

## مدل تصمیم‌یار پیشگیری از تصادفات رانندگی با بهره‌گیری از داده‌کاوی

### (مطالعه موردی در استان کرمان)

فاطمه سعادت‌جو<sup>۱</sup>، عرفان قندهاری<sup>۲</sup>، آزاده گروهی<sup>۳</sup>

از صفحه ۱۱۹ تا ۱۴۸

تاریخ دریافت: ۱۳۹۷/۱۲/۲۰ تاریخ پذیرش: ۱۳۹۸/۳/۲۵

#### چکیده

آمارها نشان می‌دهد که کشور ایران از لحاظ بروز تخلفات رانندگی و تصادفات نسبت به کشورهای جهان، سهم بیشتری را به خود اختصاص داده است. طبق آماری که توسط کارشناسان امر ترافیک به دست آمده است، سالانه حدود سه درصد از تولید ناخالص ملی کشورمان صرف آثار و تبعات ناشی از تخلفات ترافیکی می‌شود. با توجه به اهمیت بالای بحث ترافیک، در این پژوهش با در نظر گرفتن تعدادی از مؤلفه‌های مرتبط با تخلفات رانندگی که شامل اطلاعات فرد متخلف، اطلاعات خودرو و مشخصات خودرو می‌باشند و با مدل‌سازی از طریق ابزارهایی مثل درخت تصمیم و انتخاب ویژگی‌ها، به بررسی و شناخت رانندگان متخلف پرداخته شده است. داده‌های استفاده شده در این پژوهش در طول سه ماه در سال ۹۲ بر اساس تخلفات ترافیکی شهر کرمان جمع‌آوری شده است. نوآوری پژوهش بر اساس مطالعه شناسایی رفتار رانندگان حادثه‌آفرین و متدولوژی‌های ارائه شده در این زمینه شکل گرفته است که در ابتدا، ویژگی‌های مهم برای تقسیم‌بندی رانندگان دارای تخلف، بر اساس هدف، استخراج شده و سپس خوشه‌های رانندگان شکل خواهد گرفت؛ این روش روی داده‌های واقعی پیاده‌سازی شده است. نتایج پژوهش حاصل، امکان پیاده‌سازی به‌عنوان یک بسته نرم‌افزاری برای سیستم‌های ثبت و کنترل تخلفات ترافیکی را دارد.

**کلیدواژه‌ها:** مدل تصمیم‌یار، پیشگیری از تصادفات، استان کرمان، الگوریتم‌های داده‌کاوی، تخلفات رانندگی، رانندگان حادثه‌آفرین.

۱. استادیار، دانشکده مهندسی کامپیوتر، دانشگاه علم و هنر، (نویسنده مسئول)، ۰۹۱۳۳۵۳۶۴۷۱، Saadatjou@sau.ac.ir

۲. کارشناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه علم و هنر، Erfan.ghandehari@sau.ac.ir

۳. کارشناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه علوم و تحقیقات یزد، A.gorouhi@gmail.com

مقدمه

بیان مسئله

امروزه شهرها مهم‌ترین خاستگاه زیستی انسان‌ها بوده و با پدیده‌هایی مدرن نظیر آلودگی هوا، ترافیک، رانندگی پرخطر و تصادف مواجه هستند که ناشی از ورود به دنیای مدرن است (یاجان، ژاوو چنگ، سونگ بای، جی لی و ژی جان<sup>۱</sup>، ۲۰۰۹: ۲۶). رانندگی کردن و رفتارهایی که فرد در زمینه رانندگی نظیر سبقت غیرمجاز، عبور از چراغ قرمز، دورزدن در محل ممنوع، عبور وسایل نقلیه از پیاده‌رو، عبور از محل ممنوع، تجاوز از سرعت مجاز، عدم رعایت حق تقدم عبور، نبستن کمربند و موارد مشابه انجام می‌دهد، برخاسته از موقعیت‌های ساختاری و انتخاب‌های فردی می‌باشد (ولاهوجیانی، کارلافتیس و گولیاس<sup>۲</sup>، ۲۰۱۴: ۱۴). برخی، رانندگی را رفتاری می‌دانند که راننده به‌صورت الگویی برای رانندگی خود انتخاب می‌کند؛ مانند سرعت، میزان تمرکز و حفظ میزان فاصله استاندارد (اوزکان و لاجونن<sup>۳</sup>، ۲۰۰۵: ۳۶۱). این رفتارها شامل دو دسته رفتارهای مثبت و رفتارهای منفی می‌شوند. به مجموعه رفتارهای منفی در رانندگی، رانندگی تهاجمی یا رانندگی پرخطر گفته می‌شود؛ این رفتارها دو گروه اشتباهات و تخلفات را پوشش می‌دهند (آیوسو، گایلن و آلکانیز<sup>۴</sup>، ۲۰۱۰: ۷۱۳). رفتار رانندگی هر شخص به‌طورکلی درون مجموعه‌ای از رفتارها قرار می‌گیرد که بدان، شخصیت اطلاق می‌شود (سولومون، نگوین، لیبوویتز و آگرستی<sup>۵</sup>، ۲۰۰۶: ۶۲۹). در دنیای به‌شدت رقابتی امروز، تولید اطلاعات به‌عنوان یکی از فاکتورهای مهم شناخته شده است. در نتیجه، تلاش برای استخراج اطلاعات از داده‌ها، توجه بسیاری از افراد دخیل در صنعت اطلاعات و حوزه‌های وابسته را به خود جلب نموده است. حوادث رانندگی، یکی از بزرگ‌ترین مشکلات سلامت عمومی در جهان است؛ به‌طوری‌که

1. Ya-jun, Zhao-cheng, Song-bai, Jie-Li, Zhi-Jun

2. Vlahogianni, Karlaftis, Golias

3. Özkan, Lajunen

4. Ayuso, Guillén, Alcañiz

5. Solomon, Nguyen, Liebowitz, Agresti

سالانه جان ۱/۲ میلیون نفر را در جهان می‌گیرد و ۵۰ میلیون نفر نیز در جهان مصدوم می‌شوند. این آمار در کشور ایران در سال ۲۰۰۷ میلادی، ۲۷۵۶۷ نفر کشته و ۲۷۶۷۶۲ نفر مجروح بوده است

## اهمیت موضوع

یکی از حوزه‌هایی که در سال‌های اخیر به‌عنوان داده‌کاوی مورد توجه قرار گرفته، مسائل مربوط به پلیس راهنمایی و رانندگی است. شناسایی، پیش‌بینی و پیشگیری از جرایم و تخلفات ترافیکی، مسائلی هستند که با استفاده از تکنیک‌های داده‌کاوی مورد تحلیل قرار گرفته‌اند (تسنگ<sup>۱</sup>، ۲۰۱۳: ۲۳۸؛ یو، کانگ، پارک و بنگ<sup>۲</sup>، ۲۰۱۳: ۱۰۰۸؛ کاواساکی، هارا، میتانی و کواهارا<sup>۳</sup>، ۲۰۱۶: ۲۵۲؛ فوگو، گاریدو، مارتینز، کانو، کالافیت و منزونی<sup>۴</sup>، ۲۰۱۲: ۴۳). تمرکز این مقاله، پرداختن به مسائلی است که روزمره با آن در سطح شهر و جاده روبه‌رو هستیم. در این راستا، به مطالعه و بررسی بخشی از داده‌های تخلفات ترافیکی موجود از جمله شناخت نقاط قوت و ضعف داده‌های ذخیره‌شده و بهره‌برداری از آن در جهت شناخت رانندگان متخلف از نگاه علمی پرداخته می‌شود.

## اهداف

با توجه به حجم گسترده مطالب مرتبط با موضوع، به بحث در خصوص تخلفات ترافیکی بسنده شده است. نوآوری پژوهش بر اساس مطالعه شناسایی رفتار رانندگان حادثه‌آفرین و متدولوژی‌های ارائه‌شده در این زمینه شکل گرفته است که در ابتدا، ویژگی‌های مهم برای تقسیم‌بندی رانندگان دارای تخلف بر اساس هدف استخراج شده و سپس خوشه‌های رانندگان شکل خواهد گرفت؛ این روش روی داده‌های واقعی

1. Tseng

2. Yu, Kang, Park, Bang

3. Kawasaki, Hara, Mitani, Kuwahara

4. Fogue, Garrido, Martinez, Cano, Calafate, Manzoni

پیاده‌سازی شده است. به‌طور مشخص برای این پژوهش، دو هدف به‌عنوان اهداف اصلی مدنظر است که عبارتند از: الف) دسته‌بندی و خوشه‌بندی رانندگان حادثه‌آفرین و تشخیص ویژگی‌های آنان، ب) تعیین مدل انواع تخلفات بر اساس خوشه‌بندی و دسته‌بندی.

