



بررسی عوامل ایجاد خسارت در بیمه بدنی اتومبیل ها با استفاده از الگوریتم درخت تصمیم

عاطفه جعفری^۱ و علی علی جماعت^۲

^۱ دانشجوی کارشناسی رشته مهندسی کامپیوتر- نرم افزار، دانشگاه آزاد اسلامی واحد کرج، at.jafari@yahoo.com

^۲ عضو هیات علمی ، دانشگاه آزاد اسلامی واحد ابهر ، amirjamaat@yahoo.com

چکیده - بیمه بدنی اتومبیل یکی از طرح های جبران کننده خسارات ناشی از حادث رانندگی برای اتومبیل ها است. در این نوع بیمه نامه مساله قابل توجه شرکت های بیمه این است که در برابر حق بیمه دریافت شده با خسارات هنگفت از سوی بیمه شدگان مواجه نشوند. در این پژوهش برآئیم که مدلی با بکارگیری یکی از تکنیک های داده کاوی به نام درخت تصمیم، برای انواع مختلف خسارت های به بار آمده در اتومبیل ها بدست آوریم تا شرکت های بیمه در صورت مواجه شدن با چنین مدلی تصمیم بهتری در مورد بیمه کردن یا عدم بیمه کردن اتومبیل ها اتخاذ کنند. همچنین یکی از اهداف مقاله پیش رو بررسی عوامل موثر در تصادفات شدید رانندگی است که می تواند راهنمایی و رانندگی را در جهت پیش گیری از این تصادفات یاری دهد. نتایج بدست آمده از مطالعه تجربی بر روی ۱۰۵ پرونده خسارت بیمه بدنی در یکی از شعب پرداخت خسارت بیمه بدنی در تهران، نشانگر ارتباط میان عواملی چون زمان، فصل تصادف و سن راننده با شدت تصادفات و میزان خسارت می باشد.

کلید واژه- بیمه بدنی، حق بیمه، تکنیک های داده کاوی، درخت تصمیم.

۱- مقدمه

نیاز جوامع انسانی به جبران خسارت وارد به وسیله نقلیه بیمه شده و گاهی خسارات وارد به شخص ثالث می پردازند.^[۱] داده کاوی می تواند به شرکت های بیمه در جهت افزایش سودآوری و همچنین به راهنمایی و رانندگی در جهت شناسایی زمان و فصول اوج ترافیک درون شهری و جلوگیری از افزایش ترافیک در زمان های مذکور کمک کند. بر این اساس در این مقاله سعی شده است تا با شناخت الگوهای موثر در ایجاد تصادفات و با توجه به اهمیت علم داده کاوی در صنعت بیمه و راهنمایی و رانندگی به کاهش تصادفات و خسارات ناشی از آنها کمک بسزایی شود.

۲- بیمه بدنی اتومبیل

براساس شرایط این بیمه نامه، دارنده و یا مالک وسیله نقلیه اتومبیل خود را در برابر خطرات بیمه می کند تا در صورت بروز حادثه و ایجاد خسارت به وسیله نقلیه موضوع بیمه، شرکت بیمه خسارت وارد را براساس شرایط بیمه نامه جبران کند. منظور از خطرات مورد تعهد بیمه گر، همان حادثی است که در صورت واقع شدن چنانچه باعث نابودی و یا آسیب دیدگی اتومبیل شود

