پژوهشنامهٔ بیمه/ سال سی‌ام / شمارهٔ ٤/ زمستان ١٣٩٤/ شمارهٔ مسلسل ١٢٠/ صفحات ١ تا ٢٠

ارائهٔ راهکاری برای تعرفه‌گذاری پویا در صنعت بیمه با استفاده از تکنیک داده‌کاوی (مورد مطالعه: بیمهٔ شخص ثالث)

تاریخ دریافت: ١٣٩٣/١٢/٢١
تاریخ پذیرش: ١٣٩٤/١١/٢٦

داود کریم‌زادگان مقدم[١]
مجید بهروان[٢]

چکیده

بیمه‌نامهٔ شخص ثالث بیشترین سهم از بازار بیمهٔ کشور را دارا می‌باشد و فرصت مناسبی برای کاوش اطلاعات و استخراج الگوهای ناشناخته جهت تصمیمات کلان در صنعت بیمه را فراهم می‌نماید. در حال حاضر حق بیمه با کمترین توجه به عوامل ریسک بیمه‌گذاران محاسبه می‌گردد، که موجب زیانده شدن بیمه‌نامهٔ شخص ثالث برای شرکت‌های بیمه و نارضایتی بیمه‌گذاران از خدمات شرکت‌های بیمه گردیده است. بدین منظور در این پژوهش، اطلاعات خودرویی، سوابق بیمه‌ای و ویژگی‌های بیمه‌گذاران در بیش از ٣٠ میلیون بیمه‌نامه و ٧/٢ میلیون خسارت جمع‌آوری و استانداردسازی شده و در داده‌انبار ذخیره گردید. برای استانداردسازی داده‌های خودرویی با بانک اطلاعاتی نیروی انتظامی و اطلاعات هویتی بیمه‌گذاران با استفاده از داده‌های سازمان ثبت احوال کشور، اعتبارسنجی و تکمیل گردیده است. سپس ساختار کاوشی طراحی و با استفاده از سه الگوریتم خوشه‌بندی، شبکهٔ عصبی و درخت تصمیم و داده‌های آموزشی مورد آموزش قرار گرفت. در نهایت مدل‌ها با استفاده از داده‌های آزمایشی مورد بررسی قرار گرفته و نتایج به‌دست‌آمده از مدل‌ها با استفاده از ماتریس آشفتگی و نسبت خسارت، مورد اعتبارسنجی قرار گرفتند، که نتایج به‌دست‌آمده نشان‌دهندهٔ امکان استفاده از روش ارائه‌شده در تعرفه‌گذاری پویا در خصوص بیمهٔ شخص ثالث به‌صورتی کارآمد را نشان می‌دهد، به‌نحوی که نسبت خسارت، کاهش می‌یابد و ماتریس آشفتگی، صحت ارزیابی را نشان می‌دهد.

واژگان کلیدی: *داده‌کاوی، داده‌انبار، بیمه‌نامهٔ اتومبیل، حق بیمه، تعرفه‌گذاری پویا*

---

١. عضو هیئت علمی دانشگاه پیام نور، گروه مهندسی کامپیوتر و فناوری اطلاعات
d_karimzadgan@pnu.ac.ir
٢. دانشجوی کارشناسی ارشد مدیریت فناوری اطلاعات دانشگاه پیام نور(نویسنده مسئول)
majid.behravan@hotmail.com

۱. مقدمه

بیمه[1] راهکاری است که ریسک را به طرفی انتقال می‌دهد که در ریسک مزیت نسبی دارد. در مثال معمول بیمه، شخص ریسک‌پذیر[2] (مواجه با ریسک)، تمایل دارد که قیمت ثابتی را به بیمه‌گر کمتر ریسک‌گریز[3] یا بیمه‌گر دارای تنوع ریسک بپردازد، که حاضر است ریسک را به آن قیمت تقبل کند. از آن‌جا که دو طرف در مورد قرارداد توافق دارند، وضع هر دو بهتر خواهد شد (Dionne, 2005).

بیمهٔ شخص ثالث، یکی از مهم‌ترین خدمات بیمه‌ای است و برخی از آن به‌عنوان ویترین شرکت‌های بیمه یاد می‌کنند. در واقع این بیمه‌نامه به‌دلیل اجباری بودن می‌تواند به‌عنوان دروازهٔ ورود و جذب مشتری به یک شرکت بیمه برای فروش سایر پوشش‌های بیمه‌ای نیز باشد. بنابراین رضایت بیمه‌گذاران[4] در این شاخه از خدمات بیمه‌ای، می‌تواند اثر قابل توجهی بر حضور دائم بیمه‌گذار بگذارد و موجب سودآوری بیشتر گردد (رنجبر فرد، ۱۳۹۳).

ورشکستگی شرکت‌های بیمه در مقایسه با شرکت‌ها در سایر صنایع می‌تواند آثار مخرب‌تری بر مشتریان این صنعت و جامعه به‌صورت کلّی داشته باشد (دقیقی، ۱۳۹۲). لذا باید با ایجاد بستر تعرفه‌گذاری صحیح، ریسک شرکت‌های بیمه را مدیریت کرد.

با توجه به رویکرد جدید در آزادسازی نرخ‌ها در صنعت بیمه در کشور، ضرورت بررسی علمی این موضوع و ارائهٔ راهکاری برای تعرفه‌گذاری[5] متناسب و بر حسب شرایط بیمه‌گذاران اجتناب‌ناپذیر خواهد بود.

---

[1] insurance
[2] risk taker
[3] risk aversion
[4] insurer
[5] pricing

۲. بیان مسئله

قانون بیمهٔ اجباری مسئولیت مدنی دارندگان وسایل نقلیهٔ موتوری زمینی در مقابل شخص ثالث که در ایران به‌اختصار، قانون بیمهٔ شخص ثالث نامیده می‌شود، اولین بار در ١٤ ماده و سه تبصره و در تاریخ ١٣٤٧/٩/٢٦ به تصویب مجلس شورای ملی و متعاقباً در تاریخ ١٣٤٧/١٠/٢٣ به تصویب مجلس سنا رسیده و از اول فروردین ١٣٤٨ نیز به اجرا گذاشته شد.

