مهندسی صنایع و مدیریت شریف، (زمستان ۱۴۰۰)
دوره‌ی ۱-۳۷، شماره‌ی ۲، ص. ۳۹-۵۳ (پژوهشی)

Original Article

تأثیر دو عامل تصادفات و شرایط جوی بر ترافیک با استفاده از درخت تصمیم (مطالعه‌ی موردی: شهر تهران)

مینا کریمی * **(دانشجوی دکتری)**
دانشکده‌ی مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه‌نصیرالدین طوسی

محمد سعدی مسگری (استادیار)
گروه سیستم اطلاعات مکانی قطب علمی فناوری اطلاعات مکانی

دانشکده‌ی مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه‌نصیرالدین طوسی

عوامل زیادی بر ترافیک شهر تهران تأثیرگذارند. هدف این تحقیق بررسی تأثیر دو عامل تصادفات رانندگی و شرایط جوی در ایجاد ترافیک در شهر تهران و مدل‌سازی و پیش‌بینی طول ترافیک ایجاد شده است. بدین منظور ابتدا تحلیل‌های آماری و عملیات پیش‌پردازش روی داده‌ها انجام شده است. سپس با استفاده از روش داده‌کاوی درخت تصمیم، دو مدل برای پیش‌بینی در مناطق مختلف تهران به دست آمده و در نهایت قوانین مربوط به هر مدل استخراج شده است. مطابق تحلیل آماری رگرسیون خطی، طول ترافیک ناشی از تصادف تابع شدت تصادف، منطقه‌ی شهرداری محل تصادف و تعداد وسیله‌ی نقلیه‌ی درگیر است و نوع روز عادی یا تعطیلی روی آن تأثیری ندارد. طول ترافیک ناشی از شرایط جوی نیز تابع منطقه شهرداری، نوع شرایط جوی و روز عادی یا تعطیلی است. بیشترین طول ترافیک ایجاد شده که برابر ۲۰۰-۵۰۰ متر است، مربوط به تصادفات با دو وسیله‌ی نقلیه‌ی درگیر یا در شرایط بارانی است.

واژگان کلیدی: ترافیک، تصادفات، شرایط جوی، داده‌کاوی، درخت تصمیم، سیستم‌های اطلاعات مکانی.

minakarimi@email.kntu.ac.ir
mesgari@kntu.ac.ir

۱. مقدمه

امروزه با افزایش تعداد وسایل نقلیه، ترافیک و ازدحام شبکه‌های معابر درون‌شهری یکی از بزرگترین چالش‌های شهروندان و مسئولان است. اتلاف زمان، افزایش سوخت مصرفی، آلودگی‌های صوتی و آلودگی هوا موجب اهمیت کنترل ترافیک شده است. اطلاعات ترافیکی عنصری کلیدی برای آگاه کردن رانندگان هنگام تصمیم‌گیری در مورد مسیر حرکتی خود و در نتیجه پیشگیری از ازدحام درنظرگرفته می‌شود.[۱] سیستم‌های کنترل ترافیک شامل تجهیزات و فنّاوری‌های مختلفی است که در آن داده‌های مختلف ترافیکی مانند تصادفات خودرو، حجم ترافیک و تراکم خودرو در سطوح مختلف ثبت و نگهداری می‌شود.[۲] شمار خودروها و وسایل نقلیه‌ی شخصی در شهر تهران به صورت روزافزون در حال افزایش است و ترافیک به یکی از مشکلات مهم این کلان‌شهر تبدیل شده است. شرکت کنترل ترافیک تهران با استفاده از چندین مرکز کنترل ترافیک و از طریق سیستم‌های هوشمند کنترل ترافیک مانند دوربین‌های نظارت تصویری، تردد شمارها و سیستم‌های کنترل هوشمند چراغ‌ها، مسئولیت نظارت بر ترافیک شهری را بر عهده دارد. اطلاعات و داده‌های به دست آمده، توسط کارشناسان و یا به صورت خودکار در سامانه‌ی ویژه‌یی ثبت می‌شود. این سامانه که تحت وب و بر اساس سیستم اطلاعات مکانی (GIS) [۱] طراحی شده است، مبنای اصلی گزارش‌گیری و اطلاع‌رسانی ترافیک تهران است.[۳]

در ترافیک تهران علاوه بر عوامل ناشی از طراحی هندسی معابر و حجم تردد، عوامل دیگری نیز دخالت دارند؛ عواملی مانند تصادف، حریق، خرابی خودروها، شرایط جوی و نظایر آن که به صورت مقطعی باعث افزایش ترافیک می‌شوند. این عوامل پیش‌بینی نشده در زمان وقوع، بسته به پارامترهای خود، رفتارهایی قابل پیش‌بینی دارند. چنانچه تأثیرات ناشی از عوامل مشخص باشد، می‌توان با اتخاذ تمهیداتی تأثیر مخرب آنها را بر ترافیک کمرنگ کرد.

از بین عوامل مختلف مؤثر در ترافیک شهر تهران که در سامانه‌ی ثبت وقایع ثبت می‌شوند، دو عامل مهم تصادف و شرایط جوی در این پژوهش مورد بررسی قرار میگیرد. تصادفات به دلیل ایجاد اختلال در جریان ترافیک و انسداد بخش‌هایی از شبکه‌ی معابر باعث ایجاد ترافیک‌های سنگین می‌شود.[۲] تأثیر همه‌ی تصادفات یکسان نیست. هر تصادف بسته به زمان، مکان و شدت تصادف، تأثیر متفاوتی بر ترافیک خواهد گذاشت.[۴] بر اساس گزارشات سازمان بهداشت جهانی، با کشته شدن بیش از ۱/۳ میلیون نفر و جراحت ۲۰ تا ۵۰ میلیون نفر در هر سال، تصادفات جاده‌یی به عنوان یکی از علل اصلی مرگ و میر و جراحت در جهان به شمار

* نویسنده مسئول
تاریخ: دریافت ۱۳۹۸/۱۰/۲۲، اصلاحیه ۱۳۹۹/۴/۱۰، پذیرش ۱۳۹۹/۱۰/۱.
DOI:10.24200/J65.2020.54632.2066

می‌رود.[۵ و ۶] همچنین تغییرات شرایط جوی باعث اختلال در عبور و مرور وسایل نقلیه می‌شود.[۷ و ۸] تعیین تأثیر شرایط آب و هوایی بر میزان تصادفات و ترافیک، امری ضروری است.[۷] نتایج بررسی تأثیر وضعیت آب و هوا بر ترافیک ایجاد شده با استفاده از داده‌های دریافتی از مراکز کنترل ترافیک، نشان می‌دهد بارش باران باعث کاهش سرعت به میزان ۸-۱۲ درصد و کاهش ظرفیت معابر به میزان ۷-۸ درصد شده است. لغزندگی جاده باعث کاهش سرعت به میزان ۷-۸ درصد و بارش برف منجر به کاهش سطح سرویس و ترافیک به میزان ۶۵ درصد می‌شود.[۳] GIS ابزاری ضروری برای نمایش داده‌های تصادف و تحلیل نقاط داغ[۲] در بزرگراه‌هاست.[۹] بسیاری از سازمان‌های مرتبط با ترافیک از GIS برای تحلیل تصادفات استفاده می‌کنند.

اطلاعات به دست آمده از سامانه از طریق وب‌سایت و تابلوهای متغیر خبری به رانندگان اطلاع‌رسانی می‌شود؛ اما مشکل بزرگ این است که داده‌های به دست آمده وضعیت حال ترافیک را به فرمت‌های مختلف ارائه می‌کند و امکان پیش‌بینی ترافیک براثر عوامل مختلف را ندارد. به دلیل آن که تصادفات و شرایط جوی و همچنین وضعیت‌های ترافیک ناشی از آنها در سامانه‌ی ثبت وقایع شهرداری تهران ثبت می‌شود، امکان داده‌کاوی روی این داده‌ها میسر شده است. چنانچه تصمیم‌گیران بدانند به واسطه‌ی یک تصادف با یک الگوی خاص چه میزان ترافیک به وجود خواهد آمد یا این که در شرایط جوی موجود، چه میزان ترافیک و در چه بخش‌هایی از شهر خواهیم داشت، امکان پیش‌بینی وضعیت ترافیک در ساعات آتی و اطلاع‌رسانی درست آن به رانندگان فراهم خواهد شد. اگر رانندگان از وضعیت ترافیک، در زمانی که قصد تردد در معابر را دارند اطلاع داشته باشند، می‌توانند با تغییر زمان ترددشان به کاهش اتلاف وقت کمک کنند.

هدف این تحقیق بررسی میزان تأثیر تصادفات و شرایط جوی بر ترافیک شهر تهران در مناطق مختلف شهرداری و مدل‌سازی و پیش‌بینی طول ترافیک ایجاد شده بر اثر این عوامل در شرایط مختلف با استفاده از تحلیل‌های داده‌کاوی و GIS است. در این تحقیق جامعه‌ی آماری کلیه‌ی رکوردهای ثبت شده در پایگاه داده سامانه‌ی ثبت وقایع شرکت کنترل ترافیک شهر تهران از تاریخ ۱۳۹۳/۱/۱ الی ۱۹۳۴/۱۲/۲۹ با حجم تقریبی ۱٫۸۰۰٫۰۰۰ رکورد است. برای دستیابی به هدف تحقیق، ابتدا این داده‌ها مورد بازبینی و پالایش قرار گرفته و پس از جمع‌بندی، وارد نرم‌افزار اس‌پی‌اس‌اس (SPSS) [۳] شده و پردازش و تحلیل می‌شوند. آماره‌های مورد استفاده برای تحلیل، شامل آمار توصیفی، یعنی فراوانی، درصد و میانگین و آمار استنباطی شامل آزمون رگرسیون خطی و خوشه‌بندی اکتشافی خواهد بود. در نهایت نتایج وارد نرم‌افزار رپیدماینر[۴] شده و مدل پیش‌بینی بر اساس درخت تصمیم با الگوریتم C۵/۰ توسط این نرم‌افزار ایجاد می‌شود. بخش‌های دیگر این نوشتار به شرح زیر تنظیم شده است: در بخش ۲ مروری بر تحقیقات پیشین و در بخش ۳ مفاهیم پایه شامل داده‌کاوی و درخت تصمیم ارائه شده است. در بخش ۴ روش پیشنهادی تحقیق بیان شده است و بخش ۵ به پیاده‌سازی روش و نتایج حاصله اختصاص یافته است. در نهایت، در بخش ۶ نتیجه‌گیری و پیشنهادات ارائه شده است.

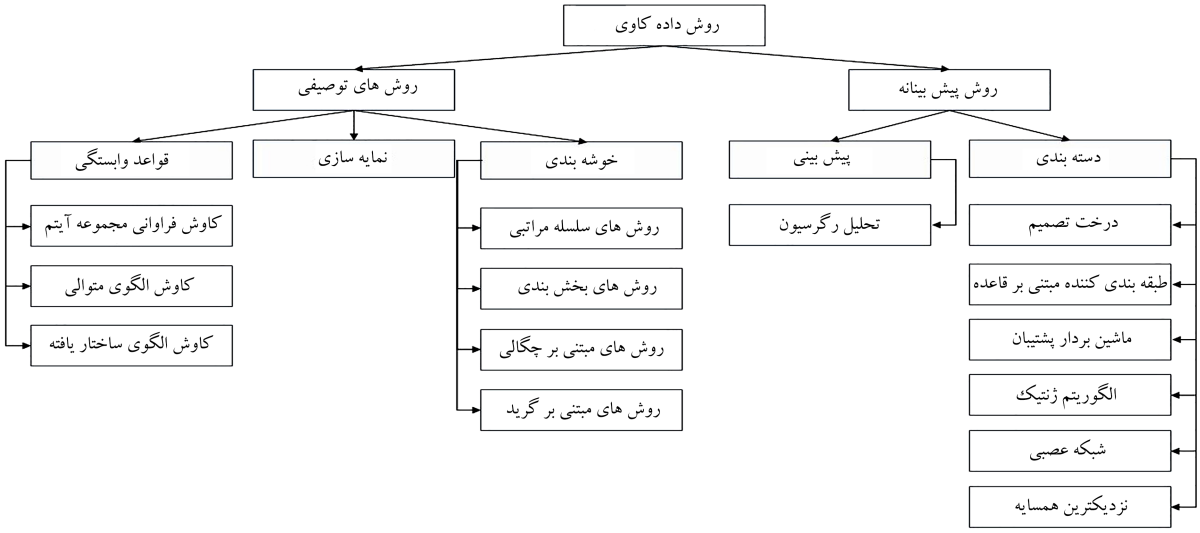
۲. مروری بر کارهای پیشین

تاکنون تحقیقات گسترده‌یی در زمینه‌ی تأثیر تصادفات و عوامل جوی بر ترافیک صورت گرفته است. پژوهش آیتی و همکاران نشان می‌دهد حدود ۲۵٪ تلفات ناشی از مرگ‌ومیرهای غیرطبیعی در ایران، ناشی از تصادفات خودرو است.[۱۰] بر اساس گزارشات ارائه شده در سال ۱۳۸۷، در ایران هر ساله به طور متوسط ۲۴۰۰۰ نفر در تصادف کشته می‌شود.[۱۱] کامران وهاس[۵] یک روش چندسطحی به منظور تشخیص ترافیک ناشی از تصادف در جاده ارائه کردند. در این روش با استفاده از داده‌های بلادرنگ به دست آمده از سیستم GPS [۶] خودرو، الگوریتم‌هایی برای تشخیص الگوی غیرمعمول ترافیکی و رفتار خودرو ارائه شده است.[۱۲] اِوسُوکُف و اِبکن مدل‌های پیش‌بینی ترافیک را ترکیب کرده و فرایندی را به منظور پیش‌بینی ترافیک در سیستم‌های اطلاعات ترافیکی ارائه کردند. آنها از مدل خوشه‌بندی برای کشف ارتباط بین سرعت ـ جریان ـ ظرفیت در ترافیک و پیش‌بینی پارامترهای ترافیکی با دقت بالا استفاده کردند.[۱۳] ابدل ـ اتی و همکاران یک مدل پیش‌بینی احتمال وقوع تصادف با استفاده از متغیرهای جریان ترافیکی بلادرنگ (از طریق مجموعه‌یی از حسگرهای زیرزمینی) و داده‌های بارش (جمع‌آوری شده در ایستگاه‌های آب و هوایی) که به طور بالقوه با وقوع تصادف همراه‌اند، ارائه کردند.[۱۴] در این تحقیقات فقط اثر تصادف بررسی شده و عامل شرایط جوی در نظرگرفته نشده است. همچنین از روش‌های داده‌کاوی استفاده نشده است.

