

(مقاله پژوهشی)

گروه‌بندی سن بیمه‌گذاران رشته بیمه شخص ثالث در ایران، با توجه به نرخ خسارت‌ها

مهناز منطقی‌پور^{۱*}

وحید قربانی^۲

محبوبه اعلائی^۳

تاریخ پذیرش: ۱۴۰۰/۰۴/۱۵

تاریخ دریافت: ۱۳۹۹/۱۲/۰۴

چکیده

مطابق قانون، بیمه مرکزی ج.ا. ایران موظف است در محاسبه حق بیمه‌های رشته بیمه‌ای خسارات وارد شده به شخص ثالث در اثر حوادث ناشی از وسایل نقلیه، ویژگی‌های راننده نیز مورد توجه قرار دهد؛ از این‌رو، در این مقاله با استفاده از داده‌های یک شرکت بیمه طی شش‌سال به بررسی نحوه اثرگذاری ویژگی سن بیمه‌گذاران بر وقوع خسارت‌ها پرداخته‌ایم. نتایج آزمون دو نمونه‌ای کولموگروف-اسمیرنف، گویای اثرگذاری ویژگی سن بر وقوع خسارت‌ها است. به دلیل تفاوت جمعیت سنی بیمه‌گذاران چهار گروه اصلی وسایل نقلیه، محاسبات، برای گروه‌های مختلف خودرو به صورت مجزا انجام شده است. با استفاده از احتمال شرطی، احتمال وقوع خسارت را برای هر رده سنی محاسبه نموده‌ایم. همچنین براساس سه ویژگی ثابت شده از بیمه‌گذاران شامل سن، جنسیت و نوع مشغولیت درخت تصمیم را محاسبه نموده‌ایم. براساس نتایج، در خودروهای سواری، بارکش و موتورسیکلت، احتمال وقوع خسارت به ترتیب برای بیمه‌گذاران با سنین کمتر از ۲۲ سال، ۳۰ سال و ۲۵ سال به میزان قابل توجهی بالاتر است. بنابراین بر اساس اصول بیمه‌سنجی، از بیمه‌گذاران این سنین حق بیمه بالاتری می‌باید دریافت شود. همچنین، بیش از ۵۰٪ از خسارت‌های منجر به آسیب‌های جانی در تمام انواع خودرو، متعلق به رانندگانی است که بیمه‌گذار نبوده‌اند؛ بنابراین خودروهایی که رانندگان متعدد دارند، می‌باید حق بیمه بالاتری پرداخت کنند. ویژگی سن بیمه‌گذاران خودروهای اتوکار تأثیر قابل توجهی بر نرخ خسارت‌ها نداشته است.

کلیدواژه‌ها: بیمه شخص ثالث وسایل نقلیه موتوری زمینی، احتمال شرطی، درخت تصمیم، ماتریس دقت رده‌بندی.

طبقه‌بندی JEL: C11, C53, C55.

Email: Manteqipour@irc.ac.ir

۱. مسئول میز داده کاوی، گروه پژوهشی بیمه الکترونیک، پژوهشکده بیمه، تهران، ایران (نویسنده مسئول)

Email: vahidghorbani@hotmail.com

۲. مشاور میز داده کاوی، گروه پژوهشی بیمه الکترونیک، پژوهشکده بیمه، تهران، ایران

Email: aalaei@irc.ac.ir

۳. استادیار گروه پژوهشی بیمه اشخاص، پژوهشکده بیمه، تهران، ایران

۱. مقدمه

یکی از مهم‌ترین موضوعات مورد بحث در صنعت بیمه، نحوه تعیین نرخ منصفانه حق بیمه است. منظور از نرخ منصفانه حق بیمه، نرخ‌هایی است که تقریب مناسبی از هزینه مورد انتظار خسارت‌ها ارائه می‌کند. هرچه نرخ‌های حق بیمه‌ها به نرخ‌های منصفانه نزدیک‌تر باشند، ضریب نفوذ بیمه در سطح جوامع افزایش می‌یابد. براساس مستندات انجمن اکچوئران استرالیا، در شرایطی که نرخ‌ها منصفانه نیستند، ریسک‌های پرخطر حق بیمه کم و ریسک‌های کم‌خطر حق بیمه بالایی پرداخت می‌کنند. این موضوع موجب جذب بیشتر بیمه‌گذاران پرخطر و بنابراین افزایش هزینه‌ها برای بیمه‌گران می‌گردد. از آنجا که شرکت‌های بیمه، بنگاه‌های اقتصادی هستند، ناچار به افزایش نرخ حق بیمه‌ها می‌شوند. با ادامه این چرخه، ضریب نفوذ بیمه در جوامع فاقد نرخ‌های منصفانه، کاهش می‌یابد که به‌وضوح اثرات نامطلوب اجتماعی و اقتصادی دربر خواهد داشت (تامز، ۲۰۰۹). به عبارتی نرخ‌های غیرمنصفانه منجر به انتخاب نامساعد ریسک در شرکت‌های بیمه می‌گردد (بدسورث^۲ و همکاران ۲۰۲۱). انتخاب نامساعد که در ادبیات اقتصادی از آن تحت عنوان «کژگزینی» نیز یاد می‌شود، موضوعی است که در تمام رشته‌های بیمه‌ای حائز اهمیت است (عربیان و همکاران، ۱۳۹۹). همچنین ضریب نفوذ بیمه بر رشد اقتصادی کشورها نیز تأثیر دارد. «محقق‌زاده» و همکاران نشان داده‌اند که توسعه بیمه‌های زندگی و غیرزندگی بر رشد اقتصادی کشورهای در حال توسعه تأثیر بیشتری دارد (محقق‌زاده و همکاران، ۱۳۹۶).

رده‌بندی مناسب ریسک‌ها در صنعت بیمه از اهمیت بسیاری برخوردار است. یکی از مهم‌ترین کاربردهای آن تعیین نرخ‌های منصفانه است (دنویت^۳ و همکاران ۲۰۱۹). در صورت امکان، رده‌بندی براساس تحلیل داده‌های ثبت شده از ریسک‌ها انجام می‌شود. برای ساختن رده‌های ریسک برخی ملاحظات آماری لحاظ می‌شود که شامل همگن و معتبر بودن رده‌ها و پایدار بودن پیش‌بینی‌ها می‌باشد. همگن بودن به معنی شبیه بودن ریسک‌های یک رده در هزینه‌های قابل پیش‌بینی است و معتبر بودن به معنی امکان پیش‌بینی آماری هزینه‌های رده‌ها است؛ ازجمله لازم است تعداد کافی عضو در هر رده وجود داشته باشد. مهم‌ترین ویژگی یک رده ریسک، امکان پیش‌بینی هزینه‌های آینده است؛ همچنین مهم است که ویژگی‌های یک رده برای اهداف پیش‌بینی مناسب باشند و امکان پیش‌بینی‌های پایدار وجود داشته باشد (پاردی^۴، ۲۰۱۴).

^۱. Thomas, 2009

^۲. Bedsworth, et al., 2021

^۳. Denuit, et al., 2019

^۴. Parodi, 2014

در اغلب کشورها بیمه خسارات وارد شده به شخص ثالث در اثر حوادث ناشی از وسایل نقلیه اجباری است. از آنجا که نحوه صدور این بیمه‌نامه‌ها توسط شرکت‌های بیمه می‌تواند تبعات اجتماعی و اقتصادی داشته باشد و طیف مشتریان و بیمه‌گذاران آن بسیار وسیع است، نحوه تعیین نرخ منصفانه در این رشته بیمه‌ای از اهمیت دو چندان برای احاد جامعه برخوردار است. علاوه بر اثراتی که چگونگی انجام عملیات بیمه‌گری در این رشته بیمه‌ای بر صنعت بیمه دارد، می‌تواند بر فرهنگ رانندگی ایمن و کاهش آمار تلفات جاده‌ای کشورها نیز تأثیرگذار باشد. شرکت‌های بیمه معتبر در دنیا به منظور تعیین نرخ منصفانه بیمه‌نامه در این رشته بیمه‌ای، ویژگی‌های رانندگان، وسایل نقلیه و نحوه استفاده ایشان از وسایل نقلیه را برای رده‌بندی ریسک‌ها و تعیین نرخ منصفانه در هر رده لحاظ می‌نمایند.

در کشور ایران، قانون‌گذار به درستی اهمیت نحوه محاسبه نرخ بیمه شخص ثالث را درک کرده است و به منظور تدقیق نحوه نرخ‌گذاری رشته بیمه شخص ثالث، سازمان‌های مرتبط را ملزم به همکاری و اقدام در این راستا نموده است. مطابق ماده ۱۸ قانون بیمه اجباری خسارات وارد شده به شخص ثالث در اثر حوادث ناشی از وسایل نقلیه، سقف حق بیمه این رشته بیمه‌ای که در ادامه این پژوهش به اختصار از آن به عنوان «رشته بیمه شخص ثالث» یاد می‌کنیم، می‌باید براساس سه دسته از ویژگی‌ها محاسبه شود که عبارتند از: ۱- ویژگی‌های وسیله نقلیه از قبیل: نوع کاربری، سال ساخت و وضعیت ایمنی آن؛ ۲- سوابق رانندگی و بیمه‌ای بیمه‌گذار شامل: نمرات منفی و تخلفات ثبت شده توسط نیروی انتظامی ج.ا. ایران موضوع قانون رسیدگی به تخلفات رانندگی، خسارت‌های پرداختی توسط بیمه‌گر یا صندوق، بابت حوادث منتسب به وی؛ ۳- رایج بودن استفاده از وسیله نقلیه برای اقشار متوسط و ضعیف شامل: موتورسیکلت و خودروهای سواری ارزان قیمت. براساس تبصره ۲ همین ماده قانونی، بیمه مرکزی با همکاری نیروی انتظامی موظف است تا حداکثر پایان برنامه پنج‌ساله ششم توسعه (۱۳۹۶-۱۴۰۰)، امکان صدور بیمه‌نامه شخص ثالث براساس ویژگی‌های راننده را فراهم نماید. به منظور پیاده‌سازی این ماده قانونی لازم است نحوه تأثیرگذاری ویژگی‌های راننده بر نرخ خسارت‌ها محاسبه شود؛ از این‌رو در این پژوهش به بحث و بررسی درباره نحوه تأثیرگذاری سن، جنسیت و نوع بیمه‌گذاران که به بیمه‌گذاران حقیقی و حقوقی تفکیک می‌شوند، پرداخته می‌شود.

در برخی از مقالات با استفاده از نظرات خبرگان صنعت بیمه ایران تلاش شده است تا نحوه اثرگذاری عوامل مختلف بر نسبت خسارت رشته بیمه شخص ثالث محاسبه شود. اگرچه نتایج چنین مقالاتی ارزشمند هستند، اما استفاده از داده‌های واقعی قابل‌اتکاتر بوده و جهت اتخاذ تصمیمات هوشمندانه کارایی بهتری دارد؛ از این‌رو در این پژوهش با استفاده از حجم قابل‌توجهی از داده‌های یک شرکت بیمه، موضوع بررسی شده است. مهم‌ترین چالش انجام چنین محاسباتی، کیفیت ثبت

داده‌ها توسط شرکت‌های بیمه است؛ علاوه بر مشکلات نحوه ثبت داده‌ها، عدم الزام شرکت‌ها به ثبت داده‌هایی چون: سال تولد بیمه‌گذاران و تعداد رانندگان، کیفیت داده‌های موردنیاز برای بررسی نحوه اثرگذاری ویژگی‌های تأثیرگذار بر نرخ خسارت‌های این رشته بیمه‌ای را کاهش داده است. تاکنون بیمه‌نامه‌های رشته بیمه شخص ثالث براساس ویژگی‌های بیمه‌گذاران و نه رانندگان، صادر شده است؛ بنابراین شرکت‌های بیمه، ویژگی‌های رانندگان را برای تحلیل و بررسی در اختیار ندارند. در این پژوهش که از داده‌های ثبت شده در پایگاه داده‌های یک شرکت بیمه طی شش سال استفاده شده است، علاوه بر ویژگی‌های بیمه‌گذاران، ویژگی‌های رانندگان مقصر حادثه را نیز لحاظ کرده‌ایم تا بتوانیم تحلیل‌های بهتری داشته باشیم.

