



توسعه چارچوب مبتنی بر الگوریتم‌های یادگیری ماشین برای بهینه‌سازی پیش‌بینی هزینه‌های سازمانی

احرار حسینی*

چکیده	اطلاعات مقاله
<p>در دنیای امروز، پیش‌بینی دقیق هزینه‌های سازمانی برای بهینه‌سازی تخصیص منابع و افزایش بهره‌وری، یکی از چالش‌های اصلی مدیران بشمار می‌رود. تغییرات سریع اقتصادی و پیچیدگی فزاینده فرآیندهای مالی، ضرورت استفاده از روش‌های پیش‌بینی داده‌محور را بیش از پیش برجسته کرده است. این پژوهش چارچوبی مبتنی بر الگوریتم‌های یادگیری ماشین ارائه می‌دهد که هدف آن بهبود دقت پیش‌بینی هزینه‌های سازمانی و تسهیل برنامه‌ریزی مالی دقیق‌تر در سازمان‌هاست. در این مطالعه، داده‌های واقعی مالی و عملیاتی از سازمان بیمه سلامت در یک بازه زمانی پنج‌ساله گردآوری و پس از انجام مراحل پیش‌پردازش، برای آموزش مدل‌های پیش‌بینی به‌کار گرفته شد. مدل اصلی مورد استفاده در این پژوهش، رگرسیون خطی بوده و عملکرد آن با استفاده از معیارهایی نظیر ضریب تعیین، میانگین خطای مطلق و ریشه میانگین مربعات خطا ارزیابی شده است. نتایج این پژوهش نشان می‌دهد که مدل رگرسیون خطی توانسته است با دقت بالایی تغییرات هزینه‌های سازمانی را پیش‌بینی کرده و عوامل کلیدی مؤثر بر هزینه‌ها را شناسایی کند. چارچوب پیشنهادی این مطالعه می‌تواند به سازمان‌ها در بهبود برنامه‌ریزی بودجه، مدیریت منابع مالی و کاهش خطاهای پیش‌بینی کمک کند. افزون بر این، یافته‌های پژوهش حاضر، پایه‌ای علمی برای توسعه مدل‌های پیش‌بینی در سایر حوزه‌های سازمانی فراهم می‌آورد و امکان ارتقای تصمیم‌گیری‌های استراتژیک را برای مدیران تسهیل می‌کند.</p>	<p>نوع مقاله: مقاله پژوهشی</p> <p>تاریخ دریافت: ۱۴۰۴/۰۴/۱۸</p> <p>تاریخ پذیرش: ۱۴۰۴/۰۶/۰۲</p>
<p>پیش‌بینی هزینه‌های سازمانی، یادگیری ماشین، مدل داده‌محور، بهینه‌سازی منابع، تحلیل داده‌ها، رگرسیون خطی، مدیریت مالی</p>	<p>کلیدواژه‌ها</p>

ناشر: دانشگاه آزاد اسلامی واحد شهرقدس

ایمیل: almastwo@yahoo.com

* نویسنده مسئول: احرار حسینی

۱. مرکز ملی تحقیقات بیمه سلامت، تهران، ایران (نویسنده مسئول)

مقدمه

مدیریت هزینه‌های سازمانی به یکی از مهم‌ترین چالش‌های مدیران در سازمان‌های دولتی و خصوصی تبدیل شده است. به اعتقاد عبدالله و همکاران^۱، پیش‌بینی دقیق هزینه‌ها نقشی کلیدی در بهینه‌سازی تخصیص منابع مالی ایفا می‌کند و مانع از هدررفت منابع می‌شود. در شرایط اقتصادی پرتلاطم و متغیر امروزی، کرلیف^۲ تأکید می‌کند که روش‌های سنتی پاسخگوی نیازهای پیچیده سازمان‌ها نیستند و تصمیمات نادرست مالی می‌تواند پیامدهای جدی برای عملکرد و بقای سازمان داشته باشد. در این میان، سازمان‌های ارائه‌دهنده خدمات عمومی مانند نهادهای سلامت، با چالش‌های بیش‌تری مواجه‌اند؛ بویژه آنکه محدودیت‌های بودجه‌ای، افزایش تقاضا برای خدمات درمانی، و تغییرات سریع در سیاست‌گذاری‌ها، پیش‌بینی دقیق هزینه‌ها را به ضرورتی انکارناپذیر تبدیل کرده است.

در پژوهش تانگ و همکاران^۳ اشاره شده است که روش‌های سنتی پیش‌بینی هزینه، از جمله تحلیل روندهای تاریخی، در مواجهه با روابط چندبعدی میان متغیرها عملکرد ضعیفی دارند. این روش‌ها نه تنها در تشخیص تغییرات ناگهانی و الگوهای غیرمعمول ناتوان‌اند بلکه معمولاً به داده‌های محدود متکی بوده و انعطاف‌پذیری لازم برای سازگاری با محیط‌های پویا را ندارند، موضوعی که لوسیوس و همکاران^۴ نیز بر آن تأکید کرده‌اند. این محدودیت‌ها، دقت پیش‌بینی‌ها را کاهش داده و می‌تواند منجر به تصمیم‌گیری‌های نادرست شود.

در نقطه مقابل، موکرچی و همکاران^۵ خاطرنشان می‌کنند که الگوریتم‌های یادگیری ماشین، با توانایی تحلیل داده‌های حجیم و شناسایی روابط پنهان، ابزارهایی کارآمد برای پیش‌بینی‌های دقیق و قابل اعتماد فراهم می‌کنند. این الگوریتم‌ها با سازگاری بالا نسبت به داده‌های جدید، دقت پیش‌بینی را در محیط‌های ناپایدار افزایش می‌دهند. از منظر آروپیتا و همکاران^۶ الگوریتم‌های یادگیری ماشین قادرند روابط پیچیده و غیرخطی میان متغیرها را درک کرده و الگوهای پنهان در داده‌ها را استخراج کنند؛ امری که موجب افزایش کارایی این روش‌ها نسبت به روش‌های سنتی می‌شود.

در همین راستا، کوهن^۷ بیان می‌کند که یکی از مزایای اصلی یادگیری ماشین نسبت به روش‌های سنتی، توان پردازش حجم بالای داده‌ها و شناسایی روابط چندلایه میان متغیرهاست. این الگوریتم‌ها می‌توانند در شرایطی که داده‌ها دچار تغییرات ناگهانی می‌شوند، ساختار خود را به‌روزرسانی کرده و عملکرد دقیق‌تری ارائه دهند. برای نمونه، در نهادهای سلامت، عواملی مانند نوسانات تعرفه‌های درمانی، تغییر در تعداد بیمه‌شدگان، یا اصلاحات سیاست‌های دولتی می‌تواند هزینه‌ها را به‌سرعت تحت تأثیر قرار دهد. یادگیری ماشین قادر است این تغییرات را پردازش و هزینه‌های آتی را با دقت بالایی پیش‌بینی کند.

این پژوهش با هدف توسعه چارچوبی مبتنی بر یادگیری ماشین برای بهینه‌سازی پیش‌بینی هزینه‌های سازمانی تدوین شده است. داده‌های واقعی مالی و عملیاتی گردآوری شده از یک سازمان خدمات سلامت در بازه زمانی پنج‌ساله، شامل اطلاعاتی مانند تعداد بیمه‌شدگان، نرخ تعرفه خدمات درمانی، بار مراجعات به پزشکان و سایر شاخص‌های کلیدی، به‌عنوان مبنای این مطالعه قرار گرفته است. این داده‌ها پس از پاک‌سازی و پردازش اولیه، به مدل‌های یادگیری ماشین وارد شدند تا روابط میان متغیرها تحلیل و پیش‌بینی‌های دقیق‌تری انجام شود.

1. Abdulla et al.

2. Krylov

3. Tang et al.

4. Lucasius et al.

5. Mukherjee et al.

6. Arpita et al.

7. Cohen

در این چارچوب، معیارهای ارزیابی گوناگونی همچون میانگین خطای مطلق^۱، ریشه میانگین مربعات خطا^۲ و ضریب تعیین^۳ بکار گرفته شده‌اند تا دقت و قابلیت اعتماد مدل‌های پیشنهادی سنجیده شود. این معیارها به مدیران سازمان‌ها کمک می‌کنند تا عملکرد مدل‌های پیش‌بینی را به طور دقیق ارزیابی کرده و از آن‌ها در تصمیم‌گیری‌های استراتژیک استفاده کنند. نتایج این مطالعه نشان می‌دهد که مدل‌های یادگیری ماشین نه تنها می‌توانند تغییرات هزینه‌های آتی را با دقت بالایی پیش‌بینی کنند بلکه قادر به تحلیل تأثیر شاخص‌های گوناگون بر هزینه‌ها نیز هستند. این چارچوب پیشنهادی می‌تواند به مدیران سازمان‌های خدماتی کمک کند تا تصمیم‌های استراتژیک بهتری در زمینه تخصیص منابع مالی اتخاذ کرده و بهره‌وری سازمانی را افزایش دهند.

افزون بر این، استفاده از چارچوب پیشنهادی در سایر بخش‌های سازمانی نظیر برنامه‌ریزی منابع انسانی، مدیریت زنجیره تأمین و بهبود فرایندهای عملیاتی نیز امکان‌پذیر است. این چارچوب می‌تواند به عنوان ابزاری برای شناسایی الگوهای هزینه‌های در بخش‌های گوناگون سازمان بکار رود و به بهینه‌سازی عملکرد کلی سازمان کمک کند. این پژوهش گامی در جهت توسعه رویکردهای داده‌محور برای مدیریت هزینه‌های سازمانی است که می‌تواند به افزایش کارایی و اثربخشی تصمیمات مدیریتی منجر شود.

هدف اصلی این پژوهش توسعه یک چارچوب داده‌محور و مبتنی بر یادگیری ماشین برای پیش‌بینی دقیق هزینه‌های سازمانی است. این چارچوب با استفاده از الگوریتم‌های یادگیری ماشین در پی بهبود دقت پیش‌بینی‌ها، شناسایی شاخص‌های تأثیرگذار بر هزینه‌ها و ارائه راهکارهایی برای بهینه‌سازی مدیریت منابع مالی سازمان است. افزون بر این، این پژوهش به بررسی میزان کارایی و دقت الگوریتم‌های گوناگون پیش‌بینی در شرایط متغیر سازمانی می‌پردازد و راهکارهایی برای بهبود عملکرد این الگوریتم‌ها ارائه می‌دهد. نتایج حاصل از این پژوهش می‌تواند به مدیران در اتخاذ تصمیمات استراتژیک و بهینه‌سازی تخصیص منابع در شرایط پیچیده و متغیر کمک کند.

مرور ادبیات

با توجه به اهمیت مدیریت ریسک و بهبود عملکرد شرکت‌های بیمه، پژوهش‌های متعددی به بررسی جنبه‌های گوناگون این صنعت پرداخته‌اند. بررسی این پژوهش‌ها نشان می‌دهد که استفاده از رویکردهای نوآورانه، از جمله چارچوب‌های مبتنی بر یادگیری ماشین، می‌تواند تأثیر بسزایی در بهینه‌سازی پیش‌بینی هزینه‌های سازمانی و افزایش توان رقابت‌پذیری این شرکت‌ها داشته باشد.

لی و همکاران^۴ به بررسی مدیریت ریسک عملیاتی در شرکت‌های بیمه چین پرداختند و سه جنبه اصلی را مورد مطالعه قرار دادند. آن‌ها ابتدا وضعیت فعلی ریسک عملیاتی در شرکت‌های بیمه چینی را توضیح داده و سپس علل وقوع این ریسک‌ها را بررسی کردند. در نهایت، روش‌هایی برای بهبود مدیریت ریسک در این شرکت‌ها ارائه کردند. یافته‌های این مطالعه نشان می‌دهد که درک دقیق علل و اتخاذ راهبردهای بهینه می‌تواند به کاهش خطرات عملیاتی کمک کرده و کارایی عملکرد شرکت‌های بیمه را افزایش دهد. در ادامه، پریموسکا و همکاران^۵ مفهوم شرکت بیمه‌گر جایگزین تحت مالکیت پزشکان را به‌عنوان جایگزینی برای بیمه‌های تجاری بررسی کرد. این شرکت که در سال ۱۹۷۵ برای مقابله با بحران مال‌پراکسیس تأسیس شد، حدود ۱۰۰ پزشک در تخصص‌های پرریسک را بیمه می‌کرد. تجربه ۱۴ ساله این شرکت نشان داد که مدل‌های مدیریتی نوآورانه می‌توانند نیازهای جامعه را با هزینه‌های کم‌تر برآورده کنند. این یافته‌ها می‌تواند به توسعه چارچوب‌های مبتنی بر یادگیری ماشین جهت پیش‌بینی و کاهش هزینه‌های عملیاتی کمک کند. هم‌چنین، جیانگ و وانگ^۶ و همکاران

¹. Mean Absolute Error | MAE

². Root Mean Square Error | RMSE

³. R-Squared | R²

⁴. Lee et al.

