مهندسی عمران شریف، (تابستان ۱۳۹۴)
دوره‌ی ۲ - ۳۱، شماره‌ی ۲/۱، ص. ۷۵-۸۵، (یادداشت فنی)

Research Note

عوامل تأثیرگذار در تبعیت از تابلوهای راهنمایی و رانندگی: مطالعه‌ی موردی تابلوهای سرعت مجاز

امیرمسعود رحیمی * **(استادیار)**
گروه مهندسی عمران، دانشگاه زنجان
مجتبی کاظمی (کارشناس ارشد)
گروه مهندسی عمران، دانشگاه آزاد اسلامی واحد رودسر و املش

تابلوهای بیشینه‌ی سرعت مجاز، براساس هندسه‌ی مسیر و شرایط محیطی، با هدف کاهش خطر برخورد و تصادف نصب می‌شوند. در این نوشتار تلاش شده است، با تکمیل ۵۲۷ پرسشنامه در مسیر چابکسر ـ رامسر، از میان ۱۶ پارامتر مختلف تأثیرگذار در رانندگی افراد، و توجه آنها نسبت به تابلو بیشینه‌ی سرعت مجاز ۵۰ کیلومتر بر ساعت، مهم‌ترین آنها شناسایی شوند. روش آماری مورد استفاده رگرسیون درختی و طبقه‌بندی (CART) است، که به منزله‌ی نسل جدید روش‌های داده‌کاوی محسوب می‌شوند. از میان ۱۶ متغیر واردشده به فرآیند مدل‌سازی، سه متغیر تحصیلات و درآمد ماهیانه‌ی راننده و نوع وسیله‌ی نقلیه‌ی وی نسبت به سایر پارامترها مهم‌تر بوده‌اند. برای اعتباریابی نتایج مدل با دنیای واقعی، تابلویی نزدیک به محل پرسش‌گری مورد آزمایش قرار گرفت. واقعیت‌های ثبت‌شده توسط دستگاه‌های سرعت‌سنج سازمان راهداری و حمل و نقل جاده‌یی، نشان از توجه نه چندان سطح بالای رانندگان نسبت به تابلوی نصب‌شده در محل عبوری‌شان دارد.

واژگان کلیدی: تابلو ترافیکی، سرعت مجاز، درخت تصمیم‌گیری، مدل رگرسیون.

amrahimi@znu.ac.ir
mojtaba.kazemi.r@gmail.com

۱. مقدمه

اگر بخواهیم تعریف جامعه‌شناختی از ترافیک ارائه کنیم، شاید یکی از تعاریف این باشد که ترافیک مجموع روابط و کنش‌های متقابلی است که خارج از محل‌های سکون میان کنش‌گران اجتماعی وجود دارد. فرهنگ ترافیکی اساساً در قالب آئین‌نامه و قوانین راهنمایی و رانندگی تبلور می‌یابد، که خود حاوی تصاویر و نوشتارهاست. مطالعه‌ی متون و شناخت‌ها، تصاویر و علائم خبری، و تصاویر و علائم اخطار و احتیاط، مهم‌ترین عناصر فرهنگ ترافیکی به دو گونه‌ی یادشده است. درکنار قوانین و مقررات یادشده، رفتارها و کنش‌های کل عناصر ترافیک نیز تکمیل‌کننده‌ی فرهنگ مکتوب هستند. بدین ترتیب، افراد جامعه به دو صورت خواندن (کسب اطلاعات و یا شناخت هنجارهای قانونی راهنمایی و رانندگی از طریق مطالعه‌ی آئین‌نامه‌ها) و نیز مشاهده‌ی رفتارهای مردم با فرهنگ ترافیک آشنا می‌شوند.[۱] عواملی که در رفتار کاربران راه تأثیرگذار است، به این شرح است:

۱. عوامل روان‌شناختی مانند: هوش، توان یادگیری، انگیزه، مهارت، خواسته و نگرش؛

۲. قدرت حواس مانند: شنوایی و بینایی؛

۳. توانایی‌های فیزیکی شامل: سن، زمان درک و واکنش؛

۴. عوامل پزشکی و وابسته به سلامت مانند: تأثیر داروهای مخدر، الکل، بیماری‌ها و معلولیت‌های فیزیکی.

عوامل زیادی در رفتار، انگیزه و مهارت‌های استفاده‌کنندگان مؤثر است و ممکن است در قابلیت‌های آنها مؤثر باشد. گاهی این عوامل به این دلایل ایجاد می‌شوند:

۱. عدم دسترسی به اطلاعات لازم و کافی؛

۲. وقوع حوادث دور از انتظار، نامعمول و غیرقابل پیش‌بینی؛

۳. ممکن است کاربران، داده‌ها را به طور نامناسب انتخاب کنند، تحلیل آنها را کُند انجام دهند و یا از نتایج به درستی استفاده نکنند.[۲]

یکی از وسایل پرکاربرد جهت در اختیارگذاشتن اطلاعات صحیح، راهنمایی به رانندگان و ایجاد تعامل و ارتباط میان انسان و جاده، علائم ترافیکی است. علائم ترافیکی عموماً به‌صورت قائم نصب می‌شوند و به‌صورت تصویری یا نوشتاری پیام‌ها را منتقل می‌کنند. همچنین علائم ترافیکی در گروه‌های مختلف شامل: انتظامی، هشداری یا اخباری (راهنمایی) وجود دارند. علائم ترافیکی زمانی که پیغام به روشنی و ساده انتقال داده می‌شوند، باید تأثیرگذار باشند و فرمان توسط رانندگان قابل درک و احترام باشد و زمان کافی برای پاسخ و عکس‌العمل مناسب برای راننده موجود باشد.[۳]

در بخش دوم این نوشتار، برخی از مطالعات انجام‌شده در داخل و خارج از کشور بررسی شده‌اند که هر یک از آنها با اهداف و روش‌های مختلف و در بخش سوم مدل استفاده‌شده در پژوهش، تشریح شده است. در بخش چهارم، نتایج حاصل از فرآیند مدل‌سازی، مطابق شیوه‌ی وصف‌شده در بخش سوم به‌صورت کامل تشریح

* نویسنده مسئول
تاریخ: دریافت ۱۳۹۱/۱۱/۳، اصلاحیه ۱۳۹۲/۸/۶، پذیرش ۱۳۹۲/۸/۱۹.

شده است. در انتها نیز بحث درباره‌ی نتایج به‌دست‌آمده و تطابق نتایج تئوری با دنیای واقعی به زبان نوشتار ارائه شده است.

۲. پیشینه‌ی پژوهش

دو تن از پژوهشگران ایرانی به ارزیابی نصب و قابلیت بهره‌برداری تابلوهای پیغام متغیر در ایران اقدام کرده‌اند. در پژوهش یادشده ابتدا مشخصات فنی تابلوهای متغیر خبری[1] بیان شده است، سپس فرآیند به‌کارگیری تابلوهای متغیر خبری شامل برداشت اطلاعات مورد نیاز جهت به‌کارگیری تابلوهای پیغام متغیر، تعیین نوع تابلو، تعیین محل نصب تابلو و جزئیات مورد نیاز جهت نصب تابلوهای متغیر خبری پرداخته شده است.[۴]

در پژوهش دیگری در ایران به مطالعه‌ی طرح جانمایی علائم و تجهیزات ایمنی و ارتباط آن با کاهش تصادفات جاده‌یی پرداخته شده است، که نتایج آن برای محورهای زنجان ـ میانه و بهبهان ـ خیرآباد به ترتیب کاهش ۲۶٫۷٪ و ۴۷٪ تصادفات هم‌سنگ در خسارات مالی شده و مقدار متوسط نسبت منافع به هزینه‌ی حاصل از آن مطالعات به ترتیب برابر ۱٫۴۷ و ۲٫۴۵ شده است، و نیز برای دستیابی به نتایج از آزمون کای مربع استفاده شده است.[۵]

همچنین پژوهش دیگری که روی رانندگان ایرانی انجام شده است، نشان می‌دهد که جنسیت رانندگان و میزان تحصیلات بر توجه رانندگان نسبت به علائم متغیر خبری بی‌تأثیر بوده است و رانندگان مرد و زن نیز به یک میزان به این نوع علائم توجه نشان می‌دهند. در این پژوهش از روش‌های تحلیل سلسله مراتبی[2]، فیلترکالمن[3]، تئوری اطلاعات متقابل[4] و معیار همبستگی[5] استفاده شده است.[۶]

درکشور انگلیس پژوهشی با استفاده از مدل CIECAM۹۷ برای شناسایی رنگ و مدل FOSTS برای ویژگی‌های شکل، به بررسی شناسایی علائم ترافیکی از نظر رنگ و ویژگی شکل آنها توسط مدل‌های دیداری انسان اقدام کرده و نتایج آن نشان داده است که نرخ شناسایی علائم بر اثر تغییرات مصنوعی بسیار بالاست، که در شرایط دیداری مختلف، نرخ شناسایی ۹۵٪ برای علائم انگلیسی (تعداد مشاهدات = ۹۸) به‌دست آمده است.[۷]

براساس پژوهشی دیگری که در ۵ کشور بحرین، قطر، امارات، کویت و عمان با استفاده از تکمیل پرسشنامه و به کاربستن روش چندمتغیره انجام شده است، این نتیجه به‌دست آمده است که رانندگان جوان با سطح تجربه‌ی کم، درآمد پایین و سطح تحصیلات کمتر نسبت به رانندگان میان‌سال با تجربه‌ی رانندگی بیشتر، درآمد بالا و مدارک تحصیلی بالاتر، دقت کمتری نسبت به علائم دارند. و نیز رانندگان مجرد و متأهل به یک نسبت علائم را درک می‌کنند. براساس نتایج این پژوهش، رانندگان پنج کشور یادشده، به طور صحیح ۵۰٪ تا ۶۰٪ علائم را درک می‌کنند.[۳]

