Archive of SID

ارایه مدل پیش بینی شدت تصادفات وسایل نقلیه با استفاده از داده‌های تصادفات (مطالعه موردی محور بابل-گنج افروز)

مقاله پژوهشی

رضوان باباگلی*، استادیار، گروه مهندسی عمران، دانشگاه علم و فناوری مازندران، بهشهر، ایران
علیرضا عاملی، مربی، گروه مهندسی عمران، واحد ملارد، دانشگاه آزاد اسلامی، ملارد، تهران، ایران
علی اصغر غلامرضاتبار، دانش آموخته کارشناسی ارشد، موسسه آموزش عالی علوم و فناوری آریان، امیرکلا، مازندران، ایران
علی پایدار، استادیار، گروه مهندسی عمران، واحد ملارد، دانشگاه آزاد اسلامی، ملارد، تهران، ایران
*پست الکترونیکی نویسنده مسئول: Rezvan_babagoli@yahoo.com

دریافت: ۹۸/۰٤/۱۲- پذیرش: ۹۸/۰۹/۱۵
صفحه ۱٤-۱

چکیده

استان مازندران با توجه به طبیعت زیبای خود همواره یکی از استان های توریستی و گردشگری می باشد که همه ساله تعداد زیادی از افراد در فصول مختلف سال به این استان سفر می‌کنند. همچنین این استان یکی از قطب های کشاورزی کشور می‌باشد که سهم زیادی از حمل و نقل این صنعت، از طریق حمل و نقل جاده ای انجام می‌گیرد که باعث ایجاد تردد ترافیکی با حجم بالا و در پی آن باعث ایجاد حوادث ترافیکی می‌گردد. آمار ارایه شده در سال های اخیر نشان می‌دهد که این استان، هفتمین استان از نظر تلفات ناشی از تصادفات در کشور می باشد. بر این اساس تحقیقات زیادی در مورد شناسایی عوامل موثر بر شدت تصادفات صورت گرفته است و پیشرفت‌های زیادی هم در این مورد حاصل شده است، ولی مطالعات انجام شده در مورد رابطه بین شدت تصادفات و نوع برخورد ناچیز بوده، لذا انجام تحقیقات بیشتر در این مورد ضروری به نظر می‌رسد. در این پژوهش با بکارگیری مدل چند جمله‌ای لوجیت از مجموعه مدل‌های انتخاب برای ارایه مدل پیش بینی شدت تصادفات بهره گرفته شده است. همچنین با استفاده از مدل پیش بینی دو تایی از مجموعه الگوریتم‌های داده کاوی شامل الگوریتم CART به عنوان یکی از الگوریتم های درخت تصمیم و الگوریتمMLP -ANN از مجموعه الگوریتم‌های شبکه های عصبی مصنوعی استفاده گردید و نتایج مورد نیاز استخراج و با یکدیگر مقایسه شده‌اند. براساس مطالعه ها صورت گرفته شده در این پژوهش نشان داده شده است که بهترین مدل از نظر درصد درست پیش بینی و قابلیت ارائه فرمول پیش بینی برای هر سطح، مدل MNL بوده است. نتایج بدست آمده در بخش مدلهای پیش بینی نشان می دهد فرمول برآورد شده قادر به پیش بینی شدت تصادفات در سطوح۰ (صفر) و ۱ با دقت کافی می باشند.

واژه‌های کلیدی: شدت تصادفات، وسایل نقلیه، مدل پیش بینی، لوجیت، شبکه عصبی

۱- مقدمه

طبق تخمین سازمان بهداشت جهانی، تصادفات جاده‌ای در بین بیش از صد عامل شناخته شده مرگ یا معلولیت در مقام نهم قرار دارد (World Health Organization, 2013). پیش بینی شده تا سال۲۰۲۰ این عامل به مقام سوم در مرگ و میر، مقام سوم در معلولیت مادام العمر و مقام دوم در سال های از دست رفته عمر تبدیل شده است. سازمان بهداشت جهانی، تصادفات در ایران را دومین عامل مرگ و میر پس از نارسایی قلبی عنوان کرده است (World Health Organization, 2004). آمارها نشان می‌دهد در دنیا به ازای هر ۱۰ هزار خودرو حدود ۹ نفر کشته می‌شوند، در حالی در ایران به ازای این تعداد خودرو ۳۷ تن کشته می شوند. اگر میانگین ۲۳ تا ۲۷ هزار کشته در حوادث رانندگی در طول سال را همراه با

www.SID.ir

Archive of SID

میانگین ۲۵۰ هزار زخمی این حوادث در نظر بگیریم، گویی هر سال در ایران یک زلزله بزرگ با این مقدار کشته و زخمی اتفاق می‌افتد. همچنین تعداد تلفات جاده ای کشور در سال با تلفات یک جنگ تمام عیار برابری می‌کند. وضعیت نامناسب حوادث جاده‌ای کشور ما باعث شده بانک جهانی در بررسی و مطالعات خود، وضعیت حوادث جاده‌ای ایران را بحرانی عنوان کند. بر اساس اعلام پژوهشکده بیمه مرکزی ایران، کشور ما از نظر تصادفات ناایمن در بین ۱۹۰ کشور جهان، رتبه بندی شده است (Persuad, Dzbik, 1993). سه عامل انسان، راه و وسیله نقلیه به عنوان عوامل موثر در ترافیک و حمل و نقل بصورت جداگانه یا در کنار یکدیگر می توانند از عوامل بروز تصادفات شدید یا خفیف در حوادث رانندگی باشند. از سوی دیگر سلامت و کارایی عوامل فوق، ضامنی معتبر برای حفظ و توسعه ایمنی، سلامت افراد و کاهش هزینه ها در این زمینه است (اداره ایمنی و ترافیک سازمان راهداری و حمل و نقل جاده های کشور، ۱۳۸٤). استان مازندران با توجه به طبیعت زیبای خود همواره یکی از استان های توریستی و گردشگری می باشد که همه ساله تعداد زیادی از افراد در فصول مختلف سال به این استان سفر می‌کنند. همچنین این استان یکی از قطب های کشاورزی کشور می‌باشد که سهم زیادی از حمل و نقل این صنعت، از طریق حمل و نقل جاده ای انجام می گیرد که باعث ایجاد تردد ترافیکی با حجم بالا و در پی آن باعث ایجاد حوادث ترافیکی می گردد. آمار ارایه شده در سال های اخیر نشان می دهد که این استان، هفتمین استان از نظر تلفات ناشی از تصادفات در کشور می باشد. در سال های اخیر تحقیقات زیادی در مورد شناسایی عوامل موثر بر شدت تصادفات صورت گرفته است و پیشرفت های زیادی هم در این مورد حاصل شده است، ولی مطالعات انجام شده در مورد رابطه بین شدت تصادفات و نوع برخورد ناچیز بوده، لذا انجام تحقیقات بیشتر در این مورد ضروری به نظر می رسد.

همچنین بررسی تصادفات جاده‌ای می‌تواند سبب کاهش شدت تصادفات جاده‌ای گردد. امید است با در نظر گرفتن عوامل موثر بر شدت تصادفات و اصلاح این عوامل در جهت کاهش شدت تصادفات به نوعی ایمنی حمل و نقل افزایش یافته و هزینه های ناشی از تصادفات کاهش پیدا کند. ارائه مدل پیش بینی جهت ارزیابی و ارتقا سطح ایمنی جاده همواره یکی از مسائل مورد نیاز به منظور کاهش خسارات جانی و تبعات روانی و اقتصادی می‌باشد.

