایی هزینه‌ای (حداکثر ۴۹۷۳ گروه) بر اساس روش با کارایی بهتر، مقادیر پیش‌بینی در نظر گرفته شده و بر اساس یکی از ۱۶ حالت ترکیبی سرجمع یا تفکیکی بودن مشخصه‌های هزینه‌ای، سرجمع هزینه‌های پیش‌بینی محاسبه شود.

در این روش برای انتخاب بهترین حالت برای انجام پیش‌بینی، معیار درصد تحقق پیش‌بینی که در فرآیندهای دستی استفاده می‌شود، به کار گرفته می‌شود. خلاصه محاسبات فوق به شرح جدول ۵ است.

مقادیر جدول ۵ بر اساس بالاترین مقدار «درصد تحقق پیش‌بینی» مرتب شده است. همان‌طور که مشاهده می‌شود، بالاترین درصد تحقق با ۹۹,۳ درصد (۰,۷ درصد انحراف پیش‌بینی) بر اساس شاخص MAPE در روش ترکیبی، مربوط به حالتی است که مشخصه‌های هزینه استان ارائه خدمت و نوع صندوق به صورت تفکیک و نوع مالکیت مؤسسه درمانی و نوع خدمت به صورت سرجمع استفاده شوند.

در این پژوهش کد نرم‌افزاری مورد نیاز به منظور ایجاد مدل، با استفاده از زبان برنامه‌نویسی پایتون به نحوی نوشته شده که با وارد کردن داده‌ها با ساختار مورد استفاده در این پژوهش و تنظیم چند پارامتر خاص، تعمیم آن به سایر موضوعات و همچنین توسعه آن برای خود موضوع پژوهش، به سهولت امکان‌پذیر است.

در اکثر پژوهشات و مطالعات صورت گرفته که کمتر در مورد موضوع این پژوهش است، با استفاده از روش‌های مورد نظر معدودی مدل ایجاد شده و مدل‌های ایجاد شده بر اساس یک شاخص به منظور انتخاب مدل مطلوب مورد ارزیابی قرار گرفته‌اند. اما در این پژوهش چنانچه داده‌ها مطابق با ساختار داده ورودی تهیه شوند، بر اساس تعداد رکوردهای داده‌ها، تعداد مدل‌های خروجی قابل تغییر و بدون محدودیت است.

باتوجه به بازه در نظر گرفته شده برای داده‌ها، تعداد داده‌ها در هر بار ایجاد مدل حداکثر ۱۶۸ (۱۴ سال در ۱۲ ماه) داده است. در زمان ارائه این پژوهش، امکان داشتن کل هزینه‌های سازمان در ابتدایی‌ترین حالت فقط به صورت ماهانه است. بعضی از هزینه‌ها که به صورت ریز نسخ در نظر گرفته می‌شوند، به صورت روزانه