⁵. Prymostka

⁶. Jiang

استراتژی‌های سرمایه‌گذاری شرکت‌های بیمه در کانال فروش اینترنتی را تحلیل کردند و گزارش کردند که در سال ۲۰۱۴، درآمد حق بیمه از راه کانال اینترنتی به ۸۵/۸۹ میلیارد یوان رسید. آن‌ها تأکید کردند که کانال اینترنتی به‌عنوان یک ابزار کلیدی در افزایش درآمد و رقابت‌پذیری اهمیت دارد و ترکیب این استراتژی‌ها با روش‌های یادگیری ماشین می‌تواند به بهبود پیش‌بینی و مدیریت هزینه‌های سازمانی کمک کند.

ییتبارک و همکاران^۱ مسئولیت اجتماعی شرکت‌ها در شرکت‌های بیمه اموال چین را با استفاده از تحلیل مؤلفه اصلی بررسی کرد. وی با بهره‌گیری از داده‌های عمومی، مسئولیت‌های شرکت‌های بیمه در قبال سهامداران، مشتریان، کارکنان، دولت، جامعه و محیط زیست را تحلیل کرد و نتایج این مطالعه نشان داد که عملکرد مسئولیت اجتماعی این شرکت‌ها به‌طور کلی ضعیف است. این موضوع نیاز به سیستم‌های ارزیابی هدفمند را برجسته می‌کند. به‌طور مشابه، تا و همکاران^۲ به بررسی نظارت دولت بر سرمایه‌گذاری شرکت‌های بیمه در شرکت‌های بوری پرداخت و بر اهمیت تطابق زمانی بین پروژه‌های سرمایه‌گذاری و منابع مالی تأکید کرد. ونگ و همکاران^۳ نیز به توسعه شرکت‌های بیمه با سرمایه خارجی در چین پرداخت و نقش این شرکت‌ها را در بهبود بازار بیمه و ارتقاء صنعت بیمه تحلیل کرد. این مطالعات نشان می‌دهند که ترکیب یادگیری ماشین با سیستم‌های نظارتی و مسئولیت اجتماعی می‌تواند به بهبود پیش‌بینی‌های مالی و مدیریت ریسک کمک کند.

کوپرولچک و همکاران^۴ تأثیر اصلاحات تعرفه‌ای و شرایط بیمه اتومبیل تجاری در چین را با تحلیل داده‌های شرکت‌های بیمه اموال در استان هبی بررسی کردند. آن‌ها دریافتند که این اصلاحات منجر به رشد پایدار درآمد حق بیمه و تغییرات در ساختار کانال‌های فروش شده است. این یافته‌ها نشان می‌دهد که شرکت‌های کوچک و متوسط در این حوزه با چالش‌های بیش‌تری مواجه هستند. در ادامه، ژی و همکاران^۵ مدل‌های بیمه‌ای مورد استفاده در شرکت‌های حوزه چوب را تحلیل کردند و تأکید کردند که تجربه خسارات سنگین می‌تواند شرکت‌ها را به استفاده بیش‌تر از خدمات بیمه ترغیب کند. هم‌چنین، ژانگ و همکاران^۶ به ارزیابی توانگری مالی شرکت‌های بیمه عمر در چین تحت نظام مقرراتی نسل دوم توانگری پرداختند و بر اهمیت مدیریت ریسک و شفافیت مالی تأکید کردند. این مطالعات نشان می‌دهند که استفاده از یادگیری ماشین در تحلیل داده‌های بیمه‌ای می‌تواند به شناسایی الگوهای خطر و ارائه راهکارهای بهینه کمک کند.

کونگ و چن^۷ استراتژی‌های توسعه شرکت‌های بیمه اینترنتی حرفه‌ای را با مقایسه شاخص‌های کلیدی مالی شرکت‌های بیمه آنلاین و آفلاین بررسی کردند و نشان دادند که شرکت‌های بیمه اینترنتی می‌توانند در توانگری مالی و سودآوری بهبود یابند، هرچند در مدیریت هزینه‌ها چالش‌هایی دارند. در ادامه، کولر^۸ به بررسی مدل‌های تصادفی در بیمه عمر پرداخت و با استفاده از شبیه‌سازی مونت‌کارلو ریسک‌های ورشکستگی شرکت‌های بیمه را تحلیل کرد. وی نشان داد که این روش می‌تواند به ارزیابی بهتر ریسک و تصمیم‌گیری‌های مدیریتی کمک کند. هم‌چنین، سان و هوانگ^۹ به تحلیل سیستم‌های هشدار مالی زودهنگام برای شرکت‌های بیمه اموال چین پرداختند و نشان دادند که عدم توجه به پیش‌بینی مالی دقیق می‌تواند به تشدید زیان‌ها منجر شود. این پژوهش‌ها نشان می‌دهند که استفاده از یادگیری ماشین می‌تواند به توسعه مدل‌های پیش‌بینی و بهبود مدیریت مالی در صنعت بیمه کمک کند.

1. Yitbarek

2. Ta

3. Weng

4. Koprolec

5. Zhi et al.

6. Wong et al.

7. Kong & Chen

8. Koller

9. Sun & Huang

پرتکووا و واورووا^۱ نسبت هزینه‌های بیمه‌نامه‌های شرکت‌های بیمه منتخب در جمهوری چک را تحلیل کردند و نشان دادند که مدل‌های بانک‌شورانس می‌توانند به کاهش هزینه‌های عملیاتی کمک کنند. ادیگزال^۲ با استفاده از پرسش‌نامه‌ای از کاربران خدمات بیمه‌ای، نقش شش عامل کلیدی مانند سطح خدمات و سرعت فرآیندها را در ایجاد تصویر مثبت سازمانی بررسی کرد و نشان داد که این عوامل می‌توانند سهم بازار شرکت‌های بیمه را افزایش دهند. در ادامه، ژو و شو^۳ رقابت‌پذیری شرکت‌های بیمه عمر چین را با استفاده از تحلیل عاملی ارزیابی کردند و بر اهمیت مدیریت ریسک و نوآوری در محصولات تأکید کردند. این یافته‌ها بیانگر آن است که استفاده از چارچوب‌های مبتنی بر یادگیری ماشین می‌تواند به بهبود رقابت‌پذیری و افزایش سهم بازار شرکت‌های بیمه کمک کند.

راماکریشنان و همکاران^۴ مسئولیت اجتماعی شرکت‌های بیمه مالزی را در سه دسته اصلی خیریه، یکپارچگی و نوآوری تحلیل کرد و نشان داد که بیش‌تر شرکت‌ها در دسته یکپارچگی قرار دارند. وینتیل^۵ با استفاده از تحلیل پوششی داده‌ها، عملکرد شرکت‌های بیمه اموال را ارزیابی کردند و نشان دادند که روش خوشه‌بندی می‌تواند به بهبود تصمیم‌گیری مدیریتی کمک کند. جینگا^۶ به بررسی تأثیر استاندارد حسابداری IFRS17 بر مدل‌های ارزش‌گذاری شرکت‌های بیمه عمر پرداخت و نشان داد که این استاندارد می‌تواند به شفافیت مالی و بهبود عملکرد شرکت‌های بیمه کمک کند. این مطالعات نشان می‌دهند که بهره‌گیری از یادگیری ماشین در تحلیل داده‌های مالی می‌تواند به شناسایی نقاط ضعف و ارائه راهکارهای بهبود کمک کند.

نیسیم^۷ به تحلیل و ارزش‌گذاری شرکت‌های بیمه بر اساس استانداردهای بین‌المللی پرداخت و نشان داد که ناسازگاری در ارائه اطلاعات مالی می‌تواند بر قابلیت استفاده از این اطلاعات تأثیرگذار باشد. ژانگ و همکاران^۸ نیز سیستم تحلیل دوپونت را برای ارزیابی سودآوری شرکت‌های بیمه اموال به کار بردند و نشان دادند که این سیستم می‌تواند عوامل مؤثر بر بازده دارایی‌ها را شناسایی کند. یافته‌های آن‌ها بر اهمیت تحلیل داده‌های مالی و استفاده از الگوریتم‌های یادگیری ماشین برای بهبود سودآوری تأکید دارد.

جمع‌بندی مقالات مورد بررسی نشان می‌دهد که صنعت بیمه به دلیل ماهیت پیچیده و پرریسک خود، نیازمند رویکردهای نوآورانه برای بهبود مدیریت ریسک، افزایش توان رقابتی و بهینه‌سازی هزینه‌های سازمانی است. مطالعات متعددی به بررسی عوامل تأثیرگذار بر عملکرد شرکت‌های بیمه در حوزه‌هایی مانند مدیریت ریسک عملیاتی، استراتژی‌های سرمایه‌گذاری، مسئولیت اجتماعی و توانگری مالی پرداخته‌اند. یافته‌های این پژوهش‌ها نشان می‌دهد که بکارگیری چارچوب‌های مبتنی بر یادگیری ماشین می‌تواند به شناسایی الگوهای پنهان در داده‌های بیمه‌ای، پیش‌بینی دقیق‌تر هزینه‌ها و کاهش ریسک‌های عملیاتی کمک کند. هم‌چنین، این رویکردها می‌توانند باعث بهبود مدیریت مالی، توسعه محصولات نوآورانه و افزایش رضایت مشتریان شوند. در مجموع، بهره‌گیری از مدل‌های داده‌محور و الگوریتم‌های هوش مصنوعی در صنعت بیمه، زمینه‌ساز ارتقای کارایی و پایداری بلندمدت شرکت‌های بیمه خواهد بود.

روش پژوهش

این پژوهش با هدف طراحی و اجرای گام‌به‌گام یک مدل مبتنی بر یادگیری ماشین برای پیش‌بینی هزینه‌های سازمان بیمه سلامت استان قم انجام شده است.

1. Přečková & Vávrová

2. Adıgüzel

3. Zhu & Shu

4. Ramakrishnan et al.

5. Vintilă et al.

6. Jinga

7. Nissim

8. Zhang et al.

در گام نخست این پژوهش، داده‌های موردنیاز از سازمان بیمه سلامت استان قم گردآوری شد. این داده‌ها شامل اطلاعات مالی، عملیاتی و آماری مرتبط با هزینه‌های درمانی، تعداد بیمه‌شدگان، تعرفه‌های خدمات درمانی و مراجعات بیمه‌شدگان به مراکز درمانی بود. بازه زمانی پنج‌ساله (۱۳۹۷ تا نیمه اول ۱۴۰۲) برای گردآوری داده‌ها انتخاب شد تا روندهای بلندمدت و تغییرات پویای مرتبط با هزینه‌های درمانی مورد بررسی قرار گیرد. افزون بر داده‌های داخلی سازمان، شاخص‌های اقتصادی نظیر نرخ تورم و تغییرات کلان هزینه‌های درمان نیز به‌عنوان متغیرهای تأثیرگذار بر هزینه‌های بیمه سلامت گردآوری شد.

داده‌های گردآوری‌شده شامل داده‌هایی متنوع مانند تعداد بیمه‌شدگان فعال در هر دوره، نرخ تعرفه‌های خدمات درمانی، تعداد مراجعات به مراکز درمانی و هزینه‌های پرداختی به این مراکز بود. این اطلاعات از سامانه‌های مدیریتی داخلی سازمان و با هماهنگی بخش‌های مالی، اطلاعات آماری و بخش فناوری اطلاعات استخراج شد. داده‌های اقتصادی مانند نرخ تورم و هزینه‌های کلان درمان نیز بمنظور تحلیل اثرات غیرمستقیم بر هزینه‌های سازمان، از منابع معتبر گردآوری شد.