تحقیقات مختلفی روی به کارگیری GIS در تحلیل تصادفات تمرکز کرده‌اند. سیدیکوئی[۷] و همکاران، عدم توجه به پارامترهای مکانی مجاور راه را یکی ازکاستی‌های مطالعات صورت گرفته در تحلیل تصادفات دانستند. آنها بر این باورند که می‌توان این تأثیرات را با استفاده از روش‌های غیرپارامتریک نظیر داده‌کاوی مدل‌سازی کرد.[۱۵] عفتی و همکاران یک استراتژی تلفیقی را برای بررسی انواع تصادفات و عوامل مکانی مؤثر بر تصادفات در راه‌های دوخطه‌ی دوطرفه‌ی برون‌شهری ارائه کردند. آنها از تحلیل‌های مکانی و روش خوشه‌بندی تجمعی سلسله‌مراتبی بر پایه‌ی K-Means استفاده کردند.[۶] اردوگان و همکاران از تحلیل‌های GIS برای شناسایی نقاط داغ تصادفات در یکی از شهرهای ترکیه استفاده کرده‌اند و تحقیق خود را روی روش‌های کنترل ترافیک انجام دادند.[۹] شفابخش و همکاران فنّاوری اطلاعات مکانی و آمار مکانی را به منظور استخراج تأثیر عوامل مکانی در شناسایی تصادفات ترافیکی ترکیب کرده و چهار تحلیل خوشه‌یی را آزمایش کرده‌اند.[۱۶] مهدویان و نیک‌نفس از داده‌های واقعی GPS خودروها به منظور پیش‌بینی مسیر و مقصد نهایی خودروها با کمک GIS و داده‌کاوی استفاده کردند.[۱]

عوامل جغرافیایی، عوامل آب و هوایی و وضعیت جاده از عوامل مهمی هستند که بر رانندگی تأثیر دارند و نقش انسان در زمان حادثه به دلیل این عوامل کم است. درگزارشی نشان داده شده است که حدود ۴۰٪ از تصادفات به هوای بارانی و برفی منتسب است.[۱۷] در استرالیا ۲۰٪ تصادفات به هوای بارانی مرتبط است[۱۸] و در آمریکا روزهای با بیشترین بارش باران و برف با بیشترین میزان تصادفات همراه است.[۱۹] کیو و نیکسون[۸] با مرور سیستماتیک و فراتحلیلی تحقیقات انجام شده در سال‌های ۱۹۶۷ تا ۲۰۰۵، تأثیر هوای نامساعد را بر تصادفات جاده‌یی بررسی کرده‌اند. نتایج نشان می‌دهد نرخ تصادف معمولاً در زمان بارش افزایش می‌یابد. برف می‌تواند اثر بیشتری نسبت به بارش باران بر روی وقوع تصادف داشته باشد: برف می‌تواند میزان تصادف را ۸۴٪ و میزان آسیب را ۷۵٪ افزایش دهد.[۷] ابراهیم وهال[۹] تأثیر شرایط نامساعد هوایی بر روابط جریان ـ مسیر و جریان ـ سرعت را مطالعه کردند و از تحلیل رگرسیون برای انتخاب مدل مناسب نمایش‌دهنده روابط استفاده کردند.[۸] در این تحقیقات تأثیر شرایط جاده‌یی نامساعد بر تصادفات بررسی شده و هیچ بررسی روی ترافیک ایجاد شده و تحلیل آن با روش‌های داده‌کاوی صورت نگرفته است.

فیروزی و مرادی مفرد تأثیر عوامل مختلف جوی بر تصادفات را از طریق توزیع داده‌های تصادفات در شرایط مختلف جوی نظیر بارانی، یخبندان، دمای بالا یا



شکل ۱. روش‌های داده‌کاوی.[۳۱]

پایین، آفتابی و وزش باد بررسی کردند.[۲۰] ون و همکاران به بررسی اثرات جاده‌ها و شرایط آب و هوایی بر تصادفات با استفاده از مدل فضایی ـ زمانی بیزین برای اندازه‌گیری ارتباط بین فراوانی تصادف و عوامل خطر احتمالی نظیر شرایط نامساعد جوی پرداختند.[۲۱] زینگ و همکاران ارتباط غیرخطی و اثرات تأخیری فاکتورهای جوی را بر تصادفات بررسی کردند.[۲۲] مالین و همکاران نیز به بررسی خطر نسبی تصادف در شرایط مختلف جوی و همچنین ترکیب این شرایط پرداختند. روش شامل محاسبه‌ی توزیع پالم[۱۰] در شرایط مختلف و مقایسه آن با توزیع شرایطی است که در حوادث مشاهده می‌شود. هنگامی که این دو توزیع متفاوت‌اند، این وضعیت از نظر آماری بر خطر تصادف تأثیر می‌گذارد.[۲۳] تئوفیلاتوس با ترکیب داده‌های بلادرنگ ترافیک و هواشناسی، احتمال و شدت تصادفات جاده‌یی در شریان‌های شهری را مشخص کرد.[۲۴] در هیچ یک از این تحقیقات به رابطه و تأثیر تصادفات و شرایط جوی بر ترافیک و پیش‌بینی طول ترافیک پرداخته نشده است.

تِسما و همکاران در تحقیق خود، قابلیت بالقوه‌ی فنّاوری داده‌کاوی را در توسعه‌ی مدلی بررسی کردند که بتواند از تحلیل شدت تصادفات ترافیکی جاده‌یی در تلاش برای پیشگیری و کنترل تصادف خودرو در شهر آدیس‌آبابا پشتیبانی کند. آنها از روش درخت رگرسیون تطبیق‌یافته برای کاوش قوانین و طبقه‌بندی تصادفات ترافیکی جاده‌یی استفاده کردند.[۲] ابلان و همکاران شدت تصادفات جاده‌یی را با استفاده از قوانین تصمیم از طریق درخت تصمیم تحلیل کردند.[۲۵] چانگ و چن از روش‌های مبتنی بر درخت شامل طبقه‌بندی و درخت رگرسیون (CART)[۱۱] به منظور تحلیل فراوانی تصادفات بزرگراه‌ها استفاده کردند.[۲۶] توکلی و مهیمنی نیز از الگوریتم CART برای تحلیل شدت جراحت تصادفات در جاده‌های دوبانده‌ی دوطرفه‌ی روستایی استفاده کردند.[۱۱] آنها نشان دادند CART روش مناسبی برای تحلیل تصادفات بزرگراه‌هاست زیرا به هیچ ارتباط از پیش تعریف شده‌یی بین متغیر هدف (وابسته) و پیش‌بینی‌کننده‌ها (متغیرهای مستقل) نیاز ندارد.[۱۱ و ۲۶] چونگ و همکاران عملکرد چهار روش یادگیری ماشین را در مدل‌سازی شدت جراحت تصادفات بررسی کردند. این روش‌ها شامل شبکه‌های عصبی ماشین بردار پشتیبان (SVM)،[۱۲] درخت تصمیم و روش ترکیبی درخت تصمیم ـ شبکه‌ی عصبی هستند.[۴] در این تحقیقات فقط عامل تصادفات با استفاده از روش‌های داده‌کاوی تحلیل و بررسی شده است و طول ترافیک مدل‌سازی و پیش‌بینی نشده است.

۳. داده‌کاوی

امروزه با گسترش سیستم‌های پایگاه داده و حجم بالای داده‌های ذخیره شده در این سیستم‌ها، به روش‌هایی که بتوان داده‌های ذخیره شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرارداد، نیازمند هستیم.[۲ و ۲۷] داده‌کاوی عبارت است از «فرایند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده‌های بزرگ به منظور کشف الگوهای پنهان».[۲۸ و ۲۹] روش‌های داده‌کاوی یکی از مجموعه روش‌های مورد استفاده برای تحلیل مقدار بسیار عظیمی از داده‌ها و تبدیل آنها به اطلاعات مفید و دانش است.[۳۰] این روش‌ها به طور گسترده‌یی در تحلیل شدت تصادفات با نتایج رضایت‌بخش مورد استفاده قرارگرفته شده است.[۲]

داده‌کاوی عمدتاً با ساختن مدل‌ها مرتبط است. یک مدل به الگوریتم یا مجموعه قوانینی گفته می‌شود که مجموعه‌یی از ورودی‌های را (معمولاً به شکل پایگاه داده‌ها) با هدف یا مقصد خاصی مرتبط می‌کند.[۲] مطابق شکل ۱ روش‌های اصلی داده‌کاوی دو دسته‌اند: توصیفی و پیش‌بینانه.[۲و۳۱] روش‌های توصیفی خواص عمومی داده‌ها را مشخص می‌کنند و هدف آنها پیدا کردن الگوهای قابل تفسیر توسط انسان برای داده‌ها است. روش‌های پیش‌بینانه، پیش‌بینی رفتار آینده‌ی آنهاست و منظور از آن به کارگیری چند متغیر یا فیلد در پایگاه داده برای پیش‌بینی مقادیر آینده یا ناشناخته‌ی دیگر متغیرهاست.[۲و۳۱] یکی از عملکردهای پیش‌بینی، «دسته‌بندی» است و آن، فرایند یافتن مدلی است که با تشخیص دسته‌ها یا مفاهیم داده می‌تواند دسته‌ی ناشناخته‌ی اشیاء دیگر را پیش‌بینی کند. در واقع دسته‌بندی یک تابع یادگیری است که یک قلم داده را به یکی از دسته‌های از قبل تعریف شده نگاشت می‌کند.[۷]

۱.۳. الگوریتم رگرسیون خطی[۱۳] (LR)

در مواردی که بررسی ارتباط بین یک متغیر وابسته با یک یا چند متغیر مستقل مد نظر باشد و هدف محقق این است که بر اساس این ارتباط و با استفاده از داده‌های تاریخی، پارامتر (پارامترهایی) برای متغیر (متغیرهای) مستقل برآورد و با

ارائه مدل اقدام به پیش‌بینی کند، می‌توان از رگرسیون استفاده کرد. رگرسیون تکنیکی آماری برای بررسی و مدل‌سازی روابط میان داده‌هاست که به محاسبه‌ی یک ارتباط خطی میان متغیرهای وابسته و مستقل می‌پردازد. محاسبه‌های انجام شده در پیش‌بینی‌ها کاربرد دارند. در حقیقت رگرسیون خطی از فرمول‌های مناسبی برای محاسبه‌ی مقادیر A و B برای رسیدن به پیش‌بینی مقدار C استفاده می‌کند. در رگرسیون خطی ساده‌ترین ارتباط یک ارتباط خطی است. هدف، ترسیم بهترین خطی است که مقدار خطا را کمینه کند. رابطه‌ی ۱ یک تابع خطی را نشان می‌دهد:

$$y = a \pm box \qquad (۱)$$

که در آن a مقدار ثابت و b شیب خط است. یک مدل رگرسیون خطی مرکب یا چندمتغیره را می‌توان به صورت رابطه‌ی ۲ نشان داد:

$$y = \beta_0 + \beta_{1x1} + \beta_{2x2} + \ldots \beta_{n-1xn-1} + \varepsilon \qquad (۲)$$

که در آن $x_1, x_2, \ldots, x_n$ متغیرهای مستقل، $\beta_0, \beta_1, \ldots, \beta_{n-1}$ ضرایب رگرسیون و ε بیان‌گر خطاست.

