

تحلیل و ارزیابی شمارش تردد وسایل نقلیه در دوربین‌های پلاک خوان درون شهری توسط مدل‌های شبکه عصبی و جنگل تصادفی (مطالعه موردی: مشهد)

مقاله علمی - پژوهشی

محمد امین ابراهیم زاده، گروه مهندسی عمران، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
نوید نخعی، گروه مهندسی عمران، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
مهديسه دباغی، گروه مهندسی عمران، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
علی عبدی کردانی*، استاد، دانشکده عمران، گروه راه و ترابری و برنامه ریزی، حمل و نقل دانشگاه بین المللی امام خمینی (ره)،
قزوین، ایران

*پست الکترونیکی نویسنده مسئول: aliabdi@eng.ikiu.ac.ir

دریافت: ۱۴۰۳/۰۹/۰۱ - پذیرش: ۱۴۰۴/۰۳/۰۱

صفحه ۵۳۷-۵۵۴

چکیده

سیستم‌های حمل و نقل هوشمند با بکارگیری تکنولوژی‌های نوین از قبیل الکترونیک، ارتباطات و سیستم‌های کنترل رایانه‌ای و تلفیق آن‌ها با علوم مهندسی ترافیک و برنامه ریزی، ابزارهای لازم برای کنترل و مدیریت را فراهم ساخته‌اند. یکی از مهم‌ترین ابزارهای سیستم هوشمند دوربین‌های کنترل سرعت می‌باشند و از متداول‌ترین دوربین‌های موجود که در شبکه‌های شهری مورد استفاده قرار می‌گیرد، دوربین‌های پلاک خوان شهری است. در این پژوهش اطلاعات ترافیکی ثبت شده توسط دو دوربین به شماره‌های ۱۳، به آدرس مشهد - بزرگراه شهید کلانتری، درب شرقی دانشگاه فردوسی، جنوب - شرقی و دوربین شماره ۱۴، به آدرس مشهد، بزرگراه شهید کلانتری، بلوار صدا و سیما، شمال - غربی در طول یک هفته اول خردادماه ۱۳۹۹ می‌باشد و شامل اطلاعات پلاک خودرو، تخلف ترافیکی، زمان دقیق، سرعت و تعداد خودرو بوده است و این دو دوربین در مقایسه با دیگر دوربین‌های موجود در شهر مشهد دارای بیشترین خطا بوده‌اند. در این تحقیق در جهت شناسایی و کاهش خطاهای موجود، دو مدل کلاس‌بندی (جنگل تصادفی، شبکه عصبی) با توجه به پارامترهای موجود مورد بررسی و تحلیل قرار گرفته‌اند. در خطای کل موثرترین پارامتر سرعت بوده و دقت مدل شبکه عصبی فوق به ترتیب ۲۲/۰ و ۶۵ درصد به دست آمده است. در مدل جنگل تصادفی متغیر سرعت و حجم تردد بیشترین تأثیر را بر احتمال وقوع خطای کل در ثبت اطلاعات خودرو دارند. در خطای ثبت پلاک موثرترین پارامتر ساعت ثبت پلاک بوده و مدل جنگل تصادفی با دقت ۱۵ درصد بهترین کلاس‌بندی برای خطای مورد نظر ارائه شده است.

واژه‌های کلیدی: دوربین‌های پلاک خوان، مدل شبکه عصبی، مدل جنگل تصادفی، وسایل نقلیه

۱- مقدمه

گزارش اخیر سازمان ملل متحد، رشد جمعیت و شهرنشینی تا سال ۲۰۵۰ با افزایش ۲٫۹ میلیارد نفری روبه‌رو خواهد بود. این افزایش جمعیت باعث افزایش تعداد خودروها به ۲٫۵ میلیارد خواهد شد. (مجید گودرزی، ۱۳۹۸) یکی از کاربردهای فناوری

مدیریت هوشمند و نظارت بر ترافیک از الزامات اساسی توسعه شهرها محسوب می‌شود. گسترش الگوی شهرنشینی از سال ۱۹۵۰ به بعد و به دنبال آن افزایش شهرنشینی، وضعیت ترافیک در بسیاری از شهرها را چالش برانگیز شده است. براساس

در مدیریت شهری، سیستم‌های هوشمند حمل و نقلی است. سیستم‌های حمل و نقل هوشمند با بکارگیری تکنولوژی‌های نوین از قبیل الکترونیک، ارتباطات و سیستم‌های کنترل رایانه‌ای و تلفیق آن‌ها با علوم مهندسی ترافیک و برنامه‌ریزی، ابزارهای لازم برای کنترل و مدیریت را فراهم ساخته‌اند (Gaber et al, 2018). به طور کلی تعریفی از حمل و نقل هوشمند در سال ۱۹۹۸ توسط انجمن حمل و نقل هوشمند آمریکا ارائه شد که به صورت عام مورد قبول قرار گرفت که به بدین شرح است: مردم از تکنولوژی در حمل و نقل برای صرفه جویی در پول و وقت در زندگی روزمره استفاده می‌کنند. تعریف رسمی تری نیز توسط حمل و نقل آمریکا در آوریل سال ۱۹۹۹ نیز منتشر شد که بدین شرح است: سیستم حمل و نقل هوشمند اطلاعات مربوط به جا به جایی مسافر و کالا را جمع‌آوری، نگهداری، پردازش و توزیع می‌نماید (قاسمی نژاد، ۱۳۸۸). یکی از مهم‌ترین ابزارهای حمل و نقل هوشمند دوربین‌های کنترل ترافیک می‌باشند. این دوربین‌ها به عنوان یک سیستم نظارتی، کاربردهای متعددی در حمل و نقل دارند که شامل: کنترل ترافیک، ثبت تخلفات، نظارت بر رعایت قوانین راهنمایی و رانندگی از جمله کنترل سرعت، اندازه‌گیری زمان سفر می‌شود. استفاده از دوربین‌ها سبب افزایش ایمنی در سیستم حمل و نقل درون شهری می‌شوند. همچنین به کمک این دوربین‌ها می‌توان موضوعاتی از جمله شرایط جوی و گزارش تصادفات را نیز به ثبت رساند. یکی از متداولترین دوربین‌های موجود که در شبکه‌های شهری مورد استفاده قرار می‌گیرد، دوربین‌های پلاک‌خوان شهری است. این سامانه کاربردهای بسیاری در سامانه‌های نظارت بر ترافیک روزانه و سامانه‌های کنترل عوارض جاده‌ای دارند. این نوع سیستم‌ها تصاویر خودروها را ضبط کرده و با استفاده از الگوریتم‌های پیشرفته، شماره پلاک خودروها را با دقت بالا ثبت می‌نماید. سپس توسط نرم افزارهای مربوطه، پردازش داده می‌شود (عبداللهی، ۱۳۹۷). سیستم تشخیص شماره اتوماتیک شماره پلاک دوربین‌های پلاک‌خوان شهری شماره پلاک وسیله نقلیه را مشخص می‌کند. این سیستم در کاربردهای مختلفی از جمله مدیریت ترافیک، مدیریت پارکینگ، جمع‌آوری هزینه پارکینگ، گذرگاه‌های مرزی، یافتن موقعیت سرقت خودروها و غیره کمک می‌کند (Khinchi et al, 2019). استخراج آمار فراتر از میزان حقیقی آن توسط شمارنده‌های ترافیکی، منجر به تصمیم‌گیری اشتباه برای توسعه راه و تحمیل هزینه‌های زیاد برای

گسترش مسیر می‌شود و چنانچه آماری بسیار پایین‌تر از ترافیک واقعی محور ثبت شود، سبب کاهش سطح کیفیت ترافیک، کاهش ایمنی، افزایش هزینه‌های تعمیر و نگهداری خواهد شد. بنابر کاربردهای مختلفی که از این دوربین‌ها یاد شده است و نیاز روزافزون به این دوربین‌ها جهت کنترل ترافیک نیاز است که خطاهای این دوربین‌ها شناسایی و سپس میزان هر یک برآورد گردد.

این امر سبب می‌شود که داده‌های اولیه برای برنامه‌ریزی با دقت بیشتری برداشت گردد تا از خسارات جبران‌ناپذیر ناشی از برنامه‌ریزی صحیح جلوگیری شود (Homayounfar et al, 2011).