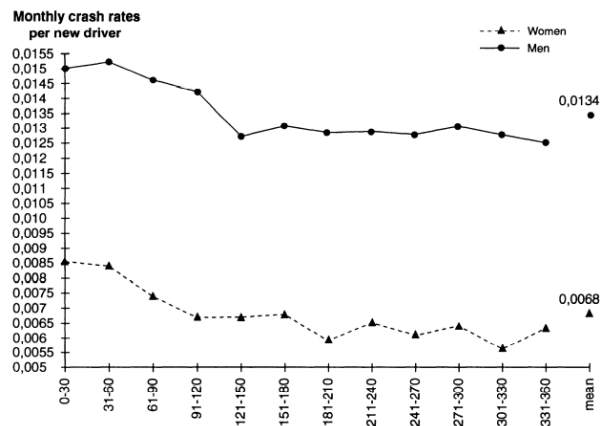
نتایج این تحقیق نشان می‌دهد ویژگی سن بیمه‌گذاران در گروه‌های مختلف خودرو شامل: اتوکار، بارکش، سواری و موتورسیکلت تأثیر متفاوتی بر نرخ خسارت‌های رشته شخص ثالث دارد و این موضوع کاملاً طبیعی است، زیرا سن بیمه‌گذاران در این گروه‌های خودرو نیز متفاوت است؛ به‌عنوان مثال، سن بیمه‌گذاران خودروهای اتوکار بیش از بیمه‌گذاران سایر گروه‌ها و سن بیمه‌گذاران موتورسیکلت به‌طور نسبی کمتر است. نتایج محاسبه احتمال وقوع خسارت برای سنین مختلف بیمه‌گذاران نشان می‌دهد که به‌ویژه برای گروه‌های سواری، بارکش و موتورسیکلت بیمه‌گذاران جوان‌تر پرخطرتر هستند. محاسبات درخت تصمیم نیز این نتیجه را تایید می‌کند. همچنین گواه بر این موضوع است که سن بیمه‌گذاران خودروهای اتوکار چندان بر نرخ خسارت‌ها تأثیرگذار نیست؛ حال آن‌که نرخ خسارت، برای خودروهای سواری، موتورسیکلت و بارکش به ترتیب برای سنین کمتر از ۲۲، ۲۵ و ۳۰ سال به میزان قابل‌توجهی بیش از سایر سنین است. در سایر کشورها، بیمه‌گذاران جوان حق بیمه‌های بالاتری پرداخت می‌کنند و در عین حال ایشان می‌توانند به‌منظور کاهش نرخ حق بیمه خود، امکان نصب ابزارهای «تلماتیک» بر خودروهایشان را برای شرکت‌های بیمه فراهم آورند. ابزارهای تلماتیک ابزارهایی هستند که بر خودروها نصب می‌شوند و می‌توانند داده‌های متعددی شامل نحوه رانندگی رانندگان را ثبت نمایند. بیمه‌گران با استفاده از داده‌های ثبت شده توسط این ابزارها، نرخ خسارت‌ها را پیش‌بینی و حق بیمه را محاسبه می‌نمایند (سلویونسکی و کوریلوویچ^۱، ۲۰۲۱).

۲- پیشینه تحقیق

نظر به اهمیت موضوع، مطالعات بسیاری برای بررسی چگونگی اثرگذاری ویژگی‌های راننده بر نرخ خسارت‌ها در کشورهای مختلف انجام شده است و نتیجه مشترک تمام مقالاتی که بر نحوه اثرگذاری

^۱. Śliwiński & Kuryłowicz, 2021

ویژگی رانندگان بر نرخ خسارت‌ها تمرکز دارند، آن است که رانندگان جوان و رانندگانی که به تازگی گواهینامه خود را اخذ کرده‌اند، پُرخطرتر هستند و در تمام رده‌های سنی نرخ خسارت‌های رانندگان خانم کمتر است. در پژوهش «مگ» و همکاران (۱۹۹۹)، نمودار سرانه تعداد خسارت‌ها در آمریکا در سال اول پس از اخذ گواهینامه، در هر دوره ۳۰ روزه به تفکیک جنسیت نشان داده شده است و قابل ملاحظه است که سرانه تعداد خسارت‌های ماهانه از ۱۵ خسارت در هر ۱۰۰۰ بیمه‌نامه به ۱۳ خسارت برای مردان و از ۸.۵ خسارت به ۶.۸ خسارت برای زنان کاهش داشته است.

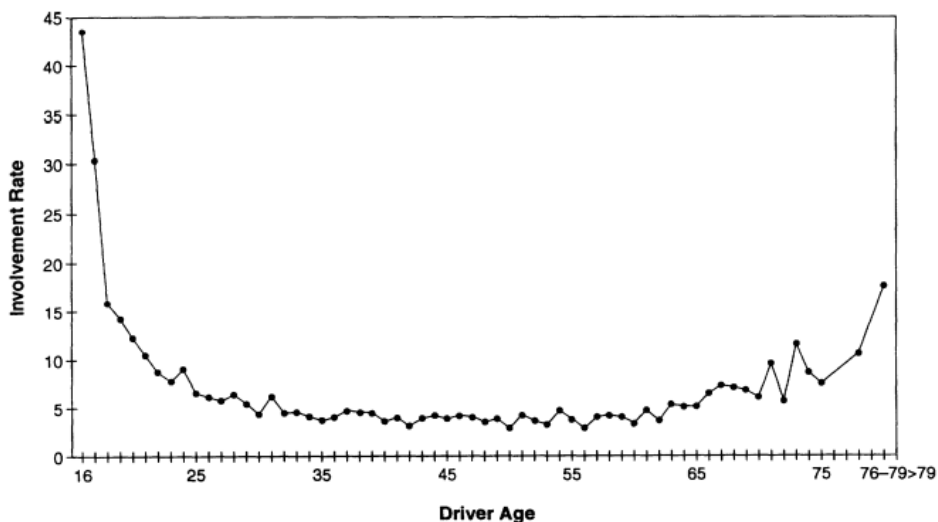


شکل ۱. سرانه تعداد خسارت‌های ماهانه در سال اول پس از اخذ گواهینامه به تفکیک جنسیت، منبع (مگ و همکاران ۱۹۹۹).

همچنین نتایج این تحقیق نشان می‌دهد، نرخ خسارت در میان افراد جوان‌تر بیشتر و در زنان در مقایسه با مردان کمتر بوده است. علاوه بر آن نرخ خسارت افراد ۱۷ الی ۱۹ سال مشابه بوده است. همچنین در پژوهش دیگری در ایالت متحده آمریکا (ویلیامز^۲، ۱۹۹۹) تعداد خسارت‌ها به ازای هر یک میلیون مایل رانندگی به تفکیک سن مطابق نمودار زیر بوده است. در این نمودار مشاهده می‌شود که نرخ خسارت افراد با سن ۱۶ سال حدود ۳ برابر افراد با سن ۱۸ الی ۱۹ سال است. همچنین نرخ سوانح رانندگی منجر به آسیب‌های جانی، برای افراد کمتر از ۲۵ سال و بیشتر از ۶۵ سال به‌طور میانگین بالاتر است.

^۱. Maag, et al., 1999

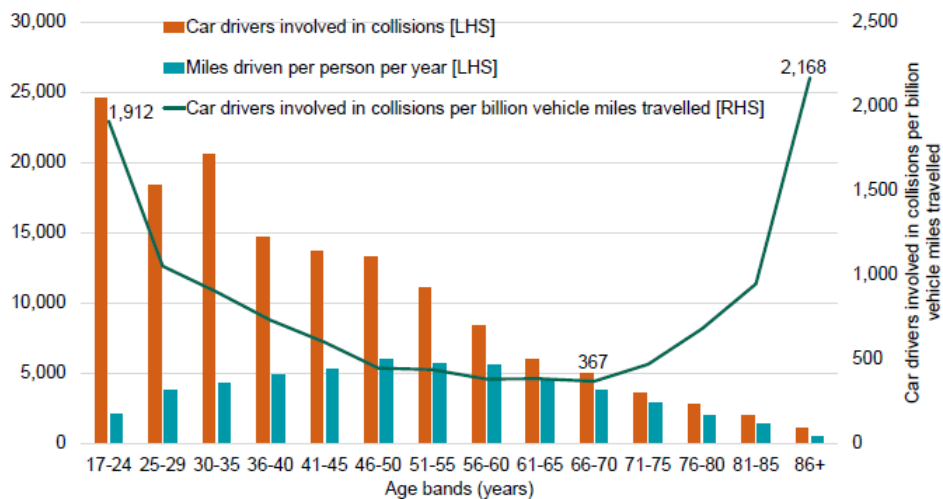
^۲. Williams, 1999



شکل ۲. مقایسه نرخ تصادفات منجر به آسیب‌های جانی برای سنین متفاوت رانندگان (ویلیامز^۱، ۱۹۹۹).

همچنین در کشور بریتانیا شاخص تعداد رانندگان حادثه‌ساز بر متوسط مسافت طی شده توسط رانندگان به تفکیک رده‌های سنی محاسبه شده است. مقادیر این شاخص در شکل ۳، قابل مشاهده است. در این شکل مشاهده می‌شود که با وجود آن‌که رانندگان با رده سنی ۱۷ الی ۲۴ سال به صورت سرانه طی سال کمتر رانندگی کرده‌اند، تعداد خسارت‌های ایشان به میزان قابل توجهی بیش از سایر سنین بوده است. به طور کلی در این نمودار نیز بالاتر بودن نرخ خسارت‌های رانندگان جوان و رانندگان بالای ۶۵ سال مشهود است.

^۱. Williams, 1999



شکل ۳. شاخص تعداد رانندگان حادثه‌ساز به متوسط مسافت طی شده توسط ایشان به تفکیک رده‌های سنی در کشور بریتانیا، (گزارش دپارتمان حمل و نقل انگلیس)^۱.

در ایران مطالعات اندکی بر این موضوع انجام شده است و بسیاری از مقالات نه بر مبنای داده‌های ثبت شده از بیمه‌نامه‌ها که بر اساس پرسش‌نامه و مصاحبه با خبرگان صنعت بیمه، (اسفندیاری و اصفهانی^۲، ۲۰۱۸) و (شمس اسفندیاری و همکاران^۳) به بررسی موضوع پرداخته‌اند. درحالی‌که ماهیت چنین پژوهشی به‌گونه‌ای است که تحلیل داده‌های واقعی و استفاده از تکنیک‌های داده کاوی می‌تواند نتایج قابل‌اتکاتری برای تصمیم‌گیرندگان فراهم نماید. «کریم‌زادگان مقدم» و همکاران (۱۳۹۴) به‌منظور تعیین نرخ‌های پویا با استفاده از داده‌ها، نشان‌دهنده‌اند که با به‌کارگیری تکنیک‌های داده‌کاوی امکان کاهش نرخ خسارت‌ها وجود دارد. در مطالعه «اسکویی» و همکاران (۱۳۹۹)، ویژگی‌های تأثیرگذار خودرو بر نرخ خسارت‌های مالی رشته بیمه شخص ثالث مورد مطالعه قرار گرفته است. در پژوهش حاضر، با استفاده از داده‌های ثبت شده در پایگاه داده‌های یک شرکت بیمه و بهره‌گیری از تکنیک‌ها داده‌کاوی، نحوه اثرگذاری ویژگی‌های ثبت شده از بیمه‌گذاران را مورد بحث و بررسی قرار خواهیم داد و چالش‌های آن را بررسی خواهیم کرد.

^۱. Department for transportation, 2018

^۲. Esfandabadi & Esfahani, 2018

^۳. Shams Esfandabadi

۳- توصیف داده‌ها

به منظور انجام محاسبات در این پژوهش، پایگاه داده‌های یک شرکت بیمه و نحوه ثبت داده‌ها در آن مورد بررسی قرار گرفته است. پایگاه‌های داده عملیاتی به منظور ذخیره‌سازی داده‌ها طراحی می‌شوند و برای تحلیل داده‌ها مناسب نیستند؛ زیرا این پایگاه‌های داده براساس اهدافی از جمله ارتقای عملکرد و سرعت پایگاه داده، ملزم به رعایت قواعد نرمال‌سازی هستند و اغلب به منظور تحلیل داده‌ها از انبارداده‌ها که غیرنرمال شده است، استفاده می‌شود؛ از این رو به منظور انجام محاسبات این مقاله، ابتدا داده‌های موردنیاز در جداولی مشابه با ساختار انبار داده توسط پژوهشگران، جمع‌آوری شده است و سپس برای تحلیل مورد استفاده قرار گرفته است. داده‌های ثبت شده در این پایگاه داده، مربوط به سال‌های ۱۳۸۷ الی ۱۳۹۲ است. از آنجا که رویه‌های دادگاهی مربوط به پرونده‌های خسارت‌های منجر به تلفات جانی زمان‌بر و ممکن است چندین سال برای حصول نتیجه نهایی زمان صرف شود؛ به کارگیری این بازه زمانی موجب می‌شود از کامل بودن تمام داده‌های خسارت مربوط به بیمه‌نامه‌های صادره طی این دوره اطمینان حاصل شود و ارزیابی ویژگی‌های مورد بررسی تحت تأثیر عدم تکمیل داده‌های خسارت قرار نگیرد.