۲.۳. الگوریتم خوشه‌بندی دومرحله‌یی [۱۴]

به طور کلی خوشه‌بندی فعالیتی برای گروه‌بندی مشاهدات در قالب زیرگروه‌های مشابه به هم و بر اساس یک یا چند ویژگی است.[۳۲] خوشه‌بندی شامل خانواده بزرگی از روش‌ها و الگوریتم‌ها است.[۳۰] یکی از الگوریتم‌های خوشه بندی، الگوریتم خوشه‌بندی دومرحله‌یی است. این روش، ابزاری اکتشافی است که برای آشکارکردن خوشه‌های ذاتی و طبیعی موجود در مجموعه داده که به طور معمول دیده نمی‌شوند، طراحی شده است.[۳۳] تفاوت اصلی روش دومرحله‌یی با روش‌های دیگر خوشه‌بندی عبارت است از:[۳۲، ۳۴ و ۳۵]

- قابلیت خوشه‌بندی بر اساس متغیرهای گسسته (رسته‌یی) و پیوسته؛
- انتخاب خودکار تعداد خوشه‌ها؛
- قابلیت تحلیل کارآمد فایل داده‌های بسیار بزرگ

این الگوریتم با فرض مستقل بودن متغیرها در مدل خوشه بندی، اندازه‌ی احتمال[۱۵] را به عنوان فاصله‌ی موجودیت‌ها در نظر می‌گیرد. همچنین، فرض بر این است که هر متغیر پیوسته دارای توزیع نرمال و هر متغیر گسسته دارای توزیع نرمال چندمتغیری است. البته آزمون‌های تجربی نشان می‌دهد که الگوریتم دومرحله‌یی نسبت به تغییر فرض‌های استقلال و نرمال بودن، به اندازه کافی پایدار[۱۶] است. مراحل الگوریتم خوشه‌بندی دومرحله‌یی چنین خلاصه می‌شود:[۳۳، ۳۴ و ۳۶]

مرحله‌ی اول) در این مرحله که مرحله‌ی پیش‌خوشه[۱۷] است، بر اساس روش خوشه‌بندی سلسله‌مراتبی تجمعی[۱۸] هر رکورد به عنوان یک خوشه درنظر گرفته می‌شود. تمامی رکوردها یک به یک بررسی شده و بر اساس معیار درنظرگرفته شده برای فاصله‌ی موجودیت‌ها، در مورد ادغام آن با خوشه‌های قبل یا آغاز یک خوشه جدید تصمیم‌گیری می‌شود. بنابراین در این مرحله زیرخوشه‌هایی به دست می‌آید.

مرحله‌ی دوم) در این مرحله زیرخوشه‌های مرحله‌ی اول به عنوان ورودی در نظر گرفته شده و بر اساس خوشه‌بندی سلسله‌مراتبی تجمعی به تعداد خوشه‌های مورد نظر تبدیل می‌شوند. در صورت عدم تعیین تعداد خوشه‌ها،این الگوریتم به صورت خودکار و با معیارهایی نظیر BIC [۱۹] و AIC [۲۰] زیرخوشه‌ها را به تعداد بهینه‌یی از خوشه‌ها تبدیل می‌کند.

۳.۳. درخت تصمیم (DT) [۲۱] و الگوریتم c۵/۰

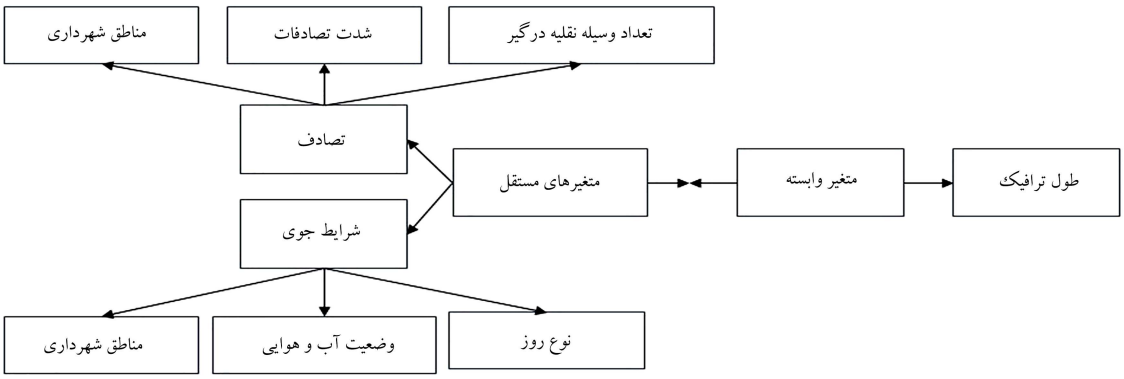
یکی از روش‌های متداول دسته‌بندی، درخت تصمیم است[۴] که برای کاوش در داده‌ها و کشف دانش کاربرد دارد.[۳۰] این الگوریتم داده‌ها را به مجموعه‌های مشخصی تقسیم می‌کند. هر مجموعه شامل چندین زیرمجموعه از داده‌های کم و بیش همگن است که دارای ویژگی‌های قابل پیش‌بینی هستند. درخت تصمیم درختی است که در آن نمونه‌ها را به نحوی دسته‌بندی می‌کند که از ریشه به سمت پایین رشد می‌کنند و در نهایت به گره‌های برگ می‌رسد. هرگره داخلی یا غیر برگ با یک ویژگی مشخص می‌شود. این ویژگی سؤالی را در رابطه با مثال ورودی مطرح می‌کند. در هر گره داخلی به تعداد جواب‌های ممکن برای این سؤال، شاخه وجود دارد که هر شاخه با مقدار آن جواب مشخص می‌شود. برگ‌های این درخت با یک کلاس یا یک دسته از جواب‌ها مشخص می‌شوند. بعد از تقسیم گره، تصمیم‌گیری در مورد برگ بودن گره‌های جدید انجام می‌گیرد. اگر برگ نباشد، این گره‌ها به عنوان ریشه برای زیر درخت‌های جدید قرار می‌گیرند و برای تشکیل زیردرخت‌های جدید استفاده می‌شوند.[۲ و ۴] مراحل فوق به روش بازگشتی بسط داده می‌شوند تا زمانی که تمام گره‌های جدید به عنوان گره‌های برگ باشند.[۳۰] مزیت درخت تصمیم، استخراج قوانین طبقه‌بندی بر اساس درخت است و می‌تواند دانش را به صورت قوانین اگر ـ آن‌گاه، با ایجاد یک قانون برای هر مسیر از ریشه به برگ نشان دهد.[۴ و ۲۵] این الگوریتم در شکل ۲ نشان داده شده است. از درخت تصمیم در مطالعه‌ی شدت تصادف استفاده می‌شود.[۲۵] به طور معمول، این روش‌ها غیرپارامتریک‌اند، یعنی به هیچ نوع تابعی بستگی ندارند و به هیچ دانش احتمالی قبلی در مورد پدیده‌های در حال مطالعه نیاز ندارند.[۲۵]

ازالگوریتم‌های پرکاربرد درخت تصمیم،الگوریتم C۵/۰ است. C۵/۰ توسعه‌یافته‌ی الگوریتم ID۳ است که یک الگوریتم برای ساخت درخت‌های تصمیم‌گیری است.[۳۷ و ۳۸] در این الگوریتم تمام نمونه‌های داده‌های آموزشی در مکان ریشه‌ی درخت قرار می‌گیرد. اگر گرهی دارای هیچ داده‌یی نباشد یا داده‌های یک گره از نوع یکسان باشند، در این صورت گره، یا به عنوان یک گره‌ی برگ خالی یا به عنوان برگی از نمونه‌های یکسان در نظر گرفته می‌شود. اگر گره دارای نمونه‌های بیشتر از یک نوع باشد، باید بتوان طبق یک تابع ارزیابی معین به تمام خصیصه‌های داده‌ها دست یافت و خصیصه‌های مناسب را انتخاب کرد. بنابراین با تقسیم خصیصه‌ها به N قسمت، تعداد N خصیصه نمونه‌های یک گره می‌شوند و هر قسمت به عنوان یک گره جدید است که به گره ریشه متصل شده است. این فرایند، تقسیم گره[۲۲] نامیده می‌شود.[۲۵، ۳۰ و ۳۸]

```
Procedure BuildTree (N, Γ)

1.  If Γ =Φ, then Exit
2.  Let D be the partition associated with node N
3.  Compute the value of the maximum gain of information for a
    feature on D (using a split criterion: SC)
    δ= max SC(C,X)
4.  If δ is lower than or equal to 0 then Exit
5.  Else
        6.  Let Xt be the variable for which the maximum δ is
            attained
        7.  Remove Xt from Γ
        8.  Assing Xt to the node N
        9.  For each possible value xt of Xt
        10. Add a node Nt
        11. Make Nt a child of N
        12. Call BuilTree (Nt , Γ )
```

شکل ۲. الگوریتم ایجاد درخت تصمیم.[۲۵]



شکل ۳. انواع متغیرهای مورد استفاده.

۴. روش تحقیق

همان طورکه اشاره شد، هدف این تحقیق کشف دانش نهفته در داده‌های مرکزکنترل ترافیک تهران با استفاده از روش‌های داده‌کاوی و اتخاذ راهکارهایی برای مدیریت ترافیک‌های سنگین در شهر تهران است. در حقیقت میزان تأثیر تصادفات و شرایط جوی بر ترافیک تهران در مناطق مختلف شهرداری بررسی و طول ترافیک ایجاد شده بر اثر این عوامل در شرایط مختلف با استفاده از تحلیل‌های داده‌کاوی و GIS مدل‌سازی و پیش‌بینی می‌شود. میزان تراکم ترافیک تأثیر مستقیمی بر وقوع تصادف دارد و هرچه تراکم ترافیک بیشتر باشد تصادفات بیشتری اتفاق می‌افتد و برعکس.

اعتبار درونی پژوهش به بررسی این پرسش مربوط می‌شود که آیا متغیرهای مستقل واقعاً در متغیرهای وابسته تغییر ایجاد کرده‌اند یا خیر؟ در داده‌های پایگاه داده ثبت وقایع مرکزکنترل ترافیک تهران، عواملی نظیر تاریخ شروع، زمان شروع، تاریخ پایان، زمان پایان، طول ترافیک، مختصات، نوع روز، منطقه‌ی شهرداری، وضعیت آب و هوایی، عنوان گزارش، موضوع مشاهده، تعداد فوتی‌ها، تعداد مجروحین، هزینه‌ی خسارت، تعداد وسیله‌ی نقلیه درگیر در تصادف، شدت تصادف و غیره ثبت می‌شود. از آن‌جا که هدف این تحقیق پیش‌بینی طول ترافیک ناشی از دو عامل تصادفات و شرایط جوی است، متغیر طول ترافیک به عنوان متغیر مستقل در نظر گرفته می‌شود. از بین سایر پارامترها، با توجه به داده‌ها و گزارش‌های مرکز کنترل ترافیک و با استفاده از مدل‌های آماری، سه پارامتر تعداد وسیله‌ی نقلیه درگیر، شدت تصادف و منطقه شهرداری و سه عامل نوع روز، وضعیت آب و هوایی و منطقه شهرداری که به ترتیب بیشترین تأثیر را در تصادفات و شرایط جوی داشتند، به‌عنوان متغیر وابسته در نظر گرفته شدند. این متغیرها در شکل ۳ نشان داده شده‌اند.

در این تحقیق جامعه آماری کلیه رکوردهای ثبت شده در پایگاه داده سامانه‌ی ثبت وقایع شرکت کنترل ترافیک شهر تهران از تاریخ ۱۳۹۳/۱/۱ الی ۱۳۹۴/۱۲/۲۹ با حجم تقریبی ۱/۸۰۰/۰۰۰ رکورد است. برای رسیدن به اهداف تحقیق، از کلیه‌ی داده‌های موجود در پایگاه داده استفاده شده است و نمونه‌گیری انجام نشده است. به دلیل وجود حجم بالای جامعه (شامل ۴۵۱، ۵۲ رکورد برای تصادفات و ۴۱۴، ۱۱ رکورد برای شرایط جوی) لازم است نحوه‌ی توزیع داده‌ها مشخص شود. طبق قضیه‌ی حد مرکزی (CLT) [۲۳] در نظریه‌ی آمار، داده‌ها دارای توزیع نرمال استاندارد هستند. قضیه‌ی حد مرکزی مرتبط با توزیع مجموع متغیرهای تصادفی مستقل و هم توزیع است. به این ترتیب که با فرض شرایطی خاص، مجموع تعدادی متغیر تصادفی مستقل با توزیع یکسان، که هریک میانگین و واریانس به خوبی تعریف شده دارند، به طور تقریبی دارای توزیع نرمال خواهد بود. هرچه تعداد این متغیرهای مستقل افزایش یابد، این تقریب بهتر می‌شود.

