قیمت‌گذاری بهینه بیمه نامه با استفاده از روش‌های داده کاوی

مینا رنجبرفرد[۱]
الهه پروانه‌وار[۲]

چکیده

صنعت بیمه یکی از بخش‌های حیاتی اقتصاد در هر کشوری است، به این دلیل که در حفظ و تضمین سرمایه نقش مهمی ایفا می‌کند. یکی از مسائلی که در بیمه از اهمیت بسیاری برخوردار است، دستیابی به رشد بازار و سودآوری است. تنظیم سطح قیمت متناسب با هزینه‌ها و هم‌زمان با آن حفظ مشتری موجود و جذب مشتریان جدید مسئله‌ای است که در صنعت بیمه همچون دیگر صنایع باید به آن توجه شود. شرکت‌های بیمه‌ای نیازمند استفاده از فناوری-هایی هستند که بر مبنای آن بتوانند داده‌ها را تبدیل به اطلاعات نمایند. یکی از انواع فناوری‌های اطلاعات قابل‌استفاده در خدمات بیمه‌ای، استفاده از فنون داده‌کاوی است. در این مقاله سعی شده است مروری بر پژوهش‌های کاربردی در حوزه بیمه که منجر به ارائه یک‌راه حل برای قیمت‌گذاری بهینه‌شده‌اند، صورت گیرد که راهنمایی برای محققان و مدیران این حوزه باشد. بر همین اساس پس از جست‌وجوی مقالات بیمه‌ای با رویکرد داده‌کاوی در پایگاه‌های علمی معتبر حدود ۲۵ مقاله مطالعه شد. درنهایت روش‌های مورداستفاده در این مطالعات استخراج و دسته‌بندی شد.

کلمات کلیدی: صنعت بیمه، داده‌کاوی، قیمت تمام‌شده، قیمت‌گذاری بهینه، اشتراک دانش

[۱] استادیار گروه مدیریت، دانشگاه الزهرا، عضو هیئت علمی دانشگاه الزهرا mina.ranjbar.ie@gmail.com
[۲] دانشجوی کارشناسی ارشد مدیریت فناوری اطلاعات، دانشگاه الزهرا elaheparvanehvar@gmail.com

www.SID.ir

مقدمه

صنعت بیمه یک منبع با حجم زیاد مشکلات کسب‌وکار است که همواره موردتوجه محققان عملیاتی قرار می‌گیرد که در مسائلی نظیر قیمت‌گذاری می‌تواند با استفاده از فن‌های پژوهش عملیاتی سنتی ازجمله رگرسیون، برنامه‌ریزی آرمانی، برنامه‌ریزی خطی و رویکردهای جدید از قبیل شبکه‌های عصبی، سامانه‌های خبره و داده‌کاوی حل شود (Smith *et al.*, 2002). بااین‌حال، شرکت‌های بیمه در سازگاری با ابزارهای داده‌کاوی نسبتاً کند بوده‌اند که عمدتاً به علت عدم رقابت ناشی از مقررات محافظتی است؛ اما اکنون، آن‌ها دیگر نمی‌توانند این موضوع را نادیده بگیرند که مواردی از قبیل اینترنت، مقررات زدایی، تثبیت و همگرایی بیمه با سایر خدمات مالی، به‌سرعت ساختار اصلی صنعت را تغییر می‌دهد (Sumathi and Sivanandam, 2006).

برای تعیین حق بیمه دقیق سالیانه در شرکت بیمه، تخمین دقیق و قابل‌اعتماد بسیار مهم است. یکی از ابزارهایی که می‌تواند به بیمه‌گر کمک کند داده‌کاوی است. داده‌کاوی مجموعه‌ای از الگوریتم‌ها است که به سازمان‌ها کمک می‌کند تا بر مهم‌ترین اطلاعات و دانش موجود در پایگاه‌های داده خود تمرکز کنند. جکسون در تعریف داده‌کاوی دو روش متفاوت در انجام داده‌کاوی بیان می‌کند که طی آن یا به دنبال ساخت مدل و یا یافتن الگو از میان داده‌ها است. در رویکرد اول استخراج مدل از میان مجموعه داده‌ها برای شناسایی و توصیف ویژگی‌های اصلی توزیع داده‌ها انجام می‌شود. به‌عبارت‌دیگر به دنبال نشان دادن روابط است. دومین رویکرد داده‌کاوی، برای شناسایی الگوهای غیرمعمول رفتار از میان قوانین به کار می‌رود (Smith *et al.*, 2002).

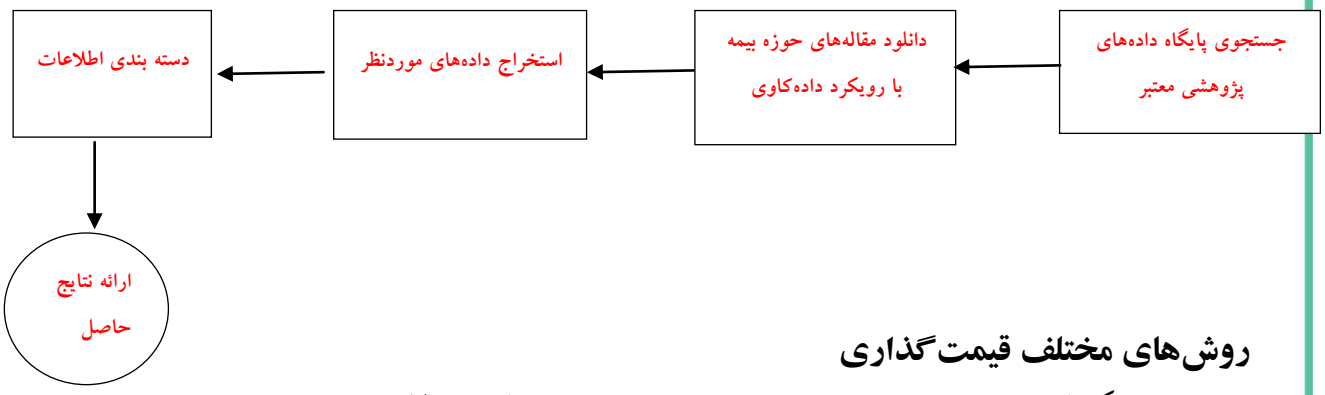
در رویکرد پژوهشی پیش رو از یافته‌های پژوهش‌های قبلی برای پیشنهاد راه‌کارهای مفید استفاده می‌شود. در این مقاله تلاش بر آن بوده که مطالعات برحسب کاربردهای داده‌کاوی توصیف و در اختیار خواننده قرار بگیرد. ازاین‌رو پژوهش‌های مطالعه شده بر اساس کاربردهای داده‌کاوی در بیمه به جهت قیمت‌گذاری ارائه‌شده‌اند. در ادامه سعی بر آن شده است که ادبیات موجود در پایگاه‌های علمی گوناگون از قبیل Scopus، Emerald، Science Direct، IEEE، Google Scholar موردبررسی قرارگرفته و پژوهش‌های مرتبط قیمت‌گذاری در بیمه با رویکرد داده‌کاوی استخراج شود.