همچنین در پژوهش دیگری علاوه بر پارامترهای سن، جنسیت و تحصیلات، در ایجاد پی در پی حوادث مرگ‌بار، پارامتر نژاد /قومیت نیز حائز اهمیت بیان شده است. در این پژوهش با استفاده از آمار توصیفی و رگرسیون لجستیک نشان داده شده است که عامل اصلی تصادفات وسایل نقلیه‌ی موتوری (MVCs) در آمریکای لاتین، کوتاهی در پیروی از قوانین و مقررات ترافیکی است. به‌طورکلی، پژوهشگران در این پژوهش دریافته‌اند که تفاوتی مستقیم میان رانندگان آفریقایی ـ آمریکایی، سفید، و اسپانیایی در مورد توجه پیوسته نسبت به تابلو ایست وجود ندارد.[۸]

نتایج پژوهش دیگری نیز نشان داده است که رانندگان جوان (زیر ۲۴ سال) نسبت به افراد مسن درک پایین‌تری از علائم دارند و درک یک شخص میان‌سال (۲۴-۳۵) نسبت به یک شخص مسن بهتر خواهد بود. موضوع حائز اهمیت دیگر آن است که تجربه‌ی سال‌ها رانندگی نمی‌تواند در بهبودبخشیدن برای درک از علائم مؤثر واقع شود.[۹]

نتایج یک پژوهش انجام‌شده درکشور استرالیا نیز که با هدف آزمایش تأثیرات عوامل راننده و ویژگی‌های طراحی علائم در قابلیت درک علائم ترافیکی انجام شده است، نشان داده است که نمره‌ی ادراک از تابلویی به تابلوی دیگر تفاوت مهمی دارد و پیرو مباحث قابلیت درک علائم ترافیکی بر حسب پارامترهای راننده و ویژگی‌های طراحی علائم متفاوت است. در این پژوهش نشان داده است که عامل سن تفاوتی در سطح ادراک راننده در سه گروه سنی بررسی شده ایجاد نمی‌کند، و نیز راننده با سطح تحصیلات بالاتر درک و فهم بهتری نسبت به علائم خواهد داشت. سال‌های اخذ گواهینامه و سال‌های کسب تجربه در رانندگی در سطح درک علائم ترافیکی با یکدیگر همبستگی دارند. به این نکته نیز باید توجه کرد که تجربه‌ی رانندگی حقیقی به تنهایی ابزار مناسبی برای پیش‌بینی سطح درک از علائم ترافیکی نیست و همچنین سطح توان درک علائم در زمان‌های به دور از رانندگی بدون تغییر باقی خواهند ماند. همچنین ترکیب تأثیر تجربه و تکرار رانندگی در درک علائم ترافیکی بسیار مهم نیست و نتایج این موضوع نشان می‌دهد که تکرار رانندگی و تجربه‌ی راننده، سطح کارایی مشابه با هم در درک از علائم ترافیکی به اندازه‌ی دیگر زیرگروه‌های راننده دارد.[۱۰]

در پژوهش دیگری، نمونه‌یی ۱۰۹ تایی از رانندگان دارای گواهینامه پس از تکمیل پرسشنامه، به منظور بررسی تأثیرات عوامل رانندگی و ویژگی‌های طرح تابلو در درک علائم ترافیکی مورد بررسی قرار گرفتند. این پژوهش با هدف موضوعات ویژگی‌های شخصی مهم، دسته‌بندی ویژگی‌های علائم، و امتیاز درک علائم طراحی شده بود. سال‌های رانندگی با گواهینامه و سطح تحصیلات، به‌صورت معناداری پیش‌بینی‌کننده‌ی درک تابلو قلمداد شدند. و برخلاف انتظار، پارامتر سن راننده، سال‌های رانندگی، ساعات رانندگی، زمان‌های گذشته‌ی رانندگی، فراوانی رانندگی و تجربه‌ی رانندگی در محیط ناآشنا تأثیری بر میزان درک نداشته‌اند. همچنین تابلوی آشنا با امتیاز درک برای رانندگان دارای گواهینامه در ارتباط معنی‌دار بوده است، در حالی‌که پیوستگی، سادگی و معناداری برای تابلو از چنین حالتی برخوردار نبوده‌اند. نتایج این پژوهش خط مشی‌یی برای طراحی علائم ترافیکی مأنوس با استفاده از مدل‌ها و روش‌های آماری و غیر آماری دیگر در آینده فراهم کرده است.[۱۱]

نیاز به شناسایی درخصوص درک بین‌المللی علائم ترافیکی پذیرفته‌شده و سطح متداول تحصیلات رانندگان، انگیزه‌ی اصلی برای انجام پژوهشی درکشور ترکیه شده است. روش پایه‌ی این پژوهش در سال ۲۰۰۹، که براساس دستیابی به نمونه‌ی ۱۴۷۸ راننده‌ی درون‌شهری در شهر آنکارا انجام شده است، تمرکز بر تعیین درک نسبت به ۳۰ تابلوی ترافیکی منتخب است، که عموماً مورد استفاده قرار می‌گیرند و برای ایمنی حائز اهمیت هستند. ایده‌ی پژوهش مذکور، پرسش در ابتدای مسیر و قبل از رؤیت تابلو و در انتهای مسیر پرسش مجدد در خصوص توجه به تابلو دیده‌شده است، که در نهایت تضاد و توافق پاسخ‌های ابتدا و انتهای مسیر در خصوص یک تابلو مشخص با یکدیگر سنجیده شده‌اند. درک بالا از ۹ تابلو موردنظر نشان‌دهنده‌ی اعتبار مطالعه است. تابلوهایی که به تازگی تغییر کرده بودند از طریق ارتباط متقابل با افراد، افزایش ریسک در ایمنی ترافیک را ثابت کرده‌اند، که در این خصوص نیاز به اطلاع‌رسانی بیشتر جهت جلوگیری از حرکات پرخاشگرانه در تقابل با آنها وجود دارد.[۱۲]

در مطالعه‌ی دیگری نیز درکشور اسپانیا نحوه‌ی تصمیم‌گیری مردم در حالت‌های اختیاری و بدون اختیار تجزیه و تحلیل و گزارشی از اطلاعات فراهم‌شده به‌وسیله‌ی یک تابلو (الزام‌آور یا منع‌کننده) یا دو تابلو (دو تابلوی الزام‌آور یا دو تابلو منع‌کننده) تنظیم

شده است. تعداد علائم که همان پیچیدگی صحنه‌ی رانندگی را تداعی می‌کند، عاملی مهم برای این مطالعه بوده است. دو آزمایش متفاوت در قالب ارائه‌ی شرایط آزمایشی (تجربی) دو تابلو به کارگرفته شده است: در آزمایش اول، دو تابلو که به‌صورت قرینه‌ی یکدیگر قرار داشتند، نشان داده شدند و در آزمایش دوم، دو تابلو به‌صورت مجزا در صفحه‌یی ترسیم شدند. نتایج مشابهی برای این دو شیوه‌ی آزمایش به‌دست آمده است. زمانی که یک تابلوی الزام‌آور در تقاطع به کار رفته است، پاسخ‌های سریع‌تر برای مانورهای اختیاری نسبت به مانورهای بدون اختیار به‌دست آمده است. زمانی‌که از تابلوی منع‌کننده استفاده شده است، نتایج برعکس شده‌اند. بنابراین، برتری حالت اجباری تابلوی منع‌کننده پس از آزمایش دو تابلوی منع‌کننده و دریافت نتایج مثبت آن تکرار نشده است.[۱۳]

تنی چند از پژوهشگران در زمینه‌ی ایمنی در کشور استرالیا، سه حالت تقاطع هم‌سطح ریل و جاده که به سه شکل، با تابلوی توقف یا چراغ چشمک‌زن قرمز یا چراغ راهنمایی کنترل شده بودند را با استفاده از شبیه‌ساز MUARC و با کمک‌گرفتن از رانندگان شرکت‌کننده در پژوهش مورد ارزیابی قرار دادند. نتایج به‌دست‌آمده نشان داده است که سرعت متوسط وسیله‌ی نقلیه‌ی نزدیک‌شونده به تقاطع هم‌سطح با شدت بیشتری در پاسخ به چراغ‌های چشمک‌زن نسبت به چراغ‌های راهنمایی کاهش یافته است. در حالی‌که سرعت نزدیک‌شدن به تقاطع در شرایط با تابلوی توقف کمتر بوده و تعداد رانندگان غیرموافق (مانند آن دسته از افراد که توقف نکردند) برای این شرایط بیشتر شده است. به‌علاوه، نتایج نشان داده است که در شرایط تجهیز تقاطع هم‌سطح ریلی به چراغ راهنمایی هیچ‌گونه نفع ایمنی بالاتر و پایین‌تر نسبت به حالت چراغ قرمز با وجود خیابان‌های پیشنهادی جهت دستیابی به نتایج بهتر مشاهده نشده است.[۱۴]

همچنین پژوهشگرانی در کشور عربستان به بررسی میزان پرت‌شدن حواس رانندگان توسط اعلام پیغام تابلوهای کنار جاده‌یی پرداخته‌اند، که برای این منظور به دو صورت شبیه‌سازی رانندگی و تکمیل ۱۶۰ برگ پرسشنامه عمل کرده‌اند. بیش از نیمی از افراد پاسخ‌دهنده به پرسشنامه‌ها اظهار داشته‌اند که دست‌کم یک بار به‌دلیل تابلوهای تبلیغاتی کنار جاده سردرگم شده‌اند. همچنین مشخص شده است که ۲۲٪ آن افراد نیز در معرض وضعیتی خطرناک قرارگرفته‌اند. شبیه‌سازی رانندگی نیز مشخص کرده است که انجام دو نوع رانندگی، جابجایی در خط و عبور خطرناک در تقاطعات، در بحرانی‌ترشدن شرایط مسیر با علائم اعلام‌کننده نسبت به رانندگی در دیگر مسیرها بسیار معنادار بوده است.[۱۵]