اما پس از اصلاحات این قانون، همچنان نحوهٔ محاسبه میزان حق بیمهٔ اتومبیل بسیار نامتوازن و بر اساس مسائل سیاسی کشور نرخ‌گذاری گردیده است و کمتر جنبه فنی نرخ‌گذاری در آن رعایت شده است که موجب زیانده شدن این رشته در صنعت بیمه گردیده است.

از دید اقتصاد بیمه، میزان ارزش افزودهٔ ناشی از تجمیع ریسک با استفاده از مفهوم حق بیمه، قابل اندازه‌گیری است. بدین معنا که اگر فرض کنیم فرد ریسک‌پذیری با درآمد Y مواجه است و یا این امکان را دارد که درآمدی معادل Y_0 را با اطمینان کامل انتخاب کند، حاضر است Y_0 را که کمتر از میانگین Y یعنی E(Y) است برگزیند، در این صورت تفاوت بین Y_0 و E(Y) در واقع حکم حق بیمه را پیدا می‌کند. به عبارت دیگر حق بیمه مبلغی است که تصمیم‌گیرنده را بین پذیرش ریسک و بیمه کردن آن بی‌تفاوت می‌نماید (خرمی، ١٣٨٢).

حق بیمه (π) عبارت از حل رابطهٔ زیر است (خرمی، ١٣٨٢):

$$U(W-\mu_L-\pi)=\int U(W-L)+F(L)\,d \qquad (١)$$

که در آن:

U(W)= تابع مطلوبیت U'>0, U''<0

W= ثروت اولیه (غیر تصادفی= غیر احتمالی)

L= L≥0 خسارت (تصادفی = احتمالی)

f(L)= تابع توزیع احتمال خسارت

$E(L)= \mu_L$= میانگین خسارت احتمالی.

بنابراین می‌توان گفت که ارزش افزودهٔ ناشی از انعقاد قرارداد بیمه، عبارت است از حداکثر ارزش حاصله بیش از خسارت انتظاری که بیمه‌گذار حاضر به پرداخت آن است؛ زیرا بیمه‌گذار مختار است که بیمه خریداری نکند و در عوض ثروت ریسکی را در اختیار داشته باشد که ارزش انتظاری است که بیمه‌گذار حاضر به پرداخت آن بوده، مبیّن ارزش بیمه است (خرمی، ١٣٨٢).

برای تعیین حق بیمه نخست باید مقدار مورد انتظار ادعای خسارت برای هر قرارداد بیمه را برآورد کرد. بدین منظور، تابع شرطی P(x) که در آن متغیر ورودی x، اطلاعات موجود هر بیمه‌گذار می‌باشد و میزان خسارت هر بیمه‌گذار را بر اساس شرایط خاص او محاسبه می‌نماید، مبنای محاسبه حق بیمهٔ خالص برای بیمه‌گذار خواهد بود (Chapados, 2010).

$$P(X) = E[A|X] \qquad (٢)$$

متغیر A میزان خسارت بیمه‌گذار به‌ازای شرایط X را نشان می‌دهد. در واقع میزان خسارت تمامی بیمه‌گذاران با شرایط یکسان، خروجی یکسانی را خواهد داشت.

٣. داده‌کاوی

موضوع داده‌کاوی شناخت چیزهای جدید و باارزش، بالقوه مفید، رابطه‌های منطقی و الگوهای موجود در داده‌ها است. در جوامع مختلف، یافتن الگوهای مفید در داده‌ها با عناوین متعدد بیان می‌شود. برای مثال از عنوان‌هایی نظیر استخراج دانش، کشف اطلاعات، برداشت اطلاعات، پردازش الگوهای داده‌ها (Fayyad, 1996) می‌توان نام برد.

هر نویسنده، محقق و کاربر با توجه به دیدگاه و نوع نگرش خود تعریف‌های مختلفی از داده‌کاوی ارائه کرده است. به‌عنوان مثال می‌توان به چند تعریف داده‌کاوی اشاره کرد که در ادامه آمده است:

- داده‌کاوی علم استخراج اطلاعات مفید از پایگاه‌های داده یا مجموعهٔ داده‌ای می‌باشد (Hand and Mannila, 2001).
- داده‌کاوی استخراج نیمه‌اتوماتیک[1] الگوها، تغییرات، وابستگی‌ها، نابهنجاری‌ها و دیگر ساختارهای معنی‌دار آماری از پایگاه‌های بزرگ داده می‌باشد (Grossman and Kamath, 2001).
- داده‌کاوی، مجموعه‌ای از روش‌ها در فرایند کشف دانش در بانک اطلاعاتی[2] است که برای تشخیص الگوها و رابطه‌های نامعلوم در داده‌ها مورد استفاده قرار می‌گیرد (Hand and Mannila, 2001).
- فرایند کشف الگوهای مفید از داده‌ها را داده‌کاوی می‌گویند (George, 1997).

۳-۱. بهره‌گیری از داده‌کاوی در تعرفه‌گذاری و ارزیابی ریسک

در روش‌های رایج تعرفه‌گذاری، بررسی تأثیر عوامل مؤثر بر ریسک به‌صورت محدود و منظم فراهم می‌شود، که جامعیت، صحت و دقت محاسبات را کاهش می‌دهند.