اتومبیل وسیله‌ای است که تقریباً از اواخر قرن نوزدهم و به شکلی همه‌گیر وارد زندگی انسان در جوامع بشری شده و نقشی بسیار مهم در حمل مسافر و جابجایی کالا دارد. که براستی باید گفت زندگی انسان بدون بهره‌گیری از این وسیله نقلیه موتوری ناممکن است. در واقع یکی از ویژگیهای بارز و مشخص قرن بیستم، پدید آمدن این وسیله نقلیه است. پدیدار شدن این وسیله نقلیه اگرچه به همراه خود رفاه، آسایش و سرعت را به ارمغان آورده اما به همان نسبت باعث بروز مشکلات بسیار و خطرات فراوان هم شده است. مهمترین خطری که جامعه امروز را تهدید می کند خطر تصادفات رانندگی است. بنابراین از علل مهم وقوع تصادفات می توان به عواملی چون علت تامه تصادفات (رعایت نکردن حق تقدم، بی توجهی به جلو، عدم رعایت فاصله طولی، تغییر مسیر ناگهانی، حرکت با دندنه عقب، عدم توانایی در کنترل وسیله نقلیه و رعایت نکردن فاصله عرضی)، نیستن کمربند ایمنی، شرایط روحی و روانی و جسمی راننده مقصص، ساعت وقوع تصادفات، فصول سال، سن و جنس راننده مقصص، شدت صدمه و نوع تصادفات، نوع وسیله نقلیه مقصص در تصادفات اشاره کرد.^[۲] شرکتهای بیمه با درک این

تکنیک های داده کاوی به دو بخش تقسیم می شوند: تکنیک های پیش بینی کننده و تکنیک های توصیف کننده. تکنیک های پیش بینی کننده تکنیک های هستند که برای ایجاد الگو با استفاده از مدل های آماده مورد استفاده قرار می گیرند. مانند دسته بندی، رگرسیون، سری های زمانی و پیش بینی. تکنیک های توصیف کننده تکنیک هایی هستند که از قبل برای آنها الگویی وجود ندارد و براساس ویژگی ها تقسیم بندی می شوند. مانند: خوش بندی، قواعد انجمنی، خلاصه سازی، الگوهای مکرر.

[۳]

۱-۳- دسته بندی

به فرایند پیدا کردن یک مدل از پیش تعیین شده برای داده ها به منظور پیش بینی دسته های جدید دسته بندی گفته می شود. دسته بندی داده ها فرایندی دو مرحله ای است، اولین مرحله ساخت مدل و دومین مرحله استفاده از مدل و پیش بینی از طریق داده های قبلی می باشد. مرحله ساخت مدل عبارتست از توصیف یک سری از دسته های از پیش تعیین شده بر مبنای مجموعه داده های آموزش مدل که این فرایند یادگیری نامیده می شود. مرحله استفاده از مدل شامل دو بخش است، در بخش نخست مدل ساخته شده مورد آزمون واقع می شود تا دقیقت پیش بینی آن بررسی شود و در بخش دوم نیز مدلی که دارای دقیقت مناسبی است برای دسته بندی داده ها به کار گرفته می شود. [۶] از میان روش های دسته بندی که شامل بیز ساده و شبکه های بیزی، نزدیکترین همسایگی، شبکه های عصبی، درخت تصمیم و رگرسیون می باشد تکنیک درخت تصمیم در این پژوهش مورد استفاده قرار گرفته است.

۱-۱-۳- درخت تصمیم

درخت تصمیم یک ابزار برای پشتیبانی از تصمیم است که از درختان برای مدل کردن استفاده می کند. درخت تصمیم، به طور خاص در آنالیز تصمیم، برای مشخص کردن استراتژی که با بیشترین احتمال به هدف برسد بکار می رود. الگوریتم ایجاد یک درخت تصمیم الگوریتمی است که در یک فرایند از بالا به پایین درخت تصمیم را ایجاد می کند. [۴] درخت تصمیم دارای سه بخش ریشه، گره و برگ است. گره های میانی در درخت تصمیم ویژگی های اشیا هستند که براساس مقادیر آنها فرزندان ایجاد می شوند و برگ ها اشیا موجود در مجموعه داده ها می باشند.

بیمه گر موظف به جبران آن خواهد بود. در این تعریف لازم است مفاهیم زیر به عنوان مفاهیم پایه ای بیمه بدنی مطرح شود.