www.SID.ir

روش تحقیق

پژوهش حاضر به بررسی مطالعات موجود در حوزه تعیین حق بیمه مطلوب با رویکرد داده‌کاوی است، به‌گونه‌ای که بتواند جنبه‌های مختلف مؤثر بر قیمت‌گذاری در بیمه را ارائه بدهد. بنابراین رویکرد این پژوهش کاربردی توصیفی است. این پژوهش کاربردی است زیرا توسعه کاربرد دانش در یک حوزه خاص را بیان می‌کند. همچنین توصیفی است به این لحاظ که توصیف نتایج پژوهش صرفاً به بازنمود مسئله تحقیق می‌پردازد و به دنبال تحلیل محتوای مقاله‌های یافت شده است.

در این پژوهش، برخی مقاله‌های موجود در قیمت‌گذاری بیمه با رویکرد داده‌کاوی مطالعه شد. نتایج حاصل از این مقالات، دسته‌بندی‌شده و اطلاعات آنان ارائه می‌شود. درنهایت روند انجام پروژه طبق فلوچارت زیر رسم شده است.



روش‌های مختلف قیمت‌گذاری

با پیشرفت تکنولوژی توقع مشتریان در خصوص ارائه خدمات مطلوب بالاتر می‌رود. ارائه این خدمات مستلزم مدیریت درآمدها و هزینه‌ها است. در روش‌های کلاسیک قیمت‌گذاری، حق بیمه بر اساس تعرفه‌های ثابت و مشخص تعیین می‌شود. طبق پژوهش‌های بررسی‌شده روش‌های مختلف داده‌کاوی برای تعیین حق بیمه مناسب در جهت جذب و حفظ مشتری در بیمه می‌تواند مؤثر باشد. در این بخش پژوهش‌هایی باهدف تعیین حق بیمه مناسب در جهت حفظ مشتریان با استفاده از الگوریتم‌های داده‌کاوی موردبررسی قرار خواهند گرفت.

در پژوهشی که توسط سونگ و همکاران در جهت حفظ مشتریان انجام‌شده، درخت تصمیم‌گیری و مدل رگرسیون لجستیک را انتخاب می‌کند. مدل تولیدشده توسط درخت تصمیم‌گیری یک ساختار درخت است که میزان مشخصی که هر متغیر بر متغیر هدف اثر می‌گذارد را پیدا می‌کند. مدل رگرسیون لجستیک احتمال روی‌گردانی هر

www.SID.ir

مشتری را محاسبه می‌کند (Song and Qiuhong, 2018). در مطالعه دوال و کولکارنی بر زمینه‌های مختلف بیمه عمر تمرکز شده است. در این مطالعه مشتریان طبق اطلاعات موجود خوشه‌بندی می‌شوند. قانون انجمن را می‌توان برای حفظ مشتری‌های موجود مورداستفاده قرارداد که با بررسی داده‌های قبلی و پیدا کردن ترکیب‌های موردنیاز با اطمینان و پشتیبانی، بیمه‌گر می‌تواند بیمه‌نامه‌های جدید را برای مشتریان موجود برای حفظ آن‌ها به فروش برساند. طبقه‌بندی می‌تواند برای هدف قرار دادن مشتریان و یا طراحی محصولات جدید مورداستفاده قرار گیرد (Kulkarni, R. V., & Devale, 2012). در پژوهش دیگری در مورداستفاده از روش‌های مدرن داده‌کاوی (DM) برای طراحی حق بیمه برای وسایل نقلیه موتوری بحث شده است. یک رویکرد دومرحله‌ای، شامل یک مدل تحلیل بقا[3] و یک مدل رگرسیون خطی، برای برآورد سطح ریسک هر مشتری و ثبت فایل ادعا طراحی‌شده است. درنهایت، نتایج حاصل از مدل‌های تصادف و سرقت برای جمع‌آوری ادعای کل مورد انتظار برای هر مشتری ترکیب‌شده است (Kahane, Levin, Meiri, & Zahavi, 2007). در پژوهش ساختیول و راجیتها بیان‌شده است که برای تعیین حق بیمه دقیق سالیانه در شرکت بیمه، تخمین دقیق و قابل‌اعتماد از تعداد ادعاهایشان و میزان کل ادعا بسیار مهم است. ازآنجاکه NN ها با استفاده از داده‌های اولیه موجود آموزش داده می‌شوند، خطا ممکن است کاهش یابد و همچنین در مقایسه با روش‌های دیگر برای پیش‌بینی اهمیت دارد، ANN با روش اعتبار بیزی، یک روش آماری برای پیش‌بینی فراوانی ادعای آینده، تطبیق داده‌شده است. نتایج حاکی از آن است که ANN برآورد قابل‌توجهی از فرآیند ادعای واقعی را در مقایسه بااعتبار بیزی در مدل پواسون / گاما برای بیمه غیرزندگی فراهم می‌کند (Sakthivel & Rajitha, 2017). در پژوهش دیوید به‌منظور ایجاد حق بیمه خالص با توجه به ویژگی‌های بیمه‌گذاران بر اساس ضرب فراوانی پیش‌بینی‌شده و هزینه ادعا، حق بیمه محاسبه‌شده است. این مقاله یک تحلیل از مدل‌های خطی کلی را در نظر می‌گیرد؛ بنابراین، به‌عنوان مرحله اول، فراوانی ادعا توسط مدل رگرسیون پواسون برآورد شده است. در مرحله بعدی، با استفاده از مدل گاما، میانگین برآورد هزینه ادعایی مربوط به هر طبقه بیمه‌شده‌ها تعیین می‌شود (David, 2015). شن و همکاران تمرکز خود را بر بیمه محصولات کشاورزی معطوف کرده‌اند. یک چارچوب آماری در برآورد بیزی را ارائه کرده‌اند که اجازه می‌دهد تا ترکیب صریح از اطلاعات پیشین، مشاهدات محصول و نظرات کارشناسان باشد. وابستگی احتمالی ناحیه تولید در مناطق مختلف توسط یک

