Predicting damage using sequential deep regression techniques in deep learning methods

F. m.patastsh[1,*], z.beheshti[2], a.raad[3]

1 Department of Computer Engineering, Technical and Engineering Faculty, South Tehran Azad University, Tehran, Iran
2 Department of Computer Engineering, Technical and Engineering Faculty, South Tehran Azad University, Tehran, Iran
3 Department of Management, University of Management and Accounting, Shahid Beheshti University, Tehran, Iran

Abstract

Objective: One of the important challenges of insurance companies is to determine the optimal rate for property insurance policies such as cars and fire. If the insurance company can detect the possibility of damage in an insurance policy, it will have effective and better decisions in determining the insurance rate, the amount of discount allocated to the insurance policy, or the decision regarding the extension of that insurance policy. Insurance companies and experts in the field of issuing and claims are looking for new methods to assess the risk of customers and insurance policies by predicting the occurrence of possible losses in this field. Traditional methods such as general linear models often fail to identify complex patterns in insurance data. 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) networks are used to model time series data.

Methodology: In this article, the data of the last 7 years of the field of fire insurance issuance and loss of Alborz Insurance Company have been used to examine and predict the loss in this field. In this article, focusing on data pre-processing and extracting the best features. In order to provide the best result and after applying different methods of feature extraction, finally, 20 features were selected from a total of 40 selected features and then trained using deep learning. The proposed approach uses a combination of deep learning and sequential models using networks (LSTM) to model time series data.

Findings: In this article, by examining different machine learning methods on the insurance company's issuance and loss data, it was concluded that the ordinal deep regression model performs better than the traditional methods. Improved prediction accuracy, higher reliability and emphasis on the importance of temporal features are key results.

Result: In order to predict the occurrence of damage in fire insurance, the 7-year loss and issuance information of Alborz Insurance Company was used with confidentiality, and the combined method of deep learning and sequential models was used, which has more accuracy and results than the previous methods. It was better.

Keywords: insurance, fire damage, deep learning, machine learning

پیش بینی خسارت با استفاده از تکنیک های رگرسیون عمیق ترتیبی در روش های یادگیری عمیق

مونا پرستش*[1]، ضیاءالدین بهشتی[2]، عباس راد[3]

۱ گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد تهران جنوب، تهران، ایران

۲گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد تهران جنوب، تهران، ایران

۳گروه مدیریت ،دانشگده ی مدیریت و حسابداری دانشگاه شهید بهشتی، تهران ، ایران

چکیده

هدف: یکی از چالش های مهم شرکت های بیمه ، تعیین نرخ بهینه بیمه نامه های اموال مانند اتومبیل و آتش سوزی است. اگر شرکت بیمه بتواند امکان وقوع خسارت را در یک بیمه نامه تشخیص دهد، تصمیمات موثر و بهتری در تعیین نرخ بیمه ، میزان تخفیف اختصاص یافته به بیمه نامه و یا تصمیم در مورد تمدید آن بیمه نامه خواهد داشت . شرکت های بیمه و خبرگان رشته صدور و خسارت ، به دنبال روش های نوین جهت ارزیابی ریسک مشتریان و بیمه نامه ها از طریق پیش بینی وقوع خسارات احتمالی در این رشته هستند. .(Manavi-f & etc ,2020)روش‌های سنتی مانند مدل‌های خطی عمومی (GLMs)[1] اغلب در شناسایی الگوهای پیچیده در داده‌های بیمه ناکام هستند. 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 تکنیک های رگرسیون عمیق ترتیبی[2] ارائه شده است. رویکرد پیشنهادی ترکیبی از یادگیری عمیق و مدل‌های ترتیبی است. از شبکه‌های حافظه طولانی کوتاه‌مدت (LSTM)۳ برای مدل‌سازی داده‌های سری زمانی استفاده می‌شود.

روش شناسی: در این مقاله از داده های ۷ سال اخیر صدور و خسارت بیمه آتش سوزی شرکت بیمه البرز جهت بررسی و پیش بینی خسارت در این رشته استفاده شده است.در این مقاله با تمرکز روی پیش پردازش داده ها و استخراج ویژگی های برتر جهت ارائه بهترین نتیجه و پس از اعمال روش های مختلف استخراج ویژگی، در نهایت از مجموع ٤٠ ویژگی انتخابی ، ۲۰ ویژگی

[1] Generalized Linear Models
[2] Sequential Deep Regression
[3] Long short-term memory

انتخاب و سپس با استفاده یادگیری عمیق آموزش داده شد. رویکرد پیشنهادی، ترکیبی از یادگیری عمیق و مدل‌های ترتیبی با استفاده از شبکه‌های را (LSTM) را برای مدل‌سازی داده‌های سری زمانی استفاده می کند.

یافته ها: در این مقاله و با بررسی روش های مختلف یادگیری ماشینی روی داده های صدور و خسارت شرکت بیمه ، این نتیجه بدست آمد که مدل رگرسیون عمیق ترتیبی نسبت به روش‌های سنتی عملکرد بهتری دارد. بهبود دقت پیش‌بینی، قابلیت اطمینان بالاتر و تاکید بر اهمیت ویژگی‌های زمانی از نتایج کلیدی هستند.

نتیجه: جهت پیش بینی وقوع خسارت در بیمه های آتش سوزی ، از اطلاعات صدور و خسارت ۷ ساله شرکت بیمه البرز با حفظ محرمانگی استفاده شده است. در این تحقیق از ترکیب روش های یادگیری عمیق و مدل های ترتیبی استفاده شده که از روش های قبلی دارای دقت و نتایج بهتری می باشد.

کلمات کلیدی : بیمه ، خسارت آتش سوزی، یادگیری عمیق، یادگیری ماشین

مقدمه

امروزه نقش صنعت بیمه در حال تغییر است. به این دلیل که شرکت‌ها از روش‌های تحلیلی جدید برای پیش بینی خسارت و ریسک استفاده می‌کنند و این روش‌ها به آن‌ها در بررسی ریسک موجود کمک خواهد کرد. در عصر حاضر مدل‌های کسب و کار سنتی و روش‌های قدیمی همواره از جانب تکنولوژی در معرض تهدید بوده اند. شرکت‌های بیمه‌ای جدید از قدرت تکنولوژی‌های نوآورانه استفاده می‌کنند تا رهبران همیشگی بازار بیمه را از بین ببرند. حمایت و حفاظتی که در مقابل ریسک‌ها از بیمه گذار می‌شود و راه حل‌های پیشنهادی که برای مقابله با ریسک ارائه می‌گردد، از طریق سرویس‌هایی که برای شناسایی ریسک های احتمالی طراحی شده بدست می آیند و این سرویس ها میتوانند جهت هشدار خطر (در موارد پر ریسک)بکار گرفته شوند؛ در نتیجه این سرویس‌ها مهم ترین وجه تمایز این شرکت‌ها و کلید موفقیت آن‌ها در آینده خواهد بود. هوش مصنوعی قوی و تحلیل حجم زیادی از اطلاعات داده های حجیم [1]به بیمه‌گران این قدرت را می‌دهد که به سمت پیش بینی خسارت و حوادث حرکت کنند. هر چه میزان اطلاعاتی که شرکت‌های بیمه از بیمه‌گذاران در اختیار دارند بیشتر باشد، آن‌ها قادر خواهند بود از این اطلاعات ارزشمند برای پیش بینی رفتار بیمه‌گذاران استفاده کرده 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 تکنولوژی‌های نوآورانه را به خدمت می گیرند، فرصت رشد بسیار زیادی در اختیار دارند. اما شرکت‌های بیمه‌ای که برای تعیین حق بیمه به سوالات ساده‌ای مثل سن، جنسیت و شغل بسنده می‌کنند، بعید است در عصر دیجیتال و با پیشرفت اینشورتک بتوانند به کار خود ادامه دهند. در حقیقت اگر بیمه‌گران «تحلیل‌های پیش بینانه» 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 کرده اند، مشتریان بیشتری جذب نموده و در طولانی مدت

