

مدلی جهت دسته‌بندی ریسکی گروه‌های مشتریان بیمه بدنه اتومبیل

بر اساس ریسک با استفاده از تکنیک داده‌کاوی

(مورد مطالعه: بیمه بدنه اتومبیل در یک شرکت بیمه‌ای)

پیام حنفی‌زاده^۱

تاریخ دریافت مقاله: ۱۳۹۰/۰۴/۲۳

ندا رستخیز پایدار^۲

تاریخ پذیرش مقاله: ۱۳۹۰/۱۰/۱۲

چکیده

در کشور ما بیمه اتومبیل از مهم‌ترین رشته‌های بیمه‌ای است که سهم عمده‌ای در پرتفوی صنعت بیمه دارد. در ایران نرخ حق‌بیمه بدنه اتومبیل با توجه به تعرفه بیمه مرکزی ج.ا.ا. تعیین می‌شود. هدف این تحقیق ارائه راهکار به شرکت‌های بیمه جهت تعیین نرخ حق‌بیمه بدنه اتومبیل با توجه به سطح ریسک هریک از مشتریان و یاری‌کردن سازمان‌های بیمه در جهت پیشبرد اهداف و به‌کارگیری استراتژی‌های مناسب در خصوص هریک از دسته‌های مشتریان و بهبود موقعیت فعلی در بازار است. در مدل ارائه‌شده در ابتدا عوامل تأثیرگذار بر ریسک بیمه‌گذاران طی دو مرحله شناسایی شد. در مرحله اول هیجده فاکتور ریسک در چهار گروه شامل مشخصات جمعیت‌شناختی، مشخصات اتومبیل، مشخصات بیمه‌نامه و سابقه راننده از بین مقالات علمی منتشرشده در مجلات معتبر در بازه سال‌های ۲۰۰۹-۲۰۰۰ استخراج شد و در مرحله دوم با استفاده از نظرسنجی از خبرگان، فاکتورهای نهایی تعیین گردید. پس از بخش‌بندی مشتریان با استفاده از شبکه خودسازمانده، ویژگی‌های مشتریان در هریک از بخش‌ها شناسایی شد. شناسایی ویژگی‌های مشتریان، اولین گام برای تعریف حق‌بیمه‌های متفاوت بر اساس سطح ریسک هر مشتری است.

واژگان کلیدی: بخش‌بندی مشتریان، شبکه خودسازمانده، بیمه بدنه اتومبیل

۱. استادیار دانشگاه علامه طباطبائی (نویسنده مسئول)

(Email: Hanafizadeh@Gmail.com)

۲. کارشناس ارشد مدیریت فناوری اطلاعات، دانشگاه علامه طباطبائی

(Email: Paydarneda@Gmail.com)

۱. مقدمه

امروزه صنعت بیمه از ضروری‌ترین بخش‌های اقتصادی هر کشوری است که وظایف و کارکردهای مهمی برعهده‌دارد. رشد صنعت بیمه در هر کشوری بیانگر توسعه‌یافتگی و افزایش پس‌اندازهای مالی است. در ایران نیز بیمه از مهم‌ترین عوامل حفظ و تضمین سرمایه است (جعفری صمیمی و مرادی، ۱۳۸۷). مشتریان تفاوت‌های اساسی و زیادی با یکدیگر دارند. به‌طورکلی مشتریان از دو دیدگاه با یکدیگر متمایزند؛ از لحاظ ارزش آنها نزد سازمان و از لحاظ تفاوت در نیازهایشان (حسین‌زاده، ۱۳۸۶). شناسایی مشتریان مستلزم تحلیل مشتریان هدف و دسته‌بندی‌کردن مشتریان است که منجر به یافتن گروه‌هایی از مشتریان سودآور براساس ویژگی‌های آنها می‌شود. به‌طورکلی بخش‌بندی مشتریان شامل تقسیم‌کردن کلیه مشتریان به گروه‌های اصلی که شباهت بیشتری به یکدیگر دارند و قراردادن آنها در یک گروه است (Ngai et al, 2009). طبقه‌بندی ریسک در حقیقت به معنای گروه‌بندی مشتریان با خصوصیات ریسک مشابه است که احتمال بروز خسارت‌های مشابهی دارند. طبقه‌بندی ریسک بیمه‌گذاران بر مبنای ویژگی‌های قابل مشاهده می‌تواند به شرکت‌های بیمه جهت کاهش زیان، افزایش نرخ پوشش بیمه و جلوگیری از وقوع انتخاب نامساعد در بازار بیمه کمک شایانی کند (ماجد، ۱۳۸۷). نرخ حق‌بیمه در بسیاری از شرکت‌های بیمه‌ای در خارج از کشور باتوجه به متغیرهای گوناگون جمعیت‌شناختی، مشخصات اتومبیل و سابقه خسارت بیمه‌گذار محاسبه می‌شود. کرایدلر^۱، عضو هیئت‌مدیره یکی از شرکت‌های بیمه‌ای در ایالت واشنگتن در گزارشی نرخ حق‌بیمه را متکی بر عواملی همچون سن بیمه‌گذار، جنسیت و وضعیت تأهل بیمه‌گذار، نوع خودرو، محل زندگی، سابقه خسارت، میزان کیلومتر مصرفی و ... اعلام کرد. این در حالی است که در کشور ما نرخ حق‌بیمه بدنه

اتومبیل باتوجه به تعرفه بیمه مرکزی ج.ا.ا تعیین می‌شود. این امر سبب می‌شود که مشتریانی با ریسک پایین‌تر، خسارت‌های مالی مشتریانی با ریسک بالاتر را جبران کنند و لذا تفاوت آنچنانی بین مشتریان با ریسک بالا و پایین وجود نخواهد داشت.

۲. تعریف مسئله

فقدان سنجش‌های تعیین ریسک افراد در بیمه اتومبیل علاوه بر ناکارآمدی قراردادهای بیمه، منتج به تعیین نرخ‌های غیرعادلانه نیز می‌شود، چرا که به جای فرد، اتومبیل بیمه می‌شود و این امر موجب شده تا اکثر شرکت‌های بیمه در زمینه بیمه اتومبیل متحمل زیان شوند. در صورتی که اکثر کشورهای توسعه‌یافته با بهره‌مندی از سیستم طبقه‌بندی ریسک افراد سعی در افزایش بهره‌وری و سوددهی صنعت بیمه خود داشته‌اند.

در فرآیند مدیریت ارتباط با مشتری^۱ به‌طور کلی چهار بخش اصلی دیده شده است که اولین و مهم‌ترین گام آن شناخت مشتری است (Ngai et al, 2009). اولین گام جهت تعیین نرخ حق‌بیمه‌های متفاوت به ازای هر مشتری باتوجه به سطح ریسک آنها، شناخت ویژگی‌های هر یک از مشتریان است.

در این تحقیق فاکتورهای مهم و تأثیرگذار بر ریسک مشتریان بیمه اتومبیل در دو فاز جداگانه شناسایی و انتخاب شد. جهت انتخاب فاکتورها در اولین گام، مطالعه و بررسی بر روی مقالات علمی منتشرشده در بازه زمانی سال‌های ۲۰۰۹ - ۲۰۰۰ میلادی در بین مجلات معتبر^۲ و گزارش‌های علمی صورت گرفت. فاکتورهای استخراج‌شده به‌طور کلی در چهار دسته اصلی شامل مشخصات جمعیت‌شناختی، مشخصات اتومبیل، مشخصات بیمه‌نامه و سابقه راننده طبقه‌بندی گردیدند. در فاز دوم پرسش‌نامه‌ای باتوجه به عوامل شناسایی‌شده در فاز اول طراحی و توسط خبرگانی تکمیل شد که از بین کارشناسان بخش خسارت مالی، رؤسا و مدیران بیمه بدنه اتومبیل در ده شرکت بیمه‌ای انتخاب شده بودند.

