



## An Intelligent Method to Identify Effective Factors in Customers' Dissatisfaction in the Insurance Industry Using Ensemble Learning

Kamran Balani <sup>1</sup>, Hossein Sadr <sup>2</sup>, Syed Ahmad Edalatpanah <sup>3</sup> and Mahnaz Manteghipour <sup>4</sup>

1. Msc. Computer Engineering, Department of Computer Engineering, Ayandegan Institute of Higher Education, Tonekabon, Iran. Email: [Balani\\_kamran@yahoo.com](mailto:Balani_kamran@yahoo.com)
2. Corresponding Author, Msc. Computer Engineering, Department of Computer Engineering and Information Technology, Rahbord Shomal Institute of Higher Education, Rasht, Iran. Email: [Sadr@qiau.ac.ir](mailto:Sadr@qiau.ac.ir)
3. Associate prof. Computer Engineering, Department of Mathematics, Ayandegan Institute of Higher Education, Tonekabon, Iran. Email: [Saedalatpanah@aihe.ac.ir](mailto:Saedalatpanah@aihe.ac.ir)
4. Assistance prof. Computer Engineering, Data mining desk leader, Insurance Research Center, Tehran, Iran. Email: [manteghipour@irc.ac.ir](mailto:manteghipour@irc.ac.ir)

### Article Info

**Article type:**  
Research Article

### Article history:

Received 2023 October 27  
Received in revised form 2023  
November 20  
Accepted 2023 January 18  
Published online 2024 March 15

### Keywords:

Customer Dissatisfaction,  
Data Mining,  
Insurance Industry,  
Machine Learning,  
Marketing Strategies.

### ABSTRACT

Given the competitive market of the insurance industry, customer retention is one of the most important goals of insurance brokers. Attracting a new customer, establishing a relationship with the insurer, and gaining trust requires much money. However, attracting new customers costs much more than retaining existing customers. Accordingly, marketing strategies have shifted from product-oriented, and many companies have turned to customer relationship management. This paper aims to use data mining methods to investigate and identify the factors affecting customer turnover in the insurance industry. Based on the results of experiments, the customer attraction channel is the most important factor affecting customer dissatisfaction, while purchase history and place of insurer are other influential factors. Companies and organizations have found that retaining their current customers as their most valuable asset is vital. Therefore, the strategy of insurance companies is first to retain existing customers and then attract new customers. In this regard, identifying the influential factors in customer turnover is essential. This paper uses data mining methods to predict the factors affecting customer dissatisfaction. Based on the empirical results, it has been determined that the customer attraction channel, purchase history, and insurer's place are essential factors affecting customer dissatisfaction in the insurance industry.

**Cite this article:** Balani, K., Sadr, H., Edalatpanah, S.A. & Manteghipour, M. (2023). An Intelligent Method to Identify Effective Factors in Customers' Dissatisfaction in the Insurance Industry Using Ensemble Learning. *Engineering Management and Soft Computing*, 9 (2). 91-105. DOI: <https://doi.org/>



© The Author(s)  
DOI: <https://doi.org/>

Publisher: University of Qom

## ارائه یک روش هوشمند به منظور شناسایی فاکتورهای موثر در رویگردانی مشتریان در صنعت بیمه با استفاده از تکنیک یادگیری جمعی

کامران بالانی<sup>۱</sup>، حسین صدر<sup>۲</sup>، سید احمد عدالت پناه<sup>۳</sup> و مهناز منطقی پور<sup>۴</sup>

۱. دانشجوی کارشناسی ارشد مهندسی کامپیوتر مؤسسه آموزش عالی آیندگان، گروه مهندسی کامپیوتر، تنکابن، مازندران، ایران. رایانامه: [Balani\\_kamran@yahoo.com](mailto:Balani_kamran@yahoo.com)
۲. نویسنده مسئول، کارشناس ارشد مهندس کامپیوتر مؤسسه آموزش عالی راهبرد شمال، گروه کامپیوتر، رشت، گیلان، ایران. رایانامه: [Sadr@qiau.ac.ir](mailto:Sadr@qiau.ac.ir)
۳. دانشیار گروه مهندسی کامپیوتر، مؤسسه آموزش عالی آیندگان، گروه ریاضی کاربردی، تنکابن، مازندران، ایران. رایانامه: [Saedalatanpah@aihe.ac.ir](mailto:Saedalatanpah@aihe.ac.ir)
۴. دانشیار گروه مهندسی کامپیوتر راهبر میز داده کاوی، پژوهشکده بیمه، تهران، ایران. رایانامه: [Mantegipour@irc.ac.ir](mailto:Mantegipour@irc.ac.ir)

چکیده	اطلاعات مقاله
با توجه به بازار رقابتی صنعت بیمه و اشباع آن، حفظ مشتریان از مهمترین اهداف کارگزاران بیمه به حساب می آید. زیرا برای جذب مشتری جدید، علاوه بر ایجاد رابطه با بیمه گذار و جلب اطمینان وی، مستلزم صرف هزینه زیادی می باشد. بطوریکه هزینه جذب مشتریان جدید بسیار بیشتر از حفظ مشتریان موجود است. بر این اساس، استراتژی های بازاریابی از محصول مداری تغییر کرده و بسیاری از شرکت ها به مدیریت ارتباط با مشتریان روی آورده اند. هدف این مقاله استفاده از روش های داده کاوی به منظور بررسی و شناسایی فاکتورهای موثر در رویگردانی مشتریان در صنعت بیمه است. براساس نتایج به دست آمده، کانال جذب مشتری مهمترین عامل پیش بینی کننده رویگردانی مشتریان است و در ادامه سابقه خرید و کاربری مکان بیمه شده از سایر عوامل تاثیر گذار می باشند. تعداد زیادی از شرکت ها و سازمان ها دریافته اند که حفظ و نگهداری مشتریان فعلی اشان به عنوان گرانترین سرمایه، ارزش بسیار بالایی دارند. استراتژی شرکت های بیمه ابتدا حفظ مشتریان فعلی و سپس جذب مشتریان جدید می باشد. در این راستا، شناسایی فاکتورهای موثر در رویگردانی مشتریان از اهمیت بالایی برخوردار است. در این مقاله از روش های داده کاوی برای پیش بینی عوامل موثر در رویگردانی مشتریان استفاده شده است. براساس تجزیه و تحلیل نتایج به دست آمده، مشخص شده است کانال جذب مشتری، سابقه خرید و کاربری مکان بیمه شده به ترتیب از عوامل مهم در رویگردانی مشتریان در صنعت بیمه است.	<p><b>نوع مقاله:</b> مقاله پژوهشی</p> <p><b>تاریخ دریافت:</b> ۱۴۰۲/۰۸/۰۵</p> <p><b>تاریخ بازنگری:</b> ۱۴۰۲/۰۸/۲۹</p> <p><b>تاریخ پذیرش:</b> ۱۴۰۲/۱۰/۲۸</p> <p><b>تاریخ انتشار:</b> ۱۴۰۲/۱۲/۲۵</p> <p><b>کلیدواژه ها:</b> استراتژی های بازاریابی، داده کاوی، رویگردانی مشتری، صنعت بیمه، یادگیری ماشین.</p>

**استناد:** بالانی، کامران؛ صدر، حسین؛ عدالت پناه، سید احمد و منطقی پور، مهناز. (۱۴۰۲). «ارائه یک روش هوشمند به منظور شناسایی فاکتورهای موثر در رویگردانی مشتریان در صنعت بیمه با استفاده از تکنیک یادگیری جمعی». *مدیریت مهندسی و رایانش نرم*، دوره ۹ (۲). صص: ۹۱-۱۰۵.

