Iranian Journal of Insurance Research
(IJIR)

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

Iranian Insurance Research Center (IIRC)

ORIGINAL RESEARCH PAPER

Life Insurance Premium Pricing Through Personalized Mortality Table Using the Wang Premium Principle

Amir Mohammad Norouzi[1], Mojtaba Ranjbar[2,*], Saman Vahabi[3], Mitra Ghanbarzadeh[4]

[1] *Master Degree, Department of Financial Mathematics and Actuarial, Faculty of Mathematical Sciences, Kharazmi University, Tehran, Iran.*

[2] *Associate Professor, Department of Financial Mathematics and Actuarial, Faculty of Mathematical Sciences, Kharazmi University, Tehran, Iran.*

[3] *Assistant Professor, Department of Financial Mathematics and Actuarial, Faculty of Mathematical Sciences, Kharazmi University, Tehran, Iran.*

[4] *Associate professor, Department of Personal Insurance, Insurance Research Center, Tehran, Iran.*

ARTICLE INFO

Article History:
Received: 28 August 2025
Final revision: 20 November 2025
Accepted: 13 January 2026
Early online access: 13 January 2026
Published: 1 July 2026

Keywords:
Fair pricing
Life insurance
Machine learning
Market price of risk
Personalized mortality table
Wang premium

*Corresponding Author:
Email: *Mranjbar@khu.ac.ir*
Phone: +98 2188942323
ORCID: 0000-0003-0491-526X

DOI: 10.22056/ijir.2026.03.01

ABSTRACT

BACKGROUND AND OBJECTIVES: The conventional approach to life insurance premium pricing relies on standardized mortality tables, which apply uniform survival probabilities to diverse policyholder populations. While this method is computationally straightforward, it overlooks the heterogeneity of individual risk profiles, often resulting in premiums that fail to reflect the true mortality risk of each insured individual. This may lead to pricing inefficiencies, a sense of unfairness among low-risk policyholders and unintended cross-subsidization between high- and low-risk groups where low-risk individuals may overpay to offset losses from higher-risk counterparts. To address these challenges, this study develops a novel framework for constructing personalized mortality tables tailor-made to the unique risk characteristics of each policyholder. By integrating these individualized tables with the Wang premium principle, enhanced by advanced machine learning techniques, the research aims to deliver fairer, more precise, and actuarially sound premium calculations. This approach not only responds to the evolving expectations of modern insurance markets but also supports long-term profitability by retaining low-risk clients through transparent and equitable pricing practices.

METHODS: This study adopts a developmental–applied approach with an analytical design to create risk-adjusted mortality tables reflecting individual policyholder characteristics. Central to the methodology is the development of a composite market risk parameter (λ), which consolidates key risk factors—including age, gender, smoking status, occupation, medical history, contract duration, and death benefit—into a single predictive index. This parameter serves as the cornerstone for the Wang distortion function, a robust mathematical tool that transforms classical survival probabilities into risk-adjusted probabilities, incorporating risk aversion and market dynamics. The estimation of λ leverages historical policyholder data, capturing a comprehensive set of demographic, health, and occupational variables. Two state-of-the-art machine learning algorithms—Random Forest (RF) and Extreme Gradient Boosting (XGBoost)—are deployed to predict λ with high accuracy, capitalizing on their ability to model complex, non-linear relationships among risk factors and handle noisy or heterogeneous data effectively. All data preprocessing, model training, and hyperparameter optimization are conducted using Python, ensuring computational efficiency, scalability, and reproducibility. The predicted λ values are then integrated into the Wang premium principle to compute individualized premium rates, ensuring that pricing reflects the unique risk profile of each policyholder and aligns with actuarial expectations.

FINDINGS: Empirical results underscore the superiority of the proposed framework in delivering equitable and precise premium estimates. By aggregating risk factors into the market risk parameter, the model generates personalized survival probabilities that replace the uniform probabilities of classical mortality tables, offering a more granular and accurate representation of mortality risk. For high-risk policyholders with elevated λ values, premiums calculated using the Wang principle are significantly higher than those derived from traditional methods, ensuring that pricing accurately reflects increased mortality risks. Conversely, low-risk individuals benefit from lower premiums, mitigating the issue of overpricing that often affects this group in conventional frameworks. Feature importance analysis, conducted through the machine learning models, reveals that health-related factors such as medical history and smoking status alongside age, exert the strongest influence on predicting λ, driving significant variability in premium rates. Occupation and gender, while less dominant, contribute meaningfully to risk assessment, highlighting the multidimensional nature of mortality risk modeling. The Random Forest model achieves a mean squared error (MSE) of 0.00028, indicating high predictive accuracy and robustness. These findings demonstrate the framework's ability to eliminate hidden subsidies between risk groups, ensuring actuarially fair pricing that aligns with individual risk profiles and market realities.

CONCLUSION: The principal contribution of this research lies in its development of a hybrid framework that seamlessly integrates personalized mortality tables with the Wang premium principle, enhanced by machine learning-driven predictions. By aligning premiums with individual risk profiles, the model ensures actuarial fairness, reduces inefficiencies, and enhances pricing transparency, fostering greater policyholder trust and reducing lapse rates in competitive markets. The findings confirm a consistent pattern: Wang-based premiums exceed classical premiums for high-risk individuals and are lower for low-risk groups, ensuring equitable differentiation across risk profiles. From a practical perspective, insurers adopting this framework can strengthen customer relationships, improve retention of low-risk policyholders, and enhance market competitiveness. Future research could extend this framework to other insurance branches, such as health or property insurance, where personalized risk assessment could yield similar benefits. Incorporating additional data sources, such as behavioral or lifestyle data from wearable devices or financial records, could further refine risk predictions. Exploring hybrid models that combine deep learning with traditional actuarial methods holds promise for enhancing predictive performance, particularly for complex or sparse datasets. A critical next step is validating the model with real-world insurance data, which would provide a robust basis for assessing its practical effectiveness and scalability in diverse market conditions. By modernizing life insurance pricing, this research paves the way for a more equitable, transparent, and customer-centric insurance industry, responding to the growing demand for fairness and precision in pricing practices.

نشریه علمی پژوهشنامه بیمه

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

مقاله پژوهشی

قیمت‌گذاری بیمهٔ عمر از طریق شخصی‌سازی جدول مرگ‌ومیر با استفاده از اصل حق‌بیمهٔ وانگ

امیرمحمد نوروزی[۱]، مجتبی رنجبر[۲،*]، سامان وهابی[۳]، میترا قنبرزاده[۴]

[۱] کارشناسی ارشد، گروه ریاضیات مالی و بیمسنجی، دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران.

[۲] دانشیار، گروه ریاضیات مالی و بیمسنجی، دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران.

[۳] استادیار، گروه ریاضیات مالی و بیمسنجی، دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران.

[۴] دانشیار، گروه پژوهشی بیمه‌های اشخاص، پژوهشکده بیمه، تهران، ایران.

اطلاعات مقاله

تاریخ های مقاله:

دریافت: ۶ شهریور ۱۴۰۴

بازنگری نهایی: ۲۹ آبان ۱۴۰۴

پذیرش: ۲۳ دی ۱۴۰۴

زودآیند: ۲۳ دی ۱۴۰۴

انتشار: ۱۰ تیر ۱۴۰۵

کلمات کلیدی:

بیمهٔ عمر

پارامتر بازاری ریسک

جدول مرگ‌ومیر شخصی‌سازی‌شده

حق‌بیمه وانگ

قیمت‌گذاری منصفانه

یادگیری ماشین

***نویسنده مسئول:**

ایمیل: *Mranjbar@khu.ac.ir*

تلفن: ۲۱۸۸۹۴۲۳۲۳ ۹۸+

ORCID: 0000-0003-0491-526X

چکیده:

پیشینه و اهداف: در چهارچوب کلاسیک اصول محاسبهٔ حق‌بیمه، بیمه‌گذاران با سطوح متفاوتی از ریسک معمولاً براساس جدول مرگ‌ومیر یکسان ارزیابی می‌شوند. هرچند این رویکرد از نظر محاسباتی ساده است، اما ناهمگونی عوامل ریسک فردی را نادیده می‌گیرد و در نتیجه، نرخ‌های حق‌بیمهٔ به‌دست‌آمده الزاماً بازتاب‌دهندهٔ دقیق احتمال واقعی فوت هر فرد نیستند. این ناهماهنگی می‌تواند به ناکارآمدی در قیمت‌گذاری و ایجاد احساس بی‌عدالتی در میان بیمه‌گذاران کم‌ریسک منجر شود. این پژوهش با هدف رفع این کاستی‌ها، به توسعهٔ جداول مرگ‌ومیر شخصی‌سازی‌شده براساس محاسبهٔ مجزای احتمالات مرگ‌ومیر برای بیمه‌گذاران با سطوح متفاوت ریسک پرداخته است. هدف اصلی، طراحی جداول مرگ‌ومیر تعدیل‌شده با ریسک است که ویژگی‌های منحصربه‌فرد هر بیمه‌گذار را منعکس کند و در قالب چهارچوب محاسبهٔ حق‌بیمهٔ منصفانه و دقیق به کار گرفته شود.

روش‌شناسی: پژوهش پیش رو ماهیتی توسعه‌ای– کاربردی با رویکرد تحلیلی دارد. با توجه به اینکه هر بیمه‌گذار عمر ممکن است با مجموعه‌ای از عوامل ریسکی مواجه باشد، تمامی این عوامل در قالب یک شاخص ترکیبی، با عنوان پارامتر بازاری ریسک تجمیع شده‌اند. این پارامتر اثر ترکیبی تمامی متغیرهای ریسکی را دربرمی‌گیرد و به‌مثابهٔ ورودی اصلی تابع انحراف وانگ عمل می‌کند. این تابع، احتمالات بقای کلاسیک را به احتمالات بقای تعدیل‌شده با ریسک تبدیل می‌کند. فرایند برآورد پارامتر بازاری ریسک با استفاده از داده‌های تاریخی بیمه‌گذاران آغاز می‌شود که شامل متغیرهایی چون سن، جنسیت، استعمال دخانیات، شغل و سابقهٔ بیماری است. پیش‌بینی این پارامتر از طریق دو الگوریتم یادگیری ماشین جنگل تصادفی و تقویت گرادیانی شدید انجام می‌گیرد. تمامی تحلیل‌ها در پایتون انجام شده است. در نهایت، مقادیر پیش‌بینی‌شده پارامتر بازاری ریسک در فرمول حق‌بیمهٔ وانگ جای‌گذاری و نرخ‌های حق‌بیمه شخصی‌سازی‌شده محاسبه می‌شوند.

یافته‌ها: با تجمیع عوامل ریسکی مطرح‌شده در قالب پارامتر بازاری ریسک، مجموعه‌ای از احتمالات بقای جدید برای هر فرد محاسبه شده که جایگزین احتمالات یکسان حاصل از جدول مرگ‌ومیر کلاسیک شد. نتایج تجربی نشان داد که ترکیب مدل‌های یادگیری ماشین با تابع انحراف وانگ، نسبت به روش سنتی، ارزیابی منصفانه‌تری از حق‌بیمه ارائه می‌دهد. برای بیمه‌گذاران پرریسک با مقادیر بالای پارامتر بازاری ریسک، حق‌بیمهٔ وانگ به‌طور معناداری بیشتر از حق‌بیمهٔ کلاسیک محاسبه شد و بالعکس. همچنین، براساس تحلیل اهمیت ویژگی‌ها، نتایج نشان داد که متغیرهای سلامت (سابقهٔ بیماری و استعمال دخانیات) و سن بیشترین تأثیر را بر پیش‌بینی پارامتر ریسک دارند.

نتیجه‌گیری: مهم‌ترین دستاورد این پژوهش، ارائهٔ چهارچوب محاسبهٔ حق‌بیمه براساس جداول مرگ‌ومیر شخصی‌سازی‌شده و اصل حق‌بیمهٔ وانگ است که امکان قیمت‌گذاری دقیق و منصفانه را فراهم می‌آورد. با تجمیع عوامل متنوع ریسک در قالب یک پارامتر بازاری و پیش‌بینی آن با استفاده از الگوریتم‌های پیشرفتهٔ یادگیری ماشین، این روش از دقت، انعطاف‌پذیری و عدالت بالایی برخوردار است.این رویکرد در مقایسه با اصول کلاسیک محاسبهٔ حق‌بیمه، واکنش‌پذیری بیشتری در مقابل تغییرات پروفایل ریسک فردی دارد و نرخ‌های حق‌بیمه را به واقعیت آماری ریسک نزدیک‌تر می‌سازد. پیامد عملی مهم این روش برای شرکت‌های بیمه، افزایش شفافیت در قیمت‌گذاری، بهبود اعتماد بیمه‌گذاران و کاهش نرخ فسخ بیمه‌نامه‌هاست. برای تحقیقات آتی، می‌توان این چهارچوب را به دیگر رشته‌های بیمه‌ای گسترش داد، داده‌های رفتاری و سبک زندگی را به مدل افزود.

DOI: 10.22056/ijir.2026.03.01

مقدمه

با توجه به تغییرات پیچیده و فراوری روزافزون در بازار بیمه، بهره‌گیری از روش‌های دقیق، کارآمد و منصفانه برای تعیین نرخ‌های حق‌بیمه به ضرورتی انکارناپذیر تبدیل شده است. مدل‌های سنتی محاسبهٔ حق‌بیمه، اگرچه در آغاز به‌دلیل سادگی استفاده و بازدهی نسبی مفید بودند، اما در مواجهه با پیچیدگی‌های رفتارهای ریسک‌پذیری و تنوع داده‌های بیمه‌گذاران امروزی قادر به تأمین نیازهای بازار نیستند. منظور از مدل‌های سنتی محاسبهٔ حق‌بیمه همان روش رایج در شرکت‌های بیمه است که طی آن ابتدا حق‌بیمهٔ خالص با استفاده از جدول مرگ‌ومیر ایران محاسبه می‌شود و سپس در صورت وجود عوامل پرریسک (مانند استعمال دخانیات یا بیماری‌های خاص)، ضریب یا مبلغی به‌عنوان بار اضافی به حق‌بیمهٔ خالص افزوده می‌شود و در نهایت، حق‌بیمهٔ نهایی به بیمه‌گذار اعلام می‌شود.

در این راستا، اصل حق‌بیمهٔ وانگ به‌عنوان رویکردی نوآورانه مطرح شده است که با اعمال تحریف بر توزیع زیان و بازتعریف انتظارات، امکان قیمت‌گذاری دقیق‌تر و مطابق با سطح واقعی ریسک را فراهم می‌سازد. نوآوری اصلی این پژوهش در ارائهٔ چهارچوبی ترکیبی است که شامل سه مرحله است: (۱) تعریف پارامتر بازاری ریسک (λ) به‌عنوان شاخصی ترکیبی از ویژگی‌های فردی؛ (۲) پیش‌بینی این پارامتر با استفاده از الگوریتم‌های پیشرفتهٔ یادگیری ماشین (جنگل تصادفی و تقویت گرادیانی شدید)؛ و (۳) به‌کارگیری آن در تابع انحراف وانگ برای ایجاد جداول مرگ‌ومیر شخصی‌سازی‌شده و محاسبهٔ حق‌بیمه‌های عادلانه‌تر. این چهارچوب ترکیبی، که مدل‌های محاسباتی کلاسیک را با هوش مصنوعی ادغام می‌کند، در ادبیات پژوهشی موجود کمتر به‌صورت یکپارچه بررسی شده است. در پژوهش وانگ (Wang, 2000)، پارامتر λ، شدت تمایل بازار به ریسک یا اجتناب از آن را نشان می‌دهد. در پژوهش حاضر، این پارامتر میزان ریسک‌گریزی بیمه‌گر را نشان می‌دهد. بیمه‌گذاران دارای ریسک‌هایی هستند که شرکت بیمه‌گر تمامی ریسک‌ها را در قالب یک پارامتر بیان می‌کند. $\lambda = ۰$ بیانگر بی‌تفاوتی به ریسک و هرچه λ بیشتر شود ریسک‌گریزی شرکت بیمه بیشتر و به‌تبع آن حق‌بیمه نیز بالاتر خواهد شد.