[3] survival analysis model

www.SID.ir

واین کاپیولا[4] گرفته‌شده است که قادر به ساخت ساختار وابسته به ابعاد بزرگ است (Shen, Odening and Okhrin, 2013). در پژوهشی دیگر مسئله‌ای موردتوجه قرار می‌گیرد که قیمت‌گذاری مناسب بیمه‌نامه است. بیمه‌های خودرو باید با استفاده از روش‌های Tweedie و ZAIG برآورد میزان ادعای کل را در نظر بگیرند. این سیستم نمره دهی شناساننده افرادی است که بیش‌ترین ریسک رادارند؛ بنابراین این مدل‌های برآورد شده ممکن است برای توسعه یک استراتژی برای قیمت‌گذاری حق بیمه استفاده شود (Bortoluzzo *et al.*, 2011). در پژوهش کیچن و هریس بیان می‌شود در بیمه، مهم‌ترین ملاحظات در قیمت‌گذاری بیمه‌گذار عبارت‌اند از: فراوانی خسارت و شدت خسارت (احتمال از دست دادن و ارزش مالی از دست دادن)؛ بنابراین شبکه‌های عصبی مصنوعی و رگرسیون خطی را برای بیمه‌نامه‌های خصوصی اتومبیل‌های مسافری در یک‌رویه سه مرحله‌ای توسعه داده است (Bortoluzzo *et al.*, 2011).

در پژوهشی که توسط بیان و همکاران انجام‌شده است، مدل قیمت‌گذاری بیمه خودرو مبتنی بر رفتار[5] (BVIP) و یک نمونه اولیه محاسبه حق بیمه خودرو برای بیمه خودرو شخص توسعه داده‌شده است (Bian *et al.*, 2018). مقاله گولمن تئوری بوستینگ گرادیان (GB) و کاربرد آن را در مسئله پیش‌بینی هزینه اتلاف تصادفات خودرو "مقصر" با استفاده از داده‌های یک بیمه‌گر کانادایی ارائه می‌دهد. دقت پیش‌بینی مدل در مقایسه با مدل خطی معمولی (GLM) مقایسه شده است (Guelman, 2012). پرداخت بیمه به ازای میزان رانندگی روشی است که در برخی کشورها جهت محاسبه قیمت بیمه خودرو مورداستفاده قرار می‌گیرد. پژوهش فریرا و مینیکل از مدل‌های پواسون و رگرسیون خطی برای تحلیل روابط بین متغیرها در این حالت استفاده کردند. یک طرح قیمت‌گذاری احتمالی می‌تواند نرخ بیمه را به‌طور کامل به ازای هر مایل با هیچ هزینه‌ای سالیانه داشته باشد (Ferreira and Minikel, 2012). مطالعه ییپ و یوان استفاده از توجیه شمارش ادعای مجموعه داده‌های بیمه اتومبیل، بر اساس مسئله zero-inflation را پیشنهاد می‌کند که از چندین مدل zero-inflation استفاده شود. بر اساس یافته‌ها، مدل توزیع دوتایی منفی (NB[6]) و مدل‌های رگرسیون zero-inflated با توجه به داده‌های بیمه در اتومبیل مناسب است. مدل رگرسیون ZIDP بهترین شیوه برای داده‌ها را فراهم می‌آورد. در این مدل بخش پواسون در ZIP با استفاده از توزیع NB، GP و DP اصلاح‌شده است (Yip and Yau, 2005). در پژوهش یونس و همکاران دودسته ادعا در نظر گرفته‌شده

[4] vine copula
[5] Behavior-centric Vehicle
[6] Negative Binominol

www.SID.ir

است، یعنی آسیب دارایی شخص ثالث (TPPD) و آسیب شخصی (OD). دو مدل، یعنی فراوانی ادعا و شدت ادعا بر روی دودسته ادعاهای TPPD و OD مورد آزمایش قرار گرفتند. توسعه مدل شبکه عصبی بازگشت به عقب (BPNN) با استفاده از آزمون‌وخطا باهدف به دست آوردن بهترین نتیجه پیش‌بینی برای فراوانی ادعا و شدت ادعا آزمون می‌شود (Yunos *et al.*, 2016). در پژوهش لیو و همکاران بیان می‌کنند که الگوریتم آدابوست، یک روش پیش‌بینی محبوب و مؤثر برای پیش‌بینی فراوانی ادعای بیمه خودرو، نقش مهمی در شرکت‌های بیمه اموال دارد (Liu, Wang and Lv, 2014). در سنجش رابطه هزینه به‌عنوان متغیر پیوسته وابسته و سایر متغیرها در پژوهش هو و همکاران یک الگوریتم پیشنهاد می‌شود که به‌طور پویا برچسب مستمر در هر گره در طی روند القایی درخت را تشریح می‌کند. آزمایش‌های گسترده نشان می‌دهد که روش پیشنهادی از رویکرد پیش‌پردازش، رویکرد درخت رگرسیون و چندین الگوریتم مبتنی بر غیر درخت استفاده می‌کند. هدف در این مقاله توسعه یک الگوریتم نوآورانه DT، یعنی طبقه‌بندی کننده با برچسب مستمر[۷] (CLC)، برای انجام طبقه‌بندی بر روی برچسب‌های پیوسته عنوان‌شده است (Hu, Chen and Tang, 2009). طبق پژوهش چادو و همکاران فرضیه اساسی مدل پواسون کلاسیک لاندبرگ[۸] برای ارزیابی خسارات کلی در یک نمونه بیمه، استقلال بین وقوع یک ادعا و اندازه ادعای آن است. یک مدل رگرسیون مشترک با یک گاما GLM حاشیه‌ای برای متغیر پیوسته از متوسط اندازه ادعا و یک GLM پواسون حاشیه‌ای برای متغیر گسسته برای تعداد ادعا ایجاد کرده است. GLM ها به‌وسیله رویکرد ترکیبی ارتباطی با یک قاعده گاوسی ترکیب‌شده‌اند که دارای یک پارامتر برای ساختار وابستگی است (Czado et al, 2012). فریز و والدز طی پژوهشی مدل‌سازی دقیق و جزئی آماری رکورد بیمه‌های خودرو را توصیف می‌کنند. آن‌ها یک مدل سلسله مراتبی را برای سه جزء پیشنهاد می‌کنند که مربوط به فراوانی، نوع و شدت ادعا می‌شود. مدل اول، یک مدل رگرسیون دوجانبه منفی[۹] برای ارزیابی فراوانی ادعا است. دومین مدل لوجیت چندجمله‌ای[۱۰] برای پیش‌بینی نوع ادعای بیمه است. مدل سوم برای جزء شدت است. در اینجا از بتا تعمیم‌یافته نوع دوم توزیع دم طولانی[۱۱] برای مقادیر ادعا استفاده و متغیرهای پیش‌بینی را نیز ترکیب می‌کند. وابستگی معنی‌دار میان انواع ادعاهای مختلف را با استفاده از T-copula اندازه‌گیری می‌کند. از توزیع GB2 برای جا دادن ماهیت دم طولانی ادعاها استفاده‌شده است

