

ارائه چارچوبی برای اصلاح نرخ حق بیمه در رشته بدنۀ اتومبیل با استفاده از مدل شبکه‌های عصبی (مطالعه موردی: شرکت بیمه آسیا)

محمد صالح ترکستانی^۱، آرمان ده‌پناه^۲، محمد تقی تقوی‌فرد^۳، شهرام شفیعی^۴

چکیده: رشته بدنۀ اتومبیل برخلاف آنچه به‌نظر می‌رسد، چندان رشته سودآوری برای شرکت‌های بیمه محسوب نمی‌شود و به سمت زیان‌دهی در حرکت است. از این رو، پژوهش حاضر به کفایت نرخ‌های حق بیمه بدنۀ اتومبیل و تدابیری برای آنها توجه کرده و در پی بهسازی و علمی‌کردن هرچه بیشتر این امور است. به این ترتیب، ابتدا با شناسایی متغیرهای تأثیرگذار بر ریسک بیمه‌گذاران و مطابقت‌دادن متغیرها با داده‌های موجود در پایگاه داده شرکت مطالعه‌شده، عوامل نهایی انتخاب شدند؛ سپس ضمن عملیات پیش‌پردازش روی داده‌ها با استفاده از مدل شبکه‌های عصبی، طبقه خسارتی و میزان خسارت بالقوه بیمه‌گذاران پیش‌بینی شد تا شرکت‌های بیمه بتوانند با در نظر گرفتن آنها و ضریب خسارت مدنظرشان، نرخ‌های بهینه‌ای برای بیمه‌نامه‌ها تعریف کنند. نتایج پژوهش نشان می‌دهد مدل ارائه‌شده می‌تواند با دقت ۹۱ درصد طبقه خسارتی را تخمین بزند و با دقت ۸۷ درصد میزان خسارت بالقوه بیمه‌گذاران را پیش‌بینی کند.

واژه‌های کلیدی: بیمه بدنۀ اتومبیل، پیش‌بینی، داده‌کاوی، مدل شبکه‌های عصبی.

۱. استادیار گروه مدیریت بازرگانی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران
۲. دانشجوی دکتری مدیریت بازرگانی، دانشکده مدیریت، دانشگاه آزاد اسلامی، بابل، ایران
۳. دانشیار گروه مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران
۴. استادیار گروه مدیریت ورزشی، دانشکده تربیت بدنی و علوم ورزشی، دانشگاه گیلان، رشت، ایران

تاریخ دریافت مقاله: ۱۳۹۴/۰۶/۱۵

تاریخ پذیرش نهایی مقاله: ۱۳۹۵/۰۲/۰۶

نویسنده مسئول مقاله: آرمان ده‌پناه

E-mail: arman.dehpanah@gmail.com

مقدمه^۱

براساس نتایج پژوهشی، تصادفات اتومبیل نهمین عامل مرگومیر در جهان در سال ۲۰۰۴ بوده است و طبق پیش‌بینی‌ها، در سال ۲۰۳۰ به پنجمین عامل تبدیل خواهد شد (کیاورز مقدم و ونگ، ۲۰۱۴).

به‌موجب این تصادفات، سالانه بیش از یک میلیون نفر در سراسر جهان جان خود را از دست می‌دهند و بیش از پنجاه میلیون نفر مجروح می‌شوند. علاوه بر این، مرگومیر ناشی از تصادفات در ایران، حدود بیست برابر بیشتر از کشورهای پیشرفته گزارش شده است (منصوری و کارگر، ۲۰۱۴).

نرخ حق بیمه بدنه اتومبیل در بسیاری از شرکت‌های بیمه‌ای در کشورهای پیشرفته با توجه به متغیرهای گوناگون جمعیت‌شناختی، مشخصات اتومبیل و سابقه خسارت بیمه‌گذار محاسبه می‌شود؛ این در حالی است که در ایران تا اواخر سال ۱۳۸۸ و اجرای آزادسازی نرخ‌ها، نرخ حق بیمه بدنه اتومبیل با توجه به تعرفه بیمه مرکزی تعیین می‌شد. این امر سبب می‌شد که مشتریان کم‌ریسک‌تر، خسارت‌های مالی مشتریان پرریسک را جبران کنند، از این رو تفاوت چندانی بین مشتریان پرریسک و کم‌ریسک وجود نداشت. در واقع، در کشور ما به‌جای فرد، اتومبیل بیمه می‌شد و این امر موجب شده بود بیشتر شرکت‌های بیمه در زمینه بیمه اتومبیل، متحمل زیان شوند (حنفی‌زاده و رستخیز پایدار، ۱۳۹۰).

نداشتن معیار سنجشی برای تعیین ریسک افراد در بیمه اتومبیل، علاوه بر ناکارآمدی قرار دادهای بیمه، به تعیین نرخ‌های غیرعادلانه نیز منجر می‌شود، در صورتی که بیشتر کشورهای توسعه‌یافته با بهره‌مندی از سیستم طبقه‌بندی ریسک خسارت افراد، در افزایش بهره‌وری و سوددهی صنعت بیمه خود تلاش می‌کنند.

بدین ترتیب، پژوهش حاضر با هدف ارائه مدل و چارچوبی برای اصلاح نرخ حق بیمه بدنه اتومبیل به عرصه تحقیق گام برداشته است. بر این اساس، پژوهش پیش رو درصدد استفاده از مدل شبکه‌های عصبی برای پیش‌بینی طبقه خسارتی و میزان خسارت بالقوه بیمه‌گذاران بیمه بدنه اتومبیل در شرکت بیمه آسیا برآمده است تا بتوان با در نظر گرفتن ضریب خسارت مدنظر شرکت، میزان بهینه حق بیمه را مشخص کرد و سیاست‌های مقتضی را درباره گروه‌های مشخص بیمه‌گذاران به کار گرفت. در این رابطه سؤال‌های پژوهش به شرح زیر مطرح می‌شود:

۱. این پژوهش با همکاری و حمایت پژوهشکده بیمه وابسته به بیمه مرکزی جمهوری اسلامی ایران اجرا شده است.

۱. چگونه می‌توان با استفاده از مدل شبکه‌های عصبی، چارچوبی برای اصلاح نرخ حق بیمه بدنه اتومبیل ارائه کرد؟
۲. چه عواملی بر ایجاد خسارت در رشته بدنه اتومبیل تأثیر می‌گذارند؟
۳. چه سیاست‌هایی را می‌توان برای بهبود فرایند تعیین نرخ حق بیمه بدنه اتومبیل و شناسایی اثربخش مشتریان ارائه کرد؟

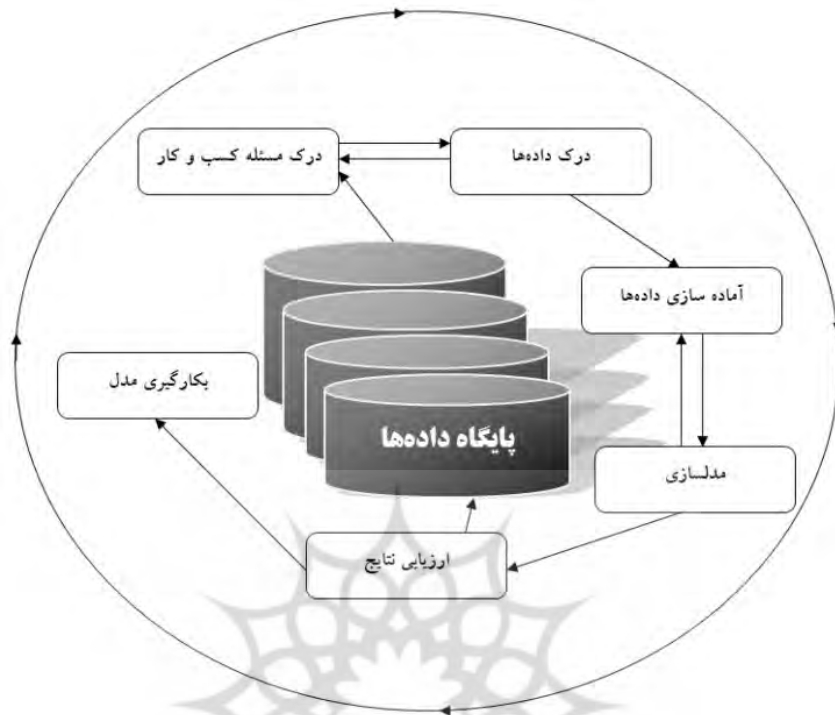
پیشینه نظری پژوهش

یکی از موارد مهمی که در سازمان‌های امروز به‌خوبی به‌چشم می‌خورد، فراوانی و حجم زیاد داده‌ها و اطلاعاتی است که در سازمان تولید می‌شود. این حجم وسیع داده و اطلاعات، در حالی که سازمان را با مسائلی در زمینه چگونگی نگهداری، جمع‌آوری و طبقه‌بندی داده‌ها روبه‌رو می‌کند، دانش نهفته‌ای دارند که هر سازمان یادگیرنده و هوشمند، از آن به‌مانند فرصتی برای ارتقای توانمندی‌هایش استفاده می‌کند (محمدی و علیزاده، ۱۳۹۳).

