Iranian Journal of Insurance Research
(IJIR)

Homepage: https://ijir.irc.ac.ir/?lang=en

Iranian Insurance Research Center (IIRC)

ORIGINAL RESEARCH PAPER

Predicting damage using sequential deep regression techniques in deep learning methods

Mona Parastesh[1,*], Ziaeddin Beheshtifard[2], Abbas Raad[3]

[1] *PhD Student, Department of Computer Engineering, Faculty of Technical and Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran.*

[2] *Instructor, Department of Computer Engineering, Faculty of Technical and Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran .*

[3] *Assistant Professor, Department of Management, Faculty of Management and Accounting, Shahid Beheshti University, Tehran, Iran.*

ARTICLE INFO

Article History:
Received: 31 October 2024
Revised: 19 January 2025
Accepted: 3 February 2025
Early online access: 5 February 2025
Published: 1 October 2025

Keywords:
Deep learning
Fire damage
Insurance
Machine learning

*Corresponding Author:
Email: *St_m_parastesh@azad.ac.ir*
Phone: +9821 29465229
ORCID: 0009-0008-2851-7702

DOI: 10.22056/ijir.2025.04.02

ABSTRACT

BACKGROUND AND OBJECTIVES: The role of the insurance industry is changing nowadays. The reason is that companies are using new analytical methods to predict losses and risks, and these methods help them assess potential risks. In this era, traditional business models and old methods have always been under threat from technology. New insurance companies are using the power of innovative technologies to eliminate the traditional leaders of the insurance market. The protection provided to the insured against risks and the proposed solutions provided to deal with risks are obtained through services designed to identify potential risks, and these services can be used to warn of danger (in high-risk cases). As a result, these services will be the most important distinction of these companies and the key to their success in the future. Powerful artificial intelligence and analysis of large volumes of big data give insurers the power to move towards predicting losses and incidents. The more information insurance companies have about their policyholders, the better they can use these valuable data to predict policyholders' behavior and create a historical profile for each individual, thereby reducing the volume of claims and associated risks. Insurance companies enjoying leverage innovative technologies have a significant opportunity for growth. However, those that continue to rely on basic questions such as age, gender, and occupation to determine premiums are unlikely to survive in the digital era and amidst the rise of insurtech. Insurers that fail to adopt predictive analytics and continue to use outdated traditional systems may experience longer delays in processing and paying claims compared to innovative companies. This gap will allow tech-driven insurers to attract more customers and cover a wider range of policyholders in the long term. Insurance data often contains nonlinear and complex relationships that simple models—such as linear regression or decision trees—cannot fully capture or model effectively. These companies are also faced with vast volumes of data. Traditional methods such as general linear models often fail to identify complex patterns in insurance data. Therefore, we seek to improve existing methods by applying modern techniques such as deep learning, since deep neural networks can more accurately identify complex patterns in insurance data, process large datasets efficiently, and uncover hidden insights. Deep learning, with the ability to identify nonlinear relationships and complex patterns, can overcome these limitations. In this paper, a method to improve the performance of deep learning using sequential deep regression techniques is presented. The proposed approach is a combination of deep learning and sequential models. Long Short-Term Memory (LSTM) neural networks are used to model time series data.

METHODS: In this study, data spanning the past seven years from Alborz Insurance Company—specifically related to the issuance and loss records of fire insurance policies—has been systematically utilized to analyze and forecast potential losses in this domain. The methodology places a strong emphasis on comprehensive data pre-processing, including cleaning, normalization, and transformation to ensure the reliability and quality of the input data. In the feature engineering stage, various techniques were applied to extract the most informative and relevant attributes from the raw dataset. Out of a total of 40 initially selected features, the top 20 features were identified through statistical analysis and machine learning-based selection methods. These refined features were then used to train the deep learning models. The proposed method is a hybrid approach that combines deep learning with sequential modeling techniques. Specifically, Long Short-Term Memory (LSTM) neural networks were employed due to their ability to capture time-dependent patterns in sequential data, making them particularly suitable for modeling the temporal dynamics inherent in insurance data over multiple years.

FINDINGS: The study involved the evaluation and comparison of multiple machine learning algorithms, including traditional models and advanced deep learning techniques. The results demonstrated that the proposed sequential deep regression approach significantly outperforms conventional models such as general linear models and decision trees. Notably, the LSTM-based model provided higher prediction accuracy and demonstrated superior performance in identifying complex, nonlinear patterns within the data. Key findings highlight the critical role of temporal features in enhancing prediction reliability and show that incorporating time series analysis is essential for improving the accuracy of damage occurrence forecasts in fire insurance.

CONCLUSION: The results of this research underscore the effectiveness of combining deep learning techniques with sequential models for predicting fire insurance losses. Using the confidential and comprehensive issuance and claims dataset from Alborz Insurance Company over seven years, the proposed hybrid model was capable of delivering better performance in comparison to previous methods. The approach not only improved the precision of predictions but also offered a more robust and scalable solution for risk assessment. Overall, the use of LSTM-based deep learning models represents a significant advancement in the insurance industry's ability to make data-driven decisions regarding premium setting, policy issuance, and risk mitigation strategies.

نشریه علمی پژوهشنامه بیمه

سایت نشریه: https://ijir.irc.ac.ir/?lang=fa

مقاله پژوهشی

پیش‌بینی خسارت با استفاده از تکنیک‌های رگرسیون عمیق متوالی در روش‌های یادگیری عمیق

مونا پرستش[۱*]، ضیاءالدین بهشتی فرد[۲]، عباس راد[۳]

[۱] دانشجوی دکتری، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

[۲] مربی، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

[۳] استادیار، گروه مدیریت، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران

اطلاعات مقاله

تاریخ های مقاله:

تاریخ دریافت: ۹ مهر ۱۴۰۳

تاریخ داوری: ۳۰ دی ۱۴۰۳

تاریخ پذیرش: ۱۵ بهمن ۱۴۰۳

تاریخ زودآیند: ۱۷ بهمن ۱۴۰۳

تاریخ انتشار: ۹ مهر ۱۴۰۴

کلمات کلیدی:

بیمه

خسارت آتش‌سوزی

یادگیری عمیق

یادگیری ماشین

*نویسنده مسئول:

ایمیل: St_m_parastesh@azad.ac.ir

تلفن: ۰۲۱ ۲۹۴۶۵۲۲۹

ORCID: 0009-0008-2851-7702

DOI: 10.22056/ijir.2025.04.02

چکیده:

پیشینه و اهداف: یکی از چالش‌های مهم شرکت‌های بیمه تعیین نرخ بهینهٔ بیمه‌نامه‌های اموال مانند خودرو و آتش‌سوزی است. اگر شرکت بیمه بتواند امکان وقوع خسارت را در یک بیمه‌نامه تشخیص دهد، تصمیمات مؤثر و بهتری در تعیین نرخ بیمه، میزان تخفیف اختصاص‌یافته به بیمه‌نامه یا تصمیم‌گیری دربارهٔ تمدید آن بیمه‌نامه خواهد داشت. شرکت‌های بیمه و خبرگان رشتهٔ صدور و خسارت، به‌دنبال روش‌های جدیدی برای ارزیابی ریسک مشتریان و بیمه‌نامه‌ها از طریق پیش‌بینی وقوع خسارات احتمالی در این رشته هستند. روش‌های سنتی مانند مدل‌های خطی عمومی (GLMs) غالباً در شناسایی الگوهای پیچیده در داده‌های بیمه ناکام‌اند. یادگیری عمیق با توانایی شناسایی روابط غیرخطی و الگوهای پیچیده می‌تواند این محدودیت‌ها را رفع کند. در این مقاله روشی برای بهبود عملکرد یادگیری عمیق با استفاده از تکنیک‌های رگرسیون عمیق ترتیبی ارائه شده است. رویکرد پیشنهادی ترکیبی از یادگیری عمیق و مدل‌های ترتیبی است. از شبکه‌های حافظهٔ طولانی کوتاه‌مدت (LSTM) برای مدل‌سازی داده‌های سری زمانی استفاده می‌شود.

روش‌شناسی: در این مقاله از داده‌های ۷ سال اخیر صدور و خسارت بیمهٔ آتش‌سوزی شرکت بیمهٔ البرز برای بررسی و پیش‌بینی خسارت در این رشته استفاده شده است. در این مقاله با تمرکز بر پیش‌پردازش داده‌ها و استخراج ویژگی‌های برتر برای ارائهٔ بهترین نتیجه و پس از اعمال روش‌های مختلف استخراج ویژگی، در نهایت از مجموع ۴۰ ویژگی، ۲۰ ویژگی انتخاب و سپس با استفاده یادگیری عمیق آموزش داده شد. رویکرد پیشنهادی، ترکیبی از یادگیری عمیق و مدل‌های ترتیبی با استفاده از شبکه‌های حافظهٔ طولانی کوتاه‌مدت (LSTM) برای مدل‌سازی داده‌های سری زمانی استفاده می‌کند.

یافته‌ها: در این مقاله با بررسی روش‌های مختلف یادگیری ماشینی روی داده‌های صدور و خسارت شرکت بیمه، این نتیجه به دست آمد که مدل رگرسیون عمیق ترتیبی نسبت به روش‌های سنتی عملکرد بهتری دارد. بهبود دقت پیش‌بینی، قابلیت اطمینان بالاتر و تأکید بر اهمیت ویژگی‌های زمانی از نتایج اصلی‌اند.