<https://doi.org/>



## ۱) مقدمه

با گسترش سیستم‌های پایگاهی و حجم بالای داده‌های ذخیره‌شده در این سیستم‌ها، به ابزاری نیاز است تا بتوان این داده‌ها را با کمترین دخالت کاربر پردازش کرده و به‌صورت خودکار الگوها و رابطه‌های منطقی آنها را استخراج و اطلاعات حاصل را در اختیار کاربران قرار داد که این عملیات را داده‌کاوی می‌نامند. امروزه فرآیندهای کشف دانش و داده‌کاوی که هدف آنها به‌دست آوردن دانش مفید از داده است، به‌عنوان یک ضرورت اساسی و مهمترین ابزار برای استفاده سودمند از منابع متنوع و فراوان داده‌ها به‌شمار می‌رود. داده‌کاوی در سال‌های اخیر تاثیرات شگرفی در محیط‌های آکادمیک و صنعتی ایجاد کرده و کاربردهای فراوانی در زمینه‌های مختلف تجاری، بازاریابی، صنعت بیمه، پزشکی، ورزشی، متن‌کاوی و وب‌کاوی و غیره دارد [۱، ۲].

یکی از مهمترین زمینه‌هایی که داده‌کاوی در سال‌های اخیر در آن نفوذ داشته‌است، صنعت بیمه است. صنعت بیمه یکی از مهمترین ارکان توسعه اقتصادی کشورهاست و با مروری بر سهم صنعت بیمه در اقتصاد کشورهای توسعه‌یافته می‌توان دریافت که بیمه در مقایسه با خدمات دیگر نقش و اهمیت بیشتر و چشمگیرتری دارد. داده‌های مشتری یکی از با ارزش‌ترین دارایی‌های صنعت بیمه است. شرکت‌های بیمه مقادیر زیادی از داده‌های مربوط به مشتریان خود را جمع‌آوری کرده‌اند اما مسئله مهم و موثر، چگونگی بهره‌بردن از این حجم داده است. در این شرایط استفاده از داده‌کاوی برای دسترسی آسان به داده‌ها برای شرکت‌های بیمه بسیار مفید است [۳].

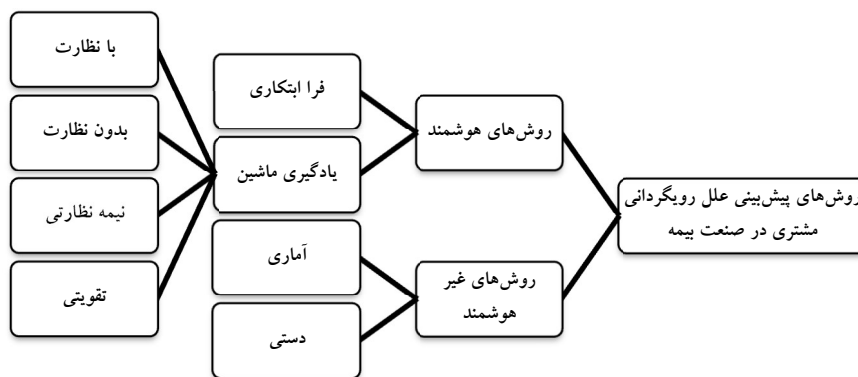
از طرق دیگر و با توجه به بازار رقابتی صنعت بیمه و اشباع آن، حفظ مشتریان از مهمترین اهداف شرکت‌های بیمه به‌شمار می‌آید [۴] زیرا برای جذب مشتری جدید علاوه بر ایجاد رابطه با بیمه‌گذار و جلب اطمینان وی، مستلزم صرف هزینه زیادی می‌باشد. بطوریکه هزینه جذب مشتریان جدید بسیار بیشتر از حفظ مشتریان موجود است که این هزینه می‌تواند ۵ تا ۷ برابر باشد [۵، ۶]. بر این اساس، استراتژی‌های بازاریابی از محصول‌مداری تغییر کرده و بسیاری از شرکت‌ها به مدیریت ارتباط با مشتریان روی آورده‌اند. در واقع تعداد بیشتری از شرکت‌ها و سازمان‌ها دریافته‌اند که حفظ و نگهداری مشتریان فعلی‌شان به‌عنوان گران‌بهاترین سرمایه، ارزش بسیار بالایی دارند [۷]. استراتژی‌های بیمه نیز ابتدا حفظ مشتریان فعلی و سپس جذب مشتریان جدید می‌باشد. از این رو سعی در شناسایی مهمترین علل رویگردانی مشتریان، از بزرگ‌ترین چالش شرکت‌ها و سازمان‌ها محسوب می‌شود [۶، ۸].

در این راستا، هدف این مقاله نیز ارائه راهکاری مبتنی بر روش‌های داده‌کاوی به‌منظور شناسایی فاکتورهای موثر در رویگردانی مشتریان است تا شرکت‌های بیمه بتوانند با استفاده از آنها به راهکارهایی به‌منظور حفظ مشتریان خود دست یابند. برای این منظور در ابتدا متغیرهای مهم و تأثیرگذار در ارزش‌گذاری مشتریان و همچنین پیش‌بینی رویگردانی، با بررسی ادبیات موضوع و نظرات خبرگان شناسایی شدند. سپس مشتریان ارزشمند براساس نمرات ترکیبی که با استفاده از متغیرهای وزن‌دار LRFMPG محاسبه شد، مشخص گشتند. سپس مدل‌های پیش‌بینی مختلف با استفاده از الگوریتم‌های متفاوت شبکه‌های عصبی مصنوعی و RBF (با چندین لایه پنهان و نرون)، درخت تصمیم (QUEST، CHAID، C&R) و C5.0، ماشین‌بردار پشتیبان (با هسته‌های مختلف) و رگرسیون لوژستیک ساخته شد و مقایسه بین عملکرد پیش‌بینی مدل‌های مختلف صورت گرفت و متغیرهای تأثیرگذار در حفظ مشتریان شناسایی شدند.

ادامه مقاله به صورت زیر سازماندهی شده است. مبانی نظری پژوهش در بخش دوم بیان شده است. بخش سوم دارای توضیحات پیرامون روش تحقیق و مدل سازی است. نتایج حاصل از پژوهش در بخش چهارم بیان شده است. بخش پنجم نیز حاوی نتیجه گیری و مسیر کارهای آینده است.

### ۱-۱) مبانی نظری و مروری بر پیشینه پژوهش

داده کاوی یک فناوری جدید قدرتمند و دارای پتانسیل بسیار خوبی برای کمک به شرکت های بیمه در استفاده از داده های جمع آوری شده در مورد رفتار مشتریان بالقوه است. داده کاوی به بخش بیمه در پیش بینی مطالبات کلاهبرداری و پوشش پزشکی و پیش بینی الگوی مشتری می پردازد که کدام مشتری از سیاست های جدید استقبال کرده و آن را خریداری می کند. امروزه پیش بینی رویگردانی یکی از بزرگ ترین چالش های شرکت های بیمه محسوب می شود. بطوریکه اعتبار و رتبه بندی مالی و رشد شرکت، مستقیماً تحت تأثیر رویگردانی مشتریان قرار می گیرد و هزینه جذب مشتریان جدید بسیار بیشتر از حفظ مشتریان موجود است. این موضوع باعث شده که تحقیقات پیرامون رویگردانی مشتریان ارزش دو چندانی داشته باشد [۹]. همانطور که در شکل ۱ نشان داده شده است، روش های پیش بینی علل رویگردانی مشتری در صنعت بیمه بطور کلی به دو دسته، روش های هوشمند و روش های غیر هوشمند تقسیم می شوند. با توجه به عملکرد خوب روش های هوشمند و یادگیری ماشین، این پژوهش بر روی روش های یادگیری ماشین بالاخص داده کاوی متمرکز است. در ادامه به برخی از برجسته ترین پژوهش های انجام شده این حوزه به طور خلاصه اشاره می کنیم.