در نهایت پس از تجزیه و تحلیل پرسش‌نامه‌ها با استفاده از روش آزمون سنگریزه^۱ فاکتورهای نهایی ریسک انتخاب گردید. در گام بعدی پس از شناسایی فاکتورهای نهایی ریسک، مشتریان بیمهٔ بدنهٔ اتومبیل با استفاده از روش شبکه عصبی خودسازمانده^۲ براساس فاکتورهای تعیین‌شدهٔ بخش‌بندی شدند.

۳. مروری بر ادبیات موضوع

۳-۱. داده‌کاوی و بخش‌بندی مشتریان

به دلیل رشد سریع تکنولوژی اطلاعات، حجم داده‌های ذخیره‌شده در پایگاه داده‌ها به سرعت در حال افزایش است. تجزیه و تحلیل داده‌های ذخیره‌شده و تبدیل کردن آنها به اطلاعات و دانش قابل استفاده برای سازمان به ابزارهای بسیار قدرتمند نیاز دارد. داده‌کاوی، ابزار ارزشمندی است که در سال‌های اخیر از آن به‌طور گسترده جهت استخراج اطلاعات، جستجو روابط و الگوها در بین حجم عظیم داده‌های استفاده شده است. براساس تحقیقات، تکنیک‌های داده‌کاوی مختلفی جهت گروه‌بندی مشتریان به‌کاررفته است.

در چرخه مدیریت ارتباط با مشتری، می‌توان بخش‌بندی مشتریان و تجزیه و تحلیل مشتریان را عناصر اصلی در گام شناسایی مشتری نام برد. تکنیک‌های داده‌کاوی به‌کاررفته در زمینه بخش‌بندی مشتریان را می‌توان به دو دستهٔ اصلی شامل دسته‌بندی مشتریان و خوشه‌بندی مشتریان تقسیم کرد. براساس تحقیق جای و همکارانش^۳ مشخص شد که از بین سی و چهار تکنیک داده‌کاوی، شبکه‌های عصبی بیشترین کاربرد را داشته است. شبکه‌های عصبی که با استفاده از مغز انسان شبیه‌سازی گردیده،

-
1. Scree Test
 2. Self – Organizing Map
 3. Ngai et al, 2009

کاربردهای زیادی در زمینه دسته‌بندی، خوشه‌بندی و پیش‌بینی داشته است جدول ۱ مطالعات صورت گرفته در زمینه شناسایی مشتریان را نشان می‌دهد.

۲-۳. مروری بر شبکه‌های عصبی خودسازمانده

شبکه‌های عصبی خودسازمانده، شبکه‌های عصبی بدون نظارتی هستند که قابلیت ارائه خروجی شبکه در قالب نقشه‌های گرافیکی گویا و قابل فهم برای مدیران سازمان‌ها را دارند. عدم حساسیت شبکه خودسازمانده به تعداد دادگان تعلیم و حساسیت کم این نوع شبکه‌ها به وجود نویز در دادگان تعلیم، توانایی نمایش روابط خطی و غیرخطی بین متغیرها و قدرت بالا در دسته‌بندی دادگان از دیگر مزیت‌های این شبکه‌هاست که در این تحقیق جهت بخش‌بندی مشتریان در بیمه بدنه اتومبیل با توجه به عوامل شناسایی گردیده در ریسک به عنوان یکی از پرکاربردترین تکنیک‌های بخش‌بندی در داده‌کاوی استفاده شد. از آنجایی که این شبکه قادر است خروجی شبکه را در قالب نقشه‌های گرافیکی نمایش دهد، سرعت درک و تفسیر نتایج برای مدیران و کارشناسان راحت‌تر خواهد بود.

جدول ۱. تکنیک‌های مختلف داده‌کاوی به کاررفته در شناسایی مشتریان

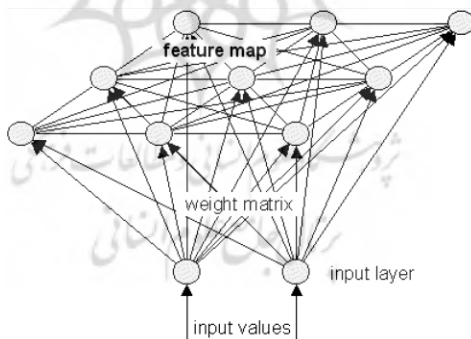
محققان	تکنیک داده‌کاوی	نوع عملیات داده‌کاوی	هدف مدیریت ارتباط با مشتری	ابعاد مدیریت ارتباط با مشتری
Kim et al, 2006	درخت تصمیم‌گیری	دسته‌بندی	بخش‌بندی	شناسایی مشتری
Ha et al, 2002	شبکه خودسازمان‌دهنده، درخت تصمیم‌گیری و مدل زنجیره مارکو			
Dennis et al, 2001	K-Means	خوشه‌بندی		
Lee & Park, 2005	شبکه خودسازمان‌دهنده و درخت تصمیم‌گیری، تحلیل پوششی داده			
Hwang et al, 2004	رگرسیون منطقی	رگرسیون	تحلیل مشتریان هدف	
Chen et al. 2003; Wu et al, 2005; Yu et al, 2005	درخت تصمیم‌گیری	دسته‌بندی		
Lee et al, 2004	شبکه خودسازمان‌دهنده	خوشه‌بندی		
Woo et al. 2005	نقشه مشتری	مصورسازی		

(Ngai et al, 2009)

اساس فلسفه شبکه‌های خودسازمانده، نگاشت فضاها با تعداد ابعاد بالا به فضایی دو یا سه بعدی است، به‌گونه‌ای که حداقل اطلاعات از بین رفته و اطلاعات نهفته در ارتباط میان داده‌ها نیز قابل کشف و نمایش باشد. این روش توانایی نمایش همبستگی میان داده‌ها و اطلاعات و اثرات متقابل و هم‌زمان آنها بر یکدیگر را دارد (Vesanto, 1997).

شبکه‌های خودسازمانده از دو لایه مجزا تشکیل می‌شوند. یک لایه ورودی و لایه دیگر، لایه نقشه نام دارد. لایه نقشه مجموعه‌ای از نرون‌هاست که با قاعده در قالب یک شبکه دو بعدی چیده شده و به لایه ورودی کاملاً متصل است. هر نرون در لایه نقشه مربوط به یک بردار اطلاعات با ابعادی برابر ابعاد فضای مورد تحلیل است. به بیان دیگر هر نرون نماینده بخشی از فضای اطلاعات است. شکل ۱ توپولوژی شبکه‌های خودسازمانده را نشان می‌دهد.

