



University of Tehran Press

Town & Country Planning

Home Page: <https://jtcp@ut.ac.ir>

Online ISSN: 2423-6268

Estimating the travel modes using machine learning algorithms for sustainable urban transportation

Omid Shamohammadi¹ | Parham Pahlavani^{2*} | Mohammad Ali Sharifi³

1. Department of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. Email: omid.shamohammadi@ut.ac.ir

2. Corresponding Author, Department of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. Email: pahlavani@ut.ac.ir

3. Department of Geodesy, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. E-mail: sharifi@ut.ac.ir

ARTICLE INFO

Article type:
Research Article

Article History:
Received December 29, 2023
Revised January 30, 2024
Accepted February 04, 2024
Published online 17 February 2025

Keywords:
Sustainable urban transportation,
Travel modes,
Streaming GNSS,
Feature extraction,
Supervised classification algorithms.

ABSTRACT

A significant portion of daily urban intra-city trips is aimed at accessing services, amenities, and goods that are not readily available in a specific area. Therefore, analyzing the frequently used trajectory and identifying the reasons for high traffic volumes on these trajectories can lead to a more accurate distribution of facilities, services, and proper land use allocation with the goal of reducing the number, distance, and time of intra-city trips. With the advent of Global Navigation Satellite Systems (GNSS) positioning sensors on smartphones, the real-time collection of individuals' positions, speed, acceleration, and more has become possible. Consequently, this research has sought to examine the possibility of using GNSS data recorded by smartphones to identify the transportation mode used by the user through four supervised machine learning models named Random Forest (RF), Gradient Boosting (GB), eXtreme Gradient Boosting (XGB), and Light Gradient Boosting Model (LightGM). For this purpose, two datasets, Microsoft Geolife and MTL 2017, which possess the necessary features for this goal, have been used as the input data. After extracting the features of each trajectory from these two datasets, with the aim of improving the models' performance and reducing processing time, among the available features, the most important ones have been identified, and classification has been applied based on them. Among the models used, the LightGM and XGB models achieved the best performance for the first and second datasets with respective F1-Scores of 92.57% and 92.67% for test data. Out of a total of 1349 trips, this algorithm accurately estimated 1250 trips, contributing to sustainable urban transportation.

Cite this article: Shamohammadi, O.; Pahlavani, P. & Sharifi, M. A. (2024-25). Estimating the travel modes using machine learning algorithms for sustainable urban transportation. *Town and Country Planning*.16 (2), 239-254. <http://doi.org/10.22059/jtcp.2024.370374.670427>



© The Author(s).

Publisher: University of Tehran Press.

DOI: <http://doi.org/10.22059/jtcp.2024.370374.670427>



انتشارات دانشگاه تهران

آمایش سرزمین

سایت نشریه: <https://jtcp@ut.ac.ir/>

شاپا الکترونیکی: ۶۲۶۸-۲۴۲۳

تخمین حالت‌های سفر به کمک الگوریتم‌های یادگیری ماشین به منظور حمل‌ونقل پایدار شهری

امید شامحمدی^۱ | پرهام پهلونی^{۲*} | محمدعلی شریفی^۳

۱. گروه GIS، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشکدگان فنی، دانشگاه تهران، تهران، ایران. رایانامه: omid.shamohammadi@ut.ac.ir
۲. نویسنده مسئول، گروه GIS، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشکدگان فنی، دانشگاه تهران، تهران، ایران. رایانامه: pahlavani@ut.ac.ir
۳. گروه ژئودزی، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشکدگان فنی، دانشگاه تهران، تهران، ایران. رایانامه: sharifi@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله:

پژوهشی

تاریخ‌های مقاله:

تاریخ دریافت: ۱۴۰۲/۱۰/۰۹

تاریخ بازنگری: ۱۴۰۲/۱۱/۱۰

تاریخ پذیرش: ۱۴۰۲/۱۱/۱۵

تاریخ انتشار: ۱۴۰۳/۱۱/۲۹

کلیدواژه:

استخراج ویژگی، الگوریتم‌های کلاسه‌بندی نظارت‌شده، جریان داده‌های GNSS، حالت‌های سفر، حمل‌ونقل پایدار شهری.

بخش زیادی از سفرهای روزانه درون‌شهری با هدف دسترسی به خدمات، سرویس‌ها، و کالاهایی است که امکان تهیه آن‌ها در یک منطقه مشخص وجود ندارد. از این رو، آنالیز مسیرهای پرتدد و علت حجم بالای ترافیک در آن‌ها می‌تواند به توزیع صحیح‌تر امکانات، خدمات، و اختصاص کاربری مناسب به زمین‌ها با هدف کاهش تعداد و مسافت و زمان سفرهای درون‌شهری منجر شود. با ظهور سنسورهای تعیین موقعیت مبتنی بر سامانه‌های جهانی ناوبری ماهواره‌ای بر تلفن‌های همراه، امکان ثبت آنی مواردی چون موقعیت و سرعت و شتاب افراد به صورت آنی فراهم شده است. بنابراین، در این پژوهش تلاش شده امکان استفاده از داده‌های GNSS ثبت‌شده توسط تلفن همراه با هدف شناسایی حالت حمل‌ونقلی که کاربر از آن استفاده کرده است توسط چهار مدل یادگیری ماشین نظارت‌شده با نام‌های *RF*، *GB*، *XGB*، و *LightGM* مورد بررسی قرار گیرد. بدین منظور، دو مجموعه داده میکروسافت *Geolife* و *MTL 2017*، که از ویژگی‌های لازم برای این هدف برخوردارند، به عنوان داده ورودی استفاده شد. پس از استخراج ویژگی‌های هر مسیر از این دو مجموعه داده، با هدف بهبود کارایی مدل‌ها و کاهش حجم پردازش، از میان ویژگی‌های موجود ویژگی‌های مهم‌تر شناسایی و کلاسه‌بندی بر اساس آن‌ها اعمال شد. بین مدل‌های استفاده‌شده، مدل *LightGM* برای مجموعه داده اول با کسب مقدار ۹۲/۵۷ درصد در *F1-Score* برای داده‌های تست و مدل *XGB* برای مجموعه داده دوم با کسب مقدار ۹۲/۶۷ درصد در *F1-Score* برای داده‌های تست بین مدل‌های موجود بهترین عملکرد را داشته‌اند. از تعداد ۱۳۴۹ سفر موجود، ۱۲۵۰ سفر توسط این الگوریتم به‌درستی تخمین زده شد که منجر به حمل‌ونقل پایدار شهری می‌شود.

استناد: شامحمدی، امید؛ پهلونی، پرهام و شریفی، محمدعلی (۱۴۰۳). تخمین حالت‌های سفر به کمک الگوریتم‌های یادگیری ماشین به منظور حمل‌ونقل پایدار

شهری. *آمایش سرزمین*، ۱۶ (۲): ۲۵۴-۲۳۹. <http://doi.org/10.22059/jtcp.2024.370374.670427>

© نویسندگان. ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <http://doi.org/10.22059/jtcp.2024.370374.670427>



مقدمه و بیان مسئله

انسان همواره برای رفع نیازهای خود، زنده ماندن، و تجسس و تفتیش در محل زندگی از ابتدای خلقت تا کنون در حال جابه‌جایی بوده است (Schlebusch & Jakobsson, 2018). گسترش محیط‌های شهری و افزایش بی‌رویه جمعیت در کلان‌شهرها سبب افزایش تقاضا برای جابه‌جایی‌های درون‌شهری شده است که از آثار آن می‌توان به ایجاد ترافیک، آلودگی هوا، هدررفت زمان و انرژی اشاره کرد. از این رو، ارزیابی دقیق رفتار حرکتی افراد در جامعه به منظور درک حرکت و تأثیر آن بر فعالیت‌های شهروندان و حفاظت از محیط زیست امری ضروری برای برنامه‌ریزان شهری به منظور حمل‌ونقل پایدار شهری است (Martín-Baos et al., 2023).

حمل‌ونقل دومین هزینه کلان در زندگی افراد است. در مطالعات اتحادیه اروپا و آمریکا، شهروندان به ترتیب به طور متوسط ۱۵ و ۲۵ درصد از کل هزینه‌های زندگی خود را صرف حمل‌ونقل می‌کنند (Wang et al., 2018). این مقدار می‌تواند معیاری از اهمیت حمل‌ونقل در فعالیت‌های انسانی در نظر گرفته شود.