استفاده از داده‌کاوی به‌عنوان یکی از ابزارهای مؤثر در پژوهش‌های متعددی از جمله در چارچوب پیش‌بینی سطح خسارت مشتریان بیمهٔ بدنهٔ اتومبیل با استفاده از راهکار داده‌کاوی (ایزدپرست، ۱۳۸۹) با به‌کارگیری چندین الگوریتم از جمله خوشه‌بندی به ارزیابی میزان ریسک مشتریان بیمه پرداخته است همچنین در مدل‌سازی ریسک بیمه‌گذاران با رویکرد داده‌کاوی (Kolyshkina and Brookes, 2002) به پیش‌بینی میزان هزینه‌های بیمارستانی بیمه‌گذاران بیمه‌های درمان پرداخته شده و در ارزیابی ریسک اعتبار از تلفیق تکنیک‌های داده‌کاوی (Chen et al., 2012) به‌عنوان رهیافتی برای ارزیابی ریسک استفاده شده است.

---

[1] semiautomatic
[2] knowledge discovery in database (KDD)

٣-٢. فرایند داده‌کاوی

فرایند داده‌کاوی موجود در SQL Server شامل تعریف مسئله، آماده‌سازی داده‌ها، بررسی و اعتبارسنجی داده‌ها، تهیهٔ مدل‌ها، بررسی و ارزیابی مدل‌ها، اجرا و روزآمدسازی مدل‌ها می‌باشد (Microsoft MSDN, 2014).

این فرایند به‌صورت چرخه است، ولی هر مرحله لزوماً به‌صورت مستقیم به مرحلهٔ بعدی منجر نمی‌شود؛ بلکه باید در نظر داشت که تهیهٔ یک مدل داده‌کاوی، فرایندی پویا و تکرارپذیر است؛ چرا که پس از بررسی داده‌ها ممکن است دریابیم که داده‌های موجود برای تهیهٔ مدل مناسب داده‌کاوی، کافی نیستند و بنابراین لازم است در جستجوی داده‌های بیشتری باشیم. ممکن است مجبور باشیم مدل‌ها را پس از اجرای آن‌ها روزآمد نماییم تا داده‌های بیشتری به دست آوریم. بنابراین درک این نکته از اهمیت ویژه‌ای برخوردار است که یک مدل داده‌کاوی، فرایندی چرخشی است و ممکن است لازم باشد هر یک از مراحل فرایند آن‌قدر تکرار شود تا مدل مناسبی تهیه گردد (Microsoft MSDN, 2014).

٤. تحلیل

در این پژوهش، فاکتورهای موثر بر میزان ریسک بیمه‌گذاران در سه حوزه اطلاعات بیمه‌نامه، اطلاعات خودرو و عوامل انسانی مورد بررسی قرار گرفته است. هدف اساسی این پژوهش استفاده از داده‌کاوی برای تعیین حق بیمه به‌نحوی است که به تحلیل و ارزیابی صحیح از میزان ریسک مشتریان صنعت بیمه و دسته‌بندی عوامل مؤثر بر ریسک بیمه‌گذاران برای بیمه‌گر بپردازد و موجبات افزایش رضایت مشتریان را با تناسب میزان حق بیمه و میزان ریسک فراهم آورد.

این پژوهش در صدد پاسخگویی به سؤالات زیر است:

چگونه می‌توان از داده‌کاوی برای تعرفه‌گذاری پویا و متناسب با ریسک بیمه‌گذاران بهره برد؟

چگونه داده‌های انبار شده در بانک‌های اطلاعاتی شرکت‌های بیمه فرصت ارائهٔ راهکار کلان بیمه‌ای را فراهم می‌نمایند؟

آیا روش‌های دانش‌محور امکان تغییر در فرایندهای تصمیم‌گیری کلان در حوزهٔ بیمه را دارند؟

پس از مشخص شدن عوامل مؤثر، از توابع خوشه‌بندی برای تفکیک ریسک بیمه‌گذاران و پیش‌بینی میزان خسارت هر یک از خوشه‌ها استفاده می‌گردد.

فیلدهای اطلاعاتی پژوهش بر اساس یافته‌های پژوهش‌های دیگر از جمله استفاده از داده‌کاوی برای یافتن ویژگی‌های مشترک بیمه‌گذاران و یافتن میزان اثر عوامل ریسک مؤثر در بیمه‌نامهٔ خودرو با استفاده از درخت تصمیم و الگوریتم خوشه‌بندی (GUO, 2002) و داده‌های موجود در صنعت بیمه انتخاب شده‌اند.

الگوریتم‌های درخت تصمیم، میزان اثرگذاری را با جزئیات بیشتر به نمایش در خواهد آورد و امکان تصمیم‌گیری بهتر از میزان تأثیر عوامل ریسک را فراهم می‌نماید.

نخست باید فیلدهای اطلاعاتی و نوع آن‌ها برای انجام این پژوهش تعیین گردد. اطلاعاتی که بیمه‌گر باید در مورد یک قرارداد بیمه داشته باشد:

برخی از این اطلاعات با رعایت حسن نیت و با هدف افشای اطلاعاتی که بیمه‌گر نسبت به آن‌ها جاهل است و به آن نیاز دارد از طرف بیمه‌گذار ارائه می‌گردد. اما پاره‌ای از اطلاعات است که بیمه‌گر نسبت به آن عالم بوده و نیازی به افشای آن نیست. همچنین پاره‌ای از اطلاعات وجود دارد که بیمه‌گر به اقتضای حرفهٔ خود باید نسبت به آن‌ها عالم باشد (انصاری، ۱۳۹۲).

مجموعه‌داده‌هایی که در این پژوهش مورد کاوش قرار می‌گیرند در سه گروه عوامل خودرویی، انسانی و اطلاعات بیمه‌نامه‌ای قابل تفکیک می‌باشند.

