فصلنامه مهندسی ترافیک، سال بیست و پنجم، شماره ۱۰۰، بهار ۱۴۰۴، ص. ۷۰-۵۷

ارزیابی الگوریتم‌های یادگیری ماشین در مدل‌سازی انتخاب (مطالعه موردی: قیمت‌گذاری پارک حاشیه‌ای در شهر تهران)

سپیده شامی، دانشجوی دکتری، گروه مهندسی برنامه‌ریزی حمل‌ونقل، دانشکده مهندسی عمران و محیط‌زیست، دانشگاه تربیت مدرس، تهران، ایران

امیررضا ممدوحی (مسئول مکاتبات)، دانشیار گروه مهندسی برنامه‌ریزی حمل‌ونقل، دانشکده مهندسی عمران و محیط‌زیست، دانشگاه تربیت مدرس، تهران، ایران، استاد وابسته دانشکده مهندسی عمران، زمین‌شناسی و معدن، دانشگاه فنی مونترال، کانادا

E-mail: armamdoohi@modares.ac.ir

چکیده

یکی از مشکلات جوامع شهری، تراکم، آلودگی هوا و آلودگی صوتی ناشی از استفاده از خودرو شخصی در سطح شهر برای سفرهای روزانه است، که محققان حمل‌ونقل، سال‌ها در پی حل این معضل هستند. یکی از راهکارهای مواجه با این مشکلات، استفاده از سیاست‌های مدیریت تقاضای سفر است. ازجمله این سیاست‌ها، قیمت‌گذاری پارک حاشیه‌ای است. برای سنجش اثربخشی سیاست‌های مدیریت تقاضای سفر و درنتیجه پیش‌بینی رفتار شهروندان در مواجه با سیاست‌های مختلف، راهکارهای مختلفی وجود دارد. یکی از این راهکارها بررسی نحوه تغییر رفتار سفر کاربران در مواجه با سیاست است. یک نوع تغییر رفتار سفر، تغییر وسیله از خودروی شخصی به سایر شیوه‌ها است. بررسی انتخاب بین چند گزینه به‌خصوص انتخاب وسیله افراد معمولاً با استفاده از مدل‌های انتخاب گسسته انجام می‌شود. در سال‌های اخیر، مدل‌های یادگیری ماشین برای مدل‌سازی انتخاب موردتوجه محققان بوده است. این مدل‌ها تنوع زیادی دارند، فاقد پیش‌فرض بوده و برای طیف وسیعی از مجموعه داده قابل‌استفاده هستند. هدف این مطالعه به‌کارگیری مدل‌های KNN، SVM، Decision Tree، Random Forest و Ada Boost برای بررسی عملکرد آن‌ها در مدل‌سازی نحوه انتخاب رفتار کاربران در صورت قیمت‌گذاری پارک حاشیه‌ای و بررسی قابلیت پیش‌بینی مدل‌ها است. در این مطالعه از اطلاعات حاصل از پرسشگری به روش رجحان بیان‌شده از ۹۰۰ نفر از کاربران خودروی شخصی که از پارک حاشیه‌ای در سطح شهر استفاده کرده‌اند استفاده شد. بررسی نمونه موردی نشان داد که تمایل به تغییر وسیله از خودروی شخصی، در صورت قیمت‌گذاری پارک حاشیه‌ای، کم (۲۷ درصد نمونه) است. در بین مدل‌ها نیز Random Forest با صحت ۷۰ درصد و مدل Ada Boost با صحت ۶۸ درصد، بهترین پیش‌بینی را دارند.

واژه‌های کلیدی: یادگیری ماشین، مدل انتخاب، قیمت‌گذاری پارک حاشیه‌ای

۱. مقدمه

کشورهای زیادی در سراسر دنیا برای مواجه با مشکل ازدحام و هزینه‌های زیست‌محیطی و اجتماعی حاصل از آن در کلان‌شهرهای خود، سال‌ها است که از انواع مختلف سیاست‌های مدیریت تقاضای سفر استفاده می‌کنند. به‌علاوه با توجه به هزینه زیاد ساخت و توسعه زیرساخت‌های حمل‌ونقل، سیاست‌های مدیریت تقاضای حمل‌ونقل علاوه بر بهینه بودن ازلحاظ مالی، می‌توانند درآمدزا نیز باشند و درآمد حاصل از آن در راستای بهبود کیفیت و افزایش زیرساخت‌های مربوط به حمل‌ونقل همگانی و غیر موتوری بکار گرفته شود.

ازجمله سیاست‌های مدیریت تقاضای سفر، قیمت‌گذاری پارک حاشیه‌ای است. قیمت‌گذاری پارک حاشیه‌ای می‌تواند منجر به کاهش مطلوبیت استفاده از خودروی شخصی، درنتیجه کاهش تراکم، افزایش ضریب سرنشین و کاهش مدت‌زمان پارک شود. مسئله اصلی بررسی تأثیرگذاری یک سیاست مدیریت تقاضای سفر مانند قیمت‌گذاری پارک حاشیه‌ای بر کاربران خودروی شخصی است. کاربران در مواجه با قیمت‌گذاری واکنش‌های متفاوتی خواهند داشت. این واکنش‌ها شامل تغییر رفتار سفر، مالکیت خودرو و انتخاب مکان سکونت یا کار است. در کوتاه‌مدت، تغییر رفتار سفر از سایر تغییرات محتمل‌تر است. تغییر رفتار سفر شامل تغییر زمان سفر، لغو سفر، کاهش تعداد دفعات سفر و تغییر وسیله سفر، عموماً از خودروی شخصی به سایر وسایل است.

برای بررسی رفتار سفر و بخصوص انتخاب وسیله، غالباً از مدل‌های انتخاب گسسته استفاده می‌شود. این مدل‌ها بر اساس فرضیه‌های آماری پرداخت‌شده و قابلیت تفسیرپذیری بالایی دارند؛ اما علی‌رغم مزایای بسیار، با محدودیت‌هایی مانند پیش‌فرض‌های آماری و لزوم انتخاب متغیرهای اثرگذار توسط مدل‌ساز در فرایند پرداخت مدل مواجه هستند. در سال‌های اخیر محققان سعی در استفاده از یادگیری ماشین در مدل‌سازی رفتار سفر افراد داشته‌اند. این مدل‌ها داده مبنا بوده و با استفاده از درون‌یابی سعی در کشف ارتباط بین متغیرها برای پیش‌بینی متغیر پاسخ را دارند.

تاکنون در مطالعاتی از یادگیری ماشین برای انتخاب وسیله استفاده شده و نتایج آن با مدل لوجیت چندگانه مقایسه شده است و در بیشتر موارد نیز مدل‌های یادگیری ماشین برآورد بهتری از مدل‌های انتخاب گسسته داشته‌اند. در غالب مطالعات تمرکز بر اطلاعات وضع موجود سفرهای شهروندان است و رفتار سفر فعلی آنان موردبررسی قرار گرفته است تا قابلیت مدل‌سازی انتخاب وسیله با استفاده از الگوریتم‌های مختلف یادگیری ماشین آزمون شود. در مواردی نیز از الگوریتم یادگیری ماشین برای انتخاب یک شیوه جدید حمل‌ونقل یا دستیابی به الگوریتم‌های قیمت‌گذاری استفاده شده است.

