

تحلیل راهبردی تخلفات رانندگان با رویکرد داده کاوی

غلامرضا شاه محمدی^۱

آرش امیری^۲

تاریخ پذیرش نهایی: ۹۸/۱۱/۳۰

تاریخ دریافت: ۹۸/۹/۲۵

فصلنامه‌ی مطالعات راهبردی ناجا / سال چهارم / شماره چهاردهم - زمستان ۱۳۹۸* ۱۵۳-۱۸۰

چکیده

یکی از عوامل بروز تخلفات و حوادث ترافیکی، بی‌توجهی به مقررات راهنمایی و رانندگی است که با توجه به بافت فرهنگی، اجتماعی، اقتصادی و جغرافیایی جوامع، میزان و نوع آن متفاوت خواهد بود. آمار بالای تخلفات ترافیکی و حوادث ناشی از آن در کشور، باعث بروز صدمات جبران‌ناپذیری می‌گردد که در این میان، عوامل انسانی اصلی‌ترین عامل بروز تخلفات قلمداد می‌گردد. هدف این تحقیق، تحلیل راهبردی تخلفات رانندگان و بررسی ارتباط آن با مشخصات فردی آنان و مراکز صدور گواهینامه با رویکرد داده کاوی می‌باشد. این تحقیق از نظر هدف، کاربردی و از نوع تحلیل محتوا است. جامعه آماری تحقیق، اطلاعات پایگاه داده صدور گواهینامه و اجرائیات و حجم نمونه اطلاعات این پایگاه‌ها در سال‌های ۱۳۹۵ و ۱۳۹۶ شهر تهران است. شیوه تحلیل داده‌ها، کشف قوانین انجمنی است. نتایج حاصل از اهم قوانین استخراج‌شده بدین شرح است: ۱. بین سن و تحصیلات رانندگان و تخلفات رانندگی آنان رابطه معنی‌داری وجود دارد؛ به نحوی که بیشترین تکرار تخلفات در قوانین کشف‌شده مربوط به رانندگان مرد با مدرک دیپلم و سن ۲۰ سال است؛ ۲. بیشترین تکرار تخلف انجام شده، تخلف ورود ممنوع می‌باشد؛ ۳. بین برخی تخلفات رانندگان و برخی آموزشگاه‌های رانندگی قوانین رانندگی ارتباط وجود دارد. اغلب قوانین کشف شده مورد تأیید کارشناسان راهور است و نتایج حاصل از قوانین، با نظریه‌های مربوط به دلایل بروز کج‌رفتاری‌ها و تخلفات انطباق دارد. تحقیق حاضر، متغیرهای موثر در تخلفات رانندگی را در نظر گرفته‌است؛ بنابراین، می‌توان شیوه ارائه آموزش به متقاضیان گواهینامه را با توجه به قوانین حاصل، برنامه‌ریزی و از بروز افزایش تخلفات پیشگیری نمود.

کلیدواژه‌ها:

داده کاوی، تحلیل تخلفات، کیفیت آموزش، تخلفات رانندگی، مشخصات فردی

Drshahmohammadi@chmail.ir

Arash_ami100@Yahoo.Com

۱ دانشیار مهندسی کامپیوتر دانشگاه علوم انتظامی امین (نویسنده مسئول)

۲ کارشناس پلیس راهور ناجا

مقدمه

آمار بالای تخلفات ترافیکی و حوادث ناشی از آن در کشور، موجب بروز صدمات جانی و مالی جبران‌ناپذیری می‌گردد که عوامل انسانی در این میان، اصلی‌ترین سبب بروز تخلفات می‌باشند. رشد فزاینده آمار تخلفات در ایران (در سال ۱۳۹۵ بیش از ۵۷ میلیون است و در سال ۱۳۹۴ بیش از ۶۸ میلیون تخلف ثبت شده است). مقایسه این آمار با سایر کشورهای جهان نشان‌دهنده وضعیت ناهنجار ایمنی عبور و مرور در کشور است. یکی از عوامل بروز تخلفات و حوادث ترافیکی در همه جوامع، بی‌توجهی به مقررات راهنمایی و رانندگی توسط راننده می‌باشد که با توجه به بافت فرهنگی، اجتماعی، اقتصادی و جغرافیایی جوامع، میزان و نوع آن متفاوت خواهد بود. از آنجا که آموزش یکی از ارکان اصلی برنامه‌ریزی ترافیک می‌باشد، توجه به آن در کسب نتیجه مطلوب و کاهش تخلفات رانندگی بسیار حائز اهمیت است. از سوی دیگر، تحلیل ارتباط اطلاعات فردی رانندگان با انواع تخلفات رانندگی آنان نیز می‌تواند رهنمودهای خوبی را درباره مشخصات رانندگان متخلف و انجام برنامه‌ریزی‌های مناسب در حین ارائه آموزش‌های رانندگی یا پس از دریافت گواهینامه به منظور اجتناب از تخلفات رانندگی ارائه دهد. فناوری اطلاعات ابعاد مختلف زندگی و کسب و کار انسان‌ها را متحول نموده است و گستره و عمق این تأثیر به سرعت در حال افزایش است. بررسی شرایط جامعه اسلامی ایران نشان می‌دهد که بخش عمده‌ای از فعالیت‌های سازمان‌ها با اتکا به فناوری اطلاعات انجام شده و بسیاری از کسب و کارهای فن‌آورمحور با استفاده از بستر اینترنت صورت می‌پذیرد. نیروی انتظامی نیز به عنوان حافظ نظم و امنیت کشور، از سال‌ها پیش به صورت منسجم و گسترده از فناوری اطلاعات در انجام مأموریت‌ها و ارائه خدمات انتظامی بهره برده و پیوسته در تلاش است تا همسو با آخرین فناوری‌های اطلاعاتی و ارتباطی در میدان وظیفه ایفای نقش نماید. رویکرد داده‌کاوی نیز یکی از رویکردهای فناوری اطلاعات است که در بخش‌هایی با حجم وسیعی از اطلاعات، امکان تحلیل داده‌ها و کشف دانش پنهان موجود در داده‌ها را فراهم ساخته و از این طریق، توانایی برنامه‌ریزی، افزایش بهره‌وری و قدرت تصمیم‌سازی مناسب در آن بخش را میسر می‌سازد. نیروی انتظامی با توجه به سامانه‌های اطلاعاتی متعددی که از سال‌ها پیش در حوزه‌های مختلف مأموریتی ایجاد نموده است، حجم عظیمی از اطلاعات را در اختیار دارد که تجزیه و تحلیل این اطلاعات به مدد رویکرد داده‌کاوی، امکان کشف روابط، روندها و الگوهای مخفی بین داده‌ها و دستیابی به دانش نوین در زمینه چالش‌های آشکار و نهان ناجا را میسر خواهد ساخت.

در نیروی انتظامی، اطلاعات مرتبط با گواهینامه‌های صادر شده برای رانندگان و آموزشگاه‌های رانندگی که متولی آموزش متقاضیان دریافت گواهینامه می‌باشند و همچنین، اطلاعات تخلفات رانندگی در بانک‌های اطلاعاتی مختلف در اختیار پلیس راهور ناجا است و با توجه به حجم زیاد این اطلاعات، امکان استفاده از رویکرد داده کاوی برای تشخیص ارتباط تخلفات رانندگان با مشخصات فردی رانندگان و آموزشگاه‌های رانندگی وجود دارد؛ از این رو، تحقیق حاضر، با هدف تحلیل راهبردی تخلفات رانندگان با استفاده از رویکرد داده کاوی به بررسی پرداخته‌است تا از این طریق، نقاط قوت و ضعف موجود شناسایی شده و به مدد برنامه ریزی‌های صحیح، از افزایش تخلفات پیشگیری گردد.

پیشینه تحقیق

برخی از پژوهش‌های انجام شده در زمینه استفاده از روش‌های داده کاوی به شرح زیر است: جعفری و صمدیان (۱۳۹۱) در مقاله‌ای با عنوان «کاربرد داده کاوی در بررسی رفتار رانندگان متخلف در کلان شهرها» به استفاده از داده کاوی در داده‌های تخلفات ترافیکی پرداخته‌اند. در این مقاله، روش داده کاوی با هدف دستیابی به قوانین تصمیم‌گیری و الگوی شناسایی رانندگان پرخطر در تخلفات مورد استفاده واقع شده‌است. در اجرای روش دسته‌بندی در داده کاوی دو روش شبکه‌های عصبی و درخت تصمیم در اطلاعات تخلفات به کار گرفته شده‌است و در انتها برای بهبود نتایج، از رویکرد ترکیبی شامل روش خوشه‌بندی و درخت تصمیم استفاده شد. احمدی و علی‌محمدی (۱۳۹۴) طی تحقیقی، تصادفات رانندگی بزرگراه کرج - قزوین را با استفاده از تراکم پنجره ای فازی مورد تحلیل زمانی و مکانی قرار دادند. در این تحقیق، تغییر الگوی زمانی و مکانی تصادفات با توجه به زمان، عامل انسانی، نوع برخورد و نوع تصادف بررسی شده‌است. نتایج نشان داده‌است که عوامل مختلف مانند ساعت، نوع برخورد، نوع تصادف، نقش عامل انسانی، روزهای هفته و وضعیت جوی، الگوی مکانی مشخصی دارند و با بررسی آنها می‌توان علت الگوهای مختلف را کشف کرد.