## سؤال اصلی

به کمک کدام‌یک از روش‌های دسته‌بندی می‌توان رانندگان متخلف را با دقت بیشتری پیش‌بینی نمود.

## پیشینه پژوهش

داده‌کاوی در مسائل مرتبط با پلیس می‌تواند به‌عنوان یکی از کاراترین ابزارها در زمینه تحلیل اطلاعات و داده‌های پلیسی موردتوجه قرار گیرد. مطالعات متعددی درمورد اهمیت داده‌کاوی در حوزه تخلفات ترافیکی انجام شده است که تمامی آن‌ها، تکنیک‌های خلاقانه‌ای را برای حل مشکلات با روش داده‌کاوی ارائه می‌دهند که در ادامه درباره آن‌ها بحث خواهیم نمود.

۱- در سال ۲۰۱۱ میلادی به‌منظور پرکردن این شکاف در ایران، تجزیه و تحلیلی روی حجم بالایی از تصادفات جاده‌ای صورت گرفت. در آن پژوهش، تصادفات ثبت‌شده در پایگاه داده پلیس ایران دارای بیش از ۳۰ ویژگی (همانند شرایط نور، جهت جاده، شرایط آب‌وهوایی و موارد مشابه) بود که از این بین فقط روی متغیرهای عامل انسانی متمرکز شدند و تنها از پنج ویژگی مؤثر که شامل سن، جنسیت، کمربند ایمنی، گواهینامه رانندگی و شدت حادثه بود، استفاده کردند. پژوهشگران با بهره‌گیری از تکنیک‌های داده‌کاوی از قبیل رگرسیون لجستیک چندگانه، طبقه‌بندی و رگرسیون درختی توانستند نقش عامل انسانی را در بروز و شدت تصادفات جاده‌ای پیدا کنند (پاک‌گوهر، سیگاری تبریزی، خلیلی و اسماعیلی<sup>۱</sup>، ۲۰۱۱: ۷۶۵).

۲- برای مدل‌سازی حجم ترافیک روزانه در سال ۲۰۱۱ میلادی، از داده‌های جمع‌آوری شده توسط حسگرهای ترافیکی به‌همراه سیاست‌های سیگنال پویا که به‌طور قابل‌توجهی باعث کاهش متوسط زمان انتظار رانندگان در تقاطع‌های چراغ‌دار در مقایسه با رویکرد برنامه‌ریزی استاتیک است، استفاده شد. با بهره‌گیری از الگوریتم خوشه‌بندی K-Means به شناسایی روزهای هفته که در ساعت مشخص، حجم ترافیکی مشابهی دارند، پرداختند. با اعمال خوشه‌بندی، خوشه‌ها به دو گروه یکشنبه تا پنج‌شنبه که مربوط به روزهای کاری بود و گروه جمعه و شنبه که مربوط به روزهای تعطیل بود، تقسیم‌بندی شدند. بیشترین تمرکز روی خوشه اول بود و از پنج ساعت از شب که عملاً عبور و مرور کاهش می‌یافت، چشم‌پوشی شد. نتایج نشان داد که شلوغ‌ترین ساعات در همه جهات بین ساعت ۸ تا ۹ صبح، ۱۳ تا ۱۴ و ۱۶ تا ۱۷ بعدازظهر بوده است (لاست، آوراومی و کاندل<sup>۱</sup>، ۲۰۱۱: ۶۰۵).

۳- در محاسبه تأثیر تخلفات ترافیکی بر احتمال وقوع یک تصادف شدید یا مرگ‌بار برای تصادفات اتفاق افتاده با توجه به ۳۰ نوع تخلف ترافیکی و انواع وسایل نقلیه (کامیون، ماشین و موارد مشابه) در بین سال‌های ۲۰۰۶ تا ۲۰۱۰ میلادی در چین، از مدل رگرسیون لجستیک چندگانه استفاده شده است. تخلفات ترافیکی در این پژوهش به سه گروه تقسیم‌بندی شده‌اند که شامل تخلفات مربوط به محدودیت سرعت، تخلفات اداری مثل نداشتن گواهینامه رانندگی و تخلفات مربوط به راننده می‌باشد (ژانگ، یاوو و چن<sup>۲</sup>، ۲۰۱۳: ۲۰).

۴- شیاوو، تاسی، هانگ و کاوو<sup>۳</sup> (۲۰۱۵: ۲) از تکنیک‌های داده‌کاوی برای ساخت یک مدل پیش‌بینی به‌منظور طبقه‌بندی حوادث ترافیکی در جاده استفاده نمودند. نتایج حاصل از داده‌کاوی در این مقاله نشان داد که روابط معناداری بین داده‌های مشخصات فردی رانندگان و تصادفات مرگ‌بار وجود دارد.

1.Last, Avrahami, Kandel  
2.Zhang, Yau, Chen  
3.Shiau, Tsai, Hung, Kuo

۵- سیستم کنترل ترافیک، یکی از سیستم‌هایی است که در آن، اطلاعات مهم (مانند تصادفات خودرو، حجم ترافیک و تراکم در سطوح مختلف) ثبت و نگهداری می‌شود. سال ۲۰۱۶ میلادی، از کاربرد ابزارهای داده‌کاوی در مدیریت ترافیک شهری مبتنی بر تجزیه و تحلیل خوشه‌ای سلسله‌مراتبی استفاده شد. این روش، توانایی طراحی یک سیستم کنترل سیگنال TOD (Time of Day) را به صورت خودکار فراهم کرد؛ چراکه می‌توانست داده‌های ترافیکی را در فواصل زمانی از روز شناسایی و جمع‌آوری کند (اسپالوویچ، آیلیک و آرسیچ<sup>۱</sup>، ۲۰۱۶: ۵۸).

۶- سیستم حمل و نقل هوشمند در حال تبدیل شدن به یکی از مهم‌ترین ابزارها برای حل مشکل ترافیک، خصوصاً مدیریت ترافیک است. از این رو گاوو<sup>۲</sup> (۲۰۱۶: ۱۰۴۳)، یک مدل سیستم تجزیه و تحلیل اطلاعات و پشتیبان تصمیم‌گیری مبتنی بر محاسبات شبکه‌ای و داده‌کاوی ارائه نمود.

۷- یکی از زمینه‌های مهم در مدیریت ترافیک، پیش‌بینی ترافیک است که ابزارهای داده‌کاوی به عنوان یک ابزار قابل اعتماد برای تولید الگوها در این زمینه انتخاب شده‌اند. بر این اساس در پژوهشی با به‌کارگیری روش‌های طبقه‌بندی به منظور یادگیری رفتار ترافیکی به پیش‌بینی حوادث آینده پرداخته شد. نتایج نشان داد که استفاده از روش دسته‌بندی در مقایسه با الگوریتم ژنتیک و شبکه‌های عصبی مصنوعی برای مدیریت ترافیک بسیار مؤثرتر خواهد بود و حتی می‌توان با ترکیب چند روش به پیش‌بینی‌های دقیق‌تری نیز دست یافت (نجداد، سیفی، احمدی و سیفی<sup>۳</sup>، ۲۰۰۹، ۶۷۶).

۸- همچنین در سال ۲۰۰۳ میلادی برای پیش‌بینی ترافیک پویا، از یک مدل تحلیلی به همراه درخت تصمیم، قوانین انجمنی و شبکه‌های عصبی استفاده شد (وو، ژاوو و ژانگ<sup>۴</sup>، ۲۰۰۳: ۳۹۷).

1. Spalević, Ilić, Arsić,

2. Guo

3. Nejad, Seifi, Ahmadi, Seifi

4. Wu, Zhou, Zhang

۹- مقدم و آیتی<sup>۱</sup> (۲۰۰۷: ۹۱) یک شاخص برآورد خطر برای رانندگان در ایران معرفی کردند. هدف از این پژوهش، بررسی روابط موجود بین ویژگی‌های رانندگان و رفتارهای نابجای رانندگی‌شان (خطاها، لغزش‌ها، تخلفات، تصادف و نرخ جریمه) بوده است و در این راستا برای رانندگان متخلف، شاخص خطر رانندگی (DRI)<sup>۲</sup> معرفی شده است. در این پژوهش برای جمع‌آوری داده‌ها و تجزیه و تحلیل آن‌ها، ۱۷۶۹ پرسشنامه از رانندگان ایرانی برای به‌دست‌آوردن اطلاعاتی درمورد شخصیت رانندگان، جنسیت، تحصیلات، تخلفات رانندگی، تصادف و مبلغ جریمه جمع‌آوری شده و برای رسیدن به این هدف، از خوشه‌بندی K-means و آزمون غیرپارامتریک استفاده شده است.