بنابراین پژوهش‌هایی در خصوص بررسی و شناسایی عوامل ایجادکننده خطا در دوربین‌های پلاک‌خوان شهری و نیز ارائه الگوریتم‌ها و روش‌هایی جهت تقلیل خطاهای موجود انجام شده‌اند که به طور مثال یکی از دلایل ایجاد خطا مکان نامناسب نصب آن‌ها است (تیموری و همکاران، ۲۰۱۹). طی پژوهشی یک مدل ریاضی ارائه کرده‌اند که با در نظر گرفتن امکان ردیابی وسایل نقلیه از طریق تشخیص پلاک و نیز حرکت آن‌ها در طول شبکه راه‌های ایران، بهترین نقاط را برای نصب دوربین‌های پلاک‌خوان وسایل نقلیه به دست آورد. این مهم از طریق معیارهایی نظیر تعداد دفعات تشخیص در مسیر حرکت نسبت به طول مسیر و همچنین حداکثر تعداد دوربین تشخیص دهنده جسم متحرک در شبکه مورد بررسی قرار گرفته است (میربها و همکاران، ۱۳۹۶). پژوهش دیگری نیز انجام داده شده است که در آن به ارائه یک مدل بهینه با هدف بیشینه سازی پوشش دوربین‌های کنترل سرعت در مسیر اقدام شده است. در این مقاله سعی بر آن شده که در یک مسیر حداقل تعداد دوربین‌ها برآورد گردد تا در حین به صرفه‌بودن حداکثر پوشش‌دهی مسیر را داشته باشیم.

در پژوهشی جهت ساماندهی و مدیریت بهتر پارکینگ‌ها اقدام به کنترل خودروها به کمک دوربین‌های پلاک‌خوان کرده‌اند. به کمک این سیستم توانستند خودروها را احراز کرده و فضای خالی داخل پارکینگ را شناسایی نمایند. در نتیجه این تحقیق دسترسی غیرقانونی به مناطق محدود، تهدید به سرقت وسایل نقلیه و تجهیزات داخل آن‌ها با این پیشنهاد به شدت کاهش یافت (Tan et al, 2013). از دیگر خطاهای می‌توان به از دست دادن سوژه تحت ردیابی اشاره نمود، که یک مشکل اساسی و شایع در سیستم دوربین‌های پلاک‌خوان شهری است. در پژوهشی با

ترافیکی شهرها را ضروری است. در این تحقیق به شناسایی خطاهای کل و خطای ثبت پلاک دوربین‌ها مورد بررسی قرار گرفته است.

۲- معرفی شهر مورد مطالعه

مشهد کلان‌شهری در شمال شرقی ایران و مرکز استان خراسان رضوی است. این شهر با ۳۵۱ کیلومتر مربع مساحت، دومین شهر پهناور ایران پس از تهران است. براساس سرشماری عمومی نفوس و مسکن سال ۱۳۹۵ مشهد با ۳,۰۰۱,۱۸۴ تن جمعیت، دومین شهر پرجمعیت ایران پس از تهران و نود و پنجمین شهر پر جمعیت دنیا به‌شمار می‌رود. این شهر به واسطه وجود حرم علی‌بن موسی الرضا(ع)، هشتمین امام مذهب شیعه، مهم‌ترین قطب زیارتی و کانون عمده گردشگری مذهبی و زیارت در ایران شناخته شده و سالانه پذیرای بیش از ۲۷ میلیون زائر از داخل و دو میلیون زائر از خارج از کشور است. با توجه به جمعیت فعلی و جمعیت زائر و مهاجر موجود در این شهر و به منظور مدیریت بهتر و کارآمد تر شهری به بررسی میزان خطا در برداشت داده‌های مورد استفاده در این تحقیق مربوط به اطلاعات ترافیکی ثبت شده توسط دو دوربین ترافیکی به شماره‌های ۱۳، به آدرس مشهد - بزرگراه شهید کلانتری، درب شرقی دانشگاه فردوسی، جنوب - شرقی و دوربین شماره ۱۴، به آدرس مشهد، بزرگراه شهید کلانتری، بلوار صدا و سیما، شمال - غربی در طول یک هفته اول خردادماه ۱۳۹۹ می‌باشد. موقعیت دوربین‌ها در شکل ۱ نشان داده شده است.

این پرسش که سوژه از کجا و چه موقع از دست رفته، مورد بررسی قرار گرفته است. در این تحقیق از الگوی تانسور که پسوند ماتریس است، برای مدل‌سازی داده‌های ترافیکی استفاده شده است. به دلیل آن که داده‌ها به طور تصادفی از دست می‌رود، الگوریتم مبتنی بر تانسور پیشنهاد شده است (Gunay et al, 2015). به کمک این الگوریتم یک مجموعه داده‌های سیستم دوربین‌های پلاک‌خوان شهری در داده‌های واقعی کنترل شده و با سایر روش‌های مدرن نیز مقایسه شده است. بنابراین با توجه به پژوهش‌های صورت گرفته در ایران و جهان و مشاهده نتایج حاصله و با توجه به آن که مدت زیادی از نصب و به کارگیری دوربین‌های پلاک‌خوان شهری در کنترل ترافیک مسیرها نمی‌گذرد، شناخت کامل و جامعی از نقاط قوت و به خصوص نقاط ضعف آن‌ها در دسترس نیست و در این زمان محدود فرصت کافی برای تحقیق و شناخت حاصل نشده است (Robinson et al, 2019).

لذا با توجه به نقش این دوربین‌ها در بدست آوردن اطلاعات صحیح ترافیکی و در نهایت مدیریت بهتر و کارآمدتر ترافیک در شهرها، از جمله سازمان‌هایی مانند: وزارت راه و شهرسازی، معاونت حمل و نقل کشور، شهرداری‌ها، سازمان‌های کنترل ترافیک شهرها، بهره برداران از سرمایه‌گذاری در پروژه‌های شهری و آزادراهی، راهنمایی و رانندگی، مهندسين مشاور، پیمانکاران حوزه راه و ترافیک و همچنین از آنجا که هنوز همه خطاها شناسایی نشده‌اند و میزان آنها مورد بررسی قرار نگرفته‌اند؛ انجام پژوهشی در مورد عملکرد این نوع از دوربین‌ها به منظور افزایش میزان دقت در شناخت صحیح از وضعیت



شکل ۱. جانمایی دوربین‌های مورد استفاده در این تحقیق خوانش دوربین‌های پلاک خوان شهری

۳-روش شناسی

در این پژوهش از دومدل شبکه عصبی و مدل جنگل تصادفی استفاده شده است که در زیر توضیحات آن ارائه شده است.

الف) شبکه عصبی

شبکه‌های عصبی روشی است که قصد دارد با استفاده از مدل‌های ریاضی و توان رایانه، برخی از جنبه‌های ساده مغز انسان را شبیه‌سازی کند. شبکه‌های عصبی به صورت یکی از بخش‌های پیچیده مغز انسان، به عنوان یک ساختار یادگیری غیر قابل درک، مشهور شده است. این ساختار پیچیده از مجموعه‌ای از نرون‌ها به وجود آمده است که این نرون‌ها ساختار ساده‌ای داشته، ولی شبکه اتصال آن‌ها وظایف یادگیری بسیار پیچیده‌ای را به انجام می‌رساند. لذا شناخت و درک ساختار بیولوژی مغز انسان می‌تواند ما را در ایجاد شبکه‌های عصبی مصنوعی به عنوان یک ابزار کارا در حل مسائل و کاربردهای علمی و شبکه‌های عصبی مصنوعی فنی یاری رساند. روش محاسبه‌ای شبکه‌های عصبی متفاوت با روش‌های متداول می‌باشد، که از روی عملکرد مغز و شبکه عصبی انسان‌ها الگوبرداری و مدل‌سازی شده است. محاسباتی که با روش‌های معمولی انجام می‌شود از نوع برنامه‌ریزی شده است و در آن‌ها الگوریتم‌ها و مجموعه‌هایی از قواعد به کار می‌روند تا مسئله را حل کنند. در این روش‌ها، اگر الگوریتمی در دست نباشد یا الگوریتم پیچیده باشد، راه حلی برای مسئله وجود نخواهد داشت. اما در محاسباتی که شبکه عصبی اجرا می‌کند به الگوریتم و مجموعه قواعد نیاز نیست. شبکه عصبی بر اساس یک سری داده‌ی آموزشی، نحوه‌ی پاسخ گویی را فرا می‌گیرد، آنگاه در مرحله عمل بر هر یک از ورودی‌ها حتی اگر جدید باشند، پاسخ مناسب می‌دهد. در مدل شبکه عصبی مصنوعی شبکه‌ای از عناصر پردازش ساده (نورون‌ها)، که می‌توانند رفتار پیچیده معینی را از ارتباط بین عناصر پردازش و پارامترهای عنصر نمایش دهند، اساس پردازش در این شبکه محسوب می‌شود. (واحدهای محاسباتی شبکه (پرسپترون‌ها) می‌توانند به دو شکل تک لایه و یا چند لایه عمل نمایند.