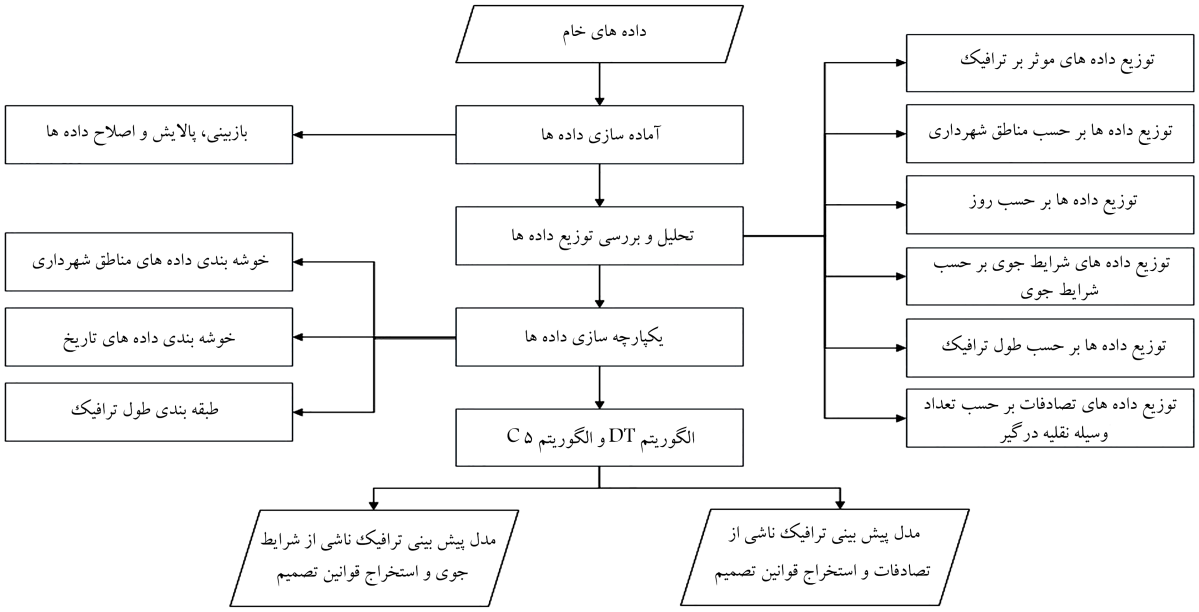
برای استخراج و تمایز داده‌ها از نرم‌افزار مدیریت پایگاه داده Sql Server استفاده شده است. ابتدا دو جدول جداگانه برای تصادفات و شرایط جوی ایجاد شد. سپس داده‌های مورد نیاز از کلیه‌ی جداول پایگاه داده با استفاده از جستجوهای SQL در پایگاه داده استخراج و در جداول متناظر تصادفات و شرایط جوی به طور مجزا ثبت شد. برای دست‌یابی به هدف تحقیق، ابتدا در مرحله‌ی آماده‌سازی داده‌ها، کلیه‌ی رکوردهای ثبت شده در پایگاه داده‌ی سامانه‌ی ثبت وقایع شرکت کنترل ترافیک شهر تهران بازبینی و پالایش و اصلاح می‌شوند. سپس با استفاده از آماره‌های مختلف به بررسی و تحلیل نحوه‌ی توزیع داده‌ها پرداخته می‌شود. در ادامه به منظور یکپارچه‌سازی داده‌ها، تغییراتی در ساختار و فرمت داده‌ها ایجاد می‌شود. بدین صورت که داده‌های مناطق شهرداری از طریق الگوریتم خوشه‌بندی دو مرحله‌یی به ۴ خوشه گروه‌بندی می‌شوند. داده‌های تاریخ نیز به نوع روز (عدی یا تعطیل) خوشه‌بندی و تبدیل می‌شوند. داده‌های طول ترافیک نیز بسته به متراژ به ۵ کلاس تقسیم می‌شوند. این داده‌ها پس از جمع‌بندی، وارد نرم‌افزار SPSS شده و با توجه به هدف‌ها و پرسش‌های تحقیق، از طریق این نرم‌افزار، پردازش و تحلیل می‌شوند. سپس خروجی این مرحله وارد نرم‌افزار رپیدماینر شده و الگوریتم C۵/۰ از الگوریتم دسته‌بندی درخت تصمیم روی داده‌ها اجرا شده و پردازش می‌شود. خروجی این فرایند، دو مدل مجزا به منظور پیش‌بینی طول ترافیک حاصل از تصادفات و شرایط جوی و همچنین استخراج قوانین تصمیم برای هر مدل است. در شکل ۴ مراحل پیشنهادی تحقیق نشان داده شده است.

۵. پیاده‌سازی و نتایج

با توجه به موارد بیان شده در بخش‌های قبل، در این بخش پیاده‌سازی روش پیشنهادی ارائه شده است.

۱.۵. آماده‌سازی داده‌ها

امروزه بانک‌های اطلاعاتی به علت حجم زیاد داده و ارتباط با منابع اطلاعاتی مختلف، گاهی شامل داده‌های متناقض، مفقود و ناقص‌اند. به وسیله‌ی روش‌های آماده‌سازی داده، می‌توان کیفیت داده‌ها و در نتیجه کیفیت نتایج خروجی را افزایش داد. به منظور تسهیل و بهبود فرایند داده‌کاوی آماده‌سازی داده‌ها یکی از مراحل



شکل ۴. مراحل روش پیشنهادی تحقیق.

اساسی تلقی می‌شود.[۳۰] در عمل تقریباً ۸۰ درصد زمان پروژه‌های داده‌کاوی، صرف آماده‌سازی داده‌ها می‌شود.[۳۷] اکثر پژوهش‌های اخیر روی الگوریتم‌های داده‌کاوی صورت گرفته و برعکس تحقیقات کمی روی بهبود روش‌های آماده‌سازی داده‌ها انجام گرفته است. این درحالی است که دستاوردهای زیادی که در مرحله‌ی داده‌کاوی به دست می‌آید، همگی به مرحله‌ی پیش‌پردازش و آماده‌سازی داده‌ها بستگی دارند.[۳۰، ۳۷ و ۳۸]

در مرحله‌ی آماده‌سازی داده‌ها لازم است ابتدا داده‌ها پالایش و پاک‌سازی شوند؛ سپس به صورت یک‌پارچه درآیند. داده‌های موجود در پایگاه داده، به دلیل خطاهای انسانی، نقص و عملکرد ناقص تجهیزات و دستگاه‌ها، دارای داده‌های غلط، ناهمگون یا ناقص بوده‌اند؛ بنابراین لازم بود که با انجام جستجوهای SQL در پایگاه داده، داده‌ها پالایش، تصحیح یا حذف شوند. به طور مثال داده‌های فاقد منطقه‌ی شهرداری برای بررسی تأثیر هر دو عامل تصادفات و شرایط جوی، برای تصادفات داده‌هایی که فاقد تعداد وسیله‌ی درگیر و شدت تصادف بودند و برای شرایط جوی داده‌هایی که فاقد نوع روز و وضعیت آب و هوایی بودند، حذف شدند. در مرحله‌ی حذف داده‌ها برای شناسایی مقادیر پرت، سعی شد تا با ایجاد رابطه‌های منطقی بین مشخصه‌ها، نمونه‌هایی که پاسخ آنها غیر قابل‌اعتماد شناخته شده، از بانک اطلاعاتی خارج شوند. برای مثال اطلاعات ثبت شده در ماه‌های خشک سال یا اطلاعات مربوط به ترافیک روان حذف شدند.

۲.۵. روش‌های آماری تحلیل داده‌ها

داده‌هایی که از پایگاه داده استخراج شده و مورد بازبینی و پالایش قرارگرفته بودند، پس از جمع‌بندی، وارد نرم‌افزار SPSS شده و با توجه به هدف پژوهش، از طریق این نرم‌افزار، محاسبه و تحلیل می‌شوند. آماره‌های مورد استفاده برای تحلیل، شامل آمار توصیفی، یعنی فراوانی، درصد و میانگین و آمار استنباطی شامل آزمون رگرسیون خطی و خوشه‌بندی اکتشافی خواهد بود. در نهایت خروجی داده‌های تغییریافته از مراحل قبل، وارد نرم‌افزار رپیدماینر شده و مدل پیش‌بینی بر اساس درخت تصمیم با الگوریتم C۵/۰ توسط این نرم‌افزار تولید خواهد شد.

۳.۵. تحلیل آمار توصیفی

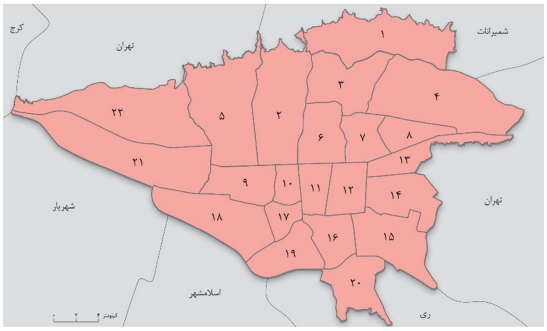
به دلیل وجود حجم بالای جامعه، لازم است نحوه‌ی توزیع داده‌ها مشخص شود. در این بخش از تحلیل آماری به بررسی چگونگی توزیع نمونه‌های آماری پرداخته می‌شود. از کل داده‌های پایگاه داده تعداد ۱/۸۰۰/۰۰۰ رکورد در بازه زمانی پژوهش قرار داشتند. پس از انجام فرایندهای استخراج، بازبینی، اصلاح و پالایش داده‌ها، ۱/۲۹۷/۲۱۷ رکورد دارای اعتبار تشخیص داده شدند.

۱.۳.۵. توزیع داده‌های مؤثر بر ترافیک

اطلاعات و داده‌های مورد استفاده در این تحقیق از مرکز کنترل ترافیک تهران به دست آمده است. کل ۱/۲۹۷/۲۱۷ رکورد در قالب ۱۴ عنوان مشاهده و شنود ثبت شده بودند که درصدهای آنها در جدول ۱ ملاحظه می‌شود.

جدول ۱. مشاهده و شنودهای ثبت شده.

مشاهدات ثبت شده	درصد کل مشاهدات
تصادف	۴/۱۱
تخلف راهنمایی و رانندگی	۱/۵۳
ازدحام مسافر	۰/۱۵
اشیاء مازاددر معابر	۰/۳۹
اطلاع‌رسانی و هماهنگی	۳/۶۰
آب گرفتگی	۰/۷۹
حریق ـ امداد ـ نجات	۰/۳۲
تجهیزات ترافیکی	۰/۲۱
موارد هماهنگی با شرکت برق	۰/۰۸
موارد اجرائی	۵/۱۱
اختلاف سطح و ناهمواری	۰/۰۹
نقص فنی خودرو	۸/۸۹
شرایط جوی	۰/۹۸
ثبت وضعیت ترافیکی	۷۳/۶۶



شکل ۵. نقشه مناطق ۲۲ گانه تهران.

با توجه به درصد داده‌ها، مشخص شد که بخش اعظم ترافیک تهران، ناشی از حجم زیاد رفت و آمدهای عادی روزمره بوده و عامل مهم دیگری به‌جز حجم تردد در آن دخیل نیست.

۲.۳.۵. توزیع داده‌ها برحسب مناطق شهرداری

شهر تهران دارای ۲۲ منطقه شهرداری است. در شکل ۵ نقشه‌ی مناطق ۲۲ گانه‌ی تهران نشان داده شده است.

پس از بارگذاری داده‌های پالایش شده در نرم‌افزار SPSS، نمودارهای زیر برای توزیع داده‌های ترافیک و شرایط جوی در مناطق شهرداری تهران به دست آمده است. توزیع داده‌های تصادفات و شرایط جوی در مناطق شهرداری تهران مطابق شکل ۶ است.

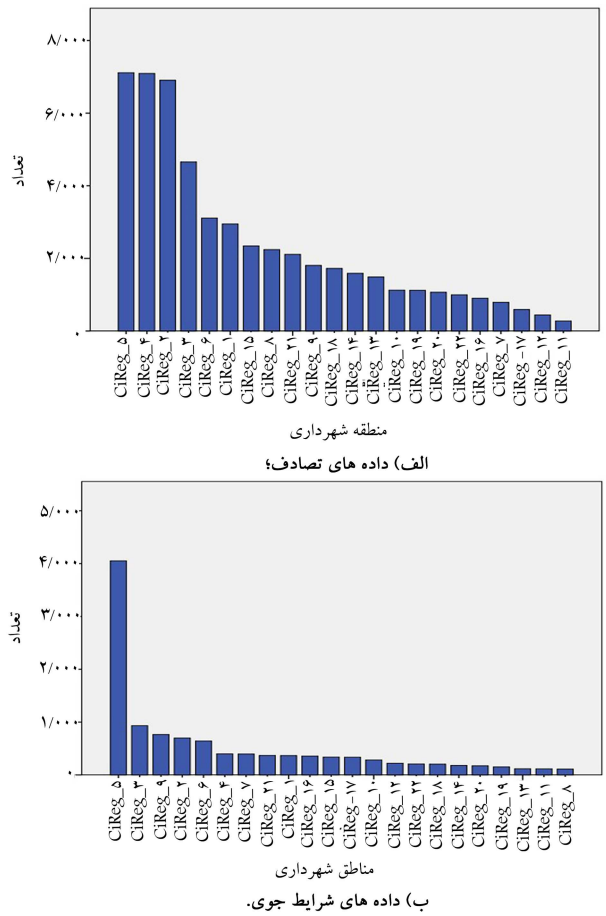
طبق نمودار در مناطق ۵، ۴ و ۲ شهر تهران، بیشترین تعداد تصادفات ثبت شده است که بین ۶۰۰۰ تا ۸۰۰۰ مرتبه بوده است و کم‌ترین تعداد رکورد تصادفات ثبت شده در مناطق ۹، ۱۸، ۱۴، ۱۳، ۱۰، ۱۹، ۲۰، ۲۲، ۱۶، ۷، ۱۷، ۱۲ و ۱۱، بوده که کم‌تر از ۲۰۰۰ سانحه بوده است. حجم تردد زیاد، وجود بزرگراه‌ها و خیابان‌های مهم و اصلی، وسعت زیاد منطقه و وجود دوربین‌ها و سیستم‌های ترددشمار بیشتر از مهم‌ترین عوامل ثبت اطلاعات تصادفی بیشتر در مناطق ۵، ۴ و ۲ است. همچنین در منطقه ۵ تهران تعداد ۴۰۰۰ رکورد مشاهده شرایط جوی ثبت شده است؛ درحالی که در تمامی مناطق دیگر تعداد این عارضه کم‌تر از ۱۰۰۰ رکورد است. علت این امر نیز وجود ایستگاه‌های هواشناسی بیشتر در این منطقه است.