جدول ۱، چکیده‌ای از کارهای مرتبط را به‌طور خلاصه نشان می‌دهد.

جدول ۱. پیشینه‌ای از کارهای صورت‌گرفته

ردیف	نام نویسنده	عنوان	سال	نتایج
۱	پاک‌گوهر و همکاران	نقش عامل انسانی در تصادفات	۲۰۱۱	استخراج میزان تأثیر نقش عامل انسانی در بروز و شدت تصادفات جاده‌ای
۲	لاست و همکاران	مدل‌سازی داده‌های ترافیکی	۲۰۱۱	شناسایی روزهایی از هفته که در ساعات مشخص، حجم ترافیکی مشابه دارند و استخراج شلوغ‌ترین ساعات در همه جهات
۳	ژانگ و همکاران	عوامل مرتبط با نقض ترافیک و شدت تصادفات	۲۰۱۳	محاسبه تأثیر تخلفات ترافیکی احتمال وقوع یک تصادف شدید
۴	شیاوو و همکاران	استفاده از داده‌کاوی به‌منظور طبقه‌بندی حوادث ترافیکی جاده	۲۰۱۵	روابط معنادار بین داده‌های مشخصات رانندگان و تصادفات مرگبار
۵	اسپالویچ و همکاران	استفاده از داده‌کاوی برای پردازش جریان در نظارت ترافیک	۲۰۱۶	ارائه راهکار به‌منظور طراحی یک سیستم کنترل سیگنال
۶	گاوو	مدل پیش‌بینی جریان ترافیک بر اساس داده‌کاوی	۲۰۱۶	ارائه مدل سیستم تجزیه و تحلیل اطلاعات و پشتیبان تصمیم‌گیری مبتنی بر محاسبات شبکه‌ای
۷	نجداد و همکاران	استفاده از داده‌کاوی در پیش‌بینی و طبقه‌بندی ترافیک شهری	۲۰۰۹	کارایی بالای روش دسته‌بندی برای مدیریت ترافیک
۸	وو و همکاران	پیش‌بینی ترافیک پویا	۲۰۰۳	ارائه نتایج در قالب استفاده از درخت تصمیم، شبکه عصبی و قوانین انجمنی
۹	مقدم و همکاران	معرفی یک شاخص برآورد خطر به رانندگان	۲۰۰۷	معرفی شاخص خطر رانندگی

یکی از مهم‌ترین مسائلی که مورد توجه پژوهشگران در ادبیات موضوع قرار گرفته، شناخت متغیرهای تأثیرگذار بر رانندگان و رفتار ترافیکی آنان است. پژوهشگران معتقد بودند که با گروه‌بندی رانندگان به وسیله مشخصات آنان، می‌توان بهترین شناسایی را در مورد رانندگی آن گروه انجام داد. ضمن اینکه یکی از شکاف‌های پژوهش‌هایی که با مطالعه ادبیات موضوع دیده می‌شود، عدم وجود یک متدولوژی جامع می‌باشد؛ به صورتی که هر کدام از پژوهشگران، بخشی از موضوع را مورد توجه قرار داده‌اند؛ به طور مثال، گروهی به بررسی تکنیک‌ها و مقایسه بین آن‌ها، و گروهی هم به بررسی متغیرهای تأثیرگذار و انتخاب ویژگی برای رانندگان پرداخته‌اند.

### مبانی نظری پژوهش

با بررسی ریشه‌های تخلفات ترافیکی در ایران به استناد آمارهای ارائه شده از جانب سازمان‌های ذی‌ربط می‌توان دید که بیش از ۷۰ درصد تخلفات ترافیکی در ایران، با علل و ریشه‌های انسانی بوده که می‌توان با برنامه‌ریزی و تدوین راهکارهایی از جمله داده‌کاوی و با دانستن سوابق تخلفاتی و مشخصه‌های فردی و محیطی، رانندگان متخلف را شناسایی و قوانین لازم را برای کاهش تخلفات ترافیکی تدوین و اجرا کرد. از این رو طیف وسیعی از تخلفات ترافیکی منجر به حوادث در ایران کاهش و یا تعدیل می‌شود و از رشد روزافزون آن که موجب بروز آسیب‌های جبران‌ناپذیر اجتماعی و زیان‌های هنگفت اقتصادی می‌باشد، جلوگیری می‌شود. از طرفی، با توجه به اینکه مدل رفتار رانندگان تحت تأثیر عوامل دیگری مانند ویژگی‌های فرهنگی هر شهر، بافت شهری از منظر شهرسازی، گروه سنی رانندگان، عوامل جغرافیایی، اقتصادی و محیطی می‌باشد، میزان و نوع تخلفات متفاوت خواهد بود؛ بنابراین نتایج پژوهش‌های هر منطقه، شهر یا کشور نمی‌تواند به طور کامل در شهرهای دیگر نیز صادق باشد. پایه اصلی پژوهش حاضر، بر اساس کشف دانش از پایگاه داده تخلفات ترافیکی

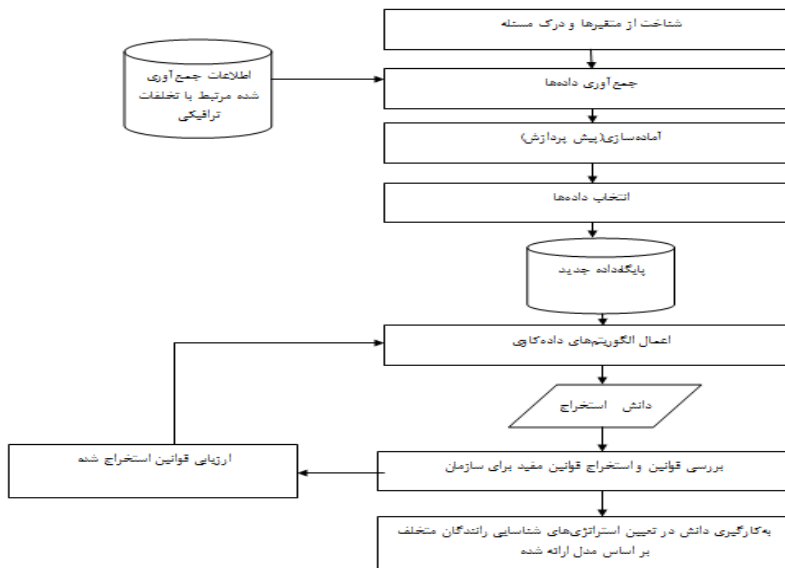


مورد مطالعه قرار گرفته است؛ از این رو استاندارد جهانی CRISP\_DM<sup>۱</sup> (ونگ و همکاران، ۲۰۱۸: ۴) به منظور انجام فرایند پژوهش مورد استفاده قرار گرفته است.

### چهارچوب نظری و فرضیه‌ها

شناسایی رفتار رانندگان بر اساس خطرآفرینی آن‌ها انجام می‌شود. همانند سایر کاربردهای داده‌کاوی، فرایند شناسایی رفتار رانندگان خطرآفرین، فرایندی تکراری و همراه با محاسبات زیاد است که قادر به ارائه راه‌حل بهینه نیست. شناسایی رانندگان خطرآفرین همانند سایر مسائل مرتبط می‌تواند به‌طور عملی به کار گرفته شود؛ از طرفی، بهترین و استانداردترین روش که در هر شرایطی برای چنین مسئله‌ای جوابگو باشد، وجود ندارد. تکنیک‌های داده‌کاوی زیادی به‌طور مداوم به وسیله متخصصین آمار و علوم رایانه ارائه و بهبود داده می‌شوند که استفاده‌های عملی از آن‌ها باید اعتبارسنجی شود. چهارچوب شناسایی رانندگان خطرآفرین مانند ابزاری برای اعتبارسنجی تأثیر تکنیک‌های جدید داده‌کاوی در کاربردهای عملی کار می‌کند. پژوهشگران می‌توانند الگوریتم‌ها و مدل‌های تازه توسعه‌یافته را به‌طوری سیستماتیک با اختیار کردن فرایند ساخت مدل استاندارد شناسایی رانندگان خطرآفرین اعتبارسنجی کنند. چهارچوب کلی به‌عنوان ابزاری مرجع برای ساخت مدل شناسایی رانندگان خطرآفرین به‌شمار می‌رود. از این چهارچوب، نه‌تنها تکنیک‌های زیادی می‌تواند برای ساخت مدل به کار رود، بلکه فرایند و استراتژی کنترل فرایند در چهارچوب کلی می‌تواند به‌طرزی سازگار و سیستماتیک از ساخت مدل پشتیبانی کند. اگرچه سعی شده است که چهارچوب ارائه‌شده، قالب عمومی داشته باشد تا بتواند بیشترین شرایط حاکم بر مسئله شناسایی رانندگان متخلف را در نظر بگیرد؛ با این حال، راهکار صددرصد کامل و قطعی برای این مسئله وجود ندارد. نمی‌توان انتظار داشت که مسائل واقعی رانندگان متخلف به‌طور کامل همانند مسائل آزمایشی تحت‌کنترل باشد.