ساختار مقاله به این صورت است: در بخش دوم مبانی نظری پژوهش و اصول حق‌بیمهٔ وانگ و انواع توابع انحراف معرفی می‌شود. در بخش سوم پیشینهٔ مطالعاتی مرتبط مرور شده است. بخش چهارم به روش‌شناسی پژوهش اختصاص دارد که در آن الگوریتم‌های یادگیری ماشین مورد استفاده برای محاسبهٔ نرخ ریسک افراد تشریح شده است. در بخش پنجم، نتایج حاصل از مدل‌ها ارائه و تحلیل شده است و در پایان، بخش ششم به جمع‌بندی یافته‌ها و پیشنهادهای آتی پژوهش اختصاص دارد.

مبانی نظری پژوهش

در بیمه‌های غیرعمر، برای محاسبهٔ حق‌بیمهٔ وانگ از توزیع خسارت استفاده می‌شود و فرمول آن به شرح زیر است (Wang, 1996).

$$P_X = \int_{۰}^{\infty} g[S_X(x)]\ \mathrm{dx}$$

که در این رابطه، $S_X(x)$ تابع بقای توزیع خسارت و برابر با $S_X(x) = ۱ - F_X(x)$ که $F_X(x)$ تابع توزیع خسارت است.

همچنین $g : [۰,۱] \to [۰,۱]$ به‌عنوان تابع انحراف شناخته می‌شود که یک تابع غیرکاهشی و دارای ویژگی‌های $g(۰) = ۰$ و $g(۱) = ۱$ است. توابع انحراف گوناگونی وجود دارند که در ادامه به آن‌ها اشاره می‌شود (Wang et al., 1997).

تابع انحراف توانی

تابع انحراف توانی[1] یکی از روش‌های تبدیل توزیع‌ها در مدل‌های مدیریت ریسک و قیمت‌گذاری است که در تبدیل وانگ کاربرد دارد. این تابع به کمک یک پارامتر انحراف (c)، توزیع احتمالات را برای در نظر گرفتن ریسک یا نگرش به ریسک تغییر می‌دهد. رابطهٔ تابع انحراف توانی به‌صورت زیر است (Wang, 2000).

$$g(x) = x^c \ , ۰ < c < ۱. \qquad (۱)$$

شکل تابع انحراف توانی با cهای مختلف به‌صورت شکل ۱ است

تابع انحراف توانی دوگانه

تابع انحراف توانی دوگانه[2] نسخهٔ تعمیم‌یافته از تابع انحراف توانی است که انعطاف بیشتری برای مدل‌سازی رفتار ریسک‌پذیری و ریسک‌گریزی در شرایط مختلف ارائه می‌دهد. این تابع به‌ویژه در مسائل مالی و بیمه‌گری برای تنظیم احتمالات توزیع به کار می‌رود. رابطهٔ تابع انحراف توانی دوگانه به‌صورت زیر است (Wang, 2000).

$$g(x) = ۱ - (۱ - x)^b \ , b \geq ۱ . \qquad (۲)$$

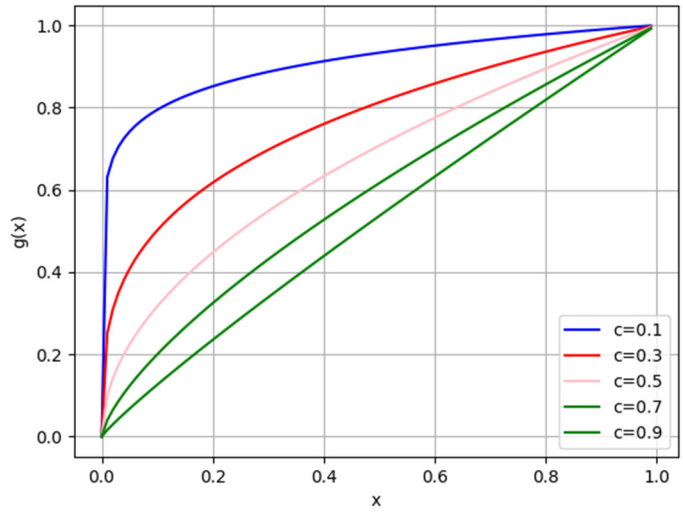
تابع انحراف توانی دوگانه با b های مختلف را می‌توان در شکل ۲ مشاهده کرد.

تابع انحراف انتقال وانگ

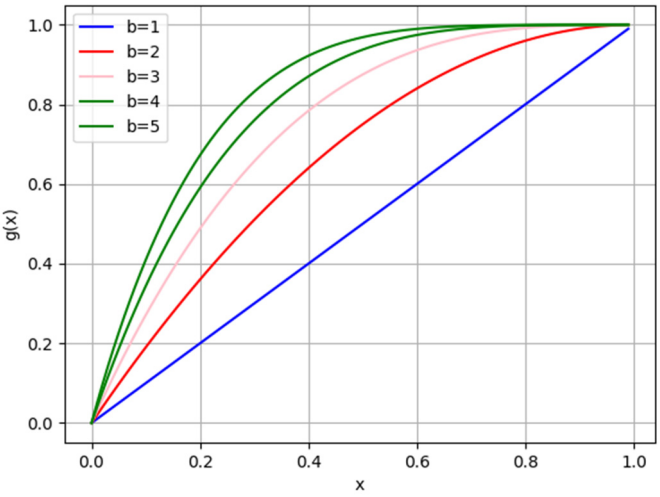
تابع انحراف وانگ[3] یکی از ابزارهای اصلی در مدیریت ریسک، بیمه و قیمت‌گذاری مالی است که از یک مفهوم تغییر احتمال برای تطبیق توزیع اصلی با نگرش به ریسک یا شرایط بازار استفاده می‌کند. این تابع به‌طور خاص در تبدیل وانگ برای قیمت‌گذاری و تغییر توزیع به‌کار می‌رود. رابطهٔ تابع انحراف وانگ به‌صورت زیر نوشته می‌شود (Wang, 2000).

$$g_\lambda(x) = \Phi(\Phi^{-۱}(x) - \lambda), \qquad (۳)$$

1 Power Distortion Function
2 Dual Power Distortion Function
3 Wang Distortion Function



شکل ۱. تابع انحراف توانی
Fig. 1. Power Distortion Function



شکل ۲. تابع انحراف توانی دوگانه
Fig. 2. Dual Power Distortion Function

که در آن λ یک عدد حقیقی است. تابع انحراف وانگ برای لانداهای مختلف را می‌توان در شکل ۳ مشاهده کرد.

در ادامه به رابطهٔ محاسبهٔ حق‌بیمهٔ خالص در بیمه‌های عمر خواهیم پرداخت.

حق‌بیمه

طبق اصل تعادل[1]، ارزش فعلی مزایا با ارزش فعلی حق‌بیمه‌ها در زمان صفر با هم برابرند. شکل ۴ روند محاسبهٔ ارزش فعلی مزایا و شکل ۵ روند محاسبهٔ ارزش فعلی حق‌بیمه‌ها در زمان صفر برای بیمه‌های تمام عمر را نشان می‌دهد.

رابطهٔ حق‌بیمه در زمان صفر به‌صورت زیر بیان می‌شود.

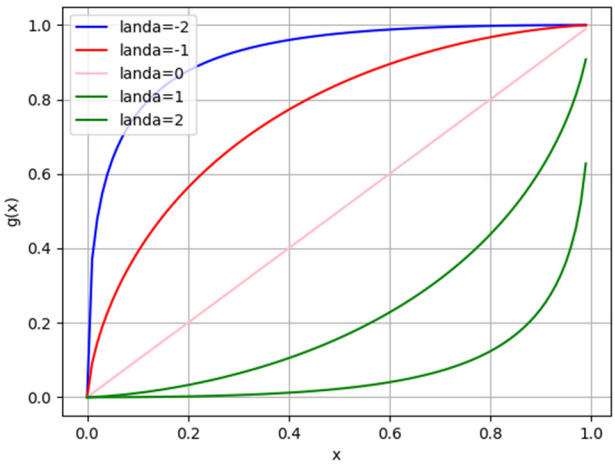
$$\Pi = \frac{A_x}{\ddot{a}_x}.\ C, \qquad (۴)$$

که Π حق‌بیمهٔ خالص، C سرمایهٔ فوت، A_x ارزش فعلی مزایا و $\ddot{a}_x$ ارزش فعلی حق‌بیمه‌هاست. در ادامه به محاسبهٔ حق‌بیمه با استفاده از مدل وانگ می‌پردازیم (Gerber, 1990).

فرض می‌شود که شخص x ساله‌ای تا k سال بعد زنده بماند، در نتیجه احتمال بقای این شخص با ${}_kp_x$ نشان داده می‌شود و به طریق مشابه احتمال فوت این شخص با ${}_kq_x$ بیان می‌شود. حال تبدیل وانگ احتمال بقا تحت تابع انحرافِ رابطهٔ (۳) که با نماد ${}_kp^*_x$ نمایش داده می‌شود به‌صورت زیر است.

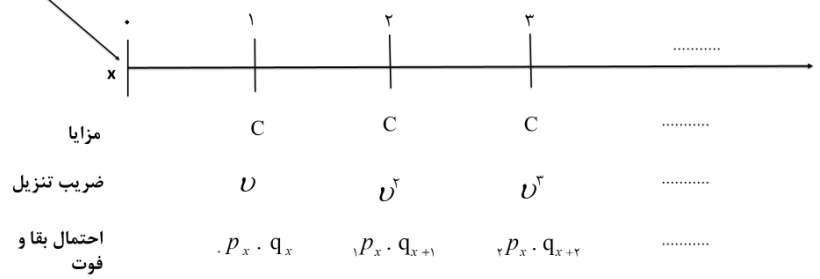
$${}_kp^*_x = g_\lambda({}_kp_x) = \Phi(\Phi^{-1}({}_kp_x) - \lambda) = \Phi(\Phi^{-1}(1 - {}_kq_x) - \lambda), \qquad (۵)$$

1 Equivalence Principle



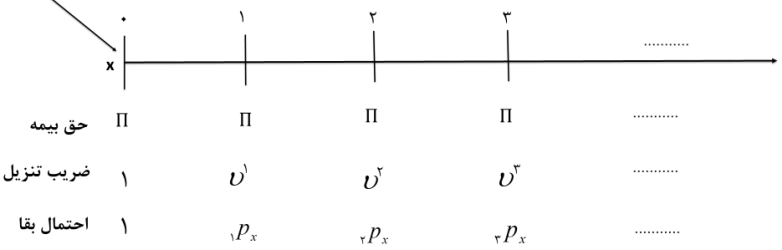
شکل ۳. تابع انحراف وانگ
Fig. 3. Wang Distortion Function

$$A_x = C.\sum_{k=0}^{\infty} \upsilon^{k+1}\ {}_k p_x\ q_{x+k}$$



شکل ۴. ارزش فعلی مزایا در زمان صفر
Fig. 4. Present Value of Benefits at Time Zero

$$\ddot{a}_x = \Pi\sum_{k=0}^{\infty} \upsilon^k\ {}_k p_x$$



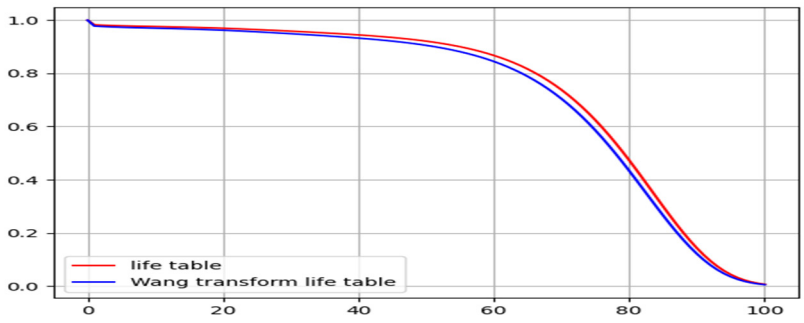
شکل ۵. ارزش فعلی حق‌بیمه‌ها در زمان صفر
Fig. 5. Present Value of Premiums at Time Zero

و به طریق مشابه ${}_k q^*_x = 1 - {}_k p^*_x$ خواهد بود. در رابطهٔ بالا، Φ توزیع نرمال استاندارد، Φ^{-1} وارون توزیع نرمال استاندارد و λ پارامتر بازاری ریسک[1] است که به پارامترهای مختلف بستگی دارد. حال در تمامی روابط ارزش فعلی مزایا و ارزش فعلی حق‌بیمه‌ها که پیش از این مطرح شد به‌جای ${}_k p_x$، ${}_k p^*_x$ و به‌جای ${}_k q_x$، ${}_k q^*_x$ قرار خواهیم داد تا حق‌بیمه در مدل وانگ را به دست آوریم.

در شکل ۶، مقایسه‌ای بین جدول مرگ‌ومیر ایران و مدل تعدیل‌شدهٔ وانگ ارائه شده که در آن پارامتر λ، ۰/۱ فرض شده است.

یکی از روش‌های دیگر برای تعدیل نرخ‌های مرگ‌ومیر، مدل کاکس است. این مدل به‌عنوان یکی از ابزارهای کلیدی در تحلیل

1 Market Price of Risk



شکل ۶. احتمال بقای جدول عمر همراه تبدیل وانگ
Fig. 6. Survival Probability of Life Table with Wang Transformation

بقا شناخته می‌شود و در بیم‌سنجی برای مدل‌سازی زمان وقوع رویدادهای ریسکی، مانند مرگ بیمه‌شده، ورشکستگی شرکت‌ها یا فسخ قراردادهای بیمه‌نامه، کاربرد گسترده‌ای دارد. تابع بقا در مدل کاکس به صورت زیر بیان می‌شود:

$$S(t \mid X) = [S_0(t)]^{\exp(\beta^T X)}$$

که در آن
$S_0(t)$ تابع بقای اولیه و $\exp(\beta^T X)$ نشان‌دهندهٔ اثر متغیرهای ریسکی بر زمان بقاست. بدین‌ترتیب، مدل کاکس با لحاظ هم‌زمان چندین عامل ریسکی، می‌تواند احتمال بقای هر فرد را به‌صورت پویا و منطبق با ویژگی‌های فردی او برآورد کند.

