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

Maryam Arab[1], Mohammad Fathian[1*], Hosein Aliahmadi Jeshfaghani[1]

[1] 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

Copyright © 2022 The Author(s); Published by National Center for Health Insurance Research. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

پیش‌بینی هزینه‌های درمانی سازمان بیمه سلامت ایران با استفاده از روش‌های مبتنی بر یادگیری ماشین

مریم عرب[۱]، محمد فتحیان[*۱]، حسین علی احمدی جشفقانی[۱]

[۱] 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.
پست‌الکترونیک: fathian@iust.ac.ir

تاریخ دریافت: ۱۴۰۰/۱۱/۲۵ **تاریخ تصحیح:** ۱۴۰۰/۱۰/۱۵ **تاریخ پذیرش:** ۱۴۰۰/۱۱/۰۶ **تاریخ انتشار آنلاین:** ۱۴۰۰/۱۲/۱۸

چکیده

مقدمه: تأمین دقیق منابع مالی به‌منظور مدیریت بهتر هزینه‌ها یکی از دغدغه‌های اصلی مدیران سازمان‌هاست. سازمان بیمه سلامت ایران با عنوان یکی از بزرگ‌ترین سازمان‌های بیمه‌گر پایه از این امر مستثنی نبوده و قطعاً برای تأمین منابع مالی و اخذ بودجه‌های لازم در حوزه درمان خود، نیازمند شناسایی و پیش‌بینی دقیق هزینه‌های درمان است. استفاده از روش‌های مبتنی بر یادگیری ماشین به‌منظور ایجاد مدل پیش‌بینی هزینه‌های درمان می‌تواند کمک بزرگی به تامین دقیق‌تر منابع مالی کند.

روش‌بررسی: این پژوهش با استفاده از داده‌های هزینه‌ای موجود در سامانه اسناد پزشکی استان‌های سازمان طی سال‌های ۱۳۸۵ تا ۱۳۹۸ و با استفاده از روش‌های SARIMAX و LSTM، مدل و روشی را برای پیش‌بینی هزینه‌های سازمان ارائه داده است. این روش می‌تواند به پیش‌بینی دقیق‌تر هزینه‌های سازمان کمک کند.

یافته‌ها: مشخص کردن روش با کارآیی بهتر براساس شاخص MAPE به تنهایی جوابگوی ایجاد مدل مطلوب نیست، بنابراین با ایجاد یک روش ترکیبی و استفاده از معیار درصد تحقق پیش‌بینی، مدل مطلوب برای پیش‌بینی هزینه‌ها ارائه شده است.

نتیجه‌گیری: باتوجه به ضرورتِ داشتن روش علمی به منظور پیش‌بینی دقیق‌تر هزینه‌های سازمان، روش و مدل پیشنهاد شده توانست با حداقل خطا نسبت به خطاهای پذیرفته شده در فرآیندهای دستی، هزینه‌های سازمان را پیش‌بینی کند.

واژگان کلیدی: سازمان بیمه سلامت ایران، هزینه‌های درمانی، پیش‌بینی، یادگیری ماشین، سری زمانی

مقدمه

رشد پرشتاب هزینه‌های بهداشتی و درمانی در کشور به‌طور عمده ناشی از عوامل مختلفی مانند رشد جمعیت، سالمند شدن جمعیت، تغییر الگوی بیماری‌ها و گسترش بیماری‌های غیرواگیر و مزمن، تغییر سطح انتظارات، گرایش روزافزون به استفاده از تجهیزات و فنّاوری نوین و گران‌قیمت درمانی است. در چنین شرایطی توجه به مکانیسم تأمین منابع مالی در بخش بهداشت و درمان بیش‌ازپیش اهمیت یافته است. باتوجه به روند اعتبارات و تنوع منابع در بخش بهداشت و درمان، بر اساس آنچه در چارچوب حساب‌های ملی بهداشت و درمان کشور مشاهده می‌شود، پایین بودن سهم اعتبارات تخصیصی از محل بودجه عمومی دولت و منابع پیش‌پرداخت در بیمه‌های اجتماعی درمانی و بالا بودن سهم هزینه‌های خانوارها از مجموع هزینه‌های مصرفی بهداشت و درمان کشور، از عوامل محدودکننده در تأمین مالی هزینه‌های مصرفی در بخش سلامت محسوب می‌شود [۱].