معماری مورد نظر در جمع‌آوری و استانداردسازی داده‌ها و استفاده از بانک‌های اطلاعاتی نیروی انتظامی و سازمان ثبت احوال جهت اعتبارسنجی، طراحی و انتقال

داده‌های گردآوری‌شده در ساختار داده‌انباری، زیرساخت فرایند داده‌کاوی در صنعت بیمه را با دقت بسیار زیادی فراهم می‌نماید. به‌کارگیری داده‌کاوی در تعرفه‌گذاری، اساس ارزیابی مناسب از ریسک بیمه‌گذاران را فراهم می‌نماید و شرایط رشد شرکت‌های بیمه و رضایت مشتریان این صنعت را شکل می‌دهد. میزان توسعهٔ بیمه در کشورهای مختلف، ارتباط مستقیمی با میزان توسعه‌یافتگی و شکوفایی اقتصادی کشور را نشان می‌دهد به‌نحوی که با تحلیل دقیق اطلاعات موجود و انتقال صحیح ریسک صنایع و افراد به ساختار بیمه‌ای، ریسک در اقتصاد توزیع خواهد شد و رشدی متوازن در جامعه را پدید خواهد آورد.

٤-١. ویژگی‌های مرتبط با بیمه‌نامهٔ خودرو

با توجه به نظر کارشناسان صنعت بیمهٔ کشور، یکی از عوامل مؤثر بر میزان ریسک هر یک از بیمه‌گذاران، نوع و ویژگی‌های وسیلهٔ نقلیه می‌باشد. از این رو برای تعیین تعرفهٔ حق بیمه، باید اطلاعات خودرویی مورد کاوش قرار گیرد.

مجموع عواملی که در این بخش مورد استفاده قرار گرفته‌اند، با استفاده از بانک اطلاعاتی نیروی انتظامی اعتبارسنجی و پالایش شده‌اند.

جدول ۱: ویژگی‌های خودرو در مدل کاوشی

نام فیلد	نوع	واحد	پیوستگی
سال ساخت	عدد صحیح[1]	سال	گسسته
کاربری	متن[2]	نوع	گسسته
رنگ اصلی	متن	نوع	گسسته
نوع سوخت	متن	نوع	گسسته
برند	متن	نوع	گسسته
مدل برند	متن	نوع	گسسته
ظرفیت	متن	نفر و کیلوگرم	گسسته
محور	عدد صحیح	تعداد	گسسته
مسافت پیموده شده	ده‌دهی[3]	کیلومتر	پیوسته
چرخ	عدد صحیح	تعداد	گسسته
تعداد سیلندر	عدد صحیح	تعداد	گسسته

٤-۲. ویژگی‌های بیمه‌گذار

رفتار رانندگان متأثر از عوامل بسیاری است که در مطالعات قابل جمع‌آوری نیستند (Higgs, 2015) و به‌دلیل عدم ذخیرهٔ صحیح این اطلاعات در بانک‌های اطلاعاتی و عدم امکان صحت‌سنجی اطلاعات در زمان انجام این پژوهش با سایر بانک‌های اطلاعاتی، از آن‌ها صرف نظر شده است. فیلدهای اطلاعاتی مربوط به بیمه‌گذاران به شرح جدول شماره ۲ می‌باشد:

جدول۲: ویژگی‌های بیمه‌گذاران در مدل کاوشی

نام فیلد	نوع	واحد	پیوستگی
جنسیت	بیت[4]	-	گسسته
تحصیلات	متن	دوره	گسسته
تعداد خودرو	عدد صحیح	تعداد	گسسته
سن	عدد صحیح	سال	گسسته
تعداد اعضای خانوار	عدد صحیح	تعداد	گسسته

[1] int
[2] nvarchar
[3] decimal
[4] bit

میزان تحصیلات افراد بر اساس اطلاعات معتبر مرکز آمار ایران به‌صورت نسبی تکمیل گردیده است. تحصیلات از بی‌سواد با اعداد ۰ ، زیر دیپلم ۱، دیپلم ۲، فوق‌دیپلم ۳، لیسانس ٤، فوق‌لیسانس ۵ و دکتری با عدد ٦ مقداردهی گردید.

جدول ۳: موقعیت جغرافیایی

نام فیلد	نوع	واحد	پیوستگی
استان محل صدور بیمه‌نامه	متن	-	گسسته
شهر محل صدور	متن	-	گسسته

محل صدور بیمه‌نامه بر اساس موقعیت جغرافیایی نمایندگی و یا شعبه صادرکننده بیمه‌نامه تعیین می‌گردد.

٤-۳. سوابق بیمه‌نامه‌ای

از مهم‌ترین عوامل تعیین‌کننده ریسک بیمه‌گذاران، عوامل مالی و بیمه‌ای افراد می‌باشد. از جمله دلایل صحت بیشتر این پژوهش می‌توان به یکپارچه‌سازی اطلاعات تمامی شرکت‌های بیمه‌ای در داده‌انبارها جهت بررسی کامل و رصد رفتار بیمه‌گذاران در کل صنعت اشاره کرد، نکته‌ای که تا قبل از این پژوهش امکان‌پذیر نبوده است.

جدول ٤: سوابق بیمه‌ای در مدل کاوشی

نام فیلد	نوع	واحد	پیوستگی
تعداد سنوات عدم خسارت	عدد صحیح	تعداد	گسسته
جریمهٔ صندوق تأمین خسارت‌های بدنی	ده‌دهی	ریال	پیوسته
خسارت جانی	عدد صحیح	تعداد	گسسته
خسارت مالی	عدد صحیح	تعداد	گسسته
دفعات زیان‌دیدگی	عدد صحیح	تعداد	گسسته
میزان جرائم رانندگی	ده‌دهی	ریال	پیوسته
تخلفات حادثه‌ساز	عدد صحیح	تعداد	گسسته

٤-٤. فرایند بررسی و اعتبارسنجی داده‌ها

با توجه به این‌که داده‌های جمع‌آوری‌شده از بانک‌های اطلاعاتی شرکت‌های بیمه از منابع اطلاعاتی گوناگون با ساختار و فرمت‌های داده‌ای متفاوت جمع‌آوری شده‌اند، نیازمند همگون‌سازی نوع و محتوای داده‌ها می‌باشند، تا استخراج مفاهیم از داده‌های با تعاریف متفاوت امکان‌پذیر شود.