ب) جنگل تصادفی

جنگل تصادفی ۴ یک الگوریتم یادگیری تحت نظارت است که از آن هم برای طبقه بندی و هم رگرسیون استفاده می‌شود. اما به طور کلی برای مسائل طبقه بندی از آن استفاده می‌شود. همانطور که میدانیم یک جنگل از درختان ساخته شده است و درختان

بیشتر به معنای جنگل مقاوم تر است. به طور مشابه، الگوریتم درخت تصادفی روی نمونه‌های داده، درختان تصمیم گیری می‌سازد و سپس از هر کدام از آنها پیش بینی می‌گیرد و در نهایت به واسطه رای گیری، بهترین راه حل را انتخاب می‌کند. این یک روش گروهی است که از یک درخت تصمیم گیری مجزا بهتر است، زیرا با میانگین گیری در نتیجه، بیش از حد مناسب را کاهش می‌دهد. عملکرد الگوریتم جنگل تصادفی با کمک مراحل زیر می‌توانیم متوجه چگونگی عملکرد الگوریتم جنگل تصادفی شویم. مرحله ۱: ابتدا، از مجموعه داده فراهم شده نمونه‌های تصادفی را انتخاب کنید. مرحله ۲: سپس، این الگوریتم برای هر نمونه، یک درخت تصمیم‌گیری خواهد ساخت و در ادامه از هر درخت تصمیم‌گیری، نتیجه پیش بینی را خواهد گرفت. مرحله ۳: در این مرحله، برای هر نتیجه پیش بینی، رای گیری انجام می‌شود. مرحله ۴: در انتها، آن نتیجه پیش بینی که بیشترین تعداد رای را داشته باشد به عنوان نتیجه پیش بینی نهایی انتخاب می‌شود (Rigatti SJ, 2017).

ج) متدولوژی CRISP-DM

متدولوژی CRISP-DM معروفترین و متداولترین متدها برای انجام پروژه‌های داده‌کاوی می‌باشد. دلیل اصلی انتخاب این متدولوژی، علاوه بر قابل فهم بودن آن برای کاربران تازه‌کار، به استناد تحقیقات انجام شده توسط انجمن تحقیقات داده‌کاوی در سال ۲۰۰۷، این متدولوژی پرکاربردتر از سایر روش‌ها است. این متدولوژی یک مدل حلقه‌ای، تکرار شونده و تعاملی است؛ بدین معنی که برخی از مراحل آن ممکن است چندین بار اجرا شوند تا نتیجه مورد نظر در مدل‌سازی حاصل گردد. تعاملی از این جهت که مراحل مختلف در این مدل به هم وابستگی زیادی دارند و خروجی‌ها و ورودی‌های آنها کاملاً بر روی یکدیگر تأثیرگذار هستند و متغیرها و شاخص‌های خارج از مدل، عموماً تأثیری بر روی مدل اجرایی ندارند. در این مدل پس از مدل‌سازی و اتمام پروژه، دوباره باید با ورود داده‌های جدید مدل‌سازی را دوباره انجام داد تا در صورت بکارگیری نتایج، مدل‌ها با دقت و صحت بیشتری مدل‌سازی را در فضای واقعی انجام دهند. وجود چرخه‌های بیرونی نشان‌دهنده همین مساله است. یکی از قوی‌ترین متدولوژی‌ها در زمینه اجرا و پیاده‌سازی پروژه‌های داده‌کاوی CRISP است.

۳-۱- تحلیل خروجی‌ها

فرآیند داده‌کاوی نقشی اساسی در مدیریت این دوربین‌ها و افزایش کارایی آنها خواهد داشت. روزانه حجم عظیمی از اطلاعات و داده‌ها توسط دوربین‌های ترافیکی ثبت می‌شود که این اطلاعات نقشی حیاتی در تصمیم‌گیری‌ها، برنامه‌ریزی‌ها و شناسایی چالش‌های موجود شهری دارد. عدم استفاده از این اطلاعات، کارایی و نقش سازمان بزرگی چون سازمان ترافیک را در برنامه‌ریزی‌های شهری کاهش خواهد داد و سازمان دچار مشکل و بعضاً بحران خواهد شد. این مراحل به ترتیب برای ۱ خطای کل دوربین‌ها و خطای ثبت پلاک مورد بررسی قرار گرفته است که شامل:

پژوهش حاضر که با استفاده از متدولوژی CRISP-DM انجام گرفته است. باتوجه به اینکه هدف اصلی تحقیق حاضر، شناخت خطاهای موجود در عملکرد دوربین‌های ANPR و بررسی تأثیر متغیرهای مستقل نظیر: اطلاعات خودرو، زمان، سرعت خودرو و خط عبوری بر میزان خطای دوربین‌های ANPR می‌باشد، هر موضوعی مرتبط با دوربین‌های ترافیکی ANPR بهره‌مند از نتایج این تحقیق بوده و در دامنه اهداف تجاری-عملیاتی این طرح داده‌کاوی قرار می‌گیرد. شناسایی خطاهای موجود، عوامل موثر بر این خطاها، کشف ارتباط بین متغیرهای ممکن و بررسی الگوی رفتاری داده‌های ثبتی دوربین‌های مذکور از طریق انجام

Y TOTAL : خطای کل دوربین‌ها

Y1 : خطای ثبت پلاک

در ادامه نتایج دو الگوریتم کلاس‌بندی (جنگل تصادفی، شبکه عصبی) به تفصیل گزارش گردیده و بر اساس معیارهای ارزیابی بهترین مدل انتخاب می‌شود. متغیرهای مستقل مورد مطالعه در این مدل‌ها که انتظار می‌رود بر خطای دوربین‌های ترافیکی، به عنوان تابعی از این متغیرها، تأثیرگذار باشند، عبارتند از:

X₁: شماره دوربین،

X₂: روز هفته،

X₃: ساعت حرکت خودرو،

X₄: سرعت حرکت خودرو،

X₅: خط عبوری،

X₆: تراکم یا حجم خودرو،

در تمامی مدل‌ها نسبت داده‌های آموزشی و آزمایشی به ترتیب ۰/۷۰ و ۰/۳۰ است.

۴- نتایج مدل‌سازی

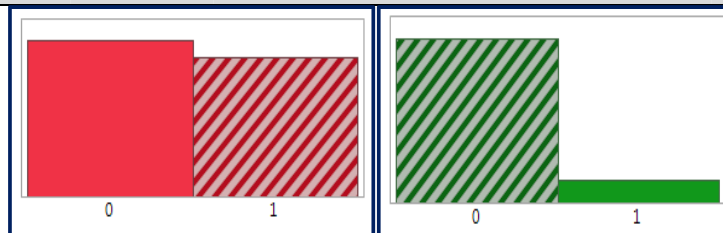
۴-۱- بررسی میزان تأثیر زمان و سایر متغیرهای ترافیکی بر خطای دوربین‌ها

انواع خطاهای خطای کل و ثبت پلاک) و سپس به تفکیک هر یک از خطاها با استفاده از الگوریتم SMOTE متوازن می‌گردد. داده‌های متوازنی که بر مبنای متغیر خطای کل و با استفاده از الگوریتم SMOTE ساخته و تولید گردید، شامل ۷۹۴۳۴۱ رکورد می‌باشد.