شکوهیار و همکاران طی تحقیقی در سال ۱۳۹۴، به خوشه‌بندی داده‌های تصادفات جاده‌ای با استفاده از فنون داده کاوی پرداختند. نتایج نشان می‌دهد که متغیرهای جاده‌ای مؤثر در بروز تصادفات به ترتیب اهمیت شامل هندسه محل، جهت حرکت راه، نوع منطقه، نقص راه، نوع جاده، مانع دید، نوع خط‌کشی، نوع روبه راه و بی‌توجهی به مقابل است. نتایج به‌دست

آمده از الگوریتم درخت دسته‌بندی و رگرسیون نشان داد که سهم هر متغیر در تصادفات منجر به جرح، فوتی و خسارتی چه میزان است که در این میان، مهم‌ترین عامل، نوع هندسه راه معرفی شد.

هوشیار و شریفی طی تحقیقی در سال ۱۳۹۵، تصادفات درون‌شهری شهر ارومیه در سال ۱۳۹۲ را مورد تحلیل فضایی قرار دادند و نتایج تحقیق آنها نشان داد که بیشترین تعداد تصادفات در مناطق ۱ و ۳ و اطراف بخش مرکزی شهر ارومیه رخ داده‌است و توزیع نقاط تصادفی برای شهر ارومیه به صورت خوشه‌ای بوده‌است.

شاه محمدی و عباسی در مقاله‌ای در سال ۱۳۹۷ به بررسی عوامل موثر در تصادفات درون‌شهری با استفاده از روش‌های داده‌کاوی در شهر اصفهان پرداختند. جامعه آماری تحقیق اطلاعات، پایگاه داده شهر اصفهان و حجم نمونه، اطلاعات تصادفات سال‌های ۱۳۹۰ تا ۱۳۹۴ می‌باشد و برای تحلیل اطلاعات، از کشف قوانین انجمنی تکنیک سبدخرید و تولید درخت اشیاء مکرر FP-Growth استفاده شد. براساس نتایج، ویژگی‌های موثر در بروز تصادف به ترتیب ۱. عجله و شتاب؛ ۲. عدم استفاده از کمربند و کلاه ایمنی؛ ۳. عدم توجه به جلو؛ ۴. نبود شانه راه؛ ۵. مجهز نبودن خودرو به تجهیزات ایمنی می‌باشد. با کمک نتایج این تحقیق امکان پیشگیری و کاهش تصادفات افزایش می‌یابد.

شاه محمدی و رجبی (۱۳۹۷) در مقاله خود، به تحلیل جرایم اخلاقی فضای سایبر با رویکرد داده‌کاوی پرداختند. جامعه آماری پژوهش، پایگاه داده جرایم سایبری پلیس فتا است و حجم نمونه، بخشی از پایگاه داده جرایم سایبری است که مربوط به جرایم اخلاقی فضای سایبر در سال‌های ۱۳۸۹ تا ۱۳۹۴ می‌باشد. نتایج نشان داد که بهترین الگو برای تحلیل و استخراج قوانین حاکم بر داده‌های جرایم اخلاقی فضای سایبر، الگوی درخت تصمیم C5 است؛ همچنین، مشخصه‌های تأثیرگذار بر وقوع جرایم اخلاقی فضای سایبر به ترتیب شغل، تحصیلات، جنسیت، سن، تأهل و نقش فرد است. در این پژوهش، قوانین حاکم بر جرم اخلاقی فضای سایبر احصا گردید که با تحلیل این قوانین می‌توان راهکارهای پیشگیرانه‌ای را ارائه نمود.

بررسی تحقیقات مرتبط نشان می‌دهد که با وجود تحقیقات متعددی که در زمینه داده‌کاوی در حوزه راهنمایی و رانندگی انجام شده‌است، هیچ یک از تحقیقات انجام‌شده مشابه تحقیق حاضر نمی‌باشد.

مبانی نظری تحقیق

الف. دلایل بروز کج رفتاری ها و تخلفات

نظر رابرت مرتن ۱ این است که جامعه، فرد را به کج رفتاری وادار می کند. به بیان خود او، کج رفتاری حاصل فشارهای ساختاری اجتماعی خاصی است که افراد را به کج رفتار شدن وا می دارد. به نظر مرتن، جوامع صنعتی جدید بر توفیقات مادی در زندگی جدید تأکید دارند که به صورت «انباشت ثروت» و «تحصیلات علمی» به عنوان مهم ترین اهداف زندگی شخص و معیارهای منزلتی تجلی می یابد. دستیابی به این اهداف مقبول اجتماعی نیاز به ابزارهای مقبولی هم دارد که البته از دسترس گروهی از افراد جامعه خارج است؛ یعنی جامعه به گونه ای سازمان یافته است که طبقات فرودست از فرصت های کمتری برای تحقق آرزوهای خود برخوردار می باشند؛ اما از آنجا که این اهداف به آرمان های اصلی زندگی همه افراد (فقیر و غنی) تبدیل شده است، افرادی که جهت تحقق این اهداف، به ابزار مشروع دسترسی ندارند، تحت فشار جامعه برای دستیابی به آنها از ابزار نامشروع استفاده می کنند (صدیق سروستانی، ۱۳۸۷: ۴۴).

یکی از نظریه های تبیین گر تخلفات، نظریه فشار است. سؤال اصلی در نظریه فشار این است که «چرا مردم کج رفتاری می کنند؟» و پاسخ کلی این سؤال این است که «عواملی در جامعه وجود دارند که برخی مردم را زیر فشار قرار می دهد و آنان را مجبور به کج رفتاری می کنند». مرتن این فشار را ناشی از عدم توانایی شخص در دستیابی به اهداف مقبول جامعه و یا اجتماعی می داند. کوئن ناکامی در رسیدن به جایگاه بالا در جامعه را عامل فشار می شمارد. کلواردوالین علت آن را عدم برخورداری اشخاص از فرصت های نامشروع برای نیل به هدف واردکننده بر افراد و راندن آنان به سوی کج رفتاری می داند (صدیق سروستانی، ۱۳۸۷: ۴۵-۴۴).

نظریه کنترل هم در تحلیل و تبیین کج رفتاری های اجتماعی مورد استفاده قرار گرفته است و از نظریه های با نفوذ حوزه جامعه شناسی انحرافات اجتماعی و جرم بوده است. موضوع اصلی در این نظریه، این است که کج رفتاری نتیجه نبود کنترل اجتماعی است. فرض اصلی در این نظریه این است که همان طور که فریود گفته است، افراد به طور طبیعی تمایل به کج رفتاری دارند و اگر تحت کنترل قرار نگیرند، مرتکب کج رفتاری خواهند شد؛ بنابراین، فقدان یا ضعف کنترل اجتماعی علت اصلی کج رفتاری است (سلیمی و داوری، ۱۳۸۵: ۳۹۱).

1 Merton

نظریه انتقال فرهنگی، تخلفات راهنمایی و رانندگی را به عنوان رفتاری توجیه می کند که مانند همنوایی آموخته می شود (رابرتسون، ۱۳۷۷: ۶۴). ساترلند^۱ در این باره از تفاوت ارتباط استفاده کرد (گیدنز، ۱۳۷۳: ۹۹). وی تفاوت ارتباط را در خصوص محیط‌هایی که مشوق افراد در انجام فعالیت‌های غیرقانونی و انحرافی هستند، به کار می برد. برخی محیط‌ها و گروه‌های اولیه نظیر دوستان، همسالان و خانواده، در انتقال رفتار انحرافی به اعضای خود بسیار تأثیرگذار هستند (کوزر و روزنبرگ، ۱۳۷۸: ۵۸).

ب. کارکرد داده کاوی

در دهه‌های اخیر، حجم داده‌های ذخیره شده در پایگاه داده یا هر منبع دیگری افزایش چشم‌گیری داشته است و پیوسته رو به افزایش است؛ به طوری که تحلیل این داده‌ها بدون استفاده از فناوری جدید و ابزارهای خودکار بسیار دشوار و غیرممکن شده است. فناوری‌های نوین اطلاعاتی و ارتباطی و همچنین، فناوری‌های پشتیبان تصمیم، با جمع‌آوری، ذخیره، ارزیابی، تفسیر و تحلیل، بازیابی و اشاعه اطلاعات و دانش به کاربران خاص، می‌توانند در اطلاع‌یابی به موقع، صحیح و موردنیاز به افراد نقش مهمی را ایفا نمایند. یکی از ابزارهای مورد استفاده در این فناوری‌ها، داده کاوی می‌باشد. داده کاوی شامل استفاده از ابزارهای پیشرفته تحلیل داده به منظور کشف الگوهای معتبر، از قبل ناشناخته و روابط در مجموعه داده‌های بزرگ است (اسماعیلی، ۱۵، ۱۳۹۴). داده کاوی ترکیبی از روش‌های یادگیری ماشین، تشخیص الگو، آمار، نظریه پایگاه داده و تخلیص و ارتباط بین مفاهیم و الگوهای جالب به صورت خودکار از پایگاه داده شرکت‌های بزرگ است. هدف اصلی داده کاوی، کمک به فرآیند تصمیم‌گیری از طریق استخراج دانش از داده‌هاست (الپایدین^۲، ۲۰۱۰). داده کاوی فرآیند کشف قوانین و دانش ناشناخته و مفید از انبوه داده‌ها و پایگاه داده است. انجام عملیات داده کاوی شامل موارد زیر است: ۱. جداسازی داده مفید از داده بیگانه؛ ۲. یکپارچه‌سازی داده‌های مختلف تحت یک قالب واحد؛ ۳. انتخاب داده لازم از میان دیگر داده‌ها؛ ۴. انتقال داده به محیط داده کاوی جهت اکتشاف قوانین؛ ۵. ایجاد الگوهای مرتبط به وسیله روش‌های داده کاوی؛ ۶. ارزیابی الگوهای ایجادشده جهت تشخیص میزان مفید بودن آنها؛ ۷. انتشار دانش استخراج شده به کاربران نهایی (لین و همکاران^۳، ۲۰۱۲).

