ارائه مدل‌های هوشمند طبقه بندی داده‌های تصادفات ترافیکی بر مبنای شبکه بیزین، درخت تصمیم و شبکه عصبی مصنوعی

مقاله علمی – پژوهشی

علی نادران، استادیار، دانشکده مهندسی عمران، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
محمدرضا احدی، دانشیار، مرکز تحقیقات راه، مسکن و شهرسازی، تهران، ایران
رسول ذبیحیان*، دانشجوی دکتری، گروه مهندسی عمران، واحد بین‌المللی کیش، دانشگاه آزاد اسلامی، جزیره کیش، ایران
*پست الکترونیکی نویسنده مسئول: r.zabihian@iau.ir

دریافت: ۱٤۰٤/۰۱/۲۷– پذیرش: ۱٤۰٤/۰۵/۰۱
صفحه ۱۷٤-۱٦۳

چکیده

هرسال تعداد زیادی از کاربران جاده درنتیجه تصادفات ترافیکی جان خود را از دست می‌دهند یا مجروح می‌شوند. تحلیل نقاط حادثه‌خیز برای و رانندگی، سازمان حمل‌ونقل و ترافیک شهرداری‌ها و سازمان‌های‌امرتبط با ایمنی‌اترافیک بسیار با اهمیت می‌باشد. برای شناسایی نقاط پرخطرتصادف از روش‌هایی مانند نزدیک‌ترین همسایگی و برازش استفاده می‌شود، اما این روش‌ها پیش‌بینی ناکافی و درصد برآورد نادرستی دارند. هدف این تحقیق، طبقه بندی داده‌های اتصادفات و ارائه یک روش بهتر ابرای شناسایی نقاط پرخطر از نظر تصادف است. برای این منظور، عملکرد درخت‌های تصمیم، مدل بیزین، و شبکه عصبی با استفاده از روش‌های متداول مانند نزدیک‌ترین همسایگی و برازش مقایسه شدند. با استفاده از ابزار GIS و روش‌های مبتنی بر تراکم همانند تابع چگالی کرنل، نقاط پرخطر تصادفات به صورت نقشه‌های کاربردی اشناسایی شدند. نتایج نشان داد که روش درخت تصمیم، شبکه بیزین و شبکه عصبی به ترتیب با ۸۳٪ و ۷۸٪ درصد برآورد صحیح، نسبت به روش‌های سنتی عملکرد بهتری دارند. همچنین، نتایج نشان داد که تعداد تصادفات در دوره‌های اوج ترافیک، به عنوان مثال صبح و عصر، نسبت به دوره‌های غیر اوج بیشتر است. با بررسی نقاط حادثه‌خیز پس از شناسایی به این نتیجه دست می‌توان یافت که نقاط حادثه‌خیز درون‌شهری بیشتر در نزدیکی تقاطعات بوده و اغلب درءاشرایط جوی مناسب و در روشنایی کافی روز اتفاق می‌افتند.

واژه‌های کلیدی: تصادفات داخل شهری، طبقه‌بندی، GIS، شبکه بیزین، درخت تصمیم و شبکه عصبی مصنوعی

۱-مقدمه

ایمنی، یـــک پارامتربســـیارمهم در کیفیت زندگی ادر بسیاری از کشورهای پیشرفته از دیدگاه مردم و مجامع مهم بین‌المللی است. تصادفات ترافیکی سالانه منجر به فوت، معلولیت دائمی و جراحت افراد زیادی از کاربران راه‌ها مخصوصاً در کشورهای درحال‌توسعه می‌شود (لینهو لی و همکارانش، ۲۰۰۷). تعریف و توصیف عوامل مؤثر در تصادفات اقدامی عملی برای کاهش و به حداقل رساندن عوامل ایجاد تصادف است که باعث بهبود ایمنی سیستم حمل‌و نقل می‌شود. برای این منظور، ضرورت دارد مکان‌های پرخطر تصادف را تشخیص داد. سیستم اطلاعات مکانی GIS ابزار مناسبی برای شناسایی و تحلیل نقاط پرخطر تصادفات می‌باشد که در چند دهه‌ی اخیر بسیار مورداستفاده قرارگرفته است. تصادفات جاده‌ای یکی از چالش‌های عمده جوامع مدرن است که سالانه هزینه‌های انسانی، اقتصادی و اجتماعی قابل توجهی به همراه دارد. با افزایش روزافزون تعداد وسایل نقلیه و پیچیدگی‌های روزافزون ترافیک، نیاز به روش‌های مؤثر و کارآمد برای تحلیل و پیش‌بینی تصادفات بیش از پیش احساس می‌شود.

در این راستا، استفاده از مدل‌های هوشمند و روش‌های پیشرفته می‌تواند به بهبود تصمیم‌گیری در زمینه مدیریت ترافیک و ارتقای ایمنی جاده‌ها کمک کند. هدف از این مطالعه، ارائه مدل‌های هوشمند برای طبقه‌بندی داده‌های تصادفات است که به‌ویژه بر مبنای سه رویکرد متفاوت یعنی شبکه بیزین، درخت تصمیم و شبکه عصبی مصنوعی بنا نهاده شده‌اند. این مدل‌ها به دلیل قابلیت‌های متفاوت خود در تحلیل داده‌های پیچیده و غیرخطی، می‌توانند به شناسایی الگوهای پنهان در داده‌های تصادفات کمک کرده و نتیجه‌گیری‌های دقیق‌تری ارائه دهند.

