تعیین عوامل مؤثر در تصادفات درون‌شهری با استفاده از روش‌های داده‌کاوی (موردمطالعه: شهر اصفهان)

غلامرضا شاه‌محمدی[1]، سعید عباسی[2]

پژوهشنامه جغرافیای انتظامی
سال ششم، شماره بیست و یکم، بهار ۱۳۹۷

از صفحه ۱۳۵ تا ۱۶۲

تاریخ دریافت: ۱۳۹۶/۰۸/۱۵
تاریخ پذیرش: ۱۳۹۶/۱۰/۲۴

چکیده

تصادفات از بزرگ‌ترین مشکلات سلامت در کشور محسوب و هرساله منجر به هزینه‌های سنگینی می‌شود. کاربرد روش‌های داده‌کاوی می‌تواند از حجم عظیم داده‌های تصادفات، الگوهای مفید را استخراج و ضمن تعیین عوامل مؤثر در بروز تصادفات، نقش قابل‌توجهی در کاهش آن‌ها داشته باشد. هدف این پژوهش بررسی عوامل مؤثر در تصادفات درون‌شهری به روش داده‌کاوی است که ازنظر هدف کاربردی و ازنظر نوع، تحلیل محتوا است. جامعه آماری پژوهش، اطلاعات پایگاه داده شهر اصفهان و حجم نمونه، اطلاعات تصادفات سال‌های (۱۳۹۰) تا (۱۳۹۴) می‌باشد. برای تحلیل اطلاعات از کشف قوانین انجمنی تکنیک سبد خرید و تولید درخت اشیاء مکرر FP-Growth استفاده شد. بر اساس نتایج، ویژگی‌های مؤثر در بروز تصادف به ترتیب، عجله و شتاب،عدم استفاده از کمربند و کلاه ایمنی، عدم توجه به جلو، نبود شانه راه و تجهیز نبودن خودرو به تجهیزات ایمنی می‌باشد. با کمک این نتایج، امکان پیشگیری و کاهش تصادفات، میسر می‌باشد.

کلید واژه‌ها: تصادفات درون‌شهری، داده‌کاوی، کشف قوانین انجمنی، الگوریتم FP-Growth.

۱- دانشیار مهندسی کامپیوتر دانشگاه علوم انتظامی امین، تهران، ایران.(نویسنده مسئول).Drshahmohammadi@chmail.ir.
۲- دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات دانشگاه علوم انتظامی امین، تهران، ایران.

Archive of SID
www.SID.ir

بیان مسئله

امروزه باوجود پیشرفت فناوری‌ها و بهبود سلامت همگانی در سطوح مختلف جامعه، بازهم می‌توان سوانح رانندگی را یکی از مهم‌ترین عوامل بروز صدمات و لطمات جبران‌ناپذیر به جامعه بشری، به‌حساب آورد. تصادفات درون‌شهری علاوه بر خسارات وارده به افراد مورد حادثه باعث افزایش ترافیک، کند شدن حمل‌ونقل و بسته شدن راه می‌شود. لذا لزوم بررسی‌ها و اقدامات لازم در جهت شناسایی عوامل مؤثر در وقوع تصادفات و به‌تبع تحلیل داده‌های تصادفات به‌عنوان منبع اطلاعاتی ارزشمند در این زمینه به‌خوبی احساس می‌شود. عوامل زیادی در بروز حوادث تصادفات دخیل هستند که در سه عنوان کلی به راه، وسیله نقلیه و عامل انسانی تقسیم می‌شوند و هرکدام از این عوامل نیز، زیر بخش‌های گوناگونی را در برمی‌گیرد (سلمانی و همکاران، ۱۳۸۷: ۳). در حالت کلی می‌توان این عوامل را به سه عامل کلی انسانی، عوامل طبیعی، و فنی و مدیریتی تقسیم‌بندی کرد؛ بنابراین برای شناسایی هر پدیده، باید علت‌های مؤثر در آن مطالعه شود (صفارزاده و همکاران، ۱۳۸۷: ۵). مدل‌های آماری در تحلیل تصادفات برای بیان روابط بین نقش تصادفات و فاکتورهای محیطی و هندسی به‌طور گسترده مورداستفاده قرار می‌گیرند. بااین‌حال مشکلات خاصی از قبیل افزایش نمایی در تعداد پارامترها (افزایش تعداد متغیرها) و عدم اعتبار آزمون‌های آماری ناشی از داده‌های پراکنده در جداول احتمالی بزرگ در تحلیل آماری سنتی روی مجموعه داده‌هایی با ابعاد بالا به وجود می‌آید. اینجا جایی است که داده‌کاوی وارد میدان می‌شود (پیتاسکی و همکاران[1]،۱۹۹۱). داده‌کاوی در این راستا می‌تواند بسیار کاربردی بوده و با توجه به افزایش روزافزون حجم داده‌ها در بانک‌های اطلاعاتی پلیس راهور در تبدیل این اطلاعات به دانش و به‌کارگیری آن به‌عنوان ابزار تصمیم‌یار مدیریتی در برنامه‌ریزی کاهش تصادفات ترافیکی، کمک شایانی به برنامه‌ریزان شهری، سیاست‌گذاران و متخصصان وادی حمل‌ونقل و نقش بی‌بدیلی را در تدوین برنامه‌ریزهای آتی شهری ایفا کند. در دهه‌های اخیر توانایی بشر برای تولید و جمع‌آوری داده‌ها به‌سرعت افزایش‌یافته است. عواملی نظیر استفاده گسترده از توانایی فناوری اطلاعات، تجهیزات آزمایشگاهی، پیشرفت در جمع‌آوری داده‌ها و سیستم‌های سنجش از راه دور ماهواره‌ای در این

1- Piatetsky et al

Archive of SID

www.SID.ir

تغییرات نقش مهمی داشته‌اند. این رشد انفجاری در داده‌های موجود، باعث پیدایش فناوری جدیدی شده تا این حجم داده را به اطلاعات و دانش تبدیل کند (تانگ[1]، ۲۰۰۸). در این میان، ابزار پشتیبانی داده‌کاوی، سازمان‌ها را قادر می‌سازد تا از سرمایه داده‌هایشان به‌درستی بهره‌برداری نمایند و از این تصمیم‌گیری استفاده کنند (شکوهیار و همکاران، ۱۳۹۴: ۴۸). نکته مهمی که بایستی در این زمینه به آن توجه شود تفاوت شهرها از جنبه‌های مختلف اقتصادی، فرهنگی و اجتماعی است. هر شهر دارای خصوصیات منحصربه‌فرد بوده که در برابر سیاست‌های مختلف مدیریت و تقاضای سفر اثرات متفاوتی می‌پذیرد. از همین رو استفاده از مدل‌های ارائه‌شده برای پیش‌بینی تصادفات و یا راه‌کارهای کاهش تصادفات و تلفات ناشی از آن در شهر یا کشوری برای دیگر شهرها و کشورها به‌طور مستقیم قابل‌استفاده نیست که دلیل آن متفاوت بودن شرایط محیطی هر منطقه با مناطق دیگر است (قاسمی نوقابی و آیتی،۱۳۹۱). بر اساس موارد ذکرشده، پژوهشگر به دنبال این مسئله است که چگونه از طریق داده‌کاوی تصادفات جمع‌آوری‌شده در فرم‌های کام (کروکی) پلیس راهنمایی و رانندگی می‌توان عوامل مؤثر در تصادفات درون‌شهری شهر موردمطالعه را شناسایی و مشخص نمود. هدف اصلی این پژوهش پیشنهاد ابزار و مدلی مناسب با استفاده از روش‌ها و الگوریتم‌های داده‌کاوی برای بررسی عوامل مؤثر در تصادفات ترافیکی است. این روش در فرآیند برنامه‌ریزی‌های کلان برای بهبود شاخص ایمنی مورداستفاده قرار می‌گیرد. با استفاده از این روش ابزار مناسبی در دست پژوهشگران قرارمی گیرد که به‌واسطه آن قادرند با بررسی و کشف رابطه میان عوامل مؤثر در تصادفات ترافیکی، در آینده سیاست‌ها و هدف‌گذاری‌های مناسب‌تری را برای کاهش تصادفات به‌کارگیرند و بتوانند نتایج حاصل از آن را به‌صورت مستمر بررسی و کنترل نمایند.

بازدهی نتایج حاصل از مطالعات تصادفات، که نشانگر زیان‌های اقتصادی و اجتماعی هنگفت است، جلوگیری از رخ دادن حتی یک تصادف را به امری مهم و مؤثر اقتصادی و اجتماعی تبدیل می‌کند و بنابراین نقش پژوهش‌های تصادفات و شناخت عوامل ایجاد آن با استفاده از ابزارهای داده‌کاوی و استفاده از نتایج آن‌ها بسیار بارز است. ازآنجاکه تمامی عوامل شامل انسان، راه و وسیله نقلیه به‌نوعی در تصادفات نقش دارند. بررسی

1- Tong

Archive of SID

www.SID.ir

آن‌ها به دلیل تعدد عوامل مؤثر و تعامل پیچیده آن‌ها کار ساده‌ای نیست (لن و لی[1]، ۲۰۰۶). تبیین و روشن شدن عوامل مؤثر بر وقوع حوادث، و ازجمله تصادفات، یکی از زیربناهای اساسی مقوله پیشگیری محسوب می‌شود. در این راستا استفاده از تحلیل داده در کشف عوامل مؤثر بر تصادفات، امری بدیع، ضروری و مطلوب در هر منطقه به شمار می‌رود. ازآنجاکه داده‌های تصادفات و ترافیک روزبه‌روز در حال افزایش است (قدیری افشار، ۱۳۹۳: ۲)؛ لذا با وسیع شدن پایگاه داده‌ای، تحلیل داده‌ها با استفاده از روش‌های آماری مرسوم بسیار مشکل می‌باشد و چه‌بسا نتایج به‌دست‌آمده خیلی قابل‌اطمینان نباشند. درحالی‌که با استفاده از ابزارهای داده‌کاوی می‌توان از میان حجم عظیمی از داده‌های تصادفات، الگوهایی مفید استخراج کرد و عوامل مؤثر در تصادفات را به‌منظور پیشگیری از وقوع تصادفات کشف و تحلیل کرد. استفاده از ابزارهای داده‌کاوی می‌تواند به‌منظور تحلیل، پیش‌بینی و شناخت عوامل مؤثر برافزایش تصادفات و اتخاذ و اجرای راه‌کارهای موردنیاز و به‌موقع در جهت کاهش تصادفات یا به حداقل رساندن هزینه‌های تصادفات خسارتی و یا جرحی و یا کاهش تلفات ناشی از آن، بسیار مفید باشد. (شکوهیار و همکاران، ۱۳۹۴: ۴۸). بنابراین هدف اصلی این پژوهش بررسی عوامل مؤثر در تصادفات درون‌شهری با استفاده از فنون و روش‌های داده‌کاوی می‌باشد.