مروری بر پیشینهٔ پژوهش

محاسبهٔ حق‌بیمه همواره از چالش‌های اساسی و کلیدی در صنعت بیمه بوده است. در چهارچوب روش‌های سنتی، مفاهیمی مانند ارزش مورد انتظار، واریانس، انحراف معیار و چندک‌ها برای ایجاد توازن میان ریسک و قیمت‌گذاری منصفانه توسعه یافته‌اند. بااین‌حال، این روش‌ها با توجه به مفروضات ساده‌انگارانه در رابطه با توزیع خسارت‌ها و رفتار ریسک، در شرایط بازارهای نوین بیمه‌ای ناکافی به نظر می‌رسند.

در سال‌های ابتدایی دههٔ ۱۹۹۰، وانگ (Wang, 1995) با بهره‌گیری از نظریهٔ توابع تحریف[1]، مدلی نوین برای حق‌بیمه معرفی کرد که به اصل حق‌بیمهٔ وانگ شناخته شد. این روش امکان قیمت‌گذاری منصفانه‌تر، به‌ویژه در مواجهه با ریسک‌های شدید و نادر را فراهم می‌کند. اصل وانگ (Wang et al., 1997) با تغییر سیستماتیک توزیع احتمال زیان‌ها، حساسیت بیشتری نسبت به ریسک‌های بزرگ ایجاد می‌کند و انعطاف‌پذیری بالاتری در تنظیم سطح ریسک‌گریزی ارائه می‌دهد. این مدل به‌سرعت در صنعت بیمه، به‌ویژه در قیمت‌گذاری قراردادهای بیمهٔ اتکایی، مورد استقبال قرار گرفت، زیرا بدون نیاز به فرض‌های آماری سنگین، امکان لحاظ کردن قضاوت‌های ریسک‌گریزانه را فراهم می‌کرد (Sepanski & Wang, 2023). ایدهٔ استفاده از توابع تحریف، ریشه در مدل‌سازی دقیق‌تر رفتارهای ریسک‌گریز افراد در شرایط عدم قطعیت دارد. در دهه‌های ۱۹۸۰ و ۱۹۹۰، محدودیت‌های روش‌های سنتی مانند میانگین و واریانس در انعکاس نگرانی‌های واقعی بیمه‌گران دربارهٔ خسارت‌های نادر اما شدید آشکار شد. در این دوره، معیارهای ریسک مبتنی بر تحریف، ابتدا به‌صورت نظری و سپس در چهارچوب‌های ریاضی دقیق‌تری نظیر انتگرال چوک توسعه یافتند (Wang, 1996). این مفاهیم با نظریه‌های مطلوبیت و تصمیم‌گیری غیرخطی پیوند خوردند و زمینه‌ساز اصول قیمت‌گذاری نوین شدند (Wang et al., 2020).

در دهه‌های اخیر، محققان این مدل را تعمیم داده و ساختارهای تحریف پیچیده‌تری را بررسی کرده‌اند. مثلاً، پژوهش‌های جین و همکاران (Jin et al., 2024) و وانگ و همکاران (Wang et al., 2020) به توسعهٔ چهارچوب‌های عمومی‌تر برای عملگرهای قیمت‌گذاری مبتنی بر تحریف پرداخته‌اند که در فضاهای احتمالی گسترده‌تر یا بدون فرض توزیع خاص عمل می‌کنند. همچنین، ارتباط این مدل‌ها با نظریه‌های تصمیم‌گیری، بهینه‌سازی محدب و یادگیری ماشین نیز مورد توجه قرار گرفته است (Ieosanurak & Moumeesri, 2025).

ازسوی‌دیگر، کاربرد معیارهای تحریف‌شده در طراحی قراردادهای بهینهٔ بیمه نیز گسترش یافته است. در مطالعهٔ جین و همکاران (Jin et al., 2024)، با استفاده از اصول حق‌بیمهٔ تحریف‌یافته تعمیم‌یافته و با تحمیل شرط عاری از خطر اخلاقی، شرایط لازم و کافی برای طراحی بهینهٔ قرارداد بیمهٔ اتکایی به دست آمده است. هدف اصلی کمینه‌سازی یک سنجه ریسک برای خسارت نگهداری‌شدهٔ بیمه‌گر است و از طریق تبدیل مسئله به یک مسئلهٔ دوگانه مانع نشان داده می‌شود که ساختار بهینهٔ واگذاری ریسک، غالباً به‌صورت توقف زیان یا توقف زیان محدود ظاهر می‌شود.

در این مقاله برای محاسبهٔ حق‌بیمه میزان ریسک هر شخص به‌صورت مجزا در نظر گرفته شده است. ازاین‌رو، ما از الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی[2] و تقویت گرادیانی شدید[3] استفاده کردیم. تحقیقات جدید نشان می‌دهد که توسعه‌های اخیر در حوزه‌های علم داده و یادگیری ماشین، تحولات چشمگیری در

1 Distortion Function

2 Random Forest
3 XGBoost

روش‌های مدل‌سازی ریسک و محاسبات حق‌بیمه به وجود آورده‌اند. در این میان، الگوریتم‌های یادگیری ماشین به‌ویژه روش‌های مبتنی بر درخت تصمیم مانند جنگل تصادفی (Breiman, 2001) و تقویت گرادیانی شدید (Chen & Guestrin, 2016) به‌دلیل قابلیت‌های منحصربه‌فرد در پردازش و تحلیل مجموعه‌های پیچیده و چندبعدی داده، کاربرد وسیعی در صنعت بیمه پیدا کرده‌اند.

این دسته از الگوریتم‌های پیشرفته با توانایی کشف و استخراج الگوهای نهفته در حجم انبوهی از داده‌ها، امکان تخمین دقیق‌تر و کارآمدتر پارامترهای کلیدی در مدل‌های محاسبهٔ حق‌بیمه را فراهم ساخته‌اند. از جمله مهم‌ترین مزایای این فنّاوری‌ها می‌توان به قابلیت شخصی‌سازی نرخ‌های بیمه بر مبنای ویژگی‌ها و مشخصات منحصربه‌فرد هر بیمه‌گذار اشاره کرد.

الگوریتم جنگل تصادفی با ایجاد ترکیبی از چندین درخت تصمیم و استفاده از روش میانگین‌گیری نتایج، دو مزیت عمده ارائه می‌دهد: اولاً از مشکل رایج بیش‌برازش[1] جلوگیری می‌کند و ثانیاً دقت پیش‌بینی‌ها را به‌طور قابل توجهی بهبود می‌بخشد. در حوزهٔ بیمه، این روش به‌ویژه در شناسایی و تحلیل عوامل مؤثر بر ریسک از جمله سن بیمه‌گذار، سوابق پزشکی و الگوهای رفتاری، نتایج بسیار مطلوبی داشته است (Hastie et al., 2009).

در مقابل، الگوریتم تقویت گرادیانی شدید رویکردی متفاوت دارد. این روش با بهینه‌سازی گرادیانی تابع هدف، توانایی منحصربه‌فردی در پردازش داده‌های ناهمگن و پیش‌بینی ریسک‌های پیچیده از خود نشان می‌دهد. قدرت اصلی این الگوریتم در مدل‌سازی روابط غیرخطی و شناسایی تعاملات پیچیده بین متغیرهای مختلف است که آن را به ابزاری ارزشمند در صنعت بیمه تبدیل کرده است (Friedman, 2001).

علاوه‌براین، تحقیقات اخیر نشان داده‌اند که ترکیب مدل‌های یادگیری ماشین با اصول تحریف‌شده مانند اصل وانگ می‌تواند دقت محاسبات را به‌طور چشمگیری بهبود بخشد. مثلاً، مطالعهٔ دنوئیت و همکاران (Denuit et al., 2019) نشان داد که استفاده از الگوریتم‌های یادگیری ماشین در کنار مدل‌های تحریف ریسک، خطای پیش‌بینی را تا ۳۰ درصد نسبت به روش‌های کلاسیک کاهش می‌دهد. همچنین، پژوهش بلایر_ونگ و همکاران (Blier-Wong et al., 2021) با تمرکز بر کاربرد یادگیری ماشین در بیمهٔ عمر، نشان داد که مدل‌های مبتنی بر داده می‌توانند با تحلیل ویژگی‌های پویا مانند داده‌های سلامت دیجیتال و رفتارهای مالی، نرخ‌های حق‌بیمه را با دقت بیشتری شخصی‌سازی کنند.

فراتر از برآورد پارامترهای مدلی، الگوریتم‌های یادگیری ماشین زمینه را برای طراحی هوشمند قراردادهای بیمه‌ای فراهم کرده‌اند. مثلاً، در پژوهش ریچمن (Richman, 2023)، شبکه‌های عصبی عمیق برای مدل‌سازی توزیع خسارت استفاده شد. این رویکرد قادر بود داده‌های ناهمگن از منابع گوناگون مانند سوابق پزشکی و داده‌های حسگرهای پوشیدنی را هم‌زمان دربرگیرد. نتیجهٔ آن نه‌تنها افزایش دقت پیش‌بینی خسارت، بلکه خلق قراردادهای بیمه‌ای تطبیقی بود که مطابق با تغییرات ویژگی‌های فردی بیمه‌گذاران طراحی می‌شوند .علاوه‌براین، پیشرفت‌های اخیر در یادگیری عمیق و روش‌های مبتنی بر هوش مصنوعی مولد، مانند شبکه‌های مولد متخاصم[2]، امکان شبیه‌سازی سناریوهای ریسک پیچیده را فراهم کرده‌اند. مثلاً، مطالعهٔ بیسواز و همکاران (Biswas et al., 2023) نشان داد که استفاده از مولد متخاصم برای تولید داده‌های مصنوعی در شرایط کمبود داده‌های واقعی، می‌تواند به بهبود عملکرد مدل‌های پیش‌بینی ریسک کمک کند. این روش به‌ویژه در بازارهای بیمه‌ای با داده‌های محدود، مانند بیمهٔ عمر در کشورهای در حال توسعه، کاربرد دارد.

یکی از اصلی‌ترین هدف‌های این مقاله قیمت‌گذاری عادلانهٔ حق‌بیمه برای بیمه‌گذاران است، به‌طوری که ریسک‌های افراد در محاسبات لحاظ شود. در مقالهٔ اعلائی (۱۴۰۰)، برای قیمت‌گذاری عادلانه محصولات بیمهٔ عمر، به استفاده از نرخ بهرهٔ فازی پرداخته شده است. در این مقاله با استفاده از این نرخ بهرهٔ فازی، به‌جای محاسبهٔ یک حق‌بیمه، بازه‌ای از حق‌بیمه‌ها را ارائه می‌دهد. از دیگر مقالاتی که به قیمت‌گذاری عادلانه می‌پردازد، مقالهٔ وهابی و پاینده نجف‌آبادی (۱۴۰۲) که با تشکیل داشتمانی از دارایی ریسکی و غیرریسکی به محاسبهٔ استراتژی و نرخ مصرف بهینه برای یک قرارداد بیمهٔ عمر طراحی شده است. همچنین با استفاده از توابع مطلوبیت میزان ریسک‌پذیری افراد در سرمایه‌گذاری‌شان بررسی شده است.

تخصیص احتمال مرگ‌ومیر با توجه به ریسک هر شخص یکی از دیگر از کارهای این پژوهش است. در مقالهٔ اعلائی (۱۴۰۲) به معرفی محصول مستمری برای بیماران دارای انواع سرطان‌ها پرداخته شده است. در این مقاله، تأثیر تعدیل احتمالات مرگ‌ومیر با استفاده از دو رویکرد مختلف ضریب تعدیل و نرخ‌گذاری مبتنی بر سن بر میزان پرداختی مستمری برای بیمه‌شدگان با سنین مختلف بررسی شده است.

در مقالهٔ قربانی و همکاران (۱۴۰۱) به بررسی ریزش مشتریان بیمه‌های زندگی با استفاده از روش‌های داده‌کاوی پرداخته است. این مقاله از الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک و شبکهٔ عصبی برای بررسی ریزش مشتریان بیمه‌های زندگی استفاده شده است. از نتایج این مقاله می‌توان به بیشتر بودن احتمال بازخرید یا ابطال بیمه‌های زندگی در زنان نسبت به مردان، کاهش بازخرید یا ابطال با افزایش سن بیمه‌شده، افزایش بازخرید با افزایش ریسک مشاغل بیمه‌شدگان اشاره کرد.

بنابراین، مجموعهٔ این پژوهش‌ها نشان‌دهندهٔ روند رو به رشد استفاده از روش‌های محاسباتی پیشرفته، شامل نظریهٔ فازی و الگوریتم‌های یادگیری ماشین، در تحلیل ریسک، تعیین حق‌بیمه و پیش‌بینی خسارت‌های بالقوه در صنعت بیمه هستند.

1 Overfitting

2 GANs

روش‌شناسی پژوهش

رابطهٔ میان قیمت‌گذاری منصفانه در بیمه و رضایت مشتریان یکی از موضوعات مهم در صنعت بیمه است. اولین مبحث در این زمینه اعتمادسازی و شفافیت است. به این معنا که قیمت‌گذاری منصفانه، یکی از عوامل اصلی برای جلب اعتماد مشتریان است. اگر مشتریان احساس کنند مبلغی که پرداخت می‌کنند با خطرات واقعی و مزایای دریافتی تناسب داشته باشد، به شرکت بیمه اعتماد بیشتری خواهند کرد. مشتریانی که احساس می‌کنند قیمت‌های بیمه منصفانه و مطابق با خدمات ارائه‌شده است، بیشتر تمایل دارند تا قراردادهای خود را با شرکت بیمه تمدید کنند. ازاین‌رو، رضایت مشتریان به افزایش وفاداری و کاهش نرخ ترک قراردادها منجر می‌شود. شرکت‌های بیمه‌ای که قیمت‌گذاری منصفانه و مطابق با واقعیت داشته باشند، در مقایسه با رقبای خود که شاید قیمت‌های غیرمنصفانه یا ناعادلانه ارائه دهند، می‌توانند سهم بیشتری از بازار به دست آورند و مشتریان بیشتری جذب کنند.

زمانی که مشتریان احساس می‌کنند حق‌بیمه‌های پرداختی منصفانه است، تمایل کمتری به انجام ادعاهای نادرست یا بی‌مورد خواهند داشت. این موضوع به نفع شرکت بیمه و همچنین موجب بهبود تجربهٔ مشتریان می‌شود. قیمت‌گذاری منصفانه که براساس نیازها و شرایط خاص هر مشتری تعیین شود، باعث می‌شود مشتریان احساس کنند که خدمات ارائه‌شده دقیقاً براساس نیازهای آن‌ها تنظیم شده و این موضوع می‌تواند رضایت بیشتری ایجاد کند.

ازاین‌رو، در ادامه با استفاده از الگوریتم‌های یادگیری ماشین و با توجه به ریسک‌های موجود در هر شخص پارامتر ریسک را جداگانه محاسبه خواهیم کرد و به آن‌ها خواهیم پرداخت. خروجی شبیه‌سازی‌ها و نتایج عددی در ادامه گزارش شده و در نهایت نتیجه‌گیری و پیشنهادهای آتی بیان شده است.