شکل ۱. توپولوژی شبکه خودسازمانده



۱-۲-۳. آموزش شبکه‌های خودسازمانده

آموزش شبکه‌های خودسازمانده بر مبنای الگوریتم یادگیری رقابتی و بدون ناظر (بدون استفاده از بردار هدف) است. در ابتدا بردار وزنی متناظر با هر نرون به‌طور تصادفی تولید شده و ساختار اولیه شبکه شکل می‌گیرد و سپس در طول فرآیند آموزش شبکه، بردار وزنی متناظر با هر نرون به‌گونه‌ای تنظیم می‌شود که بتواند

قسمتی از اطلاعات فضای مورد تحلیل را پوشش دهد. بدین ترتیب، در قسمت‌هایی از فضا که چگالی داده‌ها بیشتر است، تعداد نرون بیشتری قرار خواهند گرفت و توپولوژی شبکه مطابق با توزیع مشترک مشخصه‌های فضا شکل می‌گیرد. الگوریتم آموزش شبکه‌های خودسازمانده چهار مرحله دارد (Kim et al, 2006):

- انتخاب پارامترهای نقشه مانند ابعاد و بردار وزن ابتدایی متناظر با هر نرون؛
- ارائه داده‌های مورد تحلیل به شبکه و یافتن بهترین نرون نظیر برای هر بردار داده ورودی؛
- به هنگام کردن بردار وزنی متناظر با هر نرون؛
- بررسی شرط خاتمه الگوریتم.

چنانچه شرط بر قرار نباشد، الگوریتم از قدم دوم ادامه می‌یابد. از آنجاکه الگوریتم آموزش شبکه‌های خودسازمانده بر مبنای فاصله اقلیدسی بنا شده است، باید داده‌های هر بعد فضای مورد بررسی را مستقلاً نرمال استاندارد نمود. پس از پایان مرحله آموزش شبکه‌های خودسازمانده، نقشه‌ای از نرون‌ها به دست می‌آید که در حقیقت چکیده‌ای از فضای مورد تحلیل شبکه است. با ارائه هر بردار اطلاع جدید از فضای مورد تحلیل به شبکه، فاصله اقلیدسی بردار وزنی متناظر با هر یک از نرون‌ها تا بردار ورودی، به دست آمده است؛ بنابراین مقدار تحریک هر یک از نرون‌ها محاسبه و نرونی که بیش‌ترین مقدار تحریک را داشته باشد به عنوان نرون برنده انتخاب می‌شود.

۲-۳. نمایش فضای مورد تحلیل با استفاده از شبکه‌های خودسازمانده

پس از آموزش شبکه‌های خودسازمانده، به تعداد نرون‌های انتخاب شده برای شبکه، بردارهای وزنی بعدی به دست می‌آیند که هر یک نمایانگر بخشی از فضای مورد تحلیل است. در صورت انتخاب تعداد مناسب نرون، ابعاد شبکه و در نهایت آموزش مناسب شبکه، نمایش بردارهای وزنی متناظر با نرون‌های هر نقشه می‌تواند به خوبی

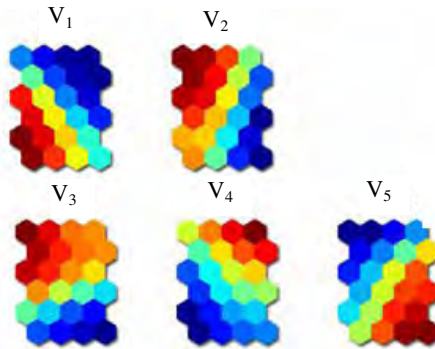
نمایانگر فضای مورد تحلیل باشد. بدین‌منظور متناظر با مقدار هر مشخصه در بردار وزنی یک بردار RGB^1 و در نتیجه یک رنگ در نظر گرفته می‌شود. به‌گونه‌ای که کلیه مقادیر با استفاده از طیف رنگی، از آبی تیره (برای کمترین مقادیر) تا قرمز تیره (برای بیشترین مقادیر)، قابل نمایش باشند. بدین ترتیب به‌ازای هر مشخصه، رنگ هر نرون تعیین شده و نقشه متناظر با آن مشخصه به‌دست می‌آید. با به‌دست آمدن نقشه‌های مشخصات، بررسی ارتباط متقابل میان آنها (تست همبستگی) ممکن می‌شود. به‌عنوان مثال، هم‌رنگ بودن قسمت‌های متناظر دو نقشه، نشان‌دهنده همبستگی مشخصه‌های نظیر با هر نقشه است. میزان همبستگی میان دو متغیر در قسمت‌های مختلف فضا نیز از شدت تفاوت یا تشابه رنگ میان نقشه‌های متناظر با هریک قابل بررسی است (Kreidler, 2008). شکل ۲، نشان‌دهنده نمونه‌ای از کاربرد شبکه‌های خودسازمانده در تحلیل الگوهای پیچیده و نمایش هم‌زمان اثرات متغیرهای مختلف بر یکدیگر است. از مقایسه نقشه‌ها با یکدیگر، موارد زیر قابل استنتاج است:

- متغیرهای V_2 و V_5 و همچنین V_1 و V_4 ، در تمام دامنه تغییرات خود همبستگی معکوس دارند هر جا V_2 دارای رنگ قرمز است (مقادیر بالا به خود گرفته)، V_5 دارای رنگ آبی است (مقادیر پایین به خود گرفته است و بالعکس). گرچه شدت همبستگی V_2 و V_5 در تمام نقاط فضا تقریباً ثابت است، اما این مطلب در مورد متغیرهای V_1 و V_4 صادق نیست.
- متغیرهای V_3 و V_5 همبستگی معکوس دارند، اما شدت همبستگی در نقاط مختلف فضا متغیر و کمتر از شدت همبستگی V_2 و V_5 است.

۱. Red-Green-Blue: از فرمت‌های استاندارد تعریف رنگ‌هاست که هر رنگی را با توجه به میزان شدت رنگ‌های

اصلی و از ترکیب آن‌ها قابل حصول می‌سازد.

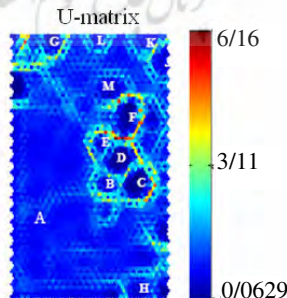
شکل ۲. کاربرد شبکه‌های خودسازمانده در تحلیل هم‌زمان روابط غیرخطی میان متغیرها



- ماتریس U

از جمله خروجی‌های دیگر شبکه‌های خودسازمانده، ماتریس دسته‌بندی و متناظر با آن نقشه دسته‌بندی است. درایه‌های این ماتریس، فاصله جبری نرون‌های همسایه را از یکدیگر نشان می‌دهند. بدیهی است انتظار داشته باشیم در صورتی که مشخصات دو قسمت از فضای مورد تحلیل به یکدیگر شبیه باشند، فاصله جبری بردارهای وزنی نرون‌های متناظر با آنها زیاد نباشد. یا به عبارت دیگر هر دو نرون، متعلق به خوشه واحدی از فضای مورد تحلیل باشند. شکل ۳ یک ماتریس U را با تعدادی خوشه و زیرخوشه از یک فضای ۶۲ بعدی نشان می‌دهد.

شکل ۳. نمایشی از یک U_Matrix



۴. مدل پیشنهادی

در این تحقیق مشتریان بیمه‌بدنه اتومبیل در یک شرکت بیمه‌ای براساس ریسک آنها بخش‌بندی گردیدند. به‌طورکلی گام‌های مدل پیشنهادی شامل تعریف متغیرها، جمع‌آوری داده‌ها، پاکسازی داده‌ها، طراحی مدل و تجزیه و تحلیل داده‌هاست.