در این مطالعه تمرکز بر رفتار سفر افراد در صورت مواجه با قیمت‌گذاری پارک حاشیه‌ای در شهر تهران و مدل‌سازی انتخاب وسیله با استفاده از الگوریتم یادگیری ماشین و بررسی قابلیت آن‌ها و مقایسه مدل‌ها با یکدیگر است. در این مطالعه از اطلاعات حاصل از پرسشگری با شیوه رجحان بیان‌شده در سال ۱۳۹۳ از ۹۰۰ کاربر خودروی شخصی (با پارک حاشیه‌ای رایگان در مقصد) استفاده شده است. نتایج این مطالعه می‌تواند مبنای استفاده از مدل یادگیری ماشین در انتخاب رفتار سفر به‌خصوص در شرایط ترجیحات بیان‌شده و سنجش اثر قیمت‌گذاری پارک حاشیه‌ای مورداستفاده قرار بگیرد.

۲. ادبیات پژوهش

موارد متعددی از کاربردهای موفقیت‌آمیز الگوریتم‌های یادگیری ماشین برای مدل‌سازی انتخاب در ادبیات موجود است. در این بخش به ارائه برخی مطالعات بر اساس کاربرد در قیمت‌گذاری پارکینگ، روش و مدل اصلی که توسط محققان در مقالات آن‌ها مطالعه شده، پرداخته خواهد شد.

سومیا[1] و همکاران (۲۰۲۳) در مطالعه خود از مدل Reinforcement Learning برای ارائه قیمت‌گذاری پویای پارک حاشیه‌ای با استفاده از اطلاعات تقاضای سفر شهروندان،

فاصله پارکینگ‌ها از مرکز شهر، وضعیت آب‌وهوا و تعطیلات استفاده کردند. نتایج مطالعات آن‌ها نشان داد الگوریتم قیمت‌گذاری پیشنهادی می‌تواند برای مدیریت پارکینگ به‌خصوص در ساعات اوج مناسب باشد.

هنگ[۲] و همکاران (۲۰۲۳) با استفاده از اطلاعات تقاضای پارک حاشیه‌ای در دو شهر سانفرانسیسکو و سیاتل و با استفاده از مدل‌های یادگیری ماشین یک مدل قیمت‌گذاری پارکینگ ارائه کردند. در این مطالعه تمرکز بر انتخاب مدلی بود که بیشترین دقت در پیش‌بینی زمانی و مکانی را داشته باشد. بر اساس مقایسات صورت گرفته مدل Neural Ordinary Differential Equations، بهترین دقت را در مقایسه با سایر مدل‌ها در پیش‌بینی فضاهای خالی و درنتیجه قیمت‌گذاری متناسب با تقاضا داشت.

دنگ[۳] (۲۰۲۱) در مطالعه خود به قیمت‌گذاری پارکینگ با استفاده از یادگیری ماشین پرداخت. او عنوان کرد که قیمت‌گذاری پارکینگ به‌عنوان ابزاری برای تنظیم منابع موجود عمل می‌کند. سؤال اصلی این مطالعه چگونگی استفاده از قیمت‌گذاری پارکینگ برای به حداکثر رساندن استفاده از منابع پارکینگ است؛ درعین‌حال که هدف درآمد پارکینگ برای مدیریت پارکینگ بهینه باشد. در این مطالعه فرایندی پیشنهاد می‌شود که از منابع عمومی موجود و درعین‌حال بهینه‌سازی درآمد با محدودیت‌های از پیش تعریف‌شده، به‌ویژه در زمینه مدیریت پارکینگ استفاده می‌کند. به‌طور خاص، ابتدا یک مدل پیش‌بینی مبتنی بر سری‌های زمانی مبتنی بر داده را طراحی، و سپس یک مدل قیمت‌گذاری پویا مبتنی بر یادگیری تقویتی برای گنجاندن محدودیت‌های قیمت طراحی شده است. علاوه بر این، معیارهایی برای ارزیابی مدل قیمت‌گذاری پویا و همچنین پیاده‌سازی و ارزیابی مدل‌های پیشنهادی با داده‌های پارکینگ واقعی ارائه شده است. داده این مطالعه داده پارکینگ از یک فرودگاه کانادا شامل ۲۵۱۴۴ ردیف است. یک مدل شبکه عصبی مبتنی بر LSTM طراحی شد که برای پیش‌بینی سری‌های زمانی واحد تخمین رزرو استفاده می‌شود. نتایج ارزیابی نشان می‌دهد که شبکه عصبی مبتنی بر LSTM از مدل سنتی ARIMA در هر سه نوع داده پارکینگ عملکرد مطلوبی دارد.

ساهاران[۴] و همکاران (۲۰۲۰) در مطالعه خود به بررسی قیمت‌گذاری پارکینگ با استفاده از الگوریتم یادگیری ماشین پرداختند. در این مطالعه یک یادگیری ماشین مبتنی بر اشغال بر اساس طرح قیمت‌گذاری پارکینگ در خیابان پیشنهاد شده است. طرح پیشنهادی از رویکردهای مبتنی بر یادگیری ماشین و با استفاده از مدل‌های Neural Network، Random Forest و Decision Tree برای پیش‌بینی اشغال پارکینگ‌ها استفاده می‌کند، که به‌نوبه خود برای استنباط قیمت‌های اشغال‌شده برای وسایل نقلیه ورودی استفاده می‌شود. به‌منظور آموزش، آزمایش و مقایسه مدل‌های مختلف یادگیری ماشینی، از داده‌های پارکینگ در خیابان شهر سیاتل استفاده شده است. نتایج به‌دست‌آمده با استفاده از یادگیری ماشین مبتنی بر اشغال پیشنهادی بر اساس طرح قیمت‌گذاری پارکینگ در خیابان، اثربخشی آن را بر سایر طرح‌های پیشرفته موجود نشان می‌دهد.

