



Travel Time Modelling of Urban Roads By Application of Coyote Optimization-based Machine Learning Method

A. Golroo * , A. H. Fani, H. Naseri

Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran.

ABSTRACT: Travel time prediction as an essential issue has been scrutinized in recent decades. To this end, various techniques are applied to estimate travel duration in dynamic networks and intelligent transportation systems. Accordingly, in this investigation, the prediction of travel time is considered by machine learning techniques. Initially, the experimental test is planned, and the travel time effective parameters are spotted. Subsequently, with the assistance of the floating car method, and My-tracks application, the data are collected in six elected roads. After data preparation, stop delay, grades, and the number of the lane are determined as the most effective travel time criteria. In this study, a novel machine learning technique based on the coyote optimization algorithm is introduced, and its precision is compared with five conventional regression models. Drawing on results, the accuracy of the coyote optimization algorithm-based machine learning technique is more than that of other prediction methods. The coefficient of determination of the introduced machine learning technique for training and testing data is equal to 0.746 and 0.724, respectively. Furthermore, coyote optimization algorithm-based machine learning estimates 73% of testing data with an error of fewer than 20 seconds. .

Review History:

Received: Feb. 25, 2020
Revised: May, 17, 2020
Accepted: May, 17, 2020
Available Online: May, 28, 2020

Keywords:

Travel time prediction
Urban road
Regression
Machine learning
Coyote optimization algorithm

1- Introduction

Travel time prediction is one of the essential parts of informing drivers. This system helps travelers to select an appropriate path, which leads to traveling duration reduction [1]. Accordingly, various machine learning techniques, including regressions, decision trees, artificial neural networks, time series, have been applied to estimate the travel time duration and to reduce the traffic in the pavement networks.

Sil et al. [2] stated that geometric characteristics of roads and the number of lanes significantly impact the vehicle's speed, and accordingly, they have to be considered in travel time prediction [2]. Vilarinho et al. [3] analyzed the influences of traffic lights on travel duration, and the results indicated that stop delay is a vital criterion in order to evaluate the travel time.

As can be seen from the above references, the novel machine learning techniques have not received enough attention in this field. Regarding the literature, essential parameters may not be considered in a model simultaneously—accordingly; a new machine learning method was introduced in this study. Moreover, various parameters are scrutinized, and the most important criteria are taken into consideration in the prediction model.

2- Methodology

2- 1- Data collection

The data were collected based on the floating car method, and this method is applied according to the procedure presented in NCHRP Handbook [4]. In this regard, Mytacks application is utilized so as to collecting data and saving them. The case study comprises nine sections, and each section is approximately 1 km.

2- 2- Parameter selection

Initially, stop delay, the number of lanes, availability of adjacent parking, grade, the number of speed bumps are taken into consideration as effective parameters. The enter method is used to analyze these parameters. The results show that stop delay, grade, and the number of lanes are the most critical parameters, and accordingly, they are applied to generate prediction models.

2- 3- Modeling

The coyote optimization algorithm-based machine learning algorithm (COA) is introduced in this investigation. That is to say, by virtue of the coyote optimization algorithm, a novel machine learning technique is developed. To this end, the coyote optimization algorithm is adjusted to solve integer programming. Afterward, the mean absolute error is set the objective function of the optimization algorithm.

*Corresponding author's email: agolroo@aut.ac.ir



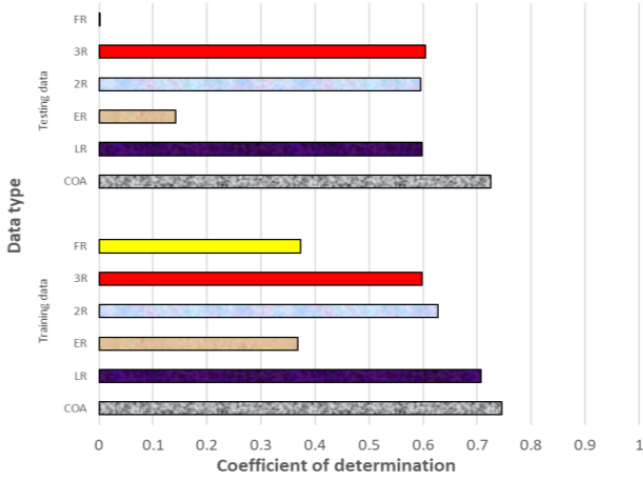


Fig. 1. The coefficient of determination

Eight thousand various modes are considered for each input. Consequently, the model is run ten times, the formula containing the lowest mean absolute error is considered the prediction model.

The introduced model is compared with five conventional regression models, including linear regression, fractional regression, exponential regression (ER), 2nd polynomial regression (2R), and 3rd polynomial regression (3R). Mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination, and the percentage of data which their error is fewer than 20 seconds (E20s) are applied to compare the introduced model (COA) with five conventional regression models. The equation travel time prediction presented by COA, LR, ER, 2R, 3R, and FR is indicated in Eq. (1) to (6).

$$COA - Time(s / km) = (0.7 \times \sqrt[3]{3} \times StopDelay)^2 + (0.7 \times \sqrt{6} \times \sqrt{Lane})^2 + (0.8 \times Grade^2)^2 + (0.7 \times \sqrt[3]{3} \times StopDelay)^3 + (0.7 \times \sqrt{6} \times \sqrt{Lane})^3 + (0.8 \times Grade^2)^3 + 3.587401 \quad (1)$$

$$LR - Time(s / km) = (1.3688 \times StopDelay) + (10.1461 \times Lane) + (14.3076 \times Grade) + 13.3413 \quad (2)$$

$$ER - Time(s / km) = (0.0001 \times \exp(StopDelay)) + (0.4291 \times \exp(Lane)) + (4.9904 \times \exp(Grade)) + 34.0197 \quad (3)$$

$$2R - Time(s / km) = (0.0598 \times StopDelay^2) + (1.8666 \times Lane^2) + (8.7428 \times Grade^2) + 24.7024 \quad (4)$$

$$3R - Time(s / km) = (0.0005 \times StopDelay^3) + (0.2976 \times Lane^3) + (4.1589 \times Grade^3) + 38.4088 \quad (5)$$

$$FR - Time(s / km) = (-92.8259 \times \frac{StopDelay}{Lane \times Grade}) + 59.9815 \quad (6)$$

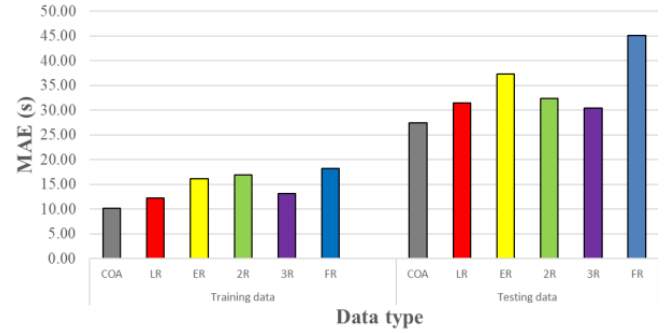


Fig. 2. The mean absolute error

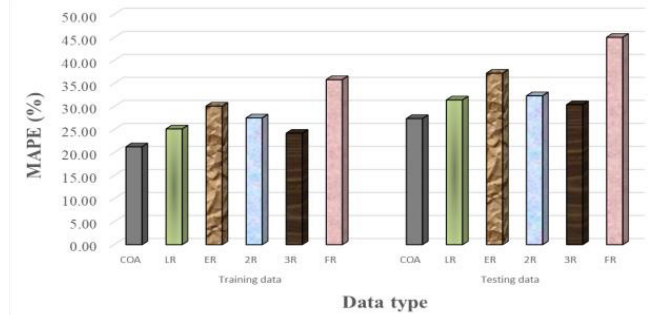


Fig. 3. The mean absolute percentage error

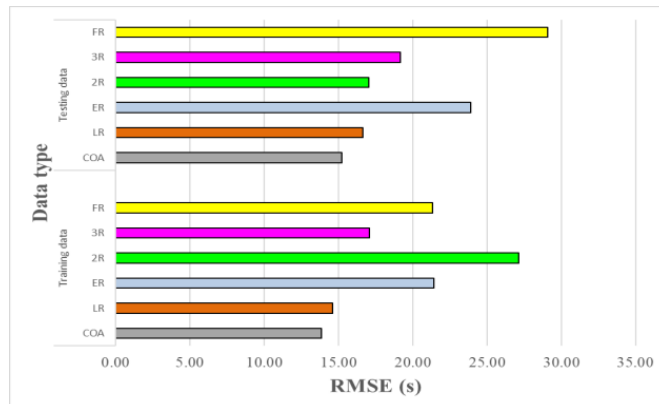


Fig. 4. The root mean square error

Results and discussions

Figure 1 displays the coefficient of determination related to the prediction models. Figures 2 and 3 illustrate the MAE and MAPE of estimation algorithms, respectively. The RMSE and E20s results are indicated in Figures 4 and 5 in the order mentioned.

3- Conclusions

As can be seen, the accuracy of COA is considerably better than that of other models for both testing and training data. Moreover, the introduced model is highly qualified to estimate travel time with high accuracy.

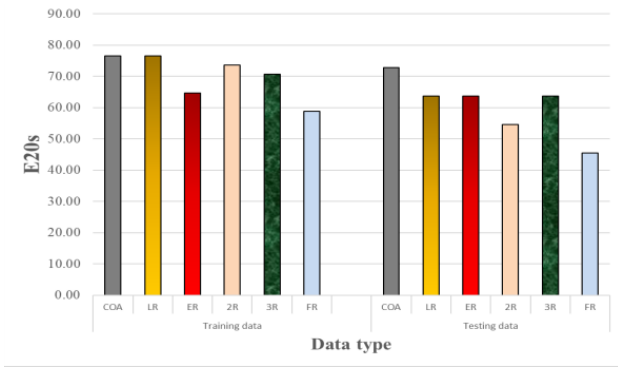


Fig. 5. The percentage of data which their error is fewer than 20 seconds

References

- [1] travel time estimation methods using sparse GPS data, *Journal of Intelligent Transportation Systems*, 20(6) (2016) 532-544.
- [2] G. Sil, A. Maji, S. Nama, A.K. Maurya, Operating speed prediction model as a tool for consistency based geometric design of four-lane divided highways, *Transport*, 34(4) (2019) 425-436.
- [3] C. Vilarinho, J.P. Tavares, R.J. Rossetti, Intelligent traffic lights: Green time period negotiaton, *Transportation research procedia*, 22 (2017) 325-334.
- [4] S.M. Turner, W.L. Eisele, R.J. Benz, D.J. Holdener, *Travel time data collection handbook*, United States. Federal Highway Administration, 1998.