۴-۱. تعریف متغیرها

از عوامل مهم در ساخت یک مدل، انتخاب درست مشخصه‌هاست. جهت بخش‌بندی مشتریان براساس ریسک، اولین و مهم‌ترین گام اصلی، شناسایی عوامل ریسک است. شناسایی و انتخاب مشخصه‌ها در این تحقیق در دو فاز صورت گرفت.

- فاز اول) مطالعه و بررسی تحقیقات موجود

در این فاز، مقالات علمی معتبر در مجلات در بازه زمانی ۲۰۰۹-۲۰۰۰ بررسی شدند.

این مرحله از حساس‌ترین مراحل انجام این تحقیق است. در این قسمت هیجده مشخصه مطابق جدول ۲ استخراج گردید. سپس این مشخصه‌ها در چهار گروه مشخصات جمعیت‌شناختی، مشخصات خودرو، سابقه راننده و مشخصات بیمه‌نامه برحسب نوع هریک از مشخصه‌ها تقسیم‌بندی گردیدند.

جدول ۲. فاکتورهای استخراج شده از ادبیات تحقیق

عوامل	محققان																										
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
مشخصات جمعیت‌شناسی	سن	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	جنسیت	✓	✓																								
	وضعیت تأهل																										
	محل زندگی	✓				✓																					
	نوع گواهینامه	✓																									
	سال اخذ گواهینامه (سابقه رانندگی)	✓	✓	✓					✓																		
	میزان تحصیلات																										
	میزان درآمد																										
	نوع خودرو		✓																								
	رنگ خودرو																										
ظرفیت موتور																											
مدل خودرو																											
مشخصات خودرو	تجهیزات ایمنی (ABS)								✓																		
	میزان کارکرد بر حسب کیلومتر		✓																								
	ساخت داخلی یا وارداتی																										
سابقه راننده	تعداد اصدای خسارت در سال قبل																										
	سرعت رانندگی																										
مشخصات بیمه‌نامه	میزان پوشش‌های بیمه‌ای																									✓	

- فاز دوم) نظرسنجی از خبرگان

پس از استخراج مشخصه‌ها از مقالات و گزارشات علمی، پرسش‌نامه‌ای طراحی و توسط خبرگان تکمیل گردید. باتوجه به این موضوع که خبرگان باید از بین افرادی انتخاب گردند که در زمینه خسارت‌های وارده به خودرو تجربه و دانش کافی داشته باشند، لذا کارشناسان خسارت مالی در بخش بیمه بدنه اتومبیل در شرکت‌های بیمه و رؤسا و مدیران آنها جهت نظرسنجی انتخاب گردیدند.

در کشور ما درحال‌حاضر هیجده شرکت بیمه در زمینه بیمه اتومبیل فعالیت دارند که تعداد کل کارشناسان، رؤسا و مدیران خسارت مالی در ده شرکت بیمه‌ای در تهران که سهم بازار آنها بیش از یک درصد است به‌عنوان جامعه آماری انتخاب گردیدند. باتوجه به بررسی‌های صورت‌گرفته، تعداد کل کارشناسان خسارت مالی و رؤسا و مدیران واحد خسارت مالی در مجموع ۴۸۵ نفر در شهر تهران برآورد گردید.

جهت تعیین اندازه نمونه باتوجه به اینکه جامعه آماری، تعداد متناهی ولی بسیاری را شامل می‌گردد و باتوجه به روش نمونه‌برداری تصادفی، از روش محاسبه حجم نمونه کوکران^۱ استفاده شد. با استفاده از فرمول کوکران اندازه نمونه ۱۱۴ نمونه انتخاب گردید. پرسش‌نامه‌ها به نسبت سهم بازار هریک از ده شرکت برگزیده توزیع گردید. از بین ۱۱۴ پرسش‌نامه توزیع گردیده در تهران، کارشناسان به ۸۷ پرسش‌نامه پاسخ دادند.

جهت انتخاب فاکتورهای نهایی و تحلیل نظرات خبرگان از روش آزمون سنگریزه استفاده شد. این روش را اولین بار کاتل^۲ در سال ۱۹۶۶ به‌منظور تخمین گرافیکی فاکتورهای بااهمیت معرفی کرد. طبق نظرات کیسر^۳ این روش از روش‌های بسیار

-
1. Cochran's Sample Size Formula
 2. Cattell, 1966
 3. Keiser

پرکاربرد است که در اغلب نرم افزارهای آماری همچون نرم افزارهای SPSS و Minitab هم گنجانده شده است. آزمون سنگریزه در حقیقت مقادیر ویژه و ارزش هریک از فاکتورها را بر روی نمودار از صعودی به نزولی به تصویر می کشد. بدین ترتیب نقطه شکست در انتخاب فاکتورها مشخص می گردد (Bener et al, 2006).

پس از تجزیه و تحلیل نتایج پرسش نامه علاوه بر هیجده فاکتور انتخاب شده، دو فاکتور شامل فاکتور سیستم خودرو (مانند پراید، زانتیا، پژو و...) و همچنین کاربری خودرو (مانند شخصی، تاکسی، اداری و...) با توجه به نظر خبرگان به سایر عوامل اضافه گردید.

فاکتورهای نهایی انتخاب شده در این تحقیق مطابق جدول ۳ دسته بندی گردیدند.

جدول ۳. فاکتورهای نهایی به دست آمده از نظر سنجی خبرگان

عوامل برگزیده	گروه عوامل
سن	مشخصات جمعیت شناختی
جنسیت	
محل زندگی	
نوع گواهینامه رانندگی	
سال اخذ گواهینامه (سابقه رانندگی)	
سیستم خودرو	مشخصات خودرو
کاربری خودرو	
نوع خودرو	
رنگ خودرو	
ظرفیت موتور	
مدل خودرو	
تجهیزات ایمنی (ABS)	
میزان کارکرد بر حسب کیلومتر	سابقه رانندگی
تعداد ادعای خسارت در سال قبل	
سرعت رانندگی	مشخصات بیمه نامه
میزان پوشش های بیمه ای	

۴-۱-۱. روایی و پایایی پرسش‌نامه

مفهوم روایی یا اعتبار به این سؤال پاسخ می‌دهد که ابزار اندازه‌گیری تا چه حد خصیصه موردنظر را می‌سنجد. بدون آگاهی از اعتبار ابزار اندازه‌گیری نمی‌توان به دقت داده‌های حاصل از آن اطمینان داشت (ماجد، ۱۳۸۷). در این تحقیق سوالات پرسش‌نامه با توجه به مطالعات و تحقیقات معتبر علمی طراحی گردید. در این تحقیق برای تعیین پایایی یا قابلیت اعتماد نتایج از روش آلفای کرونباخ استفاده شده است. این روش برای محاسبه همابندی درونی ابزار اندازه‌گیری به کار می‌رود که فرمول آن عبارت است از:

$$\text{Alpha} = N / (N-1) [1 - \sum \sigma^2(Y_j) / \sigma_x^2]$$

که در این فرمول N معرف تعداد زیرمجموعه سؤال‌های پرسش‌نامه یا آزمون، $(Y_j)^2$ واریانس زیر آزمون j ام و σ_x^2 واریانس کل آزمون است.