جمع‌آوری اطلاعات از شرکت‌های بیمه‌ای که هر یک از نرم‌افزار مستقل خود برای عملیات‌های بیمه‌گری استفاده می‌کنند، مستلزم فرایند استاندارد و پاک‌سازی داده‌ها جهت ایجاد داده‌انبار می‌باشد.

در مجامع علمی معتبر، معماری سرویس‌گرا جهت اعتبارسنجی داده‌های بین‌سازمانی در بستر فنّاوری‌های امن ارتباطات شبکه‌ای، به‌عنوان بهترین راهکار برای این فرایند توصیه گردیده است و لذا در این فرایند مورد استفاده قرار گرفت. بدین منظور، داده‌های استاندارد شده در مرحلهٔ استانداردسازی، امکان فراخوانی سرویس‌های سایر سازمان‌ها برای تکمیل و اصلاح اطلاعات جمع‌آوری‌شده را فراهم می‌نماید.

امکان اعتبارسنجی دادهای مورد نیاز بیمه‌گذاران در این پژوهش بر اساس کد ملی و تاریخ تولد افراد از سازمان ثبت احوال کشور و اعتبارسنجی اطلاعات خودرویی مورد بیمه نیز از طریق سرویس‌های نیروی انتظامی، به‌عنوان مرجع ثبت اطلاعات خودرویی کشور فراهم می‌باشد.

با انجام این مرحله، داده‌های مربوط به خودروهای پژوهش، در داده‌انبار برای ادامهٔ فرایند داده‌کاوی شامل ساخت مدل کاوش و آموزش مدل کاوش با تفکیک داده‌ها به مجموعه‌های آموزشی و آزمایشی و اعتبارسنجی، آماده می‌گردد.

نسبت داده‌های پژوهش در بخش آموزشی ٦٥ درصد ، داده‌های آزمایشی ۲۰ درصد و داده‌های اعتبارسنجی ۱۵ درصد از کل داده‌های پژوهش می‌باشد.

٤-٥. بررسی و ارزیابی صحت مدل‌های کاوش

پس از ساخت مدل و ذخیرهٔ الگوها، نتایج به‌دست‌آمده باید ارزیابی و تفسیر شوند. میزان دقت مدل‌ها به‌تنهایی برای انتخاب بهترین مدل کافی نیست و نیازمند آگاهی در خصوص انواع خطاها و هزینه‌های وابسته به آن‌ها نیز می‌باشد. در این مرحله مواردی از قبیل عملکرد درست مدل، توانایی توصیف داده‌ها، اعتماد به مدل‌های پیش‌بینی و جامعیت مدل مورد بررسی قرار می‌گیرد.

٤-٥-١. شاخص نسبت خسارت

نسبت خسارت در صنعت بیمه یکی از شاخص‌های کلیدی ارزیابی پرتفوی رشته‌های بیمه‌ای به حساب می‌آید. نسبت خسارت عبارت است از حاصل تقسیم خسارت بر حق بیمهٔ دریافتی که با استفاده از این شاخص می‌توان میزان خسارت پیش‌بینی‌شده را با وضعیت موجود در صنعت ارزیابی کرد.

٤-٥-٢. ماتریس آشفتگی[1]

این ماتریس، ابزار مناسبی برای درک میزان صحت نتایج مدل‌های دسته‌بندی و پیش‌بینی می‌باشد.

٤-٦. خوشه‌بندی ریسک مشتریان با استفاده از الگوریتم خوشه‌بندی

یک خوشه، به مجموعه‌ای از داده‌ها اطلاق می‌شود که از جهاتی شبیه به یکدیگرند. در این پژوهش با استفاده از مدل خوشه‌بندی، میزان ریسک مشتریان بیمه‌نامه‌های شخص ثالث، بر اساس عوامل مؤثر بر ریسک آن‌ها پیش‌بینی خواهد شد و با توجه به چگونگی توزیع خوشه‌ها، نحوهٔ تأثیر متقابل عوامل مؤثر بر ریسک به‌خوبی قابل مشاهده و چگونگی تأثیر این عوامل بر پیش‌بینی مدل مشخص خواهد شد.

اغلب اطلاعات و داده‌های موجود در پایگاه‌های داده‌ها توزیع‌های ناشناخته یا پیچیده‌ای دارند که به‌راحتی نمی‌توان آن توزیع‌ها را شناسایی نمود و مورد استفاده

---

[1] confusion matrix

قرار داد. بنابراین برای تحلیل داده‌ها و اطلاعات موجود در پایگاه‌های داده‌ها استفاده از روش‌هایی که نیاز به دانستن توزیع متغیرها ندارد از اهمیت خاصی برخوردار است. خوشه‌بندی یکی از روش‌هایی است که با توزیع داده‌های موجود سر و کار نداشته و اغلب با استفاده از معیارهای تشابه و عدم تشابه به خوشه‌بندی داده‌ها می‌پردازد (حائری مهریزی، ۱۳۸۲).

الگوریتم خوشه‌بندی تمام خوشه‌ها را بر اساس شباهت‌ها ایجاد می‌کند و در نهایت خوشه‌ها را بر اساس میزان شباهت به‌صورت پراکنده به نمایش می‌گذارد. فاصلهٔ بین خوشه‌ها اندازه‌گیری و به‌گونه‌ای چیده می‌شوند که خوشه‌هایی که به هم شبیه هستند نزدیک‌تر و فاصلهٔ زیاد به معنای حداقل شباهت بین داده‌های موجود در این خوشه‌ها می‌باشد. در واقع نخست خوشه‌ها بر اساس شباهت‌های بین مشخصه‌ها ایجاد می‌شوند و سپس چیدمان خوشه‌ها بر اساس فاصله آن‌ها شکل می‌گیرد.