پیشینه پژوهش

در این بخش به پژوهش‌های انجام‌شده در داخل و خارج از کشور که درزمینه استفاده از روش‌های داده‌کاوی در تحلیل تصادفات انجام‌شده است، پرداخته می‌شود.

- احمدی و علی‌محمدی طی پژوهشی در سال (۱۳۹۴)، تصادفات رانندگی بزرگراه کرج – قزوین را با استفاده از تراکم پنجره‌ای فازی مورد تحلیل زمانی و مکانی قراردادند. در آن پژوهش، تغییر الگوی زمانی و مکانی تصادفات با توجه به زمان، عامل انسانی، نوع برخورد و نوع تصادف بررسی‌شده است. نتایج نشان داده است که عوامل مختلف مانند ساعت، نوع برخورد، نوع تصادف، نقش عامل انسانی، روزهای هفته و وضعیت جوی، الگوی مکانی مشخصی دارند و با بررسی آن‌ها می‌توان علت الگوهای مختلف را کشف کرد.

1- Lan and Li

Archive of SID

www.SID.ir

- شکوهیار و همکاران طی پژوهشی در سال (۱۳۹۴)، به خوشه‌بندی داده‌های تصادفات جاده‌ای با استفاده از فنون داده‌کاوی پرداخته‌اند. نتایج نشان می‌دهد که متغیرهای جاده‌ای مؤثر در بروز تصادفات به ترتیب اهمیت شامل هندسه محل، جهت حرکت راه، نوع منطقه، نقص راه، نوع جاده، مانع دید، نوع خط‌کشی، نوع رویه راه و بی‌توجهی به جلو هستند. نتایج به‌دست‌آمده از الگوریتم درخت دسته‌بندی و رگرسیون نشان داد که سهم هر متغیر در تصادفات منجر به جرح، فوتی و خسارتی به چه میزان است که مهم‌ترین عامل، نوع هندسه راه می‌باشد.
- هوشیار و شریفی طی پژوهشی در سال (۱۳۹۵)، تصادفات درون‌شهری شهر ارومیه در سال (۱۳۹۲) را مورد تحلیل فضایی قرار داده‌اند. نتایج نشان می‌دهد بیشترین تعداد تصادفات در مناطق (۱) و (۳) و اطراف بخش مرکزی شهر ارومیه رخ‌داده است و توزیع نقاط تصادفی برای شهر ارومیه به‌صورت خوشه‌ای بوده است.

مبانی نظری پژوهش

داده‌کاوی داده‌های مستخرج از پایگاه‌های اطلاعاتی یکی از موارد مهم در فرآیند تحلیل داده‌ها است که می‌تواند برای استخراج مدل‌ها و کلاس‌بندی داده‌ها مورداستفاده قرار گیرد. در این بخش مبانی نظری پژوهش شامل داده‌کاوی، روش‌های مختلف داده‌کاوی، قوانین انجمنی و الگوهای مکرر، تشریح می‌شوند.

داده‌کاوی

در دهه‌های اخیر حجم داده‌های ذخیره‌شده در پایگاه داده یا هر منبع دیگری افزایش چشم‌گیری داشته است و روبه افزایش است، به‌طوری‌که تحلیل این داده‌ها بدون استفاده از فناوری جدید و ابزارهای خودکار بسیار دشوار و غیرممکن شده است. فنّاوری‌های نوین اطلاعاتی و ارتباطی و همچنین فنّاوری‌های پشتیبان تصمیم، با جمع‌آوری، ذخیره، ارزیابی، تفسیر و تحلیل، بازیابی و اشاعه اطلاعات و دانش برای کاربران خاص، می‌توانند در اطلاع‌یابی به‌موقع، صحیح و موردنیاز به افراد تأثیر زیادی داشته‌باشند. یکی از ابزارهای مورداستفاده در این فناوری‌ها، داده‌کاوی است. داده‌کاوی

Archive of SID

www.SID.ir

شامل استفاده از ابزارهای پیشرفته تحلیل داده به‌منظور کشف الگوهای معتبر، از قبل ناشناخته و روابط در مجموعه داده‌های بزرگ است (اسماعیلی، ۱۳۹۴: ۱۵). داده‌کاوی ترکیبی از تکنیک‌های یادگیری ماشین، تشخیص الگو، آمار، تئوری پایگاه داده، خلاصه کردن و ارتباط بین مفاهیم و الگوهای جالب به‌صورت خودکار از پایگاه داده شرکت‌های بزرگ است. هدف اصلی داده‌کاوی کمک به فرآیند تصمیم‌گیری از طریق استخراج دانش از داده‌هاست (الپایدین[1]، ۲۰۱۰). داده‌کاوی فرآیند کشف قوانین و دانش ناشناخته و مفید از انبوه داده‌ها و پایگاه داده است. انجام عملیات داده‌کاوی شامل موارد ذیل می‌باشد (لین و یه[2]،۲۰۱۲).

۱) جداسازی داده مفید از داده بیگانه؛

۲) یکپارچه‌سازی داده‌های مختلف تحت قالب واحد؛

۳) انتخاب داده لازم از میان دیگر داده‌ها؛

۴) انتقال داده به محیط داده‌کاوی جهت اکتشاف قوانین؛

۵) ایجاد مدل‌ها و الگوهای مرتبط به‌وسیله روش‌های داده‌کاوی؛

۶) ارزیابی مدل و الگوهای ایجادشده جهت تشخیص مفید بودن آن‌ها؛

۷) انتشار دانش استخراج‌شده به کاربران نهایی.

روش‌های داده‌کاوی

روش‌های مختلف داده‌کاوی را می‌توان بر اساس نوع عملیاتی که انجام می‌دهند به دودسته «پیش‌بینی کننده» و «تشریح کننده» تقسیم کرد. روش‌های پیش‌بینی کننده با ساخت مدلی برای پایگاه داده، وظیفه پیش‌بینی موارد ناشناخته را بر اساس تعریف یک ناظر یا متغیر وابسته بر عهده دارند. درحالی‌که روش‌های تشریح کننده، الگوهایی قابل‌فهم از داده‌ها را برای انسان کشف می‌کنند. با توجه به تعریف مسئله در این پژوهش یکی از دلایل انتخاب قوانین انجمنی[3] کاوش در توصیف الگوهای قابل‌فهم در

1- Alpaydin
2- Lin and Yeh
3- Assocciation Rules

Archive of SID

www.SID.ir

داده‌های واقعی تصادف در منطقه موردمطالعه می‌باشد. قوانین انجمنی شرایطی را نشان می‌دهند که در یک مجموعه داده، پی‌درپی باهم اتفاق می‌افتند. قوانین استخراج‌شده در حقیقت حضور برخی ویژگی‌ها را بر اساس سایر آن‌ها شرح می‌دهند (هور[1]،۲۰۱۰). پس برای شناخت عوامل مؤثر در تصادفات از داده‌های موجود، شناسایی مجموعه‌ای از ویژگی‌هایی که باهم مکرر در داده‌ها رخ می‌دهند و بیانگر رخداد توأمان آن‌هاست کارگشا خواهد بود؛ یعنی شناخت الگوهای پرتکرار و یافتن نظم ذاتی داده‌ها که از اهداف اصلی محقق است. بنابراین قوانین انجمنی ازجمله روش‌های مرتبط و مناسب توصیف این تشخیص مکرر محسوب می‌شود.

قوانین انجمنی و الگوهای مکرر

برای درک قوانین انجمنی و الگوهای مکرر نیاز به یادآوری برخی مفاهیم مشهود است که در اینجا مهم‌ترین آن‌ها آورده شده است.۱). مجموع اشیاء k [2]: مجموعه اقلام شامل k قلم داده را مجموعه اشیاء k تایی می‌نامند.۲). تعداد پشتیبانی [3]: فراوانی وقوع مجموعه اشیاء در تراکنش‌های موجود که آن را با حرف σ نشان می.دهیم.۳) مجموعه اشیاء مکرر[4]: مجموعه‌ای از اشیاء که تعداد پشتیبانی آن‌ها بزرگ‌تر یا مساوی مقدار آستانه[5] باشد، مجموعه اشیاء مکرر نامیده می‌شود. ۴). قوانین انجمنی کاوش در راستای کشف ارتباطات جالب و بااهمیت بین اقلام اطلاعاتی در پایگاه داده‌های بزرگ و انبارهای تراکنش می‌باشد (صنیعی آباده و همکاران، ۱۳۹۴: ۱۵۹). داده‌کاوی انجمنی و استخراج قوانین انجمنی از مجموعه داده‌ها اولین بار توسط آگراوال برای کشف دانش و الگوهای خرید کاربران یک فروشگاه ارائه شد. در طرح پیشنهادی آگراوال، سبد خرید کل کاربران در مدت‌زمانی خاص بررسی و سپس قوانین انجمنی از آن استخراج می‌شد؛ عادات و رفتار خرید مشتریان مورد تحلیل قرار می‌گرفت و الگوهای موجود در اقلام خریداری‌شده کشف می‌شدند (آگراوال[6]، ۱۹۹۳). قانون انجمنی با عبارت $X \rightarrow Y$ بیان می‌شود که در آن X و Y مجموعه اقلام غیر تهی هستند که هیچ‌گونه اشتراکی ندارند .