مرحله اول) متوازن کردن داده: متغیرهای زمان شامل روزهای هفته و ساعات مختلف شبانه‌روز است. در این بخش تأثیر متغیرهای زمانی و سایر متغیرها شامل سرعت، حجم خودرو در ساعت، شماره دوربین، خطوط عبوری خودرو بر خطای دوربین‌ها بررسی می‌شود. برای این منظور ابتدا داده‌های غیر متوازن بر اساس خطای کل در ثبت اطلاعات (مشاهده هریک از

جدول ۱. توزیع فراوانی خطای کل ثبت اطلاعات خودرو توسط دوربین‌های ترافیکی در داده‌های متوازن

خطا	کد	تعداد	درصد
ثبت کامل اطلاعات خودرو	عدم خطا=۰	۶۹۴۷۵۰	۸۸
	خطا=۱	۹۳۴۵۲	۱۲
جمع			
ثبت کامل اطلاعات خودرو	عدم خطا=۰	۴۲۰۵۳۳	۵۳
	خطا=۱	۳۷۳۸۰۸	۴۷
جمع			
		۷۹۴۳۴۱	۱۰۰



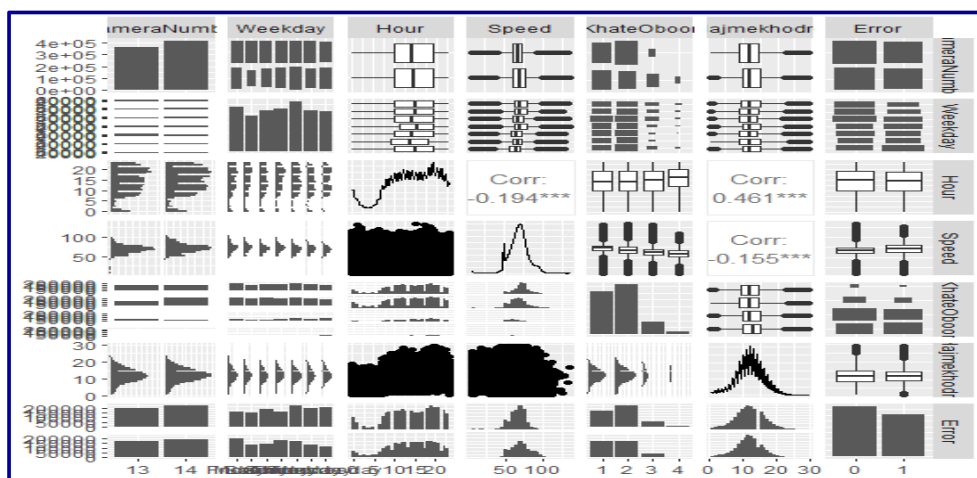
شکل ۲. نمودار ستونی توزیع فراوانی خطای کل ثبت اطلاعات خودرو توسط دوربین‌های ترافیکی

در داده‌های متوازن (سمت راست) و غیر متوازن (سمت چپ)

مرحله دوم) تعریف متغیرها: ابتدا در ماتریس نمودار پراکنش

که در شکل شماره ۳ ارائه گردیده است، رابطه بین هر دو متغیر مذکور مورد مطالعه قرار می‌گیرد. با توجه به اینکه در ۱۶۱۴۱ رکورد از داده‌های متوازن خطای ثبت خط عبور اتفاق افتاده و دقیقاً در تمامی این موارد یکی دیگر از سایر انواع خطا نیز به وقوع پیوسته است، لذا برای اجتناب از ایجاد اربابی در نتایج مدل‌های برازش شده روی داده‌ها این تعداد رکورد از بانک داده حذف می‌گردد.

جنگل تصادفی^۴: جنگل تصادفی با هدف رتبه‌بندی ویژگی‌های مورد بررسی در مدل کلاس‌بندی برای متغیر خطای کل (هرگونه خطای دوربین‌های ثبت اطلاعات خودرو) استفاده شده است، که ماتریس انطباقی آن به همراه نمودار ضرایب کاهش جینی^۵ در ادامه گزارش می‌گردد.

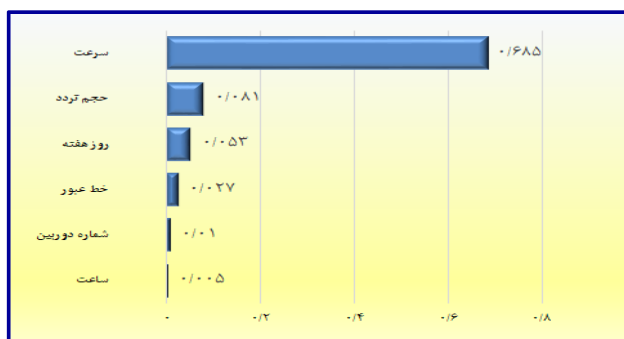


شکل ۳. ماتریس نمودار پراکنش متغیرهای حاضر در مدل‌های کلاس‌بندی مورد استفاده برای متغیر خطا در ثبت اطلاعات خودرو

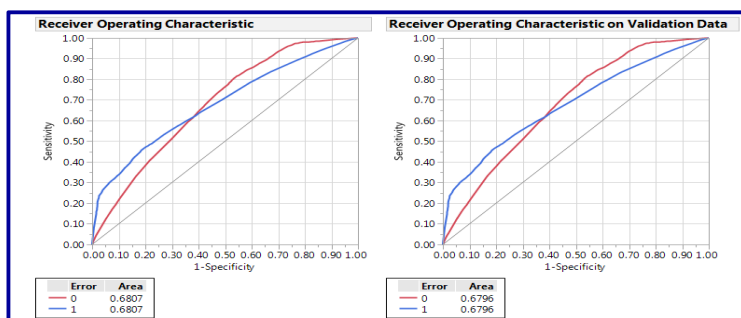
جدول ۲. ماتریس انطباقی داده‌های آزمایشی و آموزشی جنگل تصادفی برای متغیر خطای کل

واقعی		ماتریس انطباقی داده‌های آزمایشی		واقعی		ماتریس انطباقی داده‌های آموزشی	
خطا	عدم خطا	خطا	پیش‌بینی شده	خطا	عدم خطا	خطا	پیش‌بینی شده
۳۴۸۹۰	۱۰۹۶۵	خطا	پیش‌بینی شده	۸۰۸۰۶	۲۵۶۰۷	خطا	پیش‌بینی شده
۷۲۹۱۱	۱۱۴۹۷۴	عدم خطا	پیش‌بینی شده	۱۶۹۰۵۹	۲۶۸۹۸۸	عدم خطا	پیش‌بینی شده
Accuracy_Test = 64.1%				Accuracy_Train = 64.2%			
MSE = 0.219				MSE = 0.219			
Misclassification Rate = 0.359				Misclassification Rate = 0.358			

همان‌طور که شکل شماره ۴ نشان می‌دهد، متغیر سرعت و حجم تردد بیشترین تأثیر را بر احتمال وقوع خطای کل در ثبت اطلاعات خودرو دارند.



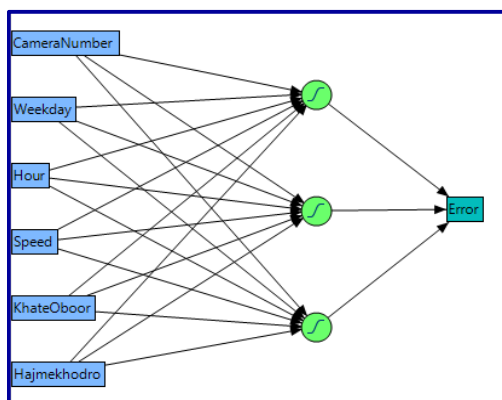
شکل ۴. نمودار میله‌ای میزان اهمیت متغیرهای مدل کلاس‌بندی جنگل تصادفی برای متغیر خطای کل در ثبت اطلاعات خودرو



شکل ۵. منحنی‌های ROC برای داده‌های آموزشی و آزمایشی در مدل جنگل تصادفی برای متغیر خطای کل

شبکه عصبی می‌باشد. مدل شبکه عصبی و ضرایب آن در ادامه گزارش می‌گردد.