در پژوهشی دیگر، تأثیرات ویژگی‌های مختلف راننده و پارامترهای طراحی تابلوها در قابلیت درک علائم ترافیکی به دلیل استفاده از روش‌های آماری و تکمیل پرسشنامه مورد ارزیابی قرارگرفته است. در این پژوهش ۲۱ تابلو از ۸۲ تابلو ترافیکی کشور هنگ‌کنگ به‌صورت تصادفی انتخاب و برای نمره‌دهی به پاسخ‌ها از روش آکیورد[۶] استفاده شده است. در این روش، نمره‌ی ۱٫۰ برای تنها پاسخ صحیح، نمره‌ی ۰٫۵ برای پاسخ‌های دوگانه، نمره‌ی ۰٫۲۵ برای پاسخ غیرمعین، و صفر برای پاسخ غیرصحیح یا پاسخ‌های ترکیبی بوده است. متوسط نمره‌ی درک کل علائم و انحراف معیار به ترتیب ۶۹٫۹۷٪ و ۲۷٫۴۲٪ شده است. توجه کمینه به تابلوی پیاده‌روی هم‌سطح با مانع (۶٫۶۵٪) و توجه بیشینه به تابلو جریان ترافیک دوطرفه (۹۹٫۰۸٪) محاسبه شده است. در پژوهش صورت‌گرفته، تجزیه و تحلیل انحراف معیار (ANOVA)[۷] و آزمون کروسکال ـ والیس[۸] برای نشان‌دادن تفاوت سطح توجه میان گروه‌های مختلف استفاده شده است. از نتایج مهم به‌دست‌آمده آن است که سطح کارایی ادراک به سال‌های فعالیت رانندگی و ساعت‌های رانندگی ارتباط نداشته است. عامل زمان گذشته‌ی رانندگی ارتباط مهمی با سطح ادراک ندارد. گروه سنی تکرار رانندگی و تجربه‌ی رانندگی غیرمحلی ارتباط مهمی با سطح ادراک رانندگان دارای گواهینامه‌ی هنگ‌کنگی نداشته است، اما سطح تحصیلات ارتباط مهمی با سطح ادراک داشته است.[۱۶]

در پژوهش‌های انجام‌شده در زمینه‌ی تبعیت از علائم راهنمایی و رانندگی از روش‌ها و مدل‌های مختلف برای ارزیابی حالات رفتاری و دیداری رانندگان استفاده شده است. پارامترهای مورد بررسی توسط پژوهشگران مختلف نیز متفاوت بوده است. در این پژوهش سعی شده است که روشی مورد بررسی قرارگیرد که در این فضا استفاده نشده است.

۳. روش‌شناسی پژوهش

مطالعه‌ی کتابخانه‌یی گام اول هر پژوهش است. با توجه به جامعه‌ی آماری مورد بررسی این پژوهش (رانندگان)، روش پژوهش میدانی و آماری (آمار استنباطی و توصیفی) استفاده شده است، و شیوه‌ی نمونه‌برداری تصادفی برای تکمیل پرسشنامه‌ی طراحی‌شده برای پژوهش اتخاذ شده است. برای تعیین تعداد افراد موردنیاز از فرمول «کوکران» استفاده شده است، که دست‌کم ۳۸۴ نمونه موردنیاز است. به منظور حصول اطمینان بیشتر، ۵۲۷ برگه‌ی پرسشنامه در بین رانندگان در حال حرکت در جاده‌ی میان شهرستان‌های رودسر و رامسر تکمیل شده است. آلفای کرونباخ محاسبه‌شده ۰٫۹۸ بوده است. بنابراین پرسشنامه، قابلیت اعتماد مناسبی داشته است. متغیر پاسخ پژوهش حالت رتبه‌یی دارد که از درخت تصمیم،[۱۷] و روش رگرسیون درختی و طبقه‌بندی (CART) در پژوهش استفاده شده است.

۱.۳. درخت تصمیم‌گیری

درخت‌های تصمیم‌گیری یکی از شیوه‌های داده‌کاوی به‌شمار می‌آیند که در دو دهه‌ی اخیر توسعه‌ی زیادی یافته‌اند. این شیوه‌ها هم برای کشف و استخراج دانش از یک پایگاه داده‌ها و هم برای ایجاد مدل‌های پیش‌بینی قابل استفاده هستند. درخت تصمیم‌گیری یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی است. درخت‌های تصمیم‌گیری قادرند از میان روابط موجود در یک مجموعه‌ی داده‌یی، توصیفات قابل درک برای انسان تولید کنند و قابلیت به‌کاربردن برای وظایف دسته‌بندی و پیش‌بینی را دارند. این ساختار تصمیم‌گیری می‌تواند به شکل روش‌های ریاضی و محاسباتی که به توصیف، دسته‌بندی و عام‌سازی یک مجموعه از داده‌ها کمک می‌کنند، نیز معرفی شوند. نقاط قوت درختان تصمیم‌گیری عبارت‌اند از:

- فهم مدل ایجادشده توسط درخت تصمیم‌گیری آسان است. به عبارتی با اینکه ممکن است الگوریتم‌هایی که درخت را ایجاد می‌کنند چندان ساده نباشند، ولی فهم نتایج آن آسان است.
- درخت تصمیم‌گیری این توانایی را دارد که پیش‌بینی‌های خود را در قالب یک‌سری قوانین ارائه دهد.
- نیاز به محاسبات خیلی پیچیده‌یی برای دسته‌بندی داده‌ها ندارد.
- برای انواع مختلف داده‌ها از قبیل پیوسته و رده‌یی قابل استفاده است.
- درخت تصمیم‌گیری نشان می‌دهد که کدام زمینه یا متغیرها تأثیرات مهمی در پیش‌بینی و دسته‌بندی دارند.[۱۸]

۲.۳. انواع درخت‌های تصمیم‌گیری

هنگامی‌که خروجی یک درخت، مجموعه‌یی گسسته از یک مجموعه‌ی مقادیر ممکن است؛ به آن درخت دسته‌بندی گفته می‌شود (مثلاً مؤنث یا مذکر، برنده یا بازنده). این درخت‌ها نگاشت $x \rightarrow C$ را بازنمایی می‌کنند که در آن C مقادیر گسسته می‌پذیرد. اهداف اصلی درخت‌های تصمیم‌گیری دسته‌بندی‌کننده را می‌توان به این صورت نام برد:

- داده‌های ورودی را تا حد ممکن درست دسته‌بندی کنند.
- دانش یادگیری‌شده از داده‌های آموزشی را به گونه‌یی عام‌سازی کنند که داده‌های دیده‌نشده را با بالاترین دقت ممکن دسته‌بندی کنند.
- در صورت اضافه‌شدن داده‌های آموزشی جدید، بتوان به راحتی درخت تصمیم‌گیری را گسترش داد (دارای خاصیت افزایشی باشند).
- ساختار درخت حاصل در ساده‌ترین شکل ممکن باشد.

هنگامی‌که بتوان خروجی درخت را یک عدد حقیقی در نظرگرفت، آن را درخت همبستگی (رگرسیون) می‌نامند. این درختان اعداد را در گره‌های برگ پیش‌بینی می‌کنند و می‌توانند از مدل رگرسیون خطی یا ثابت یا مدل‌های دیگر استفاده کنند. وظیفه‌ی یادگیری در درختان رگرسیون، شامل پیش‌بینی اعداد حقیقی به جای مقادیر دسته‌یی گسسته است.[۱۸]

مدل‌های درختی که در رده‌بندی و رگرسیون درختی استفاده می‌شوند، در سال ۱۹۶۳ میلادی برای بررسی اثرات متقابل متغیرها در داده‌های علوم اجتماعی پیشنهاد، و جنبه‌های نظری و کاربردی آن در سال ۱۹۸۴ میلادی در رساله‌یی منتشر، بسط و توسعه داده شده‌اند.[۱۹]

به‌طور کلی روش‌های مبتنی بر مدل‌های خطی، فضای متغیرهای کمکی را به ناحیه‌های مجزا تقسیم می‌کنند و داده‌ها را به گروه‌های متناظر تخصیص می‌دهند. این روش‌ها داده‌ها را به طور بازگشتی برای تعیین یا معرفی اثرات متقابل متغیرها و معرفی زیرگروه‌هایی از افراد با مشخصات دموگرافی و علائم مشخص جهت تشخیص‌های بعدی تقسیم می‌کنند. با توجه به نوع مسئله، هدف اساسی در یک مطالعه‌ی مدل‌های رده‌بندی و رگرسیون درختی می‌تواند ایجاد یک رده‌بندی‌کننده‌ی دقیق و یا کشف یک ساختار پیش‌بینی‌کننده برای مسئله‌ی موردنظر باشد. اگر هدف تعیین یک ساختار پیش‌بینی‌کننده باشد، آنگاه درک صحیح متغیرها و اثرات متقابل آنها ضروری است. معمولاً در مسائل مختلف این دو هدف به موازات هم بررسی می‌شوند.[۱۹]

یک درخت معمولاً از ریشه[۹]، شاخه‌ها[۱۰]، گره‌ها[۱۱] (جایی‌که شاخه‌ها منشعب می‌شوند)، و برگ‌ها[۱۲] تشکیل می‌شود. درخت‌های تصمیم به‌صورت مشابه از گره‌ها که با دایره نشان داده می‌شوند، تشکیل شده‌اند و شاخه‌ها که با پاره‌خط‌های اتصال بین گره‌ها نشان داده می‌شوند. درخت تصمیم را به منظور سادگی در رسم، معمولاً از چپ به راست یا از بالا به پایین رسم می‌کنند، به‌طوری‌که ریشه در بالا قرار بگیرد. در این حالت، گره‌ی اول را ریشه می‌نامند.