نتیجه‌گیری: بهینه‌سازی فرایندهای ارزیابی ریسک و خسارت از طریق مدل‌های یادگیری عمیق می‌تواند به ارائهٔ نرخ‌های حق‌بیمه دقیق‌تر و کاهش ضررهای ناشی از پرداخت‌های نامناسب منجر شود. همچنین، استفاده از این مدل‌ها می‌تواند به شرکت‌های بیمه کمک کند تا استراتژی‌های مدیریت ریسک بهتری اتخاذ کنند و فرایندهای کشف تقلب را بهبود بخشند. این موضوع به‌ویژه در بیمه‌های آتش‌سوزی، که خسارت‌های مالی سنگینی به همراه دارند، از اهمیت بالایی برخوردار است.

توجه: مدت‌زمان بحث و انتقاد برای این مقاله تا ۱ ژانویه ۲۰۲۶ در وب‌سایت IJIR در «نمایش مقاله» باز است.

مقدمه

امروزه نقش صنعت بیمه در حال تغییر است. به این دلیل که شرکت‌ها از روش‌های تحلیلی جدیدی برای پیش‌بینی خسارت و ریسک استفاده می‌کنند که این روش‌ها به آن‌ها در بررسی ریسک موجود کمک خواهد کرد. در عصر حاضر مدل‌های کسب‌وکار سنتی و روش‌های قدیمی همواره در معرض تهدید فنّاوری بوده‌اند. شرکت‌های بیمه‌ای جدید از قدرت فنّاوری‌های نوآورانه استفاده می‌کنند تا پیشگامان همیشگی بازار بیمه را از بین ببرند. حمایت و حفاظتی که در مقابل ریسک‌ها از بیمه‌گذار می‌شود و راه‌حل‌های پیشنهادی که برای مقابله با ریسک ارائه می‌گردد، از طریق سرویس‌های هشدار خطر به دست می‌آید و این سرویس‌ها مهم‌ترین وجه تمایز این شرکت‌ها و کلید موفقیت آن‌ها در آینده خواهد بود. هوش مصنوعی قوی و تحلیل حجم زیادی از اطلاعات داده‌های حجیم[1] به بیمه‌گران این قدرت را می‌دهد که به‌سمت پیش‌بینی خسارت و حوادث حرکت کنند. هرچه میزان اطلاعاتی که شرکت‌های بیمه از بیمه‌گذاران در اختیار دارند بیشتر باشد، آن‌ها خواهند توانست از این اطلاعات ارزشمند برای پیش‌بینی رفتار بیمه‌گذاران استفاده نمایند و برای هر بیمه‌گذار فایلی از اطلاعات گذشته او تهیه کنند تا به این وسیله حجم خسارات و ریسک را کاهش دهند. می‌توان گفت شرکت‌های بیمه‌ای که بخش عظیمی از فنّاوری‌های نوآورانهٔ نسل بعد را دربرمی‌گیرند، فرصت رشد بسیار زیادی در اختیار دارند. اما شرکت‌های بیمه‌ای که برای تعیین حق‌بیمه به سؤالات ساده‌ای مثل سن جنسیت و کار بسنده می‌کنند، بعید است در عصر دیجیتال و با پیشرفت اینشورتک بتوانند به کار خود ادامه دهند. در حقیقت اگر بیمه‌گران «تحلیل‌های پیش‌بینانه» را به کار نبرند و سیستم سنتی گذشته را اعمال کنند، در مقایسه با شرکت‌های نوآور در پرداخت خسارات دچار تأخیر بیشتری می‌شوند. همین امر باعث می‌شود شرکت‌هایی که این سیستم را در کار خود اعمال کرده‌اند، مشتریان بیشتری جذب کنند و در طولانی‌مدت طیف وسیعی از بیمه‌گذاران را دربرگیرند. استفاده از روش‌های سنتی یادگیری ماشین مانند رگرسیون خطی، درخت تصمیم و ... در مقادیر بالا و موارد خاص دارای خطای بالایی بوده و به همین دلیل در پی استفاده از روش‌های جدیدتر مانند یادگیری عمیق و بهبود این روش‌ها هستیم.

مبانی نظری پژوهش

ارزیابی ریسک بیمه

صنعت بیمهٔ کشور را می‌توان مؤسسهٔ بیمهٔ بزرگی فرض کرد که در همهٔ رشته فعالیت‌های بیمه‌ای مختلف فعالیت دارد. اگر بخواهیم تجمیع ریسک‌های بیمه‌گری را برای یک مؤسسهٔ بیمه معین انجام دهیم ممکن است آن مؤسسهٔ بیمه در همهٔ رشته فعالیت‌های بیمه‌ای فعالیت نداشته باشد؛ بنابراین در این حالت امکان مدل‌سازی ساختار وابستگی بین ریسک‌های بیمه‌گری همهٔ رشته فعالیت‌های بیمه‌ای با داده‌های آن مؤسسهٔ بیمه میسر نخواهد بود. البته مشابه همین تحقیق را می‌توان برای هر مؤسسهٔ بیمهٔ فرضی با داده‌های مربوطه انجام داد و از نتایج آن بهره برد (رستمی و همکاران، ۱۴۰۱).

منظور از ریسک بیمه‌گری، ریسک‌هایی است که مؤسسه بیمه به‌دلیل صدور بیمه‌نامه یا قبول اتکایی با آن مواجه است[2] مؤسسات بیمه با صدور بیمه‌نامه و در قبال دریافت حق‌بیمه از مشتریان متعهد می‌شوند که هنگام وقوع خطرات تحت پوشش بیمه، زیان وارده به مشتریان و بیمه‌گران خود را جبران و به آن‌ها خسارت پرداخت کنند. این مؤسسات به‌منظور اطمینان از توانایی مالی خود در انجام تعهدات و پرداخت خسارت در دورهٔ جاری و دوره‌های آتی، بخشی از درآمد حق‌بیمه را به‌صورت ذخایر بیمه‌ای[3] نگهداری می‌کنند؛ سپس از محل این ذخایر، دارایی‌های سرمایه‌گذاری را می‌خرند. عدم کفایت دارایی‌ها و وجوه نقد شرکت در پرداخت خسارت مشتریان، آسیب بزرگی به تداوم فعالیت مؤسسهٔ بیمه خواهد زد و در ادبیات بیمه به این وضعیت «عدم توانگری مالی»[4] گفته می‌شود (میرباقری جم و همکاران، ۱۴۰۱).

تجمیع ارزیابی ریسک برای بیمه‌نامه‌ها و خسارت‌های قبلی بر روی اقلام اطلاعاتی مؤثر، روشی کلیدی برای شرکت‌های بیمه است. مثلاً در بیمهٔ خودرو اگر فقط پرداخت خسارت برای یک بیمه‌نامه را بررسی کنیم، قطعاً برای ارزیابی ریسک نامناسب و کاملاً نادقیق است. درحالی‌که اگر این ارزیابی ریسک خسارت بر روی ۱۰۰ هزار بیمه‌نامهٔ مشابه قبلی انجام شود، می‌توان برای ارزیابی و پیش‌بینی با دقت قابل قبولی استفاده کرد (Lentz et al., 2015).

پیش‌بینی خسارت و روش‌های یادگیری ماشینی

برای پیش‌بینی خسارت در صنعت بیمه، روش‌های یادگیری ماشینی مختلفی استفاده می‌شود که هریک ویژگی‌ها و مزایای خود را دارند. در اینجا چند نمونه از این روش‌ها را بررسی می‌کنیم

الگوریتم‌های درخت تصمیم (Decision Trees)

درخت تصمیم یک مدل یادگیری ماشینی است که به‌صورت ساختار درختی تصمیم‌گیری را از داده‌های آموزشی یاد می‌گیرد. این درخت از گره‌ها و لبه‌هایی تشکیل شده است که هر گره نمایندهٔ یک ویژگی از داده و هر لبه نمایندهٔ یک تصمیم یا یک شاخه از گره‌هاست. در پایان هر شاخه (یعنی برگ درخت)، یک پیش‌بینی یا یک کلاس برای داده‌های ورودی قرار می‌گیرد و شامل مراحلی می‌شود، از جمله مرحلهٔ **انتخاب ویژگی**، که در آن براساس معیارهایی مانند انتروپی[5]، یا بهرهٔ اطلاعاتی[6]، بهترین ویژگی برای جدا کردن داده‌ها از یکدیگر انتخاب می‌شود؛ مرحلهٔ **تقسیم داده** که در آن داده‌ها براساس مقدار بهترین ویژگی به دو یا چند زیرمجموعه تقسیم می‌شوند؛ مرحلهٔ **ساخت درخت** که این فرایند به‌صورت بازگشتی

1. Big Data

۲. انجمن بین‌المللی اکچوئری (IAA) ریسک بیمه‌گری را به دو دسته «ریسک بیمهٔ زندگی» و ریسک بیمهٔ غیرزندگی» تقسیم کرده است.

3. Technical provisions
4. Insolvency
5. Entropy
6. Information gain

ادامه می‌یابد تا زمانی که شرایط پایانی برقرار شود، مانند رسیدن به حداقل تعداد نمونه در هر برگ یا عمق معین درخت. این الگوریتم‌ها به‌خوبی در تحلیل ریسک و تصمیم‌گیری دربارهٔ قیمت‌گذاری بیمه و تخصیص سرمایه استفاده می‌شوند. از مزایای آن می‌توان به قابلیت توضیح‌پذیری بالا، قابلیت ارزیابی دقیق ریسک‌ها، و قابلیت استفاده از ویژگی‌های مختلف داده اشاره کرد (Hanafy & Ming, 2021).