[1] Big Data

طیف وسیعی از بیمه‌گذاران را دربرگیرند. داده‌های بیمه‌ای اغلب شامل روابط غیرخطی و پیچیده‌ای هستند که مدل‌های رگرسیون ساده مانند رگرسیون خطی ، درخت تصمیم و .. قادر به درک و مدل‌سازی کامل آن‌ها نیستند. همچنین این شرکت ها با حجم وسیع داده مواجه‌اند. در نتیجه به دنبال راهی برای بهبود این روش ها با استفاده از روش های نوین تر مانند یادگیری عمیق هستیم. زیرا شبکه‌های عصبی عمیق می‌توانند روابط پیچیده بین داده های بیمه ای را با دقت بیشتری شناسایی کنند، توانایی پردازش و تحلیل بیشتری روی این داده‌های بزرگ دارند و می‌توانند الگوهای مخفی در آن‌ها را کشف کنند.

مبانی نظری پژوهشی

ارزیابی ریسک بیمه

صنعت بیمه کشور را میتوان به عنوان یک موسسه بیمه بزرگ فرض کرد که در همه رشته فعالیتهای بیمه‌ای مختلف فعالیت دارد. اگر بخواهیم تجمیع ریسکهای بیمه‌گری را برای یک موسسه بیمه معین انجام دهیم ممکن است آن موسسه بیمه در همه رشته فعالیتهای بیمه‌ای فعالیت نداشته باشد؛ بنابراین در این حالت امکان مدلسازی ساختار وابستگی بین ریسکهای بیمه‌گری همه رشته فعالیتهای بیمه‌ای با داده‌های آن موسسه بیمه میسر نخواهد بود. البته مشابه همین تحقیق را میتوان برای هر موسسه بیمه فرضی با داده های مربوطه انجام داد و از نتایج آن بهره برد.

منظور از ریسک بیمه‌گری، ریسک‌هایی است که موسسه بیمه به دلیل صدور بیمه‌نامه یا قبول اتکایی با آن مواجه است[1]. موسسات بیمه با صدور بیمه‌نامه و در قبال دریافت حق بیمه از مشتریان متعهد میشوند که به هنگام وقوع خطرات تحت پوشش بیمه، زیان وارده به مشتریان و بیمه‌گران خود را جبران نمایند و به آنها خسارت پرداخت کنند. این موسسات به منظور اطمینان از توانایی مالی خود در انجام تعهدات و پرداخت خسارت در دوره جاری و دوره‌های آتی، بخشی از درآمد حق بیمه را به صورت ذخایر بیمه‌ای[2] نگهداری میکنند؛ سپس‌از محل این ذخایر، دارائیهای سرمایه گذاری را می‌خرند. عدم کفایت دارائیها و وجوه نقدشرکت در پرداخت خسارت مشتریان، آسیب بزرگی به تداوم فعالیت موسسه بیمه خواهد زد و در ادبیات بیمه به این وضعیت"عدم توانگری مالی"[3] گفته میشود

ارزیابی ریسک برای بیمه نامه ها ، باتوجه به اطلاعات خسارت های قبلی علاوه بر اطلاعات بیمه نامه ، بعنوان اقلام اطلاعاتی مؤثر، روشی کلیدی برای شرکت های بیمه است.مثلاً در بیمهٔ اتومبیل اگر فقط پرداخت خسارت برای یک بیمه نامه را بررسی کنیم، قطعاً برای ارزیابی ریسک نامناسب و کاملاً نادقیق است. درحالیکه اگر این ارزیابی ریسک خسارت بر روی ۱۰۰ هزار بیمه نامهٔ مشابه قبلی انجام شود، می تواند برای ارزیابی و پیش‌بینی با دقت قابل قبولی استفاده شود (Lentz et al., 2015).

---

۱ انجمن بین المللی اکچوئری(IAA)«ریسک بیمه زندگی» و «ریسک بیمه غیرزندگی» را به دو دسته گری تقسیم کرده است.

2 Technical provisions

3 Insolvency

پیش بینی خسارت و روش های یادگیری ماشینی

برای پیش بینی خسارت در صنعت بیمه، روش‌های یادگیری ماشینی مختلفی استفاده می‌شود که هرکدام ویژگی‌ها و مزایا خود را دارند. در اینجا چند نمونه از این روش‌ها را بررسی می‌کنیم. بصورت کلی می توان این روش ها را بصورت زیر تقسیم بندی کرد:

۱. روش‌های خوشه‌بندی:[1]

این روش‌ها برای تشخیص الگوهای مشابه در داده‌ها و گروه‌بندی خطرات مورد استفاده قرار می‌گیرند. 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 بر اساس خطرات مشابه، بیمه‌گذاران می‌توانند بهتر مدیریت ریسک کنند، و از مزایای آن امکان شناسایی گروه‌های خطرات مشابه، کاهش تقلب و مدیریت بهتر ریسک است. جهت انجام خوشه بندی اولین مرحله انتخاب یک الگوریتم خوشه‌بندی، مانند k-means یا.hierarchical clustering است .سپس معمولاً نیاز به تعیین تعداد خوشه‌ها است که بهترین انتخاب بر اساس معیارهایی مانند ضریب سیلوئت ۲ می باشد. (Sen Hu & etc 2020)

۲. روش‌های آماری پیشرفته[3]:

این روش‌ها معمولاً بر پایه تکنیک‌های آماری پیچیده‌تری از جمله رگرسیون خطی یا غیرخطی، تحلیل عاملی، تحلیل خوشه‌ای و غیره می‌باشند. این روش‌ها برای تحلیل دقیق داده‌ها و استخراج اطلاعات کلیدی برای پیش بینی خسارت مورد استفاده قرار می‌گیرند. از جمله این روش ها، روشهای رگرسیون خطی ، غیر خطی و با تحلیل های خوشه ای میباشند. از این روش ها جهت کاهش ابعاد و تحلیل ساختاری و همچنین امکان مدیریت بهتر ریسک‌ها استفاده می شود. این روش‌ها به دلیل توانایی‌شان در تحلیل دقیق داده‌های پیچیده و شناسایی الگوهای پنهان، از اهمیت ویژه‌ای برخوردارند و در صنعت بیمه بهبود قابل توجهی را در فرآیندهای تصمیم‌گیری و مدیریت ریسک به ارمغان آورده‌اند. و در زمینه های مدل‌سازی و پیش‌بینی خسارت‌ها ، تحلیل ریسک‌های مختلف، بهینه‌سازی قیمت‌گذاری بیمه وکاهش تقلب و بهبود مدیریت ریسک‌ها قابل استفاده هستند.