۳-۱ توصیف آماری تعداد خسارت‌ها و بیمه‌نامه‌ها

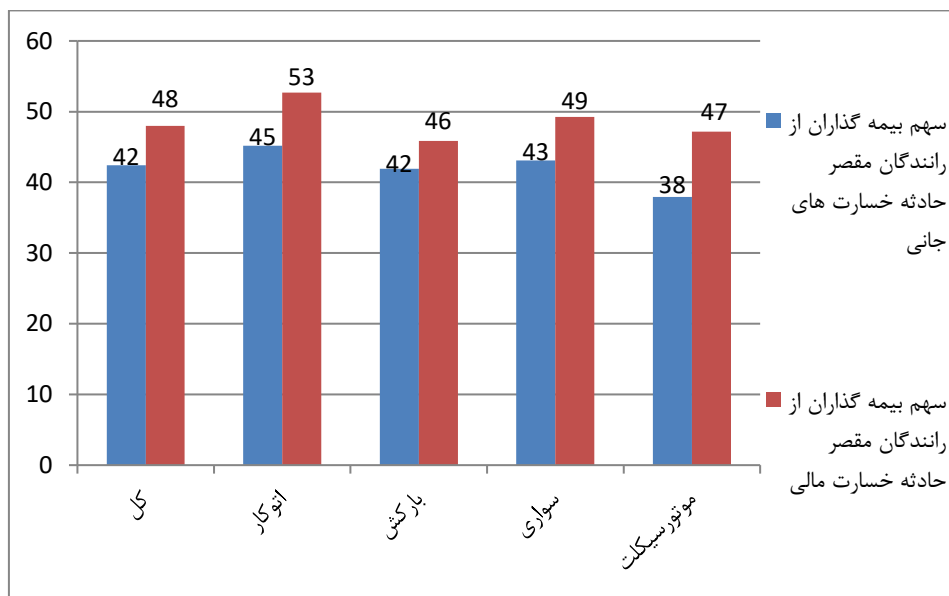
در جدول ۱، تعداد بیمه‌نامه‌ها و خسارت‌های ثبت شده در پایگاه داده‌های مورد بررسی درج شده است. همچنین درصد تعداد خسارت‌هایی که در آن بیمه‌گذاران خود راننده بوده‌اند به تفکیک گروه خودرو و خسارت‌های جانی و مالی محاسبه شده است. مطابق با این جدول، تنها حدود ۴۸٪ از خسارت‌های مالی و ۴۲٪ از خسارت‌های جانی رشته بیمه شخص ثالث متعلق به رانندگانی بوده است که خود بیمه‌گذار بوده‌اند. این آمار نشان می‌دهد که بحث راننده‌محور شدن بیمه‌نامه‌های رشته شخص ثالث از اهمیت بالایی برخوردار است. خسارت‌های جانی، خسارت‌هایی با شدت بالا هستند و آمار نشان می‌دهد در حدود ۵۸٪ از این خسارت‌ها توسط رانندگان غیربیمه‌گذار رخ می‌دهد. این موضوع نشان می‌دهد افزایش حدود ۶۰٪ نرخ حق بیمه برای بیمه‌نامه‌هایی که رانندگان متعدد را تحت پوشش قرار می‌دهد، منطقی و اصولی است. شایان ذکر است هدف قانون‌گذار از لحاظ نمودن ویژگی‌های راننده، تعیین نرخ‌های منصفانه است؛ بنابراین به تبع افزایش نرخ حق بیمه‌ها برای رانندگان پُرخطر، نرخ‌ها برای رانندگان ایمن کاهش خواهند یافت.

بیشترین نرخ خسارت رانندگان غیربیمه‌گذار مربوط به خسارت‌های جانی موتورسیکلت بوده است که حدود ۶۲٪ است و بسیار بالاست. کمترین مقدار این نسبت معادل ۵۵٪ و متعلق به خودروهای اتوکار

است. این درحالی است که براساس داده‌های مورداستفاده در این پژوهش، نسبت خسارت خودروهای اتوکار پایین‌تر از سایر انواع خودروها است. این مقادیر در شکل ۱ نیز قابل مقایسه است.

جدول ۱. آمار تعداد خسارت‌های جانی و مالی رشته بیمه شخص ثالث به تفکیک انواع گروه وسیله نقلیه.

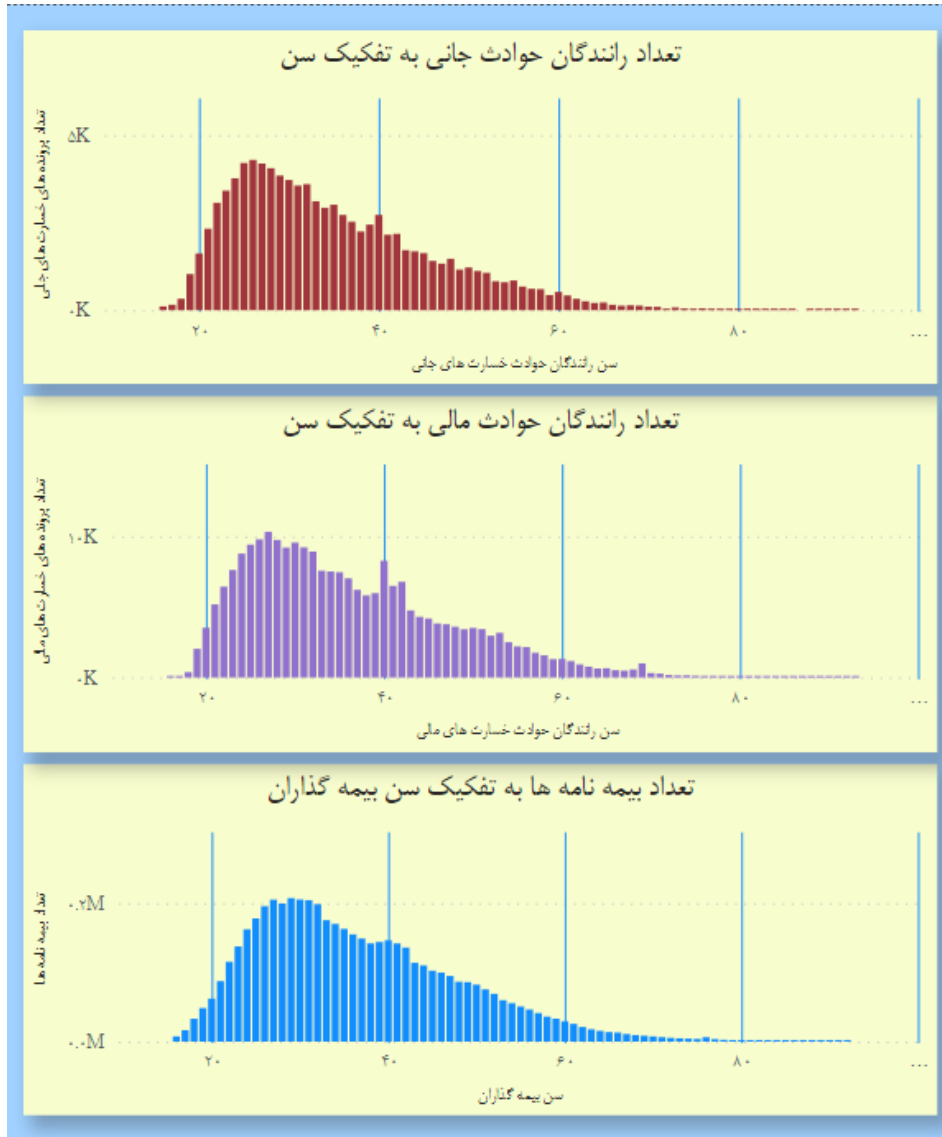
شاخص	تعداد بیمه‌نامه‌ها	تعداد پرونده‌های خسارت جانی		تعداد پرونده‌های خسارت مالی		تعداد رانندگان تصادفات منجر به خسارت مالی که بیمه‌گذار بوده‌اند		دسته‌بندی
		درصد از کل بیمه‌نامه‌ها	تعداد	درصد از کل بیمه‌نامه‌ها	تعداد	تعداد	تعداد	
کل	5693418	1	58788	4	212204	48	102972	
اتوکار	115795	2	2390	8	9820	42	5174	
بارکش	839981	1	11195	6	49869	53	22879	
سواری	2806014	1	38979	5	144206	46	71001	
موتورسیکلت	1931627	0	6224	0	8309	49	3918	



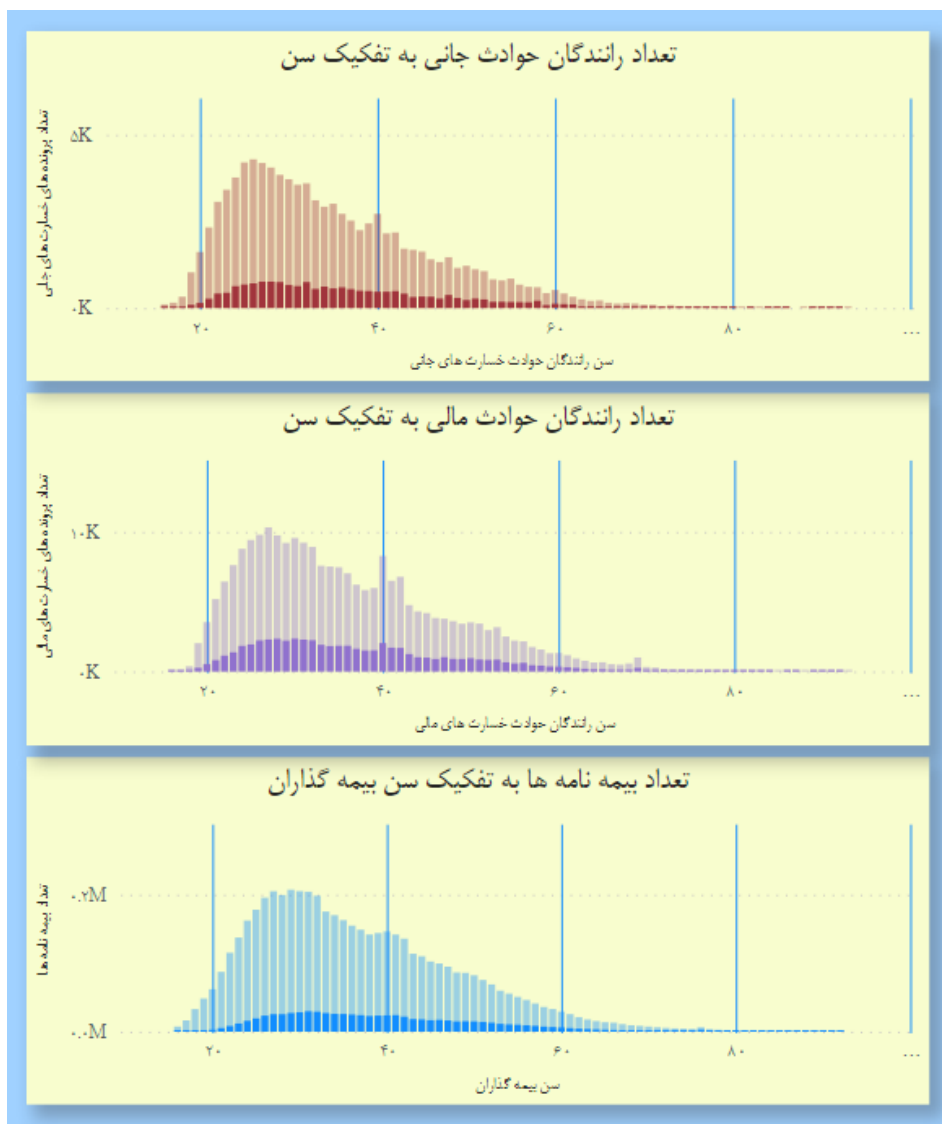
شکل ۴. درصد تعداد رانندگان منجر به حادثه از کل خسارت‌ها که خود بیمه‌گذار بوده‌اند به تفکیک خسارت‌های مالی و جانی نمایش داده شده است (براساس داده‌های تحقیق حاضر).