شکل ۱. نمودار تکنیک های مختلف پیش بینی علل رویگردانی مشتریان صنعت بیمه [۵]

کلداگ مله و کینتو آسیوتی در پژوهشی به بررسی داده های بیمه درمانی پرداختند. هدف این پژوهش ارائه روشی برای تشخیص تقلب با استفاده از داده کاوی در بیمه های درمانی بوده است. برای این هدف روش های تشخیص آنومالی و ماشین بردار پشتیبان را بکار گرفتند. تمرکز این پژوهش بر شناسایی افرادی که با فریب عمدی یا بدرفتاری برای به دست آوردن برخی از مزایا، در قالب هزینه های بهداشتی تقلب می کنند، بوده است. با نتایج حاصل از این مدل، بازرسان بیمه می توانند تحقیقات بیشتری را برای افرادی که توسط این مدل ها شناسایی شده اند، انجام دهند [۱۰].

لینگ و همکارانش در پژوهشی به بررسی داده های بانک، بیمه و شرکت بونس پرداختند. هدف این پژوهش بازاریابی غیرمستقیم و دادن راه حل برای حل مشکلات بازاریابی با استفاده از تکنیک های داده کاوی بوده است. برای رسیدن به این

هدف روش‌های بیزین ساده، نزدیکترین همسایه و شبکه عصبی را بکار گرفتند. مدل ارائه‌شده در این پژوهش دارای صحت بیش از ۷۰ درصد می‌باشد [۱۱].

اسمیت کیت و همکارانش در پژوهشی که منتشر کردند به بررسی داده‌های بیمه پرداختند. هدف این پژوهش تجزیه و تحلیل الگوهای حفظ بیمه با استفاده از تکنیک‌های داده کاوی بوده است. در این پژوهش برای نیل به این هدف، پروسه کشف دانش را در یک چارچوب جامع با استفاده از آزمون فرضیه‌ها، خوشه‌بندی، درخت تصمیم و شبکه عصبی انجام دادند. مدل ارائه‌شده در این پژوهش، راهکاری را به شرکت‌های بیمه ارائه می‌کند تا بتوانند مشتریان خود را حفظ کنند [۱۲].

چو و نای در تحقیقی مشابه به بررسی داده‌های بیمه پرداختند. هدف این پژوهش ارائه روشی برای انتخاب عوامل مؤثر در فروش بیمه بوده است. برای رسیدن به این هدف روش‌های تجزیه و تحلیل محرمانه، درخت تصمیم و شبکه عصبی را بکار گرفتند. آنها با پیش‌بینی طول مدت خدمات، حق فروش بیمه و شاخص‌های پایداری عوامل بیمه به مدیران کمک می‌کنند تا عوامل اصلی مؤثر در فروش بیمه را شناسایی کنند [۱۳].

حسین خانی و همکاران در پژوهشی به شناسایی عوامل مؤثر بر رویگردانی مشتریان در صنعت بیمه به روش میدانی و پرسشنامه در یکی از بیمه‌های خصوصی ایران پرداخته‌اند. در این پژوهش متغیرهای رفتاری، دموگرافی و ادراکی لحاظ شده‌اند و برای کاهش تعداد متغیرها از روش تحلیل عاملی استفاده شده است و سپس با روش تحلیل واریانس یک‌طرفه، تفاوت واریانس‌ها در گروه‌های مختلف، تحقیق شده است. نتایج نشان می‌دهد که متغیرهای ادراک و آگاهی مشتریان، بیش از متغیرهای دموگرافی و سطح قرارداد بر رویگردانی مشتریان مؤثرند [۱۴]. در پژوهش گانتر و همکاران با استفاده از متغیرهای اکتشافی پویا و تراکنش‌های بیمه‌ای و بهره‌گیری از روش رگرسیون لوژستیک طولی، مدلی جهت پیش‌بینی ریسک رویگردانی ماهیانه در صنعت بیمه ارائه شده است. نویسندگان در این پژوهش از مدل جمعی تعمیم‌یافته به‌عنوان یک مرحله میانی به‌منظور تشخیص رابطه غیرخطی بین لوجیت و متغیرهای اکتشافی استفاده کرده‌اند بعضی از متغیرهای مؤثر بر رویگردانی دوباره تعریف شده‌اند. سپس با فرض رابطه خطی بین متغیرهای اکتشافی بازتعریف شده و لوجیت، مدل‌سازی انجام شده است [۱۵].

مهدی فیروزی و همکاران در یک پژوهش ضمن بررسی روش‌های رایج برای شناسایی تقلب در بیمه اتومبیل از سه روش داده کاوی برای پیدا کردن الگوهایی استفاده کرده که به شرکت‌های بیمه در شناسایی تقلب‌ها کمک کند. لیلا گلیجی در پژوهشی به این نتیجه رسید که با استفاده از الگوریتم ژنتیک متغیرهای تأثیرگذار مشخص شده، مورد مقایسه قرار گرفته است. نتایج نشان می‌دهد که الگوریتم CHAID و ID3، C4.5 و در نهایت روند سه الگوریتم درصد می‌تواند به‌عنوان روش مطمئن‌تری جهت کشف الگوهای مشکوک به تقلب در صنعت بیمه بدنه اتومبیل با دقت ۹۳ محسوب می‌شود [۱۶].

موریک و کاپکه در مطالعه‌ای به بررسی بیمه عمر در یکی از شرکت‌های بیمه در کشور سوئیس پرداخته‌اند. در این پژوهش، محققان با سری‌های زمانی، مدلی با بهره‌گیری از مشخصه فراوانی سند / معکوس فراوانی سند (TF/IDF) ارائه کرده‌اند. در این پژوهش سه رویکرد مختلف ارائه شده است. رویکرد نخست، پیش‌بینی بدون داده‌های زمانی است. رویکرد دوم، پیش‌بینی بر مبنای سری‌های زمانی و رویکرد نهایی استفاده از مشخصه TF/IDF و ترکیب داده‌های زمانی با اطلاعات

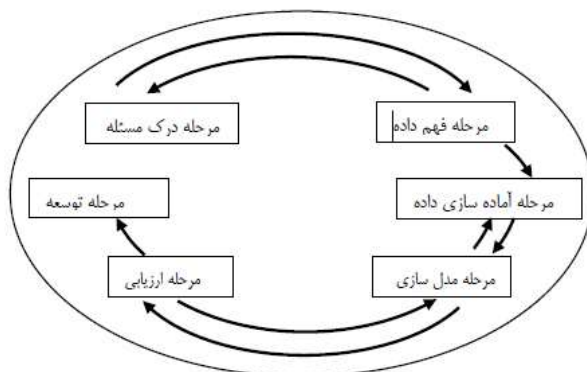
سطح قرارداد و متغیرهای رفتاری مشتریان است. مشخصه‌های جدید تعریف شده بر پایه مشخصه TF/IDF دقت، روایی و پایایی بالاتری نسبت به نتایج قبل دارند [۱۷].

توکلی و همکاران در مطالعه‌ای به کاوش در پایگاه‌های داده یکی از شرکت‌های بیمه‌ای در رشته آتش‌سوزی پرداختند. نتایج حاکی از آن است که کد حریق خریداری شده که معرف کانال جذب مشتری است، عامل اصلی پیش‌بینی کننده رویگردانی یا ماندگاری مشتری در شرکت است. بعد از آن سابقه خرید و کاربری مکان بیمه شده، عوامل پیش‌بینی کننده رویگردانی اند. همانطور که مشخص است مطالعات اندکی به حوزه تحلیل و پیش‌بینی رویگردانی در صنعت بیمه، پرداخته‌اند [۱۸].