یکی از راهکارها برای برخورد با کسری منابع مالی، مدیریت و کنترل هزینه‌ها بر اساس پیش‌بینی دقیق از هزینه‌هاست. در تمامی سازمان‌ها حجم انبوهی از داده وجود دارد که استفاده و تجزیه‌وتحلیل آنها می‌تواند کمک شایان توجهی به تصمیم‌گیری مدیران داشته باشد، اما از طرفی حجم بالای این داده‌ها 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 داده‌ها را با روش‌های قدیمی، خارج از توان افراد و ناکارآمد کند. پیشرفت فنّاوری و ایجاد

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 درنتیجه 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 درخصوص 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 دو روش کیفی و کمی تقسیم کرد. روش‌های کیفی پیش‌بینی مبتنی بر تخمین ذهنی از طریق نظرات و عقاید متخصصان در موضوع موردنظر است. ازاین‌رو در روش‌های کیفی، رویه‌های سیستماتیکی 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 درنهایت 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 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. تجزیه‌وتحلیل 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. درواقع سری زمانی، زمان پیوسته است هرگاه مشاهدات در یک‌فاصله زمانی به‌طور پیوسته جمع‌آوری‌شده باشند. یک سری زمانی زمان گسسته است، هرگاه مشاهدات فقط در زمان‌های گسسته متعلق به مجموعه‌ای شما را جمع‌آوری کرده‌باشند. اکثر سری‌های زمانی که با آن سروکار داریم از نوع گسسته

هســتند زیرا آنها به‌صورت ساعتی، روزانه، ماهانه، فصلی، سالانه و... ثبت‌شده‌اند [۴].

اولیــن و معروف‌تریــن مدل مورداســتفاده در مدل‌ســازی سری‌های زمانی، مدل ARMA است. در این مدل فرض می‌شود کــه فرآیند تولید داده‌های ســری زمانی مبتنــی 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 شــبکه‌های عصبی مصنوعی که قابلیت تخمیــن مدل پیچیده و غیرخطــی را دارند گســترش یافته، اما همچنــان به دو دلیل عمده اســتفاده از مدل‌های خطی در کارهای تجربی رواج دارد؛ اول اینکه بســیاری از اوقات تقریب خطــی مدل‌های غیرخطی ســاده‌تر بوده و به‌طورمعمول اطلاعات چندانی از دست نمی‌رود. دوم اینکه با کنار گذاشــتن 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 موردنظر دارد [۲].

از طرفی ماهیت روابط، تشــریح‌کننده بسیاری از فرآیندها و ســری‌های زمانــی اغلب غیرخطی اســت. بــه همین دلیلی پیش‌بینی رفتار چنین فرآیندها و سری‌هایی نیازمند ابزارهای پیش‌بینــی دقیــق و اثربخــش اســت. درواقع، در بســیاری از سیســتم‌های پیچیده و خصوصاً غیرخطی که مدل‌ســازی و به دنبال آن پیش‌بینی و کنترل آنها از طریق روش‌های کلاسیک رگرســیونی امری بسیار دشــوار حتی بعضاً غیرممکن است، از روش‌های غیرکلاسیک که به‌طور عمده برگرفته از رفتار طبیعت هستند، استفاده می‌شود. ازجمله این روش‌ها، شبکه‌های عصبی مصنوعی اســت که یکــی از روش‌های بدیــع و در حال تحول محســوب می‌شــود و در موضوعات متنوعی از قبیل شــناخت الگو، خوشه‌بندی، پیش‌بینی و... به‌کاررفته و نتایج مفیدی نیز داشته است [۲].

یادگیری ماشــین، کاربــرد الگوریتم‌هایی اســت که برای ما حجــم عظیمی از داده‌ها را به دانــش تبدیل می‌کنند. یادگیری ماشــین تلاش می‌کند برنامه‌هایی را طراحــی کند که از داده‌ها آموزش ببینند. به‌عبارت‌دیگر، 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 در دسترس 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].