[7] Continuous Label Classifier
[8] Lundberg
[9] Negative binomial regression model
[10] Multinomial logit model
[11] Long-tailed distribution

www.SID.ir

(Frees and Valdez, 2008). یانگ و همکاران یک الگوریتم بوستینگ درخت گرادیانت[12] برای پیش‌بینی را پیشنهاد می‌دهند و آن را برای مدل‌های پواسون ترکیبی تودی برای حق بیمه‌های خالص اعمال می‌کنند (Yang, Qian and Zou, 2018). گیلنکو و میرونوا عوامل فاجعه‌ای که تأثیر بسزایی در فراوانی ادعا و شدت ادعا در بیمه آسیب خودرو روسیه دارد را برای ارزیابی کار آیی مجموعه عوامل موجود مورداستفاده برای محاسبه‌های تعرفه بیمه مدرن، تحلیل می‌کنند. دو مدل اصلی مدل hurdle سنتی و GLM با توزیع گاما (مدل توزیع GLM-Gamma برای سادگی نامیده می‌شود) این مطالعه را تشکیل می‌دهند (Gilenko and Mironova, 2017). تاکوچی و همکاران در پژوهش خود بیان می‌کنند که روش‌های رگرسیونی باثبات سنتی فرض می‌کنند که توزیع نویز متقارن است و این به معنای نادیده گرفتن تأثیر نقاط پرت است. وقتی توزیع نویز نامتقارن است، این روش‌ها برآوردگرهای رگرسیون بایاس را منجر می‌شود. درنتیجه با توجه به مشکلات داده‌کاوی در زمینه صنعت بیمه، یک رویکرد جدید برای رگرسیون باثبات برای مقابله با توزیع نویز نامتقارن پیشنهاد می‌دهند که می‌توان آن را به هر دو رگرسیون خطی و غیرخطی اعمال کرد (Takeuchi, Bengio and Kanamori, 2002). اسمیت و همکاران در پژوهشی به حفظ مشتری و تعیین سطح قیمت پرداخته‌اند. این پژوهش به سه بخش اصلی تقسیم‌شده است. طی بخش اول برای تجزیه‌وتحلیل داده‌ها رگرسیون، درخت تصمیم و شبکه عصبی باهم مقایسه شده‌اند. درنتیجه آن شبکه عصبی بهترین نتایج را برای طبقه‌بندی بیمه‌نامه‌ها به‌عنوان احتمال اتمام یا ادامه قرارداد به دست می‌دهد. ازآنجایی‌که قیمت‌گذاری بر اساس میزان ریسک‌پذیری دارندگان بیمه‌نامه صورت می‌گیرد بنابراین در مرحله دوم گروه‌های ریسک‌پذیر با استفاده از الگوریتم‌های خوشه‌بندی مشخص شدند. برای اینکه سودآوری هر خوشه موردبررسی قرار بگیرد، شاخص‌های عملکرد کلیدی با عنوان نسبت هزینه و نسبت تکرار برای هر خوشه بررسی‌شده و گروه‌های با نرخ هزینه‌های بالا و غیرقابل‌قبول (سودآوری پایین) در این روش شناسایی می‌شوند. درنهایت، نتایج حاصل از این دو تحلیل در یک چارچوب برای تعیین قیمت حق بیمه بهینه برای یک بیمه‌نامه منحصربه‌فرد ترکیب خواهد شد (Smith, Willis and Brooks, 2000). شاخص کلیدی نسبت هزینه در اینجا هزینه‌های ادعای خسارت دارنده بیمه‌نامه را نسبت به‌حق بیمه پرداخت‌شده می‌سنجد. همچنین شاخص کلیدی نرخ تکرار، میزان نسبت تعداد ادعا به تعداد واحدهای بیمه‌نامه را اندازه‌گیری می‌کند. در مقاله دوم که اسمیت و همکاران در سال ۲۰۰۲ منتشر کردند، گام سوم این پژوهش یعنی ارائه یک چارچوب برای تعیین حق بیمه بهینه انجام شد. طی مقاله دوم سعی بر آن شده