داده‌کاوی ابزار بسیار ارزشمندی است که در سال‌های اخیر به‌طور گسترده برای استخراج اطلاعات، جست‌وجوی روابط و الگوها بین حجم عظیمی از داده‌ها استفاده شده است (حنفی‌زاده و رستخیز پایدار، ۱۳۹۰). داده‌کاوی از ترکیب چندین رشته نشئت می‌گیرد. آمار، یادگیری ماشین، روش‌های بهینه‌سازی، روش‌های تشخیص و شناخت الگو، بانک اطلاعاتی، تجسم‌سازی، شبکه‌های عصبی، مدل‌های ریاضی، بازیابی اطلاعات، الگوریتم ژنتیک و هوش مصنوعی، فنونی هستند که داده‌کاوی از آنها بهره می‌برد (رادفر، نظافتی و یوسفی اصلی، ۱۳۹۳).

داده‌کاوی اغلب می‌تواند مدل‌های اکچوئری موجود در صنعت بیمه را از طریق پیداکردن متغیرهای مهم، تعیین روابط بین آنها و کشف روابط غیرخطی متغیرها ارتقا دهد و آنها را به نتایج قابل استفاده در کسب‌وکار، شامل توسعه محصول، بازاریابی و تحلیل توزیع خسارت‌ها تبدیل کند (قره‌خانی و ابوالقاسمی، ۱۳۹۰).

با توجه به اینکه پژوهش حاضر بر پایه کشف دانش از پایگاه داده شرکت بیمه مطالعه شده استوار است، به‌منظور اجرای فرایند پژوهش از استاندارد جهانی فرایند داده‌کاوی در صنعت^۱ بهره برده شده است. شکل ۱ مراحل این استاندارد را در قالب نمودار نشان می‌دهد.



شکل ۱. مراحل داده‌کاوی براساس استاندارد CRISP-DM

پیشینه تجربی پژوهش

یئو، اسمیت، ویلیس و بروکز (۲۰۰۱) با استفاده از تکنیک خوشه‌بندی و در نظر گرفتن معیارهای جمعیت‌شناختی نظیر سن و جنسیت راننده، به خوشه‌بندی مشتریان و ریسک مربوط به آنها پرداختند؛ سپس در هر خوشه با استفاده از مدل شبکه‌های عصبی و با در نظر گرفتن معیارهای یادشده، تغییرات حق بیمه را پیش‌بینی کردند. نتایج نشان داد استفاده از تکنیک‌های خوشه‌بندی به بهبود عملکرد پیش‌بینی منجر می‌شود. استفاده ترکیبی از شبکه‌های عصبی و الگوریتم خوشه‌بندی کا - مینز^۱ توانست دقت ۷۸ درصد در پیش‌بینی را به دست آورد.

گو (۲۰۰۳) با استفاده از روش خوشه‌بندی به وسیله الگوریتم کا-مینز، به بررسی ویژگی‌های بیمه‌گذاران در رشته اتومبیل با تأکید بر مشخصه‌های بیمه‌گذار پرداخت و مشتریان را با توجه به

عواملی چون سن، جنسیت، سطح تحصیلات راننده، نوع اتومبیل، محل اقامت و... خوشه‌بندی کرد؛ سپس، با استفاده از درخت تصمیم، اهمیت هر یک از این عوامل را در ایجاد خسارت نشان داد. نیوستد و دی الیا (۲۰۰۷) در پژوهشی با استفاده از روش رگرسیون و در نظر گرفتن توزیع پواسون، رابطه بین رنگ اتومبیل و احتمال خسارت را بررسی کردند. بدین منظور، دو ایالت استرالیا به عنوان نمونه برای جمع‌آوری داده‌ها انتخاب شدند. به طور کلی، آنها نتیجه گرفتند که هنگام روز رنگ‌های مشکی، نوک‌مدادی و نقره‌ای، هنگام غروب رنگ مشکی و هنگام شب رنگ قرمز پرخطرترین رنگ‌ها محسوب می‌شوند. آنها نشان دادند بین رنگ اتومبیل و احتمال خسارت رابطه انکارنشده وجود دارد.

پارنیتزکه (۲۰۰۸) سیستم نرخ‌گذاری بیمه در آمریکا را بررسی کرد. وی در این پژوهش سیستمی را با عنوان امتیازدهی بیمه^۱ که در بیشتر ایالت‌های آمریکا استفاده می‌شود، تفسیر کرد. در این سیستم براساس سوابق اعتباری بیمه‌گذاران به آنها امتیازهایی داده می‌شود که در تعیین نرخ حق بیمه مؤثر است. طبق این سیستم، بیمه‌گذاری که سابقه اعتباری خوبی ندارد، به طور مثال قسط وام‌های خود را به موقع پرداخت نمی‌کند، فردی بی‌احتیاط در نظر گرفته می‌شود که احتمال رانندگی نامطمئن و وقوع خسارت برای او بیشتر است. وی این سیستم را با سیستم‌های رایج محاسبه نرخ در آلمان مقایسه کرد و رویکرد ترکیبی از هر دو روش را برای پیاده‌سازی در نظام بیمه اتومبیل آلمان پیشنهاد داد.

از تحقیقات انجام‌شده در کشور نیز می‌توان به پژوهش چوبدار (۱۳۸۷) اشاره کرد که با در نظر گرفتن متغیرهایی نظیر سن و جنسیت راننده، نوع و کاربری اتومبیل و... و با استفاده از تکنیک درخت تصمیم، بیمه‌گذاران بدنه اتومبیل و پیش‌بینی طبقه خسارتی آنها را شناسایی کرد. دقت پیش‌بینی مدل به کاررفته در این پژوهش به میزان ۷۵ درصد به دست آمد و قابل قبول بود. عنبری، نادعلی و اسلامی نصرت‌آبادی (۱۳۸۹) در پژوهشی با در نظر گرفتن مشخصه‌های بیمه‌گذار مانند سن و جنسیت او، در کنار مشخصه‌های اتومبیل مانند کاربری و سن آن، چند مدل (درخت تصمیم، شبکه‌های عصبی، شبکه‌های بیزین، ماشین بردار پشتیبان، رگرسیون لجستیک، تحلیل تمایزی) را به منظور پیش‌بینی طبقه خسارتی بیمه‌گذاران مقایسه کردند و به دسته‌بندی آنها در سه طبقه کم‌ریسک، ریسک متوسط و پریسک پرداختند. طبق نتایج این پژوهش، مدل درخت تصمیم بیشترین دقت را در پیش‌بینی داشت و با دقت ۶۸ درصد توانست طبقه خسارتی بیمه‌گذاران را پیش‌بینی کند.

در پژوهشی مشابه حاجی حیدری، خاله و فراهی (۱۳۹۰) با در نظر گرفتن همزمان مشخصه‌های بیمه‌گذار و اتومبیل، چند مدل (درخت تصمیم، شبکه‌های عصبی، شبکه‌های بیزین، ماشین بردار پشتیبان، رگرسیون لجستیک، تحلیل تمایزی) را به منظور پیش‌بینی طبقه خسارتی بیمه‌گذاران مقایسه کردند. طبق نتایج این پژوهش، مدل‌های شبکه‌های عصبی و درخت تصمیم با حدود ۸۲ درصد، بیشترین دقت را در پیش‌بینی داشتند.