بخش بسیاری از سفرهای صورت گرفته در طول روز با هدف دسترسی به کالاهای و امکانات و خدماتی است که افراد در نزدیکی محل سکونت خود به آن‌ها دسترسی ندارند. مثلاً بخشی از سفرهای شهری با هدف دسترسی به امکانات یک ورزش خاص، نظیر زمین تنیس، صورت می‌گیرد؛ حال آنکه توزیع مناسب این امکانات در سطح یک شهر به طرز چشمگیری می‌تواند منجر به کاهش نیاز به حمل‌ونقل شود (Sun et al., 2023). آنالیز جابه‌جایی‌های صورت گرفته در سطح شهر و همچنین فرم حمل‌ونقل مورد استفاده در هر یک از آن‌ها، علاوه بر آنکه می‌تواند درک مناسبی از مسیرهای پرتردد فراهم آورد، در شناسایی هدف جابه‌جایی صورت گرفته نیز تأثیرگذار است (Kashifi et al., 2022). از این رو مدیران شهری می‌توانند متناسب با اهداف جابه‌جایی‌ها به بهبود توزیع امکانات در سطح شهر با ابزار تخصیص کاربری زمین بپردازند. چنین رخدادی باعث کاهش حجم سفرهای روزانه، مصرف انرژی، تولید گازهای گلخانه‌ای، و صرفه‌جویی در زمان می‌شود.

افراد برای سفرهای خود از روش‌های مختلف حمل‌ونقل - مانند پیاده‌روی، دوچرخه‌سواری، ماشین، اتوبوس، قطار - استفاده می‌کنند که به آن حالت‌های سفر می‌گویند. نحوه حرکت و طریقه استفاده شهروندان از حالات حمل‌ونقلی از الگوی خاصی پیروی می‌کند. شناخت این الگوها موجب شناسایی حالت‌های حمل‌ونقلی شهروندان خواهد شد. شناسایی دقیق حالت حمل‌ونقلی عاملی ضروری برای دولت‌ها، شرکت‌ها، و مؤسسات تحقیقاتی است تا رفتارهای انسانی را درک و سیستم‌های حمل‌ونقلی هوشمندی را راه‌اندازی کنند (Yazdizadeh et al., 2021). به منظور شناسایی و تخمین حالت‌های حمل‌ونقلی، اطلاعات دقیق سفرهای انجام‌شده افراد مورد نیاز است. گسترش استفاده از تلفن‌های هوشمند در دهه دوم قرن جاری، که جمع‌آوری موقعیت لحظه‌ای را به صورت برخط با استفاده از سنسورهای GNSS محقق کرده، منبع مناسبی برای استخراج این ویژگی‌ها و اطلاعات فراهم آورده است (Hasan et al., 2022). تحلیل و آنالیز داده‌های این سنسورها با هدف شناسایی حالت‌های سفر توسط محققان مختلف صورت پذیرفته است. با این حال، چنین تحلیلی به دلیل ناکافی بودن روش‌های تشخیص حالت حمل‌ونقل موجود همچنان نیاز به مطالعه بیشتر دارد (Al Momin et al., 2022).

با توجه به توانایی الگوریتم‌های یادگیری ماشین به عنوان ابزاری مناسب برای طبقه‌بندی اطلاعات گوناگون، در این پژوهش از چهار مدل یادگیری ماشین برای شناسایی حالت سفر استفاده شد. بدین منظور، مجموعه‌ای از ویژگی‌ها از داده‌های ثبت‌شده توسط سنسور GNSS تلفن‌های همراه کاربران استخراج شد. با هدف بهبود عملکرد و کاهش حجم پردازش‌های صورت گرفته شناسایی ویژگی‌های مؤثر و کاهش تعداد ویژگی‌ها با چهار روش انتخاب ویژگی با استفاده از K-best، استفاده از مدل RFECV، استفاده از Bult-in، و در نهایت انتخاب ویژگی با استفاده از الگوریتم ژنتیک مد نظر قرار داده شد.

پیشینه پژوهش

با پیشرفت و هوشمندی تلفن‌های همراه، تعداد حسگرهای مورد استفاده در ساخت آن‌ها رو به افزایش رفت. مطالعات زیادی از داده‌ها و حسگرهای متفاوت، برگرفته از داده‌های ثبت‌شده توسط تلفن‌های همراه هوشمند، برای شناسایی حالت‌های حمل‌ونقلی

استفاده کرده‌اند. برخی از مطالعات پیشین چالش استفاده تنها از داده‌های GPS^۱ به صورت خام، بدون استفاده از داده‌های مکمل و حسگرهای دیگر، را مورد بررسی و پیش‌بینی قرار داده‌اند (Nawaz et al., 2020). مثلاً Li و همکارانش (۲۰۲۰) از مجموعه داده‌های GPS ثبت‌شده در شهر پکن کشور چین، Song و همکارانش (۲۰۱۶) از داده‌های GPS خام ثبت‌شده در کشور ژاپن، Yazdizadeh و همکارانش (۲۰۲۱) از داده‌های ثبت‌شده در شهر مونترال کشور کانادا، و Bolbol و همکارانش (۲۰۱۲) از داده‌های شهر لندن کشور انگلیس استفاده کردند.

در برخی مطالعات، از حسگرهای دیگر بدون استفاده و وابستگی به حسگر GPS استفاده شده است. مثلاً Friedrich و همکارانش (۲۰۲۰) از مجموعه داده‌های مرکز تحقیقاتی SHL^۲ شامل حسگرهای شتاب‌سنج، ژيروسکوپ، شتاب خطی، مغناطیس‌سنج، فشارسنج، جهت‌سنج، گرانش‌سنج، و موارد دیگر استفاده کردند.

مطالعات دیگری از حسگرهای هوشمند به عنوان حسگرهای مکمل در کنار داده‌های اخذشده از حسگر GPS استفاده کرده‌اند. Bantis & Haworth (۲۰۱۷) از حسگر شتاب‌سنج در کنار حسگر GPS، Byon & Liang (۲۰۱۴) از حسگرهای شتاب‌سنج و مغناطیس‌سنج و GPS، و در Jahangiri & Rakha (۲۰۱۵) از حسگرهای شتاب‌سنج و ژيروسکوپ و بردار چرخش در کنار حسگر GPS استفاده کردند.

برخی از مطالعات داده‌های محتوایی اجتماعی-جمعیتی نظیر سن، جنسیت، شغل، و اطلاعات اقتصادی-اجتماعی را در کنار داده‌های حاصل از حسگرهای تلفن‌های همراه استفاده کردند (Yazdizadeh et al., 2019).

الگوریتم‌های متفاوتی به جهت کلاسه‌بندی حالت‌های حمل‌ونقلی در مطالعات استفاده شده است که به طور کلی می‌توان آن‌ها را به چهار دسته اصلی روش‌های مبتنی بر قانون، روش‌های مبتنی بر منطق فازی، الگوریتم‌های یادگیری ماشین، و الگوریتم‌های یادگیری عمیق طبقه‌بندی کرد.

در روش مبتنی بر قانون، حالت سفر بر اساس معیارهای ازپیش‌تعیین‌شده و به صورت مرحله‌ای به کلاسه‌بندی می‌پردازد (Gong et al., 2012). مثلاً Bohte & Maat (۲۰۰۹) اعلام کردند اگر میانگین سرعت کمتر از ۲۵ کیلومتر بر ساعت و بیشینه سرعت کمتر از ۴۵ کیلومتر بر ساعت باشد حالت حمل‌ونقلی دوجرخه‌سواری پیش‌بینی می‌شود. در برخی دیگر از مطالعات منطق فازی اجرا شده است. در این مطالعات، با تعریف توابع عضویت، تمایز بین حالت‌های حمل‌ونقلی مختلف صورت گرفته است؛ به گونه‌ای که احتمال رخداد یک حالت حمل‌ونقلی خاص بین ۰ و ۱ تنظیم شده است (Sauerländer-Biebl et al., 2017).

امروزه روش‌های یادگیری ماشین و یادگیری عمیق بیش از باقی الگوریتم‌ها مورد استفاده قرار می‌گیرد. پیش‌بینی حالت‌های حمل‌ونقلی نیز از این مورد پشتیبانی می‌کند. چندین نوع الگوریتم در مطالعات بررسی شده و مورد استفاده و آزمایش قرار گرفته‌اند. در برخی مطالعات (Nawaz et al., 2020; Yazdizadeh et al., 2019) الگوریتم‌های یادگیری ماشین مدل‌های RF و در برخی دیگر (Li et al., 2021; Wang et al., 2018) مدل‌های GB استفاده شده است.