۲- پیشینه تحقیق

از مطالعاتی که در این زمینه در کشورمان انجام گرفته است می‌توان به پژوهش نصیری و طلوعی اشاره کرد که در آن به بررسی و شناسایی عوامل موثر بر شدت تصادفات جلو به جلو در جاده های دو خطه برون شهری پرداخته شده است. آنان با استفاده از مدل پروبیت ترتیبی عوامل موثر در شدت تصادفات جلو به جلو را شناسایی نمودند. نتایج تحلیل نشان داد که سطح تحصیلات راننده و زمان وقوع تصادف تاثیر زیادی بر شدت تصادفات دارد.تصادفات جلو به جلویی که یکی از وسایل نقلیه درگیر در آن ها از نوع سنگین باشد، نوعاً شدت بیشتری دارد. همچنین تصادفات جلو به جلو در آب و هوای بارانی و برفی شدت کمتری دارند (نصیری و طلوعی، ۱۳۸۰). نصیری و ادریسی با استفاده از مدل لوجیت و شبکه عصبی به شناسایی عوامل موثر بر شدت تصادف کامیون ها با خصوصیات ترافیکی، راننده، وسیله نقلیه، طرح هندسی، نحوه تصادف و شرایط محیطی بررسی شده است. نتایج به دست آمده نشان داد عواملی نظیر خستگی راننده، تصادفات جلو به جلو، عدم توانایی کنترل وسیله، تجاوز از سرعت مجاز در افزایش شدت تصادفات و عواملی نظیر هوای برفی در کاهش شدت تصادفات وسایل نقلیه سنگین در جاده های دوخطه برون شهری موثر می باشند (نصیری و ادریسی، ۱۳۸۲). دنگ و همکاران با به کار گیری یک مدل پروبیت ترتیبی به بررسی شدت تصادفات جلو به جلو در ایالت کنکتیکت پرداختند. بررسی عوامل کوناگون نشان داد سطح خیسی روسازی و وقوع تصادف در شب همبستگی زیادی با شدت تصادف دارد؛ در حالی که افزایش عرض خط باعث کاهش شدت تصادفات فوق می شود(Zuxuan, Ivan and Garner, 2006). کونونن و همکاران با استفاده از مدل لوجیت دوگانه به تحلیل شدت تصادفات بوقوع پیوسته در کشور آمریکا پرداختند. متغیرهای انتخاب شده برای مدل سازی، مواردی چون نوع

www.SID.ir

Archive of SID

برخورد، سرعت، استفاده از کمربند ایمنی و نحوه برخورد مهمترین پارامترهای دخیل در شدت تصادفات می باشند(Kononen, Flannagan and Wang , 2011).

باروا و تی به بررسی تصادفات اتوبوس های شهر داهاکا در کشور بنگلادش طی سال های ۱۹۹۸ تا ۲۰۰۵، پرداخته اند. مدل لوجیت ترتیبی برای تحلیل شدت تصادفات به کار گرفته شده است. عواملی چون وقوع تصادف در روز تعطیل، مسیر دوطرفه، درگیر با یک وسیله و در گیر با عابر از جمله عوامل افزایش دهنده شدت تصادف بوده است (Barua, Tay, 2010). کاپلان و پراتو به بررسی شدت تصادفات اتوبوس ها در سطح آمریکا، طی سال های ۲۰۰۵ تا ۲۰۰۹ پرداخته‌اند. بدین منظور آن ها از مدل لوجیت ترتیبی تعمیم یافته استفاده کرده اند. سطح شدت در پنج سطح بدون حراحت، جراحت احتمالی، جراحت بدون نقص عضو، با نقص عضو و فوت در نظر گرفته شده است (Kaplan, Prato, 2012). نصیری و ادریسی با استفاده از مدل لوجیت و شبکه عصبی به شناسایی عوامل موثر بر شدت تصادفات کامیون‌ها در جاده های دوخطه برون شهری پرداختند. مدل ترتیبی یکی از رایج ترین مدل های مورد استفاده برای بررسی شدت تصادفات است (نصیری، ادریسی، ۱۳۸۲). دنگ و همکاران با به کارگیری یک مدل پروبیت ترتیبی به بررسی شدت تصادفات در ایالت کنکتیکت پرداختند. بررسی عوامل گوناگون نشان داد سطح خیسی روسازی و وقوع تصادفات در شب همبستگی زیادی با شدت تصادفات دارد، در حالی که افزایش عرض خط باعث کاهش شدت تصادفات فوق می‌شود (Zuxuan, Ivan and Garder, 2006). هلیم و ابدل اتی برای تحلیل تصادفات بوقوع پیوسته در چهار راه‌های بدون چراغ از مدل پروبیت دوگانه با دو سطح شدت جرحی و بدون جرح پرداختند (Haleem, Abdel-Aty, 2010). ووگت و بارد ضمن توضیح و ارائه اهمیت مدل های شدت تصادف، اعتبار این مدل ها را عموما بر دقت اطلاعات فراهم شده از بانک اطلاعاتی تصادف دانست. او از مدل های لاجیت دو تایی در بررسی خود استفاده کرده و عوامل نظیر خطای راننده، شرایط بد راننده (خستگی و مریضی)، نبود دید کافی، خیس بودن جاده، تصادف در شب، خرابی وسیله نقلیه، مصرف الکل و ... را افزاینده و عواملی نظیر استفاده از کمربند ایمنی را کاهنده شدت تصادف‌ها دانست. او در ارایه مدل ها نشان داد که در صورتی که در یک دسته (مثلا تصادف های منجر به فوت) تعداد مشاهدات نسبت به کل تصادف کم باشد، ادغام آن با تصادف های جرحی و گرفتن آن ها به عنوان یک دسته، باعث معنی دار شدن متغیرهای بیشتر و نتایج بهتر در مدل‌ها می شود وگت و بارد نیز با تقسیم بندی شدت تصادف‌ها به دو گروه فوتی یا جرحی، جرح کم یا مالی و بکارگیری مدل لاجیت دوتایی، رابطه‌ای برای شدت تصادف‌های جاده‌ای دو خطه برون شهری با عوامل مختلف ارایه کردند (Voget, Bared , 1999). خطاک با استفاده از مدل پرابیت ترتیبی، علاوه بر تاکید بر تاثیر شرایط هوایی و نوع تصادف در شدت تصادف ها، عواملی نظیر عدم دقت در رانندگی، سرعت زیاد، مصرف الکل، سن کم (زیر ۲۵ سال)یا زیاد (بالای ٦٥ سال) راننده، خیس بودن جاده، وجود پیچ در جاده، شیب راه و ... را افزاینده و شرایط برفی، تصادف در روز، جاده صاف، حجم ترافیک بیشتر و ... را کاهنده شدت تصادف معرفی کرد (Khattak, Kantor and Council, 1999). مطالعات مختلفی با استفاده از مدل‌های لوجیت به موضوع تصادفات پرداخته‌اند نصیری و ادریسی با استفاده از مدل لوجیت و شبکه عصبی به شناسایی عوامل موثر بر شدت تصادفات کامیون‌ها در جاده های دو خطه برون شهری پرداختند (نصیری، ادریسی، ۱۳۸۲). راضی و صمیمی عوامل موثر بر شدت تصادفات جلو در تصادفات درون شهری را مورد بررسی قرار دادند. در این راستا، تصادفات شهر تهران با بکارگیری مدل‌های پروبیت لوجیت و شبکه عصبی مصنوعی، مدل سازی و تحلیل شده است. متغیر وابسته، شدت تصادفات (جرحی و خسارتی) و