جدول ۵: برخی از ویژگی‌های خوشه‌ها در تحلیل صورت‌گرفته از داده‌های بیمه‌ای شخص ثالث، به‌دست‌آمده از تحلیل

جمعیت خوشه (درصد)	شماره خوشه	نزدیک‌ترین خوشه	سن بیمه‌گذار	تعداد خسارت مالی	ظرفیت خودرو	مسافت پیموده‌شده (هزار کیلومتر)	نسبت جنسیت	میزان تحصیلات	میزان جرائم رانندگی (هزار ریال)	تعداد تخلفات حادثه‌ساز	تعداد سنوات عدم خسارت	تعداد خودرو
۲٤	۸	۷	۳٦,۲	۰,۲	٤,٦	۸٦	۰,۷۳	۳,۲	٤۱	۰,۰۲	٤,۲	۱,۰۳
۲۱	۹	۱	۲۷,۳	۰,۲	٤,۲	۷٦	۰,٦٥	۳	۳٥	۰,۰۱	۳,۹	۱,۰٤
۱۹,٦	٦	۸	۲۹,٦	۰,۳	٤,۹	۹۱	۰,۷۲	۲,۸	۲۷	۰,۰۳	۳,۷	۱,۰۲
۱۳	۷	۸	۳۸,۱	۰,۱	٤,۷	۷۲	۰,۸۳	۲,۳	۲٥	۰,۰۲	۳,۲	۱,۰٤
۸	٥	۱	۲٤,٥	۰,۳	٤,٦	٦۸	۰,۷۹	۲,۱	۲۲	۰,۳	۲,۹	۱,۰۱
٦,۲	۱	۹	٤۳,۲	۰,۲	۲۷,۱	۱۸٦	۰,۹۸	۲	۷۸	۰,۰٥	۳,۰	۱,۰۲
٤	٤	٥	۳۹,٤	۰,۱	۳۰,۲	۱٦۷	۰,۹۹	۱,۳	۸٤	۰,۰٤	۲,۸	۱,۰۳
۲,۱	۲	۸	٤٦,۲	۰,۰۹	۲,٦	۱۱۰	۰,۹۱	۱,۸	۳۲	۰,۰۳	۳,٦	۱,۰۱
۱,۳	۱۰	۲	۳۸,٦	۰,۳	۳,۱	۱۰٥	۰,۸۷	۱,٦	۳۷	۰,۰۲	۳,۳	۱,۰۱
۰,۸	۳	٥	۳٤,۲	۰,٤	۲,۳	۸۲	۰,۹٤	۱,۹	۲۸	۰,۰٤	۲,۹	۱,۰۱

٤-۷. درخت تصمیم

درخت تصمیم یکی از رایج‌ترین تکنیک‌های داده‌کاوی است. این الگوریتم در پژوهش‌های متنوع، از جمله در دسته‌بندی میزان خسارت بیمه‌های درمانی، به‌عنوان مؤثرترین تکنیک داده‌کاوی ارزیابی شده است (Araujo, 2015). هدف اصلی در درخت تصمیم، تقسیم داده‌ها به‌صورت بازگشتی به زیرمجموعه‌هایی است که هر زیرمجموعه دربرگیرندهٔ وضعیت همگنی از متغیر هدف می‌باشد. زمانی که پردازش بازگشتی کامل شد، درخت تصمیم شکل گرفته است.

٤-۸. شبکهٔ عصبی

شبکهٔ عصبی در مقایسه با درخت تصمیم و نیوبیز، از توانایی بیشتری در یادگیری الگوهایی پیچیده برخوردار است.

نمودار اول، شبکه‌ای شامل تعدادی گره (عصب) و خطوط ارتباطی بین گره‌ها است. انواع گره‌ها عبارت‌اند از ورودی، پنهانی و خروجی. هر یک از خطوط ارتباطی بین گره‌ها دارای وزن و جهت این خطوط نشان‌دهندهٔ گردش داده‌ها در فرایند پیش‌بینی می‌باشد. هر گره نمایانگر یک واحد پردازش می‌باشد.

گره‌های ورودی در اولین لایهٔ شبکه قرار دارد. در اغلب شبکه‌های عصبی، هر گره ورودی به یکی از صفت‌های ورودی به الگوریتم نگاشت می‌شود. مقدار هر صفت ورودی باید عددی در بازه‌ای یکسان باشد.

مجموعهٔ خروجی به‌دست‌آمده از مدل کاوشی، بر اساس میزان خسارت برآوردی به سه دستهٔ کم‌ریسک، ریسک متوسط و پرریسک طبقه‌بندی شده‌اند.

همان‌طور که در ماتریس آشفتگی قابل مشاهده می‌باشد از ۳۰ میلیون مورد پیش‌بینی، ۲۰,۳ میلیون مورد صحیح بوده است و این به این معنی است که صحت مدل تقریباً ٦۷,٦ درصد می‌باشد.

جدول ٦: ماتریس آشفتگی ارزیابی میزان صحت الگوریتم شبکهٔ عصبی به‌دست‌آمده از تحلیل (به میلیون)

پیش‌بینی	واقعی		
	کلاس کم ریسک	کلاس ریسک متوسط	کلاس ریسک بالا
کلاس کم ریسک	۱۰,۳	۲,٦	۰,۷
کلاس ریسک متوسط	۳,٤	۸,۱	۱,۳
کلاس ریسک بالا	۰,۸	۰,۹	۱,۹

٤-۹. روش پیشنهادی برای تعرفه‌گذاری پویا

روش پیشنهادی پژوهش برای محاسبه حق بیمهٔ دریافتی از بیمه‌گذاران به شرح زیر می‌باشد (محقق‌ساخته):

$$P=E[A|X]+S_c+S_d+Pf(P) \qquad (۳)$$

که در آن:

P: حق بیمهٔ دریافتی از بیمه‌گذار

E[A|X]: میزان خسارت پیش‌بینی‌شده هر بیمه‌نامه با استفاده از داده‌کاوی (متغیر A میزان خسارت بیمه‌گذار به‌ازای شرایط X را نشان می‌دهد.)