### ۱-۲) روش‌شناسی پژوهش

بکارگیری موفق داده‌کاوی مستلزم شناخت حوزه‌ای است که قرار است داده‌کاوی در آن بکار برده شود و علاوه بر آن، شناخت کافی از روش‌ها و ابزارهای داده‌کاوی نیز لازم است. مرحله شناخت سیستم شامل مشخص کردن هدف‌های موردنظر و عوامل موفقیت کلیدی سیستم است. روش‌های متعددی برای اجرای پروژه‌های داده‌کاوی وجود دارد که یکی از روش‌های قدرتمند در این زمینه رهیافت کریسپ<sup>۶۷</sup> است. این مقاله نیز بر اساس همین رهیافت به شناسایی فاکتورهای موثر در رویگردانی مشتریان می‌پردازد. این رهیافت یک رویکرد شناخته شده برای افزایش موفقیت پروژه‌های داده‌کاوی است که شامل شش مرحله است. دیاگرام پیشنهادی براساس این رهیافت در شکل ۲ نشان داده شده است [۱۹].

رهیافت کریسپ، یک فرایند چرخه‌ای است که در آن تکرارهای مختلف چرخه فرآیند داده‌کاوی منجر به ایجاد نتایج دقیق‌تر و مبتنی بر اهداف سیستم می‌شود. در این رهیافت پس از شناسایی اهداف موردنظر (مراحل شناسایی سیستم)، داده‌ها باید با توجه به اهداف مشخص شده، جمع‌آوری شوند (فهم داده‌ها). سپس روی آنها عملیات پیش‌پردازشی به منظور استخراج ویژگی‌های موثر برای ایجاد مدل صورت گیرد (آماده‌سازی داده‌ها). مرحله مدل‌سازی، مدلی را ایجاد می‌کند که نشان‌دهنده دانش به دست آمده است (مدل می‌تواند متغیرهای هدف را که نشان‌دهنده اهداف از پیش تعیین شده سیستم هستند، پیش‌بینی کند). در ادامه مدل در مرحله ارزیابی مورد تجزیه و تحلیل قرار می‌گیرد که بتواند عملکرد و کارایی سیستم را نشان دهد. اگر دقت مدل ایجاد شده کمتر از حد انتظار باشد، یک تکرار جدید در رهیافت کریسپ صورت می‌گیرد در غیر این صورت مدل در محیط واقعی پیاده‌سازی می‌شود (توسعه).



شکل ۲. فرآیند داده‌کاوی با استفاده از مدل کریسپ

67. Crisp-dm methodology.

## ۲) شناسایی سیستم و پیش‌پردازش داده‌ها

مورد مطالعه شده این مقاله، شرکت بیمه خصوصی پارسیان است و داده‌های مورداستفاده از سه پایگاه داده خسارت، بیمه‌گذار و مشتری این بیمه به‌دست آمده‌اند. بازه استخراج داده‌ها از ۱۳۹۴/۰۱/۰۱ تا ۱۳۹۹/۰۱/۰۱ (۵ سال) بوده‌است. مجموعاً ۱۳۵۱۲۱ رکورد، مربوط به ۱۰۱۱۲۲ بیمه‌گذار به‌منظور مطالعه در این پژوهش مورداستفاده، قرار گرفته‌است. در فاز پیش‌پردازش، ادغام مشتریان، حذف رکوردهای تکراری و اجماع داده‌ها انجام شد. براساس مشاهدات، تعداد زیادی از سطرهای مجموعه داده‌ها (۹۲۹۰۹ نفر)، مربوط به بیمه‌گذاران جدید هستند که در سال ۱۳۹۸ و برای اولین بار وارد شرکت شده‌اند. با توجه به اینکه این دست از مشتریان در ارزش‌گذاری نمی‌توانند به‌عنوان داده‌های مفید استفاده شوند و پیش‌بینی رویگردانی آنها نیز امکانپذیر نیست، از مجموعه داده‌ها حذف شدند. پس از حذف مشتریان جدید و انجام فرآیند پیش‌پردازش داده‌ها و حذف رکوردهای اشکال‌دار، داده‌های مربوط به ۸۱۹۹۸ نفر مشتری در مجموعه داده‌ها باقی ماند. همچنین چند فیلد به دلیل بی‌تأثیر بودن کنار گذاشته شدند.

به‌منظور برچسب‌گذاری بیمه‌گذاران، کلیه بیمه‌گذارانی که در آخرین سال از مجموعه داده‌ها بیمه‌نامه‌ای خریداری کرده‌اند و جزء مشتریان جدید نیستند، به‌عنوان مشتریان فعال در نظر گرفته شده‌اند؛ چراکه در سال آخر فعالیت داشته‌اند. برای تعریف مشتریان رویگردان، سابقه خرید دو سال اخیر بررسی شده‌است. اگر مشتری در دو سال اخیر هیچ بیمه‌نامه‌ای خریداری نکرده‌است، به‌عنوان مشتری رویگردان برچسب‌گذاری شده‌است. اما اگر تنها سال آخر، خریدی نداشته‌است و سال قبل از آن خرید کرده‌است، فعال در نظر گرفته شده‌است. علت اینکه تنها به سابقه خرید سال آخر توجه نشده‌است این است که در مجموعه داده‌ها، موارد فراوانی وجود دارند که مشتری یکسال هیچ خریدی نداشته‌است اما مجدد به شرکت بازگشته‌است. این دست از مشتریان به‌صورت موقت رویگردان شده‌اند. در شکل ۳ پنجره زمانی تعریف مشتریان رویگردان و فعال آورده شده‌است. همچنین متغیرهای جدیدی با توجه به اهداف پژوهش تعریف شده‌اند. لازم به ذکر است پس از مشورت با خبرگان سازمان و در نظر گرفتن مجموعه داده‌های موجود، در مجموع ۱۸ متغیر در دو دسته کلی به‌منظور تعیین ارزش مشتری و تحلیل و پیش‌بینی رویگردانی در نظر گرفته شد که در جدول ۱ نشان داده شده‌است.

تنها در این سال، سابقه خرید موجود است: جدید

در این سال خرید داشته است و جدید نیست: فعال



رویگردان: مشتری خریدی نداشته      فعال: مشتری سال ماقبل آخر (۱۳۹۷) خرید داشته‌است.

شکل ۳. پنجره زمانی تعریف مشتری جدید، فعال و رویگردان

### ۳) ارزش گذاری مشتریان

با توجه به اینکه هدف این مطالعه، مدل سازی هدفمند پیش بینی است و مدل سازی هدفمند، بر مشتریان با ارزش تمرکز دارد تا منابع سازمان به طور مناسب اختصاص یابد، لذا مدلی جدید برای ارزش گذاری مشتریان پیشنهاد شده است. مدل RFM<sup>۶۸</sup>، از رایج ترین و پرکاربردترین مدل ها در تعیین ارزش مشتری مبتنی بر رفتار وی است که تعداد دفعات، زمان و مقدار خرید مشتری را در نظر می گیرد. این تحلیل، برای هر مشتری امتیازی محاسبه می کند و مشتریان با

جدول ۱- متغیرهای مورد استفاده

گروه	نام متغیر	توصیف متغیر
مشخصات بیمه گذار	State	منطقه جغرافیایی بیمه گذار
	Gender	جنسیت بیمه گذار (۰ زن و ۱ مرد)
	Age	سن بیمه گذار
	Martial	وضعیت تأهل (مجرد و متأهل)
رفتار تراکنشی	G	گروه بیمه های خریداری شده
	L	تعداد روز بین اولین و آخرین خرید بیمه گذار (طول ارتباط)
	R	تعداد روز بین آخرین خرید و تاریخ ۱۳۹۸/۱۲/۲۹
	F	تعداد کل خریدهای بیمه گذار در زمان حضور در شرکت
	M	متوسط پول پرداختی بیمه گذار (حق بیمه ها) در هر سال
	P	مقدار سودآوری بیمه گذار در زمان حضور در شرکت
	F/L	فرکانس نسبی خرید (تقسیم تعداد خرید بر تعداد سال حضور بیمه گذار)
	ClaimCount	تعداد دریافت خسارت توسط بیمه گذار
	ClaimAve	مقدار متوسط مبلغ خسارت دریافتی بیمه گذار در هر سال
	Variation	میانگین تفاضل مبلغ پرداختی در دو سال متوالی
	MIN_TIME	حداقل فاصله زمانی بین دو خرید متوالی
	MAX_TIME	حداکثر فاصله زمانی بین دو خرید متوالی
	AVG_TIME	میانگین فاصله زمانی بین دو خرید متوالی
برچسب	Label	برچسب بیمه گذار (فعال یا رویگردان)