HOW TO CITE THIS ARTICLE

A. Golroo , A. H. Fani, H. Naseri , *Travel Time Modelling of Urban Roads By Application of Coyote Optimization-based Machine Learning Method*, *Amirkabir J. Civil Eng.*, 53(9) (2021) 809-812.

DOI: [10.22060/ceej.2020.17991.6730](https://doi.org/10.22060/ceej.2020.17991.6730)





مدل‌سازی زمان سفر معابر شهری با استفاده از روش یادگیری ماشین گرگ صحرائی آمریکای شمالی

امیر گل رو*، امیرحسین فانی، حامد ناصری

دانشکده مهندسی عمران و محیط زیست، دانشگاه امیرکبیر، تهران، ایران.

تاریخچه داوری:

دریافت: ۱۳۹۸/۱۱/۰۶
بازنگری: ۱۳۹۹/۰۲/۲۸
پذیرش: ۱۳۹۹/۰۲/۲۸
ارائه آنلاین: ۱۳۹۹/۰۳/۰۸

کلمات کلیدی:

پیش بینی زمان سفر
معابر شهری
رگرسیون
یادگیری ماشین
الگوریتم گرگ صحرائی آمریکای شمالی.

خلاصه: زمان سفر به عنوان یکی از تأثیرگذارترین پارامترها در تحلیل عملکرد شبکه راه‌ها به جهت مدیریت و برنامه‌ریزی آتی شبکه و همچنین مهم‌ترین متغیر در فرآیند انتخاب مسیر استفاده‌کنندگان از راه می‌باشد. پیش بینی زمان سفر در طول چنددهه به عنوان یک موضوع پرکاربرد در زمینه شرایط دینامیک شبکه و سیستم‌های هوشمند حمل و نقل به روش‌های مختلفی انجام شده است. در این راستا، در این مقاله مدل‌سازی پیش بینی زمان سفر معابر شهری شریانی درجه یک و دو در ساعت غیر اوج صورت گرفته است. در وهله اول، طراحی آزمایش جهت نمونه برداری صورت گرفته و پارامترهای موثر بر زمان سفر معابر شهری بررسی شدند. سپس داده‌ها به وسیله روش ماشین شناور و به کمک یک نرم افزار تلفن همراه در ۶ معبر منتخب برداشت شدند. پس از آماده سازی داده‌ها، متغیرهای زمان توقف تقاطع، تعداد خطوط و شیب به عنوان متغیرهای موثر در مدل پیش بینی زمان سفر شناخته شدند. یک روش جدید یادگیری ماشین مبتنی بر الگوریتم گرگ صحرائی آمریکای شمالی برای مدل‌سازی زمان سفر معرفی گردید و دقت مدل ساخته شده با پنج مدل رگرسیونی مرسوم مقایسه گردید. بر اساس نتایج به دست آمده و بر مبنای بررسی ۵ شاخص عملکردی در نظر گرفته شده برای ارزیابی صحت و دقت مدل‌ها، گرگ صحرائی آمریکای شمالی از تمامی مدل‌ها دقت بالاتری داشته و ضریب تعیین این مدل برای داده‌های آموزشی و آزمایشی به ترتیب برابر ۰/۷۴۶ و ۰/۷۲۴ می‌باشد. همچنین این مدل ۷۳٪ داده‌های آزمایشی را با خطای کمتر از ۲۰ ثانیه پیش بینی می‌کند.

۱- مقدمه

به دلیل افزایش شلوغی و تراکم در شبکه راه‌ها، پیش بینی زمان سفر معابر شبکه ترافیکی به یک بخش مهم در سیستم‌های اطلاع رسانی به رانندگان^۱ بدل شده است. این سیستمها به کاربران کمک میکنند تا مسیرهایی با زمان سفر پایین تر را انتخاب کنند. زمان سفر عبارت است از کل زمانی که یک وسیله نقلیه برای پیمودن یک مسیر مشخص از یک نقطه شبکه به نقطه ای دیگر با احتساب توقف‌ها، تاخیرهای در صف و تاخیرهای در تقاطع‌ها طی میکند [۱].

زمان سفر از اطلاعات ترافیکی استخراج می‌شود. اطلاعات ترافیکی میتوانند از روش‌های مختلفی استخراج شوند. یک روش قرار دادن یک یا چند حسگر در طول راه و برداشت اطلاعات ترافیکی از طریق آن‌ها است. این حسگرها به گونه‌های مختلفی نظیر شناساگر نقطه‌های^۲ (ثبت تعداد ماشین‌های عبوری از یک نقطه)، شناساگر بازهای^۳ (ثبت اطلاعات وسایل

- 1 Advanced Traveler Information System
- 2 Point Detector
- 3 Interval Detector

* نویسنده عهده‌دار مکاتبات: agolroo@aut.ac.ir

نقلیه از طریق GPS) و تشخیص اتوماتیک ماشین^۴ (قرار دادن دو حسگر در طول یک راه و شناسایی ماشینهای عبوری در این فاصله) هستند [۲]. روش ساده‌تر استفاده از یک وسیله نقلیه و پیمودن مسیر مشخص شده توسط آن و ثبت اطلاعات زمان سفر و سایر متغیرهای مسیر می‌باشد [۳]. در این پژوهش از روش دوم برای جمع آوری اطلاعات استفاده شده است.

پیش بینی زمان سفر معابر شبکه ترافیکی به طور کلی در دو گروه آزادراه و بزرگراه و معابر شهری و شریانی مورد بررسی قرار می‌گیرد. بسیاری از مدل‌های ساخته شده برای پیش بینی زمان سفر برای آزادراه‌ها و بزرگراه‌ها میباشند زیرا جریان در این معابر نامنقطع و ساخت مدل زمان سفر ساده تر است. همچنین متغیرهای موثر بر زمان سفر در معابر شهری به دلیل وجود تقاطع‌ها و چراغ‌های راهنمایی پیچیده تر میباشد [۴]. در این پژوهش، نمونه برداری و مدل‌سازی برای معابر شهری و شریانی انجام شده است.

روش‌های متفاوتی برای مدل‌سازی و پیش بینی زمان سفر ارائه شده است که می‌توان آن‌ها را در ۴ دسته مختلف مورد بررسی قرار داد. دسته

4 Automatic Vehicle Identification (AVI)

حقوق مؤلفین به نویسندگان و حقوق ناشر به انتشارات دانشگاه امیرکبیر داده شده است. این مقاله تحت لیسانس آفرینندگی مردمی (Creative Commons License) در دسترس شما قرار گرفته است. برای جزئیات این لیسانس، از آدرس <https://www.creativecommons.org/licenses/by-nc/4.0/legalcode> دیدن فرمائید.



روز و ساعت و وضعیت آب و هوایی [۱۲] و غیره در نظر گرفته شده است. برای تعیین معیار یادگیری مدل از روش ساده و معروف مینیم مجذور مربعات [۱۳] تا روش پیچیده مینیم آنترویی [۱۱] استفاده شده است.

در سال‌های اخیر، روش‌های جدید مبتنی بر الگوریتم‌های فرا ابتکاری^۶ ابداع شده و در علوم مختلف مورد استفاده قرار گرفته‌اند لیکن کاربرد آن‌ها در موضوع پیش بینی زمان سفر مشاهده نمی‌شود. ناصری و همکاران [۱۴] با استفاده از الگوریتم چرخه آب^۷ یک روش یادگیری ماشین^۸ جدید ارائه کردند و این مدل در پیش بینی طرح مخلوط بتن حاوی خاکستر بادی مورد استفاده قرار گرفت. نتایج آن پژوهش، بر دقت بالای روش پیش بینی مبتنی بر الگوریتم‌های فرا ابتکاری دلالت داشت. همچنین، گلفشانی و بهنود [۱۵] با استفاده از الگوریتم بیوگرافی^۹ یک روش جدید یادگیری ماشین ارائه کردند و روش مذکور در پیش بینی طرح مخلوط بهینه بتن حاوی دوده سیلیس مورد بررسی قرار گرفت و نتایج حاکی از دقت بالای این روش نسبت به روش‌های مرسوم یادگیری ماشین بودند.

متغیرهای گوناگونی در پژوهش‌های پیشین به عنوان متغیرهای موثر بر روی زمان سفر معرفی شده‌اند. اووه و همکاران [۱۶] مدل‌های پیش بینی زمان سفر و متغیرهای در نظر گرفته شده در آنها را مرور کردند. از جمله این متغیرها می‌توان به خصوصیات فردی راننده، مشخصات وسیله نقلیه، مشخصات معابر مانند سرعت حرکت آزاد، تعداد خطوط و شیب، زمان تاخیر و زمان توقف تقاطعات، شرایط جوی، زمان اوج و غیروا، روزهای خاص و مناسب‌ها، تعداد دسترسی‌ها نام برد. ابدلال و همکاران عرض راه، تعداد وسیله نقلیه سنگین عبوری در محور مورد بررسی و تعداد رانندگان را از مهم ترین عوامل انباشتگی ترافیک معرفی کردند که می‌توانند در زمان سفر تاثیرگذار باشند [۱۷]. سیل^{۱۰} و همکاران هندسه راه را از موثرترین مولفه‌های سرعت رانندگی بیان کردند و با بررسی هندسه راه‌ها به پیش بینی سرعت رانندگی و زمان سفر پرداختند [۱۸]. در این پژوهش از روش ماشین شناور جهت برداشت داده استفاده گردید و سرعت ماشین نیز علاوه بر هندسه راه در پیش بینی زمان سفر مورد استفاده قرار گرفت. ویلارینهو^{۱۱} و همکاران چراغ‌های راهنمایی و رانندگی و توقف ناشی از آن‌ها عواملی بسیار تاثیرگذار بر ترافیک بیان کردند [۱۹]. در این تحقیق، با در نظر گرفتن اطلاعات