1 Satherland

2. Alpaydin

3. Lin et al.

ج. قوانین انجمنی

قوانین انجمنی کاوش قوانین انجمنی در راستای کشف ارتباطات جالب و با اهمیت بین اقلام اطلاعاتی در پایگاه داده‌های بزرگ و انبارهای تراکنش می‌باشد (صنيعی آباده و همکاران، ۱۳۹۴: ۱۵۹). یک قانون انجمنی با عبارت $X \rightarrow Y$ بیان می‌شود که در آن X و Y مجموعه اقلام غیرتهی هستند که هیچ‌گونه اشتراکی ندارند ($X \cap Y = \emptyset$)؛ دو معیار پشتیبان و اطمینان به منظور ارزیابی قوانین انجمنی استفاده می‌شوند؛ هر چند که معیارها فقط به این دو ختم نمی‌شوند. مقدار پشتیبان نشان می‌دهد که در چند درصد از تراکنش‌های پایگاه داده‌ها می‌توان مجموعه اقلام X و Y را همراه یکدیگر پیدا کرد و مقدار اطمینان^۱ در میان تراکنش‌هایی که مجموعه اقلام X را در خود دارند، به دنبال مجموعه اقلام Y می‌گردد. تراکنش‌های حاوی X می‌تواند شامل Y نباشند و برعکس؛ بدین سبب، دو قانون انجمنی $X \rightarrow Y$ و $Y \rightarrow X$ یکسان نیستند.

د. تخلفات رانندگی

تعیین انواع تخلفات رانندگی و احصای آن توسط قانونگذار، بستگی به سیاست کیفی و سلامت‌سازی جامعه دارد که بر حسب شرایط مکان و زمان و در کشورهای مختلف تفاوت دارد. از آنجا که تخلفات رانندگی در بعضی از مواقع، منتج به تصادف و ایجاد خسارت مالی و جانی می‌گردد، بنابراین، نتیجه حاصل از تخلفات رانندگی را می‌توان به دو بخش تقسیم کرد: جانی و مالی (بهرامی، ۱۳۷۹).

ه. کیفیت آموزش‌گاه‌های رانندگی

از آنجا که بخش اعظم تخلفات ناشی از خطاها و عوامل مرتبط با رانندگان است، بر این اساس، تسلط کافی مربیان آموزشگاه‌ها به امور و قوانین حمل‌ونقل و ترافیک و کیفیت آموزش‌های نظری و عملی ارائه شده به رانندگان در آموزشگاه‌های رانندگی از اهمیت ویژه‌ای در کاهش میزان تخلفات برخوردار است؛ بنابراین، تعیین صلاحیت و آزمون آزمایش مربیان آموزشگاه‌های تعلیم رانندگی از مواردی است که بایستی مورد توجه قرار گیرد. علاوه بر این، ضروری است که با همکاری مهندسان ترافیک با تجربه و افسران کارآزموده راهنمایی و رانندگی و سایر کارشناسان روانشناسی آموزشی و کارشناسان حقوقی، با استفاده از تجربیات سایر کشورها، تجزیه و تحلیل مراحل و محتوای آموزش، آزمون و صدور گواهینامه صورت پذیرد.

روش تحقیق

این تحقیق از نظر هدف، کاربردی و از نوع تحلیل محتوا می‌باشد. جامعه آماری این تحقیق، اطلاعات پایگاه داده صدور گواهینامه و اجرائیات و حجم نمونه اطلاعات این پایگاه‌ها در سال‌های ۱۳۹۵ و ۱۳۹۶ مربوط به کلان‌شهر تهران می‌باشد و شیوه تحلیل داده‌ها، رویکرد داده‌کاوی است. در این تحقیق، با ادغام اطلاعات دو پایگاه داده و انجام آماده‌سازی داده‌ها از روش کشف قوانین انجمنی برای استخراج دانش پنهان موجود در اطلاعات استفاده گردیده‌است.

فرآیند انجام تحقیق

به منظور انجام فرآیند این تحقیق، استاندارد جهانی کریسپ^۱ مورد استفاده قرار گرفته‌است که یکی از روش‌های تحقیقی بسیار قوی در روش‌های داده‌کاوی می‌باشد و شامل گام‌های شناخت سامانه و جمع‌آوری داده‌ها، آماده‌سازی داده‌ها، الگوسازی و ارزیابی می‌باشد. شیوه تجزیه و تحلیل اطلاعات رویکرد داده‌کاوی از طریق کشف قوانین انجمنی است.

برای انجام این تحقیق، ضمن مرور بر ادبیات موضوع، تلاش شد تا با استفاده از فنون داده‌کاوی، قواعد و رابطه‌های بین آموزشگاه‌های رانندگی، رانندگان و تخلفات ترافیکی کشف گردد. ابتدا با مطالعه و بررسی داده‌های موجود، شاخص‌های موثر بر کیفیت آموزش در آموزشگاه‌ها و روابط میان مشخصه‌های فردی راننده شناسایی شد و بانک اطلاعاتی تشکیل گردید و داده‌های اطلاعاتی پیش‌پردازش و پاکسازی شد؛ یعنی داده‌هایی که برای هدف مورد نظر مناسب نیستند، حذف شده و در صورت لزوم، برای داده‌های گمشده، مقدار مناسب جایگزین گردید. در مرحله بعد، با استفاده از فنون داده‌کاوی، قوانین انجمنی الگوها کشف شده و آماده تجزیه و تحلیل گردید. در این تحقیق که بر پایه روش تحقیق داده‌کاوی و با استفاده از کشف قوانین صورت گرفته‌است، مجموعه داده‌ها با نرم‌افزارهای پایگاه داده اوراکل^۲، اکسس^۳ و اکسل^۴ آماده‌سازی گردیده و استخراج قواعد با استفاده از نرم‌افزارهای راییدمایزر^۵ انجام پذیرفته‌است. همچنین، در تحقیق حاضر، از مطالعه موردی استفاده شده

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

2-Oracle Database

3-Access

4-Excel

5-RapidMiner Studio

و جمع‌آوری داده‌ها از انبار داده سازمان انجام پذیرفته است. الگوریتم پیشنهادی برای این تحقیق به شرح زیر است:

فاز مطالعاتی

- مطالعه مفاهیم صدور گواهینامه، تخلفات رانندگی و داده کاوی؛
- مطالعه کتابخانه‌ای به منظور شناسایی عوامل؛
- مطالعه کتابخانه‌ای در زمینه قوانین انجمنی.

فاز پیاده‌سازی

- انتخاب محدوده جغرافیایی مناسب به منظور مطالعه موردی؛
- جمع‌آوری و آماده‌سازی داده‌ها؛
- کشف قوانین مرتبط با استفاده از الگوریتم FP-Growth؛
- استخراج قوانین به دست آمده و تجزیه و تحلیل آنها.

ارائه چارچوب نظری

- فرآیند کلی داده کاوی شامل چهار مرحله اساسی است که عبارتند از:
 - تعریف مسئله و آماده‌سازی داده‌ها؛
 - ادغام و پاکسازی داده‌ها؛
 - تحلیل داده‌ها و کشف قوانین؛
 - به‌کارگیری الگو و اعتبارسنجی آن.

تعریف مسئله کسب و کار

تعریف مسئله، اهداف و استفاده‌های عملی از قوانین به دست آمده از الگوریتم را روشن می‌کند. قوانین شناسایی کیفیت آموزش در آموزشگاه‌ها و روابط میان مشخصه‌های فردی راننده، می‌توانند برای اهداف مختلف ساخته شوند؛ برای مثال، این الگوریتم می‌تواند به عنوان ابزاری برای پشتیبانی تصمیم‌گیری خبرگان در زمینه شناسایی بهترین نحوه آموزش در آموزشگاه‌های رانندگی به کار رود. کاربرد قوانین شناسایی کیفیت آموزش در آموزشگاه‌ها و روابط میان مشخصه‌های فردی راننده، می‌تواند اساس کل فرآیند داده کاوی در این زمینه باشد.