مدل اجرایی پیشنهادی به صورت شکل ۱ می‌باشد. همان‌طور که در این شکل مشخص است، بعد از شناخت متغیرها و درک مسئله، به جمع‌آوری داده‌ها بر اساس اطلاعات جمع‌آوری شده مرتبط با تخلفات ترافیکی در مرحله بعد، به یکپارچه‌سازی و آماده‌سازی داده‌ها که به پیش‌پردازش معروف است، پرداخته خواهد شد و در نتیجه، پایگاه داده اصلی برای اعمال الگوریتم‌های داده‌کاوی تعیین می‌گردد. بعد از اعمال الگوریتم‌های داده‌کاوی، دانش ارائه‌شده استخراج می‌شود؛ سپس اعتبارسنجی مدل‌های ایجادشده انجام می‌گیرد و در انتها، مدل نهایی ارائه می‌گردد.



شکل ۱. مدل اجرایی پژوهش

## روش پژوهش

در هر مسئله داده‌کاوی، مرحله پیش‌پردازش داده‌ها از اهمیت ویژه‌ای برخوردار است که در ادامه به آن پرداخته می‌شود. یکی از مهم‌ترین بخش‌های هر پژوهش، تجزیه و تحلیل داده و دستیابی به نتایج است. در تجزیه و تحلیل داده‌ها، محاسبات آماری، بعد کمی پژوهش و تحلیل استدلال و استنتاج‌ها، بعد کیفی محاسبات است که بر اساس آن، یافته‌های پژوهش به جامعه آماری تعمیم داده می‌شود. یافته‌های پژوهش،



مدل تصمیم‌یار پیشگیری از تصادفات رانندگی با بهره‌گیری از داده‌کاوی

یافته و میزان صحت عملکرد مدل افزایش یابد.

برای رانندگان متخلف در پایگاه داده، سه دسته اطلاعات نگهداری می‌شود. دسته نخست شامل مشخصات تخلف، دسته دوم شامل مشخصات خودرو متخلف و دسته سوم شامل مشخصات راننده متخلف است. در جداول ۲، ۳ و ۴، فهرست ویژگی‌های توصیف‌کننده یک راننده آمده است.

جدول ۲. مشخصات تخلف

ردیف	فیلد	توضیحات
۱	نام تخلف	تخلف خودرو را نشان می‌دهد.
۲	نوع تخلف	حادثه‌سازبودن تخلف خودرو را نشان می‌دهد.
۳	مبلغ تخلف	میزان مبلغ تخلف را نشان می‌دهد.
۴	مکان تخلف	آدرس تخلف را نشان می‌دهد.
۵	تاریخ تخلف	تاریخ ارتکاب تخلف را به‌صورت روز، ماه و سال را نشان می‌دهد.
۶	زمان تخلف	زمان تخلف را به‌صورت دقیقه و ساعت نشان می‌دهد.

همان‌طور که در جدول ۲ نشان داده شده، مشخصات تخلف راننده شامل ۶ فیلد نام تخلف، نوع تخلف، مبلغ تخلف، مکان تخلف، تاریخ تخلف و زمان تخلف می‌باشد که تعریف هر فیلد در بخش توضیحات مشخص گردیده است. جدول ۳، مشخصات خودرو متخلف را نشان می‌دهد.

جدول ۳. مشخصات خودرو متخلف

ردیف	فیلد	توضیحات
۱	پلاک	پلاک خودرو را نشان می‌دهد.
۲	نام	نام خودرو مانند پراید، پژو، سمند و موارد مشابه را نشان می‌دهد.
۳	رنگ	رنگ خودرو را نشان می‌دهد.
۴	کاربری خودرو	نوع خودرو مانند سواری، وانت، کامیون و موارد مشابه را نشان می‌دهد.

همان‌طور که در جدول ۳ نشان داده شده است، مشخصات خودرو متخلف شامل ۴ فیلد پلاک، نام، رنگ و کاربری خودرو است که تعریف هر فیلد در بخش توضیحات مشخص گردیده است.

جدول ۴، مشخصات راننده متخلف را نشان می‌دهد.

جدول ۴. مشخصات راننده متخلف

ردیف	فیلد	توضیحات
۱	نام و نام خانوادگی	نام و نام خانوادگی فرد متخلف را نشان می‌دهد.
۲	کد ملی	کد ملی فرد متخلف را نشان می‌دهد.
۳	نام پدر	نام پدر فرد متخلف را نشان می‌دهد.
۴	جنسیت	جنسیت فرد متخلف را نشان می‌دهد.
۵	تاریخ تولد	تاریخ تولد فرد متخلف را به صورت روز، ماه و سال نشان می‌دهد.
۶	مدرک تحصیلی	نوع مدرک تحصیلی فرد متخلف را به صورت ابتدایی، دیپلم، لیسانس و موارد مشابه نشان می‌دهد.
۷	شماره گواهینامه	شماره گواهینامه اخذ شده توسط فرد متخلف را نشان می‌دهد
۸	نوع گواهینامه	نوع گواهینامه فرد متخلف (پایه ۲ یا پایه ۳) را نشان می‌دهد.
۹	گواهینامه دارای عینک	صاحب گواهینامه باید با عینک رانندگی نماید.
۱۰	محدودیت پزشکی	آیا صاحب گواهینامه دارای محدودیت پزشکی است یا نه؟
۱۱	مالکیت خودرو	آیا فرد متخلف خود، مالک خودرو است یا نه؟
۱۲	سال اخذ گواهینامه	تاریخ صدور اولین گواهینامه فرد متخلف را به صورت روز، ماه و سال نشان می‌دهد.
۱۳	آدرس فرد متخلف	آدرس محل سکونت فرد متخلف را نشان می‌دهد
۱۴	شغل	شغل فرد متخلف را نشان می‌دهد.
۱۵	وضعیت تأهل	وضعیت تأهل (مجرد، متأهل) فرد متخلف را نشان می‌دهد.
۱۶	درآمد	درآمد فرد متخلف (ضعیف - متوسط و خوب) را نشان می‌دهد.

همان‌طور که در جدول ۴ نشان داده شده، مشخصات فرد متخلف شامل ۱۶ فیلد می‌باشد که تعریف هر فیلد در بخش توضیحات مشخص گردیده است. عناوین تخلفات حادثه‌ساز بر اساس پایگاه اطلاع‌رسانی دولت که بنا به پیشنهاد وزارت کشور و به استناد تبصره ماده ۵ قانون اصلاح بیمه اجباری مسئولیت مدنی دارندگان وسایل نقلیه موتوری زمینی در مقابل شخص ثالث (مصوب ۱۳۸۷) می‌باشد، تعیین شده است؛ جدول ۵، این عناوین را نشان می‌دهد.