ضرورت این مطالعه نه‌تنها در ایجاد بینش‌های جدید در زمینه علل و عوامل مؤثر بر تصادفات جاده‌ای، بلکه در بهینه‌سازی استراتژی‌های پیشگیرانه و کاهش آمار تصادفات محسوس است. با توسعه و به‌کارگیری مدل‌های هوشمند، می‌توان به شناسایی نقاط خطرناک، ارزیابی تأثیر متغیرهای مختلف بر تصادفات و از همه مهمتر، ارائه راهکارهای پیوسته و مبتنی بر داده‌های واقعی برای بهبود ایمنی جاده‌ها و کاهش تلفات و خسارات ناشی از تصادفات پرداخت. به‌کارگیری این سه مدل مختلف باعث می‌شود که مقایسه‌ای جامع میان کارایی و دقت هر یک از آنها صورت گیرد و در نهایت، بهترین رویکرد برای تحلیل داده‌های تصادفات شناسایی شود. این مطالعه می‌تواند نه‌تنها به جامعه علمی کمک کند، بلکه به‌عنوان یک منبع علمی و عملی در دسترس سازمان‌های دولتی و خصوصی، پژوهشگران و متخصصان این حوزه قرار گیرد. لذا، این تحقیق دارای ارزش بالایی در ارتقای امنیت ترافیکی و کاهش تصادفات خواهد بود. بامطالعه تصادفات می‌توان عوامل مؤثر در رخداد تصادفات ترافیکی را شناسایی کرد. ازاین‌رو تحلیل نقاط حادثه‌خیز برای راهنمایی و رانندگی، سازمان حمل‌ونقل و ترافیک شهرداری‌ها و سازمان‌های مرتبط با ایمنی ترافیک بسیار با اهمیت می‌باشد. هدف از این تحقیق شناسایی اپارامترهای تأثیرگذار و محاسبه میزان اهمیت و تأثیرگذاری هر یک از عوامل مؤثر در تصادفات می‌باشد. با تشخیص این نقاط می‌توان در جلوگیری از این عوامل پیش‌قدم شد تا به میزان مطلوبی از ایمنی در راه‌ها برسیم.

۲-پیشینه تحقیق

در کشورهای مختلف، تعاریف متعددی برای نقاط حادثه‌خیز بر اساس شاخص‌هایی نظیر شدت تصادف (فوتی، جرحی، خسارتی)، بازه زمانی (یک‌ساله، سه‌ساله، پنج‌ساله) و واحد تصادف (قطعه‌راه، نقطه تصادف، مساحت و شعاع اثر) ارائه می‌شود. در ادامه شاخص‌های حادثه‌خیزی برخی کشورهای منتخب، ارائه می‌شود. در آلمان، این کشور برای مطالعه‌ی نقاط حادثه‌خیز مدت زمان یک تا سه سال در نظر گرفته می‌شود. اگر یک سال در نظر گرفته شود. در آن صورت نقاط حادثه‌خیز، نقاطی هستند که دارای ۵ تصادف شبیه به هم صرف‌نظر از شدت تصادف در فاصله ۱۰۰ متر ثبت شوند. اگر مدت‌زمان ۳ سال موردنظر باشد، نقاطی حادثه‌خیز تلقی می‌شوند که در آن یا ۵ تصادف منجر به جراحت کم یا اینکه ۳ تصادف منجر به آسیب‌های شدید رخ‌داده باشد. مجارستان از دو نوع روش تحلیل استفاده می‌کند. در مناطق مسکونی نقاطی حادثه‌خیز محسوب می‌شوند که در طول ۳ سال حداقل ٤ تصادف در طول ۱۰۰ متر ثبت‌شده باشد. در خار ج از منطقه مسکونی نیز در فاصله ۱۰۰۰ متر در همان دوره زمانی نقطه حادثه‌خیز شناسایی می‌شود.(Elvik, R, 2008)

لینهو لی و همکارانش (۲۰۰۷) از رویکرد Bayesian برای تحلیل مکانی و زمانی تصادفات درون‌شهری و محاسبه‌ی ریسک تصادفات استفاده کردند و سپس با استفاده از نرم‌افزار (GIS Arc scene) به‌صورت سه‌بعدی تراکم تصادفات را در انقشه‌هایی اسه‌بعدی به تصویر کشیدند. پُلات و دوردوران از ترکیه در سال ۲۰۱۰ روش‌های انتخاب مشخصه مبتنی بر همبستگی و ماشین بردار و شبکه عصبی را ترکیب کرده و برای پیش‌بینی تصادفات به‌کاربرده و عوامل جوی، رطوبت هوا و زمان رخداد تصادف را به‌عنوان داده‌های مطالعه‌ی خود قرار داده است. (Kemal Polat,۲۰۱۱)

مصطفی جعفری و همکاران تعیین نقاط حادثه‌خیز تصادفات موتورسیکلت با استفاده از سیستم اطلاعات جغرافیایی در شهرستان‌های گناباد و بجستان پارساناکومار (۲۰۱۱) در هندوستان از توابع تعریف‌شده در نرم‌افزار GIS که در بحث محاسبه نقاط حادثه‌خیز استفاده کرده و سه روش شاخص همبستگی مکانی موران، گتیس–اورد جی آی استار و توابع چگالی نقطه‌ای کرنل را به کار برده است که جزو مرسوم‌ترین روش‌ها در بحث تحلیل‌های مکانی هستند. (Parsana Kumar۲۰۱۱) در جدول ۱ به کارهای گذشتگان می‌پردازیم. (Romi ۲۰۱٦ Satria)

جعفری، تبریزی، اسماعیل زاده و زمانی به تعیین نقاط حادثه‌خیز تصادفات موتورسیکلت با استفاده از سیستم اطلاعات جغرافیایی در شهرستان‌های گناباد و بجستان پرداخته‌اند. (۲۰۱۱ Jafari et al)