۳.۳.۵. توزیع داده‌ها برحسب روز

بررسی داده‌های به دست آمده از مرکزکنترل ترافیک تهران نشان می‌دهد، ۹۰/۰۸٪ از تصادفات در روزهای عادی و ۹/۹۲٪ از تصادفات در روزهای تعطیل اتفاق افتاده‌اند. همچنین میزان ترافیک حاصل از شرایط جوی، در روزهای غیر تعطیل ۹۰/۸۷٪ بوده و در روزهای تعطیل ۹/۱۳٪ بوده است. برای نرمال کردن درصدهای ارائه شده، لازم است این مقادیر به تعداد روزهای عادی و تعطیل در مدت دوساله‌ی داده‌های تحقیق، تقسیم شوند. به طورکلی تعداد روزهای عادی و تعطیل از مجموع ۷۳۰ روز در سال‌های ۱۳۹۳ و ۱۳۹۴ به ترتیب برابر ۱۴۹ و ۵۸۱ روز است. بنابراین مقادیر نرمال شده با توجه به رابطه‌ی ۳ عبارت است از:

$$N'_a(x) = \frac{N_a(x)}{n(x)} \qquad (۳)$$

در این رابطه a نوع پارامتر مؤثر در ترافیک (تصادفات یا شرایط جوی)، x نوع روز (عادی یا تعطیل)، N تعداد مشاهدات ثبت شده از پارامتر مؤثر بر ترافیک، n تعداد روزهای نوع x در مدت تحقیق و N' مقدار نرمال شده هر پارامتر است. به عبارت دیگر، تعداد مشاهدات ثبت شده از پارامتر مؤثر بر ترافیک در یک نوع روز مشخص بر



الف) داده های تصادف؛

ب) داده های شرایط جوی.

شکل ۶. توزیع داده‌ها در مناطق شهرداری تهران.

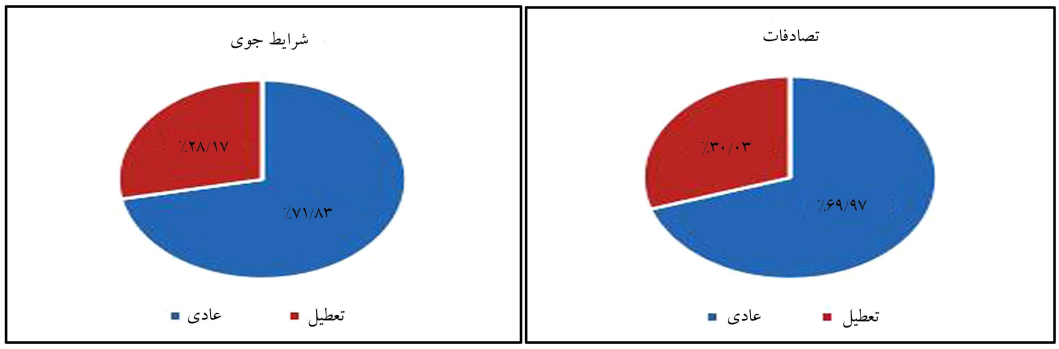
تعداد روزهای آن تقسیم می‌شود. پس از بارگذاری داده‌های پالایش شده در نرم‌افزار SPSS، نمودارهای زیر برای توزیع داده‌های ترافیک و شرایط جوی به صورت نرمال برحسب نوع روز به دست آمد. توزیع داده‌های تصادفات و شرایط جوی به صورت نرمال برحسب روز مطابق شکل ۷ است.

مطابق نمودار، میزان ترافیک حاصل از تصادفات در یک روز عادی ۲/۳۳ برابر یک روز تعطیل است. این مقدار در روز عادی برابر ۶۹/۹۷٪ است، در حالی که در روز تعطیل به ۳۰/۰۳٪ کاهش می‌یابد. همچنین میزان ترافیک حاصل از شرایط جوی در یک روز عادی ۲/۵۵ برابر یک روز تعطیل است. این مقدار در روز عادی برابر ۷۱/۸۳٪ و در روز تعطیل برابر ۲۸/۱۷٪ است. علت این امر کاهش حجم تردد در روزهای تعطیل به دلیل تعطیلی مراکز آموزشی، دولتی و خصوصی و غیره است.

۴.۳.۵. توزیع داده‌های شرایط جوی برحسب شرایط جوی

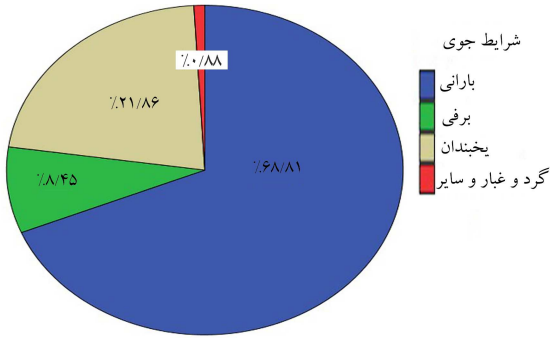
پس از بارگذاری داده‌های پالایش شده در نرم‌افزار SPSS، در شکل ۸ برای توزیع داده‌های شرایط جوی برحسب شرایط جوی به دست آمد.

طبق نمودار ۶۸/۸۱٪ ترافیک در شرایط بارانی ثبت شده است. در شرایط برفی نیز ۸/۴۵٪ میزان ترافیک گزارش شده؛ در موقعیت یخبندان با کاهش سرعت حرکت و افزایش دقت در رانندگی، ۲۱/۸۶٪ ترافیک ثبت شده است. در زمان گرد و غبار و سایر شرایط آب و هوایی تنها ۰/۸۸٪ ترافیک ثبت شده است. با توجه به نمودار فوق می‌توان به این نتیجه رسید که در زمان بارندگی میزان ترافیک بیشتر از شرایط دیگر جوی افزایش پیدا می‌کند. این شرایط به‌ویژه در شهر تهران به دلیل



الف) داده های تصادف؛ ب) داده های شرایط جوی.

شکل ۷. توزیع داده‌ها به صورت نرمال برحسب روز.



شکل ۸. توزیع داده‌های شرایط جوی برحسب شرایط جوی.

داشتن شرایط و فرهنگ خاص رانندگی، خودروهای تک‌سرنشین فراوان، فرهنگ استفاده از خودرو در شرایط نامساعد جوی، شیب زیاد برخی خیابان‌ها و لغزندگی معابر بسیار چشمگیر است. با توجه به همین موضوع می‌توان به دنبال راه‌حل مناسبی در جهت کمک به کاهش میزان ترافیک از طریق پیش‌بینی طول ترافیک حاصل از شرایط نامساعد جوی بود. همان‌طور که اشاره شد، یکی از اهداف این تحقیق نیز پیش‌بینی طول ترافیک در شرایط مختلف جوی است.

۵.۳.۵. توزیع داده‌ها برحسب طول ترافیک (TS)

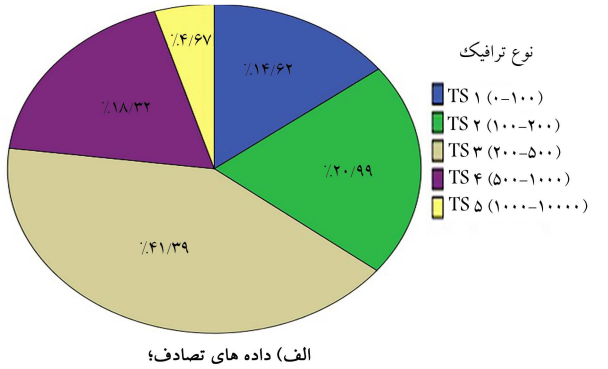
پس از بارگذاری داده‌های پالایش شده در نرم‌افزار SPSS، نمودارهای زیر برای توزیع داده‌های ترافیک و شرایط جوی برحسب طول ترافیک به دست آمده است. توزیع داده‌های تصادفات و شرایط جوی برحسب طول ترافیک مطابق شکل ۹ است.

طبق شکل ۹، ۱۴/۶۲٪ از تصادفات باعث افزایش طول ترافیک از صفر تا ۱۰۰ متر و در نتیجه افزایش ترافیک هستند. تصادفاتی که ۲۰۰ تا ۵۰۰ متر بر میزان ترافیک تأثیر می‌گذارند ۴۱/۳۹٪ هستند؛ که البته این میزان بیشترین مقدار را به خودش اختصاص می‌دهد.

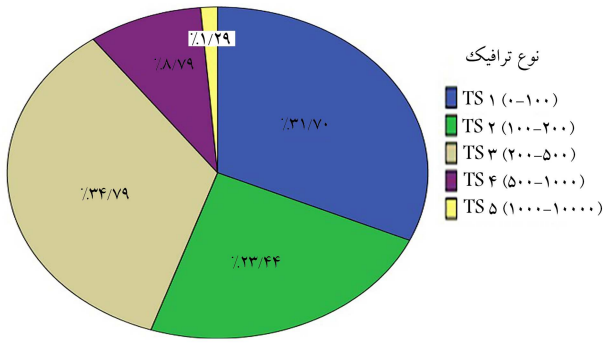
۶.۳.۵. توزیع داده‌های تصادفات برحسب تعداد وسیله‌ی نقلیه درگیر (VIC)

پس از بارگذاری داده‌های پالایش شده در نرم‌افزار SPSS، در شکل ۱۰ برای توزیع داده‌های تصادف برحسب تعداد وسیله‌ی نقلیه درگیر در تصادف به دست آمد.

طبق نمودار بیشترین میزان تصادفات که بین ۳۰،۰۰۰ تا ۴۰،۰۰۰ رکورد بوده است، ۲ وسیله‌ی نقلیه درگیر در تصادف شناخته شده‌اند. همچنین در کم‌تر از ۱۰،۰۰۰ رکورد از تصادفات، ۱ وسیله‌ی نقلیه و بیشتر از ۲ وسیله‌ی نقلیه درگیر در تصادفات شناسایی شده‌اند. بنابراین می‌توان چنین استنتاج کرد که تصادفات

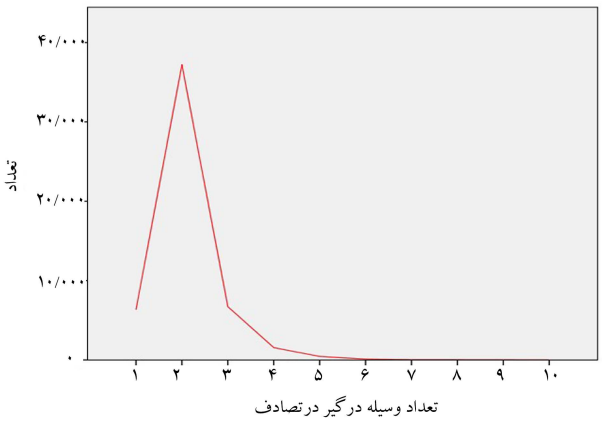


الف) داده های تصادف؛



ب) داده های شرایط جوی.

شکل ۹. توزیع داده‌ها برحسب طول ترافیک.



شکل ۱۰. توزیع داده‌های تصادف برحسب تعداد وسیله.

جدول ۲. تغییرات لازم در ساختار و یکپارچه‌سازی داده‌ها.

قلم توصیفی	ساختار ابتدایی داده	ساختار تغییریافته داده
زمان	تاریخ	۲ خوشه (روز عادی یا روز تعطیل)
مناطق شهرداری	اسم منطقه (در مناطق ۲۲ گانه)	۴ خوشه مناطق شهرداری
طول ترافیک	متراژ طول ترافیک	۵ طبقه طول ترافیک

زنجیره‌یی (برخورد همزمان تعداد زیادی خودرو به هم) وقتی پیش می‌آید که سرعت خودروها زیاد باشد. بنابراین در شرایط تهران که غالبا با ترافیک سنگین همراه است، این وضعیت کم‌تر پیش می‌آید. همچنین تصادف تک‌خودرویی معمولاً در شرایط سرعت زیاد و برخورد خودرو به گاردریل و موانع خیابانی پیش می‌آید که باز هم در شرایط تهران با ترافیک‌های غالبا سنگین کم‌تر رخ می‌دهد. در این حالت تصادف دو خودرویی بیشترین احتمال رخداد را دارد.