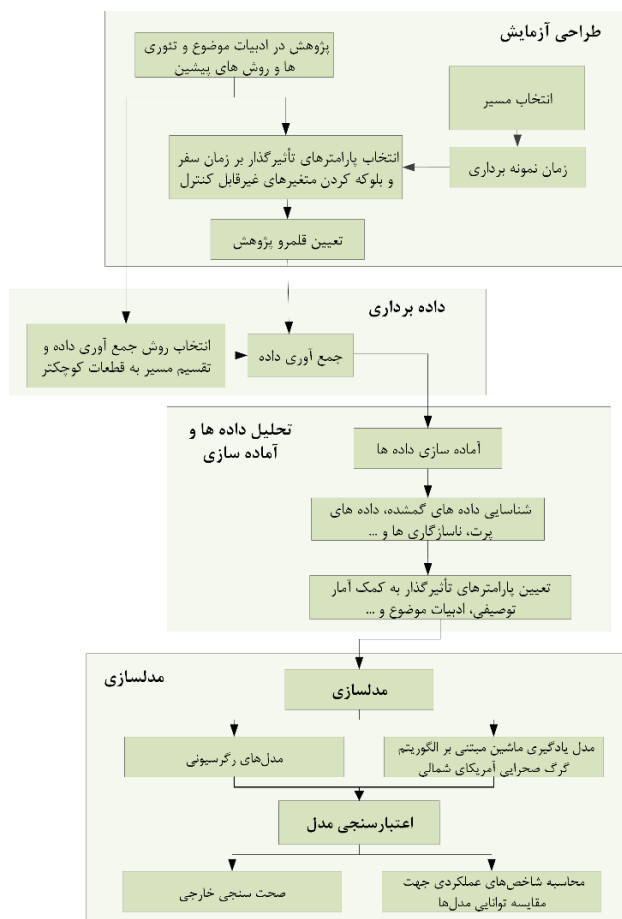
اول روش‌های ساده و لحظ‌های می‌باشند که دارای فرض‌های ساده کننده زیادی هستند و در بسیاری از موارد این فرض‌ها برقرار نمی‌باشند. به طور مثال یکی از این فرض‌ها عدم تغییر وضعیت شلوغی شبکه تا چند دقیقه یا چند ساعت بعد است. این روش‌ها غالباً در نرم افزارهای تجاری استفاده میشوند. مدل‌های دسته دوم بر مبنای بررسی وضعیت ترافیک و شبیه سازی آن به صورت خردنگر و کلان نگر هستند. این مدل‌ها به اطلاعات جریان ترافیک و حجم عبوری نیاز دارند. مدل‌های دسته سوم زمان سفر را بر مبنای اطلاعات جمع آوری شده و ویژگی‌های معابر پیش بینی می‌کنند. انواع مختلفی از این دسته روش‌ها مانند رگرسیون، سری زمانی^۱، شبکه عصبی^۲، درخت تصمیم^۳ و غیره برای مدل‌سازی زمان سفر به کار گرفته شده‌اند. در دسته چهارم مدل‌هایی با ترکیب روشهای دسته دوم و سوم زمان سفر معابر را تخمین می‌زنند [۵].

مدل‌های بر مبنای اطلاعات و روابط متغیرها به ۳ گروه پارامتریک، ناپارامتریک و نیمه پارامتریک تقسیم می‌شوند. در روش‌های پارامتریک متغیرهای موثر بر زمان سفر شناسایی شده و تابعی از زمان سفر و متغیرهای موثر تشکیل می‌شود. این مدل‌ها فهم آسانی دارند و به راحتی پیاده سازی می‌شوند لیکن ساختار ساده آنها تفسیر رفتار زمان سفر را مشکل می‌کند. از معروف‌ترین این روش‌ها می‌توان به رگرسیون خطی^۴، رویکرد بیز^۵ و سری زمانی اشاره کرد. در روش‌های ناپارامتریک ساختار مدل و متغیرها از پیش تعیین شده نمی‌باشد و شکل تابع زمان سفر از داده‌ها استخراج می‌شود. این روش‌ها به تعداد داده بیشتری نسبت به روش‌های پارامتریک نیاز دارند. روش‌های شبکه عصبی و درخت تصمیم از معروفترین روش‌های ناپارامتریک می‌باشند که در پیش بینی زمان سفر به کار گرفته شده‌اند. روش‌های نیمه پارامتریک ترکیبی از روش‌های پارامتریک و ناپارامتریک برای مدل‌سازی استفاده می‌کنند. در این مقاله از روش‌های پارامتریک برای پیش بینی زمان سفر استفاده شده است [۶].

روش رگرسیون برای مدل‌سازی زمان سفر در پژوهش‌های متعددی استفاده شده است. در رگرسیون یک مدل بین متغیر وابسته با یک یا چند متغیر مستقل تعریف می‌شود. در هر یک از این پژوهش‌ها متغیرهای مختلفی من جمله زمان تاخیر یا توقف تقاطع [۷]، حجم ترافیک ثبت شده [۸]، شیب [۹]، تعداد خطوط [۱۰]، داده‌های تاریخی^۱ و ویژگی‌های فیزیکی معبر [۱۱]،

- 6 Metaheuristic Algorithms
- 7 Water Cycle Algorithm (WCA)
- 8 Machine Learning
- 9 Biogeography-based Programming
- 10 Sil
- 11 Vilarinho

- 1 Time Series
- 2 Neural Network
- 3 Decision Tree
- 4 Linear Regression
- 5 Bayesian Approach



شکل ۱. روند انجام تحقیق

Fig. 1. The flowchart of the methodology

اوج به وسیله روش های یادگیری ماشین مبتنی بر الگوریتم گرگ صحرایی آمریکای شمالی، رگرسیون های خطی، نمایی، چند جمله ای مرتبه ۲، چند جمله ای مرتبه ۳ و کسری میباشد. به علاوه، صحت و دقت نتایج مدل های به دست آمده توسط روش های مذکور، به وسیله ۵ شاخص عملکردی مختلف، ارزیابی و مورد بررسی قرار گرفته اند.

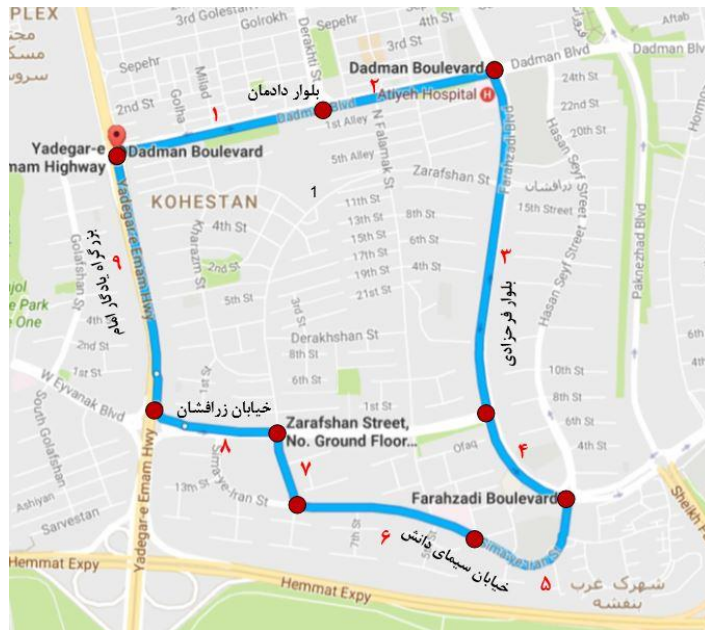
۲- روش پژوهش

در این پژوهش جهت پیش بینی زمان سفر در معابر شهری ابتدا اقدام به طراحی آزمایش شد. سپس جمع آوری داده در روز مشخص انجام شد. در مرحله بعد داده های جمع آوری شده از جهت هر گونه نقص یا خطا مورد بررسی قرار گرفت. در نهایت مدل سازی زمان سفر روی داده های حاصله انجام و مدل به دست آمده اعتبارسنجی شد. شکل ۱ روند کلی تحقیق را به صورت خلاصه نشان می دهد. در ادامه در بخش های بعدی، هر یک از

ترافیکی در مجاورت چراغ های راهنمایی و رانندگی، به برنامه ریزی زمان بندی چراغ های راهنمایی و رانندگی پرداختند که هدف اصلی آن ها کاهش تاخیر و توقف تقاطع ها بود. از این رو در این پژوهش، متغیرهای ذکر شده به همراه ویژگی های فیزیکی معبر به عنوان متغیرهایی که می توانند بر روی زمان سفر مسیر اثرگذار باشند مورد بررسی قرار گرفتند که در بخش طراحی آزمایش و داده برداری به تفصیل به آن ها پرداخته خواهد شد.

با توجه به مطالب فوق، بر اساس مطالعات صورت گرفته روش های جدید یادگیری ماشین مبتنی بر الگوریتم های فراابتکاری در پیش بینی زمان سفر معابر مورد استفاده قرار نگرفته اند. از این رو یک روش جدید به نام یادگیری ماشین مبتنی بر الگوریتم گرگ صحرایی آمریکای^۱ در این پژوهش معرفی گردیده و جهت پیش بینی زمان سفر مورد استفاده قرار گرفته است. به عبارت دیگر، هدف از این پژوهش پیش بینی زمان سفر معابر شهری در حالت غیر

1 Coyote Optimization Algorithm (COA)



شکل ۲. محدوده مورد مطالعه و قطعات در نظر گرفته شده

Fig. 2. The studied area and the considered sections

بخش‌های ارائه شده در فلوجارت شکل ۱ تشریح شده است.

گیری، قطعات حداکثر دارای طول ۱ کیلومتر باشند. محدوده مورد مطالعه و ۹ مقطع در نظر گرفته شده در شکل ۲ مشخص شده اند.