در کل، استفاده از روش‌های مبتنی بر قانون و تئوری فازی نیازمند تعیین قوانین و توابع عضویت به تفکیک هر یک از ویژگی‌های استخراجی است. این مورد در صورت افزایش تعداد ویژگی‌ها با مشکل روبه‌رو می‌شود. همین‌طور در روش‌های یادگیری عمیق ویژگی‌ها در لایه موسوم به لایه پنهان استخراج خواهد شد و مهندسی ویژگی نقش کلیدی در پیش‌بینی خواهد داشت؛ به گونه‌ای که در بیشتر موارد موجب کاهش دقت در مقایسه با استخراج ویژگی دستی می‌شود. همچنین اگرچه در تحقیقات پیشین طبقه‌بندی حالت‌های حمل‌ونقل بررسی شده است، در این تحقیق تلاش شده مقایسه‌ای از عملکرد چندین روش طبقه‌بندی مطرح همراه با انتخاب ویژگی‌های مؤثر صورت پذیرد. به منظور بهبود جامعیت مقایسه صورت‌گرفته، از تعداد ویژگی‌های بیشتری استفاده شده است تا ارزیابی دقیق‌تری بر عملکرد الگوریتم‌های انتخاب ویژگی در زمانی که کاهش ابعاد بیشتری مد نظر است صورت پذیرد.

محدوده و قلمرو مورد مطالعه

در این تحقیق برای ارزیابی روش پیشنهادی از مجموعه داده‌های پروژه Geolife که توسط مرکز تحقیقاتی مایکروسافت تهیه و

1. global positioning system

2. sussex-huawei locomotion-transportation

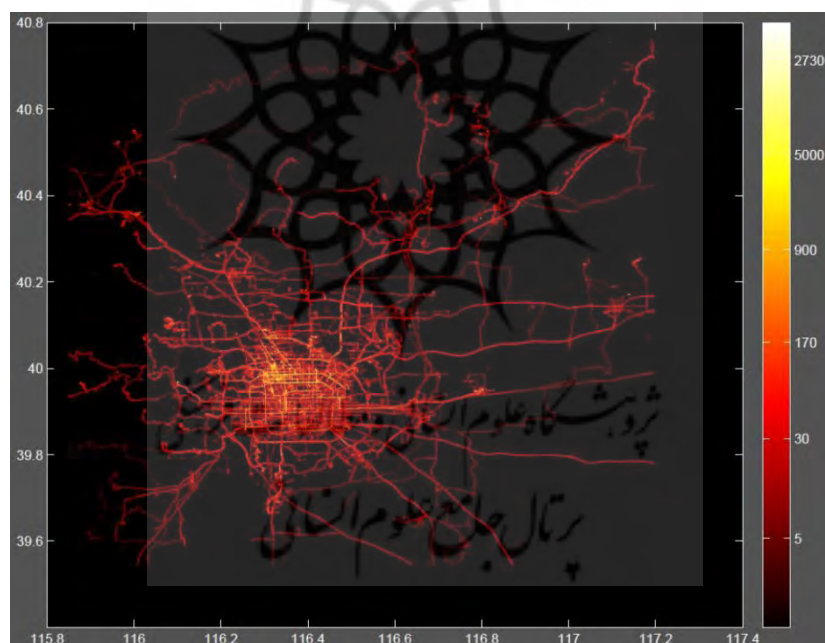
ذخیره شده است استفاده شد. این مجموعه داده طیف وسیعی از رفت‌وآمدهای کاربران با مقاصد مختلف از جمله مکان‌های کاری، استادیوم‌ها، مراکز خرید و سرگرمی را شامل می‌شود. ذخیره‌سازی این داده‌ها به صورت زمانی-مکانی در مدت پنج سال (آوریل ۲۰۰۷ تا آگوست ۲۰۱۲) و توسط ۱۸۲ نفر انجام پذیرفته است که ۶۹ کاربر در این مجموعه داده سفرهای خود را برچسب‌گذاری کرده‌اند. اطلاعات مربوط به این مجموعه داده در جدول ۱ نشان داده شده است.

این مجموعه داده به صورت منبع‌باز و رایگان در صفحه سازمان مایکروسافت قابل بارگیری و استفاده است (Geolife Dataset, Accessed June, 2022). خط سیر ثبت‌شده در این مجموعه داده توسط دستگاه‌های ثبت موقعیت مکانی PDA^۱ و تلفن همراه هوشمند با میزان نمونه‌برداری متفاوت ثبت و ذخیره شده است. این مجموعه داده در بیش از سی شهر در کشورهای مختلف، مانند چین و ایالات متحده آمریکا و برخی از کشورهای اروپایی، جمع‌آوری شده است. شکل ۱ توزیع مجموعه داده‌ها در پکن را نشان می‌دهد؛ شهری که بیشتر داده‌ها در آن جمع‌آوری شده است.

در این مجموعه داده حالت‌های سفری که برچسب‌گذاری شده‌اند عبارت‌اند از: پیاده‌روی، دوچرخه، اتومبیل، اتوبوس، قطار، مترو، هواپیما، دویدن، موتورسیکلت، تاکسی، جابه‌جایی با استفاده از قایق.

جدول ۱. مشخصات سفرهای موجود در داده Geolife

تعداد کل کاربران	تعداد کل سفرهای انجام گرفته	کل مسافت‌های پیموده شده (بر حسب کیلومتر)	کل زمان‌های پیموده شده (بر حسب ساعت)
۱۸۲	۱۷۶۲۱	۱۲۹۳۹۵۱۱	۵۰۱۷۶

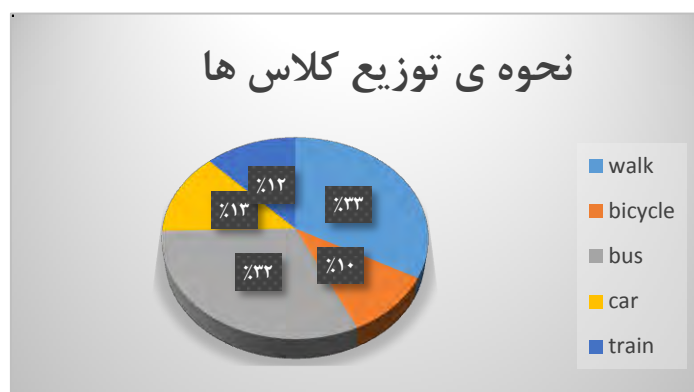


شکل ۱. توزیع داده‌ها در شهر پکن

در این تحقیق، از حالت‌های حمل‌ونقل غیرزمینی اعم از هواپیما و قایق صرف‌نظر شد. در بین حالت‌های زمینی نیز با توجه به کافی نبودن داده‌ها در حالت‌های دویدن و موتورسیکلت این دو حالت نیز کنار گذاشته شدند. همچنین با توجه به فایل راهنمای موجود خود مایکروسافت برای داده‌های Geolife، دو حالت استفاده از اتومبیل و تاکسی به عنوان یک حالت کلی استفاده از اتومبیل در نظر گرفته شد. در ادامه، با توجه به اینکه در شهر پکن مسیر قطارهای شهری و مترو به هم متصل و یکپارچه‌اند و از طرفی چون استاندارد خاصی برای تفکیک این دو وجود ندارد و شاید در برچسب‌گذاری اشتباه کنند، این دو حالت نیز به عنوان حالت حمل‌ونقل استفاده از قطار در نظر گرفته شد. در نهایت با حذف و ادغام حالت‌های گفته‌شده پنج حالت