انتهای یک زنجیره «ریشه ـ شاخه ـ گره... ـ گره» را یک «برگ» می‌نامند. از هر یک از گره‌های داخلی (یعنی هرگره‌یی که برگ نباشد)، دو یا چند شاخه‌ی دیگر می‌تواند منشعب شود. هر گره مربوط به یک خصوصیت معین است و شاخه‌ها، متناظر با بازه‌یی از مقادیر هستند. این بازه‌های مقادیر باید بخش‌های مختلف مجموعه‌ی مقادیر معلوم برای خصوصیت‌ها را بروز دهند.

با استفاده از درخت تصمیم می‌توان برای هر یک از مشاهدات X، مقدار پیش‌بینی‌شده‌ی Y را پیدا کرد. برای این منظور از ریشه‌ی درخت آغاز می‌کنند و خصوصیات مربوط به ریشه را در نظر می‌گیرند و تعیین می‌کنند که مقدار مشاهده‌شده برای خصوصیت معلوم به کدام شاخه تعلق دارد. آنگاه گره‌یی را در نظر خواهند گرفت که شاخه‌ی موردنظر به آن می‌رسد. این کار را برای این گره نیز انجام می‌دهند و به همین صورت ادامه می‌یابد تا به یک برگ دست پیدا کنند. مقدار Y_S منتسب به برگ Sاُم مقدار پیش‌بینی‌شده برای X خواهد بود. بنابراین، درخت تصمیم مدل وابستگی T را برای Y از X به‌صورت $Y = T(X)$ نشان می‌دهد.

درخت‌های تصمیمی که در یک مسئله‌ی آنالیز رگرسیون در نظرگرفته شده‌اند، درخت‌های رگرسیون نامیده می‌شوند. درخت تصمیم باید یک‌پارچه باشد، یعنی در مسیر از ریشه تا یک برگ، نباید بازه‌های تغییر در نظرگرفته‌نشده وجود داشته باشد.[۲۰]

درخت ایجادشده به گونه‌یی است که در ابتدا همه‌ی داده‌ها در گره‌ی ریشه (اولین گره در بالا) وجود دارند. سپس براساس متغیری که می‌تواند بیشترین همگنی را برای هر شاخه ایجاد کند، در ریشه انشعاب ایجاد می‌شود. این عمل آنقدر ادامه می‌یابد تا داده‌های موجود در هر گره، بیشترین همگنی را داشته باشند و به یک دسته‌ی خاص تعلق گیرند. چنین گره‌یی که در انتها قرار می‌گیرد و از آن انشعابی خارج نمی‌شود، گره‌ی نهایی یا برگ نامیده می‌شود.[۱۷]

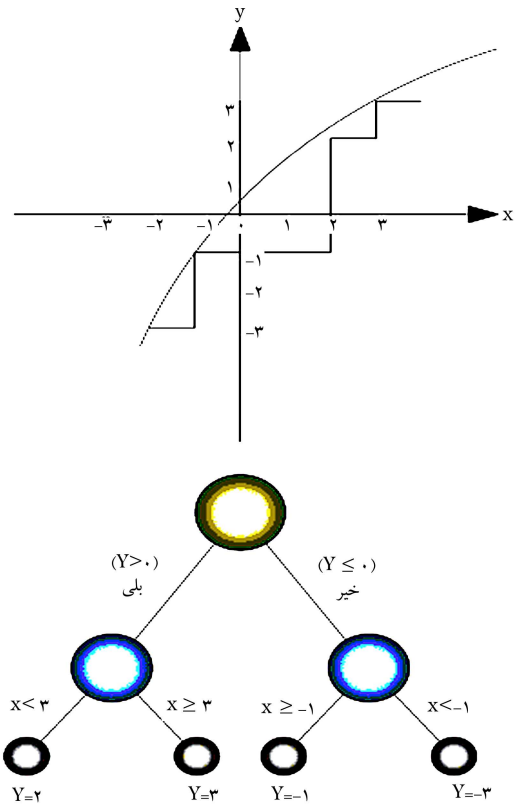
توسط درخت تصمیم می‌توان خصوصیات کمّی و خصوصیات کیفی را به طور هم‌زمان پردازش کرد، بنابراین درخت تصمیم نشان‌دهنده‌ی یک مدل منطقی از پدیده‌ی مورد پژوهش است. نقیصه‌ی درخت‌های تصمیم این است که اگر تمام خصوصیت‌ها کمّی باشند، درخت‌های تصمیم تخمین‌های غیردقیقی از جواب نهایی ارائه می‌دهند. برای مثال درخت رگرسیون شکل ۱، تقریباً تخمین ثابتی از تابع رگرسیون ارائه می‌کند. از طرف دیگر، امکان جبران این کمبود از طریق افزایش تعداد برگ‌ها، یعنی با کاهش طول «پاره خط‌ها» یا «مرحله‌ها» وجود دارد.[۲۱] در شکل ۱، منظور از بله یعنی اینکه تابع برای بخش مثبت صحیح است و در سمت راست، خیر، برای بخش منفی قابل قبول است.

یک درخت تصمیم‌گیری با M برگ در نظرگرفته شود. این درخت تصمیم‌گیری نتیجه‌ی تجزیه‌ی فضای خصوصیات به M زیرمنطقه‌ی غیرهم‌پوشان E^1 تا E^M است، به‌صورتی‌که زیرمنطقه‌ی E^S نشان‌دهنده‌ی برگ Sاُم (شکل ۲) است. در این شکل، منظور از خیر این است که $X_1 > 1$ صحیح نیست، بنابراین شاخه‌ی جدا (سمت چپ) به آن تخصیص می‌یابد، ولی برای شاخه‌ی سمت راست پاسخ مثبت و در نتیجه گسترش آن مجزاست.

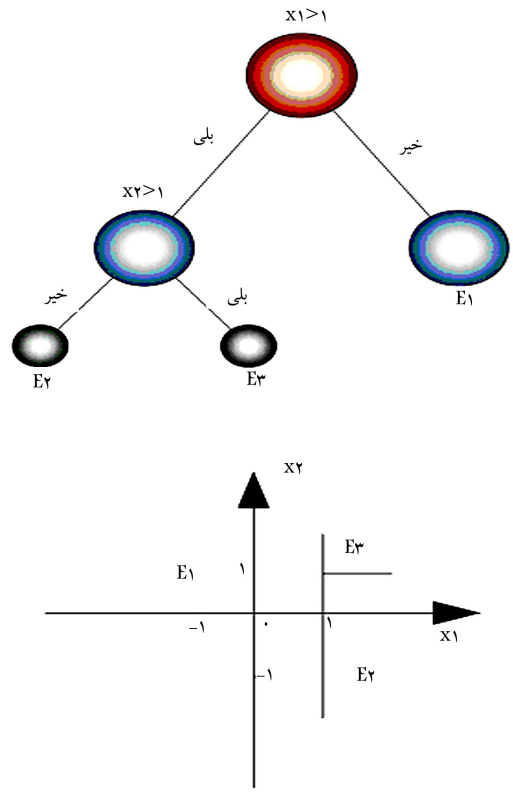
۱.۲.۳. پارامترهای کیفیت درخت

فرض شود یک درخت تصمیم‌گیری و نمونه‌یی از N شی موجود است. امکان انتخاب دو نوع اصلی از پارامترها توضیح‌دهنده‌ی کیفیت یک درخت وجود دارد. نوع اول پارامترهای دقت و نوع دوم پارامترهای پیچیدگی درخت هستند. پارامترهای دقت یک درخت را می‌توان با کمک نمونه‌ی آزمایشی تعریف کرد و کیفیت تقسیم اشیاء در کلاس‌های مختلف (در مورد یک مسئله‌ی تشخیص)، یا اندازه‌ی بزرگی خطا (در مورد یک مسئله‌ی آنالیز رگرسیون) را تعیین کرد.

فرض کنید که مشاهداتی به‌صورت $Data = (x^i, Y^i),\ i = 1...N$ داشته باشیم که هر یک از این مشاهدات (برحسب X) به یکی از زیرمجموعه‌های در نظرگرفته‌شده مربوط باشد. یعنی، $x^i \in E^S$. مجموعه‌ی داده‌های مربوط به E^S را DATA^S می‌نامیم، و تعداد مشاهدات را با N^S نشان می‌دهیم. فرض کنید



شکل ۱. نمونه‌یی از درخت رگرسیونی با تخمین ثابت از رگرسیون.



شکل ۲. درخت تصمیم‌گیری با M برگ.

N_i^S تعداد مشاهدات از DATA^S باشد که به کلاس iاُم تعلق دارد که به مسئله‌ی تشخیص الگو (PRP)[۱۳] شهرت دارد.

