پیش‌بینی حجم وسایل نقلیه ورودی به محدوده‌های ترافیکی با استفاده از یادگیری ماشین (نمونه موردی: شهر تهران)

مقاله علمی – پژوهشی

محمودرضا ناطقی، دانشجوی کارشناسی ارشد، دانشکده مهندسی عمران و محیط‌زیست دانشگاه تربیت‌مدرس، تهران، ایران
سپیده شامی، دانشجوی دکتری، دانشکده مهندسی عمران و محیط‌زیست دانشگاه تربیت‌مدرس، تهران، ایران
امیررضا ممدوحی *، دانشیار، دانشکده مهندسی عمران و محیط‌زیست دانشگاه تربیت‌مدرس، تهران، ایران
و استاد وابسته، دانشکده مهندسی عمران، زمین‌شناسی و معدن، دانشگاه فنی مونترال، کانادا
*پست الکترونیکی نویسنده مسئول: **armamdoohi@modares.ac.ir**

دریافت: ۱۴۰۳/۰۳/۲۰– پذیرش: ۱۴۰۳/۰۸/۲۰
صفحه ۵۰-۳۹

چکیده

پیش‌بینی حجم وسایل نقلیه به‌عنوان یکی از مؤلفه اصلی مدیریت ترافیک شناخته می‌شود و مدیریت آن، نقش مؤثری در بهبود کارایی ترافیک شبکه معابر شهری دارد. این پژوهش با استفاده از اطلاعات حجم وسایل نقلیه ورودی به محدوده‌های ترافیکی شهر تهران بین سال‌های ۱۳۹۶ الی ۱۴۰۰ و با بکارگیری از سه روش یادگیری ماشین (جنگل تصادفی، XGBoost و نزدیک‌ترین همسایه) به پیش‌بینی حجم وسایل نقلیه ورودی به هر یک از محدوده‌های ترافیکی (طرح ترافیک وکنترل آلودگی هوا) در گام زمانی یک ساعته پرداخته است. در این پژوهش علاوه براستفاده از متغیرهای متداول مانند مشاهدات تاریخی حجم وسایل نقلیه، داده‌های آب‌وهوا و زمان، از سیاست‌های مدیریت تقاضای سفر نیز به عنوان متغیرهای پیش‌بینی کننده استفاده کرده و با بررسی تحلیل حساسیت، سیاست‌های اثرگذارتر بر پیش‌بینی حجم وسایل نقلیه را مورد ارزیابی قرار داده است. نتایج نشان می‌دهد، مدل XGBoost و مدل جنگل تصادفی با میانگین درصد خطای مطلق ۸/۲٪ و ۱۱/۱۶٪ برای محدودهٔ طرح ترافیک و کنترل آلودگی هوا، عملکرد مناسب‌تری در مقایسه با مدل نزدیک‌ترین همسایه دارد. تحلیل حساسیت متغیرهای مربوط به سیاست‌های مدیریت تقاضای سفر نشان می‌دهد، سیاست‌های "غیرحضوری شدن مراکز آموزشی" و"اخذ عوارض محدودهٔ طرح ترافیک و کنترل آلودگی هوا "بیشترین تاثیر و سیاست" اعمال محدودیت تردد بین استانی" کمترین تاثیر را در افزایش دقت پیش‌بینی دارند.

واژه‌های کلیدی: پیش‌بینی، حجم وسایل نقلیه، مدیریت تقاضای سفر، یادگیری ماشین، XGBoost

۱- مقدمه

با افزایش جمعیت شهری، میزان استفاده از وسایل نقلیه رشد چشم‌گیری داشته است که منجر به افزایش بیش از حد حجم وسایل نقلیه نسبت به ظرفیت معابر، سطح سرویس نامطلوب و تراکم ترافیک در شبکه‌های آزادراهی، بزرگراهی و خیابان‌های شهری شده است. به همین جهت، پیش‌بینی جریان ترافیک که می‌تواند شامل حجم و یا سرعت وسایل نقلیه باشد، به تصمیم‌گیران اجازه می‌دهد تا تصمیماتی هدفمند و به‌موقع جهت رفع این معضل اتخاذ کنند. عوامل مختلفی می‌تواند بر جریان ترافیک اثرگذار باشد، متغیرهایی چون شرایط آب‌وهوا، اطلاعات تقویمی (شامل روزهای تعطیل و غیرکاری) و زمان (شامل فصل، ماه، روز و ساعت) از مواردی است که در پژوهش‌ها مورد بررسی قرار گرفته است(Razali et al., 2021). علاوه برآن

هر تغییر در شبکه یا سیاست‌های مدیریت تقاضای سفر که رفتار سفر شهروندان را تحت تأثیر قرار دهد می‌تواند بر حجم وسایل نقلیه در معابر شهری اثرگذار باشد (Shami & Mamdoohi, 2022). از سیاست‌های مدیریت تقاضای سفر تاثیر گذار بر حجم وسایل نقلیه می‌توان به سیاست قیمت‌گذاری محدوده‌های ترافیکی اشاره کرد. این سیاست در شهر تهران برای هردومحدودهٔ ترافیکی (طرح ترافیک و کنترل آلودگی هوا) به شیوه دوره مبنا از کاربران وسایل نقلیه بر اساس ساعت ورود و خروج (اوج و غیر اوج) صورت می‌گیرد. علاوه بر آن محدودیت‌های اعمال شده در دوره کووید ۱۹ نیز تاثیر مستقیمی بر حجم وسایل نقلیه و تعداد سفر افراد داشته است (Zhang et al., 2023). این محدودیت‌ها که در ایران شامل "غیرحضوری شدن مراکز آموزشی، تغییر در ساعت کاری ادارات، اعمال محدودیت فعالیت مشاغل آزاد و غیردولتی و اعمال محدودیت تردد بین استانی" بوده است را می‌توان بخشی از سیاست‌های مدیریت تقاضای سفر دانست که در بازه‌های مختلف زمانی از اواخر بهمن‌ماه ۱۳۹۸ تا پایان دوره کووید ۱۹ اجرا شده بود. برهمین اساس در این پژوهش با استفاده از مشاهدات حجم وسایل نقلیه ورودی به محدوده‌های ترافیکی شهر تهران، بین سال‌های ۱۳۹۶ الی ۱۴۰۰ که به صورت بازه‌های یک ساعته از دوربین‌های ترافیکی نصب شده به ورودی‌های محدودهٔ طرح ترافیک و کنترل آلودگی هوا، جمع آوری شده است، اقدام به پیش‌بینی حجم وسایل نقلیه ورودی در بازه زمانی یک ساعته با استفاده از سه روش یادگیری ماشین (جنگل تصادفی، XGBoost و نزدیک‌ترین همسایه) برای هردو محدودهٔ ترافیکی شهر تهران (طرح ترافیک و کنترل آلودگی هوا) کرده است. با توجه به تاثیرگذاری سیاست‌های مختلف مدیریت تقاضای سفر و تغییر این سیاست‌ها در بازه‌های مختلف زمانی، این سیاست‌ها، جزئی از متغیرهای پیش‌بینی کننده در کنار دیگر متغیرهای موثر براساس مطالعات پیشین، شامل مشاهدات تاریخی حجم وسایل نقلیه ورودی، شرایط آب‌وهوا، اطلاعات تقویمی مانند: روزهای تعطیل و غیرکاری، روزِ هفته و ساعتِ روز جهت پیش‌بینی حجم وسایل نقلیه ورودی، مورد استفاده قرار گرفته‌اند.