1. personal digital assistant

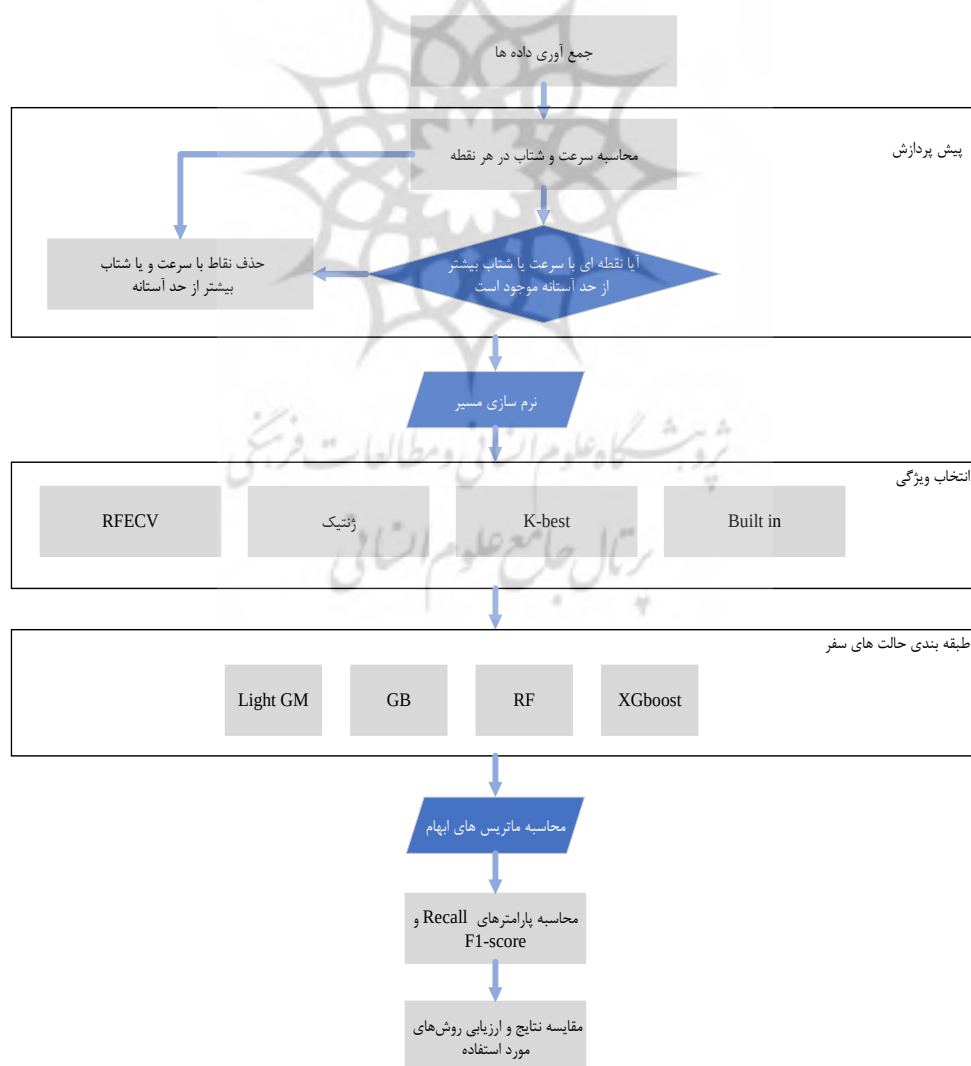
پیاده‌روی، دوچرخه، اتوبوس، ماشین، و قطار به عنوان کلاس‌های نهایی در این مطالعه مورد بررسی قرار گرفت. در شکل ۲ نحوه توزیع داده‌ها در هر یک از حالت‌ها به نمایش گذاشته شده است.



شکل ۲. نحوه توزیع کلاس‌ها

روش پژوهش

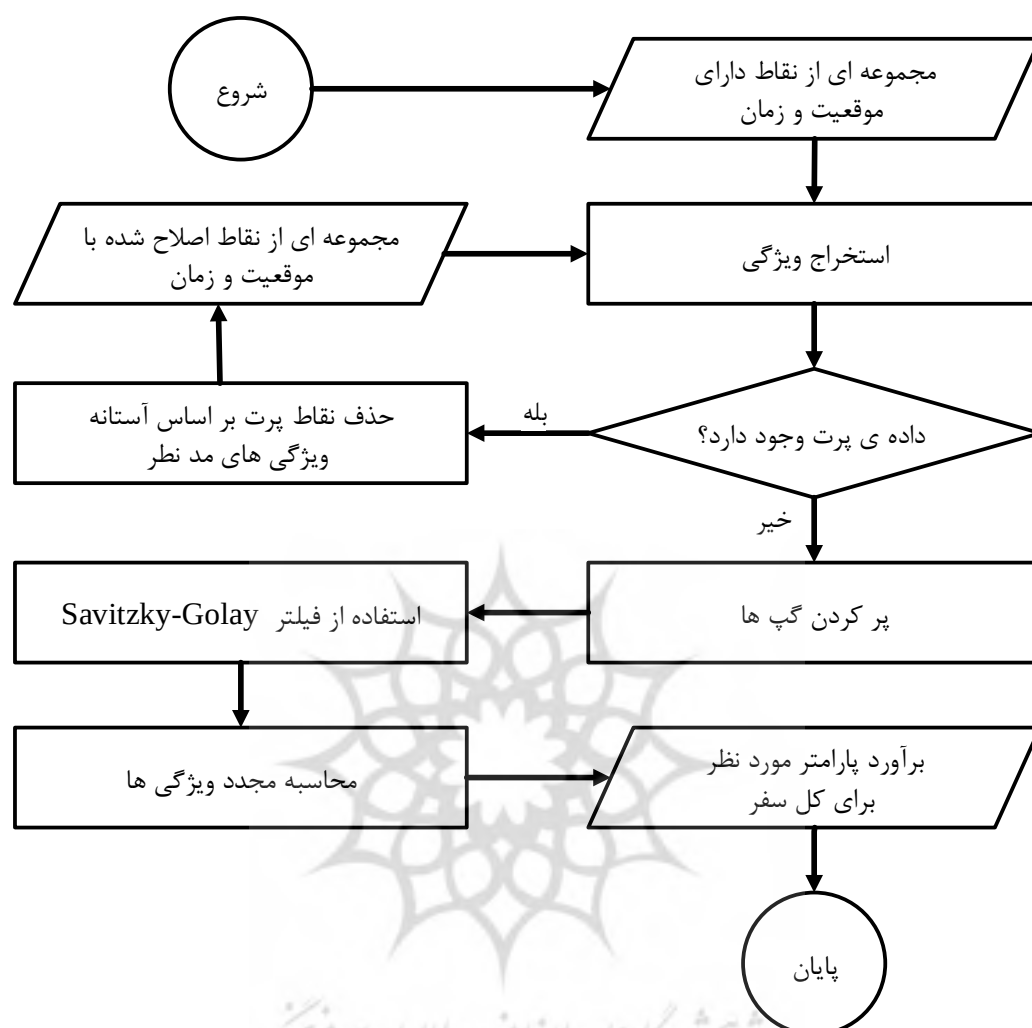
در شکل ۳ فلوجارت روش تحقیق برای پیش‌بینی و شناسایی حالت‌های حمل‌ونقل نشان داده شده است.



شکل ۳. فلوجارت مربوط به روش تحقیق

تجزیه و تحلیل داده‌ها

برای پیش‌پردازش داده‌ها به ترتیب مراحل زیر صورت گرفت:



شکل ۴. فلوچارت پیش‌پردازش‌ها

در این مرحله ابتدا با استفاده از نقاط خام موجود در داده‌ها برای هر یک از نقاط موجود سرعت و شتاب محاسبه می‌شود. سپس با استفاده از حدآستانه تعریف‌شده در جدول ۲ نقاطی از این حدآستانه که دارای سرعت و شتاب بیشتری هستند از مجموعه داده‌ها حذف شدند.

جدول ۲. مقادیر حدآستانه برای شتاب و سرعت

حالت‌های حمل‌ونقلی	حدآستانه سرعت (m/s)	حدآستانه شتاب (m/s ²)
پیاده‌روی	۷	۳
دوچرخه‌سواری	۱۲	۳
اتوبوس	۳۴	۲
اتومبیل	۵۰	۱۰
قطار	۳۴	۳

سپس نقاطی که سرعت و شتاب آن‌ها از حدآستانه بیشتر نیست بررسی می‌شوند که آیا از نظر زمانی گپ دارند یا خیر. در هر سفر این پارامتر بررسی می‌شود. اگر بین دو نقطه بیش از مدت‌زمان مشخص‌شده گذشته باشد با استفاده از تابع درون‌یابی گپ

تشخیص داده شده بین دو نقطه پر شده است. در مرحله بعد همه مسیرهای موجود با استفاده از فیلتر Savitzky-golay نرم شده‌اند. پس از این مرحله بار دیگر همه ویژگی‌های تعریف شده برای همه مسیرهای موجود محاسبه شده‌اند. در نهایت یک فایل اکسل با ۱۳۴۹ رکورد ذخیره شده است که هر رکورد نشان دهنده یک سفر مربوط به کاربر با شماره مورد نظر است.