۲-۳ مقایسه ویژگی سن به تفکیک گروه‌های خودرو

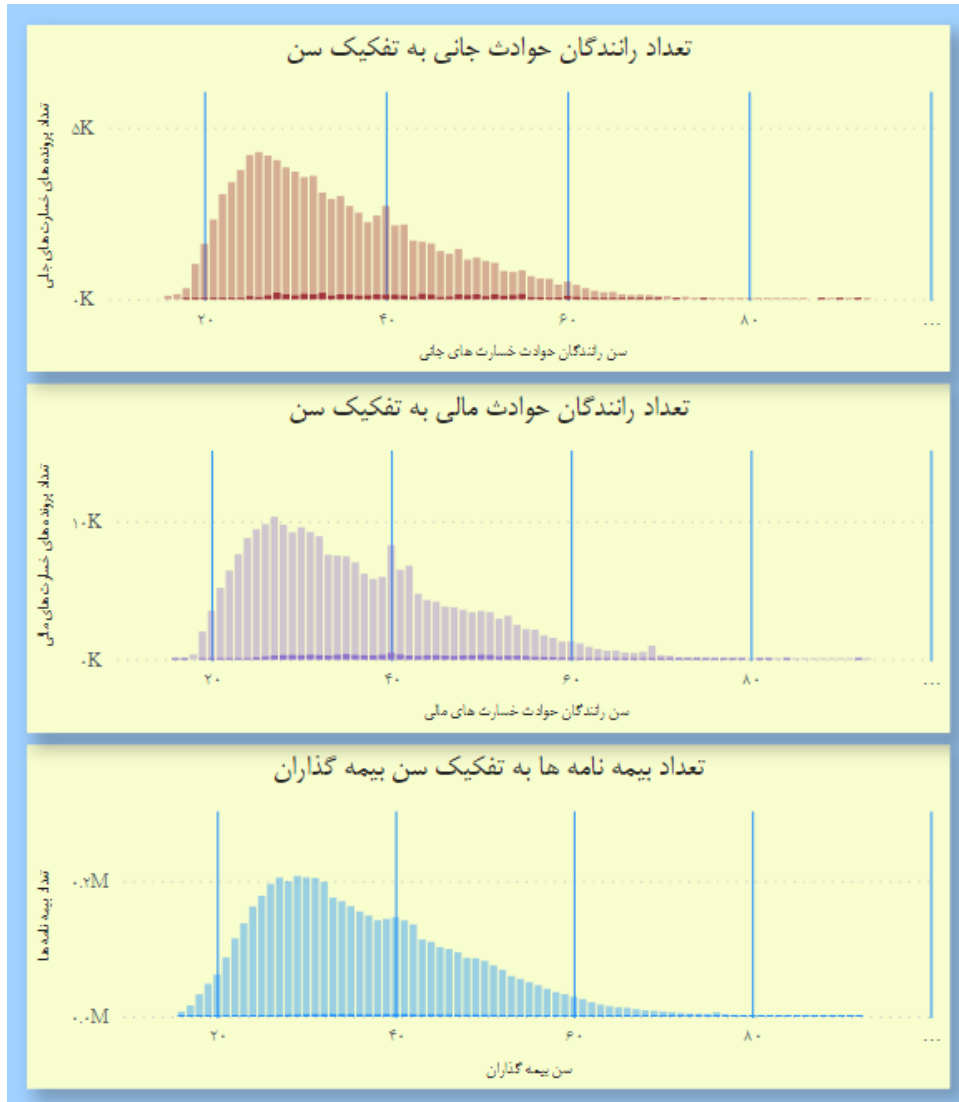
در شکل ۵، هیستوگرام تعداد بیمه‌نامه‌های صادره و تعداد پرونده‌های خسارت‌های جانی و خسارت‌های مالی یک شرکت بیمه طی شش سال، قابل مشاهده است. اگر تنها به هیستوگرام خسارت‌های جانی و یا هیستوگرام خسارت‌های مالی توجه کنیم، این ذهنیت ایجاد می‌شود که تعداد خسارت‌ها در سنین حول ۳۵ سال بیشتر است. آنچه که علاوه بر این آمار می‌باید لحاظ شود، جامعه آماری کل بیمه‌گذاران است. به عبارتی الگوی جمعیت سنی کشور در سال‌های ۱۳۸۷ الی ۱۳۹۲ به گونه‌ای بوده است که در رده‌های سنی حول ۳۵ سال جمعیت بیشتر بوده است؛ بنابراین، طبیعی است که نرخ خسارت‌های بیشتری نیز از این رده‌های سنی مشاهده شده باشد و این موضوع را نمی‌توان ناشی از پر خطر بودن این رده‌های سنی قلمداد کرد؛ بنابراین برای بررسی نحوه اثرگذاری ویژگی سن بر نرخ خسارت‌ها لازم است، الگوی جمعیتی جامعه موردنظر نیز لحاظ شود.

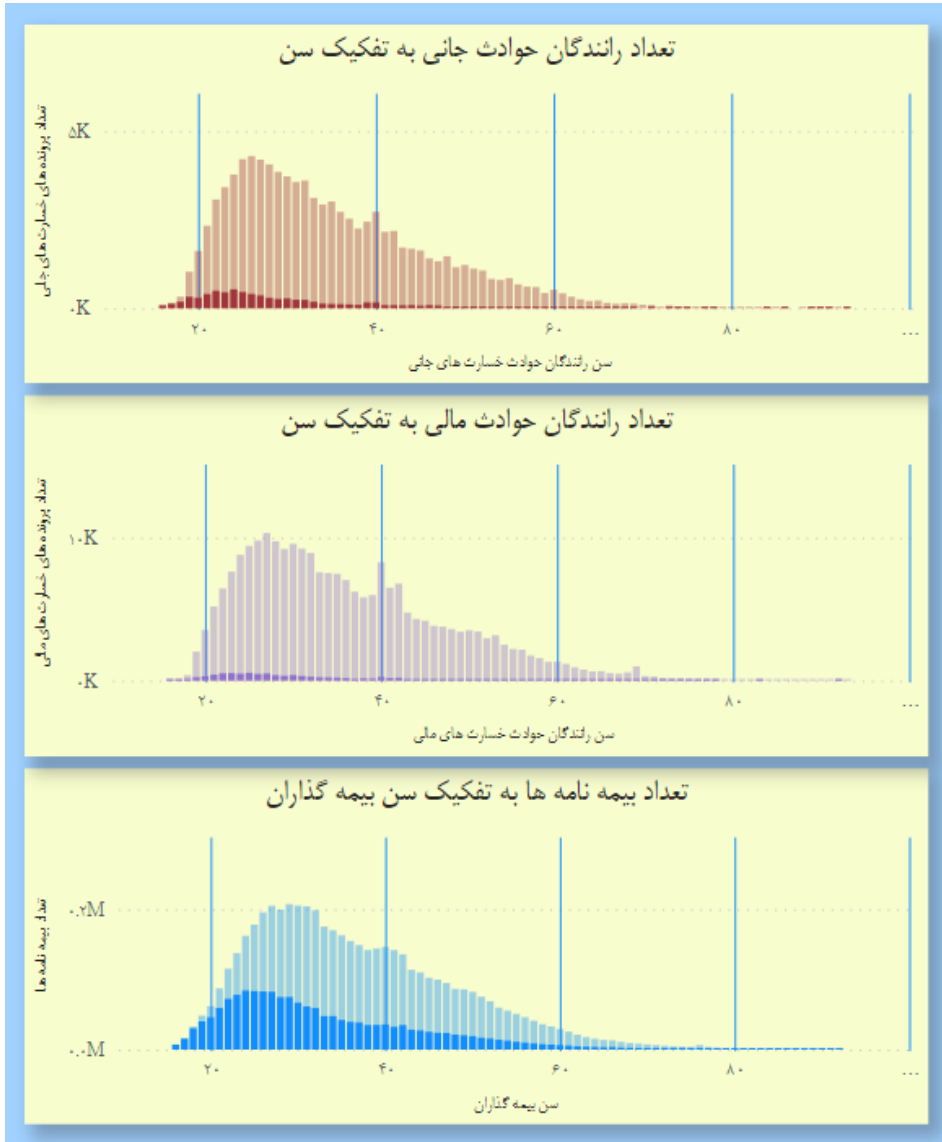


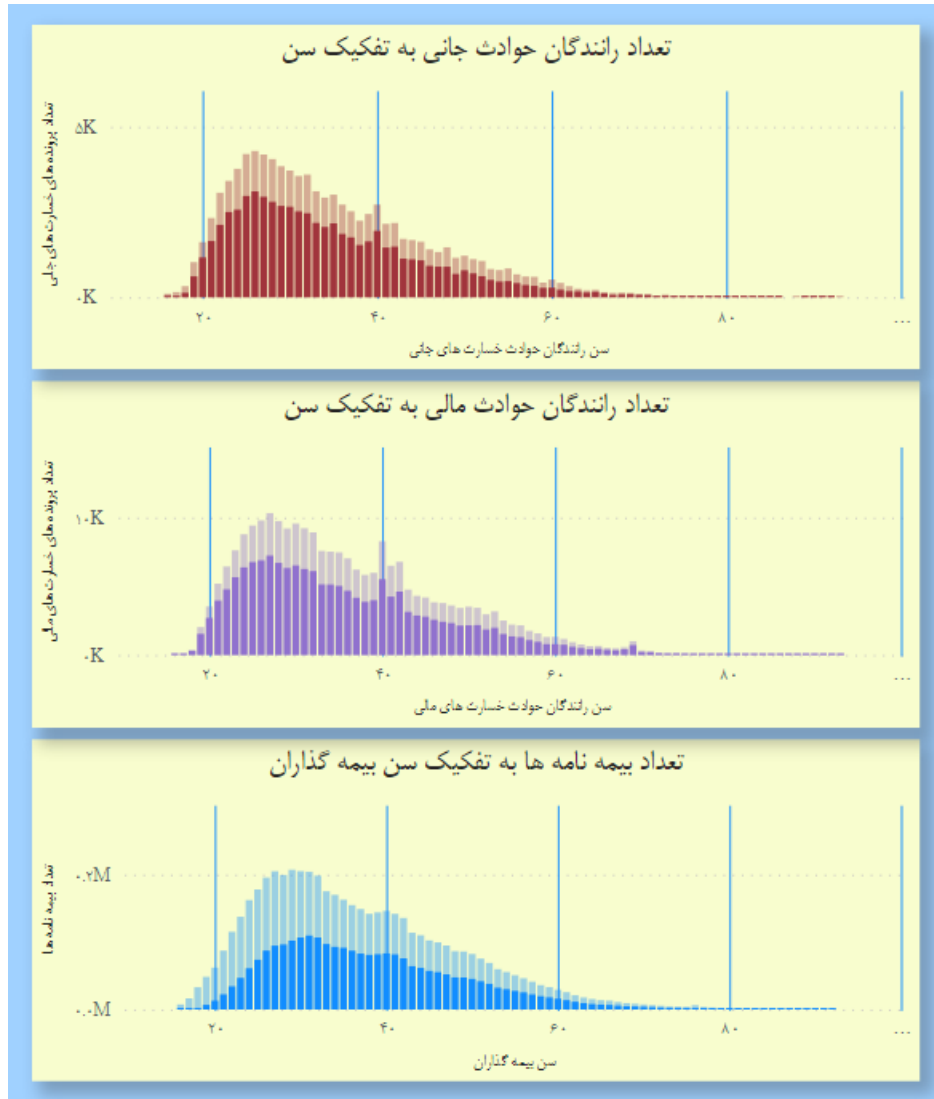
شکل ۵. هیستوگرام تعداد بیمه نامه، تعداد خسارت‌های مالی و خسارت‌های جانی به تفکیک سن.



شکل ۶. هیستوگرام تعداد بیمه‌نامه و تعداد خسارت‌های جانی و مالی رشته بیمه شخص ثالث به تفکیک انواع خودرو، به تفکیک از راست به چپ و از بالا به پایین متعلق به گروه‌های اتوکار، بارکش، سواری و موتور سیکلت (تعداد کل مشاهدات به صورت کمرنگ قابل مشاهده است).







در شکل ۵، جمعیت سنی کل بیمه‌گذاران و رانندگان حادثه‌ساز، به تفکیک حوادث منجر به خسارت‌های مالی و خسارت‌های جانی قابل مشاهده است. در شکل ۶، مشاهده می‌شود که جمعیت سنی بیمه‌گذاران در رده‌های اتوکار، بارکش، سواری و موتورسیکلت متفاوت است. در این شکل، هیستوگرام کل داده‌ها نیز پس‌زمینه نمودارها می‌باشد و می‌توان در هر گروه خودرو، داده‌های کل را با داده‌های گروه خودروی مربوطه مقایسه نمود. همان‌طور که در این شکل دیده می‌شود، سن

بیمه‌گذاران موتورسیکلت نسبت به سایر انواع خودرو کمتر است. همچنین سن بیمه‌گذاران اتوکار بیشتر از سن بیمه‌گذاران سایر انواع خودرو است.

۴- مقایسه توزیع آماری سن بیمه‌گذاران و سن رانندگان خسارت‌ها

بنا به آنچه که تاکنون بحث کردیم، شایسته است توزیع آماری سن را به تفکیک انواع گروه‌های خودرو مقایسه کنیم. در صورتی که سن بیمه‌گذاران و رانندگان حادثه‌ساز توزیع یکسانی داشته باشند، به آن معنی است که سنین متفاوت در ایجاد خسارت تفاوتی نداشته‌اند؛ زیرا توزیع سن جامعه بیمه‌گذاران و جامعه رانندگان حادثه‌ساز، یکسان بوده است. به این منظور در بخش بعد از آزمون کولموگروف-اسمیرنوف برای مقایسه توزیع سن بیمه‌گذاران و رانندگان حادثه‌ساز استفاده می‌کنیم.