عدد نسبی (فراوانی) خطاها به معنای کسری از اشیاء است، که توسط درخت به‌طور اشتباه به یک کلاس نسبت داده شده است (رابطه‌ی ۱):

$$\hat{P}_{err} = {}^{N_{err}}/_{N} \qquad (۱)$$

که در آن N_{err} از رابطه‌ی ۲ به‌دست می‌آید:

$$N_{err} = \sum_{S=۱}^{M} \sum_{i \neq \hat{Y}(S)}^{K} N_i^S \qquad (۲)$$

که در آن $\hat{Y}(S)$ نیز مقدار پیش‌بینی‌شده برای X مربوط به برگ Sاُم و K تعداد کلاس‌هاست. واریانس نسبی برای یک درخت تصمیم را می‌توان از رابطه‌ی ۳ محاسبه کرد:

$$d_{om} = d_{oc}/d_o \qquad (۳)$$

که در آن d_{oc} از رابطه‌ی ۴ به دست می‌آید:

$$d_{oc} = \frac{۱}{N} \sum_{S=۱}^{M} \sum_{i \in Data^S} (\hat{Y}(S) - y^i)^۲ \qquad (۴)$$

که در آن d_{oc} واریانس باقیمانده است. واریانس اولیه از رابطه‌ی ۵ به دست می‌آید:

$$d_o = \frac{۱}{N} \sum_{i=۱}^{N} (y^i - \bar{y})^۲ \qquad (۵)$$

که در آن $\bar{y}$ مقدار میانگین است.

پارامترهای پیچیدگی، خصوصیات شکل درخت را تعیین می‌کنند و به نمونه بستگی ندارند. برای مثال، پارامترهای پیچیدگی یک درخت به‌صورت تعداد برگ‌های درخت، تعداد گره‌های داخلی آن و بیشترین طول مسیر از ریشه تا یک برگ هستند. همچنین می‌توان از طول یک مسیر خارجی استفاده کرد، که به‌صورت تعداد شاخه‌هایی تعریف می‌شود که یک درخت کامل را تشکیل می‌دهند.

پارامترهای پیچیدگی و دقت با هم دارای پیوستگی داخلی هستند؛ به عنوان یک قانون می‌توان گفت؛ درختی که پیچیده‌تر باشد، دقت بیشتری دارد (در درختی که هر برگ آن نماینده‌ی یک شی باشد، بیشترین میزان دقت وجود دارد). «اما اگر دیگر شرایط یکسان باشند»، درختی که پیچیدگی کمتری داشته باشد، ترجیح داده می‌شود. چنین درختی مدل ساده‌تری از پدیده‌ی مورد تحقیق را به‌دست می‌دهد و تفسیرهای بعدی (توضیح مدل) را آسان می‌کند. علاوه بر این، از تحقیقات تئوری چنین بر می‌آید که در صورت کوچک‌بودن اندازه‌ی نمونه (در مقایسه با تعداد خصوصیات) درخت‌هایی که بیش از حد پیچیده باشند، ناپایدار هستند، یعنی دارای تعداد خطاهای بیشتری برای مشاهدات جدید خواهند بود. از طرف دیگر، روشن است که یک درخت خیلی ساده نیز امکان رسیدن به پیش‌بینی خوبی را فراهم نمی‌کند. بنابراین، در انتخاب بهترین درخت تصمیم‌گیری باید به یک «توافق» معین بین پارامترهای دقت و پیچیدگی رسید.

برای رسیدن به چنین توافقی مثلاً می‌توان از شرط $Q = P + \alpha M$ برای کیفیت استفاده کرد. در آن p پارامتر دقت و α پارامتر معلوم هستند. بهترین درخت با توجه به این شرط باید دارای کمترین مقدار Q باشد.

۲.۲.۳. معیارهای انتخاب صفت

معیارهای مختلفی برای تعیین صفتی که جداسازی باید براساس آن انجام شود، وجود دارد، مانند:

- بهره‌ی اطلاعاتی[۱۴]،
- نسبت بهره[۱۵]،
- شاخص جینی[۱۶].

شاخص جینی به‌صورت رابطه‌های ۶ و ۷ تعریف می‌شود:[۱۷]

$$P(j|m) = {}^{P(j,m)}/_{P(m)}$$
$$P(j,m) = \frac{(j)\,N_J(m)}{N_J} \qquad (۶)$$
$$P(m) = \sum_{j=۱}^{J} P(j,m)$$

$$\text{Gini}(m) = ۱ - \sum_{j=۱}^{J} P^۲(j|m) \qquad (۷)$$

که در آن‌ها، J تعداد دسته‌ها یا همان متغیرهای هدف، $P(j)$ احتمال اولیه‌ی مربوط به دسته‌ی j که توسط تصمیم‌گیرنده مشخص می‌شود. $N_j(m)$ تعداد مشاهدات مربوط به دسته‌ی j در گره m، N_j تعداد کل مشاهدات مربوط به دسته‌ی j در گره ریشه، $P(j|m)$ احتمال قرارگیری مشاهدات مربوط به دسته‌ی j در گره m، و $\text{Gini}(m)$ همان شاخص جینی است، که معرف عدم خلوص یا ناهمگنی در گره m است. به این معنی که مثلاً اگر همه‌ی مشاهدات در یک گره از یک دسته باشند، $\text{Gini}(m)$ برابر صفر و مبین کمترین ناخالصی، و به‌عبارت دیگر بیشترین خلوص در گره است، و برعکس، بیشترین مقدار $\text{Gini}(m)$ زمانی حاصل می‌شود که از همه‌ی مشاهدات به یک نسبت در گره وجود داشته باشند. شاخص جینی در هر گره برای تمام متغیرها محاسبه و متغیری به‌عنوان متغیر جداکننده انتخاب می‌شود که کمترین مقدار برای جینی از آن به‌دست آید. احتمال اولیه، مبین سهم هر یک از دسته‌ها در جامعه‌ی مرجع است. رشد درخت براساس شاخص جینی از همان گره ریشه، که اولین گره و در برگیرنده‌ی تمام مشاهدات است، آغاز و برای هر درختی که ایجاد می‌شود، هزینه‌ی دسته‌بندی اشتباه آن ـ که می‌توان از آن به‌عنوان شاخص خوبی برازش یاد کرد ـ طبق رابطه‌ی ۸ محاسبه می‌شود:

$$\text{misclassification...Cost} = \sum_{t=۱}^{T} P(t)\left[۱ - \sum_{j=۱}^{J} P^۲(j|t)\right] \qquad (۸)$$

که در آن، $P(t)$ سهم مشاهدات موجود در گره نهایی t از کل مشاهدات و T تعداد گره‌های نهایی است. رابطه‌ی مذکور نمایانگر آن دسته از داده‌هایی است که به اشتباه در دسته‌های غیرمرتبط با خود دسته‌بندی شده‌اند.[۱۷]

برای ارزیابی درخت ایجادشده توسط روش CART یا هر روش دیگری معیارهایی وجود دارند. از مهم‌ترین و اصلی‌ترین این معیارها نرخ خطا در درخت است. برای محاسبه‌ی نرخ خطا در درخت ابتدا باید نرخ خطا در هر شاخه به‌دست آید. نرخ خطا در هر برگ عبارت است از: نسبت تعداد رکوردهایی از کل که کلاس یا دسته‌ی آنها درست پیش‌بینی نشده است.

برای محاسبه‌ی نرخ خطای کل درخت، مجموع وزنی نرخ خطاهای برگ‌ها به‌دست آورده می‌شود (وزن هر برگ در واقع نسبت جمعیت آن برگ به کل جمعیت رکوردهاست). کیفیت درخت حاصله نیز مهم خواهد بود. به جای جلوگیری از تولید قانون‌های بی‌کیفیت در بعضی از شاخه‌ها قطع (هرس) صورت می‌گیرد. این کار با آنکه نرخ خطا را افزایش می‌دهد، ولی از ایجاد بعضی قانون‌های ناکارا جلوگیری می‌کند. همچنین باید به این نکته توجه داشت که قطع‌کردن باید به نحوی صورت گیرد که خطا از مقدار معینی بیشتر نشود.[۱۸]

۳.۲.۳. تخمینی از کیفیت یک نمونه‌ی کنترل

«نمونه‌ی کنترل (آزمون)» به نمونه‌یی گفته می‌شود که برای ساختن یک درخت به‌کار برده نمی‌شود، بلکه برای تخمین‌زدن کیفیت یک درخت ساخته‌شده به‌کار می‌رود. دو پارامتر محاسبه می‌شود: ۱. پارامتر تعداد نسبی خطاها برای مسائل تشخیص؛ ۲. واریانس نمونه‌ی کنترل برای مسائل آنالیز رگرسیون. از آنجا که این نمونه در ساخت درخت تصمیم نقشی ندارد، این پارامترها خطای نامعلوم «واقعی» را بهتر نشان می‌دهند. هر چه اندازه‌ی نمونه‌ی کنترل بزرگ‌تر باشد، دقت برآورد میزان کیفیت درخت هم بالاتر خواهد بود.