د) شبکه عصبی: یکی دیگر از الگوریتم‌های مورد استفاده برای کلاس‌بندی متغیر خطای کل ثبت اطلاعات خودرو الگوریتم



شکل ۶. مدل شبکه عصبی برای متغیر خطای کل ثبت اطلاعات خودرو به همراه ضرایب

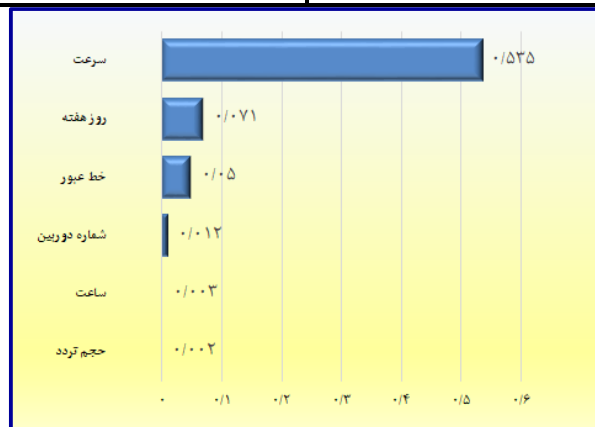
همان‌طور که شکل شماره ۷ نشان می‌دهد، متغیر سرعت و روز هفته بیشترین تأثیر را بر وقوع هرگونه خطا در ثبت اطلاعات خودرو دارد.

خطای MSE و دقت مدل شبکه عصبی فوق به ترتیب $0/22$ و 65 درصد به دست آمده است. در شکل شماره ۶ میزان اهمیت هر یک از متغیرهای ورودی مدل به تصویر کشیده شده است.

جدول ۳. ماتریس انطباقی داده‌های آزمایشی و آموزشی الگوریتم شبکه عصبی برای خطای کل

واقعی		ماتریس انطباقی داده‌های آزمایشی		واقعی		ماتریس انطباقی داده‌های آموزشی	
خطا	عدم خطا	خطا	پیش‌بینی شده	خطا	عدم خطا	خطا	پیش‌بینی شده
۲۴۷۲۰	۵۰۹۴۲	خطا	پیش‌بینی شده	۵۰۱۱۳	۱۰۱۷۲۹	خطا	پیش‌بینی شده
۱۱۵۴۴۴	۶۸۲۶۹	عدم خطا	پیش‌بینی شده	۲۳۰۲۵۶	۱۳۶۶۶۳	عدم خطا	پیش‌بینی شده

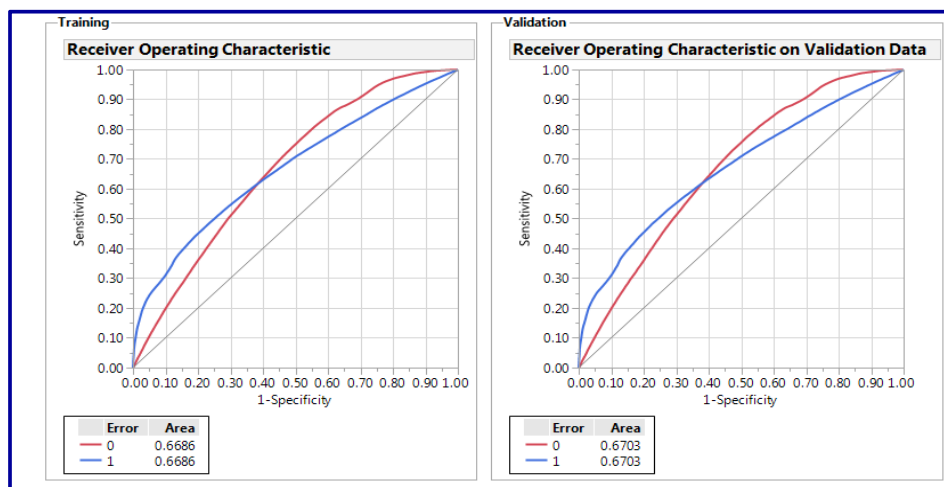
Accuracy_Test = 64.1%	Accuracy_Train = 64.0%
MSE = 0.221	MSE = 0.222



شکل ۷. میزان اهمیت متغیرهای ورودی مدل شبکه عصبی برای خطای کل در ثبت اطلاعات خودرو

مدل سطح زیر منحنی نمودار ROC است که برای داده‌های آموزشی و آزمایشی و آزمایشی مدل شبکه عصبی برابر با ۰/۶۷ به دست آمده، که تا مقدار یک نسبتاً فاصله زیادی دارد.

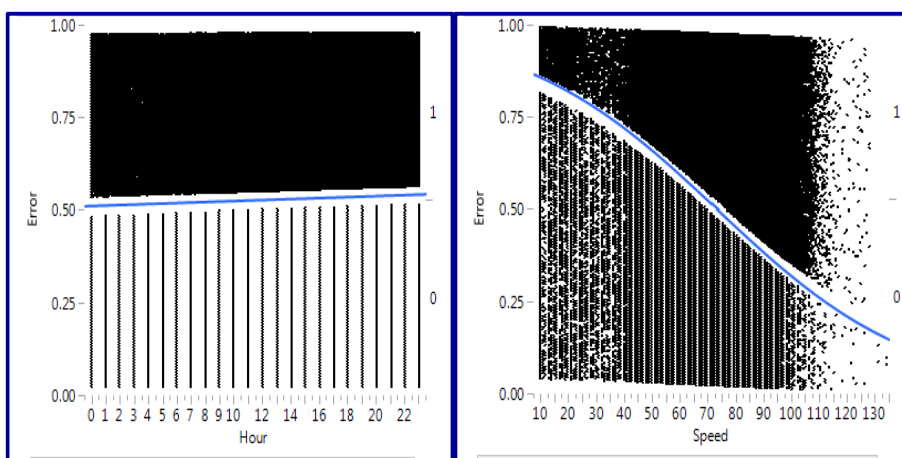
شایان ذکر است که میزان دقت و خطا در داده‌های آموزشی و داده‌های آزمایشی شبکه عصبی نیز مانند سایر مدل‌های پیشین تقریباً یکسان به دست آمد. یکی دیگر از معیارهای مناسب

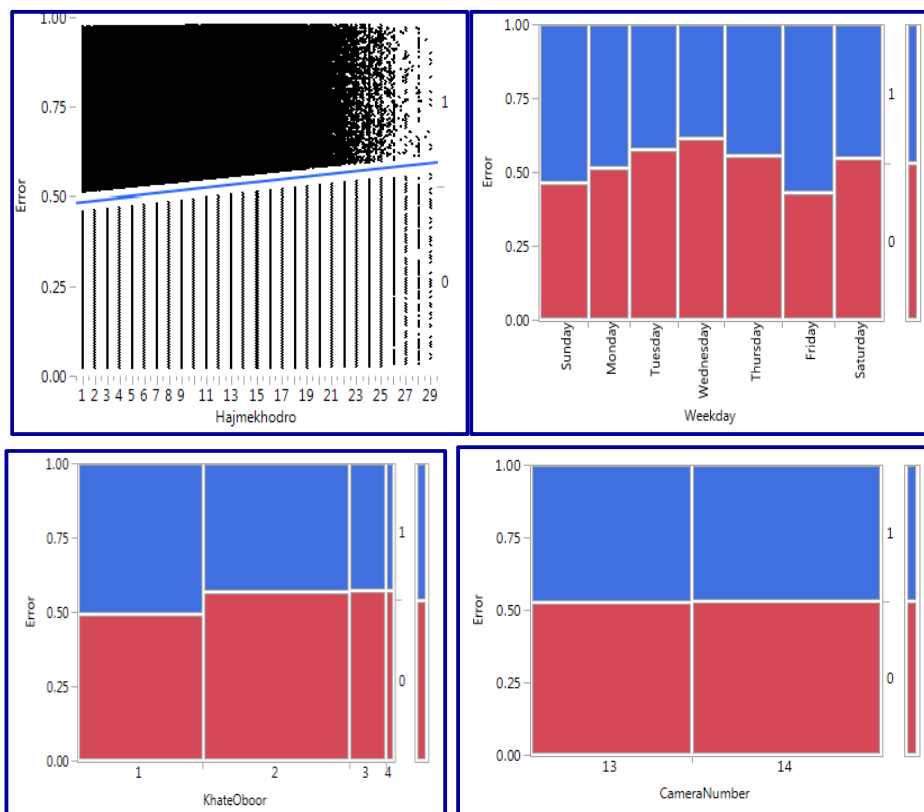


شکل ۸. منحنی‌های ROC برای داده‌های آموزشی و آزمایشی در مدل جنگل تصادفی برای متغیر خطای کل

خودرو دارد. برای مقایسه بهتر میزان تأثیر متغیرها بر احتمال وقوع خطای کل یا به عبارتی میزان اهمیت حضور متغیرها در مدل از مدل‌هایی چون جنگل تصادفی و شبکه عصبی استفاده می‌کنیم که به راحتی اندازه اهمیت متغیرها را محاسبه می‌کند.