(Jing, L & etc ,2017)

---

1 Clustering Methods

2 Silhouette Coefficient

3 Advanced Statistical Methods

برخی روش های مهم آماری پیشرفته :

- **رگرسیون خطی و غیرخطی**[1] **:** مدل‌سازی رابطه بین خسارت و متغیرهای تأثیرگذار
- **تحلیل عاملی**[2] **:** کاهش ابعاد داده و استخراج متغیرهای کلیدی
- **تحلیل خوشه‌ای**[3] **:**تقسیم‌بندی مشتریان بر اساس ویژگی‌های مشابه

۳. روش‌های مبتنی بر مدل[4]

این روش‌ها شامل الگوریتم‌هایی هستند که بر اساس داده‌های آموزشی، یک مدل ساختاریافته برای پیش‌بینی و تحلیل ریسک‌های بیمه‌ای ایجاد می‌کنند.

- **درخت تصمیم(Decision Trees - DTs)**

درخت تصمیم یک مدل یادگیری ماشینی است که به صورت یک ساختار درختی تصمیم‌گیری را از داده‌های آموزشی یاد می‌گیرد. این درخت از گره‌ها و لبه‌هایی تشکیل شده است که هر گره نماینده یک ویژگی از داده و هر لبه نماینده یک تصمیم یا یک شاخه از گره‌ها است. در پایان هر شاخه (یعنی برگ درخت)، یک پیش‌بینی یا یک کلاس برای داده‌های ورودی قرار می‌گیرد. و شامل مراحل **انتخاب ویژگی**، که درآن بر اساس معیارهایی مانند انتروپی[5]، یا بهره اطلاعاتی[6]، بهترین ویژگی برای جدا کردن داده‌ها از یکدیگر انتخاب می‌شود ،**تقسیم داده**که در آن داده‌ها بر اساس مقدار بهترین ویژگی به دو یا چند زیرمجموعه تقسیم می‌شوند. و مرحله **ساخت درخت**است که این فرآیند به صورت بازگشتی ادامه می‌یابد تا زمانی که یک شرایط پایانی برقرار شود، مانند رسیدن به یک حداقل تعداد نمونه در هر برگ یا عمق معین درخت. این الگوریتم‌ها به خوبی در تحلیل ریسک و تصمیم‌گیری در مورد قیمت‌گذاری بیمه و تخصیص سرمایه مورد استفاده قرار می‌گیرند . از مزایای آن می توان به قابلیت توضیح پذیری بالا، قابلیت ارزیابی دقیق ریسک‌ها، و قابلیت استفاده از ویژگی‌های مختلف داده اشاره کرد(Hanafy, M.& etc 2021)

1 Linear & Nonlinear Regression

2 Factor Analysis

3 Cluster Analysis

4 Model-Based Methods

5 Entropy

6 Information gain

- **ماشین بردار پشتیبان[1](SVMs)**

SVM از **مرزهای تصمیم‌گیری[2] برای تفکیک داده‌ها** استفاده می‌کند و برای **دسته‌بندی بیمه‌گذاران و تحلیل رفتارهای تقلبی** مناسب است

SVMخطی[3]:

زمانی که داده‌ها به‌صورت خطی جداپذیر هستند، SVMیک **خط (در دو بعد)** یا **یک ابرصفحه (در ابعاد بالاتر)** را برای تفکیک کلاس‌ها پیدا می‌کند.

SVM غیرخطی:[4]

وقتی داده‌ها **خطی جداپذیر نیستند،** از **توابع کرنل** [5] استفاده می‌شود تا داده‌ها را به یک فضای **بعد-بالا** [6] نگاشت کند.

کاربردها در بیمه :

- ✓ کشف تقلب در بیمه‌های سلامت و خودرو
- ✓ دسته‌بندی مشتریان بر اساس ریسک
- ✓ پیش‌بینی نرخ خسارت بر اساس داده‌های تاریخی

(Cortes, C & etc 1995)

٤. **شبکه‌های عصبی مصنوعی[7](ANN)**

این روش‌ها مدل‌هایی محاسباتی الهام‌گرفته از عملکرد مغز انسان هستند که توانایی یادگیری روابط پیچیده، تحلیل داده‌های غیرخطی، و کشف الگوهای پنهان را دارند. شبکه‌های عصبی در بیمه برای پیش‌بینی خسارت، کشف تقلب، و بهینه‌سازی فرآیندهای بیمه‌ای استفاده می‌شوند.

انواع مهم شبکه‌های عصبی:

- **شبکه‌های عصبی پیش‌خور [8]:(FNNs)** مدل‌سازی ارتباط غیرخطی بین ویژگی‌های بیمه‌ای و خسارت
- **شبکه‌های عصبی بازگشتی :(RNNs)** مناسب برای تحلیل داده‌های سری زمانی بیمه‌ای

---

[1] Support Vector Machines
[2] Hyperplanes
[3] Linear SVM
[4] Non-Linear SVM
[5] Kernel Functions
[6] Higher- Dimensional Space
7 Artificial Neural Networks

[8] Feedforward Neural Networks

- **شبکه‌های عصبی کانولوشن[1] :(CNNs)** کاربرد در پردازش تصاویر خسارت بیمه‌ای

(Alexandru Grigore & etc 2022)

۵. **شبکه‌های عصبی عمیق:(Deep Neural Networks - DNNs)**

این مدل‌ها نسخه پیشرفته‌تری از شبکه‌های عصبی هستند که شامل لایه‌های متعدد پردازشی برای **استخراج ویژگی‌های پیچیده و یادگیری الگوهای پیچیده‌تر** می‌شوند. این مدل ها **با الهام از ساختار مغز انسان**، از **شبکه‌های عصبی مصنوعی چندلایه** برای یادگیری الگوهای پیچیده در داده‌ها استفاده می‌کند. این روش توانایی بسیار بالایی در **پردازش داده‌های حجیم، شناسایی روابط پنهان و انجام پیش‌بینی‌های دقیق** دارد. به وسیله این شبکه ، یادگیری ویژگی‌ها به صورت خودکار، بدون نیاز به مهندسی ویژگی می تواند انجام شود. شبکه‌های عصبی عمیق (DNNs) از **لایه‌های متعددی از نورون‌ها** تشکیل شده‌اند:

لایه ورودی[2] : داده‌های خام را دریافت می‌کند (مثلاً تصاویر بیمه‌ای، داده‌های مشتریان)
لایه‌های پنهان[3] : شامل چندین نورون که وزن‌های ورودی را پردازش می‌کنند و ویژگی‌های پیچیده را استخراج می‌کنند.
لایه خروجی [4]: نتیجه نهایی را تولید می‌کند (مثلاً پیش‌بینی خسارت بیمه‌ای یا تشخیص تقلب
(He, K & etc ,2022)

این روش‌ها هر کدام با ویژگی‌ها و مزایای خود، در پیش بینی خسارت در صنعت بیمه اهمیت دارند و بسته به نیازهای خاص شرکت‌های بیمه و نوع داده‌های موجود، انتخاب می‌شوند.