۱. روش‌های خوشه‌بندی(Clustering Methods)

این روش‌ها برای تشخیص الگوهای مشابه در داده‌ها و گروه‌بندی خطرات استفاده می‌شوند. هدف اصلی این است که داده‌ها درون هر خوشه به‌طور مشابهی عمل کنند و خوشه‌ها با یکدیگر متفاوت باشند. خوشه‌بندی می‌تواند به شناسایی الگوها و روابط پنهان در داده‌ها کمک کند که از طریق روش‌های سنتی مشخص نمی‌شوند. با استفاده از خوشه‌بندی می‌توان ابعاد داده را کاهش داد و به تحلیل و تفسیر آن‌ها کمک کرد. با تفکیک مشتریان یا پرونده‌ها براساس خطرات مشابه، بیمه‌گذاران می‌توانند مدیریت ریسک بهتری داشته باشند، و از مزایای آن یعنی امکان شناسایی گروه‌های خطرات مشابه و کاهش تقلب بهره ببرند. برای انجام خوشه‌بندی نخستین مرحله انتخاب یک الگوریتم خوشه‌بندی، مانند k-means یا hierarchical clustering است. سپس معمولاً نیاز به تعیین تعداد خوشه‌هاست که بهترین انتخاب براساس معیارهایی مانند ضریب سیلوئت[7] است (Hu & O'Hagan, 2020).

۲. روش‌های آماری پیشرفته (Advanced Statistical Methods)

این روش‌ها معمولاً بر پایهٔ روش‌های آماری پیچیده‌تری از جمله رگرسیون خطی یا غیرخطی، تحلیل عاملی، تحلیل خوشه‌ای و مانند آن‌ها هستند. این روش‌ها برای تحلیل دقیق داده‌ها و استخراج اطلاعات کلیدی برای پیش‌بینی خسارت استفاده می‌شوند. از جملهٔ این روش‌ها می‌توان به روش‌های رگرسیون خطی، غیرخطی و تحلیل‌های خوشه‌ای اشاره کرد. از این روش‌ها برای کاهش ابعاد و تحلیل ساختاری و همچنین امکان مدیریت بهتر ریسک‌ها استفاده می‌شود. این روش‌ها به‌دلیل توانایی‌شان در تحلیل دقیق داده‌های پیچیده و شناسایی الگوهای پنهان، از اهمیت ویژه‌ای برخوردارند و در صنعت بیمه بهبود درخور توجهی را در فرایندهای تصمیم‌گیری و مدیریت ریسک ایجاد کرده‌اند. در زمینه‌های مدل‌سازی و پیش‌بینی خسارت‌ها، می‌توان به تحلیل ریسک‌های مختلف، بهینه‌سازی قیمت‌گذاری بیمه و کاهش تقلب و بهبود مدیریت ریسک‌ها اشاره کرد (Jing et al., 2018).

۳. شبکه‌های عصبی عمیق (Deep Neural Networks - DNNs)

شبکه‌های عصبی یکی از قدرتمندترین و پرکاربردترین روش‌های یادگیری ماشینی هستند که به‌عنوان یک مدل محاسباتی بر پایهٔ عملکرد سیستم عصبی انسان طراحی شده‌اند. این شبکه‌ها به‌دلیل توانایی‌شان در یادگیری الگوهای پیچیده، پردازش تصویر، تحلیل متن و پیش‌بینی داده‌ها در بسیاری از صنایع و کاربردها از جمله صنعت بیمه نیز استفاده می‌شوند. همچنین به‌دلیل قابلیتشان در یادگیری از داده‌های بزرگ و توانایی شناسایی الگوهای پیچیده، مورد توجه‌اند. معماری شبکه‌های عصبی از تعدادی نورون و لایه تشکیل شده و دارای انواع مختلفی مانند شبکهٔ عصبی بازگشتی[8] یا کانولوشن[9] هستند. از مزایای این روش دقت بالا در پیش‌بینی‌ها، قابلیت انعطاف‌پذیری برای استفاده از انواع مختلف داده‌ها، و توانایی بهبود پیش‌بینی‌ها در طول زمان است.

هریک از این روش‌ها با ویژگی‌ها و مزایای خود، در پیش‌بینی خسارت در صنعت بیمه اهمیت دارند و بسته به نیازهای خاص شرکت‌های بیمه و نوع داده‌های موجود، انتخاب می‌شوند (Abdulkadir & Fernando, 2024).

مروری بر پیشینهٔ پژوهش

در سال‌های اخیر و با مورد توجه قرار گرفتن روش‌های هوش مصنوعی و یادگیری ماشین، تمایل به استفاده از الگوریتم‌های یادگیری ماشین در صنعت بیمه نیز افزایش یافته و در تحقیق‌های متعددی برای خوشه‌بندی مشتریان، پیش‌بینی خسارت و ارزیابی ریسک، نرخ‌دهی مناسب و ... از این روش‌ها استفاده کرده‌اند. در ادامه برخی پژوهش‌های انجام‌شده در سال‌های اخیر بررسی می‌شود

رئیسی وانانی و همکاران (۱۴۰۲) در پژوهشی برای پیش‌بینی خسارت رشتهٔ آتش‌سوزی از روش‌های یادگیری ماشین و همچنین در نهایت از یادگیری عمیق استفاده کردند که در نهایت در قالب برنامه‌ای، ورودی‌هایی از بیمه‌نامه را دریافت می‌کند و پیش‌بینی احتمال وقوع خسارت را به کاربران برمی‌گرداند. معنوی و همکاران (۱۴۰۰) در پژوهش خود یک روش برای پیش‌بینی امکان رخداد خسارت در بیمه‌نامه‌های شخص ثالث خودرو ارائه کردند. در روش پیشنهادی از روش‌های یادگیری ماشین استفاده شده و پس از استخراج ویژگی‌های مؤثر برای تشخیص خسارت یا عدم خسارت از بیمه‌نامه‌ها، به کمک روش‌های یادگیری ماشین و ویژگی‌های به‌دست‌آمده یک مدل پیش‌بینی امکان وقوع خسارت، آموزش داده می‌شود که در نهایت با استفاده از این مدل می‌توان در زمان فروش بیمه‌نامه امکان وقوع خسارت و ریسک فروش بیمه‌نامه را پیش‌بینی کرد.

تجددی نودهی و همکاران (۱۴۰۲) از یادگیری جمعی برای ترکیب روش‌های رگرسیون متعدد، مانند رگرسیون لجستیک، شبکه‌های عصبی برای پیش‌بینی هزینه‌های درمانی بیمه‌گذاران استفاده کردند.

ساندرکمار و راوی یک روش ترکیبی برای مسائل با داده‌های غیرمتوازن ارائه کردند، که قادر به کشف خودکار کلاهبرداری‌ها در شرکت‌های بیمه است (Sundarkumar & Ravi, 2015).

تاکور و سینگ مطالعه‌ای د ربارهٔ مشتریان بیمهٔ خودرو انجام داده و از الگوریتم بهبو یافته‌تری نسبت به روش k-means استفاده کرده‌اند

7. Silhouette Coefficient

8. RNN (Recurrent Neural Network)

9. CNN (convolution Neural Network)

عنوان مقاله	نویسنده و سال انتشار	روش تحقیق	نتایج و یافته‌ها
Estimating the demand factors and willingness to pay for agricultural insurance	(Gulseven, 2020)	Logistic Mode	این مقاله با کمک روش‌های آماری خطی و غیرخطی به تخمین احتمال خردی بیمهٔ کشاورزی ازسوی مردم می‌پردازد.
Bias regularization in neural network models for general insurance pricing	(Wüthrich, 2019)	Neural network models	از روش‌های شبکهٔ عصبی برای پیش‌بینی قیمت در بیمه استفاده شده است.
Machine Learning Approaches for Auto Insurance Big Data	(Hanafy & Ming, 2021)	رگرسیون لجستیک XGBoost Random Forest	از جنگل تصادفی برای تحلیل داده‌های بزرگ در بیمهٔ خودرو استفاده کرده‌اند. از روش‌هایی مانند رگرسیون لجستیک و XGBoost نیز در مقایسه استفاده شده است.
Prediction of motor insurance claims occurrence as an imbalanced machine learning problem.	(Baran & Rola, 2022)	رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، xgBoost، شبکهٔ پیش‌خور	مشکلات ناشی از عدم توازن داده‌ها در پیش‌بینی ادعاهای بیمهٔ خودرو پرداخته است از روش‌های مانند SMOTE استفاده شده است
A deep learning model for insurance claims predictions. *Journal on Artificial Intelligence*	(Abdulkadir & Fernando, 2024)	loss reserving - deep learning	از مدل ترتیبی شبکه‌های عصبی عمیق (Sequential deep regression) و تأثیر عملکرد مختلف توابع فعال‌سازی (ReLU vs Swish) استفاده شده است.
Prediction of automobile insurance claims using deep neural networks	(Reddy et al., 2024)	Newral network	از شبکه‌های عصبی عمیق برای پیش‌بینی احتمال وقوع خسارت در بیمهٔ خودرو استفاده می‌کند.
Forecasting the movements of Bitcoin prices: An application of machine learning algorithms	(Pabuçcu et al., 2020)	ماشین‌های بردار پشتیبان (SVM)، شبکهٔ عصبی مصنوعی (ANN)، خلیج‌های ساده (NB)	در این مقاله برای پیش‌بینی قیمت بیت از روش‌های مختلف رگریسیون، بردار پشتیبان، جمگل تصادفی و شبکهٔ عصبی استفاده شده است.
An interactive machine-learning-based electronic fraud detection system in healthcare insurance	(Kose et al., 2015)	Machin learning metods	در این مقاله برای کشف تقلب از روش‌های مختلف یادگیری ماشین مانند رگرسیون خطی و بردار پشتیبان استفاده شده است.
Prediction of claims in export credit finance: A comparison of four machine learning	(Bärtl & Krummaker, 2020)	شبکه‌های عصبی احتمالی و روش‌های یادگیری ماشین	چهار روش یادگیری ماشین (ML) (درخت تصمیم (DT)، جنگل‌های تصادفی (RF)، شبکه‌های عصبی (NN) و شبکه‌های عصبی احتمالی (PNN)) را بر پیش‌بینی دقیق ادعاهای بیمه اعتبار صادراتی ارزیابی می‌کند.
A Comparative Study of Data Mining Algorithms in the Prediction of Auto Insurance Claims	(Weerasinghe & Wijegunasekara, 2016)	الگوریتم‌های یادگیری ماشین	در این مقاله به مقایسهٔ روش‌های مختلف یادگیری ماشین برای پیش‌بینی خسارت خودرو پرداخته شده است.
Discussion of "Machine learning applications in non-life insurance	(Banks, 2020)	الگوریتم‌های یادگیری ماشین	مقایسهٔ مدل‌های نگهداری برای بیمهٔ خانگی و سپس مشکل قیمت‌گذاری پویا برای بیمهٔ موتور آنلاین ارائه می‌شود.
Using Machine Learning to Better Model Long-Term Care Insurance Claims. North American	(Cummings & Hartman, 2022)	سری‌های زمانی	بهبود روش‌های سری زمانی در یادگیری ماشین و پیش‌بینی خسارت.