۲- پیشینه تحقیق

پژوهش‌های مختلفی با استفاده از مدل‌های یادگیری ماشین به پیش‌بینی متغیرهای مختلف ترافیکی مانند زمان سفر، حجم و سرعت وسایل نقلیه، در مناطق شهری و جاده‌ای جهت مدیریت جریان ترافیک و برنامه ریزی بهتر شهری، پرداخته‌اند (Medina-Salgado et al., 2022). روش‌های مبتنی بر یادگیری ماشین علاوه بر صرفه‌جویی در استفاده از منابع مادی و نیروی‌کار، می‌توانند با دقت بیشتری جریان ترافیک را پیش‌بینی کنند (Xie et al., 2020). از دیگر دلایل استفاده از الگوریتم‌های یادگیری ماشین می‌توان به رفتار غیرخطی و تصادفی داده‌های جریان ترافیک اشاره کرد که این مدل‌ها توانایی بیشتری در کشف این الگوها دارند (Fouladgar et al., 2017). آتا و همکاران (Ata et al., 2021) در این مقاله با استفاده از داده‌های حسگرهای تقاطع‌های شهر و الگوریتم‌های درخت تصمیم، نزدیک ترین همسایه و ماشین بردار پشتیبان، به پیش‌بینی جریان ترافیک به صورت پویا و زمان‌بندی علائم راهنمایی و رانندگی با توجه به میزان جریان ترافیک پرداختند. گنگ و همکاران (Gong et al., 2022) به بررسی تأثیر کووید-۱۹ و سیاست‌های دولتی مربوطه را بر حجم وسایل نقلیه، در مرحله قبل و بعدی همه‌گیری پرداختند. آنها از داده‌های ترافیکی بین دو شهر سالت لیک و یوتا استفاده کرده‌اند که این داده‌های شامل حجم وسایل نقلیه عبوری بین این دوشهر، سیاست‌های مختلف ترافیکی وضع شده در زمان همه گیری کووید ۱۹ و داده‌های مختلف آب و هوایی بوده است. آنها برای پیش بینی حجم وسایل نقلیه از مدل‌های مختلف یادگیری ماشین استفاده کرده‌اند که در این پژوهش از مدل‌های شبکه عصبی حافظه طولانی کوتاه مدت[1] و GCN-LSTM جهت برآورد حجم وسایل نقلیه عبوری مورد بررسی قرار گرفته است. نتایج این تحقیق نشان میدهد که مدل‌های پیش‌بینی مورد استفاده قرار گرفته عملکرد مطلوبی در پیش بینی دارند، به صورتی که مدل GCN-LSTM، بالاترین میانگین درصد مطلق خطا MAPE مدل در بین سناریوهای مختلف تنها ۱/۷۴ درصد بوده است. رسا و سیدابریشمی (Rasaizadi & Seyedabrishami, 2022) با استفاده از مدل‌های یادگیری ماشین گروهی به پیش‌بینی جریان ترافیک ساعتی در محور جاده‌ای تهران- ساوه و بالعکس پرداخته‌اند. آنها با استفاده مدل هایی شامل نزدیکترین همسایه، جنگل تصادفی، XGBoost و مدل‌های شبکه عصبی حافظه طولانی کوتاه مدت را آموزش داده‌اند. داده‌های استفاده شده در این مطالعه شامل ویژگی‌های زمانی دوره‌ای، تعطیلات

و شرایط آب و هوایی هستند. در مدل های کوتاه مدت علاوه بر ویژگی های ذکر شده، از جریان ترافیکی مشاهده شده در ۳ تا ۸ ساعت گذشته استفاده شده است. نتایج نشان می‌دهد که برای هر دو مدل کوتاه مدت و میان مدت، کمترین خطای پیش بینی توسط مدل XGBoost[۲] به دست می‌آید. با توجه به اهمیت تقاطع‌ها در استفاده همزمان مانند وسایل نقلیه، دوچرخه سواران و عابران پیاده، آلاجالی و همکاران (Alajali et al., 2018) به پیش‌بینی جریان ترافیک در تقاطع ها با استفاده از الگوریتم‌های مختلف درخت تصمیم شامل؛ درخت رگرسیون تقویت گرادیان، جنگل تصادفی، XGBoost و ماشین بردار پشتیبان[۳] پرداختند. آنها در این مقاله، علاوه بر استفاده از حجم ترافیک تقاطع از دیگرداده‌های زمینه‌ای هم‌چون روزهفته، زمان‌های تعطیلی، زمان اوج و روز یا شب بودن، اقدام پیش‌بینی جریان ترافیک تقاطع کرده‌اند. نتایج این مطالعه نشان می‌دهد، در صورت استفاده از داده‌های زمینه‌ای در کنار حجم تردد تقاطع، معیار میانگین توان دوم خطا[۴] (MSE) برای تمامی مدل‌های استفاده شده به صورت میانگین از۰,٦٤٧٤ به ۰,٦٢٧٢ کاهش یافته است. از دیگر مطالعه‌های صورت گرفته در پیش‌بینی مولفه‌های جریان ترافیک می‌توان به مطالعه توکلی و حاجی حسینلو (Tavakoli & Hadji Hosseinlou, 2023) اشاره کرد. این مطالعه با استفاده الگوریتم های یادگیری ماشین شامل ماشین بردار پشتیبان و شبکه‌ی عصبی شامل بازگشتی و حافظه طولانی کوتاه مدت به پیش بینی سرعت ترافیک پرداخته‌اند. داده های مورد استفاده در این پژوهش از شبکه شهر سیاتل ایالت متحده بدست آمده است. علاوه بر آن داده‌های مختلف زمانی شامل ساعت، روز، هفته و تعطیلات رسمی، در فرآیند پیش‌بینی استفاده شده است. این مطالعه سه فصل بهار، تابستان و پاییز در سه گام زمانی ۵ دقیقه ۱۰ دقیقه و ۱۵ دقیقه را جهت پیش‌بینی با هم مقایسه کرده است. نتایج این مطالعه نشان می‌دهد مدل پیشنهادی در فصل‌های مختلف اختلافی چشمگیری نداشته است و هر چه گام‌های زمانی بیشتر میشود خطاها بیشتر و دقت مدل کاهش پیدا می‌کند. براتساس و همکاران (Bratsas et al., 2020) با استفاده از داده‌های سرعت وسایل نقلیه در فواصل ۱۵ دقیقه‌ای که از هر بخش شبکه جاده‌ای شهر تسالونیکی در یونان جمع‌آوری شده بود و با استفاده از سه مدل یادگیری ماشین شامل جنگل تصادفی، رگرسیون بردار پشتیبان، و شبکه عصبی پرسپترون چندلایه[۵] اقدام به پیش‌بینی سرعت در سه سناریوی مختلف کردند. سناریوهای مورد بررسی در این مطالعه شامل انتخاب تصادفی جاده‌ها، زمان و تاریخ پیش‌بینی و انتخاب متفاوت توالی ساعتی شامل توالی ۱۵ دقیقه ای و ۱ ساعته بوده است. نتایج این مطالعه نشان می‌دهد، مدل شبکه عصبی تغییرات جریان ترافیک ناگهانی را با دقت بیشتری و مدل رگرسیون بردار پشتیبان تغییرات کوچک‌تر را با دقت بهتری پیش‌بینی کرده است. مدل شبکه عصبی نیز در مقایسه با سه مدل دیگر خطای کمتری و نزدیک به صفر را در پیش‌بینی های خود داشت. همان‌گونه که بررسی شد، مطالعات بسیاری پیرامون پیش‌بینی متغیرهای جریان ترافیک شامل حجم و سرعت وسایل نقلیه، با اهداف و رویکردهای متفاوت انجام شده است. متغیرهای پیش‌بینی کننده در اکثر مطالعات شامل مشاهدات گذشته جریان ترافیکی، داده‌های مربوط به زمان مانند: ساعت، روز هفته وفصل، داده‌های تقویمی شامل روزهای تعطیل و غیر تعطیل در کنار استفاده از داده‌های آب وهوایی، بوده است. اما در این میان، توجه کمتری به دیگر متغیرهای اثرگذار مانند سیاست‌های مدیریت تقاضای سفر شده است. برهمین اساس و با توجه به اهمیت محدوده‌های طرح ترافیک و کنترل آلودگی هوای شهر تهران و تاثیرگذاری سیاست‌های مدیریت تقاضای سفر بر حجم وسایل نقلیه ورودی به این محدوده‌های ترافیکی، در این پژوهش، علاوه بر استفاده از متغیرهای پیش‌بینی کننده مورد استفاده در مطالعات پیشین، سیاست‌های مدیریت تقاضای سفر نیز به عنوان بخشی از این متغیرها جهت پیش‌بینی حجم وسایل نقلیه ورودی به محدوده‌های ترافیکی شهر تهران در مدل‌های یادگیری ماشین (جنگل تصادفی، XGBoost و نزدیک‌ترین همسایه) استفاده شده است.