www.SID.ir

Archive of SID

متغیرهای مستقل، پارامترهایی چون ویژگی‌های راننده، زمان وقوع حادثه، مشخصات محیطی و ترافیکی می باشند. در این راستا مدل‌های لوجیت و پروبیت با تحلیل حساسیت شبکه عصبی مقایسه گردیده است. نتایج نشان می‌دهد که عواملی چون جنسیت سن و تحصیلات راننده در افزایش شدت تصادفات موثر بوده است (راضی اردکانی و صمیمی، ۱۳۹۰). فلانگان و همکاران با استفاده از مدل لوجیت دو گانه به تحلیل شدت تصادفات به وقوع پیوسته در کشور آمریکا پرداختند. متغیر های انتخاب شده برای مدل سازی، مواردی چون نوع برخورد، سرعت، استفاده از کمربند ایمنی و نوع خودرو بوده است. نتایج نشان داد که استفاده از کمربند ایمنی و نحوه برخورد مهم‌ترین پارامترهای دخیل در شدت تصادفات می‌باشند. پژوهشگران در سال ۲۰۱۳ تحقیقی را بر روی اثر سمت راننده و سمت ضربه بر شدت تصادف در طول آزاد راه شهری با استفاده از دو رابطه مدل لوجیت مرکب انجام داده‌اند و نتایج مهمی یافته اند که در ادامه مورد بررسی قرار گرفته شده است. در این مطالعه به شناسایی عواملی همچون شرایط هندسی، ترافیک، زیست محیطی، عوامل مرتبط به وسیله نقلیه، عوامل مرتبط به راننده و پیش بینی از شدت آسیب تصادفات در آزاد راه های شهری پرداخته شده است. در این مطالعه از توانایی های مدل لوجیت مرکب برای محاسبه اثرات مشاهده نشده‌ای که برای تعیین کمیت دشوار است و ممکن است در مدل برآورده، واکنش راننده، در زمان تصادف تاثیر گذار باشد، استفاده شده است. برای این مطالعه از داده های تصادف ۵ سال در ۸۹ آزاد راه شهری در سراسر ایالت فلوریدا در ایالت متحده استفاده شده است. نمونه‌هایی از پیش بینی شدت مورد بررسی عبارت اند از: حجم ترافیک، فاصله ترافیک از نزدیک ترین رمپ، سن دقیق راننده، نوع وسیله نقلیه، طرف‌های ضربه. برای آن که نشان داده شود که چگونه برآورد پارامترها می‌تواند متفاوت باشد، یک مدل لوجیت باینری با مدل لوجیت مرکب مقایسه شده است. از جمله نتایجی که در این مطالعه بدست آمده است می توان به این موارد اشاره کرد که سن راننده مقصر، حجم ترافیک، فاصله تصادف از نزدیک ترین رمپ، نوع وسیله نقلیه، سمت ضربه و درصد کامیون به طور قابل توجه‌ای در شدت تصادفات در آزاد راه های شهری تاثیرگذار بوده‌اند. علاوه بر این نتایج با عامل راننده جوان مقصر نیز در ارتباط می‌باشد که با افزایش قابل توجه خطر شدت تصادف در مقایسه با دیگر گروه های سنی، مشاهده شده است که برخی از متغیرها در مدل لوجیت باینری به دلیل نادیده گرفتن تغییرات تصادفی از برآورد، سبب شده است که برآورد غیر منطقی به همراه داشته باشد. از آنجا که سن راننده مقصر و سمت برخورد قابل توجه بودند، بررسی‌های عمقی بر روی پارامترهای تصادفی در مدل لوجیت مرکب انجام شد؛ و از آن متوجه شدند که ضربه‌هایی که از عقب، چپ و راست وارد شده اند. بیشترین خطر را به ترتیب در میان رانندگان میان سال، رانندگان جوان، رانندگان بسیار جوان و در نهایت رانندگان پیر و بسیار قدیمی داشته‌اند. در نهایت، در این مطالعه یک روش امیدبخش به غربالگری پیش بینی قبل از برازش مدل لوجیت مرکب با استفاده از روش جنگل‌های تصادفی ارایه شده است. علاوه بر این، اقدامات متقابل بالقوه برای کاهش شدت اثرات جانبی به علت تغییر خط، مانند ابداع سیستم اجتناب از تصادف جانبی ارائه شده است (Haleem et al., 2013).

پژوهشگران در سال ۲۰۱٤ تحقیقی را بر روی آنالیز عوامل خطر موثر بر شدت آسیب ترافیک جاده ای در ترکیه با استفاده از مدل چند جمله‌ای لوجیت انجام داده‌اند و نتایج مهمی یافته‌اند که در ادامه مورد بررسی قرار گرفته شده است. در این مطالعه به تجزیه و تحلیل ۷۷۱/۱۱ تصادفات ترافیکی که بر ۲ شهر ترکیه در بین ماه های ژانویه ۲۰۰۸ تا دسامبر ۲۰۱۳ توسط پلیس گزارش شده، پرداخته شده است؛ که در این تصادفات به ۳ دسته کشنده، آسیب، بدون آسیب طبقه بندی شده‌اند. بر اساس این طبقه بندی آنالیز چند جمله‌ای لوجیت برای تعیین عوامل خطر موثر بر شدت صدمات ترافیک انجام شده است.

www.SID.ir

Archive of SID

نتایج نشان داده است که عوامل زیر احتمال افزایش صدمات کشنده را در بر داشته است. رانندگان با بیش از ٦٥ سال سن، رانندگان تازه آموزش دیده، تصادفات تک وسیله نقلیه، حوادث رخ داده در مسیر مکان های مورد نظر، راه ها یا جاده های استانی و حضور عابر پیاده، همچنین نشان داده است که تصادفات اتومبیل و یا وسیله نقلیه خصوصی و با آن دسته که در طول اوج شب اتفاق می افتند. در شرایط آب و هوایی روشن، در خیابان های محدوده شهری و یا در حضور چراغ‌های ترافیکی سبب کاهش احتمال صدمات کشنده شده اند. این مطالعه شامل یک پایگاه جامع داده برای یک نمونه از ترکیه ایجاد شده است. این مطالعه همچنین اولین تلاش برای استفاده از مدل پیش بینی نامرتبط برای تعیین عوامل خطر موثر بر شدت جراحات ناشی از رانندگی در ترکیه بوده است (Celik et al., 2014).

۳- ساختار مدل‌های پیش بینی بکار رفته

۳-۱- مدل لوجیت دوگانه

صورت کلی مدل لوجیت چندگانه به صورت زیر است:

$$\Pr_n(i) = \frac{e^{V_{in}}}{\sum_{j \in c_n} e^{V_{jn}}} \qquad (۱)$$

$\Pr_n(i)$: احتمال وقوع حالت i

V_{1n}: بخش قابل اندازه گیری مطلوبیت حالت i

V_{jn}: بخش قابل اندازه گیری مطلوبیت حالت j

C_n: مجموعه گزینه‌های موجود است.

جدی‌ترین فرض در چارچوب چندجمله‌ای لوجیت فرض استقلال از جایگزین‌های نامرتبط (IIA) است. در این فرض به این صورت است که نسبت شانس بین هر دو نتیجه، مستقل از تعداد و ماهیت است و سایر نتایج به طور همزمان در نظر گرفته شده‌اند. فرض (IIA) همچنین به عنوان استقلال باینری یا اصل استقلال شناخته شده است که یک اصل تئوری تصمیم گیری در علوم مختلف اجتماعی است. این اصطلاح با معانی مختلف در زمینه‌های مختلف استفاده می‌شود و آنها تمام تلاش خود را برای ارایه یک برآورد منطقی برای رفتار فرد و یا تجمیع ترجیحات فردی دارند. روشن‌ترین مورد نقض این ویژگی زمانی است که نتایج خاصی به عنوان جایگزین برای دیگران مورد استفاده قرار می‌گیرد.