که با توجه به ویژگی‌های خاص مشتریان قدرت پیش‌بینی عملکرد مشتریان را افزایش می‌دهد (Thakur & Sing 2013).

بالاجی و سریواتسا روش‌های مورد استفاده برای پیش‌بینی داده‌ها برای بیمهٔ عمر را طبقه‌بندی کرده‌اند. آن‌ها الگوریتم‌های مختلفی چون روش دسته‌بندی نایو بیز و شبکه‌های بیزی را برای دسته‌بندی داده‌ها ارزیابی کرده‌اند (Balaji & Srivatsa, 2012).

پسانتز ـ ناروائز و همکاران از دو روش رقیب XGBoost و رگرسیون لجستیک، برای پیش‌بینی فراوانی خسارت‌های بیمهٔ خودرو استفاده

policyverid	maxPolicyVerNo	PolicyIssueTypeId	PolicyTypeId	InsLineId	duration	InsItemCityId	IsInspectionReqId	HasInspectionId	HasContractId	...	coverid182	coveric
13280207	3	1	23	3	366	1202.0	0	0	1	...	0	
12686299	0	1	23	3	366	2111.0	0	0	0	...	0	
12004077	1	1	21	1	365	2127.0	1	1	0	...	0	
12009709	0	1	23	3	365	2131.0	1	1	0	...	0	
11414927	1	1	21	1	365	1908.0	1	1	0	...	0	
11332656	2	1	22	2	366	701.0	1	0	1	...	0	
13287544	0	1	23	3	366	701.0	1	1	0	...	0	
12691162	2	1	27	213	366	2003.0	0	0	8	...	0	
11722612	1	1	23	3	365	701.0	0	0	2	...	0	
11340126	0	1	22	2	365	501.0	1	0	1	...	0	

s × 91 columns

شکل ۱. نمونهٔ دیتاست پژوهش
Fig. 1. Sample research dataset

کردند. این مطالعه نشان می‌دهد که مدل XGBoost کمی بهتر از رگرسیون لجستیک است (Pesantez-Narvaez et al., 2019).

عبدالهادی و همکاران در این تحقیق از چهار طبقه‌بندی‌کننده برای پیش‌بینی وقوع خسارت استفاده کرده‌اند، از جمله الگوریتم‌های ANN، J48، XGBoost و nave Bayes که بهبودی بر روش‌های طبقه‌بندی بود و در این تحقیق نیز مدل XGBoost در بین چهار مدل بهترین عملکرد را داشت (Abdelhadi et al. 2020).

در سال ۲۰۲۰، گریز و همکاران در پژوهش خود روش‌های مختلف پیش‌بینی را تلفیق کردند و مدلی ترکیبی برای پیش‌بینی دقیق‌تر خسارت‌های بیمه‌ای غیرزندگی ارائه دادند (Grize et al., 2020).

بوخر و روزنستوک مدل نوآورانه‌ای در سطح بالاتر و با بهره‌گیری از شبکهٔ عصبی برای پیش‌بینی تعداد خسارت ارائه کرده‌اند که دارای نتایج مطلوبی بوده و از سطح ماکرو برای پیش‌بینی خسارت‌های آینده استفاده کرده است (Bücher & Rosenstock, 2023).

برخی دیگر از پژوهش‌های انجام‌شده در این زمینه را در جدول زیر می‌توان مشاهده کرد:

Data Set[10] و مدل پیشنهادی

برای نمونهٔ آماری در این پژوهش از داده‌های صدور و خسارت مربوط به ۷ سال از بیمه‌نامه‌های رشتهٔ آتش‌سوزی شرکت بیمهٔ البرز استفاده شده است که اطلاعات سال‌های ۱۳۹۵ تا ۱۴۰۲ را شامل می‌شود. در این فواصل زمانی حدود ۳ میلیون بیمه‌نامهٔ آتش‌سوزی صادر و حدود ۶۰ هزار پروندهٔ خسارت ثبت شده است.

به منظور رفع مشکل داده‌های نامتوازن در دیتاست مورد نظر، از روش‌های balancing دیتا استفاده شده و همچنین با روش‌های مختلف پیش پردازش، داده‌های مورد نظر اصلاح و نواقص احتمالی برطرف گردیده است.

در ابتدا، بانظر کارشناسان و خبرگان صنعت بیمه در رشته آتش‌سوزی، اطلاعات تاثیر گزار در وقوع خسارت استخراج و از دیتابیس بیمه ای شرکت، تهیه گردید. سپس با استفاده از روش‌های استخراج ویژگی، تعداد ۱۰ ویژگی انتخاب و در الگوریتم‌های مورد نظر استفاده شد. در ابتدا با استفاده از روش‌های درخت تصمیم، رگرسیون و در نهایت یادگیری عمیق، مدل با داده‌های آموزی، آموزش و تست گردید.سپس با استفاده از روش ترکیبی مورد نظر،فرایند آموزش تکرار و دقت دوباره بررسی شد.

روش‌شناسی پژوهش

در شکل ۳ الگوریتم اجرای تحقیق را مشاهده می‌کنید که در مرحلهٔ انتخاب داده‌ها شامل پرس‌وجو دربارهٔ نحوهٔ انتخاب ویژگی‌ها، جمع‌آوری داده‌ها، از دیتابیس بیمه‌ای شرکت بیمه البرز و ... بوده و سپس به‌صورت یک فرایند تکراری چندین مرحله پیش‌پردازش داده‌ها انجام شده و به داده‌های نسبتا هماهنگ دست پیدا کردیم. برای پیاده‌سازی این الگوریتم از زبان برنامه‌نویسی پایتون استفاده شده و خروجی‌ها در ادامه قابل مشاهده‌اند. سپس داده‌های مهم‌تر و با تأثیر بیشتر استخراج و با تقسیم به دو دستهٔ آموزش و تست به مرحلهٔ مدل یادگیری ارسال شدند. در این مرحله روی داده‌های آموزشی، آموزش اجرا و سپس با استفاده از داده‌های تست، دقت روش بررسی شد.

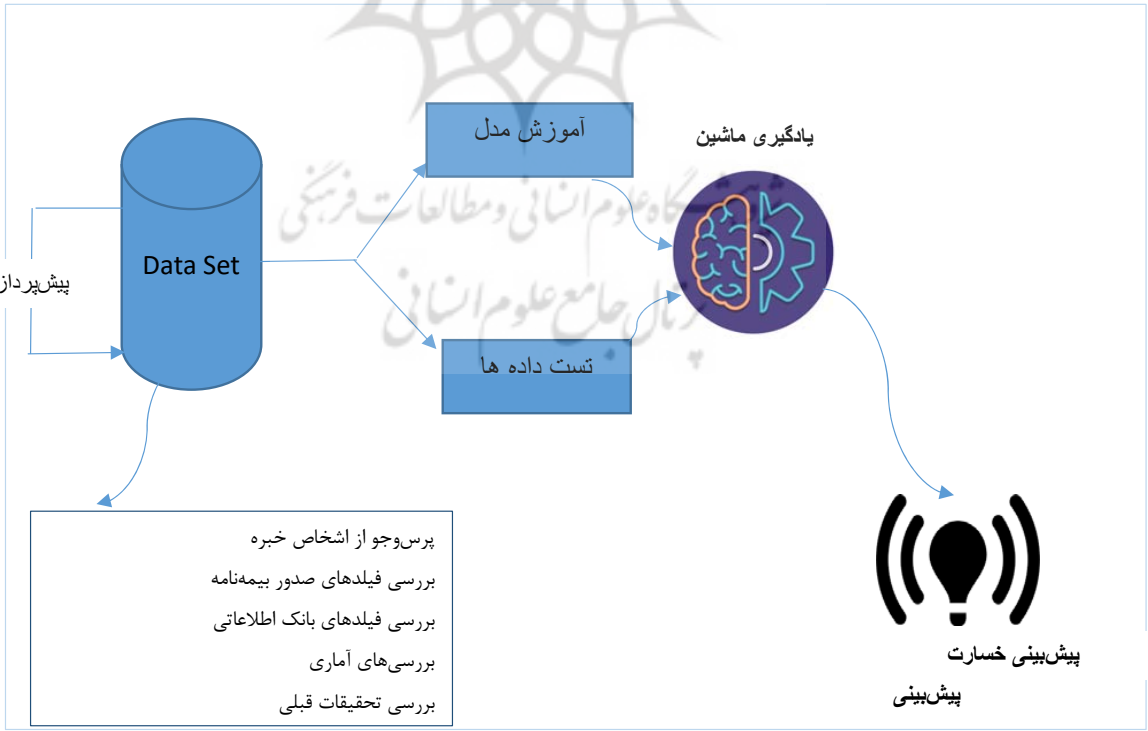
مراحل انجام تحقیق عبارت‌اند از:

- بارگذاری و پردازش داده‌ها؛
- تقسیم داده‌ها به مجموعه‌های آموزشی و تست؛
- تعریف مدل یادگیری عمیق ترتیبی؛

10. نمونهٔ استفاده‌شده در پژوهش

ردیف	نام قلم اطلاعاتی	نام استفاده‌شده در نمونه آزمایشی	شرح اطلاعات
1	نوع بیمه‌نامه	PolicyIssueTypeId	عادی/ طرح
2	نوع طرح بیمه‌نامه	PolicyTypeId	طرح جامع خانوار، طرح بیمهٔ اصناف، صنعتی، غیرصنعتی، مسکونی، انبار
3	زیررشتهٔ بیمه‌نامه	InsLineId	آتش‌سوزی صنعتی، غیرصنعتی، انبار، مسکونی
4	مدت بیمه‌نامه	duration	طول دورهٔ بیمه‌نامه به روز
5	شهر صدور بیمه‌نامه	InsItemCityId	شهر صدور بیمه‌نامه (که با استان و منطقهٔ خطر طبقه‌بندی شد)
6	بازدید بیمه‌نامه	HasInspectionId	بازدید انجام شده یا خیر
7	قرارداد	HasContractId	قرارداد دارد یا خیر
8	حق‌بیمه دستی دارد؟	HasAnnualPrmId	
9	قبلاً خسارت داشته ؟	HasLastPolicyClaimHistoryId	
10	مرهوناتی است؟	IsMortgageId	
11	نوع فعالیت	ActivityTypeId	چوب، شیشه و ...، محصولات فلزی، ماشین‌آلات و تجهیزات، و....
12	نوع سازه	ConstructionTypeId	گلی، فلزی، چوبی و ....
13	منطقهٔ خطر	DensityRiskTypeId	به 4 منطقه تقسیم شده است
14	تخفیف چندساله دارد؟	IsMultiYearDiscountReqId	
15	ذینفع دارد؟	HasBenefId	بیمه‌نامه دارای ذینفع است یا خیر
16	نوع پوشش	RiskClassId	از 1 تا 15

شکل ۲. اقلام اطلاعاتی استفاده‌شده در پژوهش
Fig. 2. Information items used in the research



شکل ۳. الگوریتم اجرای تحقیق
Fig. 3. Research implementation algorithm

- آموزش مدل و ارزیابی عملکرد؛
- بارگذاری و پردازش داده‌ها.

در ابتدا داده‌های استخراج‌شده از دیتابیس بارگذاری شدند و مراحل پیش‌پردازش روی آن‌ها انجام شد که شامل جست‌وجو و اصلاح داده‌های null، نرمال‌سازی اطلاعات عددی، گسسته‌سازی اطلاعات توصیفی و ... است. چون داده‌های بیمه‌نامه‌ها و تعداد خسارت‌های وقوع‌یافته عدم توازن داشتند (تعداد خسارت‌ها بسیار کمتر از تعداد بیمه‌نامه بود) به‌منظور برقراری توازن بین داده‌ها، از روش وزن‌دهی به مقادیر استفاده شد (داده‌های دارای خسارت با وزن بیشتری وارد دیتاست شدند) و دوباره الگوریتم‌های مورد نظر اجرا شدند.

از میان ۴۰ ویژگی استخراج‌شده برای بیمه‌نامه‌ها، ۲۰ ویژگی با استفاده از الگوریتم‌های انتها ویژگی انتخاب شدند که اثر بیشتری بر نتیجه داشتند و در ادامه کار الگوریتم‌های یادگیری روی این ویژگی‌ها آموزش داده شدند. در شکل ۳ ویژگی‌های استخراج‌شده را در زبان پایتون مشاهده می‌کنید.

سپس داده‌های پالایش‌شده به دو بخش آموزش و تست تقسیم شدند. عملیات یادگیری روی داده‌های آموزش انجام و روی داده‌های تست اعتبارسنجی شد.

مدل‌های یادگیری عمیق ترتیبی می‌توانند داده‌های پیچیده‌تر و با ویژگی‌های متنوع‌تر را پردازش کنند. این امر باعث می‌شود که این مدل‌ها بتوانند الگوهای پیچیده‌تری را در داده‌ها شناسایی کنند و پیش‌بینی‌های دقیق‌تری انجام دهند. در روش‌های یادگیری عمیق سنتی یکی از رایج‌ترین توابع فعال‌سازی تابع ReLU است که به‌دلیل سادگی و کارایی آن، به‌طور گسترده‌ای استفاده می‌شود. اما ReLU ممکن است در برخی موارد به مشکل «مرگ نورون‌ها» منجر شود که در آن برخی نورون‌ها هرگز فعال نمی‌شوند. در تحقیق پیش رو، روش به‌کاررفته، ترکیبی از یادگیری عمیق و مدل‌های ترتیبی است و از شبکه‌های حافظهٔ طولانی کوتاه‌مدت (LSTM) برای مدل‌سازی داده‌های سری زمانی استفاده می‌شود. در این روش از تابع Swish به‌جای تابع relu استفاده شده است.

تابع swish تابع فعال‌سازی جدیدتری است که محققان گوگل آن را معرفی کرده‌اند. این تابع به‌دلیل گرادیان‌های نرم‌تر و توانایی مدل‌سازی روابط پیچیده‌تر در داده‌ها، عملکرد بهتری در مقایسه با ReLU نشان داده است. Swish معادلهٔ Swish(x)=x·sigmoid(x) را دارد که باعث می‌شود به‌طور پیوسته فعال باشد و بهبود کارایی مدل را به همراه داشته باشد.

مزایای استفاده از تابع Swish

اطلاعات بیشتر: Swish مقادیر منفی را به‌طور کامل حذف نمی‌کند، بلکه به آن‌ها اجازه می‌دهد تا به مقدار کمی از جریان اطلاعات کمک کنند. این می‌تواند به یادگیری بهتر و استفادهٔ مؤثرتر از تمام داده‌ها منجر شود.

گرادیان نرم‌تر: تابع Swish گرادیان‌های نرم‌تری نسبت به ReLU تولید می‌کند. این امر به بهبود فرایند بهینه‌سازی کمک می‌کند و مشکلات مربوط به گرادیان‌های صفر را کاهش می‌دهد.

پدیدهٔ «مرگ نورون‌ها»: برخلاف ReLU که نورون‌ها ممکن است «مرده» شوند و برای همیشه غیرفعال باقی بمانند، Swish این مشکل را ندارد و نورون‌ها همچنان فعال باقی می‌مانند.

دقت بیشتر: به‌دلیل توانایی بهتر Swish در مدیریت اطلاعات و بهینه‌سازی گرادیان‌ها، مدل‌های استفاده‌کننده از Swish معمولاً دقت بالاتری نسبت به مدل‌های استفاده‌کننده از ReLU دارند.

این تفاوت‌ها به بهبود عملکرد کلی مدل‌های یادگیری عمیق منجر می‌شود و می‌تواند به دستیابی به نتایج دقیق‌تر و کارآمدتر کمک کند.

در ادامه یک مدل ترتیبی با چندین لایه Dense و Dropout تعریف و تابع فعال‌سازی Swish در لایه‌ها استفاده شد.

نتایج و بحث

در بررسی نتایج و یافته‌ها در روش‌های یادگیری ماشین عموماً پارامترهای زیر ارزیابی می‌شوند که در پژوهش پیش رو نیز بررسی و مقایسه شده است:

Accuracy (دقت کلی): نسبت نمونه‌های درست طبقه‌بندی‌شده به کل نمونه‌ها، که معیاری کلی برای ارزیابی عملکرد مدل است.

$$\text{Accuracy} = \frac{TP + TN}{FP + FN + TP + TN}$$

که در آن

- TP (True Positives): تعداد نمونه‌های مثبت که درست

```
ویژگی‌های انتخاب شده: Index(['duration', 'InsItemCityId', 'IsInspectionReqId', 'HasInspectionId',
       'HasContractId', 'IsMortgageId', 'ActivityTypeId', 'ActivitySubTypeId',
       'ConstructionTypeId', 'HasBenefId', 'itemtype1', 'coverid2', 'coverid4',
       'coverid7', 'coverid10', 'coverid13', 'coverid16', 'coverid21',
       'coverid22', 'RiskClassId'],
      dtype='object')
```

شکل ۴. ویژگی‌های استخراج‌شده پس از مرحلهٔ استخراج ویژگی
Fig. 4. Extracted features after the feature extraction

پیش‌بینی شده‌اند.