در ادامه به‌منظور مشخص‌تر شــدن روش انجام پیش‌بینی، به بررســی مطالعات انجام شــده پرداخته شده است. در بررسی مطالعات ســعی شده مطالعات انجام‌شده کاملاً مرتبط با مدل‌ها و الگوریتم‌های استفاده شــده درخصوص پیش‌بینی هزینه‌های درمان در ســازمان‌های بیمه پایه باشد اما به دلیل محدودیت در مطالعات موجود، باتوجه به اهمیت تبیین مدل‌ها و الگوریتم‌های بــه کار گرفته شــده درخصوص پیش‌بینی، برخــی از مطالعات درخصوص 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 ســلامت به موضوعاتی مانند پیش‌بینی، تشــخیص و جلوگیری از بیماری‌ها، پیش‌بینی هزینه‌های درمانی یک بیماری خاص و...
۳. گرایش اکثر موضوعات حــوزه بیمه به موضوعات حوزه بیمه تجاری مانند مدیریت ارتباط با مشتریان و مدیریت ریسک و...
۴. مــوردی بودن مطالعه و فقئتم امکان تعمیم و توســعه آن به موضوعات مشابه
۵. نبــود عوامل مؤثر سیاســی و اقتصادی خــاص ایران بر ایجاد هزینه‌های درمانی در مطالعات خارج از کشور
۶. وجود محدودیت در دسترسی به اطلاعات و دیتاهای موردنیاز جهت انجام مطالعه
۷. نبــود مدل علمــی و دقیق به‌منظور پیش‌بینــی هزینه‌ها در سازمان بیمه سلامت

از این رو باتوجه به دلایل مطرح‌شده دربندهای 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 ســازمان بیمه درمانی دولتی و توســعه مدلی علمی و دقیق و براســاس تکنیک‌های روز داده‌کاوی برای پیش‌بینی هزینه‌ها، ضرورت انجام پژوهش واضح و مبرهن به نظر می‌رسد.

همچنیــن با بررســی مطالعــات فوق، می‌توان لیســتی از روش‌ها و مدل‌های به‌کار گرفته‌شــده برای انجام پیش‌بینی را تهیه کرد. همان‌طور که مشــخص اســت، در اکثر مطالعات از روش‌های آماری مبتنی بر ARIMA به‌همراه روش‌های مبتنی بر شــبکه‌های عصبی استفاده شــده، بنابراین در این پژوهش بــا توجه اینکــه داده‌هــای موردپژوهش دارای ماهیت ســری زمانی هستند، از روش‌های توسعه‌یافته کلاسیک آماری مانند SARIMAX «میانگیــن متحرک خود توضیح فصلی با متغیر برون‌زا» و روش‌های مبتنی بر شبکه‌های عصبی مصنوعی مانند LSTM «حافظــه کوتاه‌مدت بلند» برای ارائه مدل پیش‌بینی نهایی استفاده شده است.

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 استفاده‌شده است.

آماده‌سازی داده‌ها

داده‌های مورداستفاده 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 به‌صورت ماهانه ثبت می‌شوند. هزینه‌هایی که در این پژوهش مورد استفاده قرارگرفته، 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 به‌صورت ماهانه گروه‌بندی‌شده 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 به‌منظور 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 ابدننادهر (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	چاوهان و سنق (۲۰۱۷) [۹]	داده‌های مربوط به مواد زاید بهداشتی و درمانی تولیدشده در بیمارستان‌های منطقه گرهوال 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 یوتا از اکتبر ۲۰۱۳ تا اکتبر ۲۰۱۶ شامل حدود ۹۰.۰۰۰ بیمه‌شده، ۶.۳ میلیون مطالبات درمانی و ۱.۲ 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	سیامی-نمینی و دیگران (۲۰۱۸) [۱۴]	بر اساس داده‌های مالی	ARIMA، SES، MA ، AR و LSTM
۱۰	کلازوگلو و دریتسکیس (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 ۷۵۰.۰۰۰ 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 ۱.۴۳۴.۹۱۲ بیمه‌شده شامل اطلاعات مربوط به ۶۲.۴۰۶.۳۷۹ 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	SVR، GLM، GEE، SARIMA، MLR، ARMA و SARIMA-LR
۱۵	دنگ و دیگران (۲۰۲۰) [۲۰]	داده‌های مربوط به بارمراجعه 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 ARIMA، MLP و LSTM
۱۷	یوکاسان و دیگران (۲۰۲۰) [۲۲]	داده‌های پذیرش بیماران در بخش اورژانس یک بیمارستان خصوصی در ترکیه	ANN، ARMA، LR، هموارسازی نمایی و روش‌های ترکیبی مانند ARMA-ANN و ARMA-LR