برای اطمینان از دقت و صحت داده‌ها، فرایند اعتبارسنجی در چندین مرحله انجام شد. ابتدا، داده‌ها از نظر تکمیل بودن بررسی شدند تا مقادیر گمشده شناسایی و با روش‌هایی نظیر میانگین‌گیری یا استفاده از مقادیر مشابه در دوره‌های زمانی مشابه جایگزین شوند. در مواردی که ناسازگاری میان داده‌ها مشاهده شد، داده‌ها با منابع دیگر مقایسه و در صورت نیاز اصلاح شدند. هم‌چنین، داده‌های غیرمعمول مانند تعداد مراجعات بیش‌ازحد یا مقادیر غیرواقعی بررسی و علت آن تحلیل شد.

پس از گردآوری داده‌ها، مرحله پیش‌پردازش به‌عنوان یکی از مراحل کلیدی در آماده‌سازی داده‌ها برای مدل‌های یادگیری ماشین آغاز شد. اگرچه داده‌های گردآوری شده فاقد مقادیر گمشده بودند، اما برای اطمینان از کیفیت تحلیل‌ها، بررسی‌های متعددی بر روی آن‌ها انجام گرفت. در ابتدا، داده‌ها از نظر ناهنجاری‌ها و مقادیر غیرمعمول مورد بررسی قرار گرفتند. مقادیری که از نظر آماری پرت به نظر می‌رسیدند، مانند افزایش ناگهانی در بار مراجعه یا تغییرات غیرعادی در جمعیت بیمه‌شدگان، شناسایی و اصلاح شدند. در این موارد، معمولاً مقادیر پرت با میانگین یا میانه داده‌های مشابه جایگزین شدند تا از تأثیر منفی این مقادیر بر مدل جلوگیری شود.

بمنظور مقابله با تفاوت مقیاس میان متغیرها، فرایند استانداردسازی بر روی داده‌ها اعمال شد. متغیرهایی نظیر جمعیت و هتلینگ به‌گونه‌ای تنظیم شدند که در یک بازه عددی مشابه قرار گیرند. برای این کار، از روش‌های مقیاس‌بندی به بازه صفر تا یک و نرمال‌سازی Z-Score استفاده شد. استانداردسازی داده‌ها باعث شد که الگوریتم‌های یادگیری ماشین بتوانند روابط میان متغیرها را بدون تأثیرگذاری مقیاس‌های متفاوت، به‌طور دقیق تحلیل کنند.

یکی از مراحل مهم دیگر در این فرایند، تقسیم داده‌ها به مجموعه‌های آموزشی و تست بود. ۸۰ درصد داده‌ها به‌صورت تصادفی به‌عنوان مجموعه آموزشی و ۲۰ درصد باقی‌مانده برای ارزیابی مدل به‌عنوان مجموعه تست اختصاص یافت. این تقسیم‌بندی تصادفی کمک کرد تا اطمینان حاصل شود که داده‌های هر دو مجموعه نماینده کل داده‌ها بوده و از ایجاد انحراف یا جانب‌داری در نتایج جلوگیری شود. مجموعه آموزشی برای یادگیری مدل‌ها استفاده شد، در حالی که مجموعه تست به ارزیابی توانایی مدل در پیش‌بینی داده‌های تازه اختصاص یافت.

در ادامه، فرایند مهندسی ویژگی‌ها برای بهبود دقت پیش‌بینی‌ها اجرا شد. این مرحله شامل ایجاد متغیرهای جدید از داده‌های موجود بود. برای مثال، شاخص "سرانه بار مراجعه" با تقسیم تعداد مراجعات بر جمعیت محاسبه شد تا بیانگر تعداد مراجعات به ازای هر فرد باشد. افزون بر این، روند تغییرات دوره‌ای در متغیرهایی مانند جمعیت و هتلینگ به‌عنوان متغیرهای جدید به مجموعه داده‌ها اضافه شدند تا مدل بتواند الگوهای زمانی را نیز شناسایی کند.

رگرسیون خطی یکی از ساده‌ترین و در عین حال کاربردی‌ترین الگوریتم‌های تحلیل داده و پیش‌بینی است که در بسیاری از حوزه‌های علمی و عملیاتی مورد استفاده قرار می‌گیرد. این الگوریتم با مدل‌سازی روابط خطی میان یک متغیر وابسته و یک یا چند متغیر مستقل، امکان تحلیل تأثیر متغیرهای گوناگون بر نتایج پیش‌بینی را فراهم می‌کند. در رگرسیون خطی ساده، تنها یک متغیر مستقل وجود دارد و رابطه آن با متغیر وابسته به‌صورت یک معادله خطی بیان می‌شود. این معادله (۱) است. در این معادله، (y) متغیر وابسته یا همان مقدار پیش‌بینی‌شده است، x به‌عنوان متغیر مستقل عمل می‌کند، β_0 مقدار عرض از

مبدأ یا نقطه‌ای که خط رگرسیون محور y را قطع می‌کند، و ϵ خطای مدل است که به تفاوت میان مقادیر واقعی و پیش‌بینی شده اشاره دارد.

$$y = \beta_0 + \beta_1 x + \epsilon \quad (۱)$$

در رگرسیون خطی چندگانه، که کاربرد بیش‌تری در تحلیل‌های پیچیده دارد، رابطه بین یک متغیر وابسته و چندین متغیر مستقل به صورت معادله (۲) بیان می‌شود. در معادله (۲)، متغیرهای (x_1, x_2, \dots, x_n) نشان‌دهنده متغیرهای مستقل هستند و ضرایب $\beta_1, \beta_2, \dots, \beta_n$ تأثیر هر متغیر مستقل بر متغیر وابسته را نشان می‌دهند. هدف اصلی رگرسیون خطی، یافتن ضرایبی است که بتواند خطای پیش‌بینی‌ها را به حداقل برساند. این خطا به عنوان مجموع مربعات خطا^۱ تعریف می‌شود و از راه معادله (۳) محاسبه می‌شود. در معادله (۳)، y_i مقادیر واقعی و \hat{y}_i مقادیر پیش‌بینی شده توسط مدل هستند. برای محاسبه ضرایب بهینه در رگرسیون خطی چندگانه، از روش‌های ماتریسی استفاده می‌شود که در این روش معادله به شکل معادله (۴) تعریف می‌شود. در معادله (۴)، X ماتریس داده‌های مستقل، y بردار مقادیر وابسته و β بردار ضرایب است.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (۲)$$

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (۳)$$

$$\beta = (X^T X)^{-1} X^T Y \quad (۴)$$

استفاده از رگرسیون خطی در تحلیل‌های داده‌محور به دلایل گوناگونی رایج است. این الگوریتم به دلیل سادگی در پیاده‌سازی و تفسیرپذیری بالا، یکی از بهترین گزینه‌ها برای مدل‌سازی روابط خطی میان متغیرها بشمار می‌رود. ضرایب محاسبه‌شده در رگرسیون خطی به صورت مستقیم تأثیر هر متغیر مستقل بر متغیر وابسته را مشخص می‌کنند. این ویژگی به مدیران کمک می‌کند تا با درک روابط میان عوامل گوناگون، تصمیمات آگاهانه‌تری اتخاذ کنند. از سوی دیگر، این الگوریتم در شرایطی که روابط میان متغیرها خطی یا تقریباً خطی باشد، عملکرد بسیار مطلوبی دارد و می‌تواند نتایج دقیقی ارائه دهد. همچنین، رگرسیون خطی به دلیل پیچیدگی محاسباتی کم، سرعت بالایی در تحلیل داده‌ها دارد و می‌تواند برای داده‌های حجیم و پیچیده نیز مورد استفاده قرار گیرد.

یکی دیگر از ویژگی‌های مهم رگرسیون خطی، قابلیت شناسایی و تحلیل روابط میان متغیرهاست. ضرایب بدست‌آمده از این مدل، اهمیت نسبی هر متغیر را نشان می‌دهند و به شناسایی عوامل کلیدی تأثیرگذار کمک می‌کنند. در پژوهش‌هایی نظیر پیش‌بینی هزینه‌های سازمانی، رگرسیون خطی می‌تواند روابط میان متغیرهایی مانند جمعیت، تعداد مراجعات و هزینه‌های پرداختی را به طور مؤثر مدل‌سازی کند. این ویژگی به مدیران سازمان‌ها اجازه می‌دهد تا با تحلیل دقیق‌تر این روابط، منابع مالی خود را به صورت بهینه‌تری تخصیص دهند و تصمیماتی بهتر اتخاذ کنند.

در این مطالعه، الگوریتم رگرسیون خطی برای پیش‌بینی هزینه‌های سازمان بیمه سلامت مورد استفاده قرار گرفت. با استفاده از این مدل، روابط میان متغیرهای گوناگون نظیر جمعیت، بار مراجعه و هزینه‌های پرداختی تحلیل و بررسی شدند. این الگوریتم به دلیل دقت بالا و قابلیت تفسیر آسان، یکی از ابزارهای کلیدی در تحلیل داده‌ها و بهبود تصمیم‌گیری‌های مدیریتی در این پژوهش بود. نتایج بدست‌آمده نشان داد که رگرسیون خطی می‌تواند به طور دقیق الگوهای هزینه‌ای را شناسایی کرده و پیش‌بینی‌های دقیقی برای آینده ارائه دهد که این امر می‌تواند نقشی مهم در بهبود مدیریت هزینه‌های سازمان ایفا کند.

در فرآیند ارزیابی مدل‌های یادگیری ماشین، استفاده از معیارهای گوناگون برای سنجش دقت و عملکرد مدل ضروری است. معیارهایی نظیر میانگین مربعات خطا، ریشه میانگین مربعات خطا، میانگین خطای مطلق، ضریب همبستگی و ضریب تعیین

^۱. Sum of squared error | SSE

از پرکاربردترین ابزارها در این زمینه بشمار می‌روند. هر یک از این معیارها ویژگی‌ها و کاربردهای خاص خود را دارند و معمولاً برای دستیابی به تحلیل جامع‌تر، بصورت ترکیبی مورد استفاده قرار می‌گیرند.

میانگین مربعات خطا یکی از مهم‌ترین معیارهای سنجش دقت مدل است که میانگین اختلافات مربعی بین مقادیر واقعی و پیش‌بینی شده را محاسبه می‌کند. این معیار به خطاهای بزرگ حساسیت بیشتری دارد زیرا اختلافات به توان دو می‌رسند. بنابراین، اگر مدل دچار خطاهای بزرگ شود، مقدار میانگین مربعات خطا به سرعت افزایش می‌یابد. این ویژگی باعث می‌شود که میانگین مربعات خطا ابزاری مناسب برای مقایسه مدل‌های گوناگون باشد. با این حال، به دلیل مقیاس مربعی آن، تفسیر مستقیم این معیار دشوار است.

برای رفع این مشکل، از ریشه میانگین مربعات خطا استفاده می‌شود که خطا را به مقیاس اصلی داده‌ها بازمی‌گرداند. این معیار، که از جذر میانگین مربعات خطا به دست می‌آید، به دلیل داشتن واحد مشابه با متغیر وابسته، تفسیر آسان‌تری دارد. ریشه میانگین مربعات خطا نیز مانند میانگین مربعات خطا به خطاهای بزرگ حساس است و زمانی مفید است که دقت در پیش‌بینی مقادیر بالاتر اهمیتی بیشتر داشته باشد.

میانگین خطای مطلق، معیار دیگری برای سنجش دقت مدل‌ها است که میانگین اختلافات مطلق بین مقادیر واقعی و پیش‌بینی شده را محاسبه می‌کند. برخلاف میانگین مربعات خطا و ریشه میانگین مربعات خطا، میانگین مربعات خطا به خطاهای بزرگ حساسیت کم‌تری دارد و تأکید بیشتری بر خطاهای کوچک دارد. این معیار زمانی مفید است که خطاهای کوچک مهم‌تر از خطاهای بزرگ باشند و هدف، بهبود دقت کلی مدل باشد.

ضریب همبستگی یکی دیگر از معیارهای ارزیابی است که میزان ارتباط خطی بین مقادیر واقعی و پیش‌بینی شده را نشان می‌دهد. این ضریب عددی بین -۱ و ۱ است؛ مقادیر نزدیک به ۱ نشان‌دهنده ارتباط مثبت قوی، مقادیر نزدیک به -۱ نشان‌دهنده ارتباط منفی قوی، و مقادیر نزدیک به صفر نشان‌دهنده نبود ارتباط خطی هستند. این معیار برای بررسی توانایی مدل در حفظ الگوهای خطی داده‌ها و پیش‌بینی دقیق‌تر به کار می‌رود.

