



ORIGINAL RESEARCH PAPER

## Investigating the churn of life insurance customers using data mining methods (A case Study: One of the Iran's insurance companies)

H. Ghorbani<sup>1</sup>, M. Ghanbarzadeh<sup>2,\*</sup>, R. Ofoghi<sup>1</sup>

<sup>1</sup> Eco College of Insurance, Allameh Tabataba'i University, Tehran, Iran

<sup>2</sup> Department of Personal Insurances Research, Insurance Research Center, Tehran, Iran

### ARTICLE INFO

#### Article History:

Received 16 December 2021

Revised 08 April 2022

Accepted 24 August 2022

#### Keywords:

Customer Churn

Decision Tree

Logistic Regression

Neural Network

Random Forest

### ABSTRACT

**BACKGROUND AND OBJECTIVES:** Customer retention is always considered as the most important principle in all industries, and the insurance industry is no exception. During the recent years in the Iranian society, with the increase in the sale of life insurance policies, the retention of insurance customers has become more and more important to the managers and experts of the insurance industry so that they can keep a wide range of customers. Nowadays, creating a sense of satisfaction in life insurance customers as a management art has been noticed by insurance companies. The more customers the insurance company can keep happy, the less they worry about redemptions and exits. The main goal of this research is to implement data mining methods in predicting customer churn and identifying factors affecting customer churn in the life insurance products of one of Iran's insurance companies. The purpose of customer loss forecasting is to identify the desired class or class related to insurance policies that are suspended or canceled at the request of the policyholder before the end of the insurance coverage period.

**METHODS:** In this paper, we have tried to classify life insurance customers based on abdicantion or non-withdrawal using data mining algorithms such as random forest, decision tree, logistic regression and neural network. The data used in this research include the information of life insurance policies of an insurance company in 2019 in Tehran province, which has a high and appropriate share in the portfolio of the insurance industry. To evaluate and compare these 4 methods, different criteria will be used. In the field of data mining, and in particular the problem of classification, the confusion matrix as a special tabulation makes it possible to visualize the performance of an algorithm. The confusion matrix shows how many true and false predictions have been made for each class, and based on these values, different criteria for classification evaluation and accuracy measurement can be defined.

**FINDINGS:** The results of the research show that random forest, decision tree, logistic regression and neural network algorithms have high performance in predicting the class related to customer churn. Based on the results of the research, the probability of re-buying was better in women and people with high-risk jobs and older age. On the other hand, people who initially paid the insurance premium annually or chose a lower premium and a higher percentage of capital change factor and capital risk, the probability of their redemption was less.

**CONCLUSION:** Considering that life insurance is usually long-term and also considering the liquidity needs of customers and the current economic conditions of the society, insurance companies should pay more attention to life insurance customers. Also, they should put fidelity programs in order to keep customers on their agenda by continuously monitoring the customer's behavior during the insurance policy.

\*Corresponding Author:

Email: [ghanbarzadeh@irc.ac.ir](mailto:ghanbarzadeh@irc.ac.ir)

Phone: +9821 22083895

ORCID: [0000-0003-3486-1338](https://orcid.org/0000-0003-3486-1338)

DOI: [10.22056/ijir.2022.04.04](https://doi.org/10.22056/ijir.2022.04.04)

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).





مقاله علمی

بررسی ریزش مشتریان بیمه‌های زندگی با استفاده از روش‌های داده‌کاوی (مطالعه موردی: یکی از شرکت‌های بیمه ایران)

حسن قربانی<sup>۱</sup>، میترا قنبرزاده<sup>۲\*</sup>، رضا افقی<sup>۱</sup>

<sup>۱</sup> مؤسسه آموزش عالی بیمه اکو، دانشگاه علامه طباطبائی، تهران، ایران

<sup>۲</sup> گروه پژوهشی بیمه‌های اشخاص، پژوهشکده بیمه، تهران، ایران

چکیده:

**پیشینه و اهداف:** حفظ مشتریان همواره به عنوان مهم‌ترین شاکله در همه صنایع تلقی می‌شود و صنعت بیمه نیز از این امر مستثنی نیست. طی سال‌های اخیر در ایران و با افزایش فروش بیمه‌نامه‌های زندگی، حفظ مشتریان بیمه به گونه‌ای مورد توجه مدیران و صاحب نظران صنعت بیمه قرار گرفته است که با ارائه پوشش‌های متنوع بیمه ای طیف وسیعی از مشتریان خود را راضی نگه دارد. امروزه، ایجاد حس رضایت در مشتریان بیمه‌های زندگی توسط شرکت‌های بیمه یک هنر محسوب می‌شود و هر چه شرکت بیمه مشتریان بیشتری را راضی نگه دارد، دیگر نگران بازخرید و خارج شدن مشتریان خود نیست. هدف اصلی مقاله، پیاده‌سازی روش‌های داده‌کاوی در پیش‌بینی ریزش مشتری و شناسایی عوامل تاثیرگذار بر ریزش مشتری در محصولات بیمه زندگی یکی از شرکت‌های بیمه در ایران است. منظور از پیش‌بینی ریزش مشتری شناسایی مطلوب طبقه یا کلاس مربوط به بیمه‌نامه‌هایی است که قبل از پایان یافتن زمان پوشش بیمه، به درخواست بیمه‌گذار، متوقف و پایان می‌یابد.

**روش‌شناسی:** در مقاله حاضر، سعی شده است با بهره‌گیری از الگوریتم‌های داده‌کاوی مانند جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک و شبکه عصبی به طبقه‌بندی مشتریان بیمه‌های زندگی بر اساس ریزش و یا عدم ریزش بپردازیم. داده‌های مورد استفاده در این تحقیق، شامل اطلاعات بیمه‌نامه‌های زندگی یک شرکت بیمه پایلوت در سال ۱۳۹۸ در استان تهران است که سهم مناسب و بالایی در پرتفوی صنعت بیمه دارد.

**یافته‌ها:** نتایج تحقیق حاکی از عملکرد بالای الگوریتم‌های جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک و شبکه عصبی در پیش‌بینی کلاس مربوط به ریزش مشتریان دارد. براساس نتایج حاصل از تحقیق، احتمال باز خرید در زنان و افراد دارای مشاغل پر ریسک و سن بالاتر، بیشتر است. از طرف دیگر، افرادی که در ابتدا پرداخت حق بیمه را به صورت سالانه، حق بیمه کمتر و درصد ضریب تغییر سرمایه و ریسک سرمایه بیشتری را انتخاب کرده‌اند، احتمال باز خرید آن‌ها کمتر بوده است.

**نتیجه‌گیری:** با توجه به بلندمدت بودن بیمه‌های زندگی و نیاز به نقدینگی مشتریان با توجه به شرایط اقتصادی، شرکت‌های بیمه باید توجه بیشتری به مشتریان بیمه‌های زندگی داشته باشند و ضمن رصد رفتار مشتری در طول بیمه‌نامه، برنامه‌های وفاداری به جهت حفظ مشتری را در دستور کار خود قرار دهند.

اطلاعات مقاله

تاریخ‌های مقاله:

تاریخ دریافت: ۲۵ آذر ۱۴۰۰

تاریخ داوری: ۱۹ فروردین ۱۴۰۱

تاریخ پذیرش: ۰۲ شهریور ۱۴۰۱

کلمات کلیدی:

بیمه زندگی

جنگل تصادفی

درخت تصمیم

رگرسیون لجستیک

ریزش مشتری

شبکه عصبی

\*نویسنده مسئول:

ایمیل: [ghanbarzadeh@irc.ac.ir](mailto:ghanbarzadeh@irc.ac.ir)

تلفن: ۰۲۰۸۳۸۹۵-۹۸۲۱

ORCID: 0000-0003-3486-1338

DOI: 10.22056/ijir.2022.04.04

توجه: مدت زمان بحث و انتقاد برای این مقاله تا ۱ ژوئیه ۲۰۲۳ در وب سایت IJIR در «نمایش مقاله» باز می‌باشد

مشتری در دستور کار خود قرار دهد.

در مقاله حاضر به پیش‌بینی (طبقه‌بندی) ریزش مشتری در بیمه‌نامه‌های زندگی یک شرکت بیمه ایرانی با استفاده از الگوریتم‌های داده‌کاوی (جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک و شبکه‌های عصبی) می‌پردازیم. شایان ذکر است، بیمه‌نامه‌های مورد بررسی در شرکت بیمه مورد اشاره در تحقیق، از نوع بیمه‌های عمر جامع یا همان بیمه‌های عمر و سرمایه‌گذاری هستند که علاوه بر پوشش ریسک فوت، جنبه سرمایه‌گذاری نیز دارند و بیمه‌گذار می‌تواند هر ساله حق بیمه و سرمایه فوت خود را بر مبنای نرخ تعدیل، افزایش یا کاهش دهد.

مزیت و جنبه نو بودن مقاله حاضر در داخل کشور، کاربرد چندین روش داده‌کاوی از قبیل رگرسیون لجستیک، درخت تصمیم‌گیری، شبکه‌های عصبی و جنگل تصادفی و مقایسه عملکرد آن‌ها بر پیش‌بینی ریزش مشتریان در بیمه‌های زندگی که بر مبنای داده‌های یک شرکت بیمه صورت گرفته است. با توجه به روند افزایشی بازخرید در بیمه‌های زندگی، شناسایی ویژگی‌های مشتریانی که بیمه‌نامه خود را بازخرید می‌کنند، اهمیت زیادی برای شرکت بیمه دارد زیرا می‌تواند بر مبنای آن‌ها، راهکارهایی را در جهت حفظ مشتریان به کار گیرد. این راهکارها در قالب برنامه‌های وفاداری مشتریان قابل انجام است.