مرحله چهارم) ارزیابی و انتخاب مدل مناسب: رتبه‌بندی میزان اهمیت متغیرها در الگوریتم‌های جنگل تصادفی و شبکه عصبی تقریباً یکسان است و در هر دو الگوریتم متغیر سرعت بیشترین تأثیر را بر احتمال وقوع هر یک از خطاهای ثبت اطلاعات





شکل ۹. نمودار پراکنش رابطه بین خطای کل در ثبت اطلاعات خودرو و متغیرهای تحقیق

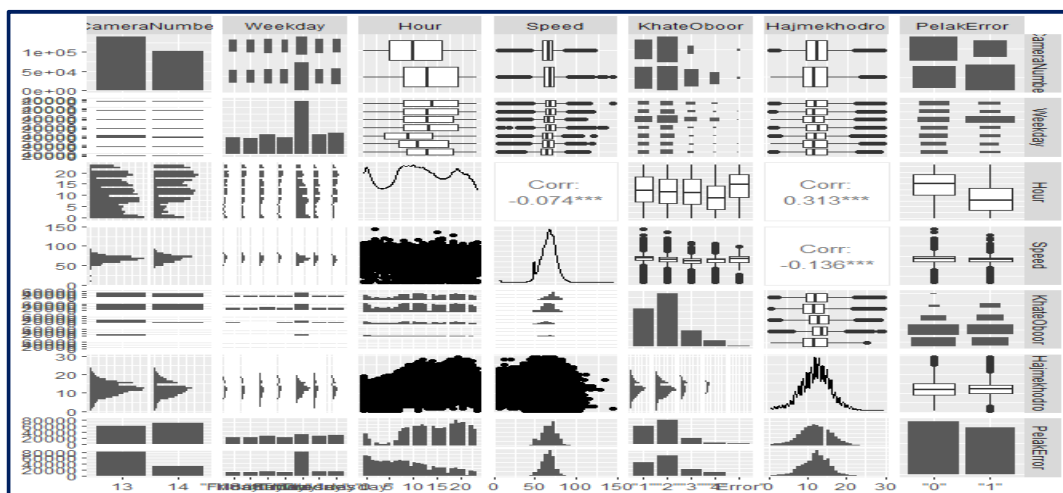
۲-۴- بررسی میزان تأثیر زمان و سایر متغیرهای ترافیکی بر خطای ثبت پلاک

مرحله اول) متوازن کردن داده‌ها: داده‌های متوازنی که بر مبنای متغیر خطای ثبت پلاک و با استفاده از الگوریتم SMOTE ساخته و تولید گردید، شامل ۲۴۲۹۳۸ رکورد می‌باشد.

جدول ۴. توزیع فراوانی خطای ثبت پلاک خودرو

خطا	کد	تعداد	درصد
ثبت پلاک خودرو	عدم خطا=۰	۱۲۸۶۱۴	۵۳
عدم ثبت پلاک خودرو	خطا=۱	۱۱۴۳۲۴	۴۷
جمع		۲۴۲۹۳۸	۱۰۰

مرحله دوم) تعریف متغیرها: ابتدا در ماتریس نمودار پراکنش که در شکل ذیل ارائه گردیده است، رابطه بین هر دو متغیر مذکور مورد مطالعه قرار می‌گیرد.



شکل ۱۰. ماتریس نمودار پراکنش متغیرهای حاضر در مدل‌های کلاس‌بندی مورد استفاده برای متغیر ثبت پلاک

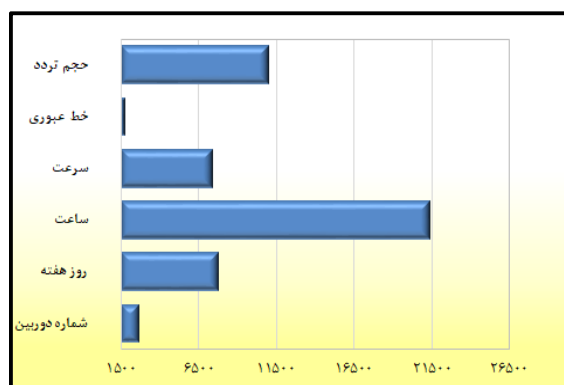
مرحله سوم) ورود اطلاعات به مدل‌ها و تحلیل نتایج: نتایج دو الگوریتم کلاس‌بندی (جنگل تصادفی، شبکه عصبی) مورد بررسی قرار گرفته است.

ج) جنگل تصادفی: جنگل تصادفی با هدف رتبه‌بندی ویژگی‌های مورد بررسی در مدل کلاس‌بندی برای متغیر خطای ثبت پلاک استفاده شده است، که ماتریس انطباقی آن به همراه نمودار ضرایب کاهش جینی در ادامه گزارش می‌گردد.

جدول ۵. ماتریس انطباقی داده‌های آزمایشی جنگل تصادفی برای متغیر خطای ثبت پلاک

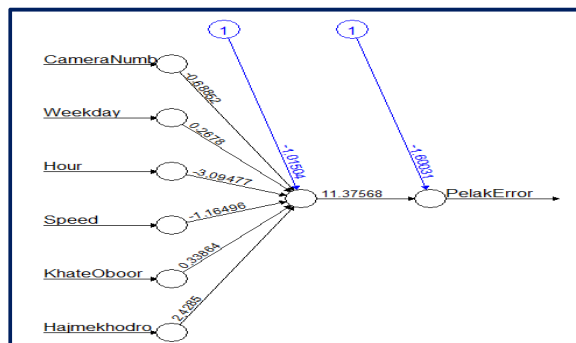
واقعی		ماتریس انطباقی داده‌های آزمایشی	
عدم خطا	خطا	خطا	پیش‌بینی شده
۲۱۵۲۴	۱۶۷۵۴	خطا	پیش‌بینی شده
۵۱۵۰۳	۳۲۷۶	عدم خطا	
$Accuracy = 0.854$ & $MSE = 0.151$			

همان‌طور که شکل شماره ۱۱ نشان می‌دهد، متغیر ساعت بیشترین تأثیر را بر احتمال وقوع خطای ثبت پلاک دارد.



شکل ۱۱. نمودار میله‌ای میزان اهمیت متغیرهای مدل کلاس‌بندی جنگل تصادفی برای متغیر خطای ثبت پلاک

د) شبکه عصبی: یکی دیگر از الگوریتم‌های مورد استفاده برای کلاس‌بندی متغیر خطای ثبت پلاک الگوریتم شبکه عصبی می‌باشد. مدل شبکه عصبی و ضرایب آن در ادامه گزارش می‌گردد.



شکل ۱۲. مدل شبکه عصبی برای متغیر خطای ثبت پلاک به همراه ضرایب

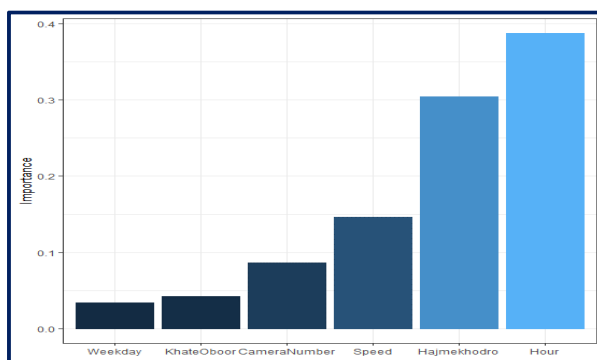
همان‌طور که شکل شماره ۱۳ نشان می‌دهد، متغیر ساعت و حجم تردد خودرو بیشترین تأثیر را بر وقوع خطای ثبت پلاک دارد. شایان ذکر است که میزان دقت و خطا در داده‌های آموزشی و داده‌های آزمایشی شبکه عصبی نیز مانند سایر مدل‌های پیشین تقریباً یکسان به دست آمد، لذا در ادامه تنها داده‌های آزمایشی را گزارش می‌کنیم.