ضریب تعیین که به عنوان نسبت واریانس توضیح داده شده توسط مدل به واریانس کل داده‌ها تعریف می‌شود، نشان‌دهنده قدرت مدل در توضیح تغییرات مقادیر واقعی است. مقدار ضریب تعیین بین ۰ و ۱ متغیر است؛ هرچه این مقدار به ۱ نزدیک‌تر باشد، مدل بهتر می‌تواند تغییرات داده‌ها را توضیح دهد. با این حال، یکی از محدودیت‌های ضریب تعیین این است که ممکن است با اضافه شدن متغیرهای جدید به مدل افزایش یابد، حتی اگر این متغیرها تأثیر قابل توجهی بر دقت پیش‌بینی نداشته باشند.

برای رفع این مشکل، از ضریب تعیین تعدیل‌یافته^۱ استفاده می‌شود که تأثیر تعداد متغیرهای مستقل و حجم داده‌ها را در محاسبات خود در نظر می‌گیرد. این معیار بویژه برای مقایسه مدل‌هایی با تعداد متغیرهای متفاوت مناسب است و می‌تواند عملکرد مدل را به صورت دقیق‌تر و واقع‌بینانه‌تر ارزیابی کند.

فرآیند پیاده‌سازی مدل‌های یادگیری ماشین در این پژوهش با هدف ارائه مدلی دقیق و کاربردی برای پیش‌بینی هزینه‌های سازمانی به صورت سیستماتیک و گام‌به‌گام انجام شده است. این مراحل شامل آماده‌سازی داده‌ها، انتخاب و تنظیم الگوریتم‌ها و ارزیابی مدل‌ها بوده است. برای انجام این تحلیل‌ها از نرم‌افزار کنایم استفاده شد. این نرم‌افزار به دلیل محیط گرافیکی پیشرفته، قابلیت‌های متنوع و امکان پیاده‌سازی فرآیندهای یادگیری ماشین به صورت بصری، بدون نیاز به کدنویسی پیچیده، انتخاب شده است.

در مرحله نخست، داده‌های مورد نیاز شامل متغیرهایی نظیر جمعیت، بار مراجعه، هزینه‌های درمان و سرانه بار مراجعه برای استفاده در مدل‌ها آماده شدند. این فرآیند شامل پاک‌سازی داده‌ها از خطاهای احتمالی، استانداردسازی متغیرها برای همگن‌سازی مقیاس‌ها و تقسیم داده‌ها به مجموعه‌های آموزشی و تست بود. برای افزایش دقت مدل‌ها، داده‌ها به صورت تصادفی به دو بخش تقسیم شدند، به طوری که ۸۰ درصد داده‌ها برای آموزش مدل‌ها و ۲۰ درصد باقی‌مانده برای ارزیابی

^۱. Adjusted R²

عملکرد آن‌ها استفاده شد. این رویکرد تصادفی کمک می‌کند تا مدل بتواند روابط میان متغیرها را بهتر یاد بگیرد و بر داده‌های جدید نیز تعمیم‌پذیری داشته باشد.

برای پیش‌بینی هزینه‌های سازمانی، از سه الگوریتم اصلی شامل رگرسیون خطی، درخت تصمیم و تقویت گرادیان استفاده شد. انتخاب این الگوریتم‌ها به دلیل توانایی آن‌ها در تحلیل داده‌های ساختاریافته و ارائه خروجی‌های قابل تفسیر بوده است. محیط گرافیکی نرم‌افزار کنایم امکان تنظیم و پیاده‌سازی این مدل‌ها را به صورت بصری فراهم کرد.

در مرحله بعد، تنظیم و بهینه‌سازی پارامترهای الگوریتم‌ها انجام شد. برای رگرسیون خطی، ضرایب متغیرها با استفاده از روش حداقل مربعات خطا محاسبه شد تا خطای پیش‌بینی به حداقل برسد. در درخت تصمیم، پارامترهایی مانند عمق درخت و حداقل تعداد نمونه‌ها در هر گره تنظیم شد تا از بروز بیش‌برازش جلوگیری شود. هم‌چنین، در مدل تقویت گرادیان، پارامترهایی مانند تعداد مراحل، نرخ یادگیری و عمق درخت‌ها تنظیم شد. این فرآیند با استفاده از روش جستجوی شبکه‌ای و اعتبارسنجی متقابل انجام گرفت، که نرم‌افزار کنایم ابزارهای لازم برای اجرای آن را فراهم کرد.

پس از تنظیم پارامترها، هر مدل با استفاده از مجموعه داده‌های آموزشی آموزش داده شد. در این مرحله، مدل‌ها روابط میان متغیرهای ورودی و خروجی را شناسایی کرده و بهبود یافتند. نرم‌افزار کنایم با ارائه نمایش گرافیکی از مراحل آموزش و ابزارهای تعاملی، فرآیند آموزش مدل‌ها را ساده‌تر کرد. پس از اتمام آموزش، مدل‌های آموزش‌دیده با استفاده از مجموعه داده‌های تست ارزیابی شدند. برای سنجش دقت پیش‌بینی‌ها، از معیارهایی نظیر میانگین مربعات خطا، ریشه میانگین مربعات خطا و میانگین خطای مطلق استفاده شد.

در نهایت، مقایسه عملکرد مدل‌ها نشان داد که مدل تقویت گرادیان در مقایسه با سایر مدل‌ها دقت بالاتری دارد. این مدل به دلیل توانایی در تحلیل روابط غیرخطی و پیچیده، به عنوان مدل نهایی انتخاب شد. تحلیل نتایج نشان داد که این مدل می‌تواند با دقت بیش‌تری هزینه‌های آینده را پیش‌بینی کند و اطلاعات ارزشمندی برای تصمیم‌گیری‌های مالی ارائه دهد.

یافته‌های پژوهش

در بخش نتایج، عملکرد مدل‌های رگرسیون خطی و گرادیان بوستینگ که در این پژوهش برای پیش‌بینی هزینه‌های سازمانی بکار گرفته شده‌اند، با استفاده از معیارهای گوناگون مورد ارزیابی قرار گرفت. معیارهای مورد استفاده شامل ضریب تعیین، میانگین خطای مطلق و ریشه میانگین مربعات خطا بوده‌اند. این معیارها بمنظور مقایسه دقت پیش‌بینی مدل‌ها و انتخاب مدل برتر تحلیل شده‌اند.

نتایج نشان می‌دهد که مدل رگرسیون خطی، با مقدار ضریب تعیین برابر با ۰/۹۳۷۰، توانسته است بخش قابل توجهی از واریانس داده‌های واقعی را توضیح دهد. این مقدار نشان‌دهنده آن است که این مدل در شناسایی و تبیین الگوهای خطی موجود در داده‌ها عملکرد خوبی داشته است. در مقابل، مدل گرادیان بوستینگ نیز با ضریب تعیین برابر با ۰/۹۲۷۲ عملکرد مناسبی نشان داده، اما در این معیار نسبت به رگرسیون خطی کمی ضعیف‌تر عمل کرده است.

نتایج سایر معیارهای ارزیابی، مانند میانگین خطای مطلق و ریشه میانگین مربعات خطا، نشان می‌دهد که مدل رگرسیون خطی دقت بالاتری در پیش‌بینی داده‌ها داشته و خطاهای کم‌تری ایجاد کرده است. مقدار میانگین خطای مطلق برای مدل رگرسیون خطی برابر با ۳۸،۷۰۰،۵ محاسبه شده، در حالی که برای مدل گرادیان بوستینگ این مقدار ۴۱،۰۶۰،۱ بوده است.

افزون بر این، مقدار ریشه میانگین مربعات خطا که به دلیل حساسیت بالا به خطاهای بزرگ، معیاری مهم برای ارزیابی مدل‌های پیش‌بینی است، برای مدل رگرسیون خطی برابر با ۵۳،۸۲۰،۷ و برای مدل گرادیان بوستینگ برابر با ۵۶،۱۰۶،۳ محاسبه شده است. این نتایج حاکی از آن است که مدل رگرسیون خطی در مدیریت خطاهای بزرگ و ارائه پیش‌بینی‌های دقیق‌تر نسبت به گرادیان بوستینگ موفق‌تر عمل کرده است.

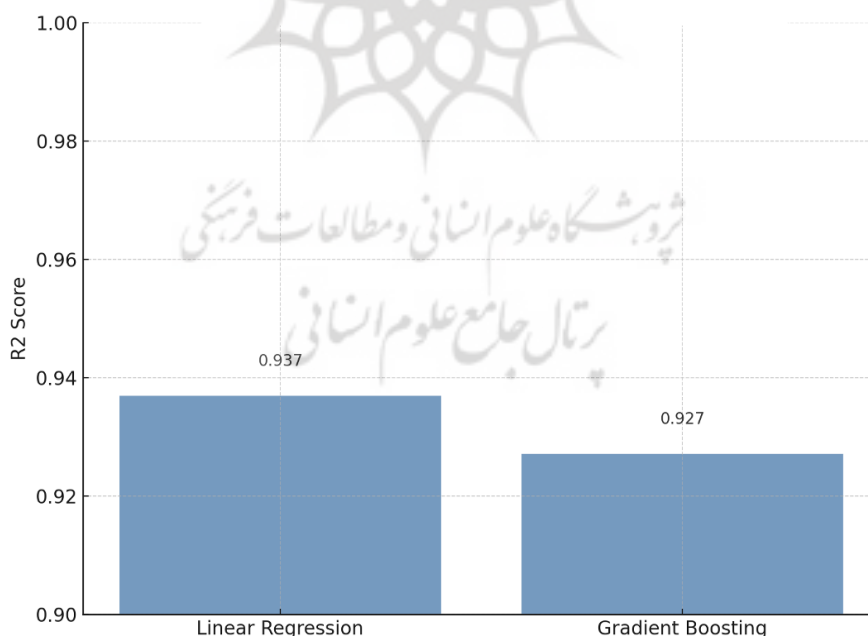
به‌طور کلی، تحلیل نتایج نشان می‌دهد که هر دو مدل در پیش‌بینی داده‌های واقعی عملکرد مناسبی داشته‌اند. مدل رگرسیون خطی به دلیل ساختار ساده و تفسیرپذیری بالای خود توانسته است روابط خطی میان متغیرها را به خوبی مدل‌سازی

کند، اما مدل گرادیان بوستینگ به دلیل توانایی در شناسایی و تحلیل روابط پیچیده و غیرخطی میان داده‌ها، دقتی بیش‌تر در پیش‌بینی داده‌ها داشته و توانسته است خطاهای پیش‌بینی را کاهش دهد. جدول ۱ خلاصه‌ای از عملکرد دو الگوریتم را نشان می‌دهد.

جدول ۱. نتایج ارزیابی مدل‌ها

مدل	ضریب تعیین	میانگین خطای مطلق	ریشه میانگین مربعات خطا
گرادیان بوستینگ	۰/۹۲۷۲	۴۱،۰۶۰،۱۰	۵۶،۱۰۶،۳۰
رگرسیون خطی	۰/۹۳۷	۳۸،۷۰۰،۵۰	۵۳،۸۲۰،۷۰

شکل ۱ ضریب تعیین دو مدل رگرسیون خطی و گرادیان بوستینگ را مقایسه می‌کند. این ضریب نشان‌دهنده توانایی مدل در توضیح واریانس موجود در داده‌های واقعی است. همان‌طور که در نمودار مشاهده می‌شود، مدل رگرسیون خطی با ضریب تعیین ۰/۹۳۷ عملکرد کمی بهتر نسبت به مدل گرادیان بوستینگ با ضریب تعیین ۰/۹۲۷۲ داشته است. این اعداد به این معنا هستند که مدل رگرسیون خطی توانسته ۹۳/۷ درصد از تغییرات داده‌های واقعی را توضیح دهد، در حالی که مدل گرادیان بوستینگ موفق به پوشش ۹۲/۷۲ درصد از این تغییرات شده است. تفاوت میان عملکرد این دو مدل در ضریب تعیین جزئی است، اما نشان می‌دهد که رگرسیون خطی به دلیل ساختار ساده و خطی خود توانسته الگوهای موجود در داده‌ها را با دقت بالاتری مدل‌سازی کند. با این حال، باید توجه داشت که هر دو مدل در این معیار، عملکرد قابل‌قبولی داشته‌اند و توانسته‌اند بخش عمده‌ای از واریانس داده‌های واقعی را توضیح دهند. این نتایج حاکی از آن است که هر دو مدل برای پیش‌بینی داده‌ها مناسب هستند، اما انتخاب مدل برتر باید با توجه به سایر معیارهای ارزیابی نیز انجام شود.