1- Hur
2- Items Set
3- Support Count
4- Frequent Item Set
5- Min Support Threshold
6- Agrawal

Archive of SID

www.SID.ir

($X\cap Y=\emptyset$) دو معیار پشتیبان و اطمینان به‌منظور ارزیابی قوانین انجمنی استفاده می‌شوند، هرچند معیارها فقط به این دو ختم نمی‌شوند. مقدار پشتیبان نشان می‌دهد که در چند درصد از تراکنش‌های پایگاه داده‌ها می‌توان مجموعه اقلام X و Y را همراه یکدیگر پیدا کرد و مقدار اطمینان[1] درمیان تراکنش‌هایی که مجموعه اقلام X را در خود دارند، به دنبال مجموعه اقلام Y می‌شود .این نکته ساده را فراموش نکنید که تراکنش‌های حاوی X می‌تواند شامل Y نباشند و بالعکس. به همین علت، دو قانون انجمنی $X\rightarrow Y$ و $Y\rightarrow X$ یکسان نیستند. از زاویه دید تئوری احتمالات برای پشتیبان و اطمینان داریم:

$$\text{Support } (X\rightarrow Y) = P(X\cap Y) \qquad \text{رابطه (۱)}$$

$$\text{Confidence } (X\rightarrow Y) = P(Y|X) = P(X\cap Y)/P(X) \qquad \text{رابطه (۲)}$$

با توجه به مجموعه‌ای از تراکنش‌ها مانند T، هدف از کاوش قوانین انجمنی یافتن تمام قوانین R_x است که از این دستورات تبعیت می‌کنند. Support $(R_x)\geq$Suppmin و Confidence $(R_x \geq$Confmin) در این دستورات منظور از (Confmin, Suppmin) به ترتیب عبارت است از کمترین مقدار برای پشتیبان[2] و کمترین مقدار اطمینان [3] که می‌بایست جهت قبول هر پاسخ نهایی به‌عنوان یک قانون باارزش موردتوجه قرار گیرد. (اسماعیلی، ۱۳۹۲: ۱۲۲).

تولید الگوهای مکرر: تولید مجموعه اقلام مکرر مرحله مهمی از کشف قوانین انجمنی را تشکیل می هد. مجموعه داده‌های که شامل k قلم داده باشد می‌تواند تعداد ۱- ۲^k الگو تولید کند. که هریک از آن‌ها می‌تواند یک الگوی مکرر باشد. با افزایش مقدار k فضای جستجو نیز به‌صورت نمایی رشد می‌کند. راهکاری که از هیچ قانون میانبر و یا مکاشفه‌ای استفاده نمی‌کند، نمی‌تواند در زمان قابل قبولی کار خود را در این فضای جستجو به پایان برساند. زیرا پس از تولید این ۱- ۲^k مجموعه اقلام برای محاسبه مقدار پشتیبان هریک از آن‌ها، پایگاه داده پیمایش و برای هر سطر آن وجود مجموعه اقلام بررسی می‌شوند. بدین ترتیب اگر تعداد تراکنش‌ها N باشد و به‌طور میانگین هر تراکنش دارای فقط تعداد A

1- Confidence
2- Suppmin
3- Confmin

Archive of SID

www.SID.ir

داده باشد الگوریتم با تعداد (۴ -1)*N* ۲^{k})) مقایسه روبرو خواهد بود. بنابراین نیاز به روش‌هایی که از حجم این محاسبات بکاهد به‌وضوح احساس می‌شود (اسماعیلی،۱۳۹۲: ۱۲۵).

دو الگوریـتم مهم و معروف برای یافتن قوانین انجمنی الگوریتم اولیه آپریوری[1] و Fp-Growth هستند. مهم‌ترین هدف الگوریتم آپریوری تعیین مجموعه اشیا مکرری است که تعداد تکرار آن‌ها حداقل برابر حد آستانه پشتیبان تعیین‌شده باشد. مزیت مهم این الگوریتم در مقابل دیگر الگوریتم‌های تولید مجموعه اشیاء مکرر سادگی پیاده‌سازی آن است. البته ایراد مهم آن نیز نگهداری تراکنش‌های ورودی در حافظه اصلی و نیز پیمایش کل این تراکنش‌ها در هر دور اجرای الگوریتم می‌باشد. الگوریتم FP-Growth در مقایسه با الگوریتم آپریوری روش کارآمدتری ارائه می‌کند. این الگوریتم با ساخت یک درخت بانام FP-Tree سرعت فرآیند تولید اشیاء مکرر را به‌طور چشمگیری افزایش می‌دهد. دلیل افزایش سرعت تولید اشیاء مکرر در FP-Growth ، کاهش قابل‌توجه تعداد دفعات مراجعه به تراکنش‌های مسئله مورد کاوش است. درواقع با یک‌بار مراجعه به مجموعه تراکنش‌های مسئله، درخت FP-Tree ساخته می‌شود (صنیعی آباده و همکاران، ۱۳۹۴: ۱۶۹). همچنین در کشف قوانین انجمنی یک الگو، تاکنون الگوریتم‌های بسیاری در جهت کاوش قواعد انجمنی ارائه‌شده که FP-Growth یکی از موفق‌ترین آن‌هاست. این الگوریتم جزو شناخته‌ترین الگوریتم‌های داده‌کاوی جهت کاوش اشیاء مکرر می‌باشد و طی دو بار پیمایش درون پایگاه داده، اشیاء مکرر را استخراج می‌کند (دینگ و یو[2]، ۲۰۰۹).

روش پژوهش

این پژوهش ازنظر هدف کاربردی و ازنظر نوع، تحلیل محتوا می‌باشد. جامعه آماری پژوهش اطلاعات پایگاه داده تصادفات شهر اصفهان می‌باشد. جامعه نمونه، اطلاعات کروکی‌های تصادفات در سال‌های (۱۳۹۰) تا (۱۳۹۴) می‌باشد. در جامعه نمونه به‌صورت تمام شمار عمل شده است. پایه اصلی پژوهش حاضر بر کشف دانش از پایگاه

1- Appriori
2- Ding and Yau

Archive of SID

www.SID.ir

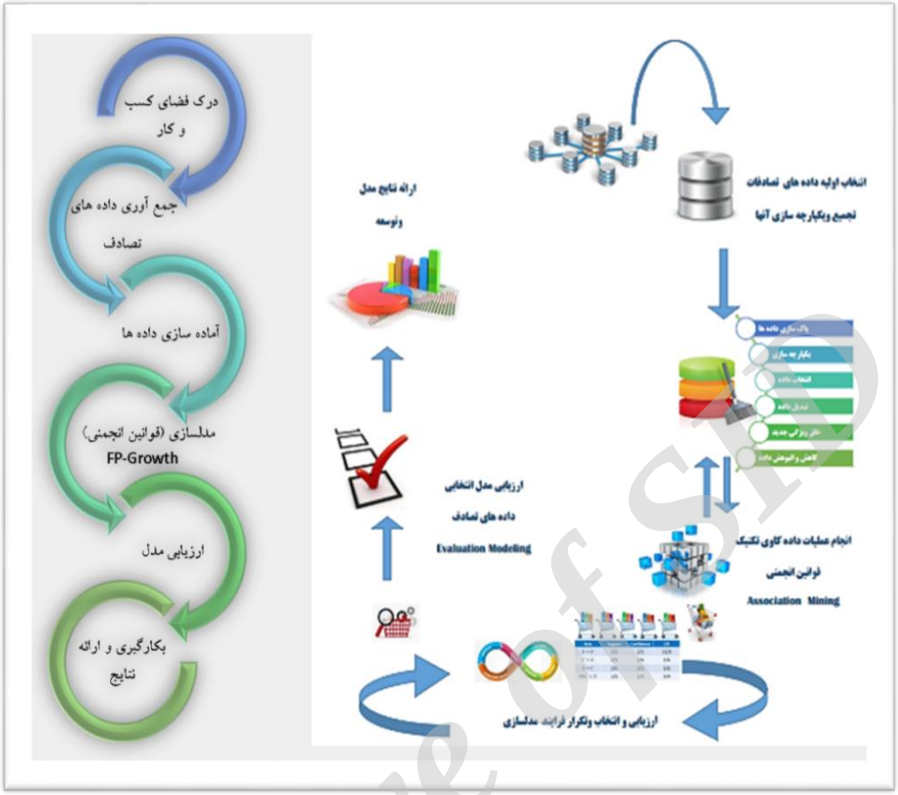
داده سازمان موردمطالعه نهاده شده است؛ ازاین‌رو استاندارد جهانی کریسپ[1] که از متدولوژی‌های بسیار قوی در روش‌های داده‌کاوی برخوردار است و شامل گام‌های شناخت سیستم و جمع‌آوری داده، آماده‌سازی داده‌ها، مدل‌سازی و ارزیابی می‌باشد، به‌منظور انجام فرآیند پژوهش، مورداستفاده قرارگرفته است. شیوه تجزیه‌وتحلیل اطلاعات رویکرد داده‌کاوی از طریق کشف قوانین انجمنی تکنیک سبد خرید و تولید درخت اشیاء مکرر FP-Growth ، برای شناخت الگوهای مکرر تصادف در چهار عامل اصلی اجزاء ترافیک (عامل انسان، وسیله نقلیه، راه و شرایط اقلیمی) و بررسی و تأثیر توأمان این ویژگی‌ها در تصادفات می‌باشد.