برای ایجاد مدل پیش‌بینی از زبان برنامه‌نویسی پایتون نسخه ۳,۹ اســتفاده 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 به‌منظور ایجاد مدل به شــرح ذیل به دو دسته آموزش و آزمون تقسیم شده‌اند:

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

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

ایجاد مدل با توابع پایتون

در این مرحله با اســتفاده از توابع موجود در پایتون و تعیین مقادیر مناســب برای پارامترهای ایــن توابع و داده‌های آموزش، مدل‌هــای موردنظر با اســتفاده از هــر دو روش SARIMAX و LSTM ایجاد می‌شود.

در ایــن پژوهش برای ارزیابی دقیق مدل و مشــخص کردن تأثیری که هر یک از مشــخصه‌های هزینه‌ای محل ارائه خدمات درمانی، نوع مالکیت مؤسسات طرف قرارداد، نوع صندوق بیمه‌ای و نوع خدمات ارائه‌شــده بر ایجاد هزینه‌های درمانی دارند، برای هــر ســری از اطلاعات بر اســاس گروه‌بندی صــورت گرفته بر اساس عناوین مشــخصه‌های هزینه‌ای، یک مدل با هر دو روش SARIMAX و LSTM ایجــاد می‌شــود. حداکثــر تعداد مدل در صورت وجود داده در هر گروه، برابر ۵۷۶۰ مدل است اما به دلیل فقــدان داده در برخی از گروه‌ها، تعداد ۴۹۷۳ مدل در هر روش ایجاد شده است.

جدول ۲| ساختار داده‌های پژوهش

تاریخ	مقادیر ۰۱-۰۴-۲۰۰۶ لغایت ۰۱-۰۳-۲۰۲۰ (۱۳۸۵-۱۳۹۸)
استان ارائه خدمات درمانی	مقادیر: -۰کلیه استان‌ها و ۱ تا ۳۱ برای ۳۱ استان کشور
نوع اسناد درمانی	مقادیر: ۱-غیر خسارت متفرقه
نوع مالکیت مؤسسات درمانی	مقادیر: -۰کلیه مالکیت‌ها -۱دولتی دانشگاهی -۲دولتی غیردانشگاهی -۳خصوصی و خیریه
صندوق بیمه‌ای	مقادیر: ۰-کلیه صندوق‌ها ۱-کارکنان دولت ۲-خویش‌فرما ۳-سایراقشار 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 پرتوپزشکی
هزینه پرداختی مؤسسات توان‌بخشی	هزینه پرداختی خدمات بستری
کل هزینه پرداختی	

نکته مهم در ایجاد هر مدل در هر روش این است که باتوجه 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 مدل‌هــا تکه کدهای مربوط به ایجاد مدل با هــر دو روش در حلقه‌های تودرتو اجراشــده و همزمان با ایجاد، برازش و پیش‌بینی مدل، شــاخص ارزیابی مدل نیز محاســبه و تمامی مقادیر لازم در جداولی به‌منظور 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 روی‌داده‌های آزمون انجام می‌شود و درنهایت خروجی‌های مدل استخراج و جمع‌آوری می‌شود.

شاخص‌های ارزیابی مدل

برای انتخاب مدل با کارآیی 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 اســتخراج آمارهای موردنیاز، اطلاعات به محیط SQL Server منتقل‌شده و در آنجا با توجه به حالت‌های مختلفی کــه برای انجام پیش‌بینــی در نظر گرفته شــده 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 اســتفاده می‌شــود. 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.

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

شاخص فوق کارآیی بیشتر یک مدل را دریکی 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 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 کارآیی بهتر مشخص می‌شود. سپس برای پیش‌بینی هزینه‌های کلی و مشخص کردن روش با کارآیی 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 درنظر 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.

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

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

تفسیر سطر شماره ۲ و ۳ جدول: در حالتی که هر ۴ مشخصه هزینه‌ای به‌صورت تفکیک درنظر گرفته شوند، ۲۳۱۷ مدل در هر روش ایجاد می‌شود که در این حالت، روش LSTM با ۱۳۳۲ دارای کارآیی بهتری است.