چندجمله‌ای لوجیت (MNL) دارای خاصیتی است که نسبت احتمال انتخاب هر دو گزینه مستقل از در دسترس بودن گزینه سوم است. مدل چندجمله‌ای لوجیت فرض می‌شود که داده‌های مورد، خاص است، به عبارت دیگر هر متغیر مستقل یک ارزش واحد برای هر مورد دارد. فرض دیگر این است که متغیر وابسته را نمی‌توان کاملاً از متغیرهای مستقل برای هر مورد پیش‌بینی کرد، مانند انواع دیگری از رگرسیون، این مدل نیاز ندارد که متغیرها از لحاظ آماری از یکدیگر مستقل باشند (برای مثال، برخلاف یک طبقه بندی ساده بیز)، با این حال نسبتاً کم فرض می‌شود که تمایز بین تأثیر چند متغیر دشوار می‌شود، در صورتی که این چنین نیست. مدل MNL با این فرض مشتق شده است از عوامل مشاهده نشده ناهمبسته که بیش از جایگزین‌ها (یا نتایج) هستند و واریانس یکسان برای همه جایگزین‌ها دارد، همچنین مدل با عنوان فرض استقلال از جایگزین‌های نامرتبط (IIA) شناخته شده است. این فرض قابل توجه‌ترین محدودیت مدل MNL است. به عنوان مثال سطح هوشیاری ممکن است یک متغیری باشد هنگام مدل سازی شدت تصادف در نظر گرفته نشود و در حالی که می‌توان آن را به عنوان یک عامل مشاهده نشده در نظر گرفت. با این حال، راننده خواب آلود که در یک تصادف مرگبار دخالت دارد، ممکن است با احتمال مشابه در یک تصادف آسیب ناتوان‌کننده دخالت داشته باشد، اگر چنین است، عوامل مشاهده نشده (سطح هوشیاری) مؤثر بر تصادف مرگبار و آسیب ناتوان کننده هستند و این عوامل می‌توانند در ارتباط با عوامل مستقل باشند. با وجود این، محدودیت فرض IIA باعث می‌شود مدل MNL بسیار مناسب و محبوب برای استفاده در مدلسازی ها باشد (Moore et al, 2011).

۳-۲- ساختار مدل‌های پروبیت

مدل‌های چند جمله‌ای پروبیت همانند مدل‌های چند جمله‌ای لوجیت هستند با این تفاوت که خطای استاندارد باید از توزیع نرمال پیروی کنند و همچنین تابع برآورد آن نیز با مدل‌های لوجیت متفاوت است. صورت کلی این مدل برای دو جمله‌ای به صورت ۲ است: فرمول تعیین مقادیر در هر سطح:

www.SID.ir

Archive of SID

$$\Phi^{-1}(p) = z - \beta_0 + \sum_{i=1}^{k} \beta_i x_i \qquad (۲)$$

$\Phi^{-1}(p)$: معکوس تابع چگالی توزیع نرمال

فرمول تعیین مقادیر فراورد در هر سطح:

$$\int_{-\infty}^{z} \frac{1}{\sqrt{2\Pi}} * e^{\frac{-x^2}{2}} dx \qquad (۳)$$

مدل پروبیت چند جمله‌ای یک جایگزین برای مدل لوجیت مرکب است. با این حال برخی از نتایج مطالعات گذشته نشان می‌دهد که مدل لوجیت مرکب ممکن است ترجیح داده شود در شرایطی که هدف این است تخمینی برای ضریب توزیع پارامترها وجود داشته باشد. مدل چند جمله‌ای پروبیت تعمیمی از مدل پروبیت است و زمانی استفاده می‌شود که ممکن است چندین دسته متغیر وابسته به وجود داشته باشد. به این ترتیب یک جایگزین برای مدل چند جمله‌ای به عنوان یکی از روش طبقه بندی چند کلاسه است. این مدل نباید با مدل چند متغیره پروبیت اشتباه گرفته شود، مدل چند متغیره پروبیت برای مدلی با نتایج باینری همبسته استفاده می‌شود که بیش از یک متغیر مستقل وجود دارد(Ye, Lord, 2011). تفاوت اصلی بین مدل OP و دو مدل لوجیت (ML , MNL) در شرایط خطا و پارامترها می‌باشد. برای مدل OP، خطای استاندارد باید دارای توزیع نرمال باشد و همچنین پارامتر هر متغیر در سراسر سطوح شدت تصادف که منجر به از دست دادن انعطاف‌پذیری در دسته احتمالات داخلی می‌شود، ثابت است. همان‌طور که توسط واشنگتن و همکاران در سال (۲۰۱۰) بیان شده است. هرچند مدل OP ترتیبی از شدت تصادفات را به رسمیت می‌شناسد اما از سوی دیگر انعطاف‌پذیری در مشخصات مدل را از دست می‌دهد. خلاصه‌ای از مدل پروبیت به صورت زیر آورده شده است (Kropko, 2007):

۱- از ویژگی‌های مدل چندجمله‌ای پروبیت این است که در این مدل فرض استقلال گزینه‌های نامرتبط وجود ندارد.

۲- نقطه ضعف بسیار آشکار در این روش این است که محاسبات فشرده‌تری نسبت به مدل‌های لوجیت دارد.

۳- توزیع خطاها در این مدل باید نرمال باشد.

٤- احتمال انتخاب با استفاده از این مدل بسیار پیچیده است علت این امر این است که ساختار مدل به این شکل است که به ازای N جایگزین (سطوح متغیر وابسته) N-1 انتگرال باید محاسبه شود.

۵- مورد فوق به حدی جدی و مهم است که برای تعداد سطوح زیاد حتماً باید کامپیوتری از نظر سخت‌افزاری قدرتمند داشت و همچنین زمان زیادی هم برای برآورد در نظر گرفت.

٦- در آزمایش‌ها مشخص شده است که حتی برای مدل پروبیت باینری هم نمی‌توان هم‌زمان هم ضرایب و هم واریانس خطاها را داشت.

۷- از نکات دیگر این مدل این است که باید به صورت جداگانه برای محاسبه P-value براساس ضرایب و انحراف استاندارد اقدام کرد.

۸- حداکثر درست‌نمایی در مدل چند جمله‌ای پروبیت با مدل چند جمله‌ای لوجیت فرق دارد و بسیار پیچیده‌تر است. هر چند روش شبیه سازی برای تقریب حداکثر درست‌نمایی وجود دارد ولی حتی این کار نیز بسیار زمان‌بر است و هرچه سطوح متغیر وابسته مدل بیشتر شود زمان نیز بیشتر می‌شود.

۹- مدل چند جمله‌ای لوجیت به علت توضیحات فوق زمانی کارآمدتر است که انتخاب‌های کمتری وجود داشته باشد که این میزان انتخاب حداکثر سه یا چهار سطح است.

۱۰- برای ارزیابی مدل نیز می‌توان از P-value استفاده کرد که در واقع میزان درستی ضرایب برآورد شده را به ما نشان می‌دهد (به عنوان تشریح جزئی‌تر هرگاه در مدلی ضرایب رگرسیونی از دقت بالاتری در برآورد برخوردار باشد نتیجه آن کمتر شدن خطای استاندارد ضرایب برآوردی می‌باشد. از سوی دیگر با توجه به رابطه مستقیم P-value با خطای استاندارد می‌توانیم مقدار مناسب P-value را به عنوان معیاری برای ارزیابی مدل در مدل‌های لوجیت در نظر گرفت.)

۱۱- همچنین می‌توان مدل چند جمله‌ای پروبیت و لوجیت را بر اساس درصد موفقیت در برآورد ضرایب با یکدیگر سنجید.