در این پژوهش برای تعیین متغیرهای تأثیرگذار در تصادفات و شناخت نوع تصادفات و همچنین تکنیک‌های داده‌کاوی، داده‌های جمع‌آوری‌شده و بعدازآن برای اینکه الگوهای موردنظر در شناسایی عوامل مؤثر در تصادفات توسط ابزارهای مختلف داده‌کاوی استخراج شود، باید داده‌ها پیش‌پردازش شده و به شکل‌های مناسب برای استفاده در نرم‌افزارهای داده‌کاوی درآید. در مرحله بعد با فنون داده‌کاوی ویژگی‌های مکرر در انواع تصادفات شناسایی و نقش این ویژگی‌ها در انواع تصادفات بررسی و از طریق آن قوانین معتبری استخراج و تجزیه‌وتحلیل می‌شود. در این پژوهش بر پایه متدولوژی داده‌کاوی و مجموعه داده‌ها با نرم‌افزار اراکل، اکسس و اکسل آماده‌سازی اولیه شده و استخراج قواعد با استفاده از نرم‌افزار رپیدماینر[2] و پیاده‌سازی فنون داده‌کاوی انجام خواهد شد. شکل شماره (۱)، مدل کشف عوامل مؤثر بر تصادف با استفاده از قوانین انجمنی تکنیک Fp-Growth را نشان می‌دهد.

1- Cross Industry Standard Process for Data Mining (CRISP-DM)
2- RapidMiner

Archive of SID

www.SID.ir



شکل شماره (۱). مدل کشف عوامل مؤثر بر تصادف با استفاده از قوانین انجمنی تکنیک Fp-Growth .

مرحله تعریف مسئله و آماده‌سازی داده‌ها

تعریف مسئله کسب‌وکار: انجام این پژوهش ارتقا سطح دانش برای شناخت و بررسی ارتباط عوامل مؤثر در وقوع تصادفات و کاربرد آن در حوزه پلیس و سایر سازمان‌های حمل‌ونقل و بیمه است و نتایج حاصل از آن به‌عنوان منبعی برای اتخاذ تصمیماتی در جهت کاهش میزان تصادفات درون‌شهری بکار می‌رود. این پژوهش داده‌های حقیقی جمع‌آوری‌شده از صحنه تصادفات را با استفاده از قوانین کشف‌شده از روش داده‌کاوی کشف قوانین توصیف و تفسیر می‌کند و به شرایط و روابط موجود، علل و عوامل تصادفات می‌پردازد.

Archive of SID

www.SID.ir

جمع‌آوری داده‌ها: داده‌های تصادفات با استفاده از فرم کروکی کام (۱۱۴) استخراج می‌شود که مرجع رسمی عمومی تمام تصادفات ایران است و به‌وسیله افسران کارشناس تصادف، اطلاعات موردنظر تنظیم و ثبت می‌شوند.

پایگاه داده: پایگاه داده کروکی‌های تصادفات شامل چهار دسته اطلاعات: ۱- مشخصات رویدادها و تصادفات شامل تاریخ تصادف، نوع تصادف، وضعیت روشنایی، وضعیت هوا، وضعیت سانحه؛ ۲- مشخصات وسیله نقلیه مقصر/غیرمقصر شامل نوع وسیله نقلیه، نوع سیستم، علت تامه تصادف، نوع بار، نوع پلاک، محل برخورد، نوع گواهینامه، سن و جنسیت راننده و شغل و غیره؛ ۳- مشخصات سرنشینان تصادف شامل سن، جنسیت، وضعیت صدمه، تحصیلات و شغل، ایمنی و غیره؛ ۴- مشخصات راه و نوع کاربری منطقه تصادف و... می‌باشد. با توجه به اهداف این پژوهش، از میان داده‌های موجود ویژگی‌های مرتبط با عوامل راه؛ انسان (راننده مقصر)، وسیله نقلیه و شرایط اقلیمی از کروکی‌های تصادفات سال‌های (۱۳۹۰) تا (۱۳۹۴)، استخراج و پس از یکپارچه‌سازی داده‌ها و جداول مرتبط، پایگاه داده نمونه با تعداد (۱۱۵۰۰۰) رکورد تشکیل شد.

آماده‌سازی داده‌ها: مجموعه عملیاتی که منجر به تولید مجموعه‌ای از داده‌های پالایش‌شده قابل کاوش خواهد شد، آماده‌سازی داده نامیده می‌شود (صنیعی آباده و همکاران، ۱۳۹۴: ۳۱۶). در طی این عملیات، که شامل پاک‌سازی داده، تبدیل و یکپارچه‌سازی داده و کاهش بُعد آن می‌باشد، داده‌های متناقض[1] ، مفقود[2] و مغشوش[3] پالایش می‌شوند (جعفری، ۱۳۸۸). پیش‌پردازش داده‌ها شامل (۹) گونه عملیات است که برخی از آن‌ها عبارت‌اند از: پاک‌سازی، انتخاب زیرمجموعه ویژگی‌ها ، کاهش ابعاد و... (صنیعی آباده و همکاران، ۱۳۹۴: ۳۲۵).

ادغام و پاک‌سازی داده‌ها: داده‌ها ممکن است دارای خطا[4]، مقادیر ازدست‌رفته[5]، مقادیر پرت و دورافتاده[6] باشند (هان و کمبر[1]،۲۰۱۰). در این پژوهش داده‌های

1- Inconsistent
2- Missing
3- Noisy
4- Noisy Data
5- Missing Values
6- Outliers

Archive of SID

www.SID.ir

ازدست‌رفته در برخی ویژگی‌ها از طریق نرم‌افزار رپیدماینر و به روش تخمین یا همان پر کردن مقادیر گم‌شده با مقدار میانگین داده‌ها و یا شبیه‌ترین داده‌ها با توجه دیگر ویژگی‌ها اصلاح و یا جایگزین شد.

انتخاب زیرمجموعه ویژگی‌ها: زمانی که قصد بررسی نظم بخش خاصی از مجموعه داده مطرح است، از این روش استفاده می‌کنیم. در این پژوهش، یکپارچه‌سازی جداول مربوطه؛ (۴۴) ویژگی در چهار نوع داده (داده‌های عوامل راه، انسان، وسیله نقلیه و محیط) وجود دارد که اگر بخواهیم نظم حاکم بر روی تنها (۲۰) ویژگی از آن‌ها بررسی شود، باید با انتخاب ویژگی‌های موردنظر عوامل راه -هندسه محل، نوع راه، خط‌کشی، وضعیت علائم ترافیکی و غیره - به بررسی و درک داده‌های این عوامل به‌صورت متمرکز بپردازیم.

فیلتر کردن نمونه‌ها: عملیات انتخاب زیرمجموعه‌ای از رکوردها برای کاوش، فیلتر کردن نمونه‌ها نامیده می‌شود. در بخش قبل، عملیات انتخاب زیرمجموعه ویژگی مشخص شد که می‌خواهیم بر روی چه فیلدها یا ستون‌هایی از جدول عملیات ادامه یابد. ولی در عملیات فیلتر کردن، نمونه‌ها سطرهای ویژه که می‌خواهیم روی آن‌ها کارکنیم انتخاب می‌شوند (صنیعی آباده و همکاران، ۱۳۹۴: ۵۷). برای فیلتر کردن به‌طور مثال نوع تصادفات خسارتی از دیگر انواع تصادف و یا فیلتر کردن ویژگی‌هایی که با مقادیر نامشخص یا سایر پرشده‌اند را از داده‌ها جدا کرده تا با استفاده از دیگر ویژگی‌ها به تصحیح آن‌ها پرداخته شود.

نمونه‌برداری: از روش‌های اصلی برای انتخاب داده‌ها است و در طی آن، از میان داده‌های موجود در مجموعه داده‌ها، با توجه به‌اندازه مجموعه داده نمونه تعدادی از داده‌ها انتخاب می‌شوند. مثل نمونه‌برداری از رکوردهای تصادف با استفاده از نوع تصادف که ازلحاظ تعداد تصادف خسارتی، جرحی و فوتی یکسان نیست.

1- Han & Kamber

Archive of SID

www.SID.ir

تبدیل داده: در عملیات تبدیل داده‌ها از تابعی استفاده می‌شود که مجموعه کل مقادیر یک ویژگی مفروض را به یک مجموعه جدید از مقادیر نگاشت می‌کند. این کار به‌گونه‌ای صورت می‌پذیرد که هریک از مقادیر قدیمی بتوانند با یکی از مقادیر جدید شناسایی شوند. (صنیعی آباده و همکاران، ۱۳۹۴: ۳۳۷). اگر درمیان داده‌های شما داده از نوع اسمی باشد می‌توان با تبدیل داده اسمی به عددی کار را دنبال نمود (اسماعیلی، ۱۳۹۲: ۷۳).

گسسته سازی: هدف عملیات گسسته سازی داده، تبدیل نوع ویژگی‌های بازه‌ای و نرخی (عددی) به نوع اسمی است. با گسسته سازی داده، سختی مسئله را کاهش داده و زمینه لازم را برای عملکرد مؤثرتر الگوریتم‌های یادگیری مدل فراهم می‌شود.

کاهش ابعاد: به‌طورکلی هرقدر ابعاد یا تعداد ویژگی‌های مسئله مورد کاوش بالا رود، پراکندگی رکوردها در فضای جستجو بیشتر می‌شود. یکی از روش‌هایی که برای کاهش ابعاد معرفی شد روش انتخاب زیرمجموعه‌ای ویژگی‌ها بود. البته در این روش ویژگی‌هایی که ارزش اطلاعاتی آن‌ها برای کاوش پایین است، برای حذف شدن انتخاب خواهند شد.