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

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

تفسیر سطر شماره ۸ و ۹ جدول: چنانچه مشخصه هزینه‌ای «مالکیت موسسه درمانی طرف قرارداد» به‌صورت سرجمع و ۳ مشخصه هزینه‌ای دیگر به‌صورت سرجمع و تفکیک درنظر گرفته شوند، ۱۳۵۷ مدل در هر روش ایجاد می‌شود که در این حالت، روش SARIMAX با ۷۵۱ مدل دارای کارآیی بهتری است.

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

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

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

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

هزینه‌ای دیگر به‌صورت سرجمع و تفکیک درنظر گرفته شوند، ۳۸۶۱ مدل در هر روش ایجاد می‌شود که در این حالت، روش LSTM با ۲۰۵۲ مدل دارای کارآیی بهتری است.

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

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

انتخاب مدل بهینه بر اساس شاخص MAPE

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

- چنانچه در انجام عملیات پیش‌بینی هر ۴ مشخصه هزینه‌ای به‌صورت سرجمع مورد استفاده قرار گیرند، روش LSTM دارای کارآیی بهتری نسبت به روش SARIMAX است.
- در حالت‌هایی که هر مشخصه هزینه‌ای به‌صورت تفکیک مورد استفاده قرار گرفته‌اند. روش LSTM دارای کارآیی بهتری نسبت به روش SARIMAX است.
- در حالت‌هایی که هر مشخصه هزینه‌ای به‌صورت سرجمع مورد استفاده قرارگرفته‌اند. روش SARIMAX دارای کارآیی بهتری نسبت به روش LSTM 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 به‌راحتی نمی‌توان یک روش قطعی برای ارائه مدل مطلوب پیش‌بینی هزینه‌ها تعیین کرد و قواعدی که از تحلیل‌های فوق به‌دست آمده، فقط یک معیار نسبی برای انتخاب مدل را می‌دهد.

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 کارآیی بیشتر با روش SARIMAX	تعداد دارای کارآیی 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

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 باتوجه 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. بدین‌صورت که در هر گروه از مشخصه‌هایی هزینه‌ای (حداکثر ۴۹۷۳ گروه) بر اساس روش با کارآیی 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، به‌کار گرفته می‌شود. خلاصه محاسبات فوق به شرح جدول ۵ است.

مقادیر جدول ۵ بر اساس بالاترین مقدار «درصد تحقق پیش‌بینی» مرتب‌شده است. همان‌طور که مشاهده می‌شود، بالاترین درصد تحقق با ۹۹٫۳درصد (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 به‌صورت تفکیک و نوع مالکیت موسسه درمانی و نوع خدمت به‌صورت سرجمع استفاده شوند.

در این پژوهش کد نرم‌افزاری موردنیاز به‌منظور 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 درمورد موضوع این پژوهش است، با استفاده از روش‌های موردنظر معدودی مدل ایجاد شده و مدل‌های ایجاد شده بر اساس یک شاخص به‌منظور 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 به‌صورت ماهانه است. بعضی از هزینه‌ها که به‌صورت ریز نسخ در نظر گرفته می‌شوند، به‌صورت روزانه وجود دارند اما قابل‌تعمیم به‌کل هزینه‌های سازمان نیستند.

نتیجه‌گیری

همان‌طور که مشاهده شد، در روش ترکیبی با استفاده از شاخص MAPE بالاترین میزان «درصد تحقق پیش‌بینی» به‌دست‌آمده 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 به‌صورت تفکیک و در نوع مالکیت مؤسسه درمانی و نوع خدمت درمانی به‌صورت سرجمع باشد و مقادیر پیش‌بینی‌شده در مدل‌ها (۱۲۴ مدل تولیدشده) از روشی (LSTM یا SARIMAX) که بر اساس شاخص MAPE دارای کارآیی بهتر است، استفاده شود، به نحوی که در این حالت و بر اساس مقطع زمانی داده‌های استفاده شده در این پژوهش، خطای پیش‌بینی فقط ۰٫۷درصد 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 ۵درصد قابل‌قبول است. اما در این روش این مقدار به حداقل مقدار خود یعنی ۰٫۷درصد 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.

تأییدیه 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

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

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.
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.
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.
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.
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.
7. Chaabouni S, Abednnadher 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.
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.
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.
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.
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. 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
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.
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.
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.
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.
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.
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.
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.
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.
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.