داده‌های پرسش‌نامه با استفاده از نرم‌افزار SPSS 17.0 تحلیل شدند تا پایایی آن مشخص گردد. با توجه به اینکه مقادیر بزرگ‌تر از ۰/۷ است، بنابراین آزمون از پایایی قابل قبولی برخوردار است (الوانی و همکاران، ۱۳۸۳).

۴-۲. جمع‌آوری داده

در این تحقیق داده‌های موجود در پایگاه داده بیمه بدنه اتومبیل در یکی از شرکت‌های بیمه داخلی استفاده شد. پنج فاکتور از فاکتورهای نهایی شامل سن، نوع گواهینامه و سال اخذ گواهینامه بیمه‌گذار و همچنین ظرفیت موتور و سرعت راننده به دلیل موجود نبودن اطلاعات در پایگاه داده در ابتدای امر حذف گردید. با توجه به اینکه فراوانی سه نوع خودروی پراید، زانتیا و وانت مزدا با سایر انواع خودروهای بیمه‌شده در این شرکت تفاوت چشمگیری داشتند، لذا تنها سه نوع خودرو نامبرده در این تحقیق تجزیه و تحلیل شد. از آنجاکه داده‌های مورد نیاز برای برخی از فاکتورها همچون میزان کارکرد (برحسب کیلومتر مصرفی) تنها در سال ۱۳۸۸ ورود

اطلاعات گردیده بود، لذا تنها پایگاه داده بیمه بدنه اتومبیل در سال ۱۳۸۸ بررسی شد. همچنین به دلیل آنکه یکی از فاکتورهای برگزیده، سابقه خسارت است لذا بیمه گذارانی مورد بررسی قرار گرفتند که سابقه خسارت آنها در سال گذشته در پایگاه داده موجود بوده است. خسارت در بیمه بدنه به طور کلی در دو دسته جداگانه خسارت‌های مربوط به تصادفات و خسارت‌های مربوط به سرقت قرار دارند. از آن جهت که این دو مسئله از یکدیگر کاملاً متفاوت اند و همچنین آمار مربوط به سرقت‌های اعلام شده در بازه یک‌ساله بسیار پایین است لذا در این تحقیق داده‌های مربوط به آمار سرقت‌های اعلام گردیده حذف شد. همچنین لازم به ذکر است که خودروهایی بررسی شدند که سال ساخت آنها در بازه زمانی بین سال‌های ۱۳۸۸-۱۳۸۰ قرار داشت. تعداد رکوردهای موجود پس از اعمال تغییرات بیان گردیده در مجموع برابر ۳۴۷۲۶ رکورد است.

۳-۴. پیش پردازش داده

پیش پردازش داده‌ها از گام‌های مهم در فرآیند داده‌کاوی است که میزان دقت نتایج به دست آمده رابطه قوی با نحوه انجام آن دارد. برای پیش پردازش داده‌های تعلیم شبکه، فعالیت‌های ذیل انجام گرفت:

- حذف دادگان نامناسب؛
- تبدیل کلیه مقادیر به مقادیر عددی؛
- نرمال‌سازی مقادیر.

باتوجه به اینکه مبنای آموزش شبکه‌های خودسازمانده براساس فاصله اقلیدسی بنا شده است، نرمال‌بودن دادگان تعلیم شبکه و وجود مقادیر خارج از الگو می‌تواند نقشه‌های خروجی را به شدت تحت تأثیر قرار دهد. به همین جهت نرمال‌سازی داده‌های ورودی، گام مهمی در آموزش شبکه‌های خودسازمانده است. در این تحقیق برای نرمال‌سازی دادگان تعلیم، روش نرمال‌سازی Z که کاربردهای فراوانی در آمار

دارد، به‌کار رفت. در این روش نرمال‌سازی از طریق تقسیم اختلاف داده‌ها با میانگین بر انحراف معیار جامعه انجام می‌شود.

۴-۴. طراحی و ساخت مدل

در این تحقیق از شبکه خودسازمانده به‌عنوان یکی از روش‌های پرکاربرد در تکنیک خوشه‌بندی در داده‌کاوی استفاده گردیده است. در ابتدا داده‌ها با استفاده از تکنیک شبکه‌های خودسازمانده آموزش داده شد و نتایج حاصل از خروجی شبکه در قالب نقشه‌های گرافیکی تحلیل شد.

۴-۴-۱. شبکه خودسازمانده

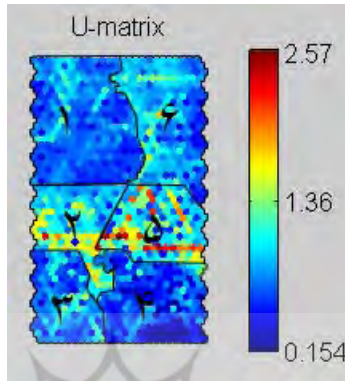
اساس شبکه‌های خودسازمانده نگاشت فضاها با ابعاد بالا به فضای دو بعدی یا سه بعدی نقشه خروجی است، به‌گونه‌ای که کم‌ترین میزان اطلاعات از بین رفته و دانش نهفته در روابط میان دادگان تعلیم آشکار شود. داده‌های تعلیم شبکه از ۳۴۷۲۶ بردار یازده بعدی تشکیل شده است که هر بردار نماینده یک رکورد از رکوردهای پایگاه داده بیمه بدنه شرکت مورد مطالعه است. ابعاد این بردار برابر تعداد فاکتورهایی است که پس از نظرسنجی از خبرگان انتخاب گردیدند. رابطه زیر مجموعه دادگان تعلیم شبکه را تعریف می‌کند:

$$X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}, X_{i8}, X_{i9}, X_{i10}, X_{i11}) \quad i: 1 \text{ to } 34726$$

در مورد تعداد نرون‌های لایه نقشه تحقیقات زیادی انجام گرفته است. کوهنن^۱ مبدع این شبکه‌های خودسازمانده در کتابی با همین عنوان، فرمول $5\sqrt{n}$ را برای تعداد نرون‌های لایه نقشه توصیه می‌کند، که در آن n تعداد داده‌های آموزش است. در این تحقیق نیز از همین رابطه استفاده شده و با توجه به اینکه تعداد داده‌های آموزش ۳۴۷۲۶ نمونه بوده است، تعداد نرون‌های لایه نقشه ۹۳۲ نرون انتخاب شده

است. شکل ۴ ماتریس U حاصل از شبکه خودسازمانده آموزش دیده به وسیله داده‌های بیمه بدنه اتومبیل را نشان می‌دهد.

شکل ۴. ماتریس U شبکه خودسازمانده پس از آموزش داده‌های شرکت بیمه مورد مطالعه



ماتریس U به دست آمده، شش بخش یا شش خوشه اطلاعاتی دارد و هریک از مشتریان بیمه اتومبیل در یکی از این شش خوشه قرار می‌گیرند. مشتریان متعلق به هریک از این بخش‌ها از لحاظ مشخصه‌های بررسی شده، دارای شباهت نزدیک به یکدیگر بوده و اختلاف زیادی با مشخصه‌های مشتریان موجود در سایر بخش‌ها دارند. به همین جهت انتظار می‌رود که ریسک مشتریان قرارگرفته در یک گروه نیز مشابه یکدیگر بوده و با ریسک سایر مشتریان در گروه‌های دیگر متفاوت باشد.