استخراج ویژگی‌ها

در ادامه پس از حذف نقاط اشتباه و پر کردن گپ‌های موجود بین نقاط حال ویژگی‌های گفته شده برای همه سفرهای موجود استخراج شده است. در نهایت برای هر یک از سفرها تعداد ۵۲ ویژگی متفاوت استخراج شده است. در جدول ۲ همه ویژگی‌های استخراج شده برای هر یک از سفرها نشان داده شده است.

جدول ۳. ویژگی‌های استخراج شده برای هر سفر

شماره	نام ویژگی	شماره	نام ویژگی
۱	TotalDistance(m)	۲۷	VelMax(m/s)
۲	StraightPAR	۲۸	VelSkew
۳	VelMean(m/s)	۲۹	VelKurtosis
۴	VelStd(m/s)	۳۰	VelIqr
۵	VelMed(m/s)	۳۱	VelChaangePar
۶	jerk60	۳۲	acc40
۷	jerk80	۳۳	acc60
شماره	نام ویژگی	شماره	نام ویژگی
۸	bear20	۳۴	acc80
۹	bear40	۳۵	jerk20
۱۰	bear60	۳۶	jerk40
۱۱	accKurtosis	۳۷	jerkMed(m/s^3)
۱۲	accIqr	۳۸	jerkMax(m/s^3)
۱۳	accChaangePar	۳۹	jerkSkew
۱۴	jerkMean(m/s^3)	۴۰	jerkKurtosis
۱۵	jerkStd(m/s^3)	۴۱	jerkIqr
۱۶	jerkChaangePar	۴۲	HCR
۱۷	bearMean(deg)	۴۳	SR
۱۸	bearStd(deg)	۴۴	WeekDay
۱۹	bearMed(deg)	۴۵	meanTime
۲۰	bearMax(deg)	۴۶	Local Mean Time
۲۱	vel20	۴۷	AccMean(m/s^2)
۲۲	vel40	۴۸	AccStd(m/s^2)
۲۳	vel60	۴۹	AccMed(m/s^2)
۲۴	vel80	۵۰	AccMax(m/s^2)
۲۵	acc20	۵۱	AccSkew
۲۶	bear80	۵۲	Duration(s)

مدل‌های کلاسه‌بندی

پس از پیش‌پردازش‌های مورد نظر و همچنین استخراج ویژگی‌های مورد نیاز، اجرای مدل‌های کلاسه‌بندی صورت پذیرفت. برای اجرای مدل‌های نام‌برده، از کتابخانه‌های Scikit-learn و XGBoost و LightGM موجود در زبان برنامه‌نویسی پایتون استفاده شد و پیاده‌سازی صورت پذیرفت. به این منظور برای هر یک از مدل‌های مورد نظر حالت‌های زیر اجرا شد:

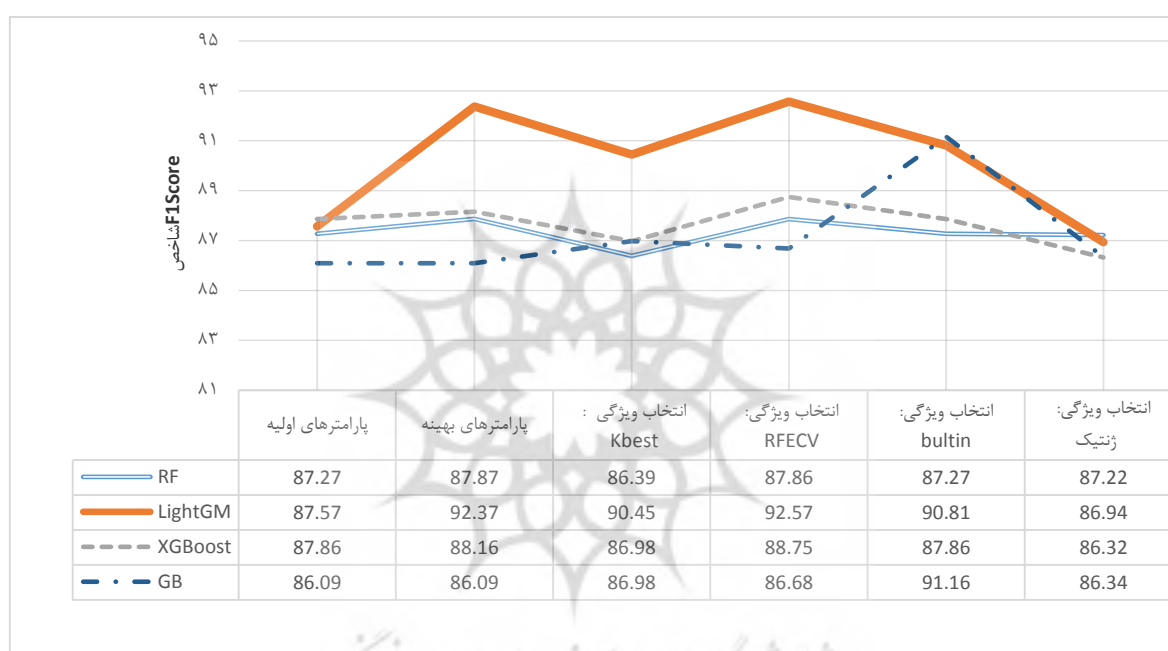
اجرای مدل با استفاده از پارامترهای پیش‌فرض؛

بهینه کردن ابرپارامترها برای هر مدل و اجرای مجدد؛

استخراج ویژگی‌های مؤثر با استفاده از Kbest و اجرای مجدد مدل؛
 استخراج ویژگی‌های مؤثر با استفاده از RFECV و اجرای مجدد مدل؛
 استخراج ویژگی‌های مؤثر با استفاده از Built-in و اجرای مجدد مدل؛
 استخراج ویژگی‌های مؤثر با استفاده از الگوریتم ژنتیک و اجرای مجدد مدل.

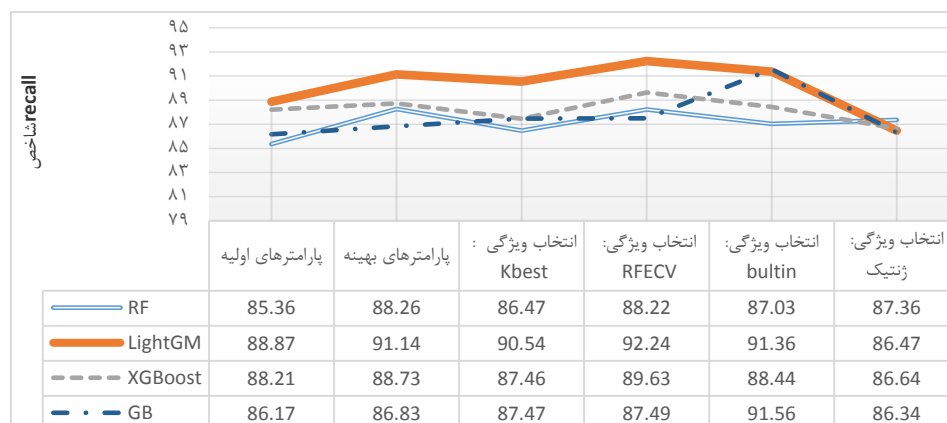
یافته‌های پژوهش

به منظور بررسی و ارزیابی نتایج به‌دست‌آمده برای مدل‌های کلاسه‌بندی از پارامترهای F1-Score و بازخوانی^۱ و دقت^۲ استفاده شد. از آنجا که برچسب حالت‌های سفر برای کلاس‌های موجود به صورت نامتعادل است، ویژگی F1-Score ارزیابی دقیق‌تری را ممکن می‌سازد. در ادامه به تفکیک مدل‌های کلاسه‌بندی مقدار هر یک از پارامترهای گفته‌شده در قالب شکل‌های ۵ تا ۷ به نمایش گذاشته شده است.