### مبانی نظری پژوهش

#### حفظ مشتری در بیمه‌های زندگی

ارزش اقتصادی حفظ مشتری در بیمه‌های زندگی، به طور گسترده‌ای شناخته شده است: حفظ موفقیت‌آمیز مشتری، نیاز به جستجوی مشتریان جدید و بالقوه با ریسک بالاتر را کاهش داده و به سازمان‌ها اجازه می‌دهد تا با ایجاد روابط مناسب‌تر، با دقت بیشتری روی نیازهای مشتریان موجود خود تمرکز کنند (HelmZadeh et al., 2019b; Mahmoudi Maymand et al., 2022). این موضوع را می‌توان از دو بعد بررسی نمود. از یک طرف، مشتریان قدیمی در بیمه‌های زندگی، ماندگاری بیشتری دارند و در صورت رضایت از خدمات بیمه زندگی، ممکن است مشتری‌های جدیدی را از طریق تبلیغات مثبت برای شرکت فراهم کنند (Ganesh et al., 2000; Paulin et al., 1998). از طرف دیگر، از دست دادن مشتری نه تنها منجر به کاهش فروش، بلکه به از دست دادن هزینه‌های فرصت برای جذب مشتری جدید نیز منجر می‌شود (Motarjem and Niakan, 2020). در واقع، هزینه فروش کالا به یک مشتری جدید، شش برابر هزینه فروش به مشتری قدیمی است و معمولاً هر مشتری ناراضی، عدم رضایت خود را

با این که بیمه‌های زندگی در رشد و توسعه صنعت بیمه اهمیت فراوانی دارند اما بر اساس اطلاعات سالنامه آماری صنعت بیمه (Statistical Yearbook of the Insurance Industry, 2020)، ضریب نفوذ بیمه‌های زندگی در کشور، پایین و با متوسط جهانی اختلاف فاحشی دارد. با توجه به بلنمدت بودن بیمه‌نامه‌های زندگی و نحوه پرداخت حق بیمه (درصد بسیار بالایی از روش پرداخت حق بیمه منظم و دوره‌ای است)، میزان حق بیمه تولیدی به نوعی حالت تجمعی پیدا کرده است. بدین معنی که هر سال علاوه بر حق بیمه تولیدی در بیمه‌نامه‌های جدیدی که به فروش می‌رسد، بخش قابل توجهی از حق بیمه تولیدی، مربوط به وصول حق بیمه‌هایی است که در سال‌های قبل صادر شده‌اند. در نتیجه، یکی از اصلی‌ترین راهکارهای قابل انجام برای رسیدن به ضریب نفوذ مطلوب بیمه‌های زندگی در وهله اول، حفظ بیمه‌نامه‌های صادره قبلی، که در حال حاضر معتبر و جاری می‌باشند، بوده و سپس افزایش تعداد بیمه‌نامه‌های صادره از طرق مختلف مانند ایجاد تنوع در محصول، توسعه شبکه فروش و ... در سال‌های آتی خواهد بود. از این‌رو، ضروری است تا علاوه بر مدنظر قرار دادن راهکارهایی جهت افزایش تقاضا توسط صنعت بیمه، راهکارهایی نیز برای نگهداری مشتری و جلوگیری از بازخرید بیمه‌نامه‌های زندگی توسط شرکت‌های بیمه در نظر گرفته شود (Mahdavi and Majed, 2017).

بر این اساس، شناسایی و تجزیه و تحلیل رفتار و استخراج الگوهای ریزش مشتری همیشه یکی از راهکارهای مدیران ارشد و میانی بسیاری از شرکت‌های بیمه جهان است تا بر مبنای آن، برنامه‌هایی در جهت حفظ مشتریان بیمه‌های زندگی در شرکت بیمه مورد استفاده قرار گیرد (HelmZadeh et al., 2019a; Khalilzadeh Talatapeh et al., 2022). از این‌رو، در کنار بررسی تغییرات اقتصادی مانند تغییرات نرخ تورم که در شرایط فعلی کشور یکی از دلایل اصلی بازخرید بیمه‌های زندگی محسوب می‌شود، نیاز به درک عوامل مؤثر در الگوی ریزش مشتری و برآورد تأثیر عوامل در ریزش نقش اساسی در تصمیم‌گیری‌های استراتژیک و بلنمدت در این زمینه دارد. یکی از رویکردهای متداول در این زمینه، استفاده از روش‌های داده‌کاوی است که ریزش مشتری را با استفاده از روش‌های مختلف پیش‌بینی و بر مبنای عوامل مؤثر بر آن، راهکارهایی را برای شرکت در نظر می‌گیرند. در واقع، در پیش‌بینی ریزش مشتری بر مبنای روش‌های داده‌کاوی، طبقه مربوط به ریزش مشتری در مشتریان شناسایی و ویژگی‌های آن طبقه مورد بررسی قرار می‌گیرند تا بر اساس آن‌ها شرکت بیمه اقداماتی را در جهت بهبود و یا حذف معضلات ناشی از ریزش

مؤثر بر آن، روش‌های داده‌کاوی مختلفی وجود دارند که شامل الگوریتم‌های یادگیری با نظارت و الگوریتم‌های یادگیری بدون نظارت است. در یادگیری با نظارت، از ابتدا برچسب دسته‌ها (ریزش/عدم ریزش) مشخص هستند و هر یک از داده‌های آموزشی به دسته‌ای خاص نسبت داده می‌شوند و اصطلاحاً گفته می‌شود نظری وجود دارد که در هنگام آموزش، اطلاعاتی علاوه بر داده‌های آموزش در اختیار یادگیرنده قرار می‌دهد. درحالی‌که، در یادگیری بدون نظارت، هیچ اطلاعاتی به جز داده‌های آموزشی در اختیار یادگیرنده قرار ندارد و یادگیرنده است که بایستی در داده‌ها به دنبال ساختاری خاص بگردد. برای مدل‌های داده‌کاوی و یادگیری ماشین با نظارت، فنون مختلفی وجود دارند که برخی از آن‌ها شامل جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک و شبکه‌های عصبی هستند (Azzone et al., 2022; Reck et al., 2022). با توجه به استفاده از این روش‌ها در مقاله حاضر، در ادامه، توضیحات مختصری در خصوص هر کدام از این روش‌ها ارائه می‌شود.

یکی از پرکاربردترین الگوریتم‌های داده‌کاوی، الگوریتم درخت تصمیم است. در داده‌کاوی، درخت تصمیم یک مدل پیش‌بینی‌کننده است به طوری که می‌تواند برای هر دو مدل رگرسیون و طبقه‌بندی مورد استفاده قرار گیرد. در ساختار درخت تصمیم، پیش‌بینی به دست آمده از درخت در قالب یک سری قواعد توضیح داده می‌شود که دارای دو مرحله ساخت درخت و هرس درخت است. نحوه هرس کردن درخت اهمیت فراوانی دارد؛ زیرا اطلاعات هر شاخه بر اساس هرس انجام شده، مشخص می‌شود. هر مسیر از ریشه تا یک برگ درخت تصمیم، یک قانون را بیان می‌کند و در نهایت برگ با کلاسی که بیشترین مقدار رکورد در آن تعلق گرفته، برچسب می‌خورد. در نهایت، برچسب متعلق به شاخه درخت می‌تواند شامل ریزش و یا عدم ریزش مشتری باشد (Yang, 2019).

دیگر الگوریتم مورد استفاده، رگرسیون لجستیک است (Han et al., 2012). در این روش، از یک مدل رگرسیونی استفاده خواهد شد که متغیر وابسته در آن دودویی و به صورت صفر و یک است و متغیرهای مستقل نشان‌دهنده، عوامل تأثیرگذار بر ریزش هستند. به عبارت دیگر، متغیر وابسته به صورت ریزش مشتری (عدد یک) و عدم ریزش مشتری (عدد صفر) و متغیرهای مستقل می‌توانند به صورت متغیرهای دموگرافیک بیمه‌گذار/بیمه‌شده و متغیرهای مرتبط با بیمه‌نامه باشند. پس از برازش مدل رگرسیون لجستیک، احتمال ریزش مشتری در بیمه‌نامه مشخص می‌شود که می‌توان این احتمال را بر اساس متغیرهای مورد بررسی، محاسبه و در خصوص ارائه راهکارهایی در صورت ریزش برنامه‌ریزی نمود.

الگوریتم دیگر، شبکه‌های عصبی است که با توجه به ماهیت بالای یادگیری در آن، از روش‌های پرکاربرد در شناسایی ریزش

با ۸ الی ۱۰ نفر به اشتراک می‌گذارد (Qarahnejad, 2009). بر این اساس، اگر شرکتی بتواند میزان نگهداری و حفظ مشتری سالیانه خود را ۵ درصد افزایش دهد؛ می‌تواند منافع و سود خود را بین ۳۰ تا ۱۲۵ درصد افزایش دهد.

بر این اساس، یک مشکل اصلی که در بخش بیمه‌های زندگی با توجه به ماهیت بلندمدت آن و فاصله طولانی مابین تعهدات بیمه‌گر و حق بیمه‌های پرداختی توسط بیمه‌گذار، برای شرکت‌های بیمه به وجود می‌آید، ریزش مشتری است. در واقع، در بیمه‌های زندگی با توجه به طولانی بودن مدت زمان پرداخت حق بیمه و عدم دریافت مزایا در این مدت، مشتریان دچار خستگی شده و یا بنا به نیازهای مالی و مشکلات اقتصادی، بیمه‌نامه‌های خود را بازخرید یا فسخ می‌کنند (Mahdavi et al., 2015). از این‌رو، می‌توان با طبقه‌بندی مشتریان و شناسایی گروه مشتریانی که به دنبال بازخرید یا فسخ بیمه‌نامه خود هستند؛ با انجام اقدامات پیشگیرانه، این موارد را کاهش داد. به منظور شناسایی این مشتریان، لازم است در ابتدا، عوامل مؤثر بر ریزش مشتری شناسایی و سپس بر مبنای آن، ریزش مشتری طبقه‌بندی گردد و در نهایت، راهکارهایی را به جهت حفظ مشتریان بیمه‌های زندگی ارائه نمود. این عوامل می‌توانند شامل مواردی چون عوامل مربوط به بیمه‌گذار/بیمه‌شده مانند سن، جنسیت، وضعیت تأهل، سطح تحصیلات، میزان درآمد و ...، عوامل مربوط به بیمه‌نامه مانند مدت بیمه‌نامه، نحوه پرداخت حق بیمه، وجود پوشش تکمیلی، سرمایه بیمه و ... و عوامل اقتصادی مانند تورم بالا، نرخ سود پایین، نرخ بیکاری و موارد این‌چنینی باشد. در ریزش مشتری عوامل دیگری همچون نحوه تعامل نماینده و یا کارگزار با مشتری و نحوه ارائه محصول به مشتری نیز دارای اهمیت است ولی با توجه به اینکه هدف از مقاله حاضر، پیش‌بینی ریزش مشتری بر مبنای داده‌های ثبتي شرکت بیمه است، این عوامل در کنار عوامل اقتصاد کلان مانند نرخ تورم و یا نرخ ارز و شاخص طلا، نسبت سالخوردگی و ... مورد بررسی قرار نگرفته‌اند. برای ورود متغیرهایی مانند تورم در مدل که یکی از دلایل اصلی بازخرید در بیمه‌های زندگی در کشور است، نیاز به سری زمانی از روند بازخرید یا اطلاعات دریافتی از بیمه‌گذار از طریق مطالعه میدانی (مصاحبه/پرسشنامه) است. بر این اساس، این رویکرد متفاوت از ماهیت فعلی تحقیق است و روش تحقیق جداگانه‌ای مانند مطالعه میدانی از طریق مصاحبه با بیمه‌گذارانی که بیمه‌نامه خود را بازخرید کرده‌اند و یا استفاده از مدل‌های اقتصادسنجی به منظور بررسی تأثیرات در طول زمان، می‌خواهد.