جهت تعیین برچسب و نام هریک از خوشه‌ها تعدادی نمونه تصادفی از هر بخش انتخاب گردید و میانگین ضریب خسارت نمونه‌های انتخابی محاسبه شد. جدول ۴ نحوه برچسب‌زنی بخش‌های نقشه خودسازمانده را نشان می‌دهد.

جدول ۴. برچسب‌زنی گروه‌های مختلف مشتریان براساس ریسک

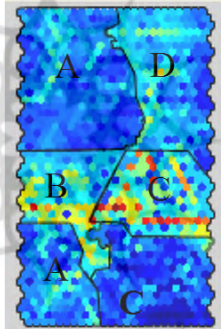
بخش	نمونه	ضریب خسارت	میانگین ضریب خسارت	میزان ریسک بخش
۱	۱	۰/۱۵	۰/۱۵	ریسک پایین
	۲	۰		
	۳	۰/۳۶		
	۴	۰		
	۵	۰/۲۴		
۲	۱	۰/۳۸	۰/۴	ریسک نسبتاً پایین
	۲	۰/۴۴		
	۳	۰/۵۴		
	۴	۰/۶۵		
	۵	۰		
۳	۱	۰	۰/۱۸	ریسک پایین
	۲	۰/۳		
	۳	۰		
	۴	۰/۳۶		
	۵	۰/۲۴		
۴	۱	۱/۴	۱/۷۲	ریسک بالا
	۲	۲/۲		
	۳	۰/۹		
	۴	۱/۶		
	۵	۲/۵		
۵	۱		۱/۳۶	ریسک بالا
	۲	۱/۲		
	۳	۰/۹		
	۴	۱/۶		
	۵	۲/۱		
۶	۱	۰/۷۷	۰/۷۳	ریسک نسبتاً بالا
	۲	۰/۶۵		
	۳	۰/۸		
	۴	۰/۷۸		
	۵	۰/۶۶		

میانگین ضریب خسارت خوشهٔ یک و سه نزدیک به هم است، لذا هر دو خوشه در طبقهٔ مشتریان با ریسک پایین جای گرفتند. همچنین دو خوشهٔ چهار و پنج نیز

میانگین ضریب خسارت نزدیک به هم دارند، لذا این دو خوشه هم باتوجه به آنکه میانگین ضریب خسارت آنها بالاتر از یک است - بدین معنا که خسارت واردشده به بیمه از سوی مشتریان این خوشه بیش از حقیبیمه پرداختی است- لذا این دو خوشه نیز در طبقه مشتریان با ریسک بالا جای گرفتند.

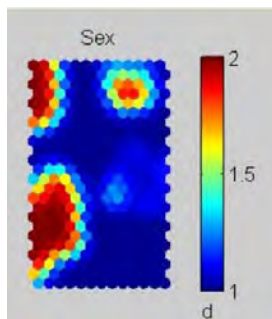
به طور کلی می توان ماتریس U را به چهار خوشه مطابق شکل ۵ تقسیم بندی کرد. مشتریان خوشه A مشتریانی با ریسک پایین، مشتریان خوشه B مشتریانی با ریسک نسبتاً پایین، مشتریان خوشه C مشتریانی با ریسک بالا و مشتریان خوشه D مشتریانی با ریسک نسبتاً بالا می باشند.

شکل ۵. تعیین برجسب هریک از خوشه های ماتریس U



به طور کلی نقشه های رسم شده امکان سه نوع تحلیل مختلف را برای ما فراهم می کنند. ابتدا می توان در مورد هریک از متغیرها تنها با استفاده از نقشه رسم شده آن اطلاعاتی را استخراج کرد. به عنوان نمونه در نقشه متغیر جنسیت بیمه گذار مطابق شکل ۶ رنگ آبی تیره و قرمز تیره به ترتیب نمایانگر جنسیت مرد و زن است. همان گونه که در نقشه مشخص است فراوانی بیمه گذاران مرد نسبت به زن بیشتر است.

شکل ۶. نقشه متغیر جنسیت بیمه‌گذار



از مقایسه ماتریس U رسم شده با نقشه تک تک متغیرها امکان تحلیل ویژگی‌های هریک از خوشه‌ها فراهم می‌گردد. در ادامه به‌عنوان نمونه خوشه با ریسک بالا و خوشه با ریسک نسبتاً پایین تحلیل می‌گردد:

- تحلیل خوشه با ریسک بالا

- سیستم خودرو: با توجه به طیف رنگی در این ناحیه متوجه می‌شویم که تعداد خودرو وانت در مقایسه با خودرو زانتیا و پراید بیشتر است. همچنین نسبت خودرو پراید در مقایسه با خودرو زانتیا بیشتر است.
- تجهیزات ایمنی (ترمز ضد قفل): اکثر خودروهای این ناحیه فاقد سیستم ترمز ABS می‌باشند. برخی از خودروهای زانتیا و پراید در این ناحیه مجهز به این سیستم ترمزند.
- رنگ خودرو: تقریباً کلیه طیف‌های رنگی در این قسمت دیده می‌شود. خودروهای با رنگ‌های تیره، رنگ‌های تند (مانند نارنجی، زرد، آلبالویی و ...) و رنگ‌هایی از خانواده رنگ آبی در این قسمت قابل مشاهده است. رنگ‌های روشن (مانند سفید و کرم) در مقایسه با رنگ‌های تیره و سایر رنگ‌ها در این ناحیه کم‌تر است.

1. Anti-Lock Break System (ABS)

- مدل خودرو: سال ساخت خودروهای این ناحیه غالباً پایین است.
- میزان کارکرد برحسب کیلومتر: میزان کارکرد بسیار بالا با رنگ قرمز تیره و کیلومترهای بالا با رنگ قرمز روشن مشخص گردیده است. اغلب خودروها در این ناحیه میزان کیلومتر مصرفی بالا دارند. در برخی از قسمت‌های نقشه هم میزان کیلومتر مصرفی پایین و در بعضی جاها حتی صفر کیلومتر هم دیده شده است.
- نوع خودرو: باتوجه به آنچه در نقشه دیده می‌شود در مقایسه با خودروهای وانت، تعداد کمی از خودروها در این قسمت از نوع سواری‌اند.
- کاربری خودرو: باتوجه به اینکه خودروهای این منطقه بیشتر از نوع وانت‌اند کاربری این نوع خودرو از نوع بارکش است. نکته قابل توجه در مورد خودروهای پراید در این ناحیه این است که اغلب خودروهای پرایدی که کاربری تاکسی یا آژانس دارند در این ناحیه قرار گرفتند.
- جنسیت بیمه‌گذار: جنسیت غالب بیمه‌گذاران این قسمت مرد است.
- شهر محل زندگی: در نقشه این متغیر طیف رنگ‌های آبی پررنگ، آبی کم‌رنگ و زرد قابل مشاهده است که به ترتیب نمایانگر کلان شهر و شهرهای بزرگ و نسبتاً بزرگ است.
- میزان پوشش بیمه‌ای خریداری‌شده: میزان پوشش بیمه‌ای خریداری‌شده در این ناحیه بسیار پایین است.
- تعداد خسارت در سال قبل: بیشترین طیف رنگ قرمز در این ناحیه مشاهده می‌شود که نمایانگر میزان سابقه خسارت‌های بالا (سابقه خسارت ۲، ۳ و ۴) است. همچنین بسیاری از خودروهای این ناحیه دارای سابقه خسارت یک می‌باشند.