جمع‌آوری داده‌ها

داده‌های استفاده شده در این تحقیق، از پایگاه داده سامانه صدور گواهینامه و سامانه اجرائیات استخراج گردیده است. اطلاعات این پایگاه داده‌ها، همه روزه بر اساس ثبت اطلاعات متقاضیان صدور گواهینامه و تخلفات رانندگان جمع‌آوری و به روز رسانی می‌شود. در پایگاه داده صدور گواهینامه، اطلاعات مربوط به متقاضیان انواع گواهینامه‌های رانندگی و در پایگاه داده سامانه اجرائیات تخلفات ارتكابی توسط راننده ثبت و نگهداری می‌گردد. لازم به ذکر است که اطلاعات مورد استفاده، تخلفات اطلاعات ثبت شده در سامانه اجرائیات و گواهینامه‌های صادره از آموزشگاه‌های کلانشهر تهران در سال‌های ۱۳۹۵ و ۱۳۹۶ می‌باشد.

پایگاه داده‌های مورد استفاده در این تحقیق، از نوع رابطه‌ای است. در این نوع پایگاه داده‌ها، اطلاعات مختلف در مورد هر رکورد در جدول‌های مختلف نگهداری می‌شوند. برای هر رکورد در جدول، یک کلید شناسایی وجود دارد که از طریق همین کلید، رکوردها به همدیگر مرتبط می‌شوند. دلیل استفاده از این پایگاه داده، علاوه بر سهولت مدیریت آنها، شکستن اطلاعات مختلف در قالب جدول‌های گوناگون و کاهش حجم و اندازه جداول می‌باشد.

با توجه به موضوع، محقق سعی نمود با مطالعه بر روی انواع اطلاعات متقاضیان دریافت گواهینامه و تخلفات آنها، فیلدها و مشخصه‌هایی که در اطلاعات پایگاه‌های داده وجود داشته‌است را به‌نحوی جمع‌آوری و تکمیل نماید تا پس از انجام عملیات داده‌کاوی، نتایج کاملی حاصل شود و میزان صحت قوانین افزایش یابد. برای متقاضیان دریافت‌کننده گواهینامه در پایگاه داده‌های صدور گواهینامه و اجرائیات، اطلاعات زیادی نگهداری می‌شود که در اینجا سه دسته از آنها مورد نیاز می‌باشد: دسته نخست، مربوط به اطلاعات فرد متقاضی است و شامل مشخصات فردی، آدرس، مدرک تحصیلی، محل تولد، تاریخ تولد، شغل و دیگر مشخصه‌های فردی است. دسته دوم، مربوط به نوع درخواست متقاضی دریافت گواهینامه می‌باشد که شامل اطلاعات نوع گواهینامه درخواستی، زمان و مکان دریافت (آموزشگاه، شهر و استان) است؛ دسته سوم نیز اطلاعات مربوط به تخلفات شامل شماره گواهینامه فرد متخلف، نوع تخلف، زمان و مکان تخلف است. در جداول ۱، ۲ و ۳ فهرست مشخصه‌هایی که در هر کدام از جدول‌های پایگاه داده، یک راننده را توصیف می‌کند، ارائه شده است.

جدول ۱. مشخصات گواهینامه متقاضیان

ردیف	فیلد	توضیحات
۱	شماره گواهینامه	شماره گواهینامه اخذ شده توسط فرد متخلف را نشان می‌دهد.
۲	نوع گواهینامه	نوع گواهینامه فرد متخلف را نشان می‌دهد.
۳	محل صدور گواهینامه	شهر صدور گواهینامه متقاضی را نمایش می‌دهد.
۴	تاریخ صدور گواهینامه	تاریخ صدور گواهینامه فرد را به صورت روز، ماه و سال نمایش می‌دهد.
۵	نام آموزشگاه	اسم آموزشگاه محل دریافت گواهینامه متقاضی را نمایش می‌دهد.

جدول ۲. مشخصات تخلف

ردیف	فیلد	توضیحات
۱	نام تخلف	تخلف خودرو را نشان می‌دهد.
۲	شماره گواهینامه متخلف	شماره گواهینامه ثبت شده در برگه های جرایم تسلیمی را نشان می‌دهد.
۳	تاریخ تخلف	تاریخ ارتکاب تخلف را به صورت روز، ماه و سال نشان می‌دهد.

جدول ۳. مشخصات فرد دریافت‌کننده گواهینامه

ردیف	فیلد	توضیحات
۱	کد ملی	کد ملی فرد متخلف را نشان می‌دهد.
۲	تاریخ تولد	تاریخ تولد فرد متخلف را به صورت روز، ماه و سال نشان می‌دهد.
۳	مدرک تحصیلی	نوع مدرک تحصیلی فرد متخلف را نشان می‌دهد.
۴	محل تولد	شهر تولد فرد متخلف را نشان می‌دهد.
۵	جنسیت	جنسیت فرد متخلف را نشان می‌دهد.
۶	شغل	شغل فرد متخلف را نشان می‌دهد.

به منظور سهولت در انجام عملیات و اجرای الگوریتم‌های داده‌کاوی مورد استفاده در این تحقیق، تمام جدول‌های مشخصات ذکر شده به یک جدول تبدیل شده است.

آماده‌سازی داده‌ها

امروزه بانک‌های اطلاعاتی به علت حجم زیاد داده و ارتباط با منابع اطلاعاتی مختلف، در معرض وجود داده‌های متناقض^۱، مفقود^۲ و مغشوش^۳ می‌باشند. صاحب‌نظران علم داده‌کاوی، یکی از علل عمده کیفیت پایین نتایج حاصل از داده‌کاوی را، کیفیت پایین داده‌های ورودی و عدم توجه به مراحل آماده‌سازی داده‌ها می‌دانند. بر این اساس، باید توجه نمود که مرحله آماده‌سازی داده‌ها، یکی از مهم‌ترین مراحل در داده‌کاوی است که شامل پاکسازی داده، تبدیل و یکپارچه کردن داده و کاهش بعد آن می‌باشد. اگرچه این مرحله، به ظاهر ساده‌ترین مرحله کار به نظر می‌رسد اما در واقع یکی از پیچیده‌ترین، حساس‌ترین و زمان‌برترین مراحل

1 Inconsistent

2 Missing

3 Noisy

فرآیند داده کاوی است. به وسیله فنون آماده سازی داده، می توان کیفیت داده ها و در نتیجه، کیفیت نتایج خروجی را افزایش داد. آماده سازی داده ها، یکی از مراحل مشکل و با اهمیت در کشف دانش و داده کاوی است. مرحله آماده سازی داده ها در داده کاوی، فرآیند زمانبر و با اهمیتی است که در حدود ۶۰ تا ۷۰ درصد زمان انجام کل فرآیند داده کاوی را به خود اختصاص می دهد (اچ او، ۲۰۰۰). از آنجا که داده های خام دریافت شده از اطلاعات رانندگان، به منظور تحلیل های توصیفی گردآوری می شوند و اکثر داده ها به صورت ترکیبی هستند، مشکلاتی را در حین کار ایجاد می کنند. پایگاه داده پدید آمده، تعداد رکوردها و حجم داده های آنها بسیار زیاد است و این در حالی است که بسیاری از رکوردها، غیر ضروری به نظر می رسند. برای اینکه حجم داده ها کاهش یابد، بسیاری از داده هایی که در تحقیق حاضر سودمند نیستند، پاکسازی می شوند؛ به طور مثال، تمامی رکوردهایی که مشخصات فرد درخواست کننده گواهینامه، در اقلام اطلاعاتی گواهینامه او به درستی ثبت نشده است و رابطه بین مشخصات و شماره گواهینامه روشن نیست.

ادغام و پاکسازی داده ها

داده ها در پایگاه داده هیچ گاه در حالت عادی برای داده کاوی آماده نیست بنابراین، بایستی با انجام اقداماتی داده ها را به شکلی تبدیل نمود که بتوان در نرم افزارهای مورد نظر، آنها را مورد استفاده قرار داد؛ به عبارت دیگر، بعد از جمع آوری داده ها، برای اینکه قوانین مورد نظر توسط ابزارهای مختلف داده کاوی استخراج شود، باید داده ها پیش پردازش شده و به شکل های مناسب برای استفاده در نرم افزارهای داده کاوی تبدیل شوند. همچنین، خطاها و داده های خالی باید به نحو مناسب پاکسازی شوند. فرآیند پاکسازی داده، شامل تکمیل مقادیر مفقود، هموارسازی داده های مغشوش، شناسایی و حذف و برطرف کردن تناقضات موجود بین داده ها می باشد. از آنجا که سامانه های صدور گواهینامه و اجرائیات در سال های اخیر به گونه ای طراحی گردیده اند که رکوردهایی با فیلدهای خالی نمی پذیرند و کاربران ثبت مشخصات متقاضیان گواهینامه رانندگی و پانچ تخلفات، از زمان کافی برای ثبت اطلاعات برخوردار هستند، داده های مفقوده در این تحقیق بسیار اندک می باشند؛ از این رو، جهت حل این مسئله، رکوردهای دارای مقادیر مفقود زیاد حذف و یا با استفاده از روشی دیگر، اصلاح می شوند.