- TN (True Negatives):: تعداد نمونه‌های منفی که درست پیش‌بینی شده‌اند.
- FP (False Positives): تعداد نمونه‌های منفی که به‌اشتباه مثبت پیش‌بینی شده‌اند.
- FN (False Negatives): تعداد نمونه‌های مثبت که به‌اشتباه منفی پیش‌بینی شده‌اند.

Precision (دقت): نسبت نمونه‌های مثبت درست شناسایی‌شده (True Positives) به کل نمونه‌هایی که به‌عنوان مثبت پیش‌بینی شده‌اند. این معیار نشان می‌دهد که چه درصدی از نمونه‌های پیش‌بینی‌شده به‌عنوان مثبت، واقعاً مثبت بوده‌اند.

$$Precision = \frac{TP}{FP + TP}$$

Recall (بازخوانی یا حساسیت): نسبت نمونه‌های مثبت درست شناسایی‌شده به کل نمونه‌های مثبت واقعی، که بیانگر توانایی مدل در شناسایی تمام موارد مثبت است.

$$Recal = \frac{TP}{FN + TP}$$

این معیار نشان می‌دهد که چه درصدی از نمونه‌های مثبت واقعی توسط مدل شناسایی شده‌اند.

F1 Score: میانگین هارمونیک Precision و Recall که تعادلی میان این دو معیار ایجاد می‌کند و برای ارزیابی مدل‌هایی که دچار عدم توازن در داده‌ها هستند، مفید است.

$$F1 = \frac{Precision \times Recal}{Precision + Recal} \times 2$$

این معیار میانگین هارمونیک Precision و Recall است و زمانی که تعادل بین این دو معیار مهم باشد، کاربرد دارد.

نتایج زیر براساس آموزش مدل با یادگیری عمیق و با دو روش استفاده از تابع فعالسازی ReLU و تابع Swish به دست آمده است:

با توجه به مقایسهٔ این دو تابع فعال‌سازی در شبکه‌های عمیق، استفاده از Swish به‌جای ReLU بهبودهای چشمگیری در عملکرد و دقت مدل‌های عمیق می‌آورد. این بهبودها عموماً به‌دلیل بهینه‌تر بودن در استفاده از اطلاعات ورودی، کاهش مشکلات گرادیان‌ها و افزایش پایداری شبکه‌های عصبی است. در شکل زیر مقایسهٔ این دو روش قابل مشاهده است.

این نمودارها نشان می‌دهند که مدل جدید با استفاده از Swish دقت بیشتری در پیش‌بینی‌ها دارد، زیرا پیش‌بینی‌ها به مقادیر واقعی نزدیک‌ترند و نوسانات کمتری دارند. این امر به‌دلیل عملکرد بهتر تابع Swish در مقایسه با ReLU است که به هموارتر شدن گرادیان‌ها و بهبود دقت پیش‌بینی‌ها منجر می‌شود. در نمودار سمت راست نقاط پیش‌بینی‌شده (خط تیره سبز) نسبت به خروجی‌های واقعی (خط ممتد آبی) نویز کمتری دارند و به داده‌های واقعی نزدیک‌ترند. این موضوع نشان‌دهندهٔ عملکرد بهتر مدل جدید است که از تابع فعال‌سازی Swish استفاده می‌کند.

در ادامه یک نمونه پیاده‌سازی برای استفاده از این روش در صنعت بیمه برای پیش‌بینی خسارت پیشنهاد می‌شود. در این نمونه یک اپلیکیشن ساده با استفاده از زبان پایتون پیاده‌سازی شده و شامل ورودی‌های استفاده‌شده در تحقیق است که پس از ورود اطلاعات و فشردن دکمهٔ پیش‌بینی، الگوریتم مورد نظر به کار گرفته می‌شود و پیش‌بینی خسارت را براساس مدل یادگیری عمیق پیشنهادشده ارائه می‌دهد.

جمع‌بندی و پیشنهادها

مدل ارائه‌شده مدلی برای ارزیابی خسارت آتش‌سوزی است که با استفاده از روش‌های یادگیری عمیق آموزش دیده و پیش‌بینی وقوع خسارت را انجام می‌دهد. این مدل در یادگیری عمیق به‌جای تابع Relu از تابع Swish استفاده می‌کند و نتایج به‌دست‌آمده نشان می‌دهد که استفاده از این روش، خروجی بهتری دارد و دقت نتایج

```
Epoch 1/10, Loss: 0.1317, Test Loss: 0.3351, Test Accuracy: 0.8468
Epoch 2/10, Loss: 0.2547, Test Loss: 0.3149, Test Accuracy: 0.8566
Epoch 3/10, Loss: 0.0754, Test Loss: 0.3074, Test Accuracy: 0.8608
Epoch 4/10, Loss: 0.0516, Test Loss: 0.3047, Test Accuracy: 0.8592
Epoch 5/10, Loss: 0.3080, Test Loss: 0.3023, Test Accuracy: 0.8576
Epoch 6/10, Loss: 0.1303, Test Loss: 0.2998, Test Accuracy: 0.8618
Epoch 7/10, Loss: 0.7462, Test Loss: 0.3003, Test Accuracy: 0.8614
Epoch 8/10, Loss: 0.1978, Test Loss: 0.2975, Test Accuracy: 0.8634
Epoch 9/10, Loss: 0.3962, Test Loss: 0.2986, Test Accuracy: 0.8627
Epoch 10/10, Loss: 0.1612, Test Loss: 0.2959, Test Accuracy: 0.8601
```

شکل ۵. اجرای الگوریتم یادگیری عمیق با تابع relue

Fig. 5. Implementing a deep learning algorithm with the relue function

```
Epoch 1/10, Loss: 0.1105, Test Loss: 0.3201, Test Accuracy: 0.8562
Epoch 2/10, Loss: 0.0902, Test Loss: 0.2999, Test Accuracy: 0.8660
Epoch 3/10, Loss: 0.0754, Test Loss: 0.2850, Test Accuracy: 0.8715
Epoch 4/10, Loss: 0.0650, Test Loss: 0.2802, Test Accuracy: 0.8738
Epoch 5/10, Loss: 0.0575, Test Loss: 0.2750, Test Accuracy: 0.8760
Epoch 6/10, Loss: 0.0503, Test Loss: 0.2701, Test Accuracy: 0.8785
Epoch 7/10, Loss: 0.0452, Test Loss: 0.2650, Test Accuracy: 0.8808
Epoch 8/10, Loss: 0.0402, Test Loss: 0.2610, Test Accuracy: 0.8832
Epoch 9/10, Loss: 0.0360, Test Loss: 0.2575, Test Accuracy: 0.8850
Epoch 10/10, Loss: 0.0325, Test Loss: 0.2540, Test Accuracy: 0.8865
```

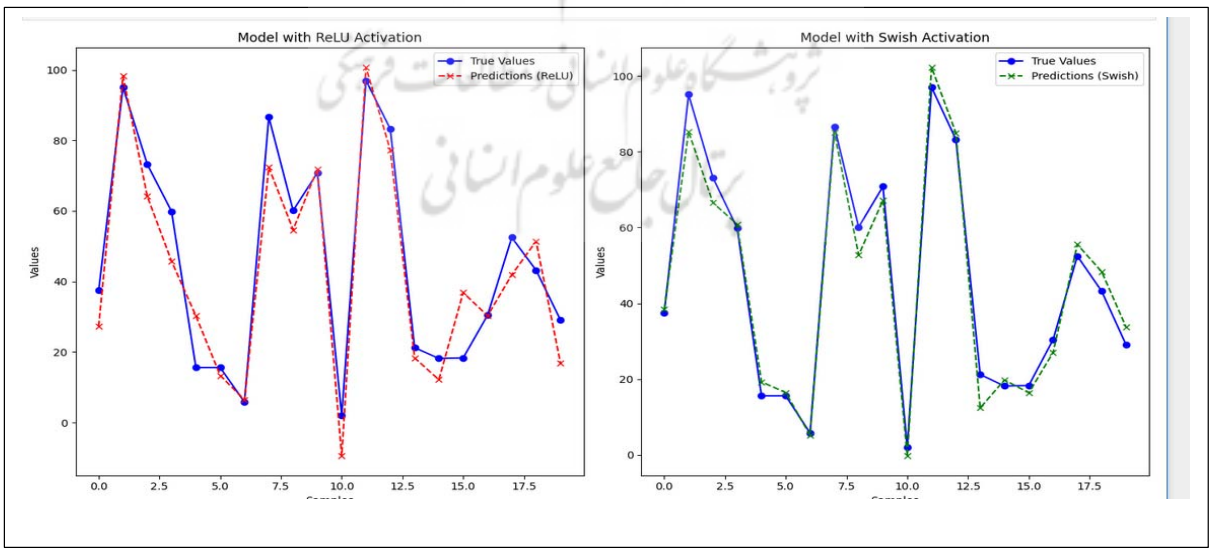
شکل ۶. اجرای الگوریتم یادگیری عمیق با تابع Swich

Fig. 6. Implementing a deep learning algorithm with the Switch function

(فرضی) Model with Swish	Model with ReLU	Metric
0.88	0.861	Accuracy
0.87	0.85	Precision
0.86	0.84	Recall
0.865	0.845	F1 Score