- 2017;2017:1312-21.
13. Richardson R, Hartman B. Bayesian nonparametric regression models for modeling and predicting healthcare claims. *Insur Math Econ.* 2018;83:1-8. doi: [10.1016/j.insmatheco.2018.06.002](https://doi.org/10.1016/j.insmatheco.2018.06.002)
  14. Siami-Namini S, Tavakoli N, Siami Namin A. A Comparison of ARIMA and LSTM in Forecasting Time Series. In *IEEE*; 2018, Pp. 1394-401. doi: [10.1109/ICMLA.2018.00227](https://doi.org/10.1109/ICMLA.2018.00227).
  15. Klazoglou P, Dritsakis N. Modeling and Forecasting of US Health Expenditures Using ARIMA Models. In: Tsounis N, Vlachvei A, editors. *Advances in Panel Data Analysis in Applied Economic Research*. Cham: Springer International Publishing; 2018, Pp. 457-72. doi: [10.1007/978-3-319-70055-7\\_36](https://doi.org/10.1007/978-3-319-70055-7_36).
  16. Zinouri N, Taaffe KM, Neyens DM. Modelling and forecasting daily surgical case volume using time series analysis. *Health Syst.* 2018;7(2):111-9. doi: [10.1080/20476965.2017.1390185](https://doi.org/10.1080/20476965.2017.1390185).
  17. Cui L, Xie X, Shen Z, Lu R, Wang H. Prediction of the healthcare resource utilization using multi-output regression models. *IISE Trans Healthc Syst Eng.* 2018;8(4):291-302. doi: [10.1080/24725579.2018.1512537](https://doi.org/10.1080/24725579.2018.1512537).
  18. Wang W, Bjarnadottir M, Gao GG. How AI Plays Its Tricks: Interpreting the Superior Performance of Deep Learning-Based Approach in Predicting Healthcare Costs. *SSRN Electronic Journal.* 2018;1-17. doi: [10.2139/ssrn.3274094](https://doi.org/10.2139/ssrn.3274094).
  19. Yousefi M, Yousefi M, Fathi M, Fogliatto FS. Patient visit forecasting in an emergency department using a deep neural network approach. *Kybernetes.* 2019;49(9):2335-48. doi: [10.1108/K-10-2018-0520](https://doi.org/10.1108/K-10-2018-0520).
  20. Deng Y, Fan H, Wu S. A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits. *J Ambient Intell Humaniz Comput.* 2020. doi: [10.1007/s12652-020-02602-x](https://doi.org/10.1007/s12652-020-02602-x).
  21. Kaushik S, Choudhury A, Sheron PK, Dasgupta N, Natarajan S, Pickett LA, et al. AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures. *Front Big Data.* 2020;3:4. doi: [10.3389/fdata.2020.00004](https://doi.org/10.3389/fdata.2020.00004).
  22. Yucesan M, Gul M, Celik E. A multi-method patient arrival forecasting outline for hospital emergency departments. *Int J Healthc Manag.* 2020;13(sup1):283-95. doi: [10.1080/20479700.2018.1531608](https://doi.org/10.1080/20479700.2018.1531608).
  - Pp. 437. [Persian]
  2. Yousefi M, Yousefi M, Fathi M, Fogliatto FS. Patient visit forecasting in an emergency department using a deep neural network approach. *Kybernetes.* 2019;49(9):2335-48. doi: [10.1108/K-10-2018-0520](https://doi.org/10.1108/K-10-2018-0520).
  3. Kaushik S, Choudhury A, Sheron PK, Dasgupta N, Natarajan S, Pickett LA, et al. AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures. *Front Big Data.* 2020;3:4. doi: [10.3389/fdata.2020.00004](https://doi.org/10.3389/fdata.2020.00004).
  4. Yucesan M, Gul M, Celik E. A multi-method patient arrival forecasting outline for hospital emergency departments. *Int J Healthc Manag.* 2020;13(sup1):283-95. doi: [10.1080/20479700.2018.1531608](https://doi.org/10.1080/20479700.2018.1531608).
  5. Deng Y, Fan H, Wu S. A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits. *J Ambient Intell Humaniz Comput.* 2020. doi: [10.1007/s12652-020-02602-x](https://doi.org/10.1007/s12652-020-02602-x).
  6. Cao Q, Ewing BT, Thompson MA. Forecasting medical cost inflation rates: A model comparison approach. *Decis Support Syst.* 2012;53(1):154-60. doi: [10.1016/j.dss.2011.12.012](https://doi.org/10.1016/j.dss.2011.12.012).
  7. Chaabouni S, Abednadhher C. Modelling and forecasting of Tunisia's health expenditures using artificial neural network and ARDL models. *Int J Med Sci Public Health.* 2013;2(3):495. doi: [10.5455/ijmsph.2013.080720131](https://doi.org/10.5455/ijmsph.2013.080720131).
  8. Xu Q, Tsui K-L, Jiang W, Guo H. A Hybrid Approach for Forecasting Patient Visits in Emergency Department: Forecasting Patient Visits. *Qual Reliab Eng Int.* 2016;32(8):2751-9. doi: [10.1002/qre.2095](https://doi.org/10.1002/qre.2095).
  9. Chauhan A, Singh A. An ARIMA model for the forecasting of healthcare waste generation in the Garhwal region of Uttarakhand, India. *International Journal of Services Operations and Informatics.* 2017;8(4):352. doi: [10.1504/IJSOI.2017.086587](https://doi.org/10.1504/IJSOI.2017.086587).
  10. 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. doi: [10.1109/JBHI.2015.2511820](https://doi.org/10.1109/JBHI.2015.2511820).
  11. Langat A, Orwa G, Koima J. Cancer Cases in Kenya; Forecasting Incidents Using Box & Jenkins Arima Model. *Biomedical Statistics and Informatics.* 2017;2(2):37-48
  12. Morid MA, Kawamoto K, Ault T, Dorius J, Abdelrahman S. Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation. *AMIA Annu Symp Proc.*