حنفی‌زاده و رستخیز پایدار (۱۳۹۰) ابتدا عوامل مؤثر بر ایجاد خسارت در بدنه اتومبیل را در ایران بررسی کردند. پس از مشخص شدن عوامل با استفاده از شبکه‌های عصبی خودسازمانده، به خوشه‌بندی بیمه‌گذاران بر اساس ریسک بالقوه آنها پرداختند.

در پژوهشی دیگر، فتح‌نژاد و ایزدپرست (۱۳۹۰) با استفاده از تکنیک خوشه‌بندی ک-مینز و درخت تصمیم و با در نظر گرفتن متغیرهای جمعیت‌شناختی نظیر سن، جنسیت، شغل، وضعیت تأهل راننده و متغیرهای مربوط به اتومبیل نظیر نوع، سال ساخت، کاربری و ظرفیت اتومبیل و... بیمه‌گذاران را خوشه‌بندی کردند و نتیجه گرفتند علاوه بر مشخصات اتومبیل، مشخصات رفتاری مشتری نیز در پیش‌بینی سطح خسارت مشتریان بیمه بدنه اتومبیل تأثیرگذار است. دقت مدل‌های استفاده‌شده در این پژوهش حدود ۶۰ درصد بوده است.

روش‌شناسی پژوهش

پژوهش حاضر به لحاظ هدف کاربردی است؛ زیرا مدل ارائه‌شده در این پژوهش راهکارهایی برای بهبود نرخ‌گذاری حق بیمه بدنه اتومبیل و به‌طور کلی بهبود عملکرد شرکت‌های بیمه، پیش‌رو می‌گذارد. از آنجا که به منظور شناخت رفتار بین متغیرها به توصیف و تحلیل روابط بین آنها پرداخته شده است، از دید ماهیت و روش توصیفی-همبستگی است. روش و ابزار گردآوری اطلاعات در این پژوهش به‌واسطه بررسی پیشینه پژوهش و استفاده از داده‌های ثانویه و مطالعه مقاله‌ها، کتاب‌ها و پایان‌نامه‌های خارجی و داخلی مرتبط، کتابخانه‌ای است.

ابتدا عوامل اثرگذار بر ریسک مشتریان بیمه بدنه اتومبیل از طریق بررسی پیشینه پژوهش شناسایی شدند و در سه دسته کلی مشخصات بیمه‌گذار، اتومبیل و بیمه‌نامه قرار گرفتند؛ سپس، از طریق مشورت با چند نفر از خبرگان صنعت بیمه و مطابقت‌دادن این عوامل با داده‌های موجود در پایگاه داده‌های شرکت مد نظر، عوامل نهایی ریسک انتخاب شدند. در گام بعد با استفاده از الگوریتم شبکه‌های عصبی مدلی ارائه شد که به‌وسیله آن می‌توان طبقه خسارتی بیمه‌گذاران و میزان خسارت بالقوه آنان را پیش‌بینی کرد. چنین مدلی می‌تواند در سیاست‌گذاری‌های آتی شرکت‌های بیمه به کار گرفته شود. شرکت‌های بیمه می‌توانند با استفاده از نتایج این داده‌کاوی،

میزان حق بیمهٔ دریافتی از بیمه‌گذاران مختلف را تعدیل کنند و با ایجاد سیستم نرخ‌گذاری مبتنی بر ریسک بیمه‌گذاران، میزان رضایت آنها را افزایش داده و سودآوری خود را ارتقا دهند.

شبکه‌های عصبی مصنوعی

بر اساس تحقیق ان‌گایی، ژیو و چاو (۲۰۰۹) مشخص شد که از بین ۳۴ تکنیک داده‌کاوی، شبکه‌های عصبی بیشترین و مؤثرترین کاربرد را داشته است. برتری شبکه‌های عصبی نسبت به روش‌های دیگر این است که می‌تواند از وقایع گذشته بیاموزد و با گذشت زمان نتایج را بهبود دهد. همچنین می‌تواند ضمن استخراج قوانین، رفتار آینده را براساس وضعیت فعلی پیش‌بینی کند (وثوق، تقوی‌فرد و البرزی، ۱۳۹۳). این شبکه‌ها با استفاده از مجموعهٔ ورودی و خروجی، روابط بین آنها را پیش‌بینی می‌کنند و به‌اصطلاح آموزش می‌بینند؛ به‌گونه‌ای که پس از آموزش، به‌ازای یک عضو جدید از مجموعهٔ ورودی، خروجی متناظر آن را تقریب می‌زنند.

شایان ذکر است که مدل شبکه‌های عصبی معماری‌های مختلفی دارد که معروف‌ترین آنها عبارت‌اند از پرسپترون^۱، پرسپترون چندلایه^۲، تابع پایهٔ شعاعی^۳ و ماشین‌های بردار پشتیبان^۴. شبکه‌های پرسپترون چندلایه با الگوریتم پس‌انتشار خطا تا چندین سال پرفرودارترین معماری شبکه‌های عصبی بود (مؤمنی، ۱۳۸۵)، اما طی سال‌های اخیر با حجیم‌تر شدن پایگاه‌های داده و پیچیده‌تر شدن ساختار آنها، معماری‌های جدیدی از شبکه‌های عصبی نظیر تابع پایهٔ شعاعی و ماشین‌های بردار پشتیبان ارائه شدند که در برخی موارد نیز توانستند عملکرد بهتری نسبت به پرسپترون چندلایه داشته باشند. هریک از معماری‌های شبکه‌های عصبی عیب‌ها و مزیت‌هایی دارد و نمی‌توان در مجموع یک معماری را بهتر از بقیه دانست؛ به‌طور مثال، ماشین‌های بردار پشتیبان، اغلب قدرت پیش‌بینی بهتری نسبت به سایر معماری‌ها نشان می‌دهند، اما آموزش آنها کمابیش زمان‌بر است و انتخاب معیارهایی نظیر کرنل^۵ یا حاشیۀ اطمینان برای ساختارشان، استفاده از آنها را نسبتاً مشکل می‌کند. به‌طور کلی، می‌توان گفت که برای داده‌هایی که ساختار بسیار پیچیده، ابعاد گسترده و اختلال فراوان دارند، بهتر است از معماری‌های جدید یا حتی ترکیبی استفاده کرد و برای داده‌هایی با ساختار نه‌چندان پیچیده و اختلال کمتر، به‌منظور راحتی کار و استفادهٔ بهینه از زمان، از معماری‌های سنتی نظیر پرسپترون چندلایه بهره برد (گانتی، پول

-
1. Perceptron
 2. Multi Layer Perceptron (MLP)
 3. Radial Basis Function (RBF)
 4. Support Vector Machines (SVMs)
 5. Kernel

و پال، ۲۰۰۹). در این تحقیق، به دلیل ساختار نه‌چندان پیچیده و اختلالات نه‌چندان زیاد داده‌ها، از شبکه‌های پرسپترون چندلایه با آموزش شبکه پس‌انتشار خطا از نوع تابع سیگموئید استفاده شده است.

تعریف متغیرها

در مرحله نخست همه متغیرهای اثرگذار بر وقوع خسارت از طریق مطالعه پیشینه تحقیق، شناسایی شدند؛ سپس، طی مشورت با چند نفر از خبرگان صنعت بیمه و نظرخواهی از آنان، متغیرهایی نظیر نوع گواهی‌نامه، سرعت رانندگی، ظرفیت موتور، نوع پرداخت حق بیمه و... حذف شدند. در مرحله بعد متغیرهای باقی‌مانده در سه دسته قرار گرفتند. مشخصه‌های نهایی در جدول ۱ آورده شده است.