۴.۵. یک‌پارچه‌سازی داده‌ها

به دلیل تعداد و تنوع زیاد داده‌ها شامل تاریخ با تنوع زیاد در روز، ماه و سال در مدت تحقیق، مناطق شهرداری که خود شامل ۲۲ عنوان است و طول ترافیک که به صورت پیوسته از صفر تا چند صد متر است و همچنین قابلیت پردازش بعضی داده‌ها در مدل، لازم است داده‌های داخل برخی اقلام توصیفی با یکدیگر یک‌پارچه شوند. بنابراین باید ساختار و فرمت داده‌ها تغییر یافته و یک شکل شوند. در جدول ۲، اقلامی توصیف شده‌اند که در این مرحله روی آنها تغییراتی صورت گرفته است.

۵.۵. خوشه‌بندی داده‌ها

قلم توصیفی مربوط به مناطق شهرداری برای هریک از پرسش‌ها (تصادف و شرایط جوی)، هرکدام دارای ۲۲ عنوان هستند. به دلیل تعداد زیاد این داده و گستردگی درخت ایجاد شده به سبب این قلم توصیفی، لازم است داده‌ی منطقه‌ی شهرداری از طریق یک تحلیل آماری، کاهش‌یافته و به صورت گروه‌بندی ارائه شود. بدین منظور، برای جلوگیری از بی‌تأثیر شدن این قلم، مناسب‌ترین روش، اعمال خوشه‌بندی بود. خوشه‌بندی قلم توصیفی فوق برحسب سایر اقلام توصیفی نظیر زمان (نوع روز)، وضعیت آب و هوایی، شدت تصادف و تعداد وسیله‌ی نقلیه درگیر در تصادف، باعث می‌شد تا مناطقی که رفتار مشابه دارند در یک گروه قرار بگیرند. این فرایند از طریق تحلیل خوشه‌بندی دومرحله‌یی به کمک نرم‌افزار SPSS روی داده‌ها انجام شد. جداول ۳ و ۴ به ترتیب گروه‌بندی مناطق برای تصادفات و شرایط جوی را نشان می‌دهند.

جدول ۳. گروه‌بندی مناطق برای تصادفات.

عنوان گروه	مناطق شهرداری
گروه منطقه‌یی GC۱	۱، ۱۴، ۱۶، ۲، ۲۰، ۲۱، ۴، ۷، ۹
گروه منطقه‌یی GC۲	۱۰، ۱۱، ۱۲، ۱۸، ۱۹
گروه منطقه‌یی GC۳	۱۵، ۳، ۶
گروه منطقه‌یی GC۴	۱۳، ۱۷، ۲۲، ۵، ۸

جدول ۴. گروه‌بندی مناطق برای شرایط جوی.

عنوان گروه	مناطق شهرداری
گروه منطقه‌یی GC۱	۱۱، ۱۲، ۱۳، ۲۰، ۳، ۴، ۹
گروه منطقه‌یی GC۲	۱۰، ۱۶، ۱۷، ۵، ۶، ۷، ۸
گروه منطقه‌یی GC۳	۱۴، ۱۸، ۲، ۲۲
گروه منطقه‌یی GC۴	۱، ۱۵، ۱۹، ۲۱

جدول ۵. طبقه‌بندی طول ترافیک.

عنوان	طول
TS۱	۱۰۰-۰
TS۲	۲۰۰-۱۰۰
TS۳	۵۰۰-۲۰۰
TS۴	۱۰۰۰-۵۰۰
TS۵	۱۰۰۰۰-۱۰۰۰

۶.۵. طبقه‌بندی داده طول ترافیک

داده طول ترافیک ایجاد شده برای هر دو پرسش پژوهش (تصادف و شرایط جوی)، داده‌یی پیوسته و از نوع اعداد حقیقی است. این نوع داده برای تحلیل‌هایی که منجر به تولید درخت تصمیم می‌شوند، مناسب نیست. برای طبقه‌بندی این داده و تبدیل آن به داده‌ی نوع کلاس، از پژوهش‌های صورت گرفته و اطلاعات مدیران مهندسی ترافیک استفاده شد. بر این اساس مشخص شد ترافیک حاصل از هر پارامتر مؤثر بر ترافیک (نظیر تصادفات یا شرایط جوی) را در طول‌های ترافیکی مشخص، تجزیه و تحلیل می‌کنند. داده‌ی به دست آمده از پایگاه داده‌ی نرم‌افزار ثبت وقایع نیز برحسب طبقه‌بندی ارائه شده توسط مهندسان ترافیک مطابق جدول ۵ طبقه‌بندی شد و نتایج در فایل خروجی نرم‌افزار SPSS و ورودی نرم‌افزار رپیدماینر اعمال شد.

۷.۵. آزمون فرضیات پژوهش

برای هریک از پرسش‌های پژوهش چند فرضیه به صورت جداگانه تعریف شد. ابتدا باید صحت و تأثیرگذاری هریک از فرضیه‌ها بر پرسش متناظر مورد بررسی قرار گیرد. برای بررسی صحت فرضیه و به دست آوردن میزان سطح معناداری، از آزمون رگرسیون خطی به کمک نرم‌افزار SPSS استفاده شد که در ادامه به شرح نتایج به دست آمده خواهیم پرداخت.

۱.۷.۵. آزمون فرضیه‌های پرسش اول (تأثیر تصادف بر ترافیک)

برای پرسش اول، چهار فرضیه مطرح شد که سه فرضیه‌ی آن تأیید و یک فرضیه رد شد. نتیجه‌ی بررسی در جدول ۶ منعکس شده است.

خروجی آزمون رگرسیون خطی چند متغیره از نرم‌افزار SPSS به شرح جدول ۷ است.

با توجه به نتایج مشاهده می‌شود که مقدار Sig مربوط به آماره کل که بیان‌گر معنی‌دار بودن کل رگرسیون است، برابر ۰/۰۰۰ بوده و حاکی از آن است که مدل در

جدول ۶. فرضیات پرسش اول.

فرضیه	نتیجه
H۰: شدت تصادف، نوع روز، منطقه‌ی شهرداری و تعداد وسیله بر میزان ترافیک ناشی از تصادف تأثیر ندارند.	عدم تأیید
H۱: شدت تصادف بر میزان ترافیک ناشی از تصادف تأثیر دارد.	تأیید
H۲: نوع روز بر میزان ترافک ناشی از تصادف تأثیر دارد.	عدم تأیید
H۳: منطقه شهرداری بر میزان ترافیک ناشی از تأثیر دارد.	تأیید
H۴: تعداد وسیله‌ی درگیر در تصادف بر میزان ترافیک ناشی از تصادف تأثیر دارد.	تأیید

جدول ۷. خروجی تحلیل رگرسیون خطی برای فرضیات پرسش اول.

		ضرایب				
	مدل	ضرایب غیر استاندارد		ضرایب استاندارد	t	sig.
		B	Std.Error	Beta		
I	(Constant)	۲۰۷/۴۴۴	۸/۳۲۵		۲۴/۹۱۸	۰/۰۰۰
	CrashSeverityID	۱۳/۹۲۴	۳/۹۸۲	۰/۰۱۶	۳/۴۹۶	۰/۰۰۰
	DayTypeID	۲/۵۵۹	۴/۳۵۷	۰/۰۰۳	۰/۵۸۷	۰/۵۵۷
	CityZoneID	۴/۷۲۴	۰/۲۰۹	۰/۰۹۸	۲۲/۶۵۱	۰/۰۰۰
	VehiclesInvolvedCount	۴۶/۱۱۹	۱/۸۷۷	۰/۱۰۹	۲۴/۵۷۷	۰/۰۰۰

جدول ۸. فرضیات پرسش دوم.

فرضیه	نتیجه
H۰: نوع شرایط جوی، نوع روز و منطقه شهرداری بر میزان ترافیک ناشی از شرایط جوی تأثیر ندارد.	عدم تأیید
H۱: نوع شرایط جوی بر میزان ترافیک ناشی از شرایط جوی تأثیر دارد.	تأیید
H۲: نوع روز بر میزان ترافیک ناشی از شرایط جوی تأثیر دارد.	تأیید
H۳: منطقه شهرداری بر میزان ترافیک ناشی از شرایط جوی تأثیر دارد.	تأیید

جدول ۹. خروجی تحلیل رگرسیون خطی برای فرضیات پرسش دوم.

		ضرایب				
	مدل	ضرایب غیر استاندارد		ضرایب استاندارد	t	sig.
		B	Std.Error	Beta		
I	(Constant)	۲۳۲/۵۱۷	۱۰/۹۱۴		۲۱/۳۰۵	۰/۰۰۰
	WeatherID	۶/۳۸۷	۲/۴۸۷	۰/۰۲۴	۲/۵۶۸	۰/۰۱۰
	DayTypeID	-۱۸/۷۷۸	۷/۳۹۰	۰/۰۰۲۴	-۲/۵۴۱	۰/۰۱۱
	CityZoneHD	۱/۱۷۸	۰/۳۸۰	۰/۰۲۹	۳/۱۰۳	۰/۰۰۲

سطح اطمینان ۹۵ درصد معنادار است. همچنین متغیر «وضعیت روز» تأثیر معناداری بر متغیر وابسته ندارد چون سطح معناداری آن بیشتر از ۰/۰۵ است ولی بقیه متغیرهای مستقل (CrashSeverityID، CityZoneID و VehiclesInvolvedCount) چون سطح معناداری کم‌تر از ۰/۰۵ دارند، اثر معناداری بر متغیر وابسته دارند.

۲.۷.۵. آزمون فرضیات پرسش دوم

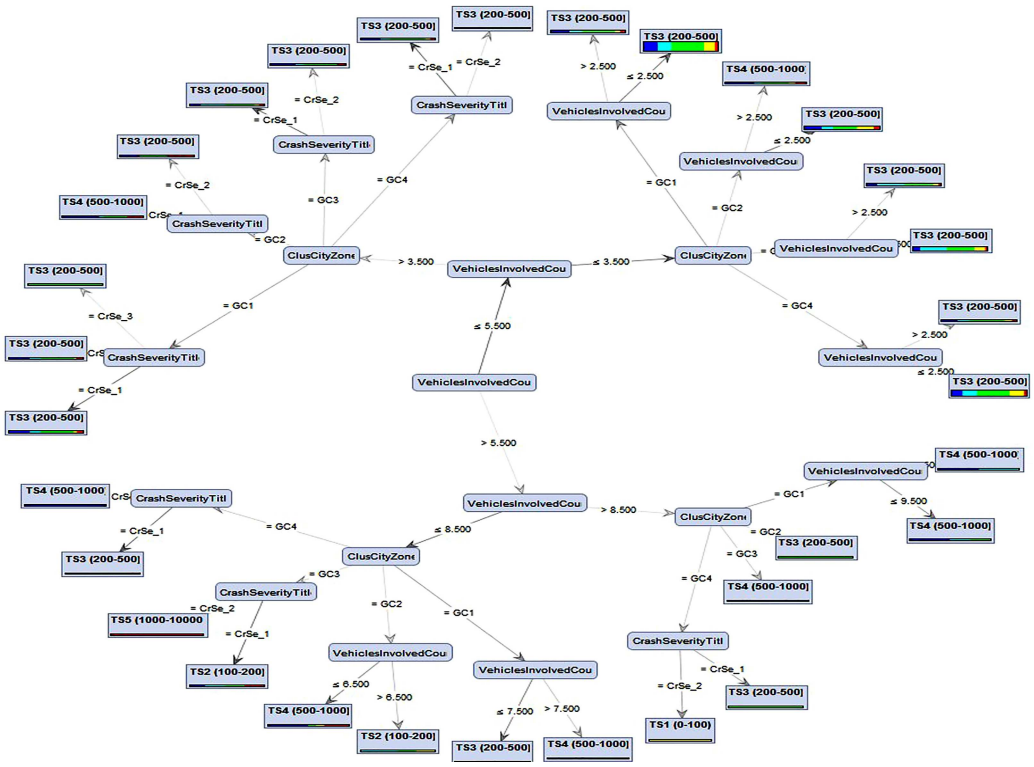
برای پرسش دوم، سه فرضیه مطرح شد که هر سه تأیید شدند. نتیجه‌ی بررسی در جدول ۸ منعکس شده است.

خروجی آزمون رگرسیون خطی چند متغیره از نرم‌افزار SPSS به شرح جدول ۹ است.

با توجه به نتایج مشاهده می‌شود که مقدار Sig مربوط به آماره کل برابر صفر بوده و حاکی از آن است که مدل در سطح اطمینان ۹۵ درصد معنادار است. همچنین تمامی متغیرهای مستقل (WeatherID، DayTypeID و CityZoneID) چون سطح معناداری کم‌تر از ۰/۰۵ دارند اثر معناداری بر متغیر وابسته دارند.