برای پیش‌بینی پارامتر بازاری ریسک داده‌های اولیه را لازم داریم. داده‌های اولیهٔ ما شامل جنسیت[1]، سن[2]، استعمال دخانیات[3]، شغل[4]، سابقهٔ بیماری[5]، مدت قرارداد[6]، سرمایهٔ فوت[7] و حق‌بیمه هستند. با توجه به اینکه حق‌بیمه و دیگر پارامترهای ریسک را داریم به محاسبهٔ پارامتر λ که به‌عنوان ضریبی از ریسک بازار است، طبق فرمول حق‌بیمهٔ وانگ خواهیم پرداخت. برای یافتن پارامتر λ برای هر داده از الگوریتم جست‌وجوی دودویی استفاده می‌کنیم. در ادامه به بررسی الگوریتم جست‌وجوی دودویی می‌پردازیم.

الگوریتم جست‌وجوی دودویی[8]

جست‌وجوی دودویی الگوریتمی کارآمد برای پیدا کردن یک عنصر در یک فهرست مرتب‌شده است. این الگوریتم به‌طور بازگشتی یا غیربازگشتی مقدار هدف را در نیمهٔ فهرست جست‌وجو کرده و هر بار نیمی از جست‌وجو را حذف می‌کند. این روند تا زمانی ادامه می‌یابد که مقدار هدف پیدا شود یا ناحیهٔ جست‌وجو به‌طور کامل کوچک شود (Cormen et al., 2009).

اصول جست‌وجوی دودویی:

۱. ابتدا یک فهرست مرتب‌شده را آماده می‌کنیم؛

۲. شروع از وسط: ابتدا مقدار میانهٔ فهرست بررسی می‌شود؛

۳. مقایسه:

- اگر مقدار میانه برابر با مقدار هدف باشد، جست‌وجو تمام می‌شود و آن مقدار پیدا شده است؛

- اگر مقدار میانه بزرگ‌تر از مقدار هدف باشد، پس هدف باید در نیمهٔ چپ فهرست باشد؛

- اگر مقدار میانه کوچک‌تر از هدف باشد، پس هدف باید در نیمهٔ راست فهرست باشد؛

۴. تقسیم فهرست: سپس فهرست به نیمهٔ چپ یا راست محدود و مجدداً همین روند تکرار می‌شود.

این الگوریتم با هر بار نصف کردن محدودهٔ جست‌وجو، سرعت بالایی دارد.

استفاده از این الگوریتم برای پیدا کردن پارامتر λ روش سریعی است. برای هر داده مراحل زیر را انجام می‌دهیم.

ابتدا یک بازه برای λ مشخص می‌کنیم، سپس مقدار λ را برابر میانگین بازه قرار می‌دهیم. اگر

- حق‌بیمه در مدل وانگ با پارامتر λ برابر حق‌بیمه در حالت کلاسیک باشد، این مقدار را λ مناسب آن داده قرار می‌دهیم.

- حق‌بیمه در مدل وانگ با پارامتر λ کمتر از حق‌بیمه در حالت کلاسیک باشد، کران پایین بازه را λ و کران بالا بدون تغییر باقی می‌ماند و مجدداً همین روند تکرار می‌شود.

- حق‌بیمه در مدل وانگ با پارامتر λ بیشتر از حق‌بیمه در حالت کلاسیک باشد، کران بالا بازه را λ و کران پایین بدون تغییر باقی می‌ماند و مجدداً همین روند تکرار می‌شود.

با توجه به محاسبات انجام‌شده به‌طور متوسط ۱۱ تکرار لازم است تا با الگوریتم جست‌وجوی دودویی، λ اولیهٔ هر داده را پیدا کنیم. تمامی λهای اولیه از رابطهٔ $\Pi = \frac{A_x(\lambda)}{\ddot{a}_x(\lambda)} \cdot C$ محاسبه می‌شوند که Π، حق‌بیمه‌های اعلامی به بیمه‌گذاران است که تمامی اضافه نرخ‌ها اعمال شده است.

الگوریتم بالا در پیوست آمده است.

در ادامه بعد از یافتن تمامی λها، از مدل جنگل تصادفی برای پیش‌بینی λهای جدید استفاده می‌کنیم. جزئیات مدل جنگل تصادفی در بخش بعدی بیان می‌شود. شایان ذکر است که حق‌بیمهٔ کلاسیک در این روش همان حق‌بیمه‌هایی است که از بازار می‌تواند استخراج شود. این حق‌بیمه‌ها در شرکت‌های بیمه موجودند و شرکت‌ها با اعمال تمامی اضافه نرخ‌ها به بیمه‌گذاران ارائه می‌دهند (Breiman, 2001).

رابطهٔ حق‌بیمهٔ کلاسیک برای هر فرد به‌صورت زیر محاسبه می‌شود

1 Gender
2 Age
3 Smoker
4 Jobs
5 History of Disease
6 Duration
7 Death Benefit
8 Binary Search

حق بیمهٔ کلاسیک = حق بیمهٔ خالص + (۰/۲ × حق بیمهٔ خالص × استعمال دخانیات) + (۰/۵ × حق بیمهٔ خالص × سابقه بیماری) + (۰/۳ × حق بیمهٔ خالص × شغل)

در این فرمول، حق‌بیمهٔ خالص براساس اصل تعادل و جدول مرگ‌ومیر ایران محاسبه شده، سپس عوامل ریسکی شامل استعمال دخانیات، سابقهٔ بیماری، و شغل پرخطر (که به‌صورت باینری ۰ یا ۱ کدگذاری شده‌اند) با ضرایب مشخص به آن افزوده می‌شوند. این ضرایب کاملاً فرضی است و بنا به شرایط مختلف در شرکت‌های بیمه اعداد متفاوتی می‌توان در نظر گرفت.

در عمل، شرکت‌های بیمهٔ عمر برای بیمه‌گذاران دارای ویژگی‌های پرریسک (مانند استعمال دخانیات، شغل‌های پرخطر یا سابقهٔ بیماری)، ضرایب تعدیل‌کننده‌ای متناسب با سیاست‌های قیمت‌گذاری خود در محاسبهٔ حق‌بیمهٔ خالص اعمال می‌کنند. این ضرایب میزان تأثیر هر عامل ریسکی را بر نرخ نهایی حق‌بیمه تعیین می‌کنند. در این پژوهش، به‌دلیل در دسترس نبودن داده‌های واقعی بیمه‌گذاران، ضرایب یادشده به‌صورت فرضی و مطابق با رویه‌های متعارف صنعت بیمه تعیین شده‌اند تا تأثیر نسبی عوامل ریسک بر حق‌بیمه شبیه‌سازی شود. بدیهی است در صورت وجود داده‌های واقعی، مدل قادر است بدون نیاز به این ضرایب ثابت، مستقیماً از روی داده‌ها اثر هر ریسک را بر حق‌بیمه بیاموزد و برای بیمه‌گذاران جدید نرخ مناسب را تخمین بزند.

در ادامه مشاهده می‌شود که مدل‌های یادگیری ماشین مورد استفاده، پارامتر بازاری ریسک برای افراد جدید را طبق پارامترهای بازاری ریسکِ داده‌های اولیه تخمین می‌زنند. پس حق‌بیمهٔ وانگ وابسته به حق‌بیمهٔ کلاسیک خواهد بود. هرچند که تغییر ضرایب در فرمول حق‌بیمهٔ کلاسیک، به تغییر حق‌بیمهٔ کلاسیک و به‌تبع آن حق‌بیمهٔ وانگ منجر خواهد شد، اما چون مدل از حق‌بیمهٔ کلاسیک الگو می‌گیرد پس نتایج به‌دست‌آمده در بخش نتیجه‌گیری، نقض نخواهد شد.

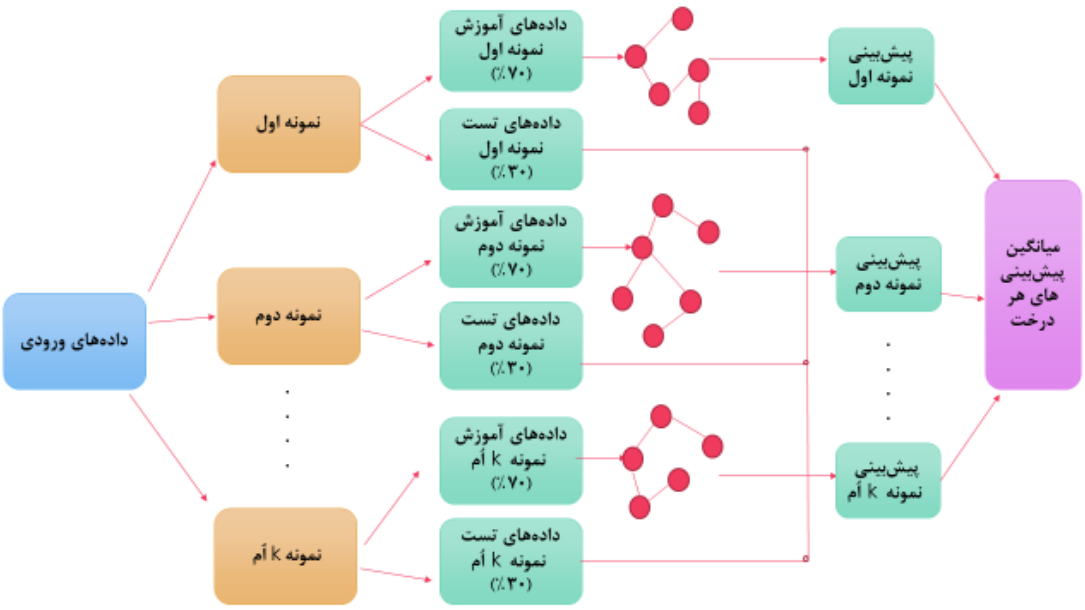
الگوریتم جنگل تصادفی

استفاده از مدل جنگل تصادفی برای پیش‌بینی پارامتر λ در مدل حق‌بیمهٔ وانگ، انتخابی مناسب برای داده‌های پیچیده و غیرخطی است. جنگل تصادفی با ایجاد مجموعه‌ای از درختان تصمیم‌گیری، می‌تواند تأثیر متقابل و غیرخطی بین ویژگی‌ها (مانند جنسیت، سن، استعمال دخانیات، شغل، سابقهٔ بیماری، مدت قرارداد و سرمایهٔ فوت) و پارامتر λ را یاد بگیرد. این الگوریتم به‌دلیل استفاده از چندین درخت تصمیم‌گیری، در برابر نویز در داده‌ها مقاومت بیشتری دارد. همچنین قادر است داده‌هایی را که شامل ویژگی‌های مختلف هستند به‌راحتی مدیریت کند. اولین مرحله برای انجام کار آماده‌سازی داده‌هاست. داده‌هایی که باید به‌عنوان ورودی مدل استفاده شوند شامل جنسیت، سن، استعمال دخانیات، شغل، سابقهٔ بیماری، سرمایهٔ فوت و مدت قرارداد است. مقدار پارامتر λ که در مراحل قبلی با الگوریتم جست‌وجوی دودویی محاسبه شده است، به‌عنوان مقدار هدف[1] برای مدل یادگیری ماشین عمل می‌کند. در گام بعدی داده‌ها را به دو مجموعه تقسیم کرده، ازاین‌رو ابتدا داده‌های آموزش[2] که برای آموزشِ مدل استفاده می‌شوند و شامل درصد زیادی از داده‌ها بوده (معمولاً ۷۰-۸۰٪) را مرتب می‌کنیم. سپس داده‌های آزمون[3] را که برای ارزیابی مدل است و معمولاً ۳۰٪-۲۰ از داده‌ها را تشکیل می‌دهند، تنظیم می‌کنیم. در گام بعد، مدل جنگل تصادفی آموزش داده می‌شود. ابتدا به ایجاد چندین درخت تصمیم‌گیری پرداخته و هر درخت تصمیم‌گیری با استفاده از نمونه‌های تصادفی داده‌ها و زیرمجموعه‌ای از ویژگی‌ها (به‌صورت تصادفی انتخاب‌شده) آموزش داده می‌شود.

این کار به افزایش دقت مدل و کاهش احتمال بیش‌برازش کمک می‌کند که در نهایت مدل از تمام درختان تصمیم‌گیری استفاده کرده تا مقدار λ نهایی را پیش‌بینی کند. نتیجه هر درخت به‌عنوان یک λ پیش‌بینی‌شده محسوب می‌شود و مدل با میانگین‌گیری از λ های پیش‌بینی‌شده، مقدار λ نهایی را تعیین می‌کند. در گام بعدی پس از آموزش مدل، می‌توان از آن برای پیش‌بینی مقدار λ برای افراد جدید استفاده کرد. به این صورت که اطلاعات مربوط به ویژگی‌های فرد جدید به مدل داده می‌شود. مدل، مقدار λ مناسب برای این فرد را پیش‌بینی می‌کند. وقتی مقدار λ برای فرد جدید به دست آمد، می‌توان از آن برای محاسبهٔ حق‌بیمهٔ منصفانه‌تر در روش وانگ استفاده کرد. فرمول وانگ که λ را به‌عنوان یک پارامتر ریسک در خود دارد، ریسک‌های فردی را در نظر می‌گیرد و حق‌بیمهٔ متناسب با شرایط واقعی در بازار به دست می‌آورد.

برای افزایش مقاومت مدل در برابر داده‌های نویزی و جلوگیری از بیش‌برازش، می‌توان به داده‌های آموزشی مقداری نویز تصادفی اضافه کرد. این کار باعث می‌شود مدل روی داده‌های تغییرپذیر بهتر آموزش ببیند. همچنین برای اطمینان از اینکه مدل جنگل تصادفی به بهترین نحو ممکن عمل می‌کند، می‌توان هایپرپارامترها را در مدل جنگل تصادفی تنظیم کرد. پارامترهایی مانند تعداد درخت‌ها، عمق هر درخت، حداقل تعداد نمونه برای تقسیم یک گره، حداقل تعداد نمونه در هر برگ و تعداد ویژگی‌هایی که در هر تقسیم استفاده می‌شود، می‌توانند بهینه شوند تا دقت مدل افزایش یابد. برای این کار از روشی به نام جست‌وجوی تصادفی[4] برای پیدا کردن بهترین هایپر پارامترها استفاده می‌شود. این روش به‌جای آزمایش تمام ترکیبات ممکن از مقادیر هایپرپارامترها (که به آن جست‌وجوی کامل[5] گفته می‌شود)، به‌صورت تصادفی تعدادی از ترکیبات را بررسی می‌کند. این کار سرعت فرایند بهینه‌سازی را وقتی تعداد هایپرپارامترها یا دامنهٔ مقادیر آن‌ها زیاد باشد، افزایش می‌دهد. در نهایت، ترکیبی از هایپرپارامترها که بهترین نتیجه را براساس معیارهای ارزیابی دارد،

1 Target
2 Training Data
3 Testing Data
4 Randomized Search
5 Grid Search



شکل ۷. روندنمای الگوریتم جنگل تصادفی
Fig. 7. Flowchart of the Random Forest Algorithm

به‌عنوان بهترین ترکیب انتخاب می‌شود. معیار میانگین مربعات خطا[1] می‌تواند به ارزیابی دقت پیش‌بینی‌های مدل کمک کند. روندنمای الگوریتم جنگل تصادفی را می‌توان در شکل ۷ مشاهده کرد.