کشیفی[۵] و همکاران (۲۰۲۲)، یک چارچوب سیستماتیک یادگیری ماشین را برای درک بهتر تصمیمات انتخاب وسایل نقلیه مسافران پیشنهاد کردند. پنج مدل مختلف یادگیری ماشین (رگرسیون لجستیک، جنگل تصادفی، درخت تصمیم، پرسپترون چندلایه، درخت تصمیم تقویت گرادیان سبک) برای مدل‌سازی انتخاب وسایل نقلیه مسافران با استفاده از داده‌های سه‌ساله نظرسنجی ملی سفر هلند توسعه داده شدند. نتایج تجربی معیارهای ارزیابی عملکرد مختلف (دقت کلی، دقت متوسط، منحنی‌های دقت-بازیابی) نشان داد که LightGBDT در مقایسه با سایر مدل‌ها عملکرد بهتری داشت. به‌منظور غلبه بر نقدهای مربوط به جعبه سیاه بودن مدل‌های یادگیری ماشین و بهبود قابلیت تفسیر آن‌ها، تحلیل اهمیت متغیر نیز انجام شد. تحلیل نشان داد که پیش‌بینی کننده‌هایی که به‌طور قابل‌توجهی بر تصمیمات انتخاب وسایل نقلیه مسافران تأثیر می‌گذارند، شامل

مسافت سفر، سن مسافران و درآمد سالانه، تعداد خودروها/دوچرخه‌های متعلق به مسافران و فراوانی سفر هستند. نتایج می‌تواند برای درک بهتر و مدل‌سازی مؤثرتر ترجیحات انتخاب وسایل نقلیه مسافران استفاده شود.

پیرا و دیانا[۶] (۲۰۱۹)، رویکرد جدیدی مبتنی بر ماشین بردار پشتیبان برای شناسایی الگوهای انتخاب وسیله سفر پیشنهاد دادند. اگرچه آن‌ها دریافتند که دقت حاصل از این روش قابل‌مقایسه با یک مدل "خوب" نیست، اما همچنین متوجه شدند که ماشین بردار پشتیبان می‌تواند یک تقریب اولیه در مطالعات موردی ارائه دهد که در آن حجم زیادی از داده‌ها نیاز به پردازش سریع دارند و راه‌حل‌های اکتشافی قابل‌قبول هستند.

چنگ[۷] و همکاران (۲۰۱۹)، در مطالعه خود یک روش جنگل تصادفی قدرتمند برای تحلیل انتخاب وسایل نقلیه پیشنهاد می‌دهند تا توانایی پیش‌بینی و تفسیر مدل را بررسی کند. با استفاده از داده‌های سفر از نانجینگ چین در سال ۲۰۱۳ که شامل متغیرهای محیط مصنوع است، اثرات پارامترهای مختلف مدل بر عملکرد پیش‌بینی بررسی شده است. نتایج مقایسه نشان می‌دهد که روش جنگل تصادفی در پیش‌بینی انتخاب وسایل نقلیه با دقت بالاتر و هزینه محاسباتی کمتر عملکرد بسیار بهتری دارد. علاوه بر این، روش پیشنهادی اهمیت نسبی متغیرهای توضیحی و چگونگی ارتباط آن‌ها با انتخاب وسایل نقلیه را تخمین می‌زند. از میان متغیرهای توضیحی، محیط مصنوع به‌طورکلی بیشتر از ویژگی‌های خانوار و فردی اهمیت دارد. ترکیب کاربری زمین، مالکیت خودرو و سن مهم‌ترین متغیرها در ویژگی‌های محیط مصنوع، ویژگی‌های خانوار و ویژگی‌های فردی هستند. این نشان می‌دهد که سیاست کاربری زمین بیشتر با نحوه سفر مرتبط است و کلید تغییر تقاضای سفر است. همچنین، هنگام ارزیابی اثرات یک سیاست حمل‌ونقل پیشنهادی، باید به واکنش‌های خانوارهای دارای خودرو و بدون خودرو و گروه‌های سنی مختلف توجه ویژه‌ای کرد.

چپلی[۸] و همکاران (۲۰۱۹)، با هدف کاهش سفرهای با خودروی شخصی در مونترال، اقدام به مدل‌سازی سفرهای ساکنین کردند. آن‌ها انتخاب ۴ شیوه اصلی سفر شامل خودروی شخصی، همگانی، اتوبوس مدرسه و پیاده را با استفاده از جنگل تصادفی و مدل‌های انتخاب گسسته مدل کردند. نتایج نشان داد الگوریتم جنگل تصادفی می‌تواند یک توصیف دقیق و مفصل از چهار شیوه سفر با استفاده از ۱۱ ویژگی خانوارها، افراد و سفرها ایجاد کند. الگوریتم همچنین در شناسایی کاربران شیوه‌های با سهم کمتر (پارک و سوار و دوچرخه) مشکل داشت، اما در این زمینه بهتر از مدل انتخاب گسسته عمل کرد. از میان متغیرها سن مسافران و مبدا–مقصد جز مهم‌ترین آنان در انتخاب وسیله بودند.

لی[۹] و همکاران (۲۰۱۸) در مطالعه خود به مقایسه مدل‌های شبکه عصبی و لوجیت چندجمله‌ای برای مدل‌سازی انتخاب شیوه سفر با استفاده از مجموعه داده‌های جمع‌آوری‌شده ردیاب سفر منطقه شهری شیکاگو از سال ۲۰۰۷ تا ۲۰۰۸ کردند. نتایج مقایسات نشان‌دهنده عملکرد بهتر مدل‌های شبکه عصبی به نسبت لوجیت چندگانه با دقت پیش‌بینی در حدود ۸۰ درصد در مقایسه با ۷۰ درصد برای لوجیت چندگانه است.

خلاصه نتایج مطالعات نشان می‌دهد که مدل‌های یادگیری ماشین دارای تنوع بوده و با مجموعه داده مختلفی قابل برازش هستند. در مواردی این مدل‌ها کارایی بهتری نسبت به مدل‌های کلاسیک همچون لوجیت چندگانه داشتند. از مدل‌های یادگیری ماشین همچنین برای ارائه مدل قیمت‌گذاری پارکینگ نیز استفاده می‌شود. در غالب مطالعات این مقایسات با استفاده از اطلاعات سفر موجود شهروندان پرداخت شده است. در این مطالعه سعی خواهد شد تا یک مقایسه بین مدل‌های یادگیری ماشین تحت شرایط قیمت‌گذاری پارک حاشیه‌ای در شهر تهران و با استفاده از ترجیحات بیان‌شده افراد انجام شود. نتیجه این مقایسه نشان می‌دهد که کدام مدل‌ها می‌توانند بهتر رفتار سفر کاربران را به‌خصوص در شرایطی که قیمت عامل تعیین‌کننده وسیله سفر است توضیح دهند و پیش‌بینی کنند.

۳. روش پژوهش

در این بخش به توضیح مختصر مدل‌های یادگیری ماشین مورداستفاده (درخت تصمیم، جنگل تصادفی، ماشین بردار پشتیبان، نزدیک‌ترین همسایه و AdaBoost) سپس پایگاه داده مورداستفاده با هدف بررسی مدل‌های مختلف یادگیری ماشین برای انتخاب کاربران پرداخته شده است.

۳-۱ الگوریتم‌های یادگیری ماشین با ناظر

الگوریتم‌های یادگیری ماشین انواع زیادی دارند که با توجه به هدف و نوع مجموعه داده قابل انتخاب و پرداخت هستند. با توجه به هدف پژوهش یعنی مدل‌سازی انتخاب شهروندان از بین گزینه‌های مختلف، از الگوریتم‌های یادگیری با ناظر استفاده خواهد شد. در این بخش توضیح مختصری در رابطه با هریک ارائه خواهد شد.