خلق ویژگی: به‌منظور افزودن ویژگی‌های جدید در کنار ویژگی‌های موجود در یک مجموعه داده برای نمایش مؤثرتر و کامل‌تر ویژگی‌های اولیه،، خلق ویژگی می‌گویند (اسماعیلی، ۱۳۹۲: ۶۲). برای خلق ویژگی‌های جدید موارد، تبدیل ویژگی تاریخ میلادی تصادف و خلق ساعت تصادف و گسسته سازی ویژگی ساعت و ایجاد دوره (۴) ساعته تصادف؛ تفکیک خودروهای مقصر (بر اساس سبک، نیمه سنگین، سنگین و موتورسیکلت و دوچرخه، با استفاده از کاربری خودروها و نوع سیستم خودروها) و ویژگی وضعیت گواهینامه (با استفاده از فیلد تاریخ گواهینامه و تاریخ شمسی تصادف، نوع گواهینامه‌ها ، مدت اعتبار و همچنین نوع خودروی مقصر در حادثه)، تفکیک و افزوده‌شده است.

انتخاب نهایی ویژگی‌ها: بر اساس مشورت با خبرگان از بین ویژگی‌های موجود، پس از مراحل پاک‌سازی داده‌ها، ویژگی‌های مشخص‌شده در جدول شماره (۲)، انتخاب شد.

Archive of SID

www.SID.ir

مرحله مدل‌سازی

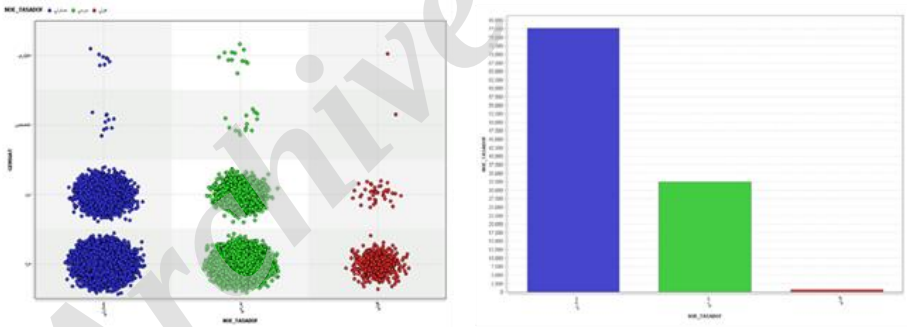
در این مرحله، انواع فنون مدل‌سازی، انتخاب و استفاده می‌شود. معیارهای بسیاری برای انتخاب یک الگوریتم مناسب وجود دارند، اما دانستن ماهیت داده‌ها در انتخاب تکنیک مدل‌سازی، اهمیت زیادی دارد (اسماعیلی، ۱۳۹۲: ۳۰).

خلق ویژگی وضعیت مجاز بودن گواهینامه: با استفاده از فیلد تاریخ گواهینامه و تاریخ شمسی تصادف می‌توان وضعیت گواهینامه را با توجه به نوع و مدت اعتبار آن و همچنین نوع خودروی مقصر در حادثه تفکیک نمود (وضعیت مجاز بودن و غیرمجاز بودن، تاریخ منقضی و فاقد گواهینامه) و ویژگی جدید وضعیت گواهینامه را مطابق جدول شماره (۱)، خلق نمود.

جدول شماره (۱). خلق ویژگی وضعیت گواهینامه رانندگان مقصر

مجاز	غیرمجاز	فاقد گواهینامه	مجاز تاریخ منقضی
۱	۲	۳	۴

شکل شماره (۲). مصورسازی توزیع داده‌ها بر اساس نوع تصادف، جنسیت



جدول شماره (۲). ویژگی‌های منتخب از پاک‌سازی داده‌ها و انتخاب ویژگی‌های مؤثر در تصادفات

ردیف		نام ویژگی/فیلد	توضیحات/ مشخصات رویداد تصادفات
۱	ویژگی‌های عوامل راه	نقایص مؤثر راه	دارد، نقص علائم عمودی، نقص علائم افقی، کم بودن عرض راه، وجود مانع و دست‌انداز، فقدان شانه خاکی و پارکینگ، اختلاف بین سطح آسفالت و شانه، فقدان حفاظ ایمنی کنار راه، غیر استاندار بودن حفاظ کنار راه، نشست

Archive of SID

www.SID.ir

ردیف		نام ویژگی/فیلد	توضیحات/ مشخصات رویداد تصادفات
			جاده‌ای، نقص رویه آسفالت، قوس یا زاویه تند، شیب عرض و طولی غیراستاندارد، نقص روشنایی راه، نقص خط کش راه، سایر موارد)
۲		تعمیرات راه	در حال تعمیر نیست، در حال تعمیر بدون علائم کافی، در حال تعمیر با علائم کافی
۳		وضع هندسه محل	مستقیم مسطح، مستقیم سربالایی، پیچ مسطح، پیچ سربالایی، پیچ سرپایینی، تونل، پل
۴		نوع رویه راه	آسفالت، شنی، خاکی، نامشخص
۵		نوع راه	آزادراه، بزرگراه، بلوار، خیابان اصلی، خیابان فرعی، کوچه، میدان، تقاطع(سه‌راهی، چهارراه، هم‌سطح و غیر هم‌سطح)
۶		کاربری محل	مسکونی، اداری، تجاری، صنعتی، کشاورزی، آموزشی، سایر
۷		نوع شانه راه	شانه آسفالته، شانه خاکی، شانه ندارد، نامشخص
۸		سمت حرکت راه	یک‌طرفه، دوطرفه، دوطرفه مجزا با جداکننده فیزیکی
۹		نوع منطقه	هموار، تپه‌ماهور، کوهستان، نامشخص
۱۰		خط کشی	ندارد، مقطع، ممتد، دوبل
۱۱	ویژگی‌های عامل انسانی تصادف	عامل انسانی مؤثر در تصادف	ندارد، خستگی و خواب‌آلودگی، نقص عضو مؤثر، ضعف ناشی از کهولت سن، استعمال موادمخدر، مصرف مشروبات الکلی، مهار نکردن محموله به طرز صحیح، عجله و شتاب بی‌مورد، عدم‌تشخیص سهم عبور سایرین، عدم آشنایی با جاده، سایر عوامل
۱۲		تحصیلات	نامشخص، بی‌سواد، خواندن و نوشتن، ابتدائی، راهنمایی، سیکل، دبیرستان، دیپلم و دانشجو، کارشناس، فوق‌دیپلم و حوزوی
۱۳		سن راننده مقصر	کمتر از ۱۸ سال،۱۸ تا ۲۸ سال،۲۸ تا ۳۸ سال،۳۸ تا ۴۸ سال،۴۸ تا ۵۸ سال،۵۸ تا ۶۸ سال و بزرگ‌تر از ۶۸ سال
۱۴		جنسیت	مرد، زن
۱۵		شغل	نامشخص، راننده، کارمند/کارگر، نظامی، آزاد، دانش‌آموز، دانشجو
۱۶		کمربند/کلاه	نامشخص، استفاده کرده، استفاده نکرده
۱۷		صدمه درصحنه	صدمه ندیده، مصدوم، متوفی
۱۸		نوع گواهینامه	ب ۱ جدیدالصدور، ب ۱ موقت، ب ۱، پایه ۳، ب ۲، پایه دوم، پایه یکم، ویژه، نامشخص
۱۹		وضعیت گواهینامه	فاقد گواهینامه، مجاز، غیرمجاز، تاریخ منقضی
۲۰	ویژگی‌های عامل وسیله نقلیه	نوع وسیله نقلیه	دوچرخه، موتورسیکلت، سواری، وانت‌بار، کامیون، اتوبوس، کامیونت، تریلر، کشنده، ادوات عمرانی، ادوات کشاورزی، آمبولانس و غیره
۲۱		مانور وسیله نقلیه	نامشخص، حرکت به جلو، گردش‌به‌چپ، گردش‌به‌راست، توقف در خارج راه و غیره
۲۲		عامل وسیله نقلیه مؤثر در تصادف	ندارد، نقص سیستم روشنایی، نقص سیستم ترمز، نقص سیستم فرمان، صاف بودن لاستیک، فقدان زنجیر چرخ در موقع ضروری، فقدان برف‌پاک‌کن در موقع ضروری، نقص دستگاه تعلیق، صاف بودن لاستیک، سایر نقایص

Archive of SID

www.SID.ir

ردیف		نام ویژگی/فیلد	توضیحات/ مشخصات رویداد تصادفات
۲۳		نوع خودرو	سبک، سنگین، نیمه سنگین، موتورسیکلت/دوچرخه
۲۴		تجهیزات ایمنی	تجهیزات ایمنی ندارد. ترمز ABS کیسه هوا. هم ترمز ABS و هم‌کیسه هوا
۲۵		کاربری خودرو	دوچرخه، موتورسیکلت، سواری، وانت‌بار، کامیون، اتوبوس، کامیونت، تریلر، کشنده، ادوات عمرانی، ادوات کشاورزی، آمبولانس و غیره
۲۶		نوع برخورد	وسیله نقلیه با وسیله نقلیه، وسیله نقلیه با موتورسیکلت، وسیله نقلیه با عابر، موتورسیکلت با موتورسیکلت، موتورسیکلت با عابر، چند برخوردی، موتورسیکلت با دوچرخه
۲۷	شرایط عمومی و اقلیمی	شرایط روشنایی	۱. روز ۲. طلوع ۳. غروب ۴. شب
۲۸		مانع دید	ندارد، درخت/بوته، ساختمان/کیوسک، تل خاک و مشابه آن، وسیله متوقف، وسیله در حال حرکت، نور خورشید، شیب، قوس قائم، مه/دود، کولاک، طوفان شن، تابلو، نور چراغ وسیله نقلیه مقابل، یخزدگی شیشه وسیله نقلیه، سایر
۲۹		وضع هوا	ابری، بارانی، صاف ، مه آلو ، برفی ، طوفانی ، غبارآلود
۳۰		شرایط سطح راه	خشک معمولی، مرطوب و خیس، یخبندان و برفی، شنی و خاکی، گل‌آلود، روغنی و کثیف، آب‌گرفتگی، قیر زدگی، سایر
۳۱		ایام هفته تصادف	شنبه ، یکشنبه ، دوشنبه ، سه‌شنبه ، چهارشنبه ، پنج‌شنبه ، جمعه
۳۲		ماه تصادف	فروردین ، اردیبهشت ، خرداد ، تیر ،مرداد ، شهریور، مهرماه ، آبان ، آذر ،دی ، بهمن ، اسفند
۳۳		نوع تصادف	خسارتی ، جرحی ، فوتی
۳۴		دوره زمانی تصادف	۰-۴،۴-۸،۸-۱۲،۱۲-۱۶،۱۶-۲۰،۲۰-۲۴
۳۵		علت تامه تصادف	عدم رعایت فاصله طولی ، عدم رعایت فاصله عرضی ، عدم رعایت حق تقدم ، عدم توجه به جلو ، عدم توانایی در کنترل وسیله نقلیه ، تجاوز از سرعت مقرره ، تجاوز به چپ ناشی از سبقت ، انحراف به چپ ، انحراف به راست ، گردش به طرز غلط ، عبور از محل ممنوع ، حرکت در خلاف جهت ، حرکت با دنده‌عقب ، نقص فنی حادث وسیله نقلیه ، نقص فنی مستمر وسیله نقلیه تغییر مسیر ناگهانی ، خطای عابر ، نقص ماده ۴ قانون ایمنی راهها ، نقص مقررات حمل بار ، یدک کشی به طرز غلط ، عبور از چراغ‌قرمز ، دور زدن در محل ممنوع ، نقص ماده ۲۱۱ و ۲۱۲ آیین‌نامه راهور ناجا ، سایر علت‌ها

منبع: یافته‌های پژوهش، ۱۳۹۶.

ساخت مدل قوانین انجمنی با استفاده از الگوریتم Fp-Growth در نرم‌افزار رپیدماینر

در طراحی پردازش مطابق شکل شماره (۳)، پردازش استخراج قوانین انجمنی از داده‌های ذخیره‌شده که در مراحل قبلی آماده‌سازی به‌عنوان داده محلی در رپیدماینر ذخیره‌شده‌اند استفاده می‌شود. زیرا داده‌های آماده‌شده در اکثر ویژگی‌ها از نوع اسمی چند مقدار بوده و پیش‌نیاز اجرای الگوریتم FP_Growth، داده‌های باینری است. لذا قبل از به‌کارگیری و انجام قوانین انجمنی و عملگر مربوطه تبدیل مقادیر عددی و اسمی

Archive of SID

www.SID.ir

ویژگی‌ها به مقادیر باینری انجام می‌شود. با این اقدام کلیه داده‌های سطری هر ویژگی به چندین ستون به تعداد مقادیر ویژگی تبدیل می‌شوند. به ازای هر مقدار از یک ویژگی ستون جدیدی از داده‌ها با مقدار باینری درست و غلط[1] تشکیل و تعداد ویژگی‌ها افزایش می‌یابد. عملگر Fp-Growth الگوریتمی است که مجموعه اشیاء متداول را نشان می‌دهد[2]و تعیین می‌کند که کدام‌یک از مجموعه اشیاء متداول به‌دست‌آمده است. سپس مجموعه تولیدشده به عملگر تولید قوانین انجمنی[3] داده می‌شود تا قوانین انجمنی موجود را تشخیص دهد. در اینجا نیز به‌صورت مکرر عملیات پاک‌سازی داده‌ها با عملگرهای نرم‌افزار انجام می‌شود و در این میان از عملگرهای دیگری مانند جداسازی برای بررسی نقش دقیق‌تر مجموعه ویژگی‌های تصادف در اجزا ترافیک به‌صورت مجزا (عامل انسانی، وسیله، اقلیمی و راه) نیز استفاده‌شده است.

پارامتر حداقل تعداد[4]: مجموعه اقلامی که باید استخراج شود را مشخص می‌کند. عدد پیش‌فرض (۱۰۰) است که پذیرفته می‌شود.

پارامتر حداقل پشتیبانی[5]: حداقل معیارهای اطمینان توسط این پارامتر مشخص می‌شود. از نسبت تعداد رکوردهایی که در آن اشیاء x و y هر دو حضور دارند، به‌کل تعداد رکودها معیار مهمی بانام پشتیبان به دست می‌آید. این مقدار عددی بین (۰ و ۱) است که هرقدر بیشتر باشد نشانگر این است که این دو شیء ارتباط بیشتری باهم دارند.

1- True & False
2- Frequent Item Set Minig
3- Create Association Rules Minig
4- Min Number
5- Min Support

Archive of SID

www.SID.ir

شکل شماره (۳). طرح پردازش قوانین انجمنی در نرم‌افزار رپیدماینر

ارزیابی مدل

الگوریتم‌های مورداستفاده برای به دست آوردن قوانین انجمنی می‌توانند الگوها و قوانین زیادی را تولید کنند. به‌ویژه با ازدیاد تعداد ویژگی‌ها (اشیاء) در یک مجموعه داده ممکن است تعداد زیادی قانون تولید شود که لزوماً همه آن‌ها جذاب نیستند. ازاین‌رو باید به دنبال قوانینی بود که بیشترین جذابیت را داشته باشند. جذابیت قوانین به عواملی مانند شخص یا سازمان استفاده‌کننده از قوانین و نیز حوزه موردبررسی، بستگی دارد. لذا آن دسته از قوانین انجمنی که درگذشته توسط افراد خبره برتر معرفی‌شده‌اند، توسط معیارهای مختلف ارزیابی قوانین انجمنی بررسی می‌شود. معیاری

Archive of SID

www.SID.ir

که این قوانین برتر را به‌درستی شناسایی نماید به‌عنوان معیار مناسب برای مسئله مورد کاوش شناسایی خواهد شد. یکی از معیارهای مهم برای مقایسه قوانین توسط الگوریتم‌های متنوع کاوش قوانین انجمنی، استفاده از معیار پشتیبان به همراه معیار اطمینان[1] است که از رابطه (۳) به‌دست‌آمده است.

$$Conf(A \rightarrow B)=\frac{SUP(AUB)}{SUP(A)}$$ رابطه (۳)

تولید قانون و نتایج مدل‌سازی

پس از اجرای الگوریتم FP_Growth آنچه در خروجی عملگرهای این بخش مشاهده می‌کنید، مجموعه‌ای است که به تحلیل‌گر یا خبره ارائه می‌شود تا قوانین بدیهی را از قوانین معتبر و باکیفیت تفکیک کند. به‌عنوان نمونه برخی از قوانین تولیدشده در پیوست شماره (۱) ارائه‌شده است. درنهایت جدول شماره (۳)، به‌عنوان بالاترین سهم ویژگی‌های مؤثر در تصادف در عامل‌های انسانی، راه، وسیله نقلیه و شرایط اقلیمی ترسیم‌شده است.

- **نقش نوع تصادف و عامل انسانی مؤثر در تصادف**: بر اساس خروجی مدل‌سازی به‌طور کامل مشخص است که در تصادفات فوتی عجله و شتاب بی‌مورد با اطمینان (٪۴۱)، در تصادفات جرحی با (٪۴۳) اطمینان عامل انسانی عجله و شتاب بی‌مورد و در تصادفات خسارتی نیز عجله و شتاب با (٪۳۹/۲) اطمینان در وقوع تصادفات از بالاترین علت مؤثر در تکرار انواع تصادفات شناخته‌شده است.

جدول شماره (۳). نتایج حاصل از قوانین انجمنی بر روی برخی ویژگی‌ها مؤثر در عاملیت‌های چهارگانه تصادفات و نوع تصادف

ردیف		نام ویژگی/فیلد	بالاترین تکرار عامل	درصد اطمینان در			
				تصادفات فوتی	تصادفات جرحی	تصادفات خسارتی	در کل تصادفات
۱	ویژگی‌های عوامل راه	نقایص مؤثر راه	ندارد	۹۶/۳	۹۷/۴	۹۸/۲	۹۷/۹
			نقص علائم افقی و عمودی	۰/۹	۱	۱/۳	۱/۱
۲		تعمیرات راه	در حال تعمیر	۹۹/۹	۱۰۰	۹۸/۳	۹۸/۸

1- Confidence

Archive of SID

www.SID.ir

ردیف		نام ویژگی/فیلد	بالاترین تکرار عامل	درصد اطمینان در تصادفات فوتی	تصادفات جرحی	تصادفات خسارتی	در کل تصادفات
			نیست				
۳		نوع رویه راه	اسفالتِ	۹۹/۵	۹۹/۷	۹۹/۸	۹۸/۸
۴		نوع راه	خیابان اصلی	۶۳/۶	۷۰/۵	۶۸/۵	۶۹/۱
۵		وضع هندسه محل	مستقیم، مسطح	۹۶/۴	۹۸/۳	۹۸/۴	۹۸/۳
۶		کاربری محل	مسکونی	۵۸	۶۹/۳	۶	۶۴/۷
۷		نوع شانه راه	شانه ندارد	۹۸/۷	۹۹/۵	۹۹/۴	۹۹/۴
۸		سمت حرکت راه	دوطرفه مجزا	۷۳/۷	۶۲/۵	۶۸	۹۹/۴
۹		نوع منطقه	هموار	۹۹/۶	۹۹/۸	۹۹/۹	۹۹/۸
۱۰		خط‌کشی	مقطع	۴۹/۶	۵۱/۷	۵۳/۷	۵۳/۱
۱۱	ویژگی‌های عامل انسانی تصادف	عامل انسانی مؤثر در تصادف	عجله و شتاب بی‌مورد	۴۱	۴۳/۳	۳۹/۲	۴۰/۴
۱۲		تحصیلات	دیپلم	۵۸/۸	۵۴/۶	۵۷/۸	۵۶/۸
۱۳		سن راننده	۲۸ تا ۳۸ سال	۲۶/۵	۲۷/۷	۳۴	۳۲
۱۴		جنسیت	مرد	۹۴/۱	۸۸/۷	۸۶/۸	۸
۱۵		شغل	آزاد	۳۶/۸	۴۴	۶۲	۵۶/۶
۱۶		کمربند/کلاه	استفاده کرده	۴۲/۸	۵۰/۸	۶۳/۸	۵۹/۹
۱۷		صدمه درصحنه	صدمه ندیده	۶۳	۷۲/۶	۱۰۰	۹۱/۷
۱۸		نوع گواهینامه	پایه دوم	۲۱/۴	۳۱/۵	۴۷/۶	۴۲/۷
۱۹		وضعیت گواهینامه	مجاز	۵۲/۹	۶۳/۴	۹۱/۶	۸۳
۲۰	ویژگی‌های عامل وسیله نقلیه	نوع وسیله نقلیه	سواری	۴۷/۹	۶۰/۷	۷۵/۲	۷۰/۸
۲۱		مانور وسیله نقلیه	حرکت به جلو	۸۹	۷۸/۵	۳/۴	۲۶
۲۲		عامل وسیله نقلیه مؤثر در تصادف	ندارد	۹۹/۳	۹۹/۹	۹۹/۹	۹۹/۹
۲۳		نوع خودرو	سبک	۵۶/۳	۶۹/۳	۸۵/۶	۸۰/۶
۲۴		تجهیزات ایمنی	ندارد	۹۶/۹	۹۸/۱	۸۹/۷	۹۲/۲
۲۵		کاربری خودرو	شخصی	۸۱/۳	۹۱/۶	۸۹/۷	۸۶/۷
۲۶		نوع برخورد	برخورد وسیله نقلیه	۴/۹	۱۱/۶	۷۵/۴	۵۶/۳

Archive of SID

www.SID.ir

ردیف		نام ویژگی/فیلد	بالاترین تکرار عامل	درصد اطمینان در تصادفات فوتی	تصادفات جرحی	تصادفات خسارتی	در کل تصادفات
			با وسیله نقلیه				
۲۷	شرایط عمومی و اقلیمی	شرایط روشنایی	روز	۶۱/۴	۶۷/۷	۷۳/۹	۷۲
۲۸		مانع دید	ندارد	۹۸/۵	۹۸/۸	۹۹/۹	۹۸/۹
۲۹		وضع هوا	صاف	۹۸/۵	۹۹/۴	۹۸/۳	۹۸/۹
۳۰		شرایط سطح راه	خشک	۹۸/۵	۹۸/۹	۹۸/۲	۹۸/۴
۳۱		ایام هفته تصادف	پنج‌شنبه	۱۸/۲	۱۵/۲	۱۵/۴	۱۵/۳
۳۲		ماه تصادف	مهرماه	۱۹/۶	۱۰/۱	۹/۳	۹/۵
۳۳		نوع تصادف	خسارتی	۷۰	۲۹/۳	۰/۷	۱۰۰
۳۴		دوره زمانی تصادف	ساعت ۱۲-۱۶	۲۰/۷	۲۵/۷	۲۸/۶	۷۷/۷
۳۵		علت تامه تصادف	عدم توجه به جلو	۳۸/۹	۲۷/۷	۲۵/۹	۲۶/۵

منبع: یافته‌های پژوهش، ۱۳۹۶.

یافته‌های پژوهش

به‌طورکلی از برخی قوانین تولیدشده نتایج حاصل مدل‌سازی این‌چنین استنباط شد که در (٪۵۶) از تصادفات برخورد دو وسیله نقلیه با یکدیگر ، در (٪۸) از سوانح رانندگی شدت صدمه راننده شدید (فوتی و جرحی) ، در (٪۹۸) تصادفات وسیله نقلیه عاملیتی در تصادف نداشته و تنها در (٪۱) بیشترین عاملیت نقص سیستم روشنایی در تصادفات دخیل بوده است. در (٪۴۰) تصادفات عجله و شتاب رانندگان علت اصلی تصادف شناخته‌شده و (٪۲۵/۳) از این تصادفات در شب اتفاق افتاده است. در (٪۹۸) این تصادفات نقص مؤثر درراه وجود نداشته و بیشترین سهم نقص راه مربوط به نداشتن علائم افقی و عمودی با (٪۱/۱) بوده است. (٪۸۷/۴) سوانح رانندگی توسط رانندگان مرد و (٪۱۳/۶) توسط رانندگان زن به وقوع پیوسته است. شغل (٪۵۶/۶) رانندگان آزاد و شغل (٪۳/۳) آنان رانندگی بوده است. (٪۸۰) تصادفات با خودروهای سبک و (٪۸/۳) با موتورسیکلت/ دوچرخه اتفاق افتاده است. (٪۱۱/۵) رانندگان فاقد گواهینامه، (٪۱/۹) گواهینامه غیرمجاز و (٪۳/۶) گواهینامه فاقد اعتبار بوده که درمجموع (٪۱۷) گواهینامه

Archive of SID

www.SID.ir

معتبر با شرایط قانونی نداشته‌اند. در (٪۱۹/۵) تصادفات رانندگان وسایل نقلیه از کمربند و کلاه مناسب استفاده نکرده‌اند. (٪۵۱/۳) تصادفات از ساعت (۱۲) ظهر تا (۲۰) شب اتفاق افتاده که بیشترین سهم مربوط به ساعت (۱۲) تا (۱۶) می‌باشد. بیشترین وقوع تصادف در مهرماه با (٪۹/۵) رخداده است. (٪۶۹) تصادفات در خیابان‌های اصلی و (٪۱۴/۵) در بزرگراه‌ها رخداده که بیشترین سهم با (٪۳/۷) از تصادفات در خیابان امام خمینی و (٪۳/۶) در بزرگراه خرازی از مکان‌های پرحادثه و (٪۹۲/۲) خودروهای مقصر فاقد تجهیزات ایمنی خاص بوده‌اند.

نتیجه‌گیری

در این پژوهش با بهره‌گیری از داده‌های تصادفات سال‌های (۱۳۹۰) تا (۱۳۹۴) با استفاده از قوانین انجمنی و استفاده از الگوریتم Fp-Growth بر روی انواع تصادف برای اولین بار ارتباط ویژگی‌ها و شناخت اشیاء و الگوهای پرتکرار آن‌ها کشف و همچنین ویژگی‌های تصادفات (الگوهای تصادف) در بیشترین تکرار شناسایی و درصد وقوع تکرار هریک از عوامل بررسی شد. برخی از مهم‌ترین یافته‌های این پژوهش عبارت‌اند از:

- مهم‌ترین عامل تصادفات، عامل انسانی عجله و شتاب بی‌مورد و مهم‌ترین علت تصادفات، عدم توجه به جلو یا حواس‌پرتی رانندگان می‌باشد.
- مشخص شد وقوع توأم کدامین ویژگی‌ها باعث افزایش شدت تصادف شده است.
- بالاترین نوع برخورد در تصادفات فوتی برخورد یک وسیله نقلیه با موتورسیکلت است.
- مکان‌های حادثه‌خیز، ضعف آموزش، انواع گواهینامه‌های رانندگی و نوع خودروهای دخیل در تصادف، شناسایی شد.
- عوامل مختلف اجزا ترافیک در بروز شدت تصادفات (خسارتی، جرحی و فوتی) تأثیرگذار می‌باشند. بهترین وضعیت هریک از ریز فاکتورها و عوامل اجزا ترافیک، وضعیتی است که ضریب مخاطره کمتر بوده و اولویت برنامه ریزان متولی هریک از عوامل باید در بهبود وضعیت موجود با درنظرگرفتن این راهبرد که هریک از

Archive of SID

www.SID.ir

ریزفاکتورهای عوامل چهارگانه اجزاء ترافیک به بهترین وضعیت تبدیل شوند، بهبود یابند.

نتایج این پژوهش نشان داد که (٪۵۶) از تصادفات مربوط به برخورد دو وسیله نقلیه بوده است. در(٪۸) از سوانح رانندگی شدت صدمه راننده فوتی و جرحی بوده و در (٪۹۸) تصادفات وسیله نقلیه عاملیتی در تصادف نداشته است. در (٪۴۰) تصادفات عجله و شتاب رانندگان علت اصلی تصادف و (٪۲۵,۳) از این تصادفات در شب رخ‌داده است. در (٪۹۸) تصادفات، راه نقص مؤثر نداشته است. سهم عمده سوانح رانندگی مربوط به مردان و شغل (٪۵۶,۶) رانندگان مقصر آزاد و شغل (٪۳,۳) آنان رانندگی بوده است. (٪۸۰) تصادفات با خودروهای سبک و (٪۸,۳) با موتورسیکلت/دوچرخه رخ افتاده است. (٪۱۷) رانندگان گواهینامه معتبر با شرایط قانونی نداشته‌اند. در (٪۱۹,۵) تصادفات رانندگان وسایل نقلیه از کمربند و کلاه مناسب استفاده نکرده‌اند. (٪۵۱,۳) تصادفات از ساعت (۱۲) ظهر تا (۲۰) شب رخ‌داده است. (٪۶۹) تصادفات در خیابان‌های اصلی و (۱۴,۵) در بزرگراه‌ها رخ‌داده است.