پیشینه پژوهش

در سالهای اخیر و مورد توجه قرار گرفتن روش های هوش مصنوعی و یادگیری ماشین ، تمایل به استفاده از الگوریتم های یادگیری ماشین در صنعت بیمه نیز افزایش یافته و در تحقیق های متعددی جهت خوشه بندی مشتریان، پیش بینی خسارت و ارزیابی ریسک، نرخ دهی مناسب و .. از این روش ها استفاده نموده اند. برخی از پژوهش های انجام شده در سالهای اخیر در زیر بررسی می شود:

رییسی وانایی و همکاران (۱۴۰۲) در پژوهشی برای پیش بینی خسارت رشته آتش سوزی از روش های یادگیری ماشین و همچنین در نهایت از یادگیری عمیق استفاده کردند که در نهایت در قالب برنامه ای، ورودی هایی از بیمه نامه را دریافت و پیش بینی احتمال وقوع خسارت را به کاربران بر میگرداند. معنوی و همکاران (۱۴۰۰) در تحقیقی یک روش برای پیش بینی کردن امکان رخداد خسارت در بیمه نامه‌های شخص ثالث اتومبیل ارائه کردند. در روش پیشنهادی از روش های یادگیری ماشین استفاده شده و پس از استخراج ویژگی های موثر برای تشخیص خسارت یا عدم خسارت از بیمه نامه ها ، به کمک روش های یادگیری ماشین و ویژگی

[1] Convolutional Neural Networks
[2] Input Layer
[3] Hidden Layer
[4] Output Layer

های بدست آمده یک مدل پیش بینی امکان وقوع خسارت، آموزش داده می شود.که در نهایت با استفاده از این مدل می توان در زمان فروش بیمه نامه امکان وقوع خسارت و ریسک فروش بیمه نامه را پیش بینی کرد.

تجددی نودهی و دیگران (۱۴۰۲) از یادگیری جمعی برای ترکیب روش‌های رگرسیون متعدد، مانند رگرسیون لجستیک، شبکه‌های عصبی برای پیش بینی هزینه های درمانی بیمه گزاران استفاده کردند .

(Sundarkumar.G & etc 2015)یک روش ترکیبی برای مسائل با داده های غیرمتوازن ارائه کردند، که قادر به کشف خودکار کلاهبرداری ها در شرکت های بیمه می باشد .(Thakur and Sing 2013) مطالعه ای بر روی مشتریان بیمه خودرو انجام داده و از الگوریتم بهبود یافته تری نسبت به روش k-means استفاده کرده اند که با توجه به ویژگی های خاص مشتریان قدرت پیش‌بینی عملکرد مشتریان را افزایش می دهد

. (Balaji and Srivatsa 2012)تکنیک های مورد استفاده برای پیشبینی داده ها برای بیمه عمر را طبقه بندی کرده اند. آنها الگوریتم های مختلفی چون روش دسته بندی نایو بیز و شبکه های بیزین را جهت دسته بندی داد هها مورد ارزیابی قرار داده اند .

(Pesantez-Narvaez& etc ۲۰۱۹) از دو روش رقیب XGBoost ورگرسیون لجستیک، برای پیش بینی فراوانی خسارت های بیمه اتومبیل استفاده کردند. این مطالعه نشان می دهد که مدل XGBoost کمی بهتر از رگرسیون لجستیک است.

(Abdelhadi et al. 2020)در این تحقیق ، چهار طبقه بندی کننده استفاده شده و برای پیش بینی وقوع خسارت استفاده شده است، از جمله الگوریتم‌های XGBoost، J48، ANN و nave Bayes که بهبودی بر روش های طبقه بندی بود و در این تحقیق نیز مدل XGBoost در بین چهار مدل بهترین عملکرد را داشت.

در سال ۲۰۲۱ ،wang در تحقیقی روش‌های مختلف پیش‌بینی را تلفیق کرده و مدلی ترکیبی برای پیش‌بینی دقیق‌تر خسارت‌های بیمه ارائه می‌دهد. (Wang, Y& etc 2021)

(Bücher and Rosenstock 2022) مدل نوآورانه ای در سطح بالاتر و با بهره گیری از شبکهٔ عصبی برای پی شبینی تعداد خسارت ارائه کرد هاند که دارای نتایج مطلوبی بوده و از تکنیک جدیدتر شبکه عصبی برای پیش بینی خسارتهای آینده استفاده کرده است.

برخی دیگر از پژوهش های انجام شده در این زمینه را در جدول زیر مشاهده می کنید :

عنوان مقاله	نویسنده	سال انتشار	روش تحقیق	نتایج و یافته ها
Estimating the demand factors and willingness to pay for agricultural insurance	Gulseven, O(2021)	۲۰۲۳	Logistic Mode	این مقاله با کمک روش های آماری خطی و غیر خطی به تخمین احتمال خردی بیمه کشاورزی از سوی مردم میپردازد
Time Series Analysis for Insurance Loss Prediction	Brown, R. G., & Mathews, M. (2019).	۲۰۱۹	Time Series Analysis	از روش های تحلیل سری زمانی برای ریسک در بیمه استفاده کردند
Machine Learning Approaches for Auto Insurance Big Data	Hanafy, M., & Ming, R. (2021)	۲۰۲۱	رگرسیون لجستیک و XGBoost , Random Forest	ازجنگل تصادفی، برای تحلیل داده‌های بزرگ در بیمه اتومبیل پرداخته‌اند. از روش هایی مانند رگرسیون لجستیک وXGBoost نیز در مقایسه استفاده شده است
Prediction of motor insurance claims occurrence as an imbalanced machine learning problem	Baran,s & Rola.p (2022)	2022	رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، xgBoost، شبکه پیش‌خور	مشکلات ناشی از عدم توازن داده‌ها در پیش‌بینی ادعاهای بیمه خودرو پرداخته است از روش‌های مانندSMOTE استفاده شده است
A deep learning approach to loss reserving	(Kuo, 2019)	2019	loss reserving - deep learning	مدل نگهداری ذخیره برای ریسک براساس شبکهٔ عصبی یادگیری عمیق
Bias Regularization in Neural Network Models for General Insurance Pricing	Mario V. Wuthrich(2019)	2019	Newral network	بررسی اشکال برازش در مدل های شبکه عصبی و راهکاری برای آن
Forecasting the movements of Bitcoin prices: An application of machine learning algorithms	Pabuçcu, H.& etc(2020)	2020	ماشین‌های بردار پشتیبان (SVM)، شبکه عصبی مصنوعی (ANN)، خلیج‌های ساده (NB)	در این مقاله که برای پیش بینی قیمت بیت کویین استفاده شده ،از روش های مختلف رگریسیون ، بردار پشتیبان ، جمگل تصادفی و شبکه عصبی استفاده شده
An interactive machine-learning-based electronic fraud detection system in healthcare insurance	Kose, & etc (2015)	۲۰۱۵	Machin learning metods	در این مقاله برای کشف تقلب از روش های 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, Mathias & Simone Krummaker	2020	شبکه های عصبی احتمالی و روش های یادگیری ماشین	چهار تکنیک 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, K and M. C. Wijegunasekara(2016)	۲۰۱۶	الگوریتم های یادگیری ماشین	در این مقاله به مقایسه روش های مختلف یادگیری ماشین برای پیش بینی خسارت اتومبیل پرداخته شده است
. Discussion of “Machine learning applications in non-life insurance	Banks, David.(2020)	2020	الگوریتم های یادگیری ماشین	مقایسه مدل‌های نگهداری برای بیمه خانگی و سپس یک مشکل قیمت‌گذاری پویا برای بیمه موتور آنلاین ارائه می شود
. Using Machine Learning to Better Model Long-Term Care Insurance Claims. North American	Cummings, Jared, and Brian Hartman	۲۰۲۲	سری های زمانی	بهبود روش های سری زمانی در یادگیری ماشین و پیش بینی خسارت