نمونه برداری در روز جمعه ۲۶ بهمن ۱۳۹۷ ساعت ۱۶:۳۰ آغاز شد. روز و ساعت نمونه برداری و همچنین مشاهده وضعیت خیابان در هنگام نمونه برداری، نشان میداد داده‌های به دست آمده مربوط به زمان غیر اوج و با توجه به تعاریف کیفی که از سطح سرویس^۱ راه‌ها در کتاب راهنمای ظرفیت راه‌ها^۲ [۲۰] موجود است می‌توان به صورت کیفی سطح سرویس مقاطع محدوده را بین B تا C ارزیابی کرد میباشند. همچنین برای جلوگیری از بروز خطاهای مختلف داده برداری با ۵ بار تکرار انجام شد. نمونه برداری به صورت طی مسیر با وسیله نقلیه و به کمک نرم افزار Mytacks انجام شد. این نرم افزار همزمان با طی مسیر، اطلاعات طول، عرض و ارتفاع جغرافیایی، سرعت لحظه‌ای و زاویه با جهت شمال جغرافیایی را ذخیره میکند. علاوه بر این زمان سفر طی شده در طول مسیر به صورت دستی نیز

۲-۱- طراحی آزمایش و جمع آوری داده

به دلیل ارتباط تنگاتنگ طراحی آزمایش و جمع آوری داده به این دو گام در یک بخش پرداخته می‌شود. برای انجام طراحی آزمایش فاکتورهای موثر روی زمان سفر در معابر شهری به کمک مرور پژوهش‌های انجام گرفته و همچنین تئوری‌های پیشین مورد بررسی قرار گرفت. همچنین برای انتخاب مسیر، با بررسی مسیرهای مختلف با ویژگی‌های متفاوت یک حلقه شامل ۶ خیابان در محدوده خیابان دادمان، بلوار فرحزادی و بخش کوچکی از بزرگراه یادگار امام در غرب تهران با رده عملکردی شریانی درجه یک (بزرگراه) و درجه دو (خیابان‌های شریانی اصلی و فرعی) انتخاب شد. برخی معابر با توجه به تغییرات در متغیرهای فیزیکی و هندسی معبر و همچنین کتابچه NCHRP Handbook [۳] به قطعات کوچک‌تر تقسیم شدند. در این کتابچه راهنمایی شده است که برای جلوگیری از بروز انواع خطاها در نمونه

1 Level of Service

2 Highway Capacity Manual (HCM)

جدول ۱. تست نرمال بودن داده‌های زمان سفر مقطع اول محدوده مورد مطالعه توسط آزمون‌های کولمگوروف اسمیرنوف و شاپیرو ویلک در نرم‌افزار SPSS

Table 1. The normality test of travel time data for the first section through Kolmogorov-Smirnov and Shapiro-Wilk test

زمان سفر	شاپیرو ویلک			کولمگوروف اسمیرنوف		
	Sig.	df	Static	Sig.	df	Static
	۰/۲۵۶	۵	۰/۸۶۷	۰/۲۰۰	۵	۰/۲۴۵

تست شد. همچنین به کمک آمار توصیفی و تست‌های بررسی نرمال بودن مانند کولمگوروف اسمیرنوف^۵، شاپیرو ویلک^۶ و تست ترسیم نمودار Q-Q نرمال بودن داده‌ها به طور تقریبی مورد وثوق قرار گرفت. به طور مثال اجرای تست کولمگوروف اسمیرنوف و شاپیرو ویلک برای داده‌های زمان سفر مقطع اول محدوده مورد مطالعه در جدول ۱ ارائه شده است. با توجه به بزرگتر بودن مقدار احتمال ($\text{Sig} = ۰/۲۵۶$ و $\text{Sig} = ۰/۲۰۰$) از احتمال خطای نوع اول ($۰/۰۵$) فرض صفر که معتبر بودن توزیع نرمال برای داده‌ها است، رد نمی‌شود. پس از آن جهت راستی آزمایی داده‌ها، ارتباط منطقی بین پارامترهای تأثیر گذار روی زمان سفر بررسی شد. به طور مثال با افزایش زمان توقف تقاطع زمان سفر افزایش پیدا می‌کند. در نهایت توزیع آماری برای داده‌های تهیه شده برآزش شد و آماره‌های هر توزیع تعیین شدند و از نتایج این قسمت برای تعیین پارامتر تأثیر گذار و پیش بینی زمان سفر استفاده می‌شود.

۲-۳- تعیین متغیرهای تأثیر گذار

برای تشخیص متغیرهای موثر در مدل، از روش گام به گام^۷ یعنی اضافه کردن یک به یک متغیرها و بررسی خوبی برازش با استفاده از پارامتر ضریب تعیین (R^2) و اهمیت متغیرها و سایر ویژگی‌های مدل استفاده شده است. همچنین از مدل رگرسیون خطی جهت شناسایی پارامترهای مهم استفاده گردید. در رگرسیون خطی مدل یک متغیره با متغیر مستقل زمان توقف تقاطع R^2 برابر ۰/۳۹۵ نتیجه می‌دهد. با اضافه کردن تعداد خطوط به زمان توقف تقاطع مدلی ۲ متغیره با R^2 برابر ۰/۵۳۷ حاصل می‌شود. مدل ۳ متغیره با متغیرهای زمان توقف تقاطع، تعداد خطوط و شیب مدل رگرسیون

ذخیره شد. همچنین نوع حرکت و نحوه رانندگی وسیله نقلیه با توجه به نکات وجود در کتابچه NCHRP Handbook [۳] به صورت ماشین شناور^۱ انتخاب شد. در حالت ماشین شناور تعداد وسایل نقلیه ای که از ماشین در حال نمونه برداری عبور می‌کنند برابر تعداد وسایل نقلیه عبور شده می‌باشد. در نهایت اطلاعات متغیرهای زمان سفر، زمان توقف تقاطع، تعداد خط، تعداد دسترسی، وجود یا عدم وجود پارکینگ حاشیه ای، کاربری، شیب، تعداد سرعتگاه، سرعت لحظه ای و طول معبر در هنگام نمونه برداری ثبت شد. لازم به ذکر است زمان توقف تقاطع مطابق تعریف کتاب راهنمای ظرفیت راه‌ها برابر زمان انتظار برای وسیله نقلیه است که سرعت آن صفر باشد [۲۰]. با توجه به طراحی آزمایش و نمونه برداری انجام شده قلمرو این تحقیق شناسایی عوامل موثر بر زمان سفر معابر شهری و توسعه مدل مناسبی برای پیش‌بینی زمان سفر آنها است. در این خصوص قلمرو این تحقیق صرفاً به ساعات غیر اوج و شرایط معمول آب و هوایی (هوای آفتابی) محدود می‌شود. همچنین طول معابر مورد پیش بینی با توجه به دستور العمل‌های قطعه بندی و سایر متغیرهای مسیر میبایست حداکثر ۱ کیلومتر باشد. زیرا در طول‌های بالاتر ممکن است رفتار رانندگی استفاده کنندگان تغییر یابد.

۲-۲- آماده سازی داده

قبل از تحلیل آماری اطلاعات، لازم است آماده‌سازی داده‌ها صورت گیرد. با یک تحلیل اولیه مانند کشیدن نمودار جعبه ای^۲، داده‌های پرت^۳ حذف می‌شوند. داده‌های گم‌شده^۴ بررسی شد و ۲ مورد پیدا شده به دلیل کم اهمیت بودن حذف شد. عدم حضور ناسازگاری بین داده‌ها تحلیل و

5 Kolmogorov-Smirnov
6 Shapiro-Wilk
7 Stepwise

1 Floating Car
2 Box Plot
3 Outlier
4 Missing Value

طبقه بندی شدند. به عبارت دیگر ۷۵٪ داده‌ها به عنوان داده‌های آموزشی^۴ در نظر گرفته شدند و روابط زمان سفر بر اساس متغیرهای تاثیرگذار توسط این داده‌ها تولید گردید. سپس، ۲۵٪ باقی‌مانده داده‌ها (داده‌های آزمایشی^۵) در روابط تولید شده قرار دادند تا قدرت مدل‌ها جهت پیش بینی داده‌های دیده نشده سنجیده شود. در ادامه روش‌های به کار گرفته شده در این پژوهش توضیح داده می‌شوند.

۲-۴-۱- مدل یادگیری ماشین مبتنی بر الگوریتم گرگ صحرایی آمریکایی شمالی

الگوریتم گرگ خاکستری (COA) یک الگوریتم فرا ابتکاری جدید می‌باشد که از تجارب تعاملی و رفتار اجتماعی گرگ خاکستری آمریکایی شمالی الهام گرفته شده است. این الگوریتم یک روش توانمند جهت یافتن جواب‌های دقیق مسایل بهینه سازی می‌باشد و در سال ۲۰۱۸ میلادی توسط پیرزن و کوالهو [۲۱] معرفی گردید. در این الگوریتم، گرگ‌ها و رفتار اجتماعی آن‌ها به ترتیب به بردارهای جواب و برازندگی جواب‌ها نسبت داده می‌شوند. ابتدا، گرگ‌ها (بردارهای جواب) به دسته‌های مختلف تقسیم می‌شوند. در هر گروه، با ارزش‌ترین بردار جواب (از لحاظ مقدار تابع هدف مسئله بهینه سازی) گرگ آلفا نامیده می‌شود. در هر گروه، گرگ‌ها با یکدیگر تعامل داشته و فرهنگ آن‌ها از یکدیگر و به خصوص گرگ آلفا تأثیر می‌پذیرد. به عبارت دیگر هر بردار جواب با برداری به سمت سایر هم گروهی‌ها و با ارزش‌ترین بردار جواب انتقال داده می‌شود. جهت عدم توقف در جواب‌های کمینه موضعی، برخی از گرگ‌ها بین گروه‌های مختلف با یکدیگر جا به جا می‌شوند و این پدیده سبب پوشش فضای بیشتری از ناحیه شدنی مسئله می‌شود. همچنین، در هر تکرار الگوریتم، ضعیف‌ترین گرگ‌ها (ضعیف‌ترین بردارهای جواب از لحاظ مقدار تابع هدف مسئله بهینه‌سازی) از جامعه از بین رفته و جمعیت جدید جایگزین آن‌ها می‌شوند. این پدیده با ایجاد جهش سبب از بین رفتن بردار جواب‌های ضعیف تر و ارتقاء مدل می‌شود.

در این پژوهش، با استفاده از الگوریتم مذکور یک روش جدید یادگیری ماشین معرفی شده است. در روش مدل‌سازی مبتنی بر الگوریتم گرگ خاکستری، هر ورودی مدل (پارامترهای موثر بر زمان سفر) به عنوان یک متغیر تصمیم عدد صحیح در نظر گرفته شد که این متغیرهای تصمیم می‌توانند از بین ۲۰ حالت انتخابی یک تابع را انتخاب نمایند. به عبارت دیگر هر یک از پارامترهای زمان توقف تقاطع، تعداد خطوط و شیب باید به

با R^2 برابر ۰/۶۵ نتیجه می‌دهد. لازم به ذکر است در تمامی ۳ مدل ذکر شده متغیرها در تست T دارای اهمیت و معنی دار می‌باشند و همچنین واریانس‌ها در تست F دارای تفاوت معنی دار می‌باشند.