- بیمه گر: بیمه گر در این قرارداد شرکت سهامی بیمه است.

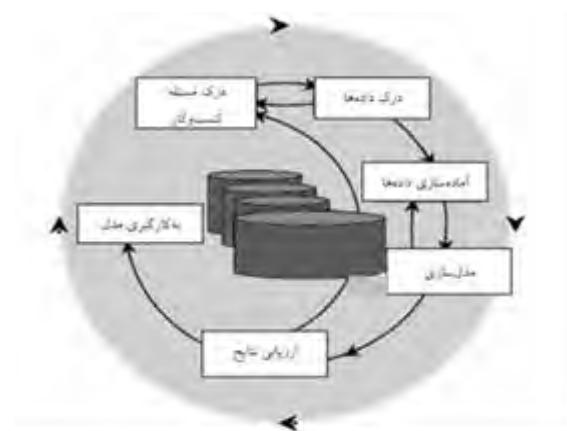
- بیمه گذار: شخص یا موسسه ای که به عنوان نماینده ای قانونی از طرف مالک وسیله نقلیه قرارداد بیمه را منعقد و پرداخت حق بیمه را متعهد می شود.

- وسیله نقلیه مورد بیمه: وسیله ای که مشخصات آن در شرایط خصوصی این بیمه نامه (بیمه بدن) تعیین و تصریح شود.

- حق بیمه: مبلغی است که بیمه گذار در مقابل بیمه کردن اتومبیل از ذینفع آن دریافت می نماید. [۲]

۳- داده کاوی

داده کاوی، اقتباس یا استخراج دانش از مجموعه ای از داده هاست. به بیان دیگر، داده کاوی فرایندی است که با استفاده از تکنیک های هوشمند، دانش را از مجموعه ای از داده ها استخراج می کند. دانش استخراج شده در قالب مدل ها، الگوها یا قواعد ارائه می شود. این الگوها، مدل ها و قواعد اشکال مختلفی برای ارائه دانش استخراج شده هستند. این دانش می تواند ملاک تصمیم گیری های آتی، عملکرد های بعدی یا تغییرات لازم در سیستم قرار گیرند. فرایند داده کاوی دارای مراحل متعددی است که این مراحل شامل درک مسئله کسب و کار، درک داده ها، آماده سازی داده ها، مدل سازی، ارزیابی نتایج و به کار گیری مدل است. [۴]



شکل ۱: مراحل مدل فرایندی داده کاوی بر اساس استاندارد CRIPS-DM [۴]



عنوانین استفاده شده برای فیلد های موجود در مدل داده کاوی بdst آمد. همچنین در این پژوهش برای برخی از فیلدها بازه هایی درنظر گرفته شده است که در جدول ۲ نمایش داده شده است. ضمناً بدلیل اینکه تمامی ۱۰۵ پرونده مربوط به تصادفاتی است که در ساعت ۶ صبح تا ۱۲ شب اتفاق افتاده است بازه ۱۲ شب تا ۶ صبح درنظر گرفته نشده است.

۴-۲-۴- مدل سازی

برای ساخت مدل لازم است ابتدا تکنیک مدل سازی انتخاب شود که در این پژوهش تکنیک درخت تصمیم انتخاب گردیده است. از آنجا که ساختار داده های پژوهش هم به صورت عددی و هم به صورت غیر عددی است، تکنیک انتخابی با این نوع داده ها هم خوانی داشته و قابل اجراست. نتایج اعمال الگوریتم درخت تصمیم بر روی داده ها در درخت های بdst آمده در شکل های ۱، ۲ و ۳ نمایش داده شده است. همچنین ابزار مورد استفاده در این پژوهش نرم افزار weka می باشد. در ادامه به بیان یافته های هریک از شکل ها و بهترین مدل های استخراجی در آنها می پردازیم.