جدول ۵. عناوین تخلفات رانندگی حادثه‌ساز

ردیف	عنوان تخلف
۱	هرگونه حرکات نمایشی مانند دورزدن درجا، حرکت روی یک چرخ و یا هرگونه حرکات آکروباتی در سطح راهها و خیابانها
۲	دورزدن در محل ممنوع
۳	عبور وسایل نقلیه از پیاده‌رو
۴	نقص سیستم روشنایی وسایل نقلیه به هنگام شب
۵	عبور از محل ممنوع (ورود ممنوع)
۶	سبقت غیرمجاز
۷	تجاوز از سرعت مجاز
۸	عبور از چراغ قرمز راهنمایی
۹	عدم رعایت حق تقدم عبور
۱۰	حرکت به‌طور ماریج در راهها
۱۱	تجاوز یا انحراف به‌سمت چپ معابر
۱۲	گردش به چپ یا به راست در محل ممنوع
۱۳	روشن‌نکردن چراغ هنگام شب و در موقع لزوم
۱۴	عدم رعایت فاصله مناسب با وسیله نقلیه و یا عدم توجه به جلو
۱۵	رانندگی با وسیله نقلیه دارای عیب و نقص فنی مؤثر (چراغ، لاستیک، فرمان، ترمز، برف‌پاک‌کن و زنجیر چرخ)
۱۶	مصرف مواد روان‌گردان و مشروبات الکلی
۱۷	عدم رعایت شرایط مندرج در گواهینامه از قبیل استفاده از سمک، عینک یا تجهیزات خاص

پیش‌پردازش داده‌ها

پس از واردشدن داده‌ها در رایانه، باید کیفیت آن‌ها بررسی شود و به‌وسیله پیش-پردازش، داده‌های بی‌کیفیت تغییر یابند یا حذف گردند. با توجه به آنکه تمام عملیات داده‌کاوی بر رویدادها اعمال می‌شود، انتخاب نادرست آن‌ها و نیز کیفیت پایین داده‌ها، کلیه نتایج کار را به مخاطره می‌اندازد. لذا عملیات پیش‌پردازش در قالب مراحل زیر از این امر جلوگیری می‌کند:

- پاک‌سازی داده‌ها<sup>۱</sup>؛
- تبدیل و یکپارچه‌سازی<sup>۲</sup>؛
- کاهش داده‌ها<sup>۳</sup>؛
- انتخاب ویژگی<sup>۴</sup>.

- 
- 1.Data Cleaning
  - 2.Transformation and Integeration
  - 3.Data Reduction
  - 4.Feature Selection

## پاک‌سازی داده‌ها

داده‌ها در پایگاه داده، هیچ‌گاه در حالت عادی برای داده‌کاوی آماده نیستند. در این مرحله بایستی با انجام اقداماتی، داده‌ها را به‌صورتی درآورد که بتوان در نرم‌افزارهای موردنظر از آن‌ها استفاده نمود. بعد از جمع‌آوری داده‌ها، برای اینکه الگوهای موردنظر توسط ابزارهای مختلف داده‌کاوی استخراج شود، باید داده‌ها پیش‌پردازش شده و به شکل‌های مناسب برای استفاده در نرم‌افزارهای داده‌کاوی درآیند. همچنین خطاها و داده‌های گمشده باید به‌نحو مناسب پاک‌سازی شوند. برای شناسایی مقادیر پرت در این مقاله سعی شد تا با ایجاد رابطه‌های منطقی بین ویژگی‌ها، نمونه‌هایی که پاسخ آن‌ها غیرقابل اعتماد شناخته شدند، از بانک اطلاعاتی خارج شوند؛ برای مثال می‌توان به موارد زیر اشاره کرد:

- حذف رکوردهای اطلاعاتی افرادی که باوجود ثبت مشخصات فرد در برگ تخلف، سایر مشخصات فردی او از قبیل مدرک تحصیلی، مالکیت خودرو و موارد مشابه در جداول موردنظر تخلف ثبت نشده بود و باعث عدم شناسایی متخلف می‌شد.
  - حذف نمونه‌هایی که باوجود ثبت شماره گواهینامه در برگ تخلف، سایر مشخصات گواهینامه فرد متخلف مانند سال اخذ، سابقه ممنوعیت، سابقه محدودیت پزشکی و موارد مشابه در جداول موردنظر ثبت نشده بود و باعث عدم شناسایی متخلف می‌شد.
- در این مقاله، پاک‌سازی داده‌ها در دو بخش اصلی انجام شد که در ادامه، هریک را موردبررسی قرار می‌دهیم:

### ▪ یک‌شکل نمودن داده‌ها

برای یکپارچه‌کردن داده‌های موجود برخی فیلدها با یکدیگر و همچنین قابلیت پردازش دادن به بعضی داده‌ها در مدل، لازم بود تا داده‌ها یک‌شکل شوند. ازجمله فیلدهایی که در این مرحله، تغییراتی روی آن‌ها صورت گرفته است، در جدول ۶ قابل مشاهده می‌باشند.

**جدول ۶. یک‌شکل‌نمودن داده‌ها**

فیلد	شکل قدیم	شکل جدید
زمان تخلف	زمان تخلف به صورت دقیقه: ساعت	روز یا شب
سن فرد متخلف	سال تولد فرد متخلف	سن فرد متخلف
گروه سنی فرد متخلف	سن فرد متخلف	گروه سنی فرد متخلف
سال اخذ گواهینامه	سال اخذ گواهینامه	مدت دارای گواهینامه
سطح تحصیلات	۴ گروه تحصیلی (بی‌سواد، ابتدایی تا دیپلم، فوق‌دیپلم تا لیسانس، فوق‌لیسانس و بالاتر)	درجه‌های مختلف تحصیلی (بی‌سواد، ابتدایی، سیکل، دیپلم، فوق‌دیپلم، لیسانس، فوق‌لیسانس)

▪ ساخت فیلد جدید در داده‌ها از روی فیلدهای قبلی

در این مرحله، ساخت یک فیلد جدید به نام نوع رانندگی، از روی فیلد مدت دارای گواهینامه صورت گرفت؛ به این صورت که پس از تعیین مدت‌زمان داشتن گواهینامه از روی سال اخذ گواهینامه، طبقه‌بندی نوع رانندگی به دو گروه مبتدی و حرفه‌ای انجام شد؛ به عبارتی، ایجاد فیلد جدید از روی سال اخذ گواهینامه به صورت زیر انجام گرفته است.

**جدول ۷. ساخت فیلد نوع رانندگی از روی فیلد مدت داشتن گواهینامه**

نوع رانندگی	مدت دارای گواهینامه
مبتدی	کمتر از سه سال
حرفه‌ای	بیشتر از سه سال

**تبدیل و یکپارچه‌سازی داده‌ها**

فرایند داده‌کاوی اغلب نیازمند یکپارچه‌کردن داده‌های خام موجود در منابع مختلف می‌باشد. یکی از کارهای رایج در اجرای فرایند داده‌کاوی، تبدیل داده‌ها به شکلی مناسب است. در یکپارچه‌سازی، داده‌های به‌دست‌آمده از منابع مختلف با یکدیگر ترکیب می‌شوند. تطابق موضوعات، حذف زواید و تشخیص مقادیر متضاد در داده‌ها، از جمله مواردی هستند که در هنگام یکپارچه‌سازی داده‌ها باید به آن‌ها توجه شود. از آنجاکه داده‌ها خیلی مناسب نبودند، بدین منظور سعی شد که مشخصات موردنیاز به‌گونه‌ای تغییر کنند که برای انجام عملیات داده‌کاوی مناسب باشند؛ از جمله این



یکپارچه‌سازی و تبدیل‌ها می‌توان به موارد زیر اشاره کرد:

- یکپارچه‌کردن اطلاعات تخلف با اطلاعات فردی که مرتکب تخلف شده است.
- یکپارچه‌کردن اطلاعات تخلف با اطلاعات و مشخصات خودرویی که فرد با آن مرتکب تخلف شده است.
- تبدیل میزان سن رانندگان به ۵ رده سنی شامل سن ۱۸ الی ۲۴ سال، ۲۵ الی ۳۵ سال، ۳۵ الی ۴۵ سال، ۴۵ الی ۵۵ سال و بیش از ۵۵ سال.
- تبدیل زمان تخلف به مقیاس‌های روز یا شب.
- تبدیل نوع خودرو به تقسیم‌بندی خودروهای سبک و سنگین.
- تبدیل فیلد سابقه رانندگی افراد به دو گروه مبتدی و حرفه‌ای بر اساس سال اخذ گواهینامه آن‌ها.
- تبدیل فیلد مدرک تحصیلی به ۴ دسته شامل بی‌سواد، ابتدایی تا دیپلم، فوق‌دیپلم تا لیسانس، فوق‌لیسانس و بالاتر.
- تبدیل فیلد شغل به دسته‌بندی شامل کارمند، راننده، شغل دولتی، شغل آزاد، خانه‌دار، دانشجو، بازنشسته، بیکار.

### کاهش داده‌ها

تکنیک‌های کاهش داده، برای دستیابی به یک کاهش در داده‌ها به کار می‌روند؛ به‌صورتی که داده‌های جدید علاوه بر کم‌حجم‌تر بودن نسبت به داده‌های اولیه، به همان میزان داده‌های اولیه، یکپارچه باشند. داده‌کاوی روی داده‌های کاهش‌یافته دارای کارایی بالاتری است. از جمله اقدامات انجام‌شده در این مقاله برای کاهش بُعد داده‌ها می‌توان به این موارد اشاره کرد:

#### ▪ ساخت ویژگی‌های ترکیبی

اطلاعات موجود در جداول، پردازش و متغیر جدیدی به نام مالکیت خودرو ایجاد گردید.