با اضافه کردن متغیرهای دیگر مدل بهتری از مدل ذکر شده که در تمامی فرض‌ها و شروط رگرسیون صادق باشد، به دست نمی‌آید. به طور مثال با اضافه شدن تعداد سرعت کاه R^2 مدل کاهش یافته و همچنین متغیر سرعت کاه در تست T معنی دار نمی‌باشد. همچنین با اضافه کردن تعداد دسترسی‌ها توزیع باقیمانده‌ها از توزیع نرمال پیروی نمی‌کند و لذا مدل قابل قبول نبود.

به علاوه، به دلیل گسسته بودن متغیر تعداد خطوط و گزینه‌های محدود آن (۲، ۳ و ۴ خط) اقدام به ساخت مدل با متغیر دامی^۱ تعداد خطوط به همراه متغیرهای زمان توقف تقاطع و شیب شد، لیکن پس از ساخت مدل مشاهده شد علی‌رغم بالا رفتن R^2 به میزان اندکی (از ۰/۶۵ به ۰/۶۷)، تابع توزیع باقیمانده‌های حاصل از این مدل از توزیع نرمال تبعیت نکرده و در تست کولمگروف اسمیرنوف رد می‌شوند و لذا این شرط رگرسیون خطی رعایت نمی‌شود.

۲-۴-۲- مدل‌سازی

مدل‌سازی پیش بینی زمان سفر از طریق روش یادگیری ماشین مبتنی بر الگوریتم گرگ صحرایی آمریکایی شمالی و ۵ مدل رگرسیونی شامل رگرسیون‌های خطی، نمایی، چند جمله‌ای مرتبه ۲، چند جمله‌ای مرتبه ۳ و کسری انجام شده است. روش رگرسیون در نرم افزار SPSS^۲ مدل شده است. همچنین مدل‌سازی روش یادگیری ماشین مبتنی بر الگوریتم گرگ صحرایی آمریکایی شمالی در نرم افزار MATLAB^۳ انجام گرفته است. نکته حائز اهمیت در این روش‌ها این است که به دلیل این که مقاطع مطالعه شده دارای طول‌های مختلفی می‌باشند و هدف ارائه مدلی واحد برای تمامی معابر می‌باشد، دو متغیر زمان سفر به عنوان متغیر وابسته و زمان توقف تقاطع به عنوان یکی از متغیرهای مستقل که دارای واحد ثانیه می‌باشند به طول هر مقطع تقسیم شده و مدل ارائه شده به ازای زمان در واحد طول می‌باشد (ثانیه در هر کیلومتر).

جهت ساختن مدل‌های پیش‌بینی پس از تعیین متغیرهای تاثیر گذار، داده‌ها به صورت تصادفی به دو دسته داده‌های آموزشی و داده‌های آزمایشی

- 1 Dummy Variable
- 2 Statistical Package for Social Sciences (SPSS)
- 3 MATrix LABoratory (MATLAB)

4 Train Data
5 Test Data

۳- بیان و تحلیل نتایج

در این بخش نتایج مدله ای ساخته شده ارائه و تحلیل می شوند. به طور کلی با وجود ۹ مقطع و ۵ تکرار انجام شده، ۴۵ ردیف داده برای مدل سازی به کار گرفته شد. همانطور که در بخش ۳-۴ مطرح شد، داده ها به دو دسته داده های آموزشی و داده های آزمایشی طبقه بندی گردیدند. مدل ها با داده های آموزشی ساخته شده و سپس توانایی پیش بینی آن ها با داده های آزمایشی مورد ارزیابی قرار گرفت. پیش از ارائه مدل ها نحوه اعتبارسنجی و شاخص های عملکردی مورد استفاده معرفی می شوند.

۳-۱- شاخص های عملکردی مورد استفاده برای اعتبارسنجی مدل ها

جهت مقایسه توانایی مدل ها و بررسی دقت جواب های حاصله از شاخص های عملکردی میانگین خطای حقیقی^۲ (MAE)، خطای جذر میانگین مربعات^۳ (RMSE)، میانگین درصد قدرمطلق خطا^۴ (MAPE)، ضریب تعیین (R^۲) و درصد داده هایی که خطای آن ها کمتر از ۲۰ ثانیه می باشد (E_{20s}) استفاده گردید. روابط شاخص های عملکردی مذکور به ترتیب در روابط (۱) تا (۵) ارائه شده است.

$$MAE = \frac{\sum_{i=1}^n |EXP_i - PRE_i|}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (EXP_i - PRE_i)^2}{n}} \quad (2)$$

$$MAPE = \frac{100}{n} \times \sum_{i=1}^n \left| \frac{EXP_i - PRE_i}{EXP_i} \right| \quad (3)$$

$$R^2 = \left(\frac{\sum_{i=1}^n (EXP_i - \overline{EXP_i}) \times (PRE_i - \overline{PRE_i})}{\sqrt{\sum_{i=1}^n (EXP_i - \overline{EXP_i})^2 \times \sum_{i=1}^n (PRE_i - \overline{PRE_i})^2}} \right)^2 \quad (4)$$

$$E_{20s} = \frac{(\text{number of data which their error is fewer than } 20s) \times 100}{n} \quad (5)$$

یک تابع تخصیص داده شوند که توابع مذکور شامل ۲۰ تابع مختلف مثلثاتی (سینوسی، کوسینوسی، تانژانتی، کوتانژانتی)، نمایی، توانی با فرجه های مختلف، لگاریتمی با فرجه های مختلف و رادیکالی با ریشه های مختلف می باشند. همچنین ۸ متغیر تصمیم جهت یافتن ضرایب ثابت رابطه در نظر گرفته شد. در نهایت ۴۰۰ تابع هدف مختلف جهت تجمیع متغیرهای تصمیم و ضرایب ثابت لحاظ شد و با در نظر گرفتن میانگین خطای حقیقی به عنوان تابع هدف، مسئله بهینه سازی حل گردید که هدف مسئله مذکور کمینه کردن میانگین خطا بر اساس یافتن مقادیر بهینه برای متغیر تصمیم می باشد. لازم به ذکر است این الگوریتم ۱۰ بار اجرا گردید و از بین ۱۰ جواب یافته شده، جوابی که دارای کمترین مقدار خطا بود تحت عنوان جواب الگوریتم گرگ خاکستری آمریکای شمالی ارائه می گردد. مبنای این روش یافتن رابطه ای می باشد که بر اساس آن مقادیر پیش بینی شده و مقادیر حقیقی بیشترین تطابق را داشته باشند و از این رو تابع هدف مسئله (میانگین خطای حقیقی) به کمترین میزان ممکن خود می رسد.

۲-۴-۲- مدل های رگرسیونی

رگرسیون یک تکنیک شناخته شده و مرسوم در یادگیری ماشین و داده کاوی^۱ است که در آن پیش بینی مقدار یک متغیر وابسته بر اساس مقادیر یک یا چند متغیر مستقل صورت می پذیرد. روش های متنوعی برای برازش خط یا منحنی روی داده ها وجود دارد که به وسیله مدل های مختلف رگرسیونی انجام می گیرند. در این پژوهش همانطور که پیشتر بیان شد، ۵ مدل رگرسیونی شامل رگرسیون های خطی، نمایی، چند جمله ای مرتبه ۲، چند جمله ای مرتبه ۳ و کسری به جهت مدل سازی داده ها استفاده شده است. در رگرسیون خطی رابطه بین متغیر وابسته و متغیرهای مستقل، یک رابطه خطی می باشد. رگرسیون غیر خطی که سایر مدل های رگرسیونی استفاده شده در این پژوهش به جز رگرسیون خطی در این دسته قرار می گیرند، به مدل هایی گفته می شود که برای تطبیق توابع غیر خطی بر روی داده های گردآوری شده به کار گرفته می شود. به طور مثال در رگرسیون نمایی سعی می شود بهترین مدل بر اساس توابع نمایی متغیرهای مستقل و ضرایب آن ها یافت شود. لازم به ذکر است تمامی مدل ها پس از ساخت نیاز به اعتبارسنجی و بررسی میزان دقت و صحت مدل معرفی شده دارند.

- 2 Mean Absolute Error (MAE)
- 3 Root Mean Square Error (RMSE)
- 4 Mean Absolute Percentage Error (MAPE)

$$FR - Time(s / km) = (-92.8259 \times \frac{StopDelay}{Lane \times Grade}) + 59.9815 \quad (11)$$

که در روابط فوق $StopDelay$ ، $Lane$ و $Grade$ به ترتیب زمان توقف تقاطع، تعداد خطوط و شیب می‌باشند.

۳-۳- دقت مدل‌های پیش بینی زمان سفر

پس از یافتن روابط ارائه شده با استفاده از داده‌های آموزشی، داده‌های دیده نشده (داده‌های آزمایشی) در روابط قرار گرفته و نتایج پیش بینی شده با مقادیر حقیقی داده مقایسه شدند و دقت مدل‌ها توسط شاخص‌های عملکردی که در بخش قبل ارائه شدند، مورد بررسی قرار گرفتند.