برای برخورد با مقادیر مفقوده، راهکارهای مختلفی وجود دارد که شامل موارد زیر است: (تورگو و هان ۲۰۰۶)^۱

۱- حذف نمونه‌های دارای مقادیر گم‌شده: این روش وقتی مطلوب است که نمونه موجود، دارای تعداد زیادی ویژگی با مقادیر مفقوده باشد. پرکردن مقدار مفقوده به صورت دستی: این روش زمانبر است و برای پایگاه داده‌های بزرگ قابل استفاده نیست.

۲- پرکردن مقادیر گم‌شده با میانگین مقادیر همان ویژگی در سایر نمونه‌ها.

۳- پرکردن مقادیر گم‌شده با مقداری که به تعداد بیشتر در بین نمونه‌ها تکرار شده است.

۴- پر کردن مقادیر گم‌شده با محاسبه ارتباط بین نمونه‌ها.

۵- پرکردن مقادیر مفقوده با بررسی شباهت بین نمونه‌ها و جای‌گذاری مقدار مفقوده با مقدار همان متغیر در شبیه‌ترین نمونه.

۶- برای شناسایی مقادیر غیرقابل استفاده، سعی شد تا با ایجاد رابطه‌های منطقی بین مشخصه‌ها، نمونه‌هایی که پاسخ آنها غیر قابل اعتماد شناخته شدند، از بانک اطلاعاتی خارج شوند؛ به عنوان نمونه، می‌توان به موارد زیر اشاره کرد:

۱- حذف اطلاعات تخلفاتی که در سامانه ثبت شده بودند اما فاقد شماره گواهینامه فرد متخلف بودند؛ که این امر باعث عدم شناسایی آموزشگاه رانندگی متخلف می‌گردد (نظیر گواهینامه‌های غیرتسلیمی).

۲- حذف رکوردهای اطلاعاتی افرادی که با وجود ثبت مشخصات شناسنامه‌ای کامل، فاقد سایر مشخصات فردی مثل مدرک تحصیلی، شغل و ... بودند.

۳- حذف نمونه‌هایی که مشخصات تخلف ثبت شده در جدول‌ها، فاقد مشخصات دقیق زمانی و مکانی تخلف بودند.

۴- حذف نمونه‌هایی که مشخصات فرد در برگ تخلف، با مشخصات گواهینامه فرد متخلف مانند سال اخذ، نوع گواهینامه و ... مغایر بود.

۵- هماهنگ‌سازی داده در این مرحله، رویه‌ای سخت و در عین حال، بسیار مهم برای رسیدن به هدف و مطالعه کیفی آموزش رانندگان در آموزشگاه‌های رانندگی و خصوصیت‌های اجتماعی آنها می‌باشد.

در این تحقیق، پاکسازی داده‌ها به روش‌های زیر صورت پذیرفته است:

۱- اصلاح اشتباهات کاربر

نکته مهم در این خصوص این است که تشخیص اشتباهات کاربران در ورود اطلاعات با مشاهده داده‌های غیرقابل قبول، تطبیق با بقیه داده‌ها و مقایسه داده‌های مشترک در پایگاه داده‌های دیگر صورت می‌پذیرد.

۲- یکپارچه‌سازی

برای یکپارچه‌سازی داده‌های بعضی فیلدها با یکدیگر و همچنین، قابلیت پردازش کردن داده‌ها، لازم بود تا داده‌ها یک شکل گردند. از جمله فیلدهایی که در این مرحله، تغییرات بر روی آنها صورت گرفته است، در جدول ۴ قابل مشاهده می‌باشند.

جدول ۴. تجمیع داده‌ها

فیلد	شکل قدیم	شکل جدید
سطح تحصیلات	درجه‌های مختلف تحصیلی (بی سواد، ابتدایی، سیکل، دیپلم، فوق دیپلم، لیسانس، حوزه و...)	۴ گروه تحصیلی (بی سواد، ابتدایی تا دیپلم، فوق دیپلم تا کارشناسی، فوق لیسانس و بالاتر)
محاسبه سن	تاریخ دریافت گواهینامه - تاریخ تولد فرد و تبدیل آن به بازه سالانه	نمایش سن به صورت سال
آموزشگاه	نام آموزشگاه	کد دهی به هر آموزشگاه با توجه به نام و محل
نام شهر	نام شهر	کد بندی شهر

۳- تبدیل و یکپارچه‌سازی داده‌ها

امروزه داده‌ها به‌عنوان گنجی پنهان در بانک‌های اطلاعاتی قسمت‌های مختلف سازمان نگهداری می‌شوند. فرآیند داده کاوی اغلب نیازمند یکپارچه‌کردن داده‌های خام موجود در منابع مختلف می‌باشد. یکی از کارهای رایج در اجرای فرآیند داده کاوی، تبدیل داده به شکلی مناسب است. در تبدیل داده، داده‌ها به شکل دیگری که برای انجام داده کاوی مناسب است،

تبدیل می‌شوند و در یکپارچه‌سازی، داده‌های به‌دست آمده از منابع مختلف با یکدیگر ترکیب می‌شوند. تطابق موضوع‌ها، حذف زواید و تشخیص مقادیر متضاد در داده‌ها، از جمله مواردی هستند که در هنگام یکپارچه‌سازی داده‌ها باید به آنها توجه شود.

از آنجا که داده‌های فوق، 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 تبدیلات، می‌توان به موارد زیر اشاره کرد:

- یکپارچه‌سازی اطلاعات و مشخصات فردی متخلف با اطلاعات فردی ثبت‌شده در تخلف؛
- یکپارچه‌سازی اطلاعات و مشخصات گواهینامه متخلف با اطلاعات گواهینامه تخلف؛
- محاسبه سن افراد در زمان گرفتن گواهینامه رانندگی با توجه به تاریخ تولد و زمان صدور گواهینامه.

۴- کاهش داده‌ها

فنون کاهش داده برای دستیابی به یک کاهش در داده‌ها به کار می‌روند؛ به صورتی که داده‌های جدید علاوه بر کم حجم‌تر بودن نسبت به داده‌های اولیه، به همان میزان داده‌های اولیه، یکپارچه باشند. داده‌کاوی روی داده‌های کاهش‌یافته دارای کارایی بالاتری است. از نظر تئوری، داشتن تعداد زیادی از مشخصه‌ها باید قدرت تمایزدهندگی بیشتری ایجاد کند؛ اگرچه داده‌های واقعی این قانون را نقض می‌کنند. پاکسازی و پیش‌پردازش داده‌ها در مرحله قبل، مواردی از دریافت‌کنندگان گواهینامه را با اطلاعات ناکافی ظاهر می‌کند که در نهایت، باید از مجموعه داده‌ها حذف شوند. مرحله بعد، گروه‌بندی اطلاعات بر اساس شماره گواهینامه آنها می‌باشد تا کار کردن با داده‌ها از طریق داشتن یک فیلد کلیدی آسان‌تر گردد. لازم به یادآوری است که در گواهینامه‌های صادره، متناظر با یک راننده، ممکن است چندین رکورد تخلف وجود داشته باشد. از جمله اقدامات برای کاهش داده‌ها، می‌توان به ساخت مشخصه‌های ترکیبی اشاره کرد که در آن، اطلاعات موجود در جداول پردازش می‌گردد و متغیر جدیدی با نام جدید ایجاد می‌گردد؛ مثل سن افراد.

۵- انتخاب مشخصه‌ها

بعد از آماده‌سازی داده‌ها باید انتخاب مشخصه‌ها انجام شود. با توجه به فیلدهای موجود در بانک اطلاعاتی و ارزیابی‌های انجام شده، مشخصه‌های دارندگان گواهینامه، آموزشگاه‌ها و تخلفات موردنظر، انتظار می‌رود که با انتخاب این مشخصه‌ها رانندگان بر اساس سوابق

موجود و دیگر موارد مطرح شده، در دسته‌های قابل استنادی قرار گیرند. مانند هر الگوسازی دیگری، انتخاب مشخصه‌هایی که بتوانند قوانین قابل فهم‌تر و ملموس‌تری را ارائه داده و ارتباط معناداری بین مشخصه‌ها ایجاد کنند، از اهمیت ویژه‌ای برخوردار است. اطلاعات مورد نیاز رانندگان برای تشخیص کیفیت آموزش و روابط میان مشخصه‌های فردی راننده را می‌توان از سابقه و گذشته آنها استخراج کرد. این اطلاعات به شکل داده‌های خام در پایگاه‌های داده ذخیره شده است؛ بنابراین، داده کاوی بخش مهمی از فرایند کشف قوانین محسوب می‌شود. به علت اینکه یکی از مهم‌ترین عوامل کلیدی موفقیت در به دست آوردن یک داده کاوی صحیح و جامع، انتخاب متغیرهای صحیح می‌باشد، در مرور ادبیات، این موضوع بیشتر مورد توجه قرار گرفته است (ودل و کاماکورا، ۱، ۲۰۰۰).