جدول ۱. عوامل نهایی انتخاب‌شده پس از مشورت با خبرگان

مشخصات بیمه‌گذار	سن، جنسیت، وضعیت تأهل، شهر محل زندگی، سال اخذ گواهی‌نامه (سابقه رانندگی راننده)، میزان تحصیلات، میزان درآمد، شغل، تعداد ادعای خسارت در سال قبل، فاصله محل کار تا محل زندگی
مشخصات اتومبیل	نوع، رنگ، سال ساخت، تجهیزات ایمنی (ABS)، میزان کارکرد به کیلومتر، کاربری، گروه، نوع پلاک، تعداد سیلندر، نوع تیپ، ارزش روز
مشخصات بیمه‌نامه	میزان پوشش بیمه‌ای

جمع‌آوری داده‌ها

در این پژوهش از داده‌های موجود در پایگاه داده بیمه بدنه اتومبیل یکی از شرکت‌های بیمه داخلی (بیمه آسیا) استفاده شد. این داده‌ها شامل داده‌های مربوط به بیمه‌نامه بیمه‌گذاران بدنه اتومبیل است که تعدادی از آنها دچار حادثه شده‌اند و خسارت دیده‌اند. از آنجا که شرکت بیمه آسیا از سال ۱۳۸۹ سیستم جدیدی را برای پایگاه داده‌هایش به کار برده است، جامعه آماری را ۷۳۳۸۷۵ داده مربوط به بیمه‌نامه‌های بدنه اتومبیل که طی سال‌های ۱۳۸۹ تا سه‌ماهه اول ۱۳۹۱، شرکت بیمه آسیا برای آنها بیمه‌نامه صادر کرده است و بخشی از آنها دچار خسارت شده‌اند، در نظر می‌گیریم.

پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها از گام‌های مهم فرایند داده‌کاوی است که میزان دقت نتایج به‌دست‌آمده تا حد زیادی به اجرای درست آن بستگی دارد؛ بدین‌منظور دو اقدام مهم کاهش داده و اعمال تغییرات در شکل داده‌ها روی آنها انجام گرفت. ابتدا ۹ متغیر از متغیرهای نهایی، شامل سال اخذ گواهی‌نامه، شغل، میزان تحصیلات، میزان درآمد و فاصله محل کار تا محل زندگی از مشخصات بیمه‌گذار؛ تجهیزات ایمنی، میزان کارکرد به کیلومتر و نوع تیپ از مشخصات اتومبیل و میزان پوشش بیمه‌ای از مشخصات بیمه‌نامه، به‌دلیل موجود نبودن اطلاعات در پایگاه داده، از بررسی خارج شدند. همچنین بعضی از متغیرهای موجود در پایگاه داده مانند کد ملی، کد پستی و نام بیمه‌گذاران، شماره بیمه‌نامه و... به‌دلیل یکتا بودن برای هر یک از بیمه‌گذاران و بی‌ارتباط بودن با هدف پژوهش حذف شدند. با توجه به اینکه فراوانی هفت نوع خودروی پراید، پژو ۴۰۵، پژو ۲۰۶، سمند، کامیون، وانت و پژو پارس با سایر انواع خودروهای بیمه‌شده در این شرکت تفاوت چشمگیری داشتند، تنها به تجزیه و تحلیل این هفت نوع خودرو پرداخته شد.

در خصوص نوع پلاک، کاربری و گروه اتومبیل نیز مواردی که فراوانی بیشتری داشتند حفظ‌شده و باقی حذف شدند. همچنین اتومبیل‌های تولیدشده در ۱۶ سال اخیر، به‌دلیل فراوانی بیشتر بررسی شدند. از آنجا که تنوع رنگ در اتومبیل‌ها زیاد بود، اتومبیل‌ها از روی رنگ در چند طبقه کلی‌تر دسته‌بندی شدند. به‌طور مثال، رنگ‌های آلبالویی، جگری، عنابی و... در گروه رنگ قرمز قرار داده شدند. داده‌های باقی‌مانده کیفیت لازم و صلاحیت ورود به مدل نهایی را نداشتند؛ به همین دلیل معیوب بودن داده‌ها از دو جنبه مطالعه شد؛ ابتدا خطاهای اپراتور در وارد کردن داده‌ها بررسی شد، بدین‌معنا که بعضی متغیرها مقادیر نامناسب و نامربوط داشتند. به‌طور مثال، در بعضی رکوردها سال تولید خودرو اعداد پنج‌رقمی یا سه‌رقمی وارد شده بود که امکان اصلاح و حدس زدن مقدار صحیح وجود نداشت و رکورد آن حذف شد، یا اینکه تعداد سیلندر برای خودروی سمند عدد ۸ درج شده بود که با مقدار ۴ جایگزین شد. همچنین در بعضی رکوردها با توجه به نام بیمه‌گذار، جنسیت بعضی از بیمه‌گذاران اشتباه وارد شده بود که تا حد امکان تصحیح شدند و اقداماتی از این قبیل در مرحله اول روی داده‌ها انجام گرفت. در جنبه دوم، رکوردها از لحاظ وجود داده‌های نامرتب بررسی شدند؛ بدین‌صورت که برای بعضی از متغیرها مقادیر نامرتب وارد شده بود. به‌طور مثال، برای نوع خودرو یا نوع استفاده آن مقادیر عددی اختصاص یافته بود. بر این اساس، تا حد امکان سعی شد رکوردهای معیوب طی مشورت با متصدیان بیمه و با توجه به متغیرهای مرتبط دیگر اصلاح شود، به‌طور کلی رکوردهایی که امکان اصلاح داده‌های معیوبشان وجود نداشت، حذف شدند.

یادآوری می‌شود که با استفاده از فیلد شهر محل سکونت بیمه‌گذاران، جمعیت هر یک از این شهرها از نتایج رسمی سرشماری عمومی نفوس و مسکن سال ۱۳۹۰ استخراج شد. سپس، شهرها براساس جمعیت به سه گروه کم جمعیت، متوسط و پرجمعیت دسته‌بندی شدند. شهرهایی که کمتر از صد هزار نفر جمعیت داشتند، شهرهای کم جمعیت؛ بین صد هزار تا یک میلیون نفر شهرهای متوسط و شهرهایی که بیشتر از یک میلیون نفر جمعیت داشتند، شهرهای پرجمعیت در نظر گرفته شدند.

انتخاب مشخصه^۱

در این قسمت با استفاده از امکان انتخاب مشخصه در کلمنتاین، صلاحیت متغیرها برای ورود به مدل و در صورت امکان کاهش ابعاد آنها بررسی شد. شکل ۲ نشان‌دهنده نتایج انتخاب مشخصه‌ها برای متغیر هدف مرحله نخست، یعنی طبقه‌بندی خسارتی بیمه‌گذاران است که طبق آن همه متغیرهای انتخاب‌شده صلاحیت لازم برای پیش‌بینی طبقه خسارتی را دارند، وابستگی خاصی به یکدیگر ندارند و ورودشان به مدل بلامانع است.

Field	Type	Importance	Value
Vehicle Group	Set	Important	1.0
Gender	Flag	Important	1.0
Vehicle Cylinder Count	Set	Important	1.0
Vehicle Usage	Set	Important	1.0
Vehicle Licence Plate Type	Set	Important	1.0
Vehicle System	Set	Important	1.0
Vehicle Color	Set	Important	1.0
Insured Value	Range	Important	1.0
Population Group	Set	Important	1.0
Vehicle Production Year	Set	Important	1.0
Policy Holder Record	Set	Important	1.0
Marital Status	Flag	Important	0.998
Age	Range	Important	0.998

شکل ۲. نتیجه انتخاب مشخصه برای متغیرها در پیش‌بینی طبقه خسارتی

همچنین شکل ۳ نشان‌دهنده نتایج انتخاب مشخصه‌ها برای مرحله دوم و پیش‌بینی میزان خسارت بیمه‌گذاران به‌عنوان متغیر هدف است. در این قسمت نیز همه متغیرهای انتخاب‌شده صلاحیت لازم برای پیش‌بینی میزان خسارت را داشتند، وابستگی خاصی به یکدیگر نداشتند و ورودشان به مدل مانعی نداشت.

خلاصه‌ای از عملیات پیش‌پردازش انجام‌شده در جدول ۲ آورده شده است.