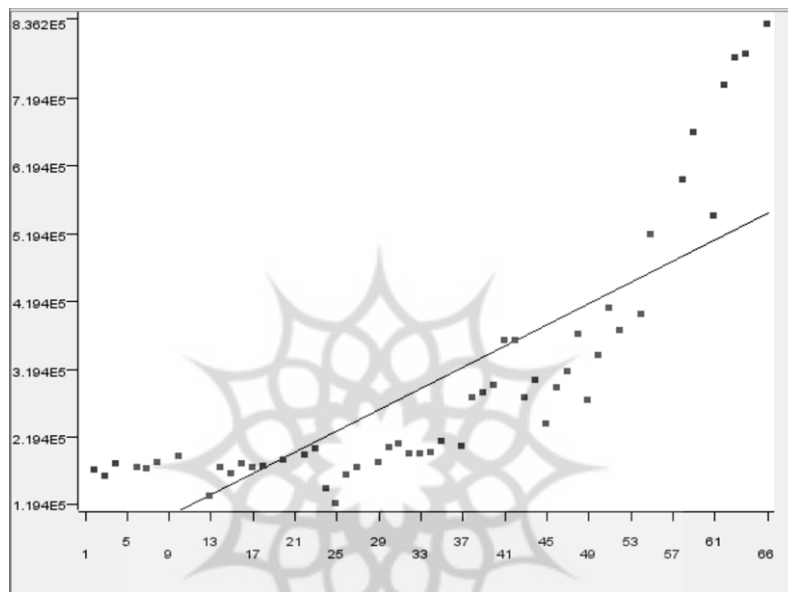


شکل ۱. ارزیابی عملکرد مدل در معیار ضریب تعیین

شکل ۲ بمنظور بصری‌سازی عملکرد مدل رگرسیون خطی و ارزیابی دقت پیش‌بینی‌های آن نسبت به داده‌های واقعی ارائه شده است. در این نمودار، مقادیر واقعی در مقابل مقادیر پیش‌بینی شده توسط مدل رسم شده و خط رگرسیون نیز برای نمایش

روند کلی داده‌ها ترسیم شده است. همان‌طور که در نمودار مشاهده می‌شود، خط رگرسیون به خوبی توانسته است الگوی خطی موجود در داده‌ها را شناسایی کرده و روند کلی تغییرات را توضیح دهد.

این شکل نشان می‌دهد که مدل رگرسیون خطی در بسیاری از موارد پیش‌بینی‌های دقیقی ارائه داده و مقادیر پیش‌بینی‌شده نزدیک به مقادیر واقعی بوده‌اند. با این حال، در برخی نقاط، انحراف‌هایی مشاهده می‌شود که نشان‌دهنده وجود خطاهای جزئی در پیش‌بینی‌ها است. این نمودار می‌تواند به عنوان مرجعی برای تحلیل رفتار مدل در پیش‌بینی داده‌های جدید و بررسی قابلیت تعمیم آن بکار رود. عملکرد مثبت مدل در این شکل نشان می‌دهد که رگرسیون خطی توانسته است روابط خطی میان متغیرهای ورودی و خروجی را با دقت مناسبی مدل‌سازی کند و مبنایی برای تصمیم‌گیری‌های مدیریتی فراهم آورد.



شکل ۲. ارزیابی عملکرد مدل رگرسیون خطی

نتایج تحلیل مدل رگرسیون خطی در این پژوهش نشان می‌دهد که چندین متغیر تأثیری معنادار بر هزینه‌های سازمانی دارند. بر اساس اطلاعات ارائه‌شده در جدول ۲ که تحلیل ضرایب را نشان می‌دهد، متغیر دوره با ضریب $4281,4806$ به عنوان تأثیرگذارترین متغیر شناخته شده است. مقدار PPP-value این متغیر برابر با $0/019$ بوده که نشان‌دهنده معناداری آماری آن است. ضریب مثبت این متغیر بیانگر این است که با گذشت زمان و افزایش دوره‌ها، هزینه‌های سازمانی نیز به صورت میانگین افزایش می‌یابد. این یافته نشان‌دهنده رشد هزینه‌های سازمانی در طول زمان است، که می‌تواند ناشی از عواملی نظیر تورم، افزایش تعرفه‌های خدمات و رشد جمعیت بیمه‌شدگان باشد.

از دیگر متغیرهای معنادار می‌توان به بار مراجعه و جمعیت اشاره کرد. متغیر بار مراجعه با ضریب $1/439$ و PPP-value برابر با $0/302$ نشان می‌دهد که هر واحد افزایش در تعداد مراجعات به مراکز درمانی منجر به افزایش هزینه‌های سازمانی می‌شود. این نتیجه بیانگر اهمیت کنترل تعداد مراجعات بیمه‌شدگان برای مدیریت بهینه هزینه‌ها است. به طور مشابه، جمعیت بیمه‌شدگان نیز با ضریب $1/1462$ و مقدار PPP-value برابر با $0/237$ تأثیر معناداری بر هزینه‌ها دارد. این یافته نشان می‌دهد که با افزایش تعداد افراد تحت پوشش بیمه، هزینه‌های سازمانی به طور متوسط افزایش می‌یابد. این رابطه منطقی به نظر می‌رسد، زیرا افزایش جمعیت معمولاً با افزایش تقاضا برای خدمات درمانی همراه است.

یکی دیگر از متغیرهای معنادار، تعداد ویزیت‌های پزشک عمومی است. این متغیر با ضریب $5/5828$ و مقدار PPP-value برابر با $0/07$ تأثیر قابل توجهی بر هزینه‌های سازمانی دارد. این یافته بیانگر این است که افزایش تعداد ویزیت‌های پزشک

عمومی می‌تواند هزینه‌های درمانی را به‌طور چشمگیری افزایش دهد. از این رو، بهینه‌سازی تعداد مراجعات به پزشکان عمومی می‌تواند یکی از راهکارهای مؤثر در کاهش هزینه‌های سازمانی باشد.

در مقابل، برخی از متغیرها نظیر سرانه بار مراجعه، ویزیت پزشک متخصص و هتلینگ از نظر آماری معنادار نبوده‌اند زیرا مقادیر PPP-value آن‌ها بزرگ‌تر از ۰/۰۵ است. برای مثال، متغیر هتلینگ با ضریب $-4/28$ و $E-5$ و PPP-value برابر با ۰/۷۶۳۵ نشان می‌دهد که این متغیر تأثیر قابل‌توجهی بر هزینه‌های سازمانی ندارد و تغییرات آن به‌صورت معنادار بر هزینه‌ها اثر نمی‌گذارد. این یافته‌ها نشان می‌دهد که تمرکز بر متغیرهای معنادار می‌تواند منجر به تصمیم‌گیری‌های مدیریتی مؤثرتری شود.

تحلیل کلی مدل رگرسیون خطی نشان می‌دهد که این مدل با ضریب تعیین برابر با ۰/۹۳۰۳ و ضریب تعیین تعدیل‌شده برابر با ۰/۹۱۹۲ توانسته است بیش از ۹۳ درصد از تغییرات هزینه‌های سازمانی را توضیح دهد. این اعداد بیانگر دقت بالای مدل در شناسایی الگوهای موجود در داده‌ها و پیش‌بینی هزینه‌ها است. به‌این‌ترتیب، می‌توان نتیجه گرفت که مدل رگرسیون خطی ابزار مناسبی برای تحلیل و پیش‌بینی هزینه‌های سازمانی است و می‌تواند اطلاعات ارزشمندی برای بهبود تصمیم‌گیری‌های مدیریتی ارائه دهد.

جدول ۲. ضرایب متغیرهای معنادار

متغیر	ضرایب	خطای استاندارد	t-value	P-value	معناداری
دوره	۴,۲۸۱/۴۸	۱,۲۹۶/۳۴	۳/۳۰۲۷	۰/۰۰۱۹	معنادار
بار مراجعه	۱/۴۳۹	۰/۶۴۲۵	۲/۲۳۹۶	۰/۰۳۰۲	معنادار
جمعیت	۱/۱۴۶۲	۰/۴۸۹۱	۲/۳۴۳۵	۰/۰۲۳۷	معنادار
سرانه بار مراجعه	-۴۲,۸۸۴/۳۵	۶۳۲,۲۹۵/۷۰	-۰/۱۱۸۴	۰/۹۰۶۳	نامعنادار
ویزیت پزشک عمومی	۵/۵۸۲۸	۱/۹۷۳۳	۲/۸۲۹۲	۰/۰۰۷	معنادار
ویزیت پزشک متخصص	-۲/۵۹۰۴	۱/۷۰۰۶	-۱/۵۲۳۲	۰/۱۳۴۹	نامعنادار
هتلینگ	-۴/۲E-05	۰/۰۰۰۱	-۰/۳۰۲۸	۰/۷۶۳۵	نامعنادار

پس از اجرای مدل رگرسیون خطی و انجام تحلیل‌های لازم برای پیش‌بینی هزینه‌های سازمان در بازه‌های زمانی گوناگون، نتایج بدست‌آمده حاکی از آن است که این مدل توانسته پیش‌بینی‌های دقیق و قابل‌اعتمادی ارائه دهد. پیش‌بینی‌های انجام‌شده نشان‌دهنده روند تغییرات هزینه‌های سازمان در دوره‌های گوناگون است که می‌تواند به تصمیم‌گیرندگان کمک کند تا منابع مالی را به‌صورت بهینه تخصیص دهند و برای دوره‌های آینده برنامه‌ریزی بهتری داشته باشند. جدول ۳ نتایج پیش‌بینی هزینه‌های سازمان را در دوره‌های پنجاه و پنجم تا شصتم نمایش می‌دهد. این داده‌ها نشان می‌دهد که مدل رگرسیون خطی توانسته است الگوهای موجود در داده‌های گذشته را با دقت بالایی شناسایی کرده و هزینه‌های آینده را بر اساس این الگوها تخمین بزند.

جدول ۳. مقادیر پیش‌بینی‌شده هزینه‌ها

دوره	هزینه پیش‌بینی‌شده
۵۵	۹۱۸,۶۷۹,۳۶۶
۵۶	۸۱۰,۲۲۱,۴۲۱
۵۷	۸۶۸,۸۳۹,۲۵۵
۵۸	۷۶۱,۴۹۷,۱۸۶
۵۹	۶۹۸,۲۸۰,۶۵۸
۶۰	۸۵۹,۶۸۱,۷۲۶

بر اساس مقادیر ارائه‌شده در جدول پیش‌بینی هزینه‌ها، مدل رگرسیون خطی توانسته است الگوهای خطی موجود در داده‌های گذشته را شناسایی کرده و پیش‌بینی‌های دقیقی ارائه دهد. همان‌طور که در جدول ۳ مشاهده می‌شود، بالاترین هزینه پیش‌بینی‌شده مربوط به دوره پنجاه و پنجم با مقدار ۹۱۸،۶۷۹،۳۶۶ ریال است، در حالی که کم‌ترین هزینه پیش‌بینی‌شده برای دوره پنجاه و نهم با مقدار ۶۹۸،۲۸۰،۶۵۸ ریال ثبت شده است. این تغییرات نشان می‌دهد که متغیرهایی نظیر دوره زمانی، جمعیت و بار مراجعه تأثیر قابل‌توجهی بر هزینه‌های سازمانی دارند.

تحلیل این داده‌ها نشان می‌دهد که مدل رگرسیون خطی ابزاری مؤثر برای پیش‌بینی هزینه‌های آتی سازمان است و می‌تواند به مدیران کمک کند تا منابع مالی را به‌صورتی بهینه‌تر تخصیص دهند. افزون بر این، بررسی این نتایج امکان شناسایی الگوهای بلندمدت و روندهای کلی هزینه‌ها را فراهم می‌کند که می‌تواند در تصمیم‌گیری‌های استراتژیک سازمانی نقشی مهم ایفا کند.

پس از گردآوری مقادیر واقعی هزینه‌های سازمان در دوره‌های موردنظر، امکان مقایسه آن‌ها با مقادیر پیش‌بینی‌شده فراهم شد. این مقایسه که در جدول ۴ ارائه شده است، به تحلیل میزان دقت مدل کمک می‌کند و نشان می‌دهد که مدل رگرسیون خطی در بیش‌تر دوره‌ها پیش‌بینی‌های نزدیکی به مقادیر واقعی ارائه داده است. برای مثال، در دوره‌های پنجاه و ششم و پنجاه و هفتم، اختلاف بین مقادیر پیش‌بینی‌شده و واقعی بسیار کم بوده و مدل توانسته است تغییرات هزینه‌ها را با دقت بالایی شبیه‌سازی کند.