S_c: هزینه‌های ثابت

S_d: هزینه‌های متغیر

Pf(P): سود بیمه‌گری

در فرمول فوق با استفاده از پرس‌وجو از مدل کاوشی پردازش‌شده، بر اساس پارامترهای ورودی مدل کاوش شامل اطلاعات بیمه‌نامه، بیمه‌گذار و اتومبیل، میزان خسارت (A) قابل پیش‌بینی می‌باشد.

سایر اجزای فرمول محاسبه حق بیمه، شامل هزینه‌های ثابت و متغیر، بر اساس هزینه‌های شرکت‌های بیمه قابل محاسبه می‌باشد و میزان سود بیمه‌گری بر اساس سیاست‌های کلان اقتصادی تبیین می‌گردد.

به‌عنوان نمونه خسارت برآوردی برای بیمه‌نامهٔ خودرو پژو ۲۰٦ مدل ۱۳۸٦ با مسافت پیموده شدهٔ ۱٤۰ هزار کیلومتر برای آقایی ۵۷ ساله در شهرستان شیراز و بدون خسارت، در مدل کاوشی ۲۵۳۰۰۰۰ ریال می‌باشد که مبالغ هزینه‌های ثابت و متغیر و میزان سود بیمه‌گری به آن افزوده و در نتیجه حق بیمهٔ خالص محاسبه می‌گردد.

جهت بررسی میزان تأثیر استفاده از روش پیشنهادی در فعالیت شرکت‌های بیمه به مقایسهٔ نسبت خسارت حاصل از روش پیشنهادی، بر اساس داده‌های موجود در داده‌انبار پژوهش و نسبت خسارت سامانهٔ نظارت و هدایت الکترونیکی بیمهٔ مرکزی پرداخته شده است:

الف) نسبت خسارت به حق بیمهٔ کل برای رشتهٔ شخص ثالث در سامانهٔ نظارت و هدایت الکترونیکی بیمهٔ مرکزی سال ۱۳۹۲:

$$\left[\frac{٤٩٠٦٨٦٨٨.٨٩}{٦٨٩٤٣٦١٨.١٤}\right] * ١٠٠ = ٧٧.١٧$$

ب) نسبت خسارت به حق بیمهٔ کل حاصل از روش پیشنهادی برای رشتهٔ شخص ثالث کل کشور به شرح زیر است:

$$\left[\frac{٤٦٥٧٥٤٧٨.٤}{٧١٢٦٠٤٨١.٩٥}\right] * ١٠٠ = ٦٥.٣$$

مقایسهٔ اعداد به‌دست‌آمده در نسبت خسارت به حق بیمهٔ کل، کاهش حدود ۱۲ درصدی این شاخص بیمه‌ای را نشان می‌دهد که این امر حاصل از تناسب بیشتر حق بیمهٔ دریافتی از بیمه‌گذاران با میزان خسارات آن‌ها، در روش پیشنهادی است و این موضوع موجب افزایش سودآوری شرکت‌های بیمه در رشتهٔ شخص ثالث را فراهم می‌نماید.

۵. نتیجه‌گیری

در این پژوهش بیش از ۳۰ میلیون بیمه‌نامه از سال ۱۳۸۵ تا ۱۳۹۳ جمع‌آوری، استانداردسازی و از سایر بانک‌های اطلاعاتی معتبر در سطح کشور از جمله بانک اطلاعاتی نیروی انتظامی و سازمان ثبت احوال، جهت اعتبارسنجی داده‌ها استفاده گردید. اطلاعات مربوط به خسارات این بیمه‌نامه‌ها نیز به در ساختار بانک اطلاعاتی رابطه‌ای به بیمه‌نامه‌های خود متصل گردید. اطلاعات مربوط به منطقهٔ جغرافیایی بیمه‌نامه‌ها، بر اساس اطلاعات شبکهٔ فروش شرکت‌های بیمه تکمیل گردید.

برخی از اطلاعات به‌خاطرِ وجود نداشتن در سامانه‌های اطلاعاتی سطح کشور، مانند میزان مسافت پیموده‌شده هر خودرو، به‌صورت اعداد تصادفی با توزیع نرمال بر اساس آمارهای موجود تهیه گردید. پس از پالایش، داده‌ها به داده‌انبار طراحی‌شده که منبع اطلاعاتی مدل کاوشی است، انتقال و عوامل ریسک با استفاده از الگوریتم‌های خوشه‌بندی، درخت تصمیم و شبکهٔ عصبی آموزش داده شد.

برای محاسبه حق بیمه با استفاده از زبان توسعهٔ داده‌کاوی،[1] مقادیر ورودی به مدل داده شد و با توجه به شرایط و ویژگی‌های ورودی، به یکی از دسته‌های ریسک تعلق گرفت و مقدار ستون قابل پیش‌بینی مدل، که میزان خسارت انتخاب شده بود، حاصل شد.

پس از پیش‌بینی میزان خسارت، با تلفیق سایر عوامل مؤثر بر حق بیمه در فرمول ارائه‌شده، میزان حق بیمهٔ هر بیمه‌گذار متناسب با میزان ریسک محاسبه گردید که به‌دلیل تناسب میزان ریسک بیمه‌گذار با حق بیمه، موجب رضایت بیمه‌گذاران فراهم خواهد شد.

پس از محاسبهٔ حق بیمه با فرمول ارائه‌شده و مقایسهٔ میزان نسبت خسارت برآوردی، اعداد به‌دست‌آمده نشان‌دهندهٔ بهبود وضعیت سودآوری شرکت‌های بیمه

[1] data mining extensions (DMX)

می‌باشد. از این رو استفاده از روش پیشنهادی برای محاسبهٔ حق بیمه، به شرکت‌های بیمه توصیه می‌گردد.