۶- فیلدهای پایگاه داده

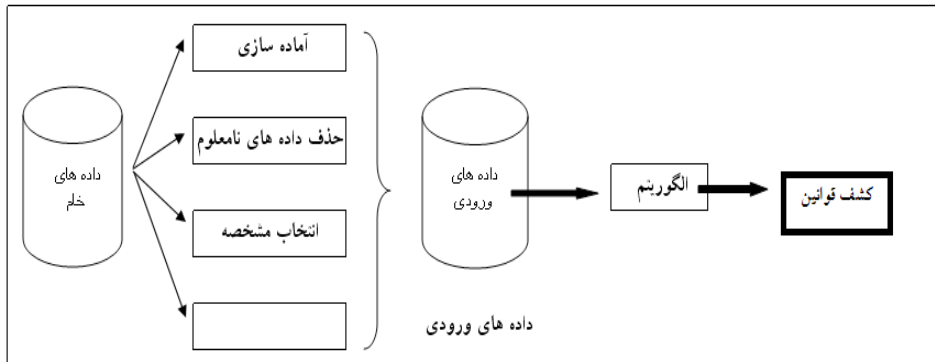
ویژگی‌های کیفیت آموزشگاه‌های رانندگی و تخلف رانندگان و روابط میان مشخصه‌های فردی راننده، همان متغیرهای مستقل تحقیق و استخراج روابط پنهان از پایگاه داده‌های مرتبط با همان متغیرها می‌باشد. برخی فیلدهای پایگاه داده تشکیل شده و جدول اصلی آن، بعد از مرحله پیش‌پردازش حذف شده است، برخی از فیلدها اضافه شده است و برخی فیلدها تغییر یافته است و پس از انجام آنچه در بالا عنوان شد، فیلدهای مورد نیاز در این تحقیق مورد استفاده قرار گرفته است.

۷- تحلیل داده‌ها و کشف قوانین

با داشتن مجموعه داده‌های مناسب می‌توان مرحله سوم را شروع کرد. هدف از این مرحله، کشف کیفیت آموزش و روابط میان مشخصه‌های فردی راننده، با استفاده از داده‌های موجود می‌باشد و چارچوب کلی آن شامل سه مرحله زیر است:

الف. بخش داده‌های ورودی

در این قسمت، پیش‌پردازش داده‌ها انجام می‌شود که به افزایش کیفیت داده‌های ورودی کمک می‌کند و مشخصه‌هایی از مجموعه داده‌های اصلی که باید انتخاب شوند، مورد استفاده قرار می‌گیرند. پیش‌پردازش‌های دیگر و بیشتری نیز ممکن است به این بخش اضافه شوند (شکل ۱).



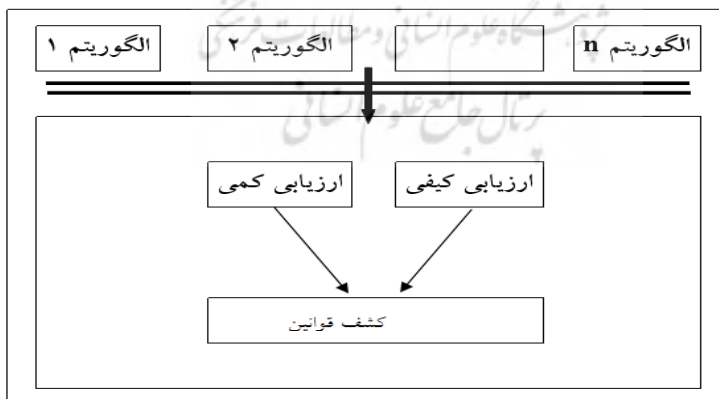
شکل ۱. بخش داده‌های ورودی

ب. اجرای الگوریتم

در این بخش با استفاده از الگوریتم، بهترین قوانین را به دست می‌آوریم و سپس نتایج حاصل از داده‌های ورودی را تجزیه و تحلیل می‌کنیم.

ج. به‌کارگیری الگو و اعتبار سنجی آن

الگوریتم در این فرآیند ارزیابی می‌شود. معیارها و روش ارزیابی الگوریتم، شناسایی کیفیت آموزش و روابط میان مشخصه‌های فردی راننده در این زیرفرآیند به کار می‌روند. هدف ارزیابی، کسب بهترین قوانین است که این فرآیند ممکن است در قسمت داده‌ها و هم در قسمت الگوریتم‌ها باشد (شکل ۲).



شکل ۲. به‌کارگیری الگو و اعتبار سنجی آن

۸- کشف قوانین انجمنی

هدف کلی، یافتن یک مجموعه از قوانین وابستگی یا انجمنی است که بر اساس آن قوانین، بتوان مشخص کرد که وجود کدام یک از مجموعه شاخص‌ها بر وجود شاخص‌های دیگر تأثیر گذار است.

چارچوب عملی برای شناسایی روابط میان مشخصه‌ها

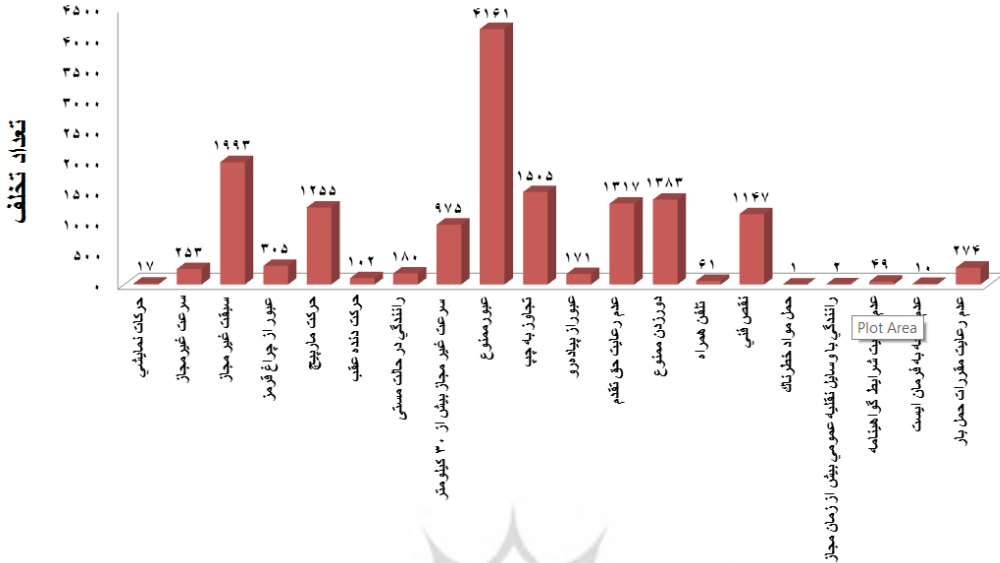
۱. محیط پیاده‌سازی: برای انجام فرآیند داده کاوی و اعمال الگوریتم‌ها، از نرم افزار رایید ماینر استدیو^۱ استفاده شد. این نرم افزار یکی از پرکاربردترین نرم افزارهای داده کاوی است که مورد استفاده قرار می‌گیرد و از بیشترین تعداد الگوریتم‌های داده کاوی نیز حمایت می‌کند.

۲. قالب‌بندی داده‌ها (برای الگو در نرم افزار): از آنجا که برای کشف قوانین نیاز است تا داده‌ها به فرمت لازم قالب‌بندی شوند، بنابراین، داده‌های موجود را برای ورود به فرمت قابل‌پذیرش آماده‌سازی می‌نماییم.

۳. آماده‌سازی داده‌ها: مجموعه داده‌های اصلی شامل رکوردهای مربوط به آموزشگاه رانندگی، متقاضیان و دریافت‌کنندگان گواهینامه‌های رانندگی و تخلفات آنها می‌باشد. با انجام کارهایی نظیر حذف اطلاعات مغشوش و برخی کارهای مقدماتی دیگر توسط نرم‌افزار، تعداد رکوردها کاهش می‌یابد؛ همچنین، اطلاعات مربوط به رکوردهایی که مشخصات راننده در آنها قابل شناسایی نبود، از مجموعه داده اصلی حذف گردید سپس داده‌های مربوط به آموزشگاه‌ها و تخلفات آنان استخراج و داده‌های اصلی در پایگاه داده جدید تکمیل گردید و در نهایت، با تعریف ویژگی‌های جدید، هر کدام از راننده‌ها به وسیله تعدادی از متغیرهای جدید توصیف شدند.

در بین مشخصه‌هایی که هر آموزشگاه را توصیف می‌کند، تعدادی از آنها از کیفیت متفاوتی در طول دوره این تحقیق برخوردار می‌باشند. متغیر نوع تخلف، وضعیت رانندگان را از نظر تخلفات انجام شده، نشان می‌دهد.

شکل ۳، نمودار نسبت تخلفات رانندگان در دو سال ۱۳۹۵ و ۱۳۹۶ در کلان شهر تهران و آموزشگاه‌های رانندگی این شهر که در سامانه ثبت شده است را نمایش می‌دهد.