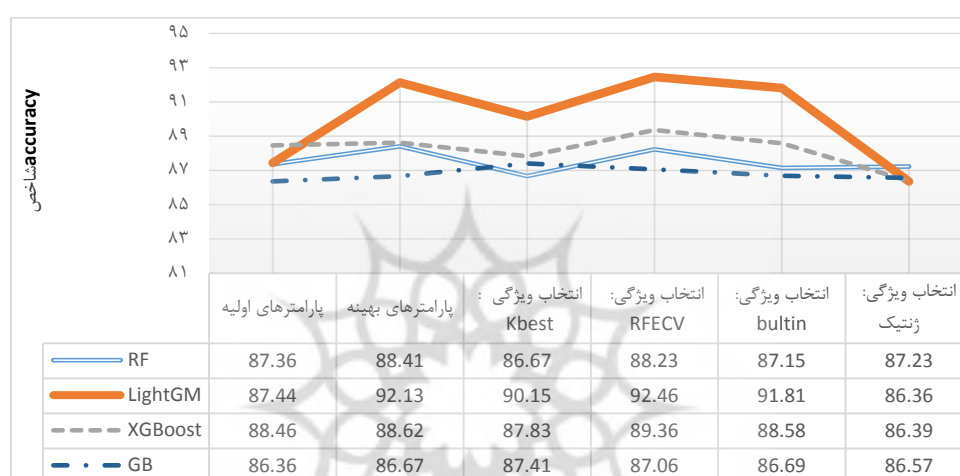


شکل ۵. میانگین وزنی F1-Score در مدل‌های کلاسه‌بندی

همان‌طور که از شکل ۵ مشخص است دقت مدل‌های به‌کاررفته بسیار نزدیک به هم است. ولی مدل LightGM توانسته در استخراج ویژگی‌های مؤثر نسبت به سایر مدل‌ها کارایی بیشتری از خود نشان دهد و با کسب ۹۲/۵۷ درصد بین داده‌های تست در حالت انتخاب ویژگی با مدل RFECV بالاترین مقدار را در این پژوهش کسب کند. این دو شاخص نیز نتایج مشابهی همچون شاخص F1-Score ارائه می‌دهند. مدل LightGM مدل برتر است و پس از آن مدل XGBoost قرار دارد که نتایج بالاتری از دو مدل دیگر کسب کرده‌اند. بدین ترتیب مدل‌های مورد بررسی در این پژوهش توانسته‌اند در پیش‌بینی حالت‌های سفر موفق عمل کنند. برای مقایسه بهتر مدل‌ها و نحوه کارایی هر مدل برای هر یک از کلاس‌ها شاخص F1-Score به تفکیک مدل و حالت انتخابی برای هر یک از کلاس‌ها همراه تعداد ویژگی انتخابی در قسمت انتخاب ویژگی در جدول ۴ نشان داده شده است.



شکل ۶. شاخص Recall در مدل‌های کلاسه‌بندی



شکل ۷. شاخص accuracy در مدل‌های کلاسه‌بندی

جدول ۴. محاسبه میزان F1-Score و ارزیابی هر یک از کلاس‌ها در مدل

مدل کلاسه‌بندی	حالت انتخابی	پایاده	دوچرخه	اتوبوس	اتومبیل	قطار	Weighted	تعداد ویژگی
Random Forest	پارامترهای اولیه	94	80	85	75	93	87	52
	پارامترهای بهینه	94	79	86	77	94	88	52
	Kbest	94	77	85	73	91	86	20
	RFECV	94	79	86	76	95	88	22
	Built-in	94	78	86	76	92	87	32
	GA	94	79	86	74	91	87	20
LightGBoost	پارامترهای اولیه	95	83	86	73	91	87	52
	پارامترهای بهینه	95	88	94	86	94	92	52
	Kbest	93	87	92	83	87	90	19
	RFECV	95	90	93	76	96	92	31
	Built-in	93	86	93	83	94	91	14
	GA	94	77	85	72	90	86	27
XGBoost	پارامترهای اولیه	94	81	87	76	91	88	52
	پارامترهای بهینه	95	84	86	74	93	88	52
	Kbest	95	80	85	72	92	87	25
	RFECV	95	82	87	78	94	89	14
	Built-in	95	84	87	73	91	88	24
	GA	95	79	85	69	87	86	22
GB	پارامترهای اولیه	94	81	83	69	94	86	52
	پارامترهای بهینه	94	81	83	69	94	86	52
	Kbest	95	78	84	71	95	87	26
	RFECV	95	80	85	70	92	87	14
	Built-in	94	83	84	72	93	87	25
	GA	95	78	84	67	89	85	27

در تعیین نوع حالت سفر حالت‌های استفاده از قطار و پیاده با دقت بالاتری شناسایی شدند که می‌تواند ناشی از تمایز سرعت پایین فرد در حالت پیاده با سایر حالت‌ها و در حالت قطار به سبب مستقیم بودن مسیر طی‌شده باشد.

در ادامه برای مقایسه بهتر کارایی هر یک از مدل‌های به‌کارگرفته‌شده، در جدول ۵ تعداد کلاس‌های تخمین‌شده برای هر یک از کلاس‌های موجود آورده شده است.

جدول ۵. بررسی تعداد سفرهای تخمین‌شده توسط هر یک از مدل‌ها

مدل کلاسه‌بندی	حالت انتخابی	پیاده (۴۴۳)	دوچرخه (۱۳۷)	اتوبوس (۴۲۷)	اتومبیل (۱۷۸)	قطار (۱۶۴)	مجموع (۱۳۴۹)
Random Forest	پارامترهای اولیه	417	110	363	134	153	1177
	پارامترهای بهینه	417	108	367	137	154	1183
	Kbest	416	105	363	130	149	1163
	RFECV	415	108	367	135	156	1181
	Built-in	416	107	367	135	151	1176
LightGBBoost	GA	417	108	367	132	149	1173
	پارامترهای اولیه	421	114	367	130	149	1181
	پارامترهای بهینه	421	121	401	153	154	1250
	Kbest	412	119	393	148	143	1215
	RFECV	421	123	397	135	157	1233
XGBoost	Built-in	416	118	397	148	154	1233
	GA	416	105	363	128	148	1160
	پارامترهای اولیه	416	111	371	135	149	1182
	پارامترهای بهینه	421	115	367	132	153	1188
	Kbest	421	110	363	128	151	1173
GB	RFECV	420	112	371	139	154	1196
	Built-in	421	115	371	130	149	1186
	GA	420	108	363	123	143	1157
	پارامترهای اولیه	417	111	354	123	154	1159
	پارامترهای بهینه	417	111	354	123	154	1159
GB	Kbest	420	107	359	126	156	1168
	RFECV	420	110	363	125	151	1169
	Built-in	416	114	359	128	153	1170
	GA	421	107	359	119	146	1152

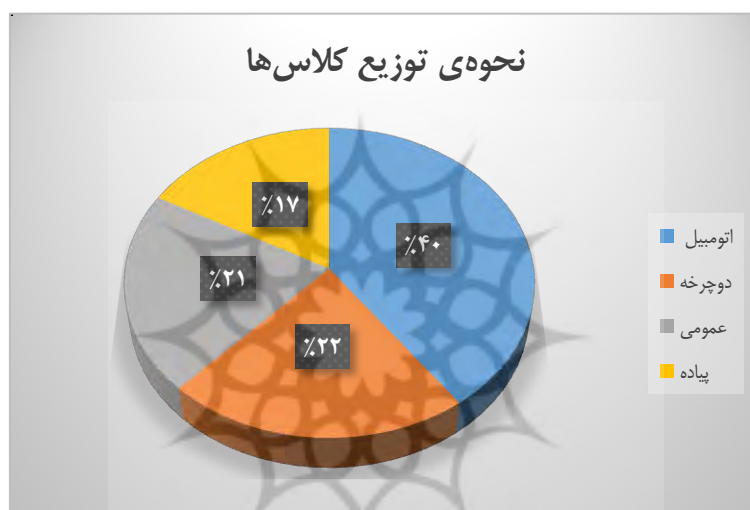
در پایان ارزیابی، دقت حاصل از روش تحقیق این مطالعه با سایر تحقیقات انجام‌شده مورد مقایسه قرار گرفت. در این قسمت هشت مطالعه اخیر که بر داده Geolife صورت گرفته‌اند همراه حالت‌های حمل‌ونقلی که مورد استفاده قرار گرفته است با روش پیشنهادی این مطالعه مقایسه شدند که نتایج حاصل از این مقایسه در جدول ۶ به نمایش گذاشته شده است. همچنین برای اینکه قیاس بهتری بین تحقیقات صورت گرفته شود شاخص وزنی F1-Score در جدول آورده شده است.