۳- روش‌شناسی

در این بخش به بررسی داده‌های پژوهش، مدل‌های یادگیری ماشین مورداستفاده شامل جنگل تصادفی، XGBoost و نزدیک‌ترین‌همسایه و معیارهای ارزیابی آنها پرداخته شده است.

۳-۱- داده‌ها

داده‌های اولیه این پژوهش، حجم وسایل نقلیه ورودی به محدودهٔ‌های ترافیکی شهر تهران (طرح ترافیک و کنترل آلودگی هوا) به تفکیک تاریخ، روز هفته و ساعت از مهرماه ۱۳۹٦ تا شهریورماه ۱٤۰۰ است که این اطلاعات توسط دوربین‌های ثبت تخلف که شامل شمارش حجم وسایل نقلیه نیز می‌شوند، در دو محدودهٔ طرح ترافیک و کنترل آلودگی هوا به‌دست‌آمده است. جدول (۱) خلاصه آماری حجم وسایل نقلیه ورودی به محدوده‌های ترافیکی را نشان می‌دهد. این داده به‌عنوان یک

سری زمانی در نظر گرفته شده است که از توالی سه ساعت گذشته جهت پیش‌بینی حجم وسایل نقلیه ورودی در ساعت آینده استفاده می‌کند.

جدول ۱. خلاصه آماری حجم ورودی به محدوده‌های ترافیکی شهر تهران (وسیله نقلیه/ ساعت)

محدودهٔ ترافیکی	میانگین	انحراف معیار	حداقل	حداکثر
طرح ترافیک	۵۲۲۰	۳۳۰۷	۲۷	۱۳۰۴۹
کنترل آلودگی هوا	۲۳۹۶۵	۱۳۴۶۱	۱۴۷	۴۹۶۳۸

از دیگر داده‌های استفاده شده جهت پیش‌بینی حجم وسایل نقلیه ورودی، اطلاعات تقویمی است که روزهای تعطیل رسمی و جمعه به‌عنوان روزهای غیرکاری و مابقی روزها به‌عنوان روزکاری در نظر گرفته شده است. بر همین اساس شرایط آب‌وهوا به دو وضعیت بارانی و آفتابی تقسیم و به مجموعه داده‌ها اضافه شده است. علاوه‌برآن از سیاست‌های مدیریت تقاضای سفر نیز به‌عنوان متغیرهای پیش‌بینی‌کننده استفاده شده است. از سیاست‌های اثرگذار می‌توان به سیاست قیمت‌گذاری محدوده‌های ترافیکی شهر تهران اشاره کرد. بر اساس این سیاست، کاربران خودرو شخصی برای تردد در ناحیه مرکزی شهر بر اساس ساعت تردد خود (اوج و غیر اوج) هزینه پرداخت می‌کنند. باتوجه‌به متفاوت‌بودن اخذ عوارض طرح کنترل ترافیک و آلودگی هوا در ساعت اوج و غیر اوج، رایگان بودن آن در بازه زمانی ۱۸:۰۰ بعدازظهر تا ۶:۳۰ صبح، متفاوت‌بودن زمان پایان طرح در روزهای پنجشنبه و رایگان بودن آن در تعطیلات رسمی و روزهای جمعه، این سیاست به‌عنوان یک متغیر پیش‌بینی‌کننده که دارای سه وضعیت رایگان، عوارض اوج و عوارض غیر اوج می‌باشد، در نظر گرفته شده است. به دلیل قرارگیری بخشی از اطلاعات حجم وسایل نقلیه در زمان کووید ۱۹ و شروع به کار ستاد ملی مبارزه با کرونا جهت اخذ تصمیمات دررابطه‌با مدیریت کشور، تصمیماتی را تحت عنوان مصوبات ستاد کرونا اعلام کرد که برخی از این مصوبات، سیاست‌هایی در ارتباط با مدیریت تقاضای سفر بوده است. این سیاست‌ها در بازه‌های مختلف زمانی از اواخر بهمن‌ماه ۱۳۹۸ تا پایان کار ستاد ملی مبارزه با کرونا اجرا شده که شامل "غیرحضوری شدن مراکز آموزشی، تغییر در ساعت کاری ادارات، اعمال محدودیت فعالیت مشاغل آزاد و غیردولتی و اعمال محدودیت تردد بین استانی" بوده است. این سیاست‌ها باتوجه‌به زمان اجرا و پایان، به صورت متغیرهای دودویی[۶] به مجموعه متغیرها اضافه شده است. جدول (۲) متغیرهای مورد بررسی در این پژوهش را نشان می‌دهد که جهت آموزش و آزمایش مدل‌های معرفی شده برای هر دو محدودهٔ (طرح ترافیک و کنترل آلودگی هوا) استفاده شده‌اند.