[12] Gradient tree-boosting algorithm

www.SID.ir

که درآمد بدون تأثیر نرخ فسخ بیمه‌نامه و سهم بازار افزایش یابد. طی این مقاله ترکیب بهینه دارندگان بیمه‌نامه برای تعیین نرخ فسخ به‌وسیله برنامه‌ریزی عدد صحیح غیرخطی بر اساس تئوری پورتفولیو مارکوویتز[۱۳] تعیین‌شده است (Smith *et al.*, 2002). پژوهش دیگری که در جهت حفظ مشتری و مدیریت ارتباط با مشتری (CRM) انجام‌شده است. خوشه‌بندی دومرحله‌ای یک فرم تجزیه خوشه‌ای را فراهم می‌کند. این می‌تواند برای خوشه‌بندی مجموعه داده‌ها به گروه‌های متمایز زمانی که ما در ابتدا نمی‌دانیم که چه کسانی هستند، مورداستفاده قرار گیرد. تجزیه‌وتحلیل خوشه‌ای دومرحله‌ای نشان می‌دهد که چهار گروه وجود دارد که با معیار اطلاعات Bayesian توسعه داده‌شده‌اند. بعدها از روش تصمیم‌گیری درخت با استراتژی CHAID برای شناسایی نمایه مشتری استفاده نمودند. در مرحله بعدی ارتباط بین رضایت مشتری و کد مؤلفه برای شناسایی بیمه‌نامه‌های نامطمئن در بازار بکار رفت (Bhatnagar, Ranjan and Singh, 2011). در پژوهش دیگری پیش‌بینی روی‌گردانی، حفظ و دگر فروشی[۱۴] بر اساس پیشنهاد قیمت‌های[۱۵] آنلاین از یک وب‌سایت بیمه‌گر در شرایط خدمت به‌عنوان یک رویداد هدف ارزیابی‌شده است. بر اساس چنین نمونه‌هایی، یک مدل طبقه‌بندی تصادفی جنگل را برای پیش‌بینی فعالیت‌های خرید استفاده می‌کنند (Mau, Pletikosa and Wagner, 2018). مقاله آنتنیو و ولیدز یک مقاله پژوهشی در مورد ابزارهای آماری برای طبقه‌بندی خطر مورداستفاده در بیمه است. در انتخاب متغیرهای رتبه‌بندی برای اهداف طبقه‌بندی خطر، ابزار آماری اولیه درون مدل‌های رگرسیون قرار می‌گیرند (Antonio and Valdez, 2012).

بحث و یافته‌ها

بر اساس بررسی و مطالعه منابع اطلاعاتی از طریق سایت‌های علمی و پژوهشی، درمجموع ۲۵ مقاله مطالعه شد. با توجه به نتایج حاصل در این مقالات به‌طورکلی دودسته تعریف شد. این دودسته شامل حفظ مشتری و تعیین حق بیمه است. ۲۰ مقاله مربوط به تعیین حق بیمه بهینه و ۵ مقاله مربوط به حفظ مشتری بود.

بنابر نتایج به‌دست‌آمده در جهت حفظ مشتریان بر اساس اطلاعات روی‌گردانی مشتریان در مورد بیمه خودرو، سیستم شاخص در سه جنبه ساخته‌شده است: اطلاعات مشتری، موضوع اطلاعات بیمه و اطلاعات مربوط به محصول (Song and Qiuhong, 2018). همچنین تعیین حق بیمه و طراحی یک بیمه‌نامه جدید می‌تواند بر اساس اطلاعات

[13] Markowits, پدر علم نوین تئوری پورتفولیو شناخته می‌شود
[14] Cross-selling
[15] Quotes

www.SID.ir

جمعیت شناختی مشتریان انجام بشود و شرکت بیمه طرح‌هایی را برای مشتری خود با مزایای بیشتری طراحی کند (Kulkarni, R. V., & Devale, 2012). برای طراحی حق بیمه برای وسایل نقلیه موتوری از پیش‌بینی احتمال و ارزش مورد انتظار برای ادعاهای آینده برای هر بیمه‌شده بر اساس چندین ویژگی موجود در پایگاه داده بر روی مشتریان و همتایان بیمه‌گر استفاده می‌شود (Kahane *et al.*, 2007). درواقع در بخش زیادی از تحقیقات تعداد ادعاها و میزان کل ادعای بیمه‌گذاران برای تعیین حق بیمه به‌کاررفته‌اند. فاکتورهایی که بر اندازه و احتمالات ادعا تأثیر می‌گذارد، شناسایی و نتایج این روش‌ها مقایسه می‌شود که هردوی آن‌ها خروجی را پیش‌بینی می‌کنند (Bortoluzzo *et al.*, 2011). در تجارت بیمه، مهم‌ترین ملاحظات در قیمت‌گذاری بیمه‌گذار عبارت‌اند از: فراوانی خسارت و شدت خسارت یعنی احتمال از دست دادن و ارزش مالی از دست دادن (Kitchens and Harris, 2015). مدل قیمت‌گذاری بیمه خودرو مبتنی بر رفتار[16] (BVIP) و یک نمونه اولیه محاسبه حق بیمه خودرو برای بیمه خودرو شخصی نوع دیگری از قیمت‌گذاری حق بیمه است که توسط پژوهشگران پیشنهادشده است (Bian *et al.*, 2018). پرداخت بیمه به ازای میزان رانندگی روشی است که در برخی کشورها جهت محاسبه قیمت بیمه خودرو مورداستفاده قرار می‌گیرد (Ferreira and Minikel, 2012).

اهمیت و مفهوم تکنیک‌ها و ابزار DM برای مدیریت ارتباط با مشتری (CRM) با پیدا کردن اطلاعات پنهان و ناشناخته از داده‌های واقعی شرکت بیمه مبحث دیگری است که در حفظ مشتری قابل‌بحث است. مباحث از طریق سطح رضایتمندی مشتریان، روی‌گردان‌ها و غیر روی‌گردان‌ها و سایر اطلاعاتی که باعث افزایش سودآوری سازمان و رضایت مشتری شده است، تحلیل می‌شود (Bhatnagar, Ranjan and Singh, 2011).

پیش‌بینی روی‌گردانی، حفظ و دگر فروشی[17] بر اساس پیشنهاد قیمت‌های[18] آنلاین از یک وب‌سایت بیمه‌گر در شرایط خدمت به‌عنوان یک رویداد هدف نیز مبحث دیگر است که ارزیابی‌شده است. این مطالعه اهمیت انتخاب منابع اطلاعات مربوطه را برجسته می‌کند و مشتریان مناسب را در زمان مناسب هدف قرار می‌دهند و از این طریق از شیوه‌های CRM تحلیلی بهره‌مند می‌شوند (Mau, Pletikosa and Wagner, 2018).