۳-۱-۱ درخت تصمیم

درخت تصمیم یکی از الگوریتم‌های ساده و قابل‌فهم است که ساختاری سلسله‌مراتبی دارد. این الگوریتم با تقسیم مکرر داده‌ها به گروه‌های کوچک‌تر بر اساس ویژگی‌های مختلف، مدلی شبیه به یک درخت ایجاد می‌کند. در این ساختار، هر گره داخلی نشان‌دهنده یک ویژگی است که بر اساس آن داده‌ها تقسیم می‌شوند، هر شاخه یک خروجی از آزمون ویژگی و هر برگ نماینده دسته نهایی (یا خروجی) است. این روش به دلیل شفافیت بالا و توانایی در مدیریت داده‌های غیرخطی بسیار موردتوجه است. بااین‌حال، حساسیت این الگوریتم به نویز و احتمال بیش برازش از چالش‌های آن است.

۳-۱-۲ ماشین بردار پشتیبان

ماشین بردار پشتیبان برای جدا کردن داده‌ها به کلاس‌های مختلف از یک مرز بهینه استفاده می‌کند. این الگوریتم با یافتن یک ابر صفحه[۱۰] که داده‌ها را با بیشترین فاصله ممکن از یکدیگر جدا می‌کند، عملکرد بسیار خوبی در تفکیک کلاس‌ها دارد. در شرایطی که داده‌ها خطی قابل‌تفکیک نباشند، از کرنل‌هایی مانند کرنل چندجمله‌ای یا گوسی برای نگاشت داده‌ها به فضای با ابعاد بالاتر استفاده می‌شود. این ویژگی SVM را به یک ابزار قدرتمند برای داده‌های پیچیده تبدیل کرده است.

۳-۱-۳ نزدیک‌ترین همسایه

الگوریتم K نزدیک‌ترین همسایه یک روش غیر پارامتری است که بر اساس شباهت یا فاصله بین نقاط عمل می‌کند. در این الگوریتم، هنگام پیش‌بینی دسته یک نمونه جدید، k نمونه نزدیک به آن در فضای ویژگی‌ها شناسایی و دسته‌ای که بیشترین تعداد نمونه در آن قرار دارد به‌عنوان دسته نهایی انتخاب می‌شود. مزیت KNN سادگی در پیاده‌سازی و عدم نیاز به مرحله آموزش است، اما در صورت وجود حجم زیاد داده، هزینه محاسباتی آن افزایش می‌یابد.

۳-۱-۴ AdaBoost

AdaBoost یا الگوریتم تقویت تطبیقی یکی از روش‌های تقویتی است که ترکیبی از چندین مدل ساده (معمولاً درخت تصمیم کوچک) برای ایجاد یک مدل قوی‌تر است. در این روش، وزن بیشتری به نمونه‌هایی اختصاص داده می‌شود که به‌اشتباه دسته‌بندی شده‌اند تا مدل‌های جدید بر روی این نمونه‌ها تمرکز کنند. با تکرار این فرآیند، مجموعه‌ای از مدل‌ها ایجاد می‌شود که درنهایت با یکدیگر ترکیب می‌شوند. این مدل در دسته‌بندی‌هایی که داده‌ها به‌طور خطی قابل‌تفکیک نیستند عملکرد خوبی دارد.

۳-۱-۵ جنگل تصادفی

جنگل تصادفی یک الگوریتم قدرتمند مبتنی بر تکنیک مجموعه‌ای[۱۱] است که از ترکیب چندین درخت تصمیم برای بهبود دقت و تعمیم‌پذیری مدل استفاده می‌کند. در این روش، مجموعه‌ای از درخت‌های تصمیم با استفاده از نمونه‌برداری تصادفی از داده‌ها و انتخاب تصادفی ویژگی‌ها در هر تقسیم‌بندی ایجاد می‌شوند. هر درخت به‌طور مستقل آموزش می‌بیند و در مرحله پیش‌بینی، نتیجه نهایی از طریق اکثریت آرای درخت‌ها (برای مسائل طبقه‌بندی) یا میانگین نتایج (برای مسائل رگرسیون) تعیین می‌شود. مزیت جنگل تصادفی در کاهش خطر

بیش برازش و افزایش دقت مدل است، به‌ویژه در مواجه با داده‌های پیچیده یا نامتوازن. این الگوریتم همچنین توانایی اندازه‌گیری اهمیت ویژگی‌ها در داده را دارد که می‌تواند برای کاهش ابعاد داده بسیار مفید باشد.

۳-۲ مجموعه داده

مجموعه داده مورداستفاده در این بخش، اطلاعات حاصل از پرسشگری از شهروندان تهران به روش رجحان بیان شده است. این مجموعه داده در سال ۱۳۹۳ با هدف بررسی اثر سیاست قیمت‌گذاری پارکینگ بر استفاده از خودروی شخصی جمع‌آوری شده است. در این پرسشنامه سه سطح قیمتی به ازای یک ساعت پارک در مقصد به شهروندان ارائه شد و شهروندان تصمیم خود را در رابطه با سفر با خودروی شخصی، از بین گزینه‌های سفر اعلام می‌کردند. شکل ۱ نحوه ارائه قیمت پارک خودروی شخصی به ازای یک ساعت و گزینه‌های قابل انتخاب شهروندان را نشان می‌دهد. اطلاعات اقتصادی اجتماعی مانند شغل، تعداد خودروی تحت مالکیت، سن و تحصیلات نیز از شهروندان دریافت شده است.

	قیمت استفاده از پارکینگ (تومان بر ساعت)	گزینه انتخابی در صورت اجرای سناریو																	
		تاکسی	اژدلس	مترو	اتوبوس تندرو	اتوبوس	سواری شخصی	موتورسیکلت	پیاده	سایر (ذکر شود):	قبل از ۶:۳۰	۶:۳۰ تا ۱۰	ساعت ۱۰ تا ۱۷	بعد از ۱۷	همراهی با سواری شخصی فردی دیگر	تغییر مقصد سفر	حذف سفر	عدم تغییر
۳-۷- سناریو۱	۲۰۰۰	○	○	○	○	○	○	○	○	.....	○	○	○	○	○	○	○	○
در صورت انتخاب گزینه "عدم تغییر" به سوال ۳-۸ پاسخ داده شود در غیر این صورت برو به سوال ۳-۹																		
۳-۸- سناریو۲	۳۰۰۰	○	○	○	○	○	○	○	○	.....	○	○	○	○	○	○	○	○
۳-۹- سناریو۳	۱۰۰۰	○	○	○	○	○	○	○	○	.....	○	○	○	○	○	○	○	○

شکل ۱. سناریوهای قیمت ساعتی پارکینگ و گزینه‌های سفر

نحوه پرسش سناریوها به این صورت است که در ابتدا سناریو قیمتی ۲۰۰۰ تومان (سطح میانی قیمت) از شهروندان پرسش شده است، درصورتی‌که شهروندان در این سطح قیمت گزینه عدم‌تغییر را انتخاب کنند، قیمت ساعتی پارکینگ افزایش یافته و قیمت ساعتی ۳۰۰۰ تومان پرسش خواهد شد. درصورتی‌که شهروندان در سطح قیمت ۲۰۰۰ تومان، تغییری در سفر خود ایجاد کنند، سناریو قیمتی ۱۰۰۰ تومان بابت یک ساعت پارک پرسش شده است.