۱-۴ نتایج آزمون کولموگروف اسمیرنوف دو نمونه‌ای

در آزمون کولموگروف-اسمیرنوف دو نمونه‌ای (پرت و همکاران^۱، ۱۹۸۱)، فرض صفر آزمون آن است که دو نمونه مفروض، توزیع یکسان دارند. با مقادیر کوچک p -value این فرض رد می‌شود. در ادامه با استفاده از زبان برنامه‌نویسی پایتون این آزمون را برای انواع گروه خودروها محاسبه کرده‌ایم و نتایج مطابق جدول زیر است.

جدول ۲. بررسی توزیع جمعیت سنی بیمه‌گذاران و رانندگان با استفاده از آزمون کولموگروف-اسمیرنوف.

نتیجه آزمون	p-value	مقدار آماره	گروه خودرو
فرض صفر رد می‌شود.	(e^{-7}) ۱.۹۱۵۷۹۳۳۲۲۶۸۲۵۷۲	۰.۰۴۰۶۶۵۶۸۳	اتوکار
فرض صفر رد می‌شود.	$(e^{-۱۶۷})$ ۲.۴۷۷۹۴۴۰۱۳۴۸۲۶۳۷۸	۰.۱۰۲۰۵۷۹۹۴	بارکش
فرض صفر رد می‌شود.	۰	۰.۱۴۹۵۲۱۰۴۵	سواری
فرض صفر رد می‌شود.	$(e^{-۱۵۶})$ ۴.۵۵۲۳۷۱۵۶۲۷۶۲۵۱	۰.۱۵۰۶۹۸۷۰۱	موتورسیکلت

براساس نتایج قابل مشاهده در جدول ۲، توزیع سن جامعه بیمه‌گذاران و رانندگان حادثه‌ساز یکسان نیست. به منظور تشخیص رده‌های سنی پر خطر، در ادامه احتمال وقوع خسارت، برای هر رده سنی به طور مجزا محاسبه شده است.

^۱. Pratt, John W; Gibbons, Jean D, 1981

۵- احتمال وقوع خسارت در سنین متفاوت

از آنجا که خسارت‌های اتومبیل که منجر به جراحات جانی می‌شوند، شدت بالاتری دارند و از نظر مالی و حتی اجتماعی بیشتر مورد توجه قرار دارند، محاسبات این بخش مبتنی بر داده‌های خسارت جانی است. احتمال وقوع خسارت جانی، برای سنین مختلف براساس قانون بیز به صورت زیر محاسبه می‌شود:

اگر A متغیر تصادفی سن رانندگان باشد و C مقدار متغیر تصادفی وقوع خسارت جانی باشد که در صورت وقوع خسارت مقدار ۱ و در صورت عدم وقوع خسارت مقدار ۰ را اخذ کند، در آن صورت احتمال وقوع خسارت برای سن i بر اساس قانون بیز برابر است با:

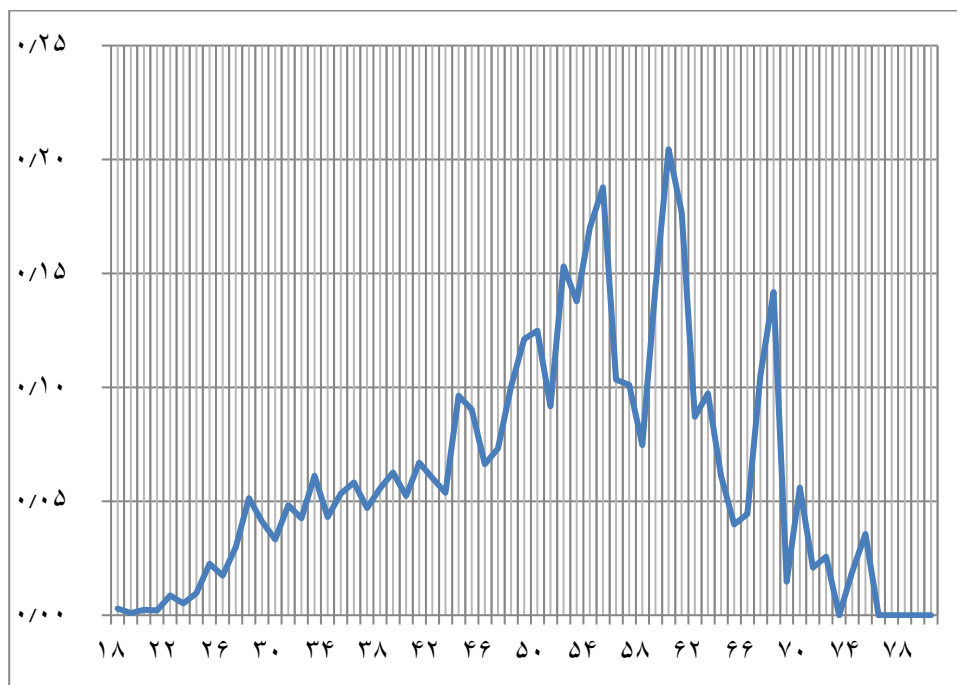
$$P(C = 1|A = i) = \frac{P(A = i|C = 1) * P(C = 1)}{P(A = i)}$$

برای برآورد مقدار $P(A = i|C = 1)$ ، از آمار مربوط به خسارت‌های جانی استفاده می‌شود و از تقسیم تعداد کل خسارت‌های مشاهده شده توسط رانندگان دارای سن i به تعداد کل خسارت‌ها برآورد می‌گردد. مقدار $P(A = i)$ با مقدار حاصل از تقسیم تعداد بیمه‌گذاران با سن i به تعداد کل بیمه‌گذاران تخمین زده می‌شود. در واقع این احتمال براساس رانندگان می‌باید، محاسبه شود؛ اما از آنجا که چنین داده‌هایی در دسترس نیست می‌توان فرض کرد که توزیع جمعیت سنی بیمه‌گذاران و رانندگان تقریباً یکسان است.

این محاسبات برای تمام رده‌های سنی و به تفکیک انواع گروه‌های خودرو محاسبه شده است. شکل‌های زیر مقادیر احتمال وقوع خسارت جانی را برای سنین مختلف در گروه‌های خودرو نشان می‌دهند.

۱-۵ احتمال وقوع خسارت برای خودروهای اتوکار

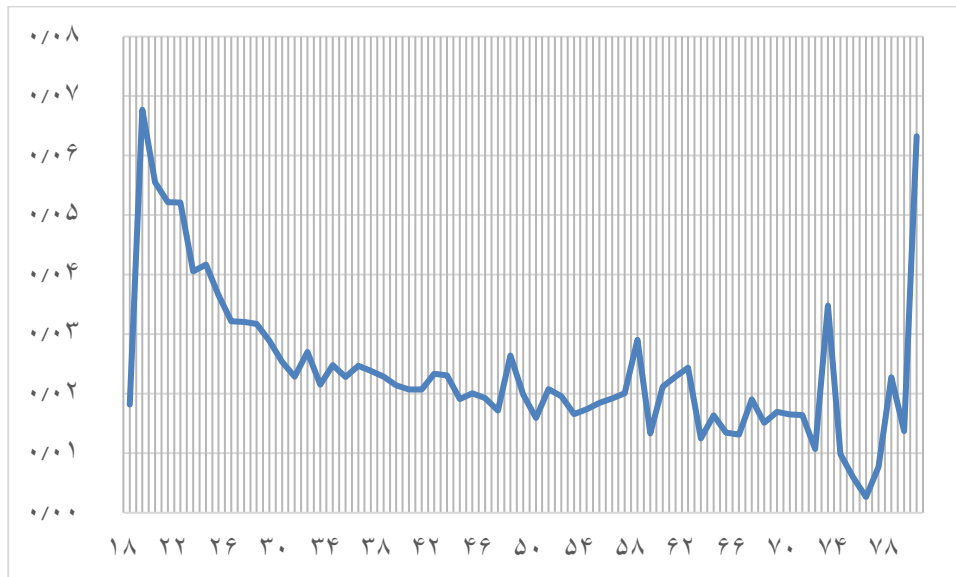
در نمودار شکل ۷، احتمال وقوع خسارت برای سنین مختلف در گروه خودروهای اتوکار قابل مشاهده است. در این نمودار مشاهده می‌شود که بیشترین احتمالات وقوع خسارت متعلق به رانندگان بازه ۴۵ تا ۶۵ سال است؛ بنابراین در این گروه خودرویی جوان‌ترها را نمی‌توان رانندگان پرخطرتر توصیف کرد. از طرفی تعداد داده‌های مورد بررسی در این گروه خودرویی بسیار اندک است و برای محاسبات دقیق‌تر نیاز است از جامعه آماری بزرگ‌تری استفاده شود.



شکل ۷. احتمال وقوع خسارت در سنین مختلف برای رانندگان خودروهای اتوکار.

۲-۵ احتمال وقوع خسارت برای بیمه‌گذاران خودروهای بارکش

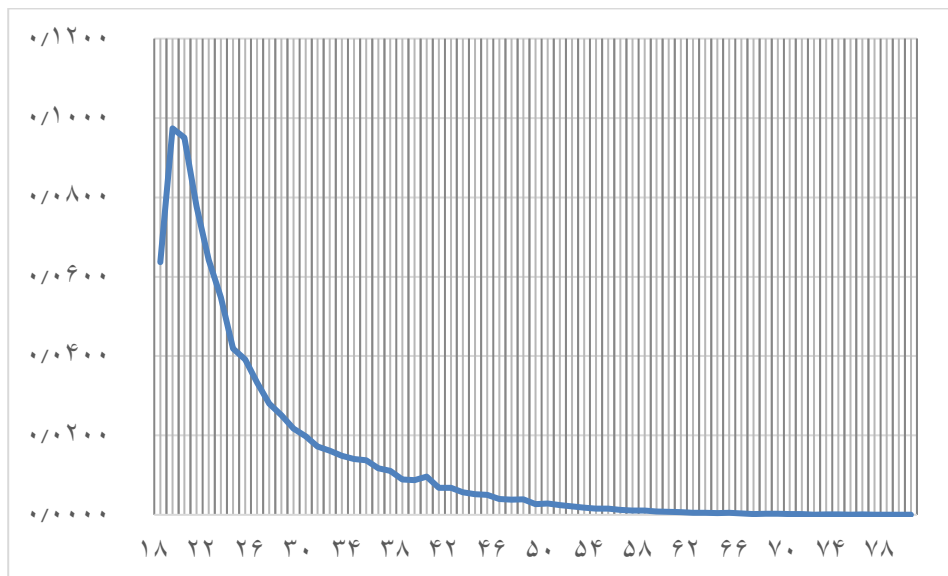
در نمودار شکل ۸، احتمال وقوع خسارت برای سنین مختلف در گروه خودروهای بارکش قابل مشاهده است. براساس این نمودار سنین کمتر از ۳۰ سال و بیش از ۷۰ سال نرخ خسارت‌های بیشتری دارند.



شکل ۸. احتمال وقوع خسارت در سنین مختلف برای رانندگان خودروهای بارکش.

۳-۵ احتمال وقوع خسارت برای خودروهای سواری

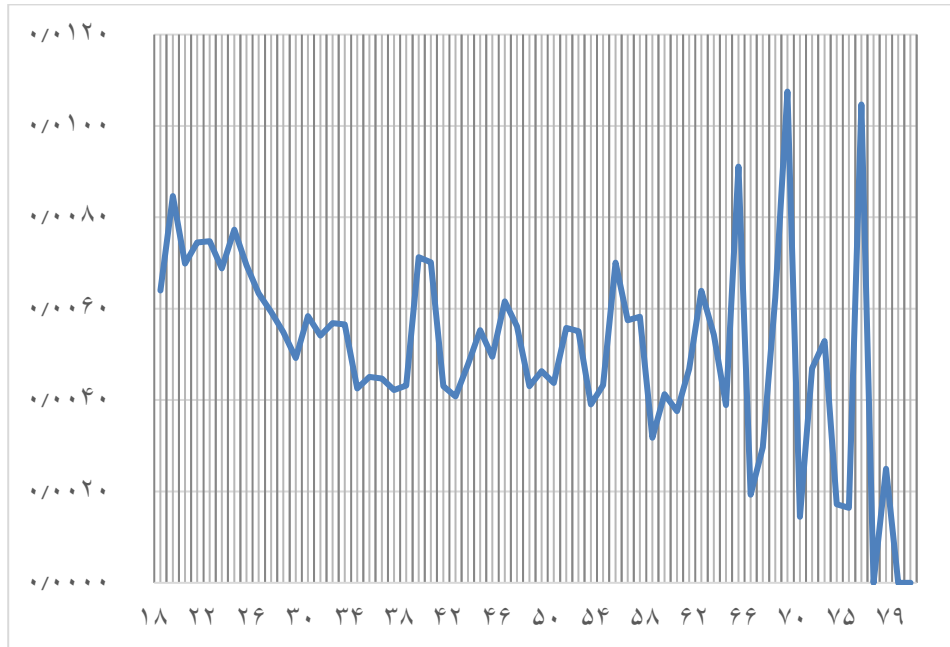
در نمودار شکل ۹، احتمال وقوع خسارت برای سنین مختلف در گروه خودروهای سواری قابل مشاهده است. براساس این نمودار سنین کمتر از ۲۲ سال نرخ خسارت‌های بیشتری دارند.