1. Feature Selection

Field	Type	Importance	Value
Gender	Flag	Important	1.0
Marital Status	Flag	Important	1.0
Population Group	Set	Important	1.0
Vehicle Usage	Set	Important	1.0
Vehicle Licence Plate Type	Set	Important	1.0
Policy Holder Record	Set	Important	1.0
Vehicle System	Set	Important	1.0
Vehicle Group	Set	Important	1.0
Vehicle Color	Set	Important	1.0
Vehicle Cylinder Count	Set	Important	1.0
Insured Value	Range	Important	1.0
Age	Range	Important	1.0
Vehicle Production Year	Set	Important	1.0

> 0.95 <= 0.95 < 0.9

شکل ۳. نتیجه انتخاب مشخصه برای متغیرها در پیش‌بینی میزان خسارت

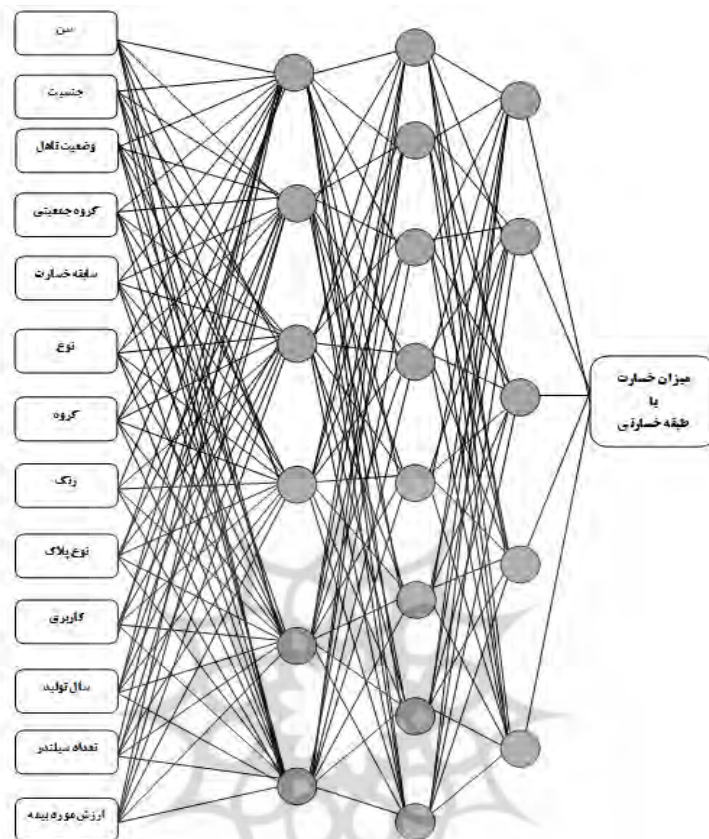
جدول ۲. خلاصه‌ای از عملیات پیش‌پردازش روی داده‌ها

نوع داده	عملیات پیش‌پردازش انجام شده	مثال
داده‌های نامرتب، بی‌معنا و خاص	حذف داده‌ها	حذف متغیرهایی مانند شماره ملی، شماره پلاک و کد پستی
داده‌های اضافی یا ناموجود در پایگاه داده	حذف داده‌ها	حذف متغیرهایی مانند میزان درآمد، میزان تحصیلات و شغل
فراوانی داده‌ها	کاهش داده‌ها	انتخاب هفت نوع خودرو با فراوانی بیشتر میان انواع خودرو
گسترده‌ی مقادیر داده‌ها	کاهش ابعاد داده‌ها	گروه‌بندی رنگ‌های مشابه مانند آلبالویی، عنابی، جگری در گروه کلی رنگ قرمز، و گروه‌بندی جمعیتی شهرهای بیمه‌گذاران
داده‌های دارای اختلال، برون‌هسته یا خطای اپراتور	در صورت امکان تصحیح از طریق مشورت با متصدیان یا دانسته‌های پیشین محقق، در غیر این صورت حذف	جای‌گذاری عدد ۴ به‌جای عدد ۸ برای تعداد سیلندر خودروی سمند
داده‌های مفقود	در صورت امکان تصحیح از طریق مشورت با متصدیان یا مقدار داده‌های مشابه، در غیر این صورت حذف	جای‌گذاری تعداد سیلندر با توجه به نوع خودروها، یا حذف داده در صورت خالی‌بودن فیلد رنگ

در نهایت، پس از پالایش، پیش‌پردازش و انتخاب مشخصه‌ها، ۲۴۳۷۰ داده شامل ۱۳ متغیر مستقل و وابسته طبقه خسارتی در مرحله اول و میزان خسارت در مرحله دوم به‌عنوان ورودی‌ها و خروجی‌های مدل انتخاب شدند. در جدول ۳ تغییرات نهایی اعمال شده، نوع متغیرهای مستقل به‌کاررفته در مدل و مقادیر هریک آورده شده است. همچنین در شکل ۴ مدل شبکه‌های عصبی به‌کاررفته در این پژوهش نشان داده شده است.

جدول ۳. متغیرهای به‌کاررفته در مدل به‌همراه مقادیر هریک

متغیر	نام متغیر در مدل	مقادیر متغیر	
مشخصات بیمه‌گذار	سن	Age	
	جنسیت	Gender	
	وضعیت تأهل	Marital Status	
	گروه جمعیتی	Population Group	
	سابقه خسارت	Policy Holder Record	
مشخصات اتومبیل	نوع	Vehicle System	
	گروه	Vehicle Group	
	رنگ	Vehicle Color	
	نوع پلاک	Vehicle License Plate Type	
	کاربری	Vehicle Usage	
	سال تولید	Vehicle Production Year	
	تعداد سیلندر	Vehicle Cylinder Count	
	ارزش بیمه‌شده	Insured Value	
			از ۱۹ تا ۹۱ سال
			مرد، زن
		مجرد، متأهل	
		کم‌جمعیت، متوسط، پرجمعیت	
		نداشتن خسارت در سال‌های اول، دوم، سوم، چهارم، پنجم و بیشتر، خسارت دارد، تخفیف ندارد، جدید	
		پراید، پژو ۴۰۵، پژو ۲۰۶، سمند، پژو پارس، کامیون، وانت	
		سواری، بارکش	
		آبی، زرد، سبز، خاکستری، مشکی، سفید، قرمز، تیره‌ای	
		شخصی، عمومی	
		شخصی، تاکسی، بارکش	
		از ۱۳۷۶ تا ۱۳۹۱	
		۴، ۶، ۸	
		از ۲۷۰۰۰۰۰۰۰ تا ۵۴۷۰۰۰۰۰ ریال	



شکل ۴. مدل شبکه‌های عصبی به کاررفته در پژوهش

ابزار اندازه‌گیری

در این پژوهش برای داده‌کاوی از نرم‌افزار کلمنتاین استفاده شده است. کلمنتاین نرم‌افزاری است که برای داده‌کاوی به کار می‌رود و شرکت آی. اس. ال.^۱ آن را ارائه کرده است. کلمنتاین به کاربران اجازه می‌دهد بدون برنامه‌نویسی با الگوریتم‌های آماری و داده‌کاوی کار کنند. دستورهای لازم برای پردازش داده‌ها به صورت آیکون انتخاب می‌شوند و برای پردازش توسط سیستم به صورت خودکار به کدهای برنامه‌نویسی تبدیل خواهند شد (نیسبت، الدر و ماینر، ۲۰۰۹). شایان ذکر است که نسخه‌های جدیدتر این نرم‌افزار پس از خریداری آن توسط شرکت آی. بی. ام، با نام IBM SPSS Modeler و در دو نسخهٔ حرفه‌ای برای تحلیل پایگاه‌های داده و ممتاز عرضه

شده است که علاوه بر ویژگی‌های نوع حرفه‌ای، از ویژگی‌های متن کاوی و تحلیل شبکه‌های اجتماعی نیز برخوردار است (رات و نات، ۲۰۱۵).

یافته‌های پژوهش

در فرایند پیش‌بینی شبکه‌های عصبی، معمولاً مجموعه‌ای از داده‌ها به‌عنوان ورودی در اختیار شبکه عصبی قرار می‌گیرند تا شبکه با تخمین رفتار سیستم پیش‌بینی شده، برای آینده برون‌یابی کند. در این پژوهش ۸۰ درصد داده‌ها به‌عنوان داده‌های آموزش و ۲۰ درصد باقی داده‌های آزمایش در نظر گرفته شدند. در جدول ۴ اطلاعاتی درباره همبستگی بین طبقه خسارتی واقعی و پیش‌بینی شده بیمه‌گذاران یا به‌بیانی، دقت و قدرت پیش‌بینی مدل آورده شده است.