۲-۳- جنگل تصادفی[۷]

مدل جنگل تصادفی، یک مدل‌های یادگیری ماشین از دستهٔ یادگیری گروهی[۸] است که بر اساس ایده Bagging مدل‌سازی انجام می‌دهد. در این ایده دسته‌بندهای ضعیف به‌صورت موازی آموزش‌دیده شده‌اند و درنهایت با برآیندگیری از کل دسته‌بندها نتیجه خروجی دسته‌بند نمایش داده می‌شود. مدل جنگل تصادفی برای ایجاد نتایجی بادقت بالا، برای مسائل رگرسیون و دسته‌بندی استفاده می‌شود. مدل جنگل تصادفی در ابتدا توسط لئو بریمن معرفی شد. این مدل از تکرار داده‌های آموزشی، برای بهبود نتایج مدل و دسته‌بندی بادقت بالاتر استفاده می‌کند(Bratsas et al., 2020). در این پژوهش از آنتروپی که در رابطه(۱) نوشته شده است برای بررسی انشعاب مجموعه داده استفاده شده است که در آن p_i نسبت داده‌های کلاس i در گره است.

$$Entropy = -\sum_{i=1}^{c} p_i \times \log(p_i) \qquad (۱)$$

جدول ۲. متغیرهای مستقل پیش‌بینی حجم وسایل نقلیه ورودی به محدوده‌های ترافیکی شهر تهران

ردیف	نام متغیر	توضیح متغیر
۱	حجم وسایل نقلیه	مشاهدات حجم ورود به محدوده‌های ترافیکی در بازه زمانی سه ساعت گذشته
۲	زمانِ روز	۲۴ ساعت در شبانه روز

۳	روز هفته	شنبه تا جمعه
٤	شرایط آب‌وهوا	دو وضعیت آفتابی و یا بارانی
٥	غیرحضوری شدن مراکز آموزشی	در صورت آموزش حضوری =۱/ دور تحصیلی =۰
٦	تغییر در ساعت کاری ادارات	عدم تغییر در ساعت کاری ادارات =۱/ تغییر در ساعت کاری ادارات =۰
۷	اعمال محدودیت فعالیت مشاغل آزاد و غیردولتی	مجوز فعالیت مشاغل آزاد =۱/ عدم مجوز فعالیت مشاغل آزاد =۰
۸	اعمال محدودیت تردد بین استانی	در صورت مجوز تردد بین استانی =۱/ اعمال محدودیت تردد بین استانی=۰
۹	روز کاری یا غیرکاری	روزهای کاری =۱/ تعطیلات رسمی و روز جمعه = ۰
۱۰	عوارض محدودهٔ طرح ترافیک و کنترل آلودگی هوا	باتوجه‌به زمانِ روز: رایگان/ عوارض اوج/ عوارض غیر اوج

۳-۳- Extreme Gradient Boosting (XGBoost)

مدل XGBoost ، نیز یک مدل‌های یادگیری ماشین از دستهٔ یادگیری گروهی است. تفاوت این مدل با جنگل تصادفی در این است که XGBoost بر اساس ایده Boosting مدل‌سازی انجام می‌دهد. در این ایده دسته‌بندهای ضعیف به‌صورت سری آموزش‌دیده می‌شوند و در هر مرحله نمونه داده‌ای احتمال انتخاب بیشتری دارد که در مرحله قبل به‌صورت اشتباه دسته‌بندی شده است. در این ایده هر دسته‌بند بر اساس خطای دسته‌بندی به آن وزنی اختصاص پیدا می‌کند و در نهایت بر اساس مجموع وزن همه دسته‌بندهای آموزش‌دیده، دسته‌بندی نهایی داده را انجام می‌دهد. تابع هدف و زیان در XGBoost به‌صورت زیر تعریف شده است، فرایند ساخت دسته‌بندهای ضعیف در XGBoost تا برآورده‌شدن معیار توقف ادامه پیدا می‌کند (Shen & Wei, 2020). رابطه (۲) نشان دهنده‌ی تابع هدف مدل XGBoost است که در آن y_i برچسب صحیح داده، $\hat{y}_i$ برچسب‌ایی که مدل پیش‌بینی انجام داده است و عبارت $\Omega(f_k)$ پیچیدگی درخت‌های تصمیم ساخته شده را اندازه‌گیری می‌کند.

$$Obj = \sum_{i=1}^{m} l(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \qquad (۲)$$

٤-۳- نزدیک‌ترین همسایه[۹]

مدل نزدیک‌ترین‌همسایه یکی از مدل‌های یادگیری ماشین است که برای مسائل طبقه‌بندی و رگرسیون استفاده می‌شود. هدف این الگوریتم این است که کلاس داده را بر اساس کلاس‌های نزدیک‌ترین همسایگان آن در فضای ویژگی‌ها پیش‌بینی کند. این الگوریتم با محاسبه فاصله بین نقطه هدف و تمامی نقاط دیگر در مجموعه‌داده و معمولاً با استفاده از فاصله اقلیدسی، کار می‌کند. پس از تعیین فاصله‌ها، نزدیک‌ترین 'k' نقطه را شناسایی کرده و به کلاس داده اختصاص می‌دهد که بین این همسایگان بیشترین فراوانی را دارد. رابطه (۳)، فاصله اقلیدسی که در الگوریتم نزدیک‌ترین همسایه استفاده می‌شود را نشان داده است.(Cai et al., 2016).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (۳)$$

۵-۳- معیارهای ارزیابی

به‌منظور ارزیابی کیفیت پیش‌بینی‌های مدل، ایجاد معیارهایی که امکان مقایسه مدل‌های مختلف را فراهم می‌کند، ضروری است. این ارزیابی شامل مقایسه بین نتایج حجم تردد پیش‌بینی شده و حجم تردد مشاهده شده، در تاریخ و زمان انتخاب شده باشد. در این پژوهش از معیارهای زیر استفاده شده است (Medina-Salgado et al., 2022).

۱-۵-۳- میانگین خطای مطلق[۱۰] (MAE)

این معیار میانگین تفاوت بین مقدار پیش‌بینی شده توسط مدل $\hat{y}$ و مقدار مشاهده y را برای داده‌ها حساب می‌کند که رابطه (٤) نشان‌دهنده این معیار است.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}| \qquad (٤)$$

۲-۵-۳-جذر میانگین مربعات خطا[۱۱] (RMSE)

بر اساس رابطه (۵) این معیار با محاسبه ریشه دوم میانگین مربعات تفاوت بین مقادیر پیش‌بینی شده $\hat{y}$ و مقادیر مشاهده y

به دســـت می‌آید. این معیار تأثیر خطاهای بزرگ را بیشـــتر از خطاهای کوچک در نظر می‌گیرد. به همین دلیل، معیار مناســبی جهت بررسی دقت مدل رگرسیونی است.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_i - \hat{y})^2} \qquad (۵)$$

۳-۵-۳- میانگین درصد خطای مطلق[۱۲] (MAPE)

این معیار، میانگین خطای مطلق را به‌صـــورت درصـــد و از تفاوت بین مقدار پیش‌بینی شده توسط مدل $\hat{y}$ و مقدار مشاهده y را بر اساس رابطه (٦) محاسبه می‌کند.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}}{y_i}\right| \times 100 \qquad (٦)$$

٤-۵-۳- ضریب تعیین[۱۳] (R^2)

این معیار نشـــان‌دهندهٔ نســـبت واریانس کل داده‌هاســـت که توســـط مدل توضـــیح داده شـــده و بیانگر درصـــد تغییرات متغیر وابســـته به متغیر مســـتقل اســـت. در رابطه (۷) مقادیر پیش‌بینی شـــده $\hat{y}$، مقادیر مشاهده y و $\bar{y}$ میانگین مقادیر، قابل مشاهده است.