شکل ۹. احتمال وقوع خسارت در سنین مختلف برای رانندگان خودروهای سواری.

۴-۵ احتمال وقوع خسارت برای بیمه‌گذاران موتورسیکلت

در نمودار شکل ۹، احتمال وقوع خسارت برای سنین مختلف در گروه خودروهای موتورسیکلت قابل مشاهده است. براساس این نمودار سنین کمتر از ۲۴ سال و بالاتر از ۵۵ سال نرخ خسارت‌های بیشتری دارند.



شکل ۱۰. احتمال وقوع خسارت در سنین مختلف برای رانندگان خودروهای موتورسیکلت.

آنچه که از نمودارهای این بخش قابل مشاهده است آن که در گروه‌های خودرو اتوکار، بارکش، سواری و موتورسیکلت رده‌های سنی پر خطر متفاوت هستند. به‌منظور تشخیص بهتر گروه‌های پر خطر سنی بر وقوع یا عدم وقوع خسارت‌های جانی درخت تصمیم مربوطه را در بخش بعد محاسبه و تحلیل می‌کنیم.

۶- درخت تصمیم براساس ویژگی‌های ثبت شده از بیمه‌گذاران در پایگاه داده

درخت تصمیم یک یادگیری ماشین است که در زمره روش‌های یادگیری تحت نظارت قرار دارد. این روش برای طبقه‌بندی داده‌ها مورد استفاده قرار می‌گیرد. درختان تصمیم می‌توانند با استفاده از فضای ویژگی‌ها، مفاهیم پیچیده را توصیف کنند. به عبارت دیگر درختان تصمیم، می‌توانند تابعی را نمایش دهند که مقادیر ویژگی‌ها را به یک رده تصمیمات (برچسب داده‌ها) نگاشت می‌کند. درخت تصمیم، یک گراف جهت‌دار بدون دور است که رئوس آن گره و یال‌های آن شعبه نامیده می‌شوند.

رئوس بدون فرزند را برگ و تنها رأس بدون والد را ریشه می‌نامند (کزاک^۱، ۲۰۱۷). ریشه درخت شامل کل داده‌ها است. با استفاده از معیار جداسازی انتخاب شده و ویژگی‌های داده‌ها، داده‌های ریشه درخت به گره‌های جداگانه تقسیم می‌شوند. سپس مجدداً در هر گره، معیاری برای جداسازی داده‌ها انتخاب می‌شود و داده‌ها جداسازی می‌شوند. این رویه تا آنجا ادامه می‌یابد که ادامه آن، منجر به افزایش دقت پیش‌بینی‌ها نگردد. الگوریتم‌های درخت تصمیم متنوع هستند. تفاوت الگوریتم‌های متفاوت در تعیین معیار جداسازی داده‌ها و نحوه جداسازی است. در هر شاخه از درخت، یک ویژگی از داده‌ها براساس معیارهای درخت تصمیم برای جداسازی لحاظ می‌شود. این ویژگی بیش از سایر ویژگی‌ها، توان جداسازی داده‌ها را براساس برچسب داده‌ها دارد. محاسبات درخت تصمیم به شیوه‌های گوناگون انجام می‌شود که پرداختن به جزئیات آن‌ها، خارج از موضوع این مقاله است. شایان ذکر است که برای ساختن درخت تصمیم داده‌ها به دو بخش تقسیم می‌شوند. یک بخش به منظور استخراج قواعد درخت تصمیم مورد استفاده قرار می‌گیرد و بخش دیگر داده‌ها که در ساختن درخت تصمیم، نقشی نداشته است، برای آزمایش درخت استفاده می‌شوند و توانایی مدل ساخته شده را در پیش‌بینی برچسب داده‌ها مورد ارزیابی قرار می‌دهند. دقت درخت تصمیم در پیش‌بینی‌ها، با استفاده از ماتریس خطا انجام می‌شود.

در این مقاله، به منظور محاسبه درخت تصمیم، از سرویس تحلیل داده اس. کیو. ال^۲، استفاده می‌شود. از آنجا که در بخش‌های قبل این تحقیق، این نتیجه حاصل شد که نحوه اثرگذاری ویژگی سن برای گروه‌های مختلف خودرو متفاوت است، در این بخش، درخت تصمیم به تفکیک گروه‌های خودرو محاسبه شده است.

در محاسبه درخت تصمیم، تحقیق حاضر برچسب داده‌ها، وقوع و یا عدم وقوع خسارت جانی برای بیمه‌نامه‌ها می‌باشد. ویژگی‌های سن، جنسیت، نوع مشتری (حقیقی و یا حقوقی) از جمله ویژگی‌های بیمه‌گذاران هستند و در پایگاه داده شرکت بیمه مورد بررسی ثبت شده‌اند. این ویژگی‌ها برای محاسبه درخت تصمیم لحاظ خواهد شد. در این محاسبات ۳۰٪ از داده‌ها برای آزمایش و مابقی برای آموزش درخت استفاده شده‌اند. داده‌های آزمایش مدل، به صورت تصادفی از کل داده‌ها انتخاب شده‌اند. به منظور جداسازی داده‌ها، روش امتیازدهی بیزین دیریکله با توزیع پیشین یکنواخت^۳ (چارانی^۴، ۲۰۲۰) مورد استفاده قرار گرفته است. همچنین مقدار جریمه پیچیدگی را که از شباهت داده‌های جداسازی

^۱. Kozak, 2017

^۲. SSAS (Sql Server Analysis Service)

^۳. Bayesian Dirichlet Equivalent with Uniform prior

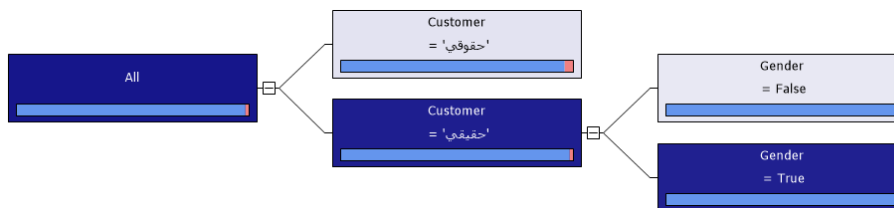
^۴. Charani, 2020

شده جلوگیری می‌کند، برابر با ۰.۹ لحاظ شده است؛ زیرا تعداد ویژگی‌ها برابر با ۳ است. برای پیاده‌سازی مدل درخت تصمیم بر داده‌هایی که شامل بیش از ۱۰ ویژگی می‌باشند، این مقدار را معمولاً ۰.۹۹ در نظر می‌گیرند.

درخت‌های تصمیم در هر چهار گروه خودرو، دقت مناسبی در پیش‌بینی‌ها ندارند. تعداد برجسب عدم وقوع خسارت در داده‌های لحاظ شده بسیار بالا است. این خاصیت مستتر در داده‌های بیمه، کاملاً مبتنی بر اصول بیمه‌گری می‌باشد. از طرفی دیگر، فقط سه ویژگی برای محاسبه درخت تصمیم در این مقاله لحاظ شده و بدیهی است که نمی‌توان براساس سه ویژگی سن، جنسیت و نوع مشتری وقوع یا عدم وقوع خسارت را با دقت بالا پیش‌بینی کرد. اما این درخت به منظور جداسازی گروه‌های سنی بیمه‌گذاران پُرخطر مفید است و هدف این مقاله نیز شناسایی رده‌های سنی پر خطر بیمه‌گذاران است.

۱-۶ درخت تصمیم خسارت‌های جانی خودروهای اتوکار

در شکل ۱۱، درخت تصمیم برای خودروهای اتوکار براساس سه ویژگی سن بیمه‌گذار، جنسیت و نوع مشتری رسم شده است. همان‌طور که در این شکل مشاهده می‌شود، اولین ویژگی لحاظ شده برای جداسازی داده‌ها نوع مشتری (حقوقی یا حقیقی) است. گروه مربوط به مشتریان حقیقی با معیار جنسیت به دو گروه تفکیک شده است. خلاصه قواعد درخت، در جدول ۲ آورده شده است. درصد وقوع خسارت‌های جانی برای کل بیمه‌گذاران خودروهای اتوکار، ۲۰.۷٪ است. این نسبت برای مشتریان حقوقی ۴.۱۹٪ و برای مشتریان حقیقی با جنسیت مؤنث ۱.۸۸٪ است. در قواعد این درخت در هیچ گروهی از معیار سن برای جداسازی داده‌ها استفاده نشده است و این موضوع نشانه آن است که ویژگی سن برای این نوع خودرو در تعیین حق بیمه‌ها حائز اهمیت نیست.



شکل ۱۱. درخت تصمیم با برجسب وقوع یا عدم وقوع خسارت جانی برای گروه خودروهای اتوکار با استفاده از سه ویژگی سن و جنسیت و نوع بیمه‌گذار (حقیقی یا حقوقی).

جدول ۲. خلاصه درخت تصمیم گروه اتوکار.

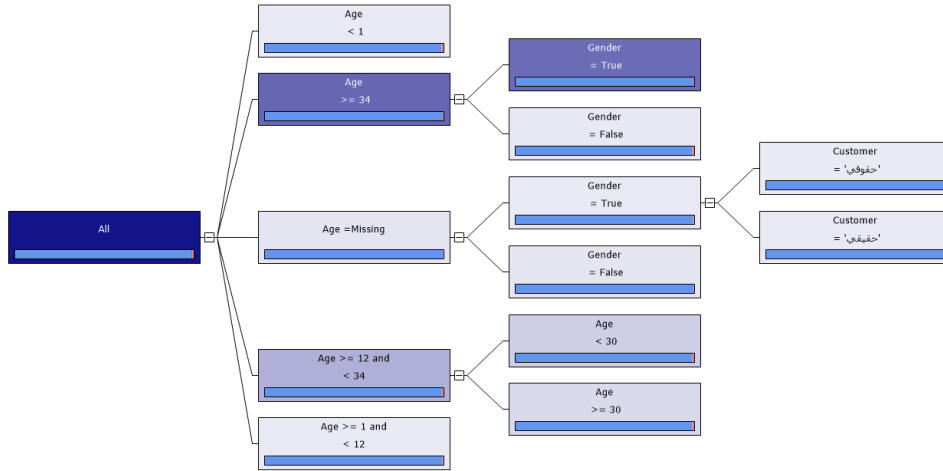
دسته های لایه اول	دسته های لایه دوم
نوع مشتری حقوقی	ندارد ۴.۱۹٪
نوع مشتری حقیقی	جنسیت مذکر ۱.۹۲٪
	جنسیت مؤنث ۱.۸۸٪

۲-۶ درخت تصمیم خسارت‌های جانی خودروهای بارکش

با کنار گذاشتن داده‌هایی که سن برای آن‌ها به‌درستی ثبت نشده است، اولین معیار جداسازی داده‌ها در درخت تصمیم شکل ۱۲ سن بیمه‌گذار است. برای خودروهای بارکش، داده‌ها به گروه‌های بیمه‌گذاران با کمتر از ۳۴ و بیشتر از ۳۴ سال در درخت تصمیم تفکیک شده‌اند و گروه مربوط به بیمه‌گذاران کمتر از ۳۴ سال نیز به دو گروه سنین کمتر از ۳۰ و بیشتر از ۳۰ سال تفکیک شده‌است. خلاصه درخت تصمیم در

بیمه‌گذاران دارای بیش از ۳۴ سال با معیار جنسیت تفکیک شده‌اند؛ جدول ۳، نمایش داده شده است بنابراین می‌توان در رده‌بندی سن، این بیمه‌گذاران را به گروه‌های بزرگ‌تر یا مساوی ۳۴ سال، بین ۳۰ تا ۳۴ سال و کمتر از ۳۰ سال تفکیک نمود. مطابق

جدول ۳، بیشترین نرخ خسارت این گروه خودرویی متعلق به بیمه‌گذاران کمتر از ۳۰ سال و برابر با ۱.۶۴٪ بوده است و کمترین نرخ خسارت متعلق به بیمه‌گذاران با سن بزرگ‌تر یا مساوی ۳۴ سال، جنسیت مذکر و نوع مشتری حقوقی و برابر با ۰.۶٪ بوده است. درصد وقوع خسارت جانی برای کل بیمه‌گذاران خودروهای بارکش ۱.۳۳٪ است.



شکل ۱۲. درخت تصمیم با برچسب وقوع یا عدم وقوع خسارت جانی برای گروه خودروهای بارکش با استفاده از سه ویژگی سن و جنسیت و نوع بیمه‌گذار (حقیقی یا حقوقی).

جدول ۳. خلاصه درخت تصمیم گروه بارکش.