اطلاعات مربوط به زمان تخلفات، موردپردازش قرار گرفته و ویژگی‌های جدیدی مانند روز و شب ایجاد گردید.

### انتخاب ویژگی

مانند هر مدل‌سازی دیگری، انتخاب ویژگی‌هایی که بتوانند خوشه‌های قابل فهم‌تر و ملموس‌تری را ارائه داده و ارتباط معناداری بین ویژگی‌های هر خوشه کشف کنند، از اهمیت ویژه‌ای برخوردار است. اطلاعات موردنیاز در مورد رانندگان برای تشخیص متخلفین را می‌توان از سابقه گذشته آن‌ها استخراج کرد. این اطلاعات به شکل داده‌های خام در پایگاه داده برای کاوش و تحلیل ذخیره شده است؛ بنابراین داده‌کاوی، بخش مهمی از فرایند بخش‌بندی محسوب می‌شود. بر اساس مطالعات و مصاحبه‌های انجام‌شده با متخصصان و رئیس پلیس راهور استان کرمان، تعاریف مختلفی از پرخطر ارائه شده است که خصوصیات فرد متخلف برای متغیر هدف (خطر آفرین بودن) به شرح زیر تعریف می‌شود:

- رانندگان حرفه‌ای دارای دو تخلف حادثه‌ساز؛
  - رانندگان حرفه‌ای دارای ده تخلف غیر حادثه‌ساز؛
  - رانندگان مبتدی دارای دو تخلف غیر حادثه‌ساز یا یک تخلف حادثه‌ساز؛
  - رانندگان مبتدی یا حرفه‌ای دارای ممنوعیت رانندگی؛
  - رانندگان دارای محدودیت پزشکی و یک تخلف حادثه‌ساز.
- اطلاعات جمع‌آوری شده در این پژوهش با کمک پرسشنامه‌های تهیه‌شده توسط پژوهشگران و چند همکار به دست آمده و سپس به صورت دستی در پایگاه داده ذخیره شده است. در کل، از دو خصوصیت فرد متخلف برای متغیر هدف (خطر آفرین بودن) استفاده شده است که به شرح زیر می‌باشند:
- رانندگان مبتدی دارای دو تخلف غیر حادثه‌ساز یا یک تخلف حادثه‌ساز؛



۲۰۱۵) که محل اقامت، تأثیر قابل توجهی بر رفتار رانندگان متخلف حادثه‌آفرین و بااحتیاط خواهد داشت. تغییرات محیط زندگی می‌تواند بر تقسیم‌بندی رانندگان تأثیرگذار باشد؛ زیرا معمولاً رانندگان در شرایط مختلف محیطی، رفتارهای متفاوتی از خود بروز می‌دهند. شهر محل زندگی راننده و محل ارتکاب تخلف و موارد مشابه، جزو این گروه از ویژگی‌ها هستند.

### یافته‌ها

برای انجام فرایند داده‌کاوی و اعمال الگوریتم‌ها، از نرم‌افزار SPSS CLEMENTINE نسخه ۱۲ استفاده شد. این نرم‌افزار، یکی از پرکاربردترین نرم‌افزارهای آماری است که مورد استفاده جامعه‌شناسان و کاربران آماری قرار می‌گیرد و از بیشترین تعداد الگوریتم‌های داده‌کاوی حمایت می‌کند.

با توجه به اهمیت ایجاد داده‌های یادگیری و آزمون، داده‌های مورد بررسی به دو دسته داده‌های آموزشی و داده‌های تست تقسیم‌بندی شده‌اند. در مرحله اول، مدل با مجموعه داده‌های آموزشی، آموزش داده و با آن پیاده‌سازی گردید و در ادامه، با مجموعه داده‌های تست، میزان صحت عملکرد مدل مورد بررسی قرار گرفت. در تقسیم‌بندی داده‌ها، ۶۵ درصد از داده‌ها و رکوردهای موجود در بانک اطلاعاتی به‌عنوان داده‌های یادگیری و ۳۵ درصد باقی‌مانده داده‌ها به‌عنوان داده‌های آزمون انتخاب شده‌اند.

مجموعه داده‌های اصلی شامل رکوردهای مربوط به تخلفات رانندگان و خودروهایی آنان می‌باشد. در پایگاه داده ایجاد شده، با تعریف ویژگی‌های جدید برای هر راننده، هر کدام از آن‌ها به‌وسیله تعدادی از متغیرها توصیف می‌شوند. برای شناسایی انواع رانندگان از نظر تخلف، از درخت تصمیم استفاده شده است. به کمک درخت تصمیم مشخص شد که در پایگاه داده مورد استفاده، ۳۷ درصد از رانندگان خطرآفرین و ۶۳ درصد از آن‌ها بااحتیاط هستند. رانندگان بااحتیاط به‌طور متوسط بین ۳ تا ۴ تخلف در طول دوره هدف دارند؛ اما رانندگان خطرآفرین بین ۱ تا ۱۰ تخلف را در طول دوره در

پروفایل خود دارا می‌باشند. با توجه به اهمیت رانندگان خطرآفرین در حوادث رانندگی، در ادامه به شناسایی آن‌ها توسط الگوریتم‌های دسته‌بندی، خوشه‌بندی و ترکیب آن‌ها پرداخته می‌شود.

به‌منظور دسته‌بندی، از الگوریتم‌های شبکه عصبی<sup>۱</sup>، CHAID، C&R Tree، QUEST، Bayesian Network، C5.0 و SVM استفاده شده است. داده‌های ورودی شامل حادثه‌سازبودن، مالکیت خودرو، محدودیت پزشکی، جنسیت، مدرک تحصیلی، وضعیت تأهل، شغل، درآمد، نوع رانندگی، و داده خروجی شامل نام تخلف می‌باشند. نتایج به‌دست‌آمده نهایی، با استفاده از اجرای الگوریتم‌های دسته‌بندی مختلف در جدول ۸ آمده است. مقایسه نتایج نشان داد که الگوریتم شبکه عصبی برای دسته‌بندی رانندگان خطرآفرین، دقت بالاتری نسبت به سایر الگوریتم‌های دسته‌بندی دارد. این الگوریتم، ۴۷ درصد از کل رانندگان به‌کاررفته برای آموزش را به‌درستی دسته‌بندی می‌کند؛ همچنین ۴۲/۱۱ درصد از رانندگان به‌کاررفته برای تست الگوریتم را به‌درستی در دسته‌ها قرار می‌دهد.

جدول ۸. مقایسه دقت نتایج الگوریتم‌های دسته‌بندی

نام الگوریتم	پیش‌بینی درست	دقت کلی
الگوریتم Neural Net	47%	42/11%
الگوریتم CHAID	39%	28/7%
الگوریتم C&R Tree	59%	40/35%
الگوریتم QUEST	40%	35/9%
الگوریتم C5.0	63%	40/35%
الگوریتم Bayes Net	55%	35/9%
الگوریتم SVM	7%	7/2%

برای افزایش میزان دقت پیش‌بینی، از ترکیب الگوریتم‌های دسته‌بندی و خوشه‌بندی استفاده شده است. برای نیل به این هدف، خروجی الگوریتم خوشه‌بندی به‌عنوان داده‌های ورودی الگوریتم‌های دسته‌بندی در نظر گرفته می‌شود. با استفاده از الگوریتم

1. Neural Net

مدل تصمیم‌یار پیشگیری از تصادفات رانندگی با بهره‌گیری از داده‌کاوی

خوشه‌بندی K-means، داده‌ها در خوشه‌های اولیه قرار می‌گیرند و نمونه‌های غیرهمگون در مجموعه داده‌ها شناخته شده و داده‌های پراکنده حذف می‌شود. نتایج به‌دست‌آمده از ده بار خوشه‌بندی نشان داد که بهترین نتایج برای تعداد خوشه، k برابر با ۵ می‌باشد.

نتایج به‌دست‌آمده از اجرای الگوریتم ترکیبی در جدول ۹ آمده است. نتایج نشان می‌دهد که الگوریتم ترکیبی خوشه‌بندی K-Means و دسته‌بندی C5.0 دارای دقت بیشتری نسبت به سایر اجرای الگوریتم‌های ترکیبی دیگر هستند.