کاربرد روش‌های داده‌کاوی در حفظ مشتریان بیمه‌های زندگی به منظور شناسایی طبقه مربوط به ریزش مشتری و عوامل

به تجزیه و تحلیل روی داده‌های مشتریان بیمه‌های زندگی پرداختند و مشتریان خود را به دسته‌های باز خرید و در معرض ریزش/باز خرید طبقه‌بندی کردند. (Jandaghi and Moradpour 2015) به تقسیم‌بندی مشتریان بیمه زندگی بر اساس ویژگی‌هایی مانند سن، جنسیت، تعداد فرزندان، وضعیت تاهل، گروه شغلی، ارتباط بیمه‌شده و بیمه‌گذار، مدت بیمه‌نامه، روش پرداخت حق بیمه، میزان حق بیمه و تعداد پوشش‌های تکمیلی با استفاده از داده‌کاوی و الگوریتم خوشه‌بندی فازی متمرکز شدند. نتایج حاکی از تقسیم‌بندی مشتریان به دو خوشه «بیمه‌نامه‌های با جزء سرمایه‌گذاری» و «بیمه‌نامه‌های پوششی و حمایتی» است. (Bash Afshar et al. 2018) به بررسی نگهداشت مشتریان بیمه‌های زندگی به منظور دستیابی به الگوی خوشه‌بندی برای ارائه خدمات پرداختند. در واقع، آن‌ها با استفاده از مدل‌های خوشه‌بندی و با استفاده از متغیرهای مربوط به اطلاعات جمعیتی بیمه‌شدگان، اطلاعات بیمه‌نامه و اطلاعات تکمیلی بیمه‌شدگان مانند سوابق بیماری، عوامل مؤثر بر ریزش مشتری و روابط بین آن‌ها را مورد بررسی قرار دادند و در نهایت، با مقایسه نتایج حاصل از مدل‌های مختلف خوشه‌بندی پیشنهاد کردند که شرکت‌های بیمه می‌توانند مشتریان بیمه زندگی را در دو گروه اصلی مشتریان سودبخش و مشتریان دارای ریسک طبقه‌بندی کنند و به هر یک از گروه‌ها، بسته‌ها و خدماتی متناسب ارائه دهند. نتایج این تحقیق نشان می‌دهد که متغیرهای جمعیتی همچون «جنسیت» و «سن» و متغیرهای بیمه‌ای همچون «حق بیمه سالیانه» و «ضریب فوت در اثر حادثه» به عنوان عوامل تأثیرگذار در شناسایی گروه‌های مشتریان هستند. (Qadadeh and Abdallah 2018) به بخش‌بندی مشتریان شرکت بیمه TIC با استفاده از الگوریتم‌های K-میانگین و نگاشت خود سازمان‌ده (SOM) پرداختند. نتایج این تحقیق نشان می‌دهد در حالی که K-میانگین، نتایج خوشه‌بندی امیدوارکننده‌ای را نشان داده است، SOM از نظر سرعت، کیفیت خوشه‌بندی و تجسم، بهتر عمل کرده است. (Groll et al. 2022) به تجزیه و تحلیل رفتار ابطال مشتریان انفرادی بیمه‌های زندگی مختلط یک شرکت بیمه آلمانی بر اساس متغیرهایی مانند حق بیمه سالانه، سرمایه بیمه و نرخ سود فنی مرتبط با قرارداد و ... بر اساس مدل‌های جنگل تصادفی، لوجیت، شبکه‌های عصبی و درخت تصمیم رگرسیون و طبقه‌بندی (CART) پرداختند. نتایج تحقیق نشان داد که رفتار ریزش قراردادهای انفرادی می‌تواند به خوبی بر اساس داده‌های موجود توضیح داده شود. هیچ یک از روش‌های طبقه‌بندی بررسی‌شده در این تحقیق، به طور قابل توجهی از روش‌های دیگر بهتر عمل نکرد یا به دقت پیش‌بینی بسیار رضایت‌بخشی دست نیافت.

مشتری محسوب می‌شود. در این مدل، یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی وجود دارد که از طریق وزن‌هایی، به یکدیگر وابسته هستند. در این روش، متغیر خروجی شامل ریزش یا عدم ریزش مشتری در بیمه زندگی و متغیرهای ورودی شامل متغیرهای مستقل تأثیرگذار بر ریزش مشتری است. تعداد متغیرهای ورودی و همچنین، تعداد لایه‌های پنهان بر اساس ماهیت داده‌ها انتخاب می‌گردند.

الگوریتم دیگر، جنگل‌های تصادفی است که یک الگوریتم یادگیری ماشین است که برای دو کاربرد طبقه‌بندی و رگرسیون قابل استفاده است. این الگوریتم با تعریف درخت‌های تصمیم فراوان، منجر به برآورد دقیق‌تری از طبقه‌بندی می‌شود. به این صورت که داده‌های اصلی به چندین زیرمجموعه تقسیم و بر روی هر کدام از زیرمجموعه‌ها، درخت تصمیم پیاده‌سازی می‌شود و در نهایت بر اساس آرای جمعی از درخت‌های تصمیم‌گیری، طبقه‌بندی به ریزش/عدم ریزش انجام می‌گردد.

### مروری بر پیشینه پژوهش

مقالات متعددی به بررسی روش‌های داده‌کاوی در بیمه‌های زندگی با نیل به اهداف متنوع پرداختند که در ادامه، برخی از آن‌ها مطرح می‌شوند. (Mehregan and Samizadeh 2012) با استفاده از الگوریتم خوشه‌بندی K-میانگین بر اساس استاندارد جهانی CRISP-DM، به ارتباط بین تعداد خریدهای مشتری و درآمد مشتری بر اساس متغیرهایی مانند نوع پرداخت، مقدار سرمایه بیمه، تعداد خریدهای بیمه‌نامه، خریدار، سن بیمه‌شده و جنسیت پرداختند. (Balaji and Srivatsa 2012) با استفاده از تکنیک‌های طبقه‌بندی از قبیل روش نیو بیز به تجزیه و تحلیل داده‌های مشتریان بیمه زندگی برای حفظ با ارزش‌ترین مشتریان بر اساس متغیرهایی مانند سن، جنسیت، تعداد فرزندان، نوع محصول، نوع طرح بیمه (پس‌اندازی/پوشش ریسک/بازنشستگی/فرزندان) پرداختند. (Allahyari Soeini and Vahidi Rodpysh 2012) با الگوریتم‌های داده‌کاوی بر روی حفظ مشتریان بیمه‌های شخص ثالث بر روی مجموعه‌ای از داده‌های هفت شعبه بیمه ایران در بندرانزلی به عنوان نمونه، متمرکز شدند که در مرحله اول، مشتریان دارای ویژگی‌های برابر را با استفاده از روش خوشه‌بندی K-میانگین انتخاب کردند و در مرحله دوم، با استفاده از شاخص چرخش و درخت تصمیم رگرسیون و طبقه‌بندی (CART)، دلایل ریزش مشتری را مورد تجزیه و تحلیل قرار دادند. (Oshini Goonetilleke and Caldera 2013) با استفاده از الگوریتم‌های داده‌کاوی مانند درختان تصمیم و شبکه‌های عصبی،

## روش شناسی پژوهش

می‌کند. ماتریس در هم‌ریختگی، نشان می‌دهد که چه تعداد پیش‌بینی صحیح و غلط برای هر کلاس، انجام شده و بر اساس این مقادیر، می‌توان معیارهای مختلف ارزیابی طبقه‌بندی و اندازه‌گیری دقت را تعریف کرد. این ماتریس به صورت **شکل ۲** در نظر گرفته می‌شود.

بر پایه این ماتریس، معیارهای مختلفی تعریف می‌گردد که تحت عناوین صحت، یادآوری، دقت و امتیاز F1 شناخته می‌شوند که در ادامه، به صورت مختصر مورد بررسی قرار می‌گیرند:

- معیار دقت: این معیار نشان‌دهنده نسبتی از مشاهدات (هم مثبت و هم منفی) است که به درستی طبقه‌بندی شده‌اند و به صورت زیر محاسبه می‌شود:

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

- معیار صحت: این معیار نشان می‌دهد اندازه درستی مدل، وقتی که نتیجه را مثبت پیش‌بینی می‌کند، چقدر است که به صورت زیر محاسبه می‌شود:

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

با توجه به اهمیت حفظ مشتری و مواردی چون طبقه‌بندی آن‌ها و پیش‌بینی الگوهای ریزش مشتری در بیمه‌های زندگی، که نقش مؤثری در تغییر الگوی تخصیص منابع فراهم می‌آورد؛ در مقاله حاضر، ضمن بررسی اطلاعات مشتریان بیمه‌های زندگی، که مربوط به بیمه‌های عمر و سرمایه‌گذاری یک شرکت بیمه ایرانی پایلوت است، عملکرد روش‌های داده‌کاوی در شناسایی طبقه ریزش مشتری مورد بررسی قرار می‌گیرد و عوامل تأثیرگذار بر ریزش مشتری شناسایی می‌شوند.

بر این اساس، از استاندارد جهانی CRISP-DM برای انجام فرایند تحقیق استفاده شده است، که ساختار اجرایی تحقیق به شرح **شکل ۱** است.