۴. تحلیل داده‌ها و نتایج مدل‌سازی

در این بخش به ارائه آمار توصیفی مجموعه داده و متغیرهای استفاده شده در مدل‌سازی و نهایتاً نتایج مدل‌سازی رفتار سفر با استفاده از الگوریتم‌های یادگیری ماشین پرداخته خواهد شد.

۴-۱ تحلیل داده

در این پژوهش، نحوه واکنش افراد به قیمت‌گذاری در دو دسته تغییر وسیله و بدون تغییر در نظر گرفته شده است. سایر گزینه‌ها به دلیل فراوانی کم از مجموعه داده کنار گذاشته شده است. فراوانی نسبی تغییر شیوه سفر ۲۶,۹۲ درصد و عدم‌تغییر ۷۳,۰۸ درصد است. به این معنا که بیشتر پاسخ‌دهندگان در هر دو سطح

قیمتی پارکینگ، تمایلی به تغییر شیوه سفر خود نداشتند. این امر وابستگی شدید افراد به خودروی شخصی را نشان می‌دهد. جدول ۱ نشان‌دهنده فراوانی، میانگین، انحراف استاندارد و مقدار کمینه و بیشینه متغیرهای قیمت پارک حاشیه‌ای، زمان سفر نهایی، سن و تعداد خودروی تحت تملک خانوار نمونه است.

جدول ۱. نتایج تحلیل آمار توصیفی متغیرهای مستقل مدل‌سازی

	قیمت پارک حاشیه‌ای (parking price)	زمان سفر نهایی (ttfinal)	سن (age)	تعداد خودرو (carsnum)
فراوانی	۸۴۷	۸۴۷	۸۳۲	۸۴۷
میانگین	۲۷۹۶,۹	۳۶,۸	۴۱,۷	۱,۱۵
انحراف معیار	۵۲۹,۳	۲۲,۴	۱۱,۳	۰,۳۹
کمینه	۱۰۰۰	۵	۲۰	۱
بیشینه	۳۰۰۰	۱۲۰	۸۴	۴

جدول ۲ نشان می‌دهد که فراوانی متغیر پاسخ در هر سطح قیمتی از پارکینگ به چه میزان است. بیشترین فراوانی مربوط به قیمت ۳۰۰۰ تومان و گزینه عدم‌تغییر است.

جدول ۲. فراوانی متغیر پاسخ در سطوح مختلف قیمت پارک (تومان)

قیمت پارک حاشیه‌ای	۱۰۰۰	۲۰۰۰	۳۰۰۰
تغییر وسیله (change)	۲۱	۴۶	۱۶۱
عدم‌تغییر (not change)	۲۹	۲۶	۵۶۴

جدول ۳ بیانگر فراوانی اهداف سفر در هر سطح از متغیر پاسخ است. بیشترین فراوانی مربوط به هدف سفر کاری و گزینه عدم‌تغییر رفتار سفر در صورت مواجه با قیمت‌گذاری است.

جدول ۳. فراوانی متغیر پاسخ به تفکیک هدف سفر

هدف سفر	همراهی	تحصیلی	شغلی	درمانی	سایر	خرید
تغییر وسیله	۲۹	۱۵	۱۳۵	۸	۲۸	۱۳
عدم‌تغییر	۵۴	۲۲	۴۲۲	۱۸	۵۷	۴۶

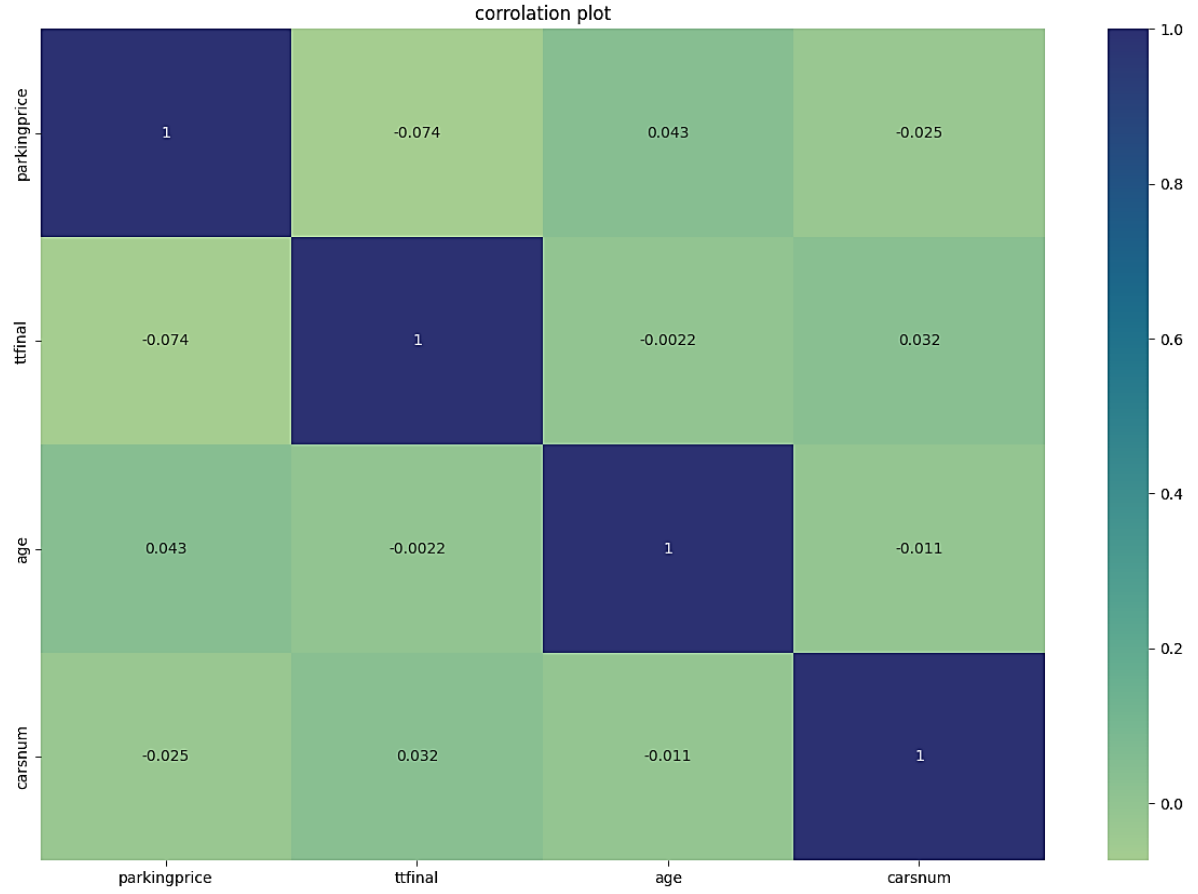
برای شناخت بیشتر نسبت به ساعت سفر افراد، فراوانی ساعت سفر به تفکیک گزینه پاسخ در جدول ۴ ارائه شد. بیشترین فراوانی مربوط به سفرهای اوج صبح است که گزینه بدون تغییر انتخاب شده است.