با این حال، در برخی از دوره‌ها مانند دوره‌های پنجاه و هشتم و پنجاه و نهم، اختلاف بیش‌تری بین مقادیر پیش‌بینی‌شده و واقعی مشاهده می‌شود. این اختلافات ممکن است ناشی از تغییرات غیرمنتظره در متغیرهای تأثیرگذار باشد که مدل قادر به پیش‌بینی آن‌ها نبوده است. عواملی نظیر تغییرات ناگهانی در سیاست‌های تعرفه‌ای یا نوسانات اقتصادی ممکن است بر این پیش‌بینی‌ها تأثیر گذاشته باشند.

به‌طور کلی، مدل رگرسیون خطی در شرایطی که عوامل تأثیرگذار پایدارتر بوده‌اند، مانند دوره‌های پنجاه و ششم و پنجاه و هفتم، عملکرد بهتری داشته و پیش‌بینی‌های دقیق‌تری ارائه کرده است. این یافته‌ها بیانگر توانایی مدل در شناسایی الگوهای ثابت و تکرار شونده در داده‌ها است. با توجه به میانگین اختلاف نسبتاً کم بین مقادیر پیش‌بینی‌شده و واقعی، می‌توان نتیجه گرفت که مدل رگرسیون خطی ابزاری کاربردی و قابل‌اعتماد برای مدیریت و برنامه‌ریزی هزینه‌های سازمانی است. این نتایج می‌تواند مبنای تصمیم‌گیری‌های مالی و تخصیص بهینه منابع در سازمان باشد.

جدول ۴. اختلاف مقادیر پیش‌بینی شده و واقعی

دوره	هزینه واقعی	هزینه پیش‌بینی شده	اختلاف
۵۵	۸۳۴،۱۱۵،۰۰۰	۹۱۸،۶۷۹،۳۶۶	۸۴،۵۶۴،۳۶۶
۵۶	۸۶۰،۶۹۰،۰۰۰	۸۱۰،۲۲۱،۴۲۱	۵۰،۴۶۸،۵۷۹
۵۷	۹۰۶،۸۷۷،۰۰۰	۸۶۸،۸۳۹،۲۵۵	۳۸،۰۳۷،۷۴۵
۵۸	۹۱۴،۳۵۵،۰۰۰	۷۶۱،۴۹۷،۱۸۶	۱۵۲،۸۵۷،۸۱۴
۵۹	۹۷۳،۵۲۳،۰۰۰	۶۹۸،۲۸۰،۶۵۸	۲۷۵،۲۴۲،۳۴۲
۶۰	۹۶۲،۴۵۹،۰۰۰	۸۵۹،۶۸۱،۷۲۶	۱۰۲،۷۷۷،۲۷۴

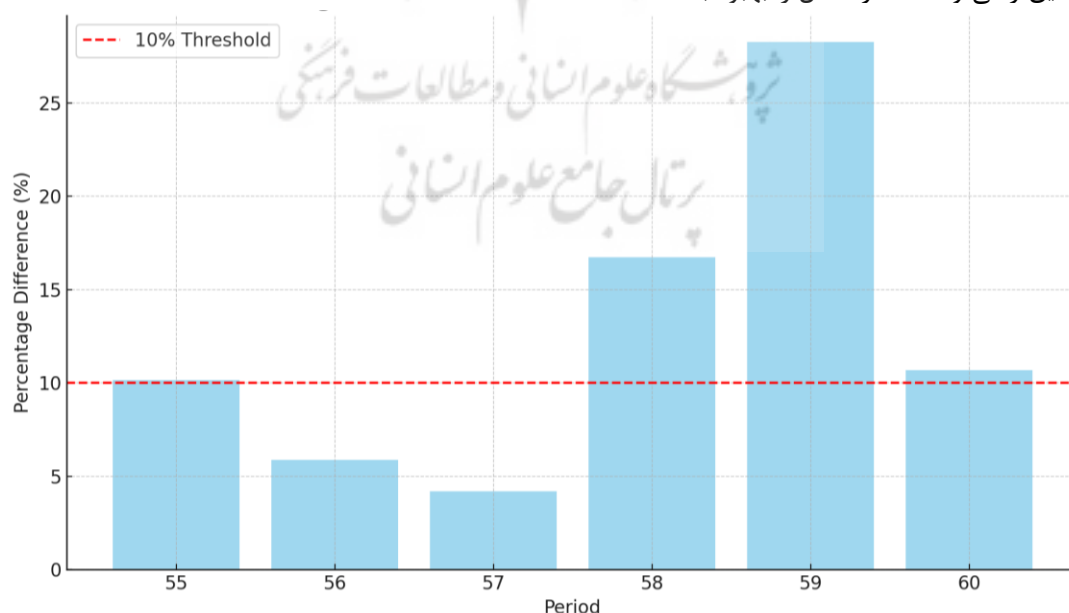
بر اساس جدول مقایسه هزینه‌های پیش‌بینی‌شده و واقعی، اختلاف مقادیر بین پیش‌بینی‌ها و مقادیر واقعی از ۳۸،۰۳۷،۷۴۵ ریال تا ۲۷۵،۲۴۲،۳۴۲ ریال متغیر بوده است. همان‌طور که در جدول مشاهده می‌شود، دوره پنجاه و نهم با اختلاف ۲۷۵،۲۴۲،۳۴۲ ریال بیش‌ترین تفاوت را نشان می‌دهد، در حالی که کم‌ترین اختلاف مربوط به دوره پنجاه و هفتم با ۳۸،۰۳۷،۷۴۵ ریال است. این تفاوت‌ها نشان‌دهنده تأثیر عوامل گوناگون در دقت پیش‌بینی‌های مدل رگرسیون خطی است.

عملکرد مدل رگرسیون خطی در پیش‌بینی هزینه‌های سازمانی قابل قبول بوده و در بسیاری از موارد اختلافات پیش‌بینی‌ها با مقادیر واقعی در حد معقول باقی مانده است. این موضوع نشان می‌دهد که مدل توانسته است روندهای اصلی و تغییرات کلی در هزینه‌ها را شناسایی کند. با این حال، در برخی از دوره‌ها مانند پنجاه‌هفتم و پنجاه‌ونهم، اختلافات بزرگ‌تری مشاهده شده که به احتمال زیاد، ناشی از تغییرات غیرمنتظره در عوامل تأثیرگذار یا شرایط خاص سازمان بوده است.

این تحلیل نشان می‌دهد که مدل رگرسیون خطی ابزار مفیدی برای پیش‌بینی هزینه‌های آتی سازمان است، اما برای بهبود دقت آن، می‌توان از داده‌های بیشتر و مدل‌های پیچیده‌تر بهره گرفت. به‌ویژه در دوره‌هایی که اختلافاتی بیشتر مشاهده می‌شود، افزودن متغیرهای تازه یا استفاده از مدل‌های پیشرفته‌تر می‌تواند به بهبود عملکرد مدل کمک کند.

در شکل ۳، درصد اختلاف پیش‌بینی‌ها نسبت به مقادیر واقعی برای دوره‌های گوناگون نمایش داده شده است. این نمودار به‌صورت بصری تفاوت‌های درصدی پیش‌بینی‌ها را نسبت به مقادیر واقعی نشان می‌دهد. خط قرمز در این نمودار به‌عنوان آستانه ۱۰ درصد در نظر گرفته شده است که به تشخیص دوره‌هایی که پیش‌بینی‌ها دارای دقتی مناسب هستند، کمک می‌کند. همان‌طور که در شکل مشاهده می‌شود، در دوره پنجاه‌هفتم درصد اختلاف کم‌تر از ۱۰ درصد است که نشان‌دهنده دقت بالای مدل در این دوره است. در مقابل، در دوره‌های پنجاه‌ونهم و پنجاه‌وهم درصد اختلاف بالاتر از آستانه تعیین شده است، که ممکن است ناشی از عواملی باشد که در مدل لحاظ نشده‌اند یا تغییرات غیرقابل پیش‌بینی در داده‌ها اتفاق افتاده است.

به‌طور کلی، مدل در حدود یک‌سوم موارد (۳۳/۳۳ درصد) پیش‌بینی‌های دقیقی ارائه داده که درصد اختلاف آن‌ها کم‌تر از ۱۰ درصد بوده است. این میزان نشان می‌دهد که مدل توانسته است روندهای اصلی داده‌ها را شناسایی کند و پیش‌بینی‌های دقیقی ارائه دهد. با این حال، در دوره‌هایی که اختلافات بیش‌تری وجود دارد، نیاز به بهینه‌سازی مدل احساس می‌شود. به‌منظور بهبود عملکرد مدل، پیشنهاد می‌شود که از مدل‌های پیشرفته‌تر یادگیری ماشین مانند تقویت گرادیان یا شبکه‌های عصبی استفاده شود، یا متغیرهایی تازه به مدل اضافه شوند تا بتوان تغییرات بیش‌تر و پیچیده‌تری را در داده‌ها توضیح داد. این یافته‌ها تأکید می‌کنند که مدل در شرایطی که داده‌ها پایدارتر هستند و تغییرات کم‌تری دارند، دقت بالاتری ارائه می‌دهد. در مقابل، در شرایط ناپایدار یا تغییرات ناگهانی، مدل نیاز به تطبیق بیش‌تر دارد و استفاده از رویکردهای بهینه‌سازی و تحلیل عمیق‌تر می‌تواند عملکرد مدل را بهبود بخشد.



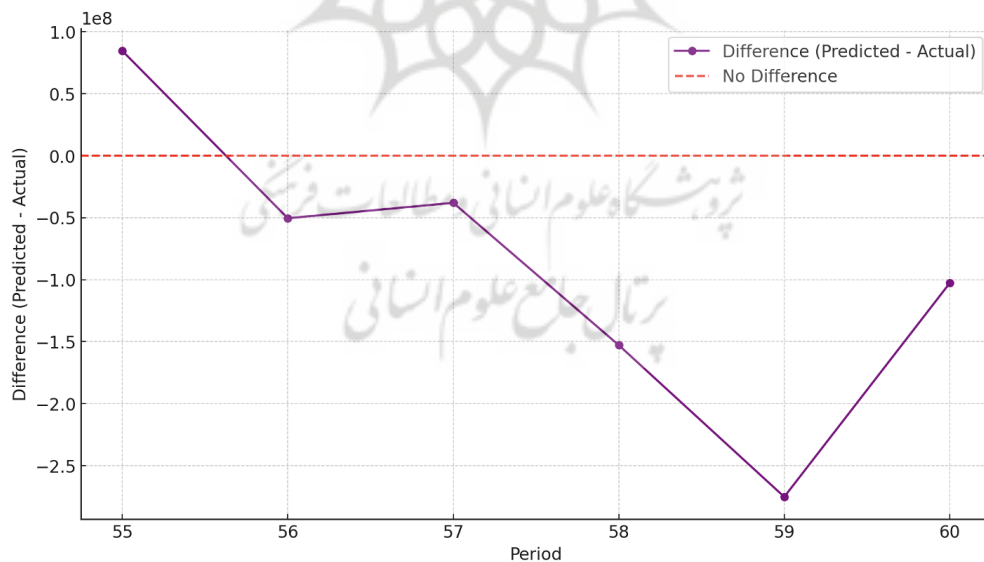
شکل ۳. اختلاف درصدی پیش‌بینی‌ها نسبت به مقادیر واقعی

شکل ۴ روند تغییرات اختلاف بین هزینه‌های پیش‌بینی‌شده و واقعی را برای دوره‌های گوناگون به تصویر می‌کشد. نقاطی که در این نمودار بالاتر از خط قرمز، که نشان‌دهنده اختلاف صفر است، قرار دارند، بیانگر پیش‌بینی‌هایی هستند که مقادیر پیش‌بینی‌شده آن‌ها بیش از هزینه‌های واقعی بوده است. در مقابل، نقاطی که پایین‌تر از این خط قرار گرفته‌اند، به پیش‌بینی‌هایی اشاره دارند که هزینه‌های کم‌تری نسبت به مقادیر واقعی ارائه داده‌اند.