در این مقاله، با استفاده از روش‌های جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک و شبکه‌های عصبی به پیش‌بینی ریزش مشتریان در بیمه‌های زندگی یک شرکت بیمه ایرانی می‌پردازیم. کاربرد این روش‌ها در شناسایی ریزش مشتریان، در بخش مبانی نظری توضیح داده شده‌اند. به منظور ارزیابی و مقایسه این ۴ روش از معیارهای مختلفی استفاده خواهد شد. در زمینه داده‌کاوی و به طور خاص مسئله طبقه‌بندی، ماتریس درهم‌ریختگی به عنوان یک جدول‌بندی خاص امکان تجسم عملکرد یک الگوریتم را فراهم

• مشخص نمودن اهداف و نیازمندی‌های پروژه	۱. درک پروژه و فهم حوزه کاربرد
• جمع‌آوری داده‌ها برای استفاده از تحلیل اکتشافی و مشخص کردن اطلاعات اولیه برای ارزیابی داده‌ها با کیفیت و انتخاب داده‌های مفید و مورد نیاز	۲. انتخاب داده‌ها
• آماده کردن داده‌های اولیه خام به داده‌های نهایی، این داده‌ها در کلیه مراحل بعدی استفاده می‌شود	۳. آماده سازی داده‌ها
• بهینه‌سازی نتایج با انتخاب و به کار بستن تکنیک‌های مناسب و روش داده‌کاوی مناسب	۴. مدل‌سازی
• ارزیابی تکنیک‌های انتخابی در مراحل قبل	۵. ارزیابی

شکل ۱: ساختار استاندارد جهانی CRISP-DM (Chapman et al., 2000)

	پیش‌بینی مثبت (ریزش)	پیش‌بینی منفی (عدم ریزش)
کلاس مثبت (ریزش)	TP	FN
کلاس منفی (عدم ریزش)	FP	TN

شکل ۲: ماتریس درهم‌ریختگی

در ابتدا، لازم است پیش‌پردازش بر روی داده‌ها صورت گیرد که نتیجه آن، مناسب شدن داده‌ها برای انجام تحلیل‌های آماری است. این پیش‌پردازش می‌تواند در ابعاد مختلف مانند پالایش، ادغام، تبدیل و یا کاهش داده‌ها صورت گیرد. در داده‌های این تحقیق، پیش‌پردازش لازم در خصوص کاهش و یا ادغام داده‌ها صورت گرفت. متغیرهایی مانند تاریخ شروع بیمه‌نامه و تاریخ پایان بیمه‌نامه به دلیل جایگزینی با مدت بیمه‌نامه و وضعیت بیمه‌نامه حذف گردیدند.

در ادامه، آمار توصیفی این داده‌ها به منظور درک بیشتر ماهیت داده‌ها مورد بررسی قرار می‌گیرد. اطلاعات دریافتی در این شرکت بیمه، مربوط به بیمه‌نامه‌های زندگی انفرادی و بیمه‌های زندگی مشترک است که ۹۵،۶۵٪ بیمه‌نامه‌ها انفرادی و ۴،۳۵٪ مشترک هستند.

هیستوگرام داده‌های مربوط به سن بیمه‌شده، مدت زمان بیمه‌نامه زندگی، سن افزوده، جنسیت بیمه‌شده، طبقه شغلی بیمه‌شده و نحوه پرداخت حق‌بیمه در **نمودار ۱** نشان داده شده است. **نمودار ۱** نشان می‌دهد که طبقه شغلی بیمه‌شدگان به صورت ۸۹،۷۴٪ در طبقه اول، ۸،۹۲٪ در طبقه دوم، ۱،۱۵٪ در طبقه سوم، ۰،۱۹٪ در طبقه چهارم توزیع شده است. نحوه پرداخت حق‌بیمه نیز در این داده‌ها به این صورت بوده که ۲۳،۸۱٪ به صورت اقساط یک ساله، ۴۶،۸۸٪ به صورت اقساط دو ماهه، ۱۴،۱۴٪ به صورت اقساط سه ماهه، ۱۲،۰۵٪ به صورت اقساط شش ماهه، ۲،۹۸٪ به صورت اقساط یک ماهه و ۰،۱۴٪ به صورت یکجا حق‌بیمه خود را پرداخت نموده‌اند. همچنین، بر اساس **نمودار ۱**، ۷۹،۵۷ درصد بیمه‌شدگان مرد و ۲۰،۴۳ درصد بیمه‌شدگان زن هستند.

در اطلاعات دریافتی از بیمه‌نامه‌های زندگی شرکت بیمه پایلوت ۴ وضعیت وجود داشت که این وضعیت‌ها در **جدول ۲** نشان داده شده است.

بر اساس داده‌های استخراج شده از این شرکت بیمه، می‌توان دریافت که حدود ۹۲،۱۵٪ بیمه‌نامه‌ها عادی، ۳،۷٪ بیمه‌نامه‌ها بازخرد، ۲،۴٪ بیمه‌نامه‌ها پایان‌یافته و ۱،۴۰٪ بیمه‌نامه‌ها فسخ (ابطال) و ۰،۰۱٪ بیمه‌شدگان فوت شده‌اند. توزیع فراوانی سن بیمه‌شده در برابر وضعیت بیمه‌نامه در **نمودار ۲** نشان داده شده است.

همان‌طور که از **نمودار ۲** مشاهده می‌شود در هر وضعیت بیمه‌نامه، توزیع مناسبی در خصوص سن بیمه‌شده وجود دارد. نکته قابل توجه توزیع تقریباً یکنواخت مدت بیمه‌نامه در وضعیت‌های فسخ و بازخرد بیمه‌نامه‌های مورد بررسی است. به منظور کسب اطلاعات بیشتر در خصوص ماهیت داده‌ها، در **نمودار ۳**

- معیار پوشش/یادآوری: بر اساس این معیار، تعداد پیش‌بینی‌های مثبت صحیح از همه پیش‌بینی‌های مثبت که می‌توانست انجام شود، تعیین می‌شود و به صورت زیر محاسبه می‌شود:

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

- معیار امتیاز F1: این معیار به صورت میانگین هارمونیک دو معیار پوشش و صحت به صورت زیر تعریف می‌شود:

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

یکی دیگر از معیارهای ارزیابی مورد استفاده در طبقه‌بندی، بر اساس سطح زیر منحنی مشخصه عملکرد سیستم (منحنی ROC) تعریف می‌شود. منحنی ROC، عملکرد طبقه‌بندی را در کلاس مثبت نشان می‌دهد و محور افقی آن، نشان‌دهنده نرخ مثبت کاذب (FPR) و محور عمودی آن، نشان‌دهنده نرخ مثبت صحیح (TPR) است. یک مدل خوب باید تا حد امکان به گوشه بالای سمت چپ نزدیک باشد و برعکس این حالت، بدترین مدل در امتداد قطر این نمودار قرار خواهد گرفت که نشانگر پیش‌بینی تصافی مدل است. به جهت عددی سازی این نمودار، شاخصی تحت عنوان AUC تعریف می‌شود که نشان‌دهنده سطح زیر منحنی ROC است که هرچه این عدد به یک نزدیک‌تر باشد، نشان‌دهنده این است که الگوریتم طبقه‌بندی با دقت و به درستی عمل کرده است و در صورتی که این عدد مساوی ۰،۵ باشد، نشان‌دهنده این است که الگوریتم به صورت کاملاً تصادفی عمل کرده است (Saniei Abadeh et al., 2012). در کنار منحنی ROC، منحنی دیگری تحت عنوان منحنی صحت-پوشش تعریف می‌شود که مقادیر معیار صحت را در مقابل مقادیر معیار پوشش رسم می‌کند و شاخصی تحت عنوان AUCPR به صورت سطح زیر این منحنی تعریف می‌گردد. مقدار بالای این شاخص، نشان‌دهنده کارایی بالای مدل است.

## نتایج و بحث

جامعه آماری این تحقیق مربوط به بیمه‌نامه‌های عمر جامع (عمر و سرمایه‌گذاری) یکی از شرکت‌های بیمه ایرانی است که اطلاعات آن‌ها در سال ۱۳۹۸ از پایگاه داده شرکت بیمه با ۱۶۳۱۲ رکورد استخراج شده است. از آنجا که داده‌های بیشتری در داده‌کاوی مناسب‌ترند، هیچ روش نمونه‌گیری انجام نشده و اندازه نمونه با کل جامعه آماری تحقیق برابر است. در این تحقیق از نرم‌افزار پایتون برای تجزیه و تحلیل داده‌ها استفاده شده است. **جدول ۱** به ارائه متغیرهای مورد استفاده در تحقیق می‌پردازد.

جدول ۱: متغیرهای مورد استفاده در تحقیق

ردیف	نام متغیر
۱	وضعیت بیمه‌نامه: دارای ۴ حالت عادی، ابطال، پایان پوشش، بازخريد
۲	سن بیمه‌شده/سن بیمه‌گذار
۳	تاریخ شروع بیمه‌نامه
۴	حق بیمه اولیه: کمترین مقدار ۱۵۰۰۰۰۰ و بیشترین مقدار ۶۰۰۰۰۰۰۰۰
۵	مدت بیمه‌نامه: از ۵ سال تا ۳۰ سال
۶	تاریخ پایان بیمه‌نامه
۷	جنسیت بیمه‌شده/بیمه‌گذار: ۲-حالت مرد یا زن
۸	تعداد بیمه شده: ۱ یا ۲ نفره شرکت بیمه در نظر گرفته شده ۲ نوع بیمه نامه دارد: بیمه نامه انفرادی و بیمه‌نامه مشترک
۹	روش پرداخت حق بیمه: دارای ۶ حالت اقساط ماهانه، ۲ ماهه، ۳ ماهه، ۶ ماهه، سالانه، یکجا
۱۰	منظور کردن پوشش‌ها در حق بیمه اولیه: دارای ۲ حالت بله، خیر
۱۱	درصد ضریب تغییر سرمایه: دارای ۴ حالت ۵، ۱۰، ۱۵، ۲۰ درصد
۱۲	درصد ضریب تغییر حق بیمه: دارای ۵ حالت ۵، ۱۰، ۱۵، ۲۰ درصد
۱۳	اندوخته اولیه: از صفر تا ۴۴۰ میلیون
۱۴	نوع و نقش بیمه‌شده: دارای ۳ حالت بیمه‌شده اصلی همان بیمه‌گذار، بیمه‌شده اصلی غیر بیمه‌گذار، بیمه‌شده فرعی
۱۵	سرمایه فوت: از ۲۰۰۰۰۰۰ تا ۳۷۵۰۰۰۰۰۰
۱۶	ضریب فوت ناشی از حادثه: دارای ۴ حالت خیر، یک برابر، دو برابر، سه برابر
۱۷	ضریب نقص عضو ناشی از حادثه: دارای ۳ حالت خیر، یک برابر سرمایه فوت به هر علت، نیم برابر سرمایه فوت به هر علت
۱۸	معافیت ناشی از «از کارافتادگی به هر علت»: دارای ۲ حالت بله، خیر
۱۹	درصد اضافه نرخ پزشکی: از ۰ تا ۲۰
۲۰	سن افزوده (سن بیمه‌شده در بدو ورود به اضافه مدت زمان سپری شده از بیمه‌نامه): از ۰ تا ۶۵ سال
۲۱	طبقه شغلی شغل بیمه شده: دارای ۴ حالت طبقه یک، طبقه دو، طبقه سه، طبقه چهار هرچه از طبقه شغلی ۱ به سمت ۴ حرکت می‌کند، میزان خطر در شغل افزایش می‌یابد. طبقه شغلی ۱ و ۲ طبقه شغلی کم ریسک و طبقه شغلی ۳ و ۴ طبقه شغلی پرریسک محسوب می‌شوند.