دسته های لایه اول	لایه دوم	لایه سوم
سن بزرگ‌تر از ۳۴	جنسیت مذکر ۱.۲۸٪	نوع مشتری حقوقی ۰.۶٪
	جنسیت مؤنث ۱.۴۷٪	نوع مشتری حقیقی ۱.۰۲٪
سن کمتر از ۳۴	۱.۵۱٪ سن کمتر از ۳۰	
	۱.۳۴٪ سن بزرگتر از ۳۰	

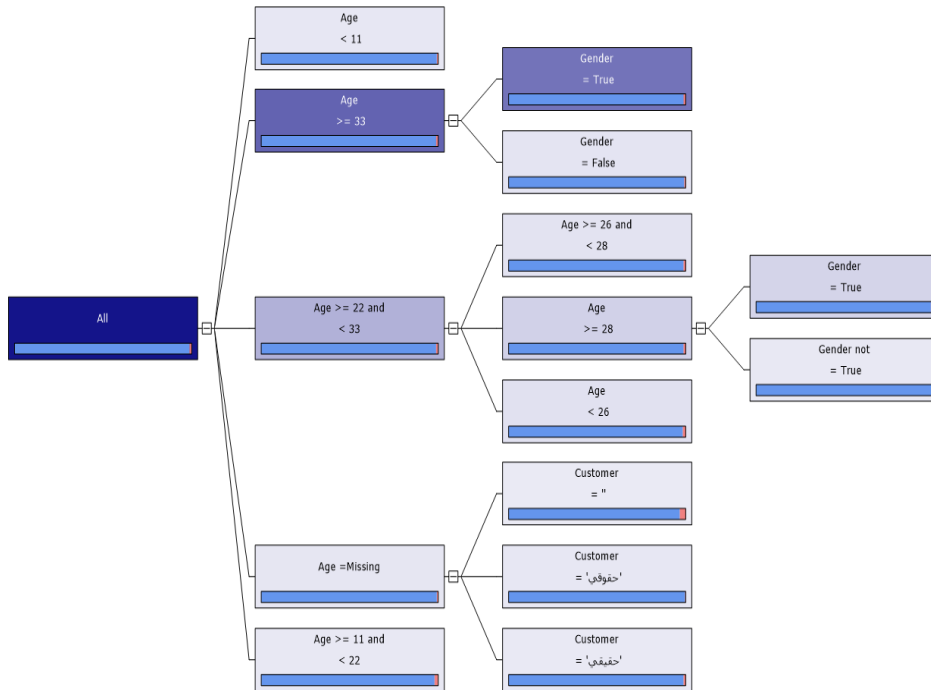
۳-۶ درخت تصمیم خسارت‌های جانی خودروهای سواری

درخت تصمیم برای خسارت‌های جانی براساس سه ویژگی سن، جنسیت و نوع مشتری (حقیقی یا حقوقی) در شکل ۱۳ رسم شده است. در درخت خودروهای سواری در اولین جداسازی، داده‌ها به سه مجموعه بیمه‌گذاران با سن کمتر از ۲۲ سال، بزرگ‌تر یا مساوی ۲۲ سال الی ۳۳ سال و بزرگ‌تر از ۳۳ سال تقسیم شده است. برای سنین بیش از ۳۳ سال درصد وقوع خسارت ۱.۲۹، در رده سنی ۲۲ الی ۳۳ سال درصد وقوع خسارت ۱.۳ و برای سن کمتر از ۲۲ سال احتمال خسارت ۲.۴ درصد بوده

است، که اختلاف قابل توجهی با سایر دسته‌ها دارد. در برخی از سطرهای داده نیز مقادیر پرتی برای سن ثبت شده است که ناشی از خطای موجود در داده‌ها است. معیار جداسازی بیمه‌گذاران دارای سن بیش از ۳۳ سال، جنسیت بوده است که برای آقایان نسبت خسارت در این دسته بیشتر است. برای خودروهایی که سن برای آن‌ها ثبت نشده، در لایه دوم نوع مشتری تعیین‌کننده بوده است. این اطلاعات به‌طور خلاصه در جدول ۵ مشاهده می‌شود. همچنین رده‌های سنی ۲۲ الی ۲۶، ۲۶ الی ۲۸ و ۲۸ الی ۳۳ به ترتیب نرخ خسارت‌ها ۱.۶۴٪، ۱.۴۶٪ و ۱.۹۱٪ درصد می‌باشد.

دسته‌های لایه اول	درصد احتمال خسارت	دسته‌های لایه دوم
سن کمتر از ۲۲ سال	۲.۴٪	ندارد
سن ۲۲ الی ۳۳ سال	۱.۳٪	سن کمتر از ۲۶ و بیشتر از ۲۲ ۱.۶۴٪
		سن بین ۲۶ الی ۲۸ ۱.۴۶٪
		بین ۲۸ الی ۳۳ ۱.۹۱٪
سن بیش از ۳۳ سال	۱.۲۹٪	جنسیت مذکر ۱.۳۰٪
		جنسیت مؤنث ۱.۱۸٪

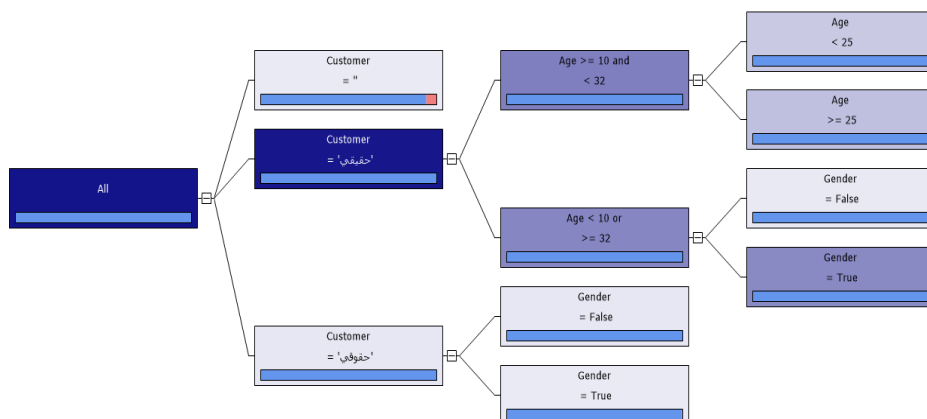
جدول ۴. خلاصه درخت تصمیم گروه سواری.



شکل ۱۳. درخت تصمیم با برچسب وقوع یا عدم وقوع خسارت جانی برای گروه خودروهای سواری با استفاده از سه ویژگی سن، جنسیت و نوع بیمه‌گذار (حقیقی یا حقوقی).

۴-۶ درخت تصمیم خسارت‌های جانی موتورسیکلت

براساس درخت تصمیم، می‌توان گفت سن بیمه‌گذاران موتورسیکلت نسبت به خودروهای سواری و بارکش تأثیر کمتری بر وقوع خسارت‌های جانی دارد. احتمال وقوع حادثه جانی، برای بیمه‌نامه‌های موتورسیکلت به‌طور کلی ۰.۳۲٪ بوده است. در لایه اول درخت تصمیم، داده‌ها براساس نوع مشتریان مجزا شده‌اند. نکته حائز اهمیت آن است که بیمه‌نامه‌هایی که نوع مشتریان برای آن‌ها ثبت نشده است، احتمال وقوع خسارت برای آن‌ها حدود ۶٪، بوده است که عدد قابل توجهی است. برای بیمه‌گذاران حقیقی با سن کمتر از ۲۵ سال نسبت خسارت ۰.۴۳٪ بوده و برای سن بیشتر از ۲۵ سال و کمتر از ۳۲ سال، ۰.۳۳٪ و برای بیمه‌گذاران بیش از ۳۲ سال با جنسیت مؤنث ۰.۴۵٪ بوده است. نکته قابل توجه دیگر درصد پایین نسبت خسارت برای خودروهای با نوع مشتری حقوقی می‌باشد.



شکل ۱۴. درخت تصمیم با برجسب وقوع یا عدم وقوع خسارت جانی برای گروه خودروهای اتوکار با استفاده از سه ویژگی سن، جنسیت و نوع بیمه‌گذار (حقوقی یا حقوقی).

جدول ۵ خلاصه درخت تصمیم موتورسیکلت.

دسته‌های لایه سوم	دسته‌های لایه دوم	درصد احتمال خسارت	دسته‌های لایه اول
ندارد	ندارد	۶.۲۵٪	نوع مشتری ندارد
۰.۴۳٪ سن کمتر از ۲۵	۰.۳۸٪ سن کمتر از ۳۲ سال	۰.۳۳٪	نوع مشتری حقیقی
۰.۳۳٪ سن بیشتر از ۲۵	۰.۲۷٪ سن بیشتر از ۳۲ سال		
۰.۴۵٪ جنسیت مؤنث	۰.۲۷٪ جنسیت مذکر	۰.۰۷٪	نوع مشتری حقوقی
۰.۲۷٪ جنسیت مذکر	۰.۰۵٪ جنسیت مؤنث		

۵-۶ دقت درخت‌های تصمیم خسارت‌های جانی در چهار گروه خودرو

در جدول ۶، ماتریس دقت پیش‌بینی درخت‌های تصمیم که در شکل‌های ۱۱، ۱۲، ۱۳ و ۱۴ نمایش داده شده‌اند، قابل مشاهده است. همان‌طور که مشاهده می‌شود، دقت درخت‌ها در پیش‌بینی وقوع یا عدم وقوع خسارت مناسب نیست. به عبارتی دیگر، با استفاده از ویژگی‌های سن بیمه‌گذار، جنسیت و

نوع مشتری امکان پیش‌بینی وقوع یا عدم وقوع خسارت با دقت مناسب امکان‌پذیر نیست؛ بنابراین نیاز به لحاظ نمودن ویژگی‌های بیشتری می‌باشد. هدف از محاسبه درخت تصمیم نیز در این پژوهش بررسی نحوه جدا سازی رده‌های سنی بیمه‌گذاران و نه پیش‌بینی وقوع و یا عدم وقوع خسارت بوده است.

جدول ۶. ماتریس دقت درخت‌های تصمیم اتوکار بارکش سواری و موتورسیکلت.

گروه خودرو	پیش‌بینی	وقوع خسارت (مقدار واقعی)	عدم خسارت (مقدار واقعی)
اتوکار	عدم خسارت	۷۱۷	۳۴۰۲۱
	وقوع خسارت	۰	۰
بارکش	عدم خسارت	۳۳۷۱	۲۴۸۶۲۳
	وقوع خسارت	۰	۰
سواری	عدم خسارت	۱۱۶۵۷	۸۳۰۱۴۷
	وقوع خسارت	۰	۰
موتورسیکلت	عدم خسارت	۱۹۰۱	۵۷۷۵۸۷
	وقوع خسارت	۰	۰

۷- نتیجه‌گیری

در رشته بیمه شخص ثالث وسایل نقلیه، مطابق قانون پنج‌ساله ششم توسعه و قانون بیمه اجباری خسارات وارد شده به شخص ثالث در اثر حوادث ناشی از وسایل نقلیه مقرر شده است که حق بیمه رشته بیمه شخص ثالث با لحاظ ویژگی‌های راننده نرخ‌گذاری شود. درحال حاضر، ویژگی‌های رانندگان اعم از: سن، جنسیت و سوابق رانندگی تأثیری بر نرخ حق بیمه ندارد. حال آن‌که به‌منظور محاسبه نرخ منصفانه حق بیمه که خود منجر به افزایش ضریب نفوذ بیمه در کشور می‌گردد، لازم است از این اطلاعات نیز در تعیین نرخ حق بیمه استفاده شود. در سایر کشورها نیز جهت تعیین نرخ حق بیمه، از چنین ویژگی‌هایی استفاده می‌گردد. این شیوه نرخ‌گذاری، منافع اجتماعی و اقتصادی بسیاری دارد؛ از جمله می‌تواند منجر به بهبود رفتارهای رانندگی و توسعه فرهنگ رانندگی ایمن شود. وجه اشتراک پژوهش‌های انجام شده در سایر کشورها این است که افراد جوان‌تر و افرادی که به تازگی گواهینامه گرفته‌اند، پرخطرتر هستند. همچنین در مجموع رانندگان خانم، نرخ خسارت‌های پایین‌تری دارند و به‌همین دلیل درحال حاضر در اغلب کشورها، نرخ بیمه‌نامه اتومبیل برای افراد جوان‌تر بیشتر است؛ از این‌رو، در این پژوهش با استفاده از داده‌های ثبت شده طی ۶ سال، در پایگاه

داده‌های یک شرکت بیمه، نحوه اثرگذاری ویژگی‌های بیمه‌گذاران بر نرخ خسارت‌ها مورد بررسی قرار گرفته است.