جدول ۴. فراوانی گزینه پاسخ به تفکیک زمان شروع سفر

متغیر پاسخ	تغییر وسیله	عدم‌تغییر
زمان شروع سفر		
اوج صبح	۱۴۲	۴۱۳

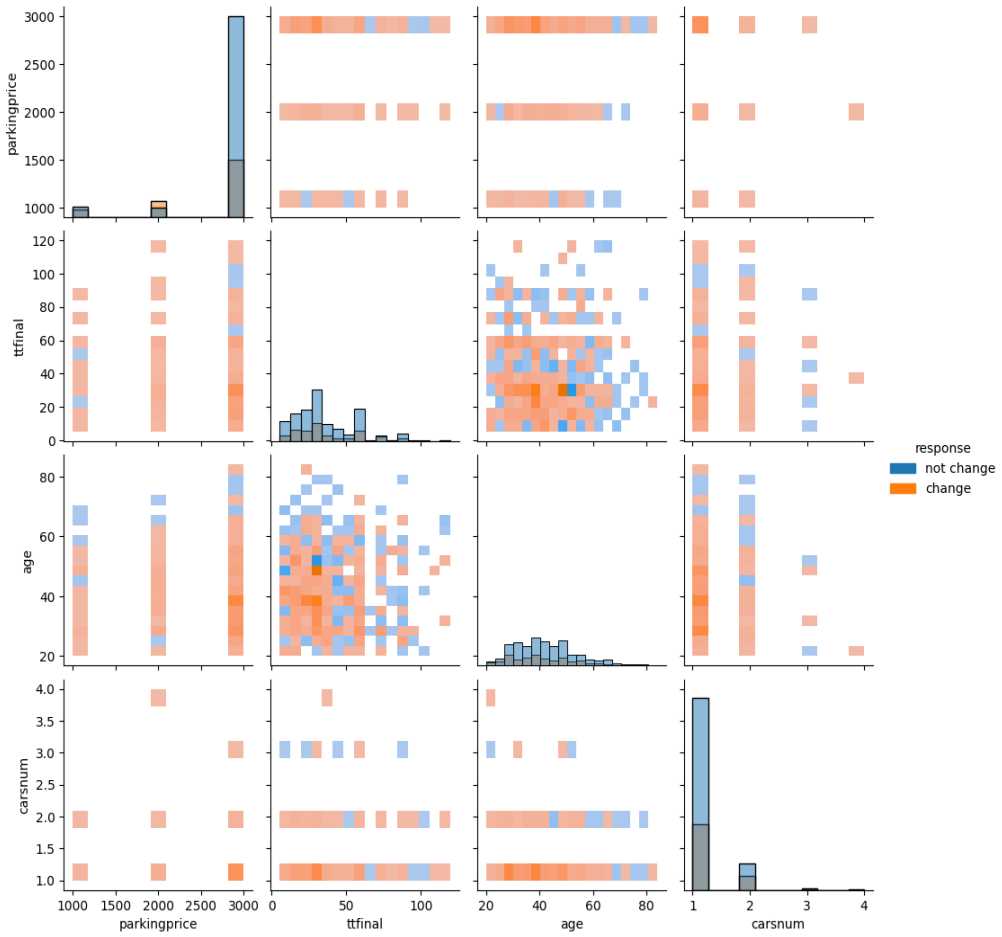
متغیر پاسخ	تغییر وسیله	عدم‌تغییر
میان روز	۳۳	۷۶
اوج عصر	۱۷	۴۱
سایر ساعات روز	۳۶	۸۹

برای بررسی همبستگی بین متغیرهای زمان سفر، تعداد خودرو، سن و قیمت پارک حاشیه‌ای از نمودار همبستگی استفاده شد. این نمودار نشان می‌دهد متغیرهای مستقل استفاده شده برای مدل‌سازی، همبستگی ناچیزی دارند که این امر لازمه مدل‌سازی بهتر است.



شکل ۲. نمودار جفتی بین سطوح مختلف متغیرهای مستقل و متغیر وابسته

شکل ۳، نشان‌دهنده ارتباط جفتی بین متغیرهای مستقل تعداد خودرو، سن، زمان سفر و هزینه پارک ساعتی، هم‌زمان با متغیر پاسخ است. نکته قابل‌توجه در این نمودار، عدم تمایل به تغییر رفتار سفر در سطوح بالای قیمت پارک ساعتی است. در سطوح سنی کمتر تمایل به ایجاد تغییر در رفتار سفر بیشتر است.



شکل ۳. نمودار جفتی بین سطوح مختلف متغیرهای مستقل و متغیر وابسته

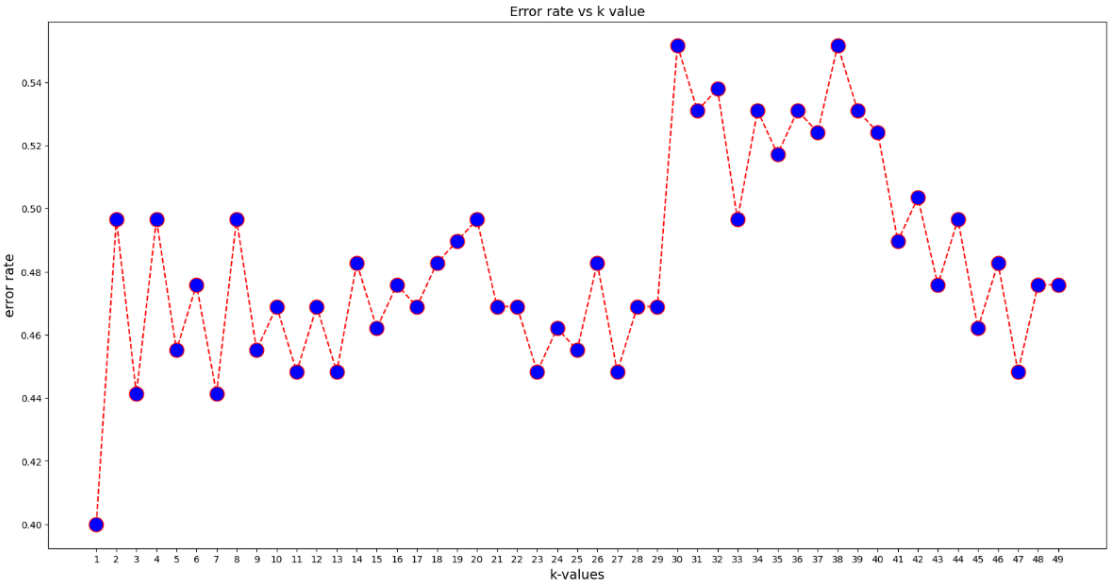
۴-۲ نتایج مدل‌سازی

مدل‌های برازش داده‌شده در این مطالعه KNN، SVM، Decision Tree، Ada Boost و Random Forest هستند. لازم به ذکر است قبل از مدل‌سازی اصلاحات مجموعه داده، مانند شناسایی و حذف داده پرت و جایگزینی داده گمشده انجام شده است. مدل‌سازی در قالب مدل‌های یادگیری باناظر به شکل دو کلاسه بدون تغییر (۰) و با تغییر وسیله (۱) در نظر گرفته شده است. خروجی‌ها شامل نمودارهای نرخ خطا، precision، recall، accuracy و F1-score است.