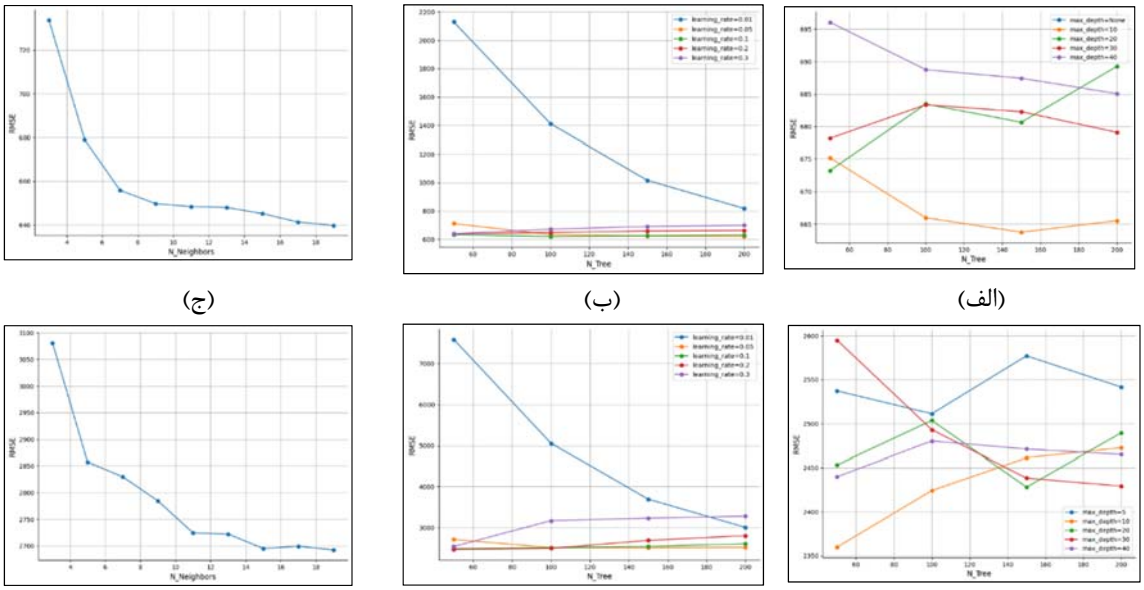
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (۷)$$

٤- تحلیل نتایج

پس از بررسی داده‌ها و متغیرهای پیش‌بینی‌کننده، جهت مدل‌سازی با استفاده از الگوریتم‌های یادگیری ماشین، ابتدا لازم است مقادیری مناسب، برای پارامترهای قابل‌تغییر جهت یادگیری بهتر مدل، انتخاب شود. هر مدل دارای پارامترهای متفاوتی است که در اینجا با استفاده از جستجوی شبکه توری[۱٤] مقادیر بهینه در کمترین معیار خطا (RMSE) انتخاب شده است. در مدل جنگل تصادفی تعداد درخت‌های ساخته شده (Nt) و عمق هر درخت (max-depth)، در مدل XGBoost نیز تعداد درخت‌های ساخته شده (Nt) و نرخ یادگیری (η) و در مدل نزدیک‌ترین‌همسایه، تعداد همسایه‌های مورداستفاده (K)، برخی از پارامترهای مهم در این مدل‌ها هستند. مقادیر بهینه انتخاب شده در جدول (۳) آورده شده است. باتوجه‌به شکل (۱) قسمت "الف"، در محدودهٔ طرح ترافیک، از مدل جنگل تصادفی با تعداد درخت (N_Tree = 150) بیشتری نسبت به این مدل در محدودهٔ کنترل آلودگی هوا (N_Tree = 50) استفاده شده است که نشان می‌دهد برای دستیابی به خطای کمتر در این محدودهٔ ترافیکی، نیاز به استفاده از تعداد درخت بیشتری در عمق (Max_Depth = 10) است. افزایش تعداد درخت از ۱۵۰ عدد با افزایش مجدد خطا مواجه شده است. بااین‌حال در محدودهٔ کنترل آلودگی هوا، تعداد درخت در این مدل به ۵۰ عدد کاهش پیدا کرده است که ممکن است نشان‌دهنده پیچیدگی کمتر داده‌های محدودهٔ کنترل آلودگی هوا باشد. در مدل XGBoost نیز با افزایش نرخ یادگیری (η) و کاهش تعداد درخت، می‌توان معیار خطا را برای محدودهٔ کنترل آلودگی هوا کاهش داد و در مدل نزدیک‌ترین همسایه، افزایش تعداد همسایه‌ها (k) به ٤ عدد، در محدودهٔ طرح ترافیک و ۵ عدد در محدودهٔ کنترل آلودگی هوا، کمترین معیار خطا را گزارش می‌دهد.

جدول ۳. مقادیر بهینه پارامترهای قابل تغییر در مدل‌ها

مدل	محدودهٔ ترافیکی	مقدار بهینه
جنگل تصادفی	طرح ترافیک	N_Tree = 150, Max_Depth=10
	کنترل آلودگی هوا	N_Tree = 50, Max_Depth=10
XGBoost	طرح ترافیک	N_Tree = 100, η =0.1
	کنترل آلودگی هوا	N_Tree = 50, η =0.2
نزدیک‌ترین همسایه	طرح ترافیک	Number of Neighbors =4
	کنترل آلودگی هوا	Number of Neighbors =5



(الف) (ب) (ج)

(د) (ه) (و)

شکل ۱. جست‌وجوی شبکه توری برای پیداکردن مقادیر بهینه پارامترهای قابل تغییر در مدل‌ها. ("الف" و"د") N_Tree و Max_Depth در مدل جنگل تصادفی. ("ب" و "ه") N_Tree و η در مدل XGBoost. ("ج" و "و") تعداد k در نزدیک‌ترین همسایه