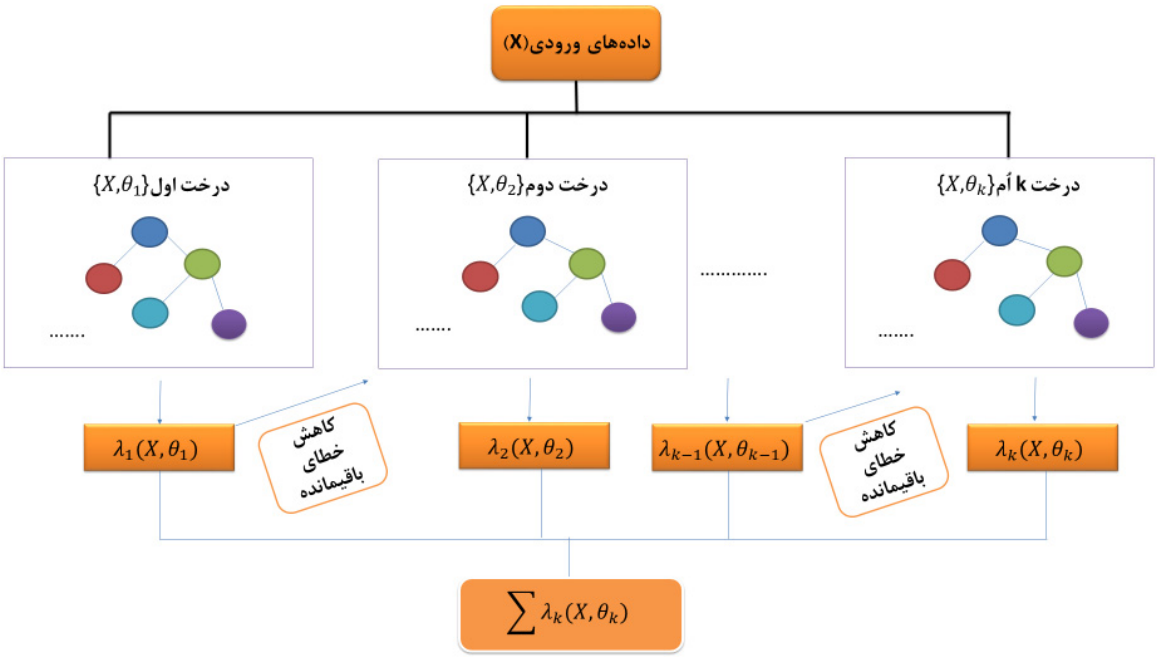
در این پژوهش، پارامتر λ نه به‌عنوان یک پارامتر ثابت در مدل آماری کلاسیک، بلکه به‌صورت تابعی از ویژگی‌های فردی بیمه‌گذاران در نظر گرفته شده است. به بیان دیگر، $f(x)=\lambda$ که در آن x بردار ویژگی‌های بیمه‌گذار است (شامل سن، جنسیت، شغل، استعمال دخانیات، سابقهٔ بیماری، مدت قرارداد و سرمایهٔ فوت). مدل جنگل تصادفی نقش تقریب‌زن تابع $f(\cdot)$ را ایفا می‌کند و هدف آن، پیش‌بینی مقدار ـ برای هر فرد جدید براساس ویژگی‌های اوست. بنابراین، پیش‌بینی ـ در اینجا به‌معنای تخمین مقادیر فردی پارامتر ریسک است، نه برآورد یک پارامتر سراسری در مدل. ازاین‌رو، چهارچوب ارائه‌شده در این پژوهش در حوزهٔ مدل‌سازی مقطعی ریسک[2] قرار می‌گیرد. از دیدگاه آماری، مدل‌های جنگل تصادفی و تقویت گرادیانی شدید در این چهارچوب، نقش توابع رگرسیونی غیرخطی داده‌محور را دارند که مقدار ـ را برای هر مشاهده به‌صورت پویا و دقیق تخمین می‌زنند.

الگوریتم تقویت گرادیانی شدید

الگوریتم تقویت گرادیانی شدید یا بوستینگ گرادیانی شدید، یک الگوریتم یادگیری ماشین بر پایهٔ روش‌های تقویتی است که برای مسائل رگرسیون (پیش‌بینی عددی) و طبقه‌بندی (پیش‌بینی دسته‌ها) استفاده می‌شود. این الگوریتم از درخت‌های تصمیم‌گیری به‌صورت متوالی استفاده می‌کند تا مدلی قوی‌تر، از ترکیب چندین مدل ضعیف‌تر بسازد. در این الگوریتم هر مدل جدید خطاهای مدل قبلی را یاد می‌گیرد و اصلاح می‌کند؛ به بیان دیگر، هر درخت به‌طور مکرر با خطاهای درخت قبلی خود به‌روزرسانی می‌شود و این چرخه ادامه می‌یابد تا بهبود دقت به حد مطلوب برسد. محاسبهٔ خطا با معیار میانگین مربعات خطا در هر مرحله سنجیده می‌شود و مدل سعی می‌کند که در هر مرحله خطا را کم کند. برای افزایش مقاومت مدل در برابر داده‌های نویزی و جلوگیری از بیش‌برازش، می‌توان به داده‌های آموزشی همانند روش جنگل تصادفی مقداری نویز تصادفی اضافه کرد. همچنین برای اطمینان از اینکه مدل تقویت گرادیانی شدید به بهترین نحو ممکن عمل می‌کند، می‌توان هایپرپارامترها را تنظیم کرد. پارامترهایی مانند حداکثر عمق درخت‌ها، نرخ یادگیری، تعداد تخمین‌گرها و دیگر پارامترها مثل جریمه‌ها[3] را که برای جلوگیری از پیچیدگی زیاد مدل استفاده می‌شوند، می‌توان بهینه کرد تا دقت مدل افزایش یابد. برای این کار از روشی به نام جست‌وجوی تصادفی برای پیدا کردن بهترین هایپرپارامترها استفاده می‌شود. مشابه الگوریتم جنگل تصادفی از معیار میانگین مربعات خطا برای ارزیابی دقت پیش‌بینی‌های مدل استفاده می‌شود (Chen & Guestrin, 2016).

در ابتدا داده‌ها به دو مجموعه آموزش و آزمون تقسیم می‌شوند تا مدل براساس مجموعهٔ آموزشی، آموزش داده و عملکرد آن با داده‌های آزمون ارزیابی شود. در نهایت، به‌دلیل کارایی بالا و سرعت پردازش، می‌تواند پارامتر λ را برای داده‌های جدید به‌طور دقیق پیش‌بینی کند و در نتیجه به محاسبهٔ حق‌بیمه‌ای منصفانه و متناسب با شرایط واقعی فرد کمک کند. الگوریتم‌های تقویت گرادیانی شدید و جنگل تصادفی هر دو از درختان تصمیم برای پیش‌بینی و مدل‌سازی

1 Mean Squared Error (MSE)
2 cross-sectional risk modeling
3 Penalty



شکل ۸. روندنمای الگوریتم تقویت گرادیانی شدید
Fig. 8. Flowchart of the Extreme Gradient Boosting Algorithm

استفاده می‌کنند، اما تفاوت‌های مهمی در نحوهٔ ساخت و یادگیری درخت‌ها دارند که بر عملکرد و کاربردهای هریک تأثیر می‌گذارد.

در روش ساخت درخت‌ها، جنگل تصادفی از تعداد زیادی درخت تصمیم‌گیری استفاده می‌کند که به‌طور مستقل و موازی ساخته می‌شوند. هر درخت با یک زیرمجموعهٔ تصادفی از داده‌ها و ویژگی‌ها آموزش می‌بیند. در پایان، پیش‌بینی نهایی به‌صورت میانگین (برای رگرسیون) یا رأی‌گیری اکثریت (برای دسته‌بندی) در بین تمام درخت‌ها انجام می‌شود. اما در تقویت گرادیانی شدید، درخت‌ها به‌صورت ترتیبی و تودرتو ساخته می‌شوند و هر درخت سعی می‌کند خطای درخت قبلی را اصلاح کند. به همین دلیل، تقویت گرادیانی شدید از یک فرایند تقویتی استفاده می‌کند که هر مرحله بهبود تدریجی درخت‌های قبلی را هدف قرار می‌دهد. از نظر مقاومت در برابر بیش‌برازش، جنگل تصادفی به‌دلیل ساختار موازی و استفاده از نمونه‌برداری تصادفی، مقاومت بیشتری در برابر بیش‌برازش دارد. هر درخت به‌صورت مجزا روی زیرمجموعه‌های متفاوتی از داده‌ها آموزش می‌بیندکه این ویژگی به مدل کمک می‌کند به داده‌های نویزدار حساسیت کمتری نشان دهد.

از نظر سرعت و کارایی، جنگل تصادفی به‌دلیل موازی بودن درخت‌ها و عدم وابستگی به‌ترتیب ساخت، در حجم داده‌های زیاد و محاسبات موازی کارایی بهتری دارد. در مقابل، تقویت گرادیانی شدید سرعت بالایی در ساخت درخت‌ها دارد و به‌دلیل بهینه‌سازی‌های سطح سخت‌افزار و نرم‌افزار به‌ویژه برای داده‌های حجیم و پیچیده مناسب است. بااین‌حال، به‌دلیل ساختار ترتیبی و نیاز به آموزش مرحله‌به‌مرحلهٔ درخت‌ها، موازی‌سازی کامل مانند جنگل تصادفی را ندارد.

در زمینهٔ دقت پیش‌بینی، برای داده‌های پیچیده و نویزدار، جنگل تصادفی می‌تواند دقت بالایی داشته باشد و با توجه به ویژگی‌های تصادفی‌سازی، برای داده‌های غیرخطی و پیچیده مناسب است. اما تقویت گرادیانی شدید به‌دلیل ساختار تقویتی و تنظیمات پیشرفته، در بسیاری از موارد می‌تواند دقت بیشتری داشته باشد. این الگوریتم برای مسابقات و چالش‌های یادگیری ماشین که نیاز به دقت بسیار بالا دارند، انتخاب رایجی است. بنابراین، اگر مسئلهٔ مورد نظر به حجم بالای داده و محاسبات سریع نیاز داشته و داده‌ها نویزدار باشند، جنگل تصادفی انتخاب مناسبی است. اما اگر هدف، دقت بالاتر و بهینه‌سازی تدریجی برای کاهش خطاست، و همچنین زمان آموزش مسئله‌ای نیست، تقویت گرادیانی شدید گزینهٔ مناسبی خواهد بود.

توضیحات شکل ۸ به شرح زیر است.

۱. ورودی (Data set X):

- داده‌های اولیه (X) به‌عنوان ورودی به الگوریتم داده می‌شوند.

۲. ساخت درخت‌های تصمیم (درخت اول، درخت دوم و ...):

- الگوریتم با ساخت درخت‌های تصمیم به‌صورت ترتیبی شروع می‌کند. هر درخت با هدف کاهش خطای پیش‌بینی[1] بهینه‌سازی می‌شود.
- در هر مرحله، الگوریتم یک درخت جدید می‌سازد که بر خطای باقی‌مانده تمرکز دارد.

۳. محاسبهٔ خطای باقی‌مانده:

- بعد از ساخت هر درخت، مدل خطای باقی‌مانده بین پیش‌بینی

1 Residual

جدول ۱. نمونه‌های شبیه‌سازی‌شده برای یک محصول عمر زمانی و محاسبهٔ پارامتر بازاری ریسک در مدل جست‌وجوی دودویی
Table 1. Simulated Samples for a Term Life Product and Market Risk Parameter Estimation in the Binary Search Model

ردیف	جنسیت	سن	استعمال دخانیات	شغل	سابقهٔ بیماری	مدت قرارداد	سرمایهٔ فوت	حق‌بیمهٔ کلاسیک (دلار)	λ
1	0	55	1	0	0	3	739	7	0.1567
2	1	75	1	1	1	5	4267	415	0.3842
3	0	52	1	0	0	28	746	11	0.1411
...	...	...	...	...	...	...	...	...	...
998	1	31	1	1	1	49	7437	55	0.3842
999	1	64	0	1	1	1	4994	130	0.2768
10000	0	43	0	0	0	11	9526	46	0.1831

فعلی و مقدار واقعی را محاسبه می‌کند. این خطا به‌عنوان ورودی به درخت بعدی داده می‌شود.

۴. به‌روزرسانی مدل:

• درخت‌های ساخته‌شده با استفاده از یک تابع هدف[1] گره‌های خود را تقسیم‌بندی می‌کنند تا بهترین پیش‌بینی ممکن را ارائه دهند

۵. ترکیب مدل‌ها:

• در پایان پیش‌بینی نهایی به‌صورت مجموع وزن‌دار پیش‌بینی‌های تمامی درخت‌ها ($\sum\lambda_k(X,\theta_k)$) محاسبه می‌شود.

در الگوریتم تقویت گرادیانی شدید، نماد λ نمایانگر یک درخت تصمیم است. هر λ به‌عنوان یک تابع عمل می‌کند که ورودی داده‌ها (X) را دریافت و خروجی پیش‌بینی را تولید می‌کند. این تابع، داده‌های ورودی را از طریق گره‌های درخت عبور می‌دهد و در نهایت یک مقدار پیش‌بینی ارائه می‌کند. پارامترهای این تابع از طریق مجموعه‌ای از پارامترها به نام θ مشخص می‌شوند که نقش مهمی در تعیین ساختار درخت و نحوهٔ پیش‌بینی آن دارند.

پارامترهای مدل، به‌ویژه پارامترهای مرتبط با هر درخت، مشخص‌کنندهٔ نحوهٔ تقسیم‌بندی گره‌ها و میزان تأثیر هر گره بر خروجی نهایی مدل هستند. این پارامترها شامل مواردی مانند مقدار پیش‌بینی گره‌ها، ساختار درخت (از جمله عمق و تعداد شاخه‌ها) و وزن‌های اختصاص‌داده‌شده به گره‌ها هستند. در فرایند بهینه‌سازی، الگوریتم تقویت گرادیانی شدید سعی می‌کند پارامترهای θ را به‌گونه‌ای تنظیم کند که تابع هدف بهینه شود. تابع هدف خود از دو بخش تشکیل شده است: بخش اول خطای مدل است که اختلاف بین پیش‌بینی و مقدار واقعی را نشان می‌دهد و باید به حداقل برسد؛ بخش دوم جریمه‌ای است برای جلوگیری از پیچیدگی بیش از حد مدل، که به منظور مقابله با بیش‌برازش طراحی شده است.

مدل کلی تقویت گرادیانی شدید به‌صورت مجموعی از چندین تابع λ_k یا همان درخت تصمیم ساخته می‌شود، که K تعداد کل درخت‌هاست. هریک از این توابع به‌تنهایی یک درخت تصمیم ساده هستند که تمرکز آن‌ها بر کاهش خطای باقی‌مانده از مراحل قبلی است. در هر مرحله از الگوریتم، یک درخت جدید ساخته می‌شود که هدف آن بهبود پیش‌بینی‌ها براساس خطاهای باقی‌مانده است. این درخت جدید با استفاده از گرادیان نزولی و تابع هدف طراحی می‌شود تا خطای مدل را کاهش دهد. درخت‌های قبلی بدون تغییر باقی می‌مانند و صرفاً درخت جدید به مدل اضافه می‌شود.