در دوره پنجاه‌وپنجم، مدل هزینه‌ای معادل ۸۴،۵۶۴،۳۶۶ ریال بیش از مقدار واقعی پیش‌بینی کرده است. این اختلاف مثبت نشان‌دهنده حساسیت مدل به برخی از عوامل خاص در این دوره است که ممکن است به‌درستی شناسایی نشده باشند. در ادامه، طی دوره‌های پنجاه‌وششم تا شصتم، مدل هزینه‌های کم‌تری نسبت به مقادیر واقعی پیش‌بینی کرده است. با این حال، این اختلافات در اکثر دوره‌ها در محدوده منطقی قرار داشته و تنها در دوره پنجاه‌ونهم به بیش‌ترین مقدار خود، یعنی ۲۷۵،۲۴۲،۳۴۲ ریال، رسیده است.

بررسی روند کلی اختلاف‌ها نشان می‌دهد که مدل رگرسیون خطی توانسته تغییرات هزینه‌ها را در بیش‌تر دوره‌ها به‌خوبی شناسایی کند و پیش‌بینی‌های نزدیکی به مقادیر واقعی ارائه دهد. این موضوع نشان‌دهنده کارایی و دقت مناسب مدل در شناسایی روندهای هزینه‌های سازمانی است. با این وجود، مشاهده اختلاف‌های بزرگ‌تر در برخی دوره‌ها، مانند دوره پنجاه‌ونهم، حاکی از تأثیر احتمالی تغییرات پیش‌بینی‌نشده در عوامل محیطی یا تغییرات ناگهانی در الگوهای هزینه‌ها است که در داده‌های ورودی مدل به‌طور کامل لحاظ نشده‌اند.

این تحلیل تأکید می‌کند که مدل رگرسیون خطی در پیش‌بینی روندهای کلی هزینه‌ها عملکرد مناسبی داشته و می‌تواند به‌عنوان ابزاری مفید در تصمیم‌گیری‌های مالی و برنامه‌ریزی منابع سازمانی مورد استفاده قرار گیرد. برای بهبود دقت مدل در دوره‌های خاص، پیشنهاد می‌شود که با افزودن داده‌های دقیق‌تر و شناسایی عوامل تأثیرگذار جدید، ورودی‌های مدل تقویت شوند. این اقدامات می‌تواند به کاهش اختلاف‌ها و افزایش دقت پیش‌بینی‌ها در شرایط پیچیده و غیرخطی کمک کند.



شکل ۴. روند اختلاف‌ها بین هزینه‌های پیش‌بینی‌شده و واقعی

بحث و نتیجه‌گیری

پیش‌بینی هزینه‌های سازمانی یکی از ضروری‌ترین فعالیت‌های مدیریتی است که نقشی مهم در تخصیص بهینه منابع و اتخاذ تصمیمات مالی مؤثر ایفا می‌کند. در دنیای امروز، سازمان‌ها در محیط‌هایی پرچالش و غیرقابل پیش‌بینی فعالیت می‌کنند که

در آن تغییرات اقتصادی و نوسانات هزینه‌ها به سرعت رخ می‌دهد. در چنین شرایطی، استفاده از ابزارهای سنتی پیش‌بینی، نظیر تحلیل روندها و رگرسیون خطی، به‌تنهایی قادر به پاسخ‌گویی به نیازهای پیچیده سازمان‌ها نیست. سازمان‌های خدمات سلامت، به‌ویژه در کشورهای در حال توسعه مانند ایران، با افزایش مداوم هزینه‌های درمان، تغییرات تعرفه‌ها و افزایش بار مراجعه بیمه‌شدگان مواجه‌اند و نیازمند پیش‌بینی‌هایی دقیق و قابل اعتماد هستند. در این میان، الگوریتم‌های پیشرفته یادگیری ماشین، با توانایی تحلیل روابط پیچیده میان متغیرها، فرصتی بی‌نظیر برای بهبود دقت پیش‌بینی‌ها و تصمیم‌گیری‌های مدیریتی فراهم می‌کنند.

مدیریت هزینه‌های عملیاتی در سازمان‌های خدمات عمومی، به‌ویژه در بخش سلامت، از اهمیت بالایی برخوردار است؛ چراکه این هزینه‌ها بخش قابل توجهی از منابع مالی سازمان را به خود اختصاص می‌دهند. در صورتی که پیش‌بینی‌های دقیق و کارآمدی در زمینه هزینه‌ها صورت نگیرد، سازمان‌ها با چالش‌هایی همچون تخصیص نامناسب منابع و کاهش سطح کیفیت خدمات مواجه می‌شوند. از این رو، پیش‌بینی دقیق هزینه‌ها نه تنها به کنترل هزینه‌ها کمک می‌کند، بلکه به بهبود برنامه‌ریزی و سیاست‌گذاری در سطح سازمانی منجر می‌شود. درک عوامل تأثیرگذار بر هزینه‌های سازمانی به مدیران این امکان را می‌دهد که با اتخاذ سیاست‌های هدفمند، ریسک‌های مالی را کاهش دهند و بهره‌وری را افزایش دهند. با توجه به محدودیت‌های روش‌های سنتی پیش‌بینی، به‌کارگیری مدل‌های پیشرفته یادگیری ماشین، که قابلیت شناسایی روابط پنهان و تحلیل داده‌های پیچیده را دارند، ضرورتی اجتناب‌ناپذیر به نظر می‌رسد. این مدل‌ها امکان شناسایی روندهای پنهان و پیش‌بینی دقیق‌تر هزینه‌ها را فراهم کرده و به بهبود تصمیم‌گیری در شرایط پویای سازمانی کمک می‌کنند.

در راستای پاسخ‌گویی به این ضرورت، مطالعه حاضر با هدف توسعه چارچوبی مبتنی بر یادگیری ماشین برای پیش‌بینی هزینه‌های سازمانی طراحی و اجرا شده است. داده‌های این پژوهش شامل اطلاعات مالی و عملیاتی سازمان بیمه سلامت در استان قم در بازه زمانی پنج‌ساله گردآوری شده است. این داده‌ها پس از پاک‌سازی و استانداردسازی، برای آموزش مدل‌های یادگیری ماشین آماده شدند. دو مدل رگرسیون خطی و گرادیان بوستینگ برای تحلیل داده‌ها و پیش‌بینی هزینه‌ها به کار گرفته شدند. عملکرد این مدل‌ها با استفاده از معیارهایی نظیر ضریب تعیین، میانگین خطای مطلق و ریشه میانگین مربعات خطا ارزیابی شده است. نتایج نشان می‌دهد که مدل گرادیان بوستینگ در مدیریت خطاهای بزرگ و ارائه پیش‌بینی‌های دقیق‌تر نسبت به مدل رگرسیون خطی عملکرد بهتری داشته است. این مطالعه تلاش می‌کند تا با ارائه پیش‌بینی‌های دقیق، ابزاری کاربردی برای مدیران سازمان‌های خدمات عمومی فراهم کند تا بتوانند منابع مالی را بهینه‌تر تخصیص دهند و هزینه‌های آینده را به‌طور دقیق‌تر مدیریت کنند.

نتایج بدست‌آمده از روش پژوهش ذکر شده نشان می‌دهد که مدل رگرسیون خطی با تحلیل داده‌های تاریخی، توانسته است روابط خطی میان متغیرهای کلیدی را به‌خوبی شناسایی کرده و پیش‌بینی‌هایی نزدیک به مقادیر واقعی ارائه دهد. ضریب تعیین مدل برابر با $0/9370$ بیانگر دقت بالای مدل در توضیح واریانس داده‌های واقعی است. بررسی مقادیر پیش‌بینی‌شده و واقعی نشان می‌دهد که در بیش‌تر دوره‌ها، اختلاف‌ها در محدوده قابل‌قبولی قرار داشته‌اند و مدل توانسته روند کلی تغییرات هزینه‌ها را به‌خوبی شناسایی کند. حدود $33/33$ درصد از پیش‌بینی‌های مدل، دارای اختلاف کم‌تر از 10 درصد با مقادیر واقعی بوده‌اند، که این موضوع حاکی از توانایی مدل در ارائه پیش‌بینی‌های دقیق در شرایط گوناگون است. این یافته‌ها نشان می‌دهد که مدل رگرسیون خطی ابزاری مناسب برای پیش‌بینی هزینه‌های سازمانی و تحلیل روندهای موجود در داده‌ها است.

بررسی جزئی‌تر اختلاف‌ها بین مقادیر پیش‌بینی‌شده و واقعی نشان می‌دهد که مدل در برخی دوره‌ها دقت بیشتری داشته و اختلاف‌ها نزدیک به صفر بوده است، مانند دوره ۵۷، در حالی که در دوره‌هایی مانند ۵۹، اختلاف بیش‌تری مشاهده شده است. این امر نشان می‌دهد که مدل رگرسیون خطی در شرایطی که تغییرات عوامل بیرونی ناگهانی نبوده و داده‌ها پایدارتر بوده‌اند، عملکرد بهتری داشته است. هم‌چنین، تحلیل ضرایب رگرسیون نشان داد که متغیرهایی مانند دوره، جمعیت و بار مراجعه تأثیر معناداری بر هزینه‌ها دارند و مدل توانسته این تأثیرات را به‌درستی شناسایی کند. به‌طور کلی، مدل رگرسیون خطی در این مطالعه ابزاری کارآمد و تفسیرپذیر برای پیش‌بینی هزینه‌های سازمانی بوده و یافته‌های آن می‌تواند به مدیران در

تخصیص بهینه منابع و مدیریت بهتر هزینه‌ها کمک کند. با این حال، برای بهبود دقت در دوره‌هایی که تغییرات بیش‌تری در عوامل تأثیرگذار مشاهده می‌شود، پیشنهاد می‌شود از مدل‌های پیشرفته‌تر یا داده‌های تکمیلی استفاده شود.

مقایسه نتایج این مطالعه با پژوهش‌های پیشین نشان می‌دهد که مدل‌های داده‌محور نقش مهمی در بهبود تصمیم‌گیری مالی ایفا می‌کنند، اما دقت به دست آمده در این مطالعه به مراتب بالاتر از بسیاری از نمونه‌های مشابه گزارش شده است. برای مثال، ژانگ و همکاران با استفاده از سیستم تحلیل دوپونت در صنعت بیمه اموال، تنها به شناسایی عوامل مؤثر بر بازده دارایی‌ها بسنده کرده‌اند، در حالی که در این پژوهش، مدل رگرسیون خطی توانسته است نه تنها تأثیر متغیرهای کلیدی مانند جمعیت و بار مراجعه را با دقت معناداری شناسایی کند بلکه با دستیابی به ضریب تعیین 0.9370 ، و ثبت اختلاف کم‌تر از ۱۰ درصد در بیش از یک سوم پیش‌بینی‌ها، عملکردی دقیق و تفسیرپذیر ارائه دهد. هم‌چنین، اگرچه کولر در تحلیل ریسک ورشکستگی از شبیه‌سازی‌های پیچیده بهره گرفته است، اما دقت نهایی مدل‌های مورد استفاده در آن پژوهش به‌طور کمی گزارش نشده و قابلیت مقایسه مستقیم با مدل حاضر را ندارد.

از سوی دیگر، یافته‌های این مطالعه از منظر دقت و پایداری پیش‌بینی‌ها نسبت به پژوهش‌هایی مانند پرتکوا و واورووا یا سان و هوانگنیز برتری قابل توجهی دارد. در حالی که پرتکوا و همکارش تنها به کاهش نسبی هزینه‌های عملیاتی اشاره کرده‌اند، مدل رگرسیون این پژوهش با ثبت عملکرد یکنواخت در دوره‌های گوناگون، حتی در شرایط نوسانی، توانسته است روند هزینه‌ها را با دقت بالا پیش‌بینی کند. هم‌چنین، در پژوهش سان و هوانگ بر اهمیت هشدارهای مالی زود هنگام تأکید شده، اما الگویی برای پیش‌بینی دقیق ارائه نشده است. در مقابل، مدل به‌کاررفته در این پژوهش ضمن حفظ سادگی و قابلیت پیاده‌سازی، توانسته است خروجی‌هایی قابل اتکا و دقیق تولید کند که این امر آن را به ابزاری اثربخش برای مدیران در تخصیص منابع مالی و مدیریت هزینه‌ها تبدیل می‌سازد.