جدول ۱. مروری بر روش‌های تحلیل تصادفات

نویسندگان	منطقه موردمطالعه	روش مورد استفاده
کیم و نیتز (۱۹۹۵)	شهری و برون‌شهری	ابزارهای مکانی که درجه تمرکز مکانی را مشخص می‌کنند و الگوهای مکانی تصادفات مختلف را تحلیل می‌کنند
استینبرگن و همکاران (۲۰۰٤)	شهری	موقعیت تصادف، خوشه‌بندی مکانی توسط تحلیل چگالی کرنل
آگرو-والورده و ژوانیس (۲۰۰٦)	شهری و برون‌شهری	مدل بیزین کامل سلسله مراتبی با اثرات زمانی و مکانی در مقایسه با روش دوجمله‌ای منفی برای تخمین فراوانی تصادف سالانه
اردوغان و همکاران (۲۰۰۸)	شهری و برون‌شهری	شناسایی لکه‌های داغ با تحلیل‌های آماری (تحلیل چگالی کرنل و پواسون)
اردوغان (۲۰۰۹)	شهری	آماره موران و گتیس-اورد (۵)
گوندوغدو (۲۰۱۱)	شهری	روش قطعات داغ برای تعیین امتداد بحرانی از تصادفات جاده‌ای و نقاط احتمالی برای پیش‌بینی نقاط بحرانی
ترونگ و سومنهالی (۲۰۱۱)	شهری و برون‌شهری	شاخص شدت، الگوهای فضایی اطلاعات تصادفات عابران پیاده و نقشه‌های نقاط داغ نقاط عابر پیاده با خودرو.
بودیهارتو و ساییدو (۲۰۱۲)	شهری	لکه‌های سیاه تصادف: سه روش (تخمین چگالی کرنل، تحلیل خوشه و ناخوشه)
چلا و همکاران (۲۰۱۳)	شهری	الگوهای مکانی تصادف با استفاده از شبکه K-function و برآورد تراکم کرنل شبکه‌ای (KDE) برای یافتن خوشه‌ها و شناسایی مکان‌های خوشه واقعی.
رانکاوات و تیواری (۲۰۱۳)	شهری و برون‌شهری	تحقیق در مورد پتانسیل استفاده از سیستم اطلاعات جغرافیایی در شناسایی موقعیت‌های ناشی از تصادفات عابر پیاده.
عفتی و همکاران (۲۰۱٤)	شهری	رویکرد فازی فضایی جغرافیایی برای شناسایی مناطق خطرناک
ما و همکاران (۲۰۱٤)	شهری	مدل پواسون-کواسی
تورتوم و آتالای (۲۰۱۵)	شهری و برون‌شهری	آماره موران و گتیس-اورد
یالچین و دوزگون (۲۰۱۵)	شهری	سه روش تحلیل الگوی مکانی در شبکه: چگالی کرنل، نزدیک‌ترین فاصله همسایگی، تابع K

۳-روش تحقیق

فرآیند انجام تحقیق بدین‌صورت است که ابتدا باید داده‌های دریافت شده از ترکیه را برای انجام مدل‌سازی‌های لازم پاک‌سازی نموده و با تحلیل‌های آماری داده‌های مهم را از میان داده‌ها را انتخاب کنیم. روش‌شناسی این پژوهش به بررسی و پیاده‌سازی الگوریتم‌های یادگیری ماشین برای تحلیل داده‌های تصادفات جاده‌ای متمرکز است. استفاده از تکنیک‌های پیشرفته یادگیری ماشین به‌ویژه در مطالعات تحلیلی مانند این تحقیق، گام مهمی در جهت بهبود دقت و قابلیت پیش‌بینی در شناسایی و طبقه‌بندی الگوهای تصادف است. در این بخش، به تشریح مراحل اصلی روش‌شناسی این پژوهش پرداخته خواهد شد.

در گام نخست، جمع‌آوری داده‌های مربوط به تصادفات جاده‌ای شامل انواع تصادفات، زمان وقوع، مکان، شرایط جوی، نوع وسایل نقلیه و عوامل انسانی صورت می‌گیرد. این داده‌ها ممکن است از منابع مختلفی مانند سامانه‌های مدیریت ترافیک، پلیس راهنمایی و رانندگی، و پایگاه‌های داده عمومی جمع‌آوری شود. مهم است که داده‌ها به‌طور کامل و دقیق باشند تا نتایج تحلیل‌ها معتبر و قابل اعتماد باشند. داده‌های جمع‌آوری‌شده ممکن بود شامل نواقص، داده‌های گمشده، و ناهماهنگی‌ها باشند. بنابراین، در این مرحله، از تکنیک‌های پیش‌پردازش مانند پاک‌سازی، نرمال‌سازی و تبدیل داده‌ها به فرمت مناسب استفاده می‌شود. همچنین، این مرحله شامل شناسایی و حذف نویزها و داده‌های غیرمعتبر نیز می‌باشد.

در این مرحله، از سه الگوریتم یادگیری ماشین شامل شبکه بیزین، درخت تصمیم و شبکه عصبی مصنوعی برای مدل‌سازی داده‌ها استفاده می‌شود.

- **شبکه بیزین:** این الگوریتم بر مبنای نظریه احتمال پایه‌گذاری شده و به‌خوبی برای شرایط عدم قطعیت مناسب است. در این تحقیق، شبکه بیزین برای شناسایی ارتباطات میان

متغیرهای مختلف تصادف، و به‌خصوص شناسایی عوامل مؤثر بر وقوع تصادفات مورد استفاده قرار می‌گیرد.

– **درخت تصمیم:** این الگوریتم برای مدل‌سازی تصمیمات و گزینه‌های مختلف استفاده می‌شود. با ساختن درخت‌هایی که میزان احتمال وقوع تصادفات را با توجه به ویژگی‌ها و پارامترهای مختلف نشان می‌دهند، می‌توان به تحلیل عمیق‌تری از عوامل مؤثر بر این پدیده دست یافت.

– **شبکه عصبی مصنوعی:** این مدل به‌ویژه در شناسایی الگوهای غیرخطی و پیچیده مؤثر است. با استفاده از این تکنیک، می‌توان به شبیه‌سازی و پیش‌بینی الگوهای پیچیده تصادفات بر اساس داده‌های ورودی مختلف پرداخته و دقت پیش‌بینی‌ها را افزایش داد. مدل‌های ایجادشده بایستی با استفاده از مجموعه‌ای از داده‌های آموزشی آموزش داده شوند. برای این منظور، داده‌ها به دو بخش آموزش و تست تقسیم‌بندی می‌شوند. پس از آموزش مدل‌ها، از داده‌های تست برای ارزیابی دقت و کارایی هر یک از مدل‌ها استفاده خواهد شد. معیارهای ارزیابی شامل دقت، صحت، یادآوری، و F1-Score خواهد بود تا به مقایسه عملکرد مدل‌ها کمک کند.