تابع هدف در تقویت گرادیانی شدید ترکیبی است از خطای پیش‌بینی، که برابر است با $Loss(y,\hat{y})$ (که $\hat{y}=\sum_{k=1}^{K}\lambda_k(X,\theta_k)$) و جریمه‌ای[2] که پیچیدگی مدل را اندازه‌گیری می‌کند. پیچیدگی مدل شامل تعداد گره‌ها در هر درخت و مقدار وزن‌های اختصاص داده‌شده به گره‌هاست. به‌طور کلی، تابع هدف به صورت زیر تعریف می‌شود:

$$\Omega(\lambda)+Loss(y,\hat{y})=Obj$$

که در آن $Loss$ نشان‌دهندهٔ اختلاف بین پیش‌بینی و مقدار واقعی است و Ω نشان‌دهندهٔ جریمهٔ مربوط به پیچیدگی مدل است. در هر مرحله، الگوریتم در جهت مخالف گرادیان تابع هدف حرکت می‌کند و به‌سمت کمینه‌سازی آن (نقطه‌ای که گرادیان برابر صفر است) پیش می‌رود.

نتایج و بحث

جدول ۱ داده‌های فرضی مورد استفاده در این پژوهش را برای ارزیابی مدل نشان می‌دهد. این داده‌ها به‌صورت تصادفی و از توزیع یکنواخت تولید شده‌اند. برای هر مشاهده، پارامتر بازاری ریسک محاسبه و گزارش شده است. به‌منظور درک بهتر محتوای جدول، در جدول ۲ ویژگی‌های ریسک به‌کاررفته در مدل معرفی شده‌اند.

در این مطالعه، به‌دلیل محدودیت در دسترسی به داده‌های واقعی صنعت بیمه، از داده‌های شبیه‌سازی‌شده استفاده شد. هدف اصلی در این مرحله، ارائهٔ چهارچوبی مفهومی و بررسی امکان‌پذیری به‌کارگیری الگوریتم‌های یادگیری ماشین در پیش‌بینی پارامتر λ و محاسبهٔ حق‌بیمهٔ شخصی‌سازی‌شده براساس مدل وانگ است. بهره‌گیری از داده‌های شبیه‌سازی‌شده این فرصت را فراهم ساخت تا روش پیشنهادی در شرایط کنترل‌شده آزموده و میزان دقت الگوریتم‌ها در بازتولید الگوهای مورد انتظار ارزیابی شود. بااین‌حال، کاملاً آگاهیم که اعتبار نهایی مدل تنها با آزمون آن بر داده‌های واقعی

1 Objective Function

2 Regularization

جدول ۲. خلاصه‌ای از ویژگی‌های ریسک در الگوریتم
Table 2. Summary of Risk Features in the Algorithm

ویژگی‌های ریسک	وضعیت
جنسیت	۰ زن/ ۱ مرد
استعمال دخانیات	۰ عدم استعمال/ ۱ استعمال دخانیات
شغل	۰ کم‌ریسک/ ۱ پرریسک
سابقهٔ بیماری	۰ عدم سابقه/ ۱ دارای سابقهٔ بیماری

جدول ۳. مقایسهٔ حق‌بیمه در حالت کلاسیک و حالت وانگ با λ پیش‌بینی‌شده از الگوریتم جنگل تصادفی
Table 3. Comparison of Premiums in the Classical and Wang Models with λ Predicted by the Random Forest Algorithm

جنسیت	سن	استعمال دخانیات	شغل	سابقهٔ بیماری	مدت قرارداد	سرمایهٔ فوت	λ	حق‌بیمهٔ کلاسیک	حق‌بیمهٔ وانگ
1	20	0	0	0	10	8000	0.3366	24	23
1	20	1	1	1	10	8000	0.4319	32	30
0	52	0	0	0	15	9000	0.0839	84	82
0	52	1	1	1	15	9000	0.3416	150	156
0	70	0	0	0	5	5000	0.0486	155	162
0	70	1	1	1	5	5000	0.3523	301	304

صنعت بیمه قابل دستیابی است. ازاین‌رو، در ادامهٔ مسیر پژوهش و به‌ویژه در مطالعات آتی، استفاده از داده‌های واقعی شرکت‌های بیمه برای ارزیابی و صحت‌سنجی چهارچوب پیشنهادی در دستور کار قرار خواهد گرفت.

داده‌های جدول ۱ براساس ویژگی‌های واقعی بیمه‌گذاران طراحی شده تا الگوهای بازار در شرایط کنترل‌شده بازنمایی شوند. هر بیمه‌گذار دارای مجموعه‌ای از ویژگی‌های ریسک همچون سن، وضعیت استعمال دخانیات و سابقهٔ بیماری است. حق‌بیمهٔ کلاسیک مطابق با روش‌های مرسوم صنعت بیمه محاسبه شد. در این روش، ابتدا حق‌بیمهٔ خالص براساس اصل تعادل تعیین می‌شود و سپس عوامل پرریسک _ همچون استعمال دخانیات، شغل پرخطر یا بیماری‌های خاص _ به‌صورت ضرایب تعدیل‌کننده در حق‌بیمهٔ خالص اعمال می‌شوند. در نهایت، مجموع این عوامل حق‌بیمهٔ کلاسیک نهایی را تشکیل می‌دهد. ازآنجاکه پارامتر بازاری ریسک در داده‌های اولیه با استفاده از حق‌بیمهٔ کلاسیک محاسبه می‌شود، هرگونه تغییر جزئی در متغیرهای شبیه‌سازی‌شده می‌تواند به تغییرات قابل توجهی در نتایج حاصل منجر شود.

در ادامه تعدادی دادهٔ جدید را به‌عنوان ورودی به سیستم می‌دهیم تا مقایسه‌ای بین حق‌بیمه در حالت کلاسیک و حالت وانگ داشته باشیم.

روش جنگل تصادفی برای این داده‌ها با انتخاب بهینه‌ترین هایپرپارامترها، MSE=۰/۰۰۰۲۸ را دارد که بیانگر دقت خوب این روش است و رابطهٔ آن به‌صورت زیر است.

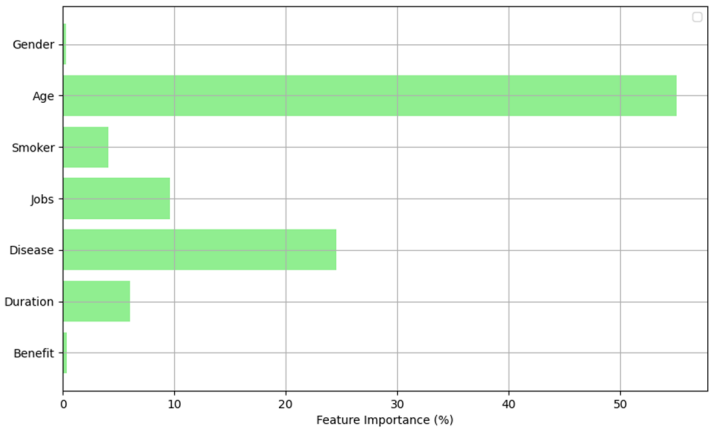
$$MSE = \frac{\sum_{i=1}^{n}(\lambda_i - \hat{\lambda}_i)^2}{n}$$

- n : تعداد داده‌ها
- λ_i : مقدار واقعی پارامتر λ برای فرد iاُم
- $\hat{\lambda}_i$: مقدار پیش‌بینی‌شدهٔ پارامتر λ توسط مدل جنگل تصادفی برای فرد iاُم

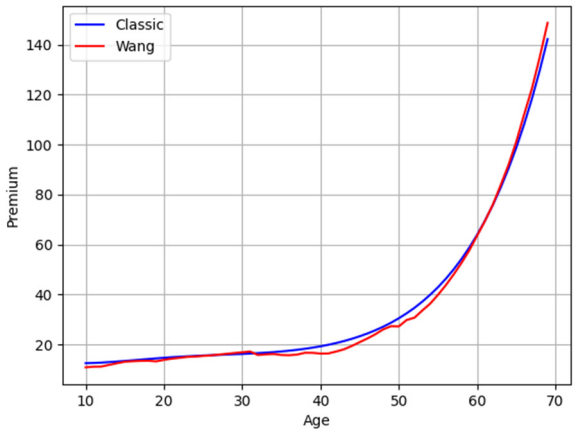
در این پژوهش، پارامتر بازاری ریسک λ به‌طور مستقیم از داده‌های واقعی قابل مشاهده نیست، اما با استفاده از روش عددی و داده‌های شبیه‌سازی‌شده از مدل کلاسیک و الگوریتم‌های یادگیری ماشین برآورد می‌شود. در واقع، مدل ابتدا براساس داده‌های آموزش که شامل حق‌بیمه‌های کلاسیک و ویژگی‌های ریسکی است، مقدار λ را برای هر نمونه محاسبه می‌کند. سپس در مرحلهٔ آزمون (۳۰ درصد داده‌ها)، مدل مقادیر $\hat{\lambda}_i$ را پیش‌بینی می‌کند و میزان انحراف پیش‌بینی‌ها از مقادیر اولیه با معیار میانگین مربعات خطا (MSE) سنجیده می‌شود.

شکل ۹ تأثیر ویژگی‌های هر فرد بر پیش‌بینی مقدار λ را نشان می‌دهد. بررسی تأثیر ویژگی‌های هر فرد در پیش‌بینی پارامتر بازاری ریسک با استفاده از شاخص اهمیت ویژگی‌ها[1] در مدل‌های یادگیری ماشین بوده که در پایتون، کتابخانه‌ای به این نام موجود است. برای این منظور، پس از آموزش مدل بر روی داده‌های تاریخی بیمه‌گذاران (شامل متغیرهایی مانند سن، جنسیت، استعمال دخانیات، شغل، سابقهٔ بیماری، مدت قرارداد و سرمایهٔ فوت)، اهمیت نسبی هر ویژگی

1 Feature Importance



شکل ۹. درصد تأثیرات ویژگی‌های ریسکی هر فرد بر پیش‌بینی پارامتر λ در مدل جنگل تصادفی
Fig. 9. Percentage Impact of Individual Risk Features on λ Prediction in the Random Forest Model



شکل ۱۰. مقایسهٔ حق‌بیمه‌ها در حالت کلاسیک و حالت وانگ برای افراد با ریسک کم در سنین مختلف در مدل جنگل تصادفی
Fig. 10. Comparison of Premiums in the Classical and Wang Models for Low-Risk Individuals at Different Ages in the Random Forest Model

در پیش‌بینی خروجی محاسبه می‌شود.

در جدول ۳ مشاهده شد که هرچه پارامترهای ریسک بیشتر باشند، حق‌بیمهٔ وانگ از حق‌بیمه در حالت کلاسیک بیشتر می‌شود و هرچه پارامترهای ریسک کمتر باشند حق‌بیمهٔ وانگ کمتر از حق‌بیمه در حالت کلاسیک خواهد شد. شکل ۱۰ مقایسهٔ حق‌بیمه‌ها در حالت کلاسیک و حالت وانگ برای افرد با ریسک کم در سنین مختلف را نشان می‌دهد.

شکل ۱۱ مقایسهٔ حق‌بیمه‌ها در حالت کلاسیک و حالت وانگ برای افرد با ریسک بالا در سنین مختلف را نشان می‌دهد.

در جدول ۴ مقایسه‌ای بین احتمال بقا در جدول مرگ‌ومیر و احتمال بقا در مدل وانگ با الگوریتم جنگل تصادفی برای افراد پرریسک با سنین مختلف انجام شده است.

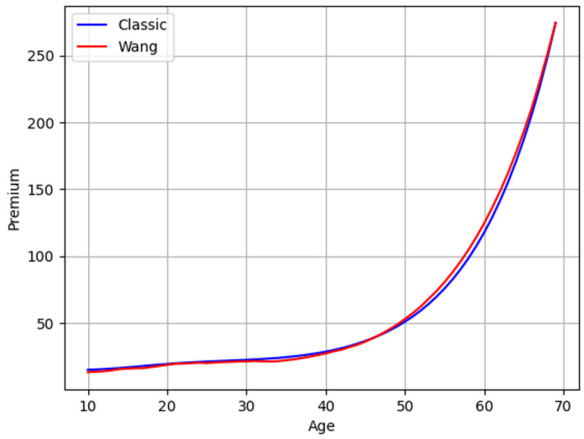
شکل ۱۲ تأثیر ویژگی‌های هر فرد بر پیش‌بینی مقدار λ در الگوریتم تقویت گرادیانی شدید را نشان می‌دهد.

در جدول ۵ مشاهده شد که هرچه پارامترهای ریسک بیشتر باشند، حق‌بیمهٔ وانگ از حق‌بیمه در حالت کلاسیک بیشتر خواهد شد و هرچه پارامترهای ریسک کمتر باشند، حق‌بیمهٔ وانگ کمتر از حق‌بیمه در حالت کلاسیک خواهد شد. شکل ۱۳ مقایسهٔ حق‌بیمه‌ها در حالت کلاسیک و حالت وانگ برای افرد با ریسک کم در سنین مختلف را نشان می‌دهد.

شکل ۱۴ مقایسهٔ حق‌بیمه‌ها در حالت کلاسیک و حالت وانگ برای افرد با ریسک بالا در سنین مختلف را نشان می‌دهد.

در ادامه به تغییر پارامترهای مدل می‌پردازیم تا قابلیت‌های مدل ارزیابی شود. در جدول ۶ پارامترهای استعمال دخانیات، شغل و سابقهٔ بیماری از حالت ۰ یا ۱ بودن به‌صورت درصدی تغییر یافته‌اند تا مدل به واقعیت نزدیک‌تر شود. در جدول ۷ توضیحات مربوط به تغییر پارامترهای ریسکی بیان شده است.