با وجود نتایج دقیق و ارزشمند به دست آمده از این پژوهش، همچنان محدودیت‌هایی در چارچوب مطالعه وجود دارد که نیاز به بررسی و بهبود در پژوهش‌های آینده دارد. یکی از محدودیت‌ها مربوط به استفاده از داده‌های تاریخی یک سازمان بیمه سلامت در یک منطقه جغرافیایی خاص است که ممکن است بر دامنه تعمیم‌پذیری نتایج تأثیر بگذارد. هرچند که داده‌های گردآوری شده از دقت بالایی برخوردار بوده و طی فرآیندهای پیش‌پردازشی استاندارد شده‌اند، اما افزودن داده‌های متنوع‌تر و مرتبط با سایر مناطق می‌تواند غنای بیش‌تری به تحلیل‌ها ببخشد. افزون بر این، در این مطالعه از مدل رگرسیون خطی به‌عنوان یک ابزار قدرتمند و تفسیرپذیر استفاده شده است که برای شناسایی الگوهای خطی در داده‌ها بسیار مؤثر بوده است. با این حال، استفاده از مدل‌های تکمیلی برای شناسایی روابط غیرخطی می‌تواند در بهبود دقت پیش‌بینی‌ها و کاهش اختلافات در برخی دوره‌های خاص مؤثر باشد. در مجموع، این محدودیت‌ها تأثیری بر اعتبار نتایج اصلی مطالعه ندارند، اما توجه به آن‌ها در پژوهش‌های آینده می‌تواند به تقویت چارچوب‌های تحلیلی و بهبود پیش‌بینی‌ها کمک کند.

کاربرد نتایج این پژوهش برای سازمان‌های بیمه، به‌ویژه در زمینه پیش‌بینی و مدیریت هزینه‌های سازمانی، بسیار گسترده است. سازمان‌های بیمه می‌توانند از مدل رگرسیون خطی ارائه شده در این مطالعه به‌عنوان ابزاری کارآمد برای تحلیل داده‌های تاریخی و پیش‌بینی هزینه‌های آتی استفاده کنند. این مدل با شناسایی الگوهای پنهان و روابط میان متغیرهای کلیدی مانند جمعیت، بار مراجعه، و هزینه‌های درمان، به مدیران سازمان‌ها کمک می‌کند تا روندهای هزینه‌ای را به‌طور دقیق‌تر پیش‌بینی کرده و برای تخصیص منابع مالی برنامه‌ریزی بهتری داشته باشند. با استفاده از این مدل، سازمان‌های بیمه می‌توانند تصمیمات استراتژیکی مانند تعیین تعرفه‌ها، مدیریت مراجعات به مراکز درمانی، و بهینه‌سازی خدمات بیمه‌ای را به‌صورت داده‌محور اتخاذ کنند. این امر می‌تواند منجر به کاهش هزینه‌های غیرضروری، افزایش بهره‌وری سازمان، و در نهایت بهبود رضایت بیمه‌شدگان شود.

افزون بر این، نتایج این پژوهش می‌تواند به سازمان‌ها در شناسایی عوامل تأثیرگذار بر هزینه‌ها و اتخاذ راهکارهای مناسب برای مدیریت آن‌ها کمک کند. برای مثال، سازمان‌های بیمه می‌توانند با تحلیل داده‌های پیش‌بینی شده، روندهای صعودی در هزینه‌ها را شناسایی کرده و پیش از وقوع مشکلات مالی، اقدامات پیشگیرانه‌ای مانند بهینه‌سازی سیاست‌های خدمات درمانی،

کنترل مراجعات غیرضروری، و مدیریت بهتر منابع انسانی و مالی انجام دهند. هم‌چنین، این مدل می‌تواند به سازمان‌ها در شناسایی بخش‌هایی از خدمات که نیاز به توجه یا اصلاح دارند، کمک کند. به‌عنوان نمونه، اگر تحلیل‌ها نشان دهد که هزینه‌های مرتبط با ویزیت پزشکان عمومی روند افزایشی دارد، سازمان می‌تواند برنامه‌هایی برای مدیریت این هزینه‌ها و هدایت بیمه‌شدگان به استفاده از خدمات جایگزین ارائه دهد. در مجموع، استفاده از مدل‌های پیش‌بینی مبتنی بر داده، به سازمان‌ها این امکان را می‌دهد که با رویکردی پیش‌نگرانه و علمی، عملکرد مالی خود را بهینه‌سازی کرده و ریسک‌های احتمالی را کاهش دهند.

برای مطالعات آتی، پیشنهاد می‌شود که پژوهش‌های پیش‌تری بر استفاده از الگوریتم‌های پیشرفته‌تر مانند شبکه‌های عصبی و جنگل تصادفی برای بهبود دقت پیش‌بینی هزینه‌های سازمانی انجام شود. هم‌چنین، افزودن متغیرهای بیشتر از جمله شاخص‌های اقتصادی، تغییرات سیاست‌های دولتی و شاخص‌های کیفی مرتبط با خدمات بیمه‌ای می‌تواند دقت مدل‌ها را افزایش دهد. از سوی دیگر، پیشنهاد می‌شود که اثرات استفاده از مدل‌های پیش‌بینی بر عملکرد بلندمدت سازمان‌ها، نظیر افزایش بهره‌وری و کاهش هزینه‌ها، در پژوهش‌های آینده بررسی شود. توسعه سیستم‌های پیش‌بینی خودکار که بتوانند داده‌های جدید را به‌صورت پویا تحلیل کرده و پیش‌بینی‌های به‌روزشده ارائه دهند نیز می‌تواند به بهبود مدیریت منابع در سازمان‌های خدمات عمومی کمک کند.

منابع

- Abdulla, A., Naim, A., Muniasamy, A., Mohammed, A. B., Bilfaqih, S. M., & Sabahath, A. (2025). Optimizing Business Insights Data Visualization Applications in Sales Forecasting, Marketing Analytics, and Financial Reporting. In *Data Visualization Tools for Business Applications*, 103-124, IGI Global.
- ADIGUZEL, Z., & YAZICI, B. (2019). ANALYSIS OF THE EFFECTS OF CORPORATE IMAGE ON PURCHASING BEHAVIORS IN INSURANCE COMPANIES. *International Journal of Eurasia Social Sciences/Uluslararası Avrasya Sosyal Bilimler Dergisi*, 10(37), 934-954.
- Arpita, H. D., Al Ryan, A., Hossain, M. F., Rahman, M. S., Sajjad, M., & Prova, N. N. I. (2025). Exploring Bengali speech for gender classification: machine learning and deep learning approaches. *Bulletin of Electrical Engineering and Informatics*, 14(1), 328-337.
- Cohen, S. (2025). Cohen, S. (2025). The evolution of machine learning: Past, present, and future. In *Artificial Intelligence in Pathology*, 3-14, Elsevier.
- Jiang, W., & Wong, J. K. (2016). Key activity areas of corporate social responsibility (CSR) in the construction industry: a study of China. *Journal of cleaner production*, 113, 850-860.
- Kong, Y. H., & Chen, S. Y. (2016). A Study on Development Strategies of Professional Internet Insurance Company: Comparative Analysis on Key financial indicators of Online & Offline Insurance Companies. In *Proceedings of the 7th China International Conference on Insurance and Risk Management*, 295-307.
- Koprolcec, M., Greger, K., Pirc, A., & Moro, M. (2009). Insurance of assets as a risk in wood sector/Osiguranje imovine kao rizik u poslovanju drvnog sektora. *Drvena Industrija*, 60(4), 235-244.
- Krylov, S. (2025). Theory and Practice of Target Financial Forecasting at Company Level. In *Encyclopedia of Information Science and Technology, Sixth Edition* (pp. 1-34). IGI Global.
- Koller, M. (2012). *Stochastic models in life insurance*. Springer Science & Business Media.
- Lee, G. F. (1991). Alternative liability insurance: A physician-owned captive insurance company. *American journal of obstetrics and gynecology*, 164(6), 1680-1685.
- Lucasius, C., Ali, M., Patel, T., Kundur, D., Szatmari, P., Strauss, J., & Battaglia, M. (2025). A procedural overview of why, when and how to use machine learning for psychiatry. *Nature Mental Health*, 1-11.
- Mukherjee, R., Ghosh, A., Chakraborty, C., De, J. N., & Mishra, D. P. (2025). Rice leaf disease identification and classification using machine learning techniques: A comprehensive review. *Engineering Applications of Artificial Intelligence*, 139, 109639.
- Nissim, D. (2010). Analysis and valuation of insurance companies. CE| ASA (Center for Excellence in Accounting and Security Analysis) Industry Study, (2).

- Prymostka, O. (2018). Life insurance companies marketing strategy in the digital world. *Insurance Markets and Companies*, 9(1), 66-73.
- Ramakrishnan, S., Alsahliy, D. K., Hishan, S. S., Keong, L. B., & Vaicondam, Y. (2017). Corporate responsibility of the listed Malaysian insurance companies. *Advanced Science Letters*, 23(9), 9279-9281.
- Sun, L., & Huang, Y. (2016). Measuring the instability of China's financial system: Indices construction and an early warning system. *Economics*, 10(1), 20160019.
- Ta, N. (2024). Analysis of the Development Status of Foreign Insurance Companies in China. *International Journal of Global Economics and Management*, 2(2), 194-199.
- Tang, J. W., Yuan, Q., Zhang, L., Marshall, B. J., Tay, A. C. Y., & Wang, L. (2025). Application of machine learning-assisted surface-enhanced Raman spectroscopy in medical laboratories: principles, opportunities, and challenges. *TrAC Trends in Analytical Chemistry*, 118135.
- Yitbarek, B. (2020). State Regulation and Supervision of Insurance Industry in Ethiopia: The Need for a Separate Regulatory and Supervisory Body. Master of Laws (LL. M) thesis, Addis Ababa University School of Graduate Studies Faculty of law, 11.
- Přechková, L., & Vávrová, E. (2019). The Expense Ratio Development of Insurance Portfolios in Selected Insurance Companies in the Czech Republic in 2010-2017. In *European Financial Systems 2019: Proceedings of 16th Annual International Scientific Conference on European Financial Systems* (pp. 452-461).
- Vintilă, A., Trucmel, I. M., & Roman, M. D. (2022). Measuring and Analyzing the Efficiency of Firms in the Insurance Industry Using DEA Techniques. *Journal of Social and Economic Statistics*, 11(1-2), 59-83.
- Weng, D. Y. (2000). China's life insurance industry: opportunities and challenges for foreign companies (Doctoral dissertation, Massachusetts Institute of Technology).
- Wong, J. (2002). A comparison of solvency requirements and early warning systems for life insurance companies in China with representative world practices. *North American Actuarial Journal*, 6(1), 91-112.
- Zhi, Y., & Hu, J. L. (2011). A cross-strait comparative study of efficiency of life insurance companies: An application of the input slack adjustment approach. *African Journal of Business Management*, 5(14), 5746.
- Zhu, J., & Shu, G. An Empirical Analysis of Small and Medium-sized China's Insurance Industry's Core Competitiveness by Three-stage DEA Method.
- Zhang, F., Han, L., & Zhang, J. (2016). The Analysis of Property Insurance Company's Profitability and its Impact Factor. In *Proceedings of 7th China International Conference on Insurance and Risk Management (CICIRM)*. Xi'an, China (pp. 235-248).

Development of a Machine Learning-Based Framework for Optimizing Organizational Cost Forecasting

Ahrar Hosseini^{1*}

Article Info	Abstract
Article type: Research Article	<p>In today's world, accurate forecasting of organizational costs is one of the key challenges managers face in optimizing resource allocation and enhancing productivity. Rapid economic changes and increasing financial complexities have highlighted the necessity of utilizing data-driven forecasting methods more than ever. This study presents a framework based on machine learning algorithms aimed at improving the accuracy of organizational cost predictions and facilitating more precise financial planning within organizations. The research involves collecting real financial and operational data from a health insurance organization over a five-year period. After pre-processing, this data was used to train predictive models. The primary model employed in this study was linear regression, and its performance was evaluated using metrics such as the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). The findings indicate that the linear regression model accurately predicted changes in organizational costs and identified key factors influencing these costs. The proposed framework can assist organizations in improving budget planning, managing financial resources, and reducing forecasting errors. Moreover, the findings provide a scientific basis for developing predictive models in other organizational areas, thereby facilitating strategic decision-making for managers.</p>
Keywords	Organizational Cost Forecasting, Machine Learning, Data-Driven Model, Resource Optimization, Data Analysis, Linear Regression, Financial Management

Publisher: Islamic Azad University Qods Branch

Corresponding Author: Ahrar Hosseini

Email: almastwo@yahoo.com

1. National Center for Health Insurance Research, Tehran, Iran (Corresponding Author)