در این مرحله، نتایج حاصل از مدل‌های مختلف مورد بررسی و تحلیل قرار می‌گیرد. نقاط قوت و ضعف هر یک از مدل‌ها به‌طور دقیق مورد ارزیابی قرار می‌گیرد تا مشخص شود کدام مدل قادر به شناسایی دقیق‌تر الگوهای تصادف است. همچنین، این مرحله شامل بررسی تأثیر ویژگی‌های مختلف بر وقوع تصادفات خواهد بود. از آنجا که این مطالعه به بررسی عملکرد سه رویکرد مختلف می‌پردازد، نتایج به‌دست‌آمده می‌تواند مبنایی برای تحقیقات آینده و توسعه مدل‌های پیشرفته‌تر باشد.

در جداول ۲ و ۳ به ترتیب آمار توصیفی داده‌های اسمی و عددی آمده است که پس از پاک‌سازی داده‌های خام می‌باشند. سپس مدل‌های مختلف را که در مراحل انجام تحقیق در شکل ۱ آمده بر داده‌ها اعمال می‌کنیم.

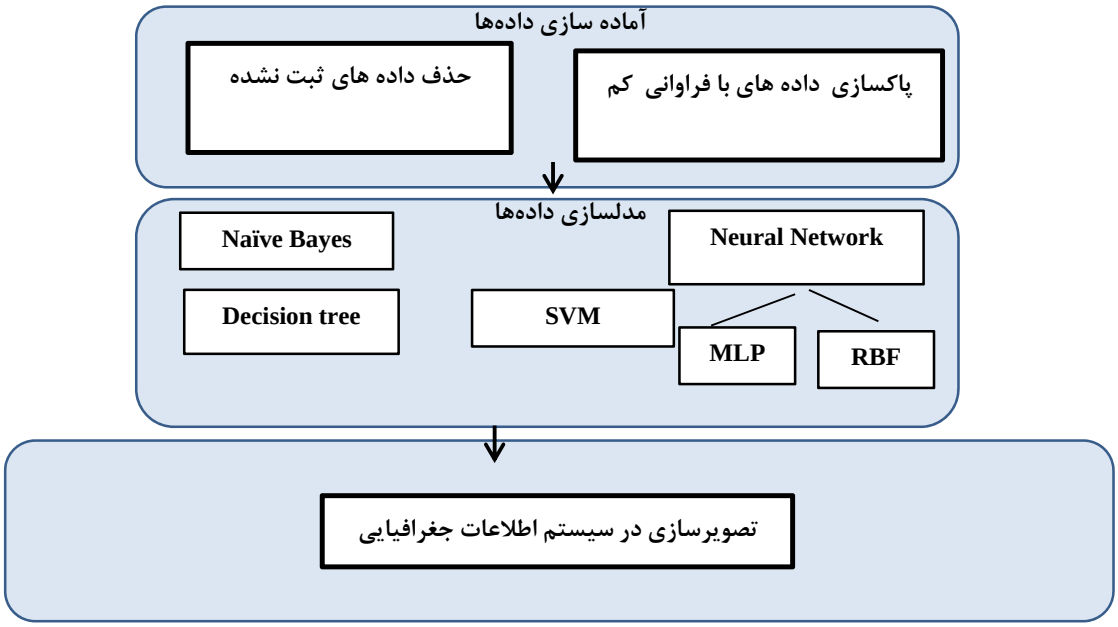
جدول ۲. فراوانی داده‌ها اسمی

متغیر	زیرشاخه	فراوانی نسبی (مطلق)	متغیر	زیرشاخه	فراوانی نسبی (مطلق)
نوع راه	جداشده	٤٨/٤ (١٣٢٣)	روز و شب	روز	٥٦/٨ (١٥٥٢)
	یک‌طرفه	٨/٨ (٢٤٠)		شب	٤٣/٢ (١١٨٢)
	دوطرفه جدا نشده	٤٢/٨ (١١٧١)	آب‌وهوا	هوای صاف	٨٤/٤ (٢٣٠٨)
طبقه راه	بزرگراه	٢٤/١ (٦٥٩)		بارانی/ برفی	١٥/٦ (٤٢٦)
	خیابان	٧٥/٩ (٢٠٧٥)	وضعیت سطح جاده	خشک	٧٩/٢ (٢١٦٤)
وجود گارد ریل	خیر	٧٨/٦ (٢١٥٠)		خیس	٢٠/٨ (٥٧٠)
	بله	٢١/٤ (٥٨٤)	هندسه افقی مسیر	قوس	١٣/٦ (٣٧٢)
وجود شانه مسیر	خیر	٧٨/١ (٢١٣٦)		مسیر مستقیم	٨٦/٤ (٢٣٦٢)
	بله	٢١/٩ (٥٩٨)	هندسه قائم مسیر	شیب‌دار	٢٧/٩ (٧٦٣)
وجود خط‌کشی	خیر	٣٢/٤ (٨٨٥)		شیب صفر	٧٢/١ (١٩٧١)
	بله	٦٧/٧ (١٨٤٩)	روشنایی مسیر	خیر	٣١ (٨٤٧)
علائم راهنمایی رانندگی	خیر	٥٥/٣ (١٥١٣)		بله	٦٩ (١٨٨٧)
	بله	٤٤/٧ (١٢٢١)	نوع تصادف	تصادف از روبرو	١٠/٧ (٢٩٣)
علائم چراغ‌دار	خیر	٨٨/٨ (٢٤٢٨)		برخورد با شیء ثابت	١٣/٩ (٣٧٩)
	بله	١١/٢ (٣٠٦)		برخورد از عقب	١٥/١ (٤١٢)
تقاطع	چهارراه	٢٢٤ (٦٦١)		برخورد زاویه‌ای	٤٦/٦ (١٢٧٣)
	دور از تقاطع	٤٥/٩ (١٢٥٥)		واژگونی	٧/٦ (٢٠٩)
	میدان	٤/٨ (١٣٢)		خروج از جاده	٦/١ (١٦٨)
	سه‌راهی T شکل	٢٠/٢ (٥٥٣)	تعداد خودرو	چند خودرویی	٧٦/٥ (٢٠٩١)
	سه‌راهی Y شکل	٤/٩ (١٣٣)		تک خودرویی	٢٣/٥ (٦٤٣)
	میانه راه	٤/٦ (١٢٦)			