جدول ۴. اطلاعات همبستگی بین طبقه خسارتی واقعی و پیش‌بینی شده

ضریب همبستگی	پیش‌بینی‌های صحیح	پیش‌بینی‌های نادرست	تعداد کل رکوردها
آموزش ۹۱ درصد	۱۷۸۷۳	۱۶۴۳	۱۹۵۱۶
آزمایش ۹۱ درصد	۴۴۴۴	۴۱۰	۴۸۵۴

بر اساس این اطلاعات، مدل می‌تواند با دقت ۹۱ درصد پیش‌بینی کند که یک مشتری خاص دچار خسارت می‌شود یا خیر. پس از این مرحله، چنانچه طبق پیش‌بینی انجام شده مشتری در دسته بیمه‌گذاران خسارتی قرار گرفت، داده‌های این مشتری خاص در پایگاه داده‌های مشتریان خسارتی موجود وارد می‌شود که در این پژوهش ۱۲۱۶۹ بیمه‌گذار است. در ادامه، عملیات مرحله دوم اجرا می‌شود تا میزان خسارت بالقوه او پیش‌بینی شود.

در جدول ۵ اطلاعاتی درباره همبستگی بین میزان خسارت‌های واقعی و پیش‌بینی شده بیمه‌گذاران یا به‌بیانی، دقت پیش‌بینی مدل و همچنین میزان خطاها درج شده است. دقت پیش‌بینی مدل برای این مرحله ۸۷ درصد است.

جدول ۵. اطلاعات همبستگی بین میزان خسارت‌های واقعی و پیش‌بینی شده

ضریب همبستگی	خطای مطلق میانگین ^۱	ضریب تعیین ^۲	تعداد رکوردها
آموزش ۸۹ درصد	۰/۱۷۶	-۰/۷۹۸	۹۷۳۱
آزمایش ۸۷ درصد	۰/۱۷۷	-۰/۷۶۳	۲۴۳۸

1. Mean Absolute Error (MAE)

2. R2

تعیین ضرایب متغیرهای اثرگذار در پیش‌بینی

جدول ۶ اهمیت متغیرهای پیش‌بین طبقه خسارتی بیمه‌گذاران بیمه بدنه اتومبیل را براساس یافته‌های پژوهش نشان می‌دهد. جدول ۷ نیز اهمیت متغیرهای پیش‌بین میزان خسارت بیمه‌گذاران بیمه بدنه اتومبیل را به نمایش گذاشته است.

جدول ۶. ضریب تأثیر متغیرهای پیش‌بین طبقه خسارتی

نام متغیر پیش‌بین	میزان اثرگذاری	نام متغیر پیش‌بین	میزان اثرگذاری
سابقه بیمه‌گذار	۲۳ درصد	تعداد سیلندر	۵/۴ درصد
جنسیت	۱۹/۳ درصد	نوع پلاک	۴/۶ درصد
گروه جمعیتی	۱۵/۸۵ درصد	گروه اتومبیل	۱/۹۵ درصد
کاربری اتومبیل	۷/۴ درصد	سن بیمه‌گذار	۱/۷۵ درصد
رنگ اتومبیل	۶/۶ درصد	ارزش بیمه‌شده	۱/۲ درصد
سال تولید اتومبیل	۵/۹ درصد	وضعیت تأهل	۱/۱۵ درصد
سیستم اتومبیل	۵/۸ درصد		

جدول ۷. ضریب تأثیر متغیرهای پیش‌بین میزان خسارت

نام متغیر پیش‌بین	میزان اثرگذاری	نام متغیر پیش‌بین	میزان اثرگذاری
ارزش بیمه‌شده	۲۲/۸۱ درصد	نوع پلاک	۶/۲۱ درصد
رنگ اتومبیل	۱۱/۱۶ درصد	سابقه بیمه‌گذار	۵/۸۹۵ درصد
سیستم اتومبیل	۱۰/۳۹ درصد	گروه جمعیتی	۳/۲۳ درصد
تعداد سیلندر	۹/۸ درصد	گروه اتومبیل	۲/۹۱۵ درصد
سال تولید اتومبیل	۸/۴۷ درصد	سن	۲/۶ درصد
کاربری اتومبیل	۷/۳۳ درصد	وضعیت تأهل	۲/۰۱ درصد
جنسیت	۷/۲۳ درصد		

نتیجه گیری

استفاده بیمه مرکزی از ابزار تعرفه برای نظارت بر شرکت‌های بیمه با فضای پیش روی صنعت بیمه کشور که آزادسازی و مقررات‌زدایی دو مؤلفه اصلی آن است، سازگاری نداشت. اکنون با توجه به اصلاح نظام تعرفه و شروع آزادسازی تدریجی نرخ‌ها، مسئولیت نظارت و تعیین نرخ رشته‌های مختلف بیمه‌ای و انواع ریسک به خود شرکت‌های بیمه واگذار شده است. بنابراین، باید این شرکت‌ها به ابزار تحلیل ریسک قدرتمندی دسترسی داشته باشند تا بتوانند ریسک دریافتی را به‌خوبی مدیریت کنند. با استفاده از روش‌های صحیح و کاربردی، شرکت بیمه می‌تواند عملکرد و بهره‌وری خود را تا حد زیادی بهبود بخشد. به‌طور قطع، این کار موجب رقابتی‌تر شدن بازار صنعت بیمه و رشد شرکت‌هایی می‌شود که خلاقیت و استفاده از دانش روز را در دستور کار خود قرار می‌دهند.

براساس نتایج این پژوهش مشاهده شد که علاوه بر مشخصه‌های اتومبیل (مثل تعداد سیلندر، سال ساخت، کاربری و...)، مشخصه‌های بیمه‌گذار نیز در بروز خسارت تأثیر عمده‌ای دارد. این نکته نشان می‌دهد پارامترهای استفاده‌شده برای نرخ‌گذاری بیمه، به‌بازنگری و توجه بیشتر به سوابق و رفتار و ویژگی‌های بیمه‌گذاران نیاز دارند؛ زیرا در هر تصادف، راننده اتومبیل و ویژگی‌های او نیز نقش بسیار مهمی دارند.

در این پژوهش پس از تعیین عوامل اثرگذار بر وقوع خسارت، طبقه خسارتی بیمه‌گذاران پیش‌بینی شد؛ بدین معنا که با در نظر گرفتن همزمان مشخصه‌های بیمه‌گذار و اتومبیل، پیش‌بینی شود آیا بیمه‌گذاری مشخص دچار خسارت می‌شود یا خیر. طبق نتایج به‌دست‌آمده در مرحله نخست، یعنی پیش‌بینی طبقه خسارتی بیمه‌گذاران، نتایج زیر به‌دست آمد:

- احتمال وقوع خسارت برای بیمه‌گذاران جوان‌تر کمتر از بیمه‌گذاران میان‌سال، و برای بیمه‌گذاران میان‌سال کمتر از بیمه‌گذاران مسن است. در واقع، می‌توان نتیجه گرفت سن بیمه‌گذار با احتمال وقوع خسارت رابطه مستقیمی دارد؛
- احتمال وقوع خسارت برای بیمه‌گذاران زن بیشتر از بیمه‌گذاران مرد و برای بیمه‌گذاران مجرد بیشتر از بیمه‌گذاران متأهل است. اصولاً بیمه‌گذاران مجرد نسبت به بیمه‌گذاران متأهل گرایش کمتری به رعایت قوانین و رانندگی صحیح دارند. در کشورهای دیگر نیز برای بیمه‌گذاران مجرد حق بیمه بیشتری تعیین می‌کنند؛
- احتمال وقوع خسارت برای بیمه‌گذاران ساکن در شهرهای پرجمعیت بیشتر از سایر گروه‌هاست. در شهرهای پرجمعیت به‌دلیل تعداد بسیار زیاد اتومبیل و ترافیک سنگین و نیز وضعیت دشوار زندگی در کلان‌شهرها که عموماً موجب شتاب در رانندگی، اضطراب و