امتیاز بالا را به عنوان مشتریان سودآور در نظر می گیرد. سه پارامتر مورد توجه در این تحلیل، تازگی، فرکانس خرید، و ارزش مالی هستند. مدل RFM توسط محققین مختلف، توسعه داده

شده است. مدل پیشنهادی این پژوهش، ترکیبی از دو مدل LRFMP<sup>۶۹</sup> که در آن متغیرهای (L) طول مدت ارتباط مشتری با شرکت و (P) سودآوری مشتری (به RFM اضافه شده است، و مدل GRFM<sup>۷۰</sup> که مشخصات گروه های کالایی

<sup>68</sup> Recency, Frequency, Monetary

<sup>69</sup> Length, Recency, Frequency, Monetary, Profit

<sup>70</sup> Group, Recency, Frequency, Monetary



که مشتری خریداری کرده است را به RFM اضافه می کند، است و مدلی جدید با نام LRFMPG<sup>71</sup> به منظور ارزش گذاری مشتریان ایجاد می کند که در آن شش متغیر در تعیین ارزش مشتریان عبارت اند از: طول مدت همکاری (L)، تأخر خرید (R)، فرکانس خرید (F)، ارزش مالی (M)، سودآوری مشتری (P) و گروه بیمه های خریداری شده (G) و سپس مقادیر LRFMPG محاسبه شده، به اعدادی در بازه [1,5] به صورت استاندارد شده نگاشت می شوند.

فرایند استانداردسازی با روش ارائه شده میگلوج انجام شده است؛ بدین صورت که در ابتدا کلیه بیمه گذاران، به ترتیب صعودی در متغیر تأخر و به ترتیب نزولی ۵ در متغیر دیگر طبقه بندی شده اند. برای متغیرهای R، M، و P ابتدا مشتریان به ۵ پنجگه برابر تقسیم شده اند. سپس بر اساس این پنجگه ها، این متغیرها به اعداد در بازه ۱ تا ۵، نگاشت شده اند. بدین صورت که مقدار ۱ نشان دهنده ارزش کمتر و مقدار ۵ ارزش بالاتر را نشان می دهد. در مورد متغیر F، با توجه به اینکه در مجموعه داده ها، تعداد بسیاری از افراد دارای  $F=1$  هستند (بیش از ۲۰ درصد) نمی توان از روش فوق استفاده کرد. لذا در زمان نمره دهی به این متغیر، خریداران با یک خرید، امتیاز ۱ دریافت می کنند. پس از آن، میانگین فرکانسهای خرید باقی مانده برای مشتریان با مقدار  $F>1$  محاسبه می شود. اگر مجموع تعداد خریدهای یک مشتری در کل کمتر از متوسط محاسبه شده باشد، نمره ۲ را دریافت خواهد کرد. این فرایند دو بار دیگر ادامه می یابد تا امتیازها برای کلیه مشتریان محاسبه شود

روش فوق در مورد متغیر L هم صادق است چراکه در مجموعه داده ها تعداد زیادی از مشتریان دارای  $L=0$  هستند؛ بدین معنی که این افراد تنها یک بار خرید داشته اند و لذا طول مدت همکاری آنان با شرکت صفر است. به منظور تعیین وزن (اهمیت) این متغیرها، مطالعه حاضر از فرایند تحلیل سلسله مراتبی AHP<sup>72</sup> بهره برده است. بدین صورت که از عوامل مؤثر استخراج شده در تعیین ارزش مشتریان، یک ماتریس نظر سنجی تهیه شد. این عوامل، تشکیل دهنده سطرها و ستونهای این ماتریس اند. از ۳ کارشناس فروش شرکت بیمه پارسیان در شعب مختلف، برای بررسی نسبی متغیرهای مزبور درخواست شد تا با مقایسه زوجی هر یک از عوامل واقع بر هر سطر ماتریس با تک تک عاملهای واقع بر ستونهای ماتریس، اهمیت هر عامل سطری بر عاملهای ستونی را به صورت عددی نمایش دهند.

پس از جمع آوری پرسشنامه ها، به وسیله ارزیابی AHP، وزن نسبی هر یک از متغیرهای LRFMPG در تعیین ارزش مشتریان (بر اساس اهمیت آنها) به دست آمد که در نمودار شکل ۴ نشان شده است. این نمودار نشان می دهد که فرکانس خرید مشتری، دارای بیشترین وزن و اهمیت است و بعد از آن طول مدت همکاری و تعداد گروه بیمه های خریداری شده بیشترین وزن را دارا هستند.

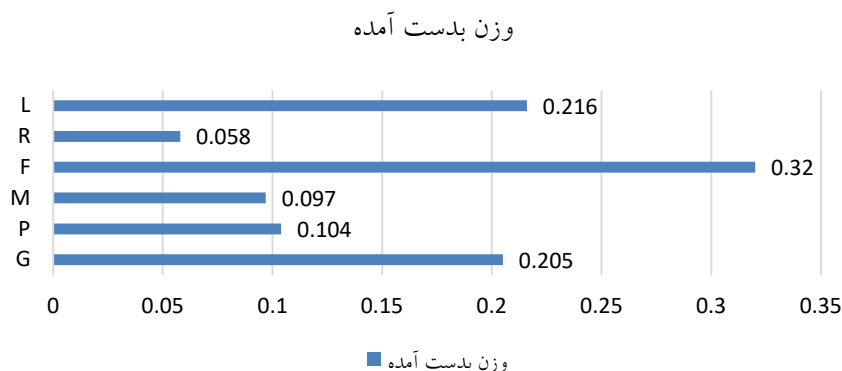
پس از تعیین ضریب تأثیر هر یک از متغیرها، ارزش و امتیاز مجموع هر یک از مشتریان تعیین می شود. فرض می کنیم که C، مجموعه مشتریان شرکت بیمه بر طبق رابطه  $C = c_1 \cup c_2 \cup \dots \cup c_n$  باشد، باشد، امتیازات استاندارد شده LRFMPG برای مشتری  $c_i$  عبارت است از:  $SL_i, SR_i, SF_i, SM_i, SP_i, SG_i$ .