متغیر	زیرشاخه	فراوانی نسبی (مطلق)	متغیر	زیرشاخه	فراوانی نسبی (مطلق)
محل تصادف	روی مسیر	۸۲/۳ (۲۲۵۱)			
	در شانه راه	۲/۲ (۵۹)			
	در کناره راه	٦/۷ (۱۸٤)			
	در پیاده‌رو	٤/۲ (۱۱٤)			

جدول ۳. آمار توصیفی داده‌های عددی

آمار توصیفی	تعداد خودروها	تعداد مجروحین	محدوده سرعت (کیلومتر بر ساعت)
کمینه	۱	۱	۲۰
بیشینه	۷	۲۳	۱۱۰
مد	۲	۱	۵۰
میانه	۱/۹۰۱	۱/۷۸	٤۹/۸۳
انحراف معیار	۰/٦۵۹	۱/٤۷۷	۱٦/٤۲۵



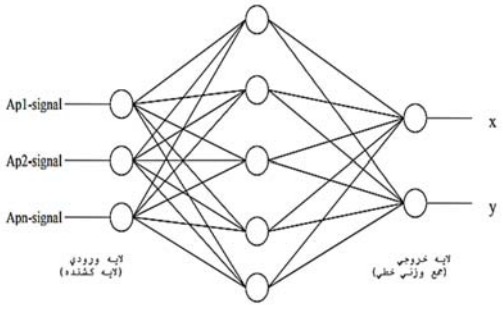
شکل ۱. مراحل انجام تحقیق

الگوریتم یادگیری درخت تصمیم پایه بیشتر الگوریتم‌های که برای درختان تصمیم یادگیری توسعه‌یافته‌اند از یک الگوریتم پایه مشتق شده‌اند که یک جستجوی حریصانه‌ی بالا به پایین را در فضای درختان تصمیم ممکن، بکار می‌گیرد. این روش توسط الگوریتم ID3 و نسخه‌ی کامل‌تر آن C4.5 نشان داده می‌شود. الگوریتم C4.5 نسل بعدی الگوریتم ID3 است و از نوعی از قانون هرس بعدی استفاده می‌کند. همچنین قادر است صفات گسسته، صفات فاقد مقدار و داده‌های نویزی را استفاده کند. این الگوریتم بهترین صفت را با استفاده از معیار بی‌نظمی انتخاب می‌کند و به دلیل استفاده از عامل Gain Ratio قادر به به‌کارگیری صفات با مقادیر بسیار زیاد می‌باشد. حتی اگر هیچ خطایی در داده‌های آموزشی وجود

نداشته باشد هرس انجام می‌شود که باعث می‌شود درخت عام‌تر شده و کمتر به مجموعه‌ی آموزشی وابسته شود (Kemal Polat۲۰۱۱)

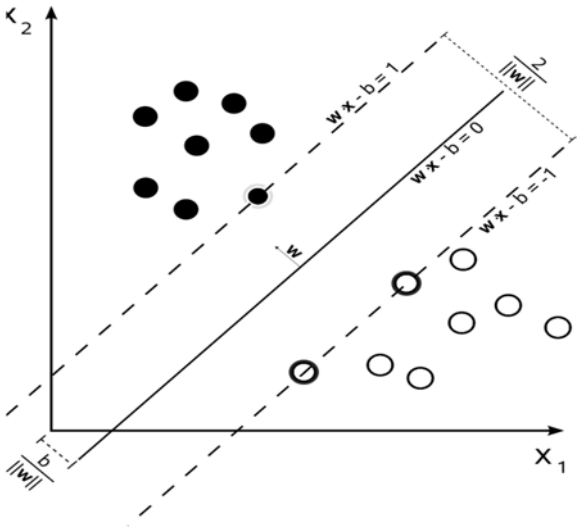
الگوریتم MLP شبکه‌های عصبی چندلایه پرسپترون شبکه‌های چندلایه‌ی پیش‌خور یکی از مهم‌ترین ساختارهای شبکه‌های عصبی مصنوعی می‌باشد. به‌طورمعمول این شبکه‌ها شامل مجموعه‌ای از واحدهای حسی (نورون‌های پایه) می‌باشند و یک‌لایه‌ی خروجی ۵، یک یا چندلایه‌ی پنهان ٤ که تشکیل‌دهنده‌ی لایه‌ی ورودی می‌باشند. ٦ سیگنال ورودی در خلال شبکه و در مسیری روبه‌جلو به‌صورت لایه به لایه منتشر می‌شود. این نوع شبکه معمولاً با عنوان پرسپترون چندلایه (MLP) نامیده می‌شود.

الگوریتم RBF شبکه‌های با تابع مدار شعاعی به‌طور گسترده‌ای برای تخمین غیر پارامتریک توابع چندبعدی از طریق مجموعه‌ای محدود از اطلاعات آموزشی به کار می‌رود. شبکه‌های عصبی شعاعی به‌واسطه آموزش سریع و فراگیر، بسیار جالب و مفید هستند و مورد توجه خاصی قرارگرفته‌اند. معماری اصلی RBF متشکل از یک شبکه سه لایه مانند شکل ۲ می‌باشد.