نتایج جدول (٤) نشان می‌دهد که برای پیش‌بینی حجم وسایل نقلیه در محدودهٔ طرح ترافیک، مدل XGBoost و برای محدودهٔ کنترل آلودگی هوا مدل جنگل تصادفی کمترین خطا را در پیش‌بینی داشته‌اند و مدل نزدیک‌ترین همسایه دارای خطای بیشتری از دو مدل دیگر در هر دو محدودهٔ ترافیکی است. براساس مطالعات پیشین، معیارخطای MAPE جهت بررسی جریان ترافیک در بازه ای بین ۵ تا ۱۵ درصد بوده است (Medina-Salgado et al., 2022).در این پژوهش کمترین خطای MAPE، در مدل XGBoost محدودهٔ طرح ترافیک و برابر با ۸/۲ درصد بوده است. در حالی که بیشترین خطا مربوط به مدل نزدیک‌ترین همسایه در محدودهٔ کنترل آلودگی و برابر با ۱۲/۱۸ درصد است. میانگین خطای در محدودهٔ طرح ترافیک ۸/۷ درصد و در محدودهٔ کنترل آلودگی به ۱۱/٦٦ درصد رسیده است، این نشان می‌دهد تمامی مدل‌های مورد بررسی خطای کمتری در پیش‌بینی حجم وسایل نقلیه در محدودهٔ طرح ترافیک داشته‌اند. ازدلایل پیش‌بینی بهتر مدل‌های جنگل تصادفی و XGBoost می‌توان به یکسان بودن مدل‌های ضعیف استفاده در این دو مدل که درخت تصمیم است، اشاره کرد. این نتیجه نشان می‌دهد، مدل درخت تصمیم، ویژگی های به کار گرفته شده در این پژوهش را بهتر از مدل های دیگر توانسته در پیش‌بینی استفاده کند. شکل (۲) و شکل (۳) نشان دهنده‌ی مقادیر مشاهده و پیش‌بینی مدل XGBoost و جنگل تصادفی از پیش‌بینی حجم وسایل نقلیه برای محدودهٔ طرح ترافیک و کنترل آلودگی هوا است. در بسیاری از نقاط نمودار، خطوط مشاهده و پیش‌بینی تقریباً بر روی هم قرار گرفته‌اند، که نشان‌دهنده خطای پایین مدل پیش‌بینی است.تفاوت‌های کوچکی بین خطوط وجود دارد که نشان می‌دهد پیش‌بینی‌ها به طور کامل دقیق نیستند، اما به طور کلی اختلافات جزئی هستند. هرچند که مدل به خوبی عمل کرده است، اما هنوز خطاها و انحرافات کوچکی در پیش‌بینی‌ها مشاهده می‌شود. این خطاها ممکن است به علت نوسانات رخ داده به علت مسائلی مانند تصادفی یا عوامل غیر قابل پیش‌بینی در داده‌ها باشد. باتوجه‌به اینکه در محدوده‌های ترافیکی شهر تهران، سیاست‌های مدیریت تقاضای سفر از اهمیت به سزایی برخوردار هستند به تحلیل حساسیت متغیرهای مربوطه پرداخته شده است تا میزان تاثیر هر متغیر در افزایش دقت پیش‌بینی مورد بررسی قرار گیرد. جدول (۵) نتیجه تحلیل حساسیت را نشان می‌دهد. تاثیر این سیاست‌ها در مدل نزدیک‌ترین همسایه، بیشترین تاثیر را در افزایش دقت پیش‌بینی برای هر دو محدودهٔ ترافیکی داشته است. میانگین این تاثیر در مدل نزدیک‌ترین‌همسایه، ۳/۳۷ درصد بوده است،

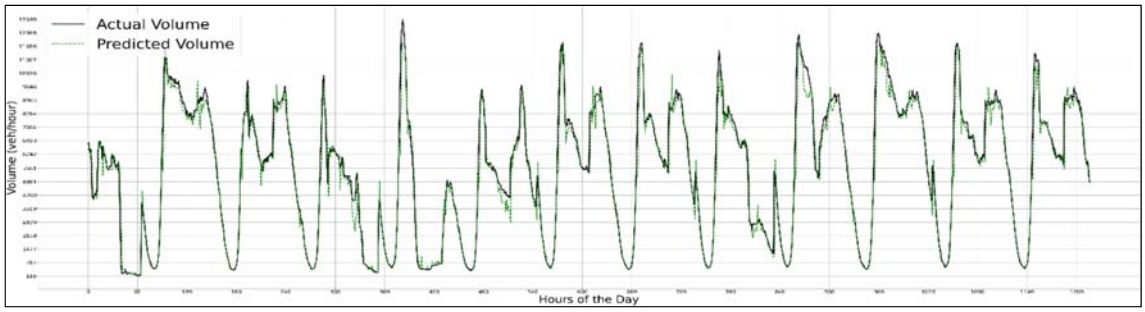
جدول ٤. معیارهای ارزیابی خطا در مدل‌های پیش‌بینی حجم وسایل نقلیه ورودی به محدوده‌های ترافیکی شهر تهران

مدل‌ها	محدودهٔ ترافیکی	RMSE	MAE	MAPE
جنگل تصادفی	طرح ترافیک	٦٧١	٣٧٧	٪٨/٨
	کنترل آلودگی هوا	٢٣٩٠	١٠٨٥	٪١١/١٦
XGBoost	طرح ترافیک	٦١٧	٣٣٦	٪٨/٢
	کنترل آلودگی هوا	٢٣٩٧	٩٨٧	٪١١/٦٥
نزدیک‌ترین همسایه	طرح ترافیک	٦٤٦	٣٥٥	٪٩/٢١
	کنترل آلودگی هوا	٣٦١٤	١٣٥٩	٪١٢/١٨

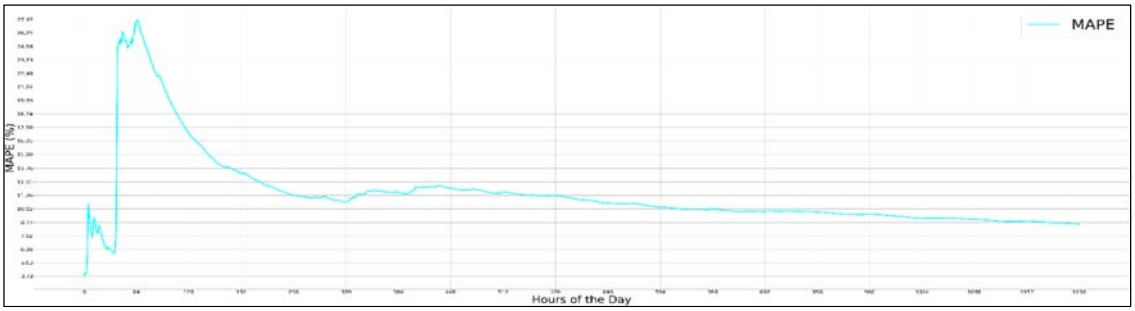
در صورتی‌که این مقدار برای مدل‌های XGBoost و جنگل تصادفی به ترتیب ۱/۲ درصد و ۲/۸۷ درصد بوده است. این نتیجه نشان می‌دهد مدل XGBoost، پایداری بیشتری در پیش بینی حجم وسایل نقلیه در مقایسه با سایر مدل‌ها داشته است. مجموع تحلیل حساسیت سیاست‌های مدیریت تقاضای سفر نشان می‌دهد، سیاست‌های "غیرحضوری شدن مراکز آموزشی" و "اخذ عوارض محدودهٔ طرح ترافیک و کنترل آلودگی هوا" بیشترین و سیاست" اعمال محدودیت تردد بین استانی" کمترین تأثیر را بر افزایش دقت پیش‌بینی مدل برای هر دو محدودهٔ ترافیکی داشته است. سیاست "غیرحضوری شدن مراکز آموزشی" که از اسفند ۱۳۹۸ تا مهر ۱۴۰۱ ادامه پیدا کرد، طولانی‌ترین سیاست مدیریت تقاضای سفر در دورهٔ کووید ۱۹ بوده که توانسته است الگویی مهم در افزایش دقت پیش‌بینی ایجاد کند که از دلایل آن می‌توان به سهم زیاد سفرهای تحصیلی انجام شده به مرکز شهر اشاره کرد. ضریب تعیین در تمامی مدل‌های مورد بررسی پس از حذف متغیرهای سیاست‌گذاری کاهش‌یافته است که این نشان می‌دهد، متغیرهای سیاست‌گذاری بر پیش‌بینی دقیق‌تر حجم وسایل نقلیه تأثیرگذار بوده است.