تکنیک‌های مورداستفاده در این پژوهش‌ها شامل طبقه‌بندی، خوشه‌بندی، رگرسیون، باهم آیی و همبستگی است. الگوریتم‌های مختلف این تکنیک‌ها در ترکیب با یکدیگر مورداستفاده قرارگرفته‌اند. همچنین الگوریتم‌های

[16] Behavior-centric Vehicle
[17] Cross-selling
[18] Quotes

www.SID.ir

جزئی‌تر به‌کاررفته در این مقالات مشخص شدند. با تکنیک‌ها و الگوریتم‌های داده‌کاوی مورداستفاده در کنار منابع مرجع در جدول پیوست دسته‌بندی‌شده‌اند. در مقایسه دو مدل درخت تصمیم و رگرسیون لجستیک برای تجزیه‌وتحلیل داده‌های مشتری، نرخ دقت تفاوت زیادی ندارد، اما درخت تصمیم‌گیری دقیق‌تر است (Mau, Pletikosa and Wagner, 2018). باید توجه داشت که محاسب‌های نرخ بیمه به‌عنوان متخصصان حرفه‌ای در ارزیابی اقتصادی حوادث نامعلوم در نظر گرفته می‌شوند و مجهز به ابزارهای آماری زیادی جهت تجزیه‌وتحلیل و تعریف متناسب تعرفه مربوط به خطرات هستند. این دسته‌بندی می‌تواند به‌عنوان راهنمایی برای خبرگان بیمه و همچنین محققان در حوزه داده‌کاوی مورداستفاده قرار بگیرد.

نتیجه‌گیری

پیشرفت مستمر فناوری در عصر حاضر سبب افزایش حجم و پیچیدگی داده‌های بیمه‌ای در حوزه‌های مختلف آن شده است. تحلیل این حجم وسیع داده‌ها نتایج جالب‌توجهی ایجاد می‌کند که می‌تواند برای بقای شرکت‌های بیمه اهمیت زیادی داشته باشد. در این راستا، شرکت‌های بیمه‌ای نیازمند استفاده از فناوری‌هایی هستند که بر مبنای آن در لایه‌های تراکنشی و همچنین عملیاتی خود بتوانند داده‌ها را استخراج کنند، انتقال دهند و تبدیل به اطلاعات نمایند. داده‌کاوی به‌عنوان یک موضوع بین‌رشته‌ای که از فن‌های مختلف برای استخراج اطلاعات از پایگاه داده استفاده می‌کند می‌تواند در تجزیه‌وتحلیل اطلاعات و سازمان‌دهی آن‌ها مورداستفاده قرار بگیرد. محاسب‌های نرخ بیمه[19] به‌عنوان متخصصان حرفه‌ای در ارزیابی اقتصادی حوادث نامعلوم در نظر گرفته می‌شوند و مجهز به ابزارهای آماری زیادی جهت تجزیه‌وتحلیل و تعریف متناسب تعرفه مربوط به خطرات هستند. درواقع داده‌کاوی ابزاری است که می‌تواند در اختیار محاسب‌های نرخ بیمه قرار بگیرد و روند ایجاد تعرفه‌های منصفانه را تسهیل نماید. بخش مهمی از روند ایجاد تعرفه‌های بیمه‌ای منصفانه، طبقه‌بندی ریسک است که شامل دسته‌بندی ریسک‌ها در کلاس‌های مختلف است که مجموعه‌ای از ویژگی‌های یکسان را به اشتراک می‌گذارند و به‌این‌ترتیب می‌توانند به‌صورت محرمانه قیمت را تبعیض نمایند.

همان‌گونه که پیش‌ازاین نیز اشاره شد، فراوانی ادعای مورد انتظار و شدت مورد انتظار در مدل‌سازی پیش‌بینی کننده برای ادعاهای بیمه اتومبیل مورداستفاده قرار می‌گیرند. سنجش این عوامل با کمک داده‌کاوی، چنانچه در پیشینه پژوهش بیان شد، می‌تواند انجام شود.

---

[19] Actuaries

www.SID.ir

ویژگی اصلی یک اقدام بیمه‌ای این است که حق بیمه را در ابتدای قرارداد بیمه تعیین کنید. برای تعیین حق بیمه دقیق سالیانه در شرکت بیمه، تخمین دقیق و قابل‌اعتماد از تعداد ادعاهایشان و میزان کل ادعا بسیار مهم است. پیش‌بینی فرآیند ادعای آینده در بیمه برای تعیین حق بیمه در ابتدای قرارداد بیمه قابل‌توجه است. تنظیم سطح قیمت متناسب با هزینه‌ها و هم‌زمان با آن حفظ مشتری موجود و جذب مشتریان جدید مسئله‌ای است که در صنعت بیمه همچون دیگر صنایع باید به آن توجه شود. در صنعت بیمه، حق بیمه با استفاده از فرمول‌های تعیین‌شده بر اساس تقسیم دارندگان بیمه‌نامه با استفاده از اطلاعات جمعیت شناختی مانند سن راننده، میزان ریسک، سن ماشین و غیره تعیین می‌شود.

در این مقاله سعی بر آن شد که کاربردهای داده‌کاوی در تعیین حق بیمه بهینه نشان داده شود. بنابراین به‌طورکلی کاربردها به دودسته تقسیم و طی جدول زیر در اختیار پژوهشگران قرار گرفت. چنانچه در جدول نشان داده‌شده، تکنیک‌ها و الگوریتم‌های مورداستفاده در پژوهش‌های پیشین دسته‌بندی و در مقابل منابع آن‌ها ذکرشده است.