۳-۳ داده‌کاوی

تکنیک‌های متنوعی در داده کاوی وجود دارند که به تولید الگوهای متفاوتی می‌پردازند. روش‌های کشف قوانین انجمنی که به دنبال رابطه‌ای میان صفات می‌گردند، روش‌های طبقه‌بندی داده‌ها که به ساخت یک مدل می‌پردازند و همچنین تکنیک‌های خوشه بندی که نمونه‌ها را بر اساس تجاوز میان آنها در گروه‌های متفاوتی قرار می‌دهند، از عمده‌ترین راهکارهایی محسوب می‌شوند که این الگوها را تولید می‌کنند. تنوع داده‌ها و کاربردهای داده کاوی چالش‌های تحقیقاتی بسیاری را برای آن ایجاد نموده است. طراحی و پیاده سازی

www.SID.ir

Archive of SID

سیستم‌های داده کاوی برای کاربردهای خاص، توسعه الگوریتم‌های مقیاس‌پذیر، استانداردسازی یک زمان برای عملیات داده کاوی، طراحی و پیاده‌سازی روش‌های جدید برای مقابله با انواع داده‌های پیچیده مانند متن، توسعه داده کاوی توزیع شده و بلادرنگ و همچنین بررسی موضوعاتی مانند امنیت و مهندسی نرم افزار در داده‌کاوی از این چالش‌ها محسوب می‌شوند (اسماعیلی، ۱۳۹۳). الگوریتم‌های متعددی برای ساخت درخت تصمیم وجود دارند. که در این بخش به برخی معروف‌ترین آنها اشاره می‌شود (اسماعیلی، ۱۳۹۳):

الگوریتم ID3

این الگوریتم یکی از ساده‌ترین الگوریتم‌های درخت تصمیم است که از معیار Information Gain استفاده می‌کند. در اجرای این الگوریتم دو شرط توقف وجود دارد. یکی این که کلیه نمونه‌های باقیمانده متعلق به یک کلاس باشند و یا اینکه پس از محاسبه مقدار معیار Information Gain بهترین آن بزرگتر از صفر نباشد هیچ‌گونه روش هرس کردنی در آن موجود نیست و می‌تواند صفات خاصه عددی و داده‌های ناقص را به عنوان ورودی بپذیرد.

الگوریتم C5

این الگوریتم یکی از تعمیم‌های الگوریتم ID3 است که از معیار Gain Ratio جهت انتخاب صفت خاصه استفاده می‌کند. الگوریتم هنگامی متوقف می‌شود که تعداد نمونه‌ها کمتر از مقدار مشخص شده‌ای باشد. این الگوریتم از تکنیک پس هرس استفاده می‌کند و همانند الگوریتم قبلی داده‌های عددی را نیز می‌پذیرد. با کمی تغییر می‌توان برای داده‌های ناقص از آن استفاده کرد.

الگویتم CART

نتیجه این الگوریتم یک درخت تصمیم دودویی است. بدین معنی که هر گره داخلی به طور دقیق دارای دو انشعاب است. از معیار TWOING استفاده می‌کند و روشی را نیز برای هرس کردن دارد. یکی از ویژگی‌های مهم الگوریتم CART توانایی تولید درختان رگرسیون است. برگ‌ها در چنین درختی یک عدد واقعی را به جای برچسب کلاس تخمین می‌زنند.

٤- تحلیل آماری و مدل پیش بینی تصادفات

٤-١- تحلیل فراوانی متغیرهای وابسته و مستقل

تحلیل فراوانی متغیرهای وابسته و مستقل با استفاده از نرم‌افزار SPSS نسخه ۲۲ گرفته است که در جداول (۱) خروجی‌های این تحلیل آورده شده است. جدول (٤) خلاصه فراوانی کلی بانک اطلاعاتی پژوهش را نشان می دهد. در این تحلیل تعداد آمار برابر ۳۰۵ و تعداد آمار از دست رفته صفر می‌باشد. سایر ویژگی‌های آماری شامل مد، محدوده تغییرات، حداقل و حداکثر مقادیر داده‌ها نیز ارایه گردیده است. در جدول (۲) تحلیل فراوانی متغیر وابسته به صورت خلاصه ارایه گردید. این جدول شامل پارامتر تکرار،درصد شامل شده و درصد تجمعی آمارها می‌باشد. بر اساس تحلیل انجام شده تصادفات خسارتی ۱۸۳ بار بیشترین تعداد تصادفات فوتی با ١٦ تکرار کمترین درصد را به خود اختصاص داده اند. جدول (۳) به ارایه خلاصه از تحلیل آماری فراوانی نوع وسیله نقلیه مقصر پرداخته است.

جدول ۱. فراوانی کلی متغیرهای مستقل کیفی و متغیر وابسته

		شدت تصادف	وسیله نقلیه مقصر	وسیله نقلیه غیرمقصر	علت تامه تصادف	زمان تصادف	جنسیت مقصر	جنسیت غیرمقصر
N	Valid	۳۰۵	۳۰۵	۳۰۵	۳۰۵	۳۰۵	۳۰۵	۳۰۵
	Missiig	۰	۰	۰	۰	۰	۰	۰
Mode		۰	۰	۰	۵	۱	۱	۱
Range		۲	۵	٦	۸	۱	۱	۱
Minimum		۰	۰	۰	۰	۰	۰	۰
Maximum		۲	۵	٦	۸	۱	۱	۱

www.SID.ir

Archive of SID

جدول ۲. جدول فراوانی متغیر وابسته

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	خسارتی	۱۸۳	٦۰	٦۰	٦۰
	جرحی	۱۰٦	۳٤/۸	۳٤/۸	۹٤/۸
	فوتی	۱٦	۵/۲	۵/۲	۱۰۰
	Total	۳۰۵	۱۰۰	۱۰۰	

جدول ۳. جدول فراوانی متغیر مستقل- وسیله نقلیه مقصر

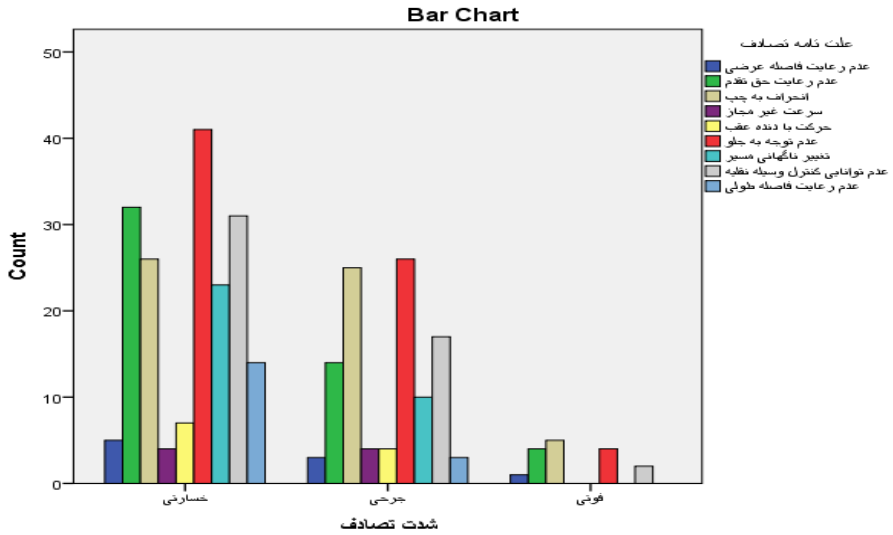
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	سواری	۱۹۹	٦۵/۲	٦۵/۲	٦۵/۲
	وانت	۵۹	۱۹/۳	۱۹/۳	۸٤/٦
	کامیون	۲۸	۹/۲	۹/۲	۹۳/۸
	موتورسیکلت	۱۱	۳/٦	۳/٦	۹۷/٤
	کشاورزی	٤	۱/۳	۱/۳	۹۸/۷
	غیره	٤	۱/۳	۱/۳	۱۰۰
	Total	۳۰۵	۱۰۰	۱۰۰	