در این پژوهش، بررسی آماری داده‌های بیمه‌نامه‌ها و خسارت‌ها، نشان می‌دهد که در تمام چهار گروه خودرو بیش از نیمی از خسارت‌ها متعلق به رانندگانی است که بیمه‌گذار نیستند و این چالش برای گروه موتورسیکلت بیش از سایر گروه‌ها است و حدود ۶۰٪ از خسارت‌ها را شامل می‌شود؛ بنابراین با هدف تعیین حق بیمه‌های منصفانه، حق بیمه بیمه‌نامه‌هایی که رانندگان متعدد را تحت پوشش قرار می‌دهند، می‌تواند حدود ۱.۵ برابر حق بیمه بیمه‌نامه‌هایی که تنها یک راننده دارند، باشد. همچنین محاسبه آزمون کولموگروف-اسمیرنوف دو نمونه‌ای نشان داد که توزیع سن بیمه‌گذاران و رانندگان حادثه‌ساز متفاوت است و بنابراین سنین مختلف بیمه‌گذاران در ایجاد خسارت‌ها رفتاری متفاوت داشته‌اند.

سپس با استفاده از احتمال شرطی، احتمال وقوع خسارت جانی در سنین مختلف به تفکیک گروه‌های خودرو محاسبه شده است. برای این محاسبات به ناچار فرض کردیم توزیع سنی جمعیت بیمه‌گذاران و رانندگان مشابه باشد. نتایج نشان دادند که به‌ویژه برای خودروهای سواری و بارکش و سپس موتورسیکلت سنین پایین‌تر احتمال ایجاد خسارت جانی بالاتری دارند. این درحالی است که سن بیمه‌گذاران خودروهای اتوکار بیشتر است و به‌صورت کلی سن بیمه‌گذاران موتورسیکلت کمتر است.

سپس به‌منظور دسته بندی سنین مختلف برای هر یک از چهار گروه خودرو، درخت تصمیم را محاسبه کردیم؛ نتایج نشان می‌دهد برای خودرو اتوکار، ویژگی سن در ایجاد خسارت‌ها تأثیر ندارد. برای خودروهای بارکش می‌توان رده‌های سنی کمتر از ۳۰، بین ۳۰ و ۳۴ و بیشتر از ۳۴ سال را با درصد خسارت‌های ۱.۲۸٪، ۱.۳۴٪ و ۱.۶۴٪ تفکیک کرد. در خودروهای سواری سن تأثیر بیشتری دارد. برای این گروه خودرو می‌توان رده‌بندی کمتر از ۲۲ سال، بین ۲۲ تا ۲۶ سال، بین ۲۶ تا ۲۸ سال، بین ۲۸ تا ۳۳ سال و بزرگ‌تر از ۳۳ سال را با درصد وقوع خسارت ۲.۲۴، ۱.۶۴، ۱.۴۶، ۱.۹۱ و ۱.۲۹ درصد گرفت و در پایان برای خودروهای موتورسیکلت سن کمتر از ۲۵، سن بین ۲۵ تا ۳۲ و سنین بیشتر از ۳۲ را با درصد وقوع خسارت ۰.۴۳٪، ۰.۳۳٪ و ۰.۲۷٪ تفکیک نمود.

به‌طور خلاصه، نتیجه این مقاله نشان می‌دهد که برای گروه‌های مختلف خودرو، سن بیمه‌گذار تأثیری متفاوت بر نرخ خسارت‌ها دارد. ویژگی سن بیمه‌گذار، بر وقوع خسارت‌های جانی، خودروهای سواری تأثیرگذار است؛ و بر وقوع خسارت‌های جانی، خودروهای اتوکار این ویژگی چندان تأثیرگذار نیست. به‌منظور محاسبه نرخ‌های منصفانه در رشته بیمه شخص ثالث نیاز است که برای افراد جوان‌تر

در گروه‌های بارکش، سواری و موتورسیکلت نرخ‌های بیشتری تعیین شود و از طرفی برای سایر بیمه‌گذاران نرخ‌ها کاهش‌یابند و به این ترتیب نرخ حق بیمه‌ها به نرخ‌های منصفانه نزدیک گردند. افزایش نرخ برای بیمه‌گذاران جوان نامطلوب خواهد بود و بسیاری از کشورها، از جمله انگلیس برای حل این چالش، و راهکار به ارگرفته شده در سایر کشورها، از جمله انگلیس آن است که برای کاهش نرخ خسارت‌ها، جوان‌ترها به منظور کاهش نرخ حق بیمه بیمه‌نامه‌هایشان می‌توانند امکان نصب ابزارهای تلماتیک را برای بیمه‌گران فراهم کنند و در غیر این صورت ایشان مجاب به پرداخت حق بیمه بیشتر می‌شوند.

استفاده از ویژگی‌های محدودی مانند: سن، جنسیت و نوع مشتری برای پیش‌بینی نحوه رانندگی افراد و نرخ خسارت‌ها چندان مؤثر نیست. از طرفی دیگر، ممکن است یک فرد جوان خاص، به مراتب بهتر از یک فرد میانسال خاص رانندگی کند، در حالی که در نرخ‌گذاری براساس ویژگی‌های راننده این استثناءها مشهود و قابل اندازه‌گیری نیستند؛ بنابراین، این شیوه همچنان با تحقق هدف تعیین نرخ‌های منصفانه فاصله خواهد داشت. در صورتی که قانون‌گذار امکان استفاده از ابزارهای تلماتیک را فراهم آورد و نحوه رانندگی با تمام جزئیات ثبت شود، براساس داده‌هایی که توسط این ابزارها جمع‌آوری می‌شود، برای هر فردی می‌توان نرخ مبتنی بر نحوه رانندگی وی تعیین نمود.

منابع

- اصغری اسکویی، م. ر.؛ خانی‌زاده، ف.؛ و بهادر، آ.، (۱۳۹۹). «کاربرد داده کاوی با استفاده از الگوریتم‌های یادگیری ماشین برای بررسی تأثیر ویژگی‌های خودرو در پیش‌بینی ریسک خسارت مالی در رشته بیمه شخص ثالث». *پژوهشنامه بیمه*، شماره ۳۵، صص: ۳۳-۶۶.
- عریبان، ن.؛ خداپرست مشهدی، م.؛ همایونفر، م.؛ فیضی، م.؛ و فکور، و.، (۱۳۹۹). «آزمون انتخاب بد در بازار بیمه پایه درمان ایران». *فصل نامه علمی مطالعات اقتصادی کاربردی ایران*، شماره ۳۳، صص: ۹۳-۱۱۱.
- کریم‌زادگان مقدم، د.؛ و بهروان، م.، (۱۳۹۴). «ارائه راهکاری برای تعرفه‌گذاری پویا در صنعت بیمه با استفاده از تکنیک داده کاوی (مطالعه موردی: بیمه شخص ثالث)». *پژوهشنامه بیمه*، شماره ۳۰، صص: ۱-۲۰.
- محقق‌زاده، ف.؛ شیرین‌بخش، ش. ا.؛ نجفی‌زاده، ع.؛ و دقیقی‌اصلی، ع.، (۱۳۹۶). «تأثیرپذیری رشد اقتصادی از بیمه‌های زندگی و غیرزندگی». *فصلنامه مطالعات اقتصادی کاربردی ایران*، شماره ۲۳، صص: ۲۰۵-۲۳۰.
- Asghari Oskoei, M. R.; Khanizadeh, F. & Bahador, A., (2020). "Application of Data Mining through Machine Learning Algorithms to Study Effect Insurance". *Iranian Journal of Insurance Research*. No. 35(1), Pp: 33-66. doi: 10.22056/jir.2020.228093.2724.
- Arabian, N.; Khodaparast Mashhadi, M., Homayounifar, M.; Feizi, M. & Fakoor, V., (2020). "Adverse selection Test in Iran's Basic Health Insurance Market". *Journal of Applied Economics Studies in Iran*, 2020, No. 9(33), Pp: 93-111. doi: 10.22084/aes.2019.17763.2768.
- Bedsworth, F.; Portillo, J. E.; Neal, D. R. & Willardsen, K., (2021). "Asymmetric information and insurance: An experimental approach". *Journal of Behavioral and Experimental Econom.*
- Karim Zadehgan Moghaddam, D., Behrvan, M., (2016). "Dynamic Tariffing in Insurance Industry With Data Mining Method (Case Study: Third Person Car Insurance)". *Iranian Journal of Insurance Research*, 2016, No. 30(120), Pp: 1-20. doi: 10.22056/jir.2016.17618.
- Mohagheghzadeh, F.; Shirinbakhsh, S.; Najafizadeh, A. & Daghighi asl, A., (2017). "Sensitivity of economic growth to life and non-life insurances". *Journal of Applied Economics Studies in Iran*, 2017, No. 6(23), Pp: 205-230. doi: 10.22084/aes.2017.13128.2415.
- Shams Esfandabadi, Z.; Ranjbari, M. & Domenico Scagnelli, S., (2020). *Prioritizing Risk Level Factors in Comprehensive Automobile Insurance Management: A Hybrid Multi-Criteria Decision-Making Model*. Global Business Review.
- Śliwiński, A. & Kuryłowicz, Ł., (2021). *Usage-based insurance and its acceptance: An empirical approach*. Risk management and Insurance Review.
- Anon., (2018). *Young car driver's road safety factsheet*, s. 1.: Department for transportation.
- Bedsworth, F.; Portillo, J. E.; Neal, D. R. & Willardsen, K., (2021). "Asymmetric information and insurance: An experimental approach". *Journal of Behavioral and Experimental Economics*, Vol. 92, Article 101696.

- Charani, N. S., (2020). "Bayesian Smoothing of Decision Tree Soft Predictions and Evidential Evaluation". *Cham*, Springer, Pp: 368-381.
- Denuit, M.; Hainaut, D. & Trufin, J., (2019). "Insurance Risk Classification". *Effective Statistical Learning Methods for Actuaries*, Pp: 3-26.
- Dussault, C. & Letendre, P., (1999). "Automobile Insurance: Road Safety, New Drivers, Risks". *Insurance Fraud and Regulation*. s.l.: Springer.
- Esfandabadi, Z. S. & Esfahani, M. M. S., (2018). "Identifying and classifying the factors affecting risk in automobile hull insurance in Iran using fuzzy Delphi method and factor analysis". *Journal of Industrial Engineering and Management Studies*, Vol. 5 (2). Pp. 84-96.
- Kozak, J., (2017). *Decision Tree and ensemble learning based on Ant colony optimization*. s.l.: Springer.
- Maag, U. et al., (1999). *An evaluation of the effects on crashes of the 1991 legislative reform on new licensees in quebec*. New York: Springer Science.
- Parodi, P., (2014). *Pricing in general insurance*. s.l.: CRC press.
- Pratt, John, W. & Gibbons, Jean D., (1981). "Kolmogorov-Smirnov Two-Sample Tests". In: *Concepts of Nonparametric Theory*. New York: Springer, Pp: 318-344.
- Thomas, G. R., (2009). "Demand elasticity, Risk classification and loss coverage: when can community rating work". *Astin bulletin*, Pp: 403-428.
- Williams, A. F., (1999). *Licensing policies for young drivers in the united states*. New York: Springer.

Classifying Age of Policyholders According to the Claim Rates in IranManteqipour, M.^{1*}, Ghorbani, V.², Aalaei, M.³**Abstract**

According to Note 1, Article 18 of Third Party Liability (TPL) Insurance Law 1395 (2016), Central Insurance of I. R. Iran (CII) in cooperation with Law Enforcement Force of I. R. Iran is obliged to work out to allow for the issuance of TPL policy based on the characteristics of the driver by the end of Sixth Five-Year Development Plan. As a result, in the premium calculation and the third-party claims, the characteristics of the driver must be taken into account. Thus, using six-year data from an insurance company, the effect of the policyholders' age on the incurred damages are investigated. In this study, initially, with the help of Kolmogorov-Smirnov two-sample test, the age distribution of the reckless policyholders and drivers were compared. The results indicated that 'age' characteristic plays a role in causing damages. Due to the age differences among policyholders and the four major vehicle types, the calculations were distinctively performed based on each set of vehicles. Using conditional probability, the possibility of damages caused by different age groups was measured. Furthermore, the decision tree model was developed based on the three registered characteristics of the policyholders including age, gender, and type of customer. The results indicated that among the passenger cars, trucks, and motorcycles, the probability of damages caused by policyholders having less than 22, 30, and 25 years of age is respectively considerably high. Thus, according to actuarial principles, these policyholders must reasonably pay more premiums. The results further showed that more than 50% of the claims that involved fatality in all types of vehicles were related to drivers that had no insurance policy. Therefore, several drivers that share the same vehicle must pay more premiums. The age characteristic of the policyholders does not have any significant effect on claims involving buses or vans.

Keyword: Third Party Automobile Insurance, Conditional Probability, Decision Tree, Accuracy Classification Matrix.

JEL Classification: C11, C53, C55.

1. Data Mining Desk leader, Electrical Insurance Research Group, Insurance Research Center, Iran

Email: manteqipour@irc.ac.ir

2. Data Mining Consultant, Electrical Insurance Research Group, Insurance Research Center, Iran

Email: vahidghorbani@hotmail.com

3. Associate Professor, Faculty of Personal Insurance Research Group, Insurance Research Center, Iran

Email: aalaei@irc.ac.ir