۵- یافته ها

برای کشف بهترین مدل در درخت های حاصل از ریشه به برگ درخت را پیمایش می کنیم و مدلی کارآمد تر است و صحت بیشتری دارد که دارای بیشترین تعداد داده در برگ خود باشد. بنابراین در درخت شکل ۲ که بیانگر ارتباط بین مبلغ خسارت و ویژگی های خودرو و راننده می باشد، بهترین الگوها دارای برگ های ۱۹ تایی، ۱۳ تایی و ۱۱ تایی می باشد و به صورت زیر بیان می شوند:

۱- خودروهایی که قیمت بیشتر از ۲۰ میلیون تومان دارند در صورت تصادف خسارت هایی بیش از ۱۰ میلیون تومان ایجاد می کنند.

۲- اگر قیمت خودرو قبل از تصادف کمتر از ۱۰ میلیون تومان باشد و علت تصادف آن از فرعی به اصلی باشد و این تصادف در فصل پاییز رخ داده باشد خسارتی در حدود پانصد هزار تومان تا یک میلیون تومان به بار خواهد آورد.

۳- اگر قیمت خودرو در حدود ۱۰ تا ۲۰ میلیون تومان باشد و سن راننده ۲۰ تا ۳۰ سال باشد خسارتی در حدود یک میلیون تومان تا ده میلیون تومان به بار خواهد آورد.

۴- مراحل اجرای پژوهش

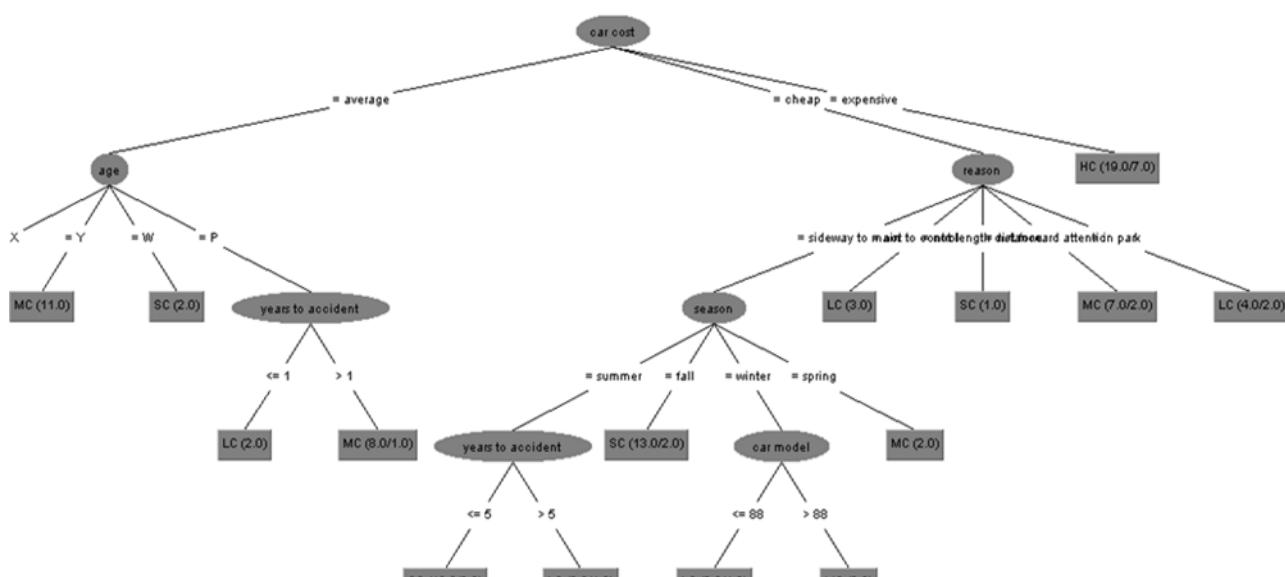
- جمع آوری داده ها از ۱۰۵ پرونده دارای خسارت بیمه بدن در یکی از مراکز پرداخت خسارت بیمه بدن در تهران، پالایش داده ها و تعیین شاخص هایی برای تعریف میزان خسارت بیمه بدن.
- اختصاص دادن ۹۵ داده جمع آوری شده به عنوان داده هایی جهت استخراج الگو ها و ۱۰ داده به عنوان داده های تست صحت الگو ها.
- استخراج الگو ها با استفاده از ۹۵ داده با تکنیک درخت تصمیم.
- اعتبار سنجی مدل با استفاده از مجموعه داده های تست صحت الگو.
- ارائه الگوی کشف شده بر اساس میزان خسارات بیمه بدن و سایر شاخص های تعیین شده.