شکل ۲. ساختار سه لایه‌ای RBF

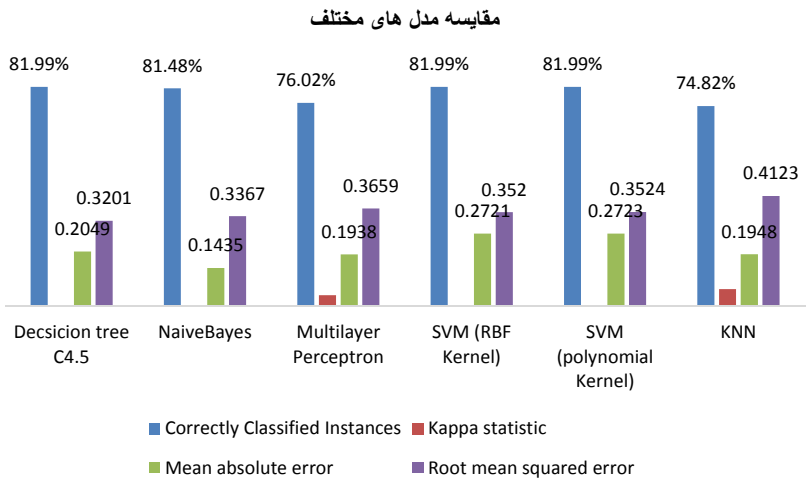
الگوریتم‌های ماشین‌های بردار پشتیبان عموماً برای مسائلی که در آن‌ها دو دسته وجود دارد، استفاده می‌شوند، اما روش‌های متفاوتی برای ترکیب چند SVM و ایجاد یک الگوریتم دسته‌بندی چند کلاس پیشنهاد شده است. در این الگوریتم دسته‌بندی دو صفحه، در مرز دو کلاس داده‌ها قرارگرفته می‌شود و مسئله یافتن مرز حداکثری بین این دو صفحه و درنتیجه بین دودسته داده‌ها می‌باشد. به این صورت که دو صفحه آن‌قدر از هم دور شوند که به داده‌ها برخورد کنند. همان‌طور که در شکل ۳ نشان داده‌شده است، هدف یافتن دوصفحه‌ای است که بیشترین فاصله را دارد و درنتیجه صفحه بین این دو صفحه بهترین جداکننده خواهند بود. (SavaşDurduran۲۰۱۰)



شکل ۳. صفحات ماشین بردار پشتیبان

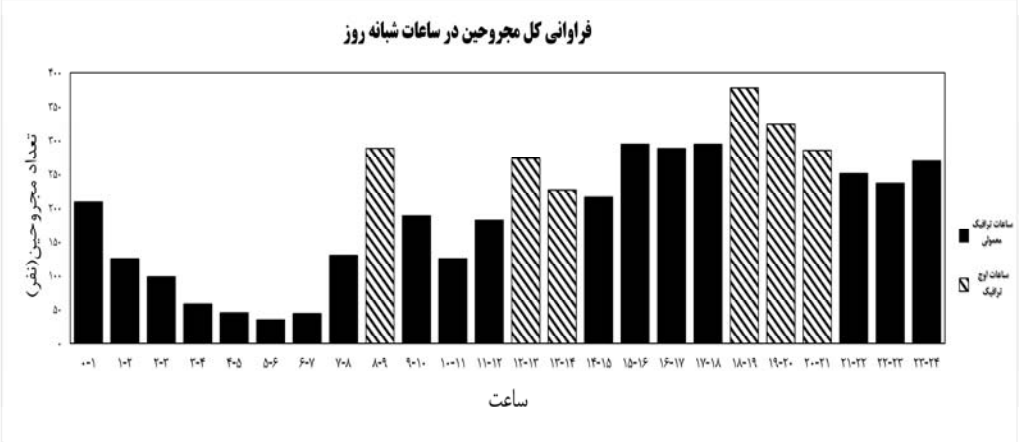
٤-یافته‌ها

با اعمال مدل‌های مختلف بر داده‌ها مشاهده می‌شود که داده‌های برداشت‌شده با دقت خوبی طبقه‌بندی می‌شوند؛ که در شکل ٤ مشاهده می‌شود که مدل‌های درخت تصمیم (C4.5) و ماشین بردارها (SVM) با درصد صحت زیادی داده‌ها را طبقه‌بندی می‌کنند.



شکل ٤. نتایج مدل‌سازی

شکل ۵ توزیع زمانی تصادفات مربوط به دو سال ۲۰۱۳ و ۲۰۱٤ در این استان نشان می‌دهد در ساعات اوج ترافیک تعداد مجروحین بسیار بیشتر می‌باشد که باید برای حل مشکلات ترافیکی باید اهمیت بیشتری به بررسی مشکلات ترافیکی در این ساعات پرداخت. (Al-Ghamdi۲۰۰۳، ۲۰۱٦، József Benedek)



شکل ۵. توزیع زمانی‌ءتصادفات

سپس این داده‌ها را در نرم‌افزار GIS وارد کرده و می‌توانیم نقاط حادثه‌خیز را در مشاهده کنیم. برای اینکه بتوانیم داده‌ها را از دیدگاه مکانی موردبررسی قرار دهیم ابتدا باید چند پارامتر را موردبررسی قرار دهیم که در قسمت زیر به آن می‌پردازیم. خودهمبستگی قوی زمانی رخ می‌دهد که مقادیر یک متغیر که ازنظر جغرافیایی به هم نزدیک هستند باهم مرتبط باشند. اگر عوارض و مقادیر متغیرهای مربوط به آن‌ها به‌صورت تصادفی در فضای توزیع‌شده باشند ظاهراً نباید بین آن‌ها ارتباطی وجود داشته باشد. آخرین ابزاری که در زمینه تحلیل الگوهای پراکنش و توزیع عوارض و پدیده‌ها در فضا و مکان توضیح داده می‌شود تحلیل خودهمبستگی فضایی که آماره موران (Moran's I) نیز معروف است، می‌باشد. آمار فضایی خود همبستگی فضایی یکی از کاربردی‌ترین و مهم‌ترین ابزارهای تحلیلی برای تحقیق در مورد داده‌های فضایی از. این تحلیل نه‌تنها به‌خودی‌خود

اطلاعات مفیدی در مورد ارتباط درونی عوارض به دست می‌دهد و نتایج آن برای بسیاری از تحلیل‌های پیچیده‌تر آماری نیز مورد استفاده قرار می‌گیرند. ابزار تحلیل خودهمبستگی فضایی موران به بررسی اخود همبستگی فضایی بر اساس مکان دو مقدار خصیصه موردنظر عوارض جغرافیایی می‌پردازد. فرض کنید تعدادی عوارض جغرافیایی با خصیصه مشخصی را در دست داریم، این ابزار نشان می‌دهد که الگوی پراکنش این عوارض با در نظر گرفتن مقدار خصیصه مورد مطالعه از الگوی خوشه‌ای و یا پراکنده برخوردار است. در ابزارهای قبلی فقط الگوی پراکنش عوارض و مقادیر خصایص آن‌ها مورد تحلیل قرار می‌گرفتند. این تحلیل الگوی توزیع عوارض در فضا را با ملاحظه هم‌زمان موقعیت مکانی و خصیصه مورد ارزیابی قرار می‌دهد. نتایج حاصل از این تحلیل نشان می‌دهد که عوارض به‌صورت تصادفی پراکنده و یا خوشه‌ای در فضا توزیع‌شده‌اند. این ابزار در حقیقت آماره و یا شاخص موران را محاسبه می‌کنند و با استفاده از امتیاز استاندارد Z و Pvalue ارزیابی معنادار بودن شاخص محاسبه‌شده می‌پردازد. شاخص موران برای خودهمبستگی فضایی به‌صورت زیر محاسبه می‌شود.