در ادامه تعدادی دادهٔ جدید را به‌عنوان ورودی به سیستم می‌دهیم تا مقایسه‌ای بین حق‌بیمه در حالت کلاسیک و حالت وانگ داشته باشیم. به‌منظور ارزیابی حساسیت مدل نسبت به تغییر پارامترهای



شکل ۱۱. مقایسهٔ حق‌بیمه‌ها در حالت کلاسیک و حالت وانگ برای افراد با ریسک بالا در سنین مختلف در مدل جنگل تصادفی

Fig. 11. Comparison of Premiums in the Classical and Wang Models for High-Risk Individuals at Different Ages in the Random Forest Model

جدول ۴. مقایسهٔ احتمال بقا در جدول مرگ‌ومیر ایران و احتمال بقای مدل وانگ

Table 4. Comparison of Survival Probabilities Between the Iranian Life Table and the Wang Model

سن	۲۰	۳۰	۴۰	۵۰	۶۰	۷۰	۸۰	۹۰	۱۰۰
احتمال بقای جدول مرگ‌ومیر	0.999	0.998	0.998	0.996	0.990	0.973	0.928	0.821	0.551
احتمال بقای وانگ	0.996	0.995	0.994	0.989	0.976	0.942	0.862	0.709	0.404

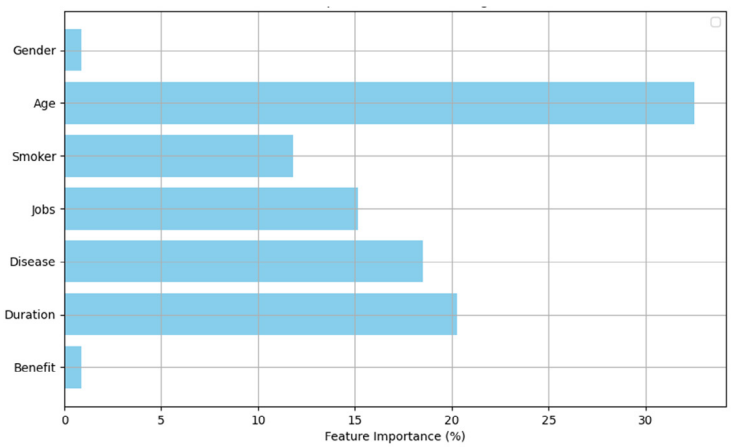
جدول ۵. مقایسهٔ حق‌بیمه در حالت کلاسیک و حالت وانگ با λ پیش‌بینی‌شده از الگوریتم تقویت گرادیانی شدید

Table 5. Comparison of Premiums in the Classical and Wang Models with λ Predicted by the Extreme Gradient Boosting Algorithm

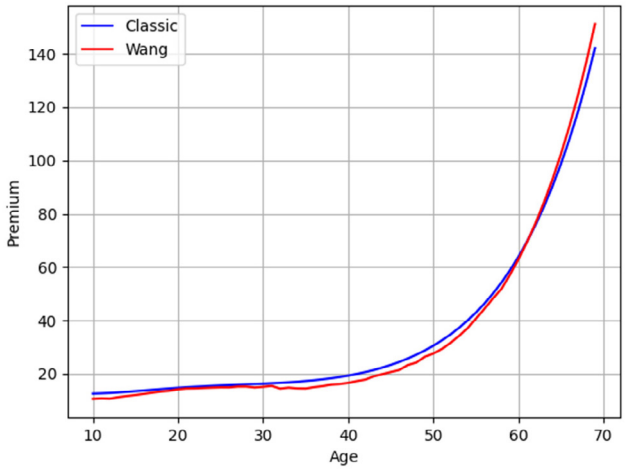
جنسیت	سن	استعمال دخانیات	شغل	سابقهٔ بیماری	مدت قرارداد	سرمایهٔ فوت	λ	حق‌بیمهٔ کلاسیک	حق‌بیمهٔ وانگ
1	20	0	0	0	10	8000	0.3393	24	23
1	20	1	1	1	10	8000	0.4506	32	34
0	52	0	0	0	15	9000	0.0819	84	82
0	52	1	1	1	15	9000	0.3491	150	159
0	70	0	0	0	5	5000	0.0605	155	166
0	70	1	1	1	5	5000	0.3595	301	307

ریسکی، مقادیر متغیرهای استعمال دخانیات، شغل پرریسک و سابقهٔ بیماری از حالت باینری به مقادیر پیوسته در بازهٔ بین صفر و یک تبدیل شدند تا رفتار واقع‌بینانه‌تری از ریسک بیمه‌گذاران شبیه‌سازی شود. نتایج حاصل، که در جدول ۸ و جدول ۹ گزارش شده‌اند، نشان می‌دهند که با افزایش مقدار پارامتر بازاری ریسک λ، اختلاف میان حق‌بیمهٔ وانگ و حق‌بیمهٔ کلاسیک به‌طور قابل ملاحظه‌ای افزایش می‌یابد. مثلاً، افزایش λ از ۰/۰۵ به ۰/۴۵ باعث رشد میانگین حق‌بیمه وانگ بین ۲۵ تا ۴۰ درصد نسبت به مدل کلاسیک شده است. این یافته‌ها بیانگر حساسیت بالای مدل پیشنهادی نسبت به تغییرات ریسک فردی و توانایی آن در بازتاب رفتارهای واقع‌بینانه بازار بیمه است. انتظار می‌رود در صورت به‌کارگیری داده‌های واقعی بیمه‌گذاران، این اختلاف‌ها با دقت و وضوح بیشتری آشکار شود.

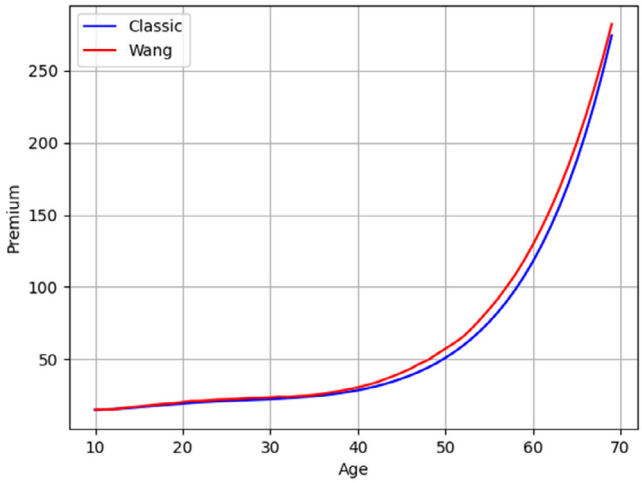
به‌طور مشابه، حق‌بیمهٔ وانگ برای افراد پرریسک بیشتر از حق‌بیمه در حالت کلاسیک خواهد شد و به‌عکس. همچنین در حالتی که مقادیر پارامترهای ریسک بیشتر می‌شود، تفاوت قابل ملاحظه‌ای بین حق‌بیمهٔ کلاسیک و وانگ مشاهده می‌شود. پس مدل



شکل ۱۲. درصد تأثیرات ویژگی‌های ریسکی هر فرد بر پیش‌بینی پارامتر λ در مدل تقویت گرادیانی شدید
Fig. 12. Percentage Impact of Individual Risk Features on λ Prediction in the Extreme Gradient Boosting Model



شکل ۱۳. مقایسهٔ حق‌بیمه ها در حالت کلاسیک و حالت وانگ برای افراد با ریسک کم در سنین مختلف در مدل تقویت گرادیانی شدید
Fig. 13. Comparison of Premiums in the Classical and Wang Models for Low-Risk Individuals at Different Ages in the Extreme Gradient Boosting Model



شکل ۱۴. مقایسهٔ حق‌بیمه‌ها در حالت کلاسیک و حالت وانگ برای افراد با ریسک بالا در سنین مختلف در مدل تقویت گرادیانی شدید
Fig. 14. Comparison of Premiums in the Classical and Wang Models for High-Risk Individuals at Different Ages in the Extreme Gradient Boosting Model

جدول ۶. نمونه‌های شبیه‌سازی‌شدهٔ پیشرفته‌تر برای یک محصول عمر زمانی و محاسبهٔ پارامتر بازاری ریسک در مدل جست‌وجوی دودویی
Table 6. More Advanced Simulated Samples for a Term Life Product and Market Risk Parameter Estimation in the Binary Search Model

ردیف	جنسیت	سن	استعمال دخانیات	شغل	سابقهٔ بیماری	مدت قرارداد	سرمایهٔ فوت	حق‌بیمهٔ کلاسیک (دلار)	λ
1	0	55	0.7	0.07	0.47	3	739	7	0.2016
2	1	75	0.58	0.75	0.94	5	4267	376	0.3276
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
999	1	64	0.11	0.85	0.51	1	4994	111	0.2133
10000	0	43	0.38	0.39	0.05	11	9526	51	0.2250

جدول ۷. خلاصه‌ای از ویژگی‌های ریسک در الگوریتم
Table 7. Summary of Risk Features in the Algorithm

نوع سطح / ویژگی ریسکی	سطح ۱ (ریسک ≤ ٪25)	سطح ۲ (٪50 ≤ ریسک > %25)	سطح ۳ (٪75 ≤ ریسک > %50)	سطح ۴ (ریسک > %75)
استعمال دخانیات	۱ تا ۵ نخ سیگار در روز مصرف کم قلیان و پیپ	۵ تا ۱۰ نخ سیگار در روز مصرف متوسط قلیان و پیپ	مصرف ۱۰ تا ۱۵ نخ سیگار در روز مصرف زیاد قلیان و پیپ	مصرف یک پاکت و بیشتر در روز مصرف خیلی زیاد قلیان و پیپ
شغل	مانند کارمند اداری، معلم یا استاد حسابدار، وکیل	مانند تکنسین آزمایشگاه، تعمیرکار تجهیزات، رانندهٔ درون‌شهری	مانند کارگر ساختمانی، مأمور آتش‌نشانی، کارگر معدن	مانند غواص صنعتی، نیروهای نظامی در مناطق پرخطر
سابقهٔ بیماری	یک بیماری	دو بیماری	سه بیماری	بیش از سه بیماری

جدول ۸. مقایسهٔ حق‌بیمه در حالت کلاسیک و حالت وانگ با λ پیش‌بینی‌شده از الگوریتم جنگل تصادفی برای نمونهٔ دوم
Table 8. Comparison of Premiums in the Classical and Wang Models with λ Predicted by the Random Forest Algorithm for Second Sample

جنسیت	سن	استعمال دخانیات	شغل	سابقهٔ بیماری	مدت قرارداد	سرمایهٔ فوت	حق‌بیمهٔ کلاسیک	حق‌بیمهٔ وانگ
1	20	0.1	0.4	0.2	10	8500	27	18
1	20	0.6	0.8	1	10	8500	32	45
0	50	0.4	0.2	0.5	10	5000	44	36
0	50	0.8	0.8	0.85	10	5000	55	72
0	70	0	0.3	0.05	5	15000	518	535
0	70	0.8	0.3	0.9	5	15000	774	819

به تغییر پارامتر به‌طور محسوسی حساس است.

هرچند اختلاف عددی میان نرخ‌های حق‌بیمهٔ کلاسیک و پیشنهادی در برخی سناریوها اندک به نظر می‌رسد، اما همین تفاوت‌های ظاهراً کوچک، در مقیاس پرتفوی‌های بزرگ می‌تواند به جابه‌جایی‌های مالی قابل توجهی منجر شود و از منظر عدالت بیمه‌ای نیز واجد اهمیت است، زیرا این رویکرد مانع از آن می‌شود که بیمه‌گذاران کم‌ریسک هزینهٔ اضافی بپردازند و بیمه‌گذاران پرریسک به‌طور غیرمستقیم از یارانه پنهان بهره‌مند شوند. افزون‌براین، با تغییر در ترکیب متغیرهای ریسک یا به‌کارگیری داده‌های شبیه‌سازی‌شده متنوع‌تر _ همچون ریسک‌های شغلی شدیدتر یا ابتلا به بیماری‌های خاص _ شکاف میان دو روش برجسته‌تر خواهد شد. بدین‌ترتیب، چهارچوب پیشنهادی نه‌تنها از قابلیت شخصی‌سازی و انعطاف‌پذیری بیشتری برخوردار است، بلکه در سناریوهای واقعی که با ناهمگونی بالای بیمه‌گذاران همراه‌اند، اختلاف معنادارتری در مقایسه با روش کلاسیک ایجاد می‌کند.

در آخر به بررسی و مقایسهٔ مدل کاکس و مدل وانگ می‌پردازیم.

جدول ۹. مقایسهٔ حق‌بیمه در حالت کلاسیک و حالت وانگ با λ پیش‌بینی‌شده از الگوریتم تقویت گرادیانی شدید برای نمونهٔ دوم
Table 9. Comparison of Premiums in the Classical and Wang Models with λ Predicted by the Extreme Gradient Boosting Algorithm for Second Sample

جنسیت	سن	استعمال دخانیات	شغل	سابقهٔ بیماری	مدت قرارداد	سرمایهٔ فوت	حق‌بیمهٔ کلاسیک	حق‌بیمهٔ وانگ
1	20	0.1	0.4	0.2	10	8500	27	19
1	20	0.6	0.8	1	10	8500	32	47
0	50	0.4	0.2	0.5	10	5000	44	34
0	50	0.8	0.8	0.85	10	5000	55	76
0	70	0	0.3	0.05	5	15000	518	538
0	70	0.8	0.3	0.9	5	15000	774	822

ریسک‌های ما شامل استعمال دخانیات، شغل و سابقهٔ بیماری هستند و در مدل کاکس آن‌ها را لحاظ می‌کنیم (پارامتر جنسیت به‌دلیل اینکه در بیمه‌های عمر ایران مطرح نیست لحاظ نمی‌شود. پارامترهای سن و مدت قرارداد مستقیماً در جدول مرگ‌ومیر و پارامتر سرمایهٔ فوت هنگام محاسبهٔ حق‌بیمه لحاظ می‌شوند).

ابتدا به مقایسهٔ احتمالات تعدیل‌شده مرگ‌ومیر در مدل کاکس و وانگ می‌پردازیم.

$$ {}_kp^{*}_x = g_\lambda({}_kp_x) = \Phi(\Phi^{-1}({}_kp_x) - \lambda) = \Phi(\Phi^{-1}(1 - {}_kq_x) - \lambda), $$

$$ {}_kp^{**}_x = ({}_kp_x)^{\exp(\beta_1 .x_1 + \beta_2 .x_2 + \beta_3 .x_3)} $$

که x_1، x_2 و x_3 به‌ترتیب پارامترهای استعمال دخانیات، شغل و سابقهٔ بیماری با ضرایب β_1، β_2 و β_3 هستند. نخستین تفاوت مدل وانگ و مدل کاکس، سهولت محاسبات در مدل وانگ است. در مدل کاکس، ضرایب β مربوط به متغیرهای ریسک از طریق بیشینه‌سازی تابع درست‌نمایی جزئی[1]، با روش‌های تکراری مانند نیوتن_رافسون[2] یا حداقل مربعات با وزن‌دهی تکراری[3] محاسبه می‌شوند. در مقابل، در مدل وانگ فقط یک پارامتر بازاری ریسک λ تخمین زده می‌شود که نمایانگر اثر ترکیبی تمام عوامل ریسک است. این تفاوت، موجب ساده‌تر شدن محاسبات و کاهش بار عددی مدل وانگ نسبت به مدل کاکس می‌شود. افزون‌برآن، درحالی‌که ضرایب β در مدل کاکس پس از برازش ثابت می‌مانند، در مدل وانگ پارامتر λ برای هر بیمه‌گذار به‌صورت شخصی‌سازی‌شده و داده‌محور تخمین زده می‌شود. تفاوت مهم دیگر در قابلیت به‌روزرسانی مدل است؛ مدل کاکس برای داده‌های جدید نیاز به بازبرازش کامل دارد، اما مدل وانگ با ساختار یادگیری ماشین، قابلیت یادگیری تدریجی و انطباق با داده‌های جدید بازار را دارد. در مجموع، مدل کاکس برای تحلیل آماری بقا مناسب است، درحالی‌که مدل وانگ_ یادگیری ماشین برای قیمت‌گذاری منصفانه و مبتنی بر ریسک بازار طراحی شده و از این منظر نسبت به مدل‌های کلاسیک مزیت کاربردی و محاسباتی دارد.

جمع‌بندی و پیشنهادها

اصلی‌ترین هدف این مقاله به دست آوردن جدول مرگ‌ومیر برای هر شخص با توجه به ریسک‌های آن شخص است. ازاین‌رو، با استفاده از مدل‌های ریاضی و الگوریتم‌های یادگیری ماشین، بهینه‌ترین و منصفانه‌ترین روش برای محاسبهٔ حق‌بیمه ارائه شد. نتایج حاصل از این پژوهش نشان می‌دهد که استفاده از مدل حق‌بیمهٔ وانگ در مقایسه با سایر اصول محاسبهٔ حق‌بیمه، دقت و انعطاف‌پذیری بیشتری دارد. همچنین یکی دیگر از نتایج مهم این مقاله این است که هرچه پارامترهای ریسک افراد بیشتر باشند، حق‌بیمهٔ وانگ از حق‌بیمه در حالت کلاسیک بیشتر محاسبه می‌شود و برعکس؛ که این نمونه‌ای از قیمت‌گذاری منصفانه حق‌بیمه توسط روش پیشنهادی را بیان می‌کند.