۱-۴- داده های پژوهش

داده های این پژوهش متشکل از ۱۰۵ پرونده مربوط به افرادی است که خودرو خود را بیمه بدن کرده اند، در سال ۹۱ تصادف داشته اند و با رجوع به مراکز بیمه اقدام به دریافت خسارت خود نموده اند است. حدود ۹۰ درصد از این پرونده ها دارای رانندگانی با جنسیت مذکور بودند بنابراین این پژوهش بر روی رانندگان مرد انجام شده است. در بین این حجم داده ، تمام داده ها از کیفیت لازم برخوردار نبودند و صلاحیت ورود به مدل نهایی را نداشتند. بدین معنی که بعضی از فرم های درون پرونده ها دارای مقادیر نامناسب بودند و در مواردی بدون مقدار بودند. در نهایت داده ها از لحاظ وجود داده های نامرتبط مورد بررسی قرار گرفته است. بر این اساس تا حد امکان سعی شد تا رکوردهای معیوب طی پرسش از متصدیان بیمه یا با توجه به فیلد های مرتبط دیگر اصلاح شود ولی در نهایت بعضی از داده ها که قابلیت اصلاح نداشتند حذف گردیدند. این مرحله از داده کاوی که به بررسی داده های اولیه قبل از مدل سازی می پردازد تا از کیفیت قابل قبولی برخوردار شوند پیش پردازش نام دارد. پیش پردازش درخصوص رکوردهای ناقص صورت گرفت که به کاهش تعدادی داده و فیلد انجامید. علت کاهش فیلد ها بی ارتباط بودن برخی از آن ها مانند تاریخ تصادف با موضوع مورد پژوهش و گاه تکراری بون مقدار فیلد برای همه رکوردها مانند نوع سوت مصرفی و جنسیت راننده و... است. در نهایت جدول ۱ جهت

نام اختصاری	نام فیلد	نام اختصاری	نام فیلد	نام اختصاری	نام فیلد
car cost	قیمت خودرو قبل از تصادف	reason	علت تصادف	age	سن راننده در زمان تصادف
car model	مدل خودرو	time	زمان تصادف	season	فصل تصادف
certificate date	سال اخذ گواهینامه	color	رنگ خودرو	day to report	زمان از تصادف تا اعلام خسارت به مراکز بیمه
		harm cost	مبلغ خسارت	years to accident	فاصله بین دریافت خودرو تا وقوع تصادف

Age (سال)	جدول ۲: بارهای درنظر گرفته شده برای فیلدهای مدل داده کاوی		
	$18 \leq W < 40$	$W \geq 40$	$P \leq 40$
Car cost (میلیون تومان)	Cheap < 10	$10 \geq Average < 20$	Expensive ≤ 20
Time (ساعت)	$6 \geq morning > 12$	$12 \geq afternoon > 6$	$6 \geq night > 12$
Harm cost (میلیون تومان)	$LC < 0.5$	$0.5 \geq SC > 1$	$1 \geq MC > 10$
			$HC \leq 10$



شکل ۲: ارتباط بین مبلغ خسارت و ویژگی های خودرو و راننده

عنوانی استفاده شده برای فیلدهای مدل داده

همچنین در درخت

شکل ۴ که بیانگر ارتباط بین فصل وقوع تصادف و میزان خسارت می باشد، بهترین مدل ها دارای برگ های ۱۳ تایی هستند و به صورت زیر بیان می شوند:

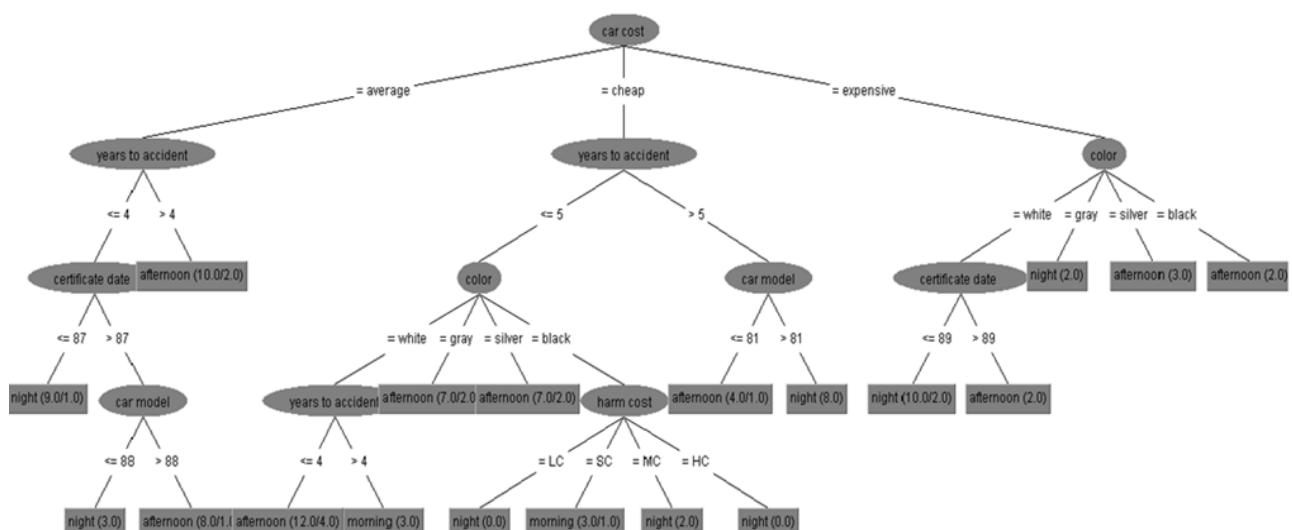
۱- اگر مبلغ خسارت ایجاد شده برای یک تصادف حدود ۱ میلیون تا ۱۰ میلیون تومان ، تاریخ اخذ گواهینامه راننده قبل از سال ۸۹ و زمان تصادف ۶ بعد از ظهر تا ۱۲ شب باشد آنگاه باید تاریخ اخذ گواهینامه در نظر گرفته شود و در صورتی که تاریخ اخذ گواهینامه قبل از سال ۸۸ باشد این تصادف در فصل تابستان اتفاق می افتد.

۲- اگر تصادف با مبلغ خسارت ایجاد شده پانصد هزار تومان تا یک میلیون تومان و با خودرویی با رنگ سفید اتفاق بیفتند آنگاه در فصل پاییز اتفاق افتاده است.