کاهش آستانه تحمل رانندگان می‌شود، این آمار طبیعی به نظر می‌رسد. همچنین احتمال وقوع خسارت برای بیمه‌گذاران ساکن در شهرهای کم‌جمعیت بیشتر از شهرهایی با گروه جمعیتی متوسط است؛

- هرچه بیمه‌گذار سابقه عدم خسارت بیشتری داشته باشد، احتمال وقوع خسارت برای او کمتر است. بیشترین احتمال وقوع خسارت برای بیمه‌گذارانی است که در سال‌های قبل دچار خسارت شده‌اند. پس از آنها بیشترین احتمال وقوع خسارت مربوط به بیمه‌گذاران جدید است؛

- احتمال وقوع خسارت برای اتومبیل‌های چهار سیلندر بیشتر از اتومبیل‌های شش و هشت سیلندر است. تعداد اتومبیل‌های چهار سیلندری که دچار خسارت می‌شوند، از اتومبیل‌های چهار سیلندری چنین نیست؛

- احتمال وقوع خسارت برای کامیون کمتر از سایر اتومبیل‌هاست. اتومبیل‌های وانت در رتبه بعدی قرار می‌گیرند. بیشترین احتمال وقوع خسارت مربوط به اتومبیل‌های پراید و پژو ۲۰۶ است. همچنین غیر از اتومبیل‌های وانت و کامیون، به نظر می‌رسد در سایر اتومبیل‌ها تعداد بیمه‌گذارانی که دچار خسارت می‌شوند، بیشتر از بیمه‌گذارانی است که دچار خسارت نمی‌شوند؛

- احتمال وقوع خسارت برای گروه سواری بیشتر از گروه بارکش است؛

- احتمال وقوع خسارت برای اتومبیل‌هایی با رنگ‌های سبز و مشکی از سایر رنگ‌ها بیشتر است. رنگ‌های نقره‌ای و خاکستری در رتبه بعد قرار می‌گیرند و در نهایت، برای رنگ‌های زرد، آبی، قرمز و سفید احتمال وقوع خسارت نسبت به اتومبیل‌های دیگر کمتر است؛

- احتمال وقوع خسارت برای تاکسی‌ها از شخصی‌ها و بارکش‌ها بسیار بیشتر است. همچنین احتمال وقوع خسارت در شخصی‌ها بیشتر از بارکش‌هاست؛

- احتمال وقوع خسارت برای اتومبیل‌های با پلاک‌های عمومی، از پلاک شخصی‌ها بسیار بیشتر است؛

- احتمال وقوع خسارت برای اتومبیل‌های مدل بالاتر (جدیدتر) از اتومبیل‌های قدیمی‌تر بیشتر است؛

در مرحله بعد، میزان خسارت بیمه‌گذاران پیش‌بینی شد. در این قسمت، پایگاه داده شامل همه بیمه‌گذاران خسارت‌دیده است. در این مرحله متغیرهای مستقل همان متغیرهای مرحله قبل

هستند، اما متغیر وابسته، میزان خسارت در نظر گرفته شده است. در این مرحله نتایج زیر به دست آمد:

- بیشتر خسارت‌های شدید مربوط به بیمه‌گذاران ۴۰ تا ۶۰ ساله است. خسارت‌های جزئی و خفیف بیشتر در بیمه‌گذاران کمتر از ۲۵ سال و بیشتر از ۷۰ سال مشاهده می‌شود؛
- بیمه‌گذاران مرد نسبت به بیمه‌گذاران زن بیشتر دچار خسارت‌های شدید می‌شوند. با مقایسه این نتیجه با نتیجه به دست آمده از متغیر جنسیت در پیش‌بینی طبقه خسارتی، درمی‌یابیم که احتمال وقوع خسارت برای بیمه‌گذاران زن بیشتر است، اما خسارت‌های شدیدتر برای بیمه‌گذاران مرد رخ می‌دهد. در نتیجه، شرکت‌های بیمه باید برای بیمه‌گذاران مرد محتاطانه‌تر عمل کنند؛
- بیمه‌گذاران متأهل نسبت به بیمه‌گذاران مجرد بیشتر دچار خسارت‌های شدید می‌شوند. این نتیجه جای بررسی بیشتری دارد؛ زیرا در کشورهای دیگر، بیمه‌گذاران متأهل نسبت به بیمه‌گذاران مجرد از تخفیفاتی بهره‌مند می‌شوند و در واقع متأهل بودن نوعی امتیاز به حساب می‌آید. در حالت کلی، بیمه‌گذاران متأهل به تبع باید گرایش بیشتری به سمت رانندگی مطمئن و رعایت قوانین داشته باشند، اما طبق نتیجه به دست آمده، بیمه‌گذاران متأهل در ایران بیشتر در معرض خسارت‌های شدید قرار دارند. این در حالی است که در نتیجه‌گیری‌های بخش طبقه خسارتی مشاهده شد که بیمه‌گذاران مجرد بیشتر در معرض خسارت قرار داشتند، اما در اینجا نتیجه می‌گیریم که آن خسارت‌ها شدید نبودند و شرکت‌های بیمه باید در برابر بیمه‌گذاران متأهل محتاطانه‌تر عمل کنند؛
- خسارت‌های بسیار شدید بیشتر برای بیمه‌گذاران ساکن شهرهای پرجمعیت رخ می‌دهد. این نتیجه می‌تواند هم به دلیل وجود اتومبیل‌های گران‌قیمت‌تر در این شهرها باشد و هم به این دلیل که تعداد اتومبیل‌ها و میزان ترافیک در این شهرها بسیار بیشتر از شهرهای دیگر است. همچنین در خصوص این نتیجه نباید از وضعیت زندگی در کلان‌شهرها که خستگی، شتاب‌زدگی در رفت‌وآمد، اضطراب، کاهش آستانه تحمل و... را در پی دارد، چشم‌پوشی کرد. بیمه‌گذاران ساکن شهرهای کم‌جمعیت نیز نسبت به تعدادشان دچار خسارت‌های به نسبت شدیدی می‌شوند. دلیل چنین خسارت‌هایی را می‌توان در کیفیت نه‌چندان مناسب جاده‌ها و بی‌توجهی به ایمن‌سازی راه‌ها در مسیرهای منتهی به این شهرها جست‌وجو کرد. همچنین در شهرهایی با تراکم جمعیتی متوسط، بیشتر خسارت‌های خفیف و جزئی مشاهده می‌شود؛
- خسارت‌های شدید برای بیمه‌گذاران جدید رخ می‌دهد. به طور کلی، می‌توان نتیجه گرفت هرچه سابقه عدم خسارت بیمه‌گذار بیشتر باشد، کمتر دچار خسارت‌های شدید می‌شود.