دقت مدل‌های یادگیری ماشین بر اساس میانگین خطای حقیقی به تفکیک داده‌های آموزشی و آزمایشی در شکل ۳ ارائه شده است. با توجه به نتایج این شکل، کمترین مقدار میانگین خطای حقیقی در هر دو دسته داده‌های آموزشی و آزمایشی متعلق به الگوریتم گرگ خاکستری آمریکای شمالی می‌باشد. مقدار میانگین خطای حقیقی الگوریتم گرگ خاکستری آمریکای شمالی برای داده‌های آموزشی و آزمایشی به ترتیب برابر ۱۰/۱ و ۲۷/۴ ثانیه می‌باشد. در بین مدل‌های رگرسیونی، کمترین مقدار میانگین خطای حقیقی برای داده‌های آموزشی مربوط به رگرسیون خطی با مقدار ۱۲/۲ و کمترین مقدار میانگین خطای حقیقی برای داده‌های آزمایشی مرتبط با رگرسیون مرتبه سه است. در نتیجه، بر اساس مقادیر میانگین خطای حقیقی، عملکرد الگوریتم گرگ خاکستری آمریکای شمالی بهتر از روش‌های رگرسیونی مرسوم می‌باشد. اگر چه، عملکرد رگرسیون خطی و رگرسیون مرتبه سه در مقایسه با سایر مدل‌های رگرسیونی قابل قبول به نظر می‌رسد. خطای جذر میانگین مربعات مدل‌های ارائه شده در شکل ۴ نشان داده شده است. بر اساس نتایج این شاخص عملکردی، بیشترین دقت متعلق به الگوریتم گرگ خاکستری می‌باشد. در داده‌های آزمایشی کمترین مقادیر میانگین خطای حقیقی به ترتیب مرتبط به الگوریتم گرگ خاکستری، رگرسیون خطی، رگرسیون مرتبه دو، رگرسیون مرتبه سه، رگرسیون نمایی و رگرسیون کسری با مقادیر ۱۵/۲، ۱۶/۶، ۱۷/۱، ۱۹/۲، ۲۳/۹ و ۲۹/۱ ثانیه است. از این رو، دقت الگوریتم گرگ خاکستری در پیش بینی زمان سفر بهتر از روش‌های رگرسیونی مرسوم به نظر می‌رسد.

که در روابط فوق PXE_i ، EXP_i ، PRE_i ، PRE_i و n به ترتیب مقادیر داده‌های حقیقی، میانگین مقادیر داده‌های حقیقی، مقادیر پیش بینی شده توسط مدل‌های مختلف، میانگین مقادیر پیش بینی شده توسط مدل‌ها و تعداد داده‌های استفاده شده در مدل (تعداد داده‌های آموزشی) می‌باشند. علاوه بر شاخص‌های عملکردی مذکور که به عنوان شاخص‌های مرسوم برای صحت سنجی و ارزیابی دقت نتایج به کار می‌رود، در بخش ۴-۴ یک رویکرد دیگر که تحت عنوان صحت‌سنجی خارجی^۱ توسط پژوهشگران این حوزه معرفی شده است به جهت تکمیل بررسی صحت مدل‌های معرفی شده به کار گرفته می‌شود.

۳-۲- شاخص‌های عملکردی مورد استفاده برای اعتبارسنجی مدل‌ها

در این بخش، روابط ارائه شده جهت پیش بینی زمان سفر توسط پارامترهای تاثیرگذار که مدل‌های یادگیری ماشین تولید کرده‌اند، ارائه می‌شود. در روابط (۶) تا (۱۱) به ترتیب روابط ارائه شده توسط الگوریتم گرگ خاکستری آمریکای شمالی (COA)، رگرسیون خطی، رگرسیون نمایی (ER)، رگرسیون مرتبه دو (۲R)، رگرسیون مرتبه سه (۳R) و رگرسیون کسری ارائه شده‌اند.

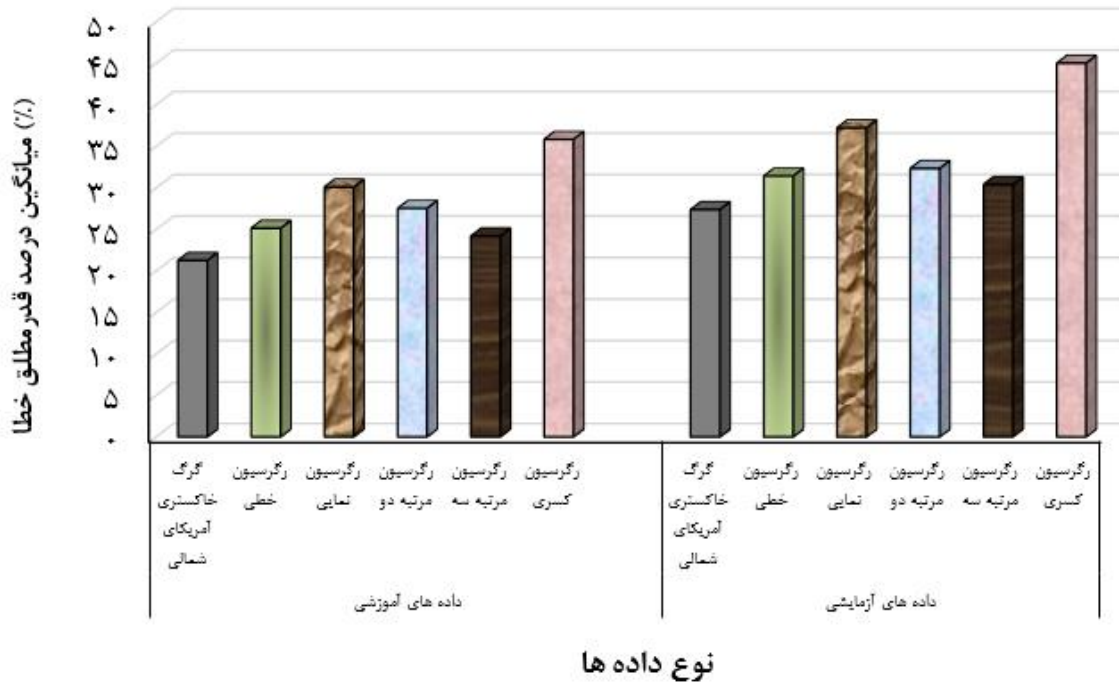
$$COA - Time(s / km) = (0.7 \times \sqrt[3]{3} \times StopDelay)^2 + (0.7 \times \sqrt{6} \times \sqrt{Lane})^2 + (0.8 \times Grade^2)^2 + (0.7 \times \sqrt[3]{3} \times StopDelay)^3 + (0.7 \times \sqrt{6} \times \sqrt{Lane})^3 + (0.8 \times Grade^2)^3 + 3.587401 \quad (6)$$

$$LR - Time(s / km) = (1.3688 \times StopDelay) + (10.1461 \times Lane) + (14.3076 \times Grade) + 13.3413 \quad (7)$$

$$ER - Time(s / km) = (0.0001 \times \exp(StopDelay)) + (0.4291 \times \exp(Lane)) + (4.9904 \times \exp(Grade)) + 34.0197 \quad (8)$$

$$2R - Time(s / km) = (0.0598 \times StopDelay^2) + (1.8666 \times Lane^2) + (8.7428 \times Grade^2) + 24.7024 \quad (9)$$

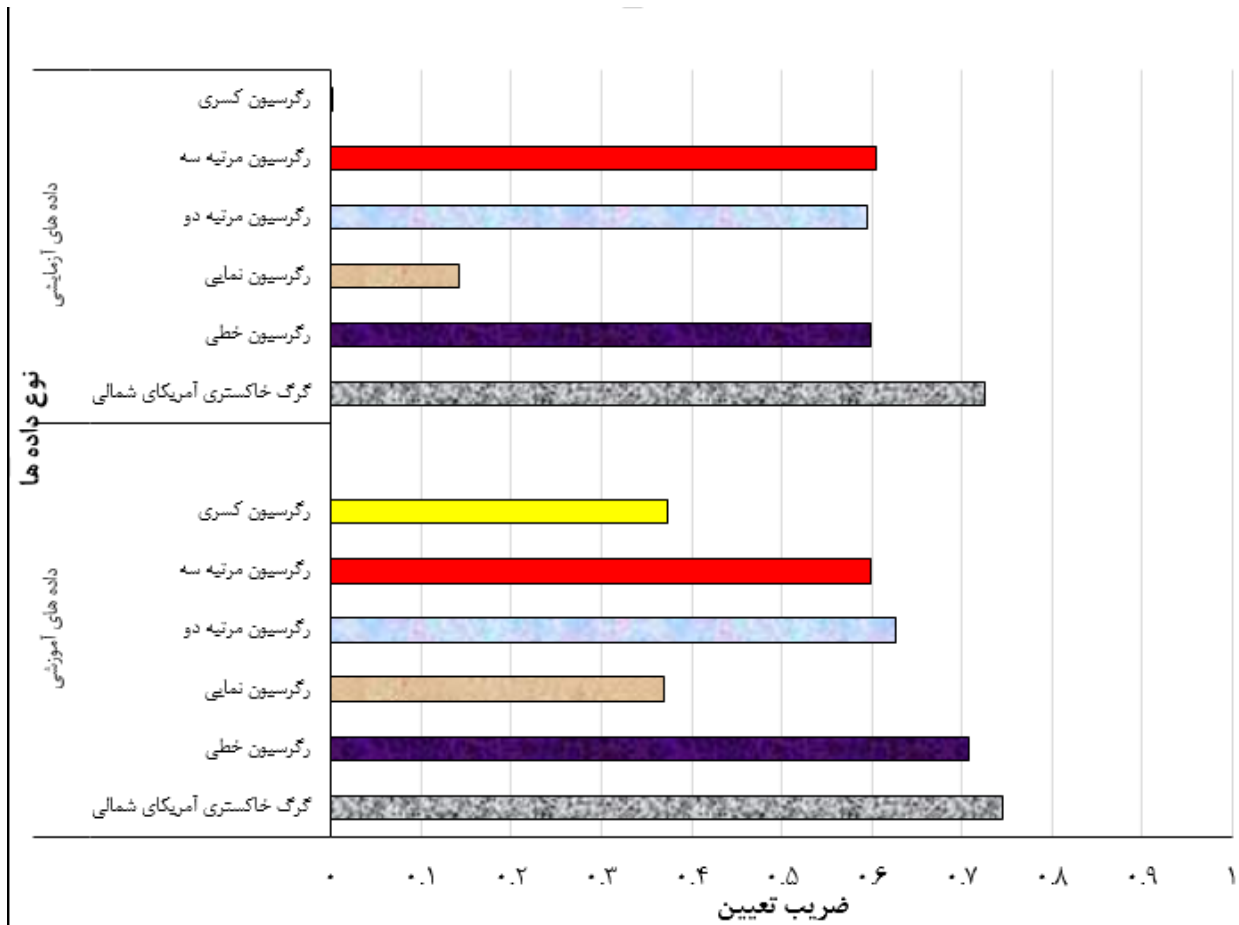
$$3R - Time(s / km) = (0.0005 \times StopDelay^3) + (0.2976 \times Lane^3) + (4.1589 \times Grade^3) + 38.4088 \quad (10)$$