جدول ۱الگوریتم ها و تکنیک ها

منبع	الگوریتم‌ها	تکنیک‌ها	کاربرد
Yeo et al., 2002	k-means، شبکه عصبی، برنامه‌ریزی عدد صحیح غیرخطی	خوشه‌بندی، طبقه‌بندی، رگرسیون	تعیین حق بیمه
Kahane, Levin, Meiri, et al., 2007	مدل تحلیل بقا و یک مدل رگرسیون خطی		
Sakthivel & Rajitha, 2017	شبکه عصبی مصنوعی (ANN) و شبکه عصبی پیش‌خور		
David, 2015	مدل‌های خطی کلی ازجمله مدل پواسون و مدل گاما		
Shen, Odening & Okhrin, 2013	یک چارچوب بیزین		
Bortoluzzo, Claro, Caetano, et al., 2011	روش تودی و گوسی معکوس تنظیم‌شده با صفر ZAIG		
Kitchens & Harris, 2015	شبکه عصبی منطبق با الگوریتم ژنتیک		

www.SID.ir

منبع	الگوریتم‌ها	تکنیک‌ها	کاربرد
Bian, Yang, Zhao, et al., 2018	مدل طبقه‌بندی سطح ریسک رفتار محور		
Guelman, 2012	گرادیانت بوستینگ		
Ferreira & Minikel, 2012	پواسون و رگرسیون خطی		
Yip & Yau, 2005	ترکیبی از چندین توزیع تعداد نامحدود پارامتری، ازجمله ZIP، ZINB، ZIGP و ZIDP		
Yunos, Ali, Shamsyuddin, et al., 2016	شبکه عصبی بازگشت به عقب (BPNN)		
Antonio & Valdez, 2012	رگرسیون تعمیم یافته و خطی، مدل های zero-inflated و hurdle		
Liu, Wang & Lv, 2014	درخت AdaBoost چند کلاسه		
Hu, Chen & Tang, 2009	رویکرد پیش‌پردازش Continuous Label Classifier، رویکرد درخت رگرسیون و چندین الگوریتم مبتنی بر غیر درخت (nontree-based algorithms)		
Czedo, Kastenmeier, Brechmann, et al., 2012	مدل رگرسیون ترکیبی ارتباطی		
Frees & Valdez, 2008	مدل رگرسیون دوجمله ای منفی، مدل لجیت چندجمله ای، توزیع دم سنگین		
Yang, Qian & Zou, 2018	الگوریتم درخت گرادیانت بوستینگ		
Gilenko & Mironova, 2017	مدل hurdle سنتی و رگرسیون خطی تعمیم یافته با توزیع، تجزیه‌وتحلیل خوشه سلسله مراتبی		

www.SID.ir

منبع	الگوریتم‌ها	تکنیک‌ها	کاربرد
Takeuchi, Bengio & Kanamori, 2002	رگرسیون روبوست با دم سنگین نامتقارن		
Smith, Willis & Brooks, 2000	الگوریتم شبکه عصبی پیشخور سه لایه و نیز الگوریتم K-means با حداقل معیارهای مربع	خوشه‌بندی، طبقه‌بندی، باهم‌آیی، همبستگی	حفظ مشتری
Bhatnagar, Ranjan & Singh, 2011	خوشه‌بندی، طبقه‌بندی و باهم آیی،CHAID		
Song & Qiuhong, 2018	درخت تصمیم و رگرسیون لجستیک		
Kulkami & Devale, 2012	قوانین انجمنی، خوشه‌بندی، طبقه‌بندی و همبستگی مناسب		
Mau, Pletikosa & Wagner, 2018	الگوریتم جنگل تصادفی		

نتایج به‌دست‌آمده از پژوهش حاضر در بهبود سیستم‌های عملیاتی و تصمیم‌گیری مدیران می‌تواند مفید واقع شود. یکی از مسائل اصلی که در صنعت بیمه مطرح است، ایجاد یک سیستم پشتیبان تصمیم برای مشاوران بیمه در تعیین یک بیمه‌نامه مناسب برای مشتریان هست که نتایج این پژوهش می‌تواند برای در این جهت مفید باشد. همچنین مدیران می‌توانند در شناسایی فاکتورهایی که بر اندازه و احتمالات ادعا تأثیر می‌گذارد، از نتایج این پژوهش استفاده کنند. در این پژوهش سعی شده جزییات ویژگی‌های رفتاری مشتریان شناسایی شوند. تمامی این ویژگی‌ها در مقالات مرجع موردبررسی قرارگرفته‌اند. این اطلاعات کمک می‌کند که مدیران بتوانند محصولات بیمه‌ای و خدمات شخصی‌سازی شده به مشتریان ارائه بدهند. درنهایت منجر به بهبود کار آیی سازمان بشود.

روش‌های مورداستفاده در هرکدام از مقالات پیشینه می‌تواند راهنمایی برای محاسب‌های بیمه در جهت تعیین حق بیمه بهینه مورداستفاده قرار بگیرد. همچنین در پژوهش‌های حوزه بیمه با رویکرد داده‌کاوی، تکنیک‌ها و الگوریتم‌هایی که طی این پژوهش معرفی‌شده است، می‌تواند توسط پژوهشگران مورداستفاده قرار بگیرد.

www.SID.ir

لازم به ذکر است که روش پیشنهادی می‌تواند فراتر از پیش‌بینی حق بیمه باشد و برای محققان در بسیاری از زمینه‌های دیگر از قبیل محیط‌زیست، هواشناسی و علوم سیاسی قابل‌استفاده است.

منابع:

1. Antonio, K. and Valdez, E. A. (2012) 'Statistical concepts of a priori and a posteriori risk classification in insurance', *AStA Advances in Statistical Analysis*, 96(2), pp. 187–224. doi: 10.1007/s10182-011-0152-7.

2. Bhatnagar, V., Ranjan, J. and Singh, R. (2011) 'Analytical customer relationship management in insurance industry using data mining: a case study of Indian insurance company', *International Journal of Networking and Virtual Organisations*, 9(4), p. 331. doi: 10.1504/IJNVO.2011.043803.

3. Bian, Y. *et al.* (2018) 'Good drivers pay less: A study of usage-based vehicle insurance models', *Transportation Research Part A: Policy and Practice*, 107, pp. 20–34. doi: 10.1016/j.tra.2017.10.018.

4. Bortoluzzo, A. B. *et al.* (2011) 'Estimating total claim size in the auto insurance industry: A comparison between tweedie and zero-adjusted inverse Gaussian distribution', *BAR - Brazilian Administration Review*, 8(1), pp. 37–47. doi: 10.1590/S1807-76922011000100004.

5. Czado, C. *et al.* (2012) 'A mixed copula model for insurance claims and claim sizes', *Scandinavian Actuarial Journal*, 2012(4), pp. 278–305. doi: 10.1080/03461238.2010.546147.