۳.۷.۵. ارائه مدل

پس از انجام آزمون بررسی فرضیات پژوهش و اصلاح داده‌ها، اطلاعات نهایی شده در یک فایل خروجی استخراج و برای بارگذاری در نرم‌افزار رپیدماینر آماده شد. رپیدماینر یکی از پرکاربردترین و آسان‌ترین نرم‌افزارها در مباحث داده‌کاوی، یادگیری ماشینی، تحلیل و پیش‌بینی و تحلیل کسب و کار به منظور آنالیز داده‌ها به صورت عددی و نمایش گرافیکی نتایج حاصل از تحلیل داده‌هاست. متن باز بودن، استفاده ازکتابخانه‌های آن برای برنامه‌نویسان، امکان به کارگیری محیط گرافیکی بدون داشتن



شکل ۱۱. مدل وضعیت ترافیک ناشی از تصادفات.

دانش برنامه‌نویسی و همچنین تنوع الگوریتم‌های آماده‌سازی و مدل‌سازی موجب شده طیف متنوعی از کاربران به این ابزار متمایل شوند. با کمک این برنامه می‌توان تمامی مراحل مورد نیاز شامل آماده‌سازی اطلاعات اولیه، مدل‌سازی، پیش‌بینی، بصری‌سازی نتایج، ارزیابی و اعتبارسنجی و بهینه‌سازی خروجی را در یک محیط یکپارچه و واحد انجام داد. درکنار قابلیت‌های تحلیلی خود، این نرم‌افزار امکاناتی برای ترکیب داده‌ها، تغییر آنها و نیز یادگیری داده توسط آنها را فراهم می‌آورد. همچنین این نرم‌افزار قابلیت ورود کدهای داده‌کاوی از سایر برنامه‌ها مانند R و Python را دارد و می‌تواند از داده‌های سایر نرم‌افزارهای داده‌کاوی مانند Microsoft Excel،Microsoft SQL Server Teradata ,IBM DB ,Access Oracle نیز استفاده کند.

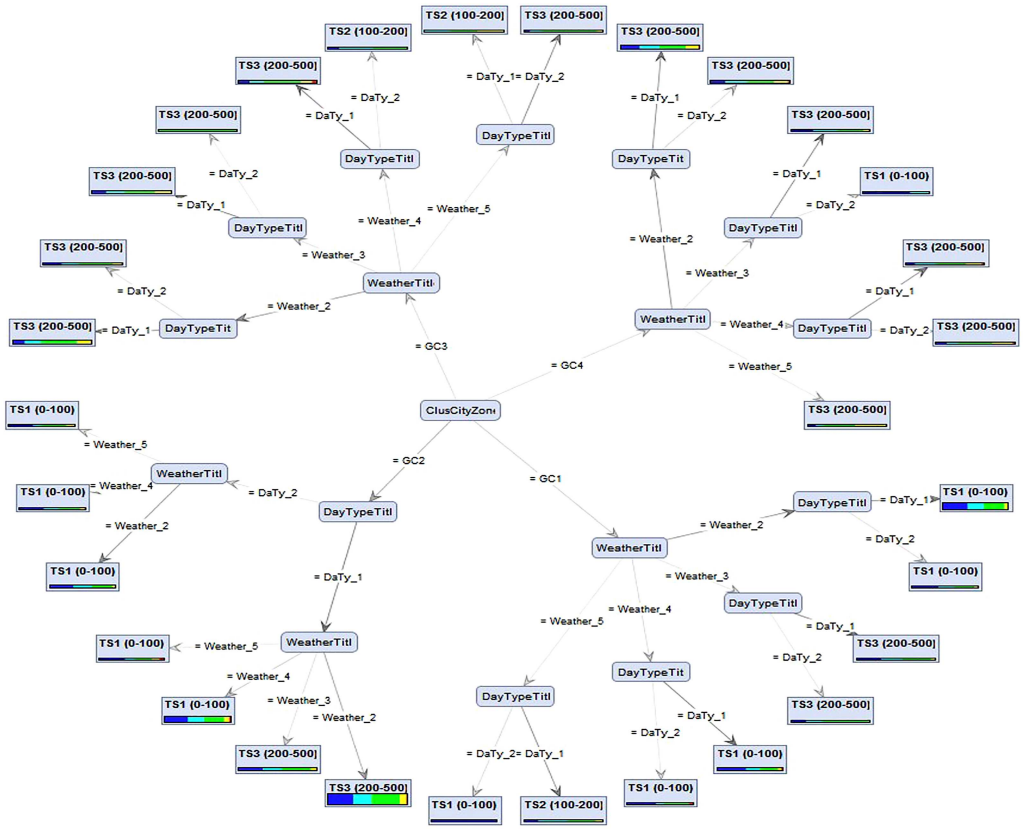
در این تحقیق خروجی نرم‌افزار SPSS شامل یکپارچه‌سازی، خوشه‌بندی و طبقه‌بندی داده‌ها که شرح مفصل آن در بخش‌های قبلی بیان شد، وارد نرم‌افزار رپیدماینر می‌شود. سپس با استفاده از درخت تصمیم و الگوریتم C۵/۰ داده‌ها دسته‌بندی شده و دو مدل مجزا برای وضعیت ترافیک ناشی از تصادفات و وضعیت ترافیک حاصل از شرایط جوی ارائه می‌شود. سپس با استفاده از قوانین استخراج شده برای هر مدل، می‌توان طول ترافیک را در شرایط مختلف پیش‌بینی کرد. در ادامه این دو مدل به همراه قوانین استخراج شده از آنها ارائه شده است. در این مدل‌ها و قوانین، VIC به تعداد وسیله‌ی نقلیه درگیر در تصادف، گروه به خوشه‌بندی مناطق ۲۲ گانه شهرداری تهران مطابق جداول ۳ و ۴، TS به طول ترافیک ایجاد شده مطابق جدول ۵، GST به شدت تصادف که شامل دو مقدار ۱ برای تصادفات منجر به فوت و ۲ برای تصادفات منجر به جراحت، DT به نوع روزکه شامل دو مقدار ۱ و ۲ به ترتیب برای روز عادی و روز تعطیل است، WT به نوع شرایط جوی که دارای ۵ مقدار ۱، ۲، ۳، ۴ و ۵ به ترتیب برای شرایط جوی بارانی، برفی، یخبندان، گرد و غبار و سایر است، اشاره دارد.

۴.۷.۵. مدل وضعیت ترافیک ناشی از تصادفات

تنظیمات در نرم‌افزار رپیدماینر انجام و درخت مدل برای وضعیت ترافیک بر اثر تصادفات مطابق شکل ۱۱ ایجاد شد.

قوانین استخراج شده از مدل ترافیک ناشی از تصادفات عبارت‌اند از:

- اگر $VIC > ۸$ (۹ یا بیشتر)، گروه ۱ و گروه ۳، تمامی حالت‌های شدت تصادف، آنگاه طول ترافیک برابر TS۴.
- اگر $VIC > ۸$ (۹ یا بیشتر)، گروه ۲، تمامی حالت‌های شدت تصادف، آنگاه طول ترافیک برابر TS۳.
- اگر $VIC > ۸$ (۹ یا بیشتر)، گروه ۴، شدت تصادف GST=۱، آنگاه طول ترافیک برابر TS۳؛ شدت تصادف GST=۲، آنگاه طول ترافیک برابر TS۱.
- اگر $۵ < VIC <= ۸$ ،(۶، ۷ یا ۸)، گروه ۱ و $VIC > ۷$ (برابر ۸)، آنگاه طول ترافیک برابر TS۴؛ $VIC <= ۷$، آنگاه طول ترافیک برابر TS۳.
- اگر $۵ < VIC <= ۸$ ، (۶، 7 یا۸)، گروه ۳، شدت تصادف GST=۱، آنگاه طول ترافیک برابر TS۲؛ شدت تصادف GST=۲، آنگاه طول ترافیک برابر TS۵.
- اگر $۵ < VIC <= ۸$، (۶، ۷ یا ۸)، گروه ۴، شدت تصادف GST=۱، آنگاه طول ترافیک برابر TS۳؛ شدت تصادف GST=۲، آنگاه طول ترافیک برابر TS۴.
- اگر $۳ < VIC <= ۵$ (۴ یا ۵)، گروه ۱ و گروه ۳ و گروه ۴، تمامی حالت‌های شدت تصادف، آنگاه طول ترافیک برابر TS۲.
- اگر $۳ < VIC <= ۵$ (۴ یا ۵)، گروه ۲، شدت تصادف GST=۱، آنگاه طول ترافیک برابر TS۴؛ شدت تصادف GST=۲، آنگاه طول ترافیک برابر TS۵.



شکل ۱۲. مدل وضعیت ترافیک ناشی از شرایط جوی.

- اگر $VIC <= ۳$، (۱، ۲ یا ۳)، گروه ۱ و گروه ۳، آنگاه طول ترافیک برابر TS۳.
- اگر $VIC <= ۳$ ، (۱، ۲ یا ۳)، گروه $VIC > ۲،۲$ (برابر ۳)، آنگاه طول ترافیک برابر TS۴؛ $VIC <= ۲$ (برابر ۱ یا۲)، آنگاه طول ترافیک برابر TS۳.
- اگر $VIC <= ۳$ ، (۱، ۲ یا ۳)، گروه $VIC > ۲،۴$ (برابر ۳)، آنگاه طول ترافیک برابر TS۳؛ $VIC <= ۲$ (برابر ۱ یا ۲)، آنگاه طول ترافیک برابر TS۱.

به طور کلی منطقه‌ی شهرداری، تعداد وسیله‌ی نقلیه‌ی درگیر در تصادف و شدت تصادف بر ترافیک ناشی از تصادفات تأثیر دارد ولی نوع روز بر ترافیک ناشی از تصادفات تأثیر ندارد. با توجه به قوانین می‌توان نتیجه گرفت که در مناطق شلوغ‌تر و با حجم تردد بیشتر، هرچه تعداد وسایل نقلیه درگیر در تصادف بیشتر باشد، طول ترافیک بیشتر خواهد بود. نکته‌ی جالب توجه این که تصادفات منجر به جراحت، موجب ایجاد ترافیک‌های طویل‌تر می‌شود که از دلایل این امر می‌توان به انتظار برای حضور نیروهای امدادی و آمبولانس برای کمک به مجروحین اشاره کرد. وجود توأمان هر دو عامل یعنی بیشتر بودن تعداد وسایل نقلیه درگیر در تصادف و تصادفات منجر به جراحت، طول ترافیک را بیشتر می‌کند. درحالی که در مناطق دورتر از مرکز شهر که از حجم ترددهای روزانه برای حضور در محل کار کاسته می‌شود، تصادفات به ترافیک کم‌تری می‌انجامد. همچنین مناطق با ساختار شهری نامناسب از نظر نوع خیابان‌ها، ترافیک طویل‌تر و سنگین‌تری مشاهده می‌شود. به علاوه تصادفات با بیشینه ۲ وسیله‌ی نقلیه‌ی درگیر در تصادف، باعث ایجاد ترافیک‌های با طول بیشینه ۱۰۰ متر می‌شود. همچنین نتایج نشان می‌دهد بیشترین میزان طول ترافیک ایجاد شده برابر TS۳ یعنی ۲۰۰-۵۰۰ متر است.

۵.۷.۵. مدل وضعیت ترافیک ناشی از عوامل جوی

تنظیمات در نرم‌افزار رپیدماینر انجام و درخت مدل برای وضعیت ترافیک بر اثر شرایط جوی مطابق شکل ۱۲ ایجاد شد.

قوانین استخراج شده از مدل ترافیک ناشی از شرایط جوی عبارت‌اند از:

- اگر گروه ۱، شرایط جوی WT=۲ و WT=۴، در هر دو نوع روز، آنگاه طول ترافیک برابر TS۱.
- اگر گروه ۱، شرایط جوی WT=۳، در هر دو نوع روز، آنگاه طول ترافیک برابر TS۳.
- اگر گروه ۱، شرایط جوی WT=۵، روز DT=۱، آنگاه طول ترافیک برابر TS۲؛ روز DT=۲، آنگاه طول ترافیک برابر TS۱.
- اگر گروه ۲، روز DT=۱، شرایط جوی WT=۲ و WT=۳، آنگاه طول ترافیک برابر TS۳؛ شرایط جوی WT=۴، آنگاه طول ترافیک برابر TS۴؛ شرایط جوی WT=۵، آنگاه طول ترافیک برابر TS۱.
- اگر گروه ۲، روز DT=۲، شرایط جوی WT=۲ و WT=۴ و WT=۵، آنگاه طول ترافیک برابر TS۱.
- اگر گروه ۳، شرایط جوی WT=۲ و WT=۳، در هر دو نوع روز، آنگاه طول ترافیک برابر TS۳.
- اگر گروه ۳، شرایط جوی WT=۴، روز DT=۱، آنگاه طول ترافیک برابر TS۳؛ روز DT=۲، آنگاه طول ترافیک برابر TS۲.
- اگر گروه ۳، شرایط جوی WT=۵، روز DT=۱، آنگاه طول ترافیک برابر TS۲؛ روز DT=۲، آنگاه طول ترافیک برابر TS۳.