شکل ۵. دقت مدل‌های پیش‌بینی بر اساس میانگین درصد قدر مطلق خطا
 Fig. 5. The mean absolute percentage error of prediction models

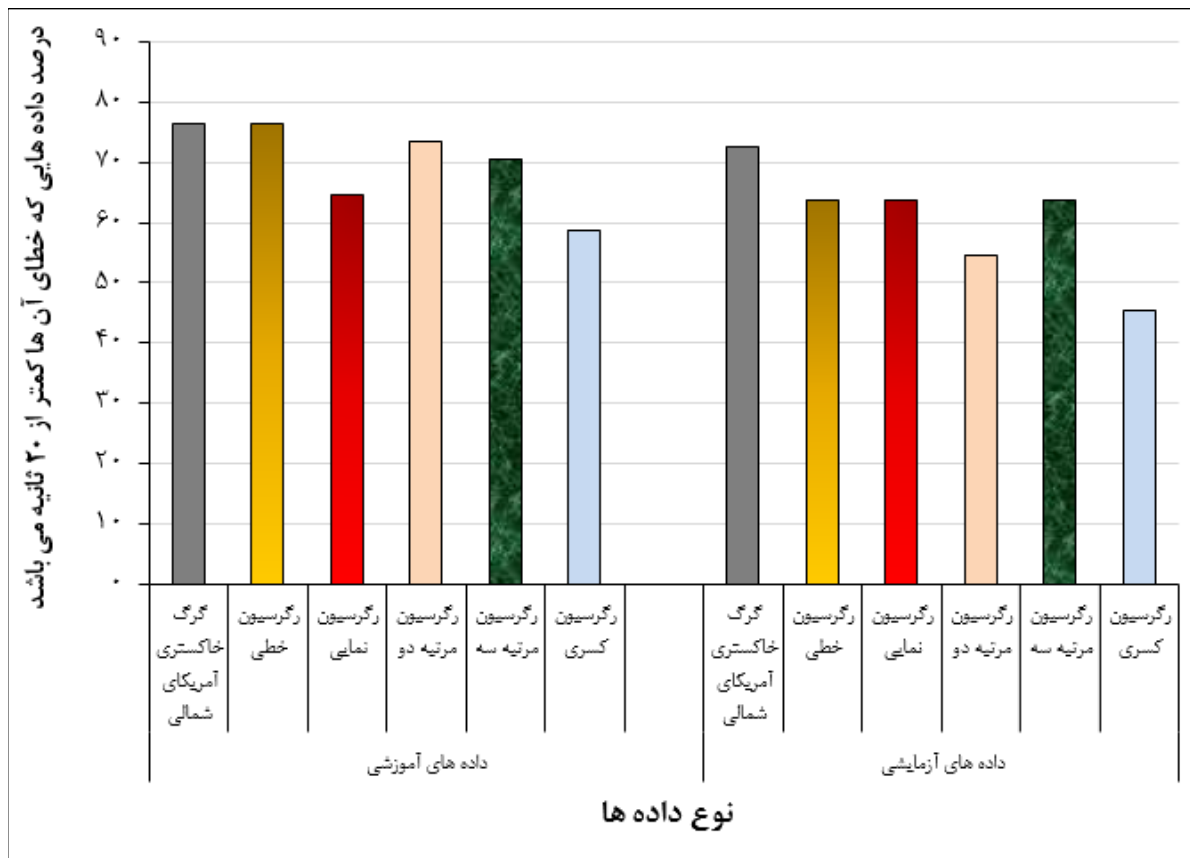
است. با توجه به نتایج این شاخص عملکردی، در داده‌های آموزشی فقط مدل‌های گرگ خاکستری آمریکای شمالی و رگرسیون خطی مقادیری بیش از ۰/۷ را دارا می‌باشند و در مرحله تولید مدل دارای دقت بیشتری نسبت به سایر روش‌ها هستند. با بررسی داده‌های آزمایشی مشخص می‌گردد که تنها ضریب تعیین الگوریتم گرگ خاکستری آمریکای شمالی بالای ۰/۷ به دست آمده است. در حالی که مقدار این شاخص برای تمامی مدل‌های رگرسیونی کوچک‌تر یا مساوی ۰/۶ می‌باشد. ضریب تعیین الگوریتم گرگ خاکستری آمریکای شمالی برای داده‌های آموزشی و آزمایشی به ترتیب برابر ۰/۷۴۶ و ۰/۷۲۵ است. از این رو می‌توان استدلال کرد که روش معرفی شده در این پژوهش بسیار دقیق‌تر از مدل‌های رگرسیونی بوده و همچنین می‌تواند برای پیش‌بینی زمان سفر که هدف اصلی این پژوهش است، نتایجی با دقت بالا ارائه کند.

نتایج دقت مدل‌های به دست آمده بر اساس میانگین درصد قدر مطلق خطا در شکل ۵ نمایش داده شده‌اند. مطابق این شاخص عملکردی هم در داده‌های آموزشی و هم در داده‌های آزمایشی عملکرد الگوریتم گرگ خاکستری آمریکای شمالی به مراتب از سایر روش‌ها بهتر می‌باشد. در داده‌های دیده نشده (آزمایشی)، الگوریتم گرگ خاکستری آمریکای شمالی تنها روشی است که میانگین درصد قدر مطلق خطای آن کمتر از ۰/۳۰ می‌باشد. در حالی که این شاخص برای تمامی مدل‌های رگرسیونی بالای ۰/۳۰ است. میانگین درصد قدر مطلق خطای الگوریتم گرگ خاکستری آمریکای شمالی برای داده‌های آزمایشی و آموزشی به ترتیب ۲۷/۴ و ۲۱/۲ درصد به دست آمده است. در بین مدل‌های رگرسیونی نیز، بیشترین دقت متعلق به رگرسیون مرتبه سه و رگرسیون خطی می‌باشد. ضریب تعیین روش‌های پیش‌بینی زمان سفر در شکل ۶ ارائه شده



شکل ۶. دقت مدل های پیش بینی بر اساس ضریب تعیین

Fig. 6. The accuracy of the prediction models concerning the coefficient of determination



شکل ۷. درصد داده‌هایی که خطای آن‌ها کمتر از ۲۰ ثانیه می‌باشد

Fig. 7. The percentage of data which their error is fewer than 20 seconds

عبارت دیگر، الگوریتم گرگ خاکستری آمریکای شمالی در ۷۲/۷٪ مواقع زمان سفر را با خطای کمتر از ۲۰ ثانیه پیش بینی می‌کند. بنابراین، نتایج این شاخص عملکردی مطابق نتایج سایر شاخص‌های عملکردی بوده و الگوریتم گرگ خاکستری آمریکای شمالی به نسبت سایر روش‌ها به مراتب بهتری از خود نشان داده است.

۳-۴- صحت سنجی خارجی

گلبریخ و تروپشا [۲۲] بیان کردند که صحت سنجی خارجی در مدل‌های پیش بینی امری ضروری است. زیرا در نظر گرفتن شاخص‌های عملکردی به تنهایی عموماً نمی‌تواند به طور کامل نشان‌دهنده صحیح بودن مدل باشد. از این رو، هنگامی مقادیر شاخص‌های عملکردی معتبر هستند که اعتبار مدل پیش‌بینی توسط صحت‌سنجی خارجی مورد پذیرش قرار گیرد.

در این تحقیق یک شاخص عملکردی جدید تحت عنوان درصد داده‌هایی که خطای آن‌ها کمتر از ۲۰ ثانیه می‌باشد (E20s) معرفی گردید. این شاخص جدید در رابطه (۵) ارائه شده است. نتایج درصد داده‌هایی که خطای آن‌ها کمتر از ۲۰ ثانیه است برای مدل‌های پیش‌بینی محاسبه شده و در شکل ۷ قابل مشاهده می‌باشد. بر اساس مقادیر این شاخص عملکردی، بیشترین دقت برای داده‌های آموزشی مرتبط به الگوریتم‌های گرگ خاکستری آمریکای شمالی و رگرسیون خطی است و درصد داده‌هایی که خطای آن‌ها کمتر از ۲۰ ثانیه می‌باشد برای داده‌های آموزشی در این دو روش ۷۴/۵٪ به دست آمده است. در داده‌های آزمایشی، الگوریتم گرگ خاکستری آمریکای شمالی دارای بیشترین دقت بوده و درصد داده‌هایی که خطای آن‌ها کمتر از ۲۰ ثانیه است، برای این روش ۷۲/۷٪ می‌باشد. به

جدول ۲. مقادیر پارامترهای شیب خط رگرسیون (k و k')، شاخص‌های عملکرد (m و n)

Table 2. The value of the regression slope parameters (k, k') and the performance indicators (m, n)

n	m	k'	k	نوع داده ها	روش
-۰/۳۳۸	-۰/۳۳۲	۰/۹۸۴	۰/۹۷۵	داده‌های آموزشی	گرگ خاکستری آمریکای شمالی
-۰/۳۵۴	-۰/۳۸۰	۰/۹۴۶	۱/۰۰۰	داده‌های آزمایشی	
-۰/۳۹۵	-۰/۴۱۴	۰/۹۵۳	۱/۰۰۰	داده‌های آموزشی	رگرسیون خطی
-۰/۶۰۲	-۰/۶۶۸	۰/۹۲۰	۰/۹۸۷	داده‌های آزمایشی	
-۱/۵۴۱	-۰/۷۱۳	۰/۸۹۹	۱/۰۰۰	داده‌های آموزشی	رگرسیون نمایی
-۵/۲۱۳	-۰/۹۶۸	۰/۸۶۷	۰/۹۸۷	داده‌های آزمایشی	
-۰/۳۶۴	-۰/۳۰۵	۱/۱۵۲	۰/۷۸۶	داده‌های آموزشی	رگرسیون مرتبه دو
۰/۹۹۹	-۰/۲۳۷	۱/۰۰۷	۰/۷۷۵	داده‌های آزمایشی	
-۰/۶۲۹	-۰/۶۷۲	۰/۹۳۶	۱/۰۰۰	داده‌های آموزشی	رگرسیون مرتبه سه
-۰/۴۸۵	-۰/۶۴۱	۰/۸۷۵	۰/۹۸۶	داده‌های آزمایشی	
-۱/۵۱۵	-۱/۶۸۳	۰/۹۰۰	۱/۰۰۰	داده‌های آموزشی	رگرسیون کسری
-۹۲۵/۶۴۴	-۱۳۶/۱۴۵	۰/۷۸۱	۰/۹۸۶	داده‌های آزمایشی	

در جدول ۲ ارائه شده‌اند. بر اساس نتایج این جدول، نتایج حاصل از رگرسیون مرتبه دو و رگرسیون کسری قابل اعتماد نیستند. اما نتایج الگوریتم گرگ خاکستری آمریکای شمالی، رگرسیون خطی، رگرسیون نمایی و رگرسیون مرتبه سه قابل اعتماد می‌باشند. از این رو روش معرفی شده در این تحقیق (الگوریتم گرگ خاکستری آمریکای شمالی) صحت سنجی گردید.