6. David, M. (2015) 'Auto Insurance Premium Calculation Using Generalized Linear Models', *Procedia Economics and Finance*, 20, pp. 147–156. doi: 10.1016/S2212-5671(15)00059-3.

7. Ferreira, J. and Minikel, E. (2012) 'Measuring Per Mile Risk for Pay-As-You-Drive Automobile Insurance', *Transportation Research Record: Journal of the Transportation Research Board*, 2297(1), pp. 97–103. doi: 10.3141/2297-12.

8. Frees, E. W. and Valdez, E. A. (2008) 'Hierarchical insurance claims modeling', *Journal of the American Statistical Association*, 103(484), pp. 1457–1469. doi: 10.1198/016214508000000823.

www.SID.ir

9. Gilenko, E. V. and Mironova, E. A. (2017) 'Modern claim frequency and claim severity models: An application to the Russian motor own damage insurance market', *Cogent Economics and Finance*. Edited by B. Spagnolo. Cogent, 5(1). doi: 10.1080/23322039.2017.1311097.

10. Guelman, L. (2012) 'Gradient boosting trees for auto insurance loss cost modeling and prediction', *Expert Systems with Applications*, 39(3), pp. 3659–3667. doi: 10.1016/j.eswa.2011.09.058.

11. Hu, H. W., Chen, Y. L. and Tang, K. (2009) 'A dynamic discretization approach for constructing decision trees with a continuous label', *IEEE Transactions on Knowledge and Data Engineering*, 21(11), pp. 1505–1514. doi: 10.1109/TKDE.2009.24.

12. Kahane, Y. *et al.* (2007) 'Applying Data Mining Technology for Insurance Rate Making: An Example of Automobile Insurance', *Asia-Pacific Journal of Risk and Insurance*. De Gruyter, 2(1). doi: 10.2202/2153-3792.1014.

13. Kitchens, F. and Harris, T. (2015) *Genetic Adaptive Neural Networks for Prediction of Insurance Claims, International Journal of Engineering and Advanced Research Technology (IJEART)*. Available at: www.ijeart.com (Accessed: 1 August 2018).

14. Kulkarni, R. V., & Devale, A. B. (2012) 'Applications of Data Mining Techniques in Life Insurance', *International Journal of Data Mining & Knowledge Management Process*, 2(4), pp. 31–40. doi: 10.5121/ijdkp.2012.2404.

15. Liu, Y., Wang, B.-J. and Lv, S.-G. (2014) 'Using Multi-class AdaBoost Tree for Prediction Frequency of Auto Insurance', *Journal of Applied Finance & Banking*. Scienpress Ltd, 4(5), pp. 45–53. Available at: http://www.scienpress.com/Upload/JAFB%2FVol 4_5_4.pdf (Accessed: 2 June 2018).

16. Mau, S., Pletikosa, I. and Wagner, J. (2018) 'Forecasting the next likely purchase events of insurance customers: A case study on the value of data-rich multichannel environments', *International Journal of Bank Marketing*, 5 June, p. IJBM-11-2016-0180. doi: 10.1108/IJBM-11-2016-0180.

17. Sakthivel, K. M. and Rajitha, C. S. (2017) 'Artificial Intelligence for Estimation of Future

www.SID.ir

Claim Frequency in Non-Life Insurance', *Global Journal of Pure and Applied Mathematics*, 13(6), pp. 973–1768. Available at: http://www.ripublication.com (Accessed: 1 August 2018).

18. Shen, Z., Odening, M. and Okhrin, O. (2013) 'Can expert knowledge compensate for data scarcity in crop insurance pricing ? Can expert knowledge compensate for data scarcity in crop insurance pricing ? a', *academic.oup.com*, pp. 1–19. Available at: https://academic.oup.com/erae/article-abstract/43/2/237/2367166 (Accessed: 1 June 2018).

19. Smith, K. A. *et al.* (2002) 'A Mathematical Programming Approach to Optimise Insurance Premium Pricing within a Data Mining Framework', *The Journal of the Operational Research Society*. Palgrave Macmillan Journals, 53(11), pp. 1197–1203. doi: 10.1057/palgravejors.2601413.

20. Smith, K. A., Willis, R. J. and Brooks, M. (2000) 'An analysis of customer retention and insurance claim patterns using data mining: A case study', *Journal of the Operational Research Society*. Taylor & Francis, 51(5), pp. 532–541. doi: 10.1057/palgrave.jors.2600941.

21. Song, H. and Qiuhong, H. (2018) 'Application of Data Mining Technology in the Loss of Customers in Automobile Insurance Enterprises', *International Journal of Data Science and Analysis*, 4(1), pp. 1–5. doi: 10.11648/j.ijdsa.20180401.11.

22. Sumathi, S. and Sivanandam, S. N. (2006) *Introduction to Data Mining and its Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg (Studies in Computational Intelligence). doi: 10.1007/978-3-540-34351-6.

23. Takeuchi, I., Bengio, Y. and Kanamori, T. (2002) 'Robust regression with asymmetric heavy-tail noise distributions', *Neural Computation*. MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046 USA journals-info@mit.edu, 14(10), pp. 2469–2496. doi: 10.1162/08997660260293300.

24. Yang, Y., Qian, W. and Zou, H. (2018) 'Insurance Premium Prediction via Gradient Tree-Boosted Tweedie Compound Poisson Models', *Journal of Business and Economic Statistics*, 36(3), pp. 456–470. doi: 10.1080/07350015.2016.1200981.

25. Yip, K. C. H. and Yau, K. K. W. (2005) 'On modeling claim frequency data in general

www.SID.ir

insurance with extra zeros', *Insurance: Mathematics and Economics*, 36(2), pp. 153–163. doi: 10.1016/j.insmatheco.2004.11.002.

26. Yunos, Z. M. *et al.* (2016) 'Predictive Modelling for Motor Insurance Claims Using Artificial Neural Networks', *International Journal of Advance Soft Computing Applications*, 8(3), pp. 160–172. Available at: http://home.ijasca.com/data/documents/ID38_Pg160-172_Predictive-Modelling-for-Motor-Insurance-Claims-Using-Artificial-Neural-Networks_2.pdf (Accessed: 1 June 2018).

www.SID.ir