در این صورت اگر مجموعه امتیازات مشتری  $c_i$  با ST نشان داده شود خواهیم داشت:

<sup>71</sup> Length, Recency, Frequency, Monetary, Profit, Group

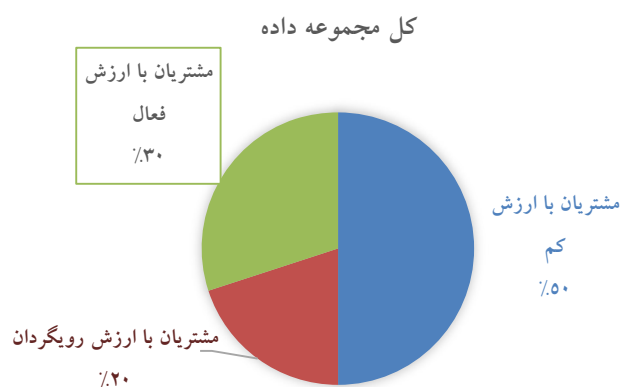
<sup>72</sup> Analytic Hierarchy Process

$$ST(c_i) = (0.216 * SLi) + (0.058 * SRI) + (0.320 * SFi) + (0.097 * SMi) + (0.104 * SPi) + (0.205 * SGi)$$



شکل ۴- وزن محاسبه شده هر یک از متغیرهای LRFMPG با استفاده از AHP

بر این اساس، بالاترین امتیاز مجموع برای یک مشتری ۵ و کمترین امتیاز مجموع، مقدار ۱ است. پس از تعیین امتیاز، به منظور نشان دادن تفاوت بین مشتریان با ارزش و کم ارزش، تحلیل نمونه تقسیم انجام شده است، بدین صورت که مشتری با امتیاز بیشتر (یا کمتر) از متوسط جمعیت، به عنوان مشتری با ارزش (یا کم ارزش) برچسب گذاری شده است. بر این اساس، تقریباً نیمی از مشتریان (۹۱۱۱۸ نفر) که امتیاز مجموع پایینی داشتند و کم ارزش بودند، از مجموعه داده ها حذف شدند. اطلاعات کلی درباره دسته بندی مشتریان کم ارزش و با ارزش (رویگردان و فعال) در شکل ۵ آمده است.



شکل ۵- فراوانی مشتریان کم ارزش و با ارزش (رویگردان و فعال)

### ۳-۱) مدل سازی

در این بخش مدل سازی پیش بینی رویگردانی را انجام می شود. گام اول، تقسیم داده به ۲ بخش آموزش و آزمون است. در این تحقیق، در ابتدا مشتریان رویگردان و فعال از یکدیگر جدا شدند، سپس به صورت تصادفی، ۸۰ درصد از هر یک از این دو مجموعه داده انتخاب شدند و به منظور ایجاد بخش آموزش با یکدیگر ترکیب شدند. سپس ۹۰ درصد باقی مانده از دو مجموعه داده مشتریان رویگردان و فعال، برای داده های آزمایشی مدل با یکدیگر ترکیب شدند. همچنین به منظور دستیابی به نتایج دقیق تر، از ارزیابی متقابل ۱۰ باره استفاده شده است.

### الف) شبکه عصبی

مدل شبکه عصبی دارای یک لایه ورودی، تعدادی لایه پنهان و یک لایه خروجی است که هر گره ورودی را دریافت کرده، آن را پردازش نموده و خروجی را تولید می‌کند. تعیین این که آیا هر ورودی به گره خروجی خواهد رسید یا خیر، به وزن آن ورودی بستگی دارد.

شبکه عصبی یک ابزار بسیار قوی برای رده‌بندی کردن و پیش‌بینی نمودن می‌باشد که به‌عنوان جزئی از متدهای یادگیری مدرن محسوب می‌شود. شبکه‌های عصبی از مدل نرون بیولوژیکی الهام گرفته شده‌است و از تعدادی از ویژگی‌های بیولوژیکی عصبها همانند غیرخطی بودن، سادگی واحدهای محاسباتی و قدرت یادگیری بهره می‌برند. شبکه‌های عصبی قادر به شناسایی روابط در بین اطلاعات موجود و نمایش آن روابط می‌باشند. ورودی‌های شبکه‌های عصبی متغیرها هستند و خروجی‌ها حالت‌هایی هستند که نیاز به پیش‌بینی یا کنترل آنها داریم. به‌منظور استفاده از شبکه عصبی و به ازای مقادیر مختلف پارامترهای آلفا (فاکتور تکانه) و اتا (نرخ یادگیری) مدل‌های مختلف ساخته و بر اساس صحت مدل‌های ساخته شده مقادیر آلفا و بتا به ترتیب ۰/۹ و ۰/۳ در نظر گرفته شد. همچنین چندین مدل مختلف با تعداد متفاوت لایه پنهان (۱ تا ۳) و نرون (۲ تا ۵ نرون در هر لایه) ساخته شد.

### ۳-۱-۱) درخت تصمیم

درخت تصمیم نیز ترکیب تعدادی استلزام منطقی است که برخلاف شبکه‌های عصبی به تولید قانون می‌پردازد. در ساختار درخت تصمیم، پیش‌بینی به دست آمده از درخت در قالب یک سری قانون توضیح داده می‌شود و در دسته بندهای مبتنی بر قانون، دانش خروجی به‌صورت یک مجموعه قوانین اگر آنگاه نشان داده می‌شود. لازم به ذکر است که درخت‌های تصمیم تنها نمایشی از فرایند تصمیم‌گیری نیستند، بلکه می‌توان از آنها در حل مسائل دسته‌بندی نیز استفاده کرد. برای ایجاد درخت و تقسیم شاخه‌ها از شاخص جینی به دلیل این که درخت تولید شده توسط آن، قابل فهم‌تر است، استفاده شده‌است. معیار اطمینان نیز برای قوانین ایجاد شده توسط درخت تصمیم ۸۰ درصد در نظر گرفته شده‌است. در آزمایشات این مقاله به‌منظور افزایش دقت مدل‌سازی‌های انجام شده، الگوریتم‌های درخت تصمیم CHAID و C5.0 به‌صورت تقویت شده نیز استفاده شده‌اند.

### ۳-۱-۲) ماشین بردار پشتیبان

الگوریتم ماشین بردار پشتیبان یکی از پرکاربردترین الگوریتم‌ها در علوم داده کاوی و متن کاوی است برای این کاربرد زیاد دو دلیل مهم ذکر شده‌است از جمله این که این الگوریتم دارای سرعت بالا در اجرا می‌باشد. بدین شکل که یکی از سریع‌ترین الگوریتم‌ها در علوم داده کاوی است در کنار این سرعت این الگوریتم دارای پیاده‌سازی آسانی است و به راحتی پیاده‌سازی می‌شود. این مورد نیز باعث شده‌است از طرف محققین مورد توجه بیشتر قرار گیرد.

این روش معمولاً دارای دو کلاس می‌باشد که این دو کلاس به وسیله یک خط از هم جدا می‌شوند. نزدیکترین نمونه‌ها به محدوده تصمیم‌گیری به نام بردارهای پشتیبانی نامیده می‌شوند. این روش کارایی بهتری را در مواقعی که مدل‌ها نیاز به مینیمم‌سازی ریسک ساختاری دارند، بوسیله ماکزیمم کردن فاصله بین دو ابر ناپایدار در بردار پشتیبانی از دو کلاس، دارا می‌باشد. و باعث کاهش میزان خطا و ریسک بهنگام عملیات می‌شود. در این روش، خط فاصل بین دو کلاس از روش زیر محاسبه می‌شود:

- تمام نمونه ها در کلاس شماره یک، در یک طرف خط فاصل و تمام نمونه های مربوط به کلاس دوم در سمت دیگر خط فاصل قرار می گیرند.
- خط فاصل واقعی مورد استفاده برای تصمیم گیری خطی است که دارای نزدیکترین فاصله به نمونه های تمرین هر دو کلاس باشد.

### ۳-۲) ارزیابی

در این مرحله دانش تولید شده در مرحله قبل ارزیابی شده و مورد تفسیر قرار می گیرد. منظور از ارزیابی دانش آن است که می بایست میزان صحت دانش تولید شده مشخص شود تا بتوان به آن اعتماد نمود و به صورت عملی از آن استفاده کرد. روش های مختلفی برای ارزیابی دانش تولید شده وجود دارند که رابطه تنگاتنگی با روش یادگیری مدل دارند. تفسیر مدل به این معنا است که دانش تولید شده را مورد بررسی قرار داده و توجیهی معنایی جهت تبیین منطق آن ارائه کند. به بیان دیگر به کمک ارزیابی می توان مدل های مختلف را مورد بررسی قرار داد و بهترین مدل را بر اساس کارایی پیش بینی گزینش کرد.