جدول ۶. مقایسه F1-Score به‌دست‌آمده با تحقیقات اخیر

منبع	حالت‌های حمل‌ونقلی	F1-Score
Dabiri et al. 2018	دوچرخه، اتوبوس، اتومبیل، قطار، پیاده	۸۴/۶۳
Xiao et al. 2017	دوچرخه، اتوبوس، اتومبیل، قطار، مترو، پیاده	۹۰/۹۲
Wang et al. 2018	دوچرخه، اتوبوس، اتومبیل، تاکسی، قطار، مترو، پیاده	۹۰/۵۱
Huang et al. 2020	دوچرخه، اتوبوس، اتومبیل، مترو، پیاده	۸۱/۴۳
Nawaz et al. 2020	دوچرخه، اتوبوس، اتومبیل، پیاده	۸۳/۷۹
Li et al. 2021	دوچرخه، اتوبوس، اتومبیل، مترو، پیاده	۹۱/۰۵
Li et al. 2020	دوچرخه، اتوبوس، اتومبیل، قطار، پیاده	۸۶/۸۴
Liang et al. 2017	دوچرخه، اتوبوس، اتومبیل، قطار، هواپیما	۸۶/۱۹
LightGM	دوچرخه، اتوبوس، اتومبیل، قطار، پیاده	۹۲/۳۷
با تمام ویژگی‌ها		
LightGM	دوچرخه، اتوبوس، اتومبیل، قطار، پیاده	۹۲/۵۷
با الگوریتم RFECV		
LightGM	دوچرخه، اتوبوس، اتومبیل، قطار، پیاده	۹۰/۴۵
با الگوریتم K best		
LightGM	دوچرخه، اتوبوس، اتومبیل، قطار، پیاده	۹۰/۸۱
با الگوریتم Built-in		

در ادامه برای ارزیابی روش پیشنهادی و کارایی آن لازم است این روش بر یک مجموعه داده دیگر نیز اجرا شود. از این رو ابتدا مجموعه داده دوم به صورت مختصر معرفی و سپس نتایج به نمایش گذاشته می شود.

این مجموعه داده در تاریخ ۲۰۱۷/۰۱/۱۸ تا ۲۰۱۷/۰۹/۱۸ در کشور کانادا، شهر مونترال، بزرگترین شهر استان کبک، توسط آزمایشگاه تحقیقات برنامه حمل و نقل^۱ با همکاری دانشگاه کنکوردیا ثبت و ذخیره شده است. جمع آوری داده ها از طریق یک برنامه قابل نصب بر تلفن همراه هوشمند (در هر دو سیستم عامل IOS و Android) انجام شد که به طور خودکار خط سیر کاربر را ثبت می کند و هنگامی که کاربر در یک مکان معین به مدت بیش از ۱۲۰ ثانیه متوقف شود از کاربر می خواهد اطلاعات سفر خود را تکمیل کند. این اطلاعات شامل برچسب دهی حالت سفر و هدف از سفر است. حالت های حمل و نقلی پس از ترجمه شامل پیاده روی، استفاده از اتومبیل / موتورسیکلت، استفاده از تاکسی، استفاده از حمل و نقل عمومی، استفاده از دوچرخه، و استفاده از اتومبیل های اجاره ای است. در این مطالعه حالت های حمل و نقلی استفاده از تاکسی، استفاده از اتومبیل / موتورسیکلت، و استفاده از اتومبیل های اجاره ای به عنوان حالت استفاده از اتومبیل در نظر گرفته شدند. شکل ۸ نحوه توزیع حالت های حمل و نقلی مورد استفاده در این مجموعه داده را نمایش می دهد.



شکل ۸. نحوه توزیع کلاس ها در مجموعه داده دوم

در این مجموعه داده نیز مانند مجموعه داده اول برچسب حالت های حمل و نقلی نامتعادل توزیع شده است. بدین منظور میزان F1-Score با استفاده از این مجموعه داده به تفکیک مدل های کلاسه بندی RF، LightGBM، XGBoost، و GB در چهار حالت پارامترهای اولیه و پارامتر بهینه و چهار حالت انتخاب ویژگی در جدول ۷ محاسبه شده است.

نتیجه و پیشنهاد

با شناسایی دقیق حالت حمل و نقل هر کاربر، می توان درک واقعی تری از تعداد وسایل نقلیه مورد نیاز برای جابجایی افراد از یک مکان خاص به مکان دیگر در هر روز ارائه داد که کمک بزرگی به کاهش ترافیک و حمل و نقل پایدار می کند. یکی از راهکارهای کاهش ترافیک ایجاد انگیزه در مسافران برای استفاده از حمل و نقل عمومی است. همچنین می توان با ایجاد خطوط پرسرنشین^۲ (HOV) در اتوبان های شهری تمایل به استفاده از اتومبیل های تک سرنشین را کاهش داد.

1. transportation research for integrated planning (TRIP)

2. high-occupancy vehicle

جدول ۷. محاسبه میزان F1-Score و ارزیابی هر یک از کلاس‌ها برای مجموعه داده دوم

مدل کلاسه‌بندی	حالت انتخابی	پایاده	دوچرخه	اتومبیل	عمومی	Weighted	تعداد ویژگی
Random Forest	پارامترهای اولیه	87.92	95.66	91.58	61.7	87.86	52
	پارامترهای بهینه	86.82	96.79	94.02	62.36	88.87	52
	Kbest	84.62	95.66	89.14	60.37	86.85	27
	RFECV	86.82	96.79	90.46	63.03	88.87	26
	Built-in	85.72	94.37	91.03	61.04	87.86	29
	GA	86.82	94.28	90.36	60.37	86.67	23
LightGBoost	پارامترهای اولیه	87.71	96.15	92.72	68.59	91.36	52
	پارامترهای بهینه	88.64	96.24	93.07	69.42	91.51	52
	Kbest	87.17	94.82	91.06	65.58	90.64	24
	RFECV	86.84	95.04	91.74	67.63	90.75	29
	Built-in	86.94	94.57	90.67	66.47	90.34	20
	GA	86.57	95.03	91.45	67.84	90.31	31
XGBoost	پارامترهای اولیه	88.19	95.53	92.74	68.58	91.45	52
	پارامترهای بهینه	90.37	95.68	92.81	70.09	92.67	52
	Kbest	87.1	93.33	87.86	68.33	90.41	23
	RFECV	87.28	93.07	91.18	67.94	90.49	18
	Built-in	87.82	94.48	87.58	68.06	90.57	27
	GA	86.01	90.25	84.2	65.57	86.37	23
GB	پارامترهای اولیه	81.57	83.59	69.49	94.67	86.61	52
	پارامترهای بهینه	82.05	83.94	70.24	94.84	86.91	52
	Kbest	78.55	82.57	68.92	93.68	86.04	28
	RFECV	80.56	82.47	69.81	92.66	85.94	19
	Built-in	80.58	83.04	69.24	93.66	86.31	31
	GA	78.55	82.61	67.48	89.63	84.82	24

چالش‌های متنوعی تا رسیدن به هدف پژوهش وجود دارد. امروزه، به لطف حسگرهای موجود در تلفن‌های هوشمند، داده‌های متنوعی توانایی اخذ و ذخیره شدن دارند. چالش اول انتخاب داده‌های مورد استفاده در پیش‌بینی است؛ به گونه‌ای که علاوه بر در دسترس بودن بالاترین سرعت و کمترین هزینه را داشته باشد. چالش دوم وجود داده‌های پرت و خطا به دلیل وجود خطاهای سیستمی و تصادفی در زمان اخذ داده‌هاست که می‌تواند دقت را به صورت مستقیم تحت شعاع خود قرار دهد. چالش بعدی استخراج ویژگی از مجموعه داده‌هاست. استخراج ویژگی یکی از مراحل مهم در پیش‌بینی است که استخراج ویژگی‌های مهم در آن موجب بهبود عملکرد پیش‌بینی خواهد شد. مهم‌ترین قسمت تحقیق صورت گرفته انتخاب ویژگی‌های مهم و به کارگیری آن‌ها در پیش‌بینی است؛ به گونه‌ای که موجب کاهش هزینه پردازشی در نتیجه افزایش سرعت و بهبود دقت در پیش‌بینی شود. چالش نهایی اجرا و انتخاب یک الگوریتم کلاسه‌بندی نظارت‌شده است که علاوه بر قدرتمند بودن بتوان با سرعت بالاتری نتیجه را پیش‌بینی کرد.