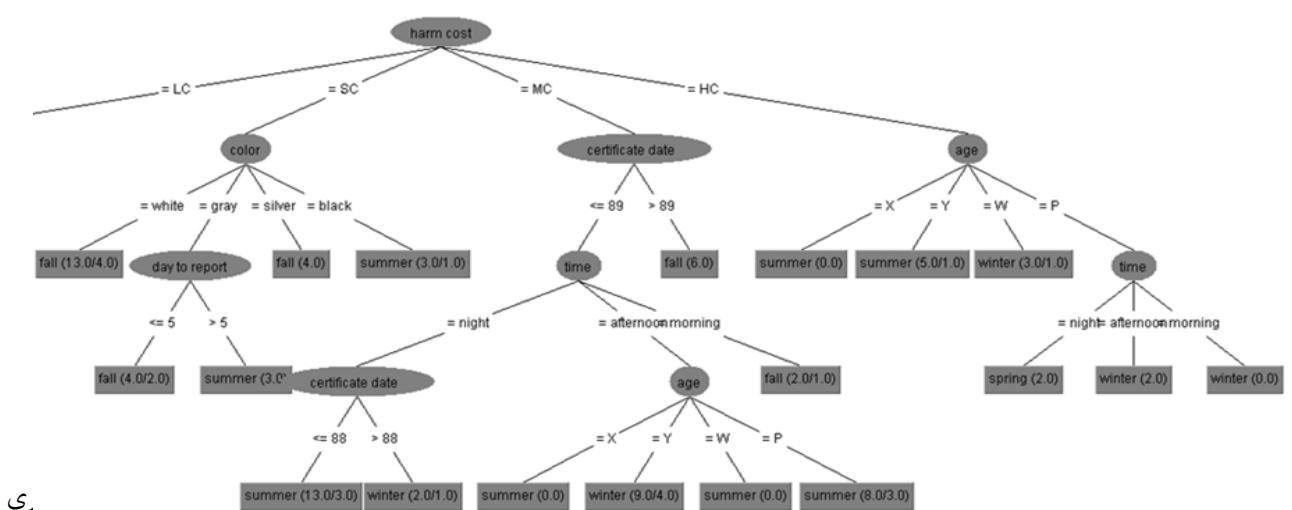
در درخت شکل ۳ که بیانگر ارتباط بین ساعت وقوع تصادف و ویژگی های خودرو و راننده می باشد، بهترین مدل ها دارای برگ های ۱۲ و ۱۰ تایی هستند و به صورت زیر بیان می شوند:

۱- اگر قیمت خودرو کمتر از ۱۰ میلیون تومان باشد و کمتر از ۵ سال از دریافت خودرو گذشته باشد و رنگ خودرو سفید باشد و اگر این تصادف در همان سال دریافت خودرو یا تا چهار سال پس از دریافت خودرو اتفاق بیفتند، این تصادف در ساعات ۱۲ ظهرالی ۶ بعد از ظهر اتفاق افتاده است.

۲- اگر قیمت خودرو بیشتر از ۲۰ میلیون تومان باشد و رنگ آن سفید باشد و راننده اش قبل از سال ۸۹ گواهینامه گرفته باشد این تصادف در ساعت ۶ بعد از ظهرالی ۱۲ شب اتفاق می افتد.



شکل ۳: ارتباط بین ساعت وقوع تصادف و ویژگی های خودرو و راننده



شکل ۴: ارتباط بین فصل وقوع تصادف و میزان خسارت. الگوهای بدست آمده انجام شد و در نتیجه



- [۱] جمشید چرایی، بیمه های اتومبیل قوانین و مقررات، روابط عمومی شرکت خدمات بیمه ایران خودرو، نوبت اول، ۱۳۸۳، صفحات ۱۰-۱۲.
- [۲] آیت کریمی، کلیات بیمه، پژوهشکده بیمه وابسته به بیمه مرکزی ایران، چاپ یازدهم : تهران ۱۳۸۷، صفحات ۲۲-۲۵.
- [۳] مهدی فیروزی، مرتضی شکوری، لیلا کاظمی، سحر زاهدی، "شناسایی تقلب در بیمه اتومبیل با استفاده از روش های داده کاوی"، پژوهشنامه بیمه(صنعت بیمه سابق)، سال بیست و ششم، شماره ۳، پائیز ۱۳۹۰، شماره مسلسل ۱۰۳، صفحات ۱۰۳ - ۱۲۸.
- [۴] نسترن حاجی حیدری، سامانند خالله، احمد فراهی، "طبقه بندی میزان ریسک بیمه گذاران بیمه بدنه خودرو با استفاده از الگوریتم های داده کاوی"، پژوهشنامه بیمه(صنعت بیمه سابق)، سال بیست و ششم، شماره ۴، رمندان ۱۳۹۰، شماره مسلسل ۱۰۴، صفحات ۱۰۷ - ۱۲۹.
- [۵] فهیمه احمدی، "علل وقوع تصادفات در معابر شهری" ، شهرداریها، سال نهم، شماره ۹۶، صفحات ۲۰-۲۵.
- [۶] مهدی غضنفری، سمیه علیزاده، بابک تیمورپور، داده کاوی و کشف دانش، دانشگاه علم و صنعت ایران، ۱۳۸۷، صفحات ۵۰-۵۲.