(۱)

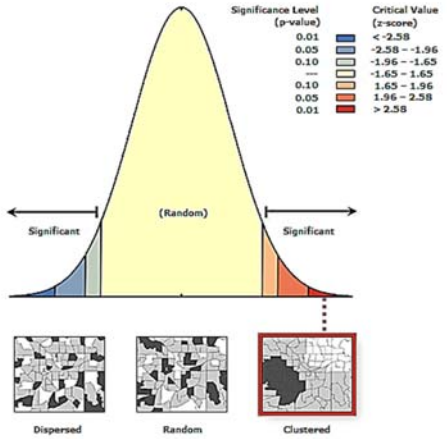
$$I=\frac{n\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{i}z_{i}z_{j}}{S_{0}\sum_{i=1}^{n}z_{i}^{2}}$$

در اینجا z_i تفاضل بین مقدار خصیصه عارضه i با میانگین آن $(x_i-\bar{x})$ می‌باشد. وزن فضایی بین عارضه i و j می‌باشد. n تعداد کل عوارض جغرافیایی موجود در لایه مورداستفاده بوده و S_0 جمع کل وزن‌های فضایی می‌باشد.

(۲)

$$S_0=\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{i,j}$$

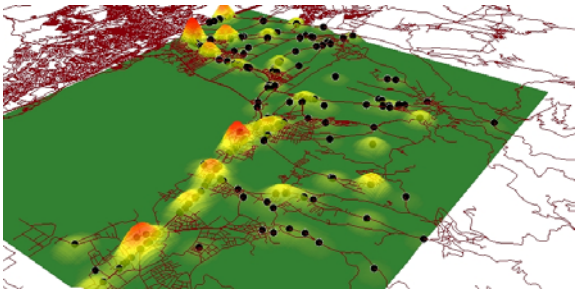
Moran's Index:	0.067422	
z-score:	2.904074	
p-value:	0.003683	



شکل ٦. الگوی خوشه‌ای

۵- نتیجه‌گیری

تحلیل خودهمبستگی فضایی به‌صورت گرافیکی و عددی ارائه می‌دهد. خروجی گرافیکی نشان می‌دهد که آیا داده‌ها پراکنده و یا خوشه‌بندی‌شده هستند به‌طورکلی اگر مقدار شاخص موران نزدیک به عدد مثبت یک باشد داده‌ها دارای خودهمبستگی فضایی و دارای الگوی خوشه‌ای بوده و اگر مقدار شاخص موران نزدیک به عدد منفی باشد آنگاه داده‌ها ازهم‌گسسته و پراکنده می‌باشند. البته این مقدار ازنظر معناداری آماری سنجیده می‌شود؛ (شکل شماره ٦) که در اینجا با توجه به عدد۰,۰٦۷ داده‌ها دارای خودهمبستگی فضایی می‌باشند و قابلیت خوشه‌بندی را دارند. (۱۰). در مورد این ابزار فرضیه صفر آن است که "هیچ خوشه‌بندی فضای بین مقادیر خصیصه‌ی مرتبط با عوارض جغرافیایی موردنظر وجود ندارد". حال زمانی که مقدار Pvalue بسیار کوچک و مقدار Z محاسبه‌شده (قدر مطلق آن) بسیار بزرگ باشد (خارج از محدوده اطمینان قرار گیرد) آنگاه می‌توان فرضیه صفر را رد کرد. اگر مقدار شاخص موران بزرگ‌تر از صفر باشد داده‌ها نوعی خوشه‌بندی فضایی را نشان می‌دهند. اگر مقدار شاخص کمتر از صفر باشد عوارض موردمطالعه دارای الگوی پراکنده می‌باشند. تصویرسازی نقاط حادثه‌خیز با استفاده از تابع چگالی کرنل انجام می‌شود (شکل ۷). با وزن دهی نقاط بر اساس تراکم نقطه‌ای هر تصادف و ارزش‌دهی نقاط بر اساس تعداد مجروحین نقاط حادثه‌خیز به‌صورت گرافیکی قابل‌مشاهده هستند (Seiji Hashimoto et al۲۰۱٦)