برای تحقیقات آتی پیشنهاد می‌شود با استفاده از مدلی هیبریدی از تلفیق سایر الگوریتم‌های داده‌کاوی و منطق فازی، امکان ایجاد سامانه‌ای جهت برآورد ریسک بیمه‌گذاران بر بستر اطلاعات موجود در صنعت بیمه، و همچنین امکان استفاده از برآوردهای حاصل از مدل در کشف تقلبات احتمالی بیمه‌گذاران بررسی گردد.

منابع

۱. رنجبر فرد، مینا، ۱۳۹۳، "رضایت بیمه‌گذاران از خدمات بیمهٔ شخص ثالث"، پژوهشنامه بیمه، سال بیست و نهم، شماره ۱، صص ۸۱.

۲. دقیقی اصل، علیرضا، ۱۳۹۲، "اولویت‌بندی سیستم‌های مختلف نظارت بر توانگری شرکت‌های بیمه با استفاده از تکنیک فرایند تحلیل سلسله مراتبی (AHP)"، پژوهشنامه بیمه، سال بیست و هشتم، شماره ۱، صص ۱۰۹.

۳. انصاری، علی، ۱۳۹۲، "مسئولیت مدنی بیمه‌گذار نقض‌کننده حسن نیست در مرحله انعقاد قرارداد بیمه"،پژوهشنامه بیمه، سال بیست و هشتم، شماره ۱، صص ۱۱۱.

٤. خرمی، فرهاد، ۱۳۸۲، "معرفی روشی نظری – کاربردی برای اندازه‌گیری داده‌ها و ستانده‌ها و قیمت آن‌ها در شرکت‌های بیمه"، *فصلنامه صنعت بیمه، پژوهشکده بیمه، سال هجدهم، شماره ۱، صص ٦٥.*

۵. حائری مهریزی، علی‌اصغر، ۱۳۸۲، داده‌کاوی: مفاهیم و روش‌ها و کاربردها، پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی، تهران.

٦. ایزدپرست، محمود، ۱۳۸۹، ارائه چارچوبی برای پیش‌بینی سطح خسارت مشتریان بیمهٔ بدنهٔ اتومبیل با استفاده از راهکار داده‌کاوی، تهران: دانشگاه پیام نور، پایان‌نامه کارشناسی ارشد مدیریت فناوری اطلاعات.

7. Chapados, N., 2010, Data Mining Algorithms for Actuarial Ratemaking, ApSTAT Technologies Inc, , Montréal (Québec).
8. Chui-Yu Chiu, 2009, "An intelligent market segmentation system using k-means and particle swarm optimization", Expert Systems with Applications, Volume 36, Issue 3, Part 1 April 2009, PP. 4558–4565
9. Dake Zhanga, Kang Jianga, 2012, "Application of Data Mining Techniques in the Analysis of Fire Incidents", International Symposium on Safety Science and Engineering in China, Volume 43, 2012, Pages 250–256.
10. Dionne,G., 2005, Foundations of Insurance Economics, London: KAP. 1992 Edition.
11. Duarte de Araujo, F.H., 2015 "Evaluation of Classifiers Based on Decision Tree for Learning Medical Claim Process", Latin America Transactions, IEEE (Revista IEEE America Latina), Volume 13 , Issue 1 ,PP. 299 - 306
12. George, H. John, 1997, "Enhancements to the Data Mining Process", PhD Thesis, Computer Science Department, School of Engineering, Stanford University.
13. Goa lijia, 2003, "applying DM in property/casualty insurance", university of Central florida.
Han, J., 2006, Data Mining Concept and Techniques, A volume in The Morgan Kaufmann Series in Data Management Systems, San Francisco, 3rd Edition.
14. Hand, D., Mannila, M., Padhraic, S., 2001, Principle of Data Mining, MIT Press.
15. Higgs,B.,Dept, Jr., 2015, "Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns", Intelligent Transportation Systems, IEEE Transactions on (Volume:16 , Issue:1) ,PP. 81 – 90.
16. Inna Kolyshkina, Richard Brookes, 2002, "Data mining approaches to modelling insurance risk", PricewaterhouseCoopers.
17. Microsoft MSDN, June 2014, Data Mining Algorithms (Analysis Services - Data Mining),2014-08-5,
From: https://msdn.microsoft.com/en-us/library/ms174949.aspx.
18. R.L. Grossman , C. Kamath , P. Kegelmeyer , V. Kumar, 2001, "Data Mining for Scientific and Engineering Applications (Massive Computing)", Springer Science & Business Media, Oct 31, 2001.
19. Renuka Panchagavi, 2012, "Survey of Electronic Health Records Data for Developing a Predictive Model of Pressure Ulcers in Critical Care Patients", Master Thesis, Computer Science Department, School of Engineering,The Ohio State University.

20. Roung-Shiunn, Wua, 2011, "Customer segmentation of multiple category data in e-commerceusing a soft-clustering approach", Electronic Commerce Research and Applications, Volume 10, Issue 3, May–June 2011, Pages 331–341.
21. Tantan Liu, M.S, 2012, "Data Mining over Hidden Data Sources", PhD Thesis, Computer Science Department, School of Engineering, The Ohio State University.
22. V.L. Miguéis, 2012, "Customer data mining for lifestyle segmentation", Expert Systems with Applications, Volume 39, Issue 10, August 2012, Pages 9359–9366.
23. Weimin Chen, Guocheng Xiang, Youjin Liu, Kexi Wang, 2012, "Credit risk Evaluation by hybrid data mining technique", Hunan University of Science and Technology, Volume 3, 2012, Pages 194–200.
24. Werner, G., 2010, Basic Ratemaking, Casualty Actuarial Society, Fourth Edition, October 2010.
25. Fayyad, U.M, Piatetsky Shapiro, Smyth P and thurusamy.(eds.),1996, "Advances in Knowledge Discovery and Data Mining", Menlo Park, California, AAAI Press.