

پیش‌بینی زمان سفر با یادگیری ماشین: رقابت رگرسیون خطی، رگرسیون چندمتغیره، جنگل تصادفی و شبکه عصبی عمیق

زهرا رضائی^۱، حسین آقامحمدی^{۲*}، محمدحسن وحیدنیا^۳، زهرا عزیزی^۲، سعید بهزادی^۴

^۱ دانشجوی دکتری سنجش از دور و سیستم اطلاعات جغرافیایی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران
Rezaee.ncc@gmail.com

^۲ استادیار گروه سنجش از دور و سیستم اطلاعات جغرافیایی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران
hossein.ghamhammadi@gmail.com
zsazizi@yahoo.com

^۳ استادیار مرکز مطالعات سنجش از دور و GIS، دانشکده علوم زمین دانشگاه شهید بهشتی، تهران
mh_vahidnia@sbu.ac.ir

^۴ دانشیار گروه مهندسی نقشه برداری، دانشکده مهندسی عمران، آب و محیط زیست، دانشگاه شهید بهشتی، تهران
s_behzadi@sbu.ac.ir

(دریافت: تیر ۱۴۰۳، تصویب: آبان ۱۴۰۳)

چکیده

پیش‌بینی دقیق زمان سفر یکی از مسائل مهم در حوزه ترافیک و حمل‌ونقل است که می‌تواند به طور قابل توجهی بر زندگی روزمره افراد و سازمان‌ها تأثیر بگذارد. در این پژوهش، چهار روش مختلف یادگیری ماشین شامل رگرسیون خطی (LR)، رگرسیون چندمتغیره (MR)، رگرسیون جنگل تصادفی (RDR) و شبکه عصبی مصنوعی عمیق (DNN) برای پیش‌بینی زمان سفر آموزش داده شدند. هدف از این پژوهش پیش‌بینی زمان سفر جهت استفاده در سیستم‌های ترافیک هوشمند است و بهره‌گیری و مقایسه چندین روش جدید شامل شبکه عصبی عمیق و رگرسیون جنگل تصادفی و همچنین دخیل نمودن پارامترهای جدید مانند وضعیت بارش، نرم جریان ترافیک، زمان و همچنین تصادفات و نقاط قفل ترافیکی نسبت به پژوهش‌های دیگر، نوآوری و جامعیت این پژوهش نسبت به سایر مطالعات می‌باشد. در طراحی و اجرای این پژوهش از داده‌های واقعی ترافیک برگرفته از Google map استفاده و آنالیز گردید. این داده‌ها شامل اطلاعاتی از جمله شرایط ترافیک، فصل سال، ساعت روز، وضعیت بارش جوی و ویژگی‌های مسیر می‌باشد. نتایج این پژوهش نشان می‌دهد که مدل DNN با R^2 برابر با ۰.۸۳۳ عملکرد خیلی خوبی را در بین مدل‌های مورد بررسی دارد. این مدل ۰.۸۳۳٪ واریانس داده‌ها را توضیح می‌دهد و توزیع باقیمانده‌ها در آن به طور نسبی مرکزی با میانگین صفر و توزیعی نزدیک به نرمال است. مدل رگرسیون خطی با R^2 برابر با ۰.۶۱۵ عملکرد ضعیف‌تری نسبت به DNN دارد و ۰.۶۱۵٪ از واریانس داده‌ها را توضیح می‌دهد. و اما مدل رگرسیون جنگل تصادفی با R^2 برابر با ۰.۹۵۵ در رقابت با DNN یکی از بهترین عملکردها را دارد و ۰.۹۵۵٪ از واریانس داده‌ها را توضیح می‌دهد. پارامترهای MSE و RMSE نیز جهت ارزیابی عملکرد مدل‌ها استفاده شدند و نهایتاً مقایسه چندبعدی بین مدل‌ها صورت گرفت و مدل جنگل تصادفی کمترین مقادیر خطا را نتیجه داد. از آنجائیکه در داده‌های ترافیکی جمع‌آوری شده، حوادث رانندگی و تبعاً نقاط قفل ترافیکی در مدل‌ها نیز استفاده شده، و با لحاظ اینکه مدل رگرسیون جنگل تصادفی با وجود نویز و آنومالی نیز به طور موثرتری با داده‌ها تطبیق یافته می‌یابند، مقدار R^2 این مدل، از شبکه‌های عصبی عمیق بدلیل داشتن ذات بیش‌پردازشی، بالاتر حاصل شده‌است.

واژگان کلیدی: زمان سفر، رگرسیون خطی، رگرسیون چند متغیره، رگرسیون جنگل تصادفی، یادگیری عمیق، شبکه عصبی عمیق

۱- مقدمه

در حال حاضر، پیش‌بینی زمان سفر یکی از مسائل اساسی و حیاتی در حوزه ترافیک و حمل‌ونقل است که تاثیر گسترده‌ای بر زندگی روزمره ما دارد. با توجه به اینکه زمان سفر برای افراد و سازمان‌ها اهمیت بسیاری دارد، توانایی دقیق در پیش‌بینی زمان سفر باعث می‌شود تا منابع و زمان به شکل بهینه‌تری مدیریت شود و از تاخیرهای ناخواسته جلوگیری شود. بنابراین، ارائه یک مدل قوی و دقیق برای پیش‌بینی زمان سفر اهمیت بسیاری دارد زیرا این مسئله در حوزه‌های مختلف از جمله حمل‌ونقل عمومی، حمل‌ونقل شخصی، مدیریت ترافیک، برنامه‌ریزی شهری و حتی تجارت و اقتصاد تأثیرگذار است [۱]. به علاوه، در حوزه مدیریت ترافیک و برنامه‌ریزی شهری، داشتن یک مدل قدرتمند برای پیش‌بینی زمان سفر به مسئولان این امکان را می‌دهد که سیستم‌های ترافیک و حمل‌ونقل عمومی را بهبود بخشند و بهینه‌سازی منابع را انجام دهند. در نتیجه، این مسئله از اهمیت بالایی برخوردار است و ارائه یک مدل قوی و دقیق برای پیش‌بینی زمان سفر مورد نظر است. پژوهشگران، تحقیقات جامع و دقیق در زمینه پیش‌بینی زمان سفر با استفاده از روش‌های مختلف الگوریتم‌های یادگیری ماشین و هوش مصنوعی برای بهبود کارایی و دقت در پیش‌بینی زمان سفر انجام داده‌اند [۲]. برای پیش‌بینی زمان سفر، داده‌های مرتبط از منابع مختلف که شامل شرایط ترافیک، فصل سال، ساعت روز، وضعیت بارش جوی و ویژگی‌های مسیر بوده، جمع‌آوری و تحلیل می‌شوند. این داده‌ها به شکل زمانی و با ویژگی‌ها و متغیرهای وابسته مناسب برای مدل‌های پیش‌بینی تبدیل شده و با استفاده از چهار روش مختلف اعم از رگرسیون خطی LR^۱، رگرسیون چندمتغیره MR^۲، رگرسیون جنگل تصادفی RFR^۳ و شبکه‌های عصبی عمیق DNN^۴ آموزش داده می‌شوند. سپس ارزیابی عملکرد مدل‌ها بر روی داده‌های آزمون و تست انجام شده تا روش مناسب‌تر برای پیش‌بینی زمان سفر مشخص گردد.

استفاده از چندین روش مختلف برای پیش‌بینی زمان سفر اهمیت زیادی دارد [۳]. تحقیقات گذشته نشان

داده‌است که استفاده از رگرسیون خطی در پیش‌بینی زمان سفر در شرایط ساده می‌تواند کارا باشد، اما در شرایط پیچیده‌تر و با داده‌های غیرخطی، عملکرد آن محدود می‌شود [۴]. اما RFR به عنوان یک الگوریتم غیرخطی و مناسب برای پیش‌بینی زمان سفر در شرایط پیچیده‌تر شناخته شده‌است [۵]. همچنین، DNNs نیز نشان داده‌اند که قابلیت خوبی در پیش‌بینی سری‌های زمانی دارند، به-خصوص در صورت دسترسی به داده‌های پیچیده و با حجم بالا، این روش به عنوان یک روش پردازش داده‌های پیچیده و غیرخطی در سال‌های اخیر به شدت مورد توجه قرار گرفته‌است [۶]. با پیشرفت تکنولوژی، به خصوص در زمینه محاسبات ابری و پردازنده‌های گرافیکی، محققان به سمت استفاده از مدل‌های پیچیده‌تر و عمیق‌تر مانند شبکه‌های عصبی مصنوعی ANNs^۵ روی آوردند. در اوایل دهه ۲۰۱۰، محققان شروع به استفاده از ANNs برای پیش‌بینی ترافیک کردند [۷]. در ادامه، مروری بر تحقیقات مرتبط آورده می‌شود.

۲- پیشینه تحقیق

تحقیقات جدید در حوزه هوش مصنوعی و یادگیری ماشین با هدف بهبود دقت و کارایی پیش‌بینی زمان سفر، قدمی مثبت در بهبود حمل‌ونقل عمومی و شخصی می‌باشد. در تحقیقی، ایزدی و همکاران (۱۴۰۱)، به پیش‌بینی متغیرهای ترافیکی و اطلاع‌رسانی به مسافران و اپراتورهای شبکه حمل‌ونقل به یک روش مدیریت تقاضا برای سفر پرداختند و سه الگوریتم یادگیری ماشین، شامل ماشین بردار پشتیبان SVM^۶، جنگل تصادفی و شبکه عصبی با حافظه کوتاه‌مدت LSTM^۷، را برای پیش‌بینی وضعیت ساعتی ترافیک بکار بردند و نشان دادند که LSTM کوتاه‌مدت دقت بالاتری در پیش‌بینی وضعیت ترافیک دارد [۸]. در تحقیق دیگری، Yifan و همکاران (۲۰۲۲)، رویکرد هوشمندی جهت مدل‌سازی و کاهش انتشار گازهای گلخانه‌ای در معابر را توسعه دادند. آنها از داده‌های نظارت‌شده ترافیک با تراکم بالا و داده‌های کاربری زمین برای آموزش مدل‌های جنگل تصادفی که قادر به پیش‌بینی دقیق حرکت خودروهای پویا هستند،

۵ Artificial Neural Network

۶ Support Vector Machine

۷ Long Short-Term Memory

۱ Linear Regression

۲ Multivariate Regression

۳ Random Forest Regression

۴ Deep Neural Network

استفاده کردند. نویسندگان از پارامترهای ویژگی‌های جاده، تراکم جمعیت، و اطلاعات کاربری زمین، در آموزش مدل استفاده نمودند و نتیجتاً یک مدل توسعه‌یافته را برای شناسایی عوامل ناهمگونی فضایی در الگوهای حمل‌ونقل و مصرف انرژی اتخاذ و روابط غیرخطی بین متغیرهای حرکت، جمعیت‌شناسی و کاربری زمین را ارائه کردند [۹]. در تحقیق دیگری Alexander و همکاران (۲۰۲۰)، از شبکه‌های عصبی عمیق برای پیش‌بینی ترافیک استفاده کردند. نویسندگان مدل‌های آماری، مدل‌های یادگیری ماشین و شبکه‌های عصبی را برای پیش‌بینی جریان ترافیک مورد ارزیابی قرار دادند. آن‌ها ساختارها، عملکرد و توانایی شبکه‌های عصبی عمیق را بررسی و مقایسه کرده و چالش‌های آینده در زمینه پیش‌بینی جریان ترافیک را شناسایی کردند. همچنین، مولفین بررسی کردند که در نظر گرفتن پارامترهای موثر بر ترافیک به طور کامل انجام نشده‌است و لازم است که پارامترهای موثر بر ترافیک به صورت جامع‌تری در مدل‌ها در نظر گرفته شود تا دقت مدل‌های خروجی افزایش یابد [۱۰]. Peng و همکاران (۲۰۲۱) در مقاله خود، رویکردی نوآورانه برای پیش‌بینی طولانی‌مدت جریان ترافیکی پیشنهاد کرده‌اند که با تمرکز بر تحلیل داده‌های گرافی به‌دست‌آمده از الگوهای ترافیک و با استفاده از شبکه‌های پیچشی گراف دینامیک (DGCN) و یادگیری تقویتی، قابلیت تطبیق با تغییرات دینامیکی در دنیای واقعی را دارا می‌باشد [۱۱]. Sayed و همکاران (۲۰۲۳)، یک مرور جامع بر روش‌ها و تکنیک‌های مختلف هوش مصنوعی به‌کاررفته در پیش‌بینی جریان ترافیک ارائه دادند و به تحلیل سیستماتیک پیشرفت‌ها و چالش‌های موجود در این زمینه پرداختند. آنها همچنین به بررسی انواع الگوریتم‌های یادگیری ماشین و یادگیری عمیق که در پیش‌بینی رفتار ترافیک و جریان وسایل نقلیه استفاده می‌شوند، پرداخته و مزایا و معایب هر کدام را برشمردند [۱۲]. Yassin و (2020) Pooja به بررسی پیش‌بینی تصادفات جاده‌ای و تفسیر مدل با استفاده از ترکیب الگوریتم K-means و الگوریتم جنگل تصادفی پرداختند. این تحقیق، به تحلیل تعاملات میان ویژگی‌های مختلف جاده‌ای و محیطی که بر روی وقوع تصادفات تأثیر می‌گذارند، پرداخته و روش‌های نوینی برای پیش‌بینی و تفسیر این حوادث ارائه می‌دهد. نتایج به‌دست‌آمده در این مطالعه نشان می‌دهد که استفاده از این الگوریتم‌های ترکیبی می‌تواند به بهبود دقت پیش‌بینی

تصادفات جاده‌ای کمک کند [۱۳]. در تحقیق دیگری، Abdullah و همکاران (۲۰۲۳) به بررسی رویکردهای مبتنی بر یادگیری عمیق برای پیش‌بینی و مدیریت ترافیک در محیط‌های شهری و شناسایی چالش‌ها و فرصت‌های موجود در بهینه‌سازی جریان ترافیک در شهرهای هوشمند پرداختند. مدل‌های توسعه‌یافته در این تحقیق، قابلیت شناسایی و پیش‌بینی وضعیت شلوغی ترافیک در زمان‌های مختلف روز و در شرایط متفاوت جوی را دارند [۱۴].

با بررسی مطالعات صورت گرفته، نیاز به استفاده از روش‌های متنوع و نوآورانه برای شبیه‌سازی زمان‌سفر احساس می‌شود. بطور کلی، روش‌های سنتی مانند رگرسیون خطی و رگرسیون چندمتغیره ممکن است در شناسایی الگوهای پیچیده ترافیک و عوامل مؤثر بر آن محدودیت‌هایی داشته باشند. از سوی دیگر، الگوریتم‌هایی مانند رگرسیون جنگل تصادفی و شبکه‌های عصبی عمیق توانایی بیشتری در مدل‌سازی روابط غیرخطی و تعاملات پیچیده بین متغیرها دارند. تحقیقات اخیر همچنین نشان داده‌است که در نظر گرفتن توپولوژی شبکه در مدل‌های پیش‌بینی زمان‌سفر، بهبود قابل توجهی در دقت و قابلیت اطمینان، به ویژه برای سفرهای طولانی‌تر، ایجاد می‌کند. اما، مشکلاتی که ناشی از نقص پارامترها در شبیه‌سازی‌های ترافیک، و در نظر نگرفتن عوامل خارجی نظیر آب و هوا، رویدادهای خاص و حوادث بوده، باعث کاهش دقت پیش‌بینی زمان‌سفر می‌شود. از این رو، لحاظ کردن این عوامل می‌تواند دقت پیش‌بینی زمان‌سفر را تقویت کند. در مطالعات جدیدتر مانند پژوهش‌های [۱۵ و ۱۶] نتایج امیدوارکننده‌ای با در نظر گرفتن مجموعه‌ی محدودی از عوامل خارجی مانند آب‌وهوا و رویدادهای برنامه‌ریزی‌شده بوجود آمد. از این رو، در نظر نگرفتن کامل عوامل خارجی موثر بر ترافیک در مدل‌های موجود، برای نمونه مواردی مانند تصادفات یا انسدادهای غیرمنتظره جاده، منجر به اختلال قابل توجه شبیه‌سازی جریان ترافیک می‌گردد. این کاستی‌ها؛ زمینه‌ساز مهمی برای تحقیقات آتی را بوجود آورد. در این پژوهش، از داده‌های زمان‌بندی شده Google Map به‌همراه سایر اطلاعات فیزیکی محیطی در یک منطقه مورد مطالعه در تهران استفاده شده‌است. از API Google Maps برای جمع‌آوری داده‌های زمان‌بندی شده ترافیک در منطقه مورد مطالعه در تهران استفاده شد. در کنار آن، پارامترهای متنوع و متعددی که در مدل‌سازی زمان‌سفر

لحاظ می‌شود، همچون حوادث رانندگی و تبعاً نقاط قفل ترافیکی نیز در محاسبات استفاده گردید.

متعاقباً مدل‌هایی که تشریح خواهند شد، با در نظر گرفتن این آنامولی‌ها و پراکندگی دامنه مشاهدات، وارد فرایند شبیه‌سازی شده‌اند. از این‌رو، نوآوری و تفاوت این پژوهش در مقایسه با سایر مطالعات صورت گرفته در این بخش، مشهود است.

۳- مواد و روش‌ها

مراحل انجام کار در این مطالعه، در سه بخش ارائه گردیده‌است. در بخش اول، مفاهیم روش‌های یادگیری ماشین ارائه شده‌است. در قسمت دوم داده‌های مورد نیاز جهت اجرای فرایند مدل‌سازی در این مطالعه، تشریح گردیده و در بخش سوم، پیاده‌سازی الگوریتم‌های مختلف یادگیری ماشین در شبیه‌سازی و پیش‌بینی زمان سفر و نتایج آنها آورده شده‌است.

۳-۱- انواع یادگیری ماشین؛ رگرسیون‌های خطی، چندمتغیره، رگرسیون جنگل تصادفی و شبکه عصبی عمیق

در یادگیری ماشین، رویکردهای مدل‌سازی با یادگیری نظارت‌شده و بدون نظارت دو رویکرد اصلی هستند. در یادگیری با نظارت، داده‌های برچسب‌گذاری شده به مدل داده شده و هدف آن، ایجاد یک تابع استنتاجی است که بردارهای ویژگی را به برچسب‌های خروجی نگاشت کند. الگوریتم‌های محبوب یادگیری با نظارت شامل ماشین‌بردار پشتیبان، K-نزدیک‌ترین همسایه KNN^۱، رگرسیون خطی، رگرسیون جنگل تصادفی و غیره هستند که برای پیش‌بینی‌های دقیق در کلاسه‌بندی و رگرسیون استفاده می‌شوند [۱۷، ۱۸]. همانطور که گفته شد در این تحقیق، چهار روش مختلف برای پیش‌بینی زمان سفر مورد بررسی قرار خواهند گرفت: رگرسیون خطی، رگرسیون چند متغیره، رگرسیون جنگل تصادفی و شبکه عصبی مصنوعی عمیق. از روش رگرسیون خطی به عنوان یک روش ساده و تفسیرپذیر معمولاً در مسائل پیش‌بینی استفاده می‌شود. این روش با استفاده از یک تابع خطی، مطابق رابطه (۱)

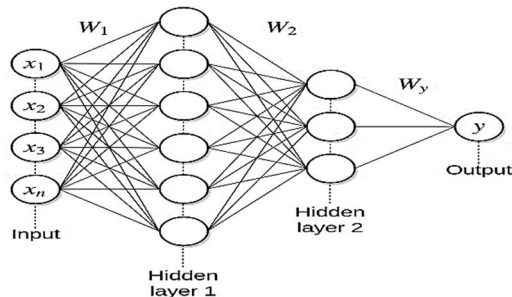
ارتباط بین ورودی‌ها و خروجی را تخمین می‌زند. این روش، به خاطر سادگی و قابل فهم بودن نتایج، برای تحلیل و تفسیر داده‌ها بسیار مناسب است. با این حال، رگرسیون خطی تنها قادر به مدل کردن روابط خطی بین ورودی‌ها و خروجی است و در مواردی که روابط غیرخطی و پیچیده‌تر وجود دارد، عملکرد ضعیف‌تری دارد [۱۹، ۲۰].

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (1)$$

بطور کلی مسأله مهمی که در داده‌های مکانی وجود دارد این است که الگوهای رگرسیون معمولی برای دو متغیر مستقل و وابسته مکانی، صرفاً نتیجه متوسطی از داده‌های غیرمکانی را ارائه می‌دهند. بنابراین قادر نخواهند بود خودهمبستگی‌های مکانی بین متغیرها را بیان کنند. آن‌ها اغلب در نشان دادن واقعیات ارتباطات مختلف مکانی ناتوان هستند و فقط برازشی از یک خط رگرسیونی کلی برای کل داده‌ها ارائه می‌دهند. از این‌رو، برخی از واقعیات محلی در این زمینه پنهان می‌ماند و در نهایت نتیجه معقول و منطقی برای داده‌های مکانی ارائه نمی‌دهند.

رگرسیون جنگل تصادفی، یک تکنیک یادگیری ماشین است که با ترکیب مجموعه‌ای از درختان رگرسیون مطابق شکل (۱) بر اساس یادگیری گروهی عمل می‌کند [۲۱]. این مدل بر اساس الگوریتم "تصادفی انجمن درخت‌ها" ساخته می‌شود که به صورت یک انجمن از درخت‌های تصمیم^۲ عمل می‌کند. این الگوریتم بر روی یک مجموعه داده بزرگ با نمونه‌های متعددی که شامل ویژگی‌های مختلف برای طبقه‌بندی یا رگرسیون هستند، کار می‌کند. در این روش، هر درخت به صورت مستقل از داده‌های آموزش، پیش‌بینی انجام می‌دهد و نتایج را ترکیب می‌کند. این الگوریتم، مجموعه‌ای از درخت‌های تصمیم مستقل را ایجاد می‌کند که هر کدام داده‌ها را به طور تصادفی نمونه‌برداری کرده و به صورت سلسله‌وار تقسیم می‌کنند تا به یک نتیجه برسند. با آموزش درخت‌ها بر روی زیرمجموعه‌ای از داده‌ها، تنوع در مدل‌ها افزایش یافته و از بیش‌برازش جلوگیری می‌شود. در انتهای هر درخت، یک نتیجه شامل طبقه یا مقدار پیش‌بینی شده وجود دارد که برای رسیدن به یک نتیجه نهایی، نتایج تمام درخت‌ها با هم

عصبی از یک تابع فعال‌سازی غیرخطی مانند ReLU، sigmoid یا tanh استفاده می‌کند تا روابط پیچیده‌تر را مدل کند. این شبکه‌ها بدلیل وجود لایه‌های عمیق، می‌توانند ویژگی‌های پنهان در داده‌ها را استخراج و تجزیه و تحلیل کنند [۲۷].

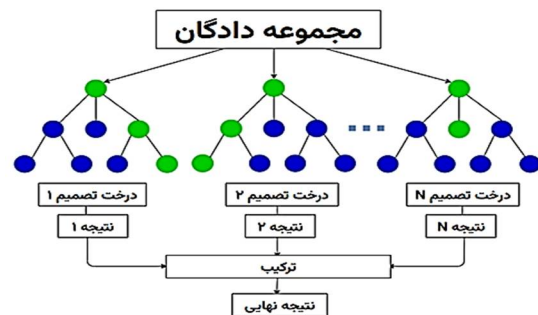


شکل (۲) معماری شبکه عصبی عمیق [۲۴]

شبکه‌های عصبی عمیق مطابق شکل (۲)، شبکه‌های پیچیده‌تر با تعداد زیادی لایه هستند که از مدل‌های شبکه عصبی تشکیل شده‌اند. شبکه‌های عصبی، شامل واحدهای محاسباتی به نام نورون‌ها^۱ هستند که به صورت دایره‌هایی در تصاویر نمایان می‌شوند. این نورون‌ها در لایه‌هایی سازماندهی شده‌اند: لایه ورودی داده‌های خام را دریافت می‌کند، لایه‌های پنهان پردازش‌های پیچیده‌تری انجام می‌دهند و لایه خروجی نتیجه نهایی را ارائه می‌دهد. ارتباطات بین نورون‌ها توسط وزن‌ها مشخص می‌شود که میزان اهمیت هر اتصال را تعیین کرده و در طول آموزش بهینه می‌شوند. هر نورون دارای یک تابع فعال‌سازی است که خروجی آن را به مقدار مشخصی تبدیل می‌کند. شبکه‌های عصبی که دارای چندین لایه پنهان می‌باشند، به عنوان "عمیق" شناخته می‌شوند زیرا توانایی یادگیری ویژگی‌های پیچیده و انتزاعی از داده‌ها را دارند که شبکه‌های با لایه‌های کمتر از آن، ناتوانند. انواع این شبکه‌ها شامل شبکه‌های بازگشتی RNN^۲، شبکه‌های عصبی کانولوشنال CNN^۳، و شبکه‌های پیشخور FFNN^۴ و ترکیبی از آن‌ها می‌باشند [۲۸، ۲۹]. این شبکه‌ها قادرند به صورت واضح انواع مختلف داده‌های ترافیکی را ضبط کنند و با افزایش تعداد لایه‌ها، قابلیت پیش‌بینی مدل‌ها بهبود یافته و دقت نوسانات ترافیک را بهبود می‌بخشند. در روش‌های دیگری مثل شبکه‌های عصبی گرافی GNN^۵ نیز، تبدیل

ترکیب می‌شوند. در رگرسیون جنگل تصادفی، تابع رگرسیون نهایی میانگین پیش‌بینی‌های تمام درختان تصمیم در جنگل است و هنگامی که تعداد زیادی از این درختان به صورت تصادفی ایجاد شده و نتایج آن‌ها ترکیب شود، یک مدل قدرتمندتر و دقیق‌تر به دست می‌آید [۱۸، ۱۹]. براین اساس، برای یک نقطه داده جدید x ، پیش‌بینی نهایی توسط رابطه (۲) محاسبه می‌شود، که در آن N تعداد درختان در جنگل و $f_i(x)$ پیش‌بینی درخت- i ام برای x است [۲۴].

$$F(x) = 1/N * \sum f_i(x) \quad (2)$$



شکل (۱) معماری رگرسیون جنگل تصادفی

رگرسیون جنگل تصادفی به عنوان یک روش غیرخطی و قوی در پیش‌بینی زمان‌سفر استفاده می‌شود. در مجموع، رگرسیون جنگل تصادفی بر اساس اصول ریاضیاتی قوی مانند آنتروپی، معیار Gini و نمونه‌گیری با جایگزینی عمل می‌کند. این الگوریتم به دلیل دقت و کارایی بالا در طیف گسترده‌ای از وظایف رگرسیون به طور گسترده‌ای مورد استفاده قرار می‌گیرد [۲۵] از ویژگی‌های شاخص این الگوریتم، می‌توان به مقاومت در برابر معضلاتی همچون بیش‌برازشی، قابلیت پردازش داده‌های کم‌کیفیت و وجود نویز در داده‌های آموزش، اشاره کرد.

شبکه‌های عصبی عمیق به عنوان روشی برای پردازش داده‌های پیچیده و غیرخطی به شدت موردتوجه قرار گرفته‌اند [۲۶]. این شبکه‌ها، با الهام از ساختار عصبی مغز و با استفاده از شبکه‌های نورون و لایه‌ها، قابلیت مدل‌سازی روابط پیچیده و غیرخطی را دارند. هر نورون در شبکه

^۴ Feedforward Neural Network

^۵ Graph Neural Network

^۱ Neuron

^۲ Recurrent Neural Network

^۳ Convolutional Neural Network

اطلاعات شبکه جاده‌ای به فرمت گرافی قابل انجام است، که در آن گره‌ها نمایانگر نقاط کلیدی مانند تقاطع‌ها و یال‌ها نمایانگر مسیرهای جاده‌ای هستند. در مدل‌های GNN، ویژگی‌های مرتبط با هر گره و یال استخراج شده و با تراکنش‌های پیام‌ها میان گره‌ها، اطلاعات گره‌های همسایه جمع‌آوری و به‌روزرسانی می‌شوند. پس از پردازش این اطلاعات، یک مدل یادگیری عمیق می‌تواند زمان سفر را تخمین بزند [۳۰، ۳۱، ۳۲]. اما در مجموع، این مدل‌ها نسبت به روش‌های قدیمی‌تر نیز معایبی دارند.

با توجه به این تفاسیر، استفاده از چندین روش مختلف برای پیش‌بینی زمان سفر مزایای خود را دارد. هر چند روش رگرسیون خطی و چند متغیره، رگرسیون جنگل تصادفی و شبکه عصبی عمیق مزایا و محدودیت‌های خود را دارند، اما با توجه به نوع داده و ساختار مسئله، ممکن است یک روش نسبت به دیگران مناسب‌تر باشد. در این تحقیق، به منظور مدل‌سازی زمان سفر و پیش‌بینی جریان ترافیک، در گام اول مجموعه داده‌های مقایسه‌ای جامع که امکان بررسی متعدد عوامل ترافیکی را فراهم آورد؛ نه تنها از جنبه‌های مکانی-زمانی، بلکه با گنجاندن داده‌های رسانه‌های اجتماعی، داده‌های آب و هوا، داده‌های حوادث و دیگر داده‌های خارجی که ممکن است بر پیش‌بینی ترافیک تأثیر داشته باشند، ایجاد گردید. سپس با مدل‌سازی‌های مبتنی بر فناوری‌های یادگیری ماشین، پیش‌بینی جریان ترافیک انجام گردید. نتیجه مورد انتظار این پژوهش، پیاده‌سازی گسترده سیستم‌های پیش‌بینی به صورت زمان‌واقعی، با هدف کمک به بهبود وضعیت ترافیک در شبکه معابر می‌باشد.

۳-۲- جمع‌آوری داده‌های ترافیکی

یکی از رایج‌ترین روش‌های دسترسی به داده‌های ترافیکی، داده‌های ترافیکی مراجع تحت وب همچون Google Map و سرویس‌های مسیریابی همچون Waze و غیره می‌باشد. Google Maps مجموعه کاملی از API ها را ارائه می‌دهند که امکان دسترسی به داده‌های ترافیک، تاریخچه این داده‌ها و تحلیل الگوهای ترافیک را فراهم می‌کند [۳۳]. Google Maps API دسترسی به داده‌های ترافیک فراهم می‌کند، اما بروزرسانی اطلاعات به صورت آنی

انجام نمی‌شود. Waze داده‌های ترافیکی آنی^۱ را به کمک کاربران خود از جمع‌آوری می‌کند، در حالی که OSM و Traffic API نیز منابع دیگری برای دسترسی به این داده‌ها هستند. مدل‌های شبیه‌سازی ترافیک می‌توانند از داده‌های تاریخچه ترافیک و داده‌های مربوط به رویدادهای آینده مانند تعطیلات یا رویدادهای ورزشی برای پیش‌بینی حجم ترافیک در آینده استفاده کنند. در این تحقیق جهت جمع‌آوری داده‌های زمان سفر، از Google Maps استفاده شد. این داده‌ها شامل زمان‌های رسیدن به مقصد در ساعات متوالی شبانه روز و در تاریخ‌های مختلف می‌باشند. این اطلاعات تحت پوشش برخی ساعات از شبانه روز، نشان‌دهنده میزان ترافیک و تاخیر در مسیرها هستند. از این داده‌ها برای تحلیل و بررسی زمان‌های مسافرت و تفاوت‌های آنها در طول روز استفاده شده‌است تا الگوهای ترافیکی مختلف شناسایی شود.

۳-۳- مقایسه و تفاوت روش‌های پیش‌بینی زمان سفر و حجم ترافیک با شبکه عصبی و روش‌های Google Maps

در پیش‌بینی زمان سفر و حجم ترافیک، Google Maps از ترکیب روش‌های مبتنی بر مدل و داده استفاده می‌کند، اما تحقیقات نشان می‌دهد که شبکه‌های عصبی می‌توانند روش‌های پیشرفته‌تری برای این موضوع ارائه کنند. این دو سامانه، یعنی Google Maps و Waze، از پیش‌بینی ترافیک برای بهبود مسیریابی استفاده می‌کنند [۳۳]. Google Maps در محاسبات خود از داده‌های GPS و حسگرهای ترافیک آنلاین بهره می‌برد، و از مدل‌های ترافیکی و یادگیری ماشین موجود استفاده می‌کند. در حالی که تحقیقات مبتنی بر شبکه‌های عصبی به تحلیل انواع داده‌ها، از جمله داده‌های تاریخی ترافیک، وضعیت آب و هوا و رویدادهای خاص می‌پردازند [۳۴].

در این تحقیق، هدف بررسی کارایی و دقت روش‌های مختلف پیش‌بینی ترافیک و شناسایی نقاط قوت و ضعف هر کدام است. روش‌های مختلف معمولاً بر اساس نوع داده‌های استفاده شده، پارامترهای ورودی و شرایط محیطی عملکرد متفاوتی دارند. بررسی این تفاوت‌ها می‌تواند به بهبود مدل‌های مختلف و درک بهتری از عملکرد الگوریتم‌ها و پیشرفت‌های

^۱ Real-time

احتمالی در این زمینه کمک کند. با توجه به بررسی زمان و نقاط حوادث، رویدادها، و ساعات مختلف جریان ترافیک، این تحقیق بر آن است تا مدل‌های عصبی جدیدی را توسعه دهد که قادر به پیش‌بینی زمان سفر و حجم ترافیک با دقت بیشتری نسبت به روش‌های سنتی باشند. نقطه قوت و نوآوری این روش، قابلیت تعمیم آن در شرایط مختلف ترافیکی و توانایی پیش‌بینی دقیق‌تر زمان سفر و حجم ترافیک است. به هرحال، پیش از استفاده گسترده از این رویکرد، لازم است چالش‌های مربوط به داده، تفسیر و محاسبات حل و فصل

شوند. در جدول (۱) خلاصه‌ای از ویژگی‌ها و توانمندی‌های مختلف این رویکردها گردآوری شده‌است. اگر نیاز به پیش‌بینی‌های دقیق و به‌روز باشد، استفاده از Google Maps ممکن است انتخاب بهتری باشد. اما در صورتی که هدف درک عمیق‌تر از ترافیک یا توسعه روش‌های جدید پیش‌بینی برای شرایط خاص باشد، روش‌های مبتنی بر روش‌های مدل سازی هوش مصنوعی از قبیل رگرسیون، یادگیری عمیق و شبکه عصبی گزینه مناسب‌تری می‌باشند.

جدول (۱) مقایسه روش‌های Google Maps و تحقیقات مبتنی بر شبکه عصبی

پارامترهای مقایسه	Google Maps	تحقیقات مبتنی بر شبکه عصبی
نوع داده‌ها	Waze, GPS, حسگرهای ترافیک	داده‌های تاریخی ترافیک، آب و هوا، رویدادها، شبکه‌های اجتماعی
مدل‌ها	مدل‌های ترافیکی و یادگیری ماشین از پیش ساخته شده	مدل‌های عصبی جدید
اهداف	پیش‌بینی‌های دقیق و به‌روز	درک عمیق‌تر از ترافیک، توسعه روش‌های جدید پیش‌بینی
مزایا	دقت، قابلیت انطباق، قابلیت تعمیم	دقت بالا، پتانسیل برای نوآوری
معایب	نیاز به داده‌های زیاد، تفسیر دشوار نتایج، محدودیت‌های محاسباتی	نیاز به داده‌های زیاد، تفسیر دشوار نتایج، محدودیت‌های محاسباتی

۴- مدل سازی پارامترهای ترافیک

در مهندسی ترافیک، مدل سازی و پیش‌بینی زمان سفر در درون شبکه‌های حمل و نقل از اهمیت ویژه‌ای برخوردار است. بهمین دلیل بررسی عوامل تاثیرگذار بر ترافیک در اینگونه تحلیل‌ها در کیفیت و دقت نتایج حاصله نقش مهمی را ایفا می‌نماید. زمان سفر، فاصله زمانی برای رسیدن وسیله نقلیه از یک نقطه به نقطه دیگر در نظر گرفته می‌شود. مدل‌های مختلف برای مدل سازی و پیش‌بینی زمان سفر استفاده می‌شود، از جمله مدل‌های آماری، مبتنی بر شبکه، مبتنی بر مکان و ترکیبی [۳۵]. برای توسعه برنامه‌های عملیاتی و راهبری سیستم‌های حمل و نقل و کنترل ترافیک، استفاده از مدل سازی ریاضی برای بررسی دقیق عوامل و پارامترهای موثر بر نقاط ازدحام و مشکلات ترافیکی بسیار حیاتی است. در یک حالت کلی و ساده، مدل ریاضیاتی پارامترهای موثر بر ترافیک سطح معابر را می‌توان به صورت رابطه (۳) زیر بیان کرد:

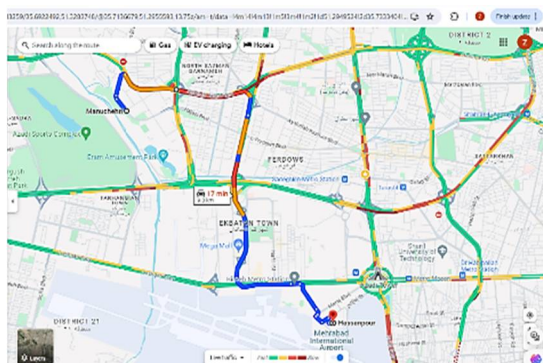
$$Q = f(A, B, C, D, E, F, \dots) \quad (3)$$

که Q میزان ترافیک در معبر، A طرح هندسی معبر، B نوع روسازی، C شرایط آب و هوایی، D وضعیت اقتصادی،

E وضعیت اجتماعی، F عوامل فرهنگی و ... سایر عوامل تاثیرگذار بر ترافیک می‌باشند. هر یک از این پارامترها می‌توانند بر میزان ترافیک تأثیر مثبت یا منفی داشته باشند [۳۶، ۳۷]. زمان سفر و جریان ترافیک رابطه پیچیده‌ای دارند و با افزایش حجم ترافیک، زمان سفر نیز افزایش می‌یابد. مدل سازی زمان سفر با در نظر گرفتن عوامل مختلف نظیر آب و هوا، تاریخ، روز، و ساعت کاری، به بهینه سازی ترافیک کمک می‌کند [۳۶، ۳۷]. اطلاعات دقیق در مورد زمان سفر، امکان بهینه سازی مسیرهای ترافیکی، جلوگیری از ترافیک بیشتر، و شبیه سازی رویدادهای ترافیکی مهم مانند تصادفات و اطلاع رسانی به رانندگان فراهم می‌کند.

۴-۱- روند جریان ترافیک شهری

جریان ترافیک یا حجم ترافیک، تأثیر زیادی بر زمان سفر دارد. حجم ترافیک در شهرها تغییر می‌کند و وابسته به عواملی نظیر ساعت، روز هفته، فصل و رویدادهاست. تغییرات روزانه حجم ترافیک در ساعات مختلف افزایش و کاهش می‌یابد. خطوط نمودار جریان ترافیک روزانه، در ساعات اوج بالا و در ساعات با فعالیت کمتر، پایین است. در



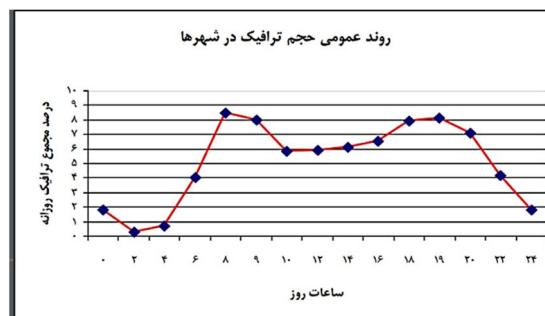
شکل (۴) منطقه مورد مطالعه و نقاط ابتدا و انتهای مسیر

لازم به ذکر است که شرایط ترافیکی طبق گزارش‌ها و مطالعات مراکز ترافیکی، در این بازه صرفاً حائز اهمیت است و در سایر ساعات شبانه‌روز (ساعات صفر بامداد تا ۷ صبح) جریان ترافیک روان، و خارج از دایره نگرانی‌های توسعه سیستم‌های ترافیکی می‌باشد. از این رو، در این پژوهش نیز بمنظور تمرکز بر ساعات ترافیک زنده شهری و پرهیز از انباشتگی اطلاعات غیرضروری، به بازه زمانی استاندارد ترافیک شهری پرداخته شد. لازم به توضیح است که منطقه مورد مطالعه محدود به این مسیر و نقاط مبدا و مقصد در نظر گرفته شد تا بتوان با تمرکز بر یک مسیر، پارامترها و مدل‌ها را مطالعه دقیق نمود و طبعاً با نتایج حاصله، بتوان از دستاوردهای این تحقیق برای تعمیم به شبکه معابر وسیع‌تر در سطح شهر استفاده نمود.

۵-۱- یارامترهای منتخب ورودی مدل‌سازیها

پارامترهایی که در این پژوهش، مورد بررسی و پایش قرار گرفتند، عبارتند از: تاریخ، زمان (ساعت و دقیقه)، روزهای تعطیل یا اداری، جریان ترافیک، تصادفات پیش آمده، نقاط قفل کننده ترافیکی و شرایط بارش جوی. توضیح چرایی انتخاب پارامترهای منتخب در این تحقیق اینست که، تمامی موارد فوق‌الذکر در مسیر ثابتی (اطلاعات پارامترهای ترافیکی در این مسیر در بازه زمانی ۱۴۰۳/۱/۱۴ تا ۱۴۰۳/۳/۱۵ در ۶۵۹۲ رکورد اطلاعاتی ثبت گردید) ثبت و پایش شدند. اینکه پارامترهای وضعیت هندسی معابر در این مدل‌ها استفاده نشد، این است که در تمام زمان و جریان ترافیک و سایر پارامتر در کل یک مسیر ثابت مورد بررسی قرار گرفته‌است، از این‌رو، تخصیص ویژگی هندسه معابر در دسته پارامترهای ورودی، تغییری را منجر نمی‌شود و ضریب ثابتی در این محاسبات به خود تخصیص می‌داد. در ادامه

این شکل (۳) که از مطالعات حجم ترافیک الهام گرفته است [۳۸، ۳۹]، حجم ترافیک در طول روز به صورت یک منحنی سینوسی نشان داده شده است. این منحنی دارای دو قله و دو دره است. قله اول در صبح زود مشاهده می شود و مربوط به ترافیک صبحگاهی است، قله دوم در عصر مشاهده می شود و مربوط به ترافیک عصرگاهی است. دره اول در ظهر مشاهده می شود و مربوط به کاهش ترافیک در طول روز است. دره دوم در نیمه شب مشاهده می شود و مربوط به کاهش ترافیک شبانه است. بررسی حجم ترافیک ساعت اوج ضروری است. مدل های شبیه سازی جریان ترافیک، می توانند این پارامترها را برای پیش بینی و کنترل بهتر ترافیک مدل سازی کنند. در این تحقیق، از مدل سازی عددی جریان ترافیک روزانه در تخمین زمان سفر نقاط مبدا و مقصد استفاده شده است.



شکل (۳) نمودار روند جریان ترافیک

۵- پیاده‌سازی و اجرا در منطقه مورد مطالعه

منطقه مورد مطالعه، مسیری در نظر گرفته شد که مطابق شکل (۴) بخشی از منطقه پنج تهران را به منطقه‌ای حوالی میدان آزادی متصل می‌کند. در اتصال نقاط ابتدایی و انتهایی مسیر منتخب، مسیرهای مواصلاتی متعددی وجود دارد که از بین آنها، یک مسیر اصلی که در برگرنده بزرگراه و شریان‌های اصلی می‌باشد، انتخاب شد و برای مدت دو ماه از تاریخ ۱۴۰۳/۱/۱۴ الی ۱۴۰۳/۳/۱۵ زمان‌سفر و پارامترهای موثر بر ترافیک همچون (تاریخ برداشت، ساعت، نوع روز، شرایط بارش جوی، تصادف، ضریب پیک ترافیک) پایش و مشاهدات بصورت هر ۱۰ دقیقه یکبار از Google Map برداشت شد [۴۰]. بازه زمانی پایش مشاهدات از ۷ صبح تا ۱۲ شب در نظر گرفته شد.

نمونه‌ای از ساعات برداشت اطلاعات به همراه پارامترهای پایش شده آورده شده است. پارامترهای ورودی، معرف مولفه‌های موثر بر ترافیک هستند که عبارتند از؛ وضعیت جوی متاثر از بارش (ضریب یک در روزهای بارانی و صفر در روزهای غیربارانی)، ضریب پیک ترافیک (با توجه به نمودار شکل ۳)، و وقوع تصادف در مسیر مورد نظر (ضریب یک برای تصادف ثبت شده و صفر بدون تصادف)، زمان برداشت اطلاعات؛ هر ده دقیقه یکبار (به مقادیر عددی متوالی تبدیل گردید)، نوع روز به لحاظ اداری یا تعطیلی و تاریخ‌های مناسبت‌دار (حد ضرایب از صفر تا یک)، روز برداشت اطلاعات که مقدار عددی شمارشی برای آن لحاظ شد، و نهایتاً ستون آخر زمان سفر آنی گرفته شده از Google Map (بعنوان متغیر وابسته). لازم به ذکر است که پس از جمع‌آوری اطلاعات و تدوین آنها، در مرحله آماده‌سازی اطلاعات، باید مقادیر ورودی را نرمال‌سازی نمود، زیرا شبیه‌سازی و مدل‌سازی پارامتر وابسته از چندین پارامتر مستقل، معمولاً با مجموعه‌ای از داده‌های مختلف سر و کار دارد که هر کدام دامنه و مقیاس خاص خود را دارند. این تفاوت‌ها می‌توانند بر روی نتایج نهایی مدل تأثیر منفی بگذارند. نرمال‌سازی داده‌ها به منظور کاهش این تأثیرات و تسهیل فرآیند یادگیری ماشین یا تحلیل داده‌ها انجام می‌شود. بسیاری از الگوریتم‌های یادگیری ماشین، مانند رگرسیون‌ها و شبکه‌های عصبی، به شدت تحت تأثیر مقیاس داده‌ها قرار می‌گیرند. نرمال‌سازی می‌تواند سرعت همگرایی این الگوریتم‌ها را افزایش دهد و در نتیجه زمان لازم برای آموزش مدل را کاهش دهد. در این تحقیق نیز، با نرمال‌سازی داده‌ها به محدوده صفر و یک، تمام پارامترها در

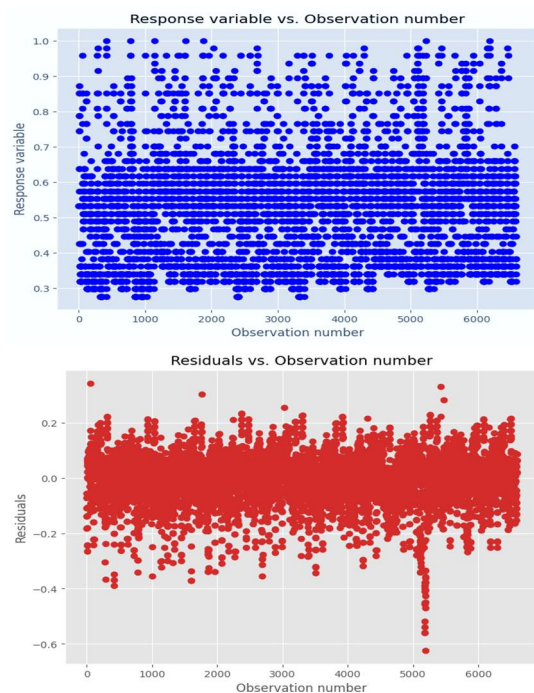
یک مقیاس یکسان قرار گرفتند که این امر منجر به الگوریتم‌های یادگیری ماشین کمک می‌کند تا به طور مؤثرتری به الگوهای موجود در داده‌ها توجه کنند و از تأثیر مقادیر بزرگ یا کوچک جلوگیری شود. این فرایند نهایتاً منجر به بهبود دقت و کارایی مدل شده و تحلیل داده‌ها را نیز ساده‌تر و قابل فهم‌تر می‌سازد. پس از نرمال‌سازی نیز، پس از اخذ نتایج، به منظور تحلیل و مقایسه خروجی‌ها، از فرایند معکوس نرمال‌سازی عددی استفاده گردید. در جدول (۲)، برشی از اطلاعات نرمال‌سازی شده آورده شده است.

۵-۲- رویکرد رگرسیون خطی و رگرسیون چند متغیره در شبیه‌سازی جریان ترافیک

در مطالعات ترافیک، عموماً داده‌های جریان ترافیک جامعیت ندارند و باید از روش‌های فناورانه مدل‌سازی برای شبیه‌سازی جریان ترافیک استفاده نمود. رگرسیون جغرافیایی یکی از بهترین روش‌های تخمین ترافیک بر اساس اطلاعات داده و زمان است و امکان ادغام معیارهای توپولوژیکی و هندسی را برای تخمین جریان ترافیک فراهم می‌کند. نتایج این مدل رگرسیون، منجر به پیش‌بینی زمان سفر و متعاقباً جریان ترافیک می‌گردد. بدین منظور، از بین روش‌های مختلف رگرسیون، مدل‌های خطی و چندمتغیره جهت ادامه کار، در نظر گرفته شدند. جهت استفاده از رگرسیون چند متغیره در تخمین جریان ترافیک، ابتدا داده‌هایی شامل مقادیر جریان ترافیک و متغیرهای مستقل (مانند ساعت روز، تعداد خودروها و ویژگی‌های توپولوژیکی و هندسی شبکه راه) جمع‌آوری شدند.

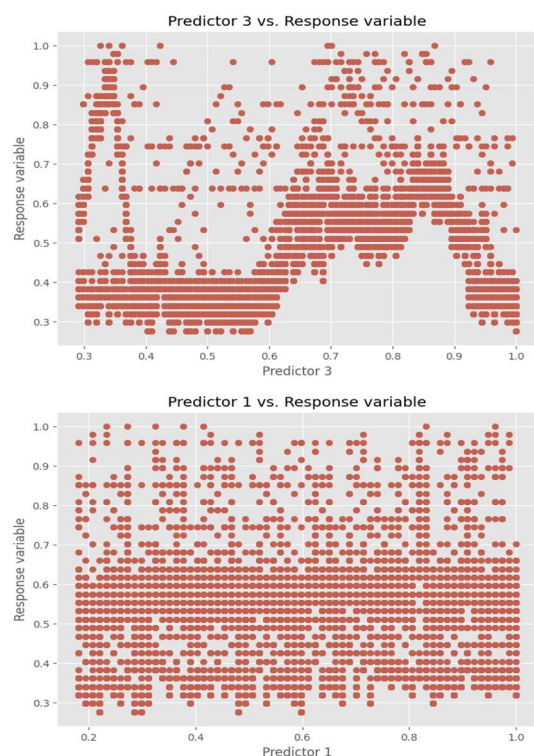
جدول (۲) نرمال‌سازی شده مقادیر پارامترهای موثر بر ترافیک جمع‌آوری شده

Weather	Traffic pick	Accident	Record's time	Type_date	Date_number	Dependent-variable
۰	۰.۵	۰	۰.۲۹۱۶	۰.۸	۰.۱۸۱۸	۰.۵۳۱۹
۰	۰.۶	۰	۰.۲۹۸۶	۰.۸	۰.۱۸۱۸	۰.۵۷۴۴
۰	۰.۷	۰	۰.۳۰۵۵	۰.۸	۰.۱۸۱۸	۰.۶۳۸۲
۰	۱	۰	۰.۳۱۲۵	۰.۸	۰.۱۸۱۸	۰.۷۴۴۶
۰	۱	۰	۰.۳۱۹۴	۰.۸	۰.۱۸۱۸	۰.۷۸۷۲
۰	۱	۰	۰.۳۲۶۳	۰.۸	۰.۱۸۱۸	۰.۷۸۷۲
۰	۱	۰	۰.۳۳۳۳	۰.۸	۰.۱۸۱۸	۰.۸۵۱۰
۰	۱	۰	۰.۳۴۰۲	۰.۸	۰.۱۸۱۸	۰.۸۷۲۳
۰	۱	۰	۰.۳۴۷۲	۰.۸	۰.۱۸۱۸	۰.۸۵۱۰



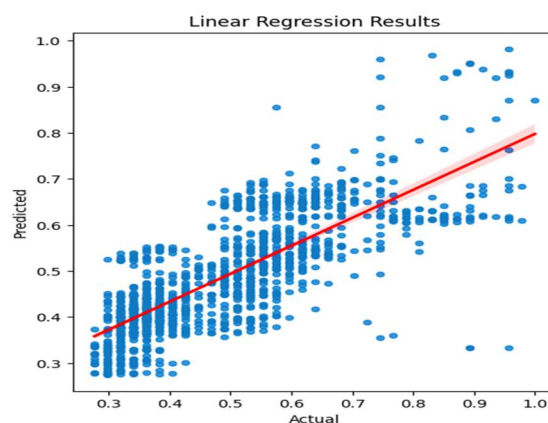
شکل (۶) مقادیر باقیمانده‌ها و پراکندگی متغیرهای مستقل و وابسته

مطابق تصاویر، نقشه‌های پراکندگی متغیرهای مستقل؛ هر نقشه نشانگر رابطه بین متغیر مستقل و متغیر وابسته است. محور X نشانگر متغیر مستقل و محور Y نشانگر متغیر وابسته است.



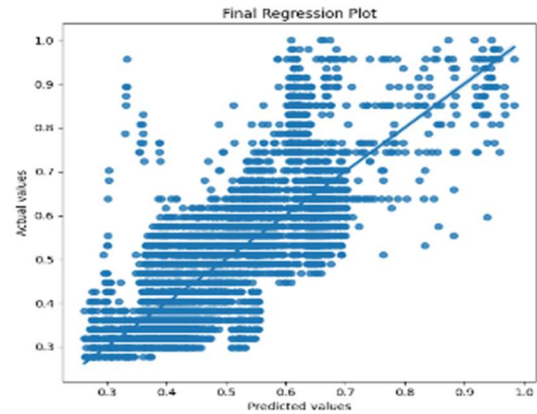
شکل (۷) پراکندگی دو نمونه از مشاهدات مستقل رگرسیون چندمتغیره

سپس داده‌ها برای تجزیه و تحلیل آماده شد و تحلیل همبستگی با استفاده از روش‌های تصویرسازی داده و ماتریس همبستگی انجام گردید. در مرحله بعد، با استفاده از اطلاعات داده‌ها، مدلی برای تخمین جریان ترافیک ایجاد شد. سپس از رگرسیون برای مدل‌سازی رابطه بین این پارامترها و زمان سفر استفاده گردید. در این فرایند، مجموعه داده به دو بخش آموزشی و آزمایشی تقسیم شد، ۸۰٪ از داده‌ها برای آموزش و ۲۰٪ برای آزمایش در نظر گرفته شدند. نتایج حاصل روش‌های رگرسیون خطی مطابق شکل (۵) و مقدار R^2 -Squared بدست آمده برابر با 0.61515 می‌باشد. در مدل رگرسیون چندگانه یا چندمتغیره نیز، متغیر وابسته (زمان سفر)، و متغیرهای مستقل (عوامل موثر بر متغیر وابسته) در نظر گرفته شدند. بنابراین در مدل چندمتغیره، ابتدا نا چند متغیر مستقل در یک مجموعه داده حاوی شش پارامتر ورودی، از جمله حجم ترافیک، زمان روز، روز هفته، آب و هوا، تصادفات جاده و زمان سفر مشاهده شده جمع‌آوری گردید.



شکل (۵) نمودار رگرسیون حاصل از مدل رگرسیون خطی

سپس، ضرایب و عبارت تقریبی در مدل‌سازی متغیرهای متغیر و وابسته ساخته و مورد آزمایش قرار گرفتند. مقادیر باقیمانده‌ها و پراکندگی متغیرهای مستقل و وابسته مطابق با اشکال ۶ می‌باشند. مدل رگرسیون خطی در نهایت، دقت متوسطی در پیش‌بینی زمان سفر با مقدار R^2 -Squared برابر با ۰.۶۲۶۴ دارد.



شکل (۸) نمودار رگرسیون نهایی مدل چند متغیره

جهت اجرای مدل رگرسیون خطی و چندمتغیره، مدلی به زبان پایتون توسعه داده شد که کد یک فایل اکسل را که مقادیر ورودی‌های مدل است، دریافت و آن را به یک فریم داده پانداس تبدیل می‌کند و مدل رگرسیون را اجرا، و چند نقشه برای تجسم داده‌ها و عملکرد مدل ایجاد می‌کند. در مدل چندمتغیره، پراکندگی دو نمونه از متغیرهای مستقل مطابق با اشکال (۷) و نمودار رگرسیون نهایی مطابق شکل (۸) می‌باشند.

مقدار حاصل Intercept این تحلیل 0.2316 نشان‌دهنده مقدار ثابت یا مقدار پیش‌بینی شده برای متغیر وابسته (در اینجا، زمان سفر) است زمانی که تمامی متغیرهای مستقل (پارامترها) صفر باشند. به عبارت دیگر، اگر تمامی متغیرهای مستقل صفر باشند، مقدار پیش‌بینی شده برای متغیر وابسته (زمان سفر) حدوداً برابر با ۰.۲۳۱۶ است.

جدول (۴) مقادیر حاصل از رگرسیون

	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰
Coefficients						

این ضرایب مطابق مقادیر مندرج در جدول (۴) نشان می‌دهند که افزایش ۱٪ اولین پارامتر تاثیرگذار در ترافیک، منجر به افزایش ۰.۰۴۴۹٪ در زمان سفر شده، تغییر ۱٪ دومین پارامتر تاثیرگذار در ترافیک منجر به تغییر 0.111٪ در زمان سفر، افزایش ۱٪ سومین پارامتر تاثیرگذار در ترافیک منجر به افزایش 0.0726٪ در زمان سفر، افزایش ۱٪ چهارمین پارامتر تاثیرگذار در ترافیک منجر به کاهش 0.0128٪ در زمان سفر، افزایش ۱٪ پنجمین پارامتر تاثیرگذار در ترافیک منجر به افزایش 0.2536٪ در زمان سفر، و نهایتاً افزایش ۱٪ ششمین پارامتر تاثیرگذار در

ترافیک منجر به افزایش 0.3048٪ در زمان سفر می‌شود. مقدار R^2 (R-squared) برابر با ۰.۶۲۶۴ نیز نشان‌دهنده ضریب تعیین است. این مقدار بیان‌گر این است که مدل توسعه داده شده، توانسته‌است حدود ۶۲.۶۴٪ از واریانس در زمان سفر را توضیح دهد. از طرفی، ۳۷.۳۵٪ از تغییرات در زمان سفر توسط متغیرهای پیش‌بینی در مدل توضیح داده نمی‌شود و احتمالاً به عوامل دیگری مانند رفتار راننده، شرایط جاده، آب و هوا، و غیره، نسبت داده می‌شود. این نتیجه نشان می‌دهد که متغیرهای پیش‌بینی در مدل تأثیر مهمی بر زمان سفر دارند، اما وجود عوامل دیگری که در مدل مورد بررسی قرار نگرفته‌اند نیز احتمالاً می‌تواند تأثیرگذار باشند. تا این بخش از مطالعه، از رگرسیون خطی و چندمتغیره برای بررسی رابطه بین زمان سفر و پارامترهای مختلف ترافیک استفاده گردید. از این نتیجه در تخمین ظرفیت جاده‌ها و توزیع مجدد جریان ترافیک در شبکه‌های جاده‌ای استفاده می‌شود. این روش می‌تواند به بهبود مدیریت ترافیک و کاهش ازدحام ترافیک کمک کند. در جدول (۵)، نیز مقادیر تقریبی جهت تفسیر کلی R^2 ، تشریح شده‌است.

جدول (۵) مقادیر تقریبی جهت تفسیر R^2

سطح مدل	مقدار R^2
بسیار ضعیف	۰.۰ تا ۰.۲
ضعیف	۰.۲ تا ۰.۴
متوسط	۰.۴ تا ۰.۶
قوی	۰.۶ تا ۰.۸
بسیار قوی	۰.۸ تا ۱.۰

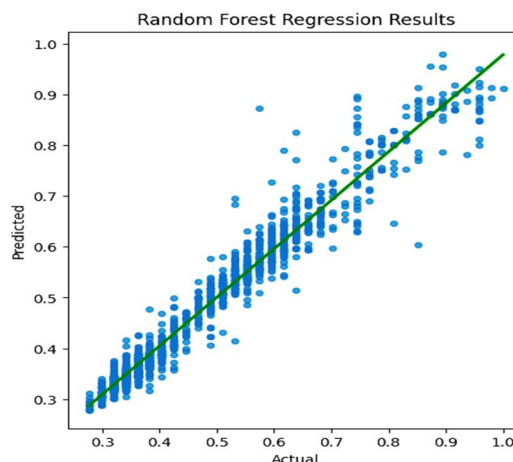
مدل رگرسیون جنگل تصادفی؛ رگرسیون جنگل

تصادفی قادر به مدل کردن روابط غیرخطی و پیچیده است و دارای مقاومت بالا در برابر داده‌های نویز دار می‌باشد. این تکنیک نیز بدلیل برخورداری از دقت بالا، برآزش بهتر داده‌های غیرخطی و تعامل با ویژگی‌های مختلف، برای پیش‌بینی زمان سفر انتخاب شد. مقدار R^2 منتج از این مدل برابر 0.95431 حاصل گردید. نمودار رگرسیون این روش نیز در شکل (۹) نشان داده شده‌است.

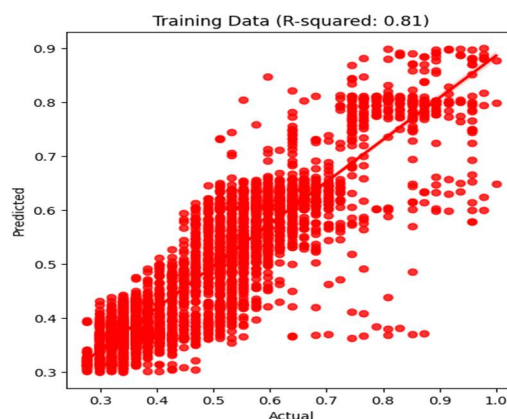
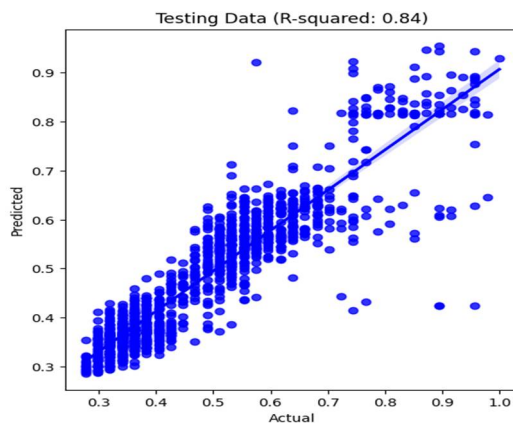
مدل شبکه عصبی عمیق؛ در روش دیگر استفاده

شده در این تحقیق، از شبکه عصبی مصنوعی عمیق برای شبیه‌سازی زمان سفر استفاده شده‌است. این شبکه، شامل

مدل MLPRegressor است که یک شبکه عصبی ژرف برای رگرسیون استفاده می‌کند.



شکل (۹) نمودار رگرسیون جنگل تصادفی

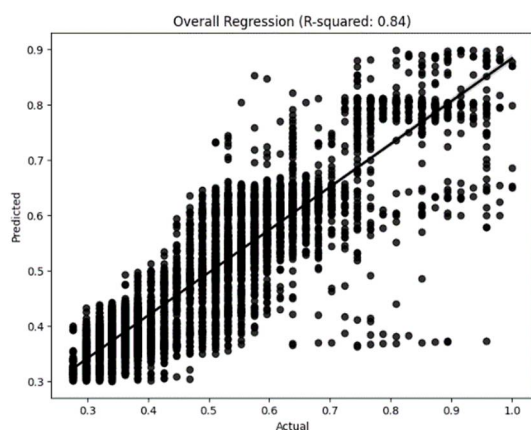


شکل (۱۰) رگرسیون مقادیر آموزشی و تست شبکه عصبی عمیق

در این مدل از کتابخانه scikit-learn برای آموزش و پیش‌بینی مقادیر بر اساس ویژگی‌های داده‌ها و از

TensorFlow نیز برای ساخت شبکه و DNN استفاده شده است. TensorFlow توسط گوگل توسعه داده شده و ابزارها و منابع بسیاری را برای یادگیری عمیق و پردازش عددی فراهم می‌کند. این روش با استفاده از شبکه عصبی توانایی حل مسائل پیچیده و غیرخطی را داراست و از مفاهیم گراف و پردازش موازی^۱ استفاده می‌کند. روش مورد استفاده TensorFlow، پردازش موازی، بهینه‌سازی گراف^۲، پشتیبانی از GPU^۳ و TPU^۴، و ابزارهای متنوع برای بارگذاری و پردازش داده می‌باشد تا بهترین عملکرد را برای محاسبات یادگیری عمیق فراهم کند. نمودارهای شکل (۱۰) نمایانگر نمودارهای رگرسیون مقادیر آموزشی و تست شبکه عصبی عمیق می‌باشد.

در این پژوهش، کد توسعه داده‌شده با الگوریتم شبکه عصبی عمیق برای مدل‌سازی بوده و عملکرد مدل نیز با استفاده از R^2 ارزیابی گردید. برنامه برای ارزیابی مدل از معیار R^2 از کتابخانه score_ و از sklearn برای ارزیابی دقت مدل در پیش‌بینی مقادیر "هدف" استفاده کرده که نتایج حاصله از شبکه عصبی عمیق در مدل‌سازی زمان‌سفر با مقدار R^2 برابر با 0.840 مطابق شکل (۱۱) می‌باشد. به طور کلی، TensorFlow ابزاری قدرتمند و همه‌کاره برای ساخت، آموزش و استقرار مدل‌های یادگیری عمیق است که به طور گسترده‌ای در تحقیقات و کاربردهای دنیای واقعی مورد استفاده قرار می‌گیرد.



شکل (۱۱) نمودار رگرسیون نهایی شبکه عصبی عمیق

در شکل (۱۲)، معماری شبکه عصبی عمیق مورد استفاده و تعداد پارامترهای ورودی، لایه‌های میانی و پارامتر خروجی نیز نشان داده شده‌است. این شبکه عصبی

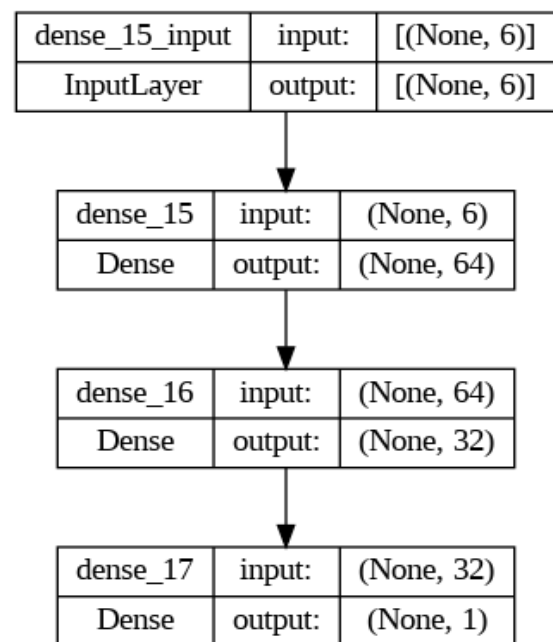
^۳ Graphics Processing Unit

^۴ Tensor Processing Unit

^۱ Parallel Processing

^۲ Graph Optimization

شامل ۴ لایه است؛ یک لایه ورودی، دو لایه پنهان و یک لایه خروجی. در لایه ورودی، ۶ نورون قرار داده شده است. "None" نشان دهنده اینست که تعداد نمونه‌های ورودی می‌تواند متغیر باشد. در لایه‌های متراکم (پنهان) هر نورون به همه نورون‌های لایه قبلی متصل است. لایه اول متراکم ۶۴ نورون دارد و خروجی آن یک بردار ۶۴ بعدی است. لایه دوم متراکم ۳۲ نورون دارد و خروجی آن یک بردار ۳۲ بعدی است. لایه آخر متراکم ۱ نورون دارد و خروجی آن یک عدد اسکالر است. با توجه به اینکه لایه خروجی تنها یک نورون دارد، این شبکه برای یک مسئله رگرسیون طراحی شده است. در مدل DNN استفاده شده، حدود ۷۰ درصد از کل داده‌ها به عنوان داده‌های آموزشی در نظر گرفته شد. این داده‌ها برای آموزش مدل استفاده شدند تا الگوریتم بتواند الگوها و روابط موجود در داده‌ها را یاد بگیرد. همچنین حدود ۳۰ درصد از کل داده‌ها به عنوان داده‌های آزمون اختصاص یافت. این داده‌ها بعد از آموزش مدل مورد استفاده قرار گرفتند تا عملکرد مدل بر روی داده‌هایی که قبلاً ندیده است، ارزیابی شود.



شکل (۱۲) معماری DNN توسعه داده شده

۶- ارزیابی نتایج

بطور کلی، در کار با مدل‌های یادگیری عمیق، به منظور بررسی و ارزیابی عملکرد مدل، داده‌ها به دو دسته اصلی تقسیم شدند: داده‌های آموزشی و داده‌های تست. برای

داده‌های آموزشی، حدود ۷۰ الی ۸۰ درصد از کل داده‌ها در نظر گرفته شدند که برای آموزش مدل استفاده شدند. از سوی دیگر، برای داده‌های تست ۲۰ الی ۳۰ درصد باقی‌مانده به منظور ارزیابی عملکرد مدل پس از مرحله آموزش در نظر گرفته شدند. این تقسیم‌بندی به مدل این امکان را می‌دهد که قادر باشد الگوهای موجود در داده‌ها را یاد بگیرد و سپس با استفاده از داده‌هایی که در طول فرایند آموزش دیده نشده‌اند، عملکرد خود را مورد بررسی قرار دهد.

در این تحقیق، پس از استفاده از مدل‌های مختلف در آموزش و آزمون داده‌های مختلف، نتایج کلی مطابق با مقادیر نشان داده شده در جدول (۶) حاصل گردید.

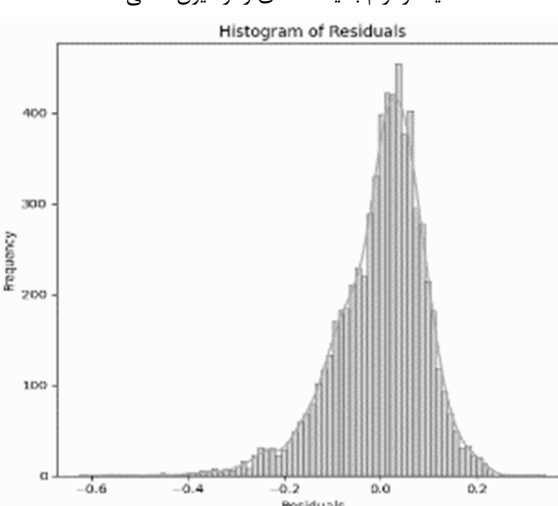
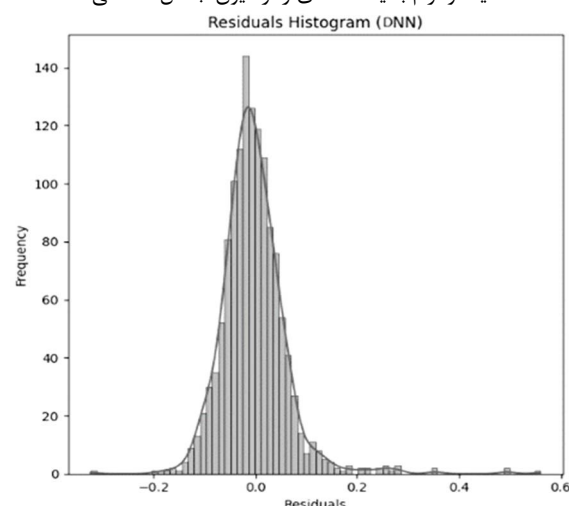
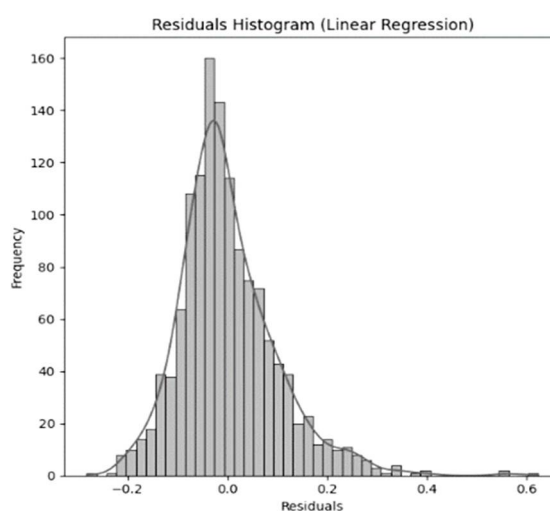
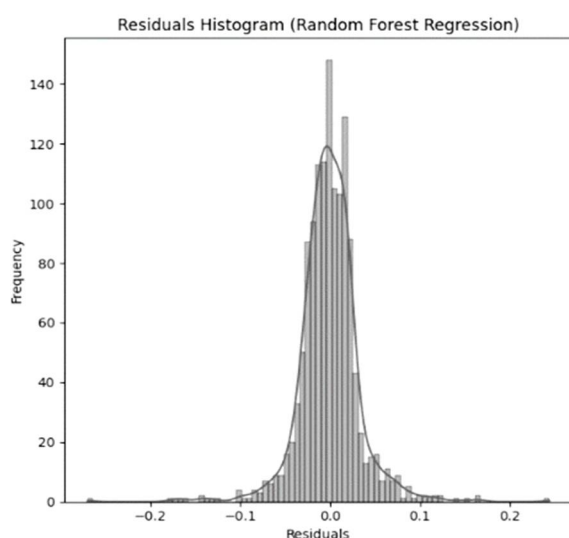
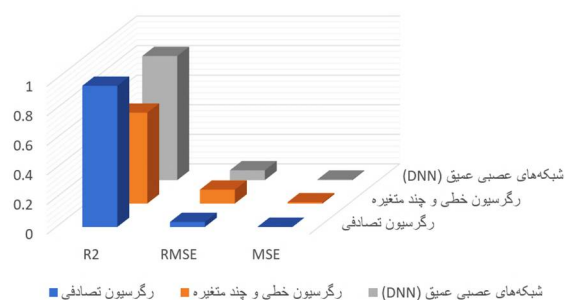
جدول (۶) جدول مقایسه پارامترهای ارزیابی مدل‌ها

پارامتر ارزیابی	شبکه‌های عصبی عمیق	رگرسیون خطی و چند متغیره	رگرسیون تصادفی
R^2	۰.۸۴	۰.۶۱۵	۰.۹۵۵
RMSE	۰.۰۶۷۱	۰.۰۹۴۳	۰.۰۳۴۰۵
MSE	۰.۰۰۴۵۱	۰.۰۰۸۹	۰.۰۰۱۱۶
تفسیر R^2	عملکرد خوب	تطبیق ضعیف‌تر	بهترین عملکرد
توزیع نمودار باقیمانده‌ها	توزیع مرکزی با انحراف کم، نزدیک به نرمال	توزیع نزدیک به نرمال با برخی اختلافات	توزیع نزدیک به نرمال با انحرافات کم

پارامترهای ارزیابی که در این مطالعه مورد استفاده قرار گرفتند، موارد R^2 ، MSE و RMSE می‌باشند. پارامتر R^2 نشان‌دهنده این است که چه درصدی از واریانس (اختلاف) متغیر وابسته (هدف) به وسیله متغیرهای مستقل (پیش‌بینی‌گرها) توضیح داده می‌شود. مقدار آن بین صفر و یک است و هر قدر به یک نزدیک باشد، یعنی مدل واریانس بیشتری از داده‌ها را توضیح می‌دهد. مقدار MSE میانگین مربعات خطا بین مقادیر پیش‌بینی شده توسط مدل و مقادیر واقعی است. مقدار RMSE نیز جذر MSE است و نشان‌دهنده میانگین مقدار خطای مطلق پیش‌بینی‌ها است. هر چقدر مقدار MSE و RSME پائین‌تر باشد، نشان می‌دهد که خطای پیش‌بینی مدل کمتر است. نتایج شبکه‌های عصبی عمیق حاکی از مقادیر R^2 برابر با ۰.۸۴، MSE و RMSE به ترتیب ۰.۰۰۴۵۱ و ۰.۰۶۷۱ می‌باشد که این مقادیر نشان‌دهنده توانایی خوب این مدل در توجیه داده‌ها است. همچنین، هیستوگرام باقیمانده‌ها مطابق شکل (۱۳) نشان می‌دهد که باقیمانده‌ها در مدل‌های مختلف به

ضعیف‌تر از دو مدل فوق‌الذکر بدست آمد. مقایسه نموداری این نتایج نیز در شکل شماره (۱۴)، آورده شده‌است.

پارامترهای ارزیابی مدل‌های یادگیری ماشین



شکل (۱۳) هیستوگرام باقیمانده‌های مدل‌های رگرسیون خطی، چندمتغیره، RFR و DNN

شکل (۱۴) نمودار مقایسه ای پارامترهای ارزیابی مدل‌ها

۶-۱- ارزیابی مستقل مدل‌های منتخب

جهت ارزیابی مستقل و مجدد نتایج، داده‌ها به دو مجموعه موقت و ارزیابی مستقل تقسیم شدند (۱۰٪ داده ارزیابی مستقل و ۹۰٪ داده موقت). سپس از مجموعه موقت، مجدداً داده‌ها به دو مجموعه آموزش و آزمون (۷۰٪ داده برای آموزش و ۳۰٪ داده برای آزمون) تقسیم گردیدند. برای اجرای ارزیابی نیز، ابتدا مدل بر روی مجموعه آزمون و سپس بر روی مجموعه ارزیابی مستقل، اجرا شد.

نتیجه این رویکرد، حصول اطمینان از این جهت که مدل نه تنها در روی داده‌های آموزشی و آزمون، بلکه در دنیای واقعی و داده‌های ناشناخته نیز کارایی دارد. با این روش، ارزیابی و صحت‌سنجی روی نتایج صورت می‌گیرد و می‌توان کارایی مدل را ارزیابی نمود. نتایج حاصل از ارزیابی مستقل بر دو مدل شبکه عصبی عمیق و رگرسیون جنگل تصادفی در جدول (۷) آورده شد. تفسیر این مقادیر مشابه با تفسیر مجموعه آزمون و آموزش می‌باشد. همانطور که نشان داده شده، مقادیر بسیار به نتایج ارزیابی مستقیم مدل‌ها نزدیک می‌باشد و همچنان مدل RFR برتری‌هایی نسبت به مدل DNN در مدل‌سازی داده‌های آنومالی دارد.

جدول (۷) جدول مقایسه پارامترهای ارزیابی مستقل مدل‌ها

پارامتر ارزیابی	شبکه‌های عصبی عمیق (DNN)	رگرسیون تصادفی RFR
R^2	۰.۸۲۲۴	۰.۹۴۲۶۲
RMSE	۰.۰۶۹۳۲	۰.۰۳۴۸۴
MSE	۰.۰۰۴۸۱	۰.۰۰۱۲۱

نتایج این بخش موید این نکته است که مدل عملکرد خوبی بر روی داده‌های آموزشی و آزمون داشته و در عین حال توانسته داده‌هایی ارزیابی مستقل را نیز به خوبی پیش‌بینی کند. لذا، پژوهشگران با این روش ارزیابی می‌توانند امکان شناسایی نقاط قوت و ضعف مدل را داشته باشند و به بهینه‌سازی آن بپردازند.

۷- نتیجه‌گیری

براساس نتایج حاصله، دو مدل شبکه عصبی عمیق و رگرسیون جنگل تصادفی توانمندی بالایی در پیش‌بینی و تطبیق با داده‌ها از خود نشان دادند. در مقایسه RFR

بهترین عملکرد را از بین مدل‌های DNN، رگرسیون خطی، و رگرسیون چندمتغیره از خود نشان داده‌است و در رتبه بعد، شبکه‌های عصبی عمیق قرار دارد. مدل‌های عمیق به طور کلی به نوبت در داده‌ها حساس‌تر هستند. از آنجائیکه مهم‌ترین محور مورد بررسی این تحقیق، حوادث رانندگی بوده، طبق این مساله اساساً نوبت قابل توجهی در داده‌های ترافیک ایجاد می‌کند. تفسیر نتایج حاصله این است که وجود نویز، منجر به کاهش دقت شبکه عصبی عمیق در پیش‌بینی زمان سفر می‌شود. در مقابل، جنگل‌های تصادفی به طور ذاتی در برابر نویز مقاوم‌تر هستند و ذاتاً می‌توانند با موفقیت بیشتری داده‌های ناقص و نویزدار را مدل نمایند. شبکه‌های عصبی عمیق مدل‌های پیچیده‌ای هستند که تعداد زیادی پارامتر دارند. این امر می‌تواند منجر به بیش‌برازشی شود، به خصوص زمانی که داده‌های آموزشی محدود باشند. در مقابل، جنگل‌های تصادفی مدل‌های ساده‌تری هستند که تعداد پارامترهای کمتری دارند. این امر به آنها کمک می‌کند تا از بیش‌برازشی جلوگیری کنند و در عین حال دقت بالایی را در وظایف پیش‌بینی حفظ کنند. جنگل‌های تصادفی از مجموعه‌ای از درختان تصمیم تشکیل شده‌اند که هر کدام از آنها با استفاده از یک زیرمجموعه تصادفی از داده‌ها آموزش داده می‌شوند، این تنوع مدل به جنگل‌های تصادفی کمک می‌کند تا از نگرش‌های مختلف در مورد داده‌ها یاد بگیرند و در نتیجه پیش‌بینی‌های دقیق‌تری ارائه دهند.

بطور کلی، نتایج بدست آمده از این مطالعه، برتری‌هایی را در مقایسه با سایر مطالعات صورت گرفته در این حوزه دارد، از جمله اینکه این مدل‌های DNN و RFR قادرند تا از داده‌های متنوع و جامع منابع مختلف که بر ترافیک اثرگذار هستند، همچون شرایط جوی، ترافیک روزانه، یا رویدادهای خاص و غیره، بهره‌برداری کنند. این امر منجر به تحلیل دقیق‌تر مدل با لحاظ کردن تاثیر این عوامل و بهبود دقت پیش‌بینی زمان سفر می‌شود. در بسیاری از تحقیقات، مدل‌های شبکه عصبی و یادگیری عمیق به دلیل توان مدل‌کردن روابط غیرخطی و پیچیده جهت شبیه‌سازی پیشنهاد شده‌است، که در این مطالعه سعی گردید، نتایج هر دو روش RFR و روش‌های DNN در شرایط یکسان بصورت هم‌زمان بررسی و مقایسه شود و نقاط ضعف و قوت هر کدام مشخص گردد.

۸- پیشنهادات

پیشنهاد می‌شود در مطالعات آتی با تحلیل عمیق‌تر باقیمانده‌ها، بررسی حساسیت مدل‌ها به تغییرات پارامترها و مقایسه بصری عملکرد مدل‌ها، به درک بهتری از نقاط قوت و ضعف هر مدل دست یافت. بطور کلی، با نتایج حاصله می‌توان نتیجه گرفت که الگوریتم‌هایی مانند جنگل تصادفی و شبکه‌های عصبی عمیق، نسبت به سایر روش‌های ساده‌تر یادگیری ماشین، هردو توانایی بیشتری در مدل‌سازی روابط غیرخطی و تعاملات پیچیده بین متغیرها دارند. اما به منظور کسب نتایج با دقت بهتر و رفع نواقص هر کدام، پیشنهاد می‌گردد با ترکیب این روش‌ها، به دقت و کارایی بیشتری

مراجع

در پیش‌بینی زمان سفر دست یافت. جنگل تصادفی می‌تواند الگوهای ابتدایی را شناسایی کند و شبکه‌های عصبی برای مدل‌سازی تعاملات پیچیده‌تر مورد استفاده قرار گیرند. این تأثیر متقابل می‌تواند منجر به بهینه‌سازی مدیریت ترافیک، کاهش زمان انتظار و افزایش ایمنی در حمل‌ونقل شود. علاوه بر این، استفاده از رویکردهای جدیدتر هوش مصنوعی، همچون یادگیری تقویتی به عنوان یک روش هوش مصنوعی، می‌تواند در بهینه‌سازی مسیرها و مدیریت ترافیک نقش بسزایی ایفا کند. این الگوریتم‌ها، با تعامل مستمر با محیط و دریافت بازخورد، قادرند به صورت برخط به تحلیل داده‌ها پرداخته و بهترین مسیرها را برای کاهش ترافیک تعیین کنند.

1. Sheng Z, Lv Z, Li J, Xu Z. Deep spatial-temporal travel time prediction model based on trajectory feature. Computers and Electrical Engineering. 2023 Sep 1;110:108868.
2. Abdollahi M, Khaleghi T, Yang K. An integrated feature learning approach using deep learning for travel time prediction. Expert Systems with Applications. 2020 Jan 1;139:112864.
3. Bai M, Lin Y, Ma M, Wang P. Travel-time prediction methods: a review. In Smart Computing and Communication: Third International Conference, SmartCom 2018, Tokyo, Japan, December 10–12, 2018, Proceedings 3 2018 (pp. 67-77). Springer International Publishing.
4. Fani, Amir Hossein, Naseri, Hamed, Golro, Amir. Modeling the travel time of urban roads using linear regression and time series methods. Traffic engineering, 1400; 1400(85): 21-30 (In Persian)
5. Zhao M, Zhang X, Appiah J, Fontaine MD. Travel Time Reliability Prediction Using Random Forests. Transportation Research Record. 2024 Mar;2678(3)
6. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data. 2021 Dec;8:1-74.
7. Talaie Khoei T, Ould Slimane H, Kaabouch N. Deep learning: Systematic review, models, challenges, and research directions. Neural Computing and Applications. 2023 Nov;35(31):23103-24.
8. Rasa Izadi, Abrishmi, Seyyed Ehsan. Forecasting the traffic situation with machine learning algorithms for short-term and medium-term time horizons. Amirkabir Civil Engineering Journal. 2022 Jun 22;54(4):1503-20 (In Persian)
9. Wen Y, Wu R, Zhou Z, Zhang S, Yang S, Wallington TJ, Shen W, Tan Q, Deng Y, Wu Y. A data-driven method of traffic emissions mapping with land use random forest models. Applied Energy. 2022 Jan 1;305:117916.
10. Tedjopurnomo DA, Bao Z, Zheng B, Choudhury FM, Qin AK. A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. IEEE Transactions on Knowledge and Data Engineering. 2020 Jun 9;34(4):1544-61.
11. Peng H, Du B, Liu M, Liu M, Ji S, Wang S, Zhang X, He L. Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning. Information Sciences. 2021 Nov 1;578:401-16
12. Sayed SA, Abdel-Hamid Y, Hefny HA. Artificial intelligence-based traffic flow prediction: a comprehensive review. Journal of Electrical Systems and Information Technology. 2023 Mar 9;10(1):13.
13. Yassin SS, Pooja. Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach. SN Applied Sciences. 2020 Sep;2:1-3.

14. Abdullah SM, Periyasamy M, Kamaludeen NA, Towfek SK, Marappan R, Kidambi Raju S, Alharbi AH, Khafaga DS. Optimizing traffic flow in smart cities: Soft GRU-based recurrent neural networks for enhanced congestion prediction using deep learning. *Sustainability*. 2023 Mar 29;15(7):5949.
15. ۱۰-۴ Yan, M., et al. (2022). "Deep learning in traffic flow forecasting: A review of methodologies and applications." *Transportation Research Part C: Emerging Technologies*, 135, 103567.
16. Chen, C.M., Liang, C.C., Chu, C.P., 2020a. Long-term travel time prediction using gradient boosting. *Journal of Intelligent Transportation Systems* 24, 109–124.
17. Li, H., Wang, Q., Xiong, W., 2021b. New model of travel-time prediction considering weather conditions: Case study of urban expressway. *Journal of Transportation Engineering, Part A: Systems* 147, 04020161.
18. Muhammad I, Yan Z. SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY. *ICTACT Journal on Soft Computing*. 2015 Apr 1;5(3).
19. Shaygan M, Meese C, Li W, Zhao XG, Nejad M. Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transportation research part C: emerging technologies*. 2022 Dec 1;145:103921.
20. Aliahmadi A, Jafari-Eskandari M, Mozafari A, Nozari H. Comparing linear regression and artificial neural networks to forecast total productivity growth in Iran. *International Journal of Information, Business and Management*. 2016 Feb 1;8(1):93.
21. Wu HC. Fuzzy estimates of regression parameters in linear regression models for imprecise input and output data. *Computational Statistics & Data Analysis*. 2003 Feb 28;42(1-2):203-17
22. Liaw A, Wiener M. Classification and regression by randomForest. *R news*. 2002 Dec 3;2(3):18-22.
23. Schonlau M, Zou RY. The random forest algorithm for statistical learning. *The Stata Journal*. 2020 Mar;20(1):3-29.
24. Bratsas C, Koupidis K, Salanova JM, Giannakopoulos K, Kaloudis A, Aifadopoulou G. A comparison of machine learning methods for the prediction of traffic speed in urban places. *Sustainability*. 2019 Dec 23;12(1):142.
25. Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas MJ. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*. 2015 Dec 1;71:804-18.
26. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*. 2009 Dec;10:1-6.
27. Rithani M, Kumar RP, Doss S. A review on big data based on deep neural network approaches. *Artificial Intelligence Review*. 2023 Dec;56(12):14765-801.
28. Kiliçarslan S, Celik M. RSigELU: A nonlinear activation function for deep neural networks. *Expert Systems with Applications*. 2021 Jul 15;174:114805.
29. B. Azman, S. Hussain, N. Azmi, M. Ghani and N. Norlen, Prediction of distant recurrence in breast cancer using a deep neural network, *Rev. int. métodos numér. Cálculo. diseño ing.* (2022). Vol. 38, (1), https://www.scipedia.com/public/Azman_et_al_2021a
30. Dhruv P, Naskar S. Image classification using convolutional neural network (CNN) and recurrent neural network (RNN): A review. *Machine learning and information processing: proceedings of ICMLIP 2019*. 2020:367-81.
31. Lee K, Eo M, Jung E, Yoon Y, Rhee W. Short-term traffic prediction with deep neural networks: A survey. *IEEE Access*. 2021 Apr 5;9:54739-56.
32. Khajeh Hosseini M, Talebpour A. Traffic prediction using time-space diagram: a convolutional neural network approach. *Transportation Research Record*. 2019 Jul;2673(7):425-35.
33. Jiang W, Luo J. Graph neural network for traffic forecasting: A survey. *Expert systems with applications*. 2022 Nov 30;207:117921.
34. Trapsilawati F, Wijayanto T, Jourdy ES. Human-computer trust in navigation systems: Google maps vs Waze. *Communications in Science and Technology*. 2019 Jul 14;4(1):38-43.

35. Jiber M, Lamouik I, Ali Y, Sabri MA. Traffic flow prediction using neural network. In 2018 International Conference on Intelligent Systems and Computer Vision (ISCV) 2018 Apr 2 (pp. 1-4).
36. Serin F, Alisan Y, Kece A. Hybrid time series forecasting methods for travel time prediction. *Physica A: Statistical Mechanics and its Applications*. 2021
37. Narayanan S, Chaniotakis E, Antoniou C. Factors affecting traffic flow efficiency implications of connected and autonomous vehicles: A review and policy recommendations. *Advances in Transport Policy and Planning*. 2020 Jan 1;5:1-50.
38. Das D. Exploring the linkage between human factors and road geometric elements influencing the road traffic accidents on the national roads of South Africa. In Canadian Society of Civil Engineering Annual Conference 2021 May 26 (pp. 15-29). Singapore: Springer Nature Singapore.
39. <https://ucivil.ir/%D8%A2%D8%B4%D9%86%D8%A7%DB%8C%DB%8C-%D8%A8%D8%A7-%D8%AD%D8%AC%D9%85-%D8%AA%D8%B1%D8%A7%D9%81%DB%8C%DA%A9-%D8%AE%DB%8C%D8%A7%D8%A8%D8%A7%D9%86-%D8%A8%D9%87-%D9%87%D9%85%D8%B1%D8%A7%D9%87-%D9%85%D8%AB/>
40. Kim C, Park YS, Sang S. Spatial and temporal analysis of urban traffic volume. In ESRI International User Conference, San Diego, CA. Accessed October 2008 Aug (Vol. 10, p. 2013).
41. <https://www.google.com/maps>