خطای MSE و دقت مدل شبکه عصبی فوق به ترتیب ۰/۱۹۲ و ۷۱ درصد به دست آمده است. در شکل شماره ۱۲ میزان اهمیت هر یک از متغیرهای ورودی مدل به تصویر کشیده شده است.

جدول ۶. ماتریس انطباقی داده‌های آزمایشی الگوریتم شبکه عصبی برای خطای ثبت پلاک

واقعی		ماتریس انطباقی داده‌های آزمایشی	
عدم خطا	خطا		
۶۱۶۱	۲۰۰۱۴	خطا	پیش‌بینی شده
۳۱۷۲۳	۱۴۲۸۳	عدم خطا	

$$Accuracy = 0.710 \ \& \ MSE = 0.192$$

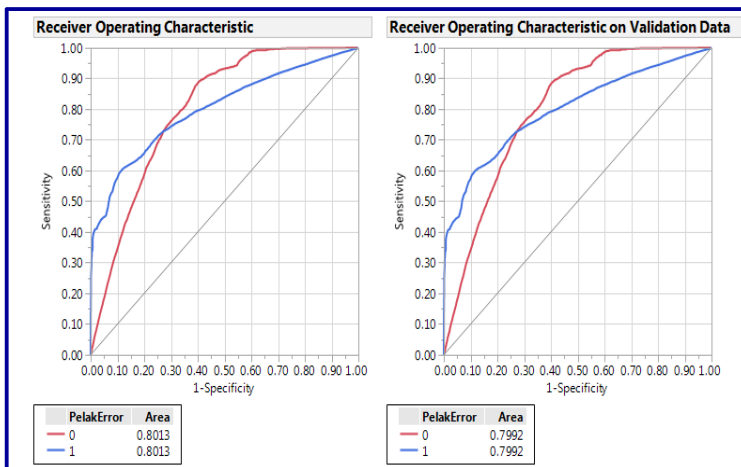


شکل ۱۳. میزان اهمیت متغیرهای ورودی مدل شبکه عصبی برای خطای ثبت پلاک

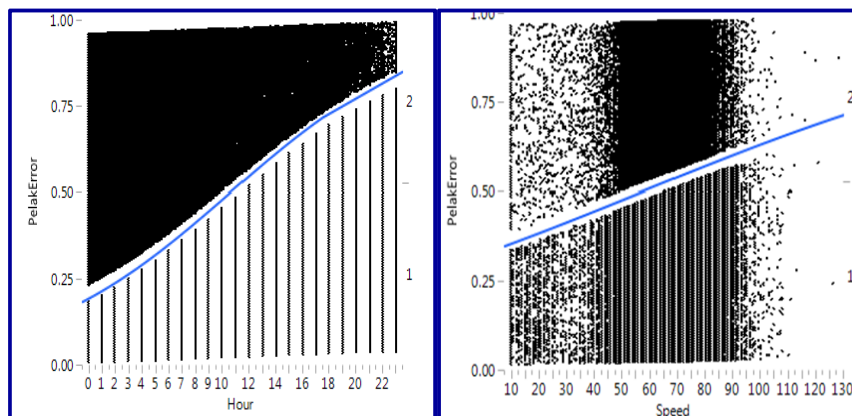
مرحله چهارم) ارزیابی و انتخاب مدل مناسب

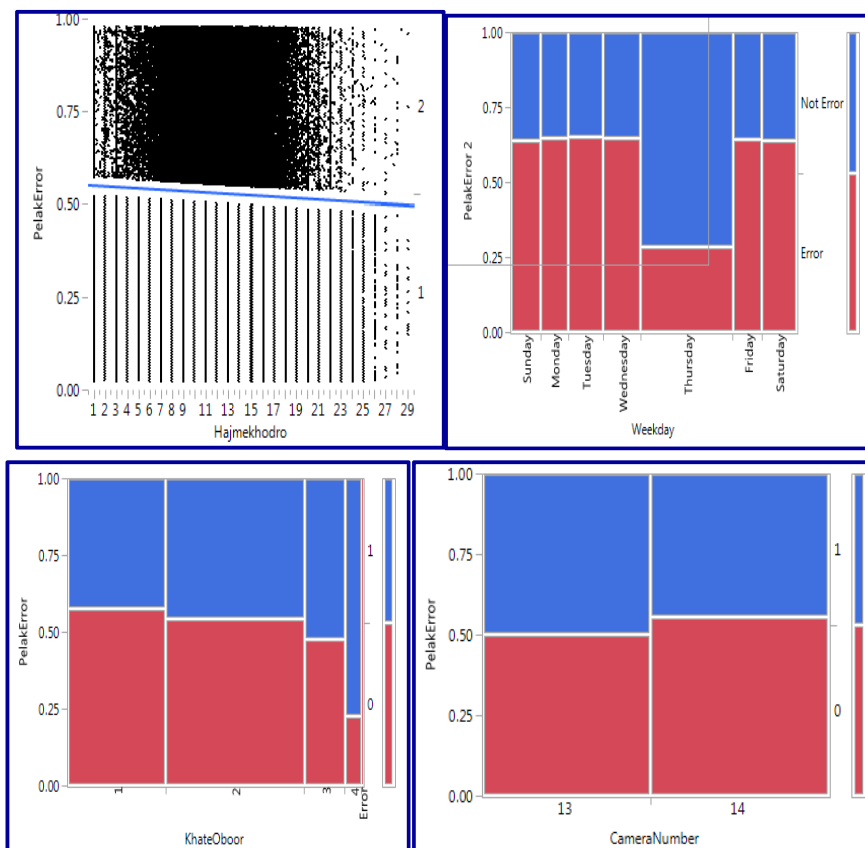
با توجه به میزان دقت و خطای مدل‌های مورد بررسی، جنگل تصادفی با دقت ۸۵ درصد بهترین کلاس‌بند برای خطای ثبت پلاک را ارائه می‌دهد. شایان ذکر است که رتبه‌بندی میزان اهمیت متغیرها در الگوریتم‌های جنگل تصادفی و شبکه عصبی یکسان است.

همان‌طور که شکل ۱۴ نشان می‌دهد، مساحت زیر منحنی‌ها تقریباً ۸۰ درصد است که نشان می‌دهد، مدل جنگل تصادفی ساخته شده برای خطای ثبت پلاک مناسب و دقیق است. در ادامه روابط دو به دوی متغیرها بر مبنای شکل شماره ۱۵ نمودار پراکنش بررسی می‌شود.



شکل ۱۴. منحنی‌های ROC برای داده‌های آموزشی و آزمایشی در مدل جنگل تصادفی برای متغیر خطای ثبت پلاک





شکل ۱۵. نمودار پراکنش رابطه بین خطا در ثبت پلاک و متغیرهای تحقیق

۵- نتیجه‌گیری

داشته است از جمله: شماره دوربین، روز هفته، ساعت حرکت خودرو، سرعت حرکت خودرو، خط عبوری، تراکم یا حجم خودرو روند مدل‌سازی از این متغیرها استفاده شد. رتبه‌بندی میزان اهمیت متغیرها در الگوریتم‌های جنگل تصادفی و شبکه‌عصبی تقریباً یکسان است و در هر دو الگوریتم متغیر «سرعت» بیشترین تأثیر را بر احتمال وقوع خطای ثبت اطلاعات خودرو دارد. در مدل جنگل تصادفی پس از سرعت به ترتیب؛ حجم، روز هفته، خط عبور، شماره دوربین و ساعت ثبت تأثیرگذاراند. خطای ثبت پلاک در این خطا رتبه‌بندی میزان اهمیت متغیرها در الگوریتم‌های جنگل تصادفی و شبکه‌عصبی یکسان است و در هر دو الگوریتم متغیر «ساعت ثبت پلاک» بیشترین تأثیر را بر احتمال وقوع هر یک از خطای ثبت پلاک دارد. در هر دو مدل پس از ساعت ثبت پلاک به ترتیب؛ حجم، روز هفته، سرعت، شماره دوربین، خط عبور تأثیرگذاراند.