جدول ۱-برخی پژوهش های انجام شده

Data Set[1] و مدل پیشنهادی

برای نمونه آماری در این پژوهش از داده های صدور و خسارت مربوط به ۷ سال از بیمه نامه های رشته آتش سوزی شرکت بیمه البرز استفاده شده است که اطلاعات سال های ۱۳۹۵ تا ۱۴۰۲ را شامل می شود. در این فواصل زمانی حدود ۳ میلیون بیمه نامه آتش سوزی صادر و حدود ۶۰ هزار پرونده خسارت ثبت گردیده است.

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- نمونه دیتاست تحقیق

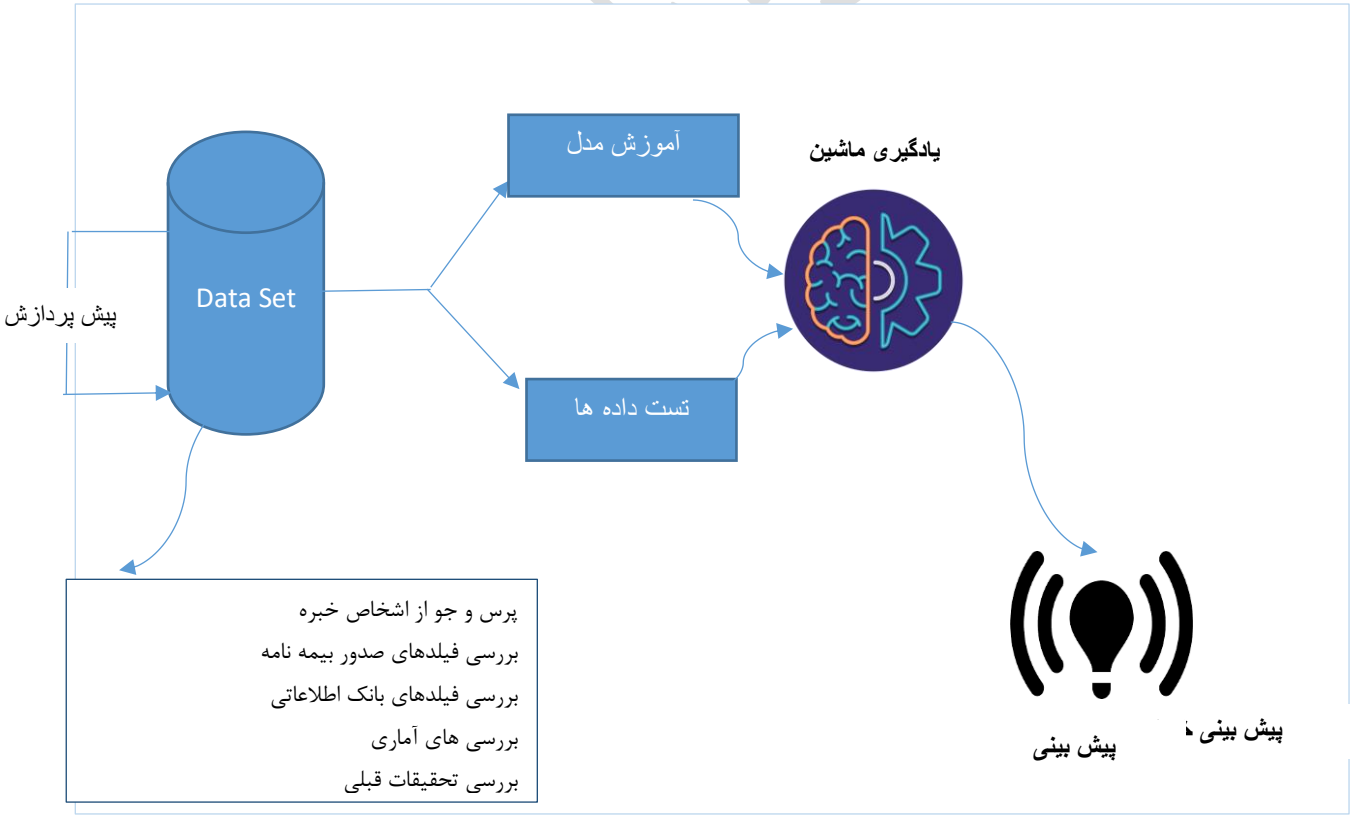
[1] نمونه مورد استفاده در پژوهش

ردیف	نام قلم اطلاعاتی	نام استفاده شده در نمونه آزمایشی	شرح اطلاعات
1	نوع بیمه نامه	PolicyIssueTypeId	عادی / طرح
2	نوع طرح بیمه نامه	PolicyTypeId	طرح جامع خانوار،طرح بیمه اصناف،صنعتی،غیر صنعتی،مسکونی،انبار
3	زیر رشته بیمه نامه	InsLineId	آتش سوزی صنعتی، غیر صنعتی ، انبار ، مسکونی
4	مدت بیمه نامه	duration	طول دوره بیمه نامه به روز
5	شهر صدور بیمه نامه	InsItemCityId	شهر صدور بیمه نامه (که با استان و منطقه خطر طبقه بندی شد)
6	بازدید بیمه نامه	HasInspectionId	بازدید انجام شده یا خیر
7	قرارداد	HasContractId	قرارداد دارد یا خیر
8	حق بیمه دستی دارد؟	HasAnnualPrmId	
9	قبلا خسارت داشته ؟	HasLastPolicyClaimHistoryId	
10	مرهوناتی است؟	IsMortgageId	
11	نوع فعالیت	ActivityTypeId	چوب ، شیشه و ... ، محصولات فلزي ، ماشین آلات وتجهیزات ، و....
12	نوع سازه	ConstructionTypeId	گلی ، فلزی ، چوبی و ....
13	منطقه خطر	DensityRiskTypeId	به ٤ منطقه تقسیم بندی شده است
14	تخفیف چند ساله دارد؟	IsMultiYearDiscountReqId	
15	ذینفع دارد؟	HasBenefId	بیمه نامه دارای ذینفع است یا خیر
16	نوع پوشش	RiskClassId	از ۱ تا ۱۵

شکل ۲- اقلام اطلاعاتی استفاده شده در پژوهش

مراحل اجرای تحقیق

در شکل زیر الگوریتم اجرای تحقیق را مشاهده می کنید که در مرحله انتخاب داده ها شامل پرس و جو در مورد نحوه انتخاب ویژگی ها ،جمع آوری داده ها ، از دیتابیس بیمه ای شرکت بیمه البرز و .. بوده و سپس بصورت یک فرایند تکراری چندین مرحله پیش پردازش داده ها انجام شده و به داده های نسبتا هماهنگ دست پیدا کردیم . برای پیاده سازی این الگوریتم از زبان برنامهنویسی پایتون استفاده شده و خروجی ها در ادامه قابل مشاهده می باشند سپس داده های مهم تر و با تاثیر بیشتر استخراج شده و با تقسیم به دو دسته ی آموزش و تست به مرحله مدل یادگیری ارسال شدند. در این مرحله روی داده های آموزشی ، آموزش اجرا شده و سپس با استفاده از داده های تست، دقت روش مورد بررسی قرار گرفت



شکل ۳- الگوریتم اجرای تحقیق

مراحل انجام تحقیق شامل مراحل زیر می باشد:

- بارگذاری و پردازش داده‌ها
- تقسیم داده‌ها به مجموعه‌های آموزشی و تست
- تعریف مدل یادگیری عمیق ترتیبی
- آموزش مدل و ارزیابی عملکرد
- بارگذاری و پردازش داده‌ها

در ابتدا داده های استخراج شده از دیتابیس بارگزاری شدند و مراحل پیش پردازش روی آنها انجام شد. که شامل جستجو و اصلاح داده های null ، نرمالسازی اطلاعات عددی ، گسسته سازی اطلاعات توصیفی و .. می باشد . چون داده های بیمه نامه ها و تعداد خسارت های وقوع پیدا کرده دارای عدم توازن بودند (تعداد خسارت ها بسیار کمتر از تعداد بیمه نامه بود) به منظور برقراری توازن بین داده ها ، از روش وزن دهی به مقادیر استفاده شد (داده های دارای خسارت با وزن بیشتری وارد دیتاست شدند) و دوباره الگوریتم های مورد نظر اجرا شدند.

از میان ٤٠ ویژگی استخراج شده برای بیمه نامه ها ، ٢٠ ویژگی با استفاده از الگوریتم های انتها ویژگی انتخاب شدند که دارای اثر بیشتری روی نتیجه بودند و در ادامه کار الگوریتم های یادگیری روی این ویژگی ها آموزش داده شدند . در شکل ٣ ویژگی های استخراج شده را در زبان پایتون مشاهده می کنید

```
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 ٤- ویژگی های استخراج شده پس از مرحله استخراج ویژگی

سپس داده های پالایش شده به دو بخش آموزش و تست تقسیم شدند . عملیات یادگیری روی داده های آموزش انجام شد و روی داده های 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 روش های یادگیری عمیق سنتی یکی از رایج‌ترین توابع فعال‌سازی تابع ReLU می باشد که به دلیل سادگی و کارایی آن، به‌طور گسترده‌ای استفاده می‌شود. اما ReLU ممکن است در برخی موارد منجر به مشکل "مرگ نورون‌ها" شود که در آن برخی نورون‌ها هرگز فعال نمی‌شوند. در تحقیق پیش رو ،روش بکار رفته، ترکیبی از یادگیری عمیق و مدل‌های ترتیبی است و از

شبکه‌های حافظه بلندمدت کوتاه‌مدت (LSTM) برای مدل‌سازی داده‌های سری زمانی استفاده می‌شود. در این روش از تابع Swish به جای تابع relu استفاده شده است

تابع swish یک تابع فعال‌سازی جدیدتر که توسط محققان گوگل معرفی شده است. این تابع به دلیل گرادیان‌های نرم‌تر و توانایی مدل‌سازی روابط پیچیده‌تر در داده‌ها، عملکرد بهتری نسبت به ReLU نشان داده است Swish .معادله Swish(x)=x·sigmoid(x) را دارد که باعث می‌شود به‌طور پیوسته فعال باشد و بهبود کارایی مدل را به همراه داشته باشد. (Ramachandran, P & etc, 2019)

مزایای استفاده از تابع Swish

Swish مقادیر منفی را به طور کامل حذف نمی‌کند، بلکه به آن‌ها اجازه می‌دهد تا به مقدار کمی از جریان اطلاعات کمک کنند. این می‌تواند منجر به یادگیری بهتر و استفاده موثرتر از تمام داده‌ها شود.

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

پدیده "مرده شدن" نورون‌ها :برخلاف ReLU که نورون‌ها ممکن است "مرده" شوند و برای همیشه غیرفعال باقی بمانند، Swishاین مشکل را ندارد و نورون‌ها همچنان فعال باقی می‌مانند.

دقت بیشتر : به دلیل 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 منجر به بهبود عملکرد کلی مدل‌های یادگیری عمیق می‌شود و می‌تواند به دست‌یابی به نتایج دقیق‌تر و کارآمدتر کمک کند(Ramachandran, P & etc, 2019 .)

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

۲- نتایج و یافته ها

در بررسی نتایج و یافته ها در روش های یادگیری ماشین عموماً پارامترهای زیر مورد ارزیابی قرار می گیرند 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.

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

که در آن:

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

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

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

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

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 عدم‌توازن در داده‌ها هستند، مفید است

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

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

نتایج زیر بر اساس آموزش مدل با یادگیری عمیق و با دوروش استفاده از تابع فعالسازی Relu و تابع Swish بدست آمده است :

```
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 relu

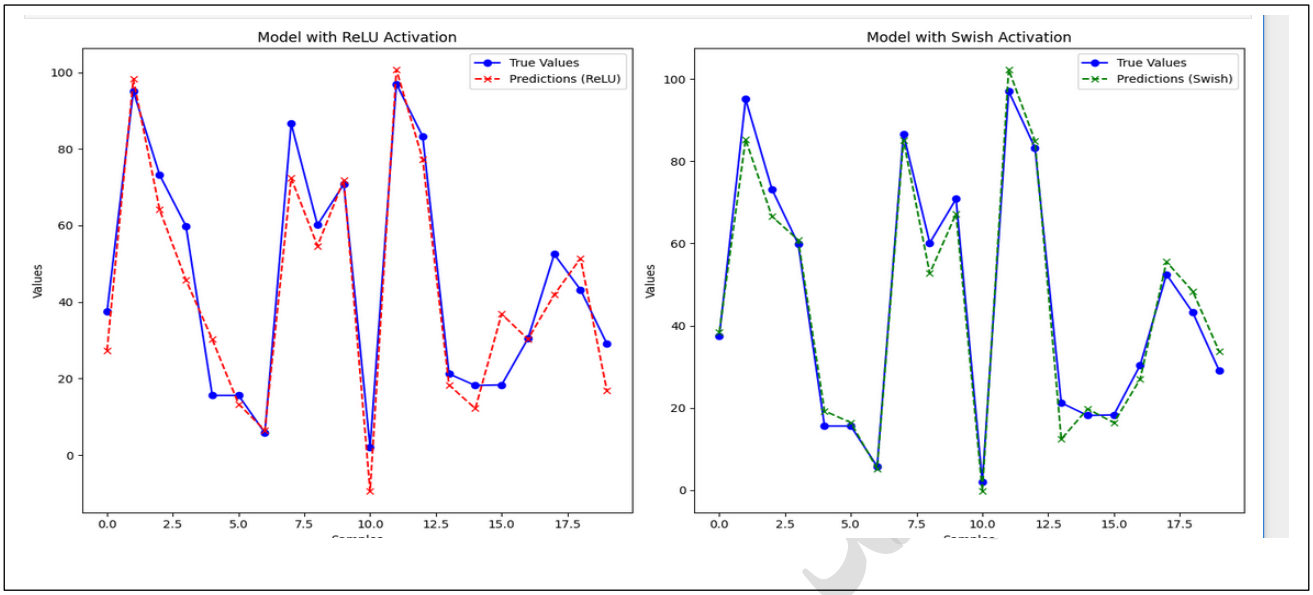
```
Epoch 1/10, Loss: 0.1105, Test Loss: 0.3201, Test Accuracy: 0.8562
Epoch 2/10, Loss: 0.0902, Test Loss: 0.2999, Test Accuracy: 0.8660
Epoch 3/10, Loss: 0.0754, Test Loss: 0.2850, Test Accuracy: 0.8715
Epoch 4/10, Loss: 0.0650, Test Loss: 0.2802, Test Accuracy: 0.8738
Epoch 5/10, Loss: 0.0575, Test 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 ٦- اجرای الگوریتم یادگیری عمیق با تابع Swich

با توجه به مقایسه‌ی این دو تابع فعال‌سازی در شبکه‌های عمیق، استفاده از Swish به جای ReLU بهبودهای قابل توجهی در عملکرد و دقت مدل‌های عمیق می‌آورد. این بهبودها عموماً به دلیل بهینه‌تر بودن در استفاده از اطلاعات ورودی، کاهش مشکلات گرادیان‌ها و افزایش پایداری شبکه‌های عصبی است. در شکل زیر مقایسه این دو روش قابل مشاهده می باشد

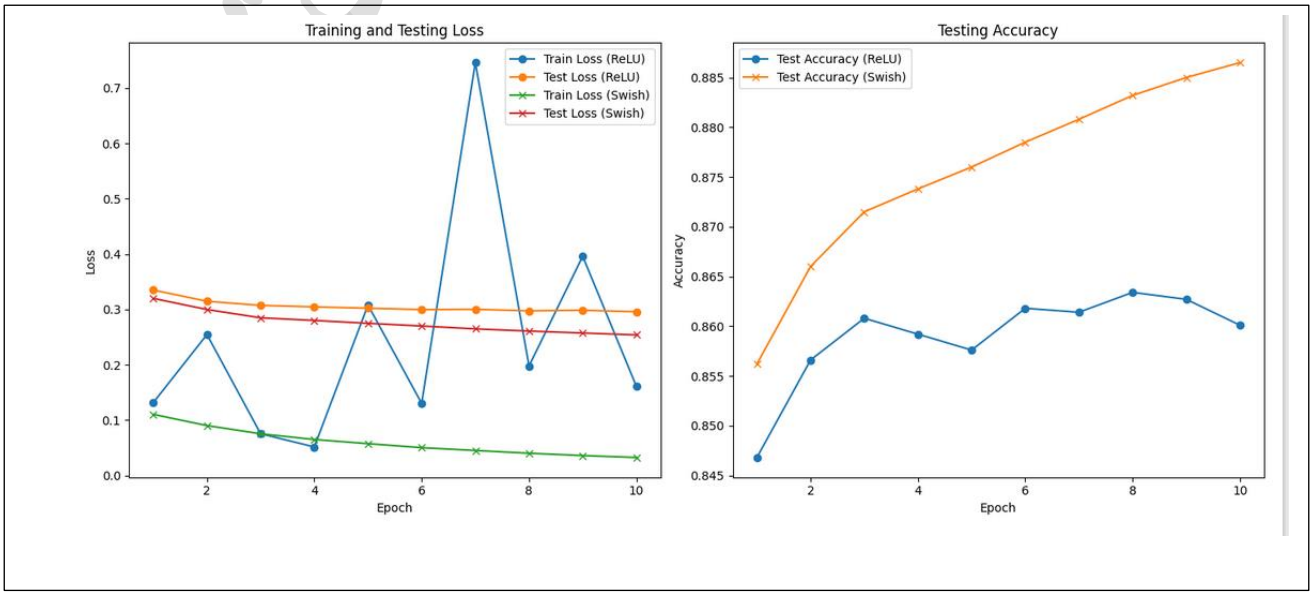
Metric	Model with ReLU	Model with Swish (فرضی)
Accuracy	0.861	0.88
Precision	0.85	0.87
Recall	0.84	0.86
F1 Score	0.845	0.865

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



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

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



شکل ۹ مقایسه عملکرد الگوریتم یادگیری عمیق با تابع relue و swish

این نمودارها نشان می‌دهند که مدل جدید با استفاده از Swish دقت بیشتری در پیش‌بینی‌ها دارد، زیرا پیش‌بینی‌ها به مقادیر واقعی نزدیک‌تر هستند و نوسانات کمتری دارند. این امر به دلیل عملکرد بهتر تابع Swish در مقایسه با ReLU است که منجر به هموارتر شدن گرادیان‌ها و بهبود دقت پیش‌بینی‌ها می‌شود. این موضوع نشان‌دهنده عملکرد بهتر مدل جدید است که از تابع فعال‌سازی Swish استفاده می‌کند.

در ادامه یک نمونه پیاده سازی جهت استفاده از این روش در صنعت بیمه برای پیش بینی خسارت پیشنهاد می گردد. در این نمونه یک اپلیکیشن ساده با استفاده از زبان پایتون پیاده سازی شده و شامل ورودی های استفاده شده در تحقیق می باشد که پس از ورود اطلاعات و فشردن دکمه ی پیش بینی، الگوریتم مورد نظر بکار گرفته شده و پیش بینی خسارت را بر اساس مدل یادگیری عمیق پیشنهاد شده ارائه می دهد .

شکل ۱۰ . نمونه اپلیکیشن خروجی برای بکارگیری روش پیشنهادی

نتیجه گیری

مدل ارائه شده یک مدل برای ارزیابی خسارت آتش سوزی است که بااستفاده از روش های یادگیری عمیق آموزش دیده و پیش بینی وقوع خسارت را انجام میدهد. این مدل در یادگیری عمیق بجای تابع Relu از تابع swish استفاده کرده و نتایج بدست آمده نشان میدهد استفاده از این روش ،دارای خروجی بهتری است ودقت نتایج را بطرز قابل ملاحظه ای افزایش داده است .برای آموزش مدل از داده های صدور و خسارت آتش سوزی ۷ سال اخیر بیمه البرز استفاده شده و از زبان برنامه نویسی پایتون برای آموزش و پیش بینی مدل استفاده شده است . بعد از آموزش مدل با استفاده از یادگیری عمیق به روش سنتی دقت حدودی ۰,۸٦ برآورد شد . پس ازبهره گرفتن از تکنیک های رگرسیون عمیق ترتیبی در یادگیری عمیق و پس از اجرای آموزش با استفاده از تابع swish دقت به ۰,۸۸ افزایش پیدا کرده است . این بهبود به دلیل ویژگی‌های تابع نرم‌تر و گرادیان‌های غیرصفر برای مقادیر منفی، به دست آمده است. نتایج نشان داد که استفاده از این روش در صنعت بیمه می‌تواند موجب کاهش میزان خطای تخمین خسارت و افزایش دقت محاسبه میزان حق بیمه شود

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

پیشنهادات آینده

نتایج این تحقیق 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 ادامه تحقیقات در این حوزه، روش‌های دقیق‌تر و کاربردی‌تری برای تحلیل ریسک‌های بیمه‌ای ارائه شوند.در سال ۲۰۲٤ تحقیقی توسط Rahman, Jamshaid و همکارانش انجام شد که از ترکیب مزایای دو تابع relue و Switch استفاده شده بود. در ادامه تحقیق می توان از این روش برای بهبود عملکرد شبکه های عصبی بهره برد. (Rahman, J & etc ,2024)

در ادامه برخی پیشنهادات برای ادامه پژوهش در این زمینه ارائه می گردد:

۱. **بهبود مدل‌های یادگیری عمیق :**پیشنهاد می‌شود تحقیقات آینده بر روی 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 تحقیق، داده‌های مورد استفاده محدود به بیمه‌های آتش‌سوزی بودند. پیشنهاد می‌شود مطالعات آینده داده‌های سایر حوزه‌های بیمه مانند بیمه عمر و بیمه 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 تحقیق به عوامل کلان اقتصادی می‌تواند به بهبود دقت مدل‌ها کمک کند.

٤. **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. **توسعه سامانه‌های هوشمند بیمه‌ای :**پیشنهاد می‌شود که نتایج این تحقیق در یک سامانه تصمیم‌یار پیاده‌سازی شود تا شرکت‌های بیمه بتوانند از آن برای ارزیابی سریع و دقیق خسارت‌ها استفاده کنند. توسعه یک نرم‌افزار مبتنی بر هوش مصنوعی می‌تواند به شرکت‌های بیمه در پردازش درخواست‌های خسارت کمک کند و فرآیندهای اداری را تسهیل نماید.

منابع

1. Insurance, Supreme Council. (Year). Bylaw No. 69 - Bylaw on how to calculate and monitor the financial viability of insurance institutions. Tehran: Central Insurance of the Islamic Republic of Iran. (In Persian)