جدول ٤. جدول فراوانی متغیر مستقل – بر اساس علت تصادف

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	**عدم رعایت فاصله عرضی**	۹	۳	۳	۳
	عدم رعایت حق تقدم	۵۰	۱٦/٤	۱٦/٤	۱۹/۳
	انحراف به چپ	۵٦	۱۸/٤	۱۸/٤	۳۷/۷
	سرعت غیرمجاز	۸	۲/٦	۲/٦	٤۰/۳
	حرکت بادنده عقب	۱۱	۳/٦	۳/٦	٤۳/۹
	عدم توجه به جلو	۷۱	۲۳/۳	۲۳/۳	٦۷/۲
	تغییرناگهانی مسیر	۳۳	۱۰/۸	۱۰/۸	۷۸
	عدم توانایی کنترل وسیله نقلیه	۵۰	۱٦/٤	۱٦/٤	۹٤/٤
	عدم رعایت فاصله طولی	۱۷	۵/٦	۵/٦	۱۰۰
	Total	۳۰۵	۱۰۰	۱۰۰	

www.SID.ir

Archive of SID



شکل۱. نمودار میله ای جدول توافقی

جدول (٤) نتایج تحلیل متغیر مستقل علت تامه تصادفات را نشان می‌دهد. تحلیل های آماری نشان می دهد که متغیر مستقل انحراف به چپ، عدم توجه به جلو و عدم رعایت حق تقدم بیشترین عامل تصادفات در محور مورد پژوهش می باشد. فراوانی سایر عوامل موثر در جدول آورده شده است. جدول (۵) و نمودار میله‌ای (۱) به وضوح فراوانی هر یک از متغیر های مستقل علت تامه در سطوح مختلف شدت تصادفات نشان می دهد. بر اساس این تحلیل انجام گرفته علت عدم توجه به جلو، عدم رعایت حق تقدم، عدم توانایی کنترل وسیله نقلیه به ترتیب بیشترین فراوانی را در سطح خسارتی به خود اختصاص داده‌اند. در عوامل سایر سطوح متغیر وابسته شدت تصادفات نیز دارای فراوانی زیادی می‌باشد.

٤-۲- بررسی نتایج مدلسازی انجام شده به روش چندجمله‌ای لوجیت

تحلیل خروجی‌های حاصل از مدلسازی انجام گرفته به صورت زیر خلاصه شده است:

۱- در مدل‌سازی انجام‌شده با استفاده از چند جمله ای لوجیت نتایج حاصل مطلوب به نظر می رسد.

۲- تعداد پارامترهای معنی‌دار شده در این روش با احتساب این‌که P-VALUE های کوچک‌تر مساوی ۰/۲۰ را در مدل نگه داشته ایم برابر ۲۳ پارامتر از ٤٦ پارامتر تخمین زده شده و مورد قبول بوده است. با توجه به اینکه پارامتر معنی دار مناسب برای تغییر مدل وجود داشته است لذا نظر دهی در مورد این مدل را ممکن نموده است.

۳- مدل چند جمله‌ای لوجیت در برآورد حداکثر درست نمایی از روش شبکه عصبی مصنوعی (ANN) استفاده می‌کند و این روش سبب می‌شود که، انحراف استانداردها در این حالت با دقت بالاتری برآورد شده و در نتیجه سبب آن می شود که، تعداد ضرایب ارزشمند برآورد شده، که بر اساس P-VALUE است؛ بیشتر و در نهایت مدل قابل تفسیر می شود.

٤- در نتیجه اولاً به علت این‌که با وجود تعداد زیاد سطوح متغیرهای مستقل کیفی این پارامترها برآورد شده‌اند و ثانیا به علت این که به تعداد سطوح متغیر وابسته پارامتر برآورد شده داریم؛ احتمالاً این مدل برای این نوع از داده‌ها مناسب است و نظر دهی در مورد این مدل امکان پذیر می باشد. باتوجه به توضیحات فوق مدل چندجمله‌ای لوجیت برای این پژوهش مطلوب است و گزینه اصلی مدل‌سازی این پژوهش محسوب می‌شود.

۵- در این مدل ۲ آماره در خروجی نرم‌افزار وجود دارد که شامل انحراف باقیمانده و آکایک است و این دو آماره هرچه کمتر باشند بهتر است که میزان این آماره‌ها برای انحراف باقیمانده برابر با ۹۱۸/ ٤۳٦ و برای آکایک برابر با ۹۱۸ /۵۲۸ است. فرمول سطوح با استفاده از مدل چندجمله‌ای لوجیت به

www.SID.ir

Archive of SID

صورت فرمول ٤ می‌باشد.

$$\ln\left(\frac{P_j}{P_2}\right)=\boldsymbol{A_j}=\beta_{j,0}+\beta_{j,1}(\mathrm{VF1})+\beta_{j,2}(VF2)+\beta_{j,3}(VF3)+\beta_{j,4}(VF4)$$
$$+\beta_{j,5}(VF5)+\beta_{j,6}(WVNF1)+\beta_{j,7}(VNF2)+\beta_{j,8}(VNF3)+\beta_{j,9}(VNF4)$$
$$+\beta_{j,10}(VNF5)+\beta_{j,11}(VNF6)+\beta_{j,12}(Cause1)+\beta_{j,13}(Cause2)+\beta_{j,14}(Cause3) \qquad (٤)$$
$$+\beta_{j,15}(Cause4)+\beta_{j,16}(Cause5)+\beta_{j,17}(Cause6)+\beta_{j,18}(Cause7)$$
$$+\beta_{j,19}(Cause8)+\beta_{j,20}(Tcrash1)+\beta_{j,21}(GF1)+\beta_{j,22}(GNF1)$$

برای متغیرهای کیفی، فقط ضریب و برای متغیرهای کمی، ضریب به همراه مقدار عددی پارامتر موردنظر قرار داده می‌شود. در کل ۳ سطح وجود دارد (سطح ۲-۰)که سطح آخر مبنا است و سایر سطوح نسبت به این سطح نوشته می‌شوند. در فرمول ۲۲ ضریب و یک عرض از مبدأ داریم که هر یک p-value های خاص خود را دارند که بر اساس آن تصمیم گرفته می‌شود که در فرمول باقی بمانند یا خیر و p-value هایی که کوچتر و مساوی ۰/۲ باشند را در فرمول نگه می‌داریم و مابقی را حذف می‌کنیم. حال برای تبدیل کردن این فرمول به فرمول پیش‌بینی کافی است از طرفین EXP گرفته شودکه در زیر خلاصه‌شده آن را می‌بینیم:

$$SEVERITY_j \rightarrow \frac{P_j}{P_2}=P(SEVERITY_j)=\frac{e^{A_j}}{1+\sum_{j=0}^{1}e^{A_j}} \text{ و } A_j=\beta_{j,0}+\sum_{i=1}^{k}\beta_{j,i}\,x_{j,i} \qquad (٥)$$

$$j=0,1 \text{ و } i=1,2,\dots\dots\dots\dots,22$$

٤-۲-۱- فرمول سطح صفر (تصادف خسارتی)

برای $\mathrm{SEVERITY}_0$ ضریب‌هایی که با عبارت مقابل‌شان حذف می‌شوند عبارت‌اند از:

$$\beta_{0,0},\beta_{0,1},\beta_{0,8},\beta_{0,12},\beta_{0,13},\beta_{0,16},\beta_{0,18},\beta_{0,20}\beta_{0,21},\beta_{0,22}$$

در فرمول زیر که مربوط به سطح صفر (تصادفات خسارتی) است ۱۳ ضریب داریم که فرمول این سطح به شکل زیر است:

$$\ln\left(\frac{P_0}{P_2}\right)=\boldsymbol{A_0}=32.37*(VF2)+31.75*(VF3)+18.81*(VF4)$$
$$+11.07*(VF5)-1.29*(WVNF1)-2.04*(VNF2)-2.93*(VNF4) \qquad (٦)$$
$$+12.21*(VNF5)+13.49*(VNF6)+25.78*(Cause3)$$
$$+20.24*(Cause4)+25.50*(Cause6)+24.20*(Cause8)$$

$$SEVERITY_0 \rightarrow \frac{P_0}{P_2}=P(SEVERITY_0)=\frac{e^{A_0}}{1+\sum_{j=0}^{1}e^{A_j}}$$
$$\text{و } A_0=\beta_{0,0}+\sum_{i=1}^{22}\beta_{0,i}\,x_{0,i} \qquad (٧)$$