۴- نتیجه گیری

اطلاعات زمان سفر شبکه حمل و نقل کاربردهای متنوعی در سطوح مختلف شامل استفاده‌کنندگان از شبکه، تصمیم‌گیران خرد و تصمیم‌گیران کلان شبکه دارد. لذا یک مدل زمان سفر با دقت بالا می‌تواند استفاده گسترده‌ای در زمینه مدیریت ترافیک و برنامه ریزی سفر داشته باشد. در این پژوهش با طراحی یک آزمایش و داده برداری در معابر منتخب، پارامترهای تأثیرگذار بر زمان سفر معابر شهری در زمان غیر اوج تعیین شدند. پس از آماده سازی داده‌ها پارامترهای موثر بر زمان سفر مورد بررسی قرار گرفتند

پارامترهای صحت سنجی خارجی بر مبنای بازه‌های قابل قبول ارزیابی می‌شوند. به عبارت دیگر، اگر پارامترهای صحت سنجی خارجی در بازه قابل قبول قرار گیرند نتایج روش یادگیری ماشین قابل اعتماد بوده و در غیر این صورت نتایج روش مورد استفاده حتی در صورت دارا بودن مقادیر قابل قبول شاخص‌های عملکردی مورد استفاده در بخش قبل قابل اعتماد نمی‌باشند.

لذا جهت صحت سنجی مدل علاوه بر شاخص‌های عملکردی که در بخش پیشین به تفصیل مورد بررسی قرار گرفتند، از پارامترهای شیب خط رگرسیون (k و k') و شاخص‌های عملکرد (m و n) استفاده شد. مقادیر این شاخص‌ها مطابق دستور العمل‌های ارائه شده توسط گلبریخ و تروپشا [۲۲] و تروپشا و همکاران [۲۳] محاسبه گردید. بر اساس جزئیات ارائه شده در این پژوهش‌ها، بازه مناسب برای پارامترهای شیب خط رگرسیون از ۰/۸۵ تا ۱/۱۵ معرفی شده است. همچنین مقدار مناسب برای شاخص‌های عملکرد (m و n) مقادیر کمتر از ۱/۰ است. مقادیر پارامترهای شیب خط رگرسیون (k و k') و شاخص‌های عملکرد (m و n) برای مدل‌های پیش بینی زمان سفر

- Federal Highway Administration, 1998.
- [4] E. Jenelius, H.N. Koutsopoulos, Urban network travel time prediction based on a probabilistic principal component analysis model of probe data, *IEEE Transactions on Intelligent Transportation Systems*, 19(2) (2017) 436-445.
- [5] C.I. Van Hinsbergen, F. Sanders, Short term traffic prediction models, in: 14th World Congress on International Transport Systems, 2007, pp. A-X.
- [6] J. Van Lint, Reliable travel time prediction for freeways, Netherlands TRAIL Research School, 2004.
- [7] F. Zheng, H. Van Zuylen, X. Liu, A methodological framework of travel time distribution estimation for urban signalized arterial roads, *Transportation science*, 51(3) (2017) 893-917.
- [8] D. Nikovski, N. Nishiuma, Y. Goto, H. Kumazawa, Univariate short-term prediction of road travel times, in: *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, IEEE, 2005, pp. 1074-1079.
- [9] Y. Chen, H.J. Van Zuylen, Y. Qipeng, Travel time prediction on urban networks based on combining rough set with support vector machine, in: 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM), IEEE, 2010, pp. 586-589.
- [10] H.-E. Lin, R. Zito, M. Taylor, A review of travel-time prediction in transport and logistics, in: *Proceedings of the Eastern Asia Society for transportation studies*, Bangkok, Thailand, 2005, pp. 1433-1448.
- [11] L. Du, S. Peeta, Y.H. Kim, An adaptive information fusion model to predict the short-term link travel time distribution in dynamic traffic networks, *Transportation Research Part B: Methodological*, 46(1) (2012) 235-252.
- [12] J. Kwon, B. Coifman, P. Bickel, Day-to-day travel-time trends and travel-time prediction from loop-detector data, *Transportation Research Record*, 1717(1) (2000) 120-129.
- [13] M. Yildirimoglu, K. Ozbay, Comparative evaluation of probe-based travel time prediction techniques under varying traffic conditions, 2012.

و در نتیجه، متغیرهای زمان توقف تقاطع، تعداد خطوط و شیب به عنوان متغیرهای تاثیرگذار شناسایی شدند.

در این پژوهش جهت پیش بینی زمان سفر، یک روش یادگیری ماشین جدید به نام مدل یادگیری ماشین مبتنی بر الگوریتم گرگ خاکستری آمریکای شمالی معرفی گردید و دقت آن با روش‌های رگرسیونی مرسوم شامل رگرسیون‌های خطی، نمایی، چند جمله‌ای مرتبه ۲، چند جمله‌ای مرتبه ۳، و کسری مقایسه گردید. جهت مقایسه این روش و روش‌های رگرسیونی به کار گرفته شده از شاخص‌های عملکردی میانگین خطای حقیقی (MAE)، خطای جذر میانگین مربعات (RMSE)، میانگین درصد قدرمطلق خطا (MAPE)، ضریب تعیین (R^2) و درصد داده‌هایی که خطای آن‌ها کمتر از ۲۰ ثانیه می‌باشد (E20S) استفاده گردید. مقادیر این شاخص‌های عملکردی برای داده‌های آموزشی در الگوریتم گرگ خاکستری آمریکای شمالی به ترتیب معادل ۱/۰/۱، ۱۳/۸، ۲۱/۲، ۰/۷۴۶ و ۷۶/۵ می‌باشد. همچنین مقادیر این شاخص‌های عملکردی برای داده‌های آزمایشی در الگوریتم گرگ خاکستری آمریکای شمالی به ترتیب برابر ۲۷/۴، ۱۵/۲، ۲۷/۴ و ۰/۷۲۵ و ۷۲/۷ می‌باشد که مقادیر روش‌های رگرسیونی در تمامی شاخص‌ها بهتر بوده و نشان دهنده عملکرد مناسب مدل معرفی شده را نشان می‌دهد. در نهایت از رویکرد توانمند صحت‌سنجی خارجی استفاده گردید و پارامترهای شیب خط رگرسیون (k و k') و شاخص‌های عملکرد (m و n) محاسبه شدند. با توجه به نتایج حاصل از این رویکرد، مقادیر به دست آمده این شاخص‌ها برای الگوریتم گرگ خاکستری آمریکای شمالی همگی در محدوده مجاز معرفی شده بوده و صحت مدل‌ها تأیید شد. به علاوه، این امر نشان دهنده این موضوع است که مدل جدید معرفی شده می‌تواند برای پیش بینی زمان سفر که هدف اصلی این پژوهش است، نتایجی معتبر ارائه کند.

منابع

- [1] I. Sanaullah, M. Quddus, M. Enoch, Developing travel time estimation methods using sparse GPS data, *Journal of Intelligent Transportation Systems*, 20(6) (2016) 532-544.
- [2] W. Chun-Hsin, H. Jan-Ming, D.T. Lee, Travel-time prediction with support vector regression, *IEEE Transactions on Intelligent Transportation Systems*, 5(4) (2004) 276-281.
- [3] S.M. Turner, W.L. Eisele, R.J. Benz, D.J. Holdener, *Travel time data collection handbook*, United States.

- geometric design of four-lane divided highways, *Transport*, 34(4) (2019) 425-436.
- [19] C. Vilarinho, J.P. Tavares, R.J. Rossetti, Intelligent traffic lights: Green time period negotiaton, *Transportation research procedia*, 22 (2017) 325-334.
- [20] H.C. Manual, HCM2010, Transportation Research Board, National Research Council, Washington, DC, (2010) 1207.
- [21] J. Pierezan, L.D.S. Coelho, Coyote optimization algorithm: a new metaheuristic for global optimization problems, in: 2018 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2018, pp. 1-8.
- [22] A. Golbraikh, A. Tropsha, Beware of $q^2!$, *Journal of molecular graphics and modelling*, 20(4) (2002) 269-276.
- [23] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, *QSAR & Combinatorial Science*, 22(1) (2003) 69-77.
- [14] H. Naseri, H. Jahanbakhsh, F. Moghadas Nejad, Developing a novel machine learning method to predict the compressive strength of fly ash concrete in different ages, *AUT Journal of Civil Engineering*, (2019).
- [15] E.M. Golafshani, A. Behnood, Estimating the optimal mix design of silica fume concrete using biogeography-based programming, *Cement and Concrete Composites*, 96 (2019) 95-105.
- [16] S. Oh, Y.-J. Byon, K. Jang, H. Yeo, Short-term travel-time prediction on highway: A review on model-based approach, *KSCE Journal of Civil Engineering*, 22(1) (2018) 298-310.
- [17] M.M.M. Abdel-Aal, Factors Affecting Road Capacity Under non-Ideal Conditions in Egypt Mounir Mahmoud Moghazy Abdel-Aal, Ahmed Ebrahim Abu El-Maaty, Hassan Abdel-rahman Abo Samra, *Nova*, 7(1) (2018) 1-13.
- [18] G. Sil, A. Maji, S. Nama, A.K. Maurya, Operating speed prediction model as a tool for consistency based

چگونه به این مقاله ارجاع دهیم

A. Golroo , A. H. Fani, H. Naseri , *Travel Time Modelling of Urban Roads By Application of Coyote Optimization-based Machine Learning Method, Amirkabir J. Civil Eng., 53(9) (2021) 3649-3664.*

DOI: [10.22060/ceej.2020.17991.6730](https://doi.org/10.22060/ceej.2020.17991.6730)