بررسی پژوهش‌های انجام‌شده در زمینه تصادفات رانندگی که به برخی از آن‌ها در پیشینه پژوهش نیز اشاره شد، نشان می‌دهد تاکنون در زمینه شناخت و تحلیل عوامل مؤثر بر تصادفات درون‌شهری با استفاده از روش داده‌کاوی برای هر چهار عامل اصلی تصادفات و بررسی توأمان این عوامل، پژوهشی انجام‌نشده است که این مورد از نوآوری‌های پژوهش محسوب می‌شود. همچنین در پژوهش‌های انجام‌شده در شناسایی یکی یا برخی از عوامل اصلی (عامل راه، انسان، وسیله نقلیه و شرایط اقلیمی) به‌صورت تک‌بعدی پژوهش‌های زیادی انجام‌شده و اکثر آن‌ها با الگوریتم‌های رگرسیون، درخت تصمیم و در کل دسته‌بندی با متغیر وابسته‌ای انجام‌شده است. از مزیت‌های دیگر این پژوهش، جامعیت و پرداختن با جزئیات کافی و به‌صورت دقیق به عوامل مؤثر در تصادفات رانندگی است. در مقایسه با پژوهش انجام‌شده توسط احمدی و علی‌محمدی و پژوهش انجام‌شده توسط هوشیار و شریفی پژوهش حاضر، عوامل و پارامترهای مؤثر بیشتری را در تحلیل تصادفات رانندگی موردتوجه قرار داده و به نتایج دقیق‌تری رسیده است که بهره‌برداری از آن‌ها در پیشگیری و کاهش تصادفات رانندگی مؤثر می‌باشد.

Archive of SID

www.SID.ir

پیشنهادهای پژوهش

با توجه به یافته‌ها و نتایج حاصل از این پژوهش، موارد زیر پیشنهاد می‌شود:

- افزایش آگاهی مردم نسبت به عوامل مؤثر در بروز تصادفات از طریق رسانه‌های مختلف، به‌خصوص نقش عامل انسانی، استفاده از کمربند ایمنی و خودروهای دخیل در تصادفات؛
- توجیه موتورسواران، از طریق رسانه‌های مختلف و آموزش روش‌های افزایش ایمنی و پیشگیری از تصادفات ؛
- توجیه مراکز صدور گواهینامه‌های رانندگی نسبت آموزش ویژه عوامل اثرگذار در تصادفات به رانندگان ؛
- استفاده از شیوه‌های مختلف ازجمله کلیپ، فیلم، تراکت و غیره، برای پیشگیری از عوامل مؤثر در تصادفات؛
- توسعه مدل به‌دست‌آمده در این پژوهش، به‌عنوان یک کمک سامانه تحلیلی تصادفات در کنار سامانه جمع‌آوری اطلاعات تصادف برای بهره‌گیری مدیران و خبرگان از علل وقوع ویژگی‌های شایع و پرتکرار تصادفات در هر منطقه و شرایط محیطی و اقلیمی و جمعیتی، در ارائه راه‌کارهای پیشگیرانه.

Archive of SID

www.SID.ir

منابع

- احمدی، مهدی؛ علی‌محمدی، عباس (۱۳۹۴). آنالیز زمانی و مکانی تصادفات رانندگی با استفاده از تراکم پنجره‌ای فازی، مهندسی حمل‌ونقل. ۷(۲). ص ۱۹۱-۲۰۵.
- اسماعیلی، مهدی (۱۳۹۲). آموزش گام‌به‌گام داده‌کاوی با رپیدماینر. تهران: نشر آتی نگر.
- اسماعیلی، مهدی (۱۳۹۴). داده‌کاوی مفاهیم و تکنیک‌ها (ویرایش سوم). تهران: نیاز دانش.
- جعفری، ادریس (۱۳۸۸). کاربرد داده‌کاوی در بررسی رفتار رانندگان متخلف در کلان‌شهرها. پایان‌نامه کارشناسی ارشد. دانشگاه تهران. تهران: دانشکده مهندسی صنایع، گروه مهندسی فناوری اطلاعات.
- سلمانی، محمد؛ رمضانزاده، مهدی؛ دریکوند، مسلم؛ ثابتی، فرخ (۱۳۸۷). بررسی عوامل مؤثر بر تصادفات جاده‌ای و ارائه راهکارهایی برای کاهش آن (موردمطالعه: منظومه روستایی جنوب خور و بیابانک). پژوهش‌های جغرافیای انسانی، ۶۵، ص ۸۷-۱۰۴.
- شکوهیار، سجاد؛ رضائیان، علی؛ ذوالفقاری، سارا (۱۳۹۴). خوشه‌بندی داده‌های تصادفات جاده‌ای با استفاده از فنون داده‌کاوی. مطالعات پژوهشی راهور. ۴(۱۴). ص۴۷ -۷۹.
- صفارزاده، امین؛ ابوالحسن نژاد، محمود؛ میرزا بروجردیان، وحید (۱۳۸۷). ارائه مدل اولویت‌بندی لیتگرا جهت تعیین قطعات خطرناک جاده‌ها برای عابران پیاده. دانشکده فنی دانشگاه تهران، ۴۲(۴). ص ۱۱-۲۴.
- صنیعی آباده، محمد؛ محمودی، سینا؛ طاهرپرور، محدثه (۱۳۹۴). داده‌کاوی کاربردی (ویرایش دوم). تهران: نیاز دانش.
- قاسمی نوقابی، مسعود؛ آیتی، اسماعیل (۱۳۹۱). تجزیه‌وتحلیل آماری فراوانی داده‌های تصادف: بررسی و ارزیابی روش‌های تحلیل. مجموعه مقالات سومین کنفرانس بین‌المللی حوادث رانندگی و جاده‌ای. دانشگاه تهران. ص (۴۲۳-۳۹۲).تهران : مرکز مطالعات راه و حمل‌ونقل دانشکده فنی دانشگاه تهران.
- قدیری افشار، مریم (۱۳۹۳). پیشگیری از حوادث ترافیکی. تهران: مرکز مدیریت حوادث و فوریت‌های پزشکی کشور. اداره پیشگیری از حوادث و ارتقای ایمنی.
- هوشیار، حسن؛ شریفی، بایزید (۱۳۹۵). تحلیل فضایی تصادفات درون‌شهری (موردمطالعه: شهر ارومیه). مهندسی جغرافیایی سرزمین. ۱(۱). ص۹۰ -۱۰۱.
- Agrawal, T. (1993). Mining Association Rules between Sets of Items in Large Databases. Proceedings of ACM SIGMOD. 207-216.

Archive of SID

www.SID.ir

- Alpaydin, E. (2010). Introduction to Machine Learning. London: The MIT Press Cambridge.
- Ding, J., Yau, S. T. (2009). TCOM, an Innovative Data Structure for Mining Association Rules among Infrequent Items. Computers and Mathematics with Applications. 57(2). 290-301.
- Han, J., Kamber, M. (2000). Data Mining: Principles and Concepts. 3th Edition. UK: Morgan Kauffman Publisher.
- Hur, R. (2010). Medical data mining based on association rules. Computer and Information Science. 3(4). 104-108.
- LAN, C. J. Li, J. (2006). Statistical evaluation of motorcycle crash injury severities by using multinomial models. Transportation Research Board. 85th Annual Meeting.
- Lin, K. C. and Yeh, C. L. (2012). Use of Data Mining Techniques to Detect Medical Fraud in Health Insurance. International Journal of Engineering and Technology Innovation. 2(2). 126-137.
- Piatetsky-Shapiro G. and Frawley W. J. (1991). Knowledge Discovery in Databases. AAAI/MIT Press. 13(3). 58-82.
- Tong, H. (2008). Random Walk with Restart: Fast Solutions and Applications. Knowledge and Information Systems. An International Journal (KAIS)). 14. 327-346.

Archive of SID

www.SID.ir

پیوست ۱. برخی از قوانین تولیدشده

.Association Rules

[NOE_TASADOF = خسارتی] --> [AMEL_ENSANI = مهار نکردن خودرو به طرز صحیح] (confidence: 0.001)

[NOE_TASADOF = جرحی] --> [AMEL_ENSANI = مهار نکردن خودرو به طرز صحیح] (confidence: 0.001)

[NOE_TASADOF = خسارتی] --> [AMEL_ENSANI = خستگی و خواب‌آلودگی] (confidence: 0.001)

[NOE_TASADOF = جرحی] --> [AMEL_ENSANI = خستگی و خواب‌آلودگی] (confidence: 0.001)

[NOE_TASADOF = خسارتی] --> [AMEL_ENSANI = عدم‌تشخیص سهم عبور سایرین] (confidence: 0.001)

[NOE_TASADOF = جرحی] --> [AMEL_ENSANI = عدم‌تشخیص سهم عبور سایرین] (confidence: 0.003)

[NOE_TASADOF = فوتی] --> [AMEL_ENSANI = ندارد] (confidence: 0.080)

[NOE_TASADOF = فوتی] --> [AMEL_ENSANI = بی‌توجهی به مقررات] (confidence: 0.099)

[NOE_TASADOF = جرحی] --> [AMEL_ENSANI = ندارد] (confidence: 0.135)

[NOE_TASADOF = خسارتی] --> [AMEL_ENSANI = ندارد] (confidence: 0.141)

[NOE_TASADOF = جرحی] --> [AMEL_ENSANI = هم عجله و شتاب و هم بی‌توجهی به مقررات] (confidence: 0.202)

[NOE_TASADOF = جرحی] --> [AMEL_ENSANI = بی‌توجهی به مقررات] (confidence: 0.224)

[NOE_TASADOF = خسارتی] --> [AMEL_ENSANI = بی‌توجهی به مقررات] (confidence: 0.228)

[NOE_TASADOF = خسارتی] --> [AMEL_ENSANI = هم عجله و شتاب و هم بی‌توجهی به مقررات] (confidence: 0.235)

[NOE_TASADOF = فوتی] --> [AMEL_ENSANI = هم عجله و شتاب و هم بی‌توجهی به مقررات] (confidence: 0.402)

[NOE_TASADOF = فوتی] --> [AMEL_ENSANI = عجله و شتاب بی‌مورد] (confidence: 0.410)

[NOE_TASADOF = جرحی] --> [AMEL_ENSANI = عجله و شتاب بی‌مورد] (confidence: 0.434)

Archive of SID

www.SID.ir