٤-۲-۲- فرمول سطح یک (تصادف جرحی)

برای SEVERITY1 ضریب‌هایی که با عبارت مقابلشان حذف می‌شوندعبارت‌اند از: در فرمول زیر که مربوط به سطح یک (تصادفات جرحی) است ۱۰ ضریب داریم که فرمول این سطح به شکل زیر است:

$$\ln\left(\frac{P_1}{P_2}\right)=\boldsymbol{A_1}=30.91*(VF2)+31.21*(VF3)+19.29*(VF4)$$
$$+12.12*(VF5)+12.49*(VNF5)+14.78*(VNF6)+26.54*(Cause3) \qquad (٨)$$
$$+20.46(Cause4)+25.31(Cause6)+23.57(Cause8)$$

www.SID.ir

Archive of SID

حال برای تبدیل کردن این فرمول به فرمـول پـیش‌بینـی کـافی است از طرفین EXP گرفته شودکه در زیر خلاصه‌شـده آن را می‌بینیم:

$$SEVERITY_1 \rightarrow \frac{P_1}{P_2} = P(SEVERITY_1) = \frac{e^{A_1}}{1 + \sum_{j=0}^{1} e^{A_j}} \quad و$$

$$A_1 = \beta_{1,0} + \sum_{i=1}^{22} \beta_{1,i}\, x_{1,i} \qquad (۹)$$

٤-۲-۳- درصد درست پیش بینی و ماتریس خطا مرتبط بـا مدل MNL

در جدول (٦) میزان درصد درست پیش بینی و ماتریس خطا بین مقادیر مشاهده شده و مقادیر پیش بینی شده متغیر وابسـته (SEVERITY) در مدل MNL برای تمامی مشاهدات بانک اطلاعاتی ارائه شده است. در ماتریس خطا اولین سـطر بیـانگر سطوح پیش بینی و اولین ستون بیـانگر سـطوح مشـاهده شـده می‌باشند. هرگاه مقادیر پیش بینـی و مقـادیر مشـاهده شـده بـا یکدیگر برابر باشند در قطر اصلی ماتریس قرار مـی گیرنـد. در جدول (۷) مقـادیر روی قطـر اصـلی بـا رنـگ سـبز مشـخص شده‌اند که بیانگر تعداد مشاهدات صحیح در سطح مـورد نظـر می باشند.

جدول ٦. خروجی بخش آنالیز نرم‌افزار R برای مدل MNL خلاصه، مقایسه بین مقادیر مشاهده شده و پیش بینی شده

صحیح	۲۰۵	٦۷/۲۱%
نادرست	۱۰۰	۳۲/۷۹%
کل	۳۰۵	

جدول ۷. ماتریس خطا بین مقادیر مشاهده شده و مقادیر پیش بینی شده متغیر وابسته

	۰	۱	۲
۰	۱٦٤	۱۹	۰
۱	٦٥	٤۱	۰
۲	۹	۷	۰

٤-۲-٤- مقایسه نتایج مدل‌سازی انجام شده

در این پژوهش برای مـدل سـازی ازالگـوریتم CART از مجموعه الگـوریتم هـای درخـت تصـمیم، الگـوریتم-ANN MLP از مجموعه شبکه های عصـبی مصـنوعی و مـدل چنـد جمله‌ای لوجیت از مجموعه مدل‌هـای انتخـاب اسـتفاده شـده اسـت کـه نتـایج آن‌هـا بـه صـورت خلاصـه در جـدول (۸) آورده شده است:

جدول ۸. خلاصه۳ مدلسازی انجام شده

نام مدل	درصد درست پیش بینی	سطوح صحیح پیش بینی شده توسط مدل	رتبه از نظر تعداد سطوح صحیح	رتبه از نظر درصد درست پیش بینی	بهترین مدل بر اساس دقت، سرعت و قابلیت تحلیل
MNL	٪٦۷/۲۱	۱, ۰	۱	۳	در مدل انتخاب مدل MNL بهترین است.
CART	٪٦۸/٥۲	۱, ۰	۱	۱	در مدل های داده کاوی بهترین الگوریتم CART است.
ANN– MLP	٪٦۸/۲	۱, ۰	۱	۲	

www.SID.ir

Archive of SID

۵- نتایج

در مدل‌سازی انجام‌شده با استفاده از چند جمله ای لوجیت نتایج حاصل مطلوب به نظر می رسد. تعداد پارامترهای معنی‌دار شده در این روش با احتساب این‌که P-VALUE های کوچک‌تر مساوی ۰/۲۰ برابر ۲۳ پارامتر از ٤٦ پارامتر تخمین زده شده که مورد قبول بوده است. با توجه به اینکه مدل چند جمله‌ای لوجیت در برآورد حداکثر درست نمایی از روش شبکه عصبی مصنوعی استفاده می‌کند مقادیر انحراف استانداردها در این حالت با دقت بالاتری برآورد می‌شود. بر این اساس تعداد ضرایب ارزشمند (P-VALUE) بیشتر و در نهایت مدل قابل تفسیرتری حاصل می‌شود. براساس مطالعه‌ها صورت گرفته شده در این پژوهش این نتیجه به دست آمده است که بهترین مدل از نظر درصد درست پیش بینی و قابلیت ارائه فرمول پیش بینی برای هر سطح مدل MNL بوده است. این مدل دارای دقت پیش بینی ٦٧/٢١ درصدی است و با توجه به نتایج مدل سازی مطلوبیت این مدل جهت برآورده کردن هدف پژوهش نسبت به سایر مدل‌های ساخته شده، بیشتر بوده است. نتایج بدست آمده در بخش مدل‌های پیش بینی نشان می‌دهد فرمول برآورد شده قادر به پیش بینی شدت تصادفات در سطوح۰ (صفر) و ۱ با دقت کافی می باشند. نتایج حاکی از آن است که دقت این مدل با افزایش سطح شدت تصادفات کاهش می یابد بنحوی که این مدل نتوانسته است شدت تصادفات را در سطح خسارت ۲ درست پیش بینی نماید که علت این امر تعداد آمارهای مربوط به خسارات فوتی می‌باشد. مدل MNL در ۲ سطح توانسته ارزیابی صحیحی از متغیر وابسته داشته باشد که شامل سطوح ۰ و ۱ می باشد و درست درصد پیش بینی که مجموع مشاهدات صحیح در سطوح فوق می باشد برابر است با ٦٧/٢١ درصد است که رتبه۳ را از نظر دقت دارا می باشد. مدل CART در ۲ سطح توانسته ارزیابی صحیحی از متغیر وابسته داشته باشد که شامل سطح ۰ و ۱ می‌باشد و درست درصد پیش‌بینی که مجموع مشاهدات صحیح در این سطح می‌باشد برابر است با ٦٨/٥٢ درصد است که رتبه ۱ را از نظر دقت دارا می‌باشد. مدل ANN– MLP در ۲ سطح توانسته ارزیابی صحیحی از متغیر وابسته داشته باشد که شامل سطوح ۰ و ۱ می‌باشد و درست درصد پیش‌بینی که مجموع مشاهدات صحیح در سطوح فوق می‌باشد برابر است با ٦٨/٢ درصد است که رتبه ۲ را از نظر دقت دارا می‌باشد. هدف پژوهش ارایه مدل پیش بینی می‌باشد که بتواند برای هر سطح نیز فرمول پیش بینی ارایه دهد؛ این هدف را مدل چندجمله‌ای لوجیت (MNL) با درصد درست پیش بینی ٦٧/٢١ برآورده کرده است. در نتیجه مدل MNL به عنوان مدل برگزیده این پژوهش انتخاب شده است.

٦-مراجع

-اداره ایمنی و ترافیک سازمان راهداری و حمل و نقل جاده‌های کشور، (١٣٨٤)، "گزارش آماری و تحلیل تصادفات جاده‌ای"، انتشارات اداره راه تهران، تهران.