جدول ٥. میزان کاهش ضریب تعیین در مدل‌های مورد بررسی پس از حذف متغیرهای مدیریت تقاضای سفر

سیاست‌های مدیریت تقاضا	محدودهٔ	مدل‌ها		
		جنگل تصادفی	XGBoost	KNN
عوارض محدودهٔ طرح ترافیک و کنترل آلودگی هوا بر اساس ساعت اوج/غیر اوج	طرح ترافیک	٪٣/٨	٪٢/٤	٪٤
	کنترل آلودگی هوا	٪٣/٢٨	٪١/١٨	٪٢/٥٠
غیرحضوری شدن مراکز آموزشی	طرح ترافیک	٪٥	٪٣	٪٩
	کنترل آلودگی هوا	٪٢/٨٣	٪٠/٦٦	٪٢/٤٤
تغییر در ساعت کاری ادارات	طرح ترافیک	٪٢/٨	٪٢/٤	٪٣/٧٢
	کنترل آلودگی هوا	٪٢/٦٥	٪٠/١٨	٪١/١٤
اعمال محدودیت فعالیت مشاغل آزاد و غیردولتی	طرح ترافیک	٪٢	٪٠/٧٥	٪٣/٢٥
	کنترل آلودگی هوا	٪١/١٦	٪٠/٣٣	٪١/٤٤
اعمال محدودیت تردد بین استانی	طرح ترافیک	٪١/٢٤	٪٠/١٦	٪٢/٣١
	کنترل آلودگی هوا	٪١/١٢	٪٠/٣٩	٪٠/٤٧

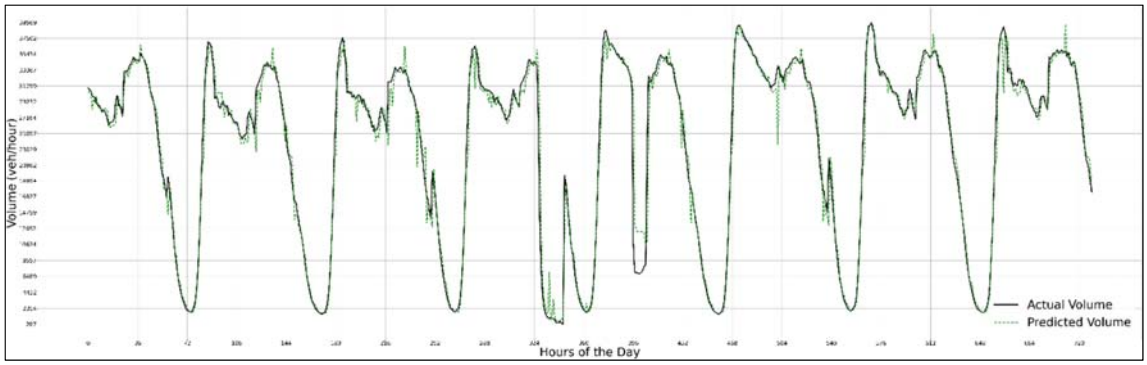


(الف)

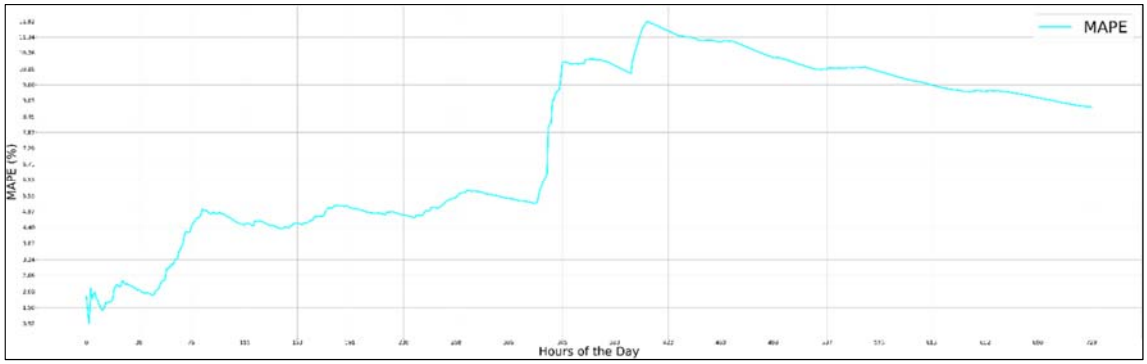


(ب)

شکل ۲. مدل XGBoost برای محدودهٔ طرح ترافیک (الف) مقادیر مشاهده و پیش‌بینی. (ب) میانگین درصد مطلق خطا



(الف)



(ب)

شکل ۳. مدل جنگل تصادفی برای محدودهٔ کنترل آلودگی هوا (الف) مقادیر مشاهده و پیش‌بینی. (ب) میانگین درصد مطلق خطا

۵- نتیجه‌گیری

اطلاع از حجم وسایل نقلیه ورودی به محدوده‌های ترافیکی شهر می‌تواند تأثیر معناداری بر سیاست‌گذاری و مدیریت ترافیک در این نواحی داشته باشد و با اطلاع‌رسانی به‌موقع به رانندگان در مورد وضعیت ترافیک می‌تواند به کاهش زمان سفر و افزایش رضایت‌مندی عمومی کمک کند. این پژوهش با استفاده از مجموعه‌دادهٔ حجم ورودی وسایل نقلیه به محدوده‌های ترافیکی شهر تهران از سال ۱۳۹۶ تا ۱۴۰۰ و با اضافه‌کردن دیگر متغیرهای تأثیرگذار شامل روز و ساعت، شرایط آب‌وهوا و اضافه‌کردن سیاست‌های مختلف مدیریت تقاضای سفر به‌عنوان متغیرهای پیش‌بینی‌کننده، اقدام به پیش‌بینی حجم تردد وسایل نقلیه ورودی به این محدوده‌های ترافیکی کرده است. مدل‌های یادگیری ماشین مورداستفاده در این پژوهش شامل جنگل تصادفی، XGBoost و نزدیک‌ترین همسایه است که در ابتدا مقادیر بهینه پارامترهای قابل‌تغییر در هر مدل، مورد بررسی قرار گرفت و مقداری که با کمترین خطا همراه بود جهت مدل‌سازی انتخاب شدند. نتایج نشان می‌دهد که در بین مدل‌های مورد بررسی، مدل XGBoost برای طرح ترافیک با میانگین درصد خطای مطلق ۸/۲٪ و مدل جنگل تصادفی با ۱۱/۱۶٪ عملکرد مناسب‌تری نسبت به مدل‌های دیگر داشته‌اند. تحلیل حساسیت سیاست‌های مدیریت تقاضای سفر نشان می‌دهد، تمامی این سیاست‌ها منجر به افزایش دقت پیش‌بینی شده است که از بین آنها، سیاست‌های "غیرحضوری شدن مراکز آموزشی" و "اخذ عوارض محدودهٔ طرح ترافیک و کنترل آلودگی هوا" بیشترین و سیاست "اعمال محدودیت تردد بین استانی" کمترین تأثیر را بر افزایش دقت پیش‌بینی داشته‌اند، علاوه‌برآن، نتایج تحلیل حساسیت نشان می‌دهد که سیاست‌های با دوره زمانی طولانی‌تر، توانسته‌اند الگوی تأثیرگذارتری جهت پیش‌بینی حجم وسایل نقلیه به وجود بیاورند که در بین سیاست‌های مورد بررسی، سیاست قیمت‌گذاری محدوده‌های ترافیکی با تغییر به روش دوره مبنا از سال ۱۳۹۷ و سیاست "غیرحضوری شدن مراکز آموزشی" از بهمن‌ماه سال ۱۳۹۸ تا مهرماه ۱۴۰۱، به دلیل دورهٔ زمانی طولانی‌تر نسبت به دیگر سیاست‌ها، اثر بیشتری بر افزایش دقت مدل داشته‌اند. نتایج این پژوهش می‌تواند در برنامه‌ریزی مؤثر مدیریت شهری و سیستم‌های هوشمند حمل‌ونقل جهت نشان‌دادن تقاضای برخط محدوده‌های ترافیکی مورداستفاده قرار گیرد. پیشنهاد می‌شود جهت ادامه پژوهش، به بررسی بازه‌های مختلف زمانی و تأثیر آن بر مقادیر پیش‌بینی را مورد ارزیابی قرارداد.