- همچنین شدت خسارت به بیمه‌گذارانی که در سال قبل دچار خسارت شده‌اند، نسبت به بیمه‌گذارانی که در دو یا چند سال قبل دچار خسارت شده‌اند، بیشتر است؛
- خسارت‌های شدید برای اتومبیل‌های شش سیلندر رخ می‌دهد. شدت خسارت‌ها در اتومبیل‌های چهار سیلندر بیشتر از اتومبیل‌های هشت سیلندر است؛
 - خسارت‌های شدید به کامیون‌ها وارد می‌شود. کامیون‌ها هم قیمت بیشتری نسبت به خیلی از اتومبیل‌ها دارند و هم اینکه بیشتر در جاده‌های خارج شهر و راه‌های نه‌چندان ایمن تردد می‌کنند. همچنین رانندگان کامیون معمولاً به دلیل پیمودن مسافت‌های طولانی، بیشتر دچار خستگی و خواب‌آلودگی می‌شوند که از اصلی‌ترین عوامل بروز تصادفات است. خسارت‌های جزئی بیشتر در اتومبیل پراید مشاهده می‌شود؛ زیرا پراید در مقایسه با سایر اتومبیل‌ها قیمت کمتری دارد و به‌نوعی اتومبیل درون‌شهری محسوب می‌شود؛
 - خسارت‌های شدید برای گروه بارکش و خسارت‌های خفیف و جزئی برای اتومبیل‌های سواری رخ می‌دهد؛
 - اتومبیل‌های سفیدرنگ بیشتر دچار خسارت‌های شدید می‌شوند و خسارت‌های نسبتاً شدید در اتومبیل‌هایی با رنگ‌های زرد و قرمز مشاهده می‌شود؛
 - اتومبیل‌های بارکش دچار خسارت‌های شدیدتری نسبت به سایر اتومبیل‌ها می‌شوند. همچنین شخصی‌ها در مقایسه با تاکسی‌ها بیشتر در معرض خسارت‌های شدید قرار دارند؛
 - خسارت‌های شدید بیشتر به اتومبیل‌های با پلاک عمومی وارد می‌شود. در پلاک شخصی‌ها نیز خسارت‌های نسبتاً شدید دیده می‌شود؛
 - اتومبیل‌هایی که مدل جدیدتری دارند نسبت به مدل‌های قدیمی‌تر بیشتر در معرض خسارت‌های شدید قرار می‌گیرند، این نتیجه می‌تواند با قیمت بیشتر اتومبیل‌های جدیدتر ارتباط مستقیمی داشته باشد؛
 - میزان خسارت در حادثه‌ای خاص برای اتومبیل‌هایی که قیمت زیادتری دارند، بیشتر است و برعکس. به‌طور مثال، اگر خودروی پژو پارس و خودروی پراید دقیقاً دچار یک نوع تصادف خاص شوند (مانند تصادف از پشت)، خسارت واردشده به اتومبیل پژو پارس بیشتر خواهد بود. این امر ناشی از دلایل متعددی چون تفاوت در قیمت لوازم جانبی، نرخ دستمزد تعمیر، افت قیمت اتومبیل و... است؛
- شرکت بیمه با دانستن میزان خسارت بالقوه مشتریان، می‌تواند با توجه به شرایط و ضوابط و با در نظر گرفتن ضریب خسارت مدنظر خود، مقدار حق بیمه‌ای که آن مشتری باید بپردازد را تعیین کرده و خدمات خود را سفرشی‌سازی کند.

پیشنهادها

- در این پژوهش با توجه به اینکه فراوانی هفت نوع خودروی پراید، پژو ۴۰۵، پژو ۲۰۶، سمند، کامیون، وانت و پژو پارس نسبت به سایر انواع خودروهای بیمه شده بیشتر بود، تنها بر این هفت نوع اتومبیل تمرکز شد. می توان در پژوهش دیگری، سایر اتومبیل هایی که فراوانی نسبتاً مشابهی دارند را بررسی کرد؛
- اقدامات و عملیات پیش بینی خسارت باید به طور دائم انجام شوند تا ارزش آنها به دامنه زمانی خاص محدود نشود و به صورت نوعی فرایند باز طی زمان ادامه یابد. در صورت استفاده شرکت های بیمه از چنین سیستمی، اطلاعاتی که برای هر بیمه گذار جدید دریافت می شود باید به پایگاه داده اضافه شود تا نتایج دقیق تری به دست آید؛
- هنگام صدور بیمه نامه، اطلاعات کامل و جامعی از مشتریان دریافت شود تا بتوان با تحلیل آنها به نتایج کامل تر و دقیق تری که موجب سودآوری بیشتر شرکت باشد، دست یافت؛
- دسترسی به داده های جرائم رانندگی بیمه گذاران، به شناسایی مشتریان و در نتیجه پیش بینی دقیق تر میزان خسارت آنها کمک شایان توجهی می کند. داده های جرائم رانندگی در اختیار پلیس راهور است، چنانچه این داده ها در سوابق بیمه گذاران موجود باشد، یعنی این دو بانک اطلاعاتی در پیوند با یکدیگر باشند، می توان نتایج بسیار دقیق تری از شناسایی رفتار بیمه گذاران و خطرپذیری آنها به دست آورد. پیشنهاد می شود شرکت های بیمه و پلیس راهور در این زمینه با یکدیگر همکاری کنند.

References

- Anbari, A., Nadali, A. & Eslami Nosrat Abadi, H. (2010). *Comparing data mining algorithms for predicting auto insurance policy holders` risk: 4th Iran data mining conference*, Tehran: Sharif University of Technology, December 1st, 1-10. (in Persian)
- Choobdar, S. (2008). *Designing a framework for the future customers of vehicle collision coverage based on data mining* (Master`s Thesis), Tarbiat Modarres University, Tehran. (in Persian)
- Fathnejad, F. & Izadparast, M. (2011). Presenting a framework for predicting damage level of vehicle collision coverage policy holders with using data mining approach. *Insurance World Updates*, 156(3), 15-32. (in Persian)
- Ghanty, P., Paul, S. & Pal, N. (2009). NEUROSVM: An architecture to reduce the effect of the choice of kernel on the performance of SVM. *Journal of Machine Learning Research*, 10(3), 591-622.

- Gharekhani, M. & Abolghasemi, M. (2011). Data mining applications in insurance industry. *Insurance World Updates*, 158(5), 5-21. (in Persian)
- Guo, L. (2003). *Applying DM in property/casualty insurance*. University of Central Florida: CAS Committee on Management Data and Information, Florida.
- Haji Heydari, N., Khale`, S. & Farahi, A. (2011). Classifying the risk of vehicle collision coverage policy holders with using data mining algorithms. *Iranian Journal of Insurance Research*, 26(4), 107-129. (in Persian)
- Hanafizadeh, P. & Rastkhiz Paydar, N. (2011). A model for risk-based clustering of vehicle collision coverage customers with using data mining technique. *Iranian Journal of Insurance Research*, 26(2), 55-81. (in Persian)
- Iran Insurance Research Center. (2011). *Insurance Research Center Annual Statistical Report*. Tehran. (in Persian)
- Kiavarz Moghaddam, H. & Wang, X. (2014). Vehicle accident severity rules mining using fuzzy granular decision tree. *Rough Sets and Current Trends in Computing*, 36(85), 280-287.
- Mansouri, M. & Kargar, M.J. (2014). Analysis and monitoring of the traffic suburban road accidents using data mining techniques: A case study of Isfahan province in Iran. *The Open Transportation Journal*, 8(1), 39-49.
- Mohammadi, Sh. & Alizadeh, S. (2014). Analyzing the problems of Ayandeh bank`s branches across the country with using data mining technique. *Journal of Information Technology Management*, 6(2), 333-350. (in Persian)
- Momeni, M. (2006). *New Approaches in Operations Research* (1st ed), University of Tehran: Faculty of Management, Tehran. (in Persian)
- Newstead, S. & D`Elia, A. (2007). *An investigation into the relationship between vehicle color and crash risk*. Monash University Accident Research Center: Report No.263, Melbourne.
- Ngai, E., Xiu, L. & Chau, D. (2009). Application of Data Mining techniques in customer relationship management: a Literature Review and Classification. *Expert Systems with Application*, 36(2), 592-602.
- Nisbet, R., Elder, J. & Miner, G. (2009). *Handbook of statistical analysis and Data Mining applications*, Burlington: Academic Press.
- Parnitzke, T. (2008). *A Discussion of Risk Assessment Methods for the German Automobile Insurance Industry* (Doctoral Dissertation), University of St.Gallen: Institute of Insurance Economics, St. Gallen.

- Radfar, R., Nezafati, N. & Yusefi Asl, S. (2014). Clustering customers of e-banking with using data mining algorithms. *Journal of Information Technology Management*, 6(1), 71-90. (in Persian)
- Raut, R. & Nathe, A. (2015). Comparative study of commercial data mining tools. *International Journal of Electronics, Communication & Self Computing Science and Engineering*, 8, 128-132.
- Vosugh, M., Taghavifard, M.T. & Alborzi, M. (2014). Detecting fraud in credit cards with using artificial neural networks. *Journal of Information Technology Management*, 6(4), 721-746. (in Persian)
- Yeo, A., Smith, K., Willis, R. & Brooks, M. (2001). Modeling the effect of premium changes on motor insurance customer retention rates using neural networks. *Computational Science*. 2074, 390-399.