داده‌کاوی، در ابتدا لازم است کلاس‌های در نظر گرفته شده متوازن شوند. با توجه به اینکه تعداد بیمه‌نامه‌های قرار گرفته شده در کلاس مثبت کم است بایستی بیش‌نمونه‌گیری صورت گیرد که با چندین روش، این کار قابل انجام است. یکی از این روش‌ها، استفاده از روش SMOTENC است که داده‌های کلاس مثبت در کل نمونه تکرار می‌شوند. در این تحقیق، از این روش جهت متوازن سازی کلاس‌ها استفاده شده است.

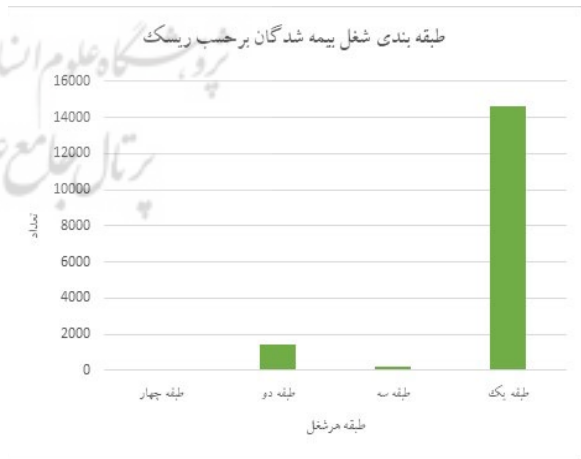
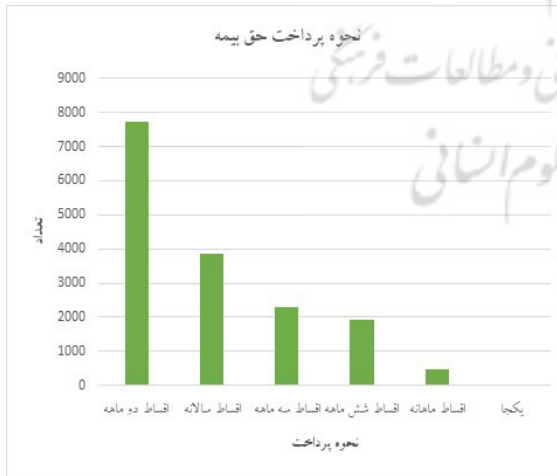
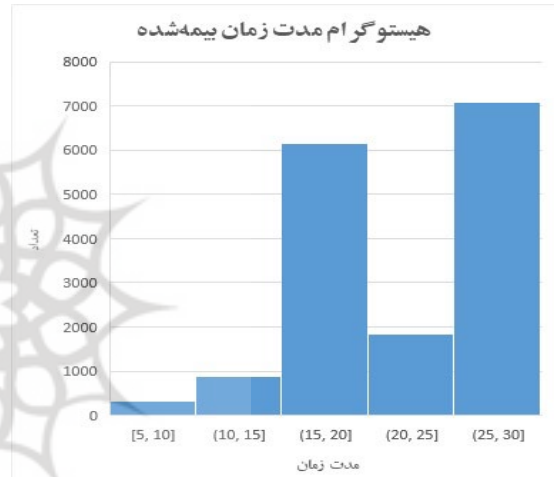
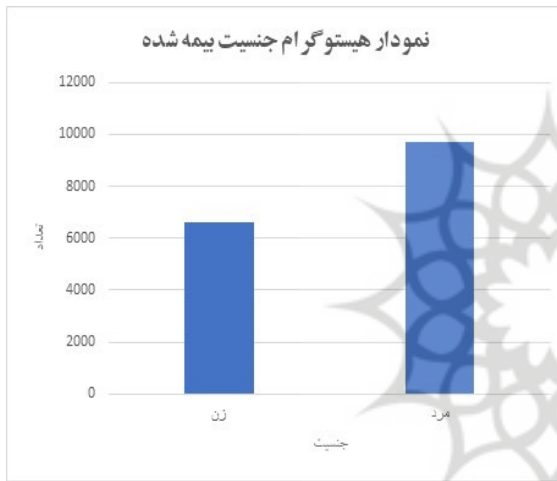
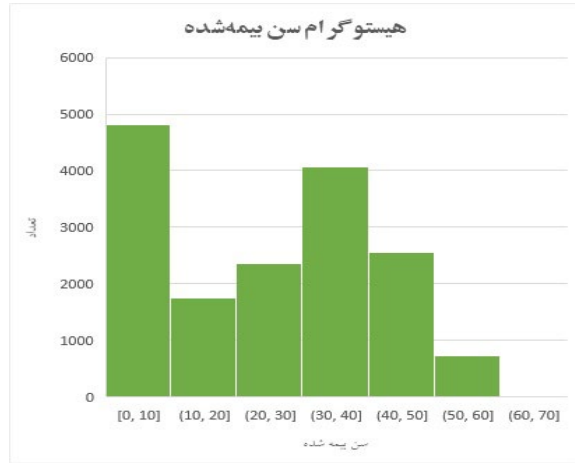
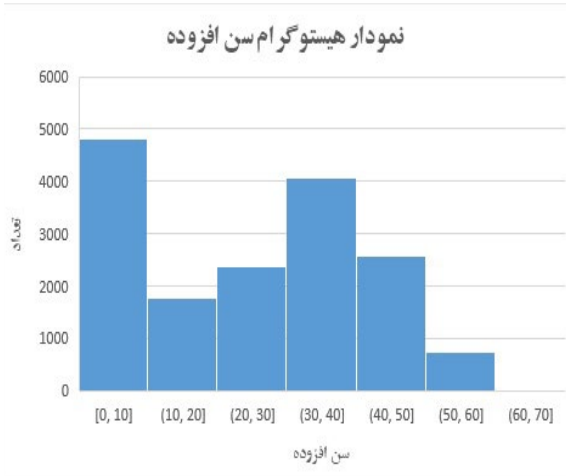
در ادامه، به منظور پیش‌بینی دقیق کلاس‌های مذکور، از روش داده‌کاوی رگرسیون لجستیک، درخت تصمیم‌گیری، شبکه‌های عصبی و جنگل تصادفی استفاده شده است. برای ارزیابی الگوریتم‌های مورد استفاده در تحقیق، از ماتریس درهم‌ریختگی و

در سمت راست توزیع فراوانی سن بیمه‌شده و نوع پرداخت حق بیمه در برابر وضعیت بیمه‌نامه زندگی و در سمت چپ توزیع فراوانی سن بیمه‌شده و طبقه شغلی بیمه‌شده (که در شرکت پایلوت، ۴ طبقه برای شغل در نظر گرفته شده است) در برابر وضعیت بیمه‌نامه زندگی نشان داده شده است.

با توجه به هدف تحقیق، که پیاده‌سازی روش‌های داده‌کاوی در پیش‌بینی طبقه مربوط به ریزش مشتریان بیمه‌های زندگی است، افرادی را که بیمه‌نامه‌های خود را بازخريد یا ابطال نموده‌اند در کلاس مثبت و بقیه بیمه‌نامه‌ها را در کلاس منفی قرار می‌دهیم. با توجه به نامتوازن بودن تعداد بیمه‌نامه‌های قرار گرفته شده در کلاس‌های مثبت و منفی، به منظور پیاده‌سازی روش‌های



ریزش مشتریان بیمه‌های زندگی

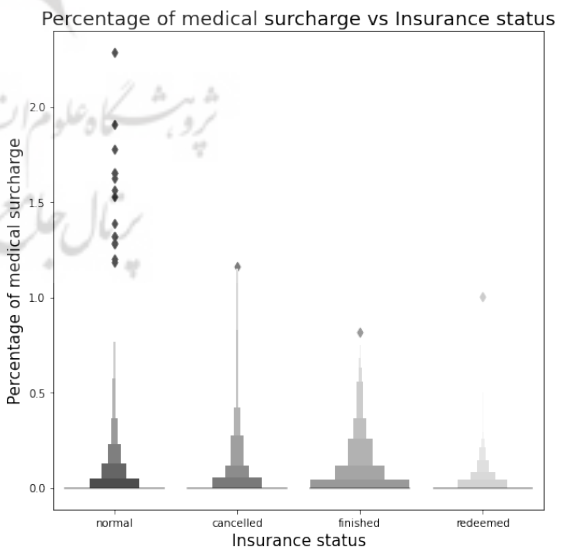
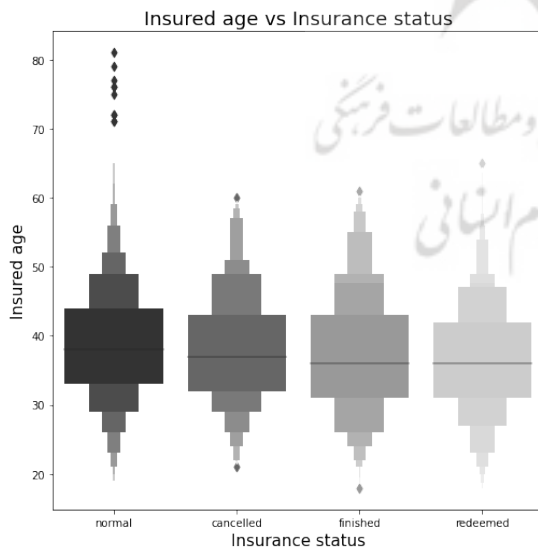
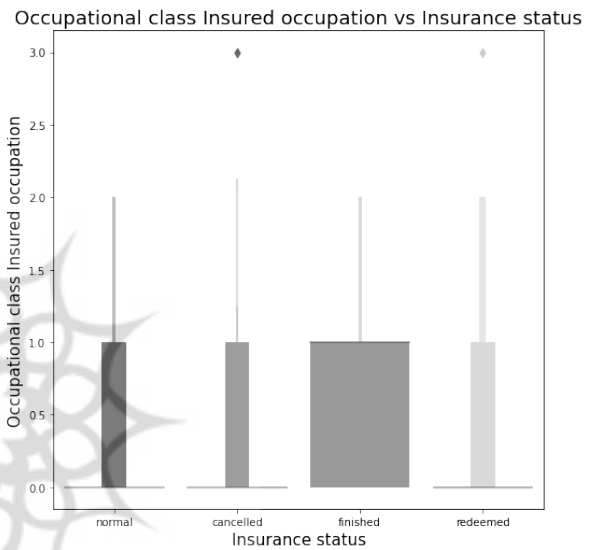
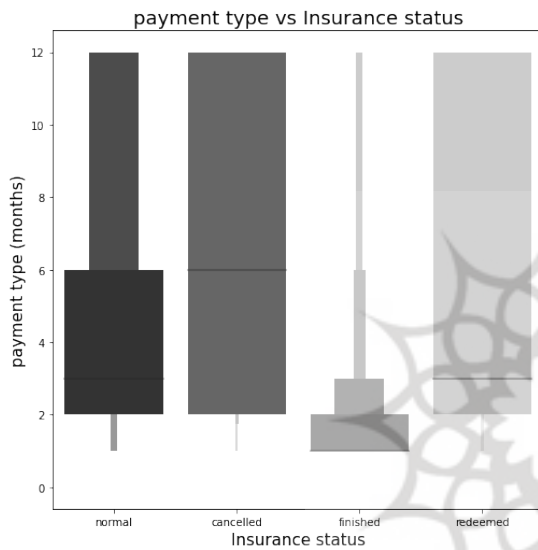