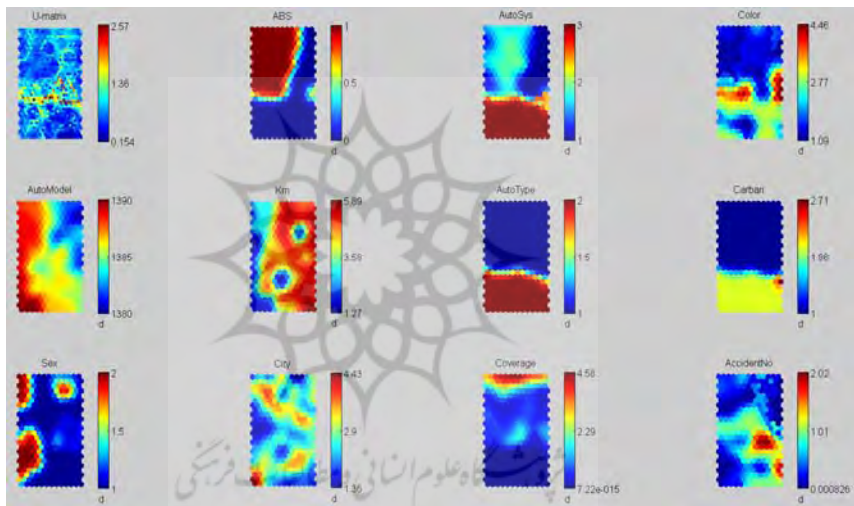
- تحلیل خوشه با ریسک نسبتاً پایین

- سیستم خودرو: هر سه نوع خودرو پراید، زانتیا و وانت در این ناحیه دیده می‌شود. تقریباً پراکندگی خودروی پراید و زانتیا در این بخش مشابه هم است. در مقایسه با این دو نوع خودرو، تعداد کم‌تری خودرو وانت در این ناحیه قرار گرفته است.
- تجهیزات ایمنی (ABS): خودروهایی که مجهز به سیستم ترمز ABS یا فاقد این سیستم‌اند، هر دو در این ناحیه قابل مشاهده‌اند.
- رنگ خودرو: کلیه طیف‌های رنگی در این ناحیه قابل مشاهده است.
- مدل خودرو: همه مدل‌های خودرو در این ناحیه وجود دارد. اغلب خودروهای این ناحیه مدل‌های ساخت بالا دارند.
- میزان کارکرد برحسب کیلومتر: طیف‌های رنگی آبی کم‌رنگ، نارنجی و زرد بیشتر در این ناحیه دیده می‌شود که نشان‌دهنده کارکردهای پایین و نسبتاً بالاست. همچنین در برخی قسمت‌ها خودروهای صفر کیلومتر هم دیده می‌شود.
- نوع خودرو: هر سه نوع خودرو در این ناحیه قرار دارد.
- کاربری خودرو: کاربری خودروهای این ناحیه شامل خودروهای شخصی و بارکش است.
- جنسیت بیمه‌گذار: هر دو نوع جنسیت در این ناحیه دیده می‌شود. با توجه به نقشه می‌توان فهمید که تعداد بیمه‌گذاران زن در این ناحیه بیشتر از مردان است.
- شهر محل زندگی: بیشتر بیمه‌گذاران این ناحیه در شهرهای بزرگ و نسبتاً بزرگ قرار دارند.
- میزان پوشش بیمه‌ای خریداری‌شده: با توجه به طیف رنگی آبی تیره و آبی روشن می‌توان فهمید که میزان پوشش بیمه‌ای خریداری‌شده در این ناحیه پایین است.

- تعداد خسارت سال قبل: با توجه به رنگ نقشه در این ناحیه می توان گفت که خودروهای با سابقه خسارت صفر و یک در این بخش دیده می شود. تعداد بسیار کمی هم خسارت بالای یک دیده می شود.

همچنین تحلیل دیگری که می توان به کمک نقشه متغیرها انجام داد، یافتن روابط معنی دار و بی معنی در بین متغیرها از طریق مقایسه نقشه دو به دو متغیرهاست. شکل ۷ نقشه هم زمان کلیه متغیرها و ماتریس U را نشان می دهد.

شکل ۷. نمایش هم زمان ماتریس U و سایر متغیرها



به عنوان نمونه از روی هم قراردادن نقشه متغیرهای جنسیت و متغیر شهر محل زندگی می توان نتیجه گرفت که فراوانی بیمه گذاران زن در کلان شهرها و شهرهای بزرگ و نسبتاً بزرگ بیشتر از شهرهای کوچک است. با توجه به نقشه متغیر جنسیت و متغیر میزان پوشش اضافی خریداری شده، رابطه معنی داری بین متغیر میزان پوشش اضافی خریداری شده و جنسیت بیمه گذار نمی توان یافت. در رابطه با متغیر تعداد تصادف در سال قبل باید به این نکته اشاره کرد که آمار تصادفات در بین بیمه گذاران مرد در مقایسه با خانمها بیشتر است.

۵. نتیجه‌گیری

در ایران نرخ حق بیمهٔ بدنهٔ اتومبیل با توجه به تعرفهٔ بیمه مرکزی ج.ا.ا. تعیین می‌شود و این امر سبب می‌گردد که مشتریانی با ریسک پایین‌تر، خسارت‌های مالی مشتریانی با ریسک بالاتر را جبران کنند و لذا تفاوت آنچنانی بین مشتریان با ریسک بالا و پایین وجود نخواهد داشت.

در این تحقیق سعی گردید فاکتورهای مهم و اثرگذار بر ریسک مشتریان شناسایی گردیده و همچنین با کمک شبکه‌های عصبی خودسازمانده، مشتریان بیمهٔ بدنهٔ اتومبیل در یکی از شرکت‌های بیمهٔ خصوصی براساس ریسکشان خوشه‌بندی شوند. نتایج این تحقیق نشان داد که مشتریان برای شرکت‌های بیمه از جنبه‌های مختلف با یکدیگر متفاوت‌اند؛ به‌عنوان مثال ریسک انواع خودرو، رنگ خودرو و جنسیت بیمه‌گذار و بسیاری از عوامل دیگر سبب می‌شود که مشتریان از یکدیگر متمایز باشند. تجزیه و تحلیل داده‌ها اهمیت شاخص‌های نهایی انتخاب‌شده را به‌خوبی نشان داد. نکته قابل توجه در مورد شاخص‌های انتخابی این است که کلیهٔ شاخص‌های ریسکی که با مطالعهٔ تحقیقات گذشته و همچنین نظرسنجی از خبرگان تعیین گردید همگی اهمیت زیادی دارند. به‌عنوان نمونه با توجه به اینکه این تحقیق تنها بر روی سه نوع خودرو تمرکز نمود در بین این سه نوع خودرو، ریسک خودرو و وانت از دو خودرو دیگر و ریسک خودرو پراید از خودرو زانتیا بیشتر است. که از بین خودروهای زانتیا و پراید با استناد به نتایج تجزیه و تحلیل داده‌ها می‌توان بیان کرد که ریسک خودروهایی که مجهز به سیستم ترمز ABS می‌باشند در مقابل سایر خودروها کم‌تر است. نتایج همچنین نشان داد که ریسک بیمه‌گذاران مرد در برابر بیمه‌گذاران زن در کشور ما بالاتر است.