- اگر گروه ۴، شرایط جوی WT=۲ و WT=۳ و WT=۵، در هر دو نوع روز، آن‌گاه طول ترافیک برابر TS۳.
- اگر گروه ۴، شرایط جوی WT=۳، روز DT=۱، آن‌گاه طول ترافیک برابر TS۳؛ روز DT=۲، آن‌گاه طول ترافیک برابر TS۱.

به طور کلی منطقه‌ی شهرداری، نوع شرایط جوی و نوع روز بر ترافیک ناشی از شرایط جوی تأثیر می‌گذارد. با توجه به قوانین می‌توان نتیجه گرفت در مناطقی که ساختار شهری خیلی مناسب نیست، شرایط جوی نامساعد نظیر بارش برف و باران و یخبندان موجب آب‌گرفتگی معابر و سختی و کندی حرکت وسایل نقلیه و در نتیجه ایجاد ترافیک‌های بسیار طویل‌تر نسبت به مناطقی که ساختار شهری مناسب‌تری دارند، می‌شود. این مورد در مناطقی نظیر مرکز شهر که حجم تردد بیشتر است نیز مشاهده می‌شود. از طرفی، شرایط نامساعد جوی در این مناطق در روزهای عادی به دلیل حجم زیاد ترددهای روزانه، نقش بیشتری در ایجاد ترافیک دارد. هرچند در برخی مناطق که ویژگی‌های تفریحی و جاذبه‌های گردشگری دارند، طول ترافیک در روزهای تعطیل بیشتر از روزهای عادی است. همچنین نتایج نشان می‌دهد بیشترین میزان طول ترافیک ایجاد شده برابر TS۳ یعنی ۲۰۰ تا ۵۰۰ متر است. از بین شرایط جوی مختلف، وضعیت بارانی و یخبندان بیشترین اثر را در ایجاد ترافیک دارد. نکته‌ی جالب توجه این است که احتمال کمی وجود دارد شرایط جوی نامساعد به ترافیک‌های با طول بیشتر از ۱۰۰۰ متر حتی در مناطق با حجم تردد زیاد یا ساختار شهری نامناسب منجر شود. در حالی که تصادفات می‌تواند چنین ترافیک‌های سنگینی را ایجاد کند.

۶. نتیجه‌گیری

امروزه با افزایش وسایل نقلیه، ترافیک به یکی از مهم‌ترین چالش‌های شهروندان و مسئولان در کلان‌شهرها تبدیل شده است. عوامل مختلفی در ایجاد ترافیک نقش دارند. در این پژوهش دو عامل تصادفات و شرایط جوی در نظر گرفته شده است. هدف این پژوهش بررسی تأثیر دو عامل «تصادفات» و «شرایط جوی» بر ترافیک شهر تهران با استفاده از روش‌های داده‌کاوی بوده است. داده‌های مورد بررسی از مرکز کنترل ترافیک تهران اخذ شد. پس از آماده‌سازی داده‌ها، تعداد ۵۲/۴۵۱ رکورد برای تصادف و تعداد ۱۱/۴۱۴ رکورد برای شرایط جوی به دست آمد.

نتایج این بررسی نشان داد، تصادفات به میزان ۴/۱۱٪ بر ترافیک تهران تأثیر دارد. همچنین شرایط جوی نیز به میزان ۰/۹۸٪ از سهم ترافیک را به خود اختصاص داده است. طبق تحلیل آماری رگرسیون خطی مشخص شد که طول ترافیک ناشی از تصادف تابع شدت تصادف، منطقه شهرداری محل تصادف و تعداد وسیله درگیر است و فرض تأثیرپذیری آن از روز عادی با تعطیلی مورد تأیید قرار نگرفته است. همچنین طول ترافیک ناشی از شرایط جوی تابع منطقه شهرداری، نوع شرایط جوی و روز عادی یا تعطیلی است. داده‌های مورد بررسی تأیید و با استفاده از الگوریتم درخت تصمیم C۵/۰ برای هرکدام مدلی مجزا جهت پیش‌بینی ترافیک ناشی از وقوع تصادف یا شرایط جوی ارائه شد. در نهایت قوانین هر مدل استخراج شد. نتایج نشان می‌دهد بیشترین طول ترافیک ایجاد شده در هر دو مدل ۲۰۰ تا ۵۰۰ متر است.

همانند بسیاری از پژوهش‌های علمی، پژوهش حاضر نیز با موانع و محدودیت‌هایی مواجه بوده است که در پژوهش‌های بعدی باید مورد توجه قرار گیرد. اطلاعات مورد استفاده در پژوهش، از پایگاه داده‌ی نرم‌افزار ثبت وقایع گرفته شده است، که بخش اعظم این اطلاعات، داده‌های ثبت شده توسط کارشناسان مراکز کنترل است. ثبت داده توسط افراد دارای خطای انسانی است، که این خطا در مرحله‌ی ثبت داده به اطلاعات این پژوهش نیز منتقل شده است. همچنین میزان ثبت داده در پایگاه داده و اولویت‌بندی کارشناسان در ثبت داده‌ها متأثر از سیاست‌های کلی مدیران مراکز و نیز نحوه‌ی توزیع جغرافیایی دوربین‌های نظارت تصویری است؛ که عدم توازن نصب دوربین‌ها و سایر حس‌گرها در سطح شهر تهران کاملاً مشهود است.

پیشنهاد می‌شود در تحقیقات آتی تصادفات رخ داده از طریق سیستم‌های مسیریاب به سایر رانندگان اطلاع داده شود. همچنین محل رخداد تصادف، معابر نزدیک تأثیرپذیر از تصادف و وضعیت ترافیک آنها، از طریق تابلوهای متغیر خبری به سایر رانندگان اطلاع‌رسانی شود.

پانوشت‌ها

1. geospatial information system
2. hot spots
3. statistical package for social science
4. RapidMiner
5. Kamran and Hass
6. global positioning system (GPS)
7. Siddiqui
8. Qiu and Nixon
9. Ibrahim and Hall
10. palm
11. classification and regression tree
12. support vector achin
13. linear regression
14. two-step clustering
15. likelihood measure
16. robust
17. pre-cluster
18. agglomerative hierarchical clustering

19. schwarz's bayesian criterion
20. akaike information criterion
21. decision trees
22. node splitting
23. central limit theorem

منابع (References)

1. Mahdavian Z, Niknafs A. "Forecasting traffic load using GPS data", *A Data Mining Approach. jgit.* **3**(2), pp. 43-59 (2015).
2. Tesema, T. B., Abraham, A. and Grosan, C. "Rule mining and classification of road traffic accidents using adaptive regression trees". *International Journal of Simulation*, **6**(10-11), pp. 80-94 (2005).
3. Tehran Comprehensive Transportation and Traffic Studies Company, "Tehran comprehensive transportation and traffic plan (general report)", *Tehran Comprehensive Transportation and Traffic Studies*, **930**, pp.800-850 (In Persian) (2007).
4. Chong, M., Abraham, A. and Paprzycki, M. "Traffic accident analysis using machine learning paradigms". *Informatica*, **29**(1), pp.89-98 (2005).
5. World Health Organization (WHO) (2009).
6. Effati, M., Rajabi, M.A. and Hakimpour, F. "Analysis of spatial factors contributing on concentration of highway corridors crashes using GIS and data mining". *JGST.* **4**(2), pp. 87-102 (2014).
7. Qiu, L., and Nixon, W. A. "Effects of adverse weather on traffic crashes: systematic review and meta-analysis". *Transportation Research Record*, **2055**(1), pp. 139-146 (2008).
8. Ibrahim, A. T., and Hall, F. L. "Effect of adverse weather conditions on speed-flow-occupancy relationships" (No. 1457) (1994).
9. Erdogan, S., Yilmaz, I., Baybura, T. and et al. "Geographical information systems aided traffic accident analysis system case study: city of afyonkarahisar". *Accident Analysis & Prevention*, **40**(1), pp. 174-181 (2008).
10. Ayati, E. "The cost of traffic accidents in Iran", *Ferdowsi University of Mashhad Press.* (In Persian) (2008).
11. Kashani, A. T., and Mohaymany, A. S. "Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models". *Safety Science*, **49**(10), pp. 1314-1320 (2011).
12. Kamran, S. and Haas, O. "A multilevel traffic incidents detection approach: identifying traffic patterns and vehicle behaviours using real-time gps data," *In Intelligent Vehicles Symposium*, pp. 912-917 (2007).
13. Evsukoff, A.G., and Ebecken, N.F. "Mining fuzzy rules for a traffic information system," *In Knowledge-Based Intelligent Information and Engineering Systems*, **2773**, pp. 237-243 (2003).
14. Abdel-Aty, M. A. and Pemmanaboina, R. "Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data". *IEEE Transactions on Intelligent Transportation Systems*, **7**(2), pp. 167-174 (2006).
15. Siddiqui, C., Abdel-Aty, M. and Huang, H. "Aggregate nonparametric safety analysis of traffic zones" *Accident Analysis and Prevention*, **45**, pp. 317-325 (2012).
16. Shafabakhsh, G. A., Famili, A. and Bahadori, M. S. "GIS-based spatial analysis of urban traffic accidents: Case study in mashhad, Iran". *Journal of traffic and transportation engineering (English edition)*, **4**(3), pp. 290-299 (2017).
17. Andrey J, Yagar S. "A temporal analysis of rain-related crash risk". *Accid Anal Prev. 1993 Aug*, **25**(4), pp. 465-72 (1993).
18. Foldvary, LA. Ashton, HT. "Road accidents and weather". *In: Proceedings of the Australian Road Research Board (ARRB) First Conference. Canberra.* **1**(1), pp. 529-83 (1962).
19. Smith K. "How seasonal and weather conditions influence road accidents in glasgow". *Scottish Geographical Magazine.* **98**, pp. 103-l14 (1982b)
20. Firouzi, M. A. and Moradimofrad, S. "Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models". *Safety Science*, **49**(10), pp. 1314-1320 (In persian) (2017).
21. Wen, H., Zhang, X., Zeng, Q. and Sze, N. N. "Bayesian spatial-temporal model for the main and interaction effects of roadway and weather characteristics on freeway crash incidence". *Accident Analysis & Prevention*, **132**, 105249 (2019).
22. Xing, F., Huang, H., Zhan, Z. and et al. "Hourly associations between weather factors and traffic crashes: non-linear and lag effects". *Analytic methods in accident research*, **24**, 100109 (2019).
23. Malin, F., Norros, I. and Innamaa, S. "Accident risk of road and weather conditions on different road types". *Accident Analysis & Prevention*, **122**, pp. 181-188 (2019).
24. Theofilatos, A. "Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials". *Journal of safety research*, **61**, pp. 9-21 (2017).
25. Abellán, J. López, G. and De OñA, J. "Analysis of traffic accident severity using decision rules via decision trees". *Expert Systems with Applications*, **40**(15), pp. 6047-6054 (2013).
26. Chang, L. Y. and Chen, W. C. "Data mining of tree-based models to analyze freeway accident frequency". *Journal of safety research*, **36**(4), pp. 365-375 (2005).
27. Safari, H. "Optimization of market basket analysis in fuzzy data mining using evolutionary algorithms". (In persian) (2012.)
28. Chen, M.-S., Han, J. and Yu, P.S. "Data mining: an overview from a database perspective", *IEEE Transactions on Knowledge and Data Engineering*, **8**, pp. 866-883 (1996).
29. Berry, M.J. and Linoff, G. "Data mining techniques: for marketing, sales, and customer support", *John Wiley & Sons* (1997).
30. Han, J. and Kamber, M. "Data mining: concepts and techniques. San Francisco", *CA: Morgan Kaufman* (2001).

31. Tan, P.N., Steinbach, M. and Kumar, V. "Introduction to Data Mining" Pearson Addison Wesley", *Boston* (2006).

32. Shih, M. Y., Jheng, J. W. and Lai, L. F. "A two-step method for clustering mixed categroical and numeric data". *Journal of Applied Science and Engineering*, **13**(1), pp. 11-19 (2010).

33. Zhang, T., Ramakrishnan, R. and Livny, M. "BIRCH: an efficient data clustering method for very large databases". *In ACM Sigmod Record* **25**(2), pp. 103-114 (June1996).

34. Chiu, T., Fang, D., Chen, J. and et al. A robust and scalable clustering algorithm for mixed type attributes in large database environment". *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 263 (2001).

35. Fraley, C. and Raftery, A.E. "How many clusters? Which clustering method? Answers via model-based cluster analysis". *Computer Journal*, **4**. p. 578-588 (1998).

36. Integral Solutions Limited., Clementine® 12.0 Algorithms Guide, (Chapter 10), **23**, pp.100-130 (2007).

37. Esmaeili, M. "Data mining: concepts and techniques". (In Persian)(2013).

38. Ghazanfari, M., Alizadeh, S. and Teymourpour, B. "Data mining and knowledge discovery". (In Persian) (2013).