اگرچه در تحقیقات پیشین تمرکز بیشتر مطالعات تنها بر استخراج ویژگی‌ها، انتخاب ویژگی‌ها، و کلاسه‌بندی بوده است، در این تحقیق تلاش شده مقایسه‌ای بر عملکرد روش‌های مختلف در انجام دادن این مهم صورت گیرد. در این تحقیق، از چهار مدل طبقه‌بندی استفاده شده است که مدل Light GM و برخی از روش‌های انتخاب ویژگی برای اولین بار با هدف طبقه‌بندی حالات حمل‌ونقل استفاده شده‌اند. به منظور بهبود جامعیت مقایسه صورت گرفته، تعداد ویژگی‌های بیشتری نیز استفاده شد تا ارزیابی دقیق‌تری بر عملکرد الگوریتم‌های انتخاب ویژگی در زمانی که کاهش ابعاد بیشتری مد نظر بود صورت پذیرد.

نتایج به دست آمده در این پژوهش از دقت مطلوبی در برآورده ساختن اهداف و طبقه‌بندی حالت‌های سفر به منظور حمل‌ونقل پایدار شهری برخوردار است. با وجود این، به دلیل استفاده از تنها یک مجموعه داده که بخش اعظمی از آن در کشور چین تهیه شده است این احتمال که مدل‌های برازش داده شده متناسب با شرایط حمل‌ونقل این کشور آموزش دیده باشند و از قابلیت تعمیم کمی به شرایط موجود در سایر کشورها برخوردار باشند وجود دارد.

با بررسی نتایج حاصل از این مطالعه و مقایسه آن با مطالعات پیشین، دقت محاسبه شده در روش پیشنهادی بهتر از باقی

مطالعات برآورد شده است. با توجه به دقت‌های برآورد شده در مجموعه داده می‌توان پیش‌بینی حالت‌های حمل‌ونقلی با استفاده از تلفن‌های هوشمند را جایگزینی بهتر از پرسشنامه‌ها در نظر گرفت که با هزینه پایین و سرعت بالا حجم عظیمی از داده‌ها را ثبت و ذخیره می‌کنند.

با توجه به اثبات کارآمدی روش ارائه شده پیشنهاد می‌شود در ادامه تحقیق و جهت جمع‌آوری داده‌های بومی هر شهر در کشور، با مراجعه و هماهنگی با سازمان‌های ذی‌ربط، برنامه‌ای در سیستم عامل اندروید تهیه شود و در اختیار کاربران مختلف قرار گیرد. به این صورت که کاربر، در ابتدای مسیر، با روشن کردن موقعیت مکانی تلفن همراه و معرفی شروع سفر برنامه، شروع به ذخیره کردن نقاط مکانی بر حسب بازه زمانی مورد نظر کند و در انتهای سفر هم با پایان یافتن سفر، برنامه برچسب حالت جابه‌جایی را از کاربر بگیرد و ذخیره کند.

همچنین چون مجموعه داده مورد استفاده دارای تعداد زیادی کاربر است که سفرهای خود را برچسب‌گذاری نکرده‌اند، در ادامه تحقیق، توانایی روش‌های نیمه‌نظارت شده مورد ارزیابی قرار گیرد و دقت‌های به دست آمده از این روش را می‌توان با استفاده از داده‌هایی که دارای برچسب حالت سفر هستند، مورد ارزیابی قرار داد.



منابع

- Al Momin, K., Barua, S., Hamim, O. F., & Roy, S. (2022). Modeling The Behavior In Choosing The Travel Mode For Long-Distance Travel Using Supervised Machine Learning Algorithms. *Komunikácie*, 24(4):A187-A197.
- Bantis T. & Haworth, J. (2017). Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics. *Transportation Research Part C: Emerging Technologies*, Vol. 80, 286-309.
- Bohte, W. & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, Vol. 17, No. 3, 285-297.
- Bolbol, A., Cheng, T., Tsapakis, I., & Haworth, J. (2012). Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Computers, Environment and Urban Systems*, Vol. 36, No. 6, 526-537.
- Byon, Y.-J. & Liang, S. (2014). Real-time transportation mode detection using smartphones and artificial neural networks: Performance comparisons between smartphones and conventional global positioning system sensors. *Journal of Intelligent Transportation Systems*, Vol. 18, No. 3, 264-272.
- Chen, T. & Guestrin, C. (2016). "Xgboost: A scalable tree boosting system" in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794.
- Dabiri, S. & Heaslip, K. (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation research part C: emerging technologies*, Vol. 86, 360-371.
- Dabiri, S., Lu, C.-T., Heaslip, K., & Reddy, C. K. (2019). Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 32, No. 5, 1010-1023.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Friedrich, B., Lübke, C., & Hein, A. (2020). Analyzing the importance of sensors for mode of transportation classification. *Sensors*, Vol. 21, No. 1, 176.
- Geolife Dataset. <https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide> (accessed).
- Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*, Vol. 36, No. 2, 131-139.
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques Third Edition [M]. *The Morgan Kaufmann Series in Data Management Systems*, Vol. 5, No. 4, 364-371.
- Hasan, R. A., Irshaid, H., Alhomaidat, F., Lee, S., & Oh, J.-S. (2022). Transportation mode detection by using smartphones and smartwatches with machine learning. *KSCE Journal of Civil Engineering*, 26(8), 3578-3589.
- Huang, Z., Wang, P., & Liu, Y. (2020). Statistical characteristics and transportation mode identification of individual trajectories. *International Journal of Modern Physics B*, Vol. 34, No. 10, 2050092.
- Jahangiri, A. & Rakha, H. A. (2015). Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. *IEEE transactions on intelligent transportation systems*, Vol. 16, No. 5, 2406-2417.
- Kashifi, M. T., Jamal, A., Kashefi, M. S., Almoshaogeh, M., & Rahman, S. M. (2022). Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. *Travel Behaviour and Society*, 29, 279-296.
- Li, J., Pei, X., Wang, X., Yao, D., Zhang, Y., & Yue, Y. (2021). Transportation mode identification with GPS trajectory data and GIS information. *Tsinghua Science and Technology*, Vol. 26, No. 4, 403-416.
- Li, L., Zhu, J., Zhang, H., Tan, H., Du, B., & Ran, B. (2020). Coupled application of generative adversarial networks and conventional neural networks for travel mode detection using GPS data. *Transportation Research Part A: Policy and Practice*, Vol. 136, 282-292.
- Martín-Baos, J. Á., López-Gómez, J. A., Rodríguez-Benitez, L., Hillel, T., & García-Ródenas, R. (2023). A prediction and behavioural analysis of machine learning methods for modelling travel mode choice. arXiv preprint arXiv: 2301.04404.
- Nawaz, A. et al. (2020). Mode Inference using enhanced Segmentation and Pre-processing on raw Global Positioning System data. *Measurement and Control*, Vol. 53, No. 7-8, 1144-1158.
- Nawaz, A., Zhiqiu, H., Senzhang, W., Hussain, Y., Khan, I., & Khan, Z. (2020). Convolutional LSTM based transportation mode learning from raw GPS trajectories. *IET Intelligent Transport Systems*, Vol. 14, No. 6, 570-577.
- Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, Vol. 24, No. 12, 1565-1567.

- Sasaki, Y. (2007). The truth of the F-measure. *Teach tutor mater*, Vol. 1, No. 5, 1-5.
- Sauerländer-Biebl, A., Brockfeld, E., Suske, D., & Melde, E. (2017). Evaluation of a transport mode detection using fuzzy rules. *Transportation research procedia*, Vol. 25, 591-602.
- Schlebusch, C. M. & Jakobsson, M. (2018). Tales of human migration, admixture, and selection in Africa. *Annual Review of Genomics and Human Genetics*, Vol. 19, 405-428.
- Song, X., Kanasugi, H., & Shibasaki, R. (2016). "Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level" in Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2618-2624.
- Sun, Y., Dong, Y. D., Waygood, E. O., Naseri, H., Jiang, Y., & Chen, Y. (2023). Machine-learning approaches to identify travel modes using smartphone-assisted survey and map application programming interface. *Transportation Research Record*, 2677(2), 385-400.
- Vapnik, V. N. (1995). The nature of statistical learning. Theory.
- Wang, B., Wang, Y., Qin, K., & Xia, Q. (2018). "Detecting transportation modes based on LightGBM classifier from GPS trajectory data" in 2018 26th International Conference on Geoinformatics, IEEE, 1-7.
- Xiao, Z., Wang, Y., Fu, K., & Wu, F. (2017). Identifying different transportation modes from trajectory data using tree-based ensemble classifiers. *ISPRS International Journal of Geo-Information*, Vol. 6, No. 2, 57.
- Yazdizadeh, A., Patterson, Z., & Farooq, B. (2019). An automated approach from GPS traces to complete trip information. *International Journal of Transportation Science and Technology*, Vol. 8, No. 1, 82-100.