۶- پی‌نوشت‌ها

1. Long-Short Term Memory
2. Extreme Gradient Boosting
3. Support Vector Machines
4. Mean Squared Error
5. Multilayer Perceptron Neural Network
6. Binary Variable
7. Random Forest
8. Ensemble Learning
9. K-Nearest Neighbors
10. Mean Absolute Error
11. Root Mean Squared Error
12. Mean Absolute Percentage Error
13. R-Squared
14. Grid Search

۷- مراجع

-Alajali, W., Zhou, W., Wen, S., & Wang, Y. (2018). Intersection traffic prediction using decision tree models. *Symmetry*, *10*(9), 1–16. https://doi.org/10.3390/sym10090386

-Ata, A., Khan, M. A., Abbas, S., Khan, M. S., & Ahmad, G. (2021). Adaptive IoT Empowered Smart Road Traffic Congestion Control System Using Supervised Machine Learning Algorithm. *The Computer Journal*, *64*(11), 1672–1679. **doi.org/10.1093/comjnl/bxz129**

-Bratsas, C., Koupidis, K., Salanova, J. M., Giannakopoulos, K., Kaloudis, A., & Aifadopoulou, G. (2020). A comparison of machine learning methods for the prediction of traffic speed in Urban places. *Sustainability (Switzerland)*, *12*(1), 1–15. **doi.org/10.3390/SU12010142**

-Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., & Sun, J. (2016). A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies*, *62*(January), 21–34. **doi.org/10.1016/j.trc.2015.11.002**

-Fouladgar, M., Parchami, M., Elmasri, R., & Ghaderi, A. (2017). Scalable deep traffic flow neural networks for urban traffic congestion prediction. *2017 International Joint Conference on Neural Networks (IJCNN)*, 2251–2258. **doi.org/10.1109/IJCNN.2017.7966128**

-Shami, S., & Mamdoohi, A. R. (2022). An effectiveness analysis of Tehran peak-based traffic scheme, a travel behavior model. *Journal of Transportation Research*, *19*(3), 149–164. **doi.org/10.22034/tri.2021.286382.2909**

-Shen, X., & Wei, S. (2020). Application of XGBoost for Hazardous Material Road Transport Accident Severity Analysis. *IEEE Access*, *8*, 206806–206819. **doi.org/10.1109/ACCESS.2020.3037922**

-Tavakoli, E., & Hadji Hosseinlou, M. (2023). Short-Term Prediction of Traffic Speed using Recurrent Neural Networks (RNN). *Quarterly Journal of Transportation Engineering*, *15*(1), 3369–3394. https://doi.org/10.22119/jte.2022.342446.2600

-Xie, P., Li, T., Liu, J., Du, S., Yang, X., & Zhang, J. (2020). Urban flow prediction from spatiotemporal data using machine learning: A survey. *Information Fusion*, *59*, 1–12. **doi.org/10.1016/j.inffus.2020.01.002**

-Zhang, Z., Fu, D., Liu, F., Wang, J., Xiao, K., & Wolshon, B. (2023). COVID-19, traffic demand, and activity restriction in China: A national assessment. *Travel Behaviour and Society*, *31*, 10–23. **doi.org/10.1016/j.tbs.2022.11.001**

-Gong, Y., Isom, T., Lu, P., Yang, X., & Wang, A. (2022). Modeling the impact of COVID-19 on transportation at later stage of the pandemic: A case study of Utah. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *0*(0), 1–11. **doi.org/10.1080/15472450.2022.2157212**

-Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., & Sierra, J. E. (2022). Urban traffic flow prediction techniques: A review. *Sustainable Computing: Informatics and Systems*, *35*(January). **doi.org/10.1016/j.suscom.2022.100739**

-Rasaizadi, A., & Seyedabrishami, S. (2022). Stacking Ensemble Learning Process to Predict Rural Road Traffic Flow. *Journal of Advanced Transportation*. **doi.org/10.1155/2022/3198636**

-Razali, N. A. M., Shamsaimon, N., Ishak, K. K., Ramli, S., Amran, M. F. M., & Sukardi, S. (2021). Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning. *Journal of Big Data*, *8*(1). **doi.org/10.1186/s40537-021-00542-7**

A Prediction of Vehicles Entering the Traffic Cordons, Using Machine Learning (Case Study: Tehran City)

MahmoudReza Nateghi, M.Sc., Student, Faculty of Civil & Environmental Engineering, Tarbiat Modares University, Tehran, Iran.
Sepodeh Shami, Ph.D., Student, Faculty of Civil & Environmental Engineering, Tarbiat Modares University, Tehran, Iran.
Amir Reza Mamdoohi, Associate Professor, Faculty of Civil & Environmental Engineering, Tarbiat Modares University, Tehran, Iran;
and Adjunct Professor, Department of Civil, Geological & Mining Eng., Polytechnique Montréal, Montréal, Canada.

E-mail: armamdoohi@modares.ac.ir

Received: August 2024- Accepted: December 2024

ABSTRACT
Predicting the number of vehicles is crucial in intelligent transportation systems and is a key component of traffic management. Effective vehicle management significantly improves the efficiency of urban road networks and reduces congestion. This study utilizes vehicle entry data for Tehran's traffic cordons from 2017 to 2021 and employs three machine learning methods (Random Forest, XGBoost, and K-Nearest Neighbors) to predict the demand for incoming vehicles in each traffic cordon (Traffic Plan and Air Pollution Control Plan) on an hourly basis. In addition to conventional variables such as weather data and time, travel demand management policies are also used as predictor variables. Sensitivity analysis is conducted to evaluate the impact of different policies on predicting traffic volume. The results indicate that the XGBoost model and the Random Forest model, with a mean absolute percentage error (MAPE) of 8.2% and 11.16%, respectively, for traffic and air pollution cordons, perform better compared to the K-Nearest Neighbors model. Sensitivity analysis of variables related to travel demand management policies shows that the policies of "online schooling" and "tolls for entering the traffic and air pollution cordons" have the greatest impact, while the policy of "inter-provincial travel restrictions" has the least impact on improving prediction accuracy.

Keywords: Prediction, Number of Vehicles, Travel Demand Management, Machine Learning, Xgboost