در یک مسئله‌ی تشخیص، تحت شرایط مستقل‌بودن مشاهدات، فراوانی خطاها از توزیع دوجمله‌یی به‌دست می‌آید. بنابراین، با دانستن تعداد خطاها در نمونه‌ی کنترل می‌توان بازه‌ی اطمینانی را پیدا کرد که به احتمال معینی تعداد خطاهای کلاس‌بندی اشتباه به آن بازه تعلق دارد. برای ایجاد درخت تصمیم در این پژوهش از نرم‌افزار SPSS ۱۸، که یکی از قوی‌ترین نرم‌افزارهای داده‌کاوی برای توسعه‌ی مدل درختی با روش رگرسیون درختی و طبقه‌بندی (CART) است، استفاده شده است.[۲۰و۲۱]

۴. فرآیند مدل‌سازی و تحلیل داده‌ها

متغیرهای مستقل برای پژوهش عبارت‌اند از:

۱. سن؛

۲. جنس؛

۳. وضعیت تأهل (مجرد یا متأهل)؛

۴. شغل (کارمند ادارات دولتی، پرسنل نیروهای مسلح، کارمند ادارات خصوصی، معلم یا استاد، محصل یا دانشجو، راننده، خدماتی، خانه‌دار، کارگر، بازنشسته، بیکار، آزاد)؛

۵. میزان تحصیلات (بی‌سواد، زیر دیپلم، دیپلم یا فوق دیپلم، لیسانس، فوق لیسانس و بالاتر)؛

۶. بیشترین زمان رانندگی (اول صبح، صبح تا ظهر، بعد از ظهر، غروب، اول شب، آخر شب، نیمه شب، هر زمانی)؛

۷. سرعت رانندگی (کیلومتر بر ساعت)؛

۸. مدت زمان رانندگی (ساعت)؛

۹. تعداد تصادفات در ۵ سال گذشته؛

۱۰. مدت زمان دریافت گواهینامه (کمتر از ۱ سال، ۱ تا ۵ سال، ۵ تا ۱۰ سال، بیشتر از ۱۰ سال)؛

۱۱. مدت تجربه‌ی رانندگی (کمتر از ۱ سال، ۱ تا ۵ سال، ۵ تا ۱۰ سال، بیشتر از ۱۰ سال)، با فرض بر اینکه مدت زمان دریافت گواهینامه و تجربه‌ی رانندگی برخی از رانندگان برابر نباشد؛

۱۲. نوع گواهینامه (پایه ۱، پایه ۲، موتورسیکلت، فاقد گواهینامه)؛

۱۳. استفاده از عینک به دستور پزشک (بلی، خیر)؛

۱۴. نوع وسیله‌ی نقلیه (سواری شخصی، تاکسی و مسافربَر شخصی، وانت بار و کامیونت، مینی‌بوس، اتوبوس، کامیون، تریلی، سایر وسایل)؛

۱۵. حدود درآمد ماهیانه (کمتر از ۳۰۰ هزار تومان، ۳۰۰ تا ۴۵۰ هزار تومان، ۴۵۰ تا ۶۰۰ هزار تومان، ۶۰۰ تا ۸۰۰ هزار تومان، ۸۰۰ هزار تومان به بالا)؛

۱۶. از بیماری خاصی رنج می‌برید (بله یا خیر)؛

۱۷. متغیر وابسته که همان توجه نسبت به تابلوی ذکرشده است، براساس «طیف لیکرت» در ۵ سطح «بدون توجه، توجه کم، توجه در حد متوسط، توجه زیاد و توجه در حد بسیار زیاد» دسته‌بندی شده است (میزان توجه به علائم نصب‌شده در مسیر).

در این پژوهش، ۹/۴۹٪ رانندگان زن و مابقی مرد بودند. به‌علاوه، سن رانندگان به گروه‌های کمتر از ۱۸ سال (۰/۲٪)، ۱۸ تا ۲۰ سال (۴/۴٪)، ۲۱ تا ۲۹ سال (۳۶/۸٪)، ۳۰ تا ۳۹ سال (۲۶/۶٪)، ۴۰ تا ۴۹ سال (۲۰/۳٪)، ۵۰ تا ۵۹ سال (۹/۳٪)، ۶۰ تا ۶۵ سال (۱/۹٪) و بیش از ۶۵ سال (۰/۶٪) تقسیم شده است.

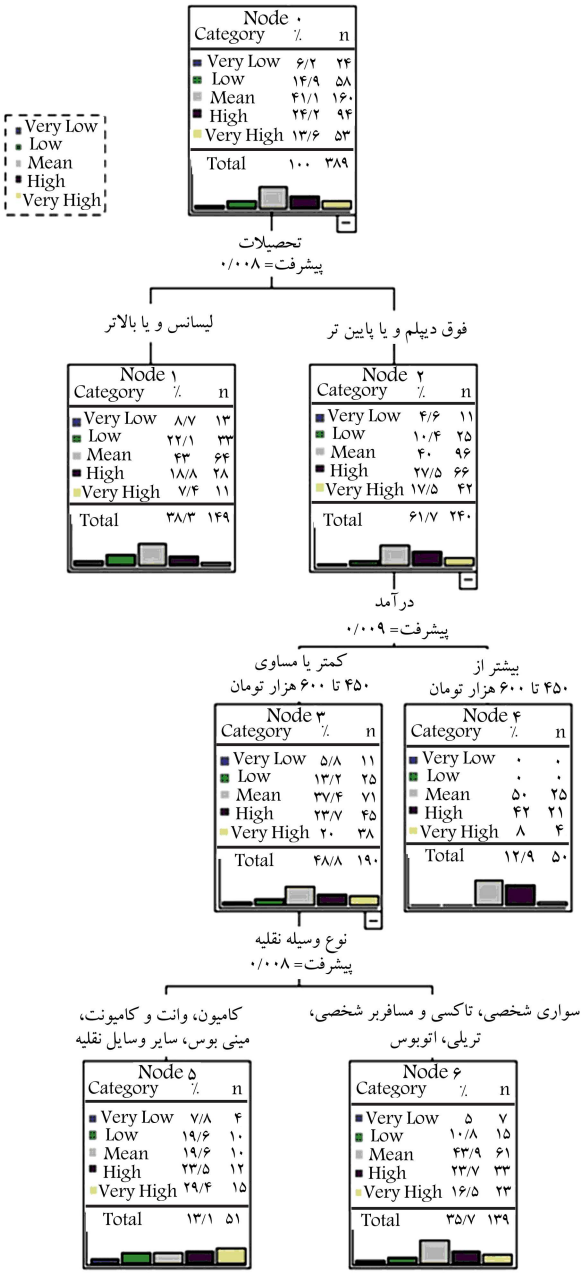
رانندگان با سطح سواد دیپلم یا فوق دیپلم از تعداد بیشتری در جامعه‌ی آماری تهیه‌شده نسبت به سایرگروه‌ها برخوردارند (رانندگان بی‌سواد ۲/۸٪، زیردیپلم ۲۴/۱٪، دیپلم یا فوق دیپلم ۳۴/۳٪، لیسانس ۲۱/۶٪ و فوق لیسانس و بالاتر ۱۷/۱٪).

فرآیند مدل‌سازی با استفاده از نرم افزار SPSS۱۸ انجام شده و در ابتدا، ۷۰٪ داده‌ها به‌عنوان داده‌ی آموزشی و ۳۰٪ به‌عنوان داده‌ی آزمایشی در نظرگرفته شده‌اند، تا نسبت به درستی نتایج به‌دست‌آمده اطمینان حاصل شود. فراوانی بیشینه برای گره والدین و فرزندان در نرم‌افزار، که نشان‌دهنده‌ی حضور جمعیت در هر گره است، به ترتیب، ۵۰ و ۲۵ بود. برای هرس درخت حاصله از «ضریب جینی» استفاده شده است. پس از تصحیح و آماده‌سازی گزینه‌های نرم‌افزاری، فرآیند مدل‌سازی آغاز و در انتها خروجی آن به‌صورت درختی در شکل ۳ برای داده‌های آموزشی، و شکل ۴ برای داده‌های آزمایشی ارائه شده است. احتمالات اولیه برای توجه نسبت به تابلو مورد بحث؛ ۰/۰۶۲، ۰/۱۴۹، ۰/۴۱۱، ۰/۲۴۲ و ۰/۱۳۶ به ترتیب برای سطح‌های خیلی کم، کم، متوسط، زیاد و خیلی زیاد است.

به‌صورت خلاصه، برای ایجاد درخت می‌توان این‌گونه توضیح داد که اولین قدم، پاسخ به این سؤال است که کدامیک از زمینه‌ها، بهترین شاخه را تولید می‌کند. بهترین ایجاد شاخه هنگامی رخ می‌دهد که شاخه‌های حاصل طوری باشند که در هر شاخه، یک کلاس بر سایر کلاس‌ها غلبه کند.

برای محاسبه‌ی گوناگونی در یک مجموعه از رکوردها، روش‌های بسیاری وجود دارد که در تمامی آنها گوناگونی زیاد عبارت است از مجموعه‌هایی که از کلاس‌های گوناگون در خود داشته باشند و گوناگونی کم عبارت است از مجموعه‌هایی که اعضاء یک کلاس در آن بر سایر کلاس‌ها غلبه کنند و بهترین نحوه‌ی ایجاد شاخه آن است که گوناگونی در مجموعه‌ها را تا حد امکان کم کنند.

در مرحله‌ی بعد دو شاخه وجود دارد که هرکدام یک‌سری رکورد دارند (هر یک از رکوردهای گره بالاتر در یکی از شاخه‌ها قرارگرفته است). حال برای هر شاخه، مانند قبل عمل می‌شود. یعنی برای هر یک از آنها دوباره فیلد طوری انتخاب می‌شود که بتوان بهترین شاخه‌های جدید را با گوناگونی کمینه ایجاد کرد. این مرحله آن‌قدر ادامه می‌یابد تا در هر زیرشاخه گره‌یی تولید شود که ایجاد شاخه‌ی جدید در آن گره، مقدار گوناگونی را کاهش قابل توجهی ندهد. به این گره نهایی، برگ گفته می‌شود. این توضیح می‌تواند خلاصه‌یی بر روابط و فرمول‌های ذکرشده در صفحات گذشته باشد، تا خواننده بتواند موضوع را به‌صورت توصیفی تجسم کند.