مدل‌های در نظر گرفته‌شده در این پژوهش، با هدف آزمایش هر مدل برای مجموعه داده مورداستفاده است.

۴-۲-۱ مدل KNN

مدل اول ساخته‌شده مدل KNN است. برای تنظیم هایپر پارامتر این مدل از نمودار نرخ خطا استفاده شد. بر اساس این نمودار می‌توان K مناسب را انتخاب کرد (شکل ۴). برازش این نمودار در تکرارهای زیاد بر روی مجموعه داده مورداستفاده نشان داد که این مدل مناسب مجموعه داده نیست. بااین‌حال خروجی آن گزارش می‌شود. برازش مدل با k=3 مدلی با صحت ۵۶ درصد شد. در ادامه معیارهای خوبی برازش مدل گزارش شده است.



شکل ۴. نمودار نرخ خطا مدل K نزدیک‌ترین همسایگی

جدول ۵. معیارهای نیکویی برازش مدل K نزدیک‌ترین همسایگی

معیارهای نیکویی برازش			
	Precision	recall	F1-score
0 بدون تغییر	۰,۳۲	۰,۴۱	۰,۳۶
1 تغییر وسیله	۰,۷۱	۰,۶۲	۰,۶۶
accuracy		۰,۵۶	

۴-۲-۲ مدل SVM

مدل بعدی SVM است که نسبت به KNN صحت بهتری داشت. برای تنظیم هایپر پارامترهای این مدل از Grid Search استفاده شد. خروجی این مدل در قالب ماتریس درهم‌ریختگی و معیارهای نیکویی برازش ارائه می‌گردد. این مدل صحتی در حدود ۶۷ درصد دارد که به نسبت مدل قبلی پیش‌بینی بهتری انجام می‌دهد اما خطا همچنان در کلاس صفر بالا است.

جدول ۶. معیارهای نیکویی برازش مدل SVM

معیارهای نیکویی برازش			
	Precision	recall	F1-score
0 بدون تغییر	۰,۴۳	۰,۴۱	۰,۴۲
1 تغییر وسیله	۰,۷۵	۰,۷۳	۰,۷۲
accuracy		۰,۶۶	

۴-۲-۳ مدل DT

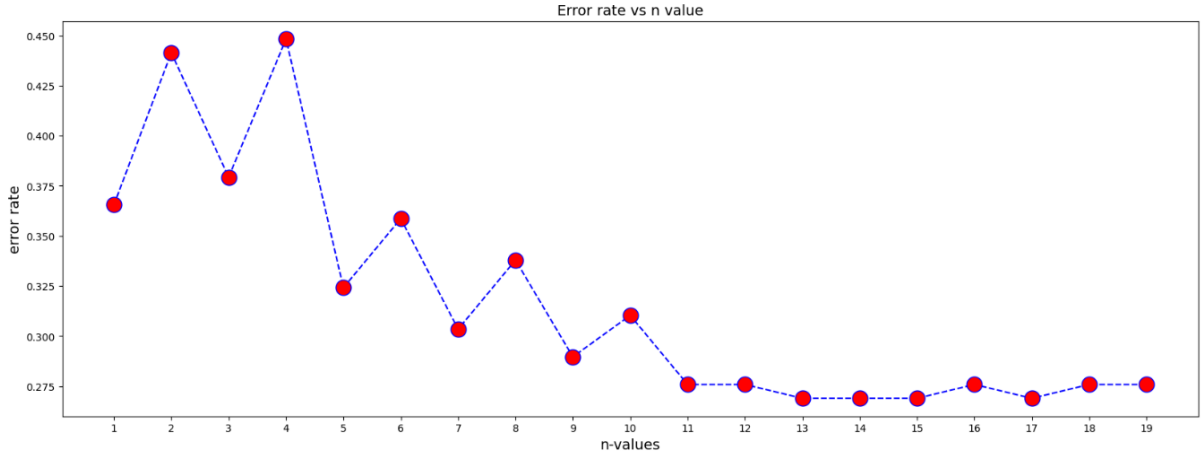
برای تنظیم هایپر پارامترهای این مدل از Grid Search استفاده شد. خروجی این مدل در قالب ماتریس درهم‌ریختگی و معیارهای نیکویی برازش ارائه می‌گردد. صحت این مدل نسبت به SVM کمتر است.

جدول ۷. معیارهای نیکویی برازش مدل DT

معیارهای نیکویی برازش			
	Precision	recall	F1-score
0 بدون تغییر	۰,۲۴	۰,۱۱	۰,۱۵
1 تغییر وسیله	۰,۶۹	۰,۸۴	۰,۷۶
accuracy		۰,۶۲	

۴-۲-۴ مدل Random Forest

هایپر پارامترهای جنگل تصادفی با استفاده از Grid Search استخراج شد. این مدل از سه مدل قبلی برازش بهتری داشته و صحتی برابر ۷۰ درصد ارائه داد. از نمودار شکل ۵ برای تنظیم نقطه تقسیم مناسب برای ساخت جنگل استفاده می‌شود. تنظیم این پارامتر به شکل درست موجب جلوگیری از بیش برازش می شود.



شکل ۵. تعیین نرخ خطا در Random Forest

جدول ۸ معیارهای نیکویی برازش مدل Random Forest

معیارهای نیکویی برازش			
	Precision	recall	F1-score
0 بدون تغییر	۰,۵۰	۰,۰۵	۰,۰۸
1 تغییر وسیله	۰,۷۰	۰,۹۸	۰,۸۲
accuracy		۰,۷۰	

۴-۲-۵ مدل Ada Boost

هایپر پارامترهای این مدل نیز با استفاده از Grid Search استخراج شد. با توجه به ماتریس درهم‌ریختگی و معیار صحت، می‌توان گفت همچنان جنگل تصادفی مدل بهتری است؛ زیرا این مدل صحتی برابر با ۶۸ درصد دارد.