توضیح: با توجه به تعداد زیاد متغیرها، تنها هیستوگرام برخی از متغیرها رسم شده است.

نمودار ۱: هیستوگرام برخی از متغیرهای تحقیق

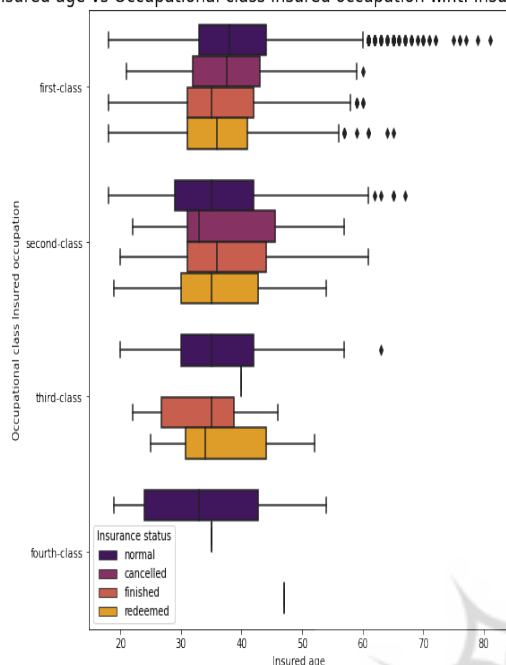
جدول ۲: وضعیت‌های مختلف در بیمه‌نامه‌های زندگی

تعریف	وضعیت بیمه‌نامه
وقتی مشتری با موفقیت در پرداخت حق بیمه در حالت توافق شده ادامه می‌دهد.	عادی
قبل از پایان بیمه‌نامه، پرداخت حق بیمه توسط مشتری متوقف و مشتری به قرارداد بیمه‌نامه پایان می‌دهد.	بازخرید
اگر قبل از اتمام سال اول، بیمه‌شده از خرید بیمه‌نامه خود منصرف گردد، می‌تواند درخواست کتبی خود را مبنی بر ابطال بیمه‌نامه به بیمه‌گر تحویل نماید یا در صورتی که بیمه‌نامه به حالت تعلیق درآمد باشد و از تاریخ معلق شدن بیمه‌نامه ۳ ماه گذشته اقدامی در جهت پرداخت حق بیمه‌های معوق و رفع تعلیق صورت نگرفته باشد، بیمه‌نامه خود به خود باطل می‌گردد.	فسخ (ابطال)
اگر تمام حق بیمه‌ها برای کل دوره، همانطور که در قرارداد مشخص شده، پرداخت و تعهدات بیمه‌گر اجرا شود به معنای پایان بیمه‌نامه است.	اتمام

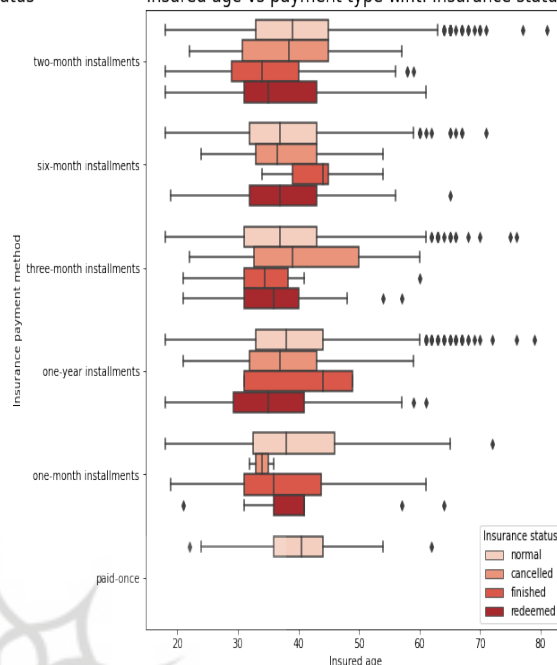


نمودار ۲: توزیع فراوانی نوع پرداخت حق بیمه، طبقه شغلی بیمه‌شده، سن بیمه‌شده و درصد اضافه نرخ پزشکی در برابر وضعیت بیمه‌نامه زندگی

Insured age vs Occupational class Insured occupation w.r.t. Insurance status



Insured age vs payment type w.r.t. Insurance status



نمودار ۳: توزیع فراوانی سن بیمه شده و نوع پرداخت حق بیمه در برابر وضعیت بیمه‌نامه زندگی (سمت راست) و توزیع فراوانی سن بیمه شده و طبقه شغلی بیمه شده در برابر وضعیت بیمه‌نامه زندگی (سمت چپ)

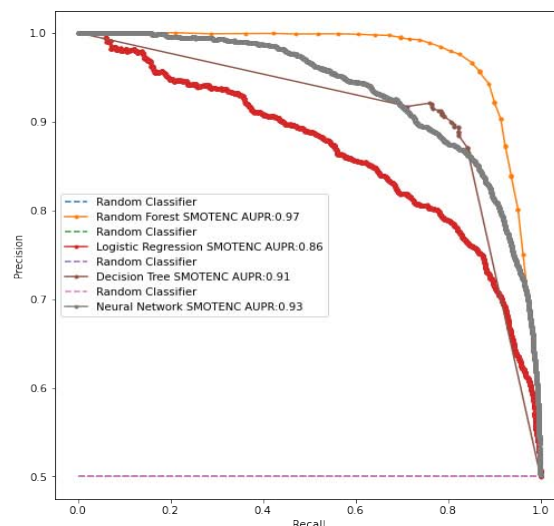
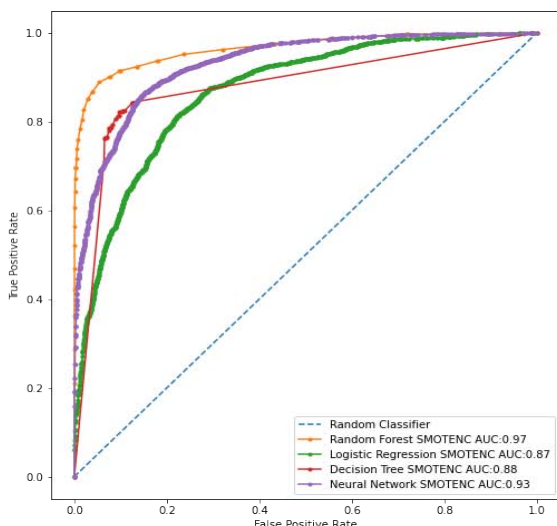
جدول ۳: ارزیابی الگوریتم‌های مورد بررسی براساس معیارهای ارزیابی ماتریس درهم ریختگی

روش	Precision	Recall	Accuracy	F1-score
جنگل تصادفی	۰٫۹۹	۰٫۷۶	۰٫۹	۰٫۸۶
درخت تصمیم	۰٫۹۰	۰٫۸۱	۰٫۸۶	۰٫۸۵
رگرسیون لجستیک	۰٫۷۷	۰٫۸۳	۰٫۷۹	۰٫۸۰
شبکه عصبی	۰٫۸۹	۰٫۷۷	۰٫۸۴	۰٫۸۲

همان‌طور که در بخش ۳ به آن اشاره شد، یکی دیگر از معیارهای ارزیابی مورد استفاده در طبقه‌بندی استفاده از منحنی مشخصه عملکرد سیستم است که عملکرد طبقه‌بندی در کلاس مثبت را نشان می‌دهد. هر چه این نمودار به گوشه سمت چپ نزدیک‌تر باشد، نشان‌دهنده مطلوبیت مدل در طبقه‌بندی است. در سمت چپ نمودار ۴، منحنی مشخصه عملکرد سیستم برای الگوریتم‌های جنگل، درخت تصمیم‌گیری، رگرسیون لجستیک و شبکه عصبی نشان داده شده است. در سمت راست این نمودار نیز، منحنی صحت-پوشش ترسیم شده که نزدیک بودن نمودار

منحنی مشخصه عملکرد سیستم استفاده شده که نتایج معیارهای ارزیابی بر اساس ماتریس درهم‌ریختگی در جدول ۳ نشان داده شده است.

براساس نتایج نمایش داده شده در جدول ۳، صحت‌داده‌ها از قدرت نسبتاً خوب مدل در تشخیص درست داده‌های مثبتی موجود در کلاس مثبت (بازخرید و ابطال) خبر می‌دهد. در میان روش‌های مورد بررسی، بر اساس معیارهای صحت و امتیاز F1، الگوریتم جنگل تصادفی بهترین عملکرد و بر اساس معیار یادآوری، درخت تصمیم بهترین عملکرد را دارا هستند.



نمودار ۴: منحنی مشخصه عملکرد سیستم (سمت چپ) و منحنی صحت-پوشش (سمت راست) برای الگوریتم‌های مورد استفاده در طبقه‌بندی بازخرید بیمه‌نامه‌های زندگی شرکت بیمه پایلوت

جدول ۴: ارزیابی الگوریتم‌های مورد بررسی براساس معیار سطح زیر منحنی ROC و AUCPR

AUCPR	AUC	روش
۰٫۹۷	۰٫۹۷	جنگل تصادفی
۰٫۹۱	۰٫۸۸	درخت تصمیم
۰٫۸۶	۰٫۸۷	رگرسیون لجستیک
۰٫۹۳	۰٫۹۳	شبکه عصبی

نمودار ۵ به ارائه این نمودار برای برخی از متغیرهای تحقیق برای بهترین روش از لحاظ عملکرد (جنگل تصادفی) پرداخته است. بر اساس نمودار ۵، اطلاعات زیر از مشاهدات و تحلیل صورت گرفته برداشت می‌شود:

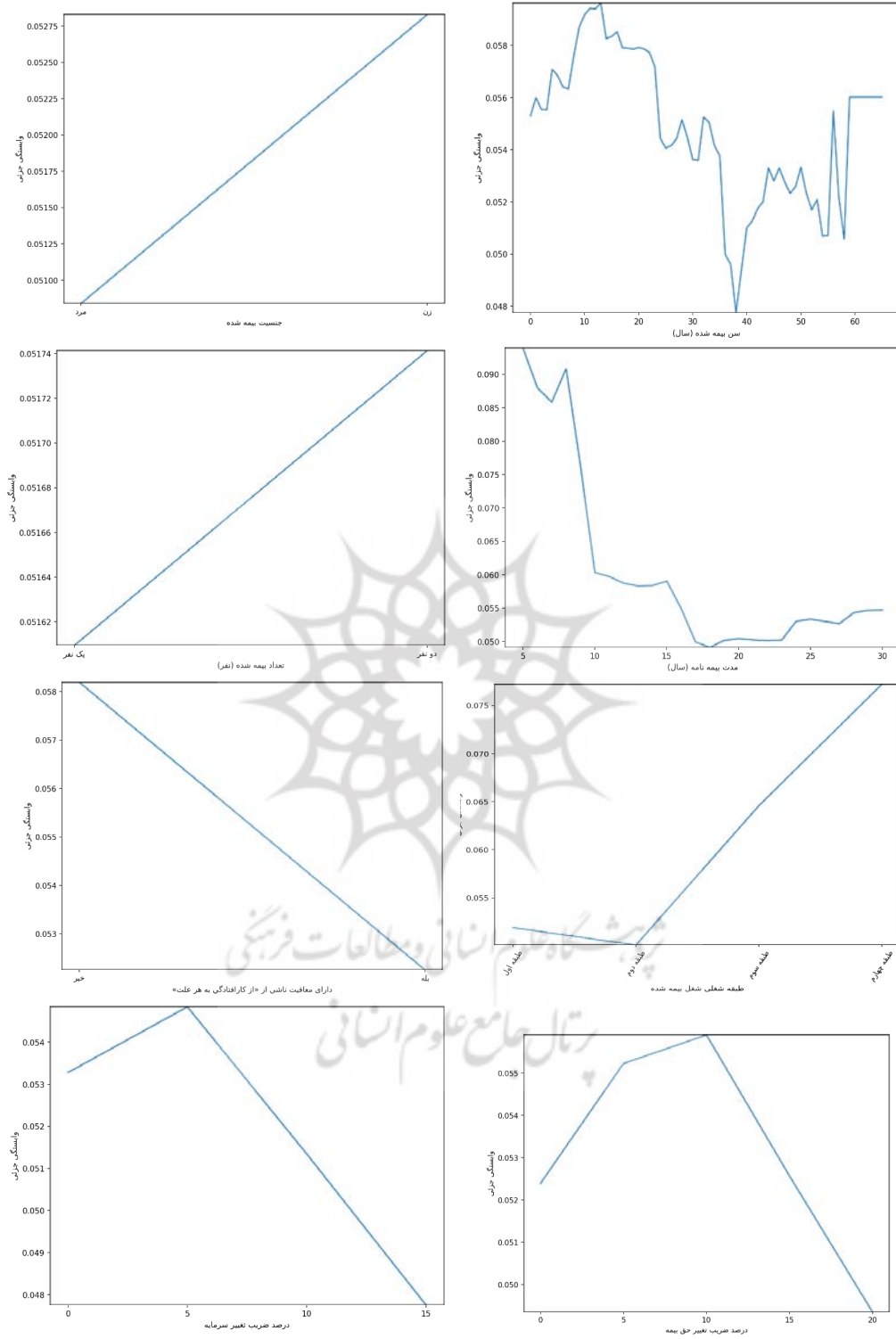
- احتمال ریزش در زنان بیشتر از مردان است.
- با افزایش سن بیمه‌شده تا ۳۹ سالگی، احتمال ریزش کاهش و از سن ۳۹ سالگی به بعد احتمال ریزش افزایش می‌یابد.
- با افزایش مدت بیمه‌نامه تا ۱۸ سال، احتمال ریزش کاهش و پس از آن با توجه به طولانی مدت شدن بیمه‌نامه احتمال ریزش افزایش یافته است.
- در صورتی که بیمه‌نامه دارای معافیت از کارافتادگی باشد احتمال ریزش نسبت به حالت عدم داشتن این ویژگی، کاهش می‌یابد.

به گوشه بالای سمت راست، نشان‌دهنده کارایی بالای مدل است. همان‌طور که از نمودار ۴ مشاهده می‌شود مدل‌های مورد بررسی کارایی خوبی در شناسایی کلاس مثبت (بازخرید و ابطال) که معنای ریزش مشتری است، دارند.

در جدول ۴ به ترتیب مقدار سطح زیر نمودار مشخصه عملکرد سیستم (AUC) و سطح زیر منحنی صحت-پوشش (AUCPR) برای هر کدام از الگوریتم‌های استفاده شده نشان داده شده است که نشان از کارایی بالای الگوریتم‌های مورد بررسی در پیش‌بینی کلاس مثبت (ریزش) دارد.

در ادامه، به منظور بررسی نحوه تأثیرگذاری متغیرها بر احتمال ریزش از نمودار وابستگی جزئی استفاده می‌کنیم. نمودار وابستگی جزئی نشان‌دهنده تأثیرات حاشیه‌ای متغیرهای مورد استفاده در روش داده‌کاوی بر متغیر خروجی (ریزش مشتری) را نشان می‌دهد.

## حسن قربانی و همکاران



نمودار ۵: نمودار وابستگی جزئی برای برخی از متغیرهای تحقیق در طبقه‌بندی انجام شده توسط روش جنگل تصادفی

صورت گذشت زمان مناسبی از بیمه‌نامه، به منظور حفظ مشتری، تخفیف‌هایی برای ادامه پرداخت‌های حق بیمه و یا تمدید یا به‌روزرسانی بیمه‌نامه اعمال شود. در واقع شرکت بیمه، می‌بایست در طول مدت بیمه‌نامه تعاملات پیوسته با بیمه‌گذار داشته باشد و خدمات جدیدی را به مرور به وی پیشنهاد دهد. با توجه به اینکه بر اساس تحلیل صورت گرفته بر داده‌های مورد بررسی در تحقیق، افرادی که ضریب افزایش حق بیمه و یا سرمایه بیمه آن‌ها بیشتر از ۵ درصد است دارای احتمال ریزش کمتری هستند، شرکت بیمه می‌تواند خدمات اختصاصی برای افراد دیگر به منظور ماندگاری آن‌ها در نظر گیرد.

### مشارکت نویسندگان

این مقاله مستخرج از پایان‌نامه کارشناسی ارشد حسن قربانی است که تحت راهنمایی دکتر رضا افقی و دکتر میترا قنبرزاده تهیه و تدوین شده است. گردآوری داده‌ها و تحلیل اولیه آن‌ها در مقاله، بر عهده نویسنده اول، ویرایش و اصلاحات علمی، تحلیل نهایی مقاله بر عهده نویسندگان دوم و سوم و مسئولیت اصلاحات و ویرایش نهایی مقاله بر عهده نویسنده مسئول است.

### تشکر و قدردانی

این پژوهش با حمایت مالی پژوهشکده بیمه انجام شده است. نویسندگان مقاله از پژوهشکده بیمه برای حمایت از این پژوهش، و همچنین، تسهیل در اختیار قرار دادن داده‌های موردنیاز تشکر و قدردانی می‌کنند.

### تعارض منافع

نویسندگان اعلام می‌دارند که در مورد انتشار این مقاله تضاد منافع وجود ندارد. علاوه بر این، موضوعات اخلاقی شامل سرقت ادبی، رضایت آگاهانه، سوءرفتار، جعل داده‌ها، انتشار و ارسال مجدد و مکرر توسط نویسندگان رعایت شده است.

### دسترسی آزاد

کپی‌رایت نویسنده(ها) ©2022: این مقاله تحت مجوز بین‌المللی Creative Commons Attribution 4.0 اجازه استفاده، اشتراک‌گذاری، اقتباس، توزیع و تکثیر را در هر رسانه یا قالبی مشروط به درج نحوه دقیق دسترسی به مجوز CC منوط به ذکر تغییرات احتمالی بر روی مقاله می‌باشد. لذا به استناد مجوز مذکور، درج هرگونه تغییرات در تصاویر، منابع و ارجاعات یا سایر مطالب از اشخاص ثالث در این مقاله باید در این مجوز گنجانده شود، مگر اینکه در راستای اعتبار مقاله به اشکال دیگری مشخص شده باشد.

- احتمال ریزش در صورتی که اقساط حق بیمه، دوماهه پرداخت شوند، نسبت به پرداخت اقساط حق بیمه در سایر دوره‌های ماهانه و سالانه افزایش می‌یابد.
- احتمال ریزش در بیمه‌نامه‌های پوشش دهنده دو نفر در بیمه‌نامه، نسبت به حالت انفرادی بالاتر است.
- احتمال ریزش در طبقه‌های پرریسک شغل بیمه‌گذار بالاتر است.