جدول ۹. نتایج حاصل از اجرای الگوریتم ترکیبی

دقت نتایج ترکیب خوشه‌بندی و الگوریتم‌های دسته‌بندی (درصد)	دقت الگوریتم دسته‌بندی (درصد)	نام الگوریتم
۳۳/۴۲	۱۱/۴۲	الگوریتم Neural Net
۱/۴۶	۷/۲۸	الگوریتم CHAID
۳۳/۴۲	۱۱/۴۲	الگوریتم C&R Tree
۱۷/۴۴	۹/۳۵	الگوریتم QUEST
۲۶/۶۶	۳۵/۴۰	الگوریتم C5.0
۴۷/۴۸	۹/۳۵	الگوریتم Bayes Net
۲۷/۱۲	۲/۷	الگوریتم SVM

از آنجاکه نتایج به‌دست‌آمده نهایی با استفاده از اجرای الگوریتم ترکیبی نشان داد که الگوریتم ترکیبی خوشه‌بندی K-Means و دسته‌بندی C5.0 دارای دقت بیشتری نسبت به اجرای سایر الگوریتم‌های ترکیبی دیگر می‌باشند، مجموعه قواعد مهم به‌دست‌آمده از این ترکیب در جدول ۱۰ آورده می‌شود.

**جدول ۱۰. مجموعه قواعد حاصل از اجرای الگوریتم ترکیبی K-Means و C5.0**

ردیف	نوع تخلف حادثه‌ساز	«ولفه‌های مؤثر در تخلف متبج از قواعد
۱	عبور از چراغ قرمز	- درآمد خوب، مدرک دیپلم، رانندگی مبتدی و مالکیت خودرو - مدرک لیسانس، رانندگی مبتدی، سن کمتر از ۲۳ سال و عدم مالکیت خودرو - مدرک دیپلم، سن بیشتر از ۲۹ سال و عدم مالکیت خودرو - درآمد ضعیف و سن بیشتر از ۲۹ سال - مدرک فوق‌دیپلم و سن بیشتر از ۲۹ سال
۲	عبور از محل ممنوع	- مدرک لیسانس، مالکیت خودرو و رانندگی مبتدی - مدرک سیکل یا لیسانس و عدم مالکیت خودرو - مدرک دیپلم، شغل آزاد و عدم محدودیت پزشکی
۳	عدم رعایت شرایط مندرج در گواهینامه	- محدودیت پزشکی و شغل آزاد
۴	عبور وسایل نقلیه از پیاده‌رو	- مدرک دیپلم و سن کمتر از ۴۱ سال - مدرک فوق‌دیپلم، سن کمتر از ۲۳ سال و عدم مالکیت خودرو - مدرک سیکل، شغل آزاد و عدم محدودیت پزشکی
۵	نقص فنی	- شغل رانندگی یا بازنشستگی
۶	سرعت غیرمجاز	- راننده مرد با درآمد خوب و سن کمتر از ۲۳ سال و بیشتر از ۴۵ سال
۷	دورزدن ممنوع	- راننده خانم و عدم محدودیت پزشکی - راننده مرد با درآمد متوسط و سن کمتر از ۲۳ سال

استخراج قوانین با استفاده از دو مدل (Generalized Rules Induction) GRI و Apriori در نرم‌افزار کلمنتاین صورت گرفته است. نتایج حاصل از این دو مدل در جدول ۱۱ آورده شده است. این نتایج می‌تواند خروجی داده‌کاوی انجام‌شده در این مطالعه موردی باشد.

**جدول ۱۱. گزارش کاربردی برای مدیران راهنمایی و رانندگی**

دقت	عنوان وابستگی	شرح	نام مدل	عنوان مؤلفه وابستگی
٪۱۰۰	تخلف دورزدن ممنوع	- رانندگان خانم خانه‌دار با درآمد خوب	GRI	فیلدهای شغل و درآمد
	تخلف سرعت غیرمجاز	- شغل کارمند و درآمد خوب		
٪۵۵	تخلف سرعت غیرمجاز	شغل آزاد و درآمد خوب	Apriori	
٪۶۹	تخلف سرعت غیرمجاز	رانندگان مرد مجرد، مدرک دیپلم و مالک خودرو.	GRI	فیلدهای جنسیت، مدرک تحصیلی، وضعیت تأهل و مالکیت خودرو
		رانندگان مرد مجرد، مدرک لیسانس و مالک خودرو	Apriori	
٪۵۴	تخلف عبور از چراغ قرمز	رانندگان با مدرک دیپلم، شغل آزاد و نوع رانندگی مبتدی	GRI	فیلدهای مدرک تحصیلی، شغل و نوع رانندگی
٪۴۰	تخلف سرعت غیرمجاز	رانندگان دانشجوی با مدرک لیسانس و نوع رانندگی مبتدی	Apriori	

به‌طور کلی، الگوریتم GRI با درصد بالاتری نسبت به Apriori توانسته به استخراج قوانین پردازد؛ بنابراین در پژوهش‌های مشابه می‌توان از این الگوریتم بهره گرفت.

## نتایج و پیشنهادها

نتایج به‌دست‌آمده از این پژوهش بیان می‌کند که فضای خالی بین تئوری و اجرا هنوز بایستی کمتر شود؛ به‌ویژه در حوزه پلیس به‌منظور تغییر مفاهیم استفاده از داده‌کاوی در جایی که اطلاعات تخلفات ترافیکی رابطه مستقیمی با افزایش تصادفات و به‌تبع آن، افزایش خسارات مالی و جانی ناشی از آن دارد. از آنجاکه این پژوهش، شواهد کاربردی را توسط داده‌کاوی در حوزه تخلفات ترافیکی بیان کرده است، می‌تواند منبع باارزشی باشد.

نجات جان انسان‌ها به‌سبب شناسایی رانندگان متخلف و کاهش ارتکاب به تخلفات توسط آنان و به‌تبع آن، کاهش تصادفات جاده‌ای، از اهمیت بالایی برخوردار است. این روش باید علاوه بر کم‌هزینه‌بودن، دارای خروجی مطلوب نیز باشد. طراحی سیستم‌های کمک تصمیم برای شناسایی رفتار رانندگان متخلف می‌تواند به‌خوبی به پلیس در تصمیم‌گیری کمک کند. اگر این سیستم‌ها با راهکارهای داده‌کاوی طراحی شوند، تمامی عوامل مؤثر بر شناسایی رفتار رانندگان را در نظر گرفته و درصد احتمال موفقیت را بالا می‌برند. همچنین باوجود اینکه بسیاری از ابزارهای داده‌کاوی در دسترس هستند؛ اما در حوزه پلیسی، استفاده از این فناوری، امری به‌شدت مرتبط به دامنه و حوزه ترافیکی مورد استفاده بوده که در کنار نیاز به دانش حرفه‌ای‌های داده‌کاوی برای موفق شدن، به حضور افراد حرفه‌ای در حوزه حمل‌ونقل و ترافیک نیز نیاز است. از دستاوردهای این پژوهش می‌توان به موارد زیر اشاره کرد:

الف) امکان پیاده‌سازی این پژوهش به‌عنوان یک بسته نرم‌افزاری برای سیستم‌های ثبت و کنترل تخلفات ترافیکی شهر کرمان وجود دارد.

ب) با بررسی‌های صورت گرفته روی داده‌ها مشخص شده است:





رانندگان پرداخت کنند؛ اما رانندگانی که بااحتیاط رانندگی می‌کنند، با احتمال کمتری در معرض تخلف و تصادف هستند. بنابراین شرکت‌های بیمه باید نرخ بیمه را بر اساس خطرپذیری رانندگان تعیین کنند؛ یعنی افراد پرخطر باید حق بیمه بیشتر و افراد بااحتیاط باید حق بیمه کمتری بپردازند.

با توجه به نتایج این پژوهش و دسته‌بندی و خوشه‌بندی رانندگان بر اساس خطرپذیری‌شان، اغلب شرکت‌های بیمه می‌توانند از این طریق، اطلاعات دقیقی در خصوص این موضوع در اختیار داشته باشند. در دیگر کشورها، حق بیمه بر اساس ویژگی‌های شخصی رانندگان تعیین می‌شود؛ یعنی سن، جنسیت، سواد، تجربه رانندگی (سال اخذ گواهینامه)، سال‌هایی که راننده از مزایای بیمه‌نامه استفاده نکرده است و موارد مشابه، در نرخ بیمه تأثیر داده می‌شود. بنابراین تاریخچه رانندگی هر راننده به شرکت بیمه کمک می‌کند تا از پرخطر یا کم‌خطر بودن راننده آگاهی یابد.