نوع تخلف انجام شده توسط جامعه آماری تحقیق

شکل ۳. تخلفات رانندگان در سال های ۱۳۹۵ و ۱۳۹۶ در کلان شهر تهران

کشف قوانین کیفی الگوریتم

در این تحقیق، برای کشف قوانین از الگوریتم FP-Growth استفاده می شود. در این تحقیق پارامتر

min confidence و max number of item در بازه های متفاوت برای مشخصه های گوناگون ورودی استفاده می شود و با مقداردهی به آنها خروجی ها از قوانین استخراج می گردد که نشان می دهد در وضعیت های مختلف، امکان وقوع چه تخلفاتی با معیار وابستگی وجود دارد؛ بنابراین، از داده های ورودی قوانین زیادی براساس متغیرهای تعریفی تولید می گردد که برای ارزیابی نهایی و کیفی قوانین کشف شده توسط الگوریتم، با استفاده از مصاحبه با متخصصین امر، قوانین مورد راستی آزمایی قرار می گیرد.

ارزیابی الگوریتم‌های کشف قوانین انجمنی

در نرم‌افزار RapidMiner Studio برای الگوریتم‌های قوانین انجمنی روش ارزیابی خاصی وجود ندارد. اگر بخواهیم الگوریتمی را ارزیابی کنیم، فقط کافی است که تعداد قوانین تولیدشده توسط این الگوریتم را نسبت به یک معیار مشخص و نیز یک حد آستانه خاص برای آن معیار مورد ارزیابی قرار دهیم. در این شرایط، هر قدر الگوریتم کاشف قوانین انجمنی تعداد قوانین بیشتری را به دست بیاورد، به طور حتم، عملکرد مطلوب‌تری را از خود بروز داده است.

نتایج به دست آمده از کشف قوانین

نتایج به دست آمده نهایی، با استفاده از اجرای الگوریتم کشف قوانین Growth-Fp در چند حالت و نسبت به معیار هدف متفاوت در زیر آمده است. در نتیجه، کشف قوانین با معیار قبولی $0/5$ به بیش از 70 مورد قانون می‌رسد. قوانینی که از این الگوریتم استخراج می‌گردد را می‌توان برای تعیین کیفیت آموزش و روابط میان مشخصه‌های فردی رانندگان و تخلفات رانندگی استفاده کرد. این قوانین را می‌بایست در اختیار خبرگان قرارداد تا داده‌های با کیفیت از آن استخراج گردد.

برخی از مهم‌ترین نتایج به دست آمده از قوانین استخراج شده، از این قرار است:

۱. آموزشگاه‌های رانندگی باید به مبحث عبور از محل ممنوع با بیشترین تعداد تخلف، توجه بیشتری داشته باشند و در این میان، آموزشگاه رانندگی مرکز صدور گواهینامه با بیشترین تخلف، نشان از کیفیت پایین دریافت گواهینامه از این مکان را دارا می‌باشد که علت آن، عدم برگزاری کلاس‌های آموزشی جهت دریافت گواهینامه از این آموزشگاه می‌باشد.
۲. رانندگان مرد با مدرک دیپلم بیشترین تخلفات را دارند.
۳. بیشترین تعداد تخلف بین رانندگان بر اساس سن آنها، به بازه ۱۹ تا ۲۲ سال تعلق دارد و در این میان، افراد ۲۰ ساله بیشترین تعداد تخلفات را دارند.
۴. بیشترین نوع تخلف رانندگان، عبور از محل ممنوع در سن ۲۰ سالگی و با مدرک دیپلم می‌باشد.
۵. تخلفات عبور ممنوع، سبقت غیرمجاز، تجاوز به چپ، عدم رعایت حق تقدم و حرکت مارپیچ می‌بایست در آموزشگاه‌ها با کیفیت بیشتری آموزش داده شود.
۶. بیشترین تعداد تخلف به دارندگان مدرک دیپلم تعلق دارد که اکثراً ۲۰ سال سن دارند.

اعتبارسنجی قوانین کشف شده

اعتبارسنجی قوانین کشف شده به دو طریق انجام شده است که عبارتند از: ۱. اخذ نظر کارشناسان راهور ناجا در مورد قوانین استخراج شده؛ ۲. انطباق نتایج حاصل از این قوانین با نظریه‌های مربوط به دلایل بروز کج رفتاری‌ها و تخلفات.

الف- قوانین کشف شده در این تحقیق، با درصد بالایی به تأیید کارشناسان راهور ناجا رسیده است. برخی از نمونه نتایج حاصل از قوانین، از این قرار است:

۱. بیشترین تعداد تخلفات حادثه‌ساز مربوط به رانندگان در بازه سنی ۱۸ تا ۲۲ سال است؛ بنابراین، سخت‌گیری‌های بیشتر در آموزش و آزمون رانندگی این افراد باید اعمال شود و برخورد قاطع‌تری با تخلفات آنان به منظور کاهش تخلفات حادثه‌ساز اعمال گردد.

۲. بیشترین افراد دارای تخلف در بین دارندگان گواهینامه را افرادی تشکیل می‌دهند که مدرک تحصیلی دیپلم دارند.

۳. هرچه مدت بیشتری از زمان اخذ گواهینامه رانندگان می‌گذرد، تعداد تخلفات آنان کمتر می‌شود ولی تخلفات حادثه‌ساز آنها همچنان ادامه دارد.

ب- نتایج این تحقیق مبنی بر اینکه بین سن و مدرک تحصیلی رانندگان و تخلفات رانندگی آنان رابطه معنی‌داری وجود دارد، با نظریه‌های فشار و کنترل‌های اجتماعی انطباق دارد؛ زیرا با افزایش سن و مدرک تحصیلی معمولاً فشار کمتری از سوی جامعه بر افراد وارد شده و به دلیل اینکه تجربه بیشتری در مواجهه با کنترل‌های اجتماعی دارند، تمایل کمتری به تخلفات رانندگی نشان می‌دهند. علاوه بر این، نتایج این تحقیق با نظریه فشار در مورد رانندگان مرد با مدرک دیپلم و سن ۲۰ سال انطباق دارد؛ زیرا این گروه به لحاظ مواجهه با فشارهای بیشتر از سوی جامعه به لحاظ سن کم و مدرک تحصیلی پایین، بیشترین تکرار تخلفات در قوانین کشف شده را دارند. همچنین، نظریه کنترل‌های اجتماعی در مورد این گروه صادق است؛ زیرا به لحاظ سنی، چندان با کنترل‌های اجتماعی مواجه نشده‌اند. نظریه انتقال فرهنگی نیز در مورد این گروه صادق می‌باشد.

نتیجه‌گیری

تخلفات رانندگی هر ساله باعث صدمات جانی و مالی زیادی می‌گردد؛ در نتیجه، به منظور پیشگیری از تخلفات رانندگی، شناخت علل بروز تخلفات از اهمیت خاصی برخوردار می‌باشد.

با توجه به حجم بالای اطلاعات تخلفات رانندگی و مشخصات گواهینامه‌های صادره که در بانک‌های اطلاعاتی پلیس راهور وجود دارد، امکان استفاده از روش داده‌کاوی برای تحلیل این اطلاعات و کشف دانش پنهان موجود در این داده‌ها وجود دارد. این تحقیق با هدف تحلیل راهبردی تخلفات رانندگان و ارتباط آن با مشخصات فردی آنان و مراکز صدور گواهینامه با رویکرد کشف قوانین انجمنی انجام شد. برخی از قوانین استخراج شده نشان می‌دهد که بین سن و مدرک تحصیلی رانندگان و تخلفات رانندگی آنان رابطه معنی‌داری وجود دارد؛ به‌نحوی که بیشترین تکرار تخلفات در قوانین کشف‌شده مربوط به رانندگان مرد با مدرک دیپلم و سن ۲۰ سال است. همچنین، بیشترین تکرار تخلف انجام شده، تخلف ورود ممنوع است و بین برخی تخلفات رانندگان و برخی از آموزشگاه‌های رانندگی ارتباط وجود دارد. به‌دیگر سخن، نتیجه ارائه آموزش ضعیف برخی از آموزشگاه‌ها موجب بروز برخی تخلفات از سوی رانندگان شده‌است. تحقیق حاضر، انواع متغیرهای موثر در تخلفات رانندگی را به‌صورت هم‌زمان در نظر گرفت و می‌توان ارائه خدمات برای متقاضیان رانندگی را بر اساس قوانین کشف‌شده، برنامه‌ریزی کرد و در بسیاری از موارد، از جمله رسیدگی و تدوین متون جدید آموزشی، نحوه آموزش، تهیه طرح درس‌ها و برخورد با آموزشگاه‌های فاقد کیفیت و کاهش سهمیه آموزش، به منظور کاهش آمار تخلفات، نتایج این تحقیق را مد نظر قرار داد.

همچنین، به منظور اعتبارسنجی قوانین کشف‌شده، نظر کارشناسان راهور اخذ گردید و اغلب قوانین به تأیید آنان رسید. از سوی دیگر، بررسی‌ها نشان داد که نتایج حاصل از این قوانین، با نظریه‌های مربوط به دلایل بروز کج‌رفتاری‌ها و تخلفات انطباق دارد.