### جمع‌بندی و پیشنهادها

برای افزایش ضریب نفوذ بیمه، دو رویکرد اصلی وجود دارد که رویکرد اول، شامل، افزایش فروش و دیگری، شامل حفظ بیمه‌نامه‌های موجود است. در دهه‌های اخیر، با توجه به آن‌که هزینه‌های جذب مشتری جدید به طور پیوسته در حال افزایش است، توجه به حفظ و نگهداری مشتریان و بالا بردن وفاداری آن‌ها، برای سودآوری شرکت‌ها بسیار مهم و حساس است و از این رو شرکت‌های بیمه، برنامه‌های مختلفی را برای حفظ مشتریان خود اجرا می‌کنند (Akhundzadeh et al., 2014). یکی از اقدامات شرکت‌های بیمه در این حوزه، می‌تواند استفاده از روش‌های داده‌کاوی به جهت شناسایی مشتریان در حال ریزش و به‌کارگیری اقداماتی به جهت حفظ آن مشتری در شرکت بیمه باشد. بر این اساس، در این مقاله، برپایه اطلاعات بیمه‌نامه‌های زندگی یکی از شرکت‌های بیمه فعال در کشور، با استفاده از روش‌های داده‌کاوی به پیش‌بینی کلاس مربوط به ریزش مشتریان پرداختیم. الگوریتم‌های مورد استفاده شامل جنگل‌های تصادفی، شبکه‌های عصبی، رگرسیون لجستیک و درخت تصمیم‌گیری است که کارایی آن‌ها با استفاده از معیارهای مختلف مورد ارزیابی قرار گرفت و جنگل تصادفی، تقریباً عملکرد بهتری را نسبت به سایر روش‌ها از خود نشان داد. از نتایج اصلی در به‌کارگیری این روش‌ها، می‌توان به بالاتر بودن احتمال بازخرید یا ابطال در زنان نسبت به مردان، کاهش بازخرید یا ابطال با افزایش سن بیمه‌شده، افزایش بازخرید با افزایش ریسک مشاغل بیمه‌شدگان اشاره نمود. در این راستا، شرکت بیمه بایستی اقداماتی را در خصوص نگهداری مشتری در بیمه‌نامه انجام دهد که برخی از آن‌ها در ادامه، مطرح می‌شوند.

با توجه به بالا بودن احتمال ریزش در زنان، پیشنهاد می‌شود برنامه‌های وفاداری مخصوص زنان در نظر گیرند تا ماندگاری آن‌ها در بیمه‌نامه بیشتر شود. همچنین، با توجه به تأثیرگذاری سن در افزایش احتمال ریزش، پیشنهاد می‌شود در کنار ارائه پاداش‌های وفاداری برای بهبود وضعیت سلامت فرد، پیشنهاد خرید یک بیمه‌نامه عمر زمانی در کنار پوشش مادام‌العمر با تخفیف به منظور نگهداشت مشتری نیز مطرح گردد. همچنین، می‌توان در

HelmZadeh, A.; Hamidi, K.; Heidarzadeh Hanzaei, K., (2019a). Investigating factors affecting the tendency to life insurance surrender. *Iran. J. Insur. Res.*, 9(2): 67-98 (32 pages). (In Persian)

Helmzadeh, A., Hamdi, K., Heidarzadeh Hanzaei, K. (2019b). Paradigm model of factors affecting the surrender of life insurance in Iran. *Iran. J. Insur. Res.*, 8(3): 63-95 (33 pages). (In Persian)

Jandaghi, G.; Moradpour, Z., (2015). Segmentation of life insurance customers based on their profile using fuzzy clustering. *Int. Lett. Soc. Hum. Sci.*, 61: 17-24 (8 pages).

Khalilzadeh Talatapeh, M., Nasehifar, V., Ghobadi Lamuki, T., Asghari Sarem, A. (2022). Analyzing factors affecting customer confusion in life insurance services using mixed method. *Iran. J. Insur. Res.*, 11(3): 444-458 (15 pages). (In Persian)

Mahdavi, G., Ofoghi, R., Abed, M. (2015). The impact of risk aversion on surrender of life insurance policies case study (Iranian life insurance market). *Iran. J. Insur. Res.*, 4(3): 61-75 (15 pages). (In Persian)

Mahdavi K., G.; Majed, V., (2017). The impact of socio-economic and psychological factors on life insurance demand in IRAN. *J. Econ. Model. Res.*, 2(5): 21-46 (26 pages). (In Persian)

Mahmoudi Maymand, M., Hossiney, M. H., Parhizgar, M. M., Azizi, I. (2022). Formulation and explanation of factors affecting the attraction of reference customers in the insurance industry, Iran. *J. Insur. Res.*, 11(2): 75-111 (37 pages). (In Persian)

Mehregan, S.; Samizadeh, R., (2012). Customer retention based on the number of purchases: A data mining approach. *Int. J. Manage. Bus. Res.*, 2(1): 41-50 (10 pages).

Motarjem, K.; Niakan, L., (2020). 'Measuring and evaluating the satisfaction of life insurance customers', *Iran. J. Insur. Res.*, 10(1): 87-119 (33 pages).

Oshini Goonetilleke, T. L.; Caldera, H. A., (2013). Mining life Insurance data for customer attrition analysis. *J. Ind. Intell. Inf.*, 1(1): 52-58 (7 pages).

Paulin, M.; Perrien, J.; Ferguson, R.J.; Salazar, A.M.A.; Seruya, L.M., (1998). Relational norms and client retention: External effectiveness of commercial banking in Canada and Mexico. *Int. J. Bank Marketing*, 16(1): 24-31 (8 pages).

Qadadeh, W.; Abdallah, S., (2018). Customers segmentation in the insurance company (TIC) Dataset. *Procedia Comput. Sci.*, 144: 277-290 (24 pages).

Qarahnejad, S., (2009). The need to maintain insurance customers using data mining tools. *News from the Insurance world*, 150-151: 15-23 (9 pages). (In Persian)

Sanii Abadeh, M.; Mahmoudi, S.; Tahripour, M., (2012). *Applied Data Mining*. Tehran: Niaz Danesh Publications. (In Persian)

Reck, L.; Schupp, J.; Reuß, A., (2022). Identifying the determinants of lapse rates in life insurance: an automated Lasso approach. *Eur. Actuar. J.*

Yang, X. S., (2019). *Introduction to Algorithms for Data Mining and Machine Learning*. Academic Press.

در صورت عدم درج مطالب مذکور و یا استفاده فراتر از مجوز فوق، نویسنده ملزم به دریافت مجوز حق نسخه‌برداری از شخص ثالث می‌باشد.

به منظور مشاهده مجوز بین‌المللی Creative Commons Attribution 4.0 به آدرس زیر مراجعه گردد:

<https://creativecommons.org/licenses/by/4.0/>

### یادداشت ناشر

ناشر نشریه پژوهشنامه بیمه با توجه به مرزهای حقوقی در نقشه‌های منتشر شده بی طرف باقی می‌ماند.

### منابع

Akhundzadeh Noghabi, E.; Al-Badawi, A.; Aghdasi, M., (2014). Exploring customer dynamics in segmentation design using data mining methods. *Inf. Technol. Manage.*, 6(1): 1-30 (30 pages). (In Persian)

Allahyari Soeini, R.; Vahidy Rodpysh, K., (2012). Applying data mining to insurance customer churn management. *Int. Proc. Comput. Sci. Inf. Technol.*, 30: 82-92 (11 pages).

Azzone, M.; Barucci, E.; Moncayo, G. G.; Marazzina, D., (2022). A machine learning model for lapse prediction in life insurance contracts. *Expert Systems with Applications*, 191, 116261.

Balaji, S.; Srivatsa, K., (2012). Naïve bayes classification approach for mining life insurance databases for effective prediction of customer preferences over life Insurance products. *Int. J. Comput. Appl.*, 51(3): 22-26 (5 pages).

Bash Afshar, M.; SaeedPanah, M.; Tireh Eidouzhi, F., (2018). Clustering pattern of life insurance customers (a case study of an insurance company). *Iran. J. Insur. Res.*, 7(2): 45-64 (10 pages). (In Persian)

Central Insurance of the Islamic Republic of Iran, (2020). *Statistical Yearbook of the Insurance Industry*. Tehran: Central Insurance of the Islamic Republic of Iran. (In Persian)


Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C., (2000). *CRISP-DM 1.0: Step-by-step data mining guide*, chapter (2)1: USA: Crispdm consortium: 13-33 (21 pages).

Ganesh, J.; Arnold M.J.; Reynolds, K.E., (2000). Understanding the customer base of service providers: An examination of the differences between switchers and stayers. *J. Marketing*, 64(3): 65-87 (23 pages).

Groll, a.; Wasserfuhr, c.; Zeldin, I., (2022). Churn modeling of life insurance policies via statistical and machine learning methods: analysis of important features. *arXiv*: 1-35 (35 pages).

Han, J.; Kamber, M.; Pei, J., (2012). *Data Mining: Concepts and techniques*. Morgan kaufmann series in data manage. Syst.

AUTHOR(S) BIOSKETCHES	معرفی نویسندگان
<ul style="list-style-type: none"><li>Email: <a href="mailto:hassan.ghobani87@gmail.com">hassan.ghobani87@gmail.com</a></li><li>ORCID: 0000-0003-2835-6933</li><li>Homepage: <a href="https://atu.ac.ir/fa">https://atu.ac.ir/fa</a></li></ul>	حسن قربانی، کارشناسی ارشد بیم‌سنجی، مؤسسه آموزش عالی بیمه اکو، دانشگاه علامه طباطبائی، تهران، ایران
<ul style="list-style-type: none"><li>Email: <a href="mailto:ghanbarzadeh@irc.ac.ir">ghanbarzadeh@irc.ac.ir</a></li><li>ORCID: 0000-0003-3486-1338</li><li>Homepage: <a href="https://www.irc.ac.ir/ghanbarzadeh">https://www.irc.ac.ir/ghanbarzadeh</a></li></ul>	میترا قنبرزاده، دکتری آمار، استادیار گروه پژوهشی بیمه‌های اشخاص، پژوهشکده بیمه، تهران، ایران
<ul style="list-style-type: none"><li>Email: <a href="mailto:ofoghi@atu.ac.ir">ofoghi@atu.ac.ir</a></li><li>ORCID: 0000-0002-0609-2888</li><li>Homepage: <a href="https://simapp.atu.ac.ir/3669">https://simapp.atu.ac.ir/3669</a></li></ul>	رضا افقی، دکتری اقتصاد بیمه، استادیار مؤسسه آموزش عالی بیمه اکو، دانشگاه علامه طباطبائی، تهران، ایران

HOW TO CITE THIS ARTICLE	
<p>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). <i>Iran. J. Insur. Res.</i>, 11(4): 305-320.</p> <p>DOI: 10.22056/ijir.2022.04.04</p> <p>URL: <a href="https://ijir.irc.ac.ir/article_155351.html?lang=en">https://ijir.irc.ac.ir/article_155351.html?lang=en</a></p>	

شپړو، ښکته علمون انساني و مطالعات فرمېځني  
پرتال جامع علمون انساني