نتایج ارزیابی ها به کمک شبکه عصبی در جدول ۲ و به کمک سایر مدل های درخت تصمیم و مشاین بردار پشتیبان در جداول ۳ و ۴ نشان داده شده است. بر اساس نتایج حاصل از آزمایشات مشخص است که الگوریتم های درخت تصمیم از قابلیت بالاتری برای پیش بینی برخوردارند.

#### جدول ۲- نتایج حاصل از پیاده سازی انواع مختلف شبکه های عصبی

معیار F	صحت	بازخوانی	دقت	لایه پنهان	روش
۷۸/۰۸	۷۴/۳۵	۷۴/۸۱	۸۱/۶۰	۱ لایه پنهان ۲ نرون	شبکه عصبی پرسپترون چند لایه MLP
۷۹/۷۱	۷۴/۷۷	۸۱/۲۸	۷۸/۲۱	۱ لایه پنهان ۳ نرون	
۷۷/۷۹	۷۱/۲۵	۸۲/۴۹	۷۳/۵۶	۱ لایه پنهان ۴ نرون	
۷۹/۷۰	۷۴/۳۰	۸۳/۰۸	۷۶/۵۹	۲ لایه پنهان ۴ نرون	
۶۴/۰۴	۶۴/۶۶	۵۱/۶۳	۸۴/۳۵	۳ لایه پنهان ۹ نرون	
۷۵/۴۷	۷۰/۶۵	۷۴/۰۵	۷۶/۹۵	-	شبکه عصبی RBF

همان طور که در جدول ۲ مشخص است، شبکه عصبی MLP با تعداد یک لایه پنهان و سه نرون، بهترین عملکرد پیش بینی را داشته است.

#### جدول ۳- نتایج حاصل از پیاده سازی انواع مختلف درخت تصمیم

معیار F	صحت	بازخوانی	دقت	لایه پنهان	روش
۸۲/۳۳	۷۷/۹۴	۸۴/۳۳	۸۰/۴۳	QUEST	درخت تصمیم
۸۵/۱۸	۸۱/۹۴	۸۵/۲۱	۸۵/۱۷	CHAID	
۸۸/۴۴	۸۵/۹۴	۸۸/۵۹	۸۸/۳۸	Boosted CHAID	
۸۶/۵۳	۸۳/۵۳	۸۶/۸۳	۸۶/۴۴	C5.0	
۸۸/۷۵	۸۶/۰۶	۹۰/۱۷	۸۷/۳۸	Boosted C5.0	
۸۱/۵۶	۷۷/۲۲	۸۲/۶۵	۸۰/۵۰	C&R	

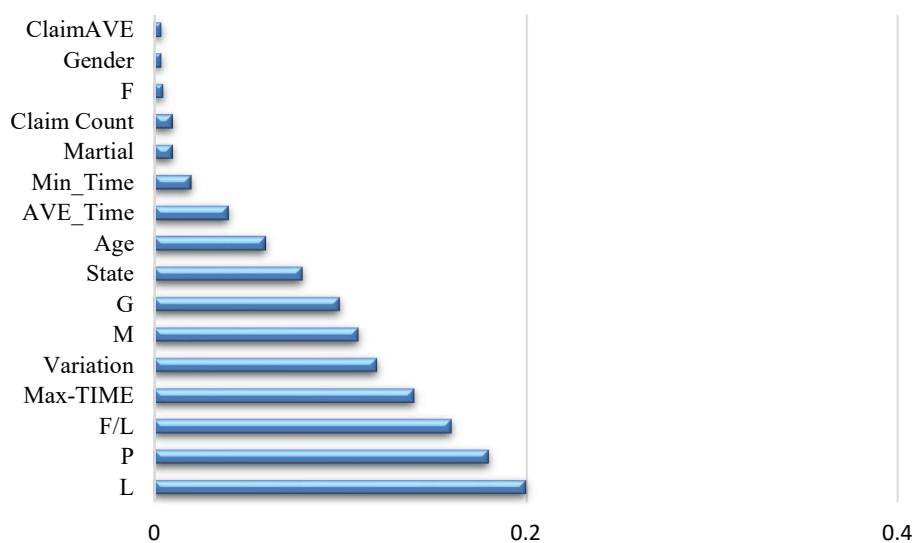
همان‌طور که در جدول ۳ مشخص است، درخت تصمیم Boosted CHAID بهترین عملکرد پیش‌بینی را داشته‌است.

**جدول ۴- نتایج حاصل از پیاده‌سازی انواع مختلف ماشین‌بردار پشتیبان**

معیار F	صحت	بازخوانی	دقت	لایه پنهان	روش
۷۹/۴۱	۷۳/۹۴	۸۲/۱۸	۷۶/۷۳	RBF Kernel	ماشین‌بردار
۷۹/۵۷	۷۴/۹۰	۷۵/۸۳	۸۱/۶۹	Poly Kernel	پشتیبان

استنتاج درخت تصمیم‌گیری می‌تواند برای توصیف گروه مشتریان رویگردان استفاده شود و نتایج حاصله از آن را می‌توان در مدیریت ارتباط با مشتری استفاده کرد. همچنین از منظر اهمیت متغیرها بر فعال یا رویگردانبودن مشتریان، با درخت تصمیم C5.0 بررسی انجام شده‌است. بر این اساس، متغیر طولمدت ارتباط مشتری، تأثیرگذارترین متغیر است. بعد از آن، سودآوری مشتری، فرکانس خرید نسبی و حداکثر فاصله بین دو خرید متوالی، به‌ترتیب متغیرهای با اهمیت بوده‌اند. شکل ۶ ترتیب اهمیت متغیرها را در پیش‌بینی انجام شده با استفاده از C5.0 نشان می‌دهد.

بر اساس قوانین استخراج شده و همچنین متغیرهای تأثیرگذار بر پیش‌بینی، آن دست از مشتریان که طول مدت ارتباط کمتری با شرکت دارند و به تعبیری به تازگی وارد شرکت شده‌اند، مشتریانی که سودآوری (تفاضل پرداخت‌های مالی و مقدار کل خسارت دریافتی) کمتر در طی حضور در شرکت دارند و مقدار خرید نسبی پایین‌تری دارند، احتمال بالاتری دارد که رویگردان شوند. همچنین آن دست از مشتریان که حداکثر فاصله بین دو خرید متوالی برای آنها بیشتر است و مقدار رشد خریدشان (تفاضل خرید دو سال متوالی) رو به کاهش است، جزو مشتریان در خطر رویگردانی‌اند. متغیرهای فرکانس خرید و تعداد و مقدار دریافت خسارت، تأثیر اندکی بر رویگردانی داشتند. از سویی برخلاف آنچه تصور می‌شد، متغیرهای مربوط به مشخصات مشتری (پروفایل) تأثیر بسیار کمی بر رویگردانی دارند و به جز منطقه جغرافیایی مشتری، سایر متغیرها جزو کم‌تأثیرترین متغیرها بوده‌اند.