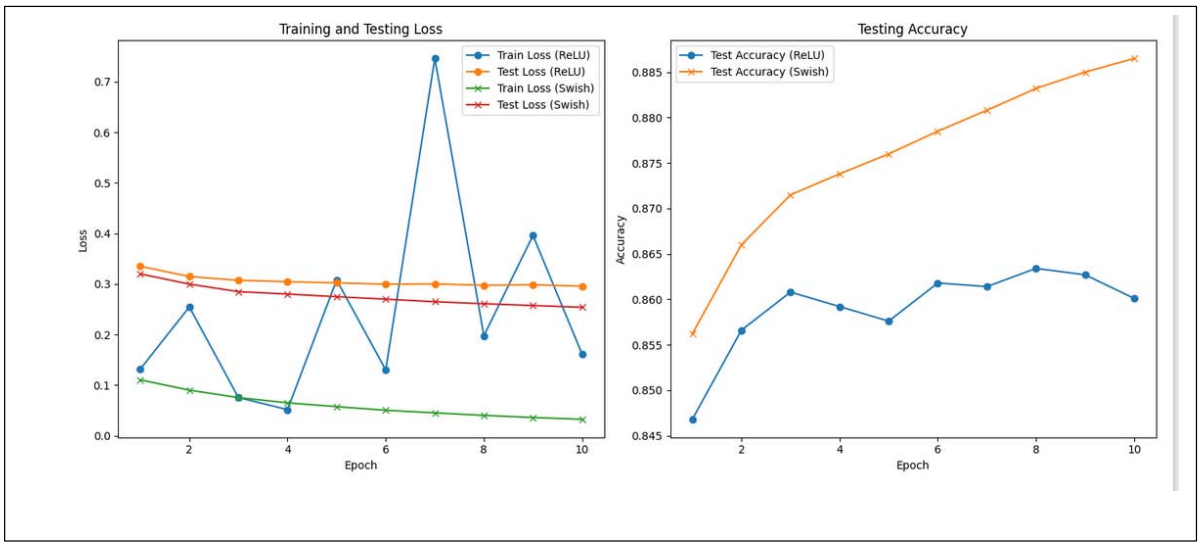
شکل ۷. مقایسه پارامترهای الگوریتم یادگیری عمیق با دو روش Relu و Swish

Fig. 7. Comparison of deep learning algorithm parameters with the two Relu and Swish methods



شکل ۸. اجرای الگوریتم یادگیری عمیق با تابع swish و relue

Fig. 8. Implementing a deep learning algorithm with the swish and relue functions



شکل ۹. مقایسهٔ عملکرد الگوریتم یادگیری عمیق با تابع Swish و ReLU

Fig. 9. Comparison of the performance of the deep learning algorithm with the swish and relue functions

شکل ۱۰. نمونه اپلیکیشن خروجی برای به‌کارگیری روش پیشنهادی

Fig. 10. Example of output application for applying the proposed method

را به‌طور قابل ملاحظه‌ای افزایش داده است. برای آموزش مدل از داده‌های صدور و خسارت آتش‌سوزی ۷ سال اخیر بیمهٔ البرز و از زبان برنامه‌نویسی پایتون برای آموزش و پیش‌بینی مدل استفاده شده است. بعد از آموزش مدل با استفاده از یادگیری عمیق به روش سنتی دقت حدودی ۰٫۸۶ برآورد شد. پس از بهره گرفتن از تکنیک‌های رگرسیون عمیق متوالی در یادگیری عمیق و پس از اجرای آموزش با استفاده از تابع Swish دقت به ۰٫۸۸ افزایش پیدا کرد. این بهبود به‌دلیل ویژگی‌های تابع نرم‌تر و گرادیان‌های غیرصفر برای مقادیر منفی، به دست آمد. نتایج نشان داد که استفاده از این روش در صنعت بیمه می‌تواند موجب کاهش میزان خطای تخمین خسارت و افزایش دقت محاسبهٔ میزان حق‌بیمه شود.

از نظر کاربردی، این یافته‌ها برای شرکت‌های بیمه اهمیت بالایی دارند. بهینه‌سازی فرایندهای ارزیابی ریسک و خسارت از طریق مدل‌های یادگیری عمیق می‌تواند به ارائهٔ نرخ‌های حق‌بیمه دقیق‌تر و کاهش ضررهای ناشی از پرداخت‌های نامناسب منجر شود. همچنین، استفاده از این مدل‌ها می‌تواند به شرکت‌های بیمه کمک کند تا استراتژی‌های مدیریت ریسک بهتری اتخاذ کنند و فرایندهای کشف تقلب را بهبود بخشند. این موضوع به‌ویژه در بیمه‌های آتش‌سوزی، که خسارت‌های مالی سنگینی به همراه دارند، از اهمیت بالایی برخوردار است.

پیشنهادهای آینده

نتایج این پژوهش نشان می‌دهد که استفاده از روش‌های هوش مصنوعی و یادگیری عمیق در صنعت بیمه می‌تواند تحولی بزرگ در پیش‌بینی و ارزیابی خسارت ایجاد کند. امید است که با ادامهٔ پژوهش‌ها در این حوزه، روش‌های دقیق‌تر و کاربردی‌تری برای تحلیل ریسک‌های بیمه‌ای ارائه شوند. در ادامه برخی پیشنهادها برای ادامهٔ‌پژوهش در این زمینه ارائه می‌شود:

۱. بهبود مدل‌های یادگیری عمیق: پیشنهاد می‌شود پژوهش‌های آینده بر ترکیب مدل‌های یادگیری عمیق با الگوریتم‌های بهینه‌سازی مانند Bayesian Optimization یا Genetic Algorithms برای بهینه‌سازی هایپرتیون‌ها تمرکز کنند.

۲. استفاده از داده‌های متنوع‌تر: در این پژوهش، داده‌های استفاده‌شده محدود به بیمه‌های آتش‌سوزی بودند. پیشنهاد می‌شود مطالعات آینده داده‌های دیگر حوزه‌های بیمه مانند بیمهٔ عمر و بیمهٔ سلامت را نیز در نظر بگیرند.

۳. بررسی نقش عوامل محیطی: شرایط جغرافیایی، وضعیت اقتصادی و سیاست‌های بیمه‌ای کشورها می‌توانند تأثیر بسزایی در پیش‌بینی میزان خسارت داشته باشند. گسترش این پژوهش به عوامل کلان اقتصادی می‌تواند به بهبود دقت مدل‌ها کمک کند.

۴. ادغام با روش‌های سنتی بیمه‌گری: برای پذیرش این روش‌ها در صنعت بیمه، لازم است مدل‌های یادگیری عمیق با روش‌های سنتی آماری و قوانین بیمه‌گری ادغام شوند تا تصمیم‌گیری‌ها بهینه‌تر و قابل اعتمادتر شوند.

۵. توسعهٔ سامانه‌های هوشمند بیمه‌ای: پیشنهاد می‌شود که نتایج این پژوهش در یک سامانهٔ تصمیم‌یار پیاده‌سازی شود تا شرکت‌های بیمه بتوانند از آن برای ارزیابی سریع و دقیق خسارت‌ها استفاده کنند. توسعهٔ یک نرم‌افزار مبتنی بر هوش مصنوعی می‌تواند به شرکت‌های بیمه در پردازش درخواست‌های خسارت کمک کرده، فرایندهای اداری را تسهیل کند.

مشارکت نویسندگان

مونا پرستش: مفهوم و طرح مقاله، جمع‌آوری و اخذ داده‌ها، تحلیل و تفسیر داده‌ها، تحلیل آماری، پیش‌نویس مقاله. ضیاءالدین بهشتی: مفهوم و طرح مقاله، تحلیل آماری، بازنگری مقاله، ارائهٔ اصلاحات. عباس راد: تحلیل آماری، بازنگری مقاله، ارائهٔ اصلاحات.

تشکر و قدردانی

از شرکت معظم بیمه البرز که خانهٔ دوم من است و با در اختیار قرار دادن اطلاعات خسارت سبب تکمیل این تحقیق شد، همچنین از تمامی همکاران و مدیران با اخلاق این مجموعه تشکر می‌کنم.

تعارض منافع

نویسنده(گان) اعلام می‌دارند که در خصوص انتشار این مقاله تضاد منافع وجود ندارد. علاوه‌براین، موضوعات اخلاقی شامل سرقت ادبی، رضایت آگاهانه، سوءرفتار، جعل داده‌ها، انتشار و ارسال مجدد و مکرر ازسوی نویسندگان رعایت شده است.

یادداشت ناشر

ناشر نشریهٔ *پژوهشنامه بیمه* با توجه به مرزهای حقوقی در نقشه‌های منتشرشده بی‌طرف باقی می‌ماند.