۶. محدودیت‌های پژوهش و پیشنهاد برای پژوهش‌های آتی

• در این تحقیق با توجه به اینکه فراوانی سه نوع خودرو پراید، زانتیا و وانت از سایر خودروها در شرکت مورد مطالعه بیشتر بود، تنها بر این سه نوع خودرو تمرکز کردیم. پیشنهاد می‌شود تحقیقات آتی بر روی انواع خودروهایی انجام شود که فراوانی نسبتاً مشابهی با یکدیگر داشته باشند؛ چرا که این امر بر نتایج تحلیل تأثیرگذار است.

• تحلیل بر روی تک‌تک انواع خودروها می‌تواند نتایج دقیق‌تر و بهتری را برای تأثیرگذاری سایر متغیرها در مورد آن خودرو خاص نشان دهد.

• برخی از فاکتورهای مهم و اثرگذار همچون سن بیمه‌گذار، سال اخذ گواهینامه و نوع گواهینامه رانندگی به دلیل موجود نبودن اطلاعات آن در پایگاه داده شرکت مورد مطالعه حذف گردید. با توجه به آنچه تحقیقات در مورد اثرگذاری و اهمیت این فاکتورها نشان می‌دهد، پیشنهاد می‌گردد در زمان صدور بیمه‌نامه برخی از این اطلاعات از مشتریان دریافت گردد تا بتوان با تحلیل آنها به نتایج مهمی دست یافت که در سودآوری شرکت مؤثر واقع گردد.

منابع

۱. الوانی، سیدمهدی، آذر، عادل و دانایی‌فرد، عادل ۱۳۸۳، روش‌شناسی پژوهش کمی در مدیریت: رویکردی جامع، ناشر اشراقی.
۲. جعفری صمیمی، احمد و مرادی، محمدعلی ۱۳۸۷، 'خصوصی‌سازی و بیمه اتومبیل در ایران، همایش بین‌المللی صنعت بیمه، چالش‌ها و فرصت‌ها.
۳. حسین‌زاده، لیلیا ۱۳۸۶، 'دسته‌بندی مشتریان هدف در صنعت بیمه با استفاده از داده‌کاوی، پایان‌نامه کارشناسی ارشد، دانشگاه تربیت مدرس.
۴. رستخیز پایدار، ندا ۱۳۸۹، 'بخش‌بندی مشتریان بیمه بدنه اتومبیل با استفاده از داده‌کاوی، مورد مطالعه، بیمه ملت، پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبائی.
۵. سرمد، زهره، بازرگان، عباس و حجازی، الهی ۱۳۸۳، روش‌های تحقیق در علوم رفتاری،

نشر آگاه تهران، چ ۱۰.

۶. ماجد، وحید ۱۳۸۷، 'راهکارهای طبقه‌بندی ریسکی بیمه‌گذاران در بازار بیمه: شواهدی از بازار بیمه بدنه اتومبیل در ایران، همایش بین‌المللی صنعت بیمه، چالش‌ها و فرصت‌ها.

7. Bener, A, Ghaffar, A, Azab, A, Sankaran-Kutty, M, Toth, F & Lovasz, G 2006, 'The impact of four-wheel drives on road traffic disability and deaths compared to passenger cars', *Journal of College of Physicians and Surgeons Pakistan*, vol. 16, no. 4, pp. 257-60

8. Cattell, RB 1966, *The scree test for the number of factors*, *Multivariate Behavioral Research*. 1, pp. 245-76.

9. Chen, YL, Hsu, CL & Chou, SC 2003, 'Constructing a multi-valued and multilabeled decision tree', *Expert Systems with Applications*, vol. 25, pp. 199-209.

10. Chou, JS & Hilty, RD 2003, *Proceedings of IPC annual meeting*, Exotic Interconnections and Advanced Process Capability, Minneapolis, MN, September 28- October 2.

11. Dennis, C, Marsland, D & Cockett, T 2001, 'Data mining for shopping centres customer knowledge management framework', *Journal of Knowledge Management*, vol. 5, pp. 368-74.

12. Dickson, PR 1982, 'Person-situation: segmentation's missing link', *Journal of Marketing*, vol. 46, no. 4, pp. 56-64.

13. Ha, SH, Bae, SM & Park, SC 2002, 'Customer's time-variant purchase behavior and corresponding marketing strategies: An online retailer's case', *Computers and Industrial Engineering*, vol. 43, pp. 801-20.

14. Hwang, H., Jung, T & Suh, E 2004, 'An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry', *Expert Systems with Applications*, vol. 26, no. 2, pp. 181-8.

15. Kim, SY, Jung, TS, Suh, EH, & Hwang, HS 2006, 'Customer segmentation and strategy development based on customer lifetime value: a case study', *Expert Systems with Applications*, vol. 31, no. 1, pp. 101-7.

16. Kohonen, T, *Automatic formation of topological maps of patterns in a self-organizing system*, Oja, E & Simula, O (Eds), *Proceedings Of SCIA Scand, Conference on Pattern Recognition*, Los Alamitos, CA. IEEE Computer Soc. Press, pp. 182-5.

17. Kohonen, T 2001, *Self-organizing maps*, Springer Series in Information Sciences, 30, Springer, Berlin, Heidelberg New York, Th³ed.

18. Kreidler, M 2008, *Guide to auto insurance*, Washington State Office of the Insurance Commissioner, viwed 2010/4/24 <<http://www.insurance.wa.gov>>.
19. Lee, JH & Park, SC 2005, 'Intelligent profitable customers segmentation system based on business intelligence tools', *Expert Systems with Applications*, vol. 29, pp. 145–52.
20. Lee, SC, Suh, YH, Kim, JK & Lee, KJ 2004, 'A cross-national market segmentation of online game industry using SOM', *Expert Systems with Applications*, vol. 27, pp 559–70.
21. Leung, Y, Zhang, J & Xu, Z 2000, 'Clustering by scale-space filtering', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, np.12, pp. 1396–410.
22. Li, Y & Bai, Y 2008, 'Development of crash-severity-index models for the measurement of work zone risk levels', *Accident Analysis and Prevention*, vol. 40, pp. 1724-31.
23. Ngai, EWT Xiu, L Chau, Dck 2009, 'Application of data mining techniques in customer relationship management: a literature review and classification', *Expert Systems with Applications*, vol. 36, pp. 2592-602.
24. Vesanto, J & Alhoniemi, E 2000, 'Clustering of the self-organizing map', *IEEE Transactions on Neural Networks*, vol. 11, no.3, pp. 586-600.
25. Vesanto, J 1997, Data mining techniques based on the Self-Organizing map', Helsinki University of Technology Web Log Classification. *IEEE Intelligent Systems*, vol. 20, pp. 55–9.
26. Woo, JY, Bae, SM & Park, SC 2005, 'Visualization method for customer targeting using customer map', *Expert Systems with Applications*, vol. 28, pp. 763–72.
27. Wu, CH, Kao, SC, Su, YY & Wu, CC 2005, 'Targeting customers via discovery knowledge for the insurance industry', *Expert Systems with Applications*, vol, 29. pp. 291-99.
28. Yu, JX, Ou, Y, Zhang, C & Zhang, S 2005, 'Identifying interesting visitors through web log classification', *IEEE Intelligent Systems*, vol. 20, pp. 55–9.