پیش از این، بررسی کیفیت آموزش رانندگی در حوزه پلیس بیشتر به‌صورت تجربی و سنتی انجام شده‌است اما تحقیق حاضر متغیرهای دخیل در این مسئله را به‌صورت هم‌زمان و غیرهم‌زمان مدنظر قرار داده‌است (مشخصه‌های فردی، جغرافیایی، آموزشگاهی) بنابراین، در نوع خود، کار جدیدی محسوب شده و درک خاصی را نسبت به کیفیت آموزش در آموزشگاه‌ها و روابط میان مشخصه‌های فردی راننده و تخلفات رانندگی ارائه می‌نماید. از دستاوردهای این تحقیق می‌توان به موارد زیر اشاره کرد:

الف) بیشترین دریافت‌کنندگان گواهینامه از مراکز صدور گواهینامه که مرتکب تخلف می‌گردند، از نظر جنسیت به‌ترتیب مردها و بعد زن‌ها می‌باشند؛ از نظر تحصیلات به ترتیب دیپلم، راهنمایی، لیسانس، فوق‌دیپلم، ابتدایی، فوق‌لیسانس، بی‌سواد و دکترا می‌باشند و از نظر سن ۲۰ و ۲۱ و ۲۲ و ۱۹ و ۲۳ سال می‌باشند.

ب) بیشترین تکرار تخلفات در قوانین کشف‌شده مربوط به رانندگان مرد با مدرک دیپلم و ۲۰ ساله می‌باشند.

ج) بیشترین تکرار تخلفات بین آموزشگاه‌های مختلف رانندگی مربوط به آقایان دریافت‌کننده گواهینامه از مرکز صدور گواهینامه می‌باشد.

د) بیشترین تکرار تخلف در بین تخلفات و مشخصه‌های فردی راننده از منظر تخلف انجام‌شده مربوط به ورود ممنوع می‌باشد.

بنابراین، این تحقیق می‌تواند یک جهت‌گیری مناسب را در زمینه نحوه آموزش و توجه به کیفیت آموزش و روابط میان مشخصه‌های فردی راننده در راستای متقاضیان گواهینامه با مدرک دیپلم و سن ۱۹ تا ۲۲ سال تعیین نماید؛ چراکه نظر خبرگان بر روی قوانین کشف‌شده و با توجه به الگوریتم استفاده‌شده از منظر کیفیت آموزش و روابط میان مشخصات فردی متقاضیان و تخلف‌نشان می‌دهد که بیشترین تمرکز در بین متخلفین مربوط به دارندگان مدرک دیپلم و سنین بین ۱۹ تا ۲۲ سال می‌باشد. همچنین، مدیران حوزه رانندگی می‌توانند فیلدهایی که حاوی مشخصات رانندگان و تخلفات ارتكابی آنهاست را بر شاخص‌های بیشتری منطبق نمایند و رفتار راننده را از زمان اخذ گواهینامه و اخذ آزمون‌های مربوطه با درج اطلاعات آن دنبال نمایند و این اطلاعات کمک موثری در شناخت بهتر راننده خواهد بود.

با استفاده از نتایج این تحقیق، می‌توان به این مهم دست یافت که ترکیب کدام یک از مشخصه‌ها در امر آموزش در فرآیند فعلی باعث تخلفات بیشتر می‌شود تا قانونگذار بتواند برای این ترکیب‌ها بیشتر از سایرین، دستور بررسی و یا اعمال مجازات را صادر نماید (مانند برخورد با متخلفین در سال اول دریافت گواهینامه)

فهرست منابع

- اسماعیلی، مهدی (۱۳۹۴)، داده کاوی مفاهیم و تکنیک‌ها (ویرایش سوم). تهران: نیاز دانش
- بهرامی، ر (۱۳۷۹)، تخلفات راهنمایی و رانندگی و علل ارتکاب آن در ایران، پایان نامه کارشناسی ارشد، دانشگاه آزاد اسلامی واحد تهران مرکز
- جعفری، ادیس؛ صمدیان، منیره السادات (۱۳۹۱)، «کاربرد داده کاوی در بررسی رفتار رانندگان متخلف در کلان شهرها»: فصلنامه علمی ترویجی مطالعات راهور، سال نهم، شماره ۱۷
- رابرتسون، بیان (۱۳۷۷)، درآمدی بر جامعه، ترجمه حسین بهروان، مشهد: آستان قدس رضوی
- سلیمی، علی؛ داوری، محمد (۱۳۸۵)، جامعه شناسی کج روی، قم: پژوهشگاه حوزه و دانشگاه
- شاه محمدی، غلامرضا؛ عباسی، سعید (۱۳۹۷)، "تعیین عوامل مؤثر در تصادفات درون شهری با استفاده از روش‌های داده کاوی (مورد مطالعه: شهر اصفهان)": پژوهشنامه جغرافیای انتظامی، سال ششم، شماره ۲۱
- شاه محمدی، غلامرضا؛ رجبی، مصطفی (۱۳۹۷)، "تحلیل جرایم اخلاقی فضای سایبر با رویکرد داده کاوی": فصلنامه پژوهشی دانش انتظامی، سال بیستم، شماره ۴
- صدیق سروستانی، رحمت الله (۱۳۸۷)، آسیب شناسی اجتماعی (جامعه شناسی انحرافات)، کتاب چهارم، تهران، سمت
- صنیعی آبادی، محمد؛ محمودی، سینا؛ طاهرپرور، محدثه (۱۳۹۴)، داده کاوی کاربردی (ویرایش دوم)، تهران، نیاز دانش
- کوزر، لوئیس؛ روزنبرگ، برنارد (۱۳۷۸)، نظریه‌های بنیادی جامعه‌شناسی، ترجمه فرهنگ ارشاد، تهران، نشر نی
- گیدنز، آنتونی (۱۳۷۳)، جامعه‌شناسی، ترجمه منوچهر صبوری، چاپ هشتم، تهران، نشر نی

منابع انگلیسی

- Alpaydin E. (2010). Introduction to Machine Learning. London: The MIT Press Cambridge.
- Health Insurance. International Journal of Engineering and Technology Innovation. 2(2). 126-137.
- Hofmann, m . (2013) "Data Mining Use Cases and Business Analytic applications".
- Klinkenberg ,r .(2013). "RapidMiner Data Mining".
- Lin. Kuo-Chung, Yeh. Ching-Long (2012). Use of Data Mining Techniques to Detect

Medical Fraud in

- Power, d, m, w., (2011) "evaluation : from precision , recall and F-measure to ROC"
- Olson, d, l, .(2008)"Advance data minintechiquesg", springer publishing company
- kim , j , .(2009)"Computational Statictices& data analysis"
- Chen, C-H, .(2010) "handbook of pattern recognition and computer vision", imperial college press
- Ayre , L. B. (2006) "Data Mining for Information Professionals"
- Barbara Mento and Brendan Rapple, (2003) "Data mining and data warehousing" , SPEC Kit 274 , Association of Research Libraries, Washington, DC
- Berson, A. S., Stephen; Thearling, Kurt (2000). Building data mining applications for CRM New York, NY [u.a.] : McGraw-Hill.
- Chen Z. , (2001) "Data Mining and Uncertain Reasoning" , A Wiley – Interscience Publication
- Chen, M. C., & Huang, S. H., (2003), Credit scoring and rejected instances reassigning through evolutionary computation techniques. Expert Systems with Applications, 24, pp. 433–441.
- Christoph F. Eick, N. Z., and Zhenghong Zhao , (2004). Supervised Clustering – Algorithms and Benefits, Department of Computer Science, University of Houston.
- Han, J. and Kamber, M. (2006). Data Mining. Concepts and Techniques. an Francisco, U.S.A, Morgan Kaufman Publishers.
- Ho, T. B. (2000). Knowledge Discovery and Data Mining -Techniques and Practice - Data, Mining Methods: Unesco Course.
- Kadambi, R. & RUPASRI R (2002). Analysis of Data Mining Techniques For Customer Segmentation and Predictive Modeling – A Case Study. Bangalore, India, Industrial Engineering & Management Bangalore University.
- Kantardzi m., Benjamin Djulbegovic and Hazem Hamdan (2002). "A data-mining approach to improving Polycythemia Vera diagnosis " Computational Statistics & Data Analysis 43(4): 765-773.
- Lee, S. C., Suh, Y. H., Kim, J. K., and Lee, K. J. (2004). "A cross-national market segmentation of online game industry using SOM." Expert Systems with Applications 27: 559-570.
- Moxon, B., (2000)"Defining Data Mining In DBMS ONLINE, DBMS Data Warehouse

Supplement”, <http://www.dbmsmag.com/9608d53.html>

- Nong Ye (2003). The Hand Book of Data Mining. New Jersey, LAWRENCE ERLBAUM ASSOCIATES.

- Olson David L.,(2002) “Data Mining”, WWW.ait.unl.edu/dolson/datamining.ppt

- Sonderson, R. (2006). COMP527: data mining.

- Tan , P.N. , (2006) “introduction to DATA MINING” , Boston

- Torgo, L. (2003). Data Mining with R:learning by case studies. Porto.

- Wedel, M., and Kamakura (2000). Market segmentation: Conceptual and methodological foundations.

- West, D., (2000), Neural network credit scoring models. Computers and Operations Research, 27, pp. 1131–1152.

- Wiess, S.M., (1998)“Predictive Data Mining: A Practical Guide”, Morgan Kaufmann Publishers

- Yueh-Min Huang, Chun-Min Hung, Hewijin Christine Jiau, (In Press), Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem, Nonlinear Analysis: Real World Applications, (In Press).





پروشکاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی