

## برآورد شاخص زمان سفر با استفاده از روش‌های یادگیری ماشین

### مقاله علمی - پژوهشی

شهریار افندی‌زاده\*، استاد، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران  
نوید عمومی خورشیدی، دانشجوی کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران  
نوید کلانتری، دانش‌آموخته دکتری، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران

\*پست الکترونیکی نویسنده مسئول: zargari@iust.ac.ir

دریافت: ۱۴۰۰/۰۵/۲۰ - پذیرش: ۱۴۰۰/۱۱/۰۵

صفحه ۲۸-۱۳

### چکیده

زمان سفر و تغییرات آن از جنبه‌های مهم حمل و نقل هستند که به عنوان شاخصی کلیدی جهت ارزیابی عملکرد شبکه در برنامه‌ریزی حمل و نقل مورد استفاده قرار می‌گیرند. مطالعات نشان می‌دهد تراکم ترافیکی به عنوان عامل مهمی در غیرقابل اعتماد بودن زمان سفر، به دو دسته تکرار شونده و غیرتکرار شونده تقسیم می‌شود. برای برآورد قابلیت اطمینان زمان سفر تحت اثر تراکم، درک و فهم علت وقوع تغییر امری ضروری است. مطالعات نشان می‌دهد تصادفات ترافیکی، نواحی تعمیر و نگهداری، آب و هوا، نوسانات ترافیک، رویدادهای خاص، تجهیزات کنترل ترافیک و تنگناهای فیزیکی هفت عامل عمده در ایجاد تغییرات زمان سفر هستند. این مقاله به دنبال بررسی آن است که چگونه تغییرات هندسه معبر (تعداد خطوط)، تصادفات، حجم ترافیک و شرایط آب و هوا بر قابلیت اطمینان زمان سفر اثر می‌گذارند. بدین منظور، طیفی از روش‌های یادگیری ماشین جهت بررسی و مدل‌سازی شبکه معابر بزرگراهی ایالت ویرجینا آمریکا مورد استفاده قرار گرفت که شامل رگرسیون بردار پشتیبان، رگرسیون نزدیک‌ترین همسایه و رگرسیون درخت تصمیم بود. نتایج این بررسی نشان داد این ابزار می‌توانند تا میزان مناسبی، تغییرات متوسط زمان سفر را منعکس کنند. با توجه به شاخص‌های آماری استخراج شده از مدل‌ها (خطای جذر میانگین مربعات، ضریب تعیین) و شاخص پایداری (نسبت ضریب تعیین مجموعه آموزش و آزمون) رگرسیون نزدیک‌ترین همسایه به عنوان مدل برتر شناسایی شد. انجام تحلیل حساسیت بر روی مدل برتر، نحوه تغییرات متغیر وابسته (شاخص زمان سفر) در مقابل تغییرات متغیرهای مستقل (متغیر نماینده حجم ترافیک، متغیر نماینده تصادفات و متغیر نماینده شرایط جوی) را نشان داد. نتایج این تحلیل‌ها در قالب نمودارهایی ارائه شد که کمک شایانی به بیان نتیجه‌گیری تحقیق نمود.

واژه‌های کلیدی: قابلیت اطمینان، زمان سفر، یادگیری ماشین، رگرسیون نزدیک‌ترین همسایه

### ۱- مقدمه

ارزیابی شبکه نیز به صورت اعداد قطعی و مشخص قابل تعیین خواهند بود، ولی در شرایط واقعی پارامترها و متغیرهای عرضه و تقاضا را نمی‌توان به صورت قطعی تعیین کرد. میزان عرضه در شبکه‌های حمل و نقل می‌تواند به علت بروز حوادث و با رفتار رانندگان دچار تغییر شود. از طرفی میزان تقاضا نیز در

زمان سفر به عنوان یکی از فاکتورهای کلیدی در برنامه‌های کاربردی سیستم حمل و نقل و مدیریت موثر ترافیک و هدایت مسیر مسافران مطرح است. در یک شبکه‌ی حمل و نقلی هنگامی که پارامترهای عرضه و تقاضا به طور قطعی معلوم باشند، می‌توان گفت که شاخص‌های محاسبه شده برای

رویدادهای خاص، تجهیزات کنترل ترافیک و تنگنای فیزیکی. [Lu, 2017 & Yazici, Kamga and Singhal, 2013]

با توجه به هفت عامل ذکر شده، انگیزه اصلی پژوهش حاضر، ارزیابی مدلی هوشمند (مبتنی بر روش‌های هوشمند یادگیری ماشین) برای پیش‌بینی قابلیت اطمینان زمان سفر با استفاده از سه متغیر حجم، آب و هوا و تصادفات در سطح کلان (سطح برنامه‌ریزی) جهت دستیابی به ابزاری برای پیش‌بینی است که می‌تواند به یک ابزار تصمیم‌گیری مفید برای مدیران تبدیل شود. مطالعات پیشین، تعاریف مختلفی را جهت مدلسازی قابلیت اطمینان زمان سفر ارائه دادند. در گزارش ارائه شده توسط اداره بزرگراه‌های فدرال ایالات متحده، قابلیت اطمینان زمان سفر در قالب ثبات زمان سفر تعریف شد، هنگامی که به صورت روزانه یا در زمان‌های مختلف در طول یک روز اندازه‌گیری می‌شوند. [Texas Transportation Institute with Cambridge Systematics, 2006]

در سری گزارش‌های برنامه تحقیقاتی راهبردی بزرگراه‌ها<sup>۴</sup>، قابلیت اطمینان به صورت تغییر زمان سفر در یک بخش از بزرگراه در یک بازه زمانی شش ماهه تا یک ساله برای برش‌های زمانی مختلف روز و یا تغییر در زمان سفر برای یک سفر یکسان در روزهای متوالی تعریف شد. (سفر یکسان به هدف، مبدا، مقصد، زمان انجام، مُد و مسیر یکسان اشاره می‌کند). [Cambridge Systematics, Inc., 2013] اکثر مطالعات قبلی بر تاثیر یک متغیر موثر از قبیل شرایط جوی یا تصادفات ترافیکی متمرکز بودند و مطالعات به نسبت کمتری پیش‌بینی چندمتغیره قابلیت اطمینان زمان سفر را بررسی کردند، در حالی که هدف این مقاله بررسی تاثیر همزمان سه متغیر موثر بر شاخص زمان سفر به عنوان متغیر وابسته است.

الفتریدو و کوی در سال ۲۰۰۷ داده‌های ترافیک را در ۱۵ کیلومتر از بزرگراهی در فیادلفیا از طریق هشت موقعیت جمع‌آوری داده، گردآوری و به طور مداوم، سرعت جریان را برای چهار ماه ثبت کردند. با استفاده از این داده‌ها، مجموعه‌ای از مدل‌ها برای برآورد زمان سفر با تکنیک رگرسیون خطی بر مبنای ۴ فاکتور اصلی توسعه داده شد. این چهار فاکتور شامل تراکم، آب‌وهوا، کارگاه‌های تعمیر و نگهداری و تصادفات بودند. این مدل‌ها، زمان سفر مورد انتظار را برای هر سناریو تولید کرده و توزیع زمان سفر بر اساس فرکانس وقوع هر سناریو را به دست می‌آورند. مدل از متغیرهای ساختگی برای

طی ساعات مختلف روز و یا حتی روزهای هفته متغیر است. [Chepuri et al., 2018] قابلیت اطمینان زمان سفر به عنوان معیاری برای تعیین میزان تغییرپذیری زمان سفر تلقی می‌شود. تحقیقات نشان می‌دهند که این معیار نقش بسزایی در تصمیم‌گیری مسافران ایفا می‌نماید، یعنی مسافران همواره مسیر یا شیوه سفری را انتخاب کنند که زمان سفر آن دستخوش تغییرات اندکی باشد. [Rajabi et al., 2018] برآورد زمان سفر فاکتوری کلیدی و مهم برای ارزیابی بهره‌وری سیستم فعالیت‌ها است و استخراج زمان سفر دقیق و قابل اعتماد، تبدیل به امری ضروری برای مهندسان حمل و نقل، مسافران و غیره شده است. [Chepuri et al., 2018].

## ۲- قابلیت اطمینان و مدلسازی آن

نتایج مطالعات نشان می‌دهد تراکم ترافیکی فاکتور مهمی در غیرقابل اعتماد بودن زمان سفر است. تراکم ترافیک اغلب در دو دسته تراکم تکرارشونده<sup>۱</sup> و تراکم غیر تکرارشونده<sup>۲</sup> بررسی می‌شود. تراکم تکرارشونده قابل پیش‌بینی است و علت آن عبور منظم و تکراری ترافیک مازاد ظرفیت راه است. در طرف مقابل تراکم غیر تکرارشونده به واسطه رویدادهای تصادفی رخ می‌دهد، به طوری که ظرفیت یک راه با رویدادهایی نظیر تصادفات، کارگاه‌های تعمیر و نگهداری راه، شرایط نامساعد آب و هوایی و مواردی که تقاضای اوج به واسطه رویدادهای خاص بالاتر از سطح نرمال است، به طور موقت کم می‌شود. [Tavasoli Holjati et al., 2016] به بیان دیگر، تراکم تکرارشونده می‌تواند ناشی از تأخیر ساعات اوج روزانه باشد که به هنگام تجاوز تقاضا از ظرفیت یک بزرگراه رخ می‌دهد، در حالی که تراکم غیر تکرارشونده به واسطه رویدادهای غیرمعمول است، از قبیل تصادفات، نقص فنی خودروها، نواحی تعمیر راه، رویداد شرایط بد جوی و رویدادهای خاص. هر دو نوع تراکم، تغییرپذیری زمان سفر و عدم قطعیت را به رانندگان تحمیل می‌کنند. [Javid and Javid, 2017] برای برآورد قابلیت اطمینان زمان سفر تحت اثر تراکم تکرارشونده و غیر تکرار شونده، درک و فهم علل وقوع تغییر امری ضروری است. هفت علت عمده برای تغییرات زمان سفر توسط اداره بزرگراه‌های فدرال<sup>۳</sup> تعریف شده است. این هفت علت عبارتند از: تصادفات ترافیکی، نواحی تعمیر و نگهداری، آب و هوا، نوسانات ترافیک،

(تجهیزات کنترل ترافیک و ظرفیت پایه ناکافی مسیر) تقسیم‌بندی کردند. [Chen and Fan, 2019]. در سال ۲۰۱۲، ژائو و چن در یک نوشتار، تحلیل جامعی از اثرات شرایط نامساعد جوی بر سرعت و قابلیت اطمینان زمان سفر بین انواع مختلف معابر و بازه‌های زمانی مختلف در طول روز انجام دادند. نتایج پژوهش آن‌ها نشان داد در قالب بازه‌های زمانی، اثرات برف در بازه‌های اوج شدیدتر از بازه‌های غیر اوج است. [Zhao and Chien, 2012]. یازیچی و همکاران در سال ۲۰۱۳ به بررسی تأثیرات آب و هوا بر زمان سفر و قابلیت اطمینان آن در شهر نیویورک پرداختند. داده‌های شرایط آب و هوایی از سایت مخصوصی جمع‌آوری شده که شامل اطلاعات مرتبط مختلفی از قبیل درجه حرارت، سرعت باد و جهت آن، بارش و غیره بود. در این پژوهش، از روش طبقه‌بندی و درخت رگرسیون<sup>۵</sup> برای تحلیل داده‌ها استفاده شد و نتایج حاصل از آن به بررسی تغییرات میانگین زمان سفر، مُد زمان سفر و ضریب تغییرات زمان سفر تحت شرایط مختلف آب و هوایی (با طبقه‌بندی مشخص) پرداخت. از جمله یافته‌های این پژوهش آن بود که اثر شرایط جوی در طول بازه‌های زمانی مختلف روز ثابت نیست. شرایط بد جوی سبب افزایش بیشتر میانگین و میانه زمان سفر در بازه‌هایی می‌شود که دارای تراکم ترافیکی نیستند. همچنین نشان داده شد که اثرات شرایط آب و هوا با افزایش ازدحام شبکه کاهش می‌یابد. [Yazici, Kamga and Singhal, 2013]

رایت و همکاران در سال ۲۰۱۵ اثر انواع مختلف تصادف را بر قابلیت اطمینان زمان سفر با استفاده از داده‌های تصادف جمع‌آوری شده در دو ایالت سیاتل و واشنگتن مطالعه کردند و دریافتند که تصادفات می‌تواند به طور معنی‌داری قابلیت اطمینان زمان سفر را کاهش دهد. در سال ۲۰۱۶، جهانبخش جاوید در یک مقاله، چارچوبی برای اندازه‌گیری مقدار، پیش‌بینی و مشارکت دادن قابلیت اطمینان در فرآیند برنامه‌ریزی حمل و نقل پیشنهاد کردند. اندازه‌گیری قابلیت اطمینان سفرها بین زوج مبدأ-مقصد با استفاده از داده‌های مشاهده شده‌ی تجربی انجام شد. با تکیه بر فرضیاتی، زمان سفر لینک‌ها به زمان سفر مبدأ-مقصد تبدیل و انحراف معیار زمان سفر به عنوان شاخص قابلیت اطمینان محاسبه شد. قابلیت اطمینان بر مبنای مبدأ-مقصد که در این نوشتار معرفی شد، مهم و کاربردی است زیرا به راحتی می‌تواند در مدل‌های

تصادفات، نواحی تعمیر و نگهداری و شرایط آب‌وهوا استفاده می‌کرد. عیب عمده‌ی این مدل آن بود که تمام اثرات رویدادهای غیرمنتظره، به ویژه حوادث را پوشش نمی‌داد.

[Bonnaire Fils, 2012]

در سال ۲۰۰۷، مرکز تحقیقات حمل و نقل دانشگاه فلوریدا تحقیقات متعددی برای آزادراه‌ها و معابر شریانی در فلوریدا انجام داد. با استفاده از ۴ فاکتور تراکم، کارگاه‌های تعمیر و نگهداری راه، شرایط آب و هوا و تصادفات که ممکن است زمان سفر را تحت تأثیر قرار دهد، مدل‌هایی برای برآورد قابلیت اطمینان زمان سفر در تسهیلات آزادراهی توسعه داده شد. علاوه بر این، سه بخش زمان سفر (زمان سفر در حرکت، زمان انتظار در صف و زمان حرکت در صف) نیز به طور مجزا برآورد و سپس به منظور تخمین زمان سفر در معابر شریانی با هم ترکیب شدند. در سال ۲۰۱۱، پارک و همکاران، اثر تصادفات ترافیکی بر قابلیت اطمینان زمان سفر را بر مبنای مدل‌های شبیه‌سازی خرد، در قسمتی از بزرگراه I-66 به طول شانزده مایل بررسی کردند. سه سناریوی تصادفات ترافیکی به صورت انسداد خطوط از یک خط تا سه خط تعریف شد. نتایج نشان داد که قابلیت اطمینان زمان سفر کاهش قابل توجهی بعد از رخداد تصادفات ترافیکی داشت. علاوه بر این، اثر تصادفات بر قابلیت اطمینان زمان سفر در تراکم زیاد بحرانی‌تر بود. هرچند، تعدادی محدودیت در این مطالعه وجود داشت. اولاً، طول مدت تصادفات به طور ثابت چهل دقیقه فرض شد که احتمالاً فرضی غیرواقعی است. دوماً، شدت تصادف در قالب تعداد خطوط مسدود شده توسط تصادف تعریف شد، هرچند برای تصادفاتی که تعداد خطوط یکسانی را مسدود می‌کنند، اثر تراکم بر مبنای شدت خود تصادف متفاوت است. سوماً، نگرانی بزرگ‌تر در مورد استفاده از مدل‌های شبیه‌سازی برای تجزیه و تحلیل حوادث است. فرآیند کالیبراسیون نیاز به استفاده از شرایط واقعی حادثه داد تا نتایج واقعی حاصل شود. [Tavasoli Holjati et al., 2016].

کوون و همکاران در سال ۲۰۱۱ یک مدل رگرسیون خطی را برای مطالعه و پیش‌بینی قابلیت اطمینان زمان سفر تحت تأثیر چند فاکتور توسعه دادند. آنها فاکتورهای مورد نظر را در سه دسته رویدادهای ترافیکی تأثیرگذار (حوادث ترافیکی و تصادفات، فعالیت‌های تعمیر و نگهداری، آب و هوا و شرایط محیطی)، تقاضای ترافیکی (نوسانات در تقاضای روزانه‌ی ترافیکی و رویدادهای خاص) و خصوصیات فیزیکی راه

آنکه مطالعات اندکی به بررسی تاثیر همزمان چند متغیر پرداختند.

### ۳- روش تحقیق

در این نوشتار، مرور مختصری بر فرضیات و مطالب اصلی مرتبط با ماشین بردار پشتیبان<sup>۶</sup>، K- نزدیک‌ترین همسایه<sup>۷</sup> و درخت تصمیم<sup>۸</sup> ذکر خواهد شد.

### ۳-۱ ماشین بردار پشتیبان

ماشین بردار پشتیبان یکی از روش‌های یادگیری ماشین است که از آن برای طبقه‌بندی کردن داده‌ها استفاده می‌شود. با استفاده از این روش، داده‌های دارای برجسب، طبقه‌بندی شده و توسط الگویی مناسب، پیش‌بینی داده‌های جدید انجام می‌شود. در این روش با استفاده از همه باندها و یک الگوریتم بهینه‌سازی، نمونه‌هایی که مرزهای کلاس‌ها را تشکیل می‌دهند بدست می‌آید و با استفاده از آن‌ها یک مرز تصمیم‌گیری خطی بهینه برای جدا کردن کلاس‌ها محاسبه می‌شود. این نمونه‌ها را بردارهای پشتیبان می‌گویند. به نمونه‌های مرزی، بردارهای پشتیبان می‌گویند و برای محاسبه مرز تصمیم‌گیری دو کلاس کاملاً جدا از هم، از روش حاشیه بهینه استفاده می‌شود.

[Steinwart and Christmann, 2008]

برای بررسی مسائل رگرسیون، جایی که مقادیر خروجی یک عدد حقیقی (مقدار پیوسته) است از رگرسیون بردار پشتیبان استفاده می‌کنیم. در این فرم، تابع  $f$  برای الگوهای آموزش طوری ساخته می‌شود که از مقادیر آموزشی بیشترین حاشیه را داشته باشد. در حقیقت رگرسیون بردار پشتیبان به دنبال برآورد پارامترهای وزن و اریب تابعی است که این تابع بهترین تطبیق را بر مجموعه‌ی داده‌ها داشته باشد. اگر فرض شود که هر ورودی  $X$  دارای  $D$  ویژگی باشد و هر نقطه دارای مقدار  $Y$  مورد نظر باشد، هدف در این جا یافتن تابعی است که رابطه‌ای به شرح زیر بین ورودی‌ها و خروجی‌ها برقرار کند.

[Parrella, 2007]

$$f(x, w) = w^T x + b \quad (1)$$

برای محاسبه‌ی تابع  $f$  محاسبه‌ی مقادیر  $w$  و  $b$  ضروری است که آن‌ها نیز از کمینه‌سازی روابط زیر حاصل می‌شوند:

$$R(C) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l L_{\epsilon}(y_i, f_i(x, w)) \quad (2)$$

سفر مورد استفاده قرار گیرد. در واقع نوآوری اصلی این مطالعه، معرفی رویکرد قابلیت اطمینان مبتنی بر مبدأ-مقصد با توجه به داده‌های مشاهده شده تجربی بود تا در فرآیند برنامه‌ریزی مورد استفاده قرار گیرد. [Javid and Javid, 2017].

ژانگ و همکاران در سال ۲۰۱۹ با جمع‌آوری داده‌های مختلف در بازه‌های ۵ دقیقه‌ای از منطقه‌ای شهری در شمال کنتاکی و استفاده از روش داده‌کاوی درخت رگرسیون و طبقه بندی، به بررسی تاثیر شرایط آب هوایی برف و باران در زمان‌های مختلف سفر در روزهای مختلف پرداختند. در مطالعه آنها، ابتدا یک درخت تصمیم برای طبقه بندی زمان و روز بر اساس خصوصیات سرعت ساخته شد. سپس تاثیر رویدادهای مختلف جوی برای هر گروه با استفاده از رگرسیون چندگانه تعیین شد. تجزیه و تحلیل نتایج نشان داد تاثیر رویدادهای جوی بر قابلیت اطمینان زمان سفر به مراتب بیشتر از متوسط زمان سفر است. [Zhang and Chen, 2019].

چن و فن در سال ۲۰۱۹ در مقاله خود به مطالعه داده‌های زمان سفر معبری در کارولینای شمالی و تحلیل قابلیت اطمینان زمان سفر آن با استفاده از شاخص زمان برنامه‌ریزی پرداختند. آنها در بدست آوردن قابلیت اطمینان، شرایطی نظیر موقعیت معبر، آب و هوا، زمان در روز و روز در هفته را لحاظ کردند. سپس با استفاده از اطلاعات بدست آمده به آنالیز توزیع زمان سفر با استفاده از توزیع‌های لوگ نرمال، بور، گاما و ویبول پرداختند و سپس با استفاده از آزمون نیکویی برازش بهترین توزیع را با توجه به شرایط معرفی کردند. [Chen and Fan, 2019]

ژانگ و همکاران در ۲۰۲۰ پژوهشی به منظور مطالعه تاثیرات شدت بارش‌های مختلف بر قابلیت اطمینان زمان سفر انجام دادند. در مطالعه ایشان، شدت بارش بر حسب میزان بارش طی یک شبانه روز (۲۴ ساعت) به ۴ دسته کم، متوسط، زیاد و شدید تقسیم شد. آنها نتیجه گرفتند به طور کلی بارش باران قابلیت اطمینان زمان سفر را کاهش می‌دهد و این تاثیر کاهشی با افزایش شدت باران رشد می‌کند. [Zhang et al, 2020]

نگاهی اجمالی به پژوهش‌های انجام شده در این زمینه نشان می‌دهد اکثر مطالعات پیشین بر تاثیر یک متغیر منفرد تمرکز کردند یا از روش‌های گوناگونی به بررسی جنبه‌های مختلف قابلیت اطمینان زمان سفر پرداختند. این مطالعات بیان کردند تصادفات، شرایط آب و هوایی و حجم تقاضا و نوسانات آن تاثیر معنی داری بر قابلیت اطمینان زمان سفر داشتند، ضمن

فضای ویژگی‌ها (خصوصیات) ایجاد می‌کند که شاخص فاصله، تابعی است با اشکال مختلف مثل زیر.

$$d_M(x_{i1}, x_{i2}) = \sum_{i=1}^n |x_{i1} - x_{i2}| \quad (4)$$

$$d_E(x_{i1}, x_{i2}) = \sqrt{\sum_{i=1}^n (x_{i1} - x_{i2})^2} \quad (5)$$

که فاصله رابطه ۴، معروف به فاصله منهن و فاصله رابطه ۵ معروف به فاصله اقلیدسی است. انتخاب الگوریتم بهینه برای یک مجموعه داده انتخابی بستگی به عواملی از قبیل تعداد نمونه و ابعاد، ساختار داده، تعداد همسایه‌ها و تعداد نقاط پرس و جو دارد. رگرسیون‌های همسایه‌مبنا می‌توانند در مواردی استفاده شوند که برچسب‌های داده‌ها پیوسته هستند. برچسب اختصاص داده‌شده به نقطه‌ی پرس و جو بر مبنای میانگین برچسب‌های نزدیک‌ترین همسایگانش محاسبه می‌شود. دو نوع مختلف از رگرسورها وجود دارد که در این روش از آن استفاده می‌شود. رگرسور همسایه K یادگیری را بر مبنای K همسایه نزدیک هر نقطه‌ی پرس و جو اجرا می‌کند که K یک عدد صحیح مشخص شده توسط کاربر است. رگرسور همسایه شعاعی یادگیری را بر مبنای همسایه‌های که در یک شعاع ثابت r از نقطه پرس و جو قرار دارند، اجرا می‌کند که r یک مقدار شناور مشخص شده توسط کاربر است. [Urso et al., 2018]

### ۳-۳- الگوریتم درخت تصمیم

درخت‌های تصمیم، روش‌های یادگیری نظارت شده‌ای هستند که در طبقه‌بندی و رگرسیون کاربرد دارند. هدف این روش، ارائه مدلی‌ای است که مقدار متغیر هدف را با یادگیری قوانین تصمیم‌گیری ساده از خصوصیات داده‌ها پیش‌بینی می‌کند. [Urso et al., 2018] نتیجه پیاده سازی الگوریتم درخت تصمیم، مجموعه‌ای از شرط‌های منطقی به صورت "اگر... آنگاه..." با ساختار درختی است که برای پیش‌بینی یک ویژگی به کار می‌رود، طوری که داده‌هایی که در برگ‌های انتهایی این درخت تصمیم قرار می‌گیرند، توسط یکی از مقادیر ویژگی هدف برچسب می‌خورند. این مدل به دلیل سهولت در تفسیر نتایج و ناپارامتری و غیرخطی بودن، نیاز به پیش‌فرض رابطه خطی بین متغیرهای مستقل و وابسته ندارد. این الگوریتم به گونه‌ای عمل می‌کند که سعی دارد گوناگونی و یا تنوع

که در آن C پارامتری ثابت بوده و مقدار آن باید توسط کاربر مشخص شود. در واقع این پارامتر برای برقراری تعادل و تغییر وزن‌های مقدار جریمه ناشی از اغماض (متغیر E) و در عین حال بیشینه سازی حاشیه‌ی جدایی است. تابع  $L_E$  معرفی شده در رابطه‌ی فوق، تابع وینیک<sup>۹</sup> است که به صورت زیر تعریف می‌شود:

$$|y - f(x, w)|_E = \begin{cases} 0 & \text{if } |y - f(x, w)| < \epsilon \\ |y - f(x, w) - \epsilon| & \text{در غیر این صورت} \end{cases} \quad (3)$$

[Gunn, 19988]

### ۳-۲- الگوریتم K- نزدیک‌ترین همسایه

الگوریتم K- نزدیک‌ترین همسایه یکی از روش‌های هوش مصنوعی است که در سال ۱۹۵۱ توسط فیکس و هودجز ابداع شد. در حقیقت این روش مبتنی بر حافظه است که یک قانون ساده دارد: تمامی نمونه‌ها را ذخیره و نمونه‌های جدید را بر مبنای شباهت دسته‌بندی می‌کند. در واقع، با در نظر گرفتن مجموعه‌ی آموزش داده‌ی D، تابعی از فاصله d و یک عدد صحیح مثل k، برای یک نقطه جدید از x که باید تصمیم اتخاذ کند، الگوریتم در فضای D، k نقطه‌ی نزدیک x را که در فاصله‌ی d قرار دارند، جستجو می‌کند. [Yahyaoui, 2018] قاعده کلی روش‌های نزدیک‌ترین همسایگی یافتن تعداد از پیش تعیین شده از نمونه‌های آموزشی است که با نقطه جدید، کمترین فاصله را دارند و بر اساس این نمونه‌های آموزشی، برچسب‌ها را پیش‌بینی می‌کند. با وجود سادگی، روش نزدیک‌ترین همسایه در تعداد زیادی از مسائل طبقه‌بندی و رگرسیون موفق بوده است. [Urso et al., 2018] این روش نمونه ناشناس را با نمونه‌های متعلق به مجموعه‌ی آموزش مقایسه می‌کند و بر مبنای شباهت با نمونه‌های آموزش، آن را به طبقه‌ی خود اختصاص می‌دهد. مجموعه‌ی آموزشی نمونه‌های زیر را در نظر بگیرید:  $(X_i, y_i)$  که در آن  $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$  یک چندتایی توصیف شده توسط n ویژگی (خصوصیت) است و  $y_i$  برچسب کلاس متناظر است. یک چندتایی توسط یک نقطه در فضای n بعدی از خصوصیت‌ها نمایش داده می‌شود. یک نمونه‌ی جدید که قرار است طبقه‌بندی شود، یکی از برچسب‌های کلاس‌های K را که نزدیک‌ترین همسایه در فضای نمایش است، می‌گیرد. با استفاده از شاخص فاصله، طبقه‌بندی کننده همسایه‌ها را در

ویژگی‌های هدف را در گره‌ها به حداقل ممکن برساند. [Bagheri et al., 2014].

#### ۴- مجموعه داده‌ها

قبل از ورود به این بحث، بررسی مختصر فلوچارت شکل ۱ به درک بهتر فرآیند انجام شده در این پژوهش کمک خواهد کرد. در گام اول، داده‌های مربوط به متغیرهای ایجادکننده تراکم (که شامل حجم ترافیک، تصادفات و آب و هوا در این مطالعه بود) جمع‌آوری شدند. با استفاده از این داده‌ها و با توجه به اهداف تحقیق (ساخت مدل در سطح برنامه‌ریزی)، متغیرهای مستقل ساخته شدند (که در این مطالعه شامل متغیر KAL، متغیر CrR و متغیر Prc است). نوآوری این پژوهش، استفاده کاربردی از طیف وسیعی از روش‌های یادگیری ماشین جهت ارائه ابزاری برای پیش‌بینی تغییرات متغیر وابسته (متغیر شاخص زمان سفر) تحت تغییر هر کدام از متغیرهای مستقل است. پس از ساخت مدل‌های مختلف (بردار پشتیبان، نزدیک‌ترین همسایه و درخت تصمیم) شاخص‌های آماری نیکویی برازش (مانند خطای جذر میانگین مربعات و ضریب تعیین) برای این مدل‌ها استخراج و بر اساس این شاخص‌ها و پایداری مدل (نزدیک بودن کیفیت عملکرد مجموعه آموزش و آزمون با استناد به ضریب تعیین) مدل برتر انتخاب شد. تحلیل حساسیت با توجه به مدل منتخب گام بعدی پژوهش بود. هدف از انجام تحلیل حساسیت، بررسی میزان تغییرات متغیر وابسته در ازای تغییرات یک متغیر مستقل (یا چند متغیر به صورت همزمان) است. در گام آخر نیز نموداری جهت نمایش کیفی و عددی تغییرات شاخص زمان سفر بر اساس تغییرات متغیرهای ایجادشده در گام دوم ترسیم شد.



شکل ۱. فلوچارت مراحل انجام پژوهش

همان‌طور که در قسمت مروری بر ادبیات ذکر شد، هفت عامل در ایجاد تراکم و تغییرپذیری زمان سفر نقش دارند. در این مطالعه از شاخص زمان سفر<sup>۱۱</sup> برای نشان دادن تغییرپذیری و ثبات زمان سفر استفاده شد. این شاخص متوسط زمان یک سفر در ساعات اوج را با شرایط جریان آزاد، مقایسه می‌کند. برای محاسبه‌ی آن نیز باید زمان سفر متوسط را بر زمان سفر جریان آزاد تقسیم کرد. هدف از انجام، مطالعه و بررسی تاثیر سه متغیر حجم ترافیک، آب و هوا و تصادفات بر شاخص قابلیت اطمینان موجود (متغیر وابسته) است. تمامی داده‌های برای ۴ سال متوالی ۲۰۱۴، ۲۰۱۵، ۲۰۱۶ و ۲۰۱۷ در سطح برنامه‌ریزی جمع‌آوری شد. بدین منظور، مقادیر متغیرهای مستقل و وابسته برای ۱۱۸ قطعه بزرگراهی مورد مطالعه در ایالت ویرجینیا، برای ۲۴ ساعت شبانه روز و ۴ سال متوالی جمع‌آوری شد. مقادیر این شاخص‌ها با استفاده از مجموعه‌ی تجزیه و تحلیل داده‌ی Probe استخراج شده است.

برای بررسی تاثیرات متغیر مربوط به حجم ترافیک، از ترکیب متوسط سالیانه حجم ترافیک روزانه<sup>۱۲</sup> و تعداد خطوط مسیر استفاده شد. داده‌های مربوط به شمارش حجم ترافیک از سایت دپارتمان حمل و نقل ایالت ویرجینیا و تعداد خطوط عبوری مسیر نیز از Google maps اخذ شده است. لازم به ذکر است قطعات راه‌های مورد بررسی در این نوشتار از نوع بزرگراه‌های میان‌ایالتی بوده و تمامی دسترسی‌ها به این بزرگراه‌ها از نوع غیرهم‌سطح می‌باشد.

$$x_1 = \frac{AADT}{1000 * No Lanes} = KAL \quad (6)$$

اثر متغیر آب و هوا در قالب سه شاخص تعداد روزهای بارانی با شدت بارش بزرگتر یا مساوی ۰/۱ اینچ (DP۰۱)، تعداد روزهای بارانی با شدت بارش بزرگتر یا مساوی ۱ اینچ (DP۱۰) و میزان بارش (PRCP) بر حسب میلی‌متر دیده شد که از بین این سه متغیر، متغیر بارش باران (PRCP) بیشترین همبستگی خطی را با شاخص‌های قابلیت اطمینان زمان سفر داشت. این داده‌ها از سایت سازمان ملی اقیانوسی و جوی ایالات متحده<sup>۱۳</sup> برداشت شد.

$$x_2 = \frac{PRCP}{1000} = Prc \quad (7)$$

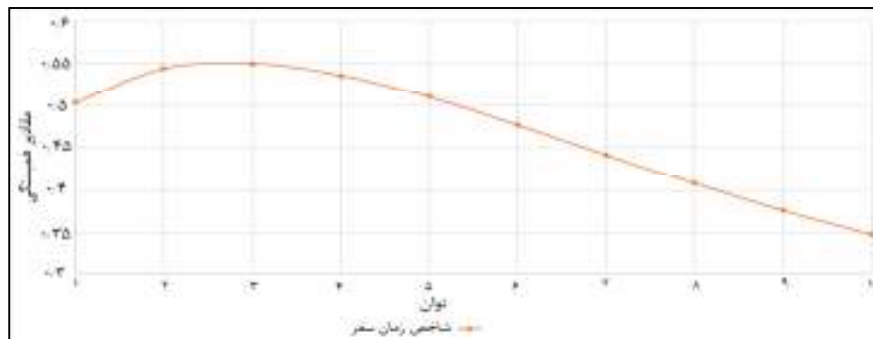
متغیر بعدی، تصادفات بود که آمار و ارقام آن از سایت تابلو عمومی<sup>۱۳</sup> اخذ شده است. برای متغیر تصادفات، تمامی قطعاتی

حسب نوع بر حاصل ضرب طول آن قطعه (Length) و متوسط سالیانه حجم ترافیک روزانه (AADT) تقسیم شد.

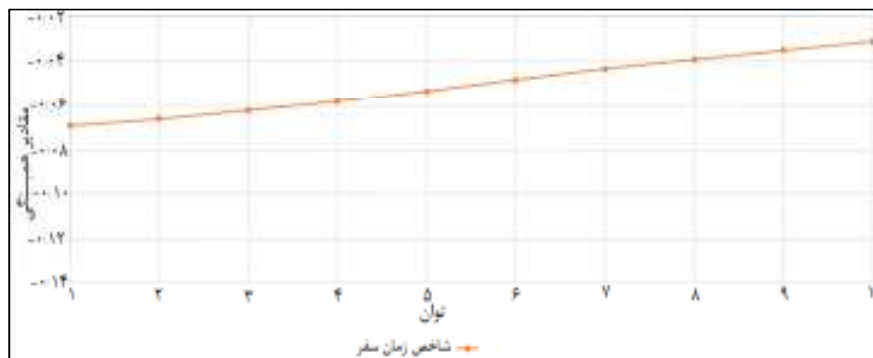
$$x_3 = \frac{No(A + B + C + K)}{AADT * Length} = CrR \quad (8)$$

از آنجا که متغیرهای مستقل به کار رفته در مدل می‌توانند بر حسب توان‌های مختلف بیان شوند، پس باید بررسی شود کدام توان از متغیرهای مستقل به کار رفته در مدل، بیشترین همبستگی را با متغیرهای وابسته دارد. بدین منظور، مقادیر ضریب همبستگی پیروسون به ازای مقادیر توان‌های از یک تا ده محاسبه شد و توانی از متغیر مستقل که دارای بیشترین همبستگی بود، انتخاب گردید. (شکل (۱)، شکل (۲)، شکل (۳)).

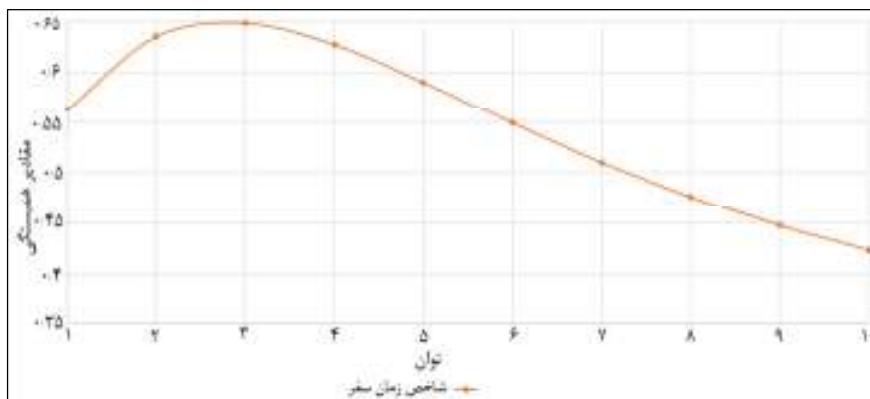
که برای آن‌ها شاخص‌های قابلیت اطمینان و حجم ترافیک محاسبه شد، وارد نرم‌افزار سیستم اطلاعات جغرافیایی شدند. بعد از ورود این نقشه‌ها به نرم‌افزار، تعداد تصادفات رخ داده در هر قسمت بر حسب نوع تصادف به تفکیک محاسبه شد. در طبقه‌بندی تصادفات در پایگاه داده، ۵ نوع تصادف فوتی (K)، جراحاتی شدید (A)، جراحاتی متوسط (B)، جراحاتی خفیف (C) و خسارتی (PDO) به تفکیک در نظر گرفته شده است که بر اساس این تعداد، نرخ تصادف محاسبه می‌شود. شیوه کار بدین صورت است که در صورت متغیر، ترکیبات خطی مختلف مربوط به نوع تصادفات اشاره شده در بالا لحاظ شدند و در نهایت، آن ترکیبی که بیشترین همبستگی با متغیر وابسته (شاخص زمان سفر) را داشت، لحاظ شد. برای محاسبه نرخ تصادف، تعداد تصادفات در هر قطعه از راه بر



شکل ۱. بررسی تاثیر توان متغیر KAL بر مقدار ضریب همبستگی



شکل ۲. بررسی تاثیر توان متغیر Prc بر مقدار ضریب همبستگی



شکل ۴. بررسی تاثیر توان متغیر CrR بر مقدار ضریب همبستگی

بردار پشتیبان انجام شد. در تابع کرنل خطی، میزان پارامتر C که در واقع پارامتر جریمه است، برابر ۱۰۰۰ انتخاب شد. در تابع کرنل چندجمله‌ای، میزان پارامتر جریمه مشابه حالت خطی و تابع چندجمله‌ای از درجه ۲ بود. در مورد علت انتخاب این میزان برای پارامتر جریمه، با بررسی مقادیر مختلف در محدوده ۱۰۰ تا ۱۰۰۰، این میزان برای پارامتر جریمه انتخاب شد. در تابع کرنل با پایه شعاعی نیز میزان پارامتر  $\gamma$  برابر ۰/۱ و میزان پارامتر جریمه مشابه حالت خطی و چندجمله‌ای است. جدول ۲ نتایج برآورد شاخص زمان سفر با رگرسیون بردار پشتیبان را نشان می‌دهد.

#### ۵-۲- برآورد شاخص زمان سفر با رگرسیون نزدیک‌ترین همسایه

در روش K-نزدیک‌ترین همسایه، مقدار K بر محاسبات ما تاثیرگذار است. روشی که در این بخش استفاده شد، ترسیم مقادیر مختلف خطای MSE بر حسب تعداد همسایه‌ها بوده است. آنچه که از شکل ۴ بر می‌آید، میزان خطا به ازای  $K=3$  کمینه می‌شود و این عدد مبنای محاسبات ما قرار می‌گیرد و نتایج برآورد در جدول ۳ ارائه شده است.

بر این اساس متغیر حجم و متغیر تصادفات با توان سوم دارای بیشترین همبستگی با متغیرهای وابسته هستند. جدول ۱ مقادیر ضریب همبستگی بین متغیرهای وابسته و مستقل را بیان می‌کند.

جدول ۱. مقادیر ضریب همبستگی پیرسون متغیرهای استفاده شده در مدل

متغیرها	TTI	$x_1^3$	$x_2$	$x_3^3$
TTI	۱	۰/۵۵	-۰/۰۷	۰/۶۵
$x_1^3$	۰/۵۵	۱	-۰/۰۸	۰/۲۲
$x_2$	-۰/۰۷	-۰/۰۸	۱	-۰/۰۵
$x_3^3$	۰/۶۵	۰/۲۲	-۰/۰۵	۱

#### ۵- ساخت مدل‌ها

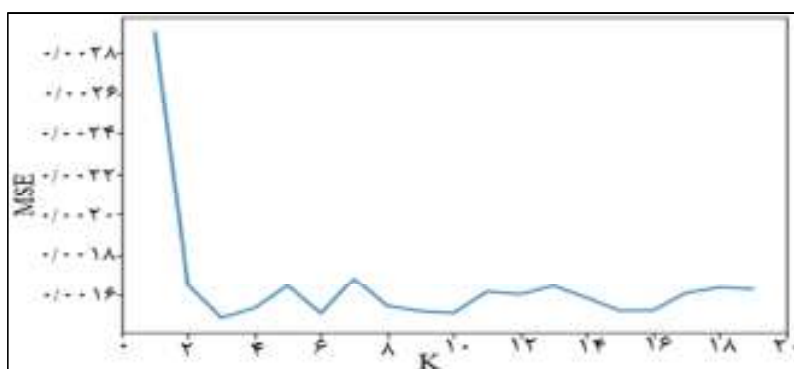
##### ۵-۱- برآورد شاخص زمان سفر با رگرسیون بردار پشتیبان

با استفاده از سه تابع کرنل به فرم‌های خطی، چندجمله‌ای و تابع پایه شعاعی، برآورد شاخص زمان سفر توسط رگرسیون



جدول ۲. نتایج برآورد شاخص زمان سفر با رگرسیون بردار پشتیبان

نوع مجموعه	نوع تابع کرنل	تعداد نمونه ها	MSE	RMSE	ضریب تعیین
آموزش	خطی	۳۳۰	۰/۰۰۳۸۲۷	۰/۰۶۱۸۶۵	۰/۶۲۷۷
آزمون	خطی	۱۴۲	۰/۰۰۲۹۸۵	۰/۰۵۴۶۳۷	۰/۳۵۷۴
آموزش	چندجمله‌ای	۳۳۰	۰/۰۰۰۴۳	۰/۰۶۵۸۶	۰/۵۷۳۸
آزمون	چندجمله‌ای	۱۴۲	۰/۰۰۰۳۳	۰/۰۵۷	۰/۲۰۳۳
آموزش	پایه شعاعی	۳۳۰	۰/۰۰۰۳۶	۰/۰۶۰۳	۰/۶۴۵
آزمون	پایه شعاعی	۱۴۲	۰/۰۰۰۳۹	۰/۰۶۲	۰/۳۳۲۳



شکل ۵. مقادیر خطای MSE به ازای مقادیر K برای شاخص زمان سفر

جدول ۳. نتایج برآورد شاخص زمان سفر با رگرسیون نزدیک‌ترین همسایه

نوع مجموعه	تعداد نمونه‌ها	MSE	RMSE	ضریب تعیین
آموزش	۳۳۰	۰/۰۰۰۱۲	۰/۰۰۳۴۸۱	۰/۸۵۳۶
آزمون	۱۴۲	۰/۰۰۰۱۴۸۷	۰/۰۰۳۸۵۶	۰/۵۷۰۷

(تاثیر این دو پارامتر از بقیه‌ی پارامترها بیشتر بود. نگارنده تغییراتی بر اکثر پارامترهای رگرسیون درخت تصمیم اعمال کرد و در نهایت به این نتیجه رسید که این دو پارامتر موثرتر هستند). کمینه‌ی مقدار MSE به ازای حداکثر عمق برابر با ۶ و مقدار وضعیت تصادفی ۱۲ بدست آمد. جدول ۴ نیز نتایج برآورد شاخص زمان سفر توسط درخت تصمیم را نشان می‌دهد.

### ۵-۳- برآورد شاخص زمان سفر با رگرسیون درخت

#### تصمیم

در این روش، پارامترهای حداکثر عمق و وضعیت تصادفی از جمله پارامترهای موثر بر نتایج هستند. با ترسیم مقادیر خطای ناشی از تغییرات پارامترهای حداکثر عمق و وضعیت تصادفی در فضای سه بعدی، مقادیری از این دو پارامتر که منجر به کمینه شدن خطا می‌شوند، شناسایی شد.

## ۶- مقایسه عملکرد مدل‌ها

آزمونی نزدیک به یک نیز باشد. با توجه به این ملاحظه، باید در انتخاب بهترین مدل با توجه به این موضوع تجدید نظر کرد. جدول ۵ بر اساس خطای میانگین مربعات و ضریب تعیین و جدول ۶ بر اساس ملاحظات ذکر شده به رتبه‌بندی نتایج به دست آمده می‌پردازد. شکل (۶) و (۵) نیز به طور کیفی نتایج مقادیر برآورد شده را در مقابل مقادیر حقیقی شاخص‌ها به تصویر می‌کشد.

در بحث یادگیری ماشین، با توجه به این که دو مجموعه‌ی آموزش و آزمون دو ضریب تعیین جداگانه دارند، در انتخاب بهترین مدل، علاوه بر بالا بودن ضریب تعیین، لازم است مقادیر ضریب تعیین مجموعه آزمون و مجموعه آموزش به هم نزدیک باشند. به بیان دیگر، مدل برتر باید علاوه بر دارا بودن مقادیر خطا و ضریب تعیین مناسب، دارای مقادیر نسبت ضریب تعیین مجموعه آموزش به ضریب تعیین مجموعه

جدول ۴. نتایج برآورد شاخص زمان سفر با رگرسیون درخت تصمیم

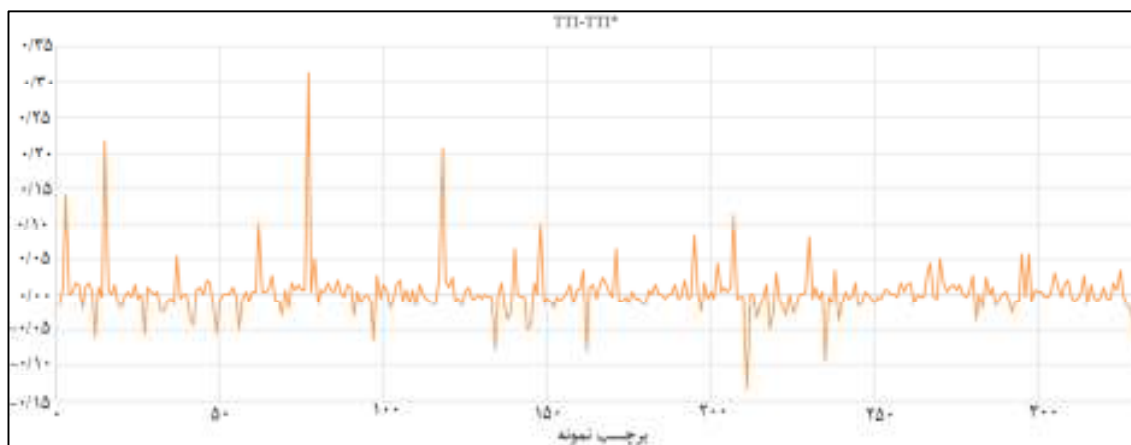
نوع مجموعه	تعداد نمونه ها	MSE	RMSE	ضریب تعیین
آموزش	۳۳۰	۰/۰۰۰۵۹۶	۰/۰۲۴۴	۰/۹۱۸۹
آزمون	۱۴۲	۰/۰۰۲۳	۰/۰۴۸۴	۰/۴۲۵۷

جدول ۵. رتبه‌بندی براساس خطای میانگین مربعات و ضریب تعیین

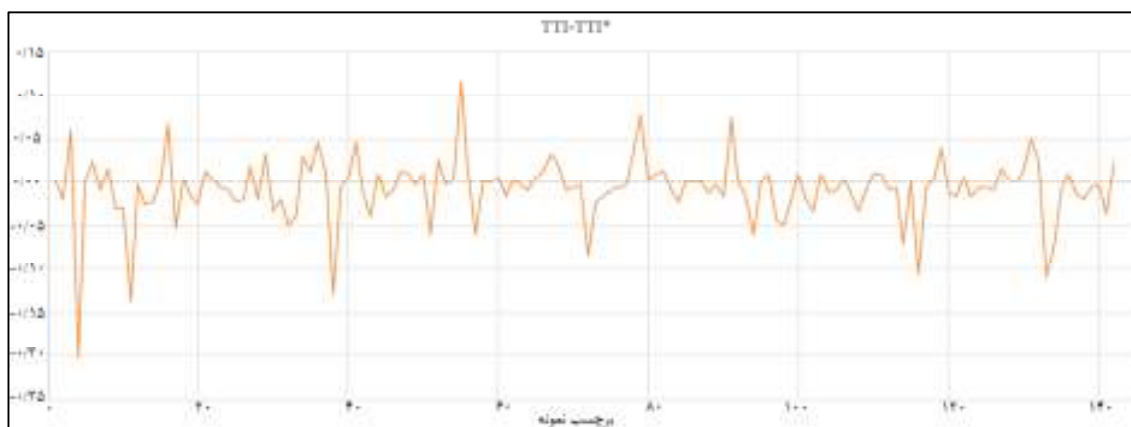
آزمون		آموزش				نوع مدل
رتبه	R <sup>2</sup>	رتبه	MSE	رتبه	R <sup>2</sup>	
۳	۰/۳۵۷۴	۳	۰/۰۰۲۹۸۵	۴	۰/۶۲۷۷	رگرسیون بردار پشتیبان خطی
۵	۰/۲۰۳۳	۴	۰/۰۰۳۳	۵	۰/۵۷۳۸	رگرسیون بردار پشتیبان چندجمله‌ای
۴	۰/۳۳۲۳	۵	۰/۰۰۳۹	۳	۰/۶۴۵	رگرسیون بردار پشتیبان پایه شعاعی
۱	۰/۵۷۰۷	۱	۰/۰۰۱۴۸۷	۲	۰/۸۵۳۶	رگرسیون نزدیک‌ترین همسایه
۲	۰/۴۲۵۷	۲	۰/۰۰۲۳	۱	۰/۹۱۸۹	رگرسیون درخت تصمیم

جدول ۶. رتبه‌بندی براساس خطای میانگین مربعات و ضریب تعیین

رتبه	نسبت ضریب تعیین آموزش به آزمون	نوع مدل
۴	۲/۱۶	رگرسیون درخت تصمیم
۱	۱/۴۷	رگرسیون نزدیک‌ترین همسایه
۳	۱/۹۹	رگرسیون بردار پشتیبان پایه شعاعی
۵	۲/۸۲	رگرسیون بردار پشتیبان چندجمله‌ای
۲	۱/۷۵	رگرسیون بردار پشتیبان خطی



شکل ۶. تفاضل مقادیر شاخص از مقادیر پیش‌بینی در مجموعه آموزش شاخص زمان سفر در رگرسیون KNN

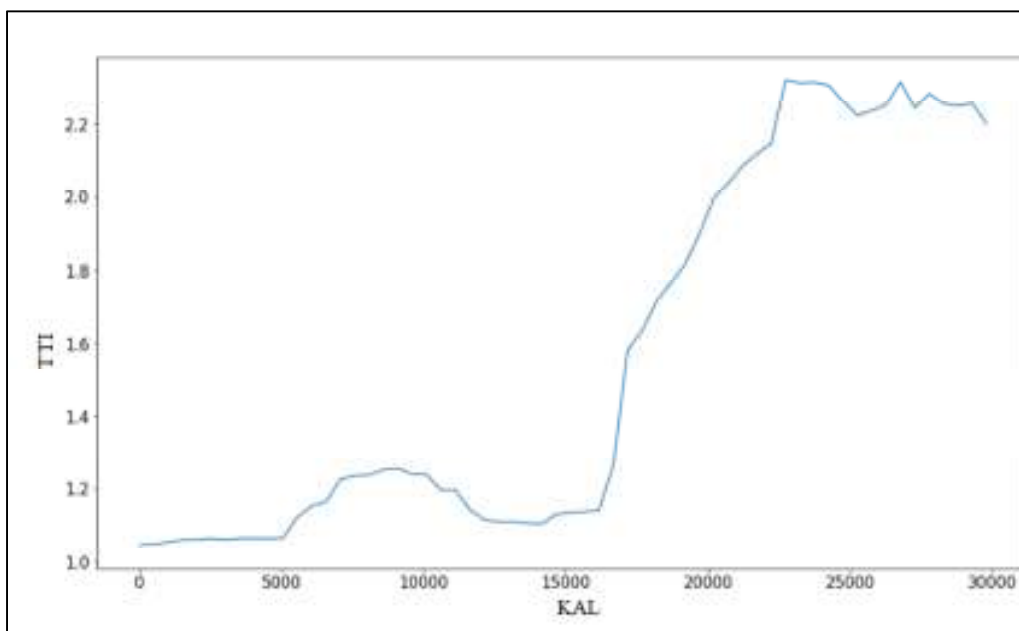


شکل ۷. تفاضل مقادیر شاخص از مقادیر پیش‌بینی در مجموعه آزمون شاخص زمان سفر در رگرسیون KNN

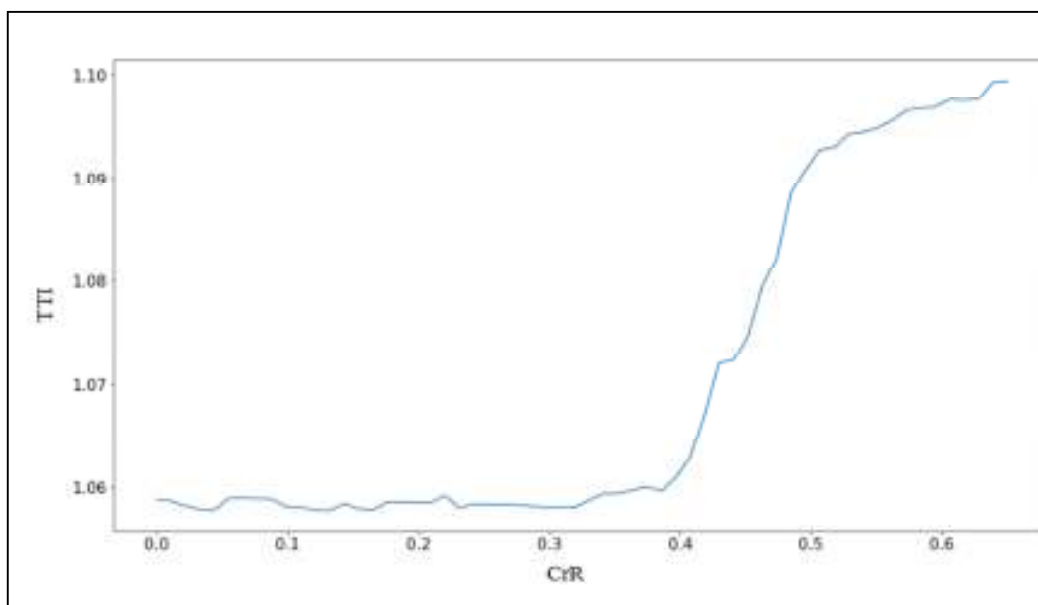
## ۷- تحلیل حساسیت و کاربرد مدل

همان‌طور که ذکر شد، هدف این مقاله فراهم کردن ابزاری جهت برآورد شاخص زمان سفر در سطح شبکه به واسطه تغییر در هندسه معبر، شرایط جوی و تصادفات است. روش‌های یادگیری ماشین اغلب به دلیل سخت بودن تفسیرشان مورد انتقاد قرار می‌گیرند. برای ارزیابی تأثیر هر یک از متغیرها بر شاخص زمان، آنالیز حساسیت انجام شد. به عبارت دیگر، تغییر متغیر وابسته در ازای تغییرات متغیرهای مستقل دیگر مورد ارزیابی قرار گرفت. تحلیل حساسیت برای هر متغیر با ثابت نگه داشتن مقدار سایر متغیرها در مقدار میانه آن‌ها انجام شد. برای نمایش نتایج، از نمودارهای ۸ و ۹ و ۱۰ برای نشان دادن نتایج تجزیه و تحلیل حساسیت هر متغیر مستقل استفاده شد.

شکل ۸ تغییرات شاخص زمان سفر به ازای تغییرات KAL را نشان می‌دهد. همان‌طور که در شکل نشان داده شده، با رسیدن مقادیر KAL به ۱۵۰۰۰ خودرو بر روز به ازای هر خط عبور، شاخص زمان سفر با سرعت بیشتری افزایش می‌یابد و با رسیدن به مقدار تقریبی ۲۳۰۰۰ خودرو بر روز به ازای هر خط عبور، منحنی به حالت افقی میل می‌کند. شکل ۹ نیز روند تغییرات متغیر شاخص زمان سفر در ازای تغییرات نرخ تصادف CrR را نشان می‌دهد. همان‌طور که از شکل بر می‌آید، افزایش نرخ تصادفات سبب افزایش شاخص می‌شود که با رسیدن میزان TCr به ۰/۴۵، سرعت رشد شاخص افزایش می‌یابد.



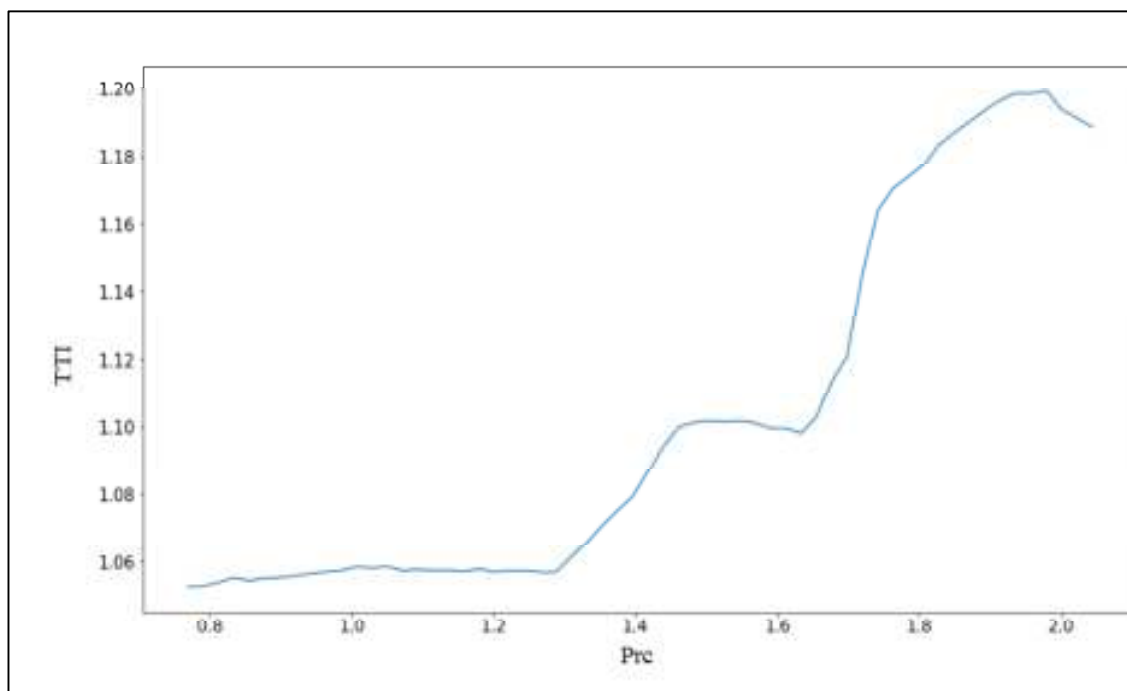
شکل ۸. تغییرات شاخص زمان سفر در قبال تغییرات متغیر KAL



شکل ۹. تغییرات شاخص زمان سفر در قبال تغییرات متغیر CrR

که در شکل نشان داده شده، مقادیر شاخص با رسیدن مقدار متغیر به ۱/۱ دچار افزایش شده و این افزایش با رسیدن مقدار متغیر مستقل بارش به ۱/۶۵ مجدداً تکرار می‌شود.

شکل ۱۰ نیز تاثیرات میزان بارش بر تغییرات متغیر وابسته را نشان می‌دهد. اگرچه برنامه‌ریزان حمل و نقل هیچ کنترلی بر سطح بارندگی ندارند، تجزیه و تحلیل تأثیر بارش بر شاخص زمان سفر می‌تواند حاوی اطلاعاتی باشد. همان‌طور



شکل ۱۰. تغییرات شاخص زمان سفر در قبال تغییرات متغیر Prc

با استناد به شاخص‌های آماری (ضریب تعیین و خطای جذر میانگین مربعات) و در نظر گرفتن دقت عملکرد دو مجموعه آموزش و آزمون (محاسبه نسبت ضریب تعیین مجموعه آموزش و آزمون) روش KNN بهترین عملکرد را داشت. پس از تعیین مدل برتر، با انجام تحلیل حساسیت و استخراج نمودارهای حساسیت، میزان تغییرات شاخص زمان سفر (متغیر وابسته) نسبت به تغییرات متغیرهای مستقل مشخص شد که این اشکال مبین نتایج این پژوهش است. همان‌طور که از شکل ۸ برمی‌آید، هنگامی که مقادیر KAL به ۱۵۰۰۰ خودرو بر روز به ازای هر خط عبور می‌رسد، شاهد افزایش TTI هستیم و این افزایش تا مقدار تقریبی ۲۳۰۰۰ خودرو بر روز به ازای هر خط عبور ادامه دارد و پس از آن منحنی به حالت افقی میل می‌کند. تفسیر این نتیجه آن است که در این بازه (۲۳۰۰۰-۱۵۰۰۰) هنگامی که متغیر KAL (لازم به ذکر است که این متغیر با توان سوم در معادلات حضور دارد، رجوع شود به جدول ۱) ۱۵٪ افزایش می‌یابد (یعنی در صورتی که هندسه معبر ثابت بماند و AADT به میزان ۱۵٪ افزایش یابد)، مقادیر شاخص زمان سفر به میزان ۸۳٪ افزایش می‌یابد. با فرض آن که تعداد خطوط نماینده مناسبی برای ظرفیت یک معبر است، در صورتی که ظرفیت یک معبر ۱۵٪ کاهش داشته باشد،

از نقطه نظر عملی در این مقاله نشان داده شد که TTI وقتی AADT در هر روز بین ۱۵۰۰۰ تا ۲۳۰۰۰ وسیله نقلیه در روز در هر خط باشد، هنگامی که نرخ کل تصادفات بیشتر از ۰/۴۵ باشد و زمانی که میزان بارش سالانه بین ۱۱۰۰ تا ۱۵۰۰ میلی‌متر در سال باشد (Prc بین ۱/۱ تا ۱/۵) حساس‌تر می‌شود.

#### ۸- نتیجه‌گیری

مطالعات بسیاری تاثیر یک عامل از عوامل هفت گانه ایجاد تراکم بر تغییرات زمان سفر را بررسی کردند که عمده‌ترین این بررسی‌ها شامل حضور متغیر تصادفات، حجم و شرایط آب و هوایی بود اما بررسی تاثیر همزمان متغیرهای مذکور تنها به چند مطالعه و استفاده از ابزارهای تحلیلی نظیر رگرسیون خطی محدود شده است. این مطالعه با هدف ارایه یک ابزار سطح برنامه ریزی (با استفاده از روش‌های یادگیری ماشین) برای تجزیه و تحلیل اثر تغییر همزمان هندسه معبر (تعداد خطوط)، تغییرات تصادفات، حجم ترافیک و شرایط جوی بر شاخص زمان سفر انجام شده است. در این مقاله روش‌های مختلف یادگیری ماشین برای تحلیل قابلیت اطمینان زمان سفر در معابر توسط شاخص زمان برنامه ریزی مورد آزمایش قرار گرفت.

#### ۹- پی‌نوشت‌ها

1. Recurring congestion
2. No-recurring congestion
3. Federal Highway Administration (FHWA)
4. Strategic Highway Research Program (SHRP)
5. Classification and Regression Tree (CART)
6. Support Vector Machine (SVM)
7. K-Nearest Neighbor (KNN)
8. Decision tree
9. Vapnik function
10. Travel Time Index (TTI)
11. Annual Average Daily Traffic (AADT)
12. National Oceanic and Atmospheric Administration (NOAA)
13. Public tableau

مجدداً شاهد افزایش شاخص زمان سفر به میزان ۸۳٪ خواهیم بود. برای بررسی نتایج متغیر تصادفات به شکل ۹ رجوع می‌کنیم. هنگامی که نرخ تصادفات از ۰/۴ به ۰/۶ می‌رسد، (نرخ تصادفات شامل متغیر تعداد تصادفات کلی، طول قطعه و AADT است) شاخص زمان سفر نیز از ۱/۰۶ به ۱/۱۰ افزایش می‌یابد. این مورد بدان معنی است که در صورت ثابت ماندن AADT یک قطعه در این بازه (۰/۶ - ۰/۴) افزایش ۵۰ درصدی تعداد تصادفات تنها سبب ۳٪ افزایش شاخص می‌شود

#### ۱۰- مراجع

- Regression", ISIS Technical Report, 14(1), pp.5-16.
- Javid, R. J. And Javid, R. J., (2017), "A Framework for Travel Time Variability Analysis Using Urban Traffic Incident Data", IATSS Research: International Association of Traffic and Safety Sciences, Vol.42, No. 1, Pp.30-38.
- Lu, Chaoru, (2017), "Estimate Freeway Travel Time Reliability under Recurring and Nonrecurring Congestion", Ph.D. Dissertation, Major Professor: Jing Dong, Ames: Department of Civil, Construction, and Environmental Engineering, Iowa State University.
- Parrella, Francesco, (2007), "Online Support Vector Regression", Master Dissertation, Genoa: Department of Information Science, University of Genoa.
- Rajabi, M., Khademi, N., Foroozandeh, R., Shariat, A., (2018), "Analysis of Bus Travel Time Reliability Using GPS Data", Quarterly Journal of Transportation Engineering, Vol.9, No. 3, pp.295-312.
- Steinwart, Ingo; Christmann, Andreas (2008), "Support Vector Machines", New York: Springer.
- Tavasoli Hojati, A., Ferreira, L., Washington, S., Charles, P And Shobeirnejad, A., (2016), "Modelling The Impact of Traffic Incidents
- Bagheri, F., Alizadeh, H., Mehrbakhsh, Z and Ziaratban, M., (2014), "Use of Data Mining Algorithms in Assessing the Affecting Factors on Predicting the Health Status of Newborns", Jorjani Biomedicine Journal, In Persian, Vol. 2, No.2, pp. 68-59.
- Bonnaire Fils, Prony, (2012), "Modeling Travel Time and Reliability on Urban Arterials for Recurrent Conditions", Ph.D. Dissertation, Major Professor: Jian John Lu, Tampa: Department Civil and Environmental Engineering, South Florida University.
- Cambridge Systematics, Inc., (2013), "Analytical Procedures for Determining the Impact of Reliability Mitigation Strategies", Washington, TRB.
- Chen, Z., Fan, W., (2019), "Data Analytics Approach for Travel Time Reliability Pattern Analysis and Prediction", J. Mod., Transport, 27, pp.250-265.  
<https://doi.org/10.1007/S40534-019-00195-6>.
- Chepuri, A., Borakanavar, M., Amrutsamanvar, R., Shriniwas, A., and Guarang, J., (2018), "Examining Travel Time Reliability Under Mixed Traffic Conditions: A Case Study of Urban Arterial Roads in Indian Cities", Asian Transport Studies, Vol. 5, No.1, pp.30-46.
- Gunn, S. R., (1998), "Support Vector Machines for Classification and

Travel Time And Travel Time Variability In New York City ", Transportation Research Board 92<sup>nd</sup> Annual Meeting, 13-17 January: Washington DC.

-Zhang, Q, Chen, H, Liu, H, Li, W and Zhang, Y., (2020), "Origin-Destination-Based Travel Time Reliability Under Different Rainfall Intensities: An Investigation Using Open-Source Data", Journal of Advanced Transportation.

-Zhang, X. And M. Chen, (2019), "Quantifying the Impact of Weather Events on Travel Time and Reliability", Journal of Advanced Transportation, pp. 1-9.

-Zhao, Linhui and Chien, I. Jy, (2012), "Analysis of Weather Impact on Travel Speed and Travel Time Reliability", 12th COTA International Conference of Transportation Professionals, 3-6 August, Beijing.

on Travel Time Reliability", Transportation Research Part C: Emerging Technologies, Vol.65, pp. 49-60.

-Texas Transportation Institute with Cambridge Systematics, (2006), "Travel Time Reliability: Making It There on Time, All the Time", Washington, US Department of Transportation, Federal Highway Administration.

-Urso, Alfonso, Fiannaca, Antonio, La Rosa, Massimo, Ravi, Valentina, And Rizzo, Riccardo, (2018), "Data Mining: Prediction Methods", Encyclopedia Of Bioinformatics And Computational Biology, Elsevier, Oxford.

-Yahyaoui, I., (2018), "Advances in Renewable Energies and Power Technologies, Vol.1, Elsevier.

-Yazici, M Anil., Kanga Camille And Singhal Abhishek, (2013), "Weather's Impact on

# Estimation of Travel Time Index Using Machine Learning Methods

*Shahriar Afandizadeh, Professor, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.*

*Navid Amoei khorshidi, M.Sc., Stud., School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.*

*Navid Kalantari, Ph.D., Grad., School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.*

*E-mail: zargari@iust.ac.ir*

Received: August 2021-Accepted: February 2022

## ABSTRACT

Travel Time and Its Variation Are Among Important Aspects of Transportation That Are Used as Key Indicators to Evaluate Network Performance in Transportation Planning. Studies Show That Traffic Congestion Is an Important Factor Effecting Travel Time Reliability and Is Divided into Two Categories: Recurrent and Non-Recurrent. Studies Show That Traffic Incidents, Work Zones, Weather, Demand Fluctuations, Special Events, Traffic Control Devices and Bottlenecks Are the Seven Major Factors Which Cause Travel Time Changes. This Article Aims to Examine How Changes in Transit Geometry (Number of Lanes), Accidents, Traffic Volume, And Weather Conditions Can Affect Travel Time Reliability. In Order to Do So, A Variety of Machine Learning Methods Were Used to Study and Model the Virginia Highway Network, Including Support Vector Regression, Nearest Neighbor Regression, And Decision Tree Regression. The Results of This Study Showed That These Tools Can Reflect the Changes in The Average Travel Time to An Appropriate Extent. Among These Methods, Performance of the Nearest Neighbor Was the Best (Coefficient of Determination 0.85 and Error 0.0012 for Training Set and Coefficient of Determination 0.57 and Error 0.0014 for Test Set and Stable Performance Equal To 1.47).

**Keywords:** Reliability, Travel Time, Machine Learning, Nearest Neighbor Regression