منابع

تجددی نودهی، م.، حسینی خطیبانی، س.، یزدی‌نژاد، م.، و زلفی، س. (۱۴۰۲). پیش‌بینی هزینه‌های بیمه درمانی افراد با استفاده از یادگیری ماشین و روش یادگیری جمعی. *پژوهشنامه بیمه*، ۱۳(۱)، ۱-۱۴. https://doi.org/10.22056/ijir.2024.01.01

رستمی، م.، بنی حسینی، س. م.، و صفائی، س. (۱۴۰۱). *الگوریتم هوشمند ارزیابی و پیش‌بینی ریسک و خسارت بیمه درمان تکمیلی با استفاده از روش شبکه عصبی مصنوعی چندلایه* [مقاله کنفرانسی]. بیست و نهمین همایش ملی و دهمین همایش بین‌المللی بیمه و توسعه، تهران، ایران. https://civilica.com/doc/1770257

رئیسی وانانی، ا.، تقوی‌فرد، م. ت، سهرابی، ب.، و امیرحسینی، م. (۱۴۰۲). طراحی سیستم ارزیابی هوشمند جهت پیش‌بینی خسارت بیمه‌های آتش‌سوزی با استفاده از یادگیری عمیق. *پژوهشنامه بیمه*، ۱۲(۴)، ۲۵۱-۲۶۴. https://doi.org/10.22056/ijir.2023.04.01

معنوی، ف.، دادگر، م.، و رحمتیان، م. (۱۴۰۰). *ارائه روش پیش‌بینی خسارت در بیمه‌نامه شخص ثالث* [مقاله کنفرانسی]. بیست و نهمین همایش ملی و دهمین همایش بین‌المللی بیمه و توسعه، تهران، ایران. https://civilica.com/doc/1390851

میرباقری جم، م.، شهیکی‌تاش، م. ن، زمانیان، غ.، و صفری، ا. (۱۴۰۱). تجمیع ریسک‌های بیمه‌گری صنعت بیمه ایران با استفاده از توابع مفصل (رویکرد توابع مفصل ارشمیدسی سلسله‌مراتبی). *پژوهشنامه بیمه*، ۴(۴)، ۴۱۲-۴۲۵. https://doi.org/10.22056/ijir.2015.04.02

Abdelhadi, S., ElBahnasy, K. A., & Abdelsalam, M. M. (2020). A proposed model to predict auto insurance claims using machine learning techniques. *Journal of Theoretical and Applied Information Technology, 98*(22), 3428–3437. https://www.jatit.org/volumes/Vol98No22/8Vol98No22.pdf

Abdulkadir, U.I., & Fernando, A. (2024). A deep learning model for insurance claims predictions. *Journal on Artificial Intelligence, 6*(1), 71–83. https://doi.org/10.32604/jai.2024.045332

Balaji, S., & Srivatsa, S. K. (2012). Naïve Bayes classification approach for mining life insurance databases for effective prediction of customer preferences. *International Journal of Computer Applications, 51*(3), 22–28. https://doi.org/10.5120/8023-0805

Banks, D. (2020). Discussion of "Machine learning applications in non-life insurance". *Applied Stochastic Models in Business and Industry, 36*(4), 538–540. https://doi.org/10.1002/asmb.2537

Baran, S., & Rola, P. (2022). *Prediction of motor insurance claims occurrence as an imbalanced machine learning problem* [Preprint]. arXiv, 2204.06109. https://doi.org/10.48550/arXiv.2204.06109

Bärtl, M., & Krummaker, S. (2020). Prediction of claims in export credit finance: A comparison of four machine learning techniques. *Risks, 8*(1), 22. https://doi.org/10.3390/risks8010022

Bücher, A., & Rosenstock, A. (2023). Micro-level prediction of outstanding claim counts based on novel mixture models and neural networks. *European Actuarial Journal, 13*, (May), 55–90. https://doi.org/10.1007/s13385-022-00314-4

Cummings, J., & Hartman, B. (2022). Using machine learning to better model longterm care insurance claims. *North American Actuarial Journal, 26*(3), 470–483. https://doi.org/10.1080/10920277.2021.2022497

Grize, Y. L., Fischer, W., & Lützelschwab, C. (2020). Machine learning applications in non-life insurance. *Applied Stochastic Models in Business and Industry, 36*(4), 545–547. https://doi.org/10.1002/asmb.2564

Gulseven, O. (2020). *Estimating the demand factors and willingness to pay for agricultural insurance* [Preprint]. arXiv, 2004.11279. https://doi.org/10.48550/arXiv.2004.11279

Hanafy, M., & Ming, R. (2021). Machine learning approaches for auto insurance big data. *Risks, 9*(2), 42. https://doi.org/10.3390/risks9020042

Hu, S., & O'Hagan, A. (2020). *Insurance claims forecasting with cluster analysis*. The Actuary, (8th July). https://www.theactuary.com/features/2020/07/08/insurance-claims-forecasting-cluster-analysis

Jing, L., Zhao, W., Sharma, K., & Feng, R. (2018). Research on probability-based learning application on car insurance data. *Proceedings of the 2017 4th International Conference on Machinery, Materials and Computer (MACMC 2017)* (pp. *59–63*) [Conference presentation]. Atlantis Press. https://doi.org/10.2991/macmc-17.2018.14

Kose, I., Gokturk, M., & Kilic, K. (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Applied Soft Computing, 36*, 283-299. https://doi.org/10.1016/j.asoc.2015.07.018

Lentz, T. J., Dotson, G. S., Williams, P. R. D., Maier, A., Gadagbui, B., Pandalai, S. P., Lamba, A., Hearl, F., & Mumtaz, M. (2015). Aggregate exposure and cumulative risk assessment—integrating occupational and non-occupational risk factors. *Journal of Occupational and Environmental Hygiene, 12*(sup1), S112–S126. https://doi.org/10.1080/15459624.2015.1060326

Manavi, F., Dadgar, M., & Rahmatian, M. (2021). *Presenting a method for predicting damage in third-party car insurance policies* [Conference presentation]. The 28th National and 9th International Conference on Insurance and Development, *Tehran, Iran*. https://civilica.com/doc/1390851/ **[In Persian].**

Mirbagheri Jam, M., Shahiki Tash, M. N., Zamanian, G., & Safari, A. (2022). Aggregation of underwriting risks in insurance industry of Iran using copulas (Hierarchical archimedean copulas approach.(*Insurance Research Journal*, *4*(4), 412–425. https://doi.org/10.22056/ijir.2015.04.02 **[In Persian].**

Pabuçcu, H., Ongan, S., & Ongan, A. (2020). Forecasting the movements of Bitcoin prices: An application of machine learning algorithms. *Quantitative Finance and Economics, 4*(4), 679–692. https://doi.org/10.3934/QFE.2020031

Pesantez-Narvaez, J., Guillen, M., & Alcañiz, M. (2019). Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. *Risks, 7*(2), 70. https://doi.org/10.3390/risks7020070

Raeesi Vanani, I., Taghavifard, M., Sohrabi, B., & Amirhosseini, M. (2023). Designing an intelligent evaluation system for predicting fire insurance claims using deep learning. *Iranian Journal of Insurance Research*, *12*(4), 251-264. https://doi.org/10.22056/ijir.2023.04.01 **[In**

Persian].

Reddy, D. B., Mounika, T., Peri, P., Rao, N. T., & Bhattacharyya, D. (2024). *Prediction of automobile insurance claims using deep neural networks* [Conference presentation]. In D. Bhattacharyya & R. Ghosh (Eds.), EAI International Conference on Computational Intelligence and Generative AI (EAI, Springer Innovations in Communication and Computing (pp. 31-41). Springer. https://doi.org/10.1007/978-3-031-76610-7_3

Rostami, M., Bani Hosseini, S. M., & Safaei, S. (2022). *An intelligent algorithm for evaluating and predicting risk and damage in supplementary health insurance using multilayer artificial neural networks* [Conference presentation]. The 29th National and 10th International Conference on Insurance and Development, Tehran, Iran. https://civilica.com/doc/1770257/ **[In Persian].**

Sundarkumar, G. G., & Ravi, V. (2015). A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Engineering Applications of Artificial Intelligence, 37*, 368–377. https://doi.org/10.1016/j.engappai.2014.09.019

Tajaddodi Nodehi, M., Hosseini Khatibani, S., Yazdinejad, M., & Zolfi, S. (2023). Predicting people's health insurance costs using machine learning and ensemble learning methods. *Iranian Journal of Insurance Research*, *13*(1), 1-14. https://doi.org/10.22056/ijir.2024.01.01 **[In Persian].**

Thakur, S. S., & Sing, J. K. (2013). Mining customer data for vehicle insurance prediction system using k-means clustering—An application. *International Journal of Computer Applications in Engineering Sciences, 3*(4), 148. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9ad9d68128e22acfd5f4897b0e1e562099792b2f

Weerasinghe, K. P. M. L., & Wijegunasekara, M. C. (2016). A comparative study of data mining algorithms in the prediction of auto insurance claims. *European International Journal of Science and Technology, 5*(1), 47–54. https://rda.sliit.lk/handle/123456789/2525

Wüthrich, M. V. (2020). Bias regularization in neural network models for general insurance pricing. European Actuarial Journal, *10*(1), 179–202. https://doi.org/10.1007/s13385-019-00215-z

AUTHOR(S) BIOSKETCHES — معرفی نویسندگان

مونا پرستش (Mona Parastesh)، دانشجوی دکتری، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران.

- Email: *St_m_parastesh@azad.ac.ir*
- ORCID: 0009-0008-2851-7702

ضیاءالدین بهشتی‌فرد (Ziaeddin Beheshtifard)، مربی، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

- Email: *Z_beheshti@azad.ac.ir*
- ORCID: 0000-0001-5363-5323

عباس راد (Abbas Raad)، استادیار، گروه مدیریت، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران.

- Email: *A-raad@sbu.ac.ir*
- ORCID: 0000-0002-1413-7126

HOW TO CITE THIS ARTICLE

Parastesh, M., Beheshtifard, Z., & Raad, A. (2025). Predicting damage using sequential deep regression techniques in deep learning methods. *Iranian Journal of Insurance Research, 14*(4), 281-294.

DOI: 10.22056/ijir.2025.04.02

URL: https://ijir.irc.ac.ir/article_160343.html