الگوهای ارائه شده و نتایج بدست آمده از ۷۰ درصد صحت برخوردارند. لذا نگاه کلی و بررسی اجمالی بر نتایج بدست آمده حاکی از آن است که با افزایش قیمت خودروها تصادفات با شدت بیشتر به وقوع می پیوندد و خسارت بیشتری را به بار می آورند که می تواند ناشی از گران تر بودن لوازم یدکی خودرو و یا به علت وسعت مالی بالای راننده باشد. چرا که خسارت های مالی چیزی از توان مالی او نکاسته و ارزش پولی این خسارت ها در نظر او ناچیز است. از موارد قابل توجه دیگر فصل و زمان وقوع تصادفات است، نتایج نشان می دهد که در فصل تابستان و در هنگام شب (۶ بعد از ظهر تا ۱۲ شب) تصادفات خسارت بیشتری را به بار می آورند و این می تواند ناشی از کاهش دید راننده در شب، افزایش ترافیک به دلیل تعطیلی مراکز آموزشی و گشایش مراکز تفریحی و پذیرایی این مراکز تا ۱۲ شب و استقبال مردم جهت استفاده از هواي مطبوع تابستانی باشد. اما در فصل پاییز و در ساعات ۱۲ ظهر تا ۶ بعد از ظهر تصادفات خفیف تر و با میزان خسارت کمتری رخ می دهد و این امر می تواند ناشی از کاهش ترافیک درون شهری نسبت به فصل تابستان باشد. سن راننده از دیگر عوامل ایجاد خسارت های هنگفت است، رانندگان که سنی در حدود ۲۰ تا ۳۰ سال دارند به علت کم تجربگی، بیاحتیاطی و عدم رعایت قوانین و مقررات راهنمایی و رانندگی احتمال ایجاد تصادفات شدیدتری دارند و خسارات بیشتری به بار می آورند. در همه موارد به علت معمول و متداول بودن خودرو هایی با رنگ های روشن مانند سفید احتمال ایجاد تصادفات برای این نوع از خودرو ها بیشتر است. با توجه به اینکه سالیانه شرکت های بیمه هزینه های بسیاری را در جهت پرداخت خسارت هایی با مبالغ بالا متحمل می شوند. این پژوهش می تواند مبنای علمی جهت کشف اتومبیل های دارای تصادفات سنگین پیش روی مدیران بیمه و پژوهشگران علاقمند به مطالعه بگذارد. از این رو پیشنهاد می گردد از ارائه سرویس بیمه بدنه به خودروهای گران قیمت بپرهیزند و تا حد امکان از عدم رانندگی رانندگان جوان و کم تجربه با خودرو مورد بیمه اطمینان حاصل کنند. بیمه کردن خودروهایی با قیمتی زیر بیست میلیون تومان برای شرکت های بیمه منافع بیشتری خواهد داشت.

امید است با ارائه این پژوهش و استفاده از نظرات خبرگان صنعت بیمه و متخصصان راهنمایی و رانندگی به کاهش هرچه بیشتر تصادفات و خسارت های هنگفت کمک بسیاری شود.

مراجع