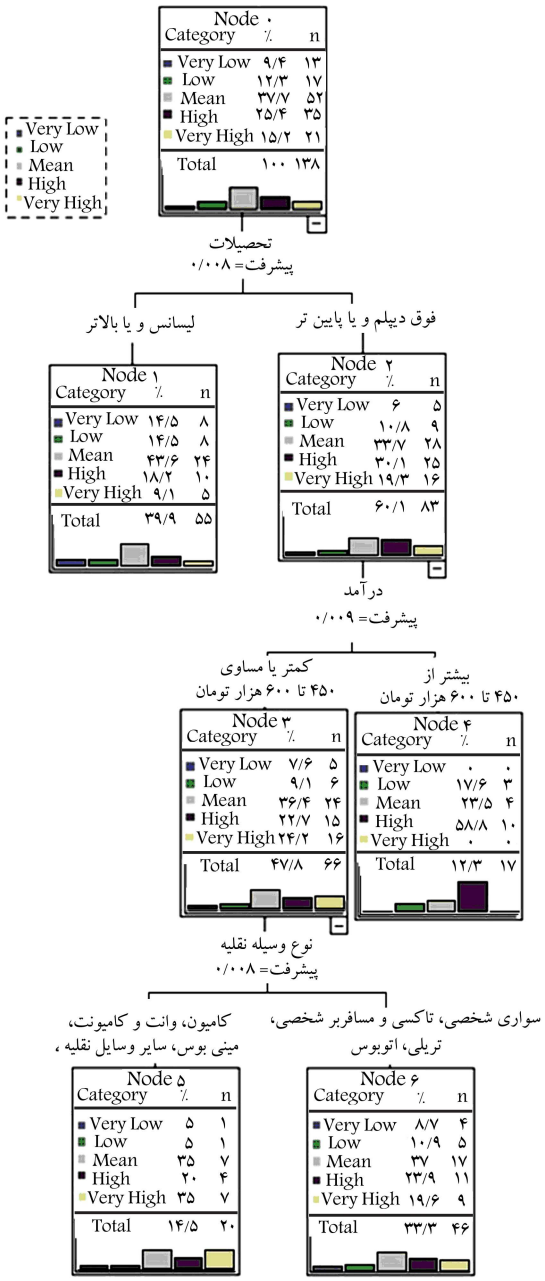


شکل ۳. داده‌های آموزشی.

با نگاهی به نتایج به‌دست‌آمده از مدل‌سازی در شکل‌های ۳ و ۴ می‌توان بیان داشت، میان درخت‌های ایجادشده برای داده‌های آزمایشی و آموزشی، شباهت زیادی از منظر پارامترهای جداساز و دسته‌بندی آنها وجود دارد و این موضوع نشان‌دهنده‌ی صحت و اعتبار مدل است.

۵. بحث درباره‌ی نتایج

پس از مشاهده‌ی درخت‌های ایجادشده، تفسیر نتایج حاصله آن است که میزان توجه رانندگان به تابلو راهنمایی و رانندگی «بیشینه‌ی سرعت مجاز ۵۰ کیلومتر بر ساعت»، در حد «متوسط» است. به گفته‌ی دیگر، این سطح توجه میان سطح توجه «کم» و



شکل ۴. داده‌های آزمایشی.

سطح توجه «زیاد» قرار دارد. اولین پارامتری که بر این میزان توجه تأثیرگذاشته است، سطح تحصیلات رانندگان است. در حقیقت می‌توان با استفاده از نتایج به‌دست‌آمده ادعا کرد که سطح تحصیلات لیسانس و بالاتر، با توجه به نوع وسیله‌ی نقلیه‌ی افراد که متأثر از درآمد ماهیانه‌شان است، باعث تغییر در توجه افراد نسبت به این تابلو خاص می‌شود (به شکل‌های ۳ و ۴ مراجعه شود).

با مشخص‌شدن پارامترهای مهم و هرس‌کردن درخت و حذف پارامترهای کم اهمیت (این پارامترها وجود دارند، اما به دلیل اهمیت کم و جلوگیری از تفسیرهای غلط حذف شده‌اند، که این موضوع معرف هرس در درخت است) در قالب جدول ۱، میزان اهمیت کلیه‌ی متغیرها به‌صورت درصد از سهم کل ۱۰۰٪ متغیر وابسته ارائه شده است. این جدول، تفسیر کاملی از شکل ۳ (که همان داده‌های آموزشی است) است. در حقیقت مقدار تأثیرگذاری متغیرهای مختلف بر سطح توجه رانندگان با صورت نزولی، برای هر گره والدین (گره‌های ۰، ۲ و ۳) نشان داده شده است، که تفسیر ریاضی نمودار درختی را آسان می‌سازد.

به منظور بررسی صحت مدل در دنیای واقعی، آمار سرعت برای نزدیک‌ترین تابلو به محل تکمیل پرسشنامه‌ها، که توسط سازمان راهداری و حمل و نقل جاده‌یی کشور با نصب دستگاه‌های سرعت‌شمار تهیه می‌شود،[۲۲] مورد ارزیابی قرار گرفته است. محل استقرار دستگاه، کیلومتر ۵ در مسیر رامسر به چابکسر است، که در آن منطقه، تابلوی بیشینه‌ی سرعت مجاز ۵۰ کیلومتر بر ساعت نصب شده است.

نتایج آمار ارائه‌شده در جدول ۲ به تفکیک انواع خودروهای عبوری از آن مسیر و بیشینه، کمینه، میانگین، میانه و انحراف از معیار ارائه شده است. پس از نشان‌دادن جدول ۲، شکل ۵ منحنی گوس برای متوسط سرعت دو گروه از وسایل نقلیه‌ی سواری و کامیون سه محور به بالا در یک نمونه‌ی ۴۱۴ برای هر گروه از وسیله‌ی نقلیه‌آمده است.

طبق نتایج ارائه‌شده در جدول ۲، رانندگان خودروهای سواری به تابلوی رؤیت‌شده در مسیر خود که آنها را از تجاوز نسبت به سرعت ۸۰ کیلومتر بر حذر می‌دارد، توجه کامل انجام نمی‌دهند و به طور متوسط نزدیک به آن رانندگی می‌کنند. از آنجایی که مسیر مورد بررسی منطقه‌یی نزدیک به شهر بوده است، می‌توان نتایج آن را قابل تعمیم به منطقه‌ی شهری و با موقعیت تابلوی نصب‌شده‌ی ۵۰ کیلومتر بر ساعت تلقی کرد.

خاطرنشان می‌شود که در آن مسیر، قبل و بعد از محل نصب دستگاه سنجش سرعت، تابلوهای مختلفی با سرعت‌های ۸۰، ۶۰، ۵۰ و ۴۰ کیلومتر نیز نصب شده است، که سرعت را به تدریج کاهش و یا افزایش می‌دهند.

براساس نظرات رانندگان مشخص شد که توجه به تابلو رانندگان در مسیر مورد مطالعه بیشتر براساس سطح تحصیلات، درآمد و به خصوص نوع وسیله‌ی نقلیه تغییر می‌یابد، اما تقریباً توجه نسبی از خود نشان می‌دهند؛ که این نتیجه‌گیری براساس سرعت‌های ثبت‌شده توسط دستگاه‌های سرعت‌سنج نیز ثابت شده است و رانندگان نشان داده‌اند با میانگین سرعت‌هایی نزدیک به سرعت ذکرشده در مسیر حرکت می‌کنند.

۶. نتیجه‌گیری

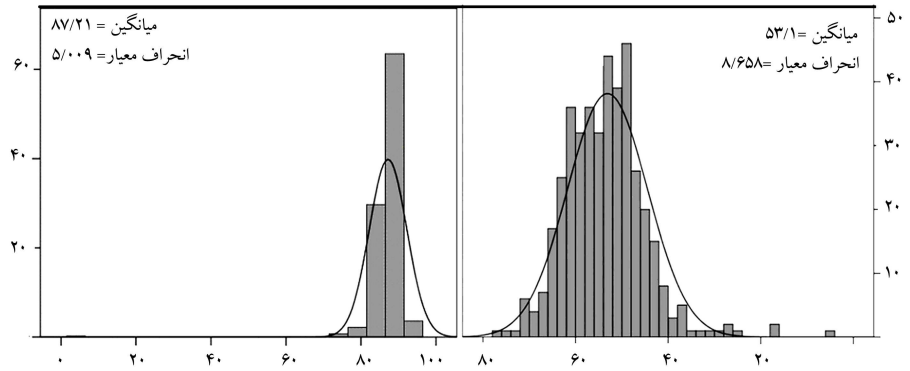
در این پژوهش، ۱۶ متغیر مستقل که هرکدام به نوبه‌ی خود بر میزان توجه نسبت به تابلوی «بیشینه‌ی سرعت مجاز ۵۰ کیومتر بر ساعت» تأثیرگذار هستند، مورد ارزیابی قرارگرفته‌اند. این تابلو بیشتر در مناطق مرتبط با حوزه‌ی شهر، شیب‌راه‌های خروجی یا ورودی، مناطق با عرض خط کم، نواحی کوهستانی، روستایی و غیره مشاهده می‌شود، و این نقاط مهم خود عامل اصلی انجام پژوهش بوده‌اند. پس از انجام فرآیند مدل‌سازی توسط الگوریتم CART برخی از متغیرهای با سطح اهمیت بالا به‌عنوان پارامترهای دارای اولویت استخراج شده‌اند، و از نمایش سایر متغیرها که نقش کم‌رنگ‌تری ایفا کرده‌اند، به منظور جلوگیری از حجیم‌شدن درخت رگرسیونی و طبقه‌بندی و تفسیر منطقی نتایج در شکل نهایی خودداری شده است. پارامترهایی که از مقدار اهمیت بیشتری برخوردار شده‌اند، عبارت‌اند از: تحصیلات، درآمد ماهیانه، و نوع وسیله‌ی نقلیه، که هرکدام به نوبه‌ی خود بر میزان توجه اثرگذارند. دگر بار خاطر نشان می‌شوند که براساس روش‌های پاک‌سازی درخت در مدل CART، متغیرهای با اهمیت کمتر از مدل حذف شده‌اند. این به مفهوم آن نخواهد بود که متغیرهای حذف‌شده کاملاً بی‌تأثیرند، بلکه این متغیرها تأثیر چندان زیادی نسبت به سه متغیر

جدول ۱. متغیرهای جایگزین و میزان اهمیت متغیرهای مستقل نسبت به تابلو «بیشینه‌ی سرعت مجاز ۵۰ کیلومتر بر ساعت».

