

## پیش‌بینی حجم ترافیک با استفاده از روش‌های یادگیری ماشین مطالعه موردی: خیابان کمال واقع در شهر اصفهان

### مقاله علمی - پژوهشی

مهسا لطیفی، دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی (GIS)، گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران و حمل‌ونقل، دانشگاه اصفهان، اصفهان، ایران  
جمشید مالکی، استادیار، گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران و حمل‌ونقل، دانشگاه اصفهان، اصفهان، ایران  
\*پست الکترونیکی نویسنده مسئول: j.maleki@cet.ui.ac.ir

دریافت: ۱۴۰۴/۰۴/۰۲ - پذیرش: ۱۴۰۴/۰۸/۰۲

صفحه ۱۸۲-۱۵۹

### چکیده

تقاضای روزافزون استفاده از وسایل حمل‌ونقل شخصی، مشکل تراکم ترافیک را به یکی از مهم‌ترین بحران‌ها در اکثر کلان‌شهرهای جهان تبدیل کرده است. توانایی پیش‌بینی میزان حجم ترافیک جاده‌ها می‌تواند در بسیاری از برنامه‌های مدیریتی و کنترل ترافیک کمک کند. اما تعیین حجم ترافیک چالش برانگیز است، زیرا شمارش وسایل نقلیه معمولاً فقط در تعداد معدودی از مکان‌هایی که دارای سنسورهای ترافیک ثابت هستند، امکان‌پذیر است. در پژوهش حاضر برای رفع این چالش از داده‌های مسیریابی سرویس نشان استفاده شده است. با مسیریابی در ساعات مختلف شبانه‌روز ثبت اطلاعات مربوط به مدت‌زمان سفر، امکان برآورد حجم ترافیک یال‌های شهری وجود دارد. برای بررسی امکان‌پذیری روش پیشنهادی، داده‌های مربوط به مدت‌زمان سفر خیابان کمال واقع در شهر اصفهان برای ساعت مختلف شبانه‌روز و به مدت ۲۴ روز اخذ گردید. همچنین به منظور پیش‌بینی حجم ترافیک ساعتی از روش‌های یادگیری ماشین شامل جنگل تصادفی، درخت تقویت کننده گرادیان شدید، شبکه عصبی حافظه کوتاه مدت طولانی و شبکه‌های عصبی بازگشتی بلندمدت‌دوطرفه بکارگرفته شده است. از تبدیل فوریه سریع نیز جهت شناسایی فرکانس‌های اصلی تغییرات حجم ترافیک و مدل‌سازی و مقایسه با روش‌های دیگر استفاده شده است. در روش‌های استفاده شده در این پژوهش، از ویژگی‌هایی مانند: وابستگی مکانی (مجموع مدت‌زمان سفر یال‌های ورودی خیابان)، تاریخچه مدت‌زمان سفر، ویژگی‌های زمانی مانند: ساعت و روزهای هفته، تعداد یال‌های ورودی و تعداد مراکز تأثیرگذار در ترافیک شهری از جمله: تعداد مدارس، تعداد بانک‌ها، هتل‌ها، شهرداری‌ها، داروخانه‌ها، اورژانس و بیمارستان‌ها، جمعیت محصل و کل جمعیت در محدوده خیابان‌ها استفاده شده است. نتایج عددی نشان می‌دهد، روش‌های یادگیری ماشین و یادگیری عمیق عملکرد بهتری نسبت به روش تبدیل فوریه سریع دارند. همچنین در بین روش‌های یادگیری ماشین و یادگیری عمیق نیز روش جنگل تصادفی با میزان بالای  $R^2$  برابر با ۰/۹۳ عملکرد بهتری را نسبت به سایر روش‌ها دارد.

واژه‌های کلیدی: پیش‌بینی حجم ترافیک ساعتی، یادگیری ماشین، تبدیل فوریه سریع، مدل‌سازی مکانی - زمانی

### ۱- مقدمه

تعیین حجم ترافیک چالش برانگیز است، زیرا شمارش وسایل نقلیه معمولاً تنها در تعداد معدودی از مکان‌هایی که سنسورهای ترافیک ثابت دارند در دسترس است (Sekula et al., 2018).

افزایش جمعیت و شهرنشینی، از دلایل اصلی تراکم ترافیک در اکثر شهرها است. پیش‌بینی جریان ترافیک نقش کلیدی در سیستم‌های حمل و نقل هوشمند دارد (Tian et al., 2018).

است که وابستگی‌های مکانی را به‌دقت در نظر می‌گیرد و این مشکل را حل می‌کند. همچنین، با بهره‌گیری از داده‌های سرویس نشان، محدودیت حسگرها و کمبود داده‌های ترافیکی رفع شده و امکان پیش‌بینی دقیق حجم ترافیک ساعتی برای ۲۴ ساعت آینده فراهم می‌شود.

## ۲- پیشینه تحقیق

در حوزه پیش‌بینی حجم ترافیک تاکنون مطالعات متعددی صورت گرفته است. در مطالعات انجام شده از روش‌های مختلفی برای انتخاب متغیرهای تأثیرگذار و همچنین تنظیم پارامترهای آن در جهت بهبود فرایند پیش‌بینی استفاده شده است. برای پیش‌بینی حجم ترافیک نیز از روش‌های مبتنی بر روابط ریاضی و یادگیری ماشین استفاده شده است. در ادامه به روش‌های مختلف پیش‌بینی حجم ترافیک اشاره شده است.

### ۲-۱- رویکرد ریاضی

استفاده از روابط ریاضی و آماری در پیش‌بینی ترافیک آینده معمولاً به یک دوره از داده‌های تاریخی نیاز است تا روندها استخراج شده و در یک معادله ریاضی قرار گیرد. در یک مطالعه انجام شده به منظور مدل سازی حجم ترافیک روزانه در یک منطقه برون شهری، روش‌های مدل‌سازی ریاضی مبتنی بر تجزیه سری‌های زمانی انجام شده است. این مقاله با اعمال ویژگی بر اساس روز هفته، اثر زمانی بر حجم ترافیک را به دست می‌آورد. نتایج این مطالعه، به میانگین خطای نسبی ۶ درصد دست یافته است. با این حال، این مدل‌ها مستعد تغییرات ناگهانی مانند تغییرات در شرایط جاده مثل تصادفات رانندگی یا شرایط محیطی مانند COVID-19 منجر به فرهنگ کار از خانه و از راه دور شده‌اند، هستند (Pechatnova & Kuznetsov, 2021). در مطالعه‌ی دیگر از یک مدل اثرات تصادفی مکانی-زمانی (STRE) با کاهش ابعاد ریاضی استفاده شده است که توسط یک تابع پایه که قادر به در نظر گرفتن الگوهای ترافیک است، ارائه می‌شود (Wu et al., 2016). با این حال، عیب رویکرد ریاضی این است که فقط در جاده‌های مشخص شده قابل اعمال است زیرا جاده‌های مختلف معادلات ریاضی متفاوتی دارند و مدل‌های ریاضی فقط می‌توانند ترافیک آینده را برای هر جاده، به طور مجزا و خاص پیش‌بینی کنند، و برای محیط‌هایی که کمتر داده از دست رفته است، مناسب هستند (Tay et al.,

توانایی پیش‌بینی به موقع و دقیق تحولات ترافیک در برنامه‌های مدیریت و کنترل ترافیک بسیار مهم است (Habtemichael & Cetin, 2016). حجم دقیق ترافیک یکی از حیاتی‌ترین اطلاعات برای سازمان‌های دولتی برای تصمیم‌گیری آگاهانه در مورد زیرساخت‌های حمل و نقل، استراتژی‌های مدیریت ترافیک و طرح‌های برنامه ریزی شهری است (Zhan et al., 2016). بنابراین مقامات می‌توانند با گنجاندن حجم ترافیک در برنامه ریزی شهری، از ازدحام بیش از حد در شبکه ترافیک جلوگیری کنند و زیرساخت‌های استراتژیک مانند بیمارستان‌ها و آتش‌نشانی‌ها را به مناطقی با ترافیک کمتر اختصاص دهند تا ازدحام کاسته شود (Lv et al., 2014) پیش‌بینی میزان حجم ترافیک به کاربران جاده کمک می‌کند تا تصمیم‌های بهتری برای سفر بگیرند (Tian et al., 2018). عدم پیش‌بینی به موقع ترافیک، باعث افزایش نرخ تصادفات، مصرف سوخت اضافی، آلودگی بیش از حد هوا، و بدتر شدن سلامت جامعه می‌شود (Shaygan et al., 2022). حسگرهای ترافیک وسیله مهمی برای به دست آوردن اطلاعات حجم ترافیکی در زمان هستند. با این حال، محدودیت‌های بودجه معمولاً مانع نصب کامل دستگاه‌های تشخیص در کل شبکه‌های جاده‌ای می‌شوند (Liu et al., 2015). به همین دلیل نصب حسگرها برای پوشش کامل همه جاده‌های شهری به منظور ثبت حجم ترافیک غیرممکن خواهد بود (Yi et al., 2021). به خصوص که الگوهای ترافیک در طول هفته و آخر هفته به دلیل عواملی مانند برنامه کاری، روزهای مدرسه، شرایط آب و هوایی، تعطیلات، متفاوت است (Gomes et al., 2023). در تحقیق حاضر، داده‌های مردمی ترافیک، از سکوی نشان اخذ شده و این امکان را می‌دهد که برای همه خیابان‌ها حجم ترافیک واقعی را در بازه‌های زمانی مشخص در اختیار داشت. از این رو با استفاده از تکنیک‌های مختلف می‌توان با در نظر گرفتن تاریخچه حجم ترافیک، پیش‌بینی دقیقی از ترافیک برای ۲۴ ساعت آتی را بدست آورد در بسیاری از مطالعات گذشته، از روش‌های یادگیری ماشین استفاده شده که هر مکان را به‌طور مستقل و بدون در نظر گرفتن وابستگی مکانی و با تکیه بر فاصله اقلیدسی آموزش می‌دهند. این روش می‌تواند منجر به اشتباه شود، زیرا جاده‌های نزدیک و متصل که از نظر مکانی وابسته هستند، ممکن است الگوهای ترافیکی متفاوتی داشته باشند. در مدل پیشنهادی این پژوهش، از شبکه گراف استفاده شده که مزیت آن نسبت به فاصله اقلیدسی این

(Wang et al., 2006) و (Grant-Muller, 2001) در مطالعه دیگر، برای پیش‌بینی وضعیت ترافیک جاده‌ای یک مدل بر اساس مدل ARIMA و فیلتر کالمن پیشنهاد شده است. میانگین مجموع خطای نسبی مربعات نشان می‌دهد که روش مبتنی بر ترکیب ARIMA و فیلتر کالمن نسبت به دو مدل ARIMA و فیلتر کالمن پایدارتر است (Xu et al., 2017). در تحقیق دیگر سیستم استنتاج فازی مبتنی بر شبکه عصبی ارائه شد. (ANFIS) ترکیبی از سیستم‌های فازی و شبکه‌های عصبی است که از طریق بهینه‌سازی پارامترها (مانند وزن‌ها و پارامترهای توابع عضویت فازی) عمل می‌کند که در این مطالعه میزان حجم ترافیک با دقت ۹۰ درصد پیش‌بینی نمود (Rahmaty et al., 2019).

#### ۲-۲-۲- مدل‌های غیرپارامتریک یادگیری ماشین

روش‌های پارامتریک معمولاً روابط پیچیده بین حجم ترافیک و متغیرهای مستقل را ساده‌سازی می‌کنند و این موضوع دقت پیش‌بینی را نسبت به روش‌های غیرپارامتریک کاهش می‌دهد. مدل‌های غیرپارامتریک برخلاف مدل‌های پارامتریک، به تعداد ثابتی از پارامترها محدود نبوده و هیچ فرضی درباره شکل یا توزیع داده‌ها ندارند. این مدل‌ها به‌طور مستقیم از داده‌ها یاد می‌گیرند و برای الگوهای پیچیده و غیرخطی مناسب‌تر هستند. مزیت اصلی آن‌ها دقت بیشتر در پیش‌بینی داده‌هایی با توزیع نامشخص است، اگرچه به داده‌های بیشتر و منابع محاسباتی بالاتری نیاز دارند. از جمله این مدل‌ها می‌توان به شبکه‌های عصبی اشاره کرد که معماری‌های مختلف آن برای بهبود پیش‌بینی ترافیک توسعه یافته است. چند نمونه این معماری‌ها عبارتند از: شبکه عصبی فید فوروارد (FFNN) (Park & Rilett, 1999)، شبکه عصبی مدولار (MNN) (Park & Rilett, 1998)، شبکه عصبی مبتنی بر طیف (SNN) (Park et al., 1999)، شبکه عصبی فازی عصبی (NFNN) (Yin et al., 2002) و شبکه عصبی بازگشتی (RNN) (Lingras et al., 2002) و (Van Lint et al., 2002). در مطالعه‌ای، یک شبکه عصبی مصنوعی برای پیش‌بینی حجم تردد در جاده‌های برون‌شهری استان‌های ایران، با در نظر گرفتن تأثیر تعطیلات و مقیاس زمانی روزانه، توسعه داده شد. میانگین درصد خطای مطلق مدل‌ها ۲۸/۹ درصد و برای استان مازندران ۸/۸۹ درصد بود. این مدل به سیاست‌گذاران کمک می‌کند تا حجم

(2023). همچنین در این روش‌ها معمولاً تنها به تاریخچه داده‌های ترافیکی توجه شده و پارامترهای تأثیرگذار در حجم ترافیک باتوجه به پیچیدگی روابط بین این پارامترها و حجم ترافیک، در نظر گرفته نمی‌شوند.

#### ۲-۲- رویکرد یادگیری ماشین

یکی دیگر از روش‌های پیش‌بینی ترافیک، استفاده از روش‌های یادگیری ماشین است. به دلیل اینکه شرایط ترافیکی به‌سرعت در حال تغییر است و ممکن است تحت تأثیر بسیاری از عوامل، مانند: کاربران جاده، شرایط آب‌وهوایی و رویدادهای خاص قرار گیرند، از این رو، یادگیری ماشین می‌تواند قوی‌تر از رویکردهای ریاضی عمل کرده و بهتر با ماهیت تصادفی بودن مقادیر پارامترهای مختلف برخورد کند (Tay et al., 2023). این رویکرد به دو دسته عمده مدل‌های پارامتریک و ناپارامتریک تقسیم بندی می‌شوند که در ادامه هر یک از این مدل‌ها جداگانه بررسی می‌شوند.

#### ۲-۲-۱- مدل‌های پارامتریک یادگیری ماشین

مدل‌های پارامتریک در یادگیری ماشین به مدل‌هایی اطلاق می‌شود که ساختار آن توسط یک تعداد ثابت از پارامترها تعریف می‌شود و این پارامترها در طول فرآیند آموزش، تنظیم می‌شوند. این مدل‌ها بر این فرض استوار هستند که داده‌ها از یک توزیع خاص پیروی می‌کنند و هدف آموزش، یافتن پارامترهای بهینه برای توزیع مفروض است. مزیت اصلی مدل‌های پارامتریک این است که سریع‌تر آموزش می‌بینند و پردازش می‌شوند، اما اگر فرض‌های اولیه‌شان درباره داده‌ها اشتباه باشد، ممکن است نتایج دقیقی ارائه ندهند. از مدل‌های پارامتریک می‌توان به رگرسیون خطی و لجستیک، مدل میانگین متحرک یکپارچه اتورگرسیون، روش رگرسیون بردار پشتیبان و مدل‌های مبتنی بر فیلتر کالمن (Okutani & Stephanedes, 1984) اشاره نمود. برای نمونه در اوایل دهه ۱۹۷۰، مدل میانگین متحرک یکپارچه اتورگرسیون (ARIMA) برای پیش‌بینی جریان ترافیک کوتاه‌مدت آزادراه مورد استفاده قرار گرفت (Ahmed & Cook, 1979). نسخه‌های متعددی از مدل ARIMA برای بهبود دقت پیش‌بینی پیشنهاد شده است، مانند Kohonen (ARIMA (KARIMA) رویکردهای مبتنی بر فیلتر کالمن نیز با موفقیت در پیش‌بینی ترافیک استفاده شده است (Chen &

روش‌های مبتنی بر درخت تصمیم اشاره نمود (Asencio-Cortés et al., 2016). برای نمونه، تامیر و همکاران (Tamir et al., 2020) یک مطالعه مقایسه‌ای برای پیش‌بینی تراکم ترافیک انجام دادند که شامل درخت تصمیم، رگرسیون لجستیک و شبکه‌های عصبی بودند. ایشان ادعا کردند که درخت تصمیم در پیش‌بینی ترافیک عملکرد بهتری دارد. در مطالعه‌ای دیگر یک روش دو مرحله‌ای یادگیری ماشین پیشنهاد داده شده است که در مرحله اول، از مدل خوشه‌بندی بدون نظارت برای استخراج الگوها از داده‌های حسگر استفاده می‌شود و در مرحله دوم، مدل‌های یادگیری ماشین نظارت‌شده برای پیش‌بینی ترافیک به کار گرفته می‌شوند. با وجود اینکه این روش به داده‌های حسگر برای آموزش نیاز دارد، می‌تواند جریان ترافیک را در مناطقی که حسگر ندارند نیز پیش‌بینی کند. در تجزیه و تحلیل معیارهای عملکرد، XGBoost و CatBoost از LightGBM و Random Forest در همه معیارهای ارزیابی بهتر عمل کردند (Berlotti et al., 2024). در بسیاری از مطالعات غیرپارامتریک، هر مکان به‌طور مستقل و بدون در نظر گرفتن وابستگی‌های مکانی آموزش داده می‌شود و فاصله اقلیدسی برای ارزیابی حجم ترافیک به کار می‌رود. این موضوع می‌تواند منجر به اشتباهاتی شود، زیرا جاده‌های همسایه که از لحاظ مکانی نزدیک هستند اما الگوهای ترافیکی متفاوتی دارند، به اشتباه مرتبط در نظر گرفته می‌شوند. این اشتباه ممکن است پیش‌بینی ترافیک را نادرست کرده و به کاهش یا افزایش حجم ترافیک تخمینی منجر شود. بنابراین، استفاده از گراف برای درک بهتر وابستگی‌های مکانی اهمیت پیدا می‌کند. سپس می‌توان با استفاده از الگوریتم‌های یادگیری ماشین میزان حجم ترافیک را با توجه به ویژگی‌های تأثیرگذار و در نظر گرفتن وابستگی مکانی به دست آورد. در مطالعه‌ای انجام شده، یک روش پیش‌بینی حجم ترافیک ساعتی، در مقیاس شبکه ارائه می‌شود که وابستگی مکانی در نظر گرفته می‌شود. به دلیل مقیاس‌پذیری قدرتمند مدل، از یک مدل مجموعه درختی پیشرفته، درخت تقویت‌کننده گرادیان شدید (XGBoost) استفاده شد. در این مطالعه، یک شبکه گراف ترافیک وزنی بدون جهت، به منظور بررسی همبستگی مکانی بین بخش‌های جاده ایجاد می‌شود. نتیجه عددی، مزیت مدل پیشنهادی را نسبت به سایر مدل‌های دیگر در هنگام استفاده از ویژگی وابستگی مکانی در یک شبکه را نشان می‌دهد. به طور

تردد را در محورهای مواصلاتی پیش‌بینی کرده و برای تراکم‌های ترافیکی در ایام خاص برنامه‌ریزی کنند (Sahebi et al., 2024). یکی دیگر از مدل‌های غیرپارامتریک مدل نزدیک‌ترین همسایه<sup>۴</sup> (K-NN) است که برای پیش‌بینی ترافیک کوتاه مدت نیز استفاده شده است. این الگوریتم با یافتن  $k$  نمونه نزدیک به داده جدید بر اساس فاصله، برای طبقه‌بندی یا پیش‌بینی مقدار استفاده می‌شود. تصمیم‌گیری بر اساس اکثریت همسایه‌های نزدیک انجام می‌گیرد. الگوریتم K-NN در مواجهه با تغییرات ناگهانی و غیر خطی بودن جریان ترافیک شهری عملکرد خوبی دارد ولی زمان اجرای این الگوریتم بالا است که منجر به کاهش رندمان پیش‌بینی خواهد شد (Chen et al., 2019). در مطالعه‌ای صورت گرفته، به منظور پیش‌بینی کوتاه مدت پارامترهای ترافیک شهری، به دلیل عدم دسترسی به داده‌های ترافیکی منظم در ایران، داده‌های این تحقیق از داده‌های سال ۲۰۱۲ تا ۲۰۱۴ در لندن با رفتار ترافیکی مشابه در طول هفته انتخاب گردید. میزان RMSE نتایج نشان می‌دهد که سه روش چندجمله‌ای کلاسیک، چندجمله‌ای بر مبنای ژنتیک و شبکه‌های عصبی برای پارامتر ترافیکی جریان به ترتیب برابر با ۱۳/۹۱، ۰/۷۸ و ۰/۲۲ و برای پارامتر سرعت به ترتیب برابر با ۰/۲۰، ۰/۷۸ و ۰/۱۹ می‌باشند. به عبارت دیگر، دقت پیش‌بینی پارامتر جریان ترافیک در روش‌های چندجمله‌ای مبتنی بر ژنتیک و شبکه‌های عصبی به ترتیب تقریباً ۱۸ و ۶۳ برابر بهتر از روش چند جمله‌ای کلاسیک و دقت پیش‌بینی پارامتر سرعت تقریباً ۷ و ۲۷ برابر بهتر بود (Emami & Rafati, 2023). همچنین شبکه‌های مبتنی بر LSTM، داده‌های بلندمدت و کوتاه‌مدت را در حافظه نگه می‌دارند و از آنها برای تصمیم‌گیری پیش‌بینی در لحظه فعلی استفاده می‌کنند. تحقیقات نشان می‌دهند که LSTM از نظر دقت و پایداری نسبت به سایر الگوریتم‌های ذکر شده در مدل‌سازی داده‌های تاریخیچه ای بهتر عمل می‌کند (Ma et al., 2015).

روش یادگیری جمعی یکی از روش‌های محبوب در یادگیری ماشین است که زیرمجموعه مدل‌های ناپارامتریک بوده و با ترکیب چندین مدل مختلف سعی در بهبود عملکرد کلی مدل‌ها دارد. به جای تکیه بر یک مدل، این روش از چند مدل استفاده می‌کند، که این ترکیب می‌تواند دقت و مقاومت مدل‌ها را در برابر داده‌های نویزی یا پیچیده افزایش دهد. از روش‌های یادگیری جمعی به‌منظور پیش‌بینی ترافیک، می‌توان به

می توان به صورت دقیق میزان حجم ترافیک را به صورت ساعتی، روزانه یا هفتگی پیش بینی کرد و در آخر می توان با لینک کردن نتایج به دست آمده به نقشه معابر اصلی اصفهان، میزان حجم ترافیک در مناطق مختلف شهر اصفهان را مشاهده کرد.

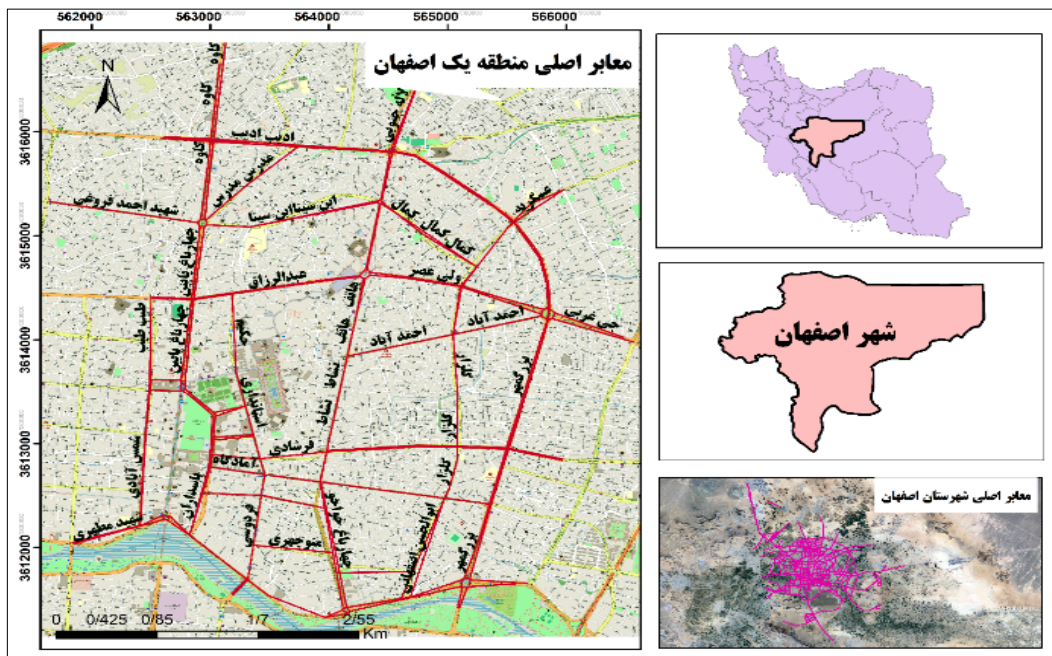
### ۳- منطقه مورد مطالعه و داده ها

#### ۳-۱- منطقه مورد مطالعه

اصفهان سومین شهر پرجمعیت ایران پس از تهران و مشهد، یکصد و شصت و پنجمین شهر پرجمعیت جهان و نهمین شهر پرجمعیت غرب آسیا به شمار می رود. اصفهان چهاردهمین کلان شهر پرجمعیت خاورمیانه است. کلان شهر اصفهان با ۱۵ منطقه شهری، پس از تهران دومین کلان شهر بزرگ ایران از نظر مساحت شهری است. همین موضوع باعث می شود که مدلسازی ترافیک و پیش بینی حجم ترافیک از اهمیت بالایی برخوردار باشد. به منظور بررسی و ارزیابی مدل های پیش بینی حجم ترافیک یکی از معابر اصلی اصفهان به نام خیابان کمال انتخاب گردیده است. (شکل ۱). داده هایی که برای این پژوهش استفاده شده است شامل موارد زیر است.

خاص، نتایج عددی مدل پیشنهادی را با معیارهای  $MAE$ ،  $R^2$  و  $MAPE$  به ترتیب  $0.09$ ،  $0.503$  و  $0.079$  نشان می دهند.

میانگین  $MAE$  و  $MAPE$   $20.3$  درصد کاهش و در مقایسه با الگوریتم  $XGBoost$  به تنهایی،  $44/8$  درصد نیز کاهش می دهد (Yi et al., 2021). این پژوهش باهدف محاسبه حجم ترافیک ساعتی یکی از معابر اصلی شهر اصفهان و با در نظر گرفتن وابستگی مکانی انجام می شود. نوآوری این تحقیق، استفاده از مدت زمان سفر به عنوان مهم ترین داده در این پژوهش خواهد بود که می تواند بدون هیچ گونه محدودیت زمانی و مکانی از سرویس نشان اخذ شود. در صورتی که در بسیاری از مطالعات انجام شده با کمبود و نقص داده برای پیش بینی مواجه هستند. همچنین یکی دیگر از نوآوری های این تحقیق استفاده از الگوریتم های مختلف غیر پارامتریک یادگیری ماشین است که علاوه بر ترافیک خیابان های مجاور، تأثیرات زمانی را نیز مدنظر قرار می دهد. همچنین پارامترهای مؤثری مانند شرایط آب و هوایی و عوامل اجتماعی - اقتصادی بررسی و انتخاب می شوند. برخلاف بیشتر مطالعات قبلی، این پژوهش از یک شبکه گراف جهت دار استفاده می کند که با مقایسه نتایج آن ها، می توان میزان دقت را بررسی و بهترین مدل را انتخاب نمود. در نهایت با استفاده از داده های پردازش شده، طراحی شبکه و آموزش داده ها،

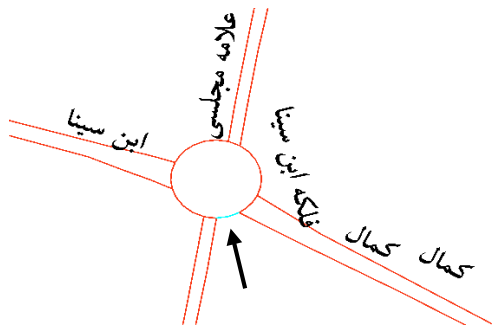


شکل ۱. موقعیت خیابان کمال در معابر اصلی منطقه یک شهر اصفهان

### ۲-۳- ارتباط بین داده های مدت زمان سفر با حجم

#### ترافیک

حجم ترافیک با مدت زمان سفر رابطه مستقیم دارد. اگر مدت زمان سفر زیاد باشد، حجم ترافیک زیاد خواهد بود و بالعکس. برای هر خیابان، کمترین زمان سفر، برابر با حجم ترافیک سفر و بیشترین زمان سفر، برابر با حجم ترافیک یک در نظر گرفته می شود. میزان حجم ترافیک برای بقیه زمان ها با انجام یک تناسب به دست می آید. با انجام این کار می توان داده ها را استاندارد کرد؛ بنابراین با داشتن مدت زمان سفر برای هر خیابان می توان حجم ترافیک را برآورد کرد. همان طور که در شکل ۲ مشاهده می کنید خیابان کمال دارای طول ۱۰۰۴/۵۱ متر است که به صورت یک طرفه و در جهت غرب به شرق است. یال ورودی به این خیابان، یکی از یال های فلکه این سینا است.



شکل ۲. یال ورودی به خیابان کمال

### ۳-۳- مفهوم ایستایی (مانایی) و نامانایی داده ها

مفهوم ایستایی یا مانایی سری های زمانی یکی از موضوعات مهم در مبحث سری های زمانی است که می تواند تأثیر شدیدی روی رفتار و ویژگی های آن داشته باشد. سری زمانی که توزیع مقادیر آن در طول زمان ثابت باشد، سری زمانی پایا نامیده می شود. در سری های پایا میانگین و واریانس متغیرها در طول زمان و کوواریانس متغیرها بین سال های مختلف ثابت است و اثرات شوک را در طول زمان از بین می برد. در مقابل، داده های نامانا به گونه ای هستند که اثر شوک های وارده، ماندگار و همیشگی است. در سری زمانی نامانا جهش شدید در مقاطعی مشاهده می شود که با تأیید آزمون های ریشه واحد می توان گفت که سری زمانی ناپایا است؛ بنابراین، اگر سری زمانی مانا باشد، می توان برای شناسایی داده های پرت از تست ۳ سیگما استفاده کرد. تست ۳ سیگما یک روش آماری است که برای

داده های مربوط به مدت زمان سفر: برای آموزش مدل های یادگیری ماشین از داده های مربوط به ترافیک شهر اصفهان که از سامانه مسیریابی نشان، دانلود شده استفاده شده است. با استفاده از قابلیت مسیریابی نرم افزار نشان می توان برای معابر اصلی در هر زمان، مدت زمان سفر را با دقت کافی به دست آورد. داده های از دست رفته با روش میانگین گیری تکمیل شده و به لینک های شبکه اختصاص داده شده است. برای هر خیابان، داده شامل شناسه منحصر به فرد، نام، جهت و طول خیابان، تاریخ و زمان ثبت شده و در نهایت مدت زمان بر حسب ثانیه است. منظور از مدت زمان، مدت زمانی است که یک ماشین طول می کشد تا خیابان مربوطه را طی کند. داده های این مطالعه از یکشنبه ۱۶ مهر ماه سال ۱۴۰۲ تا سه شنبه ۹ آبان ۱۴۰۲ به صورت هر یک ساعت دانلود شده است. ۸۲ درصد کل داده ها به عنوان داده آموزش ۱۰ درصد به عنوان داده اعتبارسنجی و ۸ درصد آن به عنوان داده آزمایش در نظر گرفته شده است.

داده های اخذ شده از Open Street Map و شهرداری: برای استفاده از داده های تأثیرگذار در ترافیک، موقعیت امکان مختلف مانند خیابان، مدارس، بانک ها، مراکز خرید و غیره از داده های Open Street Map استخراج شده و به لینک های شبکه اختصاص شد. برای هر خیابان اطلاعاتی مانند نام، یک طرفه بودن یا نبودن، نوع و سطح آن و... وجود دارد. اطلاعات مربوط به تعداد مدارس، تعداد بانک ها، هتل ها، شهرداری ها، داروخانه ها، اورژانس و بیمارستان ها، جمعیت محصل، کل جمعیت و سایر از شهرداری اصفهان اخذ گردید. اطلاعات مربوط به مدارس، بانک ها و... که از داده های مربوط به شهرداری گرفته شده است از دقت کافی برخوردار هستند.

داده های آب و هوا: داده های مربوط به بارش از سایت ناسا گرفته شده است که با رزولوشن 0.01 درجه به عبارتی تقریباً ۱ کیلومتر دانلود شده است. برای تاریخ مذکور هیچ بارشی ثبت نشده است.

ویژگی های زمانی: برای اعمال ویژگی های زمانی در مدل های مورد استفاده، ساعت در روز (ساعت ۱ تا ۲۴) و روز از هفته (شنبه تا جمعه)، به عنوان ویژگی های زمانی اضافه شد. همچنین، هیچ گونه تعطیلات رسمی در این بازه زمانی وجود ندارد.

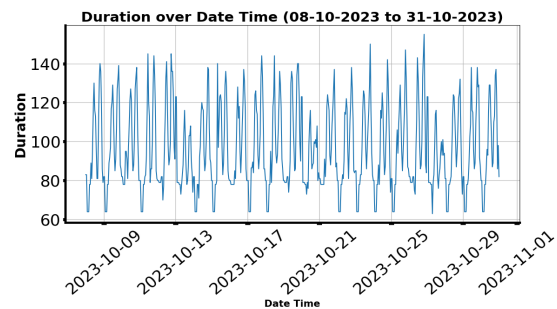
تاریخچه مدت زمان سفر: میانگین مدت زمان سفر در طی هفته های گذشته به عنوان ویژگی پیشینه مدت زمان سفر در نظر گرفته شده است.

در غیر این صورت فرضیه جایگزین ( $H_A$ ) نشان می‌دهد که سری روند ثابت دارد در نتیجه، مانا است. به عبارت دیگر، اگر فرض صفر رد شود، سری زمانی مانا است و ریشه واحد ندارد. آزمون  $KPSS^9$ : برخلاف آزمون دیکی فولر که فرضیه صفر نشان‌دهنده ریشه واحد است، فرضیه صفر  $KPSS$  نشان‌دهنده پایایی سری زمانی است. در آزمون  $KPSS$  فرضیه صفر ( $H_0$ ) نشان می‌دهد که سری روند ثابت دارند یا سری‌ها ریشه واحد ندارند و فرضیه جایگزین ( $H_A$ ) نشان‌دهنده سری غیرثابت و نامانا است.

#### ۴- روش شناسی

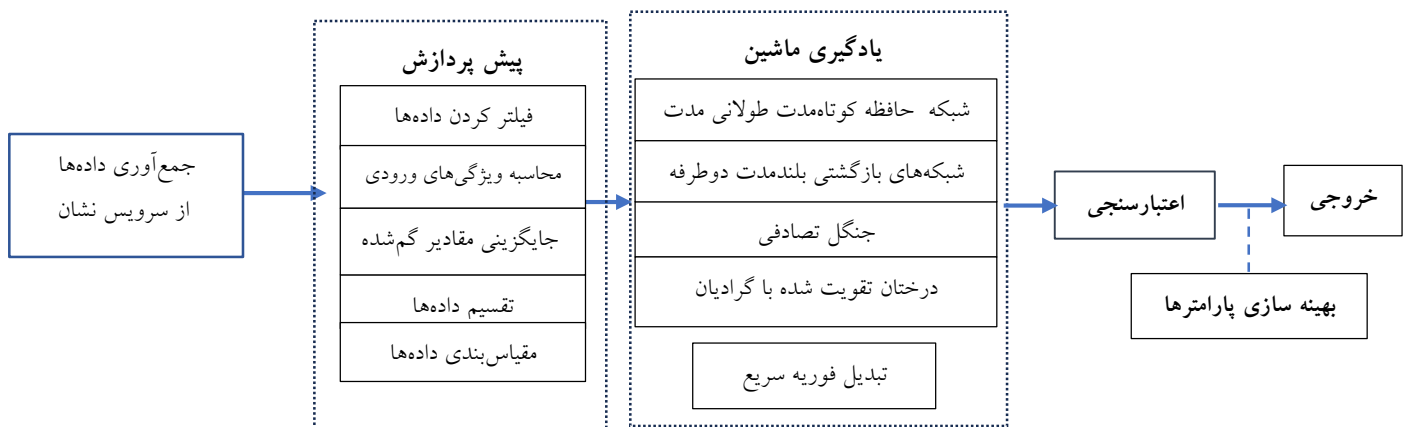
در این بخش، روش شناسی مرتبط با الگوریتم‌هایی مانند: جنگل‌های تصادفی، درخت تقویت‌کننده گرادیان شدید، شبکه عصبی حافظه کوتاه‌مدت طولانی مدت و شبکه‌های بازگشتی بلندمدت دوطرفه توضیح داده می‌شود. همچنین از تبدیل فوریه سریع که یکی از مهم‌ترین الگوریتم‌های مورد استفاده در پردازش سیگنال و آنالیز داده است، برای پیش‌بینی استفاده خواهد شد. همه الگوریتم‌ها ابتدا با استفاده از داده‌ها آموزش داده می‌شود و سپس مدت‌زمان سفر را پیش‌بینی می‌کنند. در انتهای این بخش نیز معیارهای ارزیابی به کار گرفته شده معرفی می‌شوند. محوریت اصلی این پژوهش مقایسه چهار الگوریتم یادگیری ماشین و  $FFT^{10}$  است که وضعیت ترافیک در خیابان کمال را پیش‌بینی می‌کنند. مراحل کلی انجام پژوهش در شکل ۴ نشان داده شده است.

شنا سایی و حذف داده‌های پرت در مجموعه داده‌ها بکار گرفته می‌شود. در تحلیل داده‌های ترافیک، این روش می‌تواند کمک کند تا داده‌هایی که به طور غیرمعمول از دیگر داده‌ها فاصله دارند، شناسایی شده و حذف یا تصحیح شوند تا مدل‌های پیش‌بینی دقیق‌تری ایجاد شود. در شکل ۳ می‌توان به مانا بودن سری زمانی پی برد. مانایی سری مذکور را با آزمون‌های آماری نیز می‌توان بررسی و اثبات نمود.

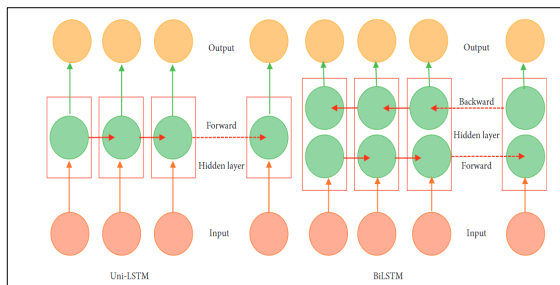


شکل ۳. سری زمانی مربوط به داده ترافیک خیابان کمال

آزمون دیکی فول تقویت شده  $ADF^8$ : آزمون دیکی فولر یکی از پرکاربردترین آزمون‌ها برای بررسی مانایی سری‌های زمانی است. ریشه واحد داشتن به معنای نامانایی است. در آزمون  $ADF$  فرض صفر ( $H_0$ ) این است که سری زمانی دارای ریشه واحد است.



شکل ۴. ساختار کلی روش تحقیق



شکل ۵. معماری بخش ۴،۲ (Abduljabbar et al., 2021)

جدول ۱. ابر پارامترهای مدل برای LSTM و BD-LSTM

مقدار	پارامترها
۳۲،۱۶	تعداد واحدهای LSTM (لایه ۱ و لایه ۲)
Relu	تابع فعال‌سازی
۰/۵	Dropout Rate
Sigmoid	تابع فعال‌سازی لایه خروجی
Adam	بهینه‌ساز
Mean Squared Error	تابع هزینه
۱۵۰	تعداد اپک‌ها
۳۲	Batch Size
۰/۱	Validation Split

#### ۴-۳- روش چهارم: جنگل تصادفی

جنگل تصادفی یک روش یادگیری ماشینی است که برای وظایف مختلف مانند رگرسیون و دسته‌بندی استفاده می‌شود. این روش یک تکنیک تجمعی (ensemble) است که از ترکیب چندین مدل یادگیری (درخت تصمیم‌گیری) برای بهبود دقت و کاهش نوسانات استفاده می‌کند (Liaw & Wiener, 2002). داده آموزش میان مدل‌های درخت تصمیم جداگانه تقسیم شده و پس از آموزش آنها، پیش‌بینی برای هر مدل درخت تصمیم، انجام می‌شود. در الگوریتم جنگل تصادفی، تصمیم نهایی بر اساس رأی اکثریت درخت‌ها (در مدل‌های دسته‌بندی) یا میانگین‌گیری (در مدل‌های رگرسیون) گرفته می‌شود. برای مدل‌های رگرسیون، از جمله مدلی که در این پژوهش استفاده شده است (Random Forest Regressor)، خروجی نهایی به صورت میانگین پیش‌بینی همه درخت‌های تصمیم‌گیری موجود، در جنگل محاسبه می‌شود. این میانگین‌گیری باعث کاهش نوسانات و افزایش دقت مدل می‌شود. گام اول الگوریتم جنگل تصادفی بدین

#### ۴-۱- روش اول: تبدیل فوریه سریع FFT

تحلیل فوریه یک ابزار رایج در پردازش سیگنال است که از آن برای به‌دست‌آوردن تمام اجزای هارمونیک یک سیگنال با استفاده از توابع پایه استفاده می‌شود. تبدیل فوریه بخشی اساسی از تحلیل فوریه است که می‌تواند سیگنال را بین حوزه زمان و فرکانس تبدیل کند. پس از تبدیل فوریه، سیگنال حوزه زمان به برهم‌نهی سیگنال‌های سینوسی متعدد تبدیل می‌شود. در حوزه فرکانس، مشخصه‌های سیگنالی که در حوزه زمان مشهود نیستند به وضوح دیده می‌شوند. از این رو، انجام تبدیل فوریه بر روی سیگنال‌ها برای تجزیه و تحلیل ماهیت آنها بسیار مهم است (Bracewell & Kahn, 1966). روش FFT برای بهبود کارایی محاسباتی DFT<sup>۱۱</sup> طراحی شده است. این روش به دلیل کارایی بهتر در محاسبات در مقایسه با DFT، برای پردازش داده سری زمانی مناسب خواهد بود.

#### ۴-۲- روش دوم و سوم LSTM و Bidirectional-LSTM

در سال‌های اخیر، LSTM یک‌طرفه توجه زیادی را به خود جلب کرده است زیرا عملکرد بهتری نسبت به شبکه‌های عصبی بازگشتی (RNNs) داشته است. حتی با وجود اینکه RNN دقت خوبی را ارائه می‌دهند، مشخص شده است که برای حافظه بلند مدت عملکرد ضعیفی دارند زیرا RNN قادر به استفاده از اطلاعات گذشته دور نیستند. بنابراین، به طور کلی مشخص شده است که LSTM در پیش‌بینی داده‌های سری زمانی از RNN ها بهتر عمل می‌کند (Yeon et al., 2019). افزودن داده‌های آموزشی اضافی منجر به ایجاد برخی از نسخه‌های توسعه‌یافته از LSTM شده است که اکنون به عنوان LSTM دوطرفه (Bidirectional-LSTM) شناخته می‌شوند. این مدل، داده‌های سری زمانی ورودی را دو بار از جهت‌های جلو و بازگشت آموزش می‌دهد. معماری شبکه LSTM و شبکه Bidirectional-LSTM در شکل ۵ نشان داده شده است. مجموعه‌های مختلفی از ابر پارامترها توسط جستجوی شبکه‌ای<sup>۱۲</sup> آزمایش شدند تا ترکیب مناسبی از مقادیر که بهترین دقت را دارند، ارائه دهند. جدول ۱ مقادیر بهینه برای پارامترهای مختلف را نشان می‌دهد.

الگوریتم با استفاده از جستجوی شبکه‌ای بدست آمده و در جدول ۳ نشان داده شده است.

جدول ۲. ابرپارامترهای مربوط به الگوریتم Random Forest

مقدار	پارامتر
۱۰۰۰	تعداد تخمین‌گرها
۱۰۰	حداکثر عمق درخت
۱۰	حداقل نمونه‌ها برای تقسیم
۲۰	حداقل نمونه‌ها در گره برگ
۰	حالت تصادفی
True	بوت‌استرپ

جدول ۳. ابرپارامترهای مربوط به الگوریتم XGBoost

مقدار	پارامترها
Squared Error	تابع هدف
۱۰۰۰	تعداد درختان
۰/۰۱	نرخ یادگیری
۱۰	عمق حداکثر
۰/۱, ۱	Subsample , Colsample bytree
۲۰	gamma
۰	Random state

#### ۴-۵- روش های ارزیابی دقت پیش‌بینی

به‌منظور ارزیابی کیفیت مدل‌های پیش‌بینی حجم ترافیک، استفاده از معیارهایی که امکان مقایسه روش‌های مختلف را فراهم می‌کند، ضروری است. این ارزیابی باید بر روی مقایسه بین نتیجه پیش‌بینی حجم ترافیک و مقدار واقعی حجم ترافیک در آن نقطه از زمان باشد که در این پژوهش ۸ درصد کل داده‌ها به عنوان داده تست در طول فرآیند الگوریتم مورد استفاده قرار می‌گیرد.

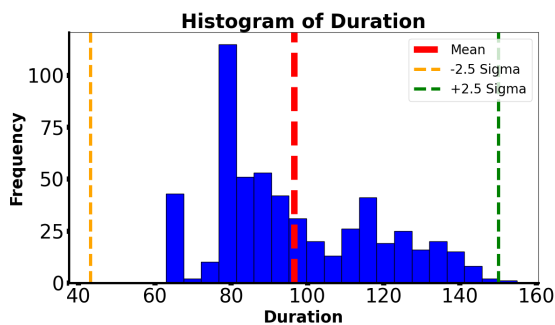
معیارهای رایج زیر در این پژوهش به کار رفته‌اند:

میانگین قدرمطلق خطا (Mean Absolute Error  $\square$ ): این معیار میانگین اختلاف مطلق بین مقادیر پیش‌بینی‌شده و واقعی را نشان می‌دهد که انحراف میانگین مطلق نیز نامیده می‌شود که تفسیر واضحی از میانگین بزرگی خطاها برای پیش‌بینی‌ها ارائه می‌دهد.

صورت است که به تعداد مدل‌های درخت تصمیم مدنظر، نمونه تصادفی از مجموعه داده آموزش برداشت می‌شود. گام دوم آن با استفاده از هر نمونه، یک مدل درخت تصمیم آموزش داده می‌شود. گام سوم نیز پیش‌بینی هر مدل درخت تصمیم برای داده آزمون انجام می‌شود و در نهایت گام چهارم آن، با استفاده از روش‌های میانگین‌گیری، وضعیت ترافیک میانگین‌گیری شده و به عنوان پیش‌بینی نهایی گزارش می‌شود (Biau & Scornet, 2016). این کار باعث می‌شود که واریانس مدل کاهش یافته و پیش‌بینی‌های دقیق‌تر و پایدارتر بدست آیند. این فرآیند باعث می‌شود که جنگل تصادفی در برابر نویز و تغییرات کوچک در داده‌ها مقاوم باشد و عملکرد بهتری نسبت به درخت‌های تصمیم‌گیری منفرد ارائه دهد. همچنین، این روش به دلیل استفاده از نمونه‌های بوت‌استرپ<sup>۳</sup> و انتخاب تصادفی ویژگی‌ها، به کاهش بیش‌برازش کمک می‌کند (Liaw & Wiener, 2002). مقادیر ابرپارامترهای این الگوریتم با استفاده از جستجوی شبکه‌ای بدست آمده است که مطابق با جدول ۲ است.

#### ۴-۴- روش پنجم: الگوریتم درختان تقویت شده با گرادیان

بسته نرم افزاری XGBoost بر مبنای الگوریتم درختان تقویت شده با گرادیان توسعه داده شده و در بسیاری از مسائل چالش برانگیز به نتایج قابل قبولی دست یافته است (Chen & Guestrin, 2016). به عنوان مثال، از میان ۲۹ راه حل برنده چالش منتشر شده در رقابت Kaggle در طول سال ۲۰۱۵، ۱۷ راه حل از XGBoost استفاده کردند (Chen & Guestrin, 2016). مدل درختان تقویت شده با ترکیب مدل‌های یادگیری پایه ضعیف، یادگیرنده قوی‌تری را تشکیل می‌دهد. این الگوریتم، با استفاده از توابع هدف و روش‌های بهینه‌سازی پیچیده، به‌طور دقیق‌تر و کارآمدتر از درخت‌های تصمیم ساده عمل می‌کند. از نظر سرعت، عملکرد و دقت، XGBoost به‌عنوان یک مدعی برتر در میان الگوریتم‌های یادگیری ماشینی قرار می‌گیرد. توانایی آن در مدیریت داده‌های پیچیده، کارایی محاسباتی و دقت بالا، آن را به ابزاری ارزشمند برای طیف وسیعی از کاربردها تبدیل کرده است. نشان داده شده است که XGBoost برای مشکلات پیش‌بینی سری‌های زمانی مؤثر است (Zhang et al., 2021). مقادیر ابرپارامترهای این



شکل ۶. نمودار هیستوگرام داده مدت زمان سفر خیابان کمال

جدول ۴. اطلاعات آماری مدت زمان سفر خیابان کمال

۵۷۶	تعداد داده
۹۶/۷۱	میانگین
۲۱/۶۷	انحراف استاندارد
۶۳/۰۰	کمترین مقدار
۸۰/۰۰	صدک ۲۵
۹۱/۰۰	صدک ۵۰
۱۱۵/۰۰	صدک ۷۵
۱۷۶/۰۰	بیشترین مقدار

برای بررسی مانایی یا نامانایی سری زمانی این پژوهش با توجه به اطلاعات وجود در جدول ۴، از آزمون های ADF و KPSS استفاده شد. برای هر دو آزمون، شرایط برای رد فرضیه صفر ( $H_0$ ) به گونه ای است که اگر آماره آزمون کوچک تر از مقدار بحرانی و مقدار  $p\text{-value} \leq 0.05$  باشد، فرضیه صفر ( $H_0$ ) رد خواهد شد. در غیر این صورت فرضیه صفر مورد قبول واقع می شود. با توجه به نتایجی که از دو آزمون ADF و KPSS نشان می دهد؛ می توان نتیجه گرفت که فرضیه صفر آزمون ADF رد می شود و سری زمانی مربوطه مانا و ثابت خواهد بود و فرضیه صفر آزمون KPSS قبول شده که مجدداً باعث اثبات مانایی سری زمانی مورد استفاده در این پژوهش خواهد شد.

مطابق با داده ترافیک خیابان کمال در شکل ۷، نمودار تبدیل فوریه سیگنال سری زمانی را نشان می دهد. محور افقی فرکانس بر حسب هرتز و محور عمودی اندازه یا دامنه فرکانس های متناظر را نشان می دهد. نرخ نمونه برداری در این پژوهش یک ساعت بود. دامنه فرکانس در نمودار تبدیل فوریه (FFT) نشان دهنده قدرت یا شدت هر فرکانس خاص در سیگنال اصلی است. فرکانس صفر هرتز، دامنه ۵۵۶۰۶ را داراست که این نشان دهنده یک مؤلفه DC قوی (سیگنال ثابت) در فرکانس

ریشه میانگین مربعات خطا: (Root Mean Square Error - RMSE) | این معیار جذر میانگین اختلاف مربع بین مقادیر

مشاهده شده و پیش بینی شده را نشان می دهد.

میانگین درصد قدرمطلق خطا - (Mean Absolute Percentage Error | MAPE) : خطا به میانگین خطای نسبی نیز معروف است. مشابه میانگین قدرمطلق خطا است، اما به جای خطا، از خطای نسبی استفاده شده است.

معیار  $R^2$ : این معیار به عنوان ضریب تعیین تعریف می شود و نشان می دهد چه کسری از واریانس متغیر وابسته (حجم ترافیک) توسط متغیرهای مستقل توضیح داده می شود. به عبارت دیگر،  $R^2$  میزان برازش مدل به داده ها را نشان می دهد.

## ۵- پیاده سازی و نتایج

در ادامه به پیاده سازی و بررسی نتایج پیش بینی حجم ترافیک با استفاده از روش های ذکر شده (بخش ۴) پرداخته می شود. داده های خام حاوی نویز و مقادیر از دست رفته هستند. به همین خاطر، به منظور تبدیل داده های خام به داده های قابل اعتماد پیش پردازش شد. با توجه به اینکه سری موجود در این پژوهش به صورت سری زمانی مانا است (بخش ۳، ۳)، بر روی داده ها، تست سه سیگما جهت حذف مقادیر پرت اعمال شد (شکل ۶). همچنین برای جایگزین داده های حذف شده، از میانگین داده های هفته قبل و بعد شده است. در طی فرآیند، داده ها بین ۰ و ۱ مقیاس بندی و مطابق با رابطه (۱) استاندارد شدند. در مرحله بعد، هر یک از الگوریتم های Random Forest، LSTM، BD LSTM، XGBoost و FFT از طریق داده آموزشی آموزش دیده و در نهایت با بهینه سازی پارامترهای مربوط به هر الگوریتم و اجرای الگوریتم بر روی داده های تست، میزان دقت آن سنجیده شد.

$$X_{norm} = \frac{X(i) - \min(x)}{\max(x) - \min(x)} \quad (1)$$

با مقدار  $5436/86$  است که نشان می‌دهد این فرکانس بیشترین تأثیر را در سیگنال دارد. نتایج به دست آمده به منظور پیش‌بینی حجم ترافیک با استفاده از روش‌های ذکر شده در بخش روش تحقیق، به شرح زیر است.

#### ۵-۱- روش اول: استفاده از مدل FFT

در این روش بر روی داده‌های مدت‌زمان سفر، سه مدل تبدیل فوریه سریع اجرا و در نهایت میزان دقت پیش‌بینی آن‌ها بررسی شد. ابتدا یک FFT ساده بر روی داده خیابان کمال اجرا می‌شود. سپس در روش "Improvement 1" که روی الگوریتم FFT (تبدیل فوریه سریع) اعمال می‌شود، ابتدا داده‌های آموزشی پیش از اعمال FFT برش داده می‌شود. به این صورت که بخش‌هایی از داده‌های آموزشی که به بازه‌های زمانی مشابه با داده‌های پیش‌بینی تعلق دارند، جدا می‌شود. این کار به این منظور انجام می‌شود که داده‌های آموزشی بیشتر با الگوهای زمانی فصلی داده‌های پیش‌بینی همخوانی داشته باشند. در نتیجه، الگوریتم FFT به شکلی دقیق‌تر بتواند الگوهای فصلی و تکراری داده‌ها را شناسایی کرده و پیش‌بینی‌های بهتری ارائه دهد.

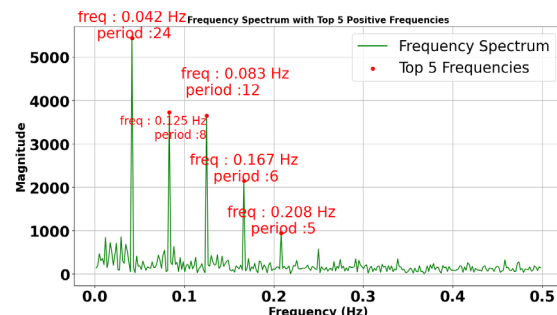
در روش "Improvement 2"، سیگنال اولیه با استفاده از تبدیل فوریه سریع (FFT) به مؤلفه‌های فرکانسی تجزیه می‌شود و سپس فرکانس‌های دارای دامنه پایین، که اغلب به عنوان نویز تلقی می‌شوند، فیلتر می‌شوند. حذف این فرکانس‌های کم‌دامنه به حفظ روندهای فصلی قوی‌تر و کاهش نویز منجر می‌شود. مطابق با جدول ۶ مشاهده می‌شود که استفاده از  $2/5$  سیگما و شناسایی داده پرت و حذف آن‌ها، دقت الگوریتم را افزایش می‌دهد. باتوجه به جدول ۶، FFT ساده دقت کمتری نسبت به سایر الگوریتم‌های FFT دارد.

با مقایسه  $2$  FFT-Improvement با یکدیگر می‌توان نتیجه گرفت که تست  $2.5$  سیگما بر روی افزایش دقت، تأثیر به‌سزایی دارد و میزان  $R^2$  را از  $0/64$  به  $0/81$  می‌رساند/میزان RMSE را به اندازه  $4/05$  ثانیه کاهش داده است و در بهترین حالت به طور متوسط  $7/54$  درصد خطا در پیش‌بینی‌های خود دارد.

صفر است. این مؤلفه در طیف فرکانسی به عنوان یک پیک در فرکانس صفر ظاهر می‌شود. برای سیگنال‌های حقیقی، تبدیل فوریه شامل فرکانس‌های مثبت و منفی است. در بسیاری از کاربردها، مؤلفه DC ممکن است به دلیل اینکه به تحلیل فرکانسی واقعی کمک نمی‌کند، حذف شود. برای این کار، معمولاً میانگین سیگنال از داده‌ها کم می‌شود. تا نوسانات فرکانسی باقی بماند. برای کاهش تأثیر مؤلفه DC و تحلیل دقیق‌تر فرکانس‌ها، می‌توان ابتدا میانگین داده‌ها را از هر مقدار کم کرد. این فرآیند کمک می‌کند تا مؤلفه DC حذف شده و سایر فرکانس‌ها بهتر دیده شوند (شکل ۷).

جدول ۵. بررسی مانایی و یا نامانایی سری زمانی

KPSS	ADF	
۰/۰۱۳	-۱۱/۸۰	میزان آماره
۰/۱	۹/۵۰-۲۲	p-value
۰/۰۱۴	-۳/۴۴۲	مقادیر بحرانی ۱درصد
۰/۴۶۳	-۲/۸۶۷	مقادیر بحرانی ۵درصد
۰/۳۴۷	-۲/۵۷۰	مقادیر بحرانی ۱۰درصد
مورد قبول	رد	وضعیت $H_0$
مانا	مانا	نوع سری



شکل ۷. فرکانس‌های موجود مدت زمان سفر خیابان کمال

مطابق با شکل ۷، فرکانس  $0/042$  هرتز مربوط به هر ۲۴ ساعت، فرکانس  $0/208$  هرتز مربوط به هر ۵ ساعت است. در اینجا، بزرگ‌ترین اندازه مربوط به فرکانس  $0/042$  هرتز

جدول ۶. نتایج بدست آمده از الگوریتم FFT

معیار ارزیابی / آزمایش	RMSE (Sec)	MAE (Sec)	MAPE (%)	$R^2$
Basic FFT	۱۶/۳۸	۱۱/۴۲	۱۱/۷۰	۰/۴۶
FFT- Improvement 1	۱۵/۷۰	۱۱/۹۲	۱۲/۱۶	۰/۵۱
FFT- Improvement 2	۱۱/۹۳	۸/۴۰	۸/۶۰	۰/۶۴
FFT- Improvement 2	۷/۸۸	۷/۹۰	۷/۵۴	۰/۸۱

۲-۵- روش دوم و روش سوم: استفاده از مدل LSTM و مدل Bidirectional LSTM

جدول ۷. جدول مربوط به معیارهای ارزیابی LSTM همراه با ویژگی‌های ورودی

ویژگی‌ها آزمایشی	مدت زمان سفر	استفاده از تاریخچه داده‌ها	مجموع حجم ترافیک یال (ها) ورودی	زمان	روز هفته	تعداد یال ورودی	سایر ویژگی‌ها شامل: تعداد بانک‌ها، شهرداری و..	RMSE (Sec)	MAE (Sec)	MAPE (%)	R <sup>2</sup>
۱	✓	✗	✗	✗	✗	✗	✗	۶/۱۶	۴/۶۹	۰/۰۴۸	۰/۹۲۱
۲	✓	✓	✗	✗	✗	✗	✗	۶/۶۳	۴/۶۲	۰/۰۴۴	۰/۹۰۹
۳	✓	✗	✓	✗	✗	✗	✗	۵/۷۹	۳/۹۷	۰/۰۳۶	۰/۹۳۰
۴	✓	✗	✗	✓	✗	✗	✗	۶/۵۴	۴/۴۰	۰/۰۴۲	۰/۹۱۶
۵	✓	✗	✗	✗	✓	✗	✗	۶/۶۸	۴/۶۰	۴/۵۹۷	۰/۹۰۷
۶*	✓	✓	✓	✗	✗	✗	✗	۱۲/۶۰	۵/۵۶	۰/۰۴۹	۰/۷۲۹
۷	✓	✓	✓	✗	✗	✗	✗	۵/۹۴	۴/۱۷	۰/۰۴۱	۰/۹۲۷
۸	✓	✓	✓	✓	✗	✗	✗	۶/۴۰	۴/۳۹	۰/۰۴۲	۰/۹۱۵
۹	✓	✓	✓	✓	✓	✗	✗	۶/۱۴	۴/۳۹	۰/۰۴۶	۰/۹۲۲
۱۰	✓	✗	✗	✗	✗	✓	✗	NAN	NAN	NAN	NAN
۱۱	✓	✗	✗	✗	✗	✗	✓	NAN	NAN	NAN	NAN

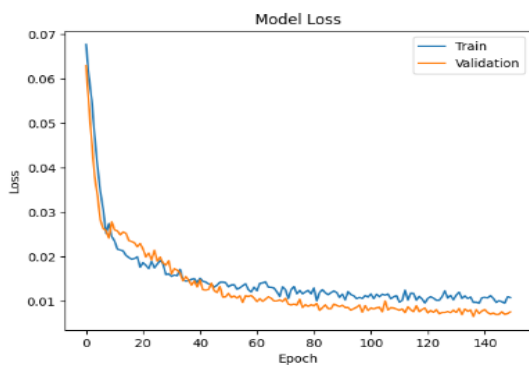
جدول ۸. جدول مربوط به معیارهای ارزیابی Bidirectional LSTM همراه با ویژگی‌های ورودی

ویژگی‌ها آزمایشی	مدت زمان سفر	استفاده از تاریخچه داده‌ها	مجموع حجم ترافیک یال (ها) ورودی	زمان	روز هفته	تعداد یال ورودی	سایر ویژگی‌ها شامل: تعداد بانک‌ها، شهرداری و..	RMSE (Sec)	MAE (Sec)	MAPE (%)	R <sup>2</sup>
۱	✓	✗	✗	✗	✗	✗	✗	۶/۰۶	۴/۴۷	۰/۰۴۵	۰/۹۲۲
۲	✓	✓	✗	✗	✗	✗	✗	۷/۲۷	۵/۶۳	۰/۰۵۳	۰/۸۸۹
۳	✓	✗	✓	✗	✗	✗	✗	۶/۴۰	۴/۶۵	۰/۰۴۶	۰/۹۱۵
۴	✓	✗	✗	✓	✗	✗	✗	۶/۰۸	۴/۳۸	۰/۰۴۳	۰/۹۲۳
۵	✓	✗	✗	✗	✓	✗	✗	۶/۶۷	۴/۸۶	۴/۸۵۸	۰/۹۰۸
۶*	✓	✓	✓	✗	✗	✗	✗	۱۲/۵۳	۵/۸۳	۰/۰۵۳	۰/۷۳۲
۷	✓	✓	✓	✗	✗	✗	✗	۶/۶۵	۴/۷۵	۰/۰۴۶	۰/۹۰۸
۸	✓	✓	✓	✓	✗	✗	✗	۶/۰۶	۴/۴۶	۰/۰۴۲	۰/۹۲۸
۹	✓	✓	✓	✓	✓	✗	✗	۶/۰۶	۴/۴۷	۰/۰۴۵	۰/۹۲۴
۱۰	✓	✗	✗	✗	✗	✓	✗	NAN	NAN	NAN	NAN
۱۱	✓	✗	✗	✗	✗	✗	✓	NAN	NAN	NAN	NAN

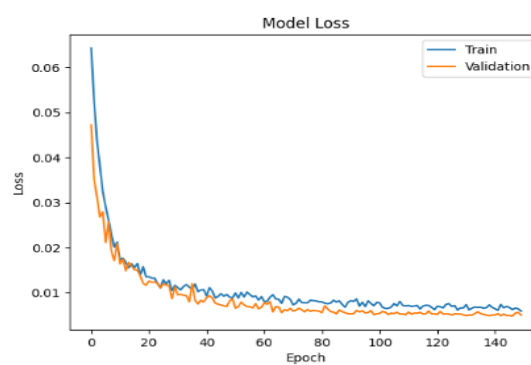
خواهد بود. استفاده از ویژگی روزهای هفته، دقت مدل‌ها را تا حدودی کاهش می‌دهد (آزمایش ۵). ویژگی‌های مدت‌زمان سفر و تاریخچه، نسبت به سایر ویژگی‌ها دقت مدل Bidirectional LSTM را کاهش می‌دهد (آزمایش ۲-جدول ۸). اضافه کردن سایر ویژگی‌ها نظیر تعداد بانک‌ها، تعداد مدارس و.... در شعاع ۱۵۰ متری خیابان کمال در الگوریتم LSTM و Bidirectional LSTM منجر به NAN شدن Loss می‌شود که در اجرای معیارهای ارزیابی دچار مشکل شود و باعث عدم اجرای آن شود (آزمایش ۱۰ و ۱۱ در جدول ۷ و جدول ۸). به طور کلی در هر دو روش LSTM و Bidirectional LSTM استفاده از ویژگی‌های مدت‌زمان سفر، داده‌های تاریخچه ترافیک، مجموع ترافیک یال(ها) ورودی به همراه ویژگی‌های زمانی (ساعت) و روزهای هفته به بهبود دقت مدل کمک می‌کند (آزمایش ۹-جدول ۷ و جدول ۸). شکل ۸ نشان‌دهنده روند کاهش خطا (Loss) در طی آموزش مدل است. محور افقی تعداد تکرارها (Epochs) و محور عمودی میزان خطا (Loss) را نشان می‌دهد. از کل داده‌ها، ۰/۸ به عنوان داده validation در نظر گرفته شده است. در اوایل آموزش برای هر دو الگوریتم LSTM و Bidirectional LSTM مقدار خطا برای هر دو مجموعه داده، بالا است که طبیعی است زیرا مدل در حال یادگیری است. با افزایش تعداد تکرارها، خطا در هر دو مجموعه کاهش می‌یابد که نشان می‌دهد مدل در حال یادگیری و بهبود است. این نمودارها نشان می‌دهد که مدل‌ها در حال یادگیری و بهبود هستند و به نظر می‌رسد که بیش از حد یادگیری رخ نداده است، زیرا خطاهای آموزشی و اعتبارسنجی به طور همزمان کاهش یافته و به سطوح پایین رسیده‌اند.

طبق جدول ۷ و جدول ۸ نتایجی که بر روی داده‌های تست توسط مدل‌های LSTM و Bidirectional LSTM بدست آمده است، نشان می‌دهد. LSTM به‌خاطر توانایی در یادگیری الگوهای زمانی، برای پیش‌بینی حجم ترافیک بسیار مؤثر هستند. این مدل‌ها از حافظه داخلی برای یادآوری اطلاعات گذشته استفاده می‌کنند و می‌توانند الگوهای تکراری در داده‌های ترافیکی را تشخیص دهند. حتی بدون ویژگی‌های ورودی اضافی، داده‌های ترافیکی خام معمولاً شامل اطلاعات کافی برای پیش‌بینی دقیق هستند. این قابلیت‌ها باعث می‌شود که LSTM دقت بالایی در پیش‌بینی‌های ترافیک داشته باشند که این موضوع را می‌توان از دقت (آزمایش ۱) در جدول ۷ و جدول ۸ مشاهده کرد. از مقایسه آزمایش شماره ۶ و آزمایش ۷ در جدول ۷ و جدول ۸ می‌توان دریافت که در نظر گرفتن تست ۲/۵ سیگما تاثیر بالایی در افزایش دقت خواهد شد. این میزان افزایش دقت  $R^2$  در LSTM ۱۹/۸۰ درصد و در Bidirectional LSTM به میزان ۱۷/۶۴ درصد است. در جدول ۷ و جدول ۸ ترکیب‌های مختلف ویژگی‌های ورودی، تأثیر قابل‌توجهی بر دقت مدل دارند. همانطور که در جدول ۷ مشاهده می‌شود، بهترین عملکرد (کمترین خطا و بالاترین میزان  $R^2$ ) برای LSTM استفاده از ویژگی مدت‌زمان سفر و مجموع حجم ترافیک یال‌های ورودی است (آزمایش ۳-جدول ۷). در Bidirectional LSTM استفاده از مدت‌زمان سفر و داده‌های تاریخی ترافیک، مجموع ترافیک یال(ها) ورودی به همراه ویژگی زمانی (ساعت)، بهترین نتیجه بدست آمده را به همراه دارد (آزمایش ۸-جدول ۸).

در تمامی جداول رنگ سبز به معنی بهترین نتیجه بدست آمده توسط مدل و رنگ قرمز به معنی ضعیف‌ترین نتیجه بدست آمده است. علامت \* به معنی عدم در نظر گرفتن تست ۲.۵ سیگما



LSTM



Bidirectional LSTM

شکل ۸. روند خطا در طی فرآیند آموزش توسط الگوریتم LSTM و Bidirectional LSTM

### ۳-۵- روش چهارم و پنجم: استفاده از روش Random Forest و XGBoost

جدول ۹. جدول مربوط به معیارهای ارزیابی Random Forest همراه با ویژگی‌های ورودی

ویژگی‌ها آزمایشی	استفاده از تاریخچه داده‌ها	مجموع حجم ترافیک یال (ها) ورودی	زمان	روز هفته	تعداد یال ورودی	سایر ویژگی‌ها شامل: تعداد بانک‌ها، شهرداری و..	RMSE (Sec)	MAE (Sec)	MAPE (%)	R <sup>2</sup>
۱	✓	✗	✗	✗	✗	✗	۶/۰۱	۴/۷۲	۰/۰۴۹	۰/۹۲۳
۲	✓	✗	✗	✗	✗	✗	۶/۰۰	۴/۶۸	۰/۰۴۹	۰/۹۲۳
۳	✗	✓	✗	✗	✗	✗	۶/۸۰	۵/۲۰	۰/۰۵۵	۰/۹۰۱
۴	✗	✗	✓	✗	✗	✗	۷/۳۸	۶/۱۸	۰/۰۶۳	۰/۸۸۴
۵	✗	✗	✗	✓	✗	✗	۲۱/۸۳	۱۸/۸۰	۰/۱۹۹	-۰/۰۱۹
۶*	✗	✗	✗	✗	✓	✗	۲۱/۷۲	۱۸/۶۵	۰/۱۹۶	-۰/۰۰۹
۷	✗	✗	✗	✗	✗	✓	۲۱/۷۲	۱۸/۶۵	۰/۱۹۶	-۰/۰۰۹
۸	✗	✗	✗	✗	✓	✓	۲۱/۷۲	۱۸/۶۵	۰/۱۹۶	-۰/۰۰۹
۹	✓	✓	✓	✗	✗	✗	۵/۸۸	۴/۳۵	۰/۰۴۶	۰/۹۲۶
۱۰	✓	✓	✓	✓	✗	✗	۵/۸۱	۴/۲۸	۰/۰۴۵	۰/۹۲۸
۱۱	✓	✓	✓	✓	✓	✗	۵/۸۲	۴/۲۸	۰/۰۴۵	۰/۹۲۸

جدول ۱۰. جدول مربوط به معیارهای ارزیابی XGBoost همراه با ویژگی‌های ورودی

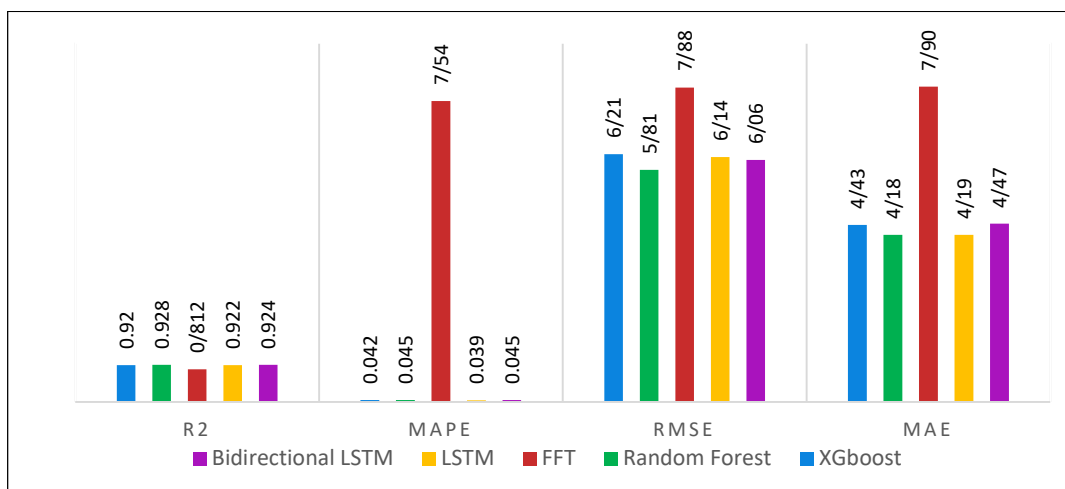
ویژگی‌ها آزمایشی	استفاده از تاریخچه داده‌ها	مجموع حجم ترافیک یال (ها) ورودی	زمان	روز هفته	تعداد یال ورودی	سایر ویژگی‌ها شامل: تعداد بانک‌ها، شهرداری و..	RMSE (Sec)	MAE (Sec)	MAPE (%)	R <sup>2</sup>
۱	✓	✗	✗	✗	✗	✗	۶/۵۲	۵/۲۴	۰/۰۵۶	۰/۹۰۹
۲	✓	✗	✗	✗	✗	✗	۶/۴۸	۵/۱۷	۰/۰۵۷	۰/۹۱۰
۳	✗	✓	✗	✗	✗	✗	۶/۷۰	۵/۸۹	۰/۰۵۵	۰/۹۰۴
۴	✗	✗	✓	✗	✗	✗	۵/۸۰	۴/۵۱	۰/۰۴۷	۰/۹۲۸
۵	✗	✗	✗	✓	✗	✗	۲۲/۰۰	۱۸/۷۳	۰/۱۹۷	-۰/۰۳۷
۶*	✗	✗	✗	✗	✓	✗	۲۱/۷۳	۱۸/۶۰	۰/۱۹۴	-۰/۰۱۴
۷	✗	✗	✗	✗	✗	✓	۲۱/۷۳	۱۸/۶۰	۰/۱۹۴	-۰/۰۱۴
۸	✗	✗	✗	✗	✓	✓	۲۱/۷۷	۱۸/۵۳	۰/۱۹۴	-۰/۰۱۵
۹	✓	✓	✓	✗	✗	✗	۶/۳۴	۴/۸۱	۰/۰۵۱	۰/۹۱۴
۱۰	✓	✓	✓	✓	✗	✗	۶/۲۱	۴/۴۳	۰/۰۴۲	۰/۹۲۰
۱۱	✓	✓	✓	✓	✓	✗	۶/۲۴	۴/۸۱	۰/۰۵۱	۰/۹۱۷

ویژگی‌های ورودی در مدل **Random Forest** و **XGBoost** مطابق با جدول ۹ و جدول ۱۰ است و متغیری که پیش‌بینی می‌کند، مدت زمان سفر خواهد بود. در هر دو مدل استفاده از ویژگی‌های روز هفته، تعداد یال ورودی و سایر ویژگی‌ها شامل تعداد بانک‌ها و شهرداری‌ها چه به صورت جدا و چه همزمان با یکدیگر، باعث منفی شدن مقدار  $R^2$  می‌شود و مدل به خوبی مدت زمان سفر را پیش‌بینی نمی‌کند. در جدول ۱۰ ویژگی زمان توانسته به طور قابل توجهی دقت مدل **XGBoost** را افزایش دهد. نتایج مربوطه نشان می‌دهند که مدل عملکرد بسیار خوبی دارد و پیش‌بینی‌های دقیق‌تری نسبت به مقادیر واقعی ارائه می‌دهد.  $R^2$  مقدار ۰/۹۳ را نشان می‌دهد که توانسته است بخش بزرگی از واریانس داده‌ها را توضیح دهد. مقادیر کم **MAPE** و **MAE** و **RMSE** نیز نشان‌دهنده خطای کم در پیش‌بینی‌ها هستند (آزمایش ۴-جدول ۱۰). در مدل **XGBoost** افزودن ویژگی تعداد یال ورودی و سایر ویژگی‌ها به ردیف ۱۰ دقت مدل ( $R^2$ ) را به میزان ۰/۳۳ درصد کاهش می‌دهد ولی در **Random Forest** هیچ تغییری ایجاد نمی‌کند. حذف داده‌های پرت (استفاده از قاعده ۲,۵ سیگما) به بهبود دقت کمک می‌کند. این موضوع را در مقایسه آزمایش ۱ و ۲ می‌توان فهمید. در مدل **Random Forest** ویژگی زمان، برخلاف **XGBoost** نتوانسته بهترین دقت را به همراه داشته باشد (ردیف ۴-جدول ۹). بهترین نتیجه بدست آمده در **Random Forest** استفاده از ویژگی‌های تاریخچه، مجموع حجم ترافیک یال(ها) ورودی

و زمان است (ردیف ۹-جدول ۹). ضعیف‌ترین نتیجه بدست آمده برای هر دو مدل **XGBoost** و **Random Forest**، استفاده از ویژگی‌های روزهای هفته است (ردیف ۵). در **Random Forest** استفاده از ویژگی‌های مانند تاریخچه داده‌ها، مجموع ترافیک یال ورودی، زمان و روز هفته به صورت همزمان، دقت مدل را نسبت به بهترین دقت آن، کاهش می‌دهد؛ اما این کاهش چشمگیر نیست. به ترتیب میزان  $R^2$  به میزان ۰/۸ درصد و ۰/۲ درصد کاهش پیدا کرده است (ردیف ۱۰). افزودن ویژگی تعداد یال ورودی و سایر ویژگی‌ها به ردیف ۱۰ دقت مدل در **XGBoost** تغییر می‌دهد ولی در **Random Forest** میزان دقت مدل را تغییر نمی‌دهد. با مشاهده جدول ۱۱ و شکل ۹ می‌توان نتیجه گرفت، دقت اکثر الگوریتم‌ها و روش‌های ذکر شده نزدیک به هم هستند ولی الگوریتم **FFT** دارای کمترین دقت و الگوریتم **Random Forest** بهترین دقت را داراست. ویژگی‌های ورودی تمامی الگوریتم‌های یادگیری ماشین، جهت مقایسه یکدیگر می‌بایستی مانند هم باشد که شامل: تاریخچه داده‌ها، مجموع حجم ترافیک یال (ها) ورودی، زمان و روز هفته است و برای الگوریتم **FFT** از هیچ گونه ویژگی استفاده نشده است. تمامی الگوریتم‌ها توسط ۸۲ درصد داده‌ها آموزش دیده و ۸ درصد پایانی داده‌ها مورد ارزیابی قرار گرفته‌اند.

جدول ۱۱. میزان دقت پیش‌بینی مدت زمان سفر توسط الگوریتم‌های ذکر شده

$R^2$	MAPE (%)	RMSE(Sec)	MAE(Sec)	روش
0/812	7/54	7/88	7/90	<b>FFT</b>
0/922	0/039	6/14	4/19	<b>LSTM</b>
0/924	0/045	6/06	4/47	<b>Bidirectional LSTM</b>
0/928	0/045	5/81	4/18	<b>Random Forest</b>
0/920	0/042	6/21	4/43	<b>XGboost</b>



شکل ۹. مقایسه شاخص‌های ارزیابی **MAE**، **RMSE**، **MAPE** و  $R^2$  الگوریتم‌های ذکر شده

نمی‌شود. این نشان می‌دهد که مدل توانسته است تا حدود زیادی تغییرات مدت زمان را به خوبی مدل‌سازی کند و هیچ رابطه سیستماتیکی بین خطاهای پیش‌بینی و مدت زمان پیش‌بینی شده وجود ندارد. ولی در مدل FFT برخی نقاط با وجود پراکندگی تصادفی نقاط، برخی نقاط با باقیمانده‌های بزرگ (مثبت یا منفی) وجود دارند. این نشان می‌دهد که مدل ما در پیش‌بینی مدت زمان برخی از رویدادها با مشکل مواجه شده است.

این نتایج نشان می‌دهند که به‌طور کلی همه الگوریتم‌های ذکر شده به جز FFT از لحاظ معیارهای ارزیابی بسیار شبیه به هم هستند. روش Random Forest عملکرد بهتری نسبت به سایر روش‌ها در کاهش خطاها RMSE، MAE،  $R^2$  دارد و روش XGBoost خطای MAPE کمتری نسبت به دیگر مدل‌ها دارد. در حالی که روش FFT دارای بالاترین مقادیر خطا است و به طور قابل توجهی ضعیف‌تر از بقیه الگوریتم‌ها عمل کرده است.

مقادیر پیش‌بینی شده داده‌های تست توسط تمامی الگوریتم‌های ذکر شده در شکل ۱۱ قابل مشاهده است. با توجه به نمودار، الگوریتم FFT نتوانسته به خوبی قله‌ها و دره‌ها را پیش‌بینی کند. همچنین، با مقایسه مقادیر پیش‌بینی شده با مقادیر واقعی، مشاهده می‌شود که پیش‌بینی‌ها دچار مقداری جابه‌جایی به سمت چپ شده‌اند.

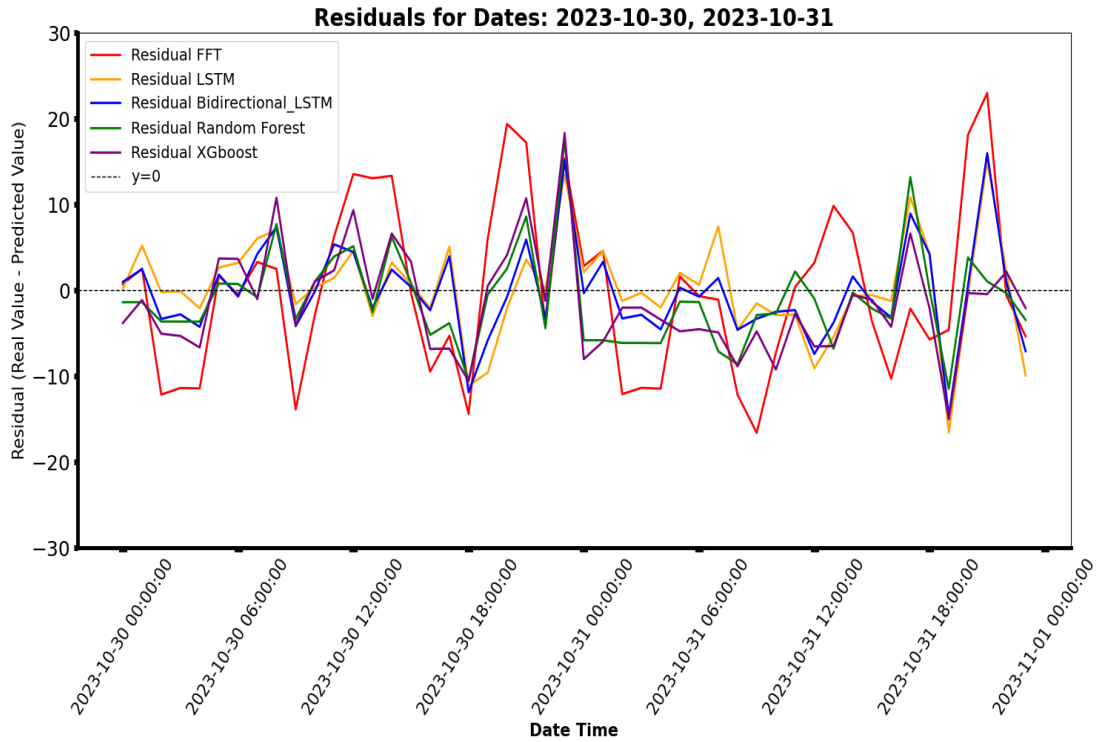
نموداری که مقادیر واقعی روی محور X و مقادیر پیش‌بینی شده روی محور Y قرار دارند، نمودار پراکندگی (Scatter Plot) است که برای بررسی کیفیت مدل‌های پیش‌بینی کننده استفاده می‌شود. یک خط ۴۵ درجه (خط ایده‌آل) از مبدا نمودار عبور می‌کند که نشان‌دهنده حالتی است که مقادیر پیش‌بینی شده دقیقاً برابر با مقادیر واقعی هستند. این خط به عنوان خط  $y=x$  شناخته می‌شود و به عنوان مرجع در نمودار استفاده می‌شود. اگر نقاط داده در نمودار پراکندگی نزدیک به این خط قرار گیرند، نشان می‌دهد که مدل توانسته پیش‌بینی دقیقی انجام دهد و پیش‌بینی‌های آن نزدیک به مقادیر واقعی است. فاصله بیشتر نقاط از این خط نشان‌دهنده خطای بیشتر در پیش‌بینی‌ها است. کمترین فاصله نقاط داده تا خط  $y=x$  مربوط به الگوریتم‌های Bidirectional LSTM، Random Forest، XGBoost و LSTM است. بیشترین خطا و بیشترین فاصله نقاط داده مربوط به FFT است (شکل ۱۲).

الگوریتم FFT پایین‌ترین مقدار  $R^2$  و Random Forest بالاترین مقدار را دارد که نشان‌دهنده قدرت پایین الگوریتم FFT است. الگوریتم‌های XGBoost، Random Forest، LSTM و Bidirectional LSTM مقادیر  $R^2$  بسیار نزدیک به هم دارند. به نظر می‌رسد توانایی مشابهی در توضیح تغییرات متغیر وابسته داشته باشند. میانگین عملکرد همه الگوریتم‌ها نیز مقداری نزدیک به مقادیر بهترین الگوریتم را دارند.

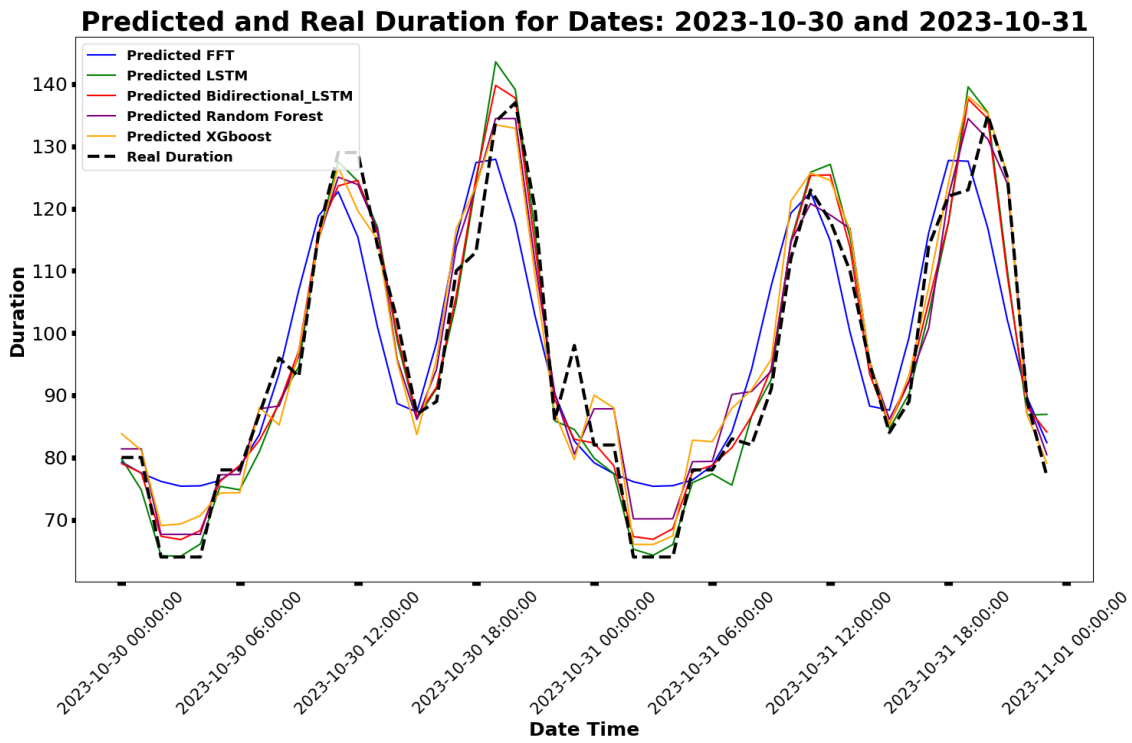
در شکل ۹ مقدار RMSE کمتر نشان‌دهنده عملکرد بهتر است؛ بنابراین در اینجا بهترین روش، الگوریتم Random Forest با مقدار  $0.05814$  ثانیه است و ضعیف‌ترین روش FFT با مقدار  $0.07877$  ثانیه است. تفاوت درصدی بین بهترین روش RandoForest و ضعیف‌ترین روش FFT حدود  $0.3548$  درصد است. این تفاوت حاکی از آن است که روش Random Forest عملکرد بهتری نسبت به روش FFT در کاهش خطای RMSE دارد.

مقدار MAE کمتر هم نشان‌دهنده عملکرد بهتر است که در شکل ۹ می‌توان مقادیر مربوط به MAE هر مدل را مشاهده کرد. در اینجا بهترین روش، Random Forest با مقدار  $0.04281$  ثانیه است و ضعیف‌ترین روش، FFT با مقدار  $0.07897$  ثانیه است. تفاوت درصدی بین بهترین روش Random Forest و ضعیف‌ترین روش، FFT حدود  $0.4684$  درصد است. این تفاوت نشان می‌دهد که روش Random Forest عملکرد بسیار بهتری نسبت به روش FFT در کاهش خطای MAE دارد. همچنین در شکل ۹، تفاوت MAPE بین بهترین و ضعیف‌ترین الگوریتم بسیار زیاد است. الگوریتم FFT با خطای پیش‌بینی بسیار بالا، عملکرد بسیار ضعیفی نسبت به سایر الگوریتم‌ها دارد. در مقابل، الگوریتم XGBoost با کمترین مقدار MAPE، بهترین عملکرد را نشان می‌دهد.

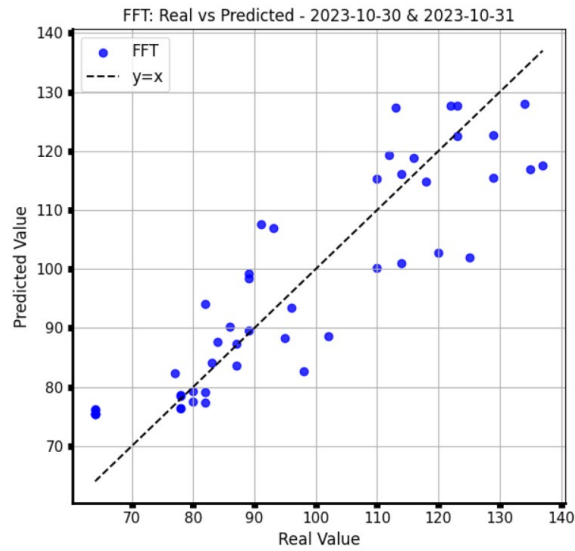
باتوجه به شکل ۱۰ محور عمودی (Residuals) است و تفاوت بین مقادیر واقعی و مقادیر پیش‌بینی شده را نشان می‌دهد. نقاط در نمودار به صورت تصادفی در اطراف خط افقی صفر پراکنده شده‌اند. اگر نقاط به صورت منظم در بالا یا پایین خط صفر قرار بگیرند، نشان‌دهنده این است که مدل دارای بایاس (Bias) است و به طور سیستماتیک مقادیر را بیش‌تر یا کمتر از مقدار واقعی تخمین می‌زند. در نمودارهای الگوی خاصی بین باقیمانده‌ها و مدت زمان پیش‌بینی شده مشاهده



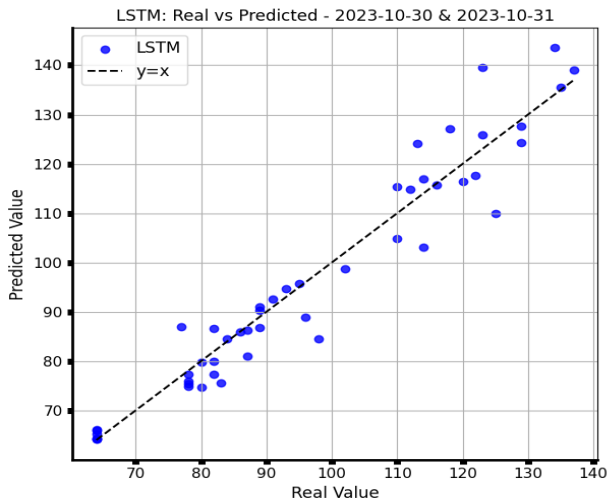
شکل ۱۰. نمودار Residual - الگوریتم‌های ذکر شده



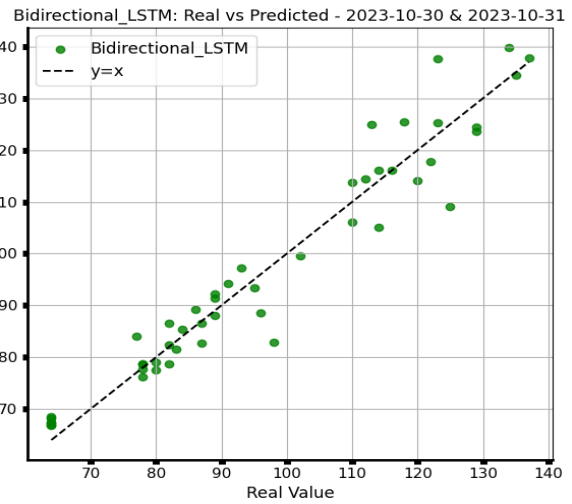
شکل ۱۱. مقدار مدت زمان سفر پیش‌بینی شده توسط مدل‌های پیشنهادی (۳۰ و ۳۱ اکتبر ۲۰۲۳)



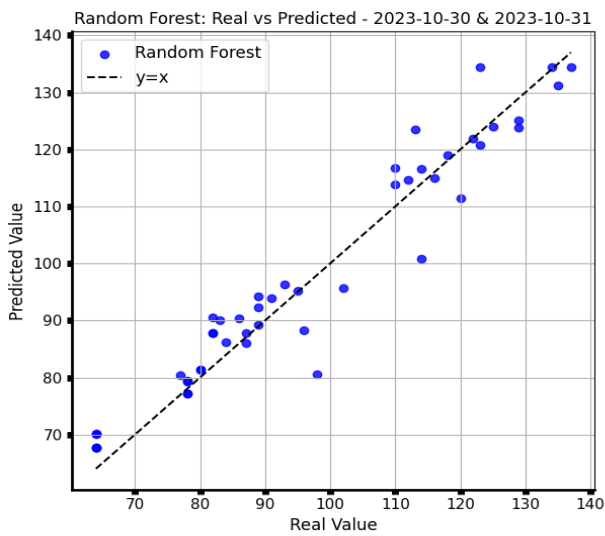
الف) روش FFT



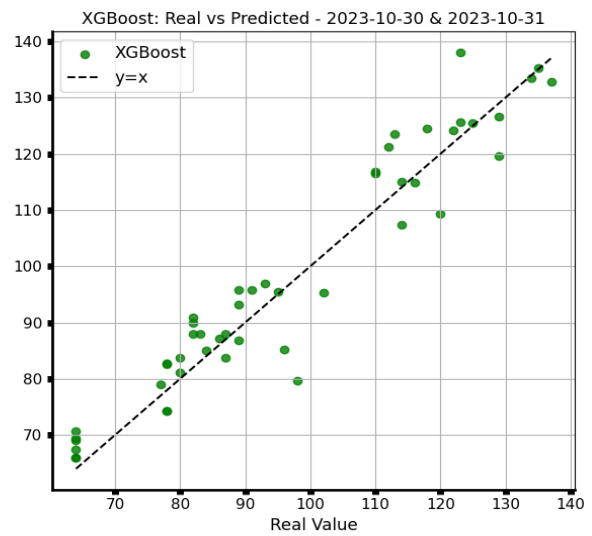
ج) روش LSTM



ب) روش Bidirectional LSTM

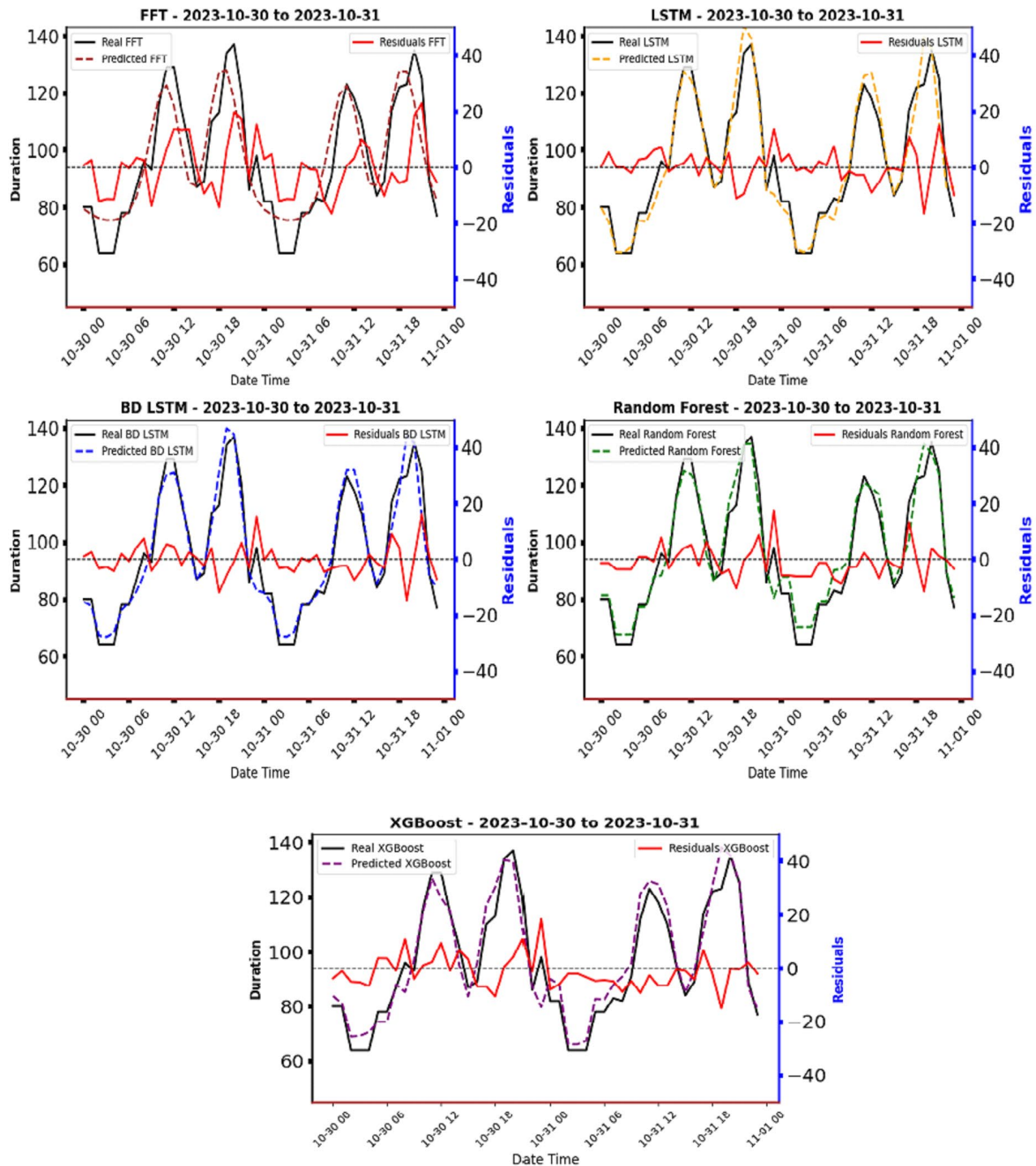


ه) روش Random Forest



د) روش XGBoost

شکل ۱۲. نمودار مقدار واقعی **Duration** در مقابل مقدار پیش‌بینی شده **Duration** توسط الگوریتم‌های مختلف



شکل ۱۳. نمودار مقدار واقعی و مقدار پیش‌بینی شده مدت زمان سفر (سمت چپ نمودار) و مقدار باقیمانده (سمت راست نمودار) برای هر مدل

ترافیک کمتر است (مانند ساعات صبح زود یا شب)، مدل‌ها عملکرد بهتری دارند و خطاها کاهش می‌یابند. برای مثال در مدل LSTM در زمان‌های افزایش ترافیک و در اوج شلوغی برای نمونه در تاریخ ۳۱-۱۰-۲۰۲۳ ساعت ۱۸، مقادیر پیش‌بینی شده دقت کمتری دارند و خطاها بیشتر می‌شوند. با این حال، در زمان‌هایی که ترافیک کاهش پیدا می‌کند (مانند ۳۰-۱۰-۲۰۲۳ ساعت ۶ صبح)، پیش‌بینی‌ها دقیق‌تر به نظر می‌رسند؛ بنابراین، افزایش ترافیک به وضوح با افزایش خطاهای

شکل ۱۳، مدل‌های مختلف LSTM، FFT، Bidirectional LSTM و Random Forest و XGBoost مقادیر واقعی و پیش‌بینی شده‌ی مدت زمان ترافیک و همچنین باقیمانده‌ها (خطاها) در دو روز (۳۰ و ۳۱ اکتبر ۲۰۲۳) نشان می‌دهند. در همه مدل‌ها دیده می‌شود که زمانی که ترافیک افزایش می‌یابد (ساعات‌های اوج‌های ترافیک)، خطاها بیشتر می‌شوند و پیش‌بینی‌ها دقت کمتری دارند. کاهش ترافیک باعث بهبود دقت می‌شود به این معنی که زمانی که

پیش‌بینی همراه است و زمانی که ترافیک کمتر است، دقت مدل‌ها بیشتر می‌شود.

## ۶- نتیجه‌گیری

یکی از کارکردهای پیش‌بینی مدت‌زمان سفر، اطلاع‌رسانی شرایط آینده ترافیک است. افراد می‌توانند با آگاهی از شرایط آینده ترافیک برنامه‌ریزی مناسب‌تری برای انتخاب انجام سفر، زمان عزیمت و انتخاب مسیر خواهد داشت. همچنین توانایی پیش‌بینی میزان حجم ترافیک جاده‌ها می‌تواند در بسیاری از برنامه‌های مدیریتی و کنترل ترافیک کمک کند. در پژوهش حاضر، به منظور پیش‌بینی حجم ترافیک ساعتی خیابان کمال شهر اصفهان از روش‌های یادگیری ماشین استفاده شده است. الگوریتم‌هایی مانند **Random Forest**، **XGBoost**، **LSTM** و **Bidirectional LSTM** و همچنین روش **FFT** بکار گرفته شد. در این پژوهش داده‌های مربوط به ترافیک شهر اصفهان (مدت‌زمان سفر) از سرویس مسیریابی نشان دانلود شده است. به منظور افزایش اطمینان از تخمین ترافیک، داده‌ها به ترتیب و متناسب با سری زمانی در سه حالت آموزشی (۸۲ درصد کل داده)، اعتبارسنجی (۱۰ درصد) و آزمایشی (۸ درصد کل داده‌ها) در مدل‌ها به کار گرفته شد.

در مدل پیشنهادی، از نظریه گراف برای ایجاد شبکه ترافیک خیابان‌های اصفهان، به ویژه خیابان کمال، استفاده شده است. در این مطالعه الگوریتم‌های یادگیری ماشین از یک سری ویژگی‌های ورودی استفاده می‌کند که می‌تواند دقت مدل را تحت تأثیر خود قرار دهد. در گام نخست پس از شناسایی مانا بودن سری زمانی بر روی داده‌ها تست ۲.۵ سیگما زده شد تا داده‌های پرت شناسایی شود و با مقادیر مشابه هفته‌های قبل جایگزین شود. عملیات پردازش داده با اضافه کردن متغیرهای پیش‌بینی‌کننده مانند ویژگی‌های زمانی (ساعت و روز هفته)، تاریخچه مدت‌زمان سفر، تعداد مدارس، تعداد بانک‌ها، هتل‌ها، شهرداری‌ها، داروخانه‌ها، اورژانس و بیمارستان‌ها، جمعیت محصل و کل جمعیت در شعاع ۱۵۰ متری به الگوریتم‌ها آغاز می‌شود. یکی دیگر از ویژگی‌های مورد استفاده این پژوهش که وابستگی مکانی بین بخش‌های جاده را مشخص می‌کند، تعیین تعداد یال‌ها) و ورودی و مجموع مدت‌زمان سفر یال‌ها) و ورودی است. این کار با جستجوی خیابان همسایه در گراف پیاده‌سازی شد. ویژگی‌هایی مانند تعطیلات و آب‌وهوا در این پژوهش

استفاده نشد، چون هیچ‌گونه تعطیل رسمی در بازه زمانی مذکور وجود نداشت و بارشی هم رخ نداده بود. در گام دوم جهت یافتن مقادیر بهینه پارامترها، مدل‌ها به‌ازای مقادیر مختلف این پارامترها آموزش داده‌شده و بر اساس دقت پیش‌بینی بر روی داده تست و استفاده از جستجوی شبکه مقادیر بهینه انتخاب شده است.

در این پژوهش پس از آموزش مدل‌ها با مقادیر بهینه، چهار مدل یادگیری ماشین (**Bidirectional Random Forest**، **LSTM**، **LSTM**، **XGBoost**) و **FFT** مقایسه شدند. در نهایت، برای بررسی تأثیر هر ویژگی بر دقت پیش‌بینی مدل‌ها، هر مدل به طور مجزا با یک ویژگی خاص و سپس با ترکیب همه ویژگی‌ها به طور هم‌زمان آموزش داده شد. سپس دقت پیش‌بینی مدل‌ها برای داده‌های تست ارزیابی شد تا مشخص شود کدام ویژگی‌ها در افزایش دقت مؤثر هستند. نتایج نشان می‌دهد، دقت اکثر مدل‌ها نزدیک به هم هستند. مدل **Random Forest** با دقت ۹۳ درصد بهترین عملکرد را داشت و پس از آن **Bidirectional LSTM** با دقت ۹۲/۴ درصد **LSTM** با ۹۲/۲ درصد و **XGBoost** با ۹۲ درصد. الگوریتم **FFT** با دقت ۷۹/۶ درصد ضعیف‌ترین نتیجه را ارائه کرد و اختلاف ۱۶/۵۸ درصدی بین **Random Forest** و **FFT** نشان از برتری مدل‌های یادگیری ماشین دارد. اضافه کردن ویژگی‌هایی مانند تعداد بانک‌ها و بیمارستان‌ها تأثیری بر افزایش دقت نداشت و استفاده از ویژگی روز هفته در همه مدل‌ها باعث کاهش دقت شد. حذف داده‌های پرت (استفاده از قاعده ۲.۵ سیگما) به بهبود دقت نیز کمک می‌کند. همچنین استفاده از ویژگی مجموع حجم ترافیک یال ورودی در فرایند به‌عنوان وابستگی مکانی در الگوریتم **LSTM** و **Bidirectional LSTM** تأثیر بهتری نسبت به حضور این ویژگی در الگوریتم‌های **XGBoost** و **Random Forest** خواهند داشت. اضافه کردن سایر ویژگی‌ها مانند تعداد بانک‌ها، بیمارستان‌ها و ... به مابقی ویژگی‌ها هیچ تأثیری در افزایش دقت نخواهد داشت. حذف داده‌های پرت (استفاده از قاعده ۲.۵ سیگما) به بهبود دقت کمک می‌کند. استفاده از ویژگی روز هفته در تمامی مدل‌ها، باعث کاهش دقت خواهد شد. تعداد یال‌های ورودی و سایر ویژگی‌ها مانند تعداد بانک‌ها، بیمارستان‌ها و ... در مدل **XGBoost** و **Random Forest** باعث منفی شدن مقدار  $R^2$  و کاهش دقت خواهد شد. به‌عنوان

استانی و یا کشوری جهت دست‌یابی به یک الگوی ملی از روش پیشنهادی این تحقیق استفاده کرد. همچنین در استفاده از مدل‌ها و الگوریتم‌های متفاوت تأثیر شرایط گوناگون محیطی از جمله مشخصات هندسی محور و شرایط روسازی جاده‌ها به‌عنوان یک ویژگی جدید در نظر گرفته شود.

پیشنهاد برای ادامه این پژوهش به‌کارگیری سایر مدل‌های یادگیری ماشین دیگر و ترکیب آنها و ارزیابی نتایج آن توصیه می‌گردد. با توجه به نتایج به‌دست‌آمده، می‌توان به‌منظور افزایش دقت از بازه زمانی طولانی‌تر استفاده کرد و به‌جای استفاده از یک خیابان، خیابان‌های دیگر را هم آزمایش کرد و در سطوح

#### ۷- پی‌نوشت‌ها

1. Spatio-Temporal Random Effects
2. Support Vector Regression
3. Adaptive Network Based Fuzzy Inference System
4. K-Nearest Neighbors
5. Ensemble Learning
6. Extreme Gradient Boosting Tree
7. Duration
8. Augmented Dickey-Fuller Test
9. Kwiatkowski-Phillips-Schmidt-Shin
10. Fast Fourier Transform
11. Discrete Fourier Transform
12. Grid Search
13. Bootstrap

#### ۸- مراجع

forecasting. *Transportation Research Part C: Emerging Technologies*, 9(5), 319-336.

-Chen, K., Zhao, S., & Zhang, D. (2019). Short-term Traffic Flow Prediction based on Data-Driven Knearest neighbour Nonparametric Regression. *Journal of Physics: Conference Series*.

-Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.

-Emami, H., & Rafati, A. (2023). Monitoring and comparing various approaches for short-term forecasting of urban traffic parameters and simulation using GIS:(Case study of the city of London). *Journal of Transportation Research*, 20(4), 443-462 (In Persian).

-Gomes, B., Coelho, J., & Aidos, H. (2023). A survey on traffic flow prediction and classification. *Intelligent Systems with Applications*, 200268.

-Habtemichael, F. G., & Cetin, M. (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns.

-Abduljabbar, R. L., Dia, H., & Tsai, P.-W. (2021). Unidirectional and bidirectional LSTM models for short-term traffic prediction. *Journal of Advanced Transportation*, 2021(1), 5589075.

-Ahmed, M. S., & Cook, A. R. (1979). Analysis of freeway traffic time-series data by using Box-Jenkins techniques.

-Asencio-Cortés, G., Florido, E., Troncoso, A., & Martínez-Álvarez, F. (2016). A novel methodology to predict urban traffic congestion with ensemble learning. *Soft Computing*, 20, 4205-4216.

-Berlotti, M., Di Grande, S., & Cavalieri, S. (2024). Proposal of a machine learning approach for traffic flow prediction. *Sensors*, 24(7), 2348.

-Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25, 197-227.

-Bracewell, R., & Kahn, P. B. (1966). The Fourier transform and its applications. *American Journal of Physics*, 34(8), 712-712.

-Chen, H., & Grant-Muller, S. (2001). Use of sequential learning for short-term traffic flow

- Sahebi, S., Meskar, M., & Bafandeh, M. (2024). Holidays Traffic Volume Prediction Using Neural Networks: a Case Study in Iran Rural Roads. *Journal of Transportation Research* (In Persian).
- Sekula, P., Marković, N., Vander Laan, Z., & Sadabadi, K. F. (2018). Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A Maryland case study. *Transportation Research Part C: Emerging Technologies*, 97, 147-158.
- Shaygan, M., Meese, C., Li, W., Zhao, X. G., & Nejad, M. (2022). Traffic prediction using artificial intelligence: review of recent advances and emerging opportunities. *Transportation Research Part C: Emerging Technologies*, 145, 103921.
- Tamir, T. S., Xiong, G., Li, Z., Tao, H., Shen, Z., Hu, B., & Menkir, H. M. (2020). Traffic congestion prediction using decision tree, logistic regression and neural networks. *Ifac-PapersOnline*, 53(5), 512-517.
- Tay, L., Lim, J. M.-Y., Liang, S.-N., Keong, C. K., & Tay, Y. H. (2023). Urban traffic volume estimation using intelligent transportation system crowdsourced data. *Engineering Applications of Artificial Intelligence*, 126, 107064.
- Tian, Y., Zhang, K., Li, J., Lin, X., & Yang, B. (2018). LSTM-based traffic flow prediction with missing data. *Neurocomputing*, 318, 297-305.
- Van Lint, J., Hoogendoorn, S. P., & van Zuylen, H. J. (2002). Freeway travel time prediction with state-space neural networks: Modeling state-space dynamics with recurrent neural networks. *Transportation Research Record*, 1811(1), 30-39.
- Wang, Y., Papageorgiou, M., & Messmer, A. (2006). RENAISSANCE—A unified macroscopic model-based approach to real-time freeway network traffic surveillance. *Transportation Research Part C: Emerging Technologies*, 14(3), 190-212.
- Wu, Y.-J., Chen, F., Lu, C.-T., & Yang, S. (2016). Urban traffic flow prediction using a spatio-temporal random effects model. *Journal of Intelligent Transportation Systems*, 20(3), 282-293.
- Xu, D. w., Wang, Y.-d., Jia, L.-m., Qin, Y., & Dong, H.-h. (2017). Real-time road traffic state prediction based on ARIMA and Kalman filter. *Frontiers of Information Technology & Electronic Engineering*, 18, 287-302.
- Transportation Research Part C: Emerging Technologies*, 66, 61-78.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Lingras, P., Sharma, S., & Zhong, M. (2002). Prediction of recreational travel using genetically designed regression and time-delay neural network models. *Transportation Research Record*, 1805(1), 16-24.
- Liu, Y., Zhu, N., Ma, S., & Jia, N. (2015). Traffic sensor location approach for flow inference. *IET Intelligent Transport Systems*, 9(2), 184-192.
- Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2014). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865-873.
- Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54, 187-197.
- Okutani, I., & Stephanedes, Y. J. (1984). Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B: Methodological*, 18(1), 1-11.
- Park, D., & Rilett, L. R. (1998). Forecasting multiple-period freeway link travel times using modular neural networks. *Transportation Research Record*, 1617(1), 163-170.
- Park, D., & Rilett, L. R. (1999). Forecasting freeway link travel times with a multilayer feedforward neural network. *Computer-Aided Civil and Infrastructure Engineering*, 14(5), 357-367.
- Park, D., Rilett, L. R., & Han, G. (1999). Spectral basis neural networks for real-time travel time forecasting. *Journal of Transportation Engineering*, 125(6), 515-523.
- Pechatnova, E., & Kuznetsov, V. (2021). Mathematical modeling of traffic volume in the suburban area based on the time series decomposition. *Journal of Physics: Conference Series*.
- Rahmaty, M., Radfar, R., Toloie Ashlaghi, A., & Pilevari Salmasi, N. (2019). Designing a Model for Prediction the Suburban Daily Traffic Volume Using Adaptive Network Based Fuzzy Inference System. *Journal of Transportation Research*, 16(1), 51-62. (In Persian)

fuzzy-neural approach. *Transportation research Part C: Emerging Technologies*, 10(2), 85-98.

-Zhan, X., Zheng, Y., Yi, X., & Ukkusuri, S. V. (2016). Citywide traffic volume estimation using trajectory data. *IEEE Transactions on Knowledge and Data Engineering*, 29(2), 272-285.

-Zhang, L., Bian, W., Qu, W., Tuo, L., & Wang, Y. (2021). Time series forecast of sales volume based on XGBoost. *Journal of Physics: Conference Series*.

-Yeon, K., Min, K., Shin, J., Sunwoo, M., & Han, M. (2019). Ego-vehicle speed prediction using a long short-term memory based recurrent neural network. *International Journal of Automotive Technology*, 20, 713-722.

-Yi, Z., Liu, X. C., Markovic, N., & Phillips, J. (2021). Inferencing hourly traffic volume using data-driven machine learning and graph theory. *Computers, Environment and Urban Systems*, 85, 101548.

-Yin, H., Wong, S. C., Xu, J., & Wong, C. (2002). Urban traffic flow prediction using a

# Traffic Volume Prediction Using Machine Learning Methods

## (Case Study: Kamal Street: Isfahan)

*Mahsa Latifi, M.Sc., Student, Department of Geomatics Engineering,  
Faculty of Civil and Transportation Engineering, University of Isfahan, Isfahan, Iran.  
Jamshid Maleki, Assistant Professor, Department of Geomatics Engineering,  
Faculty of Civil and Transportation Engineering, University of Isfahan, Isfahan, Iran.*

*E-mail: j.maleki@cet.ui.ac.ir*

Received: April 2025- Accepted: November 2025

### ABSTRACT

The increasing demand for the use of private vehicles has turned traffic congestion into one of the most critical crises in major cities worldwide. The ability to predict road traffic volume can aid significantly in many traffic management and control programs. However, estimating traffic volume is challenging, as vehicle counting is typically only possible at a limited number of locations equipped with fixed traffic sensors. In this study, routing data from the "Neshan" service has been used to address this challenge. By recording travel time data at different hours of the day, it is possible to estimate the traffic volume of urban road segments. To assess the feasibility of the proposed method, travel time data from Kamal Street, located in Isfahan, was collected for various hours of the day over a period of 23 days. Additionally, machine learning methods, including Random Forest, Extreme Gradient Boosting, Long Short-Term Memory (LSTM) neural networks, and Bidirectional LSTM networks, were employed to predict hourly traffic volume. The Fast Fourier Transform (FFT) was also used to identify the primary frequencies of traffic volume fluctuations and to model and compare with other methods. In the methods used in this study, features such as spatial dependency (total duration of the incoming edges to the street), duration history, temporal features like the hour of the day and the day of the week, the number of incoming edges, and the number of traffic-influencing centers such as schools, banks, hotels, municipal buildings, pharmacies, emergency services, and hospitals, as well as the student population and total population within the street area, were utilized. The numerical results indicate that machine learning and deep learning methods perform better than the fast fourier transform. Among the machine learning and deep learning methods, the random forest method, with a high  $R^2$  value of 0.93, shows better performance compared to other methods.

**Keywords:** Hourly Traffic Volume Prediction, Spatial-Temporal Modeling, Machine Learning, Fast Fourier Transform