با توجه به اهمیت سنجش عملکرد دوربین‌های پلاک خوان شهری به منظور بررسی و شناخت صحیحی از وضعیت ترافیک شهرها، پژوهشی در این مورد انجام شده است. در این پژوهش ابتدا نمونه داده‌ها مورد بررسی و برداشت قرار گرفته‌اند. داده‌های مورد استفاده در این تحقیق مربوط به اطلاعات ترافیکی ثبت شده توسط دو دوربین ترافیکی به شماره‌های ۱۳، به آدرس مشهد - بزرگراه شهید کلاتری، درب شرقی دانشگاه فردوسی، جنوب - شرقی و دوربین شماره ۱۴، به آدرس مشهد، بزرگراه شهید کلاتری، بلوار صدا و سیما، شمال - غربی در طول یک هفته اول خردادماه ۱۳۹۹ می‌باشد. که شامل اطلاعات پلاک خودرو، تخلف ترافیکی، زمان دقیق، سرعت و تعداد خودرو است. این دو دوربین در مقایسه با دیگر دوربین‌های موجود در شهر مشهد دارای بیشترین خطا بوده‌اند. خطاهای شناسایی شده در این تحقیق شامل: خطای کل دوربین‌ها، خطای ثبت پلاک بوده است و عوامل مختلفی بر بروز این خطاها تأثیر مستقیمی

۶- پی‌نوشت‌ها

- 1- Automatic Number Plate Recognition
- 2- Intelligent Transportation System
- 3- Artificial Neural Network
- 4- Random Forest
- 5- Mean Decrease Gini
- 6- Scatterplot Matrix
- 7- Cross-Industry Standard Process for Data Mining

۷-مراجع

- number of instrumented vehicles. *International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 379-386.
- Homayounfar, Abdollah, Anthony, TS, Ho. Naida, Zhu. George, Head, and Phil, Palmer (2011). Multi-vehicle convoy analysis based on ANPR data.
- Khinchi, M. Agarwal, C. (2019). A review on automatic number plate recognition technology and methods. *International Conference on Intelligent Sustainable Systems*, 363-366.
- Patel C, Shah D, Patel A. (2013). Automatic number plate recognition system (anpr): A survey. *International Journal of Computer Applications*, 1, 69(9).
- Qadri, MT. Asif, M. (2009). Automatic number plate recognition system for vehicle identification using optical character recognition. *International Conference on Education Technology and Computer*, 17, 335-338.
- Rigatti, S.J. (2017). Random forest. *Journal of Insurance Medicine*, 1, 47(1), 31-39.
- Roy, A. Ghoshal DP. (2011). Number Plate Recognition for use in different countries using an improved segmentation. *National Conference on Emerging Trends and Applications in Computer Science*, 1-5.
- Soole D, Fleiter, J. Watson B. (2012). Point-to-point speed enforcement [Austroads Project No.SS1649]. Austroads Ltd.
- Tan, H. Feng, G. Feng, J. Wang, W. Zhang, YJ. Li, F. A. (2013). tensor-based method for
- گودرزی، مجید، فیروزی، محمدعلی و سعیدی، امید (۱۳۹۸). بررسی و تحلیل موانع هوشمندسازی حمل و نقل درون شهری اهواز. *جغرافیا و آمایش شهری - منطقه‌ای*.
- قاسمی‌نژاد، حسین (۱۳۸۸). سیستم‌های حمل و نقل هوشمند، کنفرانس ملی تصادفات و سوانح جاده ای و ریلی.
- عبداللهی، میترا و خسروی، حسین (۱۳۹۷). طراحی و پیاده‌سازی سامانه بی‌درنگ آشکارسازی و شناسایی پلاک خودرو در تصاویر ویدئویی، پردازش علائم و داده‌ها.
- تیموری، شیخ محمدی، حسین زاده کاشان و شجاعی، عباس (۲۰۱۹). توسعه مدل مکان‌یابی دوربین‌های پلاک خوان در جاده‌های ایران جهت پوشش مسیرهای حرکتی مبدأ-مقصد وسایل نقلیه باری. *فصلنامه مهندسی حمل و نقل*، ۱۱(۱)، ۹۷-۱۲۱.
- میربها، بابک و حسینی، سیدمهرداد (۲۰۱۷). ارزیابی مدلی برای مکان‌یابی بهینه دوربین‌های کنترل سرعت در راه‌ها.
- جعفری، ادریس و صمدیان، منیرالسادات (۲۰۱۲). کاربرد داده کاوی در بررسی رفتار رانندگان متخلف در کلان شهرها.
- Anand SS, Büchner AG. (1998). Decision support using data mining.
- Fayyad, U. Piatetsky-Shapiro G. Smyth P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 1, 39(11), 27-34.
- Gaber T. Abdelwahab, S. Elhoseny, M. Hassanien, AE. (2018). Trust-based secure clustering in WSN-based intelligent transportation systems. *Computer Networks*, 9, 146, 151-158.
- Gunay, B. Saltan, M. Ilgaz, A. Albayrak Y. (2015). Area-wide ANPR coverage with a small

- Ebrahimzadeh, M. A., Javanshir, H., & Esfandiyari, M. (2024). Analysis and Evaluation of Smart Ramp Metering System on Urban Traffic Flow Control (Case Study: Tehran-Hakim Highway). *Road*, 32(121), 219-234.
- Jashnian, A. H., Ebrahimzadeh, M. A., & Yaghoobii, S. (2024). Analysis and evaluation of hourly changes in travel consumption volume and the number of severe accidents in inner-city roads (case study of Qom city). *Road*.
- Ebrahimzadeh, M. A., Nakhaie, N., & Dabbaqi, M. (2024). Investigating the Influential Personality and Demographic Variables of the Behavior of Public Transport and Freight Drivers in the Factors of Accidents (Case Study: Karaj City). *Road*, 32(120), 375-388.
- missing traffic data completion. *Transportation Research Part C: Emerging Technologies*, 1, 28, 15-27.
- Robinson A. Venter, C. (2019). Validating traffic models using large-scale Automatic Number Plate Recognition (ANPR) data. *Journal of the South African Institution of Civil Engineering*, 61(3), 45-57.
- Ebrahimzadeh, M. A., Dabbaqi, M., & Nakhaie, N. (2024). Investigating the reduction of the error caused by the counting of vehicle traffic in city license plate reader cameras by mathematical models (case study: Mashhad city). *Road*.

Analysis and Evaluation of Vehicle Traffic Counting in inner-city Automatic Number Plate Recognition by Neural Network and Random Forest Models (Case Study: Mashhad-Metropolis)

Mohammad Amin Ebrahimzadeh, Ph.D., Candidate, Majoring in Transportation Planning, Faculty of Civil Engineering, Arts and Architecture, Tehran Science and Research, Tehran, Iran.

Navid Nakhaei, Ph.D., Candidate, Majoring in Transportation Planning, Faculty of Civil Engineering, Arts and Architecture, Tehran Science and Research, Tehran, Iran.

Mahdisheh Dabbaqi, Ph.D., Candidate, Majoring in Transportation Planning, Faculty of Civil Engineering, Arts and Architecture, Tehran Science and Research, Tehran, Iran.

Ali Abdi Kordani, Professor, Department of Civil Engineering, Faculty of Technical and Engineering, Imam Khomeini International University, Qazvin, Iran.

E-mail: aliabdi@eng.ikiu.ac.ir

Received: January 2024- Accepted: April 2025

ABSTRACT

Intelligent transportation systems (ITS) have provided the necessary tools for control and management by using new technologies such as electronics, communication and computer control systems and combining them with traffic engineering and planning sciences. One of the most important tools of the smart system are speed control cameras, and one of the most common cameras used in urban networks is the automatic number plate recognition (ANPR) cameras. Therefore, considering the importance of measuring the performance of Automatic Number Plate Recognition in order to investigate and correctly understand the traffic situation of cities, a research has been conducted to identify existing errors and then provide a mathematical model to reduce camera errors. In this research, traffic information recorded by two cameras number 13, at the address of Mashhad - Shahid Kalantari Highway, east gate of Ferdowsi University, south-east and camera number 14, at the address of Mashhad, Shahid Kalantari Highway, Seda & Sima Boulevard, north-west during the first week of June 2019 and included information on license plates, traffic violations, exact time, speed and number of cars, and these two cameras had the most errors compared to other cameras in Mashhad. In this research, in order to identify and reduce existing errors, two classification models (random forest, neural network) have been investigated and analyzed according to the available parameters. Speed is the most effective parameter in the total error and the accuracy of the above neural network model is 0.22% and 65% respectively. In the random forest model.

Keywords: Automatic Number Plate Recognition, Neural Network Model, Random Forest Model, Vehicles