فصلنامه مهندسی ترافیک، سال بیست و پنجم، شماره ۱۰۱، تابستان ۱۴۰۴، ص. ۶۰-۴۵

# پیش‌بینی تقاضای دوچرخه‌های اشتراکی با استفاده از مدل‌های یادگیری ماشین: مقایسه عملکرد و تحلیل عوامل مؤثر

علی‌رضا مهین‌اقدم، کارشناس ارشد برنامه‌ریزی حمل‌ونقل، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

علی ادریسی (مسئول مکاتبات)، دکتری برنامه‌ریزی حمل‌ونقل، دانشیار دانشکده عمران، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

**E-mail: Edrisi@kntu.ac.ir**

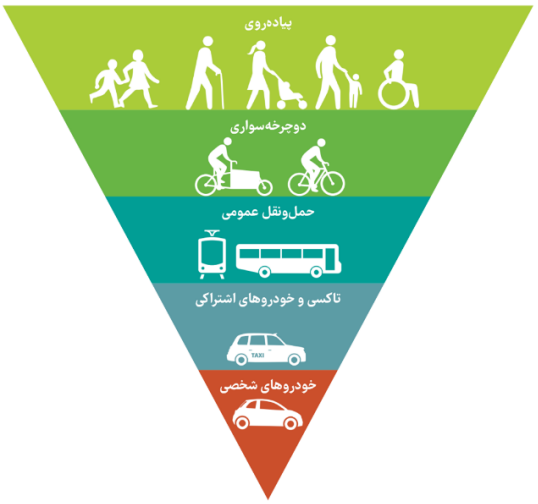
## چکیده

***سامانه‌های دوچرخه‌های اشتراکی به‌عنوان یکی از راهکارهای حمل‌ونقل پایدار، نقش مهمی در کاهش ترافیک، آلودگی هوا و مصرف سوخت ایفا می‌کنند. پیش‌بینی دقیق تقاضای این سامانه‌ها برای بهینه‌سازی تخصیص منابع و بهبود عملکرد آن‌ها ضروری است. این پژوهش با هدف بررسی و مقایسه مدل‌های یادگیری ماشین در پیش‌بینی تقاضای دوچرخه‌های اشتراکی انجام شده است. داده‌های مورداستفاده از سامانه Bluebikes در شهر بوستون آمریکا و منابع هواشناسی جمع‌آوری شده‌اند. جامعه آماری شامل ۴۷۵۱۷۹۰ سفر ثبت‌شده در سال ۲۰۲۴ است. متغیرهای موردبررسی شامل اطلاعات سفرها (زمان، مبدأ، مقصد، مدت سفر و نوع کاربر)، شرایط آب‌وهوایی (دما، بارش، سرعت باد و رطوبت)، ویژگی‌های زمانی (ساعات شبانه‌روز، روزهای هفته و فصول) و سطح دسترسی به ایستگاه‌ها هستند. پنج مدل یادگیری ماشین شامل شبکه عصبی مصنوعی (ANN)، ایکس‌جی‌بوست (XGBoost)، جنگل تصادفی (RF)، رگرسیون بردار پشتیبان (SVR) و رگرسیون خطی (LR) برای تحلیل داده‌ها به کار گرفته شدند. ارزیابی عملکرد مدل‌ها با معیارهای ضریب تعیین ($R^2$)، میانگین خطای مطلق (MAE)، میانگین مربعات خطا (MSE) و جذر میانگین مربعات خطا (RMSE) انجام شد. نتایج نشان داد که مدل‌های ANN و XGBoost عملکرد بهتری در پیش‌بینی تقاضا داشته و دقت بالاتری نسبت به سایر مدل‌ها ارائه کردند. همچنین، شرایط آب‌وهوایی، زمان سفر و سطح دسترسی به ایستگاه‌های پرتردد از مهم‌ترین عوامل تأثیرگذار بر میزان استفاده از دوچرخه شناخته شدند. این پژوهش بر اهمیت استفاده از روش‌های پیشرفته یادگیری ماشین در بهینه‌سازی تخصیص دوچرخه و مدیریت کارآمدتر سامانه‌های اشتراکی تأکید دارد. یافته‌های این مطالعه می‌توانند به سیاست‌گذاران در بهبود برنامه‌ریزی حمل‌ونقل پایدار شهری و توسعه زیرساخت‌های مرتبط کمک کنند.***

**واژه‌های کلیدی:** حمل‌ونقل پایدار، دوچرخه اشتراکی، پیش‌بینی تقاضا، یادگیری ماشین

## ۱. مقدمه

سیستم‌های دوچرخه‌های اشتراکی به‌عنوان یکی از رویکردهای نوین حمل‌ونقل شهری، امکان استفاده عمومی از دوچرخه را بدون نیاز به مالکیت شخصی فراهم می‌کنند. این سیستم‌ها نه‌تنها وابستگی به خودروهای شخصی را کاهش می‌دهند، بلکه در بهبود کیفیت هوا، کاهش تراکم ترافیکی و ارتقای سلامت عمومی نیز نقش مؤثری دارند. در چارچوب حمل‌ونقل پایدار (شکل ۱) که به حداقل رساندن اثرات منفی زیست‌محیطی و اجتماعی جابه‌جایی تأکید دارد، دوچرخه‌سواری به دلیل عدم انتشار آلاینده‌ها و کاهش مصرف سوخت‌های فسیلی جایگاه ویژه‌ای دارد.




**شکل ۱. هرم حمل‌ونقل پایدار**

مطالعات نشان داده‌اند که سیستم‌های اشتراک دوچرخه تأثیرات مثبتی بر محیط‌زیست و مصرف انرژی داشته‌اند. به‌عنوان‌مثال، در نیویورک بین سال‌های ۲۰۱۴ تا ۲۰۱۷، این سیستم‌ها موجب صرفه‌جویی ۱۳۳۷۰ تن معادل نفت و کاهش ۳۰۰۷۰ تن دی اکسید کربن و ۸۰ تن اکسیدهای نیتروژن شده‌اند. در بسیاری از شهرهای جهان، این سیستم‌ها جایگزینی برای وسایل نقلیه موتوری محسوب شده‌اند. در برخی از شهرها مانند بریزبن (۲۱٪)، ملبورن (۱۹٪)، مینیاپولیس–سنت پل (۱۹٪)، واشنگتن دی‌سی (۷٪) و لندن (۲٪)، بخشی از سفرهای درون‌شهری از خودروهای شخصی به دوچرخه تغییر یافته است. این تغییرات نه‌تنها منجر به کاهش مصرف سوخت‌های فسیلی شده، بلکه اثرات زیست‌محیطی مرتبط را نیز کاهش داده است.

از منظر اقتصادی، دوچرخه‌سواری به‌عنوان یک شیوه حمل‌ونقل کم‌هزینه و پایدار، نه‌تنها هزینه‌های شخصی سفر را کاهش می دهد، بلکه توسعه زیرساخت‌های دوچرخه‌سواری می‌تواند به ایجاد اشتغال در بخش‌های مرتبط با تولید، نگهداری و مدیریت سیستم‌های اشتراکی کمک کند. همچنین، کاهش آلودگی هوا و ارتقای سلامت عمومی منجر به کاهش هزینه‌های درمانی مرتبط با بیماری‌های تنفسی و قلبی می‌شود. از بُعد سلامت، دوچرخه‌سواری به‌عنوان یک رفتار حرکتی مثبت، تأثیرات قابل‌توجهی در کاهش بروز بیماری‌های ناشی از سبک زندگی کم‌تحرک، کاهش تنش‌های روانی و بهبود شاخص‌های سلامت روان دارد. علاوه بر این، توسعه شبکه‌های دوچرخه‌سواری در محیط‌های شهری می‌تواند به کاهش تراکم ترافیکی و بهبود کیفیت زندگی منجر شود.

با وجود مزایای متعدد، اجرای سیستم‌های دوچرخه‌های اشتراکی با چالش‌هایی همراه است. از مهم‌ترین چالش‌ها می‌توان به محدودیت‌های زیرساختی، توزیع نامتناسب ایستگاه‌ها در سطح شهر و مسائل مرتبط با ایمنی دوچرخه‌سواران اشاره کرد. غلبه بر این چالش‌ها مستلزم سیاست‌گذاری‌های جامع، سرمایه‌گذاری هدفمند در توسعه مسیرهای دوچرخه‌سواری و ترویج فرهنگ استفاده از دوچرخه است.

سیستم‌های دوچرخه‌های اشتراکی یکی از ابزارهای کلیدی در تحقق حمل‌ونقل پایدار شهری محسوب می‌شوند و می‌توانند به کاهش آلودگی هوا، کاهش هزینه‌های حمل‌ونقل و ارتقای سلامت عمومی کمک کنند. موفقیت این سیستم‌ها نیازمند برنامه‌ریزی دقیق، سرمایه‌گذاری در زیرساخت‌های مرتبط و تدوین سیاست‌های حمایتی برای افزایش مقبولیت عمومی است. توسعه و ترویج استفاده از دوچرخه‌های اشتراکی می‌تواند گامی مؤثر در جهت ایجاد شهرهای پایدار و زیست‌پذیرتر باشد.

## ۲. تعریف مسئله و اهداف پژوهش

سامانه‌های دوچرخه‌های اشتراکی به‌عنوان یکی از راهکارهای پایدار حمل‌ونقل شهری، جایگزینی مناسب برای وسایل نقلیه موتوری محسوب می‌شوند. بااین‌حال، تقاضا برای این سامانه‌ها تحت تأثیر عوامل متعددی قرار دارد که پیش‌بینی دقیق آن را به یک چالش اساسی تبدیل کرده است. درک این عوامل و توسعه یک مدل پیش‌بینی کننده کارآمد می‌تواند به بهینه‌سازی مدیریت ناوگان، بهبود دسترسی کاربران و افزایش رضایت آن‌ها کمک کند.

یکی از عوامل کلیدی مؤثر بر تقاضای دوچرخه‌های اشتراکی، سیاست‌های تشویقی است. به‌عنوان‌مثال، ارائه تخفیف در هزینه اشتراک برای کاربران عضو در مقایسه با افراد غیر عضو، می‌تواند تأثیر قابل‌توجهی بر میزان استفاده از این خدمات داشته باشد. بررسی میزان این تأثیر می‌تواند به طراحی استراتژی‌های قیمت‌گذاری کارآمدتر کمک کند.

علاوه بر این، افزودن دوچرخه‌های برقی به ناوگان دوچرخه‌های اشتراکی، بُعد جدیدی به پیش‌بینی تقاضا می‌افزاید. دوچرخه‌های برقی نیاز به‌صرف انرژی فیزیکی را کاهش داده و دامنه استفاده از این خدمات را برای گروه‌های گسترده‌تری از کاربران، به‌ویژه در مناطق دارای شیب زیاد، افزایش می‌دهند. بررسی تأثیر حضور دوچرخه‌های برقی بر میزان تقاضا، می‌تواند نقش مهمی در برنامه‌ریزی توسعه ناوگان ایفا کند.

عامل دیگر تأثیرگذار بر تقاضا، شرایط آب‌وهوایی است. متغیرهایی مانند دما، بارش، رطوبت و سرعت باد، به‌طور مستقیم بر میزان استفاده از دوچرخه‌های اشتراکی تأثیر می‌گذارند. شناسایی دقیق‌ترین شرایط آب‌وهوایی که موجب افزایش یا کاهش تقاضا می‌شوند، برای برنامه‌ریزی عملیاتی و مدیریت سیستم حیاتی است.

روزهای کاری و تعطیلات نیز ازجمله عواملی هستند که الگوهای استفاده از دوچرخه‌های اشتراکی را تحت تأثیر قرار می دهند. تقاضا در روزهای کاری که بیشتر بر سفرهای کاری و دانشگاهی متمرکز است، با الگوی استفاده در تعطیلات و آخر هفته که بیشتر جنبه تفریحی دارد، تفاوت دارد. شناخت این تفاوت‌ها می‌تواند به بهینه‌سازی توزیع دوچرخه‌ها و مدیریت ظرفیت کمک کند.

همچنین، فصل‌ها و ماه‌های سال تغییرات تقاضا را پیچیده‌تر می کنند. تغییرات دما، طول روز و رفتارهای فصلی سفر، بر میزان استفاده از دوچرخه‌های اشتراکی اثرگذار هستند. در نظر گرفتن این الگوهای فصلی برای تدوین برنامه‌های راهبردی در بلندمدت ضروری است.

با توجه به این عوامل تأثیرگذار، هدف این پژوهش یافتن یک مدل یادگیری ماشین بهینه برای پیش‌بینی تقاضای دوچرخه‌های اشتراکی است. در این راستا، چندین مدل یادگیری ماشین[۱] مقایسه شده و بر اساس معیارهای ارزیابی شامل ضریب تعیین[۲] ($R^2$)، میانگین خطای مطلق[۳] (MAE)، میانگین مربعات خطا[۴] (MSE) و جذر میانگین مربعات خطا[۵] (RMSE)، عملکرد آن‌ها بررسی می‌شود. یافته‌های این پژوهش می‌تواند به بهبود بهره‌وری سامانه‌های دوچرخه اشتراکی، تدوین سیاست‌های کارآمدتر و توسعه حمل‌ونقل پایدار شهری کمک کند.

## ۳. ادبیات پژوهش

پیش‌بینی تقاضای دوچرخه‌های اشتراکی یکی از مسائل کلیدی در مدیریت و برنامه‌ریزی سیستم‌های حمل‌ونقل شهری است. افزایش روزافزون استفاده از این سیستم‌ها، به‌ویژه در شهرهای بزرگ، نیاز به مدل‌های دقیق پیش‌بینی را بیشتر کرده است. در دهه‌های اخیر، روش‌های متنوعی برای پیش‌بینی تقاضا توسعه یافته‌اند که از مدل‌های سنتی اقتصادسنجی[۶] تا تکنیک‌های پیشرفته یادگیری ماشین را شامل می‌شوند. در این بخش، تاریخچه و توسعه سیستم‌های دوچرخه اشتراکی بررسی شده و سپس به روش‌های سنتی و مدل‌های یادگیری ماشین مورداستفاده در این حوزه پرداخته می‌شود.

### ۳-۱ پیشینه سیستم‌های دوچرخه‌های اشتراکی

سیستم‌های دوچرخه اشتراکی به‌عنوان یکی از راهکارهای کارآمد در توسعه حمل‌ونقل پایدار، در دهه‌های اخیر موردتوجه بسیاری از کلان‌شهرهای جهان قرار گرفته‌اند. این سیستم‌ها با تسهیل دسترسی شهروندان به دوچرخه و کاهش وابستگی به وسایل نقلیه شخصی، نقش مهمی در کاهش تراکم ترافیکی، کاهش آلودگی هوا و افزایش تحرک‌پذیری ایفا می‌کنند. نخستین نمونه‌های سیستم‌های دوچرخه اشتراکی در دهه ۱۹۶۰ میلادی در برخی از شهرهای اروپا راه‌اندازی شدند، اما این سامانه‌های اولیه به دلیل فقدان فناوری‌های مناسب درزمینه مدیریت و نظارت، با چالش‌هایی ازجمله سرقت دوچرخه‌ها و توزیع نامتوازن آن‌ها مواجه بودند.

تحولات فناورانه در دهه‌های اخیر منجر به توسعه سیستم‌های دوچرخه اشتراکی نسل جدید شده است که از قابلیت‌هایی همچون فناوری جی‌پی‌اس[۷]، داده‌های بلادرنگ و اپلیکیشن‌های موبایلی بهره می‌برند. این پیشرفت‌ها امکان مدیریت هوشمند ناوگان دوچرخه‌ها، بهینه‌سازی تخصیص منابع و بهبود تجربه کاربران را فراهم کرده است. همچنین، مطالعات متعددی نشان داده‌اند که پیش‌بینی تقاضای دوچرخه‌های اشتراکی نقش کلیدی در بهینه‌سازی این سیستم‌ها ایفا می‌کند. با استفاده از مدل‌های پیش‌بینی دقیق، می‌توان توزیع بهینه دوچرخه‌ها در ایستگاه‌های مختلف، مدیریت مؤثرتر ناوگان و کاهش مشکلاتی مانند عدم دسترسی یا ازدحام بیش‌ازحد در برخی ایستگاه‌ها را امکان‌پذیر ساخت است.

## ۲-۳ روش‌های سنتی پیش‌بینی تقاضای حمل‌ونقل

پیش‌بینی تقاضای حمل‌ونقل به‌طور سنتی مبتنی بر مدل‌های سری زمانی[۸]، رگرسیون خطی[۹] و روش‌های اقتصادسنجی بوده است. مدل‌های سری زمانی مانند ARIMA[۱۰] و SARIMA[۱۱] به‌طور گسترده در پیش‌بینی تقاضای حمل‌ونقل استفاده شده‌اند. این مدل‌ها بر اساس تحلیل داده‌های تاریخی و شناسایی الگوهای تکرارشونده در داده‌ها عمل می‌کنند. ARIM ترکیبی از سه جزء اصلی است: خودرگرسیون[۱۲] (AR)، تفاضل‌گیری[۱۳] (I) برای حذف غیرایستایی و میانگین متحرک[۱۴] (MA). SARIMA نیز نسخه فصلی ARIMA است که برای مدل‌سازی داده‌های با الگوهای فصلی استفاده می‌شود. این مدل‌ها به دلیل سادگی و سرعت اجرا، در بسیاری از مطالعات مورداستفاده قرار گرفته‌اند، اما در مواجهه با داده‌های غیرخطی و پیچیده عملکرد ضعیف‌تری دارند.

رگرسیون خطی و چندمتغیره[۱۵] نیز از روش‌های سنتی پرکاربرد در پیش‌بینی تقاضای حمل‌ونقل هستند. این روش‌ها بر اساس رابطه بین متغیرهای مستقل (مانند جمعیت، درآمد، قیمت سوخت و ...) و متغیر وابسته (تقاضای حمل‌ونقل) عمل می‌کنند. رگرسیون خطی ساده برای مدل‌سازی رابطه خطی بین متغیرها استفاده می‌شود، درحالی‌که رگرسیون چندمتغیره امکان مدل‌سازی روابط پیچیده‌تر بین چندین متغیر مستقل و متغیر وابسته را فراهم می‌کند. این روش‌ها به دلیل سادگی و قابلیت تفسیر، در بسیاری از مطالعات مورداستفاده قرار گرفته‌اند، اما در مواجهه با داده‌های غیرخطی و پیچیده ممکن است نتایج دقیقی ارائه ندهند.

روش‌های اقتصادسنجی نیز در پیش‌بینی تقاضای حمل‌ونقل مورداستفاده قرار گرفته‌اند. این روش‌ها شامل مدل‌هایی مانند VAR[۱۶] و VECM[۱۷] هستند که برای مدل‌سازی روابط پویا بین چندین متغیر اقتصادی استفاده می‌شوند. این مدل‌ها به دلیل توانایی در مدل‌سازی روابط پیچیده بین متغیرها و در نظر گرفتن اثرات متقابل بین آن‌ها، در پیش‌بینی تقاضای حمل‌ونقل در سطح کلان مورداستفاده قرار گرفته‌اند. بااین‌حال، این روش‌ها نیز وابستگی زیادی به داده‌های تاریخی دارند و ممکن است در مواجهه با تغییرات ناگهانی یا شرایط غیرمنتظره (مانند همه‌گیری کووید-۱۹) عملکرد ضعیف‌تری داشته باشند.

هرچند مدل‌های سنتی در برخی موارد عملکرد رضایت‌بخشی دارند، اما در شرایطی که داده‌ها پیچیدگی بالایی داشته باشند، نوسانات غیرمنتظره رخ دهد یا روابط غیرخطی بین متغیرها وجود داشته باشد، کارایی این روش‌ها کاهش می‌یابد. این

محدودیت‌ها موجب شده است که روش‌های مبتنی بر یادگیری ماشین به‌عنوان جایگزینی مناسب برای مدل‌های سنتی مطرح شوند. این روش‌ها با بهره‌گیری از قابلیت‌هایی نظیر شناسایی الگوهای پنهان در داده‌ها، قدرت پردازش حجم بالای اطلاعات و انعطاف‌پذیری در مواجهه با تغییرات ناگهانی، دقت پیش‌بینی را به میزان قابل‌توجهی افزایش داده‌اند.

## ۳-۳ کاربرد یادگیری ماشین در پیش‌بینی تقاضای حمل‌ونقل

یادگیری ماشین به دلیل توانایی در تحلیل حجم بالای داده و شناسایی روابط پیچیده، به‌طور گسترده در پیش‌بینی تقاضای حمل‌ونقل مورداستفاده قرار گرفته است. مدل‌هایی مانند جنگل تصادفی[۱۸]، رگرسیون بردار پشتیبان[۱۹]، شبکه‌های عصبی مصنوعی[۲۰] و الگوریتم‌های تقویت گرادیان[۲۱] (مانند ایکس‌جی‌بوست[۲۲]) ازجمله روش‌های رایج در این حوزه هستند. مطالعات متعددی نشان داده‌اند که بهره‌گیری از این مدل‌ها می‌تواند موجب کاهش خطا و افزایش دقت در پیش‌بینی الگوهای تقاضا شود که این امر نقش مهمی در بهینه‌سازی فرآیندهای برنامه‌ریزی و مدیریت حمل‌ونقل ایفا می‌کند.

مطالعات متعددی نشان داده‌اند که مدل جنگل تصادفی، در مقایسه با رگرسیون خطی، عملکرد بهتری در پیش‌بینی تقاضای دوچرخه‌های اشتراکی ارائه داده است. به‌عنوان نمونه، در پژوهشی، این مدل بر اساس معیارهای میانگین مربعات خطا، میانگین خطای مطلق و ضریب همبستگی[۲۳] ارزیابی شده و نسبت به روش‌های سنتی دقت بالاتری داشته است. همچنین، تحقیقات دیگر نشان داده‌اند که استفاده از یادگیری ماشین برای پیش‌بینی تعداد دوچرخه‌های اشتراکی موردنیاز در هر ساعت در شهرهایی مانند سئول، منجر به نتایج قابل‌توجهی شده است. افزون بر این، تحلیل الگوهای تحرک سفرهای اشتراکی[۲۴] و تأثیر مکان‌های جایگزین بر انتخاب مقصد مسافران نیز با موفقیت از طریق این مدل‌ها انجام شده است. استفاده از یادگیری ماشین برای پیش‌بینی الگوهای سفر و ارائه اطلاعات به رانندگان و مراکز حمل‌ونقل از دیگر کاربردهای این روش‌ها است.

علاوه بر حمل‌ونقل اشتراکی، روش‌های یادگیری ماشین در پیش‌بینی تقاضای سایر مدهای حمل‌ونقل نیز موفق بوده‌اند. به‌عنوان‌مثال، رگرسیون بردار پشتیبان در پیش‌بینی تقاضای حمل ونقل هوایی، جاده‌ای و ریلی بر اساس معیار میانگین درصد خطای مطلق[۲۵] (MAPE) عملکرد بهتری نسبت به مدل‌های خطی داشته است. علاوه بر این، پژوهش‌ها نشان داده‌اند که شبکه عصبی مصنوعی و ماشین بردار پشتیبان، با معیارهای جذر میانگین مربعات خطا و ضریب تعیین ($R^2$)، دقت بیشتری در مقایسه با مدل‌هایی مانند رگرسیون خطی چندگانه و درخت رگرسیون داشته‌اند. همچنین، در پیش‌بینی تقاضای اتوبوس در مناطق روستایی، مدل ایکس‌جی‌بوست، بر اساس معیارهای دقت[۲۶]، صحت[۲۷]، حساسیت[۲۸] و معیار F1، عملکرد بهتری نسبت به جنگل تصادفی و الگوریتم بیز ساده[۲۹] داشته است. از سوی دیگر، مدل‌های یادگیری ماشین در بهبود مدیریت ترافیک و پارکینگ‌های هوشمند نیز نقش مؤثری ایفا کرده‌اند.

افزون بر این، مطالعات مرتبط با پیش‌بینی تقاضای مسافر در سیستم‌های ریلی از مدل‌های متنوعی مانند شبکه‌های عصبی مصنوعی، درخت تصمیم، رگرسیون خطی، جنگل تصادفی، رگرسیون چندجمله‌ای و رگرسیون بردار پشتیبان استفاده کرده‌اند. یافته‌های این مطالعات نشان داده است که درخت تصمیم نسبت به سایر مدل‌ها عملکرد بهتری در پیش‌بینی تقاضای مسافری در شبکه‌های ریلی ارائه داده است.

علاوه بر این، مطالعات اخیر تأکید کرده‌اند که مدل‌های مبتنی بر تقویت گرادیان می‌توانند در بهبود مدیریت ترافیک شهری و کاهش تراکم ترافیکی مؤثر باشند. به‌ویژه، این روش که بر اساس معیار میانگین مربعات خطا ارزیابی شده است، دقت بالایی در تحلیل داده‌های ترافیکی و بهینه‌سازی کنترل سیگنال‌های ترافیکی نشان داده است.

نتایج این مطالعات نشان می‌دهد که مدل‌های یادگیری ماشین، به‌ویژه روش‌های غیرخطی، می‌توانند با کاهش میزان خطا و افزایش دقت پیش‌بینی، به بهینه‌سازی فرآیندهای مدیریت حمل‌ونقل کمک کنند. این یافته‌ها بر ضرورت به‌کارگیری مدل‌های پیشرفته برای توسعه سیستم‌های هوشمند حمل‌ونقل و ارتقای کارایی زیرساخت‌های حمل‌ونقل شهری و بین‌شهری تأکید دارند.

**جدول ۱. خلاصه پژوهش‌های مرتبط با یادگیری ماشین در پیش‌بینی تقاضای حمل‌ونقل**

| شماره مرجع | مدل‌های مورداستفاده | معیارهای ارزیابی |
|---|---|---|
| [۱۸] | رگرسیون خطی – جنگل تصادفی | میانگین مربعات خطا – میانگین خطای مطلق – ضریب همبستگی |
| [۱۹] | رگرسیون خطی – ماشین گرادیان بوست – ماشین بردار پشتیبان – درختان بوست – درختان گرادیان بوست افراطی | ضریب تعیین – جذر میانگین مربعات خطا – میانگین خطای مطلق – ضریب تغییرات |
| [۲۰] | رگرسیون بردار پشتیبان – جنگل تصادفی – درخت افزایش گرادیان – ایکس‌جی‌بوست | ضریب همبستگی پیرسون – میانگین خطای مطلق – میانگین درصد خطای مطلق – جذر میانگین مربعات خطا |
| [۲۱] | درخت تصمیم – جنگل تصادفی – شبکه عصبی مصنوعی – K – نزدیک‌ترین همسایه – رگرسیون لجستیک | دقت متوسط – انحراف معیار |
| [۲۲] | رگرسیون خطی – رگرسیون بردار پشتیبان | میانگین درصد خطای مطلق |
| [۲۳] | رگرسیون خطی چندگانه – الگوریتم ژنتیک – شبکه عصبی مصنوعی – ماشین بردار پشتیبان – درخت رگرسیون | جذر میانگین مربعات خطا – ضریب تعیین |
| [۲۴] | ایکس‌جی‌بوست – جنگل تصادفی – الگوریتم بیز ساده | دقت – معیار صحت – حساسیت – معیار F1 |
| [۲۵] | شبکه عصبی مصنوعی – تقویت گرادیان – ایکس‌جی‌بوست | دقت – ضریب همبستگی پیرسون – جذر میانگین مربعات خطا |
| [۲۶] | شبکه عصبی مصنوعی – ماشین بردار پشتیبان – درخت تصمیم – رگرسیون خطی – رگرسیون چندجمله‌ای – رگرسیون بردار پشتیبان – جنگل تصادفی | میانگین خطای مطلق – میانگین مربعات خطا – میانگین درصد خطای مطلق – جذر میانگین مربعات خطا |
| [۲۷] | تقویت گرادیان | میانگین مربعات خطا |

## ۴. روش پژوهش

در این بخش، فرآیند کلی انجام پژوهش، از جمع‌آوری داده‌ها تا ارزیابی مدل‌های یادگیری ماشین، تشریح می‌شود. هدف اصلی، طراحی و پیاده‌سازی مدل‌هایی است که بتوانند تقاضای دوچرخه‌های اشتراکی را با دقت بالایی پیش‌بینی کنند. ابتدا داده‌های مورداستفاده معرفی شده و مراحل پیش‌پردازش آن‌ها شرح داده می‌شود. سپس، الگوریتم‌های یادگیری ماشین انتخابی به‌تفصیل بررسی‌شده و در ادامه، معیارهای ارزیابی مدل‌ها ارائه می‌گردد. این چارچوب روشی سیستماتیک برای تحلیل داده‌ها و انتخاب مدل بهینه فراهم می‌کند.

### ۴-۱ جمع‌آوری و پیش‌پردازش داده‌ها

در فرآیند پیش‌بینی تقاضای دوچرخه‌های اشتراکی، جمع‌آوری و آماده‌سازی داده‌ها از اهمیت ویژه‌ای برخوردار است. کیفیت داده‌های ورودی نقش تعیین‌کننده‌ای در عملکرد مدل‌های یادگیری ماشین دارد. این بخش شامل منابع داده، مراحل پیش‌پردازش و انتخاب ویژگی‌های کلیدی برای مدل‌سازی است که به درک بهتر روابط میان متغیرهای تأثیرگذار بر تقاضای دوچرخه کمک می‌کند.

### ۴-۱-۱ منابع داده

داده‌های مورداستفاده در این پژوهش از چندین منبع مختلف جمع‌آوری شده‌اند. یکی از منابع اصلی داده، اطلاعات سیستم دوچرخه‌های اشتراکی Bluebikes در شهر بوستون، ایالت ماساچوست، ایالات متحده آمریکا است. بوستون یکی از شهرهای پرجمعیت و تاریخی ایالات‌متحده بوده و به دلیل ساختار شهری متراکم، زیرساخت‌های حمل‌ونقل عمومی گسترده‌ای دارد. سیستم دوچرخه‌های اشتراکی Bluebikes که از سال ۲۰۱۱ فعالیت خود را آغاز کرده است، در حال حاضر دارای بیش از ۴۰۰ ایستگاه و ۴۰۰۰ دوچرخه در مناطق بوستون، کمبریج، سامرویل و بروکلاین است. داده‌های این پژوهش شامل اطلاعات سفرهای انجام‌شده در سال ۲۰۲۴ بوده و متغیرهایی مانند تعداد سفرها، ایستگاه‌های مبدأ و مقصد، مدت‌زمان سفر، نوع اشتراک کاربران (کاربران عضو یا کاربران موقت)، زمان و تاریخ هر سفر و الگوی استفاده از دوچرخه در ساعات مختلف شبانه‌روز را پوشش می‌دهد. مجموعه داده موردبررسی در این مطالعه شامل ۴۷۵۱۷۹۰ سفر ثبت‌شده در سیستم دوچرخه اشتراکی است.

علاوه بر داده‌های عملیاتی Bluebikes، داده‌های جوی این شهر نیز برای بررسی تأثیر شرایط آب‌وهوایی بر تقاضای دوچرخه‌های اشتراکی جمع‌آوری شده است. این داده‌ها از طریق سرویس‌های هواشناسی دریافت شده‌اند و شامل متغیرهایی مانند دما (°C)، رطوبت نسبی (٪)، میزان بارندگی (mm)، سرعت باد (km/h)، میزان دید (km) و میزان تابش خورشید ($W/m^2$) است. داده‌های آب‌وهوایی برای یک بازه زمانی یک‌ساله، از ۱ ژانویه ۲۰۲۴ تا ۳۱ دسامبر ۲۰۲۴، به‌صورت روزانه و ساعتی جمع‌آوری شده‌اند تا تغییرات فصلی و روزانه در میزان تقاضای دوچرخه بررسی شود.

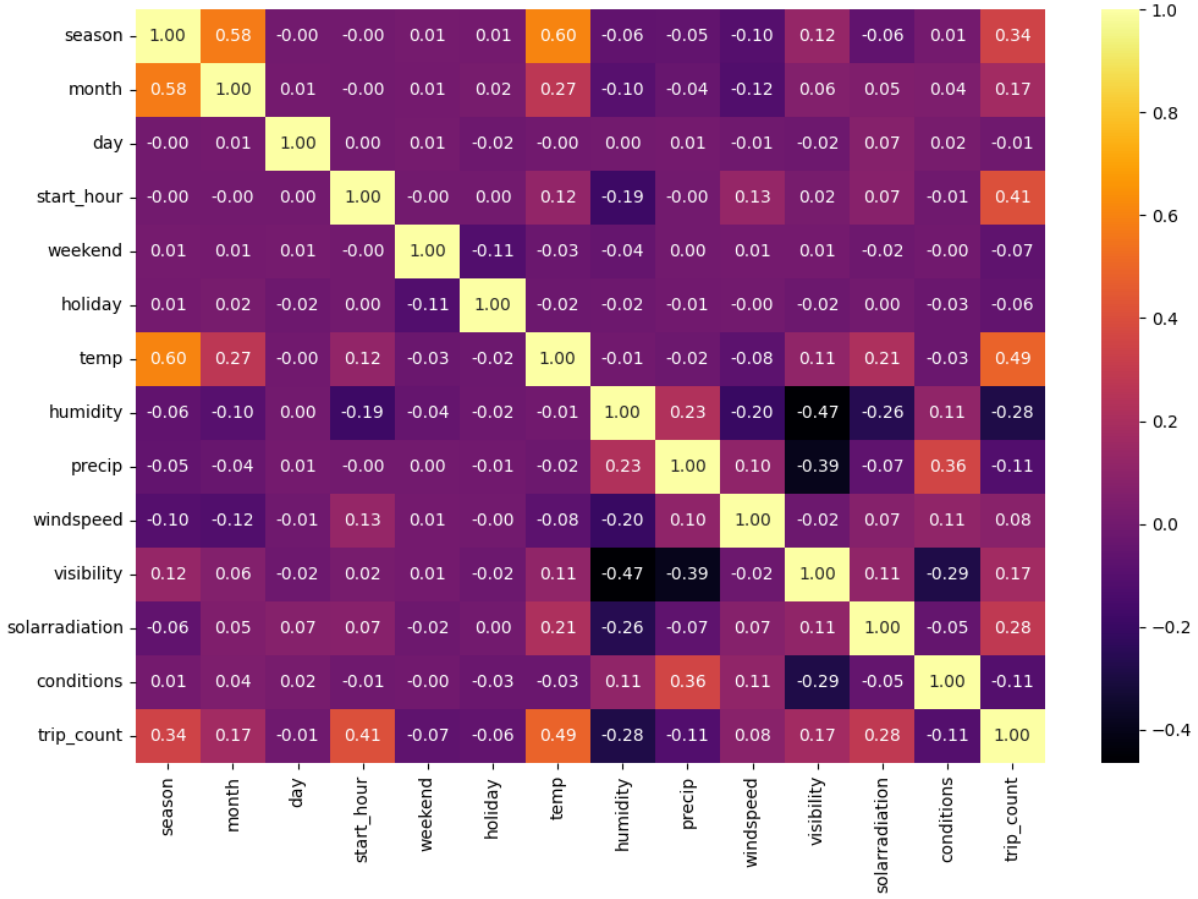
علاوه بر این، یک دیتاست سفارشی نیز برای تحلیل تأثیر روزهای تعطیل و آخر هفته بر میزان تقاضای دوچرخه‌های اشتراکی ایجاد شده است. این دیتاست شامل تمام تعطیلات رسمی ایالات متحده آمریکا و تعطیلات محلی شهر بوستون در سال ۲۰۲۴ است. همچنین، روزهای شنبه و یکشنبه به‌عنوان آخر هفته در این مجموعه داده لحاظ شده‌اند. این اطلاعات با هدف بررسی تغییرات الگوی استفاده از دوچرخه در روزهای تعطیل در مقایسه با روزهای کاری جمع‌آوری شده است.

### ۴-۱-۲ روش‌های پاک‌سازی داده‌ها

پس از جمع‌آوری داده‌ها، فرایند پیش‌پردازش به‌منظور افزایش دقت مدل‌های یادگیری ماشین انجام شده است. در این مرحله، ابتدا داده‌های پرت شناسایی و حذف شده‌اند؛ برای مثال، سفرهایی با زمان‌های غیرمنطقی (بسیار کوتاه یا بیش‌ازحد طولانی) کنار گذاشته شده‌اند. علاوه بر این، متغیرهای طبقه‌بندی شده نظیر نوع کاربر با روش کدبندی وان‌هات[۳۰] به داده‌های عددی تبدیل شده‌اند.

### ۴-۱-۳ ویژگی‌های انتخاب‌شده برای مدل‌سازی

در این پژوهش، از ماتریس همبستگی[۳۱] برای شناسایی متغیرهای تأثیرگذار بر تعداد سفرهای دوچرخه‌های اشتراکی استفاده شده است. ماتریس همبستگی یکی از روش‌های آماری متداول برای بررسی میزان وابستگی بین متغیرها است که مقادیر آن در بازه [۱ ، ۱-] قرار می‌گیرد. مقدار ۱ نشان‌دهنده همبستگی مثبت کامل، مقدار ۱- بیانگر همبستگی منفی کامل و مقدار ۰ (عدد صفر) به معنای عدم ارتباط بین دو متغیر است. تحلیل این ماتریس امکان انتخاب ویژگی‌های مؤثر در مدل‌سازی یادگیری ماشین را فراهم می‌کند و به بهبود دقت پیش‌بینی کمک می‌نماید.




**شکل ۲. ماتریس همبستگی**

نتایج تحلیل ماتریس همبستگی نشان می‌دهد که دمای هوا با مقدار همبستگی ۰/۴۹، ساعت شروع سفر با مقدار ۰/۴۱ و فصل با مقدار ۰/۳۴ از مهم‌ترین متغیرهای تأثیرگذار بر تعداد سفرها هستند. همچنین، رطوبت هوا و میزان بارش به ترتیب با ضرایب ۰/۲۸- و ۰/۱۱- تأثیر منفی بر میزان استفاده از دوچرخه دارند. این یافته‌ها تأکید می‌کند که متغیرهای زمانی و شرایط آب‌وهوایی نقش بسزایی در پیش‌بینی تقاضای دوچرخه‌های اشتراکی دارند و در فرایند مدل‌سازی باید موردتوجه قرار گیرند.

## ۴-۲ الگوریتم‌های یادگیری ماشین مورداستفاده

مدل‌های یادگیری ماشین برای پیش‌بینی تقاضای دوچرخه‌های اشتراکی باید بتوانند الگوهای پیچیده و روابط میان متغیرهای مختلف را شناسایی کنند. در این پژوهش، پنج الگوریتم یادگیری ماشین انتخاب شده‌اند که شامل مدل‌های خطی، درختی و مبتنی بر هسته، تقویت گرادیان و شبکه‌های عصبی مصنوعی هستند. این مدل‌ها به دلیل تنوع در رویکردهای پیش‌بینی، توانایی در مدل‌سازی روابط غیرخطی و عملکرد بهینه در داده‌های مختلف مورداستفاده قرار گرفته‌اند.

- رگرسیون خطی: یکی از ساده‌ترین و پایه‌ای‌ترین مدل‌های یادگیری ماشین که رابطه بین متغیرهای مستقل و وابسته را با یک معادله خطی مدل‌سازی می‌کند. این مدل برای تحلیل تأثیر ویژگی‌های مختلف بر میزان تقاضا مفید است و تفسیرپذیری بالایی دارد؛ اما در شرایطی که روابط میان متغیرها پیچیده و غیرخطی باشند، عملکرد ضعیفی خواهد داشت.
- جنگل تصادفی: یک روش یادگیری جمعی که از ترکیب چندین درخت تصمیم‌گیری برای افزایش دقت مدل استفاده می‌کند. این الگوریتم نسبت به داده‌های پرت مقاوم است و می‌تواند اهمیت ویژگی‌های مختلف را مشخص کند. جنگل تصادفی به دلیل تعادل میان دقت و تفسیرپذیری، یکی از

مدل‌های پرکاربرد در پیش‌بینی تقاضای حمل‌ونقل محسوب می‌شود. بااین‌حال، افزایش تعداد درخت‌ها باعث افزایش پیچیدگی محاسباتی مدل می‌شود.

- رگرسیون بردار پشتیبان: این مدل با استفاده از حداکثر سازی حاشیه[۳۲] برای پیش‌بینی، قادر به مدل‌سازی روابط غیرخطی میان متغیرها است. ماشین بردار پشتیبان از توابع کرنل[۳۳] برای شناسایی الگوهای پیچیده استفاده می‌کند. اگرچه دقت بالایی دارد، اما برای مجموعه داده‌های بزرگ، فرآیند آموزش آن زمان‌بر خواهد بود.
- تقویت گرادیان: یکی از روش‌های مبتنی بر درخت تصمیم‌گیری که با استفاده از تقویت گرادیان دقت پیش‌بینی را افزایش می‌دهد. ایکس‌جی‌بوست به دلیل پردازش سریع، تنظیم خودکار پارامترها و دقت بالا یکی از محبوب‌ترین مدل‌های یادگیری ماشین است.
- شبکه‌های عصبی مصنوعی: مدلی الهام گرفته از ساختار مغز انسان که توانایی یادگیری الگوهای پیچیده را دارد. این مدل از لایه‌های پردازشی شامل نرون‌های متصل‌به‌هم تشکیل شده و برای تحلیل داده‌های حجیم و غیرخطی بسیار مناسب است. تنظیم پارامترهای شبکه و نیاز به داده‌های گسترده برای آموزش، از چالش‌های اصلی این روش محسوب می‌شود.

### ۳-۴ روش‌های ارزیابی مدل‌ها

به‌منظور مقایسه مدل‌های یادگیری ماشین و تعیین مدل بهینه برای پیش‌بینی تقاضای دوچرخه‌های اشتراکی، از معیارهای ارزیابی کمی جهت سنجش عملکرد مدل‌ها استفاده شده است.

برای ارزیابی دقت مدل‌های یادگیری ماشین در پیش‌بینی تقاضای دوچرخه‌های اشتراکی، از چهار معیار عددی استفاده شده است. این معیارها میزان خطای مدل را از جنبه‌های مختلف بررسی کرده و به انتخاب مدل بهینه کمک می‌کنند. در ادامه، این معیارها همراه با فرمول ریاضی و توضیح پارامترهای مربوطه ارائه شده‌اند.

- ضریب تعیین ($R^2$): ضریب تعیین میزان تطابق پیش‌بینی های مدل با مقادیر واقعی را نشان می‌دهد و از رابطه زیر محاسبه می‌شود.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (۱)$$

$y_i$ دار واقعی تقاضای دوچرخه برای مشاهده‌ی $i$ ام است.

$\hat{y}_i$ دار پیش‌بینی‌شده توسط مدل برای مشاهده‌ی $i$ ام است.

$\bar{y}$ یانگین کل مقادیر واقعی تقاضا است.

$n$ عداد کل نمونه‌ها است.

مقدار $R^2$ بین ۰ تا ۱ متغیر است و هرچه مقدار آن به ۱ نزدیک‌تر باشد، مدل عملکرد بهتری دارد و پیش‌بینی‌های آن با مقادیر واقعی تطابق بیشتری دارد.

- میانگین خطای مطلق (MAE): این معیار میانگین اختلاف قدر مطلق بین مقادیر واقعی و پیش‌بینی‌شده را اندازه‌گیری می کند و به‌صورت زیر محاسبه می‌شود.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (۲)$$

$|y_i - \hat{y}_i|$ مق ر مطلق اختلاف بین مقدار واقعی و مقدار پیش‌بینی‌شده است.

مقدار کمتر این معیار نشان‌دهنده خطای کمتر مدل و دقت بالاتر آن در پیش‌بینی است.

- میانگین مربعات خطا (MSE): این معیار میانگین مربعات اختلافات بین مقادیر واقعی و پیش‌بینی‌شده را نشان می‌دهد و حساسیت بالایی به داده‌های پرت دارد. فرمول آن به‌صورت زیر است.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (۳)$$

اختلافات بین مقادیر واقعی و پیش‌بینی‌شده به توان دو می‌رسند، بنابراین مقادیر پرت تأثیر بیشتری بر مقدار نهایی خطا دارند. مقدار کمتر این معیار نشان‌دهنده عملکرد بهتر مدل است.

- جذر میانگین مربعات خطا (RMSE): این معیار ریشه‌ی دوم میانگین مربعات خطا است و مقدار خطا را در همان واحد

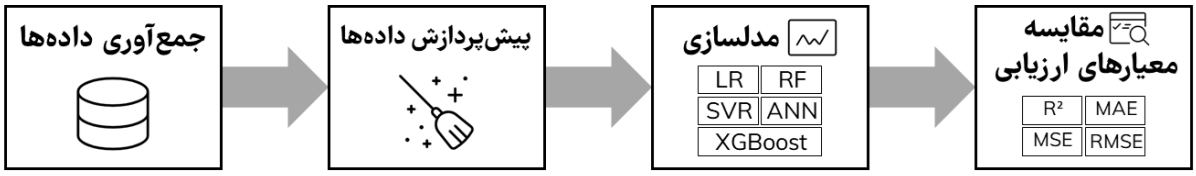
اصلی داده‌ها بیان می‌کند که باعث تفسیرپذیری بهتر آن می شود. فرمول آن به‌صورت زیر است.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (۴)$$

این معیار، مانند MSE، نسبت به داده‌های پرت حساس است اما مقدار آن به‌صورت مستقیم با مقیاس داده‌ها بیان می‌شود که درک آن را آسان‌تر می‌کند.

مقدار کمتر RMSE نشان‌دهنده دقت بالاتر مدل در پیش‌بینی است.

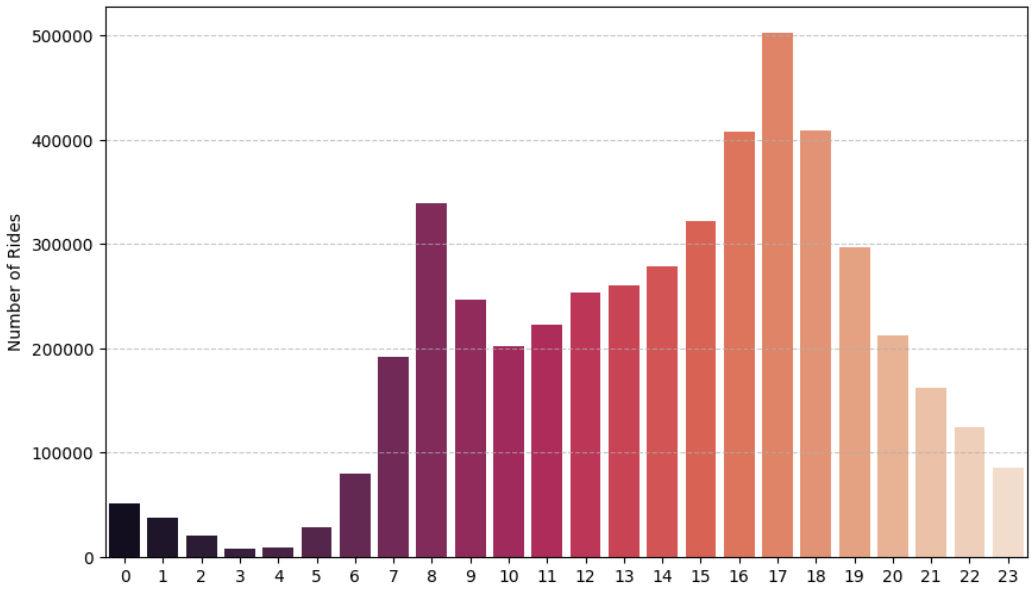
معیارهای فوق از جنبه‌های متفاوتی به ارزیابی مدل‌ها می‌پردازند. ضریب تعیین ($R^2$) نشان‌دهنده میزان تطابق مدل با داده‌های واقعی است، درحالی‌که MAE، MSE و RMSE مستقیماً میزان خطای مدل را اندازه‌گیری می‌کنند. MSE و RMSE به داده‌های پرت حساس‌تر هستند، درحالی‌که MAE میانگین کلی خطاها را ارائه می‌دهد. ترکیب این معیارها امکان ارزیابی دقیق‌تر عملکرد مدل‌ها را فراهم می‌کند.



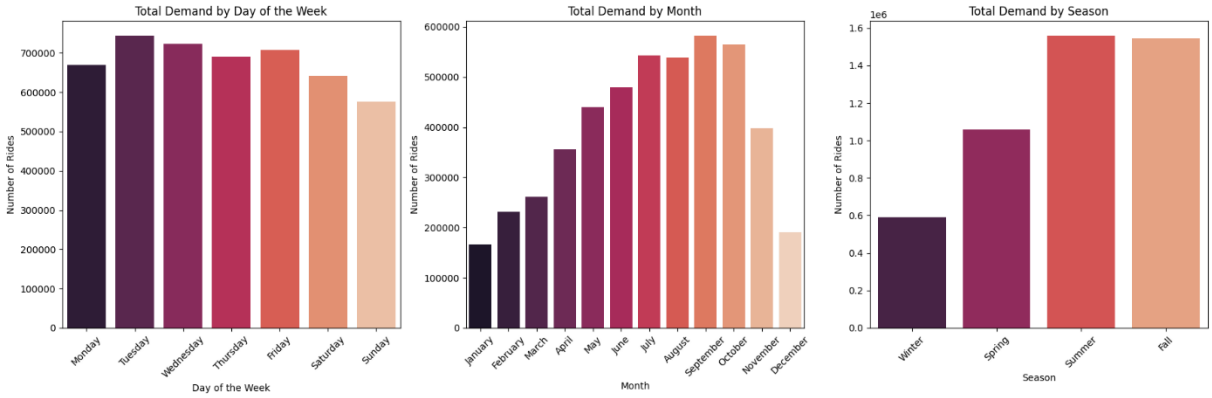

**شکل ۳. مراحل مدلسازی**

## ۵. تحلیل داده‌ها و نتایج

به‌منظور درک رفتار کاربران سامانه دوچرخه‌های اشتراکی، ابتدا الگوی کلی استفاده در ساعات شبانه‌روز، روزهای مختلف، ماه‌ها و فصول سال بررسی شد. نتایج نشان می‌دهد که بیشترین میزان تقاضا در ساعات اوج رخ می‌دهد که نشان‌دهنده تأثیر الگوهای رفت‌وآمد روزانه کاربران است. این امر می‌تواند به دلیل نیاز به جابجایی برای مقاصد کاری و آموزشی در ساعات ابتدایی روز و بازگشت به خانه در ساعات پایانی روز باشد. همچنین، تراکم ترافیکی در این ساعات ممکن است افراد را به استفاده از دوچرخه به‌عنوان یک گزینه سریع‌تر و مقرون‌به‌صرفه‌تر تشویق کند. درحالی‌که در ساعات کمترین استفاده، تعداد سفرها به حداقل می‌رسد.




**شکل ۴. تقاضای کل بر اساس ساعات شبانه‌روز**




**شکل ۵. تقاضای کل بر اساس روزهای هفته، ماه و فصل**

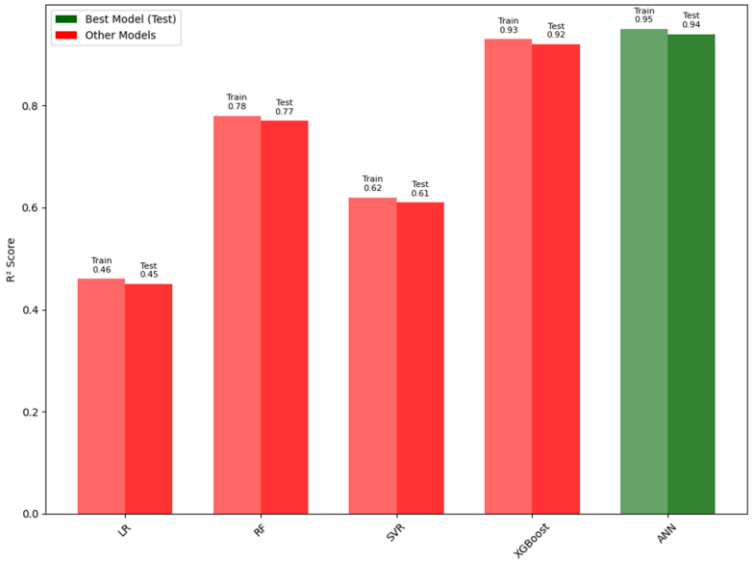
به‌منظور پیش‌بینی تقاضای دوچرخه‌های اشتراکی، پنج مدل یادگیری ماشین شامل رگرسیون خطی، جنگل تصادفی، رگرسیون بردار پشتیبان، ایکس‌جی‌بوست و شبکه عصبی مصنوعی مورد ارزیابی قرار گرفتند. مجموعه داده‌ها به دو بخش آموزش (۸۰٪) و آزمون (۲۰٪) تقسیم شد و عملکرد مدل‌ها بر اساس ضریب تعیین ($R^2$)، میانگین خطای مطلق (MAE)، میانگین مربعات خطا (MSE) و جذر میانگین مربعات خطا (RMSE) سنجیده شد. جدول زیر نتایج ارزیابی مدل‌ها را نشان می‌دهد:

**جدول ۲. نتایج ارزیابی مدل‌های یادگیری ماشین برای پیش‌بینی تقاضای دوچرخه‌های اشتراکی**

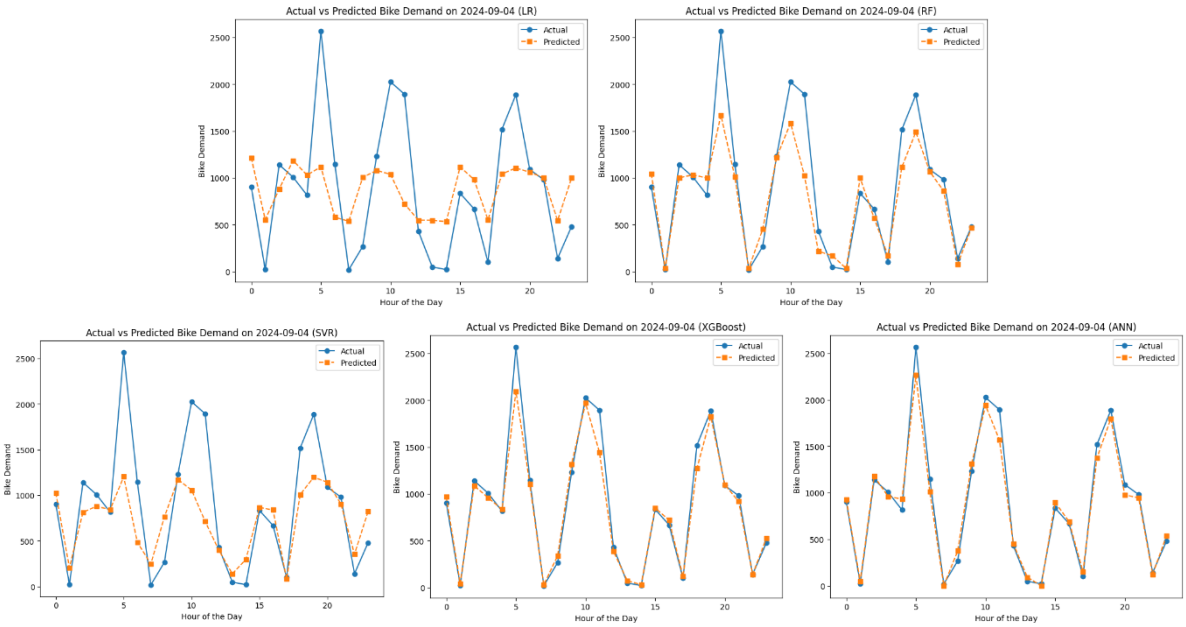
| مدل | نوع داده | ضریب تعیین | میانگین خطای مطلق | میانگین مربعات خطا | جذر میانگین مربعات خطا |
|---|---|---|---|---|---|
| رگرسیون خطی | آموزش | ۰/۴۶ | ۲۷۴/۸۱ | ۱۳۶۷۹۷/۱۴ | ۳۶۹/۸۶ |
| | تست | ۰/۴۵ | ۲۷۲/۴۷ | ۱۳۶۳۹۵/۳۶ | ۳۶۹/۳۲ |
| جنگل تصادفی | آموزش | ۰/۷۸ | ۱۵۷/۵۳ | ۵۴۳۵۹/۳۴ | ۲۳۳/۱۵ |
| | تست | ۰/۷۷ | ۱۵۵/۶۹ | ۵۶۹۵۰/۳۸ | ۲۳۸/۶۴ |
| رگرسیون بردار پشتیبان | آموزش | ۰/۶۲ | ۱۹۸/۱۹ | ۹۵۰۴۰/۴۰ | ۳۰۸/۲۹ |
| | تست | ۰/۶۱ | ۱۹۹/۹۷ | ۹۷۰۹۱/۴۵ | ۳۱۱/۶۰ |
| ایکس‌جی‌بوست | آموزش | ۰/۹۳ | ۷۶/۷۹ | ۱۶۹۳۵/۶۱ | ۱۳۰/۱۴ |
| | تست | ۰/۹۲ | ۸۵/۷۱ | ۱۹۵۶۲/۰۸ | ۱۳۹/۸۶ |
| شبکه عصبی مصنوعی | آموزش | ۰/۹۵ | ۷۱/۷۵ | ۱۱۶۰۱/۰۳ | ۱۰۷/۷۱ |
| | تست | ۰/۹۴ | ۷۹/۱۷ | ۱۴۶۱۴/۸۴ | ۱۲۰/۸۹ |

مقایسه مقادیر ضریب تعیین ($R^2$) و معیارهای خطا (MAE، MSE و RMSE) در داده‌های آموزش و آزمون نشان می‌دهد که هیچ‌یک از مدل‌های موردبررسی دچار بیش برازش نشده‌اند که این امر نشان‌دهنده تعمیم‌پذیری مناسب آن‌ها است. در این میان، مدل شبکه عصبی مصنوعی (ANN) با کمترین مقادیر خطا و بالاترین مقدار ضریب تعیین نسبت به سایر مدل‌ها، بهترین عملکرد را داشته است. پس‌ازآن، مدل ایکس‌جی‌بوست (XGBoost) عملکرد بهتری ارائه داده است. در مقابل، مدل جنگل تصادفی (RF) و رگرسیون بردار پشتیبان (SVR) دقت پایین‌تری نشان داده‌اند. رگرسیون خطی (LR) نیز ضعیف‌ترین عملکرد را داشته است که نشان می‌دهد ارتباط میان متغیرهای

ورودی و تقاضای دوچرخه‌های اشتراکی به‌صورت خطی نبوده و نیازمند مدل‌های پیچیده‌تر است.

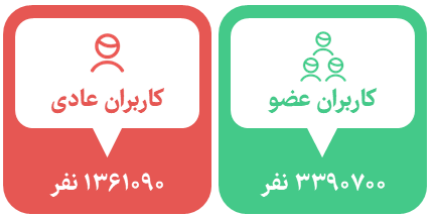



**شکل ۶. مقایسه تقاضای مدل‌ها بر اساس ضریب تعیین ($R^2$)**




**شکل ۷. مقایسه تقاضای واقعی و پیش‌بینی‌شده دوچرخه اشتراکی در تاریخ ۴ سپتامبر ۲۰۲۴ (رگرسیون خطی – جنگل تصادفی – رگرسیون بردار پشتیبان – ایکس‌جی‌بوست – شبکه عصبی مصنوعی)**

بررسی الگوهای استفاده نشان می‌دهد که کاربران عضو و کاربران عادی تفاوت‌های قابل‌توجهی در میزان استفاده از سامانه دارند. کاربران عضو سهم بیشتری (شکل ۸) در استفاده از دوچرخه‌های اشتراکی دارند و نتایج حاکی از آن است که ارائه مشوق‌هایی مانند کاهش هزینه اشتراک می‌تواند موجب افزایش تعداد کاربران وفادار شود.



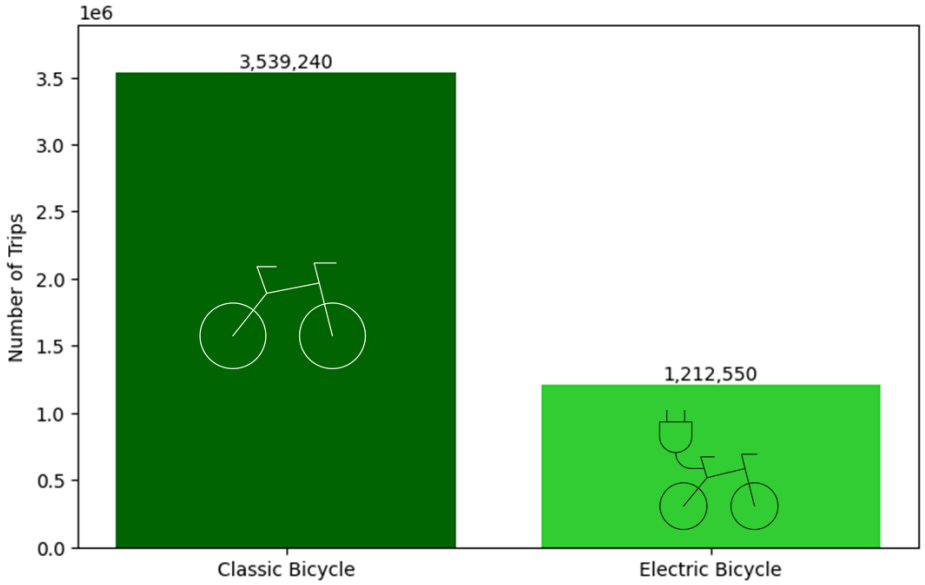

**شکل ۸. تعداد کاربران عضو و عادی**

تحلیل مکانی سفرها نشان می‌دهد که برخی ایستگاه‌ها تقاضای بالاتری نسبت به سایرین دارند. ایستگاه ام‌آی‌تی در خیابان مساچوست / خیابان امهرست بیشترین میزان استفاده را داشته، درحالی‌که ایستگاه کوچه سوان در مسیر دوچرخه‌سواری مینوت‌مَن کمترین میزان استفاده را ثبت کرده است.

**جدول ۳. مقایسه ایستگاه‌های پرتردد و کم تردد**

| پرترددترین ایستگاه‌ها | | کم ترددترین ایستگاه‌ها | |
|---|---|---|---|
| MIT at Massachusetts Ave / Amherst St | ۱۷۶۰۵۰ مراجعه | Swan Pl at Minuteman Bikeway | ۲ مراجعه |
| Central Square at Massachusetts Ave / Essex St | ۱۴۰۰۵۴ مراجعه | Shawmut Ave at Herald St | ۴ مراجعه |
| Harvard Square at Massachusetts Ave / Dunster St | ۱۲۴۹۱۴ مراجعه | Centre St at Allandale St | ۵ مراجعه |

همچنین، تعداد افراد استفاده‌کننده از دوچرخه‌های برقی در مقایسه با دوچرخه‌های معمولی قابل‌توجه بوده است که نشان می‌دهد کاربران ممکن است به دلیل راحتی بیشتر، تمایل بیشتری به استفاده از دوچرخه‌های برقی داشته باشند. بااین‌حال، تعداد دوچرخه‌های عادی بیشتر از دوچرخه‌های برقی است که نشان می‌دهد همچنان بخش عمده‌ای از کاربران از دوچرخه‌های معمولی استفاده می‌کنند.




**شکل ۹. تعداد سفرهای انجام‌شده با دوچرخه معمولی و دوچرخه برقی**

این مسئله می‌تواند به عواملی همچون هزینه کمتر اجاره دوچرخه‌های عادی، در دسترس بودن بیشتر آن‌ها و همچنین عدم نیاز به زیرساخت‌های شارژ برای دوچرخه‌های برقی مرتبط باشد. علاوه بر این، برخی کاربران ممکن است دوچرخه‌های عادی را به دلیل سبک‌تر بودن و امکان استفاده راحت‌تر در مسیرهای کوتاه ترجیح دهند.

## ۶. نتیجه‌گیری

در این پژوهش، عملکرد پنج مدل یادگیری ماشین شامل رگرسیون خطی، جنگل تصادفی، رگرسیون بردار پشتیبان، تقویت گرادیان و شبکه عصبی مصنوعی برای پیش‌بینی تقاضای دوچرخه‌های اشتراکی بررسی شد. هدف اصلی، شناسایی مدل بهینه برای افزایش دقت پیش‌بینی و بهبود مدیریت سامانه‌های اشتراکی بود. در ادامه، نتایج اصلی پژوهش ارائه می‌شود:

۱) نتایج نشان داد که شبکه عصبی مصنوعی بهترین عملکرد را داشته و پس‌ازآن، مدل تقویت گرادیان قرار گرفت. مدل‌های جنگل تصادفی، رگرسیون بردار پشتیبان و رگرسیون خطی دقت کمتری داشتند. بررسی معیارهای ارزیابی نشان داد که شبکه عصبی مصنوعی با ضریب تعیین ۰/۹۴ و مقدار جذر میانگین مربعات خطا حدوداً برابر ۱۲۰، دقیق‌ترین پیش‌بینی را ارائه کرده است.

۲) تحلیل عوامل مؤثر بر تقاضا نشان داد که شرایط آب‌وهوایی، زمان سفر و دسترسی به ایستگاه‌های پرتردد نقش کلیدی دارند. افزایش دما، کاهش بارش و افزایش ساعات پرتردد موجب افزایش تقاضا شده‌اند. همچنین، الگوهای استفاده نشان دادند که بیشترین میزان تقاضا در ساعات اوج (صبح و عصر) رخ می‌دهد که با الگوهای رفت‌وآمد کاربران همخوانی دارد. تقاضا در روزهای تعطیل کمتر بوده و استفاده از دوچرخه‌های برقی نسبت به دوچرخه‌های معمولی روند رو به رشدی داشته است اما دوچرخه‌های معمولی همچنان سهم بیشتری دارند که می‌تواند به دلایلی مانند هزینه کمتر و در دسترس بودن مرتبط باشد.

۳) بررسی مکانی ایستگاه‌ها نشان داد که برخی ایستگاه‌های پرتردد مانند مؤسسه فناوری ماساچوست بیشترین میزان استفاده را دارند، درحالی‌که برخی ایستگاه‌های کم تردد کمتر مورداستفاده قرار گرفته‌اند. این یافته‌ها تأکید می‌کند که توسعه ایستگاه‌های جدید باید با توجه به میزان تقاضا و تراکم سفرها انجام شود. همچنین، ارائه مشوق‌هایی مانند کاهش هزینه اشتراک می‌تواند موجب افزایش تعداد کاربران وفادار شود.

برای تحقیقات آینده، پیشنهاد می‌شود از یادگیری عمیق[۳۴] و مدل‌های ترکیبی برای افزایش دقت پیش‌بینی استفاده شود. همچنین، افزودن داده‌هایی مانند حجم ترافیک، شرایط اقتصادی و ویژگی‌های جمعیتی می‌تواند به بهبود پیش‌بینی‌ها کمک کند. توسعه الگوریتم‌های تخصیص هوشمند دوچرخه و بهینه‌سازی فرآیندهای جابه‌جایی نیز می‌تواند به تعادل عرضه و تقاضا در ایستگاه‌ها کمک کند.

درمجموع، این پژوهش بر لزوم بهبود زیرساخت‌های شهری، سیاست‌گذاری‌های هدفمند و بهره‌گیری از مدل‌های پیشرفته تحلیل داده تأکید دارد. این ملاحظات می‌توانند در توسعه سامانه‌های کارآمدتر دوچرخه‌های اشتراکی و ارتقای نقش آن‌ها در حمل‌ونقل پایدار شهری مؤثر باشند.

## ۷. پی‌نوشت‌ها

1. Machine Learning
2. Coefficient of Determination
3. Mean Absolute Error (MAE)
4. Mean Squared Error (MSE)
5. Root Mean Squared Error (RMSE)
6. Econometrics
7. Global Positioning System (GPS)
8. Time Series Models
9. Linear Regression (LR)
10. Autoregressive Integrated Moving Average (ARIMA)
11. Seasonal ARIMA (SARIMA)
12. Autoregression (AR)
13. Integration (I)
14. Moving Average (MA)
15. Multivariate Regression
16. Vector AutoRegression (VAR)
17. Vector Error Correction Model (VECM)
18. Random Forest (RF)
19. Support Vector Regression (SVR)
20. Artificial Neural Networks (ANNs)
21. Gradient Boosting
22. XGBoost

23. Correlation Coefficient
24. Ride-Hailing
25. Mean Absolute Percentage Error (MAPE)
26. Accuracy
27. Precision
28. Recall
29. Naive Bayes
30. One-Hot Encoding
31. Correlation Matrix
32. Margin Maximization
33. Kernel Functions
34. Deep Learning

## ۸ مراجع


– Chen, Y., Zhang, Y., Coffman, D., & Mi, Z. (2022). An environmental benefit analysis of bike sharing in New York City. Cities, 121, 103475.

– Nikitas, A. (2018). Understanding bike-sharing acceptability and expected usage patterns in the context of a small city novel to the concept: A story of ‘Greek Drama.’ Transportation Research Part F: Traffic Psychology and Behaviour, 56, 306–321.

– Fishman, E., Washington, S., & Haworth, N. (2014). Bike share’s impact on car use: Evidence from the United States, Great Britain, and Australia. Transportation Research Part D: Transport and Environment, 31, 13–20.

– Cheng, J., Hu, L., Lei, D., & Bi, H. (2024). How bike-sharing affects the accessibility equity of public transit systems—Evidence from nanjing. Land, 13(12), 2200.

– Chen, F., Zhang, W., Mfarrej, M. F. B., Saleem, M. H., Khan, K. A., Ma, J., Raposo, A., & Han, H. (2024). Breathing in danger: Understanding the multifaceted impact of air pollution on health impacts. Ecotoxicology and Environmental Safety, 280, 116532.

– Logan, G., Somers, C., Baker, G., Connell, H., Gray, S., Kelly, P., McIntosh, E., Welsh, P., Gray, C. M., & Gill, J. M. R. (2023). Benefits, risks, barriers, and facilitators to cycling: A narrative review. Frontiers in Sports and Active Living, 5, 1168357.

– Médard De Chardon, C., Caruso, G., & Thomas, I. (2017). Bicycle sharing system ‘success’ determinants. Transportation Research Part A: Policy and Practice, 100, 202–214.

– DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. Journal of Public Transportation, 12(4), 41–56.

– Gammelli, D., Wang, Y., Prak, D., Rodrigues, F., Minner, S., & Pereira, F. C. (2022). Predictive and prescriptive performance of bike-sharing demand forecasts for inventory management. Transportation Research Part C: Emerging Technologies, 138, 103571.

– Ren, S., Choi, T.-M., Lee, K.-M., & Lin, L. (2020). Intelligent service capacity allocation for cross-border-E-commerce related third-party-forwarding logistics operations: A deep learning approach. Transportation Research Part E: Logistics and Transportation Review, 134, 101834.

– González Vargas, C. A., & Elizondo Cortés, M. (2022). Automobile spare-parts forecasting: A comparative study of time series methods. International Journal of Automotive and Mechanical Engineering, 14(1), 3898–3912.

– Yuan, W.-J., Chen, J.-H., Cao, J.-J., & Jin, Z.-Y. (2018). Forecast of logistics demand based on grey deep neural network model. 2018 International Conference on Machine Learning and Cybernetics (ICMLC), 251–256.

novel ensemble model with conditional intervening opportunities for ride-hailing travel mobility estimation. Physica A: Statistical Mechanics and Its Applications, 628, 129167.

– Mohanta, B. K., Jena, D., Mohapatra, N., Ramasubbareddy, S., & Rawal, B. S. (2022). Machine learning based accident prediction in secure IoT enable transportation system. Journal of Intelligent & Fuzzy Systems, 42(2), 713–725.

– Plakandaras, V., Papadimitriou, T., & Gogas, P. (2019). Forecasting transportation demand for the U.S. market. Transportation Research Part A: Policy and Practice, 126, 195–214.

– Wild, G., Baxter, G., Srisaeng, P., & Richardson, S. (2021). Machine learning for air transport planning and management (No. arXiv: 2112.01301). arXiv.

– Stadler, T., Sarkar, A., & Dünnweber, J. (2021). Bus demand forecasting for rural areas using xgboost and random forest algorithm. In K. Saeed & J. Dvorský (Eds.), Computer Information Systems and Industrial Management (Vol. 12883, pp. 442–453). Springer International Publishing.

– He, Y., & Li, X. (2022). Feasibility of economic forecasting model based on intelligent algorithm of smart city. Mobile Information Systems, 2022, 1–9.

– Nar, M., & Arslankaya, S. (2022). Passenger demand forecasting for railway systems. Open Chemistry, 20(1), 105–119.

– Tashmetov, K., Rasulmukhamedov, M., & Tashmetov, T. (2024). Forecasting transport flows using big data and machine learning technology. E3S Web of Conferences, 531, 02012.

– Mamede, F. P., da Silva, R. F., de Brito Junior, I., Yoshizaki, H. T. Y., Hino, C. M., & Cugnasca, C. E. (2023). Deep learning and statistical models for forecasting transportation demand: A case study of multiple distribution centers. Logistics, 7(4), 86.

– Fu, W., Chien, C.-F., & Lin, Z.-H. (2018). A hybrid forecasting framework with neural network and time-series method for intermittent demand in semiconductor supply chain. In I. Moon, G. M. Lee, J. Park, D. Kiritsis, & G. von Cieminski (Eds.), Advances in Production Management Systems. Smart Manufacturing for Industry 4.0 (pp. 65–72). Springer International Publishing.

– Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. Computers & Industrial Engineering, 143, 106435.

– Kolambe, M. (2024). Forecasting the future: A comprehensive review of time series prediction techniques. Journal of Electrical Systems, 20(2s), 575–586.

– Boateng, A., Anum Adams, C., & Kofi Akowuah, E. (2023). Estimating passenger demand using machine learning models: A systematic review. E3S Web of Conferences, 418, 03002.

– Ngo, T.-T. T., Pham, H. T., Acosta, J. G., & Derrible, S. (2022). Predicting bike-sharing demand using random forest. Journal of Science and Transport Technology, 13–21.

– E, S. V., Park, J., & Cho, Y. (2020). Using data mining techniques for bike sharing demand prediction in metropolitan city. Computer Communications, 153, 353–366.

– Chen, Y., Geng, M., Zeng, J., Yang, D., Zhang, L., & Chen, X. (Michael). (2023). A