-اسماعیلی، م.، (۱۳۹۳) "داده‌کاوی (مفاهیم و تکنیک ها)"، (چاپ دوم)، تهران، نیاز دانش، پژوهشکده حمل ونقل وزارت راه و شهرسازی، (۱۳۸۹).

-راضی اردکانی، ح.، صمیمی، الف.، (۱۳۹۰)، "بررسی عوامل موثر بر شدت تصادفات درون شهری با استفاده از مدل های پروبیت، لوجیت و شبکه عصبی مصنوعی"، یازدهمین کنفرانس بین المللی حمل ونقل و مهندسی ترافیک، تهرانَ، ایران.

-نصیری، ح ا.، ادریسی، ع.، (۱۳۸۲)، " مدل سازی و شناسایی عوامل موثر در شدت تصادفات کامیون‌ها در جاده‌های دو خطه برون شهری با استفاده از مدل لوجیت و شبکه عصبی"، ششمین کنفرانس بین المللی مهندسی عمران.

-نصیری، ح. ا. طلوعی، ر.، (۱۳۸۰)، "شناسایی عوامل موثر در شدت تصادفات جلو به جلو با استفاده از مدل پروبیت ترتیبی، اولین کنفرانی بین المللی حوادث رانندگی و جاده‌ای.

-نصیری، ح. ا.، ع. ادریسی.، (۱۳۸۲)، "مدل‌سازی و شناسایی عوامل موثر در شدت تصادفات کامیون‌ها در جاده‌های دو خطه برون شهری با استفاده از مدل لوجیت و شبکه عصبی"، ششمین کنفرانس بین المللی مهندسی عمران.

-نصیریان، م.، (۱۳۸۹)، "اولویت‌بندی ایمن سازی تقاطعات شهری بر اساس مدل رگرسیون لگاریتم طبیعی"، دهمین کنفرانس بین المللی حمل و نقل و مهندسی ترافیک تهران، ایران.

www.SID.ir

Archive of SID

crashes, accident analysis and prevention, 43(1), pp.112-122.

-Kropko, J., (2007), "Choosing between multinomial logit and multinomial probit models for analysis of unordered choice data (Doctoral dissertation, The University of North Carolina at Chapel Hill)".

-Moore, D. N., Schneider IV, W. H., Savolainen, P. T., & Farzaneh, M., (2011), "Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations". Accident Analysis & Prevention, 43(3), pp.621-630.

-Persuad, B. and Dzbik, L., (1993), "Accident perdition models for freeways", transportation research record, 1401, pp. 55-60.

-Voget, A and Bared , J., (1999), "Accident models tor two lane rural segments and intersection". transportation Research Record, Issue 1635, pp. 18-29.

-WHO (World Health Organization), (2004), "Global status report on road safety (time for action)". World Health Organization. Published in Geneva, Switzerland.

World Health Organization, (2013), "Global status report on road safety (time for action)". World Health Organization. Published in Geneva, Switzerland.

-Wu, Qiong, Chen, Feng , Zhang, Guohui, Cathy Liu, Xiaoyue, Wang, Hua, M. Bogus, Susan, (2014), "Mixed logit model-based driver injury severity investigations in single and multi-vehicle crashes on rural two- land highways", Accident Analysis and Prevention, 72, pp. 105-115.

-Yasmin, Shamsunnahar, Eluru, Naveen, R. Pinjari, Abdul, Tay, Richard, (2014), "Examining driver injury severity in two vehicle crashes-A copula based all roach", Accident Analysis and Prevention, 66, pp. 120-135.

-Zeng, Qiang, Huang, Helai, (2014), "A stable and optimized neural network model for crash injury severity prediction", Accident Analysis and Prevention, 73, pp. 351-358.

-Zuxuan, d., Ivan, j.n., garder. P., (2006), "in analysis of factors affecting the severity of head on crashes two-lane rural highways in Connecticut", pp. 137-146.

–Barua, U., Tay., (2010), "Severity of Urban Transit Bus Crashes in Bangladesh". Journal of Advanced transportation, Vol. 44(I), pp. 34-41.

-Behnood, Ali, L. Mannering, Fred, (2015), "The temporal stability of factors affecting driver-injury severities in single- vehicle crashes: Some empirical evidence", Analytic Methods in Accident Research, 8, pp. 7-32.

-Celik, Ali Kemal, Oktay, Erkan, (2014), "A multinornail logit analysis of fisk factors influencing road trafficinjury severities in thr Erzurum and Kars Provinces of Turkey", Accident Analysis and Prevention, 72, pp.66-77.

-Haleem, k., Abdel-aty, m., (2010), "Examining traffic crash injury severity at signalized intersections", journal of safety research, 41(4), pp. 347-357.

-Haleem, Kirolos, Gan, Albert, (2013), "Effect of Driver's Age and Side of Impact on Crash Severity along Urban Freeways: A Mixed Logit Approach", Journal of safety Research, JSR-01087, pp. 10-11.

- Haque, M. M., Chin, H. C., & Huang, H. (2009). Modeling fault among motorcyclists involved in crashes. Accident Analysis & Prevention, 41(2), 327-335.

-Jones, A. P., Jorgensen, S. H., (2013), "The use of multilevel models for the prediction of road accident outcomes, accident analysis and prevention, 35(1), pp. 59-69.

-Kaplan, S., C. G. Prato. (2012), "Risk Factors Associated with Bus Accident Severity in the United States: A Generalized Ordered Logit Model". Journal of Safety Research, Vol. 43(3), pp. 171-180.

-Khattak, A. J., Kantor. P. and Council, F. M. (1999), "Role of adverse weather in key crash type on limited: Access roadways implications for advanced weather systems". Transportation Research Record, Issue 1621, pp.15-19.

-Kononen, d.w., Flannagan , c.a.c, Wang, s. c., (2011), identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle

www.SID.ir

Archive of SID

Presenting Model of Intensity Estimation of Vehicle Accidents Using Accident Data (Case Study: Babol-Ganj Afrooz Road)

R. Babagoli, Department of Civil Engineering, University of Science and technology of Mazandaran, Behshahr, Mazandaran, Iran.
A. R. Ameli , Department of Civil Engineering, Malard Branch, Islamic Azad University, Malard, Tehran, Iran.
A. A. Gholamrezatabar, M.Sc., Grad., Aryan Institute of Science and Technology, Amirkola, Mazandaran, Iran.
A. Paydar, Department of Civil Engineering, Malard Branch, Islamic Azad University, Malard, Tehran, Iran.

Email: Rezvan_babagoli@yahoo.com

Received: September 2019-Accepted: December 2019

ABSTRACT

Due to its beautiful nature, province of Mazandaran which is always one of the tourist and tourist provinces, where many people travel to the province every year in different seasons. The province is also one of the country's agricultural hubs, which contribute a lot to the transportation of this industry through road transport, which creates high traffic volume traffic and, as a result, traffic accidents. The statistics provided in recent years show that this province is the seventh province for casualties caused by accidents in the country. Based on this, a lot of research has been carried out on the identification of factors affecting the severity of accidents and there have been many advances in this regard, but studies on the relationship between the severity of accidents and the type of collision have been negligible, so further research in this It seems necessary. In this research, using a model of logistic polynomials from a set of selection models for predicting the intensity of accidents has been used. Also, using a prediction model of two sets of data mining algorithms including CART algorithm as one of decision tree algorithms and ANN-MLP algorithm, an array of artificial neural network algorithms was used and the required results were extracted and with each other have been compared. According to the studies carried out in this study, it has been shown that the best model was the MNL model in terms of the correct prediction and the ability to present the prediction formula for each level. The results obtained in the forecasting section show that the estimated formula is able to predict the severity of accidents at 0 (zero) and 1 levels with sufficient accuracy.

Keywords: Severity of Crashes, Vehicles, Prediction Model, Logit, Neural Network

www.SID.ir