شماره‌ی گره والدین		متغیر مستقل	پیشرفت (بهبود)	مقدار وابستگی به متغیر پاسخ
۰	اولیه	تحصیلات	۰٫۰۰۸	-
	جایگزین	شغل	۰٫۰۰۱	۰٫۴۶۳
		درآمد ماهیانه	۰٫۰۰۲	۰٫۲۴۸
		زمان رانندگی	۰٫۰۰۳	۰٫۱۰۱
		سرعت رانندگی	۰٫۰۰۵	۰٫۰۹۴
		وضعیت تاهل	۰٫۰۰۳	۰٫۰۷۴
		جنسیت	۰٫۰۰۱	۰٫۰۶۰
		تعداد تصادفات در ۵ سال گذشته	۰٫۰۰۳	۰٫۰۴۷
		تجربه رانندگی به سال	۰٫۰۰۰	۰٫۰۱۳
۲	اولیه	درآمد ماهیانه	۰٫۰۰۹	-
	جایگزین	سن	۰٫۰۰۳	۰٫۰۴۰
		زمان رانندگی	۰٫۰۰۲	۰٫۰۲۰
۳	اولیه	نوع وسیله نقلیه	۰٫۰۰۸	-
	جایگزین	شغل	۰٫۰۰۲	۰٫۱۵۷
		نوع گواهینامه	۰٫۰۰۴	۰٫۰۳۹
		زمان رانندگی	۰٫۰۰۱	۰٫۰۳۹
		سن	۰٫۰۰۳	۰٫۰۲۰
		تعداد سال دریافت گواهینامه	۰٫۰۰۲	۰٫۰۲۰

جدول ۲. تحلیل سرعت‌های عبوری بر اساس نوع خودرو.

عنوان		متوسط سرعت سواری و وانت	متوسط سرعت مینی‌بوس و کامیون دو محور سبک	متوسط سرعت کامیون دو و سه محور	متوسط سرعت اتوبوس	متوسط سرعت کامیون سه محور به بالا
متوسط		۸۷٫۲۱	۶۹٫۱۸	۵۹٫۱۳	۵۴٫۸۴	۵۳٫۱۰
خطای استاندارد از میانگین		۰٫۲۴۶	۰٫۳۵۹	۰٫۳۶۶	۰٫۳۹۴	۰٫۴۲۶
میانه		۸۷	۶۹	۵۹٫۵	۵۵	۵۳
مد		۸۷	۶۶	۵۶	۵۴	۴۹
انحراف استاندارد		۵٫۰۰۹	۷٫۳۱۳	۷٫۴۶۶	۸٫۰۲۵	۸٫۶۵۸
واریانس		۲۵٫۰۹	۵۳٫۴۸۵	۵۵٫۴۴۲	۶۴٫۴۰۴	۷۴٫۹۵۸
کمینه		۴	۴	۴	۵	۵
بیشینه		۹۴	۸۵	۷۶	۷۷	۷۶
مجموع		۳۶۱۰۴	۲۸۶۴۱	۲۴۴۸۱	۲۲۷۰۲	۲۱۹۸۴
درصدها	۲۵	۸۶	۶۶	۵۵	۵۰	۴۸
	۵۰	۸۷	۶۹	۵۹٫۵	۵۵	۵۳
	۷۵	۹۰	۷۳	۶۴	۶۰	۵۹



شکل ۵. منحنی گوس برای متوسط سرعت سواری در سمت چپ و متوسط سرعت کامیون سه محور به بالا در سمت راست.

ذکرشده ندارند و سهم اصلی به سه متغیر «تحصیلات»، «درآمد ماهیانه» و «وسیله‌ی نقلیه» تعلق دارد.

در واقع، می‌توان پیش‌بینی کرد که اگر فردی از میان رانندگان عبوری انتخاب شود و تحصیلات وی، فوق دیپلم و یا پایین‌تر باشد، و درآمد ماهیانه‌ی او کمتر از ۴۵۰ تا ۶۰۰ هزار تومان در ماه باشد و نوع وسیله‌ی مورد استفاده‌اش سواری شخصی و یا تاکسی یا مسافربَر شخصی یا تریلی یا اتوبوس باشد، به تابلوی بیشینه‌ی سرعت مجاز ۵۰ کیلومتر در ساعت، با احتمال ۰٫۱۱، توجه متوسط نشان می‌دهد (شکل ۳).

از آنجا که هدف از انجام این پژوهش، مشخص‌کردن پارامترهای تأثیرگذار با استفاده از روشی جدید در فضای ایمنی بوده است، می‌توان با مشخص‌شدن این پارامترها، به‌صورت ریزبینانه‌تر بر آنها تمرکز کرد تا با استفاده از ابزارها، قوانین موجود، آموزش‌های مرتبط و فرهنگ‌سازی پایه‌یی شاهد افزایش ایمنی، و پیامدهای آن یعنی کاهش حوادث و تصادفات بود.

پانوشت‌ها

1. variable message signs
2. analytical hierarchy process
3. Kalman filter
4. mutual information theory
5. cross correlation
6. Akeyord
7. analysis of variance
8. Wallis-Kruskal
9. root
10. branch
11. node
12. leaf
13. program recognization pattern
14. information gain
15. gain ratio
16. Gini index

منابع (References)

1. Abdol-Rahmani, R., *Sociology of Traffic*, Naja Police Department Education Sciences , Tehran (2008).
2. Shahi, J. and Eghbali Zarach, M., *Geometric Design*, Center for Academic Publishing (2005).
3. Al-Madani, H. and Al-Janahi, A.R. "Role of drivers' personal characteristics in understanding traffic sign symbols", *Accident Analysis & Prevention,* **34**(2), pp. 185-196 (2002).
4. Rahimof, K. and Amjadian, B. "Evaluate the ability of installation and operation of variable message signs on the rural road", Second the traffic and Safety Conference, Tehran (2010).
5. PourAbdol, N. "Studies designed to evaluate the effectiveness of the approach to locating signs and safety equipment to reduce road accidents", Iran University of Science and Technology, Iran (2005).
6. ZanganehPour, S., Lesani, E., Moshiri, B. and Rahimh-Kian, A. "Performance drivers in the face variable message signs and the various factors that influence", The First National Conference on Traffic Safety and Operational Solutions that Improve the of Urban and Rural Road, Bahonar University of Kerman (2011).
7. Gao, X.V., Podladchikova, L., Shaposhnikov, D., Hong, K. and Shevtsova, N. "Recognition of traffic signs based on their colour and shape features extracted using human vision models", *Journal of Visual Communication & Image Representation*, **17**(4), pp. 675-685 (2006).
8. Romano, E., Voas, R. and Tippetts, S. "Stop sign violations: The role of race and ethnicity of fatal crashes", *Journal of Safety Research*, **37**(1), pp. 1-7 (2006).
9. Al-Madani, H. and Al-Janahi, A.R. "Assessment of drivers' comprehension of traffic signs based on their traffic, personal and social characteristics", *Transportation Research Part F*, **5**(1), pp. 63-76 (2002).
10. Dobson. A., Brown, W., Ball, J., Powers, J. and McFadden, M. "Women drivers' behavior, socio-demographic characteristics and accidents", *Accident Analysis and Prevention*, **31**(3), pp. 525-535 (1999).
11. Wang, F. "Study on signs comprehension and driving safety based on drivers' psychology perception", Research Institute of Highway, MOC No. 8 Xitucheng Rd., Haidian District, Beijing, pp. 298-304 (2003).
12. Kirmizioglu, E. and Tuydes-Yaman, H. "Comprehensibility of traffic signs among urban drivers in Turkey", *Accident Analysis and Preventation,* **45**, pp. 131-141 (2012).
13. Castro, C., Moreno-Riso, S., Tornay, F. and Varga, C. "Mental representation of obligatory and prohibitory traffic signs", *Acta Psychologica,* **129**(1), pp. 8-17 (2008).
14. Lenne, M.G., Rudin-Brown, C.M., Navarro, J., Edquist, J., Trotter, M. and Tomasevic, N. "Driver behavior at rail level crossing: Responses to flashing lights, traffic signals and stop signs in simulated rural driving", *Applied Ergonomics*, **42**(4), pp. 548-554 (2011).
15. Sendak, S. and Al-Saleh, K. "The role of roadside advertising signs in distracting drivers", *International Journal of Industrial Ergonomics*, **40**(3), pp. 233-236 (2010).
16. Ng, A.W. and Chan, A.H. "The effects of driver factors and sign design features on the comprehensibility of traffic signs", *Journal of Safety Research,* **39**(3), pp. 321-328 (June 2008).
17. Shariat Mohaimany, A. and Tavakoli Kashani, A. "Severity of injury in two-road accidents using a data mining models", *Transportation Research Bulletin*, **5**(2), pp. 153-165 (2011).
18. Mahjoubi, J. and Etemad Shahidi, E. "Forecasting of wind waves parameters from the port Amirabad by regression decision trees", Fourth National Congress on Civil Engineering , Tehran University (2009).

19. Saki, E., Haji-zadeh, E. and Tehrania, N. "Analysis of risk factors for breast cancer using tree models", *Ofogh-Danesh; Quarterly Medical Sciences and Health Services Gonabad*, **17**(1), pp. 60-68 (2012).

20. Pakgohar, A. "Evaluation of safety (seat belt) and its impact on human health and reduce injuries in car accidents", Master's thesis , Department of Statistics, Islamic Azad University, Mashhad (2007).

21. Pakgohar, A. and Sadeghi-Kia, A. "Statistical data of traffic accidents by analyzing the decision tree", *Traffic Management Studies*, **3**(8), pp.27-45 (2009).

22. Road Maintenance and Transportation Organization (http://www.rmto.ir).