2. Mirbagheri, J.; Shahiki Tash, M.N.; Zamanian, G. (2015). Aggregation of underwriting risks in insurance industry of Iran using vine copula. Risk Governance and Control: Financial Markets & Institutions, 5(4), 10-19. (In Persian)
3. 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 (In Persian)
4. Manavi;M,Dadgar. M.;Rahmatian.M (2020). Presenting a method for predicting damage in third-party car insurance policies. 28th Insurance and Development Conference. (In Persian)
5. Tajaddodi.M; Hoseini khatibani.S;Yazdinejad.M;Zolfi.S (2020). Predicting the health insurance costs of individuals using machine learning and collective learning methods. Insur. Dev. J., 13(1): 1-14. (In Persian)
6. Juhari, N.; Jafari.D ;Fattahi.A(2015). Forecasting customer losses in the insurance industry using customer lifetime value and data mining techniques. Int. Conf. Management Elites. (In Persian)

7. Manavi-F. (2020). Providing a method for predicting damage in third-party car insurance policies. 28th Insurance and Development Conference. (In Persian)
8. Lentz, J.; Dotson, G.S.; Williams, P.R.; 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. J. Occup. Environ. Hyg.
9. Sundarkumar, G.G.; Ravi, V. (2015). A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. Eng. Appl. Arti. Intell.
10. Bücher, A.; Rosenstock, A. (2022). Micro-level prediction of outstanding claim counts based on novel mixture models and neural networks. Eur. Actuarial J.
11. Thakur, S.S.; Sing, J.K. (2013). Mining customer data for vehicle insurance prediction system using k-means clustering. Int. J. Comput. Appl. Eng. Sci.
12. Hu, S.; O'Hagan, A. (2020). Insurance claims forecasting with cluster analysis. Inst. Faculty Actuaries, July 2020.
13. Ramachandran, P., Zoph, B., & Le, Q. V. (2019). Swish: a Self-Gated Activation Function. arXiv preprint arXiv:1710.05941
14. Rahman, J. U., Zulfiqar, R., Khan, A., & Nimra. (2024) SwishReLU: A Unified Approach to Activation Functions for Enhanced Deep Neural Networks Performance." *arXiv preprint arXiv:2407.08232* (2024)
15. He, K., Zhang, X., Ren, S., & Sun, J. (2022). Deep learning for insurance claim prediction: A comparative study. *Springer Journal of Artificial Intelligence*, **45**(2), 98-112
16. Pesantez-Narvaez, J.; Guillen, M.; Alcañiz, M. (2019). Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. Risks.
17. Abdelhadi, S.; ElBahnasy, K.A.; Abdelsalam, M.M. (2020). A proposed model to predict auto insurance claims using machine learning techniques. J. Theor. Appl. Inf. Technol.
18. Balaji, S.; Srivatsa, S.K. (2012). Naïve Bayes classification approach for mining life insurance databases for effective prediction of customer preferences. Int. J. Comput. Appl.
19. Jing, L.; Zhao, W.; Sharma, K.; Feng, R. (2017). Research on probability-based learning application on car insurance data. Proc. 4th Int. Conf. Machinery, Materials and Computer (MACMC 2017).
20. Grigore, A. (2020). Machine learning applications in insurance. J. Comput. Appl. Math.
21. Hanafy, M.; Ming, R. (2021). Machine learning approaches for auto insurance big data. Risks.
22. Wang, Y.; Zhang, X. (2021). Combining prediction models for car insurance loss estimation. Expert Syst. Appl., 168: 114191.
23. Gulseven, O. (2020). Estimating the demand factors and willingness to pay for agricultural insurance. arXiv, arXiv:2004.11279.
24. Brown, R.G.; Mathews, M. (2019). Time series analysis for insurance loss prediction. J. Stat. Plan. Infer.
25. Hanafy, M.; Ming, R. (2021). Machine learning approaches for auto insurance big data. Insur. Risks.
26. Baran, S.; Rola, P. (2022). Prediction of motor insurance claims occurrence as an imbalanced machine learning problem. arXiv, arXiv:2204.06109.
27. Wuthrich, M.V. (2019). Bias regularization in neural network models for general insurance pricing. RiskLab, ETH Zurich.
28. Pabuçcu, H.; Ongan, S.; Ongan, A. (2020). Forecasting the movements of Bitcoin prices: An application of machine learning algorithms. Quant. Financ. Econ.
29. Kose, I.; Gokturk, M.; Kilic, K. (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. Appl. Soft Comput.
30. Weerasinghe, K.; Wijegunasekara, M.C. (2016). A comparative study of data mining algorithms in the prediction of auto insurance claims. Eur. Int. J. Sci. Technol.
31. Bärtl, M.; Krummaker, S. (2020). Prediction of claims in export credit finance: A comparison of four machine learning techniques. Risks. [DOI: 10.3390/risks8010022]
32. Banks, D. (2020). Discussion of "Machine learning applications in non-life insurance". Appl. Stoch. Models Bus.