شکل ۷. نقاط حادثه‌خیز با استفاده از روش مبتنی برتراکم کرنل

در این تحقیق ما با استفاده از تحلیل‌های مکانی- زمانی و داده‌کاوی تصادفات درون‌شهری استان کوجیلی ترکیه را موردبررسی قراردادیم و با الگوریتم‌های مختلفی با دقت بالایی این نقاط را مدل کردیم. روش‌های داده‌کاوی نقش مهمی در تحلیل تصادفات دارند. با بررسی نقاط حادثه‌خیز پس از شناسایی به این نتیجه دست می‌توان یافت که نقاط حادثه‌خیز درون‌شهری بیشتر در نزدیکی تقاطعات بوده و اغلب در شرایط جوی مناسب و در روشنایی کافی روز اتفاق می‌افتند چرا که رانندگان در شرایط ایده‌آل با احتیاط کمتری رانندگی می‌کنند. این نشان می‌دهد که عامل انسانی تأثیر بیشتری در وقوع تصادفات دارد. این تحقیق مشخص کرده‌است که تصادفات بیشتر در شرایط جوی نامناسب و در روزهای با نور کافی رخ می‌دهد. این یافته نیز همخوانی خوبی با نتایج دیگر پژوهش‌ها دارد. به عنوان مثال، مطالعه‌ای که توسط سلطانی (۲۰۱۷) انجام شد، بیان می‌کند که شرایط جوی نامساعد به‌طور مستقیم با افزایش تصادفات مرتبط است. همچنین تحلیل‌های آماری آنها از داده‌های تصادفات تصدیق می‌کند که رانندگان در شرایط مطلوب، با احتیاط کمتری رانندگی می‌کنند. خاص، مقاله‌ای از راوی و همکاران (۲۰۱۸) نشان می‌دهد که انسان‌ها به دلیل درک نادرست شرایط و یا کاهش حواس جمع در زمان‌های خاص، اغلب منجر به تصادفات می‌شوند. این موضوع در نتایج شما نیز منعکس شده است و نشان‌دهنده این است که باید به آموزش و آگاهی رانندگان توجه بیشتری شود. استفاده از تحلیل‌های مکانی-زمانی و تکنیک‌های داده‌کاوی در این تحقیق، به‌عنوان یک رویکرد نوین در شناسایی و تحلیل نقاط حادثه‌خیز برجسته بوده است. این روش‌ها می‌توانند در شناسایی الگوهای تصادفی کمک زیادی کنند. پژوهش‌های متعددی برخی از این تکنیک‌ها را تأیید کرده‌اند، از جمله مطالعه‌ای که توسط دورمان (۲۰۱۰) انجام شده و به استفاده از تکنیک‌های یادگیری ماشین برای پیش‌بینی و شناسایی نقاط خطرناک می‌پردازد. نتایج به‌دست‌آمده از این

تحقیق و تحلیل‌ها، نشان می‌دهد که حوادث جاده‌ای تحت تأثیر عوامل مختلفی هستند که باید به آنها توجه ویژه‌ای کرد. در نهایت، پیشنهاد می‌شود که پژوهش‌های آینده بر روی عوامل انسانی و شرایط محیطی تمرکز بیشتری داشته باشند و از روش‌های نوین تحلیل داده بهره‌برداری کنند تا بتوانند به بهبود ایمنی جاده‌ها و کاهش تعداد تصادفات کمک کنند.

٦-مراجع

-Elvik, R., (2008). A survey of operational definitions of hazardous road locations in some European countries. *Accident Analysis & Prevention,* 40(6).1830-1835.

-Kemal Polat, S.S.D. Subtractive clustering attribute weighting (SCAW) to discriminate the traffic accidents on Konya–Afyonkarahisar highway in Turkey with the help of GIS: A case study. *Advances in Engineering Software,* 2011. 42(7). 491-500.

-V.Prasannakumar, H.V. R.Charutha, N.Geetha, (2011). Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment. *Procedia - Social and Behavioral Sciences,* 21, 317-325.

-Jafari M, Tabrizi A, Esmaeilzadeh MH, Zamani A. (2022). Determining of Accidental Points of Motorcycle Accidents Using Geographic Information System (GIS) in Gonabad and Bajestan Counties, North-East Iran. *Irtiqa Imini Pishgiri Masdumiyat.* 10(3), 262-268.

-Romi Satria, M.C. (2016). GIS Tools for Analyzing Accidents and Road Design. A Review. *Transportation Research Procedia*, 18. 242-247.

-Saffet Erdogan, I.Y. (2008). Tamer Baybura, Mevlut Gullu, eographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Accident Analysis & Prevention,* 40(1),174-181.

-SavaşDurduran, S. A (2010). decision making system to automatic recognize of traffic accidents on the basis of a GIS platform. *Expert Systems with Applications,* 37(12). 7729-7736.

-Al-Ghamdi, A.S. (2003). Analysis of traffic accidents at urban intersections in Riyadh. 717–724.

-József Benedek, S.M.C. (2016). Titus Cristian Man, Hotspots and social background of urban traffic crashes: A case study in Cluj-Napoca (Romania). *Accident Analysis & Prevention,* 87. 117-126.

-Ali Soltani, S.A. (2017). Exploring Spatial Autocorrelation of Traffic Crashes based on Severity, in Injury. 637-647.

-Seiji Hashimoto, S. RyokoSaeki, YasuhiroMimura, RyosukeAndo, ShutaroNanba, (2016). Development and application of traffic accident density estimation models using kernel density estimation. *Journal of Traffic and Transportation Engineering,* 3(3). 262-270.

Providing Intelligent Models for Traffic Accident Data Classification based on Bayesian Network, Decision Tree and Artificial Neural Network

Ali Naderan, Assistant Professor, Department of Civil Engineering, SR.C., Islamic Azad University, Tehran, Iran.

Mohammad Reza Ahadi, Associate Professor, Transportation Research Institute, Road, Housing and Urban Development Research Center.

Rasoul Zabihian, Ph.D., Candidate, Department of Civil Engineering, Kish International Branch, Islamic Azad University, Kish Island, Iran.

E-mail: r.zabihian@iau.ir

Received: May 2025- Accepted: August 2025

ABSTRACT

Every year, a large number of road users die or are injured as a result of traffic accidents. Analyzing accident-prone points for driving, transportation and traffic organizations, municipalities and organizations related to traffic safety is very important. In order to identify high-risk points of accidents, methods such as nearest neighbor are used, but these methods have insufficient prediction and incorrect estimation percentage. The purpose of this research is to classify accident data and provide a better method for identifying high-risk points in terms of accidents. For this purpose, the performance of decision trees, Bayesian model, and neural network were compared using common methods such as nearest neighbor and fitting. Using GIS tools and density-based methods such as kernel density function, high-risk points of accidents were identified in the form of functional maps. The results showed that decision tree method, Bayesian network and neural network have better performance than traditional methods with 83%, 83% and 78% percentage of correct estimation, respectively. Also, the results showed that the number of accidents in peak traffic periods, for example by examining the accident-prone points after identification, it can be concluded that the inner-city accident-prone points are mostly near intersections and often occur in suitable weather conditions and in sufficient daylight.

Keyword: Urban Accidents, Classification, GIS, Bayesian Network, Decision Tree and Artificial Neural Network