نتایج حاصل از این پژوهش نشان می‌دهد که استفاده از مدل وانگ در مقایسه با دیگر اصول محاسبهٔ حق‌بیمه، دقت و انعطاف‌پذیری بیشتری دارد. علاوه‌براین، با بهره‌گیری از الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی و تقویت گرادیانی شدید، توانستیم پارامتر مدل وانگ را بهینه‌سازی کرده و به پیش‌بینی دقیق‌تری برای حق‌بیمه دست یابیم. این الگوریتم‌ها نشان دادند که می‌توانند در مواجهه با داده‌های پیچیده و متغیر، عملکرد قابل قبولی داشته باشند و به تصمیم‌گیری‌های دقیق‌تری در صنعت بیمه کمک کنند.

از دیگر نتایج مهم این پژوهش، تأکید بر اهمیت تنظیم حق‌بیمه به‌صورت منصفانه و متناسب با ریسک افراد است. این امر نه‌تنها باعث رضایت بیشتر بیمه‌گذاران می‌شود، بلکه اعتماد آن‌ها را به صنعت بیمه افزایش می‌دهد و به بهبود روابط میان بیمه‌گر و بیمه‌گذار کمک می‌کند.

در نهایت، این پژوهش نشان داد که ترکیب روش‌های ریاضی، تحلیل داده‌ها و الگوریتم‌های یادگیری ماشین می‌تواند به‌مثابهٔ ابزاری کارآمد برای ارتقای سیستم‌های ارزیابی و محاسبهٔ حق‌بیمه استفاده شود. قصد داریم که در پژوهش‌های آتی از داده‌های واقعی

1 log-partial likelihood
2 Newton–Raphson
3 Iteratively Reweighted Least Squares (IRLS)

شرکت‌های بیمه برای سنجش کارایی مدل وانگ نسبت به مدل کلاسیک و همچنین از مدل‌های پیشرفته‌تر یادگیری ماشین برای بهبود بیشتر نتایج استفاده کنیم. همچنین در نظر داریم در پژوهش‌های آتی با افزودن مدل‌های تصادفی به توزیع‌های خسارت رویکرد جدیدی ایجاد کنیم.

مشارکت نویسندگان

امیرمحمد نوروزی: مسئولیت اصلی مفهوم‌سازی و ساختاربندی مسئلهٔ پژوهش، تدوین روش‌شناسی تحقیق، جمع‌آوری و تحلیل مطالعات پیشینه، مدل‌سازی کمّی و کیفی را بر عهده داشته است. سامان وهابی و میترا قنبرزاده: به‌عنوان ناظر علمی در مفهوم‌سازی اولیهٔ پژوهش مشارکت داشته و در طراحی روش‌شناسی، مدل‌سازی و کنترل چهارچوب تحقیق و همچنین نظارت بر تدوین چهارچوب نظری و کنترل استانداردهای پژوهشی همکاری فعالانه داشته است. مجتبی رنجبر: به‌عنوان ناظر علمی، در تمام مراحل تحقیق از مفهوم‌سازی و روش‌شناسی تا نظارت بر چهارچوب تدوین و استانداردهای پژوهشی، راهنمایی‌های تخصصی ارائه کرده است.

سپاسگزاری و قدردانی

این پژوهش برگرفته از پایان‌نامهٔ امیرمحمد نوروزی است که با راهنمایی دکتر مجتبی رنجبر و با مشاورهٔ علمی دکتر سامان وهابی و دکتر میترا قنبرزاده انجام شده است. همچنین این پایان‌نامه با حمایت پژوهشکده بیمه انجام گرفته است. بدین‌وسیله نویسندگان قدردانی و سپاس خود را از همراهی، راهنمایی و حمایت‌های ارزشمند پژوهشکده بیمه اعلام می‌دارند.

تعارض منافع

نویسندگان اعلام می‌دارند که در خصوص انتشار این مقاله تضاد منافع وجود ندارد. علاوه‌براین، موضوعات اخلاقی، از جمله سرقت ادبی، رضایت آگاهانه، سوءرفتار، جعل داده‌ها، انتشار و ارسال مجدد و مکرر و همچنین، سیاست مجله در قبال استفاده از هوش مصنوعی ازسوی نویسندگان رعایت شده است.

دسترسی آزاد

یادداشت ناشر

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

منابع

اعلائی، م. (۱۴۰۰). قیمت‌گذاری محصولات بیمه زندگی در ایران با استفاده از نرخ بهره فازی. *پژوهشنامه بیمه*، ۱۱(۱)، ۱۵-۳۰. https://doi.org/10.22056/ijir.2022.01.02

اعلائی، م. (۱۴۰۲). معرفی محصول مستمری افزایش‌یافته و محاسبه پرداخت‌های آن برای بیمه‌شدگان دارای سرطان‌های مختلف با استفاده از رویکردهای تعدیل احتمالات مرگ‌ومیر. *پژوهشنامه بیمه*، ۱۲(۲)، ۱۴۳-۱۵۴. https://doi.org/10.22056/ijir.2023.02.05

قربانی، ح.، قنبرزاده، م.، و افقی، ر. (۱۴۰۱). بررسی ریزش مشتریان بیمه‌های زندگی با استفاده از روش‌های داده‌کاوی (مطالعه موردی: یکی از شرکت‌های بیمه ایران). *پژوهشنامه بیمه*، ۱۱(۴)، ۳۰۵-۳۲۰. https://doi.org/10.22056/ijir.2022.04.04

وهابی، س.، و پاینده نجف آبادی، الف. (۱۴۰۲). بهینه‌سازی سبد سرمایه‌گذاری برای یک محصول بیمه زندگی پویا با استفاده از ابزارهای کنترل تصادفی. *پژوهشنامه بیمه*، ۱۲(۳)، ۲۲۵-۲۳۸. https://doi.org/10.22056/ijir.2023.03.05

Aalaei, M. (2022). Pricing life insurance products in Iran using fuzzy interest rates. *Iranian Journal of Insurance Research*, *11*(1), 15-30. https://doi.org/10.22056/ijir.2022.01.02 **[In Persian].**

Aalaei, M. (2023). Introducing enhanced annuity product and calculating its payouts for the insureds with different cancers using adjustment approaches of possible mortality. *Iranian Journal of Insurance Research*, *12*(2), 143-154. https://doi.org/10.22056/ijir.2023.02.05 **[In Persian].**

Biswas, A., Al Nasim, A., Imran, A., Sejuty, A.T., Fairooz, F., Puppala, S., & Talukder, S. (2023). *Generative adversarial networks for data augmentation* [Preprint]. arXiv, arXiv:2306.02019 [cs.AI]. https://doi.org/10.48550/arXiv.2306.02019

Blier-Wong, C., Cossette, H., Lamontagne, L., & Marceau, E. (2021). Machine learning in P&C

insurance: A review for pricing and reserving. *Risks, 9*(1), 4. Retrieved May 19, 2024, from https://scholar.google.com/citations?view_op=view_citation&hl=en&user=st1DkB8AAAAJ&citation_for_view=st1DkB8AAAAJ:2osOgNQ5qMEC

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system* [Conference presentation]. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. Retrieved March 15, 2024, from https://arxiv.org/abs/1603.02754

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms* (3rd ed.). MIT Press. http://mitp-content-server.mit.edu:18180/books/content/sectbyfn?collid=books_pres_0&fn=9780262033848_pre_0001.pdf&id=8030

Denuit, M., Hainaut, D., & Trufin, J. (2019). *Effective Statistical Learning Methods for Actuaries I: GLMs and Extensions.* Springer. https://doi.org/10.1007/978-3-030-25820-7

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gerber, H. U. (1990). Life insurance. In *Life Insurance Mathematics* (pp.23–33). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-02655-7_3

Ghorbani, H., Ghanbarzadeh, M., & Ofoghi, R. (2022). Investigating the churn of life insurance customers using data mining methods (A case Study: One of the Iran's insurance companies). *Iranian Journal of Insurance Research*, *11*(4), 305-320. https://doi.org/10.22056/ijir.2022.04.04 **[In Persian].**

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Ieosanurak, W., & Moumeesri, A. (2025). Estimating ruin probability in an insurance risk model using the Wang-PH transform through claim simulation. *Emerging Science Journal, 9*(1), 188–209. https://doi.org/10.28991/ESJ-2025-09-01-011

Jin, Z., Xu, Z. Q., & Zou, B. (2024). Optimal moral-hazard-free reinsurance under extended distortion premium principles. *SIAM Journal on Control and Optimization, 62*(3), 1390–1416. https://doi.org/10.1137/23M1556046

Richman, R. (2023). *Applying Deep Learning in Actuarial Science* [Unpublished PhD thesis]. University of The Witwatersrand. SSRN. https://doi.org/10.2139/ssrn.4722671

Sepanski, J. H., & Wang, X. (2023). New classes of distortion risk measures and their estimation. *Risks, 11*(11), 194. https://doi.org/10.3390/risks11110194

Vahabi, S., & Payandeh Najafabadi, A. (2023). Investment portfolio optimization for a dynamic life insurance product by using stochastic control tools. *Iranian Journal of Insurance Research*, *12*(3), 225-238. https://doi.org/10.22056/ijir.2023.03.05 **[In Persian].**

Wang, Q., Wang, R., & Wei, Y. (2020). Distortion riskmetrics on general spaces. *ASTIN Bulletin, 50*(3), 827–851. https://doi.org/10.1017/asb.2020.14

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin, 26*(1), 71–92. https://doi.org/10.2143/AST.26.1.563234

Wang, S. (2000). A class of distortion operators for pricing financial and insurance risks. *The Journal of Risk and Insurance, 67*(1), 15–36. https://doi.org/10.2307/253675

Wang, S., Young, V. R., & Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics, 21*(2), 173–183. https://doi.org/10.1016/S0167-6687(97)00031-0

پیوست

الگوریتم 1: پیدا کردن λ های اولیه

ورودی: داده‌های اولیه شامل حق‌بیمه‌های اعلامی به بیمه‌گذار

خروجی: مقدار مناسب λ^* برای تطبیق حق‌بیمهٔ کلاسیک

1. مقدار اولیه برای λ ، $a \leftarrow \lambda_{low}, b \leftarrow \lambda_{high}$
2. قرار می‌دهیم P_c (حق‌بیمهٔ کلاسیک) و $P_W(\lambda)$ (حق‌بیمهٔ وانگ)
3. تا زمانی که $|P_c - P_W(\lambda)| \geq \varepsilon$:
   - $\lambda \leftarrow (\frac{\lambda_{high} + \lambda_{low}}{2})$
   - اگر $P_c > P_W(\lambda)$, قرار می‌دهیم $\lambda \leftarrow \lambda_{low}$
   - اگر $P_c < P_W(\lambda)$, قرار می‌دهیم $\lambda \leftarrow \lambda_{high}$
7. برگرداندن $\lambda \leftarrow (\frac{\lambda_{high} + \lambda_{low}}{2})$

الگوریتم 2: الگوریتم جنگل تصادفی برای پیش‌بینی λ های جدید و محاسبهٔ حق‌بیمهٔ وانگ

ورودی: داده‌های اولیه شامل λ های به‌دست‌آمده از الگوریتم اول

خروجی: پیش‌بینی λ برای افراد جدید و محاسبهٔ حق‌بیمهٔ وانگ

1. داده‌های اولیه شامل جنسیت، سن، استعمال دخانیات، شغل، سابقهٔ بیماری، مدت قرارداد، مزایا و λ
2. اعمال الگوریتم جنگل تصادفی به داده‌های اولیه
3. تقسیم‌بندی داده‌ها به k بخش
4. تقسیم داده‌ها به داده‌های آموزش (70٪) و تست (30٪) و ایجاد درخت تصمیم‌گیری
5. میانگین پیش‌بینی‌های همهٔ درخت‌ها
6. وارد کردن ویژگی‌های ریسکی فرد جدید مثل جنسیت، سن، استعمال دخانیات، شغل، سابقهٔ بیماری، مدت قرارداد، مزایا و محاسبهٔ λ به‌وسیلهٔ الگوریتم جنگل تصادفی
7. جای‌گذاری λ ها در رابطهٔ حق‌بیمهٔ وانگ و محاسبهٔ حق‌بیمهٔ وانگ

الگوریتم 3: الگوریتم تقویت گرادیانی شدید برای پیش‌بینی λ های جدید و محاسبهٔ حق‌بیمهٔ وانگ

ورودی: داده‌های اولیه شامل λ های به‌دست‌آمده از الگوریتم اول

خروجی: پیش‌بینی λ برای افراد جدید و محاسبهٔ حق‌بیمهٔ وانگ

1. داده‌های اولیه شامل جنسیت، سن، استعمال دخانیات، شغل، سابقهٔ بیماری، مدت قرارداد، مزایا و λ
2. اعمال الگوریتم تقویت گرادیانی شدید به داده‌های اولیه
3. تقسیم‌بندی داده‌ها به k بخش
4. تقسیم داده‌ها به داده‌های آموزش (70%) و آزمون (30%) و ایجاد درخت تصمیم‌گیری
5. پیش‌بینی هر درخت و ایجاد درخت جدید به‌منظور کاهش خطای درخت قبلی
6. میانگین وزنی پیش‌بینی‌های تمام درختان
7. وارد کردن ویژگی‌های ریسکی فرد جدید مثل جنسیت، سن، استعمال دخانیات، شغل، سابقهٔ بیماری، مدت قرارداد، مزایا و محاسبهٔ λ به‌وسیلهٔ الگوریتم تقویت گرادیانی شدید

معرفی نویسندگان **AUTHOR(S) BIOSKETCHES**

امیرمحمد نوروزی (Amir Mohammad Norouzi)، کارشناسی ارشد، گروه ریاضیات مالی و بیمه‌سنجی، دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران.

- Email: *Amir.norozi1377@gmail.com*
- ORCID: 0009-0009-0444-936X

مجتبی رنجبر (Mojtaba Ranjbar)، دانشیار، گروه ریاضیات مالی و بیمه‌سنجی، دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران.

- Email: *Mranjbar @khu.ac.ir*
- ORCID: 0000-0003-0491-526X

سامان وهابی (Saman Vahabi)، استادیار، گروه ریاضیات مالی و بیمه‌سنجی، دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران.

- Email: *Saman.vahabi70@gmail.com*
- ORCID: 0000-0003-3604-1325

میترا قنبرزاده (Mitra Ghanbarzadeh)، دانشیار، گروه پژوهشی بیمه‌های اشخاص، پژوهشکده بیمه، تهران، ایران.

- Email: *Ghanbarzadeh@irc.ac.ir*
- ORCID: 0000-0003-3486-1338

HOW TO CITE THIS ARTICLE

Norouzi, A. M., Ranjbar, M., Vahabi, S., & Ghanbarzadeh, M. (2026). Life insurance premium pricing through personalized mortality table using the wang premium principle. *Iranian Journal of Insurance Research*, *15*(3), 205-226.

DOI: 10.22056/ijir.2026.03.01

URL: https://ijir.irc.ac.ir/article_160364.html