شکل ۶- اهمیت متغیرها در پیش‌بینی با استفاده از درخت C5.0

## ۴) نتیجه گیری و کارهای آینده

امروزه با پیشرفت فناوری و حضور گسترده‌ی آن در زندگی روزمره‌مان شاهد کاربرد پررنگ داده و اطلاعات هستیم، یکی از روش‌های استخراج اطلاعات از داده‌های خام داده‌کاوی می‌باشد. داده‌کاوی یک فناوری جدید قدرتمند و دارای پتانسیل بسیار خوبی برای کمک به شرکت‌های بیمه در استفاده از داده‌های جمع‌آوری شده در مورد رفتار مشتریان بالقوه است توجه به اهمیت داده‌کاوی در صنعت بیمه در این مقاله مسئله رویگردانی مشتریان از صنعت بیمه مورد بررسی قرار گرفت. مدل پیش‌بینی رویگردانی می‌تواند به‌عنوان ابزار هشداردهنده‌ی اولیه برای کسب و کارها و استخراج‌کننده فاکتورهای حیاتی مرتبط با رویگردانی مشتری استفاده شود و امکان فراهم کردن دانش مفید افزوده برای پشتیبانی از تصمیم‌گیری را فراهم آورد. در این مطالعه، از یکسو به بررسی متغیرهای تأثیرگذار در رویگردانی مشتریان با ارزش در یک شرکت بیمه پارسیان پرداخته شد و از سوی دیگر پیش‌بینی رویگردانی مشتریان با ارزش شرکت انجام شد. برای این منظور در ابتدا متغیرهای مهم و تأثیرگذار در ارزش‌گذاری مشتریان و همچنین پیش‌بینی رویگردانی، با بررسی ادبیات موضوع و نظرات خبرگان شناسایی شدند. سپس مشتریان ارزشمند، بر اساس نمرات ترکیبی که با استفاده از متغیرهای وزن دار LRFMPG محاسبه شد، مشخص گشتند. سپس مدل‌های پیش‌بینی مختلف با استفاده از الگوریتم‌های متفاوت شبکه‌های عصبی مصنوعی و RBF (با چندین لایه پنهان و نرون، درخت تصمیم) QUEST، CHAID، C&R و C5.0 (ماشین‌بردار پشتیبان) با هسته‌های مختلف (و رگرسیون لوژستیک) ساخته شد و مقایسه بین عملکرد پیش‌بینی مدل‌های مختلف صورت گرفت. برای ساخت مدل، تأثیر ۱۸ متغیر در دو دسته کلی (مشخصات مشتری و رفتار تراکنشی) مورد بررسی قرار گرفت. بر این اساس مدل درخت تصمیم‌گیری با الگوریتم C5.0 به‌صورت تقویت شده نسبت به سایر مدل‌ها در پیش‌بینی رویگردانی مشتری عملکرد بهتری داشت. متغیر مهم و تأثیرگذار بر پیش‌بینی به‌ترتیب عبارت‌اند از طول مدت ارتباط مشتری، سودآوری مشتری، فرکانس خرید نسبی، حداکثر فاصله بین دو خرید متوالی و میانگین تفاضل پرداخت دو سال متوالی. نتایج حاکی از آن است که تمامی متغیرهای LRFMPG تأثیر بالایی بر رویگردانی ندارند و تأثیر هر یک متفاوت است. همچنین متغیرهای مربوط به مشخصات مشتری (پروفایل) در پیش‌بینی رویگردانی بسیار کم اهمیت هستند. وجوه مختلفی می‌تواند در آینده مدنظر قرار گرفته شوند. در ابتدا اینکه مجموعه داده استفاده شده، تنها توسط یک شرکت بیمه فراهم شده‌است. داده‌های شرکت‌های بیمه متعدد می‌تواند جمع‌آوری شوند و به‌منظور بالابردن تعمیم‌پذیری مدل LRFMPG، مقایسه شوند. از طرفی با توجه به اینکه تعدادی از متغیرهای در نظر گرفته شده در مدل ارائه شده، برای تحلیل ارزش مشتری و پیش‌بینی رویگردانی اهمیت کمی دارند، لذا مطالعات آینده می‌تواند مدل‌های دیگر تحلیل ارزش مشتری را به کار گیرند. همچنین می‌توان از الگوریتم‌های فراابتکاری به‌منظور پیش‌بینی رویگردانی استفاده کرد تا دقت مدل پیش‌بینی را بالا برد.

## منابع

- Karunarathna, K. and D. Rasika, *Factors Affecting Salesforce Unethical Behavior: Evidence from Life Insurance Industry*. Wayamba Journal of Management, 2021. 12(2). DOI: [10.1213/ANE.0b013e31828843ef](https://doi.org/10.1213/ANE.0b013e31828843ef)
- Osman, A.S., *Data mining techniques*. 2019. DOI: [10.1111/j.1423-0410.2010.01446](https://doi.org/10.1111/j.1423-0410.2010.01446)
- Nicholson, J.E., *Challenges for the Insurance Industry in the Future*. Journal of Insurance Regulation, 2019. 38(6). Doi: <https://doi.org/10.1016/j.tele.2013.02.002>

- Pirmohammadi, G. and M.M. Zohouri, *Comparison of Artificial Neural Network and SPSS Model in Predicting Customers Churn of Iran's Insurance Industry*. International Journal of Computer Applications. **975**: p. 8887. doi: <https://doi.org/10.13052/jicts2245-800X.12a9>
- Cao, S., et al. Deep Learning Based Customer Churn Analysis. in 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP). 2019. IEEE. Doi: <https://doi.org/10.1016/j.chb.2016.04.023>
- Saghir, M., et al. Churn prediction using neural network based individual and ensemble models. in 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST). 2019. IEEE. Doi: <https://doi.org/10.1016/j.compeleceng.2017.09.001>
- Khodabandehlou, S. and M.Z. Rahman, *Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior*. Journal of Systems and Information Technology, 2017. Doi: <https://doi.org/10.1016/j.future.2019.01.020>
- Bhatnagar, A. and S. Srivastava. A Robust Model for Churn Prediction using Supervised Machine Learning. in 2019 IEEE 9th International Conference on Advanced Computing (IACC). 2019. IEEE. Doi: <https://doi.org/10.1016/j.future.2019.01.020>
- Omerašević, A. and J. Selimović, *Risk factors selection with data mining methods for insurance premium ratemaking*. Zbornik Radova Ekonomski Fakultet u Rijeka, 2020. 38(2): p. 667-696. Doi: <https://doi.org/10.1002/sec.795>
- Kirlidog, M. and C. Asuk, *A fraud detection approach with data mining in health insurance*. Procedia-Social and Behavioral Sciences, 2012. **62**: p. 989-994. DOI: <https://doi.org/10.1016/j.eswa.2011.01.075>
- Ling, C.X. and C. Li. *Data mining for direct marketing: Problems and solutions*. in *Kdd*. 1998. <https://doi.org/10.3724/SP.J.1383.203008>
- Smith, K.A., R.J. Willis, and M. Brooks, *An analysis of customer retention and insurance claim patterns using data mining: A case study*. Journal of the operational research society, 2000. 51(5): p. 532-541. doi.org/10.1016/j.energy.2016.06.087
- Cho, V. and E.W. Ngai, *Data mining for selection of insurance sales agents*. Expert systems, 2003. 20(3): p. 123-132. <https://doi.org/10.3724/SP.J.1383.203008>
- حسین خانی، ن. س. م. حسینی مطلق، و م. خاکزار بفرویی، شناسایی عوامل موثر بر رویگردانی مشتریان در صنعت بیمه، در بیست و یکمین همایش ملی و هفتمین همایش بین المللی بیمه و توسعه. ۱۳۹۳.
- Günther, C.-C., et al., *Modelling and predicting customer churn from an insurance company*. Scandinavian Actuarial Journal, 2014. 2014(1): p. 58-71. <https://doi.org/10.3724/SP.J.1383.203008>
- فیروزی، م. و همکاران، شناسایی تقلب در بیمه اتومبیل با استفاده از روش های داده کاوی. پژوهشنامه بیمه (صنعت بیمه)، ۱۳۹۰.
- Morik, K. and H. Köpcke. Analysing customer churn in insurance data—a case study. in European conference on principles of data mining and knowledge discovery. 2004. Springer. <https://doi.org/10.3724/SP.J.1383.203008>
- توکلی، ا. و همکاران، به کارگیری فرایند داده کاوی برای پیش بینی الگوهای رویگردانی مشتری در بیمه. چشم انداز مدیریت بازرگانی، ۱۳۹۱.
- Olson, D.L. and D. Delen, *Advanced data mining techniques*. 2008: Springer Science & Business Media. <https://doi.org/10.3724/SP.J.1383.203008>