با توجه به مهم بودن مسئله شناسایی و پیش‌بینی رفتار متخلفین به‌وسیله پلیس، توسعه و بهبود مدل ارائه‌شده ضروری به نظر می‌رسد. بر اساس تجربیات و نتایج حاصل از این پژوهش، پیشنهادهای ذیل برای پژوهش‌های آتی ارائه می‌گردد:

الف) بازنگری و ارائه راهکارهای مناسب در بهبود بانک اطلاعاتی موجود و بهبود روش‌های جمع‌آوری اطلاعات بر اساس اهداف شناسایی متخلفین.

ب) از طریق بررسی ویژگی‌ها و خصوصیات که از دیدگاه پلیس، تخلف خطرآفرین و حادثه‌ساز محسوب می‌شود، متغیری به نام درجه خطرسازی رانندگان استخراج شده و سپس خوشه‌ها بر اساس این هدف تشکیل گردد.

پ) در این پژوهش، با استفاده از الگوریتم‌ها به‌صورت ترکیبی به نتایجی دست یافتیم که از نتیجه الگوریتم‌ها به‌صورت مستقل بهتر بود. اثرات ترکیب تکنیک‌ها به‌عنوان راهکاری ترکیبی در این مقاله بیشتر نمایان شده است؛ بنابراین پیشنهاد می‌شود در پژوهش‌های آتی از روش‌های ترکیبی الگوریتم‌ها و پیاده‌سازی آن‌ها در متلب استفاده شود و خروجی‌ها مورد بررسی قرار گیرد.

- Ayuso, M., Guillén, M., & Alcañiz, M. (2010). The impact of traffic violations on the estimated cost of traffic accidents with victims. *Accident Analysis & Prevention*, 42(2), 709-717. (<http://www.sciencedirect.com/science/article/pii/S0001457509002991>)
- Fogue, M., Garrido, P., Martinez, F. J., Cano, J. C., Calafate, C. T., & Manzoni, P. (2012). Using data mining and vehicular networks to estimate the severity of traffic accidents. In *Management Intelligent Systems*. 37-46. Springer Berlin Heidelberg. ([http://link.springer.com/chapter/10.1007/978-3-642-30864-2\\_4](http://link.springer.com/chapter/10.1007/978-3-642-30864-2_4))
- Guo, X. (2016). Traffic Flow Forecasting Model Based on Data Mining. *System*, 9 (10), 1043-1046. ([https://scholar.google.com/scholar?q=Traffic+Flow+Forecasting+Model+Based+on+Data+Mining&btnG=&hl=en&as\\_sdt=0%2C5](https://scholar.google.com/scholar?q=Traffic+Flow+Forecasting+Model+Based+on+Data+Mining&btnG=&hl=en&as_sdt=0%2C5))
- Kawasaki, Y., Hara, Y., Mitani, T., & Kuwahara, M. (2016). Real-Time Simulation of Dynamic Traffic Flow with Traffic Data Assimilation Approach. *Journal of Disaster Research*, 11(2), 246-254. (<https://www.fujipress.jp/jdr/dr/dsstr001100020246/>)
- Last, M., Avrahami, G., Kandel, A. (2011). Using data mining techniques for optimizing traffic signal plans at an urban intersection. *International Journal of Intelligent Systems*, 26(7), 603-620. (<http://onlinelibrary.wiley.com/doi/10.1002/int.20473/full>)
- Moghaddam, M., Ayati, E. (2007). Introducing a risk estimation index for drivers: A case of Iran. *Safety Science*, 62, 90-97. (<http://profdoc.um.ac.ir/articles/a/1036938.pdf>)
- Nejad, S.K., Seifi, F., Ahmadi, H., Seifi, N. (2009). Applying Data Mining in Prediction and Classification of Urban Traffic. *World Congress on Computer Science and Information Engineering*, 3, 674-678. (<ftp://161.24.19.221/ele/jrsantos/Leitura/05170926.pdf>)
- Özkan, T., & Lajunen, T. (2005). A new addition to DBQ: Positive driver behaviours scale. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(4), 355-368. ([https://www.researchgate.net/profile/Tuerker\\_Oezkan/publication/223324440\\_A\\_new\\_addition\\_to\\_DBQ\\_Positive\\_Driver\\_Behaviours\\_Scale/links/54f1aa6e0cf2b36214aca055.pdf](https://www.researchgate.net/profile/Tuerker_Oezkan/publication/223324440_A_new_addition_to_DBQ_Positive_Driver_Behaviours_Scale/links/54f1aa6e0cf2b36214aca055.pdf))
- Pakgozar, A., Sigari Tabrizi, R., Khalili, M., Esmaili, A. (2011). The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. *Procedia Computer Science*. 3. 764-769. (<http://www.sciencedirect.com/science/article/pii/S1877050910005016>)
- Shiau, Y. R., Tsai, C. H., Hung, Y. H., & Kuo, Y. T. (2015). The

Application of Data Mining Technology to Build a Forecasting Model for Classification of Road Traffic Accidents. *Mathematical Problems in Engineering*, 57(2), 1-9.

(<http://www.hindawi.com/journals/mpe/2015/170635/abs/>)

- Solomon, S., Nguyen, H., Liebowitz, J., & Agresti, W. (2006). Using data mining to improve traffic safety programs. *Industrial Management & Data Systems*, 106(5), 621-643.

(<http://www.emeraldinsight.com/doi/abs/10.1108/02635570610666412>)

- Spalević, Ž., Ilić, M., & Arsić, N. (2016). Legal aspects of data mining algorithms for stream processing in traffic surveillance. *Nauka, bezbednost, policija*, 21(1), 57-80. (<http://scindeks-clanci.ceon.rs/data/pdf/0354-8872/2016/0354-88721601057S.pdf>)

- Tseng, C. M. (2013). Speeding violations related to a driver's social-economic demographics and the most frequent driving purpose in Taiwan's male population. *Safety science*, 57, 236-242.

(<http://www.sciencedirect.com/science/article/pii/S0925753513000453>)

- Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3-19.

([https://www.researchgate.net/profile/Matthew\\_Karlaftis/publication/260031337\\_Short-](https://www.researchgate.net/profile/Matthew_Karlaftis/publication/260031337_Short-term_traffic_forecasting_Where_we_are_and_where_were_going/links/00b495319ce383d9a9000000.pdf)

[term\\_traffic\\_forecasting\\_Where\\_we\\_are\\_and\\_where\\_were\\_going/links/00b495319ce383d9a9000000.pdf](http://www.researchgate.net/profile/Matthew_Karlaftis/publication/260031337_Short-term_traffic_forecasting_Where_we_are_and_where_were_going/links/00b495319ce383d9a9000000.pdf))

- Wu, B., Zhou, W. J., & Zhang, W. D. (2003). The applications of data mining technologies in dynamic traffic prediction. In *Intelligent Transportation Systems, 2003. Proceedings*, 1(1), 396-401. ([http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1251984&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D1251984](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1251984&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D1251984))

- Wang, R., Ji, W., Liu, M., Wang, X., Weng, J., Deng, S., Gao, S. and Yuan, C.A. (2018). Review on mining data from multiple data sources. *Pattern Recognition Letters*, 109, 120-128. (<https://www.sciencedirect.com/science/article/abs/pii/S0167865518300199>)

- Ya-jun, W., Zhao-cheng, Y., Song-bai, H., Jie-Li, C., & Zhi-Jun, Z. (2009). A data-mining-based study on road traffic information analysis and decision support. In *Web Mining and Web-based Application*. pp. 24-27. ([http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5232459&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D5232459](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5232459&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D5232459))

- Yu, J., Kang, H., Park, D., Bang, H. C., & Kang, D. W. (2013). An in-

depth analysis on traffic flooding attacks detection and system using data mining techniques. *Journal of Systems Architecture*, 59(10), 1005-1012. (<http://www.sciencedirect.com/science/article/pii/S1383762113001562>)

- Zamani, Z., pourmand, M., saraee, M. (2010). Application of Data minig in Traffic Management: Case of city of Isfahan. International Conference on Electronic Computer Technology (ICECT). (<https://core.ac.uk/download/files/130/1667285.pdf>)

- Zhang, G., Yau, K., & Chen, G. (2013). Risk factors associated with traffic violations and accident severity in China. *Accident Analysis & Prevention*, 59, 18-25. (<http://202.114.18.201:8080/u/cms/www/201312/17155833ta1o.pdf>)