جدول ۹. معیارهای نیکویی برازش مدل Ada Boost

معیارهای نیکویی برازش			
	Precision	recall	F1-score
0 بدون تغییر	۰,۴۴	۰,۲۵	۰,۳۲
1 تغییر وسیله	۰,۷۲	۰,۸۶	۰,۷۹
accuracy		۰,۶۸	

۵. جمع‌بندی و نتیجه‌گیری

در حوزه تقاضای سفر در حمل‌ونقل، برای بررسی رفتار سفر و بخصوص انتخاب گزینه سفر بین چند گزینه، غالباً از مدل‌های انتخاب گسسته استفاده می‌شود. در این پژوهش با بررسی مطالعات پیشین، نحوه به‌کارگیری مدل‌های یادگیری ماشین در سیاست‌های مختلف مدیریت تقاضا موردمطالعه قرار گرفت. مدل‌های یادگیری ماشین انواع مختلفی دارند که با اهداف و داده‌های مختلف قابل برازش هستند. اینکه از کدام‌یک از مدل‌های یادگیری ماشین در یک تحقیق استفاده شود به عوامل مختلفی همچون هدف و نوع داده بستگی دارد. از مدل‌های دسته‌بند برای بررسی تأثیر سیاست قیمت‌گذاری پارک حاشیه‌ای بر رفتار سفر شهروندان می‌توان بهره برد.

در این پژوهش با استفاده از اطلاعات جمع‌آوری‌شده با پرسشگری به روش رجحان بیان‌شده، ترجیحات شهروندان برای رفتار سفر در صورت مواجه با قیمت‌گذاری پارک حاشیه‌ای دریافت شد. در ادامه از مدل‌های یادگیری ماشین KNN ،SVM ،Decision Tree ،Random Forest و Ada Boost برای انتخاب رفتار سفر کاربران در صورت قیمت‌گذاری پارک حاشیه‌ای استفاده شد. در این پژوهش هدف انجام یک بررسی بر روی قابلیت مدل‌سازی مدل‌های یادگیری ماشین نام‌برده برای مدل‌سازی انتخاب بود. پس از بررسی مدل‌های دسته‌بند، می‌توان گفت برای مجموعه داده توصیف‌شده، مدل جنگل تصادفی با صحت ۷۰ درصد، بهترین مدل بود. بعدازاین مدل Ada Boost قابلیت بهتری برای پیش‌بینی انتخاب دارد. برای بهبود عملکرد مدل‌ها می‌توان از راهکارهای مختلفی نظیر ورود ویژگی‌های بیشتر به مدل‌سازی، به‌کارگیری متغیرهای ترکیبی و متعادل‌سازی گزینه پاسخ استفاده کرد.

۶. پی‌نوشت‌ها

1. Sowmya
2. Huang
3. Dang
4. Saharan
5. Kashifi
6. Pirra & Diana
7. Cheng
8. Chapleau
9. Lee
10. Hyperplane
11. Ensemble

۷. مراجع

– 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.

A Characterization of Travel Modes. Https://Doi.Org/10.1177/0361198119839339, 2673(4), 173–183.

– Jing, Y., Liu, Y., Zhang, Z., & Su, Y. (2019). Passenger travel behaviour on Chinese high-speed railways using machine learning based on revealed-preference data. Expert Systems, 36(4). https://doi.org/10.1111/exsy.12422

– Sowmya, K., & Dhabu, M. M. (2023). Model free Reinforcement Learning to determine pricing policy for car parking lots. Expert Systems with Applications, 230, 120532. https://doi.org/10.1016/j.eswa.2023.120532

– Huang, K., Hong, S., Shi, Z., & Jiang, H. (2023). A Novel Reservation and Allocation Approach of Shared Parking Slots considering the Noncritical Aisle Space. Journal of Advanced Transportation, 2023, 1–18. https://doi.org/10.1155/2023/5081016

– Deng, D. (2021). Dynamic Pricing for Predictive Analytics in Parking (Doctoral dissertation, The University of Manitoba).

– Saharan, S., Kumar, N., & Bawa, S. (2020). An efficient smart parking pricing system for smart city environment: A machine-learning based approach. Future Generation Computer Systems, 106, 622–640. https://doi.org/10.1016/j.future.2020.01.031

– Tamim Kashifi, M., Jamal, A., Samim Kashefi, M., Almoshaogeh, M., & Masiur Rahman, S. (2022). Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. Travel Behaviour and Society, 29, 279–296. https://doi.org/10.1016/J.TBS.2022.07.003

– Pirra, M., & Diana, M. (2019). A study of tour-based mode choice based on a Support Vector Machine classifier. Transportation

– Loukopoulos, P., Jakobsson, C., Gärling, T., Schneider, C. M., & Fujii, S. (2004). Car-user responses to travel demand management measures: Goal setting and choice of adaptation alternatives. Transportation Research Part D: Transport and Environment, 9(4), 263–280. https://doi.org/10.1016/j.trd.2004.02.003

– Tuveri, G. (2023). Novel Neural Network Applications to Mode Choice in Transportation: Estimating Value of Travel Time and Modelling Psycho-Attitudinal Factors. UNICA University, PhD Thesis.

– Hillel, T., Bierlaire, M., Elshafie, M. Z., & Jin, Y. (2021). A systematic review of machine learning classification methodologies for modelling passenger mode choice. Journal of choice modelling, 38, 100221.

– Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. Travel Behaviour and Society, 20, 22–35. https://doi.org/10.1016/j.tbs.2020.02.003

– Wang, S., Mo, B., Hess, S., & Zhao, J. (2021). Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: an empirical benchmark. https://arxiv.org/abs/2102.01130v1

– Mohd Ali, N. F., Mohd Sadullah, A. F., P.P. Abdul Majeed, A., Mohd Razman, M. A., & Muazu Musa, R. (2020). Mode Choice Prediction using Machine Learning Technique for A Door-to-Door Journey in Kuantan City. Mekatronika, 2(1), 73–78. https://doi.org/10.15282/mekatronika.v2i1.6745

– Chapleau, R., Gaudette, P., & Spurr, T. (2019). Application of Machine Learning to Two Large-Sample Household Travel Surveys:

Planning and Technology, 42(1), 23–36. https://doi.org/10.1080/03081060.2018.1541280

– Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. Travel Behaviour and Society, 14, 1–10. https://doi.org/10.1016/J.TBS.2018.09.002

– Chapleau, R., Gaudette, P., & Spurr, T. (2019). Application of Machine Learning to Two Large-Sample Household Travel Surveys: A Characterization of Travel Modes. Https://Doi.Org/10.1177/0361198119839339, 2673(4), 173–183. https://doi.org/10.1177/0361198119839339

– Lee, D., Derrible, S., & Pereira, F. C. (2018). Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. Transportation Research Record, 2672(49), 101–112. https://doi.org/10.1177/0361198118796971

– Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification And Regression Trees. Routledge. https://doi.org/10.1201/9781315139470

– Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140. https://doi.org/10.1007/BF00058655